Exabyte Corporation
1685 38th Street
Boulder, CO 80301

www.exabyte.com

An Exabyte White Paper

The Role of Disk in Effective Data Protection Strategies

Introduction

The plunging cost of disk storage during the last decade is motivating many IT organizations to consider the contribution potential of disk storage to their data protection strategies. With the advantage of very fast backup and recovery times, employing disk storage as a backup medium appears lucrative. However, the disadvantages of appointing disk storage for long-term archival limits its role in a data protection strategy.

This paper presents practical guidance on the role of disk storage within data protection, including total cost of ownership comparisons between disk and tape media, and suggestions for the implementation of disk storage to address specific backup and restore challenges.

Is Disk Cheaper than Tape?

Disk storage is assuming a greater role in data protection schemes due to its continuously decreasing cost. At a glance, it seems that disk storage is destined to reach parity with tape media. Historically, disk storage costs have dropped 45-50% per year. Low-end desktop disk drives fell below a dollar per gigabyte in 2003[1], and are forecasted to reach 25 cents per gigabyte in 2006-2007[2].

Lagging the pace of falling low-end disk storage cost, tape media costs are decreasing 25-30% per year. This is the basis for the forecast that by 2008-2010, disk storage will cost less than tape on a media cost per gigabyte basis. Nevertheless, when considering the prevailing pattern of more dramatic cost reduction between generations of tape technology, expect tape media cost per gigabyte to maintain its status as the lower cost solution. For example, today Exabyte's VXA-2 Packet Tape format media costs 40 to 50 cents per gigabyte of compressed storage, roughly half the current cost of low-end disk. Exabyte's VXA-3 Packet Tape format, expected in 2005, will double the capacity of existing VXA-2 tape cartridges, thus reducing the cost per gigabyte by half to 20 to 25 cents.

The cost of the media is only one small component of an overall data protection strategy. Consideration must be given to other aspects, including media shelf life, archival strategies, geographic isolation, and overall ownership costs.

[1] February 16, 2003, Page 1F, San Jose Mercury News (CA).
[2] August, 2002, Gartner Dataquest

Consider the Total Cost of Ownership

A complete data protection mechanism includes all equipment and media sufficient to address backup and archival for a specific period. A basic tape backup strategy TCO includes both the drive and tapes. For a disk backup strategy, a basic TCO includes the drives and a cabinet with a RAID controller.

A common backup strategy of two, rotating sets of weekly full and incremental backups for a five day work week, twelve monthly archives, and one annual archive requires 23 tapes, provided that the full backups fit on a single tape. To store equivalent copies of primary storage on disk requires 15.8 times the primary storage size.

100GB Data Protection Schemes

2 Weekly Full Backups	2 VXA-X23 tapes	200GB
8 10GB Incremental Backups	8 VXA-X6 tapes	80GB
12 Monthly Full Backups	12 VXA-23 tapes	1,200GB
1 Annual Full Archive Copy	1 VXA-X23 tape	100GB
Total Requirements	23 tapes	1,580GB

Figure 1 – Tape and Disk Space Requirements to Backup 100GB

In terms of cost, consider the example of a company with 100GB of data requiring protection. The figures utilized in this example are based on retail pricing available in January of 2005. Although both disk and tape costs will continue to decline, the cost relationship between deploying tape versus disk storage will remain fairly constant.

Deploying a VXA-2 drive, with 15 VXA X23 tapes for full backups and monthly archives, and 8 VXA X6 tapes for incremental copies will cost just under $2,500. This assumes that the drive will be installed in an existing server's cabinetry. An externally mounted VXA-2 drive adds approximately $200.

Deploying disk storage for backup is more complex. The drives must be installed in a cabinet with power and other logic. To ensure that data is not lost if a drive crashes, a distinct possibility with multiple low-end disk drives, a RAID controller is commonly installed to provide fault-tolerance. Further, the most popular form of RAID requires an additional drive to hold the fault-tolerant "parity" information.

Deploying 1,580 GB of disk storage within a low-cost RAID controller will cost approximately $1,800 for the primary drives, $250 for the RAID parity drive, and $2,000 for a low-end RAID enclosure, totaling $4,050.

Thus, for the first year's data protection, the basic tape TCO of $2,500 is 38% less than the basic disk TCO of $4,050.

As the data volume requiring protection grows, the tape cost advantage is more significant. The Tape Technology Council, a non-profit trade association comprised of leading data-storage technology manufacturers, conducted a study[3] in 2004 to identify and compare the TCO of various data protection strategies, including hardware and media costs, installation, controllers, application software, power and cooling, floor space, security, scalability, reliability, and personnel training. The study revealed that for a company requiring 10,000 GB of archived data over three years, the cost of tape backup is less than 25% that of disk.

Three Year Total Cost of Ownership to Protect 10,000 GB

- Tape Library $102,394
- Optical Disc Jukebox $150,688
- Mirrored RAID $416,085

Includes hardware, software/controllers, media, hardware/software maintenance, floor space, and power consumption.

Figure 2 – Tape Technology Council Study Results

Consider the Level of Protection

The most significant obstacle to disk based backup strategies is the lack of portability. Because disk drives are non-removable, the backup copy remains within close proximity of the primary storage, limiting the level of protection from fires, floods, hurricanes, lightening strikes, or vandalism. Tape backup facilitates convenient and regular off-site transportation of archive copies, providing a means for recovery in case of loss of the primary hardware or a disaster preventing access to the primary location. Although methods exist for backing up to disk over long distances utilizing Internet or private networks, these techniques significantly increase the cost and complexity of the backup solution.

Unlike on-line disk backups, data archived on tape is isolated from viruses, operating system crashes, or other hardware or software failures that might corrupt connected disk drives. On-line disk backups are continuously exposed to the same hardware and software risks threatening the primary storage.

Unless a significant amount of backup disk storage is deployed, the backup array will only accommodate a few backup copies. This limits the period from which data may be restored. Consequently, if a corrupted file is not discovered for several weeks or months, the last backup containing a good copy of the file will be overwritten, and the data permanently lost.

The lower cost and portability of tape media allows companies to keep any number of past backups available to meet their recovery requirements. Monthly archives are commonly maintained for one or two years, providing an important depth of protection against undiscovered corruption.

[3] Tape Technology Council, November 1, 2004, Presented at Storage Networking World

Think Long Term

Archived data must be maintained for a minimum of several years to meet tax, governmental, and industry regulatory requirements. Tape is intended for long-term storage, with an expected shelf life of 10-30 years. Tape cartridges are designed to tolerate weekly transportation, and include cases for protection against environmental contamination. A tape created today will be readable by any drive that supports its format, including future generations of upward-compatible tape drives. Thus, accessing archived data is not dependent on the operational status of a single hardware device.

Disk drives are designed for mounting within a stationary cabinet, without relocation, with continuous power. There is no tested lifetime for disk storage when stored without power for long periods. Furthermore, the entire disk cabinetry and controller must be maintained as a single unit for the life of the archive, or the data will be inaccessible.

So Where does Disk Belong in Backup?

Disk storage offers a high data transfer speed, which can be applied to resolve specific backup challenges. In every case, disk storage is used within an overall data protection strategy that includes tape.

The question of when to apply disk-to-disk backup strategies depends on cost justifiable requirements of the organization's data protection scheme. In all cases, the addition of disk storage to the backup mechanism adds costs that must be justified by the value of resolving specific operational challenges.

Accelerated Backups

The simple speed of tape based backup or restore is insufficient for some operational requirements. Generally, backups occur while users are offline, to prevent open file conflicts. The period when backups are possible is called the backup window. As storage volume grows, the backup window may shrink to the point that a backup will not complete within the allotted period. Or, more frequently, the backup window is significantly reduced as the need for access to live data increases due to longer workdays, web-based commerce, or expansion across multiple time zones. Thereby, interim disk-to-disk backup becomes a cost-justifiable requirement to reduce the time required to produce a backup.

The addition of a disk array large enough to hold a single full backup significantly accelerates the backup process. First, each day's backup is copied directly to the disk array in a matter of minutes. When the backup is complete, at the convenience of the computer operator, the array's contents are copied to tape. By creating the tape copies off-line, the duration of the "live" backup is significantly reduced, resolving the problem of a shrinking backup window. The daily tape backups may be rotated off-site, providing disaster protection. This configuration is commonly called "disk-to-disk-to-tape."

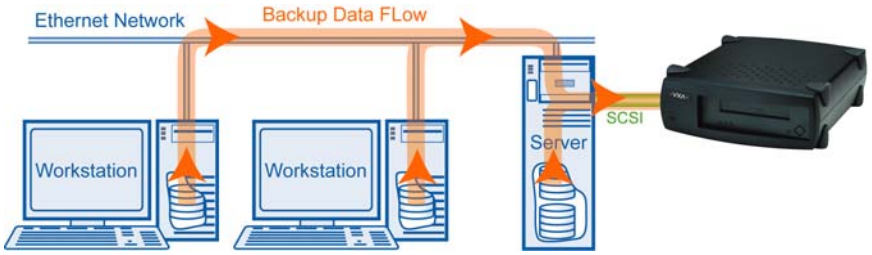

Figure 3 – Traditional Tape Backup Configuration

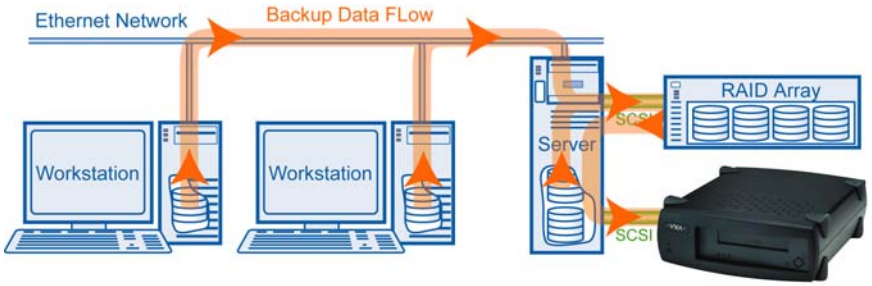

Figure 4 – Disk-to-Disk-to-Tape Backup Configuration

Restore Time Objectives

Continuous access to data is critical for on-line, transaction-oriented businesses such as stock brokerages, eCommerce Web sites, and health care facilities. Their cost of downtime is measured in thousands of dollars per hour. Expediting data loss recovery justifies the purchase of additional equipment by reducing downtime costs. These organizations have specific Restore Time Objectives (RTO), requiring that lost data is recoverable within an established timeframe, generally measured in hours or minutes. In the case of significant data loss, a tape drive may not meet the RTO, taking too long to re-read the entire tape(s).

A disk-to-disk-to-tape backup mechanism maintains the most recent backup copies on the backup disk array, prepared for rapid restoration at disk-to-disk copy speeds.

Restore Point Objectives

Some organizations also have a specific Restore Point Objective (RPO) that defines the maximum duration between backups to limit the amount data loss in the event of a hardware failure. A RPO of one hour requires that backups be performed every hour during periods of commerce. Tape backup is generally too slow to complete within the RPO period.

A disk-to-disk-to-tape mechanism with a disk array capable of holding several backups allows for an aggressive RPO. Backups are conducted frequently during the day directly to the disk array. The last backup is then copied to tape off-line from the primary systems, facilitating off-site tape rotations.

The Common Factors of Disk-to-Disk Backup

Each of the above scenarios exhibits two common factors. The requirement for implementing a disk-to-disk backup solution is cost-justified by removing a limitation on business operations. All disk-to-disk backup schemes include a tape device to archive data for off-site rotation and long-term storage.

Tape provides the only cost effective means for transporting data away from the primary site, and the only reliable and tested means for long-term archival. Although disk media pricing continues to decrease, its fragile moving components and requirement for supporting hardware limit its usability as a portable vault for data assets. Tape provides the most cost efficient and reliable means of data protection, offering off-site portability, a long-term shelf life, and scalability to meet future needs.

COPYRIGHT

TRADEMARK NOTICES

REVISION HISTORY

Revision	Date	Description
001	February 2005	

CONTACTING EXABYTE

Exabyte Corporation
2108 55th Street
Boulder, Colorado 80301

(303) 442-4333

Technical Support: (760) 305-5517

www.exabyte.com