

October 27, 2007

For Those Who Believe That Backup Does Archiving – NOT!

Faced with limited budgets and the desire to simplify operations, some IT executives believe that if they have a reliable backup procedure, it will also fulfill their archiving needs. But with new organization standards for accessing data and with legal and regulatory compliance mandates, IT executives expose themselves to greater costs and management risks by relying on this erroneous assumption. The good news is that archiving not only greatly reduces overall storage costs and addresses long-term data retention needs, but it also dovetails wonderfully with backup. When used intelligently, archiving greatly reduces the scope, expense, and complexity of backup, and thereby makes overall storage management operations much less expensive.

Backup is typically the process of copying large volumes of data onto alternative storage, which can be quickly recovered 'en masse' to rebuild an environment after a system crash or other disaster. The emphasis of backup is to rebuild a storage environment, reverting to a previous safe point-in-time, as fast as possible. To do this, backup software should be optimized for fast data transfer and managing multiple generations of points in time.

Archiving, on the other hand, is focused on data management needs. Archiving software copies or moves data off the primary store to less-expensive storage alternatives without compromising the organization's ability to access the data if needed. It can do this to reduce the cost of storing seldom accessed data. Or, it can do this to ensure safe and compliant data retention under tighter IT control. Since unused data may still have a lengthy retention requirement, archiving particularly responds to the long-term management challenges of storing specific data over a long time period for compliance or legal reasons.

When you want to rebuild a failed system, backup is your tool. When you want to find a group of related files that haven't been touched in 4 years, archiving provides the answer.

Backup is like insurance. Despite all that you invest in backing up the primary storage environment, as with insurance, you ardently hope you will never have to use your backups.

In contrast, the archive is a place you WANT and fully expect to use. It is about creating a very low cost -- yet accessible -- repository for data that is not accessed sufficiently to merit being stored on your expensive primary storage environment.

Archiving is particularly important for legal and regulatory compliance reasons. Businesses must retain and be able to recover certain electronic documents and files in their original unaltered state. With unstructured files, emails, database, and image data growing exponentially, the granular management of historical data has become a major challenge, a source of risk, and a growing expense for most organizations. Maintaining searchable historical data on primary storage is not effective, due to growing facilities and management overhead costs (security, mirroring, replication, backup) of your primary data store, despite the fact that raw disk prices have been dropping.

Backup, with the focus on the speedy handling of large blocks of data, is not designed for easy recovery, tracking, retention and security management of granular file-level information. Yes, most backup products have a file journaling system. But, the limited file tracking that most backup products offer does not include functionality for long-term and compliant retention. There is no ability to set independent file retentions; there are no systems

for ensuring security of data; there is no authentication to prove that documents are originals and have not been altered, as is often required for compliance and e-discovery. There is no long-term tracking of removable media.

Dedicated archive systems, by comparison, select and manage data at the file level and index it to be searchable by attribute and content. Because these systems operate at a granular level, the functionality for identifying and recovering specific information is built in. Moreover, archiving systems are designed with compliance in mind, so they can manage aspects of authentication, security, access control and retention.

To fully appreciate these differences, consider the following test. Try to find and recover a particular file you backed up 5 years ago. See how long it takes to locate the media your data is on. Think about the possibility that it might be stored on outdated media. Did your backup have the intelligence to suggest that active data from that era be transferred to newer media? On the flip side, you can rest assured that a highly responsive "active archive" product would give you the ability to search on attributes or contents efficiently and allow you to rapidly retrieve any files that appeared to be relevant.

Businesses can avoid problems by moving away from backup as the end-all, to professional archiving systems. In line with data lifecycle management (DLM) principles, these systems transfer data to progressively less expensive media over time, as the value and availability requirements of the underlying information change. Hardware costs are optimized and the burden taken off the primary store. And the dedicated data management features provide piece of mind when it comes to retention and compliance.

An emerging concept called protected DLM or pDLM is even more powerful. While traditional DLM relocates data to less costly storage tiers, it still requires routine backups of these secondary stores, continuing to add to storage management costs. A key difference of pDLM is that the archive is written with multiple copies, potentially to multiple media types and locations, automatically creating additional "backup copies" for disaster recovery or off-site secure storage. This greatly reduces the management burden in areas such as backup, replication and mirroring.

So many organizations who are now using backup to address their archiving requirement, will find it more cost effective, less risky and less of a management headache to use pDLM for both backup and archiving requirements.

Copyright 2007 Computer Technology Review. All Rights Reserved.