# Google Search Appliance Connectors Deploying the Connector for File Systems

Google Search Appliance Connector for File Systems software version 4.0.4 Google Search Appliance software version 7.2

January 2015



# Table of Contents

Table of Contents
About this guide
Overview of the GSA Connector for File Systems
<u>Continuous automatic updates</u>
DFS access control
Supported operating systems for the connector
Supported file system protocols
Known limitations
File System limitation
Distributed File System limitation
Before you deploy the Connector for File Systems
Windows account permissions needed by the connector
Deploy the Connector for File Systems
Step 1 Specify the IP address of the host computer
Step 2 Install the Connector for File Systems
Step 3 Configure adaptor-config.properties variables
Step 4 Run the Connector for File Systems
Advanced Topics
Not changing 'last access' of the documents on the share
Skipping File Share Access Control
Uninstall the Google Search Appliance Connector for File Systems
Troubleshoot the Connector for File Systems

# About this guide

This guide is intended for anyone who needs to deploy the Google Search Appliance Connector 4.0.4 for File Systems. The guide assumes that you are familiar with Windows operating systems, file systems, and configuring the Google Search Appliance by using the Admin Console.

See the <u>Google Search Appliance Connectors Administration Guide 4.0.4</u> for general information about the connectors, including:

- What's new in Connectors 4.0?
- General information about the connectors, including the configuration properties file, supported ACL features, and other topics
- Connector security
- Connector logs
- Connector Dashboard
- Connector troubleshooting

For information about using the Admin Console, see the <u>Google Search Appliance Help</u> <u>Center</u>.

For information about previous versions of connectors, see the <u>Connector documentation</u> <u>page</u> in the <u>Google Search Appliance Help Center</u>.

### Overview of the GSA Connector for File Systems

The Connector for File Systems enables the Google Search Appliance to crawl and index content from Windows shares. A single connector instance can support a single Windows share. The share can be a UNC path or a mapped drive. DFS links are fully supported by the connector.

The Connector For File Systems submits URLs identifying files in the file system repository to the GSA. These URLs point back to the connector, which services HTTP GET requests from the GSA crawler.

The Connector For File Systems uses a graph traversal strategy, submitting a single URL representing the root of the file system to the GSA in a metadata-and-url feed, then returning URLs for all descendants of the root via crawl requests from the GSA.

The following process provides an overview of how the search appliance gets content from the repository through the Connector for File Systems.

- 1. The Connector For File Systems generates a Docld identifying the root of the file system to traverse.
- 2. The connector constructs a URL from the Docld and pushes it and the Access Control List (ACL) of the file share to the search appliance in a metadata-and-URL feed. Take note that this feed does not include the document contents.
- 3. The search appliance gets the URL to crawl from the feed.
- 4. The search appliance crawls the repository according to its own crawl schedule, as specified in the GSA Admin Console. It crawls the content by sending GET requests for content to the connector. If the content is in HTML format, the search appliance follows links within the page.
- 5. The connector receives a crawl request from the GSA. If the requested Docld is a regular file, the connector returns that file's contents to the GSA. It also includes the file's ACL and some basic metadata in the response. If the requested Docld is for a directory, the connector generates Doclds for each file and folder contained within that directory. The connector then constructs an HTML document consisting of links to URLs constructed from those Doclds. The connector returns the generated HTML as the content and the directory's ACL as metadata.

In addition to the directed graph traversal described above, the Connector For File Systems registers a file system change notification handler. This handler receives notifications when files or folders are added, removed, moved, modified, or have changes in metadata (including ACLs). The connector generates Doclds for the changed files and folders, constructs URLs from those Doclds, and sends them to the GSA in a metadata-and-URL feed. Configuring the root of the adaptor to a DFS namespace is supported. Configuring the root to a DFS link is supported.

#### Continuous automatic updates

By default, the connector starts monitoring for changes immediately. Updates and changes to content or access-controls are immediately sent to GSA with requests to re-crawl. You can turn this feature off/on by setting the value in the connector configuration option filesystemadaptor.monitorForUpdates, as described in <u>Step 3 Configure adaptor-config.properties variables</u>.

#### **DFS access control**

The DFS system employs access control when navigating its links, and usually each DFS Link has its own ACL. One of the mechanisms employed by this is *Access-based Enumeration* (ABE). With ABE deployed, users may only see a subset of the DFS Links, possibly only one when ABE is used to isolate hosted home directories. When traversing a DFS system, the connector supplies the DFS Link ACL, in addition to the target's Share ACL as a named resource when the DFS Link is crawled. In this case, the Share ACL inherits from the DFS ACL.

#### Supported operating systems for the connector

The Connector for File Systems must be installed on one of the following supported Windows operating systems:

- Windows Server 2012
- Windows Server 2008 R2
- Windows Server 2003

The Connector for File Systems does not run on Linux.

#### Supported file system protocols

The following table lists file system protocols used to communicate with file shares and indicates if the connector supports them.

File System Protocol	Communicating with Shares on Operating System	Supported ?
Server Message Block (SMB) 1	Windows Server 2012 Windows Server 2008 R2 Windows Server 2003	Yes
SMB 2	Windows Server 2012 Windows Server 2008 R2	Yes
Distributed File System (DFS)	Windows Server 2012 Windows Server 2008 R2 Windows Server 2003	Yes
Local Windows file system	Windows Server 2012 Windows Server 2008 R2 Windows Server 2003	No
Sun Network File System (NFS) 2.0		No
Sun Network File System (NFS) 3.0		No
Local Linux file system		No

#### **Known limitations**

**File System limitation** 

This release of the file system connector does not support mapped drives and local drives.

**Distributed File System limitation** 

A mapped drive to a UNC DFS does not work correctly. Some ACLs will not be read correctly.

# Before you deploy the Connector for File Systems

Before you deploy the Connector for File Systems, ensure that your environment has all of the following required components:

- GSA software version 7.2.0.G.90 or higher To download GSA software, visit the <u>Google for Work Support Portal</u> (password required).
- Java JRE 1.7 update 6 or higher installed on computer that runs the connector
- Connector for File Systems 4.0.4 JAR executable
   For information about finding the JAR executable, see <u>Step 2 Install the Connector</u> <u>for File Systems</u>.
- Ensure that the Windows account has sufficient permissions, as described in the following section.
- When sharing a folder from a Windows platform, permission must be given at the share ACL and the NTFS ACL of the folder. Both ACLs need to give the connector appropriate access. Both ACLs are also read by the connector.

#### Windows account permissions needed by the connector

The Windows account that the connector is running under must have sufficient permissions to perform the following actions:

- List the content of folders
- Read the content of documents
- Read attributes of files and folders
- Read permissions (ACLs) for both files and folders
- Write basic attributes permissions

The connector attempts to restore the last access date for documents after it reads the document content during a crawl. In order for the last access date to be restored back to the original value before the content was read, the user account that the connector is running under needs to have write permission. If the account has read-only permission and not write permission for documents, then the last access date for documents will change as the connector reads document content during a crawl.

Membership in one of the following groups grants a Windows account the sufficient permissions needed by the connector:

- Administrators
- Power Users
- Print Operators
- Server Operators

**Note**: It is not sufficient for the user to be a member of one of these groups at the domain level. The user must be a member of one of these groups on the local machine that exports the Windows share. For more information see the Microsoft documentation on the NetShareGetinfo function.

# Deploy the Connector for File Systems

Because the Connector for File Systems is installed on a separate host, you must establish a relationship between the connector and the search appliance.

To deploy the Connector for File Systems, perform the following tasks:

- 1. Specify the IP address of the host computer
- 2. Install the Connector for File Systems
- 3. Optionally, configure adaptor-config.properties variables
- 4. Run the Connector for File Systems

#### Step 1 Specify the IP address of the host computer

For the search appliance to work with the Connector for File Systems, the search appliance needs to be able to accept feeds from the connector. To set up this capability, add the IP address of the computer that hosts the connector to the list of Trusted IP addresses:

- 1. In the search appliance Admin Console, click **Content Sources > Feeds**.
- 2. Under List of Trusted IP Addresses, select Only trust feeds from these IP addresses.
- 3. Add the IP address for the connector to the list.
- 4. Click **Save**.

#### **Step 2 Install the Connector for File Systems**

This section describes the installation process for the Google Search Appliance Connector for File Systems on the connector host computer. This connector version does not support installing the connector on the Google Search Appliance.

You can install the Connector for File Systems on a host running one of the <u>supported</u> <u>Windows operating systems</u>.

To install the Connector for File Systems:

- 1. Log in to the computer that will host the connector by using an account with sufficient privileges to install the software.
- 2. Start a web browser.

- Visit the connector 4.0.4 software downloads page at <a href="http://googlegsa.github.io/adaptor/index.html">http://googlegsa.github.io/adaptor/index.html</a>. Download the exe file by clicking on Microsoft Windows Shares in the Windows Installer table. You are prompted to save the single binary file, fs-install-4.0.4.exe.
- 4. Start installing the file by double clicking fs-install-4.0.4.
- 5. On the **Introduction** page, click **Next**.
- 6. On the **GSA Hostname and other required configuration values** page, enter values for the following options:
  - GSA Hostname or IP address of the GSA that will use the connector.
     For example, enter gsa.hostname=yourgsa.example.com or
     IP address filesystemadaptor.src=\\\\host\\share

#### Notes:

Backslashes are entered as double backslashes. To represent a single '\' you need to enter '\\'.

If you are indexing a folder on DFS, DFS links can be given as
filesystemadaptor.src: \\\\host\\dfsnamespace\\link

UNICODE, as well as non-ASCII, characters can be used in filesystemadaptor.src. Including these characters will require the adaptor-config.properties file to be saved using UTF-8 encoding.

Multiple file systems may be specified as a semicolon separated list of paths. The multiple sources may be a combination of file shares, DFS namespaces, or DFS links. For instance: filesystemadaptor.src=\\\\host\\share;\\\\host\\dfsnamespac

e1; \\\host\\dfsnamespace2 Consider using property file multi-line value support (using '\' at end of line to

indicate line-continuation) for a more readable specification of multiple sources. The delimiter character is configurable in case your pathnames actually contain semicolons.

• **Host and directory to share** (not validated) Top level directory of the file system to be crawled.

#### • Adaptor port number

Port from which documents are served. GSA crawls this port. Each instance of a Connector on same machine requires a unique port. The default is 5678.

#### • **Server dashboard port** for the Connector Dashboard.

The value is the port on which to view web page showing information and diagnostics about the connector. The default is 5679.

- Whether or not to run the connector after the installer finishes.
- 7. Click Next.
- 8. On the **Choose Install Folder** page, accept the default folder or navigate to the location where you want to install the connector files.
- 9. Click **Next**.
- 10. On the **Choose Shortcut Folder** page, accept the default folder or select the locations where you want to create product icons.
- 11. To create icons for all users of the Windows machine where you are installing the connector, check **Create Icons for All Users** and click **Next**.
- 12. On the **Pre-Installation Summary page**, review the information and click **Install**. The connector Installation process runs.
- 13. On the **Install Complete** page, click **Done**. If you selected the option to run the connector after the installer finishes, the connector starts up in a separate window.
- 14. To enable the search appliance to crawl the repository's content, add the URL provided by the connector to the search appliance's crawl configuration follow patterns:
  - In the search appliance Admin Console, click Content Sources > Web Crawl
     > Start and Block URLs.
  - Under Follow Patterns, add the URL that contains the hostname of the machine that hosts the connector and the port where the connector runs. For example, you might enter http://connector.example.com:5678/doc/ where connector.example.com is the hostname of the machine that hosts the connector.

By default the connector runs on port 5678.

• Click **Save**.

15. In the folder where you installed the connector, review, and if needed, edit logging.properties. For more information, See "Configure Connector Logs" in the Administration Guide.

#### Step 3 Configure adaptor-config.properties variables

Optionally, you can edit or add additional configuration variables to the adaptorconfig.properties file. The following table lists the most important variables that pertain to the Connector for File Systems, as well as their default values. See also "Common configuration options" in the the <u>Administration Guide</u>.

Variable	Description	Default
adaptor.namespace	Namespace used for ACLs sent to GSA	Default
filesystemadaptor. monitorForUpdates	Enable/disable filesystem change monitoring. When monitoring is disabled updates/changes to content or access-controls are not immediately sent to GSA with request to re-crawl. Turning off monitoring reduces connector's resource use significantly.	true
filesystemadaptor.src. separator	Multiple source file systems may be specified for the filesystemadaptor.src property by supplying a list of UNC sources, separated by the delimiter configured by filesystemadaptor.src.separator. The default separator is ";" (similar to how one would set the PATH or CLASS_PATH environment variable). However, if your specified source paths contain semicolons, you can configure a different delimiter that does not conflict with characters in your paths, and is not reserved by property file syntax itself.	;

	If the filesystemadaptor.src.separator is set to the empty string, then the filesystemadaptor.src value is considered to be a single pathname.	
filesystemadaptor. supportedAccounts	Accounts that are in the supportedAccounts will be included in ACLs regardless if they are builtin or not.	BUILTIN\\Administrators,\\ Everyone,BUILTIN\\Users, BUILTIN\\Guest,NT AUTHORITY\\INTERACTIVE, NT AUTHORITY\\Authenticated Users
filesystemadaptor. builtinGroupPrefix	Builtin accounts are excluded from the ACLs that are pushed to the GSA. An account that starts with this prefix is considered a builtin account and will be excluded from the ACLs.	BUILTIN\\
filesystemadaptor. crawlHiddenFiles	This boolean configuration property allows or disallows indexing of hidden files and folders. The definition of hidden files and folders is platform dependent. On Windows file systems, a file or folder is considered hidden if the DOS hidden attribute is set. By default, hidden files are not indexed and the contents of hidden folders are not indexed. Setting filesystemadaptor.crawlHid denFiles to true will allow hidden files and folders to be crawled by the search appliance.	false
filesystemadaptor. skipShareAccessControl	This boolean configuration property enables or disables sending the Access Control List (ACL) for the file share to the GSA. See <u>Skipping File Share Access</u> <u>Control</u> .	false (share ACLs are sent to the GSA)
filesystemadaptor. lastAccessedDate	Disables crawling of files whose time of last access is earlier than	disabled

	a specific date. The cut-off date is specified in <u>ISO8601</u> date format, YYYY-MM-DD. Setting filesystemadaptor.lastAccess edDate to 2010-01-01 would only crawl content that has been accessed since the beginning of 2010. Only one of filesystemadaptor. lastAccessedDate Or filesystemadaptor.lastAccess edDays may be specified.	
filesystemadaptor. lastAccessedDays	Disables crawling of files that have not been accessed within the specified number of days. Unlike the absolute cut-off date used by filesystemadaptor.lastAccess edDate, this property can be used to expire previously indexed content if it has not been accessed in a while. The expiration window is specified as a positive integer for number of days. Setting filesystemadaptor.lastAccess edDays to 365 would only crawl content that has been accessed in the last year. Only one of filesystemadaptor. lastAccessedDate Or filesystemadaptor.lastAccess edDays may be specified.	disabled
filesystemadaptor. lastModifiedDate	Disables crawling of files whose time of last access is earlier than a specific date. The cut-off date is specified in <u>ISO8601</u> date format, YYYY-MM-DD. Setting filesystemadaptor.lastModifi edDate to 2010-01-01 would only	disabled

	<pre>crawl content that has been modified since the beginning of 2010. Only one of filesystemadaptor.lastModifi edDate Or filesystemadaptor.lastModifi edDays may be specified.</pre>	
filesystemadaptor. lastModifiedDays	Disables crawling of files that have not been modified within the specified number of days. Unlike the absolute cut-off date used by filesystemadaptor.lastModifi edDate, this property can be used to expire previously indexed content if it has not been modified in a while. The expiration window is specified as a positive integer for number of days. Setting filesystemadaptor.lastModifi edDays to 365 would only crawl content that has been modified in the last year. Only one of filesystemadaptor.lastModifi edDate Or filesystemadaptor.lastModifi edDays may be specified.	disabled

#### Step 4 Run the Connector for File Systems

After you install the Connector for File Systems, you can run it on the host machine by using a command like the following example:

```
java -Djava.util.logging.config.file=logging.properties -jar adaptor-
fs-4.0.4-withlib.jar
```

Verify that the connector has started and is running by navigating to the Connector
Dashboard at http://<CONNECTOR\_HOST>:<nnnn>/dashboard or
https://<CONNECTOR\_HOST>:<nnnn>/dashboard

where <nnnn> is the number you specified as the value for the server.dashboardPort in the configuration file.

To run the connector as a service, use the Windows service management tool or run the prunsrv command, as described in "Run a connector as a service on Windows" in the Administration Guide.

**Note**: By default the Connector for File Systems service runs using the Windows Local System account. This should be fine in most cases but this can cause issues if access to documents is restricted through ACLs. In cases where the Connector for File Systems service is not able to crawl documents due to ACL restrictions, you would need to specify a user for the Connector for File Systems service through the Windows Service Control Manager that has sufficient access to crawl the documents.

# Advanced Topics

#### Not changing 'last access' of the documents on the share

The connector attempts to restore the last access date for documents after it reads the document content during a crawl. In order for the last access date to be restored back to the original value before the content was read, the user account that the connector is running under needs to have write permission. If the account has read-only permission and not write permission for documents, then the last access date for documents will change as the connector reads document content during a crawl.

#### **Skipping File Share Access Control**

The connector attempts to preserve access control integrity when sending Access Control Lists (ACLs) to the GSA. In general, only users that have access to a file share have access to the files maintained on that share, so the connector includes the share's ACL in those sent to the GSA. However, in some configurations, the connector may not have sufficient permissions to read the share ACL. In those instances, the broken share ACL will prevent all files maintained on that file share from appearing in search results. The GSA's Index Diagnostics for those will also indicate a broken inheritance chain.

If the share ACL cannot be read by the connector, the administrator may skip the attempt to read the share ACL by setting the filesystemadaptor.skipShareAccessControl configuration option to true. This feeds a highly permissive share ACL to the GSA, rather than the actual share ACL.

**WARNING:** Bypassing the file share access control may be inconsistent with your enterprise security policies. This may allow users that do not have access to the file share to see documents hosted by that file share in search results.

# Uninstall the Google Search Appliance Connector for File Systems

To uninstall the Connector for File Systems:

- 1. Navigate to the FS connector installation folder, **\_GSA\_FS\_Adaptor\_installation**.
- Click Uninstall\_GSA\_FS\_Adaptor.exe.
   The Uninstall GSA\_FS\_Adaptor page appears.
- Click Uninstall.
   Files are uninstalled.
- 4. Click **Done**.

# Troubleshoot the Connector for File Systems

For information about troubleshooting the Connector for File Systems, see "Troubleshoot Connectors," in the <u>Administration Guide</u>.