

## Creating Information Management Leverage with Veritas Storage Foundation by Symantec

**Date:** September, 2006

**Author:** *Brian Babineau, Analyst, Mark Bowker, Analyst*

**Abstract:** Veritas Storage Foundation and its Dynamic Storage Tiering (DST) facility delivers the unique ability to transparently automate file allocation and relocation onto different volumes within the same file system. Customized policies encompass flexible rules to provide a tiered storage solution across multiple volumes that aligns with individual business requirements.

### Introduction

There has been one concrete market trend as a result of the current Information Lifecycle Management fad - tiered storage. Customers realize that the value of data changes over time as information performance, accessibility and availability requirements ebb and flow based on business need. For example, an X-Ray of a car accident victim's spine may be constantly accessed during the first 6 months of patient care. After the initial 6 months, the X-Ray may not be accessed until the patient experiences another health issue. According to the HIPAA, the health care provider must retain this image for the life of the patient. The provider's IT department must find the most appropriate way to retain the X-Ray on an high-performance, highly available storage system for the first 6 months and then migrate the data to a lower cost storage device that can preserve it for an extensive period of time.

The risk in moving data between storage systems, or volumes on the same storage system, is preserving application access. If the patient starts experiencing back pains 2 years after the car accident, a doctor may want to recall the X-Ray and compare it with a more recent image. This means the older image must still be readily available even though it had not been accessed for 24 months. While some organizations can afford to purchase additional high performance storage to accommodate new data and retain files for long periods of time, this situation is not efficient due to the exponential expense of incremental capacity. Organizations can deploy storage tiers that have different performance and accessibility characteristics. Customers have invested in tiered storage architectures to archive non-transactional data onto lower-cost disk storage systems. In addition, customers have deployed various disk systems that emulate tape libraries to expedite backup and retrieval of information.

As more customers tier their storage, they need the appropriate solutions to identify, classify and manage the data across each tier. Dynamic Storage Tiering (DST), a feature of Veritas Storage Foundation, helps organizations create a file system that can store data across multiple volumes. The use of multiple volumes can also span storage systems, allowing organizations to create logical storage tiers as opposed to physical storage tiers. The key difference is that organizations can deploy heterogeneous storage systems with various RAID configurations and disk drive interfaces to construct the most economic, hardware agnostic infrastructure. Because it is one file system (supported across Solaris, HP-UX, AIX, Red Hat and SuSE), the data can be moved from one volume to another without changing the directory structure. The transparent movement between volumes and arrays preserves application access. Furthermore, the entire process can be automated via policies that dynamically move data up or down the tiers based on attributes such as its format, placement upon creation, date of last access, frequency of access (access temperature) and intensity of read-write activity (I/O temperature), enabling customers to store data more efficiently without significant manual intervention.

### Limitations of Current Tiered Storage Methodologies

As customers realize that the value of data changes over time, tiered storage strategies continue to be implemented in order to reduce capital and operating costs associated with the constant creation of more information. Initial tiered storage deployments were typically segregated at the device level, whereby a tier is composed of one or more homogenous storage systems. Each tier had one or more systems that shared the same performance, availability and acquisition cost characteristics. Tiers were differentiated by varying RAID protection schemes, disk drive interfaces (Fibre Channel, ATA, SCSI) and host connectivity (Fibre Channel, IP). Customers then needed to assign data to a tier

and a system. This forced customers to understand the business value and importance of information. A common scenario is to segregate mission critical applications, business critical applications and task critical or fixed content on a tier. ESG research suggests that the amount of non-transactional data to be archived over the next 5 years will increase almost five-fold, from 6,000 petabytes to 27,000 petabytes, suggesting that the latter tier will outgrow, in capacity terms, new data created by mission critical applications.

Segregating storage tiers by system does not help customers leverage investments in an existing storage infrastructure that may not fit into the scheme. Additionally, several hardware vendors have introduced devices that can support multiple disk drive interfaces and host connectivity options in one physical device. When customers outgrow one of these hybrid systems, they need to make a critical decision - do they continue tiering data and buy another hybrid system or do they standardize on one configuration for two separate tiers? These types of scenarios can force customers into using a single storage vendor, reducing pricing leverage.

Further complicating system based tiering is moving data once its value has changed. Because block-based storage systems may have different volume sizes, RAID schemes and other configurations, moving data from one tier to another can be extremely difficult. Ad-hoc scripting can be effective in environments that have expertise and in depth knowledge of the applications, but the effort to administer the scripting can be counterproductive and still requires some level of manual input. Other alternatives to move data are to invest in hardware based replication tools or work with one vendor and one storage system that support multiple configurations for tiering. Again, this situation leads to additional capital investment (replication software) and possible vendor lock-in. More importantly, as data is moved, file system directory information and volume mapping must be updated to ensuring application access. In some cases, updating this information requires downtime, a situation that many applications and users cannot afford.

Customers often utilize tiered storage for migrating non-transactional data from one system to another by deploying Hierarchical Storage Management (HSM) software. HSM implementations begin to address the data movement problem by creating policies that move inactive files from one storage system to another. However, HSM and other archiving tools leave links and meta-data, also called stubs, in the original file system. The file appears to the application as if still existing, but any application access requests will be delayed as the archiving software must retrieve the file from the secondary disk system and then put it back into the primary system. This process can slow application performance and users may not be able to access data in a timely manner. With HSM, the primary storage system must also maintain free space for archived files to be restored in case of an application request. This scenario leaves primary storage systems under utilized - the opposite intent of tiered storage. HSM can help organizations move data between tiers, but typically, HSM tools rely on simple policies that identify inactive files. Customers must be able to move data based on other information attributes such as access controls or security permissions in order to fully take advantage of tiered storage infrastructures.

Storage system vendors are also starting to offer compelling solutions that transparently move and store data within the storage system on different tiers. In these systems, tier one is typically synonymous with Fibre Channel (FC) drives and tier two with cost effective SATA drives. Storage system administrators define policies within the storage system that deliver the appropriate performance, capacity and bandwidth to the applications. The automated process transparently moves data within the storage system to different tiers and disks, enabling the storage system to continually run at peak performance, optimize application performance and efficiently utilize disk capacity across the entire storage system. However, block based hardware tiering lacks some flexibility as they do not have visibility into the files or directories; potentially leading to vendor lock-in.

### **Leveraging Tiered Storage with Veritas Storage Foundation**

In block based storage environments, customers rely on volume and file system management tools to abstract the physical disks from the application. Applications are built on file systems, which store data on volumes that reside on storage systems. However, current tiered storage implementations are bound by the physical device. Symantec, a reputable supplier of volume management and file system management software, enhanced Veritas Storage Foundation, by adding Dynamic Storage Tiering to dynamically abstract data movement between multiple storage tiers. Veritas File System (VxFS) can store data across multiple volumes that are created by Veritas Volume Manager (VxVM). Customers can now assign a tier to a series of volumes that reside on multiple storage systems. Volume based tiering still requires customers to understand specific application performance and accessibility requirements. For example, data that requires high availability and peak disk performance would require volumes that store and access data on a portion of a storage system that has high RPM, low capacity FC drives and is configured with RAID 1. Less critical data may only require volumes that leverage storage systems configured with high capacity SATA drives across with RAID 5 or other protection schemes. Multiple volumes within a tier constitute a placement class. Each

class is now a logical storage tier. Placement classes are used instead of individual volumes, enabling customers to add incremental volumes (and associated capacity) to a specific tier.

To automate the storage and movement of data across placement classes, customers define policies with Storage Foundation's DST. Policies are a set of defined rules to allocate and relocate files onto the appropriate placement class. Because all volumes receive data from a single file system, the file system can enforce policies upon creation or on a scheduled basis. Policies can utilize file attributes such as format, application, owner, location, size or any combination of these attributes. When a policy is executed, multiple files can be moved from one volume to another. For scalability and consistent information management practices, policies can be applied to multiple file systems that store data in separate placement classes. For example, in an Oracle Data Warehouse application that has 5 years of historical data, a customer may only want the most recent 4 quarters on tier 1 storage (and the oldest 16 quarters on tier 2 storage). A customer can then define a policy that runs every quarter and moves each Oracle data file from the 5<sup>th</sup> quarter to tier 2 storage, thus maintaining only the 4 most recent quarters on tier 1 storage. Another policy may relocate old database log files from one volume to another. Transactional database files can leverage one volume that is mirrored and replicated and store the log files on another volume that stores data on low cost ATA disk. The latter policy can be automated further as Symantec has integrated DST with Oracle and DB2 file system structures to retain highly accessed files on volumes stored on fast disk and less active data on slower storage.

As customers utilize Storage Foundation's DST, they do not have to worry about disrupting application access as the VxFS directory structure does not change; the storage is virtualized and VxVM handles the physical file placement, based on policies, to multiple storage volumes. Moreover, customers can still perform all management tasks such as volume replication, expanding/reclaiming file systems and snapshotting data for backup operations. Most importantly, customers can tier storage, maintaining the abstraction between the application and physical storage systems without compromising information accessibility. Symantec's multi-volume file system architecture effectively manages file system metadata such that the loss of a physical disk in a single tier (placement class) will not affect access to data on other storage tiers. Enhancing data availability in a tiered storage architecture is a key requirement for most customers.

### **The Bottom Line**

Tiered storage solutions are a good approach to storing data, but customers can still be confined to homogenous storage systems that limit benefits. In cases where customers can deploy heterogeneous storage systems across tiers, moving data can be challenging as each system configuration may be different. Tactical solutions, such as HSM or content archiving products, can alleviate data movement challenges but may reduce storage resource utilization and cause data retrievals to be extremely disruptive.

With broad knowledge of file system and volume management, The Dynamic Storage Tiering feature of Veritas Storage Foundation helps customers build logical storage tiers, automate data movement between tiers and ensure consistent application access. The end result is that customers can deploy heterogeneous storage systems, build tiers based on data performance and accessibility requirements and then begin storing data more efficiently. The ability to create flexible policies, add volumes to tiers and maintain application access can help organizations align business requirements with the vast amounts of information being created.

While customers must evaluate which tiered storage method, system based or volume based, will make sense for the business, existing Storage Foundation customers can utilize DST as a feature of the portfolio, not a separate product. Additionally, Symantec offers a DST Consulting practice to assist customers with architecture recommendations and deployment. Assessment utilities provide the customer with a "what-if" analysis view of their application data. Leveraging DST requires that customers deploy the industry leading VxVM and VxFS in lieu of the native operating system volume managers. Customers may weigh this investment versus hardware vendor lock-in, investments in replication or HSM software or maintenance of scripts to leverage their storage tiers. For the most part, this should be a relatively easy decision to make.