



Shedding Tiers – Creating a Simpler, More Manageable Storage Infrastructure

By Gary Orenstein, Vice President of Marketing – Gear6 [www.gear6.com]

Introduction

The concept of segmenting data storage repositories is almost as old as data storage itself. Back in the early days it was called hierarchical storage management or HSM. The technique was intended to save money by keeping only the most important, frequently accessed data on top-performing storage, and migrating less frequently used data to more cost effective mediums. Initially the split was often between disk and tape. Not so long ago, the industry promoted a term called Information Lifecycle Management which to many observers was just a fancy repackaging of HSM. And more recently, tiered storage has become a more common term as data center and storage managers strive to balance cost, capacity, and performance for different storage priorities.

No matter what you call it, the same principles and truths apply. First and foremost, managing multiple buckets (or tiers) is always going to be more difficult than managing a single storage repository. It does not matter how much fancy software you have to help you, there is simply no way that a highly segmented infrastructure is going to be easy to manage. It may appear to provide cost-savings on an overall basis, but frequently the overhead in configuring and managing a multi-tier infrastructure is more trouble than it is worth. In this article, we explore these realities and why many data storage managers are investigating how to shed tiers instead of adding them.

Forces Driving Tiered Storage Interest

Tiered storage is getting a good deal of attention these days, in many respects because of its “promise” to help enterprises reduce total storage costs. The theory goes something like this: highly sensitive, business critical data might be assigned “Tier 1” status that is stored on expensive, high-speed media such as 15K RPM Fibre Channel drives; Tier 2 data could be seldom-used files stored on less expensive media; and Tier 3 content could be maintained on even less expensive storage media for archival purposes.

Basically, the capacity distribution and dollars per Gigabyte distribution for storage are inversed. Due to the high cost of top tier storage, administrators are forced to explore alternatives to migrate data down to lower, more cost effective tiers, as indicated in Figure 1.

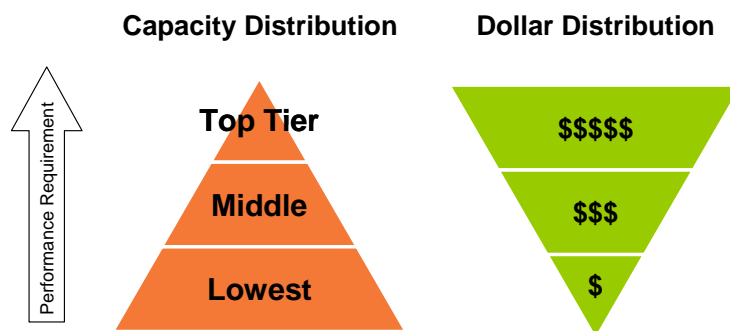


Figure 1: Storage managers are always looking for cost effective alternatives

But what is the impact of breaking up data sets into multiple tiers? First, it adds a great deal of complexity because storage administrators have to set up policies for each tier. And second, backup and recovery procedures must be established for each tier. Finally, guidelines must be created to determine when data should move up or down these tiers. So rather than alleviating a storage administrator’s workload, tiered storage strategies actually increase management overhead.

The complexity and redundant management tasks associated with a tiered storage model are shown in Figure 2.


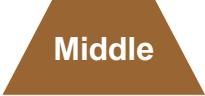

 Top Tier	 Middle	 Lowest
Each tier requires maintaining a redundant checklist		
<ul style="list-style-type: none"> • Provisioning • Snapshot • Replication • Backup • Recovery • Security • Acquisition/testing • Migration policy <ul style="list-style-type: none"> • Rules • Regulations • Exceptions • Overrides 	<ul style="list-style-type: none"> • Provisioning • Snapshot • Replication • Backup • Recovery • Security • Acquisition/testing • Migration policy <ul style="list-style-type: none"> • Rules • Regulations • Exceptions • Overrides 	<ul style="list-style-type: none"> • Provisioning • Snapshot • Replication • Backup • Recovery • Security • Acquisition/testing • Migration policy <ul style="list-style-type: none"> • Rules • Regulations • Exceptions • Overrides

Figure 2: Maintaining multiple tiers means more work

In a perfect world, if an enterprise's storage budget were limitless, companies would simply keep buying fast disk drives. And if high performance and fast response time weren't an issue, organizations could continue to buy high-capacity, low performance drives because they could reduce the overall number of spindles needed to house their constantly expanding data. In reality, however, storage administrators turn to storage tiering because they have to, not because they want to.

Ultimately storage professionals are looking for the right balance of performance, capacity, and cost, which to date has meant adding management overhead as they are forced to migrate data back and forth between tiers. Sometimes this can be semi-automated, but other times it requires manual intervention.

But things are starting to change. With the emergence of highly scalable caching appliances, the need to split data between high-performance, high-cost drives and high-capacity, low-cost drives is being eliminated. There are now new opportunities to collapse storage architectures to a single tier of high-capacity, low-cost storage by adding performance through centralized caching.

The rationale behind the centralized caching approach is simple:

- Deliver high-performance file access dynamically to the entire storage repository
- Reduce the overall cost of the infrastructure by eliminating over provisioning of storage controllers and disk drives
- Reduce the need to manually deploy and administer multiple tiers of storage

File acceleration through centralized caching delivers a number of significant benefits to data center managers and their end users. This is especially true when dealing with scalable applications that require rapid access to large libraries of objects such as photos, videos, or other user-generated content. In a typical centralized caching configuration, administrators only need to deploy a single tier of low-cost, high-capacity storage such as SATA drives instead of Fibre Channel drives.

A network-based caching appliance can deliver dramatic performance improvements for data-intensive application requests by caching frequently accessed data within the appliance. This offloads the requests from reaching the mechanical, drive-based storage, and shortens application response time 10-50 times. Since the caching appliance can access the entire storage repository, administrators are no longer required to manually move data around between tiers...often a troubling and cumbersome process. Now the caching appliance dynamically adds performance to the most frequently requested application data. A basic configuration is shown in Figure 3.

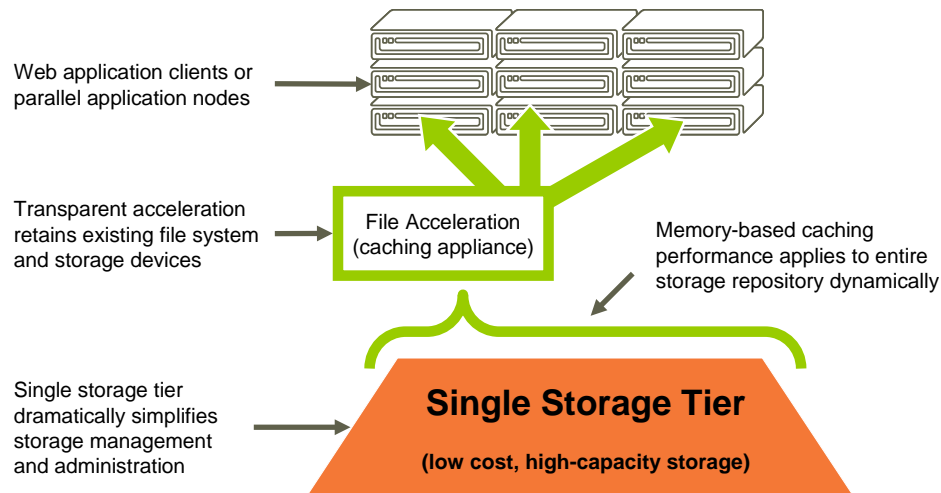


Figure 3: File acceleration with a caching appliance eliminates the need for excess storage tiers

Conclusions

For decades, data center managers have worked tirelessly to configure storage for performance using the only tools available, primarily mechanical disk drives. And while disk drives have succeeded over the years in incredible capacity increases, the speed and access time of the drive has not kept up. This has led to gross over provisioning of storage, and in an effort to stave off more top-tier storage spending, organizations have been adopting complex and hard-to-administer multi-tier storage architectures.

Now, by seamlessly integrating with existing file-based storage repositories, centralized caching enables the simplification and even elimination of multi-tiered storage. And administrators are guaranteed fast access to large file stores to boost application performance.