A WestWorld Productions, Inc. Publication
NOVEMBER 2003 VOLUME XXIII Number 11
www.wwpi.com

Computer Technology Review®

THE STANDARD OF EXCELLENCE FOR STORAGE TECHNOLOGY INFORMATION

Tape or Disk: Why Not Both?

By Steve Bagby

The topic of tape vs. disk storage has been hotly debated among technology experts, with some camps asserting that tape will become obsolete and disk will become the predominant backup medium. While there are many arguments for taking tape out of tape backup and shifting to a disk-based backup paradigm, tape will remain an integral part of backup and disaster recovery strategies. Moreover, the combination of tape and disk may be the optimal data protection strategy, especially when virtual tape technology is implemented.

Tape-Based Backup

The end of tape is not near. IDC sees the tape drive market increasing through 2006. Unit shipments of drives with 100-gigabyte or more capacity is expected to increase, sustained by growth in SDLT, LTO and SAIT drive shipments. But IDC also acknowledges that the growth of tape drives has slowed as trends in storage consolidation and storage area networks support the use of fewer, higher capacity tape, tape drives and tape libraries.

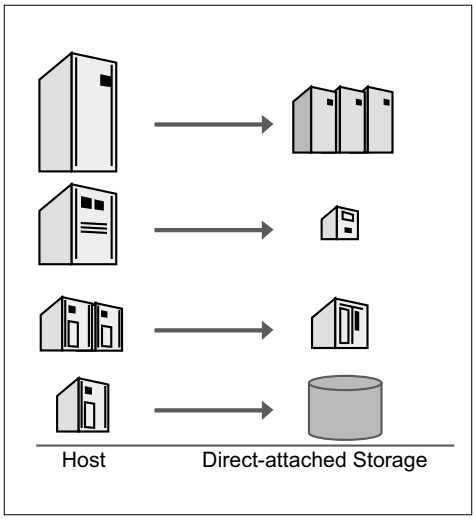

Figure 1

When comparing cost-per-gigabyte, tape is still one of the most cost-effective storage technologies. IDC states that "although pricing varies widely, disk generally costs $0.15 to $0.20 per megabyte, versus $0.001 per megabyte for tape." Even though disk prices have fallen, "ATA drives expected to drop to $0.003 per megabyte," the gap between tape and disk will probably remain as tape manufacturers continue to drive down the cost of tape while improving performance.

Among other things, tape has capacity, cost, shelf life and portability benefits over disk. These benefits, in light of regulatory requirements for data retention and the fact that most end users are unwilling to eradicate tape from their storage environments anytime in the near future, confirm the fact that tape is not yet dead. In fact, a June 2003 *Byte and Switch* poll stated that "as for the longevity of tape, most respondents believe it will remain a part of the overall data protection universe, with 40% agreeing that there will always be a need for removable media and 36% saying it will persist for long-term archiving."

The negative aspects of tape-only-based backups focus on performance and the overhead of tape handling, as well as the risk of failure introduced through human error. Tape's relatively lower seek times and head to media transfer rates makes tape comparably slower versus disk for both backup and restore. Unlike disk, tape media has to be manipulated. Operator intervention is both expensive and introduces opportunities for errors to be introduced to the backup process.

Disk-Based Backup

One of the most compelling reasons to adopt a disk-only backup strategy is the gain in backup and restore performance. Access to data on disk is almost immediate since there is no tape to mount or to rewind, and no searching through tape volumes for data files. For normal operations, data is stored on disk; so isn't the same reliable, fault tolerant media good enough for backup sets? In fact, some might argue that the risk of damaged tape media is removed when disk is leveraged as the only backup medium. Since average mean time between failures for disk drives are at least twice that of tape drives, and the difference is even greater when RAID technologies are used, the probability that backup windows will be respected and backups actually completed is much higher with disk than with a tape-only strategy.

Physical media is also eliminated. As are the costs associated with tape media, including media purchases, on- and off-site media storage, and tape handling. And let's not forget about the reduction in tape devices or the maintenance fees paid every year on tape hardware.

While there are several benefits to disk-based backup, there are also some drawbacks to a disk-only strategy. Several backup applications have adapted the structure of their backup data to disk, but still haven't perfected the selection and movement of specific backup sets to tape. Even when a low-cost disk array is deployed for backup, it will only be able to hold a few weeks of backup data. It cannot hold the vast amount of data tagged for long-term retention. Practicality will force organizations to eventually offload backup sets to removable media.

Security threats from viruses, theft, and natural disasters also need to be considered. The threat of totally losing a site, power or communications infrastructure might require physical copies of data stored away from the primary site for disaster recovery purposes, which will drive up costs of disk-based backup. For archival, the shelf-life of disk is significantly less than the 30 years expected from tape media. Idle disks may start to lose data and start to demagnetize after a year or two, unless they are spun-up every 30 days.

Deploying Both Tape and Disk for Data Protection

The role of tape in a data center's data protection strategy is changing. Tape will remain a keystone to organizations' backup strategies, especially for permanent retention of data. Disk will act as a complemen-

tary adjunct to tape, delivering the backup and restore performance needed for 24x7xforever operations. By combining both technologies and deploying disk as a staging area for tape backup, both short-term, high-performance backup and restore, and long-term archival of data are accomplished. Information Lifecycle Management is also supported by retaining more important data or data accessed within 24-48 hours on more expensive cache disk resources and retaining less important data on tape.

Actually, this two-tier strategy has been in use for years in the mainframe environment, with the deployment of virtual tape subsystems in the mid-1990s. Tape virtualization combines hardware and software to create new efficiencies in tape backup. Cache disk acts as a staging area for the tape drives and provides high-performance backup and restore for the host system. The software that drives the virtual tape engine emulates tape devices on disk and manages the movement of data from cache to tape and back. By removing the direct-attached tape drive dependency for the host, performance is radically improved because tape drive latency is eliminated and the software can present multiple logical or virtual tape devices to the host's backup application. The cache disk accepts data at any speed up to its maximum throughput.

Undeniably, virtual tape technology as an enabler for storage consolidation and resource sharing is a hot topic, especially in an economic climate that dictates productivity improvements and budgetary restraint. For data centers characterized by diverse storage infrastructure, including diverse storage topologies, multiple storage platforms, multiple operating systems, and multiple storage management solutions, the complexity is so overwhelming that storage consolidation seems an impossible task. Many organizations invest in dedicated tape hardware per platform and run stovepipe backup operations. This scenario creates distributed decision-making and operations, leading to major inefficiencies and excessive expenditures.

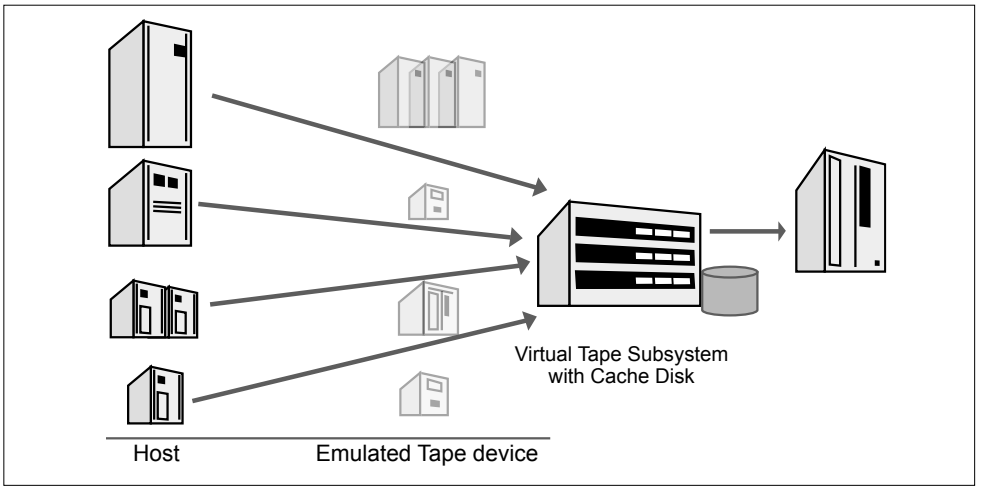

Figure 2 - *Virtual tape configuration*

Tape virtualization can reduce the complexity and expenses associated with this distributed model through storage consolidation. Organizations may elect to centralize tape operations to an enterprise-class tape library and introduce automation, or allow device sharing, thereby reducing tape hardware. By emulating tape drives on cache disk to any number and type of host systems, a centralized management capability and tape device sharing are enabled. In this scenario, the investments in disk, tape automation and virtual tape technology deliver tremendous return on investment.

By deploying a virtual tape solution, IT organizations can expect to see an impact on the total cost of ownership for tape operations. Immediate payback is seen simply in the reduction of tape hardware and its associated maintenance, as well as better tape hardware utilization. The reduction in hardware may also impact physical footprint, heating/ cooling, and power consumption. Performance improvements realized by hosts writing to any number of virtual tape devices in parallel, eliminate online application downtime for backup cycles. By streamlining operations, there may be a reduction or redeployment of operational personnel, especially if tape automation is newly introduced, since the physical handling of media is manpower-intensive. Lastly, many tape virtualization solutions offer better media utilization leading to a reduction in media and its related on-site and off-site storage costs. If the tape virtualization solution allows for the introduction of newer, high-capacity tape technologies or library automation where it wasn't available before, the impact can be significant for tape operations.

In the tape vs. disk debate, there are pros and cons to both tape and disk as independent strategies for data protection. More likely, the conclusion most will reach is that deploying both technologies will leverage the merits of each and provide a safety net for the weaknesses each impose. Furthermore, virtual tape technology is the key enabler to optimally leveraging this two-tier strategy and realizing significant business value. CTR

Steve Bagby *is senior vice president of Development & Operations for Neartek, Inc. (Westborough, MA)*

Evaluating Virtual Tape Solutions

When contemplating a virtual tape solution, consider the following:

- Will the virtual tape solution support multiple operating platforms including proprietary systems?
- Is the solution independent of platform hardware and/or tape device hardware?
- Is it compatible with current backup software applications in use?
- Does the solution have minimal impact on current backup processes?
- Is the data transfer to tape an integral part of the solution or an afterthought of disk-based backup?
- Will the virtual tape solution allow for tape resource sharing or consolidation?
- Can the solution resolve interoperability issues between host systems and tape technologies?
- Is the solution flexible enough to allow choices for disk-only, tape-only, or disk-and-tape backup/archival strategies?
- How will virtual tape impact disaster recovery plans?
- Will ROI be near-term?

www.spectralogic.com

SPECTRA
SPECTRALOGIC.COM

Spectra Logic Corporation
1700 North 55th Street
Boulder, CO 80301 U.S.A.
1-800-833-1132

Reprinted by permission from the publisher of the November 2003 edition of ***Computer Technology Review®****.* © 2003 West World Productions, Inc.