



**Data
Mobility
Group**

datamobilitygroup.com

White Paper
August 2007

Managing the Life Cycles of Backups



Years ago, we treated all data as equal. All data originated on one type of storage and stayed there until it was deleted. We now understand that not all data is created equal. Some types of data are more important than others, or accessed more frequently than others. Data Life Cycle Management (DLM), a component of Information Life Cycle Management (ILM), defines the DLM concept where data is created on one storage system, then migrated to, usually, less expensive storage systems as it ages.

Implementing tiered storage allows IT organizations to better utilize their existing storage. But, many who understand the benefits of implementing Data Life Cycle Management within their environments do not take the concept far enough into the backup space. They assume that data is backed up to some media, be it tape, disk or optical, and that is its final resting place.

However, backup data has its own lifecycle and it needs to be managed to maximize IT resources. In fact, Backup Life Cycle Management (BLM) is a critical part of the overall management of data.

Different Levels of Backup

The backup copy of data is considered by some to be the end of the line, like the caboose on a train. Data is moved from high cost, high performance disk during its creation to lower cost, lower performance disk as it ages, to SATA disk or tape for backup where it stays until it is no longer needed.

Backups must not be relegated to the “end of the line.” Backups are critical to the survival of a company. Numerous studies have shown that over half of the companies that suffered substantial unrecoverable data loss closed their doors within five years after the outage. It is critical that companies intelligently manage the backup process as an integral part of their data protection strategy.

Tiered Storage for Backup

A single tiered level of backup is not enough. For example, mission critical data must be restored within minutes to hours and requires higher performing devices; less important data can be restored in hours to days and can be placed on less expensive devices.

Backups must be directed to the proper primary or secondary tier of storage based on the application performance objectives and recovery time objectives. Assigning and managing backups to different storage tiers provide other benefits beyond meeting service level agreements. Introducing secondary disk storage, such as SATA disks, into the backup infrastructure can reduce acquisition costs, since secondary disk storage costs less than one-half of the equivalent primary disk storage. Secondary disk storage also has lower maintenance costs, reducing operating expenses.

These costs savings can be extended to remote offices/branch offices (ROBOs). Large enterprise customers may choose to back up to primary disk storage at the main data center that hosts mission-critical applications. Less critical applications that reside in remote locations may find that secondary disk storage can meet their performance objectives.

Introducing tiered disk storage into backups provides consolidation opportunities for existing tape drives and libraries. Using disk as the primary target for backups, then migrating older copies of backups to tape can reduce the number of tape drives and libraries required. This means that future tape purchases can be deferred or smaller tape libraries can be consolidated into fewer larger libraries, lowering maintenance costs and simplifying management.

Backup versus Archiving

The terms backup and archiving are used interchangeably (and incorrectly) at times to mean the same. Yet, they describe different processes with different end results.

The purpose and importance of backup has been long understood: that is, make a copy of a volume, LUN, or file that can be used to restore that data to its original condition if the data is corrupted or deleted.

Backups are designed to protect all data within a company. This data could be located in servers in the main data centers, at ROBOs or on laptops. Multiple copies or versions of backups are retained. For example, a company may retain the last three months of the weekly backups of financial data until the close of that fiscal quarter.

Archive data, on the other hand, is not a backup copy. It is data that is kept for a long period of time and is no longer "active". The purpose of archives is not to restore data in case of a corruption or accidental deletion. Archives can contain data that must be retained as dictated by government regulations. However, not all archived data is regulated data. Corporations archive data that include financial records (not under government scrutiny) or intellectual property. For example, a manufacturer may archive designs of older products that are no longer manufactured but must still be supported.

Data can be archived in several ways. Hierarchical Storage management (HSM) software is designed to scan volumes and find data has not been accessed in a long period of time. When HSM detects inactive data, the data is moved to lower cost storage and a stub file¹ is placed on the original device to specify that the data has been moved. This data, under HSM control, is considered to be "HSM archived."

Customers may want to specifically direct files to archives outside of HSM control. For example, the office of the CFO wants all quarterly files saved for ten years. This file should be sent to archival storage with a retention period of ten years. Intelligent backup software can redirect the backup file to archival storage, which means that it will not be backed up on a regular basis but retained for a specific period of time. Buyer beware! Not all backup software has the ability to redirect backups to longer term archives. And many backup software products cannot detect that the data has been migrated by HSM to lower cost storage.

In this case it is possible that the backup software is not aware of the HSM archival copy and still maintains three months of backup for the file. These backup files are no longer

¹ A stub file contains file metadata from the original file that has been moved, plus pointers that indicate where the file has been moved or migrated to. If the file is accessed after it has been moved, the data is recalled from its current location (tape, for example) rather than its original location (disk, for example).



needed. The result is storage space tied up with backups that are no longer required—an expensive proposition!

Backup data is usually not actively deleted but is overwritten. For example, if five versions of backup are retained, then version six overwrites version one, and subsequently, version seven overwrites version two. Unlike backup data, archive data should be actively deleted when its retention period has expired. Backups usually store several copies or versions of the data; archives store only one. Backups and archives have different recovery time objectives (RTOs) which define the time required to restore or retrieve the data. A backup of the customer order database, which contains all outstanding orders, may have an RTO of less than one hour because an outage greater than one hour can cost thousands of dollars in delayed orders. On the other hand, a project involved with documenting the history of the company may require access to old financial records. However, retrieval time of one to two days may be very acceptable.

It is important that backups and archives are logically separated so they can be managed properly.

Managing Backups throughout Their Life Cycle

The amount of data that IT administrators must manage continues to grow every year and that trend shows no signs of changing. As data grows, so does the number and amount of backups that exist. Treating all backups the same, that is, placing all backups on the same device type, is a simple solution. However, it can be very costly.

All data is not created equal. A major intent of Data Life Cycle Management is to ensure that data is placed on the most appropriate device at specific points in its life cycle. This same concept is also required to manage backups. That is, not all backups are equal and need to be placed on the appropriate device at specific points in its life cycle.

Consider the example of an important customer order database application deployed across various remote office/branch office systems (say a retail POS system, or medical clinic management system, public library system, police precinct system, hotel management system, golf/country club system, etc). If the database is corrupted during the day, it will be restored from last night's backup. Last night's backup should be stored on the fastest performing backup device available to meet the recovery time objectives of the customer order database. If the latest file copy is retained on a high performance device at the remote/branch office, the quickest recovery will be provided by that device rather than from a device located at a central data center. This will be particularly true if there is a lot of data to be recovered or if the communications link to the remote/branch office has been temporarily disrupted. However, older versions of the remote/branch office backups are also retained and should be moved to slower, lower cost backup devices located at a centralized, secure data center to save storage costs and to ensure that data management policies are extended consistently across all enterprise data. This practice will be of particular interest to those IT organizations that have instituted a policy of removing tapes from remote/branch offices.

Backup software, then, manages all copies of the backup and moves the backups to the less expensive storage as the backups age, and become less critical, but still “important” to the restore process.

What about Archives?

An archive, like backups, must also be managed throughout its life cycle. For example, a file that contains financial data is updated throughout the quarter. Backups are taken of this file every day and stored on backup media. Older versions of the backup are migrated to slower, less expensive backup storage as the backup ages.

At the end of the fiscal quarter, data within this file is used as input to the corporate balance sheet. However, the data within this file is no longer updated (or at least it is not supposed to change!) and the file becomes inactive. Eventually HSM will determine that this file has become inactive and it will be moved to lower cost storage as part of the HSM archive. Or, customers might determine that the file should be archived outside of HSM control.

Refreshing the Technology

Moving archives from more expensive, higher performance storage devices to less expensive storage devices saves money. But, it also provides the ability to ensure that archives reside on current technology.

A company may choose to keep archives of quarterly financial data for ten years. However, it is policy within the data center to replace disk storage every three years and tape drives every seven years. Keeping ten year old archives on the same devices that they were originally created is inefficient and costly to maintain. IT must keep older tape drives and disk drives around to support old versions of backup. Older equipment can have higher failure rates and require more expensive maintenance than current products. Archives created ten years ago on older versions of software may no longer be readable. There can be problems with the age of the media, or the format of the archive. The software that created that archive may no longer be available. Or, a much newer version of that software is not backward compatible to the version from ten years ago.

The result: a great amount of archival storage that cannot be retrieved when it is needed. If the corporation cannot retrieve data that it is required to save to meet regulatory requirements, it may become subject to heavy financial penalties (and negative publicity!)

Software should provide the ability to refresh technology – that is, move the archives from older equipment to newer equipment after a specific period of time has elapsed. This ensures that archives are never stored on outdated equipment. It ensures that archives can be located and can still be accessed and be ‘read’ from older equipment with newer levels of software. It takes advantage of lowered cost and more reliable storage devices. In fact, technology refresh should be a requirement of any archival software product. It is a part of managing the life cycle of the data.



Requirements for Backup Software

Many different backup software products are available today. Customers evaluating new software solutions need to consider the usual factors when evaluating the software. These include factors such as:

- Cost
- Reputation and reliability of software vendor
- Operating systems supported
- Ease of installation and ease of use
- Performance
- Future roadmap plans

However, meeting those objectives is not enough. Backup software needs to also manage backup data throughout its lifecycle wherever it is located. These factors include: managing all software copies, refreshing technology, archive processing and compliance, remote offices, and small businesses.

Manage all copies

The software must maintain the status of the current backup version, but also have the ability to move older versions to less expensive storage to reduce overall storage costs. Current versions of backup can reside on SATA disks, for example, while previous versions are on tape. The ability to move backup versions to different devices provides the ability to refresh the technology.

Technology Refresh

Articles have been written about the shelf life of tape, or how storing data for long periods of time on disk is “safer” than storing it on tape. However, the shelf life of the media is only one concern about data that has been stored for very long periods of time. A backup file that was stored on disk in 2004 may not be readable in 2007 if the backup software is no longer available to interpret the proprietary format. The problem is more acute with archived files that can be ten or more years old. Software should have the ability to ‘read’ into older data and rewrite the data on newer media after several years. This refreshing of the technology tests the data and media for ‘readability’ after a shelf life of several years. In addition, old, outdated hardware does not need to be kept, ‘in case’ the backup or archive must be read.

The Archive Process and Compliance

Data is stored in archives until a particular period of time elapses or a particular event occurs. For example, corporate emails that discuss financial transactions must be retained

for seven years. Here, a corporate policy or government regulation dictates how long these emails must be retained and when they should be deleted. Other data needs to be kept until a certain event occurs. For example, personnel records are kept for two years after an employee has left the company. At the end of that period, the archive data should be deleted. Corporate attorneys may insist that the data **MUST** be deleted after the event has occurred or the time period has elapsed. Isolated archive software that is not integrated with backup software can delete the data stored in its archive. However, are the backup copies also deleted? They are if the backup software is integrated with the archival process and is HSM-aware. Software that supports Backup Life Cycle Management and is integrated with archival software ensures that **ALL** copies of data are deleted. Backup software that is not archival-aware can retain data beyond time or event-specified retention periods. Saving data beyond its expiration date is costly in terms of storage space and can be very costly in terms of legal liabilities.

Remote Offices

Many companies have remote office/branch offices (ROBOs) scattered throughout countries or across continents. The result of mergers or acquisitions has forced centralized IT organizations to manage backups across numerous locations. Some backup software might work well to manage backups within a main data center; however, this same software might not provide adequate support for managing remote backups. Managing local and remote backups is not restricted to large corporations. Small companies, such as real estate and law firms, retail chains, or medical clinics might have several satellite offices that must be backed up to protect their data. It is important that backup software for both large and small organizations be able to back up data at all locations in a consistent manner often without the presence of IT expertise at each location. The data must be able to be recovered at the remote office (which provides faster recovery speeds) and at the main data center (for disaster recovery) in the same way to ensure that data can be recovered easily from an outage in either remote or local locations. The backup software must also be capable of managing all sites from one central location to reduce management complexity and cost.

Small Businesses

Small businesses with very limited IT resources can find the task of designing and implementing a backup and archive infrastructure challenging. However, these businesses do not have to hire more staff to implement a new backup scheme. They can work with managed service providers, which have expertise in backup and archive software design, implementation and management. These service providers can establish policies that will ensure all data is properly protected in accordance with regulations, should they apply.

Backup also has a Life Cycle

There has been a great deal of discussion in the press about Information Life Cycle Management or more aptly named, Data Life Cycle Management. We now accept that not all data is equal and should not be treated as such. Data should be matched to the



most appropriate storage device during its life cycle. This is also true of backup data. Not all backup data is created equal and it should not be treated as such. Backup data must be managed throughout its life cycle to provide the best data protection at the most affordable cost. If your backup vendor does not understand that, then it is time to find another vendor. ■

Asigra is a small, profitable and rapidly growing company located in Toronto, Canada. The company is not new to the backup space – actually it has been in business since 1986. However, the company has not been a ‘household’ name in backups since it previously sold its software to Storage Service Providers (SSP) that used Asigra’s Televaulting software re-branded to their own service name to back up its clients.

Asigra Televaulting software is a mature enterprise-class agent-less, backup solution with a simple capacity-based pricing model that should appeal to many customers, particularly those that have multiple remote locations.

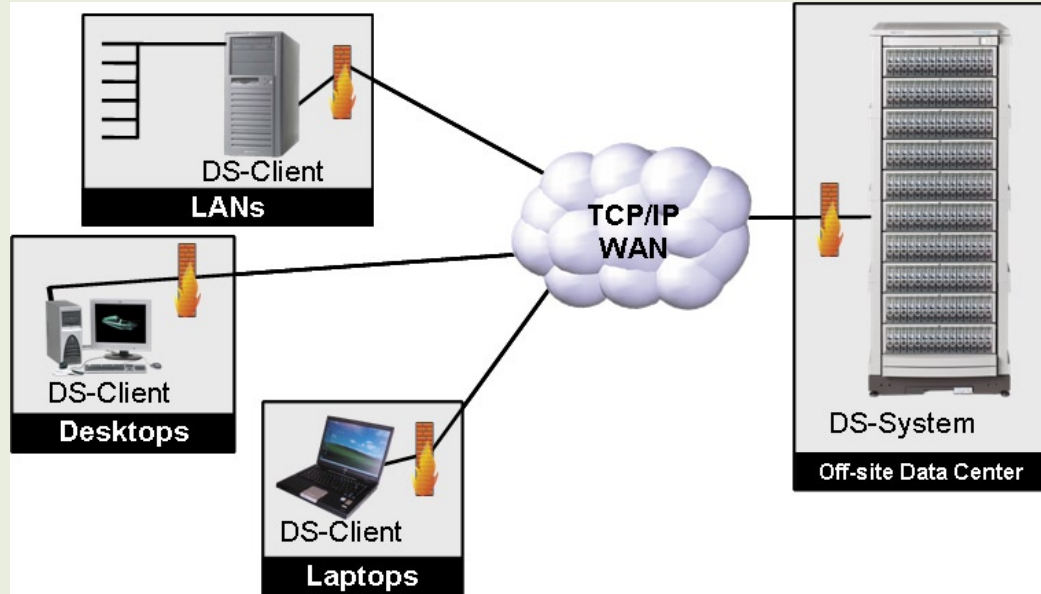
Asigra understands that backups have a life cycle. In fact, Asigra coined the term Backup Lifecycle Management or BLM.

Televaulting software

Asigra’s solution has two parts. The DS-Client sits at the remote site while the DS-System resides in the central data center. A very important feature – no agents are required on any servers at the remote or central locations. The DS-Client supports numerous flavors of UNIX, Windows, Novell, VMware, Macintosh and AS/400. When it is first implemented, the DS-Client backs up all of the data at the remote site and sends a compressed, encrypted copy of that data to the central location. Data is encrypted “in-flight” and “at-rest”. The most current version of the backup can be stored locally on the LAN where the DS-Client is installed on one of the machines.

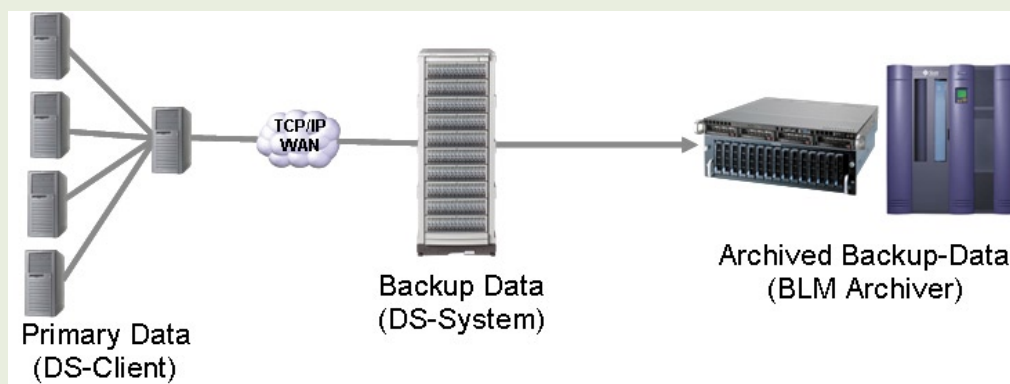
At the next backup interval, the DS-Client sends only the changes that have occurred to the DS-System, saving time and bandwidth. Another WAN optimization technique used is common file elimination across all locations.

Two levels of data protection are provided. If the data is corrupted or accidentally deleted, it can be restored from the local DS-Client storage. The local backup is compressed (using a compression method designed for very high-speed compression/decompression) and not encrypted to speed up the restore process. However, if the remote site has an outage, then the data can be restored from the centralized DS-System. Data sent from the central site is encrypted and compressed to save bandwidth and provide data security.



Because Asigra's software was initially designed to support numerous remote customers for Storage Service Providers, this software can consistently and efficiently protect data at both remote and central locations.

One of the distinguishing features of the Televaulting software is its management of the backup life cycles. Not only does Asigra understand how to manage the lifecycle of backups, the company coined the term Backup Lifecycle management (BLM). Asigra recognizes that there are multiple tiers of backup of active data. Mission critical data requires more frequent backups on higher performance devices. Less critical backups are relegated to less expensive, lower performance devices. Older versions of backup are moved to slower devices to save costs.



However, not all data is active. Inactive data, or ‘stale’ data is stored in the Televaulting ‘vault’ at the datacenter or service provider’s site. Stale data can be data that was intentionally deleted by the end user and a vaulted copy is retained in case the deletion was accidentally. It can be an older version or generation of a backup. It can be the source data that has been marked by HSM as ready for HSM archive. Or it can intentionally be ‘pushed’ to the vault to be retained in accordance with corporate policy or regulations. Televaulting can determine if it has received multiple requests to vault the same data and eliminates duplicates. Asigra understands that this vaulted data may be retained for long periods of time and saves a copy of the current DS-Client with the data in one restorable package. Now, customers can recover the data years later, when requested by government agencies, satisfying compliance regulations. Since the backup and archive software are integrated, the Televaulting solution ensures that ALL copies of the data, whether residing in backups or vaults, will be deleted as dictated by policies.

Televaulting software was originally designed to support Managed Service Providers and many of the features built in to accommodate the MSPs are available to enterprise customers. For example, an integrated billing system can charge back DS-Clients. Service Level Agreement (SLA) monitoring automatically credits end users if defined SLA objectives are not met. All DS-Clients can be remotely managed from one central site. Self-healing technology can detect and fix corrupted files. And the pricing structure is very simple. DS-Client software is free. Customers are only charged for the storage capacity utilized. Enterprises are also able to take advantage of these same features if they restructuring their storage services from a cost center to a profit center.

Asigra may not be a ‘household’ name yet in the backup space because the Asigra software, which is installed in thousands of customer’s sites, has been re-branded by service providers. However, the company now sells its software to enterprise customers looking for an intelligent solution to back up remote locations. Asigra is not a start up company, but well established with mature software that has flourished under the critical eyes of major storage service providers. Asigra leads the technology curve for Backup Lifecycle Management. ■

Copyright© 2002-2007 Data Mobility Group, LLC. All Rights Reserved. Reproduction of this publication without prior written permission is forbidden. Data Mobility Group believes the statements contained herein are based on accurate and reliable information. However, because information is provided to Data Mobility Group from various sources, we cannot warrant that this publication is complete and error-free. Data Mobility Group disclaims all implied warranties, including warranties of merchantability or fitness for a particular purpose. Data Mobility Group shall have no liability for any direct, incidental, special or consequential damages or lost profits. The opinions expressed herein are subject to change without notice.

Research sponsored by Asigra . All other brands, products, or service names are or may be trademarks, or service marks of, and are used to identify, products or services of their respective owners.