StorTrends
storage without boundaries

American Megatrends

Whitepaper

05/08/2008

Information Lifecycle Management: Optimizing Storage Performance and Protection Relevant to Data Value

American Megatrends, Inc.

5555 Oakbrook Parkway, Suite 200

Norcross, GA 30093

TRADEMARK AND COPYRIGHT ACKNOWLEDGMENTS

FOR ADDITIONAL INFORMATION

Call American Megatrends at 1-800-246-8600 for additional information. You can also visit us online at ami.com.

LIMITATIONS OF LIABILITY

In no event shall American Megatrends be held liable for any loss, expenses, or damages of any kind whatsoever, whether direct, indirect, incidental, or consequential, arising from the design or use of this product or the support materials provided with the product.

LIMITED WARRANTY

No warranties are made, either express or implied, with regard to the contents of this work, its merchantability, or fitness for a particular use. American Megatrends assumes no responsibility for errors and omissions or for the uses made of the material contained herein or reader decisions based on such use.

DISCLAIMER: Although efforts have been made to assure the accuracy of the information contained here, American Megatrends expressly disclaims liability for any error in this information, and for damages, whether direct, indirect, special, exemplary, consequential or otherwise, that may result from such error, including but not limited to the loss of profits resulting from the use or misuse of the information contained herein (even if American Megatrends has been advised of the possibility of such damages). Any questions or comments regarding this document or its contents should be addressed to American Megatrends at the address shown on the back cover of this document.

American Megatrends provides this publication "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability or fitness for a specific purpose. Some states do not allow disclaimer of express or implied warranties or the limitation or exclusion of liability for indirect, special, exemplary, incidental or consequential damages in certain transactions; therefore, this statement may not apply to you. Also, you may have other rights that vary from jurisdiction to jurisdiction. This publication could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. American Megatrends may make improvements and/or revisions in the product(s) and/or the program(s) described in this publication at any time.

Original Release: 05/25/2007
First Revision: 02/10/2008
Second Revision: 05/08/2008

Table of Contents

Introduction

Not many people would disagree that after cash, data is the lifeblood of a business, and if it is disturbed, so is the ability of the business to function properly. However, many administrators and planners fail to recognize the fact that the value of data to an organization is not constant. In fact, the value of data decreases over time, as it loses its relevance, freshness, and "popularity". One question that administrators should be asking themselves is: why should data that is decreasing in value remain in expensive front line storage, subject to the same backup, replication, and recovery policies and procedures as key data? Would it not be useful to have a system or methodology in place for analyzing and tracking data freshness, so that storage space could be made free for more fresh and relevant data, and time / bandwidth consuming data protection policies be relaxed as data loses its value? It is here that the concept of Information Lifecycle Management, or ILM, steps into this gap -to try and address and resolve some of these questions.

First, let us set the stage for this discussion, and ask: what does ILM mean? ILM is a concept that encompasses the discovery, classification, analysis, and maintenance of data, across the entire period of its useful life. It adds structure and context to data, marking the transition from information to data. ILM is a part of the larger concept of Business Continuity Planning, but has become increasingly prominent in the storage arena in recent years thanks to several factors, including advancements in data storage management techniques and the technology that underpins it, and evolution in the storage environment, including:

- Coexistence of Fibre Channel and iSCSI (IP-Storage) in the data center
- SAS and SATA storage coexisting in storage systems
- Storage consolidation practices, for reducing the use of solitary "islands of data" in direct attached storage (DAS)
- Regulatory requirements for data archiving and recall (SOX, etc.)

Though many vendors offer ILM services or modules as a part of their products, ILM is above all a concept or a strategy, rather than a product. However, for a practical explanation of what the concept embodies, we can safely generalize that many implementations of ILM encompass such components as:

- Database Management
- Storage System Performance and Monitoring
- Storage Capacity Planning and Management
- Business Controls for Data Degradation and EOL

In an ILM configuration, data is analyzed for its value, and stored accordingly. At the peak of its popularity, it is stored in the fastest, most responsive top-tier storage on hand and subject to the most stringent replication and backup controls. Since the ILM system is constantly monitoring the data's value in comparison to other data, as it loses value, it is migrated down the chain to less expensive, less powerful storage, where it may not be accessed as frequently, or protected as carefully. In the final stage, it is migrated out of the storage system completely. Data of the lowest value is either purged from the system or transferred to other media (e.g., written to tape and delivered to offsite storage) depending upon the organization's policy and regulatory requirements for data end-of-life.

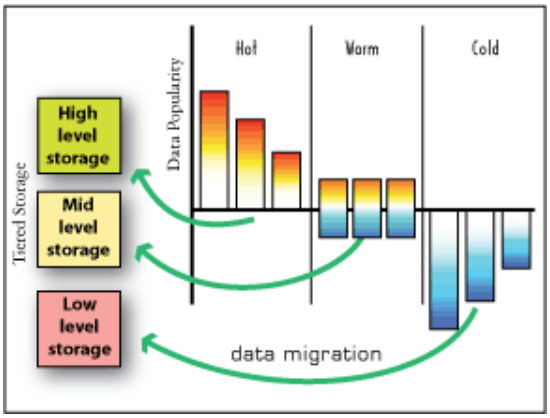

Figure 1: Data Stored and Migrated According to "Freshness"

Reasons for ILM

Having examined how an ILM system can be implemented, we should next look more closely at the reasons why more and more organizations are accepting the need for a comprehensive ILM strategy. While the needs of each organization can be wildly different, typically ILM solutions are intended to solve the problems of rapid data growth, cost (initial and operating costs) of data storage, changes in data accessibility and freshness, and increasingly stringent data backup and retrieval requirements.

In practical terms, the goals of ILM are to manage data growth and accessibility, reduce cost, improve recovery, and reduce hardware and software risk/exposure by introducing new technologies into the storage mix. We will next take a closer look at each of these different problems, to understand why ILM is needed to satisfy each of these practical concerns.

Exponential Growth of Data

With data growth averaging near 80% to 100% per year, managing storage effectively has become a challenging uphill task. Storage administrators face limited budgets, and are charged with not only expanding capacity by purchasing new hardware wisely to meet projected storage needs, but also optimize the use of existing capacity, in order to maximize the investment in current storage hardware. Moreover, any changes or additions need to be considered carefully, as the downstream effects of new hardware are often unforeseen, and can quickly wipe out any short term cost gains.

Data Accessibility / Freshness

As mentioned at the beginning of this paper, data does not have a constant value; rather that value is changing, whether it is due to time, relevance, security, or popularity. Policies and procedures must therefore be set in place to continuously shift and monitor the location (and therefore the accessibility) of data so that information that is in highest demand is in the most accessible location.

Cost (TOC) Issues

The overall cost of a storage system is measured not just in the initial price paid for the hardware and its commissioning. While there is a fixed dollar-per-gigabyte cost in acquiring the hardware, the total operating cost (TOC) includes maintenance, power and cooling expenses, together with the cost to staff and train administrators. As storage arrays grow, power usage (for server operation and cooling) is just one factor that has an enormous impact on the TOC of a storage solution.

If less expensive solutions are available, administrators should by all means devise a careful plan to incorporate these components, with some restrictions. When possible, additional storage technology should be adopted that does not require significant investment of time and resources to learn its operation. New solutions that are more power or space efficient should be integrated into the array. If there are promising, more economical technologies or vendors emerging into the market, sometimes the lower tier of storage can be a good proving ground, by testing smaller, less expensive solutions first before jumping on a new bandwagon.

Ability to Protect and Recover Lost Data

Because key data has to be protected against loss to ensure business continuity, the term Continuous Data Protection has come into being. It describes a scheme of ensuring data survival in the face of disasters such as power/network outages and natural catastrophes, and incorporates techniques such as backups, data snapshots and remote replication to do so. To add to the challenges surrounding data protection, regulatory requirements for the preservation and archiving of several types of corporate data continue to mount. Data of a particularly sensitive or critical nature must be available for recall within clearly established time limits if circumstances demand it, and kept secure as well. Therefore a successful ILM implementation integrates well with the backup solution and recovery solution of an organization along several touch points. ILM dictates that as items age they can be taken offline completely and migrated to tape storage, for example, yet some data still must be available for recall, even at this point. Since only a percentage of data has to be protected in the same manner, the ILM solution must be flexible enough to manage varying CDP requirements.

The Role of ILM in the Larger Picture of Storage Management

As noted above, the ILM policies that an organization devises must be *flexible,* in order to integrate itself with the other competing priorities such as data protection, cost control, capacity expansion, and the like. To get a better sense of this interplay, we will look at how ILM functions within the larger perspective of the overall storage management architecture.

From a wider storage management architecture perspective, ILM has an integrated relationship with the practices of **Data Tiering**, **Disaster Recovery Planning**, and **System Resource Management**. **Data Tiering** is concerned with organizing data into a hierarchy based on its value, while **Disaster Recovery Planning** (DR) addresses a storage system's ability to restore lost data as quickly and completely as possible. **System Resource Management** (SRM) encompasses a whole range of diverse storage system status and performance analysis metrics to ensure that storage resources are optimized via a number of manual and automatic controls. These four elements feed into and play off each other, as changes in one element influence the other three in a cyclical manner.

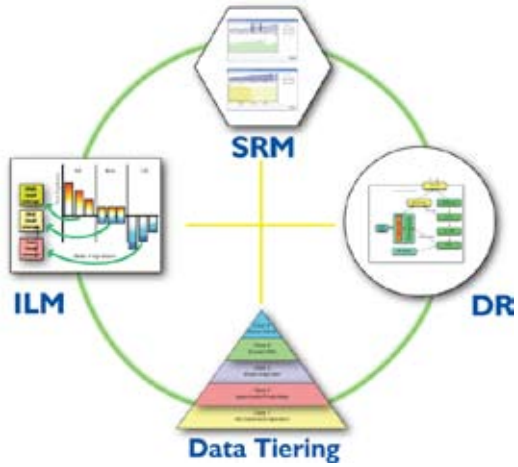

Figure 2: ILM, Data Tiering, DR,and SRM Interrelationship

Let's next look more closely at the tight interaction that exists between ILM, Data Tiering, Disaster Recovery, and SRM.

Data Tiering

Data tiering refers to the separation of data into different classes according to its value. If we consider that data can generally be divided into groups of high, medium, and low value, then we can understand how and where this data should be stored, migrated, and protected. The illustration below shows us a sample stratification of data into five classes based on its worth to the organization:

Figure 3: Data Value Hierarchy

From this diagram, it is clear that storage should be tiered to reflect the hierarchy of data. For example, data in Classes 5 through 3 would be kept on the a redundant, high performance array, while data in classes 2 and 1 would most likely be stored on a mid-range storage array with lower performance, or possibly even on tape.

Tiered Storage

Tiered storage can work in several different ways. From the physical perspective, hardware power and sophistication cascades downward, migrating data from more expensive to less expensive storage, as the data loses its freshness and relevance. From a

data protection perspective, it could mean decreasing the amount of coverage or the frequency of snapshots taken of this data set as it moves down the chain.

Note, however, that the data hierarchy shown above does not necessary have any connection with access. In other words, data that is not accessed frequently could in fact be mission critical (such as data containing system configuration instructions), while data that is not essential to operation is often frequently accessed (corporate Power Point slides) .

Therefore a combination of objective and subjective measures are necessary when evaluating the value of data in this hierarchy. The value of data could be based on the level of protection it requires, the frequency of its use or access, or whether or not it is required to be maintained for archival or regulatory purposes, and ILM must be sensitive to these sliding, objective and subjective judgments about the value of data in storage.

Disaster Recovery Planning

Once data is separated into these separate hierarchies, its recovery priority can be established. Data at the top of the pyramid should receive the highest level of protection, by being made redundant and marked for immediate recovery, for example. Since lower-end data is not nearly as critical to the organization, it may be determined that days or weeks are an acceptable amount of time before data is restored after a disaster. In fact, low-level data may not even be restored, but simply preserved for regulatory or archival requirements. As ILM migrates data, the shift in disaster recovery priority has to be synchronized with the data's new location and status.

System Resource Management (SRM)

System Resource Management (SRM) is a term used to describe the mechanism and process of data collection, data usage analysis, and trending in a storage appliance. In brief, SRM performs the following tasks in a storage system:

- Discovery of storage nodes
- Storage provisioning and capacity management
- Performance monitoring and characterization
- Backup and recovery management
- Workflow management
- Presentation and reporting mechanisms

Depending on the hardware and software involved, SRM can be anything from an automatic process managed on homogeneous hardware to one that is completely manual and left to the administrator. However, at a minimum the SRM mechanism should provide information on storage capacity and data access patterns, so that informed ILM decisions can be made based on this relevant information.

Having explained the concepts of Data Tiering, Disaster Recovery, and System Resource Management, it is easy to see their interdependence with ILM. Since ILM analyzes and sets the policies for moving data, migrating data as it changes in value from high to medium to low along the data hierarchy, data relocations made as a result of ILM policies cause shifts in the recovery priorities built around the data.

Conclusion

When expanding capacity, one of the biggest challenges is balancing cost and performance with complexity. A good ILM plan should balance the introduction of more economical, or newer technology (like SATA hard drives) with warnings against making a storage network more complex, with additional vendors, standards, and protocols to support. A good maxim to keep in mind is that the more complex a system becomes, the more expensive it is likely to be to deploy, integrate, optimize, and manage. Rather than thinking simply in terms of adding new, cheaper components, administrators should carefully consider how to use existing capacity more effectively, by improving bandwidth, capacity, and reliability.

We have also seen how ILM has to fit neatly into the larger storage solution, since there are dependent priorities at work. An ILM solution must therefore incorporate and leverage the tight interaction that exists between ILM, Data Tiering, Disaster Recovery, and SRM to be effective. A measured approach such as this will make all the difference in achieving the goals set out for the Information Lifecycle Management solution.

As mentioned at the beginning of this paper, due to its prominence as a hot topic in the storage industry, a wide-cross section of vendors offer ILM solutions, either as standalone products or incorporated into a larger storage management solution. Having first looked at how ILM is handled in general, let's now take a closer look at how ILM is implemented in StorTrends® iTX data storage software from American Megatrends.

ILM and the StorTrends® iTX Software Stack

Enhanced System Resource Management (SRM) Features

With StorTrends iTX, AMI introduced a leading-edge Storage Resource Management (SRM) module with a wide spectrum of manageability offered through a "single pane of glass". This module is a very powerful visual tool that imparts real-time manageability and enable administrators to proactively track, manage and plan capacity.

Figure 4: Storage Resource Analysis Reporting in iTX 2.75

To further boost the capability for capacity tracking, StorTrends iTX features a dynamic Information Lifecycle Management (ILM) module. This lightweight, efficient service analyzes and tracks various performance and usage attributes of the provisioned capacity. For example, users can use the ILM interface to see how far back in time a certain storage region was accessed, or to determine statistically what the best time of the day is for performing various background chores.

The ILM module in StorTrends iTX introduces a very powerful workflow management engine that allows certain tasks to be automated and scheduled in the background at the most appropriate time. As an example, unused or stale data can be automatically compacted and pushed out to cheaper storage at a statistically-determined "least busy" period.

Some advantages of the ILM implementation in StorTrends iTX include:

- Granular block level tracking of access patterns
- Bi-directional migration between tiers
- Efficiency of migration
- Automatic classification by type of storage, RPM, size of drives, RAID level and background activity
- Visual tools to monitor and analyze
- Optimizes and improves performance
- Reduces expenditure on disks by supporting multiple levels of tiered storage
- Improves continuity by conserving space for snapshots
- Built around an extensible framework that can integrate with 3rd party add-ons
- Storage unit discovery
- Capacity trending and analysis tools
- Statistical and instantaneous performance monitoring
- Backup and recovery management
- Tiered storage management
- Presentation and reporting tools

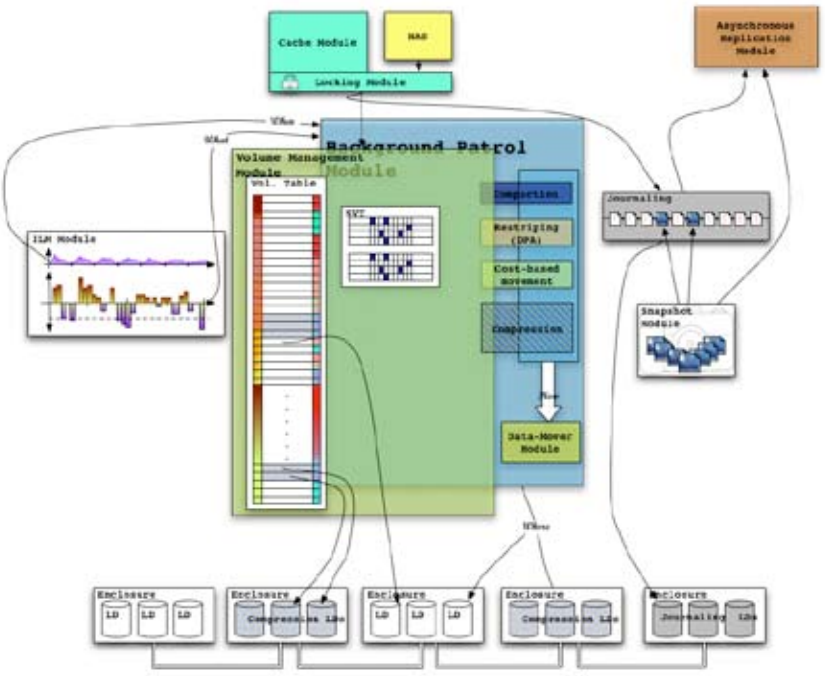

Figure 5: StorTrends iTX Architecture Diagram

Why AMI?

AMI offers a wide array of disaster recovery and high availability solutions for your business needs. We provide services that range from storage needs analysis to the design and implementation of a custom disaster recovery solution. We can help your business plan for when things are at their worst while reducing the cost and complexity of your storage environment. For more information on AMI StorTrends solutions, visit www.StorTrends.com, email to sales@ami.com, or call (800) 1-U-Buy-AMI.

American Megatrends

Whitepaper

American Megatrends

American Megatrends Inc.
5555 Oakbrook Parkway, Suite 200
Norcross, GA 30093 | t: 770.246.8600
Sales & Product Information
sales@ami.com | t: 800.828.9264
Technical Support
support@ami.com | t: 770.246.8645
www.ami.com