# NetBackup and Enterprise Vault Integration

Version 1.6 June 5, 2007

## Introduction

With ever growing volumes of data in typical organizations and increasing scrutiny around management and retention of that data, companies are beginning to seek more cost-effective ways to scale their storage environments.  In particular, with email and file server growth, IT administrators are struggling with the challenges of providing high levels of service to users while staying within realistic IT budgets.

Thousands of customers have implemented Enterprise Vault since 1999 for archiving of Microsoft Exchange email, Microsoft Windows or Network Appliance file systems, Microsoft SharePoint documents, instant messages and other content.  Enterprise Vault helped address customer challenges in this area by allowing IT organizations to automatically migrate data from primary disk storage locations to more cost-effective secondary disk storage locations, such as serial ATA environments.

Enterprise Vault customers are able to preserve a seamless user experience (so that end-users do not have to change their behavior to access archived data), while ensuring data reduction on the backend through compression and single instance storage.  IT organizations have been able to better adhere to internal and external policies around retention and destruction of this data, while providing tools for IT, legal, and end-users to securely perform searches against archived content – for information access, legal discovery and communication supervision.

However, not all customers have a need to keep archived data on disk – even if it is near-line disk such as serial ATA.  Indeed, many customers have archived data that is expected to be accessed very infrequently, and tape and other traditional backup media present a compelling and cost-effective alternative for long-term archive data storage.  Customers are observing that as information moves through its lifecycle of use, its access frequency and performance requirements diminish.

At the same time, customers have already typically invested in a large data protection infrastructure managed by VERITAS NetBackup, including backup media servers, libraries, tape drives, media and management processes.  These customers are often reluctant to setup a completely separate environment for managing archived media.

With the release of VERITAS Enterprise Vault 6.0 in July of 2005, VERITAS customers can now have the best of both worlds and meet their business data management needs in a more cost-effective fashion.  Through integration between VERITAS NetBackup and VERITAS Enterprise Vault, customers are now able to define automatic, policy-based migration strategies to move archived data from disk managed by Enterprise Vault to tape or other media managed by NetBackup – leveraging the same existing backup infrastructure customers already have today.
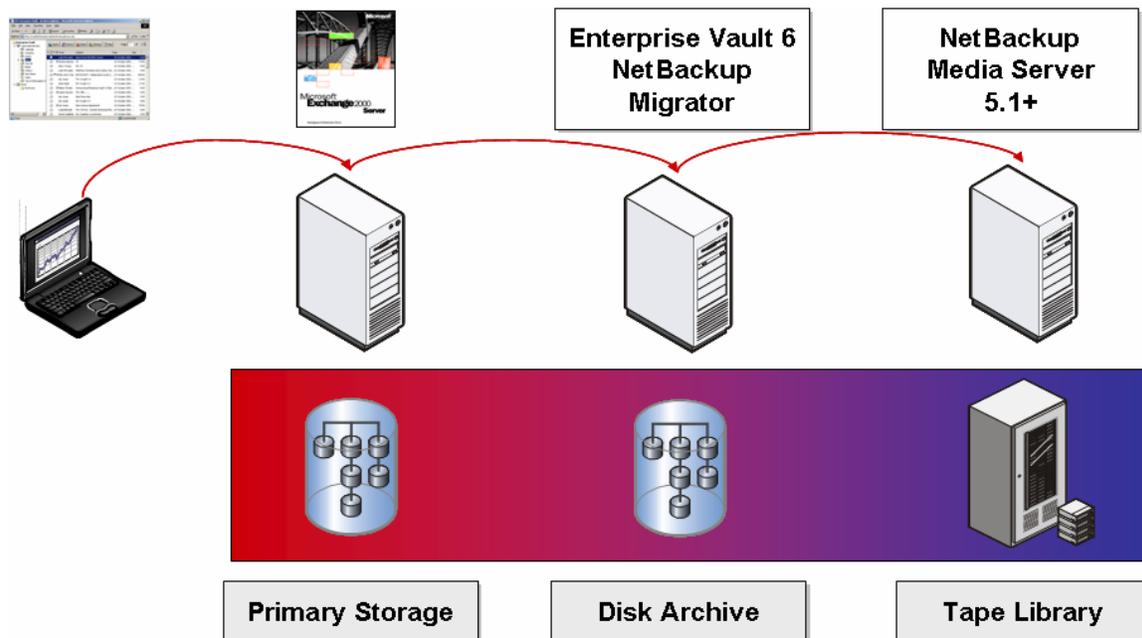
## Integration Overview



Figure 1 Overview of migration path.

Enterprise Vault 6.0 can now automatically and transparently utilize storage devices within a NetBackup 5.1 or higher environment. This includes, inline copies, disk (DSU & DSSU's), tape (including WORM), and UDO optical (on Windows via the Pegasus InveStore software).

Utilizing NetBackup 5.1 or higher with the Enterprise Vault v6.0 archived items from Enterprise Vault can now be automatically stored and retrieved on storage devices managed by NetBackup. All archived items must be stored within a Vault Store Partition within Enterprise Vault first. Once Enterprise Vault has archived the item, the collection process is run and it is placed into a CAB file. The next step is the migration of the data from Enterprise Vault to the migration process of NetBackup. Enterprise Vault's migration process calls the NetBackup migration process which starts a backup of the CAB files via a NetBackup Policy. Once the backup is complete Enterprise Vault truncates the Vault Store Partitions copy of the CAB file. This reduces the Enterprise Vault disk storage space and leverages the investments made in the NetBackup infrastructure.

During the NetBackup migration process multiple copies can be made using inline tape copies. DSU and DSSU (Disk Storage Unit, Disk Storage Staging Units) are also supported for migrations direct to disk controlled by NetBackup. It is however highly recommended to keep traditional backups separated from Enterprise Vault data since the retention requirements are most likely to be very different.

## Background on Enterprise Vault Archiving

As Enterprise Vault archives data it is stored on disk in what's called Vault Store Partitions, where it is indexed, and retention categories are applied. The directory structure for a Vault Store Partition is a Year/Month/Day/Hour hierarchy. The DVS (Digital Vault Sets) files are created in the Hour directory based on the creation time of the managed item (using GMT). If the archived item is over 50 Megabytes in size the item is placed in a DVF file and a small pointer DVS file is created. The DVF file contains the data and the DVS file contains a pointer to the DVF file. DVF files are not indexed or compressed.
The Collector allows the administrator to configure the optimization of the Vault Store Partition by collecting the typically small DVS files into a smaller number of larger files.

As DVS files age, the Collector process can be configured to automatically run on the Enterprise Vault server which collects DVS files into collections called CAB files. The default size is 200 Megabytes or 25,000 files maximum per CAB file. As DVS files are collected into CAB files the original DVS files are removed, and the SQL database is updated as to the new location for each item. The CAB files are created in the Day directory of the corresponding DVS files.

The Collection process helps optimize disk space and increase disk performance and backup efficiencies.

The Migrator automatically copies the CAB files to tertiary storage and then frees the CAB files from the Vault Store Partition thus reducing the total amount of disk space required within the Vault Store Partition.

Once a CAB file has been copied via the Enterprise Vault Migrator process, the original CAB file is either immediately deleted or the extension is changed to ARCHCAB. This file name indicates the status of the ARCHCAB as a redundant copy of a CAB file that has been migrated to another storage device. The length of time the ARCHCAB file is retained is configurable (See Figure 11). The SQL database is updated as to the new location of the CAB file in tertiary storage.

## Overview of Archiving and Migration

When items are archived to Enterprise Vault as DVS files, as described above, they can be optionally collected into CAB collection files. With Enterprise Vault 6.0, and the Migrator option, these files can now be migrated based upon age to tape or other media managed by NetBackup 5.1 or higher. The Migrator essentially "backs up" these files to NetBackup's Media Server and tracks to which image these files were written. An example of this is shown below:
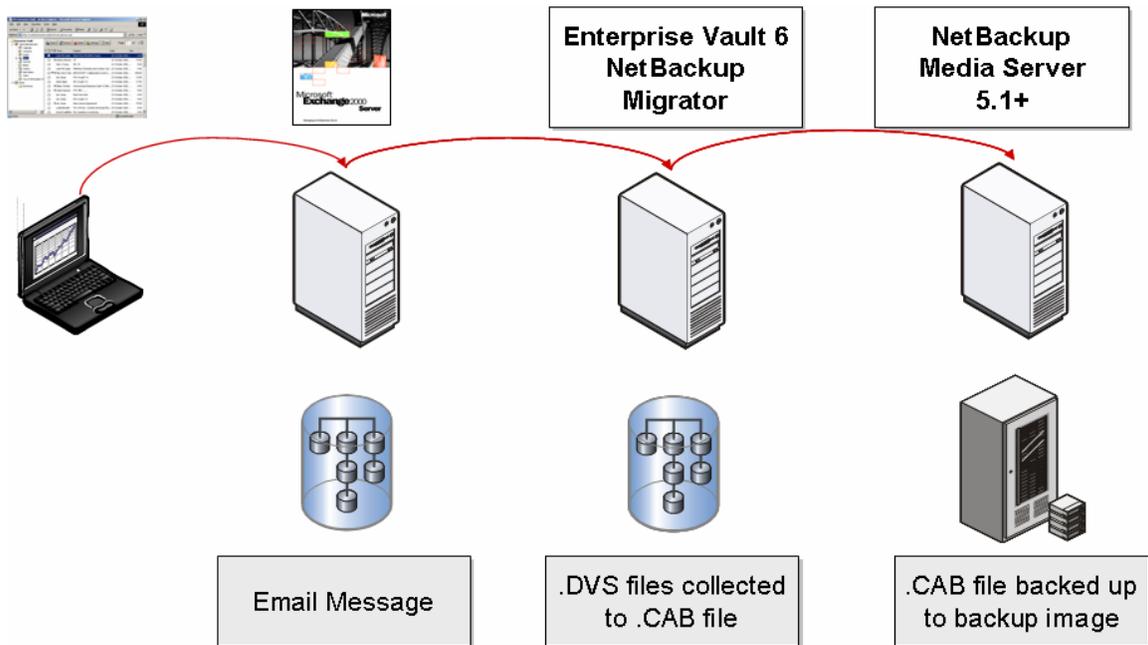


Figure 2 illustrates the process of managing and migrating items from Enterprise Vault to NetBackup:

Notice that all the managed data must first be placed on disk (in a Vault Store Partition) before being collected into a CAB file and then migrated to NetBackup.

## Overview of Retrievals

When an end-user recalls an archived file that has been migrated to NetBackup, Enterprise Vault looks up (in its SQL database) the location of the archived item.  In this case, Enterprise Vault will know that the file was collected into a CAB file and that the CAB file was migrated to media managed by NetBackup.  It will initiate a restore of the CAB file to the disk archive, where the DVS file for the archived item desired will be extracted.

The item is then returned to the user like a normal archived item.  However, given that tape access speeds can be slower and that, in some cases, the tape may no longer be in the library, Enterprise Vault 6.0 allows the administrator to configure a timeout after which time the user will get a message that the restore will take a long time and that they should check back later.  The administrator can configure a second timeout to define how long the system will continue to try to restore the data itself (See Figure 12).
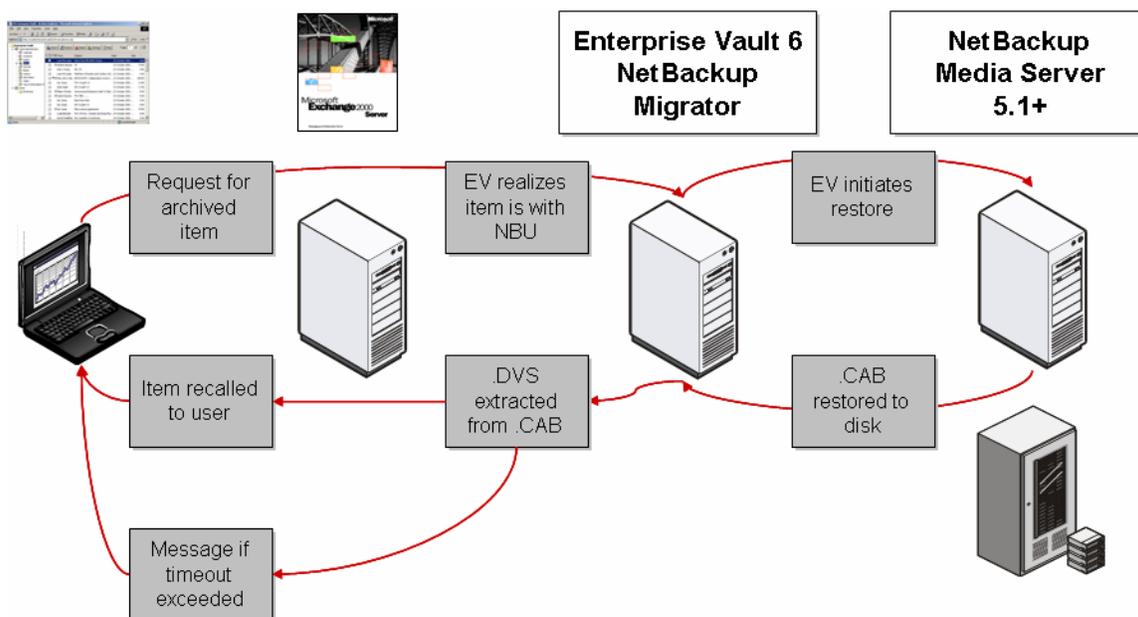
Figure 3 illustrates the overview of how retrievals work:


## Overview of Expiration

Enterprise Vault allows customers to not only archive data but also control how long that data is retained – and to also optionally expire that data after the retention period is over.  To this end, the integration between Enterprise Vault 6.0 and NetBackup allows for expiration of data archived to backup media.

Enterprise Vault tracks the retention period of each item and will notice when an item is expiring. It will mark the item as expired in the SQL database and will remove the related index data for the item.  However, since the item sits in a CAB collection file, not all items may expire at the same time.  Enterprise Vault will wait until most of the items in the CAB file have expired and it will then restore the CAB file to disk from media managed by NetBackup.

Enterprise Vault will then create a new CAB file with the contents of the old CAB file minus the expired items (DVS files).  These CAB files will then be migrated to NetBackup and Enterprise Vault will tell NetBackup to mark the old CAB file image as "expired."  NetBackup will then recycle the tape once all images have been expired on a tape.
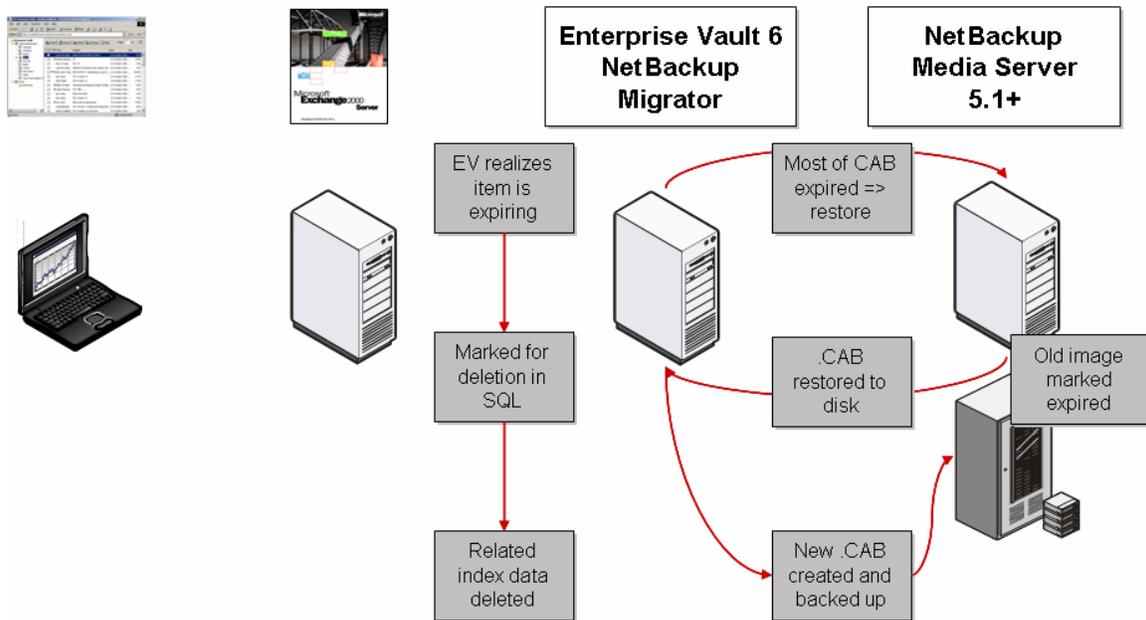
Figure 4 illustrates how expirations work:

## Flexible Migration Policies

Enterprise Vault 6.0 provides a tremendous amount of flexibility for configuring automatic migrations. Administrators can define the following policies to drive archiving and migration:

- **Traditional Enterprise Vault email, file and SharePoint archiving policies.** These policies define when items are to be archived, when and how shortcuts are to be created and whether items are to be deleted and are based upon extensive criteria such as user, folder, age, size, quota, file type, etc. (depending on the type of content being archived).
- **Collection policies.** These policies define when to create collection files (daily window), the maximum size for a CAB file, how to choose items (based upon time from last modified date) for collection and whether to migrate collections or not.
- **Migration policies.** These policies define how long after collection to migrate CAB files and whether to remove CAB files from disk after migration.
- **Expiration policies**. These policies indicate whether to expire data after the retention period is over.
- **Timeouts.** As described previously, these settings indicate how long users can wait for recalls before being directed to check back later, as well as how long the system should wait for a restore.

See the examples of timeout messages in Figures 13 – 17.

## Configuration of Enterprise Vault and NetBackup

Both Enterprise Vault and NetBackup have many different settings which control the management and movement of data.  Some examples are when items are to be archived, where to archive the item to, and how long to retain the items on which level of storage.  Since information varies in importance and retention, different types of storage devices maybe needed within site.
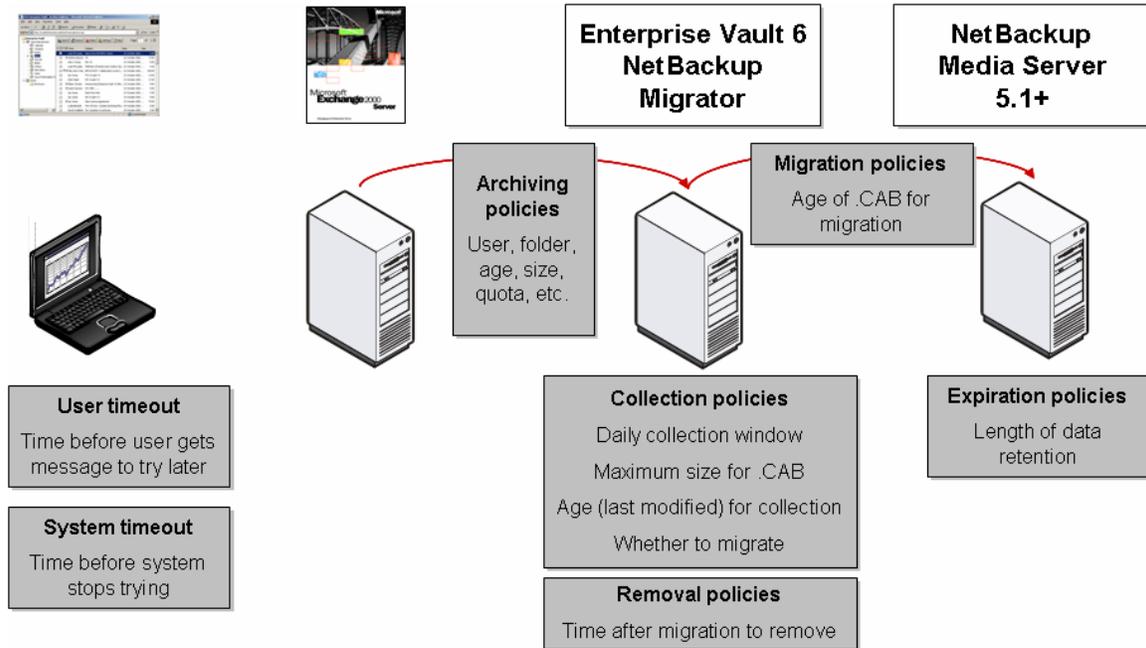


Figure 5 illustrates how policies can allow for highly configurable migration scenarios:

## NetBackup Configuration

NetBackup must have the xBSA license installed - note xBSA is part of the core product, no additional license is needed.  NetBackup needs to have a Backup Policy created.
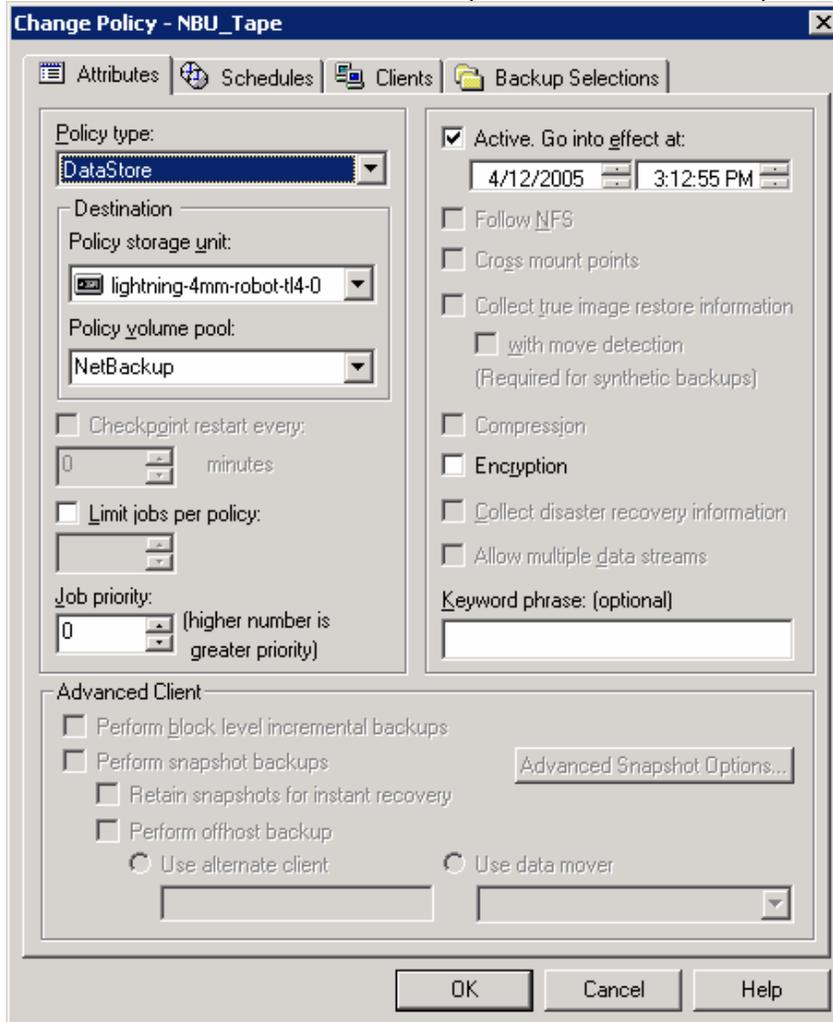


Figure 6 Screen shot from the NetBackup GUI for Creating/Changing a Policy.

When creating the policy, select the DataStore Policy type.

The schedules should be set to allow backups and restores to happen at anytime. The migration (backup) times are actually controlled by the Vault Store Partition configuration.
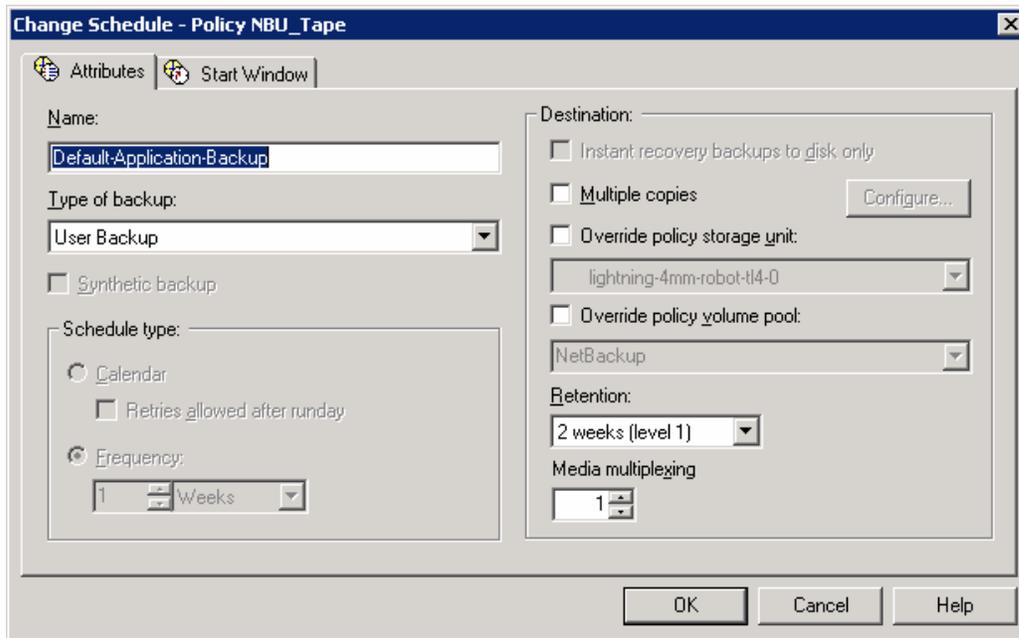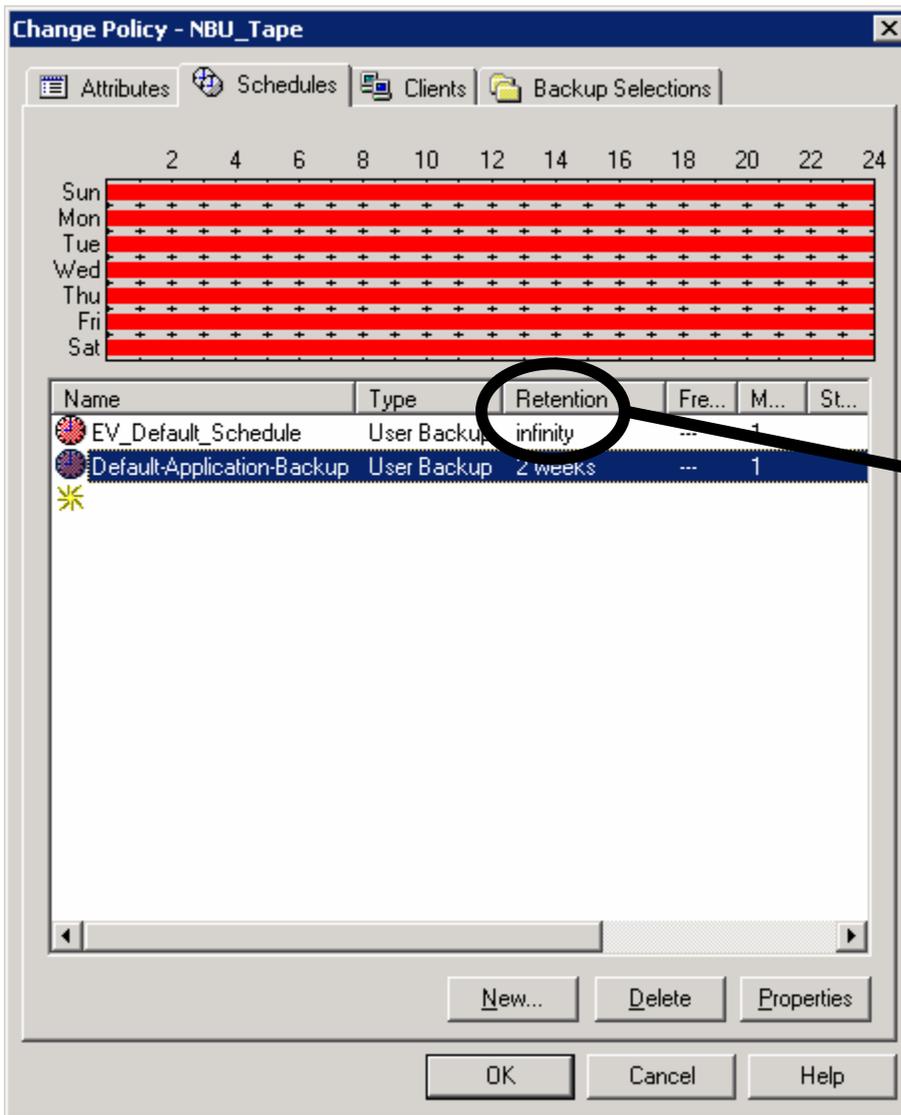


Figure 7 Screen Shot of Changing a Policy schedule.

Figure 8 Screen Shot showing a retention setting of infinity within NetBackup.

The Enterprise Vault server name is used in the NetBackup client for the policy. Set the "backup images" to infinity so they never expire.  When Enterprise Vault expires the CAB file NetBackup will be explicitly notified to delete it.

No specific policy directives need to be specified since the API's will handle the passing of file names between Enterprise Vault and NetBackup.
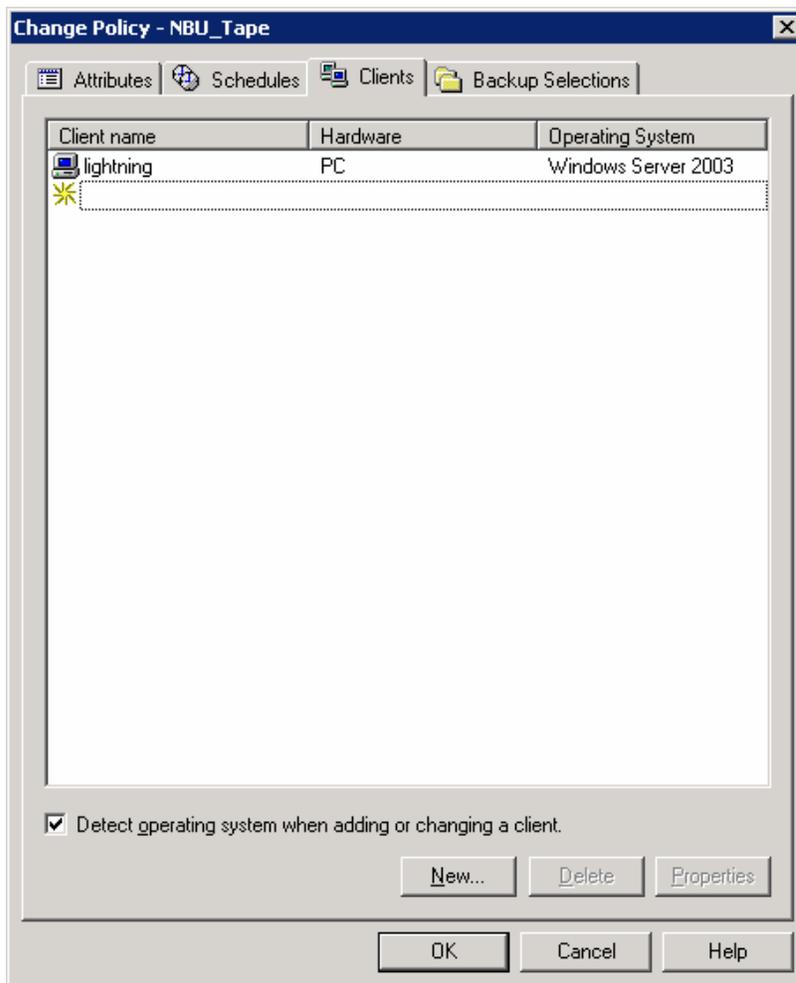


Figure 9 Screen Shot of NetBackup for selecting clients to backup.

**Notes for NetBackup Administrators**

- o **Additional tape drives and storage slots need to be considered when using the NetBackup migrator feature for storing Enterprise Vault data.**

- o **If tapes are removed from the library, timeouts will occur and users will not be able to automatically retrieve their data.**

- o **Timeouts may occur if all tape drives are in use when a Enterprise Vault user/application accesses data residing in a library.**

- o **Timeouts may occur if a migration is occurring (writing to the tape) while data on the same tape is trying to be accessed for a retrieval.**

## Enterprise Vault Configuration

Using the Enterprise Vault GUI, select the Collection page within the Properties page of the Vault Store Partition.  Select the schedule for when the collections are to be run. Typically configure quiet times when archiving and backups are not scheduled.
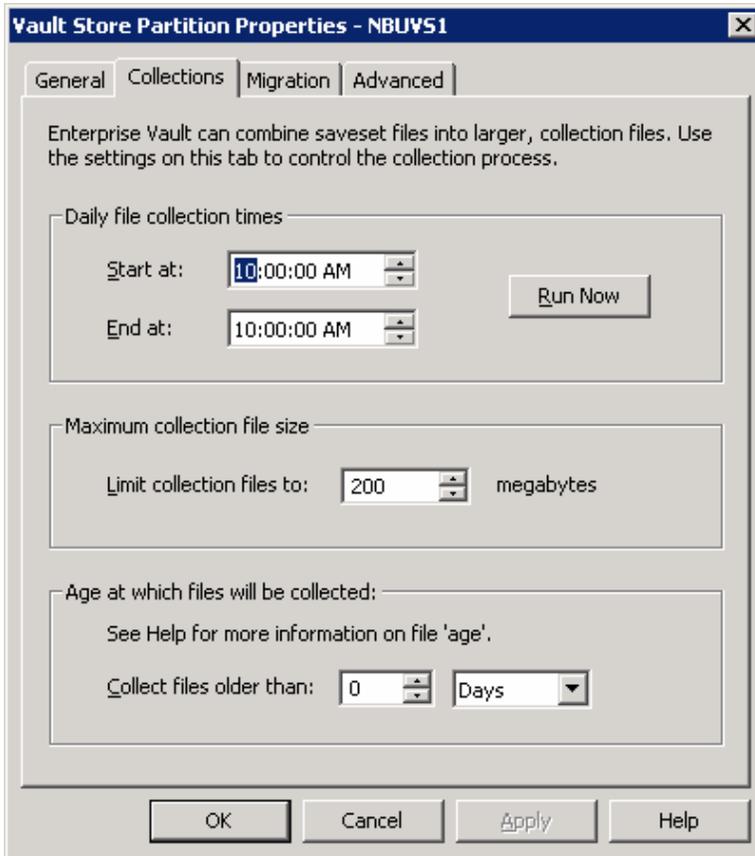


Figure 10 Screen Shot of Enterprise Vault, Vault Store Partition Properties.

The Collection process allows the administrator to specify how old the DVS files need to be before they are collected.  Once a DVS file is placed into a CAB file, any additional file archived that is identical will not be single instance stored with the copy in an existing CAB file. A typical setting would be 30, 60 or more days after a DVS file has been archived before it gets collected into a CAB file.  Depending on archiving polices this should allow the maximum single instance storage of archived files.

Select the "Migrate collection files" check box in the Vault Store Partition. Select a numeric value and then either Years, Weeks or Days. Additionally the "Remove collection files from primary storage" settings need to be specified. This value sets the amount of time the ARCHCAB files stay in the Vault Storage Partition after the collection is copied to tertiary storage.
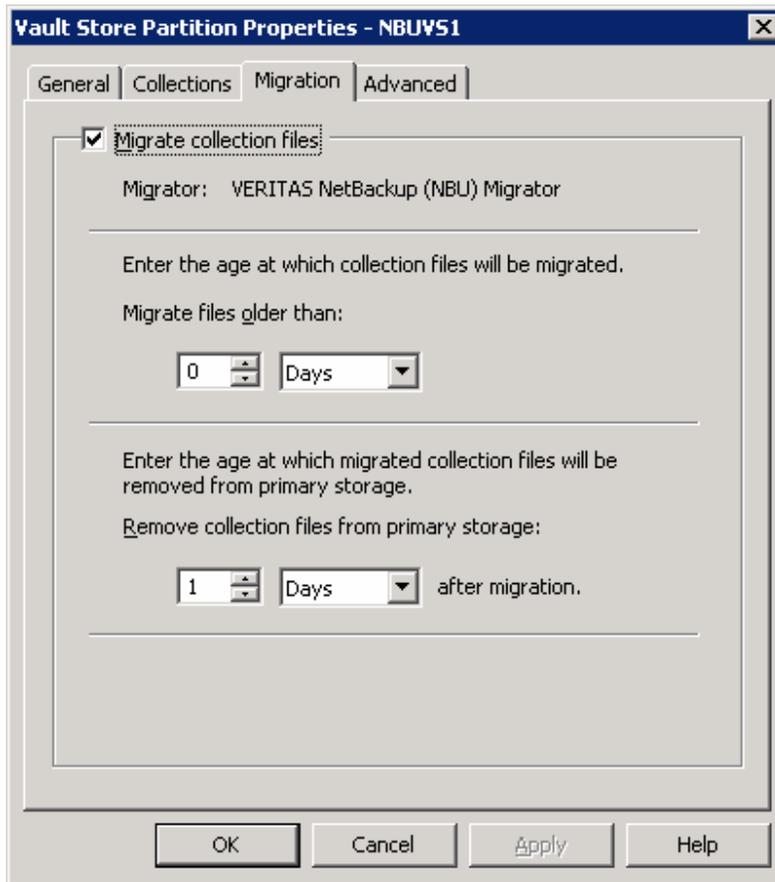


Figure 11 Screen Shot of Enterprise Vault's Vault Store Partition Migration parameters.

In the Advanced tab, two timeout values also can be tailored to the specific type of tertiary storage. They are the SystemWaitTimeout and the UserWaitTimeout. The SystemWaitTimeout specifies the number of seconds the Enterprise Vault system should wait for migrated file retrieval requests before timing out. The requested file will be removed from the retrieval queue if this value is exceeded. The UserWaitTimeout specifies the number of seconds Enterprise Vault user should wait for migrated file retrieval requests before timing out. The requested file will remain in the retrieval queue. These values may need to be increased if, NetBackup storage devices are not available (e.g. all devices are in use doing backups or restores), or the devices are slow to load and or read.
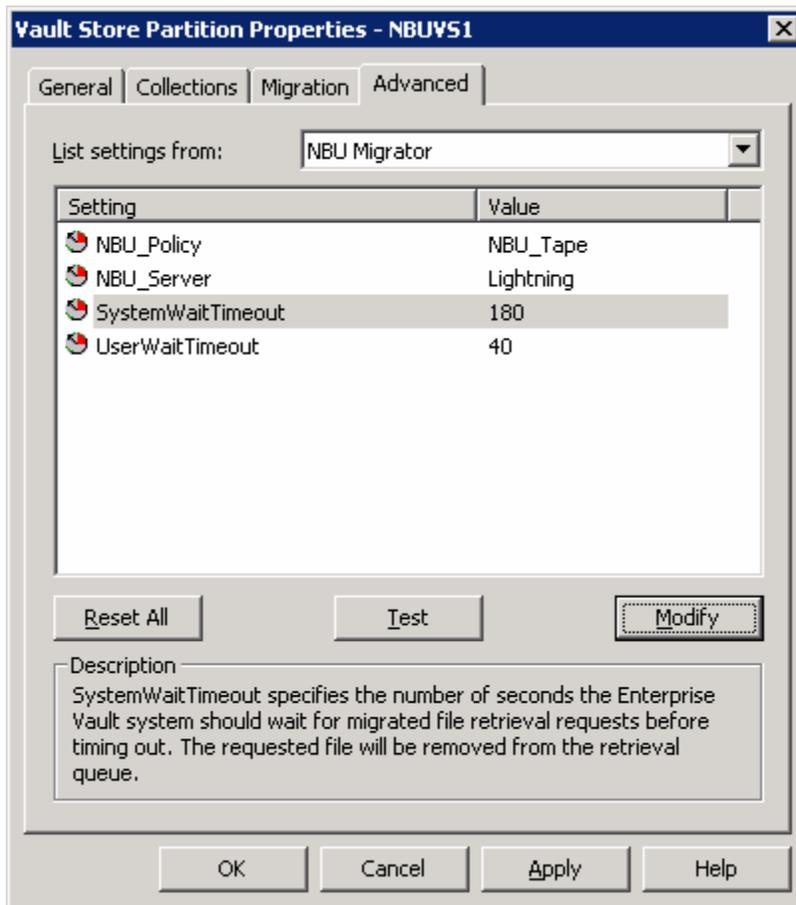


Figure 12 Screen Shot of Enterprise Vault's Vault Store Partition Migration timeout parameters.


## Access and Retrieval of Collected and Migrated Items

When accessing collected (CAB file items) that haven't been migrated, the DVS file is retrieved from the CAB file and placed in it's original Hour directory with the ARCHDVS extension, and the item is returned to the user/application that accessed it.

When accessing an item that has been collected and migrated but the ARCHCAB file is still present in the Day directory, the DVS file is retrieved from the ARCHCAB and returned to the Hour directory as an ARCHDVS file.

When accessing an item that has been collected and migrated and the ARCHCAB file has been freed. Enterprise Vault will request via the migrator API that the CAB file be restored back to the original Vault Store Partition. NetBackup then searches its catalog for the CAB file and does an alternative restore back to the original Day directory with the ARCHCAB extension. Enterprise

Vault then retrieves the DVS file, and renames it and places it in the original Hour directory as an ARCHDVS file.

The collector process deletes the ARCHDVS and ARCHCAB files when it does its scheduled runs so these file will be freed from the Vault Store as per the settings in the Vault Store Partition.

If the retrieval of the CAB file exceeds the limitations set in the System and User wait time the following messages will appear, based on the method of retrieval.   The desired file will continue to be retrieved even if the message appears.
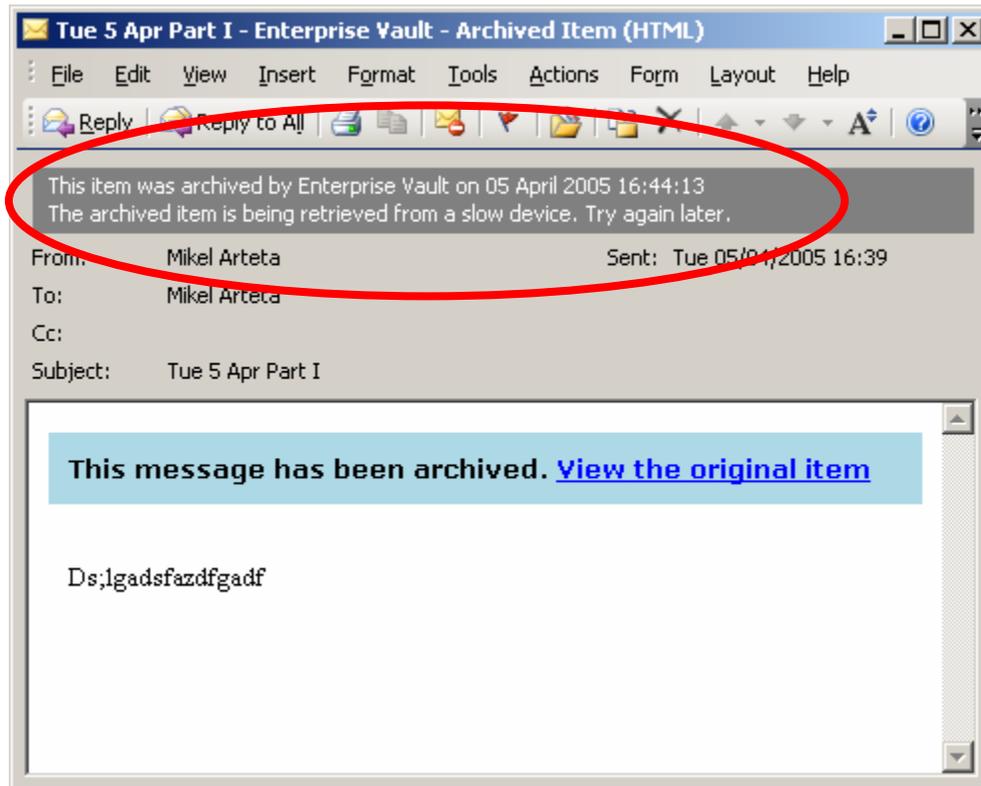

Figure 13 Timeout message for Outlook.
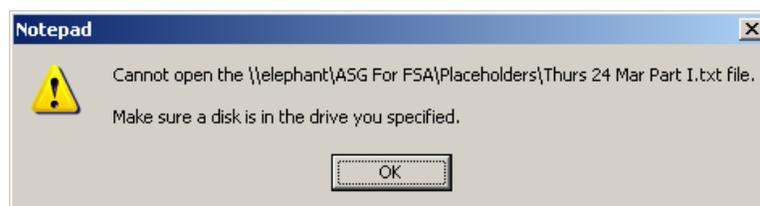

Figure 14 Timeout message for Integrated Searches.


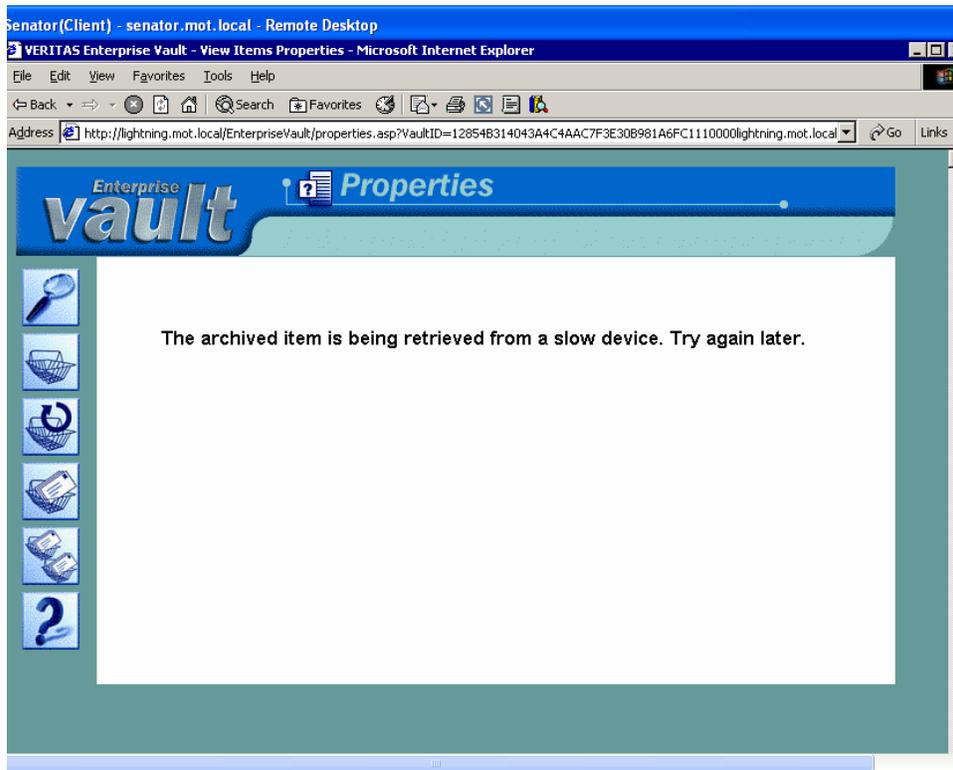Figure 15 Timeout message for File System Archiving.
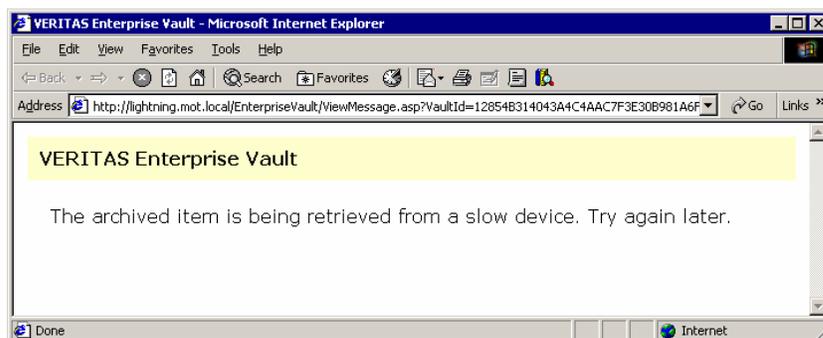
Figure 16 Timeout message for Web Searches.


Figure 17 Timeout message for Universal Short Cuts

## Expirations when Collections and Migrations are implemented

When archived data is deleted whether by the expiry process or when an original item has been deleted and the policy setting is to delete archives when the original is deleted a chain of events starts. First the SQL reference to the data is updated to reflect the deletion request.  Next the indexed data for the item is deleted. If the file hadn't been collected into a CAB file, the individual DVS file will be deleted if not on a WORM device.  If the deleted item was within a CAB file, the item is marked as deleted but physically stays within the CAB file.  Once 90% of the DVS files within a CAB have been deleted a re-CAB process can be run that will create new CAB file excluding the deleted DVS files.  If the CAB files where within NetBackup's control then the original CAB files would be returned to the disk storage partition.  Once the re-CAB process takes place the new CAB files minus the deleted DVS files could be migrated to tertiary storage within NetBackup.   When the original CAB files have been "re-CABed" into new CAB files, Enterprise Vault sends NetBackup via the migrator API a deletion request for each of the original CAB files so NetBackup can mark them for deletion.  If the CAB files are stored on tape within NetBackup,

deleting one CAB file (about 200 Megabytes by default) will not cause NetBackup to necessarily expire the tape.  NetBackup will keep the tape active until the last image has been expired (deleted) which could be quite a while.  If regulated data that under company policy must be expunged at specific times an alternative tertiary storage device may need to be used, such as a Disk Store Unit (DSU).

However, since the collection process collects DVS files from within the same Day's directory, collections should expire the same day.  Therefore, if enough data is collected and migrated to the same NetBackup tape to fill it within a few days, then all the CAB files will be deleted within a few days of each other, which will then allow NetBackup to expire and reuse that piece of media which could mitigate the whole problem.  Careful planning must be done when working with regulated data, and the collection and migration processes.

## Conclusion

Enterprise Vault environments are now able to leverage the investment of the NetBackup infrastructure to help reduce the total amount of disk space needed for long term storage of archived data within Enterprise Vault.  Backup's of the Enterprise Vault environment are also reduced by taking the older data and migrating it into an existing NetBackup environment where multiple different types of management techniques such as disk storage units, multiple copies and offsite vaulting of data can be easily done.

**VERITAS Software Corporation**
Corporate Headquarters
350 Ellis Street
Mountain View, CA 94043
650-527-8000 or 866-837-4827

For additional information about VERITAS Software, its products, or the location of an office near you, please call our corporate headquarters or visit our Web site at www.veritas.com.