IBM Content Collector
Version 2.2

Administrator's Guide

SH12-6914-01

IBM Content Collector
Version 2.2

Administrator's Guide

SH12-6914-01

Note
Before using this information and the product it supports, read the information in "Notices" on page 835.

This edition applies to version 2.2 of IBM Content Collector (product number 5724-V57) and to all subsequent releases and modifications until otherwise indicated in new editions.

This edition replaces SH12-6914-00.

© Copyright IBM Corporation 2008, 2010.
US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

ibm.com and related resources vii
How to send your comments viii
Contacting IBM. viii

Part 1. Solution overview 1

Content Collector overview 3
What's new in Content Collector Version 2.1.1? . . . 6
What's new in Content Collector Version 2.2? . . . 7

Content Collector architecture overview 13
Definition of the storage data model 15

Document archiving scenarios 21
Scenario: Document archiving for storage purposes 21
Scenario: Archiving journal email 22
Scenario: Document retention and disposition . . . 23
Scenario: Preparing the email repository for email analytics 24

Part 2. Installing 27

Installing Content Collector 29
Prerequisites for the installation 29
Hardware prerequisites 29
Software prerequisites 30
Additional prerequisites and restrictions 31
Configuration checklists 34
Configuration checklist for the Content Collector source systems 34
Configuration checklist for Content Collector repository systems 39
Configuration checklist for the Content Collector configuration database. 43
Configuration checklist for the Content Collector connectors. 45
Configuration checklists for the Content Collector general settings 52
Installing Content Collector according to your needs 58
Installing Content Collector for use with one or more source systems and Content Manager. . . 59
Installing Content Collector for use with one or more source systems and FileNet P8 60
Installing Content Collector on several servers - scale out 61
Installation steps. 62
Installing Content Collector Text Search Support 62
Installing Content Collector for Microsoft SharePoint. 63
Installing Content Collector Notes Client Extension 65
Installing Content Collector Server 67
Performing the initial configuration 69
Verifying and adjusting the initial configuration settings. 81
Setting the Content Collector environment variables 83
Installing Content Collector on several servers. . 87
Configuring the Web Application Server. 94
Replacing the Lotus Notes mail template in all mailboxes 106
Installing Content Collector Outlook Extension 107
Enabling offline repositories to allow access to archived content without network access . . . 109
Installing and configuring Content Collector Outlook Web App (formerly Outlook Web Access) support. 111
Starting, stopping, and updating your installed information center. 123

Removing Content Collector 127

Part 3. Migrating 129

Migrating to Content Collector 131
Upgrading to version 2.2 of Content Collector . . 131
Upgrading FileNet P8 task routes 134
Moving from CommonStore to Content Collector 135
Legacy update strategies 137
Installing Content Collector on a separate server 139
Installing Content Collector on the CommonStore servers 140
Installing Content Collector on a new server and manually adding CommonStore functionality. 141
Configuring Content Collector for CommonStore for Exchange Server legacy support 142
Configuring legacy support for documents archived using CommonStore for Lotus Domino 146
Enabling the access to archived data for legacy item types 148
Moving from FileNet Email Manager or FileNet Records Crawler to Content Collector 150
Moving from FileNet Connector for Microsoft SharePoint Document Libraries to Content Collector 151

Part 4. Configuring 155

Configuring Content Collector 157
The Configuration Manager 157
Enabling security in the Configuration Manager 158
Signaling changes to the configuration database 159
Adding, changing, or deleting configuration objects in the Configuration Manager 159

© Copyright IBM Corp. 2008, 2010

Keyboard commands for the Configuration Manager . . . 160
Setting up a configuration database . . . 171
Adding or editing data store connections . . . 171
Deleting a data store connection . . . 173
Exporting or importing a configuration database 173
Starting the Task Routing Engine . . . 173
Configuring the task route service . . . 174
Checking if Content Collector is running . . . 176
Configuring the settings for LDAP lookups during task route processing . . . 177
Content Collector services . . . 178
Configuring connectors . . . 184
Configuring source connectors . . . 184
Configuring target connectors . . . 199
Configuring utility connectors . . . 210
Changing the user account of a service . . . 213
Configuring general settings . . . 220
Configuring Content Collector for CommonStore for Exchange Server legacy support . . . 220
Modifying the Configuration Web Service settings . . . 224
Modifying the information center settings . . . 225
Modifying the settings for the Web Application 225
Modifying client configuration settings . . . 227
Configuring the access to archived data . . . 228
Modifying the settings for the Metadata Web Application . . . 233
Selecting the metadata form template . . . 234
Configuring the metadata form definition . . . 235
Adding and editing metadata . . . 237
Adding, editing and sorting lists . . . 238
Adding and editing user-defined metadata . . 240
Adding and editing file system metadata . . . 241
Selecting system metadata . . . 246
Configuring task routes . . . 270
Task routes . . . 270
Working with task routes . . . 271
Configuring task routes from templates . . . 283
Working with the Expression Editor . . . 404
Using extended processing functions . . . 431
Collecting documents for processing . . . 458
Configuring tasks . . . 489
Using the set-up tools . . . 572
Configuring an IBM Content Manager repository . . . 573
Configuring the Domino environment for Content Collector . . . 577
Enabling a Domino template for archiving . . 578
Enabling an IBM Content Manager repository for processing by the indexer for text search . . 579
Configuring an IBM FileNet P8 repository . . . 580
Enabling the access to archived data . . . 582
Enabling an IBM Content Manager repository for search . . . 583
Enabling an IBM FileNet P8 object store for search . . . 675
About collections . . . 680
Enabling search for email documents . . . 697
Enabling the access to archived data for legacy item types . . . 710
Changing the display format of the date value 712
Setting a default date range for the Email Search page . . . 713
Changing the preview mode for Outlook . . . 713
Handling erroneous documents . . . 713
Blacklist . . . 715
Enabling Microsoft Outlook links . . . 716
Securing Content Collector communications . . . 717
Replacing certificates for the embedded Web Application Server . . . 718
Client communication . . . 720
URL protection . . . 720

Part 5. Tutorials . . . 723

Content Collector file system tutorials 725
Archiving file system documents to FileNet P8 . . 725
Moving documents off the network into IBM FileNet P8 . . . 725
Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records . . . 726
Defining metadata to be used to process files for archiving . . . 729

Part 6. Developing . . . 731

Developing with the IBM Content Collector APIs . . . 733
Developing with the Content Collector Web Application services APIs . . . 733
Creating requests for interactive archiving . . . 734
RestoreAPI . . . 737
ViewingAPI . . . 740
Enabling security for the Web Application services APIs . . . 742
Developing with the Document Viewer . . . 747
The Document Viewer configuration files . . . 747
Document Viewer requests . . . 751
Configuring Workplace or Workplace XT for the use of the Document Viewer . . . 753

Part 7. Monitoring . . . 757

Monitoring Content Collector system performance . . . 759
Using the system dashboard . . . 759
Information monitored in the system dashboard 760
Using performance counters . . . 761
Performance counters . . . 762
Tracking system log files . . . 765
What logs to track . . . 765
File format and naming conventions for system log messages in Content Collector . . . 768
Log levels . . . 770
Using audit logs . . . 770

Using event logs 771
Interpreting event logs 772
Deleting event logs 772
Event IDs 773

Part 8. Using 775

Using Content Collector 777
Searching, viewing, and restoring documents . . . 777
Accessing archived email 777
Changes to archived documents 786
Configuring viewing in IBM FileNet Workplace XT 787
IBM Content Collector Outlook Extension 788
Installing a self-signed certificate for server authentication 788
Archiving message content 790
Restoring archived content 791
Specifying additional archiving information . . 792
Flagging messages for stubbing 792
Specifying configuration options 793
Searching in archives 795
Enabling offline repositories in Microsoft Exchange 796

Part 9. Troubleshooting and support 799

Troubleshooting Content Collector 801
Retrieving version information 801
Troubleshooting installation 801
Troubleshooting scale out mode 801
The installation of the Web applications failed 803
Creating the Content Collector configuration database on remote server fails 803
Using a different DB2 version for the configuration database on a remote server to the version installed on the Content Collector server 803
The connection to the Oracle database fails . . 804
Using DB2 Data Server Client Version 9.5 fix pack 1 or 2 804
Memory issues when running the initial configuration or the set-up tools 805
The CommonStore server and the CSLD tasks fail to start 805
Troubleshooting configuration 806
Troubleshooting source systems 806
Troubleshooting target repositories 813
Troubleshooting components 815
Troubleshooting task routes 818
Troubleshooting the indexer for text search . . . 819
Incorrect Net Search Extender configuration . . 819
The item type cannot be processed by an indexer tool 820
A native DB2 library cannot be loaded 821
DB2 installation directory cannot be located . . 821
No index is found after the index is recreated 822
Net Search Extender stops unexpectedly without issuing a COMMIT 822
Net Search Extender fails while indexer process is running 823
The indexer process stops because the AFUFetchFile UDF cannot be accessed 823
Obtaining the DB2 port number 825
Using SQL statements that are very long or complex 825
Using an incorrect JDBC setup for the indexer processes 826
The JDBC type 2 driver cannot be loaded correctly 826
Items with textual content greater than 1 MB are not indexed when using DB2 V9.5 827
The indexer process stops because the index cannot be found 827
Changes to the default directory and file name settings of the AFUFetchFile UDF configuration . 828
Error accessing XML files in the afuFetchFile UDF 829
Error accessing XML files in the afuFetchFile UDF if the item ID consists of question marks . 830
The indexer worker process stops indicating an out of memory situation 830

Part 10. Appendixes 833

Notices 835

Index 839

ibm.com and related resources

Product support and documentation are available from ibm.com®.

Support and assistance

Product support is available on the web. Simply click Support from the appropriate product website.

IBM® Content Collector
http://www-01.ibm.com/software/data/content-management/content-collector/

IBM Email Archive and eDiscovery Solution
http://publib.boulder.ibm.com/infocenter/email/v2r2m0/index.jsp

IBM CommonStore for Exchange Server
http://www.ibm.com/software/data/commonstore/exchange/

IBM CommonStore for Lotus® Domino®
http://www.ibm.com/software/data/commonstore/lotus/

IBM Content Manager
http://www.ibm.com/software/data/cm/cmgr/mp/

IBM FileNet® P8
http://www.ibm.com/software/data/content-management/filenet-p8-platform/

IBM Enterprise Records
http://www.ibm.com/software/data/content-management/filenet-records-manager/

IBM Records Manager
http://www.ibm.com/software/data/cm/cmgr/rm/

IBM WebSphere® Application Server
http://www.ibm.com/software/webservers/appserv/was/

Lotus Notes® and Domino
http://www.ibm.com/software/lotus/notesanddomino/

Information center

You can view the IBM Content Collector product documentation in an Eclipse-based information center. See the information center at http://publib.boulder.ibm.com/infocenter/email/v2r2m0/index.jsp.

PDF publications

You can view a PDF version of the IBM Content Collector installation and configuration guide by using the Adobe Acrobat Reader for your operating system. The guide is available from the IBM Publications Center. If you do not have the Acrobat Reader installed, you can download it from the Adobe website at http://www.adobe.com.

© Copyright IBM Corp. 2008, 2010

How to send your comments

Your feedback is important in helping to provide the most accurate and highest quality information.

Send your comments by using the online reader comment form at https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf.

Contacting IBM

To contact IBM customer service in the United States or Canada, call 1-800-IBM-SERV (1-800-426-7378).

To learn about available service options, call one of the following numbers:

- In the United States: 1-888-426-4343
- In Canada: 1-800-465-9600

For more information about how to contact IBM, see the Contact IBM website at http://www.ibm.com/contact/us/.

Part 1. Solution overview

© Copyright IBM Corp. 2008, 2010

Content Collector overview

IBM Content Collector archives email and other digitized content in an external, central repository. Additional functions enable users to reduce the size of their mailboxes, reclaim space on their hard drives and Microsoft SharePoint servers, search for content in the repository, and restore archived email to their original locations.

Archiving

You can archive content from various sources. These include:

- Mailboxes on Lotus Domino or Microsoft Exchange servers
- Email that is received through the Simple Mail Transfer Protocol (SMTP)
- Microsoft Exchange public folders and PST files
- Lotus Domino applications and local NSF archives
- Microsoft SharePoint sites
- Documents in NTFS, DFS, and Novell file systems

Archiving means that the content of these documents is processed and then stored in a central repository.

Terminology: IBM Content Collector uses *documents* as a generic term for email, messages, Microsoft SharePoint documents, and file system documents.

The central repository provides a single access point for all business-relevant documents, which means that sensitive data can be better controlled. Various security features are in place for the protection of business documents.

Archiving methods include automatic and interactive archiving.

- Automatic archiving means that an administrator centrally sets up an archiving schedule and selects the sources from which to archive content, such as connected email clients, applications, user groups, servers, Microsoft SharePoint sites, or storage systems.
- Interactive archiving on the client side enables Notes and Outlook client users to flag documents for archiving. Documents flagged by email client users are selected for archiving the next time the scheduled archiving process runs. Users can also specify additional archiving information before the documents are archived.

IBM Content Collector always archives the entire email content, including the attachments. You can configure which parts are removed from the original document after it is archived, and when this happens. You can select documents from all connected mail clients, or from just a subset, according to predefined criteria, such as the size of mail databases, the age of documents, and so on.

You can copy or move documents from multiple Microsoft SharePoint sites, a single site, or selected libraries and supported lists. You can filter the archive collection based on content types or through additional task route filters, and map your custom site columns to corresponding metadata in your repository.

© Copyright IBM Corp. 2008, 2010

Files system documents can be processed depending on metadata and stored in a specific repository folder structure to facilitate search and retrieval.

Accessing Content

The preview and restore functions allow your client users to view and restore archived documents from the central repository, especially in cases where the archived content has been removed from the original documents. Client users can access the archived material through links and hot spots provided in stub documents in the case of email, through direct links in the case of files system or Microsoft SharePoint documents, or through a web-based search interface.

In IBM Content Collector access to archived content is restricted. For email, access to a link is provided by the security of the user's mailbox, meaning the user will see only what the mailbox allows. For file system and Microsoft SharePoint, access to a link is determined by the user's access to the document's location within the file system or SharePoint document library.

Access to document content is also possible when using a repository client, either custom-built or out-of-the-box, where the credentials of a repository user are applied against a document's security to determine access. In IBM Content Collector, file system or SharePoint links can also be defined as secure links. Clicking a secure link prompts the user for specific user permissions to view the document content.

To remove the content of restored email in IBM Content Collector, you can define a schedule. This process is referred to as restubbing.

Search (email)

The installation wizard of IBM Content Collector adds a search interface to the connected Lotus Notes or Outlook clients. From this interface, users can start full-text queries to search for archived content. The content of archived attachments is included in the search.

For security reasons, the search capability is limited. Archive users can only search for content that was archived from their mailbox. They cannot search or restore content that is owned by other users. However, they can search content that was archived from mailboxes to which they have “delegate access”. For example, if an assistant has been given delegate access to a manager's mailbox, the assistant can search for content that was archived from this mailbox. Similarly, users can search the content that was archived from any Microsoft Exchange PST file that was assigned to them before it was archived.

Users can also search for email metadata. This is information which resides in fields of the original email, such as the sender, recipient or subject field. The information in these fields is extracted during the archiving operation, and stored in corresponding fields in the repository. You can customize the list of email fields that you want to extract metadata from. It must be said, though, that metadata searches require the user to have a deeper understanding of the data in these fields.

There is also a preview function. If a document looks promising in the result list, a user can select it to display its content in a web browser window. The search text is highlighted. If the document shows the desired content, users can click a **Restore** button to copy the content to an email document in their mailboxes.

Search (Microsoft SharePoint and file systems)

Users can view stubbed Microsoft SharePoint documents and file-system documents that were archived using the File System Connector by clicking the stub links. They can search for Microsoft SharePoint documents and file-system documents by using the search tools provided for their repository server.

For searching documents in a file-system repository, users can apply the standard search tools provided by Windows.

Document life cycle

IBM Content Collector enables you to implement a range of document retention strategies, from simple deletion after processing to a formal declaration of documents as records in IBM InfoSphere™ Enterprise Records.

You can remove parts of archived email documents or Notes application documents step-by-step from the original document until finally, the entire content is deleted. The removal of document content frees up space in the users' mailboxes or databases, and on the servers of your content management system. Generally, this increases the document throughput significantly.

In Microsoft SharePoint source systems, you can replace entire documents with links to the archived document in the target repository. You can later update outdated links and remove orphan links from target repositories.

To configure the email document life cycle, you can define a so-called stubbing life cycle. Stubbing means converting a document to a stub. A stub is a document from which parts of the content have been removed. For example, your stubbing life cycle might instruct IBM Content Collector to remove email attachments one week after the mail content has been archived. A second instruction in the schedule of the stubbing life cycle removes the main text or email body after four weeks so that just an empty shell of the original email remains. Finally, the stubbing schedule can be set up to delete the entire mail.

The stubbing function can insert links in these stub documents after archiving, which enable users to view the archived content by a mouse click. In addition, IBM Content Collector can be configured to insert brief texts in the original email to indicate that content has been removed, texts that inform users about the archiving of a particular piece of content.

Related concepts

"Content Collector architecture overview" on page 13

"Scenario: Preparing the email repository for email analytics" on page 24

"Scenario: Document archiving for storage purposes" on page 21

"Scenario: Archiving journal email" on page 22

"Scenario: Document retention and disposition" on page 23

Related information

IBM Content Collector website

What's new in Content Collector Version 2.1.1?

IBM Content Collector Version 2.1.1 is an upgrade from IBM Content Collector Version 2.1.0.

IBM Content Collector provides the following new features.

New software platforms

SQL Server 2008

Lotus Domino 8.5 and Lotus Domino 8.5.1

FileNet P8 4.5.1

FileNet RM 4.5

Content Manager 8.4.2 and Content Manager on Linux (x86)

Microsoft SharePoint Collection

You can now collect and archive documents from Microsoft SharePoint sites, enabling you to meet compliance and archiving objectives or simply to relieve the pressure on your SharePoint servers.

Multiple site collection
: You can collect from multiple SharePoint sites, subsites, and libraries, applying different schedules and collecting different metadata and content types for each.

Version collection
: You can fully track document changes by collecting any number of versions of each document.

Post-processing options
: After Content Collector copies documents to your repository, it can retain or delete documents from the SharePoint server, or replace them with links to the repository copies.

Extensions to the Email Connector

Microsoft Exchange 2003/2007 Public Folders
: You can now capture and archive documents stored in Microsoft Exchange Public Folders.

Lotus Domino applications
: In addition to email, you can now capture and archive documents from Lotus Domino applications, for example team rooms, project libraries, address books, and calendars.

Email management enhancements

Records Manager integration
You can now declare collected documents as records in IBM InfoSphere Enterprise Records (formerly IBM FileNet Records Manager).

Microsoft Exchange managed folder support
You can now apply special processing for messages in folders managed by Microsoft Exchange.

Expiration Manager
The Expiration Manager checks a repository for email that is older than the retention date. Email documents can be listed, counted, or deleted.

Collection of additional archiving information
You can enable IBM Content Collector to collect additional archiving information for email that is archived manually.

Storage space improvements
The introduction of the email compound data model greatly improves attachment deduplication by separating email body and metadata information from the embedded attachments.

Support for forms-based authentication (FBA) in Outlook Web Access

For Microsoft Exchange 2007, forms-based authentication is now supported to authenticate the Outlook Web Access Service to the Exchange server.

Read-only access to the Configuration Manager

You can set up access control for the Configuration Manager. The Configuration Manager can run in full mode or in read-only mode, depending on the permissions of the user.

New full text search indexing for Content Manager

The IBM Content Collector Indexer for Text Search replaces and combines the text search user exit and fast indexer modules used for Content Manager in the Content Collector versions before V2.1.1 to form one component that is installed and used in one step. It runs document retrieval and content conversion processes in parallel.

What's new in Content Collector Version 2.2?

IBM Content Collector Version 2.2 provides the following new features.

New software platforms

Microsoft Exchange Server 2010

Microsoft Outlook 2010

Microsoft SharePoint 2010

Configuration Manager

UI updated to the Configuration Manager
The Configuration Manager was redesigned to make it easier and more intuitive to use.

Save incomplete task routes
You can now save task routes that are incomplete or have not been configured completely.

Tabbed display of task routes
In the **Task Routes** view of the Configuration Manager, you can now open and work with more than one task route at a time.

Copy and paste task route including the error task route
You can copy an existing task route and its error task route, and use it as a basis for configuring another task route.

Make the task route active or inactive
As part of the general task route description, you can select to have the task route active or inactive.

Automatic validation of metadata used in a task
The metadata produced in a task is automatically validated. If metadata is missing, this is reported under the properties of the task.

Task route processing

Flexible configuration of audit logs
You can now include audit log tasks anywhere in the task route. The audit log of the error task route is now independent from the audit log of the main task route. All audit logs can be configured independently and log different information.

Archiving of calendar entries
In the email archiving task route templates, calendar entries are now excluded from processing. Instead, new task route templates are available that deal specifically with calendar entries.

Enhanced resilience
The master IBM Content Collector service, the IBM Content Collector Task Routing Engine service automatically restarts any services it notices have stopped. The number of times the IBM Content Collector Task Routing Engine service attempts to restart services is configurable.

Metadata

User-defined metadata
You can now import and export user-defined metadata properties.

MAPI properties and named properties
For Microsoft Exchange, you can now access additional MAPI properties and named properties and use them as metadata in further processing.

Monitoring

System Dashboard
IBM Content Collector now provides a system dashboard which lets you monitor the system performance in a scale-out environment.

IBM Content Manager

Multiple repository support
IBM Content Collector can now archive documents to or retrieve documents from more than one repository in parallel.

Document item types
File System or Microsoft SharePoint documents can now also be stored in an IBM Content Manager document item type.

ACL privilege restricts access
An Access Control List option restricts access to content to a specific list of users.

Check if a document exists
There is a new task in IBM Content Collector that checks whether a document exists in the repository.

IBM FileNet P8

Dynamic document class detection
IBM Content Collector can now dynamically determine the document class for documents that are to be archived.

Check if a document exists
There is a new task in IBM Content Collector that checks whether a document exists in the repository.

Security settings
You can now choose to have Content Collector automatically collect security information about documents in the source system and have them mapped to P8 access rights on the target document. Alternatively, you can manually configure the security settings.

Multiple repository support
IBM Content Collector can now archive documents to or retrieve documents from more than one repository in parallel.

Email data model recovery mechanism
The IBM FileNet P8 Connector has a smoother recovery when using the email data model and unexpected events happen in the repository, such as server outages and network disruptions. In these situations, FileNet P8 tasks are normally unable to perform any rollbacks or cleanup. Now the connector is able to auto-recover data and finish processing email that have been left in the repository in an unexpected state and have reached a critical age. When an unexpected data model state is found, you are alerted in the logs about which emails needed to be repaired, allowing you to choose if you need to manually inspect the results of the repair process.

File System

DFS file system support
This support enables collecting documents stored on a distributed file system (DSF) share.

Novell file system support
This support enables collecting documents stored on a file server in a Local Area Network.

Advanced filtering
You can now specify advanced filtering options for collecting documents from the file system. The options include filtering by various date values, filtering by file sizes, and filtering by file attributes.

Maintain the original file type icon for file shortcuts after archiving
If you archive to IBM FileNet P8 from a file system and stub the

document, the document icon is maintained as it was in the original location. In previous versions of Content Collector, the file shortcut icon was displayed.

Mirror the source file folder structure in the file system repository
You can now mirror the source file folder structure when you archive files into a file system repository.

Microsoft SharePoint

Blogs You can now capture and archive blogs and blog comments.

Wiki page capture
Applications that open captured wiki pages in target repositories can now render the wiki content more accurately.

Post-processing options
The new *Mark as processed* option enables you to archive documents without changing the originals in SharePoint.

More robust version retention enables you to retain only the most recent version of a document, saving space on your SharePoint server.

Simplified installation and configuration
Installation and configuration of Content Collector components on your SharePoint servers and farms is much easier, including the elimination of all other software prerequisites and all extra configuration steps that earlier versions of Content Collector required.

Link management
New link management tasks and task routes enable you to update outdated links and remove orphan links from SharePoint to target repositories.

List columns
You can now map list (library) columns to metadata.

Improved searchability in target repositories
The mirroring of metadata and security settings improve the ability to search archived documents in target repositories.

Improved transparent retrieval from FileNet P8
The Content Collector Web Service now handles retrieval of stubbed documents from IBM FileNet P8, simplifying the installation of the SharePoint component, eliminating the need for authorized users to authenticate to target repositories, and standardizing the handling of IBM Content Manager stubs for SharePoint and all file system stubs.

New task route templates
New task route templates enable you to collect statistics, audit and manage links, classify documents using IBM InfoSphere Classification Module, calculate expiration dates, and declare documents as records in IBM InfoSphere Enterprise Records

New source connector

SMTP Connector
You can now capture and archive documents from SMTP email systems. SMTP archiving can be used for compliance objectives.

Extensions to the Email Connector

Lotus Domino: Collect local NSF files

The administrator can request users in an email to upload their local NSF files to a shared directory. When a user receives the email, the user can click on a button and select the NSF file which is copied to the server where it is captured and archived.

Email management enhancements

Temporarily restore archived content when a stub document is opened

The content of a stubbed document is temporarily retrieved from the repository and displayed when the user opens the document. Users can forward the full content of an archived document instead of the stub document only. The document is not permanently restored, but only available temporarily.

Collection of additional archiving information

For IBM Content Collector Outlook Extension, the window for collecting additional archiving information is now also triggered when you copy or move email to a folder that is configured accordingly.

Specify additional archiving information on a folder level

You can now specify additional archiving information for any folder.

Blacklisting

If a document cannot be processed because it causes errors in any of the Email Connector or SMTP Connector tasks in a task route or even causes processing to be completely halted, the document is added to the *blacklist* to prevent IBM Content Collector from reprocessing this document.

Expiration Manager

Managing retention of documents from Microsoft SharePoint and the file system

The Expiration Manager now also checks a repository for documents from Microsoft SharePoint and the file system that are older than the retention date. Documents can be listed, counted, or deleted.

IBM Content Collector clients

Lotus Domino iNotes (formerly Domino Web Access (DWA))

- Now supports Domino V8.5.1 or above
- Documents that are restored from the search result list can now be stubbed
- You can now suppress adding two attachment stub links by selecting to include the attachment link in the attachment summary only.

Outlook Web App (formerly Outlook Web Access (OWA))

Now supports Outlook OWA 2010

Multiple language support in Lotus Domino

You can now select languages for the Domino template that you enable for archiving using IBM Content Collector.

Content Collector Text Search Support for IBM Content Manager

Run the indexing process in continuous mode
The `afuIndexer` process can be run either in batch mode or in continuous mode where the process continuous until it is stopped by a shutdown command.

Support working with IBM CommonStore item types for the document model BUNDLED and the archiving type ENTIRE (BRI)
CommonStore item types for the document model BUNDLED and the archiving type ENTIRE can be used in Content Collector.

Define your own table spaces
You are no longer restricted to use the default DB2® table spaces when using the indexer tools.

Specify a unique port on the command line
When indexing more than one item type at the same time, a unique port can be defined for each process.

Read user ID and password from file
The user ID and password required to run each indexer tool can be read from a file.

Built-in memory checks
The indexer process checks if the content of the XML files created after document parsing fits the buffer that is input to the indexing engine.

New configuration options
New configuration options are available to disable document conversion and to influence the way documents are processed using a specified codepage.

Additional statistics to monitor indexing process
The processing status of each item can now be monitored.

Search using IBM Content Collector

Searching enabled across more than one repository
You can now search across more than one document repository in IBM FileNet P8 and in IBM Content Manager.

Content Collector architecture overview

IBM Content Collector consists of several components, which interact with components of your Microsoft Exchange, Lotus Domino, NTFS, DFS, and Novell file systems, Microsoft SharePoint environments, and repository servers. See the diagram.

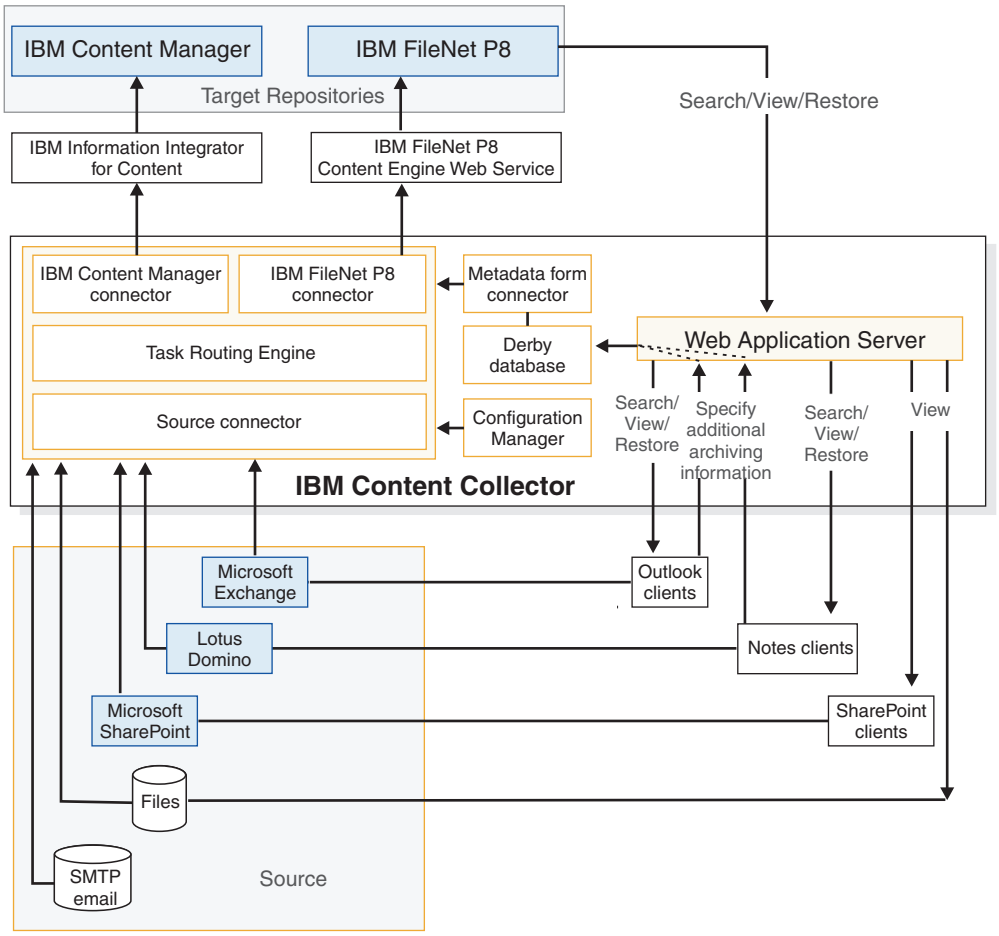


Figure 1. Interaction diagram including IBM Content Collector components, email clients, email servers, Microsoft SharePoint, file systems, and repository servers

Source system

A system that contains documents that you want to work with in IBM Content Collector. This can be Microsoft Exchange, Lotus Domino, SMTP email, NTFS, DFS, and Novell file systems, or Microsoft SharePoint environments.

Source connector

A source connector provides an interface to a third-party system that contains documents that you want to work with in IBM Content Collector. It is responsible for the communication between email servers, file servers, or Microsoft SharePoint and IBM Content Collector. Documents that are

© Copyright IBM Corp. 2008, 2010

routed to IBM Content Collector for archiving pass this layer before they are processed and stored in a repository.

Target connector
A target connector provides an interface to the third-party system that serves as the target repository for IBM Content Collector. It is responsible for the communication between a IBM Content Manager repository, a IBM FileNet P8 repository, or a File System repository, and IBM Content Collector. Documents that are routed from IBM Content Collector for archiving pass this layer before they are stored in a repository.

Task Routing Engine
A service that monitors most of the collector services that run in IBM Content Collector.

Configuration Manager
A graphical user interface for the administration of IBM Content Collector.

Web Application Server
The IBM Content Collector Web Application Server. This can be the embedded Web Application Server or an external WebSphere Application Server.

Metadata Form Connector
A connector to a database where metadata is stored temporarily.

Text Extraction Connector
An interface to the Oracle/Stellent Outside In Technology filters, which are used to convert binary data, for example from email attachments, into a plain-text representation.

Utility Connector
A container for those tasks that provide the intrinsic functions of IBM Content Collector.

Derby database
A temporary storage for any additional archiving information that a user specified when manually archiving a document.

Text Search Support
A component of IBM Content Collector that processes documents archived in a IBM Content Manager repository to enable subsequent searching in these documents. This component must be installed on the IBM Content Manager server. It contains the functionality that converts formatted documents to a text-only representation which is then passed on to the indexing engine called Net Search Extender. After the textual content of a document has been indexed, this content can be searched

Related concepts

"Definition of the storage data model"

"Content Collector overview" on page 3

"Scenario: Preparing the email repository for email analytics" on page 24

"Scenario: Document archiving for storage purposes" on page 21

"Scenario: Archiving journal email" on page 22

"Scenario: Document retention and disposition" on page 23

Related reference

"Additional prerequisites and restrictions" on page 31

Related information

IBM Content Collector website

Definition of the storage data model

The aim of defining an email storage data model in IBM Content Collector to replace the former Bundled Resource Item (BRI) format is to enable effective deduplication on email and email attachments across multiple email sources. Likewise, any Microsoft SharePoint documents that are identified as being identical are stored only once in the repository. The same applies to File System documents.

Improvements to email deduplication comprise:

- Improving the hashing and storing of varying email properties. Only one copy of an email document is stored in the repository regardless if the email source is journal, sent or received, except when Bcc recipients are included in the email document. In this case, two copies of the email document are stored, one copy for the journal and all recipients including Bcc recipients and one copy for the sender. This is because only for the sender the list of Bcc recipients is restored. For each recipient including the Bcc recipients, the restored email document does not contain a Bcc recipient list.
- Extracting the attachments from email that come from multiple email sources and storing those attachments only once in the repository.

  If the same attachment is also ingested as a document through Microsoft SharePoint or File System, no deduplication is provided.

Restriction: Email deduplication happens only if the copies of the email document have the same format. If, for example, the same email document is present in SMTP/MIME format and in a native email server format for Domino or Exchange, no deduplication will occur.

IBM Content Manager data model

All documents archived using IBM Content Collector are stored in item types in IBM Content Manager. You must have at least one IBM Content Manager item type for each source system that you configure in IBM Content Collector. Deduplication on email, Microsoft SharePoint, and File System documents is only available within one and the same item type and not across item types.

Email is stored in an email item type. The email item type is an IBM Content Manager resource item type containing one or more distinct email instances (DEIs).

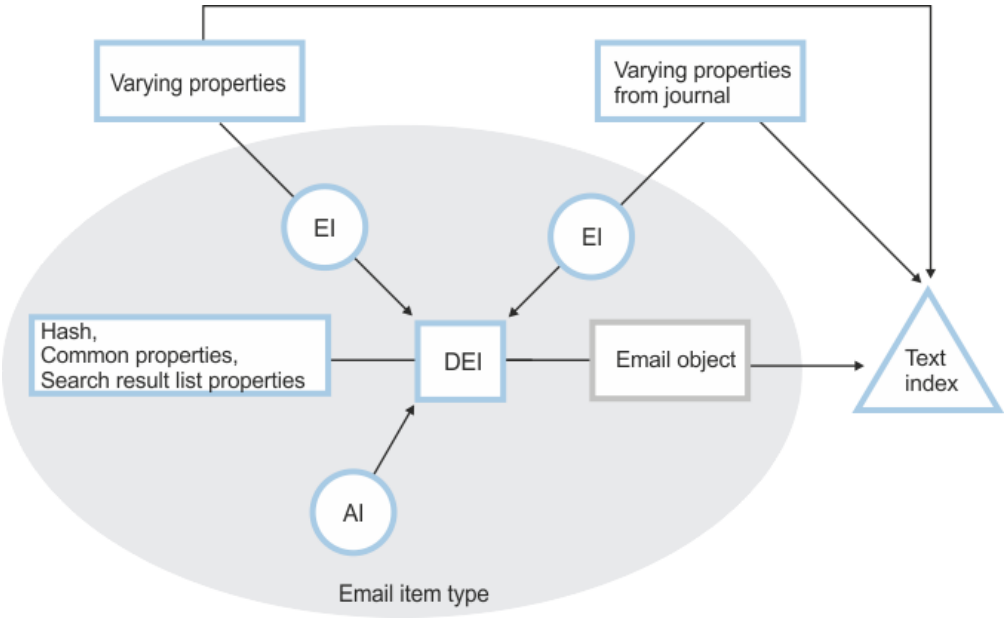
A DEI is the root item and is the common binary email object in one of these formats:

- Notes binary (CSN) format
- Multipurpose Internet Mail Extensions (MIME) format
- Microsoft Exchange mail document (MSG) format

The root holds all common email data and attributes that are shared across all instances of the email. It contains the hash that is used to ensure that the email is stored only once in the repository. The DEI is the item that is required by an application, for example, in a workflow process, for records management or for viewing purposes.

A DEI has two child components:

- The email instance (EI) child that tracks the references of all copies of the same email archived from different mailboxes or the journal. It contains the properties of each email copy which are needed to restore each individual copy of the email, the *varying properties*. For journal archiving, the varying properties contain the additional journal attributes produced during the journal process.
- The attachment instance (AI) child that tracks the references to the email attachments that are archived separately. As an email can have multiple attachments, this reference child can have none, one, or many entries pointing to attachments. Not only are the references to the attachments stored but also additional meta data required for viewing and restoring the email with its attachments, for example, the attachment file name and a correlation key which is used to restore the attachment to the original location in the email.

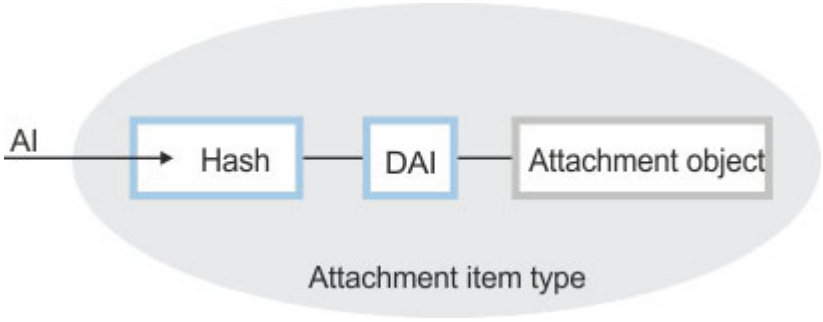


Deduplication is provided by Content Collector.

When a DEI is removed, all associated objects will be removed as well. To prevent accidental deletion of the DEI, for example, by a client user, the expiration date is monitored and only if the current date is past the expiration date, removal is allowed.

Email attachments are stored in an attachment item type. The attachment item type is a resource item type and can contain attachments from different email source system item types. The attachment item type contains one or more distinct

attachment instances (DAIs). A DAI represents the attachment object itself and is the master object that controls the deletion of the associated content and objects. A DAI is referenced by one or more AIs from an email instance (a DEI). A DAI can only be removed if no other instances are pointing to it. The only attribute required by a DAI is the hash used to calculate a unique deduplication hash key that ensures that only one copy of the attachment is kept in one item type, no matter how many times the same attachment was archived by different users.

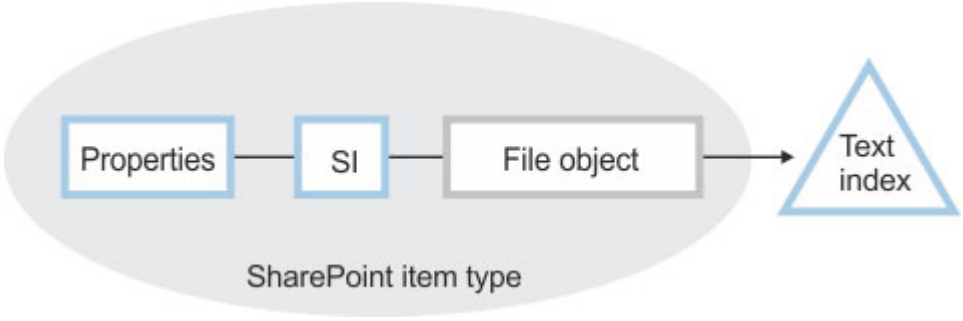


There is no formal data model for Microsoft SharePoint and File System documents. IBM Content Collector offers a sample resource item type for each. You can choose not to use the samples at all or choose to use some of the properties from the samples on a custom item type, depending on your business case.

You can store Microsoft SharePoint documents in the sample Microsoft SharePoint item type, which can be either of the following IBM Content Manager item types:

- A resource item type with one or more Microsoft SharePoint items
- A document item type with one base document part (ICMBASE) with one or more Microsoft SharePoint items. Additional document parts can be associated with the document item type but they are ignored by custom applications or IBM Content Manager clients.

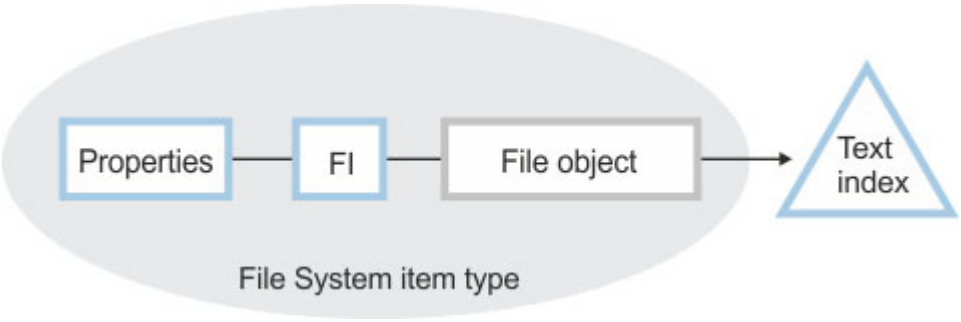
Each item has one instance (SI) that tracks all references archived from Microsoft SharePoint.



File System are stored in a File System item type. The File System item type can be either of the following IBM Content Manager item types:

- A resource item type with one or more File System items
- A document item type with one base document part (ICMBASE) with one or more File System items. Additional document parts can be associated with the document item type but custom applications and IBM Content Manager clients ignore them.

Each item has one instance (FI) that is used to keep track of all references archived from the File System, while only one copy of the actual document is stored in the repository if deduplication was manually configured for the item type in the archiving task route, and the item type is set up accordingly. Otherwise, deduplication is done on the storage device layer and through IBM Tivoli® Storage Manager (TSM).



The source instances can be deleted any time, unless they are also under the control of another application, such as a records management application or eDiscovery Manager, or before their expiration date is reached.

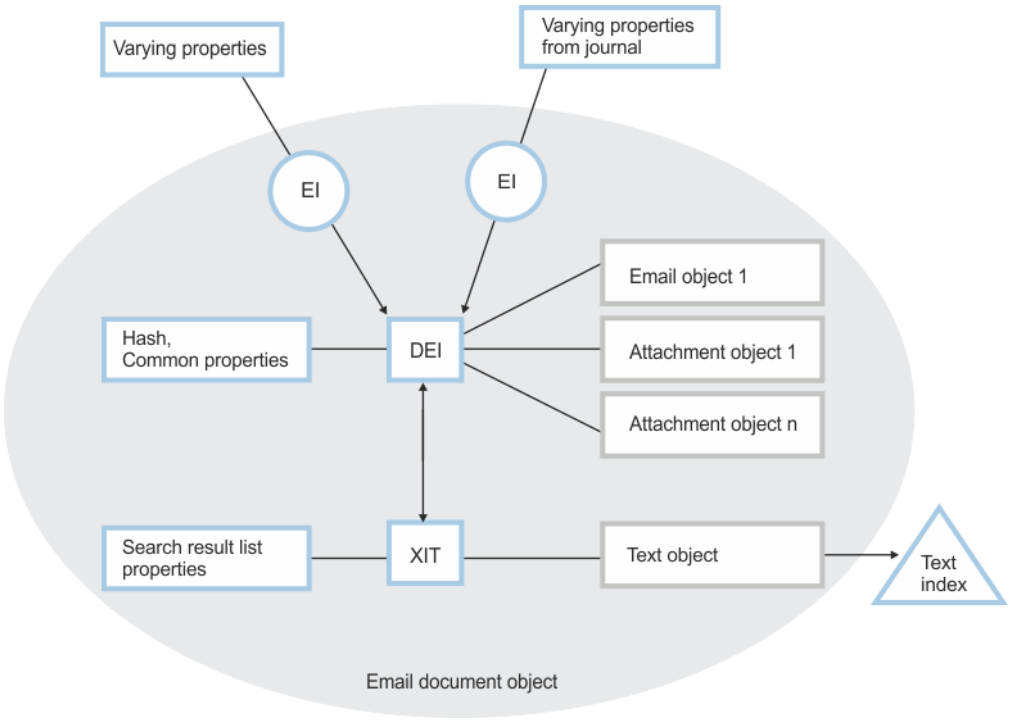
FileNet P8 data model

All documents archived using IBM Content Collector are stored as FileNet P8 document objects in an object store. The object store must be dedicated to archiving using IBM Content Collector. The same object store can be used to store email, Microsoft SharePoint, and File System documents.

An email is stored as a document object and has the following components:

- A distinct email instance (DEI) that is the root document object for the email consisting of one or more content elements:
  - The first content element is the email from different mailboxes or the journal.
  - All subsequent content elements are the attachments.

  The DEI contains the hash that is used to ensure that the email is stored only once in the repository.
- An XML Instance Text (XIT) that is an indexable XML file containing the data of the content elements of the DEI that needs to be indexed for text search.
- An email instance (EI) that is a custom object and tracks the unique data of all copies of an email (mailbox or journal instances of the DEI).



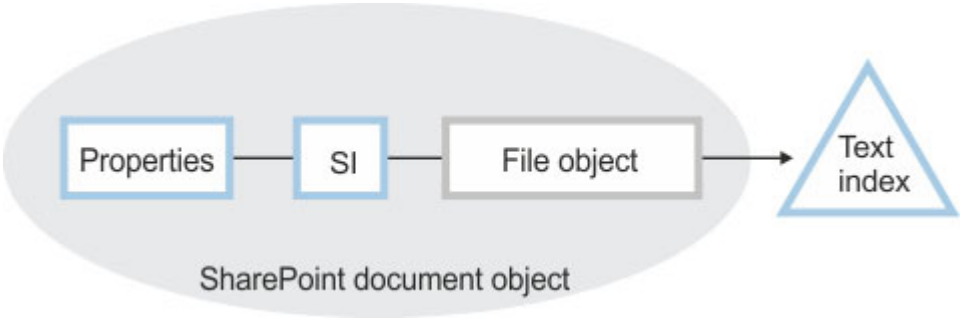
Email deduplication is provided by Content Collector, whereas attachment deduplication is managed by FileNet P8 Version 4.5.1 or on the storage device layer.

When a DEI is removed, all associated objects will be removed as well. To prevent accidental deletion of the DEI, for example, by a Workplace user, the expiration date is monitored and only if the current date is past the expiration date, removal is allowed.

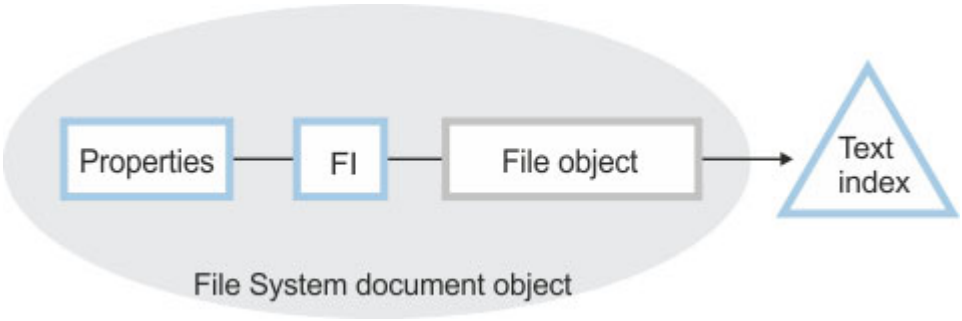
The XIT is also automatically deleted when the DEI is deleted, unless a hold flag is set on it by eDiscovery Manager. In this case, the DEI cannot be deleted. However, if you try to delete the XIT directly, an error is produced and the deletion will not complete. This ensures that the XIT is not accidentally deleted and the indexing lost.

There is no formal data model for Microsoft SharePoint and File System documents. IBM Content Collector offers a sample instance document class for each with the following components. You can choose not to use the samples at all or choose to use some of the properties from the samples on a custom document class, depending on your business case:

- For Microsoft SharePoint:
  - A Microsoft SharePoint instance (SI) that is the root document's object for the Microsoft SharePoint document with a content element which is the document itself. This Microsoft SharePoint instance document class contains probable properties that exist for archiving from this source.



- For File System:
  - A File System instance (FI) that is the root document's object for the File System document with a content element which is the document itself. This instance document class contains probable properties that exist for archiving from a file system.



The source instances can be deleted any time, unless they are also under the control of another application, such as a records management application or eDiscovery Manager.

For deduplication of Microsoft SharePoint or File System content, the system uses the native deduplication that is provided by FileNet P8 Version 4.5.1 and device level deduplication. For earlier versions of FileNet P8, deduplication is also managed by FileNet P8, unless deduplication was manually configured for in the archiving task route.

If deduplication is managed by FileNet P8, a new document object is created in FileNet P8 for each instance of a document. As a result, when browsing or searching, each duplicate will appear to be a unique document, even though at the storage level, only one copy of the file is stored. If deduplication is manually configured for in an archiving task, a new document object is not created in FileNet P8. A single document object represents all instances of a document.

Related concepts

"Content Collector architecture overview" on page 13

Related reference

"Objects deployed to FileNet P8" on page 76

Document archiving scenarios

Document archiving refers to the long-term storage of email and other documents in a central repository, and, in a broader sense, to capabilities of finding, viewing, and restoring archived content. The document archiving scenarios describe how IBM Content Collector helps companies address issues such as storage problems, regulatory compliance, and internal policy compliance. One scenario focuses on the preparation of a repository that is to be used as a knowledge base for analytics with tools such as IBM eDiscovery Manager and IBM eDiscovery Analyzer.

Scenario: Document archiving for storage purposes

This scenario describes how employees in ExampleCo. Enterprises, a fictitious company, address document storage and performance problems on client workstations and email, Microsoft SharePoint, and NTFS file servers.

ExampleCo. Enterprises decides to implement new processes to archive documents because the performance of the company's servers has degraded considerably. The volume of email and SharePoint documents has nearly doubled in the last two years. Email often contains attachments of more than 2 MB in size, so the mailboxes of most users grow rapidly. SharePoint servers can quickly fill with graphics or video files. Users sometimes wait several minutes when they search for email in their own mailbox or documents on the SharePoint server. The documents occupy a lot of disk space on the users' workstations and, more importantly, on the servers.

Increasing the server disk space will not improve, and can degrade, server performance. So ExampleCo. Enterprises decides to use IBM Content Collector to archive documents to a central repository. After copying email and documents to a central repository, the original portions of email documents can be removed from the mail system. This method of storing documents significantly reduces the disk space requirements. Less data needs to be read, scanned, and handled, and the performance of the source systems improves.

The managers discuss the archiving requirements with Judy Jameson, an IT administrator for ExampleCo. Enterprises. Judy implements the following rules and processes:

- Automatically archive email with attachments that are larger than 2 MB one week after their creation or receipt, and all other documents after four weeks.
- Retain documents on the source server for three months, to avoid impeding the work of users who are working offline.
- After three months, remove the SharePoint documents and large email attachments and replace them with links called stubs. Users can follow the links in the stubs to view and restore the documents.
- After one year, remove the stubs from the source servers. Users with access to the target repository can search and restore the documents.
- Email users can manually archive documents at any time.

To meet these requirements, Judy decides to use and modify one or more of the task-route templates that IBM Content Collector delivers. The templates provide an easy way of setting up the system and do not require in-depth system skills. She

© Copyright IBM Corp. 2008, 2010

can adapt the templates to accommodate future needs, but at present she needs to make only minor adjustments to make the templates fit the document management requirements of ExampleCo. Enterprises.

Related concepts

"Collecting documents for processing" on page 458

"Content Collector overview" on page 3

"Content Collector architecture overview" on page 13

Related tasks

"Creating a task route" on page 275

Scenario: Archiving journal email

This scenario describes how ExampleCo. Enterprises, a fictitious company, employs IBM Content Collector to archive email that is journaled by the company's email infrastructure.

For compliance purposes and to avoid accidental or intentional deletion of email, ExampleCo. Enterprises keeps a journal of all incoming and outgoing email. Currently, all email is automatically journaled to a journal mailbox on each of the company's email servers. As it is much easier for the compliance department of ExampleCo. Enterprises to maintain one single journal instead of distinct journals that are located on several email servers at different locations, ExampleCo. Enterprises decides to use IBM Content Collector to create one archive of the journal copies of all email from all email servers, so that they do not need to be retained locally.

The managers ask Judy Jameson, the IT administrator, to investigate the options for archiving journal email. She identifies two possible strategies:

- Configure Content Collector to archive directly from the existing journal mailboxes.
- Configure the email servers to send journal copies of all email to the Content Collector server, so that Content Collector can archive them.

Because the company already uses journal mailboxes, the first option is straightforward to implement. However, Content Collector must crawl the journal mailboxes on each email server to process the journal email. As the email servers are decentralized at different locations, they must be accessed on a wide area network (WAN) instead of a local area network (LAN). This is not advisable and might be very slow. To not risk performance impairment, Judy decides to implement the second option and configure the email servers to send journal copies of all future email to the Content Collector server. Content Collector then collects all journal email that it receives from the different servers and stores them in the same archive.

To receive and process the journal email in Content Collector, Judy configures the SMTP Connector, which receives email through the Simple Mail Transfer Protocol, and sets up a task route to archive the received email. Then she modifies the journaling configuration of each of the company's email servers to deliver the journal email to the Content Collector server through an SMTP connection instead of storing it in a journal mailbox.

Related concepts

"Content Collector overview" on page 3

"Content Collector architecture overview" on page 13

Related tasks

"Creating a task route" on page 275

"Configuring the SMTP Connector" on page 189

"Collecting SMTP documents" on page 478

Scenario: Document retention and disposition

This scenario describes how the real estate firm ExampleCo. Enterprises uses IBM Content Collector to retain and disposition electronic documents.

To avoid the accidental or intentional deletion of documents, the company currently journals its email and backs up every document from its Microsoft SharePoint and file servers. The method works, but not well, because it requires manual disposition of documents and is certain to overwhelm the source servers, degrading performance and eating up more and more disk space. Retrieval of backed up documents is painful if not impossible.

Before they can implement a better solution, ExampleCo. Enterprises must determine the level of control they need over retention life cycles and outcomes. Their records administrator, Alexandra Jackson, informs the chairperson that they need to declare a significant subset of documents as records. The majority of their electronic documents, however, require only simple retention: keep them for three years, then delete.

The company decides to use IBM Content Collector to retain their email and other documents. The application offers two levels of retention:

- The *Calculate Expiration Date* task schedules archived documents for deletion from the repository after a specified number of days that can vary by metadata such as user, group, or an automatic classification that IBM InfoSphere Classification Module supplies.
- The *Declare Record* task hands more complex retention tasks, such as the application of variable retention periods and disposition options, to IBM InfoSphere Enterprise Records.

Because the company requires the basic retention of some documents and the declaration of other documents as records, they decide to use both options.

Alexandra asks Judy Jameson, the IT administrator, to set up IBM Content Collector to declare as records all documents that need to be records, and to archive then delete all other documents after three years. Judy uses the task route templates to create two sets of task routes, one set to process the documents that need to be records, and another to process all other documents. To each task route in the first set she adds a *Declare Record* task that declares each document as a record in IBM InfoSphere Enterprise Records. To each task route in the other set she adds the *Calculate Expiration Date* task and sets it to mark documents as eligible for deletion after three years. She then schedules the Expiration Manager to delete eligible documents each week.

Related concepts

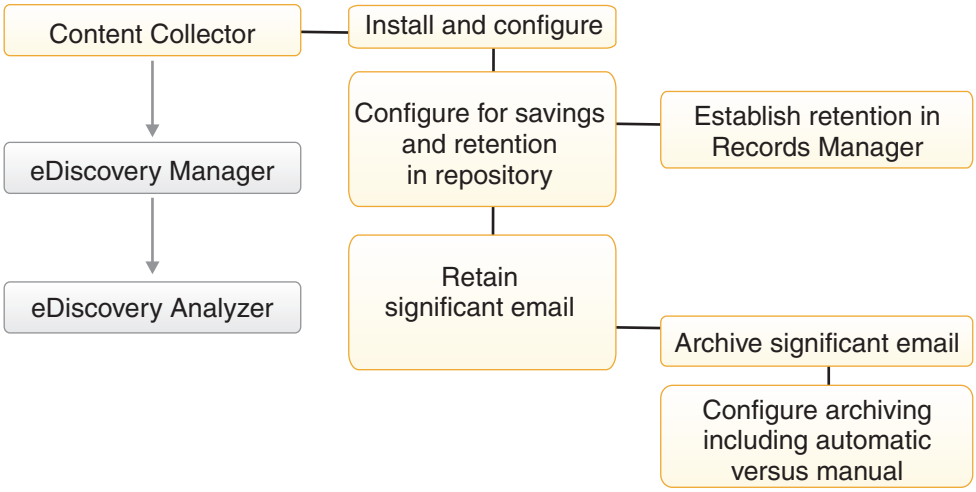
"Collecting documents for processing" on page 458
"Content Collector overview" on page 3
"Content Collector architecture overview" on page 13

Related tasks

"Creating a task route" on page 275

Scenario: Preparing the email repository for email analytics

ExampleCo. Enterprises, a fictitious company that builds electronic appliances, must go to court to contest patent claims by other companies. Email, among other evidence, can prove that ExampleCo. Enterprises is the legal owner of their inventions. This scenario describes how employees in ExampleCo. Enterprises prepare their email archiving system to be able to find email that isrelevant to a lawsuit.



Sometimes, competitors copy ExampleCo. Enterprises inventions illegally. When ExampleCo. Enterprises learns of such a case, it considers a lawsuit against the infringing party or demands compensation. The company needs to prove that it is the legal owner of these innovations. To do that, ExampleCo. Enterprises provides a law firm with blue prints, meeting minutes, product specifications, patents, patent applications, and email that date back to the time when a product was developed. The law firm analyzes the material and, based on the results, tries to negotiate a settlement with the accused party.

The email of the engineers at ExampleCo. Enterprises proves that ideas evolved at ExampleCo. Enterprises before they could have possibly been discussed by the competitor. Of special interest is email of engineers who left ExampleCo. Enterprises to work for the competition. Some of these documents contain hints that a technology was developed when the engineer still worked for ExampleCo. Enterprises and that therefore ExampleCo. Enterprises has the exclusive right to use this technology. In cases like this, the former managers and co-workers of the engineer must be identified so that they can testify if needed.

The information in the email can help the attorneys trace the departments that a person worked for and ensure that they find the right person. For that reason, Chris Marsh, the head of the corporate litigation department, and Alexandra

Jackson, the legal case administrator, want to search for department numbers, for unique employee identifiers, and for the managers of employees and departments. To collect and preserve the email, they use a tool such as IBM eDiscovery Manager, and to analyze the email, they use email analytics tools like IBM eDiscovery Analyzer. To provide the appropriate search results, these tools require additional attributes to be set in the email repository. Chris and Alexandra ask Judy Jameson, the IT administrator, to set up the email repository appropriately.

Judy creates additional attributes in the content management system that is serving as the email repository. These attributes must contain department numbers, identifiers, and manager names. IBM Content Collector will store the information for each email that is archived in the repository. Because the information cannot be found in the email, it will be extracted from the company's Active Directory when the email is archived. Judy also adds the new attribute names to the configuration file for the text-search indexer that is provided by IBM Content Collector. This causes an extraction of the attribute values when the index is built, which adds this information to the text-search index. The information in the text-search index is used by IBM eDiscovery Manager.

Part 2. Installing

© Copyright IBM Corp. 2008, 2010

Installing Content Collector

Install IBM Content Collector according to your requirements. Check the prerequisites before you start the installation.

Related information

System requirements

Prerequisites for the installation

Read the release notes and check the prerequisites on the web and in the following topics before you install IBM Content Collector.

Hardware prerequisites

Check the hardware that you need for IBM Content Collector, for the source systems that contain the documents to be archived, and for the repositories in which you want to archive the documents.

For the most current hardware requirements, see the System Requirements technote on
http://www.ibm.com/support/docview.wss?uid=swg27019702.

In addition, consider the following requirements:

- You need a separate computer or partition that runs on one of the supported Windows operating systems. This computer or logical machine must be connected by a TCP/IP network to the servers on which your repositories and your source systems are installed.

  For Lotus Domino, this computer requires that a Lotus Domino Server runtime is installed.

  For Microsoft Exchange, this computer must be in the same domain as the Microsoft Exchange server.
- Content Collector uses a multiple process architecture to use multiple GB of main memory efficiently. In addition, performance is greatly improved if there is a sufficient amount of memory available for the operating system disk cache.

  The minimum requirement of 4 GB of memory is sufficient for Content Collector servers that are used for basic document collection and archiving. However more memory (the recommended amount is 8 GB) is required:
  - For production servers that are intended for large workloads, in other words, for servers that process many, potentially large documents in parallel
  - For Content Collector servers that also provide search, viewing, and restore services through web applications, for example, in mailbox management scenarios
  - For the servers that run the SMTP Receiver component of the SMTP Connector

  More than 4 GB can efficiently be addressed on systems using a 64-bit version of the operating system.
- The working directory that the email server connector uses to create and store temporary files must be on a separate and fast disk.

© Copyright IBM Corp. 2008, 2010

- Use a Raid 5 array for your operating system and for the Content Collector components. Raid 5 array helps avoid system downtime in the case of hard-disk failures. However, do not use a Raid 5 array for the working directory of the email server connector because of the write penalty of Raid 5.

Software prerequisites

Ensure that you have the necessary software installed at version levels that this release of IBM Content Collector supports.

Check the requirements for the software that you need for IBM Content Collector, for the source systems that contain the documents to be archived, and for the repositories in which you want to archive the documents.

For the most current software requirements, including versions, see the System Requirements technote on
http://www.ibm.com/support/docview.wss?uid=swg27019702.

In addition, consider the following requirements:

- To use Lotus Domino as a source system:
  - Install Lotus Domino Server on the IBM Content Collector server and disable the Lotus Domino service.
  - Install Lotus Domino Client on client machines.
- To use Microsoft Exchange as a source system:
  1. Install Microsoft Outlook, including the latest service packs and patches.
  2. Start Microsoft Outlook and verify its connection to the email server: Create a profile and then log on to Microsoft Exchange with the user that is specified for the user account for the IBM Content Collector Email Connector service.
  3. Make Microsoft Outlook the default email client.
  4. Configure Microsoft Outlook to prompt for a profile every time Outlook is started.
  5. Stop Microsoft Outlook before you install Content Collector Server.
- To use Content Manager as your repository:
  - Install and configure the IBM Information Integrator for Content connector on the server on which Content Collector is to be installed.
  - If you want to search documents that are archived in IBM Content Manager, install IBM Content Collector Text Search Support on the IBM Content Manager server before you install Content Collector. On the Solaris Operating Environment, the text-search component requires the iconv package.
- To use IBM FileNet P8 as your repository:
  - Depending on the version of FileNet P8 that you are using, you need to install different clients and supporting software. See the System Requirements technote.
  - FileNet P8 Content Engine Server must be installed and configured for content-based retrieval (CBR). For further information, see the section on configuring Content Engine for CBR in the FileNet P8 documentation.
  - FileNet P8 Content Engine .NET must be installed to enable communication between the FileNet P8 Content Engine Server, and the IBM Content Collector Configuration Manager and the IBM Content Collector IBM FileNet P8 Connector service. This installer is integrated in the FileNet P8 Content Engine Server installer's .NET Clients option.

– Optional: Install IBM FileNet Enterprise Manager on the machine where the Content Engine Server is installed. This is a sub option under the .NET Clients option in the FileNet P8 Content Engine Server installer.
– Install the FileNet P8 Content Engine Java clients component.
  - For FileNet P8 Content Engine 4.0.1 this is an option of the FileNet P8 Content Engineinstaller.
  - For FileNet P8 Content Engine 4.5.x, you need to run the FileNet P8 Content Engine Client installer. Select **Other Applications** on the FileNet P8 Content EngineClient installer. This will trigger the installation of the FileNet Java Client API libraries. These are required by the Content Collector Initial Configuration, Content Collector Web Services and the Configuration Manager General Settings.

    FileNet P8 Content Engine Client and the FileNet Java Client API must be installed before IBM Content Collector Server is installed. If the FileNet Java Client API is not installed, a warning icon (a red exclamation mark) will be shown in the initial configuration and no FileNet P8 connection can be created.
– Ensure that the installation on the FileNet P8 Content Engine server is always in synch with the installation on the IBM Content Collector server. Note that the Java libraries in `CEClient\lib` are copied to the Content Collector Web Services deployment during installation. If a FileNet Java Client update is required, the newer versions of these files must be copied to `AFUWeb\lib`.
– Install IBM FileNet Content Search Engine on a server other than the IBM FileNet Content Engine server. Note that you cannot collocate the Autonomy K2 Master Administration Server and an additional Autonomy K2 Administration Server on the same server. For further information, see the section on installing Content Search Engine in the FileNet P8 documentation.

• Depending on the type of database that you want to use for the configuration data, you must install different clients and supporting software:

  DB2 database
  Install DB2 Runtime Client on Content Collector Server to establish a connection.

  SQL Server database
  Install a JDBC driver.

  Oracle database
  – Install a JDBC driver
  – Install the Oracle Client tools

• The WebSphere Application Server is required to provide access to configuration data and archived documents. It hosts the Content Collector web applications. If you do not want to use the embedded Web Application Server to store the configuration data, install another WebSphere Application Server or use an existing one. The IBM Content Collector web applications supports the Internet Explorer and Mozilla Firefox web browsers.

Additional prerequisites and restrictions

Consider the listed additional prerequisites and restrictions before you install IBM Content Collector.

Considerations for the source system

Table 1. Considerations for the source system

Source system	Prerequisites and restrictions
Lotus Domino	• If you want to use iNotes (formerly Domino Web Access (DWA)), configure iNotes on Lotus Domino Server. **Important:** For Lotus iNotes®, an Extension Forms File named `Forms85_x.nsf` must exist in the `iNotes` directory on the Lotus Domino server. If the file does not exist, you must create one before you can enable the Content Collector features on Lotus iNotes. For information about how to create an Extensions Forms File, see the topic about customizing the look of Lotus iNotes in the IBM Lotus Domino and Notes information center at http://publib.boulder.ibm.com/infocenter/domhelp/v8r0/index.jsp. • Ensure that the Lotus Domino server that IBM Content Collector archives from is restarted after all enablement for IBM Content Collector has been completed. • When you use IBM Lotus Domino Attachment and Object Store (DAOS) and want to restore your documents back to Lotus Notes, the attachments of the documents are not restored to DAOS.
Microsoft Exchange	• If you run Microsoft Outlook on Citrix, you must publish the application either as installed application that is accessed from a server or as application that is streamed to client. Accessing Outlook as application that is streamed to server is not supported in Content Collector. If you use Citrix streamed to client mode, you need to install Microsoft Visual C++ 2005 Service Pack 1 Redistributable Package on the client machine. • Microsoft Exchange Server 2010 only: Make sure that the client throttling policies that are turned on by default do not unintentionally restrict archiving operations on the Content Collector server as this can lead to a considerable throughput reduction. Either adapt the default client throttling policies or create a tailored throttling policy for the Content Collector archiving user accounts.

Considerations for the target system

Table 2. Considerations for the target system

Target system	Prerequisites and restrictions
Content Manager	• Use only the characters a-z, A-Z, and 0-9 of the Latin-1 character set in the names of the index directory and the index working directory. • To use IBM Content Collector Text Search Support to index and search your documents in a IBM Content Manager repository, these considerations apply: – Install the text-search component and enable the repository for search **before** you install Content Collector Server because the server uses the files and functions that are installed by this component. For more information, see the section on enabling an IBM Content Manager repository for search. – If IBM Content Manager is installed on more than one server, install Content Collector Text Search Support on the IBM Content Manager machine where the library server and Net Search Extender are installed and not where the resource manager is installed. – On Linux, use a shell that uses the `.profile` script. Otherwise, the RC file of the instance owner user ID is not updated. – On Linux and UNIX, the library server name is case-sensitive. If, during the installation of the text-search component, you create a directory in the library server administration directory with the same name as the library server, the name must match with regard to the case. – On Windows, install the text-search component on the server on which DB2 is installed. – Before you install the text-search component and run any of the indexer tools, you must define the environment variable DB2HOME. This environment variable is used to determine Net Search Extender template configuration settings and must point to the DB2 installation directory, for example on Windows, to `C:\Program Files\IBM\sqllib` It is recommended that you define this environment variable permanently on all platforms. – If you install the text-search component on a IBM Content Manager machine where the default DB2 administrator ID is not administrator, but db2admin1 for example, the installation might fail because the Net Search Extender service cannot be stopped. Stop the Net Search Extender service manually before you start installing the text-search component.
IBM FileNet P8	• Ensure that the temporary directory of the Text Extraction Connector is on a separate disk. • A FileNet P8 object store with a file storage area must exist. The object store is enabled for content-based retrieval (CBR). See the section on creating object stores in the FileNet P8 documentation and the topic “Enabling an IBM FileNet P8 object store for search” on page 675 for further information. **Important:** Set up your target object store with a file storage area as the default content store. A file storage area stores content in a network-accessible directory. To prepare your system for index area creation, each file storage area that will be full-text indexed must be accessible by both the Content Engine and the Autonomy K2 server that will perform the full-text indexing. The index area creation is required for email, as well as for File System and Microsoft SharePoint retrieval. For performance reasons it is recommend that the FileNet P8Content Engine has direct access to the file storage area and that the IBM FileNet Content Search Engine accesses this area remotely. Conversely, it is strongly recommended that the IBM FileNet Content Search Engine has direct access to Content Search Engine collection and temporary directories and that the FileNet P8 Content Engine accesses these remotely. • With Content Collector Version 2.2, the Autonomy style sheet was updated for use with FileNet P8 Version 4.5.1 or later. Therefore, you must change any existing index area and re-create the index as described in the documentation for FileNet P8 Content Search Engine.

General considerations

The Java Virtual Machine and several Content Collector components do not accept special characters in the installation path. Use only the characters a-z, A-Z, and 0-9 of the Latin-1 character set.

If you are running antivirus software on the IBM Content Collector Server machine, exclude all temporary working directories from the virus scan. If the antivirus software detects and deletes a virus in a document, Content Collector cannot process the document.

Do not set the environment variable JAVA_HOME on the IBM Content Collector Server machine.

Related concepts

"Content Collector architecture overview" on page 13

Related tasks

"Enabling an IBM Content Manager repository for search" on page 583

"Enabling an IBM FileNet P8 object store for search" on page 675

Configuration checklists

Use the checklists to gather the information that you need to configure IBM Content Collector.

Configuration checklist for the Content Collector source systems

Use the following checklists to gather the information that you need to configure the Content Collector source systems.

Lotus Domino configuration

Table 3. Information for configuring a Lotus Domino source system

Information	Notes	Record your value here
Lotus Domino home server	The name of the Lotus Domino mail server. This server is expected to provide a public Names and Address Book that enables resolving names and addresses used in IBM Content Collector. Example value: `myServer/Organization`	

Table 3. Information for configuring a Lotus Domino source system (continued)

Information	Notes	Record your value here
Administrator ID file and corresponding password	The administrator ID is used by IBM Content Collector to create the runtime environment and enable using the Lotus Domino template and iNotes form with IBM Content Collector functionality. Example value: `admin.id` Typically located at `C:\Lotus\Notes\data\` The selected ID must have sufficient privileges to change templates. Typically, this means Manager access rights. To enable templates remotely, the administrator ID must be Manager on the mail template, the iNotes (Domino Web Access) template, (there is no standalone iNotes template starting with Domino V8) and the forms database. The following rights must be configured irrespective of whether the administrator ID or whether another ID is used to enable the IBM Content Collector template for iNotes (Domino Web Access): • The rights to sign or run unrestricted methods and operations • The rights to sign or run restricted LotusScript/Java agents • The user needs to be editor with remove document access at least in order to use the IBM Content Collector functions on iNotes. The administrator password that you defined when you configured Lotus Domino for IBM Content Collector.	

Table 3. Information for configuring a Lotus Domino source system (continued)

Information	Notes	Record your value here
Domino connector ID file and corresponding password	This user ID is used by IBM Content Collector access and process the mailboxes and databases. Example value: `iccuser.id` Typically located at `C:\Lotus\Notes\data\` The selected ID must be Editor with the additional rights to : • Delete documents • Create personal folders and views • Create shared folders and views • Replicate or copy documents The user password that you defined when you configured Lotus Domino for IBM Content Collector	
Mail database template if Domino server is remote	• The name of the remote Domino server where the Notes templates are located. Example value: `myServer/Organization` • The path to the mail template relative to the Domino data directory on the Domino server. Example value: `mail8.ntf`	
Mail database template if the Domino server is local	The absolute path to the mail template in the Domino data directory on your local IBM Content Collector server. Example value: `C:\Domino\data\mail8.ntf`	
iNotes (Domino Web Access) mail template if the Domino server is local	The absolute path to the iNotes mail template in the Domino data directory on your local IBM Content Collector server. Example value for Domino 7: `C:\Domino\data\dwa7.ntf` Example value for Domino 8: `C:\Domino\data\mail8.ntf`	
iNotes (Domino Web Access) mail template if the Domino server is remote	The path to the iNotes mail template relative to the Domino data directory on the Domino server. Example value for Domino 7: `dwa7.ntf` Example value for Domino 8: `mail8.ntf`	

Table 3. Information for configuring a Lotus Domino source system (continued)

Information	Notes	Record your value here
iNotes (Domino Web Access) forms database if Domino server is local	The absolute path to the iNotes forms database in the Domino data directory on your local IBM Content Collector server, for example, `C:\Domino\data\iNotes\Form8.nsf`	
iNotes (Domino Web Access) forms database if Domino server is remote	The path to the iNotes forms database relative to the Domino data directory on the Domino server, for example, `iNotes\Form8.nsf`	

Microsoft Exchange configuration

Table 4. Information for configuring a Microsoft Exchange source system

Information	Notes	Record your value here
Microsoft Exchange server host name	The fully qualified host name of the Microsoft Exchange mail server Example value: `server.company.com`	

User ID	The user ID must have access to the Active Directory. Use:
	• SMTP format:
	Example value:
	`iccuser@server.company.com`
	• Distinguished name format:
	Example value:
	`CN=iccuser,CN=Users,DC=company,DC=com`
	
	This user ID is used by the IBM Content Collector Email Connector service.	

User password	The user password that you defined when you configured Microsoft Exchange for IBM Content Collector.	

SMTP configuration

Table 5. Information for configuring SMTP

Information	Notes	Record your value here
Message queue directory	The location where received messages are stored temporarily before they are archived. Use the UNC syntax. Example value: `\\ContentCollector\smtp_queue` Select a message queue directory that is available at all times to ensure that email is never lost. If your system is configured to run on several servers, you can use a network share that is hosted on one IBM Content Collector node and is used by all other nodes, or preferably a highly available network share that can be accessed from all nodes and is provided by a network. For best performance, use the latest version of the Windows operating system.	
User ID and password	The ID and password that you want to use in your email system to connect to the SMTP receiver. The SMTP receiver uses SMTP authentication to validate connections. The ID must only consist of letters in the English Alphabet [A-Z,a-z]. Example value: `iccuser`	

File System source configuration

Table 6. Information for configuring a File System source

Information	Notes	Record your value here
User ID	The user who starts the connector service must have permission to access the file shares.	

Microsoft SharePoint configuration

Table 7. Information for configuring Microsoft SharePoint

Information	Notes	Record your value here

User ID	This user must be a member of the SharePoint Site Collection Administrators group for this site.
	
	Optionally, you can add the fully qualified domain to this field, in one of two formats:
	• User principal Name (UPN):
	Example value:
	`iccuser@iccsp.company`
	• Universal naming convention (UNC):
	Example value:
	`iccsp.company\iccuser`	

Password	The password belonging to the User ID.	
Domain	The fully qualified domain name of the Windows server, in the format `domain.com` or `subdomain.domain.com`. Example value: `iccsp.company`	

Microsoft SharePoint site address	Web address of the site collection to process.
	
	Use the value of the site's Load balanced URL field.
	
	Use the following format, where *server* is the server name or virtual server name and *subsite* is an optional sub site:
	`http://server:port/subsite/`
	
	Example value:
	`http://mySPServer:2010/`	

Configuration checklist for Content Collector repository systems

Use the following checklists to gather the information that you need to configure the Content Collector repository systems.

IBM Content Manager server configuration

Table 8. Information for configuring the IBM Content Manager pane

Information	Notes	Record your value here
IBM Content Manager server	Library server database name on which you want to create the document repository. Default value: ICMNLSDB	
Administrator ID	The administrator user name that you defined when you configured IBM Content Manager for IBM Content Collector. Administrator authority is required to create and configure a repository. Example value: ICMADMIN If you are working with more than one Content Manager server, the administrator ID should have synchronized credentials across all servers.	
Administrator password	The administrator password that you defined when you configured IBM Content Manager for IBM Content Collector	
User ID	The user name that you defined when you configured IBM Content Manager for IBM Content Collector. This user ID is used by the IBM Content Manager connector. For the require access rights, see “Required Content Manager privileges for the connector” on page 201. Additionally, the user must be a member of the DB2USERS group and be granted DBADM authority to the library server database. This is required to set constraints on the DB2 database when archiving email to multiple item types. Example value: ICCUSER	
User password	The user password that you defined when you configured IBM Content Manager for IBM Content Collector.	
Text Search Support directory	The absolute path to the directory in which the IBM Content Collector Text Search Support component is installed on the IBM Content Manager server. Example value: `C:\Program Files\IBM\ICCIndexer`	

IBM Content Manager item type configuration

Table 9. Information for configuring the item type panes

Information	Notes	Record your value here
Source system item type	Specify at least one item type for each source system Default values for compound email item types: • ICCEmailCmpLD • ICCEmailCmpEX • ICCSMTPCmp Default values for bundled email item types: • ICCEmailLD • ICCEmailEX Default values for item types that are not email specific: • ICCFilesystem • ICCSharepoint	
Attachment item type	Default value: ICCAttachments	
Index directory for the item type	The absolute path to the directory on the Content Manager Library Server machine where the full text index is to be created for the item type. The directory must exist, must have the adequate access rights (user and group rights), and sufficient space for the index. Example value: `C:\NSE\indexes`	
Index working directory for the item type	The absolute path to the working directory of the full text index on the Content Manager Library Server machine. The directory must exist, must have the adequate access rights (user and group rights), and sufficient space for the index. Example value: `C:\NSE\work` Both the index and the index working directories should be on the same volume.	

IBM Content Manager Net Search Extender configuration

Table 10. Information for configuring the IBM Content Manager text search support in IBM Content Collector

Information	Notes	Record your value here
User ID	The user ID that runs all the indexer tools. This correlates with the DB2 instance owner user ID.	
XML file path to specify in the **GenXMLDirectory** configuration option	The absolute file path to the directory in which the indexer XML files are stored. This value MUST be identical for all item type configurations. Example value: `C:\ICCIndexer\XML`	
Log file path to specify in the **GenLogDirectory** configuration option	The absolute file path to the directory in which the indexer log files are stored. This value MUST be identical for all item type configurations. Example value: `C:\ICCIndexer\log`	
Dump file path to specify in the **GenDumpDirectory** configuration option	The absolute file path to the directory in which the original documents retrieved from the IBM Content Manager Resource Manager are stored. It is recommended that this value is identical for all item type configurations. Example value: `C:\ICCIndexer\dump`	
Work file path to specify in the **GenWorkingDirectoryForItemTypes** configuration option	The absolute file path to the directory in which the working files required by the indexer process are written. It is recommended that this value is identical for all item type configurations.	

IBM FileNet P8 configuration

Table 11. Information for configuring the IBM FileNet P8 panes

Information	Notes	Record your value here
URL to the IBM FileNet P8 Content Engine	The URL is made up of: • A connection type • A server name • A port • A path Example value: `http://server.company.com:9080/wsi/FNCEWS40MTOM/`	

Table 11. Information for configuring the IBM FileNet P8 panes (continued)

Information	Notes	Record your value here
Administrator ID	The administrator ID must be defined as the GCD administrator with the permission to create object stores, file storage areas, content cache areas, and related actions like deleting and moving. Example value: ceadmin If you are working with more than one FileNet P8 server, the administrator ID should have synchronized credentials across all servers.	
Administrator password	The administrator password that you specified when you configured IBM FileNet P8 for IBM Content Collector.	
User ID	The user ID must be able to access the object store and modify properties. Example value: iccuser For the required access rights, see "Required IBM FileNet P8 privileges for the connector" on page 207.	
User password	The task user password that you specified when you configured IBM FileNet P8 for IBM Content Collector.	
Object store	The name of the object store that is configured for IBM Content Collector **Important:** The object store must be set up with a file storage area as the default content store. Each file storage area that will be full-text indexed must be accessible by both the Content Engine and the Autonomy K2 server that will perform the full-text indexing.	

Configuration checklist for the Content Collector configuration database

Use one of the checklists to gather the information that you need to configure the database management system that you want to use for the Content Collector configuration database.

DB2 configuration

Table 12. Information for configuring the Content Collector configuration database in DB2

Information	Notes	Record your value here
DB2 server name	The fully qualified host name of the server on which the database resides or is to be created. Example value: `server.company.com`	
DB2 port	Default value: 50000	
Database name	Default value: ICCDB	
DB2 administrator ID	Default value: DB2ADMIN The DB2 administrator must be allowed to create databases.	
Administrator password	The DB2 administrator password that you defined when you configured DB2 for IBM Content Collector.	
DB2 user ID	Default value: DB2USER The DB2 user must be allowed to create, read, and write to tables.	
User password	The DB2 user password that you defined when you configured DB2 for IBM Content Collector.	

SQL Server configuration

Table 13. Information for configuring the Content Collector configuration database on SQL Server

Information	Notes	Record your value here
SQL Server name	The fully qualified host name of the server on which the database resides or is to be created. Example value: `server.company.com`	
SQL server port	Default value: 1433	
Database name	Default value: ICCDB	
JDBC driver	Default value: `C:\Program Files\Microsoft SQL Server 2005 JDBC Driver` for SQL Server 2005 and `C:\Program Files\Microsoft SQL Server 2008 JDBC Driver` for SQL Server 2008	
SQL Server administrator ID	Default value: SA The SQL Server administrator must be allowed to create databases.	
Administrator password	The SQL Server administrator password that you defined when you configured SQL Server for IBM Content Collector.	

Table 13. Information for configuring the Content Collector configuration database on SQL Server (continued)

Information	Notes	Record your value here
SQL Server user ID	Default value: ICCUSER The SQL Server user must be allowed to create, read, and write to tables.	
User password	The SQL Server user password that you defined when you configured SQL Server for IBM Content Collector.	

Oracle configuration

Table 14. Information for configuring the Content Collector configuration database on Oracle

Information	Notes	Record your value here
Oracle server name	The fully qualified host name of the server on which the database resides or is to be created. Example value: `server.company.com`	
Oracle port	Default value: 1521	
Database service name	Default value: ICCDB	
JDBC driver	Default value: `C:\Program Files\Oracle\ojdbc14.jar`	
Oracle user ID	Default value: ICCUSER The Oracle user must be allowed to create, read, and write to tables.	
User password	The Oracle user password that you defined when you configured Oracle for IBM Content Collector.	

Configuration checklist for the Content Collector connectors

Use the checklists to gather the information that you need to configure the Content Collector connectors.

Email Connector for Lotus Domino configuration

Table 15. Information for configuring the Email Connector for Lotus Domino

Information	Notes	Record your value here
Lotus Domino notes.ini location	The absolute path to the `notes.ini` file. The `notes.ini` is used by the Email Connector to authenticate against the Domino environment. It contains information about the administrator user ID which is the ID that is used by the connector to open and work with mailboxes. The `notes.ini` file is created during the initial configuration. Example value: `C:\Program Files\IBM\ContentCollector\ notesdata\admin.ini`	
Password	The administrator password that you defined when you configured Lotus Domino for IBM Content Collector.	
Server	The name of the Domino server where the address book is located. Example value: `myServerName/ OrganizationRequired` only if the default address book is not used.	
Database	The file name of the address book. Example value: `domino\ mailbox.nsfRequired` only if the default address book is not used.	

Email Connector for Microsoft Exchange configuration

Table 16. Information for configuring the Email Connector for Microsoft Exchange

Information	Notes	Record your value here
Microsoft Exchange server host name	The fully qualified host name of the Microsoft Exchange mail server. Example value: `server.company.comRequired` only if the domain default is not used.	

Table 16. Information for configuring the Email Connector for Microsoft Exchange (continued)

Information	Notes	Record your value here

User ID	The user ID must have access to the Active Directory. Use:
	• SMTP format:
	Example value:
	`iccuser@server.company.com`
	• Distinguished name format:
	Example value:
	`CN=iccuser,CN=Users,DC=company,DC=com`
	This user ID is used by the IBM Content Collector Email Connector service.	

User password	The user password that you defined when you configured Microsoft Exchange for IBM Content Collector.	
Host name	The fully qualified host name of the Microsoft Exchange Active Directory server. Example value: `server.company.com`Required only if the domain default is not used.	
LDAP port	Default value: 389 Required only if another global catalog server than the domain default is used.	
Global catalog port	Default value: 3268 Required only if another global catalog server than the domain default is used.	

SMTP Connector

Table 17. Information for configuring the SMTP Connector

Information	Notes	Record your value here
Message queue directory	The location where received messages are stored temporarily before they are archived. It is recommended to use the UNC syntax. Example value: `\\ContentCollector\smtp_queue` Select a message queue directory that is available at all times to ensure that email is never lost. If your system is configured to run on several servers, you can use a network share that is hosted on one IBM Content Collector node and is used by all other nodes, or preferably a highly available network share that can be accessed from all nodes and is provided by a network.	
User ID	The ID that you want to use in your email system to connect to the SMTP receiver. The SMTP receiver uses SMTP authentication to validate connections. The ID must only consist of letters in the English Alphabet [A-Z,a-z]. Example value: `iccuser`	
Password	The password that you define for the user.	

SharePoint Connector

Table 18. Information for configuring the SharePoint Connector

Information	Notes	Record your value here
User ID	The user name used to authenticate to the Microsoft SharePoint site collection. Example value: `iccuser`	
Password	The user password used to authenticate to the Microsoft SharePoint site collection.	
Domain	The user domain used to authenticate to the Microsoft SharePoint site collection.	
SharePoint site address	The URL to the Microsoft SharePoint site collection or subsite; if you are setting up multiple Content Collector nodes, you must use the UNC format.	

Table 18. Information for configuring the SharePoint Connector (continued)

Information	Notes	Record your value here
Log file location	Default value: • 32-bit system: `C:\Program Files\IBM\Content Collector\ctms\Log` • 64-bit system: `C:\Program Files (x86)\IBM\Content Collector\ctms\Log`	
Temporary directory	Default value: • 32-bit system: `C:\Program Files\IBM\ContentCollector\ctms\temp` • 64-bit system: `C:\Program Files (x86)\IBM\ContentCollector\ctms\temp`	

IBM Content Manager Connector configuration

Table 19. Information for configuring the IBM Content Manager Connector and connections

Information	Notes	Record your value here
User name	The user name that you defined when you configured IBM Content Manager for IBM Content Collector. Example value: ICCUSER **Important:** The user must have a minimum set of IBM Content Manager privileges. See "Required Content Manager privileges for the connector" on page 201 for details. Additionally, the user must be a member of the DB2USERS group and be granted DBADM authority to the library server database.	
User password	The user password that you defined when you configured IBM Content Manager for IBM Content Collector.	
Repository	Library server database name. Default value: ICMNLSDB	
Default item type lookup	If you do not use the default item-type names, you must map the actual names to the default names.	

IBM FileNet P8 Connector configuration

Table 20. Information for configuring the IBM FileNet P8 Connector and connections

Information	Notes	Record your value here
Server	The fully qualified host name of the Content Engine server. Example value: `server.company.com`	

Table 20. Information for configuring the IBM FileNet P8 Connector and connections (continued)

Information	Notes	Record your value here
Port	Default value: 9080	
Path	The web services end point for the Content Engine URL. Example value: `wsi/FNCEWS40MTOM`	
User name	The task user name that you specified when you configured IBM FileNet P8 for IBM Content Collector. For the require access rights, see "Required IBM FileNet P8 privileges for the connector" on page 207.	
User password	The task user password that you specified when you configured IBM FileNet P8 for IBM Content Collector.	
Object store	The object store used with the connection.	

Metadata Form Connector configuration

Table 21. Information for configuring the Metadata Form Connector

Information	Notes	Record your value here
Server host name of the temporary metadata database	The fully qualified host name of the server on which the Derby database is located. The Derby database is the repository that is used for temporarily storing any additional archiving information that a user specified when manually archiving a document. Example value: `server.company.com`	
Port	Default value: 1527	
User ID	The user ID must be listed in the `derby.properties` file in `<InstallDir>:\ContentCollector\derby\10.3.3.0\bin`. Use the following template: • Example value: `derby.user.iccuser` where `iccuser` is a listed user ID	
User password	The password for the specified user ID with which to access the Derby database.	
Log file location	Default value: `C:\Program Files\IBM\Content Collector\ctms\Log`	

File System Source Connector configuration

Table 22. Information for configuring the File System Source Connector

Information	Notes	Record your value here
User ID	The user who starts the connector service must have permission to access the file shares.	
Log file location	Default value: `C:\Program Files\IBM\Content Collector\ctms\Log`	
Temporary directory	Default value: • 32-bit platforms: `C:\Program Files\IBM\ContentCollector\ctms\temp` • 64-bit platforms: `C:\Program Files (x86)\IBM\ContentCollector\ctms\temp`	

File System Repository Connector configuration

Table 23. Information for configuring the File System Repository Connector

Information	Notes	Record your value here
Log file location	Default value: `C:\Program Files\IBM\Content Collector\ctms\Log`	
Repository name	Example value: `ICCFileRep`	
Document class display name	Example value: `ICCFiles`	
Document class system name	Example value: `ICCFiles`	

Text Extraction Connector configuration

Table 24. Information for configuring the Text Extraction Connector

Information	Notes	Record your value here
Log file location	Default value: `C:\Program Files\IBM\Content Collector\ctms\Log`	

Utility Connector configuration

Table 25. Information for configuring the Utility Connector

Information	Notes	Record your value here
Log file location	Default value: `C:\Program Files\IBM\Content Collector\ctms\Log`	
For LDAP lookup settings	See "Configuring the Utility Connector" on page 212	

Configuration checklists for the Content Collector general settings

Use the following checklists to gather the information that you need to provide when you configure the options under **General Settings** in the IBM Content Collector Configuration Manager.

Outlook Web App (OWA, formerly Outlook Web Access) Service configuration

Table 26. Information for configuring the OWA Service server

Information	Notes	Record your value here
Host name	The fully qualified name of the Internet Information Service server on which the OWA Service is installed. Example value: `iccowa.domain.company.com`	
User ID	The ID of a registered Exchange user. The ID must have administrator privileges.	
Password	The user password that you defined when you configured Microsoft Exchange for IBM Content Collector.	
Host name of the Active Directory	The fully qualified host name of the Microsoft Exchange Active Directory server used to resolve the network resources. Example value: `adserver.company.com`	
LDAP port	The port number for network communication with the Active Directory server using the LDAP protocol. Default value: 389	
Global catalog port	The port number for network communication with the Active Directory server using the LDAP protocol. Default value: 3268	

Table 26. Information for configuring the OWA Service server (continued)

Information	Notes	Record your value here

User ID	The ID is used by the OWA Service to access the Active Directory server and must be a registered Exchange Server user ID. If the Active Directory cannot be accessed by using the SMTP address, use the user distinguished name:
	• SMTP format:
	Example value:
	`iccuser@server.company.com`
	• Distinguished name format:
	Example value:
	`CN=ICCConnectorUser,CN=Users,DC=company,DC=com`	

Password	The user password that you defined when you configured Microsoft Exchange for IBM Content Collector.	
Enable tracing	This option is used to enable or disable tracing in OWA.	
Trace file name	The fully qualified file path to the trace file. If the entry field is left empty, the default is taken. Default value: `<installation_path>\logs\afuowa.trc`	

Outlook Web App (OWA, formerly Outlook Web Access) Extension configuration

Table 27. Information for configuring the OWA Extension

Information	Notes	Record your value here
Host name	The fully qualified host name of the Microsoft Exchange server that provides the IBM Content Collector buttons for the OWA users. Example value: `iccowa.domain.company.com`	
Host name	The fully qualified host name of the OWA Service server. Example value: `iccowa.domain.company.com`	
Port	The port number for communication with the OWA Service server. Default value: 443	

Legacy Restore Exchange configuration

Table 28. Information for configuring Legacy Restore

Information	Notes	Record your value here
Legacy task definition file	Browse for the legacy IBM CommonStore for Exchange Server task definition file you want to work with. After you import the task definition file, the configuration settings in the definition file are displayed in the tab pages under **General Settings → Legacy Restore Exchange**. If you have made changes to your IBM CommonStore for Exchange Server configuration, you will need to adjust settings that are taken from the definition file when you import the file. For details, see "Settings to enable restoring documents archived using CommonStore for Exchange Server" on page 143.	

The IBM Content Collector Configuration Web Service configuration

Table 29. Information for configuring the IBM Content Collector Configuration Web Service

Information	Notes	Record your value here
Host name	The fully qualified host name of the server on which the Configuration Web Service runs. Example value: `server.company.com`	
Port	The port number for communication with the Configuration Web Service server over an HTTPS connection. Default value: 11443	
JDBC driver directory	The absolute path to the JDBC driver. Example value for DB2: `C:\Program Files\IBM\SQLLIB\java` Example value for SQL Server: • SQL Server 2005: `C:\Program Files\Microsoft SQL Server 2005 JDBC Driver` • SQL Server 2008: `C:\Program Files\Microsoft SQL Server 2008 JDBC Driver` Example value for Oracle: `C:\Program Files\Oracle`	

Table 29. Information for configuring the IBM Content Collector Configuration Web Service (continued)

Information	Notes	Record your value here
JDBC port	The port number for communication with the JBDC driver. Default value for DB2: 50000 Default value for SQL Server: 1433 Default value for Oracle: 1521	
Database server	The fully qualified host name of the server on which the IBM Content Collector configuration database is located.	

The IBM Content Collector information center configuration

Table 30. Information for configuring the information center

Information	Notes	Record your value here
Host name	The fully qualified host name of the server where the IBM Content Collector information center is to run. Example value: `server.company.com`	
Port	The port number for communication with the server on which the IBM Content Collector information center is to run. Default value: 8888	

The configuration of the Web Application

Table 31. Information for configuring the Web Application

Information	Notes	Record your value here
Host name	The fully qualified host name of the server on which the Web Application is installed. Example value: `server.company.com`	
Port	The port number for communication with the server on which the Web Application is installed over an HTTPS connection. Default value: 11443	
Log file location	Default value: `C:\Program Files\IBM\Content Collector\ctms\Log`	
Repository connection	One default repository connection is listed. Add repository connections as required. You can change the default repository connection.	

The configuration of the Archived Data Access

For information of what to configure in which order for the access to archived email, see “Enabling search for email documents” on page 697

Table 32. Information for configuring the Archived Data Access

Information	Notes	Record your value here
Defined collections	A list of all collections that are defined in the configuration database. For a new installation, the Content Collector default collections that are generated during the initial configuration are listed here. Add an entry for any collection that you want to include in the archive mapping.	
Items defined in the selected collection	A list of all components that are defined for the selected collection, such as the root or child components of any IBM Content Manager item type or IBM FileNet P8 document class.	
Content server properties	A list of the available content server properties in an IBM Content Manager item type or an IBM FileNet P8 document class and that you can map to collection fields in the archive mapping.	
Text index fields	A list of text index fields that are available in the text indexer model file (IBM Content Manager) or the XIT (IBM FileNet P8), and that you can map to collection fields in the archive mapping.	

The Client Configuration

Table 33. Information for configuring the client

Information	Notes	Record your value here
Trigger mailbox	The SMTP address of the trigger mailbox required when email is archived manually. Example value: `iccjobs@company.com`	
Enable tracing	This option is used to enable or disable tracing in iNotes (formerly Domino Web Access). This information is only configurable if the email connector is configured for Lotus Domino.	

Table 33. Information for configuring the client (continued)

Information	Notes	Record your value here
Trace file location	The absolute path to the trace file for iNotes. This information is only configurable if tracing is enabled. Default value: `<temporary directory>\IBM\ ContentCollector_iNotes`	

Checklists for collecting additional archiving information for email that is archived manually

For information of what to configure in which order, see "Enabling the collection of additional archiving information" on page 431.

Metadata Web Application configuration settings

Table 34. Information for configuring the Metadata Web Application settings

Information	Notes	Record your value here
Host name	The fully qualified host name of the server on which the Derby database resides. Example value: `server.company.com`	
Port	The port number for communication with the server on which the Derby database resides. Default value: 1527	
User ID	The user ID who has access to the Derby database.	
Password	The password of the user who has access to the Derby database.	

Metadata Form Template configuration settings

Table 35. Information for configuring the Metadata Form template

Information	Notes	Record your value here
Metadata Form Template	Import a metadata form template. Select the use-defined metadata you defined under **Metadata and Lists**. Use the absolute path to the form template. Default form template: `IBM\ContentCollector\ formTemplates\form.zip`	

Metadata Form Definition configuration settings

Table 36. Information for configuring the Metadata Form Definition

Information	Notes	Record your value here
User-defined metadata	Select the use-defined metadata you defined under **Metadata and Lists**.	

Installing Content Collector according to your needs

The installation procedure depends on the source systems that contain the documents to be archived, and on the repositories in which you want to archive the documents.

Before you begin:

- Make sure the prerequisites for the installation are met. These are described in "Prerequisites for the installation" on page 29.
- If you are moving from IBM CommonStore to IBM Content Collector, read "Moving from CommonStore to Content Collector" on page 135.
- If you are upgrading from an earlier release of IBM Content Collector, read "Upgrading to version 2.2 of Content Collector" on page 131 for detailed information about the upgrading procedure.

If you run the initial configuration, you can specify in which database the configuration data for Content Collector is to be stored. If you decide to create a database, this database is created and configured with default settings during the initial configuration. If you decide to use an existing database, this database is configured during the initial configuration. You do not have to prepare DB2, SQL server, or Oracle for use with Content Collector.

In Content Collector you can specify to store archived documents in IBM FileNet P8, in IBM Content Manager, and in file system repositories.

You can select to use both IBM FileNet P8 and Content Manager at the same time during the Content Collector installation. Note however that, although you can select both content management systems, you can use the Content Collector web applications with one content system repository only. This means that you can only use one of the content management systems for document restore and search. You will not be able to use Content Collector to access documents in the other content system repository.

However, using both IBM FileNet P8 and Content Manager to store your documents could be useful, for example, if you have several source systems and want to perform different tasks on the content management systems:

- Archive email from a Microsoft Exchange journal recipient mailbox in Content Manager. As this is compliance archiving, search on email in Content Collector is done by using IBM eDiscovery Manager and not the Content Collector web applications.
- Archive documents from Microsoft SharePoint, File System, or both, in FileNet P8.
- Search the email in Content Manager by using IBM eDiscovery Manager.
- Map documents in FileNet P8 to access archived content through the IBM Content Collector web applications deployed on WebSphere Application Server.

Installing Content Collector for use with one or more source systems and Content Manager

Use this installation procedure if you want to use one or more source systems and use Content Manager as your repository.

Before you begin:

- Check the prerequisites for the installation.
- If you plan to install Content Collector on several servers, read "Installing Content Collector on several servers" on page 87. In addition, make a note of all installation options that you select when you install Content Collector on the first server. You must select the same options on the extension nodes with the exception of the type of server. You must select extension node on additional servers.

Complete these steps to install Content Collector:

1. Install the IBM Content Collector Text Search Support on your Content Manager server to be able to search in archived documents.
2. To use a Microsoft SharePoint source system, install IBM Content Collector for Microsoft SharePoint on one Web Front End server in your Microsoft SharePoint farm.
3. To provide automatic retrieve functionality on Lotus Domino clients, install IBM Content Collector Notes Client Extension on each Notes client.
4. Log on to the server where you want to install Content Collector. Use the administrator ID for the logon.
5. Install Content Collector Server on the server on which you want Content Collector to run.
6. Unless you are upgrading from an earlier release of Content Collector, perform the initial configuration to configure your source systems, Content Manager, and a DB2 database for storing the configuration data. In addition, start the Configuration Manager to store the configuration data in the DB2 database. Complete these steps:
   a. Start the initial configuration.
   b. Specify your system configuration, that is, your source systems, your repository, and your database management system.
   c. Configure the source systems.
   d. Configure Content Manager.
   e. Configure a DB2 database for the Content Collector configuration data.
   f. Run the configuration steps and store the configuration data.
7. Verify and adjust the initial configuration settings.
8. If you use an email system as a source system, complete one or more of these steps:
   - If you use Lotus Domino, replace the standard Lotus Notes mail template in all mailboxes with the mail template that contains the design changes for Content Collector.
   - If you use Microsoft Exchange and want to archive documents interactively, install either of the following programs:
     – Install IBM Content Collector Outlook Extension on Microsoft Outlook.
     – Install IBM Content Collector OWA Support on Microsoft Exchange.
9. Optional: Check the settings of the Content Collector environment variables.

10. Check the IBM Content Collector service accounts. It is recommended that you use the same user account for the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service services.
    - Determine which service accounts are required.
    - Change the account of a service.
11. Start the Task Routing Engine service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start Task Routing Engine**. This service controls all other connector services. However, the Task Routing Engine service only starts those connectors that are referenced by valid, active task routes.
12. Start the IBM Content Collector Web Application service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start ICC Web Applications**. This service is required for interactive archiving, and to enable document preview, search and restore.
13. If you configured email through SMTP, start the IBM Content Collector SMTP service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start SMTP Receiver**. This service is required for archiving email through SMTP.

You can now use Content Collector. Your next steps would be:

- Configure task routes by using the IBM Content Collector Configuration Manager.
- Configure indexing to enable searching in a Content Manager repository.

Installing Content Collector for use with one or more source systems and FileNet P8

Use this installation procedure if you want to use one or more source systems and FileNet P8 as your repository.

Before you begin:

- Check the prerequisites for the installation.
- If you plan to install Content Collector on several servers, read "Installing Content Collector on several servers" on page 87. In addition, make a note of all installation options that you select when you install Content Collector on the first server. You must select the same options on the extension nodes with the exception of the type of server. You must select extension node on additional servers.

Complete these steps to install Content Collector:

1. To use a Microsoft SharePoint source system, install IBM Content Collector for Microsoft SharePoint on one Web Front End server in your Microsoft SharePoint farm.
2. To provide automatic retrieve functionality on Lotus Domino clients, install IBM Content Collector Notes Client Extension on each Notes client.
3. Log on to the server where you want to install Content Collector. Use the administrator ID for the logon.
4. Install Content Collector Server on the server on which you want Content Collector to run.
5. Unless you are upgrading from an earlier release of Content Collector, perform the initial configuration to configure your source systems, FileNet P8, and the database for storing the configuration data. Complete these steps:

a. Start the initial configuration.
b. Specify your system configuration, that is, your source systems, your repository, and your database management system.
c. Configure your source systems.
d. Configure FileNet P8.
e. Configure the database for the Content Collector configuration data.
f. Run the configuration steps and store the configuration data.

6. Verify and adjust the initial configuration settings.
7. If you use an email system as a source system, complete one or more of these steps:
   - If you use Lotus Domino, replace the standard Lotus Notes mail template in all mailboxes with the mail template that contains the design changes for Content Collector.
   - If you use Microsoft Exchange and want to archive documents interactively, install either of the following programs:
     – Install IBM Content Collector Outlook Extension on Microsoft Outlook.
     – Install IBM Content Collector OWA Support on Microsoft Exchange.
8. Optional: Check the settings of the Content Collector environment variables.
9. Check the IBM Content Collector service accounts. It is recommended that you use the same user account for the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service services.
   - Determine which service accounts are required.
   - Change the account of a service.
10. Start the Task Routing Engine service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start Task Routing Engine**. This service controls all other connector services. However, the Task Routing Engine service only starts those connectors that are referenced by valid, active task routes.
11. Start the IBM Content Collector Web Application service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start ICC Web Applications**. This service is required for interactive archiving, and to enable document search and restore.
12. If you configured email through SMTP, start the IBM Content Collector SMTP service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start SMTP Receiver**. This service is required for archiving email through SMTP.

You can now use Content Collector. Your next step would be to configure the task routes by using the IBM Content Collector Configuration Manager.

Installing Content Collector on several servers - scale out

You can install IBM Content Collector on more than one server to distribute the workload to several servers. This distributed processing model is called *scale out*.

Before you begin, read the general information about scale out.

You must complete the following tasks to install and configure scale out:

1. On the primary node, install IBM Content Collector for use with one or more repositories:
   - Install IBM Content Collector for use Content Manager

- Install IBM Content Collector for use with IBM FileNet P8

2. Configure the primary node

   In addition, make sure that you can archive documents on the primary node. Also ensure that the host name of the server on which the Web Application is installed is specified correctly. This is necessary because the links in the stubs refer to the Web Application.
3. On the extension nodes, install Content Collector for use with the same source systems, repositories, and configuration database as on the primary node. See step 1 on page 61. Make sure that you set the same paths that you used on the primary node for all prerequisite software and for all Content Collector file locations on the extension nodes.
4. Configure the extension nodes. This task includes copying the database configuration file from the primary node to the extension nodes and testing the database connection on the extension nodes.
5. Start the IBM Content Collector Task Routing Engine service on the primary node and then on the extension nodes.

Installation steps

Depending on your setup, you must complete one or more of the described installation steps. The installation procedures for your setup show the sequence in which the individual steps must be performed.

Related tasks

"Installing Content Collector according to your needs" on page 58

Installing Content Collector Text Search Support

You can search for words, phrases, and character strings, in the content of archived documents. If you use Content Manager as your repository, you must install Content Collector Text Search Support, which is supplied with IBM Content Collector. If you use FileNet P8, you must enable the respective object store for search.

Before you begin:

- Check the prerequisites for the installation.
- If you are moving from IBM CommonStore to IBM Content Collector, or are upgrading from a Content Collector V2.1.0.x release, it is recommended that you do not uninstall the text-search user-exit component on the Content Manager server until after you have installed IBM Content Collector Text Search Support and have enabled all existing Content Collector or CommonStore item types for processing by the indexer for text search.
- If you are upgrading from a Content Collector V2.1.1.x release, you do not have to uninstall the earlier text search support installation. While the wizard is running, you are not prompted for any information. The component is automatically installed in the previous installation location. If files need to be overwritten, permit this action.
- The installation wizard tries to install IBM Content Collector Text Search Support without stopping DB2. If this is not possible, you are asked if the wizard can stop DB2. If you disallow this, the installation process is cancelled. If the wizard is unable to stop DB2, the installation process is also cancelled. In the later case, stop DB2 manually and run the wizard again. Restart DB2 after the installation.

To run the installation wizard:

1. Insert the product DVD into the computer on which you want to install the text-search component. You can also download and extract the appropriate installation package.
2. Start the installation wizard:
   - On Linux and UNIX, run **install.bin**, which is in the directory `/CM8Text/operating_system`. *operating_system* stands for the name of the operating system, for example, Linux.
   - On Windows, run **install.exe**, which is in the directory `\CM8Text\WIN`.
3. If prompted, answer the remaining prompts.
   - These are the default installation folders:

Operating system	Path
Windows (32-bit)	`C:\Program Files\IBM\ICCIndexer`
Windows (64-bit)	`c:\Program Files(x86)\IBM\ICCIndexer`
Linux and UNIX	`/opt/IBM/ICCIndexer`

   - If you are installing the component on Linux or UNIX, specify the DB2 administrator ID.

Installing Content Collector for Microsoft SharePoint

The Microsoft SharePoint connector requires the installation of web services and other components on each SharePoint server to which you want to connect or on one Web Front End server within a farm.

You can install IBM Content Collector for Microsoft SharePoint in GUI mode, console mode, or silent mode.

Installing Content Collector for Microsoft SharePoint in GUI mode

Complete these steps to install IBM Content Collector for Microsoft SharePoint in GUI mode.

If you are upgrading from IBM Content Collector Version 2.1.1, and your environment contains multiple Microsoft SharePoint Web Front End servers, you must first uninstall IBM Content Collector for Microsoft SharePoint from each Web Front End server.

The IBM Content Collector for Microsoft SharePoint installation program installs a web service and several files that IBM Content Collector requires to communicate with the SharePoint server. It also creates a feature that includes a content type and site columns used for post-processing.

To install IBM Content Collector for Microsoft SharePoint in GUI mode:

1. Verify that your Microsoft SharePoint server is running.
2. To start the IBM Content Collector for Microsoft SharePoint installation program, run `install.exe`, which is located in the `SharePoint` directory of the DVD or install image.
3. Follow the instructions in the installer. In a SharePoint farm, you need to install IBM Content Collector for Microsoft SharePoint on only one Microsoft SharePoint Web Front End server.
4. Optional: If you are upgrading from IBM Content Collector Version 2.1.1, and installed Microsoft Web Service Enhancements (WSE) for Microsoft .NET and

the IBM FileNet Content Engine API for Microsoft .NET on any of your SharePoint servers, you can remove them if no other applications need them. If you do so, you must remove their entries from the `web.config` files of each server on which you installed IBM Content Collector for Microsoft SharePoint Version 2.1.1.

You can verify that the installation completed successfully by checking for errors in the installation log file, normally stored in `C:\Program Files\IBM\Content Collector for Microsoft SharePoint\ IBM_Content_Collector_for_Microsoft_SharePoint_InstallLog.log`.

Installing Content Collector for Microsoft SharePoint in console mode

Complete these steps to install IBM Content Collector for Microsoft SharePoint in console mode.

If you are upgrading from IBM Content Collector Version 2.1.1, and your environment contains multiple Microsoft SharePoint Web Front End servers, you must first uninstall IBM Content Collector for Microsoft SharePoint from each Web Front End server.

1. Insert the product DVD into the computer on which you want to install IBM Content Collector for Microsoft SharePoint. You can also download and extract the appropriate installation package.
2. Open a Command Prompt window and navigate to the SharePoint directory of the DVD or install image.
3. Type `install.exe -i console` to start the installation.
4. Follow the instructions in the Command Prompt window. In a SharePoint farm, you need to install IBM Content Collector for Microsoft SharePoint on only one Microsoft SharePoint Web Front End server.
5. Optional: If you are upgrading from IBM Content Collector Version 2.1.1, and installed Microsoft Web Service Enhancements (WSE) for Microsoft .NET and the IBM FileNet Content Engine API for Microsoft .NET on any of your SharePoint servers, you can remove them if no other applications need them. If you do so, you must remove their entries from the `web.config` files of each server on which you installed IBM Content Collector for Microsoft SharePoint Version 2.1.1.

You can verify that the installation completed successfully by checking for errors in the installation log file, normally stored in `C:\Program Files\IBM\Content Collector for Microsoft SharePoint\ IBM_Content_Collector_for_Microsoft_SharePoint_InstallLog.log`.

Installing Content Collector for Microsoft SharePoint in silent mode

Complete these steps to install IBM Content Collector for Microsoft SharePoint in silent mode.

If you are upgrading from IBM Content Collector Version 2.1.1, and your environment contains multiple Microsoft SharePoint Web Front End servers, you must first uninstall IBM Content Collector for Microsoft SharePoint from each Web Front End server.

1. Create a response file with the following installation options. The values shown are examples only.

```
#Set to silent install
#do not change this value
INSTALLER_UI=SILENT

#Has the license been accepted
#do not change this value
#----------------------------
LICENSE_ACCEPTED=TRUE

#Select Install Folder
#Please only use ASCII characters in the path.
#Failure to do so will cause errors during the installation.
#---------------------
USER_INSTALL_DIR=C:\\Program Files\\IBM\\Content
Collector for Microsoft SharePoint

#Specify the log file destination folder
#The default value is the user installation folder as specified above.
#If an error occurs during the silent installation
and the log file is not created
#try specifying the path of an existing folder for the log file
(Remember to use \\ to escape the directory separator)
#---------------------
INSTALL_LOG_DESTINATION=$USER_INSTALL_DIR$
```

2. Save the response file to your disk.
3. Insert the product DVD into the computer on which you want to install IBM Content Collector for Microsoft SharePoint. You can also download and extract the appropriate installation package. In a SharePoint farm, you need to install IBM Content Collector for Microsoft SharePoint on only one Microsoft SharePoint Web Front End server.
4. To start the installation in silent mode, open a Command Prompt window and enter the following command:

   `install.exe -i SILENT -f <full_path_to_response_file>`

   where *<full_path_to_response_file>* is the full path to the response file; for example, `c:\temp\myresponse.txt`.

   Important: `SILENT` must be specified in uppercase characters if you have a Turkish operating system.
5. Optional: If you are upgrading from IBM Content Collector Version 2.1.1, and installed Microsoft Web Service Enhancements (WSE) for Microsoft .NET and the IBM FileNet Content Engine API for Microsoft .NET on any of your SharePoint servers, you can remove them if no other applications need them. If you do so, you must remove their entries from the `web.config` files of each server on which you installed IBM Content Collector for Microsoft SharePoint Version 2.1.1.

You can verify that the installation completed successfully by checking for errors in the installation log file, normally stored in `C:\Program Files\IBM\Content Collector for Microsoft SharePoint\IBM_Content_Collector_for_Microsoft_SharePoint_InstallLog.log`.

Installing Content Collector Notes Client Extension

Install IBM Content Collector Notes Client Extension on Lotus Notes® clients to enable automatic document retrieval on the Notes client.

Install IBM Content Collector Notes Client Extension on the Lotus Notes client workstations. It should not be installed on the system where IBM Content Collector Server is installed.

You can install IBM Content Collector Notes Client Extension in GUI mode or in silent mode.

Installing Content Collector Notes Client Extension in GUI mode

Install IBM Content Collector Notes Client Extension to enable automatic document retrieval on the client.

Before you begin, check the prerequisites for the installation.

The Notes Client Extension enables you to automatically retrieve the content of a stubbed document from the archive temporarily in your client.

Double-clicking or selecting Reply, Reply to All, or Forward temporarily retrieves the document. If you enabled the preview pane, the original content will also be displayed there.

A flag is preserved in the stubbed document if the document is replied, forwarded, or marked as follow-up. The temporary copy of the document is deleted when you close your mail database. Note that automatic retrieve still works if you select multiple documents to be forwarded.

This functionality is only available on the client if **Retrieve and display document when opened** is selected for the client configuration under **General Settings → Client Configuration** in the IBM Content Collector Configuration Manager. Depending on the number of clients in your topology, enabling this option can have a negative impact on the Content Collector Server performance. This is because requests for retrieving and displaying documents are handled by the Content CollectorWeb Application and, therefore, put load on the Web Application Server.

To install IBM Content Collector Notes Client Extension on a client, complete these steps:

1. In the Windows Explorer, change to the directory where you extracted the IBM Content Collector installation package.
2. Run the following command located in the `\NotesClient` directory of the installation package: `install.exe`
3. Answer the remaining prompts. These are the default installation folders:

Operating system	Path
Windows (32-bit)	`C:\Program Files\IBM\IBM Content Collector Lotus Notes Extension`
Windows (64-bit)	`c:\Program Files(x86)\IBM\IBM Content Collector Lotus Notes Extension`

Before you can automatically retrieve the content of a stubbed document from the archive temporarily in your client, you must have enabled your existing template to include IBM Content Collector archiving functionality. Automatic document retrieval on the client does not work if the template has not been enabled.

Installing Content Collector Notes Client Extension in silent mode

Use the silent installation if you want to install IBM Content Collector Notes Client Extension on many client machines. No user interaction is required during this installation mode.

If automatic document retrieval is enabled for very many clients, it can have a negative impact on the server performance. Installing the Notes Client Extension in silent mode enables an administrator to restrict the amount of users that can use the automatic temporary retrieve functionality.

To install IBM Content Collector Notes Client Extension on clients in silent mode:

1. Create a response file with the following installation options. The values shown are examples only.

   ```
   INSTALLER_LOCALE=en
   # Accept the license panel.
   LICENSE_ACCEPTED=TRUE
   # Set the install folder.
   USER_INSTALL_DIR=C:\\Program Files\\IBM\\IBM Content Collector Lotus Notes Extension
   ```

2. Save the response file, for example, as `myresponse.txt` under `C:\temp`.
3. To start the installation, open a Command Prompt window and enter the following command:

   `install.exe -i SILENT -f <full_path_to_response_file>`

   <full_path_to_response_file> is the full path to the response file, for example, `c:\temp\myresponse.txt`.

   Important: `SILENT` must be specified in uppercase characters if you have a Turkish operating system.

Notes Client Extension enables retrieving the content of a stubbed document from the archive temporarily by double-clicking a document or selecting Reply, Reply to All, or Forward. However, this functionality is only available on the client if **Retrieve and display document when opened** is selected for the client configuration under **General Settings** → **Client Configuration** in the IBM Content Collector Configuration Manager. Note that the client templates must be enabled to include IBM Content Collector archiving functionality before automatic document retrieval will work.

Installing Content Collector Server

You can install IBM Content Collector Server in GUI mode, console mode, or silent mode.

Before you begin:

- Check the prerequisites for the installation.
- If you want to use an external WebSphere Application Server instead of the embedded Web Application Server, configure the Web Application Server.
- If you use Microsoft SharePoint as a source system, install IBM Content Collector for Microsoft SharePoint.

Installing Content Collector Server in GUI mode

Complete these steps to install IBM Content Collector Server in GUI mode.

1. In the Windows Explorer, change to the directory where you extracted the IBM Content Collector installation package.
2. Run the following command located in the `\Server` directory of the installation package on the CommonStore server: `install.exe`
3. Answer the remaining prompts as the installer guides you through the installation process.
   - You can install Content Collector Server once per server. However, you can install Content Collector on several servers. In this case, you must select one

server as your primary node and all other servers as your extension nodes. For each node, you must select the same installation options and components. In addition, installation patches must be installed on the primary node and on each extension node.

- You can select one or more source systems and one or more target repositories. Only select source systems and target repositories that are installed and meet the prerequisites required by IBM Content Collector.
- Use an alias for the host name to be more flexible with your system setup. The alias must not be tied to the machine name and it must be resolvable to the machine that runs the Web Application Server.
- If you use the embedded Web Application Server, the IBM Content Collector Web Application and the IBM Content Collector Configuration Web Service run on the IBM Content Collector Server machine.

Installing Content Collector Server in console mode

Complete these steps to install IBM Content Collector Server in console mode.

1. Insert the product DVD into the computer on which you want to install Content Collector Server. You can also download and extract the appropriate installation package.
2. Open a Command Prompt window.
3. Type `install.exe -i console` to start the installation.
4. Follow the instructions in the Command Prompt window.

Installing Content Collector Server in silent mode

Complete these steps to install IBM Content Collector Server in silent mode.

1. Create a response file with the following installation options. The values shown are examples only.

```
#
#Has the license been accepted
#-----------------------------
LICENSE_ACCEPTED=TRUE

# NODE_A <=> primary node
# NODE_B <=> extension node
#-----------------------------
CHOSEN_INSTALL_SET=NODE_A

#Choose Install Folder
#---------------------
USER_INSTALL_DIR=D:\\IBM\\ContentCollector

#Get User Input
#--------------
# Source System (install=1/de-install=0):
#
# USER_INPUT_EXCHANGE <=> Microsoft Exchange
# USER_INPUT_LD       <=> Lotus Domino
# USER_INPUT_RC       <=> File System
# USER_INPUT_SP       <=> Share Point
# USER_INPUT_SMTP     <=> Simple Mail Transfer Protocol

USER_INPUT_EXCHANGE=1
USER_INPUT_LD=1
USER_INPUT_RC=1

# Source System:
#
# USER_INPUT_CM8     <=> IBM Content Manager
# USER_INPUT_P8      <=> IBM FileNet P8
```

```
# USER_INPUT_PANAGON <=> IBM FileNet Image Services
#                        (for File System Source only)
#
USER_INPUT_CM8=1
USER_INPUT_P8=0
USER_INPUT_PANAGON=0

# Select Deployment Method
#
# USER_INPUT_USE_EWAS      <=> Automatically to the embedded Websphere
#                              Application Server
# USER_INPUT_USE_DIFF_EWAS <=> Manually to a different Websphere Application
#                              Server
#
USER_INPUT_USE_EWAS=1
USER_INPUT_USE_DIFF_EWAS=0

USER_INPUT_EWAS_HOST=localhost
USER_INPUT_EWAS_PORT=11443
```

2. Save the response file to your disk.
3. Insert the product DVD into the computer on which you want to install Content Collector Server. You can also download and extract the appropriate installation package.
4. To start the installation in silent mode, open a Command Prompt window and enter the following command:

   `install.exe -i SILENT -f <full_path_to_response_file>`

   <full_path_to_response_file> is the full path to the response file, for example, `c:\temp\myresponse.txt`.

   Important: `SILENT` must be specified in uppercase characters if you have a Turkish operating system.

Performing the initial configuration

During the initial configuration, you provide details about your system configuration, your source systems, your repositories, and the database for the configuration data. In addition, you create the item type for Content Manager and configure the configuration database. You can perform the initial configuration only once. For additional configuration tasks, such as creating additional item types for Content Manager or configuring additional object stores for FileNet P8, use the set-up tools.

Related tasks

"Using the set-up tools" on page 572

Starting the initial configuration

Complete these steps to start the initial configuration.

You completed these tasks:

1. If you use Content Manager as your repository: "Installing Content Collector Text Search Support" on page 62.
2. If you use Microsoft SharePoint as a source system: "Installing Content Collector for Microsoft SharePoint in GUI mode" on page 63.
3. "Installing Content Collector Server" on page 67.

Continue with the following steps:

1. Open the Initial Configuration wizard in either of the following ways:

- Click **Yes** in the Launch Initial Configuration window that opens after Content Collector Server is installed.
- Click **Start → All Programs → IBM Content Collector → Initial Configuration**.

2. Click Next.

Continue with this task: "Specifying your system configuration."

Specifying your system configuration

Specify from which source systems you want to archive your documents and in which repositories you want to archive them. In addition, specify the database management system that you want to use for storing the configuration data of IBM Content Collector.

You completed this task: "Starting the initial configuration" on page 69

Continue with these steps in the Initial Configuration wizard:

1. On the System Configuration page, select one or more of the listed source systems.
2. Select the repositories in which you want to archive the documents.
3. Select one of the listed database management systems. If you use Content Manager, **DB2** is selected because IBM Content Collector supports Content Manager with DB2 only.
4. Click **Next**.

Continue with this task: "Configuring your source systems"

Configuring your source systems

For each source system that you selected on the System Configuration page of the Initial Configuration wizard, specify the details that IBM Content Collector needs for the configuration.

File System source only: The File System source connector requires no initial configuration. However, to use it with IBM FileNet P8 or with IBM Content Manager, ensure that the connectors run under a user with the correct permissions to open and read files on the target file system. The simplest method is to run the file system source and the target repository connectors under the same account. For details on the required permissions, see:

- "Required IBM FileNet P8 privileges for the connector" on page 207
- "Required Content Manager privileges for the connector" on page 201

The source systems are configured in the sequence in which they are listed under **Source System** on the System Configuration page. For example, if you selected **Lotus Domino** and **Microsoft SharePoint** as source systems, you start with the configuration of Lotus Domino. However, the Initial Configuration wizards guides you through the configuration steps in the required sequence.

Configuring Lotus Domino:

Complete these steps to configure IBM Content Collector for use with Lotus Domino.

Before you begin:

- If you want to use Lotus iNotes (formerly Domino Web Access), which provides browser-based access to Lotus Notes features, make sure that iNotes is configured on the Lotus Domino servers hosting the mailboxes.

  For Lotus Domino Version 8.5.1, also make sure that an Extensions Forms File exists on Lotus Domino Server. If the file does not exist, you must create one before you can enable the Content Collector features on Lotus iNotes. For information about how to create an Extensions Forms File, see the topic about customizing the look of Lotus iNotes in the IBM Lotus Domino and Notes information center at http://publib.boulder.ibm.com/infocenter/domhelp/v8r0/index.jsp.
- Ensure that the Lotus Domino server is running.
- You completed the following task: "Specifying your system configuration" on page 70.

Continue with these steps in the Initial Configuration wizard:

1. On the Lotus Domino Configuration page, specify the following information:
   - The name and the domain of your Lotus Domino home server, in the format *server_name/domain*. Example value: `myServer/Organization`.
   - The full path and the password of the administrator ID file.

     The administrator ID is used by IBM Content Collector to create the runtime environment and enable using the Lotus Domino template and iNotes form with IBM Content Collector functionality.

     The selected ID must have sufficient privileges to change templates. Typically, this means Manager access rights.

     To enable templates remotely, the administrator ID must be Manager on the mail template, the iNotes (Domino Web Access) template, (there is no standalone iNotes template starting with Domino V8) and the forms database.

     The following rights must be configured irrespective of whether the administrator ID or whether another ID is used to enable the IBM Content Collector template for iNotes (Domino Web Access):
     - The rights to sign or run unrestricted methods and operations
     - The rights to sign or run restricted LotusScript/Java agents
     - The user needs to be editor with remove document access at least in order to use the IBM Content Collector functions on iNotes.
   - The name, including the full path, and the password of the connector ID file. This file is used for the transactions between your Domino servers and Content Collector.
   - The version of the TCP/IP that is to be used for the connections between Content Collector and your Domino servers. IPV6 is not suitable for Lotus Domino Release 7.
2. Click **Next**.
3. On the Domino Template Configuration page, specify the location of the mail template to which the design changes are to be applied that are required for Content Collector. This modified template is to replace the standard Lotus Notes mail template in all mailboxes. Specify the following information:
   - Whether the Lotus Notes mail template that you want to use is stored on a remote Domino server or is stored locally.
   - If the template is on a remote Domino server, specify the name and the domain of this server in the format *server_name/domain*.

- The name of the template that is to be modified. If the template is stored locally, specify the full path. If the template is on a remote Domino server, specify the path relative to the Domino data directory.
- Enable the support of browser-based access and specify:
  - The template name as follows:
    - For Lotus Domino Version 7, `dwa7.ntf`
    - For Lotus Domino Version 8, `mail8.ntf`
    - For Lotus Domino Version 8.5.0, `mail85.ntf`
    - For Lotus Domino Version 8.5.1, `mail85.ntf`
    - For Lotus Domino Version 8.5.2, `mail85.ntf`
  - The forms database as follows:
    - For Lotus Domino Version 7, `iNotes\Forms7.nsf`
    - For Lotus Domino Version 8, `iNotes\Forms8.nsf`
    - For Lotus Domino Version 8.5.0, `iNotes\Forms85.nsf`
    - For Lotus Domino Version 8.5.1, `iNotes\Forms85_x.nsf`
    - For Lotus Domino Version 8.5.2, `iNotes\Forms85_x.nsf`

4. Click **Next**.
5. On the Domino Template Customization page, specify the layout of the modified mail template that is to replace the standard Lotus Notes mail template. Specify the following information:
   - The Content Collector functions to be added.
   - The name under which you want to group the selected Content Collector functions. This group will appear as a submenu of the Lotus Notes **Actions** menu.
   - The name for the Content Collector view if you decide to add the Content Collector view to the mail template.
   - The languages to be made available.
6. Click **Next**.

Depending on your system configuration, continue with one of these tasks:

- "Configuring Email through SMTP" on page 73
- "Configuring Microsoft SharePoint" on page 73
- "Configuring Content Manager" on page 74
- "Configuring IBM FileNet P8" on page 75

Configuring Microsoft Exchange:

Complete these steps to configure IBM Content Collector for use with Microsoft Exchange.

Before you begin:

- You completed the following task: "Specifying your system configuration" on page 70.

Continue with these steps in the Initial Configuration wizard:

1. On the Microsoft Exchange Configuration page, specify the following information:
   - The fully qualified host name of the server on which your Microsoft Exchange mailbox is located.

- The credentials for the user account that accesses the Active Directory information. The user ID and password are used for the transactions between your Exchange servers and Content Collector. Enter the SMTP address of the user. If Active Directory cannot be accessed by using the SMTP address, use the distinguished name of the user in the format `CN=ICCConnectorUser,CN=Users,DC=company,DC=com`.

2. Click **Next**.

Depending on your system configuration, continue with one of these tasks:
- "Configuring Email through SMTP"
- "Configuring Microsoft SharePoint"
- "Configuring Content Manager" on page 74
- "Configuring IBM FileNet P8" on page 75

Configuring Email through SMTP:

Complete these steps to configure IBM Content Collector for use with SMTP.

Before you begin:
- You completed the task: "Configuring your source systems" on page 70.

Continue with these steps in the Initial Configuration wizard:

On the SMTP Configuration page, specify the following information:
- The absolute path to the message queue directory where SMTP email that is received by the SMTP Receiver is queued until the email is archived.
- The ID and password that you want to use in your email system to connect to the SMTP Receiver. The SMTP Receiver uses SMTP authentication to validate connections. The ID must only consist of letters in the English Alphabet [A-Z,a-z].

Depending on your system configuration, continue with one of these tasks:
- "Configuring Microsoft SharePoint"
- "Configuring Content Manager" on page 74
- "Configuring IBM FileNet P8" on page 75

Configuring Microsoft SharePoint:

Complete these steps to configure IBM Content Collector for use with Microsoft SharePoint.

You completed the following task: "Specifying your system configuration" on page 70.

To configure your SharePoint connector from the Initial Configuration wizard, select Microsoft SharePoint and follow these steps in the Microsoft SharePoint Configuration window:

1. Type the **User ID**, Windows Server **Domain**, and **Password** of a user who belongs to the SharePoint Site Collection Administrators group for this site.
2. Enter a valid URL to a top-level or sub-level SharePoint site, in the format: http://*server name*:*port number*/*path*/ The address must begin with `http://` or `https://` and might require a port number or path. Type a *port number* if the server is running multiple SharePoint web applications. Type a *path* to limit

your collection to a subsite. Click **Validate** to verify that the application recognizes the address. If validation fails, see "SharePoint connector validation fails" on page 812

Depending on your system configuration, continue with one of these tasks:

- "Configuring Content Manager"
- "Configuring IBM FileNet P8" on page 75

Configuring your repositories

For each repository that you selected on the System Configuration page of the Initial Configuration wizard, specify the details that IBM Content Collector needs for the configuration.

You can select to use both IBM FileNet P8 and Content Manager at the same time during the Content Collector installation. Note however that, although you can select both content management systems, you can use the Content Collector web applications with one content system repository only. This means that you can only use one of the content management systems for document restore and search. You will not be able to use Content Collector to access documents in the other content system repository.

Configuring Content Manager:

Complete these steps to configure IBM Content Collector for use with Content Manager. In addition, create one item type for each source system as a container for the documents to be archived. You can create additional item types after the initial configuration by using the Setup Tools window of Content Collector.

Before you begin:

- Ensure that the DB2 instance for the Content Manager library server database is running.
- The IBM Content Collector Text Search Support must be installed on the IBM Content Manager server and the repository must be enabled for search. For more information, see the section on enabling an IBM Content Manager repository for search.

  If IBM Content Manager is installed on more than one server, Content Collector Text Search Support must be installed on the IBM Content Manager machine where the library server and Net Search Extender is installed and not where the resource manager is installed.
- You can also use Content Manager for z/OS® as a repository. Note, however, that you cannot search in documents that are archived in Content Manager for z/OS.
- IBM Content Collector cannot archive any objects (documents) larger than 2 GB in a IBM Content Manager repository.
- You completed this task: "Configuring your source systems" on page 70.

Continue with these steps in the Initial Configuration wizard:

1. On the Content Manager Configuration page, specify the following information:
   - The name of the Content Manager server on which you want to create the repository.

     You can define repositories and configure item types for these repositories on more than one Content Manager server. Use the Initial Configuration wizard to specify the first server. For additional servers, use **Tools** → **Set-up Tools** →

CM Repository Configuration in the IBM Content Collector Configuration Manager. Alternatively, you can select **Start → All Programs → IBM Content Collector → Set-up Tools → CM Repository Configuration**.

- The ID and password of the Content Manager administrator who is allowed to create and configure the repository.
- The ID and password of the Content Manager user who is allowed to archive and restore documents.
- The installation directory of IBM Content Collector Text Search Support. This field is not available if you selected a server on which Content Manager for z/OS runs.

2. Click **Next**.
3. On the Item Type Configuration page, configure the item type that you want to use as a container for the archived documents. For the email systems, you must also configure an item type for the attachments to be archived.
   - For Lotus Domino, Microsoft Exchange, and email through SMTP, specify the following information:
     – A name for the item type for the email. If the name already exists, a number is added to the item type when it is created.
     – Whether you want to create an item type for the attachments or use an existing item type. If you want to create an item type, type a name for it. If you want to use an existing item type, select a name from the list.
     – The directory and the working directory for the text-search index for the email item type.

     Note: The item type that you create for Lotus Domino applies only to Lotus Domino email applications. For other Lotus Domino applications, you can create an item type after the initial configuration from **Tools → Set-up Tools → CM Repository Configuration**. Alternatively, you can select **Start → All Programs → IBM Content Collector → Set-up Tools → CM Repository Configuration**.
   - For Microsoft SharePoint and File System, specify the following information:
     – A name for the item type to be created. If an item type already exists, you need not create one.
     – The directory and the working directory for the text-search index. You can also decide to disable text search for the item type.
4. Click **Next**.

Continue with either of these tasks:

- If Content Manager is your only repository, continue with "Configuring a DB2 database" on page 79.
- If you have FileNet P8 in addition, continue with "Configuring IBM FileNet P8."

Configuring IBM FileNet P8:

Complete these steps to configure IBM Content Collector for use with IBM FileNet P8.

Before you begin:

- You completed the tasks: "Configuring your source systems" on page 70.
- IBM FileNet Content Search Engine must be installed and configured.

- FileNet P8 Content Engine Server must be installed and configured for content-based retrieval (CBR). For further information, see the section on configuring Content Engine for CBR in the FileNet P8 documentation.
- A FileNet P8 object store with a file storage area must exist. The object store is enabled for content-based retrieval (CBR). See the section on creating object stores in the FileNet P8 documentation and the topic "Enabling an IBM FileNet P8 object store for search" on page 675 for further information.

  Important: Set up your target object store with a file storage area as the default content store. A file storage area stores content in a network-accessible directory. To prepare your system for index area creation, each file storage area that will be full-text indexed must be accessible by both the Content Engine and the Autonomy K2 server that will perform the full-text indexing. The index area creation is required for email, as well as for File System and Microsoft SharePoint retrieval.

  For performance reasons it is recommend that the FileNet P8Content Engine has direct access to the file storage area and that the IBM FileNet Content Search Engine accesses this area remotely. Conversely, it is strongly recommended that the IBM FileNet Content Search Engine has direct access to Content Search Engine collection and temporary directories and that the FileNet P8 Content Engine accesses these remotely.

Continue with these steps in the Initial Configuration wizard:

1. On the FileNet P8 Server Configuration page, specify the following information:
   - The information that Content Collector needs to build the URL of the Content Engine, or specify the full URL.
   - The ID and password of the FileNet P8 Content Engine administrator who is allowed to create and configure the repository.
2. Click **Next**.
3. On the FileNet P8 Task Configuration page, specify the following information:
   - The ID and password of the FileNet P8 user who is allowed to archive and restore documents.
   - The FileNet P8 object store where you want to create the repository: Click **Retrieve** to retrieve a list of object stores within eligible domains.

     You can define one or more object stores on more than one IBM FileNet P8 server. Use the Initial Configuration wizard to specify the first server and the first object store. To specify additional servers and object stores, use **Tools** → **Set-up Tools** → **P8 Repository Configuration** in the IBM Content Collector Configuration Manager or **Start** → **All Programs** → **IBM Content Manager** → **Set-up Tools** → **P8 Repository Configuration**.
4. Click **Next**.

Continue with this task: "Configuring the database for the Content Collector configuration data" on page 79.

Related information

Product documentation for FileNet P8 platform

Objects deployed to FileNet P8:

During the installation and configuration of the IBM FileNet P8 connector, IBM Content Collector deploys classes and other items to the target server.

All the objects that Content Collector creates in IBM FileNet P8 begin with the prefix *ICC*.

Important: Do not delete any of these objects.

For an overview of the Content Collector data model, see "Definition of the storage data model" on page 15

Documents and Custom Objects

Content Collector creates documents and custom objects in these hierarchies:

- *ICCDocument*

  This document class is an IBM FileNet P8 Document class.

  – *ICCMail2*

    This document class represents the email. It holds the email hash for email deduplication indirectly, as well as the email object itself as a content element.

  – *ICCMailSearch2*

    This document class represents the indexable and searchable email. The indexable mail is the content element of that object.

  – *ICCFileInstance2*

    This document class represents copies of a file from the file system holding the properties. The class is derived from the Document class and the actual file content is attached as a content element.

    - *ICCSharepointInstance2*

      This document class represents copies of a file from Microsoft SharePoint holding the properties. *ICCSharepointInstanc2* derives *ICCFileInstanc2*.

- *ICCCustomObject*

  – *ICCMailInstance2*

    This class tracks all individual copies of the same email from either user mailboxes or the journal. It contains attributes holding the varying properties of each email copy which can be used to restore the user's individual copy of the mail.

Important:

- For email, the supplied default classes should be used. You can define your own classes for these sources. If you define your own classes, these must be derived from the P8 Document class and must reflect the ICC Content Collector data model hierarchy defined for email documents.

  In a Business Process Management (BPM) scenario, however, you can use any class.

- For Microsoft SharePoint and File System, the classes are only sample classes.

Properties

IBM Content Collector defines various properties which belong to Content Collector document classes and custom objects.

Table 37. Properties created in IBM FileNet P8

Property name	Property type	Document class or custom object
ICCAttachmentCorrelationKeys	Multi-value String	*ICCMail2*

Table 37. Properties created in IBM FileNet P8 (continued)

Property name	Property type	Document class or custom object
ICCCreatedBy	Single-value String	*ICCSharepointInstance2*
ICCCreatedDate	Single-value DateTime	*ICCFileInstance2*
ICCExpirationDate	Single-value DateTime	• *ICCMail2* • *ICCFileInstance2*
ICCFileName	Single-value String	*ICCFileInstance2*
ICCFilePath	Single-value String	*ICCFileInstance2*
ICCFolderPath	Single-value String	*ICCSharepointInstance2*
ICCFrom	Multi-value String	*ICCMailSearch2*
ICCLastModifiedDate	Single-value DateTime	*ICCFileInstance2*
ICCLibrary	Single-value String	*ICCSharepointInstance2*
ICCMailboxID	Single-value String	*ICCMailInstance2*
ICCMailDate	Single-value DateTime	*ICCMailSearch2*
ICCMailFlags	Single-value Integer	*ICCMailSearch2*
ICCMailInstanceReference	Multi-value Object (*ICCMailInstance2*)	*ICCMail2*
ICCMailReference	Single-value Object (*ICCMail2*)	• *ICCMailInstance2* • *ICCMailSearch2*
ICCMailSearchReference	Single-value Object (*ICCMailSearch2*)	*ICCMail2*
ICCMailUID	Single-value String	*ICCMailInstance2*
ICCModifiedBy	Single-value String	*ICCSharepointInstance2*
ICCSharePointGUID	Single-value String	*ICCSharepointInstance2*
ICCSharePointVersion	Single-value String	*ICCSharepointInstance2*
ICCSite	Single-value String	*ICCSharepointInstance2*
ICCSubject	Single-value String	*ICCMailSearch2*
ICCTitle	Single-value String	*ICCCustomObject*
ICCTo	Multi-value String	*ICCMailSearch2*
ICCVaryingFields	Single-value Binary	*ICCMailInstance2*

Deletion-related items

Content Collector creates items that enable the programmatic deletion of *ICCMail2* instances in IBM FileNet P8.

Table 38. Deletion-related items created in IBM FileNet P8

Item type	Item name
Subscription	*ICCDataModelDeletionSubscription2*
Event action	*ICCDataModelDeletionEventAction2*
Code module	*ICCHandleDeletionofDEI2.class*

Configuring the database for the Content Collector configuration data

Depending on your system configuration, you must configure a DB2 database, a SQL server database, or an Oracle database.

Configuring a DB2 database:

You can create and configure a DB2 database for storing the IBM Content Collector configuration data or you can configure an existing DB2 database.

You completed either of the following tasks:

- "Configuring Content Manager" on page 74
- "Configuring IBM FileNet P8" on page 75

Continue with these steps in the Initial Configuration wizard:

1. On the IBM DB2 Database Configuration page, specify the following information:
   - Whether you want to store the configuration data in the repository database or whether you want to create a database with default settings or use an existing database. If you use DB2 Version 8, the database must already exist.
   - If you create a database or use an existing database, specify whether this database is to be created, or is stored, on a remote server or locally. For a remote database, specify the fully qualified host name of the server on which the database is to be created or is located. For a remote or a local database, specify the DB2 port and the database name.
   - The ID and password of an DB2 administrator who is allowed to create databases.
   - The ID and password of a user who is allowed to read, and write to, the database.
2. Click **Next**.

Continue with the following task: "Running the configuration steps and storing the configuration data" on page 80.

Configuring a SQL Server database:

You can create and configure a SQL Server database for storing the IBM Content Collector configuration data or you can configure an existing SQL Server database.

You completed this task: "Configuring IBM FileNet P8" on page 75.

Continue with these steps in the Initial Configuration wizard:

1. On the SQL Server Database Configuration page, specify the following information:
   - Whether the database is to be created, or is stored, on a remote server or locally.
   - If the database is on a remote server, the fully qualified host name of this server.

     If you are using SQL Server named instance, the host name must be `<HOSTNAME>\<NAMED INSTANCE>`, for example `VDC-VQA-SQL\N1`.
   - The port of SQL Server (if you are using SQL Server named instance, the named instance port), the name for the database, and the JDBC driver.

Note: As JDBC driver, you must specify the location of the `sqljdbc4.jar` file if you use the embedded Web Application Server or IBM WebSphere Application Server Version 7. If you use IBM WebSphere Application Server Version 6.1, specify the location of the `sqljdbc.jar` file.

- The ID and password of a SQL Server administrator who is allowed to create databases.
- The ID and password of a user who is allowed to read, and write to, the database.

2. Click **Next**.

Continue with this task: "Running the configuration steps and storing the configuration data."

Configuring an Oracle database:

You can create and configure an Oracle database for storing the IBM Content Collector configuration data. You create the Oracle database based on an existing Oracle database.

You completed this task: "Configuring IBM FileNet P8" on page 75.

Continue with these steps in the Initial Configuration wizard:

1. For a single-instance Oracle database, on the Oracle Database Configuration page, specify the following information:
   a.
   b. If the database is to be stored on a remote server, the fully qualified host name of this server.
   c. The Oracle port, the name for the new database, the JDBC driver, and the JDBC connection.
   d. The ID and password of a user who is allowed to read, and write to, the database.
2. For Oracle Real Application Clusters (RAC), specify the following information. You do not need to specify the host name, port and service name as that is provided in the JDBC connection URL.
   a. Whether the new database is to be stored on a remote server or locally.
   b. Specify the JDBC URL derived from the `tnsnames.ora` file. For example, a cluster made of two nodes with `vdc-vqa-ora1v` and `vdc-vqa-ora2v` as virtual host names and `RAC.WORLD` as server name, would have the following URL:

```
jdbc:oracle:thin:@(DESCRIPTION =
(ADDRESS = (PROTOCOL = TCP)(HOST = vdc-vqa-ora1v)(PORT = 1521))
(ADDRESS = (PROTOCOL = TCP)(HOST = vdc-vqa-ora2v)(PORT = 1521))
(LOAD_BALANCE = yes)(CONNECT_DATA =(SERVER =DEDICATED)
(SERVICE_NAME = RAC.WORLD)))
```

   c. The ID and password of a user who is allowed to read, and write to, the database.

Continue with this task: "Running the configuration steps and storing the configuration data."

Running the configuration steps and storing the configuration data

Run the configuration steps until they all complete successfully. Then start the Configuration Manager to store the configuration data.

You completed this task: “Configuring the database for the Content Collector configuration data” on page 79.

Continue with these steps in the Initial Configuration wizard:

1. On the Content Collector Configuration Steps page, click **Start Configuration**. If the initial configuration completes successfully, all of the listed processes have a green check mark.
2. If a step failed, it is marked in red. You must run this step again until it is marked in green. Proceed as follows:
   a. Click **View Log** and then search the log for information about the cause of the failure.
   b. Solve the cause of the failure.
   c. If the initial configuration failed during the creation of the document classes for IBM FileNet P8, also remove any files that were created for this purpose.
   d. Click **Restart Configuration**.
3. Select the **Start the IBM Content Collector Quick Start after the initial configuration is complete** check box to start the Configuration Manager and open the Quick Start window of the Configuration Manager. By starting the Configuration Manager, the configuration data is stored in the configuration database.
4. Click **Finish**.

Verifying and adjusting the initial configuration settings

After you have run the initial configuration, information about the datastore, connectors and parts of the Configuration Manager General Settings that you provided is displayed in the Configuration Manager. You may want to check this information. A red exclamation sign next to a field indicates that the entry is missing or needs to be changed.

You completed either of these tasks:

- If you are installing Content Collector for the first time: “Performing the initial configuration” on page 69.
- If you are upgrading from an earlier release of Content Collector: “Installing Content Collector Server” on page 67.

Note: This topic shows how to get to the most important configuration data.

1. Start the Configuration Manager: Click **Start** → **All Programs** → **IBM Content Collector** → **Configuration Manager**.

   Important: The user account that is used to run the Configuration Manager must be part of the Administrators group on the local machine.
2. Check whether the configuration database is a valid database:
   a. Click **Data Stores** to change to the Data Stores view and select your configuration database.
   b. In the middle pane, select one of the listed databases.
   c. Check the settings in the right pane. The database properties should be displayed.

   See “Setting up a configuration database” on page 171 for more information.
3. Check whether all connectors are correct:
   a. Click **Connectors** to change to the Connectors view.

b. In the **Connectors** pane on the upper left, select each connector and check its settings.

See “Configuring connectors” on page 184 for more information.

4. If you use Lotus Domino or Microsoft Exchange as a source system, check the following information:
   a. Click **Connectors** to change to the Connectors view and select **Email**.
   b. Under **Email Connector**, click the **Connection** tab.
   c. Check the settings in the right pane.

   See “Configuring the Email Connector” on page 185 for more information.
5. If you use Lotus Domino or Microsoft Exchange as a source system and you want to archive documents interactively, you must specify a trigger mailbox:
   a. Click **General Settings** to change to the General Settings view and select **Client Configuration**.
   b. Specify a trigger mailbox.

   See “Modifying client configuration settings” on page 227 for more information about how to specify the trigger mailbox.
6. If you want to retrieve, search, and restore documents, check the following information:
   a. Click **General Settings** to change to the General Settings view and select **Web Application**.
   b. Check the settings in the right pane. See “Modifying the settings for the Web Application” on page 225 for more information.
   c. Then, select **Archived Data Access**.
   d. Check the settings in the right panes. See “Configuring the access to archived data” on page 228 for more information.
7. Make help information available to users of the search application:
   a. Click **General Settings** to change to the General Settings view and select **Information Center**.
   b. Check the settings in the right pane.

   See “Modifying the information center settings” on page 225 for more information.

   Alternatively, you can also access the information center on the website http://publib.boulder.ibm.com/infocenter/email/v2r2m0/index.jsp.
8. Provide information about the Configuration Web Service:
   a. Click **General Settings** to change to the General Settings view and select **Configuration Web Service**.
   b. Check the settings in the right pane.

   See “Modifying the Configuration Web Service settings” on page 224 for more information.

Now, you can set up task routes. Content Collector supplies templates with a single task route and supplies template bundles that contain several task route templates. To deploy one or more templates follow the instructions in the topics about configuring task routes.

Related tasks

"Configuring task routes" on page 270

Setting the Content Collector environment variables

Environment variables may be set to control the runtime environment of IBM Content Collector.

The following environment variables are available:

- AFUDominoAgentPreload
- AFU_ADJUST_PID
- AFU_ADVANCEDEMAILCONFIG
- AFU_DATE_RANGE_IN_MONTHS
- AFU_DISABLE_REPOSITORY_DETECTION
- AFU_DISABLE_URL_CHECK
- AFU_DOMINO_NO_ICON_CHANGE
- AFU_END_DATE_RANGE_IN_MONTHS
- AFU_MAILBOXID_ENABLE_EXACT_MATCH
- AFU_NO_SUPPORT_INFO
- AFU_OWA_DEFAULT_EXCH2003_HTTP_VIRTUAL_SERVER_DNS
- AFU_OWA_DEFAULT_EXCH2007_WEBSITES_METABASE_PATHES
- AFU_PREVIEW_MODE_ONLY
- AFU_REVERSE_DNS
- IBMAFUEXPIRATIONMGR
- IBMAFUFNCEROOT
- IBMAFUROOT
- IBMAFUTRACELOCATION
- IBMAFU_NOTESDATABASE_HOOK_LOGLEVEL

AFUDominoAgentPreload

If this environment variable is set, the Notes Client loads the classes for interactive operations when a user opens the database.

This variable is used in the Initial Configuration for template enablement for Lotus Notes.

AFU_ADJUST_PID

If this environment variable is set, Content Collector checks the item's persistent ID for validity and, if required, adjusts it before the item is retrieved from the archive.

This variable is used by the Web Application for retrieve requests

AFU_ADVANCEDEMAILCONFIG

If you use the embedded Web Application Server and want to track processing for troubleshooting purposes, set the environment variable AFU_ADVANCEDEMAILCONFIG to 1. After you have set the variable, you must restart the Configuration Manager.

Setting this variable to 1 will lead to the creation of a log file for each script that runs during the automatic web service configuration, and also logs when the web server is started and stopped. The last configuration you saved is also logged.

All log files are written to the directory that the TEMP environment variable points to. The following naming convention is used: `AdvancedEmailConfig.ConfigName.xml` or `AdvancedEmailConfig.ConfigName.log`

This variable is used to create Configuration Web Service log files

AFU_DATE_RANGE_IN_MONTHS

By setting the system environment variable AFU_DATE_RANGE_IN_MONTHS you can define the default date range that is displayed in the date fields of the Email Search page. The environment variable is of type integer and can have the following values:

0 No default date range is displayed.

n The value must be greater than 0. The default date range is the first day of the "current minus *n*" month until the current date. For example, if the current date was November 30th, 2009 and AFU_DATE_RANGE_IN_MONTHS is set to 1, the default date range that is displayed on the search page is October 1st, 2009 until November 30th, 2009.

If this environment variable is not set or set to a value less than 0, no default date range is displayed.

This variable is used by the email search function.

AFU_DISABLE_REPOSITORY_DETECTION

This environment variable is used for requests that retrieve documents from the archive that were archived with previous versions of Content Collector. If this environment variable is set to any value, the default repository is used for the request.

If this environment variable is not set, Content Collector checks whether the request contains any repository information. If the request contains such information, this information is used. If the request does not contain any repository information, the default repository is used.

AFU_DISABLE_URL_CHECK

If your configuration includes both Lotus Notes and Microsoft Exchange collections, set the system environment variable AFU_DISABLE_URL_CHECK to ensure that all Content Collector clients can display the e-Mail Search page. You can use any value for the variable. If the system environment variable is not set and your configuration files contain only Domino collections, all Content Collector clients can start the e-Mail Search function. If your configuration files contain Exchange collections, only Exchange clients that have IBM Content Collector, Version 2.1.1, installed can start the e-Mail Search function.

This variable is used by the email search function.

AFU_DOMINO_NO_ICON_CHANGE

The Email Connector Notes interface always sets $ContentIcon along with setting the message state. Because of the display order selected in the standard Notes email template, $ContentIcon overrides other icons, such as the attachment icon (paperclip icon). You can disable the icon change on the IBM Content Collector Server side by setting the system environment variable AFU_DOMINO_NO_ICON_CHANGE to 1.

When this environment variable is set, the Email Connector Notes interface sets only the message state and not $ContentIcon. In this way, other icons that might be displayed are preserved.

Restriction: This option prevents only IBM Content Collector Server from changing the icon along with the message state. If the Lotus Notes email template was enabled with IBM Content Collector functionality, such as Mark for Archiving or Mark for Stubbing, these message state changes still modify $ContentIcon. This might lead to unwanted results, because in this case, the icon might not match the message state anymore.

AFU_END_DATE_RANGE_IN_MONTHS

By setting the system environment variable AFU_END_DATE_RANGE_IN_MONTHS you can define an end date for the default date range in the Email Search page that lies in the past. The start date of the date range is calculated based on the setting of the environment variable AFU_DATE_RANGE_IN_MONTHS.

The environment variable AFU_END_DATE_RANGE_IN_MONTHS is of type integer and must be greater than zero (0). The default end date is the current date minus *n* months.

The start date of the date range is calculated by subtracting *x* months from the calculated end date, where *x* equals the value of the environment variable AFU_DATE_RANGE_IN_MONTHS if this variable is set. Otherwise, *x* equals the value of the environment variable AFU_END_DATE_RANGE_IN_MONTHS.

Table 39. Examples for calculating date ranges

Current date	AFU_END_DATE_ RANGE_IN_MONTHS	AFU_DATE_RANGE_ IN_MONTHS	End date	Start date
October 31st, 2010	Set to 24	Not set	October 31st, 2008	October 31st, 2006
October 31st, 2010	Set to 12	Set to 5	October 31st, 2009	May 31st, 2009

This variable is used by the email search function.

AFU_MAILBOXID_ENABLE_EXACT_MATCH

When you make IBM Content Manager repositories that were previously populated by IBM CommonStore available for search and restoring, set the system environment variable AFU_MAILBOXID_ENABLE_EXACT_MATCH. You can use any value. This is required for Microsoft Exchange collections only.

This variable is used by the email search function.

AFU_NO_SUPPORT_INFO

To have the Email Connector and the SMTP Connector create the `SUPPORT` subdirectory for temporary files only if a processing error occurred for an email, set this environment variable to 1.

AFU_OWA_DEFAULT_EXCH2003_HTTP_VIRTUAL_SERVER_DNS

This environment variable denotes the distinguished names of the specified default Exchange 2003 HTTP virtual servers. On an IIS server where Exchange 2003 Server is installed, there might be several Exchange web sites. Content Collector OWA Support can support only the default web

site. If the value of a web site's distinguishedName property in the Active Directory starts with the string "CN=1,", this web site will be recognized as default web site.

You can define any web site as default we bsite by setting the value of AFU_OWA_DEFAULT_EXCH2003_HTTP_VIRTUAL_SERVER_DNS to the value of this webs ite's distinguishedName property as it is defined in the Active Directory.

To define more than one web site as default web sites, separate the values by a semicolon (;). For example:

```
set AFU_OWA_DEFAULT_EXCH2003_HTTP_VIRTUAL_SERVER_DNS=CN=101,
CN=HTTP,CN=Protocols,CN=CSX-DC2,CN=Servers,CN=First Administrative Group,
CN=Administrative Groups,CN=CSX DEV Organization,CN=Microsoft Exchange;
CN=102,CN=HTTP,CN=Protocols,CN=CSX-DC,CN=Servers,
CN=First Administrative Group,CN=Administrative Groups,
CN=CSX DEV Organization,CN=Microsoft Exchange
```

This variable must be defined as system environment variable on the server where Outlook Web Access Service is installed. After you added the variable or changed its value, Outlook Web Access Service must be restarted to enforce the new values to be used by Outlook Web Access Service.

This variable is used by Outlook Web Access Service.

AFU_OWA_DEFAULT_EXCH2007_WEBSITES_METABASE_PATHES

This environment variable controls the metabase of Exchange 2007 web sites. Use this variable to specify a set of search paths for metabase information of the specified default Exchange 2007 web sites. On an IIS server where Exchange 2007 Server is installed, there might be several Exchange web sites.Content Collector OWA Support can support only the default web site. If the value of a website's msExchMetabasePath property in the Active Directory contains the string "/W3SVC/1", this website will be recognized as default web site.

You can define any web site as default web site by setting the value of AFU_OWA_DEFAULT_EXCH2007_WEBSITES_METABASE_PATHES to the value of this web site's msExchMetabasePath property as it is defined in the Active Directory.

To define more than one web site as default web sites, separate the values by a semicolon (;). For example:

```
set AFU_OWA_DEFAULT_EXCH2007_WEBSITES_METABASE_PATHES=
IIS://ex2k7-csx84.csx.ibm.com/W3SVC/1360126940/ROOT;
IIS://csx-cas.csx.ibm.com/W3SVC/1360126840/ROOT
```

This variable must be defined as system environment variable on the server where Outlook Web Access Service is installed. After you added the variable or changed its value, Outlook Web Access Service must be restarted to enforce the new values to be used by Outlook Web Access Service.

This variable is used by Outlook Web Access Service.

AFU_PREVIEW_MODE_ONLY

When a user clicks on a preview link in a stubbed Outlook email, the Web Application sends the file content of that message to the client and the Outlook client displays it. This will not work in an environment where the Exchange server supports Unicode message file content and users work

with older Outlook clients that do not support Unicode format. Instead of an Outlook window that displays the message, users get errors when they click a preview link.

If you set this environment variable on the machine that hosts the Web Application, the Web Application will not send the message file back to the caller. Instead, the email preview page is displayed in the browser. You can set the environment variable to any value.

This variable is used for previewing in Outlook.

AFU_REVERSE_DNS
If you set this environment variable to 1, the mail connector performs a reverse DNS lookup to validate the hostname when creating a stub.

This variable is used by mail connectors.

IBMAFUEXPIRATIONMGR
On your local machine, set the environment variable IBMAFUEXPIRATIONMGR to point to the absolute path of the directory `ExpirationManager`. For example, this could be `C:\Program Files\IBM\ExpirationManager`.

This variable is used by Expiration Manager.

IBMAFUFNCEROOT
For FileNet P8, set the environment variable IBMAFUFNCEROOT on your local machine to point to the installation directory where the IBM FileNet Content Engine server or client is installed, for example `C:\Program Files\IBM\FileNet\CEClient` or `C:\Program Files\IBM\FileNet\ContentEngine`.

This variable is used by Expiration Manager.

IBMAFUROOT
This variable is set by the installer and points to the directory where you installed Content Collector.

IBMAFUTRACELOCATION
By default, Content Collector log files for a Notes client are written to the default log file directory `%UserProfile%\IBM\ContentCollector`. To change the log file location, set this environment variable in the user's `notes.ini` file to point to a different directory.

Tip: Back up the original `notes.ini` file before you edit it.

IBMAFU_NOTESDATABASE_HOOK_LOGLEVEL
This environment variable controls the trace level of the offline repository support and the trace level for automatically retrieving and displaying documents. The value for this environment variable can be set to 0, 1, or 2, where 0 provides the least details and 2 the most details.

If required for troubleshooting offline repository support or automatic retrieval, set this variable on the client.

Installing Content Collector on several servers

By installing IBM Content Collector on several servers, also known as *scale out*, you can distribute the workload to several servers.

You would consider a scale out setup using multiple Content Collector servers if the required throughput of the system exceeds the capacity of a single server. The

primary driver in a scale out setup is typically used for archiving throughput. If a scale out setup is used to provide additional capacity to service interactive requests, such as document preview, search, and restore requests from users, a load balancer is required to distribute work across the servers.

Note, however, that system performance also depends on your hardware. If the hardware on which Content Collector runs has reached its limits, an extension does not help.

You must first install the primary node and thereafter one or more secondary nodes. Collectors run on the primary node only, so the primary node searches the source systems for documents to process as defined in a task route. The primary node then either does all the processing defined in the task route for a document itself or distributes the processing of a document to an extension node. If the primary node fails or if the connection to the primary node is lost, an extension node takes over as the primary node. The only difference for the task route services and the connector services between the primary node and an extension node is that the services are started on the primary node first.

All nodes in a scale-out environment must be set up identically. This means that all nodes must have the same operating system and that all files and directories that are relevant for IBM Content Collector (such as log files, temporary files, working directories, or audit logs) must be written to the same location on all nodes. If you define using environment variables, you must make sure that the variables are the same on all nodes.

It is also important that all prerequisites are installed in the same locations on all systems. The administration and monitoring is greatly simplified if all the servers belong to the same Windows domain. Doing this, for example, ensures that the system clocks are synchronized and that the same user accounts can be used to run the Content Collector services across the entire setup.

SharePoint and SMTP only: The Microsoft SharePoint and SMTP Connectors require a shared file system location that is accessible by the user of the connector service from all of the IBM Content Collector servers. The location is typically specified using a UNC path such as *\\ComputerName\SharedFolder*. This shared location must be highly available to preserve the high-availability characteristics of a Content Collector scale-out setup. In other words, if you use a local drive on one of the Content Collector servers as your shared location, the whole IBM Content Collector setup will fail if the server providing this shared location fails. Furthermore, the performance of the IBM Content Collector setup depends greatly on the speed of this shared file system.

Tip: Do not process PST files in a scale-out environment because the nodes would compete for exclusive write access to the files. Work would not be distributed to other nodes because only a single node can write to a given file at a time. Use a single node setup to handle PST files.

Configuring the primary node

Configure the node that you want to use as primary node. The primary node is responsible for collecting documents from the source systems to be processed. In addition, it is responsible for distributing the task of processing the document by a task route.

You completed one of these tasks on the machine designated as the primary node:

- "Installing Content Collector for use with one or more source systems and Content Manager" on page 59
- "Installing Content Collector for use with one or more source systems and FileNet P8" on page 60

Complete these steps to configure the primary node:

1. Open the Configuration Manager: Click **Start → All Programs → IBM Content Collector → Configuration Manager**.
2. In the IBM Content Collector window, click **Tools → Task Route Service Configuration**.
3. In the Task Route Service Configuration window, enable database synchronization: In the **Datastore synchronization interval** field, specify a value, in seconds, greater than zero. The default value is `300` seconds.
4. To ensure that the primary node can communicate with the extension nodes, configure the IBM Content Collector Task Routing Engine service to run under a user account. Complete these steps:
   a. Open the Services window in Microsoft Windows.
   b. Select **IBM Content Collector Task Routing Engine** in the right pane and then click **Action → Properties**.
   c. In the window that opens, click the **Log On** tab.
   d. In the **This account** field on the Log On page, specify the user account under which the service is to run, and enter the password for the account.

      The user account must have a trust relationship between the primary node and the secondary node. Either the nodes must be members of a domain (recommended on Windows 2008 with User Account Control enabled) or the IBM Content Collector Task Routing Engine service user must have the following permissions:
      - Start the connector services
      - Connect to the pipes the connector services create
      - Create registry keys under `HKLM\SYSTEM\CurrentControlSet\Services\EventLog`

      You are advised to use the same user account that you used for the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service because the IBM Content Collector Web Application service needs the same privileges as the IBM Content Collector Email Connector service to restore email.

Make sure that this node can connect to the source systems and to the repositories. Check whether you can archive documents. In addition, ensure that the host name of the server on which the Web Application is installed is specified correctly. This is necessary because the links in the stubs refer to the Web Application.

Related tasks

"Configuring the task route service" on page 174

"Changing the user account of a service" on page 213

Related reference

"Content Collector services" on page 178

Configuring the extension nodes

Configure one or more nodes that you want to use as extension nodes so that several servers can handle the workload for files, a single mailbox, a set of mailboxes, or to have a backup if the primary node fails.

You completed one of these tasks on the machine designated as the extension nodes:

- "Installing Content Collector for use with one or more source systems and Content Manager" on page 59.
- "Installing Content Collector for use with one or more source systems and FileNet P8" on page 60.

You must also ensure that each planned extension node runs on the same platform as the primary node. You must install IBM Content Collector and all prerequisite software using the same paths that were used for the primary node.

Complete these steps to configure an extension node:

1. If the primary node is configured to write log files, temporary files, working directories, or audit logs to any location other that the default folders, ensure that these folders exist under the same path on the extension node.
2. Configure the extension node to connect to the same database as the primary node: Copy the `CTMSConfigStore.xml` file in the subdirectory `ctms` where IBM Content Collector is installed to the same location on the extension node.
3. If you use DB2, catalog the database locally.
4. If you use Lotus Domino as a source system, copy your Lotus Notes ID file and `notes.ini` file from the subdirectory `notesdata` where IBM Content Collector is installed on your primary node to the same location on the extension node.

   The email connector searches this location for the Lotus Notes ID and `notes.ini` files.
5. To check whether the database is connected, run the Configuration Manager. On an extension node, the Configuration Manager runs in read-only mode. A connection to the database is successfully established if the Configuration Manager loads without error messages, and displays the same configuration data as the primary node.

   Important: Never run the initial configuration on an extension node.
6. To ensure that the primary node can communicate with the extension node, configure the IBM Content Collector Task Routing Engine service to run under the same user account as on the primary node. Communication is also ensured because the configuration database is synchronized at given intervals (the default is every 300 seconds) which means that the extension nodes will contact the primary node at defined times to perceive configuration changes and to take responsibilities of the distribution of the workload. Complete these steps:
   a. Open the Services window in Microsoft Windows.
   b. Select **IBM Content Collector Task Routing Engine** in the right pane and then click **Action → Properties**.
   c. In the window that opens, click the **Log On** tab.
   d. In the **This account** field on the Log On page, specify the same user account as on the primary node, and enter the password for this account.
   e. If you use the File System Source or SharePoint connectors, you must ensure that the users of the File System Source or SharePoint services, respectively, have Read/Write access to the document content file, which for SharePoint is the connector's temporary file location. You must also ensure that users of the P8 4.x Connector service, CM 8.x Connector service, and Utility Connector service (if you use them) have Read access to the document content file.
   f.

If you use Microsoft Exchange as the source system, you can use the same account as for the IBM Content Collector Email Connector service.

7. For any additional extension nodes, repeat step 1 on page 90 to step 6 on page 90 for each node.

Make sure that each extension node can connect to the source systems and to the repositories.

Related tasks

"Changing the user account of a service" on page 213

Related reference

"Content Collector services" on page 178

Starting the IBM Content Collector Task Routing Engine service on the primary node and on the extension nodes

After you install and configure the primary node and install the extension nodes, start the IBM Content Collector Task Routing Engine service on each node.

Complete these steps:

1. On the primary node, click **Start** → **All Programs** → **IBM Content Collector** → **Start Services** → **Start Task Routing Engine**.
2. Repeat step 1 on each extension node.

The IBM Content Collector Task Routing Engine service on the primary node recognizes any registered extension nodes and begins forwarding work to these extension nodes. Each IBM Content Collector server registers itself in the configuration database. This registration is renewed every time the configuration database is synchronized.

The first node that is started (typically the intended primary node) begins processing and becomes the acting primary node. The next time the configuration database is synchronized, the acting primary node recognizes the additional nodes and starts distributing work to these newly registered nodes. If a node fails to update its availability status in the configuration database then its registration will expire. The acting primary node will notice this the next time the configuration database is synchronized and stop distributing work to that node. It may take two database synchronization intervals before the system notices that a node has expired. If the acting primary node expires, the first secondary node that realizes this situation within the next two synchronization intervals of the database, will become the active primary node and will take over the job of collecting and archiving items from the source systems. Depending on the task route schedules for collecting and archiving from the source systems, collecting may be delayed until the next scheduled interval.

Related tasks

"Configuring the task route service" on page 174

Configuring the SMTP Receiver to run on several servers

If you archive SMTP/MIME email and want the SMTP Receiver to run on several nodes instead of just one, configure a load balancer. You can set up a Microsoft Network Load Balancer cluster, for example. Alternatively, you can use other software load balancing software or a hardware load balancer.

Running the SMTP Receiver on more than one node can increase availability and performance. If the SMTP Receiver runs on the primary node only, which is the default, all incoming email must go through this node. To ensure availability in

this case, the primary node must run in a high availability cluster. Alternatively, you can run several SMTP Receiver components and use a highly available load balancing mechanism to distribute the incoming email between the available receivers.

If email frequently cannot be delivered to IBM Content Collector and the queue of outgoing messages on the originating email server grows, this indicates that there might be a bottleneck on the SMTP Receiver disk storage subsystem, processing power, or the network connection between the mail server and the SMTP Receiver. In this case, you should determine the reason for the performance problems. In some cases, it might help to increase the CPU capacity on the primary node or to dedicate one node to the SMTP Receiver, so that no other Content Collector services run on this node. The file system performance for the SMTP Receiver message queue directory is often the most likely bottleneck.

Tip: For best performance, use the UNC path to access the message queue directory only if you are running on the latest version of the Windows operating system. With older operating systems, unless Content Collector runs in a scale-out environment, use a local path to a direct attached storage drive.

To distribute receiving email, make the SMTP Receiver component available on multiple Content Collector nodes. The incoming email is then received by SMTP Receiver instances on different nodes and written to the message queue directory, a storage area that holds the email until it is processed.

To run the SMTP Receiver on several nodes, you must deploy a network load balancer that distributes the incoming SMTP stream to the multiple SMTP Receiver instances. The following topics exemplify how to set up a Microsoft Network Load Balancer (NLB) cluster, but you can use a different load balancer as well. The NLB software, which is contained in every Microsoft Windows Server, distributes the TCP/IP traffic for the IBM Content Collector SMTP Connector between the nodes in the cluster.

Setting up a Network Load Balancing Cluster:

Set up a Microsoft Network Load Balancing (NLB) cluster to distribute the TCP/IP traffic for the IBM Content Collector SMTP Receiver between several nodes, the primary node and at least one extension node.

The Microsoft NLB software is included in all Microsoft Windows Server packages. Refer to the Network Load Balancing Manager Help for detailed information about the NLB software.

You need a virtual IP address for the NLB cluster. This virtual IP address should have a virtual host name (or cluster name) assigned, which you can use in the IBM Content Collector configuration.

If possible, use a distinct network interface for the NLB cluster that is dedicated to cluster traffic from, to, and between the clients.

The message queue directory that the SMTP Connector uses to store email after it is received by the SMTP Receiver and before it is processed must be accessible from all nodes of the cluster. For example, you can use a network share that is hosted on one IBM Content Collector node and is used by all other nodes, or preferably a highly available network share that can be accessed from all nodes

and is provided by a network attached server device or a server cluster. However, note that the NLB software cannot be run on a machine in a Microsoft Server Cluster.

Important: The message queue directory must be available at all times. If data in this storage location is corrupted or lost, email that has been received by the SMTP Receiver but that has not yet been processed might be lost.

To set up an NLB cluster:

1. Start the Microsoft Network Load Balancing Manager. Click **Start → All Programs → Administrative Tools → Network Load Balancing Manager**, or run the command **nlbmgr.exe**.
2. From the menu, select **Cluster → New**.
3. In the Cluster Parameters window, specify how the cluster can be reached from the outside.
   a. Under **IP address**, specify the virtual IP address for the cluster. This must be a static IP address that is assigned to all nodes of the cluster automatically.
   b. Under **Subnet mask**, specify an appropriate subnet mask for the IP address.
   c. Unter **Full Internet name**, specify the virtual host name for the cluster. This is the name that you use for the configuration of IBM Content Collector later. All SMTP/MIME email that should be archived must be sent to this host.
4. Click **Next**.
5. Optional: In the Cluster IP Addresses window, you can specify additional IP addresses under which the cluster can be reached.
6. Click **Next**.
7. In the Port Rules window, specify the ports that are load balanced.
   a. Select the existing port rule and click **Edit...**.
   b. Define a port rule that matches your configuration. The default SMTP port for the SMTP Receiver is TCP port 25. To allow for balancing over all available nodes, set the affinity to **None**.
   c. Click **OK**.
8. Click **Next**.
9. In the Connect window, select the network interface that you want to use for the cluster communication. This interface is used for all cluster communication, internally and externally. The cluster IP address that you specified is assigned to the network interface, in addition to the original IP address.
10. Click **Next**.
11. In the Host Parameters window, specify the host parameters for the first node that you want to add to the cluster.
    a. Under **Priority**, specify a priority for the current host. A cluster can consist of up to 32 hosts. Priorities for all hosts must be unique in the cluster.
    b. Under **Dedicated IP configuration**, specify the IP configuration that is used for cluster internal communication between the cluster nodes. Specify the original IP address of the cluster network interface.
12. Click **Finish**.
13. To add more hosts to the cluster, select **Cluster → Add Host** from the menu.

Setting up Content Collector to run in a Network Load Balancing cluster:

All SMTP/MIME email that is sent to the Microsoft Network Load Balancing (NLB) cluster is automatically load balanced by the NLB cluster. All received email is written to the message queue directory. Therefore, the IBM Content Collector SMTP Connector requires the same configuration for a scale-out environment as for a single-node setup.

The SMTP Connector and the SMTP Receiver are automatically installed during the IBM Content Collector installation, so they are available on the extension nodes as well. Use the Configuration Manager on the primary node to update the configuration. The SMTP Receivers that are running on the extension nodes pick up these settings automatically during startup. Make sure to restart the services on all nodes after you change the configuration.

To configure the SMTP Connector:

1. In the Configuration Manager, switch to the Connectors view and select **SMTP**.
2. On the Connections tab, specify a message queue directory where the received email is stored before it is processed. Use the Universal Naming Convention (UNC) syntax to specify the path.
3. Save the changes and restart the IBM Content Collector SMTP Receiver service on all nodes.

Configuring the Web Application Server

Depending on the type of Web Application Server that you use with IBM Content Collector, the configuration tasks differ. Some of the configuration tasks are mandatory and some are optional.

If you are using the embedded Web Application Server with IBM Content Collector, it is recommended that, for security reasons, you replace the default Hypertext Transfer Protocol Secure (HTTPS) certificates. If you want to use a Web Application Server other than the embedded one, you have to configure that manually. Optionally, you can change the ports that are used for web applications.

IBM Content Collector installs IBM WebSphere Application Server, Version 7.0 for the embedded Web Application Server. If you want to work with a Web Application Server other than the embedded one, IBM WebSphere Application Server, Version 6.1 and Version 7.0 are supported.

Note: IBM Content Collector supports the Base edition of IBM WebSphere Application Server for Windows. The Network Deployment edition or editions for other operating systems are not supported.

Comparing the embedded Web Application Server to an external Web Application Server

The information in the following table helps you decide which type of Web Application Server to choose over the other. It is possible, however, to run several Web Application Server instances in parallel as long as you omit any port conflicts.

Table 40. Comparing embedded and external Web Application Servers

Embedded Web Application Server	External Web Application Server
All required Web Application Server components are part of the IBM Content Collector server installation. Unnecessary components are excluded so that less disk space is used for the installation and less memory is required to run the application.	A Web Application Server might already be installed and you might not want to implement another one. However, you must create a separate server profile for IBM Content Collector. This Web Application Server instance will be operating within a separate instance of the Java virtual machine (JVM). Therefore, the external Web Application Server does not provide better performance than the embedded Web Application Server.
The runtime environment for the Content Collector Web applications is already configured and is deployed when the Content Collector server is installed.	You are responsible for installing, configuring, and maintaining the Web Application Server.
The embedded Web Application Server is intended to primarily provide application support and does not support clustering. Workload management is possible by using a load balancer.	The Web Application Server can be a member of a cluster.
All administration is done via command scripts (wsadmin commands). Key management is possible only by using the IBM Key Management utility.	You can use the administrative console to manage the Web Application Server.

Using an existing Web Application Server

Read this information if you intend to use a Web Application Server with IBM Content Collector other than the embedded Web Application Server that is included in the installation. In this case, you have to create an additional application server profile and then deploy the respective EAR files as described here.

Check the prerequisites for this task:

- If you are upgrading from IBM InfoSphere Content Collector, Version 2.1, you must have completed all upgrade steps up to and including the step for upgrading the Web Application Server installation for Content Collector before you copy and install the Web Application Server files for IBM Content Collector, Version 2.2.
- IBM Content Collector Server must be installed on one of the machines in your topology
- IBM WebSphere Application Server must be installed.

  Important: IBM WebSphere Application Server must run as a 32-bit application on Windows.
- A TCP/IP connection must exist between IBM Content Collector Server and the Web Application Server.
- If the Web Application Server instance runs on a machine other than the one on which IBM Content Collector Server is installed, the following prerequisites apply:
  - Depending on the database management system that you use for the configuration database, a respective client must be installed on the computer where the Web Application Server is installed.

- Depending on your archive system, the following clients must be installed on the computer on which the Web Application Server is installed:
  - For an IBM Content Manager Enterprise Edition 8.x repository, DB2 Information Integrator for Content must be installed.
  - For an IBM FileNet P8 repository, IBM FileNet P8 Content Engine client API must be installed.
- For a Microsoft Exchange mail system, Microsoft Outlook must be installed on the computer on which the Web Application Server is installed.
- For a Lotus Domino mail system, a Lotus Domino server must be installed on the computer on which the Web Application Server is installed.
- If your mail system is Lotus Domino, the Web Application needs access to the `notes.ini` file. Therefore, the directory path to this file on the computer on which the Web Application Server is installed must be the same as on the computer on which IBM Content Collector Server is installed.

Follow these steps to enable an existing Web Application Server with IBM Content Collector:

1. "Preparing the installation of the Web applications"
2. "Installing the Web applications" on page 97
3. "Configuring the Web Application Server profile" on page 98
4. "Verifying the Web Application Server configuration" on page 99

Preparing the installation of the Web applications:

Before you can install the Content Collector Web applications, you must copy the files that are required for the installation. Depending on the version of IBM WebSphere Application Server that you use, you might also have to adapt the service name.

Follow these steps to set up the system for the installation of the Content Collector Web applications.

1. Copy all .dll files from the directory `<ICCinstallDir>\lib` on the computer on which IBM Content Collector Server is installed to a directory on the computer running the Web Application Server, where `<ICCinstallDir>` is the installation directory of IBM Content Collector Server. The directory that contains the .dll files must be included in the list of directories that the global PATH environment variable points to. Note that you must restart the WebSphere Application Server for these libraries to be available to the Web applications.
2. Make sure that the Web Application Server is started.
3. Copy all .jacl files from the `<ICCinstallDir>\ctms\script` directory to the home directory of your WebSphere Application Server installation, where `<ICCinstallDir>` is the installation directory of IBM Content Collector Server. The home directory of WebSphere Application Server is the one that the WASHOME environment variable points to. Usually this is `<WASinstall>\AppServer`, where `<WASinstall>` is the path to your WebSphere Application Server installation directory.
4. Copy files as indicated from the `<ICCinstallDir>\AFUWeb` directory to the home directory of your WebSphere Application Server installation.

   Restriction: Due to a limitation in IBM WebSphere Application Server, there is a length restriction for directory paths. The maximum length for the path to the installation root directory is 60 characters. The maximum length for the path to

the installation root directory that contains the profiles is 80 characters. If one of the paths is too long, it is possible that no profile can be created.

Copy the following files:

```
afu-configurationWebservice.ear
afu_metadata_web.ear
afu_web.ear
afuEnv.bat
afu_ewas_install.bat
afu_ewas_install.xml
afu_ewas_uninstall.bat
afu_ewas_updateJVMProperties.jacl
portdef.props
afu_ewas_deploy.bat
afu_ewas_undeploy.bat
afu_ewas_addSharedLib.bat
afu_ewas_addSharedLib.jacl
afu_ewas_exchange_cert.bat
afu_ewas_editSessionManagement.jacl
afu_ewas_editSessionManagement.bat
DocViewer.ear
```

5. If you are working with IBM WebSphere Application Server, Version 6.1, you must edit the file `afuEnv.bat` and change the internal name of the service before running the installation scripts. Adapt the command script as follows:
   a. Enable this line by deleting the rem keyword (@rem).
      ```
      @rem set serviceNameAsWASCreatesIt="IBMWAS61Service - %internalServiceName:"=%"
      ```
   b. Disable this line by preceding it with the rem keyword (@rem).
      ```
      set serviceNameAsWASCreatesIt="IBMWAS70Service - %internalServiceName:"=%"
      ```

Now, you can install the Content Collector Web applications.

Installing the Web applications:

After you have completed the steps for preparing the system for the installation of the Web applications, you can install them.

To install the Web applications for IBM Content Collector:

1. Navigate to the WASHOME directory.
2. Edit the `afuEnv.bat` file and adapt the following parameters for the repository that you use.
   ```
   @set FNCEPATH=
   @set AFUEWASHOSTNAME=
   @set ConfigWebApplicationOnly=NO
   ```
   a. Adapt the FNCEPATH parameter.
      - If your repository is IBM Content Manager, do not set a value for the parameter.
      - If your repository is IBM FileNet P8 4.0.1, set the parameter to *<FileNetInstallDir>*`\FileNet\ContentEngine`, where *<FileNetInstallDir>* is the installation path for FileNet P8.

- If your repository is IBM FileNet P8 4.5, set the parameter to <*FileNetInstallDir*>\FileNet\CEClient, where <*FileNetInstallDir*> is the installation path for FileNet P8.

b. For the AFUEWASHOSTNAME parameter, specify the fully qualified host name of the machine where the WebSphere Application Server runs.

c. If you want to install only the Configuration Web Service but not the Web applications, set the ConfigWebApplicationOnly parameter to YES. Use this setting if SMTP is the only source system for which you installed Content Collector.

3. Run the afu_ewas_install.bat command file.
4. If command file ran successfully, this or a similar message is displayed:

   ```
   BUILD SUCCESSFUL
   ```

   - In this case, proceed with setting up the Web Application Server.
   - If one of the commands failed, follow the instructions in the troubleshooting section.

Now, you have to configure the profile for the Web Application Server.

Related tasks

"The installation of the Web applications failed" on page 803

Configuring the Web Application Server profile:

After installing the Content Collector Web applications, you have to configure the profile for the Web Application Server.

Follow these steps:

1. Start the Web Application Server with the new profile by entering the following command:

   ```
   .\bin\startserver.bat afuServer -profileName AFUWeb
   ```

   Where afuServer is the name of the server that was created by running the installation scripts.
2. Run the following commands:

   ```
   .\bin\wsadmin -f removeProviderSourceSecurity.jacl <nodename> <jdbc_provider>
   .\bin\wsadmin -f setVariables.jacl <nodename> <provider_type> <jdbc_location>
   .\bin\wsadmin -f createProvider.jacl <nodename> <jdbc_provider> <provider_type>
   .\bin\wsadmin -f modifyDatasource.jacl <nodename> <jdbc_provider> <afudb>
     <jdbc_port> <db_host> <db_user> <pwd> <provider_type> <db_url>
   ```

   Where:

   <nodename>
   Is the name of the node that is currently used by the Web Application Server.

   Tip: Run the command .\bin\wsadmin -c "$AdminControl getNode" to display the node name.

   <jdbc_provider>
   Is the name of the JDBC provider to be used:
   - For DB2, specify: "DB2 Universal JDBC Driver Provider (XA)"
   - For SQL Server, specify: "Microsoft SQL Server JDBC Driver (XA)"
   - For Oracle, specify: "Oracle JDBC Driver"

Note that the JDBC provider name must be enclosed in double quotation marks.

<provider_type>
Is one of the following provider types:

DB2 For DB2

MSSQL
For SQL

ORACLE
For Oracle

<jdbc_location>
Is your directory containing JDBC files, for example `"C:\\IBM\\SQLLIB\\java"`

Note the double backslashes. They are required.

<afudb>
Is the name of your configuration database (data store)

<jdbc_port>
Is the port number to be used to connect to your configuration database

<db_host>
Is the name of the computer on which the database is located

<db_user>
Is the name of a database administrator, for example db2admin

<pwd>
Is the password of the database administrator

<db_url>
Is the URL of the database. This parameter is mandatory although it is used for Oracle databases only. For the other databases, this parameter is ignored. For DB2 and SQL Server, specify any string enclosed in double quotation marks. For Oracle, specify the URL in the form "<db_host>:<jdbc_port>:<SID>", where <SID> is the database system identifier. Note that the double quotation marks are required.

3. Stop the Web Application Server by entering the following command:
   ```
   .\bin\stopserver.bat afuServer -profileName AFUWeb
   ```
   where afuServer is the name of the server that was created by running the installation scripts.
4. Start the Web Application Server by entering the following command:
   ```
   .\bin\startserver.bat afuServer -profileName AFUWeb
   ```

Now, check the configuration.

Important: If you use **Microsoft Exchange**, the Content Collector Web Application service must be started by an account with administrator privileges for Microsoft Exchange. The administrator privileges must be the same as for the Email Connector, usually "Exchange Organization Administrators".

Verifying the Web Application Server configuration:

After having completed the installation and configuration steps for the Web Application Server, verify the proper setup.

To verify the setup of the Configuration Web Service:

1. Check your configuration by calling the following URL from a web browser:

   ```
   https://<your_AFUWebApp_Server>:11443/AFUConfig/
   Configuration?type=ibm.ctms.configWebService&unique=default
   ```

   where <your_AFUWebApp_Server> is the host name of the machine on which you installed the IBM Content Collector Web Application. If you start the browser on the same machine, <server_name> can be localhost.

   If the Web Application Server is able to access the configuration database, an XML document is returned that contains the settings of the Configuration Web Service. Depending on the web browser that you use you might be asked to open or save the resulting XML document.
2. As a second verification step, you can start an email client with the IBM Content Collector extensions. The extension should be able to read information from the IBM Content Collector configuration data store.

Changing the port for the Web Application Server

If the default ports that IBM Content Collector uses already are in use on the machine that is to host the web applications, you can change the port numbers.

If you want to work with an existing Web Application Server, it is recommended that you change the port numbers before running the installation scripts for the Web Application Server. If you must redefine the ports after installation, work with the scripts that are provided with IBM WebSphere Application Server. In this case, see the topic about updating ports in existing profiles in the IBM WebSphere Application Server documentation:

- For IBM WebSphere Application Server, Version 6.1, see http://publib.boulder.ibm.com/infocenter/wasinfo/v6r1/index.jsp?topic=/com.ibm.websphere.base.doc/info/aes/ae/tins_updatePorts.html
- For IBM WebSphere Application Server, Version 7.0, see http://publib.boulder.ibm.com/infocenter/wasinfo/v7r0/index.jsp?topic=/com.ibm.websphere.installation.base.doc/info/aes/ae/tins_updatePorts.html

To change the port numbers before installing the Web applications on an existing Web Application Server, do the following:

1. Edit the file `portdef.props`, which you copied to the home directory of your WebSphere Application Server installation, and change the port numbers to numbers that are not in use on the machine.
2. Proceed as described in the topic about using an existing Web Application Server.

Re-configuring the Web Application Server

Whenever you change any information that is related to the configuration database, for example, the user name and the password, or when you install or reinstall the IBM Content Collector Web Application, you have to re-configure the Web Application Server so that it can use the JDBC connection that is defined in your configuration database.

1. Run the following commands:

   ```
   .\bin\wsadmin -f removeProviderSourceSecurity.jacl <nodename> <jdbc_provider>
   .\bin\wsadmin -f setVariables.jacl <nodename> <provider_type> <jdbc_location>
   .\bin\wsadmin -f createProvider.jacl <nodename> <jdbc_provider> <provider_type>
   .\bin\wsadmin -f modifyDatasource.jacl <nodename> <jdbc_provider> <afudb>
     <jdbc_port> <db_host> <db_user> <pwd> <provider_type> <db_url>
   ```

   Where:

<nodename>
Is the name of the node that is currently used by the Web Application Server.

Tip: Run the command `.\bin\wsadmin -c "$AdminControl getNode"` to display the node name.

<jdbc_provider>
Is the name of the JDBC provider to be used:

- For DB2, specify: "DB2 Universal JDBC Driver Provider (XA)"
- For SQL Server, specify: "Microsoft SQL Server JDBC Driver (XA)"
- For Oracle, specify: "Oracle JDBC Driver"

Note that the JDBC provider name must be enclosed in double quotation marks.

<provider_type>
Is one of the following provider types:

DB2 For DB2

MSSQL
For SQL

ORACLE
For Oracle

<jdbc_location>
Is your directory containing JDBC files, for example `"C:\\IBM\\SQLLIB\\java"`

Note the double backslashes. They are required.

<afudb>
Is the name of your configuration database (data store)

<jdbc_port>
Is the port number to be used to connect to your configuration database

<db_host>
Is the name of the computer on which the database is located

<db_user>
Is the name of a database administrator, for example db2admin

<pwd>
Is the password of the database administrator

<db_url>
Is the URL of the database. This parameter is mandatory although it is used for Oracle databases only. For the other databases, this parameter is ignored. For DB2 and SQL Server, specify any string enclosed in double quotation marks. For Oracle, specify the URL in the form "<db_host>:<jdbc_port>:<SID>", where <SID> is the database system identifier. Note that the double quotation marks are required.

2. Restart the Web Application Server.

Replacing certificates for the embedded Web Application Server

The embedded Web Application Server creates a set of default Secure Sockets Layer (SSL) certificates with default credentials. These are used for the initial configuration of the Web Application Server. To enable a secure and trusted

environment, you must replace these certificates and credentials with certificates signed by a trusted certificate authority, especially in a production environment.

Prerequisites:

- The IBM Content Collector server must be installed.
- The IBM Content Collector Web Application service must have been deployed.

Before you can replace an SSL certificate, you have to request a new certificate. You request, receive, and replace SSL certificates for the embedded Web Application Server by using the IBM Key Management utility. You can use the same utility to add certificates for additional web servers or to change the credentials when the currently used credentials have expired.

Note: If you do not use the embedded Web Application Server, you can create a certificate authority request and receive the signed certificate by using the WebSphere Application Server AdminTask object. How to do this is described in the IBM WebSphere Application Server documentation.

1. Request a new certificate.
   a. Log on to the computer on which the IBM Content Collector server is installed.
   b. In a command prompt, go to the `<ICCinstallDir>\AfuWeb\profiles\AFUWeb\bin` directory, where `<ICCinstallDir>` is the installation directory of the IBM Content Collector server.
   c. Type `ikeyman`

      The IBM Key Management utility opens.
   d. In the IBM Key Management utility, select **Key Database File → Open**.
   e. Select **PKCS12** as key database type.
   f. In the **File name** file, specify the file name `key.p12`.
   g. In the **Location** field, specify the `<ICCinstallDir>\AFUWeb\profiles\AFUWeb\config\cells\<cell name>\nodes\<node name>` directory.

      Replace <ICCinstallDir>, <cell name>, and <node name> with the proper values of your installation.
   h. Click **OK**.
   i. When prompted for a password, enter the password. Click **OK**. The default password is WebAS. Note that the password is case sensitive. In a production environment, change the password as described in the topic about updating default key store passwords using scripting in the WebSphere Application Server (Distributed platforms and Windows), Version 7.0 Information Center.
   j. Create a new certificate request. Under **Key database content**, select **Personal Certificates Requests** and click **New**.
   k. In the **Key Label** field, specify a label for the digital certificate request, for example, `Production Certificate for Content Collector`.
   l. For the remaining fields, accept the default values.
   m. Click **OK**. A confirmation window is displayed, verifying that you have created a request for a new digital certificate.

      The **Personal Certificate Requests** field in the IBM Key Management window shows the key label of the new digital certificate request you created.
   n. Send the file to a certificate authority (CA) to request a new digital certificate, or cut and paste the request into the request forms of the CA's

website. If you have a Windows Domain CA, you can follow the procedure described in "Submitting a certificate request" on page 104 to do so. If you use a different CA to certify the certificate request, follow the procedure that applies for the respective CA.

After the CA sends you a new digital certificate, you must delete the existing certificate and add the new one to the key database from which you generated the request.

2. Delete the existing certificate. Note: Before deleting a digital certificate, create a backup copy in case you later want to recreate it.
   a. In the IBM Key Management utility, make sure that the key database file is open and that, under **Key database content**, **Personal Certificates** and **default** are selected.
   b. Click **Delete**. You are asked to confirm the deletion. The label of the digital certificate you just deleted no longer appears in the **Personal Certificates** field of the IBM Key Management window.
3. Receive the new certificate to replace the existing one.
   a. Click **Receive**. The Receive Certificate from a File window is displayed.
   b. Select **Binary DER data** as the data type of the new certificate. If the CA sends the certificate as part of an email, you might need to cut and paste the certificate into a separate file.
   c. Accept the default values for the certificate and click **OK**.
   d. Specify a label, such as `Production Certificate for Content Collector`, for the new certificate and click **OK**. The **Personal Certificates** field of the IBM Key Management window shows the label of the new certificate.
   e. Exit the IBM Key Management utility.
4. Stop and restart the service for the embedded Web Application Server (IBM Content Collector Web Application service). You can use the Start menu on a Microsoft Windows system.
   a. To stop the service, click **Start → All Programs → IBM Content Collector → Stop Services → Stop ICC Web Applications**.
   b. To restart the service, click **Start → All Programs → IBM Content Collector → Start Services → Start ICC Web Applications**.

   Important: If you use **Microsoft Exchange**, the IBM Content Collector Web Application service must be started by an account with administrator privileges for Microsoft Exchange.
5. To check if the new certificate works, open your web browser and enter the following URL in the address field:

   ```
   https://<server host name>:11443/AFUWeb/init
   ```

   where

   <server host name>
   Is the host name of the computer running the embedded Web Application Server. This is the same as the computer running the IBM Content Collector server.

   11443
   Is the default port for connections to the embedded Web Application Server

   You should be able to establish an HTTPS connection. If you receive security warnings in your browser, import the public key certificate of your certificate authority into your browser.

Related tasks

"Installing a self-signed certificate for server authentication"

Submitting a certificate request:

If you have a Windows Domain CA and web-based access to Certificate Services is enabled, you can submit a certificate request by following these steps.

1. Access the Certificates Services by specifying the following URL in your web browser:

   `http://<ca_iis_server>/certsrv/certrqxt.asp`

   where *<ca_iis_server>* is the DNS or NetBIOS name of the host server.
2. Paste the request into the form. You can browse for the .arm file that you created with the IBM Key management utility.
3. Under **Certificate Template**, select **Web Server**.
4. Click **Submit**.
5. Download the certificate.

Installing a self-signed certificate for server authentication

IBM Content Collector uses the HTTPS protocol for network communication. This requires a certificate for authentication. If you did not replace the default certificates that were created by the Web Application Server with certificates signed by a trusted certificate authority, email clients must accept a certificate for authentication when they contact the IBM Content Collector server.

The certificate testifies that a server really is the server that a client wants to connect to.

Important: Never use self-signed certificates in a production environment. Self-signed certificates can only be trusted directly while certificates signed by a trusted certificate authority have transitive trust. Transitive trust means that if the clients trust parts of the certificate chain the trust relationship is extended automatically to all parts of the certificate chain. In this case, email clients do not need to accept a certificate for authentication when they contact the IBM Content Collector server.

To install a certificate:

1. Open a web browser on the email client workstation.
2. In the address field of the browser, type the following URL:

   `http://<ICC Web Server>:11080/AFUWeb/init`

   where <ICC Web Server> is the name of the IBM Content Collector web server. A website with an error message is displayed. The message varies depending on the web browser that you use. Follow the instructions for your web browser.
3. For Internet Explorer, the following message is displayed:

   `Your connection is not secured with HTTPS.`

   After a few seconds, you are routed to another website with the following options:
   - **Click here to close this webpage.**
   - **Continue to this website (not recommended).**

   a. Click **Continue to this website (not recommended).** A reddish background is shown in the address bar your browser. Next to it on the right, the

message `Certificate Error` is shown against the same background. The page in the browser windows shows the following message:

`Your connection is secured with HTTPS.`

b. Click **Certificate Error** in the small field next to the address bar.

c. On the **Certification Path** tab of the Certificate window, select the root certificate and click **View Certificate**.

d. Click **Install Certificate** and click **Yes** when you are asked if you want to install the certificate.

e. Press F5 to refresh the page. The certificate error in the address bar disappears and a padlock icon is displayed, indicating the use of HTTPS. The main window of the browser still shows the message:

`Your connection is secured with HTTPS.`

Tip: If you have problems installing the certificate when you use Internet Explorer 7, check the setting of the **Do not save encrypted pages to disk** option. To do so, select **Tools** → **Internet Options**. Go to the security settings on the **Advanced** tab and make sure that the option is not selected.

f. Restart your web browser.

4. To verify that the certificate was properly installed proceed as follows:

a. Open a command prompt and enter the command **mmc** The Microsoft Management Console opens.

b. Select **File** → **Add/Remove Snap-in**.

c. Click certificates. The root certificate for IBM Content Collector was added to the list of certificates.

5. For Mozilla Firefox, you are routed to another website where the following message is displayed:

`This Connection is Untrusted`

a. From the options that are available on this website, select **I Understand the Risks**.

b. Click **Add an Exception** and proceed with accepting the certificate.

The page in the browser window shows the message:

`Your connection is secured with HTTPS.`

Important: Never use self-signed certificates in a production environment.

Replace the self-signed certificates by certificates issued by a trusted authority before running IBM Content Collector in a production environment.

Related tasks

"Replacing certificates for the embedded Web Application Server" on page 101

Establishing a trust relationship between the Web Application Server and IBM FileNet P8

When connecting to IBM FileNet P8, the IBM Content Collector Web Application Server uses the certificate that is sent by the IBM FileNet P8 system for authentication. If IBM FileNet P8 uses a certificate that is issued by a certificate authority (CA) that is not configured as a trusted authority in the IBM Content Collector Web Application Server, the Web Application Server cannot establish a connection because of the missing trust between the two systems.

To establish a trust relationship between the Web Application Server and the IBM FileNet P8 system, the CA certificate that is used to create the IBM FileNet P8 certificate must be imported to the WebSphere Application Server trust store.

On every machine where IBM Content Collector Web Applications run, establish a trust relationship:

1. In a command prompt, go to the <WAShome>\bin directory in the installation directory of the Content Collector server.

   Where <WAShome> is the installation directory of WebSphere Application Server and <profile name> is the name of the Web Application profile, for example, AFUWeb.
2. Type `ikeyman`

   The IBM Key Management utility opens.
3. In the IBM Key Management utility, select **Key Database File** → **Open**.
4. Select **PKCS12** as key database type.
5. In the **Location** field, specify the `<WAShome>\profiles\<profile name>\config\cells\<cell name>\nodes\<node name>\trust.p12` file.

   Replace <WAShome>, <profile name>, <cell name>, and <node name> with the proper values of your WAS installation.
6. Click **OK**.
7. When prompted for a password, enter the password. Click **OK**. The default password is WebAS. Note that the password is case sensitive. In a production environment, change the password as described in the topic about updating default key store passwords using scripting in the WebSphere Application Server (Distributed platforms and Windows), Version 7.0 Information Center.
8. Under the **Signer certificates** list, click **Add** to add the certificate of the CA that issued the FileNet P8 Server certificate as trusted authority to the trust store. For certificates that are not issued by a root CA but by an intermediate CA, you must add the complete certificate chain up to the root CA.
9. Exit the IBM Key Management utility.
10. Stop and restart the service for the embedded Web Application Server (IBM Content Collector Web Application service). You can use the Start menu on a Microsoft Windows system.
    a. To stop the service, click **Start** → **All Programs** → **IBM Content Collector** → **Stop Services** → **Stop ICC Web Applications**.
    b. To restart the service, click **Start** → **All Programs** → **IBM Content Collector** → **Start Services** → **Start ICC Web Applications**.

    Important: If you use **Microsoft Exchange**, the IBM Content Collector Web Application service must be started by an account with administrator privileges for Microsoft Exchange.

Replacing the Lotus Notes mail template in all mailboxes

Replace the standard Lotus Notes mail template in all mailboxes with the mail template that contains the design changes for Content Collector.

For information about how to replace the mail template, refer to the information about how to replace the design of a Notes application in the Lotus Domino and Notes documentation:

- For Lotus Domino Versions 8.0 and 8.5, go to the IBM Lotus Domino and Notes information center at http://publib.boulder.ibm.com/infocenter/domhelp/v8r0/index.jsp
- For Lotus Domino Version 7.0, the IBM Lotus Domino Designer 7 documentation can be obtained from http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss?CTY=US&FNC=SRX&PBL=G210-2369-00.

In addition, perform either of the following steps for Lotus iNotes (formerly Domino Web Access):

- If the forms database is stored on a remote server, restart Lotus Domino Server.
- If the forms database is stored locally, copy the forms to the server directory, then restart Lotus Domino Server.

Related tasks

"Preparing Notes Storage Facility files for archiving" on page 434

Installing Content Collector Outlook Extension

Install IBM Content Collector Outlook Extension on Microsoft Outlook to use Content Collector with Microsoft Outlook.

Install IBM Content Collector Outlook Extension on the client workstations. It should not be installed on the system where IBM Content Collector Server is installed.

You can install IBM Content Collector Outlook Extension in GUI mode, in console mode, or in silent mode. If you are upgrading from an earlier IBM Content Collector release, you do not have to uninstall the earlier Outlook Extension installation. Simply install the latest version.

The IBM Content Collector Outlook Extension uses the directory `IBM\ContentCollector_OutlookExtension` in the local application data directory to store user and machine specific information. This includes the startup trace files (`afuOEaddin<n>.trc`), the user specific configuration file `afuconfig.xml`, and all additional form and definition files. The local application data directory is a standard Windows directory that uses the path `%USERPROFILE%\Local Settings\Application Data`.

Installing Content Collector Outlook Extension in GUI mode

Use the GUI mode if you want to use the installation wizard and customize the installation according to your needs.

If you are upgrading from an earlier IBM Content Collector release, you do not have to uninstall the earlier Outlook Extension installation. Simply run the wizard. You are not prompted for any information. The component is automatically installed in the previous installation location. The default is `C:\Program Files\IBM\ContentCollector_OutlookExtension`. If files need to be overwritten, permit this action.

To run the installation wizard:

1. In the Windows Explorer, change to the directory where you extracted the IBM Content Collector installation package.
2. Run the following command located in the `\OutlookExt` directory of the installation package on the Microsoft client: `install.exe`
3. If prompted, answer the remaining prompts.

- Select which email archiving functions should be included in Microsoft Outlook. You can choose from the following options:

  Typical
  Installs the main email archiving functions, namely archiving, restoring, and searching. This is the default setting.

  Custom
  Installs the email archiving functions that you select.
- Select the functions that are to be added to Microsoft Outlook if you selected **Custom**:

  Mark for archiving
  Lets you select the messages that are to be archived. If you do not select this option, messages are automatically archived at the schedule that is set up by an administrator.

  Restoring
  Allows you to restore archived message content to your mailboxes.

  Search
  Installs a search interface that can be started from Microsoft Outlook. By using this interface, you can search for archived messages.

  Mark for stubbing
  Lets you select the messages that are to be stubbed.

  Offline repository
  Allows for maintaining a local repository, which is useful for people who are often disconnected from their company's network. To work with a local repository, you must have sufficient disk space.

  Specify additional archiving information
  Allows for specifying additional information for messages in a monitored folder. This information can be archived with the message as custom metadata.
- Specify the host name and the port of the server on which the Configuration Web Service is installed.

The menu bar of the Microsoft Outlook window now contains the **IBM Content Collector** menu from which you can use the installed functions. In addition, a tabbed page on which you can configure Content Collector has been added the Options window.

Installing Content Collector Outlook Extension in console mode

Complete these steps to install IBM Content Collector Outlook Extension in console mode.

1. Insert the product DVD into the computer on which you want to install IBM Content Collector Outlook Extension. You can also download and extract the appropriate installation package.
2. Open a Command Prompt window.
3. Type `install.exe -i console` to start the installation.
4. Follow the instructions in the Command Prompt window.

Installing Content Collector Outlook Extension in silent mode

Use the silent installation if you want to install IBM Content Collector Outlook Extension on many client machines. No user interaction is required during this installation mode.

1. Create a response file with the following installation options. The values shown are examples only.

```
LICENSE_ACCEPTED=TRUE
# Accept the license panel.
USER_INSTALL_DIR=C:\\Program Files\\IBM\\ContentCollector_OutlookExtension
# Set the install folder.
USER_INPUT_WEB_SERVER_1=www.server.com
# Set the configuration web service host name.
USER_INPUT_SERVER_PORT_1=11443
# Set the port number.
CHOSEN_INSTALL_FEATURE_LIST=Archiving, Restore, Search, Stubbing, Offline,
                            Metadata
# Set the features that you want to install. Separated by comma.
# There are six features.
```

2. Save the response file to your disk.
3. To start the installation, open a Command Prompt window and enter the following command:

   `install.exe -i SILENT -f <full_path_to_response_file>`

   <full_path_to_response_file> is the full path to the response file, for example, `c:\temp\myresponse.txt`.

   Important: `SILENT` must be specified in uppercase characters if you have a Turkish operating system.

The menu bar of the Microsoft Outlook window now contains the **IBM Content Collector** menu from which you can use the installed functions. In addition, a tabbed page on which you can configure Content Collector has been added the Options window.

Enabling offline repositories to allow access to archived content without network access

Automatic archiving of email might interfere with the needs of users who often work offline or are connected to a slow network. For example, sales people connect to the network of their company at irregular intervals but need access to their email everywhere they go.

With IBM Content Collector, you can work with offline repositories that are synchronized at defined intervals with the repository on the server. Synchronization only occurs when you are connected to the network.

The synchronization depends on the stubbing life cycle that is defined by the administrator in the Configuration Manager and on the update settings for the offline repository.

Enabling offline repositories in Lotus Domino

If you enable an offline repository in Lotus Domino, you can work with the documents that have been stubbed by IBM Content Collector even if you are offline.

Complete these steps to enable an offline repository:

1. As an administrator, take these steps:
   a. Replace the design of your mail database by the mail template that is enabled for Content Collector:
      1) Open your mail database in Lotus Notes Designer and select **Install Content Collector Offline Repository** under **Shared Code → Agents**.

2) Right-click **Design properties**.
3) Select the **Design** tab and deselect everything following **Hide design elements from**.
4) Close the window.

b. From the inbox of your mail database, select **Actions** → **Install IBM Content Collector Offline Repository** to create a memo. The memo contains the *Install IBM Content Collector Offline Repository* hotspot.
c. Right-click the **Install IBM Content Collector Offline Repository** hotspot and then click **Edit**. At the bottom of the window that opens, add a blank line in front of the *On Error Goto processError* line to suppress any `No signature` warnings.
d. Specify the recipients of the memo and add any information that you want to communicate. Then send the memo.

2. As a user who wants to work with the offline repository, take these steps:
   a. Click the hotspot in the memo that you received from your administrator.

      You are asked if you want to enable the offline repository and if the default name of the database should be used as the name for the local repository. For example, if your mail file is called `user1.nsf`, the default name for the offline repository is `user1_icclocalrep.nsf`. If you select **No**, you can specify another database name for the local repository.
   b. Click **File** → **Preferences** → **User Preferences**. On the Basics page, select the **Enable scheduled local agents** check box.
   c. Start the Lotus Notes client again.
   d. Open the local replica of your database.
   e. To enable or disable the offline repository, to change the location of the offline repository, or to recreate the offline repository, click **Actions** → **IBM Content Collector Offline Repository**.

Enabling offline repositories for Microsoft Exchange

If you enable and initialize an offline repository in Microsoft Exchange, you can work with the messages that have been stubbed by IBM Content Collector even if you are offline.

Complete these steps to enable and initialize an offline repository:

1. Install IBM Content Collector Outlook Extension. To enable an offline repository, you must select specific options during the installation:
   - Select **Custom** as installation type.
   - From the list of functions displayed, select the **Offline Repository** check box in addition to the functions that you want to install.
2. Enable the offline repository in Microsoft Outlook as follows:
   a. Start Microsoft Outlook.
   b. In the Microsoft Outlook window, click **Tools** → **Options**.
   c. In the Options window, click the **IBM Content Collector** tab.
   d. Under **Offline Repository**, select the **Enable offline repository** check box and then click **OK**.
3. Configure the offline repository:
   a. In the Microsoft Outlook window, click **Tools** → **Options**.
   b. In the Options window, click the **IBM Content Collector** tab.
   c. Under **Offline Repository**, click **Offline Repository Settings**.

d. In the Offline Repository window, you can specify the following information:
   - The location of the offline repository on your local system.
   - The maximum size of the offline repository to prevent the repository from getting too large to open.

     Tip: Specify a size less than 1 GB.
   - The maximum size and age of the messages to be stored in the offline repository. Larger or older messages are not copied to the offline repository.
   - Whether the largest messages or the oldest messages are to be deleted when the maximum size of the offline repository is reached.
   - At which interval the repository is to be updated. In this way, you can keep your repository synchronized with the repository on the server.

     Choose an interval that ensures that the offline repository is updated before the documents are stubbed. The optimal interval depends on the amount of data that is archived.

4. Initialize the offline repository: In the Microsoft Outlook window, click **IBM Content Collector → Offline Repository → Initialize**.

   If an offline repository does not exist yet, a repository is created. If an offline repository exists, you get a warning that any messages in the existing repository are removed.

The offline repository is stored in a personal folder (PST file). You can see this folder in the folders pane on the left of the Microsoft Outlook window. However, do not open or change the PST file or the messages that it contains.

Note:

- The initialization process can take a considerable amount of time. If you want to interrupt this process, click **IBM Content Collector → Offline Repository → Pause**. You can restart the initialization at any time by clicking **IBM Content Collector → Offline Repository → Resume**.
- When messages are archived in the repository on the server, complete copies are stored in the offline repository. After the copy is available in the offline repository, the message on the server is stubbed according to the life cycle that is defined for archived messages in the Configuration Manager. When you open a stubbed message, the request is redirected internally, so that the offline copy of the message is displayed.
- When you open a stubbed message, the content of the message copy in the offline repository is displayed to you. If you changed the content of the copy, but want to view the original content again, restore the message in the mailbox. When you open a restored message, the content of the message in the mailbox is displayed.

Installing and configuring Content Collector Outlook Web App (formerly Outlook Web Access) support

You can extend the Outlook Web App (OWA, formerly Outlook Web Access) capabilities to include IBM Content Collector functions which enable you to mark documents for archiving or stubbing, to search, view, or restore archived content, or to specify additional archiving information.

The OWA interface resembles that of Microsoft Outlook after IBM Content Collector Outlook Extension was installed. You select a document or document stub and then click one of the additional buttons provided by IBM Content Collector.

Note: Compared with the IBM Content Collector support for Microsoft Outlook, the functions differ in one respect: In Outlook, an action can be performed on more than one document at a time. For example, you can select a number of documents, then click and all of the documents will be archived. In the IBM Content Collector support for OWA, you can only perform an action on a single document at a time.

The IBM Content Collector OWA Support comprises the OWA Service and the OWA Extension. Before you can use the Content Collector OWA Support, you must configure both components using the Configuration Manager.

Note: The following limitations exist if you use IBM Content Collector OWA support:

- In Outlook Web Access 2003, hyperlinks to archived content do not work for RTF messages, if the stubbing option "Delete attachments" or "Delete attachments and cut body" is used. To view archived content, use the View button.
- The IBM Content Collector functions are not available in OWA for the following types of folder items:
  - Calendar items
  - Contact items
  - Task items

The OWA Service and the OWA Extension can be installed on the same machine or different machines. Typically, one or more of both the OWA Service and the OWA Extension might need to be installed based on the Exchange server topology and user requirements. Irrespective of the chosen installation setup however, the installation must be carried out as follows:

- The OWA Service must be installed where Microsoft Information Internet Services (IIS) is installed. The service is an interface that accepts requests from client browsers with a modified OWA client interface.
- The OWA Extension must be installed on all Exchange servers that provide access to end users and makes IBM Content Collector functions available on the OWA client interface.

  In an Exchange 2003 environment, this server must be a front-end Exchange server, in an Exchange 2007 and 2010 environment, this must be an Exchange server that has the Client Access Server role.

To configure and install the IBM Content Collector OWA support:

1. Configure the OWA Service
2. Configure the OWA Extension
3. Install the prerequisites for the OWA support.
4. Install IBM Content Collector OWA Support
5. Apply the OWA Extension configuration:
   a. Log on to the server on which the OWA Extension is installed.
   b. Select **All Programs → IBM Content Collector → Apply Configuration**

The configuration data that you entered using the Configuration Manager is accessed from the IBM Content Collector web server and applied to the server on which the OWA Extension is installed.

Important: If you change the configuration, you must make the changes using the Configuration Manager and then run **Apply Configuration** again to activate your changes.

The OWA Service configuration is automatically refreshed while the service is running.

Note: How often the service configuration is refreshed is defined by the parameter **timerInterval** in the `afu.bootstrap.ini` file. The default value of this parameter is set to 30 seconds. Change this interval if you find this time span too short or too long. The file is located in the Microsoft Information Internet Services (IIS) working directory `C:\Inetpub\wwwroot\afuowa`.

Prerequisites for Outlook Web App (formerly Outlook Web Access) integration

Before you install IBM Content Collector Outlook Web App (OWA, formerly Outlook Web Access), check the prerequisites for OWA integration.

Prerequisites for OWA Extension

The IBM Content Collector support for OWA adds IBM Content Collector functionality to the OWA interface, which can be accessed using Microsoft Internet Explorer.

Before you install IBM Content Collector OWA, check the prerequisites for Exchange Server at

http://www.ibm.com/support/docview.wss?rs=484&uid=swg27019702#OOS

Prerequisites for the OWA Service

Apart from the OWA Extension that is installed on the OWA Exchange server, the support package includes a component that is called the OWA Service. Check the OWA Service prerequisites at:

http://www.ibm.com/support/docview.wss?rs=484&uid=swg27019702#IIS

Prerequisites for OWA clients

Client workstations using OWA must also fulfill certain prerequisites. Check these prerequisites at

http://www.ibm.com/support/docview.wss?rs=484&uid=swg27019702#OCP

Configuring the Outlook Web App (formerly Outlook Web Access) Service

Perform the following steps to configure the Outlook Web App (OWA, formerly Outlook Web Access) Service. The service is the interface that accepts requests from client browsers using a client interface modified by OWA Extension.

Configuring basic parameters and selecting the authentication method for the Outlook Web App (formerly Outlook Web Access) Service:

Perform the steps in this section to configure basic parameters and select the authentication method for the Outlook Web App (OWA, formerly Outlook Web Access) Service.

The Web Service Extension ASP.NET v2.0.50727 must be set to **Allowed** in the Internet Information Service (IIS) Manager.

To configure basic parameters and select the authentication method for the OWA Service:

1. In the Configuration Manager, click **General Settings** to change to the General Settings view.
2. In the **General Settings** pane on the upper left, select **OWA Service**.
3. Click to create a configuration for the OWA Service. A tabbed notebook appears in the right pane.
4. On the **General** tab, type a name for your configuration in the **Name** field.
5. You can add a description in the **Description** field.
6. In the **Host name** field, type the fully qualified name of the Internet Information Service server on which the OWA Service is installed. Example:

   ```
   iccowa.domain.company.com
   ```

   This host name is used as a unique key with which the OWA Service finds the configuration data.
7. In the **User ID** field, specify the ID of a registered Exchange user. If you grant the user ID *Exchange View Only Administrator* rights, the user can only view, search, and restore documents. The user cannot specify additional archiving information, archive or stub documents.

   To start the web service on the OWA Service to enable archiving, stubbing, and specifying archiving information however, the user must have administrative privileges.

   - For Exchange 2003, the user ID must have *Exchange Administrator* or above privileges for all functions.
   - For Exchange 2007, the user ID must have the following privileges for all functions:
     a. The ID must be a registered *Exchange Organization Administrator*.
     b. The ID must have the Active Directory privilege *ms-Exch-EPI-Impersonation* on the Exchange server with the Client Access Server role.
     c. The ID must have the Active Directory privilege *ms-Exch-EPI-May-Impersonate* on all identified Exchange 2007 databases on servers with the mailbox role.

   For example: Using the Exchange Management Shell with cmdlets grant the following permissions:

   a. Grant the user the right to submit an impersonation call through the Client Access server. Adjust the settings accordingly in the following sample cmdlet to give this right to the domain.company.com/Users/iccadm user account if the impersonation call is routed through an Exchange 2007 Client Access Server named icccas.domain.company.com.

      Note: When you use the sample cmdlet, all of the information must be entered in one line or it will fail.

      ```
      Add-ADPermission -Identity (Get-ExchangeServer
          -Identity icccas.domain.company.com |
           select-object).DistinguishedName
      ```

```
-User (Get-User -Identity domain.company.com/Users/iccadm |
 select-object).identity
-extendedRight ms-Exch-EPI-Impersonation
```

b. Grant the same user the right to access any account in all identified Exchange mailbox databases. Adjust the settings accordingly in the following sample cmdlet to grant the domain.company.com/Users/iccadm user account access rights to any mailbox in mailbox database MDB_01 on server ExchSrv01:

Note: When you use the sample cmdlet, all of the information must be entered in one line or it will fail.

```
Add-ADPermission -Identity (Get-MailboxDatabase
    -Identity "ExchSrv01\SG_01\MDB_01" |
     select-object).DistinguishedName
    -User (Get-User -Identity domain.company.com/Users/iccadm |
     select-object).identity
    -extendedRight ms-Exch-EPI-May-Impersonate
```

- For Exchange 2010, the user ID must have *Exchange Impersonation* privileges for all functions.

  Use the New-ManagementRoleAssignment Exchange Management Shell cmdlets to configure *Exchange Impersonation* for specific users or groups of users in an organization:

```
-Name:impersonationAssignmentForICC
-Role:ApplicationImpersonation
-User:iccadm
```

- In a mixed environment, the user ID must have the corresponding permissions mentioned above.

8. Type the password of this user in the **Password** field.
9. Click **Validate** to check the user ID and password.
10. Select the method to authenticate the OWA Service to the Exchange server by clicking the appropriate radio button.

    A dual authentication is required for the OWA Service:

    a. The authentication of IBM Content Collector OWA to the Exchange server:

    To authenticate to the Exchange server, the following methods can be selected:

    Basic authentication
    The password is not encrypted when sent to the Microsoft Internet Information Services Server for authentication.

    This method is supported for the `afuowa` directory, which allows users to access IBM Content Collector OWA functions, and the virtual directories `OWA` (2007/2010) or `Exchange` (2003).

    Integrated Windows authentication (also known as NTLM)
    The password is encrypted when sent to the Microsoft Internet Information Services Server for authentication.

    This method is supported for the `afuowa` directory and the virtual directories `OWA` (2007/2010) or `Exchange` (2003).

    Forms-based authentication (FBA)
    Secure Sockets Layer (SSL) must be enabled on the Microsoft Internet Information Services Server. The password is encrypted when sent to the Microsoft Internet Information Services Server for authentication.

This method is supported for the virtual directories OWA (2007/2010) or Exchange (2003). For IIS 7.0, this method is not supported at all.

The authentication method Basic or Integrated Windows authentication (NTLM) is configured in the Internet Information Services (IIS) Manager or on the Exchange server. The forms-based authentication (FBA) is configured on the Exchange server. On Exchange 2003, FBA is configured on the Exchange 2003 Virtual Server in the Exchange System Manager (ESM) and on Exchange 2007 or Exchange 2010, it is configured in the Exchange Management Console. You need to check and update the authentication method on the virtual directory OWA (2007/2010) or Exchange (2003). Also check and update the authentication method on the afuowa directory.

For details about the authentication methods, refer to the Microsoft Exchange documentation.

If Microsoft Exchange uses an authentication method not listed here, you need to change the authentication method to one of the supported types. By default, basic authentication is enabled after Microsoft Exchange 2003 is installed and forms-based authentication (FBA) is enabled after Microsoft Exchange 2007 and Microsoft Exchange 2010 is installed.

b. The user authentication to the IBM Content Collector OWA Service.

Note: This must be done on the Microsoft Internet Information Services (IIS) Server and not by using the Configuration Manager. The virtual directory is AFUOWA for the OWA Service by default.

For the user to authenticate to IBM Content Collector OWA, the following methods are supported:

Basic authentication

This is the default method. The user is requested to enter authorization credentials (mailbox owner and a password) the first time an IBM Content Collector button is clicked after the OWA support was installed.

The password is not encrypted when sent to the Microsoft Internet Information Services Server for authentication.

Integrated Windows authentication (also known as NTLM)

The user must enter the URL to where OWA is running on the Exchange server.

For Exchange 2007 and Exchange 2010, the user must use an explicit logon URL to log on to mailboxes using OWA:

http<s>://*server name*/owa/*EXCHANGE user ID*/

where the *EXCHANGE user ID* is the SMTP address of the user whose account is associated with the mailbox. For example, if john.doe@company.com is the registered SMTP address of the user John Doe and the server name is myserver, the URI to the mailbox is:

https://myserver/owa/john.doe@company.com/

A mailbox can have more than one SMTP address. You can use any of them to open the mailbox.

The password is encrypted when sent to the Microsoft Internet Information Services Server for authentication.

You can now configure the Active Directory that is used by your Exchange system.

Changing the identity of the application pool for OWA:

In Microsoft Internet Information Services (IIS) 7.0 or later, the identity of the application pool where the OWA service resides must be set to NetworkService.

To change this setting in the IIS Manager:

1. In the navigation under **Sites**, select the virtual directory `afuowa`.
2. In the Advanced Settings window, check which application pool is used for the OWA service.
3. In the navigation, select **Application Pools** and check whether the identity of the application pool that is used for the OWA services is NetworkService.
4. To change the identity of the application pool if required, click **Advanced Settings** and set the identity to NetworkService.

Related tasks

"Installing Content Collector Outlook Web App (formerly Outlook Web Access) support" on page 122

Configuring the Active Directory:

Perform the steps in this section to specify the details of the Active Directory that is used by your Exchange system. The Active Directory authenticates the user credentials behind each request that reaches the Outlook Web App (OWA, formerly Outlook Web Access) Service.

On the **Active Directory** tab:

1. In the **Host name** field, type the name of the Active Directory server. The Active Directory server MUST be a Global Catalog server. You can specify the fully qualified domain name (FQDN) or an IP address, for example `adserver.company.com`.
2. In the **LDAP port** field, type the number of the port used for communication with the LDAP service on the Active Directory server. The default port is 389.
3. In the **Global catalog port** field, type the number of the port used for communication with the Global Catalog service on the Active Directory server. Each authentication request first goes to the global catalog server, from where it is routed to the relevant Active Directory server. The default global catalog port is 3268.
4. In the **User ID** field, type the ID of a user who is entitled to perform lookups on the Active Directory server. Specify the user account that is used to access the Active Directory server. If Active Directory cannot be accessed using the SMTP address, use the format `domain/userID` or the user distinguished name. The Active Directory server user ID must have READ access on the following Active Directory nodes including all child nodes.

   ```
   LDAP://<AD server host>:<Port>/CN=Partitions,CN=Configuration,DC=<Forest-Root-Domain>
   LDAP://<AD server host>:<Port>/CN=Microsoft Exchange,CN=Services,DC=<Forest-Root-Domain>
   LDAP://<Domain controller host name>:<Port>/<User Domain>
   ```

   For example:

   ```
   AD Server name: adsrv.domain.com
   Domain Name 1: domain.com DC: dc1.domain.com
   Domain Name 2: child.domain.com DC: dc2.child.domain.com

   You must grant all the nodes and their child nodes READ access:
   LDAP://adsrv.domain.com:389/CN=Partitions,CN=Configuration,DC=domain,DC=com
   ```

```
LDAP://adsrv.domain.com:389/CN=Microsoft Exchange,CN=Services,DC=domain,DC=com
LDAP://dc1.domain.com:389/DC=domain,DC=com
LDAP://dc2.domain.com:389/DC=child1,DC=domain,DC=com
```

These READ access privileges are the same for both Exchange 2003 and 2007 and must be applied on all domains in the forest.

5. Type the password of this user in the **Password** field.
6. Click **Validate** to check the User ID and password.

You can now configure the URL mappings

Configuring URL mappings to Exchange servers:

Perform the steps in this section to configure the URL mappings to Exchange servers.

There are several situations in which URL mappings are required, for example, for proxy servers which are usually employed when users access an organization's mail system through the Internet. In this case, mappings from proxy servers to the internal Exchange servers are needed.

Each mapping contains two parts:

- A *Replace* part containing the leading substring of the URL which will be replaced with the string in the With part.

  The string to be replaced must include the first character of the original URL which is to be mapped to another URL and must end with the last character before the forward slash character (/) .
- A *With* part containing the string that will replace the Replace string.

  The string in the With part cannot end with a forward slash character (/).

Both strings are case-insensitive. Wildcard characters, such as the question mark (?) or the asterisk (*), cannot be used.

URL mappings are required to:

- Map a proxy server (such as an Internet Security and Acceleration (ISA) Server) to an internal Exchange server
- Map a virtual directory to another virtual directory on website
- Map a web server to another web server internally

Mapping a proxy server to an internal Exchange server

Outlook Web App (OWA, formerly Outlook Web Access) users log on to a website on the proxy server. This requires a mapping of the Internet website address to the address of the intranet Exchange server on which the published website is located.

For example:

```
Replace: http://www.iccowa.com       With: http://iccserver
```

If the internet website uses a different port to the published intranet website, a mapping is required to enable opening attachments when viewing archived messages. For example:

```
Replace: http://www.iccowa.com:80    With: http://iccserver:81
```

Mapping a virtual directory to another virtual directory on a website

If OWA users interact with their mailboxes using an Exchange virtual directory that is not located on the default website, a mapping is required to map the directories on the non-default website to the directories on the default website.

For example: The default Exchange virtual directory is located on a default website using HTTP with host name exchsrv, default port 80, and is associated with the domain mail.company.com. You added a non-default website which is bound to the same host name (exchsrv), but uses a different port (85). On this website you defined a virtual Exchange 2003 mailbox directory named mail2003 and an Exchange 2003 public folder directory named pf2003. The virtual directory mail2003 is associated with the same domain as the default Exchange virtual directory (mail.company.com). To open mailboxes or public folders in OWA using the following URLs:

```
http://exchsrv:85/mail2003
http://exchsrv:85/pf2003
```

you must add these mappings:

```
Replace: http://exchsrv:85/mail2003    With: http://exchsrv/Exchange
Replace: http://exchsrv:85/pf2003          With: http://exchsrv/public
```

Mapping a web server to another web server internally

If you add a non-default website which is bound to another host name, for example, to exchsrv2 and not, for example, to exchsrv, the web server exchsrv2 requires a mapping. On this non-default website you defined a virtual Exchange 2003 mailbox directory named mail2003 and an Exchange 2003 public folder directory named pf2003. The virtual directory mail2003 in this case is associated with a different domain than the default Exchange virtual directory mail2.company.com. To open mailboxes or public folder in OWA using the following URLs

```
http://exchsrv2/mail2003
http://exchsrv2/pf2003
```

you must add the following mappings:

```
Replace: http://exchsrv2/mail2003     With: http://exchsrv1/exch2
Replace: http://exchsrv2/pf2003       With: http://exchsrv1/public
```

On the **Mappings** tab:

1. Click the **Add** button.
2. In the Add URL Mapping window, in the **Replace this text** field, type the leading substring of the URL which will be replaced with the string in the **With this text** field. Example:

   ```
   http://www.iccowa.com
   ```

   Note: The string to be replaced must include the first character of the original URL which is to be mapped to another URL and must end with the last character before the forward slash character (/).
3. In the **With this text** field, type the string that will replace the Replace string. Example: `http://iccserver`.

   Note: The string in the With part cannot end with a forward slash character (/).
4. Click **OK** to complete the mapping. You can see your URL mappings.

You can now configure the trace options.

Configuring tracing:

Perform the steps in this section to enable additional tracing and to specify the location of the trace file.

Most ordinary system events in your Outlook Web App (OWA, formerly Outlook Web Access) environment are documented by a brief entry in the OWA trace file. This means that a trace file is created no matter if tracing is switched on or not. However, there might be occasions where you need more information to analyze and solve a problem. In these cases, you can configure the system to give you additional information.

You can control the amount of disk space that trace files can have by setting the maximum number of trace files and the maximum size that each file can have. The oldest trace file is overwritten when the newest file has reached the size limit. For example, if you permit a maximum of four trace files with a size limit of ten MB each, no more than 40 MB of disk space can be consumed by trace files. Note that this total limit is reached sooner when tracing is switched on because the files are filled with more information in the same time.

On the **Trace** tab:

1. Select **Enable tracing** on the Trace page to switch tracing on. Clear the check box to switch it off.
2. In the **Trace file name** field, type the full path and the name of the trace file. You can also leave this field empty, in which case a default path and name is used. This is helpful because OWA support is installed on a separate machine and the installation path might be unknown. The default trace file name and path are

   installation path\logs\afuowa.trc

   where *installation path* is the installation directory that you selected for the IBM Content Collector Outlook Web App package on the machine running the OWA Service.

   If you enter a trace file name with no path information, for example `owa.trc`, this file will be created in the *installation path*\logs directory. Not specifying the trace file path ensures that the trace file is written to the \logs subdirectory. Trace files cannot be written to the root directory. If you specify a file path such as `c:\abc.trc`, the trace file will be created at `c:\abc.trc\afuowa.trc`.

   To enable the OWA Service to write trace files, grant the account that is configured to run the OWA Service application, such as the Network Service account (IIS 6.0) or the ASPNET account (IIS 5.0 and 5.1), *Full Control* of the trace-file folder. By default, this access level is given to the Network Service account or the ASPNET account for the *installation path*\logs folder during the installation of the OWA Service. For other folders, you have to manually grant this access level.
3. In the **Number of trace files** field, type the maximum number of trace files to be created. When the number of trace files that can be created is exceeded, the oldest files are overwritten.
4. In the **Size limit in MB** field, specify the maximum size of each trace file in MB.
5. The process that restores archived content to a user's mailbox overwrites all changes that the user has made to the original document after archiving. To

warn the user of this potential data loss, select **Enable warning**. Users see the warning in a message window when they restore documents. They can disable it by clicking **Do not show this message in the future** in the message window. This configuration setting is stored in a cookie on the clients' machines.

Configure the Outlook Web App (formerly Outlook Web Access) Extension

Specify the details of the Exchange server to which you want to add IBM Content Collector Outlook Web App (OWA, formerly Outlook Web Access) functions. You also select the Content Collector functions that you want to offer to client users.

OWA must be fully configured and enabled on the Exchange server that you specify. This means that a Microsoft Information Internet Services (IIS) Server must exist and be configured for OWA.

The OWA Extension is installed on:

- The front-end server for Exchange 2003. This is recommended. However, if a user does not access OWA through the front-end server but instead connects to the back-end server directly, or does not have a front-end server in the installation environment, then OWA Extension can also be installed on the back-end servers.
- The Client Access Server for Exchange 2007

To configure the OWA Extension:

1. In the Configuration Manager, click **General Settings** to change to the General Settings view.
2. In the **General Settings** pane on the upper left, select **OWA Extension**.
3. Click to a add an OWA Extension configuration.
4. Type a name for your configuration in the **Name** field.
5. You can add a description in the **Description** field.
6. Under **OWA Extension Server**, in the **Host name** field, type the fully qualified domain name of the Exchange server that provides the IBM Content Collector buttons for the OWA users. The host name is used as a unique key with which the OWA Extension can access configuration data. For example:

   ```
   owa.2007.icccas.domain.company.com
   ```
7. Under **OWA Service**, in the **Host name** field, type the fully qualified website name of the published OWA Service website.
8. In the **Port** field, type the number of the port used for communication between the Exchange server providing the IBM Content Collector buttons and the OWA Service website. The default is 443.
9. Select the protocol of the OWA Service Web site. HTTPS is the default.
10. Select the functions that you want to add to the OWA browser interface:

    Mark for Archiving

    Adds the function with which documents can be marked to be archived to the OWA browser interface. If you do not select this option, documents cannot be archived by the client users themselves, but only by means of automatic archiving.

    Note: This function is not available for messages in a monitored folder. Messages in a monitored folder are archived by means of automatic archiving.

Restore
Adds the restoring function to the OWA browser interface, which allows users to restore archived message content to their mailboxes.

View Adds the viewing function to the OWA browser interface. This function allows users to view archived content after clicking links in stub documents or in a search result list.

Mark for Stubbing
Adds the function with which documents can be marked to be stubbed to the OWA browser interface. By using this function, users can flag messages so that these are stubbed the next time that the stubbing collector runs. This way, they can free up space in their mailboxes within a short time frame and need not wait until the collector schedule triggers the processing of their mailboxes.

Search
Installs a search interface, which can be started from the OWA browser interface. Using this interface, client users can search for archived messages.

Specify Additional Archiving Information
Adds the function with which additional archiving information can be added to documents to the OWA browser interface. This functions allows users to specify additional archiving information for messages in specific folders.

Note: In OWA 2010, you must disable conversation view in the monitored folders if you want to be able to specify additional archiving information for messages.

Installing Content Collector Outlook Web App (formerly Outlook Web Access) support

Follow the steps in the installation wizard to add IBM Content Collector Outlook Web App (OWA, formerly Outlook Web Access) support.

Running the wizard adds IBM Content Collector email archiving functions and the controls to use these functions to the existing OWA capabilities. These functions enable you to mark documents for archiving or stubbing, to search, view, or restore archived content, or to add additional archiving information using the Microsoft Internet Explorer as your mail client.

IBM Content Collector OWA support comprises the OWA Service and OWA Extension. These components can be installed on the same machine or on different machines.

If you are upgrading from an earlier IBM Content Collector release, you do not have to uninstall the earlier OWA support installation. Simply run the wizard. You are not prompted for any information. The component is automatically installed in the previous installation location. If files need to be overwritten, permit this action.

To run the installation wizard:

1. In the Windows Explorer, change to the directory where you extracted the IBM Content Collector installation package.
2. Run the following command located in the `\OWA` directory of the installation package on the Exchange server: `install.exe`

3. If prompted, answer the remaining prompts. You can select to install both the OWA Extension and the OWA Service on the same machine or on different machines.

   If you install on different machines, select to install the OWA Service component on the server on which Microsoft Information Internet Services is installed, and then to install the OWA Extension component on all Exchange servers.

   Important: In an Exchange 2003 environment, this server must be a front-end Exchange server; in an Exchange 2007 environment, it must be an Exchange server that has the Client Access Server role.
4. If you use Microsoft Information Internet Services (IIS) 7.0 (this is the default version on Windows Server 2008), perform these steps after the installation.
   - Disable **Forms Authentication**.
   - In the IIS Manager, check how the identity of the application pool where the OWA Service resides in is set. For IIS 7.0 or later, it must be set to NetworkService.
5. If you work with Outlook Web Access 2007, disable the option **Use the blind and low vision experience** under **Options → General Settings → Accessibility**.

Related tasks

"Changing the identity of the application pool for OWA" on page 117

Starting, stopping, and updating your installed information center

Your installed information center runs as a service or daemon for viewing on a local network. You can also start and stop it manually in stand-alone mode so that you can add updates to it. Updates are available for download directly to a connected machine or to a CD or DVD for a machine that is not connected.

Starting or stopping an information center in stand-alone mode

You must start the information center in stand-alone mode to update the installed documentation.

You can start the information center in a local, stand-alone mode even when it is also running as a service or daemon. If you update your information center, however, you must stop and restart the information center service or daemon after the update is complete for the updated documentation to display.

Whenever you start the information center process, you should always stop it after you are done. Otherwise, the process continues to run in the background and consume system resources.

There are two ways to start or stop the stand-alone information center:

- Use the start menu to start or stop the information center in stand-alone mode.
  - To start the information center service and open the information center:
    1. To start the service, select **Start → All Programs → IBM Content Collector → Start Services → Start Information Center**.
    2. To open the information center, select **Start → All Programs → IBM Content Collector → Open Information Center**.
  - To stop the information center service:
    1. Select **Start → All Programs → IBM Content Collector → Stop Services → Stop Information Center**.

- Use the commands in the *ICC-Installation*\IEHS directory, for example C:\Program Files\IBM\ContentCollector\IEHS.
  - Enter the command: help_start to start the information center.
  - Enter the command: help_end to stop the information center.

The information center starts and starts the system default web browser to view the documentation on a randomly generated port.

Updating the information center that is installed on your computer or intranet server

The IBM Content Collector Information Center on ibm.com contains the most up-to-date information. Your locally installed information center contains the English information from the product electronic availability date.

Your locally installed information center contains the documentation in English only. If you prefer to read the documentation in your language, update your local information center to add languages other than English.

You should regularly check if updates to the information center are available to make sure that you access the most up-to-date information in your local information center for any language.

Determining whether documentation updates are available:

The IBM Content Collector Information Center hosted on ibm.com has the most up-to-date information in all available languages. Your locally installed information might not have the most current information. However, you can update your information center by downloading files from ibm.com. Before you install any files, determine whether updates are available.

See http://publib.boulder.ibm.com/infocenter/email/v2r2m0/topic/com.ibm.emailtoc.doc/infocurrency.htm for current update information. Use the **Last updated** date at the bottom of the Welcome page in your locally installed information center for comparison if necessary.

Creating a CD or DVD of documentation updates to your information center:

If the computer hosting your IBM Content Collector Information Center is not connected to the Internet, you can install the updates from a CD or DVD. From a separate computer that does have an Internet connection, download the files to create the CD or DVD.

To create a CD or DVD of updates:

1. From a web browser, go to ftp://ftp.software.ibm.com/software/cmgmt/email/v2r1m1/.

   ```
   <site name="eMail Update Site" url="ftp://ftp.software.ibm.com/software/cmgmt/email/v2r1m1/" />
   ```

2. Save the file site.xml, the subdirectory features/, and the subdirectory plugins/ to a temporary directory on your computer.
3. Copy the site.xml, the subdirectory features/, and the subdirectory plugins/ onto a CD.
4. From the computer where your information center is installed, make a backup of the file org.eclipse.help.webapp_3.1.0/bookmarks.xml.
5. Open org.eclipse.help.webapp_3.1.0/bookmarks.xml for editing.
6. Locate the following line:

```
<site name="eMail Update Site" url="ftp://ftp.software.ibm.com/software/cmgmt/email/v2r1m1/" />
```

7. Modify the `url` attribute in the line above to refer to the copy of the `site.xml` file that exists on the CD of updates in your local CD drive. For example, if you are installing the updates from a CD drive that is the `D` drive on your system and if the `site.xml` file is in a directory called `icupdates`, you specify:

   ```
   <site name="eMail Update Site"
   url="file:///D:/icupdates/site.xml" />
   ```

8. Save your changes and close the editor.

Install the updates into your information center.

Installing updates into your locally installed information center:

If your information center is installed on a computer that is connected to the Internet, you can update it directly. If your information center is on a computer that is not connected to the Internet, you can download the updates to a CD or DVD and then update the information center.

To update your locally installed information center with updates from the Internet or a CD or DVD:

1. Start the information center in stand-alone mode.
2. Click the **Update** button to display a list of installed documentation.
3. Click **Find Updates** to see a list of available updates to installed documentation.
4. Select documentation updates to download and click **Install Updates**.
5. After the download and installation process completes, click **Finish**.
6. Stop the information center in stand-alone mode.
7. Start the information center again.

The installed information center displays the new and updated information.

Related tasks

"Determining whether documentation updates are available" on page 124

Removing Content Collector

You might want to remove IBM Content Collector installation.

To uninstall Content Collector, complete these steps:

1. Stop all components and processes of your current installation of Content Collector:
   - Click **Start** → **All Programs** → **IBM Content Collector** → **Stop Services** → **Stop *process***. *process* stands for the component or process that you can stop.
   - Use the Services window of Microsoft Windows to stop the IBM Content Collector Metadata Form Database service and the IBM Content Collector Web Application service.
2. Verify that all Content Collector processes ended by using the Task Manager. Also ensure that the executable program `msiexec.exe` is not running.
3. Uninstall Content Collector by clicking **All Programs** → **IBM Content Collector** → **Uninstall IBM Content Collector**.
4. Delete all remaining files, including the log files and the temporary files, and all subdirectories from the installation directory of Content Collector. You do not need to delete the registry keys manually.
5. Uncatalog the Content Collector configuration database on the server where you installed Content Collector.

Related tasks

"Installing Content Collector according to your needs" on page 58

© Copyright IBM Corp. 2008, 2010

Part 3. Migrating

© Copyright IBM Corp. 2008, 2010

Migrating to Content Collector

To access legacy repositories created by IBM CommonStore, IBM FileNet Email Manager, or IBM FileNet Records Crawler in IBM Content Collector, refer to the topics here.

Upgrading to version 2.2 of Content Collector

To upgrade from an earlier Content Collector release to IBM Content Collector Version 2.2, you must stop components and processes, back up your data, record your user account settings, and consider several options and limitations.

Complete these steps irrespective of the release or fix pack version of Content Collector you are upgrading from:

1. Stop all components and processes of your current Content Collector installation: Click **Start** → **All Programs** → **IBM Content Collector** → **Stop Services** → **Stop *process***. *process* stands for the component or process that you can stop.
2. Verify that all Content Collector processes ended by using the Windows Task Manager.
3. During the upgrade, the configuration data of your current installation is upgraded to Version 2.2. Back up the configuration database of your current installation so that you can return to your current installation if, for example, an error occurs during the upgrade. You can back up the database by exporting it as an XML file or by using database backup methods.

   Restriction: If you decide not to upgrade the existing configuration database but to use another existing one, proceed as follows:

   a. Exit the Configuration Manager.
   b. Delete the file `CTMSConfigStore.xml` from *`ICCInstallDir`*`/ctms/`, where *ICCInstallDir* is the directory of your current Content Collector installation.
   c. Restart the Configuration Manager and select the new database.
4. If you added custom attributes to your configuration for accessing archived data, back up the `custom_label.properties` files that you created.
5. Make a note of all of the user accounts that you used for all Content Collector services. For a list of all services and required user accounts, see "Content Collector services" on page 178.
6. Install Version 2.2 of IBM Content Collector. During the installation, keep the following in mind:
   - You cannot change the installation directory.
   - The installed components are preselected. You can add or remove components as required.
   - After the installation wizard completes, do not run the initial configuration.
   - If you use Lotus Domino as source system, enable the Domino template to include the Content Collector Version 2.2 functions.
   - Upgrading does not affect the Configuration Manager security that was enabled in an earlier version of Content Collector. After upgrading, only valid users will be allowed to run the Configuration Manager.

© Copyright IBM Corp. 2008, 2010

- You are asked to upgrade the configuration database when you start the Configuration Manager. During the database upgrade, task routes are also upgraded except for the following FileNet P8 task routes:
  - Email archiving task routes with records declaration
  - Email archiving task routes that contain the P8 4.x Create Email Instance task if you want to enable deduplication across email instances

  Update these task routes manually after you installed Content Collector. You can cancel the upgrade if you need to back up your database first.

  During the update process , a log file is written to the directory *ICCInstallDir*/ctms/log. For the Email Connector and the general settings of Content Collector the name of the log file is afu_mailconnector_upgrade.log. If a file with this name already exists in the directory, the suffix *_n* is appended, for example afu_mailconnector_upgrade_2.log. For other upgrades, the information is written to Windows event logs.
- Change the account of a service to the account that you used before the upgrade.
- Enable the Configuration Web Service again by using the Configuration Manager.
- Modify the log settings for the Web Application in the Configuration Manager. During the upgrade, old log settings are overwritten. The default log file is named afu_webapplication_trace_*n*.log and is written to the directory *ICCInstallDir*/ctms/log, where *n* is a number in the range from 1 to the value that is specified as the maximum number of log files to be written. If you changed the log directory in an earlier Content Collector release, you may want to change the log directory to the previous log directory after you upgrade to Version 2.2.

7. Optional: If you use a Web Application Server with Content Collector other than the embedded Web Application Server, upgrade the IBM Content Collector Web Application on that Web Application Server:
   a. Stop the IBM Content Collector Web Application service by clicking **Start → All Programs → IBM Content Collector → Stop Services → Stop ICC Web Applications**.
   b. On the computer running the Web Application Server navigate to the directory that contains the Content Collector .dll files of the earlier Content Collector release and delete the contents of this directory.
   c. Run the uninstall script afu_ewas_uninstall.bat that was delivered with the earlier Content Collector release. Running this script removes the IBM Content Collector Web Application and deletes the application server profile.
   d. Optional: Navigate to the home directory of your WebSphere Application Server installation and delete these files:
      - All .ear files
      - All those .bat files with file names that start with afu
      - All those .jacl files with file names that start with afu
      - All those .xml files with file names that start with afu
   e. Complete the tasks that are described in the topics about using an existing Web Application Server.
8. Optional: If you are upgrading from IBM Content Collector Version 2.1.1.2 or before, delete all custom event logs (.evt files) that were created by the previous version of Content Collector. IBM Content Collector Version 2.2

supports long file names for custom event logs. Therefore, you must manually remove previously created custom event logs.

a. Make backup copies of the custom event logs in the Windows Event Viewer. The name of a custom event log starts with CTMS.
b. Delete the event logs as described in "Deleting event logs" on page 772.

   Important: You must have permission to edit the registry on the computer on which you want to delete a log.

9. Start the Task Routing Engine service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start Task Routing Engine**. This service controls all other connector services. However, the Task Routing Engine service only starts those connectors that are referenced by valid, active task routes.
10. Start the IBM Content Collector Web Application service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start ICC Web Applications**. This service is required for document preview, search, and restore.
11. If you configured email through SMTP, start the IBM Content Collector SMTP service by clicking **Start → All Programs → IBM Content Collector → Start Services → Start SMTP Receiver**. This service is required for archiving email through SMTP.

You can now use IBM Content Collector Version 2.2.

Note that documents that are archived in IBM Content Manager are stored in an item type.

- If you create item types for documents that are archived from an email system, you can select which data model should be used. Content Collector Version 2.2 supports two data models for the email item type, the bundled data model and the compound data model. You can, however, also continue using existing item types. If you installed IBM Content Collector Text Search Support and want to use the indexer for the text search, you must enable all existing item types before you can process these item types by using the indexer.

  To create item types or to enable existing item types, use the setup tools that you start by clicking **Start → All Programs → IBM Content Collector → Set-up Tools** or **Tools → Set-up Tools** in the Configuration Manager.
- When you create new item types, additional archive mapping and search configuration files are created. These files are stored in the directory `<installDir>\Configuration\initialConfig\data\search\output\cm`, where *installDir* is the directory where you installed IBM Content Collector.

  To be able to search on documents that are stored in the new item type, you must merge the contents of the additional configuration files with any configuration information for archived-data access that is stored in the Content Collector configuration database. If no archived-data access configuration exists in the configuration database, the new configuration files are imported automatically.

  While you can merge archive mapping information manually or add the new item type by using the graphical user interface of the Configuration Manager, you must merge search configuration information manually. To do so, export the existing configuration files. Check the new configuration files; most likely you will have to update the collection ID and the collection name. Remember that the collection ID and the collection name must be identical. Then, add the

contents of the new configuration files to the exported files. Import the updated files into the configuration database and save the new configuration.

Restart the Web Application Server for the changes to take effect.

- To search archived documents, you must group the email item types in collections. Bundled item types and compound item types must be in different collections. However, you can search across collections.

  For more information about collections, see "Configuring the access to archived data" on page 228.

Upgrade FileNet P8 task routes for email archiving with records declaration or for email archiving if you want to deduplicate across email instances.

Related tasks

"Changing the user account of a service" on page 213

Upgrading FileNet P8 task routes

In IBM Content Collector Version 2.2, you must manually upgrade the FileNet P8 task routes for email archiving that contain a records declaration task and the task routes that contain the P8 4.x Create Email Instance task if you want to enable deduplication across email instances.

Update the task routes:

- Reconfigure records declaration task routes. In earlier versions of Content Collector, the P8 4.x Declare Record task was invoked with the output of the P8 4.x Save Prepared Text as XML task only. A decision point controlled that only those items were passed into the P8 4.x Declare Record task that were neither attachments nor duplicates.

  In Content Collector Version 2.2, the P8 4.x Declare Record task can handle duplicate email items. When you work with the FileNet P8 email data model and declare records, each version of the distinct email instance (DEI) is declared as a separate record. If one of the versions was already declared as a record, this information is logged, and this version of the DEI is skipped.

  1. In your task route, change the output path after the P8 4.x Create Email Instance task. Connect the path that is configured with the rule for duplicate email instances to the P8 4.x Declare Record task instead of connecting this path directly to the EC Prepare Email for Stubbing task.
  2. Adapt the rule for the decision point after the P8 4.x Save Prepared Text as XML task so that only items that are not attachments are passed to the P8 4.x Declare Record task. Attachments are passed directly to the EC Prepare Email for Stubbing task.

- Enable deduplication across email instances in your task routes. You can do this in one of these ways:
  - Import one of the task route templates that are shipped with Content Collector Version 2.2 and adapt it to your needs.
  - Enable deduplication in the P8 4.x Create Email Instance task in your existing email archiving task routes and add a decision point and rules for handling duplicate and non-duplicate email instance objects immediately after this task.

    The path that is configured with the rule for duplicate email instances must point to the EC Prepare Email for Stubbing task. The path that is configured with the rule for non-duplicate email instances must point to the P8 4.x Save Prepared Text as XML task.

Moving from CommonStore to Content Collector

To move from IBM CommonStore to IBM Content Collector, you must either install IBM Content Collector on the existing CommonStore servers or install IBM Content Collector on new dedicated servers and move to production before you shut down the existing IBM CommonStore installation.

When you move from IBM CommonStore to IBM Content Collector, the two systems are not merged. The data model of archived documents is different for each product.

After you move from IBM CommonStore to IBM Content Collector, your mailboxes contain a mix of the following types of documents:

Documents that are not archived
These documents still need to be archived using IBM Content Collector.

Documents that were archived by using IBM Content Collector
These documents have already been archived using IBM Content Collector.

Documents that were archived by using IBM CommonStore
These documents are referred to as legacy documents and can be accessed only by using IBM Content Collector legacy support. Legacy documents can only be viewed and restored. Restored legacy documents cannot be restubbed. Legacy documents are not migrated to conform with the IBM Content Collector data model.

If you install IBM Content Collector on a new dedicated server, you can upgrade the IBM CommonStore clients to IBM Content Collector in stages because the CommonStore clients can still be used in IBM Content Collector with limited functionality. If you upgrade clients in stages, ensure that you group these clients in a way that the same mailboxes are not accessed for archiving at the same time by both IBM CommonStore and IBM Content Collector.

None of the CommonStore clients can access the IBM Content Collector client interface. Only the IBM Content Collector clients can access both documents that are processed by using IBM Content Collector (to archive and restore) and legacy documents (to restore only).

IBM Content Collector Outlook Extension can be installed on client machines without having to first remove CommonStore for Exchange Server Outlook Extension. IBM Content Collector will automatically remove CommonStore for Exchange Server Outlook Extension before installing IBM Content Collector Outlook Extension.

If users continue using their CommonStore for Exchange Server offline repository as the offline repository in IBM Content Collector, the messages in this repository will be preserved and new messages archived using IBM Content Collector will be added to this repository as well.

Legacy support software

Depending on the strategy that you choose to move to IBM Content Collector, you might need to install legacy support software on the CommonStore servers.

IBM Content Collector legacy support is available as:

- A component of the IBM Content Collector Server installation which is automatically installed if IBM Content Collector is installed on the existing CommonStore servers
- A separate installation package that must be installed on the CommonStore servers if IBM Content Collector is installed on a new dedicated server

Note: Legacy support for CommonStore for Lotus Domino is only available on Windows. For AIX®, you must install the IBM Content Collector Server on a new dedicated machine and manually add all CommonStore functionality.

IBM Content Collector legacy support consists of two parts:

CommonStore Server (archpro)
Provides access to archived content from backend systems. This server includes an HTTP task that enables viewing of an archived document in a stubbed message.

CommonStore for Exchange Server Task and CommonStore for Lotus Domino Task
Allows you to restore archived legacy documents to a mailbox.

The CommonStore for Exchange Server Task configuration required for legacy support is defined in IBM Content Collector using the Configuration Manager.

The CommonStore for Lotus Domino Task configuration required for legacy support is provided through a new Notes application template that replaces the design of the CommonStore for Lotus Domino configuration database with the changed IBM Content Collector design.

When you install legacy support on the CommonStore servers, archiving by using IBM CommonStore is no longer possible. You can only view and restore legacy documents.

If legacy support is installed as a separate component on the CommonStore servers, legacy support removes IBM CommonStore and installs IBM Content Collector legacy support in the directory where CommonStore was installed.

If legacy support is installed on the same server as IBM Content Collector, CommonStore is removed, the CommonStore configurations remain in the former CommonStore directories and the legacy support is installed in the IBM Content Collector directory tree. This requires adapting the existing configuration.

You must import the CommonStore for Exchange Server configuration data that is stored in the Active Directory and is required for legacy support into the IBM Content Collector configuration database by using the Configuration Manager.

Important: Do not delete any public folder documents. Existing public folder documents that were created by CommonStore for Exchange Server are required by IBM Content Collector legacy support. Existing CommonStore for Exchange Server clients and the new IBM Content Collector Web Application must have access to these documents to restore legacy documents.

The CommonStore for Lotus Domino configuration data that is required for legacy support is made available to IBM Content Collector through a template update that adapts both the CommonStore for Lotus Domino configuration for use by the Configuration Manager and the CommonStore for Lotus Domino job database.

Which legacy repositories are supported

You can only access legacy documents from repositories used by the following IBM CommonStore versions:

- CommonStore for Exchange Server Version 8.3.1
- CommonStore for Exchange Server Version 8.3.2
- CommonStore for Exchange Server Version 8.4
- CommonStore for Lotus Domino Version 8.3.2
- CommonStore for Lotus Domino Version 8.4

For older versions of CommonStore for Lotus Domino, it is recommended that you upgrade the design of the CommonStore for Lotus Domino Configuration database to CommonStore for Lotus Domino Version 8.4 level, before moving to IBM Content Collector.

Legacy documents in an offline repository

To restore legacy documents and view these in an offline repository differs depending on your CommonStore installation:

- For legacy documents archived using CommonStore for Exchange Server, all archived documents must be restored to the mailbox on the server before they can be copied to the offline repository (the local personal folder). Stubbed legacy documents are also restored to the server mailbox. However, as legacy documents restored to the server mailbox cannot be restubbed by using IBM Content Collector and should not be deleted, the size of the server mailbox might grow unpredictably large.
- For CommonStore for Lotus Domino, you must replicate the server and local mailboxes, and then copy all documents from the local replica to the offline repository. Stubbed legacy documents are restored directly from the archive to the offline repository. This means that the size of the server replica and the local replica will not increase unpredictably

Document search

You can search for legacy documents only by using the search application in IBM Content Collector if these documents were archived in a CommonStore item type for the document model BUNDLED and archiving type ENTIRE. The legacy documents that were found and are in the result list can be previewed and restored by using the search application.

If other CommonStore document models or archiving types were used, for example GENERIC_MULTIPART or GENERIC_MULTIDOC, these documents cannot be found and restored by using the IBM Content Collector search application. These documents can only be viewed and restored interactively, for example, by clicking a stub in the mailbox.

Related concepts

"Legacy update strategies"

Related tasks

"Installing Content Collector" on page 29

Legacy update strategies

You can move to IBM Content Collector by installing Content Collector on a separate server or installing it on the CommonStore server.

Use one of the following strategies:

Install Content Collector on a separate server

This is the recommended strategy. To use this strategy:

- Install Content Collector legacy support to the existing CommonStore servers. This disables archiving using CommonStore and starts the legacy support versions of the CommonStore server and the CommonStore for Exchange Server Task or the CommonStore for Lotus Domino Task.
- Install the Content Collector server on a separate machine. Start Content Collector.

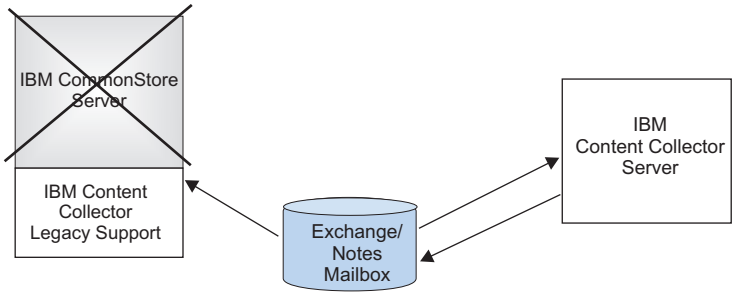


Figure 2. Installing IBM Content Collector on a separate server and installing legacy support on the CommonStore server

Install Content Collector on the CommonStore server

You can use this strategy if you plan to run Content Collector in scale out mode. To use this strategy, you must install the Content Collector server on the existing CommonStore servers. The server package includes the Content Collector legacy support. Start Content Collector and the legacy support for CommonStore.

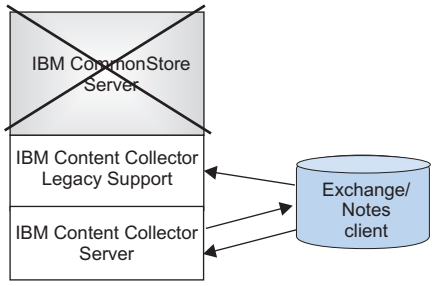


Figure 3. Installing Content Collector on the CommonStore server

Install Content Collector on a new server and manually add CommonStore functionality

You can use this strategy if you want to ensure that Content Collector has been successfully installed and configured before CommonStore is stopped. To use this strategy, you must install the Content Collector server on a new dedicated machine and manually add all CommonStore functionality instead of installing the legacy support software before you shut down the CommonStore servers. Start Content Collector and the legacy support for CommonStore.

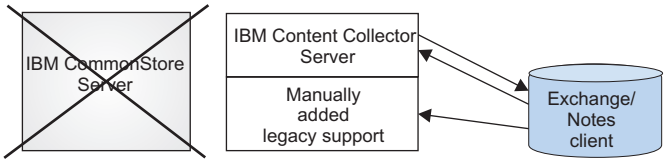


Figure 4. Installing Content Collector on a new server and manually adding CommonStore functionality

Related tasks

"Installing Content Collector on a separate server"

"Installing Content Collector on the CommonStore servers" on page 140

"Installing Content Collector on a new server and manually adding CommonStore functionality" on page 141

Installing Content Collector on a separate server

To move to Content Collector and view and restore CommonStore documents, you can install Content Collector legacy support on all existing CommonStore servers and install Content Collector on a separate dedicated server. This is the recommended strategy if you want to move to Content Collector from CommonStore.

Ensure that you back up your CommonStore servers and configuration before you install Content Collector.

After you install legacy support on the CommonStore servers, you can use a migration tool for CommonStore for Exchange and migration templates for CommonStore for Lotus Domino so that you can use CommonStore configurations in Content Collector.

To install Content Collector on a separate server, follow these main steps:

1. Suppress all CommonStore archiving functionality by unassigning policies in CommonStore for Exchange Server and stopping archiving tasks in CommonStore for Lotus Domino.
2. Install and test the Content Collector server on a dedicated machine.
3. On each CommonStore server, install IBM Content Collector Legacy support. Installing legacy support disables archiving to existing CommonStore mailboxes. Only viewing and restoring of legacy documents is enabled.

   Legacy support for CommonStore for Lotus Domino provides new CommonStore for Lotus Domino task profiles, which are used instead of the existing ones.

   For CommonStore for Exchange Server, the configuration that was stored in the Active Directory is imported into the Content Collector configuration database.
4. Configure and start legacy support:

Product	Task
CommonStore for Exchange Server	1. Configure legacy support 2. Start legacy support to view and restore documents

Product	Task
CommonStore for Lotus Domino	1. Configure and start legacy support to view and restore documents 2. Start legacy support for documents archived by using CommonStore for Lotus Domino

Installing Content Collector Legacy support

You install legacy support by running the IBM Content Collector installation wizard on the CommonStore servers. You must install legacy support on each CommonStore server.

For information about the prerequisites for IBM Content Collector legacy support see http://www.ibm.com/support/docview.wss?rs=484&uid=swg27012904#OS1

Running the IBM Content Collector Legacy support wizard enables legacy support in Content Collector. Legacy support removes CommonStore and installs Content Collector legacy support in the directory where CommonStore was installed. If legacy support is installed on a machine on which CommonStore is not installed, you prompted to enter the installation path for the legacy support.

To install legacy support:

1. In the Windows Explorer, change to the directory where you extracted the IBM Content Collector installation package.
2. Run the following command located in the `\Legacy` directory of the installation package on the CommonStore server: `install.exe`
3. Answer the remaining prompts.

Installing Content Collector on the CommonStore servers

You can install IBM Content Collector server and IBM Content Collector legacy support on the CommonStore servers.

Ensure that you back up you CommonStore servers and configuration before you install Content Collector.

After you install legacy support on the CommonStore servers, you can use a migration tool for CommonStore for Exchange and migration templates for CommonStore for Lotus Domino so that you can use CommonStore configurations in IBM Content Collector.

To install Content Collector on the CommonStore servers, follow these main steps:

1. If you are working in scale out mode, all nodes in Content Collector must be configured in the same way. Define one of the CommonStore servers as the Content Collector primary node and define all other servers as Content Collector extension nodes.
2. Install the Content Collector server on the CommonStore servers.

   No separate legacy support software needs to be installed because the legacy subsystem is installed with the Content Collector server.
3. Configure and start legacy support:

Product	Task
CommonStore for Exchange Server	1. Configure legacy support 2. Start legacy support to view and restore documents
CommonStore for Lotus Domino	1. Configure and start legacy support to view and restore documents 2. Start legacy support for documents archived by using CommonStore for Lotus Domino

Installing Content Collector on a new server and manually adding CommonStore functionality

To ensure that Content Collector is installed and configured for production before CommonStore is stopped, you can install Content Collector on a new dedicated server. Then, you manually configure the system to access documents that were archived by CommonStore.

You might need to contact IBM Software Support to do the manual configuration.

You can use a migration tool for CommonStore for Exchange and migration templates for CommonStore for Lotus Domino so that you can use CommonStore configurations in Content Collector.

The CommonStore configuration data required for viewing and restoring individual CommonStore documents must be added manually to the Content Collector server.

To install Content Collector on a new server and manually add CommonStore functionality, follow these main steps:

1. Install and test the Content Collector server on a new dedicated machine. The Content Collector server installation includes legacy support for viewing and restoring CommonStore documents. However, this legacy support does not include any task-specific CommonStore configuration data that is required to enable document restoring.
2. Copy the CommonStore Server configuration file `archint.ini` to the archpro instance directory (typically *install_dir*\server\instance01) on the Content Collector server. Adapt path names in the configuration file.
3. Suppress archiving in CommonStore by unassigning policies in CommonStore for Exchange Server and stopping all archiving jobs in CommonStore for Lotus Domino.
4. Configure legacy support:

Product	Task
CommonStore for Exchange Server	1. Configure legacy support 2. Start legacy support to view and restore documents

Product	Task
CommonStore for Lotus Domino	1. Configure and start legacy support to view and restore documents 2. Start legacy support for documents archived by using CommonStore for Lotus Domino

5. To enable viewing on the new dedicated server, update the Domain Name System (DNS) to point to the Content Collector server host and not the CommonStore Server host.
6. Test Content Collector before you move it to production.

Configuring Content Collector for CommonStore for Exchange Server legacy support

IBM Content Collector requires the configuration data related to each CommonStore for Exchange Server (CSX) task instance to restore documents archived by CSX.

After you have installed IBM Content Collector legacy support and before you begin configuring the legacy support, carry out the following:

1. Ensure that the following files were removed from the `bin` subdirectory on the server on which legacy support was installed, and if not, remove them:
   - `csx.exe` and `csxadmin.exe`
   - `csx_mapi.dll`
   - `csx.jar`
2. Open the configuration `archint.ini` file and delete the line that starts the CommonStore search server. The line contains the keyword START_SEARCHSERVER.

In CSX, all configuration data is stored in the Active Directory. In IBM Content Collector all configuration data is stored in a database that is accessed using the IBM Content Collector Configuration Manager.

To extract the configuration data for each CSX task from the Active Directory (or Active Directory Application Mode (ADAM)) and to add this data to the IBM Content Collector configuration:

1. Run `afucsx_exportTasks.exe` in a Command Prompt window on the server on which IBM Content Collector legacy support was installed. If you use an alternate AD, run `afucsx_exportTasks.exe -s ADAM_server -t ADAM_port`.

   The task description XML files and the log files are written to `<IBMAFUROOT>\Task\Export` on the IBM Content Collector server where IBMAFUROOT is an environment variable set for IBM Content Collector.

   If IBMAFUROOT has not been set, the task description and log files are written to `<CurrentDirectory>\Task\Export`.
2. Start the IBM Content Collector Configuration Manager.
3. Click **General Settings** → **Legacy Restore Exchange**.
4. Click .
5. On the General page of the tabbed notebook, enter a name and description for your configuration object.
6. Under **Import legacy task definition**

- Click **Browse** to navigate to and select the XML configuration file of the CSX Task. This is usually under `<IBMAFUROOT>\Task\Export`.
- Click **Import** to update your legacy task with the data obtained through running `afucsx_exportTasks.exe`.
- Save the legacy configuration.

  After the import, all fields in the tabbed notebook of the Legacy Restore Exchange pane are filled in automatically.

  The settings include the required CSX task parameters and all CommonStore Server and Exchange Server parameters.

  If you have not made any changes to your CSX environment, you need not adjust these settings.

  However, if you have moved or renamed any CSX servers, or want to change any trace, error, or log file paths, adjust these fields in the tabbed notebook of the Legacy Restore Exchange pane.

  Important: Check the transfer path settings. If the CommonStore Server and your CSX Task instance run on the same machine, the transfer paths in the Configuration Manager and in the server configuration profile (usually archint.ini, keyword TRANSFERPATH) MUST be identical.

  If both components run on different machines, the machines must share a common network directory. The transfer path specifications for both components must then point to the same physical directory.

  For a detailed description of these fields refer to the link under Related references at the bottom of the page.

Related reference

"Settings to enable restoring documents archived using CommonStore for Exchange Server"

Settings to enable restoring documents archived using CommonStore for Exchange Server

To restore documents archived using CommonStore for Exchange Server (CSX), you need the configuration data of your CSX Task. The configuration data includes settings pertaining to the CSX Task, the CommonStore Server and the Exchange Server.

If you have not made any changes to your CSX environment, then you do not need to adjust these configuration settings after you have imported the configuration data into the Configuration Manager. However, if you have moved or renamed any CSX servers, or want to change any trace, error, or log file paths, adjust these fields in the tabbed notebook of the Legacy Restore Exchange pane in the Configuration Manager.

These configuration settings are also described in detail in the IBM CommonStore for Exchange Server documentation.

CommonStore for Exchange Server Task settings on the General page

Task name

The name of the CSX Task instance that you want to address. Spell the name exactly as it is spelled in the CSX System Manager.

Description

A description of the task instance. This field is optional. Use it, for example, to add information about the repository.

Exchange user
The names of Exchange users to notify in case of events that require the intervention of an administrator. The selected users receive an email when an error has occurred, thus allowing them to take appropriate action. Type the email addresses of users in the **Exchange user** field and click **Add**. Selected users appear in the list box at the bottom of the property page. To delete a user, highlight the appropriate entry in the list box and click **Remove**. To remove all users, click **Delete All**.

CommonStore for Exchange Server Task settings on the Parameters page

Worker count
The number of worker threads that the CSX Task starts. The value must be an integer from 1 to 9. The worker is the subcomponent of the CSX Task that sends requests to the CommonStore Server. The requests that are sent from the worker to the CommonStore Server (archpro) are processed by agents. It is therefore recommended to configure the archpro program to start as many agents as workers.

Committer count
The number of committer threads that the CSX Task starts. The value must be an integer from 1 to 9. The committer is the subcomponent of the CSX Task responsible for changing the job-processing state in the task-job queue after a job was processed on the CommonStore Server. It also copies restored data back to a message store after a job was completed successfully on the CommonStore Server. Depending on your workload, you can improve the performance of the CSX Task by defining several committer threads to run in parallel. However, make sure that your hardware is capable of running this number of threads.

External port
The number of the port that you use to interact with a particular CSX Task instance. For example, you communicate over this port when you shut down a particular task instance. Specify an integer from 7000 to 9999. If you run more than one CSX Task instance on the same machine, specify a different port for each instance. Otherwise, conflicts arise.

Trace file name
The path and the name of the trace file. The trace file contains event messages issued by the CSX Task or the CommonStore Server, such as the startup of the CSX Task or the successful connection with the CommonStore Server.

Error file name
The path and the name of the error log file. The error log file contains all Exchange-related error messages.

Log directory
The directory to which the log files are written. To specify this directory, enter the full path. Log files are created daily. Their names are predefined, including date and time information. The file extension is log.

Trace Allows you to enable or disable tracing for operations related to the CSX Task (tracing related to the CommonStore Server must be enabled in the server configuration profile). To switch tracing on, select this box. To switch it off, clear it.

CommonStore Server settings on the CommonStore Server page

Host name

The full server name of the computer on which the CommonStore Server runs. If you only specify the machine name or IP address, and link to a user outside of your local intranet, the machine name cannot be resolved. If, on the other hand, this field contains a full server name, the Domain Name System (DNS) will resolve the server name in a link to the new IP address.

Important: Never use `localhost` as the host name.

Fixed port

The number of the port that the CSX Task uses to connect to the CommonStore Server. This port must be the same as the port that you specify in the server configuration profile (usually archint.ini) by using the ARCHPRO_PORT keyword. Specify an integer from 5000 to 9999. The default port is 8013.

Transfer path

The path to the directory that the CSX Task and the CommonStore Server use to exchange content. When you restore a document, the CommonStore Server copies it to the transfer directory, and the CSX Task picks it up from there. To specify this directory, enter the full path from the root directory.

Important: Check the transfer path settings. If the CommonStore Server and your CSX Task instance run on the same machine, the transfer paths in the Configuration Manager and in the server configuration profile (usually archint.ini, keyword TRANSFERPATH) MUST be identical.

If both components run on different machines, the machines must share a common network directory. The transfer path specifications for both components must then point to the same physical directory. To make the CommonStore Server and the CSX Task share a common network directory, use standard Windows file sharing, SAMBA, or any other Server Message Block protocol.

Exchange Server settings on the Exchange Server page

Job folder name

The name of the public folder on the Exchange server in which job documents are stored. Job documents basically identify the email that is to be restored.

Host name

Allows you to specify the Exchange servers that you want the CSX Task instance to work on. Type the fully qualified name of an Exchange server in the field, and click **Add**. The server names appear in the list box at the bottom of the page. To remove a server from the list, highlight the appropriate entry in the list box and click **Remove**. To remove all servers from the list, click **Remove All**.

Notes: A single Exchange server can only be served by one CSX Task instance. A single CSX Task instance, on the contrary, can serve multiple Exchange servers.

Starting legacy support to view and restore documents archived using CommonStore for Exchange Server

To view and restore documents that were archived using CommonStore for Exchange Server (CSX), Content Collector needs to have access to information in the CommonStore configuration file, and the CommonStore Server and CSX Task must be started.

To start legacy support:

1. Start the CommonStore Server (archpro) by entering the following command on the legacy CommonStore server instance directory:

   ```
   archpro
   ```

   See the CommonStore for Exchange Server documentation for details on starting the archiving engine and what archpro parameters to pass.

   When archpro is running, you can begin viewing documents that were archived by using CommonStore for Exchange Server.
2. Start the CommonStore for Exchange Server Task by entering the following command:

   ```
   afucsx -task taskname
   ```

   The executable file name for the CommonStore for Exchange Server Legacy Task changes from `csx` in CommonStore for Exchange Server to `afucsx`.

   The required parameters also change. The *taskname* value must be preceded by the keyword `-task`. The parameters **-p** and **-u** are removed. For example, if you started the CommonStore for Exchange Server Task by using the command `csx task1`, start the CommonStore for Exchange Server Legacy Task by using the command `afucsx -task task1`.

   In addition, the name of the administration tool changed from csxadmin to afucsxadmin.

   If you installed CommonStore as a service, you must change the service configuration file to reflect these changed legacy executable names and parameter, for example:

   ```
   PROCESS1 '"D:\CSX\bin\archpro" -i "D:\CSX\server\instance01\archint.ini"'
   PROCESS2 '"D:\CSX\bin\afucsx" -task my_task'
   ```

When the CommonStore for Exchange Server Legacy Task is running, you can restore documents in Content Collector that were archived by using CommonStore.

Configuring legacy support for documents archived using CommonStore for Lotus Domino

To view and restore IBM CommonStore for Lotus Domino documents, Content Collector legacy support must access a CommonStore for Lotus Domino job database called `CSLegacy\CSLDLegacyJobs.nsf` on each Domino server that contains documents that were archived by using CommonStore for Lotus Domino

This legacy job database does not match any existing CommonStore for Lotus Domino configuration, which means that the existing CommonStore for Lotus Domino tasks must be adapted to use this new job database.

However, not all legacy documents can use the new job database. If the legacy document that you want to restore is a document stub that contains a restore link (hotspot) that was inserted by CommonStore for Lotus Domino, the document cannot be restored by clicking the hotspot because the Lotus Domino job database

in which this hotspot creates a job is no longer monitored by Content Collector legacy support. However, the document can be restored using Content Collector.

When you install Content Collector legacy support, you also install an upgraded template for the CommonStore for Lotus Domino configuration database (`ICCConfig.ntf`) and an upgraded template for the CommonStore for Lotus Domino job database (`CSLDJobs.ntf`).

To configure legacy support for documents that were archived by using CommonStore for Lotus Domino:

1. Replace the design of the CommonStore for Lotus Domino configuration database with this new template `ICCConfig.ntf`. The new template includes a wizard that will help you to create new CommonStore for Lotus Domino Retrieve Task profiles which are based on your existing setup and which can also be used with IBM Content Collector.
2. The first time you open the Content Collector configuration database, select **Actions → Setup CS legacy Configuration**. This starts the wizard that enables you to verify and change existing CommonStore parameters. After you have finished, click **Run Setup**. Setting up the legacy configuration includes performing the following actions:
   - For each Domino server that was listed in the Domino Setup section, a new CommonStore for Lotus Domino task profile is created. Each task will be configured to process all jobs in the legacy job database on its Domino server and will run on the CommonStore server that was specified in the legacy configuration.
   - All existing CommonStore for Lotus Domino task profiles in the database are not changed. However, as the upgraded design only displays those elements that are relevant to legacy restore operations, not all of the existing documents will be visible (for example, no crawler-related documents can be seen).
   - A job database called `CSLegacy\CSLDLegacyJobs.nsf` is created on each of the Domino servers that is listed in the Domino Setup section. This job database uses the design of the `CSLDJobs.ntf` template in the data directory that is local to the Notes client from where the CommonStore legacy configuration is running.

     Important: Because the job database also includes changes that are required by Content Collector, ensure that the new `CSLDJobs.ntf` file is copied to the local data directory and not to a directory below `data`.
3. After the new task profiles are created, review the settings of the new profiles and adapt these where appropriate to better suit your environment.

Starting legacy support for documents archived by using CommonStore for Lotus Domino

To enable viewing and restoring documents that were archived by using CommonStore for Lotus Domino, Content Collector must have access to information in the CommonStore configuration file, and the CommonStore Server and CommonStore for Lotus Domino Task must be started.

To start legacy support for documents archived by using CommonStore for Lotus Domino:

1. Start the CommonStore Server (archpro) by entering the following command on the archpro instance directory (typically *install_dir*`\server\instance01`) on the server on which Content Collector legacy support was installed:

```
archpro
```

When archpro is running, you can begin viewing documents that were archived by using CommonStore in Content Collector.

2. Start the CommonStore for Lotus Domino tasks by using the newly created profiles. For example, if the given configuration database included profiles for three Domino servers, you must start three CommonStore for Lotus Domino task instances, and each task instance must use one of the newly created, server-specific task profiles.

   If you installed CommonStore as a service, you must change the service configuration file to reflect the changed legacy executable names and parameters. For example:

   ```
   PROCESS1 '"c:\program files\ibm\csld\bin\archpro.exe"
   -i "c:\program files\ibm\csld\server\instance01\archint.ini"'

   PROCESS2 '"c:\program files\ibm\csld\bin\csld.exe" -s NotesSrv1/ACME
   -n csld\CSLDConfigDB.nsf -p archive -i "c:\lotus\notes6\csld.ini"'
   ```

   Document stubs that are created by CommonStore for Lotus Domino contain a hotspot link, which when clicked, creates a restore job for the document. After moving from CommonStore to IBM Content Collector, these hotspots will no longer work because these links attempt to create the restore job in the same job database that was active when the document was archived. However, these job databases are no longer monitored by Content Collector legacy support.

3. Optional: Configure offline repository support:

Offline repository	Task
Reuse your CommonStore for Lotus Domino offline repository in Content Collector	When you install the IBM Content Collector offline repository support, point the Content Collector offline repository to your existing local database.
Use a new offline repository in Content Collector	Copy documents archived by CommonStore for Lotus Domino to your local repository as follows: • If the document is in archived status, the document will be copied from the local replica to your newly defined Content Collector offline repository after the offline repository support is installed and running. • If the document was stubbed in the CommonStore for Lotus Domino offline repository, the document must be restored to your local replica before it is picked up and copied to your local Content Collector repository.

Enabling the access to archived data for legacy item types

You can make IBM Content Manager repositories that were previously populated by IBM CommonStore for Exchange Server or IBM CommonStore for Lotus Domino available for search and restoring. The item types of the documents in these repositories are referred to as *legacy item types*.

Only legacy item types for the document model BUNDLED and the archiving type ENTIRE can be previewed and restored by using Content Collector Email Search. Document collections of other item types are not supported by the search application.

IBM Content Manager OnDemand and IBM Tivoli Storage Manager repositories and repositories that were populated byIBM FileNet Email Manager are not supported.

To configure the Content Collector to work with legacy item types:

1. Set the system environment variable AFU_MAILBOXID_ENABLE_EXACT_MATCH. You can use any value. This is required for Microsoft Exchange collections only.
2. Use the Configuration Manager to export the archive mapping file to a directory of your choice.
3. Add a collection definition.
4. Adapt the collection definition in the `search_mapping.xml` file:
   a. Change the collection name.
   b. Change the value of the <retrievable> element to false if the item type for the document model (or archive data model) is not of type BUNDLED, or the archiving type is not ENTIRE. See keyword ARCHIVETYPE in the CommonStore server configuration file. Setting the documents in a collection to be non-retrievable by setting the element to false prevents an error from being displayed in the search application if these documents are selected. Instead, a message is displayed that states that accessing the document is not supported. The following example shows the <retrievable> element set to true:

```
<doc_type_collection nm="CSMAIL" collectionType="ICC_Exchange_Email_Bundled">
  <retrievable>true</retrievable>
  <doc_types>
    <doc_type>
      <name>CSXMAIL</name>
    ...
    </doc_type>
  </doc_types>
 ...
</doc_type_collection>
```

5. Save the file.
6. Import the archive mapping file by using the Configuration Manager.
7. If you changed the collection name in the archive mapping file, change the collection name in the `search_config.xml` file to the name that you used in the `search_mapping.xml` file:
   a. Export the `search_config.xml` file by using the Configuration Manager to a directory of your choice.
   b. Change the name of the collection to the name that you used in the archive mapping file (in attribute nm="CSMAIL"). The name must be identical and is case sensitive. For example:

```
<collection>
    <name>CSMAIL</name>
    <declaration>
      <property>
        <name>EMAIL_DATE</name>
        <nls-key>jsp.searchrequest.daterange
...
```

   c. Save the `search_config.xml` file and import it by using the Configuration Manager.

8. Save the configuration.
9. Restart the Web Application Server.

Related tasks

"Adding collections" on page 694

"Viewing and restoring email from the search result list" on page 785

Moving from FileNet Email Manager or FileNet Records Crawler to Content Collector

Moving from FileNet Email Manager or FileNet Records Crawler version 4.0 or later to IBM Content Collector is a relatively standard installation process that requires the re-creation of your task routes and other settings to account for changes to tasks and connectors. Version 4, however, introduced major architectural changes that formed the basis of Content Collector, which makes a move from FileNet Email Manager version 3.7 or FileNet Records Crawler version 3.5 to Content Collector a more substantial undertaking that might require IBM Software Support input.

In addition to many administrative and functional improvements, IBM Content Collector includes virtually all of the functionality available to FileNet Email Manager and FileNet Records Crawler users, including automatic capture, custom property support, interactive archiving, and support for native email formats.

IBM Content Collector email management differs from its predecessors in several important ways:

- When email was archived, FileNet Email Manager automatically changed the icon of the email document in all mailboxes for all recipients. In IBM Content Collector, to change the icon of the email document in all mailboxes, the document must be archived from every mailbox. Icons can be changed only for Lotus Domino, not for Microsoft Exchange.
- A single Content Collector can collect from either Lotus Domino or Microsoft Exchange, but not both.
- Lotus Notes client functions depend on changes to Notes templates rather than a plug-in.
- Content Collector stores Lotus Domino email in the binary CSN format, which replaces the DXL format
- Content Collector provides legacy support for email stubs that FileNet Email Manager created, but searching and retrieval of such email must still occur in the target FileNet P8 repository, typically through Workplace XT. For example, a Content Collector user with email stubbed by FileNet Email Manager can click to retrieve them as before, but must search for them in Workplace XT instead of the Content Collector search tool. Of course, eDiscovery Manager users can retrieve all email regardless of capture method.

In all cases, you must complete most of the tasks normally associated with the fresh installation of a product, such as:

- Meeting all prerequisites, including the installation of supported versions of source and target repositories, web server software, and other required applications
- Installing IBM Content Collector, preferably on a new server, enabling you to test thoroughly before switching
- Configuring one or more connectors and connections

See “Installing Content Collector” on page 29

FileNet Email Manager and FileNet Records Crawler version 4.0 or later

Because FileNet Email Manager, FileNet Records Crawler versions 4.0 and IBM Content Collector use the same task route-based architecture, the upgrade process is relatively simple. Changes to task routes, tasks, and connectors require you to re-create your task routes in the new system, but many settings can remain the same, the Initial Configuration wizard sets up your source and target connectors and configures the included task route templates for you, and most target repository tasks are renamed primarily to identify them as specific to FileNet P8.

Tip: To facilitate the re-creation of your task routes and connectors, you can run your old and new systems side-by-side on different computers, but ensure that you are not running both systems in your production environment while you do so (two systems competing to process the same data can produce unexpected results). You can automate the transfer process somewhat by using the task route templates that Content Collector supplies.

FileNet Email Manager version 3.7 and FileNet Records Crawler version 3.5

Version 4.0 changed FileNet Email Manager and FileNet Records Crawler significantly by replacing the concepts of the index templates and profiles with a visual task route model that persists in IBM Content Collector. This new visual model supplies much more granularity and power in creating archival rules.

Given the number of new features and options available in Content Collector, you should review your current archiving use cases to determine if new functionality would benefit your organization. IBM Software Support can assist in the analysis of your templates and profiles to guide you in the generation of tasks and task routes that match your existing setup. You should upgrade directly to Content Collector; you do not need to upgrade to FileNet Email Manager or FileNet Records Crawler version 4.0 or later as an intermediate step.

Important: You can perform your own migration by building task routes based on your existing templates and profiles, but you should use the task route templates that IBM Content Collector supplies.

Contacting IBM Software Support

Moving from FileNet Connector for Microsoft SharePoint Document Libraries to Content Collector

Moving from IBM FileNet Connector for Microsoft SharePoint Document Libraries to IBM Content Collector requires the re-creation of your task routes and other settings to account for changes to tasks and other components.

Before installing IBM Content Collector for Microsoft SharePoint, you must

- record the settings, including connections, task route details, and metadata, that you want to copy to the new application
- uninstall IBM FileNet Connector for Microsoft SharePoint Document Libraries; you can do so after installing Content Collector only if you are installing the new application on a different server

Remember: You cannot run both applications on the same server, so if you installed IBM FileNet Connector for Microsoft SharePoint Document Libraries on a SharePoint farm, you must uninstall it from each Web Front End server on the farm.

In all cases, you must complete most of the tasks normally associated with the fresh installation of a product, such as:

- Meeting all prerequisites, including the installation of supported versions of source and target repositories, web server software, and other required applications
- Installing IBM Content Collector, preferably on a new server, enabling you to test thoroughly before switching
- Configuring one or more connectors and connections

See "Installing Content Collector" on page 29 and "Installing Content Collector for Microsoft SharePoint in GUI mode" on page 63

IBM Content Collector for Microsoft SharePoint and IBM FileNet Connector for Microsoft SharePoint Document Libraries share the same basic architecture but differ in several key ways. Increased modularity, which adds flexibility and streamlines data entry, has the greatest impact on migration:

- The **SP Collector** task, which is the top node on each SharePoint task route, replaces the separate SharePoint Collectors list that appeared under **Navigation** → **Collectors**.
- Connections and collectors are now separately configurable: for each **SP Collector** task in a task route, you can select any available connection.
- Custom metadata is also separately configurable: you can create multiple sets of user-defined metadata, then select a different set for each collection source in an **SP Collector** task.

To configure IBM Content Collector to match your existing IBM FileNet Connector for Microsoft SharePoint Document Libraries configuration:

1. Configure the SharePoint connector Add a connection for each connection that you copied from your collection sources in the Document Libraries application (you might have considerably fewer connections than before, because the new architecture eliminates repetition of connection information). The new **SP Create File** task requires a temporary directory, which you can change during configuration.
2. Optionally, Create one or more user-defined metadata sets that contain your custom metadata types. You must create at least one set. If you have much metadata, multiple sets can shorten the list of metadata available for mapping in a specific task route.
3. Recreate your task routes, mapping old tasks to new ones and adding decision points and rules.

Table 41. Task mapping

Old task	New task
<none> (Collector)	"Collecting from Microsoft SharePoint sites" on page 482
<none>	"SP Create File" on page 567 (optionally used for integration with IBM InfoSphere Classification Module)
Get Versions	"SP Get Versions" on page 568

Table 41. Task mapping (continued)

Old task	New task
Post processing	"SP Post-processing" on page 570
P8 4.x Capture	"P8 4.x Create Document" on page 543 or "P8 4.x Create Version Series" on page 550 (use the latter if you are collecting document versions)
P8 4.x File in Folder	"P8 4.x File Document in Folder" on page 555

4. To maintain existing SharePoint links to files archived to FileNet P8, you can run a link management task route to update link formats and eliminate links to moved or deleted content.

Verify these settings before you use them in your production environment.

Part 4. Configuring

© Copyright IBM Corp. 2008, 2010

Configuring Content Collector

Use the IBM Content Collector Configuration Manager to change the initial configuration, create connectors, task routes, collectors, and tasks.

The Configuration Manager

The Configuration Manager is the administrative user interface for configuring IBM Content Collector.

You use the Configuration Manager to manage all configurations of IBM Content Collector. The user account that is used to run the Configuration Manager must be part of the Administrators group on the local machine.

To change the language used on the interface, click **Tools** → **Language** from the Configuration Manager menu bar and select a language. Restart the Configuration Manager for the change to take effect. The first time the application is started, the language set under the operating system regional settings is taken.

The Configuration Manager contains a graphical tool called the Task Route Designer, which is used to open, create, modify, and configure task routes.

The Configuration Manager provides a view for the following configuration tasks:

- Data Stores
- Connectors
- Metadata and Lists
- General Settings
- Task Routes

You switch to a view by clicking the respective button.

The Configuration Manager interface consists of these major sections:

Left

The left side of the interface is the pane from which you initiate specific configuration tasks. Here, you have a list of options for what you can configure. When you configure task routes, you have two sections on top:

In the **Explorer** section, you can create a new task route or open an existing task route. Each task route has a context menu listing actions that can be performed on the task route, for example, open, close, delete, copy, and export.

In the **Toolbox** section, you can select tools to select items in a task route (Pointer), place a decision point in a task route (Decision Point), place a link in a task route (Link), place an audit log task anywhere in a task route (Audit Log), or select tasks specific to a connector. You can access the same tools through a context menu in the design pane or with keyboard shortcuts.

Middle

The middle section of the interface is the location for the display or design pane. You select a control or task icon on the left pane, and by clicking in

© Copyright IBM Corp. 2008, 2010

the design pane, you can take action on the control, for example, by removing it and adding another, or you can place a task, a decision point, or a rule within a task route.

Right The right side of the interface is the location for the configuration pane. You can edit configuration values in this pane. When you configure task routes, you can edit configuration values for collectors, tasks, decision points, rules, and so on.

Enabling security in the Configuration Manager

Anyone with permission to run IBM Content Collector Configuration Manager has permission to change the configuration of Content Collector. By enabling security within Configuration Manager, you can specify which users have read-only access and which users have read-write access.

If you are upgrading and Configuration Manager security was enabled, only valid users will be allowed to run Configuration Manager.

Applying security against unwanted changes to the configuration database is enabled in the IBM Content Collector Configuration Manager. Before security is applied the first time, any user can enable security. Thereafter, only a user with administrator access, or the system default user with administrator access called iccadmin can access security and set user access rights. The iccadmin user can be selected under **Tools** → **Security**.

Multiple users can have read access and start the Configuration Manager to view the configuration. However, only one user can have write access to the database at any one time, regardless of the number of users with write permission who are also logged in. The first user with write access to log in to the Configuration Manager can make configuration changes but must release this access before another user can have write access to the database.

To provide security against unwanted changes to the Content Collector configuration:

1. Select **Tools** → **Security** in the Configuration Manager.
2. Select **Enable Security**. Now you can add users, edit the rights of existing users, or delete users by clicking the respective icons above the list of users.

   User IDs can be added individually with their passwords and access rights, or, by selecting **Active Directory**, a valid Lightweight Directory Access Protocol (LDAP) user or group can be entered.

   Users can be added with *Read-only* access, *Read and write* access, or *Administrator* access.

   If you select to access registry data using LDAP you must enter the server host name, the port of the registry server, the security server ID with which to authenticate to the server and obtain privilege information about users, and the starting point for searching in the LDAP directory server.
3. To obtain write access to make configuration changes if are working in read-only mode, select **Edit** → **Acquire Write Access**.
4. To release the write access again to enable another user to make configuration changes, select **Edit** → **Release Write Access** or close the Configuration Manager. After you release the lock, the Configuration Manager will run in read-only mode.

Signaling changes to the configuration database

If you change and save configuration settings in the IBM Content Collector Configuration Manager, these changes are saved to the configuration database but will not be noticed by a running IBM Content Collector Task Routing Engine service.

To get the IBM Content Collector Task Routing Engine service to pick up configuration changes without you having to manually restart the IBM Content Collector Task Routing Engine service:

- Specify a data store synchronization interval.
- Signal that you have made configuration changes in the Configuration Manager after you have saved your changes.

If you need to change the data store synchronization interval, you can do this in the IBM Content Collector Configuration Manager under **Tools → Task Route Service Configuration**. Changing the value affects:

- The time it takes for the IBM Content Collector Task Routing Engine service to notice configuration changes.
- In a scale-out environment, the time it takes for the IBM Content Collector Task Routing Engine service to notice changes to the status of a node. Checking the lease time of a node is two times the synchronization interval.

Starting with IBM Content Collector V2.2, the default value of the data store synchronization interval is 5 minutes. Adapt this value according to your system needs:

- A lower value means that the IBM Content Collector Task Routing Engine service will query the configuration database more frequently.
- In a single node environment, a higher value means that the IBM Content Collector Task Routing Engine service may take longer to notice configuration changes.
- In a scale-out environment, a higher value means that the IBM Content Collector Task Routing Engine service may take longer to notice configuration changes and to check the status of the nodes.
- In a scale-out environment, if the primary node is overloaded and a very low value is used, there is a chance that a secondary node may incorrectly assume that the primary node has failed and may want to take over as the primary node even though it does not need to.

To get the IBM Content Collector Task Routing Engine service to pick up changes to the configuration database while it is running:

1. Ensure that a data store synchronization value is set.
2. Click the **Signal Change** tool bar button, or select **File → Signal Change**.

Adding, changing, or deleting configuration objects in the Configuration Manager

In the Configuration Manager, you create configuration objects of different types, such as data stores, connectors, user-defined metadata and lists, and so on. Sometimes, only one object of a certain type is allowed, and sometimes, you can create multiple objects of the same type. Follow the instructions here to add, change, or delete configuration objects.

- If an object is allowed only once (like an email server connector, for example), you cannot delete it. It stays as long as the Configuration Manager is installed. To provide the required details or change these, proceed as follows:
  1. Add or change the required details of the object in the configuration pane on the right.
  2. Click
- If multiple objects of a type are allowed (like user-defined metadata, for example), you can add, change, and delete objects. To do so, follow these steps:
  - To add an object, proceed as follows:
    1. Click
    2. Specify the required details or parameters in the configuration pane on the right.
    3. Click
  - To change an object, proceed as follows:
    1. Select the appropriate table row in the pane in the middle.
    2. Change the details or parameters of the object in the configuration pane on the right.
    3. Click
  - To delete an object, proceed as follows:
    1. Select the appropriate table row in the pane in the middle.
    2. Click

Keyboard commands for the Configuration Manager

The following tables lists the Configuration Manager navigation keyboard commands and Designer keyboard, mouse, and menu commands.

Configuration Manager navigation keyboard commands

Table 42. Configuration Manager navigation keyboard commands

Command	Description
Keyboard: Alt	Focus goes to the main window menu bar
Keyboard: Tab	Focus goes from the input control with focus to the next input control based on tab order. Focus should not leave the control if it is in a panel or focus should not leave the tab list control if the control is within a tab page. When the last input control has the focus and the tab key is hit again, focus will cycle back up to the first input control.
Keyboard: Shift+Tab	Same a previous description except in reverse order
Keyboard: F6	Focus goes from the panel with focus to the next panel with a selectable control. When a panel receives focus it will flash a dark border around it.
Keyboard: Shift+F6	Same as previous description except in reverse order
Keyboard: F2	Highlight the panel that currently has focus

Table 42. Configuration Manager navigation keyboard commands (continued)

Command	Description
Keyboard: Control+1 Control+2 Control+3 Control+4 Control+5	Navigate to the corresponding views: 1: Data Stores 2: Connectors 3: Metadata and Lists 4: General Settings 5. Task Routes

Designer keyboard, mouse, and menu commands

Table 43. Designer keyboard, mouse, and menu commands

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Tab	Highlights the next shape base on location. Navigation is done left to right and top to bottom.	Highlights the next shape base on location. Navigation is done left to right and top to bottom.	Highlights the selected shape.
Keyboard: Shift+Tab	Highlights the previous shape base on location. Navigation is done right to left and bottom to top.	Highlights the previous shape base on location. Navigation is done right to left and bottom to top.	Highlights the selected shape.

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Ctrl+D Menu: Detach	Detaches the shape if it is connected to a link and if the shape is allowed to be detached. The detached shape will be moved slightly to the left of the link and the link will reconnect itself to the next shape in the hierarchy. Begin and End shape cannot be detached. Audit shape can be detached. Decision Point and Task shapes cannot be detached if they have more than one link entering into the top of the shape. Decision Point shapes cannot be detached if they have more than one rule exiting from the bottom of the shape. Collector shapes cannot be detached. If the detached shape has a Decision Point upstream and downstream then the link will not reconnect because you cannot have a Decision Point follow another Decision Point. The detached shape should remain highlighted after detachment.	Nothing happens and menu item is disabled.	Nothing happens and menu item is disabled.

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Ctrl+Delete Menu: Delete	Deletes the shape if the shape is allowed to be deleted. A confirmation message is presented to the user prior to deletion so the action can be canceled. Begin and End shape cannot be deleted. Audit shape can be deleted. Decision Point and Task shapes cannot be deleted if they have more than one link entering into the top of the shape. Decision Point shapes cannot be deleted if they have more than one rule exiting from the bottom of the shape. If the deleted shape has a Decision Point upstream and downstream then the link will not reconnect because you cannot have a Decision Point follow another Decision Point. The next shape upstream will become the selected shape when the shape is deleted. If there is no shape upstream then the Begin shape will be selected.	Deletes the shape if the shape is allowed to be deleted. A confirmation message is presented to the user prior to deletion so the action can be canceled. Begin and End shape cannot be deleted. Audit shape can be deleted. Decision Point and Task shapes cannot be deleted if they have more than one link entering into the top of the shape. Decision Point shapes cannot be deleted if they have more than one rule exiting from the bottom of the shape. If the deleted shape has a Decision Point upstream and downstream then the link will not reconnect because you cannot have a Decision Point follow another Decision Point. The next shape upstream will become the selected shape when the shape is deleted. If there is no shape upstream then the Begin shape will be selected.	Nothing happens and menu item is disabled.
Keyboard: Enter Menu: Select	Selects the highlighted shape and loads the configuration GUI in the configuration pane to configure the data associated with the node.	Selects the highlighted shape and loads the configuration GUI in the configuration pane to configure the data associated with the node.	Nothing happens and menu item is disabled.

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Space	Attaches a Task, Begin, End or Decision Point shape with either a link tail, a link head (arrow) or the link center if the highlighted shape is in collision with a highlighted link or rule. If in collision with the link tail then the shape is attached to the tail. If in collision with the link head then the shape is attached to the head. If the link tail and head are already attached to shapes then the shape is attached to the middle of the link. A Decision Point cannot be attached if a Decision Point already exists either directly upstream or downstream from the attach location.	When moving the link tail it attaches the link to the task shape in collision if the shape will allow another link leading out. When moving the link head it attaches the link to the task shape in collision if the shape will allow another link leading in. If the task shape does not allow another link in or out then the designer will prevent the attachment. Rules: • Task shapes, Decision Points, Audit and End can have multiple link ins. • Decision Points can have multiple rules out. • Task shape and Begin shape can have one link out. • End and Audit can not have a link tail attached to it. • End shape will move the link head to the audit shape if the audit task is present. • Link head or tail cannot be attached to a Decision Point if the link being attached already has a Decision Point attached to it. • Link cannot be attached to a collector shape.	Nothing happens
Keyboard: Arrow keys	Shape moves in the direction of the arrow by 10 pixels.	Link will not move unless it is totally detached (no shape in or out).	Nothing happens

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Shift+Arrow keys	Task shape is resized in the direction of the arrow key. Decision Point, Audit, Begin, End and Collector are not resized.	Moves the tail portion of a link. If a link tail is attached to a shape then the connection is broken.	Nothing happens
Keyboard: Alt+Arrow keys	Nothing happens	Moves the head portion of a link. If a link head is attached to a shape then the connection is broken.	Nothing happens
Keyboard: Ctrl+Arrow keys	Nothing happens	Nothing happens	Moves the entire task route in the direction of the arrows.
Keyboard: Esc	Removes shape highlight and places context on the task route as a whole.	Removes shape highlight and places context on the task route as a whole.	Removes shape highlight and places context on the task route as a whole
Keyboard: Ctrl+ Menu: Zoom In	Zoom in the entire task route one zoom level to a maximum of 8 zoom levels. If at maximum zoom level then nothing happens and menu item is disabled.	Zoom in the entire task route one zoom level to a maximum of 8 zoom levels. If at maximum zoom level then nothing happens and menu item is disabled.	Zoom in the entire task route one zoom level to a maximum of 8 zoom levels. If at maximum zoom level then nothing happens and menu item is disabled.
Keyboard: Ctrl- Menu: Zoom Out	Zoom out the entire task route one zoom level to a minimum of 8 zoom levels. If at minimum zoom level then nothing happens and menu item is disabled.	Zoom out the entire task route one zoom level to a minimum of 8 zoom levels. If at minimum zoom level then nothing happens and menu item is disabled.	Zoom out the entire task route one zoom level to a minimum of 8 zoom levels. If at minimum zoom level then nothing happens and menu item is disabled.
Keyboard: Ctrl+Z Menu: Undo All Changes	Undo all changes made to the task route since the last save. A confirmation message is presented to the user prior to undo so the action can be canceled. If no changes have been made nothing will happen and the menu item will be disabled.	Undo all changes made to the task route since the last save. A confirmation message is presented to the user prior to undo so the action can be canceled. If no changes have been made nothing will happen and the menu item will be disabled.	Undo all changes made to the task route since the last save. A confirmation message is presented to the user prior to undo so the action can be canceled. If no changes have been made nothing will happen and the menu item will be disabled.

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Ctrl+S Menu: Save Task Route	Save changes made to the task route. If the task route is not valid the UI will not allow the user to save changes. If no changes have been made nothing will happen and the menu item will be disabled.	Save changes made to the task route. If the task route is not valid the UI will not allow the user to save changes. If no changes have been made nothing will happen and the menu item will be disabled.	Save changes made to the task route. If the task route is not valid the UI will not allow the user to save changes. If no changes have been made nothing will happen and the menu item will be disabled.
Keyboard: Home	Highlights the Begin shape regardless of what shape is highlighted in the task route.	Highlights the Begin shape regardless of what shape is highlighted in the task route.	Highlights the Begin shape regardless of what shape is highlighted in the task route.
Keyboard: End	Highlights the End shape regardless of what shape is highlighted in the task route.	Highlights the End shape regardless of what shape is highlighted in the task route.	Highlights the End shape regardless of what shape is highlighted in the task route.
Keyboard: F3 Menu: Align and Distribute	When a shape is highlighted (Begin, Task, Decision Point) the child shapes will be aligned along their tops and distributed evenly and horizontally under the parent shape. If the Begin shape is highlighted the Collector shapes will also be aligned and distributed. Align and distribute will always align the End shape and Audit shape (if present) centrally under the Begin shape and after the last shape based on the Y-location of the shape.	Nothing happens	Nothing happens
Keyboard: F4 Menu: Align Tops	Similar to Align and Distribute; however, only the tops of the child shapes are aligned. The shapes are not distributed horizontally.	Nothing happens	Nothing happens

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: F5 Menu: Distribute Horizontally	Similar to Align and Distribute; however, the child shapes are only distributed horizontally. Their tops are not aligned.	Nothing happens	Nothing happens
Keyboard: Ctrl+L Menu: Add Link	Adds a new link to the default location in the task route designer. The link will be highlighted and can subsequently moved into place.	Adds a new link to the default location in the task route designer. The link will be highlighted and can subsequently moved into place.	Adds a new link to the default location in the task route designer. The link will be highlighted and can subsequently moved into place.
Keyboard: Ctrl+I Menu: Add Link In	If the shape allows it a new link will be attached to the shape. The Link's head will enter the shape's top. If the shape does not allow it nothing will happen and the menu item will be disabled. When the Link is added it will be highlighted. Rules: • Begin shape will not allow a link in. • End shape will allow multiple links in but will move the link to the Audit shape if the audit task is included. • Task shape will allow multiple links in. • Decision Point shape will allow multiple links in; however, it will refuse a link in if the link being attached already has a Decision Point associated with it.	Nothing and menu item disabled	Nothing and menu item disabled

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Ctrl+O Menu: Add Link Out	If the shape allows it a new link will be attached to the shape. The Link's tail will exit the shape's bottom. If the shape does not allow it nothing will happen and the menu item will be disabled. When the Link is added it will be highlighted. Rules: • Begin and Task shape do not allow multiple links out. • Collector, Audit and End shape do not allow any links out • Decision Point shape will allow multiple links out (known as Rules); however, it will refuse a link out if the link being attached already has a Decision Point associated with it.	Nothing and menu item disabled	Nothing and menu item disabled
Keyboard: Ctrl+O Menu: Add Rule	Same as Add Link In but only applies to Decision Points. Note: When attaching to shapes together arrow shapes are called Links. When exiting a Decision Point arrow shapes are called Rules.	Nothing and menu item disabled	Nothing and menu item disabled

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Keyboard: Ctrl+E Menu: Add Decision Point	If the shape allows it the Decision Point will be inserted directly to the link exiting the shape. When the link is split to make room for the Decision Point the link exiting the Decision Point becomes a Rule. When the Decision Point is added it will be highlighted. Rules: • Collector, Audit and End shapes do not allow Decision Points to be added after them. • Decision Points do not allow Decision Points to follow them. • Shape must have a link leaving the shape or else the Decision Point can not be inserted.	If the link allows it the Decision Point will be inserted directly in the middle of the link. When the link is split to make room for the Decision Point the link exiting the Decision Point becomes a Rule. If the link has a Decision Point already associated with it the Decision Point will not be inserted in the Link. When the Decision Point is added it will be highlighted.	Decision Point is inserted at the default location of the designer and highlighted so it can subsequently be moved into place.
Keyboard: Insert	If the user used the tool box to add a new task and subsequently navigates to the designer control via the keyboard, the Insert key can be used to add the shape to the default location on the designer.	If the user used the tool box to add a new task and subsequently navigates to the designer control via the keyboard, the Insert key can be used to add the shape to the default location on the designer.	If the user used the tool box to add a new task and subsequently navigates to the designer control via the keyboard, the Insert key can be used to add the shape to the default location on the designer.
Keyboard: Menu key Mouse: Right-Click	Loads the context menu at the location of the shape. The menu items will be either enabled or disabled depending on what type of shape is in context and what the shape's state is.	Loads the context menu at the location of the shape. The menu items will be either enabled or disabled depending on what type of shape is in context and what the shape's state is.	Loads the context menu at the location of the shape. Enabled menu items are dependent on the type and state of the selected shape. In addition, menu items that are relevant to the Task Route as a whole will also be enabled.

Table 43. Designer keyboard, mouse, and menu commands (continued)

Command	Task shape has focus	Link shape has focus	No shape has focus
Menu: Add Collector	Adds a Collector shape to the designer above the Begin shape and to the left of the last Collector shape based on X-coordinates. The Collector shape will be highlighted after insertion.	Adds a Collector shape to the designer above the Begin shape and to the left of the last Collector shape based on X-coordinates. The Collector shape will be highlighted after insertion.	Adds a Collector shape to the designer above the Begin shape and to the left of the last Collector shape based on X-coordinates. The Collector shape will be highlighted after insertion.
Menu: Add Task	If the shape allows it the task shape will be inserted directly to the link exiting the shape. When the task shape is inserted it will be highlighted. Rules: • Collector, Audit and End shapes do not allow task shapes to be added after them. • Shape must have a link leaving the shape or else the task can not be inserted.	If the link allows it the task shape will be inserted directly in the middle of the link. When the task shape is added it will be highlighted.	Task shape is inserted at the default location of the designer and highlighted so it can subsequently be moved into place.
Keyboard: Shift Mouse: left mouse button down and movement	When the shift key is down while moving the shape with the mouse the shape will not be selected when the user releases the mouse button.	When the shift key is down while moving the link the part of the link closest to the mouse point will be moved.	Entire task route is moved.
Mouse: left mouse button down and movement	When the cursor is on the bottom-right corner of the task shape (not Decision Point, Begin, End, Collector or Audit) the shape can be resized.	When the cursor is closest to the tail the tail will move if the cursor is closest to the head the head will move.	Entire task route is moved.
Keyboard: F12	Toggles the focus between the design pane and the configuration panel to the right.	Toggles the focus between the design pane and the configuration panel to the right.	Toggles the focus between the design pane and the configuration panel to the right.
Keyboard: Control+Q	Switches between the main task route and the error task route.	Switches between the main task route and the error task route.	Switches between the main task route and the error task route.

Screen reader support for hover help

Screen readers might have problems reading out some hover help and error messages in IBM Content Collector. This hover help contains important information. You can enable Content Collector to display hover help and error messages as message boxes that can be read out by a screen reader.

To enable screen reader support for hover help, complete the following steps:

1. Open the configuration file *InstallDir*/ctms/ApplicationSettings.xml, where *InstallDir* is the installation directory of IBM Content Collector, in a text editor.
2. Locate the configuration option **EnableScreenReaderSupport** and set it to true.
3. The resulting file should contain the following line:

   ```
   <EnableScreenReaderSupport>true</EnableScreenReaderSupport>
   ```

Once you set this option, you can press Shift+F1 to display hover help as message boxes. Press Shift+F2 to display error messages as message boxes.

Setting up a configuration database

IBM Content Collector requires a connection to a data store in order to store configuration settings and data. The Content Collector Initial Configuration wizard prompts you for settings that are added to the configuration database (also referred to as data store). These settings can be viewed and modified in the Configuration Manager.

Without a configuration database, no configuration objects can be created because there is no place to store them. Only one configuration database can be active at a time. By default, this is the first data store connection that you created. If you add a new data store connection, you can make it active by selecting the checkbox **Make this my active data store**.

You can select to configure the following data store connections, one of which must be defined as the active configuration database:

DB2 database connection
: After you have named the connection, you can enter login credentials and log in to the database. A **Validate** button allows you to check if the connection works.

SQL Server database connection
: After you have named the connection, you can enter login credentials and log in to the database to verify the login credentials and view available databases.

Oracle database connection
: After you have named the connection, you can enter login credentials and log in to the database. A **Validate** button allows you to check if the connection works.

Adding or editing data store connections

You can add and edit connections to configuration databases.

Prerequisites: If your database management system is IBM DB2 or Oracle and is installed on a different machine than IBM Content Collector, a DB2 or Oracle client must be installed on the same machine as the Content Collector server. In addition,

the database or service that you want to use must be cataloged or registered using the client on this machine. Otherwise, you cannot not create data store connections to DB2 or Oracle databases.

Configuration settings are read from a single configuration database when the IBM Content Collector services start. This is the configuration database that is set to **Active**.

To add or edit a data store connection:

1. In the Configuration Manager, click **Data Stores** to switch to the Data Stores view. The available database systems appear in the explore pane.
2. In the explore pane, select the data store connection you want to add or edit. The display pane displays the currently configured connections.
3. In the display pane, do one of the following:
   - If you want to add a connection, click the **Add...** icon. A new connection is added, and you can enter the connection information in the configuration pane.
   - If you want to edit a connection, select the connection that you want to edit. The configuration pane displays current settings. Note that the database properties are displayed only for the active configuration database.
4. In the configuration pane, enter or edit a name and description for the connection.
5. To set the connection as the default configuration database to be used by the IBM Content Collector services and the Configuration Manager, select **Make this my active data store**.
6. To test the login credentials for the DB2 database connection, under **Database Information**:
   a. Enter the name of the database.
   b. Enter the login name.
   c. Enter the login password.
   d. Click **Validate database**. If the credentials are valid and IBM Content Collector can connect to the required database, the database properties associated with the connected database are displayed in the **Database Properties** section. In addition, if you make this database the active configuration database by selecting **Make this my active data store** and save the new connection, the Configuration Manager loads the configuration from the database.
7. To test the login credentials for the Oracle connection, under **Server Information:**
   a. Enter the service name.
   b. Enter the login name.
   c. Enter the login password.
   d. Click **Validate database**. If the credentials are valid and IBM Content Collector can connect to the required database, the database properties associated with the connected database are displayed in the **Database Information** section. In addition, if you make this database the active configuration database by selecting **Make this my active data store** and save the new connection, the Configuration Manager loads the configuration from the database.
8. To test the login credentials for the SQL database connection, under **Server Information**:

a. In the **Database server** text box, enter the name of the database server.
b. Enter the login name.
c. Enter the login password.
d. If you want to log in to the server and retrieve a list of databases available, click **Retrieve Databases**. A list of databases appears in the **Database Information** section.
e. In the **Database** list box, select the database you want to use with this connection. The database properties associated with the connected database are displayed. In addition, if you make this database the active configuration database by selecting **Make this my active data store** and save the new connection, the Configuration Manager loads the configuration from the database.

9. Click the **Save** icon to save, or to clear values without saving, click **Revert**.

Deleting a data store connection

You can delete configuration databases that are not marked as active. If you want to delete the active data store, you must mark another data store as active first.

Perform the following steps to delete a connection to a data store:

1. In the Configuration Manager, click **Data Store** to switch to the Data Stores view. From the **Navigation** section of the Configuration Manager, click **Data Store**. The configuration databases available for selection appear in the explore pane.
2. In the explore pane, select the type of data store that you want to delete. The display pane displays currently configured connections.
3. In the display pane, select the data store to be deleted, and click **Remove**. Before you can remove the connection to the active configuration database, you must select another data store to become the active data store.

Exporting or importing a configuration database

It is possible to export and import configuration databases. This is helpful when you need to do big changes to the configuration and want to back up your existing configuration before you do so.

- To export a configuration database:
  1. Select a data store connection.
  2. In the configuration pane, click **Export**.
  3. Specify the location and file name for the database snapshot.
- To import a configuration database:
  1. Select a data store connection.
  2. In the configuration pane, click **Import**.
  3. Select a file containing a database snapshot. The data will be imported into the selected configuration database.

Starting the Task Routing Engine

In IBM Content Collector, there is one service that monitors most of the collector services that run in IBM Content Collector. This is the IBM Content Collector Task Routing Engine service, or simply the *Task Routing Engine*.

Before you start the IBM Content Collector Task Routing Engine service for the first time, you should have created and configured a task route.

You need to start the IBM Content Collector Task Routing Engine service before configuration settings in the Configuration Manager will take effect. Whenever you make configuration changes that are monitored by the IBM Content Collector Task Routing Engine service, this service must be restarted so that the changes are considered. This also includes changes to task route configurations using the Configuration Manager.

To get the IBM Content Collector Task Routing Engine service to pick up configuration changes without you having to manually restart the IBM Content Collector Task Routing Engine service:

- Specify a data store synchronization interval. The synchronization interval determines how often the IBM Content Collector Task Routing Engine service checks for configuration changes. In a scale-out environment, the data store synchronization interval ensures that the IBM Content Collector Task Routing Engine service checks the status of the nodes.
- Click **Signal Change** in the Configuration Manager after you have saved all of your configuration changes.

Important: When the IBM Content Collector Task Routing Engine service is started, any active collector whose scheduled date has passed will start running immediately.

To start or restart the IBM Content Collector Task Routing Engine service manually, use any of the following methods:

- Go to **Start → All Programs → IBM Content Collector → Start Services → Start Task Routing Engine**.
- Go to the Services window of Microsoft Windows and start the IBM Content Collector Task Routing Engine service. For a list of which IBM Content Collector connector services are started by the IBM Content Collector Task Routing Engine service, see "Content Collector services" on page 178.

Configuring the task route service

Although the task route service is configured automatically, you can change settings after initial configuration.

To open the Task Route Service Configuration pane and change settings, click **Tools → Task Route Service Configuration**

On the **Task Route** tab, set the following configuration values:

1. Specify the **Concurrency settings**. These include:

   Queue size
   The task queue is a buffer that holds submitted units of work until a thread is available for them to be processed. The queue size determines how many items are queued. Together with the thread count, the queue size can be used to maximize system throughput and resource usage. If the queue size is too small, the task route service threads will be sitting idle. Large values can consume significant system resources because items will be waiting in the queue when the thread pool is exhausted.

   The default setting is 128. Increase this value up to 256 for servers that are sized to process more than 100 items/second.

   Thread count
   The thread count determines the maximum number of concurrent items that can be processed by the system. Each thread requires a connection

and uses memory and system resources. Together with the queue size, the thread count can be used to maximize system throughput and resource usage.

The default setting is 16. You may increase this setting up to a maximum of 30 threads for high throughput collection and archiving scenarios on servers that have 8 or more CPU cores available.

Datastore synchronization interval

This parameter is used if Content Collector is installed on more than one node and specifies how often the task route service should check the data store for the configuration change signal or for a change in node registration. This value is configured in seconds.

The default setting is 300 which means that synchronization occurs every 300 seconds (5 minutes).

If you set this value to 0, synchronization is switched off. For scale out, this value should always be set to a value greater than 0.

Maximum number of times to start connectors

This parameter specifies the number of times an attempt will be made at starting a connector. The default value is 3. If a connector unexpectedly terminates more than 3 times in an hour then it will not be restarted.

A maximum start count of 0 disables the restart feature.

The IBM Content Collector Task Routing Engine service monitors the connectors that it starts and makes sure that they are still running. The status of the monitored connectors is checked once a minute and after a connector communication failure.

If a connector terminates unexpectedly and the restart count for the connector has not been exceeded, the connector will be restarted and re-initialized. Documents that were being processed when the connector terminated are marked as failed and will be collected again. To omit duplicate documents in the repository, make sure that your task route includes a task that handles duplication.

What happens if the restart count for a connector is exceeded, depends on whether Content Collector is run on a single node or in scale out mode:

- On a single node, all task routes that reference the connector are temporarily disabled.
- In scale out mode, further processing depends on whether the connector failed on the primary node or on a secondary node:
  - On the primary node, the IBM Content Collector Task Routing Engine service disables all collectors on the node and the node switches to being a secondary node. Any work that needs to be processed by a task route that references the failed connector is rejected. Additionally, this node will not take over as the primary node should the new primary node fail.
  - On a secondary node, the IBM Content Collector Task Routing Engine service rejects any work that needs to be processed by a task route that references the failed connector. Additionally, this node will not take over as the primary node should the primary node fail.

An event is written to the main Windows event log named Application each time the IBM Content Collector Task Routing Engine service detects that a connector is not running or that a connector terminated more often than it is allowed to be restarted.

To restore normal processing after a connector failed too many times, restart the IBM Content Collector Task Routing Engine service on the node where the connector failed.

Reset start count after

After the specified period of time, the counter for the number of times each connector has been started is reset to 0.

The default value is 1 hour.

2. Specify the **Log Settings**. Task route service logs contain a detailed account of what the task route service is doing. When you are configuring the task route service log, you can select:
   - The type of data that should be written to the log file. Available options are listed here, from least to most verbose.

     Note: Log entries are cumulative as they become more verbose. A log level of the type *Information*, for example, will include *Information*, *Warning*, *Error*, and *Fatal* log entries.
   - The log file location.
   - To enable log file retention, including the number of days to retain the log file.
3. Specify the **Audit Log Settings**. The audit log contains system processing information for each item processed. When you are configuring an audit log, you can select:
   - **One log per collector** option (one file for all processing information), or many separate log files
   - The audit log folder location
   - The number of days to retain the audit logs
   - The maximum size of a log file in KB

Related tasks

"Configuring the primary node" on page 88

"Starting the IBM Content Collector Task Routing Engine service on the primary node and on the extension nodes" on page 91

Checking if Content Collector is running

IBM Content Collector is running when the respective Windows services are running.

You can check in the Services window of Microsoft Windows which services are running:

1. Go to **Start->Settings->Control Panel**.
2. Got to **Administrative Tools->Services**.
3. Check which services of IBM Content Collector are running. For a list of which services are needed and should be running, see "Content Collector services" on page 178. All IBM Content Collector services are prefixed with IBM Content Collector.

The status of a service does not necessarily indicate that IBM Content Collector is running and actually doing something. To check this, and to monitor system processing, see "Using the system dashboard" on page 759.

Configuring the settings for LDAP lookups during task route processing

In task routes, you can use rules and conditional values that depend on an LDAP group lookup in order to test property values for group membership. For example, you might want to check if the sender of an email belongs to a specific user group. To retrieve the required information from the LDAP server, you must configure the settings for LDAP lookups.

1. Start the Configuration Manager.
2. On the main menu, click **Tools → Task Route Service Configuration**.
3. To change the LDAP query options, select **Override default LDAP query options** on the **LDAP** tab. The fields under the check box, showing the default query options, become editable so that you can change the values:

   User Query
   Searches LDAP entries. By default, the query searches for user account names.

   Direct attribute
   Searches attributes of LDAP entries. By default, the query searches the *MemberOf* attribute of user account names.

   Indirect query
   Searches LDAP groups. By default, the query searches for group members.

   Group attribute
   Searches attributes of LDAP groups. By default, the query searches the name attribute of the specified group.
4. Select the appropriate type from the **Select LDAP server type** drop-down list. You can select between the following types:

   Active Directory
   For a Microsoft Active Directory LDAP server

   eDirectory
   For an eDirectory LDAP server

   SunOne
   For a SunOne LDAP server

   Domino
   For a Lotus Domino LDAP server
5. In the **Host name** field, type the host name of the LDAP server, for example:
   `myldap.xycorp.com`
6. In the **Port** field, type the proper port number for network communication with the LDAP server. The default value is 389.
7. From the **Protocol version** field, select the LDAP version that your server uses:

   2 For LDAP version 2

   3 For LDAP version 3
8. In the **Base distinguished name** field, type the prefix that identifies the proper set of account names that you want to search during authentication

requests. For example, if all LDAP user names used by IBM Content Collector start with CN=Name/O=IBM, then type this string in the **Base distinguished name** field. When the LDAP server needs to authenticate a user, it will search the set of account names starting with this prefix only, which can speed up authentication processes if many user accounts are defined on the LDAP server.

9. In the **User distinguished name** field, type the user account name of the IBM Content Collector default user.
10. In the **Password** field, type the password belonging to the LDAP user account that is specified in the **User distinguished name** field. Click **Test LDAP settings** to verify the credentials.
11. Click **OK**.

Related reference

"Operators for conditional values" on page 494

Content Collector services

Some IBM Content Collector services are automatically started after installation and initial configuration whilst others depend on imported task route templates to be started.

The table below shows at what time each service is started. Generally, connector services are started by the IBM Content Collector Task Routing Engine service as soon as their process is required in the workflow of any of the active task routes. Other services are started automatically or must be started manually, as detailed below.

You can change the user account of any of the available Content Collector services. If you change the accounts of a service that must be started manually, you must restart this service so that the changes can become effective. If the service is started by the IBM Content Collector Task Routing Engine service, click the Signal Change tool bar button or select **File → Signal Change**.**Signal** in the Configuration Manager to restart the IBM Content Collector Task Routing Engine service.

The following table lists the available Content Collector services with information on when the services are started.

Table 44. IBM Content Collector services

Content Collector service	Running status	User ID for the service
IBM Content Collector Task Routing Engine service	Must be started manually by the user after the IBM Content Collector Server installation and configuration	The user ID must have permission to start and stop the connector services monitored by the IBM Content Collector Task Routing Engine service: • Lotus Domino: Not the local system account. It can be the administrator account but may be another account as well. • Microsoft Exchange: Microsoft Exchange: An account with Exchange Enterprise Administrator rights. Exchange Server 2007 only: You can archive email from a mailbox with Full Mailbox Access right only, but public folder archiving and collecting from a server collection source will not be permitted. • SMTP: User account with full access rights to the SMTP message queue directory. • Microsoft SharePoint: User with full access rights to the temporary file location of the SharePoint Connector. • File System: Local user account with full access rights to all file locations. **Important:** In a scale-out environment, the IBM Content Collector Task Routing Engine service user on each node must have the permission to engage in pipe communication with the process on the other machines. The user right for this is Access this computer from the network.

Table 44. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector Web Application service	Must be started manually by the user after the IBM Content Collector Server installation and configuration	• Lotus Domino: Not the local system account. It can be the administrator account but may be another account as well. It is recommended that you use the same user account for the IBM Content Collector Email Connector service. • Microsoft Exchange: An account with Exchange Enterprise Administrator rights. It is recommended that you use the same user account for the IBM Content Collector Email Connector service. • SMTP: Local user account. In a scale-out environment, the user must be a domain user. • Microsoft SharePoint: User with full access rights to the temporary file location of the SharePoint Connector. • File System: Local user account with full access rights to all file locations. The IBM Content Collector File System Source Connector service account does not have to be the same as the IBM Content Collector Web Application service account as this service does not interact with the file system location that a document is captured from.
IBM Content Collector Utility Connector service	Started and monitored by the IBM Content Collector Task Routing Engine service	Not applicable

Table 44. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector Email Connector service	If Lotus Domino or Microsoft Exchange is installed and configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	• Lotus Domino: Not the local system account. It can be the administrator account but may be another account as well. • Microsoft Exchange: An account with Exchange Enterprise Administrator rights. • It is recommended that you use the same user account for the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service
IBM Content Collector SMTP Connector service	If SMTP is installed and configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	User account with full access rights to the SMTP message queue directory. In a scale-out environment, the user must be a domain user.
IBM Content Collector SMTP Receiver service	This service runs outside of the scope of the IBM Content Collector Task Routing Engine service and must be started manually.	User account with full access rights to the SMTP message queue directory. In a scale-out environment, the user must be a domain user.

Table 44. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector File System Source Connector service	Started if File System is configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	• If the File System Source Connector monitors files on other machines in the same domain, the user must have full access rights to the other machines. • The user of the IBM Content Collector Content Manager Connector service and the IBM Content Collector IBM FileNet P8 Connector service must have access to the document's content file, so that it can be archived into the repository. The simplest method is to run the file system source and the target repository connectors under the same service account. • The File System Source Connector must run as a user that has permissions to access metadata files, or there must be a trust relationship between the IBM Content Collector system and the system where the files are located.
IBM Content Collector Metadata Form Connector service	Started after the Metadata Form Connector is configured by the active task route configured to use the connector	Not applicable
IBM Content Collector Metadata Form Database service	Started automatically after the IBM Content Collector Server installation and starts the temporary metadata database (Derby database)	Not applicable

Table 44. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector SharePoint Connector service	If Microsoft SharePoint is installed and configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	• The IBM Content Collector SharePoint Connector service user must have full access rights to the temporary file location of the SharePoint Connector. • The user of the IBM Content Collector Content Manager Connector service and the IBM Content Collector IBM FileNet P8 Connector service must have access to the document's content file, so that it can be archived into the repository.
IBM Content Collector Content Manager Connector service	If Content Manager is installed and configured as target repository, this service is started and monitored by the IBM Content Collector Task Routing Engine service	For the required account privileges, see "Required Content Manager privileges for the connector" on page 201.
IBM Content Collector IBM FileNet P8 Connector service	If IBM FileNet P8 is installed and configured as target repository, this service is started and monitored by the IBM Content Collector Task Routing Engine service	For the required account privileges, see "Required IBM FileNet P8 privileges for the connector" on page 207.
IBM Content Collector Text Extraction Connector service	Started together with the IBM FileNet P8 connector service	Not applicable
IBM Content Collector File System Repository Connector service	After the connector is configured to be used in a task route, this service is started and monitored by the IBM Content Collector Task Routing Engine service	If the repositories of the File System Repository Connector are located on machines in the same domain but not on the IBM Content Collector source machine, the user must have full access rights to the other machines.
IBM Content Collector FileNet Image Services Connector service	After the connector is configured to be used in a task route, this service is started and monitored by the IBM Content Collector Task Routing Engine service	Local user account with full access rights.
IBM Content Collector Information Help Center service	Started automatically after the IBM Content Collector Server installation when the IBM Content Collector Configuration Manager is started	Not applicable

Related tasks

"Configuring the primary node" on page 88

"Configuring the extension nodes" on page 89

Configuring connectors

IBM Content Collector requires connections to third-party systems, including email servers, file servers, and repository servers. The Configuration Manager provides an interface to these third-party systems.

When you run the Initial Configuration wizard of IBM Content Collector, you select the email server, file server, or application server that you want to use (for example Microsoft Exchange, Lotus Domino, Microsoft Sharepoint, or Windows file system), as well as the type of the repositoryserver (for example IBM FileNet P8, or IBM Content Manager). The installation process then creates entries with default settings in the Configuration Manager for each selected server. You view these entries (called *connectors* as they represent a means to connect to the required server) in the Connectors window.

All connectors must be configured. Most connectors supply a default configuration that is suitable for common setups, but it is recommended that you review the default settings in case you want to change the log level or the location of temporary directories or log files.

Related tasks

"Configuring the File System Source Connector"

"Configuring the IBM FileNet P8 Connector and connections" on page 205

"Configuring an IBM Content Manager Connector and connections" on page 200

"Configuring a File System Repository Connector and repositories" on page 202

"Configuring an IBM FileNet Image Services Connector and connections" on page 204

"Configuring the Microsoft SharePoint Connector" on page 199

Configuring source connectors

A source connector provides an interface to a third-party system that contains documents that you want to work with in IBM Content Collector.

You must configure at least one source connector. You are prompted for the default configuration of each source connector that you select to install when you run the initial configuration. It is recommended that you review the default settings. To change the configuration of a connector, you must provide valid credentials for the source system first.

Configuring the File System Source Connector

Configure the File System Source Connector if you want to collect files from a directory. The File System Source Connector provides a connection to a file system, which enables IBM Content Collector to monitor source directories. When **File Systems** is selected during the initial configuration, the wizard creates a connector of this type with default settings. You might want to change these default settings or add further connectors.

To configure a connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.

2. In the **Connectors** pane on the upper left, select **File System Source**. The display pane displays the currently configured File System Source Connector.
3. In the configuration pane, in the **Log Settings** section, edit settings if required:
   a. From the **Log level** list box, select a log level.
   b. In the **Log file location** text box, edit the location, or click **Browse** to choose a location.
   c. **Attention!** By default, log files will be kept indefinitely. To select a number of days for which to retain the log file:
      1) Select **Enable log file retention**.
      2) Select the number of days to retain.
4. In the **Temporary Directory** section, edit the location in which to save temporary files, or click **Browse** to choose a location.
5. Save the configuration settings.

Related tasks

"Configuring connectors" on page 184

"Collecting file system documents" on page 479

Related reference

"Log levels" on page 770

"Required IBM FileNet P8 privileges for the connector" on page 207

Configuring the Email Connector

The Email Connector provides a connection to your email server. This can be Microsoft Exchange or Lotus Domino. The connection enables you to monitor mailboxes and other collection sources in your email system.

To configure an Email Connector:

1. Start the Configuration Manager and click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select **Email**.
3. On the General page of the configuration pane on the right, you can change the connector name and the description.
4. From the **Mail System** list, select your email system, which can be one of the following:
   - **Lotus Domino**
   - **Microsoft Exchange**

   Notes:
   - Support for the selected email system must have been selected during the installation, in the installation wizard. If, for example, Microsoft Exchange has not been selected during the installation, the option **Microsoft Exchange** will not be available in this window. That is, you will be unable to change the selection.
   - The IBM Content Collector Email Connector service and the IBM Content Collector Web Application service must run under an account that is not the local system account.

     If you select **Microsoft Exchange**, this account must have administrator privileges for Microsoft Exchange.
5. In the **Parameter Definition** section, edit settings if required:
   a. Specify the maximum number of mailboxes that the connector is allowed to process in parallel. Specify a number from 1 to 16.

b. Specify the maximum number of log file copies to keep for problem determination. If a problem occurs with the connector processes, the connector copies its log files to a separate subdirectory of the log file directory for later analysis. The `crash` directory is a subdirectory of the `support` directory and contains as many subdirectories for failures as you defined in the connector configuration. The name of the subdirectory for the failure information is the timestamp when the failure occurred, for example, `<icclog>\support\crash\<timestamp>`, where *<icclog>* is the path to the log file location. The Content Collector default path is `C:\Program Files\IBM\ContentCollector\ctms\log`.

   The value that you specify here determines the number of subdirectories that are created. This value also multiplies the storage requirements for log files, which are determined by the number of log files and their size. A value of 0 means that no log files are copied. Specify an integer from 0 to 65538.

c. In the **Retry interval** fields, specify the time interval after which documents for which processing failed are processed again. A time interval of 0 means that the collector retries to process the documents as soon as possible. Note that this does not apply for documents with a permanent problem. Specify an integer from 0 to 65538 and select a time unit.

d. Specify the maximum number of times that a collector attempts to process documents again when the processing failed before. When the specified limit is reached it is assumed that a general problem exists with an individual document. The collector then ignores the erroneous document so that it is not processed over and over again. You can find details about the erroneous document under **Tools → Blacklist**. Specify an integer from 1 to 65538.

6. In the **Log Settings** section, edit settings if required:
   a. From the **Log level** list box, select a log level.
   b. In **Log file location** type the full path to the directory in which you want to store the log files, or click **Browse** to choose a location.
   c. Select **Truncate log files** to split and truncate the log files. If **Truncate log file** is unchecked, a single log file with unlimited size is maintained.

      If **Truncate log file** is checked, enter the maximum number of log files to be created in the **Number of log files** field. In the **Log file size in MB** field, type the maximum size that a log file is allowed to reach. As soon as the first log file reaches this size level, a new log file is created. When the maximum number of log files has been reached, and all log files have also reached their maximum size, the oldest log file is overwritten with a new one. This is also known as the round-robin method.
7. Under **Logging Type**, select the format in which to store the log files. You can choose between the following options:

   Common Base Event
   Stores the log files in the Common Base Event format, which can be read by a number of tools for log analysis and reporting. Generally, these tools discover dependencies between log events and are capable of creating various reports to visualize these dependencies.

   Plain Text
   Stores the log file in a simple text format. This format requires less disk space than Common Base Event, but usually cannot be processed by analysis and reporting tools as easily.

8. Under **Working Directory**, in the **Location** field, type the full path to the directory in which you want to save temporary files. These files are created by collectors and provide the basis for further processing steps in a task route. The path must contain no other characters than a-z, A-Z, and 0-9 of the Latin-1 character set. Instead of typing the path, you can click **Browse** to choose a location.

   Note:

   - The working directory must be a local directory and not a shared network drive. If you use a shared network directory as working directory, this will decrease performance significantly.
   - The working directory that the email server connector uses to create and store temporary files must be on a separate and fast disk.

Related concepts

"Blacklist" on page 715

Related tasks

"Collecting documents for life cycle processing" on page 476

"Collecting email on request" on page 472

"Collecting documents automatically" on page 460

Related reference

"Log levels" on page 770

Specifying logon credentials for the email server:

You need to provide certain credentials so that an IBM Content Collector can connect to an email server. To connect to a Lotus Domino server, you need to specify the location of a valid `notes.ini` file and a password. To connect to an Exchange server, you need to provide the name of the server.

Depending on your selection in the **Mail System** list on the General page, perform one of the following steps:

- For **Lotus Domino**:
  1. In the **Notes.INI location** field, type the name of the directory in which a suitable `notes.ini` file can be found. The default directory in which the `notes.ini` is created by the initial configuration is `<INSTALL_DIR>\notesdata\`. The default location is set after the initial configuration has run. The `notes.ini` file needs to contain the name of the Domino server to connect to and a reference to the id file of a user with administration rights for the server.
  2. In the **Password** field, type the password that belongs to the id file which is referred to in the `notes.ini` file.
  3. Click the **Validate** button to see if the credentials are correct.

     Note: User validation only works if the `notes.ini` file contains a *MailServer* entry. In other words, Lotus Notes only authenticates a user if the user accesses a Domino server. In a local setup (where no value is assigned to *MailServer* in the `notes.ini` file) an authentication does not happen, which means that the user validation always succeeds.
- For **Microsoft Exchange**:
  1. In the **Exchange server host name** field, type the name of the Exchange server that you want to connect to.

2. Select **Open mailboxes with privileged access** if the account running the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service has Exchange administrator rights. Otherwise, the account requires full access to all mailboxes to be archived and to the trigger mailbox
3. Select **Open public folders with privileged access]** if the account running the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service has Exchange administrator rights. Otherwise, the account requires the permission level Editor for the public folders to be archived and the permission level Reviewer for the parent folders.

Because the connector that you are configuring is a Windows service, you must specify the account and the password that you want to use to access the Exchange server on the properties sheet of the IBM Content Collector Email Connector service. Additionally, the Web Application service must run under the same user ID. See the related topic for information about how to change the user account that starts the IBM Content Collector Email Connector service and the Web Application service.

Related tasks

"Changing the user account of a service" on page 213

Specifying where to look up email recipient addresses:

When IBM Content Collector archives an email, the email needs to be associated with the correct email user so that the mailbox of this user can be identified when the email is stubbed or restored.

A user can have several email addresses. This is why an address book or Active Directory lookup is performed to identify the user and associate the email with a unique ID or account.

Depending on your selection in the **Mail System** list on the General page, perform one of the following steps:

- For **Lotus Domino**: Specify the Lotus Notes address books that you want to use to resolve data in email recipient fields. Select one of the following:

  Public If selected, the public names and address book (names.nsf) on the **Lotus Domino** servers processed will be used to resolve data in email recipient fields.

  User defined
  Use this option to specify another address book to resolve data in email recipient fields. At least one address book must be selected. To add an address book (Lotus Notes database), click **Add**. In the window that opens:

  1. Specify the Lotus Notes server where the database resides. If the server is on a local machine, leave the server field blank.
  2. Specify a database that inherits its design from `pubnames.ntf`, which means, that its design is based on the public Names & Address Book (NAB), for example, `dir1\companyNAB.nsf`.

  To edit or remove a database, select a row in the table and click **Edit** or **Remove**.

Note: Even if **User defined** is selected, you may specify the public names and address book itself, by entering `names.nsf` on one of the Domino servers in your Domino environment.

- For **Microsoft Exchange**: Specify the Active Directory that you want to use to resolve data in email recipient fields.

  Tip: Make sure that the machine that runs the IBM Content Collector server is part of the Active Directory domain. Otherwise the Email Connector might fail.

  1. Select what type of Active Directory server to use:

     Domain default
     One of the Active Directory servers to which the DNS lookup of the domain name resolves. If there are several Active Directory servers, IBM Content Collector picks one of them.

     User defined
     Another global catalog server.

  2. If you selected **User defined** in step 1, also provide the following information in the appropriate field:

     Host name
     The name of the Active Directory server. The global catalog must be enabled on that server.

     LDAP port
     The port number for communication with the specified domain server. This is the default domain server of the domain that the specified global catalog server belongs to. The default is 389.

     Global catalog port
     The port number for network communication with the specified global catalog server. The default is 3268.

  3. Specify the user account that is used to access the Active Directory server. Enter the SMTP address of the user. If Active Directory cannot be accessed using the SMTP address, use the user distinguished name. The address format is:
     - For an SMTP address: `iccuser@mycompany.com`
     - For the user distinguished name: `CN=iccuser,CN=Users,DC=company,DC=com`
  4. Type the password of this user in the **Password** field.
  5. Click **Validate** to ensure that you entered the correct credentials.

- Save the configuration settings.

Configuring the SMTP Connector

Configure the SMTP Connector if you want to archive SMTP/MIME email. The SMTP Connector provides a connection to an internal SMTP Receiver, which receives SMTP/MIME email and writes the documents to a directory on the disk, from where they are processed.

Unlike the Email Connector, the SMTP Connector does not actively collect email documents from the mail server. Instead, the source SMTP server must actively send all SMTP/MIME email to be processed to IBM Content Collector. Accordingly, SMTP archiving cannot be used for mailbox management, but for compliance objectives. The Content Collector SMTP Receiver receives these SMTP/MIME email documents and writes them to a designated directory. The SMTP collector picks up the mail from the designated directory.

To configure the SMTP Connector:

1. Start the Configuration Manager and click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select **SMTP**.
3. On the General page of the configuration pane on the right, you can change the connector name and the description.
4. In the **Parameter Definition** section, edit settings if required:
   a. Specify the maximum number of log file copies to keep for problem determination. If a problem occurs with the connector processes, the connector copies its log files to a separate subdirectory of the log file directory for later analysis. The `crash` directory is a subdirectory of the `support` directory and contains as many subdirectories for failures as you defined in the connector configuration. The name of the subdirectory for the failure information is the timestamp when the failure occurred, for example, `<icclog>\support\crash\<timestamp>`, where *<icclog>* is the path to the log file location. The Content Collector default path is `C:\Program Files\IBM\ContentCollector\ctms\log`.

      The value that you specify here determines the number of subdirectories that are created. This value also multiplies the storage requirements for log files, which are determined by the number of log files and their size. A value of 0 means that no log files are copied. Specify an integer from 0 to 65538.
   b. In the **Retry interval** fields, specify the time interval after which documents for which processing failed are processed again. A time interval of 0 means that the collector retries to process the documents as soon as possible. Note that this does not apply for documents with a permanent problem. Specify an integer from 0 to 65538 and select a time unit.
   c. Specify the maximum number of times that a collector attempts to process documents again when the processing failed before. When the specified limit is reached it is assumed that a general problem exists with an individual document. The collector then ignores the erroneous document so that it is not processed over and over again. You can find details about the erroneous document under **Tools** → **Blacklist**. Specify an integer from 1 to 65538.
5. In the **Log Settings** section, edit settings if required:
   a. From the **Log level** list box, select a log level.
   b. In **Log file location** type the full path to the directory in which you want to store the log files, or click **Browse** to choose a location.
   c. Select **Truncate log files** to split and truncate the log files. If **Truncate log file** is unchecked, a single log file with unlimited size is maintained.

      If **Truncate log file** is checked, enter the maximum number of log files to be created in the **Number of log files** field. In the **Log file size in MB** field, type the maximum size that a log file is allowed to reach. As soon as the first log file reaches this size level, a new log file is created. When the maximum number of log files has been reached, and all log files have also reached their maximum size, the oldest log file is overwritten with a new one. This is also known as the round-robin method.
6. Under **Logging Type**, select the format in which to store the log files. You can choose between the following options:

   Common Base Event
   Stores the log files in the Common Base Event format, which can be read by a number of tools for log analysis and reporting. Generally,

these tools discover dependencies between log events and are capable of creating various reports to visualize these dependencies.

Plain Text
Stores the log file in a simple text format. This format requires less disk space than Common Base Event, but usually cannot be processed by analysis and reporting tools as easily.

7. Under **Working Directory**, in the **Location** field, type the full path to the directory in which you want to save temporary files. These files are created by collectors and provide the basis for further processing steps in a task route. The path must contain no other characters than a-z, A-Z, and 0-9 of the Latin-1 character set. Instead of typing the path, you can click **Browse** to choose a location.

   Note:
   - The working directory must be a local directory and not a shared network drive. If you use a shared network directory as working directory, this will decrease performance significantly.
   - The working directory that the email server connector uses to create and store temporary files must be on a separate and fast disk.

Changing the message queue directory:

The SMTP Receiver receives SMTP/MIME email that is processed by the SMTP connector. You must specify the message queue directory to which the SMTP Receiver writes the received email.

On the Connection tab of the SMTP Connector configuration, specify the directory where the received email is queued for processing.

Important: Select a directory that is highly reliable. When the SMTP Receiver receives a document and stores it in this directory, a notice is sent out to the originating server that the message was successfully received. If, for some reason, documents in the message queue directory are deleted before they are archived, the email is lost. So you must make sure that the message queue directory is safe. Also ensure that no antivirus scanning runs on this directory.

If you change the message queue directory, you must ensure that you do not lose any documents.

To change the message queue directory:

1. Stop the SMTP Receiver. Click **Start** → **All Programs** → **IBM Content Collector** → **Stop Services** → **Stop SMTP Receiver**.
2. Wait until all messages are processed. Check the current message queue directory to see if there are further messages to be processed.
3. When the message queue directory is empty, delete it. The SMTP Receiver will not start until the old directory has been removed. This is done to prevent data loss.
4. Change the message queue directory on the Connection tab of the SMTP Connector configuration.
5. Restart the SMTP Receiver. Click **Start** → **All Programs** → **IBM Content Collector** → **Start Services** → **Start SMTP Receiver**.

Enabling SMTP authentication:

IBM Content Collector can use SMTP authentication to ensure that STMP/MIME email is only processed if it is received from an authenticated sender.

Only **SMTP basic authentication** is supported in IBM Content Collector. Transport Layer Security (TLS)/Secure Socket Layer (SSL) authenticated and encrypted connections are not supported.

If you want to archive SMTP/MIME email, you must configure your SMTP server to forward all email to IBM Content Collector. Enable SMTP authentication for your SMTP server and in IBM Content Collector to make sure that only email that is sent from your server is received and processed by IBM Content Collector.

Important: You should protect the Content Collector SMTP Receiver by a firewall in addition to SMTP authentication or, if you do not enable SMTP authentication, instead of SMTP authentication. The Content Collector server should be accessible only to the mail server that forwards the email. In particular, the standard SMTP port 25 should be closed to other traffic to prevent malicious or accidental archiving of email that originates from other mail servers.

To enable SMTP authentication in IBM Content Collector:

1. On the Connection tab of the SMTP Connector configuration, select **Enable SMTP authentication**.
2. In the **User ID** field, specify the user ID that you want to use for SMTP authentication.
3. In the **Password** field, enter the password that you want to use for SMTP authentication.
4. Use the same authentication credentials in your email system.

Setting up your mail server to send journal reports:

The IBM Content Collector SMTP Connector processes journaled email that is received from an external mail server. You must configure your mail server so that it sends out journal reports to the Content Collector SMTP Connector.

Refer to the documentation of your mail server for instructions on how to set up SMTP journaling to an external host. The following sections contain instructions for setting up Microsoft Exchange 2007, Microsoft Exchange 2010, and Lotus Domino.

Setting up Microsoft Exchange to send journal reports:

Follow these instructions to set up Microsoft Exchange 2007 or Microsoft Exchange 2010 to send journal reports to the IBM Content Collector SMTP Connector.

To set up Microsoft Exchange to send journal reports to an external SMTP address:

1. Configure an SMTP connection to the external SMTP host.
2. Configure the domain where the IBM Content Collector server is running as remote domain.
3. Configure an email contact for the external SMTP recipient.
4. Enable SMTP authentication.
5. Set up a journal rule that sends journal reports for all internal and external email to the external host.
6. Adjust the retry interval for outbound messages to make sure no data is lost if the SMTP Receiver is down.

Configuring an SMTP connection in Microsoft Exchange:

To be able to send email from the Microsoft Exchange server to addresses outside of the Microsoft Exchange domain, for example to the IBM Content Collector SMTP Connector, you must define a Send connector in Microsoft Exchange.

To define a Send connector for the IBM Content Collector server:

1. In the Exchange Management Console, navigate to **Organization Configuration** → **Hub Transport**.
2. Select the Send Connectors tab.
3. Right-click and select **New Send Connector...**.
4. On the Introduction page in the New Send Connector window, specify a name for the Send connector and select **Custom** as intended use.
5. Click **Next**.
6. On the Address space page, click **Add...** to add a new address space.
7. In the SMTP Address Space window, enter the address of the domain to where the external mails should be sent, so in this case the domain of the IBM Content Collector server.
8. Click **OK** and then **Next**.
9. On the Network settings page, select **Route all mail through the following smart hosts**.
10. Click **Add...** and enter either the IP address or the fully qualified domain name of the machine where the IBM Content Collector SMTP Connector is running.
11. Click **OK**.
12. Make sure that **Use the External DNS Lookup settings on the transport server** is not selected.
13. Optional: On the Configure smart host authentication settings page, enable basic SMTP authentication. You can also do that later.
14. Click **Next** until you reach the last page of the wizard, then click **Finish**.

Configuring a remote domain for Microsoft Exchange:

A remote domain is a domain outside of the Active Directory service forest. You must define the settings for message transfer between the Microsoft Exchange server organization and the domain of the machine where your IBM Content Collector server is running.

To create and configure a remote domain for the IBM Content Collector server:

1. In the Exchange Management Console, navigate to **Organization Configuration** → **Hub Transport**.
2. Select the Remote Domains tab.
3. Right-click and select **New Remote Domain...**.
4. In the New Remote Domain window, specify a name for the remote domain and the domain name of the IBM Content Collector server.
5. Click **New**.
6. Right-click the domain that you just created and select **Properties**.
7. On the Message Format or Format of original message sent as attachment to journal report tab, select the following:
   a. Under **Exchange rich-text format**, select **Never use**.

b. Under **Character Sets**, select **Unicode (UTF-8)**.

8. Click **OK**.

Configuring the external SMTP recipient in Microsoft Exchange:

You must create an email contact for the external SMTP recipient to which Microsoft Exchange sends the journal messages and configure the settings for this recipient.

To create and configure the external SMTP journal recipient:

1. Create a new email contact to use as journal recipient.
   a. In the Exchange Management Console, navigate to **Recipient Configuration** → **Mail Contact**.
   b. Right-click and select **New Mail Contact...**.
   c. In the wizard, select **New Contact** and click **Next**.
   d. Specify a name and an alias for the journal recipient.
   e. Click **Edit...** and specify the external SMTP address for the journal recipient. This must be an SMTP address that points of the IBM Content Collector Server. You can choose any mailbox name in the SMTP address, because only the domain of the SMTP address is needed to determine the receiving SMTP server.
   f. Click **OK**.
   g. Click **Next**, then click **New**.
2. Configure the new email contact and hide it from the address book.
   a. Right-click the email contact that you just created and select **Properties**.
   b. On the General tab, change **Use MAPI rich text** to **Never**.
   c. Select **Hide from Exchange address lists**. This ensures that no Outlook user can select the contact from the address book and send messages directly to the journal address.
   d. Click **OK**.
3. Customize the message format settings for the email contact. These settings must be changed using the Exchange Management Shell.
   a. Open the Exchange Management Shell.
   b. Set the **UsePreferMessageFormat** parameter to `true`. If this parameter is set, email that is sent to this email contact always uses the message format settings that are configured for this email contact, instead of the settings for the remote domain or the email sender. Enter the following command, where *Alias* is the alias of the email contact:

      ```
      Set-MailContact -Identity Alias -UsePreferMessageFormat $true
      ```

   c. Set the **MessageBodyFormat** parameter to `TextAndHtml`. The **MessageBodyFormat** parameter specifies the message body format for messages that are sent to the mail contact. Setting this to TextAndHtml means that the MIME messages received by the IBM Content Collector Server will contain both a plain text representation of the body as well as an HTML representation.
   d. Set the **MessageFormat** parameter to `Mime`. This parameter specifies the message format. If the message body format is set to TextAndHtml, the message format must be set to Mime. Enter the following command, where *Alias* is the alias of the email contact:

      ```
      Set-MailContact -Identity Alias -MessageFormat Mime
      ```

4. Configure the email contact to accept only messages that were sent by the Microsoft Exchange recipient. Journal reports are always sent by the Microsoft Exchange recipient, so constraining the email contact to accept no email from other senders secures the journal reports that are sent to the IBM Content Collector SMTP Connector.
   a. Open the Exchange Management Shell.
   b. Set the **RequireSenderAuthentication** parameter to `true` and the **AcceptMessagesOnlyFromSendersOrMembers** parameter to `"Microsoft Exchange"`. This enables sender authentication and requires the sender to be the Microsoft Exchange recipient. Enter the following command, where *Alias* is the alias of the email contact:

      ```
      Set-MailContact -Identity Alias -AcceptMessagesOnlyFromSendersOrMembers
      "Microsoft Exchange" -RequireSenderAuthentication $true
      ```

Enabling basic SMTP authentication for Microsoft Exchange:

You should enable basic SMTP authentication to make sure that the IBM Content Collector SMTP Connector accepts only authenticated journal email and rejects all other email.

In the IBM Content Collector Configuration Manager, enable SMTP authentication and define a user ID and password. To do so, switch to the Connectors view, select **SMTP**, and select **Enable SMTP authentication**.

To configure SMTP authentication in Microsoft Exchange:

1. In the Exchange Management Console, navigate to **Organization Configuration** → **Hub Transport**.
2. Select the Send Connectors tab.
3. Select the send connector that you created and click **Properties**.
4. On the Network tab, select the smart host that you created and click **Change**.
5. In the Configure Smart Host Authentication Settings window, select **Basic Authentication**.
6. Make sure that **Basic Authentication over TLS** is not selected.
7. Enter the user ID and password that you defined in the Content Collector Configuration Manager.

Setting up a journal rule for Microsoft Exchange:

To set up journal archiving, create a journal rule that sends journal reports for all internal and external email to the IBM Content Collector SMTP Connector.

To set up a journal rule in Microsoft Exchange:

1. In the Exchange Management Console, navigate to **Organization Configuration** → **Hub Transport**.
2. Select the Journaling tab.
3. Right-click and select **New Journal Rule...**.
4. Specify a name for the journal rule.
5. As email address, select the external contact that you created before.
6. Select a journal scope. Select **Global** to journal all internal and external messages.
7. Click **New**.

Adjusting the retry interval in Microsoft Exchange:

If the SMTP Receiver is down for some reason and the journal reports cannot be delivered, Microsoft Exchange tries to deliver the email again until the retry interval has passed. If the SMTP Receiver does not accept the email within this timeframe, it is not processed.

To prevent data loss, you must make sure that the retry interval for outbound messages in Microsoft Exchange is configured such that all messages are processed. To ensure this, the retry interval must be larger than any downtime of the SMTP Receiver.

To adjust the retry interval in Microsoft Exchange:

1. In the Exchange Management Console, navigate to the transport properties.
   - On a machine that has the Edge Transport server role installed, select **Edge Transport** and click **Properties** under the server name.
   - On a machine that has the Hub Transport server role installed, select **Server Configuration** → **Hub Transport**, then select a server and click **Properties** in the action pane.
2. Select the Limits tab.
3. Specify a value for **Outbound connection failure retry interval (minutes)**. The value must be between 1 and 28800 minutes (20 days).
4. Click **OK**.

Setting up Lotus Domino to send journal reports:

Follow these instructions to set up Lotus Domino to send journal reports to the IBM Content Collector SMTP Connector.

To set up Lotus Domino to send journal reports to an external SMTP address:

1. Configure an SMTP connection to the external SMTP host.
2. Configure the domain where the IBM Content Collector server is running as foreign SMTP domain.
3. Enable SMTP authentication.
4. Set up a journal rule that sends journal reports for all internal and external email to the external host.
5. Adjust the retry interval for outbound messages to make sure no data is lost if the SMTP Receiver is down.

Configuring an SMTP connection in Lotus Domino:

To be able to send email from the Lotus Domino server to addresses outside of the Lotus Domino domain, for example to the IBM Content Collector SMTP Connector, you must define a connection to the external host in Lotus Domino.

To define a connection to the IBM Content Collector server:

1. In the Domino Administrator Client, navigate to **Configuration** → **Messaging** → **Connections**.
2. Click **Add Connection**.
3. Switch to the Basics tab.
   a. Under **Connection type**, select **SMTP**.

b. Under **Destination Server**, specify any virtual server name that does not exist. This host name is not used, and it must not resolve to any IP address.
c. Under **Destination Domain**, specify any virtual domain name that does not exist. This domain name is not used, and it must not resolve to any IP address.
d. Under **SMTP MTA relay host**, specify the host where the IBM Content Collector SMTP Connector is running. Lotus Domino must be able to resolve the IP address of the host, either through a DNS entry or by using the hosts file on the local machine.

4. Switch to the Replication/Routing tab.
   a. Disable the Replication task.
   b. Set the Routing task to **SMTP Mail Routing**.
5. Click **Save & Close**.

Configuring a foreign SMTP domain for Lotus Domino:

A foreign SMTP domain document provides servers in a Domino domain with information on where to transfer mail that is sent to external SMTP addresses. You must define the settings for message transfer between the Lotus Domino server domain and the domain of the machine where your IBM Content Collector server is running.

To create and configure a foreign SMTP domain for the IBM Content Collector server:

1. In the Domino Administrator Client, navigate to **Configuration** → **Messaging** → **Domains**.
2. Click **Add Domain**.
3. On the Basics tab under **Domain type**, select **Foreign SMTP Domain**.
4. Switch to the Routing tab.
   a. Under **Internet Domain**, specify the external domain name of the machine where the IBM Content Collector server is running.
   b. Under **Domain name**, specify the same domain name that you specified when you configured the connection.
5. Click **Save & Close**.

Enabling basic SMTP authentication for Lotus Domino:

You should enable basic SMTP authentication to make sure that the IBM Content Collector SMTP Connector accepts only authenticated journal email and rejects all other email.

In the IBM Content Collector Configuration Manager, enable SMTP authentication and define a user ID and password. To do so, switch to the Connectors view, select **SMTP**, and select **Enable SMTP authentication**.

To configure SMTP authentication in Lotus Domino:

1. In the Domino Administrator Client, navigate to **Configuration** → **Server** → **Configurations**.
2. Open the configuration document for the server that you want to configure.
3. Select **Router/SMTP** → **Basics**.

4. Under **Relay host for messages leaving the local internet domain**, specify the external domain name of the machine where the IBM Content Collector server is running.
5. Under **Use authentication when sending messages to the relay host**, select **Enabled** and specify the name and password that you defined in the Content Collector Configuration Manager.

Setting up a journal rule for Lotus Domino:

To set up journal archiving, create a journal rule that sends journal reports for all internal and external email to the IBM Content Collector SMTP Connector.

To set up a journal rule in Lotus Domino:

1. In the Domino Administrator Client, navigate to **Configuration** → **Messaging** → **Configurations**.
2. Open the configuration document for the server that you want to configure.
3. Select **Router/SMTP** → **Restrictions and Controls...** → **Rules**.
4. Create a rule that journals messages.
5. Select **Router/SMTP** → **Advanced...**.
   a. Under **Journaling**, select **Enabled**.
   b. As method, select **Send to mail-in database**.
   c. Under **Mail Destination**, specify the SMTP address of the IBM Content Collector SMTP Connector.
   d. Enable **Journal Recipients**.
6. Select **MIME** → **Advanced...** → **Advanced Outbound Message Options**.
7. Under **Always send the following Notes items in headers** specify at least the following field names:

   $JournalRecipients

   RecipientGroupsExpanded

   OriginalBcc

   Tip: You can add more fields and extract their content as custom metadata in the SMTP Connector. However, you must explicitly specify the fields that you want to include.
8. Click **Save & Close**.

Adjusting the retry interval in Lotus Domino:

If the SMTP Receiver is down for some reason and the journal reports cannot be delivered, Lotus Domino tries to deliver the email again until the retry interval has passed. If the SMTP Receiver does not accept the email within this timeframe, it is not processed.

To prevent data loss, you must make sure that the retry interval for outbound messages in Lotus Domino is configured such that all messages are processed. To ensure this, the retry interval must be larger than any downtime of the SMTP Receiver.

To adjust the retry interval in Lotus Domino:

1. In the Domino Administrator Client, navigate to **Configuration** → **Messaging** → **Configurations**.
2. Open the configuration document for the server that you want to configure.

3. Select **Router/SMTP → Restrictions and Controls... → Transfer Controls**.
4. Specify a value (in minutes) for **Initial transfer retry interval**.

Configuring the Microsoft SharePoint Connector

The Microsoft SharePoint Connector requires a connection to each SharePoint site or subsite from which you want to collect documents. When you select **Microsoft SharePoint** during the initial configuration, the wizard creates the connector with default settings. In Configuration Manager you can change these default settings and add or modify connections.

To configure the Microsoft SharePoint Connector:

1. In the Configuration Manager, navigate to **Connectors → Microsoft SharePoint**.
2. Under the **General** tab, name your connector and specify your logging requirements:
   a. From the **Log level** list box, select a log level. See descriptions of the levels at "Log levels" on page 770.
   b. In the **Log file location** text box, edit the location, or click **Browse** to choose a location.
   c. To limit the retention of log files:
      1) Select **Enable log file retention**.
      2) Select the number of days to retain log files.

      Attention: By default, log files are retained indefinitely.
3. In the **Temporary Directory** section, edit the location in which to save temporary files, or click **Browse** to choose a location. In scale-out mode, you must supply the location in UNC format, such as *\\server1\directory1*.
4. Under the **Connections** tab, select a connection to edit or delete, or click **Add Connection**.
5. Type the **User ID**, Windows Server **Domain**, and **Password** of a user who belongs to the SharePoint Site Collection Administrators group for this site.
6. Enter a valid URL to a top-level or sub-level SharePoint site, in the format: http://*server name*:*port number*/*path*/ The address must begin with `http://` or `https://` and might require a port number or path. Type a *port number* if the server is running multiple SharePoint web applications. Type a *path* to limit your collection to a subsite. Click **Validate** to verify that the application recognizes the address. If validation fails, see "SharePoint connector validation fails" on page 812
7. Save the configuration settings.

Related tasks

"Collecting from Microsoft SharePoint sites" on page 482

Related reference

"Microsoft SharePoint collection sources" on page 483

Configuring target connectors

A target connector provides an interface to the third-party system that serves as the target repository for IBM Content Collector.

You can either configure a new connector or review the configuration of a target connector that you have already defined, for example, during the initial configuration.

Configuring an IBM Content Manager Connector and connections

When IBM Content Collector is installed with the IBM Content Manager Connector repository option, the Configuration Manager will contain an entry for the IBM Content Manager Connector. The connector is installed with default log settings, which can be configured as you choose.

Prerequisites: Make sure that IBM Information Integrator for Content is installed on the IBM Content Collector server. IBM Information Integrator for Content is required for IBM Content Manager connections.

Hint: If IBM Content Manager is selected as a repository in the initial configuration wizard, the wizard creates an IBM Content Manager connection for you. By following the steps in this topic, you can create additional connections or modify settings of the existing connection. An IBM Content Manager connection must exist if you want to use the Content Manager tasks in a task route. Each connection contains a set of login credentials to a specific IBM Content Manager repository. When you configure Content Manager tasks, you must choose a configured Content Manager connection. This connection will be used by the IBM Content Manager Connector while performing the actions specified by the configured task.

For each connection, all Content Collector components or task routes that use this connection are listed in the **Connection Consumers** list section.

Steps to add or edit a connection are provided below as part of the steps to configure a connector.

Important: The name of the Content Manager repository connection is used in all stub links and cannot be changed once the connection is consumed.

To configure a connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select **IBM Content Manager**. The configuration pane displays a default name and description for the connector.
3. In the configuration pane, edit log settings if required:
   a. From the **Log Level** list box, select a log level.
   b. In the **Log file location** text box, edit the location, or click **Browse** to choose a location.
   c. To limit the retention of log files:
      1) Select **Enable log file retention**.
      2) Select the number of days to retain log files.

      Attention: By default, log files are retained indefinitely.
4. In the display pane, click the **CM 8.x Connections** tab. The display pane displays currently configured connections.
5. In the display pane, do one of the following:
   - To add a connection, click . The configuration pane is cleared of values.
   - To edit a connection, select the connection you want to edit. The configuration pane displays the current setting.

6. In the **CM 8.x Logon Information** section, enter a user name and password and the Content Manager repository to which you want to connect. The user account must have the appropriate privileges to access the repository. For further information see the topic about required user privileges for the IBM Content Manager Connector.
7. Click **Validate** to check the user privileges. If you do not click **Validate**, the privileges are checked when you save the connection. You can save the connection only if the user credentials that you specified have the required privileges.
8. In the **Item Type Lookup** section, you can provide a lookup between the default item type names and the item type names that are actually used. If you chose to override the default item type names in the Initial Configuration Wizard or if your item types are named differently in your Content Manager system, you must provide such a lookup if you want to load the default templates that are installed with IBM Content Collector.

   Note: Even if the default item type exists in your Content Manager system, you can provide an alternative lookup as long as it is compatible with the default item type.
9. Save the configuration settings.

Related reference

"Log levels" on page 770

"Required Content Manager privileges for the connector"

Required Content Manager privileges for the connector:

The user account under which the IBM Content Collector Content Manager Connector service is running requires a minimum set of privileges for the connector to be able to access the Content Manager repository and to perform all required tasks. If the user credentials that you specify do not have the minimum set of privileges associated, you cannot save the connection.

The following list contains actions and the required privileges.

Privilege	Required for this action
ItemAdd	Create items
ItemAddLink	Create a link between two items, either in the same item type or in different item types
ItemCheckInOut	Check out or lock an item, or check in or unlock an item that was checked out by the same user
ItemDelete	Delete items
ItemLinked	Set items to be linked by other items
ItemLinkTo	Link items to other items
ItemQuery	Retrieve items
ItemRemoveLink	Remove a link between two items, either in the same item type or in different item types
ItemSetSysAttr	Update system-defined attribute values of an item
ItemSetUserAttr	Update user-defined attribute values of an item

Privilege	Required for this action
ItemSQLSelect	Use SQL select statements to select items
ItemSuperCheckIn	Check in or unlock an item that was checked out by another user
ItemTypeQuery	Retrieve item type, component type, and related item type views
SystemQueryUserPrivs	Retrieve user information
UserACLOwner	Give ownership to a user ACL (access control list), allowing the owner to update and delete the user ACL

Configuring a File System Repository Connector and repositories

Use a File System Repository Connector to copy files to a location in the file system. Index files contain metadata associated with each file. You can have one index file for each file copied to the repository or one index file for a set of files.

An entry for the File System Repository Connector is created by default in the Configuration Manager. The connector is installed with default log settings, which can be configured as you choose.

Required: In addition, you must define one or more repositories in order to implement File System Repository tasks in a task route. For example, when configuring an FSR Create Document task you will be required to choose a configured file system repository. The File System Repository Connector then connects to this repository while performing the actions specified by the configured FSR Create Document task.

When defining a repository, you can specify how to create the index files containing metadata associated with the file, specifying:

- To create one index file to accompany each document. For example, an index file can be created for an email file and each of its associated attachment files.
- To create one index file for each set of documents. For example, one index file can be created for an email file and its associated attachment files.

You can also create index file classes, which are equivalent to a document class in an IBM FileNet repository. You can define properties for each class, which represent fields that will be included in the index file. You can then use the properties in rules, and in the property mappings fields in the FSR Create Document task.

Note: Steps to add or edit a repository and index file classes and properties are provided below as part of the steps to configure a connector.

To configure a connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select **File System Repository**. The configuration pane displays a default name and description for the connector.
3. In the configuration pane, edit log settings if required:
   a. From the **Log level** list box, select a log level.

b. In the**Log file location** text box, edit the location, or click **Browse** to choose a location.

c. Select a number of days for which to retain the log file:

- Select **Enable log file retention**.
- Select the number of days to retain.

Important: By default, log files will be kept indefinitely.

4. Click the **Repositories** tab. Set the following options to add or edit repositories and index files to accompany documents:

Repository name
The name to be used to refer to the repository in IBM Content Collector.

Index file per document
If selected, a single index file will be created for each "set" of documents. An example of a set of documents is an email file and its associated attachment files. If not selected, one index file will be created for each document.

Shortcut format
No value required. (this field is not used by the application)

5. In the **Document Classes** section, you can add or edit document classes, similar to a document class in a FileNet P8.x repository. Properties can be defined for each class, which represent fields that will be included in the index file. These classes and properties will be available in the property mappings section of the file system repository *Create Document* task. In the Add or Edit Document Class window, set the following options to add or edit document classes.

Display name
Document class name displayed to user when selecting a document class in a **Create Document** task.

System name
Document class name stored as the internal system name. This value cannot be edited.

6. In the **Properties** section, you can set the following options to add or edit properties for your document classes.

Display name
Property name displayed to user when selecting a document class in a *Create Document* task.

System name
Property name stored as the internal system name. This value cannot be edited.

Data type
Supported data types are String, String array, Date, Date array, Number, and Number array.

You can set a maximum string length for strings, and a date format for dates, as below.

Date format
The date format should be as follows: for date and time separators, use a backslash (\) to escape the characters. In the example below, the backslash is used to separate date values and the colon (:) is used to separate time values.

```
MM\/dd\/yyy HH\:mm\:s
```

Use the following tokens when specifying the date. Note that tokens are case-sensitive.

```
M = Months as 1 – 12
MM = Months as 01 – 12
MMM = Months as Jan – Dec
MMMM = Months as January – December
d = Days as 1 – 31
dd = Days as 01 – 31
ddd = Days as Mon – Sun
dddd = Days as Monday – Sunday
y = Years as 1, 2, ... , 99
yy = Years as 00 – 99
yyyy = Years as 1900 – 9999
h = Hours as 0 – 12
hh = Hours as 00 – 12
H = Hours as 0 – 23
HH = Hours as 00 – 23
m = Minutes as 0 – 59
mm = Minutes as 00 – 59
s = Seconds as 0 – 59
ss = Seconds as 00 – 59
tt = AM/PM
t = A/P
```

Value required

A value must be provided for the property when the File System Repository *Create Document* task is processed.

7. Save your configuration.

Related reference

"Log levels" on page 770

"CM 8.x Duplicate Detection" on page 509

Configuring an IBM FileNet Image Services Connector and connections

The IBM FileNet Image Services Connector enables IBM Content Collector to store documents in IBM FileNet Image Services (IS) repositories. When IBM Content Collector is installed with the IBM FileNet Image Services Connector option, the Configuration Manager will contain an entry for the IBM FileNet Image Services Connector. The connector is installed with default log settings, which can be configured as you choose.

Required: In addition, you must create one or more connections to the IS repository in order to implement IS tasks in a task route to process files. Each connection contains a set of login credentials to a specific IS repository. When configuring a FileNet IS task, such as IS Create Document, you will be required to choose a configured IS connection. This connection will be used by the IS connector while performing the actions specified by the configured IS task.

Steps to add or edit a connection are provided below as part of the steps to configure a connector.

To configure a connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.

2. In the **Connectors** pane on the upper left, select **FileNet Image Services Repository**. The configuration pane, **General** tab, displays the name and description for the connector.
3. In the configuration pane, edit log settings if required:
   a. From the **Log level** list box, select a log level.
   b. In the **Log file location** text box, edit the location, or click **Browse** to choose a location.
   c. By default, log files will be kept indefinitely. To select a number of days for which to retain the log file:
      - Select **Enable log file** *retention*.
      - Select the number of days to retain.
4. In the configuration pane, click the **Connections** tab. The display pane displays currently configured connections.
5. Do one of the following:
   - To add a connection, click the **Add connection** icon. The Image Services Connection window opens.
   - To edit a connection, select the connection that you want to edit and click the **Edit** icon. Current settings are displayed in the Image Services Connection window.
6. In the Image ServicesConnection window, enter settings as below to enable IBM Content Collector to access content stored in a FileNet ISrepository.

   General

   Display name
   The name used to refer to the Image Services library that you want to store documents in.

   Description
   Repository description.

   FileNet Logon

   User name
   The IS library user name that is used to connect to the library to add documents.

   Password
   The password for the user.

   Test Logon
   Verifies whether the user name and password entered are valid.
7. Save the configuration.

Related reference

"Log levels" on page 770

Configuring the IBM FileNet P8 Connector and connections

When IBM Content Collector is installed with the IBM FileNet P8 repository option, the Configuration Manager contains an entry for the IBM FileNet P8 Connector. The connector is installed with default log settings, which can be configured as you choose.

You must create one or more connections to the FileNet P8 repository in order to implement FileNet P8 tasks in a task route. Each connection contains a set of login credentials to a specific FileNet P8 object store. When configuring a task, such as Create Document or File Document in Folder, you will be required to choose a

configured FileNet P8 connection. This connection will be used by the IBM FileNet P8 Connector while performing the actions specified by the configured task.

Steps to add or edit a connection are provided below as part of the steps to configure a connector.

Important: The name of the FileNet P8 repository connection is used in all stub links and cannot be changed once the connection is consumed.

To configure a connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select IBM FileNet P8. The configuration pane displays a default name and description for the connector.
3. In the configuration pane, edit log settings if required:
   a. From the **Log level** list box, select a log level.
   b. In the **Log file location** text box, edit the location, or click **Browse** to choose a location.
   c. To select a number of days for which to retain the log file:
      - Select **Enable log file retention**.
      - Select the number of days to retain.

      Important: By default, log files will be kept indefinitely.
4. In the display pane, click the **P8 4.x Connections** tab. The display pane displays currently configured connections.
5. In the display pane, do one of the following:
   - To add a connection, click . The configuration pane is cleared of values.
   - To edit a connection, select the connection you want to edit. The configuration pane displays the current settings.
6. Enter a name and a description for the FileNet P8 connection.
7. In the **Content Engine URL** section, from the **Connection** list box, select a connection type (http or https).
8. In the **Server** text box, enter the name of the server on which the repository is installed.
9. In the **Port** field, enter or select a port number. The **Path** text box is filled in automatically.
10. From the **Path** list, select the web services end point.

    Note: For large volumes of repository ingestion, configure your connection to use only the FNCEWS40MTOM endpoint. The FNCEWS40SOAP endpoint is helpful for troubleshooting the web service transport layer and should not be used under normal circumstances.
11. In the **P8 Login Information** section:
    a. In the **User** field, type the ID of a user who can archive and restore documents from the FileNet P8 Content Engine server.
    b. In the **Password** field, type the password of that user.
    c. Click **Retrieve Object Stores** to obtain a list of selectable object stores.
    d. From the **Object Store** list, select the object store to be used with this connection.
12. Save the configuration settings.

Related reference

"Log levels" on page 770

Required IBM FileNet P8 privileges for the connector:

The user account under which the IBM Content Collector IBM FileNet P8 Connector service runs requires a minimum set of permissions for the connector to be able to access the IBM FileNet P8 repository and to perform all required tasks.

The IBM Content Collector IBM FileNet P8 Connector service does not add any security itself. If the user credentials that you specify do not have the minimum set of permissions required, you cannot save the connection. The required permissions are listed here.

Object store security

Permissions
Recommended rights: • Full Control Note that Modify certain system properties is not part of Full Control
Minimum rights: • Connect to store • Create new objects • Modify existing objects • Delete objects • Read permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Modify permissions - Only required when using the P8 4.x Modify Object Security task in a task route
Optional rights: • Modify certain system properties

Class security

The most important classes are:

- Document
- Code module
- Custom object
- Folder
- Subscription
- Event action

Rights
Recommended rights: • Minimum rights

Rights
Minimum rights: • View all properties • Modify all properties • Create instance • Create subclass • Read permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Modify permissions - Only required when using the P8 4.x Modify Object Security task in a task route
Optional rights: • Link • Delete • Modify owner

Document default instance security

Rights
Recommended rights: • Minimum rights
Minimum rights: • View all properties • Modify all properties • View content • Create instance • Minor versioning • Major versioning • Delete • Read permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Modify permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Link/Annotate - Only required when using the P8 4.x Link Documents task in a task route
Optional rights: • Publish • Unlink document • Change state • Reserved12 (deprecated) • Reserved13 (deprecated)

Custom object default instance security

Rights
Recommended rights: • Minimum rights

Rights
Minimum rights: • View all properties • Modify all properties • Create instance • Delete • Read permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Modify permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Link/Annotate - Only required when using the P8 4.x Link Documents task in a task route
Optional rights: • Modify owner • Reserved12 (deprecated) • Reserved13 (deprecated)

Folder default instance security, root folder security and code module folder security

Rights
Recommended rights: • Minimum rights
Minimum rights: • View all properties • Modify all properties • File in folder/Annotate • Create instance - Only needed when using the P8 4.x File Document in Folder task in a task route • Create subfolder - Only needed when using the P8 4.x File Document in Folder task in a task route • Delete - Only needed using the P8 4.x File Document in Folder task in a task route • Read permissions - Only required when using the P8 4.x Modify Object Security task in a task route • Modify permissions - Only required using the P8 4.x Modify Object Security task in a task route
Optional rights: • Unfile from folder • Modify owner • Reserved12 (deprecated) • Reserved13 (deprecated)

Other objects (subscriptions, event actions)

Rights
Recommended rights: • Minimum rights

Rights
Minimum rights: • View all properties • Modify all properties • Create instance
Optional rights: • Link • Create subclass • Modify owner • Delete • Read permissions • Modify permissions

Records Management security

Rights
Recommended rights: • IBM Content Collector FileNet P8 user must be a member of the group configured as a Records Administrator on the Records Manager environment.

Configuring utility connectors

A utility connector supplies additional functionality for use in IBM Content Collector.

You can configure or review the configuration of one or more utility connectors. If a **Required** note is displayed next to a connector in the Connectors window, a configuration beyond the default is necessary: To change the configuration of a connector, you must provide valid credentials first.

Configuring the Metadata Form Connector

The Metadata Form Connector provides a connection to a database where metadata is stored temporarily. The database contains metadata that a user specified additionally when archiving a message manually. This connection enables you to retrieve that information from the database.

If email for which metadata is stored in the database is not processed successfully within 24 hours, the email is reset to its initial state. In Lotus Notes, the icon that indicates that additional archiving information was specified for this document is removed. Thus, the user can specify the metadata once again. This can happen, for example, because the metadata form definition was changed. In this case, the metadata that you try to retrieve does not match the metadata that is stored in the database. Therefore, the email cannot be processed successfully unless you can provide the proper metadata.

If additional archiving information is not retrieved from the database within one month for further use in a task route, the metadata form connector removes the information from the database.

To configure the Metadata Form Connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.

2. In the **Connectors** pane on the upper left, select **Metadata Form**. The configuration pane displays a default name and description for the connector.
3. You can change the name and the description for the connector.
4. In the **Temporary Metadata Database** section, edit settings if required:
   a. In the **Host name** field, specify the host name of the server on which the Derby database is located. The Derby database is the repository that is used for temporarily storing any additional archiving information that a user specified when manually archiving a document. The primary node installation deploys the IBM Content Collector Metadata Form Database service that starts a local Derby server on the IBM Content Collector machine.
   b. From the **Port** list, select the port that remote clients can use for database connections to the Derby database instance. The default port is 1527.
   c. In the **User ID** field, specify the user account with which you want to access the Derby database.
   d. In the **Password** field, type the password for the specified user ID. The user ID and the password that can be used to access the Derby database are defined in the `derby.properties` file in the respective Derby database instance. See the related topic for detailed information about how to adapt the credentials.
5. In the **Log Settings** section, edit settings if required:
   a. From the **Log level** list box, select a log level.
   b. In **Log file location** field, type the full path to the directory in which you want to store the log files, or click **Browse** to choose a location.
   c. To configure multiple log files, select **Truncate log files**. If you do not select this option, the **Number of log files** and **Log file size in MB** are not available and a single log file of unlimited size is written.
   d. In the **Number of log files** field, type the maximum number of log files to be created. There is a dependency between this field and the **Log file size in MB** field. The **Log file size in MB** field specifies the maximum size that a log file can have. As soon as the first log file reaches this size level, a new log file is created. When the maximum number of log files has been reached, and all log files have also reached their maximum size, the oldest log file is overwritten with a new one. This is also known as the round-robin method.
   e. In the **Log file size in MB** field, type the maximum size that a log file is allowed to reach.
6. Under **Logging Type**, select the format in which to store the log files. You can choose between the following options:

   Common Base Event
   Stores the log files in the Common Base Event format, which can be read by a number of tools for log analysis and reporting. Generally, these tools discover dependencies between log events and are capable of creating various reports to visualize these dependencies.

   Plain Text
   Stores the log file in a simple text format. This format requires less disk space than Common Base Event, but usually cannot be processed by analysis and reporting tools as easily.

Related tasks

"Enabling the collection of additional archiving information" on page 431

Configuring the Text Extraction Connector

The Text Extraction Connector provides a single task that converts binary data into its text representation.

The Text Extraction Connector provides an interface to the Oracle/Stellent Outside In Technology filters, which are used in the Extract Text task to convert binary data, for example from email attachments, into a plain-text representation.

To configure the Text Extraction Connector:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select **Text Extraction**.
3. From the **Log Level** list box, select a log level. Events for this connector are logged in a Windows event log with the name IBM Text Extraction Connector.
4. In the **Temporary directory location** text box, edit the location, or click **Browse** to choose a location.

Configuring the Utility Connector

The Utility Connector acts as a container for additional IBM Content Collector functionality.

Configure the settings as required:

1. In the Configuration Manager, click **Connectors** to switch to the Connectors view.
2. In the **Connectors** pane on the upper left, select **Utility**.
3. In the configuration pane, edit log settings if required:
   a. From the **Log Level** list box, select a log level.
   b. In the **Log file location** text box, edit the location, or click **Browse** to choose a location.
   c. To limit the retention of log files:
      1) Select **Enable log file retention**.
      2) Select the number of days to retain log files.

      Attention: By default, log files are retained indefinitely.
4. On the LDAP Connection page, you can select **Use LDAP for lookups** to enable LDAP lookups. In this case, further configuration is required.
5. To change the LDAP query options, select **Override default LDAP query options** on the **LDAP** tab. The fields under the check box, showing the default query options, become editable so that you can change the values:

   User Query
   Searches LDAP entries. By default, the query searches for user account names.

   Direct attribute
   Searches attributes of LDAP entries. By default, the query searches the *MemberOf* attribute of user account names.

   Indirect query
   Searches LDAP groups. By default, the query searches for group members.

Group attribute
Searches attributes of LDAP groups. By default, the query searches the name attribute of the specified group.

6. Select the appropriate type from the **Select LDAP server type** drop-down list. You can select between the following types:

   Active Directory
   For a Microsoft Active Directory LDAP server

   eDirectory
   For an eDirectory LDAP server

   SunOne
   For a SunOne LDAP server

   Domino
   For a Lotus Domino LDAP server
7. In the **Host name** field, type the host name of the LDAP server, for example:
   `myldap.xycorp.com`
8. In the **Port** field, type the proper port number for network communication with the LDAP server. The default value is 389.
9. From the **Protocol version** field, select the LDAP version that your server uses:

   2 For LDAP version 2

   3 For LDAP version 3
10. In the **Base distinguished name** field, type the prefix that identifies the proper set of account names that you want to search during authentication requests. For example, if all LDAP user names used by IBM Content Collector start with CN=Name/O=IBM, then type this string in the **Base distinguished name** field. When the LDAP server needs to authenticate a user, it will search the set of account names starting with this prefix only, which can speed up authentication processes if many user accounts are defined on the LDAP server.
11. In the **User distinguished name** field, type the user account name of the IBM Content Collector default user.
12. In the **Password** field, type the password belonging to the LDAP user account that is specified in the **User distinguished name** field. Click **Test LDAP settings** to verify the credentials.
13. Click **OK**.

Changing the user account of a service

You must change the user accounts of IBM Content Collector services.

To change an account:

1. Determine which services you will require, what accounts to use, and how the services are started.
2. For each account that you change:
   a. Go to **Start** → **Settings** → **Control Panel** → **Administrative Tools** → **Services**.
   b. Right-click the IBM Content Collector service that in the list and select **Properties** from the context menu to change the user account of the service.
   c. Click the **Log On** tab.
   d. Select **This user account**.
   e. In the field next to the radio button, type the name of the logon account.

f. Type and confirm the password for this account.

Related tasks

"Configuring the primary node" on page 88

"Configuring the extension nodes" on page 89

Content Collector services

Some IBM Content Collector services are automatically started after installation and initial configuration whilst others depend on imported task route templates to be started.

The table below shows at what time each service is started. Generally, connector services are started by the IBM Content Collector Task Routing Engine service as soon as their process is required in the workflow of any of the active task routes. Other services are started automatically or must be started manually, as detailed below.

You can change the user account of any of the available Content Collector services. If you change the accounts of a service that must be started manually, you must restart this service so that the changes can become effective. If the service is started by the IBM Content Collector Task Routing Engine service, click the Signal Change tool bar button or select **File → Signal Change**.**Signal** in the Configuration Manager to restart the IBM Content Collector Task Routing Engine service.

The following table lists the available Content Collector services with information on when the services are started.

Table 45. IBM Content Collector services

Content Collector service	Running status	User ID for the service
IBM Content Collector Task Routing Engine service	Must be started manually by the user after the IBM Content Collector Server installation and configuration	The user ID must have permission to start and stop the connector services monitored by the IBM Content Collector Task Routing Engine service: • Lotus Domino: Not the local system account. It can be the administrator account but may be another account as well. • Microsoft Exchange: Microsoft Exchange: An account with Exchange Enterprise Administrator rights. Exchange Server 2007 only: You can archive email from a mailbox with Full Mailbox Access right only, but public folder archiving and collecting from a server collection source will not be permitted. • SMTP: User account with full access rights to the SMTP message queue directory. • Microsoft SharePoint: User with full access rights to the temporary file location of the SharePoint Connector. • File System: Local user account with full access rights to all file locations. **Important:** In a scale-out environment, the IBM Content Collector Task Routing Engine service user on each node must have the permission to engage in pipe communication with the process on the other machines. The user right for this is Access this computer from the network.

Table 45. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector Web Application service	Must be started manually by the user after the IBM Content Collector Server installation and configuration	• Lotus Domino: Not the local system account. It can be the administrator account but may be another account as well. It is recommended that you use the same user account for the IBM Content Collector Email Connector service. • Microsoft Exchange: An account with Exchange Enterprise Administrator rights. It is recommended that you use the same user account for the IBM Content Collector Email Connector service. • SMTP: Local user account. In a scale-out environment, the user must be a domain user. • Microsoft SharePoint: User with full access rights to the temporary file location of the SharePoint Connector. • File System: Local user account with full access rights to all file locations. The IBM Content Collector File System Source Connector service account does not have to be the same as the IBM Content Collector Web Application service account as this service does not interact with the file system location that a document is captured from.
IBM Content Collector Utility Connector service	Started and monitored by the IBM Content Collector Task Routing Engine service	Not applicable

Table 45. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector Email Connector service	If Lotus Domino or Microsoft Exchange is installed and configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	• Lotus Domino: Not the local system account. It can be the administrator account but may be another account as well. • Microsoft Exchange: An account with Exchange Enterprise Administrator rights. • It is recommended that you use the same user account for the IBM Content Collector Email Connector service and the IBM Content Collector Web Application service
IBM Content Collector SMTP Connector service	If SMTP is installed and configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	User account with full access rights to the SMTP message queue directory. In a scale-out environment, the user must be a domain user.
IBM Content Collector SMTP Receiver service	This service runs outside of the scope of the IBM Content Collector Task Routing Engine service and must be started manually.	User account with full access rights to the SMTP message queue directory. In a scale-out environment, the user must be a domain user.

Table 45. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector File System Source Connector service	Started if File System is configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	• If the File System Source Connector monitors files on other machines in the same domain, the user must have full access rights to the other machines. • The user of the IBM Content Collector Content Manager Connector service and the IBM Content Collector IBM FileNet P8 Connector service must have access to the document's content file, so that it can be archived into the repository. The simplest method is to run the file system source and the target repository connectors under the same service account. • The File System Source Connector must run as a user that has permissions to access metadata files, or there must be a trust relationship between the IBM Content Collector system and the system where the files are located.
IBM Content Collector Metadata Form Connector service	Started after the Metadata Form Connector is configured by the active task route configured to use the connector	Not applicable
IBM Content Collector Metadata Form Database service	Started automatically after the IBM Content Collector Server installation and starts the temporary metadata database (Derby database)	Not applicable

Table 45. IBM Content Collector services (continued)

Content Collector service	Running status	User ID for the service
IBM Content Collector SharePoint Connector service	If Microsoft SharePoint is installed and configured as source system, this service is started and monitored by the IBM Content Collector Task Routing Engine service	• The IBM Content Collector SharePoint Connector service user must have full access rights to the temporary file location of the SharePoint Connector. • The user of the IBM Content Collector Content Manager Connector service and the IBM Content Collector IBM FileNet P8 Connector service must have access to the document's content file, so that it can be archived into the repository.
IBM Content Collector Content Manager Connector service	If Content Manager is installed and configured as target repository, this service is started and monitored by the IBM Content Collector Task Routing Engine service	For the required account privileges, see "Required Content Manager privileges for the connector" on page 201.
IBM Content Collector IBM FileNet P8 Connector service	If IBM FileNet P8 is installed and configured as target repository, this service is started and monitored by the IBM Content Collector Task Routing Engine service	For the required account privileges, see "Required IBM FileNet P8 privileges for the connector" on page 207.
IBM Content Collector Text Extraction Connector service	Started together with the IBM FileNet P8 connector service	Not applicable
IBM Content Collector File System Repository Connector service	After the connector is configured to be used in a task route, this service is started and monitored by the IBM Content Collector Task Routing Engine service	If the repositories of the File System Repository Connector are located on machines in the same domain but not on the IBM Content Collector source machine, the user must have full access rights to the other machines.
IBM Content Collector FileNet Image Services Connector service	After the connector is configured to be used in a task route, this service is started and monitored by the IBM Content Collector Task Routing Engine service	Local user account with full access rights.
IBM Content Collector Information Help Center service	Started automatically after the IBM Content Collector Server installation when the IBM Content Collector Configuration Manager is started	Not applicable

Configuring general settings

You can configure the settings of several special services for archiving, searching, viewing, and restoring email, for example, if you want to make use of Outlook Web Access or if you want to restore email that was archived by CommonStore for Exchange. You can also configure the collection of additional archiving information.

Depending on your installation, not all configuration settings might be available in the Configuration Manager.

Related tasks

"Installing and configuring Content Collector Outlook Web App (formerly Outlook Web Access) support" on page 111

Configuring Content Collector for CommonStore for Exchange Server legacy support

IBM Content Collector requires the configuration data related to each CommonStore for Exchange Server (CSX) task instance to restore documents archived by CSX.

After you have installed IBM Content Collector legacy support and before you begin configuring the legacy support, carry out the following:

1. Ensure that the following files were removed from the `bin` subdirectory on the server on which legacy support was installed, and if not, remove them:
   - `csx.exe` and `csxadmin.exe`
   - `csx_mapi.dll`
   - `csx.jar`
2. Open the configuration `archint.ini` file and delete the line that starts the CommonStore search server. The line contains the keyword START_SEARCHSERVER.

In CSX, all configuration data is stored in the Active Directory. In IBM Content Collector all configuration data is stored in a database that is accessed using the IBM Content Collector Configuration Manager.

To extract the configuration data for each CSX task from the Active Directory (or Active Directory Application Mode (ADAM)) and to add this data to the IBM Content Collector configuration:

1. Run `afucsx_exportTasks.exe` in a Command Prompt window on the server on which IBM Content Collector legacy support was installed. If you use an alternate AD, run `afucsx_exportTasks.exe -s ADAM_server -t ADAM_port`.

   The task description XML files and the log files are written to `<IBMAFUROOT>\Task\Export` on the IBM Content Collector server where IBMAFUROOT is an environment variable set for IBM Content Collector.

   If IBMAFUROOT has not been set, the task description and log files are written to `<CurrentDirectory>\Task\Export`.
2. Start the IBM Content Collector Configuration Manager.
3. Click **General Settings** → **Legacy Restore Exchange**.
4. Click .
5. On the General page of the tabbed notebook, enter a name and description for your configuration object.

6. Under **Import legacy task definition**
   - Click **Browse** to navigate to and select the XML configuration file of the CSX Task. This is usually under `<IBMAFUROOT>\Task\Export`.
   - Click **Import** to update your legacy task with the data obtained through running `afucsx_exportTasks.exe`.
   - Save the legacy configuration.

     After the import, all fields in the tabbed notebook of the Legacy Restore Exchange pane are filled in automatically.

     The settings include the required CSX task parameters and all CommonStore Server and Exchange Server parameters.

     If you have not made any changes to your CSX environment, you need not adjust these settings.

     However, if you have moved or renamed any CSX servers, or want to change any trace, error, or log file paths, adjust these fields in the tabbed notebook of the Legacy Restore Exchange pane.

     Important: Check the transfer path settings. If the CommonStore Server and your CSX Task instance run on the same machine, the transfer paths in the Configuration Manager and in the server configuration profile (usually archint.ini, keyword TRANSFERPATH) MUST be identical.

     If both components run on different machines, the machines must share a common network directory. The transfer path specifications for both components must then point to the same physical directory.

     For a detailed description of these fields refer to the link under Related references at the bottom of the page.

Related reference

"Settings to enable restoring documents archived using CommonStore for Exchange Server" on page 143

Settings to enable restoring documents archived using CommonStore for Exchange Server

To restore documents archived using CommonStore for Exchange Server (CSX), you need the configuration data of your CSX Task. The configuration data includes settings pertaining to the CSX Task, the CommonStore Server and the Exchange Server.

If you have not made any changes to your CSX environment, then you do not need to adjust these configuration settings after you have imported the configuration data into the Configuration Manager. However, if you have moved or renamed any CSX servers, or want to change any trace, error, or log file paths, adjust these fields in the tabbed notebook of the Legacy Restore Exchange pane in the Configuration Manager.

These configuration settings are also described in detail in the IBM CommonStore for Exchange Server documentation.

CommonStore for Exchange Server Task settings on the General page

Task name

The name of the CSX Task instance that you want to address. Spell the name exactly as it is spelled in the CSX System Manager.

Description
A description of the task instance. This field is optional. Use it, for example, to add information about the repository.

Exchange user
The names of Exchange users to notify in case of events that require the intervention of an administrator. The selected users receive an email when an error has occurred, thus allowing them to take appropriate action. Type the email addresses of users in the **Exchange user** field and click **Add**. Selected users appear in the list box at the bottom of the property page. To delete a user, highlight the appropriate entry in the list box and click **Remove**. To remove all users, click **Delete All**.

CommonStore for Exchange Server Task settings on the Parameters page

Worker count
The number of worker threads that the CSX Task starts. The value must be an integer from 1 to 9. The worker is the subcomponent of the CSX Task that sends requests to the CommonStore Server. The requests that are sent from the worker to the CommonStore Server (archpro) are processed by agents. It is therefore recommended to configure the archpro program to start as many agents as workers.

Committer count
The number of committer threads that the CSX Task starts. The value must be an integer from 1 to 9. The committer is the subcomponent of the CSX Task responsible for changing the job-processing state in the task-job queue after a job was processed on the CommonStore Server. It also copies restored data back to a message store after a job was completed successfully on the CommonStore Server. Depending on your workload, you can improve the performance of the CSX Task by defining several committer threads to run in parallel. However, make sure that your hardware is capable of running this number of threads.

External port
The number of the port that you use to interact with a particular CSX Task instance. For example, you communicate over this port when you shut down a particular task instance. Specify an integer from 7000 to 9999. If you run more than one CSX Task instance on the same machine, specify a different port for each instance. Otherwise, conflicts arise.

Trace file name
The path and the name of the trace file. The trace file contains event messages issued by the CSX Task or the CommonStore Server, such as the startup of the CSX Task or the successful connection with the CommonStore Server.

Error file name
The path and the name of the error log file. The error log file contains all Exchange-related error messages.

Log directory
The directory to which the log files are written. To specify this directory, enter the full path. Log files are created daily. Their names are predefined, including date and time information. The file extension is log.

Trace Allows you to enable or disable tracing for operations related to the CSX

Task (tracing related to the CommonStore Server must be enabled in the server configuration profile). To switch tracing on, select this box. To switch it off, clear it.

CommonStore Server settings on the CommonStore Server page

Host name

The full server name of the computer on which the CommonStore Server runs. If you only specify the machine name or IP address, and link to a user outside of your local intranet, the machine name cannot be resolved. If, on the other hand, this field contains a full server name, the Domain Name System (DNS) will resolve the server name in a link to the new IP address.

Important: Never use `localhost` as the host name.

Fixed port

The number of the port that the CSX Task uses to connect to the CommonStore Server. This port must be the same as the port that you specify in the server configuration profile (usually archint.ini) by using the ARCHPRO_PORT keyword. Specify an integer from 5000 to 9999. The default port is 8013.

Transfer path

The path to the directory that the CSX Task and the CommonStore Server use to exchange content. When you restore a document, the CommonStore Server copies it to the transfer directory, and the CSX Task picks it up from there. To specify this directory, enter the full path from the root directory.

Important: Check the transfer path settings. If the CommonStore Server and your CSX Task instance run on the same machine, the transfer paths in the Configuration Manager and in the server configuration profile (usually archint.ini, keyword TRANSFERPATH) MUST be identical.

If both components run on different machines, the machines must share a common network directory. The transfer path specifications for both components must then point to the same physical directory. To make the CommonStore Server and the CSX Task share a common network directory, use standard Windows file sharing, SAMBA, or any other Server Message Block protocol.

Exchange Server settings on the Exchange Server page

Job folder name

The name of the public folder on the Exchange server in which job documents are stored. Job documents basically identify the email that is to be restored.

Host name

Allows you to specify the Exchange servers that you want the CSX Task instance to work on. Type the fully qualified name of an Exchange server in the field, and click **Add**. The server names appear in the list box at the bottom of the page. To remove a server from the list, highlight the appropriate entry in the list box and click **Remove**. To remove all servers from the list, click **Remove All**.

Notes: A single Exchange server can only be served by one CSX Task instance. A single CSX Task instance, on the contrary, can serve multiple Exchange servers.

Modifying the Configuration Web Service settings

Perform the steps in this section to modify the details of the Configuration Web Service.

1. You can change the name of the Configuration Web Service and its description by modifying the information in the **Name** and **Description** fields.
2. In the **Host name** field, type the host name of the server, on which the Configuration Web Service is running.
3. In the **Port** field, type the port number that is used for communication between the Configuration Web Service and the IBM Content Collector server.
4. Select **Use embedded Web Application Server** to have the Configuration Web Service configured and updated automatically when you save any changes to the configuration. This includes updates due to changes of the information in the active data store.
5. In the **JDBC driver directory** field, specify the directory in which the JDBC drivers of your database reside, for example:

   `C:\Program Files\IBM\SQLLIB\java` for DB2

   `C:\Program Files\Microsoft SQL Server 2005 JDBC Driver` for SQL Server 2005

   `C:\Program Files\Microsoft SQL Server 2008 JDBC Driver` for SQL Server 2008

   `C:\Program Files\Oracle` for Oracle

   Important:

   - If your configuration data store is an SQL Server database, you must specify the location of the `sqljdbc4.jar` file if you use the embedded Web Application Server or IBM WebSphere Application Server Version 7. If you use IBM WebSphere Application Server Version 6.1, specify the location of the `sqljdbc.jar` file.
   - If your configuration data store is an Oracle 11g database, you must update the JDBC path information in the `resources.xml` file of the Web Application Server (WAS). This file is located in one of the following directories:

     For the embedded Web Application Server:

     `<installdir>\ContentCollector\AFUWeb\profiles\AFUWeb\config\cells\<cellName>\nodes\<nodeName>\`

     For an external Web Application Server

     `<installdir>\AppServer\profiles\AFUWeb\config\cells\<cellName>\nodes\<nodeName>\`

     Replace <installdir>, <cellName>, and <nodeName> with the proper values of your WAS installation.

     The default path entry in the `resources.xml` file points to the JDBC resource file `ojdbc14_g.jar`, which does not work with Oracle 11g. Change the path information so that it points to one of the following files of your Oracle 11g installation:

     For WebSphere Application Server version 6:

     `ojdbc5.jar` or `ojdbc5_g.jar`

For the embedded Web Application Server and WebSphere Application Server version 7:

`ojdbc6.jar` or `ojdbc6_g.jar`

`ojdbc5_g.jar` and `ojdbc6_g.jar` support the tracing of database operations.

For more information, go to the WebSphere Application Server (Distributed platforms and Windows) Information Center and search for *data source minimum required settings.*

6. In the **JDBC port** field, specify the port number to be used to connect to the configuration database.
7. In the **Database server host name** field, specify the fully qualified host name of the computer on which the database is located. For example, enter:

   `server.company.com`

Related information

WebSphere Application Server Information Center

Modifying the information center settings

Perform the steps in this section to modify the configuration details of the IBM Content Collector Information Center.

1. You can change the name of the information center and its description by modifying the information in the **Name** and **Description** fields.
2. In the **Host name** field, type the fully qualified host name of the server, on which the information center is installed. For example, enter:

   `server.company.com`
3. In the **Port** field, type the port number that is used for communication with the information center.

Modifying the settings for the Web Application

You can modify the general configuration details of the Web Application and define the log settings.

To modify Web Application settings:

1. On the General page, you can change the name of the Web Application and its description by updating the information in the **Name** and **Description** fields.
2. In the **Host name** field, type the host name of the server on which the Web Application is installed.

   Important: If you use Lotus Domino as source system and you change the location of the Web Application, you must modify the Lotus Notes mail template to reflect this change.
3. In the **Port** field, type the port number that is used for communication with the Web Application. The communication protocol is HTTPS.
4. In the **Repository connection** section, select one or more connections to your content management system. This can be either connections to IBM Content Manager repositories or connections to IBM FileNet P8 object stores. You must also determine the default connection.

   One connection must be determined as the default connection. You can add, edit, or remove connections.

   All archived documents, be they email, Microsoft SharePoint, or File System documents contain information about the repository, so that IBM Content Collector can identify the repository when a user wants to view or restore an archived document.

5. On the **Log Settings** page, edit settings if required:
   a. Select the **Truncate log files** option to configure multiple log files of limited size. Otherwise, a single log file of unlimited size is maintained.
   b. From the **Log level** list box, select a log level.
   c. In **Log file location** type the full path to the directory in which you want to store the log files, or click **Browse** to choose a location.
   d. In the **Number of log files** field, type the maximum number of log files to be created. There is a dependency between this field and the **Log file size in MB** field. The **Log file size in MB** field specifies the maximum size that a log file can have. As soon as the first log file reaches this size level, a new log file is created. When the maximum number of log files has been reached, and all log files have also reached their maximum size, the oldest log file is overwritten with a new one. This is also known as the round-robin method.
   e. In the **Log file size in MB** field, type the maximum size that a log file is allowed to reach.
6. Under **Logging Type**, select the format in which to store the log files. You can choose between the following options:

   Common Base Event
   Stores the log files in the Common Base Event format, which can be read by a number of tools for log analysis and reporting. Generally, these tools discover dependencies between log events and are capable of creating various reports to visualize these dependencies.

   Plain text (internal format)
   Stores the log file in a simple text format that is specific to IBM Content Collector. This format requires less disk space than Common Base Event, but usually cannot be processed by analysis and reporting tools as easily.

   Plain text (WebSphere format)
   Stores the log file in a simple text format. The layout of the log file matches the layout of WebSphere log files. This format requires less disk space than Common Base Event, but usually cannot be processed by analysis and reporting tools as easily.
7. Restart the IBM Content Collector Web Application service to enable the changes to take effect.

Changing the location of the Web Application in the Lotus Notes mail template

If you use Lotus Domino as source system and you change the location of the Web Application, you must modify the Lotus Notes mail template to reflect this change.

To modify the location of the Web Application in the Lotus Notes mail template that is enabled for Content Collector:

1. Open the mail template in the Lotus Notes Designer.
2. Select **Shared Code → Agents**.
3. Right-click **Change IBM Content Collector Web Application Location** and select **Design properties**.
4. On the Design tab, clear everything that follows **Hide design elements from**.
5. Close the window.
6. Open the mail template in Lotus Notes.
7. Select **Actions → Change IBM Content Collector Web Application Location**.

8. Enter the new location of the Web Application.
9. Open the mail template in the Lotus Notes Designer again.
10. Select **Shared Code** → **Agents**.
11. Right-click **Change IBM Content Collector Web Application Location** and select **Design properties**.
12. On the Design tab, select everything that follows **Hide design elements from**.
13. Close the window.

Modifying client configuration settings

Perform the steps in this section to modify the details of the client configuration.

1. You can change the name of the client configuration and its description by modifying the information in the **Name** and **Description** fields.
2. In the **Trigger mailbox** field, type the name of the mailbox that is monitored for interactive archiving requests. This is the mailbox that collects the manual archiving and stubbing requests, which were initiated by clients. These requests instruct an interactive collector to process documents in the mailboxes that the requests came from. Type the mailbox name in one of the following formats, depending on your email system.

   Lotus Notes/Domino
   Use the canonical name, for example:
   `CN=ICCJOBS/O=COMPANY`

   Microsoft Outlook/Exchange
   Use the SMTP address of the mailbox, for example:
   `iccjobs@company.com`
3. Define how a stub document is displayed when the user opens it. To have the content of a stubbed document temporarily retrieved from the repository and displayed when the user opens the document, select **Retrieve and display document when opened**. The document is not permanently restored, but only available temporarily. Depending on the number of clients in your topology, enabling this option can have a negative impact on the Content Collector Server performance. This is because requests for retrieving and displaying documents are handled by the Content CollectorWeb Application and, therefore, put load on the Web Application Server.

   If you do not select this option, the document's stubbed content and the stub links are displayed when the user opens the document.
4. If the collection of additional archiving information is enabled, define the required settings for the client in the **Additional Archiving Information** section.
   a. Specify the folders for which additional archiving information must be specified when an archive request is initiated by the client. Note that you can select top-level folders only. The setting is applied to all subfolders.
   b. Specify the metadata form template and the metadata form definition that are to be used for collecting the information for the selected folder.
   c. To define if users can specify additional archiving information for documents in this folder only if a connection to the network exists, select **Only use mapping when client is online**. This setting is necessary, for example, if the metadata form relies on other web services for providing information. These web services are available only if a connection to the network exists. If you do not select this option, users can always specify additional archiving information.

5. As a Lotus Domino user, you can also enable tracing for iNotes (formerly Domino Web Access).
   a. Select **Enable tracing** to switch tracing on. Clear the check box to switch it off.
   b. In the **Trace file location** field, type the full path of the trace file location. You can also leave this field empty, in which case a default path is used. The default trace file path is *temporary directory*\IBM\ContentCollector_iNotes where *temporary directory* is the directory that is specified by the environment variables TMP or TEMP.

   For a Notes client, you enable tracing in the client's IBM Content Collector menu. The default trace file location is %USERPROFILE%\IBM\ContentCollector for Microsoft Windows users, or %USERHOME%/IBM/ContentCollector for Mac users. You can change the location by setting the environment variable IBMAFUTRACELOCATION.

Related tasks

"Enabling the collection of additional archiving information" on page 431

"Collecting email on request" on page 472

Configuring the access to archived data

Adapt the settings in the configuration database that are required for the access to archived data. For email, you enable searching, viewing, and restoring archived documents. For documents other than email you configure IBM Content Collector to enable retrieval of archived data.

The settings are based on the output of the Initial Configuration wizard. When the Initial Configuration wizard configures the repository, it creates new item types or document classes and automatically creates default settings for the access to archived data. For new installations, these settings are automatically imported into the configuration database. However, this import operation happens only once for each type of archived data access. This means that if you, for example, created multiple item types, the output of the setup tool is not automatically added to the configuration database. In this case, you must merge the contents of the additional configuration files with the existing definitions. The additional archive mapping and search configuration files are located in the \cm or \p8 subdirectory of the directory *InstallDir*\Configuration\initialConfig\data\search\output, depending on your repository, where *InstallDir* is the installation directory of IBM Content Collector.

If you are upgrading from file system archiving in Content Collector Version 2.1.0 and want to have the archiving task create shortcut links to archived documents, you must ensure that the configuration files exist. If you want to use a new item type or document class, use the Initial Configuration wizard to create it. When you start the Configuration Manager, an archived data access configuration is created and the required configuration files are imported automatically. Note that this happens only once. If you archive into an item type or document class that has been created with Content Collector Version 2.1.0, copy the required templates from *InstallDir*\Configuration\initialConfig\data\search, adapt them according to your needs, and import them. For more information about this see the related topic.

To modify the configuration for the access to archived data:

1. Select the type of access that you want to configure. Depending on your selections during the installation of IBM Content Collector, one or more configurations are available:

   Archived Data Access for Email
   This configuration is used for searching, viewing, and restoring email, and for retrieving documents archived from Notes applications.

   Archived Data Access for FileSystem
   This configuration enables retrieval of documents other than email. Note that you must also configure a shortcut link in the respective archiving task (CM 8.x Create Document, P8 4.x Create Document).

   Archived Data Access for SharePoint
   This configuration enables retrieval of Microsoft SharePoint documents. Note that you must also configure a shortcut link in one or more of these tasks:
   - CM 8.x Confirm Document
   - CM 8.x Create Document
   - CM 8.x Store Version Series
   - P8 4.x Confirm Document
   - P8 4.x Create Document
   - P8 4.x Create Version Series
2. Adapt the settings on the configuration pages on the right.
3. Save all changes.

Your changes will take effect as soon as the IBM Content Collector Web Application service is restarted.

Related concepts

"Enabling the access to archived data" on page 582

Related reference

"CM 8.x Confirm Document" on page 504
"CM 8.x Create Document" on page 506
"CM 8.x Store Version Series" on page 511
"P8 4.x Confirm Document" on page 540
"P8 4.x Create Document" on page 543
"P8 4.x Create Version Series" on page 550

Adapting collection definitions

All collections and items are listed that are currently defined in the configuration database. You can add items to existing collections or add new collections. You can also delete definitions from the configuration database.

- Add, edit, or remove item types or document classes.
  - To add an item type or a document class to an existing collection, select one of the defined collections and, under the list of defined items, click **Add**.
  - To edit an item type or a document class, select one of the defined collections. Then, select the item type or document class and click **Edit**.
  - To remove an item type or a document class from a collection, select one of the defined collections. Then, select the item type or document class and click **Remove**. Note that this is possible only if more than one item type or document class exist in the selected collection.

  To add or edit an item:

1. Select a repository connection.
2. In the **Name** field, enter or edit the name.
3. Enter or edit the names of the child components that define the subcollections for the instance-specific information of email documents or attachments.
4. If you want the item type or document class to be used for a specific period of time, select a start date and an end date.
5. Click **OK** to save your entries, click **Cancel** to leave the window without saving.

- Add or remove collections, or map collection fields to content server properties and text index fields. The **Defined collections** list contains all collections that are defined in the archive mapping. When you select a collection, the item types (IBM Content Manager) or document classes (FileNet P8) that are defined in the selected collection are listed. For all types of archived data access configuration, you can add, edit, or remove collections, and you can map collection fields to content server properties and text index fields. What data you must specify, depends on the type of configuration.

Option	Description
Add a collection for archived email	When you add a collection to the archive mappings, you must specify this data: **On the Base collection page** Specify a collection name and, if applicable, a collection type. In the **Document Type** section, select a repository connection. Enter the name of an email item type or document class that is defined in IBM Content Manager or FileNet P8. Depending on the selected collection type, you must enter the names of one or more components that will contain instance information. **On the Subcollections page** Set up the collection definitions for instance information. **Attachment instance collection** Select the attributes that contain the attachment file name, the correlation key for attachments, and the reference information to the item type that holds the attachment content. **Email instance collection** Select the attribute that contains the mailbox ID. **On the Collection fields page** Map collection fields to content server properties or text index fields.

Option	Description
Add other collections	When you add a collection to the archive mappings, you must specify this data: **On the Base collection page** Specify a collection name and a repository connection. **On the Collection fields page** Map collection fields to content server properties or text index fields.
Remove a collection for archived email	Select a collection to delete from the archive mappings. Remember to adapt the layout of the Email search page accordingly by modifying the search configuration file.
Remove other collections	Select a collection to delete from the archive mappings.
Edit fields in a collection for archived email	Set up collection fields and map these fields to content server properties (for attribute search) and to text index fields (for full-text search). These are the fields for which users can search. To address fields in other collections as required by the data model, define reference fields: 1. Add a collection field of the type reference and, from **Referenced collection**, select the collection that contains the fields that you want to address. You can define only one field of the type reference for each target collection. 2. Add other collection fields to address the specific fields in the referenced collection. To be able to retrieve archived documents, a basic set of fields must be set up and mapped to content server properties. Reserved keywords are listed in the topic about defining and exposing collections. You can work only with properties or text index fields that are defined in the configuration database. Existing entries are listed on the Properties page or the Text Index page, respectively.
Edit fields in other collections	Define collection fields and map these fields to content server properties or text index fields. Reserved keywords are listed in the topic about defining and exposing collections. To be able to retrieve documents that were archived from file system or Microsoft SharePoint sources, at least the field FILENAME must exist and must be mapped to the respective content server property.

Related tasks

"Defining and exposing collections" on page 681

Adding content server properties to the archived data access configuration

All content server properties are listed that are currently defined in the configuration database and that can be used in the archive mapping. Whenever an archive mapping file is imported, the Configuration Manager checks the content server property definitions and adds any new properties to the configuration database. You can also add properties or delete properties from the configuration database.

On the Properties page, add, edit, or remove content server properties. Note that for FileNet P8 you must use the symbolic property names and not the display names. Properties that are used in collection fields cannot be removed. However, you can edit those properties and the references are updated automatically. When you map search fields to content server properties, you can use only those fields that are defined here.

1. In the Add Property window, enter the name of a content server property.
2. Select the data type of the property.
3. Select the component for which the property is defined. You can select the root component (**Base**), the child component for the email instance (**Email instance**), or the child component for the attachment instance (**Attachment instance**).

Adding text index fields to the archived data access configuration

All text index fields are listed that are currently defined in the configuration database and that can be used in the archive mapping for email. Whenever an archive mapping file is imported, the Configuration Manager checks the text index field definitions and adds any new fields to the configuration database. You can also add fields to or delete text index fields from the configuration database

On the Text Index page, add, edit, or remove text index fields. You can add any field that is defined in the text indexer model file (for IBM Content Manager) or the XIT (for FileNet P8). When you map search fields to text index fields, you can use only those fields that are defined here. However, field names that you add are not validated so that misspelled field names will result in search errors. Text index fields that are used in collection fields cannot be removed. However, you can edit those text fields and the references are updated automatically.

Importing or exporting the archived data access configuration

Export configuration data to a file on disk to be able to adapt the data manually or to an IBM eDiscovery Manager repository for use with IBM eDiscovery Manager. Import configuration data for the access to archived data from a file on disk to the Content Collector configuration database after you changed the data manually.

To import and export the configuration files:

- Import or export the search configuration. To modify the layout of the Email search page:
  1. Export the search configuration data from the configuration database to a file.
  2. Edit the search configuration file to modify the layout as required.
  3. Import the search configuration data from the file into the configuration database.

- Import or export the archive mapping. To modify the archive mapping manually:
  1. Export the archive mapping data from the configuration database to a file.
  2. Edit the archive mapping file to modify the layout as required.
  3. Import the archive mapping data from the file into the configuration database. If the archive mapping contains repository information that cannot be resolved, the Resolve Dependencies window opens. Select an appropriate repository connection. If no repository connection is listed, you must define a repository connection in the Web Application configuration. Then, import the archive mapping again.
- To export the archive mappings to an IBM eDiscovery Manager repository, select the connection to the primary IBM eDiscovery Manager repository and click **Export**. The archive mappings are added to the IBM eDiscovery Manager repository. Existing mappings in the IBM eDiscovery Manager repository are not changed or overwritten. Any repository that is referenced in the archive mapping and that is not an IBM eDiscovery Manager repository, will be added as a secondary repository in IBM eDiscovery Manager.

Modifying the settings for the Metadata Web Application

Perform the steps in this section to modify the configuration details of the Metadata Web Application.

1. You can change the name of the Metadata Web Application and its description by updating the information in the **Name** and **Description** fields.
2. In the **Host name** field, type the host name of the server on which the Derby database is located. The Derby database is the repository that is used for temporarily storing any additional archiving information that a user specified when manually archiving a document.
3. In the **Port** field, type the port number that remote clients can use for database connections to the Derby database instance.
4. In the **User ID** field, specify the user account with which to access to the Derby database.
5. In the **Password** field, specify the password for the specified user ID. Validate the password. The user ID and the password that can be used to access the Derby database are defined in the `derby.properties` file in the respective Derby database instance. If you use the Derby database that is embedded in Content Collector, the `derby.properties` file is located in the directory `<ICCINstallPath>\derby\10.3.3.0\bin`, where `<ICCINstallPath>` is the IBM Content Collector installation directory.
   a. To add or change entries for users who are to have access to the Derby database, edit the properties file.
   b. Change the line that follows the comment `# Users definition` or add further entries. The syntax is `derby.user.<username>=<password>`. Replace <*username*> and <*password*> with proper values.
   c. Restart the IBM Content Collector Metadata Form Database service.

Related tasks

"Selecting the metadata form template"

"Configuring the metadata form definition" on page 235

Selecting the metadata form template

The metadata form template contains the code logic for the HTML form that is used for collecting additional archiving information when a user manually archives email.

To make a metadata form template available for use with IBM Content Collector, it must be stored in the configuration database. In the client configuration, the metadata form template and a metadata form definition are mapped to a monitored folder to enable the collection of additional archiving information.

To select a template to use:

1. Click **Import** to import a metadata form template into the configuration database. You are asked to select the metadata form template to import.
2. Select the template and click **OK**. The selected template is imported to the configuration database.
3. To view the content of the template, click **Preview**.
4. Save this configuration setting. Note that only one instance of a form template can exist in the configuration database. When you import the form template, a hash key for the form template is calculated. If you try to save a second instance of the same metadata form template, a message is displayed that states that this form template already exists in the configuration database.

A default metadata form template is provided with IBM Content Collector. The .zip file is located in *InstallDir*\formTemplates, where *InstallDir* is the Content Collector installation directory. If you are familiar with HTML, JavaScript programming, and the Dojo toolkit, you can adapt the metadata form template to your needs or set up your own form template. You can, for example, change the content and the design of the form template, use different web services, or add new interface controls and features.

To learn more about the Dojo toolkit, go to the Dojo Toolkit website.

To make changes to the metadata form template, you have to export it from the configuration database to a directory of your choice. Change the template as required. If you create a new form template, note that your .zip file must contain a `form.html` file.

Restriction: Due to a limitation in the database scheme for DB2, the size of the .zip file must not exceed 1 MB. It is recommended that you compress the file, for example, by using the JavaScript compression tool. The imported file must be a .zip file.

For your changes to take effect, you have to import the modified template into the configuration database and to restart the clients.

Now, you have to configure the metadata form definition to specify which information is to be collected.

Related tasks

"Configuring the metadata form definition"

Configuring the metadata form definition

Define the layout of the form for collecting additional archiving information when a user manually archives email.

Make sure that a metadata form template exists in the configuration database and that at least one user-defined metadata source is defined.

To associate additional archiving information with email, you can prompt a user for specific information and map this information to a set of metadata that you defined previously. Configure the metadata form definition accordingly. You can configure more than one metadata form definition.

Note that only one instance of a form template can exist in the configuration database. When you import the form template, a hash key for the form template is calculated. If you try to save a second instance of the same metadata form template, a message is displayed that states that this form template already exists in the configuration database.

1. In the Configuration Manager, click **General Settings** to switch to the General Settings view.
2. In the **General Settings** pane on the upper left, select **Metadata Form Definition**.
3. In the design pane, which displays any previously created metadata form definition, do one of the following:
   - If you want to add a new metadata form definition, click .
   - If you want to edit a set of metadata form definition, select the item you want to edit.
4. The configuration pane contains the following configuration options. For an existing metadata form definition, the fields contain values, which you can change. For a new metadata form definition, enter the required values.

   Name Specify a name for the current metadata form definition.

   Description
   The text that you enter here is displayed at the top of the input form. For example, you might want to provide users with instructions or additional information for completing the form.

   User-defined metadata
   Select the user-defined metadata source on which the form is based. The metadata source contains one or more metadata properties that can be used in the metadata from definition.

   Layout
   The table lists the properties that are defined in the selected user-defined metadata source. The table columns are:

   Used in form
   States if the property is used in the metadata form definition. The value can be either true or false.

   Property ID
   Uniquely identifies the property of this metadata source to the system.

Display name
Is the name of the property as it is displayed to the user.

Data type
Is the data type of the property:

- Boolean
- Byte
- Date Time
- Float
- Integer
- String

Representation type
States the representation type that was selected for this property.

Required
States if this field is mandatory.

To define if a property is used in the input form, how it is presented in the form, and what kind of information a user is to specify, select a property and click **Edit**.

a. In the **Edit Form Template Definition Layout** window, define the settings for the selected property:
   - To include the field in the metadata form template, select **Use this metadata property in the form**.
   - To make the field mandatory, select **Mark field as required**.
   - Select the representation type for the information that the user is to provide. Depending on the data type of the property, one or more of the following representation types are available:
     - **Text Entry** allows the user to enter textual information.
     - **Date Selection** allows the user to select a date.
     - **Check Box** allows the user to choose between two distinguishable states, such as "true or false" or "yes or no".
     - **Number Selection** allows the user to enter a number or to select a number using the up and down arrows.
     - **Radio Button** allows the user to select from among mutually exclusive choices.
     - **Selection** allows the user to select an item from a list.
     - **Combo Box** allows the user to select items from a list or to enter a new value.
   - For fields of the **Text Box** representation type, you have to set the **Maximum length** for the entry field. You can also define a **Default value**.
   - For fields of the **Number Selection** representation type, you have to set the **Minimum value** and the **Maximum value** for the entry field. You can also define a **Default value**.
   - For fields of the **Selection** and **Combo Box** representation types, you can also specify a **Default value**. In addition, you must specify if the list is dynamic or static:

     Populate list from web service
     If you want to provide the user with a dynamic list to pick from, you can specify the URL of the web service that is to supply the information. As a parameter, you can provide the user name, for

example, `http://www.example.com/myWebservice?user=${user name}`. The format of the user name depends on the mail system that you use:

Lotus Domino
The canonical format is used, for example: CN=John Smith/OU=Unit/O=Exampleorg

Microsoft Exchange
The SMTP address is used, for example: johnsmith@example.org

The user name is retrieved when a user selects to specify additional archiving information.

Any web service that you use with the default metadata form template to provide data for the list must support a specific format for the Dojo ItemFileReadStore, the general case type map. However, you can adapt the template to support any other kind of web service.

To learn more about the Dojo toolkit, go to the Dojo Toolkit website.

Static list
For a static list, you define a set of key-value pairs for the user to select. The key is displayed in the selection list. When a user selects a key, the associated value is assigned to the property.

b. Click **OK** to save your change and return to the configuration pane.

5. Select the **Default visible columns**. These are the columns in the input form that are shown by default.
6. Save the definitions.
7. To check the layout of the input form, select a **Metadata Form Template** and click **Preview**.

Related tasks

"Selecting the metadata form template" on page 234

Adding and editing metadata

Metadata is information about particular content. IBM Content Collector allows you to include extra metadata related to each document processed. You can then use these fields as if they were regular system metadata, for example in repository tasks property mapping or for evaluating rules.

Metadata about a library book, for example, includes title, author, and date of publication. These can be thought of as the properties of a book. Metadata about a photograph includes different properties: perhaps the date the photo was taken and details of the camera settings.

The metadata fields in a repository provide search information for user queries. For example, the repository field that corresponds to the *From* or *Sender* field of an email allows users to search for email that was sent by a specific person. Certain fields are selected by default. You can select other fields to extract metadata from to add search information to your repository.

You can make use of three different kinds of metadata:

User-defined metadata

IBM Content Collector allows you to define metadata to be used by this application when processing items.

You can define custom metadata through the **Metadata and Lists** configuration windows. You can then use the custom metadata you defined to process email by putting a *Metadata* task into a task route.

File system metadata

IBM Content Collector enables you to include extra metadata related to each document processed, then use these fields like regular system metadata properties.

The first step is to define the metadata, in the form of XMLor CSVfiles, to be used by the application.

Basic file properties may not contain all the data you need to get the most from your files. For example, a bank makes a PDFcopy of every transaction. The PDFfile contains minimal data, including only account numbers, date of processing, and a text copy of the full transaction. However, the bank also wants to associate a customer name and transaction type with the file.

Typically, one XML metadata file is used per file or item that is to be collected, whereas CSV files can contain information about thousands of files that are to be collected, one per line. The XML file exists in a monitored location along with a similarly-named file that it describes. ACSVfile exists in a monitored location along with perhaps thousands of records, each of which is described in a row in the CSV file.

Important: The File System Source Connector must run as a user that has permissions to access the metadata files or there must be a trust relation ship between the IBM Content Collector system and the system where the files are located.

You define file system metadata through the **Metadata and Lists** configuration windows. The *FSC Associate Metadata* task in a task route then adds the metadata to the file at the moment when the task is invoked. Thereafter the information can be used in rules or as input to later tasks in the task route, for example, the *P8 4.x Create Document* task.

System metadata

Different document sources create differing file formats, each with properties specific to its type. An Exchange email server, for example, creates msg files, with properties including Subject line and To field. When you choose a metadata type in a *Create Document* task, you are narrowing the properties available to just those for that metadata type.

Related tasks

"Adding and editing file system metadata" on page 241

"Adding and editing user-defined metadata" on page 240

Related reference

"FSC Associate Metadata" on page 531

"EC Extract Metadata" on page 522

Adding, editing and sorting lists

A list is a collection of strings organized into key-value pairs.

You can use these strings in rules, or you can use them to provide standard values for property mappings (for List Lookup).

Rules example

You can create a list of customers, then create a rule in an email archiving task route that captures documents where the names from the list occur in the *From* field. Then, the rule would select a specific item type or object class in the repository for these documents.

Property mappings example

Strings organized into key-value pairs can also be used in property mappings. Assume that you have a property mapping which maps the *Subject* field in email documents to the metadata field *Subject* in your repository. You could now define a condition which captures email with the words *contracts, contractor, sub-contractor, terms and conditions*, and so on, in the *Subject* field. Assume further that your repository contains an additional metadata field *Category*. You could now define that for each email with one of the keywords in the subject line, the value *Contract* has to be written to the *Category* field. The word *Contract* would be the return value for the terms *contracts*,*contractor*,*sub-contractor*, and *terms and conditions* in your key-value string. This would facilitate searches for email that have to do with contracts because users can search for the word *Contract* in the *Category* metadata field, and would find all related email, including those that have, for example, just *terms and conditions* in the subject line.

You can:

- Add or edit a list
- Delete a list
- Edit an item in a list
- Delete an item in a list
- Sort items in a list

To add, edit, or delete a list or an item in a list, or to sort items in a list:

1. In the Configuration Manager, click **Metadata and Lists** to switch to the Metadata andLists view.
2. In the **Metadata and Lists** pane on the upper left, click **Lists**.
3. The display pane displays any previously created lists.
4. In the configuration pane, enter or edit a name and description for the list.
5. Under the **Value List** box, do one of the following:
   - To add an item to the list, click the **Add** icon.
   - To edit an item in the list, click the **Edit** icon.
   - To delete an item in the list, click the **Remove** icon.
   - To sort items in the list, select the item you want to move and click the up or down arrow icon.
6. When you add or edit an item the **New Item** window appears:
   a. In the **Name** field, enter or edit the word.
   b. In the **Description** field, enter or edit a description for the list item.
   c. In the **Returned value** field, enter or edit the value to be returned if a value matches criteria defined in a rule.
   d. Click **OK** to add your entries to the list of values.
7. Save the list in the configuration database.

Adding and editing user-defined metadata

You can set up user-defined metadata to be used by this application when processing items.

To add or edit user-defined metadata:

1. In the Configuration Manager, click **Metadata and Lists** to switch to the Metadata andLists view.
2. In the **Metadata and Lists** pane on the upper left, select **User-Defined Metadata**.
3. In the design pane, you have the following options:
   - Add a new set of user-defined metadata.
   - Edit a set of user-defined metadata that already exists in the configuration database.
   - Delete a set of user-defined metadata from the configuration database.
   - Import property definitions from a file into the set of user-defined metadata that you add or edit. In this case, you can also select to overwrite existing property definitions.
   - Export property definitions from a set of user-defined metadata in the configuration database to a file. In this case, you must also specify a file name and the folder where the file is to be stored.
4. The configuration pane contains the following configuration options. For an existing set of user-defined metadata, the fields contain values, which you can change. For a new set of user-defined metadata, enter the required values.

   Property ID
   The property ID identifies the metadata source to the system. It must be unique. You can use the ID that is generated by the system, or you can assign your own property ID.

   Display name
   The name of this set of properties as it is displayed to the user (for example, `Contracts`).

   Description
   The description of the group of properties.

   Metadata Properties
   A list of the properties in the selected set of user-defined metadata. You can edit an existing item or add new items. The options for the metadata properties are:

   Property ID
   The property ID uniquely identifies the property of this metadata source to the system, for example, `ibm.icc.file.filepath`.

   Display name
   The name of the property as it is displayed to the user. For example, the display name for the property with the ID `ibm.icc.file.filepath` could be `FilePath`.

   Data type
   The data type of the property.

   Multi-value
   This setting defines if the property can have more than one value.

Restriction: This setting is not available for collecting additional archiving information nor for use with a SharePoint source system.

Description
The description of the property.

The column **Referenced** indicates whether a property is in use in a task route or a metadata form. When a property is in use, you cannot remove it or change its type.

Related tasks

"Collecting from Microsoft SharePoint sites" on page 482

Related reference

"FSC Associate Metadata" on page 531

Adding and editing file system metadata

Before you can make use of custom file metadata, you must identify the source fields to IBM Content Collector.

You begin by defining metadata. This metadata resides in XML or delimited CSV files. In the File System task route, the *FSC Associate Metadata* task then tells the system where to find the metadata files that you defined and adds the metadata to the file that is being processed by the task route at the moment when the task is invoked. This information can then be used in rules or as input to later tasks in the task route, for example, the *P8 4.x Create Document* task.

To add or edit file system metadata:

1. In the Configuration Manager, navigate to **Metadata and Lists → File System Metadata**.
2. In the display pane, click the add icon or select an entry and click the edit icon.
3. In the configuration pane, enter or edit the name and description for the metadata.
4. Select a format type of **XML** or **Delimited**, depending on the format of the files in which you store your custom metadata.
5. If your format type is **Delimited**:
   a. Configure the **Delimited File Properties**
      - Select the **Delimiter** that separates the columns.
      - If any column in the metadata file contains multiple values, select the **Multi-value delimiter** that separates each value in the column. (The value must be different from the value that you select for text qualifier).
      - If text in the metadata file is enclosed in specific characters, such as quotation marks, select a **Text Qualifier**.
      - If the first row of the metadata file contains labels defining the content of each row in the file, select **First row contains labels**. This causes the application to ignore the values in the first row.
   b. Identify delimited file system metadata.
6. If your format type is **XML**, Identify XML-based file system metadata.
7. Save the configuration settings.

Related tasks

"Archiving email from local files" on page 432

"Defining metadata to be used to process files for archiving" on page 729

Related reference

"EC Extract Metadata" on page 522

"FSC Associate Metadata" on page 531

Identifying delimited file system metadata

You can enter custom file metadata manually or use the metadata mapping wizard to help you identify delimited file system metadata to IBM Content Collector.

To add delimited metadata:

1. Navigate to **Metadata and Lists** → **File System Metadata** → **Delimited File Metadata Mappings**.
2. To use the metadata mapping wizard to import a CSV file that contains the metadata field names, click the **Wizard** icon and follow the instructions in "Importing metadata using the delimited file wizard" on page 243. To enter metadata manually, skip this step.
3. Enter or edit the display name and description for the metadata type.
4. Enter a column number.
5. Select the data type of the metadata. You might have to change the data type of any column that you imported with the wizard. For each column, the values that are in the file must be values that can be converted to the target type of the column. If the value in a file cannot be converted, an error is reported.
   a. If you select **Date Time**, you must specify a **Date Format**. Use the following case-sensitive tokens when specifying the date.

Token	Description
M	Months as 1 – 12
MM	Months as 01 – 12
MMM	Months as Jan – Dec
MMMM	Months as January – December
d	Days as 1 – 31
dd	Days as 01 – 31
ddd	Days as Mon – Sun
dddd	Days as Monday – Sunday
y	Years as 1, 2, ... , 99
yy	Years as 00 – 99
yyyy	Years as 1900 – 9999
h	Hours as 0 – 12
hh	Hours as 00 – 12
H	Hours as 0 – 23
HH	Hours as 00 – 23
m	Minutes as 0 – 59
mm	Minutes as 00 – 59
s	Seconds as 0 – 59
ss	Seconds as 00 – 59

Token	Description
tt	AM/PM
t	A/P

For example, `MM/dd/yy HH:mm:ss` displays as 11/25/10 (November 25, 2010) and 11:13:30 (30 seconds after 11:13 AM). `dddd, d MMMM yyyy` displays as Monday, 1 January 2010.

b. If the date is a UTC date, select **Date is in UTC**. If the date in the file is a UTC date, the value in the metadata file will be used as is. If the date is not a UTC date, the timezone setting in IBM Content Collector will be used to convert the date to UTC format.

6. Save the configuration settings.

Importing metadata using the delimited file wizard:

Use the delimited file wizard to import a sample file that defines the format of your custom metadata.

The sample CSV file that you import with this wizard must use a format similar to the following example (which uses | as the delimiter, though any of the available delimiters is acceptable):

```
NAME|AGE|GENDER|PRICE|DATE
```

1. Click **Browse** and select a sample metadata file in CSV format. This file requires only one row, which should contain the names of the columns that you want to map. If the results do not display the expected results, check to see that you have identified the correct delimiter and other file properties. The **Sample File** section contains headings for each column in the sample file. If you selected **First row contains labels**, the column headings shown contain the data in the first row in the file. Otherwise, the columns are named **Column** followed by the number of the column.
2. To map columns, select a column name from the **Sample File** region and click the arrow to copy it to the **Mappings** region. Repeat as needed.
3. Click **OK** to save your definitions and leave the wizard. Click **Cancel** to leave the wizard without saving the mapping definitions.
4. Configure the columns that you have mapped. The wizard identifies all columns as strings, so you might need to change the data type to another type, add a date format to a *Date Time* type, or make other adjustments. Note, for instance, that the example above contains only one column of the *String* data type.

Identifying XML-based file system metadata

You can enter custom metadata manually or use the metadata mapping wizard to help you identify XML-based file system metadata to IBM Content Collector.

1. Navigate to **Metadata and Lists** → **File System Metadata** → **XML File Metadata Mappings**.
2. To use the metadata mapping wizard to import an XML file that contains the metadata field names, click the **Wizard** icon and follow the instructions in "Importing metadata using the XML file wizard" on page 245. To enter metadata manually, skip this step.
3. Enter or edit the display name and description for the metadata type.
4. Under **XPath**, enter the full XML path for the element. XPath expressions support the Microsoft XML Parser (MSXML) syntax and are evaluated in Content Collector by using MSXML Version 6.

Name space prefixes are not supported. To work around this limitation, write an XPath in the following way:

```
/*[ local-name() = 'document' ]/*[ local-name() = 'title' ]
```

Note the use of predicates that use the local-name() function.

If the XPath evaluation results in no node values, the following values are added to the metadata:

- For property type date or a date array: no value
- For property type numeric: 0
- For property type string: an empty string
- For any other array types: an empty array

If a node value cannot be converted to a target property type, an error is reported.

A single XML file can supply metadata values for one or more documents. Values are associated with files based on the order in which they are found. Consider:

```
<documents>
<files>
<value>file1.doc</value>
<value>file2.doc</value>
</files>
<titles>
<value>title 1</value>
<value>title 2</value>
</titles>
</documents>
```

To select the file names, use `/documents/files/value/text()` and use `/documents/titles/value/text()` to select the titles.

Relative paths can be used as well. Consider:

```
<documents>
<document>
<file>test1.doc</file>
<title>title 1</title>
</document>
<document>
<file>test2.doc</file>
<title>title 2</title>
</document>
</documents>
```

To select the file name, use `/documents/document/file/text()` and use `../title/text()` to select the title.

5. Select the data type of the element. You might have to change the data type of any element that you imported with the wizard. For each XPath, the values that are in the file must be values that can be converted to the target type of the property. If the value in a file cannot be converted, an error is reported.
   a. If you select **Date Time**, you must specify a **Date Format**. Use the following case-sensitive tokens when specifying the date.

Token	Description
M	Months as 1 – 12
MM	Months as 01 – 12
MMM	Months as Jan – Dec

Token	Description
MMMM	Months as January – December
d	Days as 1 – 31
dd	Days as 01 – 31
ddd	Days as Mon – Sun
dddd	Days as Monday – Sunday
y	Years as 1, 2, ... , 99
yy	Years as 00 – 99
yyyy	Years as 1900 – 9999
h	Hours as 0 – 12
hh	Hours as 00 – 12
H	Hours as 0 – 23
HH	Hours as 00 – 23
m	Minutes as 0 – 59
mm	Minutes as 00 – 59
s	Seconds as 0 – 59
ss	Seconds as 00 – 59
tt	AM/PM
t	A/P

For example, `MM/dd/yy HH:mm:ss` displays as 11/25/10 (November 25, 2010) and 11:13:30 (30 seconds after 11:13 AM). `dddd, d MMMM yyyy` displays as Monday, 1 January 2010.

b. If the date is a UTC date, select **Date is in UTC**. If the date in the file is a UTC date, the value in the metadata file will be used as is. If the date is not a UTC date, the timezone setting in IBM Content Collector will be used to convert the date to UTC format.

6. Save the configuration settings.

Importing metadata using the XML file wizard:

Use the XML file wizard to import a sample file that defines the format of your custom metadata.

The sample XML file that you import with this wizard must use a format like the following example:

```
<xml>
 <AccountNumber>78956</AccountNumber>
  <DepositAmount>789.98</DepositAmount>
 <FirstName>Nigel</FirstName>
</xml>
```

Restriction: If your XML file contains namespaces, you cannot use the task output without manually adding an XPath that accesses XML nodes using the "local-name()" function. For example, `//*[local-name()='book' and @category='Science']/*[local-name()='name']/text()`

1. Click **Browse** to select a sample metadata file in XML format. The **Sample File** region contains all the elements in the XML file. To add an element to the Metadata Mappings table, expand the element completely until the function associated with the tag appears. For most tags the name of the function is

`text()`. The function and the elements in the path constitute an XPath. which IBM Content Collector uses to identify the location of an item in an XML file.

2. Select a function and click the arrow to copy the element into the Metadata Mappings table.
3. Click **OK** to save your definitions and leave the wizard. Click **Cancel** to leave the wizard without saving the mapping definitions.
4. Configure the columns that you have mapped. The wizard identifies all columns as strings, so you might need to change the data type to another type, add a date format to a *Date Time* type, or make other adjustments.

Selecting system metadata

The **System Metadata** screen shows what metadata sources were registered by installed components. All metadata properties provided by installed components are listed under **System Metadata Properties**.

Archiving format system metadata properties

The Archiving format metadata type contains metadata that specifies the archiving format. This metadata is used by the Text Extraction task.

Property	Description
Bundled Resource Item	Boolean indicator that states if the archiving format is BRI.

This metadata is produced by the following tasks:

EC Finalize Email for Compliance

EC Prepare Email for Archiving

SC Prepare Email for Archiving

Related reference

"EC Finalize Email for Compliance" on page 524

"EC Prepare Email for Archiving" on page 525

"SC Prepare Email for Archiving" on page 566

Attachment Deduplication system metadata properties

The Attachment Deduplication metadata type contains metadata that is added to an attachment and is available during task route processing after the extract metadata task in the task route.

Property	Description
Correlation ID	The identifier for the position of an attachment within a specific email. This identifer is unique per email.
Display Name	The display name as it appears in the email client.
Document ID	The hash key derived from the attachment binary content. This key is used for detecting duplicates.

This metadata is produced by the following tasks:

EC Extract Attachments

SC Extract Attachments

Related reference

"EC Extract Attachments" on page 522

"SC Extract Attachments" on page 564

Blacklist system metadata properties

The Blacklist metadata type contains metadata for the documents that are put on the blacklist.

Property	Description
Delete	A flag that indicates whether an entry in the blacklist is to be deleted.
Do Not Process	A flag that indicates whether an item can be processed. If the value is false, the task route service will only update the blacklist table. The submitted item will not undergo normal task route processing. Therefore, in any rule evaluation or in an audit log, this value will always be true.
Location	The location of the failure as it is defined by the connector, for example, Mailbox/MailID for the Email Connector.
Permanent	A flag that indicates whether the error is of permanent nature, that is, if a document causes the connector to fail, or if the maximum number of times that a collector is allowed to process documents again is reached.
Reason	The reason for the failure as identified by the connector.

This metadata is produced by the following tasks:

All tasks of an Email Connector or an SMTP Connector if a processing error occurs for a document.

Calculate Expiration Date system metadata properties

The Calculate Expiration Date metadata type contains metadata that governs the expiration of archived documents.

Property	Description
Expiry Date	The date on which the document becomes eligible for deletion.
Expiry Determinant	The criterion that governs the expiration of the document, such as the retention period specified for a specific user.

This metadata is produced by the following tasks:

Calculate Expiration Date

Related reference

"Calculate Expiration Date" on page 498

CM 8.x Confirm Document system metadata properties

The CM 8.x Confirm Document system metadata type contains metadata about the status of the document in the repository.

Property	Description
Connection Name	The name of the CM 8.x connection that was used to perform the CM 8.x Confirm Document task.
Document Confirmed	A flag that indicates whether the document exists in the repository.
Shortcut URL	The shortcut URL of the document in the repository (if it exists).

This metadata is produced by the following task:

CM 8.x Confirm Document

CM 8.x Create Document system metadata properties

The CM 8.x Create Document metadata type contains metadata that is added to an item after it is captured into the IBM Content Manager repository.

Property	Description
Connection Name	The name of the CM 8.x connection that was used to perform the CM 8.x Create Document task.
Duplicate	A flag that indicates whether this item is a duplicate.
Folder IDs	The persistent identifiers (PIDs) of the folders in which the document was filed. **Note:** This field can be empty if no folders were specified.
Item Type	The symbolic name of the item type used to capture an item into IBM Content Manager.
Object ID	The unique ID (the IBM Content Manager PID) of the item captured into IBM Content Manager.
Repository Name	The symbolic name of the IBM Content Manager repository to which the document has been captured.
Shortcut URL	A URL that references the item that has been captured into IBM Content Manager.

This metadata is produced by the following tasks:

CM 8.x Create Document

CM 8.x Associate Content

CM 8.x Store Version Series

Related reference

"CM 8.x Associate Content" on page 499

"CM 8.x Create Document" on page 506

"CM 8.x Store Version Series" on page 511

CM 8.x Duplicate system metadata properties

The CM 8.x Duplicate metadata type contains metadata that is added to an item by the CM 8.x Duplicate Detection task. The Duplicate property indicates whether the item was found in the repository.

Property	Description
Connection Name	The name of the CM 8.x connection that was used to perform the CM 8.x Duplicate Detection task.
Duplicate	A flag that indicates whether this item is a duplicate.
Item Type	The symbolic name of the item type used to capture an item into IBM Content Manager.
Object ID	The unique ID of the item captured into IBM Content Manager.
Repository Name	The symbolic name of the IBM Content Manager repository to which the document has been captured.
Shortcut URL	A URL that references the item that has been captured into IBM Content Manager.

This metadata is produced by the following task:

CM 8.x Duplicate Detection

Related reference

"CM 8.x Duplicate Detection" on page 509

CM 8.x Update system metadata properties

The CM 8.x Update metadata type contains metadata with information about the current state of the document in the IBM Content Manager repository.

Property	Description
Connection Name	The name of the CM 8.x connection that was used to perform the CM 8.x Confirm Document task.
Item Type	The symbolic name of the item type used to capture an item into IBM Content Manager.
Object ID	The unique ID of the item captured into IBM Content Manager.
Repository Name	The symbolic name of the IBM Content Manager repository to which the document has been captured.
Shortcut URL	A URL that references the item that has been captured into IBM Content Manager.

This metadata is produced by the following tasks:

CM 8.x Update Document

Related reference

"CM 8.x Update Document" on page 514

Collector Information system metadata properties

The Collector Information system metadata type contains metadata that provides information about the collector that submitted the item.

Property	Description
Collector ID	The unique ID of the collector that is used. This property can be used track which collector was responsible for collecting an item.

Email system metadata properties

The Email metadata type contains metadata specific to an email. The Microsoft Exchange, Lotus Domino, and SMTP collectors generate this metadata.

Properties that have "(multi)" next to the name refer to properties that consist of a list of items and can be mapped to a multi-value field in the document repository. Most of these fields have a corresponding version of the property with "(single)" next to the name that contains a representation of the list as a single value. For example, if the "To Addresses (multi)" field contains:

- dans@mycompany.com
- toms@mycompany.com

The "To Addresses (single)" version of the field would be: "dans@mycompany.com;toms@mycompany.com"

You can extract address information in various formats depending on the property that you select. See the topic about email address formats for more information.

Property	Description
Attachment Count (Lotus only)	The number of attachments in the email. This property is not available for email that was collected from Microsoft Exchange sources.
Attachment Flag	Indicates whether an email has an attachment.
BCC Addresses (multi)	A list of the email addresses of recipients in the BCC field.
BCC Addresses (single)	A single value version of BCC Addresses.
BCC Display (multi)	A list of the display names of recipients in the BCC field.
BCC Display (single)	A single value version of BCC Display.
BCC Display/Addresses (multi)	A list of both the email address and display name of recipients in the BCC field.
BCC Display/Addresses (single)	A single value version of BCC Display/Addresses.
BCC Domain (multi)	A list of the domains names of the email addresses of recipients in the BCC field.
BCC Domain (single)	A single value version of BCC Domain.
BCC Raw (multi)	A list of the email servers internal representation of recipients in the BCC field.
BCC Raw (single)	A single value version of BCC Raw.
CC Addresses (multi)	A list of the email addresses of recipients in the CC field.
CC Addresses (single)	A single value version of CC Addresses.

Property	Description
CC Display (multi)	A list of both the display names of recipients in the CC field.
CC Display (single)	A single value version of CC Display.
CC Display/Addresses (multi)	A list of both the email address and display name of recipients in the CC field.
CC Display/Addresses (single)	A single value version of CC Display/Addresses.
CC Domain (multi)	A list of the domains names of the email addresses of recipients in the CC field.
CC Domain (single)	A single value version of CC Domain.
CC Raw (multi)	A list of the email servers internal representation of recipients in the CC field.
CC Raw (single)	A single value version of CC Raw.
Conversation Topic	The subject line of the email with subject line prefixes such as RE:, FW:, FWD: removed.
Database Title (Lotus only)	The title of theLotus Notes database the email was found in.
Document Title	If the item is not an attachment, this property contains the subject line of the email. If the item is an attachment, this property contains the name of the attachment.
Folder	The folder in the mailbox or database that the email is contained in. This property is only available if the email has been collected from a folder, so you must define at least one folder to include or exclude.
Form Correlation Key	The unique key that links additional archiving information to an email document.
From Address (multi)	A multi-value version of From Address.
From Address (single)	The email address of the user owning the originating mailbox.
From Display (multi)	A multi-value version of From Display.
From Display (single)	The display name of the user owning the originating mailbox.
From Display/Address (multi)	A multi-value version of From Display/Address.
From Display/Address (single)	Both the email address and display name of the user owning the originating mailbox.
From Domain (multi)	A multi-value version of From Domain.
From Domain (single)	The domain name of the email address of the user owning the originating mailbox.
From Raw (multi)	A multi-value version of From Raw.
From Raw (single)	The email server's internal representation of the originating mailbox information.
Has Been Altered (Lotus only)	A flag indicating whether the item has been modified.

Property	Description
Instance-Specific information	Information that is unique to this instance of the email.
Is Attachment	A flag indicating whether the current item is an attachment or an email.
Is Encrypted	A flag indicating that the item is encrypted.
Is Mime (Lotus only)	For Lotus Notes email, indicates whether the email is in MIME format.
Is Signed (Lotus only)	A flag indicating that the item has been signed.
Journal Envelope Headers	The header of the journaled email, which includes information about the message ID, the recipients, and the sender.
Journal Envelope Message ID (Exchange only)	The identifier for the email.
Journal Envelope Original Message ID (Exchange only)	The identifier for the originating email.
Journal Envelope Recipients Addresses XML (Exchange only)	The recipients of the journaled email.
Journal Envelope Sender/On-Behalf-Of Address (Exchange only)	The sender of the journaled email.
Mailbox ID	The unique identifier for the originating mailbox of the email.
Managed Folder Retention Date (Exchange only)	The retention date that was calculated based on the retention settings in Exchange. This property is available for those documents only for which the action Permanently Delete is defined in the managed folder configuration in Exchange as the action to take at the end of the retention period.
Managed Folder Retention Flag (Exchange only)	A flag indicating whether retention information is available for the processed document.
Message Body	The body of the email.
Message Form	The type of item currently being processed.
Message Size	The size of the email in bytes.
Originating User	The display name of the actual sender of the email.
Parent ID	For an attachment, the Message ID of the email that attachment is contained in.
Processing State	The processing state of the email (archived or retrieved).
Received Date	The date when the email was received.
Recipients Addresses (multi)	The email addresses of all recipients of the email, including the TO, BCC, and CC recipients, and the recipients of the journaled email.
Sent Date	The date on which the email was sent.

Property	Description
Subject	The subject line of the email. If the subject line is blank, the Configuration Manager may be used to specify the value to use for the subject line.
To Addresses (multi)	A list of the email addresses of recipients in the To field.
To Addresses (single)	A single value version of To Addresses.
To Display (multi)	A list of the display names of recipients in the To field.
To Display (single)	A single value version of To Display.
To Display/Addresses (multi)	A list of both the email address and display name of recipients in the To field.
To Display/Addresses (single)	A single value version of To Display/Addresses.
To Domain (multi)	A list of the domain names of the email addresses of recipients in the To field.
To Domain (single)	A single value version of To Domain.
To Raw (multi)	A list of the email server's internal representation of recipients in the To field.
To Raw (single)	A single value version of To Raw.
Unique ID	The unique identifier for the email.

This metadata is produced by the following tasks:

EC Extract Attachments

EC Extract Metadata

SC Extract Attachments

SC Extract Metadata

Related reference

“EC Extract Attachments” on page 522

“EC Extract Metadata” on page 522

“SC Extract Attachments” on page 564

“SC Extract Metadata” on page 565

Email address formats:

The Email system metadata includes several properties that provide address information. When you extract address information, the representation of the address depends on the selected property.

Table 46. Email address formats

Property	Format
BCC Address CC Address From Address To Address	**Email Connector for Microsoft Exchange** • For email sent to or received from addresses within the Exchange domain `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com` • For email received from or sent to an Internet address (SMTP format) `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com` **Email Connector for Lotus Notes** • For email sent to or received from addresses within the Notes domain Abbreviated format without the domain name, for example, `MyName/Germany/MyCompany` • For email received from or sent to an Internet address (SMTP format) `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com` **SMTP Connector** SMTP address, such as `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com`
BCC Display CC Display From Display To Display	**Email Connector for Microsoft Exchange** • For email sent to or received from addresses within the Exchange domain Display name, for example, `MyName` • For email received from or sent to an Internet address (SMTP format) Display name, such as `MyName`, if available, otherwise `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com` **Email Connector for Lotus Notes** • For email sent to or received from addresses within the Notes domain Common name, for example, `MyName` • For email received from or sent to an Internet address (SMTP format) Common name, such as `MyName`, if available, otherwise the local part of the address, for example, `myname` of the address `myname@my_company.com` **SMTP Connector** Display name, such as `MyName`, if available, otherwise `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com`
BCC Domain CC Domain From Domain To Domain	**Email Connector for Microsoft Exchange** • For email sent to or received from addresses within the Exchange domain `fully_qualified_domain_name`, for example, `my_company.com` • For email received from or sent to an Internet address (SMTP format) `fully_qualified_domain_name`, for example, `my_company.com` **Email Connector for Lotus Notes** • For email sent to or received from addresses within the Notes domain Domain name, for example, `MYCOMPANYDE` • For email received from or sent to an Internet address (SMTP format) `fully_qualified_domain_name`, for example, `my_company.com` **SMTP Connector** `fully_qualified_domain_name`, for example, `my_company.com`

Table 46. Email address formats (continued)

Property	Format
BCC Raw CC Raw From Raw To Raw	**Email Connector for Microsoft Exchange** • For email sent to or received from addresses within the Exchange domain Canonical format, for example, `/O=Exchange_organization_name/OU=Exchange_administrative_group_name/CN=RECIPIENTS/CN=MYNAME` • For email received from or sent to an Internet address (SMTP format) `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com` **Email Connector for Lotus Notes** • For email sent to or received from addresses within the Notes domain Canonical format including the domain name, for example, `CN=MyName/OU=Sales/O=MyCompany@MYCOMPANYDE` • For email received from or sent to an Internet address (SMTP format) Full SMTP email address including the common name, for example, `"MyName"<myname@my_company.com>` **SMTP Connector** Value that is contained in the MIME header BCC, CC, FROM, or TO, for example, `MyName(myname@my_company.com)`
BCC Display/Address CC Display/Address From Display/Address To Display/Address	**Email Connector for Microsoft Exchange** • For email sent to or received from addresses within the Exchange domain Full SMTP email address including the display name, for example, `MyName(myname@my_company.com)` • For email received from or sent to an Internet address (SMTP format) Full SMTP email address including the display name, for example, `MyName(myname@my_company.com)`, if the display name is available, otherwise `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com` **Email Connector for Lotus Notes** • For email sent to or received from addresses within the Notes domain Abbreviated format including the domain name, for example, `MyName/Sales/MyCompany@MYCOMPANYDE` • For email received from or sent to an Internet address (SMTP format) Full SMTP email address including the common name, for example, `"MyName"<myname@my_company.com>` **SMTP Connector** Full SMTP email address including the display name, for example, `MyName(myname@my_company.com)`, if the display name is available, otherwise `user_name@fully_qualified_domain_name`, for example, `myname@my_company.com`

Email Deduplication system metadata properties

The Email Deduplication metadata type contains metadata that is used to detect duplicates of messages.

Property	Description
Record ID	The record ID, which is unique for each archived document.
Document ID	The document ID generated for single-instance storing.

This metadata is produced by the following tasks:

EC Extract Metadata

SC Extract Metadata

Related reference

"EC Extract Metadata" on page 522

"SC Extract Metadata" on page 565

File system metadata properties

The file system metadata type contains metadata specific to files found on the file system.

Property	Description
Accessed	The date the file was last accessed.
Access Control List	This contains the NTFS access control list. It can be used to configure security in IBM FileNet P8.
Archive Flag	A flag that indicates whether the Archive property has been selected for the file.
Captured Flag	A flag that indicates whether the file has been previously captured into the document repository. This is set by the FSC Post Processing task.
Compressed Flag	A flag that indicates whether the Compressed property has been selected for the file.
Created Date	The date the file was created.
Encrypted Flag	A flag that indicates whether the Encrypted property has been selected for the file.
File Extension	The file extension of the file. For example, for the file `C:\docs\myfile.txt`, this property would be set to `.txt`.
File Folder Path	The folder the file was found in. For example, for the file `C:\docs\myfile.txt`, this property would be set to `C:\docs`.
File Name	The name of the file including the file's extension but excluding the file's path. For example, for the file `C:\docs\myfile.txt`, this property would be set to `myfile.txt`.
File Name (without extension)	The name of the file excluding the file extension (everything after the last "." in the file) and the file path. For example, for the file `C:\docs\myfile.txt`, this property would be set to `myfile`.
File Size	The size of the file in bytes.
Filtered Owner	The owner of the file according the NTFS security "Owner" screen. The value is set to the owner's Shortname. If Shortname cannot be derived, the value is set to blank.
Full Path	The full path of the file. For example, for the file `C:\docs\myfile.txt`, this property would be set to `C:\docs\myfile.txt`.

Property	Description
Hidden Flag	A flag that indicates whether the Hidden property has been selected for the file.
Modified Date	The date the file was last modified.
Owner	The owner of the file according the NTFS security “Owner” screen. The value is either set to the owner's Shortname or the SID of the owner if Shortname cannot be derived.
Processed Flag	A flag that indicates whether the file has been previously processed. This is set by the FSC Post Processing task.
Read-only Flag	A flag that indicates whether the Read Only property has been selected for the file.

This metadata is produced by the file system collector and by the following tasks:

EC Finalize Email for Compliance

EC Prepare Email for Archiving

EC Extract Attachments

SC Prepare Email for Archiving

SC Extract Attachments

SP Create File

SP Get Versions

Related reference

“EC Extract Attachments” on page 522

“EC Finalize Email for Compliance” on page 524

“EC Prepare Email for Archiving” on page 525

“SC Extract Attachments” on page 564

“SC Prepare Email for Archiving” on page 566

“SP Create File” on page 567

“SP Get Versions” on page 568

FileNet Image Services Create Document system metadata properties

The FileNet IS Create Document metadata type contains metadata that is added to an item after it is filed into a folder in the FileNet IS library.

Property	Description
Document ID	The unique ID of the item captured into FileNet IS library.
Library Name	The symbolic name of the FileNet IS library to which the document has been captured.
Shortcut URL	A URL that references the item that has been captured into FileNet IS library.

This metadata is produced by the following task:

FileNet Image Services Create Document

Related reference

"FileNet Image Services Create Document" on page 528

FileNet Image Service File Document in Folder system metadata properties

The FileNet Image Services File Document in Folder metadata type contains metadata that is added to an item after it is filed into a folder in the FileNet IS library.

Property	Description
Document ID	The unique ID of the item captured into the FileNet IS library.
Folders	A list of folders into which the document has been captured.
Library Name	The symbolic name of the FileNet IS library to which the document has been captured.

This metadata is produced by the following task:

FileNet Image Services File Document In Folder

Related reference

"FileNet Image Services File Document In Folder" on page 529

FileNet Image Services Modify Permissions system metadata properties

The FileNet Image Services Modify Permissions metadata type contains metadata that is added to an item after permissions are assigned to it in the FileNet IS library.

Property	Description
Access Levels	The granted access levels (read, write, append/execute).
Document ID	The unique ID of the item captured into the FileNet IS library.
Grantees	The users or user groups to whom the defined access levels are granted.
Shortcut URL	A URL that references the item that has been captured into the FileNet IS library.

This metadata is produced by the following task:

FileNet Image Services Modify Permissions

Related reference

"FileNet Image Services Modify Permissions" on page 530

FSC Duplicate Management system metadata properties

The FSC Duplicate Management metadata type contains metadata to detect document duplicates. If selected, this metadata is calculated by the file system connector for each file that is collected in the task route and is used by later tasks to detect duplicate documents.

Property	Description
Hashkey	The identifying key for an archived document. The value is an MD5 hash of the content of the file.

FSC Metadata system metadata properties

The FSC Metadata system metadata type contains metadata that is specific to metadata files. The metadata information produced by the FSC Associate Metadata task describes the relationship between the metadata file and the document file.

Property	Description

Content File Line Number	The line number of the document entry in the .csv file.
	Note:
	
	The value of this property is only meaningful under the following conditions:
	• You have selected **Metadata File** as the input file type.
	• You have selected **Values in the metadata file** in the configuration of the FSC Associate Metadata task.
	• The metadata file is in CSV format.
	If these conditions are met, the value of this property is the line number in the CSV file that describes the document.
	
	For example:
	`important_doc.doc,Joe,joe@company.com`
	`important_image.jpg,Betty,betty@company.com`
	
	In this example, the value `important_doc.doc` is 1 and the value for `important_image.jpg` is 2.

Property	Description

Content File Order Number	The order number of the document entry in the CSV or XML file.
	Note:
	
	The value of this property is only meaningful under the following conditions:
	• You have selected **Metadata File** as the input file type.
	• You have selected **Values in the metadata file** in the configuration of the FSC Associate Metadata task.
	
	If these conditions are met, the value of this property indicates the position of the document name in the metadata file.
	
	For example:
	`<files_to_collect>`
	`<file name="important_doc.doc" />`
	`<file name="important_image.jpg" />`
	`</files_to_collect>`
	
	In this example, the value `important_doc.doc` is 1 and the value for `important_image.jpg` is 2.

Is Contentless	A flag that indicates if the properties of the files listed in the CSV file are to be processed and not the files themselves. Note that this means that documents named in the metadata file will not have file metadata associated with them.
Is File Grouped	A flag that indicates whether the documents are grouped by metadata file. **Note:** The value of this property is only meaningful under the following conditions: • You have selected **Metadata File** as the input file type. • You have selected **Values in the metadata file** in the configuration of the FSC Associate Metadata task. If these conditions are met, the value of this property is true if the checkbox **Group documents by metadata file** was checked in the configuration of the FSC Associate Metadata task. For some tasks, this information is used to group documents into ordered sets.
Is Metadata	A flag that indicates whether a file is a metadata file or a document file. Typically, this property is used in a rule that filters out metadata files so that they do not get archived.
Metadata File Path	The path to the directory where the metadata file is located.

Property	Description

Skipped Duplicates	The value of this property is only meaningful under the following conditions:
	• You have selected **Metadata File** as the input file type.
	• You have selected **Values in the metadata file** in the configuration of the FSC Associate Metadata task.
	• The value is associated with the metadata file.
	If these conditions are met, the value of this property indicates if documents that were named in the metadata file were skipped because they were either already being processed by the task routing service or because the metadata file named the same document twice.
	An example of the second case is a CSV file like:
	`important_doc.doc, Joe, joe@company.com`
	`important_doc.doc, Betty, betty@company.com`

This metadata is produced by the following task:

FSC Associate Metadata

Related reference

"FSC Associate Metadata" on page 531

FSR Create Document system metadata properties

The FSR Create Document metadata type contains metadata that is added to an item after is filed into a file system repository.

Property	Description
Document Class	The symbolic name of the document class used to capture an item into the file system.
Index File	A flag that indicates if one or more index files exist for this item.
Shortcut URL	A URL that references the item that has been captured into the file system.

This metadata is produced by the following task:

FSR Create Document

Related reference

"FSR Create Document" on page 536

HTTP URL system metadata properties

The HTTP URL metadata type is used to store URL specific data. An example of its usage is to pass URL file data to the P8 connector so that it can be used for adding content to a document.

The input for this metadata source is a URL in the format `http://host:port/path?query#fragment` as defined by the RFC standard.

Property	Description
RFC Fragment	The fragment identifier.
RFC Host	The IP address or the fully qualified domain name of the host server.
RFC Path	The URL path as defined by the RFC standard. It supplies the details of how the specified resource can be accessed.
RFC Port	The port number used for communication with the host server.
RFC Query	The query information.
RFC URL	The complete URL (absolute) as defined by the RFC standard.
URL Path Head	The portion of the URL path that includes the elements before the last slash (/).
URL Path Tail	The portion of the URL path that contains the element after the last slash (/). If the path ends with a slash, the path tail is an empty string.
URL Query Keys	The key-value pairs used as query information. The key-value pairs are delimited by the ampersand (&) and the equal sign (=), for example `key1=value1&key2=value2`.
URL Tail Base	The portion of the path tail that precedes a dot (.). If a period does not exist, the tail base is equivalent to the path tail.
URL Tail Extension	The portion of the path tail including and succeeding a dot (.). If a period does not exist, the tail extension is an empty string.
URL Target Path	The portion of the URL path up to but not including the query.

This metadata is produced by the following task:

SP Get Versions

Related reference

"SP Get Versions" on page 568

IBM Classification Module system metadata properties

Metadata that is generated by the IBM Classification Module. You can use this metadata in the task route to determine how documents will be processed.

Metadata type	Description
All Relevant Categories	List of top categories matched.
All Relevant Categories and Scores	Combined list of categories; scores.
All Relevant Scores	List of top scores.
Most Relevant Category	Winning category.
Most Relevant Score	Winning category score.
Decision plan results exported as XML	Decision plan results in XML format. Used only for decision plans (IBM InfoSphere Classification Module version 8.7).

This metadata is produced by the following task:

IBM Classification Module

Related reference

"IBM Classification Module" on page 537

P8 4.x Confirm Document system metadata properties

The P8 4.x Confirm Document system metadata type contains metadata about the status of the document in the repository.

Property	Description
Document Confirmed	A flag that indicates whether the document exists in the repository.
Shortcut URL	The shortcut URL of the document in the repository (if it exists).

This metadata is produced by the following task:

P8 4.x Confirm Document

Related reference

"P8 4.x Confirm Document" on page 540

P8 4.x Create Document system metadata properties

The P8 4.x Create Document metadata type contains metadata that is added to an item after it has been captured in the FileNet P8 repository.

Property	Description
Is Duplicate	A flag indicating if this item is a duplicate.
Object Class	The symbolic name of the document class used to capture an item into FileNet P8.
Object ID	The unique ID of the item captured into FileNet P8.
Object Store	The symbolic name of the FileNet P8 object store to which the document has been captured.
Shortcut URL	A URL that references the item that has been captured into FileNet P8.
Shortcut URL Mask	A mask for the URL that references the item that has been captured into FileNet P8.

This metadata is produced by the following tasks:

P8 4.x Create Content Elements

P8 4.x Create Document

P8 4.x Create Version Series

P8 4.x Save Prepared Text as XML

Related reference

"P8 4.x Create Content Elements" on page 541

"P8 4.x Create Document" on page 543

"P8 4.x Create Version Series" on page 550

"P8 4.x Save Prepared Text as XML" on page 561

P8 4.x Create Email Instance system metadata properties

The P8 4.x Create Email Instance metadata type contains metadata specific to a email document instance.

Property	Description
Object ID	The unique ID of the email document instance created in FileNet P8.

This metadata is produced by the following task:

P8 4.x Create Email Instance

Related reference

"P8 4.x Create Email Instance" on page 548

P8 4.x Declare Record system metadata properties

The P8 4.x Declare Record metadata type contains metadata that is added to an item after it is declared as a record in the FileNet P8 repository.

Property	Description
Object Class	The symbolic name of the record class used to declare an item as record in FileNet P8.
Object ID	The unique ID of the item declared as a record in FileNet P8.
Object Store	The symbolic name of the FileNet P8 object store in which the document has been declared as a record.

This metadata is produced by the following task:

P8 4.x Declare Record

Related reference

"P8 4.x Declare Record" on page 553

P8 4.x File Document in Folder system metadata properties

The P8 4.x File Document in Folder metadata type contains metadata that is added to an item after it is filed into a folder in the FileNet P8 repository.

Property	Description
File Paths	A list of folders in FileNet P8 that a document has been filed into.
Object Class	The symbolic name of the document class used to capture an item into FileNet P8.
Object ID	The unique ID of the item captured into FileNet P8.
Object Store	The symbolic name of the FileNet P8 object store the document has been captured to.

This metadata is produced by the following task:

P8 4.x File Document in Folder

Related reference

"P8 4.x File Document in Folder" on page 555

P8 4.x Link Documents system metadata properties

The P8 4.x Link Documents metadata type contains metadata specific to a link object that associates two or more separate FileNet P8 repository objects. For example, an email can be linked with its attachments.

Property	Description
Object Class	The symbolic name of the link class used to associate two or more items in FileNet P8.
Object ID	The unique ID of the link object used to associate two or more items in FileNet P8.
Object Store	The symbolic name of the FileNet P8 object store in which the link has been created.

This metadata is produced by the following task:

P8 4.x Link Documents

Related reference

"P8 4.x Link Documents" on page 557

P8 4.x Modify Object Security system metadata properties

The P8 4.x Modify Object Security metadata type contains metadata that is added to an item after permissions are assigned to it in the FileNet P8 repository.

Property	Description
Object Class	The symbolic name of the document class of the item that had its security modified in FileNet P8.
Object ID	The unique ID of the item that had its security modified in FileNet P8.
Object Store	The symbolic name of the FileNet P8 object store in which the item that had its security modified is saved.

This metadata is produced by the following task:

P8 4.x Modify Object Security

Related reference

"P8 4.x Modify Object Security" on page 558

P8 4.x Save Prepared Text as XML system metadata properties

The P8 4.x Save Prepared Text as XML metadata type contains the object ID of the XML document that was created by the P8 4.x Save Prepared Text as XML task.

Property	Description
Object ID	The unique ID of the item captured into FileNet P8.

This metadata is produced by the following task:

P8 4.x Save Prepared Text as XML

Related reference

"P8 4.x Save Prepared Text as XML" on page 561

Re-collection system metadata properties

The Re-collection metadata type contains metadata that is added after the document is collected by the SharePoint collector a second time.

Property	Description
Re-collection Flag	Indicates that the document was previously collected
Repository Type	The repository platform
Repository Name	The name of the target repository
Repository ID	The unique identifier of the document in the target repository
Repository Version Series ID	The unique identifier of the version series to which the re-collected document belongs
Repository Version Count	The number of collected versions of the document
Source Version Label	The version number of the document version within the source system

This metadata is produced by the following task:

SP Get Versions

Related reference

"SP Get Versions" on page 568

SP Blog system metadata properties

The SharePoint blog metadata type (**SP Blog**) contains metadata that is added to a blog post by a task in the task route and is available for use in rules or in other tasks before each version of the blog post and its comments is archived to the target repository.

Property	Description
Approval status	The approval status of the blog post version in SharePoint: • Approved • Draft • Rejected
Number of comments	The number of comments that the post contains
Post category	The SharePoint categories assigned to the post
Post content	The body of the post version, including its comments and embedded links
Post date	The last modified date of the post version
Post title	The title of the post version

This metadata is produced by the following tasks:

SP Collector

SP Get Versions

Related tasks

"Collecting from Microsoft SharePoint sites" on page 482

Related reference

"SP Get Versions" on page 568

SP Collection system metadata properties

The SharePoint metadata type (**SP Collection**) contains metadata that is added to a Microsoft SharePoint document by a task in the task route and is available for use in rules or in other tasks before the document is archived to the target repository.

Property	Description
Access Control List	The Access Control List (ACL) that determines who can access the document
Collection source	The source of the archived document as defined in the SP Collector (a GUID, not a name)
Content size	The size of the document, in bytes
Content type	The SharePoint content type of the document. If the content type of a document changes during its life cycle, IBM Content Collector applies the content type of the most recent version to the archived document.
Created	The document creation date
Created by (without domain)	The user name of the document creator
Created by	The user name and domain of the document creator **Tip:** To ensure correct rendering of the document creator name, use the property *Created by (without domain)*; use this property only if you are a IBM FileNet P8 user who edited the regedit file for use with a previous version of Content Collector
Folder path	The folder path of the document within the library
ID	The unique SharePoint identifier of the document
Last modified by (without domain)	The user name of the last user to modify the document
Last modified by	The user name and domain of the last user to modify the document **Tip:** To ensure correct rendering of the user name, use the property *Last modified by (without domain)*; use this property only if you are a IBM FileNet P8 user who edited the regedit file for use with a previous version of Content Collector

Property	Description
Last Version	Indicates that this document is the most recent version of the document
Library	The SharePoint library name
Library URL	The SharePoint library address
Minor Version	Indicates that the document is a minor version in SharePoint
Modified	The date and time on which the document was last modified
Name	The file name of the document
Permissions	A list of users or groups and their SharePoint rights to the document (includes specific permissions, not permission levels)
Site	The SharePoint site or subsite name
Site URL	The SharePoint site or subsite address
Title	The Title property value of a Microsoft Office document (can be empty)
Version	The version number of the document in SharePoint
Version Content Type	The SharePoint content type of each document version (can be empty)
Version Ordinal	An ordinal that the collector assigns to each document version, facilitating version ordering for tasks that require it. In many cases this number does not match the source system version number.

This metadata is produced by the following task:

SP Get Versions

Related reference

"SP Get Versions" on page 568

SP Create File system metadata properties

The SharePoint Create File (**SP Create File**) metadata type contains metadata that is added to the document by the SP Create File task in a task route when document hash mapping is enabled and is available for use in rules or in other tasks before the document is archived to the target repository.

Property	Description
Document Hash	A unique identifier that is generated for each document version, used if you have configured your repository connector to eliminate duplicate files using document hash mapping.

This metadata is produced by the following task:

SP Create File

Related reference

"SP Create File" on page 567

Task status system metadata properties

Each task updates the task status metadata with an indication of whether the task succeeded or failed.

Property	Description
Log Level	The log level that the task assigned to the message in Task Result.
Task Result	The error message if a task failed.
Task Success	A flag that indicates whether a task in the task route was completed successfully.

All tasks produce this metadata.

Text Extraction system metadata properties

Metadata that is created after a binary file has been converted to a text representation. The extracted text is used in a FileNet P8 environment when non-text formatted email attachments are stored.

Property	Description
Attachment Name	The name of the attachment as it appears in the email.
Is Entity an Attachment	A flag that indicates if the object is an attachment.
Number of Attachments	The number of attachments that were contained in the email.
Output Text File	The file that contains the extracted text from the original attachment. This is a temporary file and is cleaned up at the end of the task route.
Parent Entity ID	The ID of the originating email.

This metadata is produced by the following task:

Extract Text

Related reference

"Extract Text" on page 527

Windows Content Type Information system metadata properties

Metadata that provides information about the content of a file. This information is obtained from the Windows registry.

Property	Description
Icon Information	Icon information is used in the generated .url shortcut file. The .url file consists just of a file path and an icon index, and not the actual icon itself.
Mime Type	The representation protocol of the textual content in the file.

Configuring task routes

A task route defines a sequence of tasks and operations that perform discrete actions on documents or metadata. Task routes can be complex and offer several execution paths, so that different task sequences are carried out depending on the evaluation of one ore more rules. These rules use expressions and metadata to determine the execution path through a task route.

Each task route consists of a main task route and an error task route. A main task route contains:

- At least one collector
- One or more tasks
- **Optional:** Decision points and rules
- **Optional:** An audit log

When you create a task route, iconic representations of each processing step are added to a diagram, which serves as a visual summary of the task route setup. You can modify processing steps easily by clicking the corresponding icon in the diagram.

Related tasks

"Verifying and adjusting the initial configuration settings" on page 81

Task routes

A task route is a series of tasks to be performed on a document, most commonly to move it from an email server, a file server or some other source to a document repository. A task route can include rules that determine which task in the task route should be performed next.

To set up IBM Content Collector, you must configure one or more task routes. Every function that you want to use must be present in a task route, or it will not be carried out.

The easiest way to set up a task route is to use one of the task route templates and configure it to suit your requirements. There are task route templates for all email archiving scenarios and for the most typical document archiving setups. You can modify the task route templates and add or remove tasks to define just the series of tasks that you want IBM Content Collector to perform.

You can also configure a new task route from scratch, but this might lead to configuration errors and is thus not recommended.

Each task route consists of a main task route and an error task route. Documents are processed by the main task route. If an error occurs during the processing, the affected object is passed to the error task route.

Main task routes are made up of the following parts:

Collector

In a collector, you define where and when to find documents on the server.

For example, the File System Collector collects files from the file server as specified in the File System Collector configuration, and passes the files along the task route to be processed and perhaps added to the repository. An email collector collects email from specified locations on the email server as scheduled. It applies filtering rules to determine if the email

should be processed, and if so, passes the email along the task route to be processed and perhaps added to the repository.

Task Tasks define specific discrete actions that are carried out on the documents.

For example, a task can extract metadata from a document, save documents locally for processing, store them into the repository and so on. Usually, you will use a series of different tasks to process your documents.

Audit log
You can include one or more audit log tasks in a task route in order to log information about processing events. To log successful executions, you must include at least one audit log task to the main task route. To log unsuccessful executions, at least one audit log task must be added to the error task route.

Decision points and rules
You can insert a decision point in a task route to enable conditional processing of documents. At this point in a task route, the processing can go along different paths, depending on rules that you define. If a document that is being processed does not meet the criteria defined in the rules for the decision point, the document is dropped from the task route at that point and will not be processed further.

Related concepts

"Collecting documents for processing" on page 458

"Using decision points and rules" on page 281

Related tasks

"Working with task routes"

"Including an audit log task" on page 279

Working with task routes

Usually, you will import a task route template and adapt it to suit your requirements. You can then save the resulting task route as new template. Alternatively, you can also start from scratch with a new task route.

Working with task routes in the Configuration Manager

To create and edit task routes, you use the Configuration Manager. Watch the tour for an introduction to working with task routes in the information center.

Tip: You can use the controls at the bottom of each tour to control the speed of the tour. You can use the **Pause** button on the toolbar to study a particular section of the tour. You can also move the indicator on the time line to move forward or backward in the tour.

Working with task routes in the Configuration Manager
This 8 minute tour includes the following lessons:

- Getting help
- Importing a task route template
- Getting to know the parts of a task route
- Editing a task route
- Working with more than one task route at a time

The following topics contain the information that is presented in the viewlet in text form.

Tour: Getting help:

If you need help with the IBM Content Collector Configuration Manager at any time, click the **Help** button. This opens the IBM Content Collector information center.

Tour: Importing a task route template:

The easiest way to set up a task route is to import one of the supplied task route templates and adapt it to your needs. In the first part of the tour, you will learn how to import a task route template.

A task route defines a sequence of tasks and operations that are performed on documents or metadata. Typically, they define how to move documents from an email server, a file server or some other source to a document repository. Task routes can be complex and offer several execution paths, so that different task sequences are carried out depending on the evaluation of one or more rules. These rules use expressions and metadata to determine the execution path through a task route.

To import a task route template:

1. Click the **New** button. The New Task Route windows is displayed. In the **Choose a template** section, all task route templates in the default location are listed.
2. Select the task route template that you want to import.

   Tip: Alternatively, you can select a template bundle to import. A template bundle consists of several task route templates for typical archiving scenarios. The description of the selected template is displayed in the field below the task route templates.
3. Click **OK**.
4. If a dependency is listed in the **Resolve Dependencies** table (for example that a repository connection is required), check if this dependency is suitable for the environment in which you want to run the selected task route and click **OK**.

The task route template is loaded, and the task route is displayed in the design pane of the Configuration Manager.

Related tasks

"Creating a task route" on page 275

Tour: Getting to know the parts of a task route:

In the next part of the tour, you will learn about the parts of a task route: collectors, tasks, decision points and the audit log.

1. Click on the collector in the task route to see the collector settings. A collector interfaces with your source system. You can specify in the collector settings where and how often to collect documents. In the **Description** field in the configuration pane on the right, you can read the description of what the collector does.
2. Click the green start node to see the task route information again. In the **Description** field, you can read what the task route does.
3. Select the **Active** check box to set the task route active, or clear the check box to set it not active.

4. Click a task to get information about its configuration. Tasks define specific discrete actions that are carried out on the documents.
5. Click the **Show Help** button in the upper right corner of the configuration panel to go to the detailed information about the task in the information center.
6. Click an **i** button in the configuration pane to see additional information about a field.
7. Click a decision point to see which rules are defined. With a decision point, you can enable conditional processing of documents in a task route. At this point in a task route, processing can go along different paths, depending on the rules that you define.
8. To look at a rule, click the link that connects the decision point with the next task.

   Tip: When the link turns blue, you can click it.
9. To configure the rule, click the **Launch Expression Editor** button to launch the expression editor.
10. Include an audit log to record information about the status of each processed document. An audit log task in a task route monitors information about processing events.

Tour: Editing a task route:

In the next part of the tour, you will learn how to edit a task route.

Instead of importing task route templates, you can create task routes from scratch. However, using templates and adjusting them to your use case is easier and much quicker.

1. To delete a task from the task route, right-click the task and select **Delete**. Similarly, you can delete other elements of a task route, like collectors, decision points, or links.

   Important: You should never delete an element from a task route template if you are not absolutely sure about what it does.
2. In the Toolbox on the left, click the **+** sign to display the elements that you need to create a task route.

   is the symbol for a task.

   is the symbol for a collector.

   Tip: Instead of using the Toolbox, you can right-click in the desing pane to open a shortcut menu that offers the same functions. Alternatively, you can use keyboard shortcuts to perform all steps.
3. Select the element that you want to insert, for example a task.
4. Click the element on the toolbox and then click onto the task route on the designer area to set where the element should go. The link turns blue when the element is at a position where it can be inserted.
5. Release the left mouse button to insert the element.

   Tip: Alternatively, you can insert elements by right-clicking a link and selecting the element that you want to insert from the menu.

6. If you use specific tasks more than once in a task route or in more than one task route, right-click the tasks and select **Copy** to copy these tasks instead of creating the same tasks with the same configuration again.
7. Right-click anywhere on the canvas in the design pane and select **Paste** to insert the task. The task is inserted with a red **Disconnected** icon that indicates that the task is not yet part of the task route.
8. Drag the task to the place where you want to insert it. The task and the link turn blue when the task is at a position where it can be inserted.
9. Release the left mouse button to insert the task.
10. You can move tasks around as you want.
11. Right-click a link and insert a decision point. The red **Error** icon indicates that an element requires configuration. In this case, you must configure at least one rule.
12. Click the rule and enter a name in the configuration panel.
13. When you have finished configuring your task route, click the **Save** button to save your changes.

    Tip: An asterisk (*) behind the name of a task route in the Explorer shows that you have not saved your changes.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

Tour: Working with more than one task route at a time:

In the next part of the tour, you will learn how to work with more than one task route at a time.

Typically, you will have more than one active task route.

1. Click the **New** button to import more templates.
2. Select a task route template or a template bundle. The additional task routes are displayed in the Explorer.
3. Use the tabs to switch between the task routes.
4. After you have finished configuring the task routes, click **Save All** to save them.
5. After you have saved the task routes, you can close them by clicking the **x** buttons on the task route tabs.
6. Right-click a task route in the Explorer to set it active or not active. Inactive task routes are not run by the IBM Content Collector Task Routing Engine service.

   Tip: All saved task routes are listed in the Explorer. The icon next to the task route name indicates if a task route is active or not. Task routes that contain errors show a red **Error** icon.

   is the symbol for an active task route without errors.

   is the symbol for an active task route with errors.

   is the symbol for an inactive task route without errors.

   is the symbol for an inactive task route with errors.

7. To open a task route again, double-click the task route in the Explorer. The task route opens in the design pane again.
8. To delete a task route, right-click the task route in the Explorer, select **Delete**, and confirm.

Important aspects to consider when using email task routes

Before you begin to configure any of the email task routes, there are certain important aspects that must be considered.

When you are working with the email archiving task routes, email documents are not handled as one unit. All email is split up into multiple distinct parts, one part for the email body and one for each of the email attachments. The processing steps for the email body and its attachments are separate steps. However, although the processing steps are different, the complete email is archived, that is the email body and all attachments. Archiving only the email body, or only one or several attachments is not supported.

This is influenced by the underlying email storage data model in IBM Content Collector that stores the email body separately from the attachments in both IBM Content Manager and IBM FileNet P8. These separate processing steps are defined in tasks and, if the processing steps are more complex, in a separate execution path in the task route.

The supplied sample email task routes for archiving to IBM Content Collector only support compound resource item types. If you have bundled resource item types that you created in earlier Content Collector releases before the compound item type data model was introduced, you can only use these in the old task routes that supported bundled item types and were included in earlier Content Collector releases (all V2.1.0.x releases). You can enable these bundled resource item types for processing by the Content Collector indexer for text search if you want to stop using the fast indexer. After indexing using the fast indexer or the indexer for text search, you can search in these item types, and preview and restore email in the search result.

Item types for the document model BUNDLED and the archiving type ENTIRE (BRI) that you created in IBM CommonStore cannot be processed by any of the sample Content Collector task routes. These item types can only be enabled for processing by the indexer for text search in Content Collector. After indexing and, provided you have installed the IBM Content Collector Legacy support component, you can search in these item types, and preview and restore email in the search result.

Note that after you have enabled an item type for processing using the Content Collector indexer for text search, you cannot use the fast indexer on the item type.

Creating a task route

You can import a task route template and use it as a starting point for a new task route, or you can create task routes from scratch. Starting with existing templates is the quicker way in most cases.

To create a task route:

1. In the **Task Routes** view, click the **New** icon.
2. In the New Task Route window, choose whether you want to create a blank task route, import a task route template or import a template bundle.

- On the **Choose a template** tab, select **Blank task route** to create a blank task route. Type a name for the task route in the **Task route name** field.
- On the **Choose a template** tab, select an existing task route template to import it. All task routes saved as templates in the default location are listed. To see templates in another location, click the **...** button. The **Task route name** field is filled in automatically. You can change it as required.

  A task route template that you want to import can be in the following states:

  Structure only
  The saved template contains only the task placement within the task route and the connecting links, rules and decision points, but no task configuration or rule definitions. When you import a task route in this state, you need to edit all objects in the task route and provide appropriate configuration values.

  Structure and definition
  An existing task route is copied as is, including the values of objects in the task route. Before you import a task route in this state, you must configure connectors, repository connections, and so on appropriately. After importing the task route, it is sufficient to modify the configuration values of the task route objects as needed.

  Include environment specific configuration
  The existing task route is copied with the configuration values for the objects it contains and additionally with configuration data defined outside of the task route, such as information related to the connectors, repository connections, user-defined metadata, and so on. However, when this flag is set, any environment specific data, such as local folder paths and user names, is removed from the template configuration before it is exported. Still, setting this flag involves the least configuration effort after importing the task route as the configuration data fits the environment it is imported into overall.

  Restriction: Although you may have installed IBM Content Collector for use with Lotus Domino and Microsoft Exchange as source systems at the same time, you can configure the Email Connector for one source system only, that is, either for Lotus Domino or Microsoft Exchange. Independent of the current configuration, you can import any email-specific task route template. However, you can only configure and use those task routes properly that correspond to the current configuration of the Email Connector.
-

  Email only: On the **Choose a template bundle** tab, select an existing template bundle. Template bundles consist of several task route templates for a specific archiving scenario.

3. Click **OK**. If a dependency is listed in the **Resolve Dependencies** table, check if this dependency is suitable for the environment in which you want to run the selected task route. The task route that you selected is displayed.
4. In the **Description** field of the configuration pane on the right, you can edit the description for the task route.
5. If you created a blank task route, add one or more collectors to the task route and configure them. Otherwise, check the configuration of the collector or collectors.

6. If you created a blank task route, add one ore more tasks to the task route and configure them. Otherwise, check the configuration of the tasks. All tasks must be placed between the start node and the end node .
7. Save the task route.

Related concepts

"Collecting documents for processing" on page 458

Related tasks

"Tour: Editing a task route" on page 273

"Tips to ensure proper placement of tasks in a task route" on page 278

"Moving documents off the network into IBM FileNet P8" on page 725

"Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records" on page 726

"Defining metadata to be used to process files for archiving" on page 729

Related reference

"Keyboard commands for the Configuration Manager" on page 160

Exporting task routes as templates

When you export a task route as a template, the structure with or without the configuration can be saved so that the task route can be used as a starting point for other task routes.

Make sure that the task route that you want to export does not have unsaved changes.

To export a task route as a template:

1. Start the Configuration Manager and click **Task Routes**.
2. Double-click an existing task route in the **Explorer** on the upper left.
3. Click the **Export task route** icon.
4. In the Export window:
   a. If you want to save the template to a location other than the default, click the **...** button to select a folder.
   b. In the **File name** text box, enter a name for the template.
5. Select an export option. You can select between the following options:
   - **Structure only**
   - **Structure and definition**
   - With **Structure and definition**, you can additionally select **Include environment specific configuration**

   Here is a description of these options:

   Structure only
   The saved template contains only the task placement within the task route and the connecting links, rules and decision points, but no task configuration or rule definitions. When you import a task route in this state, you need to edit all objects in the task route and provide appropriate configuration values.

   Structure and definition
   An existing task route is copied as is, including the values of objects in the task route. Before you import a task route in this state, you must configure connectors, repository connections, and so on appropriately.

After importing the task route, it is sufficient to modify the configuration values of the task route objects as needed.

Include environment specific configuration
The existing task route is copied with the configuration values for the objects it contains and additionally with configuration data defined outside of the task route, such as information related to the connectors, repository connections, user-defined metadata, and so on. However, when this flag is set, any environment specific data, such as local folder paths and user names, is removed from the template configuration before it is exported. Still, setting this flag involves the least configuration effort after importing the task route as the configuration data fits the environment it is imported into overall.

6. Click **OK**.

Tips to ensure proper placement of tasks in a task route

Follow the guidelines in this section when you are first starting to create task routes, and you will soon find you can click and drag to place tasks with little trouble.

Tip: You will know a task is properly placed in a task route when the **Disconnected** icon disappears from the task, an arrow connects the previous task to the task you are placing, and the configuration pane displays configuration options for the task.

About icons

The icons that appear when configuring task routes are described below.

Use one of the following methods to ensure that a task is properly placed in a task route:

- Select the task from the **Toolbox** section.
  1. In the **Toolbox** section of the Configuration Manager, click to select the task.
  2. Click in designer pane. If you clicked on a link in the designer pane, the task is automatically inserted into the task route. Otherwise the task appears with a red **Disconnected** icon indicating the task is not yet part of the task route. In this case, click the task and drag it into the task route.
- Select the task from the shortcut menu.
  1. Right-click on a link in the task route. The shortcut menu opens.
  2. Select the task that you want to insert. The task is automatically placed in the task route.

The task and the route connectors above and below will become blue when the task is at a possible insertion point. Release the left mouse button or press space to drop the task into the task route.

Tasks that are part of a task route and require configuration appear with a red **Error** icon ( ).

If a task is missing required metadata a **Metadata Error** icon will appear. Right-click on the task and select **Properties**, which contains summary information about the task including information regarding the missing metadata.

Copying tasks

If you use specific tasks more than once in a task route or in different task routes, you can copy the tasks with the configuration information instead of creating new tasks and inserting the configuration information again. You can copy and paste tasks between the Main and Error task route configurations, as well as between two different task routes.

Configure the task that you want to copy.

To re-use the task with the same configuration information:

1. Right-click the task and select **Copy**.
2. Right-click the place where you want to insert the task and select **Paste**. To copy a rule, you must select a decision point first and then paste the rule.

The task is added with the same configuration information as the original task.

Including an audit log task

You can include audit log tasks in a task route in order to log whether a document was processed successfully, for example, or to track generated metadata.

1. Open a task route in the **Task Routes** view of the Configuration Manager.
2. Click **Audit Log** in the Toolbox. Then click on a link in the task route to insert an audit log task. Alternatively, you can right-click on the link and select **Add Audit Log** from the menu.
3. Select the icon of the audit log task that appears in your task route.
4. Under **Audit log file name format**, specify the log file name for the audit log task:

   Use default
   Use the default audit log file name. This name is equal for all audit log tasks that you add, so if you select this option, all audit log entries are written to one file.

   Specify file identifier
   Specify an identifier that is added to the file name. If you specify unique identifiers for all audit log tasks that you add, the audit log entries for each audit log task will be written to a unique file.

5. Select how the data is output:

   Field delimiter
   The delimiter to be used to separate fields in the log file.

   Value delimiter
   If a field in the log file contains multiple values, select the delimiter to be used to separate each value in the column. This delimiter must not be the same as the field delimiter.

   Quote character
   Select the character that you want to replace quote characters with. This is to prevent that a quotation character is misinterpreted as a field or value delimiter, which would impede the reading of the log file or the proper display of columns or values in an editor. The double-quote character, for example, is often used as the value delimiter in comma-separated variable (CSV) files. It would be useful, therefore, to replace all quote characters in the data with single quotes.

6. In the selection box that appears under the check box, select the metadata properties that you want to log. The box lists all metadata sources that you might wish to capture as part of the audit. To log a property, select the plus sign next to the check box.

   Tip: Click **Show/hide IDs** to display the IDs of the metadata properties. These IDs are displayed in the audit log header to define the columns of the audit log.
7. Save the audit log task to the task route.

   SharePoint only: Under version series processing, each version contains its own entry in the audit log, unless a decision point specifies only the last version or some other versioning restriction prevents routing to the audit log task.

Related tasks

"Creating a task route" on page 275

Configuring the error task route

For each task route, an error task route is created automatically. When an error occurs in the task route, the affected object is passed to the error task route.

1. Open a task route in the **Task Routes** view of the Configuration Manager.
2. Click the **Switch between main and error task route** icon to switch to the error task route. The text in the task route toolbar tells you whether you are currently looking at the main task route or at the error task route.
3. Add an audit log task to the error task route and configure it. All error task routes should contain an audit log task, so that errors in task route processing are logged. Before IBM Content Collector V2.2, the audit log for a task route was shared between the main task route and the error task route. Starting with IBM Content Collector V2.2, the two audit logs are independent and must be configured separately, which makes it possible to include different logging information in the audit logs of the main task route and the error task route.

   Note: If you import a task route that contains an audit log and was exported with a previous version of IBM Content Collector, the error task route will not contain an audit log. You must add one manually.
4. Optional: Configure the error task route to record information about problematic items or to make copies of the problematic items, so that they can be analyzed. You can add tasks, decision points, and rules in the same way as for the main task route.

   Tip: When configuring the error task route, consider the following tips:
   - Include an audit log task as the first task in the error task route to record the task status of the task that failed.
   - When you add a task, make sure that all required metadata for this task is available. To do so, add a decision point and a rule that checks if the metadata exists.
   - To make a copy of the item that caused the error, add a task that saves the item, for example Save Temporary File Copy or FSR Create Document.
5. Save the task route. Saving the task route saves the error task route and the main task route.

Related tasks

"Including an audit log task" on page 279

"Tips to ensure proper placement of tasks in a task route" on page 278

Using decision points and rules

Decision points are used with rules to conditionally process documents. They allow the user to define one or more rules in priority evaluation order. If a rule returns true, the task route processing continues with the task that follows this rule. If a rule returns false and there is another rule in the decision point, that rule is evaluated. If there are no more rules and the last rule returns false, processing skips to the end of the task route.

About decision points

You must insert a decision point in a task route to enable conditional processing of documents. Decision points can be inserted anywhere in the task route (except following other decision points).

About rules

You must add and configure rules following a decision point to determine what path in a task route will be followed to process a document.

If a rule returns true, the document moving through the task route is processed by the task that follows the rule. If it returns false, the next rule is evaluated. You can use an *Always True* rule as a catch all when all other rules fail evaluation. If you do not use an *Always true* rule and all rules return false, the processing of the document passes immediately to the end of the task route and stops. In this case, any Audit Log task that is included in the task after the decision point will not record a result for this document.

The order of rule evaluation that is defined in the decision point is crucial. To change the order, click the corresponding decision point. Change the evaluation order in the configuration pane of the decision point. In the task route designer pane, the rule name is preceded with an integer value that indicates its evaluation order.

If you do not apply any rules to a task route, tasks are joined together with links. Links represent a straight processing path through a task route with no alternative directions.

Adding decision points:

A decision point allows for conditional processing of documents.

You must insert a decision point in a task route to add and configure a rule. Once you have inserted a decision point and added rules, you can set the order in which the rules are evaluated.

- To add a decision point:
  1. In the explore pane of the Task Route Designer, under **Toolbox**, click **Decision Point**. Alternatively, right-click on the link location and select **Add Decision Point** from the shortcut menu to create a decision point.
  2. In the design pane, click to place the decision point and move it into the task route. Read tips to ensure proper placement.
  3. Follow these steps to add one or more rules. See Adding or editing a rule.

- To set the order in which rules are evaluated:
  1. In the design pane, click the decision point.
  2. In the configuration pane, a list of rules is displayed. Select a rule to reorder.
  3. Use the arrows to move the rule up or down in the evaluation order.

Related tasks

"Defining metadata to be used to process files for archiving" on page 729

Adding or editing a rule:

Rules determine the processing path to take along a task route.

To add a rule, you must first add a decision point. Decision points work with rules to allow conditional processing of files. Decision points denote rule processing order.

A rule is made up of a Boolean expression, which results in a value of true or false when evaluated. The default clause for a newly created rule is to return true. This is known as an *Always True* rule.

For conditional processing, you configure advanced evaluation criteria for the rule. An advanced expression for a rule consists of at least one operator and its required operands, which is the *root expression*. You can also nest expressions making up an expression tree to build comprehensive evaluation criteria.

The root expression for a rule always returns a Boolean true or false when evaluated.

To add or edit a rule:

1. Add a decision point.
2. To add a new rule, click **Link** in the toolbox, place the link in the design pane and move it into the task route, connecting to the decision point. Alternatively, right-click the decision point and select **Add Rule** from the shortcut menu. Once connected to a decision point, a link becomes a rule that can be configured.
3. In the configuration pane, enter a name and a description for the rule.
4. In the **Configure Rule** section, select the type of rule that you want to configure:
   - If you want to capture all documents in a collection source that you have specified, select **Always True**. If you use an *Always true* rule as first rule, this is equivalent to a link. *Always true* rules can be helpful as last rule in a rule set, serving as a catch-all rule.

     Note: A collection source is where a collector looks to find documents to process. You define this when you configure a collector. For example, this could be a folder, or a PST file, or a group of mailboxes, and so on.

     Tip: When you are first testing your configuration of this application, selecting **Always True** allows you to quickly see if your documents can be processed.
   - If you want to conditionally process documents, click **Advanced** and launch the Expression Editor by clicking the button to the right of the box showing the expression tree ( ).

5. Use the Expression Editor to build or adapt the expression that will serve as evaluation criteria for the rule. See the topic on editing expressions for more information about the Expression Editor.

Note:

After adding and configuring each rule, set the order in which rules are evaluated as follows:

1. In the design pane, click the decision point.
2. In the configuration pane, the list of rules is displayed. Select a rule to reorder.
3. Use the arrows to move the rule up or down in the evaluation order.

Related concepts

"Regular expressions" on page 421

Related tasks

"Adding decision points" on page 281

"Working with the Expression Editor" on page 404

Related reference

"Prototype expressions for rules and dynamic detection of document or record classes" on page 406

Configuring task routes from templates

You should use task route templates as a starting point for configuring your own task routes. The following topics document the task route templates that are delivered with IBM Content Collector and can be configured to suit your requirements.

Content Collector task route templates

IBM Content Collector provides sample task route templates for all typical document archiving setups. A task route is a series of tasks to be performed on a document, most commonly to move it from an email server, a file server, or some other source to a document repository. A task route can include rules that determine which task in the task route should be performed next.

To use any of the IBM Content Collector archiving functionality, you must configure one or more task routes. Any IBM Content Collector archiving task that you want to use must be defined in a task route, or it will not be carried out.

Each sample template typically defines an explicit task that is carried out on a document, for example, automatic archiving, stubbing, or deletion. This means that you would normally operate with several configured task route templates to accommodate your mail management requirements.

The easiest and recommended way to set up task routes is to import the provided sample task route templates and configure them to suit your requirements. The provided templates differ mainly with respect to defined collector schedule, the collection sources, filter settings, and stubbing and restubbing settings.

You can save an incomplete task route. This means that you can save a task route even if you have not configured all of a task's settings. However, the IBM Content Collector Task Routing Engine service will not start the task route until it is complete.

The collector and task names often reflect the way the collector or task was configured in the provided template although there are more configuration options available for selection. For example, the email stubbing task `EC Create Email Stub` is sometimes called `Remove Attachments`, `Remove Attachments and Body`, or `Delete Email` depending on the stubbing options selected in the stubbing task. If you change the stubbing options in an existing task, the name of the task may no longer reflect what the task actually does.

You can modify configuration settings in the tasks, for example, add collector filter settings or change stubbing definitions, but you are advised to not remove any tasks in any of the provided templates without IBM Support.

When you are configuring task routes, consider the following:

- Ensure that the collection sources in the templates are mutually exclusive
- Never set the collector schedule to **Always** in a production environment
- For testing purposes, use **Once** in the automatic archiving templates and **Always** in the interactive archiving template
- Ensure the following before using the File System templates:
  1. The templates that associate metadata can only be configured if a File System Source Connector metadata source was defined
  2. The File System Source Collector filter setting that determines which type of file to filter out should be configure in a reciprocal manner to the associate metadata settings in the `FSC Associate Metadata` task
  3. The shortcut links in the Create Document tasks must only be modified to point to the service that will provide access to the content. Typically this means adjusting the host and port portions of the URL. The rest of the link must not be changed.
- Always include the audit log in each task route to help review and analyze problems:
  - Select **Tab** as the field delimiter in the audit log
  - Include metadata that was used in rules to validate decision
  - Use the audit log to learn how metadata is used in the archiving or subbing process

Order of tasks in a task route

Task routes and task route templates typically contain common task sequences that vary by source system, repository, or main action.

Email task routes

Stubbing task routes perform tasks on the email system. They do not require interaction with the target repository, so they are independent of the repository system. There are only minor differences between the stubbing task routes for Microsoft Exchange and Lotus Domino.

Each task route needs one or more collection sources.

A stubbing task route needs only one task, namely EC Create Email Stub. This task stubs or deletes the email according to the configuration. You can also use the EC Extract Metadata task to extract metadata from the email. You can then use this metadata in one or more rules to decide on a specific stubbing action.

An archiving task route for email documents usually consists of at least the following tasks in the given order:

For Content Manager

1. EC Extract Metadata
2. Calculate Expiration Date
3. **Optional:** MC Retrieve Additional Metadata (This task is only required if additional archiving information should be obtained. The task is only used in interactive archiving task routes.)
4. CM 8.x Configure Item Types
5. **Optional:** CM 8.x Duplicate Detection (This task should be followed by the task CM 8.x Update Document for documents that are duplicates.)
6. EC Prepare Email for Archiving
7. **Optional:** IBM Classification Module
8. EC Extract Attachments
9. CM 8.x Associate Content
10. EC Prepare Email for Stubbing
11. **Optional:** EC Create Email Stub

For FileNet P8

1. EC Extract Metadata
2. Calculate Expiration Date
3. **Optional:** MC Retrieve Additional Metadata (This task is only required if additional archiving information should be obtained. The task is only used in interactive archiving task routes.)
4. P8 4.x Create Document
5. EC Prepare Email for Archiving
6. EC Extract Attachments
7. P8 4.x Create Content Elements
8. P8 4.x Create Email Instance
9. Extract Text
10. P8 4.x Save Prepared Text as XML
11. EC Prepare Email for Stubbing
12. **Optional:** EC Create Email Stub

SMTP task routes

There is only one basic compliance task route that archives email from the message queue directory that is defined in the SMTP connector automatically on a regular basis, detecting duplicates and removing attachments in email immediately.

Microsoft SharePoint task routes

SharePoint task route templates come in two flavors: one processes a single document version and the other processes multiple document versions, with a set for each repository. In order, each generally contains the SP Collector, a task that creates the documents in the repository, and typically the SP Post-processing task. The version task routes also include SP Get Versions and a task that creates the versions in the repository. The Content Manager routes also immediately follow the collector with a task that configures item types.

File System task routes

File System task route templates differ with respect to whether to associate metadata with the archived content and what post-processing chores to carry out: delete, replace with shortcuts (links), declare records or archive in a specified folder structure. The order is generally the same: the collector, a task that creates documents, then the post-processing task.

The task routes that deal with metadata typically add the FSC Associate Metadata task immediately after the collector.

Template bundles for Lotus Domino

Use the supplied sample archiving template bundles as a basis for configuring task routes if you are archiving from a Lotus Domino mailbox to an IBM Content Manager or a FileNet P8 repository.

The template bundles group together sample task routes that exist for an archiving scenario. When you select to import a template bundle to use in the IBM Content Collector Configuration Manager, all of the task routes that belong in the bundle are imported and loaded into the design pane of the Configuration Manager where you can view the task routes and begin to adapt settings to meet your needs.

Select which of the supplied archiving scenarios fit your use case:

Table 47. Typical archiving scenarios and matching templates

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Application document archiving: • Automatically archive documents from Lotus Domino applications, for example workflow documents, regularly. Only documents that are complete and have not been changed over a defined time period will be archived.	• IBM Content Manager • FileNet P8	• `CM - Lotus Domino Application Archiving template` • `P8 - Lotus Domino Application Archiving template`	• "Application Archiving" on page 317
Business process management (BPM) archiving: • Regularly archive email from monitored folders automatically into a specified business process management (BPM) folder to support searching and viewing these documents. Email is not stubbed in the mailbox and cannot be restored.	• FileNet P8	• `Business Process Management for Lotus Domino template`	• "BPM (Business Process Management)" on page 320

Table 47. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Common archiving setup: • Automatically archive all email that has been on a server for a defined period of time • Remove email content in defined stages to save space on the mail server • Archive email based on interactive archiving requests. When a user marks email for archiving, an archiving request is sent to a defined trigger mailbox. The archiving requests are collected and the marked email documents are archived, detecting duplicates, and removing attachments from the email in the mailbox.	• IBM Content Manager • FileNet P8	• `CM - Default Archiving for Lotus Domino template` • `P8 - Default Archiving for Lotus Domino template`	• "Default Archiving (Automatic)" on page 341 • "Default Archiving (Stubbing)" on page 306 • "Default Archiving (Stubbing)" on page 325 • "Default Archiving (Interactive)" on page 342
Archiving in a complex organization: • Automatically archive email older than a specified time to support users working in an organization with locations in different time zones. This task route does not contain a stubbing task. • Remove email content in defined stages. The task route also considers that users might need access to their email when they are working offline.	• IBM Content Manager • FileNet P8	• `CM - Complex Organization for Lotus Domino template` • `P8 - Complex Organization for Lotus Domino template`	• "Complex Organization (Archiving)" on page 339 • "Complex Organization (Stubbing)" on page 303 • "Complex Organization (Stubbing)" on page 322

Table 47. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Archiving with emphasis on space saving: • Archive email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed. • Archive large email only. Attachments are removed and replaced with links when the email documents are archived. • To support saving space on a mail server, remove email content after the email documents have been archived. To supports email users working offline for longer periods of time, automatic restubbing is disabled.	• IBM Content Manager • FileNet P8	• `CM - Space Saving for Lotus Domino template` • `P8 - Space Saving for Lotus Domino template`	• "Space Saving (All Email)" on page 349 • "Space Saving (Large Attachments)" on page 350 • "Space Saving (Stubbing)" on page 313 • "Space Saving (Stubbing)" on page 332

Table 47. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Archiving journal email: • Archive all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it has been archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager. • Archive all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• IBM Content Manager • FileNet P8	• `CM - Journal Archiving for Lotus Domino template` • `P8 - Journal Archiving for Lotus Domino template`	• "Journal Archiving (Email Deleted)" on page 344 • "Journal Archiving (Email Retained)" on page 364
Archiving documents in local Lotus Domino databases: • Archive documents from local Lotus Domino databases (NSF files). The task route empties the NSF files (deletes all documents in these files) as soon as the documents have been archived.	• IBM Content Manager • FileNet P8	• `CM - Lotus Domino NSF Migration template` • `P8 - Lotus Domino NSF Migration template`	• "NSF Archive Migration - Archiving" on page 310 • "NSF Archive Migration - Archiving" on page 329

Table 47. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Common archiving setup for former users of CommonStore for Lotus Domino: • Archive all email of a given size that have been in the mailbox for a defined period of time. The archiving schedule is limited to mailboxes that exceed a given size. The task route detects duplicates and deletes all attachments after the email has been archived. • Remove archived email content in defined stages from the mailbox and stubs restored email in the mailbox. • Archive email that a user marked for archiving. The task route contains all the tasks necessary for interactive archiving, such as collecting email, preparing it for storage in the repository, and stubbing it.	• IBM Content Manager	• `CSLD Migration template`	• "CSLD Migration - Automatic Archiving" on page 399 • "CSLD Migration - Automatic Stubbing" on page 400 • "CSLD Migration - Interactive Archiving" on page 401

Table 47. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Archiving using an offline repository for former users of CommonStore for Lotus Domino: • Automatically archive email by using a delayed archiving schedule. This task route is designed to support working offline. • Automatically remove email content in defined stages to support working offline. It removes email content based on the client setup. To save space, the attachments and the body are removed from all email after users have copied them to a local repository. Regardless of whether email was copied to a local repository, attachments are removed 30 days after the email was archived.	• IBM Content Manager	• `CSLD Migration for Offline Repository template`	• "CSLD Migration - Offline Repository Support (Archiving)" on page 402 • "CSLD Migration - Offline Repository Support (Stubbing)" on page 403

Template bundles for Microsoft Exchange mailbox users

Use the supplied sample archiving template bundles as a basis for configuring task routes if you are archiving from a Microsoft Exchange mailbox to an IBM Content Manager or a FileNet P8 repository.

The template bundles group together sample task routes that exist for an archiving scenario. When you select to import a template bundle to use in the IBM Content Collector Configuration Manager, all of the task routes that belong in the bundle are imported and loaded into the design pane of the Configuration Manager where you can view the task routes and begin to adapt settings to meet your needs.

Select which of the supplied archiving scenarios fit your use case:

Table 48. Typical archiving scenarios and matching templates

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Business process management (BPM) archiving: • Regularly archive email from monitored folders automatically into a specified business process management (BPM) folder to support searching and viewing these documents. Email is not stubbed in the mailbox and cannot be restored.	• FileNet P8	• `Business Pprocess Managaement for MS Exchange template`	• "BPM (Business Process Management)" on page 357
Archive messages from managed folders: • Automatically archive email in folders managed by Microsoft Exchange removing all email attachments immediately. The task route allows setting a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• IBM Content Manager • FileNet P8	• `CM - MS Exchange Managed Folder Archiving template` • `P8 - MS Exchange Managed Folder Archiving template`	• "Archiving Managed Folders" on page 338
Archive messages in personal folders: • Archive messages from personal folders (PST files) automatically. The task route removes the attachments and the body of the messages as soon as the messages are archived. The resulting stubs are moved to the Microsoft Exchange mailboxes of the PST file owners, leaving the PST files empty.	• IBM Content Manager • FileNet P8	• `CM - MS Exchange PST Migration template` • `P8 - MS Exchange PST Migration template`	• "PST Migration - Archiving" on page 347

Table 48. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Common archiving setup: • Automatically archive all email that has been on a server for a defined period of time. • Remove email content in defined stages to save space on the mail server. • Archive email based on interactive archiving requests. When a user marks email for archiving, an archiving request is sent to a defined trigger mailbox. The archiving requests are collected and the marked email documents are archived, detecting duplicates, and removing attachments from the email in the mailbox.	• IBM Content Manager • FileNet P8	• `CM - Default Archiving for MS Exchange template` • `P8 - Default Archiving for MS Exchange template`	• "Default Archiving (Automatic)" on page 341 • "Default Archiving (Stubbing)" on page 362 • "Default Archiving (Stubbing)" on page 343 • "Default Archiving (Interactive)" on page 342
Archiving in a complex organization: • Automatically archive email older than a specified time to support users working in an organization with locations in different time zones. This task route does not contain a stubbing task. • Remove email content in defined stages. The task route also considers that users might need access to their email when they are working offline.	• IBM Content Manager • FileNet P8	• `CM - Complex Organization for MS Exchange template` • `P8 - Complex Organization for MS Exchange template`	• "Complex Organization (Archiving)" on page 339 • "Complex Organization (Stubbing)" on page 341 • "Complex Organization (Stubbing)" on page 360

Table 48. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Archiving with emphasis on space saving: • Archive email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed. • Archive large email only. Attachments are removed and replaced with links when the email documents are archived. • To support saving space on a mail server, remove email content after the email documents have been archived. To supports email users working offline for longer periods of time, automatic restubbing is disabled.	• IBM Content Manager • FileNet P8	• `CM - Space Saving for MS Exchange template` • `P8 - Space Saving for MS Exchange template`	• "Space Saving (All Email)" on page 349 • "Space Saving (Large Attachments)" on page 350 • "Space Saving (Stubbing)" on page 351 • "Space Saving (Stubbing)" on page 370

Table 48. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Archiving journal email: • Archive all email in a Microsoft Exchange journal recipient mailbox automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal mailbox immediately after it has been archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager. • Archive all email in a Microsoft Exchange journal recipient mailbox automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal mailbox. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• IBM Content Manager • FileNet P8	• `CM - Journal Archiving for MS Exchange template` • `P8 - Journal Archiving for MS Exchange template`	• "Journal Archiving (Email Deleted)" on page 344 • "Journal Archiving (Email Retained)" on page 364

Table 48. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Supported repository	Name of template bundle to use	Included task routes
Common archiving setup for former users of CommonStore for Exchange Server: • Archive all email of a given size that have been in the mailbox for a defined period of time. The archiving schedule is limited to mailboxes that exceed a given size. The task route detects duplicates and deletes all attachments after the email has been archived. • Remove archived email content in defined stages from the mailbox and stubs restored email in the mailbox. • Archive email that a user marked for archiving. The task route contains all the tasks necessary for interactive archiving, such as collecting email, preparing it for storage in the repository, and stubbing it.	• IBM Content Manager	• `CSX Migration template`	• "CSX Migration - Automatic Archiving" on page 395 • "CSX Migration - Automatic Stubbing" on page 396 • "CSX Migration - Interactive Archiving" on page 397
Archiving using an offline repository for former users of CommonStore for Exchange Server: • Automatically archive email by using a delayed archiving schedule. This task route is designed to support working offline. • Automatically remove email content in defined stages to support working offline. It removes email content based on the client setup. To save space, the attachments and the body are removed from all email after users have copied them to a local repository. Regardless of whether email was copied to a local repository, attachments are removed 30 days after the email was archived.	• IBM Content Manager	• `CSX Migration for Offline Repository template`	• "CSX Migration - Offline Repository Support (Archiving)" on page 398 • "CSX Migration - Offline Repository Support (Stubbing)" on page 399

Lotus Domino with IBM Content Manager archiving templates

Use these sample archiving templates as a basis for configuring task routes if you are archiving from a Lotus Domino mailbox to a IBM Content Manager repository.

Select which of the supplied archiving templates fit your archiving user scenario:

Table 49. Typical archiving scenarios and matching templates

Typical archiving scenarios	Template files to use
Application document archiving: • This task route automatically archives documents from Lotus Domino applications, for example workflow documents, regularly if the documents are complete and have not been changed during a defined time period.	• "Application Archiving" on page 299
Common archiving setup: • This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the `Default Archiving (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Default Archiving (automatic)` task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed. • This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.	• "Default Archiving (Automatic)" on page 304 • "Default Archiving (Stubbing)" on page 306 • "Default Archiving (Interactive)" on page 305
Basic archiving setup with simple stubbing: • This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately. This task route can be used as a starting point for new task routes for email archiving.	• "Archiving" on page 300

Table 49. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Space saving on the mail server after archiving or restoring: • This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).	• "Life Cycle Template" on page 309
Archiving in a complex organization: • This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the `Complex Organizations (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Complex Organization (Archiving)` task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline.	• "Complex Organization (Archiving)" on page 302 • "Complex Organization (Stubbing)" on page 303
Archiving with emphasis on space saving: • This is one of three available task routes that emphasizes saving space on the mail server during email archiving. This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed. • This is another available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived. • This space saving task route emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.	• "Space Saving (All Email)" on page 311 • "Space Saving (Large Attachments)" on page 312 • "Space Saving (Stubbing)" on page 313

Table 49. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archiving journal email: • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager. • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• "Journal Archiving (Email Deleted)" on page 307 • "Journal Archiving (Email Retained)" on page 308
Archiving documents in Lotus Domino databases: • This task route automatically archive documents from Lotus Domino databases (NSF files). The task route empties the NSF files (deletes all documents in these files) as soon as the documents have been archived.	• "NSF Archive Migration - Archiving" on page 310
Collecting statistics on mailbox content: • This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.	• "Mailbox Statistics Collection" on page 310
Archiving calendar entries: • This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry. The task route detects duplicates and archives the same calendar entry only once.	• "Archiving Calendar Entries" on page 301

Application Archiving:

This task route automatically archives documents from Lotus Notes applications, for example workflow documents, regularly if the documents are complete and have not been changed during a defined time period.

Collectors

This template uses the following collectors:

EC Collect Documents from Application If Unchanged for at Least 30 Days and DocStatus OK
This collector collects documents from an application that is specified on the Collection Sources page. A document is collected if the value of the DocStatus field is "OK" and the document has not been changed during the last 30 days.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the document. All texts or attachment links created during stubbing are inserted in the rich text field specified in the Stubbing Options section.

File name

The file name of this template is `CM_LD Application Archiving.ctms`.

Archiving:

This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately.

This task route can be used as a starting point for new task routes for email archiving.

Collectors

This template uses the following collectors:

EC Collect All Email
This collector collects all email.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CM_LD - Archiving Template.ctms`.

Archiving Calendar Entries:

This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry.

The task route detects duplicates and archives the same calendar entry only once.

Collectors

This template uses the following collectors:

EC Collect Calendar Entries That Completed 30 Days Ago
This collector collects calendar entries that completed at least thirty days ago.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_LD - Archiving Calendar Entries.ctms`.

Complex Organization (Archiving):

This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the Complex Organizations (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 70 Days (New York)
This collector collects email that is older than 70 days from the mailboxes on a server in New York.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

EC Collect Email Older Than 70 Days (Paris)
This collector collects email that is older than 70 days from the mailboxes on a server in Paris.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_LD_4.1 - Complex Organization (archiving).ctms`.

Complex Organization (Stubbing):

This task route is designed to work with the Complex Organization (Archiving) task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline. There are two EC Create Email Stub tasks that deal with email after it was archived. The Remove Attachments and Body task removes the attachments and body after the email was copied to a local repository to save space on the mail server. The Use Life Cycle Collector task deals with email that exceeded the time limit for copying to a local repository.

Collectors

This template uses the following collectors:

Collect from Mailboxes on Servers in Paris and New York
This collector collects archived or restored email from mailboxes on servers in Paris and New York.

When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email 20 days after the email was archived. Restored email is restubbed after two months.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

Remove Attachments and Body
This task removes the attachments and the body from the email.

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_LD_4.2 - Complex Organization (stubbing).ctms`.

Default Archiving (Automatic):

This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the Default Archiving (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_LD_1.1 - Default Archiving (automatic).ctms`.

Default Archiving (Interactive):

This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.

Collectors

This template uses the following collectors:

EC Collect Email Based on Interactive Requests
This collector collects email that the users marked for archiving from the mailboxes.

Specify the Domino server and database path of the trigger mailbox on the Collection Sources tab.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CM_LD_1.3 - Default Archiving (interactive).ctms`.

Default Archiving (Stubbing):

This task route is designed to work with the Default Archiving (automatic) task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email.

When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email two months after the email was archived. In addition, the email is deleted eleven months after it was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_LD_1.2 - Default Archiving (stubbing).ctms`.

Journal Archiving (Email Deleted):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Database
This collector collects email from the journal database of a Domino server.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

Delete Email
This task deletes the email from the mailbox.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `CM_LD_3.1 - Journal Archiving (with Mail Delete).ctms`.

Journal Archiving (Email Retained):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Database
This collector collects email from the journal database of a Domino server.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `CM_LD_3.2 - Journal Archiving (without Mail Delete).ctms`.

Life Cycle Template:

This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).

It adds a status message to the original email one hour after it was archived. It removes the attachments 30 days after the email was archived. It removes the attachments and truncates the body text 60 days after the email was archived. It removes the body text 90 days after the email was archived. It deletes the entire email a year after it was archived. Restubbing is performed 14 days after the email was restored.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, stubbing is performed according to the settings defined on the Life Cycle tab. This tab also includes the time interval after which restored email is restubbed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_LD - Lifecycle Template.ctms`.

Mailbox Statistics Collection:

This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

File name

The file name of this template is `CM_LD - Mailbox Statistics Collection Template.ctms`.

NSF Archive Migration - Archiving:

This task route automatically archives documents from Lotus Domino NSF files. The task route empties the NSF files (deletes all documents in these files) as soon as the documents have been archived.

Collectors

This template uses the following collectors:

EC Collect All Documents from NSF Archives
This collector collects all documents from the NSF archives located in a folder.

Configure the collection source All NSF files in a folder tree.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

Delete Email
This task deletes the email from the NSF archive.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_LD NSF Migration - Archiving.ctms`.

Related tasks

"Archiving email from local files" on page 432

Space Saving (All Email):

This is one of three available task routes that emphasizes saving space on the mail server during email archiving.

This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 30 Days
This collector collects email that is older than 30 days.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments and Body
This task removes the attachments and the body from the email.

File name

The file name of this template is `CM_LD_2.1 - Space Saving (all email).ctms`.

Space Saving (Large Attachments):

This is one of three available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Week and Larger Than 1 MB
This collector collects email that is older than one week and larger than 1 MB.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CM_LD_2.2 - Space Saving (large attachments).ctms`.

Space Saving (Stubbing):

This is one of three available task routes that emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email three months after the email was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_LD_2.3 - Space Saving (stubbing).ctms`.

Lotus Domino with FileNet P8 archiving templates

Use these sample archiving templates as a basis for configuring task routes if you are archiving from a Lotus Domino mailbox to a FileNet P8 repository.

Select which of the supplied archiving templates fit your archiving user scenario:

Table 50. Typical archiving scenarios and matching templates

Typical archiving scenarios	Template files to use
Application document archiving: • This task route automatically archives documents from Lotus Domino applications, for example workflow documents, regularly if the documents are complete and have not been changed during a defined time period.	• "Application Archiving" on page 317
Business process management (BPM) archiving: • This task route archives email from monitored folders in mailboxes. The email can be used in a business process management (BPM) scenario. Email is not stubbed in the mailbox and cannot be restored.	• "BPM (Business Process Management)" on page 320

Table 50. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Common archiving setup: • This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the `Default Archiving (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Default Archiving (automatic)` task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed. • This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.	• "Default Archiving (Automatic)" on page 323 • "Default Archiving (Stubbing)" on page 325 • "Default Archiving (Interactive)" on page 324
Basic archiving setup with simple stubbing: • This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately. This task route can be used as a starting point for new task routes for email archiving.	• "Archiving" on page 318
Space saving on the mail server after archiving or restoring: • This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).	• "Life Cycle Template" on page 328

Table 50. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archiving in a complex organization: • This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the `Complex Organizations (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Complex Organization (Archiving)` task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline.	• "Complex Organization (Archiving)" on page 321 • "Complex Organization (Stubbing)" on page 322
Archiving with emphasis on space saving: • This is one of three available task routes that emphasizes saving space on the mail server during email archiving. This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed. • This is another available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived. • This space saving task route emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.	• "Space Saving (All Email)" on page 330 • "Space Saving (Large Attachments)" on page 331 • "Space Saving (Stubbing)" on page 332

Table 50. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archiving journal email: • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager. • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• "Journal Archiving (Email Deleted)" on page 326 • "Journal Archiving (Email Retained)" on page 327
Archiving documents in Lotus Domino databases • This task route automatically archive documents from Lotus Domino databases (NSF files). The task route empties the NSF files (deletes all documents in these files) as soon as the documents have been archived.	• "NSF Archive Migration - Archiving" on page 329
Collecting statistics on mailbox content: • This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.	• "Mailbox Statistics Collection" on page 329
Archiving calendar entries: • This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry. The task route detects duplicates and archives the same calendar entry only once.	• "Archiving Calendar Entries" on page 319

Application Archiving:

This task route automatically archives documents from Lotus Notes applications, for example workflow documents, regularly if the documents are complete and have not been changed during a defined time period.

Collectors

This template uses the following collectors:

EC Collect Documents from Application If Unchanged for at Least 30 Days and DocStatus OK
This collector collects documents from an application that is specified on the Collection Sources page. A document is collected if the value of the DocStatus field is "OK" and the document has not been changed during the last 30 days.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

P8 4.x Create Content Elements
This task copies to P8 any document content that is attached to the object that the P8 4.x Create Document task creates.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

Remove Attachments
This task removes the attachments from the document. All texts or attachment links created during stubbing are inserted in the rich text field specified in the Stubbing Options section.

File name

The file name of this template is `P8_LD Application Archiving.ctms`.

Archiving:

This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately.

This task route can be used as a starting point for new task routes for email archiving.

Collectors

This template uses the following collectors:

EC Collect All Email
This collector collects all email.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `P8_LD - Archiving Template.ctms`.

Archiving Calendar Entries:

This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry.

The task route detects duplicates and archives the same calendar entry only once.

Collectors

This template uses the following collectors:

EC Collect Calendar Entries That Completed 30 Days Ago
This collector collects calendar entries that completed at least thirty days ago.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_LD - Archiving Calendar Entries.ctms`.

BPM (Business Process Management):

This task route archives email from monitored folders in mailboxes. The email can be used in a business process management (BPM) scenario. Email is not stubbed in the mailbox and cannot be restored.

Collectors

This template uses the following collectors:

EC Collect Email from Monitored Folders
This collector collects email from monitored folders in mailboxes.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox.

Specify the folders to be monitored on the Collection Sources tab.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

Mark as Processed
This task marks the email as processed so that it will not be selected the next time.

P8 4.x Create Document for Attachment
This task creates a document in the repository for each attachment and stores the attachment in this document. It also sets the content retrieval name.

P8 4.x Create Document for Email
This task creates a document in the repository for each email and stores the email in this document. It also maps the email properties.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

P8 4.x Link Documents
This task creates link objects that associate a document with its attachments and associates files with each other.

File name

The file name of this template is `P8_LD - BPM Template.ctms`.

Complex Organization (Archiving):

This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the Complex Organizations (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 70 Days (New York)
This collector collects email that is older than 70 days from the mailboxes on a server in New York.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

EC Collect Email Older Than 70 Days (Paris)
This collector collects email that is older than 70 days from the mailboxes on a server in Paris.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_LD_4.1 - Complex Organization (archiving).ctms`.

Complex Organization (Stubbing):

This task route is designed to work with the Complex Organization (Archiving) task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline. There are two EC Create Email Stub tasks that deal with email after it was archived. The Remove Attachments and Body task removes the attachments and body after the email was copied to a local repository to save space on the mail server. The Use Life Cycle Collector task deals with email that exceeded the time limit for copying to a local repository.

Collectors

This template uses the following collectors:

Collect from Mailboxes on Servers in Paris and New York
This collector collects archived or restored email from mailboxes on servers in Paris and New York.

When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email 20 days after the email was archived. Restored email is restubbed after two months.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

Remove Attachments and Body
This task removes the attachments and the body from the email.

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_LD_4.2 - Complex Organization (stubbing).ctms`.

Default Archiving (Automatic):

This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the Default Archiving (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_LD_1.1 - Default Archiving (automatic).ctms`.

Default Archiving (Interactive):

This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.

Collectors

This template uses the following collectors:

EC Collect Email Based on Interactive Requests
This collector collects email that the users marked for archiving from the mailboxes.

Specify the Domino server and database path of the trigger mailbox on the Collection Sources tab.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `P8_LD_1.3 - Default Archiving (interactive).ctms`.

Default Archiving (Stubbing):

This task route is designed to work with the Default Archiving (automatic) task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email.

When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email two months after the email was archived. In addition, the email is deleted eleven months after it was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_LD_1.2 - Default Archiving (stubbing).ctms`.

Journal Archiving (Email Deleted):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Database
This collector collects email from the journal database of a Domino server.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

Delete Email
This task deletes the email from the mailbox.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `P8_LD_3.1 - Journal Archiving (with Mail Delete).ctms`.

Journal Archiving (Email Retained):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Database
This collector collects email from the journal database of a Domino server.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `P8_LD_3.2 - Journal Archiving (without Mail Delete).ctms`.

Life Cycle Template:

This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).

It adds a status message to the original email one hour after it was archived. It removes the attachments 30 days after the email was archived. It removes the attachments and truncates the body text 60 days after the email was archived. It removes the body text 90 days after the email was archived. It deletes the entire email a year after it was archived. Restubbing is performed 14 days after the email was restored.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, stubbing is performed according to the settings defined on the Life Cycle tab. This tab also includes the time interval after which restored email is restubbed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_LD - Lifecycle Template.ctms`.

Mailbox Statistics Collection:

This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

File name

The file name of this template is `P8_LD - Mailbox Statistics Collection Template.ctms`.

NSF Archive Migration - Archiving:

This task route automatically archives documents from Lotus Domino NSF files. The task route empties the NSF files (deletes all documents in these files) as soon as the documents have been archived.

Collectors

This template uses the following collectors:

EC Collect All Documents from NSF Archives
This collector collects all documents from the NSF archives located in a folder.

Configure the collection source All NSF files in a folder tree.

Tasks

This template uses the following tasks:

Delete Email
This task deletes the email from the NSF archive.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_LD NSF Migration - Archiving.ctms`.

Related tasks

"Archiving email from local files" on page 432

Space Saving (All Email):

This is one of three available task routes that emphasizes saving space on the mail server during email archiving.

This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 30 Days
This collector collects email that is older than 30 days.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments and Body
This task removes the attachments and the body from the email.

File name

The file name of this template is `P8_LD_2.1 - Space Saving (all email).ctms`.

Space Saving (Large Attachments):

This is one of three available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Week and Larger Than 1 MB
This collector collects email that is older than one week and larger than 1 MB.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `P8_LD_2.2 - Space Saving (large attachments).ctms`.

Space Saving (Stubbing):

This is one of three available task routes that emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email three months after the email was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_LD_2.3 - Space Saving (stubbing).ctms`.

Microsoft Exchange with IBM Content Manager archiving templates

Use these sample archiving templates as a basis for configuring task routes if you are archiving from a Microsoft Exchange mailbox to an IBM Content Manager repository.

Select which of the supplied archiving templates fit your archiving user scenario:

Table 51. Typical archiving scenarios and matching templates

Typical archiving scenarios	Template files to use
Archive messages from managed folders: • This task route automatically archives email in folders managed by Microsoft Exchange removing all email attachments immediately. It allows setting a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• "Archiving Managed Folders" on page 338
Archive messages in personal folders: • This task route archive messages from personal folders (PST files) automatically. The task route removes the attachments and the body of the messages as soon as the messages are archived. The resulting stubs are moved to the Microsoft Exchange mailboxes of the PST file owners, leaving the PST files empty.	• "PST Migration - Archiving" on page 347

Table 51. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Common archiving setup: • This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the `Default Archiving (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Default Archiving (automatic)` task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed. • This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.	• "Default Archiving (Automatic)" on page 341 • "Default Archiving (Stubbing)" on page 343 • "Default Archiving (Interactive)" on page 342
Basic archiving setup with simple stubbing: • This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately. This task route can be used as a starting point for new task routes for email archiving.	• "Archiving" on page 336
Space saving after archiving or restoring: • This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).	• "Life Cycle Template" on page 346

Table 51. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archiving in a complex organization: • This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the `Complex Organizations (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Complex Organization (Archiving)` task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline.	• "Complex Organization (Archiving)" on page 339 • "Complex Organization (Stubbing)" on page 341
Archiving with emphasis on space saving: • This is one of three available task routes that emphasizes saving space on the mail server during email archiving. This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed. • This is another available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived. • This space saving task route emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.	• "Space Saving (All Email)" on page 349 • "Space Saving (Large Attachments)" on page 350 • "Space Saving (Stubbing)" on page 351

Table 51. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archiving journal email: • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager. • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• "Journal Archiving (Email Deleted)" on page 344 • "Journal Archiving (Email Retained)" on page 345
Collecting statistics on mailbox content • This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.	• "Mailbox Statistics Collection" on page 347
Archiving calendar entries: • This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry. The task route detects duplicates and archives the same calendar entry only once.	• "Archiving Calendar Entries" on page 337

Archiving:

This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately.

This task route can be used as a starting point for new task routes for email archiving.

Collectors

This template uses the following collectors:

EC Collect All Email

This collector collects all email.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CM_EX - Archiving Template.ctms`.

Archiving Calendar Entries:

This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry.

The task route detects duplicates and archives the same calendar entry only once.

This task route requires a user defined metadata property of type Date Time. This metadata property is used to extract the appointment end date from the calendar entry. This timestamp is used in the rule Completed 30 Days Ago to avoid archiving calendar entries too early.

Collectors

This template uses the following collectors:

EC Collect Calendar Entries Older Than One Month
This collector collects calendar entries that are older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_EX - Archiving Calendar Entries.ctms`.

Archiving Managed Folders:

This task route automatically archives email in folders managed by Microsoft Exchange removing all email attachments immediately. It allows setting a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Managed Folders
This collector collects email from the mailbox folders that are managed by Microsoft Exchange.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

Set Retention Based on Folder Retention
This task sets the retention for the message to the value of the system metadata property <Email, Managed Folder Retention>. All messages that exceed the retention can be deleted from the archive using the Content Collector Expiration Manager.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `CM_EX Managed Folder Archiving.ctms`.

Complex Organization (Archiving):

This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing

options can be defined by using the Complex Organizations (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 70 Days (New York)
This collector collects email that is older than 70 days from the mailboxes on a server in New York.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

EC Collect Email Older Than 70 Days (Paris)
This collector collects email that is older than 70 days from the mailboxes on a server in Paris.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_EX_4.1 - Complex Organization (archiving).ctms`.

Complex Organization (Stubbing):

This task route is designed to work with the Complex Organization (Archiving) task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline. There are two EC Create Email Stub tasks that deal with email after it was archived. The Remove Attachments and Body task removes the attachments and body after the email was copied to a local repository to save space on the mail server. The Use Life Cycle Collector task deals with email that exceeded the time limit for copying to a local repository.

Collectors

This template uses the following collectors:

Collect from Mailboxes on Servers in Paris and New York
This collector collects archived or restored email from mailboxes on servers in Paris and New York.

When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email 20 days after the email was archived. Restored email is restubbed after two months.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

Remove Attachments and Body
This task removes the attachments and the body from the email.

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_EX_4.2 - Complex Organization (stubbing).ctms`.

Default Archiving (Automatic):

This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the Default Archiving (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CM_EX_1.1 - Default Archiving (automatic).ctms`.

Default Archiving (Interactive):

This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.

Collectors

This template uses the following collectors:

EC Collect Email Based on Interactive Requests
This collector collects email that the users marked for archiving from the mailboxes.

Specify the trigger mailbox on the Collection Sources tab.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CM_EX_1.3 - Default Archiving (interactive).ctms`.

Default Archiving (Stubbing):

This task route is designed to work with the Default Archiving (automatic) task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email.

When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email two months after the email was archived. In addition, the email is deleted eleven months after it was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_EX_1.2 - Default Archiving (stubbing).ctms`.

Journal Archiving (Email Deleted):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Mailbox
This collector collects email from a journal recipient mailbox.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

Delete Email
This task deletes the email from the mailbox.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `CM_EX_3.1 - Journal Archiving (with Mail Delete).ctms`.

Journal Archiving (Email Retained):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Mailbox
This collector collects email from a journal recipient mailbox.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `CM_EX_3.2 - Journal Archiving (without Mail Delete).ctms`.

Life Cycle Template:

This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).

It adds a status message to the original email one hour after it was archived. It removes the attachments 30 days after the email was archived. It removes the attachments and truncates the body text 60 days after the email was archived. It removes the body text 90 days after the email was archived. It deletes the entire email a year after it was archived. Restubbing is performed 14 days after the email was restored.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, stubbing is performed according to the settings defined on the Life Cycle tab. This tab also includes the time interval after which restored email is restubbed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_EX - Lifecycle Template.ctms`.

Mailbox Statistics Collection:

This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

File name

The file name of this template is `CM_EX - Mailbox Statistics Collection Template.ctms`.

PST Migration - Archiving:

This task route archives messages from personal folders (PST files) automatically. The task route removes the attachments and the body of the messages as soon as the messages are archived. The resulting stubs are moved to the Exchange mailboxes of the PST file owners, leaving the PST files empty.

Collectors

This template uses the following collectors:

EC Collect All Messages from PST files
This collector collects all messages from personal folders (PST files).

Configure one of the following collection sources: PST file on local computer, All PST files in a folder tree, All PST files on a computer, All PST files in a computer group

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Copy to Mailbox
This task copies the stub from the personal folder (PST file) to the Exchange mailbox of the PST file owner.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments and Body from Email in PST
This task removes the attachments and the body from the email in the personal folder (PST file).

Remove Email Stub from PST
This task removes the stub from the PST file.

File name

The file name of this template is `CM_EX PST Migration - Archiving.ctms`.

Related tasks

"Archiving email from local files" on page 432

Space Saving (All Email):

This is one of three available task routes that emphasizes saving space on the mail server during email archiving.

This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 30 Days
This collector collects email that is older than 30 days.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments and Body
This task removes the attachments and the body from the email.

File name

The file name of this template is `CM_EX_2.1 - Space Saving (all email).ctms`.

Space Saving (Large Attachments):

This is one of three available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Week and Larger Than 1 MB
This collector collects email that is older than one week and larger than 1 MB.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CM_EX_2.2 - Space Saving (large attachments).ctms`.

Space Saving (Stubbing):

This is one of three available task routes that emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email three months after the email was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CM_EX_2.3 - Space Saving (stubbing).ctms`.

Microsoft Exchange with FileNet P8 archiving templates

Use these sample archiving templates as a basis for configuring task routes if you are archiving from a Microsoft Exchange mailbox to a FileNet P8 repository.

Select which of the supplied archiving templates fit your archiving user scenario:

Table 52. Typical archiving scenarios and matching templates

Typical archiving scenarios	Template files to use
Archive messages from managed folders: • This task route automatically archives email in folders managed by Microsoft Exchange removing all email attachments immediately. It allows setting a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• "Archiving Managed Folders" on page 356

Table 52. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archive messages in personal folders: • This task route archive messages from personal folders (PST files) automatically. The task route removes the attachments and the body of the messages as soon as the messages are archived. The resulting stubs are moved to the Microsoft Exchange mailboxes of the PST file owners, leaving the PST files empty.	• "PST Migration - Archiving" on page 367
Business process management (BPM) archiving: • This task route archives email from monitored folders in mailboxes. The email can be used in a business process management (BPM) scenario. Email is not stubbed in the mailbox and cannot be restored.	• "BPM (Business Process Management)" on page 357
Common archiving setup: • This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the `Default Archiving (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Default Archiving (automatic)` task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed. • This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.	• "Default Archiving (Automatic)" on page 360 • "Default Archiving (Stubbing)" on page 362 • "Default Archiving (Interactive)" on page 361

Table 52. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Basic archiving setup with simple stubbing: • This is a basic task route that archives email on a server automatically on a regular basis, detecting duplicates and removing attachments in email immediately. This task route can be used as a starting point for new task routes for email archiving.	• "Archiving" on page 354
Space saving after archiving or restoring: • This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).	• "Life Cycle Template" on page 365
Archiving in a complex organization: • This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the `Complex Organizations (Stubbing)` task route that can be configured to work with this task route. • This task route is designed to work with the `Complex Organization (Archiving)` task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline.	• "Complex Organization (Archiving)" on page 358 • "Complex Organization (Stubbing)" on page 360
Archiving with emphasis on space saving: • This is one of three available task routes that emphasizes saving space on the mail server during email archiving. This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed. • This is another available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived. • This space saving task route emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.	• "Space Saving (All Email)" on page 368 • "Space Saving (Large Attachments)" on page 369 • "Space Saving (Stubbing)" on page 370

Table 52. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archiving journal email: • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager. • This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.	• "Journal Archiving (Email Deleted)" on page 363 • "Journal Archiving (Email Retained)" on page 364
Collecting statistics on mailbox content • This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.	• "Mailbox Statistics Collection" on page 366
Archiving calendar entries: • This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry. The task route detects duplicates and archives the same calendar entry only once.	• "Archiving Calendar Entries" on page 355

Archiving:

This is a basic task route that archives email automatically on a regular basis, detecting duplicates and removing attachments in email immediately.

This task route can be used as a starting point for new task routes for email archiving.

Collectors

This template uses the following collectors:

EC Collect All Email

This collector collects all email.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `P8_EX - Archiving Template.ctms`.

Archiving Calendar Entries:

This is a special task route that archives calendar entries automatically on a regular basis. The task route only archives calendar entries if at least 30 days have elapsed after the end date specified in the calendar entry. It does not remove any content from the calendar entry.

The task route detects duplicates and archives the same calendar entry only once.

Collectors

This template uses the following collectors:

EC Collect Calendar Entries Older Than One Month
This collector collects calendar entries that are older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_EX - Archiving Calendar Entries.ctms`.

Archiving Managed Folders:

This task route automatically archives email in folders managed by Microsoft Exchange removing all email attachments immediately. It allows setting a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Managed Folders
This collector collects email from the mailbox folders that are managed by Microsoft Exchange.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

Set Retention Based on Folder Retention
This task sets the retention for the message to the value of the system metadata property <Email, Managed Folder Retention>. All messages that exceed the retention can be deleted from the archive using the Content Collector Expiration Manager.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `P8_EX Managed Folder Archiving.ctms`.

BPM (Business Process Management):

This task route archives email from monitored folders in mailboxes. The email can be used in a business process management (BPM) scenario. Email is not stubbed in the mailbox and cannot be restored.

Collectors

This template uses the following collectors:

EC Collect Email from Monitored Folders
This collector collects email from monitored folders in mailboxes.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox.

Specify the folders to be monitored on the Collection Sources tab.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

Mark as Processed
This task marks the email as processed so that it will not be selected the next time.

P8 4.x Create Document for Attachment
This task creates a document in the repository for each attachment and stores the attachment in this document. It also sets the content retrieval name.

P8 4.x Create Document for Email
This task creates a document in the repository for each email and stores the email in this document. It also maps the email properties.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

P8 4.x Link Documents
This task creates link objects that associate a document with its attachments and associates files with each other.

File name

The file name of this template is `P8_EX - BPM Template.ctms`.

Complex Organization (Archiving):

This task route is designed to support users working in an organization with locations in different time zones. It automatically archives email older than a

specified time. This task route does not contain a stubbing task as the stubbing options can be defined by using the Complex Organizations (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 70 Days (New York)
This collector collects email that is older than 70 days from the mailboxes on a server in New York.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

EC Collect Email Older Than 70 Days (Paris)
This collector collects email that is older than 70 days from the mailboxes on a server in Paris.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_EX_4.1 - Complex Organization (archiving).ctms`.

Complex Organization (Stubbing):

This task route is designed to work with the Complex Organization (Archiving) task route and removes email content in defined stages. The task route also considers that users might need access to their email when they are working offline. There are two EC Create Email Stub tasks that deal with email after it was archived. The Remove Attachments and Body task removes the attachments and body after the email was copied to a local repository to save space on the mail server. The Use Life Cycle Collector task deals with email that exceeded the time limit for copying to a local repository.

Collectors

This template uses the following collectors:

Collect from Mailboxes on Servers in Paris and New York
This collector collects archived or restored email from mailboxes on servers in Paris and New York.

When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email 20 days after the email was archived. Restored email is restubbed after two months.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

Remove Attachments and Body
This task removes the attachments and the body from the email.

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_EX_4.2 - Complex Organization (stubbing).ctms`.

Default Archiving (Automatic):

This task route archives all email that has been in a mailbox for a defined period of time automatically on a regular basis. The task route detects duplicates and archives the same email only once. This task route does not contain a stubbing task as the stubbing options can be defined by the Default Archiving (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

File name

The file name of this template is `P8_EX_1.1 - Default Archiving (automatic).ctms`.

Default Archiving (Interactive):

This task route archives email based on archiving requests. When a user marks email for archiving, an archiving request is sent to the trigger mailbox that is defined in the client configuration. The archiving requests are collected and the marked email documents are archived. The task route detects duplicates and removes attachments from the email in the mailbox.

Collectors

This template uses the following collectors:

EC Collect Email Based on Interactive Requests
This collector collects email that the users marked for archiving from the mailboxes.

Specify the trigger mailbox on the Collection Sources tab.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `P8_EX_1.3 - Default Archiving (interactive).ctms`.

Default Archiving (Stubbing):

This task route is designed to work with the Default Archiving (automatic) task route and removes email content in defined stages. The older the email, the less email content is stored on the mail server. It removes the attachments from the

email two months after the email was archived. In addition, it deletes the email eleven months after it was archived. If email is restored, restubbing is not performed.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email.

When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email two months after the email was archived. In addition, the email is deleted eleven months after it was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_EX_1.2 - Default Archiving (stubbing).ctms`.

Journal Archiving (Email Deleted):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. The task route deletes email from the journal immediately after it was archived. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Mailbox
This collector collects email from a journal recipient mailbox.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

Delete Email
This task deletes the email from the mailbox.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `P8_EX_3.1 - Journal Archiving (with Mail Delete).ctms`.

Journal Archiving (Email Retained):

This task route archives all email in a journal automatically except for archiving requests that were generated when users marked email for archiving. All archived email is retained in the journal. The task route sets a retention period for email in the archive. This email can be deleted from the archive using the Content Collector Expiration Manager.

Collectors

This template uses the following collectors:

EC Collect Email from Journal Mailbox
This collector collects email from a journal recipient mailbox.

Configure the collection source Journal.

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Set Retention to 3 Years
This task sets the retention to three years. All email documents that were archived more than three years ago can be deleted from the archive using the Content Collector Expiration Manager.

File name

The file name of this template is `P8_EX_3.2 - Journal Archiving (without Mail Delete).ctms`.

Life Cycle Template:

This is a space saving task route that can be used with an archiving task route. This task route saves space by processing email that has already been archived or was restored. The task route removes content from the email in defined stages (delayed stubbing).

It adds a status message to the original email one hour after it was archived. It removes the attachments 30 days after the email was archived. It removes the attachments and truncates the body text 60 days after the email was archived. It removes the body text 90 days after the email was archived. It deletes the entire email a year after it was archived. Restubbing is performed 14 days after the email was restored.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, stubbing is performed according to the settings defined on the Life Cycle tab. This tab also includes the time interval after which restored email is restubbed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_EX - Lifecycle Template.ctms`.

Mailbox Statistics Collection:

This task route collects statistical information about the mailbox content. The task route does not archive or modify any data. The task route collects information which can be evaluated to determine or monitor users' mail behavior.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Month
This collector collects email that is older than one month.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

File name

The file name of this template is `P8_EX - Mailbox Statistics Collection Template.ctms`.

PST Migration - Archiving:

This task route archives messages from personal folders (PST files) automatically. The task route removes the attachments and the body of the messages as soon as the messages are archived. The resulting stubs are moved to the Exchange mailboxes of the PST file owners, leaving the PST files empty.

Collectors

This template uses the following collectors:

EC Collect All Messages from PST files
This collector collects all messages from personal folders (PST files).

Configure one of the following collection sources: PST file on local computer, All PST files in a folder tree, All PST files on a computer, All PST files in a computer group

Tasks

This template uses the following tasks:

EC Copy to Mailbox
This task copies the stub from the personal folder (PST file) to the Exchange mailbox of the PST file owner.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments and Body from Email in PST
This task removes the attachments and the body from the email in the personal folder (PST file).

Remove Email Stub from PST
This task removes the stub from the PST file.

File name

The file name of this template is `P8_EX PST Migration - Archiving.ctms`.

Related tasks

"Archiving email from local files" on page 432

Space Saving (All Email):

This is one of three available task routes that emphasizes saving space on the mail server during email archiving.

This task route archives email automatically after a given period of time. To save space on the mail server, all email attachments and the email body are removed.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 30 Days
This collector collects email that is older than 30 days.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments and Body
This task removes the attachments and the body from the email.

File name

The file name of this template is `P8_EX_2.1 - Space Saving (all email).ctms`.

Space Saving (Large Attachments):

This is one of three available task routes that emphasizes saving space on the mail server by archiving large email only. Attachments are removed and replaced with links when the email documents are archived.

Collectors

This template uses the following collectors:

EC Collect Email Older Than One Week and Larger Than 1 MB
This collector collects email that is older than one week and larger than 1 MB.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `P8_EX_2.2 - Space Saving (large attachments).ctms`.

Space Saving (Stubbing):

This is one of three available task routes that emphasizes saving space on the mail server after email was archived. To support email users working offline for longer periods of time, automatic restubbing is disabled.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email. When stubbing is delayed in an EC Create Email Stub task, the attachments and the body are removed from the email three months after the email was archived. Restubbing is not performed.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox, Public store

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `P8_EX_2.3 - Space Saving (stubbing).ctms`.

File System with IBM Content Manager archiving templates

Use these sample archiving templates as a basis for configuring task routes if you are archiving fFile System documents to a IBM Content Manager repository.

Select which of the supplied archiving templates fit your archiving user scenario:

Table 53. Typical archiving scenarios and matching templates

Typical archiving scenarios	Template files to use
Archive files dependent on defined metadata properties, delete the files on the file system and insert shortcut URLs: • Archive files in monitored file system folders. The files are replaced by a shortcut URL. For each file that does not have the `.xml` file extension there must be a file with same name and the `.xml` file extension. For example if there is a file named `MyDocument.doc` then in the same folder there must be a file with the name `MyDocument.xml`. The format of the `.xml` files must be defined in the File System Metadata pane.	• "FS to CM Archiving (Associate Metadata)" on page 372
Archive files automatically and delete the files to save space on the file system: • Archives all files in monitored file system folders. The files are deleted after they have been archived.	• "FS to CM8 Archiving (Delete)" on page 373
Archive files automatically and replace files with a shortcut URL: • Archives all files in monitored file system folders. The files are replaced with a shortcut URL.	• "FS to CM8 Archiving (Shortcut)" on page 374
Archive files automatically and delete files and metadata on the file system: • Archive all files in monitored file system folders that do not have the `.csv` file extension. The archived files are replaced with a shortcut URL. The task route deletes the metadata file for each archived file. For each file that does not have the `.csv` file extension there must be a file with same name and the `.csv` file extension. For example if there is a file named `MyDocument.doc` then in the same folder there must be a file with the name `MyDocument.csv`. The format of the `.csv` files must be defined in the File System Metadata pane.	• "FS to CM Archiving (Delete Metadata Files)" on page 373
Gather statistics about file system sources: • Collect statistical information about file system sources. The task route does not archive or modify any data. It only collects information which can be evaluated to determine or monitor file system usage.	• "FS Statistics Collection" on page 372

Note: There are no sample templates in IBM Content Collector archiving File System to IBM Content Manager that support file deduplication. You will need to configure your own task route that covers archiving with deduplication.

FS Statistics Collection:

This task route collects statistical information about file system sources. It does not archive or modify any data. It only collects information which can be evaluated to determine or monitor file system usage.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

File name

The file name of this template is `FS Statistics Collection.ctms`.

FS to CM Archiving (Associate Metadata):

This task route archives all files in monitored file system folders. The files are replaced a shortcut URL.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. For each file that does not have the file extension .xml there must be a file with same name and the file extension .xml. For example if there is a file named MyDocument.doc then in the same folder there must be a file with the name MyDocument.xml. The format of the .xml files must be defined in the File System Metadata pane. You must also specify the server name in the shortcut URL in the CM 8.x Create Document task.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Create Document
This task stores documents into the CM8 repository.

FSC Associate Metadata
This task associates content files with additional information, which can consist of content files or metadata files.

FSC Post Processing - replace files with shortcuts
This task replaces files on the file system with shortcuts to the archived files.

File name

The file name of this template is `FS to CM Archiving (Associate Metadata).ctms`.

FS to CM8 Archiving (Delete):

This task route archives all files in monitored file system folders. The files are deleted after they have been archived.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Create Document
This task stores documents into the CM8 repository.

FSC Post Processing
This task defines what happens to a file on the file system after processing.

File name

The file name of this template is `FS to CM Archiving (Delete).ctms`.

FS to CM Archiving (Delete Metadata Files):

This task route archives all files in monitored file system folders that do not have the file extension .csv. The archived files are replaced with a shortcut URL. The task route deletes the metadata file for each archived file.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. For each file that does not have the file extension .csv there must be a file with same name and the file extension .csv. For example if there is a file named MyDocument.doc then in the same folder there must be a file with the name MyDocument.csv. The format of the .csv files must be defined in the

File System Metadata pane. You must add a server name to the shortcut URL in the CM 8.x Create Document task.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Create Document
This task stores documents into the CM8 repository.

FSC Associate Metadata
This task associates content files with additional information, which can consist of content files or metadata files.

FSC Post Processing
This task defines what happens to a file on the file system after processing.

File name

The file name of this template is `FS to CM Archiving (Delete Metadata Files).ctms`.

FS to CM8 Archiving (Shortcut):

This task route archives all files in monitored file system folders. The files are replaced with a shortcut URL.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. You must also specify the server name in the shortcut URL in the CM 8.x Create Document task.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Create Document
This task stores documents into the CM8 repository.

FSC Post Processing - stub and delete
This task defines what happens to a file on the file system after processing.

File name

The file name of this template is `FS to CM Archiving (Shortcut).ctms`.

File System with FileNet P8 archiving templates

Use these sample archiving templates as a basis for configuring task routes if you are archiving File System documents to a FileNet P8 repository.

Select which of the supplied archiving templates fit your archiving user scenario:

Table 54. Typical archiving scenarios and matching templates

Typical archiving scenarios	Template files to use
Archive files dependent on defined metadata properties automatically, delete the files on the file system and insert shortcuts to the archived files: • Archive all files in monitored file system folders. The task route mirrors the folder structure at the source locations in the archive. The files are replaced a shortcut URL. For each file that does not have the `.xml` file extension there must be a file with same name and the `.xml` file extension. For example, if there is a file named `MyDocument.doc` then in the same folder there must be a file with the name `MyDocument.xml`. The format of the `.xml` files must be defined in the File System Metadata pane.	• "FS to P8 Archiving (Associate Metadata)" on page 377
Archive files automatically and delete the files on the file system: : • Archive all files in monitored file system folders. The task route mirrors the folder structure at the source locations in the archive. The files are deleted after they have been archived.	• "FS to P8 Archiving (Delete)" on page 379
Archive files automatically and delete the files on the file system: • Archive all files in monitored file system folders. The files are replaced with a shortcut URL.	• "FS to P8 Archiving (Shortcut)" on page 381

Table 54. Typical archiving scenarios and matching templates (continued)

Typical archiving scenarios	Template files to use
Archive files automatically and delete metadata files on the file system: • Archive all files in monitored file system folders that do not have the `.csv` file extension. The task route mirrors the folder structure at the source locations in the archive. The archived files are replaced with a shortcut URL. The task route deletes the metadata file for each archived file. For each file that does not have the `.csv` file extension there must be a file with same name and the `.csv` file extension. For example, if there is a file named `MyDocument.doc` then in the same folder there must be a file with the name `MyDocument.csv`. The format of the `.csv`files must be defined in the File System Metadata pane.	• "FS to P8 Archiving (Delete Metadata Files)" on page 380
Archive only one copy of a file in the repository and declare these archived files as records: • Archive all unique files in monitored file system folders where uniqueness is determined by hashing the contents of the files. It mirrors the folder structure at the source locations in the archive. Each file that is archived is declared as a record. Both unique files and duplicate files are replaced with a shortcut URL. • Classify documents using IBM Classification Manager, and then declare these documents as records in FileNet P8.	• "FS to P8 Archiving (Declare as Records)" on page 378 • "FS to P8 Archiving (Declare as Records) with IBM Classification Module" on page 378
Archive files from the file system in the repository in the same structure as on the file system to improve file searchability: • Archives all unique files in monitored file system folders where uniqueness is determined by hashing the contents of the files. The task route mirrors the folder structure at the source locations in the archive. Both unique files and duplicate files are replaced a shortcut URL.	• "FS to P8 Archiving (Replicate File System and Detect Duplicates)" on page 380
Gather statistics about file system sources: • Collect statistical information about file system sources. The task route does not archive or modify any data. It only collects information which can be evaluated to determine or monitor file system usage.	• "FS Statistics Collection" on page 372

FS Statistics Collection:

This task route collects statistical information about file system sources. It does not archive or modify any data. It only collects information which can be evaluated to determine or monitor file system usage.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

File name

The file name of this template is `FS Statistics Collection.ctms`.

FS to P8 Archiving (Associate Metadata):

This task route archives all files in monitored file system folders. It mirrors the folder structure at the source locations in the archive. The files are replaced a shortcut URL.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. For each file that does not have the file extension .xml there must be a file with same name and the file extension .xml. For example if there is a file named MyDocument.doc then in the same folder there must be a file with the name MyDocument.xml. The format of the .xml files must be defined in the File System Metadata pane. You must also specify the server name in the shortcut URL in the P8 4.x Create Document task.

Tasks

This template uses the following tasks:

FSC Associate Metadata
This task associates content files with additional information, which can consist of content files or metadata files.

FSC Post Processing - replace files with shortcuts
This task replaces files on the file system with shortcuts to the archived files.

P8 4.x Create Document
Archive all files.

P8 4.x File Document in Folder
This task replicates the file system structure in IBM FileNet P8.

File name

The file name of this template is `FS to P8 Archiving (Associate Metadata).ctms`.

FS to P8 Archiving (Declare as Records):

This task route archives all unique files in monitored file system folders where uniqueness is determined by hashing the contents of the files. It mirrors the folder structure at the source locations in the archive. Each file that is archived is declared as a record. Both unique files and duplicate files are replaced with a shortcut URL.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. You must also specify the server name in the shortcut URL in the P8 4.x Create Document task. To configure the record object store and the record class in the P8 4.x Declare Record task IBM Enterprise Records must be installed and configured to enable declaring of records.

Tasks

This template uses the following tasks:

FSC Post Processing
This task defines what happens to a file on the file system after processing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Declare Record
This task enables records declaration in IBM Enterprise Records.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

File name

The file name of this template is `FS to P8 Archiving (Declare as Records).ctms`.

FS to P8 Archiving (Declare as Records) with IBM Classification Module:

This template classifies documents using IBM Classification Manager, and then declares documents as records in IBM FileNet P8.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

Tasks

This template uses the following tasks:

FSC Post Processing
This task defines what happens to a file on the file system after processing.

IBM Classification Module
In order to use the IBM Classification Module, it is neccessary to have an IBM Classification Module server installed and configured using either a Knowledge Base or Decision Plan in order to gernerate classification metadata. Please see your IBM Classification Module documentation for instructions relating to installing and configuring this server. Once configured, the server can be used with this task to classify document content. The metadata produced by this classification can then be used for decision points or property mapping in later tasks.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Declare Record
This task enables records declaration in IBM Enterprise Records.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

File name

The file name of this template is `FS to P8 Archiving (Declare as Records) with IBM Classification Module.ctms`.

FS to P8 Archiving (Delete):

This task route archives all files in monitored file system folders. It mirrors the folder structure at the source locations in the archive. The files are deleted after they have been archived.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from.

Tasks

This template uses the following tasks:

FSC Post Processing
This task defines what happens to a file on the file system after processing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

File name

The file name of this template is `FS to P8 Archiving (Delete).ctms`.

FS to P8 Archiving (Delete Metadata Files):

This template processes content files differently from metadata files. The specified metadata must be created first in the File System Metadata screen before you can import this template.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

Tasks

This template uses the following tasks:

FSC Associate Metadata
This task associates content files with additional information, which can consist of content files or metadata files.

FSC Post Processing
This task defines what happens to a file on the file system after processing.

P8 4.x Create Document
This task copies documents or document metadata to IBM FileNet P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

File name

The file name of this template is `FS to P8 Archiving (Delete Metadata Files).ctms`.

FS to P8 Archiving (Replicate File System and Detect Duplicates):

This task route archives all unique files in monitored file system folders where uniqueness is determined by hashing the contents of the files. It mirrors the folder structure at the source locations in the archive. Both unique files and duplicate files are replaced a shortcut URL.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. You must also specify the server name in the shortcut URL in the P8 4.x Create Document task.

Tasks

This template uses the following tasks:

FSC Post Processing
This task defines what happens to a file on the file system after processing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

File name

The file name of this template is `FS to P8 Archiving (Replicate File System and Detect Duplicates).ctms`.

FS to P8 Archiving (Shortcut):

This task route archives all files in monitored file system folders. The files are replaced with a shortcut URL.

Collectors

This template uses the following collectors:

FSC Collector
This collector collects files from specified directories according to filter criteria.

You must specify where the collector should collect files from. You must also specify the server name in the shortcut URL in the P8 4.x Create Document task.

Tasks

This template uses the following tasks:

FSC Post Processing - stub and delete
This task defines what happens to a file on the file system after processing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

File name

The file name of this template is `FS to P8 Archiving (Shortcut).ctms`.

SMTP with IBM Content Manager or FileNet P8 archiving templates

Use the sample archiving template as a basis for configuring task routes if you are archiving from SMTP to an IBM Content Manager or a FileNet P8 repository.

There is only one SMTP sample template for compliance archiving:

Table 55. Typical SMTP archiving scenario

Typical archiving scenario	Template file to use
Basic archiving setup with deletion: • This is a basic task route that archives email from the message queue directory that is defined in the SMTP connector automatically on a regular basis. It detects duplicates and removes email immediately. This task route can be used as a starting point for new task routes for email archiving.	• Archiving to IBM Content Manager • Archiving to IBM FileNet P8

Archiving to IBM Content Manager:

This is a basic task route that archives email from the message queue directory that is defined in the SMTP connector automatically on a regular basis. It detects duplicates and removes email immediately.

This task route can be used as a starting point for new task routes for email archiving.

Collectors

This template uses the following collectors:

SC Collect All Email
This collector collects email from the message queue directory.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

SC Delete Email
This task deletes the received email from the message queue directory.

SC Extract Attachments
This task extracts the attachments from the email in the message queue directory so that the attachments can be stored as separate files in the repository.

SC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

SC Prepare Email for Archiving
This task creates working copies of the email in the message queue directory, excluding attachments, and saves them as temporary files for further processing.

SC Prepare Email for Deletion
Once a document is successfully archived, this task marks it as archived. It can then be deleted by the SC Delete Email task.

File name

The file name of this template is `CM_SMTP - Archiving Template.ctms`.

Archiving to FileNet P8:

This is a basic task route that archives email from the message queue directory that is defined in the SMTP connector automatically on a regular basis. It detects duplicates and removes email immediately.

This task route can be used as a starting point for new task routes for email archiving.

Collectors

This template uses the following collectors:

SC Collect All Email
This collector collects email from the message queue directory.

Tasks

This template uses the following tasks:

Extract Text
This task extracts the text of email attachments and prepares it for full-text indexing.

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x Create Email Instance
This task extracts metadata from email documents in mailboxes. Use this metadata to classify repository documents and provide additional search criteria for metadata-based queries.

P8 4.x Save Prepared Text as XML
This task stores the information that P8 needs for full-text search of documents and their attachments.

SC Delete Email
This task deletes the received email from the message queue directory.

SC Extract Attachments
This task extracts the attachments from the email in the message queue directory so that the attachments can be stored as separate files in the repository.

SC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

SC Prepare Email for Archiving
This task creates working copies of the email in the message queue directory, excluding attachments, and saves them as temporary files for further processing.

SC Prepare Email for Deletion
Once a document is successfully archived, this task marks it as archived. It can then be deleted by the SC Delete Email task.

File name

The file name of this template is `P8_SMTP - Archiving Template.ctms`.

SharePoint task routes

SharePoint task route templates archive SharePoint documents to a target repository.

SP Audit CM Links:

SP Audit CM Links identifies and reports broken links from your SharePoint server to migrated documents in your Content Manager repository.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

CM 8.x Confirm Document
This task attempts to confirm the existence of a document in IBM CM 8.x

File name

The file name of this template is `SP Audit CM Links.ctms`.

SP Audit P8 Links:

SP Audit P8 Links identifies and reports broken links from your SharePoint server to migrated documents in your FileNet P8 repository.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

P8 4.x Confirm Document
This task attempts to confirm the existence of a document in P8.

File name

The file name of this template is `SP Audit P8 Links.ctms`.

SP Manage CM Links:

SP Manage CM Links identifies and resolves broken links from your SharePoint server to migrated documents in your Content Manager repository. It deletes links to missing content and can update resolvable links.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

CM 8.x Confirm Document
This task attempts to confirm the existence of a document in IBM CM 8.x

SP Manage Link
SP Manage Link deletes or updates unresolved migrated document links on the SharePoint server, depending on your configuration and the retrievability of linked documents in the target repository. The task deletes all links that point to missing content.

File name

The file name of this template is `SP Manage CM Links.ctms`.

SP Manage P8 Links:

SP Manage P8 Links identifies and resolves broken links from your SharePoint server to migrated documents in your FileNet P8 repository. It deletes links to missing content and can update resolvable links.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

P8 4.x Confirm Document
This task attempts to confirm the existence of a document in P8.

SP Manage Link
SP Manage Link deletes or updates unresolved migrated document links on the SharePoint server, depending on your configuration and the retrievability of linked documents in the target repository. The task deletes all links that point to missing content.

File name

The file name of this template is `SP Manage P8 Links.ctms`.

SP Statistics Collection:

This task route collects SharePoint site statistics but does not archive or modify any data. The properties that you select in the audit log task determine the statistics to be collected.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

File name

The file name of this template is `SP Statistics Collection.ctms`.

SP to CM - Calculate Expiration Date:

This task route archives versions of SharePoint documents to Content Manager, creating a document in Content Manager with all the versions of each SharePoint document, filing it in a folder, and removing all but the most recent version of the document on the SharePoint server. The expiration date of the Content Manager document is set to the last modified date of the SharePoint document plus one year.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

Calculate Expiration Date
This task sets a retention (expiry) date on content based on user name or LDAP group membership

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Store Version Series
This task stores 1-n versions of a document as CM versions.

SP Create File
SP Create File creates a temporary local file copy for each SharePoint document version to be processed.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to CM - Calculate Expiration Date.ctms`.

SP to CM - Classify:

This task route archives versions of SharePoint documents to Content Manager, creating a document in Content Manager with all the versions of each SharePoint document, filing it in a folder, and removing all but the most recent version of the document on the SharePoint server. The classification task must be configured before this task route can be used.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Store Version Series
This task stores 1-n versions of a document as CM versions.

IBM Classification Module

SP Create File
SP Create File creates a temporary local file copy for each SharePoint document version to be processed.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to CM - Classify.ctms`.

SP to CM - No Versions:

This task route archives SharePoint documents to Content Manager, creating a document in Content Manager for the most recent version of each SharePoint document, filing it in a folder, and marking the original document on the SharePoint server.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Create Document
This task stores documents into the CM8 repository.

SP Create File
SP Create File creates a temporary local file copy for each SharePoint document version to be processed.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to CM - No Versions.ctms`.

SP to CM - With Versions:

This task route archives versions of SharePoint documents to Content Manager, creating a document in Content Manager with all the versions of each SharePoint document, filing it in a folder, and removing all but the most recent version of the document from the SharePoint server.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Store Version Series
This task stores 1-n versions of a document as CM versions.

SP Create File
SP Create File creates a temporary local file copy for each SharePoint document version to be processed.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to CM - With Versions.ctms`.

SP to P8 - Calculate Expiration Date:

This task route archives versions of SharePoint documents to FileNet P8, creating a document in P8 with all the versions of each SharePoint document, filing it in a folder, and removing all but the most recent version of the document from the

SharePoint server. The expiration date of P8 document is set to the last modified date of the SharePoint document plus one year.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

Calculate Expiration Date
This task sets a retention (expiry) date on content based on user name or LDAP group membership

P8 4.x Create Version Series
This task creates a version series in P8 for the item being processed.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to P8 - Calculate Expiration Date.ctms`.

SP to P8 - Classify:

This task route archives versions of SharePoint documents to FileNet P8, creating a document in P8 with all the versions of each SharePoint document, filing it in a folder, and removing all but the most recent version of the document from the SharePoint server. You must configure the classification task before using this task route.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

IBM Classification Module

P8 4.x Create Version Series
This task creates a version series in P8 for the item being processed.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

SP Create File
SP Create File creates a temporary local file copy for each SharePoint document version to be processed.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to P8 - Classify.ctms`.

SP to P8 - Declare as Record:

This task route archives versions of SharePoint documents to FileNet P8, creating a document in P8 with all the versions of each SharePoint document, filing it in a folder, declaring the most recent version as a record, and removing all but the most recent version of the document from the SharePoint server.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

P8 4.x Create Version Series
This task creates a version series in P8 for the item being processed.

P8 4.x Declare Record
This task enables records declaration in IBM FileNet Records Manager.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to P8 - Declare as Record.ctms`.

SP to P8 - No Versions:

This task route archives SharePoint documents to FileNet P8, creating a document in P8 for the most recent version of each SharePoint document, filing it in a folder, and marking the original document on the SharePoint server.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

P8 4.x Create Document
This task copies documents or document metadata to P8. You can use the P8 4.x Create Content Elements task to defer the storage to a later point in the task route.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to P8 - No Versions.ctms`.

SP to P8 - With Versions:

This task route archives versions of SharePoint documents to FileNet P8, creating a document in P8 with all the versions of each SharePoint document, filing it in a folder, and removing all but the most recent version of the document from the SharePoint server.

Collectors

This template uses the following collectors:

SP Collector
SP Collector collects documents from collection sources, which are configured to crawl and filter documents from SharePoint sites and libraries.

Tasks

This template uses the following tasks:

P8 4.x Create Version Series
This task creates a version series in P8 for the item being processed.

P8 4.x File Document in Folder
This task saves document objects to specific repository folders to facilitate search and retrieval.

SP Get Versions
SP Get Versions specifies how many versions of each SharePoint document the task route processes.

SP Post-processing
SP Post-processing determines how and if migrated documents appear in SharePoint after processing.

File name

The file name of this template is `SP to P8 - With Versions.ctms`.

Archiving templates for former IBM CommonStore users

Use these archiving templates as a basis for defining new task routes if you are a former IBM CommonStore user and now want to archive by using IBM Content Collector.

The migration task route templates for CommonStore for Exchange Server and for CommonStore for Lotus Domino simulate the archiving behavior defined in the IBM CommonStore tasks.

The migration templates cannot be used to migrate documents archived using IBM CommonStore to IBM Content Collector. The migration task route templates must be used on IBM Content Manager item types created by using IBM Content Collector.

If you used IBM CommonStore and have upgraded to IBM Content Collector, you do not have to use the migration templates. All of the archiving scenarios defined in the migration task route templates are also defined in the following templates:

- For Lotus Domino users in the “Lotus Domino with IBM Content Manager archiving templates” on page 297
- For Microsoft Exchange users in the “Microsoft Exchange with IBM Content Manager archiving templates” on page 333

If you prefer using the migration task route templates to define your archiving scenarios, select which of the supplied archiving templates best fit your use case. They all combine the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules.

- For former CommonStore for Exchange Server users:

Table 56. Typical archiving scenarios and matching templates for former users of CommonStore for Exchange Server

Typical archiving scenarios	Template files to use
Common archiving setup: • Archive all email of a given size that have been in the mailbox for a defined period of time. The archiving schedule is limited to mailboxes that exceed a given size. The task route detects duplicates and deletes all attachments after the email has been archived. • Remove archived email content in defined stages from the mailbox and stubs restored email in the mailbox. • Archive email that a user marked for archiving. The task route contains all the tasks necessary for interactive archiving, such as collecting email, preparing it for storage in the repository, and stubbing it.	• "CSX Migration - Automatic Archiving" on page 395 • "CSX Migration - Automatic Stubbing" on page 396 • "CSX Migration - Interactive Archiving" on page 397
Archiving using an offline repository: • Automatically archive email by using a delayed archiving schedule. This task route is designed to support working offline. • Automatically remove email content in defined stages to support working offline. It removes email content based on the client setup. To save space, the attachments and the body are removed from all email after users have copied them to a local repository. Regardless of whether email was copied to a local repository, attachments are removed 30 days after the email was archived.	• "CSX Migration - Offline Repository Support (Archiving)" on page 398 • "CSX Migration - Offline Repository Support (Stubbing)" on page 399

• For former CommonStore for Lotus Domino users:

Table 57. Typical archiving scenarios and matching templates for former users of CommonStore for Lotus Domino

Typical archiving scenarios	Template files to use
Common archiving setup: • Archive all email of a given size that have been in the mailbox for a defined period of time. The archiving schedule is limited to mailboxes that exceed a given size. The task route detects duplicates and deletes all attachments after the email has been archived. • Remove archived email content in defined stages from the mailbox and stubs restored email in the mailbox. • Archive email that a user marked for archiving. The task route contains all the tasks necessary for interactive archiving, such as collecting email, preparing it for storage in the repository, and stubbing it.	• "CSLD Migration - Automatic Archiving" on page 399 • "CSLD Migration - Automatic Stubbing" on page 400 • "CSLD Migration - Interactive Archiving" on page 401

Table 57. Typical archiving scenarios and matching templates for former users of CommonStore for Lotus Domino (continued)

Typical archiving scenarios	Template files to use
Archiving using an offline repository: • Automatically archive email by using a delayed archiving schedule. This task route is designed to support working offline. • Automatically remove email content in defined stages to support working offline. It removes email content based on the client setup. To save space, the attachments and the body are removed from all email after users have copied them to a local repository. Regardless of whether email was copied to a local repository, attachments are removed 30 days after the email was archived.	• "CSLD Migration - Offline Repository Support (Archiving)" on page 402 • "CSLD Migration - Offline Repository Support (Stubbing)" on page 403

CSX Migration - Automatic Archiving:

This task route archives email automatically by using Content Collector. The task route contains all the tasks necessary for automatic archiving, such as collecting email, preparing it for storage in the repository, and stubbing it. It thus combines the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules.

The task route archives email documents of a given size that have been in the mailbox for a defined period of time. Archiving is limited to mailboxes that exceed a given size. The task route detects duplicates and deletes all attachments after the email was archived.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 30 Days and Larger Than 50 KB from Mailboxes Larger Than 100 MB

This collector collects email that is older than 30 days and larger than 50 KB. Only mailboxes with a storage size larger than 100 MB are selected.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

CM 8.*x* Associate Content

This task is used to store attachments and link them back to parent (email) items.

CM 8.*x* Configure Item Types

This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CSX Migration - Automatic Archiving.ctms`.

CSX Migration - Automatic Stubbing:

You can use this task route to supplement the CommonStore Migration - Automatic Archiving task route. The task route removes archived email content in defined stages from the mailbox and stubs restored email in the mailbox. The task route removes the attachments from the email 30 days after the email was archived. Restubbing is performed two months after the email was restored.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email.

When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email 30 days after the email was archived. Restored email is restubbed after two months.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CSX Migration - Automatic Stubbing.ctms`.

CSX Migration - Interactive Archiving:

This task route archives email that the user marked for archiving by using Content Collector. The task route contains all the tasks necessary for interactive archiving, such as collecting email, preparing it for storage in the repository, and stubbing it. It thus combines the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules.

Collectors

This template uses the following collectors:

EC Collect Email Based on Interactive Requests
This collector collects email that the users marked for archiving from the mailboxes.

Specify the trigger mailbox on the Collection Sources tab.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CSX Migration - Interactive Archiving.ctms`.

CSX Migration - Offline Repository Support (Archiving):

This task route is designed to support working offline. It contains all the tasks necessary for archiving, such as collecting email and preparing it for storage in the repository. It thus combines the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules. This task route does not contain a stubbing task as the stubbing options can be defined by using the CommonStore Migration - Offline Repository Support (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 14 Days from Offline Group
This collector collects email that is older than 14 days from a group of mailboxes that are used with offline repositories.

Configure the collection source All mailboxes in a group and specify the name of the group that holds the users who enabled offline repository support.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CSX Migration - Offline Repository Support (Archiving).ctms`.

CSX Migration - Offline Repository Support (Stubbing):

This task route is designed to work with the CommonStore Migration - Offline Repository Support (Archiving) task route. It removes email content based on the client setup. To save space, the attachments and the body are removed from the email after users have copied the email documents to a local repository. Regardless of whether email was copied to a local repository, attachments are removed 30 days after the email was archived.

Collectors

This template uses the following collectors:

Collect from Offline Group
: This collector collects archived or restored email from a group of mailboxes that are used with offline repositories. When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email 30 days after the email was archived. Restored email is restubbed after two months.

  Configure the collection source All mailboxes in a group and specify the name of the group that contains the users who enabled offline repository support.

Tasks

This template uses the following tasks:

EC Extract Metadata
: This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

Remove Attachments and Body
: This task removes the attachments and the body from the email.

Use Life Cycle Collector
: This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CSX Migration - Offline Repository Support (Stubbing).ctms`.

CSLD Migration - Automatic Archiving:

This task route archives email automatically by using Content Collector. The task route contains all the tasks necessary for automatic archiving, such as collecting email, preparing it for storage in the repository, and stubbing it. It thus combines the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules.

The task route archives email documents of a given size that have been in the mailbox for a defined period of time. Archiving is limited to mailboxes that exceed a given size. The task route detects duplicates and deletes all attachments after the email was archived.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 30 Days and Larger Than 50 KB from Mailboxes Larger Than 100 MB
This collector collects email that is older than 30 days and larger than 50 KB. Only mailboxes with a storage size larger than 100 MB are selected.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CSLD Migration - Automatic Archiving.ctms`.

CSLD Migration - Automatic Stubbing:

You can use this task route to supplement the CommonStore Migration - Automatic Archiving task route. The task route removes archived email content in defined stages from the mailbox and stubs restored email in the mailbox. The task route removes the attachments from the email 30 days after the email was archived. Restubbing is performed two months after the email was restored.

Collectors

This template uses the following collectors:

Collect from Mailboxes
This collector collects archived or restored email.

When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email 30 days after the email was archived. Restored email is restubbed after two months.

Configure one of the following collection sources: All mailboxes in a group, All mailboxes on a server, Mailbox

Tasks

This template uses the following tasks:

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CSLD Migration - Automatic Rearchiving.ctms`.

CSLD Migration - Interactive Archiving:

This task route archives email that the user marked for archiving by using Content Collector. The task route contains all the tasks necessary for interactive archiving, such as collecting email, preparing it for storage in the repository, and stubbing it. It thus combines the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules.

Collectors

This template uses the following collectors:

EC Collect Email Based on Interactive Requests
This collector collects email that the users marked for archiving from the mailboxes.

Specify the Domino server and database path of the trigger mailbox on the Collection Sources tab.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

Remove Attachments
This task removes the attachments from the email.

File name

The file name of this template is `CSLD Migration - Interactive Archiving.ctms`.

CSLD Migration - Offline Repository Support (Archiving):

This task route is designed to support working offline. It contains all the tasks necessary for archiving, such as collecting email and preparing it for storage in the repository. It thus combines the configuration aspects used in IBM CommonStore, like policies, policy assignment, archive mappings, and task schedules. This task route does not contain a stubbing task as the stubbing options can be defined by using the CommonStore Migration - Offline Repository Support (Stubbing) task route that can be configured to work with this task route.

Collectors

This template uses the following collectors:

EC Collect Email Older Than 14 Days from Offline Group
This collector collects email that is older than 14 days from a group of mailboxes that are used with offline repositories.

Configure the collection source All mailboxes in a group and specify the name of the group that holds the users who enabled offline repository support.

Tasks

This template uses the following tasks:

CM 8.x Associate Content
This task is used to store attachments and link them back to parent (email) items.

CM 8.x Configure Item Types
This task is used to specify the item type and attribute-to-metadata mappings to be used by all CM8 tasks in the task route

CM 8.x Duplicate Detection
This task uses hash keys calculated by the source collectors to determine if an item is a duplicate.

CM 8.x Update Document
This task creates a record of duplicate items identified by the 'CM 8.x Duplicate Detection' task.

EC Extract Attachments
This task extracts the attachments from the documents so that they can be stored as separate files in the repository.

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

EC Prepare Email for Archiving
This task copies all email in the mailboxes and saves it as temporary files for further processing.

EC Prepare Email for Stubbing
This task saves information about the repository to the original email.

File name

The file name of this template is `CSLD Migration - Offline Repository Support (Archiving).ctms`.

CSLD Migration - Offline Repository Support (Stubbing):

This task route is designed to work with the CommonStore Migration - Offline Repository Support (Archiving) task route. It removes email content based on the client setup. To save space, the attachments and the body are removed from the email after users have copied the email documents to a local repository. Regardless of whether email was copied to a local repository, attachments are removed 30 days after the email was archived.

Collectors

This template uses the following collectors:

Collect from Offline Group
This collector collects archived or restored email from a group of mailboxes that are used with offline repositories. When stubbing is delayed in an EC Create Email Stub task, the attachments are removed from the email 30 days after the email was archived. Restored email is restubbed after two months.

Configure the collection source All mailboxes in a group and specify the name of the group that contains the users who enabled offline repository support.

Tasks

This template uses the following tasks:

EC Extract Metadata
This task extracts metadata from the documents. In addition, it creates unique hash keys for the documents so that each document is archived only once.

Remove Attachments and Body
This task removes the attachments and the body from the email.

Use Life Cycle Collector
This task removes content according to the settings in the life cycle collector.

File name

The file name of this template is `CSLD Migration - Offline Repository Support (Stubbing).ctms`.

Working with the Expression Editor

Use the Expression Editor to create or modify expressions that are used in rules or for determining document or record classes dynamically.

Using the Expression Editor, you can set up advanced rules for conditional processing of documents. You can also set up expressions in the P8 4.x Create Document and the P8 Create Version Series tasks to determine document classes dynamically, and, in the P8 4.x Declare Record task, to determine record classes dynamically.

The Expression Editor interface consists of these sections:

Expression
This section shows the expression that you are working with. This might be a single expression or an expression tree. Depending on which expression fragment you select, different edit actions are available.

Description
This section contains a description of the selected expression fragment. The description comprises the name and type of the operator, a summary of what the operator does, the number of operands for this operator, and the operand types.

Edit Actions
In this section, you can select **Edit**, **Replace**, or **Insert** from the edit actions that are available for the selected expression leaf. Depending on the selected action, different prototype expressions become available that you can use when building the expression.

Prototype Expressions
This section lists the prototype expressions that can be used for the current type of operator and the edit action that you selected.

When the selected action is **Edit**, this section is replaced with an appropriate **Edit** section, where you can enter expression values.

From the Expression Editor toolbar, you can select the following actions:

- Replace the selected expression fragment with an expression template that was saved previously
- Save part of an expression
- Save the entire expression
- Undo or redo changes
- Test the entire expression
- Test part of the expression

To edit an expression follow these steps:

1. Select an expression fragment in the expression tree. The expression fragment can be an operator or an operand.
2. Select an edit action.

Option	Description
Edit	Edit the literal or metadata value of the selected expression fragment. For Test metadata and Dynamic metadata expressions, you can also edit test values. To edit an expression value, continue with step 3.
Replace	Replace the selected expression fragment with another prototype expression. Only those expressions are available that have the same data type as the selected fragment. To replace an expression fragment in the tree, continue with step 4 on page 406.
Insert	Insert a prototype expression as a parent to the selected expression fragment. Only those expressions are available that have the same data type as the selected fragment. In addition, at least one operand of the expression that you want to insert must have the same data type as the return value of the selected expression fragment. To insert an expression in the tree, continue with step 5 on page 406.

3. Edit the selected expression value. The layout of the **Edit** section depends on the data type of the operand.

   Boolean values
   Select **True** or **False**.

   Byte, Float, or Integer values
   Enter a value or use the up and down arrows to select a value.

   Metadata
   Select a metadata source and a property to be evaluated. Different document sources create different formats, each with properties specific to its type. The properties available change depending on the selected metadata source.

   String values
   Enter a value.

Arrays

Enter a value and click the Plus button to add the value to the list of values. You can manipulate the list with the buttons to the right. To prevent duplicate values from being added to the list, select **Prevent duplicates**. Or, you can remove duplicates from the array before the expression tree is updated by clicking the respective button to the right.

To update the expression tree with the selected value or values, click **Update**.

4. To replace an expression in the expression tree, complete these steps:
   a. In the **Prototype Expressions** section, select the expression that you want to use.
   b. To update the expression tree with the selected operator, either double-click your selection or click the button at the bottom of the **Prototype Expressions** section ( ).
5. To insert an expression as a parent to the selected expression fragment, complete these steps:
   a. In the **Prototype Expressions** section, select the expression that you want to use.
   b. To update the expression tree with the selected operator, double-click your selection.
6. Repeat steps 3 on page 405, 4, or 5 for each expression that you want to modify.
7. Click **OK** to save your changes, or **Cancel** to close without saving.

If the rule or the expression has any error, the entire task route is rendered invalid and is, therefore, not run.

Prototype expressions for rules and dynamic detection of document or record classes

You can use different expressions to construct rules or to have IBM Content Collector determine document or record classes dynamically. Which operators are available depends on the type of expression that you edit.

Prototype expressions include the following items:

Literal A literal is a literal representation of a specific data type. In an expression, the value is set exactly as you specify it.

Metadata

Metadata references the property of a system or user-defined metadata source or an element in a list. Different document sources create differing formats, each with properties specific to its type. An email document, for example, has properties including the Subject and To fields. When you choose a specific metadata source, you are narrowing the properties available to just those for that source. Therefore, when you select a metadata source, you are choosing the kind of properties that will be available during processing. For **Metadata** references, the expression value is determined at run time when the property is populated by a collector or a task.

Operator

An operator defines what operation is performed on other parts of an expression. Some operators have different meanings based on the data types of the expression. Only the operators that are suitable for the data type of the selected expression leaf and for the selected edit action are

available. In a complex expression, the operators are evaluated in order of precedence. In an expression, the operands and result usually must be of the same data type.

An operator can have one or more operands. An operand can be **Literal**, **Metadata**, a value, or another expression.

Values can have the following data types:

Boolean
A Boolean value is either TRUE or FALSE.

Boolean Array
A Boolean Array contains zero or more Boolean elements.

Byte A Byte value is an 8-bit value from 0 to 255.

Byte Array
A Byte Array contains zero or more Byte elements.

Date Time
A Date Time value is expressed in UTC format and is represented in ISO 8601 date format when it is displayed as a string. The format is [+|-]YYYY-MM-DDThh:mm:ss:sTZD. For example: +2010-01-14T09:30:00:187-08:00-Pacific Standard Time

Date Time Array
A Date Time Array contains zero or more Date Time elements.

Float A Float value is a 64-bit value from -1.79769313486232E+308 to 1.79769313486232E+308.

Float Array
A Float Array contains zero or more Float elements.

Integer
An Integer value is a 64-bit value from -9223372036854775808 to 9223372036854775807.

Integer Array
An Integer Array contains zero or more Integer elements.

String A String value is represented as UTF-16 Unicode. The number of string characters is not limited; however, size will be limited by memory resources on the hosting system.

String Array
A String Array contains zero or more String elements.

Add expressions:

In **Add** expressions, the operands are summed up for Byte, Integer, and Float values. String values are concatenated.

The **Add** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 58. **Add** *expression*

Operator	Operands
Add	(Byte, Byte) (Integer, Integer) (Float, Float) (String, String)

Example

`Add(2,3)` returns 5
`Add("Snow","ball")` returns "Snowball"

AddDays expressions:

The **AddDays** expression adds the specified number of days to the value of the Date Time operand. If the number of days is a negative number, the specified number of days is subtracted from the Date Time value. The first operand is the Date Time value and the second Integer operand indicates the number of days to be added.

The **AddDays** operator has two operands.

Table 59. **AddDays** *expression*

Operator	Data types of the operands
AddDays	(Date Time, Integer)

Example

`AddDays(+2010-10-31T09:30:00:187+00:00-UTC,5)` returns `+2010-11-05T09:30:00:187+00:00-UTC`

AddMonths expressions:

The **AddMonths** expression adds the specified number of months to the value of the Date Time operand. If the number of months is a negative number, the specified number of months is subtracted from the Date Time value. The first operand is the Date Time value and the second Integer operand indicates the number of months to be added.

The **AddMonths** operator has two operands.

Table 60. **AddMonths** *expression*

Operator	Data types of the operands
AddMonths	(Date Time, Integer)

Example

`AddMonths(+2010-10-31T09:30:00:187+00:00-UTC,7)` returns `+2011-05-31T09:30:00:187+00:00-UTC`

AddYears expressions:

The **AddYears** expression adds the specified number of years to the value of the Date Time operand. If the number of years is a negative number, the specified number of years is subtracted from the Date Time value. The first operand is the Date Time value and the second Integer operand indicates the number of years to be added.

The **AddYears** operator has two operands.

Table 61. **AddYears** *expression*

Operator	Data types of the operands
AddYears	(Date Time, Integer)

Example

`AddYears(+2010-10-31T09:30:00:187+00:00-UTC,3)` returns `+2013-10-31T09:30:00:187+00:00-UTC`

Age expressions:

The **Age** expression returns the number of milliseconds between now (the current date and time) and the specified Date Time operand.

Table 62. **Age** *expression*

Operator	Operands
Age	(Date Time)

And expressions:

The **And** expression returns the value TRUE if both of the Boolean operands are TRUE; otherwise, a value of FALSE is returned.

The **And** operator has two operands. Each operand can be either a value or another expression, where the expression must return a Boolean value.

Table 63. **And** *expression*

Operator	Operands
And	(Boolean, Boolean)

Append expressions:

The **Append** expression returns a new array that contains the values from the first operand followed by the value or values from the second operand.

The **Append** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 64. **Append** *expression*

Operator	Operands
Append	(Boolean Array, Boolean) (Boolean Array, Boolean Array) (Byte Array, Byte) (Byte Array, Byte Array) (Date Time Array, Date Time) (Date Time Array, Date Time Array) (Float Array, Float) (Float Array, Float Array) (Integer Array, Integer) (Integer Array, Integer Array) (String Array, String) (String Array, String Array)

CaseInsensitiveEqual expressions:

The **CaseInsensitiveEqual** expression does a comparison between the two operands. If both operands are equal, the return value is TRUE; otherwise, it is FALSE.

The operator is not case-sensitive. You cannot use wildcard characters with this operator. When you compare arrays, both operands must have the same number of elements. Each element in the first operand must be equal to its corresponding element in the second operand, regardless of the case. The comparison operation uses the system locale to determine the casing rules for strings.

The **Case insensitive equal** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 65. **CaseInsensitiveEqual** *expression*

Operator	Operands
CaseInsensitiveEqual	(String, String) (String Array, String Array)

Ceiling expressions:

The **Ceiling** expression returns the smallest integer that is greater than or equal to the specified Float operand value.

The **Ceiling** operator has one operand, which can be either a value or another expression, where the result of the expression must be a Float value.

Table 66. **Ceiling** *expression*

Operator	Operands
Ceiling	(Float)

Example

`Ceiling(4.1234)` returns 5

Conditional expressions:

The **Conditional** expression contains a ternary operator where the first operand is a Boolean condition and the second and third operands are of one data type. The return value depends on the evaluation of the condition. If the condition equates to TRUE, the value of the second operand is returned. If the condition equates to FALSE, the value of the third operand is returned.

The **Conditional** operator has three operands, where the second and third operands can be either a value or another expression. The result of the expression must match the data type of the operand in the parent expression.

Table 67. **Conditional** *expression*

Operator	Operands
Conditional	(Boolean, Boolean Array, Boolean Array) (Boolean, Byte, Byte) (Boolean, Byte Array, Byte Array) (Boolean, Date Time, Date Time) (Boolean, Date Time Array, Date Time Array) (Boolean, Integer, Integer) (Boolean, Integer Array, Integer Array) (Boolean, Float, Float) (Boolean, Float Array, Float Array) (Boolean, String, String) (Boolean, String Array, String Array)

Contains expressions:

The **Contains** expression checks if the value of the first operand equals to any element value of the second operand. If the value of the first operand is found within the array of the second operand, the return value is TRUE; otherwise, the return value is FALSE.

The **Contains** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 68. **Contains** *expression*

Operator	Operands
Contains	(Date Time, Date Time Array) (Integer, Integer Array) (String, String Array)

ContainsSome expressions:

The **ContainsSome** expression returns TRUE if at least one element value of the first operand is exactly equal to an element value of the second operand. If there is no match, the return value is FALSE.

The **ContainsSome** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 69. **ContainsSome** *expression*

Operator	Operands
ContainsSome	(Date Time Array, Date Time Array) (Integer Array, Integer Array) (String Array, String Array)

Divide expressions:

The **Divide** expression divides the first operand by the second operand. If, for Byte or Float values, the quotient results in a fractional value, it is rounded accordingly to match the return type. Dividing by zero results in an error.

The **Divide** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 70. **Divide** *expression*

Operator	Operands
Divide	(Byte, Byte) (Float, Float) (Integer, Integer)

DynamicMetadata expressions:

The **DynamicMetadata** expression looks up a String metadata value based on the values of its operands. The first String operand is the reference to the metadata source or list that is to be searched. For the second operand, special considerations apply.

If the first operand is a metadata source, the following considerations apply for the second operand:

- A String operand must match the property ID of the metadata source's property. On a successful match, the metadata value is returned; otherwise, the expression results in an error.
- A String Array operand must match a collection of property IDs of the metadata source's property. The value of the first matching property is returned. If no match is found, the expression results in an error.

If the first operand is a list, the following considerations apply for the second operand:

- A String operand must match a List Lookup value. On a successful match, the matching list value is returned; otherwise, the expression results in an error.
- A String Array operand must match a collection of List Lookup values. On a successful match, the first matching list value is returned.

The **DynamicMetadata** operator has two operands. Each operand can also be another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 71. **DynamicMetadata** *expression*

Operator	Operands
DynamicMetadata	(String, String) (String, String Array)

Element expressions:

The **Element** expression returns the value of the first operand at the element location specified by the second operand. Because indexes are zero-based, you must specify an index of 0 to get the first element value. To get the second element value, specify an index of 1. Any specified index that is out of range causes an error.

The **Element** operator has two operands. Each operand can also be another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 72. **Element** *expression*

Operator	Operands
Element	(Byte Array, Integer) (Date Time Array, Integer) (Float Array, Integer) (Integer Array, Integer) (String Array, Integer)

Equal expressions:

The **Equal** expression returns TRUE if one of the operands are exactly equal; otherwise, FALSE is returned.

One of the following conditions must apply for the expression to return TRUE:

- The Boolean operands are exactly equal.
- The Integer operands are exactly equal.
- The Date Time operands are exactly equal. When comparing two Date Time values like these, `+2010-01-14T19:18:59:000+00:00-UTC` and `+2010-01-14T19:18:59:000-00:00-UTC`, the values are equivalent because, despite the bias being different, the offsets are both set to 00:00. In addition, if the only difference between two Date Time values is the descriptive time zone, the values are also deemed equivalent. Therefore, the following two values are also equal: `+2010-01-14T19:18:59:000+00:00-UTC` and`+2010-01-14T19:18:59:000+00:00-Pacific Time Zone`
- The String Array operands are exactly equal. Both operands must have the same number of elements. Each element in the first operand must be exactly equal to its corresponding element in the second operand.

  Here, the operator is case-sensitive. You cannot use wildcard characters.

The **Equal** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 73. **Equal** *expression*

Operator	Operands
Equal	(Boolean, Boolean) (Date Time, Date Time) (Integer, Integer) (String Array, String Array)

Floor expressions:

The **Floor** expression returns the largest integer that is less than or equal to the specified Float operand value.

The **Floor** operator has one operand, which can be either a value or another expression, where the result of the expression must be a Float value.

Table 74. **Floor** *expression*

Operator	Operands
Floor	(Float)

Example

`Floor(4.9876)` returns 4

FromString expressions:

The **FromString** expression converts the string representation of an operand to its data-type equivalent.

For the conversion, these considerations apply:

- The string representation of a Date Time operand must exactly match the required ISO 8601 format, which is [+|-]YYYY-MM-DDThh:mm:ss:sTZD.
- When you convert delimited strings to an array, use braces ({}) to delimit each entry. A delimited string looks like the ones in these examples:
  - For Boolean values: {TRUE}{FALSE}
  - For Byte values: {1}{255}{100}{0}{64}
  - For Integer values: {1}{-345}{987342}{0}{-9219384}
  - For Float values: {1}{-345.1489300012}{987342.541234}{0}{1.019999999999999900e+001}{-1.999E+123}{1.999E-123
  - For String values: {The}{quick}{brown}{fox}{jumps}{over}{the}{lazy}{dog}
  - For Date Time values: {+2010-01-15T12:11:22:000+00:00-UTC}{+2010-01-15T04:11:22:000-08:00-Pacific Standard Time}

If the conversion fails, an error results.

The **From string** operator has one operand, which can be either a value or another expression, where the result of the expression must be a String value.

Table 75. **FromString** *expression*

Operator	Operands
FromString	(String)

GreaterThan expressions:

The **GreaterThan** expression returns TRUE if the first operand has a greater value than the second operand; otherwise, FALSE is returned. String comparisons are case-sensitive.

For array comparison, the comparison operation uses dictionary-style sorting for comparison between the two array operands. The elements are compared in order until an unequal pair is found. If the element from the first operand is greater than the element from the second operand, the return value is TRUE; otherwise, it is FALSE. If the first operand has more elements and all pairs are equal, TRUE is returned. If the first operand has less elements and all pairs are equal, FALSE is returned.

The **GreaterThan** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 76. **GreaterThan** *expression*

Operator	Operands
GreaterThan	(Date Time, Date Time) (Date Time Array, Date Time Array) (Float, Float) (Integer, Integer) (Integer Array, Integer Array) (String, String) (String Array, String Array)

GroupLookup expressions:

The **GroupLookup** expression performs an LDAP query. The query returns the groups that are associated with the user or the list of users defined in the first String operand.

If you set the second Boolean operand to TRUE, a recursive search is done for all groups associated with the user. If you set the second Boolean operand to FALSE, a search is done for only those groups of which the user is a direct member. For Group Lookup, the LDAP lookup settings for the Task Routing Engine must be configured accordingly.

The **GroupLookup** operator has two operands, where the first operand can be either a value or another expression. The result of the expression must be a String value or a String Array value.

Table 77. **GroupLookup** *expression*

Operator	Operands
GroupLookup	(String, Boolean) (String Array, Boolean)

Intersection expressions:

The **Intersection** expression checks both operands for equal elements. A new array is created with the values that exist in both arrays.

The **Intersection** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 78. **Intersection** *expression*

Operator	Operands
Intersection	(Boolean Array, Boolean Array) (Date Time Array, Date Time Array) (Integer Array, Integer Array) (String Array, String Array)

IsLikeIn expressions:

The **IsLikeIn** expression checks if the value in the first String operand is similar to any element value in the second String Array operand.

The operator is not case-sensitive and you cannot use wildcard characters with this operator. The comparison returns TRUE on the first match. Therefore, the order of the array elements might be important.

The **Is like in** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 79. **IsLikeIn** *expression*

Operator	Operands
IsLikeIn	(String, String Array)

Example

`IsLikeIn ("bC","ABcdEF")` returns TRUE

Length expressions:

The **Length** expression returns the number of elements found in the array.

Table 80. **Length** *expression*

Operator	Operands
Length	(Boolean Array) (Byte Array) (Date Time Array) (Float Array) (Integer Array) (String Array)

LessThan expressions:

The **LessThan** expression returns TRUE if the first operand has a lesser value than the second operand; otherwise, FALSE is returned. String comparisons are case-sensitive.

For comparison between the two array operands, the comparison operation uses dictionary-style sorting. The elements are compared in order until an unequal pair

is found. If the element from the first operand is less than the element from the second operand, the return value is TRUE; otherwise, it is FALSE. If the first operand has less elements and all pairs are equal, TRUE is returned. If the first operand has more elements and all pairs are equal, FALSE is returned.

The **Less than** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 81. **LessThan** *expression*

Operator	Operands
LessThan	(Date Time, Date Time) (Date Time Array, Date Time Array) (Float, Float) (Integer, Integer) (Integer Array, Integer Array) (String, String) (String Array, String Array)

Like expressions:

The **Like** expression returns TRUE if the first String value exists within the second String value.

This operator is not case-sensitive. You cannot use wildcard characters with this operator.The **Like** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 82. **Like** *expression*

Operator	Operands
Like	(String, String)

Example

`Like("bC","ABcdEF")` returns TRUE

Modulo expressions:

The **Modulo** expression divides the first operand by the second operand and returns the remainder of the division.

The **Modulo** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 83. **Modulo** *expression*

Operator	Operands
Modulo	(Byte, Byte) (Integer, Integer)

Example

`Modulo(7,3)` returns 1
while
`Modulo(9,3)` returns 0

Multiply expressions:

The **Multiply** expression multiplies the first operand by the second operand. An error results if the product is larger than the bounds defined by the data type.

The **Multiply** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 84. **Multiply** *expression*

Operator	Operands
Multiply	(Byte, Byte) (Float, Float) (Integer, Integer)

Narrow expressions:

The **Narrow** expression down converts the operand to a value that fits the data type that is defined by the parent expression.

The conversion is done by cutting off the underlying bits:

- A Float operand to an Integer value
- An Integer operand to a Byte value or a Float value

The **Narrow** operator has one operand.

Table 85. **Narrow** *expression*

Operator	Operands
Narrow	(Float) (Integer)

Not expressions:

The **Not** expression contains a unary negation operator that returns TRUE if its operand is false, and FALSE if its operand is true.

Table 86. **Not** *expression*

Operator	Operands
Not	(Boolean)

Or expressions:

The **Or** expression contains a logical operator that returns the value TRUE whenever one or more of its Boolean operands are true; otherwise, a value of FALSE is returned.

Table 87. **Or** *expression*

Operator	Operands
Or	(Boolean, Boolean)

Example

`Or(A,B)` is true if *A* is true or if *B* is true, or if *A* and *B* are true.

RegexSearch expressions:

The **RegexSearch** expression uses a Perl regular expression to check whether a certain string exists.

The first operand is the regular expression to match, that is, the pattern to search for. The second operand is the value to be analyzed. If the second operand is an array, the elements of the array are searched for the regular expression in sequence until the first match is found.

If a match is found, the return value is TRUE. Otherwise, the return value is FALSE. Missing or invalid regular expressions in the first operand result in an error.

For examples, see the topic about regular expressions.

Table 88. **Regex search** *expression*

Operator	Operands
RegexSearch	(String, String) (String, String Array)

Related concepts

"Regular expressions" on page 421

RegexSubstitute expressions:

The **RegexSubstitute** expression uses a Perl regular expression to substitute the specified string with a replacement string.

The first operand is the pattern to be searched for, the second operand is the String value to be searched, and the third operand is the replacement string.

For examples, see the topic about regular expressions.

Table 89. **RegexSubstitute** *expression*

Operator	Operands
RegexSubstitute	(String, String, String)

Related concepts

"Regular expressions" on page 421

Slice expressions:

The **Slice** expression returns a section of the first operand.

If the first operand is a string, the operation cuts out a section of a string and returns a new string containing that section. The first operand is the string from which the section is cut. The section that is cut out starts at the specified zero-base index of the second Integer operand. The third Integer operand specifies the length of the section to be cut.

If the start index is a negative value, an error occurs. If the start index is a value greater than the number of characters in the string, an empty string is returned. If the length value is a negative number, the value is ignored and all characters specified after the start index are returned.

If the first operand is an array, the operation creates a new array by cutting out a selection of elements of the first operand. The section of the array that is cut out starts at the specified zero-base index of the second Integer operand. The third Integer operand specifies the number of elements to be cut.

If the start index is a negative value, an error occurs. If the start index is a value greater than the number of elements in the array, an empty array is returned. If the length value is a negative number, the value is ignored and all elements specified after the start index are returned.

Table 90. **Slice** *expression*

Operator	Operands
Slice	(Boolean Array, Integer, Integer) (Byte Array, Integer, Integer) (Date Time Array, Integer, Integer) (Float Array, Integer, Integer) (Integer Array, Integer, Integer) (String, Integer, Integer) (String Array, Integer, Integer)

Subtract expressions:

The **Subtract** expression subtracts the first operand from the second operand.

The **Subtract** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 91. **Subtract** *expression*

Operator	Operands
Subtract	(Byte, Byte) (Float, Float) (Integer, Integer)

TripleDESEncrypt expressions:

The **TripleDESEncrypt** expression uses Triple Data Encryption Algorithm (TDEA) block cipher encryption to encrypt a string value. The first operand is the string value to be encrypted. The second operand is the password to be used as the encryption key.

The **TripleDESEncrypt** operator has two operands. Each operand can be either a value or another expression, where the result of the expression must match the data type of the operand in the parent expression.

Table 92. **TripleDESEncrypt** *expression*

Operator	Operands
TripleDESEncrypt	(Date Time Array, Date Time Array) (Integer Array, Integer Array) (String , String)

TestMetadata expressions:

The **TestMetadata** expression checks whether a specific metadata property exists. The first operand specifies the metadata source. The second operand specifies the metadata property. Evaluation is done at run time. If the metadata property exists, TRUE is returned; otherwise, the return value is FALSE.

Table 93. **TestMetadata** *expression*

Operator	Operands
TestMetadata	(String, String)

ToString expressions:

The **ToString** expression converts the operand to a string representation of the respective data type.

Float values are always shown in scientific notation.

Table 94. **ToString** *expression*

Operator	Operands
ToString	(Boolean) (Byte) (Date Time) (Float) (Integer) (String)

Example

`ToString(10.2)` returns 1.019999999999999900e+001

Widen expressions:

The **Widen** expression up converts the Byte operand to a value that fits the Float data type.

Table 95. **Widen** *expression*

Operator	Operands
Widen	(Byte)

Regular expressions

A regular expression is a text string that describes a search pattern (also known as *regexp* or *regex*).

Using regular expressions is similar to using typical keyword searches with wildcard characters. For example, you can use wildcard characters in a search

string, such as *.doc, to find documents that have the .doc extension. Using regular expressions is also similar to using wildcard characters. However, by using regular expressions, you can get more precise results or a broader range of results depending on the regular expression that you use.

In IBM Content Collector, you can use regular expressions when you configure rules or property mappings in task routes.

The Configuration Manager and the underlying architecture rely on the open source Boost C++ Libraries, Version 1.37, Perl syntax, for evaluating regular expressions.

You can use the following types of regular expressions when you configure rules or property mappings in task routes:

- Regular expressions for finding matching patterns
- Regular expressions for replacing strings

Matches regular expressions

With a matches regular expression, you can search for a specified pattern and check whether a certain string exists within the metadata property that you are analyzing. Such an expression evaluates to true if a match is found for the defined pattern and to false if not. In regular expressions, some characters require the backslash as an escape character to mark them as literal characters.

You can use this type of regular expressions when you configure rules for conditional processing of email or files and when you configure property mappings.

Replacement regular expressions

With a replacement regular expression, you can take part of the specified string and replace it with something else. For example, to replace the forward slash (/) in dates with a hyphen (-), specify the forward slash as the replacement expression and the hyphen as the replacement string. The replacement expression is the pattern to be searched for. The replacement string is the string with which you want to replace the returned text string.

You can use this type of regular expressions only when you configure property mappings and only for system and custom metadata properties that have a string data type.

You can also use a replacement regular expression to remove part of the string by simply replacing it with nothing. For example, to remove hyphens, parentheses, or other characters from a phone number, you can use the following characters in your regular expression: `[- +\(\)\.]`. Use the brackets to search for any of the characters that are typed inside those brackets. Some of the characters, such as the period, require the backslash as an escape character because these characters are special characters in regular expressions. They need the escape character to mark them as literal characters. In the phone number example, you want to match a period literally.

If you want to keep part or all the original string, use the *$n syntax*. Using the `$n` value in the replacement string denotes that you want to keep whatever is enclosed in the nth set of parentheses and use that as part of your replacement

string. For example, the format of contract numbers in your company might be six numbers, a hyphen, and the company code, which consists of four letters, such as `123456-ABCD`. The contract number is part of the email that is to be archived and is mapped to the custom metadata property `ContactNumber`. In your archiving task route, you want to use the company code to define the folder where the email will be stored. Therefore, you must keep the letters, which is the second part of the string, remove the numbers and the hyphen, and add more information to build the folder path. The regular expressions that you use in your task route are as follows:

Replacement regular expression

```
.*\d{6}-([A-Z]{4}).*
```

This expression finds a contract number, which is six digits followed by a hyphen followed by four letters, and anything that might appear before or after the pattern.

Replacement string

```
Contracts/$1/Email
```

This replacement string is how the folder path is to be built. Because the $1 value is used, whatever the regular expression matched within the parentheses is preserved and returned.

Result A match is found for `ICC/Contracts/DEF/123456-ABCD/Sales`. `ABCD` is the value that is returned for $1. Therefore, the folder path is `Contracts/ABCD/Email`.

You can also use a regular expression to rearrange the parts of the contract number and build a more advanced folder structure as in the following example:

Replacement regular expression

```
.*(\d{6})-([A-Z]{4}).*
```

This expression finds a contract number, which is six digits followed by a hyphen followed by four letters, and anything that might appear before or after the pattern.

Replacement string

```
Contracts/$2/$1/Email
```

This replacement string is how the folder path is to be built. For the $1 value, whatever the regular expression matched within the first set of parentheses is preserved and returned. For the $2 value, whatever the regular expression matched within the second set of parentheses is preserved and returned.

Result A match is found for `ICC/Contracts/DEF/123456-ABCD/Sales`. `123456` is the value that is returned for $1, `ABCD` is the value that is returned for $2. Therefore, the folder path is `Contracts/ABCD/123456/Email`.

Regular expression syntax:

Regular expressions are written in a formal language that can be interpreted by a regular expression processor. You can use special characters and character sets in regular expressions to configure rules or property mappings in task routes.

Characters in regular expressions match a single instance of themselves with these exceptions:

- The special characters: \ [^ $ | ? * + ( )
- Characters that are defined in character classes

A character class is a defined set of characters and is enclosed in brackets. Any character that is specified in a character class except the special characters \ ^ -] adds that character to the possible matches for the character set.

A bracket expression can contain any of these elements:

Any combination of single characters
For example, `[abc]` matches any of the characters `a`, `b`, or `c`.

Character ranges
For example, `[a-c]` matches any single character in the range of a to c.

Negations
For example, `[^a-c]` matches any character that is not in the range of a to c.

Predefined character classes
For example, `[[:lower:]]` matches any lowercase character.

Escape characters
For example, `[\^]` matches ^.

The range of a character class also depends on the locale of the machine where you run the task route.

You can use the following operations to construct regular expressions:

Alternation
A vertical bar (`|`) separates alternatives.

Grouping
Parentheses are used to define the scope and precedence of the operators.

Lookaround
Lookaround constructs are also called *zero-width assertions*. They actually match characters, but return only the result match or no match. The lookaround constructs are *lookahead* and *lookbehind*.

Positive and negative lookahead
Use positive lookahead to match something that follows a given pattern without making the pattern part of the match: `match(?=pattern)`. Use negative lookahead if you want to match something that is not followed by something else: `match(?!pattern)`.

Lookahead is typically used to create the logical AND of two regular expressions, for example, if a password must contain a lowercase letter, an uppercase letter, no punctuation marks, and be at least 8 characters long, you could use the following expression to validate the password:

```
(?=.*[[:lower:]])(?=.*[[:upper:]])(?!.*[[:punct:]]).{8,}
```

Positive and negative lookbehind
Lookbehind has the same effect as lookahead but works backwards. Positive lookbehind matches something that is preceded by a given pattern: `(?<=pattern)match`. Negative lookbehind matches something that is not preceded by a given pattern: `(?<!pattern)match`.

Quantification

A quantifier, such as the question mark (?), the asterisk (*), or the plus sign (+) after a token, such as a character or group, specifies how often the element that precedes it is allowed to occur. The standard quantifiers in regular expressions are *greedy*, meaning they match as much as they can. For example, a pattern of `/.*/` that is applied to `abc/123/xyz_6/7` returns `123/xyz_6` instead of `123` because with greedy quantification, as many characters as possible are returned. To avoid this problem, you can specify a quantifier as *lazy*, which is also known as *nongreedy*, by putting a question mark after the quantifier. With lazy quantification, the expression tries the shortest match first.

You can combine these constructions to form complex expressions.

When you set up regular expressions for matching operations, you can use several modifiers to determine how a regular expression is interpreted:

i Match a pattern regardless of the case.

m Treat the string as multiple lines. In this mode, the caret and the dollar sign match the start or end of any line anywhere within the string.

s Treat string as single line. In this mode, the period matches any character, even a newline character.

x Extend your pattern's legibility by permitting whitespace and comments.

These are usually written as `/modifier`, even though the delimiter in question might not really be a slash. You can also use any of these modifiers within the regular expression itself by using a `(?modifier)` construct, for example:
`(?i)car matches car and CAR.`

Operators are evaluated in the following order:

1. Collation-related bracket symbols: `[==] [::] [..]`
2. Escaped characters: `\`
3. Character set (bracket expression): `[]`
4. Grouping: `()`
5. Quantifiers: `* + ? {m,n}`
6. Concatenation
7. Anchoring: `^ $`
8. Alternation: `|`

Remember: The case of the character or character class matters in some cases.

Table 96. Regular expression syntax

Character or character class	Description	Example
`\`	The backslash escapes special characters so that they are treated as literals.	`\+` matches +
`[ ]`	The brackets enclose a character class.	`[abc]`

Table 96. Regular expression syntax (continued)

Character or character class	Description	Example
^	The caret matches at the start of the string to which the regular expression is applied. If a bracket expression begins with the caret, it matches the complement of the characters it contains (negation).	`^.` matches `a` in `abc/xyz` `[^x-z]` matches any character that is not in the range x to z
$	The dollar sign matches the end of the string to which the regular expression is applied.	`.$` matches `z` in `abc/xyz`

|	The vertical bar is used for alternatives and matches either of its arguments.
	
	Parentheses are used to group alternatives.
		`abc|def|xyz` matches `abc`, `def`, or `xyz`
		
		`abc|(def|xyz)` matches `abcdef` or `abcxyz`

?	The question mark makes the item that directly precedes it optional. It is also used to set up lazy quantification.	`abc?` matches `ab` or `abc`
*	The asterisk indicates zero or more occurrences of the item that directly precedes it.	`ab*c` matches `ac`, `abc`, `abbc`, or `abbbc`
+	The plus sign indicates one or more occurrences of the item that directly precedes it.	`ab+c` matches `abc`, `abbc`, or `abbbc`, but not `ac`
()	Parentheses are used to group the parts of the regular expression.	`gr(a|e)y` matches `gray` or `grey`
-	The hyphen specifies a range of characters unless it is specified immediately after an opening bracket. In this case, it is used literally.	`[A-Za-z0-9]` matches any letter or digit
.	The period matches any single character.	
(?=*pattern*)	This construct matches something that is followed by a given pattern.	`a(?=b)` matches the `a` and only the `a` in `cab`, but does not match `bath` or `bar`.
(?!*pattern*)	This construct matches something that is not followed by a given pattern.	`a(?<!b)` matches the `a` and only the `b` in `bath` or `bar`, but does not match the `a` in `cab`.
(?<=*pattern*)	This construct matches something that is prededed by a given pattern.	`(?<=a)b` matches the `b` and only the `b` in `cab`, but does not match `bed` or `debt`.
(?<!*pattern*)	This construct matches something that is not prededed by a given pattern.	`(?<!a)b` matches the `b` and only the `b` in `bed` or `debt`, but does not match the `b` in `cab`.
\<	This escape sequence matches a the start of a word.	`\<ton` matches `tons` but not `button`.
\>	This escape sequence matches a the end of a word.	`\>ton` matches `button` but not `tons`.

Table 96. Regular expression syntax (continued)

Character or character class	Description	Example
\Q...\E	This sequence matches the characters between \Q and \E literally.	\Q+-*/\E matches +-*/
\d, \w, and \s	These shorthand character classes match the digits 0 - 9, word characters (letters, digits, and the underscore), and white space. \d is equivalent to [:digit:]. \w is equivalent to [:word:]. \s is equivalent to [:space:].	[\d\s] matches a character that is a digit or white space
\D, \W, and \S	These shorthand character classes are negated versions of the character classes that match digits, word characters, or white space. \W is equivalent to ^ [:word:]. \S is equivalent to ^ [:space:].	\D matches a character that is not a digit
\b	This shorthand character class matches a word boundary (the start or end of a word) unless it is specified inside a character class. In this case, \b is a backspace character.	.\b matches c in abc
\B	This shorthand character class matches only when it is not at a word boundary.	\B.\B matches y in xyz
\A	This shorthand character class matches at the start of the string to which the pattern is applied.	\A. matches a in abc
\z	This shorthand character class matches the end of the string to which the pattern is applied.	\z. matches c in abc
{n}	This quantifier repeats the item that directly precedes it exactly *n* times, where *n* is an integer equal to or greater than 1.	a{3} matches aaa
{n,m}	This quantifier repeats the item that directly precedes it between *n* and *m* times, where *n* is an integer equal to or greater than 0 and *m* is an integer equal to or greater than *n*.	a{2,4} matches aaaa, aaa, or aa
{n,}	This quantifier repeats the preceding item at least *n* times, where *n* is an integer equal to or greater than 0.	a{2,} matches aaaaa in aaaaabc
[:alphanum:]	This character class matches any alphanumeric character.	
[:alpha:]	This character class matches any alphabetic character.	

Table 96. Regular expression syntax (continued)

Character or character class	Description	Example
[:blank:]	This character class matches any white space that is not a newline character.	
[:cntrl:]	This character class matches any control character, for example, the newline character or the backspace character.	
[:digit:]	This character class matches any decimal digit. [:digit:] is equivalent to \d.	
[:graph:]	This character class matches any graphical character: alphanumeric or punctuation.	
[:lower:]	This character class matches any lowercase character.	
[:print:]	This character class matches any printable character: alphanumeric, punctuation, or space.	
[:punct:]	This character class matches any punctuation character.	
[:space:]	This character class matches any white space, such as the blank character, the newline character, or the tab character. [:space:] is equivalent to \s.	
[:upper:]	This character class matches any uppercase character.	
[:word:]	This character class matches any word character: letters, digits, and the underscore. [:word:] is equivalent to \w.	
[:xdigit:]	This character class matches any hexadecimal digit.	

Regular expression examples:

Examples of common regular expressions show you how you can set up regular expressions to find text patterns or to find patterns and replace parts of the returned strings. You can use these sample patterns and adapt them to your needs.

Matches regular expressions

The following table contains examples of regular expressions that can be used in match mode.

Table 97. Matches regular expression examples

Purpose	Regular expression	Sample text	Sample match
Match a string of numbers of fixed length.	`\d{3}`	Contract Number 12-345 AB12	345
Match a string of any characters of a specific length. The string can consist of the characters a-z and the digits 0 - 9.	`\w{8}`	Contract Number 12-345 AB12	Contract
Match a string of any characters of a specific length at the beginning. The string can consist of the characters a - z and the digits 0 - 9.	`^\w{6}`	Contract Number 12-345 AB12	Contra
Match a string of any two characters of fixed length followed by two digits.	`\w{2}\d{2}`	Contract Number 12-345 AB12	AB12
Match a word that is of fixed length with the assumption that the word is followed by a space.	`\w{8}\s`	Contract Number 12-345 AB12	Contract
Match a string of numbers of fixed length that also contains specific characters, for example, a contract number that consists of six characters with a hyphen after the second number.	`\d{2}-\d{3}`	Contract Number 12-345 AB12	12-345
Match a string of at least three numbers.	`\d{3,}`	Contract Number 12-345 AB12	345
Match the first folder in a path.	`^([[:word:]]|\ s)*(\\\|\/)`	Folder 1\Folder 2\Folder 3\Folder 4	Folder 1

Replacement regular expressions

The following table shows examples of regular expressions that can be used in replacement mode.

Table 98. Replacement regular expression examples

Purpose	Regular expression	Replacement string	Sample text	Sample result
Get a folder path without a drive letter.	`^[^\\]*`	`$1`	C:\folder 1\folder 2	\folder 1\folder 2
Get a drive letter from a folder path.	`(\\.*)`		C:\folder 1\folder 2	C:

Table 98. Replacement regular expression examples (continued)

Purpose	Regular expression	Replacement string	Sample text	Sample result
Get a specific folder in a path with a drive letter. For each section in a path with a drive letter in it, repeat the expression:`([^\\]*)\\?` with the value:`.*` at the end of the expression. Use `$<section number>` to get the specific level required, where section one is the drive letter, section two is the first folder in the path, section three is the second folder in the path, and so on.	`([^\\]*)\\?.*`	`$1`	C:\one\two\three	C:
	`([^\\]*)\\?([^\\]*)\\?.*`	`$2`	C:\one\two\three	one
	`([^\\]*)\\?([^\\]*)\\?([^\\]*)\\?.*`	`$3`	C:\one\two\three	two
Get the second folder in a path.	`^[\\/]?[^\\/]+[\\/]([^\\/]+)([\\/][^\\/]+)*`	`$1`	Folder 1\Folder 2\Folder 3\Folder 4	Folder 2
Get the last two folders in a path.	`((\\\|\/)([[:word:]]|\s*)*){2}$`	`$1`	Folder 1\Folder 2\Folder 3\Folder 4	Folder 3\Folder 4
Get all email with a case number matching the pattern "eight digits followed by a hyphen followed by three uppercase letters." Replace the case number with the phrase `Automobile claim`.	`(.*)(\d{8}-[A-Z]{3})(.*)`	`$1Automobile claim$3`	98765432-DEF, your email dated August 17, 2008	Automobile claim, your email dated August 17, 2008
Get all email with specific originator addresses and add the respective company name.	`(.?\Q@example.\E)(com|org|net)`	`$1$2 (Example Company)`	Message forwarded by `X@example.org`	Message forwarded by `X@example.org (Example Company)`
Search for IDs starting with `AB` and replace them with the department title `Controlling`.	`AB\w{2,4}@example\.com`	`Controlling`	Sent by `AB12@example.com`	Sent by Controlling
Remove any strings that match the pattern "four or more digits enclosed in parentheses."	`\(\d{4,}\)`		Item number (12345) 6789	Item number 6789
Delete the forward prefix `Fw:` or the reply prefix `Re:` from the subject.	`^(Fw:|Re:)(.*)`	`$2`	Fw: Your email regarding case number 98765432-DEF, dated August 17, 2008	Your email regarding case number 98765432-DEF, dated August 17, 2008

Table 98. Replacement regular expression examples (continued)

Purpose	Regular expression	Replacement string	Sample text	Sample result
Truncate the value of the selected property to 80 characters, for example, the subject of an email.	`^(.{0,80}).*$`	`$1`	Network maintenance - Service interrupt Wednesday Feb. 6, 2008 - Monthly service interrupt February	Network maintenance - Service interrupt Wednesday Feb. 6, 2008 - Monthly service

Using extended processing functions

IBM Content Collector provides some intrinsic functions that you can use by inserting and configuring the respective tasks in your task routes. Additionally, you can integrate functions of other products into IBM Content Collector. The following topics give some background information on these functions.

Enabling the collection of additional archiving information

You can enable IBM Content Collector to collect additional archiving information for email that is archived manually.

To configure IBM Content Collector accordingly you have to complete these tasks. Perform steps 1 to 9 on page 432 on the machine where Content Collector is installed. Step 10 on page 432 is required for the proper setup on client side.

1. Adapt the credentials for the Derby database that is used to temporarily store the additional archiving information. The user ID and the password that can be used to access the Derby database are defined in the `derby.properties` file in the respective Derby database instance. If you use the Derby database that is embedded in Content Collector, the `derby.properties` file is located in the directory `<ICCINstallPath>\derby\10.3.3.0\bin`, where `<ICCINstallPath>` is the IBM Content Collector installation directory.
   a. To add or change entries for users who are to have access to the Derby database, edit the properties file.
   b. Change the line that follows the comment `# Users definition` or add further entries. The syntax is `derby.user.<username>=<password>`. Replace *<username>* and *<password>* with proper values.
   c. Restart the IBM Content Collector Metadata Form Database service.
2. Create new user-defined metadata sources for all additional archiving information that you want to collect when a user manually archives email. To make the information available in the repository, add the metadata to the IBM Content Manager item type or FileNet P8 document class:
   a. Create the additional IBM Content Manager attributes or FileNet P8 properties.
   b. Add the attributes to the ICCEmailInstance child component of the item type or document class for the compound email model.

   The IBM Content Manager attributes or FileNet P8 properties must have the same data type as the user-defined metadata. Otherwise, you will not be able to map the attributes or properties to the metadata in the CM 8.x Configure Item Types task or P8 4.x Create Document tasks in the task route.
3. Configure the Metadata Form connector.

4. Verify the settings for the Metadata Web Application. If required, modify the settings.
5. Import the metadata form template. A default template is provided with the product. It is named `form.zip` and is located in the `formTemplates` folder of your Content Collector server installation.
6. Configure the metadata form definition.
7. Edit the client configuration to specify the folders for which a user can specify additional archiving information The folders are created by Content Collector. You must close and reopen your mail file before you can see the folders.
8. Create a task route where additional archiving information is collected.
   a. Set up the collector for this task route. Specify one or more folders for which additional archiving information will be collected as the collection source.
   b. Include the MC Retrieve Additional Metadata task in the task route.
   c. Include the CM 8.x Configure Item Types task or the P8 4.x Create Document task in the task route.
9. If you added the user-defined metadata to the IBM Content Manager item type or the FileNet P8 document class and you want to enable search on these custom attributes, you have to configure the access to archived data accordingly.
   - For IBM Content Manager, add the new fields to the model and the configuration files of indexer for text search and recreate the index for the item type. Then, update the configuration files for the archived data access for email.
   - For FileNet P8, adapt the `style.xml` file and make sure to include a P8 4.x Save Prepared Text as XML task in your task route. On the **Custom XIT Metadata** tab, map the user-defined metadata to XIT elements so that the information is added to the index for full-text search.
10. Select **Specify additional archiving information** when you install the Content Collector Outlook Extension or when you enable a Domino template for Content Collector.

Related concepts

"Enabling search on custom attributes" on page 698

Related tasks

"Adding and editing user-defined metadata" on page 240
"Configuring the Metadata Form Connector" on page 210
"Modifying the settings for the Metadata Web Application" on page 233
"Selecting the metadata form template" on page 234
"Configuring the metadata form definition" on page 235
"Modifying client configuration settings" on page 227
"Collecting email on request" on page 472
"Adding attributes to the model file" on page 614
"Adding attributes to the configuration file of the item type" on page 615
"Collecting SMTP documents" on page 478

Related reference

"MC Retrieve Additional Metadata" on page 539

Archiving email from local files

To save space on the email servers, users often create local archive files for their email. As these files are located on the client machines and not on the email server,

they are not under the control of IBM Content Collector. To archive email from existing local files, you must prepare the files for Content Collector and configure Content Collector to collect them.

Email can be stored locally in personal storage (PST) files for Microsoft Exchange or in Notes Storage Facility (NSF) archives for Lotus Domino. To archive these files for compliance, they must be made accessible to IBM Content Collector, so that they can be collected by a task route. After processing, the old files should be deleted or marked as archived.

To archive email from local files:

1. Prepare the local files for archiving. This step differs for Microsoft Exchange PST files and Lotus Domino NSF files. See the related topics for detailed information.
2. Create a task route that collects the NSF or PST files and processes them. Use the NSF Archive Migration - Archiving template for Lotus Domino or the PST Migration - Archiving template for Microsoft Exchange.
3. Optional: On the **Collection Sources** tab of the collector configuration, select **Create status information for local NSF files** or **Create status information for local PST files**. If you select this option, Content Collector creates an XML metadata file for every email file that is archived. These metadata files are written to the source directory and contain status information about the processing result.
4. Optional: Create a task route to do postprocessing on the local files, for example to delete the empty files, according to the information in the XML metadata files. To create a task route for postprocessing:
   a. Add a file system collector and configure it to collect from the folder that contains the local files.
   b. Create a set of XML file system metadata that contains the status information from the XML metadata files that you want to use. The following information is available in the XML metadata files:

XPath	Data type	Description
`/containermetadata/state/lastchange`	Date Time	The time when the file has been changed
`/containermetadata/state/empty`	Boolean	A flag that indicates whether the local file (NSF or PST) is empty. Documents in the Drafts folder and, for PST files only, in the Trash folder are ignored.
`/containermetadata/state/processed@id`	String	The ID of the collector that last worked on the file
`/containermetadata/state/processed`	Boolean	A flag that indicates whether the collector that last worked on the file found new content to process
`/containermetadata/state/errors`	Numeric	The number of documents that caused errors and are still contained within the file

   c. Add an FSC Associate Metadata task to extract the status information from the XML metadata files.

d. Add decision points, rules, and FSC Post Processing tasks to process the local files according to their status.

Related tasks

"Preparing Notes Storage Facility files for archiving"

"Preparing personal storage files for archiving" on page 435

"Adding and editing file system metadata" on page 241

Related reference

"PST Migration - Archiving" on page 347

"PST Migration - Archiving" on page 367

"NSF Archive Migration - Archiving" on page 310

"NSF Archive Migration - Archiving" on page 329

Preparing Notes Storage Facility files for archiving:

Users must upload their local Notes Storage Facility (NSF) archives to a shared directory from where they are processed. They can do this by clicking a button in a note sent out by the administrator.

Tip: When using the automatic upload feature, no manual work is required to prepare the NSF files for archiving. However, programs for manually finding and tagging NSF files are provided and can be used to inspect the files. The programs **findNsf** and **tagNsf** are located in *InstallPath*/ctms. The usage is parallel to the programs **findPst** and **tagPst**.

To request users to upload their local NSF archives for archiving:

1. If the action **Enable Local Archive Databases for IBM Content Collector Archiving** is not available in the menu of the administrator's Lotus Notes mail database, enable it:
   a. Select the mail database of the IBM Content Collector administrator and replace the database design by the mail template that is enabled for IBM Content Collector.
   b. Open the administrator's mail database in the Domino Designer.
   c. Go to **Shared Code** → **Agents**, right-click **Enable Local Archive Databases for IBM Content Collector Archiving** and select **Design properties**.
   d. On the **Design** tab, clear all check boxes under **Hide design elements from**.
2. From the Lotus Notes **Actions** menu, select **Enable Local Archive Databases for IBM Content Collector Archiving**.
3. Specify the required information:
   - Specify the user name that the Email Connector uses to connect to Lotus Domino. For hierarchical names, enter the abbreviated hierarchical name.
   - Specify the location of the shared directory that is used to temporarily store the local NSF archives during processing. Use the Universal Naming Convention (UNC) format for the folder path for Microsoft Windows users, or a path to a mounted directory for Mac users. The share directory must be accessible by all workstations that upload NSF files and by IBM Content Collector (must have read, write, and modify rights).

   An email document that contains a button for uploading local NSF archives is generated and displayed.
4. Edit the text of the email document and send it out to the recipients that you want to upload their local NSF archives.

5. When users receive the email, they can click on the button and select an NSF file. This file is then copied to the server. If the file already exists on the server, the user is prompted to confirm if the file should be replaced. If the file has been processed before, only documents that have been added or changed since then are processed. Encrypted files are automatically decrypted before they are uploaded. To make sure that only Content Collector can access the files that are uploaded to the server, all files in the shared directory that is used to temporarily store the local NSF archives are encrypted using the user account of the Email Connector service.

   Important: When copying an NSF file to the server, the client should be in the same local network as the server. Otherwise, copying can take a long time.

Related tasks

"Archiving email from local files" on page 432

"Replacing the Lotus Notes mail template in all mailboxes" on page 106

"Collecting documents automatically" on page 460

Related reference

"findPst" on page 436

"tagPst" on page 437

Preparing personal storage files for archiving:

IBM Content Collector can identify the owner or creator of a document only through the mailbox account. Typically, personal storage (PST) files do not contain this information. Therefore, you must prepare the PST files in a way that this information is included in the PST files before they can be archived.

To enable the owners or creators of PST files to view or restore archived PST documents, the PST files must be associated with their email addresses. However, if PST files are added to a user's Outlook mailbox, the Outlook extension automatically associates PST files with the email address when the PST files are added to the user's Outlook client mailbox. Thus, the owner can still be identified even if the PST file is removed from the mailbox.

You use two programs to prepare personal storage files for archiving. The first program finds the PST files and lists them in a CSV file. In this CSV file, you can add the appropriate email addresses so that the PST files are associated with the proper users.

The second program uses the output of the first program to add the email addresses of the owners or creators to the PST files. This information is stored with the PST files in the repository so that users who want to view or restore PST documents can be authenticated.

To prepare PST files for archiving:

1. Run the **findPst** program to locate the PST files. The program adds the PST file names to a comma-separated value (CSV) file.
2. Open the CSV file with a suitable application such as a spreadsheet program.
3. Add a second column that contains the appropriate email address for each PST file name. If you can derive the addresses from the location, computer or corporate LDAP directory, you can use a script.
4. Save the CSV file. You can now use the CSV file as input for the **tagPst** program.

5. Run the **tagPst** program to add the email addresses of PST file owners or creators to the PST files so that Content Collector can archive these files.

Now, you can automatically transfer the message stubs of PST files to the owner's mailbox.

Related tasks

"Archiving email from local files" on page 432

"Collecting documents automatically" on page 460

Related reference

"findPst"

"tagPst" on page 437

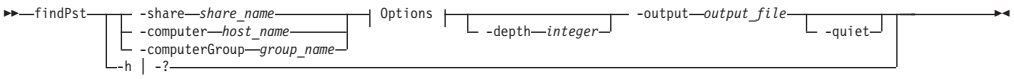
findPst:

The **findPst** program searches for PST files and writes their names to a comma-separated value (CSV) file.

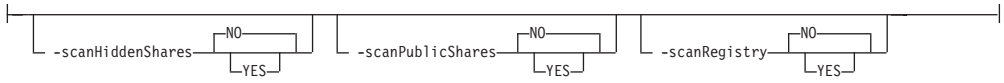
You run the **findPst** program from a Windows command prompt. The program file is located in *InstallPath*/ctms.

Syntax

Syntax



Options:



Parameters

-share *share_name*
Searches for PST files on the specified drive. You can specify any local drive or shared network drive. However, you must be able to access the drive from the computer that you use to run the program.

-computer *host_name*
Searches for PST files on the computer with the specified host name.

-computerGroup *group_name*
Searches for PST files on all computers in the group with the specified name. The computer group must be defined in the Active Directory that is used by your Exchange servers.

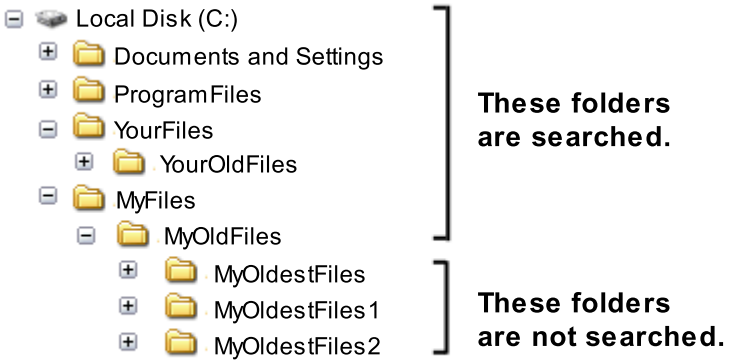
-scanHiddenShares YES | NO
Includes hidden network drives in the search, such as C$. Generally, the owners have not given other users permission to access these drives, but you can access them if you have administrator privileges. The default value is NO.

-scanPublicShares YES | NO
Includes shared network drives that are public (intended to be accessed by

multiple users). You must have *Change* permission for any public network drive that you want the findPst program to search. The default value is NO.

-scanRegistry YES | NO
Identifies the drives to search by scanning the Windows registry of the computer or the computers in a group. This option looks for MAPI profiles in the registry. It will find only those PST files that are part of Outlook client mailboxes. The default value is NO.

-depth *integer*
Determines the folder levels to search. If the default value 0 is used, only the root folder is searched. If you specify 2, the root folder is searched and all folders down to a maximum of two folder levels. For example with depth 2:



-output *output_file*
Specifies the name of the comma-separated output file that the program creates.

-quiet
Suppresses the output of processing information while the program is running.

-h | -?
Displays help information.

Example

The following command searches for PST files on the computer with the host name server1.company.com and writes the file names and the email addresses of their creators to a CSV file. The program will search three folder levels below the root folder. The output is written to a file named `PSTscan.csv`.

```
findPst -computer server1.company.com -scanPublicShares YES -scanHiddenShares YES
-depth 3 -output PSTScan.csv
```

Related tasks

"Preparing Notes Storage Facility files for archiving" on page 434

"Preparing personal storage files for archiving" on page 435

tagPst:

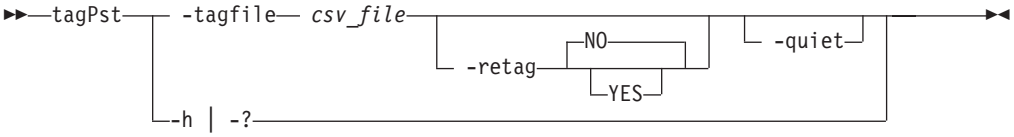
The **tagPst** program adds the email addresses of PST file owners or creators to the PST files so that these files can be archived by IBM Content Collector.

Note:

The program takes its input data from a CSV file. You must first create this file by using the **findPst** program. You run the tagPst program from a Windows command prompt. The program file is located in *InstallPath*/ctms.

Syntax

Syntax



Parameters

-tagfile *csv_file*
Specifies the CSV file that provides the input data.

-retag YES | NO
Permits or prohibits a reprocessing of PST files that are processed by the program.

-quiet
Suppresses the output of processing information while the program is running.

-h | -?
Displays help information.

Example

The following command adds the email addresses listed in the PSTscan.csv file to the corresponding PST files. The **-retag** parameter is used so that the program can be run again to update the email addresses associated with the PST files.

```
tagPst -tagfile PSTscan.csv -retag YES
```

Related tasks

"Preparing Notes Storage Facility files for archiving" on page 434

"Preparing personal storage files for archiving" on page 435

Managing document retention

You must often retain documents for a minimum period of time before you can delete them. In IBM Content Collector, you can automatically assign a retention date to each document and use the Expiration Manager tool for your repository to delete all documents that have passed their expiration date.

IBM Content Collector offers two methods of retention management:

Using the *Calculate Expiration Date* **task**
The *Calculate Expiration Date* task (formerly known as *Base Retention*) sets an expiration date for each document, based on the time period you supply and on the user name or the LDAP group membership of the recipient, or on a date expression that you can derive from a metadata property, expression, or literal value. The expiration date is set as a property when the document is archived.

A retention date is never updated: even if you move the document into a monitored folder with a longer retention period, the retention date maintains the original value that the *Calculate Expiration Date* task set.

To delete documents from the archive, you run the Expiration Manager. This tool checks for documents that have passed their retention date and deletes them from the archive.

See "Calculate Expiration Date" on page 498.

Using the *Declare Record* task

For more complex document retention management (FileNet P8 only), you can declare collected documents as records in IBM InfoSphere Enterprise Records.

See "P8 4.x Declare Record" on page 553.

Related information

IBM Records Manager information center

Running the Expiration Manager on Microsoft Windows:

Run the Expiration Manager tool to list and delete expired documents. You can run the tool on the IBM Content Collector server or remotely on a Microsoft Windows workstation.

The Expiration Manager checks the repository for documents that are older than their retention date. You can list, count, or delete these documents.

If you want to retain documents only for the retention time span and delete them afterward, run the Expiration Manager regularly to delete all documents with an expiration date prior to the current date.

To run the Expiration Manager on a Microsoft Windows machine:

1. If you run the Expiration Manager remotely on a different machine than IBM Content Collector, prepare your machine to run the Expiration Manager remotely.
2. Edit the properties file of the Expiration Manager according to your requirements.
3. For FileNet P8, customize the .bat file that starts the Expiration Manager.
4. Open a command promt and run the Expiration Manager to list the documents that will be deleted.
   - For Content Manager, run the following command:

     ```
     CM8ExpirationMgr.bat -propfile PropFileName -password Password -list
     ```

     where *PropFileName* is the complete filename of the property file and *Password* is the password for the Content Manager user. You need to specify the password only once. It is encrypted and stored in a file, so that you can omit the password the next time you run the Expiration Manager.
   - For FileNet P8, run the following command:

     ```
     P8ExpirationMgr.bat -propfile PropFileName -password Password -list
     ```

     where *PropFileName* is the complete filename of the property file and *Password* is the password for the P8 domain user. You need to specify the password only once. It is encrypted and stored in a file, so that you can omit the password the next time you run the Expiration Manager.

   You can substitute `-list` in the command with `-count` to output the number of documents that qualify for deletion.
5. Confirm that the documents that are listed in the report file should be deleted.

6. Delete the documents from the archive.
   - For Content Manager, run the following command:
     `CM8ExpirationMgr.bat -propfile PropFileName -delete`

     where *PropFileName* is the complete filename of the property file.
   - For FileNet P8, run the following command:
     `P8ExpirationMgr.bat -propfile PropFileName -delete`

     where *PropFileName* is the complete filename of the property file.

Running the Expiration Manager remotely:

If you want to run the Expiration Manager on a different Microsoft Windows machine than IBM Content Collector, you must prepare you machine to run the Expiration Manager remotely.

You can run the Expiration Manager from any machine that can access the repository system. For Content Manager, IBM Information Integrator for Content must be installed on the local machine. For FileNet P8, the IBM FileNet Content Engine server or client must be installed on the local machine. Make sure that the Java clients package for IBM FileNet Content Engine is installed. If you use the IBM FileNet Content Engine client, the client version must match the version of the IBM FileNet Content Engine server.

1. Copy the Expiration Manager tool to the local machine where you want to run the tool. The tool is located in *ICCdir*`/tools/ExpirationManager` on the IBM Content Collector machine, where *ICCdir* is the installation directory of IBM Content Collector, for example `C:\Program Files\IBM\ContentCollector`.
2. Copy the .jar files that are required to run the tool from the IBM Content Collector machine to your local machine. Copy the following files from the directory *ICCdir*`/lib`, where *ICCdir* is the installation directory of IBM Content Collector, into your local `ExpirationManager` directory:
   - hl14.jar
   - tlcore.jar
   - hlcbe101.jar
   - afu-logging.jar
   - org.eclipse.emf.common_2.3.0.v200710030400.jar
   - org.eclipse.emf.ecore_2.3.1.v200710030400.jar
3. On your local machine, set the environment variable *IBMAFUEXPIRATIONMGR* to point to the absolute path of the directory `ExpirationManager`. For example, this could be `C:\Program Files\IBM\ExpirationManager`.
4. On your local machine, set the environment variable *JAVA_HOME* to the installation directory of the Java Runtime Environment or the Java Development Kit (version 1.5 or later). For example, this could be `C:\IBM\java50`.
5. For FileNet P8, set the environment variable *IBMAFUFNCEROOT* on your local machine to point to the installation directory where the IBM FileNet Content Engine server or client is installed, for example `C:\Program Files\IBM\FileNet\CEClient` or `C:\Program Files\IBM\FileNet\ContentEngine`.

Customizing the Expiration Manager properties file:

Customize the Expiration Manager properties file and set the properties according to your requirements.

1. Navigate to *ICCdir*/tools/ExpirationManager, where *ICCdir* is the installation directory of IBM Content Collector, for example C:\Program Files\IBM\ContentCollector.
2. Open the properties file of the Expiration Manager. For Content Manager, the properties file is called afu-CM8ExpirationMgr-config.properties. For FileNet P8, the properties file is called afu-P8ExpirationMgr-config.properties.
3. Edit the properties file and set the properties according to your requirements. You can set the following options:

   For Content Manager and FileNet P8

   UseArchiveDate

   Specifies what to do when the expiration date in the document is not set.

   - If *UseArchiveDate* is set to no, the tool operates only on documents with a valid expiration date. If the expiration date is before the current date, the specified action is performed on the document.
   - If *UseArchiveDate* is set to ifExpirationDateIsNull, the tool operates on documents with a valid expiration date and on documents with an expiration date that is set but not valid. If the expiration date is not valid, the action is based on the archive date.
   - If *UseArchiveDate* is set to ifExpirationDateFieldDoesNotExist, the tool operates on documents with a valid expiration date and on documents on which the expiration date field does not exist. If the expiration date field does not exist, the action is based on the archive date.

   You can specify both ifExpirationDateIsNull and ifExpirationDateFieldDoesNotExist at the same time:

   ```
   UseArchiveDate  = ifExpirationDateIsNull|ifExpirationDateFieldDoesNotExist
   ```

   The default value is no.

   DeleteReport

   Specifies the verbositiy of the information that is logged in the report file.

   - If *DeleteReport* is set to normal, the following information is logged:
     - The time when the delete action starts.
     - The time when the delete action ends.
     - The number of documents or items that are eligible for deletion.
     - The number of documents or items that have been deleted.
   - If *DeleteReport* is set to verbose, additional information is logged, for example the run time, the user ID, and the document or item ID and status.

   The default value is normal.

ExpireDays
Specifies the number of days until expiration. This information is used if the expiration date is not set.

ReportFilePath
Specifies the location of report files. If this property is not set or if it contains an invalid path, the report files are written to the directory to which the TMP or TEMP environment variable points.

If the trace level INFO or a lower trace level is configured, the trace file itself mentions the exact directory where the report files can be found.

LinesPerFile
Specifies the number of lines per report file.

For Content Manager

Server The name of the library server.

UserID
The user name to access Content Manager.

ItemTypes
You can specify one or multiple item types and use the "*" and "." wildcard characters.

For FileNet P8

UserName
The user name to access the P8 domain.

CEServerHost
Specifies the hostname or IP of the CE server.

P8DomainName
Specifies the FileNet P8 domain name.

ObjectStoreName
Specifies the object store name.

ICCDEIClassName
Specifies the symbolic name of the DEI class.

ICCXITClassName
Specifies the symbolic name of the XIT class.

ICCNonEmailClassName
Specifies the symbolic name of the class for documents other than email, such as documents from Microsoft SharePoint or the file system.

TargetClassType
Specifies the type of documents to process.

- Set *TargetClassType* to `EMAIL` to process email documents. In this case, you must specify values for the properties *ICCDEIClassName* and *ICCXITClassName*.
- Set *TargetClassType* to `NON-EMAIL` to process documents other than email. In this case, you must specify a value for the property *ICCNonEmailClassName*.

DeleteBatchSize
Specifies the maximum number of documents that can be

deleted in one batch operation. If any of the documents in the batch cannot be deleted, none are deleted and an exception occurs. The default value is 50 documents. The maximum value is 100 documents. If the value that you set is larger than the default value, you must adjust the transaction timeout value of the application server on which the IBM FileNet Content Engine server is deployed.

Customizing the Expiration Manager start file for FileNet P8:

For FileNet P8, you must customize the Expiration Manager start file before you can run the Expiration Manager tool.

1. Navigate to *ICCdir*/tools/ExpirationManager, where *ICCdir* is the installation directory of IBM Content Collector, for example C:\Program Files\IBM\ContentCollector.
2. Open the file P8ExpirationMgr.bat in a text editor.
3. Locate the line **set AppServer_Type="<AppServer_Type>"** and replace <*AppServer_Type*> with the application server type.

   Possible values are:

   WL For WebLogic

   JBoss For JBoss

   WAS For WebSphere Application Server
4. Locate the line **set ConnProtocol_Type="<ConnProtocol_Type>"** and replace <*ConnProtocol_Type*> with the protocol type.

   Possible values are http and https.
5. Locate the line **set AppServer_Port="<AppServer_Port>"** and replace <*AppServer_Port*> with the application server port that is used for HTTP or HTTPS connections.

   The default values are:

Table 99. Default application server ports

	WebLogic	JBoss	WebSphere Application Server
HTTP	7001	8080	9080
HTTPS	7002	8443	9443

Running the Expiration Manager on UNIX:

Run the Expiration Manager tool to list and delete expired documents. You can run the tool remotely on a UNIX system.

The Expiration Manager checks the repository for documents that are older than their retention date. You can list, count, or delete these documents.

If you want to retain documents only for the retention time span and delete them afterward, run the Expiration Manager regularly to delete all documents with an expiration date prior to the current date.

To run the Expiration Manager on a UNIX system:

1. Copy the Expiration Manager tool to the local machine where you want to run the tool. The tool is located in *ICCdir*/tools/ExpirationManager on the IBM

Content Collector machine, where *ICCdir* is the installation directory of IBM Content Collector, for example `C:\Program Files\IBM\ContentCollector`.

2. Copy the .jar files that are required to run the tool from the IBM Content Collector machine to your local machine. Copy the following files from the directory *ICCdir*`/lib`, where *ICCdir* is the installation directory of IBM Content Collector, into your local `ExpirationManager` directory:
   - hl14.jar
   - tlcore.jar
   - hlcbe101.jar
   - afu-logging.jar
   - org.eclipse.emf.common_2.3.0.v200710030400.jar
   - org.eclipse.emf.ecore_2.3.1.v200710030400.jar
3. On your local machine, set the environment variable *IBMAFUEXPIRATIONMGR* to point to the absolute path of the directory `ExpirationManager`. For example, this could be `/opt/IBM/ExpirationManager`.
4. On your local machine, set the environment variable *JAVA_HOME* to the installation directory of the Java Runtime Environment or the Java Development Kit (version 1.5 or later). For example, this could be `/usr/java5`.
5. Edit the properties file of the Expiration Manager according to your requirements.
6. Open a shell and run the Expiration Manager to list the documents that will be deleted.
   - For Content Manager, run the following command:

     ```
     java -classpath ClassPath com.ibm.afu.purger.repository.cm8.ExpiredCM8Documents
     -propfile PropFileName -password Password -list [-h/help]
     ```

     where *ClassPath* is the full path to the .jar files that you copied in step 2. Enclose the path in double quotes if it contains space characters. *PropFileName* is the complete filename of the property file, and *Password* is the password for the Content Manager user. You need to specify the password only once. It is encrypted and stored in a file, so that you can omit the password the next time you run the Expiration Manager.
   - For FileNet P8, run the following command:

     ```
     java -Djava.ext.dirs=Djava_ext_dirs [-Dwasp.location=Dwasp_location]
     com.ibm.afu.purger.repository.p8.ExpireP8Documents -cp ConnProtocol_Type
     -ws AppServer_Type -tp CEWS -wp AppServer_Port
     -propfile PropFileName -password Password -list
     ```

     where:

     Djava_ext_dirs
     : is the path to the lib directory of the IBM FileNet Content Engine server or client, for example *CEInstallRootDir*`/lib` (where *CEInstallRootDir* is the installation directory of the IBM FileNet Content Engine server or client) and the path to the directory `ExpirationManager`. The path names are separated by semicolon. If you use IBM FileNet Content Engine version 4.5 or below, you must add the path *CEInstallRootDir*`/wsi/lib` as well.

     Dwasp_location
     : is the path to the directory *CEInstallRootDir*`/wsi` (where *CEInstallRootDir* is the installation directory of the IBM FileNet Content Engine server or client). Set this if you use IBM FileNet Content Engine version 4.5 or below.

ConnProtocol_Type
is the type of connection protocol.

Possible values are `http` and `https`.

AppServer_Type
is the type of the application server on which IBM FileNet Content Engine is deployed.

Possible values are:

WL For WebLogic

JBoss For JBoss

WAS For WebSphere Application Server

AppServer_Port
is the application server port that is used for HTTP or HTTPS connections.

The default values are:

Table 100. Default application server ports

	WebLogic	**JBoss**	**WebSphere Application Server**
HTTP	7001	8080	9080
HTTPS	7002	8443	9443

PropFileName
is the complete filename of the property file

Password
is the password for the P8 domain user. You need to specify the password only once. It is encrypted and stored in a file, so that you can omit the password the next time you run the Expiration Manager.

You can substitute `-list` in the command with `-count` to output the number of documents that qualify for deletion.

7. Confirm that the documents that are listed in the report file should be deleted.
8. Delete the documents from the archive. In the command from step 6 on page 444, substitute the option `-list` with `-delete`.

Deduplication

Deduplication ensures that only one copy of a document or an embedded attachment is kept in the archive, no matter how many times the same document or attachment was archived by different users.

Deduplication depends on the calculation of a unique deduplication hash key. Each source connector type calculates its hash keys differently.

For email connectors, the calculation takes document elements or metadata as input data. Attachments archived in Content Manager in item types created using the compound data model are also subject to deduplication. While this is handled by FileNet P8 internally, this support has been added when using compound item types in Content Manager.

The file system and Microsoft SharePoint connectors use a standard MD5 hash. If a task route includes a task that creates hash keys, it considers any two documents with the same hash key as identical and thus stores only one of them in the repository.

Restriction: Do not attempt to use deduplication with the Microsoft SharePoint collector if you either:

- Collect document versions, or
- Collect Microsoft Office documents (SharePoint changes the metadata that the collector uses to identify identical documents)

Doing so slows your system and results in no deduplication.

Example

Suppose an executive has sent an email to 30 employees. Because the content of the email is important, all 30 users archive the email using the IBM Content Collector archiving function of their email client.

Normally, the email would be saved in the archive 30 times, once for each user. Deduplication avoids this waste of archiving space. IBM Content Collector checks if the document belonging to an archiving request already exists in the archive and if it has been modified since it was archived. If it exists and has not been modified, it is not archived again. IBM Content Collector merely performs the post-archiving actions as configured for the clients. That is, it creates a document stub for each original document.

Related tasks

"Collecting file system documents" on page 479

Calculation of deduplication hash keys for Microsoft Exchange mail documents:

IBM Content Collector stores a deduplication hash key for each message. Based on the key, it is decided whether two or more messages are the same. That is, messages with the same key are treated as identical messages, and only one copy of these will be archived. A basic set of message properties goes into the calculation of the deduplication hash key for a message.

Only messages that belong to certain message classes are eligible for deduplication. These are the standard email message classes and the report message classes.

- IPM
- IPM.NOTE and children
- REPORT.IPM.Note.DR (delivery receipt)
- REPORT.IPM.Note.NDR (receipt of non-delivery)
- REPORT.IPM.Note.IPNRN (read notification)
- REPORT.IPM.Note.IPNNRN (not-read notification)
- IPM.Recall.Report.Success (report of a successful recall)
- IPM.Recall.Report.Failure (report of an unsuccessful recall)

Required message properties

The following message properties must contain values. Otherwise, a unique key based on the (unique) message ID is generated. This has the effect that no other message will have the same deduplication hash key. All messages are considered to be distinct.

- PR_CLIENT_SUBMIT_TIME
- PR_INTERNET_MESSAGE_ID
- PR_MESSAGE_CLASS
- PR_SENDER_EMAIL_ADDRESS
- PR_SUBJECT

Messages with blind-carbon-copy (BCC) recipients (`PR_RECIPIENT_TYPE = BCC`) need a unique identifier, or otherwise other users can read the list of BCC recipients.

Messages that contain tracking information also require a unique identifier. Otherwise, other users can read, for example, electronic votes.

Common message properties

The values of the following message properties, which are included in all messages belonging to the eligible classes, contribute to the calculation of the deduplication hash key:

Table 101. Common message properties whose values contribute to the calculation of the hash key

Message property	Type	Comment/Example
PR_CLIENT_SUBMIT_TIME	PT_SYSTIME	Timestamp in coordinated universal time (UTC) representing the *Sent* or *Posted* time, for example `1:51 PM, 5/2/2005`
PR_DISPLAY_BCC	PT_TSTRING	Display names of the Bcc recipients as shown in the **Bcc** field in a message
PR_DISPLAY_CC	PT_TSTRING	Display names of the Cc recipients as shown in the **Cc** field in a message
PR_DISPLAY_TO	PT_TSTRING	Display names of the To recipients as shown in the **To** field in a message
PR_MESSAGE_CLASS	PT_TSTRING	Exchange message class, for example `IPM.NOTE`
PR_RTF_COMPRESSED	PT_BINARY	Message-body text
PR_SENDER_EMAIL_ADDRESS	PT_TSTRING	Email address associated with the sender as shown in the **From** field of a message.
PR_SENDER_NAME	PT_TSTRING	Display name of the sender as shown in the From field in a message
PR_SENT_REPRESENTING_EMAIL_ADDRESS	PT_TSTRING	Email address classified `Sent_Representing`, as shown in the **From** field of the message.
PR_SENT_REPRESENTING_NAME	PT_TSTRING	Display name of the user representing the sender of a message
PR_SUBJECT	PT_TSTRING	Message subject

Note: Messages properties of the type PTSTRING are hashed as follows:
- If the property contains a Unicode string, a hash value is calculated for the Unicode string.
- If the property contains an ASCII string, a hash value is calculated for the ASCII string.
- A hash value is also calculated if the property is not set.

Message properties of report messages

If a deduplication hash key is calculated for report messages, the values of the following message properties are included in the calculation:

Table 102. Message properties of report messages whose values contribute to the calculation of the hash key

Message property	Type	Comment/Example
0x1046 (optional)	PT_TSTRING	Message identifier of the original message
PR_ORIGINAL_DISPLAY_TO	PT_TSTRING	Display name of the original recipient
PR_ORIGINAL_SUBJECT	PT_TSTRING	Subject of the original message
PR_ORIGINAL_SUBMIT_TIME	PT_TSTRING	*Sent* time of the original message
PR_REPORT_TIME (optional)	PT_TSTRING	Time when the reported event occurred

Recipient properties of standard email messages

If a deduplication hash key is calculated for standard email messages, the values of the following message properties are included in the calculation:

Table 103. Recipient properties of email messages whose values contribute to the calculation of the hash key

Message property	Type	Comment/Example
PR_RECIPIENT_TYPE	PT_LONG	Recipient type (*To*, *Cc*, or *Bcc*)
PR_SMTP_ADDRESS	PT_TSTRING	Recipients' SMTP addresses

Recipient properties of report messages

If a deduplication hash key is calculated for report messages, the values of the following message properties are included in the calculation:

Table 104. Recipient properties of report messages whose values contribute to the calculation of the hash key

Message property	Type	Comment/Example
PR_DISPLAY_NAME	PT_TSTRING	Display name of the original recipient
PR_RECIPIENT_TYPE	PT_LONG	Recipient type (*To*, *Cc*, or *Bcc*)
PR_REPORT_TIME	PT_TSTRING	Time when the reported event occurred
PR_REPORT_TEXT (optional)	PT_TSTRING	Additional report text
PR_SUPPLEMENTARY_INFO (optional)	PT_TSTRING	Additional report information

Attachment properties

There are numerous attachment message-properties because each attachment method uses a set of its own. Therefore, the attachment properties that influence

the calculation of the hash key are listed in a separate section. The calculation is also influenced by the attachment size, which consists of multiple values.

Which attachment properties are available depends on the attachment method that is used. The following methods are available:
- ATTACH_BY_VALUE
- ATTACH_BY_REFERENCE
- ATTACH_BY_REF_RESOLVE
- ATTACH_BY_REF_ONLY
- ATTACH_EMBEDDED_MSG
- ATTACH_OLE

The hash calculation includes the property PR_ATTACH_METHOD and a superset of attachment properties that are relevant for the individual attachment methods.

Attachment by value

If an object is attached to a message by value (method ATTACH_BY_VALUE), the following message properties is contribute to the hash key:

Table 105. Message properties used by the method ATTACH_BY_VALUE

Message property	Type	Comment/Example
PR_ATTACH_DATA_BIN	PT_BINARY	Binary data of the attachment. Since it might contain binary 0, it consists of a size and binary data, for example: `cb:3, lpb: 41 74 72`
PR_CREATION_TIME	PT_SYSTIME	Coordinated universal time (UTC) timestamp indicating when the attachment was created (this might be different from the time that the attachment was added to the message)
PR_LAST_MODIFICATION_TIME	PT_SYSTIME	UTC timestamp indicating when the attachment was last modified.
PR_DISPLAY_NAME	PT_TSTRING	Display name of the attachment. This is the name that is displayed next to the icon when you open the message or select the message and click **View Attachments**
PR_ATTACH_LONG_FILENAME	PT_TSTRING	File name of the attachment, as opposed to the display name. The message shows the display name. The file name is disclosed when the attachment is opened.

Note: Including PR_ATTACH_DATA_BIN in the hash calculation is very time-consuming. It is therefore excluded from the hash calculation for ATTACH_BY_VALUE. It is included only if one of the other properties is not included.

For other attachment methods, the binary data must be included in the hash calculation because there are not enough properties to uniquely identify the attachment.

Attachment by reference

An object can be attached to a message by referring to the object. There are different types of references, and therefore different attachment-by-reference methods:

- ATTACH_BY_REFERENCE
- ATTACH_BY_REF_RESOLVE
- ATTACH_BY_REF_ONLY

However, the same message properties contribute to the calculation of the hash key, no matter which of these methods is used.

Table 106. Message properties contributing to the hash calculation if an attachment-by-reference method is used

Message property	Type	Comment/Example
PR_ATTACH_PATHNAME	PT_TSTRING	Fully qualified path and file name. Each directory name is restricted to a length of eight characters. File names are restricted to eight characters and an extension of three characters.
PR_ATTACH_LONG_PATHNAME	PT_TSTRING	Fully qualified path and file name.

Embedded messages

Embedded messages (attachment method ATTACH_EMBEDDED_MSG) are another attachment type. If a message contains embedded messages, the following message properties influence the calculation of the hash key:

Table 107. Message properties that influence the calculation of the hash key if a message contains embedded messages

Message property	Type	Comment/Example
PR_ATTACH_DATA_OBJ	PT_OBJECT	Pointer to the embedded message. This pointer is resolved and the object referenced by the pointer is included in the hash calculation.
PR_DISPLAY_NAME	PT_TSTRING	Display name of the attachment. This is the name that is displayed next to the icon when the message is opened.

Attachment by object-linking and embedding (OLE)

If an attachment is linked with a message using OLE (method ATTACH_OLE), the following message properties contribute to the calculation of the hash key:

Table 108. Message properties that influence the calculation of the hash key if an attachment is linked through OLE

Message property	Type	Comment/Example
PR_ATTACH_DATA_BIN	PT_BINARY	Binary data of the attachment. Since it might contain a binary 0, it consists of a size and binary data, for example: `cb:3, lpb: 41 74 72`
PR_ATTACH_DATA_OBJ	PT_OBJECT	Pointer to the OLE attachment. This pointer is resolved and the OLE object referenced by the pointer is included in the hash calculation.
PR_DISPLAY_NAME	PT_TSTRING	Display name of the attachment. This is the name that is displayed next to the icon when the message is opened.

Calculation of deduplication hash keys for Lotus mail documents:

A deduplication hash key is stored for each user who tried to archive the same document. The key is used to identify the users with a right to retrieve the document and ensures that the document is kept in the archive as long as there are users who might want to retrieve a copy, that is, until all users have decided to delete the document.

Only mail documents are eligible for deduplication. A document is identified as a mail document if the following fields or items are present and have the values indicated:

- **$MessageID**
- **Form**
- **$TITLE**
- **$Title**

The field **$MessageID** just has to be present and contain a value. **Form**, **$TITLE**, and **$Title** must have one of the following values:

- *Memo*
- *Reply*
- *Reply With History*

Documents that are not considered email documents, or whose form has not been added to the set of forms to consider for deduplication will be archived individually. Their hash key will be computed based on the document's Universal ID and the Replica ID of the database it resides in.

To make other documents and not only mail documents eligible for deduplication, you can extend the existing set of document **Form** values so that if a document has any of the new form values, it will be considered a mail document eligible for deduplication.

Calculation of the hash key for mail documents

The deduplication hash key for an instance of the same mail document is calculated on the basis of the following values for existing fields:

- Value of **Form**, **$TITLE**, or **$Title**
- Value of **$MessageID**
- Value of **From**
- Value of **Principal**
- Value of **Subject**
- Value of **SendTo**
- Value of **CopyTo**
- Value of **PostedDate** if present (in endian-neutral representation)
- Any value if the field **DeliveredDate** is not present or does not have a flag, indicating the sender copy of an email document, if no DeliveredDate item is present, or has an empty value
- Value (content) of the document parts **TYPE_COMPOSITE**, **TYPE_MIME_PART**, or **TYPE_SEALDATA**
- For all document parts of the type **TYPE_OBJECT** (if present):
  - Attachment name
  - Size of attachment (in endian-neutral representation)
  - Creation date of attachment (in endian-neutral representation)
  - Last modification date of attachment (in endian-neutral representation)

Endian-neutral representation means that before a hash key is calculated, numerical values are converted to Big-Endian format to ensure that a platform-independent byte order is used in the hash key calculation

Important: If you add a form that has other items that contribute towards its distinguishing features than those listed above, you must also add these items to the set of items that are included in the hash computation. Otherwise a unique deduplication hash key cannot be guaranteed.

Calculation of deduplication hash keys for email received through SMTP:

IBM Content Collector stores a deduplication hash key for each email received through SMTP. Based on this key, it is decided whether two or more email are the same. That is, email with the same key are seen as being identical, and only one copy of the email is archived. A set of email properties forms the basis of the calculation of the deduplication hash keyl.

Email header and certain email content properties are considered for the calculation of the hash key.

The following header properties contribute to the computation:

- Value of **From**
- Value of **To**
- Value of **Sender**

- Value of **CC**
- Value of **BCC**
- Value of **Message-Id**
- The received date, usually the value of **Date**
- Value of **Subject**

In addition, the following email content contributes to the deduplication hash key:

- The message body
- For each attachment:
  - The attachment metadata
  - The attachment content

Using Classification Module to classify documents

The IBM InfoSphere Classification Module is an enterprise platform for a wide range of applications that require unstructured content to be automatically categorized. For more information see the IBM InfoSphere Classification Module information center.

IBM InfoSphere Classification Module automates the organization of unstructured content by analyzing the full text of documents and email and applying rules. When Content Collector is integrated with IBM InfoSphere Classification Module version 8.6 or higher, you can leverage the classification schema available in an existing Classification Module knowledge base for use with your repository and thus use a knowledge base to classify documents according to their content. With IBM InfoSphere Classification Module version 8.7 or higher, you can benefit from decision plans that use a combination of rules and classification schema to analyze content or to associate digital content against multiple knowledge bases. You can use this to extract metadata from the documents, for example.

Classification Module provides a classification or additional metadata, which can then be used in Content Collector, for example to determine if an item should be captured or not or how the item should be processed as it is captured and filed. With Classification Module, you can use advanced classification rather than relying solely on document metadata or source location metadata.

For example, to meet compliance and record management initiatives, or to prepare collections of data for legal discovery, Classification Module can distinguish important email from email that has no business value. Based on the content analysis, Content Collector determines the appropriate action to be taken. An email that discusses a *patent application* might be copied to an IBM FileNet P8 folder and be declared as a record in IBM FileNet Records Manager. To contrast, an email that discusses *patent leather* might be filtered out and not archived.

If you use a Classification Module decision plan to classify your documents, Classification Module can provide further data in addition to a categorization. You can access all metadata fields that are filled by Classification Module and use them in Content Collector.

For example, if Classification Module extracts a reference number from all documents in one category, this reference number can be used in Content Collector, for example to file a document into a specific folder for this reference number.

To leverage Classification Module in Content Collector, add an IBM Classification Module task to any task route. You can configure the task to use a Classification Module knowledge base or a Classification Module decision plan.

Related reference

"IBM Classification Module" on page 537

"IBM Classification Module system metadata properties" on page 262

Related information

IBM InfoSphere Classification Module information center

Setting up Classification Module for use with Content Collector:

To use IBM InfoSphere Classification Module classification in IBM Content Collector, you must integrate Classification Module into Content Collector. If you plan to classify Microsoft Exchange email, you must configure the Classification Module server to support this.

Integrating Classification Module into Content Collector:

Before you can add the IBM Classification Module task to a task route, you must configure the Content Collector server.

Make sure that the following software prerequisites are met:

- IBM Content Collector is installed.
- IBM InfoSphere Classification Module is installed.
- You deployed a knowledge base on IBM InfoSphere Classification Module.

For better scaling and performance, install Content Collector and Classification Module on separate servers.

1. If the IBM InfoSphere Classification Module client modules are not installed on the server where Content Collector is installed, install them.
   a. Run the Classification Module installation program on the Content Collector server.
   b. Select the option to install Custom components.
   c. Select the **Classification Module Client only** check box.
2. Configure Content Collector for using Classification Module. Content Collector requires access to Classification Module .dll files, which are located in the `/Bin` directory of the Classification Module client installation.

   Note: These files are not provided in the IBM Content Collector installer.

   a. Copy the required libraries from the `/Bin` directory of the Classification Module client (the default directory is `C:\IBM\classificationModule`) to the `/ctms` directory of Content Collector (the default directory is `C:\Program Files\IBM\ContentCollector`). The required libraries are listed below.

      For IBM InfoSphere Classification Module version 8.6:

      ```
      PackageDll23.dll
      stlport_ban46.dll
      bnsClient86.dll
      ```

      For IBM InfoSphere Classification Module version 8.7:

      ```
      PackageDll87.dll
      stlport_ban46.dll
      bnsClient87.dll
      ```

b. Register Classification Module as a utility connector task. To do so, open a DOS command window and enter the following commands, where *<ICCdir>* denotes the installation directory of Content Collector (the default is `C:\Program Files\IBM\ContentCollector`):
   - `cd <ICCdir>/ctms`
   - `utilityConnector.exe -u`
   - `utilityConnector.exe -r`

c. Confirm that IBM InfoSphere Classification Module exists as one of the utility connectors tasks in the Configuration Manager. Launch the Configuration Manager and make sure that IBM Classification Module shows up in the list of utility connectors tasks.

Configuring Classification Module for classifying Microsoft Exchange email:

If you want to use IBM InfoSphere Classification Module to classify Microsoft Exchange email, you must configure the Classification Module server to support this.

1. Install Microsoft Office Outlook 2007 or Microsoft Office Outlook 2003. The email archiving filter in Classification Module leverages the Messaging Application Program Interface (MAPI) for parsing email from a Microsoft Exchange server. To configure this support, complete the following steps:
   a. Install Microsoft Office Outlook 2007 or Microsoft Office Outlook 2003 on the server that hosts Classification Module.
   b. Select **Microsoft Office Outlook** as the default email application in your web browser.
2. For IBM InfoSphere Classification Module version 8.6: Configure the Classification Module server for email archiving. You do not need to complete these steps for IBM InfoSphere Classification Module version 8.7.
   a. Log on to the server hosting IBM InfoSphere Classification Module version 8.6.
   b. Stop the Classification Module services. Launch Windows services and stop the two services labeled "IBM InfoSphere Classification Module Process Manager" and "IBM InfoSphere Classification Module Trace".
   c. If email documents are being archived, overwrite the default document filter. Open a DOS command window and change to the `\Filters` directory of the Classification Module directory, for example `C:\IBM\ClassificationModule\Filters`. Enter the following commands:
      - `copy docFilterManager.xml docFilterManager.xml.orig`
      - `copy docFilterManager.email.xml docFilterManager.xml`
   d. Start the Classification Module services again. Launch Windows services and start the two services labeled "IBM InfoSphere Classification Module Process Manager" and "IBM InfoSphere Classification Module Trace".
3. Launch the Classification Module Management Console to load and start a knowledge base. See the IBM InfoSphere Classification Module information center for detailed information on uploading, launching, starting, and stopping an existing knowledge base.

Related information

IBM InfoSphere Classification Module information center

Using an Classification Module knowledge base:

With IBM InfoSphere Classification Module version 8.6 or higher, you can use a Classification Module knowledge base to access the classification information for the documents.

If you add the IBM Classification Module task to a task route and configure it to use a knowledge base, it produces metadata as described in the table below.

Metadata type	Description
All Relevant Categories	List of top categories matched.
All Relevant Categories and Scores	Combined list of categories; scores.
All Relevant Scores	List of top scores.
Most Relevant Category	Winning category.
Most Relevant Score	Winning category score.

You can use these metadata properties later in the task route to determine how documents will be processed. For example, you can use decision points and rules that depend on Classification Module properties, or you can use the properties in a property mapping.

The following scenarios give an idea of how you can use the metadata that is determined by IBM InfoSphere Classification Module to process and classify your documents:

Archive email or documents into predefined folders determined by the winning category

For example, you can assign the metadata "Most Relevant Category" as the folder path destination for the archived email or documents.

Identify mission-critical email or documents and declare them as records, depending on the relevancy score

For example, you can add decision points to the task route and use the metadata "Most Relevant Score" in the rule that determines the following action.

Populate metadata of the archived instance with category names that might be used for parametric search or classification of search results

For example, you can enrich property fields with the metadata "Most Relevant Category". In this way, you can for example add the proposed category name to the subject of an email message.

Related reference

“IBM Classification Module” on page 537

“IBM Classification Module system metadata properties” on page 262

Using an Classification Module decision plan:

With IBM InfoSphere Classification Module version 8.7 or higher, you can use a Classification Module decision plan in Content Collector instead of a knowledge base. Decision plans allow for more elaborate classification and analysis of documents. If you use a decision plan, you can access the metadata fields that Classification Module provides for the documents, on top of the system metadata that Classification Module provides.

If you add an IBM Classification Module task that makes use of a decision plan to the task route, the system metadata is populated in the same way as for a

knowledge base. In addition, the metadata property "Decision plan results exported as XML" can be filled with the decision plan results in XML format. This can be used as a history of classification activities when you want to review classifications later with the Review Tool that is shipped with IBM InfoSphere Classification Module. The following metadata is available:

Metadata type	Description
All Relevant Categories	List of top categories matched.
All Relevant Categories and Scores	Combined list of categories; scores.
All Relevant Scores	List of top scores.
Most Relevant Category	Winning category.
Most Relevant Score	Winning category score.
Decision plan results exported as XML	Decision plan results in XML format. Used only for decision plans (IBM InfoSphere Classification Module version 8.7).

In addition to the system metadata, you can access the Classification Module document metadata properties. For each Classification Module property that you want to use, you must define a metadata property in IBM Content Collector and map the Classification Module property to the Content Collector user-defined metadata.

Related reference

“IBM Classification Module” on page 537

“IBM Classification Module system metadata properties” on page 262

Mapping Classification Module properties to Content Collector metadata properties:

To access the Classification Module document metadata properties for a document in Content Collector, you must map them to Content Collector user-defined metadata properties.

To add a user-defined metadata property in Content Collector and map a Classification Module property to this metadata property for an IBM Classification Module task in a task route, complete the following steps:

1. In the Configuration Manager, click **Metadata and Lists** to switch to the Metadata andLists view.
2. In the **Metadata and Lists** pane on the upper left, select **User-Defined Metadata**.
3. Add the metadata properties that you plan to use. These properties are not related to the properties in Classification Module, so you can choose any name. However, it is convenient to use the same name as in Classification Module.
4. In the **Navigation** section of the Configuration Manager, click **Task Routes** and select the task route that contains your IBM Classification Module task.
5. Select the IBM Classification Module task. Make sure that the task uses a decision plan.
6. Go to the **Map Decision Plan Results** tab.
7. Select the metadata set that you want to use. The mapping table is populated with the metadata properties.
8. Click the metadata property that you want to map and select a Classification Module decision plan property.

Related tasks

"Adding and editing user-defined metadata" on page 240

Collecting documents for processing

To be able to process documents with IBM Content Collector, you must include a collector in your task route. A collector connects to your source system and collects the documents to be processed in a task route.

When you configure a collector, you define the source from which the collector collects the documents and you set the schedule on which the collector collects the documents. Depending on the type of the documents, you can specify additional filter criteria to determine which documents a collector collects. You can add other conditions by including decision points and rules in your task routes.

Define a collector for the source system from which you want to collect documents. The source system can be a mail system, a file system, or a SharePoint site. If the source system is a mail system, you can also define different collectors for automatic or interactive archiving, or a collector for life cycle processing.

Collector schedules

To set up a collector schedule, specify when and how often a collector checks the monitored collection source for documents or files to collect.

In the section **This collector runs**, select a time interval for the collection. The following table describes the available options:

Table 109. Time interval configuration options

Time interval	Description
Always	The collector scans the collection sources continuously. Microsoft SharePoint only: The collection occurs indefinitely at one-minute intervals. **Recommendation:** This selection can result in high load on your source server. Furthermore, contention can occur between different collectors that try to access the server. For example, if you have interactive, automatic, and life-cycle task routes deployed and all collectors are set to run always, some collectors might run less frequently than you expect. Therefore, do not select **Always** in production environments.
Daily	The collector runs once a day depending on the values that you set for **Repeat collection every**, at the time that you specify.
At intervals	The collector runs at specified intervals, regardless of date boundaries.
Monthly	The collector runs on specified days of a month, that is, on the first, the second, the third, and so on, starting at the time that you specify.
Once	The collector runs only on the day and time that you specify.

Table 109. Time interval configuration options (continued)

Time interval	Description
Weekly	The collector runs once a week.

Time frame

To postpone the start of the monitoring period, change the value in the **Start date** field. The first collection does not necessarily occur on the start date. For example, if the start date is a Monday, but you scheduled collections for Tuesdays during a weekly interval, then the first collection occurs on the day after the start date.

Tip: For all intervals except **Always**, setting a start date in the past executes the collection immediately when you start IBM Content Collector services.

Specify when to end the monitoring period by selecting between the following options:

Run endlessly
To never end the monitoring period.

Until To end the monitoring period on the specified date. The collector no longer runs after this date. Note that even though no more documents are collected, all work generated up to this date is still processed.

End after running
To make the end of the monitoring period dependent on the number of collections. For example, to end the monitoring period after five collections, change the value in the **times** field to **5**.

Run time

Specify when to start and stop each collection run. Under **Start first collection at**, select a time of day. Under **Stop collection**, select an option:

When task completes
Stops the collection when all mailboxes, databases, or files in the collection source are processed.

When task completes or at
Stops the collection at the specified time regardless of whether all mailboxes, databases, or files could be processed. If you select this option, an unfinished collection resumes at the start of the next run.

Repeat collection every

Set the frequency of collections with regard to the selected interval. For example, if the selected interval is **Weekly**, and you want to collect on Mondays and Thursdays, but only every second week, specify the following values:

- **Repeat collection every**: 2
- **Monday**
- **Thursday**

Related tasks

"Collecting file system documents" on page 479

"Collecting from Microsoft SharePoint sites" on page 482

"Collecting email on request" on page 472

"Collecting SMTP documents" on page 478

"Collecting documents automatically"

Collecting from mail-system document stores

To collect email and other documents from Lotus Notes or Microsoft Exchange sources for processing, you must include an email collector in your task route and configure it accordingly. IBM Content Collector can collect these documents automatically on a schedule or interactively, that is, at a user's request. For life cycle processing of these documents, you must include and configure a stubbing collector. To collect SMTP/MIME email, include and configure an SMTP collector.

Collecting documents automatically:

To have Content Collector automatically collect email and other documents, you must configure an **EC Collect Email by Rules** collector. This type of collector runs at preset intervals and collects documents from predefined sources. Email and other documents are collected based on the age, size, mailbox size, presence of attachments, and other criteria that you specify.

Prerequisites: Before IBM Content Collector can collect documents, you must configure a source connector. Otherwise, Content Collector cannot access the source system.

To configure a collector for automatic collection of documents:

1. Open the Configuration Manager and click **Task Routes**.
2. Create or select the task route to which you want to add the collector.
3. In the **Toolbox**, click **Email** → **EC Collect Email by Rules** and add it to the task route diagram.
4. On the General page in the right pane, define general settings.
   a. Specify a name and a description for the collector.
   b. If you do not want to use the collector right away, deselect **Active**.
5. On the Schedule page, set a collector schedule by specifying when and how often the collector checks the monitored libraries for documents to collect.
6. On the Collection Sources page, configure one or more collection sources.

Option	Description
Lotus Domino	Add one or more collection sources: • To collect email from client mailboxes • To collect email from a journaling database • To collect documents from Notes applications, such as TeamRooms, project libraries, address books, and calendars You can specify this collection source in addition to mailbox collection sources. • To collect all Notes Storage Facility files (NSF files) in specific folders on the local computer or a share

Option	Description
Microsoft Exchange	Add one or more collection sources: • To collect email from client mailboxes • To collect email from a journaling mailbox • To collect personal storage files (PST files) PST files, also known as personal folders, are local files that were created by Outlook users. These files are not under the control of the Exchange email servers and are therefore not associated with the email addresses of their owners or creators. Before PST files can be archived, you must prepare the PST files so that the owners or creators of PST files can view or restore archived PST documents. • To collect documents from public folders You can combine mailbox and PST file collection sources depending on how you want to process the PST files, but a collector for public folders cannot collect documents from any other sources.

7. Specify filter criteria to collect a subset of documents from the selected collection sources. The default filter settings depend on the source system.

Option	Description
Lotus Domino	Exclude the monitored folder Draft including its subfolders. You cannot change this setting. Exclude these message types: • Calender entry • Calender event • ToDo

Option	Description
Microsoft Exchange	Collect from all folders, regardless of whether they are managed by Exchange. Exclude these monitored folders: • Outbox • Drafts • Sync issues • Trash • Junk • RSS feeds • Calendar • Contact • Journal • Note • Task You cannot change the settings for the folders Outbox, Drafts, and Sync issues. Exclude these message types: • Conflict • Meeting request • Meeting response • Recall message • Task request • Task response You can change the settings only for the message types Meeting request and Meeting response.

8. Save your settings.

Related tasks

"Preparing personal storage files for archiving" on page 435

"Preparing Notes Storage Facility files for archiving" on page 434

Related reference

"Collector schedules" on page 458

"Lotus Notes collection sources for automatic archiving"

"Microsoft Exchange collection sources for automatic archiving" on page 464

"Collection filter for email collectors" on page 468

Lotus Notes collection sources for automatic archiving:

You can specify the several collection sources for automatically collecting documents from Lotus Notes sources. Such a collection source can consist of client mailboxes, journaling databases, Notes applications, or NSF files.

You can set a size limit for mailboxes in your collection sources, so that documents are collected only if the size of the mailbox exceeds the specified limit. Select **Store size greater than** and specify the size in MB or GB.

When local NSF files are collected, you can have Content Collector create status information for each file that was processed. The information is written to a metadata file named <source_filename>.icc.xml in the directory from which the NSF file was collected. The metadata file can be used to further process the local NSF files with the File System Collector.

Table 110. Lotus Notes collection sources for automatic archiving

Collection source	Description	Further options
Domino database	Collects documents from Notes applications, such as TeamRooms, project libraries, address books, and calendars. You can specify this collection source in addition to mailbox collection sources.	Enter the name of the Domino server in the **Domino server** field. Use the Lotus Notes abbreviated format, for example, `ATE75TS/D/ATE`. Then enter the path to the database on that server in the **Database path** field. Use the same format as in this example: `e_dir/131456.nsf`.
All Domino databases in a directory	Collects documents from all Notes applications that can be found in the specified folder on a Domino server. You can specify this collection source in addition to mailbox collection sources.	Enter the name of the Domino server in the **Domino server** field. Use the Lotus Notes format, for example `ATE75TS/D/ATE`. Also enter the name of the folder in the **Folder name** field. The folder path must be relative to the Domino Data directory. If you leave this field empty, the Domino Data directory is used. You can also include items in subfolders.
All mailboxes in a group	Collects email from the mailboxes of users belonging to a certain user group.	In the **Group name** field, enter the name of a user group as shown in the Lotus Notes address book.
All mailboxes on a server	Collects email from all the mailboxes that can be found on the specified email server. **Restriction:** Journaling databases will be excluded from the collection.	Enter the name of the email server in the **Domino server** field. Use the Lotus Notes abbreviated format, for example, `ATE75TS/D/ATE`.
Mailbox	Collects email from the specified mailbox only.	Enter the address of the mailbox owner in the **Mailbox address** field. Use either the SMTP format for example, `iccuser@mycompany.com`, or the Lotus Notes abbreviated format, for example, `iccuser/Germany/mycompany`.

Table 110. Lotus Notes collection sources for automatic archiving (continued)

Collection source	Description	Further options
Journal	Collects email from a journaling mailbox or database. Journaling creates a copy of all email and the corresponding email header data to and from users on a mailbox and sends the message copy to a specified mailbox. If you have multiple journaling mailboxes and a round-robin mechanism to fill them, journals are written to a new mailbox or database each time the previously used mailbox has reached its size limit. When all journaling mailboxes or databases are filled, the one with the oldest entries is enabled and the old entries are overwritten. In this case, select **Exclude the active journal** to exclude the journaling mailbox that currently records incoming and outgoing mail traffic, meaning that documents are archived only from the inactive journaling mailboxes or databases.	Enter the name of the email server on which the journaling mailbox is located in the **Domino server** field. Use the Lotus Notes abbreviated format, for example, `ATE75TS/D/ATE`.
All NSF files in a folder tree	Collects all NSF files in the specified folder or in subfolders on the local computer or a network share. A network share is a location on a computer network that allows multiple computer users on the same network to have a centralized space on which to store files.	Also enter the name of the folder in the **Folder name** field. For files on the local computer, enter the fully qualified path. For files on a network share, enter the path in Universal Naming Convention (UNC) syntax: `\\servername\sharename\path\filename` You can also include items in subfolders.

Related tasks

"Collecting documents for life cycle processing" on page 476

"Collecting documents automatically" on page 460

Microsoft Exchange collection sources for automatic archiving:

Select the collection sources for automatically collecting documents from Microsoft Exchange sources. A collection source can consist of client mailboxes, journaling mailboxes, public stores, or personal storage files (PST files).

You can set a size limit for mailboxes in your collection sources, so that documents are collected only if the size of the mailbox exceeds the specified limit. Select **Store size greater than** and specify the size in MB or GB.

When PST files are collected, you can have Content Collector create status information for each file that was processed. The information is written to a metadata file named <source_filename>.icc.xml in the source directory. The metadata file can be used to further process the PST files with the File System Collector.

Table 111. Microsoft Exchange collection sources for automatic archiving

Collection source	Description	Further options
PST file on local computer	Collects the messages from a single PST file.	Enter the name of the PST file in the **PST file** field.
All PST files in a folder tree	Collects messages from all PST files that can be found in the specified folder or in subfolders.	Enter the name of the folder in the **Folder name** field. Additionally, enter the number of folder levels to search under the specified root folder. For example, if you specified `Private_Mails` in the **Folder name** field, and you want to search for PST files in `Private_Mails\Drafts`, enter `2` in the **Folder tree levels to search** field. The default level is `0`, which means that no folders are searched.
All PST files on a computer	Collects messages from all PST files that can be found on the local hard drives of a computer.	In the **Mail server or computer name** field enter the name of the email server as shown in the Active Directory Users and Computers window, under **<Domain Name>** → **Computers**. Select any options from "PST Search Options" on page 467. Additionally, enter the number of folder levels to search under the specified root folder. For example, if you specified `Private_Mails` in the **Folder name** field, and you want to search for PST files in `Private_Mails\Drafts`, enter `2` in the **Folder tree levels to search** field. The default level is `0`, which means that no folders are searched.

Table 111. Microsoft Exchange collection sources for automatic archiving (continued)

Collection source	Description	Further options
All PST files in a computer group	Collects messages from all PST files that can be found on the local hard drives of computers belonging to a computer group.	In the **Group name** field, enter the display name of a group that exists in the Active Directory. Enter the name as shown in the Active Directory Users and Computers window, under **<Domain Name>** → **Groups**. Additionally, enter the number of folder levels to search under the specified root folder. For example, if you specified `Private_Mails` in the **Folder name** field, and you want to search for PST files in `Private_Mails\Drafts`, enter `2` in the **Folder tree levels to search** field. The default level is `0`, which means that no folders are searched.
All mailboxes in a group	Collects email from the mailboxes of users belonging to a certain user group.	In the **Group name** field, enter the display name of a group that exists in the Active Directory or the display name of a dynamic group. Enter the name as shown in the Active Directory Users and Computers window, under **<Domain Name>** → **Groups**.
All mailboxes on a server	Collects email from all the mailboxes that can be found on the specified email server.	Enter the name of the email server in the **Mail server or computer name** field, for example, `server1.company.com`. If IBM Content Collector runs in the same domain, you can enter `server1`.
Mailbox	Collects email from the specified mailbox only.	Enter the Internet address of the mailbox owner in the **Mailbox SMTP address** field, for example, `iccuser@mycompany.com`.

Table 111. Microsoft Exchange collection sources for automatic archiving (continued)

Collection source	Description	Further options
Journal	Collects email from a journaling mailbox. Journaling creates a copy of all email and the corresponding email header data to and from users on a mailbox and sends the message copy to a specified mailbox.	Enter the internet address of the journal recipient mailbox in the **Mailbox SMTP address** field, for example `journal@company.com`. **Important:** For Microsoft Exchange envelope journaling mailboxes, IBM Content Collector supports the Exchange MAPI Message Format (TNEF) for the copied message that is attached to the envelope message. Outlook Message Format (.msg) is not supported, and archiving of envelope messages with attachments in Outlook Message Format fails.
Public store	Collects email from public folders. This selection is available only if no other collection sources are defined for this collection. When you select **Public store**, the entire public-folder tree is included in the collection. To limit the set of public folders that the collector processes, you can include or exclude folders from collection by defining a collection filter. Messages in public folders can be marked for stubbing by any user that has write access to the message. To restore or view a message that was archived from a public folder, read access to the message is required. Messages in public folders are archived only once. This means that if you change an archived message, the change is not reflected in the repository, and any change will be lost when the archived message is restored or restubbed.	

PST Search Options

The following PST search options are available:

Search hidden shares
In addition to searching for PST files on the local drives that are open to most users, also looks for PST files on local drives or shared network drives that are hidden. For example, these drives include shares such as C$ or D$ that can be viewed only by administrators.

Search public shares
In addition to searching for PST on the local drives of computers, also looks for PST files on shared network drives unless those drives are hidden.

Search registry
When a user creates a PST file, this is registered in the Outlook logon profile, and keys that identify the PST file are added to the Windows registry. IBM Content Collector can search the registry and find those keys, and locate PST files on a computer.

Disable PST file creation
After existing PST files have been archived, no new PST files can be created on the scanned computer.

Related tasks

"Collecting documents for life cycle processing" on page 476

"Collecting documents automatically" on page 460

Collection filter for email collectors:

By using filtering, you can define which documents a collector includes or excludes when it searches the collection source for documents to collect. You can use a combination of different filtering criteria.

For automatic archiving, the criteria can be the age or the size of a document, or whether a document is in a certain folder. You can also specify a custom search expression to select the documents that you want.

Message constraints

You can combine these options:

Table 112. Message constraints

Option	Description
Filter email by age	Filter on the basis of the document age: • Select the reference date and time for the age calculation: – The date and time when the document was created, such as the time when the document was drafted and saved, or sent by a user. – The date and time when the document was last modified, such as the time when the document was last opened, edited, or moved to a different folder in the mailbox. – The date and time when the document was received by a user. • Specify an interval or a fixed date: **Older than** An interval that uses the reference date and time as the starting point and the current date and time as the end point. For example, selecting **modified date** and **Older than 3 months** will collect documents whose last modification is at least three months in the past. **On or before** A fixed date to be associated with the reference date **created date** or **modified date** date. For example, selecting **modified** and **On or before Saturday, August 9, 2008** will collect documents whose last modification occurred on or before that day.
Filter email by size	Filter on the basis of the document size. A document will be collected only if its size is above the specified limit.
Ignore encrypted items	Exclude encrypted documents from the collection.
Ignore items previously processed	Exclude documents from the collection that were already processed by a collector. You must also specify whether this option applies only to documents that were processed by the current collector or whether it applies to all documents that were processed by any collector.

Table 112. Message constraints (continued)

Option	Description
Lotus Notes only: **Custom search expression**	Select documents by a custom search expression. Type the expression in the field under the check box. The query must be defined as a Notes formula.
Microsoft Exchange only: **Managed Folder Constraints**	Select the folders for collection: **Collect from folders not managed by Exchange** Content Collector collects documents from folders other than managed folders. You can explicitly exclude folders from collection. Unless you select **Collect from folders managed by Exchange**, you can also specify a list of folders to be included instead. **Collect from folders managed by Exchange** All managed folders are added to the list of folders that are included in the collection. If the **include subfolders** option is set in Exchange, subfolders of any managed folder are also included. You can explicitly exclude any of these folders in the list of monitored folders.

Monitored folders

You can exclude or include documents in specific mailbox, public, or PST-file folders in the collection. To do so, select **Exclude** or **Include** and enter the folder to exclude or include in the **Folder/View name** field in the Add Exclude Condition or Add Include Condition window.

If you select **Exclude**, the following folders and their subfolders are excluded by default:
- Outbox for Outlook
- Drafts for both Outlook and Lotus Notes
- Sync issues for Outlook

If you select **Include** and you do not specify any folders in the list, the following folders and their subfolders are included:
- all folders for Outlook
- all folders except for the Drafts folder for Lotus Notes

Important: You should never archive documents from the Drafts folder. Archived draft documents are not archived again after they have been sent, and as a result, they cannot be retrieved. The only context to archive from the Drafts folder is PST Migration, where it is required to archive all content. In Lotus Notes, draft documents (and tasks without a due date, which are also considered drafts) are always excluded from archiving. In Outlook, if you select **Include**, make sure that you do not explicitly add the Drafts folder and do not leave the folder list empty to include all folders.

When you add folders to the list, enter the following placeholders so that the correct folder is identified for both supported email clients, no matter which language the client is set to:

- /%OUTBOX% for the *Outbox* folder in the English version of Outlook
- /%SENT% for the *Sent Items* folder in the English version of Outlook or the *Sent* folder in the English version of Lotus Notes
- /%DRAFTS% for the *Drafts* folder in the English version of Outlook or Lotus Notes
- /%TRASH% for the *Deleted Items* folder in the English version of Outlook or the *Trash* folder in the English version of Lotus Notes
- /%JUNK% for the *Junk E-mail* folder in the English version of Outlook or the *Junk Mail* folder in the English version of Lotus Notes
- /%RSS_FEEDS% for the *RSS feeds* folder in the English version of Outlook
- /%SYNC_ISSUES% for the *Sync issues* folder in the English version of Outlook
- /%INBOX% for the *Inbox* folder in the English version of Outlook or Lotus Notes.

Make sure to use a forward slash (/) with the placeholders for both Outlook and Lotus Notes.

You can still specify folders in the format that is specific to the email client, for example, /Sent Items for Outlook or ($Sent) for Lotus Notes. However, this notation works only for one type of email client. You must use the native notation if a generic placeholder does not exist for the folder that you want to specify.

To exclude or include the subfolders of the specified folder, select **Also include items in subfolders**. For example, you might not want to collect documents in the *Junk R-mail* and *Deleted Items* folders of Outlook mailboxes for archiving. In this case, you must add two excluding conditions. In one of these, you specify /%JUNK%, and in the other, you specify /%TRASH%.

Excluded Message Types

To exclude documents from the collection that use certain standard Lotus Notes forms or Exchange message classes, add the message type to the list of excluded message types. The message types that you can select in the Add Message Type window are related to a default list of commonly used forms and message classes. Selecting one of these will exclude the Lotus Notes form or Exchange message class. Alternatively, you can enter the exact name of a Lotus Notes form or Exchange message class in the display field of the list, for example:

Lotus Notes
```
Person
```

Microsoft Exchange
```
IPM.Report
```

If you enter a form or message-class name, the radio buttons under the **Message type** list are enabled. To exclude just the specified form or class, but not its children, select **Exclude base message type**. To exclude just the children, select **Exclude all message types derived from the specified base type**.

For example, to exclude the parent message class for delivery reports without its children (IPM.REPORT), enter IPM.Report in the **Message type** field and select **Exclude base message type**. To exclude the children (IPM.REPORT.*), enter IPM.Report and select **Exclude all message types derived from the specified base type**.

You can exclude more than one message type.

Related tasks

"Collecting documents automatically" on page 460

Collecting email on request:

If you want to allow users to decide when IBM Content Collector is to collect email for further processing, you must configure a collector for interactive archiving. With interactive archiving, Lotus Notes and Microsoft Outlook client users can flag documents for archiving or stubbing. Documents flagged by email client users are selected for processing the next time the collector runs.

Prerequisites: Before IBM Content Collector can collect documents, you must configure a source connector. Otherwise, Content Collector cannot access the source system.

When users flag a document for processing by Content Collector, an email is sent to a specific mailbox, the *trigger mailbox*. This email instructs an interactive collector (**EC Collect Email by User Selection** collector) to process documents in the client mailbox that the request came from. The collector searches the trigger mailbox for archiving or stubbing requests that were submitted by client users or picks up email directly from certain folders in the users' mailboxes, according to the schedule that you define. As the administrator, you define the trigger mailbox to which processing requests are sent or the folders into which users can put the email to be archived. Content Collector creates these folders in the client mailboxes.

With this type of collector, you can also let users specify additional archiving information for a document when they manually submit it for archiving. This additional archiving information can later be used when the document is processed in a task route.

To configure an email collector for interactive archiving:

1. Open the Configuration Manager and click **Task Routes**.
2. Create or select the task route to which you want to add the collector.
3. In the **Toolbox**, click **Email** → **EC Collect Email by User Selection** and add it to the task route diagram.
4. On the General page in the right pane, define general settings.
   a. Specify a name and a description for the collector.
   b. If you do not want to use the collector right away, deselect **Active**.
   c. Define the location that the collector monitors for archiving or stubbing requests. Select one of the following options:

      Collect from a trigger mailbox
      Monitor a trigger mailbox for archiving or stubbing requests that were initiated by client users. You define the trigger mailbox when you define the collection sources. Using the information in the request documents, Content Collector can identify the mailboxes to collect the email from.

      Collect from folders
      Monitor specific folders periodically and collect the email in those folders. You define these folders when you define the collection

sources. Content Collector creates the folders in the mailboxes of your client users. Users drag email that they want to archive into these folders.

If you enabled Content Collector for gathering and using additional archiving information, you can also specify whether such information is associated with the email in the monitored folders. In this case, only those documents are collected for which additional archiving information was specified.

To exclude documents from the collection that were already processed by a collector, select **Ignore items previously processed**. Specify also whether this option applies only to documents that were processed by the current collector or whether it applies to all documents that were processed by any collector.

5. On the Schedule page, set a collector schedule by specifying when and how often the collector checks the trigger mailbox for archiving requests or the specified folders for documents to collect.
6. On the Collection Sources page, configure one or more collection sources. Depending on your selection on the General page, you must define a trigger mailbox or folders as collection source:
   - Specify the mail server on which the trigger mailbox resides. Note that you must also specify the trigger mailbox in the client configuration, so that clients can send archiving and stubbing requests to the trigger mailbox. The collector processes the requests that it finds in the trigger mailbox.

Option	Description
Lotus Domino	When you specify the Domino server, use the Lotus Notes format `ATE75TS/D/ATE`. Also specify the path to the Notes database that serves as the trigger mailbox in the **Database path** field. For example, enter `z_dir/iccjobs.nsf`.
Microsoft Exchange	When you specify the Exchange server, use the format `server1.company.com`. If IBM Content Collector runs in the same domain, you can enter `server1`. Also specify the Internet address of the trigger mailbox in the **Mailbox SMTP address** field. For example, enter `iccjobs@company.com`.

   - Specify the mailboxes in which to create drag-and-drop folders for interactive archiving. The collector searches these folders for selectable documents.

Option	Description
Lotus Domino	In the Add Collection Sources window, select one of the following options: **Domino database** Collects email from the specified Domino database. Enter the name of the Domino server in the **Domino server** field. Use the Lotus Notes abbreviated format, for example, `ATE75TS/D/ATE`. Then enter the path to the database on that server in the **Database path** field. Use the same format as in this example: `e_dir/131456.nsf`. **All mailboxes in a group** Collects email from the mailboxes of users belonging to a certain user group. In the **Group name** field, enter the name of a user group as shown in the Lotus Notes address book. **All mailboxes on a server** Collects email from all the mailboxes that can be found on the specified email server. **Note:** Journaling databases will be excluded from the collection. Enter the name of the email server in the **Domino server** field. Use the Lotus Notes abbreviated format, for example, `ATE75TS/D/ATE`. **Mailbox** Collects email from the specified mailbox only. Enter the address of the mailbox owner in the **Mailbox address** field. Use either the SMTP format for example, `iccuser@mycompany.com`, or the Lotus Notes abbreviated format, for example, `iccuser/Germany/mycompany`.

Option	Description
Microsoft Exchange	In the Add Collection Sources window, select one of the following options: **All mailboxes in a group** Collects email from the mailboxes of users belonging to a certain user group. In the **Group name** field, enter the display name of a group that exists in the Active Directory or the display name of a dynamic group. **All mailboxes on a server** Collects email from all the mailboxes that can be found on the specified email server. Enter the name of the email server in the **Mail server or computer name** field, for example, `server1.company.com`. If IBM Content Collector runs in the same domain, you can enter `server1`. **Mailbox** Collects email from the specified mailbox only. Enter the Internet address of the mailbox owner in the **Mailbox SMTP address** field, for example, `iccuser@mycompany.com`.

Define the drag-and-drop folders that are to be used for interactive archiving in the **Monitored Folders** section. In the **Folder name** field of the Add Monitored Folders window, enter a folder name. This folder will be created in all user mailboxes.

Important: Do not include the root folder (/) in the list of monitored folders. If you want to include the root folder and thus all folders, provide an empty include list. However, you should never archive documents from the Drafts folder. Archived draft documents are not archived again after they have been sent, and as a result, they cannot be retrieved. In Lotus Notes, draft documents (and tasks without a due date, which are also considered drafts) are always excluded from archiving. In Outlook, you should explicitly specify the monitored folders so that the Drafts folder is not included.

When you add folders to the list, enter the following placeholders so that the correct folder is identified for both supported email clients, no matter which language the client is set to:

- `/%OUTBOX%` for the *Outbox* folder in the English version of Outlook
- `/%SENT%` for the *Sent Items* folder in the English version of Outlook or the *Sent* folder in the English version of Lotus Notes
- `/%DRAFTS%` for the *Drafts* folder in the English version of Outlook or Lotus Notes
- `/%TRASH%` for the *Deleted Items* folder in the English version of Outlook or the *Trash* folder in the English version of Lotus Notes

- /%JUNK% for the *Junk E-mail* folder in the English version of Outlook or the *Junk Mail* folder in the English version of Lotus Notes
- /%RSS_FEEDS% for the *RSS feeds* folder in the English version of Outlook
- /%SYNC_ISSUES% for the *Sync issues* folder in the English version of Outlook
- /%INBOX% for the *Inbox* folder in the English version of Outlook or Lotus Notes.

Make sure to use a forward slash (/) with the placeholders for both Outlook and Lotus Notes.

You can still specify folders in the format that is specific to the email client, for example, /Sent Items for Outlook or ($Sent) for Lotus Notes. However, this notation works only for one type of email client. You must use the native notation if a generic placeholder does not exist for the folder that you want to specify.

For Lotus Notes the following considerations apply:

- Folders created by IBM Content Collector are not available for users until they close and reopen their mailboxes. You should specify only folders that already exist.
- If Content Collector automatically creates a folder, and a client user manually creates a folder with the same name, two folders with identical names exist in the client user's mailbox. IBM Content Collector might not find messages that were added to the monitored folder because IBM Content Collector recognizes only the folder that was created first, while users see only the folders that they created.

7. Save your settings.

Related reference

"Collector schedules" on page 458

Collecting documents for life cycle processing:

For life cycle processing of documents, you must set up a task route that contains a stubbing collector (**EC Process Email Stubbing Life Cycle**). With such a stubbing task route, you reduce the amount of content in the email, its attachments, or both. When this content is reduced, IBM Content Collector leaves a stub document in the source location. You define the content of the stub document in the collector settings, for example, it might contain only the header of the original note and a link to the archived content.

Prerequisites: Before IBM Content Collector can collect documents, you must configure a source connector. Otherwise, Content Collector cannot access the source system.

A stubbing collector (**EC Process Email Stubbing Life Cycle**) collects email at regular intervals to remove:

- Attachments
- The body text
- The remaining email documents

The content, the attachments, or the entire email is removed based on the date when the email was archived, modified, received, or restored, and other status information in the email.

For example, a stubbing collector is configured in a way that attachments are removed from the source document three months after archiving. Six months after archiving, Content Collector deletes the entire email. An email that is not yet archived or stubbed contains no status information, so the stubbing collector will not collect it for processing. After the email is archived in an archiving task route, the status of the email is archived. The stubbing collector now collects the email, and, if the document was archived three months ago, removes the attachments, which leaves an email with body text and links to the archived attachments. The status of the email is now archived and attachments removed.

The next time the collector processes the email, the collector checks the status of the email and how much time passed since it was archived. As soon as the specified interval of six months after archiving passed, Content Collector deletes the remaining stub document from the source location.

To configure a stubbing collector:

1. Open the Configuration Manager and click **Task Routes**.
2. Create or select the task route to which you want to add the collector.
3. In the **Toolbox**, click **Email** → **EC Process Email Stubbing Life Cycle** and add it to the task route diagram.
4. On the General page in the right pane, define general settings.
   a. Specify a name and a description for the collector.
   b. If you do not want to use the collector right away, deselect **Active**.
5. On the Schedule page, set a collector schedule by specifying when and how often the collector checks the mailboxes for email to collect for stubbing.
6. On the Collection Sources page, configure one or more collection sources, depending on your source system. Typically, this will be a mailbox collection source. However, you can also select a journaling collection source, if you do not want to remove journal mails right after archiving them.
7. On the Lifecycle page, define a stubbing life cycle by selecting stubbing options. Select stubbing options for email that is in one of the following states:
   - Email that was archived
   - Email that users marked for stubbing
   - Email that was restored, but was not restored from a search result list
   - Email that mobile users copied to an offline repository after it was archived. When the email is copied, its status is set to mobility done. This status means that IBM Content Collector does not wait for the delayed stubbing interval to pass, but stubs the original email the next time that the stubbing collector runs.

   Restriction: A stubbing sequence as defined in a document life cycle cannot be applied to documents that were restored from a repository that was fed from IBM CommonStore for Exchange Server or from IBM CommonStore for Lotus Domino. The information about the document state in these documents cannot be interpreted by IBM Content Collector because it is in an incompatible format.

   Important: The address information for the IBM Content Collector server becomes an unchanging part of the link in the stub document. To avoid problems with the generated links, specify the fully qualified host name (for example, `ICCServer.example.com`) of the machine that runs the Web Application Server, or the respective alias, in the Web applications configuration under

General settings. Ensure that this host name or alias is resolved properly by the DNS. If the host name of the server changes, stub links will not work.

For each selected stubbing option, set the time when you want this to happen and select whether this time is calculated relative to the date the email was received, archived, or modified.

Choose from these options:

Remove nothing and add text
Add text to original email indicating that content was archived by Content Collector.

Remove attachments
Remove the attachments of email after archiving.

Remove attachments and cut body
Shorten the body text of email after archiving. Content Collector replaces the original formatted text with a plain text representation that is cut off at the specified length, where line breaks are preserved. Other formatting, however, is not preserved.

Remove attachments and body
Remove all of the body text in the email after archiving.

Delete entire email
Delete the email after archiving.

Select documents to re-create stubs
Stub email again after users restored the content.

Important: Email that was archived with CommonStore for Lotus Domino or CommonStore for Exchange Server and is restored with IBM Content Collector will not be restubbed.

Select restored documents for deletion
Delete documents that were restored from a search result list.

You specify the text that indicates archiving or stubbing operations in the postprocessing task for stubbing (EC Create Email Stub task). You must include a postprocessing task in your task route even if you use a stubbing collector because stubbing operations must occur after archiving. Through its position in the task route, the postprocessing task for stubbing ensures that the stubbing actions that are selected in the stubbing collector are done at the right time.

8. Save your settings.

Related reference

"Lotus Notes collection sources for automatic archiving" on page 462

"Microsoft Exchange collection sources for automatic archiving" on page 464

Collecting SMTP documents:

Configure an SMTP collector (**SC Collect All Email**) to collect SMTP/MIME email that is received by the SMTP Receiver. This collector runs at preset intervals and collects all documents from the message queue directory that is defined in the SMTP Connector.

Prerequisites: Before IBM Content Collector can collect documents, you must configure a source connector. Otherwise, Content Collector cannot access the source system.

To configure an SMTP collector:

1. Open the Configuration Manager and click **Task Routes**.
2. Create or select the task route to which you want to add the collector.
3. In the **Toolbox**, click **SMTP → SC Collect All Email** and add it to the task route diagram.
4. On the General page in the right pane, define general settings.
   a. Specify a name and a description for the collector.
   b. If you do not want to use the collector right away, deselect **Active**.
5. On the Schedule page, set a collector schedule by specifying when and how often the collector checks the message queue directory for documents to collect.
6. Save your settings.

Related tasks

"Enabling the collection of additional archiving information" on page 431

Related reference

"Collector schedules" on page 458

Collecting file system documents

Configure a file system collector to collect documents from specified locations in your file system, according to the schedule and the filter criteria that you define.

Prerequisites: Before IBM Content Collector can collect documents, you must configure a source connector. Otherwise, Content Collector cannot access the source system.

A file system collector (**FSC Collector**) collects files from the file server and submits the files to the appropriate task route for further processing. The collected files are processed as defined in the task route, most commonly they are archived into a repository.

To configure a file system collector:

1. Open the Configuration Manager and click **Task Routes**.
2. Create or select the task route to which you want to add the collector.
3. In the **Toolbox**, click **File System Source → FSC Collector** and add it to the task route diagram.
4. On the General page in the right pane, define general settings.
   a. Specify a name and a description for the collector.
   b. If you do not want to use the collector right away, deselect **Active**.
   c. To calculate a hash key for each file collected, select **Generate hash key**.

      The hash key is a unique identifier that prevents duplicates from being collected. If files have identical hash keys, only one of the files is collected for processing during one collector run. The duplicate files will be collected during the following collector runs.

      If you check this box, FSC Duplicate Management metadata will be associated with the collected items. The hash key value will be available in the Hashkey property of this metadata. Leaving this box unchecked means that, should you later add deduplication to your collecting, you cannot retroactively deduplicate documents archived prior to your generating hash keys.
   d. If you want to collect information about the content type of the collected document, select a value for **Content type information**. The following options exist:

Do not collect
No content type information will be collected. Therefore, a default icon will be shown for the archived files.

Collect from source system
The file system collector attempts to get content type information from the registry of the system where the file is located. If icon information exists in the registry, the respective icons will be shown for the archived files.

- If the file is on a local drive, for example, `C:\` or `D:\`, the source for the content type information is the local Windows registry.
- If the file is on a network share, for example, `\\machinename\sharename\`, the source for the content type information is the Windows registry of the machine where the network share is located.

If content type information is not available for any reason, the value for the properties will be an empty string.

Collect from source system and local system
The file system collector attempts to get the information from the registry of the system where the file is collected as well as from the registry of the system where IBM Content Collector is installed. If no content type information is available and the file is on a network share, the collector attempts to get the content type information from the local registry. If icon information exists in either registry, the respective icons will be shown for the archived files.

Collect from local system
The file system collector attempts to get the information from the local registry. If icon information exists in the registry, the respective icons will be shown for the archived files.

To reduce performance cost associated with collecting content type information, the file system collector tries to cache the information it retrieves from the registry.

The File System Source Connector must be configured with user credentials with the appropriate access permissions.

5. On the Schedule page, set a collector schedule by specifying when and how often the collector checks the monitored folders for files to collect.
6. On the Collection Sources page, specify the file system folders to be monitored by the collector. These folders can be local folders or folders on a shared NTFS network drive.
   - To collect files from a local folder, enter the complete path from the root drive to the folder that is to be monitored, for example, `C:\My_Documents\Folder1`.
   - To collect files from a folder on a network drive, you must enter a Universal Naming Convention (UNC) path to the folder, in the format `\\<servername>\<sharename>\directory`.

     Important: You cannot collect files from mapped network drives such as `F:\folder_name\foldername`. These drive mappings are specific to an individual user and are not recognizable on a network.

Tip: When you configure a task route for archiving to FileNet P8 and want to configure the P8 Create Document task with the **Reference external content** option, the collection source must be a shared directory with the appropriate

permissions on the share to allow access for the FileNet P8 Content Engine and Application Engine (Workplace) servers. When you configure the path for collection from a shared directory, the collection-source directory must be in the format `\\servername\shared_folder\`, so that the FileNet P8 content reference is in a recognizable format for the FileNet P8 system.

If you want the collector to search for files in subfolders, select **Monitor sub-folders** and enter the number of folder levels to search under the specified root folder in the **Folder depth** field. For example, if you specified `C:\My_Documents` in the **Monitored folder** field and you also want to search for files in `C:\My_Documents\2008`, type `1` in the **Folder depth** field. If you do not select **Monitor sub-folders**, only the folder that is specified in the **Monitored folder** field is searched.

Also specify where the flags for the postprocessing of the collected files are stored:

NTFS post processing
In Windows NTFS, an alternate stream set at the file system layer is provided so that extra information can be persisted with the file. Content Collector sets postprocessing flags on the secondary streams of the collected files. This is the default.

Control folder post processing
Content Collector writes postprocessing tags to a control file. You must also specify the name of the folder where Content Collector stores the control files. The folder is created in the folder from which the source was collected. For each file that Content Collector processes, a control file with the same name is written to the control file folder.

When you delete the collected file or the shortcut file, the control file will remain in the control file folder. When you move or rename the collected file or the shortcut file, the postprocessing tags for that file are lost because the control file is not automatically moved or renamed.

7. Specify filter criteria to collect a subset of file system documents from the selected collection sources.
8. Save your settings.

Related tasks

"Moving documents off the network into IBM FileNet P8" on page 725

"Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records" on page 726

"Defining metadata to be used to process files for archiving" on page 729

Related reference

"Collector schedules" on page 458

"Collection filter for a file system collector"

Collection filter for a file system collector:

Filtering enables you to include or exclude files from a collection, based on criteria such as file extension or whether the file is read-only, can be accessed, or was processed or captured before.

You can combine filtering criteria to collect a precise subset of file system documents.

Table 113. Filter criteria for files

Option	Description
Filter file extensions	You can define inclusion lists or exclusion lists for file extensions, but not both at the same time. Therefore, to include or exclude files on the basis of their extension, select **Include** or **Exclude**. In the **File extension** field of the Filter Extension window, enter the file extension, for example, xml for XML files. **Important:** If you are collecting custom metadata, you must include the extension for your metadata file format: xml for XML files or csv for delimited files.
Ignore documents	Select one or more of the following characteristics: **Already processed** Exclude files that were already processed by the collector whether or not they were captured. **Already captured** Exclude files that were previously marked as captured by the post processing task. **Where access is denied** Exclude files that cannot be accessed by the collector, for example, because the user ID that the collector uses has no permission to do so.
File attribute filter	Include or exclude files with specific attributes. For example, you can select to process files that are specified as read-only. Select one or more of the listed attributes and specify whether you want to collect files with this attribute or not.
Date filters	Exclude files based on their date properties. Select one or more of the listed date properties and specify an absolute or a relative date value.
Minimum file size in bytes	Exclude files that are smaller than the specified size.
Maximum file size in bytes	Exclude files that are larger than the specified size.

Related tasks

"Collecting file system documents" on page 479

Collecting from Microsoft SharePoint sites

Configure a SharePoint collector to collect documents from SharePoint sites. A SharePoint collector can process multiple collection sources at specified intervals.

Prerequisites: Before IBM Content Collector can collect SharePoint documents, you must configure a SharePoint connector.

To configure a SharePoint collector:

1. Open the Configuration Manager and click **Task Routes**.
2. Create or select the task route to which you want to add the collector.
3. In the **Toolbox**, click **Microsoft SharePoint → SP Collector** and add it to the task route diagram. Each SharePoint task route template already includes a collector.
4. On the General page in the right pane, define general settings.
   a. Specify a name and a description for the collector.
   b. If you do not want to use the collector right away, deselect **Active**.
   c. To process custom metadata, select **Add user-defined metadata to output** and select a collection of metadata that you previously defined in **Metadata and Lists → User Defined Metadata**.
   d. To collect only documents that a task route has previously collected, select **Collect only previously migrated documents**.

      Important: Use this option only under well-defined scenarios such as link management task routes, audits of previous collections, and re-collection.
5. On the Schedule page, set a collector schedule by specifying when and how often the collector checks the monitored libraries for documents to collect.
6. On the Collection Sources page, configure one or more collection sources.
7. Save your settings.

Related reference

"Collector schedules" on page 458

"Microsoft SharePoint collection sources"

Microsoft SharePoint collection sources:

A SharePoint collector can process multiple collection sources, each with its own set of libraries and lists, content types, and metadata mappings.

A SharePoint collector (**SP Collector**) collects any document that meets these conditions:

- Belongs to the selected site or subsite and, if configured, its children
- Belongs to a selected folder or its subfolders
- Belongs to a selected library or blog post list
- Is of, or is derived from, a selected content type
- Is not checked out
- Was not created or modified within the past 30 seconds
- Has a *Migrated* column value of `No`, unless you set your collector to collect previously migrated documents

  If you migrated from IBM FileNet Connector for Microsoft SharePoint Document Libraries to Content Collector, the column *P8ArchiveDate* must be empty.

By default, the collector includes only SharePoint system metadata, but you can also add your own user-defined metadata to the output.

Site connection

Each collection source can collect from just one configured SharePoint connection, so you must add a collection source for each SharePoint site that you want to process. An empty connection menu means that you must configure a valid SharePoint connection.

Libraries

You can collect documents of all SharePoint library types and the Blog Post list type.

You can collect from all libraries and supported lists or a subset of libraries and supported lists.

Table 114. Library and list selection

Collection source	Description
Collect from all libraries and supported lists	Collects from all libraries and lists on a site and, optionally, all libraries and lists on all child sites.
Collect from selected libraries and supported lists	Collects from the libraries and lists that you select, starting with the root folder or a child folder that you specify and including all its children. If you select multiple libraries or lists, you must ensure that each library or list contains the same path to the folder that you specify here. Libraries and lists that do not contain this path are not processed but do not interfere with the processing of other libraries and lists. The log records an error that specifies the unprocessed libraries and lists. **Blogs:** To collect SharePoint blog posts, you must select the *Posts* list and the *Post* content type.

Important: Collecting from certain libraries, such as Form Templates and other system-like libraries that contain style sheets or other data that the system uses, can render the SharePoint site inoperable by deleting or modifying the security settings of critical documents. Before setting up your task routes, especially those that delete original documents, you must carefully analyze your SharePoint site collection contents to determine what to collect and what not to collect.

Content types

You can collect one or more SharePoint content types. Selecting a parent content type also collects all its children. Selecting *Document*, which is typically the parent of many content types, selects several other content types. To collect from libraries that are not configured to manage multiple content types, select *Document*. To collect SharePoint blog posts, you must select the *Post* content type.

The collection of multiple content types such as blogs and wikis requires a method for keeping such content attached to its metadata while remaining distinct and searchable. The collector processes blogs, which typically merge many distinct entries that contain their own metadata, by combining each blog post, its comments, and its metadata into an individual HTML document for storage and

retrieval. The HTML document uses the default styles that SharePoint 2007 uses to render blogs. To ensure that graphics render properly and links do not break, the collector converts relative URLs (which SharePoint uses if the graphic or other linked content belongs to the same application as the blog) to absolute URLs. Removal of the embedded content from SharePoint breaks the links.

If you collect multiple versions of blog posts and comments, the collector associates the most recent version of the post with the most recent version of a comment, with one exception: the collector appends all comment versions that the commenter added after the most recent version of the post. The collector does not archive attachments to blog comments. To remain consistent with native SharePoint behavior, Post removal does not remove comments.

Restriction: Do not select **Replace with link** if you are collecting blog posts or any SharePoint library, such as Slide or Wiki, that does not support multiple content types. The attempt to replace blog posts with links in SharePoint fails and the content remains unchanged. You can use this option with wiki content, but the wiki content remains in SharePoint, giving this option the same effect as Mark as processed.

If document versions switch content types, Content Collector collects the document based on the content type of the most recent version. To ensure that you collect all such documents, identify the most likely content type sequences and select all of those content types for collection. For example, select Draft, Review Draft, and Final.

Remember: You can collect documents of the *Report* content type within a *Report* library type, but the collector does not collect *report history* documents.

Columns

You can map multiple columns and user-defined metadata properties that your repository recognizes. For example, you can map a SharePoint site column called *Home Address City* to a user-defined metadata property called *Home city*, then map *Home city* to the *Home* property in your repository.

You can map only columns and user-defined metadata properties of the same data type.

The **Metadata source** is the value that you selected from the **Metadata** menu on the **General** page. You can apply mappings to the entire site and override the site column mappings for any lists or libraries that you selected on the **Libraries** page. In the table the overriding mappings display with an **Inherited** value of *No*, while mappings that persist across all site columns have an **Inherited** value of *Yes*. To remove an inherited mapping you must change the source to *Site* and remove the site column mapping.

You cannot generally map or collect:

- Site or list columns of the data types *Calculated*, *Lookup*, or *Summary Links data*.
- Site or list columns created specifically within child sites, unless you create a separate connection and collection source that specifies the child site at the top level. However, you can map inherited site columns.

Lock Down

You can grant one or more users and groups full control over SharePoint documents that were archived and locked to restrict changes to the originals. Any users or groups who are eligible to edit the original document and are not listed can only read the document. Site Collection Administrators have Full Control over all documents.

To load a list of the site collection's users or groups, select **Suggest existing users** or **Suggest existing groups**, or both.

To lock SharePoint documents that this task route processes, select `Lock down` in the **SP Post-processing** task.

Document security

In a task route that archives to FileNet P8, you can add a P8 Modify Security task to map SharePoint document permissions to FileNet P8 access rights:

Table 115. SharePoint to FileNet P8 permissions

SharePoint permissions	FileNet P8 access rights
Full Control	Full Control
Edit Items	Modify all properties, Major versioning, Minor versioning, Create instance
Delete Items, Delete Versions	Delete
View Items	Read permissions, View all properties, View content
Open Items	View all properties, View content
View Versions	Read permissions, View all properties, View content
Manage Permissions	Modify permissions

Archiving an entire SharePoint site

You cannot use a SharePoint task route to archive an entire SharePoint site. Instead, first you use SharePoint tools (commonly the stsadm or powershell tools) to create a single backup file, ensuring that you observe any restrictions that the SharePoint documentation contains. Then you configure a file system task route, typically *FS to P8 Archiving (Delete)* or *FS to CM8 Archiving (Delete)*, to archive the backup file to your target repository. Alternatively, you can use repository tools to archive the backup file.

Restriction: Ensure that the SharePoint backup file does not exceed the maximum file size recommended for your target repository.

Related tasks

"Collecting from Microsoft SharePoint sites" on page 482

Re-collecting archived Microsoft SharePoint documents:

You can enable re-collection of previously archived documents from SharePoint through minimal configuration of existing task routes.

Users sometimes want to edit documents already collected from SharePoint and archived in a target repository. If you have previously configured a version-enabled task route and selected either the *Mark as processed* or *Lock down* option in the SP Post-processing task, you can enable your users to edit SharePoint documents for re-collection.

Re-collection strategies

The best method for enabling re-collection is to select *Mark as processed* and configure a new collector to use the *Collect only previously migrated documents* option. This strategy ensures the seamless archiving of all targeted SharePoint documents by requiring minimal configuration and, more important, no special action from your SharePoint users. The other three methods require your users to perform one or more actions that, while relatively simple, can add to your administrative workload and result in incomplete archiving of documents from SharePoint. The table shows the four strategies and the actions that each requires of administrators and users:

Table 116. Re-collection configuration strategies

Strategy	Post-processing option	**Adds *SP Collector* for re-collection?**	Administrator actions	User actions
1	*Mark as processed*	Yes	Add a re-collecting *SP Collector* to each active SharePoint task route.	None
2	*Lock down*	Yes	1. Add a re-collecting *SP Collector* to each active SharePoint task route. 2. Distribute re-collection instructions to your SharePoint users. 3. Grant or deny edit permissions when users request them.	If you cannot edit the document, ask the administrator to grant permission.
3	*Mark as processed*	No	1. Add the *Migrated* column to existing SharePoint content types. 2. Distribute re-collection instructions to your SharePoint users.	1. Verify that the content type of your document contains the *Migrated* column, and add the column if necessary. 2. When you are ready to have the document re-collected and archived to your content repository, deselect the *Migrated* check box.

Table 116. Re-collection configuration strategies (continued)

Strategy	Post-processing option	Adds *SP Collector* for re-collection?	Administrator actions	User actions
4	*Lock down*	No	1. Add the *Migrated* column to existing SharePoint content types. 2. Distribute re-collection instructions to your SharePoint users. 3. Grant or deny edit permissions when users request them.	1. If you cannot edit the document, ask the administrator to grant permission. 2. Verify that the content type of your document contains the *Migrated* column, and add the column if necessary. 3. When you are ready to have the document re-collected and archived to your content repository, deselect the *Migrated* check box.

Though they can result in more user errors and calls to support, strategies 3 and 4 can give your users manual control over re-collection, typically to conserve network bandwidth and server space by reducing unnecessary archiving of incremental saves. Otherwise, strategies 1 and 2 are far better choices, requiring little setup and little or no user action.

Configuring re-collection

Depending on the strategy you select, you must perform one or more tasks to enable re-collection:

- Add a second instance of *SP Collector* to each of your active SharePoint task routes (strategies 1-2). To re-collect the same document collection that your task route collects, you copy and paste the existing collector, which maintains the same collection sources and metadata mapping. You configure the collector for re-collection by selecting *Collect only previously migrated documents* on the General page of the collector. You might want to change the schedule to re-collect less frequently.
- Add the *Migrated* column to existing SharePoint content types (strategies 3-4). If your SharePoint users can create new content types, they must add the *Migrated* column and select *Migrated* to enable re-collection. You can, however, minimize user confusion by adding the column to all existing content types that you collect.
- Distribute re-collection instructions to your SharePoint users (strategies 2-4). The User actions column of the table details the actions required for each strategy. If your users are unfamiliar with columns in SharePoint, you can supplement the instructions with the help that Microsoft provides. Obviously, this step is especially critical to the success of strategies 3 and 4.

- Grant or deny edit permissions when users request them (strategies 2 and 4). You can reduce the number of requests by granting Full Control privileges to individual users or groups from the Lock Down page of each collection source.

Configuring tasks

Tasks can perform transformations of different types on collected documents. They can also be used to read metadata from a document, to look up metadata on external systems, or perform actions that do not necessarily modify the original document. You add tasks to task routes in the Configuration Manager.

Various tasks can be applied to a document while it is moved through a task route. Some tasks transform work on the collected source, while others prepare documents for the storage in a specific repository. The source-related tasks typically occur earlier in a task route or as post-processing task; the repository-related tasks are closer towards the end. You can find reference information for all tasks in "Task reference" on page 496.

Verifying task settings and adding further processing

Some tasks require additional configuration before they are valid and can perform an action on documents.

When you load a task route from a template, the configuration settings are stored in the task route template are displayed. You should check if these settings are correct for your setup. Configuration settings that rely on external system metadata of the source system for example will not be filled in.

Check the following kinds of configuration settings:

- Collection sources
- File paths
- LDAP locations
- Additional custom metadata that must be defined before the task route can be configured.

  For example, the user-defined metadata property called **Appointment End Date** of type Date Time must be defined when archiving Microsoft Exchange calendar entries.
- The document classes to use in FileNet P8 task routes
- The item type to use in Content Manager task routes
- In file system task routes, you must specify the server name in the shortcut URL in the P8 4.x Create Document task.

The sample task route templates include typical document archiving setups. However, the email archiving templates do not include tasks that influence how documents can be organized in an archive:

- To classify archived documents into defined categories, see "Using Classification Module to classify documents" on page 453.
- FileNet P8 only: To declare collected documents as records, see "Managing document retention" on page 438.

To adapt task settings:

1. Click the task in the task route. The configuration options for the task are shown in the configuration pane.

2. Edit the configuration values according to your requirements. See the "Task reference" on page 496 for detailed information on each task and its configuration options.
3. Ensure that the configuration is valid.
   - If an error icon ( ) is displayed, something is missing in the configuration.
   - If the metadata error icon ( ) is displayed, dependent metadata is missing.

Assigning property values

You can configure property mappings for document classes or folder options. A property mapping is an expression that populates the source document class, object class, link class property, or item type attribute with a value. When you have to provide values for properties, you have several options to do so.

1. In the table listing the properties, select the table row of the value you want to set, and click **Edit**.
2. Select one of these options in the pop-up window:
   - To provide the same value for every document added using this task or this path in the task route, select **Literal** and enter a value to assign to the property.
   - To use a metadata property as the value, select **Metadata**.
   - To use a regular expression to search for patterns in metadata to use as the property value, select **Regular Expression**.
   - To concatenate one or more metadata values with one or more static values to create a new value, select **Calculated value**.
   - To create a value that is conditional upon criteria you specify being met, select **Conditional value**.
   - To do a case-insensitive search to see if a property contains an item in a list and return the first match found, select **List Lookup**. This option is available only if you have previously configured lists in the **Metadata and Lists** section of this application.
3. Click **OK** to save your settings. In some cases, you have the option to undo changes you have made by clicking **Reset to default value**. If a default value for the property has been set in the repository, that value will be displayed.

Display of property mappings:

This topic describes how property mappings of a document class or an item type display.

- **Bold**, *italic* = required
- <hidden> = hidden in FileNet P8

  Note: Hidden properties will be shown only if:
  - The **Show "Hidden Properties"** button is clicked in the P8 4.x Create Document task.
  - A value has been mapped to the property in the current window.
  - No value has been mapped, but the field is required and has no default value in FileNet P8.

  Only configurable hidden properties will be available for display. If no hidden properties are shown when you click **Show "Hidden Properties"**, no hidden properties are available for configuration for that particular FileNet P8 object class.

- <system> = FileNet P8 system properties that can be set

  Note: Certain system properties will be shown only if:

  – The **Show "System Properties"** button is clicked in the P8 4.x Create Document task.

    The button to show the system properties is only available if you have permission to modify system properties on the FileNet P8Content Engine. To be able to map certain Content Engine system properties, your object store security must be set to **Modify certain system properties**. This right is not granted to Content Engine administrators by default, it must be granted explicitly. After granting this right, the IBM Content Collector Configuration Manager must be restarted.

Using regular expressions in property mappings:

When configuring property mappings, you can create regular expressions to parse system and user-defined metadata properties.

The properties must have a string data type. This is because the regular expression will return only a string as a result of the expression being evaluated.

For example, if you want to find a contract number contained in an email and use it as a value of a property such as *Contract Number*, you can use a regular expression to extract the contract number from the email property you specify. To do that, however, the *Contract Number* property must be of the type *string*.

The Configuration Manager supports two types of regular expressions for property mappings:

- Replacement regular expressions - this is akin to using search and replace
- Matches regular expressions - this allows you to check if a certain value exists within the metadata property being tested

To use a regular expression to obtain a value:

1. Select **Regular Expression**.
2. Click the **...** button.
3. In the **Select Metadata** section, select a metadata type and property you want to be searched for the regular expression.
4. In the **Define Regular Expression** section, select an option:

   Define replacement regular expression

   This option allows you to provide a pattern for which to search, a value to substitute if that pattern is found, as well as a default value to substitute if that pattern is not found.

   For example, an insurance company with different case number formats for different claim types could search for email with a case number matching the pattern specified in the subject line. Once found, the case number could be substituted with the phrase *Automobile Claim*. In addition, the default title of *Invalid Claim Number* could be assigned to all email with a subject not fitting the pattern. This title could be used as a flag to kick off a workflow whereby a claim number in the valid format is found and assigned to the email (note that such a workflow is outside the scope of this system).

   a. In the **Replacement regular expression** text box, enter the regular expression that will be applied to the metadata property.

b. In the **Replacement string** text box, enter the string to substitute if the pattern defined by the regular expression is found.
c. In the **Default value** text box, enter the string to substitute if the pattern defined by the regular expression is not found.

Define matches regular expression

This option allows you to provide a pattern for which to search, as well as a default value to substitute if that pattern is not found.

For example, you could search for email with a contract number matching the pattern you specify in the subject line, and use that contract number as the document title in the repository. In addition, you could specify the default title of "Invalid Contract Number" to all email not fitting the pattern, and use this title to kick off a workflow whereby a contract number in the valid format is found and assigned to each email (note that this workflow is outside the scope of this system).

a. In the **Matches regular expression** text box, enter the regular expression that will be applied to the metadata property.
b. In the **Default value** text box, enter the string to substitute if the pattern defined by the regular expression is not found.

Related concepts

"Regular expressions" on page 421

Testing the regular expression:

- To test a replacement regular expression:
  1. In the **Test value** section, in the **Test value** text box, enter a value that should be flagged as matching the pattern indicated in the regular expression you entered.
  2. Click **Test**. The following will result:
     - The **Matches** text box displays in bold the portion of the value that matches the pattern specified in the regular expression and any non-matching remainder text in non-bold, or if no match is found, displays `{None}`.
     - The **Result** text box displays the replacement string (concatenated with the non-matching portion of the string, if any), or if no match is found, displays the default value (if you entered one).
- To test a matches regular expression:
  1. In the **Test value** section, in the **Test value** text box, enter a value that should be flagged as matching the pattern indicated in the regular expression you entered.
  2. Click **Test**. The following will result:
     - The **Matches** text box displays in bold the portion of the value that matches the pattern specified in the regular expression and any non-matching remainder text in non-bold, or if no match is found, displays `{None}`.
     - The **Result** text box displays the replacement string concatenated with the matching portion of the value, or if no match is found, displays the default value (if you entered one).

Calculated values:

Calculated values allow for the manipulation of email or file properties to create a new value.

For example, you could concatenate a contract number with a user name to create a value such as "123ABC_Smith".

To create a calculated value:

1. Select **Calculated value**.
2. Click the **...** button.
3. In the Edit Calculated Value window:
   a. To edit a value, select the value in the row and click **Edit**, or to add a value, click **Add**.
   b. In the Edit Expression window:
      1) Select one of the following:
         - To enter a static value on which to base the calculation, select **Literal**.
         - To select a metadata type and property on which to base the calculation, select **Metadata**.

           In most cases when selecting a metadata type, you can select **Email** or **File**. The properties available will change depending on the metadata type selected.
      2) There is only one available operator (+). This operator is a concatenation operator to join the selected values together.
      3) Click **OK**.
      4) Repeat steps 3b1 to 3b3 to create a second value for the calculation.
      5) Repeat to add as many values as you want, clicking **OK** in the Edit Calculated Value window when complete.

Conditional Values:

Conditional values allow you to specify a condition to be met in order to assign a value, and to assign an alternate value if the condition is not met. You can think of this as a *If x, then y, else z* expression.

To create a conditional value:

1. Select **Conditional value**.
2. Click the **...** button.
3. In the Edit Conditional Value pop-up window, do one of the following:
   a. To edit a value, select the value in the row and click **Edit**.
   b. To add a value, click **Add**.
4. In the Configure Expression Clause pop-up window, from the **Metadata type** list box, select a metadata type and property. In most cases, you can select **Email** or **File**. The properties available will change depending on the metadata type selected.
5. From the **Operator** list box, select an operator for the expression.
6. In the **Value** section, do one of the following:
   - Select **Literal** and enter the value.
   - Select **Metadata** and select the Metadata type and property.
   - Select **List** (available only if you have created a list and the operator is one of "Is contained in", "Not is contained in", "Is like in", or "Not is like in"), and choose a list from the list box.

7. In the **Return Value** section, select one of the following:
   - Select **Literal** and enter the value.
   - Select **Metadata** and select the Metadata type and property.
8. Click **OK**.
9. Enter a default value to be used if the condition you've specified never resolves to true. This is the final *else* portion of the conditional clause.
10. Click **OK**
11. Save the configuration settings.

Related tasks

"Adding and editing metadata" on page 237

Operators for conditional values:

You can use different operators to construct conditional values. Which operators are available depends on the data type of the metadata property.

The operators are used to compare the value or values of a metadata property (referred to as value *A* in the table below) to some value that you specify (value *B* below). After you select the metadata property, only the operators that are suitable for this data type are available.

Table 117. Operators for conditional values

Operator	Data types (*A*, *B*)	Description
IsLikeIn	(string, string) (string, string array)	IsLikeIn (*A*, *B*) is true if value *B* or one of the values in array *B* is contained in value *A*. This operator is not case-sensitive. For example, the following test would return true: `IsLikeIn ("ABcdEF","bC")` You cannot use wildcard characters with this operator.
NotIsLikeIn	(string, string) (string, string array)	NotIsLike in (*A*, *B*) is true if value *B* or all of the values in array *B* are not contained in value *A*. This operator is not case-sensitive. For example, the following test would return true: `NotIsLike in ("ABcdEF","bCa")` You cannot use wildcard characters with this operator.
IsContainedIn	(string, string) (string, string array)	IsContainedIn (*A*, *B*) is true if value *A* is a substring of value *B* or equal to one of the values of array *B*. For example, the following test would return true: `IsContainedIn ("bc","abcdef")`
NotIsContainedIn	(string, string) (string, string array)	NotIsContainedIn (*A*, *B*) is true if value *A* is not a substring of value *B* and not equal to any of the values of array *B*. For example, the following test would return true: `NotIsContainedIn ("bC","abcdef")`
Equal	(bool, bool) (int, int) (string, string)	Equal (*A*, *B*) is true if value *A* is equal to value *B* (they must be an exact match). This operator is case-sensitive. You cannot use wildcard characters with this operator.
NotEqual	(int, int) (string, string)	NotEqual (*A*, *B*) is true if value *A* differs from value *B*. This operator is case-sensitive. You cannot use wildcard characters with this operator.

Table 117. Operators for conditional values (continued)

Operator	Data types (*A*, *B*)	Description
IEqual	(bool, bool) (int, int) (string, string)	IEqual (*A*, *B*) is true if value *A* is equal to value *B* irrespective of the case. You cannot use wildcard characters with this operator.
NotIEqual	(bool, bool) (int, int) (string, string)	NotIEqual (*A*, *B*) is true if value *A* is not equal to value *B*. This operator is not case-sensitive. You cannot use wildcard characters with this operator.
RegexSearch	(string, regex) (string array, regex)	RegexSearch (*A*, *B*) is true if the regular expression *B* matches in value *A* or in one of the values in array *A*. For examples, see the topic about regular expressions.
NotRegexSearch	(string, regex) (string array, regex)	NotRegexSearch (*A*, *B*) is true if the regular expression *B* does not match in value *A* or in any of the values in array *A*. For examples, see the topic about regular expressions.
Like	(string, string)	Like (*A*, *B*) is true if value *A* is contained in value *B*. This operator is not case-sensitive. For example, the following test would return true: `Like ("bC","ABcdEF")` You cannot use wildcard characters with this operator.
NotLike	(string, string)	NotLike (*A*, *B*) is true if value *A* is not contained in value *B*. This operator is not case-sensitive. For example, the following test would return true: `NotLike ("bCa","ABcdEF")` You cannot use wildcard characters with this operator.
GroupLookup	(string, list) (string array, list)	GroupLookup (*A*, *B*) is true if the values in list *B* appear in the list of groups that was returned by an LDAP lookup of value *A*.
NotGroupLookup	(string, list) (string array, list)	NotGroupLookup (*A*, *B*) is true if the values in list *B* do not appear in the list of groups that was returned by an LDAP lookup of value *A*.
GreaterThan	(date, date) (float, float) (int, int)	GreaterThan (*A*, *B*) is true if value *A* is greater than value *B*.
NotGreaterThan	(date, date) (float, float) (int, int)	NotGreaterThan (*A*, *B*) is true if value *A* is not greater than value *B*.
LessThan	(date, date) (float, float) (int, int)	LessThan (*A*, *B*) is true if value *A* is less than value *B*.
NotLessThan	(date, date) (float, float) (int, int)	NotLessThan (*A*, *B*) is true if value *A* is not less than value *B*.
Age	(date, int)	Age (*A*, *B*) is true if date *A* is more than *B* days in the past.
Contains	(date array, date) (string array, string)	Contains (*A*, *B*) is true if value *B* is equal to one of the values in array *A*.
NotContains	(date array, date) (string array, string)	NotContains (*A*, *B*) is true if value *B* is not equal to any of the values in array *A*.

Table 117. Operators for conditional values (continued)

Operator	Data types (*A*, *B*)	Description
ContainsSome	(date array, date) (date array, date array) (string array, string) (string array, string array)	ContainsSome (*A*, *B*) is true if value *B* or at least one value in array *B* is equal to one of the values in array *A*.
NotContainsSome	(date array, date) (date array, date array) (string array, string) (string array, string array)	NotContainsSome (*A*, *B*) is true if value *B* or all of the value in array *B* are not equal to any of the values in array *A*.

Related concepts

"Regular expressions" on page 421

Related tasks

"Configuring the settings for LDAP lookups during task route processing" on page 177

List Lookup:

To do a case-insensitive search to see if a metadata property contains a value in a list, and obtain a return value of the first match found, use the List Lookup option.

The List Lookup option is available only if you have previously configured lists in the **Metadata and Lists** section of this application.

For example, if client names always appear in the Subject line of all email, you can select to evaluate the email subject against a list of client names, and return the first name matched, which will be mapped as the value of the property. If customer names may appear in the Subject line, but do not always, you can place an initial *Is Like In* rule in the task route to evaluate whether any customer names appear in the selected list, and send for processing only those that do. You can then use the List Lookup feature later in the task route to obtain the exact value matched in the list, and use this, for example, to file the item into its own folder.

1. Select **List lookup**.
2. Select the list to use in the lookup, and click the **...** button.
3. In the **Edit List Lookup** screen, select the metadata type and property against which to evaluate the items in the selected list.

The selected list name, metadata type and property will be displayed as a List Lookup expression as follows: `List name| <Metadata type, Metadata property>`, or as in the example above, `Clients| <Email, Subject>`.

Task reference

IBM Content Collector provides a variety of tasks that you can use to perform transformations of different types on collected files or email.

Table 118. IBM Content Collector source-related tasks

Source-related	
File System Source tasks	"FSC Associate Metadata" on page 531
	"FSC Post Processing" on page 534

Table 118. IBM Content Collector source-related tasks (continued)

Source-related	
Email tasks	"EC Copy to Mailbox" on page 515
	"EC Create Email Stub" on page 516
	"EC Extract Attachments" on page 522
	"EC Extract Metadata" on page 522
	"EC Finalize Email for Compliance" on page 524
	"EC Prepare Email for Archiving" on page 525
	"EC Prepare Email for Stubbing" on page 526
Microsoft SharePoint tasks	"SP Create File" on page 567
	"SP Get Versions" on page 568
	"SP Manage Link" on page 569
	"SP Post-processing" on page 570
Email received through SMTP tasks	"SC Prepare Email for Archiving" on page 566
	"SC Extract Metadata" on page 565
	"SC Extract Attachments" on page 564
	"SC Prepare Email for Deletion" on page 567
	"SC Delete Email" on page 564

Table 119. IBM Content Collector repository-related tasks

Repository-related	
IBM Content Manager repository tasks	"CM 8.x Associate Content" on page 499
	"CM 8.x Configure Item Types" on page 502
	"CM 8.x Confirm Document" on page 504
	"CM 8.x Create Document" on page 506
	"CM 8.x Duplicate Detection" on page 509
	"CM 8.x Store Version Series" on page 511
	"CM 8.x Update Document" on page 514
File System Repository tasks	"FSR Create Document" on page 536
FileNet Image Services repository tasks	"FileNet Image Services Create Document" on page 528
	"FileNet Image Services File Document In Folder" on page 529
	"FileNet Image Services Modify Permissions" on page 530

Table 119. IBM Content Collector repository-related tasks (continued)

Repository-related	
IBM FileNet P8 repository tasks	"P8 4.x Confirm Document" on page 540
	"P8 4.x Create Content Elements" on page 541
	"P8 4.x Create Document" on page 543
	"P8 4.x Create Email Instance" on page 548
	"P8 4.x Create Version Series" on page 550
	"P8 4.x Declare Record" on page 553
	"P8 4.x File Document in Folder" on page 555
	"P8 4.x Link Documents" on page 557
	"P8 4.x Modify Object Security" on page 558
	"P8 4.x Save Prepared Text as XML" on page 561
Metadata Form tasks	"MC Retrieve Additional Metadata" on page 539
Text Extraction tasks	"Extract Text" on page 527
Utility tasks	"IBM Classification Module" on page 537
	"Calculate Expiration Date"
	"Save Temporary File Copy" on page 563

Calculate Expiration Date:

This task calculates an expiration date for the retention of document content, a date that is based on user name or LDAP group membership, or on any of a range of literal or calculated metadata property values.

Task summary

Table 120. Calculate Expiration Date task summary

Characteristic	**Value**
Task name	Calculate Expiration Date
Main purpose	Calculates a date on which retained document content is eligible for deletion from the repository. The date is based on user name, LDAP group membership, or a value that is obtained from a metadata property.
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Optional in archiving task routes
Placement in task route	Must appear before any task that applies the retention metadata, such as Configure Item Types or Add Email Instance
Produces which metadata?	Calculate Expiration Date, Task Status

Table 120. Calculate Expiration Date task summary (continued)

Characteristic	Value
Configuration options	"Configuration"

Configuration

The task calculates the expiration date from one of two methods:

- **User and date** calculates the expiration date by matching the user associated with the document to a retention period you specify. Metadata selection enables you to specify the metadata that contains the user name and the metadata that contains the date from which the retention period begins. Retention periods map users and groups to a particular retention period.
- **Date expression** gives you a wide range of options for determining the expiration date. You can simply apply a particular date and time to all documents, obtain a date from a metadata property (which could allow users to set an expiration date), calculate a value based on a metadata property, or set a conditional value based on combinations of these properties.

Related reference

"Calculate Expiration Date system metadata properties" on page 247

"Task status system metadata properties" on page 269

CM 8.x Associate Content:

This task is used to archive the parent document and all child documents (both email and attachments) within a single transaction. It can be used only with the compound email data model. The task is an extension to the CM 8.x Create Document task.

Task summary

Table 121. CM 8.x Associate Content task summary

Characteristic	Value
Task name	CM 8.x Associate Content
Main purpose	Archives the email and all its attachments
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM Content Manager Connector
When needed?	Required in email archiving task routes if the compound email data model is used
Placement in task route	Can appear only after the tasks EC Prepare Email for Archiving and EC Extract Attachments
Produces which metadata?	CM 8.x Create Document, Task Status
Configuration options	"Connection" on page 500 "Attachment Item Types" on page 500 "Map attributes to metadata" on page 500 "Checkin Options" on page 500

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Attachment Item Types

Specify the item type in which attachments are stored. Click **Select Item Types** to add a new item type. All item types that you add must be structurally equivalent. This means that they must contain the same number and type of attributes, including the attribute names. You can then specify end dates and metadata to determine which item type is used.

Map attributes to metadata

The table lists the attributes of the selected item types. Because the attributes are the same for all selected item types, they appear only once. Any mapping that you define for an attribute is therefore valid for all selected item types.

Checkin Options

With these options, you determine the way a document is stored in Content Manager.

- An access control list (ACL) is used to restrict access to content to a specific list of users. It must be created on the Content Manager server. To use a different ACL for the document than the default ACL for the item type, select one of those listed under **Access Control List**. The list contains all access control lists that are available on the Content Manager server.
- In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the repository. Based on this definition, Content Collector can generate a shortcut URL for each document that the task is applied to. With this URL, users can access archived documents using the IBM Content Collector Web Application.

  Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
  Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

  ```
  https://host_name:port_number/AFUWeb/RD.do?
  r=%PID_ENCRYPTED&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%
  ```

  To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

  ```
  https://host_name:port_number/AFUWeb/SRD.do?
  r=%PID_ENCRYPTED%&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%&am=%CHALLENGE_MODE%
  ```

  In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.

  – Replace *host_name* and *port_number* with the name and port number of the Content Manager application server.
  – Do not alter any of the tokens *%token_name%* and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
  – If you have customized a user client and require the URL to contain more parameters, you can use the following tokens in the URL:

- %PID_ENCRYPTED%: the encrypted persistent identifier of the item in Content Manager
- %URL_CHECKSUM%: the checksum of the encrypted values
- %REPOSITORY_ID_ENCRYPTED%: the encrypted unique identifier of the Content Manager repository
- %CHALLENGE_MODE%: the access mode for the repository when secure links are used

– The ENCRYPTED tokens are encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service. This means that you cannot use %PID_ENCRYPTED%, %ITEMTYPE_ENCRYPTED%, or %URL_CHECKSUM% with applications that do not use the IBM Content Collector Web Application service.

• Review the default MIMEtype that will be applied to documents checked into the repository by clicking **MIME type configuration**.

The Configuration Manager contacts the Content Manager server to obtain the default MIME types. If the Content Manager 8.x connection is working properly, you see a number of predefined rules in the MIME type configuration window. Briefly summarized, these rules work as in the following example: If the extension of an email attachment or file is bmp, assign it the MIME type image/bmp in the Content Manager repository. Typically, default MIME types are defined for the most common attachment types and file types. However, it may be that you want to create a rule for an additional MIME type or change one of the default rules.

Note: If you add a new MIME type, you must add it to the Content Manager server first, because MIME types that are unknown to Content Manager result in errors.

Metadata type
The object that contains the property which you want the rule to evaluate. For example, select **File** to read the property of a file.

Metadata property
The property to be evaluated by the rule. For example, if the **Metadata type** is **File**, and you want to assign a MIME type to files of a certain type, select **File Extension**.

Operator
Select an operator. See the related topic for a description of all operators.

Specify the values to compare the selected property with. Possible choices are:

Literal A constant value that you type in a text field.

Metadata
The values of another property.

Regex search
The results of a regular expression search.

List A list of values. This option is available only if you have previously configured lists in the **Metadata and Lists** section of this application.

Specify the MIME type to be used. You can choose between the following methods to do so:

– **Literal** assigns the MIME type that you type in the text field
– **Metadata property** reads the MIME type from a property (the property has only one value and this is the MIME type)

Related reference

"Operators for conditional values" on page 494

"CM 8.x Create Document system metadata properties" on page 248

"Task status system metadata properties" on page 269

CM 8.x Configure Item Types:

This task allows you to select item types for email archiving and duplicate detection. In addition, you can select and manipulate email metadata to be stored in repository attributes.

Task summary

Table 122. CM 8.x Configure Item Types task summary

Characteristic	Value
Task name	CM 8.x Configure Item Types
Main purpose	Defines the item type that is used to archive content and to perform duplicate searches and maps document metadata to repository attributes
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM Content Manager Connector
When needed?	Required in archiving task routes
Placement in task route	Any task that produces metadata that is to be mapped to a Content Manager attribute must precede this task.
Produces which metadata?	Task Status
Configuration options	"Connection" "Configure Item Types" "Map attributes to metadata" on page 503

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Configure Item Types

Click **Select Item Types** to add a new item type. The available item types are listed alphabetically. In the Choose Item Types window, type all or part of the item type name that you are looking for in the **Filter** field and click **Apply**.

The box on the left lists the item types whose names contain the specified string. Select one or more item types and click **Add**. All item types that you select must have the same set of attributes. The selected item types are listed in the box on the right. To correct your selection, you can select one or more item types on the right and then click **Remove**.

Click **OK** when your selection is complete. You return to the configuration pane of the CM 8.x Configure Item Types task. The selected item types are listed in a table with the heading **Selected item types**.

Note: If item types are added while the Configuration Manager is running, you will need to restart the application to see them in the list.

If you selected more than one item type, a column with the heading **Last Valid Date** is added to the table listing the selected item types. This allows you to file documents into different item types based on date values. For example, if the last valid date for an item type is set to *Wednesday, September 30, 2009*, and to *Thursday, December 31, 2009* for another item type, all documents with a reference date on or before September 30 are moved to the first item type, and all documents with a later reference date (but earlier than January 1, 2010) are moved to the second item type.

The list of selected item types can be sorted alphabetically by clicking **Item Type** or by date by clicking **End Date**.

Select the **Metadata source** and the **Metadata property** that are to be used for determining the item type.

Note: If a document is processed that exceeds the last valid date of the last item type in the list, an error is logged and the document is not archived. In this case, you must create a new item type, identical to the existing types, on the Content Manager server. Then, update the task to use this new item type.

Map attributes to metadata

The table lists the attributes of the selected item types. Because the attributes are the same for all selected item types, they appear only once. Any mapping that you define for an attribute is therefore valid for all selected item types. To map a property to an attribute, select the attribute in the table and click **Edit**. Inthe Configure Property Mapping window, select a method to obtain the attribute values:

Literal A constant value that will always be stored in the selected attribute. For example, you might want to indicate that an archived document is an email, and therefore type `Email` in the text field.

Metadata
Variable values obtained from a field, such as an email property. For example, by mapping an attribute named Sender to the From email property, you determine that sender email addresses are to be stored in the Sender attribute.

Regular expression
Values resulting from the application of a regular expression to the values of certain properties, including values obtained by running a search and replace operation using regular expressions. Assume, for example, that all the email addresses of the managers of a certain company start with the string mgr. This would allow you to define a regular expression that searches for the string mgr in the From property of captured email. If the string is found, it is replaced with the word Manager, which is then stored in the selected repository attribute to indicate that an archived email was originally sent by a manager.

Calculated value

Values obtained by joining property values or literals. For example, assume that you want to store information about the sender and the recipient of an archived email in a single attribute. To this end, you could define a calculated value that combines the values of the From property with that of the To property. In addition, you could define to place blank characters and a dash between the two so that a single value of the attribute looks like

Jones@xyz.com - Peters@xyz.com

indicating that Jones once sent the archived email to Peters.

Conditional value

Values that are stored in the selected attribute if certain conditions are met. Assume that you want to add some information about the size of an archived email to the repository. You could define a rule that evaluates the number of attachments in an email (the Attachment Count property). Each time the value of this property is 0, the string Small Doc is written to the selected attribute.

Note: When Lotus Notes email is captured, the Folder metadata field is only filled with a name if the corresponding email was obtained through a folder, that is, if a collector for manual archiving is configured to monitor drag-and-drop folders or if a collector for automatic archiving is configured to include folders. For email that is collected in another way, this field will always be empty.

If applicable, click the **...** button to formulate the rule or expression for the chosen method.

Related tasks

"Assigning property values" on page 490

Related reference

"Task status system metadata properties" on page 269

CM 8.x Confirm Document:

For use in SharePoint connector link management and auditing task routes, the CM 8.x Confirm Document task attempts to confirm the existence of a document in IBM Content Manager.

Task summary

Table 123. CM 8.x Confirm Document task summary

Characteristic	Value
Task name	CM 8.x Confirm Document
Main purpose	Checks if a document exists in the IBM Content Manager repository
Usable with which source connectors?	SharePoint Connector
Usable with which target connectors?	IBM Content Manager Connector
When needed?	Required in link management task routes such as **SP Manage CM Links** and **SP Audit CM Links**
Placement in task route	Must appear before the SP Manage Link task
Produces which metadata?	CM 8.x Confirm Document, Task Status

Table 123. CM 8.x Confirm Document task summary (continued)

Characteristic	Value
Configuration options	"Connection" "Shortcut Link"

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Shortcut Link

In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the repository. Based on this definition, Content Collector can generate a shortcut URL for each document that the task is applied to. With this URL, users can access archived documents using the IBM Content Collector Web Application.

Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

```
https://host_name:port_number/AFUWeb/RD.do?
r=%PID_ENCRYPTED&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%
```

To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

```
https://host_name:port_number/AFUWeb/SRD.do?
r=%PID_ENCRYPTED%&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%&am=%CHALLENGE_MODE%
```

In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.

- Replace *host_name* and *port_number* with the name and port number of the Content Manager application server.
- Do not alter any of the tokens *%token_name%* and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
- If you have customized a user client and require the URL to contain more parameters, you can use the following tokens in the URL:
  - %PID_ENCRYPTED%: the encrypted persistent identifier of the item in Content Manager
  - %URL_CHECKSUM%: the checksum of the encrypted values
  - %REPOSITORY_ID_ENCRYPTED%: the encrypted unique identifier of the Content Manager repository
  - %CHALLENGE_MODE%: the access mode for the repository when secure links are used
- The ENCRYPTED tokens are encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service. This means that you cannot use %PID_ENCRYPTED%, %ITEMTYPE_ENCRYPTED%, or %URL_CHECKSUM% with applications that do not use the IBM Content Collector Web Application service.

Related reference

"Task status system metadata properties" on page 269

CM 8.x Create Document:

To save a document in the repository, you need to specify where to save it. You do so by selecting a repository, for example *icmnlsdb*. This will cause item types available in the selected repository to be displayed. You can then select the item type and map the document attributes to the attributes defined in the item type selected. For example, you can specify that the File Name property of a file is mapped to the File Name attribute of the selected item type in the repository. You can also select to use folders in Content Manager.

SharePoint only:
Document security is based on the configured item type, not on SharePoint document permissions.

Task summary

Table 124. CM 8.x Create Document task summary

Characteristic	Value
Task name	CM 8.x Create Document
Main purpose	Map document attributes to item-type attributes and save the document in the repository
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM Content Manager Connector
When needed?	One of the following tasks is required in archiving task routes where the target repository is Content Manager: • CM 8.x Create Document • CM 8.x Associate Content • CM 8.x Store Version Series
Placement in task route	Can appear only after these tasks: • EC Extract Metadata • CM 8.x Configure Item Types • EC Prepare Email for Archiving or EC Finalize Email for Compliance
Produces which metadata?	CM 8.x Create Document, Task Status
Configuration options	"Connection" "Checkin Options" on page 507 "Create Folder " on page 508

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Checkin Options

With these options, you determine the way a document is stored in Content Manager.

- An access control list (ACL) is used to restrict access to content to a specific list of users. It must be created on the Content Manager server. To use a different ACL for the document than the default ACL for the item type, select one of those listed under **Access Control List**. The list contains all access control lists that are available on the Content Manager server.
- In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the repository. Based on this definition, Content Collector can generate a shortcut URL for each document that the task is applied to. With this URL, users can access archived documents using the IBM Content Collector Web Application.

  Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
  Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

  ```
  https://host_name:port_number/AFUWeb/RD.do?
  r=%PID_ENCRYPTED&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%
  ```

  To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

  ```
  https://host_name:port_number/AFUWeb/SRD.do?
  r=%PID_ENCRYPTED%&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%&am=%CHALLENGE_MODE%
  ```

  In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.
  - Replace *host_name* and *port_number* with the name and port number of the Content Manager application server.
  - Do not alter any of the tokens *%token_name%* and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
  - If you have customized a user client and require the URL to contain more parameters, you can use the following tokens in the URL:
    - %PID_ENCRYPTED%: the encrypted persistent identifier of the item in Content Manager
    - %URL_CHECKSUM%: the checksum of the encrypted values
    - %REPOSITORY_ID_ENCRYPTED%: the encrypted unique identifier of the Content Manager repository
    - %CHALLENGE_MODE%: the access mode for the repository when secure links are used
  - The ENCRYPTED tokens are encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service. This means that you cannot use %PID_ENCRYPTED%, %ITEMTYPE_ENCRYPTED%, or %URL_CHECKSUM% with applications that do not use the IBM Content Collector Web Application service.
- Review the default MIMEtype that will be applied to documents checked into the repository by clicking **MIME type configuration**.

  The Configuration Manager contacts the Content Manager server to obtain the default MIME types. If the Content Manager 8.x connection is working properly, you see a number of predefined rules in the MIME type configuration window.

Briefly summarized, these rules work as in the following example: If the extension of an email attachment or file is bmp, assign it the MIME type `image/bmp` in the Content Manager repository. Typically, default MIME types are defined for the most common attachment types and file types. However, it may be that you want to create a rule for an additional MIME type or change one of the default rules.

Note: If you add a new MIME type, you must add it to the Content Manager server first, because MIME types that are unknown to Content Manager result in errors.

Metadata type
The object that contains the property which you want the rule to evaluate. For example, select **File** to read the property of a file.

Metadata property
The property to be evaluated by the rule. For example, if the **Metadata type** is **File**, and you want to assign a MIME type to files of a certain type, select **File Extension**.

Operator
Select an operator. See the related topic for a description of all operators.

Specify the values to compare the selected property with. Possible choices are:

Literal A constant value that you type in a text field.

Metadata
The values of another property.

Regex search
The results of a regular expression search.

List A list of values. This option is available only if you have previously configured lists in the **Metadata and Lists** section of this application.

Specify the MIME type to be used. You can choose between the following methods to do so:

– **Literal** assigns the MIME type that you type in the text field
– **Metadata property** reads the MIME type from a property (the property has only one value and this is the MIME type)

Create Folder

To work with Content Manager folders, select **Create folder**. In this case, further options become available:

Folder errors prevent document creation
If you select this option, the document is not archived in case of a failure to create a folder or a failure to file a document in a folder. If you do not select this option, the document is archived even if a folder error occurs. In this case, a warning message is written to the log file.

Folder name delimiter
Specify the character that separates the elements of the folder path.

Metadata Properties
Specify a set of string and string array properties that provides folder paths.

The strings are used as follows:

• Empty strings are ignored

- Non-empty strings are split on the delimiter character.
- Any empty leading or trailing elements of a string are ignored

For example, if the delimiter is \ and one of the string values is `\\mymachine\myshare\folder1\folder2`, the result is:

mymachine

myshare

folder1

folder2

This represents the set of names that will be used to build a folder path. If after removing any empty elements the result is an empty list, the result is ignored.

Related tasks

"Configuring an IBM Content Manager Connector and connections" on page 200

"Tips to ensure proper placement of tasks in a task route" on page 278

"List Lookup" on page 496

Related reference

"Operators for conditional values" on page 494

"CM 8.x Create Document system metadata properties" on page 248

"Task status system metadata properties" on page 269

CM 8.x Duplicate Detection:

This task uses the hash keys calculated by the *EC Extract Metadata* task to determine if a document is a duplicate, that is, if it has already been stored in the repository.

Task summary

Table 125. CM 8.x Duplicate Detection task summary

Characteristic	Value
Task name	CM 8.x Duplicate Detection
Main purpose	Determines if the document has already been stored in the repository
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM Content Manager Connector
When needed?	Required for deduplication in archiving task routes
Placement in task route	Can appear only after these tasks: • EC Extract Metadata • CM 8.x Configure Item Types
Produces which metadata?	CM 8.x Duplicate, Task Status
Configuration options	"Connection" on page 510 "Configure Shortcut Link" on page 510 "Configure Hash Key" on page 510

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Configure Shortcut Link

In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the repository. Based on this definition, Content Collector can generate a shortcut URL for each document that the task is applied to. With this URL, users can access archived documents using the IBM Content Collector Web Application.

Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

```
https://host_name:port_number/AFUWeb/RD.do?
r=%PID_ENCRYPTED&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%
```

To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

```
https://host_name:port_number/AFUWeb/SRD.do?
r=%PID_ENCRYPTED%&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%&am=%CHALLENGE_MODE%
```

In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.

- Replace *host_name* and *port_number* with the name and port number of the Content Manager application server.
- Do not alter any of the tokens *%token_name%* and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
- If you have customized a user client and require the URL to contain more parameters, you can use the following tokens in the URL:
  - %PID_ENCRYPTED%: the encrypted persistent identifier of the item in Content Manager
  - %URL_CHECKSUM%: the checksum of the encrypted values
  - %REPOSITORY_ID_ENCRYPTED%: the encrypted unique identifier of the Content Manager repository
  - %CHALLENGE_MODE%: the access mode for the repository when secure links are used
- The ENCRYPTED tokens are encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service. This means that you cannot use %PID_ENCRYPTED%, %ITEMTYPE_ENCRYPTED%, or %URL_CHECKSUM% with applications that do not use the IBM Content Collector Web Application service.

Configure Hash Key

Define the **Hash key source**:

- From the **Metadata source** list, select **Email deduplication**. This selection ensures that the hash keys to be compared with those stored in the repository are taken from the previous EC Extract Metadata task in the task route. It is the only choice that works in this context.

- From the **Metadata property** list, select **Document ID**. This is the name of the property in which the EC Extract Metadata task stores the hash keys of incoming email.

Define the **Hash key attributes in CM**:

- From the **Filter attributes by item type** list, select the target email repository (item type). This must be one of the item types configured in the previous CM 8.x Configure Item Types task.
- From the **Hash key attribute** list, select the repository attribute in which the hash keys are stored. IBM Content Collector compares the hash keys in this attribute with those that were calculated for the incoming email documents. You specified this attribute as the repository location for hash keys in the CM 8.x Configure Item Types task.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

Related reference

"CM 8.x Duplicate system metadata properties" on page 248

"Task status system metadata properties" on page 269

CM 8.x Store Version Series:

This task creates a version series in Content Manager for each Microsoft SharePoint document that is being processed. You can also select to use folders in Content Manager.

Restriction: Content Manager does not recognize the SharePoint concept of minor versions, so each collected document version receives a version ordinal, the first collected version being 1, the second 2, and so on. If you choose to collect fewer than all versions, your SharePoint and Content Manager version numbers can become out of sync.

Document security is based on the configured item type, not on SharePoint document permissions.

Task summary

Table 126. CM 8.x Store Version Series task summary

Characteristic	Value
Task name	CM 8.x Store Version Series
Main purpose	Creates a version series for a document in the Content Manager repository
Usable with which source connectors?	SharePoint Connector
Usable with which target connectors?	IBM Content Manager Connector
When needed?	When you want to add multiple versions of SharePoint documents to Content Manager, or re-collect processed SharePoint documents.
Placement in task route	Can appear only after these tasks: • EC Extract Metadata • CM 8.x Configure Item Types
Produces which metadata?	CM 8.x Create Document, Task Status

Table 126. CM 8.x Store Version Series task summary (continued)

Characteristic	Value
Configuration options	"Connection" "Checkin Options" "Create Folder " on page 513

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Checkin Options

With these options, you determine the way a document is stored in Content Manager.

- An access control list (ACL) is used to restrict access to content to a specific list of users. It must be created on the Content Manager server. To use a different ACL for the document than the default ACL for the item type, select one of those listed under **Access Control List**. The list contains all access control lists that are available on the Content Manager server.
- In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the repository. Based on this definition, Content Collector can generate a shortcut URL for each document that the task is applied to. With this URL, users can access archived documents using the IBM Content Collector Web Application.

  Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
  Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

  ```
  https://host_name:port_number/AFUWeb/RD.do?
  r=%PID_ENCRYPTED&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%
  ```

  To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

  ```
  https://host_name:port_number/AFUWeb/SRD.do?
  r=%PID_ENCRYPTED%&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%&am=%CHALLENGE_MODE%
  ```

  In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.
  - Replace *host_name* and *port_number* with the name and port number of the Content Manager application server.
  - Do not alter any of the tokens *%token_name%* and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
  - If you have customized a user client and require the URL to contain more parameters, you can use the following tokens in the URL:
    - %PID_ENCRYPTED%: the encrypted persistent identifier of the item in Content Manager
    - %URL_CHECKSUM%: the checksum of the encrypted values
    - %REPOSITORY_ID_ENCRYPTED%: the encrypted unique identifier of the Content Manager repository

- %CHALLENGE_MODE%: the access mode for the repository when secure links are used

– The ENCRYPTED tokens are encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service. This means that you cannot use %PID_ENCRYPTED%, %ITEMTYPE_ENCRYPTED%, or %URL_CHECKSUM% with applications that do not use the IBM Content Collector Web Application service.

- Review the default MIMEtype that will be applied to documents checked into the repository by clicking **MIME type configuration**.

  The Configuration Manager contacts the Content Manager server to obtain the default MIME types. If the Content Manager 8.x connection is working properly, you see a number of predefined rules in the MIME type configuration window. Briefly summarized, these rules work as in the following example: If the extension of an email attachment or file is bmp, assign it the MIME type image/bmp in the Content Manager repository. Typically, default MIME types are defined for the most common attachment types and file types. However, it may be that you want to create a rule for an additional MIME type or change one of the default rules.

  Note: If you add a new MIME type, you must add it to the Content Manager server first, because MIME types that are unknown to Content Manager result in errors.

  Metadata type
  The object that contains the property which you want the rule to evaluate. For example, select **File** to read the property of a file.

  Metadata property
  The property to be evaluated by the rule. For example, if the **Metadata type** is **File**, and you want to assign a MIME type to files of a certain type, select **File Extension**.

  Operator
  Select an operator. See the related topic for a description of all operators.

  Specify the values to compare the selected property with. Possible choices are:

  Literal A constant value that you type in a text field.

  Metadata
  The values of another property.

  Regex search
  The results of a regular expression search.

  List A list of values. This option is available only if you have previously configured lists in the **Metadata and Lists** section of this application.

  Specify the MIME type to be used. You can choose between the following methods to do so:

  – **Literal** assigns the MIME type that you type in the text field
  – **Metadata property** reads the MIME type from a property (the property has only one value and this is the MIME type)

Create Folder

To work with Content Manager folders, select **Create folder**. In this case, further options become available:

Folder errors prevent document creation

If you select this option, the document is not archived in case of a failure to create a folder or a failure to file a document in a folder. If you do not select this option, the document is archived even if a folder error occurs. In this case, a warning message is written to the log file.

Folder name delimiter

Specify the character that separates the elements of the folder path.

Metadata Properties

Specify a set of string and string array properties that provides folder paths.

The strings are used as follows:

- Empty strings are ignored
- Non-empty strings are split on the delimiter character.
- Any empty leading or trailing elements of a string are ignored

For example, if the delimiter is \ and one of the string values is `\\mymachine\myshare\folder1\folder2`, the result is:

mymachine

myshare

folder1

folder2

This represents the set of names that will be used to build a folder path. If after removing any empty elements the result is an empty list, the result is ignored.

Related reference

"Operators for conditional values" on page 494

"CM 8.x Create Document system metadata properties" on page 248

"Task status system metadata properties" on page 269

CM 8.x Update Document:

The *CM 8.x Update Document* task creates a record of all the duplicates that could be identified by the *CM 8.x Duplicate Detection* task. The record ensures that users who wanted to archive an email can restore a copy, although the email has only once been stored in the repository. This is necessary if IBM Content Collector is configured to stub email after archiving. The record also ensures that access to an email in the repository remains restricted because only the users who were able to archive it can restore it.

Task summary

Table 127. CM 8.x Update Document task summary

Characteristic	Value
Task name	CM 8.x Update Document
Main purpose	Creates a record of all the duplicates that could be identified
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM Content Manager Connector

Table 127. CM 8.x Update Document task summary (continued)

Characteristic	Value
When needed?	Required in archiving task routes to enable single-instance storing
Placement in task route	The following tasks must have been configured and placed before this task in the task route: • EC Extract Metadata • CM 8.x Configure Item Types • CM 8.x Duplicate Detection
Produces which metadata?	CM 8.x Update, Task Status
Configuration options	"Connection" "Shortcut Link"

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Shortcut Link

In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the repository. Based on this definition, Content Collector can generate a shortcut URL for each document that the task is applied to. With this URL, users can access archived documents using the IBM Content Collector Web Application.

Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

```
https://host_name:port_number/AFUWeb/RD.do?
r=%PID_ENCRYPTED&sum=%URL_CHECKSUM%&repositoryID=%REPOSITORY_ID_ENCRYPTED%
```

Related reference

"CM 8.x Update system metadata properties" on page 249

"Task status system metadata properties" on page 269

EC Copy to Mailbox:

This task copies documents from a personal folder (PST file) to the Exchange mailbox of the PST file user after the archiving of these documents. Thus, information that might exist outside of the email system is brought into the system. Such information consists of PST file documents that have never been sent or received as an email, but has none the less been created and stored in a PST file. The EC Copy to Mailbox task subjects such documents to the control of the email system.

Task summary

Table 128. EC Copy to Mailbox task summary

Characteristic	Value
Task name	EC Copy to Mailbox
Main purpose	Copies documents from PST files to mailboxes
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required for processing PST files
Placement in task route	Can appear only after these tasks: • EC Prepare Email for Archiving • EC Finalize Email for Compliance • EC Create Email Stub task (to stub the documents in the PST files, resulting in stubs being copied to the mailbox) and before theEC Create Email Stub task (to delete the stubs from the PST files)
Produces which metadata?	Task Status
Configuration options	None

Note: You need a collector that looks for documents in PST files (PST files as collection source).

The PST file user is the user whose ID (account ID or email address) is associated with a PST file. This is not necessarily the owner because you can create PST files on behalf of other users.

You can copy entire documents to user mailboxes or stub documents if stubbing has been applied after archiving.

Related reference

"Microsoft Exchange collection sources for automatic archiving" on page 464

"Task status system metadata properties" on page 269

EC Create Email Stub:

This task removes content from the original email in client mailboxes or from Notes documents after these have been archived by IBM Content Collector and inserts links by which users can view the archived content. Which links are inserted depends on the specifications for the stubbing options.

Task summary

Table 129. EC Create Email Stub task summary

Characteristic	Value
Task name	EC Create Email Stub

Table 129. EC Create Email Stub task summary (continued)

Characteristic	Value
Main purpose	Removes content from the original email in client mailboxes or from Notes documents after these have been archived by IBM Content Collector and inserts links by which users can view the archived content. Links are inserted according to the specifications for the stubbing options.
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in archiving and stubbing task routes for email and Notes application documents
Placement in task route	Can appear only after an EC Prepare Email for Stubbing task or an EC Process Email Stubbing Life Cycle collector
Produces which metadata?	Task Status
Configuration options	"Stubbing Options" "Format Definition" on page 519 "Stubbing Definitions" on page 519 "Excluded Message Types" on page 520

Stubbing Options

Stubbing can only be done after the successful archiving of documents. Therefore, you need to place the stubbing task after the EC Prepare Email for Stubbing task in the task route or after an EC Process Email Stubbing Life Cycle collector. There are different methods to perform stubbing: In addition to using a simple stubbing task, you can use a stubbing life cycle. The latter requires more system resources because it uses its own collector. A stubbing life cycle is thus only recommended if you want to reduce the content of documents step-by-step, for example, by first removing attachments, then reducing or removing the body texts, and so on. For a onetime removal of content, a simple stubbing task is sufficient. Note, however, that both methods require a stubbing task object in the task route.

Important: The address information for the Content Collector server becomes an unchanging part of the link in the stub document. To avoid problems with the generated links, specify the fully qualified server name (for example, ICCServer.example.com) for the Content Collector server in the Web Application under **General Settings → Web Application**. Make sure that this host name is resolved properly by the DNS. If the host name of the server that hosts the Web Application changes, all stub links become useless.

Content Collector does not treat MSG attachments as attachments but as part of the Exchange message. An embedded MSG file is not stored separately in the archive and no separate link is created for the embedded MSG. It remains part of the archived Exchange message. The embedded MSG attachment remains in the message stub if you select the stubbing type **Remove attachments**. It is only remove from the stub when **Remove attachments and cut body** is selected.

As a Lotus Notes user, select the type of documents to which the stubbing method is applied:

Email Email documents are stubbed according to the selected stubbing method.

Application
Documents from Lotus Notes applications are stubbed according to the selected stubbing method. You must also specify the rich text field into which the stubbing information is to be inserted.

Note: When documents from Lotus Notes applications are archived, no preview link for the document body is inserted. Therefore, it is recommended that you select the **Remove attachments** stubbing option. With this stubbing option, the full content of the document's rich text fields is kept.

Select the stubbing method that you want to use:

- **Stub immediately** creates a simple stubbing task without a stubbing life cycle. Select also what to remove from the original documents:

  Remove nothing and add text
  This does not remove anything from the original client documents, but adds a text message. When users open a stub document, the text message informs them about the fact that the content has been archived by IBM Content Collector. You can define the text message to be inserted; this is done in the **General information** field.

  Remove attachments
  This option removes the attachments of the original documents immediately after archiving. At the same time, links to the archived attachments and a text message that informs the user that the attachments have been archived are inserted. You can define the text message in the **Text for stubbed attachments** field. Lotus Notes users can specify the text that precedes the links in the **Attachment link text** field.

  With Lotus Domino, attachments can be inserted anywhere in the document body. These attachments are replaced with attachment links and, additionally, an attachment link for each removed attachment is included in the *attachment summary* at the end of the stub document. To have Content Collector insert links in the attachment summary only, select the option **Include attachment link in attachment summary only**.

  Remove attachments and cut body
  In addition to removing the attachments (see entry under **Remove attachments**), this option reduces the length of the body text to the specified number of characters. You set the character limit in the field labeled **Number of characters the body is reduced to**.

  The actual number of letters that remain in an email body is lower than this number if multiple Unicode code points are needed to represent a letter.

  Remove attachments and body
  This is similar to the option **Remove attachments and cut body**. The difference is that the body text is removed entirely rather than reduced.

  Delete entire email
  This option deletes the original documents after archiving. This frees up the maximum amount of space in the users' mailboxes and reduces the

number of entries in views and folders of the email client. However, as there will not be any stub documents, users can only find archived email by using the search function.

- **Delay stubbing and use EC Process Email Stubbing Life Cycle collector** uses a stubbing life cycle. Select this option only in task routes that are fed from the EC Process Email Stubbing Life Cycle collector.

Depending on the stubbing method that you selected, the following additional options are available:

- Microsoft Exchange only: If you selected any of the options for removing the attachments or if you selected to delay stubbing, you can also select **Preserve paper clip icon in message view when removing attachments**. For further information about this option, click the appropriate link at the bottom of this topic.
- If you selected any of the options for removing part or all of the body text or if you selected to delay stubbing, you can also select the option **Do not add body link to stub document**. With this option, Content Collector does not insert a link to archived body text in the stub document but only attachment links.
- Lotus Domino only: If you selected any of the options for removing the attachments or if you selected to delay stubbing, you can also select the option **Include attachment link in attachment summary only** to prevent Content Collector from inserting additional attachment links in the body of the stub document.

Format Definition

Specify the format of date and time entries:

Date format

For example, you could type `MM/dd/yyyy` to specify a date format starting with a two-digit number for the month, followed by a two-digit number for the day, and ending with a four-digit number for the year.

You can insert a placeholder `%d%` for the date in the fields mentioned in "Stubbing Definitions." These fields contain the texts to be inserted in the stub documents, indicating that parts of the content have been removed. If the placeholder `%d%` is inserted in one of these fields, a date stamp will be inserted in the stub messages that conforms to the `MM/dd/yyyy` pattern. For a list of possible formats to insert in the **Date format** field, click the link at the bottom of this topic.

Time format

For example, you could type `HH:mm:ss` to specify a date format starting with a two-digit number for the hours, followed by a two-digit number for the minutes, and ending with a two-digit number for the seconds.

You can insert a placeholder `%t%` for the time in the fields mentioned in "Stubbing Definitions." The effect is similar to the one described for the date format.

Stubbing Definitions

Type the texts to be inserted in places where content has been removed. Use plain text only. Markup like, for example, HTML is not supported. The availability of entry fields depends on the stubbing options that you selected:

General information

Text that appears in original documents, no matter which option is selected. Example:

```
The content of this email has been archived by %u% on %d% at %t%.
```

where

%u%
Stands for the user who runs the Email Connector.

%d%
Stands for the archiving date, which is inserted in the format specified in the **Date format** field.

%t%
Stands for the archiving time, which is inserted in the format specified in the **Time format** field.

Text for stubbed attachments

Text that appears at the end of the stub document after attachments have been removed. Example:

```
Attachments removed.
```

Attachment link text

The text that precedes the link to an archived attachment in the stub document. Example:

```
Click here to view the archived attachment.
```

This option is available for Lotus Notes users only.

Body link text

Text that appears on or as the link to an archived body text in the stub document. Example:

```
Click here to view archived email content.
```

Note: For Notes application documents, no body link text is inserted.

Number of characters the body is reduced to

Specify a number of characters. For example, if you type in 100, only the first 100 characters of the body text in a document will remain after the cutting.

Note: The actual number of letters that remain in an email body is lower than this number if multiple Unicode code points are needed to represent a letter.

Name of attachment used to preserve the paper clip icon

Specify the text that is used as file name for the dummy attachment in a stub document. The text is limited to 28 characters.

This option is available for Microsoft Exchange users only.

Excluded Message Types

This configuration option is available for Microsoft Exchange users only.

Documents of the Exchange message that are listed under **Excluded message types** classes are excluded from stubbing. You can add further message types to the list. In the Add Message Type window, type the exact name of an Exchange message class, for example, `IPM.Report` in the **Message type** field.

To exclude just the specified class, but not its children, leave **Exclude base message type** selected. To exclude just the children, select **Exclude all message types derived from the specified base type**.

For example, to exclude the parent message class for delivery reports without its children (IPM.REPORT), type `IPM.Report` in the **Message type** field and leave **Exclude base message type** selected. To exclude the children (IPM.REPORT.*), type `IPM.Report` and select **Exclude all message types derived from the specified base type**.

Related concepts

"Paper clip icon for removed attachments"

Related tasks

"Collecting documents for life cycle processing" on page 476

Related reference

"Task status system metadata properties" on page 269

Related information

Possible date and time formats

Paper clip icon for removed attachments:

Starting with IBM Content Collector 2.1.0.2, the way of handling attachments for which links were added has changed. Previous versions of IBM Content Collector only removed the binary content from the attachment and left the attachment control information untouched. If the body was removed, all attachments were completely removed from the message and the paper clip icon was lost. Now, IBM Content Collector completely removes the attachments from the attachment table and adds a .gif file as a dummy attachment to preserve the paper clip icon. This file shows an archive icon with a paper clip to represent the fact that the message that was archived by IBM Content Collector contained attachments that are now accessible using links in the message stub. This dummy attachment is preserved even if the body is removed.

This behavior is controlled in the *EC Create Email Stub* task. You select the option **Preserve paper clip icon in message view when removing attachments** to have IBM Content Collector add an attachment placeholder whenever it removes attachments. The dummy attachment is marked as hidden. While Outlook respects this setting and does not show the attachment placeholder, Outlook Web Access always shows attachments whether they are marked as hidden or not. If you do not select this option, IBM Content Collector does not add an attachment placeholder to preserve the paper clip icon.

If you select to preserve the paper clip icon when attachments are removed, you must also specify a name for the dummy attachment. The file name can have up to 28 characters. It cannot contain any of the following characters: `\/:*?<>|"`

When you import a task route that was created with IBM Content Collector 2.1.0.1 or before and the stubbing option includes removing attachments, **Preserve paper clip icon in message view when removing attachments** is selected in the *EC Create Email Stub* task and a default file name is set.

Messages that have already been stubbed will not be changed automatically, only when they are restored and then later restubbed, will restubbing use the new setting.

EC Extract Attachments:

This task allows attachments to be saved as files separate from the document. If you do not add this task to the task route, the attachments will not be saved in the repository as separate objects.

Task summary

Table 130. EC Extract Attachments task summary

Characteristic	Value
Task name	EC Extract Attachments
Main purpose	Extracts attachments from email, Notes application documents, or PST files
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in archiving task routes for the compound data model
Placement in task route	Can appear only after the EC Prepare Email for Archiving task
Produces which metadata?	Attachment Deduplication, Email, File, Task Status
Configuration options	None

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

Related reference

"Attachment Deduplication system metadata properties" on page 246

"Email system metadata properties" on page 250

"File system metadata properties" on page 256

"Task status system metadata properties" on page 269

EC Extract Metadata:

This task extracts metadata from fields in an document to store this metadata in corresponding fields in the repository.

The metadata fields in a repository provide search information for user queries. For example, the repository field that corresponds to the From or Sender field of an email allows users to search for email that was sent by a specific person. Certain email fields are selected by default, such as the Subject, To, or Sender (From) fields. You can select other fields to extract metadata from if you think that these fields add valuable search information to your repository.

For Exchange messages, this task also extracts managed folder information if this information is available.

Encrypted Lotus Notes documents are decrypted before any metadata is extracted. This is required to allow for search later. All encrypted Lotus Notes documents are archived in decrypted format, and they will be restored as decrypted copy. Because encrypted Microsoft Exchange messages consist of a container and an encrypted

attachment, these messages can be archived in encrypted format. So the EC Extract Metadata does not decrypt Microsoft Exchange messages.

Task summary

Table 131. EC Extract Metadata task summary

Characteristic	Value
Task name	EC Extract Metadata
Main purpose	Extracts metadata from fields in a document to store this metadata in corresponding fields in the repository.
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in email or application processing task routes
Placement in task route	Must be the first task in a task route
Produces which metadata?	Email, Email Deduplication, Task Status
Configuration options	"Associate Metadata" "Additional Forms Definition (Notes only)" on page 524

Associate Metadata

To extract metadata from additional document fields, select the appropriate set of fields from the **User defined metadata** list. Such sets of fields must have been defined earlier, in the **User Defined Metadata** section of the **Metadata and Lists** configuration.

For Microsoft Exchange, you can also access MAPI properties and named properties. Edit the custom field to which you want to map a MAPI property or a named property. Specify if you want to have the property extracted from the message so that you can use the metadata in rules or when you assign property values and select the property type:

MAPI property
: Microsoft Messaging API (MAPI) properties are a standard set of predefined properties from Microsoft. Select a property name to refer to a specific MAPI property. For the selected property, the hexadecimal property identifier and the MAPI property type are displayed.

Named property
: Named properties are properties that were defined by a user or an application. They usually serve a purpose for which a MAPI property cannot be used. Named properties are referenced by a property ID, which is a hexadecimal value, or a name, which is a string value. Select **MNID_ID** for reference by property ID or **MNID_STRING** for reference by name and specify an appropriate value in the **ID** field. Also select the property set that contains the selected named property.

Additional Forms Definition (Notes only)

Lotus users can also specify additional forms or properties to enlarge the basis for the calculation of the deduplication hash key:

- To add a form for the calculation of the deduplication hash key, type its name in the **Form name** field and click **Add**. For example, to add the form for calendar entries in Lotus Notes 8, type `Appointment`. The form name appears in the list under the **Form name** field.
- To add a property for the calculation of the deduplication hash key, type its name enclosed in backslash (\) characters in the **Form name** field and click **Add**. The backslash (\) characters are required to denote that the entry refers to a property, not to a form. For example, to add the property BCC, which is available only in email of BCC recipients and in the sent copy, type `\BlindCopyTo\`. The entry appears in the list under the **Form name** field.

Related tasks

"Adding and editing user-defined metadata" on page 240

Related reference

"FSC Associate Metadata" on page 531

"Email system metadata properties" on page 250

"Email Deduplication system metadata properties" on page 255

"Task status system metadata properties" on page 269

EC Finalize Email for Compliance:

This task converts a temporary file created by the EC Prepare Email for Archiving task to a bundled resource item (BRI) file. A file in this format can be stored in the repository as is, that is, no further conversion is needed. This task can be used only with the bundled email data model.

Task summary

Table 132. EC Finalize Email for Compliance task summary

Characteristic	Value
Task name	EC Finalize Email for Compliance
Main purpose	Converts a temporary file created by the EC Prepare Email for Archiving task to a bundled resource item (BRI) file
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in email archiving task routes if the bundled email data model is used
Placement in task route	Can appear only after the EC Prepare Email for Archiving task
Produces which metadata?	Archiving Format, File, Task Status
Configuration options	None

Related reference

"Archiving format system metadata properties" on page 246

"File system metadata properties" on page 256

"Task status system metadata properties" on page 269

EC Prepare Email for Archiving:

This task creates a temporary file for each email that has been captured by a collector. The temporary file is the basis for further transformations.

Task summary

Table 133. EC Prepare Email for Archiving task summary

Characteristic	Value
Task name	EC Prepare Email for Archiving
Main purpose	Creates a temporary file for each email that has been captured by a collector
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in email archiving task routes
Placement in task route	Can appear only before an EC Finalize Email for Compliance task
Produces which metadata?	Archiving Format, File, Task Status
Configuration options	"Archiving Format"

Archiving Format

If IBM Content Collector connects to a Lotus Notes/Domino mail system, you can choose between the following options:

Notes native format
: To store the temporary file in the IBM CommonStore for Lotus Domino (CSN) format. This format ensures that the original appearance is preserved when the email is restored to a Lotus Notes database.

Plain format
: To extract the plain text from the email body and store that as a temporary file. No attachment information will be available in temporary file and, therefore, in the archived document. If you select this format, the email cannot be stubbed or restored. It is only recommended for FileNet P8 business processes.

Note: If IBM Content Collector connects to an Exchange email system, the file format of the temporary file is always MSG.

Select **Save native message file without attachments** to enable IBM Content Collector to archive attachments separately (compound data model).

Related reference

"Archiving format system metadata properties" on page 246

"File system metadata properties" on page 256

"Task status system metadata properties" on page 269

EC Prepare Email for Stubbing:

This task marks the document as archived and writes repository information to the document to be stubbed so that later, the document can be linked with the archived content. Also, to satisfy BPM scenarios, it marks the document as processed, so that it is processed only once.

Task summary

Table 134. EC Prepare Email for Stubbing task summary

Characteristic	Value
Task name	EC Prepare Email for Stubbing
Main purpose	• Writes repository information to the document to be stubbed and marks the document as archived • Marks the document as processed, so that it is processed only once • Checks if the document and all attachments have been archived
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in email and application archiving task routes
Placement in task route	Can appear only before an EC Create Email Stubtask
Produces which metadata?	Task Status
Configuration options	"Processing Options"

After this task has been completed, an icon appears next to documents in Lotus Notes. This icon marks documents as *archived*. In Outlook, such an icon is not displayed.

This task also performs a consistency check to ensure the document and all attachments have been archived. If one part of the document has not been archived, an exception occurs. Error messages are written to the log files and the document is not marked as archived. Therefore, it is not possible to filter attachments during archiving, for example, to exclude .mp3 files from archiving.

Processing Options

Select one of these options:

- **Mark as processed** marks a document as *processed* by a collector. This option can be used in task routes that process email documents (for example, move them to specific folders) but do not archive the documents.

- **Mark as compliance archived** marks a document as *archived*. You must also specify the archiving format:
  - With the **Bundled email data model**, the complete document content, that is, the text, the attachments, and all other document properties, is archived in one single file.
  - With the **Compound email data model**, the document content is not stored in one single file, but decomposed into separate attachment files and the rest into a single file. For example, if an email document contained three attachments, the message would logically be divided into four parts.

  If you select this option, you can also select **Write base retention to message**. In this case, the retention date that is set by the Calculate Expiration Date task is written to the message.

Related reference

"Task status system metadata properties" on page 269

Extract Text:

This task extracts the text of email attachments and prepares it for full-text indexing.

If text extraction fails, the Extract Text task writes an error notification to the text-search indexing document. Refer to the related topic for a list of possible error strings.

Task summary

Table 135. Extract Text task summary

Characteristic	Value
Task name	Extract Text
Main purpose	Extracts the text of files or email attachments and prepares it for full-text indexing
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Required in email archiving task routes when processing attachments that must be full-text indexed
Placement in task route	Can appear only after the EC Extract Attachments task in email archiving task routes
Produces which metadata?	Task Status, Text Extraction
Configuration options	"File Extension Filter"

File Extension Filter

Define a filter for file extensions. When you define an exclude filter, the Extract Text task will skip files with the listed extensions for text indexing. If the list is empty, the task will render all files that are passed in. When you define an include filter, the Extract Text task will process files with the listed extensions for text indexing. If the list is empty, the task will render none of the files that are passed in.

For all files that are skipped, the task writes the string "IcmFceWarning:IcmConfigFilteringFile" to the icc_attachment and icc_attachment_text fields of the text-search indexing document, so that you can search for all documents that contain attachments that were not indexed.

To restore the default list of extensions, click **Load Default Extension List**. To empty the list, click **Clear Extension List**.

Related reference

"Searching for error strings" on page 662

"Text Extraction system metadata properties" on page 269

"Task status system metadata properties" on page 269

FileNet Image Services Create Document:

To save a document in an IBM FileNet Image Services repository, you need to specify where to save it and how to index it. You specify where to save a document by selecting a previously configured connection to the IS library. You specify how to index it by choosing a *class* for the item and by setting the values to be assigned to each property of that class.

Task summary

Table 136. FileNet Image Services Create Document task summary

Characteristic	Value
Task name	FileNet Image Services Create Document
Main purpose	Archive a document in a FileNet IS repository
Usable with which source connectors?	File System Source Connector
Usable with which target connectors?	IBM FileNet Image Services Connector
When needed?	Required for task routes that archive document in FileNet IS
Placement in task route	Must appear before the following tasks: • FileNet Image Services File Document In Folder • FileNet Image Services Modify Permissions
Produces which metadata?	FileNet Image Services Create Document, Task Status
Configuration options	"Connection" "Checkin Options" on page 529 "Property Mappings" on page 529

Prerequisites: Create any classes you will need in the FileNet IS repository.

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

Checkin Options

You can enter a URL to be used when adding a shortcut to a document in the repository.

Property Mappings

The **Class** list box displays all document classes in the selected repository. Select the document class you want to use when capturing the item.

Note: If document classes are added while the Configuration Manager is running, you will need to restart the application to see the new classes in the list.

Add or edit properties of the selected document class.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Assigning property values" on page 490

"List Lookup" on page 496

Related reference

"FileNet Image Services Create Document system metadata properties" on page 257

"Task status system metadata properties" on page 269

FileNet Image Services File Document In Folder:

You must add a document to a folder in the repository to be able to browse for the document at a later point. If you do not add the document to the folder, you can access it only by searching for it.

Task summary

Table 137. FileNet Image Services File Document In Folder task summary

Characteristic	Value
Task name	FileNet Image Services File Document In Folder
Main purpose	Filing documents into a FileNet IS repository folder
Usable with which source connectors?	File System Source Connector
Usable with which target connectors?	IBM FileNet Image Services Connector
When needed?	Required in file archiving task routes
Placement in task route	Can appear only after the FileNet Image Services Create Document task
Produces which metadata?	FileNet Image Services File Document In Folder, Task Status
Configuration options	"Connection" "File in Folder Options" on page 530

Connection

Select the appropriate connection to enable access to the target repository by IBM Content Collector.

File in Folder Options

In the **Folder Path** text box, provide the complete path to the folder in IS. Use one of the following methods:

- Browse to select the folder from the folder tree.
- Type in a literal value of a folder in which you want to file the document.
- Use a metadata expression for the folder path.
- Use a regular expression for the folder path.
- Use a calculated value for the folder path.
- Use list lookup to determine the folder path.

Check **Create folder if it does not exist** to have the folder created automatically if it does not exist. If you do not check this option and the specified folder does not exist, an error occurs.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Assigning property values" on page 490

"List Lookup" on page 496

Related reference

"FileNet Image Service File Document in Folder system metadata properties" on page 258

"Task status system metadata properties" on page 269

FileNet Image Services Modify Permissions:

This task modifies user permissions for files being saved to a FileNet IS repository.

Task summary

Table 138. FileNet Image Services Modify Permissions task summary

Characteristic	Value
Task name	FileNet Image Services Modify Permissions
Main purpose	Modifies user permissions for files in a FileNet ISrepository
Usable with which source connectors?	File System Source Connector
Usable with which target connectors?	IBM FileNet Image Services Connector
When needed?	Required in file archiving task routes
Placement in task route	Must appear after the FileNet Image Services Create Document task
Produces which metadata?	FileNet Image Services Modify Permissions, Task Status
Configuration options	"Connection" "User/Group allowed to" on page 531

Connection

Select theIS repository for which you want to set security permissions.

User/Group allowed to

Set configuration options for each field:

Read Select users or groups to whom *Read* permissions should be assigned.

Write Select users or groups to whom *Write* permissions should be assigned.

Append/Execute
Select users or groups to whom *Append/Execute* permissions should be assigned.

Related reference

"FileNet Image Services Modify Permissions system metadata properties" on page 258

"Task status system metadata properties" on page 269

FSC Associate Metadata:

The *FSC Associate Metadata* task enables you to use control files (in XML or CSV format) to add custom metadata to archive documents.

Sometimes, not all the information that is required for processing a document is available in the document itself. In these cases, you can use metadata files to provide additional information. For example, a company might want to archive PDF copies of order confirmations. These files contain some information, but not necessarily all required information, for example they might be missing a customer name or order type. This information can be provided in a separate file, and the contents of this metadata file can be used as metadata for the document.

You can use metadata files in two different ways:

Collect documents
To collect the documents that you want to archive and locate one metadata file for each of the documents:

1. Create a file system metadata type that defines the information in your metadata files.
2. Configure the file system collector to collect the documents that you want to archive.
3. In the FSC Associate Metadata task, select input file type **Document**.
4. As metadata source type, select the file system metadata type that you created for this purpose.
5. In the **Metadata File Name** section, define how to derive the name of the metadata file from the name of the document that has been collected.
6. The information that is contained in the metadata file is associated with the document, so that you can access it during further processing as if it was part of the original document.

Collect metadata files
To collect metadata files that contain additional information and locate the documents that you want to archive according to this metadata file:

1. Create a file system metadata type that defines the information in your metadata files.
2. Configure the file system collector to collect the metadata files that contain the information that you want to add to the documents.

3. In the FSC Associate Metadata task, select input file type **Metadata File**.
4. As metadata source type, select the file system metadata type that you created for this purpose.
5. In the **Document Name** section, define how to derive the name of the documents to be processed from the metadata file. You can base the definition on the metadata file name (to process one document) or on values in the metadata file (to process one or more documents).
6. During further processing, configure rules to distinguish between the metadata files that were collected by the file system collector and the documents that were located by the FSC Associate Metadata. You can use the FSC Metadata property Is Metadata to distinguish between metadata files and document files.

Task summary

Table 139. FSC Associate Metadata task summary

Characteristic	Value
Task name	FSC Associate Metadata
Main purpose	Matches metadata files to content files and enables you to map custom metadata
Usable with which source connectors?	File System Source Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in file processing task routes if you want to access fields defined in the metadata file associated with a file, and use the fields in property mappings
Placement in task route	Can appear only before repository tasks
Produces which metadata?	Task Status
Configuration options	"Monitor Options" "Document Name options" on page 533 "Metadata File Name options" on page 533

Monitor Options

The **Input File Type** determines the direction of the search:

- Selecting **Metadata File** tells the application that you are collecting metadata files and that the task must search for document files that match them. You specify your search criteria in the **Document Name** region.
- Selecting **Document** tells the application that you are collecting content files and that the task must search for metadata files that match them. You specify your search criteria in the **Metadata File Name** region.

The **Metadata source type** is the set of custom metadata that you defined from the metadata items you defined under **Metadata and Lists**. If you are collecting files that contain financial data, for instance, then you would select a source type named *Financial*, which might include fields for account numbers and balances.

The **Maximum wait time** is the number of seconds the system is to wait for a document file associated with a metadata file. Set a value of at least one second, because associated document and metadata files might not arrive in source folders at the same time. The log records an error for each file that does not arrive within the allotted time.

Document Name options

These options are available when you select the **Metadata File** input file type and tell the application how to locate content files associated with a metadata file.

- **Metadata name + file extension:** The application determines the document file name to search for by replacing the file extension of the metadata file with the file extension entered. For example, if the metadata file name is `document.xml`and you enter the file extension .pdf, the application searches for the metadata file named `document.pdf`.

  Important: Selecting this option restricts processing to only those documents with the file extension you supply, so do not select this option if you want to process files of all types.
- **Regular expression applied to document file name:** You can use a regular expression to convert the metadata file name to the name of a document file.
- **Values in the metadata file:** The metadata file contains the names of one or more content files.
  - Select the property to use in the file name.
  - *Content Elements configuration only:* Select **Group documents by metadata file** to use a .csv or .xml file that contains file properties (field or column names) on the first line of the file, and file paths and names on subsequent lines. This will set the metadata property Is File Grouped to true. This property is used when invoking any task that takes a group of items and results in the creation of one target repository object with multiple content elements for multiple files. If you do not select this option, documents are grouped by line. That is, each line in the file represents a repository object with one or more content elements.
  - *Content Elements configuration only:* Select **Documents are contentless** to use only the properties of the files listed in the .csv or .xml file, not the files themselves.

Metadata File Name options

These options are available when you select the **Document** file type and tell the application to determine how to locate the metadata file associated with a content file.

- **Metadata name + file extension:** determines the metadata file name for which to search by replacing the file extension of the document with the file extension entered. For example, if the document name is “document.pdf” and the file extension “xml' is entered, the metadata file name should be “document.xml”.
- **Regular expression applied to Document (file) Name:** You can use a regular expression to convert the document file name to the name of a metadata file.

Related concepts

"Regular expressions" on page 421

Related tasks

"Adding and editing file system metadata" on page 241

"Defining metadata to be used to process files for archiving" on page 729

Related reference

"FSC Metadata system metadata properties" on page 259

"Task status system metadata properties" on page 269

FSC Post Processing:

This task defines what happens to a file on the file system after it has been processed. These settings apply only to the document on the file system, not to the document that is added to the repository.

Task summary

Table 140. FSC Post Processing task summary

Characteristic	Value
Task name	FSC Post Processing
Main purpose	Specifies what remains in the file system after archiving
Usable with which source connectors?	File System Source Connector
Usable with which target connectors?	IBM FileNet P8, IBM Content Manager, File System Repository Connector
When needed?	Optional in file processing task routes; required if you want to preserve the original icon for the shortcut file
Placement in task route	Can appear only at or near the end of a task route after one of these tasks: • FSR Create Document • CM 8.x Create Document • P8 4.x Create Document
Produces which metadata?	CM 8.x Update, Task Status
Configuration options	"Post Processing Options"

Post Processing Options

You can select one of these options:

- To delete the file from the file system after capture, select **Delete file** to delete the file permanently. You can also replace the deleted file with a shortcut. To assign the recommended shortcut:
  1. Select a **Metadata type** that corresponds to your target repository:

Repository	Metadata type
IBM FileNet P8	P8 4.x Create Document
IBM Content Manager	CM 8.x Create Document
File system repository	FSR Create Document

2. From the **Property** menu, select *Shortcut URL*.

   If icon information is available from the Content Type Information metadata source, the icon information will be used in the generated .url (shortcut) file.

   The .url shortcut file consists just of a file path and an icon index, and not the actual icon itself. If you are viewing the shortcut on a system where the icon source is not available, you will see the default icon for shortcut files.

- To keep the file in the file system after capture, select **Do not delete file**. If you select this option, you can also select one or more of these options:

  Rename file by adding the file extension
  Adds the specified file extension to the file name. For example, if you enter the file extension "done", the file "data.txt" is renamed "data.txt.done". If a file with this name already exists, a number in parentheses is added to the file name to make the file name unique. To prevent multiple processing of the same file, set the file system collector for this task route to filter for and exclude files with this extension.

  Move file to folder
  Moves the file to the specified folder. Selecting this option enables the **Replicate folders** option, which replicates the original file location (the path includes the drive letter or server name plus the folder hierarchy) to the destination folder.

  Mark file as processed
  Adds a property to the file that indicates that the collector has already processed (but not necessarily captured) this file. The file system collector uses this property to omit files from subsequent collection efforts, so if you are collecting continuously or repeatedly, you should select this option.

  When you select this option, the collector configuration determines how the file is marked. If you selected the option **NTFS post processing**, which is valid for NTFS files only, Content Collector stores the marking information in an alternate data stream of the file. If you selected the option **Control folder post processing**, Content Collector creates a control folder in the source folder of the collected file and writes a control file with the marking information to that folder.

  Mark file as captured
  Adds a property to the file that indicates that this file has already been archived. The file system collector uses this property to omit files from subsequent collection efforts.

  When you select this option, the collector configuration determines how the file is marked. If you selected the option **NTFS post processing**, which is valid for NTFS files only, Content Collector stores the marking information in an alternate data stream of the file. If you selected the option **Control folder post processing**, Content Collector creates a control folder in the source folder of the collected file and writes a control file with the marking information to that folder.

  Create shortcut
  Creates a shortcut for the file, using the specified **Metadata type** and **Property** .

  Note that the web browser that retrieves the archived message and attempts starting the associated client does not request any logon authentication. If the security settings of the browser are set to not allow this, the archived content will not be displayed.

Access: If the application can retrieve the access control list (ACL) from the original file, it applies the ACL to the shortcut. Otherwise, it applies the Windows default ACL. In either case, changes you make under **Modify security** override these settings.

Change file security
Enables you to modify the file-security settings. Edit security permissions as indicated in the window that opens when you click **Modify security**.

Related tasks

"Moving documents off the network into IBM FileNet P8" on page 725

"Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records" on page 726

"Defining metadata to be used to process files for archiving" on page 729

Related reference

"Task status system metadata properties" on page 269

FSR Create Document:

To save a document in the file system, you need to specify where to save it and how to index it. You specify where by selecting a folder in which to save. You index the document by choosing a document class for the item and by specifying what values should be assigned to each property of that class. An index in a repository points to a document, to facilitate searching (and finding) that document.

Task summary

Table 141. FSR Create Document task summary

Characteristic	Value
Task name	FSR Create Document
Main purpose	Archive documents in a file system repository
Usable with which source connectors?	File System Source Connector
Usable with which target connectors?	File System Repository Connector
When needed?	Required in file archiving task routes
Placement in task route	Must appear before the postprocessing task
Produces which metadata?	FSR Create Document, Task Status
Configuration options	"Repository" "Captured File Options" on page 537 "Index File Options" on page 537 "File Security" on page 537 "Property Mappings" on page 537

Prerequisites: Create any document classes you will need in the repository.

Repository

Select the connection to be used when capturing.

Captured File Options

For **Destination folder**, specify the location of the folder in which to add captured files.

The option **Mark file read only** restricts the ability to make changes to a document.

The option **Mark file as hidden** hides the document from view

The option **Mirror source file folder path** mirrors the source file location in the repository folder.

Index File Options

For **Destination folder**, specify the location of the folder in which to add index files that contain additional metadata for files.

The option **Mark file read only** restricts the ability to make changes to the index file.

The option **Mark file as hidden** hides the index file from view.

File Security

Select **Change file security** to be able to modify the file-security settings. Edit security permissions as indicated in the window that opens when you click **Modify file security**.

Property Mappings

In the section, the **Class** list box displays all classes previously defined as part of creating a connection to the file system repository. Select the document class you want to use when capturing the file.

Add or edit properties for the selected class.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Assigning property values" on page 490

"Configuring a File System Repository Connector and repositories" on page 202

"List Lookup" on page 496

Related reference

"FSR Create Document system metadata properties" on page 261

"Task status system metadata properties" on page 269

IBM Classification Module:

This task uses IBM InfoSphere Classification Module to generate additional metadata for a document. You must integrate IBM InfoSphere Classification Module before this task shows up in the list of utility connector tasks in the Configuration Manager.

Task summary

Table 142. IBM Classification Module task summary

Characteristic	Value
Task name	IBM Classification Module
Main purpose	Uses IBM InfoSphere Classification Module to generate additional metadata for a document
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Optional in any task route
Placement in task route	Must appear after EC Prepare Email for Archiving in email archiving task routes. EC Prepare Email for Archiving must not remove the attachments though, because they are needed for the classification. You might need another EC Prepare Email for Archiving task after the IBM Classification Module task to archive the document without attachments. No restriction for file system task routes. Must appear after the SP Create File in SharePoint task routes.
Produces which metadata?	IBM Classification Module, Task Status
Configuration options	"Server" "Classification" "Result Set" on page 539 "Map Decision Plan Results to Metadata" on page 539

Server

Specify the IP address of the server that hosts the IBM InfoSphere Classification Module server and the port number of the Classification Module listener component. The default port number is 18087.

Click **Verify** to verify the connection to the server and to enable or disable the support for decision plans. Decision plans are supported started with version 8.7 of IBM InfoSphere Classification Module.

Classification

With IBM InfoSphere Classification Module version 8.6, only knowledge bases are supported. For versions higher than 8.6, select the kind of classification that you want to use:

Knowledge Base
If you want to use a Classification Module knowledge base, select this and

choose the knowledge base that you want to use. Click the explore button to retrieve the list of knowledge bases that are available on the Classification Module server.

Decision plan

If you want to use a Classification Module decision plan, select this and choose the decision plan that you want to use. Click the explore button to retrieve the list of decision plans that are available on the Classification Module server.

Check **Populate metadata with XML results** to populate the system metadata property "Decision plan results exported as XML" with the decision plan results in XML format.

In the **Content field** dropdown, select the content field that you want to use to represent the content that is to be analyzed. Click the explore button to retrieve the list of available content fields.

Note: Content items such as documents or email messages that are classified by Classification Module are represented as collections of fields. Content fields store the textual content of each item, other information such as the delivery channel or author's name, and categorization information. For example, the content fields of an email message store different components of the message such as To, From, or Body. The content type of each field determines how Classification Module analyzes and classifies content items. For example, to process the content of a document, select a content field of type **Document**. If you select **File name**, Classification Module will process the file name. It is recommended to use type **Document** for both files and email.

Result Set

Specify the relevance threshold and the maximum results returned. Classification Module will return only results with scores above the relevance threshold and a maximum number of relevant categories as specified by **Maximum results returned**.

Map Decision Plan Results to Metadata

If you select **Decision Plan** as instance type, specify the metadata mapping on the **Map Decision Plan Results** tab.

Select the metadata source and map the metadata properties to Classification Module decision plan properties.

Related tasks

“Using Classification Module to classify documents” on page 453

Related reference

“IBM Classification Module system metadata properties” on page 262

“Task status system metadata properties” on page 269

MC Retrieve Additional Metadata:

This task is used to retrieve archiving information from the temporary metadata database and to associate this metadata with a specific message.

Task summary

Table 143. MC Retrieve Additional Metadata task summary

Characteristic	Value
Task name	MC Retrieve Additional Metadata
Main purpose	Retrieves archiving information from the temporary metadata database and associates this metadata with a specific message
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Optionally in email archiving task routes; can be used only if specifying additional archiving information is configured for interactive archiving
Placement in task route	Can appear only after an EC Extract Metadata task and before a CM 8.x Configure Item Types task or a P8 4.x Create Document task
Produces which metadata?	Task Status
Configuration options	Associate Metadata

Associate Metadata

Select the metadata that is to be used as correlation key. IBM Content Collector uses the correlation key to locate that record in the metadata database that belongs to the email for which the task is invoked. Thus, Content Collector can retrieve and map the custom metadata that belongs to a specific document:

- **Metadata source** and **Metadata property** determine the correlation key. For email archiving, the Email system metadata property Form Correlation Key is preselected. Do not change this selection.
- **User-defined metadata** defines the set of custom metadata to which the retrieved metadata is mapped. A metadata source can have multiple properties. Select the metadata properties that you want to extract.

Related reference

"Task status system metadata properties" on page 269

P8 4.x Confirm Document:

For use in SharePoint connector link management and auditing task routes, the P8 4.x Confirm Document task attempts to confirm the existence of a document in FileNet P8.

Task summary

Table 144. P8 4.x Confirm Document task summary

Characteristic	Value
Task name	P8 4.x Confirm Document
Main purpose	Checks whether a document exists in the FileNet P8 repository
Usable with which source connectors?	SharePoint Connector

Table 144. P8 4.x Confirm Document task summary (continued)

Characteristic	Value
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Required in link management task routes such as **SP Manage P8 Links** and **SP Audit P8 Links**
Placement in task route	Must appear before the SP Manage Link task
Produces which metadata?	Task Status
Configuration options	"Connection" "Shortcut Link"

Connection

Prerequisites: A connection to a FileNet P8 object store must exist.

Select the IBM FileNet P8 object-store connection that you want to use.

Shortcut Link

In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the object store.

Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

```
https://host_name:port_number/AFUWeb/RetrieveDocument.do?
r=<%DOCID_ENCRYPTED%>&repositoryID=<%REPOSITORY_ID_ENCRYPTED%>&sum=<%URL_CHECKSUM%>
```

To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

```
https://host_name:port_number/AFUWeb/SecureRetrieveDocument.do?
r=<%DOCID_ENCRYPTED%>&repositoryID=<%REPOSITORY_ID_ENCRYPTED%>&sum=<%URL_CHECKSUM%>
&am=<%CHALLENGE_MODE%>
```

In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.

- Replace *host_name* and *port_number* with the name and port number of the IBM Content Collector Web Application service.
- Do not alter any of the tokens <%*token_name*%> and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
- The ENCRYPTED tokens include the same source information as the unencrypted tokens that Content Collector previously used, but are now encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service.

Related reference

"Task status system metadata properties" on page 269

"P8 4.x Confirm Document system metadata properties" on page 263

P8 4.x Create Content Elements:

A multi-content document has one document ID assigned, but the document consists of more than one file, the *content elements*. Each version of a document can have different files as content elements. Use this task in a task route that is configured for archiving from a file system source, if you set up metadata files containing paths to content that you want to archive into FileNet P8. When used in an archiving task route that is configured for email compliance, this task adds the email and its attachments, created by previous tasks in the compliance task route, as individual content elements on the document archived in FileNet P8.

Task summary

Table 145. P8 4.x Create Content Elements task summary

Characteristic	Value
Task name	P8 4.x Create Content Elements
Main purpose	Creates content elements for content to which a metadata file points, or adds an email and its attachments as individual content elements
Usable with which source connectors?	Email Connector, File System Source Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Optional in task routes for archiving from a file system source; required in email archiving task routes that were created with IBM InfoSphere Content Collector Version 2.1.1.1 or before and are configured for email compliance
Placement in task route	Can be used only together with the following tasks in task routes for archiving from a file system source: • P8 4.x Create Document • FSC Associate Metadata Can be used only together with the following tasks in email archiving task routes: • P8 4.x Create Document • EC Prepare Email for Archiving • EC Extract Attachments
Produces which metadata?	P8 4.x Create Document, Task Status
Configuration options	"Connection" on page 543 "Checkin Options" on page 543

Important: Do not use this task in a task route for the purposes of email compliance and attempt to configure it by yourself. If you require a task route for email compliance, use one of the templates provided with the product. The template contains a task route with tasks pre-configured as required for electronic discovery.

The term *email compliance* is used to describe the need for an organization to capture and retain email for a period of time for the purpose of demonstrating compliance with an external or internal regulation (for example SEC-17a). Email captured for email compliance is typically subject to electronic discovery – any

process whereby electronic data is searched for use as evidence in a civil or criminal legal case. To enable effective electronic discovery on email stored in IBM FileNet P8, IBM Content Collector must process email using a special task route. This email compliance task route has dependencies on various P8 tasks. These tasks enable email to be properly content-indexed for legal discovery.

Connection

Prerequisites: A connection to a FileNet P8 object store must exist.

Select the IBM FileNet P8 object-store connection that you want to use.

Checkin Options

- Select **Skip first content element** to have IBM Content Collector ignore the first content element. When you set up metadata files with paths to multiple content elements, the first entity is the metadata file itself. You can use this option if you do not want archive the metadata file but only the content to which the file points. Select this option only when you configure a task route for archiving from a file system source and the task route also contains an FSC Associate Metadata task.
- Select **Set content retrieval name** and specify the **Retrieval name metadata mapping** to configure a specific retrieval name on a content transfer element. If no value is set then the retrieval name defaults to the file name including the file extension.

Related reference

"P8 4.x Create Document system metadata properties" on page 263

"Task status system metadata properties" on page 269

P8 4.x Create Document:

Email and files are stored as documents in the FileNet P8 4.x repository. To create a document in the repository, you need to specify where to create the document, and how to index it. You specify where by selecting a repository from a list of those configured. You index the document by choosing a document class for the item and by specifying what values should be assigned to each property of that class.

Task summary

Table 146. P8 4.x Create Document task summary

Characteristic	Value
Task name	P8 4.x Create Document
Main purpose	Creates a document in the FileNet P8 repository
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	P8
When needed?	Required in archiving task routes
Placement in task route	Usually appears before post-processing tasks and after metadata and version tasks
Produces which metadata?	P8 4.x Create Document, Task Status

Table 146. P8 4.x Create Document task summary (continued)

Characteristic	Value
Configuration options	"Connection" "Checkin Options" "Document Deduplication" on page 545 "Property Mappings" on page 546 "Data Correction" on page 547

Connection

Prerequisites: A connection to a FileNet P8 object store must exist.

Select the IBM FileNet P8 object-store connection that you want to use.

Checkin Options

With these options, you determine the way a document is stored in IBM FileNet P8. For more detailed information about each option, see the IBM FileNet P8 Enterprise Manager documentation:

- Select **Auto classify** to enable a script you create, usually in Visual Basic or XML, to manage the processing of the file for check in.
- Select **Defer checkin** if you are configuring a task route with the P8 4.x Create Content Elements. This will create a "reservation object" in P8, which will later be checked into P8 when the object gets the files it requires to be complete later in the task route. Documents left in a reservation state at the end of the task route are automatically checked in.
- Under **Version**, select an option:
  - Select **Major**. This is a P8 term for a document that has been "released". Typically, a major version's security makes the document available to a wide range of users.
  - Select **Minor**. This is a P8 term for a document version that has not been released. Typically, a minor version's security makes the document available only to the authors and reviewers.
- Under **Content capture options**:
  - Select **Transfer content (default)** to save the document in P8.

    Important: Only select this option if an EC Prepare Email for Archiving task appears before this P8 4.x Create Document task in the task route. The EC Prepare Email for Archiving task extracts the email from the mail system, and content can only be transferred if it has been extracted.
  - Select **Reference external content** to allow a reference to the document to be created in FileNet P8, and the physical document to be located on the source system. This option is valid only if this task route is using a file-system collector.

    The collection source must be a shared directory with the appropriate permissions on the share to allow access for the FileNet P8 Content Engine and Application Engine (Workplace) servers. When you configure the path for collection from a shared directory, the collection-source directory must be in the format `\\servername\shared_folder\`, so that the FileNet P8 content reference is in a recognizable format for the FileNet P8 system.

– Select **Do not transfer content (contentless)** to allow document metadata only to be saved in P8. This option must be selected if the P8 4.x Create Document task appears before the EC Prepare Email for Archiving task in the task route.

- Select **Set content retrieval name** to set a metadata property to use for the Retrieval Name on the content element. This option is enabled only if you have selected the **Transfer content (default)** option.
- In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the object store.

  Important: The Email Connector uses its own link format and ignores the contents of the **Shortcut link** field.
  Enter the URL in the same format as the sample URL in the text box. Modify the sample URL provided:

  ```
  https://host_name:port_number/AFUWeb/RetrieveDocument.do?
  r=<%DOCID_ENCRYPTED%>&repositoryID=<%REPOSITORY_ID_ENCRYPTED%>&sum=<%URL_CHECKSUM%>
  ```

  To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

  ```
  https://host_name:port_number/AFUWeb/SecureRetrieveDocument.do?
  r=<%DOCID_ENCRYPTED%>&repositoryID=<%REPOSITORY_ID_ENCRYPTED%>&sum=<%URL_CHECKSUM%>
  &am=<%CHALLENGE_MODE%>
  ```

  In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.

  – Replace *host_name* and *port_number* with the name and port number of the IBM Content Collector Web Application service.
  – Do not alter any of the tokens <*%token_name%*> and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
  – The ENCRYPTED tokens include the same source information as the unencrypted tokens that Content Collector previously used, but are now encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service.

  Tip: Users of previous versions of Content Collector can retain URLs in the previous format:

  ```
  http://server_name/Workplace/getContent?
  ObjectStoreName=<%OBJSTORE%>&id=<%DOCID%>&objectType=<%OBJTYPE%>
  ```

  – Replace *server_name* with the name and port of the FileNet P8 Application Server that is running Workplace.
  – Do not alter any of the tokens <*%token_name%*> and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.

Document Deduplication

Select **Detect duplicates** to enable duplicate detection.

From the **Hash key metadata mapping** lists, select a metadata class and property to use to detect duplicates. In task routes using email collectors, a hash key is automatically generated and available in these lists for duplicate detection. For task routes using a file system collector, you must enable hash key generation when

configuring the collector to enable the hash key to be available in these lists. For task routes using a SharePoint collector, you must enable hash key generation in the SP Create File task.

Select one of the following algorithms to detect duplicates:

Always create document
This is the default. Content Collector tries to create a document in FileNet P8 without checking for duplicates first. If an object with the same ID already exists in the repository, creating another one would violate a uniqueness constraint, so the underlying database produces an error. The errors are recorded in the log file of the Web Application container. This results in a processing overhead.

Select this option if you expect only a low number of duplicate documents.

Check before creating document
Enable Content Collector to search the repository for an object with the same ID before trying to create the document. In this case, no errors are produced, so that no logging activity is created. However, a database query is run for every document that is processed.

Select this option if you expect a large number of duplicate documents.

If a low number of duplicate documents is expected, leaving this option unchecked improves performance by reducing the database load. Content Collector will attempt to create a new document without checking for duplicates first. If a duplicate is found in the repository, the underlying database will throw a uniqueness constraint violation exception and incur the processing overhead. These exceptions are not logged as this behavior is expected.

Tip: Set a decision point and rules immediately after the P8 4.x Create Document task to specify what to do with duplicates and non-duplicates.

SharePoint only: Do not attempt to use deduplication with the SharePoint collector if you either:

- Collect version series documents, or
- Collect Microsoft Office documents (SharePoint changes the metadata that the collector uses to identify identical documents)

Doing so slows your system and results in no deduplication.

Property Mappings

The **Document class** list box lists all document classes in the selected object store. Select the document class you want to use as base document class when creating the object in the repository.

Important:

- If document classes are added while Content Collector is running, you will need to restart the application to see the new classes in the list.
- The default document class for email archiving in the P8 4.x Create Document task is *ICCMail2*.

If you have permission to modify system properties on the FileNet P8Content Engine, you can click to display and map settable FileNet P8 system properties such as:

- Date Created, the date when the document is added to the repository
- Creator, the login name of the user adding the document to the repository
- Last Modifier, the name of the user who last modified the document
- Date Last Modified, the date the document was last modified

To be able to map certain Content Engine system properties, your object store security must be set to **Modify certain system properties** in FileNet P8 Enterprise Manager. Restart the IBM Content Collector Configuration Manager after granting this right.

Click to display configurable properties that have been marked as hidden in the repository.

Edit the property mappings for the selected document class as required.

Lotus Notes only: When Lotus Notes email is captured, the Folder metadata field is only filled with a name if the corresponding email was obtained through a folder, that is, if a collector for manual archiving is configured to monitor drag-and-drop folders or if a collector for automatic archiving is configured to include folders. For email that is collected in another way, this field will always be empty.

If you want Content Collector to determine the class dynamically, click **Advanced** and select to use an expression to determine the class. Configure the expression by using the Expression Editor. Content Collector can also dynamically create property mappings for these classes based on user-defined metadata sources. In this case, the name of the user-defined metadata source must match the symbolic name of the repository property. You can select the properties for dynamic mapping in the Advanced Options window.

Data Correction

Select **Truncate strings** to issue a warning and truncate any string metadata values set in Property Mappings to fit inside the maximum length of the string property in P8. Leaving this box unchecked causes the task to fail if a string metadata value cannot fit inside the P8 property it is mapped to.

Select **Ignore choice list properties on error** to issue a warning and skip setting a metadata value on a P8 property that has a choice list associated with it. Leaving this box unchecked causes the task to fail if a metadata value does not match any of the values in the choice list associated with the P8 property it is mapped to.

Related tasks

"Configuring the IBM FileNet P8 Connector and connections" on page 205
"Tips to ensure proper placement of tasks in a task route" on page 278
"Assigning property values" on page 490
"Working with the Expression Editor" on page 404
"List Lookup" on page 496
"Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records" on page 726
"Defining metadata to be used to process files for archiving" on page 729

Related reference

"P8 4.x Create Document system metadata properties" on page 263
"Task status system metadata properties" on page 269

P8 4.x Create Email Instance:

This task is used in a task route configured for email compliance. The task extracts from email metadata-specific instance data that is required for electronic discovery, such as locations in which email was found. Note that the **Check for duplicates** option must be enabled in a P8 4.x Create Document task placed prior to this task in the task route to allow for this location check. You can use this data to classify documents in the repository and provide additional search criteria for metadata-based queries.

Task summary

Table 147. P8 4.x Create Email Instance task summary

Characteristic	Value
Task name	P8 4.x Create Email Instance
Main purpose	Creates a FileNet P8 instance record for the copy of an email document in the user's mailbox
Usable with which source connectors?	Email Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Required in email archiving task routes that are configured for email compliance
Placement in task route	Can be used only together with the P8 4.x Save Prepared Text as XML and P8 4.x Create Content Elements tasks; can appear only after a P8 4.x Create Document task
Produces which metadata?	P8 4.x Create Email Instance, Task Status
Configuration options	"Connection" on page 549 "Expiration Metadata Mapping Options" on page 549 "Hashkey Metadata Mapping Options" on page 549 "Property Mappings" on page 550 "Data Correction" on page 550

Important: Do not use this task in a task route for the purposes of email compliance and attempt to configure it by yourself. If you require a task route for

email compliance, use one of the templates provided with the product. The template contains a task route with tasks pre-configured as required for electronic discovery.

The term *email compliance* is used to describe the need for an organization to capture and retain email for a period of time for the purpose of demonstrating compliance with an external or internal regulation (for example SEC-17a). Email captured for email compliance is typically subject to electronic discovery – any process whereby electronic data is searched for use as evidence in a civil or criminal legal case. To enable effective electronic discovery on email stored in IBM FileNet P8, IBM Content Collector must process email using a special task route. This email compliance task route has dependencies on various P8 tasks. These tasks enable email to be properly content-indexed for legal discovery.

Connection

Prerequisites: A connection to a FileNet P8 object store must exist.

Select the IBM FileNet P8 object-store connection that you want to use.

The object store needs to contain the object class that you want to add the metadata to.

Expiration Metadata Mapping Options

Select **Set an expiration date** to set a metadata date value on the email that was archived in the P8 4.x Create Document task. This option is available in this task because the expiration date can differ across email instances. If the expiration dates between email instances differ, the task selects the date that is furthest into the future, that is, the longest retention period.

Hashkey Metadata Mapping Options

In the P8 4.x Create Document task, deduplication is done by merging the data of a set of uniquely identifiable email documents, based on property values. The P8 4.x Create Email Instance task creates an email instance object that also stores all data that is not shared among the copies of an email that were determined by the P8 4.x Create Document task. Select **Detect duplicates** to enable duplicate detection for email instance objects that are stored in FileNet P8.

Tip: Set a decision point and rules immediately after the P8 4.x Create Email Instance task to specify what to do with duplicates and non-duplicates.

From the **Hash key metadata mapping** lists, select a metadata class and property to use to detect duplicates. In task routes using email collectors, a hash key is automatically generated and available in these lists for duplicate detection. For task routes using a file system collector, you must enable hash key generation when configuring the collector to enable the hash key to be available in these lists. For task routes using a SharePoint collector, you must enable hash key generation in the SP Create File task.

The task route templates that are provided with IBM Content Collector Version 2.2 are already configured accordingly.

Property Mappings

From the **Custom object class** list, select the object class to which you want to add the metadata. The table at the bottom of the pane is populated with the properties (table columns holding metadata) of the selected object class.

Important: The default document class for email archiving in the P8 4.x Create Email Instance task is *ICCMailInstance2*.

You probably do not want to write values to hidden properties because these are usually reserved for system-related or non-public information. To prevent accidentally adding a metadata value for a hidden property, hidden properties are, by default, not shown in the table. However, if you do want to add metadata to hidden properties at archiving time, select **Display hidden properties**.

Select a property in the table and click **Edit**. In the **Value** field of the Edit window, type the value (metadata information) that you want to store in the property when a document is archived.

Data Correction

Select **Truncate strings** to issue a warning and truncate any string metadata values set in Property Mappings to fit inside the maximum length of the string property in P8. Leaving this box unchecked causes the task to fail if a string metadata value cannot fit inside the P8 property it is mapped to.

Select **Ignore choice list properties on error** to issue a warning and skip setting a metadata value on a P8 property that has a choice list associated with it. Leaving this box unchecked causes the task to fail if a metadata value does not match any of the values in the choice list associated with the P8 property it is mapped to.

Related tasks

"Assigning property values" on page 490

Related reference

"P8 4.x Create Email Instance system metadata properties" on page 264

"Task status system metadata properties" on page 269

P8 4.x Create Version Series:

This task creates a version series in FileNet P8 for the Microsoft SharePoint document that is being processed.

Task summary

Table 148. P8 4.x Create Version Series task summary

Characteristic	Value
Task name	P8 4.x Create Version Series
Main purpose	Creates a version series for a document in the FileNet P8 repository
Usable with which source connectors?	SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	When you want to add multiple versions of SharePoint documents to FileNet P8, or re-collect processed SharePoint documents

Table 148. P8 4.x Create Version Series task summary (continued)

Characteristic	Value
Placement in task route	Cannot appear before SP Get Versions task and cannot appear beforeSP Post-processing task
Produces which metadata?	P8 4.x Create Document, Task Status
Configuration options	"Connection" "Checkin Options" "Property Mappings" on page 552 "Data Correction" on page 553

Connection

Prerequisites: A connection to a FileNet P8 object store must exist.

Select the IBM FileNet P8 object-store connection that you want to use.

Checkin Options

With these options, you determine the way a document is stored in IBM FileNet P8. For more detailed information about each option, see the IBM FileNet P8 Enterprise Manager documentation:

- Select **Auto classify** to enable a script you create, usually in Visual Basic or XML, to manage the processing of the file for check in.
- Under **Content capture options**:
  - Select **Transfer content (default)** to save the document in P8.
  - Select **Reference external content** to allow a reference to the document to be created in P8, and the physical document to be located on the source system. This option is valid only if this task route is using a file system collector.
  - Select **Do not transfer content (contentless)** to allow document metadata only to be saved in P8.
- Select **Set content retrieval name** to set a metadata property to use for the Retrieval Name on the content element. This option is enabled only if you have selected the **Transfer content (default)** option.
- In the **Shortcut link** text box, enter the URL to be used when adding a shortcut to a document in the object store. Modify the sample URL provided:

  ```
  https://host_name:port_number/AFUWeb/RetrieveDocument.do?
  r=<%DOCID_ENCRYPTED%>&repositoryID=<%REPOSITORY_ID_ENCRYPTED%>&sum=<%URL_CHECKSUM%>
  ```

  To provide secure links to archived documents that require users to log on to the repository before they can access the content, provide the URL in this format:

  ```
  https://host_name:port_number/AFUWeb/SecureRetrieveDocument.do?
  r=<%DOCID_ENCRYPTED%>&repositoryID=<%REPOSITORY_ID_ENCRYPTED%>&sum=<%URL_CHECKSUM%>
  &am=<%CHALLENGE_MODE%>
  ```

  In this case, the repository connection is established with the user's credentials and access to the item in the repository is granted based on the user's access rights.
  - Replace *host_name* and *port_number* with the name and port number of the IBM Content Collector Web Application service.

– Do not alter any of the tokens <%*token_name*%> and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.
– The ENCRYPTED tokens include the same source information as the unencrypted tokens that Content Collector previously used, but are now encrypted with an algorithm that is compatible with the IBM Content Collector Web Application service.

Tip: Users of previous versions of Content Collector can retain URLs in the previous format:

```
http://server_name/Workplace/getContent?
ObjectStoreName=<%OBJSTORE%>&id=<%DOCID%>&objectType=<%OBJTYPE%>
```

– Replace *server_name* with the name and port of the FileNet P8 Application Server that is running Workplace.
– Do not alter any of the tokens <%*token_name*%> and adhere to the order of the parameters except for the parameter &sum. This parameter can appear anywhere in the parameter list.

Property Mappings

The **Document class** list box lists all document classes in the selected object store. Select the document class you want to use as base document class when creating the object in the repository.

Important: If document classes are added while Content Collector is running, you will need to restart the application to see the new classes in the list.

If you have permission to modify system properties on the FileNet P8Content Engine, you can click to display and map settable FileNet P8 system properties such as:

- Date Created, the date when the document is added to the repository
- Creator, the login name of the user adding the document to the repository
- Last Modifier, the name of the user who last modified the document
- Date Last Modified, the date the document was last modified

To be able to map certain Content Engine system properties, your object store security must be set to **Modify certain system properties** in FileNet P8 Enterprise Manager. Restart the IBM Content Collector Configuration Manager after granting this right.

Click to display configurable properties that have been marked as hidden in the repository.

Edit the property mappings for the selected document class as required.

If you want Content Collector to determine the class dynamically, click **Advanced** and select to use an expression to determine the class. Configure the expression by using the Expression Editor. Content Collector can also dynamically create property mappings for these classes based on user-defined metadata sources. In this case, the name of the user-defined metadata source must match the symbolic name of the repository property. You can select the properties for dynamic mapping in the Advanced Options window.

Data Correction

Select **Truncate strings** to issue a warning and truncate any string metadata values set in Property Mappings to fit inside the maximum length of the string property in P8. Leaving this box unchecked causes the task to fail if a string metadata value cannot fit inside the P8 property it is mapped to.

Select **Ignore choice list properties on error** to issue a warning and skip setting a metadata value on a P8 property that has a choice list associated with it. Leaving this box unchecked causes the task to fail if a metadata value does not match any of the values in the choice list associated with the P8 property it is mapped to.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Working with the Expression Editor" on page 404

"Assigning property values" on page 490

Related reference

"P8 4.x Create Document system metadata properties" on page 263

"Task status system metadata properties" on page 269

P8 4.x Declare Record:

This task allows you to configure properties related to declaring a record in IBM InfoSphere Enterprise Records.

Task summary

Table 149. P8 4.x Declare Record task summary

Characteristic	Value
Task name	P8 4.x Declare Record
Main purpose	Configure properties related to declaring a record in IBM InfoSphere Enterprise Records (formerly FileNet P8 Records Manager)
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Whenever you want to declare records in IBM InfoSphere Enterprise Records
Placement in task route	Can appear after a P8 4.x Create Document task or, if applicable, after one of these tasks: • P8 4.x Create Content Elements • P8 4.x File Document in Folder • P8 4.x Save Prepared Text as XML in email archiving task routes when you work with the compound email data model. The P8 4.x Declare Record task can be used only when processing non-duplicate email. Make sure to set an appropriate decision point in your task route so that only non-duplicate email is passed along the path where this task is included.
Produces which metadata?	P8 4.x Declare Record, Task Status

Table 149. P8 4.x Declare Record task summary (continued)

Characteristic	Value
Configuration options	"Connection" "Property Mappings" "Data correction" "Configure Classifications"

IBM InfoSphere Enterprise Records must be installed and configured to declare a record.

Note that the Can Declare property of the record's document class must be set to True to be able to declare a record against the document.

Connection

From the **Connections** list, select the object store in the repository in which to declare the record.

Property Mappings

In the **Property Mappings** section, click **Browse** to select a record class to use when declaring the record. The property list screen will be populated with properties and values of that class.

The property mappings table displays a list of properties (including custom properties) configured for the selected record class. An asterisk (*) next to the name is used to indicate properties that are mandatory. Edit the property mappings for the selected document class as required.

If you want Content Collector to determine the class dynamically, click **Advanced** and select to use an expression to determine the class. Configure the expression by using the Expression Editor. Content Collector can also dynamically create property mappings for these classes based on user-defined metadata sources. In this case, the name of the user-defined metadata source must match the symbolic name of the repository property. You can select the properties for dynamic mapping in the Advanced Options window.

Data correction

- Select **Truncate strings** to issue only a warning and to truncate any string metadata values set in the Property Mappings to fit inside the maximum length of the string property in FileNet P8. If you do not select this, the task fails if a string metadata value cannot fit inside the FileNet P8 property it is mapped to.
- Select **Ignore choice list properties on error** to issue only a warning and to skip setting a metadata value on a FileNet P8 property that has a choice list associated with it. If you do not select this, the task fails if a metadata value does not match any of the values in the choice list associated with the FileNet P8 property it is mapped to.

Configure Classifications

To classify a record, click **Add**, or to remove a classification, click **Remove**. The classification can be a static value or can be dynamically assigned (that is, based on the name of the folder in which an item is located).

Important:

Any categories or folders to which you want to classify records must already exist in the object store. This task cannot create new classification paths.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Working with the Expression Editor" on page 404

"List Lookup" on page 496

"Assigning property values" on page 490

"Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records" on page 726

Related reference

"P8 4.x Declare Record system metadata properties" on page 264

"Task status system metadata properties" on page 269

P8 4.x File Document in Folder:

This task files document objects that were created in the repository in specific folders, so that you can browse for them at a later date.

Tip: When you are mapping metadata to be used as FileNet P8 folder names, refer to the Content Engine documentation for reserved characters. Most reserved characters are removed from the mapped metadata values when processed. However, \ or / are used to determine new folder creation.

Task summary

Table 150. P8 4.x File Document in Folder task summary

Characteristic	Value
Task name	P8 4.x File Document in Folder
Main purpose	Files document objects in specific folders
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Whenever you want to file a document created in the P8 4.x Create Document task into a FileNet P8 folder
Placement in task route	Can appear only after the P8 4.x Create Document task
Produces which metadata?	P8 4.x File Document in Folder, Task Status
Configuration options	"Connection" "File in Folder Options" on page 556

Connection

Select the IBM FileNet P8 object-store connection that you want to use.

File in Folder Options

In the **Folder Path** text box, provide the complete path to the folder in IBM FileNet P8.

Use the **Calculated value** feature when adding FileNet P8 Content Engine folder paths. Using a static value (\) between metadata or other static values provides a flexible way to add folder paths.

SharePoint only: You can typically recreate a SharePoint document's folder path by using a calculated value, as in this example:

1. Click **Add** then **Calculated value**.
2. In the Edit Calculated Value window, add the following items in sequence:
   - Literal: `MySharePointArchives\`
   - Metadata: Metadata type: `SP Collection`; Property: `Site`
   - Literal: `\`
   - Metadata: Metadata type: `SP Collection`; Property: `Library`
   - Literal: `\`
   - Metadata: Metadata type: `SP Collection`; Property: `Folder Path`

Documents residing in a SharePoint site named *ABC Company*, library named *HR*, and folder path of *Internal\Meeting Minutes* would have a folder path created and filed to: `MySharePointArchives\ABC Company\HR\Internal\Meeting Minutes`

Select **Inherit folder security** to allow folder security to be inherited, and specify how inheritance is applied:

- Select **Set security parent of document to folder** to modify the security of the document so that it also inherits the current security of the folder it is located in. For example, if the security of the parent folder has GroupA with **Full Control** and GroupB with **View Properties**, the GroupB security is inheritable. This additional security configuration will be inherited and applied to the document's current security.
- Select **Add folder security to document security** to merge existing folder security with document security. Select **Add folder security to document security** to merge existing folder security with document security. If the same grantee is included on both, the merging of security will be handled by the FileNet P8 Content Engine security model.

Selecting **Inherit folder security**:

- applies the security settings of only the parent folder and not of folders higher up, and
- overwrites the values of the system metadata properties *Modified* and *Last modified by*, which an upstream archival task might have set.

SharePoint only: To implement version series processing without risking the filing of a document multiple times in the same folder, you must add a decision point and rule to your task route, using the following criteria:

- Metadata type: `SP Collection`; Property: `Last Version`; Operator: `Equal`; Value: Literal: `True`
- Metadata type: `Re-collection`; Property: `Re-collection Flag`; Operator: `Equal`; Value: Literal: `False`

Link them using the conjunction operator `And`. If you are not re-collecting documents, you can omit the second criterion.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Assigning property values" on page 490

"List Lookup" on page 496

"Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records" on page 726

"Defining metadata to be used to process files for archiving" on page 729

Related reference

"P8 4.x File Document in Folder system metadata properties" on page 264

"Task status system metadata properties" on page 269

P8 4.x Link Documents:

This task creates a link object that associates email with its attachments, or files with other files you want to be linked to determine what email and files were originally associated with each other.

Task summary

Table 151. P8 4.x Link Documents task summary

Characteristic	Value
Task name	P8 4.x Link Documents
Main purpose	Links two or more documents that are saved in the repository
Usable with which source connectors?	Email Connector, File System Source Connector
Usable with which target connectors?IBM FileNet P8 Connector	FileNet P8
When needed?	Whenever you want to create a link item in FileNet P8 to associate two or more documents together
Placement in task route	Can appear only after the P8 4.x Create Document task
Produces which metadata?	P8 4.x Link Document, Task Status
Configuration options	"Connection" "Property Mappings" "Data correction" on page 558

Connection

Select the IBM FileNet P8 object-store connection that you want to use.

Property Mappings

From the **Link Class** list box, select the link class to use to link the items together. In the table listing properties of the link class selected, add, edit, or remove property values.

Select to display hidden properties in the table listing properties of the document class.

Data correction

- Select **Truncate strings** to issue only a warning and to truncate any string metadata values set in the Property Mappings to fit inside the maximum length of the string property in FileNet P8. If you do not select this, the task fails if a string metadata value cannot fit inside the FileNet P8 property it is mapped to.
- Select **Ignore choice list properties on error** to issue only a warning and to skip setting a metadata value on a FileNet P8 property that has a choice list associated with it. If you do not select this, the task fails if a metadata value does not match any of the values in the choice list associated with the FileNet P8 property it is mapped to.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

"Assigning property values" on page 490

"List Lookup" on page 496

Related reference

"P8 4.x Link Documents system metadata properties" on page 265

"Task status system metadata properties" on page 269

P8 4.x Modify Object Security:

You can specify the security to apply to a document, a record, or a list of folders when it is created within the repository. You do so by selecting the user or group to whom you want to grant or deny document access, then setting security privileges on the object that you want the selected user or group to have. You can select users or groups based on metadata that contains a list of user IDs that may be in the LDAP directory, such as the email properties To, From, and CC. This is a simple way to restrict access only to senders or recipients of the email.

Use this task whenever you want to change the direct permissions on any document, record, or folder that is created by IBM Content Collector. However, when you declare records in your task route, changing the security on the document will have no effect because, in this case, the permissions are determined by the record. Therefore, you must modify the permissions on the record to change the security settings on the document.

Task summary

Table 152. P8 4.x Modify Object Security task summary

Characteristic	Value
Task name	P8 4.x Modify Object Security
Main purpose	Sets FileNet P8 object security
Usable with which source connectors?	Email Connector, File System Source Connector, SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Optional but needed to modify the access permission list for an object created within the task route

Table 152. P8 4.x Modify Object Security task summary (continued)

Characteristic	Value
Placement in task route	Can appear only after the P8 4.x Create Document, P8 4.x File Document in Folder, P8 4.x Create Version Series, or P8 4.x Declare Record task.
Produces which metadata?	P8 4.x Modify Document Security, Task Status
Configuration options	"Connection" "Security Settings" "Security Setting Type" on page 561

Connection

Select the connection to the IBM FileNet P8 object store connection that contains the document, folder, or record for which you want to modify the permissions.

Input Object Mapping

Select the type of FileNet P8 object for which you want to modify the security settings and the corresponding **Input object IDs**. The object IDs allow the task to find the objects in FileNet P8. The selected type of FileNet P8 object determines the security settings:

Object type	Source	Property
Document	P8 4.x Create Document	Object ID
Folder	P8 4.x File Document in Folder	File Path
Record	P8 4.x Declare Record	Object ID

Security Settings

You can choose to have Content Collector automatically collect security information about source objects, or you can manually configure the security settings. When you select both options, both sets of security settings are applied to the FileNet P8 object. FileNet P8 Content Engine then determines the final security settings to be applied according to the documented security inheritance model. Setting security dynamically is supported only by the File System Source Connector and the SharePoint Connector. Setting security manually is supported for all source connectors for which you can use this task.

When you automatically collect security information about items in the source system, your task route must contain an upstream task that extracts the security metadata and provides it for use. The rules for how security on the source system is mapped to the FileNet P8 security on the target object are predetermined and do not require configuration. The source object's property that represents the grantee must be specified in an LDAP short name format to have the security successfully mapped, otherwise the permission is skipped. Automatic collection of security information is available only for use with FileNet P8 documents and records.

You can manually configure the security settings by selecting the grantees and the access rights that are to be applied to the P8 document, folder, or record. However,

when you modify folder security, only the last folder in each file path that is processed receives the modification. When you manually configure the security settings for FileNet P8 objects, you can add or remove grantees and select appropriate settings for the access type, access level, and access rights.

To add a user or group to whom to assign or deny access to the document, click **Add**. In the Find Users and Groups window, you can search for a user or group by name, or by the email properties To, From, or CC.

- To search by name:
  1. Click the **Find Users and Groups** tab.
  2. From the **Select Realm** list, select a location in the LDAP directory to search. The realm displayed by default is the current directory.
  3. From the **Select Object Type** section, select any or all object types on which to search. The *special accounts* are the #CREATOR-OWNER and #AUTHENTICATED-USERS accounts. For more information about these special accounts see the section about required users and groups in the IBM FileNet P8 information center.
  4. In the **Search Criteria** section:
     a. Select **Short name** or **Display name**, select an operator for the search, and enter the value for which to search.
     b. From the **Sort order** list box, select an order in which to display search results.
     c. Enter or select a maximum number of results to display.
     d. Click **Find**.

  Results display in the table at the bottom of the screen.
- To search by email properties:
  1. Click the **Select Metadata** tab.
  2. Select a metadata type and property on which to search.

     The property that represents the grantee name must be in an LDAP short name format for the task to successfully set the grantee. If the grantee name cannot be resolved, that user is skipped.

Click **OK** to save, or **Cancel** to close without saving.

Access Type

You can allow or deny access for the specified users or groups.

Access Level and Access Rights

When you configure security settings manually, you can set the access available to the specified users or groups.

Access levels represent a group of access rights you want to allow or deny a user. You can configure an access level explicitly by selecting a predefined FileNet P8 access right group, or through a single-valued metadata reference. For example, if you select the **Full Control** access level on the left, by default all access rights that this access level comprises become selected on the right, and the selected users or groups are granted or denied this access. You can maintain the default level, or customize by selecting only those access rights you want to apply.

Security Setting Type

You can choose to overwrite or to merge previously defined object security with the settings you just specified. To overwrite the users and groups previously granted or denied access permission by adding the selected users and groups, select **Overwrite object security with selected users and groups**. To merge the security just defined with previously defined security, select **Add selected users and groups to object security**.

The following considerations apply when you are re-collecting SharePoint documents and set security dynamically:

- When you select **Add selected users and groups to object security** and the user of a SharePoint document exists on the access control list (ACL) of the FileNet P8 document, the SharePoint document permissions are merged with the existing FileNet P8 permissions. Therefore, if a user has lesser permissions for re-collected versions of the SharePoint document, this will not be reflected on the ACL of the FileNet P8 document. That is, the permissions will not be removed from the FileNet P8 document's ACL with the re-collected versions.
- When you select **Overwrite object security with selected users and groups**, all SharePoint permission modifications are applied during re-collection.

With this task you can modify only those permissions that are assigned directly to the object, that is, which are contained in direct Access Control Entries (ACEs). You cannot modify any permissions of default, template, or inherited ACEs. Content Collector creates an ACE and adds it to the ACL. When those new permissions are saved, FileNet P8 merges all direct permissions as far as possible. FileNet P8 Content Engine then determines the security settings to be applied according to the given evaluation order for the ACE source and type. For more information about the evaluation order, see the topic about access rights in the IBM FileNet P8 information center.

Related tasks

"Tips to ensure proper placement of tasks in a task route" on page 278

Related reference

"P8 4.x Modify Object Security system metadata properties" on page 265

"Task status system metadata properties" on page 269

P8 4.x Save Prepared Text as XML:

This task saves text that was extracted previously in the task route into XML format in a document in FileNet P8, when the email is archived. The values specified in this task are associated with existing properties of a selected object store. The entire XML data is then added to the text-search index. You can use these values to classify documents in the text-search index and thus provide additional search criteria to the index.

Task summary

Table 153. P8 4.x Save Prepared Text as XML task summary

Characteristic	Value
Task name	P8 4.x Save Prepared Text as XML
Main purpose	Stores all information required for full text search on email in the P8 repository
Usable with which source connectors?	Email Connector

Table 153. P8 4.x Save Prepared Text as XML task summary (continued)

Characteristic	Value
Usable with which target connectors?	IBM FileNet P8 Connector
When needed?	Required in email archiving task routes that are configured for email compliance
Placement in task route	Can be used only together with the P8 4.x Create Email Instance task
Produces which metadata?	P8 4.x Create Document, P8 4.x Save Prepared Text as XML, Task Status
Configuration options	"Connection" "Property Mappings" "Data correction" on page 563 "Custom XIT Metadata" on page 563

Important: Do not use this task in a task route for the purposes of email compliance and attempt to configure it by yourself. If you require a task route for email compliance, use one of the templates provided with the product. The template contains a task route with tasks pre-configured as required for electronic discovery.

The term *email compliance* is used to describe the need for an organization to capture and retain email for a period of time for the purpose of demonstrating compliance with an external or internal regulation (for example SEC-17a). Email captured for email compliance is typically subject to electronic discovery – any process whereby electronic data is searched for use as evidence in a civil or criminal legal case. To enable effective electronic discovery on email stored in IBM FileNet P8, IBM Content Collector must process email using a special task route. This email compliance task route has dependencies on various P8 tasks. These tasks enable email to be properly content-indexed for legal discovery.

Connection

Prerequisites: A connection to a FileNet P8 object store must exist.

Select the IBM FileNet P8 object-store connection that you want to use.

The object store must contain the document class to which you want to add the metadata.

Property Mappings

Select the **Document class** that you want to use. The table at the bottom of the pane is populated with the properties (table columns holding metadata) of the selected document class.

Important: The default document class for email archiving in the P8 4.x Save Prepared Text as XML task is *ICCMailSearch2*.

You probably do not want to write values to hidden properties because these are usually reserved for system-related or non-public information. To prevent accidentally writing data to a hidden property, these properties are, by default, not

shown in the table. However, if you do want to add XML data to hidden properties at archiving time, select **Display hidden properties**.

Select a property in the table and click **Edit**. In the **Value** field of the Edit window, type the value (metadata information) that you want to save as XML information when a document is archived.

Data correction

- Select **Truncate strings** to issue only a warning and to truncate any string metadata values set in the Property Mappings to fit inside the maximum length of the string property in FileNet P8. If you do not select this, the task fails if a string metadata value cannot fit inside the FileNet P8 property it is mapped to.
- Select **Ignore choice list properties on error** to issue only a warning and to skip setting a metadata value on a FileNet P8 property that has a choice list associated with it. If you do not select this, the task fails if a metadata value does not match any of the values in the choice list associated with the FileNet P8 property it is mapped to.

Custom XIT Metadata

To add your own custom string metadata to the XML file of XIT:

1. Select the **Custom XIT Metadata** tab at the top of the task configuration pane.
2. Click the **Add...** icon.
3. Enter in a custom XIT element name. This name cannot start with the system prefix `icc_`
4. Select the datatype of your custom values. Only single-value and multi-value strings are supported.
5. Select the metadata source and property.
6. Select **Write property as raw XML** if you want to write your custom metadata as true XML. Be careful when enabling this option: If your custom metadata is not valid XML, the XML validation in the task causes it to fail and the email is not archived or full-text searchable.
7. Click **OK**.

To edit your custom metadata settings, select the custom metadata you need to edit and click the **Edit...** icon.

To remove your custom metadata settings, select the custom metadata you need to remove and click the **Remove** icon.

Related tasks

"Assigning property values" on page 490

Related reference

"P8 4.x Create Document system metadata properties" on page 263

"P8 4.x Save Prepared Text as XML system metadata properties" on page 265

"Task status system metadata properties" on page 269

Save Temporary File Copy:

This task copies every document that was processed to the specified location in the file system. Use it for troubleshooting or temporary backups.

Task summary

Table 154. Save Temporary File Copy task summary

Characteristic	Value
Task name	Save Temporary File Copy
Main purpose	Copies every document that was processed to the specified location in the file system
Usable with which source connectors?	Email Connector, File System Source Connector
Usable with which target connectors?	File System Repository Connector
When needed?	Optional
Placement in task route	Can appear anywhere before postprocessing
Produces which metadata?	Task Status
Configuration options	"Save Temporary File Copy Options"

Save Temporary File Copy Options

Select **Enable File Retention** to have IBM Content Collector copy every document that was processed by a collector to the specified **Retention folder**. Also specify the retention period by selecting the number of **Days to retain files**.

Related reference

"Task status system metadata properties" on page 269

SC Delete Email:

This task deletes the received email from the message queue directory.

Task summary

Table 155. SC Delete Email task summary

Characteristic	Value
Task name	SC Delete Email
Main purpose	Deletes the received email from the message queue directory.
Usable with which source connectors?	SMTP Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in SMTP email archiving task routes
Placement in task route	Can appear only after an SC Prepare Email for Deletion task
Produces which metadata?	Task Status
Configuration options	None

Related reference

"Task status system metadata properties" on page 269

SC Extract Attachments:

This task extracts the attachments from the email in the message queue directory so that the attachments can be stored as separate files in the repository. If you do not add this task to the task route, the attachments will not be saved in the repository as separate objects.

Task summary

Table 156. SC Extract Attachments task summary

Characteristic	Value
Task name	SC Extract Attachments
Main purpose	Extracts the attachments from the email in the message queue directory.
Usable with which source connectors?	SMTP Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in SMTP email archiving task routes for the compound data model
Placement in task route	Can appear only after the SC Prepare Email for Archiving task
Produces which metadata?	Attachment Deduplication, Email, File, Task Status
Configuration options	None

Related reference

"Attachment Deduplication system metadata properties" on page 246

"Email system metadata properties" on page 250

"File system metadata properties" on page 256

"Task status system metadata properties" on page 269

SC Extract Metadata:

This task extracts metadata from fields in an document to store this metadata in corresponding fields in the repository.

The metadata fields in a repository provide search information for user queries. For example, the repository field that corresponds to the From or Sender field of an email allows users to search for email that was sent by a specific person. Certain email fields are selected by default, such as the Subject, To, or Sender (From) fields. You can select other fields to extract metadata from if you think that these fields add valuable search information to your repository.

Task summary

Table 157. SC Extract Metadata task summary

Characteristic	Value
Task name	SC Extract Metadata
Main purpose	Extracts metadata from fields in an document to store this metadata in corresponding fields in the repository.
Usable with which source connectors?	SMTP Connector

Table 157. SC Extract Metadata task summary (continued)

Characteristic	Value
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in SMTP email archiving task routes
Placement in task route	Must be the first task in a task route
Produces which metadata?	Email, Email Deduplication, Task Status
Configuration options	"Associate Metadata"

Associate Metadata

To extract metadata from additional document fields, such as the SMTP header fields, select the appropriate set of fields from the **User defined metadata** list. Such sets of fields must have been defined earlier, in the **User Defined Metadata** section of the **Metadata and Lists** configuration.

Related reference

"Email system metadata properties" on page 250

"Email Deduplication system metadata properties" on page 255

"Task status system metadata properties" on page 269

SC Prepare Email for Archiving:

This task creates working copies of the email in the message queue directory, excluding attachments, and saves them as temporary files for further processing.

Task summary

Table 158. SC Prepare Email for Archiving task summary

Characteristic	Value
Task name	SC Prepare Email for Archiving
Main purpose	Creates working copies of the email in the message queue directory, excluding attachments, and saves them as temporary files for further processing.
Usable with which source connectors?	SMTP Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in SMTP email archiving task routes
Placement in task route	Must appear before a task that archives the document
Produces which metadata?	File, Archiving Format, Task Status
Configuration options	None

Related reference

"Archiving format system metadata properties" on page 246

"File system metadata properties" on page 256

"Task status system metadata properties" on page 269

SC Prepare Email for Deletion:

This task marks a document as archived once it has been successfully archived. The document can then be deleted by the SC Delete Email task.

Task summary

Table 159. SC Prepare Email for Deletion task summary

Characteristic	Value
Task name	SC Prepare Email for Deletion
Main purpose	Marks documents as archived once they have been successfully archived.
Usable with which source connectors?	SMTP Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed?	Required in SMTP email archiving task routes
Placement in task route	Can appear only before an SC Delete Email task
Produces which metadata?	Task Status
Configuration options	None

This task also performs a consistency check to ensure the document and all attachments have been archived. If one part of the document has not been archived, an exception occurs. Error messages are written to the log files and the document is not marked as archived. Therefore, it is not possible to filter attachments during archiving, for example, to exclude .mp3 files from archiving.

Related reference

"Task status system metadata properties" on page 269

SP Create File:

The SP Create File task creates a temporary local file copy for each Microsoft SharePoint document version to be processed.

Task summary

Table 160. SP Create File task summary

Characteristic	Value
Task name	SP Create File
Main purpose	Creates a temporary local file copy for each SharePoint document version, and enables you to create the document hash that deduplication requires
Usable with which source connectors?	SharePoint Connector

Table 160. SP Create File task summary (continued)

Characteristic	Value
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed	When using document hash-based deduplication, or when downstream tasks such as IBM InfoSphere Classification Module classification or IBM Content Manager tasks need a local file. IBM FileNet P8 tasks require a local file only when you are using document hash-based de-duplication.
Placement in task route	Must appear before repository tasks that require local file copies, or after "SP Get Versions" (if the task route contains the task)
Produces which metadata?	File, SP Create File, Task Status
Configuration options	"Document hash"

Document hash

If you have configured your repository connector to eliminate duplicate files using the *Document Hash* mapping, you must select the **Create document hash** check box to create a unique identifier for each document version.

Restriction: Do not attempt to use deduplication with the Microsoft SharePoint collector if you either:

- Collect document versions, or
- Collect Microsoft Office documents (SharePoint changes the metadata that the collector uses to identify identical documents)

Doing so slows your system and results in no deduplication.

Related reference

"P8 4.x Create Document" on page 543

"CM 8.x Create Document" on page 506

"File system metadata properties" on page 256

"Task status system metadata properties" on page 269

"SP Create File system metadata properties" on page 268

SP Get Versions:

The SP Get Versions task sets the number of versions of each Microsoft SharePoint document to retrieve for processing.

Task summary

Table 161. SP Get Versions task summary

Characteristic	Value
Task name	SP Get Versions
Main purpose	Specifies the number of SharePoint document versions to retrieve
Usable with which source connectors?	SharePoint Connector

Table 161. SP Get Versions task summary (continued)

Characteristic	Value
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed	Whenever you want to archive more than the most recent version of any SharePoint document
Placement in task route	Must appear before any target repository tasks; must appear before "SP Create File" on page 567, if used
Produces which metadata?	File, HTTP URL, Re-collection, SP Blog, SP Collection, Task Status
Configuration options	"Versions"

Versions

You can collect any number of document versions:

- **All versions** collects all versions of each SharePoint document
- **[Number of] versions** processes the specified number of versions, beginning with the highest version number. For example, if you set this value to 3 and a document contains 5 versions, the task route processes versions 5, 4, and 3.

  Tip: Selecting **All versions** gives you the best chance to keep version numbers in sync, though several circumstances (such as a user deleting an intermediate version from SharePoint) can push versions out of sync anyway.

If the content type of a document changes during its life cycle, IBM Content Collector applies the content type of the most recent version to the archived document.

Restriction: Do not attempt to use deduplication if your task route contains this task. Doing so slows your system and results in no deduplication.

Related reference

"File system metadata properties" on page 256

"HTTP URL system metadata properties" on page 261

"Re-collection system metadata properties" on page 266

"SP Blog system metadata properties" on page 266

"SP Collection system metadata properties" on page 267

"Task status system metadata properties" on page 269

SP Manage Link:

The SP Manage Link task deletes or updates unresolved migrated document links on the Microsoft SharePoint server, depending on your configuration and the retrievability of linked documents in the target repository. The task deletes all links that point to missing content.

Task summary

FileNet P8: If a SharePoint user deletes the most recent version of a migrated link document from SharePoint, the link breaks and any task route that contains the SP Manage Link task deletes the orphaned link.

Table 162. SP Manage Link task summary

Characteristic	Value
Task name	SP Manage Link
Main purpose	Deletes or updates unresolved migrated document links on the Microsoft SharePoint server
Usable with which source connectors?	SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector
When needed	If you use the *Replace with link* option in your SharePoint archiving task routes, to resolve broken links you must periodically run a link management task route that contains this task
Placement in task route	Must follow a **Confirm Document** task in every link management task route **Important:** Do not add this task to any archiving task route.
Produces which metadata?	Task Status
Configuration options	"Link Update"

Link Update

By default, this task updates a migrated document link to match the Shortcut link that the task receives from the Confirm Document task. Selecting the **Do not update link** option preserves the link in a previous format, enabling you to eliminate links to moved or deleted content without updating any valid links. You might select this option if some of your links use the Content Collector Version 2.1.1 or other format and another of your SharePoint sites or servers still uses them.

Related reference

"Task status system metadata properties" on page 269

SP Post-processing:

The SP Post-processing task determines whether and how archived documents appear in Microsoft SharePoint.

Task summary

Table 163. SP Post-processing task summary

Characteristic	Value
Task name	SP Post-processing
Main purpose	Specifies what remains in Microsoft SharePoint after archiving

Table 163. SP Post-processing task summary (continued)

Characteristic	Value
Usable with which source connectors?	SharePoint Connector
Usable with which target connectors?	IBM FileNet P8 Connector, IBM Content Manager Connector, File System Repository Connector (which does not support the **Replace with link** option)
When needed	Optional, but recommended to prevent duplication of documents in the target repository
Placement in task route	Typically last; must follow any target repository tasks
Produces which metadata?	Task Status
Configuration options	"Document retention in SharePoint" "Version retention in SharePoint" on page 572 "Linking" on page 572

Document retention in SharePoint

You can select one of four options to govern what happens to a document in SharePoint after archiving:

- **Mark as processed** marks a document as processed to prevent future collection but leaves the document otherwise untouched. Users retain their existing access.
- **Lock down** leaves the original document in SharePoint but prevents users from changing it. To grant selected users full privileges on a locked down document, you add users or groups to the list in **SharePoint Collector → Lock Down**.

  See "Microsoft SharePoint collection sources" on page 483.
- **Replace with link** deletes the original document from SharePoint, replacing it with a hyperlink to the archived document. Archiving changes access permissions to link documents, depending on the original permissions:

Permission level in SharePoint	Permission in target repository
Full Control	Read only (administrators retain write and delete permissions)
Contribute or Design	Read only
Read	Read only
None	None

  Restriction: Do not select **Replace with link** if you are collecting blog posts or any SharePoint library, such as Slide or Wiki, that does not support multiple content types. The attempt to replace blog posts with links in SharePoint fails and the content remains unchanged. You can use this option with wiki content, but the wiki content remains in SharePoint, giving this option the same effect as **Mark as processed**.
- **Delete** deletes the original document from SharePoint and leaves no trace.

If you select **Lock down** or **Replace with link**, do not delete the Read permission level in SharePoint.

Version retention in SharePoint

If you select either the **Mark as processed** or **Lock down** options, you can retain all the versions of a document or only the most recent version. The **Most recent version only** option can save disk space and avoid performance decreases without removing the document from the SharePoint server.

Exception: If the most recent version is a Minor Version, and a Major Version exists, the **Most recent version only** option retains both document versions.

These options do not apply to the **Replace with link** or **Delete** retention options.

Linking

The **Replace with link** option replaces the SharePoint document with a document whose content is the link and content type is *Migrated Document Link*. This action:

- Sets the *Migrated* column value to Yes, which effectively marks the document as processed
- Copies the original values for *Created at*, *Created by*, *Last modified at*, and *Last modified* to the new link document
- Copies the list of users and groups who can access the original to the new link document, changing their permissions from Read-Write or other permission set to Read-only
- Copies all other metadata except for inapplicable SharePoint properties such as *Order*, *Image size*, *Thumbnail*, *Preview*, and *[Image] File type*
- Does not replace aspx files with links, giving this option the same effect as **Mark as processed**

Tip: Content that you retrieve by a link opens according to your client browser configuration and your installed applications. If a file does not open as expected, you might need to:

- install an application that can open the file, and
- add the application to your browser's list of helper applications

Related reference

"Microsoft SharePoint collection sources" on page 483

"Task status system metadata properties" on page 269

Using the set-up tools

The IBM Content Collector set-up tools enable you to configure repositories in Content Manager (creating item types and enabling existing item types) and IBM FileNet P8, enable Domino templates for IBM Content Collector archiving functions, and configure the Lotus client environment if IBM Content Collector and the Lotus Notes runtime environment are on different computers.

The set-up tools can be called up by selecting **All Programs** → **IBM Content Collector** → **Set-up Tools** or **Tools** → **Set-up Tools** in the IBM Content Collector Configuration Manager.

Related tasks

"Performing the initial configuration" on page 69

Configuring an IBM Content Manager repository

All documents archived by using IBM Content Collector are stored in an item type in IBM Content Manager.

You must have at least one IBM Content Manager item type for each source system that you configure in IBM Content Collector.

The first time you install Content Collector and run the Content Collector initial configuration, the configuration wizard will guide you through the steps required to create an item type for each of the source systems that was selected during the installation. For example, if you selected Microsoft Exchange, File System and Microsoft SharePoint, you will be guided you through the process of creating item types for these three source systems.

Important: All item types that are created during the Content Collector initial configuration are IBM Content Manager resource item types. For File System and Microsoft SharePoint source systems, Content Collector also supports IBM Content Manager document item types. However, you must create document item types manually in IBM Content Manager. Furthermore, these document item types cannot be processed by the IBM Content Collector indexer for text search or by the standard IBM Content Manager indexer.

All additional resource item types that are required after the first installation of Content Collector must be created by using the Content Collector set-up tools.

Item type creation varies depending on the selected source system and the item type data model for email source systems:

- If you are upgrading to IBM Content Collector V2.2 and have been using the bundled data model for email item types, you can select both data models for archiving from Lotus Domino or Microsoft Exchange mailboxes:

Table 164. Item type creation criteria for an upgrade

Source system	Data model
Email source systems: • Lotus Domino • Microsoft Exchange	You can select which email data model to use. IBM Content Collector supports two data models: • The **compound data model** that archives documents in parts. This model exists since IBM Content Collector V2.1.1. If selected, two item types are created when the item type is configured; an *email item type* for the email body and meta data, and an *attachment item type* for the embedded email attachments. An attachment item type can be shared by more than one email item type. This division greatly improves attachment deduplication. • The **bundled data model** that archives all data, including embedded attachments, as one unit in a single item type. This model was used in all Content Collector versions before V2.1.1. If you are upgrading from an earlier Content Collector version that created items types of type bundled, you can select to create item types of both data models. However, because of the improved attachment deduplication that is supported using the new compound data model, you are strongly encouraged to use the compound data model when you create new item types for archiving and indexing email. This is especially recommended if you split item types; enable the existing item types for processing by the IBM Content Collector indexer for text search and, when the next item type is due to be created, change to the new model.
Email through SMTP	No data model selection is permitted. The compound data model is used.
Microsoft SharePoint	No data model selection is permitted. The compound data model is used.
File System	No data model selection is permitted. The compound data model is used.

- If this is a first installation of IBM Content Collector, or if you are upgrading from a V2.1.1.x of IBM Content Collector, you cannot select the email data model:

Table 165. Item type creation criteria if Content Collector is installed the first time

Source system	Data model
Email source systems: • Lotus Domino • Microsoft Exchange • Email through SMTP	No data model selection is permitted. All email item types use the compound data model.
Microsoft SharePoint	No data model selection is permitted. The compound data model is used.
File System	No data model selection is permitted. The compound data model is used.

All newly created resource item types (compound and bundled) are automatically enabled for processing by the IBM Content Collector indexer for text search. These enabled item types can only be processed by the IBM Content Collector indexer for

text search. These item types cannot be processed by the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit.

To create an IBM Content Manager resource item type:

1. Plan your item type creation.
2. Select **Set-up Tools → CM Repository Configuration**
3. Select the source system for which you want to create an item type.

Source system	Task
Email source systems: • Lotus Domino • Microsoft Exchange • SMTP	1. Select the IBM Content Manager server. 2. Enter the administrator ID and password. 3. If you want to search document content, select to enable the item type for text search. If the item is enabled for text search, enter the path where the IBM Content Collector Text Search Support package was installed. 4. Select the data model for the item type. If you are new to IBM Content Collector, you cannot select a data model as only compound item types are created. 5. Enter the name of the email item type. 6. Select to create a new attachment item type and enter a name, or select to use an existing attachment item type and select which one to use. 7. If the item type is enabled for text search, enter the relative path to the index directory and the index working directory.
Microsoft SharePoint, or File System	1. Select the IBM Content Manager server. 2. Enter the administrator ID and password. 3. If you want to search document content, select to enable the item type for text search. If the item is enabled for text search, enter the path where the IBM Content Collector Text Search Support package was installed. 4. Select to create an item type. Enter the name of the item type. All File System , Microsoft SharePoint, and SMTP item types are automatically created as compound item types. 5. If the item type is enabled for text search, enter the relative path to the index directory and the index working directory.

Creating an IBM Content Manager repository for documents from Notes applications

Documents that are not mail documents in Lotus Notes or Lotus Domino databases and are archived in IBM Content Manager using IBM Content Collector must be stored in an item type.

When you use the IBM Content Collector set-up tools to create an item type and select Lotus Domino as the source system, you can only create an email item type. The information in a Notes application however, for example, in a TeamRoom, in project libraries, in address books, and in calendars, might differ significantly from mails and requires a different item type in which to store this information. At the

same time, to fully exploit the deduplication feature in IBM Content Collector that ensures that only one copy of a document is archived, this item type must resemble the email item created using the set-up tools.

To create an item type for documents archived in Notes applications:

1. Create an email item type for Lotus Notes but ensure that you give it a distinct name that closely matches your application. Make sure to use the same attachment item type for this item type as for email archiving. Using a different attachment type is not supported.
2. Copy the item type by using the IBM Content Manager administration client. You must copy the item type so that you can remove attributes from it.
3. Open the copied item type and remove all attributes that do not match the Notes application items that you want to map. To save the item type, you must rename the references to child components and delete the foreign key mapping. The foreign key is created again when you save the item type.
4. Define new attributes that you want to map from your Notes application and add these attributes to the item type.

Table 166. Notes field types and matching IBM Content Manager attribute types

Lotus Notes field type	IBM Content Manager attribute type	Comment
Text	Character Variable character	Multiple values are concatenated by the IBM Content Manager connector. Make sure that the maximum length of the concatenated Notes item value is shorter or equal to the length defined for the attribute. Alternatively, you can use a RegEx to cut off the value at a maximum length.
Number	Short integer Long integer Decimal Double	Only single values are supported
Date only	Date	A date only value must be mapped to a Date attribute
Time only	Time	A time only value must be mapped to a Time attribute
Date and time	Time stamp	A date and time value with a date and a time component must be mapped to a Time stamp attribute.

5. Save the item type.

Creating an IBM Content Manager document item type for File System or Microsoft SharePoint documents

Documents that are collected from File System or from Microsoft SharePoint sites and are archived in IBM Content Manager using IBM Content Collector must be stored in an item type. This can either be an IBM Content Manager resource item type or an IBM Content Manager document item type. The default item type created in Content Collector is the resource item type.

If you intend to use document item types for File System or Microsoft SharePoint documents, you must consider the following constraints:

- If you use document item types in any of the File System or Microsoft SharePoint sample task routes, the document item types must match the structure of the resource item types used in IBM Content Collector. To make this possible, you must either create a resource item type in Content Collector and then copy the item type, rename it, and update its properties in IBM Content Manager or manually create a document model item type with the same properties as an IBM Content Collector resource item type. Depending on the tasks performed in the task routes, ensure that you add the required properties to the item type, for example, for File System deduplication add the ICCHash property to the item type.
- If you have your own custom solution and are not using the Content Collector task routes, create the document item types directly in IBM Content Manager. However, you must include an ICMBASE part when you create the document item types to enable these item types to be used by Content Collector.

To create an IBM Content Manager document item type to use in one of the sample Content Collector task routes:

1. Select **Set-up Tools → CM Repository Configuration** to create a resource item type.
2. Select the source system for which you want to create an item type.
3. Select the IBM Content Manager server.
4. Enter the administrator ID and password.
5. Do not enable the item type for text search. Document item types cannot be indexed by the IBM Content Collector indexer for text search nor by the built-in IBM Content Manager indexer.
6. In the IBM Content Manager administration client:
   a. Select the item type you just created, right-click, and select **Copy**.
   b. On the Definition page, enter a new name for the item type and change the item type classification to **Document**.
   c. On the Document Management page, add document parts. The only document part that is required and used by Content Collector is the base document part ICMBASE. In the Define Document Management Relations Set window, you must select an Access Control List. The default that is used by Content Collector is DocRouteACL.
7. Save the item type.

Configuring the Domino environment for Content Collector

If IBM Content Collector and the Lotus Notes runtime environment are on different computers, you must configure the Content Collector server and the task runtime environment.

This includes creating a server .ini file and scripts to run the server, the tasks and the crawler. To configure the Domino environment:

1. Select **Set-up Tools → Lotus Runtime Configuration**
2. Enter the Domino home server.
3. Browse for the administrator ID file and enter a valid password.

   The selected ID must have sufficient privileges to change templates. Typically, this means Manager access rights.
4. Browse for the Lotus Notes ID that you want to use for the Content Collector processes and enter a password.

   The selected ID must be Editor with the additional rights to :

- Delete documents
- Create personal folders and views
- Create shared folders and views
- Replicate or copy documents

The Lotus Notes ID must have the permission to change and delete documents in mail boxes.

Enabling a Domino template for archiving

To be able to select IBM Content Collector archiving functions in a Domino mailbox or a Domino application, you must enable an existing template to include this functionality.

To enable a template for archiving:

1. Enter the Domino home server, in the format *server_name/domain*. Example value: `myServer/Organization`.
2. Browse for the ID file that you want to use for administrator functions and enter a valid password.

   The administrator ID is used by IBM Content Collector to create the runtime environment and enable using the Lotus Domino template and iNotes form with IBM Content Collector functionality.

   The selected ID must have sufficient privileges to change templates. Typically, this means Manager access rights.

   To enable templates remotely, the administrator ID must be Manager on the mail template, the iNotes (Domino Web Access) template, (there is no standalone iNotes template starting with Domino V8) and the forms database.

   The following rights must be configured irrespective of whether the administrator ID or whether another ID is used to enable the IBM Content Collector template for iNotes (Domino Web Access):
   - The rights to sign or run unrestricted methods and operations
   - The rights to sign or run restricted LotusScript/Java agents
   - The user needs to be editor with remove document access at least in order to use the IBM Content Collector functions on iNotes.
3. Browse for the Lotus Notes ID file that you want to use for the Content Collector processes and enter a password.
4. Enter the location of the Domino template you want to enable. Note that there is a limitation if you are using a multilingual mail database.
5. Customize the template dependent on the application type.

Application type	Task
Email	1. Enter the template name. 2. If you want to include the same IBM Content Collector archiving functions for browser-based access, enter: a. The template name b. The name of the forms database **Important:** For Lotus iNotes, an Extension Forms File named `Forms85_x.nsf` must exist in the `iNotes` directory on the Lotus Domino server. If the file does not exist, you must create one before you can enable the Content Collector features on Lotus iNotes. For information about how to create an Extensions Forms File, see the topic about customizing the look of Lotus iNotes in the IBM Lotus Domino and Notes information center at http://publib.boulder.ibm.com/infocenter/domhelp/v8r0/index.jsp. 3. Select which IBM Content Collector elements to add or remove from the template for interactive archiving. 4. Enter the name of the Content Collector Actions submenu. 5. Check if you want to add the Content Collector view to the template and enter a view name.
Other Domino applications	1. Enter the template name. 2. Select to add the Restore function to the template for interactive restore. 3. Enter the name of the Content Collector Actions submenu.

Enabling an IBM Content Manager repository for processing by the indexer for text search

All existing item types created in earlier Content Collector versions before V2.1.1 must be enabled for processing by the indexer for text search before indexing can be run on these item types.

Important: Before you can use an item that was created before Content Collector V2.1.1 in a Content Collector version after V2.1.1, the item type must be enabled for processing by the indexer for text search. Before you enable the existing item type, you must stop all further archiving to this item type and check that all items in this item type that were archived were indexed by using the fast indexer or the standard IBM Content Manager indexer. The Net Search Extender log table must be empty.

If you choose not to enable an existing item type for processing by the indexer for text search, you can only process it using the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit. You cannot process the item type by using the IBM Content Collector indexer for text search.

Only existing Content Collector item types can be enabled using the Content Collector Set-up Tools. To enable IBM CommonStore item types for processing by the indexer for text search, see “Enabling IBM CommonStore item types for processing by the indexer for text search” on page 589.

To enable existing IBM Content Manager item types for processing by the IBM Content Collector indexer for text search:

1. Select **Set-up Tools → CM Repository Enablement**
2. Select the IBM Content Manager sever on which the item types are located and enter the administrator ID and password.
3. Select which item types you want to enable for text search from the list of loaded item types.

   - Item types marked by can be selected.
   - Item types marked by are already enabled and cannot be selected again.
   - Item types marked by cannot be enabled because the Net Search Extender log table for the item type is not empty. This means that the item type was not completely indexed by the fast indexer or the standard IBM Content Manager indexer. Before you can enable the item type for processing by using the IBM Content Collector indexer for text search, you must complete indexing the item type using the fast indexer or the standard IBM Content Manager indexer.

   Both single and multiple selection are supported.
4. To enable the selected item types, click **Enable Selected Item Types**.
5. To refresh the list of item types, click **Reload Item Types**.

Important:

- Enabled item types can only be processed by the IBM Content Collector indexer for text search. These item types cannot be processed by the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit.
- After you have enabled an existing item type for processing by the indexer for text search, you can continue using the same full text index that you used for the item type with the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit. You do not have to recreate the full text index.
- The IBM Content Collector indexer for text search can coexist with the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit on the same IBM Content Manager machine and can run at the same time, but not on the same item type.

Configuring an IBM FileNet P8 repository

All documents archived using IBM Content Collector are stored in an object store in IBM FileNet P8.

Because you can only use a single content-based retrieval configuration for an object store, the store must be dedicated to archiving using IBM Content Collector. It must be empty so that it can be configured for IBM Content Collector and cannot be used for other purposes.

It is not necessary to configure different object stores for each source system in IBM FileNet P8. For example, you can configure one object store for email, File System and SharePoint documents. This single object store approach has advantages with respect to deduplication as in a single object store duplicate files from various sources (email attachments and documents from File System or SharePoint) can be detected, thus reducing the storage requirements. However, as IBM Content Collector uses a customized indexing configuration to allow for the

fine-grained indexing of email documents, custom indexing configurations for other documents (like XML documents from a File System source) may be limited or require manual integration steps.

To configure an object store:

1. Select **Set-up Tools → P8 Repository Configuration**
2. Enter the URL to the Content Engine.
3. Enter the Content Engine administrator ID and password.
4. Enter the FileNet P8 user ID and password with which to log on to the object store. **Validate** this information.
5. Click **Retrieve** to fill the list with all of the available object stores. Select an object store to configure. The FileNet P8 domain name is displayed.

Configuring an IBM FileNet P8 repository for documents from Notes applications

Documents that are not mail documents in Lotus Notes or Lotus Domino databases and are archived in IBM FileNet P8 using IBM Content Collector must be stored in an object store.

When you use the IBM Content Collector set-up tools to configure an object store and select Lotus Domino as the source system, you can only configure the object store for email. The information in Notes applications however, for example, in team rooms, project libraries, address books, and calendars, might differ significantly from that in mails and requires a different object store configuration.

To configure an object store for documents archived in Notes applications:

1. Open the object store using the FileNet P8 Enterprise Manager.
2. Create a document class for the documents from Notes applications.

   Important: Add the ICCAttachmentCorrelationKeys property.
3. Define new properties that you want to map from your Notes application and add these properties to the object class.

Table 167. Notes field types and matching FileNet P8 property types

Lotus Notes field type	FileNet P8 attribute type	Comment
Text	Character Variable character	Make sure that the maximum length of the concatenated Notes item value is shorter or equal to the length defined for the attribute. Alternatively, you can use a RegEx to cut off the value at a maximum length.
Number	Short integer Float	
Date and time	DateTime	Lotus Notes interprets a time value of 00:00:00 as not set.
Boolean	Boolean	

Enabling the access to archived data

To enable access to documents that were archived by IBM Content Collector, you must configure your system accordingly. Search in Content Collector is available for email documents archived in both IBM FileNet P8 and IBM Content Manager repositories. Viewing and restoring in Content Collector is available for all documents that were archived from a mailbox, a personal storage file (PST file), a local Notes archive (NSF file), a file system, or from Microsoft SharePoint into FileNet P8 or IBM Content Manager repositories.

In IBM Content Collector, you can search email documents that were archived from your mailbox, a personal storage file (PST file), or the local database into a central repository. Using the email search function of Content Collector's web client, you can search documents by date, sender, recipients, subject, contents, attachment contents, or any combination of these. You can also customize the search function to make other attributes available for search. You can then preview the documents that met your search criteria and restore them to your mailbox. Although it is not possible in Content Collector to search documents that were archived from other sources, for example, Lotus Notes applications, a file system, or Microsoft SharePoint, it is possible to view those documents. However, you can search these documents with the means provided by the respective content server or by using IBM eDiscovery Manager. Before you can search, view, or restore archived documents with Content Collector, you have to adapt the configuration settings for accessing archived data.

For searching, previewing, or restoring email that was archived with IBM CommonStore for Exchange Server or IBM CommonStore for Lotus Domino, however, there are several restrictions:

- If you configure Content Collector to use both IBM Content Manager and FileNet P8, search, preview, and restore of archive documents using Content Collector can only be carried out on one of the content management systems, not on both. However, you can configure to use more than one repository on a content management system, and search, preview, and restore across these repositories simultaneously.
- You cannot search for email that was archived to an IBM Content Manager OnDemand repository.
- You cannot search for email that was archived to an IBM Tivoli Storage Manager repository.
- To be able to search for email that was archived to an IBM Content Manager 8 repository you must create a specific collection as described in “Enabling the access to archived data for legacy item types” on page 148.
- You can preview and restore email in the search result only if the item type was created with the CommonStore archiving model BUNDLED. For the archiving model BUNDLED, there are additional restrictions if the CommonStore archiving type was not **Entire**. If the archiving type was **Attachment** or **Component**, you can search for the email but you cannot preview or restore the email.

Remember: The number of users running the IBM Content Collector email search application at the same time cannot be higher than the value of DB2_APP_CONNECTION.

Enabling an IBM Content Manager repository for search

The IBM Content Collector indexer for text search processes documents that are archived in an IBM Content Manager repository to enable subsequent searching in these documents.

The indexer for text search is an IBM Content Collector component called IBM Content Collector Text Search Support which must be installed on the IBM Content Manager server before the IBM Content Collector Server code is installed. For details on requirements and the correct order in which components must be installed, see the topics under installing.

The indexer for text search replaces and combines the text search user exit and fast indexer modules used in the Content Collector versions before V2.1.1. It runs document retrieval and content conversion processes in parallel which significantly speeds up the process of full text indexing.

Related reference

"Additional prerequisites and restrictions" on page 31

The indexer for text search

The IBM Content Collector indexer for text search prepares all documents for indexing, and sets up the indexing engine, Net Search Extender, to perform indexing.

All documents archived in IBM Content Manager using IBM Content Collector are stored in document collections called item types which must be text-searchable. Documents in these item types are marked for indexing the moment they are stored in the item type.

Before processing can begin on an item type, the following tasks are performed to set up the system to support the indexer for text search. These tasks are carried out during the initial configuration the first time you install IBM Content Collector, or after installation, by using the IBM Content Collector set-up tools. These tasks include:

- Creating new item types or reconfiguring existing item types to enable processing using the indexer for text search in IBM Content Collector
- Adjusting the indexer for text search configuration that is used by the item type. This includes the configuration of the indexing engine called IBM Net Search Extender.
- Preparing the item type for the indexer for text search (formerly the FetchContentSetup process that was run separately). This is now done automatically and includes for each item type:
  - Creating a database table for the indexing open tasks

    The table of open tasks contains all items that have not been processed and indexed.
  - Creating a database table for the indexing completed tasks

    The table of completed tasks contains both items that were processed and indexed successfully as well as items that could not be processed for what ever reason.
  - Creating triggers that create rows in the open tasks table for each item archived

After the set-up tasks have been performed and if the item type contains documents, indexing can begin.

This indexing process called the `afuIndexer` process (formerly the `IndexDocuments` process) retrieves the archived documents from the Resource Manager for a specific item type and then extracts all textual information stored in each document. The textual information for each document is stored in a temporary XML file that is read by the indexing engine Net Search Extender with the help of a user defined function (UDF) called `AFUFetchFile`. Net Search Extender digests the content of each XML file and creates the full text index.

Briefly, the `afuIndexer` steps comprise:

- Reading and checking the indexer configuration including the item type specific settings
- Checking the available disk space and ensuring exclusive access rights to the index
- Automatically determining the batch size of the items (documents) to process during an indexing run
- Creating a new worker process with a pool of worker threads to process the batch of items
- Reading the items from the table of open tasks
- Retrieving documents from the Resource Manager
- Extracting all textual information stored in the items and writing this information to XML files. All embedded attachments in email and all archived SharePoint and File System documents are parsed using the Oracle/Stellent Outside In Technology filters.
- Triggering Net Search Extender to digest the XML files using the `AfuFetchfile` UDF and adding the file content to the full text index
- Moving all indexed items from the table of open tasks to the table of completed tasks

The main indexing steps

You can start indexing items in the item type immediately after the item type configuration has been created and as soon as the batch of items to index exceeds the defined minimum size required to start indexing.

The main indexing steps include:

1. Creating a new item type or enabling existing item types for processing by the indexer for text search. Skip this step if you have already created item types or have enabled existing item types for processing by the indexer for text search.
2. Creating a configuration for the item type.
3. Indexing items.
4. Checking the indexer process return code.
5. Checking the status of indexed items.

Configuring an IBM Content Manager repository:

All documents archived by using IBM Content Collector are stored in an item type in IBM Content Manager.

You must have at least one IBM Content Manager item type for each source system that you configure in IBM Content Collector.

The first time you install Content Collector and run the Content Collector initial configuration, the configuration wizard will guide you through the steps required to create an item type for each of the source systems that was selected during the

installation. For example, if you selected Microsoft Exchange, File System and Microsoft SharePoint, you will be guided you through the process of creating item types for these three source systems.

Important: All item types that are created during the Content Collector initial configuration are IBM Content Manager resource item types. For File System and Microsoft SharePoint source systems, Content Collector also supports IBM Content Manager document item types. However, you must create document item types manually in IBM Content Manager. Furthermore, these document item types cannot be processed by the IBM Content Collector indexer for text search or by the standard IBM Content Manager indexer.

All additional resource item types that are required after the first installation of Content Collector must be created by using the Content Collector set-up tools.

Item type creation varies depending on the selected source system and the item type data model for email source systems:

- If you are upgrading to IBM Content Collector V2.2 and have been using the bundled data model for email item types, you can select both data models for archiving from Lotus Domino or Microsoft Exchange mailboxes:

Table 168. Item type creation criteria for an upgrade

Source system	Data model
Email source systems: • Lotus Domino • Microsoft Exchange	You can select which email data model to use. IBM Content Collector supports two data models: • The **compound data model** that archives documents in parts. This model exists since IBM Content Collector V2.1.1. If selected, two item types are created when the item type is configured; an *email item type* for the email body and meta data, and an *attachment item type* for the embedded email attachments. An attachment item type can be shared by more than one email item type. This division greatly improves attachment deduplication. • The **bundled data model** that archives all data, including embedded attachments, as one unit in a single item type. This model was used in all Content Collector versions before V2.1.1. If you are upgrading from an earlier Content Collector version that created items types of type bundled, you can select to create item types of both data models. However, because of the improved attachment deduplication that is supported using the new compound data model, you are strongly encouraged to use the compound data model when you create new item types for archiving and indexing email. This is especially recommended if you split item types; enable the existing item types for processing by the IBM Content Collector indexer for text search and, when the next item type is due to be created, change to the new model.
Email through SMTP	No data model selection is permitted. The compound data model is used.
Microsoft SharePoint	No data model selection is permitted. The compound data model is used.
File System	No data model selection is permitted. The compound data model is used.

- If this is a first installation of IBM Content Collector, or if you are upgrading from a V2.1.1.x of IBM Content Collector, you cannot select the email data model:

Table 169. Item type creation criteria if Content Collector is installed the first time

Source system	Data model
Email source systems: • Lotus Domino • Microsoft Exchange • Email through SMTP	No data model selection is permitted. All email item types use the compound data model.
Microsoft SharePoint	No data model selection is permitted. The compound data model is used.
File System	No data model selection is permitted. The compound data model is used.

All newly created resource item types (compound and bundled) are automatically enabled for processing by the IBM Content Collector indexer for text search. These enabled item types can only be processed by the IBM Content Collector indexer for text search. These item types cannot be processed by the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit.

To create an IBM Content Manager resource item type:

1. Plan your item type creation.
2. Select **Set-up Tools → CM Repository Configuration**
3. Select the source system for which you want to create an item type.

Source system	Task
Email source systems: • Lotus Domino • Microsoft Exchange • SMTP	1. Select the IBM Content Manager server. 2. Enter the administrator ID and password. 3. If you want to search document content, select to enable the item type for text search. If the item is enabled for text search, enter the path where the IBM Content Collector Text Search Support package was installed. 4. Select the data model for the item type. If you are new to IBM Content Collector, you cannot select a data model as only compound item types are created. 5. Enter the name of the email item type. 6. Select to create a new attachment item type and enter a name, or select to use an existing attachment item type and select which one to use. 7. If the item type is enabled for text search, enter the relative path to the index directory and the index working directory.

Source system	Task
Microsoft SharePoint, or File System	1. Select the IBM Content Manager server. 2. Enter the administrator ID and password. 3. If you want to search document content, select to enable the item type for text search. If the item is enabled for text search, enter the path where the IBM Content Collector Text Search Support package was installed. 4. Select to create an item type. Enter the name of the item type. All File System , Microsoft SharePoint, and SMTP item types are automatically created as compound item types. 5. If the item type is enabled for text search, enter the relative path to the index directory and the index working directory.

Planning your item type creation:

Before you create an item type, you must decide if you want to keep the default attribute description of an item type defined by IBM Content Collector or if you want to add other custom attributes and additional IBM Content Manager attributes to the item type. The attribute description cannot be changed after the item type has been created without recreating the full text index.

The attributes represent the content of the document that you can subsequently query when searching for the document.

You can extend the Net Search Extender full text index by adding additional information from documents or from IBM Content Manager attributes. By adding the field definitions that contain this information to configuration files, the data in these fields is added to the index when the index is built. This additional information in the index increases the search scope which means that information from more document fields can be used to find archived documents.

The document attributes are defined in a model file. Three model files exist and can be used by Net Search Extender:

- The `afu_mail_entire.xml` file for bundled Content Collector item types
- The `afu_mail_entire_cmp.xml` file for compound Content Collector item types
- The `cs_mail_entire.xml` for CommonStore item types for the document model BUNDLED and the archiving type ENTIRE

All model files are located in the subdirectory `model` where you installed the component for text search support in IBM Content Manager. These model files contain the complete set of field definitions required for text search indexing by IBM Content Collector.

For Lotus Domino documents, you can add attributes for fields of the following type:

- Any text field
- Any text-list field
- Any RFC822 text field

For Exchange documents, all user-defined fields can be indexed.

Additionally, all attributes defined in IBM Content Manager can be added to the index for all document types.

Important:

1. If you define additional attributes to add to the index, you must add the same definitions to the model file and to the indexer configuration file for the item type.
2. The indexer configuration file is created when you run `afuConfig`. Before you can begin indexing items in an item type, you must update the indexer configuration file manually to include the same attribute definitions that you added to the model file. When the indexer process is started on the item type, the process checks that the model file and the indexer configuration file are consistent.

Before you create an item type:

1. Determine which additional user data to add as custom attributes to the model file. For example, you might wish to store the company record number to use in later search queries.
2. Add attributes to the model file that corresponds to the data model of the item type you want to create.

Enabling an IBM Content Manager repository for processing by the indexer for text search:

All existing item types created in earlier Content Collector versions before V2.1.1 must be enabled for processing by the indexer for text search before indexing can be run on these item types.

Important: Before you can use an item that was created before Content Collector V2.1.1 in a Content Collector version after V2.1.1, the item type must be enabled for processing by the indexer for text search. Before you enable the existing item type, you must stop all further archiving to this item type and check that all items in this item type that were archived were indexed by using the fast indexer or the standard IBM Content Manager indexer. The Net Search Extender log table must be empty.

If you choose not to enable an existing item type for processing by the indexer for text search, you can only process it using the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit. You cannot process the item type by using the IBM Content Collector indexer for text search.

Only existing Content Collector item types can be enabled using the Content Collector Set-up Tools. To enable IBM CommonStore item types for processing by the indexer for text search, see “Enabling IBM CommonStore item types for processing by the indexer for text search” on page 589.

To enable existing IBM Content Manager item types for processing by the IBM Content Collector indexer for text search:

1. Select **Set-up Tools → CM Repository Enablement**
2. Select the IBM Content Manager sever on which the item types are located and enter the administrator ID and password.
3. Select which item types you want to enable for text search from the list of loaded item types.

- Item types marked by can be selected.
- Item types marked by are already enabled and cannot be selected again.
- Item types marked by cannot be enabled because the Net Search Extender log table for the item type is not empty. This means that the item type was not completely indexed by the fast indexer or the standard IBM Content Manager indexer. Before you can enable the item type for processing by using the IBM Content Collector indexer for text search, you must complete indexing the item type using the fast indexer or the standard IBM Content Manager indexer.

Both single and multiple selection are supported.

4. To enable the selected item types, click **Enable Selected Item Types**.
5. To refresh the list of item types, click **Reload Item Types**.

Important:

- Enabled item types can only be processed by the IBM Content Collector indexer for text search. These item types cannot be processed by the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit.
- After you have enabled an existing item type for processing by the indexer for text search, you can continue using the same full text index that you used for the item type with the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit. You do not have to recreate the full text index.
- The IBM Content Collector indexer for text search can coexist with the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit on the same IBM Content Manager machine and can run at the same time, but not on the same item type.

Enabling IBM CommonStore item types for processing by the indexer for text search:

Items types that were created by using IBM CommonStore must be enabled for processing by the Content Collector indexer for text search.

Only CommonStore item types for the document model BUNDLED and the archiving type ENTIRE (BRI) can be enabled. All other item types must be processed by using the fast indexer or the standard IBM Content Manager indexer with the IBM Text Search user exit.

Important: Before you can use a CommonStore item in Content Collector, the item type must be enabled for processing by the indexer for text search. Before you enable the CommonStore item type, you must stop all further archiving to this item type and check that all items in this item type that were archived were indexed by using the fast indexer or the standard IBM Content Manager indexer. Alternatively, you can recreate the entire index by using `afuRecreateIndex`.

CommonStore item types cannot be enabled using the Content Collector set-up tools. To enable existing CommonStore item types for processing by the IBM Content Collector indexer for text search:

- On UNIX and Linux, run:

  ```
  afuEnableItemType.sh
  ```

- On Windows run:

  ```
  afuEnableItemType.cmd
  ```

Note: When you run the `afuEnableItemType` process:

- All command line arguments are case sensitive.
- The command line arguments can appear in any order.
- The item type name is not case sensitive.

If you run `afuEnableItemType` without specifying an argument, or if you run `afuEnableItemType` with the argument `-?` the following is displayed:

```
afuEnableItemType: Usage:
  [-?|-h|-copyright]
  -it <itemtype>
  -database <database>
  [-user <userID|@environmentVariable|:filename>]
  [-password <password|@environmentVariable|:filename>]
  [-schema <schema>]
  [-tablespace <tablespace>]
  [-udftransferbuffersize <buffer size in MB>]
  [-verbose] [-showsql] [-trace] [-simulate]
  [-quiet]
```

Where:

-? Lists all of the arguments that you can pass to the indexer tool `afuEnableItemType`.

-h Lists all of the arguments that you can pass to the indexer tool `afuEnableItemType`.

-copyright
Lists the copyright notice and version information.

-it <*itemtype*>
Required.

The Content Manager item type (document collection) that you want to index. Item type names are not case sensitive.

-database <*database*>
Required.

The name of the library server database in which the item type is stored. The name of the database can be specified in two formats which affects the way the indexer tool connects to the database. The database connection is implemented using Java Data Base Connectivity (JDBC) which defines how a Java client may access a database. JDBC features different connection types and IBM Content Collector supports JDBC type 2 and JDBC type 4.

To use a JDBC type 4 connection, the specification for the database name of the argument `-database` must include a server name, a port number, and a database name in the following format `<databaseserver>:<portnumber>/<database name>`, for example `-database localhost:50000/icmnlsdb`. JDBC type 4 is the recommended database access method but always requires fully specifying the user ID and password.

To use a JDBC type 2 connection, the specification for the database name of the argument `-database` must name an existing DB2 connection name, in other words, the database name must be cataloged in DB2, for example `-database icmnlsdb`. As JDBC type 2 uses an existing database connection, no user ID and password need to be specified. However, a native 32-bit version of the DB2 JDBC Type 2 driver (shared library) must be available and defined in the class path for using the JDBC type 2 connection.

-user <*userID* | *@environmentVariable* | *:filename*>
Optional.

Provide a user ID. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a user ID:

1. Specify the user ID directly on the command line
2. Specify the name of an environment variable from which the user ID is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_USERID=user8

   afuEnableItemType -user @INDEXER_USERID
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuEnableItemType -user :myUserID.txt
   ```

   The text file must contain a single line with the user ID (no quotes and no spaces).

The advantage of using an environment variable or a file is that the user ID is not visible in any process status information.

The user must have access rights to IBM Content Manager.

-password <*userID* | *@environmentVariable* | *:filename*>
Optional.

The password of the user ID with database administrator authority. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a password:

1. Specify the password directly on the command line
2. Specify the name of an environment variable from which the password is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_PASSWORD=12345678

   afuEnableItemType -password @INDEXER_PASSWORD
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuEnableItemType -password :myPassword.txt
   ```

   The text file must contain a single line with password (no quotes and no spaces).

The advantage of using an environment variable or a file is that the password is not visible in any process status information.

Passwords are not logged in a readable plain text format.

-schema <*schema*>
Optional.

The name of the database schema if the schema is different from the current database schema.

-tablespace <*tablespace*>
Optional.

Enables specifying which table space to use for the tables used by the indexer process. If this argument is not passed, the default DB2 table space is used.

-udftransferbuffersize <*buffer size in MB*>
Optional.

Enables specifying the size of the buffer which the `AFUFetchFile` UDF uses to load the temporary XML files for access by Net Search Extender. The default transfer buffer size is 32 MB. For more detail, see "Changing the default transfer buffer size of the AFUFetchFile UDF" on page 640.

-verbose
Optional.

Displays detailed processing information on the screen. Using **-verbose** is helpful at the beginning to get detailed information on all processing. If **-verbose** is specified the log files also contains additional information.

-showsql
Optional.

Displays all of the DB2 SQL commands executed by the system on the screen. Using this argument is helpful at the beginning to verify the configuration and execution of all DB2 system processing. If this argument is specified, all of the SQL statements are also added to the log file.

-trace
Optional.

Goes beyond verbose by displaying more detail in log files to support error determination. This is only required for IBM Support and requires the use of the trace property file.

-simulate
Optional.

Runs the create index process but does not update the databases. Use this setting to validate the index configuration.

-quiet
No processing steps are displayed on the console.

Creating an indexer configuration file for the item type:

Before you can begin indexing items in an item type using the indexer for text search, you must create an indexer configuration for the item type. An item type indexer configuration can be spread across more than one configuration file or multiple item types can share a configuration.

If the item type is an email compound item type, only create an indexer configuration for the email item type. Do not create a configuration or try to index an attachment item type. Attachment item types cannot be processed on their own; they are automatically processed as part of the email item type.

An indexer configuration file for an item type is created by calling the `afuConfig` tool or by creating a configuration file manually using your favorite text editor. `afuConfig` automatically recognizes if the item type is a Content Collector or a enabled CommonStore item type and creates a default configuration depending on the data model of the item type. The name of the new configuration file is `<item_type_name>.ini`.

The default configuration contains all of the configuration options required for indexing and does not need to be modified. Changes to configuration options are

only required to fine-tune the indexer process or for debugging and reindexing purposes. Indexing can run using the default values.

The only options that have no defaults because they are user dependent are the following absolute file path values for which you are prompted when you run `afuConfig` the first time:

- **GenXmlDirectory**: this is the directory in which the intermediate XML files which are the input to indexing are stored.

  This value MUST be identical for all item type configurations.
- **GenLogDirectory**: this is the file path in which the indexer log files are stored.

  This value MUST be identical for all item type configurations.
- **GenDumpDirectory**: this is the directory in which the original documents retrieved from the IBM Content Manager Resource Manager are optionally stored.

  It is recommended that this value is identical for all item type configurations.
- **GenWorkingDirectoryForItemTypes**: this is where the working files required by the indexer process are written.

  It is recommended that this value is identical for all item type configurations.

The file path information that you enter is valid for all item type configurations and you are only prompted for this information the first time you create an item type configuration. The `afuConfig` tool checks if the directories that you specified actually exist and if the directory names contain only the supported alphanumeric characters. If the directories do not exist, you are asked whether you want to create the directories.

Important: If you change the **GenLogDirectory** or **GenXmlDirectory** file paths at a later stage, you must update all of the item type configurations as these file paths must be consistent across all item types. The indexer process will automatically register these changes and will create the new directories if necessary.

If the values of **GenLogDirectory** or **GenXmlDirectory** are changed, the `AFUFetchFile` UDF configuration is updated automatically. This is disclosed in a message that might request restarting DB2 so that the changes can become effective.

As these directories are shared by all of the indexer processes that are running at the same time and can become very large, you must ensure that enough disk space is allocated to these directories, in particular to the temporary XML files in the directory specified by the configuration option **GenXMLDirectory**.

To avoid having to maintain many similar item type configurations, it is recommended that you distribute the indexer configuration of an item type across more that one configuration file. For details, see “Changing the default indexer configuration” on page 622.

To create an indexer configuration file for the item type:

- On UNIX and Linux, run

  ```
  afuConfig.sh
  ```
- On Windows, run

  ```
  afuConfig.cmd
  ```

Note: When you run the `afuConfig` process:

- All command line arguments are case sensitive.
- The command line arguments can appear in any order.
- The item type name is not case sensitive.

If you run `afuConfig` and do not specify an argument or enter the command followed by `-?`, the following is displayed:

```
afuConfig: Usage:
  [-?|-h|-copyright]
  [-it <itemtype>]
  -database <database>
  [-user <userID|@environmentVariable|:filename>]
  [-password <password|@environmentVariable|:filename>]
  [-schema <schema>]
  [-verbose] [-showsql] [-trace]
  [-quiet]
```

where:

-? Lists all of the arguments that you can pass to the `afuConfig` tool.

-h Lists all of the arguments that you can pass to the `afuConfig` tool.

-copyright
Lists the copyright notice and version information.

-it *<itemtype>*
Optional.

The name of the IBM Content Manager item type (document collection). If `afuConfig` is called without an item type name, the user is prompted for it.

Important: If your item type is an email compound item type, do not enter the attachment item type. Attachment item types cannot be processed on their own but are automatically processed as part of the email item type.

-database *<database>*
Required.

The name of the library server database in which the item type is stored. The name of the database can be specified in two formats which affects the way the indexer tool connects to the database. The database connection is implemented using Java Data Base Connectivity (JDBC) which defines how a Java client may access a database. JDBC features different connection types and IBM Content Collector supports JDBC type 2 and JDBC type 4.

To use a JDBC type 4 connection, the specification for the database name of the argument `-database` must include a server name, a port number, and a database name in the following format `<databaseserver>:<portnumber>/<database name>`, for example `-database localhost:50000/icmnlsdb`. JDBC type 4 is the recommended database access method but always requires fully specifying the user ID and password.

To use a JDBC type 2 connection, the specification for the database name of the argument `-database` must name an existing DB2 connection name, in other words, the database name must be cataloged in DB2, for example `-database icmnlsdb`. As JDBC type 2 uses an existing database connection, no user ID and password need to be specified. However, a native 32-bit version of the DB2 JDBC Type 2 driver (shared library) must be available and defined in the class path for using the JDBC type 2 connection.

-user <*userID* | *@environmentVariable* | *:filename*>
Optional.

Provide a user ID. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a user ID:

1. Specify the user ID directly on the command line.
2. Specify the name of an environment variable from which the user ID is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_USERID=user8

   afuConfig -user @INDEXER_USERID
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuConfig -user :myUserID.txt
   ```

   The text file must contain a single line with the user ID (no quotes and no spaces).

The advantage of using an environment variable or a file is that the user ID is not visible in any process status information.

The user must have access rights to IBM Content Manager.

-password <*password* | *@environmentVariable* | *:filename*>
Optional.

The password of the user ID with database administrator authority. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a password:

1. Specify the password directly on the command line.
2. Specify the name of an environment variable from which the password is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_PASSWORD=12345678

   afuConfig -password @INDEXER_PASSWORD
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuConfig -password :myPassword.txt
   ```

   The text file must contain a single line with password (no quotes and no spaces).

The advantage of using an environment variable or a file is that the password is not visible in any process status information.

Passwords are not logged in a readable plain text format.

-schema <*schema*>
Optional.

The name of the database schema if the schema is different from the default database schema.

-verbose
Optional.

Displays detailed processing information on the screen. Using **-verbose** is helpful at the beginning to get detailed information on all processing. If **-verbose** is specified the log files also contains additional information.

-showsql
Optional.

Displays all of the DB2 SQL commands executed by the system on the screen. Using this argument is helpful at the beginning to verify the configuration and execution of all DB2 system processing. If this argument is specified, all of the SQL statements are also added to the log file.

-trace
Optional.

Goes beyond verbose by displaying more detail in log files to support error determination. This is only required for IBM Support and requires the use of the trace property file.

-quiet
No processing steps are displayed on the console.

If you made any changes to the item type model files when the item types were created, you must also add these to the indexer configuration file for the item type before you can begin indexing

Add the attribute definitions to the index configuration file.

Related reference

"Indexer configuration options" on page 624

Indexing items in an item type:

If the item type has been enabled for processing by the indexer for text search, and the indexer process and item type have been configured, the indexer process can be started.

The indexer for text search process (formerly the `IndexDocuments` process) retrieves the archived documents from the IBM Content Manager Resource Manager for a specific item type and extracts all textual content from each document. This textual content is stored in a temporary XML file that is read by the indexing engine Net Search Extender with the help of the user defined function (UDF) called `AFUFetchFile`. Net Search Extender digests the content of each XML file and creates the full text index.

You can run the indexer for text search in one of two modes:

- Batch mode: use a scheduler like cron or Windows Task Scheduler to run the indexing process on an item type in a schedule
- Continuous mode: run the indexing process in an endless loop until it is interrupted by an explicit shutdown command

The indexer for text search process can only run on one item type at a time, but multiple item types can be processed simultaneously. For details, see "Processing more than one item type at the same time" on page 648.

If the item type is an email compound item type, only run the indexer for text search on the email item type. Do not index an attachment item type. Attachment item types cannot be processed on their own; they are automatically processed as part of the email item type.

Before you start indexing, ensure that enough disk space is allocated for the temporary XML files in the directory specified by the configuration option **GenXMLDirectory**. The XML directory is shared by all of the indexer processes that are running at the same time for multiple item types and can therefore become very large.

Indexing requires an indexer configuration file for the item type. By default, the indexer tool looks for a default configuration file with a name `<item_type_name>.ini` in the current directory and if found, loads it. If no default configuration file is found and no configuration file is passed as an argument when the indexing process is started, the indexer process will display an error message and processing will stop.

Important:

- The indexer process will only start processing if the number of items to index in the next batch of items is at least the number specified by the configuration option **IdxItemMinBatchSize** or the number specified by the indexer command line argument `-minbatchsize`.

  The default value of **IdxItemMinBatchSize** is 0 which forces the process to automatically calculate the minimum size of the batch of items that is waiting to be processed.
- Only the number of items specified by the configuration option **IdxItemMaxBatchSize** or by the indexer command line argument `-maxbatchsize` is processed each time the indexer is scheduled to run.

  For details on the interaction between **IdxItemMinBatchSize** and **IdxItemMaxBatchSize** see "Constraints when indexing in batch mode" on page 644.

To retrieve, convert, and index archived documents:

- On UNIX and Linux, run:

  `afuIndexer.sh`
- On Windows run:

  `afuIndexer.cmd`

Note: When you run the `afuIndexer` process:

- All command line arguments are case sensitive.
- The command line arguments can appear in any order.
- Use the following delimiters for arguments that can pass more than one value:
  - On Windows, a semicolon (;)
  - On UNIX and Linux machines, a colon (:)
- The item type name is not case sensitive.
- The indexer configuration options passed through an argument are not case sensitive.

The item type and the name of the library server database in which the item type is stored must be passed as mandatory arguments to the indexing process.

If you run `afuIndexer` without specifying an argument, or if you run `afuIndexer` with the argument `-?` the following is displayed:

```
afuIndexer: Usage:
  [-?|-h|-copyright]
  -it <itemtype>
  -database <database>
```

```
[-user <userID|@environmentVariable|:filename>]
[-password <password|@environmentVariable|:filename>]
[-schema <schema>]
[-port <port number>]
[-maxbatchsize <number of items>]
[-minbatchsize <number of items>]
[-continuousmode]
[-reindexdate <yyyymmdd>]
[-reindexwarnings] [-reindexerrors]
[-reindexsevereerrors] [-reindexrc <IDXRC>]
[-reindexitems <itemid>;<...>]
[-debug <itemid>;<...>]
[-nseindexreorg]
[-shutdown complete|force]
[-useconfig <.ini file>;<...>]
[-usesettings "<option=value>;<...>"]
[-exportconfig <.ini file>]
[-listconfig]
[-verbose] [-showsql] [-trace]
[-simulate]
[-quiet]
```

Where:

-? Lists all of the arguments that you can pass to the indexer tool `afuIndexer`.

-h Lists all of the arguments that you can pass to the indexer tool `afuIndexer`.

-copyright
Lists the copyright notice and version information.

-it *<itemtype>*
Required.

The Content Manager item type (document collection) that you want to index. Item type names of both the bundled and compound data models are not case sensitive.

Important: If your item type is an email compound item type, do not enter the attachment item type. Attachment item types cannot be processed on their own but are automatically processed as part of the email item type.

-database *<database>*
Required.

The name of the library server database in which the item type is stored. The name of the database can be specified in two formats which affects the way the indexer tool connects to the database. The database connection is implemented using Java Data Base Connectivity (JDBC) which defines how a Java client may access a database. JDBC features different connection types and IBM Content Collector supports JDBC type 2 and JDBC type 4.

To use a JDBC type 4 connection, the specification for the database name of the argument `-database` must include a server name, a port number, and a database name in the following format `<databaseserver>:<portnumber>/<database name>`, for example `-database localhost:50000/icmnlsdb`. JDBC type 4 is the recommended database access method but always requires fully specifying the user ID and password.

To use a JDBC type 2 connection, the specification for the database name of the argument `-database` must name an existing DB2 connection name, in other words, the database name must be cataloged in DB2, for example `-database icmnlsdb`. As JDBC type 2 uses an existing database connection, no user ID and password need to be specified. However, a native 32-bit version of the DB2

JDBC Type 2 driver (shared library) must be available and defined in the class path for using the JDBC type 2 connection.

-user <*userID* | *@environmentVariable* | *:filename*>
Optional.

Provide a user ID. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a user ID:

1. Specify the user ID directly on the command line.
2. Specify the name of an environment variable from which the user ID is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_USERID=user8

   afuIndexer -user @INDEXER_USERID
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuIndexer -user :myUserID.txt
   ```

   The text file must contain a single line with the user ID (no quotes and no spaces).

The advantage of using an environment variable or a file is that the user ID is not visible in any process status information.

The user must have access rights to update text search indexes. For exact information on what user rights are required to update text search indexes, see http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html for DB2 Net Search Extender V9.5 and http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html for DB2 Net Search Extender V9.7.

-password <*password* | *@environmentVariable* | *:filename*>
Optional.

The password of the user ID with database administrator authority. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a password:

1. Specify the password directly on the command line.
2. Specify the name of an environment variable from which the password is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_PASSWORD=12345678

   afuIndexer -password @INDEXER_PASSWORD
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuIndexer -password :myPassword.txt
   ```

   The text file must contain a single line with password (no quotes and no spaces).

The advantage of using an environment variable or a file is that the password is not visible in any process status information.

Passwords are not logged in a readable plain text format.

-schema <*schema*>
Optional.

The name of the database schema if the schema is different from the current database schema.

-port <*number of port*>
Optional.

The indexer process communicates with the worker process through a port.

The default port number is 4242 if no value is specified by using `-port` or if the environment variable **AFU_INDEXER_RMI_PORT** is not set for the instance.

To process multiple item types at the same time however, each indexer process instance must use a free and unique port number. This can be guaranteed either by passing a unique port number using the command line argument `-port` or by setting the environment variable **AFU_INDEXER_RMI_PORT** for each instance before `afuIndexer` is started.

-maxbatchsize <*number of items*>
Optional.

The maximum number of documents in the item type that are processed during an index run. After the specified number has been processed, Net Search Extender is called. The value overwrites the batch size specified in the configuration option **IdxItemMaxBatchSize** and disables the automatic computation of the value.

-minbatchsize <*number of items*>
Optional.

The minimum number of items in the item type required before the indexer process will begin processing. The value overwrites the minimum batch size specified in the configuration option **IdxItemMinBatchSize** and disables the automatic computation of the value.

-continuousmode
Optional.

The indexer process runs in a loop without stopping until an explicit shutdown request is issued for the item type. For details see "Running `afuIndexer` in continuous mode" on page 646.

The following arguments are all related to reindexing documents. They cannot be used if the indexer is run in debug mode. All reindexing processing can take considerable time to complete. Refer to the topic on reindexing documents under Related topics for details.

-reindexdate <*yyyymmdd*>
Optional.

Reindexes documents that have already been indexed but need to be indexed again. All documents that have an archive date of <*yyyymmdd*> or later will be reindexed.

-reindexwarnings
Optional.

Reindexes all documents that were indexed during a previous indexing run and that resulted in a warning IDXRC with a value of between 1 and 99.

-reindexerrors
Optional.

Reindexes all documents that were indexed during a previous indexing run and that resulted in an error IDXRC with a value between of between 100 and 199.

-reindexsevereerrors
Optional.

Reindexes all documents that were indexed during a previous indexing run and that resulted in a severe error (IDXRC = 200). Refer to the topic on what steps to take if indexing stops unexpectedly for more detail.

This command line argument replaces the command line argument `-reindexfatalerrors` used in Content Collector V2.1.1.

-reindexrc <*IDXRC*>
Optional.

Reindexes all documents that were indexed during a previous indexing run and that produced a specific <*IDXRC*> value.

-reindexitems <*itemid*>;<*itemid*>
Optional.

Reindexes only those documents that are listed. List the documents (the <*itemid*>) separated by the delimiter for your operating system. Only items in the item type that is being processed can be reindexed.

Item IDs can be listed with and without a version number. If the version number is omitted, `-V1` (version 1) is the default. Versioning is not relevant when dealing with email as for email there only ever is one version. However, for application documents for example, or Microsoft SharePoint and File System documents, the version number is relevant and must be indicated in the item ID to ensure that the correct item is processed.

The following optional arguments can also be passed to the indexer tool:

-debug <*item id*>;<*item id*>
Optional.

Switches processing to a debugging mode that enables passing individual documents. List the documents (the <*itemID*>) separated by the delimiter for your operating system. The debug option cannot be run in batch mode as user interaction is required. For details see "Debugging the processing of an item" on page 658.

Item IDs can be listed with and without a version number. If the version number is omitted, `-V1` (version 1) is the default. Versioning is not relevant when dealing with email as for email there only ever is one version. However, for application documents for example, or Microsoft SharePoint and File System documents, the version number is relevant and must be indicated in the item ID to ensure that the correct version of the item is processed.

-nseindexreorg
Optional.

Invokes an immediate reorganization of the Net Search Extender index as soon as Net Search Extender has completed indexing. This command line argument overwrites all index reorganization computations set elsewhere.

The command line arguments `-nseindexreorg` and `-debug` cannot be used at the same time.

The command line arguments `-nseindexreorg` and `-continuousmode` cannot be used at the same time.

-shutdown *complete|force*
Optional.

Requests stopping a running indexer instance for the specified item type.

Select *force* to stop processing as soon as possible. If Net Search Extender had not yet been started, it will not be called and all processing done on documents will be lost as nothing will be indexed. If Net Search Extender was already started, the items waiting to be indexed will first be indexed before the process is stopped.

Select *complete* to stop processing as soon as the remaining items waiting to be processed have been processed. This is dependent on **IdxItemMinBatchSize**. Typically, a maximum of 256 items could still be waiting to be indexed. Net Search Extender will run.

Note: Do not expect the indexer process to stop immediately even if you force a shutdown. Stopping the running instance can take some time.

-useconfig *<.ini file>;<.ini file>*
Optional.

Enables using alternate configuration .ini files or configuration files from a different path. Multiple .ini files can be specified and they will be loaded in the order in which they have been specified.

-usesettings *"<option=value>;<option=value>"*
Optional.

Enables the direct temporary overwriting of the persistent indexer configuration values stored in the .ini file. You can add any number of option value pairs *<option=value>* separated by the delimiter for your operating system. The option value pairs must be enclosed in double quotes. The configuration options you list (*<option>*) are not case sensitive. Using this argument is helpful at the beginning to verify the configuration and execution of the indexer.

-exportconfig *<.ini file>*
Optional.

Exports the indexer configuration to the .ini file specified in *<.ini file>*.

-listconfig
Optional.

Only lists all the configuration options and their values that are set for the current indexer run. No other operation is performed.

-verbose
Optional.

Displays detailed processing information on the screen. Setting `-verbose` is helpful at the beginning to get detailed information on all processing. If `-verbose` is specified the log files also contains additional information.

-showsql
Optional.

Displays all of the DB2 SQL commands executed by the system on the screen. Using this argument is helpful at the beginning to verify the configuration and execution of all DB2 system processing. If this argument is specified, all of the SQL statements are also added to the log file.

-trace
Optional.

Goes beyond verbose by displaying more detail in log files to support error determination. This is only required for IBM Support and requires the use of the trace property file.

Important: Do not specify the `-trace` command line argument in combination with `-shutdown`.

-simulate
Optional.

Runs the indexer process but does not update the databases. Net Search Extender is not run. This argument is useful to verify the processing of files and the configuration.

-quiet
No processing steps are displayed on the console.

Related tasks

"Debugging the processing of an item" on page 658

"Reindexing documents" on page 655

Related reference

"Timing information created during the indexer process" on page 610

"Log files created during the indexer process" on page 608

Checking the indexer process return code:

The indexer process return code shows the result of the indexing process. If the indexer process is run automatically in a schedule, the process return code must be checked before the next indexer process starts.

IBM Content Collector provides a sample script that lists all probable indexer system process return codes that require user interaction before the indexer process is called again.

You can use the sample script as a basis for recognizing error situations and determining what action to take. The returned error codes are grouped as follows:

Table 170. Indexer system process return codes

Returned value	Description
0	Indexing completed with no errors. No action is required before the next scheduled indexer process runs.
1	No items were found to process. Either the process is restarted after a defined sleep interval or no action is taken until the next scheduled indexer process runs.

Table 170. Indexer system process return codes (continued)

Returned value	Description
> 1	An error was detected that must be checked before the indexer process scheduler is called again. These errors are serious errors caused, for example, by an incorrect system configuration, by calling the indexer process incorrectly, or by DB2 database processing problems. **Note:** These errors are not related to the actual processing of an item. For a list of these errors see "Index return codes stored in the table of completed tasks."

To evaluate the process return code if the indexer process is run automatically in a schedule:

1. Define your error handling by adapting the sample script `runAfuIndexer` in the subdirectory `samples`.
2. Run the script:
   - On UNIX and Linux you can use:
     ```
     runAfuIndexer.sh
     ```
   - On Windows use:
     ```
     runAfuIndexer.cmd
     ```

Checking the status of indexed items:

You must check the status of all indexed items regularly to determine which items were not indexed successfully because items that were not indexed cannot be searched and found in the repository.

After processing, an item is added to the table of completed tasks with an IDXRC value. Items in the table of completed tasks are grouped as follows:

- An item was successfully indexed with an IDXRC value of 0
- An item was indexed but resulted in a warning. The IDXRC value is between 1 and 99. The item is indexed with a warning notification string
- An item produced an error and was not indexed. The IDXRC value is between 100 and 199
- An item produced a severe error and was not indexed. The IDXRC value is 200

Check the indexing status of items by querying the IDXRC column of the database table of completed tasks to determine the number of items that caused a warning, an error, or a severe error, and possibly the cause of these warnings and errors. The name of the tables of completed tasks are displayed when the indexer or indexer check tools are run in verbose mode.

1. Query the number of items with warnings, errors, or severe errors.
   - For all warning, type:
     ```
     db2 "select * from <tableOfCompletedTasks> where IDXRC >= 1 and idxrc < 100"
     ```
   - For all errors, type:
     ```
     db2 "select * from <tableOfCompletedTasks> where IDXRC >= 100 and idxrc < 200"
     ```
   - For all severe errors, type:
     ```
     db2 "select * from <tableOfCompletedTasks> where IDXRC = 200"
     ```
2. Reindex items with warnings or errors.

Index return codes stored in the table of completed tasks:

All processed items are added to the table of completed tasks with an IDXRC value indicating whether the item was successfully indexed.

An IDXRC value of 0 means that the item was successfully processed and indexed.

An IDXRC value of between 1 and 99 indicates a warning, an IDXRC value of between 100 and 199 indicates an error, and an IDXRC value of 200 indicates a severe error.

The term document in the following descriptions refers to embedded email attachments, and File System and SharePoint documents.

Table 171. Indexing return codes for reindexing items

IDXRC	Return code description
0	The item was successfully indexed.
1	The item was indexed but one or more of its documents were not indexed entirely as their content was truncated due to memory restrictions.
2	The item was indexed but one or more of its documents were not indexed because the documents could not be parsed by the Oracle/Stellent Outside In Technology filters. These documents could also not be considered as plain text documents because they contained more extended chars (code point > 0x7F) than configured in the configuration option **TxtCnvMaxExtendedCharactersRatio**.
3	The item was indexed but one or more of its documents or its body parts were not indexed completely because they were encoded in a code page that is not supported.
4	The item was indexed but one or more of its documents were not indexed because they could not be processed by the Oracle/Stellent Outside In Technology filters. No further attempt was made to examine and reindex these documents because the configuration option **TxtCnvPlainTextHandlingForUnsupportedFormat** was set to 0 which means that no attempt was made to process the document as a plain text document.
5	The item was indexed but one or more of its documents were not indexed because they could not be handled by the Oracle/Stellent Outside In Technology filters. These documents could also not be considered as plain text documents because they contained at least one control character (code point < 0x06) that indicates non-textual content.
6	The item was indexed but one or more of its documents were not indexed because these documents have a file size of zero bytes.
7	The item was indexed but one or more documents were not indexed because the Oracle/Stellent Outside In Technology filters ran into a serious error parsing them.
101	The item was not indexed because Net Search Extender reported an error code 101 for the item.
102	The item was not indexed because Net Search Extender reported an error code 102 for the item.
103	The item was not indexed because Net Search Extender reported an error code 103 for the item.
104	The item was not indexed because Net Search Extender reported an error code 104 or 105 for this item.
105	The item was not indexed because Net Search Extender reported an error code that does not match any of the reason codes above for the item.
110	The item was not indexed because its download processing time from the Resource Manager exceeded the specified timeout.

Table 171. Indexing return codes for reindexing items (continued)

IDXRC	Return code description
111	The item was not indexed because the processing time of the Oracle/Stellent Outside In Technology filters exceeded the specified timeout.
120	The item was not indexed because an error occurred while retrieving the item from the Resource Manager.
121	The item was not indexed because an item was retrieved from the Resource Manager that has zero bytes. This usually indicates a Resource Manager error because all items archived by IBM Content Collector have a size greater than zero bytes.
130	The item was not indexed because the textual content of an item exceeds the transfer buffer size of the `AFUFetchFile` UDF.
199	The item was not indexed because an error occurred while processing the item that does not match any of the errors above for the item.
200	The item was not indexed because it was being processed when the indexer worker process stopped or failed. The item might be the reason for the unexpected error. The item is not processed again.

Monitoring text indexing

You can monitor the indexing process and the indexing performance.

IBM Content Collector provides the following ways in which indexer processing and throughput information can be accumulated and used for monitoring purposes:

Logging

Logging is enabled by setting the indexing configuration option **IdxUseLogging**. If logging is enabled log files for all types (event, system, and trace log files) are created.

Each line in the log files contains a timestamp, a severity, a thread ID, and an operation prefix. This log file information can be extracted for monitoring purposes for example by using the grep tool.

You can filter the type of information you want to monitor by searching for specific operation prefixes, for example you can extract all item processing information from the log files by searching for the entries prefixed by `ITEM:` or extract overall processing statistics by searching for the entries prefixed by `STATISTICS:`.

If the command line argument `-verbose` is specified in addition to logging when any of the indexer tools (`afuIndexer`, `afuCreateIndex`, `afuConfig`, `afuEnableItemType` or `afuCheck`) are run, additional processing information is displayed both on the console or added to the log files.

Visual display

You can monitor the indexer process directly on the screen while it is running by enabling the configuration options **IdxDisplayProgressIndicator**, **IdxDisplayPerformanceMonitor**, and **IdxDisplayNSEMonitor**. Using visual display can be helpful especially during the system setup phase to check that the indexer process is running correctly.

If the command line argument `-verbose` is specified and the configuration option **IdxDisplayProgressIndicator** is enabled, more detailed progress information is displayed on the screen while the indexer process is running.

Timing

If timing is enabled by setting the indexing configuration option **IdxUseTiming**, additional timing information is displayed on the screen while the indexer process is running.

If both timing and logging are enabled, then timing information is not only displayed on the screen but also written to the log files.

In addition, if timing is enabled, a timing file with individual timing values for each item is created. The timing file is in comma separated values (CSV) format that facilitates information extraction for monitoring purposes.

If you want to monitor and measure performance data, enable **IdxDisplayPerformanceMonitor**. Performance data is displayed on the screen and written to the log files (prefixed by `Time:`).

Selecting information to monitor:

It is recommended that you always enable logging during indexer processing by setting the configuration option **IdxUseLogging** to 1. However, logging is not required because the information that the items of an item type have been processed is also recorded in the table of completed tasks. The only difference is that the table of completed tasks does not contain any error details. You should enable logging while you are setting up the text search support component.

Which configuration options you should use to monitor processing differs depending on whether you are still setting up the system or running in production mode. The following settings are recommended.

1. To monitor processing information while you are still setting up the text search support component:
   a. Enable the following configuration options:
      1) **IdxUseLogging**
      2) **IdxDisplayProgressIndicator**
      3) **IdxDisplayPerformanceMonitor**
      4) **IdxDisplayNSEMonitor**
      5) **UdfUseLogging**
   b. Use the indexer processing argument `-verbose`.
2. To monitor processing information after the indexer environment has been set up correctly:
   a. Do not use the indexer processing argument `-verbose`
   b. If you do not want visual feedback or the indexer process is scheduled to run in unattended batch mode, you can disable the following configuration options:
      1) **IdxDisplayProgressIndicator**
      2) **IdxDisplayPerformanceMonitor**
      3) **IdxDisplayNSEMonitor**

c. If you want visual feedback from time to time, especially if you are running the indexer process in continuous mode as opposed to in batch mode, it is recommended that you enable:

   1) **IdxDisplayNSEMonitor**

   The option **IdxDisplayProgressIndicator** is automatically enabled if you are running the indexer process in continuous mode.

d. It is recommended to disable **UdfUseLogging** as otherwise the UDF log file could become very large.

3. To monitor performance data, enable:
   a. **IdxUseTiming**
   b. **IdxDisplayPerformanceMonitor**

You can change the configuration options that determine the time intervals for these monitors if necessary. The options are: **IdxDisplayProgressDelaySecs**, **IdxDisplayPerformanceMonitorDelaySecs**, and **IdxDisplayNseMonitorDelaySecs**.

Log files created during the indexer process:

Logging can be enabled by setting the indexing configuration option **IdxUseLogging** to true.

If logging is enabled in the indexer configuration, log files are created in the directory set in the configuration option **GenLogDirectory**. The name of these log files depends on the setting of the configuration option **GenLogFileNameTimeStamp**.

The general format for log file names is `<itemTypeName><GenLogFileNameTimeStamp>.<type>.[<generation>].log`. The following log files are generated:

- System log files with all of the messages associated with the processing of items and errors
- Trace log files. Only recommended if requested by IBM Support.

If the command line argument `-verbose` is specified when any of the indexer tools is called, a verbose log file with the suffix `verbose.log` containing the system messages and verbose messages is generated in addition to the system log files.

Additionally, the worker process creates its own set of log files with the same file names as the indexer log files but with an additional `worker` infix.

If the command line argument `-verbose` is not specified, only the system log files are created if logging is enabled.

If the indexer is run in batch mode, a single set of log files is created for each indexer run that contains all of the logging information. This means that, in batch mode, log files are not truncated or limited in size and can grow to any size.

If the indexer is run in continuous mode on the other hand, a generation marker is automatically added to the names of the log files. The generation marker consists of the hash (#) character and a consecutive number. Whenever a log file reaches a file size of 128 MB, the log file is closed and logging continues in a new log file with a new generation marker and a new number. The date and timestamp of all the log files however is identical. For example, before processing is shut down, the following log files might have been generated:

```
TESTIT_20100607_1339.system.#0.log
TESTIT_20100607_1339.system.#1.log
TESTIT_20100607_1339.system.#2.log
TESTIT_20100607_1339.system.#3.log
TESTIT_20100607_1339.system.#4.log
```

Note that the number of log files generated using this schema is limited to 16. After 16 log files, the older log files are automatically removed.

Each line in the log file contains a timestamp, a severity, a thread ID, and an operation prefix.

The following prefixes are used:

- ITEM: information about the processing of an item
- NSE: information about the Net Search Extender processing
- TIME: timing information
- STATISTICS: statistical information
- CFG: information about configuration settings
- SYS: information about the operating system
- CMD: information about external commands that are launched
- ERROR: information related to processing errors

The following additional prefixes are used if the command line arguments `-verbose` or `-showsql` are specified when `afuIndexer` is called:

- SQL: information about SQL operations
- RES: results of SQL database operations
- RM: information about IBM Content Manager Resource Manager operation
- LS: information about the IBM Content Manager library server operation

Log file information can be extracted for monitoring purposes, for example, all item processing information (entries prefixed by ITEM) or overall processing statistics (entries prefixed by STATISTICS).

Trace files created when the indexer tools run:

If the command line argument `-trace` is specified when any of the indexer tools (`afuConfig`, `afuIndexer`, `afuCheck`, `afuEnableItemType`, and `afuRecreateIndex`) is called trace files are created.

The trace properties file includes the directory to which the trace files are written. The default is the current directory.

Note:

- A worker process in `afuIndexer` creates its own set of trace files with the same file names as the indexer log files but with an additional `worker` infix.
- Ensure that the current directory (the directory where you run the tool) has write access.
- Unless otherwise instructed, it is recommended that you do not change any settings in the trace properties file.

Running an indexer tool with the argument `-trace` requires a trace properties file in the `cfg` subdirectory that is provided when the text search component is installed. This trace properties file is accessed by the environment variable

AFU_INDEXER_TRACE_PROPERTIES which points to this file. This variable is automatically set when the text search support component is installed.

If further support is required, see Contacting IBM Software Support for more detail on obtaining support.

Timing information created during the indexer process:

Timing can be enabled by setting the indexing configuration option **IdxUseTiming** to true.

General timing information is written to the log file with the prefix TIME. A timing file with individual timing values for each item is written to a file in comma separated values (CSV) format from where this information can be extracted for monitoring purposes.

The timing file name has the same base name as all other log files in IBM Content Collector only the file extension is `timing.csv`.

The first line in the CSV file contains the column names separated by commas. All subsequent rows in the file contains the timing information for an item.

The minimum width of each timing field in the CSV file is set by the option **GenTimingFormatFieldWidth**.

The first column lists the item ID and the following columns contain the operating times. The following operations are measured in milliseconds:

Table 172. Operations for which timing information is gathered

Operation	Description
DATE	The date when the item type was processed
TIME	The time when the item type was processed
TASK	Overall time spent processing the item
RM_RETRIEVAL	Time spent retrieving and loading the item and item attachments from the IBM Content Manager Resource Manager
LS_RETRIEVAL	Time spent retrieving and loading the item's attributes from the IBM Content Manager library server
XML_GENERATION	Time spent generating the XML structure for the item
XML_PARSING	Time spent parsing the IBM Content Manager varying field attribute (ICCVaryingFields)
XML_FILESAVE	Time spent saving the XML file
DOC_PARSER	Time spent running the document parsing including the Oracle/Stellent Outside In Technology filters
DOC_FILEDUMP	Time spent dumping the original item content to the dump directory if the configuration option **GenDump** is enabled.

Statistics created during the indexer process:

Processing statistics are collected whenever the indexer process runs.

If logging is enabled by setting the indexing configuration option **IdxUseLogging** and one or more log files are created, processing statistics are written to these log files and can be extracted for subsequent processing. All statistic information logged is prefixed by STATISTICS.

If the configuration option **IdxDisplayPerformanceMonitor** is also enabled, additional performance data such as the item processing rate is also logged and prefixed by STATISTICS.

The following statistics are collected and written to the log files:

Table 173. Statistics collected during an indexer process

Type of statistics	Description
ITEMS_SCANNED	The number of items taken from the database table of open tasks and sent to the worker process
ITEMS_BYTES	The number of bytes loaded for items from the IBM Content Manager Resource Manager
ITEMS_MISSED	The number of items in the database table of open tasks that could not be sent to the worker process because the worker process terminated unexpectedly and was automatically restarted
WORKER_ITEMS_WITH_NO_ERRORS	The number of items successfully completed by the worker process with no warnings or errors
WORKER_ITEMS_WITH_WARNINGS	The number of items that were successfully completed by the worker process but produced a processing warning (an IDXRC value of between 1 and 99)
WORKER_ITEMS_WITH_ERRORS	The number of items that could not be completely processed by the worker process due to a processing error (an IDXRC value of between 100 and 199)
WORKER_ITEMS_WITH_SEVERE_ERRORS	The number of items that could not be processed by the worker process due to a severe processing error (an IDXRC value of 200)
WORKER_ATTACHMENTS_LOADED	The number of deduplicated attachments that were loaded from the Content Manager Resource Manager
ATTACHMENTS_BYTES	The number of bytes of the deduplicated attachments loaded from the Content Manager Resource Manager
NSE_ITEMS_TO_INDEX	The number of items sent to Net Search Extender for indexing. This is the sum of WORKER_ITEMS_WITH_NO_ERRORS and WORKER_ITEMS_WITH_WARNINGS.
NSE_ITEMS_WITH_NO_ERRORS	The number of items that were indexed successfully by Net Search Extender
NSE_ITEMS_WITH_ERRORS	The number of items that could not be indexed successfully by Net Search Extender.

Table 173. Statistics collected during an indexer process (continued)

Type of statistics	Description
NSE_BYTES	The number of bytes sent to Net Search Extender as textual content in XML files
WORKER_INSTANCES_LAUNCHED	The number of worker processes that have been started by the indexer process
NSE_INSTANCES_LAUNCHED	The number of Net Search Extender instances that have been started by the indexer process

Note: Inconsistencies in the overall statistics can arise due to the following conditions:

- If the worker process stops unexpectedly, the value of WORKER_ITEMS_WITH_NO_ERRORS can be less than the value of ITEMS_SCANNED.
- If items were not processed successfully due to a time out issue, they will not be sent to Net Search Extender for indexing. This means that the value of NSE_ITEMS_TO_INDEX can be less that the value of WORKER_ITEMS_WITH_NO_ERRORS.
- If the worker process stops unexpectedly, the actual number of bytes loaded from the IBM Content Manager Resource Manager can be higher than the value of ITEMS_BYTES.
- If items were not processed successfully, the actual number of bytes processed by Net Search Extender and the AFUFetchFile UDF may be less than the value of NSE_BYTES.
- WORKER_ATTACHMENTS_LOADED and ATTACHMENTS_BYTES will only show positive numbers if the item type uses the compound data model with deduplicated attachments.
- The value in ITEMS_MISSED is the difference of the value in ITEMS_SCANNED minus the sum of the values of NSE_ITEMS_TO_INDEX, WORKER_ITEMS_WITH_ERRORS and WORKER_ITEMS_WITH_SEVERE_ERRORS.
- The value in WORKER_INSTANCES_LAUNCHED is usually 1 unless the worker process stopped unexpectedly and was automatically restarted by the indexer.
- The value in NSE_INSTANCES_LAUNCHED is usually 1 in batch mode and can be any number if the indexer process is run in continuous mode.
- Items counted as ITEMS_MISSED will be processed during a subsequent run of the indexer.

Advanced operations and settings

The procedure to enable searching for text in documents archived in IBM Content Manager after installing the Text Search Support component and creating or enabling item types includes steps with predefined default settings. If you want to add attributes to the default model files, want to change environment variables or adjust indexing configuration defaults, or want to learn more about running indexing in debug mode or about the available reindexing options, check the advanced operations and settings topics.

Preparing for indexing:

Before you can begin indexing, check that the environment has been set up and configured correctly.

Consider the following before you begin indexing documents in a text-searchable item type:

1. The database that you define for archiving in IBM Content Manager must be a UTF-8 database. If the database was not created as a UTF-8 database, you must unload the database, recreate it, and reload the data using the `db2look` and `db2move` commands.
2. Ensure that the Net Search Extender work and index file directories are on the same volume. If the directories are on different volumes, a large amount of data must be copied at the end of each index update, and the entire index must be copied when the index is reorganized. If the files are on the same volume, they are simply renamed instead of being copied.

   If a Net Search Extender index was created using different directories, this can be changed using the `db2text alter index` command.
3. The directory in which the Net Search Extender index is located must have at least twice as much free space as the expected index size. The total space in the file system must be at least 3 times the expected index size. The index size varies by customer, but for planning purposes assume that the index size will be at least 15% of the total size of all email that is indexed.

   For example, if 1 TB of email will be indexed, the index size might be as large as 150 GB. The total space in the file system must be at least 450 GB.

   The text search indexer tracks the size of the indexer files, and if there is not enough free space, processing is stopped and a message is displayed.
4. Configure DB2 functions and procedures on AIX, Linux, or Solaris to run under the ID of the instance owner, not a separate fenced ID. Change the fenced ID by setting the owner on `/home/db2instl/sqllib/adm/.fenced`

The model file:

The model file is a configuration file that contains field definitions specifying what document data (document attributes and metadata) or which IBM Content Manager attributes to include when the index for the item type is built.

The model file is accessed when the item type is created. This is done implicitly by the Content Collector set-up tools or when the initial configuration is run. The model file that is used depends on the data model of the item type.

The model files are in the subdirectory `model` where you installed the component for text search support in IBM Content Manager and are named as follows:

- `afu_mail_entire.xml` for bundled Content Collector item types
- `afu_mail_entire_cmp.xml` for compound Content Collector item types
- `cs_mail_entire.xml` for IBM CommonStore item types for the document model BUNDLED and the archiving type ENTIRE (BRI)

The attribute definitions in a model file must match the attribute definitions in the corresponding indexer configuration file for the item type.

After the full text index has been initialized using a particular model file, the attribute definition for the index cannot be changed anymore. This happens when the item type is created or when the full text index is recreated.

You can only search Content Collector attributes in archived documents after the documents and their attributes have been indexed. IBM Content Manager attributes in archived documents can be search before indexing. However, as these attributes have not been indexed, this process can be very time consuming.

Adding attributes to the model file:

To include additional information from fields available in documents or stored in IBM Content Manager attributes in the index, you must add the corresponding field definitions to the model file.

The model file provides the link to the index engine in IBM Content Manager. It is registered when you create the text-searchable item type.

If you make changes to the model file, you are strongly encouraged NOT to remove any entries from the model file that have been added by Content Collector. Only add new entries to the model file.

To add new attributes:

1. Open the model file in the subdirectory `model` where you installed the component for text search support in IBM Content Manager and add the following definition block:

   ```
   <XMLFieldDefinition
       name   =
       exclude = "NO"
       locator=
    />
   ```

   XMLFieldDefinition
   Introduces a new field definition

   name
   Is the name that you want to give the additional field in the model file

   exclude
   Defines an exclusion condition. As you want to include the additional field rather than exclude it, `exclude` must be set to `= "NO"`.

   locator
   Is the XML path that the indexing engine uses to locate the indexing information in its internal XML structure.

   Note: All paths must begin with the prefix that is defined for the model file. The prefix is dependent on the data model that is selected for the item type.
2. Save your changes and close the file.

Important: When an item type is created, a full text index is initialized using the model file. The attribute definition for the index cannot be changed anymore after the item type was created.

Example for a Content Collector compound item type

```
<XMLFieldDefinition
name = "Record number"
exclude = "NO"
locator = "/icc_document/icc_email/icc_content/record_nr" />
```

In this example, an attribute is added named **Record number**. The search paths that can be used to locate this information are: icc_document, icc_email, icc_content, and record_nr (the same as in the configuration file for the item type).

Adding attributes to the configuration file of the item type:

To include additional information from fields available in documents or in IBM Content Manager attributes in the index, you need to add the corresponding field definitions to the configuration file of the item type.

Additional user-defined IBM Content Manager attributes can be defined for the child components of an item type or for the root component of an item type. You cannot define an attribute for both the child and the root component of an item type with identical names. In this case the indexer will select the attribute defined for the child component of the item type and the attribute for the root component cannot be accessed.

The attribute definitions in the indexer configuration file must match the attribute definitions in the model file for the item type. In particular, the moddef locators defined for attributes in the model file must also be defined in the indexer configuration file. The indexer verifies that the attribute definitions in both files match and issues an error if this is not the case.

You are encouraged not to remove any entries from the configuration file that were added by IBM Content Collector. Only add the attribute definitions to the configuration files that you added to the model file for the item type.

To add field definitions, open the indexer configuration file for the item type that you created:

1. Add the field definitions. Add document field definitions to the section that begins with `[Attributes]` and IBM Content Manager attribute definitions to the section that begins with `[CMAttributes]`.

   Important:

   - Fields added to the [Attributes] section of the indexer configuration file are not case-sensitive.
   - Attributes added to the [CMAttributes] section of the indexer configuration file are case-sensitive. This means that the case must match the case used in the IBM Content Manager item type attribute definition.

   The syntax in both sections is:

   ```
   attribute_name = moddef_locator
   ```

   where

   attribute_name
   Specifies the name of the document field or IBM Content Manager attribute from which the information is to be extracted.

   If the data of the IBM Content Manager attribute is of type CLOB or BLOB, then the client field name must be prefixed by CLOB or BLOB, for example,

   ```
   clob:ICCEJRcpAddrs = /document/envelope/recipient_addresses
   ```

   or

   ```
   blob:ICCVaryingFields = /icc_document/icc_journal
   ```

IBM Content Manager attributes of type String, Integer, or VARCHAR can also be used and do not require a prefix. Any other data types cannot be used.

moddef_locator
Lists the path in the XML structure used for indexing by the indexing engine. You must specify the same `moddef_locator` as you did for `locator` in the model file for the item type. The prefix must match the value of the indexer configuration setting **XmlModdefLocatorBase**.

2. Save your changes and close the file.

Example of a document field definition

```
[Attributes]
Record_number = /icc_document/icc_email/icc_content/record_nr
```

In this example, the indexer would look for information to index in the following paths of its internal XML structure: `icc_document`, `icc_email`, `icc_content`, and `record_nr`. The name of the attribute that the information is extracted from is **Record_number**.

Item type attributes:

Indexer configuration files for IBM Content Collector bundled and compound item types and IBM CommonStore item types are provided by IBM Content Collector. The files contain the default attributes that are stored in the index.

When you create an indexer configuration file for an item type using the `afuConfig` process, the content of the default item type configuration is copied to a new configuration file for the item type.

The default attributes for a compound item type configuration include:

Table 174. Default document and Content Manager attributes added to a compound item type configuration

Attribute names	Moddef_Locator
__ITEMID__	/icc_document/icc_item_id
__@INDEXDATE__	/icc_document/@icc_index_date
__@RECEIVED_DATE__	/icc_document/icc_email/@icc_received_date
__ATTACHMENT__	/icc_document/icc_email/icc_content/icc_attachment
__@ATTACHMENT_NAME__	/icc_document/icc_email/icc_content/icc_attachment/@name
__CONTENT__	/icc_document/icc_generic/icc_content
__BODY__	/icc_document/icc_email/icc_content/icc_body
Attributes used by all parsers	
Body	/icc_document/icc_email/icc_content/icc_body
Subject	/icc_document/icc_email/icc_subject
Attributes used by the SMTP parser	
From	/icc_document/icc_email/icc_from
To	/icc_document/icc_email/icc_recipients/icc_to
CC	/icc_document/icc_email/icc_recipients/icc_cc
BCC	/icc_document/icc_email/icc_recipients/icc_bcc

Table 174. Default document and Content Manager attributes added to a compound item type configuration (continued)

Attribute names	Moddef_Locator
Attributes used by the Microsoft Exchange parser	
Sender	/icc_document/icc_email/icc_from
SentBy	/icc_document/icc_email/icc_from
To	/icc_document/icc_email/icc_recipients/icc_to
CC	/icc_document/icc_email/icc_recipients/icc_cc
BCC	/icc_document/icc_email/icc_recipients/icc_bcc
Attributes used by the Lotus Domino parser	
From	/icc_document/icc_email/icc_from
Principal	/icc_document/icc_email/icc_from
SMTPOriginator	/icc_document/icc_email/icc_from
SendTo	/icc_document/icc_email/icc_recipients/icc_to
CopyTo	/icc_document/icc_email/icc_recipients/icc_cc
INetFrom	/icc_document/icc_email/icc_from
INetSendTo	/icc_document/icc_email/icc_recipients/icc_to
INetCopyTo	/icc_document/icc_email/icc_recipients/icc_cc
BlindCopyTo	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses/bcc
INetBlindCopyTo	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses/bcc
$JournalRecipients	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_1	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_2	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_3	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_4	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_5	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_6	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_7	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_8	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_9	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
$JournalRecipientsExpanded_10	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_1	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses

Table 174. Default document and Content Manager attributes added to a compound item type configuration (continued)

Attribute names	Moddef_Locator
EMSJournalRecipients_2	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_3	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_4	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_5	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_6	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_7	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_8	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_9	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_10	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
CM attribute names	
ICCMailBoxID	/icc_document/icc_mailbox_id
blob:ICCVaryingFields	/icc_document/

The default attributes for a bundled item type configuration include:

Table 175. Default document and Content Manager attributes added to a bundled item type configuration

Attribute names	Moddef_Locator
__ITEMID__	/document/itemid
__@INDEXDATE__	/document/@index_date
__@RECEIVED_DATE__	/document/email/@received_date
__ATTACHMENT__	/document/email/content/attachment
__@ATTACHMENT_NAME__	/document/email/content/attachment/@name
__CONTENT__	/document/email/content
__BODY__	/document/email/content/body
Attributes used by all parsers	
Body	/document/email/content/body
Subject	/document/email/subject
Attributes used by the SMTP parser	
From	/document/email/from
To	/document/email/recipients/to
CC	/document/email/recipients/cc
BCC	/document/email/recipients/bcc
Attributes used by the Microsoft Exchange parser	

Table 175. Default document and Content Manager attributes added to a bundled item type configuration (continued)

Attribute names	Moddef_Locator
Sender	/document/email/from
SentBy	/document/email/from
To	/document/email/recipients/to
CC	/document/email/recipients/cc
BCC	/document/email/recipients/bcc
Recipients	/document/email/recipients
Attributes used by the Lotus Domino parser	
From	/document/email/from
Principal	/document/email/from
SMTPOriginator	/document/email/from
SendTo	/document/email/recipients/to
CopyTo	/document/email/recipients/cc
BlindCopyTo	/document/email/recipients/bcc
INetFrom	/document/email/from
INetSendTo	/document/email/recipients/to
INetCopyTo	/document/email/recipients/cc
INetBlindCopyTo	/document/email/recipients/bcc
CM attribute names	
ICCMailBoxID	/document/mailboxid
CM attributes used for Exchange envelope journaling	
ICCEJSndAddr	/document/envelope/sender_address
ICCEJMsgId	/document/envelope/message_id
ICCEJOrigMsgId	/document/envelope/original_message_id
clob:ICCEJRcpAddrs	/document/envelope/recipient_addresses

The default attributes for a CommonStore item type configuration include:

Table 176. Default document and Content Manager attributes added to a CommonStore item type configuration

Attribute names	Moddef_Locator
__ITEMID__	/document/itemid
__@RECEIVED_DATE__	/document/email/@received_date
__ATTACHMENT__	/document/email/content/attachment
__@ATTACHMENT_NAME__	/document/email/content/attachment/@name
__BODY__	/document/email/content/body
Attributes used by all parsers	
Body	/document/email/content/body
Subject	/document/email/subject
Attributes used by the SMTP parser	
From	/document/email/from

Table 176. Default document and Content Manager attributes added to a CommonStore item type configuration (continued)

Attribute names	Moddef_Locator
To	/document/email/recipients/to
CC	/document/email/recipients/cc
BCC	/document/email/recipients/bcc
Attributes used by the Microsoft Exchange parser	
Sender	/document/email/from
SentBy	/document/email/from
To	/document/email/recipients/to
CC	/document/email/recipients/cc
BCC	/document/email/recipients/bcc
Recipients	/document/email/recipients
Attributes used by the Lotus Domino parser	
From	/document/email/from
Principal	/document/email/from
SMTPOriginator	/document/email/from
SendTo	/document/email/recipients/to
CopyTo	/document/email/recipients/cc
BlindCopyTo	/document/email/recipients/bcc
INetFrom	/document/email/from
INetSendTo	/document/email/recipients/to
INetCopyTo	/document/email/recipients/cc
INetBlindCopyTo	/document/email/recipients/bcc
CM attribute names	
CSLDOrigDB	/document/csmailboxid
CSORIGINATOR	/document/csmailboxid
CM attributes used for Exchange envelope journaling	
CSXEJSNDADDR	/document/envelope/sender_address
CSXEJMSGID	/document/envelope/message_id
CSXEJORIGMSGID	/document/envelope/original_message_id
clob:CSXEJRCPADDRS	/document/envelope/recipient_addresses

The attributes that are enclosed by a double underscore are the built-in attributes. They include:

- `__ITEMID__` : The actual item ID.
- `__@INDEXDATE__` : The date when the XML file is read and processed by the indexing engine Net Search Extender.
- `__@RECEIVED_DATE__` : The date when the document was received in the mail box.

  The date format is controlled by the configuration setting **XmlReceivedDateAndTime**. The default value of the configuration setting is 0 which correlates with the date format `YYYYMMDDhh` which is correct for a Content Collector compound item type. This supports forward compatibility.

For a bundled item type (Content Collector and CommonStore), the value of the configuration option **XmlReceivedDateAndTime** must be changed to 1 which correlates with the date format `YYYYMMDDhhmmss`.

- `__ATTACHMENT__` : The textual contents of one embedded attachment. An item can have more than one attachment.
- `__@ATTACHMENT_NAME__`: The file name of one embedded attachment. An item can have more than one attachment.
- `__CONTENT__` : The textual representation of non-email content, in other words SharePoint and File System document content.
- `__BODY__` : The textual representation of email content.

The at symbol (@) in front of a node in a moddef_locator value indicates that what follows is not a new node but an attribute to the preceding node in the generated XML structure, for example, `<email received_date="2009123123">`.

Text indexer environment variables:

Environment variables may be set to control the execution of the indexer for text search in different environments.

The following environment variables are supported for use by the different indexer tools:

AFU_INDEXER_RMI_PORT

The indexer process communicates with the worker process, and this communication requires a free port number.

The default port number used by the indexer process is 4242. If you experience communication problems between the indexer and the worker process, choose a different port number that is free either by using **AFU_INDEXER_RMI_PORT** and assigning it a unique port number, or by using the `-port` argument when the indexer process is started.

Although the indexer process can only run on one item type at a time, multiple item types can be processed simultaneously. However, to process multiple item types at the same time, each indexer instance must use a unique port. This can be ensured either by using the command line argument `-port` when the indexer process is started or by setting the environment variable **AFU_INDEXER_RMI_PORT**.

AFU_INDEXER_PROMPT_RUN

The indexer detects whether an instance is running for a specific item type by creating a file in the file system. This prevents multiple runs for the same item type. If the indexer is started for the same item type again, the tool reports that it has located the file which indicates that another instance is already running. This results in a message with the name of the file found and the process return code 100.

Normally, this control file is removed as soon as the indexer has completed processing. However, if the indexer terminates abnormally or is interrupted, the file may still reside in the file system and must be removed manually before the indexer will run again for the item type.

If **AFU_INDEXER_PROMPT_RUN** is set to 1, the indexer process will not terminate with a return code of 100 if it detects that an indexer process is already running for the same item type but instead will prompt the user about whether it should continue and ignore the return code. Naturally,

this inhibits batch processing but might be useful during the phase in which the indexer process is being set up.

AFU_INDEXER_PROMPT_RECREATE
The default handling of the `afuRecreateIndex` tool is to prompt the user for confirmation during the index recreation process. This prevents `afuRecreateIndex` from running in a batch mode.

If **AFU_INDEXER_PROMPT_RECREATE** is set to 0, the index recreate tool does not prompt the user but instead assumes `Yes` to be the answer to all prompts. Use this setting with care.

AFU_INDEXER_XML_VIEWER
When the indexer runs in debug mode and processes each document individually, the XML files that are generated for each item can be displayed if **AFU_INDEXER_XML_VIEWER** is set to an application that can display XML files. The indexer will then prompt whether to display the XML file and run the specified application with the name of the XML file if the user chooses to display the file.

For example, on Windows set the environment variable to `C:\Mozilla\FireFox3\firefox.exe` to use the Mozilla Firefox web browser as XML file viewer.

Note: Under UNIX or Linux, it is recommended that the XML viewer application is configured to use a separate window for the display of the XML contents, for example, the Mozilla Firefox browser.

AFU_FETCHFILE_CONFIG_PATH
Typically, the `AFUFetchFile` UDF configuration file that is automatically created by the indexer for text search is located in the directory for temporary files, usually denoted by the environment variables TMP or TEMP on Windows, and in the `/tmp` directory on UNIX or Linux. By specifying **AFU_FETCHFILE_CONFIG_PATH** you can use a directory of your choice to store the UDF configuration file.

For more detail, see “Changes to the default directory and file name settings of the AFUFetchFile UDF configuration” on page 828

AFU_INDEXER_TRACE_PROPERTIES
This environment variable controls the name of the trace properties file if tracing is enabled using the command line switch `–trace`.

If trace is enabled and this environment variable is defined, the indexer process will look for a file with a file name as was specified in the environment variable.

If trace is enabled and this environment variable was NOT defined, the indexer process will look for the default trace properties file named `afuIndexerTrace.properties` in the current directory. If this default file cannot be found, an error message is displayed and the indexer process will stop.

Note: This variable is automatically set during the installation of Content Collector Text Search Support in all of the indexer tool script files in the `bin` directory.

Changing the default indexer configuration:

You can work with more than one configuration file when processing an item type, skip files during document processing, and can adjust the default values of most configuration options.

Working with more than one configuration file

When `afuConfig` is started, it creates a default configuration for the item type. This default configuration is enough to run indexing but is not necessarily fine-tuned to achieve the best system performance and throughput. This means that you might want to change some of the default values.

To avoid having to make the same changes in more than one configuration file, it is recommended that you distribute the indexer configuration of an item type across more that one configuration file. For example, you can have one file containing all the generic item type options that are shared by all item types (for example, the file paths for the XML and log files) and another file with the file filter definitions.

To create more that one configuration file:

1. Make copies of the default configuration file that was created for an item type when you ran `afuConfig`.
2. Edit those files to contain different types of configuration options, for example, add the generic options to one file, the options that you want to monitor into another file, and those that you want to change in yet another file.

   If required options are not found in any of the configuration files, the default values for these options are used.

You can also create your own configuration file. If you create a configuration from scratch, it is recommended that the configuration file begins with the line:

```
[AFU_V1]
```

You can ignore the following two lines that also appear in the sample below as they are used only for reference purposes:

```
CreationDate                                = 2009/07/21 13:55
CreatedForItemType                          = TEST
```

All of the configuration options that you add must be preceded by `[Settings]`.

To specify multiple configuration files when `afuIndexer` is started, use the command line argument:

```
-useconfig <.ini file1>;<.ini file2>
```

The configuration file names are separated by a semicolon (;) on Windows and a colon (:) on UNIX and LINUX machines.

Skipping files during document processing

A set of default file types that is used for filtering when processing email attachments or SharePoint and File System documents is provided by Content Collector in a file called `afu_filetypefilter.ini`. This file represents the blacklist meaning the list of file extensions to skip during processing and is automatically copied to the default configuration when `afuConfig` is run.

If you want to add file extensions to skip during document processing, you can edit this file before you create the indexer configuration file or edit the indexer configuration directly at a later time. The file type filters can be modified after the item type has been created.

A recommended alternate method of handling files to skip during processing is to remove the set of file types from the default configuration and to maintain a separate configuration file with this blacklist of file extensions which is shared across all item types. The configuration file would include **GenFileTypeFilterIsBlackList** = true under [SETTINGS] and list the file extensions to skip under [FILETYPEFILTER].

You can also maintain a whitelist of file extensions representing only those file types that you wish to have processed. If **GenFileTypeFilterIsBlackList** is disabled, the list of file extensions is considered to be the whitelist.

Changing default configuration option values

The values of configuration options in the configuration file created after afuConfig is run are persistent. If you need to change values, there are two ways do this:

- Edit the configuration file directly before afuIndexer is called
- Pass the configuration option values by using the command line argument -usesettings when afuIndexer is called. The configuration values passed using this command line argument overwrite the configuration option values temporarily during the indexing run.

If logging is enabled, all configuration options showing which ones were changed and which ones no longer have the default setting are logged.

Indexer configuration options:

The indexer configuration file for an item type contains all the configuration options that are required for indexing items in the item type.

Note:

- The names of the configuration options are not case-sensitive.
- If the option value is of type Boolean, the value must be either 0 (disabled) or 1 (enabled), or true or false.
- If the option value is of type String, and the value is empty, leave the string empty. Do not indicate an empty string by using double quotes.
- If the option value is of type Integer, you can specify the value as a number or a number with a suffix. The suffix is a factor that is multiplied by a value. The following suffixes are supported:
  - k = * 1000
  - K = * 1024
  - m = * 1000 * 1000
  - M = * 1024 * 1024
  - g = * 1000 * 1000 * 1000
  - G = * 1024 * 1024 * 1024

  For example: 1g = 1000m = 1000 000k = 1000 000 000

  You must round off all numbers, for example, you cannot use 1.5m only 1m.

Generic index process configuration options

The following GenCommandXXX configuration options can be used to add user-defined processing steps directly before Net Search Extender indexing or Net Search Extender index reorganization begins, or immediately after indexing or reorganization has completed. Inserting user-defined commands at these key processing stages enables monitoring index processing options or backing up relevant data.

Note: On Windows, if the user command is a .BAT or .CMD command file, the command must include `cmd /c`. For example, `cmd /c mycommand.bat`.

All of the configuration GenCommandXXX options can use the following variables, which will be replaced by their actual values in the command strings.

- %ITEMTYPE%: name of the item type
- %ITEMTYPEID%: the item type identifier
- %INDEXNAME%: Net Search Extender index name
- %INDEXID%: Net Search Extender index identifier
- %INDEXDIR%: Net Search Extender index directory
- %INDEXWORKDIR% : Net Search Extender index working directory

This enables using backup commands that can be embedded in the indexing processes and can then be passed arguments that these commands require to access item specific information. For example:

```
GenCommandAfterNseIndexing =
  myBackup.exe -itemtype %ITEMTYPE%  -indexdirectory %INDEXDIR%
```

or

```
GenCommandAfterNseIndexing =
  cmd /c myBackup.cmd -itemtype %ITEMTYPE%  -indexdirectory %INDEXDIR%
```

GenCommandAfterNseIndexing
Value type: String

Default value is no value.
If set, the user command will be executed directly after the Net Search Update Index process has completed.

GenCommandAfterNseReorg
Value type: String

Default value is no value.
If set, the user command will be executed immediately after the Net Search Extender index reorganization operation completes.

GenCommandBeforeNseIndexing
Value type: String

Default value is no value.
If set, the user command will be executed directly before the Net Search Update Index process begins.

GenCommandBeforeNseReorg
Value type: String

Default value is no value.
If set, the user command will be executed immediately before the Net Search Extender index reorganization operation begins.

The remaining generic configuration options include:

GenCommandErrorRcAborts
Value type: Boolean

Default value: 0
If set to 0, the indexer will continue processing even if a user command in any one of the GenCommandXXX options completed with a return code that was not zero.
Only if it is set to 1, will the indexer process stop if one of the user commands completed with a return code that is not zero. The indexer process itself then stops with a return code of 9 which indicates that an error occurred in one of the user commands in a GenCommandXXX option.

Note: If an error occurred during the execution of one of the GenCommandXXX commands and **GenCommandErrorRcAborts** is set to 1, the temporary XML files are not deleted. After you have fixed the error, you must delete these files manually. Bear in mind that, because of the error, item processing was not completed and the table of complete tasks is not updated.

GenRmRetryAllErrors
Value type: Boolean

Default value: 1
Normally the indexer will only attempt to retrieve those items again that produced a timeout error.
If **GenRmRetryAllErrors** is set to 1, all of the errors returned from the Resource Manager for items are treated as if they were timeout errors and the indexer process will try retrieving these items again the next time the indexer process runs. The overall number of times that the indexer will attempt retrieving these items is set by the **GenMaxNumberOfRetriesForRmErrors** option.
If **GenRmRetryAllErrors** is set to 0, the indexer will only attempt retrieving the items that produced a timeout error again. The items that produced other Resource Manager errors are not retrieved again and are added to the table of completed tasks with an error. The items that could not be retrieved from the Resource Manager are also added to the log file.

GenDump
Value type: Boolean

Default value: 0
If set to 1, the original documents retrieved from the Resource Manager are stored in the dump directory set by the **GenDumpDirectory** option. If you have difficulties processing certain documents and require support analyzing the documents, set this value to 1.

GenDumpDirectory
Value type: String

Default value: $(TMP)/afu/dump
The directory where the original documents retrieved from the Resource Manager are written to if **GenDump** is set to 1. To help to differentiate files, the file names in the dump directory contain the item ID.

Note:

- Dump operations that result in errors are not logged or handled in any way. The dump directory is also not cleaned up by the indexer process. It is the responsibility of the user to delete files in the dump directory and to reset the dump option after debugging.

- It is recommended that the value of **GenDumpDirectory** and **GenWorkingDirectoryForItemTypes** be identical for all item type configurations.

GenFileTypeFilterIsBlackList
Value type: Boolean

Default value: 1
Determines which attachments or documents are to be parsed using the Oracle/Stellent Outside In Technology filters, and which to skip. If set to 1, the list of file name extensions listed in the section FileTypeFiles in the configuration is considered to be the blacklist. This means that all documents with matching file name extensions are skipped.
If set to 0, the list of file name extensions in the section FileTypeFiles is considered to be the whitelist. This means that only those documents that have a file name extension listed in the section FileTypeFiles will be processed. All other documents are ignored.

GenLogDirectory
Value type: String

Default value: $(TMP)/afu/log
The directory where the log files are written if logging is enabled by setting the option **IdxUseLogging** to 1.

Note: Only the following alphanumeric characters can be used when determining the directory name:

- All letters in the alphabet (A to Z, and a to z) and the numbers between 0 and 9
- On UNIX and Linux: `/_-.@#`
- On Windows: `\:~_-.@#$` and the space character

The maximum length of the string is 255 characters.

Restriction: The value of **GenXmlDirectory** and **GenLogDirectory** MUST be identical for all item type configurations.

GenLogFileNameTimeStamp
Value type: String

Default value: _yyyyMMdd_HHmm
Log files use timestamps in their names. It is recommended that you keep to the default timestamp for the log file names so that there is a new set of log files for each run of the indexer process.
If you do not specify a value for **GenLogFileNameTimeStamp**, all log files will have the same file name and will be overwritten the next time the indexer runs for the same item type.
If you want to change the timestamp in the file names for log files, use the following letters. Any other letter is treated as is.

Note: Only the following characters are supported:

- On UNIX and Linux: `/_-.@#`
- On Windows: `\:~_-.@#$` and the space character

Table 177. Letters to use for timestamps in log file names

Letter	Date or time	Example
y	Year	1997; 97

Table 177. Letters to use for timestamps in log file names (continued)

Letter	Date or time	Example
M	Month in year	July; Jul; 07
w	Week in year	27
W	Week in month	2
D	Day in year	189
d	Day in month	31
F	Day of week in month	2
E	Day in week	Tuesday; Tue
a	Am/pm marker	PM
H	Hour in day (0-23)	0
k	Hour in day (1-24)	24
K	Hour in am/pm (0-11)	0
h	Hour in am/pm (1-12)	12
m	Minute in hour	30
s	Second in minute	57
S	Millisecond	978
z	Time zone	PST; GMT-08:00

GenRmConnectionTimeOutSecs
Value type: Integer

Default value: 60
The timeout for retrieving items from the Resource Manager. If retrieving an item from the Resource Manager takes longer than the value specified in the option **GenRmConnectionTimeOutSecs**, the retrieval is treated as having produced an error. This error can result in the item being retried the next time the indexer runs, or in the item being added to the table of completed tasks with the error message.
.

GenMaxNumberOfRetriesForRmTasks
Value type: Integer

Default value: 3
The number of times an attempt is made at retrieving items that returned with a timeout error and, if the option **GenRmRetryAllErrors** is set to 1, returned with any other Resource Manager error. 10 times is the maximum number of times that processing can be retried.

GenWorkingDirectoryForItemTypes
Value type: String

Default value: $(TMP)/afu
This directory is required for all working files. A subdirectory is automatically created for each item type. The OutsideIn filters also use this working directory. Ensure that you use a high performance file system for this directory.

Note: It is recommended that the value of **GenDumpDirectory** and **GenWorkingDirectoryForItemTypes** be identical for all item type configurations.

GenXmlDirectory
Value type: String

Default value: $(TMP)/afu/xml
One XML file is generated for each item. The XML files are intermediate files that are automatically deleted after indexing unless the configuration option **XmlDeleteFilesAfterNseProcessing** is set to 0. Ensure that there is sufficient space in the directory for all of the XML files.

Note: Only the following alphanumeric characters can be used when determining the directory name:

- All letters in the alphabet (A to Z, and a to z) and the numbers between 0 and 9
- On UNIX and Linux: `/_-.@#`
- On Windows: `\:~ _-.@#$` and the space character

The maximum length of the string is 255 characters.
The XML directory is shared by all of the indexer processes that are running at the same time for multiple item types and can therefore become very large.

Restriction: The value of **GenXmlDirectory** and **GenLogDirectory** MUST be identical for all item type configurations.

GenXmlDirectoryMinFreeSizeRequiredInMB
Value type: Integer

Default value: 512
When the indexer process checks the designated file system, at least the number of MB that is specified in the configuration option **GenXmlDirectory** must be available in the XML directory.

GenNseIndexDirectoryMinFreeSizeRequiredInMB
Value type: Integer

Default value: 512
When the indexer process checks the designated file system, at least the specified number of MB must be free in the Net Search Extender index directory.

GenDatabaseTransactionSize
Value type: Integer

Default value: 200,000
The value of **GenDatabaseTransactionSize** determines the maximum number of rows that will be updated during a database operation. The number depends on the IBM DB2 transaction log size. If you experience SQL errors during update operations, decrease this number. Alternatively, you can increase the IBM DB2 transaction log size.

Index process configuration options

IdxMaxNumberOfSubsequentRmErrors
Value type: Integer

Default value is 100
The minimum value is 10 and the maximum is 9999. This option specifies the maximum number of Resource Manager retrieval errors that are permitted in a row with no successful retrieval attempts between these retrieval errors. This is

useful to detect if the Resource Manager is not running and requires administrator attention to prevent endless indexer processing on a large batch with no items to index.

IdxUseTiming
Value type: Boolean

Default value is 0
Enables timing and the creation of timing files in CSV format in the log directory.

Note: If the **IdxDisplayPerformanceMonitor** configuration option is enabled, the indexer processing rate in items processed per second s also displayed irrespective of the value of **IdxUseTiming**.

IdxUseLogging
Value type: Boolean

Default value is 1
Enables logging and creates log files in the log directory specified in the **GenLogDirectory** configuration option. By default, the log files will contain only important information, for example, which items were processed successfully, which items were processed with an error, and system configuration data.

Note: If the command line argument `-verbose` is specified when `afuIndexer` is called, additional information is added to the log file if logging is enabled.

IdxLogItemIds
Value type: Boolean

Default value is 1
If logging is enabled, by default, the item IDs of all items are logged twice; once when the items are sent to the workers and again when the items have been processed. This means that two lines are written to the log file for each item which has an impact on the size of the log file especially for very large batches.
As the status information of an item is also stored in the table of completed tasks, the logging of the processing status of each item in the log file can be suppressed by setting **IdxLogItemIds** to 0. Only items that produce processing errors will be logged.
It is recommended that this option is disabled when the indexer is run in continuous mode to reduce the amount of data added to the log files.

IdxItemMaxBatchSize
Value type: Integer

Default value is 100 000
The number of documents that are indexed in one indexing run.

IdxItemMinBatchSize
Value type: Integer

Default value is 0
If set to 0, the indexer process automatically determines the size of the batch of items that are waiting to be processed. As soon as this minimum batch size is reached, the items in the batch will be processed. If this minimum size is not reached, the indexer will stop with a return code implying that there are no documents to process.
If you want indexing to run at all costs even if the batch contains only a few items set the option to 1.

IdxAutoIndexingEveryNDays
Value type: Integer

Default value is 7 (days). Allowed is a value between 0 and 30.
IdxAutoIndexingEveryNDays only affects the indexer process if `afuIndexer` is run in continuous mode.
Setting **IdxAutoIndexingEveryNDays** to a value greater than 0 days ensures that all items that were archived during this time period are indexed irrespective of any batch size specifications. This enables searching for archived content quicker, and not having to wait for the maximum batch size to be reached before Net Search Extender is started and the item content is indexed. If Net Search Extender is started because the number of days set in this option has elapsed, a message is written to the console and the indexer log file. Setting the option to 0 means that this is not done and Net Search Extender is only started when the maximum batch size is reached.

Note: The effective time is always a few seconds or minutes more than `N days * 24 hours`.

IdxDaysConsideredForMinBatchSize
Value type: Integer

Default value is 10 (days)
The number of days to use to calculate the average number of items that were added to the item type per day. This average is used to compute the minimum batch size.
For details of how this options is used, see "Constraints when indexing in batch mode" on page 644.

IdxReorgBatchSize
Value type: Integer

Default value is 1.5 million
The number of items that are indexed before an index reorganization operation is automatically performed.

IdxMaxNumberOfUpdatesWithoutReorg
Value type: Integer

Default value is 10 million
The absolute upper limit of items that can be indexed after which all processing stops until the index is reorganized. As soon as the number of items added to the index has reached the number specified in this option and if no index reorganization has been performed so far, the indexer will refuse to index any further items and will stop processing with an error. The index reorganization operation must be started manually by running the `afuIndexer` with the command line argument `-nseindexreorg`.

IdxReorgTimeBlockIntervalBegin
Value type: String in time format `HH:mm` (hours in day 0:23)

Default value is no value
This configuration option is used with **IdxReorgTimeBlockIntervalEnd** to define a specific time frame during which no index reorganization is allowed to run. This option sets the start of the time frame. As the index reorganization operation is time intensive and also blocks searching the index for a period of time, defining a time frame during which an index reorganization is not permitted to run may be advisable, for example, during the main business hours. The default is to define no start time.

IdxReorgTimeBlockIntervalEnd
Value type: String in time format `HH:mm` (hours in day 0-23)

Default value is no value
This configuration option is used with **IdxReorgTimeBlockIntervalBegin** to define a specific time frame during which no index reorganization is allowed to run. This option sets the end of the time frame. The default is to define no end time.

IdxDelayDocumentProcessingInDays
Value type: Integer

Default value is 0
If the value is 0 the item is processed without delay. Setting a delay of a few days might be useful to prevent the redundant processing of the same items that are added to the item type through archive requests by different users during this time window of a few days. By delaying the immediate processing of these items, the chances are greater that these items will need to be processed only once. The disadvantage is that these documents cannot be searched for immediately because they will only be added to the index after the specified number of days has elapsed.

IdxNumberOfParallelIndexerTasks
Value type: Integer

Default value is 0
If set to 0, the system automatically determines the appropriate number of threads based on the amount of system CPU available. The maximum number is 128.
Use the default setting if the machine on which the indexer is running can use the majority of the available system resources for the indexer workload, for example, when running the indexer during light-load times.
Configure the number of processing threads if the system can only use a limited amount of the system resources (notably CPU resources) for the indexer workload.
In general, the number of indexer threads should correspond to the number of CPU processing cores that you want the indexer workload to be using. For example, if only 60% of the CPU resources may be used by the indexer, allocate a number of threads that corresponds to 60% of the CPU processing cores.
The system should always be monitored to ensure adequate response times for all offered services. Lowering the number of parallel indexing tasks will free up system resources for other workloads. The higher the number of tasks, the greater the impact on memory consumption.

IdxDisplayProgressIndicator
Value type: Boolean

Default value is 1
If set to 1, a graphical progress indicator is displayed for all longer operations at regular intervals. The delay between these intervals is set in the configuration option **IdxDisplayProgressDelaySecs** and can be changed. The default is display the progress every 30 seconds.
It is recommended to disable this option if the indexer process is running in an environment in which the console output is redirected to a file.
When `afuIndexer` is run in verbose mode, the graphical progress indicator is replaced by a textual progress indicator.
When `afuIndexer` is run in continuous mode, the graphical progress indicator is replaced by an overall status monitoring that displays process statistics,

batch sizes, and the time elapsed since the indexer process was started. This option is automatically enabled when the indexer is run in continuous mode.

IdxDisplayProgressDelaySecs
Value type: Integer

Default value is 30 seconds
The time interval during which processing progress is displayed. The minimum is 5 seconds and the maximum is 600 seconds.

IdxDisplayPerformanceMonitor
Value type: Boolean

Default value is 1
If set to 1, the processing rate in items processed per second is displayed for longer operations at regular intervals. The delay between these intervals is set in the **IdxDisplayPerformanceMonitorDelaySecs** configuration option and can be changed. The default is every 60 seconds. The processing rate is also written to the log file as statistical information if logging is enabled. The performance monitor works independently of the setting of the **IdxUseTiming** option.

IdxDisplayPerformanceMonitorDelaySecs
Value type: Integer

Default value is 60 seconds
The time interval during which the item monitoring performance processing rate is displayed. The minimum is 5 seconds and the maximum is 600 seconds.

IdxDisplayNseMonitor
Value type: Boolean

Default value is 0
IdxDisplayNseMonitor displays the status of the Net Search Extender update index process at regular intervals by showing the number of items that were added, updated, or deleted by Net Search Extender.
If set to 1, the progress of the Net Search Extender index update operations is displayed at regular intervals after Net Search Extender has been started. The delay between these intervals is set in the configuration option **IdxDisplayNseMonitorDelaySecs** and can be changed. The default interval is every 300 seconds (every 5 minutes).

IdxDisplayNseMonitorDelaySecs
Value type: Integer

Default value is 300 seconds (5 minutes)
The time interval during which the Net Search Extender index update operations is displayed. The minimum is 60 seconds and the maximum 3600 seconds (1 hour).

IdxWorkerMaxMemorySizeInMb
Value type: Integer

Default value is 1024 MB
The maximum size of the worker's JVM process heap memory. If you notice that the worker processes are running short of memory and terminating as a consequence, increase the size of **IdxWorkerMaxMemorySizeInMb** gradually. The maximum size of the worker's JVM process heap memory is 3072 MB.

XmlModdefLocatorBase
Value type: String

Default value is `/icc_document` for item types of the compound data model. If the data model of the item type is bundled, the value must be changed to

/document.
This value must match the prefix of the moddef locators used in the model file associated with the indexer configuration for the item type.

Note: Only the following alphanumeric characters can be used in a moddef locator string:

- All letters in the alphabet (A to Z, and a to z) and the numbers between 0 and 9
- The characters: /_@ (whereby @ has a special meaning)

XmlReceivedDateAndTime
Value type: Boolean

Default value is 0
If 0, the item type attribute received_date contains only the date and hour. This is the default date format for the received_date attribute in compound item types. This is the recommended date format for all new item types.
If 1, the item type attribute received_date contains the date and precise time. This is the default date format for the received_date attribute in bundled item types.

Document parsing configuration options

The following options directly influence document parsing. This options also affect the document parsing carried out by the Oracle/Stellent Outside In Technology filters. These filters are called for all embedded email attachments, and File System and SharePoint documents. The term document is used to refer to both email attachments, and File System and SharePoint documents.

MsgMapAllUnsupportedCodepagesToDefault
Value type: Boolean

Default value is 1
If set to 1, all documents, which are encoded in an unsupported or an invalid code page listed in the configuration option **MsgCodepageMapTable**, are treated as if they were encoded in the code page specified by the option **MsgDefaultCodePageForNonUnicodeFiles** and converted.
This configuration option is only relevant when processing Microsoft Exchange documents

MsgDefaultCodePageForNonUnicodeFiles
Value type: Integer

Default code page is 1252
Specifies which codepage to use as the default code page if an Exchange document is encoded in unsupported code page. This option is required only if either **MsgCodepageMapTable** or **MsgMapAllUnsupportedCodepagesToDefault** is also specified. If the option is not specified, the code page of the current locale is used as the default code page.
This configuration option is only relevant when processing Microsoft Exchange documents

MsgCodepageMapTable
Value type: String

Default value is no value.
Specifies a set of code pages (separated by spaces) that are not supported by the indexer. If a documents is encoded in one of these code pages, it is

converted as if it were encoded in the code page specified by the option **MsgDefaultCodePageForNonUnicodeFiles**. This enables maintaining a list of code pages that can be converted to the code page specified by the option **MsgDefaultCodePageForNonUnicodeFiles** because their character sets are almost identical or similar to the code page specified by the option **MsgDefaultCodePageForNonUnicodeFiles**.

Note: All code pages listed must be numeric values only. Only IBM CCSIDs can be used, no ISO code pages. For example: iso-8859-1 = CCSID 1252.
This configuration option is only relevant when processing Microsoft Exchange documents

MsgCodepageIgnoreTable
Value type: String

Default value is 42
Specifies a set of code pages (separated by spaces) that are to be ignored by the indexer. If a documents is encoded in one of these code pages, it is not converted. This allows ignoring code pages such as Windows 42 which is a pure symbol code page that is not supported by the indexer and for which a conversion into textual content would not make sense.

Note: All listed code pages must be numeric values only. Only IBM CCSIDs can be used, no ISO code pages. For example, iso-8859-1 is CCSID 1252.
This configuration option is only relevant when processing Microsoft Exchange documents.

TxtCnvEnabled
Value type: Boolean

Default value is 1.
If set to 1, document conversion is enabled.
If **TxtCnvEnabled** is set to 0, documents are not converted to their textual representation which is required to enabling indexing. This configuration option should only be set to 0 if you do not want to index email attachments when working with email item types. Do not set this configuration option to 0 when working with Microsoft SharePoint and File System item types.

TxtCnvPlainTextHandlingForUnsupportedFormat
Value type: Boolean

Default value is 1

If set to 1, all documents in unsupported formats will be attempted to be processed as plain text documents. For details refer to the topic on embedded email attachment and document processing.

TxtCnvAutoCodepageDetection
Value type: Integer

Default value is 1
If set to 1, the indexer process will attempt to automatically detect the code page of the documents that could not be parsed by the Oracle/Stellent Outside In Technology filters. Depending on the level of confidence attained detecting the code page, the document is processed or tagged with an error stating that the automatic code page detection was not successful. For details refer to the topic on embedded email attachment and document processing.

TxtCnvAutoCodepageDetectionMinConfidence
Value type: Integer

Default value is 70
Automatic code page detection is considered successful if a minimum level of confidence is reached which is equal to or higher than the value set in the **TxtCnvAutoCodepageDetectionMinConfidence** option. If the confidence level is above the level set, the document is converted from its code page and processed as a plain text document. The minimum level of confidence (between 1 and 100 percent) is calculated if the option **TxtCnvAutoCodepageDetection** is set to 1. For details refer to the topic on embedded email attachment and document processing.

TxtCnvDefaultCodepageForPlainText
Value type: Integer

Default value is 0.
If **TxtCnvDefaultCodepageForPlainText** is not specified or is set to 0, all documents for which no code page could be detected are considered to be plain text and indexed as plain text.
If the value set in **TxtCnvDefaultCodepageForPlainText** is anything other than 0, all documents that are not handled in another way are considered to be encoded in the specified code page and converted from this code page for processing.

TxtCnvMaxExtendedCharactersRatio
Value type: Integer

Default value is 30
The percentage of extended characters (code point > 0x7F) that a document may contain. This percentage is calculated if automatic code page detection fails or is disabled. If the ratio is exceeded, the IDXRC of the document is set to 2. For details, see "Searching for error strings" on page 662.

TxtCnvTargetBufferSizeInMB
Value type: Integer

Default value is 10
The maximum amount of space for each thread that the extracted text section of a parsed document can fill. If the pure textual content of the document does not fit into this buffer, it will be truncated and marked that it was truncated. The minimum is 1 MB for each thread and the maximum is 256 MB.

Important: The value of this configuration option must be less than the size of the UDF transfer buffer. The UDF transfer buffer is used when the AFUFetchFile UDF loads the temporary XML files into memory (which is the transfer buffer) for access by Net Search Extender. The default size of the UDF transfer buffer is 32 MB. For details about how to change the UDF transfer buffer size, see "Changing the default transfer buffer size of the AFUFetchFile UDF" on page 640.

TxtCnvTimeOutInSecs
Value type: Integer

Default value is 120
The time in seconds during which the document must be parsed. The minimum value is 30 seconds and the maximum is 3600 seconds (1 hours). After a timeout is reached, the document is tagged with a timeout error.

Less important indexer configuration options

Some configuration options are used for backward compatibility purposes only and are always configured with a default value. As with all configuration options,

you can change the default value if necessary. However, these configuration options are only exported or listed as part of the configuration if the value is no longer the system default value.

The less important configuration options are the following:

GenTimingFormatFieldWidth
Value type: Integer

Default value: 8
The minimum width of each timing field (column) in the timing file in CSV format.

IdxIgnoreUnknownCmAttributes
Value type: Boolean

Default value is 1
An unknown IBM Content Manager attribute is an attribute that is listed in the text search indexer configuration but is not defined in the target IBM Content Manager item type. If set to 1 and an unknown IBM Content Manager attribute is encountered, processing continues with a warning. If set to 0, processing stops with an error.

IdxMaxNumberOfErrorsDisplayed
Value type: Integer

Default value is 0
Specifies the maximum number of error messages that are displayed on the screen during indexer processing. If the number of error messages exceeds the specified number, a warning is displayed that all subsequent errors will no longer be displayed on the screen. However, the indexer processing will continue writing all errors to the log file if logging was enabled.
If **IdxMaxNumberOfErrorsDisplayed** is set to 0, all errors are displayed on the screen.

IdxDisplayTimeStamp
Value type: Boolean

Default value is 0
If set to 1, all informational messages are preceded by a time stamp.
The option is automatically enabled when the indexer process is run in continuous mode.

IdxWorkerMinMemorySizeInMb
Value type: Integer

Default value is 1024 MB
The minimum size of the worker's JVM process heap memory.
If you set the value to 0, the JVM will determine a minimum value automatically.

XmlDeleteFilesAfterNseProcessing
Value type: Boolean

Default value is 1
If set to 1, automatically deletes the generated XML files after Net Search Extender has finished.
If set to 0, the files are not automatically deleted and must be manually deleted. The XML files are located in the directory specified by the option **GenXmlDirectory**.

Note: The XML directory is shared by all of the indexer processes across all item types, so delete XML files with caution if `afuIndexer` instances are running for other item types at the same time. If no other `afuIndexer` instances are running, the complete directory tree specified in the **GenXmlDirectory** configuration option can be deleted. The indexer process will recreate the directory structure automatically if necessary.

XmlKeepFilesCausingNseErrors
Value type: Boolean

Default value is 0
Usually, if the option **XmlDeleteFilesAfterNseProcessing** is set to 1, all XML files are automatically deleted. If **XmlKeepFilesCausingNseErrors** is set to 1, the XML files that caused a Net Search Extender error will not be deleted.
If **XmlKeepFilesCausingNseErrors** is set to 0, and the option **XmlDeleteFilesAfterNseProcessing** is set to 1, all XML files are deleted, including those that caused a Net Search Extender error.

XmlInsertAttachmentProcessingWarnings
Value type: Boolean

Default value is 1
If a warning (and not an error) occurred during document processing (an embedded email attachment, or a File System or SharePoint document) and the option is set to 1, a warning string `IcmFceWarning: <a warning message>` will be indexed with the document.
If the option is set to 0, the warning string `IcmFceWarning: <a warning message>` will not be indexed with the document.
Refer to the topic on searching for error strings for a list of the warning strings that are stored with the item.

TxtCnvGeneralErrorAborts
Value type: Boolean

Default value is 0
If set to 1 and if a severe error occurred in the Oracle/Stellent Outside In Technology filters processing, the corresponding item is flagged with an error value of 199 in column IDXRC in the table of completed tasks and no attempt is made to process the item. Any remaining attachments or documents belonging to the item will not be processed.
If set to 0 and if an error occurs during document processing, the error is ignored and the document is indexed with an error that can be searched for. This way, all documents that were flagged as having caused an error can be identified. For a list of the error return values, refer to the topic on index return codes.

TxtCnvOutsideInExtendedTest
Value type: Boolean

Default value is 1
If set to 1, the Oracle/Stellent Outside In Technology filters run an extended test to identify the document type. This supports detecting the type of document more precisely but this extended analysis requires more of the available CPU resources.

TxtCnvOutsideInFilterGroups
Value type: Boolean

Default value is 1
Use the Oracle/Stellent Outside In Technology filters document groups for

filtering specific file types. These documents are filtered out and not processed because they include content that is not suitable for indexing, such as graphic files or binary files.

MimeGeneralErrorAborts
Value type: Boolean

Default value is 0
If set to 1 and if a severe error occurs during MIME parsing, the corresponding item is flagged with an error value of 199 in column IDXRC in the table of completed tasks and no attempt is made to process the item. Any remaining attachments or documents belonging to the item will not be processed.
If this option is not set, the error is ignored and document parsing will be continued if possible.

MimeTreatHtmlAsAttachment
Value type: Boolean

Default value is 1
If you want to process HTML content in alternate MIME parts of a document correctly, you must set **MimeTreatHtmlAsAttachment** to 1 which triggers processing by the Oracle/Stellent Outside In Technology filters. Although parsing HTML content using the Oracle/Stellent Outside In Technology filters is time consuming, the content that is subsequently indexed will not contain any HTML tags and will as such be easier to search. However, this content is indexed as attachment content with the file name `HTML_PART` and not as body content. This is significant when searching because the HTML content will not be found if the search criterion is limited to the content stored in the body attribute. Note that this generated attachment `HTML_PART` file will never be filtered by a file type filter definition.
If **MimeTreatHtmlAsAttachment** is set to 0 and if an alternate MIME part with content of type HTML is detected in a document, the complete MIME part is ignored and will not be indexed.

`AFUFetchFile` UDF related configuration options

The `AFUFetchFile` UDF has its own configuration which is automatically created by the indexer process. The UDF configuration uses the following `afuIndexer` configuration options:

- **UdfMaxNumberOfWarnings**
- **UdfUseLogging**
- **UdfAutoConfigReloadInMin**
- **GenXmlDirectory**
- **GenLogDirectory**

Whenever the value of any one of these configuration options is changed, the UDF configuration is automatically updated. After the UDF configuration is updated, the user is informed about this in a message. Depending on the value of the configuration option **UdfAutoConfigReloadInMin**, the user is asked to either restart DB2 so that the changes can become effective, or to wait some time during which the UDF reloads the configuration itself.

After the configuration has been updated, the indexer process terminates with a process return code of 4. The indexer process must be restarted manually.

The following `AFUFetchFile` UDF related configuration options are shared across all item type configurations:

UdfMaxNumberOfWarnings
Value type: Integer

Default value is 100. The maximum value is 9999
If more than the specified number of warnings occurs during the Net Search Extender index run, the UDF returns a severe error to Net Search Extender and Net Search Extender stops processing.
This means that the items that have already been processed as input to Net Search Extender will not be removed from the table of open tasks and added to table of completed tasks. All of the items processed during this run will be processed again the next time `afuIndexer` runs.

UdfUseLogging
Value type: Boolean

Default value is 0
If set to 1, the `AFUFetchFile` UDF adds logging information to the file `AfuFetchFile.log` in the log file directory specified in the configuration option **GenLogDirectory**.

UdfAutoConfigReloadInMin
Value type: Integer

Default value is 0
The number of minutes before the `AFUFetchFile` UDF automatically reloads its configuration. The configuration is located either in the default temporary subdirectory of the IBM Content Manager installation directory or in the file path defined by the environment variable AFU_FETCHFILE_CONFIG_PATH. A default of 0 seconds means that the configuration is not automatically reloaded.
As the `AFUFetchFile` UDF is started after DB2 has started and Net Search Extender is running for the first time, any configuration changes to the UDF require a DB2 restart. Forcing the UDF to reload the configuration automatically at regular intervals avoids having to restart DB2.
Automatic reloading of the configuration enables repeated modifying and testing of any one of the following configuration options without having to restart DB2:

- **UdfMaxNumberOfWarnings**
- **UdfUseLogging**
- **GenXmlDirectory**
- **GenLogDirectory**

As soon as the configuration is stable and requires no further changes, automatic reloading should be disabled by setting **UdfAutoConfigReloadInMin** to 0.

If you need to reconfigure the transfer buffer size of the `AFUFetchFile` UDF because the buffer size is too small for the textual content that is extracted from an item, refer to "Changing the default transfer buffer size of the AFUFetchFile UDF" for details.

Changing the default transfer buffer size of the AFUFetchFile UDF:

The `AFUFetchFile` UDF loads the temporary XML files into memory for access by Net Search Extender.

The size of this memory buffer is set to the default value of 32 MB when the `AFUFetchFile` UDF is registered for the item type when the item type is enabled for processing by the indexer for text search. This is sufficient most of the time.

However, the situation can arise that this size is too small if the textual content that is extracted from an item is larger than 32 MB.

The `afuIndexer` process checks that there is enough space on the transfer buffer for each XML file. If there is not enough space, the item is not processed any further. The item is put in the table of completed tasks and is assigned an IDXRC of 130.

If this happens to many items, it is recommended that you increase the transfer buffer size by redefining the UDF to be able to handle larger amounts of data.

To redefine the `AFUFetchFile` UDF buffer size:

1. Edit the DB2 data definition language script `AFUFetchFile.ddl` in the `bin` subdirectory of the Content Collector Text Search Support installation. Change the CLOB value of 32 MB to a value that is appropriate.
2. Run the script with the following command:

   ```
   db2 -tvf AFUFetchFile.ddl
   ```

   The script drops the UDF and redefines it with a new return data size limit.

Running the configuration check tool:

The indexer for text search process `afuCheck` validates the indexer installation and configuration

The configuration check process looks for a default configuration for the item type name in the current directory and if found, will load it. If no default .ini file is found or if no .ini file is passed in an argument, the configuration check process will display an error message and stop processing.

It checks for all dependent libraries and that the item type has a valid configuration .ini file. In particular, it checks the Net Search Extender configuration, matches configuration settings against the Net Search Extender model file settings, and checks that the AFUFetchFile UDF is installed correctly.

`afuCheck` does not correct anything in the configuration, it only displays its findings.

Note: This validation process, without the additional checking for the dependent libraries, is automatically carried out each time `afuIndexer` and `afuRecreateIndex` are run. This means that `afuCheck` does not have to be called explicitly before `afuIndexer` and `afuRecreateIndex` are run.

To check an item type configuration:

- On UNIX and Linux, run

  ```
  afuCheck.sh
  ```

- On Windows, run

  ```
  afuCheck.cmd
  ```

Note: When you run the `afuCheck` process:

- All command line arguments are case sensitive.
- The command line arguments can appear in any order.
- Use the following delimiters for arguments that can pass more than one value:
  - On Windows, a semicolon (;)

– On UNIX and Linux machines, a colon (:)

- The item type name is not case sensitive.
- The indexer configuration options passed through an argument are not case sensitive.

If you run `AfuCheck` and do not specify an argument or enter the command followed by `-?`, the following is displayed:

```
afuCheck: Usage:
  [-?|-h|-copyright]
  -it <itemtype>
  -database <database>
  [-user <userID|@environmentVariable|:filename>]
  [-password <password|@environmentVariable|:filename>]
  [-schema <schema>]
  [-useconfig <.ini file>;<.ini file>]
  [-usesettings "<option=value>;<...>"]
  [-exportconfig <.ini file>]
  [-listconfig]
  [-verbose] [-showsql] [-trace]
  [-quiet]
```

where:

-? Lists all of the arguments that you can pass to the `afuCheck` tool.

-h Lists all of the arguments that you can pass to the `afuCheck` tool.

-copyright
Lists the copyright notice and version information.

-it <*itemtype*>
Required.

The Content Manager item type (document collection) that you want to index.

Important: If your item type is an email compound item type, do not enter the attachment item type. Attachment item types cannot be processed on their own but are automatically processed as part of the email item type.

-database <*database*>
Required.

The name of the library server database in which the item type is stored. The name of the database can be specified in two formats which affects the way the indexer tool connects to the database. The database connection is implemented using Java Data Base Connectivity (JDBC) which defines how a Java client may access a database. JDBC features different connection types and IBM Content Collector supports JDBC type 2 and JDBC type 4.

To use a JDBC type 4 connection, the specification for the database name of the argument `-database` must include a server name, a port number, and a database name in the following format `<databaseserver>:<portnumber>/<database name>`, for example `-database localhost:50000/icmnlsdb`. JDBC type 4 is the recommended database access method but always requires fully specifying the user ID and password.

To use a JDBC type 2 connection, the specification for the database name of the argument `-database` must name an existing DB2 connection name, in other words, the database name must be cataloged in DB2, for example `-database icmnlsdb`. As JDBC type 2 uses an existing database connection, no user ID and password need to be specified. However, a native 32-bit version of the DB2

JDBC Type 2 driver (shared library) must be available and defined in the class path for using the JDBC type 2 connection.

-user <*userID* | *@environmentVariable* | *:filename*>
Optional

Provide a user ID. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a user ID:

1. Specify the user ID directly on the command line.
2. Specify the name of an environment variable from which the user ID is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_USERID=user8

   afuCheck -user @INDEXER_USERID
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuCheck -user :myUserID.txt
   ```

   The text file must contain a single line with the user ID (no quotes and no spaces).

The advantage of using an environment variable or a file is that the user ID is not visible in any process status information.

The user must have access rights to IBM Content Manager.

-password <*userID* | *@environmentVariable* | *:filename*>
Optional

The password of the user ID with database administrator authority. Only required if no connection exists to the library server database or if JDBC connection type 4 is used for establishing a connection. There are three ways in which you can enter a password:

1. Specify the password directly on the command line.
2. Specify the name of an environment variable from which the password is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_PASSWORD=12345678

   afuCheck -password @INDEXER_PASSWORD
   ```
3. Specify the name of a text file that contains the password, for example,

   ```
   afuCheck -password :myPassword.txt
   ```

   The text file must contain a single line with password (no quotes and no spaces).

The advantage of using an environment variable or a file is that the password is not visible in any process status information.

Passwords are not logged in a readable plain text format.

-schema <*schema*>
Optional.

The name of the database schema if the schema is different from the current database schema or the configuration option **GenJbdcType** is set to 4.

-useconfig <*.ini file*>;<*.ini file*>
Optional.

Enables using alternate configuration .ini files or configuration files from a different path. Multiple .ini files can be specified and they will be loaded in the order in which they have been specified.

-usesettings *"<option=value>;<option=value>"*
Optional.

Enables the direct overwriting of indexer configuration values stored in the .ini file. You can add any number of configuration option value pairs <*option=value*> separated by the delimiter for your operating system. The option value pairs must be enclosed in double quotes. The configuration options you list (<*option*>) are not case sensitive. Using this argument is helpful at the beginning to verify the configuration and execution of the indexer.

-exportconfig <*.ini file*>
Optional.

Exports the indexer configuration to the .ini file specified in <*.ini file*>.

-listconfig
Optional.

Lists all the configuration options and their values that are set for the current indexer run.

-verbose
Optional.

Displays detailed processing information on the screen. Using **-verbose** is helpful at the beginning to get detailed information on all processing. If **-verbose** is specified the log files also contains additional information.

-showsql
Optional.

Displays all of the DB2 SQL commands executed by the system on the screen. Using this argument is helpful at the beginning to verify the configuration and execution of all DB2 system processing. If this argument is specified, all of the SQL statements are also added to the log file.

-trace
Optional.

Goes beyond verbose by displaying more detail in log files to support error determination. This is only required for IBM Support and requires the use of the trace property file.

-quiet
No processing steps are displayed on the console.

Constraints when indexing in batch mode:

Running the indexer process in a schedule on a batch of items is influenced by limitations that are defined for the batch size.

If the configuration option **IdxItemMinBatchSize** is set to 0, the indexing process attempts to automatically determine the minimum number of documents that must be in the table of open tasks before indexing will start. If the number of documents in the table of open tasks is smaller than this computed value, the indexer process stops with a return code indicating that there is nothing to process at present. If the number of documents found in the table of open tasks is higher than the computed value, the indexer will begin processing these items right away.

The formula for automatically computing the threshold for when to start processing items is as follows:

```
The arithmetic mean of the number of items added to the item type
for the last N days where N is the value of the configuration option
IdxDaysConsideredForMinBatchSize.
The default value (number of days) is 10.
```

The computation of this formula is based on content in the IBM Content Collector item type database table and might be expensive to compute because, if the database table does not have an index, a complete table scan might be required. To avoid a complete table scan, it is recommended that you create a database index for the item type database table to speed up subsequent computations of this formula.

To create a database table index, you need to know the database table name of the underlying item type. This table name is displayed next to the label `Item type component id table name` when the `afuIndexer` or `afuCheck` tool is invoked for the respective item type. Invoke `afuIndexer` or `afuCheck` using the `-verbose` command line option.

For example, the following information is displayed when one of the tools is invoked:

```
Item type component id table name.......: ICMUT01363001
```

Using this database table name you can create a database index using the following DB2 command:

```
DB2 CREATE INDEX <schema>.<indexname> ON
 <schema>.<tablename> ("CREATETS" ASC) ALLOW REVERSE SCANS;
```

If the configuration option **IdxItemMinBatchSize** is set to a value greater than 0, the indexer uses the specified value as the minimum batch size. If the number of documents found in the table of open tasks is smaller than the specified number, the indexer process stops with a return code indicating that there is nothing to process at present. If the number of documents found in the table of open tasks is greater than or equal to the specified value, the indexer will immediately start processing items.

If **IdxItemMinBatchSize** is set to 1, indexing will run even if the batch only contains as little as one item.

Note: The command line argument `–minbatchsize` temporarily overwrites the value in the configuration option **IdxItemMinBatchSize** and disables the automatic computation of the item batch size.

The `afuIndexer` process triggers Net Search Extender to begin indexing when the following conditions are met:

- The number of items to process exceeds the defined minimum batch size
- There are no more items in the table of open tasks
- The number of items that have been read from the table of open tasks equals the maximum item batch size set in the configuration option **IdxItemMaxBatchSize** or the maximum batch size passed through the `–maxbatchsize` argument when `afuIndexer` was called.

If an index reorganization operation was performed as part of the last index update operation, the maximum batch size will be reassessed automatically and increased by approximately a factor of 1.5 of the original maximum batch size.

Running `afuIndexer` in continuous mode:

The indexer process `afuIndexer` can be configured to run in a schedule on a defined batch size of items or to run in continuous mode until it is interrupted by an explicit shutdown command.

Running the indexer process in a schedule on a batch of items is the default processing mode.

To run the indexer process in continuous mode:

1. Run `afuIndexer` with the additional argument `-continuousmode`. The indexer process runs in endless iterations until a shutdown command is issued.

   In each iteration, the indexer process scans the table of open tasks for new items that need to be indexed and processes these. If the table of open tasks does not contain any new items during an iteration, the indexer process waits for about a minute before it automatically starts another iteration.

   The indexer behavior changes when the indexer process is started in continuous mode:

   - The configuration option **IdxDisplayTimeStamp** is automatically enabled.
   - **IdxDisplayProgressIndicator** is also automatically enabled. This means that progress monitoring is replaced by an overall status monitoring that displays:
     - Current timestamp
     - Current indexer statistics
     - Current batch size: The number of items that have already processed and are waiting to be indexed by Net Search Extender. This corresponds to the number of rows in the Net Search Extender log table.
     - Minimum batch size
     - Maximum batch size
     - Time elapsed since the indexer process started
   - The names of the generated log files include a generation marker showing when the log file was generated, and are limited in file size and in the number that is generated.
   - If a worker process stops, the process is restarted automatically without stopping the indexer process. If the indexer process runs in continuous mode, this is allowed to happen up to 1000 times before the indexer process is stopped. If the indexer process is run in batch mode in contrast, restarting the worker process automatically is limited to 100 times.
   - The maximum batch size that is defined in the configuration option **IdxItemMaxBatchSize** is recomputed after each Net Search Extender indexing phase for each batch. Recomputing the maximum batch size regularly is necessary because the index may have been reorganized during any one of the many indexing phases when `afuIndexer` runs in continuous mode.

     If you specify the maximum batch size using the `afuIndexer` command line argument `-maxbatchsize`, the maximum batch size is not recomputed. Hence, it is recommended that you do not use the command line argument `-maxbatchsize` when the indexer runs in continuous mode.

- The minimum batch size that is defined in the configuration option **IdxItemMinBatchSize** has a different affect on the indexer processing when `afuIndexer` runs in continuous mode:
  - The minimum batch size is only relevant if `afuIndexer` is stopped using the argument `-shutdown complete`. Net Search Extender will run if the current batch size is higher or equal to the computed minimum batch size.
  - The minimum batch size is automatically recomputed every 24 hours if **IdxItemMinBatchSize** is set to 0 and the computed value is dependent on the average item load during a defined period of N days.

    It is recommended that you do not use the command line argument `-minbatchsize` to set a value, as this will inhibit the minimum batch size from being recomputed automatically.
- The configuration option **IdxAutoIndexingEveryNDays** allows starting Net Search Extender irrespective of any batch size specifications. The default value of **IdxAutoIndexingEveryNDays** is 7. If the archiving load suddenly drops, **IdxAutoIndexingEveryNDays** ensures that all archived content is indexed even if the computed minimum and maximum batch sizes which typically starts Net Search Extender have not been reached. If **IdxAutoIndexingEveryNDays** is set to 0, Net Search Extender only starts when the maximum batch size is reached which means that, if the archiving load is low, items might not be indexed for quite some time.
- The command line arguments `-continuousmode` and `-debug` cannot be used at the same time.
- The command line argument `-continuousmode` and `-nseindexreorg` cannot be used at the same time.
- If the indexer process is shut down while it is running in continuous mode, the indexer process return code is not 102 which means that the indexer process was stopped due to a shutdown request. Instead 0 or 1 is returned reflecting if items were processed or not processed.
- Items that produced an error or an unrecoverable error (for example, if processing is stopped unexpectedly) are moved to the table of completed tasks during the next processing iteration after Net Search Extender has run. This corresponds to two subsequent indexer runs in batch mode.

2. Stop the process by using one of the following arguments:

Option	Description
`afuIndexer -shutdown complete`	• If the current batch size is greater than or equal to the minimum batch size, the indexer will complete processing all items in the current batch and index these items using Net Search Extender before processing is stopped. • If the current batch size is smaller than the minimum batch size, the indexer will stop processing as soon as possible. Net Search Extender is not called.
`afuIndexer -shutdown force`	Irrespective of the current batch size, the indexer will stop processing immediately. Net Search Extender is not called.

Processing email without attachments:

If your use case emphasis is on email content only and not on email attachment content, you can configure the indexer process to omit processing and indexing email attachment content.

To process email without its attachments:

Process email content ignoring all embedded attachment content by defining one of the following configuration options:

Option	Description
• **GenFileTypeFilterIsBlackList** = 0 • The `[filetypefilter]` section in the indexer configuration file is empty	• The email item is indexed but all embedded attachments are skipped. The item in the table of completed tasks is assigned an IDXRC value of 0. • If **XmlInsertAttachmentProcessingWarnings** = 1, the notification string `IcmFceWarning: IcmConfigFilteringFile` is indexed with the item. • If **XmlInsertAttachmentProcessingWarnings** = 0, no notification string is indexed.
TxtCnvEnabled = 0	• The email item is indexed but all embedded attachments are skipped. The item in the table of completed tasks is assigned an IDXRC value of 50. • If **XmlInsertAttachmentProcessingWarnings** = 1, the notification string `IcmFceError: IcmConversionDisabled` is indexed with the item. • If **XmlInsertAttachmentProcessingWarnings** = 0, no notification string is indexed.

Processing more than one item type at the same time:

The indexer for text search process can only run on one item type at a time, but multiple item types can be processed simultaneously.

To process multiple item types at the same time however, each indexer process instance must use a free and unique port number.

To guarantee that each instance has a unique port:

Set a distinct port value for each `afuIndexer` instance:

Option	Description
Process argument	Run: `afuIndexer -port <port number>` The value passed through the argument `-port` overwrites the value set in the environment variable **AFU_INDEXER_RMI_PORT**.

Option	Description
Environment variable	Set **AFU_INDEXER_RMI_PORT** for each instance before starting `afuIndexer`

Embedded email attachment and document processing:

To index documents for text search, the textual content must first be extracted from the documents.

The Oracle/Stellent Outside In Technology filters are used to extract textual information from all embedded email attachments, and File System and Microsoft SharePoint documents.

The term document is used in this topic to refer to both embedded email attachments, and File System and Microsoft SharePoint documents.

Not all documents in the IBM Content Manager Resource Manager are documents that contain textual information that must be indexed to enable text search. This means that those documents that do not contain textual information must be identified and filtered before they are indexed by Net Search Extender. This can be done in one of two ways:

- By file name extension:

  A list of file name extensions can be configured for document processing. Depending on the setting of the configuration option **GenFileTypeFilterIsBlackList**, the list of file name extensions is considered to be a blacklist, or a whitelist.

  If the configuration option is set to 1, the list of file name extensions listed in the section `[FileTypeFilter]` in the configuration is considered to be a blacklist. This means that all documents in the IBM Content Manager Resource Manager with file name extensions in this list are skipped and not processed by the text extraction filters.

  If the option is set to 0, the list of file name extensions in the section `[FileTypeFilter]` is considered to be a whitelist. This means that only those documents in the IBM Content Manager Resource Manager that have a file name extension listed in the section FileTypeFiles will be retrieved.

  The set of file types that is used for filtering must be stored in a separate section in the configuration .ini file called `[FileTypeFilter]`. For each file type or extension, a separate option named `extension =` must be specified. For example.

  ```
  [FileTypeFilter]
  extension = .JPE
  extension = .JPEG
  extension = .JPG
  extension = .PJP
  extension = .PJPEG
  ...
  ```

  The sample file type filter file that is included in the text search component is a blacklist.

- By the Oracle/Stellent Outside In Technology filters file type:

  If the configuration option **TxtCnvOutsideInFilterGroups** is enabled, all documents that are identified by the Oracle/Stellent Outside In Technology filters as belonging to one of the following groups of file types will be filtered, in other words, those documents will not be processed for text extraction:

– BMP, TIFF, PCX, GIF, EPSTIFF, CCITTGRP3, MACPICT2, WPG, WINDOWSMETA, LOTUSPIC, MACPICT1, AMIDRAW, TARGA, GEMIMG, OS2DIB, WINDOWSICON, WINDOWSCURSOR, MICROGRAFX, MACPAINT, CORELDRAW2, CORELDRAW3, HPGL, HARVARDDOS3, HARVARDDOS2, HARVARDDOS3PRS, FREELANCE, WPG2, CGM
– MultiMedia
– Graphic2
– Bitmap
– The Oracle/Stellent Outside In Technology filters group *Reserved*
– The Oracle/Stellent Outside In Technology filters group *OEM*
– The Oracle/Stellent Outside In Technology filters group *Other*

Document conversion using the Oracle/Stellent Outside In Technology filters is influenced by enabling or disabling the configuration option **TxtCnvEnable**:

Table 178. Document conversion using the Oracle/Stellent Outside In Technology filters when the document format is identified or no conversion is requested

Configuration option **TxtCnvEnable** = 1 or not specified **Result** • The document is converted. • The item is added to the table of completed tasks with IDXRC = 0.
Configuration option **TxtCnvEnable** = 0 **Result** • No document conversion takes place. • The notification string `IcmFceWarning: IcmConversionDisabled` is added to the textual content that is indexed for the item.

When **TxtCnvEnable** is enabled, the following configuration option come into play:

- **TxtCnvPlainTextHandlingForUnsupportedFormat**
- **TxtCnvAutoCodepageDetection**
- **TxtCnvAutoCodepageDetectionMinConfidence**
- **TxtCnvMaxExtendedCharactersRatio**
- **TxtCnvDefaultCodepageForPlainText**

The values of these options have no significance when **TxtCnvEnable** is disabled.

The following table shows what happens if the document format cannot be identified by the Oracle/Stellent Outside In Technology filters or if the document cannot be processed at all.

Table 179. Document conversion when the Oracle/Stellent Outside In Technology filters cannot identify the document format

Configuration option • **TxtCnvPlainTextHandlingForUnsupportedFormat** = 0 **Result** • No document conversion takes place. • The item is added to the table of completed tasks with IDXRC = 4, (`WARN_FORMAT_NOT_SUPPORTED`) • The notification string `IcmFceError: InsoErrorNoFilterAvailable` is added to the textual content that is indexed for the item.
Configuration option • **TxtCnvPlainTextHandlingForUnsupportedFormat** = 1 • **TxtCnvAutoCodepageDetection** = 1 • **TxtCnvDefaultCodepageForPlainText** = 0 **Result** • The document is processed and an attempt is made at recognizing the character encoding (code page) of the document. • If the confidence level with which the code page was detected is equal to or higher than the value specified in **TxtCnvAutoCodepageDetectionMinConfidence**: – The document is converted. – The item is added to the table of completed tasks with IDXRC = 0. – The notification string `IcmFceWarning: IcmConversionAutoCodePage` is added to the textual content that is indexed for the item. • If the confidence level with which the code page was detected is less than the value specified in **TxtCnvAutoCodepageDetectionMinConfidence**: – The item is processed as if **TxtCnvAutoCodepageDetection** is set to 0. **Important:** Test results show that the default value of **TxtCnvAutoCodepageDetectionMinConfidence** is optimal and you are encouraged not to change it. However, if you want to ensure that no document is ever incorrectly converted, increase the value to around 90.

Table 179. Document conversion when the Oracle/Stellent Outside In Technology filters cannot identify the document format (continued)

Configuration option

- **TxtCnvPlainTextHandlingForUnsupportedFormat** = 1
- **TxtCnvAutoCodepageDetection** = 0
- **TxtCnvDefaultCodepageForPlainText** = 0

Result

- If the file size is 0:
  - No document conversion takes place.
  - The item is added to the table of completed tasks with IDXRC = 6.
  - The notification string `IcmFceError: IcmEmptyFile` is added to the textual content that is indexed for the item.
- If the file size is greater than 0:
  - The document is processed and all binary characters (code point < 0x05) and extended characters (code point > 0x7f) in the document are counted.
  - If binary characters are found:
    - No document conversion takes place.
    - The item is added to the table of completed tasks with IDXRC = 5, (`WARN_BINARY`)
    - The notification string `IcmFceError: InsoErrorNoFilterAvailable` is added to the textual content that is indexed for the item.
  - If no binary characters are found:
    - If extended characters are found and the percentage is lower than the value in **TxtCnvMaxExtendedCharactersRatio**:
      - The document is converted.
      - The item is added to the table of completed tasks with IDXRC = 0.
      - The notification string `IcmFceWarning: IcmConversionPlainText` is added to the textual content that is indexed for the item.
    - If extended characters are found and the percentage is higher than the value in **TxtCnvMaxExtendedCharactersRatio**:
      - No document conversion takes place.
      - The item is added to the table of completed tasks with IDXRC = 2.
      - The notification string `IcmFceError: IcmConversionFailedTooManyExtendedChars` is added to the textual content that is indexed for the item.

Table 179. Document conversion when the Oracle/Stellent Outside In Technology filters cannot identify the document format (continued)

Configuration option

- **TxtCnvPlainTextHandlingForUnsupportedFormat** = 1
- **TxtCnvAutoCodepageDetection** = 0 or 1
- **TxtCnvMaxExtendedCharactersRatio** = 100
- **TxtCnvDefaultCodepageForPlainText** = <*CCSID*>

Result

- If **TxtCnvAutoCodepageDetection** is disabled or code page detection fails:
  - If the document contains no binary characters:
    - The document is converted using the code page specified in **TxtCnvDefaultCodepageForPlainText**.
    - The item is added to the table of completed tasks with IDXRC = 0.
    - The notification string `IcmFceWarning: IcmConversionDefaultCodePage` is added to the textual content that is indexed for the item.
  - If the code page specified in **TxtCnvDefaultCodepageForPlainText** is not supported:
    - No conversion takes place.
    - The item is added to the table of completed tasks with IDXRC = 3.
    - The notification string `IcmFceError: IcmConversionFailedCodepageNotSupported` is added to the textual content that is indexed for the item.
  - If the code page specified in **TxtCnvDefaultCodepageForPlainText** is supported:
    - The document is converted using the specified code page.
    - The item is added to the table of completed tasks with IDXRC = 0.
    - The notification string `IcmFceWarning: IcmConversionDefaultCodePage` is added to the textual content that is indexed for the item.

Important:

- Setting **TxtCnvDefaultCodepageForPlainText** to a default code page is only recommended if the embedded plain text attachments are likely to all be of the same character encoding. This setting is not recommended if you expect to have to process plain text documents with varying character encodings.
- If you set **TxtCnvDefaultCodepageForPlainText** to a default code page, you should set **TxtCnvMaxExtendedCharactersRatio** to 100.

Sample script for evaluating the indexer process return codes:

IBM Content Collector provides a sample script to use to evaluate the indexer process return code.

The sample scripts for different platforms can be found in the subdirectory `samples` where IBM Content Collector Text Search Support is installed on the IBM Content Manager server.

The script lists the error codes that can be returned by the indexer system process. If you decide to use the sample script, complete the script by adding the steps of action to take for each error or for a group of errors. Only the return codes 0 and 1 do not require any user action.

This is a UNIX sample script and is structured as follows:

```
# This sample script file can be used for embedding the afuIndexer script
# and branch according to the return code ("$?")
#
# ************************************************************************
AFU_INDEXER_RC_NO_ERROR=0
AFU_INDEXER_RC_NO_ACTION_PERFORMED=1
AFU_INDEXER_RC_CONFIG_ERROR=2
AFU_INDEXER_RC_LOGGING_ERROR=3
AFU_INDEXER_RC_INDEXING_ERROR=4
AFU_INDEXER_RC_NSE_ERROR=5
AFU_INDEXER_RC_SQL_ERROR=6
AFU_INDEXER_RC_IO_ERROR=7
AFU_INDEXER_RC_GENERAL_EXCEPTION=8
AFU_INDEXER_RC_RMI_ERROR=9
AFU_INDEXER_RC_OUT_OF_DISK_SPACE=10
AFU_INDEXER_RC_OUT_OF_MEMORY=11
AFU_INDEXER_RC_USER_COMMAND_ERROR=50
AFU_INDEXER_RC_TOO_MANY_ERRORS=60
AFU_INDEXER_RC_ALREADY_RUNNING=100
AFU_INDEXER_RC_NOT_RUNNING=101
AFU_INDEXER_RC_SHUTDOWN=102
AFU_INDEXER_RC_GENERAL_COMMAND_LINE_ERROR=200

# launch the indexer
afuIndexer $*
AFU_INDEXER_RC=$?

# echo afuIndexer RC: $AFU_INDEXER_RC

# determine action depending on return code
if [[ $AFU_INDEXER_RC >= $AFU_INDEXER_RC_GENERAL_COMMAND_LINE_ERROR ]] ; then
 echo IBM afuIndexer was launched with incorrect command line arguments
   # ...

elif [[ $AFU_INDEXER_RC == $AFU_INDEXER_RC_NOT_RUNNING ]] ; then
 echo IBM afuIndexer terminated with: NOT_RUNNING ($AFU_INDEXER_RC_NOT_RUNNING)
  # ...

elif [[ $AFU_INDEXER_RC == $AFU_INDEXER_RC_ALREADY_RUNNING ]] ; then
 echo IBM afuIndexer terminated with: ALREADY_RUNNING
                                  ($AFU_INDEXER_RC_ALREADY_RUNNING)
 # ...

elif [[ $AFU_INDEXER_RC == $AFU_INDEXER_RC_TOO_MANY_ERRORS ]] ; then
 echo IBM afuIndexer terminated with: TOO_MANY_ERRORS
                                  ($AFU_INDEXER_RC_TOO_MANY_ERRORS)
      # ...
...
```

The following codes are returned by the indexer process when the process completes. For more detail about the returned system process code, refer to the log files as well.

Table 180. Text search indexer process return codes

Return code	Description
0	Indexing process was successful, there were no errors of any kind.
1	No items were found to process. This is returned if no new documents were archived since the last time that the indexing process ran for the item type, if **IdxDelayDocumentprocessingInDays** was set to a number of days that has not yet elapsed, or if the minimum batch size that triggers the indexing process and is set in the option **IdxItemsMinBatchSize** was not satisfied.

Table 180. Text search indexer process return codes (continued)

Return code	Description
2	A severe error occurred due to a indexing process setup or configuration issue.
3	A severe error occurred in the context of logging.
4	A severe error occurred during document processing before Net Search Extender is called.
5	A severe Net Search Extender error was encountered.
6	A severe DB2 processing error was encountered.
7	A severe file access (file path or file rights) error was detected.
8	A severe general processing error was encountered that does not match any of the other errors from above.
9	A severe communication error was encountered. The indexer process could not set up communication with the worker process.
10	The indexer process detected that there is not enough file system space.
11	The indexer process or the worker process have detected an out of memory situation.
50	One of the user-defined commands that can be called before or after Net Search Extender is started, returned with a code that is not zero.
60	Too many errors have occurred. The number of subsequent IBM Content Manager Resource Manager retrieval errors that occurred exceeded the limit that is specified in the configuration option **IdxMaxNumberOfSubsequentRmErrors**
100	The indexer process is already running for the specified item type.
101	The indexer was run with a shutdown command line argument but no running instance was found for the specified item type.
102	The indexer process was stopped due to a shutdown request.
200	The indexer process was called using an incorrect command line syntax.
201	The indexer process was called with the -copyright argument. No action was performed.
202	The indexer process was called with an unknown argument.
203	A required argument was missing when the indexer process was called.
204	An argument that requires a string value was passed to the indexer process without a value.
205	An argument that requires an integer value was passed to the indexer process without a value.
207	The indexer process was called with the same argument more than once.
208	The value of an argument that should have been numeric was not numeric.
210	The indexer process was called with conflicting arguments.
211	The indexer was called with missing arguments.

Reindexing documents:

Use the indexer process reindexing arguments when you next run the indexing to attempt reprocessing items that ran against specific errors, or were indexed during a particular period before significant changes or adaptations were made to the system.

After you have grouped your errors into warning errors and more serious errors, you can choose between different reindexing options that you can pass `afuIndexer` the next time the process is scheduled to run.

Remember: Reindexing items that caused warnings or errors is only effective if index configurations were changed or corrected, for example, if timeout intervals were adjusted, or new updates were installed, such as if additional document filters were added, existing filters were enhanced, or a new version of Net Search Extender was installed. Severe errors typically occur if the indexer process stopped unexpectedly while processing an item because the item was corrupted or otherwise malformed in an unexpected way. These items should be analyzed and possibly corrected before a new attempt is made to index them. Severe errors can also result if the process ran out of memory

The reindexing options include:

- Reindexing individual items. Only items for the item IDs listed will be reprocessed.
- Reindexing items that have a specific IDXRC value. For a list of all values stored in column IDXRC of the table of completed tasks, see "Index return codes stored in the table of completed tasks" on page 604.
- Reindexing all items that have a warning IDXRC value as a group. This serves as a shortcut to the warning IDXRC range.
- Reindexing all items that have an error IDXRC value as a group. This serves as a shortcut to the error IDXRC range.
- Reindexing all items that have an IDXRC value of 200 as a group. This serves as a shortcut to all severe errors. See "Steps to take if the indexer process stops unexpectedly" on page 657.

Reindexing always requires searching the entire database table of completed tasks for items that meet the requested reindexing conditions and can take considerable time to complete depending on the amount of documents that have already been indexed. A complete table scan is necessary unless additional database indexes are created manually.

If you specify any of the reindexing arguments, the items that match the reindexing criteria will be added to the table of open tasks, ready for processing. They are not removed from the table of completed tasks. These entries in the table of completed tasks are overwritten after the items have been reprocessed.

There is however no guarantee that these items will be processed during the same indexer run in which they were identified to be reindexed. They might also be processed during a later indexer run. If the reindexing criteria do not match any items in the table of completed tasks, a warning is issued that no items could be found that match the reindexing criteria but indexing will continue.

Important: Only documents that have been processed using the indexer for text search can be reindexed. Documents that were processed using the fast indexer or the standard Content Manager indexer with the IBM Text Search user exit cannot be reindexed using these reindexing arguments. These documents can only be reindexed manually by adding a new row to the table of open tasks for each item.

Processing items again includes:

- Reindexing items with warnings or errors:

Option	Description
Individual items	Run `afuIndexer -reindexitems <itemID>;<itemID>`
All items with warnings	Run `afuIndexer -reindexwarnings`
All items with errors	Run `afuIndexer -reindexerrors`
All items with severe errors	Run `afuIndexer -reindexsevereerrors`
Items with a specific IDXRC value	Run `afuIndexer -reindexitems <IDXRC>`

- Reindexing items that were indexed during a particular time period after which changes were made to the processing environment that influence document processing and indexing outcome. Run

  `afuIndexer -reindexdate <yyyymmdd>`

  where *<yyyymmdd>* is the archive date of the item.

  All items with an archive date of *<yyyymmdd>* or later are reindexed.

Related reference

"Index return codes stored in the table of completed tasks" on page 604

Steps to take if the indexer process stops unexpectedly:

If the indexer process stops unexpectedly before all the items in the batch of items have been processed, or if CTRL-BREAK is pressed during processing, the indexer process must be restarted manually.

To ensure robust 24x7 performance, the indexer process uses a separate worker process to deal with all documents even if some of these documents are malformed and as a consequence destroy the process's memory structure or cause the process to terminate unexpectedly. This worker process uses Oracle/Stellent Outside In Technology filters to convert documents to their textual content. Using a separate worker process is explicitly recommended by Oracle/Stellent Outside In Technology filters because the quality of the documents that are processed is unknown and malformed documents can lead to unexpected process behaviour.

Note: If the indexer worker process stops unexpectedly on UNIX or Linux, the operating system usually creates a core file. By default, the indexer process automatically removes the core file. To suppress the creation of a core file, issue the `ulimit -c 0` command within the shell in which the indexer process is started. For details, refer to the operating system manual.

To prevent multiple instances of the same item type, the indexer creates a file for each item type in the file system as soon as the indexer process is started. Normally, this control file is removed as soon as processing has completed. However, if the indexer process stops unexpectedly or is interrupted, this file is not deleted. If the indexer is started again for the same item type, the indexer process displays a message stating that an instance is already running. The message includes the name of the control file and the process return code 100.

The environment variable **AFU_INDEXER_PROMPT_RUN** influences this behavior. If the variable is set to 1, the indexer process will not terminate with a return code of 100 if it detects that an indexer process is already running for the same item type but instead will prompt the user about whether to continue and ignore the return code. Naturally, prompting the user holds up the process.

If the indexer process stops unexpectedly or CTRL-BREAK is pressed during processing:

1. Delete the instance control file that appears in the error message displayed by the indexer process.
2. Run `afuIndexer` again with the same arguments as before the process was interrupted.

Processing items with a severe error:

Items that have a severe error might be the cause for the worker process stopping unexpectedly during indexing resulting in the indexer process stopping. During a subsequent run, those items are identified by the indexer as potentially being responsible for the worker process stopping.

The next time the indexer process is run, those items that could have been the cause are identified.

To try and find which items produced a severe error, follow these steps:

1. Stop archiving to the item type.
2. Run `afuIndexer` again with the same arguments as before the indexer process stopped. Run the indexer until all of the items that are in the table of open tasks are processed (the table of open tasks must be empty). The items that might have caused the process to stop will not be reprocessed. They will be added to the table of completed tasks with an IDXRC of 200.
3. To attempt reprocessing the items with an IDXRC of 200, run `afuIndexer -reindexsevereerrors -usesettings "IdxNumberOfParallelIndexerTasks=1; GenDump=1"`

   Tip: It is recommended that you reduce the number of processing threads to one while reindexing the items with severe errors.
4. Restart archiving to the item type if processing completes successfully. Otherwise report this error to IBM Support and include the original documents that caused the severe errors. These files are located in the directory specified in the **GenDumpDirectory** configuration option.

Debugging the processing of an item:

When the indexer process is run in debug mode, the indexer switches into a processing mode that supports a subsequent careful examination of the documents that were processed.

As the debug option requires user interaction, `afuIndexer` cannot be run in batch mode. The debug processing mode affects the indexer process as follows:

- The indexer runs in single process mode. No separate worker process is started.
- The indexer worker runs in single thread mode, not in parallel mode. This means that tasks are processed one at a time.
- The following configuration options are automatically changed:
  - The dump option is enabled.
  - Logging is enabled. The log file names do not include a timestamp which means that all log data is written to one file which is distinct from all other log files and is overwritten by the next debug session.
  - The processing time for each item is not recorded as timing is disabled.

  - The temporary XML files are not deleted after indexing. You can delete these files manually later in the directory specified by the configuration option **GenXmlDirectory**.
  - Timeouts are increased and internal timing values are adapted.
  - No backup command or any other configured commands are run.
- If the environment variable AFU_INDEXER_XML_VIEWER is defined, the user is prompted for each item about if the temporary XML files should be displayed. For details, see the description of the indexer environment variable AFU_INDEXER_XML_VIEWER.
- Additional comments are added to the content of the XML files that include the IDXRC value and attribute names. These comments may help you to understand how the indexer processed the original document. These comments are ignored by Net Search Extender and not added to the full text index.
- The user is prompted for each item for confirmation about whether the item is to be added to the Net Search Extender full text index, in other words, if the item is to be indexed.
- Debug mode switches on the effects of the **-verbose** command line arguments.
- Exceptions and errors are displayed in the full context. This may include a stack trace.
- As specified items are automatically reindexed when `afuIndexer` is run in debug mode, non of the reindex arguments can be used.
- The command line arguments `-debug` and `-nseindexreorg` cannot be used at the same time.
- The command line arguments `-debug` and `-continuousmode` cannot be used at the same time.

Related reference

"Text indexer environment variables" on page 621

Retrieving items from the IBM Content Manager Resource Manager:

To debug items that cannot be processed correctly by the indexer process, the original documents might be required to help determine the cause particularly if IBM Support is requested.

The original documents reside in the IBM Content Manager Resource Manager.

To access these documents:

Extract the original documents from the IBM Content Manager Resource Manager to a local disk:

Option	Description
Using the `-debug` argument	1. Run `afuIndexer -debug <item ID>;<item ID>` where <item ID> is the item that could not be processed correctly. If you run `afuIndexer` with the argument `-debug`, the **GenDump** configuration option is automatically enabled.

Option	Description
Using the `-reindexitems` argument	1. Enable the configuration option **GenDump** or use the indexer argument `-usesettings "GenDump=1"`. 2. Run `afuIndexer -reindexitems <item ID>;<item ID>` where <item ID> is the item that could not be processed correctly. 3. Disable the configuration option **GenDump**.

Note: The original documents are written to the directory specified by the configuration option **GenDumpDirectory**.

Stopping a running IBM Content Collector indexer process prematurely or at a specific point in time:

If you need to stop a running process for a specific item type prematurely, for instance at a specific time or after a specific time interval, you can issue a shutdown request against a running indexer process or an indexer process running in continuous mode.

To stop the process, start a second instance of the `afuIndexer` process for the same item type and include the shutdown command line argument. Two different shutdown modes are available:

- Complete
- Force

For a complete shutdown, start a new `afuIndexer` process by using the following command line argument:

```
-shutdown complete
```

This instance of the `afuIndexer` process will carry out the following tasks:

- Stop reading more items from the table of open task
- Complete processing all items for all active tasks
- Run Net Search Extender to update the full text index with the items that were processed so far if the current batch size exceeds the minimum batch size

This means that the indexer tool will end processing the item type as if the processing task had reached the maximum batch size limit.

For a forced shutdown, start a new `afuIndexer` process by using the following command line argument:

```
-shutdown force
```

This instance of the `afuIndexer` process will carry out the following tasks:

- Stop reading more items from the table of open task
- Stop processing items for all active tasks
- Terminate without running Net Search Extender to update the full text index

This means that all work performed for the item type during this indexer run is lost and will be repeated the next time the indexer is run.

Attention: If you plan to shut down a running indexer process for an item type at a specific point in time and you issue a `-shutdown complete` request (for example, through a cron job), remember that a considerable amount of time can elapse before indexing will stop after the shutdown request has been recognized by the running indexer process:

- Completing all active tasks might take time, even a few minutes, depending on the system performance, the number of active tasks with documents in progress, and the setting of the configuration option **TxtCnvTimeOutInSecs**.
- The subsequent Net Search Extender call to update the full text index also takes time. Depending on the number of items that need to be processed, this call can take considerable time, sometimes up to a few hours.
- The shutdown command has no impact on the execution of the Net Search Extender update process. This means that if a shutdown command is issued after Net Search Extender has already started, Net Search Extender will continue processing the batch. Only after Net Search Extender has finished processing the batch, will the shutdown request be executed.

If the `afuIndexer` process is shut down while it is running in batch mode, the process will terminate with an process return code of 102 which indicates that the indexer process was stopped due to a shutdown request. If the `afuIndexer` process is shut down while running in continuous mode, the process will terminate either with an process return code 0 indicating that processing was successful or with another return code indicating that no items were processed or that an error occurred.

Index reorganization:

The more updates that are made to a Net Search Extender index and the more index data accumulates, the more the internal index structures become fragmented and require more disk I/O operations for noncontiguous data structures. As a consequence the index search performance drops.

To improve the index search performance, an index must be reorganized periodically. The `afuIndexer` process attempts to determine when to begin an index reorganization automatically using configuration settings which help to track the number of documents that were indexed by the indexing engine Net Search Extender for each item type.

If and when an index reorganization process must be started is influenced by the values of the following configuration options:

- **IdxReorgBatchSize**:

  If the number of documents indexed by Net Search Extender in previous indexer runs exceeds the number of items set in the configuration option **IdxReorgBatchSize**, an index reorganization is attempted.

  The default value for **IdxReorgBatchSize** is 1.5 million items.
- **IdxMaxNumberOfUpdatesWithoutReorg**:

  The value of **IdxMaxNumberOfUpdatesWithoutReorg** prevents the number of items that are added to the index from growing endlessly without an index reorganization being performed. This can happen accidently if the indexer is scheduled to run during the blocking time frame defined by the options **IdxReorgTimeBlockIntervalBegin** and **IdxReorgTimeBlockIntervalEnd**

As soon as the number of items added to the index has reached the number specified in the option **IdxMaxNumberOfUpdatesWithoutReorg** and if no index reorganization has been performed so far, the indexer will refuse to index any further items and will stop processing with an error. The index reorganization operation must be started manually by running the `afuIndexer` with the command line argument `-nseindexreorg`.

- **GenNseIndexDirectoryMinFreeSizeRequiredInMB**:

  Before a reorganization is attempted, the indexer process checks if there is enough free disk space. The minimum amount that must be available on the Net Search Extender index directories is determined by the value of **GenNseIndexDirectoryMinFreeSizeRequiredInMB**.

  If there is enough free disk space on these index directories, an index reorganization is attempted. If there is not enough free disk space, the index cannot be reorganized. An error message is displayed and logged.

- **IdxReorgTimeBlockIntervalBegin** and **IdxReorgTimeBlockIntervalEnd**:

  These configuration options can be used to prevent an index reorganization during a specific time frame. As the index reorganization operation is time intensive and also blocks searching the index for a period of time, it might be useful to prevent running an index reorganization at certain times, for example, during the main business hours. By setting these options, a time frame can be defined during which no index reorganization is allowed to run.

  If these options are specified, the reorganization operation is only performed if the current processing time is not within the time interval specified by the two options. If the processing time is in this time frame, a warning is displayed and logged, and the reorganization operation is postponed until the next time the indexer runs.

  If these options are not specified, the reorganization operation is not blocked by a time frame.

If the `afuIndexer` command line argument `–nseindexreorg` is specified, an index reorganization will be performed as part of the index update operation irrespective of the value of **IdxReorgBatchSize** or options specifying a time frame block. Only the value of the option **GenNseIndexDirectoryMinFreeSizeRequiredInMB** is checked and could prevent an index reorganization from taking place.

Note: If a reorganization was performed as part of the last index update operation, the maximum batch size is adjusted (increased) automatically.

Searching for error strings:

In addition to the value stored in column IDXRC of the table of completed tasks indicating whether indexing was successful or resulted in an error or warning, a notification string is indexed with the item if a warning or error was detected.

The notification string that is produced as the result of a warning begins with `IcmFceWarning:` followed by a warning message, for example `IcmFceWarning: IcmConfigFilteringFile`. A warning notification string is indexed along with an item if the configuration option **XmlInsertAttachmentProcessingWarnings** is set to 1 which is the recommended setting.

The notification string that is produced as the result of a document error begins with `IcmFceError:` followed by an error message, for example `IcmFceError: IcmConversionFailedTooManyExtendedChars`. Error notification strings are always indexed along with an item.

These warning and error strings were also generated by the text search user exit in earlier Content Collector and CommonStore versions and were stored with the item in the index.

You can use these message strings in a search query to identify documents that were indexed with these message strings and can then reindex these documents, for example, if new document filters were installed that should resolve former document conversion problems.

The following warning message strings are indexed with an item and can be searched:

Table 181. The indexer warning strings

Notification string `IcmFceWarning: IcmConfigFilteringFile` **Return code and description** • IDXRC = 0 • The document, or one of the embedded attachments of the item, was filtered with the help of a file type filter specification or the configuration option **TxtCnvOutsideInFilterGroups** was enabled.
Notification string `IcmFceWarning: IcmConversionPlainText` **Return code and description** • IDXRC = 0 • The automatic code page detection or the analysis of extended characters showed that the document, or one of the embedded attachments of the item, is a plain text document.
Notification string `IcmFceWarning: IcmConversionAutoCodePage` **Return code and description** • IDXRC = 0 • The document was subjected to automatic code page detection.
Notification string `IcmFceWarning: IcmConversionFailedNoText` **Return code and description** • IDXRC = 0 • One of the embedded attachments, or a document, did not contain any text, for example it was a graphics file.

Table 181. The indexer warning strings (continued)

Notification string `IcmFceWarning: IcmConversionDefaultCodePage` **Return code and description** • IDXRC = 0 • The document encoding was assumed to be in the default code page and converted.
Notification string `IcmFceWarning: IcmConversionFailedProtected` **Return code and description** • IDXRC = 0 • The document, or one of the embedded attachments of the item, could not be handled by Oracle/Stellent Outside In Technology filters because it was protected or encrypted.
Notification string `IcmFceWarning: IcmUnhandledOLEObject` **Return code and description** • IDXRC = 0 • The item was indexed but the item contained embedded OLE objects which could not be processed.
Notification string `IcmFceWarning: IcmUnhandledEmbeddedMsg` **Return code and description** • IDXRC = 0 • The item was indexed but the item contained embedded MSG objects which could not be processed.
Notification string `IcmFceWarning: IcmConversionDisabled` **Return code and description** • IDXRC = 0 • The document, or one of the embedded attachments of the item, is not processed at all because the configuration option **TxtCnvEnabled** is disabled.

The following error message strings are indexed with an item and can be searched:

Table 182. The indexer error strings

Notification string `IcmFceError: InsoErrorNotEnoughMemory` **Return code and description** • IDXRC = 1 • The document, or one of the embedded attachments of the item, could not be processed completely due to memory constraints and was only partially indexed.
Notification string `IcmFceError: IcmConversionFailedTooManyExtendedChars` **Return code and description** • IDXRC = 2 • The document, or one of the embedded attachments of the item, could not be handled by the Oracle/Stellent Outside In Technology filters. The configuration option **TxtCnvPlainTextHandlingForUnsupportedFormat** was enabled but the ratio of extended characters (characters > 0x7F) exceeded the value specified in the configuration option **TxtCnvMaxExtendedCharactersRatio**.
Notification string `IcmFceError: IcmConversionFailedCodepageNotSupported` **Return code and description** • IDXRC =3 • The document encoding was specified to be in a default code page that is not supported.
Notification string `IcmFceError: InsoErrorNoFilterAvailable` **Return code and description** • **TxtCnvPlainTextHandlingForUnsupportedFormat** = 0 • IDXRC = 4 • The document, or one of the embedded attachments of the item, could not be handled by the Oracle/Stellent Outside In Technology filters because the document format is not supported by the Oracle/Stellent Outside In Technology filters.
Notification string `IcmFceError: InsoErrorNoFilterAvailable` **Return code and description** • **TxtCnvPlainTextHandlingForUnsupportedFormat** = 1 • IDXRC = 5 • The document, or one of the embedded attachments of the item, could not be handled by the Oracle/Stellent Outside In Technology filters and, although the configuration option **TxtCnvPlainTextHandlingForUnsupportedFormat** was enabled, the characters recognized in the document or attachment indicated a non-textual (binary) content.

Table 182. The indexer error strings (continued)

Notification string `IcmFceError: IcmEmptyFile` **Return code and description** • IDXRC = 6 • The document, or one of the embedded attachments of the item, could not be processed because the file is empty (size 0 bytes).
Notification string `IcmFceError: InsoErrorConvertingAttachment` **Return code and description** • IDXRC = 7 • The document, or one of the embedded attachments of the item, could not be handled by the Oracle/Stellent Outside In Technology filters due to a general error.

Recreating an index:

In most cases, you only need to ensure that the item type is text-searchable, in other words that it has been enabled for indexing, and do not have to concern yourself with creating the text index. However, if the Net Search Extender index configuration is not correct, or if the index is corrupted and cannot be restored from a backup, the index must be recreated.

The index recreation process can only be started if an indexer configuration file exists for the item type. If the index recreation process is called without a configuration file, the process looks for a default configuration for the item type in the current directory and if found, will load it. If no default configuration file is found or if no configuration file is passed in an argument to the index recreation process, the process will display an error message and processing will stop.

Recreating the index itself does not take long. However, filling the index is a very expensive procedure especially if a large amount of documents needs to be reindexed. Do not recreate an index unless there is absolute proof that the index is definitely corrupted or that recreating the index is clearly required.

Important: Before the index recreation process is started, the IBM Content Collector web application services must be stopped. Archiving to the item type must be stopped.

As a default, index recreation is run in interactive mode. This means that the user is prompted, for example, for confirmation about whether index recreation is really desired and whether archiving to the item type has been stopped. To run the index recreation process in batch mode without user prompting, set the value of the **AFU_INDEXER_PROMPT_RECREATE** environment variable to 1.

Recreating the index (formerly the RecreateIndex function) comprises the following:

- Deleting the Net Search Extender full text index
- Creating a new empty Net Search Extender full text index

- Creating a new empty database table of completed tasks
- Creating a new empty database table of open tasks
- Resetting the monitoring data for the item type in the table for monitoring tasks
- Deleting old triggers and creating new ones in the table of open tasks
- Filling the table of open tasks with all the item IDs in the item type

Attention:

- If you want to change attribute field definitions for the existing item type, make these changes to the item type model file and the item type configuration file before you recreate the index.
- The `afuRecreateIndex` process only creates the index; it does not fill the index. To fill the index, you must run `afuIndexer` which, depending on the size of the original index, can be a very time-consuming procedure.

To recreate the index:

- On UNIX and Linux, run

  ```
  afuRecreateIndex.sh
  ```

- On Windows, run

  ```
  afuRecreateIndex.cmd
  ```

Note: When you run the `afuRecreateIndex` process:

- All command line arguments are case sensitive.
- The command line arguments can appear in any order.
- Use the following delimiters for arguments that can pass more than one value:
  - On Windows, a semicolon (;)
  - On UNIX or Linux machines, a colon (:)
- The item type name is not case sensitive.
- The indexer configuration options passed through an argument are not case sensitive.

If you run `afuRecreateIndex` and do not specify an argument or enter the command followed by `-?`, the following is displayed:

```
afuRecreateIndex: Usage:
  [-?|-h|-copyright]
  -it <itemtype>
  -database <database>
  [-user <userID|@environmentVariable|:filename>]
  [-password <password|@environmentVariable|:filename>]
  [-schema <schema>]
  [-indexdirectory <directory>]
  [-workdirectory <directory>]
  [-modelfile <filename>]
  [-tablespace <tablespace>]
  [-useconfig <.ini file>;<...>]
  [-usesettings "<option=value>;<...>"]
  [-exportconfig <.ini file>]
  [-listconfig]
  [-verbose] [-showsql] [-trace]
  [-simulate]
  [-quiet]
```

where:

-? Lists all of the arguments that you can pass to the `afuRecreateIndex` tool.

-h Lists all of the arguments that you can pass to the `afuRecreateIndex` tool.

-copyright
Lists the copyright notice and version information.

-it <*itemtType*>
Required.

The IBM Content Manager item type (document collection) that you want to index. Item type names of both the bundled and compound data models are not case sensitive.

Important: If your item type is an email compound item type, do not enter the attachment item type. Attachment item types cannot be processed on their own but are automatically processed as part of the email item type.

-database <*database*>
Required.

The name of the library server database in which the item type is stored. The name of the database can be specified in two formats which affects the way the indexer tool connects to the database. The database connection is implemented using Java Data Base Connectivity (JDBC) which defines how a Java client may access a database. JDBC features different connection types and IBM Content Collector supports JDBC type 2 and JDBC type 4.

To use a JDBC type 4 connection, the specification for the database name of the argument `-database` must include a server name, a port number, and a database name in the following format `<databaseserver>:<portnumber>/<database name>`, for example `-database localhost:50000/icmnlsdb`. JDBC type 4 is the recommended database access method but always requires fully specifying the user ID and password.

To use a JDBC type 2 connection, the specification for the database name of the argument `-database` must name an existing DB2 connection name, in other words, the database name must be cataloged in DB2, for example `-database icmnlsdb`. As JDBC type 2 uses an existing database connection, no user ID and password need to be specified. However, a native 32-bit version of the DB2 JDBC Type 2 driver (shared library) must be available and defined in the class path for using the JDBC type 2 connection.

-user <*userID*|*@environmentVariable*|*:filename*>
Optional

Provide a user ID. Only required if no connection exists to the library server database. There are three ways in which you can enter a user ID:

1. Specify the user ID directly on the command line
2. Specify the name of an environment variable from which the user ID is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_USERID=user8

   afuRecreateIndex -user @INDEXER_USERID
   ```

3. Specify the name of a text file that contains the password, for example,

   ```
   afuRecreateIndex -user :myUserID.txt
   ```

   The text file must contain a single line with the user ID (no quotes and no spaces).

The advantage of using an environment variable or a file is that the user ID is not visible in any process status information.

The user ID must have access rights to delete and update text search indexes. For exact information on what user rights are required to delete and update text search indexes, see http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html for DB2 Net Search Extender V9.5 and http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html for DB2 Net Search Extender V9.7.

-password <*userID* | *@environmentVariable* | *:filename*>
Optional

The password of the user ID with database administrator authority. Only required if no connection exists to the library server database. There are three ways in which you can enter a password:

1. Specify the password directly on the command line
2. Specify the name of an environment variable from which the password is read. The name of the environment variable must be prefixed by an At (@) character, for example,

   ```
   export INDEXER_PASSWORD=12345678

   afuRecreateIndex -password @INDEXER_PASSWORD
   ```

3. Specify the name of a text file that contains the password, for example,

   ```
   afuRecreateIndex -password :myPassword.txt
   ```

   The text file must contain a single line with password (no quotes and no spaces).

The advantage of using an environment variable or a file is that the password is not visible in any process status information.

Passwords are not logged in a readable plain text format.

-schema <*schema*>
Optional.

The name of the database schema if the schema is different from the current database schema or the configuration option **GenJbdcType** is set to 4.

-indexdirectory<*directory*>
Optional.

The absolute path of the directory on the Content Manager library server where Net Search Extender is to create the full text index for the item type.

Enables recreating the index in a different directory. If this argument is not passed, the index is recreated in the same directory in which the old index was located.

-workdirectory<*directory*>
Optional.

The absolute path of the working directory on the Content Manager library server that is required by Net Search Extender for temporary index files

Enables specifying a different directory that the index process is to use as a work directory. If this argument is not passed, the work directory is the same as the index directory.

Note: It is recommended that the index directory and the work directory are on the same volume.

-modelfile<*filename*>
Optional.

Enables recreating the index for the item type using a model file that you have modified.

If a model file is not specified, the model file that was used for the existing Net Search Extender index will be used to recreate the index.

-tablespace<*tablespace*>
Optional.

Enables specifying which table space to use during index creation. If this argument is not passed, the default DB2 table space is used.

-useconfig <*.ini file*>;<*.ini file*>
Optional.

Enables using alternate configuration .ini files or configuration files from a different path. Multiple .ini files can be specified and they will be loaded in the order in which they have been specified.

-usesettings *"<option=value>;<option=value>"*
Optional.

Enables the direct temporary overwriting of indexer configuration values stored in the .ini file. You can add any number of configuration option value pairs <*option=value*> separated by the delimiter for your operating system. The option value pairs must be enclosed in double quotes. The configuration options you list (<*option*>) are not case sensitive. Using this argument is helpful at the beginning to verify the configuration and execution of the indexer.

-exportconfig <*.ini file*>
Optional.

Exports the indexer configuration to the .ini file specified in <*.ini file*>.

-listconfig
Optional.

Lists all the configuration options and their values that are set for the current indexer run.

-verbose
Optional.

Displays detailed processing information on the screen. Using **-verbose** is helpful at the beginning to get detailed information on all processing. If **-verbose** is specified the log files also contains additional information.

-showsql
Optional.

Displays all of the DB2 SQL commands executed by the system on the screen. Using this argument is helpful at the beginning to verify the configuration and execution of all DB2 system processing. If this argument is specified, all of the SQL statements are also added to the log file.

-trace
Optional.

Goes beyond verbose by displaying more detail in log files to support error determination. This is only required for IBM Support and requires the use of the trace property file.

-simulate
Optional.

Runs the create index process but does not update the databases. Net Search Extender is not called. Use this setting to validate the index configuration.

-quiet
No processing steps are displayed on the console.

Backing up index files:

Index files can be corrupted during the final merging or reorganization stages in the indexing process if either the `db2text` or `cteprupx` processes are stopped unexpectedly.

The index can also be corrupted if files cannot be written to because there is not enough file system space. It is therefore important to design a strategy to restore a point-in-time backup and then to update this backup with documents that were indexed prior to the backup.

If you do not back up the index, you will have to recreate the index which is an expensive and time consuming task.

Backup processing can be included during indexer processing using the GenCommandXXX configuration options. This can be useful if the defined backup processing steps are intended to run on Net Search Extender data only, and if backing up the index data is not part of a complete disk backup solution that is run periodically.

Using these option, user-defined processing steps can be added directly before indexing or index reorganization begins, or immediately after indexing or reorganization has completed.

Database table layout for the indexer task tables:

There are two database indexer task tables, one for the open tasks and another for the completed tasks.

Both tables are created for each item type. The content of these database tables can be used for reference purposes in addition to the available indexer process command line arguments and configuration options.

The table of open tasks contains all items that have not been processed and indexed. The name of the database table of open tasks is formed by the combination of `AFU1FTIOPEN<item type id>` where `<item type id>` is the numeric item type ID. Enter preceding zeros if necessary to make the ID 5 digits long.

It has the following layout:

Column name	Type schema	Type name	Length	Scale	Nulls
ITEMID	SYSIBM	CHARACTER	26	0	No
VERSIONID	SYSIBM	SMALLINT	2	0	No
ADDDATE	SYSIBM	TIMESTAMP	10	0	Yes
OP	SYSIBM	SMALLINT	2	0	Yes
IDXSTATUS	SYSIBM	SMALLINT	2	0	Yes

The name of the database table of completed tasks is formed by the combination of AFU1FTICOMP<item type id> where <item type id> again is the numeric item type ID and must be 5 digits long. The database table of completed tasks contains all items that were indexed or were skipped for what ever reason. It has the following layout:

```
Column                      Type      Type
name                        schema    name              Length   Scale Nulls
--------------------------- --------- ----------------- -------- ----- ------
ITEMID                      SYSIBM    CHARACTER               26     0 No
VERSIONID                   SYSIBM    SMALLINT                 2     0 No
ADDDATE                     SYSIBM    TIMESTAMP               10     0 Yes
IDXDATE                     SYSIBM    TIMESTAMP               10     0 Yes
IDXRC                       SYSIBM    SMALLINT                 2     0 Yes
```

See “Index return codes stored in the table of completed tasks” on page 604 for a list of the available IDXRC values.

Deleting an item type:

Before you delete an existing item type, run the indexer check tool to determine the database table names which you need to delete database tables and the full text index.

To delete an item type:

1. Run the afuCheck tool with the argument -verbose. This mode will display the following:
   - The name of the database table for the open tasks (under **Table of open tasks**)
   - The name of the database table for complete tasks (under **Table of complete tasks**)
   - The name of the Net Search Extender full text index (under**NSE Properties → NSE index name**)
2. Delete the table of open tasks for the item type using the DB2 DROP TABLE command and the name of the table of open tasks.
3. Delete the table of completed tasks for the item type using the DB2 DROP TABLE command and the name of the table of completed tasks.
4. Delete the Net Search Extender full text index as it is described in the Net Search Extender documentation by running db2text drop index command.

Indexer for text search error messages:

This is a list of the indexer error messages that may occur with a description of what action to take.

1 **An instance of the indexer is already running for this item type. The file "{0}" indicates this. A second instance cannot be started.**

User action: This can occur if the indexer process stopped unexpectedly or was interrupted.

Check that the indexer process is no longer running. After the process has been stopped, delete the specified file, and restart the indexer

2 **A communication problem occurred: The indexer process cannot communicate with the worker process.**

User action: Choose a new port number by setting the environment variable AFU_INDEXER_RMI_PORT to a different value or using the -port indexer process argument. Make sure that the port number is free and is not used by another running indexer process instance or another application.

3 **The current Net Search Extender configuration file does not contain the recommended configuration values.**

User action: It is recommended that you configure Net Search Extender using the recommended settings. For details on which configuration settings require which values, see "Incorrect Net Search Extender configuration" on page 819. To use new settings, the Net Search Extender index must be recreated using the `afuRecreate` tool.

4 **The item type cannot be enabled because it still contains items that have not been indexed by Net Search Extender. The Net Search Extender log table must be empty before the item type can be enabled for use with the indexer.**

User action: Disable archiving on the item type and then index the items that have not been processed using the fast indexer or the standard Content Manager indexer with the IBM Text Search. Then the item type can be enabled for processing by the indexer for text search.

5 **The Net Search Extender text indexing service cannot be queried.**

User action: Check the Net Search Extender installation. The Net Search Extender commands must be available and defined in PATH, and they must have sufficient permissions to run.

Check that you are logged in as the instance owner and have sufficient permissions to run the `db2text` command in the current session.

6 **The Net Search Extender UDF registered for this item type cannot be identified. Verify the installation of the Net Search Extender UDF.**

User action: Check that the AFUFETCHFILE UDF was copied to the DB2 function subdirectory.

7 **The section [{0}] in the Net Search Extender configuration file "{1}" could not be determined.**

User action: Check that the Net Search Extender index configuration file is valid as described in the Net Search Extender documentation.

8 **The value of the Net Search Extender index configuration option {0} is not the recommended value for correct indexing (current value: {1}). The Net Search Extender index must be recreated using an index configuration in which the value of the configuration option {0} is set to {2}**

User action: You must change the value of the index configuration option in the Net Search Extender index configuration template file to the recommended value. The Net Search Extender index configuration template file is called `cteixcfg.ini` and is located in the Net Search Extender installation directory (usually `.../sqllib/db2ext`).

For details on which configuration settings require which values, see "Incorrect Net Search Extender configuration" on page 819.

Then the Net Search Extender index must be recreated using the afuRecreateIndex tool.

The Net Search Extender index configuration template file is automatically copied from the Net Search Extender installation directory to the new text index directory during the text index recreation process.

9 **The Net Search Extender index configuration option {0} is not defined. The Net Search Extender index must be recreated using an index configuration in which the configuration option {0} is defined and set to {1}.**

User action: You must change the value of the index configuration option in the Net Search Extender index configuration template file to the recommended value. The Net Search Extender index configuration template file is called `cteixcfg.ini` and is located in the Net Search Extender installation directory (usually `.../sqllib/db2ext`). Then the Net Search Extender index must be recreated using the afuRecreateIndex tool.

The Net Search Extender index configuration template file is automatically copied from the Net Search Extender installation directory to the new text index directory during the text index recreation process.

10 **The Net Search Extender configuration file "{0}" cannot be loaded for verification.**

User action: Check the file permissions and that the file is available.

11 **The Net Search Extender text indexing service cannot be started.**

User action: Check the Net Search Extender installation. The Net Search Extender commands must be available and defined in PATH.

Check that you are logged in as the instance owner and have sufficient permissions to run the `db2text` command in the current session.

12 **The status of Net Search Extender cannot be determined because of unexpected output: "{0}".**

User action: Check the Net Search Extender installation. The Net Search Extender commands must be available and defined in PATH.

Check that you are logged in as the instance owner and have sufficient permissions to run the `db2text` command in the current session.

13 **The Net Search Extender process did not output any status messages ("{0}").**

User action: Check the Net Search Extender installation. The Net Search Extender commands must be available and defined in PATH.

14 **The Net Search Extender search engine could not be started: Reason: {0}**

User action: Check the Net Search Extender installation. The Net Search Extender commands must be available and defined in PATH.

Check that you are logged in as the instance owner and have sufficient permissions to run the `db2text` command in the current session.

15 **The Net Search Extender model file "{0}" could not be loaded.**

User action: Check the file permissions and that the file is available and valid as described in the Net Search Extender documentation. If this does not solve the problem, recreate the index.

16 **The Net Search Extender model file "{0}" could not be parsed.**

User action: Check the file permissions and that the file is available and valid as described in the Net Search Extender documentation.

17 **The DB2 installation directory could not be located: The environment variable "{0}" is not defined.**

User action: Set the specified environment variable to the correct value and restart the command.

18 **The Net Search Extender configuration file ("{0}") could not be located in the DB2 installation directory ("{1}") that was set in the environment variable "{2}".**

User action: Check that the specified environment variable points to the directory in which DB2 was installed. The Net Search Extender configuration file ("{0}") is searched in all subdirectories of the specified directory.

19 **No Net Search Extender text search index was found for the specified item type ("{0}"). This item type cannot be processed.**

User action: Check that the correct item type was specified when the indexer process was called and that it has been enabled for processing by the indexer for text search.

20 **A Net Search Extender index reorganization is overdue! The number of items that were updated since the last Net Search Extender index reorganization ({0}) exceeds the number specified in the configuration option "{1}". No more index updates are possible until an index reorganization has been performed.**

User action: The indexer process cannot be run again without the command line argument `-nseindexreorg` and a batch size of 1.

21 **An item was found that is missing a value for a defined Content Manager attribute.**

User action: If any of the following attributes are defined for an item type, the indexer process expects a value for the attribute in each item it processes:

- ICCMailboxID
- CSORGINATOR (when accessing a CommonStore bundled resource item)
- CSLDOrigDB (when accessing a CommonStore bundled resource item)

Check your task route configuration. If these values are missing, the task route was not able to extract the values for these attributes.

22 **The `afuIndexer` process cannot be started.**

Explanation: The following error message is displayed when the `afuIndexer` indexer process is started:

```
afuIndexer: An error occurred:
A configuration or setup problem occurred:
The version of the Net Search Extender UDF
registered for this item type is not
supported.
Version '[V2]' is required.
Verify that the installation of the Net
Search Extender UDF AFUFetchFile is correct
```

This indicates that an older version of the `AFUFetchFile` user defined function (UDF) is still installed on your system. This can happen if problems are encountered when Content Collector Text Search Support is installed.

User action: You must update the `AFUFetchFIle` UDF manually:

1. Stop DB2.
2. Copy the file `AfuFetchFile.dll` from the subdirectory `lib` in the indexer installation directory (default: `/opt/IBM/ICCIndexer/lib`) to the `function` subdirectory of the DB2 instance directory (default: `$DB2HOME/function`).
3. Restart DB2.
4. Restart the `afuIndexer` process.

Enabling an IBM FileNet P8 object store for search

Before users can search on email that was archived in a FileNet P8 object store by using IBM Content Collector, you must set up the FileNet P8 Content Search Engine accordingly.

The general procedure to enable an object store for search with IBM Content Collector is as follows. The respective steps vary depending on the version of FileNet P8 that you use.

1. Enable the object store for full-text indexing
2. Verify your Content Search Engine configuration
3. Optionally: Clean the object store

Related tasks

"Configuring an IBM FileNet P8 repository" on page 580

Related reference

"Additional prerequisites and restrictions" on page 31

Enabling object stores in FileNet P8 versions before Version 4.5.1 for full-text indexing

Configure the FileNet P8 Content Search Engine to enable content-based retrieval of email that was archived using IBM Content Collector in an object store in FileNet P8 versions before Version 4.5.1.

Prerequisites:

- An installation of IBM FileNet P8 Content Engine and Content Search Engine
- An empty FileNet P8 object store using a File Storage Area as the default Content Store. See the steps in "Cleaning object stores" on page 679 if the object store already has an index area. For information about defining an object store refer to the topic about configuring an IBM FileNet P8 repository.

To enable an object store for Content Collector:

1. Change to the `ctms` folder of the Content Collector installation directory.
2. Extract the contents of the file `ICC_FileSystem_PushAPI2.zip` to the FileNet P8 Content Search Engine `stylesets` directory. For example:

   AIX `/opt/verity/data/stylesets`

   The resulting directory structure looks like this:

   `/opt/verity/data/stylesets/ICC_FileSystem_PushAPI2`

   Windows
   `<FileNetInstallDir>\FileNet\ContentEngine\Verity\data\stylesets\`

The resulting directory structure looks like this:

`<FileNetInstallDir>\FileNet\ContentEngine\Verity\data\stylesets\ICC_FileSystem_PushAPI2`

Where <*FileNetInstallDir*> is the installation path for FileNet P8.

3. In the Verity K2 Dashboard, select **Collections** → **Manage Style Sets** → **Import** to import this style set.
   a. In the **Style Set Alias** field, enter: `ICC_FileSystem_PushAPI2`
   b. In the **Gateway Type** field, enter: `File System Gateway`
   c. In the **Source Path** field, enter the path to the `stylesets` directory.
   d. Click **Import**.
4. If you want to include custom attributes in the index, adapt the `style.xml` file accordingly. The `style.xml` file contains a definition for a zone with the name `icc_custom_metadata`. By default, all user-defined metadata is stored in that zone. To have each custom attribute indexed separately in a zone of its own and to enable searches in this specific zone, add a `<preserve xmltag="xxx" >` element for each attribute.

   Important: If you modify any of the style files that are used to create an index, you must remove the index and then recreate the index for an existing object store. Re-indexing will not incorporate the style file modifications in an existing index.
5. In IBM FileNet Enterprise Manager, create an index area for the IBM Content Collector object store using the Content Collector style set.

   Important: Only an index area created using the template type ICC_FileSystem_PushAPI2 must exist for the object store that you want to use for email archiving with Content Collector. If the object store index area has been configured using another template, you must clean up the old index area collections. See the section on cleaning object stores.
6. Now your object store is ready for creating Content Collector document classes. You can proceed with installing the IBM Content Collector Server on a dedicated machine. After you have installed IBM Content Collector Server, start the initial configuration wizard to create document classes in the object store.

Verifying your Content Search Engine configuration for IBM FileNet P8 versions before Version 4.5.1

After you have installed IBM Content Collector Server and have run the initial configuration wizard, verify your Content Search Engine configuration using FileNet Enterprise Manager (FEM).

1. Create a new document of class ICCMailSearch2 that contains a text file. The text file should be have a .xml file extension (for example `searchtest.xml`) and should contain the following XML statement:

   ```
   <icc_content>test</icc_content>
   ```

   If the ICCMailSearch2 document class does not exist in your object store, you must rerun the IBM Content Collector initial configuration wizard to enable the object store for email archiving with IBM Content Collector.
2. Wait for the content-based retrieval (CBR) dispatcher interval to elapse (the default is 180 seconds). Then, check if an index area collection was created for your object store index area. If no index area was created, check the Content Engine log files for errors.

3. Use the Verity K2 Dashboard to view the details of your newly created collection. There should be one document index. Wait a moment for the index to appear.
4. Open the Text Search function on your Search Server in the Verity K2 Dashboard and issue the following query by using the Verity Query Language query parser:

   ```
   test <in> icc_content
   ```

   FileNet P8 Content Search Engine should find the test document.

Enabling object stores in FileNet P8 Version 4.5.1 for full-text indexing

Configure the FileNet P8 Content Search Engine to enable content-based retrieval of email that was archived using IBM Content Collector in an object store in FileNet P8 Version 4.5.1.

Prerequisites:

- An installation of IBM FileNet P8 Content Engine and Content Search Engine
- An empty FileNet P8 object store using a File Storage Area as the default Content Store. See the steps in "Cleaning object stores" on page 679 if the object store already has an index area. For information about defining an object store refer to the topic about configuring an IBM FileNet P8 repository.

To enable an object store for Content Collector:

1. Change to the `ctms` folder of the Content Collector installation directory.
2. Extract the contents of the file `ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1.zip` to the FileNet P8 Content Search Engine `stylesets` directory. For example:

   AIX `/opt/verity/data/stylesets`

   The resulting directory structure looks like this:

   `/opt/verity/data/stylesets/ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1`

   Windows
   `<FileNetInstallDir>\FileNet\ContentEngine\Verity\data\stylesets\`

   The resulting directory structure looks like this:

   `<FileNetInstallDir>\FileNet\ContentEngine\Verity\data\stylesets\ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1`

   Where *<FileNetInstallDir>* is the installation path for FileNet P8.
3. In the Autonomy K2 Dashboard, select **Collections** → **Manage Style Sets** → **Import** to import this style set.
   a. In the **Style Set Alias** field, enter: `ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1`
   b. In the **Gateway Type** field, enter: `File System Gateway`.
   c. In the **Source Path** field, enter the path to the `stylesets` directory.
   d. Click **Import**.
4. For a new object store, you must set a CBR locale value, such as `uni` for Unicode, before you can create an index area. You do that on the CBR tab on the object store property sheet. For a complete list of valid CBR locale values, see the Autonomy K2 documentation.
5. If you want to include custom attributes in the index, adapt the `style.xml` file accordingly. The `style.xml` file contains a definition for a zone with the name `icc_custom_metadata`. By default, all user-defined metadata is stored in that

zone. To have each custom attribute indexed separately in a zone of its own and to enable searches in this specific zone, add a <preserve xmltag="*xxx*" > element for each attribute.

Important: If you modify any of the style files that are used to create an index, you must remove the index and then recreate the index for an existing object store. Re-indexing will not incorporate the style file modifications in an existing index.

6. In IBM FileNet Enterprise Manager, create an index area for the Content Collector object store using the Content Collector style set.

   Important: Only an index area created using the template type ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1 must exist for the object store that you want to use for email archiving with IBM Content Collector. If the object store index area has been configured using another template, you must clean up the old index area collections as described in the topic on cleaning object stores.
7. Install the IBM ICC Compliance Data Model for the object store. You can import it either by using the Content Collector setup tools for repository enablement or by using the AddOn Installer of the IBM FileNet Enterprise Manager.
8. Configure the partition settings of the object store. You do that on the CBR tab on the object store property sheet. Select Change Collection Partitioning Settings and define an interval for the ICCMailDate date property, for example, one month.
9. Set the settability for ICCMailDate to **Settable only on create**. You do that on the More tab on the ICCMailDate properties sheet.

Verifying your Content Search Engine configuration for IBM FileNet P8 Version 4.5.1

After you have installed IBM Content Collector Server and have run the initial configuration wizard, verify your Content Search Engine configuration using FileNet Enterprise Manager (FEM).

1. Create two new documents of class ICCMailSearch2 that contain text files. The text files should be have a .xml file extension and should contain the following XML statement: <icc_content>*testtext*</icc_content>. For example, searchtest1.xml with <icc_content>test partition one</icc_content> and searchtest2.xml with <icc_content>test partition two </icc_content> . If the ICCMailSearch2 document class does not exist in your object store, you must rerun the initial configuration wizard to enable the object store for email archiving using IBM Content Collector.
2. Import the files as new P8 documents of type ICCMailSearch2. Be sure to attach each file as a content element and set the ICCMailDate property for each file to a date/time value. For example, set the ICCMailDate property of the first document to "4/20/2008 9:53:00 AM" omitting the quotes. Set the ICCMailDate property of the second document to "1/10/2009 8:01:00 AM" omitting the quotes. Make sure that the two dates fall in two different collection partitions based on the collection partition configuration of the object store.
3. Wait for the content-based retrieval (CBR) dispatcher interval to elapse (the default is 180 seconds). Then, check if index area collections were created for your object store index area. If no index area collection was created, check the Content Engine log files for errors.

4. Use the Autonomy K2 Dashboard to view the details of your newly created collection. There should be one document index. Wait a moment for the index to appear.
5. Open the Text Search function on your Search Server in the Autonomy K2 Dashboard and issue the following query by using the Verity Query Language query parser:

   ```
   testtext <in> icc_content
   ```

   where *testtext* is the text you specified in one of your .xml text files. FileNet P8 Content Search Engine should find the test documents.

Cleaning object stores

If an index area already exists and if collections are associated with the index area, you might need to clean up the object store.

1. Start FileNet Enterprise Manager (FEM).
2. Deactivate content-based retrieval (CBR) for all document classes. Typically, CBR is only enabled for ICCMailSearch2.
   a. From the Enterprise Manager tree view, expand the object store node and right-click the main document class.
   b. Select **Properties** from the menu.
   c. On the General tab on the object store property sheet, clear the **CBR Enabled** check box.
3. Reindex the index area by committing the job automatically created in the Index Jobs Manager.
   a. Right-click the object store.
   b. From the context-menu, select **All Tasks → Index Jobs Management**.
   c. Click **View Class Definition CBR Enabled Changes** to open the List of Class Definitions Changed window.
   d. Select the class definitions you want to reindex.
   e. Click **Create Index Job**.

   The existing index area must not show associated collections anymore. If collections still exist, go back to step 2 and check your FileNet Content Engine Server logs.
4. Modify the index area configuration to make sure the Content Collector style file is used instead of FileNet_FileSystem_PushAPI (the standard for IBM FileNet Content Engine CBR setups). The proper style file depends on the version of FileNet P8 that you use:

   ICC_FileSystem_PushAPI2
   FileNet P8 versions before Version 4.5.1

   ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1
   FileNet P8 Version 4.5.1
5. Verify your IBM FileNet Content Search Engine configuration.

Related tasks

"Verifying your Content Search Engine configuration for IBM FileNet P8 versions before Version 4.5.1" on page 676

"Verifying your Content Search Engine configuration for IBM FileNet P8 Version 4.5.1" on page 678

About collections

Each archived document belongs to an IBM Content Manager *item type* or an IBM FileNet P8 *document class*. An item type or a document class is essentially a template for defining and later locating similar items. Each item type or document class is used to store documents according to specific criteria. For search, item types or document classes are grouped in *collections*. Use an item type or document class only once in a given collection and, if possible, do not use it in more than one collection. If an item type or document class is used in more than one collection, all field definitions of the collections containing it must be the same.

Before users can search collections, you must define the collections in the archive mapping and search configuration files. Your company might have only one collection, or it might have multiple collections. There are many reasons for having multiple collections, for example:

- You need multiple collections if you have content with different properties. For example, files typically have creation dates, file names, and authors. Email typically has recipients, subjects, bodies, and attachments.
- If your company uses Lotus Domino and acquires another company that uses Microsoft Exchange, your company ends up with two collections, one of Lotus Domino email and another of Microsoft Exchange email.
- Your company might find it useful to separate email into different collections based on user role, geography, or some other factor (as defined in the archiving task route).
- If your company changes the way in which it archives email (for example, it adds fields), this creates old and new collections that must be searched independently.
- If your company needs to make repositories previously fed by IBM CommonStore for Exchange Server or IBM CommonStore for Lotus Domino available for search and restoring, a separate collection is required.

If you define multiple collections, you can group them into *collection sets*.

Users can perform a single search across multiple item types or document classes that are grouped in the same collection. The item types or document classes can be located in more than one repository. If the search application is configured accordingly, users can also perform searches across collections by selecting a specific collection set.

When you configure email search, consider the impact that the search scope has on performance. The more item types or document classes are in a given collection and the more collections are in a given collection set, the longer a query might run.

Important: One installation of the email search application of Content Collector can run queries against only one type of repository. Therefore, all collections that are defined in one archive mapping or search configuration file must reside in the same type of repository. If you change the type of repository (for example, from an IBM Content Manager repository to a FileNet P8 repository), you have to adapt the search configuration files accordingly.

If your configuration includes both Lotus Notes and Microsoft Exchange collections, set the system environment variable AFU_DISABLE_URL_CHECK to ensure that all Content Collector clients can display the email search page. You can use any value for the variable. If the system environment variable is not set and your configuration files contain only Domino collections, all Content Collector clients can start the email search function. If your configuration files contain Exchange collections, only Exchange clients that have IBM Content Collector, Version 2.1.1, installed can start the email search function.

Defining and exposing collections

Users can search the collections that are defined in the archive mapping and search configuration files. These files are part of the archived-data access configuration in the Content Collector configuration database. The archive mapping file defines the collections that exist in the IBM Content Collector data store. The search configuration file defines what collection sets are displayed on the Email Search page of the IBM Content Collector web client. These configuration files are also required for viewing or restoring documents.

The archive mapping file and the search configuration file are stored in the Content Collector configuration database. Both files are created during the initial configuration of IBM Content Collector and initially contain default definitions for one collection set with the respective collections and item types or document classes.

If the default collection set is sufficient for your company, no configuration is required. However, you can define further collections and collection sets. You can also add item types or document classes to a collection as long as the new definition includes the same attributes and references as one of the existing ones.

If your repository is IBM Content Manager and you want to include legacy item types into the search scope, you must define a new collection by using the templates provided with the product.

You can adapt the archive mapping file either by using the graphical user interface of the Configuration Manager or by editing the file. The search configuration file must be adapted manually in any case.

Before you can modify the archive mapping or search configuration files manually, you have to export the respective file to a directory of your choice by using the Configuration Manager. Edit the file with a text editor. After saving your changes, import the file to the IBM Content Collector data store using the Configuration Manager and save the changed configuration. Then, restart the Content Collector Web Application Server for the changes to take effect.

Important:

- Names that you specify in the configuration files are case sensitive. The names defined in the archive mapping file must match the names as defined in the search configuration file.
- Names and IDs must be identical.
- Template names and template IDs must consist of alphabetic characters. Do not use any special characters, digits, or blanks in template names or template IDs.
- String values for elements in the configuration files must not contain leading or trailing blanks.
- The following keywords in the default archive mapping templates are reserved and must not be changed:

Table 183. Reserved keywords in archive mapping files

Keyword	Repository	Document type
Application Archiving	• IBM Content Manager • IBM FileNet P8	• All document types
ATTACH_REF	• IBM Content Manager	• Email documents • Notes application documents
ATTACHMENT_NAME	• IBM Content Manager	• Email documents • Notes application documents
CONTENT	• IBM Content Manager • IBM FileNet P8	• All document types
CONTENT_REF	• IBM Content Manager	• Email documents • Notes application documents
CORRELATION_KEY	• IBM Content Manager • IBM FileNet P8	• Email documents • Notes application documents
EMAIL_REF	• IBM Content Manager	• Email documents
FILENAME	• IBM Content Manager	• File System documents • Microsoft SharePoint documents
File System	• IBM Content Manager • IBM FileNet P8	• All document types
ICC Document	• IBM Content Manager • IBM FileNet P8	• All document types
MAIL_REF	• IBM FileNet P8	• Email documents
MAILBOX_ID	• IBM Content Manager • IBM FileNet P8	• All document types
MIMETYPE	• IBM FileNet P8	• File System documents • Microsoft SharePoint documents
Sharepoint	• IBM Content Manager • IBM FileNet P8	• All document types

Configuration files for enabling access to archived data:

Depending on the type of source document and on the type of your repository, IBM Content Collector requires different definitions in the configuration files for enabling the access to archived data from a Content Collector web client. The configuration files are the archive mapping and search configuration files.

Entries in the archive mapping file define and map these elements, depending on the repository that you use:

IBM Content Manager repository

- Define logical names for attributes
- Map search fields and document types to attributes, to fields in the full-text index, and to item types that are available in IBM Content Manager

FileNet P8 repository

- Define logical names for attributes
- Map search fields and document types to properties, to zones in the full-text index, and to document classes that are available in FileNet P8

Note: For File System and Microsoft SharePoint, you do not need to perform these tasks. However, you must map any custom (user-defined) metadata properties to your document classes. See "Adding and editing metadata" on page 237. You might also need to identify additional collections or document classes. See the topics about metadata and about collections."About collections" on page 680.

Documents from File System, Microsoft SharePoint, or Notes application sources are archived into a given item type or document class. Therefore, for each of these sources only one collection is defined that contains the field for the archived properties. This collection does not refer to any other collection. You must define an archive mapping file for these systems to be able to retrieve such documents from the archive. If your system is set up for File System or SharePoint only, you must also provide a basic search configuration.

When you configure the archive mapping file, you also have to consider the email data model that is used for archiving:

Bundled email data model
An archived email document consists of a set of attributes that are common to all instances of the email document, such as the email document itself or its subject, and multiple sets of attributes that are unique to each instance of the email document, such as the mailbox from which the email document was archived. Attachments are part of the email document, they are not stored separately.

Compound email data model
An archived email document consists of a set of attributes that are common to all instances of the email document, such as the email document itself or its subject, and multiple sets of attributes that are unique to each instance of the email document, such as the mailbox from which the email document was archived. Attachments are stored separately. When attachments are archived, Content Collector checks for duplicates and ensures that each attachment is stored only once in the archive (de-duplication).

The search configuration file consists of templates that define the layout of the Email search page. Each template comprises these major elements:

<collections> section
Defines the search scope. Queries that use this template search all collections that are listed here. For each collection, the properties that are defined in the declaration section are mapped to fields in archive mapping file. Note that you can use only those properties that are defined in the declaration section when you define the actual page layout.

<declaration> section
Defines which properties are used by this template

<form> section
Defines which properties are used in the input fields on the Email search page

<result> section
Defines which fields are shown on the result page and in which order those fields are shown

For all configuration files, templates are provided with the product. These template files are stored in the directory <*installDir*>\Configuration\initialConfig\data\search. The configuration files that are created during the initial configuration of IBM Content Collector are stored in subdirectories of the directory <*installDir*>\Configuration\initialConfig\data\search\output. The subdirectories are \cm for IBM Content Manager and \p8 for FileNet P8. The templates for use with legacy item types, documents archived from Notes applications, and for customized search are stored in the directory <*installDir*>\AFUWeb\afu\config\templates, where *installDir* is the directory where you installed IBM Content Collector.

Template configuration files	Description
afu_cm_search_config_template.xml	Template search configuration file for a IBM Content Manager repository.
afu_bundled_cm_search_mapping_template.xml	Template archive mapping file for a IBM Content Manager repository with the bundled email data model.
afu_compound_cm_search_mapping_template.xml	Template archive mapping file for a IBM Content Manager repository with the compound email data model.
afu_p8_search_config_template.xml	Template search configuration file for a FileNet P8 repository.
afu_compound_p8_search_mapping_template.xml	Template archive mapping file for a FileNet P8 repository.

Configuration files for use with legacy item types	Description
CS_CM_attribute_collection_search_config.xml	Template search configuration file for attribute search on legacy item types (stored by IBM CommonStore in IBM Content Manager).
CS_CM_full_text_collection_search_config.xml	Template search configuration file for full-text search on legacy item types (stored by IBM CommonStore in IBM Content Manager).
CS_CM_doc_type_collection_search_mapping.xml	Template archive mapping file for a collection containing legacy item types (stored by IBM CommonStore in IBM Content Manager).

Configuration files for customized search	Description
afu_customized_sample_search_config_template.xml	Template search configuration file for customized search.
afu_customized_sample_cm_search_mapping.xml	Template archive mapping file for customized search in IBM Content Manager.

Configuration files for customized search	Description
afu_customized_sample_p8_search_mapping.xml	Template archive mapping file for a customized search in FileNet P8.
afu_p8_bundled_search_mapping.xml	Template archive mapping file for the email data model for FileNet P8 that was used in versions before IBM Content Collector, Version 2.1.1.

Configuration files for viewing documents archived from Notes applications	Description
afu_application_archiving_search_config_template.xml	Template search configuration file for documents archived from Notes applications.
afu_cm_application_archiving_search_mapping_template.xml	Template archive mapping file for documents archived from Notes applications in IBM Content Manager.
afu_p8_application_archiving_search_mapping_template.xml	Template archive mapping file for documents archived from Notes applications in FileNet P8.

Configuration files for viewing documents other than email	Description
afu_cm_filesystem_search_config.xml	Template configuration file for file-system documents that were archived in a IBM Content Manager repository.
afu_cm_filesystem_search_mapping_template.xml	Template mapping file for file-system documents that were archived in an IBM Content Manager repository.
afu_cm_sharepoint_search_config.xml	Template configuration file for SharePoint documents that were archived in an IBM Content Manager repository.
afu_cm_sharepoint_search_mapping_template.xml	Template mapping file for SharePoint documents that were archived in an IBM Content Manager repository.
afu_p8_filesystem_search_config.xml	Template configuration file for file-system documents that were archived in a FileNet P8 repository.
afu_p8_filesystem_search_mapping_template.xml	Template mapping file for file-system documents that were archived in a FileNet P8 repository.

You can modify the archive mapping file by using the graphical user interface of the Configuration Manager. To change the configuration of the archived-data access manually, you must export the configuration files to a directory of your choice by using the Configuration Manager. Make sure to save a backup copy of the configuration files before changing or adding any entries. Edit the configuration files and change or add entries as required. Import the changed configuration files into the Content Collector configuration database by using the Configuration Manager. Save the configuration and restart the Web Application Server for the changes to take effect.

Related concepts

"Definition of the storage data model" on page 15

"About collections" on page 680

Related tasks

"Adding and editing metadata" on page 237

Configuration files for email search in IBM Content Manager repositories:

The archive mapping file maps search fields and document types to information in the repository. The search configuration file defines the layout of the Email Search page in the IBM Content Collector web client, such as the search scope and the contents of the result list.

IBM Content Collector provides templates for both configuration files, the archive mapping file and the search configuration file. Depending on the email data model that you selected during the initial configuration of IBM Content Collector, a default archive-mapping file for the bundled email data model or one for the compound email data model is installed. A default search configuration file is also installed. In the default files, all values that appear in the template files as variables that start and end with percent signs (%), such as *%ICC-EMAIL-PROVIDER%*, were set during the initial configuration.

You can reuse the definitions for the ICCEmailInstance collection to model further collections as long as they share the same attributes.

You can reuse the definitions for the ICCAttachmentInstance collection to model email collections that share the same item type for storing attachments.

The following field names in the default archive mapping templates must not be changed:

- ATTACH_REF
- ATTACHMENT_NAME
- CONTENT_REF
- CORRELATION_KEY
- EMAIL_REF
- MAILBOX_ID
- CONTENT

Archive mapping file for the bundled email data model

The bundled email data model requires one IBM Content Manager item type. This item type contains all attributes that are common to all instances of the email document and the content of the email document. This item type is a resource item type. It has a child component that contains the varying attributes of each instance of the email document. This information is used to access a user's individual copy of the email document.

The template for the corresponding archive mapping file contains default collection definitions for the email item type as required for the bundled data model:

Default Mail
: Is the collection definition for the email item type that contains the content of the email document

ICCEmailInstance
Is the collection definition for the child component that contains the varying attributes of each instance of the email document

The <doc_type> definition in the collection for the email item type contains a reference to the child component:

```
<children>
  <child ref_coll="ICCEmailInstance">
    %ICC_CONFIG_ITEMTYPECHILD_EI%
  </child>
</children>
```

Where *ICCEmailInstance* is the name of the collection for the child component. In addition, there are field definitions that address fields in the child component. For example, to access the Content Collector attribute holding the user's mailbox, define a reference field for the collection that defines the email instance child component:

```
<field nm="EMAIL_REF" type="REFERENCE" ref_coll="ICCEmailInstance" multivalue="true">
  <relationship type="CHILD"></relationship>
</field>
```

Then, use the reference field EMAIL_REF to address the MAILBOX_ID attribute:

```
<field nm="MAILBOX_ID" type="STRING">
  <search>mailboxid</search>
  <reference>EMAIL_REF.MAILBOX_ID</reference>
</field>
```

The collection ICCEmailInstance has the collection type dependent. You cannot select it for search.

Archive mapping file for the compound email data model

The compound email data model requires one IBM Content Manager item type for all attributes that are common to all instances of the email document and the content of the email document. This item type is a resource item type. It has two child components.

The first child component contains the varying attributes of each instance of the email document. This information is used to access a user's individual copy of the email document. The second child component contains attachment-specific attributes, such as the file name of an attachment and the correlation key that links back to the attachment position in the email document. In addition, this child component contains a reference to the attachment item type. The attachment item type is also a resource item type and holds the attachment content.

The template for the corresponding archive mapping file contains default collection definitions for the email item type as required for the compound email data model:

Default Mail
Is the collection definition for the email item type that contains the content of the email document

ICCEmailInstance
Is the collection definition for the child component that contains the varying attributes of each instance of the email document

ICCAttachmentInstance

Is the collection definition for the child component that contains the unique attributes of the attachments and a reference to the actual attachment item type

The definition of the <doc_type> element in the collection for the email item type contains references to both required child components:

```
<children>
  <child ref_coll="ICCEmailInstance">
    %ICC_CONFIG_ITEMTYPECHILD_EI%
  </child>
  <child ref_coll="ICCAttachmentInstance">
    %ICC_CONFIG_ITEMTYPECHILD_AI%
  </child>
</children>
```

Where the values in enclosed in percent signs (%) represent the names of the child components.

In addition, there are field definitions that address fields in the child component. For example, to access the Content Collector attribute holding the correlation key, define a reference field for the collection that defines the attachment child component:

```
<field nm="ATTACH_REF" type="REFERENCE" ref_coll="ICCAttachmentInstance" multivalue="true">
  <relationship type="CHILD"></relationship>
</field>
```

Then, use the reference field ATTACH_REF to address the correlation key attribute:

```
<field nm="CORRELATION_KEY" type="STRING">
  <reference>ATTACH_REF.CORRELATION_KEY</reference>
</field>
```

All collections other than Default Mail have the collection type dependent. You cannot refer to them in your search configuration.

Search configuration file

The template for the search configuration file contains definitions for the search scope, such as the collections and the attributes that are available for search.

Configuration files for email search in IBM FileNet P8 repositories:

The archive mapping file maps search fields and document types to information in the repository. The search configuration file defines the layout of the Email Search page in the IBM Content Collector web client, such as the search scope and the contents of the result list.

Default configuration files are installed during the initial configuration of IBM Content Collector. Additionally, templates are provided with the product.

The following field names in the default archive mapping templates must not be changed:

- CORRELATION_KEY
- MAIL_REF
- MAILBOX_ID
- CONTENT

Archive mapping file for the bundled email data model

The bundled email data model requires several FileNet P8 document classes. One document class contains the properties that are common to all instances of the email document. The content element of this class holds an XML representation of the email. This XML representation is added to the full-text index. The default name for this document class is ICCMailSearch. One document class contains the properties that are specific to an instance of the email document. The default name for this document class is ICCMailInstance. One document class contains the content of the email document. The default name for this document class is ICCMail. Objects of this class refer to objects of the classes ICCMailSearch and ICCMailInstance.

The template for the corresponding archive mapping file contains default collection definitions for the document classes as required for the bundled data model:

Default Mail
: Is the collection definition for the document class that contains common properties and the XML representation for the full-text index

ICCMail
: Is the collection definition for the document class that contains the content of the email document

ICCMailInstance
: Is the collection definition for the document class that contains the unique properties of an email instance

The collection Default Mail contains field definitions that address fields in other collections. If such fields are in the full-text index, define a search element for each of them in the collection Default Mail. For example, to access the email content or the Content Collector property holding the user's mailbox, define a reference field for the collection ICCMail that contains the email content and a reference to the collection that contains the unique properties of an email document:

```
<field nm="MAIL_REF" type="REFERENCE" ref_coll="ICCMail">
  <attr>ICCMailReference</attr>
</field>
```

Then, use the reference field MAIL_REF to address data in the referenced collection:

```
<field nm="MAILBOX_ID" type="STRING">
  <search>icc_mailbox_id</search>
  <reference>MAIL_REF.MAILBOX_ID</reference>
</field>

<field nm="CONTENT" type="STRING">
  <search>icc_content</search>
  <reference>MAIL_REF.CONTENT</reference>
</field>
```

To access the unique properties of email instances, you have to define a reference to the collection ICCMailInstance:

```
<field nm="MAIL_INST_REF" type="REFERENCE" ref_coll="ICCMailInstance" multivalue="true">
 <attr>ICCMailInstanceReference</attr>
</field>
```

Additionally, you must define a field that refers to the collection that contains the unique properties for the email instance:

```
<field nm="MAILBOX_ID" type="STRING">
  <reference>MAIL_INST_REF.MAILBOX_ID</reference>
</field>
```

All collections other than Default Mail have the collection type dependent. You cannot refer to them in your search configuration.

Archive mapping file for the compound email data model

The compound email data model requires several FileNet P8 document classes. One document class contains the properties that are common to all instances of the email document. The content element of this class holds an XML representation of the email. This XML representation is added to the full-text index. The default name for this document class is ICCMailSearch2.

One document class contains the email document and its attachment data in the content element. The default name for this document class is ICCMail2. One document class contains the properties that are specific to an instance of the email document. This document class does not have a content element. Default name for this document class is ICCMailInstance2.

Objects of class ICCMail2 refer to objects of class ICCMailSearch2 and have a multi-valued reference to objects of class ICCMailInstance2. Objects of the class ICCMailSearch2 and ICCMailInstance2 have a reference to objects of class ICCMail2.

The template for the corresponding archive mapping file contains default collection definitions for the document classes as required for the compound email data model. The archive mapping file is basically the same as the one for the bundled email data model with these exceptions:

- The names of document classes and collections have changed.
- The mapping file contains an additional field definition for the correlation keys of attachments. This field also refers from collection Default Mail to collection ICCMail2, like the field definition for MAILBOX_ID does.

```
<field nm="CORRELATION_KEY" type="STRING" multivalue="true">
<reference>MAIL_REF.CORRELATION_KEY</reference>
</field>
```

This property is available in collection ICCMail2 and is defined as follows:

```
<field nm="CORRELATION_KEY" type="STRING" multivalue="true">
<attr>ICCAttachmentCorrelationKeys</attr>
</field>
```

Search configuration file

The template for the search configuration file contains definitions for the search scope, such as the collections and the attributes that are available for search.

Configuration files for documents archived from Notes applications:

The configuration files are used to determine the item types or document classes in the repository that are used for archiving documents from Notes applications. This information is required for retrieving the archived document when a user clicks the respective stub link or restores the document.

Tip: Configure access to documents archived from Notes applications only in combination with access to email documents. If you do not configure email search, you might encounter problems.

The archive mapping template files are available in the directory `<installDir>\AFUWeb\afu\config\templates`, where *installDir* is the directory where you installed IBM Content Collector. Use `afu_cm_application_archiving_search_mapping_template.xml` if your repository is IBM Content Manager, or `afu_p8_application_archiving_search_mapping_template.xml` if your repository is FileNet P8. You do not need to change the search configuration.

The archive mapping template files contain variables that start and end with percent signs (%), such as %P8_DOCUMENT_CLASS%. You must replace the following variables in the template file for your repository with the values for the item type or document class that you created:

Table 184. Variables that must be replaced in the archive mapping file

Repository	Variable to be replaced	Sample value
IBM Content Manager	%ICC_CONFIG_ITEMTYPE_NAME%	ICCEmailCmp
	%ICC_CONFIG_ITEMTYPECHILD_AI%	AFUAChild0001
	%AFU_CONFIG_ATTR_FILENAME%	AFUFileName
	%ICC_CONFIG_ATTR_CORRELATIONKEY%	AFUCorrelationKey
	%ICC_CONFIG_ATTR_CONTENTREF%	AFUContentRef
FileNet P8	%P8_DOCUMENT_CLASS%	ICCMail

The following field names in the default archive mapping template must not be changed:

For IBM Content Manager:

- ATTACH_REF
- ATTACHMENT_NAME
- CONTENT_REF
- CORRELATION_KEY

For FileNet P8:

- CORRELATION_KEY

After you replaced all variables, add the configuration from the template file to your current configuration:

1. In the Configuration Manager, go to **General Settings**->**Archived Data Access**->**Archived Data Access for Email**.
2. On the Advanced page, export the current archive mapping by clicking **Export**.
3. Open the exported XML file and add the data from the template file that you modified. In the template file, copy everything between the start tag `<doc_type_collections>` and the end tag `</doc_type_collections>` and paste it into the exported archive mapping file. Insert the copied text at the end of the file just before the end tag `</doc_type_collections>`.
4. Save the archive mapping file and import it to the Configuration Manager by clicking **Import**.
5. Restart the IBM Content Collector Web Application service.

Configuration files for file system documents:

The configuration files are used to determine the item types or document classes in the repository that are used for archiving file system documents. This information is required for retrieving the archived document when a user clicks the respective stub link.

For IBM Content Manager repositories, the file system item type can be a simple resource item type or a document item type.

For FileNet P8 repositories, the document class that holds file system documents is derived from the FileNet P8 class Document. The actual file content is attached as content element.

The following field names in the default archive mapping template must not be changed:

- CONTENT
- FILENAME
- MIMETYPE (only FileNet P8)

Important: The search configuration file for file system is an internally used file. It cannot be used for configuring search for file system documents.

Configuration files for Microsoft SharePoint documents:

The configuration files determine the IBM Content Manager item types that are used for archiving Microsoft SharePoint documents. This information is required for retrieving the archived document when a user clicks the respective stub link.

For IBM Content Manager repositories, the SharePoint item type can be a simple resource item type or a document item type.

The following field names in the default archive mapping template must not be changed:

- CONTENT
- FILENAME
- MIMETYPE (only FileNet P8)

Important: The search configuration file for Microsoft SharePoint is an internally used file. It cannot be used for configuring search for Microsoft SharePoint documents.

Adding item types or document classes to a collection:

For email search, you can expand the search scope by adding further item types or document classes to a collection definition. For file system or SharePoint documents, you can enable retrieval of archived documents for additional item types or document classes by adding them to a collection definition.

Important: String values for elements in the configuration files must not contain leading or trailing blanks.

You can add an item type or a document class to an existing collection in the archive mapping file either by editing the configuration file or by using the

graphical user interface of the Configuration Manager. If you want to use the Configuration Manager, follow the instructions in the topic about configuring the access to archived data.

To add item types or document classes manually:

1. In the Configuration Manager, select the type of configuration that you want to change:
   - Archived Data Access for Email
   - Archived Data Access for FileSystem
   - Archived Data Access for SharePoint
2. On the **Advanced** page, export the archive mapping file to a directory of your choice.
3. Open the `search_mapping.xml` file with a text editor to modify the archive mapping definition.

   Note: Make sure to save a backup copy of the file before changing or adding any entries.
4. Add an item type or document class.

Option	Description

The item type or document class is located in a repository that is already defined in the archive mapping file.	Copy an existing <doc_type> entry within the existing <repository> definition and change the name, for example:
	`<repository id="abc">`
	`  <doc_types>`
	`  <doc_type>`
	`  <name>ICCMail</name>`
	`  :`
	`  </doc_type>`
	`  <doc_type>`
	`  <name>ICCMailNew</name>`
	`  :`
	`  </doc_type>`
	`  </doc_types>`
	`</repository>`

Option	Description

The item type or document class is located in a repository that is not yet defined in the archive mapping file.	Copy an existing <repository> entry and adapt the repository ID. Then, select one <doc_type> entry, change the name, and delete all unnecessary <doc_type> entries within the new repository definition, for example:
	`<repositories>`
	`  <repository id="abc">`
	`  <doc_types>`
	`  <doc_type>`
	`  <name>ICCMail</name>`
	`  :`
	`  </doc_type>`
	`  </doc_types>`
	`  </repository>`
	`  <repository id="def">`
	`  <doc_types>`
	`  <doc_type>`
	`  <name>ICCMailNew</name>`
	`  :`
	`  </doc_type>`
	`  </doc_types>`
	`  </repository>`
	`</repositories>`

If the <doc_type> entry that you copied contains definitions for child components, make sure that all references are correct. If you have archive mappings where different collections contain the same <doc_type> definition, you must ensure that all field definitions in those collections are the same. To omit problems that might occur because of definition mismatches, use an item type or document class in one collection only.

Remember: All collections within one collection set must have the same attributes and references.
If you want to enable sorting the result by EMAIL_DATE, make sure that, if item types or document classes in a given collection are defined with date ranges (start and end dates), the date ranges do not overlap and that the attribute `partitionkey` is set to true for the field EMAIL_DATE.

5. Save the file.
6. Use the Configuration Manager to import your changes to the database.
7. Save the changed configuration.

Your changes will take effect as soon as the Content Collector Web Application Server is restarted. Until then, Content Collector sessions continue to operate under the old configuration settings.

Related tasks

"Configuring the access to archived data" on page 228

Adding collections:

If your company requires more than the collections that were defined in the initial configuration of IBM Content Collector, you can add collection definitions to the archive mapping and search configuration files.

The following scenarios require that you add collections to the existing configuration files.

- You might have an earlier release of IBM Content Collector and want to upgrade to Version 2.2.

  If you select the bundled email model in the initial configuration of the Content Collector, the previously used configuration files are adapted for use with this model.

  If you select the compound email model in the initial configuration of the Content Collector, you define new item types (for a IBM Content Manager repository) or new document classes (for a FileNet P8 repository). The respective configuration files are created and stored in the\cm or \p8 subdirectory of the directory <*installDir*>\Configuration\initialConfig\data\search\output, where *installDir* is the directory where you installed IBM Content Collector. They are not imported into the configuration database. You must merge the definitions of the old and the new configuration files manually.
- You want to use a different item type for archiving specific documents, for example, all documents that are collected on Friday from a given domain, whereas for all other documents the items types of the existing collection are used.

You can add collections to the archive mapping file either by editing the configuration file or by using the graphical user interface of the Configuration Manager. If you want to use the Configuration Manager, follow the instructions in the topic about configuring the access to archived data. Before you can define further collections manually, you must export the configuration files to a directory of your choice using the Configuration Manager.

Important: String values for elements in the configuration files must not contain leading or trailing blanks.

To add a collection definition manually:

1. In the Configuration Manager, select the type of configuration that you want to change:
   - Archived Data Access for Email
   - Archived Data Access for FileSystem
   - Archived Data Access for SharePoint
2. Export the configuration files to a directory of your choice.
3. Open the search_mapping.xml file and the appropriate template file with a text editor. For configuring archived-data access for email, select the archive mapping template file that corresponds to the repository that you work with, which is either IBM Content Manager or FileNet P8, and addresses the document model of the new collection, which can be bundled or compound. For configuring archived-data access for file system or SharePoint, select the archive mapping template file that corresponds to the repository that you work with and that has filesystem or sharepoint, respectively, as part of the filename.
4. In the template mapping file, select all collection definitions pertaining to an email, attachment, file system, or SharePoint item type or document class and copy them to the search_mapping.xml file, either before or after any existing collection definition. Any new collection can reuse existing collections if those collections contain the correct definitions for fields, references, and <doc_type> elements. For more information see the topic on configuration files that applies to your setup.
5. Exit the template mapping file without saving.

6. To adapt the collection definition in the search_mapping.xml file, change the collection name and the collection ID. The collection name and the collection ID must be identical. Make sure that all references are correct. For example,

   ```
   <doc_type_collection id="OtherMail" nm="OtherMail" collectionType="ICC_Domino_Email_Compound">
   ```

   Replace all entries in between percent signs with valid names that are used in the IBM Content Manager repository. For FileNet P8, check the property and class names that are used in the new collection and change them if required.
7. Save the file.
   - If you change the configuration for archived-data access for file system or SharePoint, proceed with step 14.
   - If you change the configuration for archived-data access for email and you want the new collection to become part of every search, proceed with step 8.

     In this case, sorting the result list by EMAIL_DATE will be disabled. All properties that are defined in the <declaration> element of the collection set to which you add the new collection must have been defined in the archive mapping of that collection.
   - If you change the configuration for archived-data access for email and you do not want the new collection to become part of every search but to offer a choice of collection sets, import the updated configuration file to the database and proceed as described in the topic about providing more than one collection set for search.
8. Open the search_config.xml file and the appropriate template file with a text editor. Your search_config.xml file can contain one or more collection sets. One collection set is defined in one <search-template> section.
9. In the template file, select the collection definition. To make the new collection part of an existing collection set, copy the collection to the search_config.xml file, either before or after any existing collection definition in the selected collection set.
10. Exit the template file without saving.
11. Adapt the collection name and the collection ID in the search_config.xml, for example,

    ```
    <collection name="Other Mail" id="Other Mail">
    ```

    The collection name must match the collection name that you specified in the archive mapping definition. In addition, the collection name and the collection ID must be identical.
12. Make sure that all field definitions in the collection correspond to the definitions that you used when you defined the collection in the archive mapping file. This applies to these elements:

    ```
    <property name="EMAIL_DATE">
      <field>EMAIL_DATE</field>
    </property>
    ```

    If you customized your collection definition, you probably used new names for the fields. If you did not, the default names should match.
13. Save the file.
14. Import the updated configuration files to the database by using the Configuration Manager.
15. Save the changed configuration.

Your changes will take effect as soon as the Content Collector Web Application Server is restarted. Until then, Content Collector sessions continue to operate under the old configuration settings.

Related tasks

"Providing more than one collection set for search" on page 709

"Configuring the access to archived data" on page 228

Enabling search for email documents

Configure the access to archived email in a way that users can search the repository for archived email, considering restrictions and special set up, such as including custom metadata in the search scope.

The following prerequisites apply, depending on the type of repository that you use:

Table 185. Prerequisites for enabling search

Repository	Prerequisites
IBM Content Manager	• The repository is configured and enabled for search. • You configured at least one repository connection for the IBM Content Manager Connector. • The configuration files for the text search indexer contain attribute and field definitions for all custom attributes that you want to make available for search. • An email item type was configured during the initial configuration of Content Collector and is enabled for text search. • These components are configured properly: – Configuration Web Service – Information Center – Web Application
IBM FileNet P8	• The object store must be enabled for content-based retrieval. • You configured at least one repository connection for the IBM FileNet P8 Connector. • The `style.xml` file contains zone definitions for all custom attributes that you want to make available for search, and the attributes must be included in the index for full-text search. • These components are configured properly: – Configuration Web Service – Information Center – Web Application

Tip: If your configuration includes both Lotus Notes and Microsoft Exchange collections, set the system environment variable AFU_DISABLE_URL_CHECK to ensure that all Content Collector clients can display the email search page.

During the initial configuration of IBM Content Collector, configuration files for the selected target system are created and imported into the configuration database. For new installations, these files contain basic definitions for one collection set. For upgrade installations, previously used configuration files are adapted automatically.

To enable search for email documents:

1. Check the definitions for the archived data access for email.
   a. In the Configuration Manager, select **General Settings → Archived Data Access** and select **Archived Data Access for Email**. On the General page, all defined collections and their associated storage templates are listed. For a new installation, this list will contain an entry for the newly defined items only. For an upgrade installation, all collections that are defined in the configuration database are listed.
   b. Optional: Add collection definitions as required.
   c. Check the list of content server properties on the Properties page. Consider that you can use only the fields that are defined here when you map collection fields to content server properties. If required, add, edit, or remove content server properties. You can add any field that is defined on the IBM Content Manager item type or the IBM FileNet P8 document class.
   d. Check the list of text index fields on the Text Index page. Consider that you can use only the fields that are defined here when you map collection fields to text index fields. If required, add, edit, or remove text index fields. You can add any field that is defined in the text indexer model file (IBM Content Manager) or the XIT (IBM FileNet P8).
   e. Save your settings.
2. If necessary, adapt the layout of the search page by updating the search configuration file. You might have to update the <declaration> section, the <form> section, and the <result> section.
   a. On the Advanced page, export the search configuration file to disk.
   b. Update the file as required and save your changes.
   c. On the Advanced page, import the search configuration file from disk.
   d. Save your changes.
3. If you added custom attributes to the search scope, define labels for the search fields.

Restart the IBM Content Collector Web Application service for any changes to take effect.

If you use Microsoft Exchange, this service must be started by an account with administrator privileges for Microsoft Exchange. If you use Lotus Domino, this service must be started by an account that is not the local system account.

Enabling search on custom attributes

If all IBM Content Manager item types or the FileNet P8 document classes ICCMailSearch or ICCMailSearch2 in a given collection contain custom attributes (user-defined metadata) that you want to be available for search or for display in the result list, you can adapt the search configuration files for that collection so that these additional data are used when searching.

To enable search on user-defined metadata, these conditions must be met:

- For automatic archiving, ensure that the required metadata properties are extracted (with the EC Extract Metadata task).
- For interactive archiving, the Metadata Web Application must be configured.
- In IBM Content Manager, you must have added the appropriate attribute and field definitions to the indexer for text search. How to do this is described in the topics on the text-indexer model file.
- In FileNet P8, the `style.xml` file contains a definition for a zone with the name `icc_custom_metadata`. By default, all user-defined metadata is stored in that zone. For example, when you add the user-defined metadata FOLDER and DEPARTMENT, these are both stored in the `icc_custom_metadata` zone. So, when you search for the folder name Contracts, the search will return all email documents that are in the folder Contracts but also all email documents that have the department name Contracts, even if they are not stored in the folder Contracts.

  To have each custom attribute indexed separately in a zone of its own and to enable searches in this specific zone, add a `<preserve xmltag="xxx" >` element for each attribute, for example:

  ```
  <preserve xmltag="icc_custom_metadata" />

    <preserve xmltag="part_number" />
  ```

  Important: If you modify any of the style files that are used to create an index, you must remove the index and then recreate the index for an existing object store. Re-indexing will not incorporate the style file modifications in an existing index.

  To add user-defined metadata to the index for full-text search, you must include a P8 4.x Save Prepared Text as XML task to your archiving task route and configure the mappings according to your needs.
- The configuration for accessing archived email must be customized accordingly.

The following restrictions apply:

- Metadata of the type STRING can be searched in the full-text index, regardless of whether they are common to all instances of an email document.
- Metadata of the types INTEGER and DATE cannot be searched in the full-text index, regardless of whether they are common to all instances of an email document.
- Metadata of the types STRING, DATE, and INTEGER can be searched in the database if they are common to all instances of an email document.

To allow for the search on custom attributes, you have to add these attributes to the archive mapping and search configuration files. You also have to define the labels for the respective fields on the Email Search page of the client application.

Important:

- The name and the data type of any custom attribute must be the same for all item types or document classes in one given collection.
- User-defined metadata that are added to the full-text index are indexed as string values. This also applies to numeric values, such as 123, or date values, such as 03/03/2009. Therefore, range search in the full-text index is not supported for user-defined metadata even if the data type is NUMERIC or DATE. Full-text search is based on string lookup only.

- String values for elements in the configuration files must not contain leading or trailing blanks.
- String values for elements in the configuration files are case sensitive.

Related tasks

"Planning your item type creation" on page 587

"Adding attributes to the model file" on page 614

"Adding attributes to the configuration file of the item type" on page 615

Related reference

"EC Extract Metadata" on page 522

"MC Retrieve Additional Metadata" on page 539

Adding custom attributes to the archive mapping file:

If you defined metadata to be stored in the archive in addition to the metadata provided by the system, you might want to enable search on the user-defined metadata or to have these metadata displayed in the search result list. To do so, you must adapt the archive mapping file by adding a field definition for each custom attribute that you want to use in the search panel.

You can add field definitions to the archive mapping file either by editing the configuration file or by using the graphical user interface of the Configuration Manager. If you want to use the Configuration Manager, follow the instructions in the topic about configuring the access to archived data. Before you can modify the archive mapping file manually, you must export the file to a directory of your choice using the Configuration Manager. Make sure to save a backup copy of the file before changing or adding any entries.

Important: The only date attribute for which date-range queries can be done in the full-text index is the system-defined attribute EMAIL_DATE. This is also the recommended attribute for use as a partition key when item types are split.

Starting with FileNet P8, Version 4.5.1, the system-defined received date as it is defined for EMAIL_DATE is the partition key that is used to organize indexes in FileNet P8. If the object store is set up with date-partitioned collections, the archive mapping definition for the received date must no longer contain a <search> field definition for EMAIL_DATE, but define the <attr> element only. The FileNet P8 repository internally routes searches to the full-text index.

To modify the archive mapping definition manually:

1. Open the `search_mapping.xml` file with a text editor.
2. In the <fields> section of the collection definition that contains the attributes that are common to all email instances, add a <field> element for each custom attribute.

   ```
   <field nm="CUSTOMER_DEFINED_NAME" type="STRING">
    <attr>attributeName</attr>
   </field>
   ```

   where:

   nm Defines the logical name of the new field. It is the name by which the field is addressed from within the search configuration file.

   type Defines the data type of the new field. This applies to search in the

database only. In full-text search, queries are based on comparing string values. No date range or other numeric comparison is supported. The following types are supported:

STRING
EQUAL and LIKE comparisons are supported. LIKE comparison is used if the search term contains wildcard characters.

DATE Date-range search is supported. Note that for IBM Content Manager item types, the referenced attribute must have the IBM Content Manager date type TIMESTAMP.

Per collection, exactly **one** field of the data type DATE can have the additional attribute `partitionkey` set to true. If the attribute is set, this custom date can be used as partition key.

INTEGER
EQUAL comparison is supported.

attributeName
Is the name of the attribute as it is defined in Content Manager or FileNet P8.

The <attr> element is for used for searching the new attribute in the database. This element must be defined if the attribute is to be shown in the search result list or if search is to be done in the database not in the text search index, or both.

3. If you set up the full-text index to include custom attributes, you can define a field for searching on the index.

```
<field nm="icc_custom_metadata" type="STRING">
 <search>icc_custom_metadata</search>
</field>
```

icc_custom_metadata
Is the name of the field in the index. The <search> element defines that a query on this field runs in the full-text index, not the database.

Remember: For IBM Content Manager repositories, you must have added the appropriate attribute and field definitions to the indexer for text search.

In FileNet P8, the `style.xml` file must have been adapted accordingly, otherwise the <search> element must always reference the `icc_custom_metadata` zone and the search will return all documents where any of the metadata matches the search term.

4. Save the file.

You must import the file to the IBM Content Collector data store by using the Configuration Manager and save the changed configuration. Then, restart the Content Collector Web Application Server for the changes to take effect.

Related tasks

"Configuring the access to archived data" on page 228

Adding custom attributes to the search configuration file:

To make custom attributes available for search or display in the result list, you must adapt the search configuration file.

Before you can modify the search configuration file, you have to export the file to a directory of your choice by using the Configuration Manager. Make sure to save a backup copy of the file before changing or adding any entries.

1. Open the search configuration file, `search_config.xml`, with a text editor and modify the search configuration definitions for a given collection.
2. Define each new field that you added to the `search_mapping.xml` file as a property in the <declaration> section of the search configuration file.

   ```
   <property>
    <name>CUSTOMER_DEFINED_NAME</name>
    <nls-key>jsp.searchrequest.CUSTOMER_DEFINED_NAME</nls-key>
   <property>
   ```

   The value *CUSTOMER_DEFINED_NAME* should match the field name that is defined in the archive mapping file.

   The value of the <nls-key> element defines the header of an input field when a group containing the respective property is expanded. It must refer to a property definition in the `custom_label.properties` file (see "Defining labels for the search fields" on page 705).

   Note that you can use each property declaration only once per template.
3. In all collection definitions, add mappings to field names for the newly defined properties. The field name must match the name used in the archive mapping file. The field name is case sensitive.

   ```
   <collections>
     <collection name="CUSTOMIZED" id="CUSTOMIZED">
     <property name="EMAIL_DATE">
      <field>EMAIL_DATE</field>
     </property>
        :
     <property name="CUSTOMER_DEFINED_NAME">
      <field>CUSTOMER_DEFINED_NAME"</field>
     </property>
     </collection>
   </collections>
   ```
4. If you want to use the custom fields as search input fields, add group elements to the <form> section. The <form> section defines the search mask that is shown on the Email Search page.

   Groups bundle input fields. In a collapsible group the fields can be used as one unit for the search. For example, for the group named `Addresses` you can enter one search term, which is looked up in both the Sender and the Recipients fields. If the group is expanded, you can select to search just one of the fields belonging to the group, or you can enter different search terms in the input fields of this group. A non-collapsible group contains just one or more input fields, which are not used as one unit for the search.

   When you define a group, consider that a group cannot be collapsed if it contains fields of incompatible data types (such as DATE and STRING) or if it contains one or more fields of the type DATE.

   ```
   <group id="g4" nls-key="jsp.searchrequest.group4.title"
     collapse-field-nls-key="jsp.searchrequest.group4.collapse"
     collapse-field-tooltip-nls-key="jsp.searchrequest.group4.tooltip"
     collapse-field-example-nls-key="jsp.searchrequest.group4.example">
    <field>
     <name>CUSTOMER_DEFINED_NAME</name>
     <hidden>false</hidden>
     <size>100</size>
     <max-length>500</max-length>
     <tooltip-nls-key>jsp.searchrequest.CUSTOMER_DEFINED_NAME.tooltip</tooltip-nls-key>
     <example-nls-key>jsp.searchrequest.CUSTOMER_DEFINED_NAME.example</example-nls-key>
    </field>
    <field>
     <name>CUSTOMER_DEFINED_NAME_2</name>
   ```

```
  <hidden>false</hidden>
  <size>100</size>
  <max-length>500</max-length>
  <tooltip-nls-key>jsp.searchrequest.CUSTOMER_DEFINED_NAME_2.tooltip</tooltip-nls-key>
  <example-nls-key>jsp.searchrequest.CUSTOMER_DEFINED_NAME_2.example</example-nls-key>
 </field>
</group>
```

- Group attributes

  nls-key
  Defines the header of the group. The attribute must refer to a property definition in the `custom_label.properties` file.

  This attribute is mandatory.

  collapse-field-nls-key
  Defines the header of the input field if the group is collapsed. The attribute must refer to a property definition in the `custom_label.properties` file. If this attribute is defined, the group can be collapsed or expanded. Note that a group containing only one input field cannot be collapsed.

  This attribute is optional.

  collapse-field-tooltip-nls-key
  Defines the hover help that is displayed for the input field if the group is collapsed. The attribute must refer to a property definition in the `custom_label.properties` file.

  This attribute is optional but you should define it if the `collapse-field-nls-key` attribute is defined.

  collapse-field-example-nls-key
  Defines the query example displayed for the input field if the group is collapsed. The attribute must refer to a property definition in the `custom_label.properties` file.

  This attribute is optional but you should define it if the `collapse-field-nls-key` attribute is defined.

- Elements for fields that have a data type other than DATE

  name Is the name of a property that is defined in the <declaration> section.

  This element is mandatory.

  hidden
  Defines a hidden property. If you use this element, set it to false.

  This element is optional. It is for internal use only.

  size Defines the size of the input field.

  This element is optional. If you do not set the size, the default value 100 is used.

  max-length
  Defines the maximum length of the string that you can enter in the input field.

  This element is optional. If you do not set the size, the default value 100 is used.

  tooltip-nls-key
  Defines the hover help that is displayed for the input field if the

group is expanded or if the group is not collapsible. The attribute must refer to a property definition in the `custom_label.properties` file.

This element is optional. If you do not define it, no hover help is available for the field.

example-nls-key
Defines the query example displayed for the input field if the group is expanded or if the group is not collapsible. The attribute must refer to a property definition in the `custom_label.properties` file.

This element is optional. If you do not define it, no example is displayed for the field.

- Elements for fields that have the data type DATE

input-type
Defines the type of input. For fields with the data type DATE, this attribute must be set to daterange.

This attribute is mandatory.

name Is the name of a property that is defined in the <declaration> section.

This element is mandatory.

hidden
Defines a hidden property. If you use this element, set it to false.

This element is optional. It is for internal use only.

size Defines the size of the input field.

This element is optional. If you do not set the size, the default value 100 is used.

max-length
Defines the maximum length of the string that you can enter in the input field.

This element is optional. If you do not set the size, the default value 100 is used.

startdate-tooltip-nls-key
Defines the hover help that is displayed for the start-date selection. You can refer to the value for jsp.searchrequest.date.startdate.prompt that is defined in the IBM Content Collector properties file but you can also define a different value in the `custom_label.properties` file.

This element is mandatory.

enddate-tooltip-nls-key
Defines the hover help that is displayed for the end-date selection. You can refer to the value for jsp.searchrequest.date.enddate.prompt that is defined in the IBM Content Collector properties file but you can also define a different value in the `custom_label.properties` file.

This element is mandatory.

5. Define the columns that are displayed for the result list. Add the required column definitions to the <result> section.

```
<result>
 <column>
  <name>CUSTOMER_DEFINED_NAME</name>
   <width>10%</width>
```

```
  <nls-key>jsp.searchresult.column.CUSTOMER_DEFINED_NAME</nls-key>
  <caption>false</caption>
 </column>
</result>
```

Where:

name Is the name of a property that is defined in the <declaration> section.

This element is mandatory.

width Defines the width of each column to be displayed. The width definitions for all elements in the <result> section must add up to 100%. This value may not be exceeded.

This element is optional. If you do not define it, the value is set by the browser.

nls-key
Defines the header of the result table. The attribute must refer to a property definition in the `custom_label.properties` file.

This element is mandatory.

caption
Adds the attribute to the hover help for a row in the result list if the element is set to true.

This element is optional.

6. Save the file.

You must import the file to the IBM Content Collector data store using the Configuration Manager and save the changed configuration. Then, restart the Content Collector Web Application Server for the changes to take effect.

Defining labels for the search fields:

Before users can search on custom attributes, you have to define the labels for the respective fields on the Email Search page.

The `custom_label.properties` file contains the label definitions. You can either copy the sample file that is delivered with the product, or you can create a new file with that name. The file must be stored in the `afu\config` subdirectory of your WebSphere Application Server installation. For the embedded Web Application Server, this is the directory `<installDir>\AFUWeb\afu\config`, where *installDir* is the directory where you installed IBM Content Collector Server.

If you need to support different languages, translate the properties file and name it in the form `custom_label_language code.properties`, for example, `custom_label_de.properties` for German. Translated versions of the properties file must also be stored in the `afu\config` subdirectory of your WebSphere Application Server installation.

Tip: Define a default file and test the definitions to make sure the Email Search page displays properly before you translate the property file. Make sure to save a backup copy of these files.

1. Open the file `custom_label.properties` with a text editor. If you copied the sample file, make sure that these statements are contained in the file. If you created a new file, include these statement in your file. These statements must be at the top of the properties file.

```
# NLS_ENCODING=UTF-8
# NLS_MESSAGEFORMAT_NONE
```

2. Add or modify the declaration section where you define the headers for the new fields.

```
#####
# declaration
#####
jsp.searchrequest.CUSTOMER_DEFINED_NAME  customer-defined header for the new field
```

3. Add or modify the definitions for the search form.

```
#####
# search form
#####
```

   a. Add or modify the group definitions.

```
# group definitions
#####
jsp.searchrequest.group4.title  Label of the customer-defined group
jsp.searchrequest.group4.tooltip ToolTip for the input field of the collapsed customer-defined group
jsp.searchrequest.group4.example Example text for the input field of the collapsed customer-defined group
jsp.searchrequest.group4.collapse Header for the input field of the collapsed customer-defined group
```

   b. Add or modify the field definitions.

```
# field definitions
#####
jsp.searchrequest.CUSTOMER_DEFINED_NAME.tooltip  customer-defined ToolTip for the new field
jsp.searchrequest.CUSTOMER_DEFINED_NAME.example  customer-defined example for the new field
```

   c. Add or modify the definitions for the result-column headers.

```
jsp.searchrequest.column.CUSTOMER_DEFINED_NAME.tooltip  customer-defined header for the column in the result list
```

4. Save the file. Make sure to use UTF-8 encoding.

Make sure to adapt the configuration for accessing archived data accordingly.

Enabling search on multiple item types or document classes

Configure the access to archived email in a way that users can perform searches across item types or document classes, considering restrictions and special set up, such as including custom metadata in the search scope.

The following prerequisites apply, depending on the type of repository that you use:

Table 186. Prerequisites for enabling search

Repository	Prerequisites
IBM Content Manager	• One or more repository connections are defined. • The item type that you want to add to your search scope must already exist and must be enabled for text search. For more information, see the topic on enabling item types for text indexing. • If you want to make custom attributes available for search, the appropriate attribute and field definitions must exist in the configuration for the indexer for text search. For more information, see the topics on the text-indexer model file.

Table 186. Prerequisites for enabling search (continued)

Repository	Prerequisites
IBM FileNet P8	• One or more repository connections are defined. • If you want to make custom attributes available for search, your `style.xml` file must contain the required definitions, and the attributes must be included in the index for full-text search. For more information, see the topic on enabling the search on custom attributes.

To enable search across multiple item types or document classes in one or more repositories, the definitions must be grouped in the same collection. Define an archive mapping file that contains one collection with multiple item types, either for the bundled email data model or for the compound email data model, or that contains multiple document classes. The item types or document classes can be located in more than one repository. Also add any custom attribute that you want to be searchable to the definitions. Note that the name and the data type of any custom attribute must be the same for all item types or document classes in one given collection. Adapt the search configuration file accordingly. If you included custom attributes, define the labels for the respective fields on the Email search page of the client application.

Restriction: When IBM Content Manager item types are defined with overlapping date ranges, users cannot sort the result list by date.

1. Adapt the archive mapping file.
   - Use the graphical user interface of the Configuration Manager to add item types or document classes and to add definitions for the custom attributes that you want to include in the search scope.
   - Edit the archive mapping file to add item types or document classes and to add definitions for the custom attributes that you want to include in the search scope.
2. If necessary, adapt the layout of the search page by updating the search configuration file. You might have to update the <declaration> section, the <form> section, and the <result> section.
3. If you added custom attributes to the search scope, define labels for the search fields.

Related concepts

"The model file" on page 613

"Enabling search on custom attributes" on page 698

Related tasks

"Enabling an IBM Content Manager repository for processing by the indexer for text search" on page 579

"Adding item types or document classes to a collection" on page 692

Enabling search on multiple collections

Configure the access to archived email in a way that users can perform searches across collections, considering restrictions and special set up, such as including custom metadata in the search scope.

The following prerequisites apply, depending on the type of repository that you use:

Table 187. Prerequisites for enabling search

Repository	Prerequisites
IBM Content Manager	• One or more repository connections are defined. • All item types that are referenced in a collection definition must be enabled for text search. For more information, see the topic on enabling item types for text indexing. • If you want to make custom attributes available for search, the appropriate attribute and field definitions must exist in the configuration for the indexer for text search. For more information, see the topics on the text-indexer model file.
IBM FileNet P8	• One or more repository connections are defined. • If you want to make custom attributes available for search, your `style.xml` file must contain the required definitions, and the attributes must be included in the index for full-text search. For more information, see the topic on enabling the search on custom attributes.

To enable searches across collections, you define multiple collections and group them into a collection set. Defining multiple collections is necessary if you use different item types for archiving documents specifically, for example, all documents that are collected from a given domain are archived in one item type.

Define an archive mapping file that contains all collections that you want to be searchable. You can combine collections that contain item types or document classes for the bundled email data model and collections that contain item types or document classes for the compound email data model. Also add any custom attribute that you want to be searchable to the definitions. Note that the name and the data type of any custom attribute must be the same for all item types in all collections in the collection set. Adapt the search configuration file accordingly. If you included custom attributes, define the labels for the respective fields on the Email search page of the client application.

Restriction: With cross-collection search, users cannot sort the result list by date.

1. Adapt the archive mapping file.
   - Use the graphical user interface of the Configuration Manager to add collections and to add definitions for the custom attributes that you want to include in the search scope.
   - Edit the archive mapping file to add collections and to add definitions for the custom attributes that you want to include in the search scope.
2. Adapt the layout of the Email Search page by updating the search configuration file. You might have to update the <declaration> section, the <form> section, and the <result> section.

3. If you added custom attributes to the search scope, define labels for the search fields.

Related concepts

"The model file" on page 613

Related tasks

"Enabling an IBM Content Manager repository for processing by the indexer for text search" on page 579

"Adding collections" on page 694

"Adding custom attributes to the archive mapping file" on page 700

"Adding custom attributes to the search configuration file" on page 701

"Defining labels for the search fields" on page 705

Providing more than one collection set for search

For email search, it can be appropriate to define more than one set of collections that users can search, depending on your company's requirements. When you group collections into collection sets, you define a search mask for each collection set. For each collection set, you can have a different layout of the Email Search page to enable users to search for different types of information.

When you define several collection sets, the Email Search page offers a list of collection sets from which users can choose. Searches will be done on all collections that are defined within the selected set. Depending on the definitions in the search configuration file, the search fields vary. For example, you can define a collection set that provides search fields for sender and recipient information and for the date and time on which the email was received, a collection set that provides the possibility to search for the subject of an email or for specific text in the email content, and a collection set that allows for searches on custom attributes that you defined earlier.

To add further sets of collections to the search scope:

1. In the Configuration Manager, select to configure archived-data access for email.
2. Export the search configuration file to a directory of your choice. The configuration file is here referred to as `search_config.xml` file.
3. Open the `search_config.xml` file and one of the template files for email with a text editor. The file names of these template files contain the string `search_config_template`.
4. In the template file, select the collection set definition. The collection set is defined by the content of the <search-template> element. Copy the collection-set definition to the `search_config.xml` file, either before or after any existing collection-set definition but in any case before the </search-template-list> element. Include a <search-template> element with the required definitions for each collection set that you want to make available for selection.

```
<search-template name="YourSet" id="YourSet">
   <form-index>0</form-index>
   <declaration>
    <property>
     <name>EMAIL_DATE</name>
     <nls-key>jsp.searchrequest.daterange</nls-key>
     <sortable>true</sortable>
    </property>
     .
   </declaration>

   <collections>
```

```
        .
        </collection>
    </collections>

    <form>
     .
    </form>

    <result>
     .
    </result>
</search-template>
```

5. Exit the template file without saving.
6. Adapt the definition in the `search_config.xml` file to your needs. At least, change the name of the collection set. For example,

   ```
   <search-template name="YourSet" id="YourSet">
   ```

   Important:
   - All names are case sensitive.
   - Names and IDs must be identical.
   - Template names and template IDs must consist of alphabetic characters. Do not use any special characters, digits, or blanks in template names or template IDs.

   a. Specify a name that is meaningful to you and your users. Users must select a set before they can search.
   b. Check the <declaration> section. Only properties that are defined in this section are available within the template.
   c. Change the property mappings in the collection definition All field names that are used here to map the properties of this template to a field name must match field names that are defined in the archive mapping file for the given collection.
7. Save the file.
8. Import the updated configuration file to the database by using the Configuration Manager.
9. Save the changed configuration.

Your changes will take effect as soon as the Content Collector Web Application Server is restarted. Until then, Content Collector sessions continue to operate under the old configuration settings.

Enabling the access to archived data for legacy item types

You can make IBM Content Manager repositories that were previously populated by IBM CommonStore for Exchange Server or IBM CommonStore for Lotus Domino available for search and restoring. The item types of the documents in these repositories are referred to as *legacy item types*.

Only legacy item types for the document model BUNDLED and the archiving type ENTIRE can be previewed and restored by using Content Collector Email Search. Document collections of other item types are not supported by the search application.

IBM Content Manager OnDemand and IBM Tivoli Storage Manager repositories and repositories that were populated byIBM FileNet Email Manager are not supported.

To configure the Content Collector to work with legacy item types:

1. Set the system environment variable AFU_MAILBOXID_ENABLE_EXACT_MATCH. You can use any value. This is required for Microsoft Exchange collections only.
2. Use the Configuration Manager to export the archive mapping file to a directory of your choice.
3. Add a collection definition.
4. Adapt the collection definition in the `search_mapping.xml` file:
   a. Change the collection name.
   b. Change the value of the <retrievable> element to false if the item type for the document model (or archive data model) is not of type BUNDLED, or the archiving type is not ENTIRE. See keyword ARCHIVETYPE in the CommonStore server configuration file. Setting the documents in a collection to be non-retrievable by setting the element to false prevents an error from being displayed in the search application if these documents are selected. Instead, a message is displayed that states that accessing the document is not supported. The following example shows the <retrievable> element set to true:

```
<doc_type_collection nm="CSMAIL" collectionType="ICC_Exchange_Email_Bundled">
  <retrievable>true</retrievable>
  <doc_types>
    <doc_type>
      <name>CSXMAIL</name>
    ...
    </doc_type>
   </doc_types>
 ...
</doc_type_collection>
```

5. Save the file.
6. Import the archive mapping file by using the Configuration Manager.
7. If you changed the collection name in the archive mapping file, change the collection name in the `search_config.xml` file to the name that you used in the `search_mapping.xml` file:
   a. Export the `search_config.xml` file by using the Configuration Manager to a directory of your choice.
   b. Change the name of the collection to the name that you used in the archive mapping file (in attribute nm="CSMAIL"). The name must be identical and is case sensitive. For example:

```
<collection>
    <name>CSMAIL</name>
    <declaration>
      <property>
        <name>EMAIL_DATE</name>
        <nls-key>jsp.searchrequest.daterange
...
```

   c. Save the `search_config.xml` file and import it by using the Configuration Manager.
8. Save the configuration.
9. Restart the Web Application Server.

Related tasks

"Adding collections" on page 694

"Viewing and restoring email from the search result list" on page 785

Changing the display format of the date value

If you want to change the way in which a date value is displayed on the Email Search page, you must adapt the message file.

The pattern that defines the display format for date values in the Email Search page is defined in the language-specific message file. For the English language, this is the `messages_en.properties` file. It is located in the `afu\config` subdirectory of your WebSphere Application Server installation. For the embedded Web Application Server, this is the directory `<installDir>\AFUWeb\afu\config`, where *installDir* is the directory where you installed the IBM Content Collector.

The message file contains a definition for the pattern that is used to format the string that is displayed as input selection for the date picker, for example:

```
jsp.searchrequest.date.pattern MM/dd/yyyy
```

Where:

dd Is day in month

MM Is month in year

yyyy Is the year

You can rearrange the values and use different separators, for example, periods, but you must use the listed tokens when specifying the date. Note that these tokens are case-sensitive.

You can change the format either globally for a given language, for example, for the English language, or for a given language in a specific country, for example, for the English language if the country is the United Kingdom.

By default, the language code is part of the name of the message file: `messages_<languageCode>.properties`. If you want to change the pattern globally for a given language, change the pattern in the respective message file, for example, the `messages_en.properties` file.

If you want to change the pattern for a given language only for a specific country, copy the existing `messages_<languageCode>.properties` file to a file named `messages_<languageCode>_<countryCode>.properties` and change the pattern string in the newly created file. For example, if you want to keep the US pattern for date format as the default format but want to change the display format for the UK, create a new file named `messages_en_gb.properties` and change the pattern in that file.

Adapt the strings for the tooltip accordingly:

```
jsp.searchrequest.date.invalidformat    The date format is not correct. Enter a date in
                    the format dd/MM/yyyy.
jsp.searchrequest.date.startdate.prompt  Select a start date from the calendar or enter a
                    start date in the format dd/MM/yyyy.
jsp.searchrequest.date.enddate.prompt   Select an end date from the calendar or enter an
                    end date in the format dd/MM/yyyy.
```

Important:
Create backup copies of the changed of newly created files because a new installation of the Web Application will overwrite any changes. In this case, you must redo the changes.

Setting a default date range for the Email Search page

To make use of date partitioning and to reduce the number of queries that are performed against multiple item types or segmented indexes, restrict searches to a given time interval.

By setting the system environment variable AFU_DATE_RANGE_IN_MONTHS you can define the default date range that is displayed in the date fields of the Email Search page. The environment variable is of type integer and can have the following values:

0 No default date range is displayed.

n The value must be greater than 0. The default date range is the first day of the "current minus *n*" month until the current date. For example, if the current date was November 30th, 2009 and AFU_DATE_RANGE_IN_MONTHS is set to 1, the default date range that is displayed on the Search page is October 1st, 2009 until November 30th, 2009.

If this environment variable is not set or set to a value less than 0, no default date range is displayed.

Changing the preview mode for Outlook

To prevent errors when users click a preview link in an environment where the Exchange server supports Unicode message file content and users work with older Outlook clients that do not support Unicode format, you can set the environment variable AFU_PREVIEW_MODE_ONLY.

When a user clicks on a preview link in a stubbed Outlook email, the Web Application sends the file content of that message to the client and the Outlook client displays it. This will not work in the described environment. Instead of an Outlook window that displays the message, users get errors when they click a preview link.

If you set the environment variable AFU_PREVIEW_MODE_ONLY on the machine that hosts the Web Application, the Web Application will not send the message file back to the caller. Instead, the email preview page is displayed in the browser. You can set the environment variable to any value.

Handling erroneous documents

If a document cannot be processed successfully, it is added to the blacklist. Using the information in the blacklist entry, you can decide how you want to handle erroneous documents.

To handle erroneous documents:

1. Select **Tools** → **Blacklist**. The blacklist entry contains the following information:
   - For which connector the processing failed.
   - Location information for the erroneous document.
   - In which task the error occurred.

- The cause of the error.
- The dates when the entry was added to the blacklist and when the processing failed the last time.
- The failure count.

Important: A maximum of 32,000 entries can be displayed in the Blacklist window. If the number of blacklist entries exceeds 32,000 and you want to view all of the entries, you can use your database client software to directly query the Black_List table in the Content Collector configuration database. Do not edit or remove entries from the Black_List table.

2. Double-click an entry to view the detailed location information for the blacklisted document. The Blacklisted Document window provides the following information:
   - The name of the server on which the mail box is located
   - The identifier of the mailbox
   - The address of the mailbox, if available
   - The unique id of the blacklisted document:
     - The document UNID for Lotus Domino documents
     - The document EntryId for Microsoft Exchange documents
     - The file location for SMTP email
   - A link to the blacklisted document
3. Optional: Click the link to open the blacklisted document or to copy the document link. The resulting action depends on the type of connector that processed the erroneous document:

Connector	Result
Lotus Domino Email Connector	Copy the link to a command prompt on a machine where a Notes client is installed and run the command. The Notes client is opened and you are prompted for credentials. The ID file that you use for log on must have at least read access to the mailbox where the document is located.
Microsoft Exchange Email Connector	An Outlook client is opened and you are prompted for a profile. You can use an existing profile, or you can create a new profile for accessing the mailbox. However, the profile must include sufficient access rights for the mailbox where the document is located. You can also copy the link to a different machine or send the link to another user to work with it. **Restriction:** Outlook hyperlinks do not work in applications other than Outlook unless you apply the changes to the registry as described in the topic about enabling Outlook links.
SMTP Connector	The program that is associated with EML files is launched, and the document is displayed.

4. Check the document and the task route that processed the document to possibly solve the problem and proceed with the next step.
5. Depending on the result of the previous step, choose one of these processing options:

Option	Description
Except the document from further processing	Delete the erroneous document yourself, or copy the document link and send it to the document owner to have the owner delete the document. The document cannot be processed any more.
Resubmit the document for processing	Delete the blacklist entry for the erroneous document if the maximum number of times was reached that a collector is allowed to process documents again. The failure count is reset to zero and the document qualifies again for processing.

When you encounter a problem with an erroneous document that you cannot solve, contact IBM Software Support.

Related concepts

"Blacklist"

Related tasks

"Configuring the Email Connector" on page 185

"Configuring the SMTP Connector" on page 189

Blacklist

If a document cannot be processed because it causes errors in any of the Email Connector or SMTP Connector tasks in a task route or even causes processing to be completely halted, the document is added to the *blacklist* to prevent IBM Content Collector from reprocessing this document. This can be temporary or, if the processing of a document fails repeatedly or causes the failure of a connector, permanent. Monitor the blacklist for documents that are marked as being permanently in error or that have reached the retry limit and take the appropriate action to handle these errors.

Managing blacklist entries

A blacklist entry contains the following information:

- The name of the connector for which the processing failed. This can be the ID of an Email Connector for Lotus Domino or Microsoft Exchange, or the ID of an SMTP Connector.
- Location information for the erroneous document. This information provides access to the erroneous document so that you can check its content to possibly identify the cause of the error. The location information also helps you to identify patterns in processing errors. For example, if a corrupted attachment caused a processing error, archiving of this attachment will probably fail for all of the recipients' mailboxes. If a large number of documents cannot be processed for a certain mailbox, consider checking the consistency of the mailbox.
- The task where the processing failed.
- The reason why the document was blacklisted, that is, if applicable, the exact cause of the error.
- The date when the entry was added to the blacklist.

- The date when the processing failed the last time.
- The failure count. This column states the number of times that processing for the document failed. If the limit that you specified in the connector configuration is reached, IBM Content Collector assumes that a general problem exists with this document. The failure count is then set to permanent, and the collector ignores the erroneous document. If a document causes the connector to fail, the failure count is also set to permanent.

The blacklist is kept in the configuration database. All mail connectors can add, update, or delete blacklist entries. At connector startup or when the configuration database is synchronized, the Task Routing Engine notifies the connector of any blacklist entries. When Content Collector collects documents, it checks the unique document ID of each document against the blacklist. Note that if the size of the blacklist for a specific mailbox exceeds 40,000 entries, the process for checking the blacklist changes. This has an impact on the performance. If no blacklist entry exists for a document, it is passed to the task route for processing. If a blacklist entry exists for a document, Content Collector checks the data in the entry against the respective settings in the connector configuration:

1. The failure count against the specified maximum number of times that a collector attempts to process documents again. As long as the specified limit is not reached, Content Collector checks the retry interval and, if applicable, the collector processes the document again. If the processing fails again, the failure count is increased by one. When the specified limit is reached and the document could not be processed successfully, the failure count is set to permanent. Then, the erroneous document is ignored.
2. Unless the failure count for the document has reached the specified limit, Content Collector checks the time that passed since the date of the last failure against the specified retry interval. If the retry interval for the document has passed, the document is processed again. If the retry interval did not yet pass, the document is skipped.

Blacklist entries are deleted automatically in these cases:

- If the document was processed successfully.
- If the entry is outdated.

  Content Collector internally keeps track of when a blacklisted document was last collected by a collector. Once a week, Content Collector checks for documents that were not collected for at least four weeks and then deletes the blacklist entries for such documents. This is especially important for Microsoft Exchange mail sources because the identifier of an email changes when it is moved to a different folder. Therefore, orphaned entries might be generated if users move documents that failed to process.

Manual changes to the blacklist, such as deleting an entry, are updated to the connector services when the datastore is synchronized.

In cases where you have to contact IBM Software Support for assistance, copy the contents of the `SUPPORT` directory, which is a subdirectory of the connector's log file directory, and provide that information to IBM Software Support.

Enabling Microsoft Outlook links

To enable Microsoft Outlook hyperlinks for applications other than Outlook you have to define a new key in the Windows registry.

To modify the Windows registry:

1. Run **regedit.exe** to start the Windows Registry Editor.
2. Locate and select the root key HKEY_CLASSES_ROOT.
3. Click **Edit** → **New** → **Key**.
4. Enter the name `outlook` for the new key.
5. Adapt the settings for this new key:
   a. Select the entry (Default), which is a string value and click **Edit** → **Modify**.
   b. In the **Value data** field, enter `URL:Outlook Folders` and click **OK**.
   c. Select the key outlook and click **Edit** → **New** → **String value**.
   d. Enter the name `URL Protocol`. This string value does not require value data.
   e. Add the subkey DefaultIcon to the key outlook. Select the key outlook and click **Edit** → **New** → **Key** and enter the name `DefaultIcon`.
   f. Select the key DefaultIcon and enter as value data for the (Default) value the complete path to your `outlook.exe` , for example, `C:\Program Files\Microsoft Office\Office12\OUTLOOK.EXE`.
   g. Add the subkey shell to the key outlook.
   h. Add the subkey open to the key shell.
   i. Add the subkey command to the key open.
   j. Select the key command enter as value data for the (Default) value the following string:

      `"outlook_path" /select "%1"`Where *outlook_path* is the complete path to your `outlook.exe` as defined for the key DefaultIcon. The double quotation marks are a required part of the string.
6. Close the Windows Registry Editor.

Securing Content Collector communications

When you run IBM Content Collector in a production environment, you must ensure that all communication is secure and trusted.

There are three different kinds of connections in an IBM Content Collector system setup. All of them must be secure.

Connections between servers

An example for a connection between servers is the communication between the Content Collector server, the email server, and the repository server.

This security must be provided by the data center IT environment by using secure transport, for example over VPN connections. Additionally, all systems must be managed in a way that they can be considered secure, which means they must for example be secured by access control to the system and on the system and adhere to a suitable password policy.

Connections between clients and source servers

An example for a connection between clients and source servers is the communication between the email client and the email server.

This security must be provided by the respective clients and servers.

Connections between Content Collector clients and the Content Collector server

Content Collector clients are the email and SharePoint clients that are enabled to provide IBM Content Collector functions and the user workstations that access file system stubs.

This security must be controlled by the IBM Content Collector system. If you use a trusted certificate issued by a certificate authority, all IBM Content Collector communications are secure.

To ensure that the communication between the Content Collector clients and the Content Collector server is secure, you must replace the Secure Sockets Layer (SSL) certificates for the Web Application Server by certificates that are issued by a certificate authority.

Important: Never use self-signed certificates in a production environment.

Replacing certificates for the embedded Web Application Server

The embedded Web Application Server creates a set of default Secure Sockets Layer (SSL) certificates with default credentials. These are used for the initial configuration of the Web Application Server. To enable a secure and trusted environment, you must replace these certificates and credentials with certificates signed by a trusted certificate authority, especially in a production environment.

Prerequisites:

- The IBM Content Collector server must be installed.
- The IBM Content Collector Web Application service must have been deployed.

Before you can replace an SSL certificate, you have to request a new certificate. You request, receive, and replace SSL certificates for the embedded Web Application Server by using the IBM Key Management utility. You can use the same utility to add certificates for additional web servers or to change the credentials when the currently used credentials have expired.

Note: If you do not use the embedded Web Application Server, you can create a certificate authority request and receive the signed certificate by using the WebSphere Application Server AdminTask object. How to do this is described in the IBM WebSphere Application Server documentation.

1. Request a new certificate.
   a. Log on to the computer on which the IBM Content Collector server is installed.
   b. In a command prompt, go to the `<ICCinstallDir>\AfuWeb\profiles\AFUWeb\bin` directory, where `<ICCinstallDir>` is the installation directory of the IBM Content Collector server.
   c. Type `ikeyman`

      The IBM Key Management utility opens.
   d. In the IBM Key Management utility, select **Key Database File** → **Open**.
   e. Select **PKCS12** as key database type.
   f. In the **File name** file, specify the file name `key.p12`.
   g. In the **Location** field, specify the `<ICCinstallDir>\AFUWeb\profiles\AFUWeb\config\cells\<cell name>\nodes\<node name>` directory.

      Replace <ICCinstallDir>, <cell name>, and <node name> with the proper values of your installation.
   h. Click **OK**.
   i. When prompted for a password, enter the password. Click **OK**. The default password is WebAS. Note that the password is case sensitive. In a production environment, change the password as described in the topic

about updating default key store passwords using scripting in the WebSphere Application Server (Distributed platforms and Windows), Version 7.0 Information Center.

j. Create a new certificate request. Under **Key database content**, select **Personal Certificates Requests** and click **New**.

k. In the **Key Label** field, specify a label for the digital certificate request, for example, `Production Certificate for Content Collector`.

l. For the remaining fields, accept the default values.

m. Click **OK**. A confirmation window is displayed, verifying that you have created a request for a new digital certificate.

   The **Personal Certificate Requests** field in the IBM Key Management window shows the key label of the new digital certificate request you created.

n. Send the file to a certificate authority (CA) to request a new digital certificate, or cut and paste the request into the request forms of the CA's website. If you have a Windows Domain CA, you can follow the procedure described in "Submitting a certificate request" on page 720 to do so. If you use a different CA to certify the certificate request, follow the procedure that applies for the respective CA.

After the CA sends you a new digital certificate, you must delete the existing certificate and add the new one to the key database from which you generated the request.

2. Delete the existing certificate. Note: Before deleting a digital certificate, create a backup copy in case you later want to recreate it.
   a. In the IBM Key Management utility, make sure that the key database file is open and that, under **Key database content**, **Personal Certificates** and **default** are selected.
   b. Click **Delete**. You are asked to confirm the deletion. The label of the digital certificate you just deleted no longer appears in the **Personal Certificates** field of the IBM Key Management window.
3. Receive the new certificate to replace the existing one.
   a. Click **Receive**. The Receive Certificate from a File window is displayed.
   b. Select **Binary DER data** as the data type of the new certificate. If the CA sends the certificate as part of an email, you might need to cut and paste the certificate into a separate file.
   c. Accept the default values for the certificate and click **OK**.
   d. Specify a label, such as `Production Certificate for Content Collector`, for the new certificate and click **OK**. The **Personal Certificates** field of the IBM Key Management window shows the label of the new certificate.
   e. Exit the IBM Key Management utility.
4. Stop and restart the service for the embedded Web Application Server (IBM Content Collector Web Application service). You can use the Start menu on a Microsoft Windows system.
   a. To stop the service, click **Start → All Programs → IBM Content Collector → Stop Services → Stop ICC Web Applications**.
   b. To restart the service, click **Start → All Programs → IBM Content Collector → Start Services → Start ICC Web Applications**.

Important: If you use **Microsoft Exchange**, the IBM Content Collector Web Application service must be started by an account with administrator privileges for Microsoft Exchange.

5. To check if the new certificate works, open your web browser and enter the following URL in the address field:

   ```
   https://<server host name>:11443/AFUWeb/init
   ```

   where

   <server host name>
   Is the host name of the computer running the embedded Web Application Server. This is the same as the computer running the IBM Content Collector server.

   11443
   Is the default port for connections to the embedded Web Application Server

   You should be able to establish an HTTPS connection. If you receive security warnings in your browser, import the public key certificate of your certificate authority into your browser.

Submitting a certificate request

If you have a Windows Domain CA and web-based access to Certificate Services is enabled, you can submit a certificate request by following these steps.

1. Access the Certificates Services by specifying the following URL in your web browser:

   ```
   http://<ca_iis_server>/certsrv/certrqxt.asp
   ```

   where *<ca_iis_server>* is the DNS or NetBIOS name of the host server.
2. Paste the request into the form. You can browse for the .arm file that you created with the IBM Key management utility.
3. Under **Certificate Template**, select **Web Server**.
4. Click **Submit**.
5. Download the certificate.

Client communication

The communication between the IBM Content Collector server and the Content Collector clients uses Hypertext Transfer Protocol Secure (HTTPS) connections.

HTTPS connections are secure if they use a trusted certificate issued by a certificate authority. This means that you must replace the self-signed certificates that are created by the Web Application Server with trusted certificates. For detailed information on how to do that, see the related topic.

IBM Content Collector clients are:

- Email clients that are supported by IBM Content Collector and that are enabled to provide Content Collector functions, like for example Lotus Notes, iNotes, Microsoft Exchange, and Outlook Web Access
- User workstations that access file system stubs
- SharePoint clients

Related tasks

"Replacing certificates for the embedded Web Application Server" on page 718

URL protection

To prevent URLs from manipulation, they must be protected. How they are protected depends on whether the URLs are dynamic or static.

URL protection for dynamic URLs

Dynamic URLs are generated for temporary requests. They do not need to be persistent. Most typically, dynamic URLs are used for interactive requests where a user requests an interactive web application from the Web Application Server.

Dynamic URLs are protected by the following means:

Expiration time stamp
: An expiration time stamp invalidates the URL after a certain amount of time. This applies only to URLs that are used for interactive search in the archive.

URL signature
: A URL signature ensures that a manipulation of any URL parameter invalidates the URL.

Functions that make use of dynamic URLs in IBM Content Collector are:

- Interactive restore of documents
- Interactive viewing of documents
- Interactive search in the archive

URL protection for static URLs

Static URLs are used to access archived content in the repository. Because the URLs are added to the document stubs and provide access to the archived content, they must be static and valid as long as the stub exists. It is thus not possible to protect the URL with an expiration time stamp or dynamic parameter replacement.

Static URLs are protected by the following means:

Access control
: Access to the document containing the URL is restricted according to the access control for the original document.

URL signature
: A URL signature ensures that a manipulation of any URL parameter invalidates the URL.

Functions that make use of static URLs in IBM Content Collector are:

- Stub URLs for email and attachments
- Stub URLs for documents archived from the file system
- Stub URLs for documents archived from Microsoft SharePoint
- URL redirection for viewing documents in applications like for example Workplace XT

The access to the document containing the URL is managed by the access control method of the source system and client. For example, access is controlled by the email system, by access control lists for file system, or by the SharePoint system.

The URL provides direct access to archived content. URLs for email and attachments do not require an additional logon to the repository, this means that the client workstations do not need to have repository clients or APIs installed. URLs for documents archived from the file system or from Microsoft SharePoint to Content Manager can require a logon to the repository, depending on how the link is configured in the task route.

IBM Content Collector provides access to archived content through the Web Application running on the Web Application Server. All content access requests pass through the IBM Content Collector Web Application, which verifies that the URL requesting the content is valid.For most retrieval requests, Content Collector uses the repository connection that was established for the user ID that is defined in the connection configuration in the IBM Content Collector Configuration Manager. If file system or Microsoft SharePoint task routes are configured accordingly, a logon page is displayed when users click a stub URL. In this case, a user-specific repository connection is established by using the user's credentials and access rights.

The retrieved content is then sent back to the Web Application and passed on to the client. The URL is never redirected directly to the repository, and the repository never sends the content directly to the client. The SharePoint link handler provides an extra layer of security by requiring users to retrieve linked documents from within SharePoint.

Tip: To prevent other users from accessing protected content, users should clear their web browser cache and history after working with static URLs, because the content of the documents might be cached.

Part 5. Tutorials

© Copyright IBM Corp. 2008, 2010

Content Collector file system tutorials

Use the file system tutorials to learn how to set up task routes to meet your archiving needs.

Archiving file system documents to FileNet P8

You can use the File System Source Connector to move documents off a network into an IBM FileNet P8 repository, detect and process duplicates, and define metadata to be used to process files for archiving.

The File System Source Connector provides a connection from IBM Content Collector to a file system. You can then set up a file system collector to retrieve files from the file server on a scheduled time and submit the files to task route.

A task route defines a series of tasks that process documents, including moving the document from the file server to a document repository. A task route also includes rules and decision points that determine which task in the task route is processed next. For example, you might want a document deleted from the file server after it is moved to the document repository only if it is over a certain age.

Moving documents off the network into IBM FileNet P8

Storing documents in a folder on a shared network can make your documents unsecured because they can be changed or deleted by anyone. To avoid this, you can archive all of your documents in a repository, and delete the copy on the file system.

After a file is archived in the repository, the file can be read but not modified.

Content Collector can copy documents to a central repository. When files are archived, the original file is stubbed. When users return to a network folder to find a document that was archived, users can click a link to access the document in its new repository location and view a read-only copy of the file.

The task route includes an error task route. If an error occurs when moving documents off the network, such as if FileNet P8 is not reachable and a file is not archived, that file can be preserved with a flag that indicates it was not archived.

To move documents off the network:

1. Modify the `FS to P8 Archiving (Delete).ctms` task route to your environment. This template contains a simple task route for archiving email automatically.
   a. Open the IBM Content Collector Configuration Manager and click the **Task Route** tab.
   b. In the **Task Routes** view, click the **New** icon.
   c. In the **Choose a template** view and select the `FS to P8 Archiving (Delete) - Complete.ctms` template and click **OK**. This will create a new version of the task route that you can modify to your environment.
   d. Configure a collection source (a place to find files to process) by clicking **FSC Collector** in the Main Task Route window. File system task routes require a file collector. The job of a file collector is to monitor locations that you specify for files to be processed by IBM Content Collector.

© Copyright IBM Corp. 2008, 2010

e. Configure the **P8 4.x Create Document** task to specify the repository (document class) in which to save files. Replace HOST:PORT in the shortcut link with the server name and port of the Web Application that retrieves the objects.

f. Set postprocessing options for files on the file system after archiving, such as delete the file, rename the file, or mark the file as processed, by modifying the **FSC Post Processing** task. You must specify what to do with files that are left on the file system after they are archived.

2. Replace files on the file system with shortcuts to archived files by modifying the `FS to P8 Archiving (Shortcut).ctms` template.

   When you set up automatic archiving of files on the file system, IBM Content Collector copies the content of these files to the repository that was specified during the initial configuration. The original files are deleted after they are archived, and each file is replaced with a shortcut to the archived document. This saves space on the file system, and users can quickly view archived files from their usual place on the file system.

   a. In the **Task Routes** view, click the **New** icon again and in the **Choose a template** view, select the `FS to P8 Archiving (Shortcut).ctms` template. This will create a new version of the task route that you can modify to your environment.

   b. Configure a collection source (a place to find files to process) by clicking **FSC Collector** in the Main Task Route window.

   c. Configure the **P8 4.x Create Document** task to specify the repository (document class) in which to save files. Replace HOST:PORT in the shortcut link with the server name and port of the Web Application that retrieves the objects.

   d. Set postprocessing options for files on the file system after archiving, such as delete the file, rename the file, or mark the file as processed, by modifying the **FSC Post Processing** task. You must specify what to do with files that are left on the file system after they are archived.

3. Check the error task routes of both the FS to P8 Archiving (Delete) and FS to P8 Archiving (Shortcut) task routes by clicking the **Switch between main and error task route** icon. Files that result in errors are moved to a folder by the **FSC Post Processing** task so that the files can be reviewed by an administrator. To include different logging information in the audit logs of the main task route and the error task route, all error task routes should contain an audit log task.

   Note: If you import a task route that contains an audit log and was exported with a previous version of IBM Content Collector, the error task route will not contain an audit log. You must add one manually.

Related tasks

"Creating a task route" on page 275

"Collecting file system documents" on page 479

Related reference

"FSC Post Processing" on page 534

Detecting and processing duplicates, searching for archived and stubbed documents, and declaring documents as records

If you do not want to archive duplicate files in the repository, you can set up the file system collector to archive only one copy of a duplicate file. In addition, you can set up a shortcut to access an archived document by replicating the source file system directory structure in the document repository.

IBM Content Collector can create links on the file system for each copy of a duplicate file and each link points to the same archived file.

To ensure that a document cannot be changed or deleted, declare each archived file as a record, in accordance with your retention policies.

During this task route process:

- Duplicate files are detected. Only one copy of the duplicates is archived; additional copies are deleted. Each file on the original system is stubbed, each link pointing to one archived file.
- The source file system directory structure is replicated in the document repository.
- As part of the archive process, each file saved in the repository is declared as a record.

To set up the file system collector to detect and process duplicates:

1. Create a task route for duplication detection by using:

Option	Description
A task route template:	1. Open the IBM Content Collector Configuration Manager and click the **Task Route** tab. 2. In the **Task Routes** view, click the **New** icon. 3. In the **Choose a template** view and select `FS to P8 Archiving (Detect Duplicates and Delete).ctms` and click **OK** to modify the template to your environment.
A new task route:	1. Open the IBM Content Collector Configuration Manager and click the **Task Route** tab. 2. In the **Task Routes** view, click the **New** icon. 3. In the **Choose a template** view, select `Blank task route` and enter a name. 4. Click **OK** to create a new blank task route.

   a. Configure a collection source (a place to find files to process) by clicking **FSC Collector** in the Main Task Route window. File system task routes require a file collector. The job of a file collector is to monitor locations that you specify for files to be processed by IBM Content Collector.
   b. The **P8 4.x Create Document** task checks for duplicates and passes nonduplicates and duplicates down different paths. To add a requirement that only nonduplicate files are processed, set a decision point, and a rule in the task route to detect if a document is a nonduplicate, and if so to pass it along the route.
   c. The **FSC Post Processing** task sets options to delete archived files and replaces them with shortcuts to files in the repository.
2. Select the `FS to P8 Archiving (Replicate File System and Detect Duplicates).ctms` template or create a new task route for deduplication. With the deduplication task route, you can archive files into the `root` directory of the

FileNet P8 repository. As searching for a file in the repository can be difficult, the task route archives files in the same structure as the file system.

a. Configure a collection source (a place to find files to process) by clicking **FSC Collector** in the Main Task Route window.
b. Configure the **P8 4.x Create Document** task to specify the repository (document class) in which to save files. Replace HOST:PORT in the shortcut link with the server name and port of the Web Application that retrieves the objects.
c. The **P8 4.x File Document in Folder** task re-creates the file system hierarchy in the FileNet P8 repository. If you created a new task route:

Location:	Selection:
Add Folder window	**Regular Expression**
Edit Regular Expressions window	**Metadata type File** and **Property File Folder Path**
Define Regular Expressions section	**Define matches regular expression**
Matches regular expression text box	Enter the following regular expression: `[^\\\/]*$`.

d. Optional: Export the File in Folder task route to be used as a template.

3. Declare the archived files as records by selecting the `FS to P8 Archiving (Declare as Records).ctms` template. This task route sets up automatic records declaration for nonduplicate files that are archived to a repository. The postprocessing options for each arm of the task route remain: nonduplicates are archived and replaced with a shortcut to the item in repository; duplicates replaced with a shortcut to the original archived document in repository.

   Restriction: You must be connected to IBM InfoSphere Enterprise Records.

   a. Configure a collection source (a place to find files to process) by clicking **FSC Collector** in the Main Task Route window. File system task routes require a file collector. The job of a file collector is to monitor locations that you specify for files to be processed by IBM Content Collector.
   b. Configure the **P8 4.x Create Document** task to specify the repository (document class) in which to save files. Replace HOST:PORT in the shortcut link with the server name and port of the Web Application that retrieves the objects.
   c. The **P8 4.x Declare Record** task selects a Records Manager Class and Record Class to assign to each file being archived.

Related tasks

"Creating a task route" on page 275

"Collecting file system documents" on page 479

Related reference

"P8 4.x Create Document" on page 543

"P8 4.x File Document in Folder" on page 555

"P8 4.x Declare Record" on page 553

"FSC Post Processing" on page 534

Defining metadata to be used to process files for archiving

When you set up an archiving process, you can use CSV or XML metadata files to define how documents are added to the repository. You must create the required classes and properties in the repository before you map values to the properties from the metadata files.

To define metadata to process files for archiving:

1. Add or edit the file system metadata. Open the IBM Content Collector Configuration Manager and click the **Metadata and Lists** tab.
2. Create a collector to monitor for the metadata files:
   a. On the **Task Route** tab, click the **New** icon and select the `FS to P8 Archiving (Associate Metadata).ctms` template.
   b. Configure the **FSC Collector** task to monitor locations that you specify for content files to be archived by IBM Content Collector. File system task routes require a file collector.
3. Set up automatic archiving for files that you want associated with metadata read from XML files:
   a. Configure the **FSC Associate Metadata** task by selecting the metadata source type.
   b. Configure the **P8 4.x Create Document** task to your task route to specify the repository (document class) in which to save files. Replace HOST:PORT in the shortcut link with the server name and port of the Web Application that retrieves the objects.
   c. The **P8 4.x File Document in Folder** task re-creates the file system hierarchy in the FileNet P8 repository.
   d. The **FSC Post Processing** task set options to delete archived files and replace them with shortcuts to files in the repository. For this, **File** is selected as the metadata type and **File Name** as the property.
4. To process content files differently from metadata files, use the `FS to P8 Archiving (Delete Metadata Files).ctms` template. The task route includes a decision point that allows for conditional processing. In this case, only content files are archived and filed. The metadata file is deleted to free up file system space. The metadata file is no longer needed after IBM Content Collector has read the metadata file and used the contents to populate the metadata source.

Related tasks

"Adding and editing file system metadata" on page 241
"Creating a task route" on page 275
"Collecting file system documents" on page 479
"Adding decision points" on page 281

Related reference

"FSC Associate Metadata" on page 531
"P8 4.x Create Document" on page 543
"P8 4.x File Document in Folder" on page 555
"FSC Post Processing" on page 534

Part 6. Developing

© Copyright IBM Corp. 2008, 2010

Developing with the IBM Content Collector APIs

IBM Content Collector ships with these sets of application programming interfaces (APIs): Web Application services APIs and the Document Viewer. These APIs enable other applications to integrate with Content Collector. You can use the Web Application services APIs to trigger archiving of email and to restore or view archived email documents by using information in the stub document in the mailbox. The Document Viewer allows you to view archived documents in a customized format, without using a Content Collector client.

Developing with the Content Collector Web Application services APIs

The IBM Content Collector Web Application services application programming interfaces (APIs) enable other applications to integrate with Content Collector. You can use the APIs to trigger archiving of email and to restore or view archived email documents.

Calling Content Collector Web Application services is supported by servlets on the Content Collector Web Application Server. These servlets process all Content Collector Web Application services API requests and verify that requesters have the appropriate authority to call the service.

The Content Collector API for archiving allows you to request email archiving without using a Content Collector client.

The Content Collector APIs for restoring or viewing email conform to representational state transfer (REST) principles. URIs (uniform resource identifiers) are used to exchange information between client applications and the Content Collector Web Application Server. Because REST is based on HTTP, you can use any client or programming language that is capable of submitting an HTTP or HTTPS request. However, these Content Collector APIs require HTTPS requests and, therefore, specific security setup. Communication is secured by the Transport Layer Security (TLS) protocol. For mutual authentication, a client certificate is required. See the topic on enabling security for the Web Application services APIs.

Important: To be able to configure the required security settings for using these APIs, an external Web Application Server must be installed and deployed on the Content Collector Server machine.

The following information is provided for each Content Collector REST API:

Purpose
Brief information about using the API.

Parameters
Descriptions of the API parameters.

Responses
Descriptions of the API responses. Responses include status codes and, if an error code is returned, an error message is returned too.

© Copyright IBM Corp. 2008, 2010

Related tasks

"Enabling security for the Web Application services APIs" on page 742

Creating requests for interactive archiving

To be able to trigger archiving from a client application, you have to set up a request for interactive archiving. Such a request includes marking one or more email documents for archiving and sending a trigger mail to the job mailbox that is monitored for interactive archiving requests. The documents that were marked for archiving are then collected and processed according to the settings in the respective collector.

The documents that are to be archived are identified by the client application. To find out which email documents were not yet archived, check the document state of the email documents in the mailbox.

To set up the request for interactive archiving:

1. Mark one or more email documents for archiving. A document is marked for archiving by setting a specific property in the document. IBM Content Collector uses this property to identify documents that are to be archived.

   For Lotus Domino, you can optionally set the $ContentIcon property, so that the status of documents is immediately reflected in the client application.

   You can set the following properties:

   - Message properties for Microsoft Exchange:

     PR.AFU.MESSAGE.STATE
     Sets the document state to MARKED_FOR_ARCHIVING.

     Type PT_STRING8

     Value MA

   - Notes items for Lotus Domino:

     IBMAfuMessageState
     Sets the document state to MARKED_FOR_ARCHIVING.

     Type Text

     Value MA

     $ContentIcon
     Reflects the document status in the client application.

     Type Number or Text

     Value The internal number or the name of the icon that is used to reflect the document state in the client application, for example, `marked-for-archiving.gif`.

2. Send a trigger mail. Marking documents for archiving does not cause Content Collector to collect documents for archiving. You have to trigger Content Collector explicitly to check a specific mailbox, to collect documents that were marked for archiving, and to archive them.

   To do so, send a so-called trigger mail to the job mailbox that is defined as the collection source for the collector for interactive archiving. The trigger mail contains all information that Content Collector requires to be able to find and archive the documents in questions. So, instead of checking each user mailbox for new archiving requests, the collector checks only this job mailbox (trigger mailbox).

The trigger mail defines a number of properties, some of which are mandatory and some of which optional. The optional properties can be used to make the trigger mail more usable, for example, for an administrator who is monitoring the trigger mailbox. You can set the following properties in the trigger mail:

- Message properties for Microsoft Exchange trigger mails

  PR_MESSAGE_CLASS
  Mandatory property.

  Type PT_STRING8

  Value IPM.Note.AFU.Trigger.Archive

  PR.AFU.MESSAGE.TARGET.MAILBOX.DN
  Mandatory property.

  Type PT_STRING8

  Value <*mailbox DN*>, which is the distinguished name of the mailbox containing documents that were marked for archiving. For example:

  ```
  /o=EX/ou=First Administrative Group/cn=Recipients/cn=user
  ```

  PR.AFU.MESSAGE.TARGET.SERVER.DN
  Mandatory property.

  Type PT_STRING8

  Value <*server DN*> , which is the distinguished name of the server where the mailbox is located. For example:

  ```
  /o=EX/ou=First Administrative Group/cn=Configuration/cn=Servers/cn=server
  ```

  PR.AFU.MESSAGE.TRIGGER.EXEC.TIME
  Mandatory property.

  Type PT_SYSTIME

  Value <*timestamp*>, which is the absolute time when the collector should consider this trigger mail. When users are working with their Microsoft Outlook client in cached mode, there will be a delay between when a email is marked for archiving and when this modification is synchronized to the Exchange server, while the trigger mail might be delivered immediately. To prevent Content Collector from checking the mailbox on the server too early when users are working in cached mode, specify a timestamp that defines the time when the trigger mail was created plus the synchronization interval, for example, current time plus 5 minutes.

  PR.AFU.MESSAGE.TRIGGER.VERSION
  Mandatory property.

  Type PT_I2

  Value 2

  PR_SUBJECT
  Optional property.

  Type PT_STRING8

  Value <*subject text*>, which is a value for the mail subject to enhance readability of a trigger mail. For example: `Interactive archiving request for mailbox <user> on server <server>`

- Notes items for Lotus Domino trigger mails

  Form Mandatory property.

  Type Text

  Value IBMAfuArchiveTrigger

  IBMAfuRequestMailboxPath
  Mandatory property.

  Type Text

  Value <*DBPath*>, which is the relative path to the mailbox on the Lotus Domino server, for example,`mail\user42.nsf`

  IBMAfuRequestMailboxServer
  Mandatory property.

  Type Text

  Value <*DominoServer*>, which is the name of the server where the mailbox is located, specified in abbreviated or canonical format, for example,`mySrv1/mail/foo`

  IBMAfuTriggerExecutionDate
  Mandatory property.

  Type Date/Time

  Value <*timestamp*>, which is the absolute time when the collector should consider this trigger mail. When users are working from a local replica of their mailbox, there will be a delay between when email is marked for archiving and when this modification is replicated back to the Domino server, while the trigger mail might be delivered immediately. To prevent Content Collector from checking the mailbox on the server too early when users are working from a local replica, specify a timestamp that defines the time when the trigger mail was created plus the replication interval, for example, current time plus 5 minutes.

  Note: You can set the timestamp to the current time but only if the request is triggered on the server replica.

  IBMAfuTriggerVersion
  Mandatory property.

  Type Number

  Value 2

  Subject
  Optional property.

  Type Text

  Value <*subject text*>, which is a value for the mail subject to enhance readability of a trigger mail. For example: `Interactive archiving request for mailbox <DBPath> on server <DominoServer>`

Document states

The document state is a value that indicates the processing state of an email document.

The document state is reflected by the value of this property:

- PR.AFU.MESSAGE.STATE for Microsoft Exchange documents
- IBMAfuMessageState for Lotus Domino documents

Table 188. Processing states of email documents

Name of the processing state	Property value	Description
MARKED_FOR_ARCHIVING	MA	The document is marked for archiving.
MARKED_FOR_ARCHIVING_METADATA_PROVIDED	MAMP	The document is marked for archiving. Additional archiving information for the document was provided.
ARCHIVED	A	The document is archived.
MARKED_FOR_STUBBING	MS	The document is archived and marked for stubbing.
STUB_NOTHING_ADD_TEXT	STXT	The document is archived. Nothing is removed from the original documents, but the specified text message was added to the original document.
STUB_ATTACHMENTS	SATT	The document and any attachments are archived. The attachments were removed from the original document. The stub document stub contains links to archived attachments and, if applicable, the specified link text.
STUB_ABBREV	SABB	The document and any attachments are archived. In addition to removing the attachments from the original document, the length of the body text is reduced. The stub document stub contains links to the archived content and the archived attachments and, if applicable, the specified link text.
STUB_BODY	SDEL	The document and any attachments are archived. Attachments and the body text are removed from the original document. The stub document stub contains links to the archived content and the archived attachments and, if applicable, the specified link text.
RESTORED	R	An archived document was restored from the mailbox.
SEARCH_RESTORED	RSEARCH	An archived document was restored from the search result list.
MOBILITY_DONE	L	The document is archived and available in the offline repository.

RestoreAPI

Use this API to call the Web Application service for restoring email documents from a client application.

Purpose

Use a client application to request restoring archived email documents. Provide the following information with the call:

- The mail provider
- The name of the email server that hosts the mailbox
- The name of the mailbox
- The number of email documents that are to be restored with this request
- A list of unique identifiers for the email documents that are to be restored. The list is processed in batch mode.

The Web Application for restoring email documents uses HTTPS connections for communication. Access to the API is restricted. Therefore, it requires specific security setup. See the topic on enabling security for the Web Application services APIs.

Submit the API call as a POST request. The URL for the request must be set up as follows:

`https://<server_name>:<port_number>/AFUWeb/RestoreAPI`

where <*server_name*> is the name of the computer that hosts the Content Collector Web applications and <*port_number*> is the port number that is used for HTTPS communications. For Content Collector, the default port number is 11443.

Tip: To omit problems with URL length limits, send the request parameters in the request body. The request is performed synchronously. Therefore, a response is not sent before all restore request that are contained in the HTTPS request have been processed.

If the request includes email documents that were archived with IBM CommonStore, a job for processing those email documents is created. The restore will then be performed asynchronously by the CommonStore for Exchange Server or CommonStore for Lotus Domino tasks.

Parameters

These parameters are all required but can be specified in any sequence. Note that the parameters and their values are case sensitive.

mailServer
ID of the mail server

Microsoft Exchange
The distinguished name of the mail server where the mailbox is located into which the email documents are to be restored. The ID of the mail server is the value that is assigned to msExchHomeServerName in the Active Directory for the user.

Example:

```
/o=EX2003/ou=First Adminstrative Group/cn=Configuration/cn=Servers/cn=myServer
```

Lotus Domino

Example:

```
dominoServer/dominoDomain
```

mailboxName
The distinguished name of the mailbox into which the email documents are to be restored.

Microsoft Exchange
The name of the mailbox is the value that is assigned to legacyExchangeDN in the Active Directory for the user.

Example:

```
/O=EX2003/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=username
```

Lotus Domino

Example:

```
mail\\user01.nsf
```

mailProvider
A string identifying the email provider. Select one of the following values:

Microsoft Exchange
`EXCHANGE`

Lotus Domino
`DOMINO`

paramCount
An integer value that defines the number of email documents that are to be restored with this request.

Note: If the value of **parmCount** is less than the actual number of email IDs listed in the request, only as many email documents as specified with **parmCount** are restored. If the value of **parmCount** exceeds the actual number of email IDs listed in the request, the request is not processed and an error status code is returned.

d*n* This parameter name serves as separator for the email IDs. Each parameter name consists of the character d and the counter *n*, which is an integer value starting with 0 and is incremented by one. The counter is not used for sequencing or any other purpose than providing unique parameter names.

The parameter value is the identifier of the email that is to be restored. This identifier is specific to the email provider.

Microsoft Exchange
The entry ID of the email, which is the value of the PR_ENTRYID property in hexadecimal representation.

Lotus Domino
The unique ID of the email, which is the value of the UNID property in hexadecimal representation.

Responses

The following status codes, return codes, or error message can be returned in the response:

SC_OK
All email documents were successfully restored. This status code is accompanied by a success message.

SC_ACCEPTED
Not all email documents could be restored. This status code is accompanied by an error message.

SC_BAD_REQUEST
The request could not be processed because input parameters were missing or contained invalid values. This status code is accompanied by an error message.

errNum=*n***&msg=***message text*
Success or error message, where *n* can be -1, 0, or any positive integer value.

- A value of -1 indicates that an error prevented processing of any email:
  ```
  errNum=-1&msg=An error occurred. Contact your system administrator.
  ```
- A value of 0 indicates that all email documents were successfully restored:
  ```
  errNum=0&msg=All items were successfully processed.
  ```
- Any other value indicates the number of email documents that could not be restored because an error prevented their processing.

Related tasks

"Enabling security for the Web Application services APIs" on page 742

ViewingAPI

Use this API to call the Web Application service for viewing archived email documents from a client application. The client application accesses archived email by using information in the stub document in the mailbox.

Purpose

Use a client application to request viewing archived email documents. Based on properties that Content Collector sets when archiving documents, the API identifies the email in the archive, retrieves it, and displays the preview page for the email.

Provide the following information with the call:

- The mail provider
- The name of the email server that hosts the mailbox
- The name of the mailbox
- A unique identifier for the email documents that you want to view

The Web Application for viewing archived email documents uses HTTPS connections for communication. Access to the API is restricted. Therefore, it requires specific security setup. See the topic on enabling security for the Web Application services APIs.

Submit the API call as a POST request. The URL for the request must be set up as follows:

```
https://<server_name>:<port_number>/AFUWeb/ViewingAPI
```

where <*server_name*> is the name of the computer that hosts the Content Collector Web applications and <*port_number*> is the port number that is used for HTTPS communications. For Content Collector, the default port number is 11443.

Tip: To omit problems with URL length limits, send the request parameters in the request body. The request is performed synchronously. Therefore, a response is not sent before the viewing request that is contained in the HTTPS request was processed.

Restriction: For documents that were archived with IBM CommonStore, viewing is supported for those documents only that were archived in a CommonStore item type for the document model BUNDLED and the archiving type ENTIRE.

Parameters

These parameters are all required but can be specified in any sequence. Note that the parameters and their values are case sensitive.

mailServer
ID of the mail server

Microsoft Exchange
The distinguished name of the mail server where the mailbox is located that hosts the email document to be viewed. The ID of the mail server is the value that is assigned to msExchHomeServerName in the Active Directory for the user.

Example:

```
/o=EX2003/ou=First Adminstrative Group/cn=Configuration/cn=Servers/cn=myServer
```

Lotus Domino

Example:

```
dominoServer/dominoDomain
```

mailboxName
The name of the mailbox that hosts the email document to be viewed.

Microsoft Exchange
The distinguished name of the mailbox is the value that is assigned to legacyExchangeDN in the Active Directory for the user.

Example:

```
/O=EX2003/OU=FIRST ADMINISTRATIVE GROUP/CN=RECIPIENTS/CN=username
```

Lotus Domino

Example:

```
mail\\user01.nsf
```

mailProvider
A string identifying the email provider. Select one of the following values:

Microsoft Exchange
`EXCHANGE`

Lotus Domino
`DOMINO`

d The identifier of the email that is to be viewed. This identifier is specific to the email provider:

Microsoft Exchange
The entry ID of the email, which is the value of the PR_ENTRYID property in hexadecimal representation.

Lotus Domino
The unique ID of the email, which is the value of the UNID property in hexadecimal representation.

Responses

If an error occurs, an error message is displayed in a separate browser window.

Related tasks

"Enabling security for the Web Application services APIs"

Enabling security for the Web Application services APIs

Secure Web Application services to provide protection for messages exchanged in a web service environment. IBM Content Collector restricts the access to the Web Application services APIs by using a specific security role *iccUser_Role*. To enable security checks, you must configure the Web Application Server in a way that application security is enabled and you must map the security role to one or more users that are defined in the LDAP that will be used for authentication.

Complete the following steps to enable security for API calls:

1. Configure WebSphere Application Server security. The following instructions apply to WebSphere Application Server Version 6.1. For WebSphere Application Server Version 7, the procedures might be different (see the WebSphere Application Server Version 7.0 Information Center).
   a. Start the WebSphere Application Server administrative console. In a browser, enter this web address:

      ```
      http://<serverName>:11060/ibm/console/login.do log-in
      ```

   b. Enable administrative security
   c. Enable application security
2. Provide certificates for API calls

Enabling WebSphere Application Server administrative security

Administrative security requires users to authenticate before obtaining administrative control of the application server. In this case, access control is performed when the resource is requested by a web client and it is determined whether the authenticated user has the required security role.

Complete the following steps to enable WebSphere Application Server administrative security:

1. In the WebSphere Application Server administrative console, select **Security** → **Secure administration, applications, and infrastructure**.
2. Enable administrative security. Application security can be in effect only when administrative security is enabled.
3. Enable application security.
4. Disable Java 2 security by clearing the check box.
5. Under **Authentication**, select **Use domain-qualified user names** if the user is a domain user.
6. Under **User accounts repository**, select **Standalone LDAP registry** and click **Set as current**.
7. Click **Configure** to set up that LDAP registry. For Microsoft Exchange systems, that LDAP registry can be the Microsoft Active Directory managing users in the domain. For Lotus Domino systems, that LDAP registry can be the IBM Lotus Domino LDAP managing users.
   a. Enter the credentials for the user who is to log on to the administrative console as well as the host name and the port to use when connecting to LDAP. This must be a user with the appropriate permissions to start the IBM Content Collector Web Application service, for example, the Microsoft Exchange or Lotus Domino administrator. Specify the following information:

Host Is the LDAP server

Base distinguished name (DN)
Consists of the domain components of the user DN. For example: dc=ibm, dc=com

Bind distinguished name (DN)
Is the full distinguished name of the user. For example: cn=*adminUsername*, cn=users, dc=ibm, dc=com

Bind password
Is the user's password in the LDAP.

b. Test the connection to ensure that the values related to connecting to LDAP are valid.

c. Click **OK** to save the changes.

d. Save your changes to the master configuration to change the settings on disk.

e. Verify the settings for mapping certificates:

1) Under **Additional Properties**, click **Advanced Lightweight Directory Access Protocol (LDAP) user registry settings**.

2) Verify that the map mode for the certificate is EXACT_DN. You can also define a certificate filter for mapping the attributes in the client certificate to entries in the LDAP registry

3) Click **Apply**.

4) Save the changes to the master configuration.

8. Define the administrator user by selecting **Users and Groups → Administrative User Roles**. Make sure that the administrator role is selected for the user for whom you just defined LDAP access.

9. Save the changes to the master configuration.

10. Enable authentication via client certificate.

a. Click **Security → SSL certificate and key management**.

b. Under **Related Items**, click **SSL configurations**. A list of SSL configurations is displayed.

c. Click **NodeDefaultSSLSettings** in the list of SSL configurations. The configuration tab for this SSL configuration is displayed.

d. Under **Additional Properties**, click **Quality of Protection (QoP) settings**.

e. From the list under **Client authentication**, select **Supported**.

f. Click **OK** and save the changes to the master configuration.

11. Import the root CA for the client certificate to the default node truststore.

a. Click **Security → SSL certificate and key management**.

b. Under **Related Items**, click **Key stores and certificates**. A list of key stores is displayed.

c. Click **NodeDefaultTrustStore** in the list of key stores. The configuration tab for this key store is displayed.

d. Under **Additional Properties**, click **Signer certificates**.

e. Add the root CA certificate that was used when signing the client certificate.

Now, the security settings for the administrative console are complete.

12. Restart the Web Application Server for these settings to become active. Make sure that the credentials of an administrator user (as defined in step 8) are

used. If you want to start the server from the command line, navigate to the <*WAS_HOME*>\bin directory and enter the following command:

```
startserver afuServer -profileName AFUWeb -username <user> -password <password>
```

Now, WebSphere Application Server is configured to connect to LDAP for authentication, and there is an administrator user who can log on to the administrative console.

Note: If problems occur when you start the server or when the newly configured administrator user information is used for logon, you can disable administrator security as follows:

1. Launch the wsadmin tool by entering the following command:

   ```
   wsadmin -conntype NONE -profileName AFUWeb
   ```

2. From the wsadmin prompt, enter `securityoff` to disable administrative security.
3. Enter `quit` or `exit` to leave the wsadmin session.
4. Restart the Web Application Server.

   All configuration settings except the selections on the **Secure administration, applications, and infrastructure** page do still exist.
5. Correct the configuration of the LDAP connection and re-enable administrative security as described in steps 1 on page 742 and 2 on page 742.
6. Restart the Web Application Server.

The next step is to enable WebSphere Application Server application security.

Enabling WebSphere Application Server application security

When application security is enabled, access to applications in your environment is restricted. When accessing a protected resource, a web client is prompted for authentication. IBM Content Collector restricts the access to the Web Application services APIs by using the security role *iccUser_Role*. You must map the security role to one or more users or groups that are defined in the LDAP that will be used for authentication.

Make sure that the Content Collector Web Application is installed and deployed. You must re-enable application security each time you deploy the Content Collector Web Application.

Complete the following steps to enable application security:

1. In the WebSphere Application Server administrative console, select **Applications → Enterprise Application**.
2. In the list of applications, click **afu_web** to view the settings for this application.
3. Under **Detailed Properties**, click **Security role to user/group mapping**. A list of the roles that are defined in the application is displayed.
4. Select **iccUser_role** and click either **Look up users** or **Look up groups**.
5. Search for one or more users or groups that you want to grant access to the API servlet.
6. Add the users or groups to the list of selected entries. Should the search fail with an error message, check the LDAP settings you defined previously. Should the user or group that you want to add not show up in the list, check and change the search string or the limit shown in the panel, or both.
7. If the selected list is populated as required, click **OK**.

8. Verify that your selection is displayed in the user mappings.
9. Verify that the general settings for web security are set the way you need them to be.
   a. Select **Security** → **Secure administration, applications, and infrastructure**.
   b. Under **Authentication**, expand **Web security** and click **General settings**. The web authentication settings that are associated with a web client are displayed.
   c. Set the scope for the security settings.
      - To enforce usage of client certificates only for the API method while all other clients remain able to call services of the Web Application, such as search, without sending a client certificate, select **Authenticate only when the URI is protected**.
      - To have WebSphere Application Server show a login panel if no matching client certificate is sent by the client request and use credentials given there for authentication to LDAP, select **Default to basic authentication when certificate authentication for the HTTPS client fails**. You might want to provide this authentication method for users calling the protected web address from a system when the required client certificate is not installed. With this setting, however, a logon window will be shown for each call that requires a client certificate.
   d. Click **OK** and save your changes to the master configuration.
10. Restart the Web Application Server for these changes to become active.

Providing certificates for Web Application services API calls

Mutual authentication is required for calls to the Web Application services APIs. Therefore, the caller must hold a client certificate to be able to establish a TLS connection with the server and to call the APIs. You can create a client certificate yourself or you can request one from an external certificate authority (CA) provider.

For an example of how to obtain a client certificate, see the sample procedure for creating a client certificate for the Web Application services API. Depending on the way in which an API request is submitted, the client certificate must either be installed on the machine that hosts the web client or the client certificate must be provided with the request.

- Provide certificates for web clients. Import the client certificate.

  Internet Explorer 7

  1. Select **Tools** → **Internet Options**.
  2. On the **Content** tab, click **Certificates**.
  3. On the **Personal** tab, click **Import**.
  4. Select the file that holds the client-certificate keystore. Make sure that the store contains the certificate along with its key and import it.
  5. On the **Trusted Root Certification Authorities** tab, there must be an entry for the CA that created the client certificate. The trusted entry for that authority will be in the keystore file if you exported the certificate chain and should be imported automatically. If not, import it manually.

  Mozilla Firefox 3.5

  1. Select **Tools** → **Options** → **Advanced**.
  2. On the **Encryption** tab, click **View Certificates**.

3. On the **Your certificates** tab, click **Import** to import the certificate from the keystore file on your disk.
4. On the **Authorities** tab, there must be an entry for the CA that issued the certificate you just imported. If it is missing, import that one as well.

- Provide certificates for Java clients.
  1. Copy the file containing the client certificate to the machine that hosts the client calling the Web Application services API. The file must contain the complete certificate chain, including the trusted entry for the CA that issued the certificate.
  2. Add the server certificate, which the Web Application Server sends to its clients for identification, as a trusted entry to the keystore. You can use any browser that already has imported that server certificate to export the certificate to disk.
  3. Use the ikeyman tool to import the keystore holding the server certificate to the keystore that holds the client certificate. As a result, the keystore will contain trusted entries for the machine hosting the Web Application Server and the CA that issued the client certificate as well as an entry for the client certificate itself.

  When you develop the Java client for the Web Application services API call, use that keystore when you create the KeyManager object that will be used in your SSLContext class.

Related tasks

"Sample procedure for creating a client certificate for the Web Application services API"

Sample procedure for creating a client certificate for the Web Application services API:

A client certificate is required for access to the IBM Content Collector Web Application services APIs.

If you use a Windows Domain CA and web-based access to Certificate Services is enabled, you can submit a certificate request by following these steps. If you use a different CA to certify the certificate request, follow the procedure that applies for the respective CA.

1. Log on to the Windows domain as the user for whom the certificate is to be created.
2. Access the Certificates Services by specifying the following URL in your web browser:

   ```
   http://<ca_iis_server>/certsrv
   ```

   where *<ca_iis_server>* is the DNS or NetBIOS name of the CA host server.
3. Click **Request a certificate**.
4. Select the type of certificate. The subject name in the certificate template that is used for the request must be built from the DNS name information in the LDAP that you use for authentication:
   - Click **User Certificate** if the default certificate template defines the subject name as described. If required, enter identifying information for the certificate request. Click **Submit**. Install this certificate.

- Click **advanced certificate request** to select an appropriate certificate template. Click **Create** to create a certificate request to this CA. Enter all required identifying information.

5. Optional: Enter identifying information for the certificate request.
6. Submit the request.
7. Install the certificate.
8. Export the new certificate. Note that these descriptions apply to Windows installations.
   - If you want to send the certificate to clients so that the clients can import the certificate into their browser, export the certificate including the key.
     - If you want to set this up for many users, you can use automatic enrollment. For details, see the Microsoft documentation on auto-enrollment of certificates.
     - If you want to manually propagate a client certificate to one or more client machines, create an advanced certificate request. Make sure that **Mark keys as exportable** is selected. Otherwise, the exported certificate cannot be used for logon. If that option is not available, check the definition of the selected certificate template: Start the Microsoft console `certtmpl.msc` and check the template definition. Then, copy the certificate to the machine where you need it and import it to your browser.
   - If you code a Java client, export the certificate to a keystore.

Related tasks

"Providing certificates for Web Application services API calls" on page 745

Developing with the Document Viewer

The Document Viewer allows you to view email documents, email attachments, or files that were archived to an IBM Content Manager repository or a IBM FileNet P8 repository in a customized format when you work with IBM FileNet Workplace or IBM FileNet Workplace XT.

The Document Viewer API conforms to representational state transfer (REST) principles. URIs (uniform resource identifiers) are used to exchange information between client applications and the Document Viewer. Because REST is based on HTTP, you can use any client or programming language that is capable of submitting an HTTP or HTTPS request. However, Content Collector APIs require HTTPS requests and must, therefore, contain a user name and password as request parameters.

Calling Document Viewer services is supported by servlets on the Content Collector Web Application Server. The Document Viewer is deployed as a web application either during the installation of Content Collector when you select to work with the embedded Web Application Server, or when you configure an external Web Application Server for use with Content Collector.

The information that the Document Viewer requires as input consists of the configuration settings in the `docviewer.config` file and the request parameters for identifying the document to be viewed and for authentication.

The Document Viewer configuration files

The `docviewer.config` file contains configuration settings for the Document Viewer. The `ral.properties` file contains information about the repository connection. Adapt these settings according to your needs.

The docviewer.config file and the ral.properties files are in the AFUWeb\DocViewer\config subdirectory of one of these directories:

- The installation directory of IBM Content Collector Server if you work with the embedded Web Application Server.
- The home directory of WebSphere Application Server if you use an external Web Application Server. The home directory of WebSphere Application Server is the one that the WASHOME environment variable points to. Usually this is <WASinstall>\AppServer, where <WASinstall> is the path to your WebSphere Application Server installation directory.

Configuration settings

Adapt the configuration settings for the Document Viewer in the docviewer.config file.

Lotus Domino Server configuration
Configure the connection to the Lotus Domino Server. These default configuration settings are used when Lotus Notes email is viewed in its native format.

DOMINO_ENABLED
To enable Lotus Notes viewing, set this parameter to true. The default value is false.

DOMINO_SERVER
Specify the host name of the Lotus Domino server on which the viewing database is created.

DOMINO_PORT
Specify the HTTP port number of the Lotus Domino server.

DOMINO_MAILTEMPLATE
Specify the mail database template for creating the viewing database, for example, mail8.ntf or myTemplates/mail8.ntf.

Default conversion format for email
Define the default conversion format with the **defaultConversionFormat** parameter. Select one of these formats:

0 HTML format

1 PDF format

2 TIFF format

Default conversion format for email attachments
Email attachments are files in the native format. Usually, you can use the native application to view them. If this is not possible, the default conversion format that you specified with the **defaultAttachConversionFormat** parameter.

Stylesheet specification
Define the Extensible Stylesheet Language Transformation (XSLT) files to use for converting documents to the selected format. The available XSLT files are in the same directory as the docviewer.config file.

Lotus Notes
For each Notes message type, you can specify a different XSLT file to use for the conversion. Use the notation

`FORMTYPE.msg_type.format stylesheet`, where *msg_type* is the Notes message type, *format* is the viewing format, and *stylesheet* is the XSLT file to use, for example:

```
FORMTYPE.Memo.0 EmailTemplate.xsl
FORMTYPE.Memo.1 PDFIMAGETemplate.xsl
```

Microsoft Outlook

For each Outlook message class, you can specify a different XSLT file to use for the conversion. Use the notation `MESSAGECLASS.msg_type.format stylesheet`, where *msg_type* is the Outlook message class, *format* is the viewing format, and *stylesheet* is the XSLT file to use, for example:

```
MESSAGECLASS.IPM.0 EmailTemplate.xsl
MESSAGECLASS.IPM.1 PDFIMAGETemplate.xsl
```

Default settings

These are the default XSLT mappings. These are used if you do not define any specific mappings for Lotus Notes or Microsoft Outlook, or if one of the specified mappings cannot be applied.

```
XSLT.DEFAULT.0 EmailTemplate.xsl
XSLT.DEFAULT.1  PDFIMAGETemplate.xsl
XSLT.DEFAULT.2  PDFIMAGETemplate.xsl
```

MIME type mapping

With the MIME type mapping, you define whether documents of a specific MIME type are treated like email documents, for example, `application/x-filenet-filetype-msg true`.

Message translation

Specify the properties files that contain the information for language-specific messages.

Specify a file name for **email.nls.keys.file**, for example:

```
email.nls.keys.file emailnlskeys.properties
```

This file lists the set of language-specific variables that are defined as key-value pairs in the message properties file.

Specify the name of the message properties file that contains the key-value pairs for the locale support, for example:

```
messages.fileprefix messages
messages.filesuffix properties
```

The resulting file name is `messages.properties`. For a language-specific file, also specify the locale to use:

```
messages.fileprefix messages
userlocalcountry de
messages.filesuffix properties
```

The resulting file name is `messages_de.properties`. This file must be in the same directory as the `docviewer.config` file.

Logging

Configure logging to facilitate troubleshooting:

console.logLevel

Specify the logging level. Select one of these values:

Logging level	Content
SEVERE	Serious failure

Logging level	Content
WARNING	Potential problem
INFO	Informational messages
CONFIG	Configuration messages
FINE	General trace information
FINER	Detailed trace information
FINEST	Highly detailed trace information

The default value is SEVERE.

console.logFile
Specify the full path and a file name for the Document Viewer log file.

console.numLogFiles
Specify the maximum number of log files to be created. When you specify a value, also consider the size limit of each log file, the amount of disk space that your system has, and the amount of history that you want to keep.

The default value is 2.

console.logFileSize
Specify the maximum size that a log file is allowed to reach. As soon as the first log file reaches this size level, a new log file is created. When the maximum number of log files has been reached, and all log files have also reached their maximum size, the oldest log file is overwritten with a new one. This is also known as the round-robin method.

When you specify a value, also consider the number of log files, the amount of disk space that your system has, and the amount of history that you want to keep.The default value is 10.

Repository connection information

Adapt the repository connection information for the Document Viewer in the `ral.properties` file. Uncomment the parameters that are applicable for your repository system and specify the appropriate values.

IBM Content Manager repository connection

```
#Content Manager connection information
REPOSITORY_TYPE=ICM

cmDatabase=database_name
cmSchema=schema_name
adminUser=administrator_ID
adminPassword=administrator_password
```

database_name
The Library server database name for the document repository.

schema_name
The Library server schema name.

administrator_ID **and** *administrator_password*
The administrator user name and administrator password that you defined when you configured IBM Content Manager for Content Collector.

IBM FileNet P8 repository connection

```
#P8 connection information
REPOSITORY_TYPE=P8

adminUser=administrator_ID
adminPassword=administrator_password
p8Domain=P8_domain_name
p8Uri=CE_URI
p8ObjectStore=object_store
p8Host=host_name
p8Protocol=protocol
```

administrator_ID **and** *administrator_password*
: The administrator user name and administrator password hat you defined when you configured FileNet P8 for Content Collector.

P8_domain_name
: The name of the FileNet P8 to which the object store belongs.

CE_URI
: The Content Engine URI that is configured in Process Task Manager, for example, `http://CEServer:9080/wsi/FNCEWS40MTOM/`, where *CEServer* is the host name of the Content Engine server and `FNCEWS40MTOM` is the attachment protocol.

object_store
: The name of the object store.

host_name
: The host name of the Content Engine server.

protocol
: The protocol that the Document Viewer uses to access Content Engine, for example, FileNetP8WSI.

Document Viewer requests

Create a request for viewing archived email documents, email attachments, or files in a customized format.

Purpose

Use a client application to call the Document Viewer for viewing archived files, email documents, or email attachments in a customized format. Determine how the content of archived documents is presented when users view them in a client or when the Document Viewer is called from another application.

You can submit the API call as either a GET or a POST request. The URL for the request must be set up as follows:

- To view files or email documents in a customized format:
  `https://<server_name>:<port_number>/DocViewer/ViewItem.do?<request_parameters>`
- To view attachments of email documents in a customized format:
  `https://<server_name>:<port_number>/DocViewer/ViewAttachment.do?<request_parameters>`
- To view documents in their native format:
  `https://<server_name>:<port_number>/DocViewer/NativeViewItem.do?<request_parameters>`

Where <*server_name*> is the name of the computer that hosts the Content Collector Web applications and <*port_number*> is the port number that is used for HTTPS communications. For Content Collector, the default port number is 11443.

For a GET request, include the <*request_parameters*> key-value pairs in the URL of the request. For a POST request, include the <*request_parameters*> in the request body.

Parameters

With the <*request_parameters*> you identify the repository and the document or attachment to be viewed, provide credentials for authentication with the repository, and determine the format and the highlighting of the document that is viewed. These parameters can be specified in any sequence. Note that the parameters and their values are case sensitive.

Document identification

This request parameter is required. You can specify document and repository information in either of the following formats. When you work with multiple IBM eDiscovery Manager repositories where all repositories are registered with the primary repository, the Document Viewer uses the repository information that is provided by eDiscovery Manager. Otherwise, the repository information that is contained in the `ral.properties` file is used.

reposID=*PID*
Specify the persistent identifier (PID) of the item in the repository. In this case, the default repository is used.

reposID=*repositoryID*,*PID*
Specify the repository ID and the PID of the item in the repository.

reposID=*PID*&**ls**=*ls_name*
Specify the PID of the item in the repository and the name of the IBM Content Manager library server.

When the Document Viewer connects to an eDiscovery Manager repository with registered servers, the repository name or repository ID that is provided by eDiscovery Manager is used. Otherwise, the information listed in the `ral.properties` file is used.

p8ObjectStore=*store_name*&**reposID**=*PID*
Specify the name of an IBM FileNet P8 object store other than the one provided by eDiscovery Manager or listed in the `ral.properties` file, and the PID of the item in the repository. In this case, the default FileNet P8 domain of the primary repository is used.

p8Domain=*domain_name*&**p8ObjectStore**=*store_name*&**reposID**=*PID*
Specify the names of the FileNet P8 domain and of an object store other than the one provided by eDiscovery Manager or listed in the `ral.properties` file, and the PID of the item in the repository.

Attachment identification

For viewing attachments of email documents, you must specify one of these request parameters in addition to the document identification.

correlationKey=*key*
Specify the correlation key of the attachment.

attachId=*attachment_ID*
Specify the attachment ID. This ID is an integer that indicates the offset in the attachment list for email documents where the data model type is not compound.

Authentication

You can specify these request parameters in addition to the document identification.

userid=*userID*
Specify the user ID with which you want to access the repository.

pwd=*password*
Specify the password for the user ID.

Viewing format

You can specify this request parameter in addition to the document identification.

format=*value*
Select one of these formats:

0 To view the document in HTML format. This is the default unless you changed the setting of the **defaultConversionFormat** parameter in the `docviewer.config` file.

1 To view the document in PDF format.

2 To view the document as an image file in TIFF format.

Highlighting

You can specify these request parameters in addition to the document identification.

htcount=*n*
Specify the maximum number of terms that are to be highlighted when a document is viewed.

htkeyN=*term1*,*term2*
Specify one or more terms to be highlighted.

Responses

By default, the Document Viewer responds to requests in XML format. Responses include either a success code of 0 (zero) or an error code. If an error code is returned, an error message is returned too. The error message is returned in the locale that was specified in the request.

Example:

```
<res>
   <stat>1</stat>
   <errCode>error_code</errorCode>
  -
  <err>
     error_message
   </err>
  <errorClass>error_class</errorClass>
</res>
```

Where the value of the <stat> element is always 1 for error messages.

Configuring Workplace or Workplace XT for the use of the Document Viewer

Configure IBM FileNet Workplace or IBM FileNet Workplace XT in a way that users can view archived documents with the Document Viewer.

The following prerequisites apply:

- IBM FileNet Workplace or IBM FileNet Workplace XT must be installed and set up properly.
- The IBM Content Collector Web Application must be running.
- Archived-data access must be configured for each repository in which you want to access archived documents.

To be able to view archived documents in Workplace or Workplace XT by using the Document Viewer, you must configure Workplace or Workplace XT accordingly.

1. Create a Java Server Pages (JSP) page with a name that is meaningful to you, for example, `DocViewerAPIRedirect.jsp`. The JSP page must have the following contents:

   ```
   <%response.sendRedirect("https://<<ICCServerName>>:11443/DocViewer/ViewItem.do?
   p8ObjectStore="+request.getParameter("objectStoreName")+"&reposID="+request.getParameter("id"));%>
   ```

   Where *<ICCServerName>* is the name of the Content Collector Web Application Server.
2. Save the file to the `redirect` directory. This directory is located in one of these directories:
   - The *<WASinstall>*`\profiles\<profile name>\installedApps\<cell name>\WorkplaceXT.ear\web_client.war` directory
   - The *<FileNetInstallDir>*`\FileNet\Config\WebClient` directory
   - The *<FileNetInstallDir>*`\FileNet\Config\AE` directory

   Where *<WASinstall>* is the path to the installation directory of WebSphere Application Server, *<profile name>* is the name of the web application profile, and *<cell name>* is the respective cell name. *<FileNetInstallDir>* is the installation directory of FileNet P8.
3. Edit the `content_redir.properties` file. This file is located in one of these directories:
   - The *<FileNetInstallDir>*`\FileNet\Config\WebClient` directory
   - The *<FileNetInstallDir>*`\FileNet\Config\AE` directory

   Where *<FileNetInstallDir>* is the installation directory of FileNet P8.
4. Map all MIME types that you want to view with the Document Viewer as in this example:

   ```
   image/gif=/redirect/DocViewerAPIRedirect.jsp?
   vsId={VERSION_SERIES_ID}&objectStoreName={OBJECT_STORE_NAME}&id={OBJECT_ID}
   ```

   Where *DocViewerRedirect.jsp* is the name of the file that you created in step 1.
5. Save your changes.
6. Adapt how the retrieved content is presented by modifying the viewer configuration. To do so, edit the `docviewer.config` file that is in the `AFUWeb\DocViewer\config` subdirectory of one of these directories:
   - The installation directory of IBM Content Collector Server if you work with the embedded Web Application Server.
   - The home directory of WebSphere Application Server if you use an external Web Application Server. The home directory of WebSphere Application Server is the one that the WASHOME environment variable points to. Usually this is `<WASinstall>\AppServer`, where `<WASinstall>` is the path to your WebSphere Application Server installation directory.

7. Restart the Web Server on which Workplace XT is running for the changes to take effect.

Now, users can view archived documents in Workplace or Workplace XT by double-clicking the document or by using the view function if the MIME type for this document is redirected to the Document Viewer. A browser window with a URL similar to the one in the following example will display the content of the selected document:

```
https://<<ICCServerName>>:11443/DocViewer/ViewItem.do?
p8ObjectStore={759E9C6F-2B90-432F-A12F-A82949545747}&reposID={CD25741B-C3EA-4A26-90E3-F6E9B1373012}
```

Where *<ICCServerName>* is the name of the Content Collector Web Application Server.

If any redirection errors occur, make sure that the redirection .jsp file that you created is located in the proper directory. If the URL of the browser window does not correspond to what you configured in the redirection file, check the MIME type mappings in the `content_redir.properties` file and correct them if necessary.

For further information about content redirection see the section about content redirection properties in the IBM FileNet P8 information center.

Tip: On the first redirection to the Document Viewer, users are prompted to log on to the repository. To avoid this, you can include the user ID and password in the URL that you configure in the redirection file.

Part 7. Monitoring

© Copyright IBM Corp. 2008, 2010

Monitoring Content Collector system performance

You have various possibilities to monitor the performance of your IBM Content Collector server, collectors, and task routes.

Using the system dashboard

The IBM Content Collector system dashboard is a tool with which to quickly determine which nodes and which task routes are actively processing documents, and which are not. It's aimed to provide a quick insight to the health of the IBM Content Collector system.

The system dashboard is deployed with the installation of IBM Content Collector Server. It must be run from a machine on which Content Collector is installed.

The system dashboard is intended to be run from the primary node in the Content Collector system configuration. It can run with or without the IBM Content Collector Configuration Manager started.

The status of each node and each task route is shown, along with the total number of errors a node has experienced. Optionally, the user may expand the view to show various performance counters for a given task route and the number of errors that task route has experienced on that node.

All node and task route information is obtained and updated dynamically. This means that the display reflects all changes, for example, if the primary node changes or task routes are added or deleted.

The status information of up to four nodes and several task routes for these nodes can be displayed on a screen without having to scroll. The dashboard window can be set to always appear as the top window by selecting **Options → Always on Top** on the system dashboard interface.

To start the system dashboard:

Select:

Option	Description
All Programs → IBM Content Collector → System dashboard	The IBM Content Collector start menu folder
Tools → System Dashboard in the IBM Content Collector Configuration Manager	The IBM Content Collector Configuration Manager
`SystemDashboard.exe` and run the executable	The IBM Content Collector installation directory in the `ctms` subfolder

In some cases, a counter value may not be a number but may display the string `Unknown`. This may be the case if performance counter information cannot be accessed, or if a task route's event log could not be found. The Windows event log will not have been created for the task route if the task route is newly-created and has not been run yet.

© Copyright IBM Corp. 2008, 2010

All error information gathered from the system dashboard while it is running is logged in the system dashboard log file, located in the `ctms\Log` directory.

Information monitored in the system dashboard

The IBM Content Collector system dashboard runs alongside Content Collector and provides real time performance information for each task route on each Content Collector node.

It uses a subset of all the available counters you see in "Performance counters" on page 762. A more advanced user has the option of obtaining further information by monitoring additional performance counters by using a profiling tool, for example, perfmon.

The following type of information is displayed on the dashboard interface:

- All nodes in the current IBM Content Collector configuration
- All task routes, both active and inactive ones
- Performance counter information for each task route including:
  - Total documents accessed: the total number of items that were accessed by the collector since the node was last restarted
  - Documents accessed/sec: the number of items per second that the collector is currently processing
  - Total unique documents created: the total number of unique documents that were created by a given task route since the node was last restarted
  - Unique documents created/sec: the number of unique documents per second that the task route is currently creating
  - Total errors: the number of errors that this task route has experienced on that node since the task route was created
- An error count for each node

The name of each machine is displayed with one of the following status label:

Services Running
The IBM Content Collector Task Routing Engine serviceis running.

Services Stopped
The IBM Content Collector Task Routing Engine service has stopped.

No Connection
The machine cannot be reached.

Additionally, when the services are running, the primary node in the IBM Content Collector configuration is identified by the **Primary Node** label.

Each node that is displayed contains a list of task route names. A task route can have one of the following task route status labels:

Processing
The task route is currently processing documents.

Idle The IBM Content Collector Task Routing Engine service is running and the task route is currently monitoring collection sources but is not processing any documents. This does not imply that the collector is not running. The collector is running and is searching mailboxes for content to archive, but cannot found anything eligible for archiving.

Stopped
The IBM Content Collector Task Routing Engine service is stopped.

No Host Connection
The host machine cannot be reached and thus no task route performance counter information can be obtained.

Error Task route performance counter information cannot be read because this information does not exist on the host machine, or permissions are restricted.

If performance counter information cannot be accessed, the counter number will change from `0` to `Unknown`.

Inactive
The task route is not active.

Information about the status of the nodes and task routes is retrieved by examining the task route tables in the IBM Content Collector datastore. The status of a task route is determined by examining whether the throughput-related counter value is increasing in value, and whether the IBM Content Collector Task Routing Engine service is started on the node in question.

The status of a node is determined by pinging the node, and by examining the statuses of the task routes for that machine. If task routes are processing or idle, the machine is labeled as **Services Running**, and if the task routes are stopped then the machine is labeled as **Services Stopped**.

Performance counter information for each task route is displayed below each task route name when you expand the view. The performance counters related to document processing are system performance counter, while the **Total Errors** counter monitors the total number of errors in that task route's Windows event log for the machine on which the dashboard is running.

Using performance counters

Performance counters provide an in depth monitoring of the system, recording the details about the Content Collector processing progress, such as the number of documents accessed, the number of documents submitted for processing, and the number of unique and duplicate document instances.

IBM Content Collector provides a variety of performance counters that can be used to monitor what the system is doing. Refer to the table of IBM Content Collector performance counters for more information about which counters are available.

If all you want is a quick insight to the health of your Content Collector system without having to configure a profiling tool, use the Content Collector system dashboard. It uses a subset of the available Content Collector counters.

If there are other performance counters that you want to monitor that are not included in the system dashboard, there are various tools that you can use to monitor these counters, for example the Microsoft Windows performance monitor or IBM System Director.

Tip: On 64-bit versions of Windows Server, you should use the 32-bit version of the performance monitor. You can start it with the command `mmc /32 perfmon.msc`.

The Performance Logs and Alerts service must be active. On Windows Server version 2008 you must ensure that the Performance Counter DLL Host Service is also running.

To add performance counters to the performance monitor that comes with Windows:

1. Run **perfmon.exe**. This will open the Windows performance monitoring utility.
2. Right-click in the graph area and select **Add Counters...**.
3. Select the **Performance object** to which the counter that you want to add belongs, for example CTMS Collector.
4. Select which counters you want to add.
   - Select **All counters** to monitor all counters.
   - Select **Select counters from list** to monitor only the counters that you select.
5. In the **Instances** list, select the instance for which you want to monitor the counts. An instance can be a task route or a collector, for example.

   Note: There are instance counters and instanceless counters. In Windows performance monitor, you cannot monitor instance performance counters if there is no instance. This means that you cannot add these counters until IBM Content Collector has been started and the counters have been instantiated. For example, you cannot configure the tool to monitor the counter Documents Created before a document has been created. You can work around this problem by either using another tool to monitor counters or by completing the following steps:
   a. Save the current set of counters. Right-click in the graph area and select **Save as...** to save the counter set in an HTML file.
   b. Open the HTML file in a text editor and manually add the instance counters that you want to monitor.
6. Click **Add**. The graph begins updating every couple of seconds to show the current counts.

Performance counters

IBM Content Collector provides a variety of performance counters.

The following counters are useful for regular system monitoring. Some of these counters are monitored by the Content Collector system dashboard.

- Accessed Entities (monitored by the dashboard)
- Accessed Entities/sec (monitored by the dashboard)
- Documents Created (monitored by the dashboard)
- Documents Created/sec (monitored by the dashboard)
- Duplicate Documents (if archiving to FileNet P8)
- Additional instances (if archiving to Content Manager)
- Run Count
- Searched Locations

Table 189 on page 763 lists the available performance counters to monitor IBM Content Collector.

Table 189. Performance counters

Counter name	Performance object	Counter description	Recorded on
Accessed Entities	CTMS Collector	Number of items accessed by the collector. If no errors have occurred, this will be equal to the submitted items plus the skipped items.	Primary node
Accessed Entities/sec	CTMS Collector	Number of items accessed per second.	Primary node
Connector Failure	CTMS Connector Health	Number of times a connector failed. If you select the instance **_Total** for this counter, the counter shows the total value for all connectors.	Any node
Documents Created	CTMS Target	Number of documents created in the repository.	Any node
Documents Created/sec	CTMS Target	Number of documents created in the repository per second.	Any node
Entity Errors	CTMS Collector	Number of items that were skipped by the source connector due to errors.	Primary node
Entity Errors/sec	CTMS Collector	Number of items that were skipped per second due to errors.	Primary node
Forwarded Items	CTMS Core	Number of items that were forwarded to a secondary task route service.	Primary node
Forwarded Items/sec	CTMS Core	Number of items that were forwarded to a secondary task route service per second.	Primary node
Forwarding Errors	CTMS Core	Number of errors that occurred while attempting to forward to secondary nodes.	Primary node
Forwarding Errors/sec	CTMS Core	Number of errors that occurred per second while attempting to forward to secondary nodes.	Primary node
Item Backlog	CTMS Instanceless Core	Number of submitted items for which processing has not yet started.	Any node
Location Errors	CTMS Collector	Number of locations that were skipped due to errors.	Primary node
Location Errors/sec	CTMS Collector	Number of locations that were skipped per second due to errors.	Primary node
Rejected Items	CTMS Core	Number of items that were rejected because they were already being processed at submission time.	Any node
Rejected Items/sec	CTMS Core	Number of items that were rejected per second because they were already being processed at submission time.	Any node
Restart Count	CTMS Connector Health	Number of times a connector was restarted. If you select the instance **_Total** for this counter, the counter shows the total value for all connectors.	Any node
Run Count	CTMS Collector	Number of times the collector has run.	Primary node
Runs/sec	CTMS Collector	Number of times the collector has run per second.	Primary node
Searched Locations	CTMS Collector	Number of locations searched.	Primary node
Searched Locations/sec	CTMS Collector	Number of locations searched per second.	Primary node
Secondary Nodes	CTMS Instanceless Core	Number of secondary nodes.	Any node

Table 189. Performance counters (continued)

Counter name	Performance object	Counter description	Recorded on
Submitted Items	CTMS Core	Number of items that were submitted to the task route service.	Primary node
Submitted Items/sec	CTMS Core	Number of items that were submitted to the task route service per second.	Primary node

Table 190 lists the available performance counters to monitor information about the Content Manager connector.

Table 190. Performance counters for Content Manager

Counter name	Performance object	Counter description
Additional Instances	CMV8 Connector	Number of additional instances that have been recorded. For example, if two people receive the same email, both have an instance of the email. For the first instance to be archived, the Documents Created counter is incremented. For any additional instances, the Additional Instances counter is incremented.
Instance Uniqueness Constraint Violations	CMV8 Connector	If the record of an additional instance cannot be created because one of the attribute values of the instance violate a uniqueness constraint, then this counter is incremented.
JVM Free Memory	CMV8 Connector JVM	Portion of total memory that is not in use.
JVM Max Memory	CMV8 Connector JVM	Maximum memory that the JVM will use.
JVM Total Memory	CMV8 Connector JVM	Amount of memory that the JVM currently uses.
JVM Used Objects	CMV8 Connector JVM	Number of objects on the heap. This counter is updated if memory dumps have been enabled or if the environment variable IBM_CTMS_CMV8_SHOW_OBJECT_COUNT has been set.

Table 191 lists the available performance counters to monitor information about the FileNet P8 connector.

Table 191. Performance counters for FileNet P8

Counter name	Performance object	Counter description
Duplicate Documents	P8 Connector	Number of duplicate documents that the archive task detected.
Duplicate E-mail Instances	P8 Connector	Number of duplicate email instances that the P8 4.x Create Email Instance task detected.
E-mail Instances Created	P8 Connector	Number of email instances created.

Tracking system log files

Monitor the system log files for the various components of IBM Content Collector to detect problems and errors that might occur during processing.

What logs to track

IBM Content Collector uses log messages to record information about events.

The table includes the most important logs in IBM Content Collector. When you send these logs to IBM Software Support for faster problem determination, remember that the logs might contain sensitive information.

Table 192. Log files for the most important IBM Content Collector services

Service	Location	Description
Task route service log files	Default: *ICC-install-directory*\ctms\Log\ibm.ctms.taskrouteservice*	Monitors all archiving processes and records messages from the task routing service. For example, if there is an error when a connector invokes a task, it is recorded here. Errors that occur while crawling a mailbox are not recorded, but detailed connector specific information is contained in the connector log files. Begin viewing this log file and the connector log files to resolve archiving related problems in the system. **SharePoint only:** You can disregard the warning: `Failed to initialize COM security in WMIServiceController. Error code -2147417831`

Table 192. Log files for the most important IBM Content Collector services (continued)

Service	Location	Description
Back-end log files of the email server connectors	Default directory: *ICC-install-directory*\ctms\Log\ If the environment variable AFU_NO_SUPPORT_INFO is not set, a subdirectory called support is created in the *ICC-install-directory*\ctms\Log directory that contains information packages listing the email that was involved in error situations. This occurs in addition to the logs that are written to the default directory. You can send these packages to IBM Software Support for faster problem determination, however, remember that email in these packages might contain confidential information. To avoid generating these support packages, set the environment variable AFU_NO_SUPPORT_INFO to true. If a connector subprocess fails, a *log-directory*/SUPPORT/CRASH/*timestamp* directory is created that contains error information. This directory is created irrespective of the setting of the environment variable AFU_NO_SUPPORT_INFO. If such an error occurs, check the blacklist, save the information in the *log-directory*/SUPPORT/CRASH/*timestamp* directory and provide the information in this directory to IBM Software Support for faster problem determination.	Monitors email through SMTP, Microsoft Exchange or Lotus Domino mailboxes and other collection sources in your email system.

Table 192. Log files for the most important IBM Content Collector services (continued)

Service	Location	Description
Web Application log files	• Web Application log files default directory: *ICC-install-directory*\AFUWeb\afu\logs. The afu-logging.properties file lists the number of log files that are produced and the names of these files created under \logs. • Embedded WebSphere Application Server log files and Web Application start up errors default directory: *ICC-install-directory*\AFUWeb\profiles\<server>logs. • Configuration Web Service log files: If you chose to use the embedded WebSphere Application Server during the Web Service configuration and want to track processing for troubleshooting purposes, set the environment variable AFU_ADVANCEDEMAILCONFIG to 1. After you have set the variable, you must restart the Configuration Manager. Setting this variable to 1 will lead to the creation of a log file for each script that runs during the automatic Web Service configuration, and also logs when the web server is started and stopped. The last configuration you saved is also logged. All log files are written to the directory that the TEMP environment variable points to. The following naming convention is used: AdvancedEmailConfig.ConfigName.xml or AdvancedEmailConfig.ConfigName.log	Monitors Web applications.
Configuration Manager logs	Use the Windows Event Viewer for error messages.	Monitors Configuration Manager activities.
Content Manager connector log files	Default directory: *ICC-install-directory*\ctms\Log When you send these log files to IBM Software Support for faster problem determination, remember that the log files might contain sensitive information.	Monitors Content Manager connector activities and errors. You can use this log to infer information about the communication between Content Manager 8 and IBM Content Collector.
FileNet P8 connector log files	Default directory: *ICC-install-directory*\ctms\Log When you send these log files to IBM Software Support for faster problem determination, remember that the log files might contain sensitive information.	Monitors FileNet P8 connector activities and errors. You can use this log to infer information about the communication between FileNet P8 and IBM Content Collector.

Table 192. Log files for the most important IBM Content Collector services (continued)

Service	Location	Description
Microsoft SharePoint connector log files	Default directory: *ICC-install-directory*\ctms\Log When you send these log files to IBM Software Support for faster problem determination, remember that the log files might contain sensitive information.	Monitors Microsoft SharePoint connector activities and errors. Errors can range from connectivity and authentication errors during document retrieval to post-processing errors that concern document locking and link creation.
SMTP Receiver log files	Default directories: • *ICC-install-directory*\james\james-2.3.2\logs • *ICC-install-directory*\james\james-2.3.2\apps\james\logs When you send these log files to IBM Software Support for faster problem determination, remember that the log files might contain sensitive information.	Monitor the SMTP Receiver services. The log files in *ICC-install-directory*\james\james-2.3.2\logs contain information about the startup and shutdown processing of the service. The log files in *ICC-install-directory*\james\james-2.3.2\apps\james\logs contain information about message processing.
Audit log files	Default directory: *ICC-install-directory*\audit When you send these log files to IBM Software Support for faster problem determination, remember that the log files might contain sensitive information.	Audit logs can be added to a task route and monitor the final status of each item that is submitted to the task route service for processing. You can configure which metadata properties are recorded.
Outlook Web Access log files	Outlook Web Access Extension: Default directory: *Outlook-Web-Access-install-dir*\AFUOWAExtension\configOWA.trc	Monitors the Outlook Web Access Content Collector enablement processing.
	Outlook Web Access Services: 1. Default: *Outlook-Web-Access-install-dir*\logs\afuowa.trc 2. Default: *Outlook-Web-Access-install-dir*\logs\afuowaconfigs.trc 3. C:\Inetpub\wwwroot\afuowa\owaapp.log	1. Monitors the Outlook Web Access Services. 2. Monitors the synchronization of the Outlook Web Access Services settings and the Content Collector server. 3. Monitors the final startup processing of the services.

File format and naming conventions for system log messages in Content Collector

To make logging more consistent, most system log messages that are recorded when IBM Content Collector looks for and processes email have a specific format and follow the same naming convention.

File format

The log message format is as follows:

Table 193. Log message format

Timestamp	Level	Message	Method (offset)	Thread
2009-01-20T13:38:12Z	Trace2	Initialized task queue with 8 threads, and 64 max tasks	core.dll:0x234cd	0x10d4

The column definitions are:

Timestamp
UTC format (ISO8601)

Level For example, Warning, Error, or Fatal

Message
A description of the log message

Method (offset)
Either the current method's binary object and offset (if available) or the method name

Thread
Tread ID or thread name

File naming conventions

Except for the exceptions that are mentioned below, a log file name has the following naming convention, where subidentifier is an optional parameter (for example, trace or sysout):

```
connector_name/component_name-YYYY-MM-DD[_subidentifier]
```

The following components use a different format:

Connectors (except for the mail connector and the text extraction connector)
In general, connector log files have the following naming convention:

ConnectorID-YYYY-MM-DD.log

For example:

```
ibm.ctms.p8connector.p84x.P84xConnector-2009-04-17.log
```

Mail connector
The mail connector log files have the following naming convention:

```
afu_mailconnector_sysout|trace_nn
afu_monitor_sysout|trace_nn
```

Text extraction connector
The text extraction connector logs to the Windows event log "IBM Text Extraction Connector".

Web Application
The Web Application log files use the default naming convention for the WebSphere Application Server log.

You can configure the system to work with up to n rotated log files that are numbered in the order in which they are created. After the number of log files reaches the configured number, the subsequent file rotation deletes the oldest log file.

Log levels

When you are configuring a log, you can select the type of data that should be written to the log file. Available options are listed here, from least to most verbose.

Because most error occur during the initial setup, you should perform a test run of your system with a log level of type Information or higher. This will give you sufficient information to fix the errors. When your system is configured correctly, restore the log level to Warning or Error.

Note that log entries are cumulative. A log level of type Information, for example, will include Information, Warning, Error, and Fatal log entries.

Table 194. Log levels

Log level	Description
Fatal	An event is written to the log whenever a severe problem or critical condition occurs. This log level provides minimum logging detail.
Error	An event is written to the log whenever an error condition occurs - such as when a connection attempt to a server fails.
Warning	An event is written to the log whenever a warning condition occurs - such as when the server cannot understand a communication sent to it.
Information	An event is written to the log with every significant action that takes place - such as when a document is collected.
Trace	Events are written to the log at individual steps. This log level provides verbose logging. It is useful only for debugging purposes. **Important:** This will slow down your system.
Trace 2	Extended debugging information is logged. This log level provides the most verbose logging. Use with caution as the log file will grow very large very quickly. **Important:** This will slow down your system.

Using audit logs

Items that are processed by a specific task route can be monitored in an audit log. Include an Audit Log task in your task route to record information about the final status of every item that is processed by the task route.

Audit logs also record the arbitrary meta data of processed documents. This can be used to gather statistics on the document size and type. The audit log can also be

used to understand the history of the system throughput without having to monitor performance counters for long durations.

Every item that is processed by the task route creates one or multiple entries in the log file that indicate whether the item was processed successfully or not. Additionally, some metadata information about the item is recorded.

These items are typically the documents that are collected by a collector. However, other tasks can also submit items to the task route service, for example the EC Extract Attachments task or the FSC Associate Metadata task.

You can configure which metadata properties are logged for each task, depending on what you want to analyze. For example, you can use audit logs to analyze the number, type, and size of file items or email items that are processed by either a specific task route (if one audit log per task route is configured), or by all task routes (if one audit log for the whole system is configured). By default, the audit log records amongst other information the document size, whether the document is a duplicate of another document that has been archived already, and, if applicable, the attachment file names, types, and sizes. This information can for example be used to analyze the size and type distribution of the processed documents and their attachments. This information may help to better estimate storage needs for the content repository.

Audit logs can be used as an alternative to performance counters to monitor proper system operation and throughput.

Related tasks

"Including an audit log task" on page 279

Using event logs

Configuration Manager events are logged in event logs. Monitor these logs to detect problems and errors that might occur during processing.

Configuration Manager events are logged in various Windows event logs. An event log record consists of an event type (success, warning, informational, or error), an event ID, a description, and an optional binary blob of data. Refer to the table of event IDs for detailed information on what the different collector events are.

Event log records provide more detail than performance counters. However, this detail comes at a cost in memory and processor time. Therefore some events are recorded by incrementing a performance counter instead of creating an event log record.

There are many ways to view, manage, and monitor event logs. The easiest way is to use the basic event viewer that comes with Microsoft Windows. It is part of the system or administrative tools. Run `eventvwr.exe` to access it.

You can configure a maximum log size and specify the action that should be taken when this limit is reached. To configure the size of an event log file:

1. Run **eventvwr.exe** to open the Windows Event Viewer.
2. Select the event log that you want to configure. All error events that occur independent of a task route or collector are logged in the main Windows event log named Application. All events that occur in the context of a task route are

logged in one event log for this task route. This event log is named CTMS - *TaskRouteName*, where *TaskRouteName* is the name of the task route.

3. Right-click the event log and select **Properties**.
4. Specify the maximum log size.
5. Select what action should be taken when the maximum log size is reached.
   - Select **Overwrite events as needed** when event records are used only as triggers.
   - Select **Overwrite events older than ? days** to keep event records for the specified number of days.
   - Select **Do not overwrite events (clear log manually)** to retain all event records. You should set up a script or utility to back up and clear the event log regularly, if you select this option.

Interpreting event logs

Event log records contain a lot of information. In the event log of a task route, you can monitor when a collector was running and how many documents were processed.

When a collector runs, the start and end event are recorded in the event log of the respective task route. You can use this information to determine whether a collector is currently running or when it last finished.

When a collector skips or collects documents, this is also recorded, and you can use this information to count for example how many documents have been processed, or could not be collected.

1. In the Windows Event Viewer, click on the **Time** field to sort the events by the time they occurred.
2. Locate the latest event ID that you are interested in. Refer to the list of event IDs to find out what the different IDs stand for.
3. Double-click the event record to see detailed information about the event. For example, the description contains the information to which collector the event is related.
4. If there is a start event (event ID 128) but no corresponding finish even (event ID 128 through 132), the collector is currently running.
5. Other event IDs give information about documents that have been collected, skipped, or could not be collected.

Deleting event logs

For each task route, an event log is created automatically. You can delete it manually.

Event logs are stored as files and accessed using for example the Windows Event Viewer. To delete an event log:

1. Make sure that no collectors are running.
2. Run **regedit.exe**. This will open the Windows Registry Editor.
3. Locate the registry entry for the event log that you want to delete. The entry is stored under the registry key `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Eventlog\`*`EventLogName`*, where *EventLogName* is the name of the event log, for example CTMS - Task Route.

4. Determine the name of file that is used to store the event log. The name is stored in the registry value `File`. For example: `C:\WINDOWS\system32\config\em00001b.evt`
5. Remove the registry key `HKEY_LOCAL_MACHINEHKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Eventlog\`*`EventLogName`*.
6. Restart the machine.
7. Delete the file that was used to store the event log. The file is stored in `%SystemRoot%\System32\config`.

Event IDs

In the event log, events are recorded with different event IDs.

Table 195 lists the collector events that can appear in the event log of a task route. The following variables are used in the event message descriptions:

Collector
The collector name.

StartTime
The time when the collector started running.

Location
The location specified by the user.

Document
The document ID.

Reason The reason.

Error The error message.

Table 195. Event IDs of collector events

Event type	Event ID	Event message
Informational	119	Collector "*Collector*" skipping entity "*Document*" from location "*Location*" in run started on *StartTime*. Reason: "*Reason*"
Error	120	Collector "*Collector*" experienced error "*Error*" while trying to collect entity "*Document*" from location "*Location*" in run started on *StartTime*.
Informational	128	Collector "*Collector*" started on *StartTime*.
Informational	129	Collector "*Collector*" finished search started on *StartTime*.
Informational	130	Collector "*Collector*" stopping search started on *StartTime* due to a scheduled stop.
Informational	131	Collector "*Collector*" stopping search started on *StartTime* due to service shutdown.
Error	132	Collector "*Collector*" halted run started on *StartTime* due to the following error: *Error*
Informational	144	Collector "*Collector*" beginning search of location "*Location*" as part of run started on *StartTime*.
Informational	145	Collector "*Collector*" has finished search of location "*Location*" as part of run started on *StartTime*.
Informational	146	Collector "*Collector*" stopped search of location "*Location*" as part of run started on *StartTime* due to a scheduled stop.
Informational	147	Collector "*Collector*" stopped search of location "*Location*" as part of run started on *StartTime* due to service shutdown.

Table 195. Event IDs of collector events (continued)

Event type	Event ID	Event message
Informational	148	Collector "*Collector*" skipping search of location "*Location*" in run started on *StartTime* finished. Reason: *Reason*
Error	149	Collector "*Collector*" stopped search of location "*Location*" in run started on *StartTime* finished. Search stopped due to error: *Error*

Table 196 lists the task route events that can appear in the event log of a task route. The following variables are used in the event message descriptions:

CollectorTask
The name of the collector task.

Document
The document ID.

Node The node name.

Error The error message.

Table 196. Event IDs of task route events

Event type	Event ID	Event message
Informational	160	Task Routing Engine received a submission for entity "*Document*" from collector "*CollectorTask*".
Informational	161	Task Routing Engine rejected a submission for entity "*Document*" from collector "*CollectorTask*".
Informational	162	Task Routing Engine forwarded entity "*Document*" from collector "*CollectorTask*" to secondary node "*Node*".
Error	163	Error "*Error*" while forwarding entity "*Document*" from collector "*CollectorTask*" to secondary node "*Node*".

Part 8. Using

© Copyright IBM Corp. 2008, 2010

Using Content Collector

From your client application, you can manually archive documents with IBM Content Collector, search for archived email, and view or restore archived documents.

Searching, viewing, and restoring documents

In IBM Content Collector, you can search email documents that were archived from your mailbox or personal storage file (PST file) or the local database into a central repository. You can then preview the documents that met your search criteria and restore them to your mailbox. Although it is not possible to search documents that were archived from other sources, for example, Lotus Notes applications, a file system, or Microsoft SharePoint, it is possible to view those documents.

You can search for archived email using the Content Collector search function from the Outlook main window or in Lotus Notes. The search application opens in your web browser.

You can preview and restore archived attachments or document content. To do so, you use the stubs provided by Content Collector. If a user delegates access to the mailbox, it depends on the permissions that the delegates are granted whether they can search, view, or restore archived email.

For documents archived from a file system or Microsoft SharePoint, each document is replaced with a shortcut to the archived document. You use this shortcut to access the archived document in its new repository location and view a read-only copy of the file.

Accessing archived email

You can search, view, and restore the archived content of email documents and attachments. Using the Email Search function of Content Collector's web client, you set up queries for searching archived email. You can view or restore archived content from links in the document stubs in the mailbox or from a search result list.

Delegate access to archived email

In Microsoft Exchange and Lotus Domino, it is possible to delegate access to a mailbox. Whether the delegates can search, view, or restore archived email depends on the permissions that they are granted.

Lotus Domino

For searching, previewing, or restoring email that was archived from Lotus Domino, the delegate must have the following access permissions:

Action	Access permission required
View archived email	"Read any document" or higher

© Copyright IBM Corp. 2008, 2010

Action	Access permission required
View and restore archived email	"Read and create any document, send mail on your behalf" or higher **Note:** With the access permission "Read and create any document, send mail on your behalf", the user might encounter error messages when restoring an email. These messages can be ignored. The email is restored successfully.
View, restore, and search archived email, manually archive email	"Read, edit, create, and delete any document, send mail, enable Out-of-Office" or "Read, edit, and create any document, send mail, enable Out-of-Office" **Note:** The access permission "Read, edit, and create any document, send mail, enable Out-of-Office" is sufficient for full access to archived email. However, the user might encounter error messages when restoring an email, searching, or marking an email for archiving. These messages can be ignored. The actions are carried out successfully. The error messages do not appear with access permission "Read, edit, create, and delete any document, send mail, enable Out-of-Office".

With permission "Read, create, edit, and delete any Calendar Entry, To Do, and Contact" or lower, delegates have no access to archived data.

Microsoft Exchange

If a user delegates access to an Exchange mailbox, delegates can search and view content that was archived from this mailbox under certain conditions:

- Delegates that use IBM Content Collector Outlook Extension can search for archived content if all of the following conditions are met:
  - The search function is installed.
  - The delegate has read access to the inbox folder.
- Delegates that use IBM Content Collector Outlook Web Access can access the archived data if the search function has been enabled in Outlook Web Access and if they can access either the mailbox top level folder or one of the following folder types:
  - RSS feeds
  - Mail and Post Items (including the predefined folders Deleted Items, Drafts, Inbox, Junk Email, Outbox, and Sent Items)
  - Journal Items
  - Note Items

In both Microsoft Exchange and Lotus Domino, access to email can be controlled not only on mailbox level, but also on folder level and document level. Access control lists define which users can view or modify which documents. When a document is archived, this access control is not reflected in the archive. This means that if delegate users have the required permissions to search and restore documents from a specific mailbox, they can find and view all documents that were archived from this mailbox, irrespective of the original access control. So a delegate user might be able to view documents that were marked private or that were archived from a folder that the delegate user is not allowed to access.

For Lotus Domino, the access control lists are reinstated when a document is restored. This means that delegate users might find and view a document that they can no longer access once it has been restored, because the delegate users do not have permission to access this document.

Searching for archived email

You compose and submit your archive queries on the Email Search page. The Email Search page allows you to enter dates, names of senders and recipients, subject information, and lines of text to be found in archived email bodies and attachments. You can use one or more of these search criteria in combination to find documents in a repository. If the search function is configured accordingly, further attributes are available for search.

To perform a search:

1. If more than one collection set was defined, select a collection set. Note that you can search only one collection set at a time.
2. To search for archived documents that were originally created or received on a certain date or during a certain period, follow these steps:
   a. Click the icon next to the first field under **Date range** and select a date from the date picker that opens.
   b. Click the icon next to the second field under **Date range** (further to the right) to select another date. The first date marks the start of the period; the second date its end.
   c. Select the sort order. This option is not available in these cases:
      - The current collection set contains more than one collection.
      - There is only one collection defined but this collection contains multiple item types or document classes that were defined with overlapping date ranges.
3. To search for archived documents that were sent or received by certain users, follow these steps:
   a. Enter the email addresses or Lotus Notes addresses of the senders in the **Sender** field. Separate multiple entries by commas.

      A match is shown in the result list if at least one of the specified addresses is included in the **Sender** field of a document.
   b. Enter the email addresses or Lotus Notes addresses of the recipients in the **Recipients** field. Separate multiple names by commas.

      A match is shown in the result list if at least one of the specified addresses is included in the **Recipients** field of a document.

   If both, the **Sender** field and the **Recipients** field contain values, you can widen the scope of your search by selecting **OR**. The search then returns documents containing one of the specified users in the **Sender** field or in the **Recipients** field. If you select **AND**, only such documents are returned that show at least one of the users in both, the **Sender** and the **Recipients** fields.

   Restriction: If you are searching for documents in non-indexed archive, the rules for attribute search apply (see "Guidelines for searching in the database (attribute search)" on page 781.)
4. To search for archived documents containing a certain text in the subject field, the document body, or in one of the attachments, follow these steps:
   a. To search for text in the subject fields of archived documents, enter a word or a phrase in the **Subject** field.

b. To search for text in the document bodies and attachments of archived documents, enter a word or a phrase in the **Body and attachments** field.

If both, the **Subject** field and the **Body and attachments** field contain text, you can widen the scope of your search by selecting **OR**. The search then returns documents that contain the search text of either entry field in just one of the document parts referred to. If you select **AND**, only such documents are returned that show the search text of either entry field in the subject field and in the body or attachments field, respectively.

5. To submit your search, click **Search**.

   Note that if you select a document in the search results, a preview of that document is displayed in the Preview pane.

   Limitation: The Content Collector can display up to 10000 search results. With an average of 20 to 25 documents visible at a time, this means that you can scroll the results set window approximately 400 times before reaching this limitation.

6. To clear all entries and start over, click **Clear**.

Guidelines for searching:

Choosing the right search terms is the key to finding the information you need. Often a series of searches is more effective at getting the results you want. For example, you can start with a general search and perform progressively more restrictive searches to narrow the results based on patterns in the results. Alternatively, if you are looking for very specific information, you can start with a very restrictive search, and if you do not get any results, you can perform progressively less restrictive searches to broaden the search.

You can search for text (words, phrases, character strings) in the text content sections of archived documents or for attribute values of archived documents. The text or the attribute value that you search for is called the search term or query term. When the search term is found in an archived document, the document is put on a result list, from which it can be accessed.

Guidelines for searching a full-text index (text search)

- Search terms are not case sensitive. For example, searches for legal, Legal, and LeGaL will all return the same results.
- To find an exact match, put double quotation marks around your search term, for example, `"Alaska pipeline"` or `"endangered species"`.
- To combine search terms within a search field, use the logical operators AND, OR, and NOT. Use AND to find documents that contain multiple terms, OR to find documents that contain at least one of the terms, and NOT to find documents that do not contain the specified term. For example:
  - To find documents that contain both of the words plants and animals, specify: `plants AND animals`.
  - To find documents that contain either of the words plants or animals, specify: `plants OR animals`.
  - To find documents that contain either of the words fauna or flora, but do not contain the words plants and animals, specify: `fauna OR flora NOT plants NOT animals`.
- You can use wildcard characters within search terms. Use a question mark (?) to match a single character or use an asterisk (*) to match multiple characters, preceded by at least two characters. For example, specifying `tax*` might find

documents containing the words Taxodium, taxation, or taxonomy. Note, however, that using wildcards in text searches has a significant impact on the overall search performance.

Filenet P8 systems only: Wildcards are only supported for searches in the full-text index.

- To enter separate search terms for the sender and recipients, expand the **Addresses** section.
- To enter separate search terms for the subject and body content, expand the **Content** section.

Guidelines for searching in the database (attribute search)

- Search terms are case sensitive. For example, searches for legal, Legal, and LeGaL will not return the same results.
- Searches return exact matches only. For example, if you specify `Alaska pipeline` in the Subject field, only documents with the exact subject are found.
- To combine search terms within a search field, use the logical operators AND, OR, and NOT. Use OR to find documents that contain exact matches for one of the terms, and NOT to find documents that do not contain exact matches for the specified term. For example:
  - To find documents that contain an exact match for either of the words plants or animals, specify: `plants OR animals`.
  - To find documents where the subject starts with tax or tex but to exclude all documents with a subject beginning with texas, specify: `(tax* OR tex*) AND NOT texas`. The result list then includes documents with the subjects "tax reduction" and "text" but does not include documents with the subject "texas oil".
- You can use wildcard characters within search terms. Use a question mark (?) to match a single character or use an asterisk (*) to match multiple characters, preceded by at least two characters. For example, specifying `tax*` might find documents containing exact matches for a phrase starting with tax (for example, tax reduction or taxonomy). Note, however, that using wildcards in searches has a significant impact on the overall search performance.
- To enter separate search terms for the sender and recipients, expand the **Addresses** section.

Search tips for text search:

Each field on the Search pane has some additional options that can help refine your search.

Date range search

To perform a date range search, enter both a start date and an end date.

The date of a document is the date on which the document was created, posted, or modified using Greenwich Mean Time (GMT). If you search for a document matching a specific date, there is a chance that the document might not show up in the search results due to the time difference between your local time, the time stored in the archive, and GMT. If you do not find the document that you expect to find, extend the search to include the day before and the day after.

Example of valid date range searches:

Start 01/01/2000

End 12/12/2007

Sender or recipients search

The sender and recipients fields are the To and Cc fields.

When entering search terms for the sender or recipients fields, you can enter complete or partial names or email addresses. Typically, searches on names are more likely to obtain results than searches on email addresses.

In general, use double quotation marks when entering search terms for the sender or recipients fields. If you do not use double quotation marks, any space in the search terms is interpreted as the Boolean operator AND.

Examples of valid sender or recipients searches:

Common name search
"john q. smith"

Common name search with wildcards
john sm*

Common name search with a Boolean operator
john AND smith

Common name search with multiple Boolean operators
(john OR bill) AND smith

Email address search (Internet format)
"jqsmith@loc.com"

Email address search (Internet format) with wildcards
jqsmith@*

Email address search (Abbreviated format used by Lotus Domino)
"john q.smith/orgunit/org"

Email address search (Canonical format)
"cn=john q.smith/ou=orgunit/o=org"

Simple keyword search

To perform a simple keyword search, enter one word or require all of the words with the Boolean operator AND. For example, if you enter `king`, all documents that contain an exact match for `king` are returned. If you enter the search terms `king AND lear`, documents that contain both `king` and `lear` are returned.

To see more precise results, use more specific keywords. For example, use `French AND roast AND coffee` rather than `coffee`, and use `Kauai AND hiking AND tours` rather than `Hawaiian AND vacations`.

If a simple keyword search returns results that are not useful to you, exclude some terms or add more terms to refine your search. If the search returns too few results, use wildcard characters.

Requiring at least one of the search terms to appear

The Boolean operator OR specifies that at least one of the search terms must appear in the document. For example, searching for `(othello OR otello)` returns results that contain either `othello` or `otello`.

A space between terms is interpreted as the Boolean operator AND. For example, `othello otello` is interpreted as `othello AND otello`.

You can combine OR with other boolean operators. Group the expressions with parentheses to make the combinations clear. For example, `(king AND (lear OR edward))` returns all documents that contain `king` and `lear` or contain `king` and `edward`.

Excluding terms

Use NOT (in capital letters) to exclude documents that contain certain terms. For example, if you want to find documents that contain the term `lear` but do not contain the word `edward`, enter the search terms as `lear AND (NOT edward)`.

Examples of valid searches that exclude terms:

(cougar OR jaguar) AND (NOT car)
: Returns all documents that contain `cougar` or `jaguar` unless a document also contains `car`.

(cougar OR jaguar) OR (NOT car)
: Returns all documents that contain `cougar` or `jaguar`, and all documents that do not contain `car`.

(cougar OR jaguar) NOT car
: Equivalent to `(cougar OR jaguar) OR (NOT car)`

Phrase search

If you want to ensure that terms appear in documents in the same sequence in which you entered them, use double quotation marks. For example, if you want to find documents that contain the words `King Lear` with the words appearing together and in that order, enter `"king lear"`. In contrast, entering `king lear` without double quotation marks returns any documents that contain either of those words anywhere in the content.

Phrases are not sensitive to case. For example, `"king lear"` and `"King Lear"` return the same documents.

Wildcard characters

IBM Content Collector supports two wildcard characters: the asterisk (*) and the question mark (?).

Restriction:

- You cannot use wildcard characters inside double quotation marks.
- For IBM Content Manager repositories, wildcard characters in search terms consisting of double-byte characters only (for example, words or phrases in Chinese) are treated as literals. Therefore, such queries will not return the expected results, or they will even return no result at all.

asterisk (*)

The asterisk (*) wildcard matches zero or more characters. It can help you find words when you do not know the full spelling, or if you want to find words that are similar to a term. For example, the search term `nominat*` returns documents with the terms `nominate`, `nomination`, `nominated`, and other possible results.

question mark (?)

The question mark (?) wildcard matches a single character. It can help you find terms that differ by only one character. For example, `to?s` returns documents that contain the terms `toss`, `toes`, and `tots`, but not `toads`.

Note, however, that using wildcard characters in searches has a significant impact on the overall search performance.

Searching for special characters

Ensure that you use a backslash (\) in front of the listed characters. If a single backslash character appears before any character that is not listed in the following table, it is ignored. For example, entering `\$10` produces the same results as entering `$10`.

Restriction: The result of searching for special characters depends on what is stored in the index. Most special characters are treated as token delimiters and are not stored in the index unless you redefine the tokenization rules. These rules are set up in the `uni.cfg` file for FileNet P8 and in the `cteixcfg.ini` file for IBM Content Manager.

In an indexed archive, you cannot search for < (U+003C, "less-than sign"), > (U+003E, "greater-than sign"), the ampersand (U+0026), or the percent sign (U+0025). Whenever one of these characters appears in a document processed by the indexer for text search or in an embedded attachment, the character is ignored and will not be added to the full-text index. These characters do not exist in the text-search index because they were converted to space characters at the time the full-text index was created.

Special considerations for searching FileNet P8 repositories:

Special characters are escaped with a backslash as usual when you want to search for these as literals. In some cases, however, these escaped sequences still produce errors in the search application and do not return results. For example, properly escaped searches for literal question marks will fail.

Table 197. Special characters that require a backslash

Special character	Description	Example of searching email for this character
+	plus sign	6\+7
-	minus sign	6\-7
!	exclamation point	Stop\!
(	opening parenthesis	Juan \(Slugger\) Morton
)	closing parenthesis	Juan \(Slugger\) Morton
[	opening bracket	X\[YZ\]
]	closing bracket	X\[YZ\]
{	opening brace	X\{YZ\}

Table 197. Special characters that require a backslash (continued)

Special character	Description	Example of searching email for this character
}	closing brace	X\{YZ\}
^	caret	6\^2
"	quotation mark	John AND \"Bugsy\" AND Laws
~	tilde	\~11 million units
*	asterisk	Dial *69
?	question mark	What\?
:	colon	the following\:
\	backslash	the path C:\\Program Files
&&	double ampersands	A \&& B
||	double vertical bars; a single vertical bar does not need a backslash	A \\|| B

Viewing and restoring email from the search result list

Email Search returns a result list with all documents that matched your search request. From this list, you select the documents that you want to preview or to restore.

You can sort the result list by date if one of the following applies:

- You searched a single collection that contains just one item type.
- You searched a single collection that contains multiple item types. These item types are partitioned by date and the date ranges do not overlap.

  If date ranges for item types in a given collection overlap, the result list that is displayed is a concatenation of the sorted result lists for each item type.

You cannot sort the result list by date if one of the following applies:

- In a given collection, item types are defined with overlapping date ranges.
- Cross-collection search is configured.

When you select a document in the search result list, as much of the document content as can be displayed by the application without actually opening the document is shown. This is referred to as previewing. After restoring the document, you can open it in a native client and view it in full.

To preview a document in the search result list, click the document. The Preview pane opens to display the document's content.

Alternatively, double-click the document that you want to preview. A new browser window opens to display a full-screen preview of the document's content.

Search terms are highlighted as follows:

- Search terms that appear in the From, Subject, or Content fields are highlighted.
- Search terms that appear in the Recipients fields are highlighted only if the format of the property value for the email address that is used in the Preview pane (for example, the SMTP address or the display name) matches the search term. So, although an email document would qualify as a search result because

the search term matches the display name of the recipient (for example, search term "Mary Poppins" matches display name "Mary Poppins"), it would not be highlighted if, for the preview, the Recipient field was mapped to another property value, for example, the SMTP address (Mary@Banks.Family).

- Search terms that appear in attachment names or attachments themselves are not highlighted.
- Search terms that were used in searches on custom attributes are highlighted only if they refer to information that is contained within the email and that is visible when the email is previewed.

Preview provides an HTML representation of email documents. However, some rich text features are not supported.

You can also restore the documents in the search result list to your mailbox.

1. Select one or more of the documents from the search result list.
2. To restore the documents to a folder other than the default folder enter the name of that folder in the **Folder** field to the right of the **Restore** button. Note that folder trees are not supported, so you cannot restore documents into subfolders.
3. Click **Restore**. The selected document or documents are restored to your mailbox.

Restriction for documents that were archived with CommonStore: Only documents that were archived in a CommonStore item type for the document model BUNDLED and the archiving type ENTIRE can be previewed or restored using the search application. Documents that were archived in CommonStore item types for the document models GENERIC_MULTIPART or GENERIC_MULTIDOC must have the value of the <retrievable> element set to false in the search_mapping.xml file for the collection.

Changes to archived documents

Never change the content of a document after it was archived as this may lead to unexpected behavior.

A document is only archived once. Any change that is made to a document after it was archived is not saved to the archive because, for IBM Content Collector, the document is marked as having been archived.

A document might be stubbed after you have made changes to it. This means that the document content including the changes made to the content will be removed after some time depending on the document life cycle defined in the IBM Content Collector Configuration Manager. The changed content will be lost.

Do not change the content of an email after it was archived

Do not make changes to the body of an email or to its attachments after the email was archived as these changes will definitely be lost.

Do not resend a changed email that was already archived. This resent email is still marked as having been archived and will not be archived again. This applies to the sender's copy as well as to the recipient's copy. Additionally, the recipients are not able to search for resent email in the archive.

Do not archive draft documents

You should never archive a draft document. If you do, the document is marked as having been archived, and further changes will not be reflected. This means that if you later send out such a draft document, neither the sender's copy nor the recipient's copy is archived again.

As a consequence, a search for these email documents in the archive might return the original draft, but it will never return the email that was sent.

Configuring viewing in IBM FileNet Workplace XT

Configure IBM FileNet Workplace XT in a way that users can view email that was archived by IBM Content Collector.

The following prerequisites apply:

- IBM FileNet Workplace XT must be installed and set up properly.
- The IBM Content Collector Web Application must be running.
- Archived-data access must be configured for each repository in which you want to access archived documents.

To be able to view archived email in Workplace XT, you must configure Workplace XT accordingly.

1. Create a Java Server Pages (JSP) page with a name that is meaningful to you, for example, `DocViewerRedirect.jsp`. The JSP page must have the following contents:

   ```
   <%response.sendRedirect("https://<<ICCServerName>>:11443/DocViewer/ViewItem.do?
   p8ObjectStore="+request.getParameter("objectStoreName")+"&reposID="+request.getParameter("id"));%>
   ```

   Where *<ICCServerName>* is the name of the Content Collector Web Application Server.
2. Save the file to the `redirect` directory. This directory is located in one of these directories:
   - The *<WASinstall>*`\profiles\<profile name>\installedApps\<cell name>\WorkplaceXT.ear\web_client.war` directory, where *<WASinstall>* is the path to the installation directory of WebSphere Application Server, *<profile name>* is the name of the web application profile, and *<cell name>* is the respective cell name
   - The *<FileNetInstallDir>*`\FileNet\Config\WebClient` directory, where *<FileNetInstallDir>* is the installation directory of FileNet P8
3. Edit the `content_redir.properties` file. The file is located in the *<FileNetInstallDir>*`\FileNet\Config\WebClient` directory, where *<FileNetInstallDir>* is the installation directory of FileNet P8.
4. Map all MIME types that you want to view with the Document Viewer as in this example:

   ```
   image/gif=/redirect/DocViewerRedirect.jsp?
   objectStoreName={OBJECT_STORE_NAME}&id={OBJECT_ID}
   ```

   Where *DocViewerRedirect.jsp* is the name of the file that you created in step 1.
5. Save your changes.
6. Restart the Web Server on which Workplace XT is running for the changes to take effect.

Now, users can view an archived email document in Workplace XT by double-clicking the document or by using the view function if the MIME type for

this document is redirected to the Document Viewer. A browser window with a URL similar to the one in the following example will display the content of the selected document:

```
https://<<ICCServerName>>:11443/DocViewer/ViewItem.do?
p8ObjectStore={759E9C6F-2B90-432F-A12F-A82949545747}&reposID={CD25741B-C3EA-4A26-90E3-F6E9B1373012}
```

Where *<ICCServerName>* is the name of the Content Collector Web Application Server.

If any redirection errors occur, make sure that the redirection .jsp file that you created is located in the proper directory. If the URL of the browser window does not correspond to what you configured in the redirection file, check the MIME type mappings in the `content_redir.properties` file and correct them if necessary.

For further information about content redirection see the section about content redirection properties in the IBM FileNet P8 information center.

Tip: On the first redirection to the Document Viewer, users are prompted to log on to the repository. To avoid this, you can include the user ID and password in the URL that you configure in the redirection file.

Related tasks

"Installing a self-signed certificate for server authentication" on page 104

"Adding collections" on page 694

IBM Content Collector Outlook Extension

IBM Content Collector Outlook Extension is a Content Collector component that enables you to use Content Collector archiving functionality in Microsoft Outlook.

The Outlook Extension enables you to mark messages for archiving, specify additional archiving information that is archived with a message, search for messages in the archive, restore messages, and temporarily restore archived content when a stub is opened.

The help system that is part of the component installation provides information on IBM Content Collector Outlook Extension.

If you selected to install the offline repository functionality when you installed IBM Content Collector Outlook Extension, the first time you open Outlook, you will be asked if you want to enable and configure the offline repository functionality.

Selecting **Yes** will open the window on which you can enable working with an offline repository and adjust offline repository settings.

No will leave the offline repository functionality disabled. To enable working with an offline repository from the Outlook client, click the IBM Content Collector tab on the Options dialog of your Outlook client and select **Enable offline repository**

If you select **Cancel**, you will be prompted again the next time you start Outlook.

Installing a self-signed certificate for server authentication

IBM Content Collector uses the HTTPS protocol for network communication. This requires a certificate for authentication. If you did not replace the default certificates that were created by the Web Application Server with certificates signed

by a trusted certificate authority, email clients must accept a certificate for authentication when they contact the IBM Content Collector server.

The certificate testifies that a server really is the server that a client wants to connect to.

Important: Never use self-signed certificates in a production environment. Self-signed certificates can only be trusted directly while certificates signed by a trusted certificate authority have transitive trust. Transitive trust means that if the clients trust parts of the certificate chain the trust relationship is extended automatically to all parts of the certificate chain. In this case, email clients do not need to accept a certificate for authentication when they contact the IBM Content Collector server.

To install a certificate:

1. Open a web browser on the email client workstation.
2. In the address field of the browser, type the following URL:

   `http://<ICC Web Server>:11080/AFUWeb/init`

   where <ICC Web Server> is the name of the IBM Content Collector web server. A website with an error message is displayed. The message varies depending on the web browser that you use. Follow the instructions for your web browser.
3. For Internet Explorer, the following message is displayed:

   `Your connection is not secured with HTTPS.`

   After a few seconds, you are routed to another website with the following options:

   - **Click here to close this webpage.**
   - **Continue to this website (not recommended).**

   a. Click **Continue to this website (not recommended).** A reddish background is shown in the address bar your browser. Next to it on the right, the message `Certificate Error` is shown against the same background. The page in the browser windows shows the following message:

      `Your connection is secured with HTTPS.`

   b. Click **Certificate Error** in the small field next to the address bar.

   c. On the **Certification Path** tab of the Certificate window, select the root certificate and click **View Certificate**.

   d. Click **Install Certificate** and click **Yes** when you are asked if you want to install the certificate.

   e. Press F5 to refresh the page. The certificate error in the address bar disappears and a padlock icon is displayed, indicating the use of HTTPS. The main window of the browser still shows the message:

      `Your connection is secured with HTTPS.`

      Tip: If you have problems installing the certificate when you use Internet Explorer 7, check the setting of the **Do not save encrypted pages to disk** option. To do so, select **Tools** → **Internet Options**. Go to the security settings on the **Advanced** tab and make sure that the option is not selected.

   f. Restart your web browser.
4. To verify that the certificate was properly installed proceed as follows:

   a. Open a command prompt and enter the command **mmc** The Microsoft Management Console opens.

   b. Select **File** → **Add/Remove Snap-in**.

c. Click certificates. The root certificate for IBM Content Collector was added to the list of certificates.

5. For Mozilla Firefox, you are routed to another website where the following message is displayed:

   `This Connection is Untrusted`

   a. From the options that are available on this website, select **I Understand the Risks**.

   b. Click **Add an Exception** and proceed with accepting the certificate.

   The page in the browser window shows the message:

   `Your connection is secured with HTTPS.`

Important: Never use self-signed certificates in a production environment.

Replace the self-signed certificates by certificates issued by a trusted authority before running IBM Content Collector in a production environment.

Archiving message content

This section describes how to select and archive message content from Microsoft Outlook.

The main window of Microsoft Outlook must be open and display your mailbox.

1. Select one or more messages in the right pane of the Outlook main window.
2. Click on the toolbar or **IBM Content Collector → Mark for archiving** on the main menu.

   Important:

   - The archiving button and the menu choice are only available if archiving was selected at the time the IBM Content Collector Outlook Extension was installed.
   - The button and the **Mark for archiving** menu choice are grayed out if you select messages of the following type or messages residing in one of the following folders:
     - **Outbox**. Messages in the Outbox folder cannot be archived at all because this folder is a temporary folder. Messages are deleted from this folder after they were sent. Copies can be found in the **Sent Items** folder, which is included in the archiving process. Excluding messages from the **Outbox** folder prevents duplicate archiving.
     - **Sync Issues**. Messages in this folder contain conflict resolution information and cannot be archived at all because this might delete critical problem-solving information that the user needs to try to solve the conflict.
     - **Junk E-mail**. By default, messages in this folder are not archived because it is assumed that their content is not worth archiving.
     - IBM Content Collector search result folders. These are folders that you specify as the target folder for search and retrieval operations. Messages in these folders cannot be archived at all.

The selected messages are archived immediately.

If you selected **Display report message box** on the Advanced Options dialog, a report is displayed with the following information when the archiving process has finished:

- The number of selected messages.
- The number of messages that were successfully archived.
- The number of messages that could not be processed due to errors.

Restoring archived content

When you restore archived attachments or message content, the message stub is replaced with the original message body and attachments. If the automatic retrieve function is enabled, you can view the original message body of a stubbed message without manually restoring it.

The main window of Microsoft Outlook must be open.

If the automatic retrieve function is enabled, the content of stubbed message is retrieved from the archive and temporarily stored in your file system when you open the message by double-clicking it or by selecting **Reply**, **Reply All**, or **Forward**. If you enabled the preview pane, the original content will also be displayed there. You can change the flags, categories, the importance, or other properties as you like. The changes will be preserved in the stubbed message.

The temporary copy of the message will be deleted when you exit Microsoft Outlook. Note that automatic retrieve and display works for single messages only. You cannot select multiple messages.

The automatic retrieve functionality is only available on the client if IBM Content Collector Notes Client Extension is installed on the client and the functionality has been enabled by the system administrator in the IBM Content Collector Configuration Manager. Also, the client user must select **Temporarily retrieve message when opened** under Advanced options to enable the automatic retrieve functionality.

Note: If this functionality is enabled for very many clients, it can have a negative impact on the IBM Content Collector server performance.

To restore the original message body and the attachments:

1. Select one or more messaged stubs in the right pane of the Outlook main window. You can restore manually archived messages and messages that were archived automatically.
2. Click or **IBM Content Collector** → **Restore** on the main menu. The restore button and the menu choice are only available if the restoring was selected at the time the IBM Content Collector Outlook Extension was installed.

If you selected to restore a message that is too big to be restored on the server side within 210 seconds, a warning message will be displayed. However, the operation will complete.

It is possible that you cannot see the full content of a restored message immediately after a successful restore operation. You can either wait for the cache to expire, or you can switch to another message and then switch back to refresh the cache.

If you selected **Display report message box** on the Advanced Options dialog, a report is displayed with the following information when the restore process has finished:

- The number of selected messages
- The number of messages restored successfully
- The number of messages that could not be restored due to errors

Important: You should never manually restore a message if the automatic retrieve function is enabled or if you are using an offline repository. In this case, if you are working in cached mode, manually restoring a message might lead to a synchronization conflict.

Specifying additional archiving information

You can specify additional archiving information for messages in specific folders if this option was installed and was configured using the Configuration Manager. This additional archiving information can be used for the further processing of messages.

Specifying additional archiving information is possible only if this option was selected at the time the IBM Content Collector Outlook Extension was installed.

To specify additional archiving information:

1. Select one or more messages and copy or move them to a folder in Microsoft Outlook that is set up for messages with additional archiving information. This will work only if offline repository support is enabled.
2. Select a folder in Microsoft Outlook and click or **IBM Content Collector** → **Specify additional archiving information** on the main menu. The button and the **Specify additional archiving information** menu choice are grayed out if you select an empty folder or a folder that contains only messages to which you cannot add archiving information.

A window opens where you can select one or more messages and then specify additional archiving information for those messages. The messages for which you can specify additional archiving information are marked as unread. When you have completed the information for the messages, you can submit the messages for processing, which means that the next time the collector becomes active the messages will be archived and the additional information will be stored for further processing.

If you are working offline, the message is marked for archiving and the additional archiving information is stored locally. The additional information will be submitted for processing the next time when you are working online.

Flagging messages for stubbing

You can set a flag on selected messages so that content is removed from these messages during the next work cycle of the collector (“stubbing” or “restubbing”). This usually frees up space in your mailbox much more quickly because the stubbing activity is triggered immediately and thus independent of the stubbing timeout period defined using the Configuration Manager.

The mark for stubbing button and the menu choice are only available if stubbing was selected at the time the IBM Content Collector Outlook Extension was installed.

1. Select one or more messages in Microsoft Outlook.
2. Click or **IBM Content Collector → Mark for Stubbing** on the main menu.

The content of the messages flagged in this way is removed the next time that the collector becomes active. What parts of a message are removed and in which order depends on the stubbing life cycle settings that were made using the Configuration Manager.

If you selected **Display report message box** on the Advanced Options dialog, a report is displayed with the following information when the mark-for-stubbing process has finished:

- The number of selected messages
- The number of successfully marked messages
- The number of messages that could not be marked due to errors

Related concepts

"Message stubs" on page 796

Specifying configuration options

On this page, you can select to enable to work with an offline repository, determine controls that influence offline access, and select a number of additional options like enabling tracing.

1. To set advanced options like tracing for example, click the **Options** button in the Outlook Extension pane. Doing so opens the Advanced Options dialog. The Advanced Options dialog contains controls for tracing, process reporting, messaging, and the startup screen.
2. To enable offline archiving to a local repository (Outlook PST file), select **Enable offline repository**. This enables a submenu, which becomes available by clicking **IBM Content Collector → Offline Repository** on the main menu. The submenu offers you the following choices:
   - **Initialize**
   - **Resume**
   - **Pause**

   In addition, the **Offline Repository Settings** button next to the checkbox is enabled. Clicking this button gives you access to a set of additional settings for handling offline archiving.
3. Click **Apply** to save your settings or **OK** to save the settings and close the Options property sheet.

Related concepts

"Enabling offline repositories in Microsoft Exchange" on page 796

Related tasks

"Creating an offline repository" on page 797

"Starting and stopping periodic updates" on page 798

Setting advanced options for offline archiving

In this window, you can specify for how long messages are kept in an offline repository. In addition, you can set a maximum size for messages to be archived,

the time interval for periodic updates, the location of the offline repository, and whether the content of a stubbed message is automatically retrieved when the message is opened.

1. To change the location of your offline repository, enter another path for the offline repository in the **Offline repository path** field. By default, offline repositories are created in this directory:

   C:\Documents and Settings\<user name>\<product name>where <user name> is the Outlook logon user name and <product name> IBM Content Collector Outlook Extension.

   If you change the offline PST file location, the messages in mailbox will be initialized and the previous offline PST file will be deleted the next time Outlook is started.
2. The current size of the offline repository is shown. To change the maximum size limit for the offline repository, enter a number of megabytes in the **Limit the offline repository size to** field. By default, this limit is set to 2000 megabytes (2 GB). When the limit is reached, the oldest messages in the repository are deleted to make room for new ones. Setting such a limit is highly recommended because it ensures that the offline repository is kept within the limits imposed by Microsoft Outlook. If the recommended maximum is exceeded, the repository file might be damaged, thus making it unusable. The recommended maximum varies with respect to the Outlook version that you use. Check Microsoft information sources for the recommended maximum and set the limit accordingly.
3. Select which messages to delete first when the repository limit is reached. Either select to delete the largest messages first or the oldest messages first.
4. To set a size limit for messages to be copied to the offline repository, select **Do not store messages larger than** and enter a number of megabytes in the corresponding field. Messages with a size greater than the specified value will not be copied to the offline repository. If the check box is not selected, this means that there is no limit; messages of all sizes are copied to the offline repository.
5. To change the retention period for messages in your offline repository, select **Do not store messages older than** and enter a number of months in the corresponding field. The number specifies for how many months from the date of entry a message will be kept in the offline repository before it can be overwritten by newer messages. This helps to keep the overall size of your offline repository under control. If the check box is not selected, the messages are kept in the offline repository indefinitely.
6. To set a time interval for periodic updates, select **Run the periodic update every** and enter a number of minutes in the corresponding field. The number specifies after which time the offline repository is synchronized with the central backend archive. For example, a value of 40 specifies that an update process is automatically started every 40 minutes. Every 40 minutes, a background program checks your local mailbox to see if messages have been archived in the central backend archive. If messages were archived, the same messages will also be copied to your offline repository.

   Clearing the check box disables periodic updates, meaning that the offline repository is synchronized with the central archive only if the user starts the process manually by clicking **IBM Content Collector** → **Offline Repository** → **Resume**.

   This menu choice becomes available after **Enable offline repository** has been selected on the Options page and Outlook has been restarted. Otherwise the menu choice is not visible or grayed out.

7. Enable or disable the automatic retrieve of message content when a stubbed message is opened. If a message has not yet been added to the offline repository and you disabled this function, you can see only the stubbed content of a message when you open it. If you enabled this function, the content of a stubbed message is automatically retrieved when you open it and stored in the local file system as an MSG file. The message will be added to the offline repository when the update process runs.

   This option is available when both offline repository and automatic retrieve are enabled on the client.
8. Click **OK**. You return to the IBM Content Collector page of the Options property sheet.

Related concepts

"Enabling offline repositories in Microsoft Exchange" on page 796

Related tasks

"Creating an offline repository" on page 797

"Starting and stopping periodic updates" on page 798

Specifying advanced settings

This topic describes how you can switch on additional settings in IBM Content Collector Outlook Extension. This includes tracing and process reporting.

1. To switch tracing on, select **Enable tracing to file**. To switch tracing off, clear the box. Tracing allows you to collect information for error recovery. The trace information is written to a file on the local workstation, that is, the workstation on which the Outlook client is installed. You are asked to submit the trace file when you address a problem to the IBM Content Collector support team. The path and the name of the trace file are displayed under the check box.
2. To display a report after an archiving or restore process, select **Display report message box**. The report box tells a user how many archiving or restore jobs were created successfully, how many jobs could not be created, and so on.
3. To display a message that warns you of a possible loss of changes caused by overwriting existing content during a restore, select **Display a warning when restoring a message**.
4. To show a startup screen with the product name and version at each time you start Outlook, select **Show product information screen on startup**.
5. To enable automatic retrieve and display of stubbed messages, select **Temporarily retrieve message when opened**.
6. Click **OK**. You return to the IBM Content Collector page of the Options property sheet.

Searching in archives

You can search for archived messages using the IBM Content Collector search function from the Outlook main window.

To launch the search mask, click on the toolbar or **IBM Content Collector** → **Search** on the main menu.

Note: The search button and the menu choice are only available if the Search component was selected at the time IBM Content Collector Outlook Extension was installed.
The search application opens in your web browser.

Message stubs

Message stubs are the leftovers of an archiving operation.

When the content of a message has been archived, a stub is left in your Outlook system. The stub is a message that contains just the header or the header and the message body. The header consists of message items such as the Sender, Subject, and Receiver properties. The stub is a reference to the archived content, and serves as a handle for restoring it. A stub message contains one or more hyperlinks to archived content. Clicking a link shows the archived content in a web browser.

Note that the browser retrieves the archived message and attempts starting the associated client without any logon authentication. If the security settings of the browser are set to not allow this, the archived content will not be displayed.

Related tasks

"Flagging messages for stubbing" on page 792

Prerequisite for the use of the IBM Content Collector Client Components

A user's default inbox must be part of the Exchange server mailbox. See your Microsoft Outlook documentation for instructions on how to move a default inbox if it is not part of the Exchange server mailbox.

Enabling offline repositories in Microsoft Exchange

With IBM Content Collector, you can work with archived documents in your Microsoft Exchange databases even if you are offline.

Offline repositories are stored in personal folders (PST files). To set up and manage these offline repositories, various functions have been added to the IBM Content Collector Client Components for Outlook.

- To be able to use offline archiving on an Outlook workstation, you must select the feature at the time you install the IBM Content Collector Outlook Extension on the workstation.
- You enable working with an offline repository by clicking the IBM Content Collector tab on the Options dialog of your Outlook client and selecting **Enable offline repository**.
- Define your offline repository settings by clicking **Advanced** where you can specify the following:
  - The location of the offline repository on your local machine.
  - A maximum size limit for the offline repository. This prevents the repository from growing too big to be opened. When the limit is reached, either the oldest messages or the largest messages are deleted from the offline repository to make room for incoming messages.
  - A time interval that specifies when the update process is to run. This helps to keep the offline repository in synch with the central backend archive. The specified interval initiates updates to the offline repository within the stubbing timeframe (the stubbing lifecycle) defined for the central backend repository using the Configuration Manager.
  - An explicit size limit for messages to be stored in the offline repository. Messages that are larger will not be copied to the offline repository.
  - How long messages are kept in the offline repository before they are "overwritten" by new incoming messages. This allows you to control the size of the offline repository.

- You create a local offline repository by selecting **IBM Content Collector → Offline Repository Settings → Initialize** on the Outlook toolbar. You only need to initialize once. If no local repository exists, a new one will be created. If a local repository does exist, it will be cleared.
- You can update the offline repository immediately, that is, start an update process irrespective of the time interval that you set when you defined the offline repository settings by selecting **IBM Content Collector → Offline Repository Settings → Resume**. You can also interrupt the update process at any time by selecting **IBM Content Collector → Offline Repository Settings → Pause**.

Note:

- When you are working offline, only **Archive**, **Mark for Stubbing**, and **Specify Additional Archiving Information** can be selected for a message on the toolbar or on the main menu when **IBM Content Collector** is selected.
- If you changed the content of a message in the offline repository, and want the original content again in your offline repository, restore the original message in the mailbox. This also copies the restored original message to the offline repository and overwrites the changed copy. Opening the restored message in the offline repository opens the original message in the mailbox.

Related tasks

"Setting advanced options for offline archiving" on page 793

"Specifying configuration options" on page 793

"Creating an offline repository"

"Starting and stopping periodic updates" on page 798

Creating an offline repository

This section explains how to create an offline repository, and how to start the initial synchronization with the central backend archive.

The choice to work with an offline repository is only available if offline repository was selected at the time the IBM Content Collector Outlook Extension was installed.

In the Outlook main window, click **IBM Content Collector → Offline Repository → Initialize**. A warning is displayed if a local repository already exists because initializing will clear the repository.

A personal folder (PST file) is created to serve as the offline repository. The personal folder appears in the folders view of the Outlook client.

Important:

- The initialization process might take considerable time. If for any reason you wish to interrupt the process, click **IBM Content Collector → Offline Repository → Pause**. You can restart the initialization at any time by selecting **IBM Content Collector → Offline Repository → Resume**.
- The offline repository is built by copying all messages to the local offline repository at the time when they are archived on the server repository. First complete copies of these messages are added to the offline repository, and then later these messages are stubbed based on the stubbing lifecycle defined for archived messages using the Configuration Manager.

- To find out if a message has been archived in the central archive, it is required that a message stub belonging to the original message still exists in your mailbox. If the stub has been deleted, an already archived message cannot be retrieved.

Related concepts

"Enabling offline repositories in Microsoft Exchange" on page 796

Related tasks

"Setting advanced options for offline archiving" on page 793

"Specifying configuration options" on page 793

"Starting and stopping periodic updates"

Starting and stopping periodic updates

If you use a local offline archive, run periodic update processes in the background to keep your local archive in synch with the central backend archive.

Before you continue, make sure that you have enabled the offline archiving functions by selecting **Enable offline repository** on the IBM Content Collector page of the Outlook Options property sheet. From this page, you can also access an advanced options dialog, which allows you to specify further offline archiving options including the time schedule for automatic periodic updates of your local archive.

- To run an update process immediately, click **IBM Content Collector → Offline Repository → Resume**. This option is available only if the background process is not running. A background program will check if further messages have been archived in the central backend archive since the last update. If that is the case, the corresponding messages are stored in your local offline archive before the stubbing process sets in.
- If for any reason you wish to discontinue the update process, click **IBM Content Collector → Offline Repository → Pause**. This option is available only if the background process is running. The currently running update process will be interrupted.

Related concepts

"Enabling offline repositories in Microsoft Exchange" on page 796

Related tasks

"Setting advanced options for offline archiving" on page 793

"Specifying configuration options" on page 793

"Creating an offline repository" on page 797

Part 9. Troubleshooting and support

© Copyright IBM Corp. 2008, 2010

Troubleshooting Content Collector

Check the following topics if you encounter problems in IBM Content Collector.

If your problem is not mentioned here, try to search live, web-based support resources.

Before you contact IBM support with your problem, read MustGather: Read first for all IBM InfoSphere Content Collector Issues on ibm.com.

Retrieving version information

Use the Component Version Report tool to retrieve component version information.

To create a report that contains the version information for all installed IBM Content Collector components:

1. In the Configuration Manager, select **Help → Component Version Report**.
2. Click **Retrieve** to retrieve the version information. This can take several minutes.
3. When the report is available, click **Report** to open the report in the default text editor.

Troubleshooting installation

These topics can help you solve common installation problems such as third-party application errors, failure to create the configuration database, and difficulty installing on multiple servers.

Troubleshooting scale out mode

Options available to troubleshoot a scale out configuration are described in this topic. Keep in mind when troubleshooting node issues that logs on each individual node must be checked, as IBM Content Collector logs are not centrally stored.

To troubleshoot scale out configuration, take the following steps:

- Verify whether each extension node has database connectivity by starting the IBM Content Collector Configuration Manager on each node. If the Configuration Manager starts successfully and the configuration settings and task routes can be viewed, database connectivity has been successfully established. Note that on each node, only one instance of the Configuration Manager runs in write mode.
- If the permissions on the account that are used to the run the Task Route Engine service are insufficient, warning messages will appear in the task route log indicating this.
- Each node must be able to resolve the name of any other node in the scale out scenario that it is attempting to communicate with. To ensure a specific node is able to resolve the name of another node, use the “ping” command with the name of the machine representing the node in question.
- On the primary node, set the Task Route service log level to trace2 as follows: Open the Configuration Manager; from the **Tools** menu, select **Task Route Service Configuration**, and in the **Log Settings** section, set **Log level**. Once set,

© Copyright IBM Corp. 2008, 2010

check logs for “forwarding node” entries which indicate to which nodes work is being distributed. The following is an example of a forwarding node log entry:

```
Forwarded processing of entity 207@151@147@32@B31BEB6
71DD90C10882574C700593123-1@-1@102@IMPL==DOMINO;
TYPE==DATABASE;SERVERID==x.x.x.x;STOREID==mailjrn.nsf;
EMAILADDR==null;FILEPATH==null36@2f781f4d-e321-4e37-
a046-cb951e8d41b72@{}7@INITIAL to secondary node
[server name] ibm::ctms::connector::
CollectorSubmitStub::forwardTaskOutput
```

`Secondary node [server name]` refers to the name of the extension node to which work is being distributed.

- If there are no "forwarding node" entries in the logs, check the datastore synchronization interval. Open the Configuration Manager and select **Task Route Service Configuration** from the **Tools** menu. If **Datastore synchronization interval** is set to 0, the configuration database is never synchronized with the extension node. For scale out, you must set it to a value greater than 0.
- In the database, query the database table TROUTE_OPMODE to verify that a record exists in the table for the primary node and each extension node. Each record in the table contains the following fields:

Field	Description
trom_machine_name	Name of the server that is registered as an extension node.
trom_mode	Indicates whether the server is a primary or secondary node (0 = primary, 1 = secondary)
trom_heartbeat	The last time the server checked in, in order to indicate it is still available to perform work.
trom_leasetime	The lease time for the IP connection.
trom_reg_date	The date the server was registered.

- To assist with troubleshooting, the following flags may be passed to the Task Route Service as in this example: `C:\Program Files\IBM\ContentCollector\ctms\TaskRoutingService.exe` -m=*machinename*, where *machinename* is the machine name of the node. The table below contains a complete list of flags available:

Task Route Service Flag	Description
-r=r	Adds or updates a node's registration as an extension node.
-r=u	Unregisters a node as an extension node.
-n=machine1;machine2	For a primary node, overrides extension node registrations and forces the primary node to use the machines specified as extension nodes.
-m=machinename	Forces a node to register with the specified machine name instead of the 'Windows Computer Name'. For example, the IP address could be used to work around DNS issues.

The installation of the Web applications failed

If the install command for the Web Application did not complete successfully, you must check several settings, run the uninstall command, and rerun the install command.

Proceed as follows:

1. Check the output file that was written by the install script for error messages. This file is named `afu_ewas_install.txt`. It is located in the home directory of your WebSphere Application Server installation.
2. Depending on your repository, check the following settings:
   - If your repository is Content Manager 8.x, the system environment variable IBMCMROOT must point to the installation directory of the Content Manager client. The default value is:`C:\Progam Files\IBM\db2cmv8`
   - If your repository is IBM FileNet P8, FNCEPATH must be the path to the FileNet P8 installation directory that contains the subdirectory `\lib`. Check the setting of the FNCEPATH parameter in the `afuEnv.bat` file to ensure that it points to the correct directory.
3. Enter the following command:

   ```
   afu_ewas_uninstall.bat
   ```
4. Delete the `AFUWeb` directory in `<WASHOME>\profiles\`.
5. Rerun the installation of the Web applications.

Related tasks

"Installing the Web applications" on page 97

Creating the Content Collector configuration database on remote server fails

You cannot create a default Content Collector configuration database with DB2 Enterprise Server Version 9.5 on a remote server if the available DB2 fix packs have not been installed on the Content Collector server machine.

If the available DB2 fix packs for DB2 Enterprise Server Version 9.5 are not installed on the Content Collector server machine, the following error might occur:

```
DB21018E  A system error occurred. The command line processor could not
continue processing
```

To avoid this error:

Upgrade the DB2 runtime environment (server or client) on the Content Collector server machine to at least fix pack 1 of DB2 Version 9.5.

Using a different DB2 version for the configuration database on a remote server to the version installed on the Content Collector server

If you use DB2 Version 9.1 on a remote database server and you want to create a default configuration database, use the same DB2 version on the Content Collector server machine.

Using different versions might lead to a processing error. For example, when DB2 Version 9.1 is installed on the remote database server and DB2 Version 9.5 is installed on the Content Collector server machine, the following error may occur:

```
SQL1088W  The database was created, but an error occurred while
binding the utilities. The utilities are not bound to the database.
```

To correct this error and set up the configuration database correctly, follow these steps:

1. Copy the bind directory ([SQLLIB]\bnd) from the DB2 9.1 machine to the DB2 9.5 machine.
2. Run the following DB2 statements on the DB2 9.5 machine: <*path*> denotes the path to the bind directory that you copied from the DB2 9.1 machine.
   a. `db2 catalog tcpip node <nodename> remote <hostname> server <dbport>`
   b. `db2 catalog database <dbname> as <dbname> at node <nodename> authentication server`
   c. `db2 CONNECT TO <dbname> user <db2_adminuser_name> using <db2_adminuser_password>`
   d. `db2 BIND <path>\db2schema.bnd BLOCKING ALL GRANT PUBLIC SQLERROR CONTINUE`
   e. `db2 BIND <path>\@db2ubind.lst BLOCKING ALL GRANT PUBLIC ACTION REPLACE`
   f. `db2 BIND <path>\@db2cli.lst BLOCKING ALL GRANT PUBLIC ACTION ADD`
   g. If the user who ran the `db2 connect` statement is not the DB2 administrator run `db2 GRANT CREATETAB ON DATABASE TO <db2_connectuser_name>`.
   h. `db2 terminate`
   i. `db2 connect to dbname user <db2_connectuser_name> using <db2_connectuser_password>`
   j. `db2 -tvf "<IBMAFUROOT>\ctms\Database Scripts\db_ctms_400_db2.sql"`
   k. `db2 terminate`

The connection to the Oracle database fails

When you try to establish a connection to the Oracle database in the Configuration Manager, you might get the error `OleDbException "ORA-12638: Credential retrieval failed"`.

In this case, change the value of the SQLNET.AUTHENTICATION_SERVICES key in the Oracle configuration file `sqlnet.ora` as follows: change `SQLNET.AUTHENTICATION_SERVICES=(NTS)` to `SQLNET.AUTHENTICATION_SERVICES=(NONE)`.

The SQLNET.AUTHENTICATION_SERVICES key is used to enable one or more authentication services. NTS specifies that Oracle can use NT native authentication to authenticate users. The disadvantage of changing the value to NONE is that you can no longer log on as sysdba without a password.

Using DB2 Data Server Client Version 9.5 fix pack 1 or 2

If you use DB2 Data Server Client Version 9.5 fix pack 1 or fix pack 2 on the Content Collector server machine and DB2 Version 9.5 on a remote database server, and you want to create a default configuration database on the remote database, you must copy the missing bind file from the remote server with DB2 Version 9.5 to the server on which DB2 Data Server Client Version 9.5 fix pack 1 or fix pack 2 is installed.

If DB2 Data Server Client Version 9.5 fix pack 1 or fix pack 2 is installed on the Content Collector server machine and DB2 Version 9.5 is installed on the remote database server, the following error message might occur:

```
SQL1065W  The database was created, but an error occurred
on one or more bind files in list "db2ubind.lst".
The file(s) in the following position(s) in the bind
list file were not bound: "31".
```

To fix this problem:

Copy the file *SQLLIB*\bnd\db2ats_sps.bnd from the remote database server to the Content Collector server machine at *SQLLIB*\bnd.

Memory issues when running the initial configuration or the set-up tools

When running the initial configuration or the set-up tools, you might run up against memory problem because the heap size has not been set to a larger size.

If you are running the initial configuration or one of the configuration set-up tools and run into a memory problem like the following:

```
SEVERE <?xml version='1.0'?>
<DXLImporterLog>
<error id='263'>Insufficient memory.</error>
<error id='263'>Insufficient memory.</error>
<warning id='7031'>Import operation incomplete; 1 notes(s)
       imported successfull</warning>
<error>DXL importer operation failed</error>
```

extend the JVM maximum heap size:

Add the following parameters to the end of the file *ContentCollector-Install-dir*\Configuration\initialConfig\initialConfig.ini:

```
-vmargs
-Xmx<MaxHeapSizeInMB>m
```

For example, for a maximum heap size of 2 GB, add the following and make sure that each parameter is on its own line:

```
-vmargs
-Xmx2048m
```

The CommonStore server and the CSLD tasks fail to start

The CommonStore server (archpro) and the CommonStore for Lotus Domino (CSLD) tasks for legacy support fail to start if there is a conflict between libraries that are used by CommonStore for legacy support and libraries that are used by Content Manager OnDemand.

Symptoms

The CommonStore server (archpro) fails to start and reports an error similar to the following:

```
C:\Program Files\IBM\CSLD\server\instance01>archpro


        ****************************************************************
        * IBM CommonStore - Server 8.4.0.33                            *
        * Copyright IBM Corporation, 1997, 2007. All Rights Reserved.    *
        * Build 8.4.0.33,   Compiled at Jan 21 2009.                   *
```

```
          ****************************************************************

CSS0030I: ArchPro is using INI file 'C:\Program Files\IBM\CSLD\server\instance01
\archint.ini'.
CSS0910I: Trying to get a LUM Production License for IBM CommonStore for Lotus D
omino
CSS0929I:
          ****************************************************************
          *                 Got a Production License for                 *
          *               IBM CommonStore for Lotus Domino               *
          ****************************************************************
CSS0158I: ArchPro 1868 started on UNICODE Port 6707.
CSS0157I: ArchPro 1868 is waiting for external connections on fixed port 8012.
CSS9011I: Setting up master connection to CommonStore server on Socket[addr=loca
lhost/127.0.0.1,port=6707,localport=6747].
CSS9019E: Initialisation of Notes/Domino Dispatcher failed. Reason: <ESD9505E Th
e message could not be read from the underlying stream.>. Exiting...
```

The CommonStore for Lotus Domino tasks fail to start and report an error similar to the following:

```
C:\Program Files\IBM\CSLD\server\instance01>csld -s <servername> -n
<configdatabasename> -p <profilename> -i <notesinifile>
------------------------------------------------------------
  IBM CommonStore for Lotus Domino
  Version 8.4.0.37
  Copyright IBM Corporation, 1997, 2007.
  All Rights Reserved.
------------------------------------------------------------
CSS7402E: Notes Error. Cannot read configuration parameters. Reason: '........."

          Aborting task.
```

Causes

This problem can be caused by a conflict between the libraries that are provided by the CommonStore server (archpro) and the CommonStore for Lotus Domino (CSLD) tasks for legacy support and the libraries that are provided by Content Manager OnDemand. If the libraries that are provided by Content Manager OnDemand are used by CommonStore for legacy support, the CSLD tasks and the CommonStore server fail to start.

Resolving the problem

If possible, install IBM Content Collector legacy support and Content Manager OnDemand on different machines.

If IBM Content Collector legacy support and Content Manager OnDemand must be installed on the same machine, make sure that the *PATH* environment variable contains the directory where CommonStore is installed before the directory where Content Manager OnDemand is installed, so that IBM Content Collector legacy support uses the CommonStore libraries.

Troubleshooting configuration

These topics can help you solve common problems in configuring source systems, target repositories, task routes, and other components.

Troubleshooting source systems

These topics can help you solve common problems in configuring source systems.

User ID for the IBM Content Collector Email Connector service

If Lotus Domino is your email system, the IBM Content Collector Email Connector service must not be run under a local system account.The error AFUC0004E indicates that the settings for a service are incorrect.

Error message

```
AFUC0004E: COS runtime exception occurred:
1001501: SHAREDMEMORY_COULD_NOT_CREATE
(cos_8_0SHM32:AFUPWDMGR:AFU:xxxxxxxxxxxxxxxxx
Error as reported by Operating System: 5 (ERROR_ACCESS_DENIED)
[Origin: \cosroot\cos_base\cos_shared_memory.cpp:290]
```

Solution

For information on how to change the user account, click the link at the bottom of this topic.

Related tasks

"Changing the user account of a service" on page 213

Using a multilingual mail database

If you installed the Domino Language Pack on the Domino server and are using multilingual mail templates on the mail database, there is a limitation in that email can only be delivered to one Inbox folder at a time. Only users working with different locales set on their Notes clients will see the email for all of the different languages in their Inbox folders.

Symptoms

A user typically only sees the email that is delivered to the Inbox folder for the language configured on the Notes client. For example, if the mail database has design elements for German, French, and Dutch. but only German is set on the Notes client, the user cannot view the email that is delivered to the French or Dutch Inbox folders.

Resolving the problem

Contact IBM Support for the necessary steps to take to simultaneously set different locales on a Notes client.

Lotus Notes ID file cannot be validated

The Lotus Notes ID file specified for the Lotus Domino administrator or Lotus Notes repository user cannot be validated if the path to the folder of this file contains characters of the extended ASCII set, language-specific characters, or UTF-8 characters. If the ID file folder contains such characters, an error message is displayed. This is a known limitation.

Make sure that the ID files of the Lotus Domino administrator or the Lotus Notes repository user are located in a folder path that contains only characters from the basic 128-character ASCII code page. Additionally, ensure that the following Content Collector path values are listed in the PATH variable. For example, <*ICC_INSTALL_DIR*>\bin;<*ICC_INSTALL_DIR*>\lib;<*ICC_INSTALL_DIR*>\ctms where <*ICC_INSTALL_DIR*> is C:\Program Files\IBM\ContentCollector.

Duplicate attachments in Lotus Notes documents are not removed

Duplicate attachments in a Lotus Notes document are not removed when the document is stubbed

Symptoms

If a Lotus Notes document contains several attachments with the same display name (but different internal names), the duplicate attachments are not removed when the document is stubbed. The attachments remain in the document and are not deleted. The document is still intact after the stubbing, so the attachments are still available as part of the document.

Causes

This is a limitation of theLotus Domino Server Runtime before version 8.5.1.

Resolving the problem

To resolve this issue, upgrade the Lotus Domino Server Runtime on the system that runsIBM Content Collector Server to Lotus Domino Server Runtime 8.5.1 or later.

To fix documents with duplicate attachments that were stubbed incorrectly, restore them after you upgrade to Lotus Domino Server Runtime 8.5.1. They will be stubbed correctly after this.

Only one process can work with a PST file at a time

A PST file can only be accessed by one process at a time.

If a PST file is accessed by more than one process, the file can either not be archived or it cannot be viewed or edited by a client user.

Therefore, consider the following when dealing with PST files:

- If you want to archive a PST file, it must not be accessed by an Outlook client
- To enable client users to view or edit the PST file from an Outlook client, the archiving collector scheduler is not allowed to be set to Always. If it is set to Always, the file will be in constant access by Content Collector, and client users will not be able to view or edit it.

Microsoft Exchange envelope journal messages are not archived

Archiving of envelope messages with attachments in Outlook Message Format is not supported.

Symptoms

When you try to archive messages with attachments in Outlook Message Format (.msg) from a Microsoft Exchange envelope journaling mailbox, the messages are not archived. The trace log contains the following line:

```
SEVERE  AFUM0088E: An error occurred: Invalid attachment method in envelope
journal message (expected is 5): '1'
```

Additionally, the trace log contains the entry:

```
ICC does not support Envelope Journal messages with "Outlook Message Format
(.msg) attachments.
```

Causes

Envelope messages contain the original message as attachment. This attachment can be in Exchange MAPI Message Format or in Outlook Message Format. IBM Content Collector supports Exchange MAPI Message Format, but not Outlook Message Format. Envelope messages with attachments in Outlook Message Format cannot be archived.

Resolving the problem

To archive from envelope journaling messages, ensure that attachments to the envelope messages are in Exchange MAPI Message Format.

Changing the timeout if the Outlook Extension cannot be loaded

If Outlook Extension cannot be loaded when a client user starts Microsoft Outlook, Outlook Extension probably did not have enough time to read configuration data from the IBM Content Collector Configuration Web Service. You can set a timeout value in the `afuconfig.xml` file to overcome this problem.

Workaround

1. On each affected client workstation, open the `afuconfig.xml` file in an editor. The file is located in the file path `IBM\ContentCollector_OutlookExtension` in the client user's local application data directory. The local application data directory is a standard Windows directory at `%USERPROFILE%\Local Settings\Application Data`.
2. Locate the following section:

```
<SymphonyAddin>
 <SymConfigServer>
   <ServerName>9.123.120.229</ServerName>
   <ServerPort>11443</ServerPort>
   <Protocol>HTTPS</Protocol>
   <NumRetries> 1</NumRetries>
   <Timeout> 3</Timeout>
```

3. Change the values of the following parameters to solve the problem:

   <NumRetries>1</NumRetries>
   The maximum number of times that the Outlook Extension can attempt to read configuration data from the Configuration Web Service. To increase the likelihood of a successful read operation, replace the number between the start and end tags with a higher value to allow more attempts. For example:

   ```
   <NumRetries>3</NumRetries>
   ```

   <Timeout>3</Timeout>
   The current timeout in seconds. To give the Outlook Extension more time to read configuration data, set this to a higher value, that is, replace the number between the start and end tags with a higher value, for example:

   ```
   <Timeout>10</Timeout>
   ```

4. Save the changes to the `afuconfig.xml` file.
5. Restart the Configuration Web Service.

Starting Outlook can be very slow if you are not connected to the network

Each time you start Outlook, the Outlook client attempts to connect to the Content Collector Configuration Web Service to retrieve configuration data it requires on the client workstation.

The default client server connection timeout is set to 3 seconds. Connecting to the server is attempted twice, which means that 6 seconds might elapse before Outlook is started even if you have chosen to work offline.

To avoid having to wait for Outlook to start if you are working in offline mode, change the timeout value in the Outlook Extension configuration file directly after you installed IBM Content Collector Outlook Extension on the client workstation and before you start Outlook the first time.

Set this value so that it is long enough to enable a connection to your server in case you change to connected mode (dependent on your connection speed), yet short enough to avoid long waiting times before Outlook starts.

If you have already started Outlook, also change the setting in the your copy of the afuconfig.xml file located in the application data directory under `IBM\ContentCollector_OutlookExtension`.

For example, for Windows XP at `C:\Documents and Settings\`*`user`*`\Local Settings\Application Data\IBM\Content Collector_OutlookExtension` and for Vista at `C:\Users\`*`user`*`\AppData\Local\IBM\ContentCollector_OutlookExtension`.

To adjust the timeout value:

Change the value between the timeout tag (<Timeout>) in the configuration file afuconfig.xml in the Outlook Extension installation directory (default directory: `C:\Program Files\IBM\ContentCollector_OutlookExtension`).

Running Content Collector Outlook Extension on Citrix

On Citrix, temporary message files cannot be seen by other Content Collector Outlook Extension users in the same environment.

Symptoms

The temporary message files that are generated when Content Collector Outlook Extension is deployed on one server in a Citrix environment are only visible to the Citrix administrator. These files cannot be seen by other users using Content Collector Outlook Extension in the same environment.

Documents archived using CommonStore for Exchange Server cannot be restored

Sometimes documents archived using CommonStore for Exchange Server cannot be restored because the job folder name is invalid or does not exist.

If the error message:

```
The selected messages could not be restored
```

is displayed when you try to restore documents that were archived using CommonStore for Exchange Server and the systemout.log file of the Web Application contains the entry:

```
WebContainer SEVERE AFUM0007E: An error
occurred: MAPI call failed, return code '0x800B0001'
```

check that the job folder path in the legacy configuration is valid and that the folder exists.

After restoring legacy documents, these are not displayed in Outlook

If documents have been restored from an email repository that was filled by CommonStore for Exchange Server, these documents are listed in the inbox or in other Outlook folders, but their content is not displayed after a user clicks the **View** icon. This is a known limitation.

Workaround

Users need to close and then reopen Outlook. After reopening, the **View** button will work.

Not all archived messages are copied to a new offline repository

If you are using Windows Vista and Outlook 2007 and want to change your offline repository, initializing this new offline repository might not work as expected as all archived messages might not be copied to this new offline repository.

To prevent this from happening, change the profile to run in Cached Exchange Mode.

Microsoft Exchange: Offline repository support stops working after 5000 documents have been retrieved or copied

If a large amount of documents must be copied to the offline repository, offline repository support stops working.

Symptoms

If offline repository support is enabled for Microsoft Exchange, Content Collector Outlook Extension retrieves all stubbed messages from the archive and copies them into the offline repository. If there is a large amount of documents and Cached Exchange Mode is enabled, offline repository support stops working after 5000 documents have been retrieved or copied. If Cached Exchange Mode is not enabled, Microsoft Outlook users cannot deal with a large number of messages in one batch. This is a limitation for Microsoft Exchange Server.

Resolving the problem

If Cached Exchange Mode is enabled, users must restart Microsoft Outlook so that offline repository support works again.

If Cached Exchange Mode is not enabled, change the maximum number of resources that a MAPI client can use at the same time. Follow the instructions in article 830829 on the Microsoft support page to increase the value for the object type objtMessage.

The Email Connector for Microsoft Exchange cannot connect to the Active Directory

Symptoms

The Email Connector for Microsoft Exchange fails to connect to the Active Directory. It reports the following error message:

```
AFUP0010E Could not create global session.
Error message: com.ibm.afu.mailinterface.MailException:
javax.naming.AuthenticationException:
```

```
[LDAP: error code 49 - 80090308: LdapErr: DSID-0C090334,
comment: AcceptSecurityContext error, data 531, vece ]
[com.ibm.afu.mailconnector.MailConnectorService initMailinterface]
[pool-2-thread-1 1]
```

Causes

The domain controller that the Email Connector connects to in order to resolve names and groups reports error 531, which means that the user account of the IBM Content Collector Email Connector is not allowed to log on to the domain controller.

Diagnosing the problem

Check the information for the user account of the IBM Content Collector Email Connector in the Active Directory. If the information does not list the domain controller that the IBM Content Collector Email Connector connects to as user workstation, the user is not allowed to log on to the domain controller.

Resolving the problem

Resolve the problem in one of the following ways:

- Grant the user account of the IBM Content Collector Email Connector permission to access all required domain controllers.
- Remove the logon restriction from the domain controller.

SharePoint connector validation fails

Validation fails when creating or editing a connection to Microsoft SharePoint.

Symptoms

When you click **Validate** during the initial configuration or editing of a connection to Microsoft SharePoint, you receive one of several messages that validation failed.

Causes

Validation failures can result from several causes, most commonly an invalid site address or credentials.

Resolving the problem

Verify and change as needed:

Credentials:

- User name
- Password
- Domain
- Domain: fully qualified, containing the full path, including subdomain?

Site address:

- Protocol: http or https
- Server name
- Port, if needed

- SharePoint path, including site prefix and subsite, if needed; for example, `http://spsite/sites/sitecoll2`
- SharePoint alternate access mapping might require a different address
- SharePoint load balancer might require a different address
- Is there a firewall?
- Cannot include library or list names
- Specifying the SharePoint Central Administration site collects from that site, not from all site collections

SharePoint column names display incorrectly

The Content Collector for Microsoft SharePoint column names display internal names.

Symptoms

The Content Collector for Microsoft SharePoint column names display their internal names rather than their display names:

- *ICCSPMigrated* or *$Resources:ICCSPMigrated* rather than *Migrated*
- *ICCSPMigratedInformation* or *$Resources:ICCSPMigratedInformation* rather than *Migrated Information*

Causes

The two most likely causes are:

- Uninstalling Content Collector for Microsoft SharePoint deleted the localization files from the SharePoint Web Front End (WFE) server, or
- The localization files are missing or no longer contain the localization resources

Resolving the problem

To restore the correct column display names, reinstall Content Collector for Microsoft SharePoint on one SharePoint Web Front End server in the farm.

Troubleshooting target repositories

These topics can help you solve common problems in configuring target repositories.

Content Manager documents persist despite post-archiving failure

When a source-to-IBM Content Manager post-archiving task fails on one or more documents, IBM Content Manager does not delete the documents that the document creation task created.

Symptoms

The audit log (if configured) shows post-archiving errors, and duplicate documents appear in the repository each time collection occurs.

Causes

Any failure during post-archiving can cause the problem.

Resolving the problem

Follow these steps:

1. Correct the post-archiving error or change the post-archiving option.
2. Use the audit log or other method to identify the incorrect documents.
3. Manually delete the documents.

When restarting the Content Collector server the connection to the FileNet P8 object is lost

After rebooting the IBM Content Collector server, the FileNet P8 connector may lose its connection to the FileNet P8 object the first time the Configuration Manager is started.

There are two workarounds:

- In the Configuration Manager, select the FileNet P8 connector, retrieve the object store, and save it. This must be done each time the Content Collector server is rebooted.
- Change the timeout for the FileNet P8 object store from 10 seconds to 60 seconds in the *ContentCollector-install-dir*\ctms\p8Config.xml file. This only needs to be done once.

IBM FileNet P8 validation fails using HTTPS connection in Initial Configuration/Setup Tools

Symptoms

When configuring an IBM FileNet P8 repository for IBM Content Collector by using the Content Collector initial configuration or the set-up tools, the credentials that are used to access the FileNet P8 object store cannot be validated and the list of available object stores cannot be retrieved. This happens when the FileNet P8 Content Engine is accessed via an HTTPS connection.

Causes

The certificate for the FileNet P8 server does not exist in the keystore of the Content Collector Java Runtime Environment.

Diagnosing the problem

If the credentials that are used to access the FileNet P8 object store cannot be validated and the list of available object stores cannot be retrieved, check the log file *ICCInstallDir*\Configuration\initialConfig\log\AfuInitialConfigTrace.log (where *ICCInstallDir* is the directory of your Content Collector installation) on the Content Collector server for the following error message:

```
Caused by: java.security.cert.CertPathBuilderException:
PKIXCertPathBuilderImpl could not build a valid CertPath
```

Resolving the problem

Import the certificate for the FileNet P8 server into the keystore of the Content Collector Java Runtime Environment. Issue the following command, where *ICCInstallDir* is the directory of your Content Collector installation and *P8CertificatFile* is the certificate for the FileNet P8 server:

```
"ICCInstallDir\java\jre\bin\keytool.exe" -import -file P8CertificateFile
-keystore "ICCInstallDir\java\jre\lib\security\cacerts" -alias afup8
```

To retrieve the FileNet P8 certificate:

- On the Content Collector server, access the secure Content Engine URL (for example `https://P8CEHostname:9443/wsi/FNCEWS40MTOM/`) in a browser.
- Install the certificate in the browser.
- Export the certificate from the browser.

CONTENT_CA_READ_FAILED or CONTENT_PC_WRITE_FAILED errors appear frequently in the IBM FileNet P8 Connector logs

Symptoms

CONTENT_CA_READ_FAILED or CONTENT_PC_WRITE_FAILED errors appear frequently in the IBM FileNet P8 Connector logs.

Causes

The values for the total transaction lifetime timeout and the maximum transaction lifetime timeout properties and the values for the content stream time-to-live (TTL) that are set on the IBM FileNet Content Engine server might be too low.

Resolving the problem

Set the values of the transaction timeout properties for the Content Engine profile to 600 seconds.

Adapt the values of the content stream TTL properties:

1. Create a file named `FileNet.properties`.
2. Include these statements:

   ```
   Content.StreamIdleTTLSeconds=30
   Content.StreamTTLSeconds=30
   ```

3. Save the file to the directory that also contains the `FileNet` directory. Depending on the web application server that you use, this directory must be in the web application server profile or in the domain root directory, for example:

   Oracle WebLogic Server

   ```
   C:\bea\user_projects\domains\base_domain
   ```

   IBM WebSphere Application Server

   ```
   C:\IBM\WebSphere\AppServer\profiles\AppSrv01
   ```

4. Restart your web application server.

For more information, see the IBM FileNet Content Engine documentation and the documentation for the web application server that you use.

Troubleshooting components

These topics can help you solve common problems in configuring components, not including sources, targets, or task routes.

Monitoring folders with very long paths does not work

IBM Content Collector cannot monitor folders with very long paths.

Symptoms

If Content Collector is configured to monitor a folder with a very long folder path (several thousand characters), the respective collector does not run.

Collection source "Server" is not supported for Microsoft Exchange Server 2010

With IBM Content Collector, the administrator can specify a Microsoft Exchange Server as a source for automatic collection. In this case, the collector asks the Microsoft Exchange Server by means of Outlook MAPI for all mailboxes it hosts and then starts collecting from each mailbox except for journal mailboxes. This works for Microsoft Exchange Server 2003 and Microsoft Exchange Server 2007. Starting with version 2010, Microsoft Exchange Server no longer provides an implementation for that function for Outlook MAPI.

Symptoms

If you add a Microsoft Exchange Server 2010 as collection source for automatic collection, Content Collector does not collect email from the mailboxes on this server and the sysout log file contains the following error message:

```
AFUP0016E Could not resolve collection sources of collector CollectorName
Error message: AFUM0007E: An error occurred: MAPI call failed, return code
'MAPI_E_CALL_FAILED'
```

Resolving the problem

As workaround for collecting from all mailboxes on a server running Microsoft Exchange Server 2010, create a new Windows user group and add all users that have their mailboxes on this server. Specify the new user group as collection source for your collector. Note that this group must be maintained manually; this means that whenever a user mailbox (not journal mailbox) is added to or removed from the server, the corresponding user must be added to or removed from the Windows user group.

Task routes with a collector that includes the root folder / in the list of monitored folders do not work

If the filter of a collector for automatic archiving includes the root folder / in the list of monitored folders, the task routes do not work.

Symptoms

Task routes with an "EC Collect Email by Rules" collector do not work if they use a filter that includes the root folder / in the list of monitored folders.

Resolving the problem

If you want to include the root folder and thus all folders, provide an empty include list.

Searches for DBCS characters fail if DB2 Version 9.5 is used

You cannot search for double-byte characters or strings (for example, Chinese words or phrases) in documents in an IBM Content Manager Enterprise Edition 8.x repository if the database management system used by Content Manager is IBM DB2 Version 9.5.

Purpose

APAR IC58821 addresses this issue.

DBCS character search in the IBM Content Collector Web Applicationdoes not work for Content Manager 8 repository archive systems using an underlying DB2 Version 9.5.

APAR IC58821 will also address this issue. It is available for customer download from the following location:

ftp://ftp.software.ibm.com/ps/products/db2extenders/fixes/v95win/NSE_95_INTERIM_IP22989/

The fix can be applied to 32-bit Net Search Extender Version 9.5 installations together with DB2 Version 9.5 GA, FP1, or FP2. This interim fix will become obsolete with the availability of Net Search Extender Version 9.5 FP3.

Some temporary files disappear and cannot be processed

Symptoms

Some documents or attachments cannot be archived, viewed, or restored. The connector or the IBM Content Collector Web Application returns an error message stating that the document or attachment does not exist.

Causes

The document or attachment contained a virus and was deleted by the antivirus software before it could be processed. To process certain types of content, IBM Content Collector creates files in temporary working directories that contain the content. If a virus scanner deletes any of these temporary files because it seems to contain a virus, the document cannot be processed.

Diagnosing the problem

To diagnose the problem, check the log files and settings of your antivirus software to see whether the software modified or deleted any temporary files that were created by IBM Content Collector.

Resolving the problem

To resolve the problem, configure your antivirus software to not scan any of the IBM Content Collector temporary working directories on the server. If a document containing a virus is archived and restored, or if it is displayed by the Web Application, the local virus scanner on the client will detect and delete the virus.

The following table lists the temporary directories that Content Collector uses:

Table 198. Temporary directories used by Content Collector

Component	Temporary working directory
• Email Connector • File System Source Connector • SharePoint Connector • Text Extraction Connector	The working directory is defined in the connector configuration.

Table 198. Temporary directories used by Content Collector (continued)

Component	Temporary working directory
Text indexer	The working directories are defined by the configuration options **GenWorkingDirectoryForItemType** and **GenDumpDirectory**.
Outlook Extension	Temporary files are stored on the client machine in the user application folder (such as `C:\Documents and Settings\`*`user_name`*`\Local Settings\Application Data\IBM\ContentCollector_OutlookExtension\`) and in the Microsoft Windows temporary directory (such as `C:\Documents and Settings\`*`user_name`*`\Local Settings\Temp\`).
Notes Client Extension	Temporary files are stored on the client machine in the Lotus Notes data directory and in the Microsoft Windows temporary directory (such as `C:\Documents and Settings\`*`user_name`*`\Local Settings\Temp\`).

Some Microsoft Outlook email attachments cannot be previewed

Some Microsoft Outlook email attachments that are encoded in MacBinary format cannot be previewed because the file extension might not indicate that the file is in MacBinary format.

Symptoms

When a user attempts to preview an archived attachment within the Outlook Extension, no error messages are shown, but the link to the attachment does not open or it shows garbled content.

Causes

Optional. Clearly state the causes of the problem.

When archiving attachments, IBM Content Collector stores attachment data together with the file name and extension. However, the file extension might not reflect the correct encoding of the data. Attachments that were created from MIME email that was sent from MAC OS computers might be encoded in MacBinary format, but the file extensions might not reflect this encoding. In such situations, the wrong application is used to display the archived attachment. For example, if the attachment is a JPG file named `test.jpg` and it is encoded in MacBinary format, previewing the attachment shows garbled content, because the content is not really JPG.

Therefore, Content Collector appends a .bin suffix to the attachment's file name to be able to identify MacBinary encoded files after they were archived. For example, Content Collector archives the file `test.jpg` with file name `test.jpg.bin`. When you view such an attachment, it is passed to the default application that is associated with the suffix .bin. If this application can handle the MacBinary encoded file, the attachment is displayed correctly.

Resolving the problem

If previewing the attachment still shows garbled content, view or restore the entire message to display the attachment correctly.

Troubleshooting task routes

These descriptions can help you solve common problems when working with task routes.

Evaluating metadata when using the audit log

If using the audit log to evaluate metadata, select the minimum amount of metadata required, as otherwise the values can be difficult to process.

What to do if content is routed in an unexpected order

If content seems to be routed in a different way than expected, check the rule order on decision points to be sure that the rules are evaluated in the order expected. This order can be changed by clicking the decision point, and modifying the order in which rules are evaluated.

What to do when errors occur

When errors or issues occur, ALWAYS set the task route log level in the Configuration Manager by selecting the Task Route tab under **Tools menu → Task Route Service Configuration** and setting the log level to Trace2 (verbose logging). In addition, set the connector that is failing to the same log level. This information may help to diagnose the problem, and if not, will be essential for IBM Support staff to help diagnose the problem.

What to do if the CM8 Duplicate Detection task does not detect duplicates

If the CM8 Duplicate Detection task is not detecting duplicates, make sure that the **Filter Attributes by Item Type** field is set to an item type that has been used in the preceding CM 8.x Configure Item Types task. If there is a mismatch here, no duplicates will be detected.

Troubleshooting the indexer for text search

These topics can help you solve common problems when using the text search support in IBM Content Manager.

Incorrect Net Search Extender configuration

IBM Content Collector text search support requires that all text indexes managed by Net Search Extender be created with specific index configuration settings. If these settings are not configured correctly, the indexer for text search does not work correctly because either the text search result set is not complete or no search result set is found.

Symptoms

Incorrect settings can also lead to significant performance degradation and data loss. Therefore, it is essential that the index configuration is correct for each Net Search Extender full text index of each item type.

The following configuration options are essential because these options control how text is indexed and whether this text can be found again:

- **RESPECTCASE**
- **TREATNUMBERSASWORDS**
- **BLOCKMODE**

These values are checked by the indexer for text search. If they are not correctly configured in the Net Search Extender full text index configuration, the indexer issues one or more error messages with the following disclaimer:

```
A configuration or setup problem occurred:
The Net Search Extender index configuration is not as recommended and cannot be
used with IBM Content Collector.
The Net Search Extender index must be recreated using a corrected
Net Search Extender index configuration.
Modify the Net Search Extender index configuration template file 'cteixcfg.ini'
in the Net Search Extender installation directory and then recreate the index.
Net Search Extender will then copy the index configuration template file to the
new index.
```

Causes

You cannot change the Net Search Extender index configuration after the full text index has been created. Also, Content Collector cannot correct Net Search Extender index configuration settings before the indexer uses these settings.

This restriction exists because, when an item type is created by using the IBM Content Collector initial configuration wizard or the IBM Content Collector set-up tool, it is created on the IBM Content Collector server machine and not on the machine on which IBM Content Manager and the Net Search Extender applications usually run.

This means that the tools that create the item type cannot access to the Net Search Extender index configuration template file called `cteixcfg.ini`. Configuration settings can only be checked after the full text index is created and the indexer is run on the item type. However, at that point in the process, it is too late to change a setting because the index is already created based on the settings in the template configuration file.

Resolving the problem

If a Net Search Extender full text index is created with incorrect configuration settings, the complete index must be recreated and rebuilt. To correct this problem, you must perform the following steps:

1. Change the values of the index configuration options in the Net Search Extender index configuration template file to the recommended values. The Net Search Extender index configuration template file called `cteixcfg.ini` is located in the Net Search Extender installation directory (usually `.../sqllib/db2ext`).
2. Recreate the Net Search Extender index by using the `afuRecreateIndex` tool.

   All items in the item type are reprocessed the next time that the `afuIndexer` process runs. Depending on the size of the item type, processing all items again can take a considerable amount of time.

The item type cannot be processed by an indexer tool

The indexer tools (afuIndexer, afuCreateIndex, afuConfig, or afuCheck) cannot process the item type that was passed.

Symptoms

The following error is reported:

```
A configuration or setup error has occurred with the following
reason:
The required database table 'AFU1FTIOPEN01057' does not exist. Check
if this item type has been enabled for processing by the indexer
```

Causes

The reason for this error is that the indexer tool was called with an item type that is not enabled for processing by the tool.

Resolving the problem

Check that you entered the correct item type and that the item type has been enabled for processing by the indexer for text search.

If your item type is an email compound item type, do not enter the attachment item type name. Attachment item types cannot be processed on their own but are automatically processed as part of the email item type.

A native DB2 library cannot be loaded

The native DB2 library db2jcct2 cannot be loaded

Symptoms

The DB2 library db2jcct2 cannot be loaded because the path to the library does not exist. The following error is reported:

```
Connecting to database.................: icmnlsdb
com.ibm.db2.jcc.a.SqlException: [jcc][10389][12245][3.52.95] Failure in loading
native library db2jcct2, java.lang.UnsatisfiedLinkError: db2jcct2 (A file or dir
ectory in the path name does not exist.):  ERRORCODE=-4472, SQLSTATE=null
        at com.ibm.db2.jcc.a.ad.a(ad.java:650)

...

at com.ibm.afu.ftsidx.createconfig.Main.main(Main.java:101)
--------------------------------------------------------------------------------
afuConfig: An error occurred:
           An SQL problem occurred:
           [jcc][10389][12245][3.52.95] Failure in loading native library
           db2jcct2, java.lang.UnsatisfiedLinkError: db2jcct2 (A file or
           directory in the path name does not exist.):  ERRORCODE=-4472,
           SQLSTATE=null
```

Causes

The LIBPATH settings are incorrect.

Resolving the problem

The path information to the 32-bit libraries must precede the path information to the 64-bit libraries. For example:

```
LIBPATH=/opt/IBM/db2cmv8/lib:/usr/lib:/lib:/home/db2inst1/sqllib/lib32:
/home/db2inst1/sqllib/lib64:/home/db2inst1/sqlib/lib:...
```

DB2 installation directory cannot be located

Symptoms

When the indexer tool `afuRecreateIndex` is run, the following error message is reported:

```
The DB2 installation directory could not be located: The environment variable
DB2HOME is not defined.
```

Causes

The cause is that the environment variable DB2HOME was not defined.

Resolving the problem

Before you can run the indexer tool `afuRecreateIndex`, you must define the environment variable DB2HOME.

This environment variable is used by `afuRecreateIndex` to determine Net Search Extender template configuration settings and must point to the DB2 installation directory, for example, on Windows to `C:\Program Files\IBM\sqllib` or on UNIX to `/home/db2inst1/sqllib`.

It is recommended that you define this environment variable permanently on all platforms.

No index is found after the index is recreated

Recreating an index using the same name and location as the dropped index results in an error stating that the index cannot be found.

Symptoms

If you run a query against a Net Search Extender index using CONTAINS successfully, then drop the index, recreate it using the same index name, and run the exact same query again, the following error is reported:

```
CTE0198 No corresponding text index
```

Causes

The cause is that the old index location is still cached. DB2 keeps the results of all internal subqueries, including the index location, in its query cache to avoid having to rebuild and run potentially expensive scans.

Resolving the problem

Clear the query cache by using the following command:

```
db2 flush package cache dynamic
```

Net Search Extender stops unexpectedly without issuing a COMMIT

The indexer process stops unexpectedly because Net Search Extender detected a fatal error.

Symptoms

If the `AfuFetchFile` UDF in the indexer process cannot process the XML files, it sends Net Search Extender a fatal error and Net Search Extender performs an immediate shutdown without issuing a COMMIT.

The following error is reported:

```
Net Search Extender indexing stopped abnormally (CTE0005_INDEX_COMMIT was not found)
```

Causes

The `AfuFetchFile` UDF cannot read the XML file that was created by the indexer process containing the textual content of the item that should be indexed.

Resolving the problem

To check why the XML file was not read:

- Enable the `AfuFetchFile` trace and logging feature by setting the configuration option **UdfUseLogging** to 1.
- Restart DB2.
- Check the value of the configuration option **UdfMaxNumberOfWarnings** and compare this value with the actual number of warnings that were reported. If more than the specified number of warnings occurred, Net Search Extender stops processing.
- Verify if the directory set in the configuration option **GenXmlDirectory** exists and is accessible, and that the directory has enough space for processing.
- Check the UDF configuration file. It is named `AfuFetchFile.config`. The directory for this file is the directory for temporary files, usually denoted by the environment variables TMP or TEMP on Windows, and `/tmp` on UNIX or Linux. Alternatively, you can have defined a directory using the environment variable AFU_FETCHFILE_CONFIG_PATH.

  You should be able to read the file name and file path of the XML file in a binary file editor. The XML path must match the value in **UdfMaxNumberOfWarnings**.

Net Search Extender fails while indexer process is running

Symptoms

When the indexer tool `afuIndexer` is run, the following error message is reported:

```
CTE0103 An internal error occurred. Location: "72", "1".
```

Causes

The cause is that Fix Pack 3 must be applied for DB2 Version 9.7.

Resolving the problem

Apply DB2 Version 9.7 Fix Pack 3 before you run the indexer tool `afuIndexer` again. `afuIndexer` will process the batch of items again.

The indexer process stops because the AFUFetchFile UDF cannot be accessed

The `AFUFetchFile` UDF used by Net Search Extender must be registered in DB2.

Symptoms

The indexer process stops and the following message is displayed:

```
The Net Search Extender UDF registered for this item type cannot
be identified.
Verify the correct installation of the Net Search Extender UDF {0}
```

Causes

The cause is that the AFUFetchFile UDF cannot be accessed because it has not been registered in DB2.

Resolving the problem

- Check if the AFUFetchFile UDF library exists. When Content Collector Text Search Support is installed, the AFUFetchFile UDF library is copied, for example on Windows, to the SQLLIB\FUNCTION\ directory. If you cannot find the file AfuFetchFile.dll there, copy it from IBM\Indexer\lib\ to your SQLLIB\FUNCTION directory.
- Check if the AFUFetchFile UDF is registered in DB2 in the DB2 Control Center under *library_ server_name* → **Application objects** → **User Defined Functions**.
- Log on to the IBM Content Manager server as the DB2 instance owner.
- Run:

  ```
  DB2 SELECT FUNCNAME FROM SYSCAT.FUNCTIONS WHERE FUNCNAME LIKE 'AFU%'
  ```

  This results in a list which, if it is not empty, lists and shows that the UDF is registered.
- If the UDF is not registered run the following in the bin subdirectory of the Content Collector Text Search Support installation directory:

  ```
  db2 -tvf AfuFetchFile.ddl
  ```

  After you have run the script, when the indexer process starts it automatically verifies that the UDF is working correctly.
- To verify if the UDF functions, run:

  ```
  db2 "values ( AFUFETCHFILE ( '_AFU_' ) )"
  ```

  The UDF is registered and running correctly if you see a table entry similar to the following:

  ```
  AfuFetchFile: [V2] Build Id: 2.2.0.0-1 [Compile TOD: Sep 21 2010 15:50:42]
  Copyright (C) 2009 by IBM Corporation
  ```
- To obtain the index identifier for a given item type, run the query:

  ```
  db2 select EVENTTABLENAME, LOGTABLENAME, INDEXIDENTIFIER from DB2EXT.TTEXTINDEXES
  WHERE INDEXNAME = upper('<IndexName>')
  ```

  where the following parts make up the index name of the given item type:

  ```
  "ICMUU" + item type comoponent id (5 digits) + "001TIE"
  ```
- To obtain the function name, run the query:

  ```
  db2 select FUNCTIONNAME from DB2EXT.TTEXTCOLUMNS WHERE
  INDEXIDENTIFIER = '<IndexIdentifier>'
  ```

  If the result is not AFUFETCHFILE, run:

  ```
  db2 UPDATE DB2EXT.TTEXTCOLUMNS SET FUNCTIONNAME = 'AFUFETCHFILE' WHERE
  INDEXIDENTIFIER = '<IndexIdentifier>'
  ```

Obtaining the DB2 port number

You must know the DB2 database name and the connection port number to configure a connection between the system administration client and the remote DB2 database.

Symptoms

To obtain the DB2 connection port number:

Resolving the problem

1. Run the following command:
   ```
   db2 get dbm cfg
   ```
2. In the output produced by the command, search for `TCP/IP-Servicename (SVCENAME) = XXX` where XXX is the service name.
3. Using the found service name, enter the following:
   - On UNIX machines in `/etc/services`:
     ```
     cat /etc/services | grep XXX
     ```
   - On Windows in `<WINDOWSINSTALLDIR>\system32\drivers\etc\services`:
     ```
     type c:\WINDOWS\system32\drivers\etc\services  | find XXX
     ```

   The output could look something like the follow:
   - For UNIX: `db2c_db2inst1 50000/tcp`
   - For Windows: `db2c_DB2 50000/tcp`

   The DB2 port number is 50000.

Using SQL statements that are very long or complex

SQL statements that are either too long or very complex result in an SQL0101N error message.

Symptoms

The SQL010N error message appears in the log file and the SQL statement is not processed.

Causes

This problem is caused because the DB2 statement heap size is too small.

Diagnosing the problem

Check if the SQL statement:

- Exceeds either a server system limit or a data source system limit for length or complexity.
- Violates some other data source specific limit.

Resolving the problem

Follow the steps in this SQL troubleshooting technote.

Using an incorrect JDBC setup for the indexer processes

The SQL error SQL1224N is caused by an incorrect JDBC setup.

Symptoms

The connection to the library server database in which item types are stored is implemented using Java Data Base Connectivity (JDBC) which defines how a Java client accesses a database. JDBC features different connection types and IBM Content Collector supports JDBC type 2 and JDBC type 4.

If an incorrect JDBC setup is used, the indexer process stops and the following SQL error is reported:

```
afuIndexer[worker]: An error occurred:
An SQL problem occurred:
[jcc][4038][12241][3.53.95] T2LUW exception: SQL1224N The
database manager is not able to accept new requests, has
terminated all requests in progress, or has terminated your
particular request due to an error or a force interrupt.
SQLSTATE=55032
ERRORCODE=-1224, SQLSTATE=55032
```

Resolving the problem

Switch to a JDBC type 4 connection or increase the maximum number of client connections in DB2. The configuration parameter for this limit is called **max_connections**. For details see the DB2 documentation.

The JDBC type 2 driver cannot be loaded correctly

The `afuIndexer` process cannot be started, displaying an SQL error message stating that a native library could not be loaded.

Symptoms

After the `afuIndexer` process is started, the following error message appears:

```
[jcc][10389][12245][3.52.95] Failure in loading native library
db2jcct2, java.lang.UnsatisfiedLinkError: db2jcct2 (A file or
directory in the path name does not exist.): ERRORCODE=-4472,
SQLSTATE=null
```

Causes

This may be caused by an incorrect library search path on a 64-bit operating system.

Resolving the problem

Check your LIBPATH settings for the indexer for text search. The path to the 32-bit versions of the shared libraries must precede the path to 64-bit versions of the shared libraries. For example:

```
LIBPATH=...:/home/db2inst1/sqllib/lib32:/home/db2inst1/sqllib/lib64:
/home/db2inst1/sqlib/lib:...
```

Items with textual content greater than 1 MB are not indexed when using DB2 V9.5

If you are using DB2 Universal Database V9.5 and the indexer for text search, items which have a textual content that is greater than 1 MB might not be indexed.

Symptoms

No data is passed to Net Search Extender to be indexed.

Resolving the problem

Carry out these steps before you call DB2 and start indexing. The items that were not indexed must be reindexed:

- Set the DB2 environment variable DB2_RESTRICT_DDF to true.
- Include this variable in DB2ENVLIST on UNIX machines.
- Defined this variable as a system environment variable on Windows.

The indexer process stops because the index cannot be found

The indexer `afuIndexer` process stops and an error message is displayed stating that the index does not exist.

Symptoms

When the indexer `afuIndexer` process attempts to run the Net Search Extender update index command, the following error occurs:

```
CTE0132 The text index "ADMINISTRATOR"."<NSE indexname>" does not exist.
```

Note that the index name consists of the current schema name, a dot, and the index name.

Causes

This is most likely caused by a missing schema specification on the command line when `afuIndexer` is run.

Resolving the problem

Restart the `afuIndexer` process using the correct value for the `-schema` command line argument.

To determine the default schema for your current connection use the following DB2 command:

```
db2 select current_schema from sysibm.sysdummy1
```

To determine the correct schema for an item type and its text index use the following DB2 command:

```
db2 select indexschema from db2ext.ttextindexes where indexname = '<NSE indexname>'
```

where <NSE indexname> is the index name that appears in the error message.

Changes to the default directory and file name settings of the AFUFetchFile UDF configuration

The indexer process automatically creates and maintains a binary configuration file which is used by the AFUFetchFile UDF.

Symptoms

The guidelines used for creating the default directory and file name of the AFUFetchFile UDF configuration file have changed to conform with the latest Windows program files installation guidelines.

Environment

- Default directory settings in Wndows:

  Formerly, the default directory for this file on Windows was the TMP subdirectory of the Content Manager installation directory defined in the environment variable IBMCMROOT. Starting with this indexer for text search version, the default directory for this file is the directory for temporary files, usually denoted by the environment variables TMP or TEMP. Alternatively, you can use a directory of your choice to store the configuration file by defining the environment variable AFU_FETCHFILE_CONFIG_PATH and assigning it a value. For example:

  ```
  SET AFU_FETCHFILE_CONFIG_PATH=C:\IccIndexerData\UdfConfig
  ```

  Important:

  – If you use the environment variable AFU_FETCHFILE_CONFIG_PATH to access the UDF configuration file, make sure that you have defined the environment variable as part of the system settings on the Windows Control Panel. (**Control Panel** → **System** → **Advanced** → **Environment variables** → **Add**) . Simply defining the environment variable using the SET command in the command line shell is not sufficient and will not work within the context of DB2.

  – It is recommended that you use the environment variable AFU_FETCHFILE_CONFIG_PATH if you run and control the indexer process through a Windows Terminal Service or a Windows Remote Session. In such an environment, the environment variables TMP and TEMP may not point to the directory defined in the system settings (for example: C:\DOCUME~1\icmadmin\LOCALS~1\Temp) but instead may point to a subdirectory of the original directory (for example: C:\DOCUME~1\icmadmin\LOCALS~1\Temp\1). This results in the indexer using a different environment variable value compared to DB2 and produces a CTE0100 error.

- Default directory settings in UNIX and Linux:

  The default directory for this file on UNIX or Linux is still the /tmp directory. As a new option however, you can now also use a directory of your choice to store the configuration file by defining the environment variable AFU_FETCHFILE_CONFIG_PATH and assigning it a value. For example:

  ```
  export AFU_FETCHFILE_CONFIG_PATH=/IccIndexerData/UdfConfig
  ```

  The environment variable must be defined both in the context in which the indexer is run as well as where DB2 is started. Also note that if you use the environment variable AFU_FETCHFILE_CONFIG_PATH to access the UDF configuration file, you must include AFU_FETCHFILE_CONFIG_PATH in the DB2ENVLIST so that the environment variable is accessible in the DB2 environment.

- Default file name settings on all platforms:

  Formerly, for all platforms, the file name for the configuration file was `.AfuFetchFile.control`. Starting with this version of the indexer for text search, the file name of the configuration file has been changed to `AfuFetchFile.config`. After you have installed the latest version of the indexer for text search, all old instances of the `.AfuFetchFile.control` file can be removed.

Resolving the problem

If the `AFUFetchFIle` UDF cannot load the configuration file because the UDF could not identify the directory in which the configuration file was created, the following message is displayed:

```
afuIndexer: An error occurred:
            CTE0100 A DB2 operation failed. DB2 information: "38C90"
            "[IBM][CLI Driver][DB2/NT] SQL0443N  Routine "AFUFETCHFILE"
            (specific name "*2151400") has returned an error SQLSTATE with
            diagnostic text "CFG: ERR 2: \AfuFetchFile.config".
             SQLSTATE=38C90  ".
```

In this case, it is most likely that the environment variable TEMP, TMP, or AFU_FETCHFILE_CONFIG_PATH was not set correctly and was not unique.

Check the environment variable settings and ensure that the values are consistent, both in the DB2 context and in the shell or command line session in which the indexer process runs.

Error accessing XML files in the afuFetchFile UDF

The indexer process is unable to access the XML files which contain the document content that is to be indexed.

Symptoms

The `afuFEtchFile` UDF cannot open the XML files created during document processing resulting in the following error message:

```
afuIndexer: An error occurred:
            CTE0100 A DB2 operation failed. DB2 information: "01H99"
            "[IBM][CLI Driver][DB2/AIX64] SQL0462W  Routine "*ETCHFILE"
            (specific name "") has returned a warning SQLSTATE, with diagnostic
            text "XML: ERR: 22: /path/ICC211_XML/01d1/afu#A100100B2".
             SQLSTATE=01H99 ".
```

Causes

A possible cause for this error message are inadequate file access permissions. For example, the XML file listed in the error message only has user read/write access:

```
-rw------- 1 db2inst1 db2iadm1 7775 Feb 23 19:03 afu#A1001001A10B19C35634F05123-V1#.xml
```

Diagnosing the problem

This situation arises if the XML files and the XML directory are created using an overly restrictive umask setting. As a result, the `afuFetchFile` UDF which is run by DB2 as the fenced user cannot access the XML files that are created by the indexer process which is run as the instance owner. The owner of the `db2fmp` process determines which permissions are needed for the XML files.

Resolving the problem

1. Check the owner and group ID access permissions for the `afuFetchFile` UDF in the DB2 control file `~/sqllib/adm/.fenced`. The `afuFetchFile` requires the access permissions of the fenced user and not the instance owner.
2. If the file access permissions of the Net Search Extender index files are not the same as for the XML files (400), change the permissions. The instance owner and fenced user must share the same _primary_ group. A umask for the instance owner of 077 must be changed to a umask of at least 027 for the instance owner.
3. After these permission changes, delete the XML files and the directory structure defined in the configuration option **GenXmlDirectory**.
4. Rerun the indexer process.

Error accessing XML files in the afuFetchFile UDF if the item ID consists of question marks

The indexer process is unable to access the XML files which contain the document content that is to be indexed if the item ID consists of question marks (?).

Symptoms

The `afuFEtchFile` UDF cannot open the XML files created during document processing resulting in the following error message:

```
afuIndexer: An error occurred:
            CTE0100 A DB2 operation failed. DB2 information: "01H99"
            "[IBM][CLI Driver][DB2/AIX64] SQL0462W  Routine "*ETCHFILE"
            (specific name "") has returned a warning SQLSTATE, with diagnostic
            text "XML: ERR: 22: /path/ICC211_XML/01d1/afu#??????????".
             SQLSTATE=01H99 ".
```

Causes

The cause for this error message is that the TIEREF for the original item is empty. The item must have a valid TIEREF value to be processed properly.

Diagnosing the problem

This situation arises if the item that is being processed does not contain a TIEREF value, which means that the value in the TIEREF column for the item is empty. This can occur if the item type has been processed by the fast indexer before and was left in an incorrect state.

Resolving the problem

For all items that do not have a TIEREF value, use the GENTIE2 tool of the fast indexer to re-create the TIEREF. See the fast indexer documentation for details.

The indexer worker process stops indicating an out of memory situation

Sometimes the indexer worker process stops and assumes that there is not enough memory to continue processing.

Symptoms

This situation is not caused by a batch size that is too large to allow any further processing.

Causes

The reason for this error situation can be one of the following:

- An item contains a large attachment that cannot be processed.
- Too many parallel tasks have been specified.
- The configured memory settings are too high.

Resolving the problem

This is not a severe issue and does not impact the overall indexer processing. Unfortunately, this situation and its consequences cannot be avoided in Java.

For this reason, the indexer process will continue processing the other items in the batch. Items that could not be processed because of this memory issue will be stored in the table of completed tasks and can be reprocessed after the respective configuration options have been modified.

Part 10. Appendixes

© Copyright IBM Corp. 2008, 2010

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:
INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

© Copyright IBM Corp. 2008, 2010

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Deutschland GmbH
Department M358
IBM-Allee 1
71139 Ehningen
Germany

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Microsoft, SharePoint, and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

The Oracle Outside In Technology included herein is subject to a restricted use license and can only be used in conjunction with this application.

Other product and service names might be trademarks of IBM or other companies.

Index

A

access to archived data
  See archived data access
Active Directory 188
additional archiving information
  defining the entry fields 235
  layout of the screen 235
  Outlook Extension 792
  overview 431
  user-defined metadata 240
address lookup
  Active Directory 188
  Names & Address Book 188
  recipient address 188
afuIndexer
  See indexer for text search
archive mapping files 682
archived data
  previewing 777
  restoring 777
  searching 777
  viewing 777
archived data access
  configuring 228, 232
  content server properties 232
  delegates 777
  general page 229
  text index fields 232
archiving
  preparing email 525
archiving tasks
  Content Manager 506
  file system repository 536
  FileNet IS 528
  FileNet P8 543
associate metadata task 531
attachments
  archiving scenarios 21
  extracting 522, 565
  linking 557
  placeholders 521
  removing 516
audit logs 279, 770
automatic archiving
  collectors 460, 476, 478, 479, 483
  scenarios 21

B

blacklist 715

C

Calculate Expiration Date task 498
capture tasks
  Content Manager 506
  file system repository 536
  FileNet IS 528
  FileNet P8 543
certificates
  embedded Web Application Server 102, 718
  for server authentication 104, 789
  replacing 102, 718
  trust relationship 106
changing the WAS port 100
checking for documents
  FileNet P8 540
  IBM Content Manager 504
Classification Module
  classifying Microsoft Exchange email 455
  decision plans 456
  IBM Classification Module task 538
  integration 454
  knowledge bases 456
  metadata properties 457
  setting up 454
  using Classification Module with Content Collector 453
client capture with prompt for metadata
  defining the entry fields 235
  layout of the screen 235
  Outlook Extension 792
  specifying additional archiving information 431
  user-defined metadata 240
client configuration settings 227
CM 8.x Associate Content task 499
CM 8.x Configure Item Types task 502
CM 8.x Confirm Document task 504
CM 8.x Create Document task 506
CM 8.x Duplicate Detection task 509
CM 8.x Store Version Series task 511
CM 8.x Update Document task 514
collecting documents
  email 460, 472, 476
  files 479
  mail-system documents 460
  SharePoint 483
  SMTP/MIME email 478
collection sets 680
collections
  adding 229
  defining 694
  editing 229
  multiple document classes 706
  multiple item types 706
  overview 680
  removing 229
  searching 708
collectors
  automatic archiving 460
  by rules 460
  by selection 472
  configuring
    collection filters 468, 481
    schedule 458
  file system 479
  interactive archiving 472
collectors *(continued)*
  life cycle 476
  mail system 460, 472
  SharePoint 483
  SMTP 478
  stubbing 476
  types 458
CommonStore
  item types 149, 710
  legacy support 131, 135, 139, 140, 141
  legacy update strategies 138
CommonStore for Exchange Server
  legacy restore 146
  legacy support 142, 220
  legacy support configuration 143, 221
  legacy viewing 146
  restored documents not displayed in Outlook 811
CommonStore for Lotus Domino
  legacy support 146, 147
compliance 524
components 13
configuration checklists
  configuration databases 44
  connectors 46
  general settings 52
  overview 34
  repository systems 40
  source systems 34
configuration databases
  adding 171
  configuration checklists 44
  deleting 173
  editing 171
  exporting 173
  importing 173
  setting up 171
configuration files
  archive mapping files 682
  file system documents 692
  Notes applications 691
  search configuration files 682
  SharePoint documents 692
Configuration Manager
  adding, changing, and deleting configuration objects 160
  enabling security 158
  overview 157
  signaling changes to the configuration database 159
configuration parameters
  CXS legacy restore 143, 221
  indexer for text search 624
  verifying and adjusting 81
Configuration Web Service settings 224
confirm tasks
  FileNet P8 540
  IBM Content Manager 504
connectors
  changing the user account 213

© Copyright IBM Corp. 2008, 2010

connectors *(continued)*
configuration checklists 46
configuring 184
Email Connector 185
File System Repository Connector 202
File System Source Connector 184
IBM Content Manager Connector 200
IBM FileNet Image Services Connector 204
IBM FileNet P8 Connector 205
Metadata Form Connector 210
Microsoft SharePoint Connector 199
overview 184
required privileges
Content Manager 201
IBM FileNet P8 207
SMTP Connector
configuring 189
configuring Lotus Domino 196, 197, 198
configuring mail server 192
configuring Microsoft Exchange 192, 193, 194, 195, 196
message queue directory 191
SMTP authentication 192
source connectors 184
target connectors 199
Text Extraction Connector 212
Utility Connector 212
utility connectors 210
content elements 542
content management scenarios 24
Content Manager 8.x
configuring item types 502
detecting duplicates 509
updating the record of duplicates 514
Content Manager privileges
ItemAdd 201
ItemAddLink 201
ItemCheckInOut 201
ItemDelete 201
ItemLinked 201
ItemLinkTo 201
ItemQuery 201
ItemRemoveLink 201
ItemSetSysAttr 201
ItemSetUserAttr 201
ItemSQLSelect 201
ItemSuperCheckIn 201
ItemTypeQuery 201
SystemQueryUserPrivs 201
counters 761, 762
create version series 511, 550
custom attributes
search configuration file 702
search mapping files 700
custom metadata tasks
email 522
SMTP 565

D

data stores
adding 171
data stores *(continued)*
deleting 173
editing 171
exporting 173
importing 173
setting up 171
date ranges 781
date values
changing display format 712, 713
DBCS search terms 781
decision points
adding 281
overview 281
declaration elements 702
declare record task 553
deduplication
attachments 445
detecting duplicates 509
email 445
hash key calculation 446, 451, 452
hash keys 522, 565
display formats
date values 712, 713
document archiving scenarios 21
document classes
adding 229, 692
collections 680
editing 229
removing 229
document instances 548
document management 3
document retention
overview 438
scenarios 23
Document Viewer
authentication parameters 751
configuration options 748, 751
enabling viewing in Workplace XT 754, 787
highlighting parameters 751
message translation parameters 751
overview 747
request parameters 751
documents
preparing for stubbing 526
previewing 785
restoring 785
stubbing 516
duplicates
attachments 808
Content Manager 8.x 509
hash keys 522, 565
processing 727
recording 514

E

EC Copy to Mailbox task 516
EC Create Email Stub task 516
EC Extract Attachments task 522
EC Extract Metadata task 522
EC Finalize Email for Compliance task 524
EC Prepare Email for Archiving task 525
EC Prepare Email for Stubbing task 526
email
analytics preparation
IBM eDiscovery Analyzer 24
IBM eDiscovery Manager 24
scenarios 24
finalizing for compliance 524
journaling scenarios 22, 23
preparing for archiving 525
preparing for stubbing 526
retention 438
searching 697, 777
services 220
stubbing 516
Email Connector 185
email searches 777
Enterprise Records 553
envelope journal messages 808
environment variables 83, 713
erroneous documents 713, 716
error task routes 280
escape characters 781
event logs
deleting 772
event IDs 773
interpreting 772, 773
expiration dates 498
Expiration Manager
.bat file 443
properties file 441
running on UNIX 443
running on Windows 439
running remotely 440
exporting tasks routes as templates 277
Extension for Outlook
See Outlook Extension
extension nodes
configuring 90
SMTP Receiver 91, 92, 94
starting task route services 91
Extract Text task 527

F

field definitions 700
file in folder
FileNet IS 529
FileNet P8 555
file system
archive mapping files 682
archiving 725
configuration files 682
metadata 256, 531
post-processing 534
file system repository
archiving task 536
File System Repository Connector 202
File System Source Connector 184, 725
FileNet Image Services Create Document task 528
FileNet Image Services File Document In Folder task 529
FileNet Image Services Modify Permissions task 530
FileNet IS repository
archiving task 528
FileNet P8
archive mapping files 682

FileNet P8 *(continued)*
content elements 542
document instances 548
full-text indexing 675, 677
P8 Content Search Engine 675, 677
saving prepared text as XML 561
search configuration files 682
FileNet P8 task routes
upgrading 134
FileNet P8 Version 4.5.1
full-text indexing 677
P8 Content Search Engine 677
filtering
filter criteria 468, 481
findPst 436
form elements 702
FSC Associate Metadata task 531
FSC Post Processing task 534
FSR Create Document task 536
full-text search
FileNet P8 675, 677
FileNet P8 Version 4.5.1 677

G

general settings
configuration checklists 52
configuring 220
generally available metadata types 269
guidelines
attribute search 780
searching a non-indexed archive 780
searching an indexed archive 780
text search 780

H

hardware prerequisites 29
hash keys
calculation
for email received through SMTP 452
for Exchange 446
for Lotus 451
extracting metadata 522, 565
HTML form for collecting additional archiving information 234
HTTPS certificates
embedded Web Application Server 102, 718
replacing 102, 718
SSL certificates 104, 789

I

IBM Classification Module task 538
IBM Content Collector APIs
developing 733
Document Viewer 733
Web Application services APIs 733
IBM Content Collector for Microsoft SharePoint
installing
console mode 64
GUI mode 63
IBM Content Collector for Microsoft SharePoint *(continued)*
installing *(continued)*
overview 63
silent mode 64
IBM Content Collector Notes Client Extension
installing
GUI mode 66
overview 65
silent mode 67
IBM Content Collector Outlook Extension
installing
console mode 108
GUI mode 107
overview 107
silent mode 109
IBM Content Collector Outlook Web App (formerly Outlook Web Access) support
installing
GUI mode 122
overview 112
IBM Content Collector Server
installing
console mode 68
GUI mode 67
overview 67
silent mode 68
IBM Content Collector services 178, 214
IBM Content Collector Text Search Support
installing 62
IBM Content Manager
archive mapping files 682
search configuration files 682
IBM Content Manager Connector 200
IBM FileNet Image Services Connector 204
IBM FileNet P8 Connector 205
IBM FileNet P8 privileges
class security 207
code module folder security 207
custom object default instance security 207
document default instance security 207
event actions 207
folder default instance security 207
object store security 207
Records Management security 207
root folder security 207
subscriptions 207
IBM InfoSphere Classification Module 457
classifying Microsoft Exchange email 455
decision plans 456
IBM Classification Module task 538
integration 454
knowledge bases 456
setting up 454
using Classification Module with Content Collector 453
IIS server 117
indexer for text search
adding attributes
to the configuration file for the item type 615
to the model file 614
advanced operations 612
afuIndexer 596
arguments 596
attachment processing 649
backing up files 671
checking status of indexed items 604
configuration options 624
configuring the item type 592, 623
constraints 644
debugging 658
deleting item types 672
environmental variables 621
error messages 672, 673, 674
existing CommonStore item types 589
forced shutdown 660
gathering statistics 611
index reorganization 661
indexing email with attachments 648
item type attributes 616
layout of database tables 671
log files 608
main steps 584
model files 613
monitoring 606
overview 583, 584
preparing for indexing 613
process return codes 603, 653
processing items with severe errors 658
processing multiple item types 648
recreating an index 666
reindexing items 656
retrieving items form the IBM Content Manager Resource Manager 659
return codes 605
running afuIndexer in continuous mode 646
running the configuration check tool 641
searching for error strings 662
steps to take if indexing stops unexpectedly 657
stopping processing 660
timing information 610
trace files 609
transfer buffer size 640
information center
adding languages 124
modifying settings 225
starting 123
stopping 123
updating 124, 125
updating from CD 124
initial configuration
configuration database 79
Content Manager item type 74
DB2 database 79
IBM FileNet P8 75, 77
Lotus Domino 70
Microsoft Exchange 72
Microsoft SharePoint 73

initial configuration *(continued)*
Oracle database 80
overview 69
repositories 74
running 81
SMTP 73
source systems 70
SQL Server database 79
starting 69
system configuration 70
installation
additional prerequisites 32
for use with Content Manager 59
for use with IBM FileNet P8 60
IBM Content Collector for Microsoft SharePoint
console mode 64
GUI mode 63
overview 63
silent mode 64
IBM Content Collector Notes Client Extension
GUI mode 66
overview 65
silent mode 67
IBM Content Collector Outlook Extension
console mode 108
GUI mode 107
overview 107
silent mode 109
IBM Content Collector Outlook Web App (formerly Outlook Web Access) support
GUI mode 122
overview 112
IBM Content Collector Server
console mode 68
GUI mode 67
overview 67
silent mode 68
IBM Content Collector Text Search Support 62
on several nodes 61, 87
overview 29
prerequisites 29
additional 32
hardware 29
overview 29
software 30
scale out 61
services 178, 214
uninstalling 127
upgrading 131
Web applications 97
interactive archiving
collectors 460, 472
item types
adding 229, 692
collections 680
configuring 502
editing 229
legacy 149, 710
removing 229
searching 706

J

JDBC
provider 95, 98, 100, 803
journaling
configuring Lotus Domino 196, 197, 198
configuring mail server 192
configuring Microsoft Exchange 192, 193, 194, 195, 196
scenarios 22, 23

K

keyword search 782

L

labels
search fields 705
LDAP settings 177
legacy item types 149, 710
legacy support
installing 131, 139, 140, 141
legacy support for CSLD
configuring 147
legacy support for CSX
configuration parameters 143, 221
configuring 142, 220
restored documents not displayed in Outlook 811
legal tracing scenarios 24
link texts 516
link to attachments task 557
linking documents with their attachments 557
links in Microsoft Outlook 716
List Lookup 496
lists
adding 239
editing 239
List Lookup 496
overview 237
sorting 239
logs
files 765
formats 769
indexer for text search 608
levels 770
Lotus Notes mail templates
replacing 106
Web Application 226

M

MacBinary format 818
matches regular expressions
samples 428
MC Retrieve Additional Metadata task 540
message queue directory 191
metadata
adding and editing
file system metadata 241
XML file metadata 243
associate metadata task 531
metadata *(continued)*
defining to process files 729
delimited file metadata 242
extract metadata task
email 522
SMTP 565
file system 237
manual archiving 210
mapping Classification Module properties 457
search 699
system 237
user defined 237
wizard
delimited file metadata 243
XML file metadata 245
Metadata Connector task 540
Metadata Form Connector 210
metadata form template 234
Metadata Web Application
modifying settings 233
Microsoft Exchange
envelope journal messages 808
PST files 516
Microsoft SharePoint
See SharePoint
migration 131, 135
modifying object security 558
modifying permissions
FileNet IS repository 530
monitoring system performance
audit logs 770
counters 761, 762
event logs 771, 772, 773
indexer for text search 606
log levels 770
overview 759
performance counters 761, 762
system dashboard 759, 760
system log files 765
system log format 769

N

Names & Address Book 188
names.nsf 188
Notes Client Extension
installing
GUI mode 66
overview 65
silent mode 67

O

objects
adding 160
changing 160
deleting 160
security 558
offline repositories
enabling
Lotus Domino 109
Microsoft Exchange 110
Microsoft Exchange 811
overview 109

operators
Add 407
AddDays 408
AddMonths 408
AddYears 409
Age 409
And 409
Append 409
CaseInsensitiveEqual 410
Ceiling 410
Conditional 411
conditional values 494
Contains 411
ContainsSome 411
Divide 412
DynamicMetadata 412
Element 413
Equal 413
Floor 414
FromString 414
GreaterThan 415
GroupLookup 415
Intersection 416
IsLikeIn 416
Length 416
LessThan 416
Like 417
Modulo 417
Multiply 418
Narrow 418
Not 418
Or 419
property values 406
RegexSearch 419
RegexSubstitute 419
rules 406
Slice 420
Subtract 420
TestMetadata 421
ToString 421
TripleDESEncrypt 420
Widen 421
order of tasks in a task route 284
Outlook Extension
additional archiving information 792
advanced configuration 794, 795
archive queries 795
archiving 790
client prerequisites 796
complex search expressions 795
configuration options 793
custom metadata 792
default inbox 796
Exchange server mailbox 796
inbox 796
installing
console mode 108
GUI mode 107
silent mode 109
local archiving 794, 796, 797, 798
mailbox 796
manual archiving 790
manual stubbing 793
mark for stubbing 793
message stubs 796
metadata prompt 792
Outlook Extension *(continued)*
mobile user support 794, 796, 797, 798
number of retries 809
offline archiving 794, 796, 798
offline repository 794, 796, 797, 798
overview 788
periodic updates 798
problems loading 809
queries 795
restoring 791
search terms 795
searching 795
stubs 796
timeout 809
tracing 795
Outlook links 716
Outlook Web App
archiving functions 121
configuring
basic parameters 114
tracing 120
URL mappings 118
Exchange server 121
prerequisites 113, 117
selecting the authentication method 114
Outlook Web App Service
Active Directory page 117
configuring 113
Outlook Web App support
installing
GUI mode 122
overview 112
overview 3

P

P8 4.x Confirm Document task 540
P8 4.x Create Content Elements task 542
P8 4.x Create Document task 543
P8 4.x Create Email Instance task 548
P8 4.x Create Version Series task 550
P8 4.x Declare Record task 553
P8 4.x File Document in Folder task 555
P8 4.x Link Documents task 557
P8 4.x Modify Object Security task 558
P8 4.x Save Prepared Text as XML task 561
paper clip icons 521
performance counters 761, 762
personal folders (PST files)
copy to mailbox 516
phrase search 783
placing tasks in task routes 278
prerequisites
additional 32
hardware 29
installation 29
software 30
preserve paper clip icon 521
preview 785
preview mode
Outlook 713
primary node
configuring 88
starting task route services 91
privileges
IBM Content Manager Connector 201
IBM FileNet P8 207
Records Management security 207
property mappings 490
PST files
copy to mailbox 516
public NAB 188

R

recipient address 188
records
declaring 727
regular expressions
matching 421
overview 421
replacing 421
syntax reference 423
removing attachments 516
replacement regular expressions
samples 428
repository systems
configuration checklists 40
requests for interactive archiving
setting up 734
resilience 713
restored docments not displayed in Outlook 811
restoring 785
restrictions
previewing 785
restoring 785
result elements 702
retention management
Expiration Manager 439, 440, 441, 443
overview 438
scenarios 23
retrieving additional archiving information 540
retrieving information from the temporary metadata database 540
rules
adding or editing 282
always true rule 282
overview 281
rules' clause 282
setting the order of evaluation 281

S

samples
matches regular expressions 428
regular expressions 428
replacement regular expressions 428
Save Temporary File Copy task 564
saving text as XML 561
SC Delete Email task 564
SC Extract Attachments task 565
SC Extract Metadata task 565
SC Prepare Email for Archiving task 566
SC Prepare Email for Deletion task 567

scale out
configuring
extension nodes 90
primary node 88
installing 61
insufficient permissions 801
overview 87
primary node failover 801
SMTP Receiver 91, 92, 94
starting task route services 91
troubleshooting 801
write access to configuration database 801
scenarios
attachment archiving 21
automatic archiving 21
content management 24
document archiving 21
document retention 23
email analytics preparation 24
email journaling 22, 23
journaling 22, 23
legal tracing 24
overview 21
retention management 23
space-saving 21
stubbing 21
search
across collections 680, 708
basic setup 697
custom attributes 699
date range 781
exact match 781
excluding terms 781
field labels 699, 705
including terms 781
overview 777
sender or recipients 781
terms
excluding from search 781
including in search 781
tips 781
user-defined metadata 699
search configuration files
custom attributes 702
for email search in IBM Content Manager 686
for email search in IBM FileNet P8 688
overview 682
search mapping files
custom attributes 700
search scopes 680
search templates 682
choosing 709
collections 694
searching across collections 708
searching across document classes 706
searching across item types 706
secure communications
client communications 720
overview 717
URL protection
dynamic URLs 721
overview 721
static URLs 721
security
client communications 720
for objects in the repository 558
secure communications 717
URL protection 721
self-signed certificates 104, 789
server authentication 104, 789
services
changing the user account 213
overview 178, 214
Task Routing Engine 173
set-up tools
before creating an item type 587
configuring
FileNet P8 repository 580
IBM Content Manager repository 573, 584
IBM Content Manager repository for documents from Notes applications 575, 576
IBM FileNet P8 repository for documents in Notes applications 581
enabling
Domino template for archiving 577
Domino template for Content Collector processing 578
existing item type for processing by the indexer for text search 579, 588
using 572
sets of collections 680
shared network
moving documents off 725
SharePoint
archive mapping files 682
blog properties 266
configuring 73
connector 199
create version series
Content Manager 511
FileNet P8 550
installing
console mode 64
GUI mode 63
overview 63
silent mode 64
migration 151
search configuration files 682
SharePoint tasks
SP Create File 567
SP Get Versions 568
SP Manage Link 570
SP Post-processing 570
SMTP Connector 189
SMTP scenarios 22
software prerequisites 30
source connectors 184
source systems
configuration checklists 34
SP Create File task 567
SP Get Versions task 568
SP Manage Link task 570
SP Post-processing task 570
space-saving scenarios 21
special characters 781
storage data model
description 15
email data model 15
File System data model 15
Microsoft SharePoint data model 15
storing metadata temporarily 210
stubbed documents
searching 727
stubbing
delayed 516
immediately 516
preparing documents 526
scenarios 21
syntax reference
regular expressions 423
system dashboard
overview 760
starting the interface 759
system metadata
selecting 246
system metadata properties
Archiving format 246
Attachment Deduplication 246
Blacklist 247
Calculate Expiration Date 247
CM 8.x Confirm Document 248
CM 8.x Create Document 248
CM 8.x Duplicate 249
CM 8.x Update 249
Collector Information 250
Email 250, 254
Email Deduplication 255
File 256
FileNet Image Services Create Document 257
FileNet Image Services File Document in Folder 258
FileNet Image Services Modify Permissions 258
FSC Duplicate Management 259
FSC Metadata 259
FSR Create Document 261
HTTP URL 261
IBM Classification Module 262
P8 4..x Confirm Document 263
P8 4.x Create Document 263
P8 4.x Create Email Instance 264
P8 4.x Declare Record 264
P8 4.x File Document in Folder 264
P8 4.x Link Documents 265
P8 4.x Modify ObjectSecurity 265
P8 4.x Save Prepared Text as XML 265
Re-collection 266
SP Blog 266
SP Collection 267
SP Create File 268
Task status 269
Text Extraction 269
Windows Content Type Information 269
system performance
See monitoring system performance

T

tagPst 437
target connectors 199
Task Route Designer 157
task route service
  checking 176
  configuring 174
  starting 173
task route templates
  Application Archiving 300, 318
  Archiving 301, 318, 336, 354
  Archiving Calendar Entries 302, 319, 337, 355
  Archiving Managed Folders 338, 356
  Archiving to FileNet P8 383
  Archiving to IBM Content Manager 382
  BPM (Business Process Management) 321, 358
  CommonStore Migration - Automatic Archiving 395, 400
  CommonStore Migration - Automatic Stubbing 396, 401
  CommonStore Migration - Interactive Archiving 397, 401
  CommonStore Migration - Offline Repository Support (Archiving) 398, 402
  CommonStore Migration - Offline Repository Support (Stubbing) 399, 403
  Complex Organization (Archiving) 303, 322, 340, 359
  Complex Organization (Stubbing) 304, 323, 341, 360
  Default Archiving (Automatic) 304, 323, 341, 361
  Default Archiving (Interactive) 305, 324, 342, 362
  Default Archiving (Stubbing) 306, 325, 344, 363
  File System with FileNet P8 375
  File System with IBM Content Manager 370
  FS Statistics Collection 372, 377
  FS to CM Archiving (Associate Metadata) 372
  FS to CM Archiving (Delete Metadata Files) 373
  FS to CM8 Archiving (Delete) 373
  FS to CM8 Archiving (Shortcut) 374
  FS to P8 Archiving (Associate Metadata) 377
  FS to P8 Archiving (Declare as Records) 378
  FS to P8 Archiving (Declare as Records) with IBM Classification Module 378
  FS to P8 Archiving (Delete Metadata Files) 380
  FS to P8 Archiving (Delete) 379
  FS to P8 Archiving (Replicate File System and Detect Duplicates) 380
  FS to P8 Archiving (Shortcut) 381
  Journal Archiving (Email Deleted) 307, 326, 344, 363
task route templates *(continued)*
  Journal Archiving (Email Retained) 308, 327, 345, 364
  Life Cycle Template 309, 328, 346, 365
  Lotus Domino with the FileNet P8 314
  Lotus Domino with the IBM Content Manager 297
  Mailbox Statistics Collection 310, 329, 347, 366
  Microsoft Exchange with the FileNet P8 351
  Microsoft Exchange with the IBM Content Manager 333, 382
  NSF Archive Migration - Archiving 310, 330
  overview 283
  PST Migration - Archiving 348, 367
  SharePoint task routes 384
  similar to the IBM CommonStore tasks 393
  SP Audit CM Links 384
  SP Audit P8 Links 385
  SP Manage CM Links 385
  SP Manage P8 Links 386
  SP Statistics Collection 386
  SP to CM - Calculate Expiration Date 387
  SP to CM - Classify 387
  SP to CM - No Versions 388
  SP to CM - With Versions 389
  SP to P8 - Calculate Expiration Date 390
  SP to P8 - Classify 390
  SP to P8 - Declare as Record 391
  SP to P8 - No Versions 392
  SP to P8 - With Versions 392
  Space Saving (All Email) 311, 331, 349, 368
  Space Saving (Large Attachments) 312, 332, 350, 369
  Space Saving (Stubbing) 313, 333, 351, 370
task routes
  bundled data model 275
  configuring 270, 283
  creating 275
  error task routes 280
  exporting 277
  important aspects 275
  importing 275
  order of tasks 284
  overview 270
  placement of tasks 278
  workflow 275
Task Routing Engine
  checking 176
  configuring 174
  starting 173
tasks
  associating content 499
  associating file system metadata 531
  calculating expiration dates 498
  checking for documents
    FileNet P8 540
    IBM Content Manager 504
tasks *(continued)*
  configuring 489
  configuring Content Manager item types 502
  copying 279
  copying PST files to mailbox 516
  creating content elements 542
  creating document instances 548
  creating documents
    Content Manager 506
    file system repository 536
    FileNet P8 543
  creating email stubs 516
  creating FileNet Image Services document 528
  creating records of duplicates 514
  creating temporary file copies 567
  creating version series
    Content Manager 511
    FileNet P8 550
  declaring records 553
  deleting email 564
  detecting duplicates in Content Manager 8.x 509
  extracting attachments 522, 565
  extracting metadata
    email 522
    SMTP 565
  extracting text 527
  file system post-processing 534
  filing documents in folders
    FileNet IS 529
    FileNet P8 555
  finalizing email for compliance 524
  IBM Classification Module 538
  linking documents with their attachments 557
  managing document links 570
  modifying object security 558
  modifying user permissions 530
  order in a task route 284
  overview 496
  placement in task routes 278
  preparing documents for stubbing 526
  preparing email for archiving 525
  preparing email for deletion 567
  retrieving additional archiving informations 540
  saving prepared text as XML 561
  saving temporary file copies 564
  SC Prepare Email for Archiving 566
  SharePoint post-processing 570
  specifying number of versions to retrieve 568
  verifying settings 489
templates
  See also search templates
  See task route templates
temporary metadata database 210
Text Extraction Connector 212
text for stub links 516
text search
  Content Search Engine 675, 677
  FileNet P8 675, 677
  installing 62

text search in an IBM Content Manager repository
See indexer for text search
text search indexer
See indexer for text search
text search indexing
searching for error strings 662
troubleshooting
AFUFetchFile UDF not registered 823
cannot load DB2 library 821
cannot process the item type 820
collection source not supported 816
collectors 815, 816
CommonStore 805
components 815
configuration 806
connection to the FileNet P8 object is lost 814
content greater than 1 MB not indexed 827
creating a configuration database on a remote server 803
CSX legacy restore 811
DB2 installation directory cannot be located 821
envelope journal messages 808
error accessing XML files in afuFetchFile UDF 829, 830
heap size too small 825
incorrect Net Search Extender configuration 819
index cannot be found 827
indexer for text search 819
indexer worker process stops indicating an out of memory situation 830
installation 801
JDBC type 2 driver cannot be loaded 826
list of monitored folders 816
memory issues when running the initial configuration or the set-up tools 805
Microsoft Exchange email attachments in MacBinary format 818
Microsoft Exchange envelope journal messages 808
Microsoft Exchange offline repository 811
Microsoft Exchange Server 2010 816
Net Search Extender fails while indexer process is running 823
Net Search Extender stops abnormally 822
Net Search Extender stops CTE0005 822
Net Search Extender stops without COMMIT 822
no AFUFetchFile UDF configuration found 828
no index found after index recreation 822
not all archived messages are copied to a new offline repository 811
obtaining the DB2 port number 825
troubleshooting *(continued)*
only one process with a PST file at a time 808
Oracle connection fails 804
overview 801
restored docments not displayed in Outlook 811
restoring documents archived using CommonStore for Exchange Server 810
root folders 816
running Outlook Extension on Citrix 810
scale out mode
insufficient permissions 801
primary node failover 801
write access to configuration database 801
searches for DBCS characters fail 816
source systems 807
SQL error SQL0101N 825
SQL error SQL1224N 826
starting Outlook is very slow if you are not connected to the network 810
target repositories 813
textual content not indexed 827
using DB2 Data Server Client Version 9.5 fix pack 1 or 2 805
using different DB2 versions 803
trust relationships 106
trusted certificates 106
types of collectors 458

U

uninstalling 127
upgrading 131, 134
URL protection
dynamic URLs 721
overview 721
static URLs 721
user help 777
user-defined metadata 240
Utility Connector 212
utility connectors 210

V

verifying task settings 489
version series 511, 550

W

Web Application
changing the location 226
installing 97
Lotus Notes mail template 226
modifying settings 225
preparing the installation 96
Web Application Server
configuration tasks 94
embedded 94
existing 94
external 94
port 100
Web Application Server *(continued)*
resynchronization 100
Web Application services APIs
administrative security 742
application security 742, 744
authentication 745
client certificate 745
developing 733
interactive archiving requests 734
key store 745
LDAP authentication 742, 744
message states 737
obtaining client certificates sample procedure 746
REST principles 733
RestoreAPI 737, 740
security role 742, 744
WebSphere Application Server
configuration 95, 97, 98, 100, 803
profile 98, 100
wildcard characters 781
working with archived documents 786
Workplace 754
Workplace XT 787

X

XML
saving prepared text 561

Program Number: 5724-V57

SH12-6914-01