Enterprise Vault Best Practices

Implementing File System Archiving

This document contains information on best practices when implementing Enterprise Vault File System Archives (FSA)

If you have any feedback or questions about this document please email them to **EV-TFE-Feedback@symantec.com** stating the document title.

This document applies to the following version(s) of Enterprise Vault: 9 & 10

This document is provided for informational purposes only. All warranties relating to the information in this

.

Document Control

Contributors

Who	Contribution
Darren Locke	Content
Evan Barrett	Content & author

Revision History

Version	Date	Changes
1.0	November 2011	Initial release

Related Documents

Version	Date	Title

Table of Contents

Introduction

This document will cover best practices for implementing successful Enterprise Vault File System Archiving (FSA) and will include hint and tips. This document is not an introduction to FSA and assumes that the reader will have prior knowledge of Enterprise Vault and FSA.

Infrastructure Basics

Operating System and Hardware

Starting with Enterprise Vault 10, only Windows 2008 R2 and later will be supported on the Enterprise Vault server itself. The hardware requirements for Enterprise Vault 10 have also changed when compared to previous versions. For absolute hardware minimums, there must be at least four CPU cores and 8GB RAM in the Enterprise Vault server.

For best indexing performance, it is recommended to increase the RAM to 16GB or higher. A benefit of increased memory is that an Enterprise Vault server can keep more index data in memory which can reduce the amount of disk I/O during searches.

It is highly recommended that Index Volumes be placed on high speed disk such as a storage area network (SAN) or direct attached storage (DAS). This will reduce the amount of time required when searching archived file content due to the fact that Enterprise Vault will be able to obtain data from disk more rapidly.

The temp location for the Vault Service Account (VSA) should also be placed on high speed disk. This will increase the archiving throughput. The size of the disk volume should be large enough to handle the largest files in the environment that will be archived. Any file larger than 50MB in size will be temporarily transferred to this area for fingerprinting (a file less than 50MB in size will be fingerprinted in place on the file server).Thus, if the environment has 5 GB files that will be archived, more than 5GB of disk space will be needed for the VSA temp location.

By default the VSA temp area will be location on the operating system drive (normally C:). As well as ensuring this is on high speed disk, it is recommended to move the VSA temp area away from the operating system drive.

Pass Through Recall

If Pass Through Recall will be implemented, it is highly recommended that the cache location for this feature be placed on high speed disk on the file server target and away from the operating system disk. This will increase the recall speed for when archived files are accessed.

Domain Name Service (DNS)

It is necessary to ensure that all Enterprise Vault and all target file servers are correctly registered in DNS. A properly configured DNS environment will reduce the amount of errors when setting up FSA. DNS reverse lookup zones must also be configured with all file server targets registered if FSA Reporting will be used.

Indexing

Enterprise Vault offers the ability to defer indexing with FSA, but this is not recommended for a couple of reasons. Some FSA tools, such as FSAUtility, require indexes in order to function. For these to function, it is recommended to enable indexing with at least a 'brief' level. Brief indexing will index all file meta-data, but none of the contents.

If search (using Browser Search or Discovery Accelerator) and/or the use of Archive Explorer are required, it is recommended to enable 'full' indexing. Full indexing will index all the file meta-data as well as the file contents. Even if search or Archive Explorer is not being deployed to your end-users, it is beneficial to enable so that IT administration staff can perform searches and self-service file recovery on behalf of the install base.

The index storage requirements for FSA are less when compared to other content (such as Exchange) that can be archived by Enterprise Vault. For FSA, a Brief level for indexing will typically require around 2% of the original file size and a Full level will typically require around 6% of the original file size.

Archive Points

Archive Points are markers that are placed on the file system and identify the start of an archive for Enterprise Vault. All folders and files below an archive point will go into the same archive.

It is highly recommended that a volume have more than one archive point. A single archive point on a large file volume can result in a very large index which will increase the time needed to perform searches and will compromise the end-user experience with Archive Explorer. Recommended archive locations include (but not limited to):

- At the root of each user's home folder
- At the root of a department or project folder
- At the root of any common or shared folder

Use auto-enabled archive points for home folders. This will automatically create archives when new user home directories are created.

Enterprise Vault also includes a command line utility, ArchivePoints.exe, which allows the FSA administrator to manually perform Archive Point functions. This utility can be used for the following:

- Naming convention
- Setting the indexing level
- Create archive points
- Delete archive points
- Set AutoEnable on folders

ArchivePoints.exe can also use an XML template file to assist with defining Archive Point properties. For more information, refer to the Enterprise Vault Utilities guide.

Permissions on the File Target

Folder and File Permissions

Enterprise Vault will keep track of permissions at the folder level and not for individual objects such as files when archiving from file server targets. Files whose permissions are different than that of the parent folder will have the option of either being archived with the parent folder permissions or not archived at all. This option can be set in the Permissions tab of the FSA Volume or Folder policy.

To determine if files have different permission than that of the parent folder, run the FSA Task against the desired target volume in Report Mode (with verbose logging enabled). The report, which will be stored in the \Enterprise Vault\Reports\FSA\<name_of_FSA_task>, will identify any files that have explicit permissions.

If files with explicit permissions are archived, then the permissions on the resulting placeholder will still be the explicit permissions. However, the permissions of the file in the archive will be the permissions from the parent folder. This means that someone who does not have access to the file on the file system may have access to the file within the archive.

By default, FSA will synchronize the permissions at the share level. If the folder structure being archived contains permissions which differ to the share, then it is recommend to synchronize the NTFS permissions. A registry change is required on the Enterprise Vault server to enable this:

- HKLM\SOFTWARE\KVS\Enterprise Vault\
    - DWORD value: SynchroniseFSASharePermissions
    - Value data: 0

Vault Service Account (VSA) Permission on the File Server Target

It is recommended that the VSA have local administrator rights on a Windows file server target. However, there are situations where the VSA cannot be granted these rights. In these particular situations, it is recommended that the following be configured for the VSA on the Windows file server target:

- Add the VSA to the Power Users Group
- VSA either added to the backup operators group or granted backup privileges
- Grant the VSA Remote Launch and Remote Activation DCOM permissions
- Authorize the VSA account in WMI Control with the Remote Enable permission at the root level (see **http://technet.microsoft.com/en-us/library/cc787533%28WS.10%29.aspx**) and propagate to all subfolders
- Ensure that the VSA has Full Control on the share being targeted

Creating FSA Archiving Policies

Know the Content before Archiving

Before implementing FSA, it is a good idea to know the nature of the files that reside on the target file servers. Implementing a policy that will essentially "archive everything" (e.g.: using *.* as the file filter in the policy) is not recommended as files may be archived that are not great candidates for archival. A utility from Symantec, File System Analyzer, is available that will scan file servers (Windows and CIFS-enabled filers) and collect metadata on all discovered files. This metadata includes file type, file size, dates (creation, access, modified), as well as file attributes. The output from this utility can be used to design proper FSA Volume and Folder policies.

Archiving "Everything" Policies

While FSA can archive any file type, there are many files in an environment that should not be archived such as system files, application files (such as .exe and .dll files), as well as mail files such as PST or NSF (recommend to eliminate with the email archive offerings of Enterprise Vault). Implementing an "archive everything" policy may result in numerous retrieval requests from the Enterprise Vault server. This can lead to applications poorly performing and ultimately a request to the IT team to recall certain files.

Starting with Enterprise Vault 10, newly created FSA policies will include two rules to not archive common Windows and Mac files which typically should not be archived. These pre-defined file groups, Windows Files and Mac Files, can be modified in the Enterprise Vault Administration Console (VAC).

There are times when using an archive “everything” rule makes sense. Use cases for this include:

- File shares that only contain specific application data
- A copy of files destined for preservations purposes where shortcuts or placeholders are not required
- If there are other “do not archive” rules higher in the hierarchy of an archiving policy

Deleting Files

FSA has the ability to delete unneeded files during archiving that can take up a lot of space but offer little or no practical, compliance or discovery benefit. These files typically include temporary files created by an application or the operating system that were not deleted. Enterprise Vault includes a File Group entitled “File Types to Remove” and include the following file types by default (this group can be edited as needed):

- *.bak
- *.chk
- *.gid
- *.log
- *.old
- *.tmp
- ~*.*

FSA administrators may also find it necessary to delete other file types in order to comply with company or government regulations. An example of this may be to delete all MP3 files as there may be a policy to not store these types of files on company hardware resources.

It should be noted that the deletion action only looks at the file extension when determining what to delete. So if an MP3 file had been renamed as a TXT file it would not be deleted.

File blocking can be used to prevent unwanted file types from being saved to the file server in the first place. File blocking can block based on the file extension, or optionally it can look at the file signature which is contained within the file. This is a slower process but will allow detection of files that have been disguised as something else, such as the MP3 file masquerading as a TXT file. File blocking is currently available for Windows file servers and NetApp filers.

Small Files

Unless there is a need to archive a small file for compliance purposes, there is no benefit to archiving small files. A placeholder will take up one disk cluster. On Windows systems, a disk cluster is typically

4K, but can vary up to 64KB. Thus, archiving a file that is 4KB or less will not result in any disk space savings.

Any file archived by Enterprise Vault will consume around 10KB of space on the Enterprise Vault server itself. This space is divided between SQL, index, and saveset. Generally, files that are less than ~15KB will not save any disk space in the environment overall unless there are many duplicates of the file. An exclusion rule for small files should be placed in the top of the archiving policy.

Large Files

Enterprise Vault has the ability to archive files of any size. However, it should be noted that large files will take time when being retrieved from Enterprise Vault. End users should be educated on this fact in order to set expectations. Pass Through Recall (when applicable) can shorten the retrieval time of certain files (such as video or .zip files) as only the accessed chucks of data will be restored instead of the whole file.

If space reduction on file servers is the main goal of implementing FSA, it is recommended that larger files be targeted initially for archiving. This will show the greatest space savings on the file server. Once these larger files have been archived, the archiving rules can be changed to allow smaller files to be archived.

Dates of Files

Many FSA administrators typically like to use last accessed date when creating archiving rules. However, the last accessed date can be reset by poorly written applications that scan the file system. This may result in no files being archived. If the last accessed date cannot be relied upon, it is recommended that the last modified date be used instead.

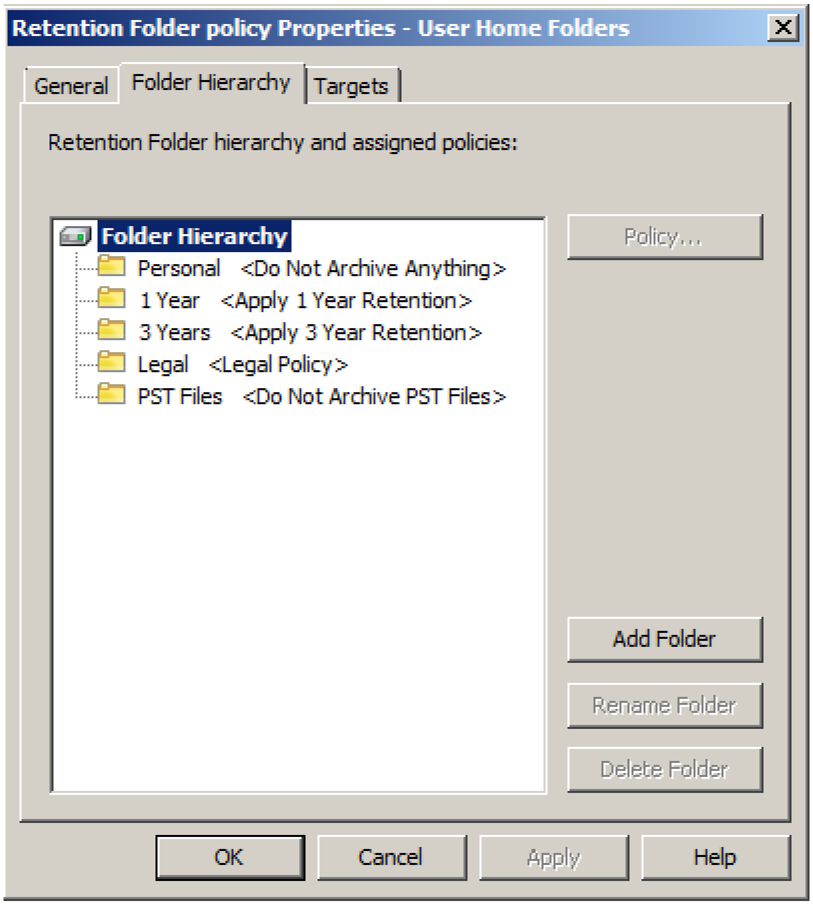
When using last accessed date in the archive policy, it should be noted that any recalled file may not be archived again until the date criteria has been met. For example, if the policy contains a rule to only archive files that were last accessed over 90 days ago, the recalled file will not be re-stubbed until 90 days after the last access date. In situations like this, using the last modified date instead could result in the file being re-stubbed almost immediately (during the next FSA archival run) if the file was not modified.

For best practices, it is recommended that a non-aggressive policy be used initially by only archiving older files (such files that are two or three years old). Once these older files have been archived, the age of the files can be reduced (e.g. two years to one year, one year to six months, etc.). This type of policy will provide the least impact to end users while the system is introduced and any while any implementation issues are resolved. It will also allow the FSA administrator to tune the policy based on the actual space reclaimed.

Retention Folders

Retention Folders allow the FSA administrator to deploy a pre-defined folder structure to users' home areas or to shared areas for archiving. Each defined folder can have a unique archiving policy along with a differing retention period if desired. An often overlooked use of retention folders is to apply a 'do not archive' policy to a common set of folders. Some use cases for Retention Folders include:

- User home directories – Create a predefined folder structure (as shown in the following screenshot). The example shows five separate folders each with their own archiving policy assigned. Two of the folders are associated with policies which would not archive files.



Retention Folder policy Properties - User Home Folders

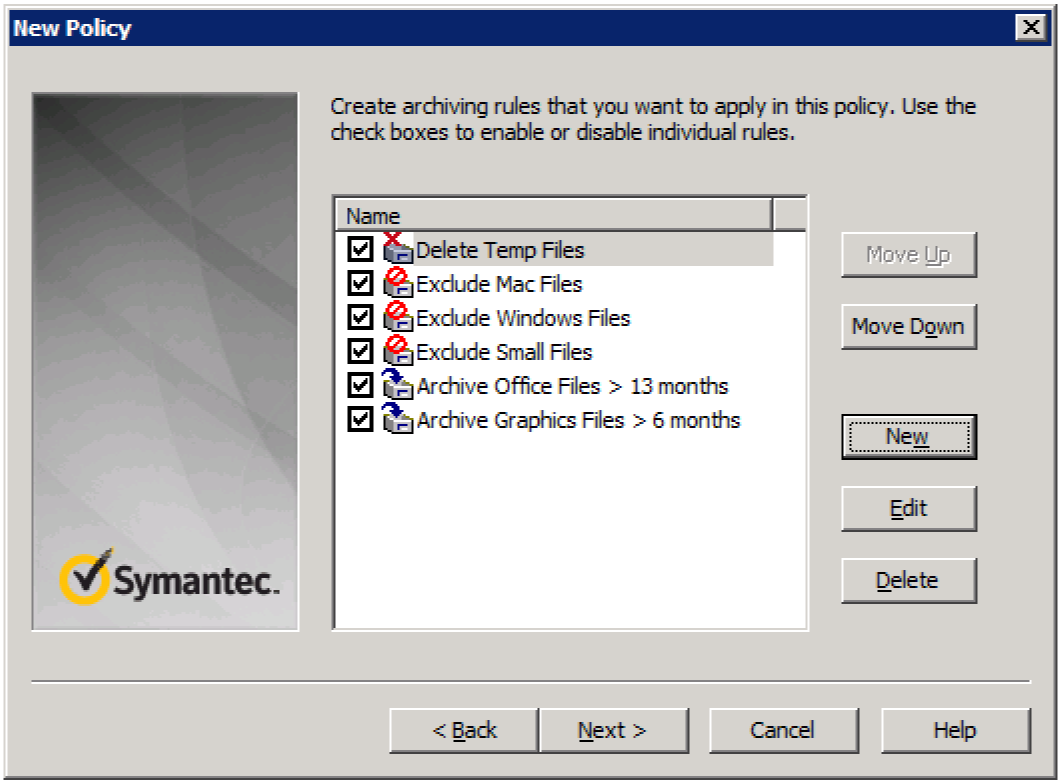
Retention Folder hierarchy and assigned policies:

Folder Hierarchy
- Personal <Do Not Archive Anything>
- 1 Year <Apply 1 Year Retention>
- 3 Years <Apply 3 Year Retention>
- Legal <Legal Policy>
- PST Files <Do Not Archive PST Files>

- Group or project folders – Create pre-defined folder structures that would contain different aspects of a project such as project plans, meeting notes, code, etc., which could all have unique policies assigned.

So What Does a Good File Policy Look Like?

The following screenshot illustrates many of the topics that have been previously discussed. File exclusions rules are set higher in the hierarchy of the policy while the actual archiving rules are towards the bottom.



Create archiving rules that you want to apply in this policy. Use the check boxes to enable or disable individual rules.

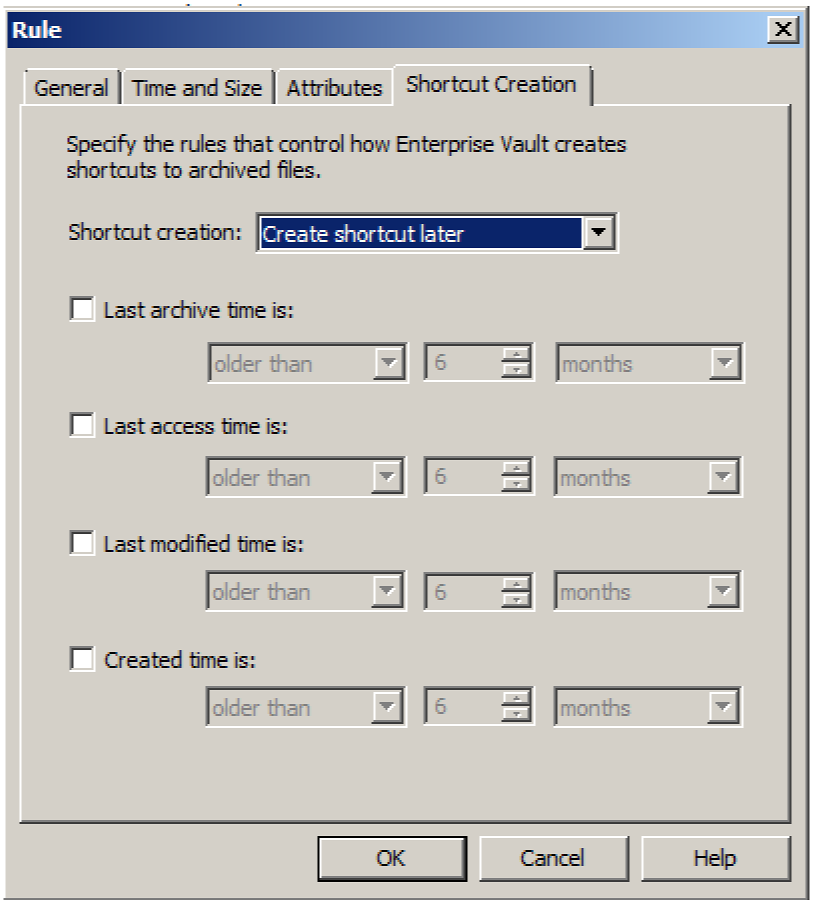
Name

- ☑ Delete Temp Files
- ☑ Exclude Mac Files
- ☑ Exclude Windows Files
- ☑ Exclude Small Files
- ☑ Archive Office Files > 13 months
- ☑ Archive Graphics Files > 6 months

Archiving for eDiscovery

Archive Now, Shortcut Later

When using FSA to archive files for compliance or discovery, a copy of all files on the file system or share can be preserved with Enterprise Vault. The "Archive Now, Shortcut Later" feature allows files to be archived immediately. Shortcuts do not necessarily need to be created at the time of archiving. This option is defined in the "Shortcut Creation" tab of rule's property window (as shown in the following screenshot). Not selecting a date for shortcut creation would ensure that files are archived, but that a placeholder would never be created for the archived files. It is recommended that the rule should archive all files (*.*) when using FSA for compliance or discovery purposes.



Specify the rules that control how Enterprise Vault creates shortcuts to archived files.

Shortcut creation: Create shortcut later

□ Last archive time is: older than 6 months

□ Last access time is: older than 6 months

□ Last modified time is: older than 6 months

□ Created time is: older than 6 months

FSA Filter Points

Starting with Enterprise Vault 10, a new feature entitled “FSA Filter Point” allows files to be passed to an external application for advanced archival, retention, and indexing options. Some of the use cases for the filter point are:

- Content-based classifications (assists with identifying business records)
- Voice-to-text conversion of audio files
- Optical Character Recognition (OCR) of faxes or scanned documents

These external applications for FSA Filter Point are developed by Symantec partners but not Symantec itself.

When to Use Pass Through Recall

Pass Through Recall, a feature introduced with Enterprise Vault 8.0SP1, allows files that are being opened for read-only purposes, to be sent directly to the recalling application or workstation, without having to be restored to the primary file server volume. There are many benefits with this feature including the following:

- Read-only file systems – Placeholders located on a read-only file system (such as a Windows VSS or NetApp / Celerra snapshot) can be recalled
- The FSA administrator desires that archived files not be recalled directly back to the original file system. Some points to note:
    - Any file being opened for modification (whether or not changes would be made) will still be recalled to the file server (such as using Microsoft Excel)
    - The recall behavior is determined by the recalling application and not FSA
    - Great for static types of content such as graphics, videos, and PDF files.
- Quicker access to certain file types such as video or audio files (as only the accessed portions of the data will be restored instead of the whole file).

It should be noted that Pass Through Recall can result in slower recalls compared to full recalls in certain situations.

Performance Tuning

There are several methods available to tune the performance of FSA. The first option is to move the VSA "TEMP" area to an array of multiple spindles. This will increase the overall disk performance and reduce the amount of time needed to index files and process large files.

Adjust the number of "worker" threads available to the FSA archiving tasks. This can be accomplished by modifying the NumberOfThreads column in the Task table of the EnterpriseVaultDirectory database on the SQL server. The default setting is 5. Gradually increase this thread count while monitoring system usage. Generally speaking, a setting over 20 sees no additional benefit. Before modifying anything in SQL, ensure a good backup is available just in case anything untoward occurs.

When possible, schedule archiving tasks during quiet hours. During quiet hours, networks and file servers are generally used less frequently which provides more through put for archiving. Ensure that archiving tasks are not scheduled at the same time as backup windows.

As archiving only places a small load on the actual file server, it may be possible to archive through the day. This will give more archiving hours which may be needed at the start of an implementation when there is a large backlog of files to archive. If archiving through the day it would be recommended to reduce the number of “worker” threads so that the Enterprise Vault server has plenty of capacity to process recall and search requests as well as to perform the archiving.

Moving and Migrating Placeholders

During file server moves or consolidations, avoid directly copying placeholders from one server to another. During the copy process, placeholders will be recalled resulting in slower throughput and initiating a mass recall situation which could fill up the source file server.

To overcome the recall issue, use FSAUtility to move placeholders with one of the following options:

- -pm – Used when moving whole archives when the source and destination servers are being processed by the same Enterprise Vault server (new with Enterprise Vault 9).
- -m – Used for partial archive moves or when the source and destination servers are being processed by different Enterprise Vault servers

Keep in mind that FSAUtility will only move placeholders. To move non-archived files, use robocopy with the /XA:O flag to only move non-archived files. If possible, use FSAUtility first to move placeholders before using robocopy.

FSAUtility can also be used to migrate placeholders between different platforms. For example, an FSA administrator may have a project where placeholders need to be migrated between Windows and a NetApp filer. Placeholders cannot be directly copied from one platform to another. FSAUtility will ensure the placeholders are converted to correct format when used in this type of migration.

Recalling Archived Files

There are several methods available to recall archived files back to the file system:

- FSAUtility –b – The –b option can be used to recall all placeholders below a given folder path
    - What to recall is determined by *walking the folder structure*
    - Any placeholder encountered will be recalled
    - New with Enterprise Vault 9
- FSAUtility –t – The –t option can be used to restore all placeholders below a given folder path
    - What to restore is determined by *looking in the Enterprise Vault databases*
    - Files are restored to their original location (even if the placeholder was moved to another subdirectory)
    - Files are restored even if a placeholder no longer exists
    - This option is typically used in disaster recovery situations
- Archive Explorer can also be used to restore files and allows end users to perform self-service file recovery

Preventing Mass Recall

Often times badly behaving applications that scan the file system may cause mass recalls of archived content. To prevent mass recalls, there are two options that are available:

- Executables can be blocked from triggering recalls by modifying the registry on the file server target (Windows file server targets only):
    - HKLM\SOFTWARE\KVS\Enterprise Vault\FSA\PlaceholderService
    - String value: ExcludedExes
    - Value data: <semi-colon separated list of executables>, e.g.: notepad.exe;winword.exe
- Service or user accounts can be blocked from triggering recalls on Windows, NetApp, and Celerra devices
    - For Windows and NetApp, use EvFsaBackupMode.exe
    - For Celerra, use Celerra fs_dhsm backup mode
    - More details are available in the FSA Installation and Configuration guide

FSA Reporting

FSA Reporting offers many benefits when used before and after archiving. Before archiving has been implemented, FSA Reporting can analyze potential archiving targets to provide a breakdown of file types, ages, and size. Once archiving has been implemented, it can be used to show the impact of archiving and possibly for chargeback.

When reviewing reports generated, specific file types will only be explicitly summarized if they are part of a file group. Ensure that your file groups are configured correctly within Enterprise Vault prior to initiating reporting.

If you have a remote NetApp or Celerra device, utilize a remote reporting proxy that is on the same local area network as the NAS device for quicker scans and collections. This new feature was introduced with Enterprise Vault 9.

If five or more simultaneous reporting scans are scheduled, consider using multiple reporting databases. The multiple FSA Reporting database feature was introduced with Enterprise Vault 9 to increase reporting performance. Limiting the number of simultaneous scans by using staggering scan schedules can negate the need for additional reporting databases.

About Symantec:

Symantec is a global leader in providing storage, security and systems management solutions to help consumers and organizations secure and manage their information-driven world.

Our software and services protect against more risks at more points, more completely and efficiently, enabling confidence wherever information is used or stored.

For specific country offices and contact numbers, please visit our Web site: **www.symantec.com**

Symantec Corporation
World Headquarters
350 Ellis Street
Mountain View, CA 94043 USA
+1 (650) 527 8000
+1 (800) 721 3934