WHITE PAPER: TECHNICAL

symantec.

Confidence in a connected world.

Using Virtual Tape Libraries with Veritas NetBackup™ Software

Don Peterson, Principal Product Manager
April 2009

Table of Contents

What is a Virtual Tape Library?

A Virtual Tape Library (VTL) is typically a dedicated appliance, which emulates a physical tape library and tape drives and stores backup images on disk. The VTL allows existing tape backup software to be used and the paradigm for managing backups remains fairly similar to that of using physical tape. However, there are some differences and we'll discuss the implication of thinking the paradigm is identical later in the paper.

The VTL consists of three components:

- computer hardware, usually Intel processor-based and powered by the Linux operating system or a close variant;
- application software, which emulates a tape robot and tape drives; and
- a RAID-based array of disk drives, which ensures no backup data is lost if a hard drive fails.

These components are frequently bundled by a single vendor into an appliance. However, it is possible to purchase the components separately.

The VTL allows a customer to configure virtual tape drives (NetBackup supports specific drive types) and virtual tape cartridges, and to specify cartridge capacity. The maximum number of supported virtual tape drives varies by vendor. Unlike physical tape libraries, which require additional tape drives be purchased and installed, virtual tape drives can be configured in a VTL with no additional hardware costs.

When using NetBackup software, new storage units must be configured to use the VTL's virtual drives. Backup policies must then be modified to use the new storage units. A NetBackup media server must also be configured as the robot control host. The NetBackup configuration process is identical to that needed for a physical tape library and drives. If multiplexing is being used with physical tape, it may not be necessary with a VTL, but this will be discussed later in the paper.

How Is a VTL Qualified with NetBackup Software?

Many VTL vendors claim that they are able to emulate any physical library and drive type. However, with over a decade of experience testing tape libraries and drives, the NetBackup HCL team knows vendor tape libraries operate differently from one another.

NetBackup software requires that a VTL provide a unique library inquiry string so both the NetBackup software and NetBackup customer support know when virtual tape libraries and drives are being used. VTL drives use the inquiry string of the physical drive they are emulating because they need to use existing physical tape drivers. The NetBackup Hardware Compatibility Lab (HCL) qualifies each different tape drive with every operating system on which support is desired. Because of the amount of time and effort this testing takes, Symantec limits NetBackup support to a small number of tape drives specified by the VTL vendor. If the drive type is an important consideration, a VTL qualified for use with NetBackup with that specific drive type should be chosen. However, as discussed below, drive type is not significant when deploying a VTL. Symantec allows VTL vendors to self-qualify new drive emulations if the OS platform for which a new emulation will be qualified is already supported by NetBackup with the VTL.

NetBackup software currently supports VTLs for the open systems market from 20 different vendors. See the Virtual Tape Libraries/Drives section of the NetBackup 6.5 HCL at the following link:

ftp://exftpp.symantec.com/pub/support/products/NetBackup_Enterprise_Server/284599.pdf

For the NetBackup 6.0 HCL, use the following link:

ftp://exftpp.symantec.com/pub/support/products/NetBackup_Enterprise_Server/278692.pdf

How are VTLs licensed with NetBackup?

In September 2008, NetBackup licensing of VTLs (and all disk-based backup options such as OpenStorage and AdvancedDisk) changed. There are now two options available for licensing VTLs. Both of these are based on the amount of primary storage (Front End Terabytes) that are backed up to the VTL. This capacity is the total sum of data, from a NetBackup perspective, included in the largest set of full backups to the VTL. If the VTL can perform deduplication and/or replication, the Enterprise Disk Option license is required. If the VTL cannot perform deduplication or replication, the Standard Disk Option license is required. The suggested list price of these options is $1,795 and $995 respectively, per front end terabyte. These licenses allow a customer to configure as many virtual tape drives as desired and allows sharing of these virtual tape drives among media servers. Sharing of virtual tape drives is typically not necessary or wise when using a VTL. This was useful when VTLs were first introduced and some of them supported only four to six drives, but support for configuring hundreds of virtual drives has alleviated the need for sharing the drives.

NOTE: The Library Based Tape Drive Option for NetBackup is **ONLY** for licensing physical tape drives with NetBackup, not virtual tape drives.

Comparing the VTL and a Standard Tape Library

Problems with physical tape drives and media lead to the failure of a fair number of backup jobs, and these problems can be difficult to diagnose. Write errors, reported by an operating system, do not indicate whether the media or drive is at fault, leaving backup applications to make an educated guess — sometimes incorrectly — at the cause. The administrator must then spend time determining whether the media or drive caused the problem. In addition, restoring from physical tape can involve multiple tape cartridges. If one of these cartridges is unreadable due to improper storage, lost tape, etc., the restore may be unsuccessful.

Because all VTLs use RAID storage, read and write failures are extremely unlikely so the VTL effectively eliminates drive and media issues from the backup and recovery process.

Most VTLs are capable of an aggregate 500 MB/sec throughput or higher. If deduplication is performed in real-time, this can limit the data input rate to something less than this. If deduplication is done as a post-backup process, the input rate is not affected unless deduplication on one backup is performed while another is in process. The design of how the VTL implements this latter type of deduplication will determine the performance impact.

Base VTL throughput can be improved by adding more capacity (disk drives), controllers, and Fibre Channel (FC) ports. However, with newer tape drives capable of native speeds at over 100MB/sec and over 200 MB/sec with compression, backing up large amounts of data to physical tape with multiple drives being used may still be faster than a VTL.

Performing restores from the VTL may be faster than using physical tape. This is likely to be the case when recovering specific files, due to the random access of disk compared to the mount and load times and sequential access of tape. However, if huge amounts of data are being restored, and multiple tape drives are reading data in parallel, physical tape may be faster than a VTL.

With no cost imposed when configuring additional VTL drives — assuming the maximum allowable number of drives has not been reached — virtual drives can be allocated specifically for restore operations. This ensures that restores will be initiated quickly. Overall speed of the restore operation will still depend on available bandwidth and the size of the restore.

Comparing the VTL and Disk Storage Units

Backups to both VTLs and Disk Storage Units (DSUs) are written on disk and have similar benefits when compared to tape. Both remove tape drive and media issues from the primary backup process, minimize the need for multiplexing (high levels of which can slow down restores) and sometimes provide higher performance backups and restores than using physical tape. However, there are important differences between these types of storage units. NetBackup currently supports four general types of DSUs, which include:

- Basic Disk – A DSU, which doesn't perform deduplication and/or replication. It can't utilize the advanced functionality discussed in the following paragraph.
- AdvancedDisk – A DSU, which may perform duplication and/or replication. It can utilize the advanced functionality discussed in the following paragraph.
- OpenStorage – A DSU, which is an intelligent disk appliance and uses the NetBackup OpenStorage API for deeper integration with NetBackup software. It typically supports deduplication and/or replication and can utilize the advanced functionality discussed in the following paragraph. A NearStore DSU is specific to NetApp's SnapVault for NetBackup product, which uses a predecessor to the OpenStorage API, and for which NetApp has announced End of Availability.
- PureDisk – A Disk Storage Pool (DSP), which stores data deduplicated by NetBackup media servers, globally across domains. Replication from one DSP to another DSP is managed by NetBackup.

NetBackup 6.5 added new functionality specifically for backups to disk, except for Basic Disk. Because individual backup images may be removed from a DSU to free up space once they expire, intelligent capacity management can be used. Customers can configure NetBackup to retain images for a specific period of time, until duplication of an image has been completed (e.g., if it's being duplicated to tape) or the image can remain on the DSU as long as there is enough free space (intelligent capacity management). When using intelligent capacity management, customers can assign a service level to backups to indicate the importance of the data. NetBackup will delete the less-important data first as it needs to free up space. This allows the more important data to remain on the DSU for a longer period of time to improve restore times. This functionality does not work with a VTL because all backup images on a tape cartridge must expire before the space on the cartridge can be reclaimed

The VTL is optimized for moving large blocks of sequential data as the customer is not required to install a file system on the backup target. In many configurations, this provides higher performance than using an AdvancedDisk DSU, which must have a file system installed on it. However, the NetBackup OpenStorage (OST) API, which is used when backing up to an OpenStorage DSU, minimizes this difference as the customer does not need to install a file system on an OpenStorage DSU, which is an appliance.

The VTL also alleviates the need to determine how big to make each AdvancedDisk DSU. For example, if three heterogeneous NetBackup media servers have 6 TB of disk space available, a separate DSU is required for each media server, and each DSU must have a file system installed and be allocated a specific amount of the 6 TB of disk capacity. By using a VTL, all three media servers share the 6 TB, effectively providing dynamic sharing because the virtual cartridges would be used by each media server as needed. This type of sharing does not require the use of the NetBackup Shared Storage option (SSO).

With an OpenStorage DSU, logical storage units (LSUs) are configured, which can be shared across media servers, and the OpenStorage appliance vendor's file system manages that storage, so storage management and sharing of the available disk capacity is much more comparable to that of a VTL vs. the AdvancedDisk DSU.

However, VTLs do have some limitations compared to using DSUs with NetBackup. Because backup images are stored on virtual tape cartridges, until all images on a virtual tape have expired, the "tape" cannot be reused and thus disk space is not freed up on the VTL. In fact, even when all images have expired, the media has been unassigned, and the virtual cartridge is returned to the NetBackup Scratch Pool or the volume pool to which it was originally assigned, the VTL is not aware of this and does not free up the space. Only when NetBackup writes from BOT (beginning of tape) does the VTL become aware the space can be freed up.

Most VTL vendors allow customers to configure tape cartridges of smaller capacity than the physical tape cartridge provides. This helps to increase the amount of free space because a smaller amount of capacity is tied up when all images on a piece of media have not expired. In addition, most VTL vendors do not allocate the entire capacity of the cartridge when it's configured (only a relatively small amount) and only allocate more space as needed to minimize the free space issue. This does not cause any problem because NetBackup is not aware of media capacity, but just keeps writing to the media until it reaches End-of-Tape (EOT).

Another way to free up space on a VTL more quickly is to configure virtual tapes with very small capacities. However, this results in more NetBackup overhead in managing a larger number of virtual media and for vendors that copy virtual to physical tape, this would result in a huge amount of wasted space on physical tape cartridges.

While it is not a good idea, it's possible for a customer to force the VTL to free up space on unassigned media by relabeling the media. A customer could write a script that looked for unassigned media (or any media returned to the scratch pool), verifies it is media in a VTL, and then re-labels the media. However, once media is relabeled the backup images can no longer be accessed and the media can no longer be imported. Relabeling unassigned media or media returned to the scratch pool is also a good way to make any backup images on the tape inaccessible, if company or industry regulations mandate doing so.

With backups to a DSU, an administrator can view the free capacity of the DSU in the NetBackup Admin GUI. This becomes more complicated if deduplication is being used, but the OpenStorage API is able to query OpenStorage appliances that perform deduplication to find out how much logical space (takes deduplication into account), is still available within the DSU.

With a VTL, it is very difficult to determine how much free space is available for the next set of backups to be run. See the last section in this paper titled *Why Does the VTL Indicate the Wrong Amount of Free Space* for more information regarding this.

One of the major advantages of using a VTL vs. a DSU comes into play when backing up NAS systems using NDMP. Because the NDMP specification only supports writing from disk to tape, a VTL can be connected directly via FC to a NAS system using NDMP and backup data does not need to be sent across the LAN. If one wants to use a DSU as the target for a NDMP backup of the NAS device, the data must be sent across the LAN to the media server before being sent to the DSU. For customers with a lot of NAS primary storage, the implication of having to send all this data across their LAN for backups, especially if it's not a dedicated backup LAN, can be pretty severe.

Can I use my VTL with NetBackup OpenStorage?

NetBackup OpenStorage, typically referred to as OST, utilizes the NetBackup OpenStorage API, a vendor's OpenStorage plug-in and the vendor's *intelligent* appliance to implement a disk-based-backup solution.

When OST is used, NetBackup writes to Logical Storage Units (LSUs) that are disk targets that reside on the vendor's appliance and communicates with the vendor OST plug-in installed on the media server.

Because VTLs are intelligent disk-based appliances that emulate tape you can think of them as having a front-end consisting of software providing tape drive and library emulation and a back-end providing a RAID controller and disk, deduplication, replication, etc. Because of this, Symantec talked to the VTL vendors and suggested they implement an alternate way for their intelligent appliances (VTLs) to integrate with NetBackup. This would be accomplished by integrating with the OpenStorage API, thereby eliminating the limitations encountered when dealing with tape (discussed in the previous section) as well as providing integration with NetBackup for some of their enhanced capabilities, such as deduplication and replication.

To do this, a vendor has two choices. One method is to develop an OST plug-in that communicates directly with their VTL software. In this case, the plug-in will perform a translation from NetBackup telling the plug-in to write to a specific LSU to the plug-in writing to a specific virtual tape library and/or virtual tape drive and media. This method is implemented by FalconStor.

The second method is for the vendor to develop an OST front-end for their intelligent appliance, replacing their VTL front-end, and develop an OST media server plug-in that communicates with the OST front-end of their appliance. Using this method, there is no conversion from disk to tape. This is the method chosen by Data Domain and Quantum.

It's also possible the vendor may allow their intelligent appliance to be partitioned such that part of it is represented to NetBackup as a VTL and part as an OST device.

In summary, depending on the vendor's implementation, it may or may not be possible to directly connect a VTL to a NetBackup media server and utilize OST. However, if the vendor does implement an OST solution, and this can be determined via the NetBackup HCL, their plug-in and some version of their intelligent appliance and front-end software will be able to be used with OST.

Does it Matter How Many Virtual Tape Drives are Configured?

Most VTLs allow a large number of virtual tape drives to be configured. Because there is no added cost for configuring a large number of virtual tape drives, customers sometimes perceive "the more, the better". As mentioned earlier, one benefit of doing this is to dedicate specific tape drives for restore operations, alleviating any concern about having drives available for doing restores that are requested while backups are in progress. This can, however, affect the available bandwidth for backup jobs.

Customers like to treat virtual tape drives like physical tape drives and have a tendency to develop their architectures accordingly. This can create bandwidth and performance issues. When using physical drives, the goal is to keep the drives active as much as possible, so they're typically shared among media servers. This is a wise idea because you want to pump as much data to the physical drive as possible to take advantage of its speed and you want it being used for backups as much as possible during your backup window. As long as there is adequate SAN bandwidth for each drive, and enough bandwidth from the system being backed up, the drive can run continuously and as fast as possible.

However, a VTL has a specific amount of throughput, which typically is around 500 MB/sec or greater in total (in-line dedupe may limit this to a lower number). Unlike a physical tape library, where adding another tape drive (assuming you have or can add the appropriate amount of SAN bandwidth) may add

another 200 MB/sec of throughput (for a LTO4 tape drive and compressible data), configuring/adding another virtual tape drive does NOT increase the VTL throughput. Only by adding another VTL controller head, the appropriate number of SAN ports and possibly disk drives can VTL throughput be increased.

To demonstrate this, assume a physical tape library with ten LTO4 physical tape drives. Writing at 100% capacity (no compression) would be an overall throughput of 1200MB/sec. Configuring ten LTO4 tape drives on a VTL would yield only maximum throughput of the VTL, or in the paragraph above, 500 MB/sec.

Therefore, it is critical to configure the number of virtual tape drives such that overall performance of the system is not adversely impacted and the specific requirements and capabilities of the backup environment and the VTL must be used to dictate how many virtual drives are needed.

One must consider the implications when configuring (and concurrently using) a large number of virtual tape drives. A VTL is comprised of some number of disk drives and each disk drive has a single spindle. VTLs also implement RAID to provide redundancy to protect against drive failures. When using RAID, whenever a write operation is performed, parity information for those new data blocks must also be updated and this can impact write performance if the eixsiting parity information must first be read from the disk. Intelligent RAID controllers attempt to prevent having to first read the old parity information before writing new parity information, but this depends on the design of the RAID controller.

Because disk drives are random-access devices, a VTL does not encounter the issue of tape shoe-shining (meaning media must stop, be repositioned and start moving again), which occurs if a physical tape drive doesn't receive data quickly enough to keep the tape streaming. Because of this, customers sometimes perceive they can configure tens or even hundreds of virtual tape drives, so each NetBackup clients' data can be written to a separate drive, removing the need for multiplexing, which when used at high levels has the potential to slow down the recovery time.

For example, a client that can only transmit data at 5 MB/sec can't keep a physical LTO3 or LTO4 tape drive streaming, so multiplexing would be required to get acceptable performance out of the drive. However, with a VTL, because disk is used, the non-streaming issue is no longer considered. However, that thought process can cause a more significant issue.

Because each disk drive has only one spindle, if a customer attempts to concurrently use more virtual tape drives than there are spindles, performance degradation will occur due to thrashing of the drive heads (seek and rotational latencies). If a customer tries to run 40 concurrent backup jobs with only 10 disk drives, this will noticeably impact performance.

There are some alternatives to circumvent this potential issue:

- If jobs are from the same policy, the Limit jobs per policy attribute can be used to limit the number of jobs NetBackup performs concurrently when the policy is run.

- The Maximum concurrent write drives attribute specifies the number of tape drives that NetBackup can use at one time for jobs to a storage unit. The number of tape drives available is limited to the maximum number of tape drives in the storage device. This could be used to limit the number of virtual tape drives being used concurrently.

- Multiplex client data, so less drives can be used, which will result in less disk thrashing.

NOTE: This same limitation exists for backups to disk storage units, however, most customers would probably not configure a large number of volumes on a small number of disk drives as that would cause a good deal of storage management pain. In addition, NetBackup software does NOT support multiplexing backups to disk.

A potential issue then becomes if you limit how many jobs can run concurrently or how many drives can be written to concurrently, the VTL throughput may be very poor. If NetBackup clients can only send data at 10 MB/sec and you can run only 20 concurrent jobs, throughput is 200 MB/sec, well below the 500 MB/sec expected.

To address this issue, multiplexing could be used, which is discussed in the following section of this paper.

One thing you might do is ask the VTL vendor what their maximum I/O rate is for a VTL with a specific number of virtual tape drives. That number is the maximum number of virtual drives you want being accessed at any one time. So, you can configure more virtual tape drives, but you need to limit how many are concurrently being used to avoid disk thrashing.

In summary, achieving optimal performance will require considering client data speeds, media server throughput, VTL throughput, the number of virtual drives, and the number of concurrent backup jobs.

Should Multiplexing Be Used with a VTL?

Multiplexing of client backup jobs to a single physical tape drive is often used to keep a tape drive streaming. If the tape drive cannot maintain streaming, it must stop, reposition the tape, and start writing again, which has a huge impact on performance and affects reliability and the amount of data that can be stored on the tape. LTO2, LTO3 and LTO4 tape drives have been designed to stream at multiple speeds to try to minimize this shoe-shining. In addition, with a physical tape drive you want to maximize the speed at which data is sent to it, so the drive is being efficiently used.

Multiplexing is typically used when client data rates are slow, networks are slow or LAN traffic slows down data transfer rates from clients to media servers or many small backups (e.g., incrementals backups) are being performed.

However, high levels of multiplexing impact restore performance. Restoring data from a multiplexed backup may take longer because one client's data is intermixed with that of other clients and spread over a larger area on the tape cartridge.

If you've read the previous section of this paper, you now realize you need to carefully consider the number of disk drives in the VTL, the number of virtual tape drives you configure, how many backups jobs you will be running concurrently, and client data rates, to determine whether multiplexing is required in your environment to be able to complete your backups within your backup window.

Find out from the VTL vendor how many I/Os /sec your specific VTL is capable of performing, configure NetBackup so only that number of concurrent jobs will run, then increase the maximum number of streams (1 – 32) per drive (configured per storage unit) until you maximize your VTL throughput.

How Does Deduplication Affect Backups to VTLs?

While not all VTLs provide deduplication, when it is available it is transparent to NetBackup. The VTL will perform deduplication either in real-time, as the data is coming into the VTL, or after the backup process completes, so that the data rate of the data coming into the VTL is maximized.

Deduplication allows more data to be stored and kept for a longer period of time on the VTL which is important as the VTL has a finite amount of disk to use for storage unlike a physical library that can be completely refreshed by removing used tapes and adding new tapes. Deduplication at the VTL level may alleviate the need to migrate or copy some of the data to physical tape. A potential limitation is as the storage capacity of the VTL fills, image expiration may free up little space because so much of the data referenced in expired images is still referenced in images that have not expired. Worst case, this will cause backup jobs to fail because no space is available, which will require expiring more backup images, migrating some backup images to another type of media or adding storage capacity to the VTL.

Will Different Virtual and Physical Drive Types Cause Problems or Affect Performance?

In the world of physical tape, drive types are chosen based on a combination of performance, media capacity, and reliability, and each has a bearing on price. These characteristics do not translate to the VTL.

There are three factors to physical tape drive performance: the mount time, the load time, and tape streaming speed. When VTLs were originally developed, NetBackup software incorporated delays for different drive types to mimic individual drive mount times and load times. NetBackup software has been modified to remove those delays, and the drive type configured in the VTL no longer matters. From the perspective of data throughput, the VTL does not simulate the performance of the physical drive type selected for emulation. Read and write operations from and to the VTL take place as fast as the disk array can execute those operations, regardless of the chosen drive type.

With regard to capacity, the virtual cartridges in the VTL can be configured to whatever size is desired. Although a physical LTO drive uses a cartridge holding 100 GB of uncompressed data, a virtual LTO cartridge can be configured to store 30 GB, 300 GB, 1 TB, or whatever capacity is appropriate to the application. Virtual cartridges should not be configured so large that they limit the number of concurrently running backup jobs — a VTL with 20 TB of storage can only support 20 concurrently running backup jobs if virtual cartridges are sized at 1 TB. Also keep in mind the disk image expiration issue noted previously in this document when configuring the size of the tape media capacity.

NetBackup software is not aware of the capacity of specific media. During a backup, NetBackup software writes data until notified by the tape drive that it is nearing the end of the media. Backup images in the VTL can be duplicated from one media to another by NetBackup software. A 300 GB LTO virtual cartridge can be duplicated to a physical SDLT, LTO2 or T9940 cartridge — or any other type supported by NetBackup software. If the 300 GB virtual tape is being duplicated to a T9840C drive, with media that holds 40 GB of uncompressed data, NetBackup software will span physical tapes to duplicate the virtual image. Sizing the VTL cartridge similar to that of the physical media likely makes sense, but many VTLs do not compress data and tape drives use compression. If a virtual cartridge holds 100 GB of data and is duplicated to a physical cartridge that holds 100 GB of uncompressed data, depending on how much data compression occurs when it's written to the tape drive, there may be only 30% to 50% of the physical media used. This is not an issue if media are not being vaulted, as NetBackup software can use the same media when duplicating another backup image. However, some VTLs support ejecting media from the library by duplicating the virtual tape to a physical tape (see more on this in the following section). In this circumstance unused capacity on the physical cartridge is wasted, resulting in inefficient media usage.

The final factor influencing physical drive type selection is reliability, and, once again, this has little correspondence in the virtual world of the VTL. The disk array of the VTL does not know whether it is configured as a low or high quality tape drive, and all drive types will benefit from the reliability of the RAID-based disk array. In this respect, the reliability of a virtual DLT 7000 drive is identical to that of a

virtual T9840 drive — even though there is a huge reliability difference between the two physical drive types.

There is one caveat to using different virtual and physical drive types. At least one vendor's VTL will not copy virtual media to physical media if the drive types are not identical. However, as discussed in a following question, using the VTL to make copies of virtual tapes is not necessarily wise to do.

Using NetBackup to Duplicate Backup Images from a VTL to Physical Tape

In many cases, VTLs are deployed as a front-end to a physical tape library whereby backup images can be copied from the VTL to physical tape using the duplication capabilities of NetBackup. This can be accomplished via the NetBackup GUI, the command line interface (CLI), by using a Storage Lifecycle Policy (SLP) or by using the NetBackup Vault option. In a nutshell, backups are sent to the VTL for short term backup and recovery, and are duplicated to physical tape so they can be sent offsite for long term storage.

NetBackup 6.5 introduced a feature referred to as VTL direct-to-tape, which uses the NetBackup NDMP direct copy functionality. If a VTL has embedded NDMP tape server software, NetBackup duplication can be configured from a NDMP storage unit in the VTL to a NDMP storage unit in the physical tape library (NetBackup sees the physical tape drive in the library via the NDMP path provided through the VTL). This is beneficial because the data path is not through the media server and thus off-loads the I/O from the media server. This also reduces the overall SAN traffic as the backup image is copied from the VTL to the physical tape library vs. copying the backup image from the VTL to the media server to the physical tape library.

When the duplication job kicks off, NetBackup mounts the virtual tape cartridge in a virtual tape drive in the NDMP storage unit in the VTL and positions it for the duplication, mounts a physical tape cartridge in a tape drive in a NDMP storage unit in the physical tape library, positions it, writes a backup header, and then tells the VTL to copy the backup image from the VTL to physical tape. Because NetBackup is managing the duplication, the drive type in the VTL and the physical tape drive can be different. NetBackup will span tapes if necessary to perform the duplication. The 2nd copy is then added to the NetBackup catalog. NetBackup controls the robotics in the physical tape library through either a direction connection to the library or via SCSI pass-thru commands through the NDMP tape server in the VTL.

The above functionality may also be used to copy backup images from one VTL to another VTL.

To determine whether a specific VTL vendor supports the NDMP direct copy functionality, see the VTL section of the NetBackup HCL listings.

For NetBackup 6.5:

ftp://exftpp.symantec.com/pub/support/products/NetBackup_Enterprise_Server/284599.pdf

For NetBackup 6.0:

ftp://exftpp.symantec.com/pub/support/products/NetBackup_Enterprise_Server/278692.pdf

The EMC DL4000 series of products embeds a NetBackup media server within it. A customer must buy a license from EMC that enables using the media server software in addition to buying the media server license from Symantec (or a reseller). Because this media server is used only for duplication, not for backups, a cheaper SAN media server license can be used. Therefore, a customer can buy a Tier 2

Enterprise Client license for this. This is because the VTL is dual processor Linux box (Tier 2) and the SAN media server is now part of the Enterprise Client license.

NetBackup software has multiple methods of migrating data from a VTL, or DSU, to tape. By using the NetBackup Vault option, Vault policies can be configured to schedule and automate the duplication of backup images on a VTL or DSU. In addition, Vault provides the tracking of media, and containers of media, for off-site media protection, and is integrated with Iron Mountain's web-based software to facilitate pickup, delivery, and tracking of the physical tape cartridges.

NetBackup 6.5 software added a feature called Storage Lifecycle Policies (SLP). This allows a customer to create a SLP that specifies a backup and duplication(s) of an image. When a backup policy is created, it can then be assigned to use the SLP (versus a storage unit or storage unit group), so NetBackup software will automatically duplicate the backup image after the backup is completed.

Manual duplication can also be performed through the NetBackup Admin GUI or via the CLI using the bpduplicate command. Scripts can also be written to automate the duplication process.

With the more recent addition of deduplication to VTLs, customers may be able to keep data on disk until it expires, so they may copy fewer backup images to physical tape than they previously did.

Can I Use the VTL Software to Copy Virtual Media to Physical Tape?

Allowing a VTL to copy a tape from the VTL to a physical tape, in a tape library controlled by the VTL, is tempting to do. It frees the NetBackup media server from performing the task, which relieves I/O bandwidth and processor cycles on the media server and frees up SAN bandwidth. Most VTLs providing this functionality make a byte-by-byte copy of the tape, so it retains the NetBackup tape format. In addition byte-by-byte copies are very fast.

However, a VTL duplicates media, while NetBackup software duplicates backup images. When a VTL duplicates media, NetBackup software is not aware of these second copies of the backup images. In addition, because NetBackup software is not responsible for making the physical copy, if the physical tape cannot be read by NetBackup software, it is the VTL vendor's issue.

Some VTLs copy a virtual tape as part of a VTL policy configured by the user while other VTLs make copies based on a NetBackup eject command.

In either case, when a VTL copies a virtual tape directly to physical media, the same NetBackup media identifier will be written on the tape. The NetBackup catalog will not accept two media cartridges with identical media identifiers. If a customer puts this physical tape into a library NetBackup controls and attempts to inventory the library, the inventory will fail and NetBackup will tell the customer to resolve the issue (the tape must be removed from the library) before the inventory can run successfully. If for some reason the barcode on the physical tape is different from that on the virtual tape, NetBackup will perceive the tape as new media and will overwrite the data on the tape, unless the customer imported the tape before it was overwritten.

In addition, when the VTL copies a virtual cartridge to physical tape, it MUST fit on the physical tape. Because media, not images, are being copied by the VTL, it cannot span tapes like NetBackup is able to do. If a 30 GB virtual tape is copied onto a 100 GB physical tape, space is wasted and media costs increase. VTL vendors do a number of things to attempt to make certain the copy process goes smoothly and minimizes tape waste.

If the VTL provides this copy capability, compression must be taken into account. If the VTL does not compress data, compression by the physical tape drive may result in unused space on the physical cartridge, which increases the media expense.

For VTL vendors offering compression in their VTL, it can be either hardware or software based. If software based, it will impact on performance. Because the VTL compression algorithm may be slightly different than that of a physical drive, the VTL vendor will likely make sure the virtual cartridge holds a slightly smaller capacity than what they expect the physical cartridge will hold. Some VTL vendors require the virtual tape drive emulation be identical to that of the physical tape drive (i.e., if the physical tape drive is LTO3, the VTL tape drive emulation must also be LTO3).

When NetBackup is not in control of the duplication process, it will not know about the secondary copy. Therefore, no catalog entries will exist for these second copies (copy2 of the backup image from the NetBackup perspective). When the primary backup image on the virtual tape in the VTL expires, NetBackup software will remove the image relating to this primary copy from its catalog. To restore data from the secondary physical tape copy, it must either be migrated back to virtual tape in the VTL or placed into a tape library NetBackup controls, and then NetBackup software must import the backup image into the catalog. The physical media must have an identical format to the NetBackup virtual media, and, if not, the VTL must be able to convert the physical tape copy back to the NetBackup format so that it can be imported. Most VTLs perform byte-by-byte copies of virtual media to physical media, so the tape format is a non-issue.

Once again, the above-mentioned problem only occurs when a VTL bypasses NetBackup software to duplicate the virtual cartridge, or when a VTL supports ejecting virtual cartridges. Because NetBackup software copies backup images, not media, a single virtual backup image can span multiple physical media cartridges to alleviate this potential problem.

Another issue may be encountered if the VTL migrates a virtual tape to physical tape without the knowledge of NetBackup software. A request by NetBackup to access the virtual media will require the VTL software to restore the physical tape copy to the VTL. This could occur if NetBackup software attempts to append additional backup images to a virtual cartridge. This delay would significantly impact the performance of backup jobs as new backup images were appended to the virtual media. Therefore, make certain the media is completely full before migrating it to physical media. If the tape is not completely full, suspending or freezing the media would limit the media to being used for restore operations, so NetBackup would not attempt to append to the media.

Some VTL vendors provide the ability to “eject” media from the VTL. This eject capability is accomplished by copying the virtual tape to a physical tape library attached to the VTL and ejecting the physical tape from the tape library. The eject operation can take multiple hours, depending on the type of tape drive used.

The process of removing physical media from the tape library works fine when using NetBackup software to eject the media, because NetBackup software will catalog the media as no longer existing in the VTL. However, the barcode label on both the virtual tape and physical tape must match; otherwise, NetBackup software will not recognize the physical tape when a restore is attempted. To restore the same tape directly from the VTL, the VTL software must inject the physical tape media back into the virtual library.

When using this type of eject capability, it is important to remember that the VTL is copying the virtual tape, not NetBackup backup images, to a physical tape. The capacity of the physical tape must meet or exceed the capacity of the virtual tape cartridge if the procedure is to succeed.

The NetApp VTL provides a feature called Shadow Tape, which hides the virtual tape from NetBackup software when a physical tape copy is made as a result of an eject command sent to the VTL. NetApp

keeps the virtual tape on the VTL as long as there is space available. If NetBackup software needs to restore from a tape ejected from the VTL, NetApp checks to see whether the virtual tape is still in the VTL. If so, it automatically mounts it in the VTL and triggers NetBackup software to inventory the VTL. NetBackup software will then determine the tape has been returned to the VTL and commence the restore operation. The result is a much faster restore than having to obtain the backup image from the physical tape.

Is There a Reason to Use the Shared Storage Option When Using a VTL?

As mentioned earlier, because VTLs support configuring a large number of virtual drives sharing virtual drives on the VTL is both unnecessary and unwise. Using SSO also requires additional NetBackup administration and inter-server communications.

If you properly configure your VTL and the NetBackup policies, so you aren't trying to write to too many virtual tape drives concurrently, there is no need for use SSO. However, if you can't afford to spend the time properly configuring your environment, you could configure the same number of tape drives as there are disk spindles, and share those virtual drives. This will prevent disk thrashing,

However, experience indicates many customers believe "the more, the better" and end up configuring not only too many virtual drives, but also sharing them to try to keep all of them in use. The result is reduced performance and slower backups.

Remember, a VTL has a limited amount of throughput. Other than possibly configuring a drive(s) only used for restores, you shouldn't configure more virtual tape drives than disk spindles and if those drives will be kept busy during your backup window without being shared, it would be wise to not share them.

Why Does the VTL Indicate the Wrong Amount of Free Space?

It can be very difficult for a customer to determine how much free space there actually is on a VTL. The VTL can indicate how much free space it is aware of, but this doesn't mean it is usable. For example, NetBackup is normally configured to write backup images with the same expiration period onto a specific tape (piece of media). It's possible the backups planned for the next day may have an expiration period that doesn't match available media with free space.

As mentioned previously, when NetBackup software unassigns a virtual tape cartridge, the VTL is not aware that data on the cartridge is no longer needed. It is not until NetBackup software begins writing to the beginning of that same virtual tape cartridge that the VTL becomes aware that the remainder of the data on that cartridge is no longer needed. The VTL then frees the disk space. Between the time the backup images on the virtual cartridge expire and the time NetBackup software writes a new backup image to the cartridge, the VTL will view the cartridge capacity as unavailable.

Although a customer could run a report to determine which media fall into this category, there is no way to know how much physical capacity these virtual tapes were consuming because:

- The VTL was likely configured to only allocate space as required and unless NetBackup had reached EOT when writing to it, the entire capacity of the cartridge was not used.
- Because the images have already been expired, the NetBackup catalog no longer has information on the capacity of data written to the cartridge and even if it did, if the VTL compresses data, NetBackup has no idea how much physical space the data actually consumes.

Although a customer might determine how many cartridges are unassigned, there isn't a way to figure out how much physical disk space was being consumed by those images.

Summary

NetBackup qualifies and supports VTLs from all of the major vendors and has implemented specific features, such as the VTL direct-to-tape, to enable a VTL to duplicate backup images under NetBackup management. This allows NetBackup software to catalog both copies of the backup image.

When backing up NAS storage using NDMP, VTLs definitely have an advantage in that the backup images can go directly from the NAS device to the VTL across FC, whereas sending the backup images to a DSU requires the data be sent across the LAN to the media server before going to the DSU.

It's important to properly configure your VTL within your backup environment and make certain you are not concurrently using more virtual tape drives than there are disk drives in the VTL to optimize performance.

However, because NetBackup software treats a VTL as a tape library and tape drive, there are inherent limitations that arise. These include not being able to free up space when a backup image expires, which prevents NetBackup from being able to optimize capacity management when it is being used for staging backups to physical tape. Knowing how much free space is available on a VTL is also very difficult if not impossible to figure out.

The NetBackup OpenStorage (OST) API has allowed VTL vendors to take their intelligent disk subsystem and provide an OST front-end in place of (or in addition to) a VTL front-end. This enables tighter integration between NetBackup and the vendor's appliance, as NetBackup is made aware of its advanced capabilities and can utilize those capabilities in performing backups and duplications. This joint integration through the OST API provides customers with a solution that provides even more capability than when using the vendor's VTL with NetBackup.

About Symantec

Symantec is a global leader in providing security, storage and systems management solutions to help businesses and consumers secure and manage their information. Headquartered in Cupertino, Calif., Symantec has operations in 40 countries. More information is available at www.symantec.com.

For specific country offices and contact numbers, please visit our Web site. For product information in the U.S., call toll-free 1 (800) 745 6054.

Symantec Corporation
World Headquarters
20330 Stevens Creek Boulevard
Cupertino, CA 95014 USA
+1 (408) 517 8000
1 (800) 721 3934
www.symantec.com