Data Mobility Group

datamobilitygroup.com

Quantum

Compliments of Quantum

White Paper

May 2005

Dianne McAdam

Choosing the Right Disk-based Backup Solution

 76 NorthEastern Blvd. Suite 29A Nashua, NH 03062 Phone: 877.463.6456 Fax: 603.563.8282

Choosing the Right Disk-based Backup Solution

Shrinking backup windows and changing recovery time objectives have many customers investigating adding disk to traditional tape backups to accelerate the backup-and-restore process.

Vendors have responded to the demand for a speedier backup-and-restore process with a variety of disk-to-disk (D2D) backup-and-recovery options. Users can now choose to implement disk as a disk target (the disk system appears to the host as disk, not tape) or as virtual tape (the disk system emulates tape). Customers evaluating disk-based backups should understand the benefits and disadvantages of each approach in order to determine which best meets their needs.

The Background

One of the principal jobs of every IT department is to provide both short term and long term protection for their organizations' data. Years ago, it was common for end users to use one process—backup—as their only data protection procedure. The situation is much different today, especially for larger organizations. Short term data protection is being provided in many cases by some combination of data replication strategies, typically clones and snapshots, with backup providing longer term protection and defending data against site loss.

Clones and snapshots are data copying functions that can reside in the disk array, in an appliance within the SAN, or as software on the host. They are designed to store point-in-time images of volumes throughout the day. These temporary copies can be used to restore a volume rapidly if a problem is detected within hours, but they are not replacements for backup since they are not designed to retain permanent copies of volumes and can easily be overwritten. Customers implementing these types of copying technologies often use the copied volume, not the primary volume, as the source volume for the backup.

Traditional Backup Methodology

Backup itself has changed over the years. Operators used to back up data every night writing volumes to manual tape drives. Tape operators were kept busy running from one tape drive to another, mounting and dismounting tape media in order to keep the backup jobs running. If the operator could not find the right tape, then the job waited until the right tape was located.

The arrival of automated tape libraries replaced human tape operators with fast robotics. The robots could locate and mount tapes within the libraries in seconds, faster than any tape operator in running shoes. And tape libraries provide additional benefits. Capacity can be shared across numerous backup processes by combining tape cartridges into logical pools. Administrative processes, such as exporting tapes daily to a remote location, can be automated. And bandwidth between backup servers can be shared when the libraries are attached to a storage area network (SAN).

Tapes are not only used to store current backups, but are also used for disaster recovery and to store data with long-term retention policies.

Products that Add Disk to Backup

Tape has several important strengths as data protection technology—high capacity, low cost, portability, and good long term retention. But the tape-head interface and linear access characteristics also create latency. Disk, with its random access technology, has long been recognized as a good way to speed up parts of the backup and restore process, but it has been much too costly for mainstream use. New lower cost disk systems are making the technology practical for more widespread use, and vendors have responded with a variety of disk-to-disk backup-and-restore solutions. Users can now choose from several different options where disk is the target device. In some cases, the disk system appears to the operating system as disk; in others, the disk system emulates tape drives and libraries. Some of these solutions are software based, others are integrated appliances. Some of these solutions support physical tape devices and others do not. Each option has a different mix of benefits and disadvantages and meets the needs of a different group of customers.

These solutions can be very effective at speeding up the backup and recovery process. But disk solutions are not by themselves disaster recovery solutions and can be expensive for long-term retention. Tape continues to maintain its role as the most cost-effective medium for long-term retention and off-site storage for disaster recovery.

Disk as Disk

One segment of the D2D market—"disk-as-disk"—consists of using a disk array as the target of backup software instead of a tape drive or a tape library. This methodology is supported by all the major backup applications. Backed up data is written in disk, not tape, format. Disk-as-disk products are available in several varieties. In general, these products are not integrated with tape drives. Two approaches are common—using a conventional disk array as the target and using a special-purpose disk-backup appliance.

Conventional Disk Arrays

In this approach, secondary disk (usually higher-capacity SATA disk) serves as the target device for backups, and no specially built disk arrays need to be purchased. The backup application now directs the backup to disk rather than to tape.

Benefits:	Disadvantages:
If customers have extra disk capacity on hand, they do not have to purchase any additional hardware. This can be the lowest-cost solution. All of the major backup vendors support disk-based targets and publish how-to manuals that step administrators through the process of changing the backup targets from tape to disk.	In most cases, using native disk as a backup target means changing the backup procedure, including how data is moved off site for long term storage. Disk systems must be partitioned to support different operating systems and capacity cannot be shared across partitions. If the backup requirement for one operating system increases, then the disk system may need to be re-partitioned. Native disk arrays have basic interfaces that allow IT administrators to configure disks; however they cannot monitor the backup activity. Customers need to monitor these disks to determine if they are running out of capacity or if the disks have become fragmented.

Specialized Backup Disk.

Some vendors have developed specialized software that ships with disk arrays to help them serve as backup targets and provide other options, such as filtering (to eliminate redundant copies) or compression. Some of these solutions look like NAS filers to the operating system and support CIFS and NFS file systems; others look like Fibre Channel disks. These vendors have developed their "appliances" to address specific problems that have plagued IT administrators for years. For example, if the file system that is currently being backed up is corrupted, the backup will simply mirror the file system's corrupted state. One appliance runs consistency checks against the file systems to flush out integrity problems.

Benefits:	Disadvantages.
Appliances can help prevent backup problems by alerting IT administrators when disk capacity is reaching a critical threshold. Some appliances automatically invoke processes to move disk-resident backups to tape in order to free up disk space before an "out of space" condition can occur. Appliances have graphical interfaces that can monitor the status of the backups, which allows administrators to quickly determine the status of numerous backup tasks. Some backup appliances offer specialized functions, such as compression, that can save disk space and money. When additional capacity is required, many appliances allow disk arrays to be added into the existing storage pool.	These appliances have many of the same installation requirements as native disks; that is, the process changes, disks have to be configured and partitions or directories (in the case of NAS filers) must be created, and creating removable media is a separate and potentially more complex process. Special purpose appliances have many features not available with native disks, but usually come with a higher price tag.

In general, using conventional arrays and special purpose disk/software solutions as backup targets can reduce restore times. Special purpose appliances that offer specialized functions, such as compression, can save disk

space and money. You need to compare the cost and capacity of native uncompressed disks with the cost and capacity of an equivalent appliance.

One important note

Using either native disks or appliances requires changes to the backup application and process. Disks must be partitioned and the backup streams currently writing to physical tape must be changed to write to disk. Utility programs that eliminate disk fragmentation must be run regularly to reclaim unused disk space. These considerations make disk-as-disk target backup better suited for systems with smaller data volumes, slower growth, and fewer backup servers. And these solutions, by themselves, are not a disaster recovery solution. Customers must consider taking additional steps to ensure that copies of data are stored in a remote secure location.

Disk as Tape, or Virtual Tape (VT)

With virtual tape implementations, the disk array responds to software commands just as if it were a tape drive or a tape library, hence the name virtual tape. That gives virtual tape products an important benefit: they require few, if any, changes to the existing backup infrastructure. Usually, the disk array and software can be easily plugged into existing backup products. The normal process of directing the backup stream to tape is simply redirected to disk. Since the disk arrays emulate tape drives, not disk drives, many installation and maintenance tasks, such as setting up partitions and running disk utility programs to eliminate disk fragmentation, are not required.

Vendors have implemented virtual tape in different ways. Some only supply software while some ship disk arrays with software already installed. They also vary in how they support tape drives and libraries. Some do not provide direct support for tape drives at all, while others provide direct connections to tape drives for physical media creation

In general, these products fall into two categories: software only, and disk/software appliances. *Software-only* solutions are just that: software only. A vendor supplies the software to make the attached disk look like a tape drive or tape library to the operating system while the customer supplies the disk hardware. The software runs on a server which sits between the hosts and the disk system. Software-only products allow customers with excess storage capacity to put that disk storage to good use, but they require more active management and may increase the complexity of support.

Disk/software appliance implementations integrate virtual tape software and disk (usually a SATA array). Like software-only implementations, the virtual tape software that comes with these products responds to operating system commands as if the attached disk devices were tape drives, not SATA disk storage. But, unlike software-only products, disk/software solutions don't support heterogeneous disk. They only work with disk systems provided by the solution vendor. The software is developed to run on a specific storage platform, which allows vendors to optimize performance or implement additional features, such as hardware or software compression. The appliance approach usually

simplifies installation, management and support, but the acquisition costs are normally higher.

Whether they are deployed as software-only solutions or appliances, there are three general architectures available for virtual tape systems: tape drive emulation, disk-only library systems, and integrated disk-tape library approaches.

Tape Drive Emulation.

In these products, the disk system presents itself as one or more standalone tape drives, and the backup software writes to them in tape format.

Benefits:	Disadvantages:
This approach provides the simplest interface and is likely to have comparatively lowest cost than other virtual tape systems.	Tape drive emulation systems do not provide automated processes and do not allow the disk within the virtual tape system to share its capacity or throughput between different hosts. Users who have previously automated functions within libraries will lose those capabilities, and scalability will be limited. With tape drive emulation systems, creation of physical tapes is generally a separate process and may not be automated.

Virtual Library—Disk Only.

In these systems, the disk capacity is presented to the backup software as a tape library, including drives, cartridges, and robotics. In disk-only style libraries, the appliance itself consists of controllers, software and disk storage, but no tape components are integrated.

Benefits:	Disadvantages:
Because the disk is presented as a tape library, users can operate the disk using library processes. That means they can share the system's capacity and throughput between different hosts, can take advantage of backup automation, and can scale to higher capacities.	Since the integration is limited to disk storage, users who want to create physical tapes for export with disk-only virtual libraries are required to set up a separate process. The backup application that initially creates the disk backup image is used to copy, or clone, the image to physical tape for longer-term storage or off-site disaster recovery. This extra step consumes additional I/O bandwidth and backup server cycles.

Virtual Library with Direct Tape Support.

The third virtual tape architecture includes an integrated connection between the disk-based virtual library and a physical tape library, using processors in the appliance to move data between disk and tape. Some of these systems use the tape as an export device—a way of getting virtual volumes onto removable media. Others merge the capacities of

disk and tape so that the backup software sees them as a combined pool, and manage the movement of data between disk and tape tiers.

Benefits:	Disadvantages:
Virtual libraries with direct tape support require the least change for end users accustomed to tape library operation. Since disk and tape capacities are managed as one large pool, capacity can easily be increased by adding more tape cartridges to the library. Backups can be migrated from disk to tape without requiring any resources of the backup server. The systems that merge disk and tape capacity also provide very high scalability by letting the backup application write to the total capacity of the disk and tape storage without actually having to know where the backup copy resides. The virtual tape software keeps track of the physical location of the backed up data and can migrate backup copies between disk and tape over dedicated paths as business needs dictate. Some systems offer additional data protection levels, such as management of duplicate copies of virtual volumes in different locations or on different media types.	Systems that combine disk and tape support are likely to be more costly than systems that don't build that capability in. And because they offer more options, they may be more complex to configure. In addition, users need to be certain that the features they want are supported by the products they select. If users expect to create multiple tapes regularly, they need to be sure that their system has high enough bandwidth for the job. Likewise, if users want to track offsite media through the backup software and restore off-site tapes in a conventional library, they have to be sure that the product they select offers those features. In some products, the physical cartridges remain under the control of the virtual tape application, not the backup application. In others, exported tapes are written in a proprietary format and can only be restored in another virtual library. Customers who plan to remove tapes from the library for disaster-recovery purposes must determine if the tape can be used outside the control of the virtual tape software or must be returned to the library before it can be read.

Questions to Ask Vendors–

These questions will help you evaluate the numerous disk-based backup solutions on the market today:

Performance and throughput.

- What is the maximum throughput of the disk-based backup system?
- Is the backup environment architected to deliver this throughput to the virtual tape system?
- Are there existing bottlenecks—for example, in the network?
- Does the vendor have benchmarks that describe the performance of the system in different environments?
- Does the vendor have tools that can predict the performance of the system in a specific environment?

Compression.

- Does the system support hardware or software compression? Compression saves space but can affect performance, depending on how it is implemented. For example, hardware compression is more efficient than implementing compression via software.
- How does compression impact the performance of the system?
- Can compression be turned off if performance is degraded?

Installation.

- How long does it take to install?
- Is it customer-installable?

Training.

- How much training is required?
- Is training provided by the vendor?
- What is the cost?

Support.

- What levels of support will be available from the vendor?

Market acceptance.

- How many systems have been installed?
- How many systems are in production?

Product maturity.

- How mature is the product?
- Has it been available for years or is it a recent addition to the market?

Integration with existing environment.

- Does the disk-based solution integrate with existing tape drives/libraries or is it a standalone product?
- Does it integrate with currently installed backup software applications or must a new backup application be installed?

Capacity and scalability.

- What is the capacity of the disk system?
- If additional disk storage is required later, is an upgrade path available?

Number of backups.

- How many different backups can be stored?
- How many different versions of each backup?
- Are there vendor tools to help estimate the number of backups that can be stored?

Disk management.

- How will disk space be managed?
- Will disk copies be automatically migrated to tape as they age or will they be migrated manually or through scripts?
- Can selected backups be written immediately to tape if required?
- Can thresholds be set to send alerts when disk storage is reaching capacity?
- What happens if the disk space is full and cannot store any more backups?

Media management.

- How are physical tape cartridges managed? In most disk-only solutions, the backup application copies the disk backup to a physical tape cartridge and the cartridges remain under the control of the backup application. Some integrated disk/tape solutions automatically migrate disk copies to tape and the cartridges remain under the control of the virtual tape software.
- When the tape cartridge is ejected from the library and moved off-site for disaster recovery purposes, can it be read in a standalone drive or must it be re-entered into the library?
- Is the tape written in a proprietary format that can only be read through the virtual tape software? If so, are utility programs available to convert the tape to standard format to recover data in a disaster recovery site?

New backup architecture.

- Should the solution be used for all backups or only selected backups?
- Will some backups be directed to disk and the rest written directly to tape?
- How will the disk-based backup solution fit into the existing backup infrastructure?

Number of server-to-disk paths.

- How many paths are supported between the host server(s) and disk storage? The more paths available, the more concurrent backup streams can be supported.
- Number of disk-to-tape paths.
- How many paths are supported between the disk storage and the real tape drives? The more paths available, the less time it takes to migrate the backup copies to tape when the disk system reaches capacity. Tightly integrated disk/library systems migrate disk backup copies to tape using dedicated paths between the disk systems and tape drives. Disk-only systems use the backup server to move the copies to tape, increasing utilization of existing channels and networks.

Network connections.

- What types of paths or network connections are supported (SCSI, Fibre Channel, ESCON, FICON, IP, iSCSI)?

Management tools.

- What management tools are available?
- What policies are pre-set and what policies can be changed?
- How granular are the policies?
- What types of reporting tools are available?
- What features are available today and what will be available in the future?

Remote management.

- Can the system be managed remotely?
- What levels of security are provided to prevent unauthorized access?

Virtual tape solutions.

- Which tape libraries does the virtual tape technology emulate?
- Which tape drives? Some virtual tape solutions present themselves as tape drives only, while others present themselves as multiple libraries with multiple drives. While both emulations work, library solutions are a better choice to support three or more backup servers.
- Which types of cartridges are emulated?
- Does the product support only one type of library, drive, and cartridge or multiple types of devices and media?
- If the emulated devices are different from the currently installed devices, are any changes required to the backup application?

Recommendations

Which solution is the right one? The answer depends on your environment and your budget.

Small customers with one or two backup servers and very limited data-growth have a multitude of choices since any of these solutions will work for them.

Customers with limited or overextended IT staff should consider virtual tape solutions, since they require few, if any, changes to the current backup processes. Also, if the existing backup application does not support writing disk-based backups effectively, then virtual tape is the best solution.

If the IT organization has been plagued with file system integrity problems, then specially built appliances that provide file system integrity checks can go a long way to resolve that difficulty.

Organizations that have very little update activity will find that appliances that provide compression and factoring (which eliminates redundant copies) can save a large amount of storage.

Current and planned data retention policies also factor into the decision process. If the company has policies that dictate that data is kept for months or years, then the best solution is one that can automatically migrate backups to physical tape based on pre-defined policies.

Virtual tape solutions with more direct connections to tape drives are best suited for complex environments with high data-growth. Large virtual tape appliances which emulate multiple tape drives (and libraries) can assign, for example, a virtual library and several virtual drives, or several tape drives, to each backup environment, providing consolidation without adding additional complexity. Customers with high rates of data growth can easily and inexpensively scale by adding more cartridges to the library without having to add more disk arrays. Customers with three or more backup servers will find that virtual tape solutions allow resources, such as tape drives, to be easily shared.

Disk-based backup solutions are effective for reducing the backup window and speeding up the restore time. But disk-based solutions are not disaster recovery solutions. Many D2D backup products will reside in the same SAN as primary disk, rendering them ineffective for disaster recovery. A data center outage, such as a power failure, would make both the primary disk and the D2D backup inaccessible. Disaster recovery best practices dictate that a copy of data be stored at an off-site location. Each solution must be carefully evaluated to determine which processes, if any, need to be changed to satisfy the requirement of off-site storage.

Conclusions

There are numerous disk-based backup solutions on the market today. These products can enhance the backup-and-recovery process, provided the right solution is chosen for the environment. Disk-based backup solutions that use conventional disk arrays as targets are well suited to smaller, less complex, budget-constrained environments; disk-as-disk appliances provide additional features over native disk; virtual tape solutions require the fewest changes to overall data management and backup infrastructures.■