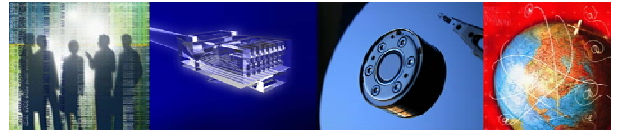




IT Knowledge • Business Results



White Paper

Separating Backup and Archive Processes to Meet New Challenges of Information Management

By Brian Babineau
Analyst
Enterprise Strategy Group
Intelligent Information Management

September, 2006

Table of Contents

Table of Contents	1
Introduction	2
The Differences Between Data Protection and Archiving	2
Using Separate Backup and Archive Processes for Your Benefit	4
Conclusion	5

Introduction

IT departments hardly need another process dictating how they do their jobs. However, in some cases, due to external factors or inefficiencies, IT should reevaluate existing processes to see if improvements can be made. If IT has not already done so, now is the time to separate backup (data protection and recovery) from archiving (data retention and retrieval) because organizations need to differentiate between copying data for recovery and retaining data for future reference and retrieval.

When it comes to protecting corporate information, IT usually has well-engrained procedures to back up data nightly and, depending on the business requirements, send the data offsite daily for business continuity purposes. As an insurance policy, IT usually retains historical backups for months or years, designating this data 'corporate archives.' In some cases, IT uses these archives to meet regulatory requirements or satisfy electronic records retention programs. The archiving process equates to saving old backups, usually on tape media.

The combined process of backup and archiving served IT departments well up until the turn of the 21st century.. At this time, record retention regulations, such as rules 17a-3 of the 1934 Securities Exchange Act (SEC rule 17a-3) and the Health Insurance Portability Accountability Act (HIPAA), required organizations to keep certain electronic business records, including e-mail, for specified periods of time. In addition, as a result of the increased scrutiny on corporate governance because of Enron, ImClone, and several other high profile incidents, organizations began implementing electronic records management programs to deter executive malfeasance, amongst other things. As organizations were being required to retain more information, litigators and regulators began to target these formal data repositories, seeking a 'smoking gun' for specific legal or regulatory matters under investigation.

Around the same time, external factors, such as regulatory compliance and an increase in electronic discoveries, challenged traditional information management processes, while IT departments also struggled with completing data protection operations. The glut of digital capacity created by enterprise applications meant that backups took longer and longer to complete. In some cases, IT had to split databases into multiple instances just to hit backup windows.

Because of shrinking backup windows, the implementation of electronic records management programs, and the recognition that all data is 'discoverable' at any time, IT departments now need to reevaluate existing data protection and archiving processes, specifically separating them to meet these new information management challenges while mitigating the risk of data loss and application downtime.

The Differences Between Data Protection and Archiving

Before organizations separate archiving from backup, they should understand the objectives of each process in order to most efficiently utilize supporting technologies. Backup solutions should be used for copying data from a primary storage system (or server) to a tertiary system. Backups enable IT to protect data from corruption to the primary copy of data or to prevent data loss if the primary copy is accidentally or maliciously deleted.

Backup software creates a full copy of data and then identifies changes made since the last full backup and makes an incremental copy. Traditionally, backup processes dictate creating a copy of data on tape storage media, but more recently, IT is using large capacity, inexpensive disk-based storage systems as backup targets to speed up the process. However, IT departments still create copies of data on tape as the media is portable and can be sent offsite to protect the data in the event of a disaster to the primary data center.

To help decipher what backup technologies to deploy, including disk versus tape targets, IT should develop Recovery Time Objectives (RTOs) and Recovery Point Objectives (RPOs) for each application or data type.

RTOs measure the amount of time it takes for an organization to recover data and bring applications online while RPOs determine how much data an organization can afford to lose. IT must balance RTOs and RPOs with the cost of an integrated backup solution that includes how data is backed up, the frequency of backups, the backup target and whether or not data is sent offsite.

Because backups are designed to protect data, all changes to transactional information should be backed up regularly. The flexibility of backing up full data sets and then creating incremental copies enables IT to protect and recover data more efficiently. By utilizing incremental backups, the data subsets are smaller and can be recovered much more quickly than restoring an entire database or application data set. IT should make full copies of large data sets, such as an entire database, only when absolutely necessary. IT must also determine how long to keep full data sets and incremental copies. In many cases, full backups are created at the beginning of every week, with incremental copies made daily. Once the next week's full copy is completed, all of the previous week's incremental copies are discarded.

Backup copies of data can also be stored offsite to protect against an even worse situation in which a primary data center and its contents are destroyed or otherwise rendered inaccessible. In such a situation, IT would need to restore data to another set of applications and supporting systems at a different location. This recovery operation would take significant resources and time to restore the data but, because the data is being regularly sent offsite, the loss of data would be minimal.

Restoring large amounts of data from high capacity media can satisfy recovery requirements, but increasing requirements to keep aged data available for specific periods of time for compliance or litigation support reasons mandate that organizations archive information. ESG defines archiving as the longer term retention of historical data that is no longer needed for current business operations in order to satisfy regulatory compliance, corporate governance, litigation support, records management or other information management requirements. Archived data needs to be searchable and accessible as IT may need to find certain files in response to a discovery request. In addition, other departments may utilize the older data for business analytics or to generate trending reports. For example, customer service organizations may use older bank account transactions to train staff on how to handle customer inquiries.

Despite the fact that archived data is no longer transactional, an organization should consider how frequently the information will be accessed to determine where and how it is stored. When using archived data for business analytics such as reporting or trending analysis, IT should retain archived data on a disk-based storage system. Further, if IT must support regular electronic discovery requests or if business records are constantly requested as part of ongoing regulatory inquiries, an indexed archive that can quickly search and retrieve content is most appropriate.

When designing archive processes, organizations need to consider several internal and external factors. First, organizations should identify whether specific data is governed by record retention regulations. Financial services firms, universities and health care providers all create business records that must be retained according to law. Whether those records are e-mails, x-rays or students' report cards, IT must identify the appropriate technology which can capture these data types, establish a retention period, and then store the information where it can still be accessed but not deleted before the retention period expires. IT may also need to consider who should have access to certain archived files and set access control permissions on the data, to prevent unauthorized usage. There are several software solutions available on the market today that can help IT identify messages, files and database transactions that may be subject to specific regulations. These solutions can set retention periods, establish access permissions and store the archived data on the storage system that aligns with the data accessibility requirements.

Archiving solutions also create an index to help IT and business users retrieve information. Doing so requires archiving solutions to understand certain attributes about data such as the creator, the format, the contents and the date of last access. These attributes, as well as new archiving instructions inclusive of a retention period and access controls, are indexed, enabling organizations to search large amounts of data based on very specific criteria. The rich index created by archiving solutions can be used to find data quickly in the event of an electronic discovery or regulatory inquiry. The index can also automatically identify duplicate content and only

store the attributes of the copies. This reduces archive storage cost and enables organizations to gain control over their data.

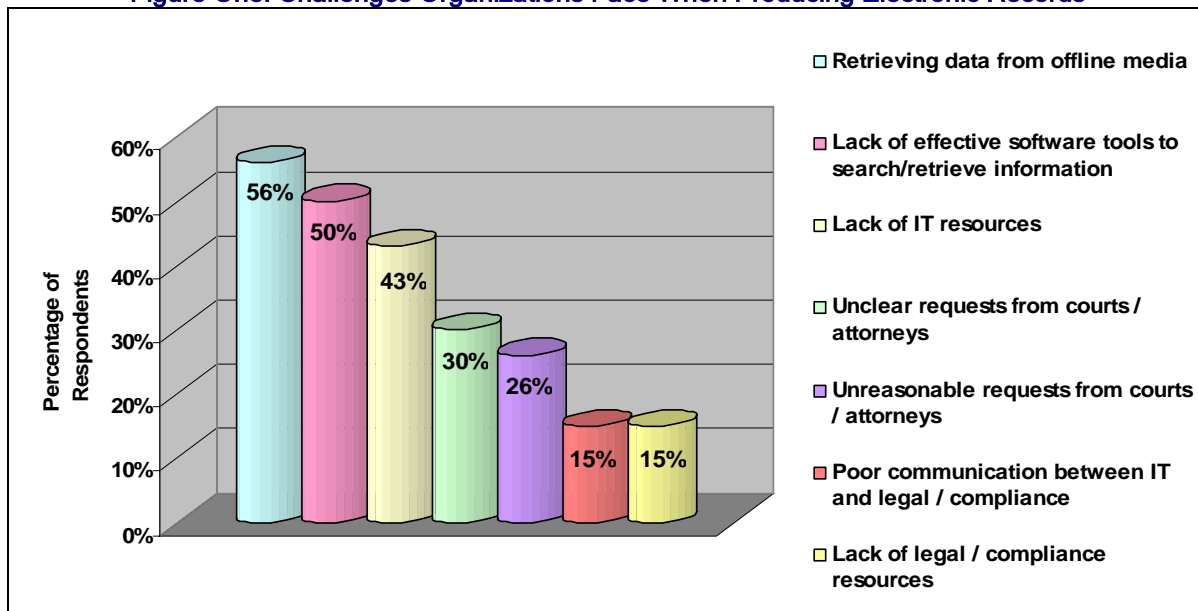
Backup solutions are optimized to create a copy of data from one storage system (or server) to another. In many cases, backup software changes and compresses file formats to expedite the copying process. Smaller data sets can be restored to bring applications online. Archiving technology helps organizations permanently move information, in native format, from one storage system to another while building a rich index for fast retrieval.

Using Separate Backup and Archive Processes for Your Benefit

Many IT departments may struggle to understand why archiving should be separate from backup mostly because the definitions of these terms are often misconstrued. In fact, the true definition of archiving was hardly ever known or debated because it was immediately added to the end of data protection operations. As a result, institutionalized thinking accepted the fact that archives were copies of historical data created from old backups. Only compliance issues or an electronic discovery that requested aged data would create the need to use an archive. Because of the lack of use, businesses rarely perceived archives as accessible information that could be leveraged in everyday activities.

There are several reasons that traditional backups should no longer serve as corporate archives. Regulators and litigators have found several ‘smoking guns’ within unmanaged, out of control backups that constitute an archive. ESG estimates that 46% of organizations have experienced an electronic discovery request in the past 12 months, causing IT departments and in-house counsel to find a better way to quickly locate relevant information. Figure One highlights the challenges organizations face when producing electronic records. By utilizing archiving technology, organizations can gain control over their information by storing fewer copies of the same data, setting retention and disposition schedules to delete information when it is no longer needed by the business, and indexing all information to make it easier to locate relevant data. Attorneys can spend more time reviewing the appropriate information rather than looking for it. Searching through old backup media to find a file or e-mail from a year ago began to drain IT operations resources as this is the only group that knew how to restore the data and search for the requested files.

Figure One: Challenges Organizations Face When Producing Electronic Records



Whatever processes and technologies IT departments determine best meet their RPO and RTO objectives, backups should be used for data protection and recovery. As data gets older, it becomes much more difficult to restore, as it is usually stored in larger data sets, such as full monthly backups. Recovering large data sets is slower and these backups should only be used in a 'worst case' scenario. Backup processes create copies of data to be used for recovery purposes and the copies are retained for extensive periods of time, mitigating any risk of data loss. However, these older copies of data are stored in large data sets often on offline media such as tape, making regular access to granular files within the data difficult.

Figure Two displays the top objectives organizations have for archiving specific content. The most common objectives involve resource management and optimization, where customers can improve storage capacity utilization, improve the speed of data backups and achieve other data protection objectives. One example of how IT can use archiving in conjunction with backup is to remove stagnant data from production systems, facilitating faster backups. If the data being regularly protected is old, unchanging or rarely accessed, but needs to be retained, there is no reason to keep the information on production servers and storage. IT can archive this data, removing it to lower cost storage where it is still accessible but at lower cost. As a result, this data is no longer part of the recurring backup operations. Organizations can complete backups much faster and save money on tertiary media by archiving. The alternative is to delete this data from primary systems, which could cause an organization to be out of compliance with regulations and limits the opportunity to leverage the information for other business purposes, such as reporting.

IT must also understand that its archive, regardless of the size or use (compliance, information sharing, etc) also needs to be backed up. Efficient archiving mandates that the data does not reside anywhere else (because it was moved from primary systems). As such, IT must add the archive system to the backup schema so that, if data corruption or deletion occurs, the information can be recovered.

Backup and archiving processes should be split. However, they should intersect at two points. IT should archive inactive data to free up capacity on primary storage and servers and reduce the amount of data to be backed up from these systems. IT should also add information archive systems to the backup schema protecting this information, which is retained for compliance, sharing or litigation support purposes.

Conclusion

Long standing backup and archive processes served IT departments well from mainframe days through the dot-com frenzy. As more information was created, backup windows began to shrink and IT needed to find new technologies to protect data from corruption or deletion. Because backup and archiving were never separated, IT departments did not realize that they were copying much of the same data. This lack of insight also hurt IT and legal departments that paid upwards of \$3,000 a tape to have historical data restored in response to legal and regulatory inquiries.

Figure Two - Why Organizations Archive **Top Objectives of Organizations That Have Deployed Archiving Applications for Each Respective Content Type**

E-mail

- ✓ Improve compliance with regulatory record retention mandate
- ✓ Improve litigation support capabilities
- ✓ Reduce storage management complexity

Database

- ✓ Increase speed of database backups
- ✓ Improve database performance
- ✓ Reduce storage capacity

Unstructured (Files)

- ✓ Reduce time and cost associated with recovering data in event of disaster/outage
- ✓ Reduce time required to search and locate archived content
- ✓ Improve management of and access to corporate intellectual property
- ✓ Improve compliance with record retention mandates

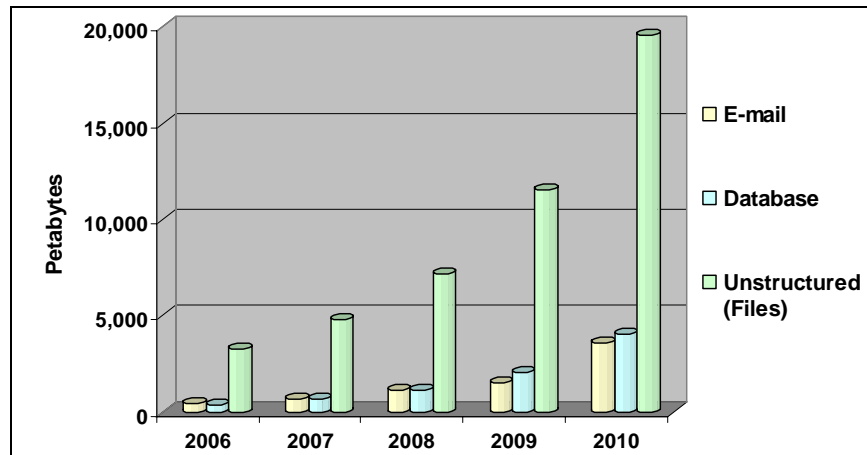
Source:

ESG Research Report: *Digital Archiving: End-User Survey & Market Forecast 2006-2010*

By segregating the archive process from backup, IT does need to manage another procedure, which requires its own infrastructure and resources. However, the benefits of taking one large, laborious process and splitting it into two are too hard to ignore. By archiving information more diligently, IT can remove non-transactional data from production systems, reducing the amount of data to be regularly backed up. In addition, online archives can be indexed, making it much easier to search, locate and review information requested as part of an electronic discovery. IT can also retain business records online for compliance purposes, while also making this information accessible by others to use for trending, reporting and other analytics that were impossible when archived data was stored offline and often offsite. As a result of these benefits, ESG expects that organizations will archive over 60,000 petabytes of e-mails, files and database transactions over the next 5 years, as shown in Figure Three.

Figure -Three: Total Worldwide Digital Archive Capacity by Content Type, 2005-2010

(ESG Research Report: *Digital Archiving: End-User Survey & Market Forecast 2006-2010*)



IT should realize that adding new processes or modifying existing ones is necessary as new information management challenges come to bear. One place to start is to separate backup from archiving to make data protection more efficient, keep information online and accessible at lower costs for longer periods of time and achieve compliance with external regulations and internal governance policies.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. and is intended only for use by Subscribers or by persons who have purchased it directly from ESG. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.