Google Search Appliance Connectors

Deploying the Connector for SharePoint

Google Search Appliance Connector for SharePoint software version 4.0.3
Google Search Appliance software version 7.2

October 2014

Table of Contents

About this Guide

This guide is intended for anyone who needs to deploy the Google Search Appliance Connector 4.0.3 for SharePoint. The guide assumes that you are familiar with Windows or Linux operating systems and configuring the Google Search Appliance by using the Admin Console.

See the Google Search Appliance Connectors Administration Guide 4.0.3 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

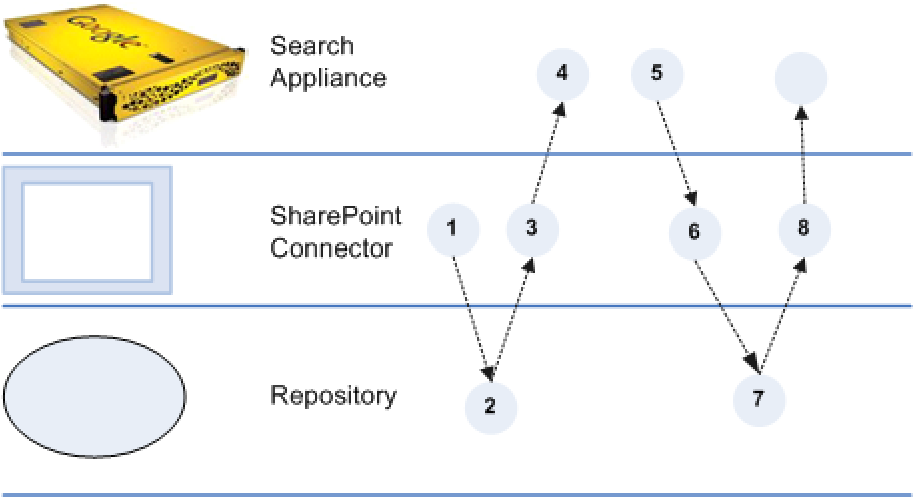
For information about using the Admin Console, see the Google Search Appliance Help Center.

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

Overview of the GSA Connector for SharePoint

The Connector for SharePoint 4.0 enables the Google Search Appliance to crawl and index content from Microsoft SharePoint. Each connector instance can support only one SharePoint Web Application. If you have more than one Web Application, you need to deploy one connector instance for each one.

The following diagram provides an overview of how the search appliance gets content from SharePoint through the connector. For explanations of the numbers in the process, see the steps following the diagram.



1. The Connector for SharePoint starts communicating with the repository by presenting authentication credentials.
2. The repository sends a limited number of Doc IDs of documents in the repository to the connector.
3. The connector constructs URLs from the Doc IDs and pushes it to the search appliance in a metadata-and-URL feed. Take note that this feed does not include the document contents.
4. The search appliance gets the URLs to crawl from the feed.
5. 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.
6. The connector requests the content from the repository.
7. The repository sends the content to the connector.

8. The connector pushes the content to the search appliance for indexing in a content feed. If the content is in HTML format, the search appliance will follow links within the page and send more GET requests for the linked content to the connector.

Automatic updates every 15 minutes

After the initial process completes, the connector periodically informs the search appliance of new documents and deltas, according to the value set in the connector configuration option `adaptor.incrementalPollPeriodSecs`. The default interval value is 15 minutes, but you can configure it to suit your needs. For more information, see “Common configuration options” in the Administration Guide.

Supported SharePoint versions

The Connector for SharePoint 4.0 supports the following versions:

- SharePoint 2010
- SharePoint Foundation 2010
- SharePoint 2013
- SharePoint Foundation 2013

Supported operating systems for the connector

The Connector for SharePoint 4.0 must be installed on one of the following supported operating systems:

- Windows Server 2012
- Windows Server 2008 (32 and 64 bit)
- Windows Server 2003 (32 and 64 bit)
- Ubuntu
- Red Hat Enterprise Linux 5.0
- SUSE Enterprise Linux 10 (64 bit)

Supported authentication mechanisms

The Connector for SharePoint 4.0 supports the following authentication mechanisms:

- Kerberos
- NTLM V1

  Note: NTLM V2 is not supported.
- Forms authentication
- HTTP Basic authentication

Known connector limitations

- Only one connector instance is allowed per Virtual Server / SharePoint Web Application.
- The connector requires full read permissions at the Web Application Policy level. Because this is not feasible when multiple tenants share the same web application, it might not be possible for the connector to support a multi-tenant configuration.
- The number of content databases will affect document change detection latency.
- The number of unique users and groups used in ACLs for each site collection will affect memory consumption.

Multi-Tenant configurations

The connector requires full read permissions at the Web Application Policy level. Because this is not feasible when multiple tenants share the same web application, it might not be possible for the connector to support a multi-tenant configuration. Multi-Tenant configurations will be supported in Connectors 4.0.4.

Before you deploy the Connector for SharePoint

Before you deploy the Connector for SharePoint, 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 Google for Work Support Portal (password required).
- Java JRE 1.7u6 or higher installed on the Windows computer that runs the connector
- Java JRE 1.6u27 installed on the Linux computer that runs the connector
- Connector for SharePoint 4.0.3 JAR executable
  For information about finding the JAR executable, see Step 2 Install the Connector for SharePoint.
- User account for the connector, with Full Read permissions to SharePoint Web Application in the User Policy
- If running the connector on Linux, the user account used for running the connector should belong to same domain as the SharePoint server. A user from a child domain or from the same forest is not sufficient.
- If there are any write-locked site collections, run the PrepareWriteLockedSitesForAdaptor.ps1 script on SharePoint using an account that has Admin privileges before installing the connector.

Optionally, configure the search appliance for the authentication method in use (typically LDAP for Active Directory). For detailed information about configuring authentication, see Managing Search for Controlled Access-Content.

Deploy the Connector for SharePoint

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

To deploy the Connector for SharePoint, perform the following tasks:

1. Configure the search appliance
2. Install the Connector for SharePoint
3. Optionally, configure adaptor-config.properties variables
4. Run the Connector for SharePoint

Step 1 Configure the search appliance

For the search appliance to work with the Connector for SharePoint, the search appliance needs to be able to crawl SharePoint content and accept feeds from the connector. To set up these capabilities, perform the following tasks by using the search appliance Admin Console:

1. Add the URL provided by the connector to the search appliance's crawl configuration follow patterns.
2. Add the IP address of the computer that hosts the connector to the list of Trusted IP addresses so that the search appliance will accept feeds from this address.
3. Set up connector security.

Add the URL

To add the URLs provided by the connector to the search appliance's crawl configuration follow patterns:

1. In the search appliance Admin Console, click **Content Sources > Web Crawl > Start and Block URLs**.
2. 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.
3. Click **Save**.

Add the IP address

To 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**.

Set up security

For information about setting up security, see "Enable connector security" in the Administration Guide.

Step 2 Install the Connector for SharePoint

This section describes the installation process for the Google Search Appliance Connector for SharePoint 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 SharePoint on any host running one of the supported operating systems, however, the host must be in the same domain as the SharePoint installation.

As part of the installation procedure, you need to edit some configuration variables in the configuration file. Take note that you can encrypt the value for `sharepoint.password` before adding it to the file by using the Connector Dashboard, as described in "Encode sensitive values" in the Administration Guide.

To install the Connector for SharePoint:

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.
3. Visit the connector 4.0.3 software downloads page at http://googlegsa.github.io/adaptor/index.html.
4. Download the exe file to the host by clicking on **MicrosoftSharePoint** in the Windows Installer table.
   You are prompted to save the single binary file, `sp-install-4.0.3.exe`.

5. Start installing the file by double clicking `sp-install-4.0.3`.
6. On the **Introduction** page, click **Next**.
7. On the **GSA Hostname** page, enter the hostname or IP address of the GSA that will use the connector and 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 **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**.
14. In the folder where you installed the connector, edit `adaptor-config.properties` by providing the following configuration (replacing bolded items with your real configuration) within the file:

    ```
    gsa.hostname=yourgsa.example.com or IP address
    sharepoint.server=http://yoursharepoint.example.com/
    ```

    where **`yourgsa.example.com`** is a fully-qualified domain name. If it is not a fully-qualified domain name, then you must set DNS override on the connector host.
    Linux: Add these additional configuration options to `adaptor-config.properties`:

    ```
    sharepoint.username=YOURDOMAIN\\ConnectorUser
    sharepoint.password=user_password
    ```

    NTLM and **Kerberos:** (**Windows**): When SharePoint and the current user domain is the same or from the same domain hierarchy, Windows operating systems automatically use the credentials of the person currently signed on to Windows.

    If not, you need to specify a username and password. For example, suppose that you run the connector from a coprorate domain such as @mycompany.com against a SharePoint instance from another domain, such as GSA-CONNECTORS. In this case, you need to specify user credentials for the GSA-CONNECTORS domain.

Forms Authentication (**Linux** and **Windows**): Always specify the username and password.

Note: See Step 3 for optional variables that you can also configure for the connector.

15. In the same folder, review, and if needed, edit `logging.properties`.
    For more information, See “Configure Connector Logs” in the Administration Guide.
16. In the same folder, run the `run.bat` file.

If SharePoint is configured to use HTTPS, get a SharePoint certificate to add it as a trusted host for the connector by performing the following steps:

1. Navigate to SharePoint in a browser.
   A warning page appears with a message such as "This Connection is Untrusted." This message appears because the certificate is self-signed and not signed by a trusted Certificate Authority. Click, "I Understand the Risks" and "Add Exception."
2. Wait until the "View..." button is clickable, then click it.
3. Change to the "Details" tab and click "Export...".
4. Save the certificate in your connector's directory with the name "`sharepoint.crt`".
5. Click Close and Cancel to close the windows.
6. To allow the connector to trust SharePoint, enter the following command:

```
keytool -importcert -keystore cacerts.jks -storepass changeit -
file sharepoint.crt -alias sharepoint
```

17. When prompted Trust this certificate?, answer yes.

Step 3 Configure adaptor-config.properties variables

Optionally, you can edit or add additional configuration variables to the `adaptor-config.properties` file. The following table lists the most important variables that pertain to the Connector for SharePoint, as well as their default values. See also "Common configuration options" in the the Administration Guide.

Variables	Description	Default
`server.dashboardPort`	Port on which to view web page showing information and diagnostics.	5679
`adaptor.namespace`	Namespace used for ACLs sent to GSA	Default
`sharepoint.xmlValidation`	Whether to enable strict checking of XML responses using the expected schema.	False
`sharepoint.maxIndexableSize`	Number of bytes of a document that GSA indexes.	2097152

Step 4 Run the Connector for SharePoint

After you install the Connector for SharePoint, you can run it by using `cmd.exe` on the host machine:

```
java -Djava.util.logging.config.file=logging.properties -jar adaptor-
sharepoint-4.0.3-withlib.jar
```

To Verify that the connector has started and is running, navigate 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.

Uninstall the Google Search Appliance Connector for SharePoint

To uninstall the Connector for SharePoint:

1. Click the **Change GSA_SP_Adaptor Installation** icon on your desktop.
   The **Uninstall GSA_SP_Adaptor** page appears.
2. Click **Next**.
3. On the **Uninstall Options** page, select an option:
   a. **Complete Uninstall.** Google recommends selecting **Complete Uninstall**.
   b. **Uninstall Specific Features.** If you click **Uninstall Specific Features**, select **Application**.
4. Click **Uninstall**.
   Files are uninstalled.
5. On the **Uninstall Complete** page, select **No, I will restart my system myself.**
6. Click **Done**.

Troubleshoot the Connector for SharePoint

For information about troubleshooting the Connector for SharePoint, see “Troubleshoot Connectors,” in the Administration Guide.