CA DataMinder

Stored Data Integration Guide

Release 14.6

CA Technologies Product References

This document references the following CA Technologies products:

- CA DataMinder™

Contact CA Technologies

Contact CA Support

For your convenience, CA Technologies provides one site where you can access the information that you need for your Home Office, Small Business, and Enterprise CA Technologies products. At http://ca.com/support, you can access the following resources:

- Online and telephone contact information for technical assistance and customer services
- Information about user communities and forums
- Product and documentation downloads
- CA Support policies and guidelines
- Other helpful resources appropriate for your product

Providing Feedback About Product Documentation

If you have comments or questions about CA Technologies product documentation, you can send a message to techpubs@ca.com.

To provide feedback about CA Technologies product documentation, complete our short customer survey which is available on the CA Support website at http://ca.com/docs.

Contents

Chapter 1: About Stored Data Integration

This section contains the following topics:

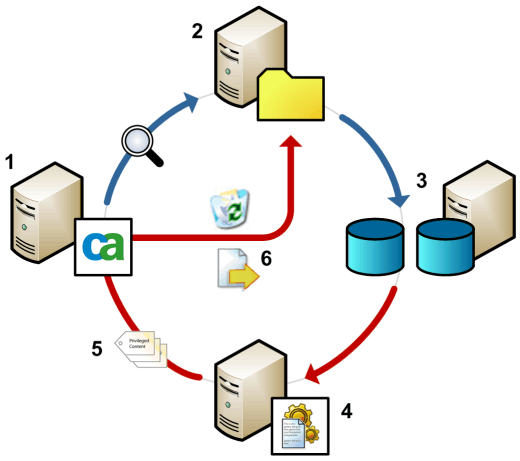
About the FSA

The File Scanning Agent (FSA) integrates CA DataMinder with the stored data used by your organization. It is designed to scan, analyze and apply appropriate policy controls to:

- Files saved in local or remote file systems
- Items stored in Exchange Public Folders
- Items hosted on Microsoft SharePoint sites
- Text entries and documents stored in SQL Server or Oracle databases.

The FSA must be deployed in conjunction with CA DataMinder policy engines and a policy engine hub. It uses Data At Rest triggers to analyze and apply policy to scanned items. It only extracts text content if it is required for policy processing, making the scans highly efficient.

When a scanning job runs, an event is generated for each scanned item that causes a Data At Rest trigger to fire. All file events captured by the FSA can be viewed alongside existing e-mail, Web and IM events using the iConsole or Data Management console.



FSA: scanning procedure for files

The FSA (**1**) scans target folders (**2**) and checks files against records in the Scanned File database and, optionally, a NIST database (**3**). New or changed files are passed to a policy engine (**4**) for processing. Data At Rest triggers analyze the files and, if required, apply smart tags (**5**). When processing is complete, the policy engine calls back to the FSA with the processing results. Finally, the FSA then implements any applicable control actions, such as deleting specific files or copying them to a new location (**6**).

Scanning Jobs

A scanning job definition defines which machines, folders and files are scanned, and optionally when and how often the file scan runs. It also specifies which user's policy is applied to scanned files and, optionally, which users are stored in the CMS database as event participants.

Applying Policy to Scanned Items

CA DataMinder uses Data At Rest triggers to analyze and apply policy to scanned items. These triggers can detect specific file names, analyze the text content, add smart tags, and (by using XML Attribute lookup commands) identify file attributes such as size, date last modified, and the document author. In addition, Data At Rest control actions can be configured to delete scanned files or copy or move them to an alternative location.

More information:

Set Up CA DataMinder Policy Triggers (see page 40)

File System Scans

These are identified in the Job Definition wizard as 'Drives and folders' scans.

Which Files Can Be Scanned?

The FSA can monitor files in specified folders on specified machines and can manage multiple scanning jobs simultaneously. It can scan all currently supported file types and apply policy to those files using Data At Rest triggers.

Which File Types Can Be Excluded?

When you define a scanning job, you can specify whether the scan includes system, hidden and offline folders and files. You can also specify whether to scan subfolders.

- **Offline files and folders**

  The 'offline' job attribute indicates that the text content of a file (or files within a folder) has been archived by a third party application and the original file replaced with a stub file. Scanning offline files and folders can take a long time because the third party archive application must first retrieve the original files. You can therefore configure the FSA to ignore offline files and folders.

- **System files and folders**

  If required, you can exclude system files and folders. But note that very few files and folders have this Windows attribute. To omit specific system files or folders from your scans, we recommend you exclude them individually in the job definition.

Which Files Are Not Scanned?

The FSA ignores the following files:

- **Files on removable drives**

  If a scanning job is configured to scan all ‘local hard drives’, be aware that this will **not** include removable disk drives. That is, the FSA will not scan files on DVDs or CDs, USB flash drives, SD cards, optical drives, and so on.

  However, the FSA **will** scan a removable drive if it is specified explicitly in the job definition as a scan location (that is, the job specifies a path to a removable drive). But note that the FSA cannot perform Delete or Replace actions on read-only media such as CDs and DVDs.

- **Unchanged files**

  The FSA ignores files that have not changed since the last scan. That is, it does not scan unchanged files. This greatly speeds up regular file scans, particularly if only a small proportion of files have been modified since the previous scan. The FSA can quickly identify these files by checking the file hashes in the scanned file database.

- **Excluded files**

  When you set up a scanning job, you can explicitly exclude files in specific folders and files. Alternatively, you can configure jobs to only include specific folders and files.

- **Large files (above a maximum size)**

  You can specify that files over a certain size are not scanned. To do this, you edit the Maximum Size of Files setting in the user policy System Settings folder.'

  Note: To ensure that files of any size are scanned, set Maximum Size of Files to a value of zero.

More information:

Set Up CA DataMinder Policy Triggers (see page 40)

SharePoint Scans

SharePoint sites can contain a variety of items, such as document libraries, pictures, tasks and discussion boards. CA DataMinder puts these items into two categories: files or documents, and 'generic items'. The FSA Remote Connector can scan the text content of all SharePoint items but it only applies policy to files or documents.

Generic Items

Generic items include Announcements, Discussion Boards, Events, Issue Tracking, and Tasks. The FSA Remote Connector can scan the text content of generic items but cannot apply policy to them.

However, it is possible for generic items to have file attachments and the FSA Remote Connector *can* analyze and apply policy to these attachments.

Do Not Scan Template Files

When scanning SharePoint sites, we recommend you exclude template files. SharePoint comes with its own set of template files. To ensure that SharePoint users do not encounter potential problems from moved or deleted template files, we recommend that you exclude such files when setting up SharePoint scans.

Exchange Public Folders Scans

The FSA can scan the text content of items in Microsoft Exchange Public Folders (for example, calendar events, tasks, or journal entries). In particular, it can scan documents attached to the items.

The FSA can then apply policy to these scanned items and, if necessary, delete them or copy or move them to an alternative location.

Extracting MAPI Properties

The FSA can also extract MAPI properties from items scanned in Exchange Public Folders. It adds these MAPI properties to the metadata for the resulting file event. In turn, this allows Data At Rest policy triggers to test these MAPI properties when applying policy. For example, you may choose to move an attachment to a secure location if the ‘Last Author’ property does not match a list of approved authors.

Why must you specify the individual MAPI properties that you want to extract from scanned items? After all, this is not necessary when scanning file systems for, say, Microsoft Office documents. The reason lies in the method used to store item attributes in Exchange Public Folders.

For Microsoft Office documents stored in a file system, the FSA attempts, as far as possible, to extract all the relevant document properties and add them to the metadata of the resulting file event (the metadata is stored in XML format).

But this approach is not possible for items posted to Exchange Public Folders. The attributes of these items (such as 'Last Author' or 'Sensitivity') are saved as MAPI properties and are too numerous for the FSA to routinely extract all available properties for each item. Instead, you need to configure the FSA scanning job to detect and extract the specific MAPI properties that are relevant to your policy triggers. The triggers can then test these extracted MAPI properties by using XMLATTR data lookup commands.

Database Scans

You can set up FSA jobs to scan Oracle and SQL Server databases. Specifically, the FSA can scan columns in database tables that contain text or binary data. Binary data refers to documents (such as Microsoft Office files).

What Are Database Events?

When the FSA scans a database table, the scanned data is written to XML blob files. The FSA then passes these blob files to policy engines for processing.

Policy engines then apply Data At Rest triggers to the blob files. If a trigger fires, a database event is generated and stored on the CMS. You can search for these events in the iConsole.

Data At Rest control actions (such as deleting or replacing scanned items) are **not** applied to scanned database records.

How Does the FSA Create Database Events?

By default, all scanned data in the table (each column for every row) is written to a single XML blob file and stored as a single event. However, for very large database tables, this is clearly undesirable. Instead, you can configure the FSA to slice the scanned data into smaller, more easily manageable events.

For example, you can specify a maximum number of rows per event, or a maximum size (in KB) per event. You can also configure special handling for binary data; for example, if a database table contains MS Word documents, you can specify that these documents are stored as attachments to database events.

Where Do I Install the FSA?

The target location for the FSA depends on which items you want the FSA to scan:

Files saved in local or remote file systems

If you want to a file system, you can install the FSA directly on the file server that you want to scan. Alternatively, you can install the FSA on a network server and run remote scans of multiple file servers.

If you need to scan sensitive folders on a remote server that cannot be accessed by a UNC path, you must install the FSA on a network server and deploy an FSA Remote Connector on the remote server. The FSA Remote Connector runs on the remote server and passes scanned items back to the FSA for policy processing.

We also recommend that you deploy an FSA Remote Connector if your scanning jobs are configured to exclude a large proportion of files. For these scanning jobs, it is more efficient to identify these exclusions on the host server instead of retrieving all files across the network to the FSA for analysis.

Items stored in Microsoft Exchange Public Folders

If you want to scan the text content of items in Exchange Public Folders, or documents attached to these items, install the FSA on a network server that has a local MAPI client. Specifically, install the FSA on a server that has the 'Messaging API and Collaboration Data Objects 1.2.1' component installed.

This component can be downloaded from the Microsoft Web site. Before you install it, ensure that Microsoft Outlook is not also installed; if it is, you must remove it.

Note: We recommend that you do not install the FSA directly on an Exchange server. The FSA is resource-intensive and may affect Exchange performance.

Items hosted on Microsoft SharePoint sites

If you want to scan SharePoint items, either install the FSA directly on the SharePoint server, or deploy an FSA Remote Connector on the SharePoint server.

We recommend that you use an FSA Remote Connector to reduce the processing required on the SharePoint server. The FSA Remote Connector runs on the SharePoint server and passes scanned items back to the FSA for policy processing.

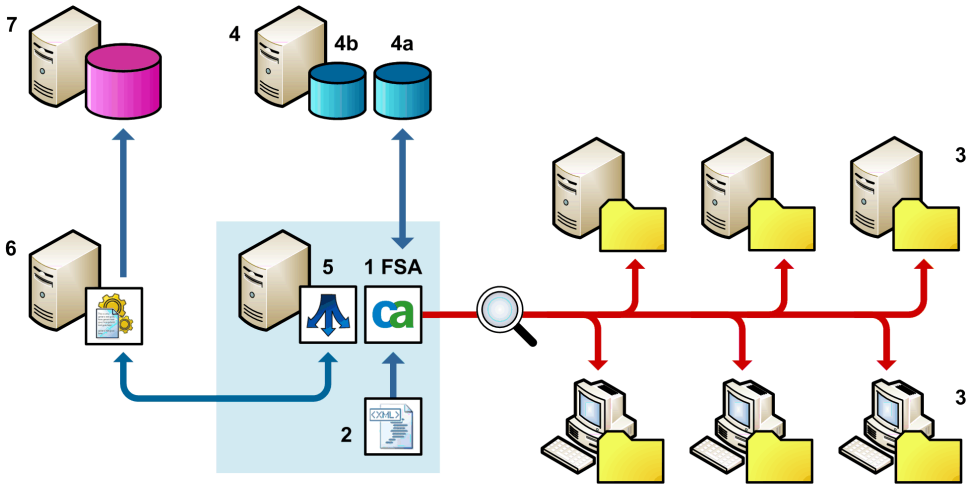
Note: You cannot use the FSA to scan a remote SharePoint server.

Text entries and documents stored in SQL Server or Oracle databases

If you want to scan a database, install the FSA on a network server. (You specify the DBMS connection details when you set up a scanning job.)

FSA Architecture

The FSA can scan, analyze and apply policy to files saved in file system, items in Exchange Public Folders and on SharePoint sites, and database records. It can run multiple scanning jobs simultaneously, with each job scanning separate folders or machines and, if required, using a separate 'run as' account to access remote machines. An example FSA deployment architecture for scanning remote file systems is shown below.



Example FSA deployment: file scanning

1. **FSA:** The FSA runs scanning jobs to monitor files saved in remote file systems. It can delete or copy specified files.
2. **Scanning job file:** An XML file containing the job definition.
3. **Scanned folders:** The job file defines which items you want to scan and the target host machine.
4. **Database:** The scanned files database (**4a**) contains records for each file already scanned. The FSA queries this database to check whether a file needs to scanned in the current job.

   For files that **do** need scanning, the FSA can then optionally query a NIST database (**4b**) to identify system files which can be omitted from the scanning job.
5. **Policy engine hub:** Files that need to be analyzed are passed to the hub by the FSA. The hub then allocates files to policy engines for processing (**6**).

   When processing is complete, the hub also passes back to the FSA details of any actions that must be taken (for example, to delete a file or copy it to a new location).
6. **Policy engines:** When the policy engine hub receives a new file from the FSA, it allocates the file to the least heavily loaded policy engine (that is, the policy engine that can process the new file most quickly).

   The policy engine then analyzes the file and applies Data At Rest triggers as necessary. Any resulting control actions (for example, to delete or copy a file) are passed back to the hub (**5**). The hub then relays these actions to the FSA (**1**).
7. **CMS:** Each policy engine replicates processed file events up to the CMS. File events can be viewed alongside existing e-mail, Web and IM events using CA DataMinder consoles.

FSA Terminology

Note the following terminology:

Scanned file database

The FSA uses a scanned file database to track the status of each file in a scanning job. For each scanned file, the database contains a hash (see below) to uniquely identify that version of the file plus its 'last scanned' date. For scheduled scanning jobs, the FSA checks the hashes in the scanned file database to see whether a file has changed or moved since the last scan. This allows the FSA to skip files which have not changed.

Similarly, if a scanning job gets interrupted before it is finished (because, say, there is a network or system failure), the FSA checks the hashes in the database when the job next runs and skips any files which have not been modified since they were last scanned.

Note: Files (binary data) found during database scans are **not** stored in the scanned file database.

NIST database

Also known as the National Software Reference Library (NSRL), the NIST database is a list of known benign and malicious files, maintained by the National Institute of Standards and Technology (NIST). Its purpose is to ease the burden of investigating computer files. Desktop computers can contain over 100,000 files, so investigators need to eliminate as many known files as possible from having to be reviewed. From the NIST Web site:

"The NSRL provides a repository of known software, file profiles, and file signatures for use by law enforcement and other organizations in computer forensics investigations."

The FSA can use this database to identify files that do not need scanning. It checks scanned files for specific profiles and signatures; if any match known files in the NSRL database, the FSA can omit these files from the scan. The NSRL also enables the FSA to identify files that are not what they claim to be (for example, a file with the same name, size, and date of a known file, but not the same content).

Note: Files (binary data) found during database scans are not checked against the NIST database.

File hashes

Before scanning a file, the FSA applies a SHA-256 cryptographic hash function to the file based on the file name, path, size and last modified date. From this, it generates a string that represents a digital fingerprint of that file. In FSA terms, this fingerprint is the **file hash**, also sometimes referred to as a hash code, hash sum, or hash value. File hashes are stored in the FSA scanned file database (see above).

Each time a scanning job is repeated, the FSA compares a file's newly-generated hash with the hash from the previous scan. If these differ, the FSA infers the file has changed since and so scans it again; if they are the same, it infers the file is unchanged and ignores it.

Note: Hashes are **not** generated for files (binary data) found during database scans.

DoD deletion

This is forensic deletion, so called because the storage media are purged or 'sanitized' to guarantee that data cannot be recovered and used to obtain evidence in legal discovery. 'DoD' is a reference to Department of Defense approved methods for purging storage media.

Unlike conventional delete operations where the file header is overwritten, a DoD deletion overwrites the disk sector multiple times in a prescribed pattern to ensure that deleted files cannot be recovered.

Note: DoD deletions are not available for scanned items in Exchange Public Folders and SharePoint sites, or for scanned database records.

DSN

A Database Source Name (DSN) describes a connection to a specific database through an ODBC driver. The DSN specifies all parameters of the connection, including the host machine, port and database name, host server, other information. You can use a DSN in an application to query the database. The FSA installer uses DSNs to connect to the scanned file database and NIST database.

More information:

How Do I Delete a Scanned File Database? (see page 61)

Chapter 2: Deploying the FSA

This section contains the following topics:

FSA Requirements

Before deploying the FSA, note the following requirements.

FSA Operating System

The FSA is included in the integration.msi and integration_x64.msi installation packages.

Supported Versions

Integration.msi supports a 32-bit version of:

- Windows Server 2003 (see note 1)
- Windows Server 2008

Integration_x64.msi supports 64-bit versions of these operating systems:

- Windows Server 2003 (see note 1)
- Windows Server 2008 (see note 1)
- Windows Server 2008 R2
- Windows Server 2012

Note 1: We have not tested these operating systems with the current versions of integration.msi and integration_x64.msi.

FSA User Accounts

The FSA requires the following user accounts:

FSA job setup user

Log on to the Administration console using this Windows user account. The Administration console then uses this account to connect to the FSA server when you create or manage scanning jobs.

FSA service user

The FSA service runs as this user account. The user must be a local administrator on the FSA host machine.

FSA Run As user

A scanning job runs as this user account. There are two types of Run As user: a limited access user and a full access user. These can be used to test different aspects of data security on your network.

List and Site Permissions

If you intend to use the FSA to scan items on Microsoft SharePoint sites, the user that the Remote FSA Connector runs as on the SharePoint host machine must have sufficient List and Site permissions to the Microsoft SharePoint site being scanned. For example, a capture-only policy requires at least the 'View Items' and 'View Application Pages' List permissions and the 'Browse Directories' Site permission.

PE domain user

As part of the policy engine hub installation, the wizard prompts you for the credentials of the PE domain user. This user account must be a member of the local Administrators group on the FSA host server. Confirm this is so before installing the hub.

More information:

FSA Job Setup User (see page 27)
FSA Run As User (see page 29)
FSA Service User (see page 28)

Windows Remote Registry Service

The Remote Registry service must be running on the FSA server if you want to manage scanning jobs from a remote Administration console.

You can manage FSA scanning jobs in the Administration console. But if the Administration console is on a different server to the FSA, the Remote Registry service must be running on the FSA server. If the service is not running, start the service manually,

If an administrator tries to schedule a scanning job on a remote FSA server but the Remote Registry service is not running, the Administration console displays a warning.

Scanned File and NIST Databases

The FSA uses a scanned file database to track the status of each file in a scanning job. It can optionally use a NIST database (see page 25) to identify files that do not need scanning.

The scanned file database and the NIST database both use SQL Server. Therefore, the FSA must have access to either a local or remote SQL Server instance. This SQL Server instance must already be installed before you install the FSA.

Scanned File database

The FSA supports SQL Server 2008 and 2012.

The FSA installation wizard prompts you for the database host server and an existing SQL Server login.

NIST database (optional)

The NIST database supports SQL Server 2008 and 2012.

You must manually attach the NIST database to a SQL Server before installing the FSA. The FSA installation wizard prompts you for the SQL Server hosting the NIST database.

Note: Do not confuse these DBMS requirements with the databases that can be scanned by the FSA.

More information:

FSA Terminology (see page 16)
Database Scans (see page 23)

Exchange Public Folder Scans

Microsoft Exchange

The FSA can scan public folders in the following versions of Exchange:

Microsoft Exchange Server 2003, 2007 or 2010

Note: The FSA is unable to scan public folders in Exchange 2013. This limitation is caused by changes in Exchange 2013 support for MAPI. We anticipate that this limitation will be fixed in a future CA DataMinder release.

Exchange-compatible Email Application

The user account that the FSA runs as (such as the FSA full access user) requires a default email application compatible with Microsoft Exchange, such as Microsoft Outlook.

Local MAPI Client

The FSA host server must have a MAPI client installed.

We recommend the 'Messaging API and Collaboration Data Objects 1.2.1' component. This software is typically referred to as the 'MAPI client and CDO 1.2.1' component. Download the 'MAPI client and CDO 1.2.1' component from the Microsoft Web site. We recommend using the latest build, which is unrelated to the 1.2.1 version number.

SharePoint Scans

Note the following requirements:

Supported Versions of SharePoint

The FSA can scan SharePoint sites based on the following versions of Microsoft SharePoint:

- SharePoint 2003
- SharePoint 2007
- SharePoint 2010
- SharePoint 2013

SharePoint Scan Requirements

FSA Remote Connector

If you want to scan Microsoft SharePoint items, either install the FSA on the SharePoint server, or deploy an FSA Remote Connector on the SharePoint server. We recommend using the FSA Remote Connector to reduce the processing required on the SharePoint server. You cannot use an FSA to scan a remote SharePoint server.

List and Site Permissions

If you intend to use the FSA to scan items on Microsoft SharePoint sites, the user that the Remote FSA Connector runs as on the SharePoint host machine must have sufficient List and Site permissions to the Microsoft SharePoint site being scanned. For example, a capture-only policy requires at least the 'View Items' and 'View Application Pages' List permissions and the 'Browse Directories' Site permission.

Load-Balanced SharePoint Environments

A load-balanced SharePoint environment has multiple front-end SharePoint web servers. However, you need to install the FSA Remote Connector on a single front-end web server.

For details, see Load-Balanced SharePoint Environments (see page 35).

Remote Domain Scans

Scanning jobs in remote domains are not supported. That is, if the FSA host server is in domain A, you cannot use an FSA Remote Connector to configure a scanning job in domain B.

Database Scans

Supported databases

The FSA can scan records stored in the following databases:

- Microsoft SQL Server 2008 and 2012
- Oracle 10g (10.2.0.4), and 11g (11.1.0.7) or later

Database client tools

Verify that the Oracle or SQL Server client tools are installed on the same server as the FSA.

More infomation:

Install Database Client Tools (see page 27)

FSA Deployment Overview

This section describes how to install and configure the FSA. The key deployment tasks are as follows:

1. Deploy your policy engines.
2. Deploy the FSA. You need to:
   a. (Optional) Deploy a NIST database. If you use a NIST database to filter file scans, you must install it before deploying the FSA.
   b. (Optional) Install database client tools. If you want to run database scans, you must ensure the appropriate client tools are installed on the server hosting the FSA or FSA Remote Connector.
   c. Specify FSA user accounts.
3. Install the FSA.
   a. (Optional) Install a Remote Policy Engine Connector.
   b. (Optional) Install a NIST Database Connector.
4. (Optional) Deploy FSA Remote Connectors.
5. Configure the FSA.
6. Set up CA DataMinder user policies.
7. Configure FSA scanning jobs.

More information:

Deploy FSA Remote Connectors (see page 33)
Set Up CA DataMinder Policy Triggers (see page 40)
Install Database Client Tools (see page 27)
Specify FSA User Accounts (see page 27)

Deploy Policy Engines

For installation and configuration instructions, see the Policy Engines chapter in the *Platform Deployment Guide*.

In particular read the instructions for specifying the PE domain user. You must specify this user account when you install a policy engine hub.

Deploy a NIST Database

(A NIST database is optional.)

The FSA can optionally use a NIST database of known files to identify files that do not need scanning.

If you intend to use a NIST database to filter file scans, CA distributes a NIST database that is pre-configured for use by the FSA. This database ships separately from the FSA installation package. You must manually install the NIST database before installing the FSA. Specifically, you must attach the WGN_NIST database to any supported instance of SQL Server.

More information:

FSA Service User (see page 28)

What is a NIST Database?

Also known as the National Software Reference Library (NSRL), the NIST database contains known benign and malicious files. The database is maintained by the National Institute of Standards and Technology (NIST). From the NIST Web site:

"The NSRL provides a repository of known software, file profiles, and file signatures for use by law enforcement and other organizations in computer forensics investigations."

The purpose of the NSRL is to ease the burden of investigating computer files. A typical desktop computer can contain over 100,000 files, so investigators need to eliminate as many known files as possible from having to be reviewed. For further details, see the NIST Web site: http://www.nsrl.nist.gov

RDS Version

The file signatures and identifying information in the database, called the Reference Data Set (RDS), is distributed through NIST's Standard Reference Data Group as NIST Special Database 28. The RDS version included in this image is:

Version 2.38, September 2012

Note: Although this version of the NIST database was released in conjunction with CA DataMinder r14.5, you can deploy it with other versions of CA DataMinder if required.

Attach a NIST Database

You must install the NIST database before installing the FSA. Specifically, you must attach the WGN_NIST database to any supported instance of SQL Server.

To attach a NIST database

1. Run installNIST.cmd.

   Find this file in the root of the distribution image for the NIST database.

   Run this command file on the SQL Server host machine.

2. When prompted, enter the path to the target folder for the NIST database.

   The installation utillity attaches a WGN_NIST database to the target instance of SQL Server. To do this, it installs the following data and transaction log files:
   `WGN_NIST.mdf`
   `WGN_NIST_log.ldf`

   The WGN_NIST database is successfully attached to SQL Server.

3. Install the FSA.

   As part of the FSA installation, you must also install the NIST Database Connector.

More information:

Install the FSA (see page 31)

Update a NIST Database

If you have an earlier version of the WGN_NIST database, you must uninstall it before installing a newer version.

To remove a NIST Database

1. Verify that there are no FSA jobs running that use the NIST database.
2. Start the SQL Server Management Studio and connect to the RDBMS where the WGN_NIST database is located.
3. Select the WGN_NIST database and choose the option to delete it.

You are now ready to attach a new NIST database.

Install Database Client Tools

Before you can use the FSA to scan databases, verify that the appropriate Oracle or SQL Server client tools are installed on the server hosting the FSA (or, if used, the FSA Remote Connector).

These client tools are available from your database vendor and include the drivers (that is, the OLE DB Providers) that the FSA uses to access the target database. For example, Oracle Data Access Components (ODAC) include the necessary drivers for accessing Oracle databases.

When you create a new scanning job using the FSA Job Definition wizard in the Administration console, you must choose from a list of SQL Server or Oracle OLE DB Providers. This allows the FSA to identify what type of database it must connect to and which drivers it must use.

In technical terms, the FSA uses the specified OLE DB Provider type plus other details supplied by yourself (such as the database name and authentication details) to generate a connection string that it uses to connect to the target database.

Specify FSA User Accounts

To create an FSA scanning job, you need to provide credentials for the following Windows domain users:

- FSA Job Setup User
- FSA Service User
- FSA Run As User

FSA Job Setup User

You use the FSA job definition wizard in the Administration console to create and manage scanning jobs stored on the FSA server. The console therefore needs to connect to the FSA, which typically runs on a remote server. The FSA then creates or updates the actual scanning job definitions.

To connect to the FSA, the Administration console uses the Windows account that you used to log on to the Administration console host machine. This is your **FSA job setup user**. Throughout this section, the term ‘FSA job setup user’ refers to the domain user that the Administration console uses to connect to the FSA.

When you log on to Windows on the Administration console host machine, you must do so as the FSA job setup user. This user must have write access to the \FSA\Jobs subfolder on the FSA host server. Find this subfolder in CA's folder in the Windows All Users profile on the machine hosting the FSA.

FSA Service User

When you install the FSA, the installation wizard prompts you for the logon account that is used by the FSA service. This account is your **FSA service user**. The FSA service runs as this user.

This user account must be a member of the local administrators group on the FSA host server and (if used) the Remote FSA Connector host server.

Verify that the FSA Service User is Authenticated in SQL Server

The FSA connects to SQL Server using Windows Authentication, passing the logon account for the FSA service user to SQL Server. But note these requirements:

- The FSA service user must be a valid SQL Server login with permissions to read and write to the scanned file database and, if used, the NIST database:
  - If the FSA service user is a domain administrator, it is automatically a valid SQL Server login.
  - If the FSA service user is not a domain administrator, you must manually add it as a new login on the server or servers hosting the scanned file database and NIST database. You do this using SQL Server Management Studio.
- The FSA service user must be listed explicitly under the 'Users' for the scanned file database and if used, the NIST database. In both cases, these database-specific users must have the db_owner role.

FSA Run As User

The FSA Run As user is the user account that FSA scanning jobs run as. This guide identifies two types of Run As user, reflecting their different purposes:

- **Limited Access FSA Run As User:** You can use the FSA to test whether sensitive documents stored on your network are accessible to unauthorized users. To do this, you set up a scanning job to run as a user with limited access to sensitive network locations. This is your 'limited access FSA Run As user'.

  This 'limited access' user is representative of your ordinary network users. If, while running as this user, the FSA detects and scans documents that should not be accessible to ordinary users, this indicates that your network security needs to be tightened.

- **Full Access FSA Run As User:** You can use the FSA to scan the content of files stored on your network to determine if sensitive information is stored in the correct location or to identify files or documents with unauthorized content. (Indeed, this is often regarded as the 'classic' use for the FSA.)

  To do this, the FSA must be able to access remote file systems and delete files when instructed to do so as a result of policy processing. Specifically, the FSA must run as a domain user with permission to delete and copy files on any machines that you want to scan. This user is your 'full access FSA Run As user'. See the account requirements below.

In both cases, we recommend that the FSA Run As user is a custom account created for exclusive use by the FSA. *This is because it is essential that nobody logs in to a scanned machine using the 'Run As' account while a scheduled scan is running!*

You typically specify an FSA Run As user when you schedule a scanning job using the Job Definition wizard in the Administration console (see step *3* of Schedule a scanning job) or when you log in to Windows before running a scanning job from a command line.

Requirements for Run As User

Note the following requirements for the FSA Run As User:

Local Administrator on FSA Server

The FSA Run As User must be a member of Local Administrators group on FSA server.

Microsoft Exchange Public Folders

For Exchange Public Folder scans, the FSA Run As user requires:

- An Exchange mailbox. (You can typically create a mailbox at the same time as you create a new user account in Active Directory.)
- Read access to Exchange Public Folders for the limited access user.
- Read and Write access to Exchange Public Folders for the full access user. This permits the FSA to copy or delete Public Folder items.

The FSA also requires a default email application compatible with Microsoft Exchange, such as Microsoft Outlook.

Microsoft SharePoint sites

For SharePoint scans, the FSA Run As user:

- Must be a domain user with local administrator rights on the SharePoint host machine.
- Must be a member of the MSSQL db_owner role on the SharePoint configuration database and any SharePoint content databases that contain the site data being scanned.
- Requires sufficient List and Site permissions to the Microsoft SharePoint site being scanned. For example, a capture-only policy requires at least the 'View Items' and 'View Application Pages' List permissions and the 'Browse Directories' Site permission.

Install the FSA

You install the FSA using the CA DataMinder Integration Agents installation wizard.

If required, you can choose to install a policy engine hub when you install the FSA. If you choose not to install a hub, the FSA will automatically passed scanned items to a local policy engine for processing. You must ensure this policy engine is already installed before you run a scanning job.

To install the FSA

1. Find setup.exe in the root of your CA DataMinder distribution image. Run setup.exe to launch the CA DataMinder installation wizard.

   The Installation Type screen opens.

2. Click Advanced Installation.

3. In the Advanced Install Options screen, choose Server Agents and then click Install.

   This launches the CA DataMinder Integration Agents installation wizard in a separate window.

4. In the Integration Agents installation wizard, navigate to the Customer Setup screen.

5. In the Custom Setup screen, choose:

   - File Scanning Agent
   - (Optional) NIST Database Connector

     Important! If you want to use a NIST database, you must install it *before* you install the FSA.

   - (Optional) Remote Policy Engine Connector. Install this feature if you want to use a policy engine hub to distribute scanned items to remote policy engines for processing.

   Note: Do *not* choose the File Scanning Agent Remote Connector. You install this feature separately on the machine you want to scan. See Deploy FSA Remote Connectors (see page 33).

6. In the Data Location screen, specify the name and network location of the data folder. Accept the default location or specify a different location.

   The FSA stores XML scanning job definition files in an \FSA subfolder below the specified data folder. If you use the Content Registration feature, CA DataMinder stores content indexes in an \FSA\PRC subfolder.

   Note: To build content indexes, CA DataMinder runs a specialized FSA scanning job.

7. In the File Scanning Agent Account screen, specify the logon accounts used by the FSA service.

   This service must run as the **FSA service user**. Click the Browse button  for the FSA service, then enter the domain, name and password of the FSA service user.

8. In the Scan Database Location screen, you specify where you want to create your scanned file database; this database tracks the status of each item in a scanning job.

   Server

   Click the 'Server' button  to select the DBMS host server in the Database Server dialog. This dialog lists any servers found to be hosting SQL Server. In the case of multiple SQL Server instances running concurrently on the same computer, the dialog identifies each instance as:

   `<machine name>\<Instance name>`

   Where <machine name> is the name of the server on which SQL Server is running and <Instance name> is the name of the SQL Server instance. For example:

   `MyDBServer\Instance_1`

   Username and Password

   Specify the credentials for the SQL Server login that the FSA uses to access the scanned files database.

9. If you chose the NIST Database Connector in step 2, the NIST Database Location screen prompts for the name or IP address of the server hosting your NIST database.

   Click the 'Server' button  to select the host server in the Database Server dialog. This dialog lists any servers found to be hosting SQL Server.

10. If you chose to install a Remote Policy Engine Connector in step 2, the Policy Engine Hub Configuration screen prompts for credentials for the PE domain user.

11. In the final wizard screen, click Install to start the file transfer.

12. You now need to configure the policy engine hub and the FSA. You may also need to:

    - Install an FSA Remote Connector. This is essential for SharePoint scans and may also be necessary for other scanning jobs for reasons of security or efficiency.
    - Securely store the logon credentials used by the FSA when scanning database records. This avoids the need to store these credentials in the actual job definition file on the FSA host server.

    See the references below for details about these tasks. When these tasks are complete, you can start a scanning job.

More information:

FSA Service User (see page 28)
Deploy FSA Remote Connectors (see page 33)
Securely Store Logon Credentials for Database Scans (see page 33)
Configure the PE Hub (see page 33)
Configure the FSA (see page 36)
Where Do I Install the FSA? (see page 13)

Configure the PE Hub

This is described in Configure the policy engine hub.

Securely Store Logon Credentials for Database Scans

When scanning database records, the FSA connects to the target database using the authentication method specified in the scanning job definition.

As a secure alternative to storing a database user's password in plain text in the XML job definition file, you can use wgncred.exe to cache the password; the required component ID is: **FSADBSCAN**

Deploy FSA Remote Connectors

The FSA Remote Connector is a comparatively small utility that enables the FSA to run remote scans. In effect, it runs a local scan on the remote machine and passes scanned items back to the FSA for policy processing. It is needed for scanning SharePoint sites but may also be your preferred method for scanning databases or files and folders on remote machines.

When you configure a scanning job to run on a remote server (such as a DBMS or SharePoint host machine), you specify the server hosting the FSA Remote Connector in the General Scan Options screen of the FSA job definition wizard.

Why Deploy an FSA Remote Connector?

It may be necessary for reasons of security or efficiency to run the scanning job on a remote machine. Installing the FSA Remote Connector on a remote machine allows you, in effect, to run a scan on that machine. Scanned items that require policy processing are sent back across the network to the FSA.

For example:

- If you want to scan Microsoft SharePoint items, either install the FSA on the SharePoint server, or deploy an FSA Remote Connector on the SharePoint server. We recommend using the FSA Remote Connector to reduce the processing required on the SharePoint server. You cannot use an FSA to scan a remote SharePoint server.
- If scanning a large database, it may be more efficient to install the FSA Remote Connector on the machine hosting the DBMS. This prevents the bottleneck of having to retrieve database records across the network to the FSA for analysis.
- If you need to scan sensitive folders on a remote machine that cannot be accessed by a UNC path, deploy an FSA Remote Connector on that machine.
- If a scanning job is configured to exclude a large proportion of files, then it is more efficient to identify these exclusions on the host machine, rather than retrieving all files across the network to the FSA for analysis.

Load-Balanced SharePoint Environments

In a load-balanced SharePoint environment, you do *not* need to install the FSA Remote Connector on each front-end SharePoint web server. Just install the FSA Remote Connector on the front-end web server that you want CA DataMinder to use when running SharePoint scanning jobs. When a user creates a scanning job in the Administration console, they specify the front-end web server hosting the FSA Remote Connector.

You only need to install the FSA Remote Connector on multiple front-end SharePoint web servers if the front-end web servers are serving up different content.

Also, you may want to install the FSA Remote Connector on multiple front-end SharePoint web servers so you can run multiple scanning jobs. By running smaller scanning jobs on multiple front-end web servers, you can reduce the load on each front-end web server. This may benefit other users trying to access your SharePoint site.

Examples

- Front-end web servers A, B, C, and D each serve management documents and research documents. Install the FSA Remote Connector on server A only to scan all management and research documents.
- Front-end web servers A and B serve management documents, while front-end web servers C and D serve research documents. Install the FSA Remote Connector on server A to scan all management documents and on server C to scan all research documents.

Install the FSA Remote Connector

You install the FSA Remote Connector using the CA DataMinder Integration Agents installation wizard:

To install the FSA Remote Connector

1. Find setup.exe in the root of your CA DataMinder distribution image. Run setup.exe to launch the CA DataMinder installation wizard.

   The Installation Type screen opens.

2. Click Advanced Installation.
3. In the Advanced Install Options screen, choose Server Agents and then click Install.

   This launches the CA DataMinder Integration Agents installation wizard in a separate window.

4. In the Integration Agents installation wizard, navigate to the Customer Setup screen.
5. In the Custom Setup screen, choose the File Scanning Agent Remote Connector.
6. In the final wizard screen, click Install to start the file transfer.

Note: If you install an FSA Remote Connector, you must still install the FSA itself somewhere on your network as part of your CA DataMinder enterprise.

Configure the FSA

You configure the FSA by modifying the registry. For example, you can specify a default scanning job run by the FSA service. Other registry values determine how to handle failed attempts to apply policy to files. To configure the FSA, you modify values in the following registry key:

```
HKEY_LOCAL_MACHINE\SOFTWARE\ComputerAssociates\CA DataMinder
    \CurrentVersion\FSA
```

FSA Registry Values

The following registry values are created automatically in the FSA registry key when you install the FSA:

LogLevel

Type: REG_DWORD

Data: Defaults to 2. This determines the level of logging for file processing. For example, you can configure the FSA to only log errors or important system messages.

Log entries are written to the wgnfsa_<date>.log file. The log file is saved in CA's \data\log subfolder of the Windows All Users profile on the machine hosting the FSA.

The supported logging levels are:

1 Errors only

2 Errors and warnings

3 Errors and warnings, plus informational and status messages

Note: Setting LogLevel=3 will cause the log file to grow extremely rapidly. This level of logging is provided for testing purposes only.

LogMaxNumFiles

Type: REG_DWORD

Data: Defaults to 10. This specifies the maximum number of log files. When the maximum number of log files exists and the maximum size of the latest is reached (see below), the oldest log file is deleted to enable a new one to be created.

LogMaxSizeBytes

Type: REG_SZ

Data: Defaults to 1,000,000. This specifies the maximum size for each log file. When the current log file reaches its maximum size, the FSA creates a new log file.

ThrottlePercent

Type: REG_SZ

Data: Defaults to zero. This specifies how much time (as a percentage of total time) the FSA spends waiting, rather than reading file data from a disk. For example, to set waiting time to 30%, set this parameter to 30 (do not include a '%' character).

This parameter enables you to restrict how much time the FSA spends reading data from disk. For example, if you specify 30% waiting time, then the FSA can only spend 70% of its time reading data. This can be useful to prevent network and system performance problems during intensive scanning operations (for example, when multiple jobs are running simultaneously or when multiple worker threads are scanning a single volume).

WorkerThreadCount

Type: REG_DWORD

Data: Defaults to 10. This specifies the number of concurrent 'worker' threads used by the FSA to analyze files.

AnalysisRetryAttempts

Type: REG_DWORD

Data: Defaults to 0. If the first attempt to pass a file to the policy engine hub fails, this registry value determines how many times the FSA retries before writing a 'file failure' entry to the log. After a first failed attempt, files waiting to be retried are moved to the retry queue.

Such failures can occur when, for example, the hub has suspended operations because it has exceeded its maximum memory allocation. Likewise, a file can be timed out by the hub and passed back to the FSA if no policy engines are available to process the file.

ActionRetryAttempts

Type: REG_DWORD

Data: Defaults to 0. If the first attempt to execute a control action on a file (typically 'delete') fails, this registry value determines how many times the FSA retries before writing a file failure entry to the log. For example, the FSA may be unable to delete a file because the file is open.

Data At Rest control actions can stipulate that a file is deleted or moved:

- File deletions are carried out by the FSA when it receives an instruction to do so from the policy engine via the hub.
- File moves are actually consecutive 'copy and delete' operations. First, a file is copied to a new location; then the original version is deleted.

If the FSA is unable to delete a file, the file is moved to a retry queue. The FSA tries to delete the file again after the retry period expires (this defaults to five minutes; see below).

Note: If, as part of a file move operation, the FSA successfully copies a file but is then unable to delete the original version, this is recorded in the log file and the FSA only retries the delete operation.

RetryPeriodSeconds

Type: REG_DWORD

Data: Defaults to 300 (equivalent to five minutes). This parameter defines how long (in seconds) the FSA waits before retrying failed files in the retry queue. These files include:

- **Analysis failures:** These are files that could not be passed to the hub for processing.
- **Action failures:** These are typically files that could not be deleted. More accurately, they are files for which the FSA could not execute a control action.

AnalysisTimeoutSeconds

Type: REG_DWORD

Data: Defaults to 600 (equivalent to ten minutes). This defines how long (in seconds) the FSA waits for a file to be successfully analyzed by a policy engine. The timeout starts when a file is added to the hub input queue.

If the FSA does not receive an acknowledgment from the hub that a file has been successfully processed before this timeout expires, the FSA flags the file as a failure and writes an entry to the log.

Note that files are processed asynchronously. The hub acknowledgement may call for a file to be deleted. This deletion is performed by the FSA and is not governed by this analysis timeout.

FileOverwritePassesFixedMedia

Type: REG_DWORD

Data: Defaults to 3. This specifies the number of overwrite operations for DoD deletions of files saved on fixed storage media (that is, hard disks).

FileOverwritePassesRemovableMedia

Type: REG_DWORD

Data: Defaults to 3. This specifies the number of overwrite operations for DoD deletions of files saved on removable storage media, typically a USB portable hard drive.

ScanDatabaseDSN

Type: REG_SZ

Data: This specifies a DSN used by the FSA to connect to the Scanned File database. A default DSN is created when you install the FSA.

ScanDatabaseServer

Type: REG_SZ

Data: This specifies the name of the server hosting the Scanned File database. This registry value is set during installation.

NISTDatabaseDSN

Type: REG_SZ

Data This specifies a DSN used by the FSA to connect to the NIST database. A default DSN is created when you install the FSA.

NISTDatabaseServer

Type: REG_SZ

Data: This specifies the name of the server hosting the NIST database. This registry value is set during installation.

UseLocalPolicyEngine

Type: REG_DWORD

Data: This specifies whether scanned items are passed to a local policy engine or a local policy engine hub. This value is set automatically when you install the FSA; it is set to zero if you install a Remote Policy Engine Connector; otherwise it is set to 1.

More information:

FSA Terminology (see page 16)

Set Up CA DataMinder Policy Triggers

This section describes how the FSA uses Data At Rest triggers to analyze and apply policy to scanned items. It also describes which user policy gets applied and you can configure this policy to apply smart tags to scanned items.

Data At Rest Triggers

To apply policy to items scanned by the FSA, you need to edit the Data At Rest triggers in the user policy. Data At Rest triggers enable the FSA to:

- **Detect specific file names or formats** (for example, Microsoft Word documents).
- **Analyze an item's text content** to detect the presence or absence of key phrases, or to determine whether the item matches a particular document classification.
- **Add smart tags** to events generated by the FSA.
- **Analyze a file's attributes** such as its size, date created, date last modified, and the document author. To do this, you configure the trigger to use XML Attribute data lookup commands. See the Administration console online help for details; search for 'XML Attribute lookup'.

More information:

Apply Smart Tags to Scanned Items (see page 42)

Data At Rest Control Actions

Using Data At Rest control actions, the FSA can copy, move, delete, replace and categorize scanned items:

Copy files to an alternative location

If required, you can copy scanned files to alternative folders. When used in combination with a 'delete file' action, a copy action effectively becomes a 'move file' action. See the Administration console online help for details; search for 'scanned files: copying'.

Delete files, including DoD deletions

You can delete scanned files if, for example, policy processing determines them to be unauthorized. DoD deletions ensure deleted files cannot be recovered.

Note: DoD deletions are only supported for file system scans.

Replace deleted or moved files

You can optionally replace deleted files or files that have been moved to a new location with an explanatory stub file to alleviate any user concerns.

Categorize scanned items

If required, you can ensure that scanned items are automatically categorized. If a scanned item causes multiple triggers to fire, CA DataMinder automatically chooses the category with the highest score.

Note: For full details about Data At Rest control actions, see the *Policy Guide*; search for 'Intervention setting'.

More information:

Control Action Exceptions (see page 41)

Control Action Exceptions

DoD deletions are not supported for scanned items in Exchange Public Folders or SharePoint sites, or for scanned database records.

Also, no Data At Rest control actions can be applied to scanned database records or generic items on SharePoint sites (such as Announcements and Discussion Boards). However, if SharePoint generic items have file or document attachments, the FSA *can* apply control actions to those attachments.

More information:

SharePoint Scans (see page 10)

Which User Policy Gets Applied?

When you create a scanning job using the FSA Job Definition wizard, you must specify the policy participant. This is the CA DataMinder user account whose policy you want to apply to all scanned items.

For each scanning job, you can choose to apply the Default Policy for Files (defined in policy engines' machine policy) or you can apply a specific user's policy; to do this, you specify the user's email address. See the wizard's online help for details.

Note: You also specify the event participants when you create a scanning job. These are the users that you want to associated with the scanned items.

More information:

How Are Scanned Items Associated with CA DataMinder Users? (see page 61)

Apply Smart Tags to Scanned Items

CA DataMinder can apply smart tags to events generated by the FSA (that is, 'scan events' stored on the CMS). In the case of scanned Microsoft Office documents, it can also apply smart tags to the original document. For further details, see the Administration console online help; search for 'smart tags'.

Apply Smart tags to CA DataMinder Events Generated by the FSA

You can configure Data At Rest triggers to add smart tags to events generated by the FSA. For example, you can use smart tags to classify scanned files for data classification purposes. When the trigger activates, each smart tag is saved with the event metadata in the CMS database.

Apply Smart Tags to Original Microsoft Office Documents

You can also configure Data At Rest control actions, to apply smart tags to scanned Microsoft Office documents. In fact, you can apply smart tags either to the original document, or to a copy of the scanned document. Each smart tag is then saved as a new property of the scanned document or, if applying smart tags to scanned items in Exchange Public Folders, each smart tag is added as a MAPI property. For details, see the Administration console online help; search for: 'smart tags: file events'.

Note: You cannot apply smart tags to original items when scanning SharePoint sites or when scanning database records.

Uninstall the FSA

Important! If a policy engine is installed on the same computer as the FSA, you must uninstall the FSA before uninstalling the policy engine.

Use Add/Remove Programs to manually uninstall the FSA. This applet is part of the Control Panel.

1. In Add/Remove Programs, select CA DataMinder Integration Agents and click Change.
2. When the wizard starts, go to the Program Maintenance screen and choose Modify.

   Note: If you choose Remove, this removes all CA DataMinder components, not just the Exchange or Domino server agents.
3. In the Custom Setup screen, choose the File Scanning Agent.
4. In the final wizard screen, click Install to begin the uninstallation.

Chapter 3: Scanning Jobs

This section describes how to add, run and schedule scanning jobs. It also describes how to purge the scanned file database and how CA DataMinder users get associated with scanned items.

This section contains the following topics:

Data At Rest Scans

Data At Rest scans use the File Scanning Agent (FSA) to scan and analyze data and apply Data At Rest policy triggers. Data At Rest scans can analyze the text content of:

- Files saved in local and remote file systems.
- Microsoft Exchange Public Folder data, such as, calendar events, tasks, or journal entries.
- Microsoft SharePoint items, such as tasks or discussion boards.
- Database records.

Note: For full details about deploying the FSA, see the *Stored Data Integration Guide*; search for 'FSA'.

Manage Scanning Jobs

In the Administration console, you can perform the following tasks.

Create a scanning job

Right-click a file scanning server and click Create New Job.

Run or restart a scanning job

Right-click a scanning job and click Run Job Now.

If you restart a stopped job, CA DataMinder ignores items that have already been scanned in the current job. To do this, CA DataMinder performs hash checks to identify those files that can be excluded from the scan when it restarts. However, these hash checks can take a long time for very large scanning jobs.

Note: For each scanned item, the scan database contains a hash to uniquely identify that versin of the file plus its 'last scanned' date.

Note: You can also run scanning jobs from a command line.

Stop a scanning job

Right-click a scanning job and click Stop Job.

Reschedule a scanning job

Right-click a scanning job and click Schedule.

The Schedule Job dialog appears.

Clone or copy a scanning job

The Clone and Copy tools let you quickly create multiple scanning jobs. Right-click a scanning job and click:

Clone

Creates a new scanning job on the current file scanning server.

The new job has identical settings to the existing job, but with a different name.

Copy To

Copies the current scanning job to a different file scanning server. The new job has the same name and identical settings to the existing job.

After both Copy and Clone operations, you can customize the settings in the new job. For example, you can specify different scan locations.

Import a scanning job

Click a file scanning server and click Import Job File. You can select the XML scanning job definition that you want to import.

You may need to import scanning jobs if you have upgraded from CA DataMinder 6.0.

View scan logs

CA DataMinder maintains scan logs for each scanning job, each file scanning server, and each FSA stub, wgnfstub.exe. (The FSA Remote Connector uses an FSA stub.) To view:

Job logs

Right-click a scanning job and click View File Scanning Logfiles.

Server logs

Right-click a file scanning server and click View File Scanning Logfiles.

FSA stub logs

You cannot view these logs from the Administration console. These logs are saved to a WgnFStb_<date>.log file on the machine hosting wgnfstub.exe. These logs are saved in CA's \data\log subfolder of the Windows All Users profile.

More information:

Database Scanning Jobs (see page 54)
Scheduled Scanning Jobs (see page 49)

Create a Scanning Job

To create a scanning job

1. On the Administration console host machine, log on to Windows as the FSA Job Setup User.
2. Click Data At Rest Scans or CCS Preclassification Scans.
3. Right-click a file scanning server and click Create New Job.

   This launches the Scanning Job Definition wizard. The wizard steps you through the job configuration.
4. Enter the job settings in the wizard screens.

   The number of screens depends on the type of scanning job.
5. Click Finish in the final wizard screen.

   The Schedule Job dialog appears.
6. Specify when and how often the scanning job runs.

More information:

Database Scanning Jobs (see page 54)
Scheduled Scanning Jobs (see page 49)

Scheduled Scanning Jobs

You can schedule scanning jobs for the FSA and CFSA.

FSA scanning jobs

You define a job schedule when you create a scanning job. After you click Finish in the final wizard screen, the Schedule Job dialog appears.

To modify the schedule for an existing scanning job, right-click a scanning job and choose Schedule.

In both cases, you must specify the FSA Run As user.

Important! When a scheduled scanning job is running, it is essential that nobody logs onto the target machine using the same account as the FSA Run As user!

CFSA scanning jobs

The Client File System Agent (CFSA) can scan local files and folders. You define the CFSA scan schedule in the local machine policy.

In the Client File System Agent, Data At Rest Protection, File System Scan Configuration policy folder you can specify the start day, start time, and frequency.

More information:

Schedule Job Dialog (see page 52)

Schedule a Scanning Job

A scanning job must have a schedule defined before it can run. This allows the job to run repeatedly at regular intervals. You can also override the schedule and run a job immediately.

You typically set up a scan schedule immediately after creating a new job. You can modify this schedule subsequently.

To schedule a scan

1. Log on to the Administration console host machine as the **FSA Job Setup User**. For details about this user, account see the reference below.
2. In the Administration console, create a scanning job or modify a schedule for an existing job. See Manage Scanning Jobs (see page 46) for details.

   In both cases, the Schedule Job dialog appears.
3. Go to the **Task tab** and enter the **FSA Run As User** in the Run As field. For details about this user, account see the reference below.

   Important! When a scheduled scanning job is running, it is essential that nobody logs onto the target machine using the same account as the FSA!
4. Go to the **Schedule tab** and specify when and how often the scanning job runs.
5. (Optional) Go to the **Settings tab** and configure further settings that define when the scanning job runs.

   For example, you can stop the job if it overruns, or you can set it to only run if the target computer is idle.

More information:

FSA Job Setup User (see page 27)
FSA Run As User (see page 51)
Schedule Job Dialog (see page 52)

FSA Run As User

This is the account that a scanning job runs as. You specify the FSA Run As user when you schedule a scanning job using the Job Definition wizard or when you log on to Windows before running a scanning job from a command line.

Choosing the Run As user

Important! When a scheduled scanning job is running, it is essential that nobody logs onto the target machine using the same account as the FSA!

If a user and the FSA are logged on to the same machine, at the same time, and using the same Windows user account, this can adversely affect the scanning job. The job may terminate prematurely or it may even fail to respond to attempts to terminate it manually.

For this reason, you must be careful when choosing the Run As user for a scheduled scanning job. Specifically, if you want jobs to run as your FSA limited access user or FSA full access user, we recommend that these users are bespoke accounts created for exclusive use by the FSA. Do not choose a Run As user that corresponds to a real user account if there is any possibility that this user will be logged on to the target machine while a scheduled scanning job runs.

Avoid this situation! The classic mistake is when a network administrator schedules a scanning job and enters their own domain account as the Run As user (because they know that they have access to all the required scan locations). While performing an unrelated task, they subsequently log on to the target machine while the scan is in progress, so causing the scan to fail.

Types of Run As user

There are two types of FSA Run As user, reflecting their different purposes:

'Limited Access' FSA Run As User

You can use the FSA to test whether sensitive documents stored on your network are accessible to unauthorized users. To do this, you set up a scanning job to run as a user with limited access to sensitive network locations. This is your 'limited access FSA Run As user'. In effect, the limited access user is a proxy for your ordinary network users.

'Full Access' FSA Run As User

You can use the FSA to scan the content of files stored on your network to determine if sensitive information is stored in the correct location or to identify files or documents with unauthorized content. (Indeed, this is often regarded as the 'classic' use for the FSA.) To do this, the FSA must be able to access remote file systems and delete files when instructed to do so as a result of policy processing. Specifically, the FSA must run as a domain user with permission to delete and copy files on any machines that you want to scan. This user is your 'full access FSA Run As user'.

Requirements for Run As user

If you want to scan:

Microsoft Exchange Public Folders

When scanning Exchange public folders, the FSA Run As user requires:

- An Exchange mailbox. You can typically create a mailbox at the same time as you create a new user account in Active Directory.
- Read access to Exchange Public Folders. However, the full access user needs Read and Write access to be able to copy or delete Public Folder items.

The FSA also requires a default e-mail application compatible with Microsoft Exchange, such as Microsoft Outlook.

Microsoft SharePoint

When scanning a SharePoint site, the FSA Run As user must be a domain user with:

- Local administrator rights on the SharePoint host machine.
- Read access to the SharePoint site being scanned. However, the full access user will need Read and Write access.

For full details, see the *Stored Data Integration Guide*; search for 'FSA user accounts'.

Schedule Job Dialog

Use this dialog to define a scanning job schedule.

Task tab

The Run and Start In fields are populated automatically; you do not need to edit these fields.

In the Run As field, enter the name and password for the FSA Run As User. But note the requirements for this user.

Important! When a scheduled scanning job is running, it is essential that nobody logs onto the target machine using the same account as the FSA Run As user! If a user and the FSA are logged on to the same machine, at the same time, and using the same Windows user account, this can adversely affect the scanning job. The job may terminate prematurely or it may even fail to respond to attempts to terminate it manually.

Schedule tab

Specify when and how often the scanning job runs.

Settings tab

This tab includes optional settings to further configure when the scanning job runs. For example, you can stop the job if it overruns, or you can set it to only run if the target computer is idle.

More information:

FSA Run As User (see page 51)

Command Line Scanning Jobs

You can run FSA scanning jobs from a command line. First, you create a job using the scanning job definition wizard. Then you reference this job in an FSA command.

In practice, you only run scanning jobs from a command line when scanning a database. This is for security reasons. The command line method allows you to avoid storing database logon credentials in the job definition.

Example command

The command syntax is shown below. This command runs a database scan and prompts you for the logon credentials that the FSA will use to access the target database objects:

```
wgnfstub /job:<job name> /user /password
```

Where:

<job name>

Is the name of an existing scanning job. You specify job names in step 1 of the job definition wizard.

/user and /password

(Optional) The purpose of these parameters is to avoid the security risk of storing database credentials in a scanning job definition.

If these parameters are used, the FSA prompts you for the logon credentials that it will use to access the target database objects.

More information:

Database Scanning Jobs (see page 54)

Database Scanning Jobs

You can set up FSA jobs to scan databases. Specifically, the FSA can scan columns in database tables that contain text or binary data. Binary data refers to documents (such as MS Word files).

What are database events?

When the FSA scans a database table, the scanned data is written to XML blob files. The FSA then passes these blob files to policy engines for processing.

Policy engines then apply Data At Rest triggers to the scanned data. If a trigger fires, a database event is generated and stored on the CMS. You can view these events in the iConsole; they are indicated by  icons in the Search Results screen.

How does the FSA create database events?

By default, all scanned data in the table (each column for every row) is written to a single XML blob file and stored as a single event. However, for very large database tables, this is clearly undesirable. Instead, you can configure the FSA to slice the scanned data into smaller, more easily manageable events.

For example, you can specify a maximum number of rows per event, or a maximum size (in KB) per event. You can also configure special handling for binary data; for example, if a database table contains MS Word documents, you can specify that these documents are stored as attachments to database events.

How does the FSA connect to a database?

When you create a scanning job, the FSA generates a database connection string, based on the detail that you supply in the Job Definition wizard.

Important: If scanning Oracle databases, be aware that Oracle user, schema and table names are case-sensitive when you specify the connection string!

What data is captured?

When setting up a database scanning job, you can specify how binary data is handled.

Similarly, when you set up user policies, the Capture File Details? setting in each Data At Rest capture action determines what data is captured and stored on the CMS.

More information:

Binary Data Handling (see page 55)

Binary Data Handling

When setting up a database scanning job, you can specify special handling for binary data. For example, if a database table contains MS Word documents, you can specify that any documents found are stored as attachments to the resulting database event.

Binary data options

The following binary data options affect the events created by the FSA:

- Store binary data as event attachments = True or False
- Only store database rows with columns containing binary data = True or False
- Store rows containing binary data as individual events = True or False

Example Database Table

In this example table, column C contains binary data (in this case, Microsoft Word documents of varying size):

	Col.A	Col.B	Col.C	Col.D
Row 1	data	data		data
Row 2	data	data		data
Row 3	data	data		data
Row 4	data	data	X	data
Row 5	data	data		data
Row 6	data	data	X	data

Store Binary Data as Event Attachments = True

If you store binary data on the CMS as event attachments, the binary data options affect the events created by the FSA in the example database table as follows:

Only store database rows with columns containing binary data = True

Only rows with binary data are stored.

Store rows containing binary data as individual events = True

Only rows with binary data are stored and they are stored as individual events. The FSA generates four events:
Event1 contains data for row 1; plus attachment
Event2 contains data for row 2; plus attachment
Event3 contains data for row 3; plus attachment
Event4 contains data for row 5; plus attachment

Store rows containing binary data as individual events = False

Only rows with binary data are stored, but they are not stored as individual events. The FSA generates one event:
Event1 contains data for rows 1, 2, 3, and 5; plus four attachments

Only store database rows with columns containing binary data = False

All rows are stored.

Store rows containing binary data as individual events = True

All rows are stored, but rows with binary data are stored as individual events. The FSA generates five events:
Event1 contains data for rows 4 and 6
Event2 contains data for row 1; plus attachment
Event3 contains data for row 2; plus attachment
Event4 contains data for row 3; plus attachment
Event5 contains data for row 5; plus attachment

Store rows containing binary data as individual events = False

All rows are stored, but rows containing binary data are not stored as individual events. The FSA generates one event:
Event1 contains data for rows 1 through 6; plus four attachments

Store Binary Data as Event Attachments = False

If you do not store binary data on the CMS as event attachments, the binary data options affect the events created by the FSA in the example database table as follows:

Only store database rows with columns containing binary data = True

Only rows with binary data are stored.

Store rows containing binary data as individual events = True

Only rows with binary data are stored and they are stored as individual events. The FSA generates four events:

Event1 contains data for row 1; no attachment

Event2 contains data for row 2; no attachment

Event3 contains data for row 3; no attachment

Event4 contains data for row 5; no attachment

Store rows containing binary data as individual events = False

Only rows with binary data are stored, but they are not stored as individual events. The FSA generates one event:

Event1 contains data for rows 1, 2, 3, and 5; no attachments

Only store database rows with columns containing binary data = False

All rows are stored.

Store rows containing binary data as individual events = True

All rows are stored, but rows containing binary data are stored as individual events. The FSA generates six events:

Event1 contains data for row 1; no attachment

Event2 contains data for row 2; no attachment

Event3 contains data for row 3; no attachment

Event4 contains data for row 4

Event5 contains data for row 5; no attachment

Event6 contains data for row 6

Store rows containing binary data as individual events = False

All rows are stored, but rows containing binary data are not stored as individual events. The FSA generates one event:

Event1 contains data for rows 1 through 6; no attachments

Import a Version 6.0 Scanning Job

For CA DataMinder r12.0, the underlying XML schema for scanning job definition files has been amended. If you have scanning jobs defined using version 6.0 of the CA DataMinder product (previously called Orchestria APM), you **must** first import these jobs onto the FSA before you can run them. Use the Administration console to do this.

When you import 6.0 scanning jobs, the FSA automatically updates them to use the new r12.0 job schema. You can then run these jobs as normal.

To import 6.0 job files

1. Ensure you have upgraded your FSA to r12.0. You must also install a r12.0 Administration console.
2. Log on to the Administration console host machine as the FSA job setup user.
3. In the Administration console, expand the File Scanning Agents branch.
4. Select an FSA server and choose Edit, Import Job File.
5. This launches the Import FSA Job Definitions dialog. Browse to the XML job file that you want to import.
6. Each individual job defined in the 6.0 job file is added to the FSA screen. Note that the FSA creates a separate new job file for each individual job extracted from the 6.0 job file.

More information:

FSA Job Setup User (see page 27)

Scanning Job Log Files

About FSA logs

The FSA logs the outcome of scanning jobs, such as details of replaced files, connections to the scan database, and when jobs started and completed. You define the log level in the registry.

Most FSA logs can be viewed in the Administration console.

Viewing scan logs

CA DataMinder maintains scan logs for each scanning job, each file scanning server, and each FSA stub, wgnfstub.exe. (The FSA Remote Connector uses an FSA stub.) To view:

Job logs

Right-click a scanning job and click View File Scanning Logfiles.

Server logs

Right-click a file scanning server and click View File Scanning Logfiles.

FSA stub logs

You cannot view these logs from the Administration console. These logs are saved to a WgnFStb_<date>.log file on the machine hosting wgnfstub.exe. These logs are saved in CA's \data\log subfolder of the Windows All Users profile.

More information:

Configure the FSA (see page 36)

Purge the Scanned File Database

The FSA uses a scanned file database (also known as the 'scan database') to track the status of each file in a scanning job.

You can purge the scanned file database of all file records. Or you can purge the database of file records associated with a specific scanning job.

You can purge the scanned file database from the Administration console or from a command line.

Adminstration Console Purges

To purge the scan database completely

Right-click a file scanning server and click Purge Scan Database.

The FSA purges *all* records from the scan database on the specified server.

To purge a scnning job

Right-click a scanning job and choose Purge Scan Database.

The FSA only purges records for the selected scanning job.

Command Line Purges

Use the following command syntax to purge:

- The entire scanned file database:

  ```
  <path>\wgnfstub /purge
  ```

- A specific scanning job:

  ```
  <path>\wgnfstub /purge <job name>
  ```

  Where <job name> specifies a specific job. If the job name contains spaces, use quotes:
  ```
  <path>\wgnfstub /purge "Asia Sales"
  ```

More information:

FSA Terminology (see page 16)

Scanning Job FAQs

How Are Scanned Items Associated with CA DataMinder Users?

When you create a scanning job using the FSA Job Definition wizard, you must enter the email address of any CA DataMinder user that you want to associate with the files or items being scanned.

For each event, the FSA assigns the event participants specified in the scanning job definition.

Then, when a scan runs and events are generated, these addresses get stored as event participants in the CMS database. When a reviewer subsequently searches for scanning events in the iConsole or Data Management console, these addresses are mapped onto CA DataMinder user accounts.

Note: For details about what happens if no event participants are specified, see the next section. Further information about mapping file events to CA DataMinder users is available in the 'Event Participants' technical note, available from CA Technical Support.

How Do I Delete a Scanned File Database?

A scanned file database can only be deleted manually, by configuring the machine it is hosted on:

1. Open the relevant application, according to the type of database you are running. For example, use SQL Server Management Studio for SQL Server 2005.
2. In the Object Explorer, select the Database Engine that contains the database you want to delete, and connect to it.
3. Browse to the actual database object location within the Databases folder structure.
4. Right-click on the database name and select delete.

   Note: We recommend you accept the default setting to delete the database backup.

Do I Use Multiple Scanning Jobs or Multiple FSAs?

The FSA is highly flexible. You can have a single FSA running many scanning jobs or multiple FSAs each running a few jobs. Which deployment is appropriate for you depends on the volume of data you need to scan, your network configuration, available hardware and so on. However, there are some practical considerations that can inform your decision:

- A single FSA running multiple scanning jobs is easier to deploy. But all scanned files that need policy processing are channeled through the same hub. Hub capacity is therefore a key determinant.
- Multiple FSAs permit load balancing and can share a SQL Server database server. Each FSA can use a separate database hosted in the same instance of a SQL Server database engine. Or each FSA can use a separate instance of SQL Server, with all instances hosted on the same machine. For further guidance, see the *Database Guide*; search for 'shared database'.
- If you deploy multiple FSAs, each running local scans, then all scanned files are automatically associated with the source machine when the file event is stored in the CMS database.

More information:

What Happens If No Event Participant Is Specified? (see page 63)

Can Scanning Jobs Overlap?

Yes. You can have two overlapping scanning jobs. That is, both are configured to scan a common folder. This allows you apply different policies to same folder. For example, one scanning job may apply DoD deletions to files matching a particular classification, while another job saves copies of specific files to a new location.

What Happens If No Event Participant Is Specified?

Typically, you specify one or more event participants when you run the job definition wizard in the Administration console. Specify the email addresses for these participants and CA DataMinder maps those addresses to CA DataMinder user accounts.

However, you do not need to associate scanned files with a CA DataMinder user. This is because each scanned file is automatically associated with the machine hosting the source folder.

Specifically, an address matching the machine's domain name in Active Directory is associated with each scanned file and stored in the CMS database. This machine 'address' takes the form:

```
cn=<computer name>,cn=computers
```

For example:

```
cn=UX-MILAN-W2K3,cn=computers
```

Even if the job file does not specify a file event participant, each scanned file is associated with a 'host machine' address. In this situation (based on the example above), to ensure that files scanned on machine UX-MILAN-W2K3 can be retrieved during an iConsole event search, you must add the above machine address to the list of addresses specified for an appropriate CA DataMinder user account. Add new addresses in the User Properties dialog in the Administration console.

Database Scanner Performance Monitor Counts

Due to the way rows are packaged into events, the Items Scanned perfmon counter cannot be used to predict the number of rows imported.

Index