Mesabi Group

Commentary

April 27, 2011

Rethink Data Protection and Retention Now

In an ongoing race to remain competitive and grow, enterprises are making greater demands upon IT. To better service these needs, IT itself is evolving to eventually provide true "IT as a service." For both enterprises and IT, a top challenge is to provide the scalability and agility required for 24x7 data availability, as well as the ability to service evolving retention needs for compliance, legal discovery, and business analytics. But in light of explosive growth and distribution of information, cost efficiency is essential to provide the necessary level of protection affordably, for the life of the data.

Business is Changing: Data Protection and Retention Practices Need to Keep Up

The future holds major challenges for all enterprises. Businesses face great pressure to not only preserve but take market share with both existing and new products and services. But businesses have a heightened awareness that events outside their control (macroeconomic, political, and natural disaster) also pose threats.

The ability to deal well with change (both planned and unplanned) will be the hallmark of successful enterprises. The twin pillars by which success will be measured in managing change are agility and efficiency. Agility is the ability to respond to meet changing business objectives rapidly. Efficiency is the ability to use economic resources in as close to an optimal manner as possible to meet those business objectives.

To better help the business meet its challenges of change, IT is in a process of transformation itself. IT is moving through stages such as virtualization and cloud computing to true IT as a service, which can be measured (guess what) by agility and efficiency.

Data protection and retention are core IT competencies that must keep up with ever increasing challenges along a number of dimensions. These include increased demand for 24x7 availability, increasing need to manage and protect more data, and the no longer ignorable need to have true data retention management policies that balance compliance and data disposal needs.

The manual, non-integrated approach to data protection and retention needs to be rethought and replaced with an automated and integrated approach that meets the necessary agility and efficiency requirements.

26 Country Lane
Westwood, Massachusetts 02090
Telephone: 781 326 0038
Fax: 781 326 0038
www.mesabigroup.com

Data Protection Is Not Keeping Up

The problem is that data protection processes and technologies that were designed to meet past requirements are not able to keep up with the changing times.

Just a couple of key ways that the changing times affect how enterprises are using IT are:

- **Integration of key applications** — take, for example, enterprise resource planning (ERP), customer relationship management (CRM), and supply chain management (SCM); they now have key points of integration (orders, inventory, billing).
- **New modes of doing business** — the rise of the Web as a business network has led to online e-business as well as social media and collaboration (including e-mail) from disparate devices (smartphones tablets, and laptops among them) without regard to location or time.
- **Distribution of data –** Today critical information resides both inside and outside a single data center, including remote servers, additional data centers, PCs and laptops.

The implication of integration from a data protection perspective is that downtime in one critical system impacts the ability of other key systems to function as well as they should.

One implication of new modes of business operation is that almost everything IT is 24x7 including the integrated systems (which may use the Web for SaaS-type CRM).

The net result is that no downtime is tolerable for key applications and "always on" is becoming highly desirable for any application that is accessed for business reasons on a 24x7 basis.

Among the impacts of these trends on data protection technologies are the need to make sure that replication approaches (such as remote mirroring) transparently fail over from physical failures (such as the loss of a disk array) and that snapshots are properly integrated with backup software. The traditional backup window where applications were brought down for a period of time is now an anachronism.

IT is also making key changes, such as:

- **Server virtualization** — a key objective of server virtualization is to better utilize physical-server CPU and memory resources by consolidating multiple applications using virtual machines (VMs) on the same physical server. An unintended possible consequence is that backing up a number of VMs may lead to contention problems with physical servers or network resources.
- **Cloud computing —** the cloud promises flexibility and efficiencies, such as for resource provisioning, as well as offsite protection for disaster recovery. Currently, the best cloud solutions are addressing concerns such as how to provide security and multi-tenancy properly.

For both server virtualization and cloud computing, combining automation and integration with data protection technologies is essential for achieving the levels of data protection and availability that each must deliver.

26 Country Lane
Westwood, Massachusetts 02090

Telephone: 781 326 0038
Fax: 781 326 0038
www.mesabigroup.com

Employing automation properly is necessary in order to deal with new levels of complexity. In conjunction with automation, integration ensures that multiple tools, such as snapshots, backup software, and storage virtualization, run in harmony with one another.

Data protection is unlikely to have kept up with structural changes (such as the effective use of an active archive) or new needs for discipline (such as a thorough, enforceable, and enforced data retention management policy). And what is often lacking is the planning and management process that enables structure *and* discipline — the process called information governance.

Formal Information Governance Policy Must Link to Data Retention Practices

Business processes (and that includes data retention processes) tend to run the same way today as they did yesterday and will run the same way tomorrow in most cases.

How can an enterprise rethink those data retention processes and employ the necessary data retention technologies?

Well, an IT architect can sit down and rethink what needs to be redone for data retention. That is a necessary condition, but not a sufficient condition. Why not? There are at least three fundamental reasons.

The first is that IT is the physical custodian of business processes and data, but does not set the business rules for the enterprise. For example, IT cannot make arbitrary rules for the destruction of data. The legal department and the business unit have to make those decisions.

The second is that the business department that benefits and the one that pays are not necessarily the same. For example, a business department may not care that unnecessary data is not destroyed, but IT bears the burden of the storage costs.

The third reason is the hierarchical management of enterprises results in well-defined roles with specific tasks for individuals, but that can leave gaps in accepting responsibilities as things change.

For example, the exploding growth in data has resulted in retaining data that serves no useful business purpose, even compliance. Legal is concerned about retention rules only for data involved in legal matters and business units generally pay little attention to older data that does not have an ongoing business value.

Information governance is necessary to overcome the three problems cited above. Information governance (a.k.a., data governance) includes the people, policies, practices, and procedures to ensure the preservation, availability, confidentiality, and usability of an enterprise's data.

Information governance involves the formal collaboration of all the data protection stakeholders in an enterprise. The information governance team can deal with the policy, cost, and responsibility issues that might otherwise sidetrack an effective data protection strategy, including planning and guidance on service level agreements, retention management practices, and compliance.

26 Country Lane
Westwood, Massachusetts 02090

Telephone: 781 326 0038
Fax: 781 326 0038
www.mesabigroup.com

IT can work with the information governance team to define a comprehensive (not one-off or piecemeal) approach (with the necessary technologies) that can be implemented in a planned sequence to meet prioritized business objectives within the necessary financial restraints.

Getting Data Protection and Retention to Keep Up with Business Needs

Three key aspects of data protection and retention that together illustrate the need to rethink data protection are: continuous data availability for high service levels, retention management and archiving, and managing the growth and cost of the data footprint.

Continuous Data Availability Is Required for High Service Levels

How are IT organizations going to deliver always on (i.e., 24x7) availability of data to those applications that demand it? The answer is, on the physical data protection side, to be able to failover transparently and as nearly instantaneously as possible to an active copy of the data.

Now, both local and remote mirroring have probably been done for years, but one still has to perform a careful review to ensure that the right type of mirroring and the necessary automation software is applied appropriately.

On-site local mirroring and short distance (less than 60 miles) synchronous remote mirroring protect against operational failures (say, two nearly simultaneous disk failures on a RAID 5 array) at the original production data site. Long distance (300 miles or greater) separation of facilities for disaster recovery purposes needs to use short-latency-acceptable asynchronous remote mirroring. The total mirroring process needs to be automated to assure proper failover to an active copy of the data, and fail-back when appropriate.

Note that in the old days, mirroring had to be done to identical vendor disk storage systems and to the same tier of storage, typically Tier 1 FC or SAS storage. That no longer has to be the case. For example, Tier 2 SATA storage may be used at the non-production site. Yes, performance degradation might take place, but then again it might not as some critical applications may not be I/O intensive in fail-over mode. That makes mirroring more affordable (think cost-efficient).

But mirroring provides robustness against physical failures, not logical problems (data base corruption, viruses). Snapshots are an answer for quick restoration after a logical problem, but snapshots today are not your father's snapshots. Today incremental snapshots can be taken over time on the original full snapshot. These space-efficient snapshots improve the economics of taking snapshots by shrinking the recovery point, data-set size, and the recovery time for restoring data. Snapshots are now the first line of backup defense for business applications (although there is still a need for the other lines of defense).

So the addition of modern mirroring and snapshotting provide the agility to respond to failure (negative changes) as well as greater efficiency than in the past.

26 Country Lane
Westwood, Massachusetts 02090
Telephone: 781 326 0038
Fax: 781 326 0038
www.mesabigroup.com

Retention Management and Archiving

Retention management requires clear policies, the ability to precisely identify what policy rules apply to every piece of data in an enterprise, and an enforcement mechanism to ensure the retention policies are actually carried out.

A necessary consequence of retention management is the need to build an active archive to house the data that is retention-managed. For the most part, retention management impacts data that is fixed, which means no longer changing. In fact, fixed data is most of the data in an enterprise. Also, the retention-management application has to have control over what can and cannot be deleted or changed (which is necessary to preserve chain-of-custody for legal purposes).

So retention management involves moving all fixed content (not just e-mail) to an active archive that can still be read by the originating application. Once this is done, a lot of good things happen for both the business and IT.

First, the active production systems shrink dramatically in size, which saves bandwidth and time when failing back from a remote mirror, and greatly reduces the size of a full snapshot. Moreover, full backups to disk and/or tape are smaller and faster. And on the production side, less system storage means better performance.

On the archiving side, more cost efficient disk (SATA) can typically be used. Moreover, either on the way into the archive or shortly after insertion into an archive, one can dispose of data that no longer has any business or legal value. Deleted data costs the least — none.

The active archive is not just about cost efficiencies; it is also about business value. The archive provides the basis for meeting compliance audits and aids greatly in the legal discovery process. But the real business value comes from using the archive for the ongoing business needs of the enterprise. These needs can include the retrieval of key information, such as an MRI medical image, an old contract, or customer history information. However, the archive can serve not only for the retrieval of individual information items, but also as the foundation for analytics that detect patterns that can give key insights to management.

So retention management leads to archiving, which has data protection benefits, but also provides other business benefits as well.

Managing the Growth and Cost of the Data Footprint

Data space management is about efficiency, which is necessary to deal not only with the increasing volume of original data, but also with all the spun off copies that are needed for data protection.

One of the hottest topics in data protection today is data deduplication, but deduplication is only one in what should be a comprehensive approach to data space management.

The first step in data space management is to properly apply retention policies to dispose of all unnecessary data. Propagating deletion to all copies of the original as well as the original saves the most space. The second step is copy management. A presumed benefit of data deduplication is to reduce the number of data copies

26 Country Lane
Westwood, Massachusetts 02090

Telephone: 781 326 0038
Fax: 781 326 0038
www.mesabigroup.com

needed (read storage space). An obvious, but sometimes overlooked, step is to determine whether extra copies are really needed. Note that data deduplication technology can operate at a file level (single instance storage) or at the subfile level (a finer level of granularity than a file) to generate space savings and reduce the bandwidth required for online backup.

Finally, data space management efforts should consider compression. Note that deduplicated data can still benefit from compression, and vice versa.

In the past, compression was typically applied to data on tape. Today, an evolving strategy is to be able to compress on primary production storage. A potential problem is performance degradation, but at least one solution has overcome this objection, so watch that space carefully.

In cases where data growth continues even with efficient management and deduplication, a solution should scale automatically to meet the changing needs without significant investment expenditures.

Tying It All Together

The robustness to handle shocks that would otherwise disrupt a business is the key agility benefit that data protection provides when it ensures continuous data availability. In addition, making an archive active improves agility when a business needs to find ongoing value from fixed content data as well as to meet evolving compliance and legal discovery requirements.

To do this efficiently in the face of increased data growth with constrained budgets requires storage management strategy that can take advantage of storage tiering, retention management, and data space reduction management.

Conclusions

An old familiar saying is "that if isn't broke, don't fix it." Well, data protection is broke and needs fixing. For many organizations, the fix is not just a patch, but a rethinking of the total data protection strategy. That does not mean that IT organizations are going to have to undertake a massive rip and replace process. It does mean that IT is going to have to create a more comprehensive strategy that goes beyond a piecemeal approach, and must prioritize the introduction of new and extended data protection and retention technologies as time goes on.

All in all, enterprises need to rethink their data protection and retention practices, and they need to do it now.

David Hill

Analyst Name: David Hill
Topic Area: Data Protection

Mesabi Group LLC
26 Country Lane
Westwood, MA 02090
www.mesabigroup.com

This document was developed with IBM Funding. Although the document may utilize publicly available information from various vendors, including IBM, it does not necessary reflect the position of such vendors on the issues addressed in this document.

Phone: (781) 326-0038
email the author: davidhill@mesabigroup.com