



---

## **Coveo Platform 7.0**

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](http://www.coveo.com).

© Coveo Solutions Inc., 2013

Coveo is a trademark of Coveo Solutions Inc. This document is protected by intellectual property laws and is subject to all restrictions specified in the Coveo Customer Agreement.

Document part number: PM-120818-EN

Publication date: 1/3/2019

## Table of Contents

<b>1. File Connector</b> .....	<b>1</b>
<b>2. File Connector Deployment Overview</b> .....	<b>3</b>
<b>3. File Connector Requirements</b> .....	<b>4</b>
<b>4. Planning Your File Connector Collections and Sources</b> .....	<b>5</b>
<b>5. Setting up a File System Crawling Account</b> .....	<b>7</b>
<b>6. Adding a User Identity</b> .....	<b>8</b>
<b>7. Configuring and Indexing a File Connector Source</b> .....	<b>10</b>
7.1 Modifying Hidden File Connector Source Parameters .....	16
7.2 Adding an Explicit Connector Parameter .....	17
7.3 Toggling Live Monitoring for a Source .....	20
<b>8. Mail Archive Indexing with the File Connector</b> .....	<b>21</b>
8.1 About Permissions .....	21
8.2 Live Monitoring Limitation .....	21
8.3 About the Mail Archives Modified Date Attribute .....	22
8.4 File Connector Mail Archive Indexing Deployment Overview .....	22
8.5 Installing the Microsoft MAPI Component for Mail Archive Indexing .....	23
8.6 Setting up a Document Type for Mail Archive Indexing .....	29
8.7 Creating a Mail Archive Mapping File .....	30
8.7.1 Mail archive mapping file format .....	31
8.8 Selecting a 32-bit or 64-bit Process for a Connector .....	32
<b>9. Troubleshooting File Connector Issues</b> .....	<b>34</b>



# 1. File Connector

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.

The File connector features are:

## Supported file shares

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

- Microsoft Windows Server (2012/2008) and Microsoft Windows (8/7)
- Microsoft Windows Distributed File System (DFS)
- File share on other operating systems (ex.: UNIX, Linux, Mac) accessible through the Windows network. Paths and filenames must be compliant with the Windows naming conventions (see [Naming Conventions](#)).

**Note:** **CES 7.0.7183+ (November 2014)** The File connector supports long Unicode file paths up to 32,767 characters for both refresh and live monitoring operations, breaking the original 260 character maximum path length.

## 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.

## Incremental refresh

The File connector uses file live 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 21).

**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.

### Connector Feature History

CES version	Monthly release	Features
7.0.7183	November 2014	Support for long file paths <a href="#">[more]</a>
7.0.5425	May 2013	Support of permission levels and sets

## 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 7](#)).

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 21](#)).

**Note:** When Coveo runs on a 64-bit server and none of your File connector sources are 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 32](#)).

5. In the Coveo 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 7](#)). You will then assign this crawling account to a user identity in the `domain\username` form (see ["Adding a User Identity" on page 8](#)), and assign the user identity to the source (see ["Configuring and Indexing a File Connector Source" on page 10](#)).

- 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 ["Configuring and Indexing a File Connector Source" on page 10](#)).

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 34](#) and ["Modifying Hidden File Connector Source Parameters" on page 16](#)).

### 3. File Connector Requirements

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

- Coveo license for the File Connector

Your Coveo 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 23](#)).

- When the access to communication ports between the Coveo Master server and the file share server(s) is restricted, the appropriate ports must be opened in the network infrastructure such as in firewalls (see [Understanding Shared Folders and the Windows Firewall](#)).

## 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 generally only visible by Coveo administrator in the Administration Tool.

**Note:** The source names can appear in the search interface for example when a custom facet presents `@syssource` elements.

- You can configure a search interface to include a **Collection** facet or collection check boxes 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 want 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 ["Adding a User Identity" on page 8](#)), and you will assign the user identity to a source (see ["Configuring and Indexing a File Connector Source" on page 10](#)).

**Important:** When indexing PST mail archive files, the crawling account must also have write and modify 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. Adding a User Identity

A user identity is a set of credentials for a given repository or system that you enter once in CES and can then associate with one or more sources or security providers.

A user identity typically holds the credentials of an account that has read access to all the repository items that you want to index. It is a best practice to create an account to be used exclusively by the Coveo processes and for which the password does not change. If the password of this account changes in the repository, you must also change it in the CES user identity.

To add a user identity

1. On the Coveo server, access the Administration Tool.
2. In the Administration Tool, select **Configuration > Security**.
3. In the navigation panel on the left, click **User Identities**.
4. In the **User Identities** page, click **Add**.
5. In the **Modify User Identity** page:

- a. In the **Name** box, enter a name of your choice to describe the account that you selected or created in the repository to allow CES to access the repository.

**Note:** This name appears only in the Coveo Administration Tool, in the **Authentication** or **User Identity** drop-down lists, when you respectively define a source or a security provider.

- b. In the **User** box, enter the username for the account that you selected or created to crawl the repository content that you want to index.
- c. In the **Password** box, enter the password for the account.
- d. In the **Options** section, the **Support basic authentication** check box is deprecated and not applicable for

most types of repositories. You should select it only when you need to allow CES to send the username and password as unencrypted text.

- e. Click **Save**.

**Important:** When you use Firefox to access the Administration Tool and it proposes to remember the password for the user identity that you just created, select to never remember the password for this site to prevent issues with automatic filling of username and password fields within the Coveo Administration Tool.

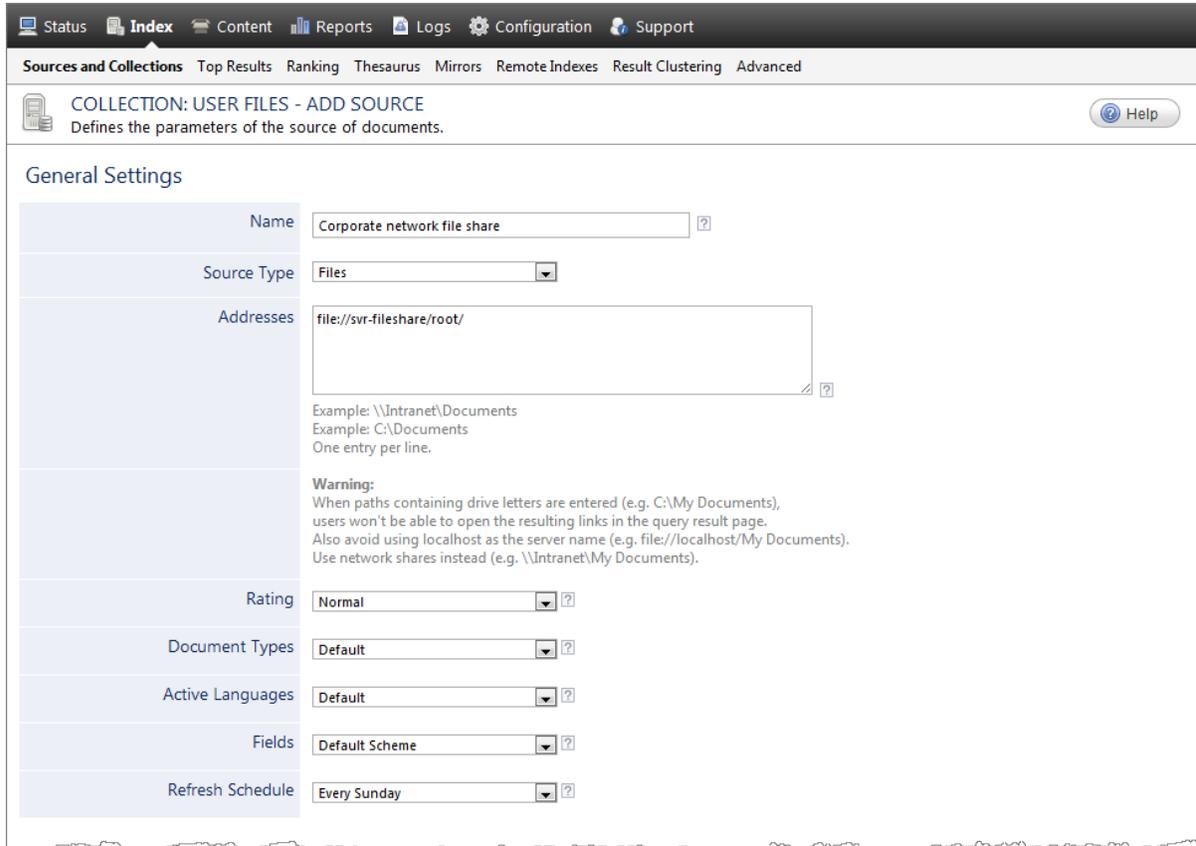
## 7. Configuring and Indexing a File Connector Source

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 configure and index a File connector source

1. On the Coveo server, access the Administration Tool.
2. Select **Index > Sources and Collections**.
3. In the **Collections** section:
  - a. Select an existing collection in which you want to add the new source.  
OR
  - b. Click **Add** to create a new collection.
4. In the **Sources** section, click **Add**.  
The **Add Source** page that appears is organized in three sections.
5. In the **General Settings** section of the **Add Source** page:



a. Enter the appropriate value for the following required parameters:

**Name**

Enter a descriptive name of your choice for the connector source.

**Example:** Corporate network file share

**Source Type**

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

**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/mails/jsmith.pst</code>
Folder in a mail archive:	<code>file://svr-fileshare/mails/jsmith.pst/work</code>
IP address	<code>file://192.168.1.2/share</code>

**Important:** When you use paths containing drive letters as starting addresses (ex.: `C:\fileshare`), 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`).

**Note:** For Windows Server 2008, an access denied folder and the files it contains are not indexed. For Windows Server 2003, when a folder was set in the file system to deny access to the crawling account, the folder name was indexed, but not the files it contained.

**Refresh Schedule**

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

**Note:** You can create new or modify existing source refresh schedules.

- b. Review the value for the following parameters that often do not need to be modified:

**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.

**Document Types**

If you created a custom document type set for this source, select it. Otherwise, select **Default**.

**Active Languages**

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

**Fields**

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

6. In the **Specific Connector Parameters & Options** section of the **Add Source** page:

- a. 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 21 and "Creating a Mail Archive Mapping File" on page 30).

**Example:** C:\CES7\Config\Coveo.CES.CustomCrawlers.File.MailArchives.config

**Expand Mail Archives**

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

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

**Number of Live Monitoring Threads**

Determines the number of file system changes that the connector live monitoring 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 want 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.

### Index Share Permissions

By default this option is cleared. Select this option to index both the share and NTFS permissions (see the Microsoft document [Share and NTFS Permissions on a File Server](#)).

### Parameters

Click **Add Parameter** when you want to show advanced hidden source parameters (see "[Modifying Hidden File Connector Source Parameters](#)" on page 16).

- c. 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 selected, CES indexes all the document metadata, even metadata that are not associated with a field. The orphan metadata are added to the body of the document so that they can be searched using free text queries.

When cleared (default), only the values of system and custom fields that have the **Free Text Queries** attribute selected will be searchable without using a field query.

**Example:** A document has two metadata:

- `LastEditedBy` containing the value `Hector Smith`
- `Department` containing the value `RH`

In CES, the custom field `CorpDepartment` is bound to the metadata `Department` and its **Free Text Queries** attribute is selected.

When the **Index the document's metadata** option is cleared, searching for `RH` returns the document because a field is indexing this value. Searching for `hector` does not return the document because no field is indexing this value.

When the **Index the document's metadata** option is selected, searching for `hector` also returns the document because CES indexed orphan metadata.

**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.

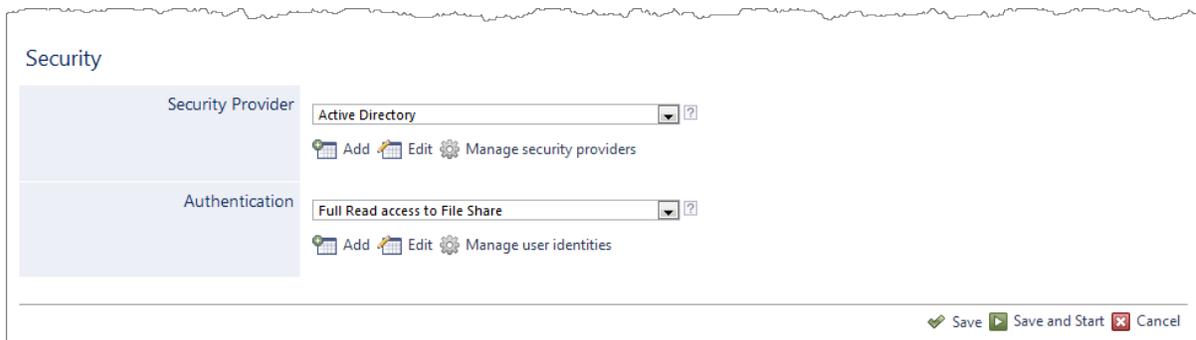
Consider clearing this check box only if you do not want to use Quick View links or to save resources when building the source.

**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** check box.

**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 29).

7. In the **Security** section of the **Add Source** page:



- a. In the **Security Provider** drop-down list, select **Active Directory** or a custom Active Directory security provider that you created for a specific domain.
- b. In the **Authentication** drop-down list, when you chose to use a specific account to crawl the file system (see "[Setting up a File System Crawling Account](#)" on page 7), select the user identity that you created for this account. Leave this parameter empty when you want the connector to crawl the file system using the

CES service identity.

- c. Click **Save and Start** to save the source configuration and start indexing this source.
8. Validate that the source building process is executed without errors:
- In the navigation panel on the left, click **Status**, and then validate that the indexing proceeds without errors.
- OR
- Open the CES Console to monitor the source building activities.

## 7.1 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.

### **EnableCrawDFSReferralLink (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 34). The default value is `False`.

### **IgnoreUnresolvedDeniedSecurities (Boolean)**

Set to True to ignore unresolved denied permissions. This option is useful to voluntarily ignore unresolved denied security errors. The default value is `False`.

**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 modification events to store for each monitored starting address before discarding them. Discarded modification events will be indexed the next time the source is refreshed. This parameter is useful to prevent queuing a large number of modification events that would take a large amount of memory on the server when many files under the starting address are modified in a very short period of time. The default value is `100,000`.

### **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 34](#)).

### **TempFileRegex (String)**

Regular expression (regex) used to exclude unwanted temporary files from indexation. By default this parameter is empty. 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.

### **FileSystemWatcherBufferSize (Integer)**

Buffer size of the File System Watcher instance used by the File Crawler to monitor file system events. For best performance, use a multiple of 4 KB (4096). Increasing this buffer size is expensive because it comes from non-paged memory that cannot be swapped out to disk. Keep the buffer as small as possible. The default value is 8192.

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 ["Adding an Explicit Connector Parameter" on page 17](#)).
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 want 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 want to modify.
  - c. Under **Sources**, click the existing File connector source in which you want to modify the newly added advanced parameter.
  - d. In the **Source: ... General** page, edit the newly added advanced parameter value.

## 7.2 Adding an Explicit Connector Parameter

Connector parameters applying to all sources indexed using this connector are called explicit parameters.

When you create or configure a source, the Coveo Enterprise Search (CES) 7.0 Administration Tool presents parameters with which you can configure the connector for most setups. For many connectors, more advanced and more rarely used 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.

### To add an explicit connector parameter

1. On the Coveo server, access the Administration Tool.
2. Select **Configuration > Connectors**.
3. In the list on the **Connectors** page, select the connector for which you want to show advanced hidden parameters.
4. In the **Parameters** section of the selected connector page, click **Add Parameter** for each hidden parameter that you want to modify.

**Note:** The **Add Parameter** button is present only when hidden parameters are available for the selected connector.

5. In the **Modify the parameters of the connector** page:

- a. In the **Type** list, select the parameter type as specified in the parameter description.
- b. In the **Name** box, type the parameter name exactly as it appears in the parameter description. Parameter

names are case sensitive.

- c. In the **Default Value** box, enter the default value specified in the parameter description.

**Important:** Do not set the value that you want 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.

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

**Example:** To easily link the label to the hidden parameter, you can simply use the parameter name, and if applicable, insert spaces between concatenated words. For the **BatchSize** hidden parameter, enter `Batch Size` for the label.

**Note:** To create multilingual labels and quick help messages, use the following syntax: `<@ln>text</@>`, where *ln* is replaced by the language initials—the languages of the Administration Tool are English (en) and French (fr).

**Example:** `<@fr>Chemin d'accès du fichier de configuration</@><@en>Configuration File Path</@>` is a label which is displayed differently in the French and English versions of the Administration Tool.

**Tip:** The language of the Administration Tool can be modified by pressing the following key combination: `Ctrl+Alt+Page Up`.

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

**Tip:** Copy and paste key elements of the parameter description.

- f. When **Predefined values** is selected in the **Type** parameter, in the **Value** box that appears, enter the parameter values that you want 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.
- g. Select the **Optional parameter** check box when you want to identify this parameter as an optional parameter. When cleared, CES does not allow you to save changes when the parameter is empty. This parameter does not appear for **Boolean** and **Predefined values** parameter types.
- h. Select the **Sensitive information** check box for password or other sensitive parameter so that, in the Administration Tool pages where the parameter appears, the typed characters appear as dots to mask them. This parameter appears only for the **String** type.

**Example:** When you select the **Sensitive information** check box for a parameter, the characters typed appear as follows in the text box:



- i. 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.

- j. In the **Maximum length** box, enter the maximum number of characters for the string. This parameter appears only for the **String** type. When you enter 0, the length of the string is not limited.
  - k. Click **Save**.
6. Back in the **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. Refer to the documentation for each connector for details.

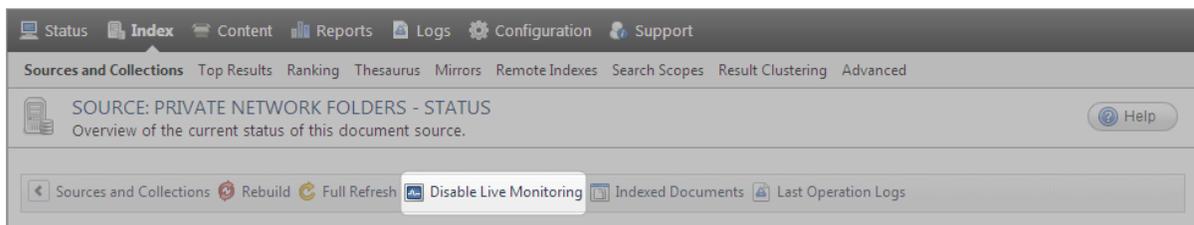
**Note:** When you want to modify a hidden source parameter, you must first delete it, and then redefine it with the modified values.

### 7.3 Toggling Live Monitoring for a Source

Some repository types can detect that a document was added, modified or deleted and notify external systems. The File connector uses this technique called *live monitoring* to maintain the index up-to-date with the source content. When supported, live monitoring is enabled by default. You can however manually enable or disable live monitoring.

To toggle live monitoring for a source

1. On the Coveo server, access the Administration Tool.
2. Access the **Sources and Collections** page (**Index > Sources and Collections**).
3. In the **Collections** section, click the collection that contains the source for which you want to toggle the live monitoring state.
4. In the **Sources** section, click the appropriate source.
5. On the horizontal button bar:



- When live monitoring is active, click **Disable Live Monitoring** to turn it off.
- OR
- When live monitoring is turned off, click **Enable Live Monitoring** to activate it.

## 8. 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 ["File Connector Mail Archive Indexing Deployment Overview" on page 22](#)).

**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.

### 8.1 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 30](#)).

**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 30](#)).

**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.

### 8.2 Live Monitoring 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 monitoring on mail archives. Consider turning off live monitoring on the source created to index your mail archives (see ["Toggling Live Monitoring for a Source" on page 20](#)).

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.

### 8.3 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 archiving software actively monitors the repository to detect changes to files based on the modified date attribute.

### 8.4 File Connector 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 23](#)).

2. Create a custom document type set.

In the search results, the URI of 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 29](#)).

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 30](#)).

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 32](#)).

## 8.5 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. You therefore need to install this Microsoft component on the Coveo server when you want to index mail archive content.

You must install the MAPI component using the Microsoft Office installer on the Coveo server. You can install only the MAPI component using the installation options.

**Note:** Do not use the Microsoft standalone installation package for MAPI and CDO. This package does not install all the components that the File connector needs.

You can use the installer for various versions of Microsoft Office:

- ["Installing the MAPI components with Microsoft Office 2010" on page 23](#)
- ["Installing the MAPI components with Microsoft Office 2007" on page 25](#)
- ["Installing the MAPI components with Microsoft Office 2003" on page 27](#)

### Installing the MAPI components with Microsoft Office 2010

1. Using an administrator account, connect the Coveo Master server on which a File connector needs to access mail archive content.

**Note:** By default, the Microsoft Office Professional Plus 2010 installer does not offer options to select specific components to install. You can however use the Office Customization Tool that is part of the setup program to specify components to install.

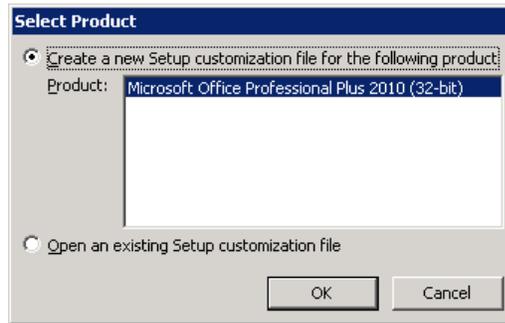
2. Launch the Microsoft Office 2010 installer using the `/admin` option to start the Office Customization Tool:
  - a. Open a **Command Prompt** window.
  - b. Type `setup.exe /admin` at the command line from the root of the network installation point that contains the Office 2010 source files.

**Example:** `\\server\share\Office14\setup.exe /admin`

The Office Customization Tool starts.

3. In the **Select Product** dialog box that appears:
  - a. Select **Create a new Setup customization file for the following product**.
  - b. Ensure Microsoft Office is selected in the **Product** list.

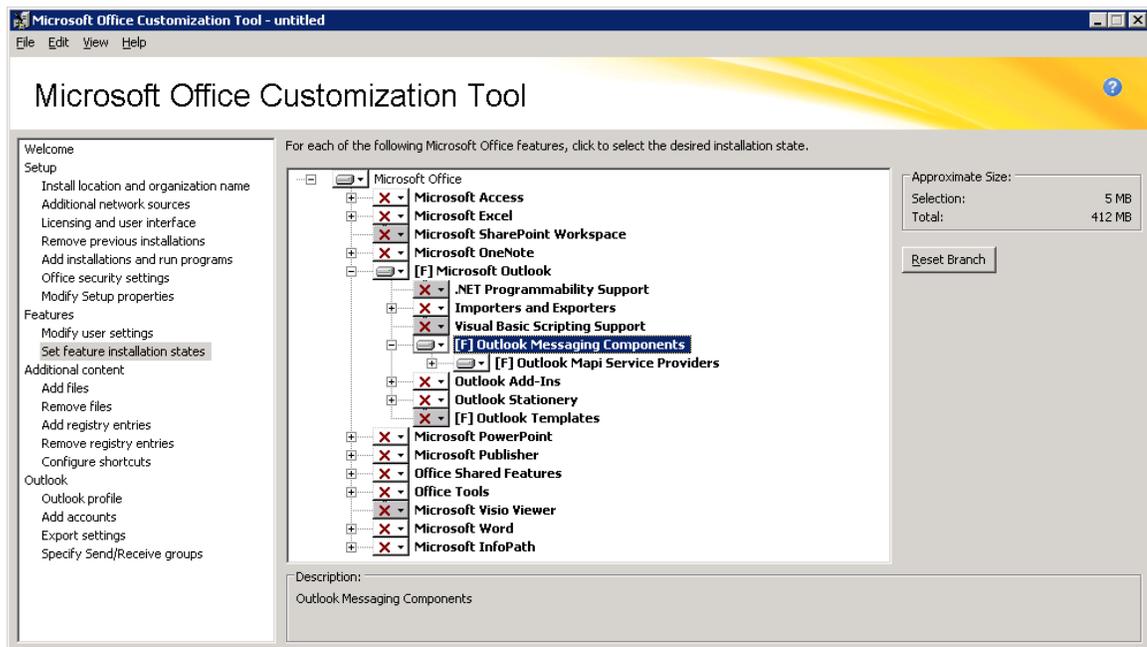
- c. Click **OK**.



4. In the **Microsoft Office Customization Tool**:

- a. In the navigation panel on the left, under **Features**, select **Set feature installation states**.
- b. In the central panel, expand the product tree to show **Microsoft Office > Microsoft Outlook > Outlook Messaging Components**.
- c. Right-click **Outlook Messaging Components**, and then select **Run from My Computer**.
- d. For all other product tree elements, right-click the element and then select **Not Available**.
- e. On the menu, select **File > Save As** to save a customization .msp file with a name and to a location of your choice.

**Example:** C:\user\username\Desktop\MyMSOfficeConfig.msp



- 5. Start the Microsoft office installer using the setup command-line option `/adminfile` to specify the fully-

qualified path of the location of the .msp file:

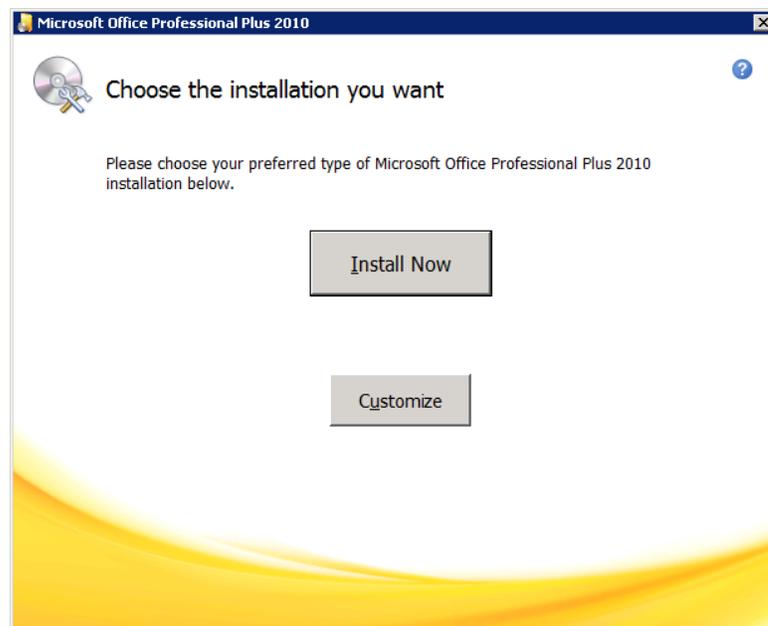
- In a **Command Prompt** window, type a command in the form:

```
setup.exe /adminfile [path]\[customization_file]
```

**Example:**

```
setup.exe /adminfile C:\user\username\Desktop\MyMSOfficeConfig.msp
```

6. In the **Microsoft Office Professional Plus 2010** dialog box:
  - a. Read and accept the Microsoft Software License Terms, and then click **Continue**.
  - b. In the **Choose the installation you want** screen, click **Install Now** to install the components selected in the customization file.

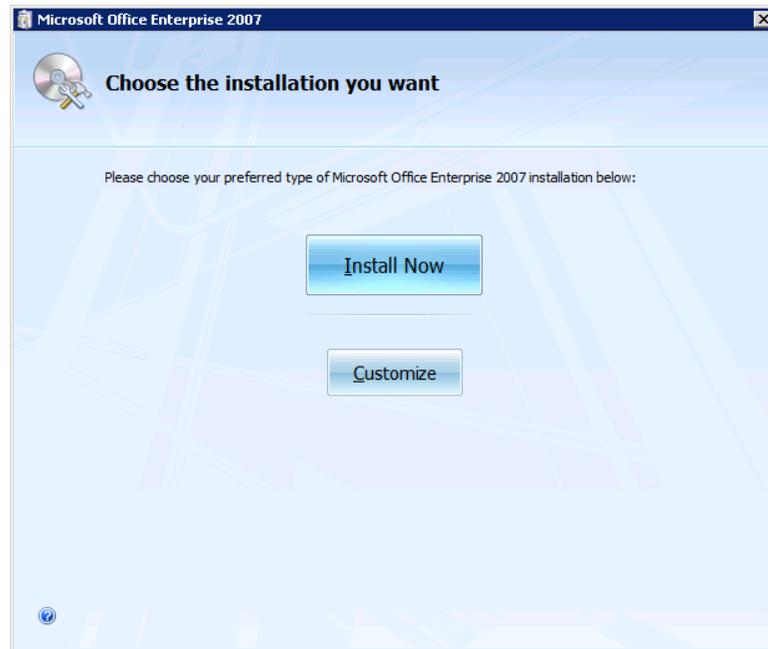


The **Installation Progress** screen appears while the installation is performed.

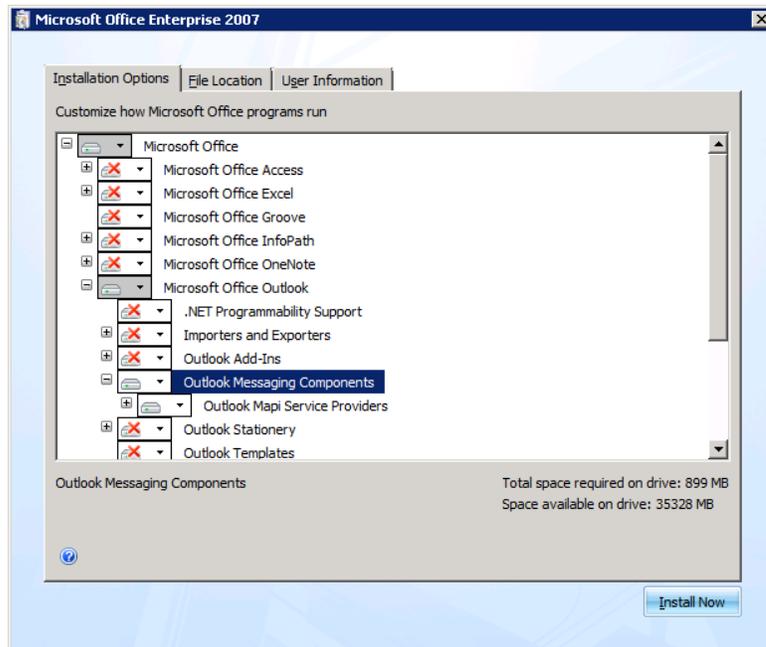
- c. In the **Complete your Office experience** screen, click **Close**.

### Installing the MAPI components with Microsoft Office 2007

1. Using an administrator account, connect the Coveo Master server on which a File connector needs to access mail archive content.
2. Launch the Microsoft Office 2007 installer.
3. Read and accept the Microsoft Software License Terms, and then click **Continue**.
4. In the **Choose the installation you want** screen, click **Customize**.



5. In the **Installation Options** tab:
  - a. Expand the product tree to show **Microsoft Office > Microsoft Office Outlook > Outlook Messaging Components > Outlook Mapi Service Providers**.
  - b. Click **Outlook Mapi Service Providers**, and then select **Run from My Computer**.
  - c. For all other product tree elements, click the element and then select **Not Available**.



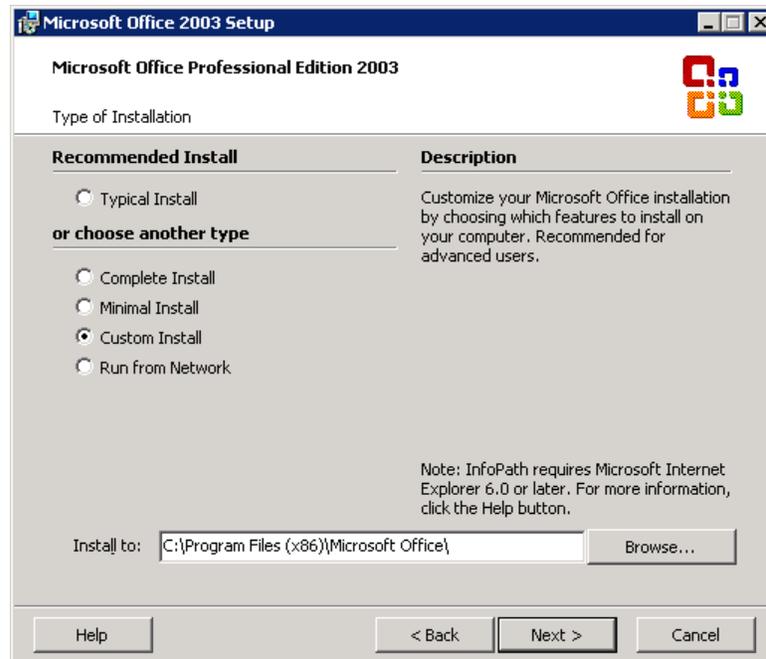
d. Click **Install Now**.

The **Installation Progress** screen appears while the installation is performed.

6. In the last screen, click **Close**.

### Installing the MAPI components with Microsoft Office 2003

1. Using an administrator account, connect the Coveo Master server on which a File connector needs to access mail archive content.
2. Launch the Microsoft Office Professional Edition 2003 installer.
3. In the **User Information** screen, you can leave all parameters empty, and then click **Next**.
4. In the **Type of Installation** screen, select the **Custom Install** option, and then click **Next**.

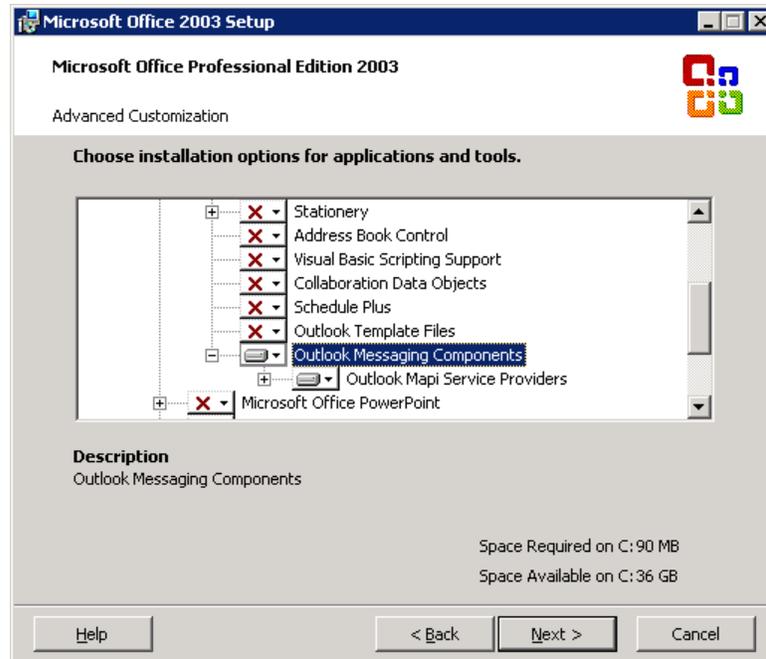


5. In the **Custom Setup** screen:
  - a. Unselect all applications except **Outlook**.
  - b. Select the **Choose advanced customization of applications** check box, and then click **Next**.



6. 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 **Outlook Messaging Components** to **Run from my computer**.
- c. Click **Next**.



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

## 8.6 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 a Quick View version of the mail archives items is created when CES crawls the source (see "[Configuring and Indexing a File Connector Source](#)" on page 10).

To set up a document type set for mail archives indexing

1. On the Coveo server, access the 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** check box.
8. Click **Apply Changes**.

**Important:** Ensure that every source used to index mail archives uses this new document type set (see "[Configuring and Indexing a File Connector Source](#)" on page 10).

## 8.7 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 want to index from a given source (see "[Mail archive mapping file format](#)" on page 31).

**Tip:** You can start with the sample mail archive mapping file available in the `[CES_Path]\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 `[Index_Path]\Config`.

## 8.7.1 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>
    <AllowedMapping>
      <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 archive 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.

## 8.8 Selecting a 32-bit or 64-bit Process for a Connector

On a 64-bit Coveo server, many connectors can run either in a 32-bit or 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.

**Important:** 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** check box 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, access the Administration Tool.
2. Select **Configure > Connectors**.
3. In the **Connectors** page, select the connector that you want to modify.
4. In the page for the connector:
  - a. In the **Option** section, select or clear the **Run in 64 bits** check box when you want the connector to run respectively in a 64-bit or 32-bit process.
  - b. Click **Apply Changes**.

**Note:** You may need to restart the CES service to make changes effective (see [Starting the CES Service](#)).

5. Refresh all the sources for this connector.

## 9. 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 16).

### 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 16).

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 30](#)).

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 30](#)). 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">