Technology Insight Paper

Economic Considerations for Long-Term Data Retention

By Randy Kerns

January 2015

Evaluator Group

Enabling you to make the best technology decisions

Executive Summary

The amount of information created continues to increase and the data must be handled as a valuable asset. Data is being stored for longer periods of time because it may become more valuable with further analysis or because operation rules dictate that it be retained. The combination of increasing capacity demands and lengthening retention periods is a major challenge for information technology organizations to store data and provide the management and protection required in the most efficient manner possible.

An optimal approach in determining the most effective means for storing, managing, and protecting data uses economic analysis that accounts for the variety of factors that contribute to overall cost. Economic analysis aids in developing a technology strategy that will meet requirements and stay within defined budget parameters for capital and operational expenses. The analysis must cover a timescale driven by the lifespan of the data rather than physical devices in order to provide an understanding of the full economic effect of a technology strategy.

IT professionals use different technologies to address different problems in storing, managing, and protecting data. Developing a new strategy based on economic analysis brings a clear understanding of the value of introducing new solutions and operational processes. Storing information in content repositories - integrated with management and protection practices while retaining the access properties for data - provides an economical means to address these demands.

The Impact of Continuing Information Demands

Organizations are creating more valuable information that must be stored and managed as a corporate resource. Information is the result of processing data with an understanding of what the data represents. Increasingly, data has gained value over time with further analysis. At the same time, the practice of retaining data for longer periods is increasing as well. The overwhelming amount of increasing data to be stored, managed, and protected and its effects on IT operations require critical decisions when architecting solutions to address the challenges. Usually the initial evaluation focuses on the simple issue of the primary storage required. However, the impacts are more pervasive across IT and require a broader investigation overall. The factors causing the impact must be considered in the larger strategic decision.

Understanding Types of Data Causing Greatest Demands

First, it is important to understand the types of information driving increased capacity and retention demands. The primary task of many corporate data centers has been protecting structured files and database applications where multiple versions of data are created and saved. Now those same data centers must cope with an influx of unstructured digital content such as videos, imagery and audio files. Different information has different value to an organization, and characterizing the type of data should include an understanding of the value and management requirements.

Increased Requirements for Storing Data

Second, the longevity and availability for access requirements for data will determines the storage resources needed. Primary storage has certain performance and reliability characteristics, which may be best-suited for some data. Secondary storage or content repositories with different performance and expense characteristics that still maintain online access to information may be the correct storage resource for other data. The task of matching the cost of storage resources to the value of data adds to administrators' workloads. Complicating that fact is how the value of data or usability changes over time, and all of this means automated tools for managing data location and movement are critical.

Management and Protection Requirements

Storage resource considerations include the requirements for both managing and protecting data. Managing data means ensuring it is placed on the appropriate storage resource, retaining it according to the operational or business requirements, and ensuring the access availability required is maintained. Making data available based on its business value and operational environment is a specific consideration where costs for implementing availability mechanisms can vary. Data protection addresses how to cope with damaging operational conditions such as inadvertent deletion or corruption, and from physical problems like device failures. In both data management and data protection, storage resources and operational processes incur expenses and other potential impacts on the organization. At

some point when a practical capacity threshold is reached, increasing capacity typically requires redesigning the data management and data protection implementations.

Addressing Capacity and Long-Term Retention

There are various approaches to addressing the impacts of greater capacity requirements and the related problem of long-term retention. An economic analysis modeling the economic benefits using short-term return on investment (ROI) and the broad-view total cost of ownership (TCO) calculations will show the most effective solutions.

From an economics standpoint, using automation to move data to less expensive storage based on a set of criteria coupled with a change in the methods used for data protection delivers the greatest value for IT investment. Moving data to different systems fits into a category usually defined as tiered storage systems.

Moving Data to Match its Probability of Access

The simplest way to describe data movement is with a very high-level diagram. In this diagram, moving data between different storage resources based on criteria focuses on the data value or probability of access.

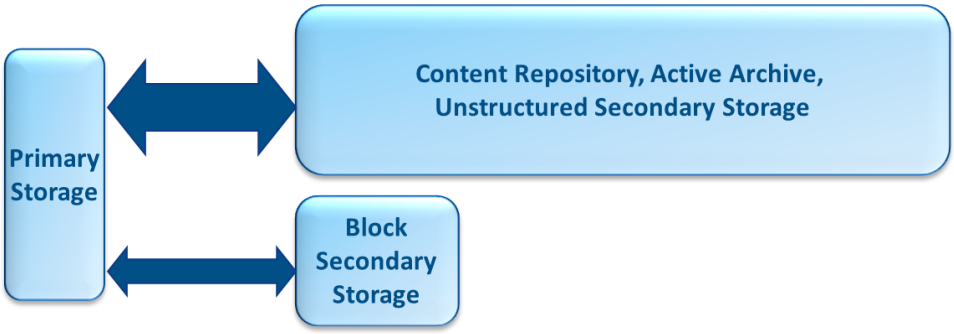

Figure 1: External Storage Tiers

To illustrate a potential set of data access requirements and the data protection change that occurs, the following diagram shows how the probability of access can be used, as one measure, to make a decision about moving data. The time scale at the bottom is just a common example of how the probability changes over time.

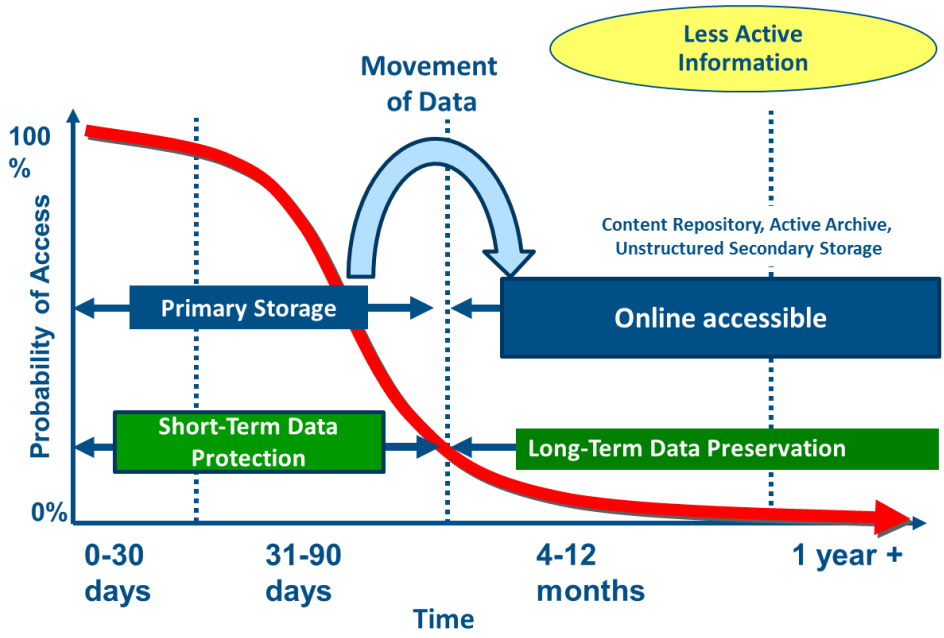

Figure 2: Probability of Access Changes Requirements

Active archiving is about moving data that is not required for immediate work and is not expected to change to a more economical storage system such as a content repository, and protecting it according to a set of rules when it is moved while maintaining the online accessibility. Anticipating when data will no longer need to be immediately accessible could be case-specific, such as when a project is finished and data is archived as a final record. The decision to move could also be based on a policy – for example, if data has not been accessed in six months it should be moved. The organization that owns the data may set the rules for storing information, or they may be determined by regulatory requirements.

Changing Data Protection

Data protection methods must be adjusted to handle increasing capacity demands and meet the requirements based on the value of data and the probability of access. Current practices may be unsustainable due to required time or expense. Greater efficiency and lower costs can be achieved using the same or similar criteria as for moving data between tiers of storage resources and the different storage technologies available. The following diagram matches different storage resources with data protection and availability requirements.

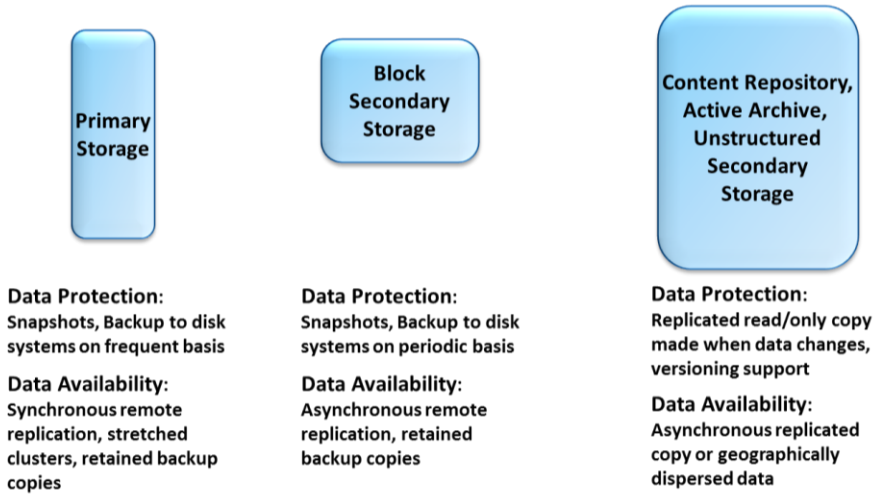

Figure 3: Protection and Availability Requirements for Storage Tiers

Economic Evaluations

Economic models that compare alternative data protection solutions give a projection of the value of those solutions. With the nature of storing some information for very long periods, a longer economic analysis period shows the financial impact of an effective strategy to address the capacity increase and long-term retention. Economic models are more than just a simplistic measure of price per capacity (represented as $/GB), as they include a variety of elements contributing to the true long-term cost.

Following are some of the leading factors that can influence an economic decision on a new strategy for growth and long-term retention. In some situations, there will be additional expenses incurred for specific requirements.

Factor	Explanation
Cost of storage system	This multi-part consideration depends on the implementation of storage tiering with data movement. Each system will have a different cost and lifespan. Savings come from shifting the data storage requirements to less expensive systems with longer useful lifespans. Many storage technologies include the ability to reduce the amount of data stored, usually through compression and deduplication. In effect, this reduction allows more data to be stored in a given amount of storage capacity, thus reducing the cost of storage.
Software costs for storage systems	Software features such as replication have charges that persist for the lifespan of the storage system. Usually the maintenance or support charges are among the top storage system expenses incurred.
Maintenance and	Some storage systems have longer warranty periods and different coverage than

support costs	others. The costs may escalate sharply, causing movement to a new technology.
Environmental costs	The costs of space, power, and cooling add up over time, but generally are dwarfed by other costs.
Administrative costs	Administration time can vary widely between systems. While administrative time is very visible, the actual financial impact is very modest compared to other costs.
Data movement or copy software costs	Additional software to move or copy data between systems - including backup and archiving software - is necessary for managing and protecting data. Usually charges are based on the amount of data moved or backed up and can grow commensurate with capacity. Capacity-based licenses for software can quickly dominate total cost of ownership and represent opportunity for changes to protection and availability.
External services expenses	Services used to retain copies of data - either as tapes in a vault, or in a public cloud - have associated costs that aggregate over time to significant amounts as capacity grows. Transfer costs such as bandwidth for moving data continue to increase with more data moved.
Technology replacement and migration costs	As technology ages, storage must be replaced. Replacing a storage technology requires moving stored data from the earlier technology to the new one. Some systems may have incorporated migration as a seamless capability. Without that capability, additional variables need to be included: • Overlap of new and old systems during the migration • Administrative time to monitor and manage the migration • Interruptions for configuration changes and switching between systems Administrative time is typically required, and often is not budgeted for the correct amount, impacting operations.

The following diagrams from an Evaluator Group project illustrate cost reduction areas and percentage of savings for each area over three- and ten-year periods by implementing a strategy to move data with lower performance and access demands to a content repository along with a change in data protection while still maintaining online access requirements. The basis for this example was a 1.5 PB primary storage environment with a growth rate of 35% per year. Movement of data to a content repository used probability of access criteria. The resulting savings are $8.5 million over a 3-year period and $74 million over 10 years.

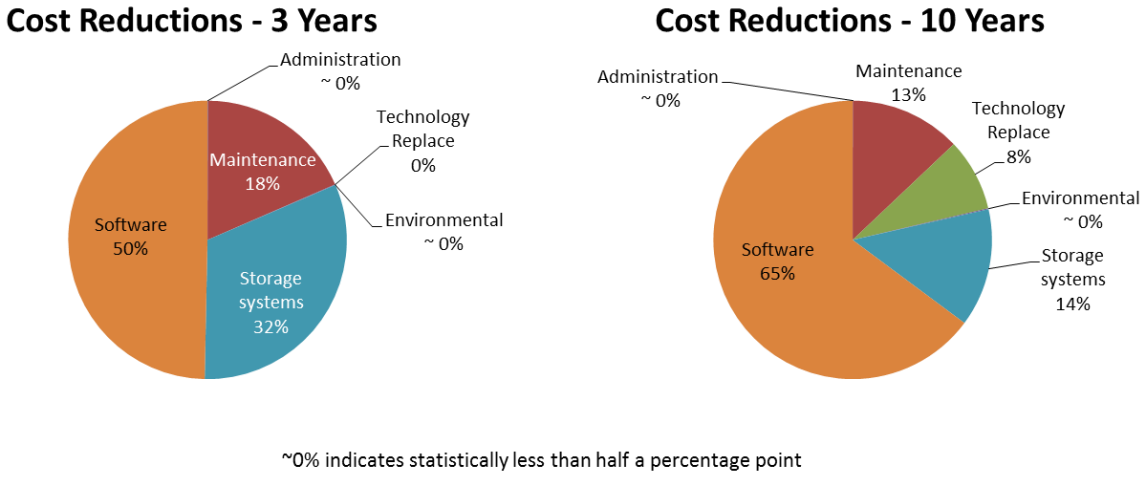

Figure 4: Cost Reduction Comparisons

In the example above, opportunities for cost reductions were greatest for the areas of software (backup and data movement), storage system purchases, and maintenance and support of storage system software. No external services or data movement software costs were required. The economic analysis produced information that led to a strategic technology decision to implement a content repository.

Technologies to Implement New Strategies

A new strategy to address capacity demands and long-term retention must consider the complete storage environment. There may be beneficial effects in other areas for storing data, management and protection. The following diagram depicts a potential change to the current practice. Text on the diagram indicates the data protection and data availability processes along with the movement of data.

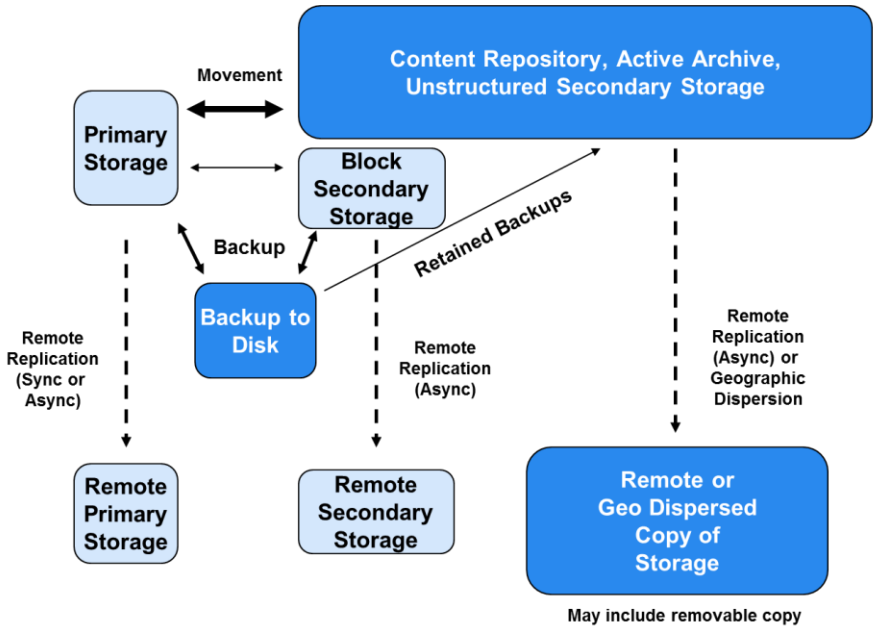

Figure 5: Strategy for Capacity and Protection

Developing a strategy to address capacity growth and long-term protection starts with a review of the environment where data protection and capacity changes can be made as part of the technology strategy.

Backup to Disk: While not part of long-term data protection, backup to disk is an integral part of data protection and increased capacity demand and must be considered in any strategy. Disk-based backup solutions operate as a target for backup software to protect primary and block secondary storage. Integrated deduplication and compression technology allows greater amounts of data to be stored, reducing the costs of data protection.

Content Repository, Active Archive, or Unstructured Secondary Storage: A storage system capable of scaling very large is a good candidate for use as a content repository or secondary storage for online access of unstructured data. There are several alternatives where data is online and available for access:

- Storage systems used for content repositories such as object-based private cloud systems offer cost-effective storage with global access, built-in data protection and required access performance. File access and object access are necessary to meet both current and developing needs.
- Tape library systems with Linear Tape File System (LTFS) management permit LTO tape devices to function as archiving systems. Using LTFS brings the longer lifespan of tape systems to the

economic equation. Tape library systems can scale to store massive amounts of data and remain directly accessible where immediate, sub-second response time is not required.

- Public clouds provide the archive storage resource either as the protected copy or for the basic content repository. Organizations can use a two-tier archiving implementation where the content repository or active archive is the on-premise system and a copy of data is sent to a public (or hosted) cloud for protection. Later, data can be aged from the on-premise system in the two-tier archive as a means to reduce the costs of storage. It should be noted that public cloud solutions need to be carefully examined to understand the potentially hidden costs associated with them. For example, many public cloud solutions charge additional fees for actively accessing data, and if the data is to be retained for many years, moving that data from a public cloud to another storage option can become cost prohibitive.

Management Systems: Part of a strategy to deal with capacity growth and long-retention includes data management. Software solutions on appliances or servers manage the storage and movement of information across multiple storage technologies according to automated policies established.

Summary

Increasing capacity demand and long-term retention of data are directly linked problems for IT organizations. A strategy needs to be comprehensive in addressing capacity and retention while meeting data protection and data accessibility needs. Economic value is the ultimate measure of different solutions, and an economic model is the best means to compare alternatives and understand the long-term implications. Creating a content repository is an effective means to address the challenges in the most economic manner.

Backup to disk is critical for timely data protection of primary storage and block-based secondary storage. Quantum DXi-Series appliances have data deduplication to increase the amount of data capable of being stored when serving as a backup target.

According to the IDC Tape OEM Market Share report for 2013, Quantum is the market share leader in tape automation. Quantum Scalar LTO Libraries can be partitioned to support both backup and archiving capabilities and include proactive diagnostics to ensure long-term data integrity. Combined with LTFS, Scalar tape libraries provide small or departmental archives that keep content directly accessible to users with simple drag and drop functionality.

The Quantum solutions of Lattus for content repositories and archive, Scalar LTFS and Scalar Tape Libraries for archiving and data protection, and DXi-Series systems for efficient disk backup and deduplication are key products for providing increased data capacity and long-term retention. The addition of StorNext as an over-arching, high-performance storage management solution adds to the strategy for dealing with increasing capacity demand and long-term retention.

About Evaluator Group

*Evaluator Group Inc. is dedicated to helping **IT professionals** and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver **in-depth, unbiased analysis** on storage architectures, infrastructures and management for IT professionals. Since 1997 Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and **education**. **www.evaluatorgroup.com** Follow us on Twitter @evaluator_group*

This document was developed with Quantum funding. Although the document may utilize publicly available material from various vendors, including Quantum, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.