

## **Long-Term Data Protection and Retention**

### **Finding the Correct Balance**

**By Greg Schulz**

**Founder and Senior Analyst, the StorageIO Group**  
**Author “The Green and Virtual Data Center” (CRC)**



**March 18, 2009**

**This Industry Trends and Perspectives Paper is Compliments of:**

**Quantum®**  
**www.quantum.com**

## **Long-Term Data Protection and Retention: Finding the Correct Balance**

### **Introduction**

IT organizations of all sizes contend with a growing data footprint with more data to manage, protect and preserve for longer periods of time. Online primary storage, has focus a on fast low-latency, 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 footprint - for inactive data.

### **Background and Issues**

The emphasis of short-term data retention is on data protection focused disk-to-disk 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 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. Another focus of long-term data retention is the ability to rapidly recover data with performance as well as with industry standard technologies for future proof and investment protection.

With the continued adoption of disk-based data protection, the demise of tape as a viable data storage medium has been greatly exaggerated. Instead, tape is currently undergoing a renaissance of sorts as a complimentary technology to disk and other tiered data storage mediums. Tape remains relevant for long-term data retention needs while supporting compliance, green energy efficiency, cost containment and other business sustainability requirements.

Technology improvements in disk-based 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. However, 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:

- Business and regulatory requirements – regulatory compliance and data preservation
- Economic and budgetary concerns – doing more with 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

## **Long-Term Data Protection and Retention: Finding the Correct Balance**

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. The notion is to think beyond regulatory compliance. In other words, organizations of all sizes need and rely on information, both current and past.

In addition to compliance, long-term data retention is needed for many other purposes. For example, information is needed to spot trends, support litigation or other business functions as well as normal functions including research & development, manufacturing and service support, sales and marketing - including customer relationship management (CRM) - along with human resource and accounting or financial activities. Magnetic tape (tape) continues to be relevant as a cost effective, business sustainable and energy efficient (green) proven technology for long-term reliable data retention needs.

As mentioned during the introduction of this paper, there are common myths about backup, data protection, and archiving both for long-term and short-term as well as about different types of data storage mediums and technologies. For example:

**Myth:** Disk drives have rendered tape technology obsolete.

**Reality:** Tape evolves (performance, availability, capacity, energy efficiency, density).

**Myth:** Tape is unreliable and not well suited for long-term data retention.

**Reality:** Digital tape is field proven with the ability to preserve data for decades.

**Myth:** Tape requires regular maintenance adding to cost and management complexity.

**Reality:** Over time, data on any digital storage media including disk will need to be migrated.

**Myth:** Tape is too slow for use compared to disk-based technologies.

**Reality:** Tape excels at streaming data at high performance rates for bulk retention.

**Myth:** Tape compression ratios are inferior to disk-based data deduplication.

**Reality:** Deduplication trades performance to improve disk density capacity benefit vs. tape.

**Myth:** There is no new development or technology innovation taking place with tape.

**Reality:** Vendors continue to evolve tape performance, availability, capacity and economics.

**Myth:** Debates are about disk vs. tape, solid state disk (SSD) vs. disk and tape.

**Reality:** Instead of debates, discuss what technology is best suited to given tier of service.

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 backups followed by a copy of data also being made to tape. Technology alignment, that is aligning the applicable type of storage medium and devices to the task at hand to meet application service requirements, is essential to achieving an optimized and efficient IT environment. For example, for very I/O intensive active data, as shown in Figure 1, leveraging high-performance SSD (FLASH or RAM) Tier-0, or for high I/O active data, Tier-1 fast 15.5K SAS and Fibre Channel storage based systems.

**Long-Term Data Protection and Retention: Finding the Correct Balance**

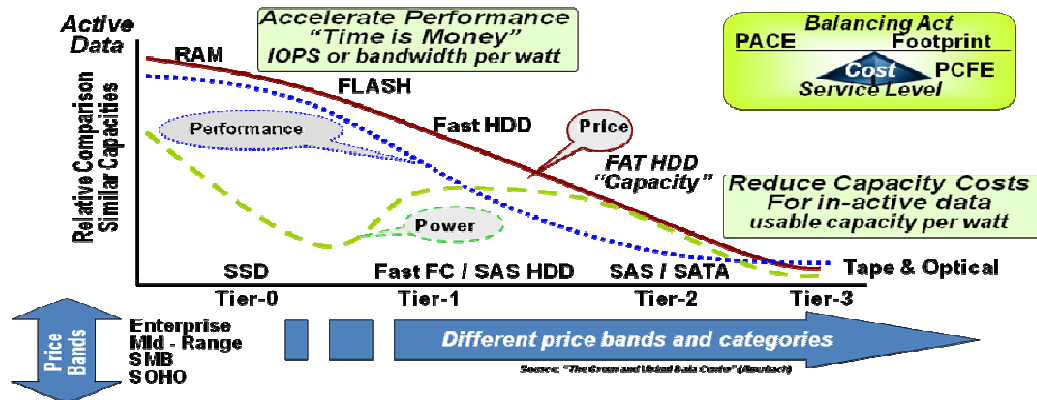


Figure 1 Tiered Storage: Balancing Performance, Availability, Capacity and Energy to QoS<sup>1</sup>

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.

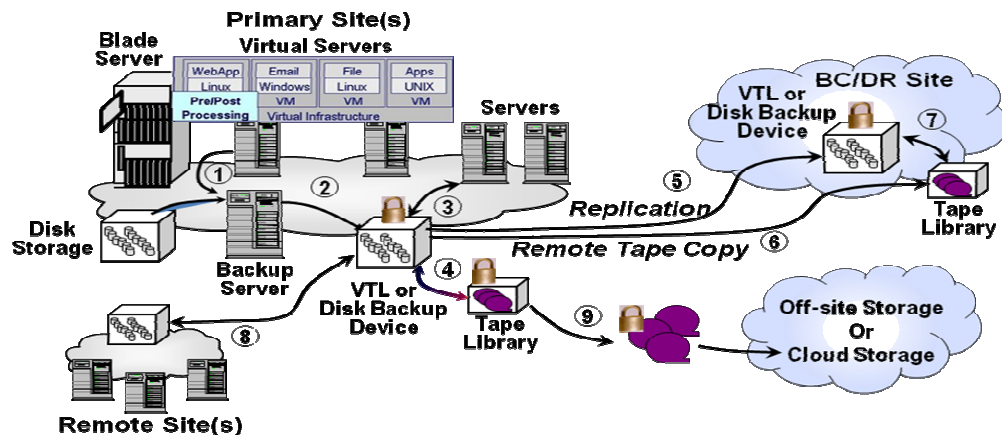


Figure 2 Evolving data protection and retention paradigm: *Disk and Tape co-existence*

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. Disk-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. The emphasis is on RTO and RPO for speed of restoration for business recovery and survivability along with maximizing tape drive and media utilization. In order to maximize disk space, policy based data deduplication immediate mode can be used to reduce the data footprint of backed-up data. On the other hand, in order to meet RTO and backup window or performance requirements, policy based data deduplication deferred mode can be used.

<sup>1</sup> Quality of Service (QoS) to meet business service requirements and objectives

## **Long-Term Data Protection and Retention: Finding the Correct Balance**

For example, disk 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.

### **Reasons Tape Remains Relevant**

Best in class energy efficiency for storing large amounts of data for long periods of time with rapid restore capabilities. In addition to low cost, energy efficiency and high performance for storing ultra-dense data over time, tape is a proven technology for preserving data integrity with industry standard compatibility that continues to be evolved to compliment disk and other technologies thus enabling tiered data and storage services.

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.

### **Putting it All Together - Building a Solution with Quantum’s Help**

An example of showing 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 shown in figure 2 and tiered storage model shown in figure 1, 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.

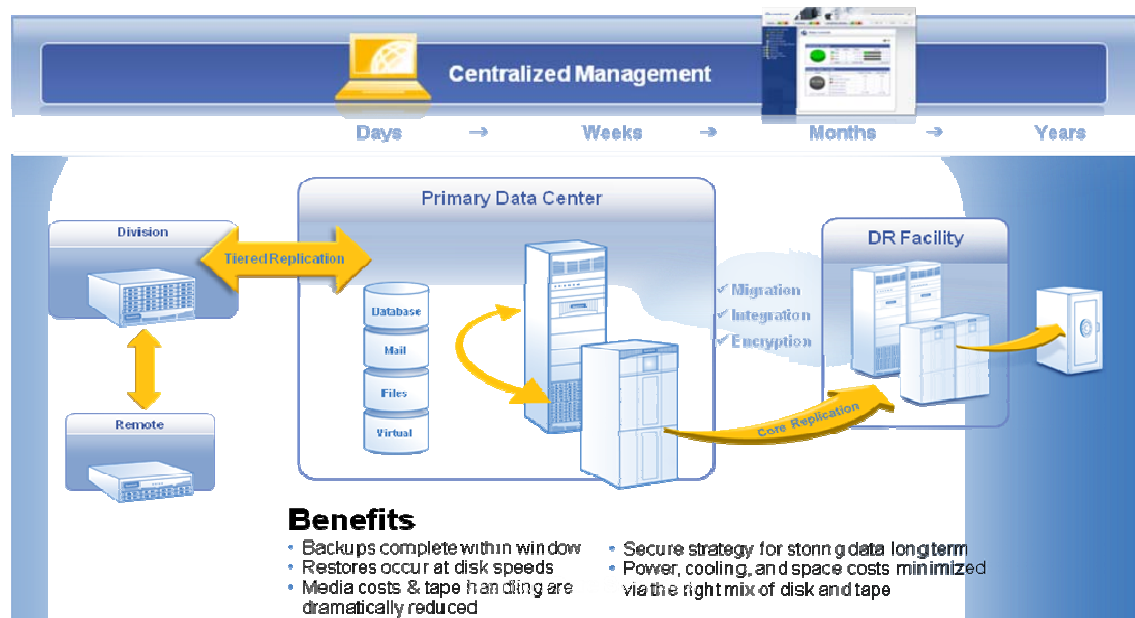


Figure 3 Example Quantum based Long-Term Data Retention Solution

The disk-based DXi-Series data protection systems shown in Figure 3 provide rapid and flexible data protection and recovery for remote or distributed sites while enabling BC/DR with data replicated back to a primary site. 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. 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. Learn more about Quantum data protection and long-term data retention solutions along with other related topics at [www.quantum.com](http://www.quantum.com).

## Conclusion

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

## Long-Term Data Protection and Retention: Finding the Correct Balance

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.

	<b>Remote Office Branch Office (ROBO)</b>	<b>Short-term Data Retention</b>	<b>Long-term Data Protection and Retention</b>
<b>Retention</b>	Hours, Days, Weeks	Hours, Days, Weeks	Weeks, Months, Years,
<b>Storage</b>	Disk to Disk	Disk and Tape	Tape
<b>PCFE or Green</b>	Disk to Disk Compression / Dedupe	Snapshots, backups, tape, disk, dedupe, compress disk and tape	Disk to tape, replication or copies of tape off-site, compress and encrypt tape media

Table 1 - Technology Co-Existence: What Data Storage Technology to Use When

A combination of tape and disk-based data protection, including archiving for data preservation, coupled with a data footprint reduction strategy can help to address PCFE or green concerns while meeting other needs and issues.

### 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.
- Networks can move more data than in the past; however, there is more data to move.
- Incorporate a media and data migration plan as part of an overall data retention strategy.
- Align the applicable technology, for example storage tier, to the task at hand.
- If you can't go back, you can't go forward

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.

### About the author

Greg Schulz is founder of the StorageIO Group, an IT industry analyst and consultancy firm as well as author of the books *The Green and Virtual Data Center* (CRC) and *Resilient Storage Network* (Elsevier). Learn more at [www.storageio.com](http://www.storageio.com).

*All trademarks are the property of their respective companies and owners. The StorageIO Group makes no expressed or implied warranties in this document relating to the use or operation of the products and techniques described herein. The StorageIO Group in no event shall be liable for any indirect, consequential, special, incidental or other damages arising out of or associated with any aspect of this document, its use, reliance upon the information, recommendations, or inadvertent errors contained herein. Information, opinions and recommendations made by the StorageIO Group are based upon public information believed to be accurate, reliable, and subject to change.*

WP00130-v01