VERITAS™

ARCHITECT NETWORK

Using Virtual Tape Libraries with VERITAS NetBackup™ Software

December 9, 2004

TABLE OF CONTENTS

What is a Virtual Tape Library?

A Virtual Tape Library (VTL) is typically a dedicated computing appliance that emulates the drives of a physical tape library and stores backup images on disk. The VTL allows existing tape backup software to be used, and administrators are interested in these appliances because the paradigm for managing backups remains the same as when using tape.

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, to emulate a tape library and tape drives; and, a RAID-based array of disk drives, to ensure 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 computer hardware and software from one vendor and the disk array from a different vendor.

The VTL allows a customer to configure virtual tape drives, virtual tape cartridges, and to specify cartridge capacity. The maximum number of supported virtual tape drives varies by vendor, ranging from single digits to an unlimited number of drives. And, unlike physical tape libraries, which require additional tape drives be purchased and installed, virtual tape drives can be added to the VTL by changing the software configuration, with no additional hardware costs.

Because the VTL emulates a tape library and its drives, it does not require a change to the backup paradigm. When using VERITAS 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 VERITAS NetBackup media server must also be configured as the robot control host. The VERITAS NetBackup configuration process is identical to that needed for a physical tape library.

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 VERITAS NetBackup engineering team knows that vendor libraries and drives operate very differently from one another, and VTL emulation of a wide variety of libraries and drives is not easily achieved. In fact, VERITAS has yet to test a VTL that successfully emulated a tape drive and library without requiring software modification by the vendor.

VERITAS NetBackup software requires that a VTL provide a unique library inquiry string so that 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 the existing physical tape driver software. The VERITAS NetBackup Compatibility Lab 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, VERITAS limits NetBackup support to one or two tape drives specified by the VTL vendor. If drive type is an important consideration, a VTL qualified for use by VERITAS with that specific drive type should be chosen. However, as discussed below, drive type is not significant when deploying a VTL.

NetBackup software currently supports VTLs designed for the open systems market from 11 different vendors, and additional products continue to be qualified. See the listings for NetBackup 4.5 software at the following link under either the vendors' names or *View All Supported Libraries*.

http://support.veritas.com/dsl/vndrselect_ddProduct_NETBACKUPDC_prod1_NBUDC45.htm

For NetBackup 5.0 and 5.1 software, use the following link, click on the Compatibility tab, then clicking on "Device Support List for all supported libraries" for either NetBackup version.

http://support.veritas.com/menu_ddProduct_NBUESVR.htm

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 fails 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. Backups will often be migrated from the VTL to physical tape, introducing the possibility of physical media read and write errors. However, failures encountered during the secondary backup stage are a lesser concern.

Most VTLs are capable of 150 MB/sec throughput as a local SAN backup or when aggregating the throughput of multiple networked clients. This may improve performance and allow backups to be completed within a specified backup window. Base VTL throughput can also be improved by adding more capacity (disk drives), controllers, and Fibre Channel (FC) ports. However, with newer tape drives capable of backing up data, with compression, at greater than 50 MB/sec, backing up large amounts of multi-streamed data to physical tape may still be faster than the VTL.

Performing restores from the VTL may also be faster than using physical tape. This is likely to be the case when recovering specific files, due to the random access of disk as compared to the 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 the VTL.

Multiplexing of client backup jobs to a single tape drive is often used to keep a tape drive streaming. But, if the tape drive cannot continue streaming, it either has to stop, reposition the tape, and start writing again — which has a huge impact on performance and reliability — or the drive has to slow down and write data at reduced speed. Either way, backup performance takes a hit. Multiplexing also impacts restore performance. Restoring data from a multiplexed backup takes longer because one client's data is intermixed with many others and spread over a larger area on the tape cartridge. A VTL uses disk and provides random access to data. Rather than multiplex backups, each client can be allocated a separate virtual drive. If the disk backup is then copied to physical tape, it will not be multiplexed. Restoring from this tape will be faster than restoring from a multiplexed backup.

With no penalty 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.

In most instances the VTL will be deployed as a front-end to a traditional physical tape library. Backup data can be archived from the VTL to physical tape using the duplication capabilities of the VERITAS NetBackup graphical user interface (GUI) and the command line interface (CLI), or by using the NetBackup Vault feature. Because multiplexing is unnecessary, restoring from the physical tape is likely to be faster.

Comparing the VTL to Disk Storage Unit and Disk Staging Storage Unit

The VTL, Disk Storage Unit (DSU), and Disk Staging Storage Unit (DSSU) — introduced in the NetBackup 5.0 release — all write to disk and have similar benefits when compared to tape (they

provide higher performance backup and recovery processing, eliminate the need for multiplexing, and alleviate tape drive and media issues from the primary backup process). However, there are important differences between these methods.

The VTL is optimized for moving large blocks of sequential data with no file system involvement. In many configurations, this provides higher performance than using a DSU or DSSU, which writes data through the file system.

When using a DSU, if the volume becomes full during a backup operation, the backup will fail. With a DSSU, if the disk contains data that has already been relocated to the Final Storage Unit (FSU) this data can be removed and, as long as enough space has been freed, the backup process will continue without error. When using a VTL, VERITAS NetBackup software is presented with virtual tape cartridges and, unless the entire VTL runs out of free space, when one cartridge fills another is loaded into the virtual tape drive and the backup proceeds.

The VTL also alleviates the need to determine how big to make each DSU. Unlike the DSU and DSSU, the VTL allows all allocated disk space to be dynamically shared by all backup clients. For example, if three heterogeneous VERITAS NetBackup media servers share 6 TB of disk space, the DSU, or DSSU, requires a separate storage unit for each media server, and each unit must be configured with a file system and allocated a specific amount 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 VERITAS NetBackup Shared Storage option (SSO).

Important Questions and Answers about Using VTLs

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: 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 process, 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 DLT 7000 drive uses a cartridge holding 35 GB of uncompressed data, a virtual DLT 7000 cartridge can be configured to store 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. If backup jobs are being multiplexed, very large virtual cartridges can be configured. However, the ability to avoid multiplexing is considered by many to be a significant benefit of using a VTL.

VERITAS 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 VERITAS NetBackup

software. A 300 GB DLT 7000 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. 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.

Does it Matter How Many Virtual Drives I Configure?

Depending on the capabilities of the VTL, anywhere from single-digit to an unlimited number of virtual drives can be supported. However, unlike a physical tape library, configuring an additional virtual tape drive does not increase available bandwidth. 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 of the backup environment must dictate how many virtual drives are needed.

For example, assume a VTL has 150 MB/sec bandwidth and has virtual drives spread across various storage units and policies. There are two backup environments. The first contains many clients with limited amounts of data. The second uses SAN media servers so clients with large amounts of data do not need to transfer this data across the LAN when doing a backup.

In the first backup scenario, a physical tape configuration would likely multiplex the client data to keep the tape drives streaming. In the VTL environment, a separate virtual tape drive can be configured for each client, and the virtual library will likely be able to handle combined data rates generated by the backups. If the data rate overwhelms the VTL it will result in minimal performance degradation as disk is a random access device. Assuming each client does not provide data at more than 5 MB/sec, up to 30 virtual drives — 150 MB/sec divided by 5 MB/sec — can be supported before performance is impacted. With so many concurrent backups accessing the disk array, disk thrashing may create performance problems, although it will largely depend on the design of the VTL.

The second backup environment, containing a number of SAN media servers, would be expected to generate sufficient data to keep physical tape drives streaming. Assuming a media server provides data at 30 MB/sec, it would take only five backup jobs on five virtual drives to saturate the 150 MB/sec bandwidth of the VTL. If 30 virtual drives were configured and 30 backup jobs were scheduled to start at the same time, each job would average 5 MB/sec compared to the 30 MB/sec that would be possible if there were enough bandwidth. In this scenario, allocating more virtual drives could actually reduce the

speed of the backups. Adding another VTL controller, or more FC ports, to increase the bandwidth to the VTL would alleviate this performance bottleneck.

Because real-world backup environments are likely to be a mix of the two scenarios above, careful consideration must be given to the configuration of virtual drives in the VTL. Achieving optimal performance will require a balance between the bandwidth available to the VTL, the number of virtual drives allocated, and the desired number of concurrently running backup jobs.

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

Performing a direct transfer of data between the VTL and a physical tape in the tape library is tempting because it promises to free the VERITAS NetBackup media server from performing the task — relieving I/O bandwidth and processor cycles on the media server— and keeps the data transfer off the SAN. However, there is a downside to bypassing the NetBackup media server.

If NetBackup software is not involved in duplicating the backup images — NetBackup software duplicates backup images, not media — NetBackup software does not know that a second media cartridge has been made. When the backup image on the virtual tape in the VTL expires NetBackup software will remove the image from its catalog. To restore data from this physical tape it must be migrated back to virtual tape in the VTL and then NetBackup software must import the backup image into the catalog. If NetBackup software is able to see the tape library hosting the physical media, the media can be imported into the NetBackup catalog. However, 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.

When a VTL copies a tape from virtual media directly to physical media, the same media identifier will be written on the tape. VERITAS NetBackup software will not accept two media cartridges with identical media identifiers.

Many VTLs do not support compressing data when writing to a virtual tape cartridge. However, data is compressed by a tape drive when being written to physical tape. If the VTL copies a virtual tape cartridge to a physical cartridge, depending on how much compression occurs, there may be a significant amount of unused space on the physical cartridge, even if the virtual cartridge was full. An attempt can be made to take data compression into account by configuring virtual cartridges with a larger capacity than the physical cartridge. However, this assumes that it is possible to predetermine the compressibility of the data being backed up. If the data cannot be compressed sufficiently, then the data on the virtual cartridge may not be able to be copied to the physical media. If it can be compressed more than expected, then there is wasted space on the physical media. VTLs that are able to compress data are unlikely to use the same compression algorithms used by physical tape drives, and the same issues will be encountered.

The above-mentioned problem only occurs when a VTL bypasses VERITAS 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.

If the VTL migrates a virtual tape to physical tape, without the knowledge of NetBackup software, a request by NetBackup software 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.

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.

VERITAS 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.

Because NetBackup software duplicates backup images, rather than backup tapes, media capacity concerns are irrelevant. If the media to which backup images are being duplicated is too small, NetBackup software will simply span the backup image to another piece of media. This frees administrators to configure large virtual cartridge sizes without regard for physical media compatibility. For example, a virtual DLT 7000 cartridge — uncompressed capacity of a physical cartridge is 35 GB — can be configured with a capacity of 400 GB. The 400 GB virtual cartridge might then be duplicated to a physical LTO-2 cartridge, which holds roughly 400 GB compressed — depending on how compressible the data is. If all the data does not fit on a single physical cartridge, NetBackup software will span tape cartridges. Duplication can be performed through the NetBackup CLI, or via a script, using the `bpduplicate` command.

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

Many VTLs support the configuration of a large number of virtual drives and, in most environments, the NetBackup Shared Storage Option (SSO) is not needed. However, under certain circumstances SSO may prove useful.

Consider two NetBackup configurations. In the first setup, three NetBackup media servers are sharing eight virtual drives. In the second configuration, three media servers have each been configured with eight dedicated virtual drives in the VTL. The configuration with eight shared drives will support a maximum of eight concurrently running NetBackup backup jobs — assuming multiplexing is not used. Other backup jobs will be queued, waiting for a virtual tape drive to become available. However, the configuration with 24 virtual drives is able to support 24 concurrent NetBackup backup jobs. Depending on how fast client data is being provided by the media servers, the 24 backup jobs may overwhelm the VTL and cause enough disk thrashing that the overall throughput is less than that achieved by sharing the drives and having only eight jobs running concurrently.

If physical tape drives are currently configured to support multiplexing, to accommodate relatively slow client data rates, then a VTL configured with a large number of virtual drives is likely to be a better solution.

Analyzing the amount of data and data rates that the clients can maintain, and testing configurations with various numbers of shared and unshared virtual drives, is the best method of determining the configuration that provides the highest performance for any specific environment.

For VTLs that allow configuring only a small number of virtual drives, the NetBackup SSO feature may be provide improved performance.

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

When VERITAS NetBackup software expires 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 that the backup images on the virtual cartridge expire and the time that NetBackup software writes a new backup image to the cartridge, the VTL will view the cartridge capacity as unavailable. This processing can mislead administrators by indicating that the VTL is much closer to running out of space than it really is, possibly creating short-term panic.

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, VERITAS Architect Network, or the location of an office near you, please call our corporate headquarters or visit our Web site at www.veritas.com.