Google Search Appliance Connectors

Administration Guide

Google Search Appliance Connectors software version 4.0.3
Google Search Appliance software version 7.2

October 2014

Table of Contents

- About this Guide
- 1 About Connectors 4.0
  - The Lister/Retriever model
- 2 What's New in Connectors 4.0?
  - Resiliency
  - Easy monitoring and reporting
  - Off-Board installation
  - Connector configuration
  - Crawl configuration
  - Automatic updates every 15 minutes
  - Support for early binding only
  - SAML security messages
  - Simplified troubleshooting
  - Giving all users access to all documents
  - Saving feeds sent to the GSA
- 3 General Information
  - Supported connectors
  - Supported Google Search Appliance versions
  - Configuration properties file
  - Repository content relevancy
  - Secure crawling and serving configurations
  - Admin Console access
  - Secure mode
  - Supported ACL features
  - Mark all documents as public
  - Archive feeds
  - Reverse proxy setup
  - Download the connector software
  - Run a connector as a service on Windows
    - Register the connector as a service
    - Add optional parameters
      - Apache Daemon Logging
      - Service Username and Password
      - Jvm options
    - Run the connector as a service

- Stop running a connector

4 Enable Connector Security

- Certificate Authorities
- Self-signed certificates
- Create a self-signed certificate for the GSA
  - Firefox
  - Chrome
  - OpenSSL (command line)
- Create a self-signed certificate for the connector
- Exchange certificates
- Turn on security with the server.secure property
- Run in secure mode with self-signed certificates
- Enable stricter security

5 Configure Connector Logs

- Logging properties file
- Change the location of logs
- Change the logging level
- Change the log file size
- Change the number of log files

6 Monitor Connectors with the Dashboard

- Supported browsers
- Dashboard port number
- Log in to the Connector Dashboard
- View information about a connector
  - Version
  - Status
  - Statistics
  - Connector configuration
  - Recent log messages
- Start or restart feeds
- Encode sensitive values
- Download rich data about the connector

7 Troubleshoot Connectors

- Debug a connector by using a web browser
- Troubleshooting scenario
- Troubleshooting quick reference

8 Common configuration options

About this Guide

This Administration Guide is intended for anyone who needs to understand how to manage Google Search Appliance (GSA) Connectors 4.0. It provides overview information about the Connectors, as well as procedures that you can follow to install, configure, or monitor each of the Connectors.

The guide assumes that you are familiar with Windows or Linux operating systems and configuring the Google Search Appliance by using the Admin Console.

For information about installing and configuring the connectors, see the following guides:

- Deploying the Connector for SharePoint
- Deploying the Connector for SharePoint User Profiles
- Deploying the Connector for File Systems
- Deploying the Connector for Active Directory

These guides, as well as information about using the Admin Console are available from the Google Search Appliance Help Center.

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

1 About Connectors 4.0

Google Search Appliance connectors enable the Google Search Appliance to acquire content from external repositories and provide that content in search results. A Google Search Appliance with configured connectors can perform fast, unified, secure search across multiple systems and document repositories.

A fundamental strength of the search appliance is discovering enterprise content in web pages and indexing it. The GSA accomplishes this by crawling the web pages over HTTP/HTTPS, following hyperlinks within the pages to interrelated web pages, and adding the content it discovers to the search index. Ultimately, the GSA serves content from its index as search results to end users.

However, many organizations have content that is stored in repositories, such as SharePoint and Windows file shares, rather than on web pages. Because documents in repositories are not usually interrelated through hyperlinks, the search appliance cannot find this content through normal crawling.

Connectors 4.0 exploit the search appliance’s strengths by enabling it to crawl non-web content in repositories over HTTP/HTTPS. Additionally, connectors can feed groups information to the search appliance. Groups information can restrict the visibility of certain content to members of particular groups by using Access Control Lists (ACLs).

The search appliance adds content acquired through connectors to the search index and uses credentials provided by connectors to protect secure content.

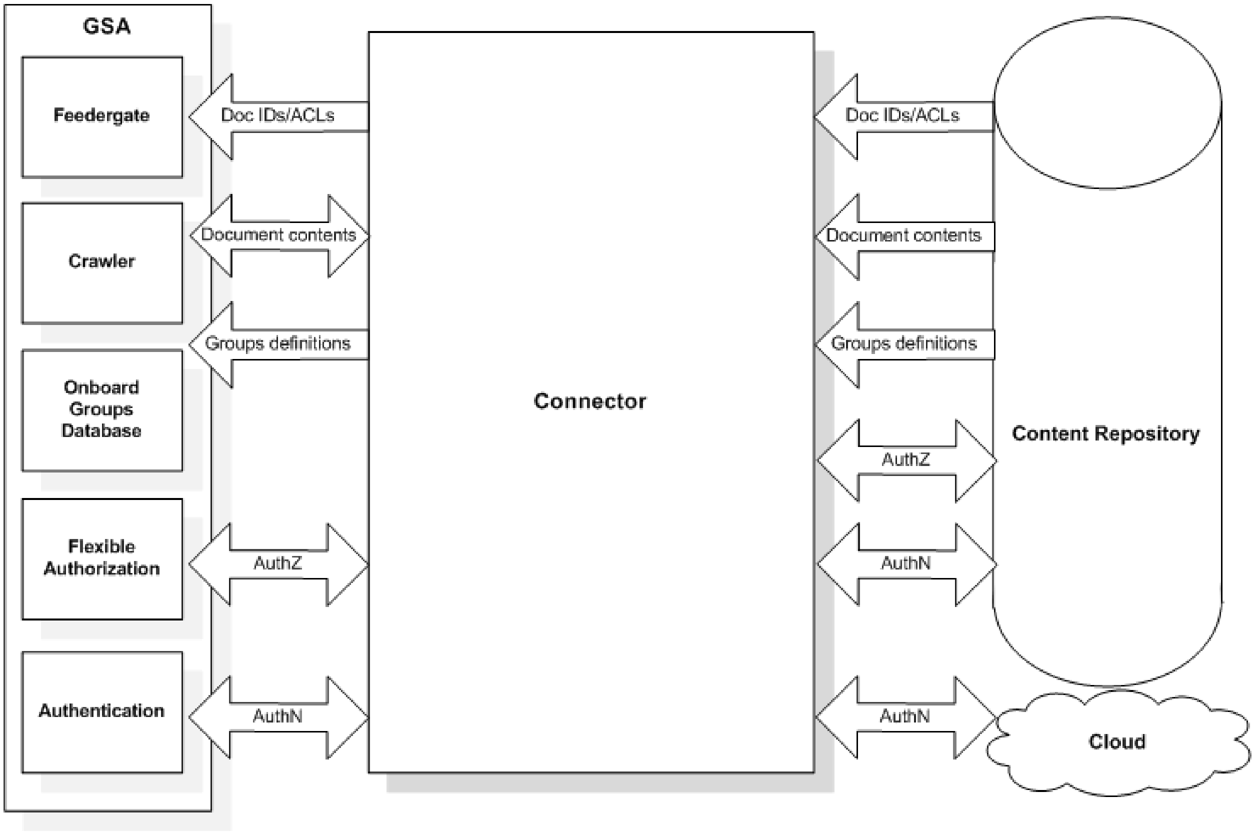
There are several ways to model and communicate your repository's contents to the GSA, and Adaptors are one of them. For other possible solutions, look into Connectors 3.x and Content Feeds. Connectors 3.x support older GSA versions. Content Feeds should be used when the repository does not provide random document access and instead only provides changes occurring in the repository.

The Lister/Retriever model

Connectors 4.0 are based on the lister/retriever model. In this model, the lister notifies the search appliance of the names of documents in the repository and the retriever sends URLs and ACLs in feeds to the search appliance. The search appliance uses the URLs to crawl documents in the repository over HTTP/HTTPS. Each document in a repository is identified by a unique document identifier (DocId).

The search appliance performs the task of determining which DocIds the user can access and handles authentication requests between the search appliance and the repository.

The following diagram shows how the connectors interact with the search appliance and the repository.



2 What's New in Connectors 4.0?

In release 4.0, connectors work seamlessly with more search appliance features than previous releases. Noteworthy features of Connectors 4.0 include:

- Resiliency
- Easy monitoring and reporting
- Off-Board installation
- Connector configuration
- Crawl configuration
- Updates every 15 minutes
- Support for early binding only
- SAML security messages
- Simplified troubleshooting
- Giving all users access to all documents
- Saving feeds sent to the GSA

Resiliency

Because Connectors 4.0 are "stateless," they are more resilient than previous versions. For example, in previous versions, you had to restart the connectors when errors occurred. In this version, when an error occurs, you can ignore it because it will be corrected without administrator intervention.

Easy monitoring and reporting

The Connector Dashboard is a new, web-based interface that eases monitoring and reporting. It displays the connector's version, status, statistics, configuration variables and values, response time and available throughput in real time (including the previous hour and the previous day), and recent log messages.

Also, from the Connector Dashboard, you can download the **Diagnostics zip** archive, which contains logs and other rich data about the connector.

For more information about the Connector Dashboard, see Monitor Connectors with the Dashboard.

Off-Board installation

All Connectors 4.0 are installed on a separate host server rather than the search appliance itself. For more details on this topic, see Download the connector software.

Connector configuration

Because Connectors 4.0 are not built-in the search appliance, they are not configured through the search appliance Admin Console, as in previous releases. Configuration is handled in the `adaptor-config.properties` file. In release 4.0, “adaptor” is the internal name of the architecture used by the connectors.

Crawl configuration

As part of deploying a connector, you configure crawl of repository content by using the search appliance Admin Console. Once crawling starts, you can use the search appliance’s robust features for monitoring and fine-tuning crawl. Take note that Connectors 4.0 crawl content according to the crawl schedule that is configured in the Admin Console.

Because some repositories make it easy to notice what has changed recently, adaptors for those repositories can inform the GSA immediately when things happen, and even force the GSA to immediately recrawl a document because the adaptor knows that document was modified recently. This can allow the GSA to pick up changes to a repository with very low latency, but is not required for correct operation.

For information about configuring crawl by using the Admin Console, see Administering Crawl.

Automatic updates every 15 minutes

Every 15 minutes, the Connectors for SharePoint, File Systems, and Active Directory provide recent changes to the GSA. You can change the default time interval between updates by editing the `adaptor.incrementalPollPeriodSecs` configuration option.

Support for early binding only

Connectors 4.0 enable support for authorization by early binding only; late binding is not supported. If your implementation requires late binding, Google recommends using a previous version of connectors.

SAML security messages

In this release, SAML is the communication protocol between the search appliance and the connector for user authentication and authorization. This protocol replaces XML, which was used in previous releases.

Simplified troubleshooting

A strength of Connectors 4.0 is simplified troubleshooting. If you encounter issues, you have several options for troubleshooting them, including the Connector Dashboard, as well as index and real-time diagnostics on the search appliance, and logs on the connector machine. For detailed information, see Troubleshoot Connectors.

Giving all users access to all documents

Connectors 4.0.3 introduces the capability of giving all users access to all documents. For information about this capability, see Mark all documents as public.

Saving feeds sent to the GSA

Connectors 4.0.3 introduces the capability of archiving all feeds sent to the GSA on the local drive. For more information, see Archive feeds.

3 General Information

This section contains general information about Connectors 4.0, including:

- Supported connectors
- Supported Google Search Appliance versions
- Configuration properties file
- Repository content relevancy
- Secure crawling and serving configurations
- Admin Console access
- Secure mode
- Supported ACL features
- Mark all documents as public
- Archive feeds on the local drive
- Reverse proxy setup
- Download the connector software
- Run a connector as a service
- Stop running a connector

Supported connectors

In version 4.0, Google supports the following connectors:

- Connector for SharePoint
- Connector for SharePoint User Profiles
- Connector for File Systems
- Connector for Active Directory

Guides for deploying each connector are available from the Google Search Appliance Help Center.

Supported Google Search Appliance versions

Connectors 4.0.3 work with Google Search Appliance version 7.2.0.G.90 or higher. If you plan on deploying the Connector for Active Directory and you need to support over 1 million group memberships (which includes nested groups and users), then use GSA software version 7.2.0.G.230 or higher.

Configuration properties file

Configuration is handled in the `adaptor-config.properties` file. Each connector installation procedure in this documentation contains steps for creating its `adaptor-config.properties` file and adding minimal variables to make the connector operational. For example, after you create the configuration file, you need to add the `server.port` variable to point to the retriever port.
For more information about configuration, see the section that pertains to your connector.

Additionally, there are common configuration variables, which are used by all connectors. If you do not indicate values for these variables, defaults are used. For more information about this topic, see Common configuration options.

Repository content relevancy

The search appliance determines the relevancy in search results of a document that it crawls on the web by using a "page rank" algorithm, which is based on an analysis of hyperlinks among documents. A search appliance administrator can view the relative page rank of a document by using the **Index > Diagnostics > Index Diagnostics** page in the Admin Console.

By default, the search appliance sets the relevancy of a document that it crawls in a repository by using a "content rank" algorithm, which is based on analysis of the document's content. If a document has a content rank, its relative page rank on the **Index > Diagnostics > Index Diagnostics** page is zero.

The configuration variable `gsa.scoringType` controls whether the search appliance uses a content rank algorithm or page rank algorithm for repository content. Valid values for this variable are:

- `content` (content rank)
- `web` (page rank)

If the connector uses the `web` scoring type, you must add the connector root URL as a **Start URL** on the **Content Sources > Web Crawl > Start and Block URLs** page in the GSA Admin Console. Otherwise, all documents will have a page rank of 1.

Secure crawling and serving configurations

Connectors 4.0 support the authentication and authorization configurations for crawling and serving that the GSA administrator configures for the search appliance.

For information about secure crawling and serving configurations, see Managing Search for Controlled-Access Content.

Admin Console access

If the search appliance only allows HTTPS access to the Admin Console, then the connector must be running in secure mode. In secure mode, use HTTPS to access the Connector Dashboard.

To disable HTTPS only access to the Admin Console, select **Enable HTTP (i.e. non SSL) Admin Console and Version Manager access** on the **Administration > System Settings** page in the Admin Console. When HTTP access to the Admin Console is enabled, you can use the Connector Dashboard with or without security enabled for the connector.

Secure mode

Connectors 4.0 support communication in secure mode over HTTPS. You can enable secure mode for any connector, but Google strongly suggests that you enable security for the Connector for SharePoint, and the Connector for SharePoint User Profiles.

For more information on this topic, see Enable connector security.

Supported ACL features

Access Control Lists (ACLs) control which documents a user can see. The search appliance needs to crawl and index all documents, but still rapidly determine which documents a specific user is allowed to view in a search.

All the 4.0 connectors feed ACLs at crawl time, using a separate channel from the other feeds. Access Control Lists (ACLs) may be inherited from a parent. This reduces the number of ACLs that require re-indexing. Connectors send full-fidelity ACLs, which include inheritance and can contain local groups.

The connectors also support Deny ACLs - ACLs which deny access to specific individuals or groups, local and global namespaces for ACL users and groups and Active Directory groups.

ACLs are stored in the search appliance’s group database.

Mark all documents as public

Adding the variable `adaptor.markAllDocsAsPublic=true` to the `adaptor-config.properties` file enables you to treat all users as if they are members of all groups, thereby giving them access to all documents. The default value for `adaptor.markAllDocsAsPublic` is “false.”

Take note that if this option is set to “true” for the Connector for Active Directory, the connector does not send a feed specifying group membership to the search appliance.

Archive feeds

Adding the variable `feed.archiveDirectory` with a valid path to the `adaptor-config.properties` file enables you to save feeds to the specified directory on the local drive as they are sent to the GSA. All feeds successfully and unsuccessfully sent to the GSA are archived. Failed feeds are tagged with FAILED in the archive feed file name. The feeds contain listed document-ids, named resources, and group definitions

Reverse proxy setup

You can add a reverse proxy to your configuration as an intermediary for crawl requests from the search appliance to the connectors. For example, you might configure a proxy server (`PROXYHOST`) between the search appliance and multiple connectors in a round-robin setup.

To set up a reverse proxy:

1. Configure `server.hostname` as the proxy server instead of the server running the connector. In the previous example, you would configure `server.hostname=PROXYHOST`
2. Optionally configure `server.reverseProxyPort` (defaults to server.port). This option controls the port used in retriever URLs.
3. Optionally configure `server.reverseProxyProtocol` to either http or https, depending on proxy traffic (defaults to https in secure mode and http otherwise).

Download the connector software

All connectors 4.0 must be installed on a host machine. This connector version does not support installing connectors on the Google Search Appliance.

To download the software for a connector, visit http://googlegsa.github.io/adaptor/index.html. Executables are available for all the 4.0 connectors. Google provides the installation software for each 4.0 connector in a single binary file, as listed in the following table.

Repository	Connector	Executable
SharePoint	SharePoint	`sp-install-4.0.3.exe`
SharePoint User Profile Service Application	SharePoint User Profiles	`spup-install-4.0.3`
Microsoft Windows Shares	File System	`fs-install-4.0.3.exe`
Microsoft Active Directory	Active Directory	`ad-install-4.0.3.exe`

For information about installing your connector, see the appropriate connector deployment guide, as listed in About this Guide.

Run a connector as a service on Windows

When you run a connector as a service, you do not need to run it manually. The connector runs when the host server starts up, and shuts down when the host is shut down. Before running the connector as a service, register it, as described in the following section

Register the connector as a service

You register the connector as a service by running the `prunsrv` command, as shown in the following procedure. Take note that you can increase the power of the `prunsrv` command by adding optional parameters. For example, you can specify logging for the Apache daemon by using the `LogPath` parameter. For detailed information , see “Add optional parameters.”

To register a connector as a service:

1. Download and extract prunsrv.exe from the latest Windows binary download of Apache Commons Daemon. If you are running on 64-bit Windows and will use a 64-bit JVM, then you should use the prunsrv.exe in the amd64/ directory.
2. Place prunsrv.exe in the same directory as the connector you would like to run as a service.
3. In the same directory where the connector `.jar` files are installed, run the following command:

   ```
   prunsrv install <CONNECTOR-NAME> ^
    --StartPath="<STARTPATH>" ^
    --Classpath=<CONNECTOR-JAR> ^
    --StartMode=jvm ^
    --StartClass=com.google.enterprise.adaptor.Daemon ^
    --StartMethod=serviceStart ^
    --StartParams=<FULL-CONNECTOR-CLASSNAME> ^
    --StopMode=jvm ^
    --StopClass=com.google.enterprise.adaptor.Daemon ^
    --StopMethod=serviceStop ^
    --StdOutput=<OUTPUT-LOG> ^
    --StdError= <ERROR-LOG> ^
    --Jvm=<JVM-DLL> ^
    --Startup=auto
   ```

   Where:

   `<CONNECTOR-NAME>` is the name of the connector in the list of running services:
   SharePoint: `adaptor-sharepoint`
   SharePoint User Profiles: `adaptor-sharepoint-user-profile`
   File Systems: `adaptor-fs`
   Active Directory: `adaptor-ad`

   `<STARTPATH>` is the absolute path of the `StartPath`, for example `"C:\Users\administrator.GSA\Desktop\Connector"`

   `<CONNECTOR-JAR>` is the the name of the connector `.jar`, for example, `adaptor-sharepoint-4.0.3-withlib.jar`

   `<FULL-CONNECTOR-CLASSNAME>` is one of the following values:
   `com.google.enterprise.adaptor.sharepoint.SharePointAdaptor`

```
com.google.enterprise.adaptor.sharepoint.SharePointUserProfileAdaptor
com.google.enterprise.adaptor.fs.FsAdaptor
com.google.enterprise.adaptor.ad.AdAdaptor
```

`<OUTPUT-LOG>` is the full path to the output log, for example, `C:\sp\logs\stdout.log`

`<ERROR-LOG>` is the full path to the error log, for example, `C:\sp\logs\stderr.log`
`<JVM-DLL>` is the path to where Java Virtual Machine dynamic link library is installed, for example, `C:\Java\jdk1.7.0_67\jre\bin\server\jvm.dll`

An alternative to specifying the JVM on the command line with the `Jvm` parameter is to configure the default JVM with the Java Control panel `(javacpl.exe).` Be sure to update the service registration each time you update the JVM.

Add optional parameters

You can add important optional parameters to the prunsrv command to specify:

- Apache daemon logging
- Service username and password
- JVM options

Apache Daemon Logging

You can specify logging for the Apache Daemon by using the `LogPath` parameter`.` The default value is:

```
%SystemRoot%\System32\LogFiles\Apache
```

where `SystemRoot` is a root path, for example `C:\Windows`

Service Username and Password

Use the `ServiceUser` parameter to specify the name of the account under which the service should run, as shown in the following example:

```
  --ServiceUser DOMAINname\username ^
```

Use the `ServicePassword` to specify the password for the account designated by the ServiceUser parameter, as shown in the following example:

```
  --ServicePassword password ^
```

Jvm options

Use the `JvmOptions` parameter to specify a `JvList` of options in the form of `-D` or `-X` that will be passed to the JVM, as shown in the following example:

```
 ++JvmOptions=-Djava.util.logging.config.file=logging.properties
```

Run the connector as a service

To run a connector as a service, run the following command in the same directory where the connector .jar files are installed:

```
prunsrv start <CONNECTOR-NAME>
```

Stop running a connector

To stop running a connector in Windows or Linux, close the connector command prompt on the host.

You can stop running a connector as a service on Windows from either the service list or the command line.

To stop running a connector as a service on Windows from the service list:

1. On the connector host, choose **Start > Run > services.msc**
2. Select the connector service.
3. Click **Stop**.

To stop running a connector as a service on Windows from the command line, enter the following command on the host:

```
prunsrv stop <CONNECTOR_NAME>
```

Where `<CONNECTOR-NAME>` is the internal name of the connector:

- SharePoint: `adaptor-sharepoint`
- SharePoint User Profiles: `adaptor-sharepoint-user-profile`
- File Systems: `adaptor-fs`
- Active Directory: `adaptor-ad`

To stop running a connector by using the search appliance Admin Console, perform either or both of the following actions:

1. On the **Content Sources > Diagnostics > Crawl Status** page, click **Pause Crawl**.
2. On the **Content Sources > Web Crawl > Start and Block URLs** page, remove the connector URL from the **Follow Patterns**.

4 Enable Connector Security

In secure mode, the connectors communicate with the Google Search Appliance over HTTPS. You can enable security for any connector by configuring certificates and turning on security.

Take note that you must enable security for the Connector for SharePoint and the Connector for SharePoint User Profiles.

Secure mode supports using either of the following types of certificates:

- Certificate Authorities (CA's)
- Self-signed certificates

In either case, you can also choose options to enable stricter security.

Certificate Authorities

The GSA and the connector executable both have default Certificate Authorities; public keys are already in the GSA and connector trust stores. For the connector, you can find the default keystore CAs under jre\lib\security\.

If you are using the default CA's only, complete the tasks described in the following sections:

- Exchange certificates
- Turn on security

By default, the search appliance alias is "gsa" and the connector alias is "adaptor." Optionally, you can configure either alias.

Self-signed certificates

If you need to create self-signed certificates before turning on security, complete the tasks described in the following sections:

- Create a self-signed certificate for the GSA
- Create a self-signed certificate for the connector

- Exchange certificates
- Turn on security with the server.secure property

Create a self-signed certificate for the GSA

For information about creating a self-signed certificate for the search appliance, see the GSA Admin Console help page for Administration > SSL Settings.

To get the GSA's freshly-created certificate to add it as a trusted host for the connector, follow the procedure for your preferred browser or the command line.

Firefox

1. Navigate to the GSA's secure search: https://gsahostname/.
   A warning page appears with the following message: "This Connection is Untrusted." This message appears because the certificate is self-signed and not signed by a trusted Certificate Authority.
2. Click, "I Understand the Risks" and "Add Exception."
3. Wait until the "View..." button is clickable, then click it.
4. Change to the "Details" tab and click "Export...".
5. Save the certificate in your connector's directory with the name "gsa.crt".
6. Click **Close** and **Cancel** to close the windows.

Chrome

1. Navigate to the GSA's secure search: https://gsahostname/.
   A warning page appears with the following message: "The site's security certificate is not trusted!" In the location bar, there should be a padlock with a red 'x' on it.
2. Click the padlock and then click "Certificate Information."
3. Change to the "Details" tab and click "Export...".
4. Save the certificate in your adaptor's directory with the name "gsa.crt".
5. Click **Close** and **Cancel** to close the windows.

OpenSSL (command line)

1. Execute the following command:
   ```
   openssl s_client -connect gsahostname:443 < /dev/null
   ```
2. Copy the section that begins with `-----BEGIN CERTIFICATE-----` and ends with `-----END CERTIFICATE-----` (including the `BEGIN` and `END CERTIFICATE` portions) into a new file.
3. Save the file in your connector's directory with the name "gsa.crt".

Create a self-signed certificate for the connector

Generate a self-signed certificate for the connector and export the newly created certificate.

1. Within the connector’s directory, run the following command:
   ```
   keytool -genkeypair -keystore keys.jks -storepass changeit -keypass changeit -alias adaptor -keyalg RSA -validity 365
   ```
2. For "What is your first and last name?", enter the hostname of the connector’s computer. You are free to answer the other questions however you wish (including not answering them).
3. Answer "yes" to "Is CN=yourcomputershostname, OU=... correct?"
4. Still in connector’s directory, run the following command:
   ```
   keytool -exportcert -alias adaptor -keystore keys.jks -storepass changeit -keypass changeit -rfc -file adaptor.crt
   ```
5. Copy cacerts from Java to the connector's directory:

   For Windows, run the following command:
   ```
   copy PATH\TO\JRE\lib\security\cacerts cacerts.jks
   ```

   For Linux ,run the following command:
   ```
   cp PATH/TO/JRE/lib/security/cacerts cacerts.jks
   ```

6. To allow the connector to trust itself, run the following command:
   ```
   keytool -importcert -keystore cacerts.jks -storepass changeit -file adaptor.crt -alias adaptor
   ```

7. When prompted **Trust this certificate?**, answer yes.

Exchange certificates

To allow the connector to trust the search appliance:

1. On the connector host, run the following command:
   ```
   keytool -importcert -keystore cacerts.jks -storepass changeit -file gsa.crt -alias gsa
   ```
2. When prompted **Trust this certificate?**, answer yes.

To allow the search appliance to trust the connector:

1. In GSA Admin Console, click **Administration > Certificate Authorities**.

2. Under **Add more Certificate Authorities**, click **Browse**.
3. Navigate to the connector’s directory and select adaptor.crt.
4. Click **Save**.

Turn on security with the server.secure property

You can turn on security for the connector by using `server.secure` property, which enables HTTPS and certificate checking. Add the following line to your `adaptor-config.properties` file:

```
server.secure=true
```

When `server.secure=true`, the connector uses the GSA's authentication configuration and HTTPS for all communication. Also, when the value of `server.secure` is `true`, the following conditions apply:

- You need to add the key to the connector keystore with an alias defined in the connector config file, `server.keyAlias`.
- The connector runs on the configured port enforcing SSL.
- The Connector Dashboard runs on the configured port enforcing SSL.
- Feeds from the connector are forced to the search appliance secure Feedergate port (19902), even if the search appliance accepts feeds over HTTP.
- The connector validates the search appliance’s certificate during the SSL handshake.

Run in secure mode with self-signed certificates

If you are using one or more self-signed certificates in your configuration, you must run the connector with SSL settings, as shown in the following example command:

(Windows):

```
  java ^
    -Djava.util.logging.config.file=src/logging.properties ^
    -Djavax.net.ssl.keyStore=keys.jks ^
    -Djavax.net.ssl.keyStoreType=jks ^
    -Djavax.net.ssl.keyStorePassword=<password> ^
    -Djavax.net.ssl.trustStore=<truststore>.jks ^
    -Djavax.net.ssl.trustStoreType=jks ^
    -Djavax.net.ssl.trustStorePassword=changeit ^
    -classpath adaptor-name-4.0.3-withlib.jar ^
    com.google.enterprise.adaptor.name.NameAdaptor
```

(Linux / Unix systems):

```
java \
  -Djava.util.logging.config.file=src/logging.properties \
  -Djavax.net.ssl.keyStore=keys.jks \
  -Djavax.net.ssl.keyStoreType=jks \
  -Djavax.net.ssl.keyStorePassword=<password> \
  -Djavax.net.ssl.trustStore=<truststore>.jks \
  -Djavax.net.ssl.trustStoreType=jks \
  -Djavax.net.ssl.trustStorePassword=changeit \
  -classpath adaptor-name-4.0.3-withlib.jar \
  com.google.enterprise.adaptor.name.NameAdaptor
```

Enable stricter security

Optionally, you can improve security by choosing stricter security features on the **Administration > SSL Settings** page in the Admin Console, as described in the following table. However, using any of these options require the connector to be configured for security and have `server.secure=true` in its configuration.

Option	**Setting**	**Description**
Enable HTTP (non-SSL) access for Feedergate	Uncheck	When this option is unchecked, only HTTPS communications will be accepted by feedergate. Connectors send document ids to feedergate.
Enable Client Certificate Authentication for Feedergate	Check	When this option is checked, the Feedergate SSL port (19902) only accepts connections from IP addresses in the trusted IP addresses list and clients who present a valid x509 certificate when connecting. Valid means that the certificate is signed by a certificate in the CA keystore on the search appliance (or a certificate in the certificate chain).
Enable Server Certificate Authentication	Check	When this option is checked, it is a requirement for the crawler to authenticate certificates presented by servers that contain secure content.

You must include `server.secure=true` in the connector configuration before enabling these stricter features.

To enable stricter security, perform the following steps by using the GSA Admin Console:

1. Click **Administration > SSL Settings**.
2. Make any of the following changes on this page:
    a. Uncheck **Enable HTTP (non-SSL) access for Feedergate**.
    b. Check **Enable Client Certificate Authentication for Feedergate**.
    c. Check **Enable Server Certificate Authentication**.

3. Click **Save**.

5 Configure Connector Logs

The connectors log processing messages, including exceptions and warnings. Log messages appear in the Connector Dashboard and you can download the logs, as described in Download rich data about the connector.

Messages contain information about thread processing, including:

- Date stamp--Date and time the message was logged
- Name-of-thread--The thread that generated the message
- Last-30-characters-of-method--Code source for connector request
- Logging level--Filter log messages by level of severity
- Log-message--Text message for log entry

The following example shows a log message:

```
06-12 18:20:08.839 background URLConnection.getInputStream() FINE:
sun.net.www.MessageHeader@b20ccbf12 pairs:...
```

Logging properties file

Log configuration is controlled by the `logging.properties` file. Each connector installation procedure in this documentation contains a step for editing `logging.properties`. By editing values in this file, you can configure the following settings:

- Location of logs
- Logging level
- Log file size
- Number of log files

The following example shows a `logging.properties` file with default values.

```
.level=INFO
handlers=java.util.logging.FileHandler,java.util.logging.ConsoleHandler
java.util.logging.FileHandler.formatter=com.google.enterprise.adaptor.CustomF
ormatter
java.util.logging.FileHandler.pattern=logs/adaptor.%g.log
java.util.logging.FileHandler.limit=10485760
java.util.logging.FileHandler.count=20
java.util.logging.ConsoleHandler.formatter=com.google.enterprise.adaptor.Cust
omFormatter
com.google.enterprise.adaptor.CustomFormatter.useColor=true
```

Change the location of logs

By default, the logs are saved in `logs/adaptor.*.log`, in the same directory where the connector is running.

To change the location of log files, edit the `java.util.logging.FileHandler.pattern` value in the `logging.properties` file:

```
java.util.logging.FileHandler.pattern=logs/adaptor.%g.log
```

Change the logging level

You can filter messages written to log files by the following Java log levels:

- FATAL
- WARNING
- INFO
- FINE
- FINER
- FINEST

By default, the log level is INFO. The number of messages generated increases with each level, where FATAL logs the smallest number of messages and FINEST logs the largest.

To change the level of log files, edit the `.level` value in the `logging.properties` file:

```
.level=INFO
```

Change the log file size

By default, the size of connector log files is 10485760 bytes. Restarting the connector will create a new log file, regardless of how large the previous one had been.

You can change the size to suit your needs. The limit must be specified as a 32-bit integer, and thus has an upper limit of 2,147,483,647 (2 gigabytes, about 205 times as large as the default size).

To change the size of log files, edit the `java.util.logging.FileHandler.limit` value in the `logging.properties` file:

```
java.util.logging.FileHandler.limit=10485760
```

Change the number of log files

The connector writes to a log file until the size limit is reached, then starts writing to a new log file. By default, the connector writes to 20 log files, but you can change the number to suit your needs. There is no upper limit to the number of log files. After it finishes writing to the last log file, it starts writing over the first file.

To change the number of log files, edit the `java.util.logging.FileHandler.count` value in the `logging.properties` file:

```
java.util.logging.FileHandler.count=20
```

6 Monitor Connectors with the Dashboard

The Dashboard is a web-based interface that provides information about the connector's operation, with easy access to logs and error history.

Use the Connector Dashboard to perform the following tasks:

- View information about the connector
- Start or restart feeds
- Encode sensitive values
- Download rich data about the connector

You must start the connector to use the Dashboard.

Supported browsers

The Connector Dashboard runs in the following browsers:

- Google Chrome 22
- Internet Explorer 8 and 9
- Firefox 15 and 16
- Safari 5 and 6

Dashboard port number

By default, the Connector Dashboard uses port 5679. The port number is determined by the value of the variable `server.dashboardPort` in the `adaptor-config.properties` file for the connector. You can change the Connector Dashboard port number by changing the default value in this file. Every instance of a connector running on a host must have a unique value for `server.dashboardPort`.

Log in to the Connector Dashboard

To display the Connector Dashboard, open a browser and navigate to the following HTTP or HTTPS address:

```
http://<CONNECTOR_HOST>:<nnnn>/dashboard
```

or

```
https://<CONNECTOR_HOST>:<nnnn>/dashboard
```

where:

- `HTTP or HTTPS`--If you run the connector in secure mode, use HTTPS to log in to the Dashboard.
- `<CONNECTOR_HOST>` is the hostname or IP address of the host that is running the connector
- `<nnnn>` is the dashboard port number, as specified in the `adaptor-config.properties` file for the connector

To log in to the Connector Dashboard, use your search appliance user or administrator login credentials. You cannot log in to the Connector Dashboard with search appliance manager login credentials.

View information about a connector

You can use the Dashboard to monitor the connector by viewing up-to-date information, including:

- Version
- Status
- Statistics
- Connector (Adaptor) configuration
- Recent log messages

Version

In the Version section, the Dashboard displays information about the currently installed Java version, Connector (“Adaptor”) library version, Connector type, and Connector version.

Status

In the Status section, the Connector Dashboard displays the current status of the Java version (supported or not), feed pushing, the error rate of document retrieval from the repository (derived from logs), and search appliance crawling.

For each item, a signal indicates the status by color:

- Green for OK. The item is functioning.
- Yellow for alert. The item is not currently functioning, but no action is required. For example, the Dashboard displays yellow when the GSA is not currently crawling.
- Red for warning. The item is not functioning and requires attention.

Statistics

In the Statistics section, the Connector Dashboard displays the following information:

- A datestamp for when the connector program was started.
- Datestamps for the last successful push (full or incremental) start or end. The push can either be started automatically or manually.
- Status of the current push, if any.
- Total number of DocIDs pushed from the repository to the connector since the program started.
- Total number of requests for documents and unique documents from the GSA and the connector.
- Time resolution

The Statistics section also displays graphs showing throughput and response time for the last minute, last hour, and last day.

Connector configuration

In the configuration section, the Connector Dashboard displays the values for all the configuration variables in the `adaptor-config.properties` file.

Recent log messages

In the Recent Log Messages section, the Dashboard displays connector log messages. For more information on this topic, see Download rich data about the connector.

Start or restart feeds

The Connector Dashboard enables you to start or restart a full feed or an incremental feed as often as needed or when errors are detected. To start or restart a feed, click either **Run Incremental Push** or **Run Full Push**.

Encode sensitive values

You can encode passwords and other sensitive configuration values and copy them to the `adaptor-config.properties` file. Values can be specified in the configuration as prefix:data, where the prefix specifies how the value is stored.

You can encode the listed sensitive values for the following connectors:

- Connector for SharePoint--`sharepoint.password`
- Connector for SharePoint User Profiles--`sharepoint.password`
- Connector for Active Directory--`ad.defaultPassword`
  <any overriding password of any particular server>
- The Connector for File Systems does not support encoding sensitive values.

The value can be stored as:

- **Plain text** allowing the password or other information to be read by anybody who can read the configuration. Denoted by "pl" prefix.
- **Obfuscated** where the information is in a highly unreadable format, but it is possible for anyone to retrieve the original text. Denoted by "obf" prefix.
- **Encrypted** which uses your HTTPS encryption key to encrypt the value. Denoted by "pkc" prefix.

To encode a sensitive value:

1. Under **Storing Sensitive Values**, enter the sensitive value in the field.
2. Click a storage option.
3. Click **Encode Sensitive Value**.
   The encoded value appears.
4. Copy and paste the sensitive value into the `adaptor-config.properties` file.

Download rich data about the connector

The Diagnostics zip archive contains rich data about the connector, including:

- Current configuration settings (in the `config.txt` file)
- Connector version, status, and statistics (in the `state.txt` file)
- Thread details (in the `threaddump.txt` file)
- Logs folder

This data that can help you to diagnose connector issues. To download the archive, click **Diagnostics zip file** on the Dashboard.

7 Troubleshoot Connectors

Connectors 4.0 provide several options for troubleshooting issues, including:

- Connector Dashboard for checking the status of feeds and document retrieval
- Logs on connector machine for checking messages about thread processing
- Search appliance index diagnostics for checking crawl status
- Search appliance real-time diagnostics for checking HTTP headers for a specific URL at any time without having to wait for the crawler to ingest it
- Web browser with the connector host

Additionally, you can troubleshoot issues by examining URL-and-metadata feed files. Because these types of feed files are relatively small, troubleshooting them does not require significant effort.

Debug a connector by using a web browser

A connector, by default, will deny all document accesses, except from the search appliance. To allow debugging and testing a connector by using a browser without a search appliance, you can add a hostname to the `server.fullAccessHosts` configuration option to allow that computer full access to all connector content.

In addition, this setting allows that computer to see metadata and other GSA-specific information as HTTP headers. This capability can be very useful when combined with Firebug or the Web Inspector in your browser to observe a connector's behavior.

Troubleshooting scenario

In this scenario, users cannot find a specific document in search results, even though it is assumed to be in the search appliance index. To troubleshoot this issue, the administrator can track the document through the system by following the path a document takes to get into the search appliance index.

The administrator might perform one or more of the following steps:

1. Make sure that the search appliance is set to follow and crawl the Connector's URLs by checking the **Content Sources > Web Crawl > Start and Block URLs** page in the Admin Console.

2. Make sure GSA crawling is not paused by using the **Content Sources > Diagnostics > Crawl Status** page.
3. Check the Connector status and recent log messages by using the Dashboard.
4. Ensure that the Connector fed the document URL to the search appliance by examining the feed file.
5. Ensure that the search appliance got the document by using the **Index> Diagnostics > Index Diagnostics** page in the Admin Console.
6. Check the HTTP header for the document by using the **Content Sources > Diagnostics > Real-time Diagnostics** page in the Admin Console.
7. Ensure that the connector logged the document by checking connector log files. The Lister logs a file when it feeds it to the search appliance. The Retriever logs the file when the crawler requests the document.
8. Find out if the connector has information about the document by using a web browser to access the file information on the connector host.

If the document isn't located, the administrator can request a recrawl of the missing document by restarting the crawl from the Connector Dashboard, or recrawling the URL by using the **Content Sources > Web Crawl > Freshness Tuning** page in the Admin Console.

Troubleshooting quick reference

Error message/Issue	**Resolution**	**Type of Connector**
Logs: Unathorized request. Status code:200	Add host IP to the GSA's feeds' list of trusted IP addresses by using the **Content Sources > Feeds** page in the Admin Console.	SharePoint, SharePoint User Profiles, File Systems
Index diagnostics: Error: Permanent DNS failure.	Add a DNS override by using the **Administration > DNS Override** page in the Admin Console.	SharePoint
Index diagnostics: Retrying URL: Connection reset by peer during fetch.	DNS override is wrong. Correct it by using the **Administration > DNS Override** page in the Admin Console.	SharePoint
Errors in the logs for some	Host load is too high, try to	SharePoint

documents:The server sent HTTP status code 503: Service unavailable	reduce host load	
Feeds are not coming through	• Make sure GSA can accept feeds from the connector host machine. • Check connector logs for errors, such as failure to connect to look-up GSA, or failure to communicate with the repository.	SharePoint, SharePoint User Profiles, File Systems, Active Directory
Documents are not getting indexed	• Make sure GSA is set to follow and crawl the Connector's URLs by checking the **Content Sources > Web Crawl > Start and Block URLs** page in the Admin Console. • Make sure GSA crawling is not paused by using the **Content Sources > Diagnostics > Crawl Status** page. • Check for error messages on the **Index> Diagnostics > Index Diagnostics** page. • Take a look at connector's log messages. • Check the **Content Sources > Diagnostics > Real-time Diagnostics** page for the particular URL that you expect to be indexed.	SharePoint, SharePoint User Profiles, File Systems

Crawling is slow	Use the Dashboard to find: ● What is the mean duration of a request (Response Time)? A couple hundred milliseconds would be good. ● What is the max duration of a request? A file taking over a couple of minutes would be bad.	SharePoint, File Systems
Document retrieval times out	The connector gives a document retrieval request 30 seconds to start and 3 minutes to complete. If you want to give your repository more time you can adjust `adaptor.docContentTimeoutSecs` and `adaptor.docHeaderTimeoutSecs`.	SharePoint, File Systems
The Google Search Appliance Index Diagnostics shows many documents with a Crawl Status of "Document not found (404)."	Files and folders that are marked as hidden are not fed to the GSA. However, they may be listed on the **Index> Diagnostics > Index Diagnostics** page with a crawl status of "Document not found (404)."	SharePoint, File Systems
SharePoint is returning 401 (unatuhorized)	Ensure that the full read permissions are given on the SharePoint Web Application policy.	SharePoint
Renamed user names are not reflected in ACLs.	Run User Profile Synchronization job for incremental updates.	SharePoint

8 Common configuration options

The following table lists common configuration options, which are used by all connectors. If the administrator doesn't set these options, defaults are used. The only required option is `gsa.hostname`. All others are optional.

Name	Meaning	Default
`gsa.acceptsDocControlsHeader`	Use X-Gsa-Doc-Controls HTTP header with namespaced ACLs. Otherwise ACLs are sent without namespace and as metadata. If not set, then an attempt to compute from gsa.version is made.	true
`adaptor.fullListingSchedule`	When to invoke Adaptor.getDocIds, in cron format (minute, hour, day of month, month, day of week).	0 3 * * *
`adaptor.incrementalPollPeriodSecs`	Number of seconds between invocations of PollingIncrementalLister.getModifiedDocIds.	900
`adaptor.docContentTimeoutSecs`	Number of seconds a connector has to complete sending content before it is interrupted. Timing starts when sending content starts.	180
`adaptor.docHeaderTimeoutSecs`	Number of seconds connector has to start sending content before it is interrupted.	30
`adaptor.markAllDocsAsPublic`	When the value is "true," all	false

	documents are marked as “public.” Take note that if this option is set to “true” for the Connector for Active Directory, the connector does not send users/groups to specify group memberships.	
`adaptor.pushDocIdsOnStartup`	Whether to invoke Adaptor.getDocIds on process start (in addition to `adaptor.fullListingSchedule`).	true
`docId.isUrl`	If your connector document ids are already URLs, prevent them from being inserted into connector generated URLs.	false
`feed.archiveDirectory`	Save feeds of listed document-ids, named resources, and group definitions to the specified directory on the local drive as they are sent to the GSA. All feeds successfully and unsuccessfully sent to the GSA are archived. Failed feeds are tagged with FAILED in the archive feed file name.	
`feed.crawlImmediatelyBitEnabled`	Send bit telling GSA to crawl immediately.	false
`feed.maxUrls`	Set the maximum number of URLs included per feed file.	5000
`feed.name`	Source name used in feeds. Generated if not provided.	

`feed.noRecrawlBitEnabled`	Send bit telling the GSA to crawl your documents only once.	false
`gsa.version`	Version number used to configure expected GSA features.	Defaults to acquiring from GSA. Uses 7.0.14-114 if acquiring fails.
`gsa.characterEncoding`	Character set used in feed files.	UTF-8
`gsa.hostname`	Machine to send feed files to. Process errors if not provided.	
`gsa.samlEntityId`	The SAML Entity ID that identifies the GSA.	http://google.com/enterprise/gsa/security-manager
`journal.reducedMem`	Avoid tracking per URL information in RAM; suggested with over five hundred thousand documents.	true
`gsa.scoringType`	Type of relevance algorithm GSA utilizes to rank documents. Either content or web. Is sent when adaptor.sendDocControlHeader is true.	content
`server.dashboardPort`	Port on a connector 's machine for accessing a connector's dashboard. Every instance of a connector running on a machine must have a unique value for `server.dashboardPort`.	5679

`server.docIdPath`	Part of URL preceding encoded document ids.	/doc/
`server.fullAccessHosts`	Hosts allowed access without authentication (certificates still needed when in secure mode).	empty, but implicitly contains gsa.hostname
`server.hostname`	Hostname of a connector machine for URL generation. The GSA will use this hostname to crawl the connector.	lowercase of automatically detected hostname
`server.keyAlias`	Keystore alias where encryption (public and private) keys are stored.	connector
`server.maxWorkerThreads`	Number of maximum simultaneous retrievals allowed.	16
`server.port`	Retriever port. Every instance of a connector running on a machine must have a unique value for `server.port`.	5678
`server.queueCapacity`	Maximum retriever queue size.	160
`server.reverseProxyPort`	Port used in retriever URLs (in case requests are routed through a reverse proxy).	`server.port`
`server.reverseProxyProtocol`	Can be either http or https, depending on proxy traffic.	http in secure mode or http otherwise
`server.samlEntityId`	The SAML Entity ID that the connector uses to identity itself.	http://google.com/enterprise/gsa/adaptor
`server.secure`	Enables https and certificate	false

	checking.	
`server.useCompression`	Compress retrieval responses.	true
`transform.acl.X`	Where X is an integer, match and modify principals as described.	no modifications
`transform.pipeline`	Sequence of transformation steps.	empty string (no pipeline)