

I D C V E N D O R S P O T L I G H T

Making the Economic Case for Data Preservation

October 2008

Adapted from *Worldwide Removable Hard Disk Drive 2008–2012 Forecast and Analysis: The Quest for a Viable Tape Replacement* by Robert Armatruda, John Rydning, and David Reinsel, #210169; *Worldwide Archiving Software 2007 Vendor Shares: Can Archiving Go Mainstream?* by Laura DuBois, #213311; and *Worldwide Archiving Software 2008–2012 Forecast* by Laura DuBois, #212216

Sponsored by ProStor Systems

Most organizations realize that the requirements for accessing specific types of corporate content change over time, in particular the requirements for fixed content, which may need to be preserved and retained but which is infrequently accessed. Factors such as how long to keep fixed content and the implications of how that content is retained are influenced by operational, legal discovery, and regulatory mandates. How fixed content is retained can have a positive impact on backup windows, power and cooling requirements, storage costs, and administrative overhead. More organizations are implementing solutions that meet increasingly stringent business mandates for record retention and retrieval requirements while also providing greater efficiency and lower costs.

The prevailing wisdom of data storage is mired in the concept that the cheapest long-term storage medium is tape. Yet for all its perceived benefits, tape has shortcomings when it comes to satisfying the need for rapid access to critical data. The sequential nature of tape, while inexpensive as a medium, does not offer prompt and precise access to fixed content. This paper examines lower-cost, intelligent, nonsequential media alternatives, along with the concept of "data preservation" and the higher value obtainable from an economics-based solution. The paper also looks at the role of ProStor Systems and its RDX® hard disk storage cartridges and companion ProStor InfiniVault® solution for addressing the economic, business, and technology issues surrounding the short- and long-term storage of static data.

The Business of Information Management

Every business is in the business of information management. Sound data storage practices are the foundation of effective information management. In fact, IDC studies point out that small- to medium-sized businesses (SMBs) have the same data storage issues and concerns as large enterprises. Diverse stakeholders, many of whom are nontechnical, are changing company data storage and retention requirements such as the following:

- Compliance with business rules and policies for record retention requirements
- Improved efficiency and accuracy of ediscovery
- Continuing IT storage optimization improvements (including a "green" emphasis)

Businesses are finding they need to store more and more unstructured content, such as email messages. This often places greater demands on storage optimization in the areas of performance, primary storage capacity, and backup procedures. This trend will continue, and future solutions will

need to address an increasingly broad spectrum of content types. For example, frequent ediscovery searches on structured data represent a significant driver, but searches on unstructured data pose the greatest risk, disruption, and unknown costs to a business. IDC data suggests that as business size or complexity increases — and with it the promise of greater revenue — so does the possibility of increased legal risk and quicker responsiveness to discovery.

These considerations place greater demands on data storage systems, both primary and backup. The amount of data on disk-based primary storage and on tape backup continues to increase every year because of the ever-growing demand from the business, based on the needs elucidated earlier. In addition, there are the following IT infrastructure issues:

- Continued penetration of disk storage in the data protection path
- The need for selecting different recovery services based on application priority
- Protecting the system image for both physical and virtualized systems
- Protection of remote or edge data, often managed centrally
- Strategies for deduplication and optimization of storage facilities
- Inadequacy of current replication and continuous data protection (CDP) backups with no reporting capabilities
- Evaluation of emerging storage-as-a-service offerings

The Growing Need for More Immediate Access

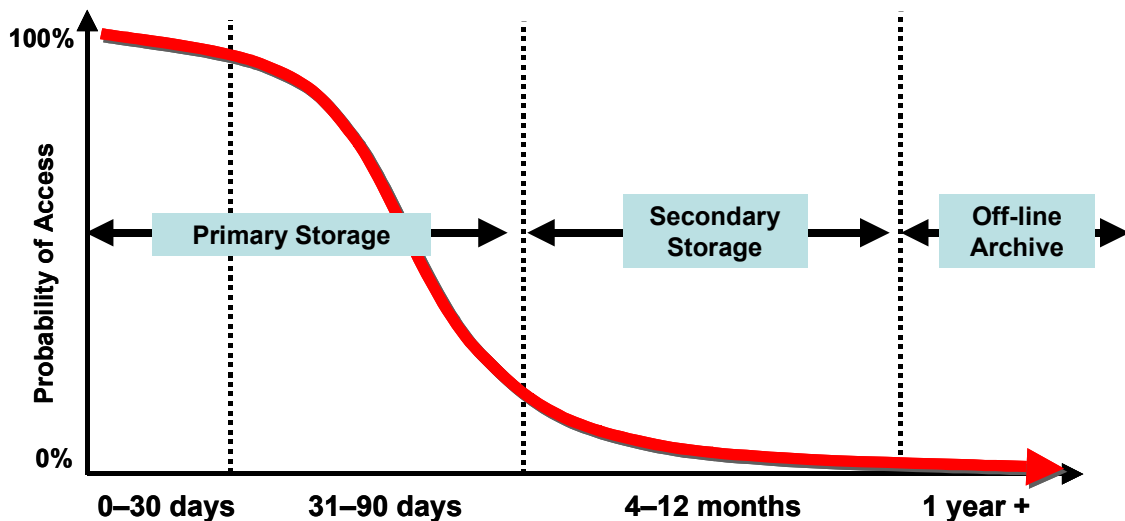
Data stored on disk drives is rapid and nonsequential, making disk an ideal primary storage medium. Conversely, storing data on tape is effective, but it lacks the performance of disk due to the sequential manner in which data is written on physical tape cartridges. Historically, tape storage cost far less than conventional fixed-disk storage, especially when new or incremental capacity was added. This cost differential of adding capacity on tape has made it the preferred medium for backup and archive. Tape storage is well suited for long-term preservation and disaster recovery from system crashes. Nevertheless, when retaining data for short- or near-term business access, such as compliance or ediscovery, software tools are often required to locate and manage tape-based data for presentation as information. Consider the following:

- In most businesses, static data — that which is not in current use — represents between 50% and 80% of all stored data
- Dynamic, or in-use, data is newer and in higher demand than static data and has a higher probability of need for access
- Access inflection points occur with specific frequency, usually within 30–90 days of data creation, and determine when data is shifted from one storage medium to another

Interestingly, probability of access patterns can be predicted, as shown in Figure 1.

Figure 1

Data Access According to Probability



Source: ProStor Systems, 2008

Once understood, probability of access patterns can be examined and used to determine data preservation strategies and to obtain economic advantage from storage systems.

The Economic Advantages of Data Preservation

Traditional storage and backup tactics are based on the simplest form of protection: make a copy. Conventional tape backup, however, has become arduous and cumbersome: weekly, monthly, quarterly, annually, full, incremental, multiple, onsite, offsite. There can be hidden costs as well: the cost of ever more tape cartridges in rotation or archive, changes in capacity or format, and so forth. Tape being sequential by design means it will always be difficult to retrieve data quickly. Moreover, new generations of tape cartridges and technology migrations create new budget line items and force higher total cost of ownership (TCO).

Today, a business' primary storage capacity requires between seven and 20 times as much tape backup storage space. Consider the following example, supplied by ProStor Systems, of backing up using tape cartridges when adding more primary storage. This scenario examines the total protected cost of adding 10TB of primary storage to an existing datacenter:

- A new 10TB disk must be added at a cost of \$50,000 (industry average for controller, disk, and software).
- Backup for new drive will need minimum of seven times its capacity, or 70TB.
- Four new tape drives to serve this capacity are required, costing \$4,000 each, for a total of \$16,000 (life: six years).
- 63 tape cartridges are required, costing \$100 each, for a total of \$6,300.

The total cost without backup licenses, administration, or maintenance:

- 5 years of storage: \$72,300
- 30 years of storing: \$411,500

At a 7:1 backup-to-primary storage cost, this is hardly a cost-effective solution. Nor is backing up to fixed-disk storage the answer; it's even more expensive and requires at least three times the primary disk space.

From Data Storage to Data Preservation

Even if they were not so cost-prohibitive, traditional data retention and backup tactics are inadequate for today's needs. Businesses must think about information management, and that requires a different approach to storage: data preservation.

Data preservation is a strategy that takes into account both short-term and long-term data requirements:

- Short-term data preservation is focused on disaster recovery or protecting against a malignant loss of data.
- Long-term data preservation is intended to preserve static data for business rules, regulatory or compliance requirements, and historical records management.
- Preservation techniques for short-term and long-term data are identified by analyzing patterns of usage and determining the inflection points mentioned earlier.

Being able to understand and predict inflection points determines the business strategy for data preservation, which in turn determines storage optimization. This is the basis for exploiting an economic advantage for storing and retaining data.

A Strategy for Cost-Effective Data Preservation

Some may claim tape backup is "good enough," but in truth the time for good-enough backup has passed. Business, regulatory, and user demands, and especially costs, have grown exponentially just since 2000. Well-practiced data preservation methods can turn around the rising expense of backing up primary storage, making frequently used data more accessible, more quickly. Deploying a data preservation strategy begins by answering the following questions:

- How much data do you manage?
- What types of data do you store?
- What types of data are problematic?
- What are the primary storage requirements for various data (e.g., business records, taxation, compliance)?
- How long must you keep the data?
- How much does storing data over time cost?
- What data can't be deleted because it may be needed within the next 90 days?
- What are the power costs for your backup systems, annually and at fixed amortization points in the future?

Considering ProStor Systems

Answering the preceding questions will lead to a determination of your firm's capital expenses for long-term static data storage and its TCO. Using this information and the tape backup example cited earlier, you can now compare the cost benefits of data preservation and storage optimization using an up-to-date storage solution. Consider the same example from ProStor of a datacenter needing to add 10TB of disk capacity, but deploying the ProStor InfiniVault solution instead of adding primary storage and tape cartridges:

- ProStor InfiniVault Model 30 at a cost of \$33,000
- 20 each 500GB/1TB RDX disk cartridges required (for two copies) at a cost of \$7,800

Total cost, without backup licenses, administration, or maintenance:

- 5 years of storage: \$40,800 — savings over tape: \$31,500
- 30 years of storing: \$205,800 — savings over tape: \$205,700

In addition to delivering a significant economic advantage over conventional backup for data protection of primary storage, the ProStor InfiniVault solution provides the following additional benefits:

- Better data management: no application goes unserved; shrinks the backup window; full access to metadata
- Improved data preservation and storage methods: completely searchable; quick access; all storage is on high-capacity, nonsequential fixed disks
- Green: fixed-disk cartridges can be stored offline so they are not consuming power 24 x 7

The ProStor InfiniVault system comprises the RDX removable hard disk drive cartridges, RDX bays in removable disk units, and an intelligent operating system designed to automate the data management on the disks.

The RDX Disk Technology

ProStor's patented RDX disk cartridge technology, introduced in 2005, uses removable hard disk drive cartridges for high-speed, active access to retained data as well as optimized backup procedures. Similar in size to a tape cartridge, the RDX cartridge contains a ruggedized, removable 2.5-inch SATA hard disk drive in capacities ranging up to 500GB. The RDX cartridge stores data nonsequentially, compressed at a 2:1 ratio, at a native transfer rate of up to 45MBps. A single RDX cartridge can be used in a desktop docking unit. RDX cartridges are available from select VARs and OEMs such as Dell and Hewlett-Packard.

ProStor plans to introduce cartridges of 1TB capacity and higher in early 2010. All current cartridges are, and future cartridges will be, forward- and backward-compatible and thus never obsolete. The minimum expected service life is 5,000 load/unload cycles for cartridges and 10,000 load/unload cycles for docks. Recently completed third-party testing has proven a shelf life of at least 30 years for the RDX medium, equaling or exceeding the most optimistic expectations for tape as an archival medium.

ProStor InfiniVault

In 2007, ProStor Systems introduced ProStor InfiniVault, an enterprise-class solution for midmarket businesses wishing to take full advantage of optimizing their IT infrastructure while also ensuring they are able to cost-effectively retain static data to meet a company's business rules for retaining data, including the most rigorous regulatory compliance requirements. ProStor InfiniVault is a data preservation storage system that combines advanced storage intelligence on the technology side with a data preservation operating system to automatically manage a variety of business services, including retention and legal discovery requirements.

ProStor InfiniVault installs as a network mount point or network drive. The data preservation operating system provides interoperability with applications that use standard network CIFS or NFS protocols, including any archive or HSM applications such as those used for email archiving or unstructured data. Files retained in the ProStor InfiniVault system may be accessed over the network whenever needed. The retention period can be set for any length of time. Multiple copies of data can be made and sent to offsite storage for archival and disaster recovery purposes.

Unlike with other products, with ProStor InfiniVault, a file will remain visible to end users and applications even after the data has been removed from the ProStor system and the RDX cartridges are shipped to a safe offsite location. This means it is always easy to search through the metadata and retrieve information from a ProStor InfiniVault system. Storage intelligence software creates independent "vaults" (similar to a Microsoft folder) on a single system for common business documents from applications such as Microsoft Office, as well as for business records, human resources records, databases, and email messages.

The ProStor InfiniVault system consists of the following components:

- A system controller with policy-based data preservation software that manages retention and data actions (file deduplication, content indexing, authentication, access logging, compression, encryption, and deletion) and automates all interaction between the network disk and removable RDX drives
- A network disk with a standard network interface
- RDX bays and disk cartridges in removable disk units (RDUs) providing for nearline storage with removability for data vaulting
- Browser-based management

The systems are also connected to multiple RDX cartridges in stackable RDUs that hold up to 10 removable RDX hard disk drive cartridges each. Data is written to the RDX cartridge with WORM, ensuring immutability of the data. Systems can be configured with multiple RDUs, allowing up to 100 RDX cartridges to be connected to each system. ProStor InfiniVault systems all feature some amount of permanent RAID-protected storage, providing immediate access to stored information.

Having a choice of configurations allows a business to design the best system for its size and needs, from preservation-level storage with infinite capacity to information movement and protection-level offsite archiving using multiple RDX cartridges. Another important feature is that all data stored on a ProStor InfiniVault is written to the network disk and to RDX cartridges simultaneously, meaning that the data is always doubly protected.

The ProStor InfiniVault, coupled with removable RDX hard disk drive cartridges, is designed to provide the most productive and economical solution for data preservation. It uses the computer's native interface and appears on the network simply as a drive. Because it uses 2.5-inch hard disk drives, it's faster than traditional backup or archiving systems. Among its other technical attributes are the following:

- Data that is exported in standard interchange format for ediscovery purposes
- As economical as tape, but with greater reliability and longer shelf life
- AES 256 encryption and key management for security
- Content-addressable security using a hash key
- Content index for each stored file and text-level search capabilities
- Retention management and chain of custody for legal issues

- Information audit trail for tracking actions and data locations
- Single-instance storage — will not reduplicate at the file level in a given archive
- Data compression to maximize capacity
- Secure delete using digital shredding for documents whose retention period has expired
- Immutability ensured by WORM cartridge format and WORM-protected file area

Challenges and Opportunities for ProStor

Tape storage is very prevalent in customer environments. It's reliable and cost-effective for backup and archive functions. Over the years, many businesses have found tape backup cumbersome but "good enough." Tape storage is slower due to its sequential design of storing data on tape cartridges. However, today's businesses demand greater performance than tape can provide due to the increased awareness placed on regulatory and compliance requirements for access to information. Moreover, once carefully analyzed, tape can be easily proven too costly to continue using.

ProStor's challenge is to convince customers that its data preservation solution is superior to current tape-based backup and delivers greater economies of scale for their storage needs. We expect OEM and VAR agreements for the RDX technology will enhance the validity of the solution going forward.

If ProStor educates customers so that they understand a data preservation solution helps reclaim primary storage while efficiently retaining static data, that it reduces backup size along with power and space needs, and that it lowers administration costs, the company will have the potential to achieve considerable success.

Conclusion

The gap between business policy and technology execution is closing slowly. The point where they meet is information management, which is predicated on sound data management and preservation practices. End users are becoming more involved in decisions regarding how data is managed and preserved — of necessity because of administrative or compliance issues.

This trend is likely to continue, and as it does so, tape backup will prove less and less desirable. Users will be impatient with tape's slow and laborious retrieval and access characteristics. And as storage capacities continue to increase — and assuredly they will — IT will find it more and more expensive to justify, operate, and maintain a tape backup system. When the economic case for the removable hard disk drive cartridge can be made so simply — it costs less to buy, maintain, and sustain — and given the ease of transition and additional benefits that accrue, it does not seem like such a difficult decision.

ABOUT THIS PUBLICATION

This publication was produced by IDC Go-to-Market Services. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Go-to-Market Services makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

COPYRIGHT AND RESTRICTIONS

Any IDC information or reference to IDC that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. For permission requests, contact the GMS information line at 508-988-7610 or gms@idc.com. Translation and/or localization of this document requires an additional license from IDC. For more information on IDC, visit www.idc.com. For more information on IDC GMS, visit www.idc.com/gms.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com