IBM InfoSphere Global Name Scoring

Installation Guide

Version 4 Release 1
Note

Before using this information and the product it supports, read the information in the Notices section.

Edition

This edition applies to Version 4, Release 1 of IBM Global Name Scoring (product number 5724-Q20) and to all subsequent releases and modifications until otherwise indicated in new editions.

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.
Contents

Preface .................................. v
What's new in Version 4.1 .................. v
Contacting IBM Software Support ......... vi

Chapter 1. Overview of IBM InfoSphere
Global Name Scoring ............... 1
Overview of IBM InfoSphere Global Name Reference
Encyclopedia ................................ 1
Product architecture .................... 2
Component APIs .......................... 2
IBM NameWorks .......................... 2
Client applications ...................... 3
NameHunter Server ...................... 3
Distributed Search ...................... 3

Chapter 2. Detailed System
Requirements ......................... 5
System requirements when running on IBM AIX ..... 5
System requirements when running on Linux x86 ... 5
System Requirements when running on Linux x86_64
or AMD Opteron ........................ 6
System requirements when running Linux on zSeries .. 7
System requirements when running on Sun Solaris ... 7
System requirements when running on Microsoft
Windows Server .......................... 8
System requirements when running on Microsoft
Windows Server x86_64 or AMD Opteron ......... 8

Chapter 3. Hardware requirements .... 11
Client workstations ..................... 11
Search server machines ................ 11
Performance planning ................... 12

Chapter 4. Installing and configuring
the product. ............................ 15
Supported upgrade versions .............. 15
Installing IBM InfoSphere Global Name Scoring
IBM InfoSphere Global Name Scoring installation
panel worksheet ......................... 16

Chapter 5. Uninstalling IBM InfoSphere
Global Name Scoring .................. 17

Notices ................................ 19
Trademarks .............................. 23
Terms and conditions .................. 25
Index .................................. 27

© Copyright IBM Corp. 2001, 2009
Preface

The IBM® InfoSphere Global Name Scoring product contains components to identify identical and fuzzy clear text and phonetic name matches effectively, overcoming transliteration, pronunciation, and naming and syntactical scheme issues.

About this publication

This installation guide provides information about how to install and configure IBM InfoSphere Global Name Scoring.

Intended audience

This installation guide is intended for installers, system administrators, and IBM Professional Services personnel to successfully deploy the product in your environment.

Prerequisite and related information

Before using this installation guide, become familiar with the product information available at the following locations:

- IBM Global Name Scoring information center
- Product PDF books on the installation DVD.

You can also install the product information center on a local machine.

How to send your comments

Your feedback is important in helping to provide the most accurate and high-quality information. If you have any comments about this book or any other IBM InfoSphere Global Name Recognition documentation, use the following form to send us your comments:

http://www.ibm.com/software/data/rcf/

What’s new in Version 4.1

This version of IBM InfoSphere Global Name Scoring contains many new features and product enhancements.

For the most recent information about IBM InfoSphere Global Name Scoring Version 4.1, go to the IBM InfoSphere Global Name Recognition Web site located at http://www-306.ibm.com/software/data/ips/products/masterdata/globalname/

Platform additions and deprecations

IBM InfoSphere Global Name Scoring now supports the latest versions of several platforms. Other platforms were deprecated as part of this release. For more information, see the IBM InfoSphere Global Name Scoring system requirements at http://www-01.ibm.com/software/data/globalname/scoring/requirements/index.html?S_CMP=rnav
Contacting IBM Software Support

IBM Software Support provides assistance with product defects.

Before contacting IBM Software Support, your company must have an active IBM software maintenance contract, and you must be authorized to submit problems to IBM. For information about the types of maintenance contracts available, see “Enhanced Support” in the Software Support Handbook at techsupport.services.ibm.com/guides/services.html

Complete the following steps to contact IBM Software Support with a problem:
1. Define the problem, gather background information, and determine the severity of the problem. For help, see the “Contacting IBM” in the Software Support Handbook at techsupport.services.ibm.com/guides/beforecontacting.html
2. Gather diagnostic information.
3. Be prepared to provide the following information in the problem report to assist IBM Software Support:
   - Product name and version
   - Operating system name and version
4. Submit your problem to IBM Software Support in one of the following ways:
   - By phone: For the phone number to call in your country, go to the Contacts page of the IBM Software Support Handbook at techsupport.services.ibm.com/guides/contacts.html

If the problem you submit is for a software defect or for missing or inaccurate documentation, IBM Software Support creates an Authorized Program Analysis Report (APAR). The APAR describes the problem in detail. Whenever possible, IBM Software Support provides a workaround that you can implement until the APAR is resolved and a fix is delivered. IBM publishes resolved APARs on the Software Support Web site daily, so that other users who experience the same problem can benefit from the same resolution.
Chapter 1. Overview of IBM InfoSphere Global Name Scoring

The IBM Global Name Scoring product contains components that can perform both exact and fuzzy matching on personal names, effectively overcoming not only noise and typographical errors, but also differences in orthography, phonetic spelling, script, Romanization, and syntax.

The components within IBM InfoSphere Global Name Scoring enable users to:
- Search for multicultural names and see their most likely matches and name variations
- Match names even if they are affected by typical spelling and cultural variations, related by sound but not by spelling, or damaged by spelling and typing errors
- Match names on both pronunciation and orthography, with the closest matches returned first
- Adjust search parameters for highly tunable and application-specific results
- Separate personal names from organization names
- Compare date values, which can be useful when searching for a name that has an associated date value (such as date of birth), or compute differences between date values

API components and server processes in this bundle include:
- IBM NameWorks, an integrated high-level API
- NameParser®
- NameClassifier™
- NameClassifier - Country of Association
- Country of Association
- NameHunter®
  - NameHunter Server
  - Distributed Search process
- NameSifter
- DateCompare

Overview of IBM InfoSphere Global Name Reference Encyclopedia

IBM Global Name Reference Encyclopedia (GNRE) is a Web-based tool that includes much of the detailed information that is required to perform name analysis work. This functionality compliments the entire suite of IBM InfoSphere Global Name Recognition products and offers valuable cultural information that is important for name analysis.

You can learn about a specific name by entering the parsed name (given name and surname) and clicking the Analyze button. IBM InfoSphere Global Name Reference Encyclopedia displays a list of the countries in which that name occurs most frequently, the likely culture and gender of the name, and possible spelling variations of the name.

You can also learn more about names from a particular culture by exploring the provided cultures, maps, and related country information that is listed in the navigation pane. Choose a culture from the list in order to learn more about the
name parts, name syntax, spelling variations, regional and dialectal differences, gender patterns, and other cultural influences on names.

Product architecture

The product architecture of IBM InfoSphere Global Name Scoring consists of the component APIs, IBM NameWorks APIs, and the client and server applications that communicate with these APIs.

Server applications are applications on the server side that are built upon and provide the functionality of the component APIs. IBM InfoSphere Global Name Scoring includes the NameHunter and Distributed Search server applications. You can also develop your own server applications by using the component and IBM NameWorks APIs.

Component APIs

IBM InfoSphere Global Name Scoring component APIs are C++ libraries that can be integrated into any C++ application.

All of the component APIs perform an analytical function of a single name, but NameHunter and DateCompare take two or more things and compare them. Each of the component APIs are presented in the following list:

NameParser
The NameParser package (com::las_inc::name::parser) parses personal names into their constituent parts (given name and surname).

NameSifter
The NameSifter package (ibmgnr::NameSifter) separates organization names from personal names.

NameClassifier-Country of Association (NC_COA)
The NameClassifier-Country of Association package (ibmgnr::cc class) uses Country of Association (COA) in conjunction with NameClassifier to produce highly accurate results for an associated name culture.

Country of Association (COA)
The COA package (ibmgnr::coa class) references the data that is contained in the IBM Name Data Archive (NDA) to list the countries in which each of the components of a personal name have been observed to occur.

NameHunter
The NameHunter package (LAS::NH) compares pairs of personal and organization names and also searches lists with these name types.

DateCompare
DateCompare (ibmgnr::DateCompare class) compares two date values and returns a similarity score. DateCompare can only compare dates in the Gregorian, 12-month calendar.

IBM NameWorks

IBM NameWorks combines the individual IBM InfoSphere Global Name Recognition components into a single, unified, easy-to-use application programming interface (API), and also extends this functionality to Java applications and as a Web service.

IBM NameWorks comprises a single API class:
Scoring class

Includes the functions that are necessary to compare two names or to search for a name in one or more data lists, along with ancillary tasks such as date comparison and name categorization that might be used to refine search results. Preparation for searching (parsing and culture classification) can be performed separately or included in a search operation.

You can access IBM NameWorks in three ways, either through the C++ functions, Java functions or through Web services. The C++ and Java interfaces can be used directly on any of the supported platforms and the Web service interface can be used either locally or remotely in SOA environments. Any programming environment that can utilize Web services can take advantage of the name analysis and comparison tools provided by IBM NameWorks. Similarly, the Java interface can be used to build custom SOA applications.

Client applications

Client applications are built upon the component APIs or IBM NameWorks. These applications can communicate with server-side applications that are built upon the same framework.

You can use either API package to build applications that display a wide range of physical architectures, ranging from simple standalone solutions that operate on a single host platform, to more complex solutions that operate as independent processes on multiple networked host platforms, such as in a client-server environment. Two major types of client applications exist:

end-user applications

Applications that are built upon the component APIs or the IBM NameWorks package and are compiled to run on the user’s machine.

client-side applications

Client-side applications that communicate with server applications that are built upon the component APIs or the IBM NameWorks package. For example, an IBM NameWorks Web server client that is built from the SOAP APIs.

NameHunter Server

NameHunter Server provides support for multiple data lists to extend the name searching functionality of the NameHunter API. The application accepts query and list management XML messages from a client over a TCP/IP connection.

NameHunter Server is beneficial when loading multiple small data lists (tens of thousands of names), but provides limited performance when loading and searching large data lists. If you need to load and search a large data list, you should use Distributed Search instead.

Attention: NameHunter Server is being replaced by Distributed Search. If you are not already using NameHunter Server, you should use Distributed Search instead because it utilizes additional message interfaces and should be enhanced and supported over time.

Distributed Search

Distributed Search exposes the functionality of the NameHunter API in the form of a single server process that can accommodate complex and performance-intensive...
search requirements due to the size of data lists to be searched or the number of
search transactions that occur at a given time.

Distributed Search is best suited for loading large data lists, comprising millions of
names. However, the application is unable to load multiple data lists into a single
search. This limitation prevents clients from searching multiple data lists from a
single XML message. If you need to load multiple small data lists into a single
server application, you should use the NameHunter Server application.

You can interface with Distributed Search directly or through end-user client
applications and server applications that are built upon IBM NameWorks.
Chapter 2. Detailed System Requirements

This section identifies the required hardware, software, and supported operating systems for IBM InfoSphere Global Name Recognition products from which you can select specific detailed system requirements.

This information describes the minimum product levels that you should have installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, we cannot test every possible configuration. In general, you can install and run with updates to supported products if those updates are forward-compatible.

Your entitlement to support, if any, is dependent upon your license and maintenance agreements for IBM InfoSphere Global Name Management, IBM InfoSphere Global Name Analytics, IBM InfoSphere Global Name Scoring, IBM InfoSphere Global Name Reference Encyclopedia, and is limited to your use of the relevant prerequisite with a supported product.

System requirements when running on IBM AIX

This document contains system requirements when running on the IBM AIX® operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the AIX operating system.

Table 1. System requirements when running on IBM AIX

<table>
<thead>
<tr>
<th>Operating Systems</th>
<th>IBM AIX 5.3L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware Requirements</td>
<td>POWER4™</td>
</tr>
<tr>
<td></td>
<td>POWER5™</td>
</tr>
<tr>
<td>Web Browsers</td>
<td>Mozilla Firefox 2.0 or later (no beta versions)</td>
</tr>
<tr>
<td>Compiler Support</td>
<td>XL C++ v7</td>
</tr>
<tr>
<td></td>
<td>XL C++ v8</td>
</tr>
</tbody>
</table>

System requirements when running on Linux x86

This document contains system requirements when running on a Linux® x86 operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team.

Your entitlement to support, if any, is dependent upon your license and maintenance agreements for IBM InfoSphere Global Name Management, IBM InfoSphere Global Name Analytics, IBM InfoSphere Global Name Scoring, IBM InfoSphere Global Name Reference Encyclopedia, and is limited to your use of the relevant prerequisite with a supported product.
Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the Linux x86 operating system.

**Table 2. System requirements when running on Linux x86**

| Operating Systems       | • Red Hat Enterprise Linux AS, Version 4.0  
|                        | • SuSE Linux Enterprise Server 9          |
| Hardware Requirements   | • Intel® x86 (IA32)                        |
| Web Browsers            | • Mozilla Firefox 2.0 or later (no beta versions) |
| Compiler Support        | • Red Hat Enterprise Linux AS, Version 4.0: Free Software Foundation gcc v3.4.6  
|                        | • SuSE Linux Enterprise Server 9: Free Software Foundation GNU Compiler Collection (GCC) v3.3.3 |

**System Requirements when running on Linux x86_64 or AMD Opteron**

This document contains system requirements when running on a Linux x86_64 or AMD Opteron operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the Linux x86_64 or AMD Opteron operating system.

**Table 3. System Requirements when running on Linux x86_64 or AMD Opteron**

| Operating Systems       | • Red Hat Enterprise Linux AS, Version 4.0  
|                        | • SuSE Linux Enterprise Server 9          |
| Hardware Requirements   | • Intel x86_64  
|                        | • AMD Opteron                         |
| Web Browsers            | • Mozilla Firefox 2.0 or later (no beta versions) |
| Compiler Support        | • Red Hat Enterprise Linux AS, Version 4.0: Free Software Foundation gcc v3.4.6  
|                        | • SuSE Linux Enterprise Server 9: Free Software Foundation GCC v3.3.3 |
System requirements when running Linux on zSeries

This document contains system requirements when running Linux on a zSeries operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the Linux on zSeries operating system.

Table 4. System requirements when running Linux on zSeries

<table>
<thead>
<tr>
<th>Operating Systems</th>
<th>SUSE Linux Enterprise Server 9 for IBM zSeries (s390x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware Requirements</td>
<td>IBM System z® 64-bit s390x</td>
</tr>
<tr>
<td>Web browsers</td>
<td>Mozilla Firefox 2.0 or later (no beta versions)</td>
</tr>
<tr>
<td>Compiler Support</td>
<td>Free Software Foundation gcc v3.3.3</td>
</tr>
</tbody>
</table>

System requirements when running on Sun Solaris

This document contains system requirements when running on the Sun Solaris operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the Sun Solaris operating system.

Table 5. System requirements when running on Sun Solaris

<table>
<thead>
<tr>
<th>Operating Systems</th>
<th>Sun Solaris 9.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sun Solaris 10.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hardware Requirements</th>
<th>UltraSPARC III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UltraSPARC IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web browsers</th>
<th>Mozilla Firefox 2.0 or later (no beta versions)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Compiler Support</th>
<th>Sun Studio 11</th>
</tr>
</thead>
</table>
System requirements when running on Microsoft Windows Server

This document contains system requirements when running on the Microsoft® Windows® Server operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the Microsoft Windows Server operating system.

Table 6. System requirements when running on Microsoft Windows Server

<table>
<thead>
<tr>
<th>Operating Systems</th>
<th>• Microsoft Windows Server 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware Requirements</td>
<td>• Intel x86 (IA32)</td>
</tr>
<tr>
<td>Web Browsers</td>
<td>• Microsoft Internet Explorer v6.0.2900 and v7.0</td>
</tr>
<tr>
<td></td>
<td>• Mozilla Firefox 2.0 or later (no beta versions)</td>
</tr>
<tr>
<td>Software Requirements</td>
<td>• Microsoft Visual C++ 2005 Redistributable Package for x86</td>
</tr>
<tr>
<td>Compiler Support</td>
<td>• Microsoft Visual Studio 2003</td>
</tr>
<tr>
<td></td>
<td>• Microsoft Visual Studio 2005</td>
</tr>
</tbody>
</table>

System requirements when running on Microsoft Windows Server x86_64 or AMD Opteron

This document contains system requirements when running on the Microsoft Windows Server x86_64 or AMD Opteron operating system.

This document describes the minimum product levels that need to be installed before opening a problem report with the IBM InfoSphere Global Name Recognition support team. Because other products frequently ship fixes, updates, and new releases, testing every configuration is not possible. In general, you can install and run with updates to supported products if those updates are forward compatible.

The following list identifies the products that are supported when IBM InfoSphere Global Name Scoring runs on the Microsoft Windows Server x86_64 or AMD Opteron operating system.

Table 7. System requirements when running on Microsoft Windows Server x86_64 or AMD Opteron

<table>
<thead>
<tr>
<th>Operating Systems</th>
<th>• Microsoft Windows Server 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware Requirements</td>
<td>• Intel x86_64</td>
</tr>
</tbody>
</table>
| **Web Browsers** | • Microsoft IE v6.0.2900 and v7.0 or later  
• Mozilla Firefox 2.0 or later (no beta versions) |
|------------------|--------------------------------------------------|
| **Software Requirements** | • Microsoft Visual C++ 2005 SP1  
Redistributable Package for x64 |
| **Compiler Support** | • Microsoft Visual Studio 2005 SP1 |
Chapter 3. Hardware requirements

Hardware requirements vary for IBM InfoSphere Global Name Scoring and depends upon the network and the actual client workstations and server machines that run the applications.

Client workstations

Client workstations run the IBM InfoSphere Global Name Scoring client end-user and client-side applications.

The WebSphere Application Server machine hosts the WebSphere application server instance of the IBM NameWorks Web Service. The minimal requirements to support WebSphere application server are included in the product package.

Physical hardware requirements for client workstations vary for each of the two application types:

end-user applications
Physical hardware requirements vary for end-user applications. These requirements are entirely dependent on how the end-user application is designed to function. Each end-user application has its own set of requirements that are dependent not only on the resources that the component APIs or the IBM NameWorks package use, but also depend on the other functions that the application performs.

client-side applications
Because processing occurs on the server machine, physical hardware requirements for client-side applications that communicate with server applications are minimal. These applications provide the ability to support the communication protocol (TCP/IP or Web services) that connects with the associated server process, as well as the ability to produce messages that are expected by the target server process, such as XML messages and Web service requests.

Search server machines

The search server machines host the NameHunter Server and Distributed Search server applications. Much of IBM’s name-searching functionality is exposed through server processes, which allow greater architectural flexibility, easy scaling, and efficient allocation of processing resources.

Each server application has its own set of server machine requirements:

NameHunter Server
NameHunter Server is memory and processor intensive. NameHunter Server can be multi-threaded; however each thread requires a processor to be useful.

Each NameHunter Server process must load its entire corresponding data list into memory. If a process has to use virtual memory, performance will be severely degraded. Estimating the amount of physical memory consumed by each NameHunter Server process is affected by many factors, including the average length of the input names and whether or not regularization is used.
If you know how many names are in the entire data list, you can use the following equation to estimate how much memory each NameHunter Server requires:

\[ 50 \text{ MB} + [\# \text{ of names in the entire data list}] \times 180 \text{ bytes} \]

= physical memory required

50 MB is the amount of memory required if you load all NameHunter support files. 180 bytes is a conservative estimate of the memory required per name in the data list.

**Distributed Search server**

A Distributed Search server application is memory and processor intensive. A Distributed Searcher process, when processing a search request, consumes 100% of a processor’s capacity. A Distributed Search server machine should have at least as many processors as there are Distributed Searcher processes.

Each Distributed Searcher process must load its entire corresponding data list into memory. If a process must use virtual memory, performance will be severely degraded. Estimating the amount of physical memory consumed by each Distributed Searcher process is affected by many factors, including the average length of the input names, whether or not regularization is used, and how the names have been converted by NamePreprocessor.

If you know how many names are in each Distributed Searcher data list, you can use the following equation to estimate how much memory the Distributed Search server requires:

\[ 100 \text{ MB} + [\# \text{ of names in each data list}] \times 180 \text{ bytes} \times [\# \text{ of Distributed Searcher processes}] \]

= physical memory required

100 MB is the amount of memory required if you load all Distributed Search support files. 180 bytes is a conservative estimate of the memory required per name in the data list.

**IBM InfoSphere Global Name Scoring Version 4.1 memory requirements**

Required memory per name in each data list has increased from 150 bytes in version 3.2 to 180 bytes in version 4.1. Load times are accordingly slower.

Configuring organizational name searches to include personal name data can take twice as much memory and load time as organizational name data alone, depending on the ratio of the two types of data. For example, 1 million organizational names and 1 million personal names in a search would require 100% more load time than 2 million organizational names only.

**Performance planning**

Performance for the IBM InfoSphere Global Name Recognition products are hardware dependent.

**Performance factors**

Performance and throughput for IBM InfoSphere Global Name Recognition products are typically proportional to three key factors:

- Number of processors available
- Clock speed of the processor used
- RAM resources

Applications that involve the IBM InfoSphere Global Name Analytics product generally require fewer processor cycles and RAM, while applications that involve the IBM InfoSphere Global Name Scoring product require a larger number of processors, processor cycles, and RAM.

When using IBM NameWorks on a Microsoft Windows x86 installation, you should change the default JVM RAM space setting to 150 MB to prevent runtime errors.

**Performance considerations for version 4.1**

Loading large collections of organizational names can take as much as ten times longer than loading personal names only. For example, if loading 1 million personal names takes 3-4 minutes, loading 1 million organizational names can take 30 to 40 minutes.

Configuring organizational name searches to include personal name data can take twice as much memory and load time as searching organizational name data alone, depending on the ratio of the two types of data. For example, 1 million organizational names and 1 million personal names in a search would require 100% more memory and load time than 2 million organizational names only.

Name Preprocessor can take over 30 hours to preprocess 200 million Personal names. If your data list consists of only personal names, then you can set `doCategorize = false` in the npp.config file to instruct Name Preprocessor to skip name categorization, effectively reducing processing time.
Chapter 4. Installing and configuring the product

You can install and configure IBM InfoSphere Global Name Scoring using the following steps.

Supported upgrade versions

You can use the product installer to upgrade only versions 3.1 or greater. If you are upgrading from an earlier version, please contact IBM Software Support.

When upgrading your installation, the installation program will only upgrade the components already installed in the previous version.

Installing IBM InfoSphere Global Name Scoring

You must complete the following steps to start the IBM InfoSphere Global Name Scoring installation program.

You must run the installer from the product media, or copy the product installer package including the executable to a local drive. The product installer cannot be run from a network drive.

1. Obtain the IBM InfoSphere Global Name Scoring product media.
2. Run the install program:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>in GUI mode from a command line</td>
<td>• Navigate to the /Disk1/InstData/VM/ directory on the product media.</td>
</tr>
<tr>
<td></td>
<td>• Run the install executable.</td>
</tr>
<tr>
<td>in command line mode</td>
<td>• Navigate to the /Disk1/InstData/VM/ directory on the product media.</td>
</tr>
<tr>
<td></td>
<td>• Run the install executable with the -i console option.</td>
</tr>
</tbody>
</table>

3. Follow the instructions on the installation program wizard.
4. To verify that your installation was successful, check for error messages in the following directory:
   /<product_install_location>/installer/logs/
5. Optional: If you have selected to install Web services, run the following script to start the Embedded WebSphere Application Server:

   Windows system
   /<product_install_location>/startGNR.bat

   UNIX system
   /<product_install_location>/startGNR.sh

   Note: If the startGNR script does not exist in your installation directory, check for errors in the installation and configuration log files, which can be found in the following directory:
   /<product_install_location>/installer/logs/
IBM InfoSphere Global Name Scoring installation panel worksheet

These worksheets include all of the installation panel settings. Use this worksheet to keep a record of your settings.

Embedded WebSphere Application Server panel

Table 8. Embedded WebSphere Application Server panel

<table>
<thead>
<tr>
<th>Setting</th>
<th>Description</th>
<th>My setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully Qualified Host Name</td>
<td>The fully qualified hostname of the server or the IP address of the server hosting the embedded WebSphere Application Server.</td>
<td>(Default setting: 14520)</td>
</tr>
<tr>
<td>Memory used during deployment (MB)</td>
<td>(Default setting: 1024)</td>
<td></td>
</tr>
<tr>
<td>Web server port number (http)</td>
<td>(Default setting: 14520)</td>
<td></td>
</tr>
<tr>
<td>Secure web server port number (https)</td>
<td>(Default setting: 14521)</td>
<td></td>
</tr>
<tr>
<td>Administration port number</td>
<td>(Default setting: 14522)</td>
<td></td>
</tr>
<tr>
<td>Secure administration port number</td>
<td>(Default setting: 14523)</td>
<td></td>
</tr>
<tr>
<td>SOAP port number</td>
<td>(Default setting: 14524)</td>
<td></td>
</tr>
<tr>
<td>Application server port number</td>
<td>(Default setting: 14525)</td>
<td></td>
</tr>
</tbody>
</table>

Documentation panel

Table 9. Documentation panel

<table>
<thead>
<tr>
<th>Setting</th>
<th>Description</th>
<th>My setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documentation port</td>
<td>The TCP port number to install the IBM InfoSphere Global Name Scoring documentation information center on.</td>
<td>(Default setting: 14526)</td>
</tr>
</tbody>
</table>
Chapter 5. Uninstalling IBM InfoSphere Global Name Scoring

You remove the product installation by running the uninstallation program.

1. Navigate to the <install_location>/_uninst directory.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| On an AIX or UNIX-based system | Enter the following command:        
GNS.bin                        |
| On a Microsoft Windows system | Run the GNS.exe executable to start the uninstallation program, or use the Windows Add or Remove Programs option. |

2. Follow the instructions on the uninstall program wizard.
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:

IBM World Trade Asia Corporation
Licensing 2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, 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 Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

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 Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

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 is for planning purposes only. The information herein is subject to change before the products described become available.

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.
Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.
Trademarks

IBM trademarks and certain non-IBM trademarks are marked on their first occurrence in this information with the appropriate symbol.

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., 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 [www.ibm.com/legal/copytrade.shtml].

The following terms are trademarks or registered trademarks of other companies:

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.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

ITIL is a registered trademark and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACSLink, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACSLink licensee of the United States Postal Service.

Other company, product, or service names may be trademarks or service marks of others.
Terms and conditions

Permissions for the use of these publications is granted subject to the following terms and conditions.

**Personal use:** You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

**Commercial use:** You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.
Index

A
architecture, product
  IBM InfoSphere Global Name Scoring 2

C
client applications 3
client workstations 11
comments
  sending v
component
  APIs 2
contacting
  IBM software support vi

D
description 2
Distributed Search
  server applications 4

F
features
  new in version 4.1 v

H
hardware requirements
  IBM InfoSphere Global Name Scoring 11

I
IBM InfoSphere Global Name Reference Encyclopedia
  Encyclopedia 1
IBM InfoSphere Global Name Scoring
  Description 1
IBM NameWorks 2
  APIs 2
IBM software support
  contacting vi
  installation
  uninstalling 17
installation program
  starting 15
  system requirements
    worksheet 16
  installing 15

L
legal guidelines
  publications
    reusing 25
  terms and conditions 25

N
NameHunter
  server applications 3
  new in version 4.1 v

O
overview
  IBM InfoSphere Global Name Reference Encyclopedia 1
  IBM InfoSphere Global Name Scoring 1

P
performance
  planning 12
physical architecture 11
prerequisite information
  version 4.1 v

R
related information v

S
search server machines 11
  sending comments v
support
  contacting vi
  system requirements
    details 5
  IBM AIX 5
  Linux AMD Opteron 6
  Linux x86 5
  Linux x86_64 6
  Linux, zSeries 7
  Microsoft Windows Server 8
  Sun Solaris 7

T
trademarks 23

U
uninstalling 17
upgrading
  supported versions 15