

IBM DB2 Information Integrator
OmniFind Edition



Release Notes

Version 8.2 Fix Pack 1

IBM DB2 Information Integrator
OmniFind Edition



Release Notes

Version 8.2 Fix Pack 1

Before using this information and the product it supports, be sure to read the general information under "Notices."

This document contains proprietary information of IBM. It is provided under a license agreement and Copyright law protects it. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

You can order IBM publications online or through your local IBM representative:

- To order publications online, go to the IBM Publications Center at www.ibm.com/shop/publications/order
- To find your local IBM representative, go to the IBM Directory of Worldwide Contacts at www.ibm.com/planetwide

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 2004, 2005. All rights reserved.

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

Contents

1	About the release notes	v
1	Technical support information on the Web	v
1	Installing DB2 II OmniFind Edition,	
1	Version 8.2.1.	1
1	Uninstalling DB2 II OmniFind Edition	3
1	New in this release	5
1	New supported operating system and data source	
1	type	5
1	Specifying a custom collection ID	5
1	Crawling password-protected Web sites	5
1	Crawling framesets in Web pages	8
1	Configuring security tokens from META tags in Web	
1	documents	8
1	Indexing the anchor text in links to forbidden	
1	documents	9
1	Overriding no-index and no-follow directives in Web	
1	pages	9
1	Configuring a global crawl space for Web crawlers	10
1	Configuring multiple Lotus Notes servers	12
1	Specify the Notes server to crawl page	12
1	Performance enhancements for the VeniceBridge	
1	crawler	14
1	Specifying login information for Windows file	
1	systems	15
1	External name is now Display name	15
1	Configuring and monitoring the data listener	15
1	System view: Data Listener page	16
1	Monitor system view: Data Listener page	17
1	Data listener details page	17
1	Stopping and starting the data listener	
1	component	18
1	Estimating the number of documents in a collection	18
1	Checking system resources is automatic with the	
1	collection wizard	19
1	Monitoring parser activity	19
1	Parser details page	19
1	Mapping HTML metadata to search fields	21
1	HTML metadata field mappings page	22
1	Removing new line characters from white space in	
1	Japanese and Chinese documents	23
1	Configuring document types for a collection parser	
1	and Stellent session	23
1	Configure document types for the collection	
1	parser	24
1	Configure document types for the Stellent session	26
1	Enhancements to query processing	30
1	Collapsing results from the same Web site	30
1	Extended highlighting in search results	30
1	Timeout support in search applications	31
	Known limitations, problems, and	
	workarounds	33
	Unpacking the DB2 UDB for AIX tar file	33
	Shared memory segments in AIX	33
	Restart the system after you uninstall on Windows	33
	Orphan URLFetcher process for Linux and AIX	34
	Index location on the server	34
	Heap size limits	35
	DB2 UDB tables with large objects	35
	Index status does not display	35
	User ID permissions for Linux and AIX	35
1	Enabling direct document access for FileNet Content	
1	Services	35
1	Limitations for crawling nickname tables	35
	Notices	37
	Trademarks	39
	Contacting IBM	41
	Obtaining product information	41
	Providing comments on the documentation	41

1

1 **About the release notes**

1 The release notes document describes what is new in this release, known
1 limitations, problems, workarounds, and documentation updates for IBM® DB2®
1 Information Integrator OmniFind™ Edition, Version 8.2.1 (version 8.2 plus Fix Pack
1 1).

1 IBM DB2 Information Integrator OmniFind Edition, Version 8.2.1 was not
1 translated. Changes made to the product interfaces and messages are provided in
1 English only.

1

1 **Technical support information on the Web**

1 The DB2 II OmniFind Edition product was recently renamed WebSphere®
1 Information Integrator OmniFind Edition. You might see references to WebSphere
1 Information Integrator OmniFind Edition on product-related Web pages, but the
1 product interfaces, documentation, and troubleshooting information (technotes) still
1 reflect the DB2 brand.

1 You can find technical support information on the following Web sites:

1 **IBM DB2 Information Integrator OmniFind Edition**

1 www.ibm.com/software/data/integration/db2ii/supportomni.html

1 **IBM WebSphere Information Integrator**

1 <http://www.ibm.com/software/data/integration/db2ii/support.html>

1 **IBM WebSphere Application Server**

1 <http://www.ibm.com/software/webservers/appserv/was/support/>

1 **IBM DB2 Universal Database™ for Linux™, UNIX®, and Windows® Version 8**
1 **products**

1 www.ibm.com/software/data/db2/udb/winos2unix/support

1

1 Installing DB2 II OmniFind Edition, Version 8.2.1

1
1
1

With the DB2 II OmniFind Edition Installation Launchpad, you can install the full product (version 8.2.1) or install only the version 8.2.1 upgrade on your existing installation.

1
1
1
1
1

Before you install the version 8.2.1 upgrade, you should back up your existing installation. See *Administering Enterprise Search* for information about backing up and recovering your system. During installation, specify the path to the existing data directory. The Installation Launchpad will use the existing installation information to install the upgrade.

1
1
1
1
1

If you want to install the full version 8.2.1 product, and you previously installed DB2 II OmniFind Edition version 8.2, you must first uninstall DB2 II OmniFind Edition. You cannot install the full product on top of an existing installation, nor can you back up your existing system, install version 8.2.1, then restore version 8.2 configuration settings to a version 8.2.1 system.

1

To install version 8.2.1:

1

1. Start the Installation Launchpad.

1

Option	Description
Linux	Log in as the root user and type <code>./install-linux.bin</code>
AIX	Log in as the root user and type <code>./install-aix.bin</code>
Windows	You must have administrator authority. Double-click the file <code>install-win32.exe</code> or type <code>install-win32.exe</code>

1
1

1
1

1
1
1
1

1

2. After the Installation Launchpad opens, click **Install Products**.

1

3. To install the version 8.2.1 upgrade:

1
1
1

- a. Select **Upgrade an existing DB2 II OmniFind Edition installation** in the upgrade installation window. If you select this option, your existing version 8.2 system configuration will be migrated to a version 8.2.1 system.

1

- b. Type the path to the existing data directory. Click **Next**.

1

- c. Review the settings on the Summary window. Click **Next** to start the installation.

1

4. To install the full version 8.2.1 product:

1
1
1

- a. Select **Install a new version of DB2 II OmniFind Edition**. If you previously installed DB2 II OmniFind Edition version 8.2, be sure to uninstall it before you select this option.

1

- b. Follow the procedures in *Installation Guide for Enterprise Search* to install the product.

1

1

1 Uninstalling DB2 II OmniFind Edition

1

The uninstallation program will remove the enterprise search components. You can also remove all data system configuration.

1

1

You should back up your system before you uninstall DB2 II OmniFind Edition.

1

For a multiple server configuration, run the uninstallation program on each server. To uninstall DB2 II OmniFind Edition:

1

1

1. For Linux and AIX, log in as the root user.

1

2. Browse or change to the _uninst directory in the installation directory (ES_INSTALL_ROOT) and start the uninstallation program:

1

1

Option	Description
Linux or AIX	Type ./uninstall.bin
Windows	Double-click the file or type uninstall.exe

1

1

1

3. Optional: To remove all the enterprise search administrator configuration information, select the Remove all data and configuration files check box.

1

1

Attention: If you select this check box, you will remove all data for all collections on the system.

1

1

On Windows, you must restart the computer before you can reinstall DB2 II OmniFind Edition.

1

1

1 New in this release

1

New features and functions for DB2 II OmniFind Edition include improvements to search administration, enhancements to search quality and performance, and support for Microsoft® Windows 2003.

1

1

1 New supported operating system and data source type

1

DB2 II OmniFind Edition supports the 32-bit Windows Server 2003 operating system (Standard Server, Advanced Server, and Datacenter Server), and can crawl Microsoft SQL Server databases.

1

1

1

To install Windows Server 2003, follow the same instructions for Windows Server 2000 in the *Installation Guide for Enterprise Search*.

1

1

To crawl SQL Server databases, use the enterprise search administration console to create a DB2 crawler and configure the data sources that you want to crawl.

1

1

1 Specifying a custom collection ID

1

When you create a collection, you can choose to use the internal identifier that the system creates for you, or you can specify a custom collection identifier.

1

1

If you specify a custom collection identifier, your search applications can use enterprise search application programming interfaces to call the collection with this identifier instead of the potentially cryptic identifier that the system creates.

1

1

1

If you specify a custom collection identifier, you must do so when you create the collection. You cannot later edit the collection and change this identifier.

1

1

To specify the type of collection identifier that you want to use:

1

- 1 Start the enterprise search administration console. If security is enabled, log in as a user with enterprise search administrator authority.
- 1 Click **Create Collection** or **Collection Wizard**, depending on whether you want to use the wizard to create a collection.
- 1 Select one of the following options in the **Collection ID** field:

1

1

1

1 **Default collection ID**

1

Select this option to allow the system to create an internal identifier for the collection.

1

1

1 **Custom collection ID**

1

Select this option if you want to specify your own identifier for the collection. The value that you specify can contain only alphanumeric characters and the underscore character.

1

1

1

1 Crawling password-protected Web sites

1

DB2 II OmniFind Edition Web crawlers can crawl password-protected Web sites by using HTTP basic authentication and by retaining information in cookies.

1

HTTP basic authentication support

As the enterprise search administrator, you can identify password-protected Web sites that you want to crawl and specify the required credentials in a configuration file named `httpauth.ini` before you start the crawl.

The `httpauth.ini` file must follow these formatting rules:

- The content of the `httpauth.ini` file is text and cannot contain nonprinting characters except for white space. Data entries are separated by new lines.
- The file can contain authorization entries, comments, and blank lines in any order.
- A comment is a line whose first nonwhite space character is `#`.
- Comments cannot appear on a line with non-comment text.
- Each authorization entry must contain the following information in the following order on a single line:
 - A target URL (the root of the server file system for which authorization is needed)
 - A realm name that is meaningful to the server
 - A user name for logging in
 - A password for logging in
 - A reauthentication URL
 - A list of semicolon-separated security tokens, which can be empty, but the field must still be present, with a comma separator after the fifth field
- Fields on a line are separated by commas. White space is permitted around the commas.
- The password in each entry is encrypted with the same encryption method that is used in the rest of DB2 II OmniFind Edition.
- Authorization for a URL is assumed valid for all URLs below the specified URL in the server's file system.

You can modify the `httpauth.ini` file by using a command line utility called `httpauthedit`. With this utility, you can type passwords in clear text (unencrypted) when you create or edit an entry, but the utility encrypts the passwords before writing them to disk. To change a password, you must delete an entry and add it again.

To start the `httpauthedit` utility, type the following command on one line:

```
java -classpath ES_INSTALL_ROOT/lib/URLFetcher.jar:esinstall/lib/es.oss.jar  
com.ibm.es.wc.uf.HTTPAuthEdit configdir
```

ES_INSTALL_ROOT is the DB2 II OmniFind Edition installation directory, and *configdir* is the master configuration directory of the crawler instance where the `httpauth.ini` file is to be created (or modified).

The master configuration directory is in `ES_NODE_ROOT/master_config/CR_INST_DIR`, where *ES_NODE_ROOT* is the value of `NodeDirectory` in the `ES_CFG` file, and *CR_INST_DIR* is the directory that was created for a new Web crawler instance (which is given an arbitrary name by the administration software). Be sure that you operate on the correct instance configuration. The changes will not affect the crawler until the next time it is started.

The httpauthedit utility accepts the following commands:

Table 1. Commands accepted by the HTTPAUTHEDIT utility

Command	Description
?	Help: Describes the command set.
b	Browse: Prints the contents of the file.
a	Add: Adds a new authorization entry to the file. This command requires five or six arguments that are comma-separated on a single line. For example: http://www.ibm.com/solutions/, solutions, wwwsearch, foobar, xyz123, http://www.ibm.com/solutions/, group1;group2;admin3 <ul style="list-style-type: none">• Target URL: http://www.ibm.com/solutions/• Realm name: solutions• User name: wwwsearch, foobar• Password: xyz123• Reauthentication URL: http://www.ibm.com/solutions/• Optional: Security tokens: group1;group2;admin3
d	Delete: Removes an authorization entry from the file. This command requires one argument: the target URL of the authentication entry that is to be removed from the file, such as http://www.ibm.com/solutions/.
q	Quit: Exits the program.

The reauthentication URL can be the same as the target URL:

- The reauthentication URL must be a URL that, if requested without authorization, triggers a 401 response from the server, with an HTTP basic challenge, for the realm for which the client wants to submit authorization.
- The reauthentication URL must trigger this response without redirection (the reauthentication URL must be the true destination page, not a 301, 302, and so on, page that redirects the client to another page). This reauthentication URL is sometimes not the same as the root URL of the realm that it authorizes. Whether the reauthentication URL is the same as the root URL depends on the internal logic of the Web site. The Web site administrator can provide details.

Leading and trailing white space will be removed from field values. If you type invalid input, httpauthedit shows an error message.

Changes are saved when you type the quit command. There is no explicit save command. To cancel a change, you can kill the program by exiting without using quit.

Cookies

The Web crawler will retain cookies that are received from servers and use them for the life of the crawler instance. When the crawler shuts down, it saves all unexpired cookies to the disk and reloads them at the start of the next crawling session. The cookie file is called cookies.ini, and it is in the crawler data directory.

Cookie handling in the crawler is independent of HTTP authentication (the httpauth.ini file), but if cookies are being used by Web servers to protect content, authentication might also be required, even with manually supplied cookies. Web site administrators should be able to supply information about what is required to crawl their sites.

1 As the enterprise search administrator, you can add cookies in addition to the ones
1 that were found by the crawler by using a text editor to append them to the
1 cookies.ini file. Because this file is rewritten by the crawler every time it shuts
1 down, keep your manually entered cookies in another file, and add them if needed
1 before starting a crawl. (The crawler does not discard unexpired cookies, but if a
1 problem prevents the writing of the entire cookie collection, you do not want to
1 lose the cookies that you manually specified.)

1 To specify cookies for the Web crawler, you edit the cookies.ini file. In the
1 cookies.ini file, each line is an entry. Blank lines and comments are permitted, but
1 they will not be preserved. Each entry has the following format. Each entry must
1 be on one line:

```
1 Cookie(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSIDNFLSNL;path=/  
1 https://www.ibm.com:443/help/solutions/
```

1 *Cookie* is a required keyword indicating the start of an entry. The number 53 is the
1 length in characters of the cookie's name and value pair and any other content
1 (path, secure setting, and so on) to be sent with the cookie. The number 40 is the
1 length of the associated URL, for validation.

1 For ASPSESSION. . . , these 53 characters are the content of the cookie to be sent to
1 the originating server. This string is followed by a comma (,) separator.

1 The URL `https://www.ibm.com:443/help/solutions/` is used to validate the cookie
1 by supplying a domain name, for example. With manually added cookies, you
1 should include a validating URL that will satisfy the security and privacy
1 restrictions on cookies.

1 **Crawling framesets in Web pages**

1 The Web crawler is now able to crawl Web pages that were created with HTML
1 frames.

1 In DB2 II OmniFind Edition Version 8.2, the Web crawler was not able to parse
1 framesets or extract links from frames to other HTML pages.

1 **Configuring security tokens from META tags in Web documents**

1 New security options for the Web crawler enable you to specify the names of
1 META tags in a configuration file. The contents of those tags are used as security
1 tokens for controlling access to the documents in which the tags are found.

1 In DB2 II OmniFind Edition, Version 8.2, security tokens were globally defined,
1 and they applied to all documents or none, according to whether document-level
1 security was enabled for the Web crawler.

1 You can now control access to documents by specifying META tags as security
1 tokens. To use META tag-based security, you must enable document-level security
1 for the Web crawler, and you must edit a configuration file to specify the META
1 tag names (you cannot specify the META tag names in the administration console).

1 To specify the META tags that you want to use as security tokens:

- 1 1. Log in on the crawler server as the enterprise search administrator (this user ID
1 and password were configured during the installation of DB2 II OmniFind
1 Edition).

1 2. Use a text editor to edit the `crawl.properties` file in the `master_config`
1 directory tree.

1 3. If necessary, insert the following parameter: `security_meta_tags`. (If the
1 `security_meta_tags` parameter already exists, the default value is `nil`.)

1 4. Specify the META tag names in a comma-delimited list. For example:
1 `security_meta_tags=security,acl,access`

1 **Rules for tag names:**

- 1 • Tag names read from the configuration file are converted to lowercase.
- 1 • Tag names seen in crawled documents are converted to lowercase before
1 comparison with the names in the configuration file.
- 1 • The only separator character in the name list is the comma (white space does
1 not delimit a name).
- 1 • Embedded blanks in names are retained; leading and trailing blanks are
1 removed.

1 5. Save and exit the file.

1 If tag names are configured for the `security_meta_tags` parameter, then during the
1 processing of every downloaded HTML page, the document's META tags will be
1 checked against the `security_meta_tags` list.

1 For complete information about using security tokens to control access to
1 documents, see the topic *Document-level security* in *Administering Enterprise Search*.

1 Indexing the anchor text in links to forbidden documents

1 If a document includes links to documents that the Web crawler is not allowed to
1 crawl, you can specify whether you want to retain the anchor text for those links
1 in the index.

1 Settings in a `robots.txt` file or in metatag specifications can prevent the Web
1 crawler from accessing certain documents on a Web site. If a document that the
1 Web crawler is allowed to crawl includes links to forbidden documents, you can
1 specify how you want to handle the anchor text for those links on the Advanced
1 Web crawler properties page:

- 1 1. If you want to include the anchor text for links to forbidden documents in the
1 index, select the **Index the anchor text in links to forbidden documents** check
1 box. (Anchor text is the descriptive text that appears between the `<A>` and ``
1 tags in Web documents.) If you select this check box, forbidden pages can then
1 be found by searching for text that is in the anchor text for links that point to
1 them.
- 1 2. If you want to exclude the anchor text for the forbidden links from the index
1 (along with excluding the forbidden documents), clear the **Index the anchor
1 text for links to forbidden documents** check box.

1 You can specify Advanced Web crawler properties when you create a Web crawler
1 or by editing an existing Web crawler's crawl space.

1 Overriding no-index and no-follow directives in Web pages

1 You can specify rules in a configuration file to control whether the Web crawler
1 follows links to pages, or indexes pages, that contain no-follow or no-index
1 directives. There is no support for this feature in the enterprise search
1 administration console.

1 Some Web pages are designated with no-index or no-follow directives, which
1 instruct robots (such as the Web crawler) to not include the contents of those pages
1 in an index, not follow links found in those pages, or to not do either of these
1 actions.

1 In DB2 II OmniFind Edition version 8.2, these restrictions are enforced the Web site
1 administrator or content creator. Having control over these settings, however, can
1 improve the quality of the crawl. For example, there are sets of directory pages
1 that contain thousands of links, but no other useful content; those pages should be
1 crawled, and their links followed, but the pages themselves should not be indexed.

1 There are also times when the crawler administrator wants the crawl to go no
1 lower in a hierarchy, but the desired leaf pages contain links, and do not contain a
1 no-follow directive. It may or may not be possible to get the content owners of
1 such pages to insert the required directives because some of these pages are
1 autogenerated and have no owners.

1 The version 8.2.1 crawler looks for a new configuration file, called
1 followindex.rules, in its configuration directory. This file can contain any number
1 of follow and index rules:

```
1 forbid follow URLprefix  
1 allow follow URLprefix  
1 forbid index URLprefix  
1 allow index URLprefix
```

1 You cannot specify IP addresses and DNS host names in the follow and index
1 rules; you can specify URL prefixes (including wildcards) only. Order is significant,
1 just as is when you specify crawling rules. If a rule is defined that explicitly allows
1 or forbids following or indexing, it overrides other settings, including those in the
1 target document.

1 **Configuring a global crawl space for Web crawlers**

1 You can specify rules in a configuration file to create a global crawl space for Web
1 crawlers and better control the removal of URLs from the index. There is no
1 support for this feature in the enterprise search administration console.

1 Each Web crawler is configured with a crawl space that defines the URLs that are
1 to be crawled or not crawled. Discovered URLs that are in the crawl space are
1 retained (in a database) for later crawling; those that are not in the crawl space are
1 discarded. If the crawler starts with an empty database, the crawl space definition
1 and database remain consistent as long as the crawler runs.

1 Sometimes a crawler is stopped, and its crawl space is reduced (by adding new
1 rules that forbid pages). When the crawler is restarted, its crawl space definition
1 and database will be inconsistent: the database will contain URLs - some crawled,
1 some not - which are not in the new, smaller crawl space. In DB2 II OmniFind
1 Edition Version 8.2, the Web crawler restores consistency by changing the HTTP
1 return codes for these URLs to 760 (crawl space exclusion) and writing instructions
1 to remove the now-excluded pages from the index. It does this because it assumes
1 that it is the only Web crawler in the collection, and that its crawl space identifies
1 all Web documents that are allowed to be in the index.

1 Sometimes, after a crawler has been operating on a large crawl space for a while,
1 the administrator decides to divide the crawl space among two or more crawlers
1 (perhaps to ensure some parts are crawled more often than the rest). New crawlers

1 are created, each having its own independent database tables (initially empty), and
1 each having a different part of the original crawl space. The original crawler's
1 crawl space is then reduced to whatever is left after removing the parts to be
1 crawled by the new crawlers.

1 If the original crawler is a Version 8.2 crawler, it will restore consistency to its
1 database in the manner described above. This is not what the administrator wants:
1 the moved URLs are not to be removed from the index; they are simply being
1 crawled by other crawlers.

1 DB2 II OmniFind Edition Version 8.2.1 provides a second, higher-level crawl space,
1 which represents URLs that are not to be crawled by the current crawler, but
1 which are not to be deleted, either. (URLs that are not in either space continue to
1 be rejected during discovery, and removed from the index when re-scanned, just as
1 before.)

1 The second, global crawl space is defined by the contents of a file called
1 global.rules, which is in the crawler's configuration directory. If it exists, the file is
1 read in during crawler initialization. If there is no global.rules file, the crawler
1 operates with a single-level crawl space (as in Version 8.2). If a global space is
1 defined, the crawler will rule URLs in or out as before, but will only send
1 instructions to remove a URL from the index if the URL is not in either space.

1 The global.rules file has the same syntax as the local crawl.rules file, but only
1 domain name rules are significant. (This enables a crawl space to be partitioned
1 between crawlers only by using DNS host names, not IP addresses or URL prefix
1 patterns.)

1 The presence of a global.rules file enables (and its absence disables) the global
1 crawl space feature. URLs that are excluded by URL prefix or IP address rules in
1 the local crawl space (as defined in the crawl.rules file) are unaffected by the
1 global crawl space; they are still excluded. The global crawl space can be used only
1 to prevent the removal from the index of URLs which are excluded from one
1 crawler's crawl space by a local domain rule.

1 Here is how the rules work:

- 1 1. If a URL from the crawler's database is excluded by a local prefix rule or
1 address rule, it is assigned return code 760 and it is removed from the index.
1 That is the end of its crawl space handling. The URL will not be crawled again.
- 1 2. After step 1: If a URL from the crawler's database is excluded by a local
1 domain rule, and there is no global crawl space, it is assigned return code 760,
1 and it is removed from the index. That is the end of its crawl space handling.
1 The URL will not be crawled again.
- 1 3. After step 1: If a URL from the crawler's database is excluded by a local
1 domain rule, but explicitly allowed by a rule in the global crawl space, it is
1 assigned return code 761. The crawler will not crawl it again, but it is not
1 removed from the index (it is assumed to be in some other crawler's local crawl
1 space).
- 1 4. After step 1: If a URL from the crawler's database is excluded by a local
1 domain rule, and not explicitly allowed by a rule in the global crawl space, it is
1 assigned return code 760, and removed from the index. That is the end of its
1 crawl space handling.

1 This is an important point: because the global crawl space is consulted only to
1 prevent the deletion of URLs that have already been excluded by the local

1 crawl space, the default result from the global crawl space, if no rule applies to
1 a candidate URL, is to forbid. This is the opposite of the default for the local
1 crawl space.

1 Note that the global.rules file needs to be present in the master_config directory of
1 every crawler that shares a global crawl space, and all copies of the global file, and
1 the individual local crawl.rules files, need to be carefully edited to be mutually
1 consistent.

1 **Configuring multiple Lotus Notes servers**

1 You can add databases from multiple Lotus Notes® servers to the same crawl
1 space without configuring additional crawlers to crawl them. In DB2 II OmniFind
1 Edition Version 8.2, each Notes crawler was limited to crawling a single Lotus
1 Notes server.

1 After you configure a Notes crawler to crawl databases on one Lotus Notes server,
1 take the following actions to add databases from other Lotus Notes servers to the
1 same crawl space:

- 1 On the Crawl page, select the Notes crawler that you want to configure and
1 click the **Crawl space** icon.
- 1 On the Notes crawl space page, click **Add Database**.
- 1 Use the following window to specify information about the Lotus Notes server
1 that you want to crawl: “Specify the Notes server to crawl page”

1 If you need assistance with the subsequent pages displayed by the crawler wizard,
1 click **Help**.

1 **Tip:** To change the password that the crawler uses to access a Lotus Notes server,
1 select the server on the Notes crawl space page, then click **Edit server**. (When
1 the Notes crawler was limited to crawling a single Lotus Notes server, this
1 action was provided at the top of the Notes crawl space page.)

1 **Specify the Notes server to crawl page**

1 Use the Specify the Notes server to crawl page to identify a Lotus Notes server
1 that you want to configure for an existing crawl space. If you are adding a server
1 to the crawl space, specify information that enables the crawler to access databases
1 on that server.

1 For information about setting up your Notes® environment so that it can be
1 crawled by the Notes crawler, see *Administering Enterprise Search*.

1 **Important:** After you click **Next**, the system validates its ability to access the Notes
1 server that you specify. If it is unable to access the server, an error
1 message is displayed. Verify and correct the server name and protocol
1 information and click **Next** again.

1 **Fields and controls**

1 **Existing server**

1 Specifies that you want to add databases from a Lotus Notes server that
1 already exists in the crawl space.

1 If you select this option, select the Lotus Notes server that you want to add
1 databases from.

1 **New server**

1 Specifies that you want to add information about a new Lotus Notes server
1 to the crawl space, and select databases from that server to crawl.

1 If you select this option, specify options that enable the Notes crawler to
1 access databases on the server.

1 **Lotus Notes server name**

1 Type the fully qualified host name of the Domino® server that
1 hosts the Notes databases (.nsf files) that you want this crawler to
1 crawl, such as server1.ibm.com.

1 The default port for Notes servers that use the NRPC protocol is
1 1352. The default port for Notes servers that use the DIOP
1 protocol is 80. If the server does not listen at the default port,
1 specify the port number (for example, server1.ibm.com:1353 or
1 server1.ibm.com:81).

1 **Protocol**

1 Specify information about the communications protocol that the
1 server uses:

1 **Notes Remote Procedure Call (NRPC)**

1 Select this option if the Domino server uses the Notes
1 remote procedure call protocol (NRPC) to exchange
1 information between the server and Web browsers.

1 **Important:** If you previously configured any Notes
1 crawlers for this collection that use the NRPC
1 protocol, stop them before you configure
1 information about this Notes crawler. The
1 enterprise search discovery processes that use
1 NRPC, and the crawler processes that use
1 NRPC, cannot run concurrently. (The discovery
1 processes provide the crawler with information
1 about the sources available for crawling.)

1 If you select the NRPC option, specify the following
1 information:

1 **Lotus Notes user ID file**

1 Type the absolute path for a Notes user ID file that
1 has the authority to access databases on this server.
1 For example, on a UNIX machine you might type
1 /home/username/username.id.

1 Every Notes user has an ID file that uniquely
1 identifies them. Ensure that a copy of the ID file
1 that you specify here exists on the crawler
1 machine. The enterprise search administrator ID
1 must have read access to the Notes user ID file and
1 the directory where the file is stored. (The user ID
1 for the enterprise search administrator is specified
1 during the installation of DB2 II OmniFind
1 Edition.)

1 **Password**

1 Type the password for the specified Notes user ID
1 file.

Domino Internet Inter-ORB Protocol (DIIOP)

Select this option if the Domino server uses the Domino Internet inter-ORB protocol (DIIOP) to exchange information between the server and Web browsers. To use this protocol, you must enable the HTTP and DIIOP tasks on the Domino server.

If you select this option, specify the following information:

Lotus Notes user ID

Type a Notes user ID that has the authority to access databases on this server, such as user1/Hometown/IBM.

Password

Type the Notes Internet password that is specified for this user ID in the Domino Directory. (When DIIOP is configured for a Notes server, one of the steps is to create Internet passwords for users who are authorized to access the server.)

Performance enhancements for the VeniceBridge crawler

Ventica VeniceBridge is now an IBM product called DB2 Information Integrator Content Edition (DB2 II Content Edition). Because the DB2 II OmniFind Edition product interfaces and documentation still refer to VeniceBridge, these release notes also use the term VeniceBridge to describe enhancements to the user interface.

To improve recrawling performance, you can now select a data map when you configure rules for individual VeniceBridge item classes. A data map is a DB2 II Content Edition construct that provides the VeniceBridge crawler information about the item class, such as the creation date, revision date, sort key, and content size. By associating a data map with an item class in the crawler's configuration, you can help ensure that the crawler is able to rapidly recrawl content in that item class.

For Documentum and FileNet Content Services, you can take advantage of this feature without specifying a data map because the VeniceBridge crawler has the default data map.

Important:

To take advantage of these performance enhancements, you must be running DB2 II Content Edition Fix Pack 1.

To associate a data map with an item class in an existing VeniceBridge crawl space:

1. Edit the collection that the VeniceBridge crawler belongs to.
2. Click the Crawl page.
3. Locate the VeniceBridge crawler name and click the **Crawl space** button.
4. On the VeniceBridge crawl space page, locate the item class that you want to configure and click **Edit**.
5. In the **Data map** field on the Configure rules for a VeniceBridge item class page, select the name of the data map that you want to use with this item class.
6. Click **OK**.

1 If you are creating a VeniceBridge crawler, or adding an item class to an existing
1 crawl space, you select the data map name when you specify other options on the
1 Configure rules for a VeniceBridge item class page.

1 Specifying login information for Windows file systems

1 When you configure crawling rules for a Windows subdirectory, you can specify a
1 user ID and password that enables the crawler to access the data in that
1 subdirectory.

1 To configure login credentials for a subdirectory in an existing Windows file
1 system crawl space:

- 1 1. Edit the collection that the Windows file system crawler belongs to.
- 1 2. Click the Crawl page.
- 1 3. Locate the Windows file system crawler name and click the **Crawl space**
1 button.
- 1 4. On the Windows crawl space page, locate the subdirectory that you want to
1 configure and click **Edit**.
- 1 5. On the Configure rules for a Windows subdirectory page, type a valid user ID
1 and password for this subdirectory in the **User ID** and **Password** fields.
- 1 6. Click **OK**.

1 If you are creating a Windows file system crawler or adding a subdirectory to an
1 existing crawl space, you specify the login credentials when you specify other
1 options on the Configure rules for a Windows subdirectory page.

1 External name is now Display name

1 In the enterprise search administration console, the label **External name** was
1 changed to **Display name** in the crawler configuration wizards to better convey
1 the fact that the values specified in this field are for display purposes only.

1 When you assign a display name, you do not change the internal name of the
1 object. For example, you might assign a database field named `lastUpdateDateTime`
1 the more intelligible display name `Last updated`.

1 Having meaningful display names makes it easier for users to recognize the fields
1 that they want to search. It also enables users to search on one field, such as
1 `Author`, and find documents that have different internal field names in their
1 respective data sources (such as `Author`, `Creator`, `Producer`, and so on).

1 Configuring and monitoring the data listener

1 You can use the enterprise search administration console to configure support for
1 your custom data listener applications.

1 In DB2 II OmniFind Edition, Version 8.2, you had to manually edit several files
1 before your applications were able to update enterprise search collections. The
1 following information replaces the instructions in *Configuring support for external*
1 *crawlers* in *Administering Enterprise Search*.

1 A data listener application can add data to a collection, remove URIs from a
1 collection, or instruct a Web crawler to visit or revisit URLs. When the user (client)
1 connects to the data listener component, it needs to pass in a client ID, a password,

1 and the ID of the collection to be updated. The data listener component
1 authenticates the client ID and password and, if valid, authorizes the application to
1 update the specified collection.

1 To configure support for your data listener applications, select the **System** option
1 on the administration console toolbar. Use the following windows to specify
1 information about your data listener applications and to monitor data listener
1 activity:

- 1 • “System view: Data Listener page”
- 1 • “Monitor system view: Data Listener page” on page 17
- 1 • “Data listener details page” on page 17

1 **System view: Data Listener page**

1 Use the Data Listener page of the System view to specify information that enables
1 your custom data listener applications to update collections.

1 A data listener application enables you to crawl data source types that cannot be
1 crawled by the default crawlers provided with DB2 Information Integrator
1 OmniFind Edition. Your application can add data to a collection, remove data from
1 a collection, or instruct a Web crawler to visit or revisit URLs.

1 **Fields and controls**

1 **Number of threads**

1 Type the maximum number of working threads that the data listener
1 component can create for processing requests from data listener
1 applications. The default value is 6.

1 **Data listener port**

1 Type the port number for the data listener component. The default value is
1 6668.

1 **Important:** If you change the port number, you must restart the Data
1 Listener component for the change to become effective.

1 **Add Data Listener Client ID**

1 Adds a blank row to the list of data listener client identifiers so that you
1 can specify information about a new data listener application.

1 **Data Listener client ID**

1 Type the client ID for the data listener application that you want to
1 authorize to access enterprise search collections. Each ID must be unique
1 within the enterprise search system.

1 **Password**

1 Type a password for the data listener application. To access an enterprise
1 search collection, the data listener application must use the credentials (ID
1 and password) that you specify.

1 **Collection name**

1 Select the collection that you want this data listener application to update.
1 Each data listener application can update content in one collection only.



1 **Remove**

1 Removes the selected data listener application from the enterprise search
1 system. This application will not be used to add data to, or remove data
1 from, any enterprise search collections.

Monitor system view: Data Listener page

Use the Data Listener page while you are monitoring the system to view details about the data listener component and data listener application activity.

Fields and controls

Refresh

Refreshes the information on this page. A status message indicates when this information was last refreshed.

Name Shows the component name (**Data Listener**).

Details

Shows detailed information about data listener activity. (Details are not available if the data listener is stopped.)

Status The following icons indicate the current status of the data listener:

Running

Indicates that the data listener component is active. The data listener component is started automatically when the enterprise search system is started. Typically, you do not need to stop it or restart it unless you change the port number that you configured for it.

Stopped

Indicates that the data listener component is stopped.

Data listener details page

Use the Data listener details page to monitor details about current data listener activity, such as information about thread state and data listener application requests.

Fields and controls

Refresh

Refreshes the information on this page. A status message indicates when this information was last refreshed.

Status The following icons indicate the current status of the data listener:

Running

Indicates that the data listener component is active. The data listener component is started automatically when the enterprise search system is started. Typically, you do not need to stop it or restart it.

Stopped

Indicates that the data listener component is stopped.

Number of requests to be processed

Shows the number of data listener application requests that are queued to be processed.

Total number of threads

Shows the total number of threads that the data listener component can create to process requests from data listener applications.

Thread state

Shows the possible thread states:

- Threads initialized
- Threads are reading a client request
- Threads are processing a client request
- Threads are sending a response to the client
- Threads completed sending a response to the client
- Threads are waiting for the next client request
- Threads are in pause state
- Threads are in the process of stopping

Number of threads

Shows how many threads are active for a given thread state (for example, 5 threads are actively fetching data).

Stopping and starting the data listener component

If you change the port number that is configured for the data listener component, you must stop the component and restart it.

Procedure

To stop and restart the data listener component:

1. On the index server, log in as the enterprise search administrator. This user ID was specified during the installation of DB2 II OmniFind Edition.
2. Change to the ES_NODE_ROOT/master_config/datalistener directory.
3. Enter the following commands to stop and restart the data listener:

```
esadmin datalistener stop  
esadmin datalistener start
```

Estimating the number of documents in a collection

When you create a collection, you can provide an estimate for how many documents you expect the collection to hold. The resource manager in DB2 II OmniFind Edition uses this number to estimate the memory and the disk resources that are required for a collection.

The resource manager issues warnings when your estimate of memory and disk resource requirements is more than is currently available in the system. The warnings allow you to prevent future out-of-resource problems.

The Monitor uses the estimated number of documents in conjunction with a threshold percentage that you configure for receiving alerts to determine when to send notifications related to the maximum number of documents being approached by a collection.

To provide an estimate for the potential size of a collection, type a number in the **Estimated number of documents** field when you create the collection. The default value is 1 million documents (1000 000).

After you create a collection, you can change this estimate by editing the collection and specifying a new value on the General page.

1 When the collection grows to the size that you estimate, the system does not stop
1 adding documents to the index. However, if you configure alerts for the collection
1 and enable the option to notify you when the number of documents in the index
1 exceeds a limit, you will receive notifications when the size of the collection
1 reaches the notification percentage that you specify.

1 **Note:** When you configured this alert in version 8.2, you specified a limit for the
1 number of documents in the index by typing the number of documents that
1 the index could contain. With version 8.2.1, the number that you specify in
1 the **Estimated number of documents** field is displayed automatically in the
1 **Limit** field, and you configure the notification threshold by specifying that
1 you want to be notified when the number of documents in the index reaches
1 a percentage of this limit.

1 **Checking system resources is automatic with the collection wizard**

1 If you use the collection wizard, the system automatically estimates whether
1 system resources are adequate for adding a new collection or crawler to your
1 enterprise search system. In DB2 II OmniFind Edition, Version 8.2, you could check
1 system resources only by explicitly clicking the **Check system resources** button.

1 If you check system resources after you create a collection or crawler but before
1 you start crawling documents, you can determine whether the available system
1 resources are adequate for running DB2 II OmniFind Edition at full capacity based
1 on your current configuration settings.

1 For complete information about this feature, see the following technote (reference
1 number 1193761) on the DB2 II OmniFind Edition support Web site: Checking DB2
1 Information Integrator OmniFind Edition system resources.

1 **Monitoring parser activity**

1 While you are monitoring a collection, you can view detailed information about
1 parser activity.

1 In DB2 II OmniFind Edition Version 8.2, you could monitor the parser to start or
1 stop processing, but you could not view detailed status information or obtain
1 information about the number of documents that the parser parsed.

1 To monitor the parser for a collection and view detailed status information:

- 1 1. In the enterprise search administration console, click the **Monitor** icon for the
1 collection that you want to monitor.
- 1 2. Click the Parse page.
- 1 3. If the status of the parser is **Running**, click the **Details** icon.
- 1 4. Use the following window to view details about parser activity:
 - 1 • “Parser details page”

1 **Parser details page**

1 Use the Parser details page to see detailed status information about the documents
1 parsed for a collection. Options enable you to review statistics and administer
1 parser activity.

Fields and controls

Refresh

Refreshes the information on this page. A status message indicates when this information was last refreshed.

Parser status summary

The options in this area show you the status of the parser and provide statistical information about the documents that are parsed but not yet indexed.

Status The first icon on the **Status** line indicates the current status of the crawler:



Running

Indicates that the parser is running.

Depending on the actual state of the parser, you might see a message that provides more information about current parsing activity. For example, the parser might be actively parsing documents, idle (sleeping until there are more documents available to parse, waiting to be restarted because of an error condition, or waiting for the parser service to restart), or paused (waiting, for example, for an index reorganization to be completed).



Stopped

Indicates that the parser is stopped.

Depending on the current status of the parser, and your administrative role, the following icons enable you to administer parser activity:



Start

Starts the parser.

After the parser is started, it typically parses crawled documents until there are no more documents to be parsed. The parser then enters a sleeping mode, and will not run until it next checks for the availability of newly crawled documents to parse.



Stop

Stops the parser.

You might need to stop and restart the parser, for example, if you change the categorization type or add category rules.

Number of documents in the data store that need to be indexed

Shows the number of documents that the parser parsed and wrote to the data store. These documents have not yet been added to the enterprise search index.

The following statistics are available for each crawler in the collection:

Snapshot statistics

The options in this area show you the current status of parser activity and provide statistical information about the documents collected by individual

1 crawlers. To achieve the best search results, do not build the index until the
1 parser has parsed all of the documents crawled by the various crawlers in
1 the collection.

1 **Snapshot start**

1 Shows the date and time that the statistics in the current view of
1 the parser's status were collected.

1 **Parse rate**

1 Shows how many pages the parser is parsing per second.

1 The following statistics are available for each crawler in the collection:

1 **Crawler name**

1 Shows the name of each crawler in the collection.

1 **Crawler type**

1 For each crawler, shows the type of data that the crawler crawls.

1 **Documents parsed**

1 For each crawler, shows how many documents are available for
1 parsing and how many have been parsed so far.

1 **Status** For each crawler, shows the status of document parsing activity:



1 **Running**

1 Indicates that the parser is actively parsing the documents
1 crawled by this crawler.



1 **Stopped**

1 Indicates that the parser is not parsing any documents for
1 this crawler (perhaps because no documents are available
1 for parsing).



1 **Paused**

1 Indicates that the parsing of documents crawled by this
1 crawler is paused.



1 **Errors**

1 Indicates that errors occurred while the documents crawled
1 by this crawler were being parsed. To review the errors,
1 open the Log page while you are monitoring the collection
1 and view the log files for the parser.



1 **Complete**

1 Indicates that the parser finished parsing the documents
1 crawled by this crawler.

1 Mapping HTML metadata to search fields

1 The parser can map HTML metadata elements to field names. When you configure
1 the parser for a collection, you can specify which HTML metadata elements are to
1 be indexed as search fields. Users can then search specific parts of HTML
1 documents by specifying these field names in queries.

1 When you map HTML metadata to fields, you specify whether you want users to
1 be able to search the field, search the field with a parametric query, or view the
1 field in the search results.

- To map HTML metadata elements to search fields, take the following actions:
1. In the administration console, click the **Edit** icon for the collection that you want to change.
 2. Select the Parse page.
 3. Click **Map HTML metadata to fields**.
 4. Use the following window to specify information about the metadata fields that you want to search:
 - “HTML metadata field mappings page”

HTML metadata field mappings page

Use the HTML metadata field mappings page to map HTML metadata elements to field names. By specifying the mapped field names in queries, users can search specific parts of HTML documents and obtain more precise search results.

When you create an HTML field mapping, the change becomes effective the next time that you stop and restart the parser. The new mapping has no effect on data that is already parsed and indexed.

Fields and controls

Add Field

Adds a blank line in which you can map another HTML metadata element to a search field name.

For each HTML metadata element that you want to map to a search field name, provide the following information.

Field name

Type a name that you want to associate with the HTML metadata element. Users can specify this field name when they query HTML documents in this collection. For example, to map an HTML metadata tag named publisher, you might type publisher as the search field name.

This value must be a single word that contains only alphanumeric characters.

HTML metadata field

Type the name of the HTML metadata element that maps to the field name that you specified in the **Field name** field. To use the same example, type publisher in this field because you are mapping the HTML metadata element publisher to a search field named publisher.

This value must be a single word that contains only alphanumeric characters.

Search by field name

Select this check box if you want users to be able to search HTML documents by specifying the field name that you specified in the **Field name** field. If this check box is clear, users cannot specify the field name when they search HTML documents in this collection.

Parametric search

Select this check box if you want users to be able to formulate parametric queries when they search this field. Parametric queries enable users to evaluate numeric and date values. For example, a user might want to determine whether the value of this field is less than or greater than another value.

1 If you select this check box, ensure that the data type of the HTML
1 metadata element is DECIMAL, DOUBLE, INTEGER, SHORT, TIME, or
1 TIMESTAMP.

1 **Show in search results**

1 Select this check box if you want to show this field in the search results. If
1 this check box is clear, this field will not be included in the search results.



1 **Delete**

1 Deletes the selected HTML metadata field mapping.

1 **Removing new line characters from white space in Japanese and 1 Chinese documents**

1 In languages where white space is not used to delimit word boundaries, such as
1 Japanese and Chinese, you can configure the parser to remove certain white space
1 characters that cause line breaks.

1 To enable this feature for a collection:

- 1 1. Log in as the enterprise search administrator (this user ID and password were
1 established during the installation of DB2 II OmniFind Edition).
- 1 2. Use a text editor to edit the file
1 `collection_ID.parserdriver/collection.properties`, where `collection_ID` is
1 the ID that was assigned to the collection by the system when you created it.
- 1 3. Change the value of the `removeCjNewlineChars` property from `false` to `true`.

1 If this feature is enabled, and a document in the specified collection is in Japanese
1 or Chinese, then the parser will remove any sequence of new line or tab characters
1 that separate two letter characters from the non-ASCII character range. The
1 following characters are removed : Tab (0x09), LF or line feed (0x0A), and CR or
1 carriage return 0x0D.

1 **Configuring document types for a collection parser and Stellent 1 session**

1 In DB2 II OmniFind Edition Version 8.2, file types and content types are
1 pre-associated with a particular collection parser and Stellent document filtering
1 session. You can now create configuration files to specify how particular types of
1 documents are to be parsed.

1 In version 8.2, certain types of documents are not parsed (such as PostScript
1 documents), and some are hard-coded to be sent to the Stellent session for
1 processing (Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Lotus
1 Freelance, Lotus 123, PDF, RT, and Ichitaro). All other document formats are
1 processed by the built-in HTML or XML parsers. Because metadata can be
1 misleading, plain text and HTML documents might be sent to the Stellent session
1 in error, and then sent back to one of the built-in parsers (a situation that can
1 impact performance).

1 To ensure that documents in your crawl space are accurately and efficiently parsed,
1 you can create configuration files to specify which types of documents are to be
1 parsed by the collection parser and which are to be parsed by the Stellent session.

1 Associating document types with the collection parser and Stellent session involves
1 the following tasks:

- 1 1. Configure document types for the collection parser. This step involves creating
1 a configuration file that maps document types to the parser that is used by a
1 collection. You can create one of these configuration files per collection.
- 1 2. Configure document types for the Stellent session. This step involves creating a
1 configuration file that maps document types to the Stellent document filters
1 that are used by a collection. You can create one of these configuration files per
1 collection.
- 1 3. For the changes to become effective, use the enterprise search administration
1 console to monitor the collection for which you configured document types,
1 then stop and restart the parser.

1 **Configure document types for the collection parser**

1 To associate particular types of documents with the collection parser, you create a
1 configuration file. There is no support for this feature in the enterprise search
1 administration console.

1 The configuration file specifies:

- 1 • Which documents you want to send to the Stellent session, depending on the file
1 extension or the content type.
- 1 • How to parse documents whose type is unknown because of incomplete
1 metadata.

1 If the configuration file does not exist, the collection parser uses the default
1 configuration settings, which correspond exactly to the behavior of DB2 II
1 OmniFind Edition, Version 8.2. For a description of these default settings, see
1 “Default parser service rules” on page 25.

1 To associate document types with the collection parser:

- 1 1. On the index server, log in as the enterprise search administrator. This user ID
1 was specified during the installation of DB2 II OmniFind Edition.
- 1 2. Create the configuration file as follows, where *collection_ID* identifies the
1 collection that you want to configure:

1 ES_NODE_ROOT/master_config/*collection_ID*.parserdriver/parserTypes.cfg

1 This file lists the file extensions and content types to send to the Stellent
1 session. The format of the file is a sequence of lines, where each line is one of
1 the following rules:

1 **EXTENSION** *extension parser*

1 All documents whose URL ends on the specified extension will be
1 processed by the specified parser. Do not include the period in the
1 extension; comparison is case-insensitive.

1 **CONTENTTYPE** *type/subtype parser*

1 All documents whose content type matches the specified type/subtype
1 will be processed by the specified parser. Given the content type t/s of
1 a document, a match occurs if t equals type, and either s equals
1 subtype or the subtype is a wildcard (the asterisk, *).

1 **UNKNOWN** *parser*

1 All documents whose extension and content type are not known (that
1 is, not made available by the crawler), will be processed by the
1 specified parser.

1 **DEFAULT** *parser*

1 All documents that are not covered by any of the other rules will be
1 processed by the specified parser.

1 In all cases, *parser* must be `html`, `xml`, `stellent`, or `none`, where `none` means that
1 the document is not to be parsed.

1 **Rule priority:** If more than one rule matches a document, then the more
1 specific rule prevails, disregarding the order in which the rules appear. The
1 specificity of a rule decreases in order of appearance in the list above.

- 1 • An EXTENSION rule is more specific than a CONTENTTYPE rule.
- 1 • A CONTENTTYPE rule that includes a subtype is more specific than one
1 with a wildcard. For example, a rule for content type
1 `application/postscript` has priority over a rule for `application/*`.
- 1 • There should not be two rules for the same extension or content type. In that
1 case, it is up to the implementation which of the rules is given priority.

1 Example

1 In this example, the built-in HTML parser processes all documents with the
1 extension `txt`, `htm` or `html`, with a content type that begins with `text/`, or with an
1 unknown extension and content type. The built-in XML parser processes all
1 documents with extension `xml` or with content type `text/xml`. All other documents,
1 including those with a content type that starts with `application/`, are sent to the
1 Stellent session.

```
1 EXTENSION doc stellent  
1 EXTENSION txt html  
1 EXTENSION htm html  
1 EXTENSION html html  
1 EXTENSION xml xml  
1 EXTENSION ps none  
1 CONTENTTYPE text/xml xml  
1 CONTENTTYPE text/* html  
1 CONTENTTYPE application/* stellent  
1 UNKNOWN html  
1 DEFAULT stellent
```

1 Default parser service rules

1 If you do not create a configuration file to map file types and content types to the
1 parser for a collection, default rules are used to parse documents.

1 The default rules used by the collection parser are as follows:

```
1 EXTENSION pdf stellent  
1 EXTENSION ppt stellent  
1 EXTENSION prz stellent  
1 EXTENSION lwp stellent  
1 EXTENSION doc stellent  
1 EXTENSION rtf stellent  
1 EXTENSION xls stellent  
1 EXTENSION 123 stellent  
1 EXTENSION vsd stellent  
1 EXTENSION vdx stellent  
1 EXTENSION jxw stellent  
1 EXTENSION jsw stellent  
1 EXTENSION jtw stellent  
1 EXTENSION jaw stellent  
1 EXTENSION juw stellent  
1 EXTENSION jbw stellent  
1 EXTENSION jvw stellent  
1 EXTENSION jfw stellent  
1 EXTENSION jtt stellent  
1 EXTENSION jtd stellent  
1 EXTENSION jttc stellent  
1 EXTENSION jtdc stellent  
1 EXTENSION jtdx stellent
```

```

1      EXTENSION ps none
1      EXTENSION xml xml
1      EXTENSION txt text
1      EXTENSION htm html
1      EXTENSION htm1 htm1
1      EXTENSION shtml html
1      EXTENSION xhtml html
1      EXTENSION asp htm1
1
1      CONTENTTYPE application/postscript none
1      CONTENTTYPE application/* stellent
1      CONTENTTYPE text/rtf stellent
1      CONTENTTYPE text/richtext stellent
1      CONTENTTYPE text/xml xml
1      CONTENTTYPE text/html html
1      CONTENTTYPE text/plain text
1
1      UNKNOWN html
1      DEFAULT html

```

1 **Configure document types for the Stellent session**

1 To specify which types of documents are to be parsed by Stellent document filters,
1 you create a configuration file. There is no support for this feature in the enterprise
1 search administration console.

1 The configuration file specifies:

- 1 • Which file types are to be parsed by the Stellent session. A file type corresponds
1 to one of the file types recognized by the Stellent library.
- 1 • Which file types are to be sent back to the collection parser for processing with
1 one of the built-in parsers. (This action is needed because the collection parser
1 might send a document to the Stellent session in error, due to misleading
1 metadata.)
- 1 • Which file types are to be rejected because they are not supported by DB2 II
1 OmniFind Edition.

1 If the configuration file was specified but does not exist, the parser will fail to
1 start. If no configuration file was specified for the `OutsideInSupportedTypes`
1 property in the `stellent.properties` file, then the default configuration settings will
1 be used. For a description of these default settings, which correspond to the
1 behavior of DB2 II OmniFind Edition Version 8.2, see “Default parsing rules for
1 Stellent sessions” on page 28.

1 To configure document types for the Stellent session:

- 1 1. On the index server, log in as the enterprise search administrator. This user ID
1 was specified during the installation of DB2 II OmniFind Edition.
- 1 2. Edit the
1 `ES_NODE_ROOT/master_config/collection_ID.stellent/stellent.properties`
1 file, where `collection_ID` identifies the collection that you want to configure.
- 1 3. For the `OutsideInSupportedTypes` property, specify the absolute path of the
1 configuration file that you are creating.

1 For example, you might create the following configuration file for a single
1 collection, and store it with other collection-specific files:

```
1 ES_NODE_ROOT/master_config/collection_ID.stellent/stellenttypes.cfg
```

To use the same settings for all collections, you might create the following configuration file and store it with other system-level files. (If you use this approach, be sure to specify this path in the `stellent.properties` file for each collection, as specified in step 2 on page 26.)

```
ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```

4. Create the configuration file. This file lists document types and how they are to be handled. The format of the file is a sequence of lines, where each line is one of the following rules:

```
accept DEFAULT
accept ALL doctype
accept type doctype
native DEFAULT
native type doctype
reject type
```

Where:

doctype

Is the value to be used for the doctype query token. The doctype token is searchable by specifying, for instance, `doctype:pdf` in a query.

type Is one of the FI_ values in the Stellent library, and *doctype* is the value to be used for the doctype token if a rule is applied.

DEFAULT

Means that the list of accepted or native types, depending on the type of the rule, includes the default list. This option enables you to extend the default configuration instead of replacing it.

All Means that all types that are not explicitly listed are accepted with the specified doctype token.

Rule processing:

- If there is an accept rule for *type*, including the default list if accept DEFAULT was specified, it is accepted.
- Else, if there is a reject rule for *type*, it is not accepted.
- Else, if accept ALL was specified, it is accepted.
- Otherwise, it is not accepted.

If the document type is accepted, the *doctype* value that was specified in the rule that was applied is used. This value is sent back to the collection parser along with the parsed content. If the document type is not accepted, the following behavior occurs:

- If there is a native rule for *type*, including the default list if native DEFAULT was specified, the document is sent back to the built-in parser, indicating the value for the *doctype* token that is specified by this rule. The value of *doctype* must be either `txt`, `htm` or `xml`, indicating plain text, HTML or XML, respectively.
- Otherwise, the document is rejected and will not be parsed.

Example

Given the following configuration file, the Stellent session accepts, in addition to the default list, the Microsoft Visio format.

```
accept DEFAULT
accept FI_VISI03 visio
accept FI_VISI04 visio
accept FI_VISI05 visio
accept FI_VISI06 visio
```

1 Given the following configuration file, Postscript documents will be supported and
1 searchable with a document type of ps; the X pixmap format (XPM) will be sent
1 back to the built-in text parser, the PNG image format will be rejected; and all
1 other file types will be accepted and made searchable with a document type of
1 other.

```
1 accept DEFAULT
1 accept FI_POSTSCRIPT ps
1 native FI_XPIXMAP txt
1 accept ALL other
1 reject FI_PNG
```

1 **Default parsing rules for Stellent sessions**

1 If you do not create a configuration file to map file types to Stellent session
1 document filters, default rules are used to parse documents.

1 The default rules used by the Stellent session are as follows:

```
1 ACCEPT FI_WORD4 doc
1 ACCEPT FI_WORD5 doc
1 ACCEPT FI_RTF rtf
1 ACCEPT FI_WINWORD1 doc
1 ACCEPT FI_WINWORD1COMPLEX doc
1 ACCEPT FI_WINWORD2 doc
1 ACCEPT FI_WORD6 doc
1 ACCEPT FI_WINWORD6 doc
1 ACCEPT FI_ICHITAR03 jxw
1 ACCEPT FI_ICHITAR04 jsw
1 ACCEPT FI_WINWORD1J doc
1 ACCEPT FI_WINWORD5J doc
1 ACCEPT FI_RTFJ rtf
1 ACCEPT FI_WINWORD7 doc
1 ACCEPT FI_WORDPRO lwp
1 ACCEPT FI_WINWORD97 doc
1 ACCEPT FI_ICHITAR08 jtd
1 ACCEPT FI_WORDPRO97 lwp
1 ACCEPT FI_WINWORD2000 doc
1 ACCEPT FI_WINWORD2002 doc
1 ACCEPT FI_WINWORD2003 doc
1 ACCEPT FI_123R1 123
1 ACCEPT FI_123R2 123
1 ACCEPT FI_123R3 123
1 ACCEPT FI_EXCEL xls
1 ACCEPT FI_EXCEL3 xls
1 ACCEPT FI_EXCEL4 xls
1 ACCEPT FI_123R4 123
1 ACCEPT FI_EXCEL5 xls
1 ACCEPT FI_123R6 123
1 ACCEPT FI_EXCEL97 xls
1 ACCEPT FI_123R9 123
1 ACCEPT FI_EXCEL2000 xls
1 ACCEPT FI_EXCEL2002 xls
1 ACCEPT FI_EXCEL2003 xls
1 ACCEPT FI_FREELANCE prz
1 ACCEPT FI_POWERPOINT4 ppt
1 ACCEPT FI_POWERPOINT3 ppt
1 ACCEPT FI_POWERPOINT7 ppt
1 ACCEPT FI_FREELANCE3 prz
1 ACCEPT FI_POWERPOINTMAC3 ppt
1 ACCEPT FI_POWERPOINTMAC4 ppt
1 ACCEPT FI_PDF pdf
1 ACCEPT FI_EXTPOWERPOINT4 ppt
1 ACCEPT FI_EXTPOWERPOINTMAC4 ppt
1 ACCEPT FI_POWERPOINTMACB3 ppt
1 ACCEPT FI_POWERPOINTMACB4 ppt
1 ACCEPT FI_POWERPOINT97 ppt
```

1	ACCEPT FI_PDFMACBIN	pdf
1	ACCEPT FI_POWERPOINT9597	ppt
1	ACCEPT FI_POWERPOINT2000	ppt
1	ACCEPT FI_POWERPOINT2	ppt
1		
1	NATIVE FI_HTML	htm
1	NATIVE FI_HTML_LATIN2	htm
1	NATIVE FI_HTML_JAPANESESJIS	htm
1	NATIVE FI_HTML_JAPANESEEUC	htm
1	NATIVE FI_HTML_CHINESEBIG5	htm
1	NATIVE FI_HTML_CHINESEEUC	htm
1	NATIVE FI_HTML_CHINESEGB	htm
1	NATIVE FI_HTML_KOREANHANGUL	htm
1	NATIVE FI_HTML_CYRILLIC1251	htm
1	NATIVE FI_HTML_CYRILLICKO18	htm
1	NATIVE FI_CYRILLIC1251	txt
1	NATIVE FI_CYRILLICKO18	txt
1	NATIVE FI_W2KHTML	htm
1	NATIVE FI_XL2KHTML	htm
1	NATIVE FI_PP2KHTML	htm
1	NATIVE FI_XML	xml
1	NATIVE FI_WML	xml
1	NATIVE FI_HTML_JAPANESEJIS	htm
1	NATIVE FI_WML_CHINESEBIG5	xml
1	NATIVE FI_WML_CHINESEEUC	xml
1	NATIVE FI_WML_CHINESEGB	xml
1	NATIVE FI_WML_CYRILLIC1251	xml
1	NATIVE FI_WML_CYRILLICKO18	xml
1	NATIVE FI_WML_JAPANESEJIS	xml
1	NATIVE FI_WML_JAPANESESJIS	xml
1	NATIVE FI_WML_JAPANESEEUC	xml
1	NATIVE FI_WML_KOREANHANGUL	xml
1	NATIVE FI_WML_LATIN2	xml
1	NATIVE FI_HTMLUNICODE	htm
1	NATIVE FI_XML_DOCTYPE_HTML	htm
1	NATIVE FI_XHTML	htm
1	NATIVE FI_ASCII	txt
1	NATIVE FI_ANSI	txt
1	NATIVE FI_UNICODE	txt
1	NATIVE FI_ASCII8	txt
1	NATIVE FI_ANSI8	txt
1	NATIVE FI_MAC	txt
1	NATIVE FI_MAC8	txt
1	NATIVE FI_SHIFTJIS	txt
1	NATIVE FI_CHINESEGB	txt
1	NATIVE FI_HANGEUL	txt
1	NATIVE FI_CHINESEBIG5	txt
1	NATIVE FI_LATIN2	txt
1	NATIVE FI_JAPANESE_EUC	txt
1	NATIVE FI_HEBREW_OLDCODE	txt
1	NATIVE FI_HEBREW_PC8	txt
1	NATIVE FI_HEBREW_E0	txt
1	NATIVE FI_HEBREW_WINDOWS	txt
1	NATIVE FI_ARABIC_710	txt
1	NATIVE FI_ARABIC_720	txt
1	NATIVE FI_ARABIC_WINDOWS	txt
1	NATIVE FI_7BITTEXT	txt
1	NATIVE FI_JAPANESE_JIS	txt
1	NATIVE FI_CENTRALEU_1250	txt
1	NATIVE FI_UTF8	txt
1	NATIVE FI_EBCDIC_37	txt
1	NATIVE FI_EBCDIC_273	txt
1	NATIVE FI_EBCDIC_277	txt
1	NATIVE FI_EBCDIC_278	txt
1	NATIVE FI_EBCDIC_280	txt
1	NATIVE FI_EBCDIC_284	txt
1	NATIVE FI_EBCDIC_285	txt

```
1 NATIVE FI_EBCDIC_297 txt
1 NATIVE FI_EBCDIC_500 txt
1 NATIVE FI_EBCDIC_870 txt
1 NATIVE FI_EBCDIC_871 txt
1 NATIVE FI_EBCDIC_1026 txt
```

1 Enhancements to query processing

1 In a multiple server configuration, failover protection is available at the collection
1 level, not just at the server level. If a collection on one search server becomes
1 unavailable for any reason, then the queries for that collection are routed
1 automatically to the other search server.

1 Collapsing results from the same Web site

1 When you edit the SiteDefs.txt file to specify options for collapsing search results
1 from the same Web site, you can specify a string as the site identifier along with
1 the URL prefix.

1 In DB2 II OmniFind Edition, Version 8.2, the SiteDefs.txt file contained a list of
1 URL prefixes only. Now, you can specify a string as an identifier for each site. As
1 shown in the following example, you type the URL prefixes at the start of the line
1 and optionally type a string identifier to the right:

```
1 http://mycompany.server1.com/hr/ hr
1 http://mycompany.server2.com/hr/ hr
1 http://mycompany.server3.com/hr/ hr
1 http://mycompany.server1.com/finance/ finance
1 http://mycompany.server1.com/news news
1 http://mycompany.server1.com/USA
```

1 Search applications can use either the URL prefix or the identifier to group and
1 collapse documents in the search results. For example, the hr identifier enables
1 search results from three servers that host human resource documents to be
1 collapsed together in the search results.

1 The enterprise search query syntax was extended to support this feature. For
1 example, to retrieve documents that belong to the same group, use the following
1 syntax, where *group* is either the group identifier or the fully qualified URL for a
1 document that belongs to the group:

```
1 samegroupas:group
```

1 For example:

```
1 samegroupas:hr
1 samegroupas:http://mycompany.server3.com/hr/
```

1 For complete information about updating the SiteDefs.txt file, see the topic
1 *Collapsing results from the same Web site* in *Administering Enterprise Search*.

1 Extended highlighting in search results

1 A property in the Query interface enables query terms to be highlighted in several
1 areas of the search result details. In DB2 II OmniFind Edition, Version 8.2,
1 highlighting was limited to the summary field.

1 When you create a search application, you can manipulate query processing by
1 using the getProperty and setProperty methods of the Query interface. A new

property that you can use with these methods, `HighlightingMode`, enables you to control how query terms are highlighted in the search results. This property has the following values:

DefaultHighlighting

Highlights query terms in the summary field only. If your search application omits the `HighlightingMode` property, this is the default action.

ExtendedHighlighting

Extends the highlighting of query terms to other areas of the search results, such as the title, URL, and other fields.

You can retrieve the highlighted URL field from the `Result` object's properties by using the `getProperty` or `getProperties` methods. The property name is `HighlightedDocumentID`, and its value is the highlighted URL. The `Result` method `getDocumentID` will continue to return the non-highlighted URL.

Timeout support in search applications

A property in the Remote application programming interface (API) enables you to specify a timeout value for remote requests.

Search applications can pass in the timeout property with the `Properties` object on the following methods:

```
SearchFactory.getSearchService(Properties)  
BrowseFactory.getBrowseService(Properties)
```

For example:

```
// create a new Properties object.  
Properties config = new Properties();  
config.setProperty("hostname", hostname);  
config.setProperty("port", portNumber);  
config.setProperty("locale", "en_US");  
config.setProperty("timeout", "60");  
SearchService searchService = factory.getSearchService(config);
```

The timeout value, which represents the number of seconds to wait before the remote request times out, must be an integer (such as 60, not 60.5 or sixty). If you do not specify a timeout value, the Remote API uses the default value, 30 seconds.

Known limitations, problems, and workarounds

This information describes known limitations, problems, and workarounds for DB2 Information Integrator OmniFind Edition, Version 8.2.

Any limitations and restrictions might or might not apply to other releases of the product.

Unpacking the DB2 UDB for AIX tar file

For DB2 Universal Database Enterprise Server Edition, Version 8.2 for AIX, you must untar the CD image before you can install DB2 UDB.

Unpack the CD before you start the enterprise search installation.

For more information about unpacking the CD, go to the DB2 Information Center, Version 8.2: <http://publib.boulder.ibm.com/infocenter/db2help/index.jsp>. In the DB2 Information Center, click **Installing** → **Database Systems** → **DB2 Universal Database for Linux, UNIX, and Windows** → **DB2 servers** → **DB2 UDB Enterprise Server Edition (non-partitioned)** → **AIX** → **Starting the DB2 setup wizard**.

This file must be untarred before the enterprise search installation program can install DB2 UDB.

To untar the CD image:

1. Copy product.tar.Z, where product represents the product you are licensed to install, to a temporary file system.
2. Enter the following command to start the DB2 Setup wizard:

```
zcat product.tar.Z | tar -xf - ;./product/db2setup
```

For example, if the product name for DB2 UDB Enterprise Server Edition is ese, then enter the following command:

```
zcat ese.tar.Z | tar -xf - ;./ese/db2setup
```

During the enterprise search installation, provide the directory in the untarred file path that contains the file db2setup. For example, the directory for the untarred file might be .\ese.sbcsaix1.

Shared memory segments in AIX

If you see message SQL1224N on AIX, you might have exceeded the number of shared memory segments that are allowed by AIX.

To resolve this problem, see the topic SQL1224N Shared Memory Segments on the IBM Support Web site.

Restart the system after you uninstall on Windows

If you uninstall DB2 II OmniFind Edition on a Windows computer, you must restart the system to complete the removal of all program files.

You must also manually delete the enterprise search installation directories. If you do not restart the system or delete the program files and then attempt to install enterprise search again, problems might occur.

Orphan URLFetcher process for Linux and AIX

Problem: The Web crawler uses two processes. When one process exits, it signals a second process to exit. One process might sometimes exit unexpectedly and leave the other process running indefinitely.

You can determine that this happened if you see an orphan process like the following process from the output of the UNIX ps command:

```
F S UID          PID  PPID  C  PRI  NI ADDR      SZ  WCHAN  STIME TTY          TIME CMD
0 S user1      26222    1  0  75   0   - 152641 schedu 10:34 pts/8      00:00:00
/opt/IBMJava2-141/jre/bin/java -classpath ../../URLFetcher.jar:...
```

Notice that:

- The PPID (parent process ID) of 1 indicates that the process was orphaned.
- The presence of URLFetcher.jar in the classpath indicates that this is a Web crawler process.

Solution

Find the process ID (PID) of the orphaned process. In the previous example, it is 26222 and kill it. On Linux and AIX, use the kill command with no argument. You might also see “child” entries in the process table. The child entries represent threads that are created by the orphan process. These child entries should be deleted when you kill their ancestor process.

When you see an orphan URLFetcher process, you should kill the process because if it is in the process table, the operating system reserves resources for it. You should regularly check whether these orphan processes exist, especially if you suspect that the Web crawler has terminated abnormally. If the processes exist, kill them.

Note: In DB2 II OmniFind Edition Version 8.2.1, orphaned processes rarely occur because the URL fetcher now self-terminates if it notices that the C++ component is gone. If the problem does occur, take care to not accidentally kill another user’s httpauth.ini editing session (the HTTPAuthEdit utility is a JVM that runs from the URLFetcher JAR file, and might look similar in the process table).

Index location on the server

When you create a collection, if you want to specify an index location other than the default location, the disk space usage for this nondefault index location will not be monitored if you specify an index location that is not on the server data root directory.

For a multiple server installation, when you create a collection, if you want to specify an index location other than the default location, the index location must first exist on all four servers.

Heap size limits

The document parser Java™ session (also known as parserservice) has a 400 MB heap size limit.

The heap size value is stored as follows:

```
ES_NODE_ROOT/config/services.ini:session7.max_heap=400
```

If OutOfMemory messages are displayed in the system error log for the parserservice session, stop the parserservice session, increase the max_heap value in the services.ini file for the parserservice, and restart the parserservice session.

DB2 UDB tables with large objects

DB2 Universal Database tables with LOB data that is 128 MB or more cannot be crawled. Documents that are larger than 8 MB cannot be indexed.

Index status does not display

If you re-install a new version of DB2 II OmniFind Edition without removing the previous installation, the index monitoring status might not display correctly.

When you start the index build, the status of the build does not display correctly on the Index page on the administration console. However, the index is building, and when it is finished building, the Index page will show the correct number of documents that were indexed and the correct times of the index build.

It is recommended that you remove the previous version of DB2 II OmniFind Edition before you install a new version.

User ID permissions for Linux and AIX

You might need to add write permissions for a new enterprise search user ID.

If you re-install DB2 II OmniFind Edition with a different Linux or AIX user from the enterprise search administrator, then when you add a DB2 crawler and specify that you want to crawl remote, uncataloged databases (those that use the DB2 Universal Database Type 4 Java database connectivity (JDBC) driver), an error might occur. In that case, you must add write permissions for the new enterprise search administrator user to the `/home/db2inst1/sqllib/tmp/0.spf` file.

1 Enabling direct document access for FileNet Content Services

1 When you use the VeniceBridge crawler to crawl FileNet Content Services
1 repositories that are managed by the DB2 Information Integrator Content Edition
1 (formerly VeniceBridge) server, the target FileNet Content Services library and
1 FileNet Web Services work with the FileNet Content Services connector and must
1 be configured to use direct document access. See the FileNet Web Services
1 *Administrator's Guide* and FileNet Content Services administration documentation
1 to set up direct document access support.

1 Limitations for crawling nickname tables

1 There are limitations to crawling IBM DB2 Information Integrator nickname tables
1 for IBM DB2 Universal Database for z/OS® and Microsoft SQL Server.

1 If the target table is a DB2 Information Integrator nickname table for DB2
1 Universal Database for z/OS or Microsoft SQL Server and if document content
1 field is specified, unique identifiers cannot include any columns with the following
1 data types:

- 1 • LONG VARCHAR
- 1 • LONG VARCHAR FOR BIT DATA
- 1 • LONG VARGRAPHIC

1 If the target table is a DB2 Information Integrator nickname table for Microsoft
1 SQL Server, and if the nickname table is created on a database with UTF-8 codeset,
1 document content from columns with the BLOB data type might be corrupted.

1 If the target table is a DB2 Information Integrator nickname table for DB2
1 Universal Database for z/OS or Microsoft SQL Server and if the document content
1 field is specified, follow these steps to improve the performance of the DB2
1 crawler:

- 1 1. Create a server with COLLATING_SEQUENCE option set to Y.
- 1 2. Create the nickname under this server.

1 If the nickname table was already defined before you set the option to Y, drop the
1 nickname and create it again.

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 all 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 give 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 (DBCS) information, contact the IBM Intellectual Property Department in your country/region 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/region 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 that 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 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. 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 IBM's application programming interfaces.

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

Outside In (®) Viewer Technology, © 1992-2004 Stellent, Chicago, IL., Inc. All Rights Reserved.

Trademarks

This topic lists IBM trademarks and certain non-IBM trademarks.

The following terms are trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM
AIX
AIX 5L
DB2
DB2 Universal Database
Domino
Informix
Lotus
Lotus Notes
Notes
OmniFind
WebSphere
xSeries
z/OS

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

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. 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.

Intel, Intel Inside (logos), MMX and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

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

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

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

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

To locate an IBM office in your country or region, see the IBM Directory of Worldwide Contacts on the Web at www.ibm.com/planetwide.

Obtaining product information

Information about DB2 Information Integrator is available by telephone or on the Web.

Information about DB2 Information Integrator is available by telephone or on the Web. The phone numbers provided here are valid in the United States.

1. To order products or to obtain general information: 1-800-IBM-CALL (1-800-426-2255)
2. To order publications: 1-800-879-2755
3. Visit the Web at www.ibm.com/software/data/integration/db2ii/support.html.

This site contains the latest information about:

- The technical library
- Ordering books
- Client downloads
- Newsgroups
- Fix packs
- News
- Links to Web resources

Providing comments on the documentation

Please send any comments that you have about this book or other DB2 Information Integrator documentation.

Your feedback helps IBM to provide quality information. Please send any comments that you have about this book or other DB2 Information Integrator documentation. You can use any of the following methods to provide comments:

1. Send your comments using the online readers' comment form at www.ibm.com/software/data/rcf.
2. Send your comments by e-mail to comments@us.ibm.com. Include the name of the product, the version number of the product, and the name and part number of the book (if applicable). If you are commenting on specific text, please include the location of the text (for example, a title, a table number, or a page number).



Printed in USA



Java[™]
COMPATIBLE