



Coveo Platform 6.5

File Connector Guide

Notice

The content in this document represents the current view of Coveo as of the date of publication. Because Coveo continually responds to changing market conditions, information in this document is subject to change without notice. For the latest documentation, visit our website at www.coveo.com.

Copyright © 2011, Coveo Solutions Inc.

All rights reserved.

Coveo is a trademark of Coveo Solutions Inc. This document is protected by copyright and other intellectual property law and is subject to the confidentiality and other restrictions specified in the Coveo License Agreement.

Document part number: PM110627-EN Revision A

Table of Contents

1. File Connector Features	1
2. File Connector Deployment Overview	3
3. File Connector Requirements	4
4. Planning Your File Connector Collections and Sources	5
5. Setting up a File System Crawling Account	6
6. CES Configuration for the File Connector	7
6.1 Configuring a File Connector User Identity.....	7
6.2 Creating and Indexing a File Connector Source.....	8
6.3 Modifying Hidden File Connector Source Parameters.....	12
6.4 Showing Advanced Hidden Source Parameters.....	13
6.5 Toggling Live Indexing for a Source.....	15
7. Mail Archive Indexing with the File Connector	17
7.1 Mail Archive Indexing Deployment Overview.....	18
7.2 Installing the Microsoft MAPI Component for Mail Archive Indexing.....	18
7.3 Setting up a Document Type for Mail Archive Indexing.....	21
7.4 Creating a Mail Archive Mapping File.....	21
7.5 Selecting a 32-bit or 64-bit Process for a Connector.....	24
8. Troubleshooting File Connector Issues	26

1. File Connector Features

The Coveo connector for file shares allows you to index the content of files stored on local or network drives. The content of the files is integrated into the Coveo unified index, making the files easily searchable by end-users.

Features

The File connector features are:

Supported file shares

The File connector can index files on file shares of the following types:

- Microsoft Windows Server and Microsoft Windows (NT/2000/XP/Vista/7)
- Microsoft Windows Distributed File System (DFS)
- File share on other operating systems (ex.: UNIX, Linux, Mac) accessible through the Windows network.

Security

The permissions associated with a file in the Coveo unified index are the same as the ones found in the file system.

Identity impersonation

You can configure each connector source to impersonate a different identity allowing to index several repositories that require different access credentials.

Live indexing

The File connector uses file monitoring to identify modifications of indexed files. When this feature is enabled, CES processes modifications as soon as they are detected, thus keeping the unified index synchronized with the file system without requiring a source refresh.

Mail archives

You can optionally configure the File connector to open Microsoft Exchange Personal Folders (.pst) files and index the content of individual emails so that they become easily searchable by end-users. The File connector supports the Unicode and the legacy ANSI format of PST files (see ["Mail Archive Indexing with the File Connector" on page 17](#)).

Note: Microsoft Exchange Personal Folders (.pst) files are referred to as *mail archives* in the File connector documentation.

Note: The File connector is completely independent from the Desktop connector. While both connectors can crawl local and network drives, the File connector is configured by the Coveo administrator and the crawling process runs on the Coveo server. The Desktop connector is configured by end-users using the Desktop Integration Package (DIP) and the crawling process runs on their computers. Both connectors send content to the unified index on the Coveo server.

Table File connector feature history

Coveo product version	Features
CES 6.1.4019	Addition of three hidden source parameters (IgnoreUnresolvedDeniedSecurities , LiveMonitoringEventsQueueMaxSize , and TempFileRegex) and deprecation of one (Ignore temporary files).
CES 6.0	Connector introduction - Replacement of the File (Legacy) connector

2. File Connector Deployment Overview

The following procedure outlines the steps needed to bring content from file shares into the Coveo unified index using the File connector. The steps indicate the order in which you must perform configuration tasks.

1. Validate that your environment meets the requirements (see ["File Connector Requirements" on page 4](#)).
2. Determine how you will organize your File connector sources and collections within the Coveo unified index (see ["Planning Your File Connector Collections and Sources" on page 5](#)).

3. Select or create one or more necessary crawling accounts for the file share.

The File connector needs an account with which it can crawl the complete content (see ["Setting up a File System Crawling Account" on page 6](#)).

4. Optionally, configure the File connector to index Microsoft Exchange mail archives.

The File connector needs specific configuration to be able to open PST files and efficiently index their content (see ["Mail Archive Indexing with the File Connector" on page 17](#)).

Note: When Coveo runs on a 64-bit server and none of your File connector sources is configured to index Microsoft Exchange mail archives, change the default File connector process type from 32-bit to 64-bit to take advantage of the improved 64-bit performance (see ["Selecting a 32-bit or 64-bit Process for a Connector" on page 24](#)).

5. In the CES Administration Tool, for each planned source:

- a. Optionally, configure the user identity.

By default the File connector crawls the file share with the CES service identity. It is generally better to rather select or create a file share account with appropriate permissions to be used by the connector to crawl the file share (see ["Setting up a File System Crawling Account" on page 6](#)). You will then assign this crawling account to a user identity (see ["Configuring a File Connector User Identity" on page 7](#)), and assign the user identity to the source.

- b. Configure and index the File connector source.

The File connector must know details about the file shares to be able to index their content (see ["Creating and Indexing a File Connector Source" on page 8](#)).

6. In the Interface Editor, ensure that the collections containing the new File connector sources are included in the scope of the appropriate search interfaces.

7. Verify that the target content is available from the appropriate search interfaces.

8. Optionally, modify hidden source parameters

Once your File connector source is up and running, if you encounter specific issues, consider modifying some hidden source parameters to resolve the issues (see ["Troubleshooting File Connector Issues" on page 26](#) and ["Modifying Hidden File Connector Source Parameters" on page 12](#)).

3. File Connector Requirements

Your environment must meet the following requirements to be able to use the File connector:

- CES 6.0+

The connector was introduced with CES 6.0 to replace the File (Legacy) connector that is still available up to CES 6.5 for backward compatibility.

- CES license for the File Connector

Your CES license must include support for the File Connector to be able to use this connector.

- When indexing PST mail archives, Microsoft MAPI component on the Coveo server

The File connector needs the Microsoft MAPI component to open PST files (see "[Installing the Microsoft MAPI Component for Mail Archive Indexing](#)" on page 18).

4. Planning Your File Connector Collections and Sources

The content of the Coveo unified index is organized in collections and each collection contains one or more sources. Before starting to deploy the File connector, you should determine how to organize collections and sources for the content of your file shares.

Consider the following facts:

- End-users can see collections names in search interface elements, while sources are only visible by Coveo administrator in the Administration Tool.
- You can configure a search interface to include a **Collection** facet or collection checkboxes below the search box so that end-users can refine search results based on collections.
- Each search interface has a specific scope that is defined by including one or more collections in which to search.
- When you create a collection, you can set permissions on the collection by specifying users or groups allowed to search the content of the collection.
- Similarly, when you create a source, you can set permissions on the source by specifying users or groups allowed to search the content of the source.

Consider the following recommendations when planning collections:

- Separate your file share content in collections that are meaningful to end-users and that are useful to refine results.

Example: When you have network file servers for different locations in your organization, create a collection for each file server:

- New York file share
- San Francisco file share
- Houston file share

- When creating a collection, choose a name that is clear and meaningful to end-users.
- Consider creating separate collections for separate audiences when you define specific search interfaces for specific audiences.

Consider the following recommendations when planning sources:

- Create separate sources when you need different impersonators to fully crawl different file shares or file share sections.
- When you choose to index mail archive (.pst) files, create a source to exclusively crawl mail archive files and exclude mail archive files from the source that crawls all other file types within the same file share.
- Consider creating separate sources when you want to set different permissions to different sections of a file share.
- Avoid grouping local and remote servers on the same source to prevent delaying source refresh on all servers when one server stops responding.

Note: Use the Desktop connector when you want to index files (including mail archives) located on hard drives in end-user desktop and laptop computers.

5. Setting up a File System Crawling Account

The File connector needs to connect to the file system using an account that has read access to all the content that you wish to bring into the Coveo unified index.

By default, the File connector crawls the file share with the CES service identity. You can also select or create a file share account with appropriate permissions to be used by the connector to crawl the file share. This is typically done in Active Directory by creating an account that has full read permissions throughout the file shares to index. A best practice is to create a dedicated account for this purpose with a strong password that never changes.

In CES, you will assign this account to a user identity (see "[Configuring a File Connector User Identity](#)" on [page 7](#)), and you will assign the user identity to a source (see "[Creating and Indexing a File Connector Source](#)" on [page 8](#)).

Important: When indexing PST mail archive files, the crawling account must also have write permissions. When the mail archive files are stored in a given folder, you can set up the account so it only has write access to that folder; however, when the mail archive files are scattered through different locations, give the account write access to the entire repository being indexed.

When you want or need to use different accounts for various files shares or file shares sections, consider creating two or more sources and assign a different user identity to each source.

Example: You can index the complete content of a file share except mail archive files with one source using an account with full read permissions and use a second source pointing to the folders containing PST files and use an account with read and write permissions to index only the content of PST files.

6. CES Configuration for the File Connector

6.1 Configuring a File Connector User Identity

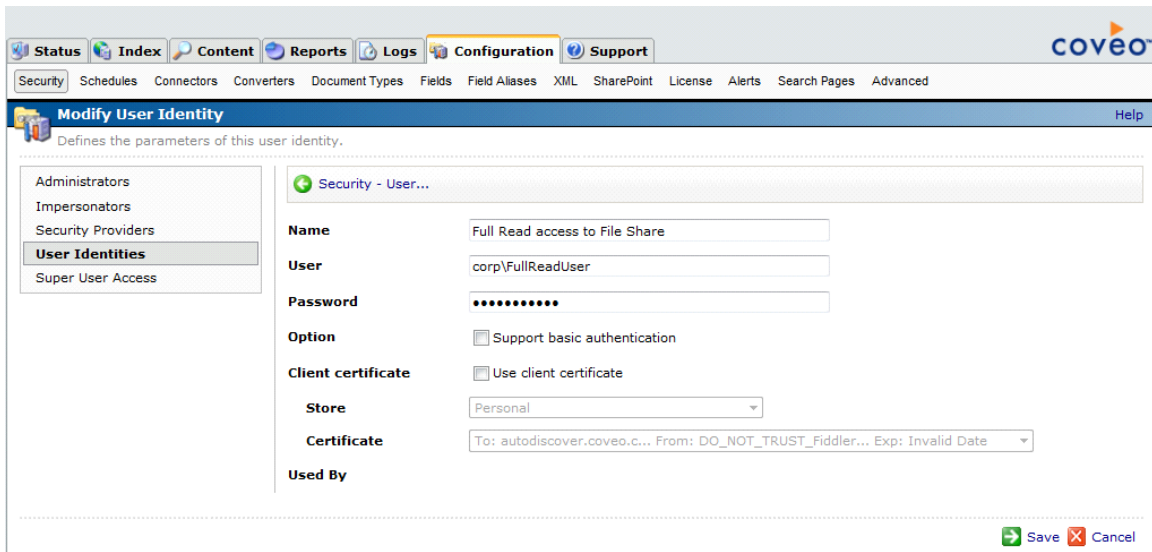
By default, the File connector crawls the file share with the CES service identity. When you rather choose to select or create a file share account with appropriate permissions to be used by the connector to crawl the file share (see "Setting up a File System Crawling Account" on page 6), you must assign this crawling account to a user identity.

To configure a File connector user identity

1. On the CES server, open the Administration Tool (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **Administration Tool**).
2. Select **Configuration** > **Security**.
3. In the **Security** page, in the left navigation pane, click **User Identities**.
4. In the **User Identities** page, click **Add**.
5. In the **Modify User Identity** page:
 - a. In the **Name** field, enter a name of your choice to describe the file share account that you selected or created to crawl the content that you want to index. This name appears only in the Coveo Administration Tool when you assign a user identity to a source.
 - b. In the **User** field, enter the username in the `domain\username` form for the file share account that you selected or created to crawl the content that you want to index.
 - c. In the **Password** field, enter the password for the file share account.

Note: The **Option**, **Client certificate**, **Store**, and **Certificate** parameters do not apply to the File connector.

- d. Click **Save**.



The screenshot displays the 'Modify User Identity' configuration interface. On the left, a navigation pane lists 'Administrators', 'Impersonators', 'Security Providers', 'User Identities' (selected), and 'Super User Access'. The main area is titled 'Security - User...' and contains the following fields:

- Name:** Full Read access to File Share
- User:** corp\FullReadUser
- Password:** Masked with 10 dots
- Option:** Support basic authentication
- Client certificate:** Use client certificate
- Store:** Personal
- Certificate:** To: autodiscover.coveo.c... From: DO_NOT_TRUST_Fiddler... Exp: Invalid Date
- Used By:** (Empty field)

At the bottom right, there are 'Save' and 'Cancel' buttons.

6.2 Creating and Indexing a File Connector Source

A source defines a set of configuration parameters for one or more file shares or file share sections.

Note: Create two or more sources when file shares or file share sections need different parameters sets. A source uses one or more starting addresses to determine locations to crawl and index.

To create and index a File connector source

1. On the CES server, open the Administration Tool (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **Administration Tool**).
2. Select **Index** > **Sources and Collections**.
3. In the **Collections** section:
 - a. Select an existing collection in which you wish to add the new source.
 - OR
 - b. Click **Add** to create a new collection.
4. In the **Sources** section, click **Add**.
5. In the **Add Source** page:
 - a. Enter the appropriate value for the following required parameters:

Name

A descriptive name of your choice for the source. This name is only visible to a Coveo administrator within the Administration Tool.

Example: Corporate network file share

Source Type

The connector used by this source. In this case, select **Files**.

Note: Do not select the deprecated **Files (Legacy)** connector.

Addresses

The list of starting address URIs indicating locations to index, one entry per line. You can specify the URIs as local or network paths. Addresses can represent a file system folder or file, a mail archive, or even a folder within a mail archive.

Examples:

Network folder:	<code>file://svr-fileshare/root</code>
Local folder:	<code>file:///c:/fileshare/root/</code>
Local file:	<code>file:///c:/fileshare/root/docs/work.doc</code>
Mail archive:	<code>file://svr-fileshare/emails/jsmith.pst</code>
Folder in a mail archive:	<code>file://svr-fileshare/emails/jsmith.pst/work</code>

Important: When you use paths containing drive letters as starting addresses (ex.: C:\fileshare), Windows XP SP2+ users will not be able to open the resulting links in the search result page. A better practice is therefore to rather index network file shares (ex.: \\Intranet\fileshare).

Authentication

If you chose to use a specific account to crawl the file system (see "Setting up a File System Crawling Account" on page 6), select the user identity that you created for this account (see "Configuring a File Connector User Identity" on page 7). Leave this parameter empty when you want the connector to crawl the file system using the CES service identity.

The screenshot shows the 'Collection: File Shares - Add Source' configuration page. The fields are as follows:

- Name:** File Share
- Source Type:** Files
- Addresses:** file://svr-fileshare/root/ (with examples: \\Intranet\Documents, C:\Documents)
- Rating:** Normal
- Document Types:** Default
- Fields:** Default Scheme
- Refresh Schedule:** Every day
- Mail archives configuration file:** C:\CES6\MailArchives.config
- Number of Live Indexing Threads:** 1
- Max Number of Retries:** 2
- Number of Refresh Threads:** 2
- Expand before filtering:**
- Expand mail archives:**
- Parameters:** Add Parameter
- Option:**
 - Index subfolders
 - Index the document's metadata
 - Document's addresses are case-sensitive
 - Generate a cached HTML version of indexed documents
 - Open results with cached version
- Authentication:** Full Read access to File Share

Buttons at the bottom: Save, Save and Start, Cancel

- b. Enter the appropriate value for the following parameters when you optionally want to index the content of mail archive files:

Mapping archives configuration file

When you decide to use a mail archive mapping file, enter the absolute full path pointing to your mapping file (see "Mail Archive Indexing with the File Connector" on page 17 and "Creating a Mail Archive Mapping File" on page 21).

Example:

```
C:\CES6\Config\Coveo.CES.CustomCrawlers.File.MailArchives.config
```

Expand mail archives

Select to index the content of mail archives (.pst). The default is false.

Document Types

Select the custom document type set that you created for mail archive sources.

- c. The default values for the following parameters generally do not need to be changed:

Rating

Change this value only when you want to globally change the ranking associated with all items in this source relative to the rating of other sources.

Example: If this source was for a legacy system, you may want to set this parameter to **Low**, so that in the search interface, results from this source appear later in the list compared to those from other sources.

Fields

If you defined custom Field sets, ensure to select the most appropriate for this source.

Refresh Schedule

Time interval at which the index is automatically refreshed to keep the index content up to date. By default, the **Every day** option instructs CES to refresh the source everyday at 12 AM.

Number of Live Indexing Threads

Determines the number of file system changes that the connector can process simultaneously. The default and recommended value is 1.

Max Number of Retries

Number of retries to perform when indexing fails for a file that is opened by another application. The default and recommended value is 2.

Number of Refresh Threads

Determines the number of files that the connector can refresh simultaneously. The default and recommended value is 2.

Expand before filtering

By default this option is not selected so that the crawler applies inclusion and exclusion filters on files but also on folders before crawling so that it only expands folders that you wish to index. In rare cases where an inclusion or exclusion filter should only be applied to files (ex. *.tif), you need to select this option so that the crawler fully expands folders to see all files and effectively applies the filters.

Note: Selecting this option can have a significant performance cost. The best practice is to use inclusion or exclusion filters to specify folders, not file types. Rather use document type sets to specify the file types to be indexed.

Parameters

Click **Add Parameter** when you want to show advanced hidden source parameters (see ["Modifying Hidden File Connector Source Parameters" on page 12](#)).

- d. The **Option** check boxes generally do not need to be changed:

Index Subfolders

Check to index all subfolders below the specified starting addresses.

Note: You can control more precisely specific folders or files to crawl using inclusion or exclusion filters.

Index the document's metadata

When you select this check box, any metadata value associated with the document becomes searchable using free text query. Leave the check box cleared to prevent seeing metadata values appear as separate items in the search results.

Document's addresses are case-sensitive

Leave the check box cleared. This parameter needs to be checked only in rare cases for case sensitive systems in which distinct documents may have the same file name but with different casing.

Generate a cached HTML version of indexed documents

When you select this check box (recommended), at indexing time CES creates HTML versions of indexed documents and saves them in the unified index. In the search interfaces, users can then more rapidly review the content by clicking the Quick View link to open the HTML version of the item rather than opening the original document with the original application.

When the source includes mail archives files, you must select this option to ensure users can view the content of mail archives items.

Consider clearing this check box only if you do not want to use Quick View links or to save resources when building the source.

Open results with cached version

Leave this check box cleared (recommended) so that in the search interfaces, the main search result link opens the original document with the original application. Consider selecting this check box only when you do not want users to be able to open the original document but only see the HTML version of the document as a Quick View. When this option is selected, you must also select the **Generate a cached HTML version of indexed documents** checkbox.

Note: When you index mail archive files, a custom document type set handles how mail archive items are opened from the search interfaces (see ["Setting up a Document Type for Mail Archive Indexing" on page 21](#)).

- e. Click **Save** to save changes to the source configuration.
-

Note: Before starting indexing the source content, consider using exclusive or inclusive filters to fine tune files that will be crawled.

OR

Click **Save and Start** to save the source configuration changes and immediately start indexing the source.

6. Validate that the source building process is executed without errors:

- In the left navigation pane, click **Status**, and then validate that the indexing proceeds without errors.

OR

- Start the CES Console to monitor the source building activities (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **CES Console**).

6.3 Modifying Hidden File Connector Source Parameters

The **Add Source** and **Source: ... General** pages of the Administration Tool present the parameters with which you can configure the connector for most file share setups. More advanced and more rarely used parameters are available but hidden. You can choose to make one or more of these parameters appear in the **Add Source** and **Source: ... General** pages of the Administration Tool so that you can change their default value. Consider changing values of hidden parameters only when you encounter time out error messages or performance issues.

The following list describes the available advanced hidden parameters for File connector sources. The parameter type (integer, string,...) appears between parentheses following the parameter name.

EnableCrawlDFSReferralLink (Boolean)

Set to True to enable crawling of Distributed File System (DFS) referral links. This option is useful when Windows perceives the crawling by the connector as a Denial-of-Service attack (see "[Access denied when crawling through a Distributed File System \(DFS\)](#)" on page 26). The default value is False.

IgnoreUnresolvedDeniedSecurities (Boolean)

Set to True to ignore unresolved denied securities. This option is useful to voluntarily ignore unresolved denied security errors. The default value is False. This option is only available for CES 6.5.4019+.

Example: When a user or group no longer exists, accessing their documents is denied and causes unresolved security exceptions with a message like: Unexpected error occurred while retrieving content from directory. - Access to the path [path] is denied.

Important: Enable this parameter with caution as it can create a security hole.

LiveMonitoringEventsQueueMaxSize (integer)

Maximum number of items to store for each starting addresses being monitored before discarding them. Discarded items will be indexed the next time the source is refreshed. This parameter is useful when the file path being monitored is being modified frequently, to prevent queuing a large number of modification events that would take a large amount of RAM on the server. The default value is 100000. This option is only available for CES 6.5.4019+.

RetryDelay (integer)

Delay (in seconds) before retrying to process a document that failed to be indexed. The default value is 30. Consider increasing the value when you think that this can increase chances for the file to be available for crawling.

RetryQueueMaxSize (integer)

Maximum number of items to store in the retry queue before discarding them. The default value is 100. Consider increasing the value when you experience frequent sharing violation when crawling and want to ensure no document is discarded (see ["Some items are not added to the retry queue when they failed to be indexed" on page 26](#)).

TempFileRegex (string)

Regular expression (regex) used to exclude unwanted temporary files from indexation. By default this parameter is empty. This option is only available for CES 6.5.4019+. This option is useful when exclusion filters are not precise enough to exclude specific files such as temporary files. The option can also be used to filter other types of files using a custom regular expression.

Use the following procedure only when you want to modify one or more of the above hidden source parameters.

To modify hidden File connector source parameters

1. Add one or more hidden source parameters (see ["Showing Advanced Hidden Source Parameters" on page 13](#)).
2. For a new File source, access the **Add Source** page of the Administration Tool to modify the value of the newly added advanced parameter:
 - a. Select **Index > Sources and Collections**.
 - b. Under **Collections**, select the collection in which you wish to add the source.
 - c. Under **Sources**, click **Add**.
 - d. In the **Add Source** page, edit the newly added advanced parameter value.
3. For an existing File source, access the **Source: ... General** page of the Administration Tool to modify the value of the newly added advanced parameter:
 - a. Select **Index > Sources and Collections**.
 - b. Under **Collections**, select the collection containing the source you wish to modify.
 - c. Under **Sources**, click the existing File connector source in which you wish to modify the newly added advanced parameter.
 - d. In the **Source: ... General** page, edit the newly added advanced parameter value.

6.4 Showing Advanced Hidden Source Parameters

When you create or configure a source, the CES Administration Tool user interface presents source parameters with which you can configure the connector for most setups. For many connectors, more advanced and more rarely used source parameters also exist but are hidden by default. CES then uses the default value associated with each of these hidden parameters.

You can however choose to make one or more of these parameters appear in the **Add Source** and **Source: ... General** pages of the Administration Tool so that you can change their default value.

Note: Refer to the documentation of each connector to get information on available hidden parameters.

To show advanced hidden source parameters

1. On the CES server, open the Administration Tool (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **Administration Tool**).


2. Access the **Modify Additional Connector** page:
 - a. Select **Configuration > Connectors**.
 - b. In the panel on the left, select **Additional Connector**.
 - c. In the list on the right, select the connector for which you wish to show advanced hidden parameters.
3. In the **Modify Additional Connector** page, for each hidden parameter that you wish to modify, perform the following steps:
 - a. Click **Add Parameter**.
 - b. In the **Modify the parameters of the additional connector** page:
 - i. In the **Type** list, select the parameter type as specified in the parameter description.
 - ii. In the **Name** box, type the parameter name exactly as it appears in the parameter description, noting that parameter names are case sensitive.
 - iii. In the **Default Value** box, enter the default value specified in the parameter description.

Important: Do not set the value that you wish to use for a specific source. The value that you enter here will be used for all sources defined using this connector so it must be set to the recommended default value. You will be able to change the value for each source later, in the **Add Source** and **Source: ... General** pages of the Administration Tool.

- iv. In the **Label** box, enter the label that you wish to see for this parameter.

To easily link the label to the hidden parameter, you can simply use the parameter name, and if applicable, insert spaces between concatenated words.

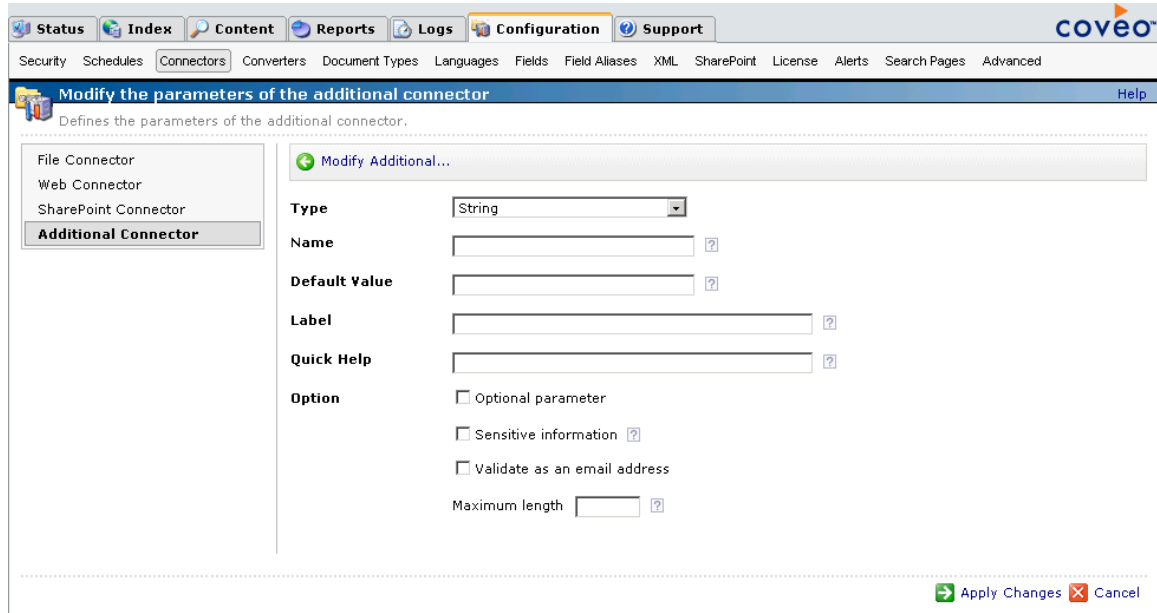
Example: For the **BatchSize** hidden parameter, enter `Batch Size` for the label.

- v. Optionally, in **Quick Help**, enter the help text that you wish to see for this parameter when clicking the question mark button  beside the parameter value.

Tip: Copy and paste the key elements from the parameter description.

- vi. For the **Predefined values** parameter type only, in the **Value** box, enter the parameter values that you wish to see available in the drop-down parameter that will appear in the Administration Tool interface. Enter one value per line. The entered values must exactly match the values listed in the hidden parameter description.
- vii. Select the **Optional parameter** check box when you wish to identify this parameter as an optional parameter. When cleared, CES does not allow to save changes when the parameter is empty. This parameter does not appear for **Boolean** and **Predefined values** parameter types.
- viii. Select the **Sensitive information** check box for password or other sensitive parameter so that in the Administration Tool pages where the parameter appears, the parameter value appears hidden (ex.:). This parameter appears only for the **String** type.
- ix. Select the **Validate as an email address** check box when you want CES to validate that the text string that a user enters in this parameter respects the format of a valid email address. This parameter appears only for the **String** type.
- x. In the **Maximum length** box, enter the maximum number of characters for the string. This parameter appears only for the **String** type.

xi. Click **Apply Changes**.



c. Back in the **Modify the parameters of the additional connector** page, click **Apply Changes**.

The hidden parameter now appears in the **Add Source** and **Source: ... General** pages of the Administration Tool for the selected source. You can change the parameter value from these pages.

6.5 Toggling Live Indexing for a Source

Several Coveo connectors support live indexing, a feature that allows to continuously index changes to the content of a source by detecting addition, removal, or updates to source documents, and rapidly bring these changes within the unified index.

Live indexing is a very useful feature and is generally active by default. You may however encounter cases where you need to turn it off temporarily or permanently.

Example: Mail archive files cannot be shared when they are opened in Microsoft Outlook and therefore, cannot be indexed during that period. It is recommended to turn off live indexing on the source pointing to mail archives, relying on source refresh scheduled at off-hours to update indexing of these files.

To toggle the live indexing for a source

1. On the CES server, open the Administration Tool (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **Administration Tool**).
2. Access the **Source and Collections** page (**Index** > **Source and Collections**).
3. In the **Collections** section, select the collection containing the source for which you want to toggle the live indexing.
4. In the **Sources** section, select the source for which you want to toggle the live indexing.
5. In the toolbar:

- a. When live indexing is active, click **Disable Live Indexing** to deactivate live indexing.
- b. When live indexing is inactive, click **Enable Live Indexing** to reactivate live indexing.

7. Mail Archive Indexing with the File Connector

The File connector can index Microsoft Exchange mail archive files (.pst) that reside on crawled file shares. The File connector supports the legacy ANSI PST file format that was used up to Microsoft Outlook 2003 as well as the Unicode format that was introduced in Microsoft Outlook 2003 and is the only format used since Microsoft Outlook 2007.

Indexing PST mail archive files requires some specific configuration. Consequently, a best practice is to create and configure one source that exclusively index mail archive files on a file share (see "[Mail Archive Indexing Deployment Overview](#)" on page 18).

Note: You can also deploy the Desktop Integration Package (DIP) together with the Desktop connector. End-users can then configure the DIP to crawl mail archive files stored on the local hard drives on their computer or on private network folders so that their content is searchable from the unified index on the Coveo server.

About permissions

The following list describes how the File connector manages permissions on items retrieved from mail archive files:

1. Mailbox

When you use an optional mail archive mapping file, you can associate a specific mailbox to a specific mail archive. The permissions associated to the mailbox in Active Directory are assigned to the mail archive. This type of permission is used first when it exists (see "[Creating a Mail Archive Mapping File](#)" on page 21).

Example: For a mail archive file containing emails of one user, you can associate the mailbox of the specific user to the mail archive file.

2. Mapping file security

When you use an optional mail archive mapping file, you can define allowed users (`AllowedUser`) in the `CommonMappings` and `Mapping` sections of the mapping file for a specific mail archive file. This type of permission is added to the mailbox permissions (see "[Creating a Mail Archive Mapping File](#)" on page 21).

Example: For a mail archive file containing shared emails from the support department, you can allow all users working in the support department to be able to search items from this file.

3. File system

When the permissions of a mail archives file is not defined in a mapping file, the File connector uses the NTFS permissions for the mail archive file to set the permissions on each mail archive item in the unified index.

Live indexing limitation

The File connector cannot index mail archives that are currently opened in a Microsoft Outlook profile. Microsoft Outlook always opens a mail archive in exclusivity mode. Any File connector attempt to open mail archives file during that time fails. Consequently, it is not possible to effectively implement live indexing on mail archives (see "[Toggling Live Indexing for a Source](#)" on page 15).

It is recommended to only index repositories containing mail archives that are not used and to schedule periodic refresh schedules to pick up any changes that could be made to the archives.

About the mail archives modified date attribute

The File connector needs to first add the mail archive to a temporary Microsoft Outlook profile to be able to make MAPI calls to open and process the archive content. Unfortunately, this operation causes Microsoft Outlook to update the modified date attribute of the mail archive file to the current date and time. When the indexing of the mail archive file is completed, the File connector sets the modified date attribute back to its original value. Consequently, for the time required to process the mail archive file, the modified date is changed to the current one.

Important: A temporary change of the modified date attribute for mail archive files could have consequences when a backup or an archiving software actively monitors the repository to detect changes to files based on the modified date attribute.

7.1 Mail Archive Indexing Deployment Overview

When you choose to index the content of mail archive files using the File connector, you need to perform the following tasks.

1. Install the Microsoft MAPI component.

The File connector uses the Microsoft MAPI component to access the content of mail archive files. This component must be installed on the Coveo Master server (see ["Installing the Microsoft MAPI Component for Mail Archive Indexing" on page 18](#)).

2. Create a custom document type set.

In the search results, the URI of an archived email results cannot be opened in the original mail application. The solution is to use the Quick View to open a cached HTML version created in the unified index when the item was crawled. You need to create a special document type set to do so (see ["Setting up a Document Type for Mail Archive Indexing" on page 21](#)).

3. Consider creating an optional mapping file.

By default, the file crawler uses the NTFS permissions on the mail archive file to set the permissions for each mail archive item in the unified index. You need to create and use a mail archive mapping file when you want to override these permissions, index password protected mail archive files, or make mail archive items appear in email search interfaces (see ["Creating a Mail Archive Mapping File" on page 21](#)).

4. Ensure the File connector runs in a 32-bit process.

On a 64-bit server, the File connector must run in a 32-bit process to be able to index PST mail archives. This is the default configuration because the connector uses a third party library that does not support 64-bit processes. The parameter for the connector process type (32-bit or 64-bit) affects all sources for this connector. Consequently, you must ensure that the File connector runs in a 32-bit process (see ["Selecting a 32-bit or 64-bit Process for a Connector" on page 24](#)).

7.2 Installing the Microsoft MAPI Component for Mail Archive Indexing

The File connector uses the Microsoft Messaging API (MAPI) client libraries to access mail archive content. This Microsoft component needs to be installed on the Coveo server.

Note: You can verify if MAPI is installed on the Coveo server by checking in the Registry if the `HKEY_LOCAL_MACHINE\SOFTWARE\Clients\Mail\ExchangeMAPI` key exists.

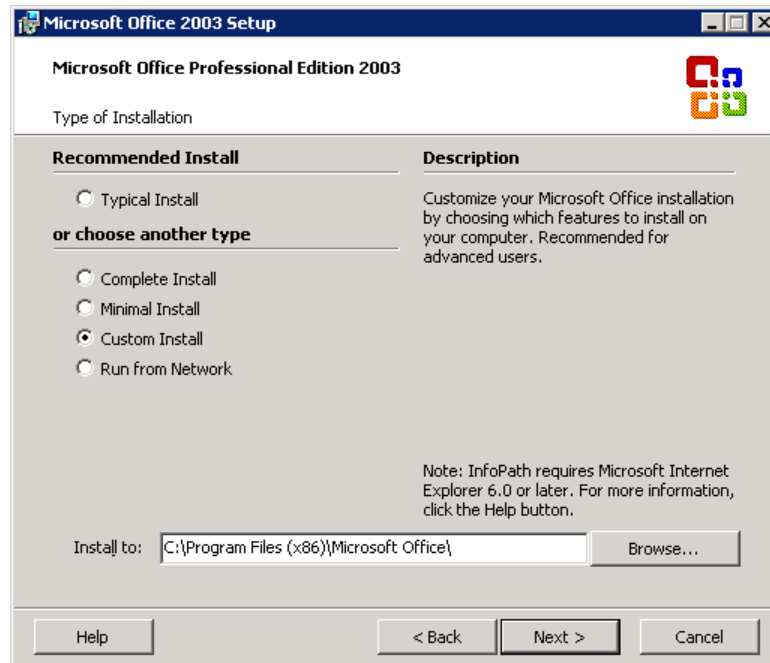
The MAPI component is automatically installed with Microsoft Outlook. This component is available as a standalone download from Microsoft, but the version of MAPI installed by the standalone package only

supports mail archives saved in the Unicode format (introduced in Microsoft Outlook 2003 and the only supported format since Microsoft Outlook 2007), not the legacy ANSI format.

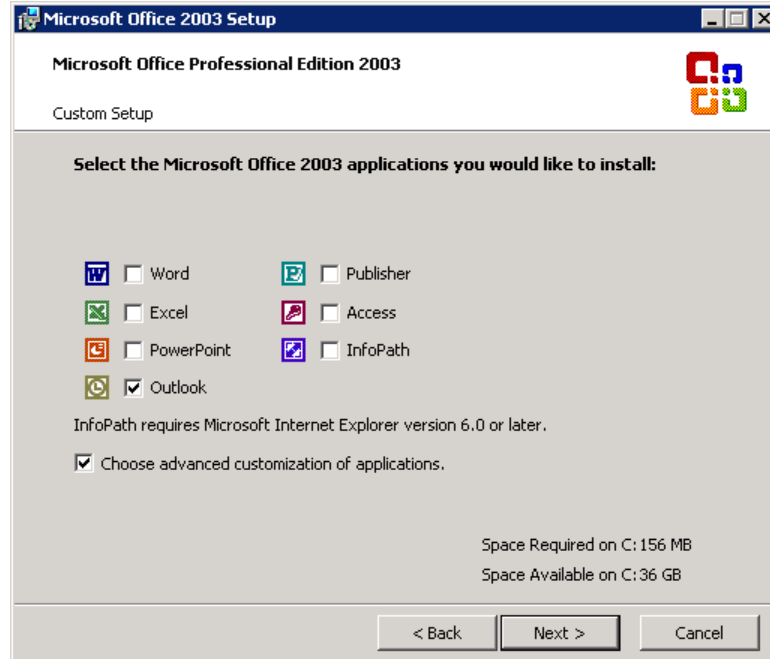
The proposed solution is to install the Collaboration Data Objects (CDO) component that includes MAPI by running the Microsoft Office Professional Edition 2003 installer on the Coveo server but only install the MAPI/CDO components using the advanced installation options.

To install the required MAPI/CDO components

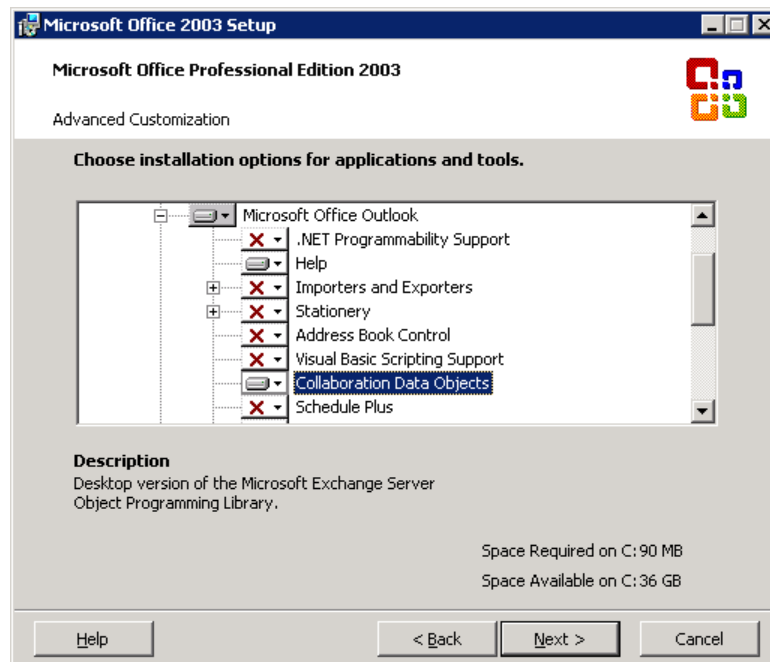
1. Launch the Microsoft Office Professional Edition 2003 installer.
2. In the **User Information** screen, you can leave all parameters empty, and then click **Next**.
3. In the **Type of Installation** screen, select the **Custom Install** option, and click **Next**.



4. In the **Custom Setup** screen:
 - a. Unselect all applications except **Outlook**.
 - b. Select **Choose advanced customization of applications**, and then click **Next**.



5. In the **Advanced Customization** screen:
 - a. Set all installation options to **Not available** for applications other than Microsoft Office Outlook.
 - b. Under **Microsoft Office Outlook**, set **Collaboration Data Objects** to **Run from my computer**.
 This option automatically selects **Outlook Messaging Components** that includes the MAPI for installation.
 - c. Click **Next**.



6. In the **Summary** screen, click **Install**.

7.3 Setting up a Document Type for Mail Archive Indexing

In the search results, the URI of an archived item result cannot be opened in the original mail application. The solution is to use the Quick View to open an HTML cached version of the content created in the unified index when the item was crawled. You need to create a special document type set for mail archives that instructs CES to open results with the HTML cached version.

Note: You need to verify that the **Generate a cached HTML version of indexed documents** option is selected for the source to ensure that HTML cached version of the mail archives items are created when CES crawls the source (see "[Creating and Indexing a File Connector Source](#)" on page 8).

To set up a document type set for mail archives indexing

1. On the CES server, open the Administration Tool (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **Administration Tool**).
2. In the Administration Tool, select **Configuration** > **Document Types**.
3. In the **Document Type Sets** page, click **Add**.
4. In the **Add Document Type Set** page:
 - a. In **Name**, enter a name representing the document type set:

Example: QuickViewMailArchives

- b. In **Description**, optionally enter a description of the purpose of the document type set.
 - c. Click **Save**.
- The new document type set is displayed in the **Document Type Sets** list.
5. Click on the newly created document type set.
 6. In the page that appears, in the **Name** list, click **Exchange Items**.
 7. In the page that appears, in the **Option** section, select the **Open results with cached version** checkbox.
 8. Click **Apply Changes**.

Important: Ensure that every source used to index mail archives uses this new document type set (see "[Creating and Indexing a File Connector Source](#)" on page 8).

7.4 Creating a Mail Archive Mapping File

The File connector can use a mail archive mapping file to get detailed instructions on how to open and index the content of mail archive files. Using a mail archive mapping file is not mandatory, and if you do, having a mapping file entry for each mail archive is not mandatory.

Associating a mail archive mapping file to a File connector source provides the following advantages:

- Allows indexing of password protected mail archive files.
- Allows to associate a Microsoft Exchange mailbox with a mail archive file so that the items it contains are

indexed with the permissions associated with the mailbox. This also sets the `sysmailbox` field for the mail archive items, allowing the items to appear in email search interfaces.

- Can explicitly specify the permissions to the content of a mail archive file or to the content of a folder in a file by setting allowed users or groups.

To create a mail archive mapping file

1. Connect to the Coveo Master server using an administrator account.
2. Using a text editor, create an XML mapping file that respects the mail archive mapping file format and that describes the mail archive file that you wish to index from a given source (see "[Mail archive mapping file format](#)" on page 22).

Tip: You can start with the sample mail archive mapping file available in the `[Installation_path]\Coveo Enterprise Search 6\Bin\Coveo.CES.CustomCrawlers.File.MailArchives.zip` file on the Coveo server.

3. Save the mapping file on the Coveo master server with a name of your choice (ex.: `NetworkShareMailArchivesMappingFile.config`). The recommended folder is `C:\CES6\Config`.

Mail archive mapping file format

The mail archive mapping file can be divided into two sections:

CommonMapping

Settings that apply to all mail archives, whether they are defined in the mapping file or not.

Mapping

Settings for a specific mail archive. A specific mapping overrides a mapping defined in the `CommonMapping` section.

The following sample of a mail archive mapping file illustrates how it can be organized and how to use the various XML elements.

```

<?xml version="1.0" encoding="utf-8" ?>
<MailArchives>
  <CommonMapping>
    <AllowedUsers>
      <AllowedUser type="Windows" allowed="true">
        <Name>corp\administrators</Name>
        <Server></Server>
      </AllowedUser>
    </AllowedUsers>
  </CommonMapping>
  <Mapping type="\svr-archives\mail\employees\jsmith.pst">
    <Fields>
      <Password>12345</Password>
    </Fields>
    <Mailbox active="true">
      <LDAPSearchRoot>LDAP://OU=companynameOU, DC=corp, DC=companyname,
DC=com</LDAPSearchRoot>
      <Name>jsmith@corp.com</Name>
    </Mailbox>
  </Mapping>
  <Mapping type="\svr-archives\mail\employees\jdow.pst">
    <!-- Jane Dow mailbox does not exists anymore, set mailbox active attribute to
false -->
    <Mailbox active="false">
      <Name>jdow@corp.com</Name>
    </Mailbox>
  </Mapping>
</MailArchives>

```

In a mail archives mapping file, you can use the following elements:

Fields

The only field that you can specify for mail archives is the `Password` field. Since a mail archive can be password protected by a user when it is created, this field holds the password used when attempting to open protected archives. If the password of a protected archive is not defined in the mapping file, the archive will not be opened; hence, not indexed.

Important: Special care must be taken when specifying a mail archive password. When you specify a password in the mapping file for a mail archive file that has currently no password, the Microsoft MAPI component opens the mail archive and permanently sets the specified password to the mail archive file.

Mailbox

This is where a Microsoft Exchange mailbox can be associated to a mail archive. This association enables mail archive items to appear in the results of email search interfaces. Without a mailbox association, mail archives items can only appear in the results of generic search interfaces such as the All Content search interface.

The `Mailbox` element requires the following information:

Active

When this attribute is set to true, the security for the mailbox is resolved from Active Directory and is set on each item retrieved from the mail archive. When set to false, it blindly associates the mailbox to the archive items without retrieving its security or validating that the mailbox exists in Active Directory.

Name

Element used to specify the name of the mailbox and set the `sysmailbox` field.

Example: `jsmith@corp.com`

LDAPSearchRoot

This optional element specifies to the connector where to start looking in Active Directory. When this parameter is not specified, the connector looks at the root of Active Directory, which can be extremely large. By specifying a value, you can refine the search and speed up the mapping process.

Example: To search only within the organizational unit (OU) `companynameOU` within the domain `corp.companyname.com`, enter: `LDAP://OU=companynameOU, DC=corp, DC=companyname, DC=com`.

AllowedUsers

Use this element to grant or deny access to the mail archive content. These security settings complement existing ones retrieved from Active Directory when an active mailbox is specified for the archive.

The `AllowedUser` element requires the following information:

Type

Attribute used to set the type of users specified in the `name` element. The two possible values are `Windows` and `WindowSid`.

Name

Element used to specify the name of the Windows User or Group in the form `domain\username` (ex.: `corp\administrators`).

Server

Element used to specify the name of the local machine when referring to local users or groups. For domain users, you should leave this element empty.

7.5 Selecting a 32-bit or 64-bit Process for a Connector

On the 64-bit server, many connectors can run either in a 32-bit or in a 64-bit process. Setting a connector to run in 64-bit allows to take advantage of the 64-bit performance. However, in some cases, connectors need to run in a 32-bit process.

Example: When indexing PST mail archive files, the File connector uses a third party library that does not support 64-bit processes and must therefore run in a 32-bit process.

Note: Selecting a 32-bit or 64-bit process for a connector affects all sources for this connector. Changing the state of the **Run in 64 bits** checkbox requires a refresh of all the sources of this connector.

To select a 32-bit or 64-bit process for a connector

1. On a 64-bit Coveo server, open the Administration Tool (Windows **Start** menu > **All Programs** > **Coveo Enterprise Search 6** > **Administration Tool**).
2. Select **Configure** > **Connectors**.
3. In the navigation panel on the left, select **Additional Connectors**.
4. In the panel on the right, under **Additional Connectors**, select the connector for which you wish to change the process type.
5. In the **Modify Additional Connector** page:

- a. Do one of the following:
 - Select the **Run in 64 bits** checkbox when you want the connector to run in a 64-bit process.
OR
 - Clear the **Run in 64 bits** checkbox when you want the connector to run in a 32-bit process.
 - b. Click **Apply Changes**.
6. Refresh all the sources for this connector.

8. Troubleshooting File Connector Issues

Access denied when crawling through a Distributed File System (DFS)

Possible cause

When the File connector performs a Rebuild or a Refresh operation through a DFS, depending on the number of threads used and the overall crawling speed, several connections can be opened simultaneously to the targeted server. In some instances where the number of connections grows to a large number, Microsoft Windows can see this as a Denial-of-Service attack on the server and start refusing to create new connections to that server. When this situation arises, the File connector logs will display `Access Denied Errors` for any items located on that server. You can confirm this situation by monitoring Windows Event Viewer logs for Event ID 2027.

Possible solution

If you encounter this problem, you can change your source starting address so it targets one of the DFS active referral links instead of going through the DFS itself. This could help resolve the problem if the server experiencing the connection problem is the DFS server and not the file share server where the resource being crawled is located.

Example: Replace the DFS starting address `\\DFSName\Rootname\Ressource` by `\\ServerName\RessourceFileShare`.

The File connector also has the ability to automatically attempt to detect and crawl a DFS active referral link instead of going through the DFS. You can add the `EnableCrawlDFSReferralLink` hidden File connector source parameter and set it to `True` to enable this feature which is disabled by default (see "[Modifying Hidden File Connector Source Parameters](#)" on page 12).

Some items are not added to the retry queue when they failed to be indexed

Possible cause

The File connector automatically retries to index any item that failed to be indexed because it was opened by another application during the initial crawling (sharing violation). To keep the retry queue to a reasonable size, the default maximum number of items in the queue is set to 100. When the connector encounters frequent sharing violations, the limit may be exceeded.

Possible solution

If you experience frequent sharing violations when crawling, you can increase the default value of the `RetryQueueMaxSize` hidden File connector source parameter (see "[Modifying Hidden File Connector Source Parameters](#)" on page 12).

When crawling a Network Share, the "CGLSecurity::SecurityInvalidUserOrGroup: No mapping between account names and security IDs was done" error is displayed for every file

Possible cause

Even though the starting address provided is valid, an error occurred when attempting to resolve the permissions on files from that address. This problem can occur with some network configurations where CES can't properly interpret the host from the supplied starting address.

Possible solution

If you are using the network share fully qualified name, try to use its shortened version.

Example: Use `file://fileshare/root/` instead of `file://fileshare.domain.com/root/` and vice versa.

When crawling a mail archive with the Expand Mail Archives source option enabled, the "Failed to open mail archive, it is in use and cannot be accessed. Make sure it is not still opened in Outlook." error message is returned

Possible cause

The File connector cannot index the content of a mail archive that is currently opened in Microsoft Outlook.

Possible solution

Close the Archive in Microsoft Outlook and retry. When you encounter this error even after closing the archive in Microsoft Outlook, restart Microsoft Outlook to ensure it releases all handles on the archive.

When crawling a mail archive with the Expand Mail Archives source option enabled, the "Failed to open mail archive, it is password protected. The password specified in the Mail Archives Mapping File for this archive is incorrect." error message is returned

Possible cause

This message indicates that a password for the archive was found in the mail archive mapping file but the password is incorrect. Passwords are case sensitive.

Possible solution

Ensure that you specify the passwords in the File connector mapping file with the same casing as when you open the archives in Microsoft Outlook (see "[Creating a Mail Archive Mapping File](#)" on page 21).

When crawling a mail archive with the Expand Mail Archives source option enabled the "Failed to open mail archive, it is password protected. You need to add this archive password to the Mail Archives Mapping File, please refer to the Connector documentation for further details." error message is returned

Possible cause

This means that there is no password specified for the protected archive in the mail archive mapping file

Possible solution

Ensure that you specify a password for each protected archive file in the mail archive mapping file (see "[Creating a Mail Archive Mapping File](#)" on page 21). If you did specify a password for the archive, ensure the mapping type for the archive was properly entered and that it contains the full path to the archive.

Example: `<Mapping type="\svr-archives\mail\employees\jdow.pst">`

