Today’s archiving solutions help rein in expanding data storage costs while satisfying the growing demands of compliance and e-discovery.
Executive Summary

The shift to digital content and communications means organizations of all sizes are creating and saving more data than ever before. The increasing use of e-mail and messaging technologies, social networking, multimedia content, web pages and PDF files means the explosive growth of data generation – particularly unstructured data – will not merely continue, but multiply dramatically.

In addition to dealing with this ever-growing mass of data, businesses are also facing the challenge of storing data for longer periods of time. This is needed in order to satisfy increasingly stringent requirements for compliance and e-discovery.

What’s more, no longer does the strategy of simply adding more storage devices to take on more data make sense. In addition to being prohibitively expensive, it is inefficient and does not address the growing need to manage data so that it can be identified when needed, and accessed and delivered in a timely, efficient and cost-effective manner.

Fortunately, data archiving is emerging as the go-to technology for efficiently managing growing data stores. With today’s scalable archiving solutions, businesses can save money by managing their storage infrastructure more efficiently, satisfy both compliance and e-discovery demands, and treat data as the strategic corporate asset that it is.

Managing Data Strategically

Not long ago, many organizations solved the challenges of data growth by simply purchasing additional storage devices. Today, this is no longer practical, efficient or strategic.

Even though the storage industry has a long history of reducing the costs of storing data, according to industry experts, most businesses still spend as much as 15 to 20 percent of their IT budgets on storage. And without a plan in place to address storage growth, those numbers will surely go up.

When businesses add storage, they generally purchase high-performance units to handle mission-critical applications, but this approach winds up being costly and leads to inefficient usage of storage devices. In addition, simply installing more devices adds to energy costs and takes up more physical space in the data center at a time when IT departments are under a clear mandate to reduce costs, reduce space requirements and produce better results without necessarily spending more money.

What’s more, storing data that does not need to be stored is a waste and an expense. And storing data on high-performance devices, when it can and should be archived, is even more wasteful and expensive.

For organizations of all sizes, data archiving is emerging as one of the key solutions to overwhelming data storage challenges. With archiving, organizations can automatically manage and move data that has lost value with age so that it is not only stored within less-expensive storage units but...
is still readily available if needed. In addition to addressing some of the serious cost challenges in managing data growth, archiving helps enterprises deal with other IT challenges as well, including:

- Improving overall performance and making it easier to back up applications that tend to get overloaded with unstructured data, such as Microsoft Exchange
- Enhancing the ability of the business to respond to e-discovery or compliance requests by making it simpler to identify, find and access critical data — even if this data is later in its lifecycle
- Easing the migration of data as newer versions of applications are added. This saves IT time, expense, maintenance costs and possible complications in rolling out new apps across the enterprise
- Maximizing the use of the data storage infrastructure by eliminating data that does not need to be saved and by storing data on devices that are more appropriate to the level of performance required to access and recapture that data
- Building a scalable storage infrastructure that enables companies to save, store, create and access additional amounts of data in this increasingly digital world

What Is Archiving?

Archiving provides a store of data in a central storage device. It also frees up disk space across the network by eliminating duplicate files and allows quick retrieval of data at any point in its lifecycle. This type of device on which data is stored is based on whether the data is mission-critical or dated and less strategic.

Archiving also enables the enterprise to provide a level of structure to growing volumes of unstructured data. It does this by providing a systematic and automated approach to storing, managing and searching for files, e-mails, instant messages and other applications. In addition, archiving allows the business to define data retention policies for various applications and users, ensuring proper storage of data and its deletion when no longer needed.

Archiving also enables users to perform advanced searches, which means e-mails or instant messages can be searched and flagged for legal review during an e-discovery process. And it provides organizations with a foundation and framework for rules and policies to prevent data from being deleted or altered, which is critical to processes such as e-discovery and compliance.

Why Archiving?

IT professionals in businesses of all sizes recognize that the growth of unstructured data is not only going to continue, but is likely to increase dramatically. The expansion of multimedia, messaging, rich media and other data-intensive applications ensures that the challenges created by data growth will have to be addressed and, in most cases, addressed immediately and with a long-term solution.

From a practical standpoint, IT professionals generally require the following from their storage infrastructure:

- Dedicate the highest performance storage to applications that are most closely associated with company profitability.
- Maximize storage investment by not saving duplicate data and also moving data that is older or deemed less critical to less expensive storage devices.
- Utilize an automated system for identifying and tagging data so that IT staff do not have to manually sort through potentially millions of files, e-mails, texts, etc., extract the data, manually fetch what is valuable and deliver it to the user who has requested it.

Because of quickly expanding data storage requirements, most organizations eventually experience a pain point

---

**Data Backup vs. Data Archiving: What is the Difference?**

Research suggests that data archiving is sometimes confused with data backup. This often stems from assuming that if needed, data backup could be a substitute for data archiving. Keep in mind, there are different objectives for these two data storage technologies.

Data backups are typically used over the short run to restore data that may be lost, corrupted or destroyed due to a number of reasons including simple disk failure or system crash. In addition, data backups often copy data to a sequential access recovery medium and multiple copies of the data are typically protected.

On the other hand, archiving is considered the systematic approach to providing structure to unstructured data. It enables the storing, managing, retrieving and eventual discarding of data throughout its lifecycle. Archiving can also be considered the foundation for e-discovery and a way to help meet compliance regulations.

While the distinction seems clear, be aware that today manufacturers are integrating more functionality into backup software. For example, data deduplication and data lifecycle management with storage tiering — typically associated with archiving — are now part of some backup software packages.
where they know they must do something to get things under control. For example, a business has upgraded storage devices several times in a recent period just to keep up with data growth.

Suddenly, IT chiefs realize that even more storage must be added for a number of reasons. For example:

- Certain applications, such as Microsoft Exchange, become so large that e-mail is difficult to back up
- Data storage has become so unwieldy that users can’t get to valuable data and administrators are unable to complete backups
- Corporate attorneys are unable to locate key files
- Organizations are constantly searching large file systems and struggling to respond in a timely manner to legal or regulatory requirements
- Much of the data is static or typically used for reference material and does not get updated

For the company that is experiencing or has experienced any or all of these issues, it is clearly time to start investigating an archiving solution. Ideally, this is one that offers a number of capabilities including reducing costs, ensuring compliance, eliminating storage silos and increasing content-reuse capabilities.

Getting Started

Choosing an archive solution that offers a secure yet accessible environment is important. When determining which archiving solution is best for a particular business, start with a thorough understanding of the firm’s data storage and management infrastructure.

What are the pain points? Is IT management continually adding new storage devices, failing to meet service level agreements (SLAs) or having trouble locating important data?

Also, look into whether there is a storage strategy in place that will address e–discovery or compliance requirements quickly and efficiently. Does the storage process allow for the identification of data location? Are policies in place to label and name data? Are there procedures to manage data as it gets older and becomes less critical to the business?

As is true with any significant IT initiative, it is important to do a full evaluation and assessment. Look at the entire storage infrastructure and assess what is available and what is needed.

Five Top Benefits of Archiving

IT professionals attribute many benefits to data archiving. Some of the most highly touted include the following:

1. **SAVING MONEY.** In most cases, when companies need more storage, they buy more drives, typically high-performance devices for production environments. Archiving allows moving older disks to second– or third–tier storage, thus extending their lifecycle. Another option is to simply purchase less–expensive devices for lower–tier storage.

2. **REDUCING STORAGE.** With archiving, single–instance storage enables data to be deduplicated, meaning the firm will have less data to store and thus reduced storage requirements. That means less data to back up as well.

3. **COMPLIANCE.** Archiving offers the ability to easily search for relevant data. Furthermore, write–once and time–stamp technology offer much more insight into what data is available and where it is located. Data can also automatically be deleted once it has hit its retention life.

4. **OPERATIONAL EFFICIENCY.** There are huge efficiencies to be gained by automating processes and not having IT professionals doing them manually. This lets IT focus on more strategic initiatives, helping the business to save money and helping IT to “move at the speed of business.”

5. **SMALLER PRODUCTION FOOTPRINT.** Archiving data from production stores creates a smaller production footprint resulting in faster backups and less “crowding” of production spindles, which postpones production spindle purchases and saves energy costs.

Keep in mind, there are reasons to be archiving data — operations, financial, accounting, compliance, legal, etc. So even though some data is not considered highly valuable, that doesn’t mean it has little or no value. Thinking that it is acceptable to lose or misplace such data is likely a misguided assumption.

At this point, it is critical to understand what policies are in place for labeling and managing data and begin thinking about how to create new policies that can be easily implemented and enforced. This assessment is vital to making sure the right solution is chosen.

It can also be a good idea to bring in a knowledgeable third–party partner, in order to help assess the archiving option without being tied to one specific vendor or solutions provider. Questions to answer include:
• How often is data accessed?
• Where does data reside and what costs are associated with that storage?
• How many duplicate files are stored?
• How much of the data is searchable?
• How tamper-resistant are the files?

It is important to recognize where the business is vulnerable and put in place an archiving strategy that addresses both existing and potential pain points.

Archiving Solutions

An effective archiving solution is typically comprised of a mix of hardware and software. It will likely include:

• ARCHIVING SOFTWARE. Archiving software automates the movement of data through various tiers of storage during the data’s lifecycle. It enables the firm to establish the criteria for determining where and when the data moves from one tier to another. It could be based on the age of the data, its content or other criteria that can be established by policy.

Archiving software can move the data automatically between tiers. And it can delete files at the end of their retention period. It also provides single-instance storage of each file, freeing up disk space and processing power for backup and storage systems.

• E-DISCOVERY SOFTWARE. Using archiving software as a base, e-discovery software employs advanced search features and systematic archiving inventories, enabling users and administrators to quickly search all files, e-mails, texts and other messages that are related to a specific topic in response to legal inquiries.

Data retention policies can be established and enforced through archiving and e-discovery software by making sure that all files that may have legal significance are flagged to prevent them from being altered or deleted, based on whatever policies are put in place.

• ARCHIVING STORAGE DEVICES. Archiving storage devices move data off production servers and onto lesser-performance devices — from a Fibre-Channel drive to a Serial Advanced Technology Attachment (SATA) drive, for example — which simplifies administrative management. Performance is also enhanced if high-performance drives are handling fewer I/O requests. Because they are less clogged with unimportant data, their overall performance will increase.

Law Firm Moves from Tape to Disk for Archiving

A large, global law firm had all of its older data on tapes. When searching for an e-mail in reference to a case, the IT department would literally take all of the tapes into a conference room and load the data onto a disk appliance. They would put the appliance on a server and the lawyers would manually go through the data – at times running into several terabytes — to search for the specific files they needed.

When this firm moved to an archiving system, it was able to take the data online and set policies, keywords, phrases, dates and other parameters. This allowed lawyers to more easily find and access the files they were looking for with a few simple clicks on their computers.

At the same time, the performance of other drives will improve if they are not being used for daily, mission-critical functions. Having data stored on the right device can help boost performance overall, even if each drive is being utilized to a greater capacity. Archiving storage devices can be as simple as a dedicated server, network attached storage (NAS) or, for enterprise needs, a storage area network (SAN).

Because there is a wide range of options available, it is often recommended to look at archiving requirements on an application-by-application basis. The advantage of doing an analysis by each is that policies for automated data movement can be established for each app based on the type of data it generates and stores.

An IT department, working with business managers and endusers, has the opportunity to set policies and parameters for moving data from a production tier to a secondary or even tertiary tier, based on how and when it is appropriate for that data to be moved. Once again, a third-party partner and solutions provider who understands the opportunities and challenges presented by each can be valuable.

Archiving Benefits

A successful archiving solution can simplify data backup, which, for many companies, still means moving data to a tape device that is either managed by the company itself or by a third-party tape service provider. This type of backup becomes less efficient as more data is being stored and companies are becoming increasingly concerned with issues such as business continuity and disaster recovery.
In today’s environment, more IT professionals are moving toward disk-based backup. Archiving should be considered a major part of this solution, along with data deduplication and continuous data protection. By incorporating an archiving system, many organizations are successfully reducing the amount of data that even enters the backup cycle.

Archiving lends itself to more of a strategic approach for data storage by not only storing the data on disks, but in making the data readily available for users to access. This sets a strong foundation for e-discovery and data compliance.

Keep in mind, while disk-based backup offers an efficient and secure process, for years tape has been the de facto standard for data backup and recovery. Being portable and slightly more economical than disk storage, it can serve as an effective medium for long term, offsite backup. While tape is no longer the one-size-fits-all answer for every backup situation, it still has a place in keeping data safe. Nevertheless, if sending data offsite, even if it is backup data, the ability to automatically delete it or call it up when needed in an emergency may not be immediately available.

**Automated Tiered Storage**

One of the keys to deploying any effective archiving solution is automated or intelligent tiered storage. With automated tiered storage, policies are put into place to identify all data stores according to attributes and parameters determined by the business — usually a combination of users, business decision-makers and the IT department. These attributes can include the age of the data, when it was created, its format (an e-mail, Word document or an audio file, for example) and its size. In fact, most archiving systems will enable a wide variety of labeling criteria.

This sounds like a more difficult process than it really is. In a survey by Storage magazine, more than 40 percent of storage professionals indicated that the biggest pain point related to their tiered storage system is classifying data so that it is sent to the right tier. Another 18 percent said their biggest challenge was keeping track of data where data currently resides. Of the survey respondents, 36 percent said they still manually move data from tier to tier. (See chart below.)

**Hurdles to Archiving ROI**

The business case for archiving is clear: archiving less-frequently used files reduces consumption of primary storage space. Inactive data is offloaded from primary storage to improve performance and reduce backup and recovery times. What’s more, automation simplifies management and applies consistent retention rules. This in turn can reduce discovery cost.

While effective archiving offers many benefits, this is not to say that it comes without issues. For example, the sheer amount of data to be dealt with far outstrips any other application in today’s enterprise.

This means archiving systems must offer a high level of scalability and be flexible enough to accommodate data growing at exponential rates. According to industry insiders, intelligent archival presents four major challenges for IT organizations. These include:

1. **Simply getting a handle on and assessing the volume and relevance of data.** Outdated information protection policies that retain multiple (daily, weekly and monthly) copies of documents for recovery purposes steal primary storage capacity. Policy-based archiving migrates data from primary to tertiary (archive), and reduces the incidence of redundant backup copies.

2. **Properly identifying data governance (or the lack thereof).** Many organizations are unclear about proper data governance procedures and err on the side of caution by saving everything on premium storage real estate.

3. **Care and feeding of different data.** Different islands of information in Microsoft Exchange servers, BlackBerry Enterprise servers, SharePoint and other unstructured files require different types of archiving — depending on the industry’s compliance and regulatory environment.

4. **The threat of litigation.** Companies in litigious industries contend with the relentless burden of legal discovery — the process of identifying, collecting, reviewing, analyzing and producing information for legal actions. E-discovery is the part of the discovery process that applies to information stored electronically, including e-mail, instant messages (IMs), documents, financials and a variety of other data.

**What’s your biggest pain point related to tiered storage?**

- **42%** Classifying data so that it’s sent to the right tier
- **24%** Moving data between tiers
- **18%** Keeping track of where data currently resides
- **8%** Poor performance on lower tiers

*Source: Storage magazine*
In reality, however, with an effective archiving solution, all of the policies can be established up front and data in the system can be identified by specific attributes — such as the data or the application — and replaced with a pointer file. This helps in identifying files that need to be deduplicated as well.

When looking at archiving solutions, consider flexible filtering classification and retention options, among other features. Also consider solutions that support both user-initiated and automated classification.

One of the other advantages of classifying data for an automated tiered storage solution is that the process can be highly flexible. So if the business decides for any reason that it needs to change the criteria for classifying data, it can generally do so on the fly and without a lot of impact on the IT department. Once the data is identified, tagged and policy parameters are in place, the IT department doesn’t have to manage the data as it moves from tier to tier in the storage hierarchy.

### Archiving in the Cloud

A fickle economy has data storage administrators looking for the most expedient and cost-effective approach to data storage. This includes considering archiving data stores in the cloud as a budget-saving alternative to traditional in-house archiving processes.

The global market intelligence firm International Data Corporation or IDC has earmarked the surging adoption of public and private clouds as one of the three emerging technologies to go mainstream in 2011. Low upfront costs, scalability and ease of use are only some of the benefits being touted by cloud storage vendors.

According to the experts, the elasticity of cloud services, coupled with the highly dynamic nature of digital data growth, makes archiving well suited for the cloud. Furthermore, the process plays well to the cost efficiencies gained through Infrastructure as a Service (IaaS) implementation.

The IaaS service model enables user organizations to forgo deployment of new data center equipment to handle growing operational needs. Rather, the enterprise obtains needed IT Infrastructure from a cloud services provider, often via a self-service catalog.

### Archiving as a Competitive Advantage

Using an automated tiered storage solution as part of an overall archiving strategy is a way for businesses to gain competitive advantage, no matter what size or what industry. The reality is that all businesses will be dealing with data growth, and those that don’t deploy the proper solutions are running significant risks of not being able to control spending on storage, not operating efficiently and not being able to meet the demands of users, customers and, in some cases, lawyers, accountants and regulators.

Consider the competitive advantages of archiving:

- The business is maximizing its storage infrastructure and saving money on energy costs and improved utilization of storage resources.
- Applications are running efficiently and not breaking down because they are overrun with too much data.
- The IT department is less bogged down with managing inefficiencies and thus more responsive to business needs.
- The company has a strategy in place to deal with compliance and e-discovery challenges before they become crises.
- The data storage environment is maximized for backup, disaster recovery and business continuity.

If a storage archiving solution can deliver these benefits, how does that not translate into competitive advantage? By contrast, if the organization is struggling with data management challenges, perhaps it is time to start thinking about a long-term solution. Simply waiting for the data to stop growing is not a solution. It will likely not happen.

With a smart archiving environment, access to information for compliance, e-discovery or just to get work done is possible without draining IT operating budgets. What’s more, intelligent archiving will reduce the cost of storage and the energy footprint of the data center.

Smart archiving is attainable. It begins by understanding what information needs to be archived and how the enterprise values that information. With context in hand, IT can better evaluate the processes, risks, applicable regulations and laws, and then the technology necessary to achieve the desired results.

For the knowledge workers in the organization, archiving does more than simply solve regulatory or discovery challenges. With access to information on all storage tiers, workers can make better business decisions.
CDW: A Data Storage and Archiving Partner that Gets IT

There are many options in dealing with data storage and archiving. By nature, it will require a multivendor solution, possibly involving software solutions and appliances, as well as NAS or SAN storage and various types of drives such as SATA, Serial attached SCSI (SAS) and Fibre Channel.

When dealing with this type of complex challenge — one that is so critical to the success of the business — it is important to work with a partner that understands all angles and is capable of providing a wide array of solutions, often across different manufacturers and disciplines.

It is also critical to have help in evaluating the different approaches and seeing how they apply to the specific challenges that today’s companies and IT departments are facing.

CDW brings together this wealth of expertise and knowledge, plus the experience of working with leading manufacturers of storage and archiving solutions. CDW’s multifaceted approach includes:

• An initial discovery session to understand your goals, requirements and budget
• An assessment review of your existing environment and definition of project requirements
• Detailed manufacturer evaluations, recommendations, future environment design and proof of concept
• Procurement, configuration and deployment of the final solution
• Ongoing product lifecycle support

Back up, recover, discover and retain information faster and more efficiently with Symantec

Unforeseeable events should not bring your business to a halt. Symantec delivers top backup, recovery and archiving technology to help you recover systems, data and e-mail — in minutes, not hours. Symantec allows you to archive e-mails and reduce storage costs by minimizing the base of information that must be protected. Also, safeguard information wherever it is stored — on desktops, notebooks, servers or e-mail.

A good data protection strategy minimizes the risk of downtime and data loss as well as the risk of a compliance breach. Competitive data protection implementations can be complex and costly, requiring a number of specialty products. NetApp has a better solution: NetApp Integrated Data Protection.

NetApp Integrated Data Protection delivers high availability, backup, archive/compliance and disaster recovery/business continuity directly from storage — replacing multiple products with a single, high-efficiency platform.

The right balance of technologies can help you minimize total cost of ownership (TCO) and achieve a rapid payback. By integrating Quantum’s DXi disk-based deduplication for short-term retention and Scalar iLayer tape libraries for your long-term backup and archiving strategy, you ensure rapid access to the most likely restores, eliminate storing of costly redundant data and cost-effectively retain older data for disaster recovery and compliance. Plus, simplified management across the tiers cuts down on costly administration time.

HP’s Business Continuity and Availability Solutions help ensure companies continue their business despite the circumstances — ranging from an application failure to an operator problem, a security breach or full-scale disaster. These solutions are typically used for critical business processing, as defined by the customer, such as call centers, financial reporting and even e-mail.

CDW.com/symantec  CDW.com/netapp  CDW.com/quantum  CDW.com/hp