Evaluation Report: HP 3PAR StoreServ 7450 with 8Gb Express Writes

Evaluation report prepared under contract with HP

Executive Summary

The availability and deployment of flash storage in the industry is expanding by leaps and bounds. I/O performance numbers measured in multiple gigabytes of data transferred per second and hundreds of thousands to millions of I/O operations per second across corporate SANs are becoming increasing achievable and common. However, there are situations in which raw bandwidth and IOPs provide only so much benefit to a compute infrastructure.

I/O request latency is being increasing identified as a key component in application performance. Highly transactional applications, particularly those in which subsequent transactions are dependent on prior ones, or transactions that operate against very dynamic data are beginning to demand response times in microseconds. Robust flash-based storage, such as the HP 3PAR StoreServ 7450 is able to deliver this type of performance out of the box, along with high IOPs and bandwidth.

Demartek recently published a report evaluating the performance of the All-Flash HP 3PAR StoreServ 7450 with 16Gb FC targets[1], in which we strongly advocated end-to-end 16Gb FC infrastructure when deploying high-performance all-flash storage. However, there are cases where upgrading storage targets or other SAN infrastructure is not feasible. For these situations, HP has developed 8Gb Express Writes, a technology which accelerates write I/O to the HP 3PAR StoreServ array (available across all shipping models), improving the already impressive performance of these storage systems. HP commissioned Demartek to evaluate the All-Flash HP 3PAR StoreServ 7450 storage system with the 8Gb Express Writes feature installed and enabled to confirm these claims of enhanced performance. Demartek ran an industry standard online transaction processing workload on a server connected to the HP 3PAR storage system with a single

[1] Demartek, "Demartek Evaluation of HP 3PAR StoreServ 7450 16GFC All-Flash Storage", http://www.demartek.com/Demartek_HP_3PAR_StoreServ_7450_16GFC_2015-03.html (March 30, 2015)

8Gb FC HBA. We then activated Express Writes and compared the difference in performance, focusing on I/O latency.

Key Findings

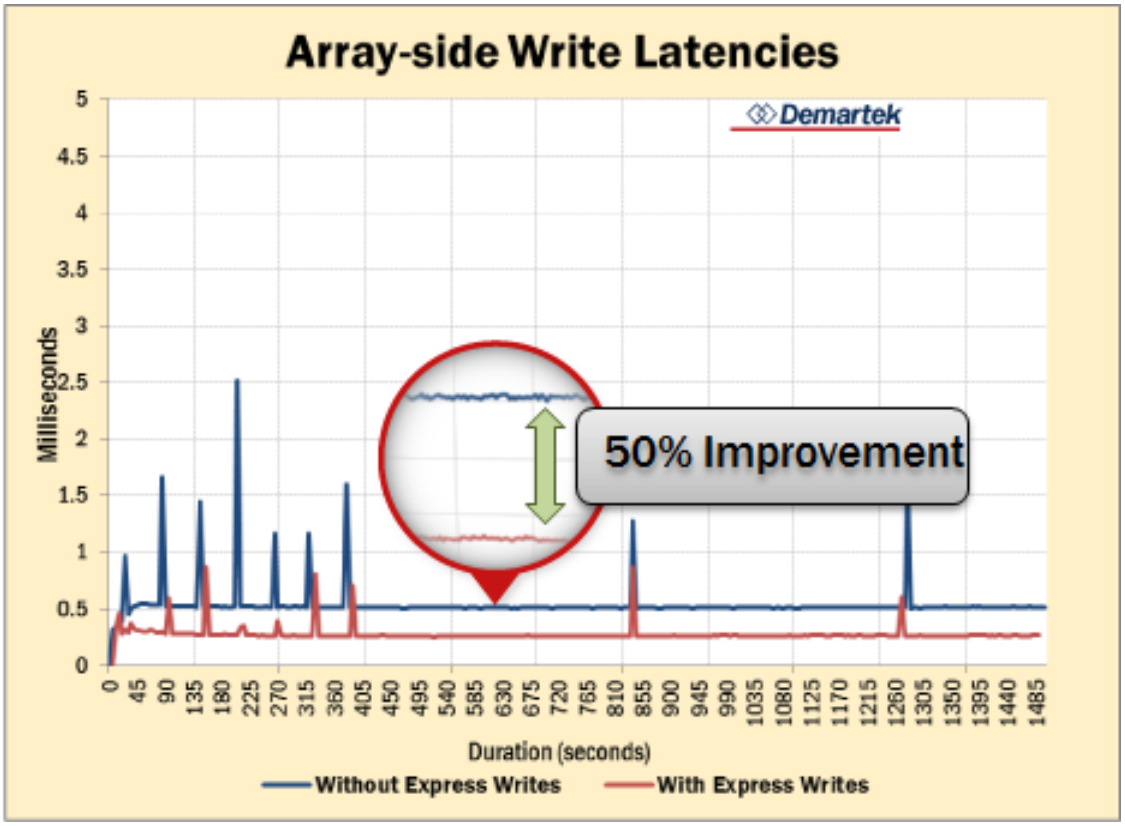

Figure 1– Latency improvements with 8Gb Express Writes

We found that 8Gb Express Writes, a feature available to 8Gb FC targets on HP 3PAR StoreServ systems running OS release 3.2.1 or better, decreased the latency of database writes in a TPC-E workload on the storage system by up to 50% (Figure 1). Deploying Express Writes was a simple operating system upgrade to the storage system and was accomplished without impact to the live system involved. In our opinion, Express Writes is a simple performance enhancement that shows immediate beneficial results.

The Advantage of HP Express Writes for 8Gb Fibre Channel

HP 3PAR 8Gb Express Writes is a built-in HP 3PAR OS write acceleration feature that optimizes the storage system's CPU utilization, delivering fewer interrupts per I/O transaction to increase throughput and IOPS, and reduce latency for write operations. 8Gb Express Writes is part HP 3PAR OS 3.2.1.

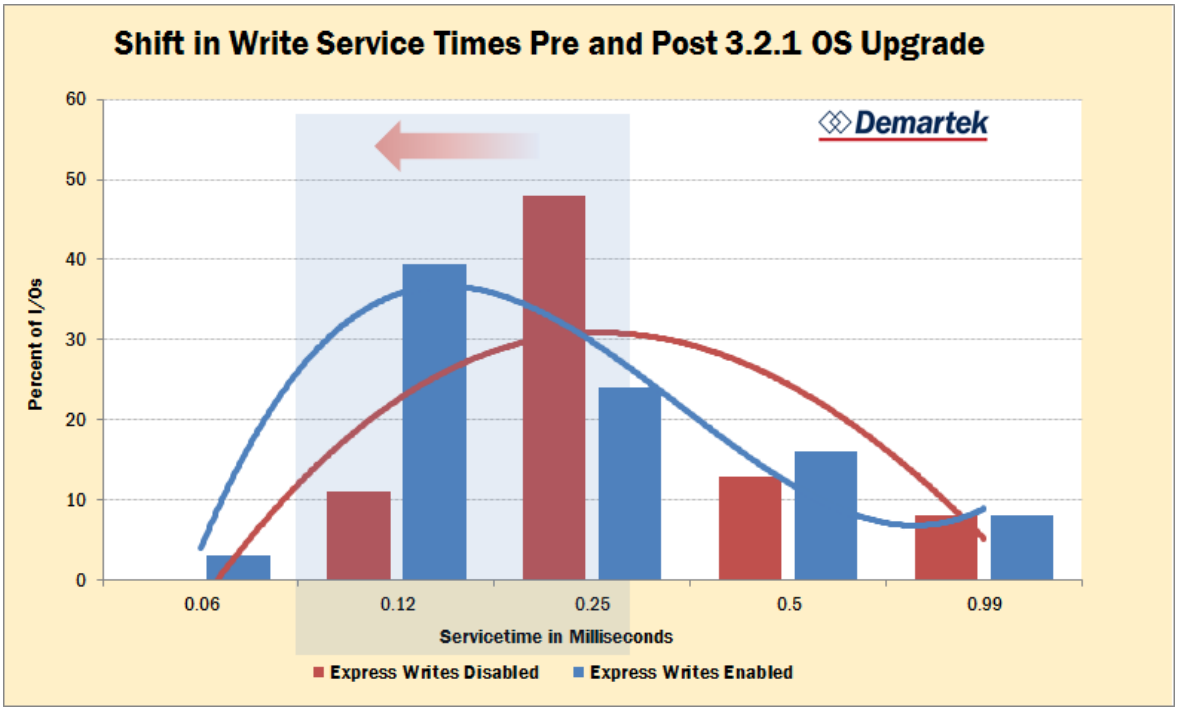

Figure 2 – Service times histogram before and after enabling Express Writes

Telemetry data from operational HP 3PAR StoreServ storage systems upgraded from OS version 3.1.3 to OS version 3.2.1 (Figure 2) show a positive shift in the histogram of write service times with Express Writes enabled. In addition to improvements in average latency, Express Writes also help improve write service times of all the write I/Os. Write service times across the install base have moved to the left, as highlighted by the trend lines on the chart above, indicating that a higher percentage of write I/O is completed in less time than before the upgrades. More precisely, nearly 40% of writes are being serviced at .12 milliseconds post upgrade compared to about 10% pre-upgrade – a 3X improvement.

Demartek was commissioned to review just what that improvement would look like to write I/O on provisioned LUNS supporting a real-world workload.

Transactional Database Workload Description

The TPC-E Database Workload

In order to show the performance of the storage system and to highlight the benefits of 8Gb FC Express Writes, Demartek ran the TPC Benchmark™ E (TPC-E), to generate a real-world workload for measuring the performance of the storage system. It was not the intention of this exercise to produce TPC-E scoring reports for publication and, in fact, the

database was limited to a very low amount system memory to force storage I/O at the expense of the database transactions.

TPC-E is a variation of an On-Line Transaction Processing (OLTP) workload that models a financial brokerage firm with customers who generate transactions related to trades, account inquiries, and market research. The brokerage firm in turn interacts with financial markets to execute orders on behalf of the customers and updates relevant account information. This workload consists of a mixture of mostly reads with some writes to its database.

Real vs. Synthetic Workloads

The workloads employed in this test used a real database (Microsoft SQL Server) with database tables, indexes, etc., performing actual database transactions. When using real database workloads, I/O rate will vary as the workload progresses because the database performs operations that consume varying amounts of CPU and memory resources in addition to I/O resources. These results more closely resemble a real customer environment.

These are unlike benchmarks that use synthetic workloads that perform the same I/O operations repeatedly, resulting in relatively steady I/O rates which, although potentially faster, do not resemble real customer environments.

Workload Definition and Evaluation Objectives

The TPC-E workload is read-heavy, but we wanted enough write I/O to demonstrate just what 8Gb Express Writes would do to write service times on HP 3PAR StoreServ 7450. We provisioned 100 virtual users, generating a respectable read I/O bandwidth with enough write traffic to highlight some interesting results.

Knowing that logical relationships between transactions within the workload would likely place a limit on write bandwidth and IOPs, we chose to focus on I/O latency as the key metric for our analysis as opposed to bandwidth or IOPs.

Results and Analysis

We executed the TPC-E workload and took performance measurements with the 3PAR system reporter tool.

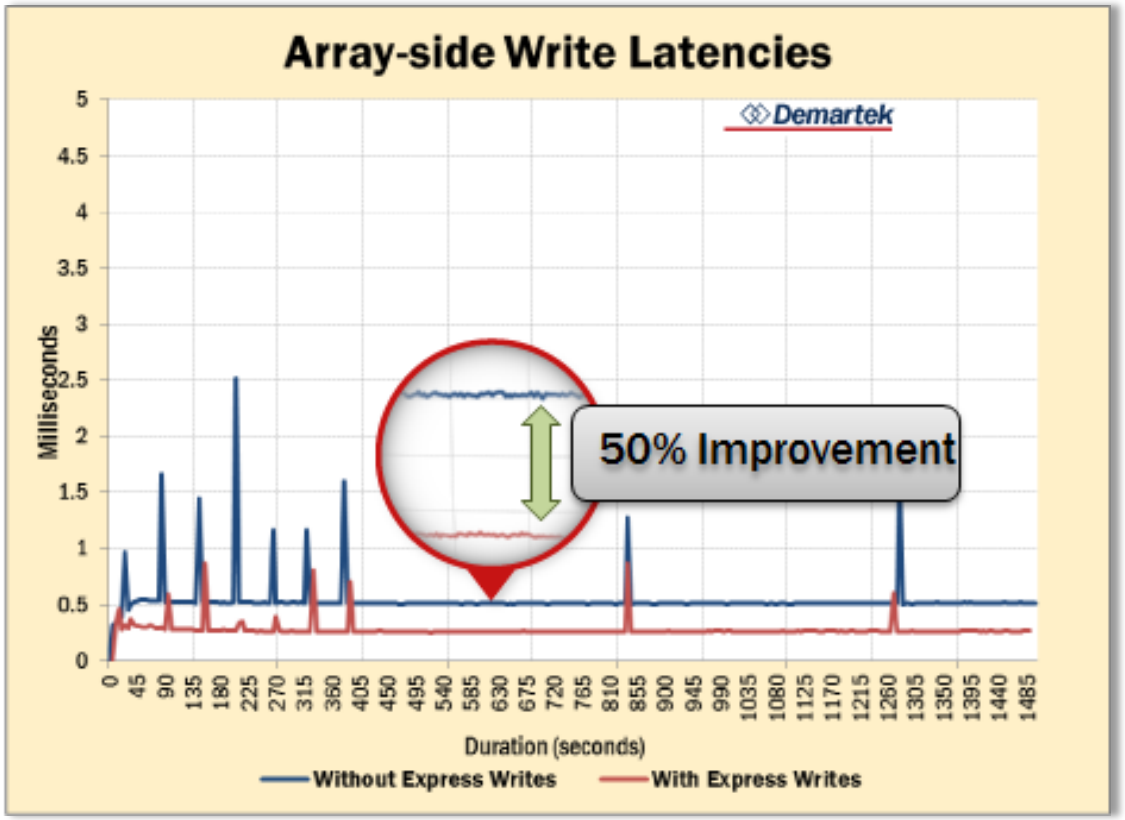

Figure 3 – Write latency

To exclude host-side contributions to latency we measured the write latency at the array where we saw a very impressive 50% improvement, from .5 milliseconds down to .25 milliseconds (Figure 3). We also noted that where sudden bursts of write activity caused fleeting spikes in latency, Express Writes technology reduced those spikes considerably. With Express Writes enabled these rare, sudden spikes never breached the 1 millisecond mark.

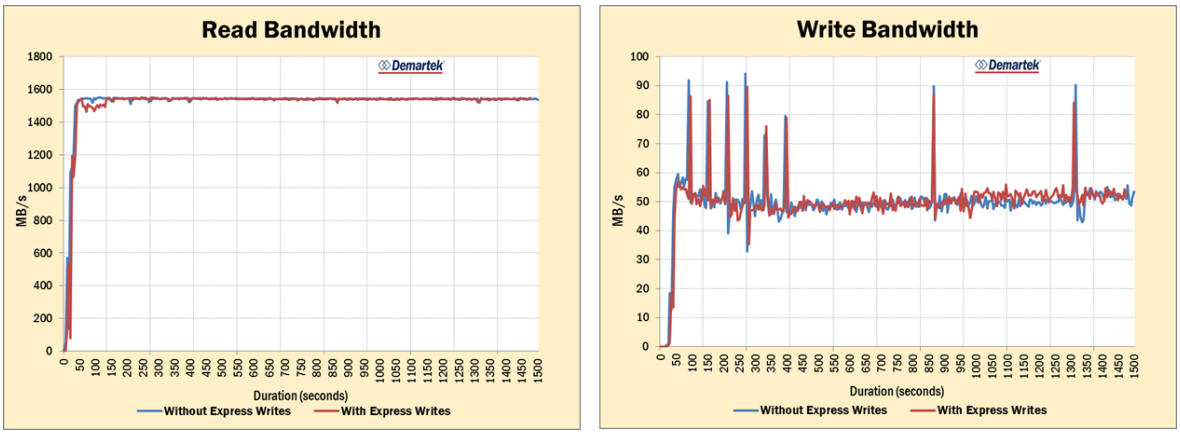

Figure 4 – Bandwidth measurements before and after 8Gb Express Writes

Figure 4 demonstrates read and write bandwidths measured with the 8Gb Express Writes feature first disabled and then enabled. We saw little difference in these metrics regardless of whether the 8Gb Express Writes feature was enabled, as expected since the storage system will not deliver more bandwidth than demanded by the host.

A Brief Commentary on Latency

Before flash storage became commonplace in the datacenter, storage I/O latencies of 10 to 20 milliseconds were generally acceptable for many applications. In fact, latencies lower than 2 milliseconds are almost unachievable on spinning hard disk drives, simply because of the time it takes to perform the mechanical motions of the platters and heads. Flash storage has been a game-changer in this area with sub-millisecond latency now the expectation for all-flash arrays. As with all technology advances, applications and user expectations have changed in response to this capability.

The impact of higher latencies depends greatly on the workload. High bandwidth streaming or very sequential workloads might be more or less unaffected, especially where read-ahead buffering grabs more data than I/O requests actually demand. Data warehousing and video streaming are two examples of these types of workloads. However, if latencies become too high, even these jobs begin to suffer from noticeable lags. For optimal user experiences, lower latency is always better.

Online transactional workloads can generate high numbers of IOPS and consume a respectable amount of bandwidth. Latency becomes especially important, particularly in very highly transactional workloads, when database requests are time sensitive and have a great deal of dependency on prior transaction results. Consider applications that perform real-time trend analysis and/or process vast amounts of data. Stock trading, such as modelled by the TPC-E workload, weather forecasting, geological survey modelling, and biometric analysis are examples of workloads that can be extremely sensitive to latency. As storage systems have improved, reducing I/O latency in the process, the performance expectations for these types of applications have likewise adapted to expect very fast response times.

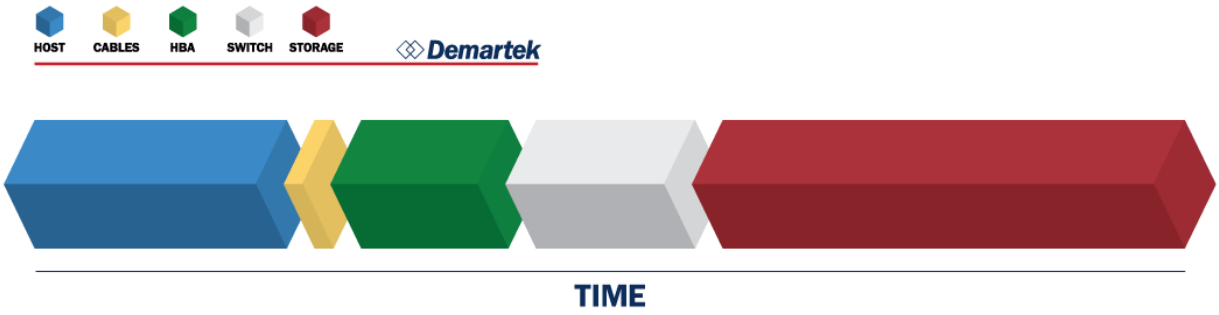

Figure 5 – Contributors to overall latency

Latency in a SAN is introduced from several sources (Figure 5). Total latency will never be lower than the storage system's best case scenario. However, varying amounts of latency are also introduced at the host HBA, and this can grow rapidly when IOPS or bandwidth limitations are met. As soon as host HBAs reach a saturation point, I/Os begins queuing up on the host, resulting in increased latency. Each SAN switch between the host and storage contributes to latency, typically in the low tens of microseconds. Fiber optic cables will add a few microseconds per kilometer length as well.

The additive effect of these latency sources has an impact on application performance and user experience to varying degrees, depending on the application requirements. If the aggregate exceeds agreed upon QoS standards or service levels demanded by the business, a savvy storage administrator will consider all of these potential points of latency insertion. The ability to measure performance at the host, the storage system, and in between (in switches and other storage networking elements) is essential to pinpoint where action needs to be taken.

Summary and Conclusion

Latency is important, and the computing industry is recognizing that modern applications are demanding ever lower latency from compute systems. Storage devices have long been the slowest components in computing environments, though modern technologies have been making considerable improvements in that area. As aggregate system latency can never be lower than its largest contributor, the latency at the storage system provides a baseline that can be used for tuning the rest of the hardware and software stack to meet application and user service level requirements. The lower the storage system latency is, the better the overall service time for the application experience (obviously a poorly tuned server and application have the potential to squander any latency improvements delivered by the storage, but that is a topic for another report).

Write operations stress flash storage more than reads. While the All-Flash HP 3PAR StoreServ 7450 already services write requests quickly, there is room for improvement, and HP has delivered that improvement through Express Writes. HP's 8Gb Express Writes technology is a simple upgrade that provides an immediate performance boost for storage systems, particularly those that may not be in a position to commit to other performance enhancing upgrades such as 16Gb Generation 5 Fibre Channel. Write latency can see improvements of 50% by a non-intrusive array operating system upgrade to 3.2.1 or better.

The demand for extremely low latency coupled with high bandwidth and IOPS is only increasing. Owners of applications in which microseconds count and who are looking to squeeze the absolute best performance out of their storage system are advised to strongly consider the HP 3PAR StoreServ 7450 for its raw performance as an all-flash storage system and for the added benefits of HP's Express Writes technology.

Appendix A – Test Description and Environment

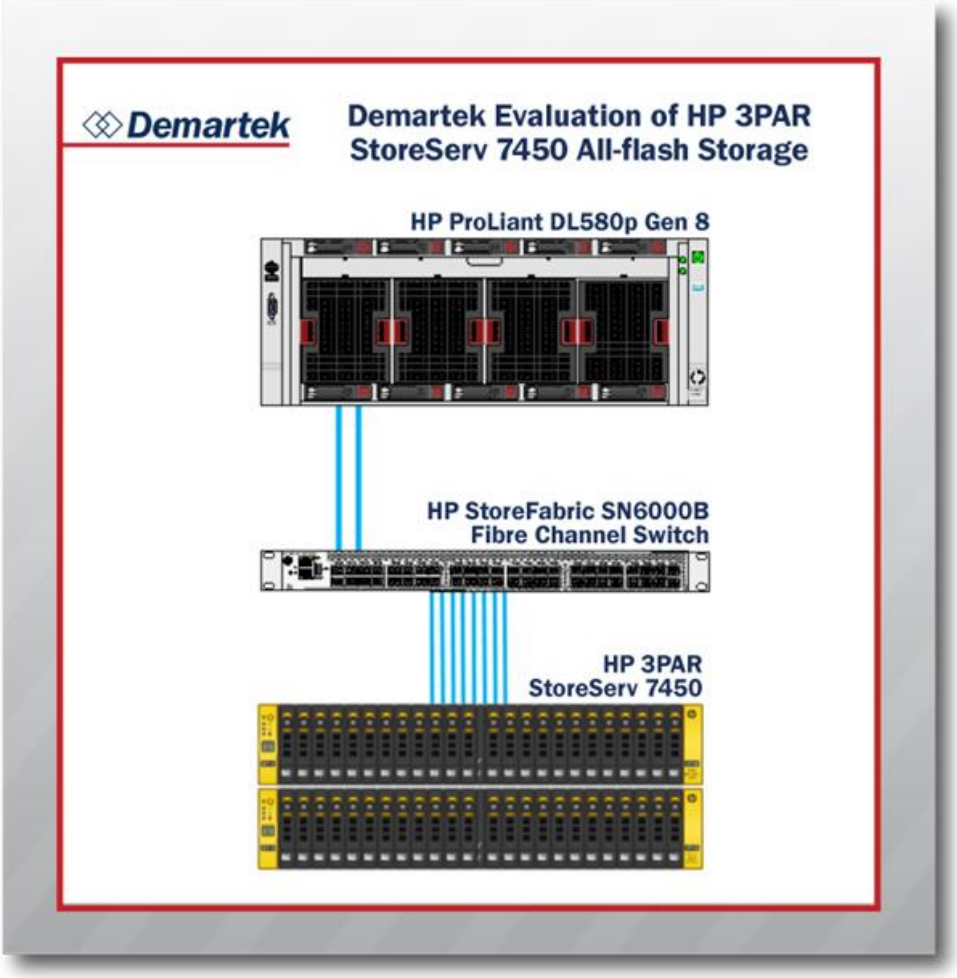

Figure 6 – Test Infrastructure

Server

- HP ProLiant DL580 Gen 8
- 4 Intel E7-4890v2 2.8GHz CPUs
- 256 GB RAM (MS SQL Server constrained to 16GB RAM to force storage I/O)
- Emulex LPe12002 (8Gb FC) dual port HBA
- Microsoft Windows Server 2012 R2
- Microsoft SQL Server 2012, Microsoft TPC-E benchmark kit

Fibre Channel Switch

- Brocade 6510 (sold by HP as the HP StoreFabric SN6000B Fibre Channel Switch)

Storage Array

- HP 3PAR StoreServ 7450 Storage
- 3PAR Operating System 3.2.1 MU2
- 48x 400GB 6Gb SAS SSD
  - RAID5 configuration for database volume
  - RAID10 configuration for log volume
- 8x 8GFC target ports or 8x 16GFC target ports (HP OEM'ed Emulex target adapters)

Appendix B – HP 3PAR StoreServ 7450 All-flash Storage Architecture

The HP 3PAR StoreServ 7450 All-flash storage system provides high-speed performance with low latency for mission-critical applications such as database applications. It provides high-performance and the same enterprise-grade levels of resiliency (up to 4 nodes), rich Tier 1 data services, efficiency, data mobility and disaster tolerance that are available on other HP 3PAR StoreServ platforms. It does this with a flash-optimized architecture that reduces the performance bottlenecks that can choke general-purpose disk arrays that have been retrofitted by stuffing them with solid-state disks (SSDs), and includes enterprise-grade features that are not always available with some of the newer all-flash offerings from others.

HP 3PAR StoreServ 7450 storage system has a number of design optimizations to leverage performance from flash media while keeping the costs down. Some of them are described below briefly:

Adaptive Read and Write

Adaptive Read and Write is a software innovation that enables a more granular approach than with spinning media by matching host I/O size reads and writes to flash media to avoid unnecessary data reads and writes. This significantly reduces latency and optimizes back end performance to enable more applications to be consolidated.

Autonomic Cache Offload

Autonomic Cache Offload is another new flash software-based optimization that eliminates cache bottlenecks by automatically changing the frequency at which data is offloaded from cache to flash media, based on utilization rate and without requiring any user intervention. This ensures consistently high performance levels as you scale workloads to hundreds of thousands of IOPS.

Another important aspect of the cache offload algorithm is the decision around which cache data should be flushed to the back end, and which should not be. HP 3PAR StoreServ Storage keeps track of read cache hits and keeps hot data in cache itself, thereby lowering latencies of frequently accessed data. In addition, for handling flash, flusher threads have been added to 3PAR cache management so it can perform more operations in parallel.

Multi-tenant I/O Processing

New multi-tenant I/O processing innovations enable performance optimization for mixed workloads and VDI deployments by breaking large I/O into 32 KB sub-I/O blocks. This prevents small read I/O chunks from getting held up behind large I/O requests, therefore assuring reduced latency. Breaking large sequential read I/O into sub-I/O chunks distributes these reads and lowers the possibility of smaller transactional read I/O operations of get held up because of a previous large read I/O operation, therefore ensuring consistently low latency for transactional I/O, even in mixed workload scenarios.

With the multiple workloads (OLTP and Decision Support) running on 7450 in this evaluation, enhancements like multi-tenant I/O processing help 7450 to deliver consistently good latency for the OLTP application while deliver good bandwidth for the Decision Support application.

Specialized ASIC for Mixed Workload Support and Zero Detection

In the HP 3PAR StoreServ Architecture, transactional and throughput-intensive workloads are able to run on the same storage resources without contention because of the HP 3PAR StoreServ Gen4 ASIC, which offloads some of the work of the controllers, allowing them to process smaller I/O without inordinate delay due to large block I/O. This is important, because precious CPU cycles are not wasted for data movement; instead the CPU cycles are available for delivering advanced Tier 1 data services.

The HP 3PAR StoreServ ASIC features an efficient, silicon-based zero detection mechanism. This hardware feature removes allocated but unused space without impacting performance. Every block of flash storage that gets reclaimed immediately becomes available for other applications requiring space. This ensures that existing flash storage is utilized efficiently, possibly delaying the purchase of additional capacity.

In addition, with zero detection built into the ASIC, a stream of zeroes that may be present in a write I/O can be eliminated before being written to flash. In the world of flash, where there is a penalty for every single write (in terms of media endurance), this write elimination helps elongate the life of flash-based media.

System-wide Striping

Data and I/O for each volume are striped widely across all system resources, including CPUs, ports, cache, and drives. This wide striping enables the system to deliver

accelerated performance levels (with all resources supporting each volume) while avoiding any single point of contention. Even a small volume can leverage the performance of hundreds of flash devices and all the system's controller nodes for optimal performance.

Because the system autonomically manages this system-wide load balancing, no extra time or complexity is required to create or maintain an optimally configured system. With system-wide striping, data is distributed across all SSDs in a granular fashion, keeping the command queues low on any individual SSD.

System-wide Sparing

The HP 3PAR StoreServ Architecture reserves spare chunklets in all flash media. In contrast to traditional architectures that enforce the need to reserve dedicated spares that sit idle, the HP 3PAR StoreServ Architecture uses every single flash device, reserving spare chunklets in each of them. This ensures a balanced load and wearing across all flash media. Should there be a media failure, system-wide sparing also helps protect against performance degradation by enabling many-to-many rebuild, resulting in faster rebuilds.

Moreover, with Adaptive Sparing, 3PAR StoreServ is able to release the spare chunklets present in each SSD media to the SSD itself, allowing SSDs to use that spare capacity as additional over-provisioned capacity. This increase in over-provisioned capacity enhances endurance of SSDs within a 3PAR StoreServ array

Media Wear Gauge

The HP 3PAR Operating System monitors the wear of each flash device and allows users to see the wear level of each device at any given time. This keeps users informed as to the amount of media wear taking place and helps them replace SSDs in a planned fashion.

Appendix C – Performance Metrics

When measuring the performance of storage systems, three key metrics are I/Os per second (IOPS), bandwidth and latency. Online transaction applications tend to generate a high numbers of IOPS and can consume a respectable amount of bandwidth. Latency is especially important in highly transactional workloads where database requests can be quite time sensitive.

- **IOPS** – I/O's per second – a measure of the total I/O operations (reads and writes) issued by the application servers.
- **Bandwidth** – a measure of the data transfer rate, or I/O throughput, measured in bytes per second or MegaBytes per second (MBPS).
- **Latency** – a measure of the time taken to complete an I/O request, also known as response time. This is frequently measured in milliseconds (one thousandth of a second).

Other Demartek reports[2,3] have examined the overall performance of the HP 3PAR StoreServ 7450 storage system and its capacity for bandwidth and IOPs. This project did not attempt to exercise the array to its full bandwidth and IOPs potential. Instead we chose to focus on latency as the metric to best highlight the effect of 8Gb Express Writes on a workload.

Metrics can be gathered at many points within a SAN. Host based metrics offer a picture of the end-user experience. Switches, arrays, and to some extent adapters also have monitoring tools which can provide a more targeted view of the performance at those particular subsystems within the SAN. Gathering data from all of these locations is critical for troubleshooting performance problems. However, we were not trying to fix or enhance application performance through server and application tuning. We merely wanted to determine the benefit of a single storage system upgrade. In order to get the clearest picture of the latency improvements that 8Gb Express Writes could offer, we decided to exclude all the "upstream" contributors to latency by measuring performance at the array.

---

[2] Demartek, "Demartek Evulation of HP 3PAR StoreServer 7450 All-flash Storage", http://www.demartek.com/Demartek_HP_3PAR_StoreServ_FlashStorage_2014-05.html (May 21, 2014)

[3] Demartek, "Demartek Evaluation of HP 3PAR StoreServ 7450 16GFC All-Flash Storage", http://www.demartek.com/Demartek_HP_3PAR_StoreServ_7450_16GFC_2015-03.html (March 30, 2015)

The original version of this document is available at:
http://www.demartek.com/Demartek_HP_3PAR_StoreServ_7450_8GFC_Express_Writes_2015-05.html on the Demartek website.