Not All Data is Created Equal

Designing and Managing a Tiered Storage Environment

emtec
Transforming IT

Executive Summary:

Enterprise storage costs are spiraling out of control. Demands for storage capacity have been exponentially increasing across multiple industries, and IT departments are too rapidly *reacting* to these demands—purchasing hardware and software before assessing what they actually need.

This amplified spending is evidenced alone by the projected growth of the global storage management software market; it is estimated that in 2008 the market will be approximately $11.1 billion in total revenue, and it will reach approximately $14.8 billion by 2012.[1] When you consider that software is only half of the equation (hardware and implementation costs will compound these numbers), companies will spend an astounding amount of money on their storage infrastructures in the next five years.

While IT departments will likely need to invest in some additional storage during this time, it is critical they first get a real handle on the storage assets that they already have before engaging in reactive spending. In far too many cases, storage administrators can't answer with confidence whether their storage infrastructure is truly providing business value and meeting IT/storage objectives. They don't know if their environments are unnecessarily overtaxing the financial and personnel resources of the organization. The reality is, what they don't know is hurting them—"knee jerk" spending is costing companies tens- to hundreds of thousands of dollars each year. Don't make that same mistake.

By performing a proactive assessment, you can determine up front how to more efficiently manage your existing assets. Many companies find through this evaluation process that tiered storage is an effective strategy for aligning storage hardware, software, services and processes to achieve greater performance, high availability and scalability. Those who have a tiered infrastructure already in place learn through an evaluation how to best manage it and drive greater efficiencies. This paper examines the benefits and strategies behind storage assessments and the advantages of tiered storage environments.

[1] *Gartner Dataquest Storage Management Software, Worldwide 2007-2012 Report.*

Storage Assessment—The Evaluation

Storage consumption is skyrocketing—by 2011, users are expected to amass almost 6.5 times the amount of terabytes of data installed in 2007, and the number of terabytes per array will increase by 400 percent in the next five years[2]—in part because companies aren't effectively leveraging their current storage environments. Take, for example, some recent statistics on one piece of the equation: utilization. A study revealed that 51% of open system data is unnecessary, duplicate, or non-business related; 68% of data has not been accessed for 90 days or more.[3] The proof is in numbers like these. There is a serious need for storage assessment.

So what is storage assessment? By definition, it can be likened to the method of establishing disaster recovery plans where companies determine what data is mission critical, what is business critical and what tiers of applications should be up and running at all times—first, second, third, fourth. Through the assessment process, companies reveal the true "health" of their storage environments via analysis of six primary areas:

- Storage capacity utilization
- Storage capacity growth
- Trending, file ageing trends
- Stale file identification
- File ownership and distribution
- Storage housekeeping

Armed with this new found knowledge, companies can put strategies in place for extracting greater value from their current infrastructure and augment where needed.

Some companies find the results of their assessments nothing short of astounding. One of the nation's largest savings and loan organizations, a multi-billion dollar institution, recently performed a storage and data profile assessment of 35 servers and four Storage Area Networks to support the integration and alignment of data storage for an upcoming server and storage consolidation effort. The assessment examination included a survey of the total amount of data stored, total amount of free space, occupied space on all volumes, file-type, user occupied space, file-aging, duplicate files and wasted space. The institution obtained clarity around the

[2] *Source: Gartner's 26th Annual Data Center Conference*
[3] *Source: SNIA/Source Consulting*

utilization and underutilization of its existing storage, and learned that it was storing 2.6 million files, representing 281 gigabytes of data (including 1.6 million duplicate files) that hadn't been accessed in more than three years. Some of its servers were at 90 percent utilization housing unnecessary data and were in severe danger of exceeding capacity.[4] These results were a call to action for the institution to implement a tiered storage environment. Its current production data now resides on faster performing RAID arrays, while archived materials are stored on more cost effective storage, such as offline tape.

The results of assessments like these are increasingly making quantifiable arguments for tiered storage as a strategy for meeting multiple performance needs of databases, backups, archives and many other enterprise applications. This is also driving down the overall cost of storage. Validating the strategy, tiered storage build out was recently cited as a top three storage initiative for Fortune 1000 companies in 2008[5]. How do you go about designing and managing a tiered storage environment?

Tiered Storage—The Approach

Tiered storage enables IT departments to break down their storage requirements into digestible, manageable pieces. Specifically, tiered storage is the implementation of two or more storage schemes with distinctive cost/performance characteristics with the reorganization of corporate data onto the most appropriate platform. The objective is to match the data's relative value to a particular tier, placing more recent or valuable data on the faster performing storage, while relegating the older, less critical or infrequently accessed data to less expensive storage. Shifting less valuable data to less-expensive storage media can allow for very high storage capacities (usually measured in hundreds of gigabytes per drive) at a lower cost per gigabyte.

In most cases, organizations can realize a dramatic cost savings upfront, before building out their tiered storage infrastructures, by conducting storage assessments. For example, a large county

[4] *Source: Emtec Storage Assessment Case Study*
[5] *Source TheInfoPro, "Wave 10 Storage Study."*

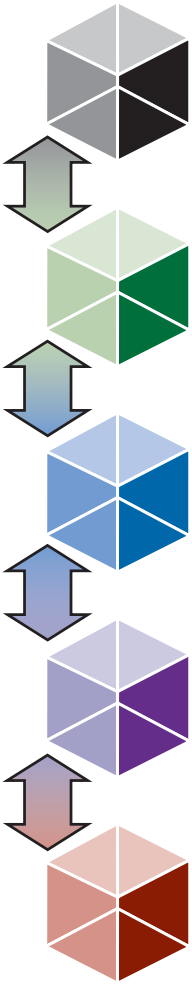
government in New York was looking to make a significant storage purchase because it was running out of capacity and needed to provide replication capabilities for its disaster recovery site. The county's storage vendor provided a "free" evaluation of its existing infrastructure; the result of which was a recommendation for 6.5 terabytes of high-end storage that would cost the local government one million dollars. The county sought a second, vendor-neutral opinion from Emtec, which conducted an extensive analysis, including a hardware audit, staff interviews, backup log reviews and storage data collection.

The conclusions drawn from the non-biased assessment addressed both under and over utilized storage and general housekeeping issues, such as duplicate and old files. The final recommendation differed significantly from the manufacturer's. The county could effectively meet its immediate and long-term storage objectives by implementing a tiered infrastructure, Hierarchical Storage Management (HSM), for an additional investment of only one terabyte (versus 6.5 terabytes) of mid tier storage. This recommendation saved them at least $500,000 in hardware purchases.

Organizations are also seeking to optimize performance. Although tiering involves lower performance drives, many companies see an immediate improvement in storage performance, because network traffic is spread among several platforms rather than competing for access on a single storage system. This enhances the storage service experience for all users.

Tiered storage is not without its challenges. The biggest of which is data management—matching the data to each tier, migrating the initial data and then moving data between tiers over time. This is where data classification comes in—the action taken to identify data and determine its value to the organization. Data classification is the first critical step in placing the right data into appropriate storage tiers.

*Figure 1 shows the categories into which data is commonly classified.**



Mission-critical Data

- Most valuable to an enterprise, high access
- High performance, high availability, near zero downtime, highest cost

Business-critical Data

- Important to the enterprise, average cost
- Reasonable performance, good availability, less than eight-hour recovery

Accessible Online Data

- Cost sensitive, low access, often compliance of fixed content
- Online performance, high availability, less than eight-hour recovery

Nearline Data

- Cost sensitive, low access, large volumes
- Less than one-hour access time, automated retrieval

Offline Data

- Archived data, backup or compliance related
- Very cost sensitive, limited access, ~72-hour seek time

**Source: IDC*

Data classification can be, but does not have to be a manual process. There are software products that can help to identify and move data types, but there is no *automatic* way to determine the value of specific files to your organization; comprehensive data classification involves input from key departments across the enterprise and it should never be approached as an IT-only function.

Managing a Tiered Storage Environment

There are many benefits to tiered storage, but it is important to acknowledge that once data has been classified and tiered appropriately, the approach can create some ongoing challenges. More tiers create greater complexity and the migration process takes time and must be repeated regularly (e.g.; daily, weekly,

monthly, etc.) as more data is created or existing data ages or changes in relative value. Software tools can help to facilitate the migration process, but storage administrators must establish data movement and management policies.

Another consideration is that tiering can potentially create interoperability issues between the software tools and storage platforms. Not all storage tiering tools are fully heterogeneous across the physical storage platforms, and this can create serious hurdles for organizations that must use multiple tools or forego the use of certain storage systems. Therefore, conducting due-diligence up front is extremely important to ensure interoperability.

That being said, another benefit of tiered storage is having the ability to chargeback storage users based on the service level of each tier. For example, top-level Tier-1 storage might command the highest value, while Tier-2 SATA storage might be significantly cheaper; archival disk or tape platforms would be even less expensive. Keep in mind that the software tools that enable tiering also have chargeback capabilities built in.

Figure 2. ABC Company Sample Storage Tier Metrics
In further examining the management of a tiered environment, the data classification categories can be mapped to each tier as follows. This illustrates what the metrics for each category could look like for your organization.

Tier Name	Mission Critical	Business Critical	Accessible Online	Nearline	Archival
Availability	99.999%	99.999%	99.99%	99.90%	Offline
RPO (recovery point objective – how far back you need to go)	2 Hours	2 Hours	10 Hours	24 Hours	2-3 Days
RTO (recovery time objective – how long it takes to recover)	1 Hour	4 Hours	14 Hours	24 Hours	2-3 Days
Point in Time Snapshots (how often copies of data are made)	1/2 Hour	1/2 Hour	1 Hour	3 Hours	NA
Tiered Snapshots (ability to aggregate snapshots)	No	Yes	Yes	Yes	NA
Offering Cost/GB (Relative – how much storage allocated to snapshots)	100%	80%	60%	40%	10%

Lastly, it is also important to note that the majority of tiered storage tools center on Storage Resource Management (SRM) software products. SRM products can typically collect, store, backup, recover, provision, virtualize and forecast data storage. SRM software may be offered as a standalone product or as part of an integrated program suite. However, not all tiering tools are generic, and key storage manufacturers are paying particular attention to tiering functions.

Stop. Think. Assess.

Before you spend another dollar on your storage environment, be sure that you know how to best leverage your existing infrastructure and/or gain the knowledge you need to quantify your rationale for a next purchase. But also, don't reserve assessments for a particular storage "event." There is real monetary value in conducting this type of analysis on a regular basis to determine your day-to-day storage needs and gain a better understanding of how your storage architecture works.

The reality is that in-house assessments can be time-intensive, and often storage administrators do not have a methodology in place or the time to spare to conduct an assessment. Frequently, companies rely upon the manufacturers of the storage technologies they are using to conduct free assessments and make recommendations based on their findings. But let's face it: manufacturers are motivated and compensated to up-sell their products. The good news is that there are companies that provide vendor-agnostic, objective recommendations of how an IT organization can maximize their current storage infrastructure and accommodate future demands.

Conclusion:

Ultimately, it is crucial to understand that all data is not created equal and therefore should not be treated as such. A tiered storage environment—mapping your data by category to the proper tier—can ensure that your data will remain available, while minimizing both risk and cost as well as increasing performance.

Emtec's Storage Assessment service provides a vendor-agnostic view into an organization's current storage environment and possible needs. After monitoring current and historical data usage for three to four weeks, Emtec provides an accurate, complete report of utilization, strengths, shortcomings, bottlenecks and redundant operations. Taking the results obtained from various Storage Resource Management tools and applying intelligence gained both through experience and collaboration with the client teams, Emtec storage architects then develop an effective storage strategy—focused on present and future needs rather than on any specific platform or technology. Through this process, companies gain invaluable insight into how to reduce both capital expenditure and on-going storage management costs, while improving data availability and infrastructure performance. If Emtec can't save you at least the cost of the assessment, the Storage Assessment is free.