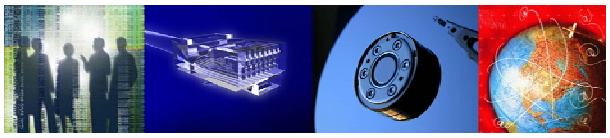
Enterprise Strategy Group

IT Knowledge • Business Results



White Paper

Lowering the Cost of Information Retention with Effective Disk-based Archives

ProStor's InfiniVault Helps Simplify Archiving with Portable, Removable Disk Systems

By Brian Babineau
Senior Analyst
Enterprise Strategy Group
Intelligent Information Management

October, 2007

Table of Contents

Introduction

Digital information is growing at an explosive rate. Collaboration, supply-chains, electronic commerce and several other common business activities are catalysts for this continuous deluge of information. This phenomenon carries large direct costs, which include the purchase of systems that store the data, as well as the software and personnel that manage it all. Organizations must find additional budget resources to protect all of their information.

Further complicating matters, various regulations and corporate governance policies require that certain information is to be kept for specific periods of time. As such, organizations need to be able to identify what data constitutes a business record and ensure that it is kept for the appropriate duration.

Another more recent challenge that all organizations—regardless of size or industry—must deal with is electronic discovery. This process starts with a legal discovery request and requires organizations to identify, preserve, review and produce electronic information. In the electronic discovery process, the preservation step mandates that organizations retain certain electronic information and prevent it from being deleted or modified until a matter is resolved. Given the tempered pace of the legal system, preservation notices can last for several years.

For large organizations, the information onslaught and related compliance considerations may be much more considerable. However, enterprise IT departments typically have access to capital budgets and operational staff, making the problem manageable. For Small to Midsize Enterprises (SMEs), the same information growth problems exist, but their limited IT budgets and lack of operational resources make it extremely challenging to handle all of the data and address information retention requirements.

In order to address information retention requirements, some enterprises and most SMEs simply copy data onto backup tapes and store them for the appropriate period of time. This process is ineffective because an organization must buy more primary storage as information grows, procure a similar amount of tape capacity and then separate the backup process into two data types: regular corporate information and those business records that are subject to regulatory or legal retention requirements. Additionally, when a piece of information is requested as part of a discovery or inquiry, the data is not easily accessible as the tapes must be restored, indexed and searched before anyone can find what they are looking for.

Making multiple copies of information during the backup process can be expensive, and retrieving them later can be extremely time-consuming. Both processes can be resource-intensive from a capital and operational budget standpoint, which is something that affects SMEs more than other organizations. SMEs must understand the changing regulatory and legal landscape and take advantage of emerging technology solutions that facilitate information retention on more accessible media. The ultimate benefit, of course, would be achieving compliance with regulatory and legal mandates in a much simpler manner—without breaking the bank.

Compliance: Changing the Rules of Information Retention

Record Retention Regulations Apply to Organizations of All Sizes

There are several laws mandating that organizations store information for specified periods of time. Each law typically designates what information must be kept and for how long. Retention periods range widely: from one year to forever, depending on the type of data and the regulation that applies to it. For example, HIPAA requires healthcare entities to retain patient records until the patient dies. HIPAA

applies to large hospitals and insurance providers as well as local doctor's offices. Some regulations—such as the Securities and Exchange Act of 1934—are a bit dated, but recent interpretations require organizations to apply these rules to electronic information including e-mails, database records and general files. Specifically, SEC Rule 17a-4(f) allows for the use of electronic storage systems on the condition that these devices prevent overwriting, erasing or altering a record during its required retention period.

Proper record retention programs should incorporate the appropriate storage and expiration of information. This can help organizations reclaim capacity and use it for more recently created business records. There are some rules that stipulate the manner in which data is destroyed when it exceeds its required retention period or is at the end of its useful life. Most notably, the Department of Defense 5015.2 Standard, which provides guidelines on how United States defense agencies manage certain information, mandates that once a record is destroyed, there should be no way in which to physically reconstruct it. Though this is a DOD standard, many non-government entities use these rules as guidelines in their own environments.

Increased Focus on Corporate Governance

In the early part of this decade, executive malfeasance, insider trading and other undesirable activities severely impacted many businesses. As such, senior leaders and boards of both publicly-traded and privately-held companies have reinstituted corporate governance policies and processes. Many of these rules require organizations to keep certain business records—such as stock option plan documentation—for specified periods of time.

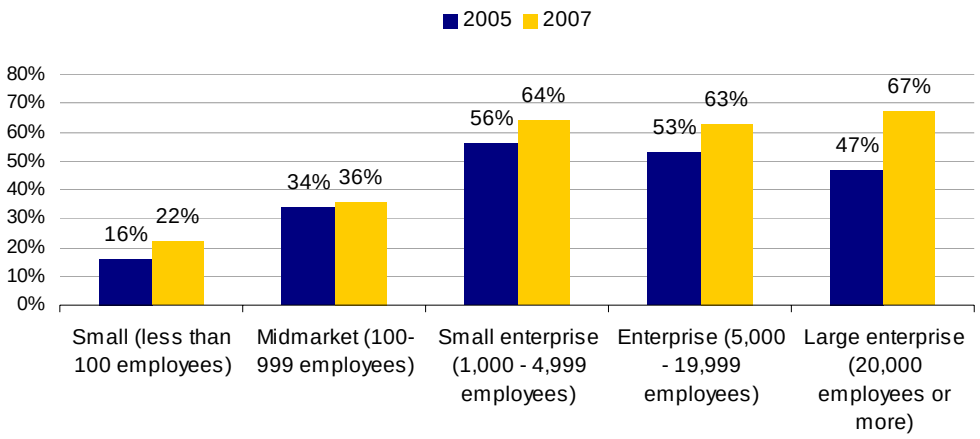
General records management programs—which typically covered paper documents—now encompass human resource files, executive e-mail correspondence and other digital information. While there is no legal requirement that forces organizations to retain this information, proper corporate governance—which ensures that organizations are doing the right things for the right reasons—incorporates electronic records management policies.

Electronic Discovery is Pervasive

As more and more business is conducted electronically, regulators and litigators must seek out historical e-mails, database records and application files to support cases ranging from insider trading to employee discrimination. As shown in Figure One, all organizations are subject to electronic discovery—regardless of size. Because of this, a large number of companies have had to preserve information due to a legal hold. The information must be retained and remain unaltered during the legal hold process until a case is ultimately resolved.

Recent changes to the Federal Rules of Civil Procedure (FRCP)—which govern the process of civil litigation in the United States—emphasize the ongoing trend of increased electronic discovery events and set forth requirements regarding how organizations must manage electronically stored information. Companies must be prepared to disclose sources and accessibility of electronically stored information. For SMEs that do not have a large IT or legal staff but are subject to electronic discovery requests, addressing the new FRCP amendments is a significant, potentially costly challenge.

Figure One: Electronic Data Discovery Trends, By Organization Size



Source: ESG Research, 2007

Security

The line between information storage and security grows blurrier all the time, especially as more and more sensitive and confidential information is being kept for legal or regulatory reasons. Additionally, many record retention regulations also include some measure of information security relative to private data, ensuring that only authorized users have access. Section 164.306 of HIPAA's Security Rules outlines security requirements and even specifies encryption as a viable means of protecting confidential records.

Organizations must find ways to securely retain information due to information privacy regulations that abound. Failing to keep and properly protect information can lead to fines and unmanageable legal risk.

Separating Archiving From Backup

If an organization desires to improve its information retention practices, it must first delineate its backup and archival processes. Backups are full copies of information that are used in the event of data corruption or deletion. The process involves creating batch files and sending the information to secondary media, most commonly tape.

Archiving should be a separate process. Non-transactional information–including business records, potential electronic evidence and other static data–is moved from a primary storage system to a secondary device. Archived information should remain in its native file format and be indexed as it is moved from one system to another, facilitating easy identification and search. During the archiving process, a retention period can be established as the information is moved. This allows organizations to comply with record retention, corporate governance and legal preservation requirements. An additional benefit of archiving is the recovery of expensive primary storage and a subsequent reduction in the amount of data that must be backed up.

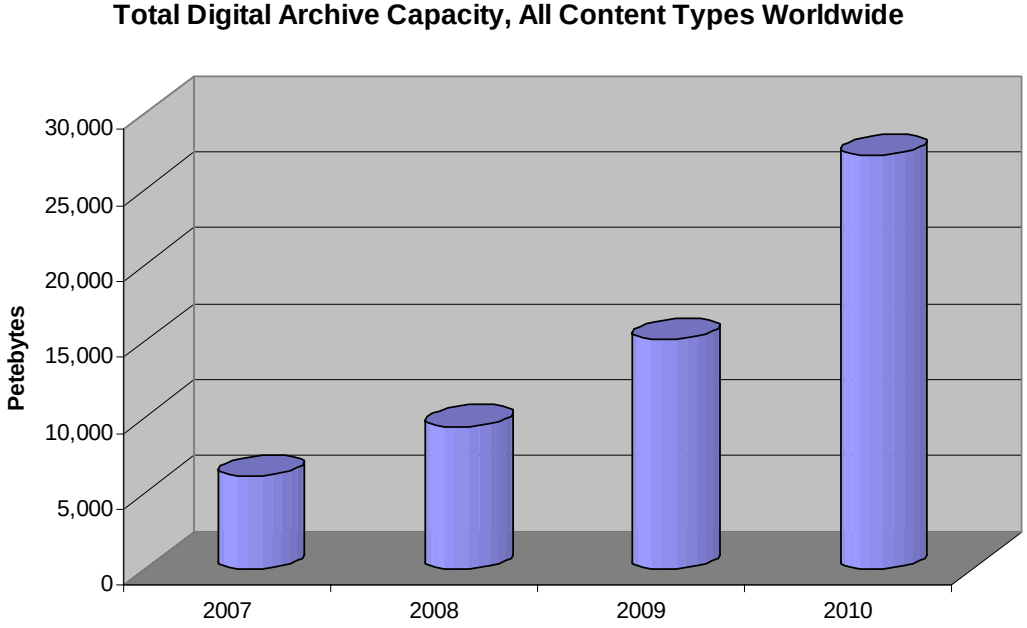
To make archives more accessible and information readily retrievable, the secondary storage media should be disk as these devices are optimized for random access of data. The secondary system should

also have some form of built-in capacity optimization capability, most likely de-duplication, to ensure that only one instance of data is retained.

Archiving and backup processes can intersect at some discrete points. For example, information should be archived before it is backed up so that there is less to copy, which allows data protection operations to be completed faster. This also reduces the amount of backup storage media capacity an organization must buy and manage. Corporate archives should also be backed up to ensure that business records and other non-transactional information are protected. However, these backups can occur less frequently than those on primary systems since the information is not changing on a regular basis.

The compliance benefits—combined with expedited and more efficient backups—have led organizations to embrace archiving. ESG expects archiving adoption rates to increase as the confluence of data growth and compliance lead to more information being retained for longer periods of time. According to ESG research, as shown in Figure Two[1], users expect archived data to grow from 5,000 petabytes to nearly 30,000 petabytes by 2010. This represents over a five fold growth in just three years time, and confirms that digital archiving will be one of the most significant priorities of organizations going forward.

Figure Two. Total Worldwide Digital Archive Capacity, 2007-2010



Leveraging ProStor's InfiniVault for Information Archiving

Disk-based Removable Storage Media is Critical for Scalable Archives

There are a number of reasons that tape has been the preferred media for long-term data storage since computing took off in the mid 20th century. The argument could really begin and end with the fact that tape media is a portable and low cost solution that can help organizations get data offsite for disaster

[1] ESG Research Report: *Digital Archiving: End-User Survey & Market Forecast 2006-2010*, January, 2006

recovery purposes. Additionally, tape systems can scale as more media is added to a system after full devices are removed.

When attempting to comply with retention and preservation requirements, tape was a logical choice because it offered a central place for all information. However, as already discussed, archive information needs to be readily accessible and trying to locate a few business records amongst hundreds or thousands of tapes is not feasible.

Disk systems are orders of magnitude faster than tape. Disk drive random access performance is also far superior to that of tape drives. Disk systems also allow storage in native file format—a luxury that tape does not afford. However, for SMEs, many disk systems are hard to manage and difficult to scale, as adding capacity often means adding a whole new system, which compounds the operational burden.

ProStor's InfiniVault provides the best of both worlds in terms of data protection. ProStor's InfiniVault combines removable disk media and policy based retention management—enabling scalable, easy to manage corporate archiving.

Inside ProStor's InfiniVault

ProStor's InfiniVault combines the best of both worlds: The availability and performance of disk, with the mobility and cost of tape. InfiniVault is comprised of an intelligent appliance that connects to the data network via the CIFS or NFS protocols. Appearing as a drive letter on corporate LANS, users can access information based on permission level. The appliance stores a copy of all archived information on permanent disk, known as the Active Archive, while also storing one to four copies on removable media for long term retention. As the Active Archive fills, older files are stubbed, but still remain visible to the user—improving information access. The system controller serves as the archiving management software, configures the archiving policies and assists with the establishment of retention periods.

The storage elements that hook into the InfiniVault system controller and enable both active and off-line archiving are permanent RAID disks (Active Archive) and the ProStor Removable Disk Array (RDA) with RDX removable disk cartridges. The fixed RAID disks can be integrated with the system controller or provided by an external disk array. Regardless of the implementation, the purpose of this layer is to present an active online archive. This offers visibility to all files in the InfiniVault system, as well as providing immediate access to more recent files that are more likely to require instant retrieval.

Where ProStor completely differentiates itself, however, is with the RDA and accompanying RDX removable disk media. There are ten removable RDX cartridges per RDA, which provides portability and, more significantly, infinite scalability. Even with one or two RDA units, users would never have to buy additional systems that would just consume more real estate and suck more power. When an RDX cartridge reaches capacity, just remove it, ship it to the off-site archive location of choice and replace it with a new RDX cartridge. Lather, rinse, repeat. Another key consideration is the fact that RDX cartridges are also forward and backward compatible, eliminating the periodic forward migration issues faced by other removable technologies.

In addition to the unique functionality that the RDX removable cartridges bring to the table, the RDA—coupled with the InfiniVault software—has some appealing features. For starters, it provides automatic data protection (up to four copies), which is one less thing that system administrators have to contend with when bringing the system up. The RDA units were designed to draw minimum power, and subsequently generate less heat. The RDX cartridges also park and power down whenever they are inactive. Power and cooling are big issues, and for SMEs that may not have access to large power grids, investing in energy efficient IT solutions is a must.

What InfiniVault really provides is a tiered archiving infrastructure. The RAID disks provide instant online access to information that isn't transactional in nature, but still important. A good example is a patient's records, which might require instant access in a life or death situation. At the point that a record becomes static (it is no longer regularly needed), it is migrated to an RDX cartridge in the RDA for nearline storage. Once the RDX is full, it can be removed and shipped off-site. Sticking with the medical theme, a three year old X-ray of a broken arm would be a candidate for this tier of archiving.

InfiniVault is an integrated, easy to set up system that does not require purpose-built archiving software and specialized storage systems. However, the InfiniVault can be used as a storage target by a variety of archiving applications such as e-mail, HSM, ECM and PACS. It provides users instant access to information, while offering removable, off-site deep archiving capabilities from the same appliance. This is a key consideration for SME customers who lack the budget to buy separate solutions, as well as the manpower to manage them. SMEs can archive before they run their backups, allowing for the reclamation of primary storage and reduction in tape media utilization, which equates to greater savings.

Compliance Simplified

For data that cannot be altered or deleted, customers can use the WORM-based RDX disk cartridges, which can be particularly useful to users that need to abide by SEC Rule 17a-4. In addition, WORM can preserve potential electronic evidence that is on legal hold during a specific matter, thus ensuring immutability of data. InfiniVault also generates an audit trail, which helps prove authenticity and integrity of information, whether the audience is internal or not.

Information is also secured through AES-256 strength encryption—with the accompanying key management—which should appeal to companies that must meet HIPAA and other information security and privacy requirements. A key management system also provides users with an option for data destruction when information is no longer needed. However, for those not comfortable doing so, RDX cartridges enable data to be shredded according to the DOD 5015.2 Standard.

Increasing Access without Increasing Costs

InfiniVault builds an index of archived information, enabling search and retrieval. For information that is no longer available on local storage, the removable disk drive simply needs to be reloaded and read. Information access is critical if organizations need to respond to regulatory inquiries or simply need to reference information during the normal course of business. With tape, this is very difficult to do.

To reduce costs, the system performs single instance storage at the file level by creating a unique content address for every file it sees. If the same address is generated, the system knows that it has already stored the data and creates a reference pointer. This helps optimize archive storage capacity, which in turn improves disk utilization. By reclaiming existing capacity, organizations no longer have to bring additional systems online—systems that will consume not only valuable data center real estate, but also power and cooling resources.

Conclusion

Storing and protecting information always presented a big enough quagmire for most organizations, regardless of size. Now, these challenges have been multiplied by compliance mandates that include records retention regulations, corporate governance policies and electronic discovery. Factor in the staggering growth rate of digital data, and many organizations find themselves hard-pressed to identify

simpler, more cost-effective ways to retain information. Nowhere is this truer than in SME organizations, where these problems are exacerbated by a lack of resources.

Current processes, which typically involve saving backups for longer periods of time, do not scale and are inefficient. The first thing that IT staffs must do is separate the backup and archive processes. A majority of the information that is being backed up on a nightly basis has not changed in more than 90 days, which means there are terabytes of duplicate data throughout corporate computing networks. Implementing archiving will allow for reclamation of a lot of that capacity, which will improve utilization of existing hardware assets while negating the need for additional storage resources.

Customers need easy to implement archive solutions and ProStor's InfiniVault can provide that. It is inexpensive to acquire and manage and easy to implement, so it enables capital expenditure savings and reduces the burden on operations resources—which are especially crucial to SMEs. It is also more cost-effective thanks to de-duplication capabilities and decreased power consumption.

In addition, the removable RDX drives provide unlimited scalability and more specifically, an integrated archiving solution with features that can address compliance requirements such as data immutability, accessibility and security. IT environments that are looking to consolidate and manage their growing volumes of archived data—while simultaneously satisfying the slew of compliance and legal regulations with little or no expertise and all for a low cost—can benefit from evaluating the ProStor InfiniVault solution.