

The Importance of Removable Disk for Archiving

By: Randy Kerns — Chief Technology Officer, ProStor Systems

Archiving is often confused with backup and taking backup tapes offsite somewhere. IT and storage professionals who have re-examined archiving principles have found the advantages in economics for capital expenditures and operational expenses to provide overwhelming value. The value provided is even greater when using removable disk as the archive media and it also meets many of the business requirements that may have been overlooked previously.

Archiving is about moving data from primary storage to another media that is protected and that requires less ongoing expense for storage. When data is moved from primary storage to removable archive media, the primary storage capacity is reclaimed and the backup process is reduced by the amount of data that was moved. Backup is about making copies of data usually on disk storage systems where there is some degree of risk of the data being lost due to device failures or inadvertent or malicious changing or deleting of the data. The backups are done on a regular basis with a certain number of copies maintained in a rotation.

Effective archiving systems provide the data in the context of the application that uses the data or that moves the data to the backup device. This means that the data is usable by the required application without any intervening software encapsulation of the data. Backup software typically encapsulates the data in some format specific to the backup software and for the data to be recovered, the backup software must be used to “unwrap” and access the data. Transitions with backup software, either to another vendor or to a different generation, become a major issue when considering recovery of data that has been archived for a long period of time.

Archiving data is about preserving information that is not usually required immediately and may not ever be needed for access but must be retained. There are business rules for companies around how the data is preserved and there are regulatory compliance laws around some of the data. The compliance laws are more stringent in certain industries. All companies are subject to discovery regulations regarding access to information in case of civil or criminal proceedings and all data, regardless of where it is stored, is subject to discovery.

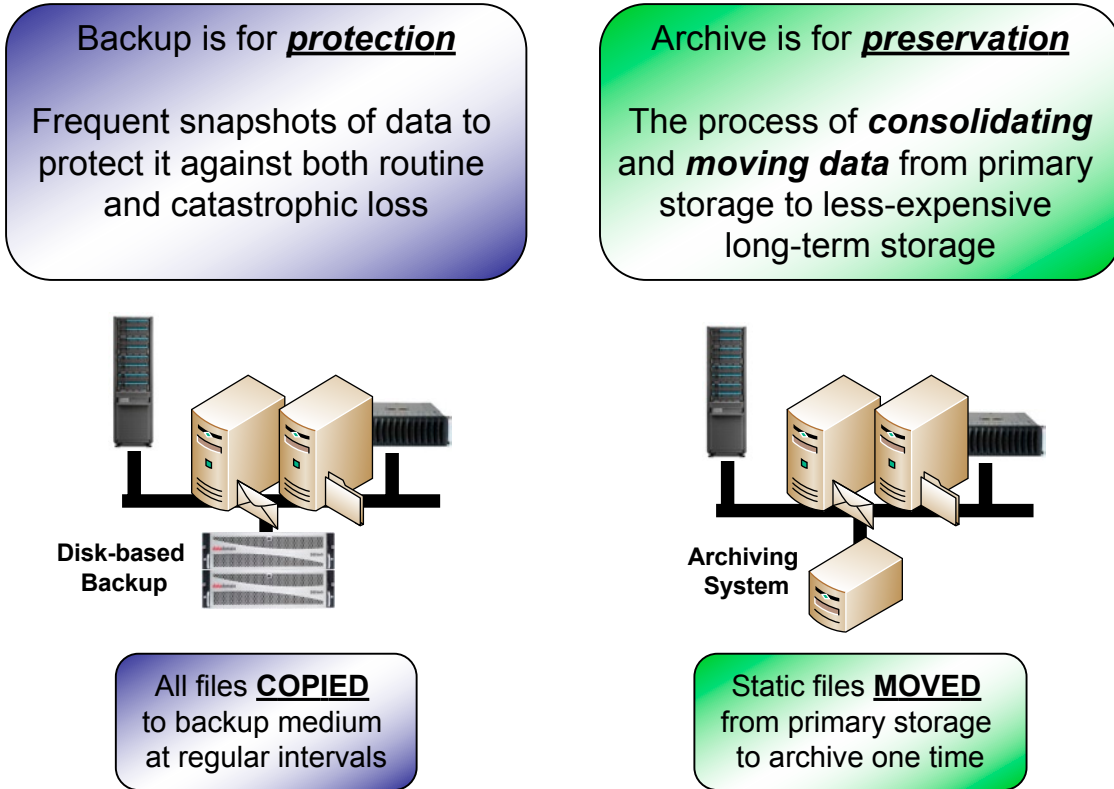


Figure 1: Backup vs. Archiving

The preservation of the data is the focus of archiving and several factors must be considered:

- *Longevity* – how long is it reasonable to store data on any particular media – can the media technology support the storage for that length of time? In addition to the media longevity, the systems that have written the data and the applications that will use the data will need to be considered for that entire period of time.
- *Costs* – the long term costs of the storage are a major concern for archiving data. As more data is archived, the costs multiply. The infrastructure required, technology transitions forward that may be required due to either the obsolescence of the storage media or drives chosen or the wear-out characteristics of devices may cause precipitous expenses at regular intervals.

- **Access** – the movement of data to an archiving system is based on a probability of access. Retrieving data that has been archived is a matter how often and over what period of time. An archiving system that provides an active archive where all data is visible to an application and can be accessed directly for a given period of time and indirectly through a recall mechanism provides flexibility for implementation. The random access nature of retrieval of archived data as opposed to restores done from backup tapes may dictate some expectation on the speed of access for archive data.

Economics for the Long Term

Archiving can have overwhelming, immediate economic benefits:

- *Reclaiming of primary capacity* that will allow new primary storage capital expenditures to be deferred.
- *Continual reduction in capacity* required as more data becomes archived over time. Overall administration of primary storage is reduced commensurate to the reduction in consumption of new capacity.
- *Reduction in the amount of backup required.* The administrative time performing the backup, the potential impacts to online systems, and the consumption of backup media are all reduced.

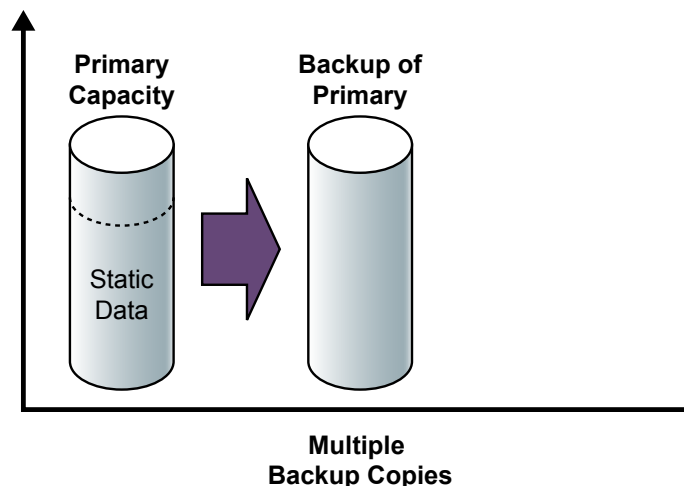


Figure 2: *Typical Data Protection using Backup*

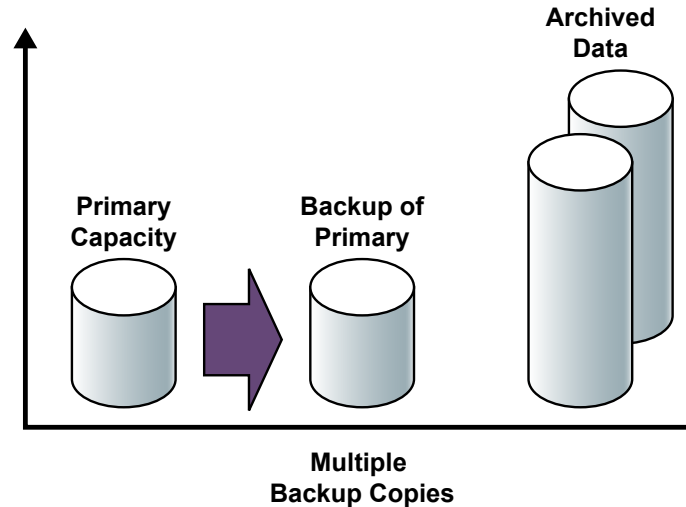


Figure 3: *Effect of archiving*

The long-term economics become very interesting when considering continual archiving and the different technologies that may be used as an archive target. The cost for long-term archiving must consider many different factors and some may be less than obvious when making determinations about what archiving technology to choose. For the case of looking at long-term storage, two checkpoints seem to be used by most businesses regarding data preservation: ten years and twenty years. For considerations beyond twenty years, the discussions regarding the availability of an application that can execute on whatever systems are available at that time and the recognition of a format that would have changed over time become larger considerations than the archiving media. Consider the popular media choices when it comes to archiving for the potential of the ten to twenty year checkpoints and their long-term costs for archiving:

Online disks – Using disk systems available today as the media for archiving require that the data be backed up for protection. The target is to use online disks as another tier of storage for archiving and to provide protection another online disk system is required at a remote location with the appropriate IT infrastructure. Online disk systems are typically replaced on a regular interval based on time installed – usually in four or five years. The cost implications when using online disk for archiving include:

- *Cost of the disk systems* – times 2 for the local and remote systems. Included in the cost of the disk systems must be the replication software which is typically an extra charge.

- *The cost of the remote link* between the two disk systems. This may be a shared fiber connect of some type but will require some type of physical connect. The additive costs for the IT infrastructure would also have to be included with the remote link and remote systems.
- *Administration of the online disk systems.* This will be a typical charge for disk storage systems that is an ongoing expense.
- *Migration costs* for when the local and remote disk systems are replaced on the regular system schedule. This will be done twice in the 10-year window and four or five times in the 20-year window.
- *Environmental costs* include the physical space occupied and the power and cooling required for the systems. Some disk systems may spin the disk down even though disk drives may be used that leave the heads in contact with the media which helps to mitigate the power and cooling charges. These systems typically have to spin the drives up and exercise them periodically to determine if there have been any effects from leaving the heads in contact with the media.

Using online disk systems also has a cascading effect on costs. As more data is archived, more disk systems will be required which multiply an already significantly high number.

Removable disk systems – Removable disk archiving systems, specifically the ProStor InfiniVault™ systems, use removable disk cartridges. In the case of the RDX® removable disk cartridge, the disk inside the cartridge is a mobile hard disk drive with ramp-loaded heads. The ramp-loaded heads in the removable cartridge park the heads safely on a ramp when the drive is powered down. The hard disk drive powers down when it is not transferring data and with the ramp-loaded heads, the heads never come into contact with the media.

The ProStor InfiniVault uses the RDX cartridges that have been tested at an independent lab to confirm a 30-year longevity for data storage. With the use of the removable RDX disk cartridges, data can be transferred to the cartridges (in addition to the active archive of the InfiniVault) and removed from primary storage based on business and regulatory requirements. The costs to be considered using the removable RDX disk cartridges include:

- *Cost of the InfiniVault archive system* – a single system is required. Included in the cost of the archive system is all software running on the system to manage the intelligent archiving.
- *The cost of the removable RDX cartridges.* These cartridges will contain the archive data and will be usable for the entire and 10- and 20-year archive time-frames. For protected archives, at least two copies will be required.
- *Environmental costs* will be nominal as the removable RDX cartridges will be powered down when not transferring data and while stored.

Removable RDX disk cartridges represent a scaling of costs directly with the amount of data stored. As more data is archived, more cartridges will be consumed. No forward migration will be required during the entire planned archive period because the drive and the media are both contained in the removable RDX cartridges.

Tape systems – Tape systems used for archiving would include a server running backup or other software with either individual drives attached or drives attached in a tape library. In either case, the tape media is the removable storage. With the use of tape systems, the tape drive is separate from the tape media. Tape media is purchased separately as needed. Tape media has a use count that necessitates replacement of the tape media but that is typically an issue for backup and not for archival. Costs to consider for using tape systems for archive include:

- *Cost of the tape systems* – this will include the drives and potentially tape libraries as well as the servers and software for performing the movement of data. Tape technologies change on a regular basis and typically a tape drive will be replaced on a five year cycle with a newer drive. The tape drive replacement will be necessitated because of the fact that the previous tape drive is no longer offered for sale as a new product or may not be supported any longer.
- *The long-term cost must include forward migration of the data* from one generation of tape technology to another. With the tape drive change, even though the data may be readable from the tape media, the data must be migrated to new media that is used for the new tape drives. For the 10-year archive period, there may be two migrations and may be four migrations for the 20-year archive requirement.
- *The tape media will need to be purchased for the data* to be archived. Because of the forward migration required, the media will need to be purchased again for each migration required. For a 20-year archive requirement, the media may need to be purchased four times. For protection, at least two copies of the media would need to be produced which would double the number of tapes to purchase.

- *Environmental costs* for tapes are limited to the storage costs of the media and the costs of the systems to write the media and any library mechanisms in use.

Tape systems meet the economic (and business and regulatory requirements) needs of removable media and may have an initial low cost for media. Subsequent forward migrations of data, the purchase of new tape drive technologies on a regular basis and purchase of new media drive up the long-term archive costs.

Optical disk systems – Similar to tape systems, optical disk systems may use individual drives or a library mechanism where data is moved to the optical device using software running on a server. As with tape, the optical drive is separate from the optical media that is used. Optical technology typically changes at a slower pace than tape. Costs to consider for using optical disk systems for archive include:

- *Cost of the optical systems* – this will include the optical disk drives and potentially optical disk libraries as well as the servers and software for performing the movement of data. The optical disk technology change may occur on a cycle of seven years or greater. The optical drive replacement will be required because of the fact that the previous generation optical drive is no longer offered for sale as a new product or is no longer supported.
- *The long-term cost must include forward migration of the data* from one generation to another. With the optical disk drive change, even though the data may have a much longer life on the optical media, the data must be migrated to new media that is used for the new optical disk drives. For the 10-year archive period, there may be one migration and there may be two migrations for the 20-year archive requirement.
- *The optical disk media will need to be purchased for the data* to be archived. Because of the forward migration required, the media will need to be purchased again for each migration required. For a 20-year archive requirement, the media may need to be purchased two additional times. For protection, two copies of the media would need to be made which would double the amount of media to purchase.
- *Environmental costs* for optical disk used for archive is limited to the storage costs of the media and the costs of the systems to write the media and any library mechanisms in use.

In many cases, optical disk systems meet the economic (and business and regulatory requirements) needs of removable media and may have an initial low cost for media. Subsequent forward migrations of data, the purchase of new optical disk drive

technologies on a regular basis and purchase of new media will drive up the long-term archive costs in the same manner as tape technology.

Cost Item	Cost	# / 10 years	# / 20 years	Total
<i>Archiving System</i> (or servers & software) Admin / op expense				
<i>Archive target cost</i> Media or Storage system (factor in growth %)				
<i>Forward migration (if required)</i> Media or Storage system Op expense to migrate				
<i>Remote site charges (if any)</i> Leased network Storage system(s) Infrastructure Admin / op. expense				

Table 1: Sample Cost Consideration Table

Protection of Data

Archiving moves data off the primary storage systems and when used with removable storage systems, out of the regular backup process. The archive data must still be protected however. The protection level may be dictated by business practices or by regulatory compliance. For removable storage systems, the protection is accomplished by making multiple copies of the data. The removable systems that are efficient such as the ProStor InfiniVault make simultaneous copies when the data is first archived. Fixed disk systems require that the data be protected by making a copy to another fixed disk system or by backing up to a tape or optical system.

Protection rules generally require that at least one copy of the archive data be stored at a remote site where it is outside of potential regional disasters and can still be utilized in the case recovery is required. For removable media, storing the media at remote locations is simple and economical. Fixed disk systems require an infrastructure around the disk systems at the remote site to be able to receive and maintain the data. When more than two copies of data are required, the costs and complexity of the fixed disk solution are increased.

The ProStor InfiniVault makes simultaneous copies as data is ingested and can utilize any combination of the different capacity RDX disk cartridges. By accumulating copies of data on cartridges in multiple slots of the InfiniVault, the maximum capacity of a disk cartridge can be utilized thereby minimizing the number of cartridges required to be used (and stored).

Summary

The overall economics of archiving data provide a compelling case for Information Technology to implement an archiving solution. Archiving is a long-term proposition however with requirements for data preservation based on business needs or regulations. Consequently the economics need to be considered over the long term. The advantages of using RDX removable disk cartridges provide an overwhelming advantage in economics over the long term as well as meeting the immutability, performance, and reliability required. The proven longevity and the unique characteristic of having both the drive mechanism and the media in the removable cartridge provide distinguishing advantages.

Compliance is a factor with some archived data and must be considered in protection and other requirements such as hardware enforced WORM and security. The RDX removable disk provides hardware enforced WORM, encryption, and digital signatures to meet or exceed compliance requirements.

The newest archiving technology, RDX removable disk cartridges, has both economic and technical advantages over those in prior use. For archiving, using removable disk technology will be strategically important for companies.