

INFORMATION LIFECYCLE MANAGEMENT

Can you afford to ignore this topic?

- Do you have less data now than you did last year?
- If you ignore the issue, will it go away?
- If you cross your fingers and close your eyes, will it get easier to manage your storage and your backup/restore load as data volumes grow over time?
- Are your internal staff and managers, and also external parties (auditors, industry bodies, government/regulators) getting any less concerned with data lifecycles?
- Is it easier to untangle a mess later if you ignore it today?
- As data volumes grow over time, will the backup window just get longer as if by magic, and give you the time you need to sort it out each night?
- Can't you just throw money at the problem to resolve it?

Definition of ILM:

The Storage Networking Industry Association (SNIA) defines ILM as “the policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost-effective IT infrastructure from the time information is conceived through its final disposition. Information is aligned with business processes through management of policies and service levels associated with applications, metadata, information and data.”

Put simply, it is the idea of managing an enterprise's data throughout its useful life.

Why ILM?

Economical, high-capacity media will be used to facilitate backup and rapid restore, leaving higher performance storage for transactional data and tape archiving or other media for archiving. Soaring data growth, high cost of managing data & increasing compliance make ILM a must do on every savvy IT Manager's list.

- The average size of corporate storage per employee has risen from less than 100 megabytes in 1993 to more than 3 gigabytes in 2003, according to META Group Inc. Furthermore, a 5,000-user organisation has at least 15 terabytes of data to manage—a number that will jump to 80 terabytes by 2008.
- Storage is no longer just about file servers and databases. It now includes e-mails, Web content, images and other data with different storage characteristics. Some information has high transaction & high importance rates; other data requires quick read-write of files but no overwriting. E-mail for example, is accessed very often in the first 30 days of life, but still needs to be accessed very quickly but infrequently later in its life cycle.

By managing the information life cycle, you realise hardware cost savings, administrative cost savings and, more importantly, better data protection.

Developing and executing an ILM strategy is not easy. Analysts say it is a five- to seven-year endeavour, & no user has yet implemented a full-blown enterprise-wide ILM strategy. But with all the data migration, storage resource management and SAN management software on the market, as well as the policy engines, document management systems and archival tools for databases, e-mail and files, the technology pieces to support full-blown ILM are still evolving and growing in sophistication.

INFORMATION LIFECYCLE MANAGEMENT (ILM) STRATEGIES FOR YOUR STORAGE

3 step strategy to executing ILM

ILM execution can be very simplistically split into three stages

- Classifying information according to your company policy
- Storing it in a tiered structure
- Moving the information among these tiers

Classification of information according to your company policy

The key to Information Life Cycle Management is classifying data by its age and business value. Doing the upfront work of developing an enterprise-wide ILM strategy is a challenging job. You may need professionals for this strategy. Although it might be tempting to skip this step - which analysts admit is painful and tedious - it is a crucial one.

Ray Paquet, an analyst at Gartner, advises approaching your data classification as three separate categories: unstructured data, such as files; semi-structured data, such as e-mail; and structured data, such as databases. "*These are three distinct problems that need three distinct technologies and three distinct tools,*" he says. Define the data needs in terms of reliability, disaster recovery, backup, retention, availability and performance, and then map those criteria to your storage infrastructure.

There are tools to automate parts of the process. This is where our data profiling tools pay off. You can get a clear picture of your data, - location, age, usage, duplication etc. Globalnet Solutions Australia provides software solutions that can identify and map the relationship between physical and logical devices, including servers, SANs and storage. Customers have recovered more than 40% of their disk space, reduced backup and restore time by 18 hours and improved storage resource utilisation saving \$150,000.

Talk to Globalnet Solutions Australia if you would like a data profiling exercise done for you.

But this can't be a one-time exercise. Data values change, so you need to create business policies that support the movement of data onto higher or lower storage tiers as needed. You also need to maintain a central repository of metadata, or "data about your data." Policy engines, discovery tools and other systems are being developed to recommend what data should be moved and then automate data movement. We advocate an initial assessment and ongoing monitoring.

Storing it in a tiered structure

Tiered storage is the assignment of different categories of data to different types of storage media. Essentially, ILM systems and processes will manage and organise information according to a hierarchy as follows:

- **Online:** High performance-High availability enterprise arrays – for high-performance applications, Data can be accessed automatically and instantaneously – such as within milliseconds. to access it. Eg enterprise RAID (mirrored and replicated) for mission-critical transaction data, Fibre channel, SCSI/SAS
- **Nearline:** midrange disk or NAS for other active data– offering good performance, capacity and rapid access for streaming media, disk-to-disk backup and staging. Data can be accessed automatically but at a slower speed, such as a second or more. Eg SATA. (Serial ATA storage is technically like the disk interfaces used in laptops. It's much cheaper than hard-disk storage--about the same price as tape systems--but easier to pull information from than tapes.)
- **Offline:** Lower performance, slower performance archival storage – for storage, removable media, and archival storage. Data requires human intervention (i.e., finding and loading a tape)

Evaluate each tier according to several performance factors; the performance of the storage should then be matched to the business value of the information it contains as well as the information's availability requirements. Economical, high-capacity media will be used to facilitate backup and rapid restore, leaving higher performance storage for transactional data and tape archiving or other media for archiving.

A recent IDC survey shows that most large storage environments (defined as 5 TB or more) have yet to adopt tiered storage. About 25 percent have implemented it, another 5 percent are exploring it and the rest have, so far, left it alone.

Globalnet Solutions Australia can work with you towards simplifying your storage & achieving your ILM goal.

Moving the information among storage tiers -SRM

Movement based on data value, business process needs, user access needs and retention/deletion requirements is typically managed by specialist software. The job of deciding what to move when and where is achieved by using Storage Resource Management (SRM) and similar tools.

SRM manages the storage infrastructure—hardware, switches, files and file systems—and optimises the efficiency with which available storage space is used. The goal is to automate tasks—data collection, backup and recovery, SAN performance analysis, virtualisation, provisioning, forecasting and so on. This reduces the labour and complexity involved in supporting multiple tiers of storage service. More importantly, these tools effectively bind tiers to applications enforcing compliant service delivery.

SRM is essentially a prerequisite for ILM.

Here are a few features you should look for in SRM software

- **Enterprise wide.** The best option is to use a common workflow-based policy engine that tracks and moves data across multiple applications across the corporation. Ideally, whether the data comes in from SAP or an e-mail database, it should have the same policy engine.
- **Reporting.** You should be able to monitor the environment from end-to-end and manage multiple assets and processes from a centralised dashboard. Can your SRM analyse events to provide intelligence useful in understanding performance issues, asset utilization, provisioning, and other planning functions? Or report on how storage assets are utilized by applications and departments?
- **Optimised Storage.** According to a recent Gartner survey, only 60% of storage is utilised, the rest is wasted space. Setting policies & thresholds to determine when data has reached a critical stage & needs to be moved to second or third tier storage, or deleted means you can reclaim wasted space.
- **Automated movement.** Moving data among the tiers can be done manually. Automatic and transparent movement of data between production storage media and secondary storage media according to business data needs would be ideal. Ask for transparent implementation of your rules (to move data between storage tiers so data is stored on the most appropriate media), automatic policy driven migration and recalls

- **Path aware.** Path management, as most storage vendors call it, is key. Some information, such as legally required records, maintains high value for its entire life, but is seldom going to be accessed. Other information, such as a lot of accounting information, declines in value slowly, although the uses to which it is put changes as it ages. Often it stops being current working information and becomes the basis for reports and analysis, the kind of thing better kept in a data warehouse. Most data shows a steep decline in value and accesses over the first 60 days after it is recorded, settling into a seldom-used limbo after about 90 days. Obviously this can lead to a pretty complicated data retrieval situation, because data that will age at different rates is ideally kept on different storage cycles. The ability to find a particular piece of data no matter what file, volume or disk or tape it happens to be on is important.

There is a lot more to choosing an SRM - we haven't even touched disaster recovery techniques yet! This is something that Globalnet Solutions Australia can review in accordance your individual requirements.

