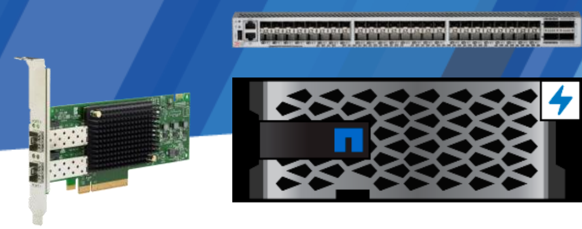
Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

*NVMe™ over Fibre Channel delivered **58% higher IOPS** and **34% lower latency** than SCSI FCP. (What's not to like?)*



Executive Summary

NetApp's ONTAP 9.4 is the first generally available enterprise storage offering enabling a complete **NVMe™ over Fibre Channel (NVMe/FC)** solution. NVMe/FC solutions are based on the recent T11/INCITS committee **FC-NVMe** block storage standard, which specifies how to extend the NVMe command set over Fibre Channel in accordance with the NVMe over Fabrics™ (NVMe-oF™) guidelines produced by the NVM Express™ organization.

Fibre Channel is **purpose-built for storage** devices and systems and is the de facto standard for storage area networking (SAN) in enterprise datacenters. Fibre Channel operates in a lossless fashion with hardware offload Fibre Channel adapters, with hardware-based congestion management, providing a reliable, credit-based flow control and delivery mechanism, meeting the technical requirements for NVMe/FC.

Today's Fibre Channel adapters have the added benefit of being able to run traditional Fibre Channel Protocol (SCSI FCP) that uses the SCSI command set **concurrently** with the NVMe over Fibre Channel command set in the same adapter, the same Fibre Channel Network, and the same Enterprise All Flash Arrays (AFAs). The NetApp AFF A700s is the first array to support both SCSI FCP and NVMe/FC concurrently on the same port. This provides **investment protection** for existing FC adapters while offering the ***performance benefits of NVMe/FC with a simple software upgrade***. Modern Fibre Channel switches and host bus adapters (HBAs) already support both traditional SCSI FCP and NVMe/FC concurrently.

For this test report, Demartek worked with NetApp and Broadcom (Brocade and Emulex divisions) to demonstrate the benefits of NVMe over Fibre Channel on the NetApp AFF A700s, Emulex Gen 6 Fibre Channel Adapters, and Brocade Gen 6 Fibre Channel SAN switches.

Key Findings and Conclusions

> ***NVMe/FC enables new SAN workloads:*** Big data analytics, Internet of Things (IoT) and A.I. / deep learning will all benefit from the faster performance and lower latency of NVMe/FC.

> ***NVMe/FC accelerates existing workloads:*** Enterprise applications such as Oracle, SAP, Microsoft SQL Server and others can immediately take advantage of NVMe/FC performance benefits.

> ***Test results:*** in our tests, we observed up to **58% higher IOPS** for NVMe/FC compared to SCSI FCP ***on the same hardware***. We also observed minimum differences, depending on the tests, of 11% to 34% lower latency with NVMe/FC.

> ***NVMe/FC is easy to adopt:*** All of the performance gains we observed were made possible by a software upgrade.

> ***NVMe/FC protects your investment:*** The benefits we observed were with existing hardware that supports 32GFC.

> ***NVMe/FC Datacenter consolidation***: More work can be completed in the same hardware footprint with increased IOPS density.

Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

What is NVMe over Fibre Channel?

NVMe over Fibre Channel is a solution that is defined by two standards: NVMe-oF and FC-NVMe. NVMe-oF is a specification from the NVM Express organization that is transport agnostic, and FC-NVMe is an INCITS T11 standard. These two combine to define how NVMe leverages Fibre Channel. NVMe over Fibre Channel was designed to be backward compatible with the existing Fibre Channel technology, supporting both the traditional SCSI protocol and the new NVMe protocol using the same hardware adapters, Fibre Channel switches, and Enterprise AFAs.

Purpose-Built for Storage

Fibre Channel storage fabrics provide consistent and highly reliable performance and are a separate, dedicated storage network that completely isolates storage traffic. FC fabrics have a built-in, **proven method to discover host initiators and storage devices** and their properties on the fabric. These devices can be initiators, such as host application servers with FC host bus adapters (FC HBAs) and storage systems, also known as storage targets.

Rapid access to data is critically important to today's enterprise datacenters. Traditional Fibre Channel fabrics are typically deployed with redundant switches and ports that support **multi-path I/O**, so that in the event of a link failure, an alternate path is available, maintaining constant access to data. NVMe/FC also supports multi-path I/O and supports preferred path with the addition of **Asymmetric Namespace Access** (ANA). ANA was added to the NVMe specification and ratified in March 2018 as a technical proposal (TP 4004). This requires both initiators and targets to implement ANA. Demartek believes that preferred path support (via ANA mechanisms) will become available in some NVMe solutions during this calendar year.

Note: ANA applies only to NVMe – other storage protocols have their methods to implement multi-path and preferred path support.

The technology used in FC fabrics is backwards compatible with at least the two previous generations. This provides long-term **investment protection** for an organization's critical data assets and aids in long-term capital budgeting planning.

Fibre Chanel fabrics are designed to support multiple protocols including NVMe over Fibre Channel concurrently with SCSI over Fibre Channel. This provides organizations the ability to easily deploy NVMe over Fibre Channel on their current servers with Emulex Fibre Channel cards, Brocade Fibre Channel Switches, and NetApp All Flash Arrays.

Why Move to NVMe over Fibre Channel?

The vast majority of enterprise datacenters use Fibre Channel SANs to store mission-critical data. Many of the customers running these datacenters already have the hardware necessary to run NVMe/FC, including Fibre Channel switches, adapters and storage. For this test, moving to NVMe/FC with this existing hardware requires only a software upgrade on the host initiators and the storage targets. Because SCSI FCP and NVMe/FC can run on the same wire at the same time, NVMe namespaces can be created as needed to replace existing application SCSI LUNs and applications can reference the NVMe namespaces to get immediate performance benefits.

NVMe/FC Benefits – NetApp Storage System

In this test, the lion's share of the performance improvement comes from adding NVMe over Fibre Channel to the storage array – ***The primary performance benefit is faster AFAs***. Because NVMe is more efficient than older protocols, a number of benefits are available with NVMe/FC fabrics. These benefits pertain to the traffic carried over the fabric and are independent of the type of storage devices inside the storage system connected via NVMe/FC.

NetApp's ONTAP 9.4 includes several new features with respect to automatic cloud tiering of cold data, support for 30TB SSDs and new compliance and security features including compliance with GDPR. However, the main new feature highlighted in this report is support for NVMe/FC.

Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

IOPS Benefits

A more efficient command set can deliver higher IOPS. In our tests, we observed up to a 58% increase in IOPS by simply moving over to NVMe/FC from the traditional SCSI FCP command set.

Latency Benefits

NVMe/FC has lower latency than traditional SCSI FCP. We also observed minimum differences, depending on the tests, of 11% to 34% lower latency with NVMe/FC.

Better Performance with Existing Hardware

NetApp achieves these benefits by simply applying a software upgrade license to the A700s. By moving to NVMe/FC with the same storage hardware, dramatic increases in performance are available. The back-end flash SSDs use existing interfaces.

NVMe/FC Benefits – FC Switches

Brocade Gen 6 Fibre Channel fabrics transport both NVMe and SCSI (SCSI FCP) traffic concurrently with same high bandwidth and low latency. Overall, the NVMe performance benefits are in the end nodes – initiators and targets. NVMe/FC provides the same proven security that traditional Fibre Channel protocol has provided for many years. Fibre Channel provides full fabric services for NVMe/FC and SCSI FCP such as discovery and zoning. Finally, NVMe over FC is the first NVMe-oF transport that meets the same high bar as SCSI over FC with full-matrix testing as an enabler and essential for enterprise level support.

Brocade switches include ***IO Insight***, which proactively monitors I/O performance and behavior through integrated network sensors to gain deep insight into problems and ensure service levels. This capability non-disruptively and non-intrusively gathers I/O statistics for both SCSI and NVMe traffic from any device port on a Gen 6 Fibre Channel platform, then applies this information within an intuitive, policy-based monitoring and alerting suite to configure thresholds and alarms.

NVMe/FC Benefits – FC HBAs

The test data in this report represents the performance improvement of NVMe over Fibre Channel for the complete solution. To better explain the performance benefits of NVMe over Fibre Channel, it helps to describe the performance improvements for workloads on the server. NVMe over Fibre Channel brings native parallelism and efficiency to block storage that SCSI FCP cannot and delivers meaningful performance improvement for application workloads. We reviewed test results from Broadcom (Emulex division).

When testing initiator performance for characteristics such as maximum IOPs, it is essential to use either an extremely fast target or multiple targets to remove any bottlenecks that may distort the test results.

The data shows the following results:

- The target-side efficiency of NVMe enables a single initiator to exceed 1 million IOPS with fewer targets than with SCSI FCP targets.
- 2x improvement in IOPS at 4KB I/Os with moderate workloads.
  - 2x improvement in PostgreSQL transaction rate
  - 50% or more reduction in latency
  - At least 2x higher IOPS when normalized to CPU utilization

Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

Test Configuration – Hardware

This section describes the servers, storage and storage networking configuration for this study. It is important to note that although all of the elements in the configuration are capable of supporting **concurrent** NVMe/FC and SCSI/FC, for this study they were configured separately in order to simplify modification and optimization of specific parameters for one protocol without impacting the behavior for the other protocol.

Servers (qty. 4)

> Fujitsu RX300 S8
> 2x Intel Xeon E5-2630 v2, 2.6 GHz, 6c/12t
> 256 GB RAM (16x 16GB)
> BIOS V4.6.5.4 R1.3.0 for D2939-B1x
> SLES12SP3 4.4.126-7.ge7986b5-default

Fibre Channel Switch

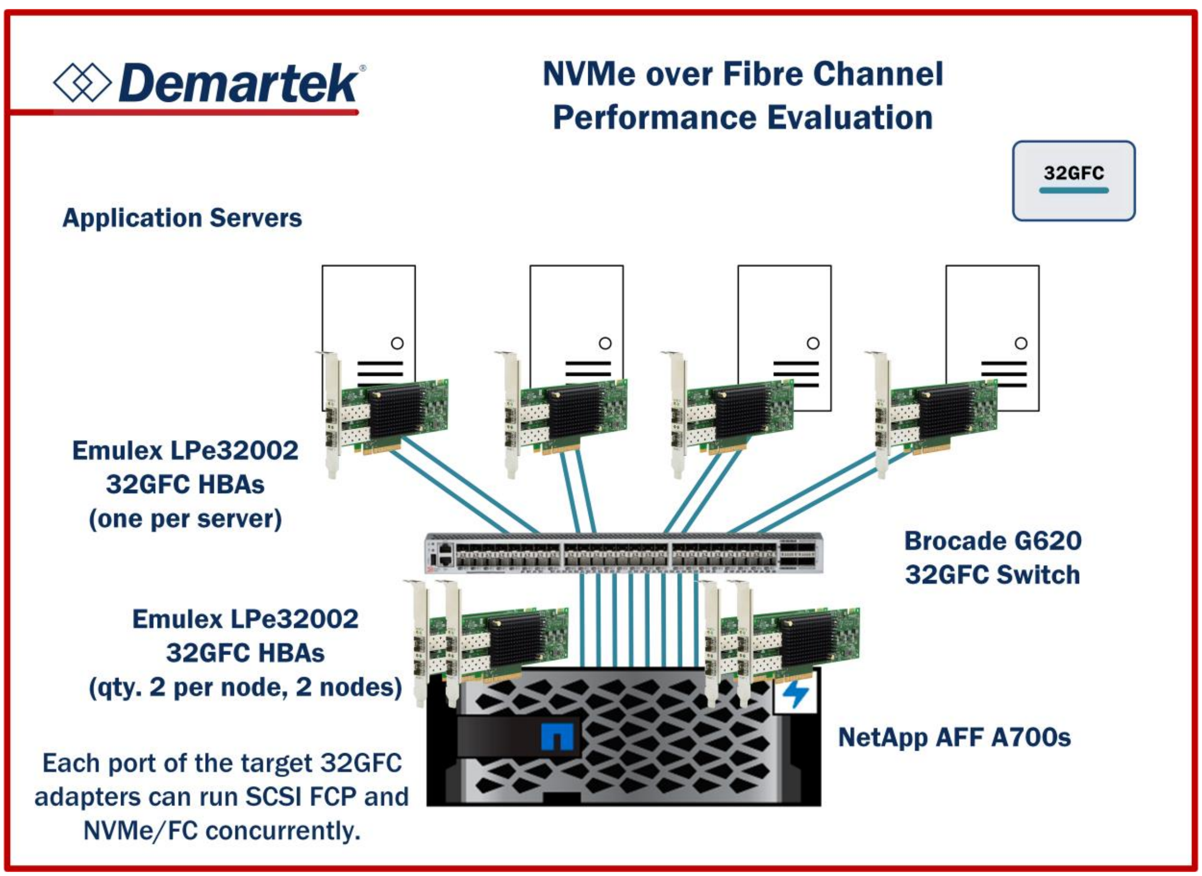
> Brocade G620, 48 ports, 32GFC
> FOS 8.1.0a

Storage System

> NetApp AFF A700s
> ONTAP 9.4
> 4 target ports on each of two nodes, 32GFC
> 24x SAS SSD, 960 GB each

Fibre Channel HBA

> Emulex LPe32002 32GFC supporting SCSI FCP and NVMe/FC
> Firmware version: 11.4.204.25
> Driver version 11.4.354.0



Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

Test Methodology

The purpose of our test was to compare performance metrics of NVMe/FC against SCSI FCP on the AFF A700s storage system. Assessing the maximum overall IOPS for the storage system was not a focus of this study. The following sections describe the test methodology and design considerations used to measure the performance of these two protocols while running a suite of synthetic workloads.

In our study, we configured four servers running SUSE Enterprise Linux 12.3 to a single A700s 2-node HA storage controller via a Brocade G620 network switch.

The A700s storage controller in our testbed contained two storage nodes. For the purposes of this test, one storage node was used to host the storage for NVMe/FC containers and one storage node for the SCSI FCP containers. This test design was used to guarantee the full performance for each protocol.

Table 1 provides the details of the NetApp storage controller configuration.

Storage system Active Pair	AFF A700s configured as a highly available (HA) active-active-pair
ONTAP version	ONTAP 9.4 (pre-release)
Total number of drives per node	24
Drive size	960GB
Drive type	SAS SSD
SCSI FCP target ports	4 – 32Gb ports
NVMe/FC target ports	4 – 32Gb ports
Ethernet ports	4 – 10Gb ports (2 per node)
Ethernet logical interfaces (LIFs)	4 – 1Gb management LIFs (2 per node connected to separate private VLANs)
FCP LIFs	8 – 32Gb data LIFs

During our testing, only one protocol and workload was active at a given time. Note that although every component (the servers, the HBAs, the switch and the AFF A700s) involved in this test is capable of supporting concurrent FC-NVMe and FC-SCSI production traffic, the protocols were isolated during the testing in order to enable gathering of independent metrics for each protocol and to simplify the tuning of independent specific parameters for each protocol.

We created one aggregate in ONTAP on each of the two storage nodes, named NVMe_aggr and FCP_aggr, respectively. Each aggregate consumed 23 data partitions spanning 23 of the 24 SAS-attached SSDs, leaving one partition spare for each data aggregate.

The NVMe_aggr contained four 512GB namespaces. Each 512GB namespace was mapped to a single SUSE host to drive IO. Each namespace was contained in its own FlexVol. Each namespace was associated with its own subsystem.

The FCP_aggr contained 16 LUNs, each contained within its own FlexVol. Total container size was the same as the NVMe namespaces. Each LUN was mapped to each of the four SUSE hosts to receive IO traffic evenly.

We used the Vdbench load generation tool to generate workload mixes against an A700s storage target. Vdbench is an open source workload generator provided by Oracle that can be found at http://www.oracle.com/technetwork/server-storage/vdbench-downloads-1901681.html. Vdbench generates a variety of IO mixes, ranging from small random IOs, large sequential IOs, and mixed workloads designed to emulate real application traffic.

We first conducted an initial write phase to populate the thin-provisioned LUNs and namespaces. This phase writes through each LUN/namespace exactly one time with non-zero data. This ensures that we are not reading uninitialized portions of LUN or namespace that can be satisfied from the A700s without due processing.

We designed our Vdbench workloads to highlight a range of use cases. These use cases provided a general overview of performance and demonstrate the performance differences between SCSI FCP and NVMe/FC in ONTAP 9.4.

1. Synthetic "4-Corners" Testing: 16 Java Virtual Machines (JVMs), 128 threads for SCSI FCP, 512 threads for NVMe/FC
    a. **Large Sequential Reads (64K)**
    b. Large Sequential Writes (64K)
    c. **Moderate Sequential Reads (32K)**
    d. Moderate Sequential Writes (32K)
    e. **Small Random Reads (4K)**
    f. Small Random Writes (4K)
    g. Mixed Random Reads and Writes (4K)
2. Emulated Oracle OLTP Workload: 16 JVMs, 100 threads
    a. **80/20 8K Read/Write mix**
    b. 90/10 8K Read/Write mix
    c. **80/20 8K Read/Write mix with a separate stream of 64K Sequential Writes emulating redo logging**

Note: performance results are provided for the items in **bold text** above.

Workload Design

We used Vdbench 5.04.06 and Java 1.8.0_66-b17 to drive different IOPS mixes against SCSI FCP and NVMe/FC storage. These mixes include an emulation of SLOB2 workloads by using profiles that mimic the storage load of an Oracle 12c database running an 80/20 select/update mix. We included other synthetic IO patterns to give a general indication of the difference in performance between SCSI FCP and NVMe/FC.

Note: We took care in these test steps to simulate real database and customer workloads, but we acknowledge that workloads vary across databases. In addition, these test results were obtained in a closed lab environment with no competing workloads on the same infrastructure. In a typical shared-storage infrastructure, other workloads share resources. Your results might vary from those found in this report.

Network Design

This section provides the network connectivity details for the tested configurations.

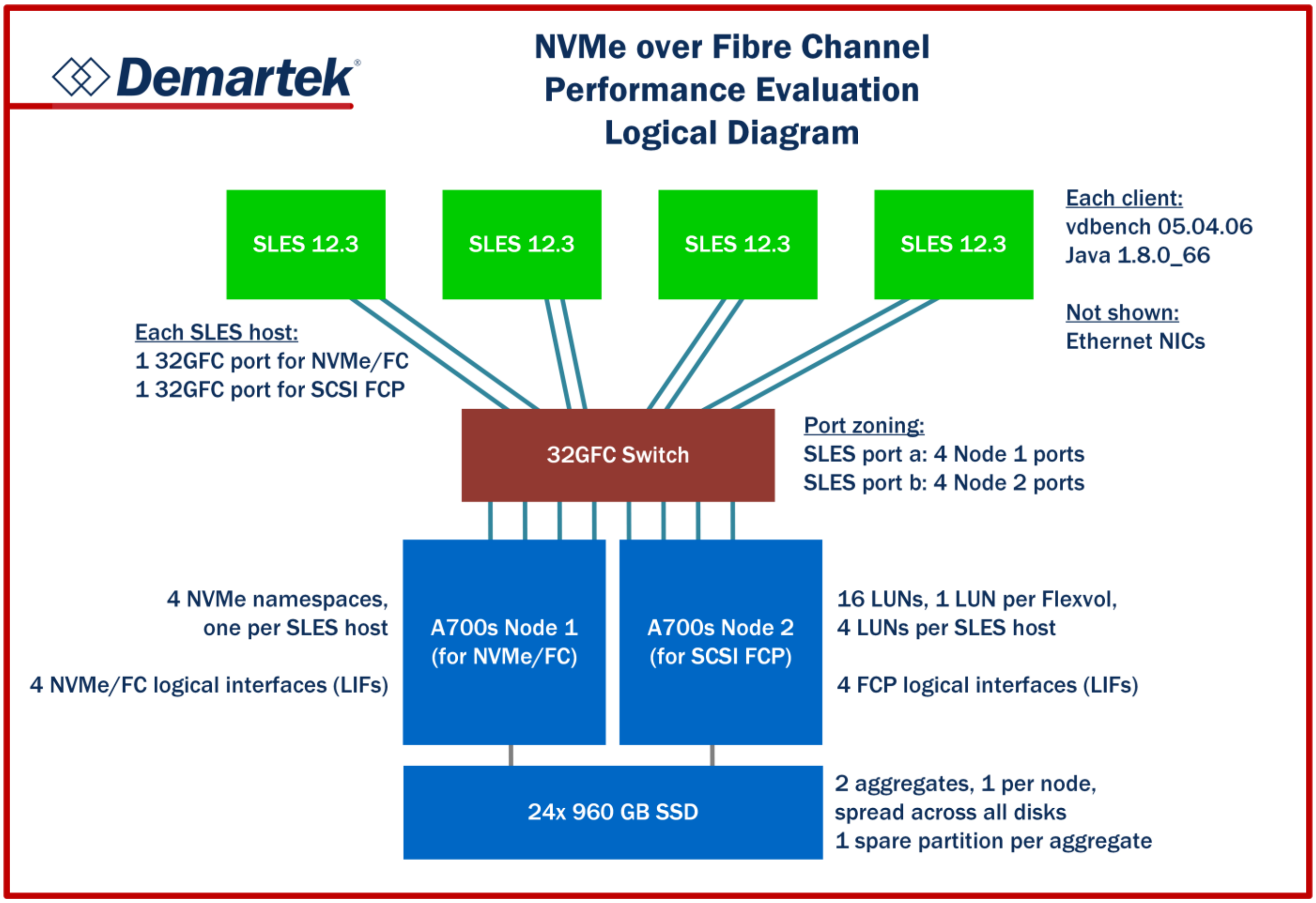
The network diagram shows that the FCP SAN was deployed with a Brocade G620 32Gb FCP switch. Each storage node had four ports connected to the FCP switch. Each server had two ports connected to the switch. At no point in the testing did the network connectivity create a bottleneck.

For Ethernet connectivity, each of the four hosts has a 1Gbps link for external access and to manage Vdbench coordination between nodes.

We used one igroup per server to contain the FCP initiators. We then used the "latency-performance" tuned profile to manage the SUSE hosts. We manually modified the FCP DM devices to use the "deadline" scheduler that improves performance for SCSI FCP.

Each of the four SUSE servers had a dual-port FC HBA that supports both protocols simultaneously. Both ports were connected to the Brocade switch. Each A700s node had four FC ports also connected to the same switch for eight total connected ports. We configured the Brocade switch with port zoning to map port 1 of each SUSE host to all four ports of the A700s storage node 1. Similarly, we mapped port 2 of each SUSE host to all four ports of the A700s storage node 2.

Test Environment Logical Diagram



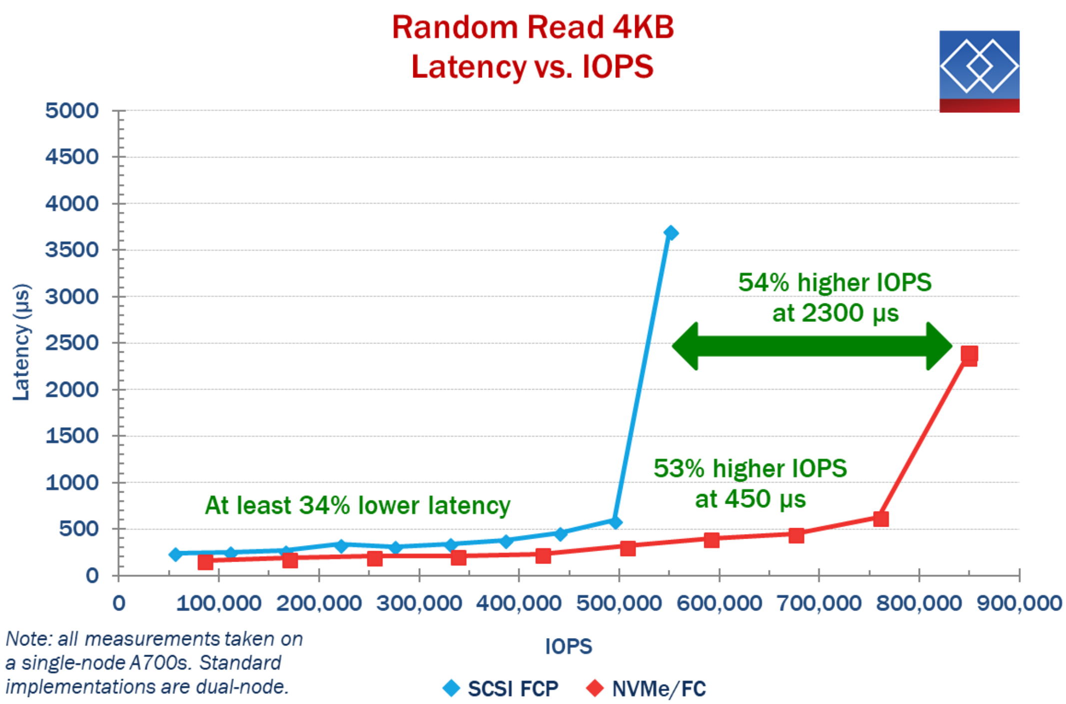
Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

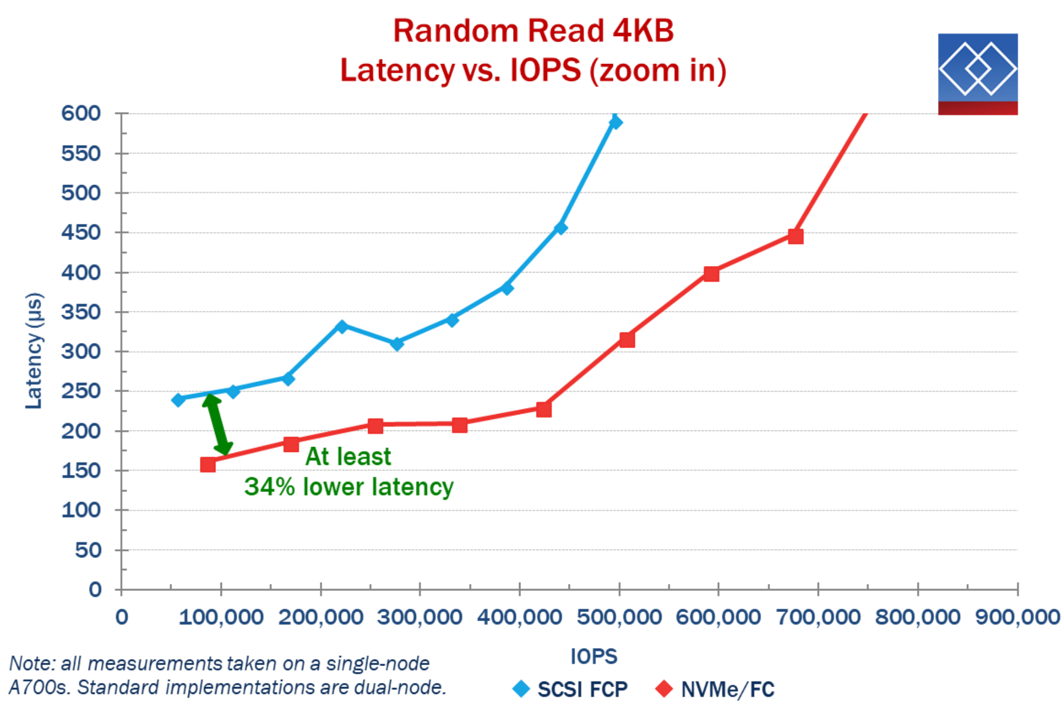
Performance Results

Selected results are shown on this page and the two following pages. All measurements were taken on a single-node A700s. Standard implementations use a dual-node configuration.

Random Read 4KB

For 4KB random reads, NVMe/FC achieved ***53% higher IOPS*** at 450 µs latency. Latency was at least 34% lower (better) for NVMe/FC. The second chart on this page "zooms in" on the latencies below 600 µs.

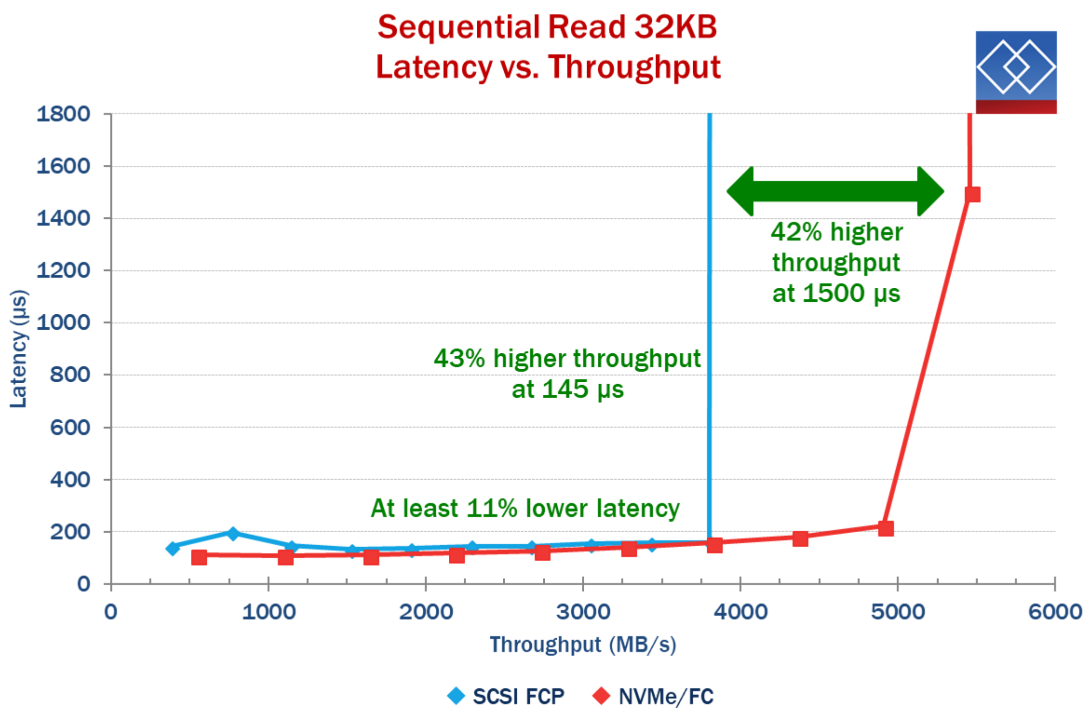


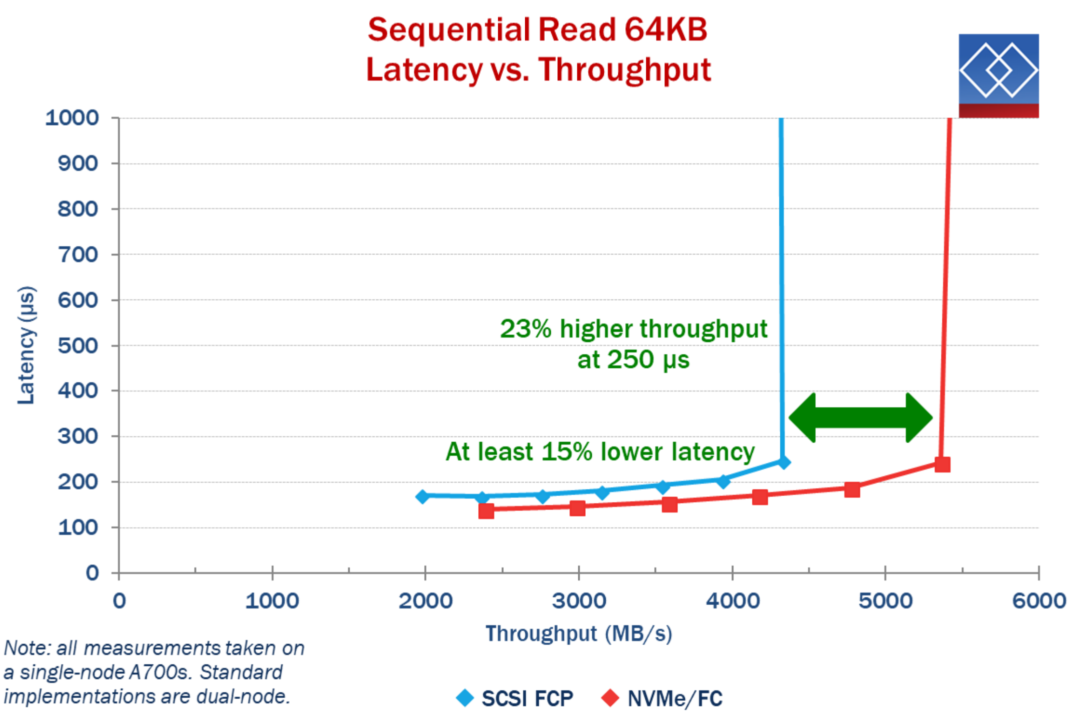


Sequential Read: 32KB and 64KB

For sequential reads at 32KB block size, NVMe/FC achieved 43% higher throughput at 145 µs. Latency was at least 11% for NVMe/FC.

For sequential reads at 64KB block size, NVMe/FC achieved 23% higher throughput at 250 µs. Latency was at least 15% lower for NVMe/FC.



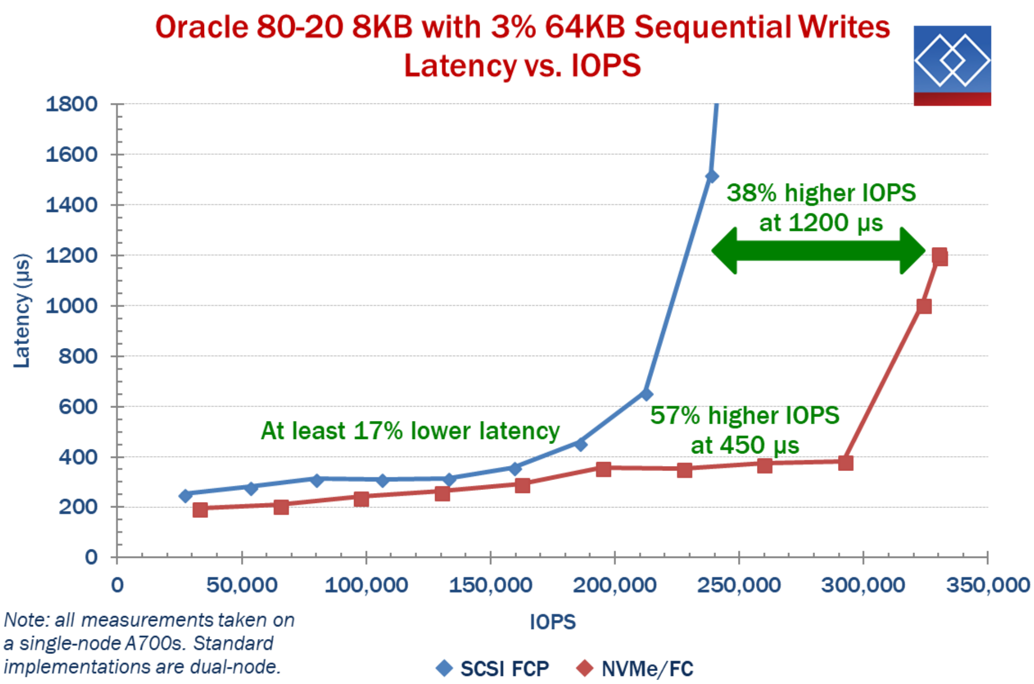


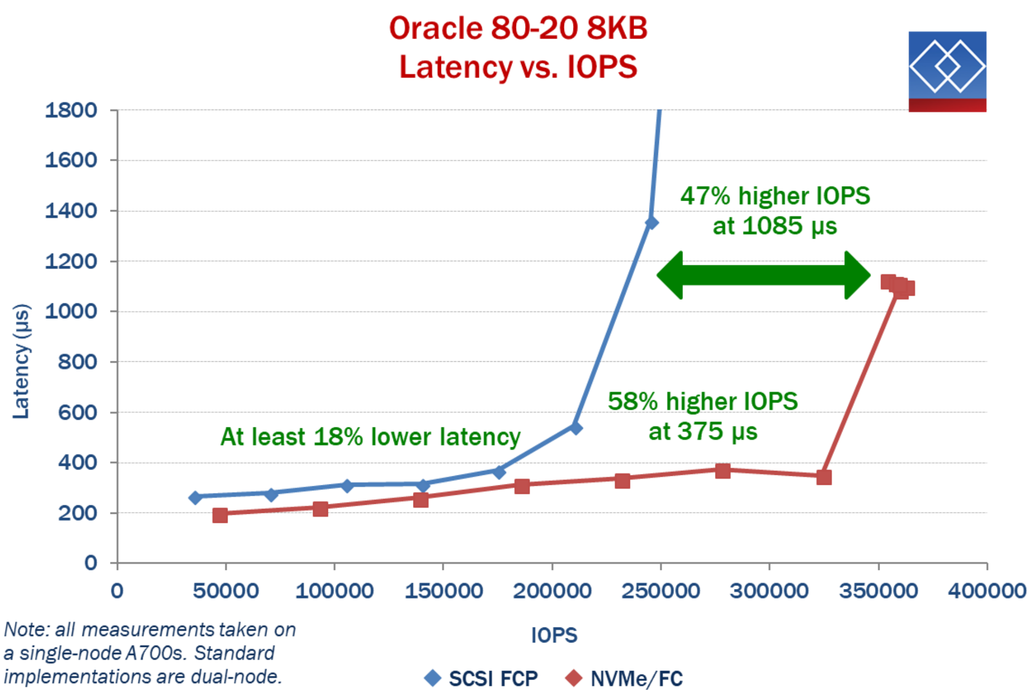
Note: all measurements taken on a single-node A700s. Standard implementations are dual-node.

Simulated Oracle 80-20 8KB Workloads

For the simulated Oracle workload with an 80/20 read/write mix at 8KB (typical OLTP database I/O) plus a small amount of 64KB sequential writes (typical redo logs), NVMe/FC achieved ***57% higher IOPS*** at 450 µs latency. Latency was at least 17% lower for NVMe/FC.

For the simulated Oracle workload with an 80/20 read/write mix at 8KB (typical OLTP database I/O)), NVMe/FC achieved ***58% higher IOPS*** at 375 µs latency. Latency was at least 18% lower for NVMe/FC.

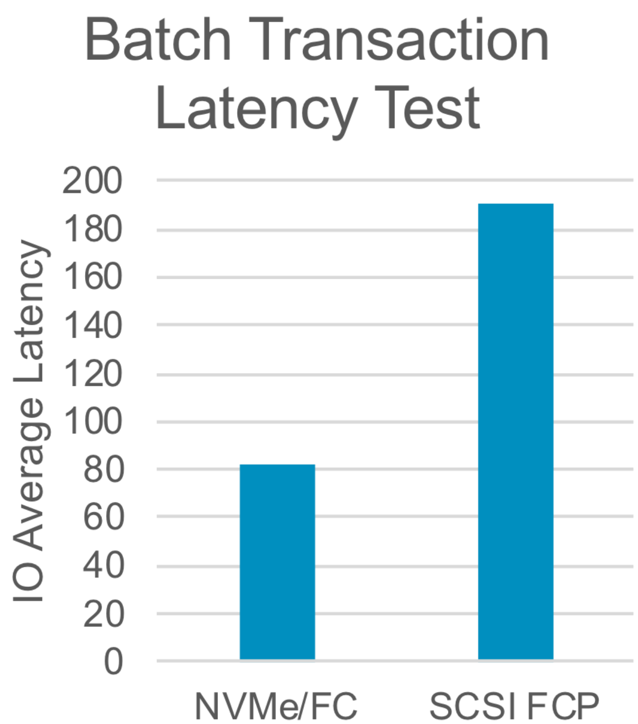




Performance Benefits of NVMe™ over Fibre Channel – A New, Parallel, Efficient Protocol

NetApp Performance Demos

In this report, we examined the performance improvement in the NetApp AFF A700s for a single node. NetApp can demonstrate NVMe/FC running on an A300 with ONTAP 9.4 for their enterprise customers. NetApp showed Demartek the following performance data with 4KB Random Read IOs, eight Threads, and a Queue Depth of 1. This FIO test configuration simulates multiple types of workloads, with this example being batch transactions.



Source: NetApp

The data from the NetApp demonstration shows that their NVMe/FC latency drops by half in the NetApp A300 – a level of latency only seen before from internal SATA and SAS SSDs. NetApp invites you to contact your NetApp Representative to schedule your NVMe over Fibre Channel demo today.

Summary and Conclusion

NVMe/FC leverages the parallelism and performance benefits of NVMe with the robust, reliable enterprise-grade storage area network technology of Fibre Channel.

In our tests, by using NVMe/FC we observed up to a 58% improvement in IOPS over traditional SCSI FCP with the same hardware. For the configuration tested, only a software upgrade was required in the host initiators and storage targets. This means that investments already made in Fibre Channel technology can be adopted easily without requiring the purchase of new hardware. This also means that more performance per square foot is possible, providing consolidation opportunities. Furthermore, by adopting NVMe/FC, there may be opportunities to delay purchases of new server and storage hardware, saving on potential hardware and software licensing costs.

NVMe/FC enables existing applications to accelerate performance and organizations to tackle demanding new applications such as Big data analytics, IoT and A.I. / deep learning with their existing infrastructure. For the configuration tested, all of this was possible with a software upgrade to the host initiators and storage targets. This makes NVMe/FC easy to adopt, at an organization's own pace, without requiring a forklift upgrade or learning the nuances of an entirely new fabric technology.

Demartek believes that NVMe/FC is an excellent (and perhaps obvious) technology to adopt, especially for those who already have Fibre Channel infrastructure, and is a good reason to consider Fibre Channel technology for those examining NVMe over Fabrics.