Data Mobility Group

www.datamobilitygroup.com

White Paper

October 2006

Maximizing the Value of Tiered Storage

 76 Northeastern Blvd. Suite 29A, Nashua NH 03062 Phone: 603.835.6141 Fax: 877.254.4886

Executive Summary

Storage tiering once meant adding a complement of SATA disk as an intermediate storage tier between high performance disk and nearline tape. Not any more. With the multiplicity of storage tiering options now available expressed in both hardware and software, storage administrators have the ability to create tiered storage environments that are optimized to their own applications and operational circumstances. This paper explores these options and makes specific recommendations regarding deployment from the highest to the lower tiers.

Driving the Need for Tiered Storage

It has become something of a given for enterprise storage administrators that stored data volumes will grow at an average rate of 50% per year. Rapid growth stresses both the storage systems that must handle the increasing volume, and the administrators that must maintain application performance while also dealing with new compliance and legal data retention imperatives. Applications and other factors now driving growth include:

- Email and attachments (and the associated regulatory/legal retention requirements)
- The proliferation of multiple (and often unnecessarily redundant) copies of data
- ERP and CRM applications
- Data warehousing

Beyond impacting storage, this growth overloads primary production storage subsystems and creates a negative impact on application performance. Unfortunately, such is often the case when IT departments are tasked with doing more with less and the prevailing perception is that raw storage is relatively inexpensive on a cost per GB basis.

What Exactly is Tiered Storage?

Two years ago, storage tiering meant moving data contained in high performance, high availability, and higher cost disk array subsystems off to lower performance/lower cost per GB devices. Data sets that were candidates for this migration process were those whose activity levels had diminished over some predetermined period of time. Data sets that supported critical applications and saw the most I/O activity received the highest priority and were therefore allocated to the "gold standard" arrays within the IT environment. Data sets with less activity or less importance in terms of application criticality were moved to lower performing/higher capacity arrays (SATA for example.) By adding slower, capacity-weighted disk arrays based on SATA and FATA to its normal complement of high performance, high availability arrays, an intermediate disk storage tier (tier 2) that lived between "gold standard" arrays (tier 1) and nearline tape (tier 3) was created.

Today, additional tiering strategies have emerged that are based on other factors including availability and data protection levels as well as regulatory compliance and corporate governance requirements. Alternative tiering architectures are also emerging. For example, storage tiering may now be enabled by storage virtualization; that is, multiple disk arrays composed of the same drives[1] may be managed by a virtualized array controller. It also enables disk array generations[2] to be managed by a virtualized, heterogeneous storage controller.

Today, a good working definition for tiered storage is:

> *"The use of storage systems that have different performance, scale and cost characteristics. It is the assignment of different categories of data to different types of storage devices and media based on access, availability, and scalability needs as well as business policies including, in some cases, compliance needs. Policy-based data migration functions move the data across tiers in accordance with business policies while backup works throughout the storage environment to meet business continuity and disaster recovery requirements."*

Tiered storage means more than adding arrays of intermediate disk such as SATA arrays. Here we summarize the options now available to storage administrators.

Data Mobility Group

www.datamobilitygroup.com

	Tier 1: Online Storage	Tier 2: Near-Online Storage	Tier 3: Nearline
Disk Drives	- 10-15K RPM – - FC/SAS/FATA	7.5K RPM SATA	
Disk Arrays	- FC/FATA/SAS disk based high availability - high performance - strong feature set - Best data protection and recoverability - High cost per GB	- Pure SATA drive - Medium availability and performance - Minimal feature set - Minimal data protection/recovery capability - Low cost per GB	
RAID Groups	RAID 1+0 RAID 1	RAID 5	
Storage Media	- Online high performance FATA disk	- Near-online Virtual Library Systems - Tape and optical - Online SATA disk	Off-line tape and off-site vaulting
Array Generation	- Latest generation still on warranty/ maintenance	- Latest generation of lower cost array - Next previous generations of Tier 1 devices (on minimal maintenance or self maintained)	
Archive Requirements	- Data migration to SATA-based systems	- Data migration to nearline tape	
Backup/ recovery requirements	- Backup to disk-based Virtual Library Systems	- Backup to nearline tape	- No backup protection required
Local/Offsite replication requirements	- Disk-to-disk, three-way replication (synchronous local copy plus asynchronous remote copy)	- Snapshot	- No replication, may have tape duplication

 76 Northeastern Blvd. Suite 29A, Nashua NH 03062 Phone: 603.835.6141 Fax: 877.254.4886

Tiered Storage, Tiered Data, Unified Management

Historically, significant savings can be realized from minimizing capacity growth on the gold-standard disk arrays by moving data off to higher capacity, lower cost-per-GB storage devices. Today, the challenge is to be able to do so in such a way that:

- Application performance is not negatively impacted. In fact, performance gains should be expected from moving less frequently accessed data off of high-performance disk arrays.
- Management complexity is not magnified by creating multiple, dis-integrated management consoles.
- Unnecessarily redundant copies of data are not created. Rather, one of the goals of a storage tiering project should be to reduce the number of data copies.
- Data is not left in an unprotected - or perhaps worse - in an unsecured state as may be the case with backup tapes stored in remote offices.
- Retrieval of data that must be promoted back to a higher tier as required by an application or application user is not hampered.

Early storage tiering implementations exposed these issues, suggesting the dual need for a data classification scheme coupled with some way to move data non-disruptively up and down the physical storage tiers over time. This realization created a critical need for unified storage management software. It is not enough to partition the storage environment into “buckets” ranging from high- to intermediate- to low-performance and availability. One must also have an integrated view of the tiered infrastructure as well as some way to classify data and automate the allocation of data over time based on classification.

To do this, storage administrators need centralized data management views and functions layered on top of the physical tiered storage architecture. This may require vendors to move some data and content management functionality that has traditionally lived at the application level down into the storage infrastructure as well as introduce tighter integration between performance monitoring and migration so that data sets can be temporarily demoted to a lower physical tier or promoted to a higher tier as production processing and recovery needs dictate.

For many organizations, the process of designing and deploying an effective tiered storage system can be accelerated and simplified by working with vendors who can provide a complete portfolio of storage devices, media, management software, and services as well as optimize the organization's existing technology investments. The right vendor can help an organization understand how data aligns with changing business requirements, translate them into multiple tiers of storage with varying cost and performance profiles, and simplify the process of implementation.

The result will be a tiered data environment where storage administrators can:

- View and search through structured and unstructured data types residing on any storage tier
- Provision tiered storage resources dynamically and non-disruptively
- Migrate data (data mobility) among tiers non-disruptively

All of the above functions should be manageable from a central point of control.

Data Movement, Software, and Tiered Storage

Tiered storage has begun to deliver more strategic business benefits through the use of software. Organizations are subdividing enterprise storage into tiers on the basis of performance, availability and other factors and using data movement software to automatically redistribute the data across tiers in accordance with business policies. Backup software is also used – working throughout the storage environment to meet business continuity and disaster recovery requirements. And to ensure that the administration of the tiered system doesn't increase in complexity, thus offsetting its benefits, heterogeneous storage management software, as discussed above, is helping keep overall TCO under control.

Data movement technologies, and especially database pruning/tuning applications, have developed into important tools in the design and implementation of a tiered storage environment. Data movement software allows IT administrators to allocate files to different tiers based on business policies defined by metadata. Data is then automatically promoted or relegated to storage tiers based upon application and user needs.

Production datasets can be pared and pruned based on application-driven rules and policies using specialized applications that perform this function. The process, also referred to as active archiving, solves many of the performance and manageability challenges posed by contemporary databases that now increase in volume at the rate of 40% or more per year.

For example, a storage administrator can set a policy such that data sets reaching 20GB in volume are pruned according to access frequency. This less frequently accessed data is then migrated to secondary disk storage or even to tape to be held in an "active" archive. If properly implemented, historical transactions can be maintained as complete transactions, enabling the data to be accessed long after the source application is retired.

The benefits of this technique are that the data is still immediately accessible and overall performance of the application improves yet TCO is reduced.

Conclusion

It is a given that the volume of stored data will continue to grow and the rate of growth will continue to accelerate. Growth will be generated by the need to retain even seemingly useless data for longer periods of time, particularly for the purpose of responding in a timely and complete manner to legal discovery – an increasingly common and unwelcome occurrence. But growth will also be generated by new applications that leverage new data types – RFID for example. And while data growth can't be avoided, it can be managed.

Storage tiering can help manage growth while holding down overall costs related to capacity and management. However, care should be taken when architecting the tiered storage infrastructure not to create too many tiers as this may in fact result in increased complexity and cost. On the other hand, attributes of the storage environment that are particularly useful in a tiered storage architecture include:

- Multiple classes of storage (classes can be defined by a combination of features)
- Virtualization
- Heterogeneous, non-disruptive data migration
- Data and operational management software
- A data classification scheme

- Data management policies based on data classification
- Automated data management policy

An effective tiered storage environment will incorporate storage tiers with varying cost and performance profiles, data movement software that automatically moves information to the correct tier of storage in accordance with policy and service level agreements, and storage management software that provides a holistic view of the tiered storage environment. It may also include services to help with the discovery, profiling and alignment of information with business requirements. Further, it will require an understanding that data retrieval is as important to application users as data storage -- perhaps even more important. When data sets are migrated to lower performance, lower cost, higher capacity tiers, provision must be made such that, if a data set needs to be promoted back to a higher level tier, it can be done so transparently and expeditiously.

Finally, a well implemented tiered storage solution also allows businesses to evolve their infrastructure with changing business requirements – while supporting the level of data accessibility needed now and in the future.

(Footnotes)

[1] For example, all FC arrays as tier 1 vs. all SATA arrays as tier 2

[2] The latest release generation is tier 1 vs. next previous release generation which could then be allocated to tier 2

Data Mobility Group

www.datamobilitygroup.com

for example

HP StorageWorks offers a wide range of storage solutions that can deployed in a tiered storage architecture. Leveraging HP's hardware, software, and services portfolio, incorporating a variety of storage technologies and platforms, and utilizing deep application and information management expertise, HP delivers cost effective, manageable tiered storage that meets business and IT requirements.

Online Storage Solutions

For Operational (transactional) data, dedicated storage or shared storage arrays reduce management and maintenance costs while improving overall IT flexibility. HP online storage solutions incorporate fibre channel (SATA and FATA) and SCSI technology for transactional and some operational information

- HP XP, EVA, and MSA Array Families
- HP StorageWorks Clustered Gateway
- HP ProLiant Storage Servers
- HP SVS 200 Virtualized Controller

Near-Online Storage Solutions

For reference information or as a staging platform for faster application recovery, these HP storage solutions offer lower cost storage technologies

- HP Virtual Library System
- HP Data Protection Storage Server
- HP Reference Information Storage System (HP RISS)
- HP File System Extender (FSE)

Nearline Storage Solutions

Tape systems reduce your backup window and lower tape backup costs while improving data protection and recovery time. HP Nearline Storage solutions offer low cost media for data that does not require immediate access and can be restored in disaster recovery situations.

- HP EML, ESL and MSL Series Tape Libraries
- HP DAT and Ultrium Tape Drives
- HP DAT and Ultrium Tape Autoloaders

 76 Northeastern Blvd. Suite 29A, Nashua NH 03062 Phone: 603.835.6141 Fax: 877.254.4886

Data Mobility Group

www.datamobilitygroup.com

Data Movement Solutions

Improve quality of service and reduce data management costs through policy driven automatic data discovery, classification, migration and backup – across storage tiers – based on organization policies.

- HP Reference Information Manager for Messaging (HP RIM)
- HP Reference Information Manager for Databases (HP RIM)
- HP Data Protection Storage Server
- HP Data Protector

Infrastructure Management

Integrated management tools keep operational costs down. HP offers solutions that enable IT organizations to meet service levels, control costs, and maximize IT staff and employee productivity. HP's Unified Infrastructure Management enables an organization to to manage the entire heterogeneous tiered storage environment from a single, unified console.

- HP Storage Essentials

Tiered Storage Services

HP Services help customers explore and implement a flexible architecture that can automatically move information to the correct tier of storage with varying cost and performance that align and evolve with changing business requirements.

4AA0-8507ENW, October 2006