How to Optimize Data Preservation

By: Randy Kerns — Chief Technology Officer, ProStor Systems

The most basic function since computing began has been to preserve information. That requirement will continue and will become even more complicated with business rules and compliance demands as well as the sheer increase in the amount of data that must be preserved.

While data preservation is a fundamental practice, the purpose of the preservation can be divided into two primary categories: short-term data protection and long-term data preservation. The short-term data protection is primarily about disaster recovery or protection from some inappropriate action of some type. Long-term data preservation has a time element; the data must be preserved according to business rules or requirements for a long period of time and must be available when required.

With the two differing demands for data preservation, the methods of doing the preservation and the economics of providing access become a focus area and must be considered when developing the overall strategy for managing information.

Probability of Access

The most critical part of determining how to optimize the trade-offs for economical preservation and access needs for information is to understand the probability of access for that information. There are many special circumstances regarding types of data, but for the majority of information there is a very predictable pattern for the accessing of information. Newly created or introduced information will be the subject of processing activities almost immediately but will decline fairly rapidly. Inflection points regarding the access occur at specific times, the most general being 30 and 90 days.

These patterns for probability of access can be exploited for economic advantage both in the types of storage predicated on the access required and the data preservation required. The probability of access curve is shown in figure 1.

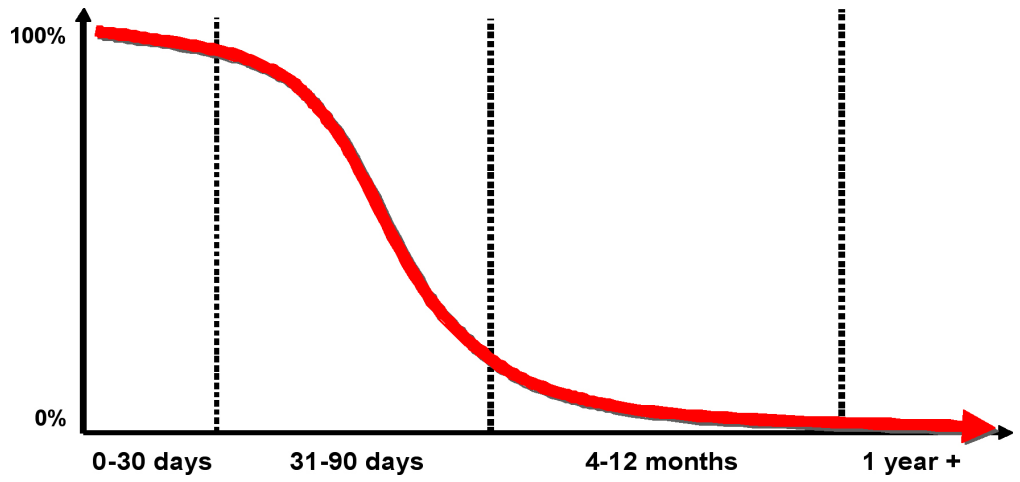

Figure 1: *Probability of Access*

Exploitation of the pattern of probability can be broken down into the location of where the data is stored (the implication being that there are different costs for the different access requirements) and the short-term data protection vs. the long-term data preservation costs.

The access of information during the stages where the information is accessed the most is usually critical for information processing and, consequently, the information is stored on primary storage. These primary storage systems meet the needs for supporting the application usage of the information in performance demands. After the access patterns have changed to where the data is unlikely to be needed with the same level of frequency or performance demand, the information can be stored on less expensive secondary storage, but still be available for online access. After the probability of access has declined further to a point where the economics do not warrant the retention of the information for online access, the information can be stored on an offline management system where it can be recalled as needed. Figure 2 shows the access characteristics needed at the key points in the probability of access curve.

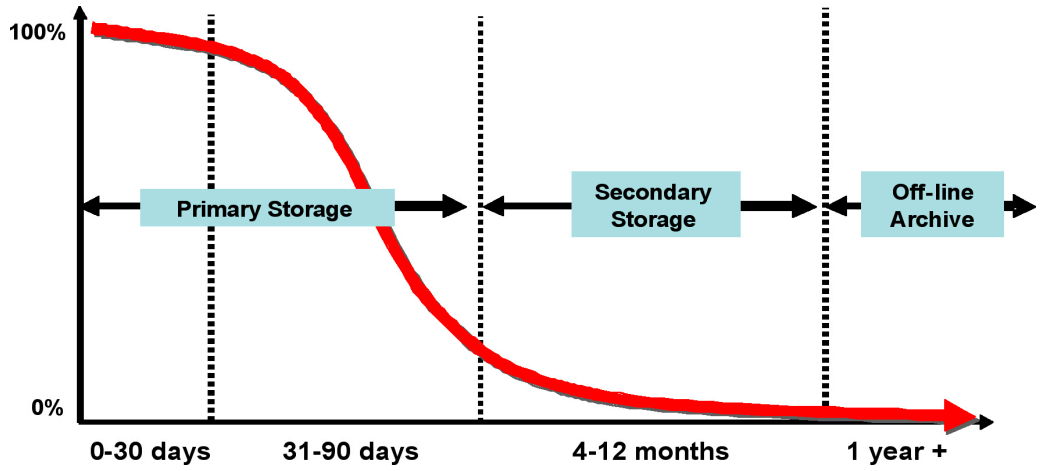

Figure 2: *Access According to Probability*

The other major economic benefit to exploit based on the probability of access is regarding the data preservation to be performed. With high probability of access there is the greater potential for some interruption of business if the data is corrupted or lost for some reason. Additionally, the potential for normal updating or changing of the information is much higher. Along with storing information on primary storage, short-term data protection must be performed. The short-term data protection or backup may have stringent requirements regarding the time to do the backup and the recovery time, which will dictate the technology to be used.

Once the access probability is further down the curve, the data preservation transitions to long-term preservation where a more economical solution that takes the data out of the traditional backup cycle can be utilized. Long-term data preservation protects an instance of the data one time and does not require additional protection unless a new instance is created. The time element and the management of retention (and deletion) become major economic advantages as well. Data typically has a life span that may be for decades and automating the retention and deletion and the rules governing the information greatly reduces the ongoing costs of preserving the information. Figure 3 shows the data preservation based on the probability of access.

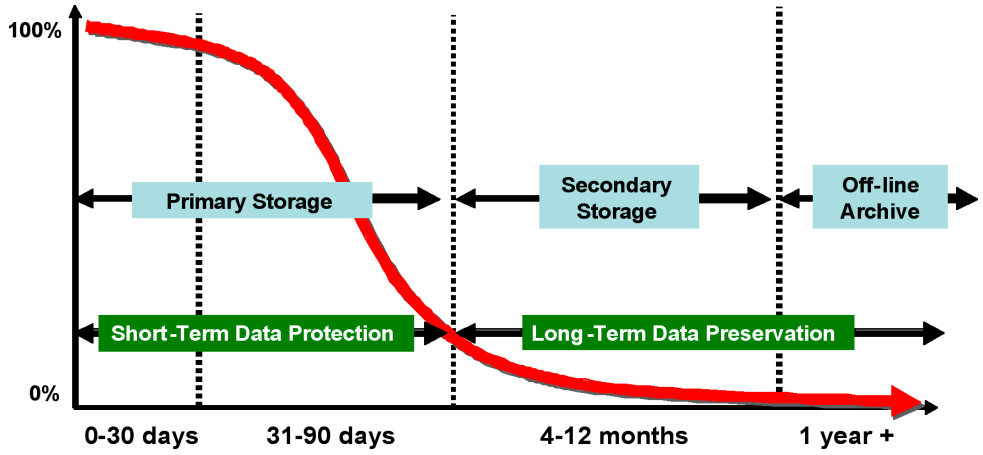

Figure 3: *Data Preservation*

The ProStor InfiniVault system provides the most economic solution to meet both the access requirements for information when the probability passes the critical primary stage, as well as the preservation of data for up to 30 years (with no forward migration required while meeting regulatory requirements). The ProStor InfiniVault provides on-line access to the information for a period of time depending on the choice of configuration and online removable disk capacity, and will manage the long-term data preservation retention and deletion automatically. The economics of using the removable disk technology with ProStor InfiniVault provide an overwhelming advantage over other technologies. Figure 4 shows the usage of the ProStor InfiniVault to exploit the probability of access.

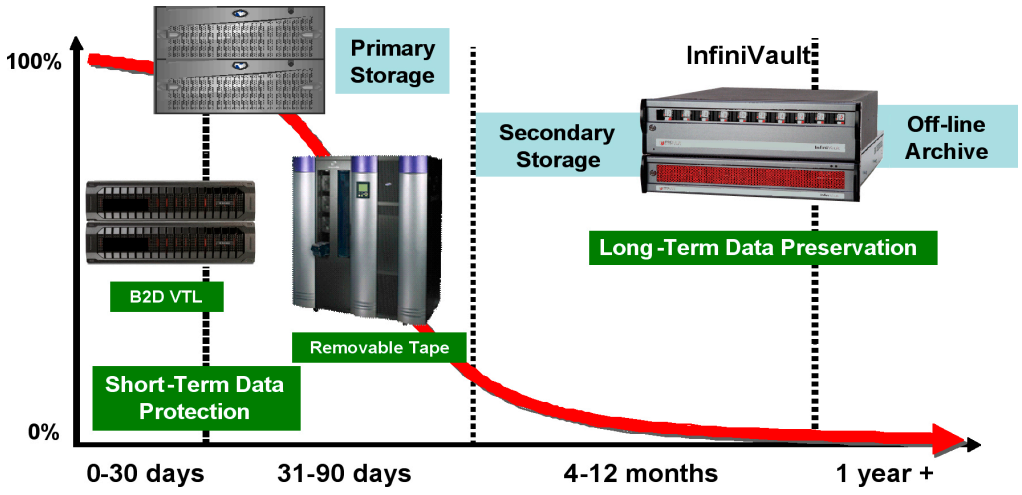

Figure 4: *ProStor InfiniVault Usage*

As a comparative, the economics of short-term data protection and long-term data preservation will show the reasons behind the economic value of InfiniVault for data preservation. The considerations for costs are in the table in Figure 5.

Long-Term Preservation Cost Considerations		
Storage Device – Lifespan		
Examples:	Fixed disk systems	5 years
	Tape drives	6 years
	Optical drives	6 years
	RDX removable disks	30 years
Removable Media Costs – highly dependent on drive lifespan and technology generation		
Maintenance and Support Costs		
Forward Migration Costs – costs to migrate to new technology/media when required		
Software Licenses		
Examples:	Backup software if used	
	Fixed disk system feature licenses	
Administration Costs		
	Power / Cooling / Floor space Requirements	

Figure 5: *TCO Considerations for Data Preservation*

A specific case comparing ProStor InfiniVault vs. tape drives vs. backup to disk VTL is available in a technical note from ProStor Systems at (www.prostorsystems.com). The outcome of that specific comparison is shown in Figure 6.

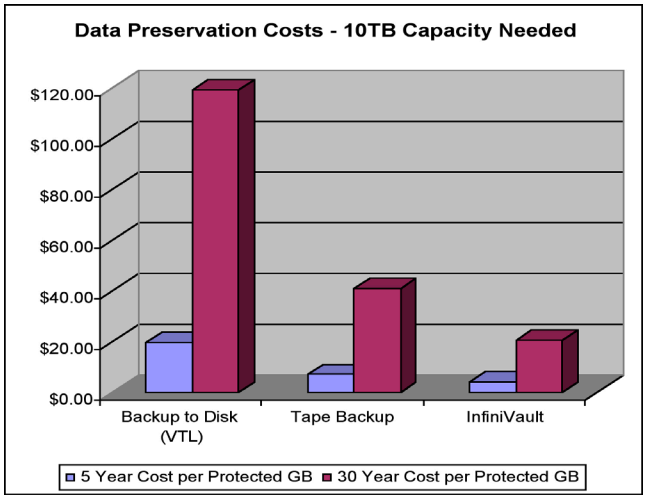

Figure 6: *Long-Term Preservation Costs*

Summary

Demands on storing information have continued to increase and new technologies to address the demands have been developed. The economic reasons are best illustrated by understanding the fundamental requirements for accessing the information and the data preservation requirements. Data has persistence: it stays on storage systems for a long time. It can be interminable if the data preservation system does not manage retention and deletion.

The ProStor InfiniVault system provides overwhelming economic advantages in being the target system for both online and offline access and for long-term preservation. It is the obvious choice when the economic analysis is performed.

June 2008

PROSTOR SYSTEMS

ProStor Systems, Inc.
5555 Central Ave., Suite 100
Boulder, Colorado 80301
Main: 303.545.2535
www.prostorsystems.com
www.rdxstorage.com

ProStor Systems Sales:
sales@prostorsystems.com
Tel: 866.560.6467