

Real-Time Storage Tiering for Real-World Workloads



INTRODUCTION

The performance of many of today's server-based applications is limited by the relatively slow input and output (I/O) to disk storage. A common way to mitigate this limitation and improve performance is to replicate active data from disks to higher-speed memory in a cache in servers and storage area networks (SANs).

While caching content does help, it too has limitations, however. Caching primarily accelerates reads, and the size of the cache is normally only a small fraction of the disk capacity available. This makes it necessary to change the content constantly, and the algorithms employed to determine which content to cache may fail to deliver meaningful performance gains for critical applications.

Advances in technologies have created a new layer or tier of storage between cache memory and traditional hard disk drives: the solid-state drive or SSD. SSDs use high-speed flash memory and offer capacities on a par with those of hard disk drives (HDDs). SSDs can be used either to create a very large cache (that substantially improves "hit" rates) or a very fast and full tier of disk storage. Either way, the performance gains can be significant for many applications.

The purpose of this white paper, intended for both business and technical decision-makers, is to describe the use of flash memory in SANs, in general, and how Dot Hill Systems has implemented SSDs in its next-generation SAN. The document is organized into three sections followed by a brief conclusion. The first section explores storage tiering as a technology, both in theory and in practice. The second section describes how Dot Hill has implemented storage tiering with advanced RealStor™ technology in the AssuredSAN Pro 5000 Series. The third section provides an assessment of the performance gains that should be expected from using Dot Hill's autonomous, real-time storage tiering technology.

STORAGE TIERING: THEORY VS. CURRENT PRACTICE

Every application is limited to some extent—sometimes significantly—by I/O to disk storage. Consider this: accessing data in the server's memory takes only 10 nanoseconds while accessing data on a hard disk drive takes about 10 milliseconds—a difference of 1,000,000 times or six orders of magnitude. In between is solid state storage using flash memory with an I/O latency of 3 to 5 microseconds, making them 2000 to 3000 times faster than HDDs.

The performance impact of this difference varies depending on the application. I/O-intensive applications, especially those accessing information randomly in a database, experience the most adverse impact with HDDs. Read-intensive and computation-intensive applications, especially those that access information sequentially, are generally the least impacted. In between are most applications with a mix of random and sequential input (reads) from and outputs (writes) to disk storage.

The traditional measure of performance for transaction-oriented workloads is the number of I/Os per second, or IOPS. Historically, certain rules of thumb have evolved to estimate the number of IOPS

needed for a particular application, such as the number of IOPS per mailbox in an email application. This approach completely misses the temporal dimension of performance, however, which is how the workload changes over time.

Consider the case where transactions generated by server applications are actually being generated by human interaction. This is the case when a Web server is responding to what is trending on the Internet, or when an email server is processing messages that are being sent and received. The point is that the data generating the most I/Os is the data that is “hot” at that moment in time. In other words, the workload is not uniform across the entire data set; rather, the workload is focused on a small subset of the data set, and that subset changes constantly. A storage tiering system capable of delivering a significant improvement in IOPS must, therefore, be able to adapt to these changing workloads in real-time.

Storage tiering holds the potential to improve performance for all applications by locating “Hot Data” in the fastest tier consisting of solid state drives. Hot Data is that which is needed by the applications currently running on the servers. The next tier consists of traditional hard disk drives, and the drives used are normally fast-spinning to deliver the best possible performance for this medium. Some solutions offer a third tier for archiving infrequently used data on lower-cost, lower-performance (but often higher capacity) HDDs. The better solutions also include a memory-based cache, enable the storage tiers to be pooled and virtualized, and provide data protection with support for various RAID (Redundant Array of Independent Disks) configurations.

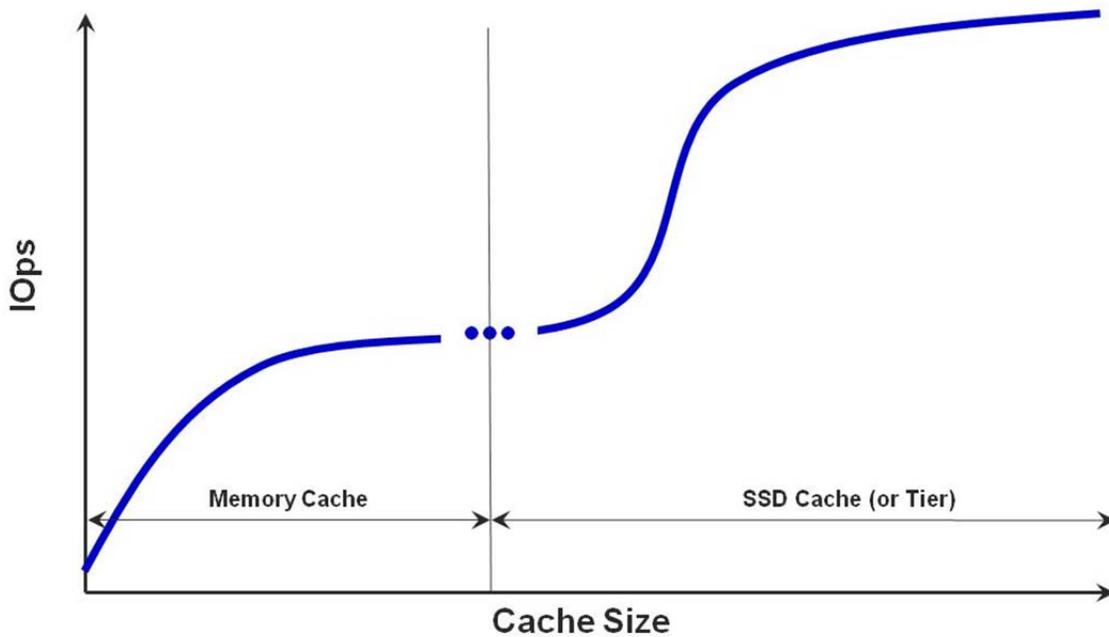
The advantage of storage tiering is that it offers an effective balance of high performance (with the very fast SSDs) and low cost (with the relatively inexpensive HDDs). In the distant future when the cost of SSDs approaches the cost of HDDs on a per-byte basis, it may be possible to deploy a purely SSD-based SAN. But that day is a long way away, which makes storage tiering the most cost-effective way of achieving significant performance gains today.

In theory, storage tiering always delivers excellent results, improving the performance of all applications. This is not the case in practice, however. The biggest challenge is identifying the Hot Data to move to the SSD tier because Hot Data can be fleeting, changing by the hour or even by the minute for each application. This challenge is exacerbated in a storage area network where every application being served has its own individual and fluid set of Hot Data.

In practice, storage tiering solutions suffer from two common and sometimes significant limitations. The most crippling limitation is the frequency of the data migration between tiers. Many SAN solutions are capable of moving data to the SSD tier on a daily basis only owing to the disruptive nature of the migration itself. Because halting an application to migrate the data is normally unacceptable, the migration must occur during periods of minimal or no demand, which is usually overnight. Such a daily or nightly migration may be sufficient for applications where the datasets can be known in advance (and fit entirely in the SSD tier), but it is of little or no value for all other applications.

The second limitation involves how both the system itself and the data migration process are managed. Most systems require the IT department to determine which data sets will be migrated at what times between the tiers. This might be part of a regular schedule (e.g. a certain day of the week or month), or on an ad-hoc basis as needed (e.g. an event or project scheduled for the following day). Either way, a sub-optimal determination inevitably results in a minimal improvement in performance.

Overcoming these limitations will extend the performance gains afforded by caching into the SSD storage tier as shown in the diagram below. Note how the practical size of a cache limits its ability to achieve gains beyond a certain point. The SSD's far larger capacities (in the hundreds of Gigabytes range), which are also available at a significantly lower cost per byte than cache memory, make it possible to scale the performance gains considerably. But this will only be the case if the migration of data between tiers can be made sufficiently intelligent and dynamic to keep pace with the real-time changes in Hot Data.



Caching technology inevitably reaches a point of diminishing return, and this is where SSDs, with their far higher capacities, can take over to extend the performance gains into either a substantially larger cache or a much faster storage tier.

REAL-TIME STORAGE TIERING WITH REALSTOR™

Dot Hill Systems has advanced the state-of-the-art in storage tiering with the introduction of the AssuredSAN™ Pro 5000 Series. This next-generation offering utilizes Dot Hill's patent-pending RealStor technology that integrates five separate capabilities, which together extend the performance gains achievable as the storage tiers scale.

The five capabilities, each explained in turn in this section, include:

- Autonomic Real-time Tiering
- Thin-Provisioning
- Rapid Rebuild
- Disruptively Simple User Interface
- SSD Flash Cache Option

Autonomic Real-time Tiering

Dot Hill's RealTier™ technology overcomes the two major limitations found in most tiered storage systems today by (1) automating the migration of data in (2) real-time. The system virtualizes both the SSDs and HDDs at the sub-LUN level using 4 MB pages distributed across multiple RAID sets. Intelligent RealTier algorithms then continuously monitor I/O access patterns and automatically move Hot Data to the SSDs to maximize I/O operations and, therefore, improve performance of the aggregate application workload.

RealTier utilizes three separate processes, all of which operate in an autonomic manner in real-time, including:

- *Scoring* to maintain a current page ranking on each and every I/O using an efficient process that adds less than one microsecond of overhead. The algorithm takes into account both the frequency and recency of access. For example, a page that has been accessed 5 times in the last 100 seconds would get a high score.
- *Scanning* for all high-scoring pages occurs every 5 seconds, utilizing less than 1.0% of the system's CPU. Those pages with the highest scores then become candidates for promotion to the higher-performing SSD tier.
- *Sorting* is the process that actually moves or migrates the pages: high scoring pages from HDD to SSD; low scoring pages from SDD back to HDD. Less than 80 MB of data are moved during any 5 second sort to have minimal impact on overall system performance.

Thin-Provisioning

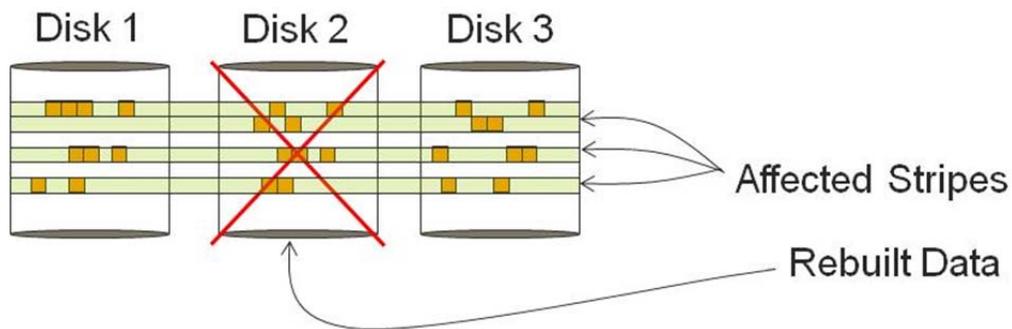
The increase in the volume and velocity of data can cause storage costs to exceed available budgets without some prudent provisioning. With Dot Hill's RealThin™ thin-provisioning capability, IT managers can dedicate available storage space to volumes only when actually needed and add storage capacity transparently to any application, also as needed. This approach enables organizations to minimize capital expenditures initially and scale capacity incrementally over time. RealThin also enables LUN (volume) size to be configured independently of physical disk space and supports LUNs up to 128 TB.

With any thinly provisioned configuration, it is important to know when physical storage capacity is running low. For this reason, RealThin enables IT managers to establish the alert thresholds needed to receive adequate warning. These alerts are especially critical when capacity is becoming "overbooked" or configured larger than the available physical space. Should physical space ever become at imminent

risk of being exhausted, RealStor switches to write-through mode to ensure that all data is able to be written to disk.

Rapid Rebuild

When a disk drive fails in a RAID-protected storage volume, it exposes the application to data loss until the drive can be replaced. Dot Hill RealQuick™ rapid rebuild minimizes this exposure by accelerating the recovery time needed to fully rebuild a failed drive in a RAID set. Because RealStor spreads LUNs across multiple RAID sets, rebuilding a single RAID set affects only a fraction of all disk I/Os. With less work involved during the rebuild, the disk(s) affected return to full operation rapidly, resulting in volumes becoming fully fault tolerant more quickly. Because the acceleration afforded by RealQuick is directly proportional to amount of unused disk space, a trade-off may be required with RealThin's thin-provisioning capability.



By spreading LUNs across multiple RAID sets, Dot Hill's RealQuick rapid rebuild is able to recover from disk failures more quickly with minimal impact on overall system performance.

Disruptively Simple User Interface

Most tiered storage solutions place the burden on administrators to determine how to configure the controllers, including the creation of all storage pools and the specification of all low-level configuration details. Dot Hill minimizes these tasks by performing most of them automatically and makes the remaining ones as simple and intuitive as possible.

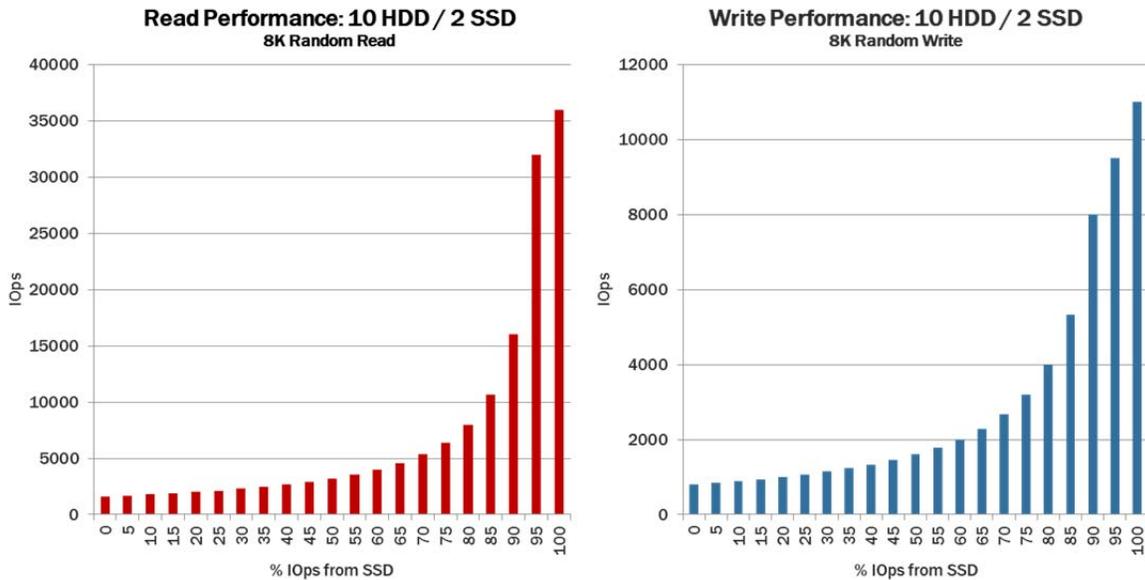
Dot Hill's RealPool™ auto-pooling technology automatically creates all storage pools, thereby eliminating the usual difficulties involving the determination of RAID levels, chunk sizes, which disks to use for which vdisks, etc. Ongoing management tasks are then all streamlined so that each can be performed quickly and efficiently with no need to create and maintain complicated policies. The system is remarkably easy to navigate and use, and all of the information required to perform any task is always readily available.

SSD Flash Cache Option

SSD Flash Cache is a lower cost option for accelerating read-intensive applications. As the name implies, this option uses the flash memory in the solid state drive to extend the system’s cache from 4 GB to the full capacity of the SSD, which is typically hundreds of Gigabytes. Flash Cache is, therefore, able to deliver substantial improvement in performance for read-intensive applications with half the number of SSDs needed for storage tiering. It works by copying or replicating Hot Data to the SSDs from the RAID-protected HDDs—automatically and in real-time, just as with tiering. Data is then read from the SSD Flash Cache, and written to both the SSD Flash Cache and the HDD tier.

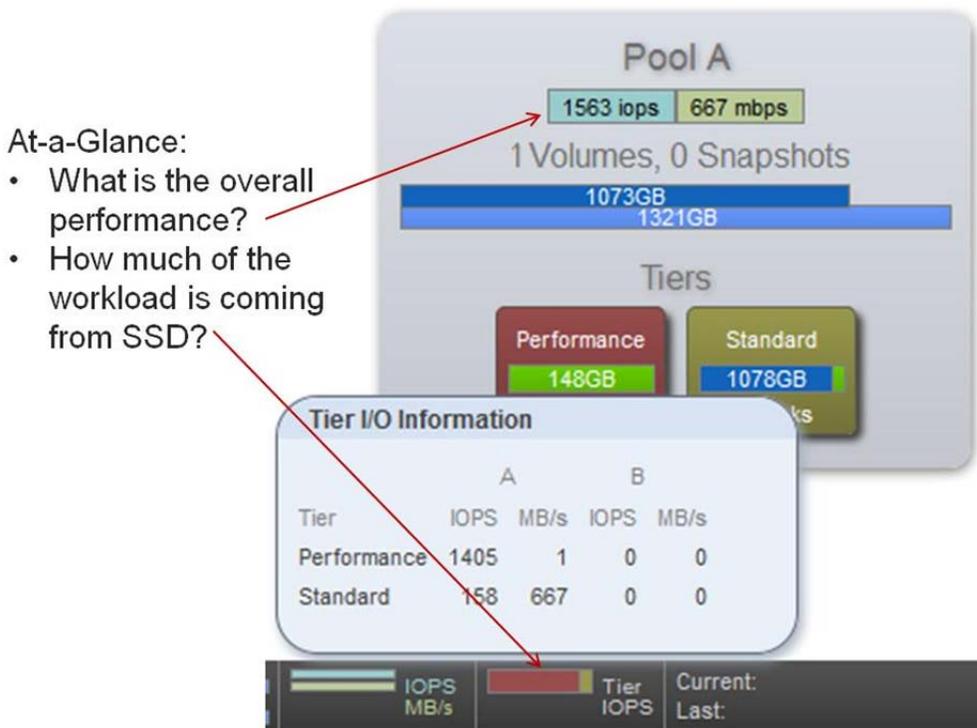
SSD TIER PERFORMANCE

Dot Hill’s AssuredSAN Pro 5000 Series with RealStor technology delivers up to 100,000 random read and 32,000 random write I/Os per second. The chart below shows the potential performance gains achievable with tiered storage. Naturally, the higher the “hit rate” in the SSD tier, the higher the gains. But even a conservative hit rate of 70% (easily attainable in most environments) can deliver a three-times improvement in application performance. Far greater performance gains can be realized when the SSD tier handles 80% or more of the I/O load, even though that tier represents only 5-10% of the system capacity.



Thorough testing of the AssuredSAN Pro 5000 Series shows the significant gains possible in read and write IOPS, especially in environments where the application mix can achieve an SSD tier “hit rate” of 80% or more.

With autonomic, real-time tiering, noticeable performance gains are experienced almost immediately, even with challenging applications. But because performance gains are dependent on the SSD tier hit rate, RealStor's management system includes an application that enables IT managers to tune the performance. The application constantly monitors IOPS, displays pertinent performance statistics, and optionally exports these to Excel-compatible .csv files for reporting, archiving and trending analysis needs. Changing the mix of applications on servers and/or the distribution of data across the SAN can increase (or decrease!) the hit rate.



RealStor's management system includes an application that monitors and displays pertinent performance statistics to enable IT managers to track and optionally fine-tune the performance gains achievable with tiered storage.

OTHER SALIENT FEATURES OF THE ASSURED SAN PRO 5000 SERIES

In addition to the capabilities provided by RealStor, the AssuredSAN Pro 5000 Series offers these other capabilities:

- Scalable to 240 drives in ten 2U shelves
- SSD tier employs RAID level 1 with mirroring for data protection
- HDD tier employs RAID level 6 with block-level striping and double-distributed parity to provide good performance with fault tolerance of two drive failures
- SSD/HDD storage pooling with support for up to 1024 LUNs

- 4 GB of cache with SimulCache™ low latency cache mirroring and EcoStor™ battery-free cache backup
- 8 Gbps Fibre Channel or 10 Gbps iSCSI interfaces
- Intuitive, browser-based Storage Management Console
- Compliance with NEBS and MIL-STD-810 standards
- AssuredSnap™ to take point-in-time snapshots for backup and recovery processes
- AssuredCopy™ to create full volume copies or backups of disk volumes, and quickly restore whole volumes, folders or individual files
- AssuredRemote™ to easily replicate volumes between two Dot Hill storage systems at different locations as part of a full disaster recovery plan

CONCLUSION

Many if not most applications today are I/O-constrained, which limits their performance when using traditional HDDs, whether directly attached or in virtualized storage area networks. Caching helps improve performance, but fails to scale because it quickly reaches the point of diminishing return on the investment.

SSD technology that uses fast flash memory increases I/O rates by 2000 to 3000 times compared with HDDs, and a combination of the two in a tiered storage configuration offers the most cost-effective way of achieving significant performance gains today. But most tiered storage systems deliver only marginal improvements in performance because they cannot accommodate dynamic workloads where the Hot Data is constantly changing.

Dot Hill's AssuredSAN Pro 5000 Series with RealStor technology overcomes the limitations of other tiered storage solutions to deliver a combination of industry-leading high performance and high availability on a scalable platform with advanced features that are easy to manage with minimal IT resources. Intelligent RealTier algorithms operate continuously, automatically and in real-time to deliver the best possible improvement in performance for today's rapidly-changing real-world workloads. And the low initial capital expenditure, combined with streamlined management, result in a surprisingly low total cost of ownership, while the performance gains yield a high return on investment.

To learn more about how your organization can benefit from real-time storage tiering using the AssuredSAN Pro 5000 Series, please visit Dot Hill Systems on the Web at www.dothill.com or call 303-845-3200.

Scan for more
details.



© 2012 Copyright Dot Hill Systems Corporation. All rights reserved. Dot Hill Systems Corp., Dot Hill, the DotHill logo, AssuredSAN, AssuredRemote, RealStor, RealTier, RealSize, RealQuick and RealPool are trademarks or registered trademarks of Dot Hill Systems. All other trademarks are the property of their respective companies