TABLE OF CONTENTS

1   TIVOLI STORAGE MANAGER QUICK PRIMER

1.1   TSM OVERVIEW

Unlike other storage management products available in the market today, which protect your data with full backups performed usually on the weekends and incremental backups during the rest of the week, TSM employs a technique called progressive backups forever. This backup method saves time and storage by backing up only new and modified files, which are tracked by TSM's own relational database. This method is efficient and provides many benefits for simplifying and taming the backup environment. Figure 1 shows the lifecycle of a typical TSM backup.

Figure 1) Basic TSM architecture.

IBM Tivoli Storage Manager is client-server architecture. The TSM client is any system that communicates and backs up to the TSM server by using a specialized TSM client. After the backup is initiated, the TSM server decides where to place that data according to the policy engine that drives TSM device handling. The most common configuration in today's TSM environments is:

1. Client initiates the backup to the TSM server.

2. The TSM server receives that data into a primary disk-based storage pool. This storage pool is used only as a staging area and typically has only enough space to hold all or part of the incremental daily backup.

    While the data resides on disk, any restore requests from the clients (if their requested data is resident in the disk pool) are satisfied from this disk pool.

3. The next step is to make sure that a disaster recovery copy of the data exists. Under ideal conditions, most TSM installations make an off-site copy of the data from disk pool to a copy tape storage pool (commonly called an off-site tape pool). These tapes are usually under the control of TSM Disaster Recovery Manager (DRM). They are ejected from the tape library at the end of the copy process and are sent off-site to a remote location on a regular basis. Optionally, data can be backed up to a copy storage pool at a remote location using virtual volumes on a remote TSM server over an IP connection.

4. When the copy to the off-site tape pool is completed, the TSM server triggers a migration process at scheduled intervals to move the contents of the disk pool onto a primary tape pool (commonly known as an on-site tape pool). When the data is in the on-site tape pool, all restores are performed from tape.

   The migration process frees up the space in the primary disk pool so that it can receive more data from the next backup iteration.

When the cycle is complete, the TSM server has several background server-side housekeeping processes to make sure that the environment is run optimally. Expiration and reclamation are two server-side processes that have a great effect on the efficiency and cost of running a TSM environment.

Over time, as data is deleted from the client or expires as a result of polices set on the server, the TSM database needs to be cleaned up to make room for new entries and to remove references that point to data on media that are no longer valid. This is accomplished through the expiration process. Expiration marks data as expired and deletes metadata for the expired data from the database.

As a result of expiration, the media that hold the data become fragmented and accumulate unused space. Reclamation is a server process that enables you to consolidate unexpired data from several fragmented media onto a few. This process makes sure of the automated recycling of media, minimizing the number of media required to be available.

1.2 TSM AND DEDUPLICATION

One of the key new features introduced in TSM V6.1 is deduplication. Deduplication is the identification and elimination of redundant copies of files or blocks. This feature can allow TSM to store more data in sequential or FILE type disk pools by removing duplicate blocks. Within TSM, deduplication is a two-stage process. The first stage identifies duplicate data within storage pools. TSM only supports FILE type disk storage pools for deduplication. This duplicate data identification process requires extra disk I/O and processor resources. Once the duplicates have been identified, they are removed in the second stage of the deduplication process. The removal is done via reclamation of the storage pool, or the movement or migration of data to another deduplication enabled storage pool.

Since TSM deduplication is another background process within TSM that can be highly CPU and I/O intensive, it might not be feasible for all environments, especially those that are hard-pressed to finish all of the TSM maintenance processes described above. The benefits of TSM deduplication might be limited as it is storage pool based. In other words, data in one storage pool is not compared against data in a separate storage pool. Additionally, since TSM is designed not to back up duplicate data by following an incremental forever backup methodology, the benefits achieved might not outweigh the extra processing required.

The NetApp VTL introduced support for deduplication in VTLOS 6.0 and offers the flexibility to run in both dedupe and nondedupe modes, allowing the administrator to optimize performance and storage efficiency to meet specific backup needs.

A significant benefit for using a NetApp VTL with TSM would be to remove the load from a TSM server by having the VTL perform deduplication. TSM has processing tasks that are performed throughout the day, and using the deduplication feature can add significantly to the load of the server. Performing deduplication on the VTL allows the TSM server to complete other tasks, and deduplication on the VTL is more optimized to deduplicate data across all of the data stored on the VTL, not just within a storage pool.

2   COMMON PROBLEM AREAS

The architecture described earlier works well as long as it is implemented correctly and enough physical resources are available to the TSM server. However, TSM administrators face several problems that eat up valuable operational cycles and escalate operational costs. Identifying these problems makes it easy to correlate which NetApp technologies are the best fit.

2.1   UNSCHEDULED MIGRATIONS

Unscheduled or uncontrolled migrations can happen because the defined disk pools do not have the capacity to hold one night's worth of backup data. This unscheduled migration from the disk pool to the next storage pool, usually the on-site tape pool, can have a domino effect on the environment. It can interfere with backups, causing them either to fail or to run longer, which in turn delays the copying process by forcing a tape-to-tape copy for off-site tape pools and also keeps the tape drives busy for on-site tape copies. Most companies have a service-level agreement (SLA) for when the backup tapes need to be shipped off-site, and unscheduled migrations can prevent an organization from meeting its SLAs.

2.2   TAPE STORAGE POOL COLLOCATION

Collocation refers to the ability of TSM to keep data together according to established rules. The most common form of collocation is by node, in which data from a client is kept on as few tapes as possible. TSM will only back up data to a particular set of tapes for each node or group of nodes depending on the type of collocation configured. In the absence of collocation, however, restore times can be extremely high and the number of parallel restores limited due to the large amount of mixing of data from several nodes onto the same set of tapes.

2.3   TAPE LIBRARY SHARING

A powerful feature of TSM is a native library manager that allows a TSM server to control a physical tape library and to share it among multiple TSM servers. Sharing a physical tape library has both advantages and disadvantages.

In an ideal world, only an optimal number of TSM servers would share a physical tape library, making sure that the physical resources are not overtaxed. In reality, however, physical tape libraries are often oversubscribed. One of the major disadvantages of this approach to sharing a physical tape library is the limited amount of time that each TSM server can have exclusive access to the tape library resources. The backup operations as well as the housekeeping functions need to happen within the allocated time slot for a particular TSM server. Any uncontrolled event in such an environment is potentially disruptive. This is why a VTL is useful as multiple virtual drives can be created to more efficiently perform parallel operations.

2.4   TAPE LIBRARY CAPACITY

Tape library capacity is a critical component that forces TSM administrators to come up with creative solutions to their backup challenges. A tape library has several resources, including robotic medium changers, tape slots, and tape drives. Many common TSM issues can be tracked to the capacity constraints of these resources.

The most important resource constraint on TSM environments is the number of slots available in the tape libraries that they connect to. When a tape library in a TSM environment reaches its slot limit, trouble sets in at several levels. TSM servers require new scratch tapes to be inserted regularly for writing new backup data; this needs to be done quickly to keep the TSM servers running. The most common temporary remedy for the slot problem is "local vaulting." This technique determines which tapes can be easily ejected from the library without breaking immediate functionality that the TSM server needs. After the tapes are ejected, they are kept on-site and handled manually from that point on. TSM can determine whether or not the tape is in the library. If it cannot find a tape that is required for an operation, it generates a message for the tape to be made available, and the operator can manually insert the required tape. This approach sounds simple and workable; however, it has serious long-term repercussions in a TSM environment, as discussed in the next sections.

2.5 LONGER CLIENT RESTORE TIMES

Because of the "progressive forever" nature of TSM backups, TSM has no full backups. To perform a full restore of a client, a full backup of the client is required, and in the worst case the data to be restored might reside on all or a large number of tapes in the library. Mounts and dismounts of these tapes, as well as seek times, can greatly increase restore times for a client. Also, if the tapes reside outside of the library, a restore request requiring those tapes will proceed at a snail's pace because the tapes must be manually inserted and then mounted. This can quickly escalate into a customer satisfaction issue.

2.6 INABILITY TO RECLAIM TAPES

As discussed earlier, reclamation is a server process that enables you to consolidate unexpired files from several fragmented media onto a few. This process makes sure of the automated recycling of media, minimizing the number of media required to be available. However, because this process does consolidation, it requires multiple media mounts to read the data off the old tapes and write it to new tapes. This becomes a problem in environments where tapes are outside the library because the reclamation process is painfully slow. Because the reclamation of tapes cannot occur fast enough for reuse, backup administrators are forced to keep buying new tapes.

3 VIRTUAL TAPE LIBRARIES

A virtual tape library is a disk-based storage system that appears like a tape library to a backup application but provides the superior reliability and performance of disk technologies. The VTL backup solution complements conventional tape drives and libraries by seamlessly inserting disk into the backup process, allowing existing tape libraries to be used for the creation of physical tapes for off-site storage.

Modern tape drives have high throughputs that can handle most backup and restores from clients; however, tape suffers from the inherent inefficiency of being a sequentially accessed medium, which means that in order to keep the drives performing at the highest throughput, data needs to be streamed to it at a high rate. Also, backups are only one side of the equation; restores are equally important, especially in a disaster situation. Disk-based solutions provide fast access compared to tapes while backing up, and they also provide random access to data, which means that all incremental backups can be restored more quickly from disk than from tape.

Because of the inherently "progressive forever" nature of TSM backups, there is a tendency for TSM to create fragmented tapes over time as the normal process of expiration deletes references to files on tapes. This in turn reduces the amount of useful data actually residing on any given tape, resulting in longer restore times. When a restore request comes in, the backup server determines which tape to mount in the tape drive and then locates the desired file. The effects of locating and relocating files become noticeable and time consuming when a full restore of a server is needed. Therefore, having incremental backups on disk can reduce the restore time when compared to tape.

To alleviate the problems just discussed, TSM introduced a FILE device class that creates a poor man's virtual tape library. It enables a TSM server to use disk as sequential tape media. However, a NetApp VTL solution provides several advantages over the FILE device class available in TSM.

Although both FILE device class disk pools and NetApp VTL use SATA disks to lower costs, using a NetApp VTL instead of FILE-type disk pools has the following benefits:

- Administration overhead: There is no need for host-side administration of the FILE-type disk pools destination file systems, which can be very time consuming and difficult to manage in a complex TSM environment.
- Performance: Unlike traditional file systems, the VTL OS is optimized for large-block sequential writes. This provides gains in cost and performance versus FILE-type disk pools, which sit on top of traditional file systems.
- Storage efficiency: Deduplication together with high-speed hardware-accelerated compression available on NetApp VTL provides better space utilization than disk pools as deduplication occurs on an appliance-wide level, resulting in lower cost per gigabyte.
- LAN-free backups: Backups are designed to work with tape-type devices. LAN-free backups with FILE-type disk pools are not common and not easily implemented. In an environment where LAN-free

backups to physical tape are present, the VTL virtual drives can easily be configured for LAN-free backups in the same way as the physical tape drives.

- Scalability: NetApp VTL allows simple, no-management addition of disk space and automated RAID group configuration.

3.1   NETAPP VTL SOLUTION OVERVIEW

NetApp VTL is a disk-based storage system that appears like a tape library to a backup application but provides the superior reliability and performance of disk technologies. The VTL backup solution complements conventional tape drives and libraries by seamlessly inserting disk into the backup process, allowing existing tape libraries to be used to create physical tapes for off-site storage. Based on proven technology developed specifically to address the needs of open systems storage networks, the NetApp VTL system delivers an industry-leading combination of scalability, cost savings, simplicity, and performance. It delivers all the benefits of disk-to-disk backup and recovery while seamlessly integrating into your current storage infrastructure.

NetApp VTL products are based on the popular FAS3000 series hardware platforms, although they run a different software microkernel than the FAS systems. The NetApp VTL is purpose built and tuned for backup and restores, providing up to 8.2TB/hr of throughput to meet tightening windows. The commonality of hardware platform allows support to cross-grade NetApp VTL systems with FAS3000, which means tremendous flexibility for platform repurposing—converting VTL systems into FAS storage or vice versa.

NetApp VTL added deduplication for TSM with VTLOS 6.0.1. The native deduplication is an intrinsic part of the operating system and runs natively on the VTL disk controller. It stores only unique data, letting you reduce disk requirements by up to 95% or more. This allows more backups to be retained online, and more servers can be protected while minimizing power and cooling requirements. NetApp VTL uses variable block deduplication combined with the security of byte comparison for maximum efficiency and data integrity. NetApp VTL enables deduplication to be selectively enabled on a per-virtual library basis. This allows the administrator to customize throughput and storage efficiency to meet specific backup requirements.

FULL TECHNOLOGY OWNERSHIP

Many VTL vendors build their products by using a variety of third-party OEM components and therefore do not have the direct control over their VTL technology that is necessary to rapidly respond to technical support cases and future customer requirements. NetApp has invested heavily to acquire and own all technological aspects of its VTL offering and has full control over both the customer satisfaction and technology roadmap aspects of what is delivered to market. Because the engineering expertise is behind the products in house, NetApp technical support can respond rapidly to customers, and NetApp can also evolve the product in line with the needs of customers. Customers control the roadmap of NetApp VTL.

In addition to these key points, NetApp VTL delivers many other benefits and is backed by the full portfolio of supporting infrastructure and services (maintenance and support, product education, and so on) that you would expect from an industry-leading vendor like NetApp.

4   INTEGRATING A NETAPP VTL IN AN EXISTING TSM ENVIRONMENT

The information presented here is in the context of a TSM server running on an IBM AIX machine. However, the same concepts apply generally to any platform running a TSM server. This section explains how the backup server is configured and then shows how easy it is to incorporate the NetApp VTL library in the mix without any disruptions to the existing environment.

4.1   METHODOLOGY OVERVIEW

VTL AS TAPE POOL REPLACEMENT

The preferred approach to integrating the NetApp VTL with TSM is to coexist with the existing primary disk pool (or a portion of it) and the off-site tape copy pool. Large client backups can benefit by going directly to the VTL; however, clients with a lot of small files can continue to go to a disk pool to be consolidated before going to the VTL. File size limits in the storage pool hierarchy can also be used to accomplish this by specifying a MAXSIze limit on the disk stgpool so that files larger than that limit will go to the next storage

pool, the VTL. As for the off-site tape copy pool, to preserve TSM's media lifecycle management, it is best to continue to have TSM control the physical tape resources and create the copy pool tapes. Typically, the physical tape library is not displaced by the NetApp VTL and continues to be the copy pool for off-site vaulting, managed directly by TSM, resulting in no change to the disaster recovery plan. There is no reason to have the VTL control physical tape creation with DTC[1].

The NetApp VTL typically replaces the primary tape pool. LAN-free clients are enabled by the cost-effective and easy configuration of virtual tape drives. This reduces the load on the LAN. The TSM daily maintenance tasks of reclamation and migration are enhanced, making TSM more reliable and improving restore performance. With the fast throughput of the NetApp VTL, file migration from the primary disk pool to the primary (NetApp VTL based) tape pool is also accelerated.

VTL AS DISK POOL REPLACEMENT
The NetApp VTL can also be used as a disk pool replacement. Data is still copied to physical tape for disaster recovery. For environments with many very slow clients, NetApp recommends keeping a portion of the disk pool.
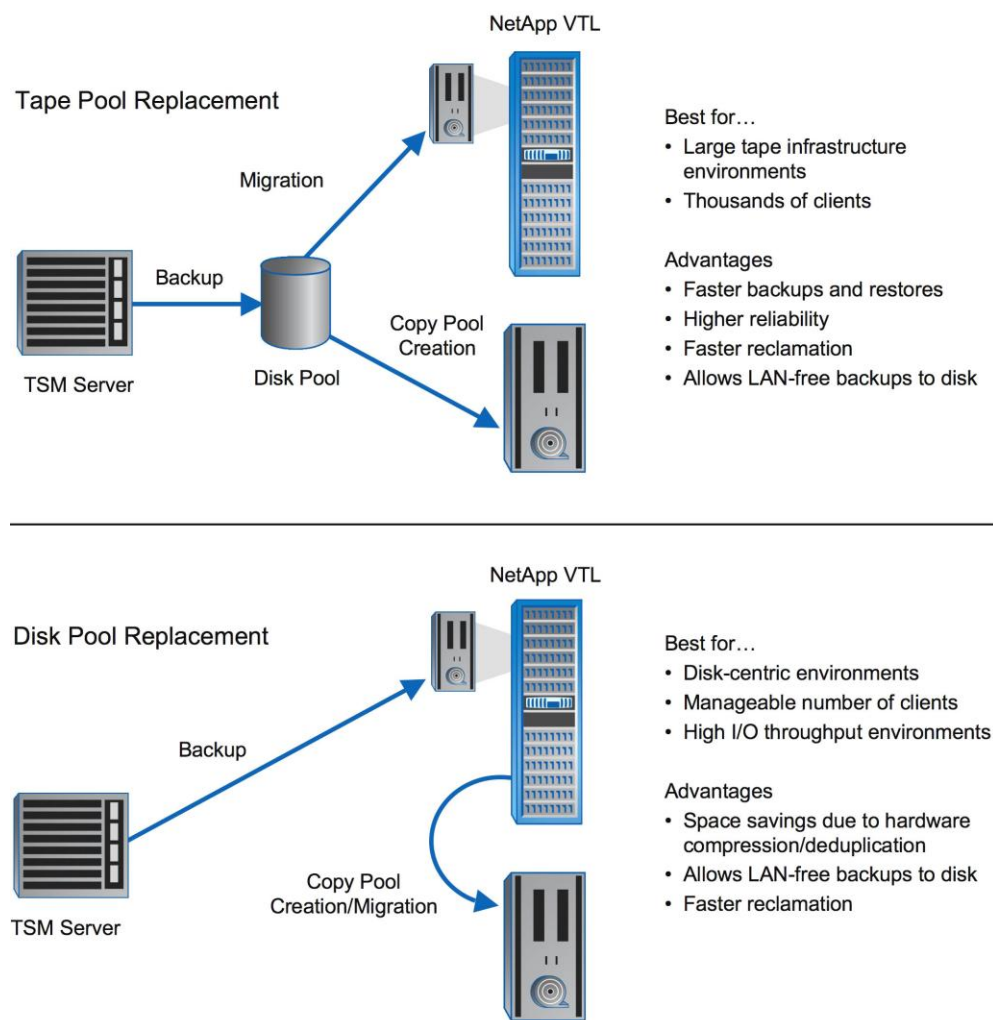

Figure 2) NetApp VTL use cases with TSM.

---

[1] Direct tape creation (DTC) is the feature of the NetApp VTL to create physical tape directly without needing a backup server.

TSM AND REPLICATION

Using any VTL to replicate data to another VTL is not recommended because there is the possibility that data at the remote site can get out of sync with the TSM database. Unlike other backup applications, the tapes created by TSM cannot be scanned and re-indexed into a new TSM database. If the TSM database is lost, the data on the tapes cannot be recovered. Similarly, if the database is not in sync with the tapes, some of the data will not be recoverable. Because TSM backs up only changed or new data, it can efficiently copy data to a copy pool at a remote data center. If dark fibre exists between data centers, the local TSM server can be configured to back up a storage pool to a copy pool on a VTL at the remote site. For IP connections between the two sites, the local TSM server can back up a storage pool to a copy pool, which points to virtual volumes (which can be a VTL) on a TSM server at the remote site.

EXAMPLE TSM HIERARCHY

The following example setup shows the typical storage hierarchy that is used in TSM environments

```
tsm: VTL_TEST>q stg

Storage       Device       Estimated   Pct   Pct   High  Low  Next Stora-
Pool Name     Class Name   Capacity    Util  Migr  Mig   Mig  ge Pool
                                                    Pct   Pct
-----------   ----------   ----------  ----  ----  ----  ---  -----------
ARCHIVEPOOL   DISK            8.0 M     0.0   0.0    90   70
COPYPOOL      LTO3            0.0 M     0.0
DISKPOOL      DISK              8 G     0.0   0.0    80   20   TAPEPOOL
TAPEPOOL      LTO3        9,999,000     0.0   0.0    90   70
                                  G
```

- DISKPOOL: A primary disk storage pool. Any backup data from clients is first received in the DISKPOOL. Data is held in the DISKPOOL for a predetermined period, after which it is migrated to the TAPEPOOL.
- TAPEPOOL: A primary tape storage pool that holds all of the on-site data from the clients. This pool is defined as the next storage pool for the DISKPOOL and therefore is the default target for all migrations. All restores from clients eventually occur from this pool, because the DISKPOOL is serving only as a short-time buffer and does not hold all data for clients.
- COPYPOOL: A copy tape storage pool. This pool is used for making off-site copies of the data in the DISKPOOL and the TAPEPOOL. After the data is written to tapes in this storage pool, those tapes are typically ejected from the library and sent off-site for DR purposes.

The following example setup uses a SCSI library with a medium changer. For the sake of completeness, the library, drives, and path definitions are also included.

```
tsm: VTL_TEST>q libr

        Library Name: TAPELIB
        Library Type: SCSI
              ACS Id:
    Private Category:
    Scratch Category:
WORM Scratch Category:
    External Manager:
              Shared: No
             LanFree:
   ObeyMountRetention:

tsm: VTL_TEST>q path

Source Name      Source Type      Destination      Destination      On-Line
                                  Name             Type
-----------      -----------      -----------      -----------      -------
VTL_TEST         SERVER           TAPELIB          LIBRARY          Yes
VTL_TEST         SERVER           TAPEDRIVE1       DRIVE            Yes
VTL_TEST         SERVER           TAPEDRIVE2       DRIVE            Yes

tsm: VTL_TEST>q dr

Library Name     Drive Name       Device Type      On-Line
-----------      -----------      -----------      -------------------
TAPELIB          TAPEDRIVE1       LTO              Yes
TAPELIB          TAPEDRIVE2       LTO              Yes
```

This example setup suffers from the common problems that were mentioned in the preceding sections. Keeping the advantages of a VTL solution in mind, to alleviate those problems VTL is introduced into the existing TSM environment. To complete that seamless integration, follow the steps outlined in the following sections.

4.2   LIBRARY CONFIGURATION ON THE NETAPP VTL

The NetApp VTL fits seamlessly into a TSM environment. As mentioned earlier, it appears to TSM just like another physical tape library, but with superior reliability and performance because it leverages the speed of disk. Using the NetApp VTL as the main on-site primary tape pool alleviates the aforementioned problems that exist in many TSM environments.

When implementing a VTL in a TSM environment, it is generally best to continue to use the existing disk pool as the first landing place for the client backups. Large file backups can go directly to the VTL. When the backups are complete, the off-site copies can be made by copying from the primary disk pool or VTL to physical tape. The data in the disk pool is then migrated to the VTL instead of to physical tape to make room for the next night's backup.

There are many benefits to having the primary data on VTL instead of physical tape. Because tape mounts, dismounts, and seeks happen instantaneously on the VTL, client restore times are dramatically reduced. And because a large number of virtual drives can be defined on the NetApp VTL, more TSM processes, such as reclamation, can run simultaneously, allowing the processes to complete more quickly. Physical tape resource contention is also alleviated; because the migration and reclamation maintenance processes occur on the VTL, the physical tape resources can be used only to create the disaster recovery copies.

Figure 3 shows the port assignments for a NetApp VTL700 controller. Simply connect one or more of the four host-facing ports to the SAN so they can be zoned to the TSM servers.
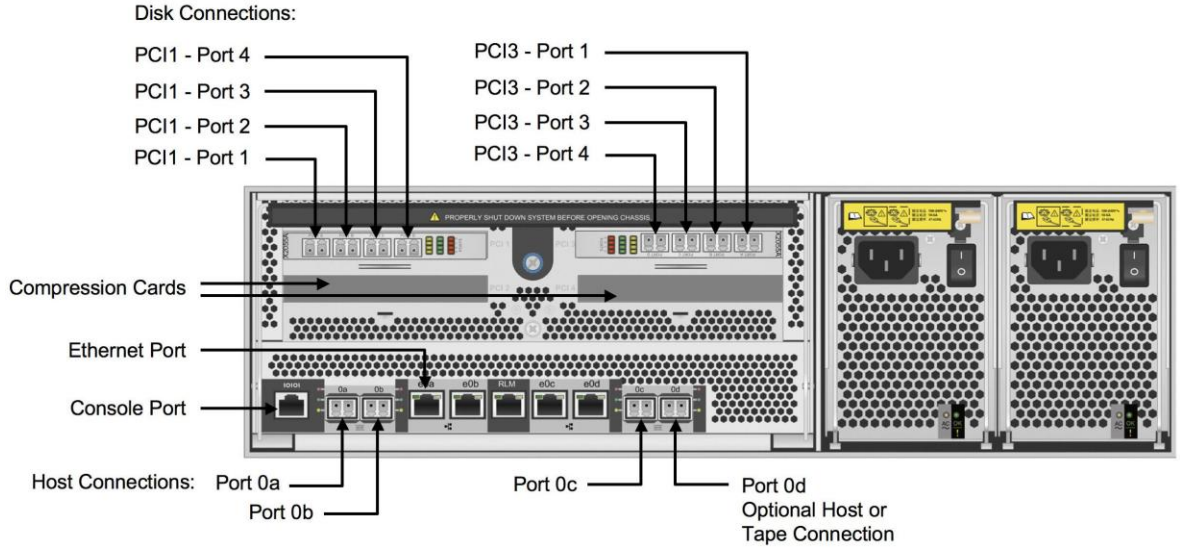
Figure 3) NetApp VTL700 port assignments.

When the initial setup of the NetApp VTL and connections to the backup SAN are complete, configuring and deploying virtual libraries are easy using the intuitive NetApp VTL GUI. All that is required is going through a few steps using the GUI, and a virtual library is ready to use by TSM within a few minutes.

VTL DEDUPLICATION CONFIGURATION

The NetApp VTL introduced deduplication support for TSM starting with VTL OS 6.0.1. To make sure of optimal storage efficiency for TSM environments, the backup application must be set to TSM in the deduplication configuration screen. NetApp recommends adjusting the deduplication post processing window to start after the TSM maintenance processes (migration, reclamation) have completed and leaving the maximum post process request size as the default, 5000GB (unless otherwise instructed by NetApp Professional Services/Support).
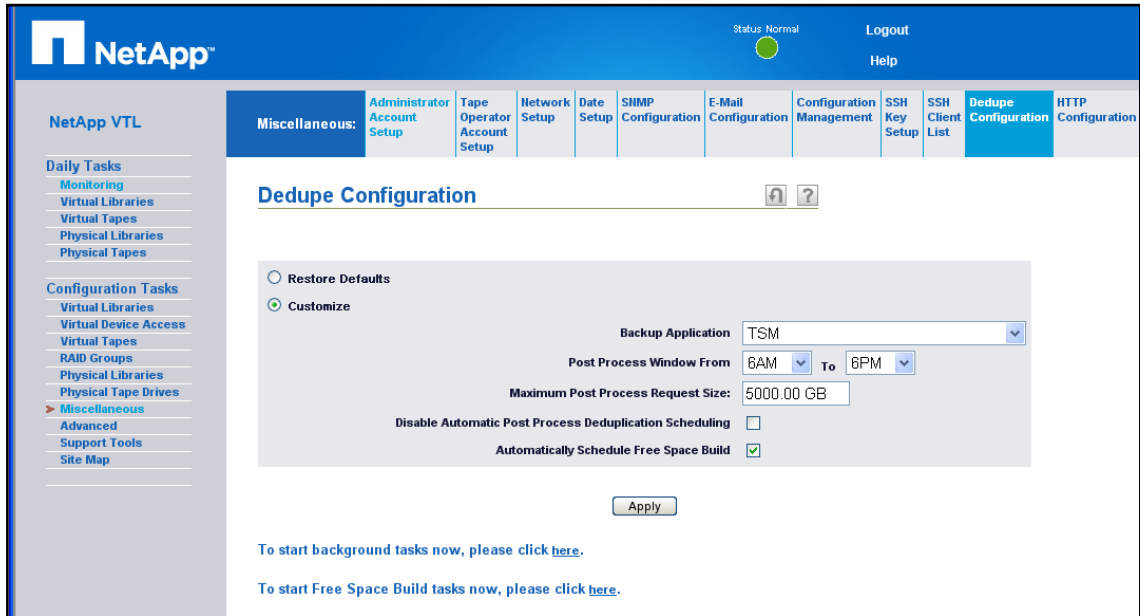

Figure 4) NetApp VTL V6.0.3 deduplication configuration screen.

CREATE CUSTOM VIRTUAL DRIVE TYPE

Before a virtual library can be created, it must first be decided what type of drive emulation to use. For TSM v5.5.1 and higher environments, NetApp recommends creating a custom STK T10000A virtual drive type with size of 100GB raw capacity.

Field Name	Value
Virtual Tape Drive Type Name	STK T10000A non-DTC
Raw Capacity (GB)	100
Vendor ID	STK
Product ID	T10000A
Tape Smart Sizing	None
Disk Allocation Per Tape (GB)	100

This means that the virtual tapes created will report that they are full (end-of-media) when TSM has written 100GB to a tape. The raw capacity can be set to a higher or lower value. It is not recommended to use a raw capacity that is very small in size (less than 50GB) since the maximum number of tapes that can be created on a single VTL controller is 10,000. For example, on larger systems, if the tape size is too small, the VTL disk space might not be effectively utilized even if 10,000 tapes have been created. A raw capacity of 100GB is optimal for most environments.

The vendor ID and product ID must match the values above. Tape Smart Sizing is a NetApp VTL feature that optimizes tape usage when it is configured to create physical tapes. This feature is not used and thus set to NONE in TSM environments since TSM, not the VTL, is configured to create physical tapes. The disk allocation per tape specifies the maximum amount of disk space that can be allocated for this tape.

Figure 5) NetApp VTL create virtual tape drive type screen

Optionally, a similar custom template can be created based on the IBM LTO4 inquiry strings. This would be for environments where it is desired to use the IBMTape driver for the virtual drives.

Field Name	Value
Virtual Tape Drive Type Name	IBM LTO4 non-DTC
Raw Capacity (GB)	100
Vendor ID	IBM
Product ID	ULTRIUM-TD4
Tape Smart Sizing	None
Disk Allocation Per Tape (GB)	100

CREATE VIRTUAL LIBRARY

Now that a virtual drive type has been created, virtual libraries can be defined. Whether or not data is eligible for deduplication is set at the virtual library level. A virtual library can be set as deduplication enabled or hardware compression only. If set to deduplication, all data backed up to virtual tapes in this virtual library will go through the deduplication process where duplicate data is identified and eliminated. If set to hardware compression, all data backed up to virtual tapes in this virtual library will be stored compressed on disk, but will not go through the deduplication process.

For TSM environments, NetApp recommends creating two or more virtual libraries; a virtual library set to deduplication and a separate one set to hardware compression. Clients backing up file system data should be configured to back up data to a virtual library that is set to hardware compression. Since TSM follows an incremental forever backup methodology and does not do repeated full backups of file system data, it is not expected for there to be a great deal of duplicate data for these normal backups. The bulk of the data reduction will be due to compression of the data, and any additional data reduction due to deduplication is expected to be marginal. Clients using client encryption or compression should also be directed to the hardware compression virtual library as this type of data is not expected to dedupe well. In contrast, application server (such as Oracle® and Exchange) and NDMP backups where repeated full backups usually occur should be configured to back up data to a virtual library that is set to deduplication.

Note: When creating a virtual library, it is best to create one with extra empty slots to minimize the amount of reconfiguration required in both TSM and the VTL when storage is added to the VTL.

Figure 6) NetApp VTL create virtual library screen.

Field Name	Value	Description/Notes
Library Nickname		Name of the virtual library.
Library Type	NetApp	The robot type of virtual library. For TSM v5.5.1 or later, select NetApp. For earlier versions please contact NetApp Support.
Virtual Tape Drive Type	STK T10000A non-DTC	Select the newly created template.
Automatically assign "First …" values	Not checked	Clear this checkbox to manually assign the element address ranges for the virtual library.
Slot Count	*Number of slots desired*	NetApp recommends creating extra empty slots to reduce the amount of reconfiguration needed when storage is added to the VTL. Note: The Virtual Library GUI will only display virtual libraries with 1,024 slots or less. Manipulation of virtual libraries with more than 1,024 slots cannot be done using the GUI but can be done using the CLI.
First Slot	100	The element address for the first slot. NetApp recommends making this range sufficiently higher than the drive element address range.
Drive Count	*Number of drives desired*	The number of tape drives for this library. NetApp recommends creating a total of 60 virtual drives for all defined virtual libraries to make sure maximum concurrent stream count is not exceeded for a single appliance.
First Drive	4	The element address for the first drive. NetApp recommends making this range lower than the slot element address range.
Mailbox Count	1	A mailbox is the entry/exit port for the virtual library. A virtual library can have only one entry/exit port. This field cannot be modified.
First Mailbox	0	The element address for the mailbox. The element address for the mailbox should not fall in the drive or slot element address range.
Transport Count	1	A transport is a picker. Because this is a virtual library, there is no benefit to defining more than one transport. However, this gives the option of matching the geometry of a physical library.
First Transport	1	The element address for the first transport. The element address for the transport should not fall in the drive or slot element address range.
Assigned to Port – WWPN		Select the host port through which the virtual library is accessed. The virtual device host masking and port assignment can be changed after the virtual library is created.
Fully Loaded with Virtual Tapes	Not checked	Specifies whether the NetApp VTL will attempt to create a new virtual tape for each slot in the virtual library. If this checkbox is selected, the starting label for these virtual tapes must be specified. The labels of additional volumes increment automatically. Since virtual tapes will be created after creating the virtual library, clear this checkgox.
Starting Tape Label	Leave Blank	The starting label of the virtual tapes. Always use uppercase (all capitals) for any letters in the virtual barcodes. However, since virtual tapes will be created later, leave this field blank.
Disk Compression Type		Select the type of compression desired: deduplication or hardware compression.
Tape Smart Sizing	Not Checked	This feature optimizes physical tape usage when the VTL is creating physical tapes. Because TSM, not the VTL, creates the off-site physical tape copies, clear this checkbox.
Automatically Move Tapes from EEP to slot	Checked	When this checkbox is selected, the virtual library automatically moves virtual tapes from the entry/exit port to an available slot when new virtual tapes are created. This is selected by default, and NetApp recommends leaving it selected.
Enable Tape Cloning	Not Checked	For use only in cases where the VTL creates physical tape copies. Because TSM creates the off-site physical tape copies, clear this check box.
Disk Allocation Per Tape	Default (100)	Specifies the maximum amount of disk space is allocated per tape. Leave this option at the default, 100GB.

CREATE VIRTUAL TAPES

NetApp recommends initially creating a number of tapes that will be equivalent to allocating 70% of the total capacity of the VTL after RAID group creation (total disk capacity/disk allocation per tape * 70%). For example, for a system with 50TB of total disk capacity as reported in the VTL monitoring window, create 50000GB/100GB * 70% = 350 virtual tapes. After data has been backed up and deduped, more virtual tapes can be created to take advantage of the space freed up by dedupe and compression.

Note: When creating virtual tapes, any letters in the virtual barcodes must be uppercase (all capital letters). TSM does not function properly if barcodes contain lowercase letters.

4.3   CONFIGURING THE VIRTUAL LIBRARY ON A TSM SERVER

When the virtual library has been configured on the NetApp VTL, you must incorporate the newly defined VTL into the existing or new environment by configuring the TSM environment as shown in Figure 7.
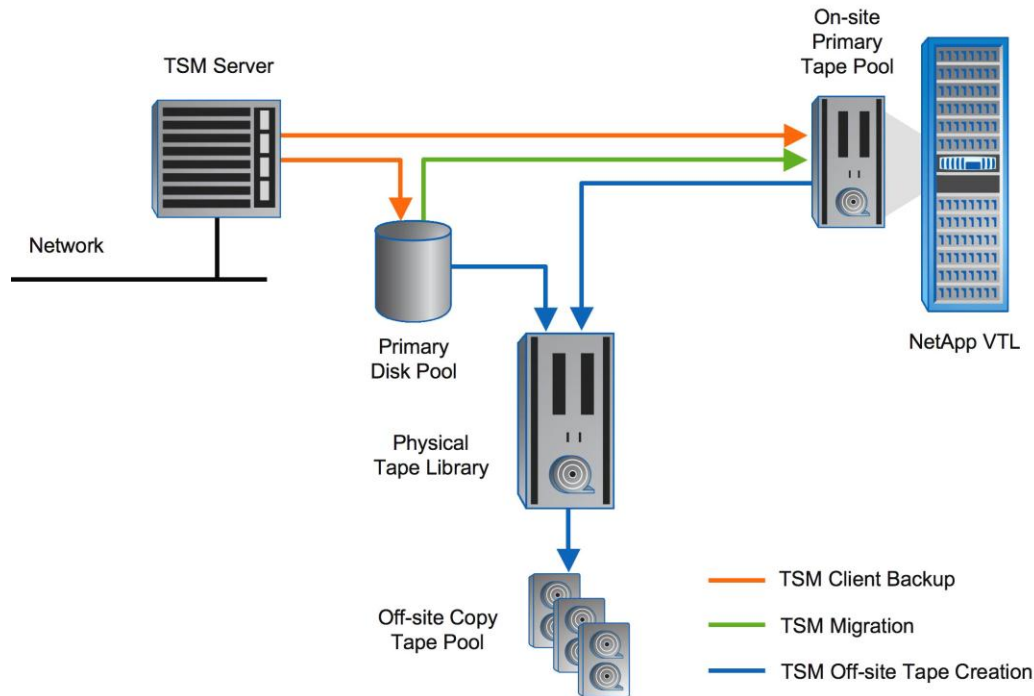

Figure 7) Reconfigured TSM architecture.

INSTALLING AND CONFIGURING DEVICE DRIVERS

Appropriate device drivers must be installed before you can use a tape or optical device on any platform that supports IBM Tivoli Storage Manager. For information on which drivers to install and how to install them, see the IBM Tivoli Storage Manager and platform documentation. The example setup used IBM AIX, which required the atape, atldd, as well as the TSM device drivers to be installed. For information on device drivers for TSM, see the IBM Web site.

ZONING THE VIRTUAL LIBRARY TO THE HOST

With the virtual library created, the TSM servers must be zoned properly to the VTL. The NetApp VTL700 has four 4Gb/sec host-facing ports through which virtual devices can be defined. It is possible to provide host access control by restricting access of certain devices to one or more initiator World Wide Port Names (WWPNs). Similarly, virtual devices can be mapped to one or more host-facing ports. In an environment where there are multiple TSM servers (or the TSM server has multiple HBAs), using a combination of host masking and port mapping on the NetApp VTL to spread the virtual drives across the host-facing ports yields the best results. For more information about virtual device mapping, see the NetApp VTL System Administration Guide.

After the virtual devices are defined across the host-facing ports, simply zone the TSM server HBAs to a specific host port on the NetApp VTL. See Figure 8 for an example zoning configuration.
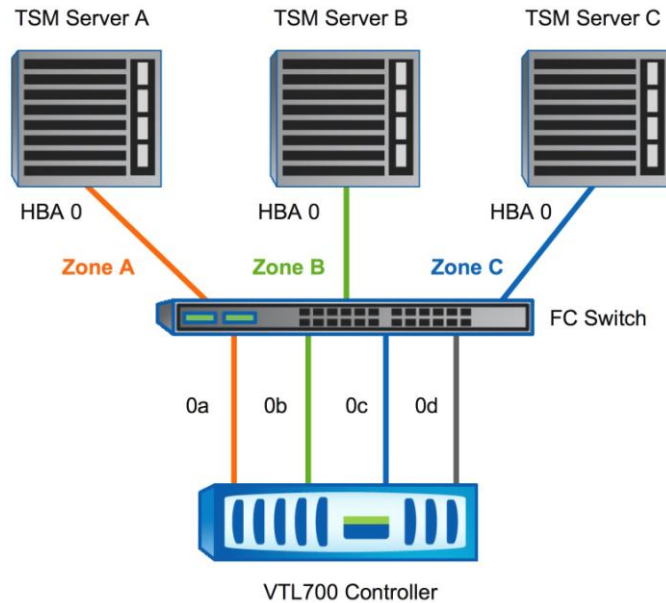

Figure 8) Example zoning configuration.

DISCOVERING THE DEVICES FROM THE HOST

The detailed discovery process is beyond the scope of this paper. For details on how to discover new devices, please refer to the appropriate product documentation. This process varies depending on what platform and operating system you are using. After running the device discovery process, you should see a device for the library controller and devices for all the tape drives that were configured in the virtual tape library and connected to this host.

DEFINING THE LIBRARY TO TSM

Setting up a virtual tape library is the same as setting up any non-IBM SCSI tape library. The following steps describe the setup of a library virtualized in the NetApp VTL with TSM 5.5.2 or later running.

First start by defining the library and then defining virtual tape drives (repeating for every virtual tape drive by running the commands each time changing the drive name [VTAPE1 in the example] and `device=` arguments). The output below shows adding four drives. The DEFINE DRIVE and DEFINE PATH commands would need to be repeated for each drive in the virtual library.

```
tsm: VTL_TEST>define library vtllib libtype=scsi shared=no
               relabelscratch=yes
ANR8400I Library VTLLIB defined.

tsm: VTL_TEST>define path vtl_test vtllib srct=server destt=library device=
               /dev/lb1 online=yes
ANR1720I A path from VTL_TEST to VTLLIB has been defined.

tsm: VTL_TEST>define drive vtllib vtape1 element=100
ANR8404I Drive VTAPE1 defined in library VTLLIB.

tsm: VTL_TEST>define path vtl_test vtape1 srct=server destt=drive libr=
               vtllib device=/dev/mt4
ANR1702I A path from VTL_TEST to VTLLIB VTAPE1 has been defined.

tsm: VTL_TEST>define drive vtllib vtape1 element=101
ANR8404I Drive VTAPE2 defined in library VTLLIB.

tsm: VTL_TEST>define path vtl_test vtape2 srct=server destt=drive libr=
               vtllib device=/dev/mt5
ANR1702I A path from VTL_TEST to VTLLIB VTAPE2 has been defined.

tsm: VTL_TEST>define drive vtllib vtape1 element=102
ANR8404I Drive VTAPE3 defined in library VTLLIB.

tsm: VTL_TEST>define path vtl_test vtape3 srct=server destt=drive libr=
               vtllib device=/dev/mt6
ANR1702I A path from VTL_TEST to VTLLIB VTAPE3 has been defined.

tsm: VTL_TEST>define drive vtllib vtape1 element=103
ANR8404I Drive VTAPE4 defined in library VTLLIB.

tsm: VTL_TEST>define path vtl_test vtape4 srct=server destt=drive libr=
               vtllib device=/dev/mt7
ANR1702I A path from VTL_TEST to VTLLIB VTAPE4 has been defined.
```

Note: TSM introduced the RELEABELSCRATCH parameter in V5.5.1, which automatically relabels a tape when it goes back to the scratch pool. This feature is useful for freeing up space on expired tapes as soon as it is expired rather than waiting until it is reused. However, TSM V5.5.1 had some issues with this parameter, and IBM recommends using this parameter for TSM V5.5.2 and later (APAR IC58862).

CREATING A NEW DEVICE CLASS

After creating the virtual tape library and defining the tape drives, follow these steps to create a new device class.

1. Define a device class to be used to define a new storage pool of this device type class. The new pool will be used to migrate data to the VTL. To determine what type of device class you need to define, run the following command.

```
tsm: VTL_TEST>q dr f=d

 Library Name: VTLLIB
   Drive Name: VTAPE1
  Device Type: ECARTRIDGE
      On-Line: Yes
 Read Formats: T10000A-C,T10000A
Write Formats: T10000A-C,T10000A
      Element: 100
```

2. To define the new device class, issue the following command.

```
tsm: VTL_TEST>define devc ecart libr=vtllib format=T10000A-C estcapacity=
                    100g mountretention=0 mountwait=20 mountlimit=
                    drives devt=ecartridge
ANR2203I Device class ECART defined.
```

Note: Set mountretention to 0 so that the virtual tapes are unloaded immediately after TSM has finished using them. This setting allows the tape to reside in the drive for a certain amount of time so that the TSM server does not have to remount the tape. Mounting and unmounting of tapes are instantaneous for the VTL, so there is no need to set a delay in unmounting a tape.

3. Follow these steps to define a new storage pool by using the newly created ECART device class. This storage pool will be used to migrate data from the disk pool and local physical tapes to VTL virtual tapes.

```
tsm: VTL_TEST>define stg vtlpool ecart pooltype=primary description
              ="VTL Pool" maxscratch=9999
ANR2200I Storage pool VTLPOOL defined (device class ECART).
```

The TSM storage hierarchy on your server now looks like this.

```
tsm: VTL_TEST>q stg

Storage      Device       Estimated    Pct    Pct   High  Low  Next Stora-
Pool Name    Class Name   Capacity     Util   Migr  Mig   Mig  ge Pool
                                                     Pct   Pct
-----------  ----------   ----------   -----  ----- ----  ---  -----------
ARCHIVEPOOL  DISK            8.0 M      0.0    0.0    90    70
COPYPOOL     LTO3            0.0 M      0.0
DISKPOOL     DISK              8 G      0.5    0.5    80    20  TAPEPOOL
TAPEPOOL     LTO3        9,999,000      0.0    0.0    90    70
                                G
VTLPOOL      ECART           0.0 M      0.0    0.0    90    70
```

4. Label VTL volumes and check them into the library for immediate use.

```
label libvol vtllib search=yes labelsource=barcode checkin=scratch
ANS8003I Process number 18 started.
```

Query the library to verify that the tapes have been labeled and are ready to use.

```
tsm: VTL_TEST>q libvol

Library Name  Volume Name  Status      Owner      Last Use   Home
Device
                                                             Element
Type
------------  -----------  ---------   ----------  ---------  -------  ---
---
VTLLIB        VT0000       Scratch                                    200
VTLLIB        VT0001       Scratch                                    201
VTLLIB        VT0002       Scratch                                    202
```

You can run a quick database backup to check how the VTL tape appears to the TSM server. Notice that from the activity log the VTL appears and operates just like a physical tape library. All of the TSM messages indicate the same thing.

```
tsm: VTL_TEST>backup db type=full devc=lto vol=VT0000
ANR2280I Full database backup started as process 19.
ANS8003I Process number 19 started.

tsm: VTL_TEST>q pr

 Process Process Description Status
  Number
-------- ------------------- --------------------------------------------
----
      19 Database Backup     Full backup: 0 pages of 3368 backed up.
Waiting
                                for mount of output volume VT0000 (1
                                seconds).
```

Messages similar to the following appear on the TSM console.

```
ANR0984I Process 19 for DATABASE BACKUP started in the BACKGROUND at 00:07
ANR2280I Full database backup started as process 19.
ANR20171 Administrator CHAUDRY issued command: QUERY PROCESS
ANR8337I ECARTTRIDGE volume VT0000 mounted in drive VTAPE2 (/dev/mt5).
ANR05131 Process 19 opened output volume VT0000.
ANR13601 Output volume VT0000 opened (sequence number 1).
ANR4554I Backed up 1664 of 3368 database pages.
ANR4554I Backed up 3328 of 3368 database pages.
ANR1361I Output volume VT0000 closed.
ANR05151 Process 19 closed volume VT0000.
ANR4550I Full database backup (process 19) complete, 3368 pages copied.
ANR0985I Process 19 for DATABASE BACKUP running in the BACKGROUND complete
with completion state SUCCESS at 00:07:47
```

Whether you are defining the NetApp VTL as an additional tape library into an existing TSM environment or defining the library for a new installation, the process just described is the same. It is generally desirable to keep a disk storage pool to receive data on a daily basis and use the VTL as the next storage pool.

4.4   USING THE VTL

The next step is to start using the newly created VTL. First, you need to eliminate the on-site tape pools. This requires migration of the existing data that is residing on the local physical tapes to the VTL and changes in the policies to make sure that the new data gets written to the VTL. There are several ways to accomplish this migration. The following sections describe some of the possible ways to use the VTL tape pools.

4.5   REDIRECTING NEW BACKUPS TO THE VTL TAPE POOL

Although the VTL tape pools can be used without a front-end TSM disk pool, it is generally a good practice to send large backups directly to the VTL tape pool and use a front-end disk pool for backups with a large number of small files to optimize performance.

To send backups directly to a VTL tape pool, you must change the copy group definition on the TSM server for the domain to point to a VTL tape pool. In the following example, the copy destination in the backup copy group of a domain called VTLTEST is changed from DISKPOOL to VTLPOOL. The copy group being modified is part of the default management class of the domain, and therefore it affects all new backups.

```
tsm: VTL_TEST>q co vtltest standard standard f=d

                  Policy Domain Name: VTLTEST
                     Policy Set Name: STANDARD
                     Mgmt Class Name: STANDARD
                     Copy Group Name: STANDARD
                     Copy Group Type: Backup
                Versions Data Exists: 2
               Versions Data Deleted: 1
               Retain Extra Versions: 30
                  Retain Only Version: 60
                            Copy Mode: Modified
                   Copy Serialization: Shared Static
                       Copy Frequency: 0
                     Copy Destination: DISKPOOL

tsm: VTL_TEST>upd co vtltest standard standard destination=vtlpool
ANR1532I Backup copy group STANDARD updated in policy domain VTLTEST, set
STANDARD, management class STANDARD.

tsm: VTL_TEST>validate po vtltest standard
ANR1515I Policy set STANDARD validated in domain VTLTEST (ready for
activation).

tsm: VTL_TEST>activate po vtltest standard
Do you wish to proceed? (Yes (Y)/No (N)) y
ANR1514I Policy set STANDARD activated in policy domain VTLTEST.
```

These commands direct all backups coming into the VTLTEST domain to the VTLPOOL (VTL tape pool in the test setup). You can easily verify this by running a backup from a client defined in that domain (in the example, VTLTEST domain). Large file clients can be registered to this domain.

In contrast, if you use a disk pool to receive data and later decide to migrate the data to a VTL pool, the next storage pool definition must be changed to point to a VTL tape pool for the existing disk pool, as shown in the following example.

```
tsm: VTL_TEST>q stg

Storage         Device       Estimated    Pct    Pct   High  Low  Next Stora-
Pool Name       Class Name   Capacity     Util   Migr  Mig   Mig  ge Pool
                                                        Pct   Pct
-----------     ----------   ----------   -----  ----- ----  ---  -----------
ARCHIVEPOOL     DISK              8.0 M    0.0    0.0    90   70
COPYPOOL        LTO3              0.0 M    0.0
DISKPOOL        DISK                8 G    0.5    0.5    80   20   TAPEPOOL
TAPEPOOL        LTO3         9,999,000     0.0    0.0    90   70
                                   G
VTLPOOL         ECART        2,047,795     0.0    0.0    90   70
                                   G

tsm: VTL_TEST>upd stg diskpool nextstg=vtlpool
ANR2202I Storage pool DISKPOOL updated.
```

These commands direct data from DISKPOOL to VTLPOOL when a migration is triggered on DISKPOOL.

4.6 MIGRATING DATA FROM PHYSICAL TAPES TO VTL TAPE POOL

The next step in the process is to make sure that the data that is currently written to physical tapes gets migrated to the VTL tapes. There are several ways to accomplish this migration; this section describes the three most common methods. For details on the syntax of the commands discussed in this section, refer to the product documentation.
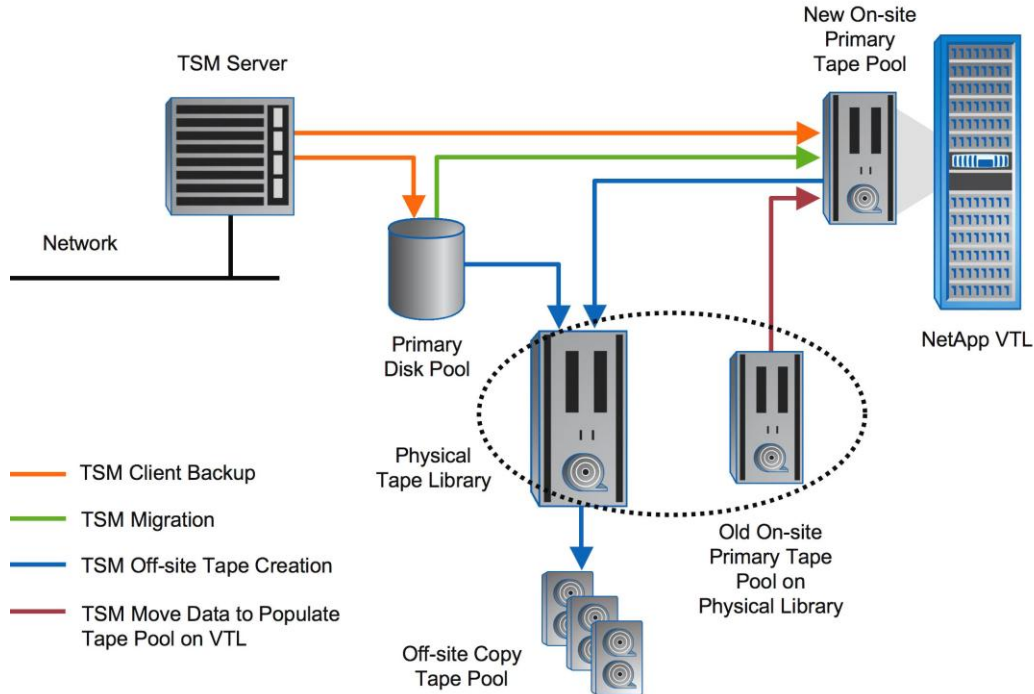

Figure 9) Migrating data from physical to VTL tapes.

- **Move Data:** Use this command to move physical tape contents to the VTL tape pool. This method is ideal when some tapes reside outside the library and more granular control over the migration process is desired. You can easily automate this process by writing some simple scripts.

- **Move Node Data:** Use this command to move physical tape contents to the VTL tape pool. This method is good when you need all tapes for a certain node or group of nodes to be migrated to VTL at the same time. A practical consideration for using this method is to see whether the data has been collocated for a node or group of nodes. If collocation is not turned on, this method might perform a large number of media mounts as opposed to reducing them. Since data related to a node or a group of nodes is moved together, this method provides more predictability to the TSM administrator for managing TSM resources. For a complete syntax of the command and the granularity that you can choose to move the contents of nodes, see the TSM documentation.

- **Migrate Storage Pool:** You can also use storage pool migration to move contents from physical tapes to a VTL tape pool. This method is ideal if no tapes that reside outside the library. This method provides less predictability for the TSM administrator because it moves the contents of a tape storage pool sequentially to the VTL storage pool. Therefore, any restores in between the transition end up getting some data from the VTL tape pool and some from the old physical tape pool.

5 CONCLUSION

The NetApp VTL provides an intuitive, nondisruptive, easy way to implement disk-to-disk backup solutions. The process of introducing a NetApp VTL is totally nondisruptive to your existing TSM environment.

A virtual tape library solution can help alleviate the most common problems associated with TSM by removing the physical bottlenecks, and it can also lower the recurring costs associated with tape or even eliminate tape media altogether and lower operational costs significantly by copying to a copy tape pool located in a remote NetApp VTL connected over a long-distance FC/IP link.

The NetApp VTL dramatically improves the ability to complete backups within the backup window and delivers enhanced reliability by removing mechanical handling of tapes from the backup window. The addition of optimized RAID VTL protection provides another level of protection to TSM users who use physical tape exclusively or who use FILE device classes on unprotected storage. Deduplication in conjunction with high-speed, hardware-based compression dramatically increases disk-based retention for backups, enabling more backup data to be made available for high-speed disk-based restores. NetApp VTL optimization of physical tape performance and capacity extends the useful lifespan of existing tape library hardware and media.

For more information on NetApp VTL, visit *www.netapp.com*.