Virtual Tape: An Overview

BY LLOYD CHRISTENSEN

Virtual tape systems are a viable solution to the problem of continued growth in customer tape libraries. This article addresses two virtual tape solutions: IBM's Virtual Tape Server and StorageTek's Virtual Storage Manager.

I am a storage consultant for StorageTek's Solutions Business Group. Since joining STK a year ago, most of my projects have entailed helping our clients implement our Virtual Storage Manager (VSM). Prior to joining STK I worked as a contractor doing general storage administrator functions and as a subcontractor through IBM implementing IBM's Virtual Tape Server (VTS) for a large government agency. Although I will try to treat both vendors fairly, I will primarily use StorageTek terminology throughout this article.

TAPE GROWTH AND TMM

The continuing growth in mainframe data storage includes an increase in the number of tapes data centers have to manage. Hardware vendors have increased the reliability, speed and storage capacity of tape media. In 1985, IBM introduced its first cartridge tape, the 3480, which without compression had a capacity of about 200MB. Fifteen years later, IBM's Magstar and StorageTek's 9840 cartridges exceed 20GB without compression and with compression can exceed 80GB.

With the increased capacity came a new problem. Although much more data is being retained longer, there is substantially no difference in the size of individual files and the number of files on a tape has remained constant. Surveys show that in most data centers 50 percent of all tapes store less than 50MB and 80 percent of all tapes store files less than 800MB.

The tapes themselves require storage in tape racks. Data centers with more than 300,000 tape volumes are by no means rare, and the floor space necessary to store these tapes is tremendous. The cost of that floor space can also be tremendous. Additionally, physical handling of these tapes is manpower-intensive and expensive. A part of the solution to the problem of "where do we put these tapes" is to more effectively use the tapes.

TAPE MOUNT MANAGEMENT

In his October 1999 "Storage Strategies" column (*Technical Support*) Steve Pryor writes about a strategy that became common in the late 1980s and early 1990s called Tape Mount Management (TMM). In TMM, JCL changes or DFSMS routines are used to direct small tape data sets to a disk buffer and then migrate all of the data sets on the buffer to one tape via DFHSM or similar products.

However, TMM has problems of its own. Maintenance of DFSMS routines to direct the data to disk is required. The data sets must be in the ICF Catalog. Honoring the data owner's retention criteria becomes difficult and may involve politically difficult retention standards. The migration from disk to tape takes a lot of CPU cycles, as does the recalling of data sets if they are needed. Because the data sets will not expire at the same time, de-fragmentation of the migration volumes (a DFHSM recycle of the migration tapes) is required and also burns CPU cycles. Disk buffers exceeding 100GB are not unusual.

VIRTUAL TAPE

In 1997, IBM introduced a new solution: the Virtual Tape Server. In December 1998, StorageTek made its answer to VTS available, the Virtual Storage Manager. Other vendors with their own virtual tape solutions include Sutmyn, EMC², and a software-based solution from Sterling Software called SAMS:VTAPE. The remainder of this article will address only IBM's VTS and StorageTek's VSM.

Each solution involves connecting a disk buffer to that vendor's automated tape library. The IBM VTS uses the 3494 library while the StorageTek VSM uses their libraries (4410, 9310, 9360, and

9740). They each offer options on the size of the disk buffers. IBM uses a RISC processor to manage its VTS while StorageTek uses enhancements to the Host Software Component (HSC), host-based software that controls the automated tape library. The enhancements to HSC are called Virtual Tape Control System (VTCS). VTCS does use some host CPU cycles but with not nearly the impact as TMM. IBM allows up to two of its control units per physical tape library (3494) while Storage-Tek allows up to 16 of its VTSSes per ACS (an ACS is a group of one to 16 silos interconnected). There may be up to 16 ACSses so they can get as many as 256 VTSSes in one MVS environment.

Each solution involves defining up to 64 virtual tape drives (VTDs) to the host system (MVS). The controllers are used to emulate 3490 model tape drives. The disk buffers are used to emulate 3490 tape volumes. These tape volumes are defined to your tape management system. For StorageTek, each Virtual Tape Volume (VTV) is defined as a standard length 3490 cartridge (400MB). IBM allows you to specify either standard or enhanced length 3490s (400MB or 800MB) and mix them. Both vendors have a compression feature that is comparable to IDRC/ICRC and should see at least a 3:1 compression ratio. IBM's base VTS model allows defining up to 32 virtual tape drives. IBM's ESCON High Performance Option (EHPO) allows defining up to 64 virtual tape drives. All StorageTek models allow 64 virtual tape drives to be defined. The IBM 3494-tape library has a limit of 300,000 virtual volumes. StorageTek has no limit as to how many virtual volumes can be defined; however, they have a limit of 100,000 virtual volumes residing on the disk buffer.

Virtual volumes are defined differently for each vendor. The VTV for an IBM VTS is defined through panels on the RISC processor and it's necessary to define a new catalog structure called a VOLCAT. VOLCAT entries relate the VTV to the VTS controller. For StorageTek, VTVs are defined via a CONFIG utility that updates the HSC control data set (CDS) to insert new volume records and associate which VTV is affiliated with which VTSS. Virtual volumes in an IBM VTS cannot be used by another VTS. For StorageTek, virtual volumes created in one VTSS may be used on virtual drives associated with another VTSS. A VTSS may also be connected to up to two ACSses.

Alphabet Soup

ACS:	Automated Cartridge System (one or more attached libraries)
CA-1:	Computer Associates' Tape Management System
CDS:	StorageTek HSC Control Data Set
DBU:	Disk Buffer Utilization (how full is the disk buffer?)
DFHSM:	IBM's Hierarchical Storage Manager
DFSMS:	IBM's System Managed Storage
ExPR:	StorageTek's Expert Performance Reporter
HSC:	StorageTek's Host Software Component
MVC:	Multiple-Volume Cartridge (a physical tape)
RTD:	Real Tape Drive
TMM:	Tape Mount Management
VTCS:	Virtual Tape Control System (part of StorageTek's HSC)
VTD:	Virtual Tape Drive
VTS:	IBM's Virtual Tape Server
VTSS:	StorageTek's Virtual Tape Storage Subsystem
VTV:	Virtual Tape Volume

Data is directed to a VTD by either a JCL specification for the esoteric name (i.e., VTAPE) or optionally by DFSMS. StorageTek provides a third option, TAPEREQ statements, which are used by the HSC software. StorageTek also allows the use of HSC User Exits to direct data to virtual volumes. A VTD is selected by MVS (for StorageTek, HSC is also part of the VTD selection) and a virtual volume is mounted. Virtual tape drives may be shared by multiple MVS hosts using whatever software is used for sharing other drives (i.e., SYSPLEX Auto-tape sharing or CA-MIM). From the user's perspective there is nothing different in the JES log; he or she will see a tape drive being selected and a volume being mounted. The user may notice the speed in which the mount is satisfied. Both vendors can say that scratch mount requests receive sub-second response times. Mounts for specific volumes that reside on the disk buffer also will see sub-second response times.

MIGRATION, RECALL AND RECLAIM

For the IBM VTS, four minutes after the volume is closed it is copied to a physical tape. As with the virtual volumes, the physical tape is defined for IBM by a panel on the RISC processor. The physical volume is mounted on a real tape drive (RTD). The RTD is not defined to the host system for an

IBM VTS. IBM allows a minimum of three and a maximum of six RTDs per VTS controller. In a multiple VTS environment the RTDs are dedicated to specific VTS subsystems. The RTDs are always Magstar (3490) drives and are dedicated to the VTS Controller. IBM always leaves an output tape in at least one RTD to speed up the copy process.

For StorageTek the physical tape, called a Multi-Volume Cartridge (MVC), is defined via the same CONFIG utility used to define the virtual volumes. StorageTek allows the installation a choice of device types for the RTDs. These choices include Timberline drives (9490s), Redwood drives and 9840s, and may be mixed. StorageTek allows up to eight RTDs per controller (VTSSes) that can be shared between multiple VTSSes or between a VTSS and the MVS host(s). When there is a mixture of RTD device types the fastest devices are selected first for migration and the greater capacity devices are selected first in the reclaim process (these processes will be described shortly). StorageTek allows sharing RTDs between VSM and the MVS host(s); however, no mechanism is available to prevent an operator from varying a drive online to a host system while in use by VSM and vice-versa, which can cause problems with media integrity. My own opinion is don't share RTDs with MVS.

Initially, StorageTek did not copy the Virtual Tape Volume to an MVC immediately

after the volume was closed. The copy process, referred to as migration, was done automatically when the disk buffer utilization (DBU) reached a high migration threshold. VTCS then copied the least recently used volumes from the disk buffer to an MVC and deleted the disk version. Migration continued until the DBU reached a low migration threshold. The migration thresholds are set by the CONFIG utility and can be changed via an operator command. This migration process is still available; however, a recent enhancement to VTCS allows the installation to determine the migration process. Using the new Management Class feature of TAPEREQ statements the installation can cause the virtual volume to be migrated immediately after creation and optionally leave the VTV on the disk buffer. The ACSLIST parameter will allow the duplex copies to be created on MVCs in different ACSses. Operator commands can also initiate the migration of volumes, either by volser or to reach a specified threshold level. The automatic migration feature will delete the disk copy of the volume. The operator command allows you to choose whether to delete the disk copy or leave it on the buffer.

As mentioned earlier, StorageTek provides a third way of directing data to virtual tape. The TAPEREQ statements provide an option unique to StorageTek's VSM: duplexing. When an eligible VTV is migrated from the disk buffer to an MVC a second copy can be made on a different MVC. The first copy is considered the primary and will be used for recalls; however, if the MVC with the primary copy is unavailable or damaged, a recall action will automatically use the duplex copy. A duplex copy of the VTV is not a second volume serial number. The HSC control data set has an entry for the VTV and a pointer to the primary location of that volume and a pointer to the location of a secondary copy of the volume. It is not available to MVS or for offsite transport for disaster recovery purposes. It will only be used by VTCS if the primary copy of the VTV is unavailable. It is possible, however, to have duplexed VTVs written to MVCs that reside in different libraries (ACSses).

Regardless of the vendor, any attempt to read a virtual volume that still resides on the disk buffer will be satisfied by the disk buffer copy. If the volume needed is not resident on the disk buffer the physical tape (MVC) on which the virtual volume resides

will be mounted on an available RTD and the contents of the entire virtual volume will be re-copied to the disk buffer. Pointers to the MVC copy remain in place. Note that the user's job will not indicate that the tape mount has been satisfied until the entire virtual volume's contents are back on the disk buffer. StorageTek will leave the MVC mounted for 10 minutes (the time is set via the CONFIG utility's "RETAIN" parameter) in case other virtual volumes migrated at the same time to the same MVC are also needed for recall.

Retention periods for virtual volumes are determined the same way as traditional volumes (i.e., JCL specified expiration dates, SMS Management class, or via the CA-1 Retention Data Set). As a result of the migration process there may be hundreds of virtual volumes stacked on a physical tape. Both vendors have processes in place for recycling the physical tapes to remove gaps caused by expired virtual volumes. Storage-Tek provides, via the CONFIG utility, some control over that process. This process is called RECLAIM. StorageTek customers using the TAPEREQ Duplexing feature need not worry about the reclaim process. The primary and duplex copies of a virtual volume will never be on the same MVC.

REPORTING

IBM provides a set of SAS programs called VTS TOOLS that read SMF type 94 records cut hourly by the VTS sub-system. Among the reports generated are average and maximum mount times for virtual tapes (scratch and specific mounts are separated), average and maximum number of virtual and real tape drives allocated, and megabytes of data transferred for read and write actions.

StorageTek provides two reports, the MVCRPT and VTVRPT. The MVCRPT output lists the MVCs used, the media type, how full they are and their fragmentation level, and optionally lists the VTVs residing on the physical tape. The VTVRPT shows the status of virtual volumes, if they reside on an MVC (and if so, which MVC), if they reside on the disk buffer, and if there is a duplex copy and if so on which MVC.

StorageTek has bundled with VSM their ExPert Performance Reporter (ExPR). ExPR is also sold separately and is used to generate reports on tape drive utilization for manual and automated drives. Some of the

ExPR reports are comparable to the output of IBM's VTS Tools reports. ExPR has batch reporting as well as a real-time monitor using an SMF exit.

WORKLOADS AND DFHSM

In preparing their proposals for virtual tape solutions, each vendor will analyze your SMF data and recommend the data sets that should be directed to virtual tape. The appropriate workload will vary by client. In some cases they will include existing TMM candidates. Generally they will recommend data sets that are small with an average size of 200MB being the maximum data set size.

In my opinion you should ignore data set size and concentrate instead on a write/read ratio. Concentrate on data sets that are either unlikely to be read again or are likely to be read within a few hours of their creation (while the VTV is still disk buffer resident). I think this approach makes implementation quicker and makes it less difficult to maintain the selection criteria for DFSMS or HSC/TAPEREQ statements.

There are differing thoughts on directing DFHSM data to virtual. I have not seen any site direct DFHSM data to virtual tape. DFHSM has a limit of four million blocks of data on any tape volume and it is possible DFHSM, at sites defining the Magstar or 9840 drives as 3490s, may not fully exploit the capacity of these "fat tapes." Most data centers seldom use a DFHSM DUMP tape for recovery of a data set. I believe it makes sense to direct DFHSM DUMP tapes to virtual, especially in a StorageTek environment where the TAPEREQ MGMTCLAS will cause the volume to be immediately migrated to an MVC and deleted from the buffer. Similarly, directing DFHSM BACK-UP tapes to virtual with the same MGMT-CLAS makes sense if there are few times when a DFHSM RECOVER function is needed.

DFHSM Level-2 Migration volumes are another story. I would be reluctant to use virtual tape for ML-2 volumes. If I were directing level-2 tapes to virtual I'd want the virtual volumes to be migrated immediately but not deleted from the disk buffer until the disk buffer utilization required it. If the site uses the DFHSM TAPECOPY function I'd leave this management class as DUPLEX=NO; otherwise, I'd want the management class to be DUPLEX=YES. (I've been told that the most recent version of DFHSM allows for creating a duplex copy without using

TAPECOPY). I would also set the DFHSM SETSYS parameter of PARTIALTAPE-(MARKFULL) so that an ML-2 volume used for output tonight won't be recalled at the start of tomorrow night's DFHSM migration; however, this will lead to creating more volumes for DFHSM. I would want to be certain that either there are very few DFHSM RECALLS or that the majority of the recalls will take place in the anticipated disk buffer residency period. Remember that when a migrated volume is needed the MVS tape mount as seen by the user is not satisfied until the entire virtual volume has been recalled. An MVS user recalling a data set that DFHSM migrated to a level-2 volume that the virtual tape system has migrated to an MVC (and is no longer resident on the buffer) may see a longer wait time on their recall than if conventional volumes were used.

WRAP UP

Virtual tape systems are a viable solution to the problem with growth in customer tape libraries. Vendors have increased the capacity, reliability, and speeds of tape but allocations of many small datasets lead to a large number of poorly utilized tapes. Tape Mount Management and after-the-fact tape stacking utilities can help reduce the total number of tapes in a data center but involve additional CPU cycles and increased manpower requirements. Virtual tape solutions can improve tape utilization and offload the maintenance of the stacked tape volumes. ⓚ

NaSPA member Lloyd Christensen is a Storage Consultant with Storage Technology Corporation (STK). Lloyd can be reached via email at lloyd_christensen@storagetek.com or at lcwrite@aol.com.