

Understanding JIC Data and Information Lifecycle Management

*by Alan Stuart
Chief Strategist and
Business Line Executive
IBM Data Retention Solutions*



Overview

The management and protection of information is one of the most important tasks facing IT organizations today. Information's value to the business can fluctuate based on the type of data and the point in its lifespan. Critical data needs to be well protected and readily accessible, whereas less critical data can tolerate longer access (latency) and recovery times. The value of employing Information Lifecycle Management (ILM) technology is in correctly managing information based on its value to the business at a given moment in its lifecycle. Customer requirements for ILM typically center on achieving the lowest total cost of ownership (TCO), providing the appropriate latency and recovery consistent with the data's value to the organization. It is important to note that ILM is not a storage centric methodology. Rather, it encompasses comprehensive and efficient content and records management, which are leveraged to meet corporate and regulatory data retention and protection governance requirements. Data becomes information when its content can be used within a context.

IBM's History of Delivering the Foundation for ILM

Before the term information lifecycle management was invented, IBM was delivering solutions for integrated information management and storage to manage the lifecycle of data from creation to disposal. For example, IBM introduced hierarchical storage management (HSM) software for mainframes (DFSMS) in 1974. Over the years, IBM has worked with its vast customer install base to refine the original HSM concepts. One particularly important addition to HSM was data migration, the automated movement of data from older media types to newer ones. Frequently, a piece of data outlives the usefulness of the media it is stored on. Today, the transition from older media to newer media can be automated, allowing organizations to quickly gain new cost benefits.

Additional highlights of IBM's deep history and leadership in ILM include the advent of hierarchical storage management software for open systems in 1993; the delivery of the IBM DB2 Content Management solution in 1988; and the full integration of content management with storage management technologies in 1996.

Managing the Value of Information – Integration with Business Processes

Line-of-business (LOB) managers are increasingly implementing comprehensive content management solutions that can help them to:

- Search for and find information
- Reduce business process cycle times
- Improve worker collaboration
- Manage content on the Web

Corporate officers, driven by the need to show business control are also looking to content management systems to effectively manage the proper retention and disposition of business information and to comply with government regulations. To the CIO, these business imperatives represent additional data to be stored and managed. As an industry leader in content management, IBM provides the building blocks that help LOB managers meet their business imperatives. IBM has also fully integrated its content management offerings with its storage infrastructure and HSM offerings so that IT managers can cost effectively store this new information without the need to retrain administrators or implement a separate storage infrastructure.

To some extent, this data differentiation can be over simplified into answering 5 basic questions:

1. What records or data must be retained?
2. How long should it be retained?
3. Why is it being retained?
4. How is your organization going to find it when you need it?
5. What happens to the records or data when they are no longer needed?

Only IBM has the comprehensive solutions and services to help customers build an IT infrastructure that is flexible enough to change as quickly as business requirements. The ultimate vision for this type of flexible infrastructure can be described as an enterprise whose business processes — integrated end-to-end across the company and with key partners, suppliers and customers — can respond with speed to any customer demand, market opportunity or external threat.

Key to managing the value of information in this environment is the integration of business processes and information. This can mean leveraging formerly fragmented data and making it fully accessible across an organization to enable more accurate and timely business decisions. For example, content management solutions can be used to integrate content for suppliers, customers and employees while records management solutions can be used to meet compliance and regulatory demands by linking retention based business processes with application content management and storage management.

Second to this is aligning the management of the IT infrastructure with business needs. This extends to optimizing storage management to protect data, enhance utilization, and reduce costs. This can be accomplished across heterogeneous environments to reduce the complexity in IT infrastructure.



IBM DB2 Content Manager provides a foundation for managing, accessing and integrating critical business information. It helps integrate all forms of content (email, document, web, image and rich media) across diverse business processes and applications, including Siebel, PeopleSoft and SAP. Content Manager also helps deliver powerful information and services by providing a single repository for all content which can also be managed.

As compliance requirements and regulations become more complex, customers require more control for content lifecycle. IBM Records Manager applies formal records management capabilities to electronic as well as physical records. It is integrated with IBM DB2 Content Manager to provide one repository for records across the organization, but is also available as an engine that can be embedded into any business application and leverage that application's repository. This is especially valuable for companies that want to keep their critical records separate from other systems. IBM Records Manager offers several advantages to recordkeeping such as a building blocks approach to designing and managing corporate file plans, time and

event-based retention of records, and legal holds (suspensions) of records for pending lawsuits or potential audits.

Finally, the IBM vision for leveraging the value of information helps organizations innovate the business to differentiate themselves and deliver new value while making better use of the resources they have and becoming more productive.

Managing the Value of Information – The Concept of Tiered Storage

Once an organization has this better knowledge of how business processes and information interact, various technologies, in particular focus for this article, storage technologies, can be employed to execute on that strategy.

HSM is typically used to allow the over allocation of a disk data pool. A policy is established to move data that has not been accessed in, say, 90 days to less expensive disk or tape. As data meets these criteria no data is moved until the data pool reaches its maximum size. Then based on algorithms in the HSM software, some amount of data that has met the aging policy is moved to make space for new data. A pointer is left in the original pool since the application is expecting to find its data at the original location. When the application calls for a data object that has been staged out, the HSM software brings it back in to the original pool for use. This can cause a problem if the pool is relatively full, in that, before an object can be retrieved, the HSM software may need to move data out of the original data pool to make room. So while this is a great way to manage storage and keep costs down, it does not provide the user with a consistent response time (latency). The algorithm typically used for HSM is called "least recently used" (LRU).

In early 2004, IBM introduced the concept of tiered storage when it announced the IBM TotalStorage® Data Retention 450 . The policies here reference the age of the data, not whether or not it has been used or referenced in a period of time. The policy might say that when the data is 6 months old move it to tape. Note that if the data had been used the day before or there is still space in the storage pool, it makes no difference. The data is moved when the policy is met. Further, and this is important, when the data is accessed after it has been moved, it is accessed directly from the media where it currently resides. No staging back to some other media is necessary.

What this does is provide the ability to set service levels on data and to be able to clearly tie their associated storage costs to them. So, for example, when you mail out your customer statements you might expect some percentage of customers calling your support center in the first 30 days after the statement is mailed. For the next 30 days that activity drops by 50%. By the time the statement is 6 months old, you are hardly getting any calls at all. So at this point in time, you decide to move the statements to tape. Basically you are saying, any statement that is 12 months old or younger we want subsecond response time from the storage system. Any data that is older than that can be accessed in 1 minute. This can result in a significant reduction of TCO. So why is this technique so important now?

Managing the Value of Information – The Concept of JIC Data

ILM has frequently been subtitled, 'managing information from its creation through its destruction'. It is not hard then to realize that during its lifecycle an object's value to an organization changes. For example, a brokerage research report on a corporation may have very low enterprise value during its creation, editing and pre-publishing phases. Once published, however, the report comes to full value, and its preservation as an intact object becomes important. After a period of time, the information contained in this report becomes out of date (say six months) and its value as a report falls. However, the organization may be required to

keep this report for a specific period of time. Contrast this with the customer statement scenario described earlier in this article. While there comes a point where its value to the company has dropped significantly and/or the object is likely to never be accessed again, it has to be kept for a much longer period of time to meet corporate or government retention requirements.



Many vendors in the marketplace have taken an extremely simple (and naïve) approach to ILM, to wit, ILM is the movement of data from disk to disk to disk. It ignores the fact that, during the lifecycle of some data objects, they can tolerate significantly longer latency than disk provides (and hence significantly lower cost). Tape, optical, even paper are viable storage media for ILM data. Many vendors deny the value of this approach mainly because they do not provide that level of flexibility and choice in their product line. So how can an organization determine if the usage of these less expensive media makes sense for its ILM data? Well, one way to make that determination is to understand the concept of JIC data.

To help conceptualize a point in time where data has reached its lowest value, while it is still necessary to retain it, we coin the term JIC Data, which stands for Just In Case Data. It refers to the last stage of an object's life, when the likelihood that it will be accessed is very small or zero and it is being retained "Just In Case" someone or some authority might ask for it. Once an organization can apply the concept of JIC data, they can reap significant reductions in TCO.

In the two examples above, 6 months after the report is written and say 12 months after the customer statement was delivered, these two classes of data may become JIC data. Further consider check images that are xx months old. While there may be requirements for an organization to save these for years, the likelihood of a request for a particular check after xx months is extremely small. An organization might reasonably set an SLA with a 24 hour turn around for any check older than xx months old (each organization will have different requirements and will pick a different number for xx, but again, we have found that 12 months is a typical break point).

Managing the Value of Information – New Considerations for Total Cost of Ownership

To better understand why using a tiered storage approach to manage long term data retention, we need to observe some interesting industry trends.

1. The amount of storage that is being used for long term data retention – This is regularly measured in terabytes. Many organizations already have more than 5-10TB of data they are retaining for years.
2. Data outlives the media it is stored on – Most organizations refresh ATA or SATA disk drives every 3-4 years, 5 years at the most. Long term retention managed data can be kept for 7-10 years or longer. This means that an organization will have to repurchase many terabytes of storage 2-3 times during the life of the data. Further, each time the storage is replaced, a costly and time consuming data migration is required.
3. The difference between the cost of ATA/SATA disk storage and tape – Many vendors will tell you that their disk storage system is less expensive than tape. It's very hard to see how this is possible. At the very bottom end of the scale, ATA/SATA disk storage sells for \$5-7/GB. The IBM 3592 300GB Tape Cartridge sells for about 40 cents/GB. That is approximately a 15:1 ratio and this is a comparison to the least expensive disk on the market. Further, disk storage systems require power to keep the disks spinning, and cooling to dissipate the heat generated by the power. A tape cartridge sitting in a tape library requires neither.
4. So, now the big difference is when you take all of these into consideration. Sure, at very low amounts of storage, the price differential in favor of tape may not be enough to offset the speed of the disks. However, when you consider (a) the large amounts of data being preserved, (b) the long period of time the data is being retained and (c) the differential of media price and operational cost between disk storage and tape, the difference between the TCO of disk storage vs. tape storage can be enormous.

IBM has developed an extensive TCO modeling tool to allow organizations determine for themselves the cost differences between an all disk and a disk/tape hybrid model.

Managing the Value of Information – The IBM TotalStorage DR550, the Industry's First ILM Offering for Long Term Data Retention and Protection

The IBM TotalStorage DR550 has been designed to meet the requirements of efficient and effective long term data retention and protection at a significantly lower TCO. It is also designed to be compliant with SEC 17a-4 which is considered one of the most stringent regulations with respect to data protection, retention and data integrity. It can be configured with up to 56TB of SATA disk storage. What makes it different than its competitors is its ability to incorporate as tiered storage many other kinds of media seamlessly into its storage pools. So, for example, an organization can attach a 3494 Tape Library that contains IBM 3592 Tape Drives. These drives are enabled for normal tape media or WORM tape media. In addition, the IBM DR550 can support the attachment of Optical and/or DVD-ROM libraries. This gives an organization an incredible amount of flexibility and choice. If an organization is already committed to a particular kind of media or vendor, it is likely that the DR550 can support the attachment of those devices.

What this means is that the IBM DR550 has been designed to allow an organization to implement the kinds of policies and

processes discussed in this article. Policies can be established to store an object on the DR550's disk storage for a period of time, say 6 months, and then retained for the rest of its lifecycle on WORM tape. At the end of its lifecycle, the DR550 can automatically delete the object. The DR550 is built with IBM's latest server technology (POWER5), storage technology (DS4100), and data protection technology (Tivoli Storage Manager for Data Retention). As a result, the DR550 can deliver high performance at a low cost with extremely low long term TCO.

The entry model of the DR550 provides 3.5TB of disk storage and lists for around \$80,000. The disk storage on the DR550 scales up to 56TB, and by adding optional tape, optical and other media, scales to multiple PBs of data that can be managed in a single DR550 archive.



Summary

If an organization reviews its policies, practices, procedures and processes around long term data retention it is likely that they will find a significant amount of JIC Data, or data that will become JIC data at some point in time. This presents an opportunity to lower TCO for the storage the data resides upon. IBM has designed the IBM TotalStorage DR550 to help alleviate the costs of storing, managing and retrieving this data from its inception through its destruction. At the same time, it is designed to be compliant with the most stringent of regulatory requirements, providing policy-based non-eraseable and non-rewriteable storage on a variety of media types. Its POWER5 processors provide outstanding performance, while its server-based architecture provides a competitive price. JIC Data and the IBM TotalStorage DR550: Perfect Together!

Datatrend's TrendSetter eNewsletter
January 15, 2005