

Coveo Enterprise Search 6.0

SharePoint Connector

Coveo's *SharePoint* connector allows users to index the information stored on a *SharePoint* server. The connector can be configured to index a complete *SharePoint* farm or a specific list of a *SharePoint* Web sites.

Features

The *SharePoint* connector can be used to index content from one or multiple *SharePoint* servers. It allows *Coveo Enterprise Search* (CES) to crawl and index the content of *SharePoint* farms, Web applications, site collections, Web sites, lists, list items and documents.

The following details the features available in the *SharePoint* connector:

- **Content indexing**
 - Indexing all *SharePoint* content:
 - Farms and Web applications;
 - Site collections and Web sites;
 - Portals, areas and area listings;
 - Lists, list items and list item attachments;
 - Document libraries and documents;
 - User profiles and personal Web sites.
 - **Automatic site collection discovery:** It is possible to configure the CES *SharePoint* connector to index all site collections located on a *SharePoint* farm without having to list them manually;
 - **HTTP over SSL (HTTPS) support:** The *SharePoint* connector can be use to index a *SharePoint* site that uses HTTPS.
- **Security:** The *SharePoint* security model is fully supported, only the results the user has access to are returned when aquery is performed in the CES search interface.
- **Live indexing:** Once live indexing is enabled on a CES *SharePoint* source, the *SharePoint* connector automatically refreshes the content modified since the last live indexing run. This way, no more source rebuilds or refreshes are required to keep the index up to date.

Supported Versions

The *SharePoint* connector supports these versions:

- **SharePoint 2003:**
 - *Windows SharePoint Services 2.0* (WSS 2)
 - *Microsoft Office SharePoint Portal Server 2003* (SPS 2003)
- **SharePoint 2007:**
 - *Windows SharePoint Services 3.0* (WSS 3)
 - *Microsoft Office SharePoint Server 2007* (MOSS 2007)

Requirements

- **Administrative account:** The crawling user must have the proper rights to retrieve the information from the *SharePoint* server. For more information, refer to [How to Setup an Administrative Account](#).
- **CES SharePoint web services (optional):** If you need to index a SharePoint 2007 server (WSS 3 or MOSS 2007), or if you want to be able to use the Live indexing feature, the CES SharePoint web services must be installed on the SharePoint server. For more information, refer to [How to Install the CES SharePoint Web Services \(optional\)](#).

Installation

▶ How to Setup an Administrative Account

SharePoint 2003 (WSS 2 and SPS 2003)

The following table displays the minimal required permission the CES administrative account must have to perform the specified action:

Actions to Perform	Minimal Required Permissions
Content indexing	Web Site Contributor
Security indexing	Web Site Administrator
Personal sites and user profiles indexing, live indexing and site collection discovery	Member of the <i>SharePoint</i> Administrator Group or of the <i>SharePoint</i> server local Administrator group.

Perform the following procedure in order to add the **Web Site Contributor** or the **Web Site Administrator** permission:

1. Access the *SharePoint* Web site of the CES source starting address.

Microsoft Office
SharePoint Portal Server 2003

My Site Site Settings Help

Home Topics News Sites

Intranet
Add Users: Intranet

Use this page to add users to this site, list, or document library.

Step 1: Choose Users
You can enter e-mail addresses, user names (e.g., DOMAIN\name), or cross-site group names. Separate them with semicolons.

Users: companyname\jsmith

[Select users and groups](#)

Step 2: Choose Site Groups
Choose the site groups you want these users to have.

Site groups:

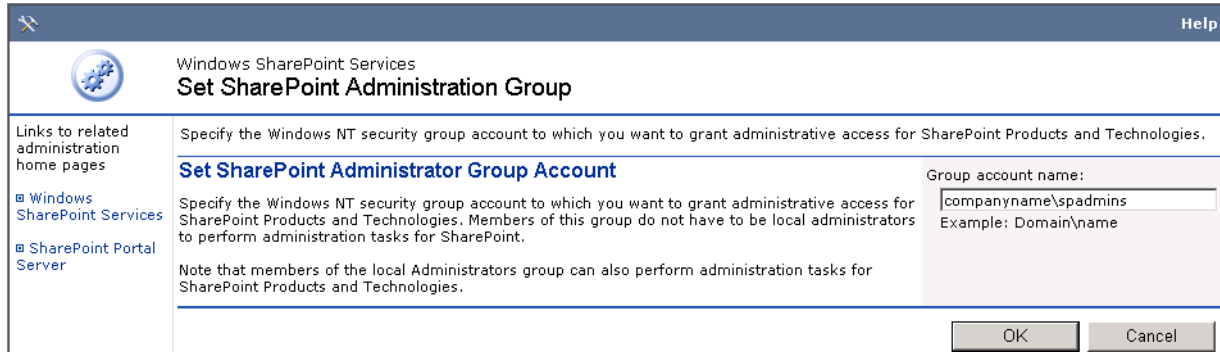
- Reader - Has read-only access to the Web site.
- Contributor - Can add content to existing document libraries and lists.
- Web Designer - Can create lists and document libraries and customize pages in the Web site.
- Administrator - Has full control of the Web site.
- Content Manager - Can create and manage areas, lists, libraries, and sites.
- Member - Can view and personalize portal site content and create sites.

Next >

2. Click **Site Settings**.
3. Click **Manage Users** in the **General Settings** section.
4. Click the **Add Users** button.
5. Add the CES administrative account in the **Users** text box, select the **Contributor** or the **Administrator** site group and click **Next >**.

Perform the following procedure in order to add the CES administrative account to the **SharePoint Administrator Group**:

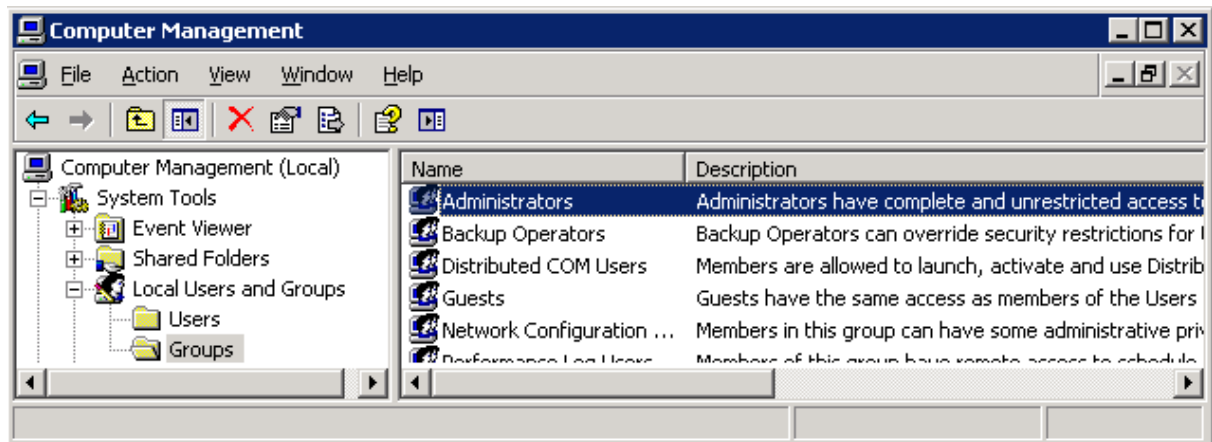
1. Access **SharePoint Central Administration** (*Windows Start* menu > All Programs > Administrative Tools).



2. Click **Set SharePoint administration group** in the **Security Configuration** section. The *SharePoint* administrator domain group is displayed in the **Group account name** text box.
3. Ask the domain administrator to add the CES administrative account to the specified *SharePoint* administrator domain group.

Perform the following procedure in order to add the CES administrative account to the **SharePoint server local Administrator group**:

1. Access the **Computer Management** console (*Windows Start* menu > All Programs > Administrative Tools).
2. Expand **System Tools > Local Users and Groups** and click **Groups**.



3. Right-click **Administrators** and then click **Add to Group**.
4. Click **Add**.
5. In the **Select Users, Computers, or Groups** window, enter the CES administrative account and click **OK**.
6. Click **OK** to close the Administrators properties sheet.

SharePoint 2007 (WSS 3 and MOSS 2007)

To extract the content of a *SharePoint 2007* server, the CES administrative account must have a *Full Read* permission for each *SharePoint* Web application to crawl. This permission is required because the *SharePoint* connector uses the *SharePoint* Web services to access the *SharePoint* content, and these Web services require the calling user to have the Web application *Full Read* permission to successfully process the request. Note that the *SharePoint* built-in indexing service also has this requirement. For more information, refer to the **Windows SharePoint Services Search accounts** sections of the Web pages:

- <http://technet.microsoft.com/en-us/library/cc288210.aspx> (WSS 3)
- <http://technet.microsoft.com/en-us/library/cc179348.aspx> (MOSS 2007)

Furthermore, to crawl *SharePoint* user profiles and personal sites content (MOSS 2007), the CES administrative account must also have the **Read** permission for the CES source starting address' site collection and have the **Manage User Profiles Shared Service** right.

Users can either let the CES *SharePoint* connector setup these permissions automatically or setup them manually.

Automatic Setup

When a *SharePoint* source refresh or rebuild is executed, the *SharePoint* connector automatically setups the required permissions for the CES administrative account:

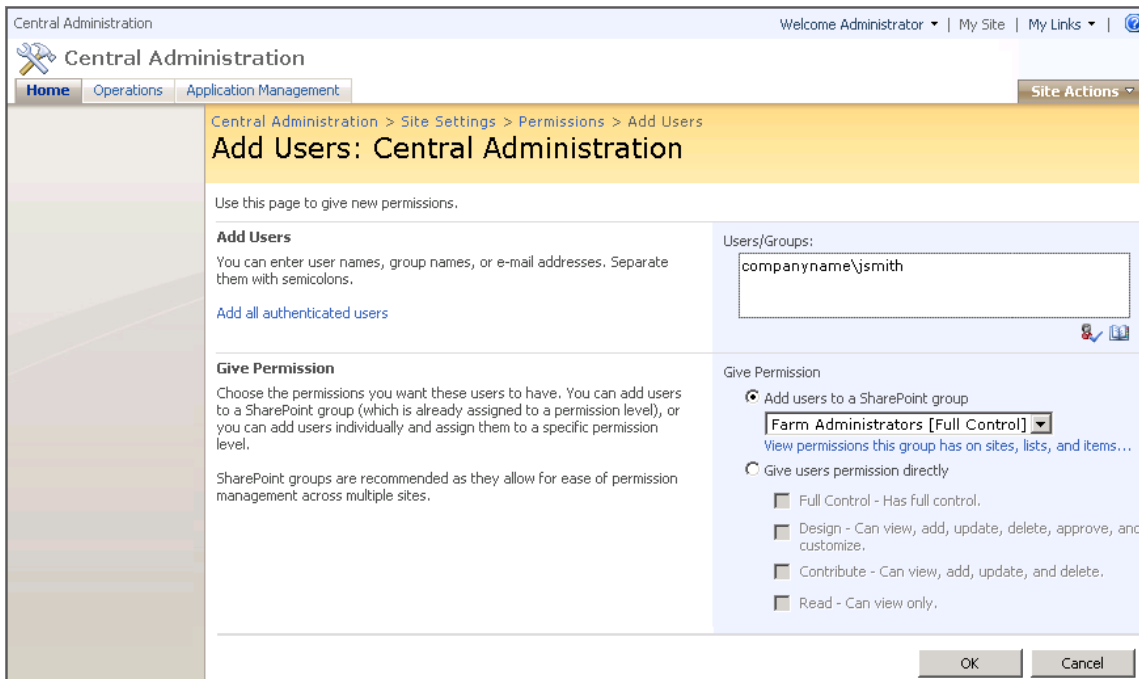
- The *Full Read* permission is automatically granted for the CES administrative account for all farm Web applications;
- The **Manage User Profiles Shared Service** right is automatically granted for the CES administrative account (MOSS 2007 only).

Requirements:

- The CES *SharePoint* Web services must be installed on the *SharePoint* server. For more information, refer to [How to Install the CES SharePoint Web Services \(optional\)](#);
- The CES administrative account must be a member of the *SharePoint* farm administrator's group;
- The CES administrative account must have the **Read** permission for the starting address' site collection.

Perform the following procedure in order to manually add the CES administrative account to the *SharePoint* farm administrator's group:

1. Access **SharePoint 3.0 Central Administration**.



2. Click **Operations**.

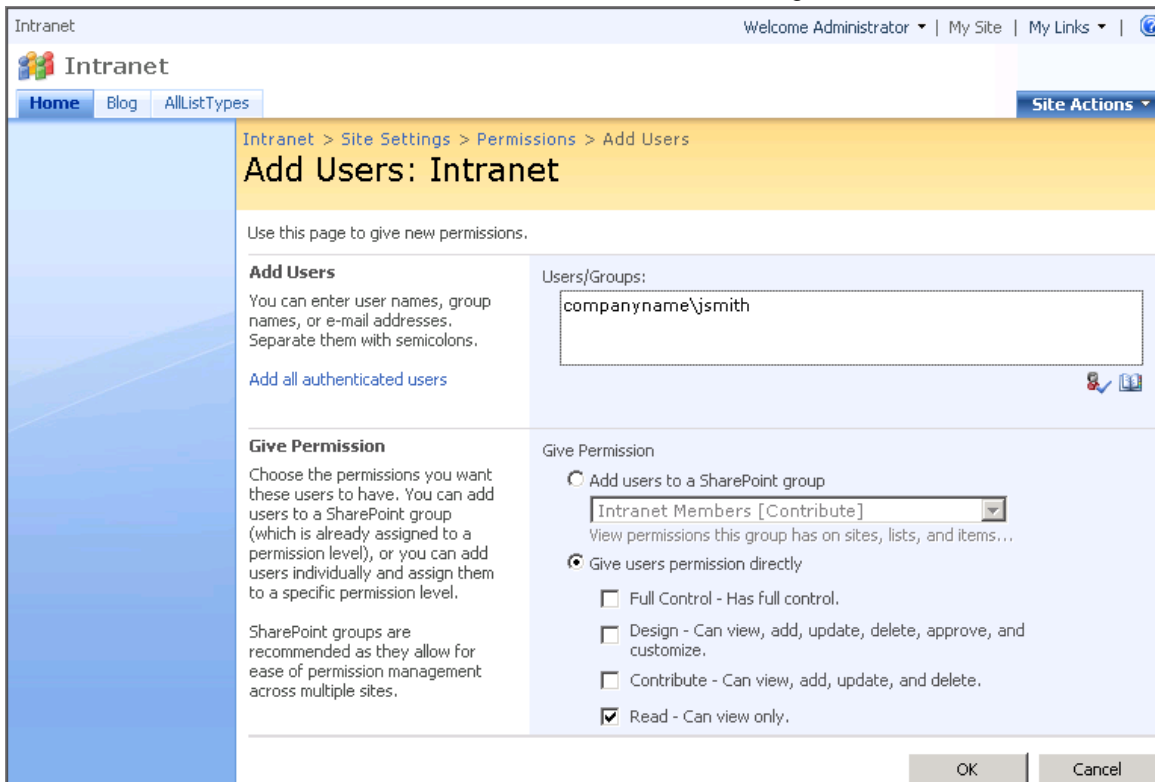
3. Click **Update farm administrator's group**.

4. Click **New**.

5. Add the CES administrative account, select **Farm Administrators [Full Control]** and click **OK**.

Perform the following procedure in order to manually add the **Read** permission:

1. Access the *SharePoint* site collection of the CES source starting address.

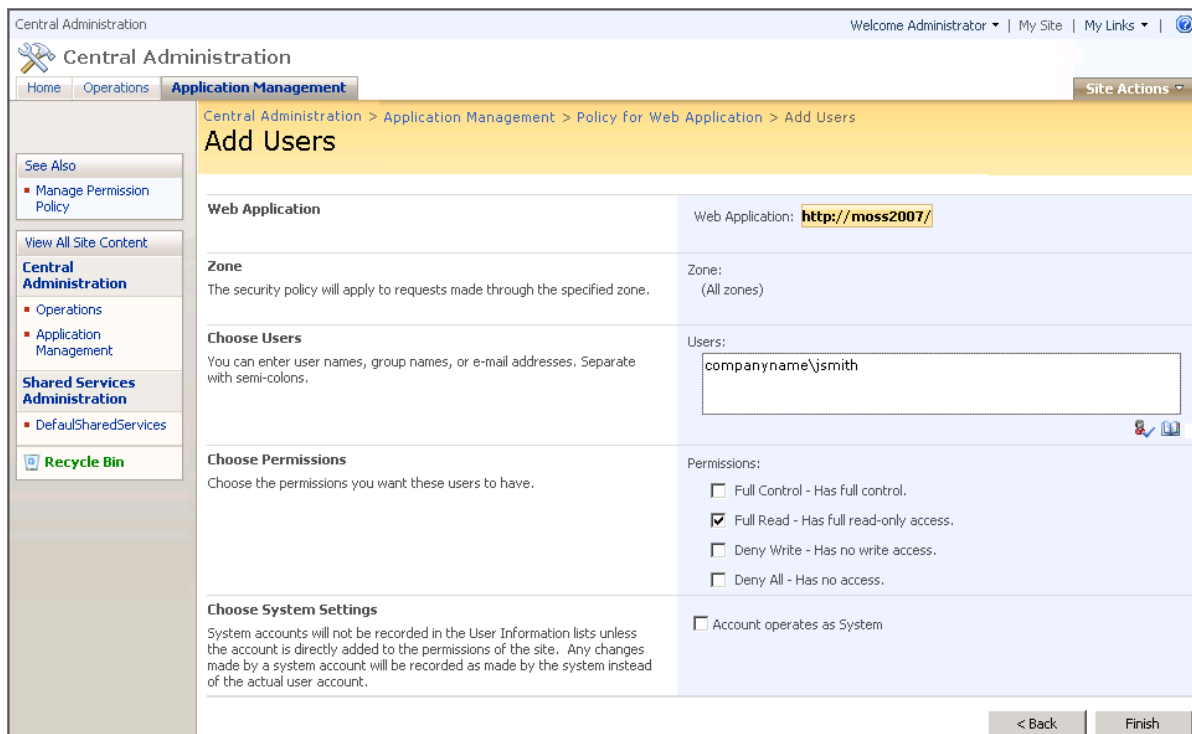


2. Click **Site Actions** and select **Site Settings**.
3. Click **Advanced Permissions** in the **Users and Permissions** column.
4. Click **New**.
5. Add the CES administrative account, select **Give users permission directly**, then the **Read** permission and click **OK**.

Manual Setup

Perform the following procedure in order to manually add the *Full Read* permission to the CES administrative account for all farm Web applications:

1. Access **SharePoint 3.0 Central Administration**.
2. Click **Application Management**.
3. Click **Policy for Web application**.



The screenshot shows the 'Add Users' page in SharePoint Central Administration. The breadcrumb trail is: Central Administration > Application Management > Policy for Web Application > Add Users. The page is titled 'Add Users' and contains several sections:

- Web Application:** Web Application: `http://moss2007/`
- Zone:** Zone: (All zones)
- Choose Users:** Users: `companyname\jsmith`
- Choose Permissions:** Permissions:
 - Full Control - Has full control.
 - Full Read - Has full read-only access.
 - Deny Write - Has no write access.
 - Deny All - Has no access.
- Choose System Settings:** Account operates as System

At the bottom right, there are '< Back' and 'Finish' buttons.

4. For every Web application to crawl:
 - Click **Add Users**.
 - Select **(All Zones)** and click **Next >**.
 - Add the CES administrative account, select the **Full Read** permission and click **Finish**.

Perform the following procedure in order to manually add the **Read** permission:

1. Access the *SharePoint* site collection of the CES source starting address.

Intranet > Site Settings > Permissions > Add Users

Add Users: Intranet

Use this page to give new permissions.

Add Users
You can enter user names, group names, or e-mail addresses. Separate them with semicolons.

[Add all authenticated users](#)

Give Permission
Choose the permissions you want these users to have. You can add users to a SharePoint group (which is already assigned to a permission level), or you can add users individually and assign them to a specific permission level.

SharePoint groups are recommended as they allow for ease of permission management across multiple sites.

Users/Groups:
companyname\jsmith

Give Permission

Add users to a SharePoint group
Intranet Members [Contribute]
View permissions this group has on sites, lists, and items...

Give users permission directly

Full Control - Has full control.

Design - Can view, add, update, delete, approve, and customize.

Contribute - Can view, add, update, and delete.

Read - Can view only.

2. Click **Site Actions** and select **Site Settings**.
3. Click **Advanced Permissions** in the **Users and Permissions** column.
4. Click **New**.
5. Add the CES administrative account, select **Give users permission directly**, then the **Read** permission and click **OK**.

Perform the following procedure in order to manually add the **Manage User Profiles Shared Service** right (MOSS 2007 only):

1. Access **SharePoint 3.0 Central Administration**.

Shared Services Administration: DefaultSharedServices

Shared Services Administration: DefaultSharedServices

Shared Services Administration: DefaultSharedServices > Manage Permissions

Add Users/Groups: Shared Service Rights

Warning: This page is not encrypted for secure communication. User names, passwords and any other information will be sent in clear text. For more information, please contact your administrator.

Choose Users
You can enter person names, group names, or email addresses. Separate them with semicolons.

Users/Groups:
companyname\jsmith

Choose Permissions
Choose permissions to assign to these users/groups. If you add a user that already exists, that user's rights will be overwritten.

Create personal site

Use personal features

Manage user profiles

Manage audiences

Manage permissions

Manage usage analytics

2. Click **Shared Services Administration**.
3. Click the shared service link hosting the user profiles and personal sites data.
4. Click **Personalization service permissions**.
5. Click **Add Users/Groups**.
6. Add the CES administrative account, select the **Manage user profiles** permission and click **Save**.

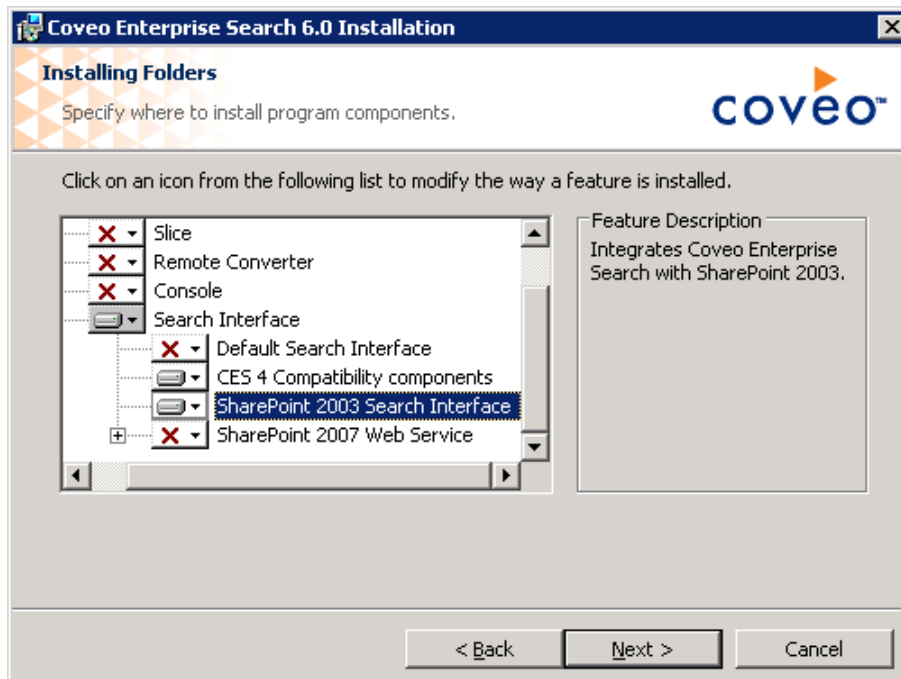
▶ **How to Install the CES SharePoint Web Services (optional)**

Installing the CES *SharePoint* Web services on the *SharePoint* server provides many advantages:

- **SharePoint 2003:**
 - Enables the *SharePoint* connector live indexing functionality;
 - Enables the indexing of Area Listing HTML content;
 - Enables the *SharePoint* site collections discovery.
- **SharePoint 2007:**
 - Enables the indexing of content security;
 - Enables the automatic CES administrative account setup;
 - Enables the *SharePoint* site collections discovery.

Perform the following procedure in order to install the CES *SharePoint* Web services on a *SharePoint* server:

1. Launch the CES installation kit on the *SharePoint* server.
2. In the **Select Setup Type** page, select **Custom**.
 - **For SharePoint 2003:**
 - a. Unselect all features except for **SharePoint 2003 Search Interface**.





- b. Click **Next >**.
- c. In the **Search interface server** field, enter the name of the CES server. It is possible to enter any server name if you do not have a search server or you do not want to use a CES search server in the *SharePoint* search interface.

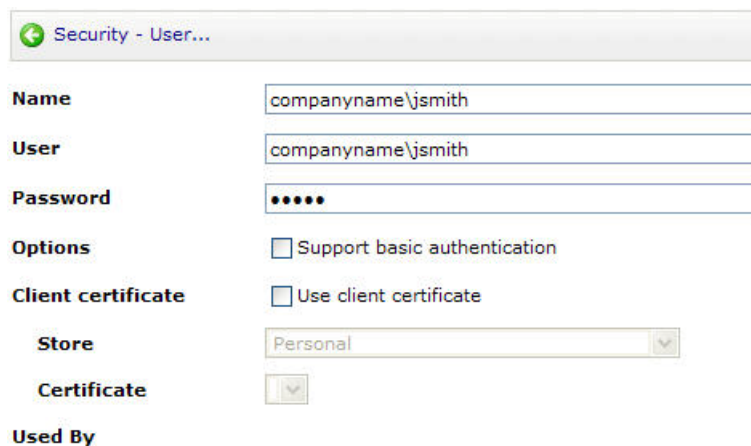
- d. In the **Integrate to the pages in the SharePoint 2003 database field**, select whether to add the **Coveo Search Bar** to all the *SharePoint* pages stored in the *SharePoint* database. It is possible to disable this functionality if you only want to install the CES *SharePoint* Web services.
- e. Click **Next >**.
- f. Click **Install**.
- g. Click **Finish**.
- **For SharePoint 2007:**
 - a. Unselect all features except for **SharePoint 2007 Web Service**.
 - b. Click **Install**.
 - c. Click **Finish**.

Configuration

▶ How to Configure a User Identity based on an Administrative Account (optional)

Once the connector's administrative account has been setup, configure a CES user identity if the administrative account is different than the CES service log on account:

1. In the Administration Tool, access the **Administrators** page (Configuration > Security).
2. In the left navigation pane, click **User Identities**. The **User Identities** page is displayed.
3. Click  **Add**. The **Modify User Identity** page is displayed.
4. In the **Name**, **User** and **Password** fields, enter the credentials of the administrative account.
5. If the SharePoint server is accessible through HTTP over SSL (HTTPS), select the **Use client certificate** checkbox and to select a certificate from the list.
6. Click  **Save**.



Security - User...

Name

User

Password

Options Support basic authentication

Client certificate Use client certificate


Store

Certificate

Used By

▶ How to Index a SharePoint Source


A *SharePoint* source can index *SharePoint 2003* content, *SharePoint 2007* content or both.

1. In the Administration Tool, access the **Sources and Collections** page (Index> Sources and Collections).
2. In the **Sources** section, click  **Add**. The **Add Source** page is displayed.
3. Enter the appropriate values of the selected *Microsoft Exchange* server.

Field	Description	Example
Name	Any descriptive name.	SharePoint Sites
Source Type	The connector used by this source.	SharePoint
Addresses	List of starting points for the connector, one address per line.	<p>To index the content of a <i>SharePoint</i> site and its subsites, enter the <i>SharePoint</i> site URL. For example:</p> <p>http://server/ http://server:8080/sites/Support/default.aspx https://server/</p> <p>To index the content of a <i>SharePoint</i> list, enter the <i>SharePoint</i> list URL. For example:</p> <p>http://server/Document Library/ http://server:8080/sites/Support/Lists/Contacts/AllItems.aspx</p>
Rating	Specifies whether documents from the source should receive a higher or lower ranking than average.	Normal
Document Types	Indicates which document type set is used to index the source.	Default
Fields	Indicates which field set is used to index the source.	Default Scheme
SharePoint	Indicates which <i>SharePoint</i> references set is used to index the source.	Default For more information on how to configure the way <i>SharePoint</i> documents are indexed, refer to What Are SharePoint Reference Sets .
Refresh Schedule	Indicates which refresh schedule is used to update the source content.	//*Every day (everyday at 12 A.M).
Index personal sites	Indexes all the personal sites linked to the <i>SharePoint</i> portal. Note that this option is selected by default.	
Index user profiles	Indexes all the <i>SharePoint</i> portal user profiles. Note that this option is selected by default.	

Index subsites	Indexes subsites recursively, starting with the main addresses and proceeding to deeper levels until every document has been indexed. Note that this option is selected by default.	If http://intranet/pages/ is the main Web site and Index subsites is selected, the http://intranet/pages/news/ and http://intranet/pages/news/2007/ Web sites are also indexed.
Index the document's metadata	Indexes metadata (e.g. modification date and author of documents). Note that this option is selected by default	
Generate a cached HTML version of indexed documents	Produces a <i>Quick View</i> of documents. Note that this option is selected by default.	
Generate a cached mobile HTML version of indexed documents	Produces a <i>Quick View</i> for the mobile devices of documents. Note that this option is selected by default.	
Open results with cached version	Displays the <i>Quick View</i> instead of the actual document when a title is clicked in the result list. The purpose of this option is to prevent accidental document modification. Note that Generate a cached HTML version of indexed documents must be selected in order for Open results with cached version to be available; moreover, it is not selected by default.	
SharePoint Site Discovery	Specifies which <i>SharePoint</i> site collections to index.	<p>Index only starting addresses: Indexes only the sites (and subsites) whose addresses are entered in the Addresses field.</p> <p>Index the starting addresses and all sites listed in the Portal Site Directory: Indexes the sites (and subsites) whose addresses are entered in the Addresses field as well as all sites, related to these addresses, listed in the portal site directories.</p> <p>Index the Web application of the starting addresses: Indexes all site collections of the Web applications (virtual directories) related to the starting addresses. Note that CES must be integrated to <i>SharePoint</i>; otherwise, the connector is unable to locate the site collections of the Web applications (virtual directories).</p>

Authentication	Specifies the crawling account to use.	<p>Index all the Web applications having the same host name: Indexes all site collections of all Web applications (virtual directories) related to the starting addresses, as long as the Web application has the same host as its related starting address. Note that CES must be integrated to <i>SharePoint</i>; otherwise, the connector is unable to locate all Web applications (virtual directories).</p> <p>To use the CES service log on account, select (none). To use a specific administrative account, select if from the list of click Add. For more information concerning the administrative accounts, refer to How to Configure a User Identity based on an Administrative Account (optional).</p>
----------------	--	--

4. Click  [Save and Start](#).

It is also possible to configure more crawling option in the source **Advanced** tab.

Field	Description	Example
Index hidden lists and their items	Specifies whether hidden lists should be indexed or not. <i>SharePoint</i> uses hidden (or internal) lists to store non content information, such as Web parts, workflows state, etc. These lists and their list items are not indexed by default.	
Index redundant issue list items	Specifies whether redundant issues should be indexed or not. When an issue is updated in a <i>SharePoint</i> issue list, <i>SharePoint</i> creates a new issue item, leaving the old one unchanged. These redundant old issues are not indexed by default.	
Index survey responses	Specifies whether survey responses should be indexed or not. When someone responds to a <i>SharePoint</i> survey, this response is stored as a list item into the survey list. These responses are not indexed by default.	

Index documents uploaded with WebDAV

Specifies whether orphan documents should be indexed or not. Using the WebDAV functionality, it is possible to store documents that do not belong to any *SharePoint* document libraries. These documents are indexed by default.

Disable cookies

Specifies whether the connector should use the cookies sent by the crawled site. The cookies are enabled by default.

Expand sites and lists before applying filters

Specifies whether the connector should retrieve the *SharePoint* Web sites and list children before applying the source filters. This option is required to index a specific part of a *SharePoint* Web site or list. Note that enabling this option slows down the connector as it retrieves all children of all Web sites and lists from *SharePoint* and rejects filtered document just before indexing them. This option is disabled by default

If <http://intranet/> is the main Web site, http://intranet/pages/* is the only source include filter and **Expand sites and lists before applying filters** is selected, <http://intranet/pages/> and <http://intranet/pages/news/> Web sites are indexed. If **Expand sites and lists before applying filters** is not selected, nothing will be indexed as <http://intranet/> will be filtered out.

Use the author extracted from the document instead of the SharePoint author

Sometimes, a document *SharePoint* stored in a *SharePoint* documents library has two different authors, the author of the document stored in the document properties and the author stored in the *SharePoint* database, which represents the person who has modified or added the document into *SharePoint*. In such case, CES uses the *SharePoint* author instead of the document author. Use this option to modify this behavior.

A side effect of selecting this option is that the CES *SharePoint* search interface cannot link the search result author to the *SharePoint* user, as the document author may not be a *SharePoint* user.

Enable Microsoft Office integration

Specifies whether *MS Office* documents should be opened with *MS Office* directly using *SharePoint ActiveX* when the search result is clicked. If this option is not selected, the document is opened normally.

Server Name Alias	Specifies a server name that overrides the one from which documents are downloaded in the index. This is useful to have query results point to another server as the one used for indexing.
Priority	Specifies the download process priority.
Disable advanced text layout analysis for PDF documents	For better content indexing, CES can analyze PDF documents in order to identify text columns and restore appropriate content ordering. This process can be disabled to save CPU resources while indexing.
Disable advanced duplicate document filtering	Disables the filtering of duplicate documents in order to save CPU resources and speed up indexing. The purpose of this advanced filtering is to display only one copy of each document in the result list. Note that the filtering of duplicate documents is enabled by default.
Conversion Timeout	Determines the number of minutes after which the converter proceeds to another document even if the conversion is not complete (the document for which the conversion has not been finished is considered corrupted). By default, the conversion timeout is of 10 minutes; however it is suggested to increase it to 30 minutes or even 60 minutes, when converting large audio video files.