Hybrid Storage Arrays Lead the Way

TECHNOLOGY PAPER

Hardware and software options make for a cost-effective HFA system solution

Businesses of all sizes and in virtually every industry need to solve a common problem—where and how to store their business data. Data storage technology has advanced significantly, and determining the ideal storage solution usually means sorting through a barrage of conflicting information. However, one trend is obvious: there is a lot of hype around the perceived benefits of solid state drives (SSDs) for buisness applications. This often leaves customers evaluating multiple options—traditional hard drives, all-flash arrays (AFA), and hybrid flash arrays (HFA).

Clearly, SSDs provide increased performance measured in IOPS (input/output operations per second) and latencies (application response times) when compared to traditional hard drives. And for business-critical applications, the IOPS performance of SSDs make AFA solutions the most compelling. Furthermore, SSDs prices have dropped in recent years, which has led to an increase in the number of AFA technologies available in the market.

But while the AFA market is projected to have the highest growth rate over the next four years, HFAs (combination of SSD and HDDs) remain the largest segment of the market. For those customers that demand AFA performance, Seagate's AFA solution provides one of the lowest $/IOPs in the market. But for budgets that can't quite make that work, the Seagate HFA becomes the more attractive solution.

A false dichotomy: traditional arrays vs. AFAs

In storage buying decisions, most business customers typically consider the amount of storage required. For example, imagine that an SMB with 50TB of data is budgeting for a new storage purchase. Imagine, too, that this particular SMB is planning to have up to 100TB of data within 5 years. If the SMB customer approaches this decision from a purely capacity standpoint, they could simply purchase Seagate® Exos™ 10TB hard drives and start building a system. In fact, twelve 10TB hard drives installed in an Exos X 2U12 RAID enclosure would provide 120TB of raw capacity. Apply RAID protection in the form of a RAID 6 (10+2) configuration and the SMB would have up to 100TB of configurable system capacity.

In some environments—such as archival data, backup target, or applications that don't require a lot of IOPs—the Exos 10TB is a compelling, high-capacity, and economical solution. However, a storage system configured in this manner would be limited in performance to the spindle count and number of IOPS each hard drive could deliver. In this example: 100 IOPS per hard drive × 10 hard drives with RAID 6 equals 1000 IOPS. Unfortunately for higher performance business applications, such a configuration would be lacking albeit the most economical.

Hybrid Storage Arrays Lead the Way

If this same SMB customer wanted to solve the problem with a performance solution and sized their initial 50TB of storage using Nytro® 3.2TB SSDs in an AFA, it would take approximately 16 of these SSDs to reach 50TB (with RAID 5). This solution obviously does not include a plan for data growth. However, from a performance perspective, the Nytro solution would be much better than with a pure Exos 10TB hard drive solution and more desirable for the most business-critical solutions.

A smart compromise: hybrid flash array technology

Aside from capacity and performance, however, there is a clear factor that drives most IT expenditure—cost. AFAs provide much higher performance, but at a much higher cost. By contrast, hybrid array technology combines flexible hardware options with powerful software. Customers can have the amazing performance needed for specific applications along with the cost/TB they need to stay within budget.

Seagate offers a complete portfolio of hard drives and SSDs in a full range of capacity points, plus enclosures to allow for the combination of 2.5-inch SSDs along with 2.5- and 3.5-inch hard drives. Customers can design combined systems with combinations for the RAID head and expansion JBODs to reach their capacity and performance goals. For example, a customer might want to start with an Exos X 2U24 RAID enclosure, add some SSDs, hard drives, or some combination thereof, and then increase their overall capacity by adding a 2U12 system with 10TB drives.

Real-time tiering? Yes, it's essential.

Before delving into Seagate software in more detail, it is important to understand a basic premise of hybrid array solutions. Not all customer data is ***hot*** all the time—it does not always require the performance that SSDs can offer. As an example, Figure 1 below shows a 3D graph depicting host I/Os across an entire pool of available capacity.

At first glance it might appear there is a fairly even distribution of I/O across the entire LBA range; however, if we graph this data as a plot of cumulative IOPs vs. the total capacity (Figure 2), we can see that 80% of the IOPs are coming from only 6.6% of the total capacity. In this specific example the overall capacity of the storage pool was 30TB; 6.6% of that number equates to 1.98TB.

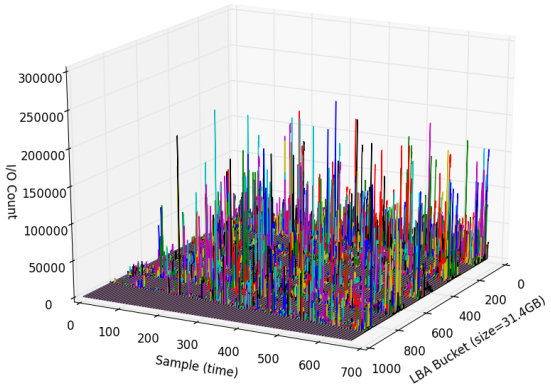

Figure 1. 3D graph depicting host I/Os across an entire pool of available capacity

Real-time tiering automatically tiers between SSDs and hard drives–as opposed to a batch-migration process. Seagate's tiering software recognizes incoming data patterns, and distributes data on the most appropriate tier to provide optimal performance.

Hybrid Storage Arrays Lead the Way

4MB paging

Seagate's tiering software is based on a concept called *paging*, meaning user volumes are logically broken down into small, 4MB chunks called pages. Pages are then ranked based upon a sophisticated algorithm, which is used
to efficiently select and allow good pages to move between tiers. The result is that pages migrate between tiers automatically, meaning that I/Os are optimized in real time (Figure 4).

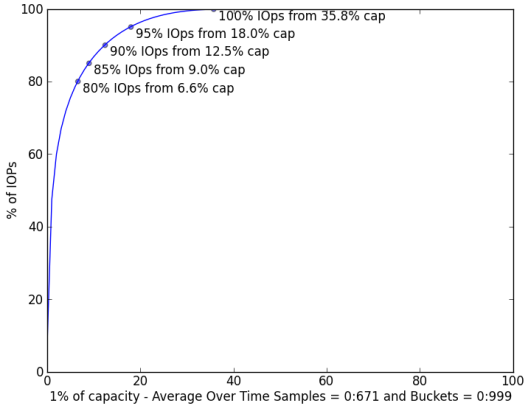

Figure 2. CDF Plot of I/Os over entire pool capacity

Figure 3 shows that three pages have been moved from the Enterprise SAS Tier to the SSD Tier; one page from the midline SAS Tier to the SSD Tier; and a single page has been moved down from the SSD Tier to the SAS Tier.

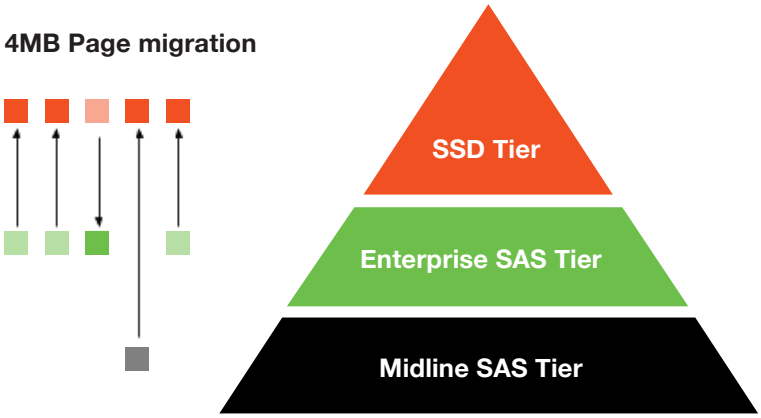

Figure 3. Pages moving between tiers

This tiering is done at the sub-LUN level which provides efficient data movement with a minimum of CPU and memory resources. Movement can happen in real time rather than in offline batch movements, providing a more reactive, efficient means of helping to ensure that the most active data is promoted to the highest available tier. Many competitive offerings require a batch migration done after a 24-hour period. Most customers know that a 24-hour period is a long time in terms of what data is the most active and critical to their business. Often with these batch migration approaches (by the time the data has been migrated) it is no longer as critical and therefore customers are never able to fully utilize their SSD investment.

Key concepts:

The tiering algorithm runs every 5 seconds. The algorithm ranks, scans, and migrates pages.

- Pages are ranked according to access patterns
- A scan looks for highly-ranked pages
- Highly ranked pages are migrated up/down
- Pages are only migrated down if room is needed for a highly-ranked page
- Only 80MB migrated from any one 5-second interval to avoid degrading system throughput

Hybrid Storage Arrays Lead the Way

SSD read cache

In addition to tiering, Seagate offers a second method to leverage SSDs—SSD read cache. Read cache is a special-purpose disk group dedicated to hold volatile copies of data within one or more SSDs. Read cache also has the following advantages:

Key Concept:
Read cache keeps copies of data that already exist in the hard disk tier; therefore, it doesn't require more SSD disks to be fault-tolerant. The SSD disk(s) provide the performance and the HDD disks provide the fault tolerance. This is a cost-effective method of providing performance.

- Improves performance in read-centric workloads where maximum IOPS is crucial and is intended for customer workloads that are highly read-intensive. Note this does not benefit writes.
- Requires fewer SSDs when compared to a dedicated SSD tier. A single SSD disk can be used as a read cache for a storage pool, whereas an SSD tier requires multiple disk to provide fault tolerance.
- Takes advantage of tiering-like algorithms so that only active pages are copied into the read cache. Active pages can be from any volume within the pool.

The read cache becomes active when the controller cache is saturated. The controller cache software gives hints to the controller software about which pages it would like to keep, but does not have space for in the controller cache. The system then copies these pages into the SSD read cache. Reads are served from the SSD; writes are written to both SSD and HDD.

Test Results

Seagate worked with a third-party vendor to conduct a series of tests to compare how its hybrid flash array performed against a traditional all hard drive array. There were several workloads tested: Microsoft SQL Server, Microsoft Exchange, and a VMware workload environments with mixed applications running. The baseline all-hardware test was performed first, followed by a read-cache test. A full SSD tiering test followed.

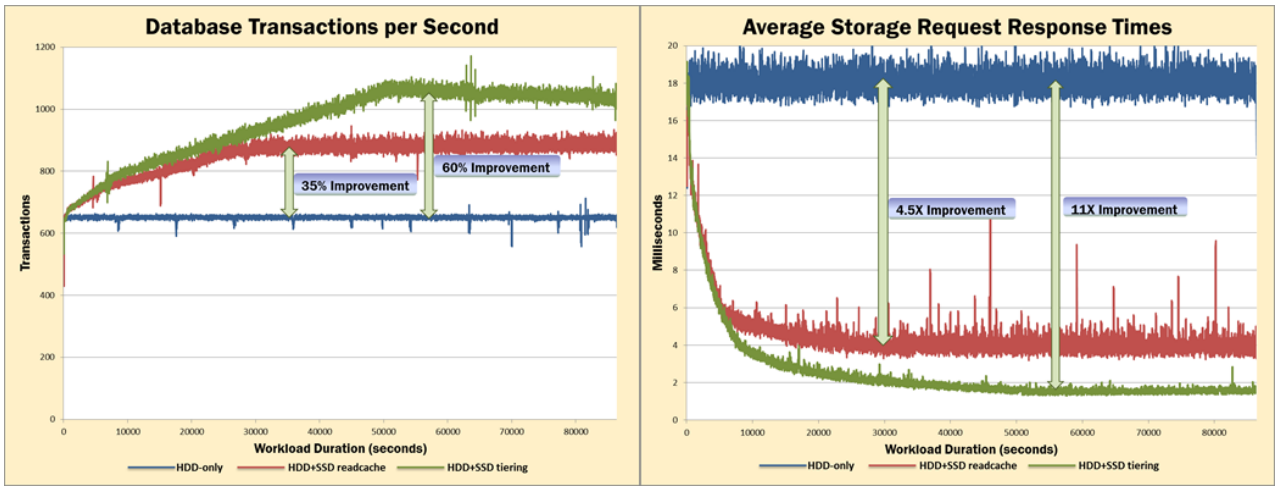

Figure 4. Results from Microsoft SQL OLTP Testing

Hybrid Storage Arrays Lead the Way

Figure 4 shows the results obtained when comparing the baseline all hard drive configuration (blue), to the read-cache configuration (red), and the full SSD tier testing (Green). Adding the SSDs, along with the tiering software, resulted in increased database transactions per second, as well as lowered response times.

It should be noted, that aside from provisioning the SSDs, as either read-cache or an SSD-tier, there is no additional user intervention required. Seagate software automatically analyzes the incoming workloads, and makes intelligent decisions on where to place data to provide the maximum performance. As an example, most database workloads are small, random transactions. Those workloads are generally IOPS—intensive, and better suited for an SSD. Therefore, when the Seagate array detects this workload, it automatically migrates that data from the hard drive up to an SSD tier. This migration happens immediately and reaches a steady-state in less than 14 hours. In comparison, competitive batch-migration solutions require significantly more time before providing performance benefits.

Conclusion

The use of SSDs in arrays has clearly increased and led to a number of innovative AFA solutions in the marketplace. Coupled with decreasing SSD costs, customers are finding it more practical to look at AFA storage arrays – particularly for their business-critical applications. Evaluating array access patterns however, shows us that in many instances, the majority of data most frequently accessed is a smaller percentage of the overall storage capacity. On the other hand, with Seagate's unique software features and flexible hardware options, customers can implement cost-effective, high-performance hybrid flash arrays.

For more information, on the Seagate Storage Systems, go to:
seagate.com/enterprise-storage/systems/

seagate.com

SEAGATE

AMERICAS Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, 408-658-1000
ASIA/PACIFIC Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, 65-6485-3888
EUROPE, MIDDLE EAST AND AFRICA Seagate Technology SAS 16–18, rue du Dôme, 92100 Boulogne-Billancourt, France, 33 1-4186 10 00