Tape's evolving data storage role

Balancing Performance, Availability, Capacity, Energy for long-term data protection and retention

By Greg Schulz

Founder and Senior Advisory Consultant, the Server and StorageIO Group
Author "The Green and Virtual Data Center" (CRC) and
"Resilient Storage Networks: Designing Flexible Scalable Data Infrastructures" (Elsevier)

StorageIO

February 17, 2011

This Industry Trends and Perspectives Paper is Compliments of:

Quantum.

www.quantum.com

Introduction

There is no such thing as a data recession and no end in sight for the amount of data that will need to be stored and protected for longer periods of time. Organizations of all sizes are faced with a growing data footprint that results in more data to manage, protect and preserve for longer periods of time. Magnetic tape media (tape) is a data storage technology that has been routinely declared dead over the past couple of decades. With recently released tape roadmaps along with product enhancements, it is safe to say there is plenty of life left in this important data storage medium that enables more inactive data to be stored in a cost effective manner.

Background and Issues

Traditionally, backups have been done from disk to tape, either a standalone tape drive or a drive in a tape library system. Recent technology evolution is seeing an increase in disk-to-disk (D2D) backups followed by a copy of data also being made to tape. With D2D backups serving as a staging or buffer area to hold current data copies, tape's role is changing to holding larger copies of weekly or full backups, retaining gold or master backup copies and supporting data archiving and preservation.

Online primary storage has a focus on fast low-latency, or throughput reliable access to data while near-line secondary storage has a focus on low cost and high capacity. Long-term data retention requires a combination of ultra-low cost, good performance during storage and retrieval, and reduced footprint in terms of power, cooling, floor-space and economics (PCFE) - also known as a small green efficient and effective footprint - for inactive data.

The emphasis of short-term data retention is on data protection focused D2D snapshots, data replication locally or to a remote facility, and backup to disk or tape for restoration to support business continuance (BC) and disaster recovery (DR). The "short-term" of short-term data protection is measured in days, weeks or, perhaps, months. Emphasis for short-term data retention is on recovery time objectives (RTO) and recovery point objectives (RPO) for business

> ### Industry Trends and Perspectives
>
> StorageIO experience along with other industry research[1] confirms that many mid to large size organizations continue to use a combination of disk and tape. In discussions with IT professionals and VARs, StorageIO routinely hears and sees the evolving use of tape for long term data retention combined with disk to disk (D2D) and other data footprint reduction (DFR[2]) technique.
>
> A common theme found with IT managers is the alignment of the right technology for the task at hand making HDD and tape an effective storage combination providing a balance of performance, availability, capacity, economics and energy efficiency for backup/restore along with long term data retention.

survivability. Long-term data retention consists of monthly or other long-term backups, also known as a master or "gold" copy of data, along with off-line copies of static, fixed content reference and other data as well as archiving of both compliance and non-compliance strategic information . The emphasis of long-term data retention is safe, secure and ultra-low cost for large amounts of data stored for long periods of time ranging from months to years to decades.

---

[1] Refer to Quantum (www.quantum.com) and LTO (www.ultrium.com) sites to view examples.

[2] DFR = Data Footprint Reduction includes archive, backup/restore modernization, compression, dedupe along with storage tiering among other technology and techniques. Learn more at http://storageioblog.com/?p=1532

www.storageio.com   P.O. Box 2026   Stillwater, MN 55082   651-275-1563   info@storageio.com

Tapes evolving data protection and retention role

Contrary to some beliefs, tape is far from being a dead technology, particularly for long-term data retention and bulk data storage protection. While some may claim disk drives have "killed" tape, the reality is that disk drives are actually helping to keep tape relevant by off-loading random type access and serving as a buffer or cache for tape. The net effect is that by using a combination of disk and tape (Table 1), backup data can be staged to disk-based technologies and more effectively streamed to tape to boost tape drive and media utilization leveraging compression and encryption built-in capabilities.

	Short-term data protection	Long-term data retention
Retention polices	Hours, Days, Weeks	Weeks, Months, Years,
Storage mediums	D2D, Tape, local and remote	Local and offsite tape, cloud including tape services
Data Footprint Reduction (DFR), Green IT optimization	Archive, backup modernize, compression, dedupe, space saving snapshots, tiered storage including disk and tape	D2D to tape (D2D2T), copies of tape sent off-site, compress, verify and encrypt tape media

Table 1 - Technology Co-existence: Determining storage techniques and technology to use

Technology improvements in D2D backup have had a positive impact on the ability to address short-term data protection, BC and DR requirements. In addition to protecting data for BC and DR, data also needs to be preserved or retained (also known as archiving) for varying lengths of time to meet different near- and long-term business and service requirements. In addition to the increasing amounts of data, business requirements and compliance regulations have resulted in even larger amounts of data to be retained in more locations for longer periods of time.

Considerations and drivers for long-term data retention include:

- Data finding value as information over longer periods of time
- Business and regulatory requirements – regulatory compliance and data preservation
- Economic and budgetary concerns – doing more with what you have or less
- Data loss prevention and information protection – protect, preserve and serve
- Environmental and business sustainment – green and economically efficient
- Maximize IT resource effectiveness and return on investment (ROI)
- Reduce total cost ownership (TCO) of IT resources and service delivery

There are many common myths about tape, data protection and archiving, one, for example, being that archiving and long-term data retention are only for regulatory compliance purposes. The reality is that while regulatory compliance data, including Sarbox and CFR financial or HIPAA medical, require long-term retention, many other common application data for almost every business, including those that do not fall under regulatory requirements, can benefit from - if not require - long–term data retention. Think beyond regulatory compliance instead looking at the wide opportunities and DFR benefits of leveraging archive as a storage optimization and Green IT enabling technology across all types of data or applications.

Tape past, present and future

With the continued enhancements in native, compressed storage capacities on tape media and improved resiliency for long term data survivability or shelf life, tape not only remains relevant, its role is changing to that of being used for large scale bulk data protection as well as long term data preservation. A common misperception is that the hard disk drive (HDD) combined with data deduplication and other forms of data footprint reduction (DFR) along with different interfaces including NAS and VTL (Virtual Tape Library) has killed off tape as a storage medium.

While D2D based backups continue to gain popularly, tape usage, particularly in mid to larger sized environments, continues to be relied upon expanding roles. For example, to boost utilization of tape drives while reducing the number of devices required to accomplish backup, archive or other data protection functions in a given amount of time, D2D solutions are functioning as a buffer or cache for tape. By performing a D2D backup, perhaps from a snapshot or replication copy, data is buffered such that when it is ready to be written to tape, a drive with adequate configuration can actually stream the data in a more effective manner. The net result is that the tape drives themselves can be used closer to 100% utilization other than for changing media or routine maintenance.

Tape continues to evolve in terms of its usage or deployment scenarios as well as a technology. Table 1 shows how tape, specifically LTO continues to evolve with its planned roadmap. The roadmap in Table 1 shows improvement in terms of native as well as compressed data storage space capacity in the same form factor along with performance enhancements.

> **Myth and Reality: HDD and Tape**
> Ironically, the HDD has not killed tape as a medium, rather, it has helped reinvigorate and keep the technology active in a complimentary manner. What is even more ironic is that the HDD has also been declared dead by some at the hands of Solid State Devices (SSD) which in fact are helping to give HDD new and extended life by off-loading some storage functions similar to how HDD are helping to keep tape relevant.

LTO	Introduced	Capacity native	Performance native[3]	Capacity compressed	Performance compressed	Functionality
1	2001	100GB	20MB/sec	200 GB	40MB/sec	2:1 Compress
2	2003	200GB	40MB/sec	400 GB	80MB/sec	2:1 Compress
3	2005	400GB	80MB/sec	800 GB	160MB/sec	2:1 Compress, WORM
4	2007	800GB	MB/sec	1.6 TB	240MB/sec	2:1 Compress, WORM, Encryption
5	2010	1.5TB	140MB/sec	3 TB	280MB/sec	2:1 Compress, WORM, Encryption, Partitions
6	TBA	3.2TB	210MB/sec	8 TB	525 MB/sec	2.5:1 Compress, WORM, Encryption, Partitions
7	TBA	6.4TB	315 MB/sec	16 TB	788MB/sec	2.5:1 Compress, WORM, Encryption, Partitions
8	TBA	12.8TB	472MB/sec	32 TB	1,180MB/sec	2.5:1 Compress, WORM, Encryption, Partitions

Table 1 – LTO technology roadmap showing tapes evolving future

---

[3] Technology is capable of supporting data transfer speeds up to these rates, your performance may vary

In addition to the performance and capacity improvements also shown are enhanced functionality such as the previously introduced Write Once Read Many (WORM) and encryption. Two other functionality enhancements shown in table 1 include media partitions along with compression enhancements. LTO partitions have the ability to store more frequently used data near the beginning of the media. This capability lays the foundation for tape to function more like traditional HDDs in the future for example placement of a dedupe catalog or index at the beginning of a tape in a partition to help enable safe, secure and efficient use of deduped tape.

Storage tiering and technology alignment

Technology alignment, that is aligning the applicable type of storage medium and devices to the task at hand in order to meet application service requirements, is essential to achieving an optimized and efficient IT environment. For very I/O intensive active data, Figure 1 shows leveraging high-performance SSD (Flash or RAM) Tier-0, for high I/O active data, Tier-1 fast 15.5K SAS or Fibre Channel based media are examples of aligning the right technology to the task at hand.

For low activity applications or inactive (idle) data with a focus of storing as much data as possible at the lowest cost is the objective, such as disk-based backup, slower, high capacity SATA-based storage systems are a good fit. For long-term bulk storage to meet

Figure 1 - Balancing storage PACE[4] to service and cost requirements

archiving, data retention or other retention needs as well as storing large weekly or monthly full backups, tape is the ticket (Figure 1) with the best combination of performance, availability capacity and energy efficiency per footprint.
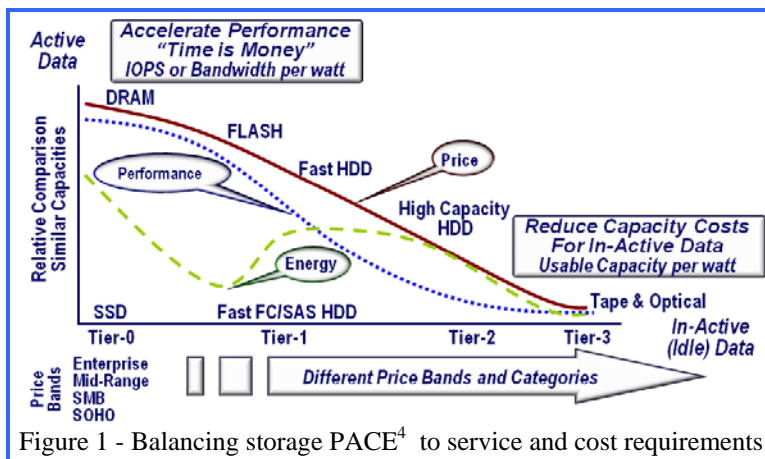
For low activity applications where storing as much data as possible with the lowest cost is the objective, slower, high capacity SATA based storage systems are a good fit. For long-term bulk storage to meet archiving, data retention or other retention needs as well as storing large weekly or monthly full backups, tape is the ticket (Figure 1) with the best combination of performance, availability capacity and energy efficiency per footprint.

Figure 2 shows how disk and tape combined can enable efficient and effective data protection by leveraging each technology's strength. Short-term data retention is shifting from tape to disk-based data protection. D2D based data protection technologies and techniques include snapshots, replication, backup and restore for BC and DR where disk is used as a cache or staging area before sending data to tape along with compression along with policy-based deduplication. For short-term data protection, emphasis is on RTO and RPO for speed of restoration for business recovery and survivability along with maximizing tape drive and media utilization.

---

[4] PACE = Performance Availability (including reliability) Capacity and Energy efficiency
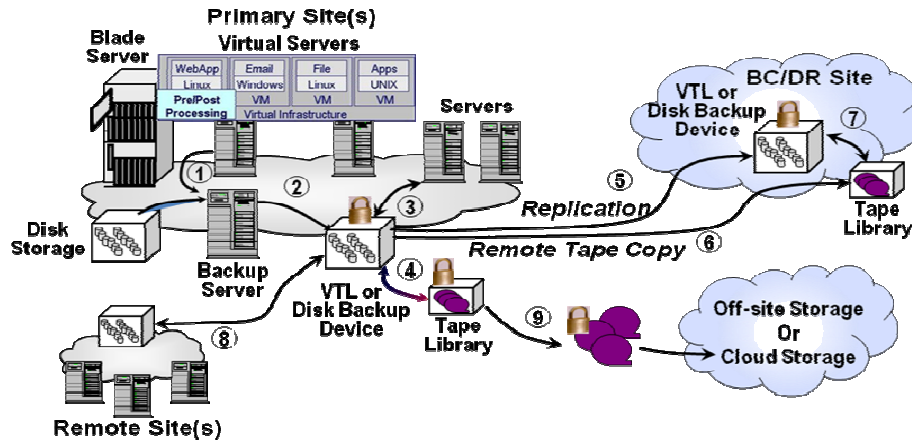
Figure 2 - Evolving data protection and retention paradigm: *Disk and Tape co-existence*

As an example shown in Figure 2, D2D is well suited for staging initial backups too (steps 1, 2, 3 and 8 in figure 2) and then replicating to another disk-based backup device or virtual tape library (VTL) (step 5 in figure 2) or making subsequent tape copies (step 4 in figure 2) on or off-site (steps 6, 7 and 9 in figure 2) for BC/DR purposes. Backups staged to disk enable rapid restore of data leveraging disk as a cache or holding area, with data being streamed to tape more efficiently.

Long-term data retention includes weekly, monthly or other long-term backup, primary backup copy of data, off-line copy of static or fixed content data, archive and strategic data preservation. The emphasis is on low cost, long-term durability, compatibility, and energy efficiency for lengthy data retention. In Figure 2, tape is leveraged as a high performance bulk storage medium to off-load the disk cache, boosting the effectiveness and utilization of disk-based systems. From a green and economic efficiency standpoint, data staged off-line to tape consumes no energy while enabling exceptional performance during bulk restore operations. The combination results in both very green and economically efficient storage in addition to supporting business sustainability and enabling compliance.

A tape copy operation may be made locally and then physically transported to another location for safe off-site storage, or data may be replicated as part of the backup and data protection process to a remote VTL or tape library where a removable tape copy is made. Hybrid solutions also leverage disk- to-disk locally with snapshots or other point-in-time copies that are then replicated to another location or to a cloud-based storage managed service provider (MSP). Data and network bandwidth optimization techniques and technologies, including compression and deduplication among others, enable more data to be moved on available networks or to reduce networking requirements.

For example, backups, snapshots or other time-based data copies are initially made to disk where data resides for some relatively short period of time for rapid restoration while a copy is also streamed to tape from the disk backup for better utilization of tape drives and tape media. For added security, an extra copy of the data can also be made to a secondary tape, a so-called "gold" copy, which goes to an off-site facility for an extended duration. After a period of time, data on the disk backup is removed to free up space for additional backups. Essentially, the disk backup provides a cache or staging area for rapid restore, improved tape drive and media usage as well as for facilitating faster backups to meet data protection windows and compliance requirements.

Tips and recommendations

Leveraging the right technology, tool and best practice techniques are important for an optimized data storage environment. In order to obtain maximum reliability, routine maintenance should be performed on all magnetic media including disk and tape. Routine maintenance includes regular proactive data or media integrity checks to proactively detect potential errors before they become a problem. For disk based online primary as well as secondary and D2D solutions media maintenance involves proactive drive integrity checks or powering up spun down disks along with background RAID parity checks. Since tape is often used as a portable medium to get data to off-site facilities for long term retention, it can also be exposed to different environments during transit. With the role of tape shifting to longer term data retention, routine maintenance is a best practice to achieve media life of 20 years. Media verification can be accomplished using software, appliances as well as functionality found in some tape libraries.

In addition to media management, another import best practice is securing data during transmission as well as at rest. This means leveraging encryption to provide data security and information protection compliance which for some geography locations is a regulatory requirement. As part of a long term data retention strategy and data protection, verify that encryption keys are also safely secured as well as available when needed.

General tips and comments:
- Factor in total cost of ownership (TCO) and return on investment (ROI) considerations
- Archiving is a useful technique for managing compliance and noncompliant data
- Long-term data retention applies to all types of data that has business value
- Implement tape and media tracking along with data protection management
- Audit and periodically test all data protection media, processes and procedures
- Adhere to vendor recommended media management and handling techniques
- Incorporate a media and data migration plan as part of an overall data retention strategy
- Perform regular verification and maintenance to prolong life of magnetic media devices
- Align the applicable technology, for example storage tier, to the task at hand.

Putting it All Together - Building a Solution with Quantum's Help

An example of how magnetic tape continues to evolve and co-exist with disk-based data protection technologies is shown in figure 3. Building on the previous example figures, disk and tape technologies co-exist being used for different purposes to leverage their respective capabilities to meet service level requirements and economic objectives (IT budget management). For example, disk-based data protection devices such as the Quantum DXi-Series provide virtual tape library (VTL) tape emulation along with file or NAS based backup capability for co-existence with existing software and data protection process.

The disk-based DXi-Series also incorporate data replication and policy based data deduplication are shown in figure 3 being deployed in remote office branch office (ROB), work group or departmental sites with data replicated to a larger DXi-Series based system in a primary data center. At the primary data center disk-based DXi-Series systems also support local backups and data protection in addition to being a collection point for distributed backups.
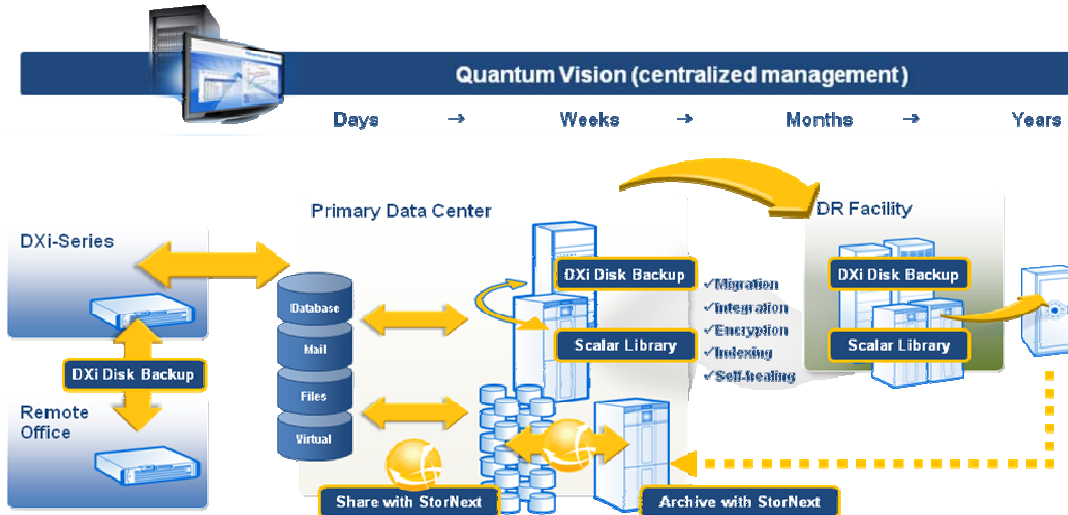
Quantum's Edge-to-Core approach

Figure 3 Example Quantum based Long-Term Data Retention Solution

To maximize disk buffer or cache space on the disk-based DXi-Series systems, data can be compressed and reduplicated using policy based algorithms to balance performance and capacity to meet specific application RTO and RPO requirements. To further maximize disk space and reduce cost of storing data, Quantum DXi-Series based systems also support path to tape (PTT) for moving backup as well as archive and other long-term retention data off-line to LTO and other tape-based mediums either on, or off-site. Tying the solution together are data protection management tools including Quantum Vision.

Quantum Scalar tape libraries with integrated iLayer management complete the example solution. iLayer management simplifies management tasks and helps manage growing data volumes. One way iLayer can help IT staff is by automating report creation and distribution, for example tape drive utilization and media usage reports. iLayer also saves time and improved reliability by proactively monitoring hundreds of system components, intelligently presenting information to users in plain English and offering steps to resolution.

Scalar libraries offer many features that improve business continuance and disaster recovery operations. Features like Control Path and Data Path Failover help reduce downtime and improve BC. Media Data Integrity Analysis (MeDIA) offers customers the capability to check the data integrity of tapes being stored for DR or long-term archive purposes, protecting their data at rest.

Scalar tape libraries support multiple options for encryption key management. Customers can choose to manage their encryption keys via the backup application, RSA from EMC, or with Quantum's Scalar Key Manager (SKM). SKM provides a highly secure key management solution (FIPS Level 1 certified) in a flexible and easy to manage package. SKM can be deployed as an appliance pair or as a pair of virtual machines, both offering redundancy. SKM is integrated with the iLayer management software for easier administration and reporting. Learn more about Quantum data protection and long-term data retention solutions along with other related topics at www.quantum.com.

Summary

Tape's role is changing and being complimented by magnetic hard disk drive (HDD) coupled with virtualization in the form of virtual tape libraries (VTL) and data footprint reduction (DFR) impact technologies (compression and dedupe among others). Given how no energy is required to store data when it is not being accessed, tape remains the most effective and efficient medium for retaining inactive data from both a Green environmental as well as an economic standpoint. An example of how HDD and tape co-exist and complement each other is having smaller more frequent data restoration coming from routine backups that have been deduped to save space while larger full volume or large scale restoration comes from a master gold or tape copy.

With its economic and environmental including power and cooling benefits, time-tested reliability combined with new and emerging enhancements, it is safe to say that tape still has plenty of life left in it. With more data being stored on tape today than in past history requiring fewer media cartridges and drive devices combined with an evolving roadmap and shift towards long term data preservation usage, the reality is that tape is alive.

What is dead is the myth that tape is no longer relevant.

The bottom line is that tape remains the most cost effective, energy efficient and reliable program technology for storing large amounts of inactive, off-line or near-line data while complimenting disk-based data protection and retention solutions. Keep in mind that you cannot go forward if you cannot go back: as a business, in order to provide sustainably, being able to go back in time and access preserved and protected data insures business sustainability.

Why tape remains relevant
- No such thing as a data recession
- Support data footprint reduction (DFR)
- Increased data density to reduce costs
- Proven media reliability and longevity
- Energy efficiency enabling Green IT
- Regulatory and other compliance needs
- Routine backups shift to a D2D model
- D2D backups serve as a buffer for tape
- More efficient use of tape devices
- Technology enhancement roadmap

About the author
Greg Schulz is founder of the Server and StorageIO Group, an independent IT industry advisory consultancy and research firm. Greg is also author of the books *The Green and Virtual Data Center* (CRC), *Resilient Storage Network: Designing Flexible Scalable Data Infrastructures* (Elsevier) and coming summer 2011 *Cloud and Virtual Data Storage Networking* (CRC). Learn more at www.storageio.com, www.storageioblog.com or on twitter @storageio.