sageza

Strategic Snapshot

The Business Advantage of a Unified Backup and Archival Strategy

By Clay Ryder and Rob Kidd

The Sageza Group, Inc.
March 2005

The Sageza Group, Inc.
32108 Alvarado Blvd #354
Union City, CA 94587
650·390·0700 fax 650·649·2302
London +44 (0) 20·7900·2819
Milan +39 02·9544·1646

sageza.com
info@sageza.com

The Business Advantage of a Unified Backup and Archival Strategy

ABSTRACT

In our January 2005 paper "Mapping the Business Value of Data Backup and Archival Solutions" we discussed the differences between backup and archive. This paper expands upon that thinking, and highlights the business benefits that can be realized from leveraging both, together, in a unified backup/ recovery and archival strategy.

Recent changes in the business and technology environment have ignited reexamination of enterprise data backup, extraction, and archiving. It has become clear that there is a compelling need for a strategic alignment of backup and archival processes within organizations. Traditional thinking and paradigms are evolving to more unified, yet granular approaches to data management that leverage Information Lifecycle Management (ILM) concepts. It is no longer sufficient to perform complete backups of production data without consideration for cost and operational efficiency, as this will only lead an organization to a competitive disadvantage relative to more progressively thinking organizations. Regulatory, competitive, and physical limitations have narrowed backup/ recovery windows, and government compliance and corporate governance have placed new demands on archiving, such as more extensive archiving with expedient, precision-point data recovery requirements.

Although backup, extraction, and archiving remain discrete business processes and technological solutions, there are competitive and strategic advantages to be gained by organizations that align these processes with the value of the business data being stored in order to minimize costs and maximize efficiency across the enterprise. Failure to bring an organization's storage policies into alignment with business value will result in the continuing unchecked and costly growth of the production, backup, and archive data environments.

This paper discusses the current state of backup, recovery, and archiving in organizations today, and offers some definitions and clarity to the often-misused terminology of extraction, archive, and backup. It considers the minimum technological capabilities required by enterprises in order to create cost-effective backup and archival strategies that align the overall cost of data storage with the business value of the data being stored. It also provides an understanding of current backup and archival issues and the expected benefits of a unified backup and archival strategy.

The Business Advantage of a Unified Backup and Archival Strategy

TABLE OF CONTENTS

The State of Backup and Archiving in Organizations Today

While the terms *backup*, *archive*, and *recovery* are frequently interchanged, each defines a specific business process and technological implementation that collectively allow an organization to create and deploy a cost-effective and efficient data storage strategy. These combined processes leverage ILM concepts to improve production performance and reliability, reduce the length of backup windows, provide faster application recovery and information retrieval, and improve the security of the enterprise's electronic data.

Backup

Backups are secondary copies of production data that are usually updated/overwritten on a regular or automated schedule. The fact that these copies have a comparatively short life span allows them to be useful for point-in-time recovery operations but disqualifies them for long-term data retrieval needs that an archive provides.

The goal of backup is the short-term protection of data to assist or protect critical business processes and provide disaster recovery. That backups are periodically overwritten after relatively short intervals makes the process inadequate for regulatory and compliance issues.

Recovery

The purpose of backup recovery is to restore business operations by repopulating operational data as quickly as possible.

Recoveries can be required for reasons, including disaster-level outages, security risks, natural disasters, data destruction, corruption, viruses, or unintentional deletion. Regardless of why the data or applications are no longer available, recovery focuses on speed and accuracy. The standard media for this process today are various forms of tape. With today's business pressures, certain performance and architectural inefficiencies of tape-based environments are apparent. Many organizations are no longer willing to justify further investment in a technology that cannot reach higher service levels.

A recent development for Backup/Recovery environments employs a more effective and efficient strategy leveraging disk technology for solutions that look like a tape drive to the server and backup software, maintaining tape in a much reduced role. No process or backup application changes are required, yet dramatic increases in recovery speed can be achieved and the management and media issues around tape are avoided.

Archive

The primary purpose of archiving is the long-term protection and accessibility of critical information for business, governance, or regulatory compliance. Typically, archived data consists of information that is not needed on a day-to-day basis, but is important enough to be retained. Archival data may also include "fixed content" material that is referenced frequently, but is not expected to change often. For many, having an archive simply means keeping backup tapes for a longer period.

However, today's business requirements to fully leverage an organization's information assets dictate that an archive must be "active" and must provide fast random access to individual pieces of information all while assuring its authenticity and the ability to enforce applications retention and disposition policies intrinsic within the storage layer. These requirements are vastly different from what tape technology was designed for, i.e., backup and recovery. With this realization, many users are investigating and deploying intelligent disk-based archiving solutions.

The Critical Question

The pressure on today's backup/archive strategies and technologies derives from the reality that production data is growing at an ever-increasing pace and that "normal" business hours for many organizations are now 24x7. The result is that the number and length of backup windows are decreasing or disappearing. At the same time nearly all organizations have greatly increased the volume and breadth of data that must be retained, making the high-maintenance, slow-recoverability nature of tape inadequate for enterprises of all sizes.

If backup windows are being squeezed, how does an organization make backup and archive processes significantly faster without adding costs? The solution is to create a unified backup and archiving strategy that extracts and archives a considerable volume of the information before the backups are done. This new approach requires that the organization adopt principles to ensure proper long-term data retrieval capabilities while increasing the performance and success of recovery operations.

Although there are some leading-edge organizations that have fully embraced a unified backup and archival strategy, for the bulk of the IT community, this is not the case. How then should an organization think about or move toward this optimal unified strategy?

What Is the Optimal Backup and Archival Strategy?

There is a range of backup and archive solutions today that incorporate, to varying degrees, the previously mentioned components of information management. To maximize the business value of data assets, organizations need to develop and implement a unified backup and archival strategy that offers fast, online, standards-based access to content from any application on any platform.

An optimal unified backup and archival strategy would:

- Discover and classify all information and data types in the environment, rank the criticality of the information and set policies based on business, internal governance, or regulatory requirements.
- Establish a tiered storage strategy that aligns the cost of storage to the business value of the information.
- Provide for the extraction of certain data prior to backup in order to increase backup performance and eliminate redundant operations, e.g., extracting 7- or 30-day data from the production environment.
- Guarantee content authenticity to the standards in force for day-to-day business, internal governance, or regulated environments.
- Continue to support legacy tape library processes where necessary or appropriate, regardless of whether the "tape" library is actually a tape-based solution or a form of faster media with embedded tape emulation software.
- Transparently execute very fast information restores, for business continuance and disaster recovery.

The specifics for any unified backup and archival implementation will vary according to enterprise need, but the basic strategy and requirements will remain constant. The above points are merely a starting point for organizations to begin contemplating a path toward an integrated Backup, Recovery, and Archival solution. It does not represent the full list of elements expected of any unified backup and archive strategy.

Beyond the cost of human resources, unnecessary expenditure can be reduced by minimizing storage of content duplicates, reduction of systems management, and transfer of inactive or final content off primary storage and into an archive. Overall, backup processes and infrastructure should be sized and tuned for the protection and restore needs of production data. Extraction serves an important ILM function to streamline backup and leverage an active archive to deliver content as needed for a variety of business reasons.

Understanding Extraction and Replication

Part of an optimal unified Backup/Recovery and Archive strategy is the use of extraction and replication. While these are commonly used terms, as ILM concepts they can be leveraged to improve performance and reliability, reduce the length of backup windows, and provide faster application recoveries and retrieval of archival data. As techniques, they help align the business value of information with its storage costs while improving service-level performance.

Extraction

One of the goals of extraction is to reduce the quantity of information being backed up by reducing the amount of information stored in production environments. This process also improves end-user application performance and reduces the cost of IT infrastructure. The result is fewer application servers, backup server cycles, backup media, and personnel efforts.

Extraction requires a level of data classification of information that many organizations have not yet developed. Tools exist that allow file-level review of data which can help this process, but as a general rule, a consulting engagement with a trusted vendor is the most reliable way to help classify enterprise information.

Upon evaluation, inactive data that has value to the business, or that is required to be kept from a governance/compliance standpoint, should be archived. This archival process should be deployed on tiered storage infrastructure to gain maximum performance levels at the most appropriate cost points. Data uncovered in the classification processes that does not meet these criteria can be deleted.

Replication

Many forms of replication can be utilized in an integrated Backup/Recovery and Archive environment. Replication is a technique to maintain updated copies of data, and is widely deployed for disaster recovery or business continuity. There are different techniques for replication ranging from host-based software options to bi-directional replication of complete storage environments.

In the most common use case for Backup and Recovery, a snapshot or clone copy is made of a data set in a very short time frame, and that copy of data is used for creating a backup image. The value of this technique is that the backup window is lessened to the time required to make a consistent copy with the replication software of choice.

Extraction complements replication as it benefits a “regular” backup process. The amount of data to be replicated is lessened, meaning that the replicas can be created more quickly and will exert less stress on the system responsible for their creation.

For archiving, business continuity, and disaster recovery scenarios the best replication approach is bi-directional asynchronous replication over a wide area network. This automatically creates a mirror image of the information at a remote site and addresses concerns with offsite and long-term data protection. This technique is the domain of disk-based storage solutions.

Benefits of a Unified Backup and Archival Strategy

The following examples illustrate some of the tangible IT benefits from adopting some or all of the unified backup and archival strategy components we have outlined.

Dealing with Growing Quantities of Information

Organizations across the board are handling ever-increasing quantities of information, including billing records, email, Web sites, and more. With continued growth in their information base and usage patterns, backup and recovery windows can be squeezed beyond the point of manageability. A desire to maximize the business value of information, improve the "total customer experience," and a new focus on internal governance or regulatory requirements are but a few of the reasons that many organizations are reassessing their backup and storage resource management policies. For example, virtually every organization today has email. With a unified Backup/Recovery and Archive strategy, organizations can reduce the size of primary email stores, extracting and archiving finalized email and attachments prior to back-up, all while having immediate access to said information. Reclaimed primary storage capacity could be allocated to new application purposes thus delaying the need for additional storage expenditures. Any organization with a sizable email environment would be well advised to explore the opportunities afforded by a unified backup and archival strategy.

Other enterprise IT applications that can potentially benefit from deployment of unified backup and archival strategies include enterprise content management (ECM), medical or business imaging, mainframe output, and DBMS, to name a few.

Enhancing ECM

In the case of ECM, the two important factors are interface continuity with production systems and rapid archive retrieval, manipulation, and management. For example, it is very important for customer service agents to be able to rapidly access target customer information, cross sells, and up sells. In these instances production system extractions that are closely correlated with archives are important to the overall efficiency of the ECM process. For example, a call center representative cannot know in advance if the present caller last contacted the firm six days or six months ago yet must be able to quickly get at archived business records if the customer's questions are to be answered in one interaction. Unified backup and archive have an important role to play in ECM as businesses seek to reduce the quantity of "current" information while at the same time enabling rapid access to historic customer data.

Medical Imaging/Records Management

Medical imaging/records management and patient data is another area where the principles of unified backup and archival are critical. Emerging medical management systems rely heavily on rapid and diverse archive data retrieval, to support operational production systems and tactical patient care goals. Physicians and practitioners use these myriad images and records for day-to-day health care decisions and care management. The key is to have information from the patient medical archives, such as x-rays or cat-scans, rapidly available to administer the most effective patient medical care. Yet the cost of maintaining such records in primary data stores would prove enormous. The effective delivery of this information depends on a strategic archival and retrieval process. This process can be substantially improved by implementing a selective extraction strategy in conjunction with backup and archive.

Streamlining the Production DBMS Datastore

Enterprise production DBMS environments have always been subject to rigorous business continuity, backup-and-recovery regimens, and service-level expectations. It is only recently that with increasingly more demanding compliance and governance regulation, correspondingly more rigorous archive requirements have been brought into operation. DBMS environments are complex, with each requiring individual review, but in general extracting 30% of the final form information prior to backup and archive would reap significant benefits in production systems performance, providing more rapid retrieval and allowing data value alignment with storage asset costs.

As with all enterprise strategies, unified backup and archiving strategies are key to overall enterprise IT success, with each environment requiring its own detailed analysis and review.

Unified Backup and Archive: Pulling It All Together

Organizations that undertake a unified backup and archival strategy can expect to enhance their overall cost structure and competitiveness with the following benefits:

- Improved production system and application performance with less human intervention, tuning, and related personnel costs.
- Reduction in the time and management needed for backup and recovery, while providing more robust business continuity protection and decreased costs associated with business interruption that may occur in a recovery scenario.
- Faster retrieval of information from the archive for more timely responses to business, governance/regulatory issues, reduced manpower expended, and lesser exposure to possible embarrassment, litigation, or fines.
- More efficient management of the backup, processes, and associated storage.
- Hardware cost savings achieved through higher utilization of all infrastructure resources by matching information value to appropriate tiered storage and the commensurate deferral in additional hardware expenditures (on both servers and storage).
- Improved overall organizational efficiency and opportunity through enhanced deployment, service levels, and management of valued information assets.

Traditional and contemporary technologies provide IT with a variety of options for virtually every backup and archival need. However, to achieve maximum utilization, efficiency, and competitive advantage, organizations should undertake a coordinated and unified strategy with respect to disk-based backup and archival. Such a strategy demands careful consideration of all business needs as well as those of IT to provide the most cost effective and business enhancing solutions possible.

What Does It All Mean?

Although there are some leading-edge organizations that have fully embraced a unified backup and archival strategy, for most organizations, this is not the case. At present, most storage managers are playing a game of catch-up, hoping to avoid short-term operational restores while seeking to comply with an ever-changing world. Business, competitive, and physical limitations have narrowed back-up/recovery windows, and corporate governance and regulatory requirements have placed new demands on enterprises' archive strategies. The inefficiencies of past backup and archival solutions cannot be incrementally overcome through further investment in the same: it is no longer sufficient to simply copy production

data to tape at the end of each day, and store these tapes offsite as an archive. A new unified strategy for backup, recovery, and archive is needed that maximizes the alignment of business data value with the cost of its storage, while improving performance service levels.

A unified backup, recovery, and archival strategy at a minimum should include:

- The discovery, classification, and ranking of all information and data types in the environment based upon policy.
- The extraction of unneeded data prior to backup and the retention of needed "fixed content" to an archive to increase performance and eliminate redundant operations.
- The support of legacy tape processes, regardless of whether the "tape" library is actually a tape or disk-based solution.
- Deployment of tiered disk-based storage resources, including an "active" archive that provides online access with assured content authenticity to information.

We believe the implementation of such an objective will help align business costs with information value and provide organizations with a strategy to reduce overall costs while achieving higher levels of information utilization, customer service, and ROI.

Although backup, extraction, and archive remain discrete processes, there are strategic advantages to be gained by organizations that align these processes with the value of the business data being stored in conjunction with ILM initiatives. Failure to do so will result in the continued unchecked and costly growth of the production, backup, and archive data environments. We believe that organizations who undertake a unified backup and archival strategy are well positioned to reap the benefits of intelligent archiving and generate a sustainable competitive advantage.