White Paper

Seven Essential Strategies for Effective Archiving

Sponsored by:

Abstract

Archiving helps organizations effectively retain, manage, and leverage their information assets. It is also complimentary to backup and broader data protection activities. This paper outlines seven essential strategies for archiving that drive cost savings, risk reduction and IT transformation.

Disclaimer

contoural

Table of Contents

Introduction

IDC have recently estimated the size of the digital universe at 1.8 zettabytes (1.8 trillion gigabytes in 500 quadrillion files), a volume that is more than doubling every two years. In the face of such growth, managing information seems to be getting harder to accomplish. The difficulty is exacerbated by:

- The need to retain larger volumes of data for longer periods of time to meet governance and compliance requirements.

- The need to maintain accessibility to data for business intelligence and eDiscovery requirements.

- Limited information technology (IT) budget dollars.

Organizations and vendors have responded with technologies based on magnetic tape, specialized storage platforms, optical media, automated storage tiering and deduplication. Each of these can play a role in managing retention and controlling costs, but given today's explosive growth of information, some are better than others.

This paper outlines seven critical strategies for effective archiving that drive cost savings, reduce risk and enable "information transformation."

The Importance of Archiving

When customers are asked what "archiving" means to them, the answers depend on the perspective of the respondent. To IT managers, archiving often means the placement of electronically stored information (ESI) on the most cost-effective media throughout the information's lifecycle. Compliance officers want archives to contain ESI that can be indexed, located and controlled, is stored authentically, and has retention policies applied. And attorneys require historical data stores that can be quickly searched for information to support an immediate legal challenge. We believe there are three key drivers for archiving.

To Improve Data Storage Efficiency

Typically, organizations have implemented formal archiving processes and technologies to improve IT cost control. With archiving, capacity on expensive primary storage can be reclaimed by moving infrequently-accessed information to lower-cost tiers. Reducing the size of primary data stores can reduce backup and recovery times and boost database and application performance.

> Definition of an archive: "…a **specialized repository**…used to…**preserve**, protect, control, maintain authenticity and integrity, accommodate physical and logical migration, and **guarantee access** to information and data objects over their required retention period."
>
> (Source: SNIA)

To Promote "Information Transformation"

Archiving can help organizations use growing volumes of information in potentially new and unanticipated ways. For example, new product innovation can be fostered

if engineers can access archived project materials such as designs, test results, and requirement documents. Customer service can be improved by providing ready access to historical customer records, email, and correspondence. Healthcare records can be used as reference material for new research. Archiving such information with secure data retention capabilities keeps it secure, accessible and ready to provide business value today and in the future.

To Cope with eDiscovery and Regulatory Requirements

Regulations force organizations to secure certain types of data and make them accessible. It's important for companies to know where that information resides so that they can protect it, and to eliminate stale data so that it doesn't get them in trouble. eDiscovery – the requirement that customers collect, preserve, review and produce information for legal proceedings and investigations – remains a big concern because of its urgency and expense, both of which can disrupt IT operations. Capabilities for efficient retention and searching of data stores – enabled by archiving – can substantially reduce risk and cut the cost of compliance and eDiscovery.

Seven Essential Strategies for Effective Archiving

Archiving helps organizations more effectively retain, manage, protect, and utilize their information assets. It should be considered as an essential element of an overall data protection and storage management strategy.

Strategy 1: Clearly Define the Roles of Backup and Archiving

Many organizations depend on backup tapes for long-term retention. For such organizations, they are considered to be an "archive." We believe that this is misguided.

Backups are secondary copies of active production data. Since backups deal with constantly changing business information, they are generally short-term focused and often overwritten. This makes them a poor choice for retaining data for compliance reasons. Further, retrieval of fine-grained information (for example, a single email or SharePoint document) from backups can be time-consuming and expensive, especially if the tapes need to be brought back from off-site storage.

Archive	Backup
Primary information	**Copy** of information
Used for **compliance and efficiency**: – data retained in original form – enables response to legal or regulatory action – offloads information from production systems and storage	Used for **recovery**: – improves availability, allowing applications to be restored to a point in time
Typically **long-term**: – months, years or even decades	Typically **short-term**: – information held for days or weeks

By contrast, archives focus on information retrieval, usually at the level of a file, e-mail or other individual piece of content. When a piece of information stops changing or is no longer frequently used and long-term retention must be applied, it is best to move it to a searchable archive, from where it can be retrieved in less time and at lower cost compared to backups.

If magnetic tapes are still in widespread use, formal archiving procedures can be a "bridge" technology to help eliminate or at least reduce their role. In fairness, the dynamics of tape usage for backup have changed in recent years. Leading IT departments have embarked on conscious tape elimination efforts that have yielded significant benefits. For example, Virtual Tape Libraries (VTLs) use disks to emulate industry standard tape libraries. VTLs can help IT departments reduce Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) significantly, consolidate storage and keep backup information readily accessible. But the sheer increase of information volume means that this approach fills the capacity of storage arrays in many organizations.

Our advice is to *move away from the paradigm of "backup as archive" for long-term retention*. Backups – especially magnetic tapes, because of the cost of physical storage and the latency involved in retrieving information – are a poor choice for managing data to enable eDiscovery, to comply with regulatory requirements and to promote information transformation as we've defined it.

Strategy 2: Adopt a Disk-Based Archive

In addition to meeting governance and compliance requirements, organizations retain data for business intelligence and competitive advantage. A data set's present minimal or unknown value might become apparent in some future and unforeseen context; just consider what's happening with "big data" and predictive analytics. Both active and archived information can help data scientists – and rank-and-file employees – drive new innovations or help to improve current business processes. In this way, archiving can help organizations "expect the unexpected".

Furthermore, with a robust archiving system in place, employees will be less likely to engage in "underground archiving", whereby documents are saved in private repositories, laptops or USB drives, all of which are outside of the control of the organization.

To make sure that information is readily accessible, *archived information needs to be available online*. Organizations relying upon traditional tape-based archiving will be challenged to address business needs due to the inherent performance and accessibility characteristics of tape. It might also be tempting simply to devote high-capacity arrays of inexpensive drives (or "just a bunch of disks," also known as JBOD) to the archiving task. But "cheap disk" is only inexpensive to buy. A far greater operational cost comes from the management of such resources – provisioning, backup/protection, replication, etc.

A better approach is to deploy a disk-based solution to address the needs of archival workloads. Some of the characteristics of such a solution include, but are not limited to:
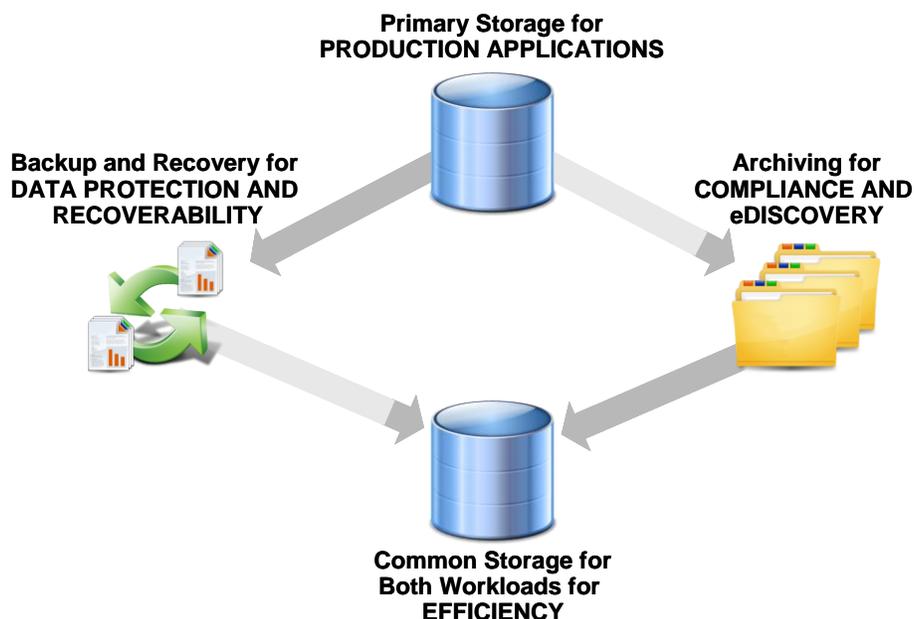
- Employs deduplication to maximize storage efficiency.

- Ensures data integrity, protection and security.

- Guarantees that data can be restored, even after a lengthy time period, taking into account changing drive technology, software versions, etc.

- Enables administrators to enforce multiple retention periods for different archive data sets stored on the same system.

Customers and industry observers seem to agree that an online approach to archiving is gaining momentum:

> "The growth of digital archive capacity stored on external disk systems will also exceed the overall market growth rate, increasing at a 62% CAGR between 2010 and 2015…rapid growth in the use of external disk-based storage systems for digital archiving means that the share of all archive capacity stored on external disk will increase from 29% in 2010 to 35% in 2015." (Source: Enterprise Strategy Group, June 2010)

Strategy 3: Converge Backup and Archive Storage

Backup is driven by the need for recoverability and disaster protection while archiving is driven by the need for improved efficiency, information re-purposing and to address compliance and eDiscovery challenges. Real cost savings can be realized by adopting a strategy for the physical storage of *both* backup and archiving workloads. To accomplish this, a common storage target must be able to handle the throughput and inline deduplication requirements of backup workloads and secure and long-term retention requirements of archive workloads. In addition, the storage target should provide built-in capabilities for network-efficient replication for disaster recovery needs, enterprise features such as encryption, and allow for easy integration with existing application infrastructure.

Primary Storage for PRODUCTION APPLICATIONS

Backup and Recovery for DATA PROTECTION AND RECOVERABILITY

Archiving for COMPLIANCE AND eDISCOVERY

Common Storage for Both Workloads for EFFICIENCY

Backup and archiving are necessary and complimentary IT procedures. By leveraging a common infrastructure for both, organizations can greatly ease the burden of eDiscovery, data recovery, business continuity, and compliance and achieve these goals in the most cost-efficient manner.

Strategy 4: Use Storage and Application Intelligence to Classify Information

Organizations *must* know much more about the information that they possess than they did just a few years ago. Of course, it is important to know how many petabytes are in place and where they reside, but to stay relevant and competitive; organizations must understand *why* they have them.

Clearly, information has intrinsic value. Otherwise, it wouldn't be captured and retained. But the value of most data sets is subjective; deciding how much effort and resources should be spent on keeping, expunging or archiving them can be a confusing and contentious issue.

The good news is that determining the value of information can be made easier by taking a good inventory of what's in place and classifying it effectively. As the volume of information grows, so does the amount of metadata – "data about the data" – which makes value judgments about information stores possible. Metadata includes things like file and block size, date of last access, location and many other storage-related parameters that enable "storage intelligence" about what exists, what resources it is consuming and whether it should be moved to lower-cost storage tiers.

Core applications like Enterprise Resource Planning (ERP), Microsoft SharePoint, structured databases, social software (like Twitter and Google+) and document authoring tools generate rich and varied application-specific metadata like keywords, subject, versions, security attributes and linkage to other applications. Such metadata can be regarded as "application intelligence" at the level of the data item or business object and not just at the file or block level.

Classification – made possible by combining storage and application intelligence – helps organizations identify the importance and value of information. But classifying everything all at once can be a daunting task. We suggest that customers do the following:

- Identify the data sets that carry the most potential for exposure to regulatory, legal or eDiscovery risk and the associated costs. Some common examples include email, SharePoint, and networked file shares. Make these the initial sources for "intelligence-gathering."

- Use automated tools for indexing, auto classification, text and content analytics, search and data mining to extract meaning and value from the data deluge.

Classifying information is a crucial element of a sound archiving strategy that drives policies for long-term retention and defensible deletion.

Strategy 5: Adopt Value-Based Retention Policies

Typical decisions about the long-term retention of information are based on IT operational needs. Sure enough, holding on to content based on file system metadata – such as age or file size – makes it possible to capture and migrate content to lower-cost tiers of storage. But such a capacity-based approach makes no allowance for the importance or confidentiality of such information. Perhaps there's no good reason to save it at all!

A better approach is to supplement capacity-based retention rules with policies aligned with *business value*. For example, retention policies can be based on any or all of the following:

- External regulations or legal mandates that define what kinds of information to save and for how long.

- The requirement of departments (e.g., finance, manufacturing, sales) or business units to save different information for varying lengths of time.

- Requirements to preserve certain historical information for operational continuity reasons.

The notion of value-based retention dovetails nicely with the concepts of records series, file plans and legal holds that are essential to the discipline of "records and information management" (RIM). To make life easier, we recommend that IT managers reach out to their RIM colleagues who have already dealt with issues about value-based retention policies.

Strategy 6: Enable Enterprise-wide Defensible Deletion

While some organizations get in trouble for not saving documents for the required retention period, most face the opposite problem: their *de facto* policy is to save nearly all information forever.

A defensible deletion policy ensures that organizations do not delete documents they need to retain or preserve. It provides some level of protection against litigants and regulators who, in the future, may ask uncomfortable questions about why specific documents have been deleted. Finally, it affords justification for the removal of unneeded information that would otherwise drive up storage costs and increases the risks and costs of eDiscovery.

To achieve defensible deletion, we recommend that customers:

- Put effort into determining what needs really needs to be saved. A good archiving system based on the value of information to be retained can greatly aid this process.

- Make sure that policies include both the business justification and process for deleting documents.

- Create a clear and consistent legal hold process that clearly identifies information being held, allowing for routine deletion of data not under hold.

Strategy 7: Plan for Today and the Future

So far, we've only alluded to this important and sometimes overlooked aspect to the archiving program: gaining early organizational agreement and maintaining it over time. Decisions about long-term retention, defensible deletion and information accessibility are important beyond the IT department. They should be made up-front and with the agreement of major stakeholders.

We advise regular consultation with groups such as the legal and records management teams as well as executives and end-users. Incorporate archiving into the shared service plans for storage management, regardless of the topology of the computing and storage network. Make sure that policies and service-level agreements meet business needs and are realistic.

Carefully consider the direction and "velocity" of technological changes. For example, cloud storage today is mature and reliable; plans for backup and archival should consider platforms that can support public and private cloud environments. "Big data" means high volumes of very large (e.g., surveillance video) and very small (e.g., Telco call detail records) file sizes, forcing organizations to determine the value of information to be retained. And expectations of end users means that access to active and archived information needs to be available to smartphones, tablets, laptops and desktop systems. Make sure that what is put in place works today and can accommodate the future.

Conclusion: Stay Ahead of the Curve with Archiving

Managing information *seems* to be getting harder to accomplish, but it doesn't have to be that way. An effective archiving strategy will help organizations retain, manage, protect, and utilize their information assets effectively by:

- Centralizing views of retained information in a digital, online archive so that it is searchable and readily accessible by end-users, either in routine operations or in totally unexpected ways.

- Taking advantage of advances in object-based storage, deduplication and tiered storage topologies to direct both backup and archived information workloads to a common storage target.

- Leveraging storage and application intelligence to measure the value of information at the level of the data item or business object.

- Simplifying retention and defensible deletion policies based on information value, rather than driving toward a costly "save everything forever" approach.

- Staying close to the needs of the business and trends in technology.

Challenged by growth in volume and variety of information, archiving has become an essential IT activity. With a well-designed archiving strategy in place, organizations need not be victims to rising costs, compliance challenges and legal risk.

About EMC

EMC Corporation is a global leader in enabling businesses and service providers to transform their operations and deliver IT as a service. Fundamental to this transformation is cloud computing. Through innovative products and services, EMC accelerates the journey to cloud computing, helping IT departments to store, manage, protect and analyze their most valuable asset — information — in a more agile, trusted and cost-efficient way. Additional information about EMC can be found at www.EMC.com.

About Contoural, Inc.

Contoural is the largest independent provider of Information Governance Consulting Services with expertise in Records & Information Management, Litigation Readiness and Sensitive Information Control. Having worked with more than 20% of the Fortune 500, as well as a number of Federal agencies, they enable companies to become proactive and defensible in controlling, managing and deleting their business information to reduce risk, increase compliance and lower costs. Contoural is independent and provides proactive consulting services; they do not sell any products, provide "reactive" discovery services, or warehouse any information, and as such our clients truly view us as trusted advisors.

With an average of 14 years industry experience, Contoural's team is comprised of attorneys, former compliance officers and records managers who have a deep understanding of the legal, compliance and business requirements for retaining and managing information -- as well as seasoned IT professionals with expertise in document archiving, search, litigation management systems, data classification and data storage, all focused on effective program execution.

Contoural services include:

- RIM and Litigation Readiness Assessment and Roadmap Development Services
- Records and Information Management Policy Development Services
- Data Classification Services
- Litigation Readiness Services
- Solution Design, Technology Evaluation and Vendor Selection Services
- Solution Implementation Services
- Change Management and Education Services
- Ongoing Program Management Services

With these services, Contoural helps enterprises ensure compliance and reduce risk, while also achieving litigation readiness and reducing costs.

Contoural, Inc.
5150 El Camino Real, Ste D30
Los Altos, CA 94022
650-390-0800
www.Contoural.com
info@contoural.com