Evaluation Report: Improving Oracle Database Response Times with All-Flash HP 3PAR StoreServ 7450c

Evaluation report prepared under contract with HP

Executive Summary

Enterprise applications have extremely demanding performance requirements and the Oracle database is a prime example of a high performance application used to run critical business workloads. Careers are built on managing and tuning Oracle databases. However, no matter how well tuned a database is, its performance will still be limited by the capabilities of the hardware it runs on. Storage systems have traditionally been the slowest part of a compute infrastructure in terms of IOPs, bandwidth, and particularly response times. Businesses are increasingly turning to flash storage to improve these performance metrics.

Flash storage I/O bandwidth numbers of multiple gigabytes of data per second and hundreds of thousands to millions of I/O operations per second across corporate SANs are both achievable and common. Response time, or latency, is becoming an increasingly critical component of enterprise-class performance in all-flash arrays. In highly transactional workloads low latency can be the primary driver of the application experience. Sub-millisecond I/O latency is expected from enterprise all-flash arrays and is a critical component in optimizing the performance of Oracle workloads. Robust flash-based storage, such as the HP 3PAR StoreServ 7450c delivers this type of performance out of the box.

Demartek recently published two reports evaluating the performance of the All-Flash HP 3PAR StoreServ 7450c. The first with MS-SQL database applications and 16Gb FC targets[1] and the second with a MS-SQL database application and 8Gb FC targets with Express Writes technology[2]. This time Demartek has evaluated the All-flash HP 3PAR StoreServ

[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)
[2] Demartek, "Demartek Evaluation of HP 3PAR StoreServ 7450 8GFC Express Writes", http://www.demartek.com/demartek_HP_3PAR_StoreServ_7450_8GFC_Express_Writes_2015-05.html (May 12, 2015)

7450c storage system supporting an Oracle database application and analyzed the I/O response time improvements offered by both HP StoreFabric Gen 5 16Gb FC technology and 8Gb Express Writes to an Oracle workload. For this project, Demartek ran a transactional workload application on an Oracle 12c database in an Oracle Enterprise Linux environment and analyzed the performance difference between end-to-end StoreFabric 8Gb FC and Gen 5 16Gb FC SANs. Recognizing the some enterprises may be forced to remain at 8Gb FC for various business reasons, we also considered the performance effect of the HP 3PAR 8Gb Express Writes technology on the same workload.

Key Findings

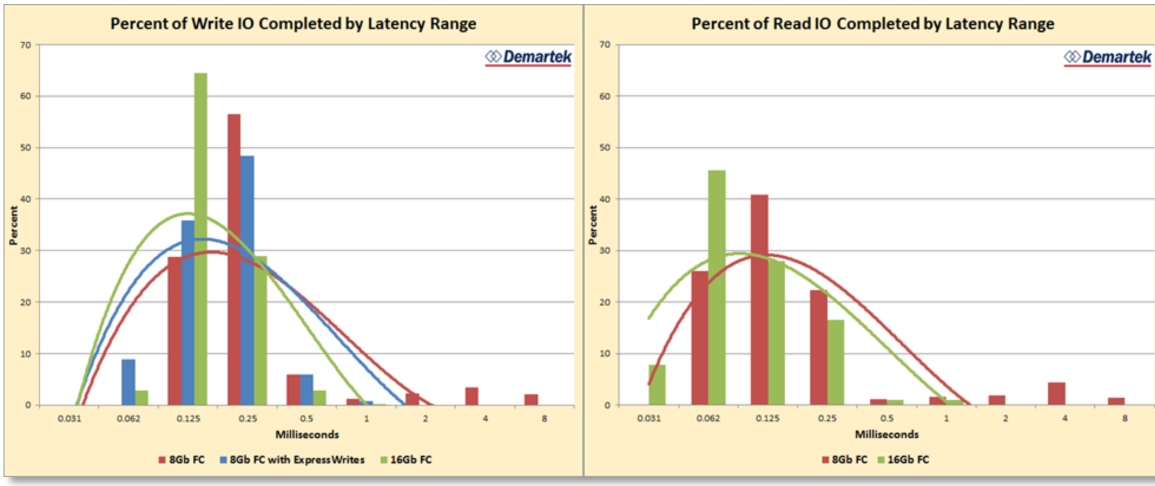

Figure 1 - Latency improvement histograms

Both technologies, end-to-end Gen 5 16Gb FC SAN and 8Gb Express Writes in an 8Gb SAN environment, reduced I/O latencies. The histograms in Figure 1 display this shift to the left, or toward lower values, of I/O latency with the application of these upgrades. These upgrades also eliminated the long, tailing I/Os evident with 8Gb FC technology.

16Gb FC Results

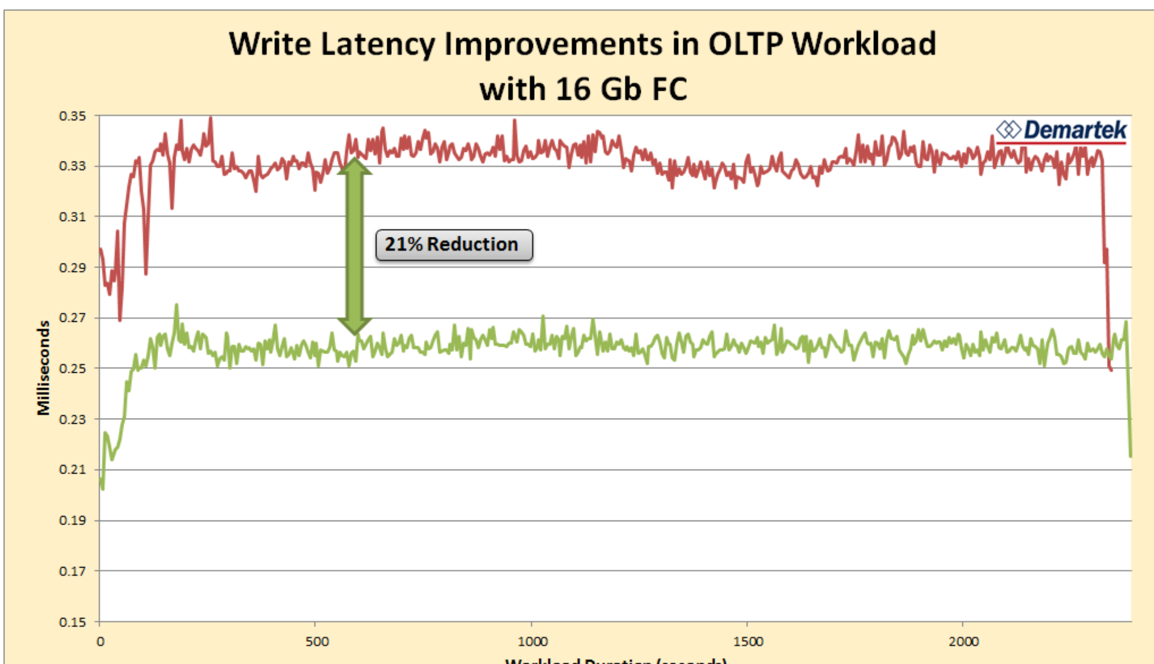

Figure 2 - Write latency improvement with 16Gb FC

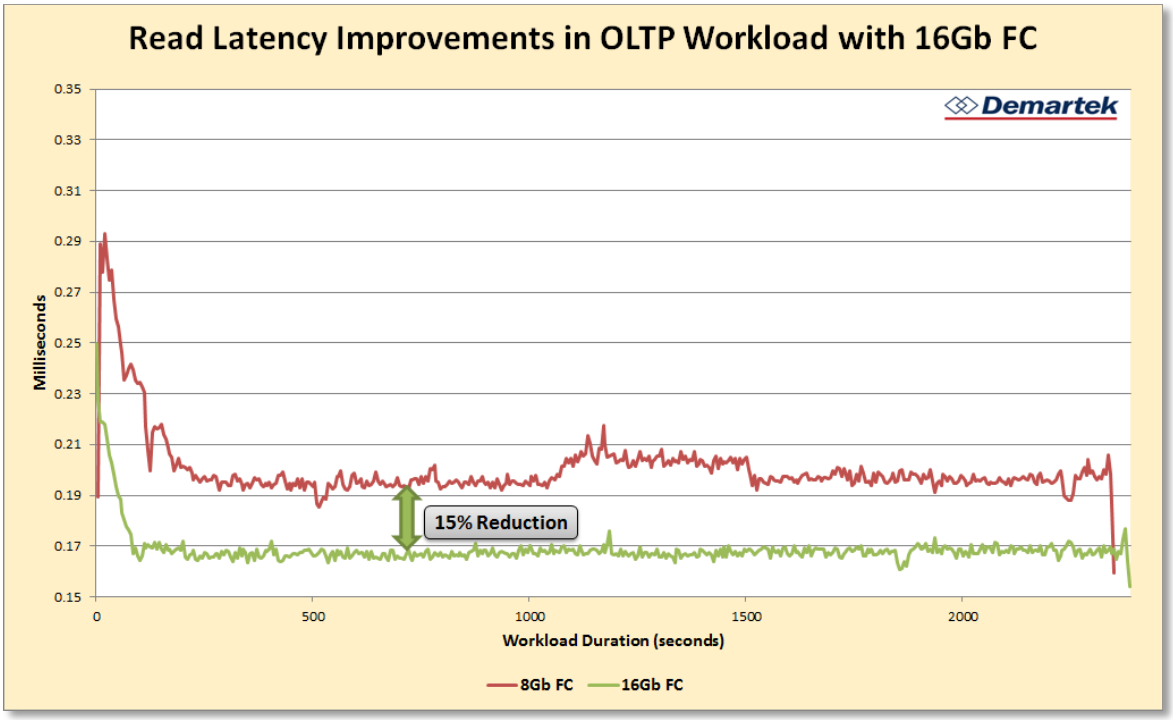

Figure 3 - Read latency improvement with 16Gb FC

The All-Flash HP 3PAR StoreServe 7450c, out of the box in a legacy 8Gb FC SAN, delivered response times to I/O requests from an Oracle OLTP workload modeled on the familiar Transaction Performance Council™ Benchmark C with write and read averages of 330 microseconds and 200 microseconds respectively (Figures 2 and 3). Deploying end-to-end StoreFabric Gen 5 16Gb FC SAN infrastructure (HP StoreFabric HBAs, switching, and storage targets) as the storage environment for the same workload resulted in I/O response time improvements averaging 21% for write I/O and 15% for read I/O.

8Gb Express Writes Results

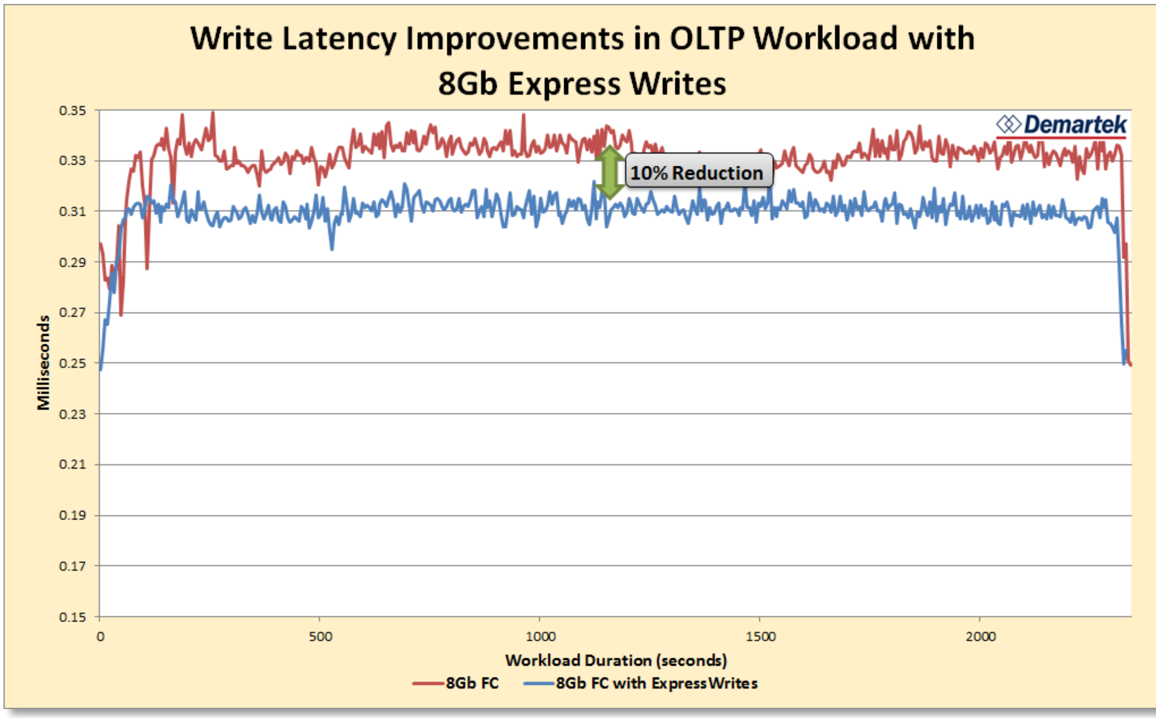

Figure 4 – Write latency improvements with 8Gb Express Writes

HP 3PAR 8Gb Express Writes, is a feature available to 8Gb FC targets on HP 3PAR StoreServ systems running OS release 3.2.1 or better. As the name implies this is a write optimization for 8Gb FC targets. Express Writes reduced write latency by10% in this environment (Figure 4).

The Advantage of HP Express Writes for 8Gb Fibre Channel

Upgrading SAN infrastructure to StoreFabric Gen 5 16Gb FC unlocks optimal performance from all-flash storage systems. This type of upgrade can be non-trivial and that may delay some enterprises from deploying it. Recognizing this, HP has provided owners of HP 3PAR storage systems with an option to improve performance of 8Gb FC storage targets. 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 and is included in the base license. In short, if a 3PAR StoreServ array supports 3.2.1, it gets 8Gb Express Writes for free.

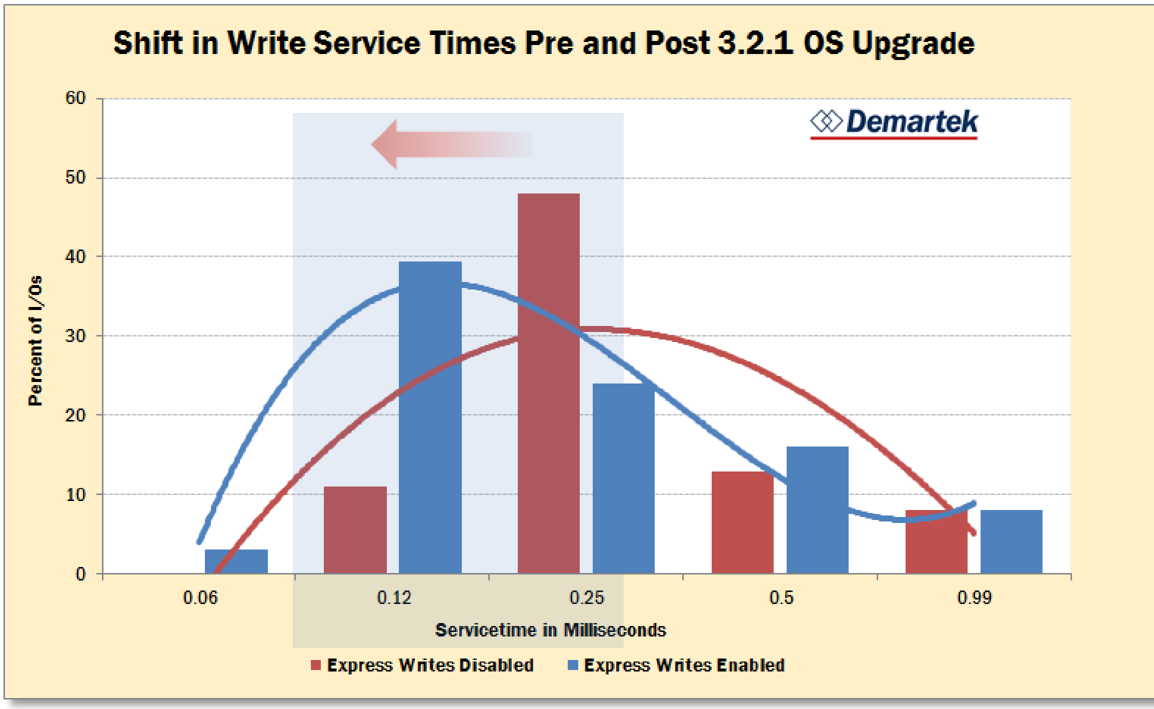

Figure 5 - Service times histogram before and after upgrading to 3.2.1

Telemetry data from operational HP 3PAR StoreServ storage systems upgraded to OS version 3.2.1 (Figure 5) show a positive shift in the histogram of write service times with Express Writes enabled. 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.

Transactional Database Workload Description

Demartek ran an Oracle database workload modeled on the TPC –C benchmark, to generate a real-world workload for measuring the performance of the storage system. It should be recognized that this workload is not the official TPC-C benchmark and it was not the intention of this exercise to produce TPC-C scores for publication. Nor did we wish to turn this into an Oracle tuning exercise to produce the highest scores possible on the test platform. Therefore, minimal database tuning was implemented. In fact, Oracle was limited to a very low amount of system memory to force storage I/O, highlighting the benefits of HP StoreFabric infrastructure and array-based features such as 16Gb FC targets and 8Gb Express Writes at the expense of the database transactions. This would not be the case in true production environments where we would take advantage of the full power of the HP server and extensive database tuning. Readers interested in optimizing the performance of their HP Proliant servers are advised to refer to HP's technical whitepaper on *Configuring and tuning HP Proliant Servers for low-latency applications*[3]. We would expect that in a true production environment applications owners would combine the very best server and application tuning with the additive benefits delivered by modern storage technologies like StoreFabric Gen 5 16Gb FC SAN infrastructure or 8Gb Express Writes.

TPC-C is an industry standard On-Line Transaction Processing (OLTP) workload that simulates customer orders of products supplied from a company's warehouses and the fulfillment of those orders. We chose to deploy the HammerDB open source TPC-C implementation as the workload generator.

Real vs. Synthetic Workloads

The workload employed in this test used a real database (Oracle 12c) 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. The differences in real vs synthetic workload I/O can be seen by comparing the magnitude of the irregularities in the I/O patterns of the real-world workload (such as Figure 6) with a synthetic workload (such as Figure 8) in the results and analysis portion of this report.

From our previous testing of the HP 3PAR 7450c in a Gen 5 16Gb FC SAN, we had an idea of the benefits the database workload would see from upgrading to Gen 5 16Gb FC. However, since Express Writes is a software rather than a hardware upgrade we hypothesized that its benefits might be affected by the nature of the workload being executed. Therefore before running the Oracle workload, we also ran a synthetic workload with the vdbench I/O generator. This workload was designed to resemble I/O patterns Oracle transactions might produce to get a preliminary idea of what to expect from 8Gb Express Writes. The vdbench workload was a mixture of read and write I/Os in 8 kilobyte blocks, varying the read-write ratios and level of randomness in the workload. As a synthetic workload, we were able to control the I/O pattern very closely for modeling purposes, an unlikely scenario in the real world, but useful for general analysis. (The vdbench parameter file can be seen in Appendix F.)

---

[3] HP Technical Whitepaper, "Configuring and tuning HP Proliant Servers for low-latency applications", http://h10032.www1.hp.com/ctg/Manual/c01804533.pdf (November 2014)

Workload Definition and Evaluation Objectives

TPC-C is fairly write-heavy compared to other TPC database benchmarks. We chose this workload to generate enough write I/O to demonstrate the effect of 8Gb Express Writes on the HP 3PAR StoreServ 7450c. 46 virtual users were enough to just about exhaust the processing capacity of the server used as the database and workload engine host while generating a respectable I/O bandwidth.

We chose to focus on I/O latency as the key metric for our analysis as opposed to bandwidth or IOPs. The reasoning for this being that the server used as the workload engine was run at close to the limits of its processing capacity and would not be able to drive much more in the way IOPs and bandwidth regardless of the capacity of the SAN. In short, the storage array was not going to deliver more than the server could request, but it could potentially complete each request more quickly.

Results and Analysis

In the course of this testing, the Oracle workload was not retuned or resized to take advantage of any benefits gained by the application of newer technology (Gen 5 16Gb FC infrastructure or HP 3PAR 8Gb Express Writes). Our intention was only to measure the immediate effects of these upgrades on an existing workload configuration.

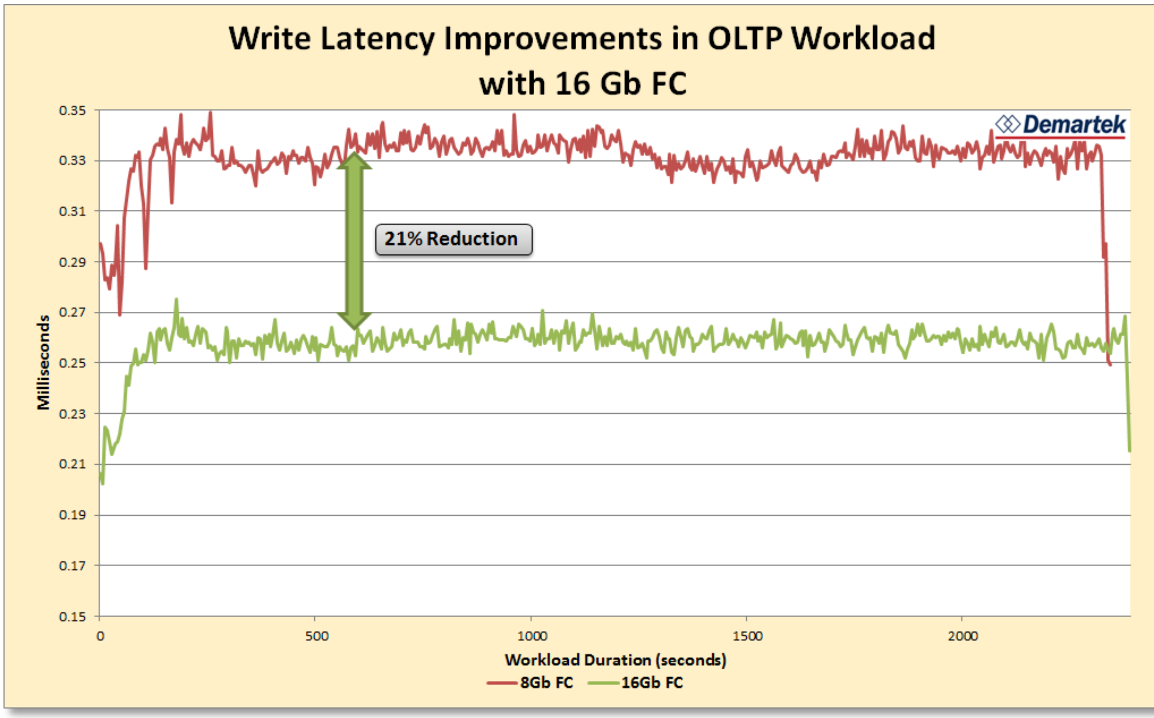

Figure 6 - Write latency improvement with 16Gb FC

In an 8Gb FC end-to-end SAN (8Gb FC HBAs, 8Gb FC switch, and 8Gb FC storage targets in the array) the Oracle workload under test experienced an average write I/O latency of 330 microseconds (Figure 6). When evaluating latency improvements lower values are better. That is clearly achieved by the upgrade to an HP StoreFabric Gen 5 16Gb FC SAN environment. Write I/O latency improved by 21%, down to an average of 260 microseconds.

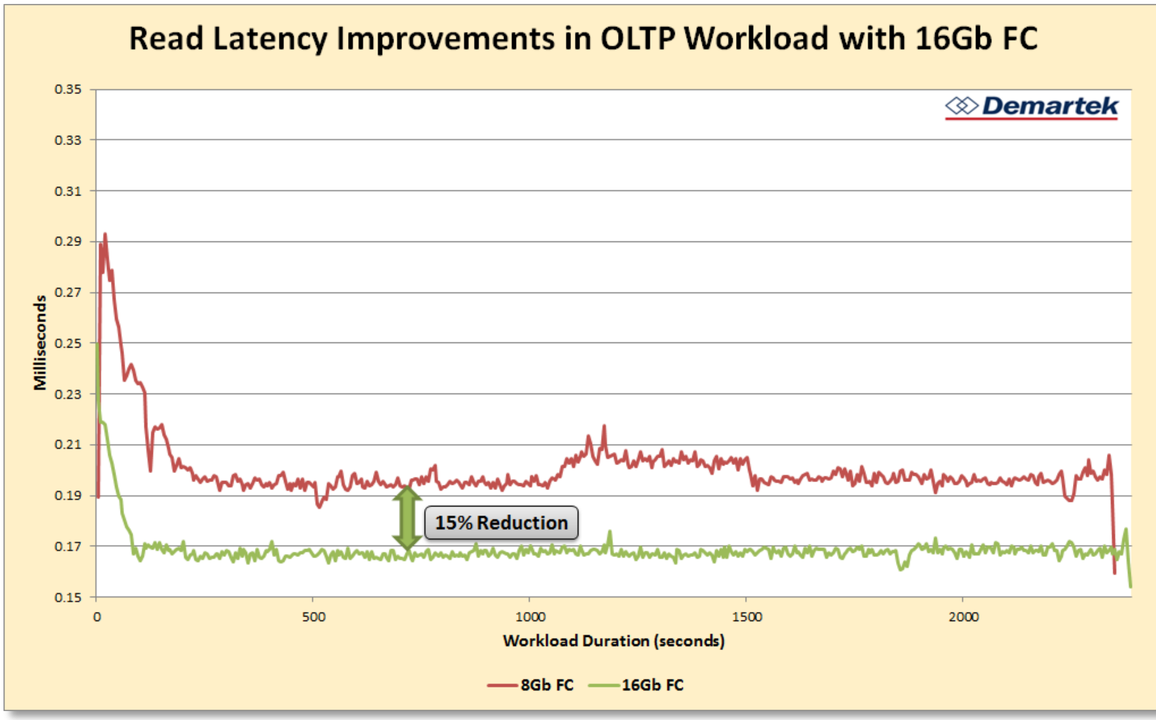

Figure 7 – Read latency improvement with 16Gb FC

Read I/O also saw a measurable improvement, decreasing from 200 microseconds in the 8Gb FC environment to an average of 169 microseconds after deploying Gen 5 16Gb FC, a 15% improvement (Figure 7). These write and read response time values highlight the very low latencies that the HP 3PAR StoreServ 7450c is capable of out the box, even with legacy 8Gb FC, and Demartek's recommendation that Gen 5 16Gb FC is the technology that owners of high-end flash storage should be deploying if they wish to get the absolute best out of their storage investment.

We then turned our attention to HP 3PAR 8Gb Express Writes to evaluate the benefits to the workload in an environment that can't make the transition to Gen 5 16Gb FC just yet. For this portion of the testing we only looked at write latencies. Read metrics were captured earlier and are expected to be unaffected by a write optimization.

Before running the Oracle TPC-C-like database workload, we ran a synthetic database I/O modeling workload as a preview of how 8Gb Express Writes might affect the write latency of a workload performing 8Kb transfer size I/O, which is the default I/O size for Oracle databases.

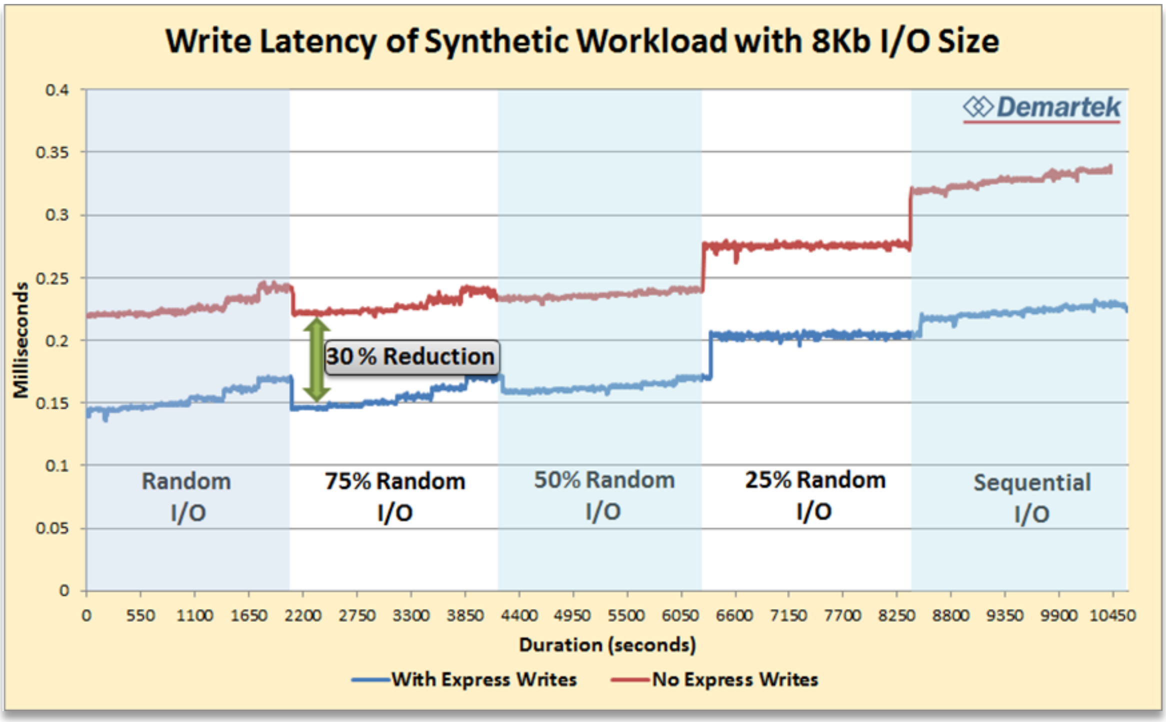

Figure 8 - Write latency of the vdbench workload

Within each variant of random I/O graphed in Figure 8 the workload performed read-write I/O with mixes beginning at 95% read moving down to 70%read. As a result write bandwidth varied considerably, from 30 MB/s to 200 MB/s. Write latency was fairly stable throughout each major grouping. Latency reductions from 30% – 35 % were recorded when Express Writes was enabled.

Being a synthetic workload we were able to control the I/O tightly, resulting in very little variance between data points. We did not expect this tightly constrained behavior to be repeated in the Oracle OLTP workload, nor could we control the precise size of each I/O and the degree of randomness between subsequent I/Os in the real database workload. The results from the vdbench workload likely represent a best case scenario of the workload as defined. However, it demonstrated that Express Writes does reduce latency of 8Kb write I/Os, which are common in many database applications.

The TPC-C-like Oracle database workload also experienced a reduction in write I/O latency (Figure 9).

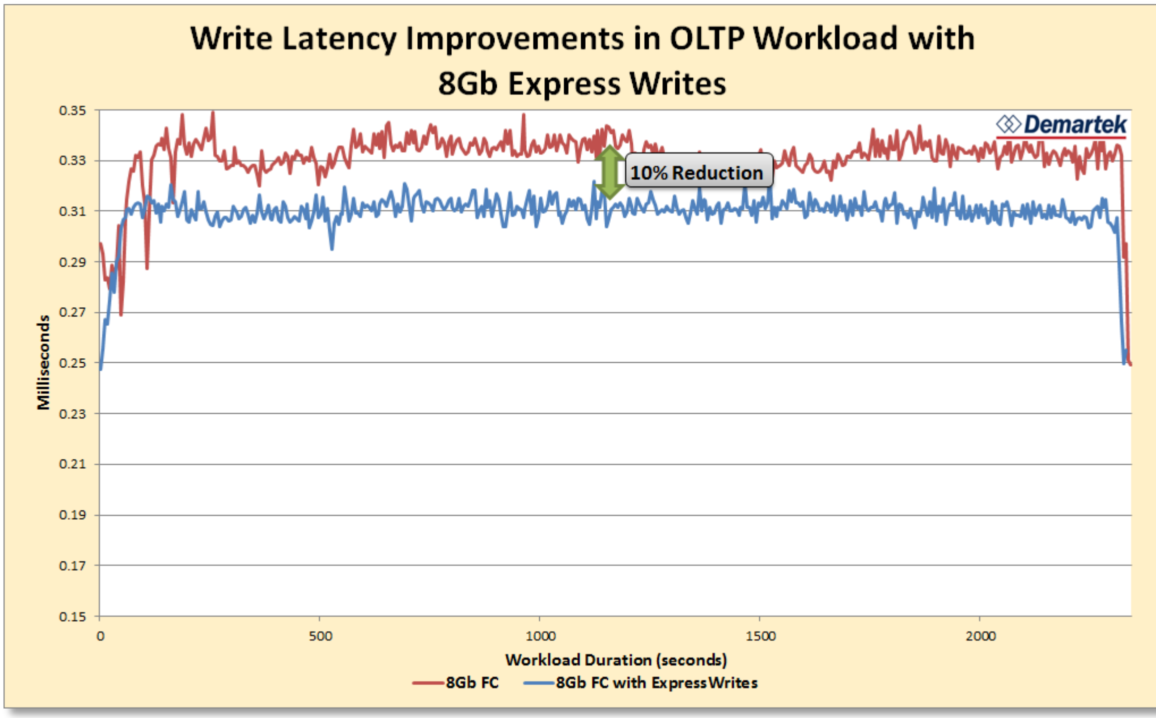

Figure 9 - Write latency improvement with 8Gb Express Writes

The average was about a 10% improvement. This was lower than our synthetic model as we expected, but is still a clear improvement for this workload. It is important to note that these results can't be directly compared with the vdbench workload, or other database workload testing Demartek has performed with HP 3PAR StoreServ 8Gb Express Writes[4] as the workloads differ with respect to control of I/O sizes, random versus sequential I/O, and the read-write ratios. Nor can we declare that certain software is more amenable to Express Writes benefits as the software is strictly the engine that sends I/O to the array. What we can conclude is that there is a measurable reduction to write latencies of I/O requests created by this TPC-C-like workload in an Oracle 12c database environment, and this reduction was realized simply by upgrading the array to take advantage of a no-cost feature in HP 3PAR OS .

[4] Demartek , "Demartek Evaluation of HP 3PAR StoreServ 8Gb 7450 with 8GbExpress Writes", http://www.demartek.com/Demartek_HP_3PAR_StoreServ_7450_8GFC_Express_Writes_2015-05.html (May 2015)

Using HP 3PAR StoreServ System Reporter, we also captured array latency histogram data (Figure 10).

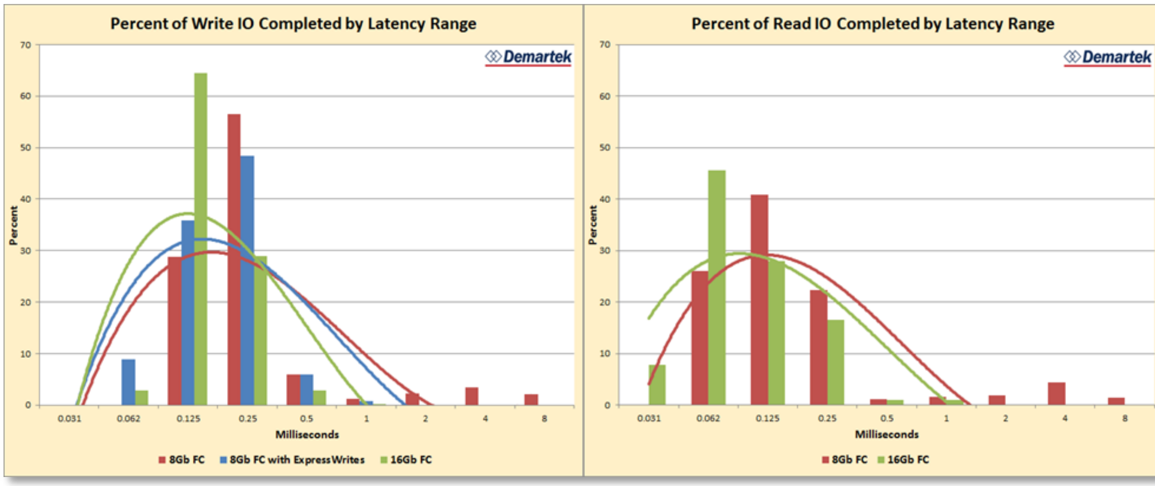

Figure 10 - Latency Histograms

The histogram data shows the distribution of I/Os in each of the latency buckets on the x-axis. This data demonstrates a bit more graphically the positive shift to the left (lower values) of response times at the device level for 16Gb FC storage targets or 8Gb Express Writes. Specifically, I/Os serviced within 62 microseconds were only achieved after Gen 5 16Gb FC SAN infrastructure or Express Writes were deployed. Another point to call out, that may be overlooked, is that there are no read I/Os at all with service times of more than one millisecond when 16Gb FC targets are in place and no writes serviced in more than a millisecond with Gen 5 16Gb FC SAN infrastructure or 8Gb Express Writes.

While any latency at a millisecond or above are rare on the HP 3PAR StoreServ 7450c in general, the loss of these long, tailing I/O latencies means that processor cores are not going idle waiting for I/Os to complete during these intervals. Instead, these cores are available to execute the next transaction in queue. In Oracle environments the reduction of these sporadic but long response times can improve the transaction throughput of a busy system. If this improved server efficiency is enough to allow the workloads to be run with fewer processor cores, fewer or smaller servers may accomplish the same work. Fewer core-based software licenses and related support expenses will then be required to support the workload, resulting is very real cost savings. If we extrapolate this out to the entire data center it can translate to significant savings in licensing and capital purchases.

Summary and Conclusion

Businesses purchase high-end flash storage because their processes demand extreme performance. However, throwing flash at a storage bottleneck doesn't guarantee that performance bottlenecks go away. An enterprise application, such as Oracle database, running demanding workloads requires a well architected and well-tuned compute environment, including servers, storage, and networking to deliver the service needed by the business. Without giving proper consideration to the full compute infrastructure and making informed capital investments performance potential from all-flash storage can be wasted. This in turn can create other bottlenecks such as underperforming servers due processor cores being idle while waiting on I/O .

High IOPs and bandwidth are fairly easy to achieve with today's all-flash arrays and they are important benchmarks of a high performance storage system but they are also related to latency. Lower service times allow storage systems to come closer to achieving the best IOPs, and thereby bandwidth, the storage media is able to deliver. Conversely, higher I/O latencies will constrain IOPs more than the hard limits imposed by the media which can reduce bandwidth, particularly for smaller size I/Os. Perhaps more importantly, as the latency aggregated from the entire compute system increases, application performance decreases. All things equal, the storage system is typically the biggest contributor to total latency. This makes it essentially a baseline for tuning all the components of an application environment. Performance improvements delivered by the storage are passed along to the rest of the environment (though these improvements can be wasted by poor tuning at the host).

The HP 3PAR StoreServ 7450c performs very well out of the box, but in an 8Gb FC infrastructure there is a limit as to how low I/O latencies can go. If cost and effort are not factors, we recommend upgrading the entire SAN to 16Gb FC to reap the full benefits of all-flash storage systems. However, this type of upgrade may not be possible for some data centers. For these situations, HP developed 8Gb Express Writes as an incremental performance boost, which can be applied with a non-intrusive array operating system upgrade. Deploying any of these solutions lowered I/O latency from 10% – 21% for the Oracle database workload demonstrated in this report. Other workload and I/O patterns have seen varying levels of benefit as well, some in excess of 50% improvement.

As the demand for extremely low latency increases, microseconds become more and more important. We strongly recommend that owners of HP 3PAR StoreServ storage systems investigate deploying 16Gb FC technology, or least 8Gb Express Writes, to get the very best out of their storage systems.

Appendix A – Test Description and Environment

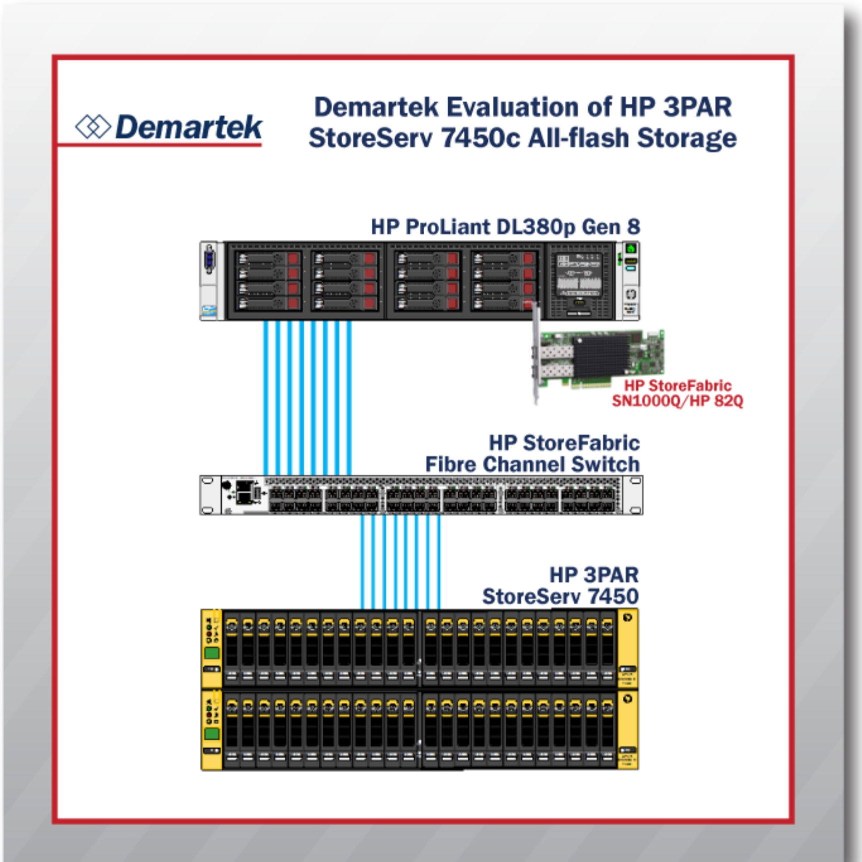

Figure 11 - Test Infrastructure

Server

- HP ProLiant DL380 G8
- 2 Intel E5-2630, 2.3GHz CPUs
- 15 GB RAM
- 4 Qlogic 8Gb FC dual port HBA (OEM'ed by HP as HP 82Q HBAs)
- 4 HP 16Gb FC HBAs – SN1000Q (OEM'ed Qlogic 16Gb FC dual port HBA)
- Microsoft Windows Server 2012 R2
- Oracle 12c, HammerDB 2.16 TPC-C workload

Fibre Channel Switch

- Brocade 300e 8Gb FC switch
- Brocade 6510 16Gb FC switch (sold by HP as the HP StoreFabric SN6000B Fibre Channel Switch)

Storage Array

- HP 3PAR StoreServ 7450c 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

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

The HP 3PAR StoreServ 7450c 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 7450c 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 7450c in this evaluation, enhancements like multi-tenant I/O processing help 7450c 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 - HP StoreFabric Gen 5 Fibre Channel Infrastructure

HP StoreFabric Gen 5 Fibre Channel Infrastructure is built on the OEM relationship between HP, QLogic, and Brocade. Qlogic 16Gb Gen 5 Fibre Channel host bus adapters are sold by HP as the HP SN1000Q HBA and Brocade Gen 5 Fibre Channel switches as the HP B-series. The HP StoreFabric Gen 5 Fibre Channel infrastructure has been engineered by HP, QLogic, and Brocade to deliver several key benefits to the storage network.

Enhanced Diagnostics

Gen 5 Fibre Channel improvements extend Brocade Fabric Vision™ from the QLogic Adapters deployed in HP servers across the SAN to storage targets on HP arrays. This is accomplished through Brocade ClearLink™ Diagnostics support in the HP SN1000Q 16Gb Fibre Channel Adapters and HP B-series Fibre Channel switches. Clearlink Diagnostics takes advantage of features in the ASIC and optics to test and report on the following, end-to-end, across the SAN:

- Electrical loopback
- Optical loopback
- Link distance and latency

Along with support for Fibre Channel Ping and Traceroute, these benefits provide:

- Fault detection in the FC SAN infrastructure
- Improved uptime and performance
- Reduction in operation expenses associated with maintaining FC SAN infrastructure

Rapid SAN Deployment

The Clearlink Diagnostics supported by HP StoreFabric Gen 5 Fibre Channel Infrastructure provides improved deployment of SAN resources by avoiding or identifying issues while in the pre-deployment phase. In addition, dynamic fabric provisioning with HP SN1000Q HBAs acquire port WWN addressed from HP B-series fabrics. These features deliver the following benefits:

- Reduce or prevent fabric reconfiguration when adding or replacing servers

- Reduce or eliminate manual deployment processes for deploying and administering the FC SAN by use of a single console (Brocade Network Advisor) to monitor and manage all SAN infrastructure components
- Reduce or eliminate the need for modifying zoning and LUN masking

Improving Quality of Service for Virtualization Scalability

HP SN1000Q HBAs work in conjunction with the CS_CTL QoS feature on HP B-series switches and supported targets. This extend fabric QoS to the host, which is good for physical environments, but particularly allows virtual environments to scale rapidly without compromising business SLAs through an end-to-end priority classification of Fibre Channel traffic by port or virtual port.

More information on the 16Gb Gen 5 Fibre Channel SAN technology delivered by QLogic and Brocade is available in the QLogic solution sheet *QLogic and Brocade Technology Alliance Drives Gen 5 Fibre Channel to New Levels for HP Customers*[5].

[5] QLogic," QLogic and Brocade Technology Alliance Drives Gen 5 Fibre Channel to New Levels for HP Customers", http://www.qlogic.com/OEMPartnerships/HP/Documents/Storage/SS_QLogic_and_Brocade_Drives_Gen5_FC.pdf (April 2015)

Appendix D - 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[6,7] have examined the overall performance of the HP 3PAR StoreServ 7450c 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 and in fact the single server deployed as the workload host did not possess the processing power to do so. Instead we chose to focus on latency as the metric to best highlight the effect of 8Gb Express Writes on a workload. (See Appendix D of this report for more information of latency in SAN environments.)

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 didn't want to turn this test into a tuning exercise. 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

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

[7] Demartek, "Demartek Evaluation of HP 3PAR StoreServ 7450c 16GFC All-Flash Storage", http://www.demartek.com/Demartek_HP_3PAR_StoreServ_7450c_16GFC_2015-03.html (March 30, 2015)

Writes could offer, we decided to exclude all the "upstream" contributors to latency by measuring performance at the array.

Appendix E – 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.

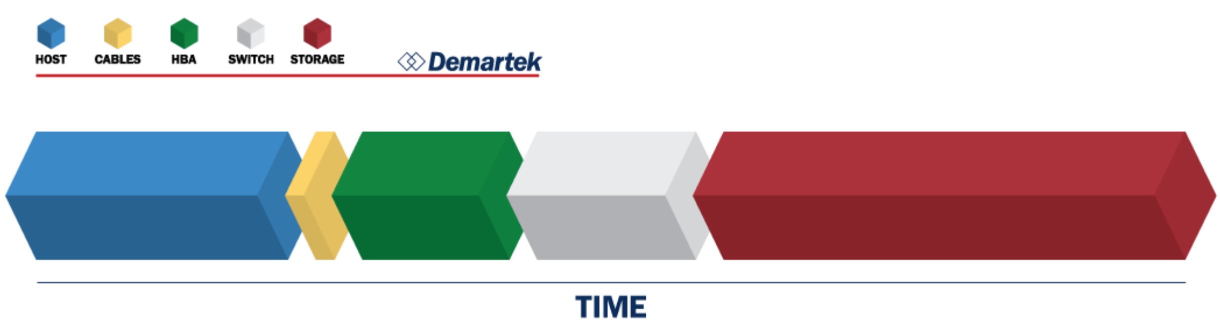

Figure 12 - Contributors to overall latency

Latency in a SAN is introduced from several sources (Figure 12). 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.

Appendix F – Vdbench Parameter File for 8Kb I/O Testing.

```
#concatenate=yes

sd=sd1,lun=/dev/mapper/mpathb,size=100GB,openflags=o_direct
sd=sd2,lun=/dev/mapper/mpathc,size=100GB,openflags=o_direct
sd=sd3,lun=/dev/mapper/mpathd,size=100GB,openflags=o_direct
sd=sd4,lun=/dev/mapper/mpathe,size=100GB,openflags=o_direct
sd=sd5,lun=/dev/mapper/mpathf,size=100GB,openflags=o_direct
sd=sd6,lun=/dev/mapper/mpathg,size=100GB,openflags=o_direct
sd=sd7,lun=/dev/mapper/mpathh,size=100GB,openflags=o_direct
sd=sd8,lun=/dev/mapper/mpathi,size=100GB,openflags=o_direct
sd=sd9,lun=/dev/mapper/mpathj,size=100GB,openflags=o_direct
sd=sd10,lun=/dev/mapper/mpathk,size=100GB,openflags=o_direct

#Read only workload
wd=wd_read,xfersize=8k,sd=sd*,seekpct=0,rdpct=100

#Pre-fill workload
wd=wd_fill,xfersize=8k,sd=sd*,seekpct=0,rdpct=0

# Workload Definition for random I/O
wd=wd_rand_read95,xfersize=8k,sd=sd*,seekpct=rand,rdpct=95
wd=wd_rand_read90,xfersize=8k,sd=sd*,seekpct=rand,rdpct=90
wd=wd_rand_read85,xfersize=8k,sd=sd*,seekpct=rand,rdpct=85
wd=wd_rand_read80,xfersize=8k,sd=sd*,seekpct=rand,rdpct=80
wd=wd_rand_read75,xfersize=8k,sd=sd*,seekpct=rand,rdpct=75
wd=wd_rand_read70,xfersize=8k,sd=sd*,seekpct=rand,rdpct=70

# Workload Definition for 75% random I/O
wd=wd_25perseq_read95,xfersize=8k,sd=sd*,seekpct=75,rdpct=95
wd=wd_25perseq_read90,xfersize=8k,sd=sd*,seekpct=75,rdpct=90
wd=wd_25perseq_read85,xfersize=8k,sd=sd*,seekpct=75,rdpct=85
wd=wd_25perseq_read80,xfersize=8k,sd=sd*,seekpct=75,rdpct=80
wd=wd_25perseq_read75,xfersize=8k,sd=sd*,seekpct=75,rdpct=75
wd=wd_25perseq_read70,xfersize=8k,sd=sd*,seekpct=75,rdpct=70

# Workload Definition for 50% random I/O
wd=wd_50perseq_read95,xfersize=8k,sd=sd*,seekpct=50,rdpct=95
wd=wd_50perseq_read90,xfersize=8k,sd=sd*,seekpct=50,rdpct=90
wd=wd_50perseq_read85,xfersize=8k,sd=sd*,seekpct=50,rdpct=85
wd=wd_50perseq_read80,xfersize=8k,sd=sd*,seekpct=50,rdpct=80
wd=wd_50perseq_read75,xfersize=8k,sd=sd*,seekpct=50,rdpct=75
wd=wd_50perseq_read70,xfersize=8k,sd=sd*,seekpct=50,rdpct=70

# Workload Definition for 25% random I/O
wd=wd_75perseq_read95,xfersize=8k,sd=sd*,seekpct=25,rdpct=95
wd=wd_75perseq_read90,xfersize=8k,sd=sd*,seekpct=25,rdpct=90
wd=wd_75perseq_read85,xfersize=8k,sd=sd*,seekpct=25,rdpct=85
wd=wd_75perseq_read80,xfersize=8k,sd=sd*,seekpct=25,rdpct=80
wd=wd_75perseq_read75,xfersize=8k,sd=sd*,seekpct=25,rdpct=75
wd=wd_75perseq_read70,xfersize=8k,sd=sd*,seekpct=25,rdpct=70

# Workload Definition for sequential I/O
wd=wd_seq_read95,xfersize=8k,sd=sd*,seekpct=0,rdpct=95
wd=wd_seq_read90,xfersize=8k,sd=sd*,seekpct=0,rdpct=90
wd=wd_seq_read85,xfersize=8k,sd=sd*,seekpct=0,rdpct=85
wd=wd_seq_read80,xfersize=8k,sd=sd*,seekpct=0,rdpct=80
wd=wd_seq_read75,xfersize=8k,sd=sd*,seekpct=0,rdpct=75
wd=wd_seq_read70,xfersize=8k,sd=sd*,seekpct=0,rdpct=70
```

```
#Run Definitions

#Read-only
#rd=rd_read,wd=wd_read,iorate=max,warmup=1m,elapsed=60m,interval=5,thread=10


#Pre-fill
rd=rd_fill,wd=wd_fill,iorate=max,warmup=1m,elapsed=60m,interval=5,thread=10

# Random I/O
rd=rd_random_5w,wd=wd_rand_read95,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_random_10w,wd=wd_rand_read90,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_random_15w,wd=wd_rand_read85,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_random_20w,wd=wd_rand_read80,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_random_25w,wd=wd_rand_read75,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_random_30w,wd=wd_rand_read70,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10

# 75% random I/O
rd=rd_75per_random_5w,wd=wd_25perseq_read95,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_75per_random_10w,wd=wd_25perseq_read90,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_75per_random_15w,wd=wd_25perseq_read85,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_75per_random_20w,wd=wd_25perseq_read80,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_75per_random_25w,wd=wd_25perseq_read75,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_75per_random_30w,wd=wd_25perseq_read70,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10

# 50% random I/O
rd=rd_50per_random_5w,wd=wd_50perseq_read95,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_50per_random_10w,wd=wd_50perseq_read90,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_50per_random_15w,wd=wd_50perseq_read85,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_50per_random_20w,wd=wd_50perseq_read80,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_50per_random_25w,wd=wd_50perseq_read75,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_50per_random_30w,wd=wd_50perseq_read70,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10

# 25% random I/O
rd=rd_25per_random_5w,wd=wd_75perseq_read95,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_25per_random_10w,wd=wd_75perseq_read90,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_25per_random_15w,wd=wd_75perseq_read85,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_25per_random_20w,wd=wd_75perseq_read80,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_25per_random_25w,wd=wd_75perseq_read75,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_25per_random_30w,wd=wd_75perseq_read70,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10

# Sequential I/O
rd=rd_sequential_5w,wd=wd_seq_read95,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_sequential_10w,wd=wd_seq_read90,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_sequential_15w,wd=wd_seq_read85,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_sequential_20w,wd=wd_seq_read80,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_sequential_25w,wd=wd_seq_read75,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
rd=rd_sequential_30w,wd=wd_seq_read70,iorate=max,warmup=1m,elapsed=5m,interval=5,thread=10
```

The original version of this document is available at:
http://www.demartek.com/demartek_HP_3PAR_StoreServ_7450_Improving_Oracle_Database_Response_Times_2015-06.html on the Demartek website.

Intel and Xeon are registered trademarks of Intel Corporation.

Microsoft, Windows, and Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

TPC Benchmark and TPC-H are trademarks of the Transaction Performance Council.

Demartek is a trademark of Demartek, LLC.

All other trademarks are the property of their respective owners.