Technical Validation

Optimizing Oracle Database Efficiency and Performance with 32G End-to-End NVMe

IBM FlashSystem 9100 with b-type Storage Networking and Broadcom Emulex Fibre Channel Technology

By Brian Garrett, VP IT Validation Services; and Kerry Dolan, Senior IT Validation Analyst

October 2019

Contents

ESG Technical Validations

The goal of ESG Technical Validations is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Technical Validations are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.
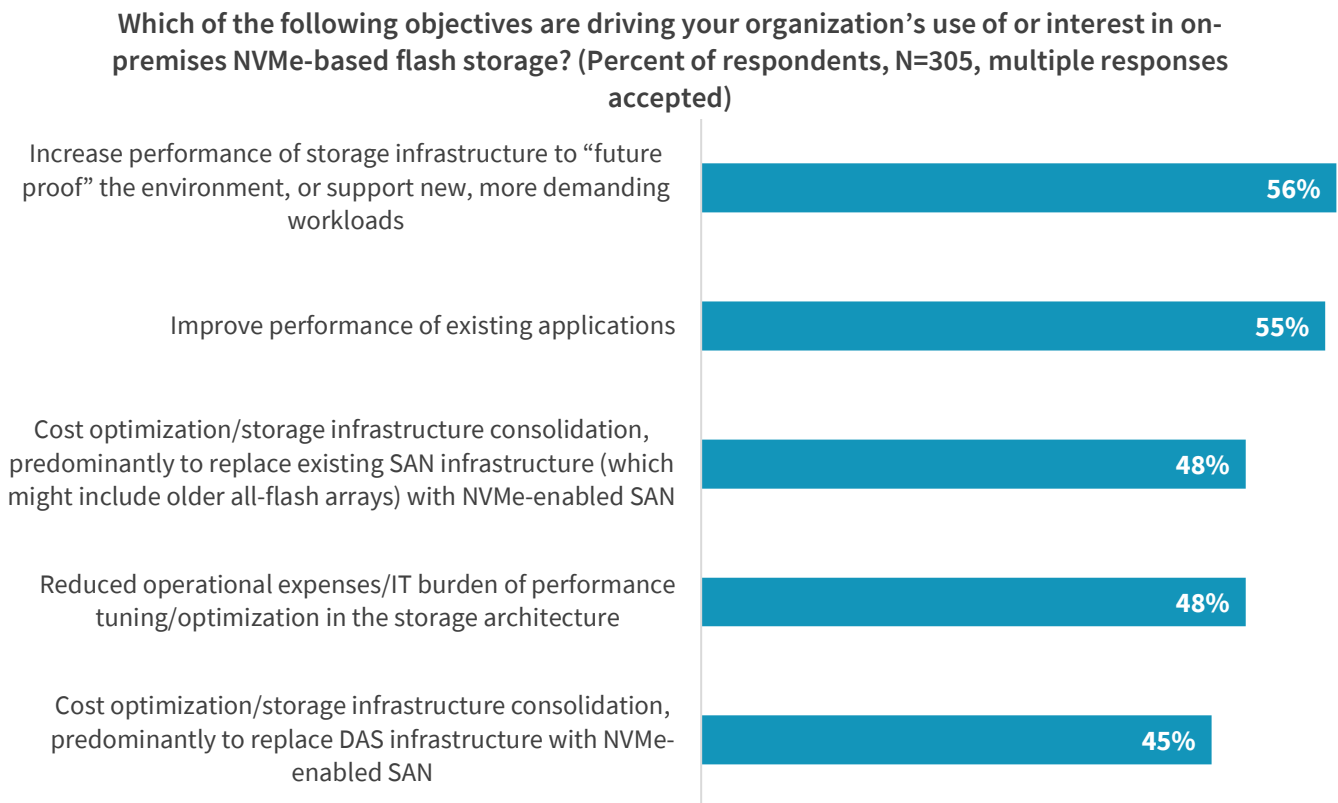
Introduction

This report documents performance testing of the IBM FlashSystem 9150 with end-to-end NVMe, including a comparison of 16GB vs. 32GB Fibre Channel (FC) Host Bus Adapters (HBAs) in the server and storage array along with the b-type storage network switch.

Background

Data drives business processes, making speed of access a business priority. Organizations are executing transactions and analyzing trends in real time to make better business decisions. Artificial intelligence (AI)- and machine learning (ML)-led innovations are also delivering business insights. These capabilities depend on fast access to data.

The advent of solid-state drives improved storage performance by orders of magnitude, trading hard drive spinning platters for speedy flash. However, SAS and SATA interfaces to solid-state drives have become a bottleneck. Drives that connect via Non-Volatile Memory Express (NVMe) reduce this connectivity bottleneck, enabling lower latency and faster data access directly to flash memory. Delivering higher performance in a smaller footprint also enables infrastructure consolidation for cost efficiency. ESG research reveals that organizations are selecting NVMe not just to increase performance of existing and future applications, but also to optimize costs.[1]

Figure 1. Objectives Driving the Adoption of NVMe

Which of the following objectives are driving your organization's use of or interest in on-premises NVMe-based flash storage? (Percent of respondents, N=305, multiple responses accepted)

Objective	Percent
Increase performance of storage infrastructure to "future proof" the environment, or support new, more demanding workloads	56%
Improve performance of existing applications	55%
Cost optimization/storage infrastructure consolidation, predominantly to replace existing SAN infrastructure (which might include older all-flash arrays) with NVMe-enabled SAN	48%
Reduced operational expenses/IT burden of performance tuning/optimization in the storage architecture	48%
Cost optimization/storage infrastructure consolidation, predominantly to replace DAS infrastructure with NVMe-enabled SAN	45%

Source: Enterprise Strategy Group

However, NVMe drives alone cannot solve all the potential I/O bottlenecks. Enabling *end-to-end* NVMe connectivity—from the host through the switch to the flash array—with NVMe over Fabric (NVMe-oF) can eliminate the bottlenecks and speed performance.

---

[1] Source: ESG Master Survey Results, *2019 Data Storage Trends,* to be published.

End-to-end NVMe with IBM FlashSystem 9150, IBM b-type 32G Storage Networking, and Emulex 32G Fibre Channel Server and Storage Adapters

The IBM FlashSystem 9150 was built to deliver fast data access and storage efficiency for secure, highly available environments. It offers end-to-end NVMe with Broadcom Emulex HBAs delivering high performance and the optimal NVMe experience. It boosts performance for physical, virtual, multi-cloud, and container environments. Among its newest features are:

- **NVMe over Fabric (NVMe-oF)**, which provides end-to-end NVMe with compatible hosts and includes native multi-pathing.

- **IBM b-type 32G FC storage network switching** to increase performance and extend bandwidth.

- **Emulex 32G FC HBAs** to improve performance, reliability, and management.

- **The option to use IBM FlashCore modules.** These are NVMe SSDs with FlashCore innovations for consistent microsecond latency, extreme reliability, and operational/cost efficiencies. These 3D TLC flash drives offer high density and include a hardware-accelerated NVMe architecture with inline, performance neutral hardware compression. Advanced flash management includes IBM Variable Stripe RAID, error-correction codes, and proprietary garbage collection algorithms that increase flash durability, accelerate performance, and reduce latency.

Figure 2 demonstrates the advantages of end-to-end NVMe using IBM FlashSystem 9150 with IBM b-type 32G storage networking. It provides frictionless data movement throughout the data path, from the host through the switch to the storage and back-end drives. This increased performance lets organizations accelerate the execution of business processes to drive revenue. It also enables more efficient use of storage with AI-powered storage insights; more efficient server usage including the opportunity to reduce CPU and memory requirements; and future-proof ROI, offering a rich set of data services across all storage systems with IBM Spectrum Virtualize.

Figure 2. IBM End-to-end NVMe over Fabric
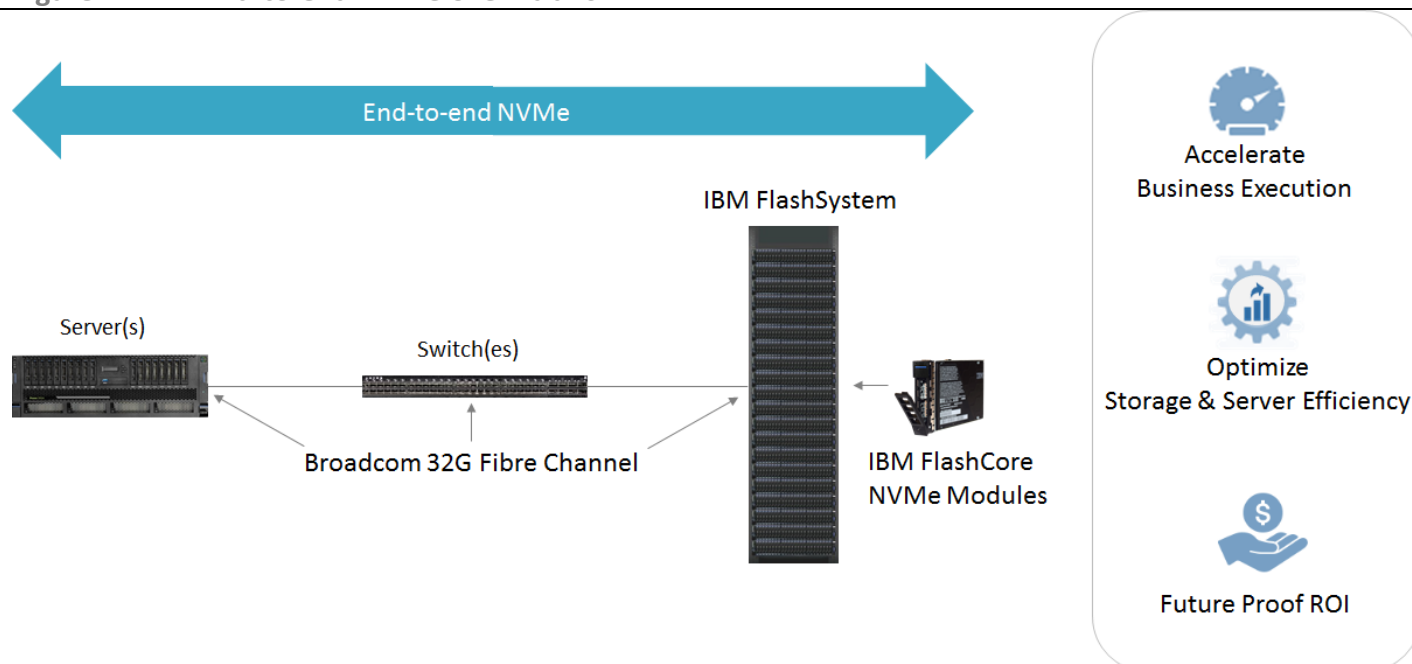
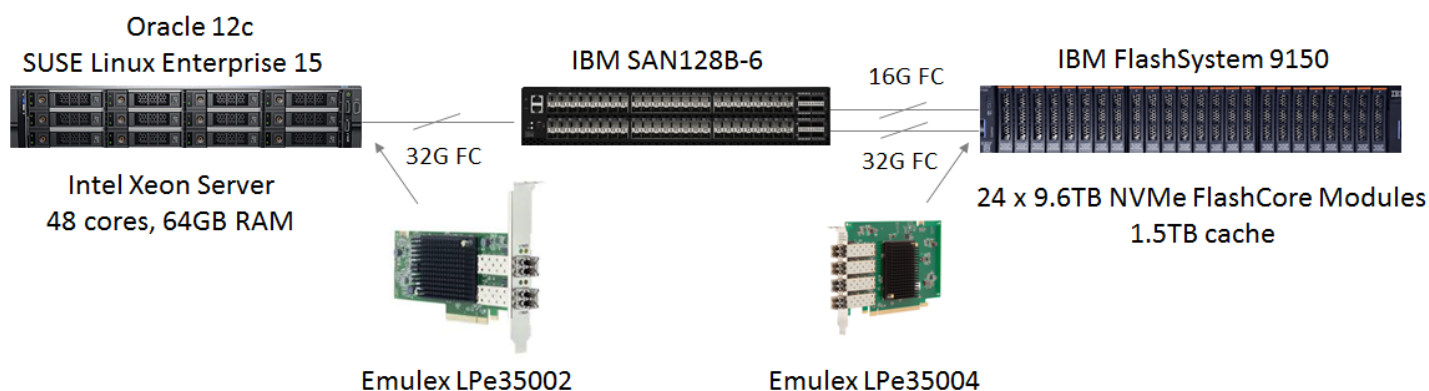*Source: Enterprise Strategy Group*

ESG Technical Validation

ESG audited the results of testing conducted at an IBM data center in the Washington Systems Center (WSC) in Herndon, VA. Typical Oracle application workloads were tested with a goal of quantifying the performance acceleration and server efficiency benefits of end-to-end NVMe powered by IBM FlashSystems with IBM b-type storage networking and Emulex Fibre Channel technology.

Oracle Database Performance Testing

Typical database workloads were tested with [HammerDB](#), an industry-standard, open source database load testing and benchmarking tool. As shown in Figure 3, Oracle 12c applications were tested on a 48-core Intel Xeon server with an Emulex LPe35002 host bus adapter that was connected through an IBM SAN128B-6 storage network switch to an IBM FlashSystem 9150.[2] The storage network was configured with two 16G and two 32G Fibre Channel connections between the FlashSystem 9150 and the storage network switch. This configuration was zoned to make it easy to compare the performance of applications that were deployed on different storage networks (e.g., the latest 32 G end-to-end NVMe over Fabric versus previous generation 16 G Fibre Channel).

Figure 3. ESG Validation Test Bed

Source: Enterprise Strategy Group

The ESG validation test bed also demonstrates the future-proof benefits of leveraging existing Fibre Channel infrastructure while enabling the performance and efficiency benefits of NVMe over fabric. Two types of NVMe fabrics are emerging: NVMe over RDMA (e.g., RoCE, iWarp, and Infiniband) and NVMe over Fibre Channel (NVMe/FC). In this case, testing was performed with a dual-protocol storage network that supports traditional FC and next-generation NVMe/FC. This dual-protocol storage network approach reduces risk in a variety of ways including:

- Uses existing Fibre Channel infrastructure (host bus adapters, switches, and storage arrays).

- Utilizes common names services, discovery, zoning, and flow control.

- Provides a cost effective and future-proof upgrade path.

- Leverages familiar tools and team expertise.

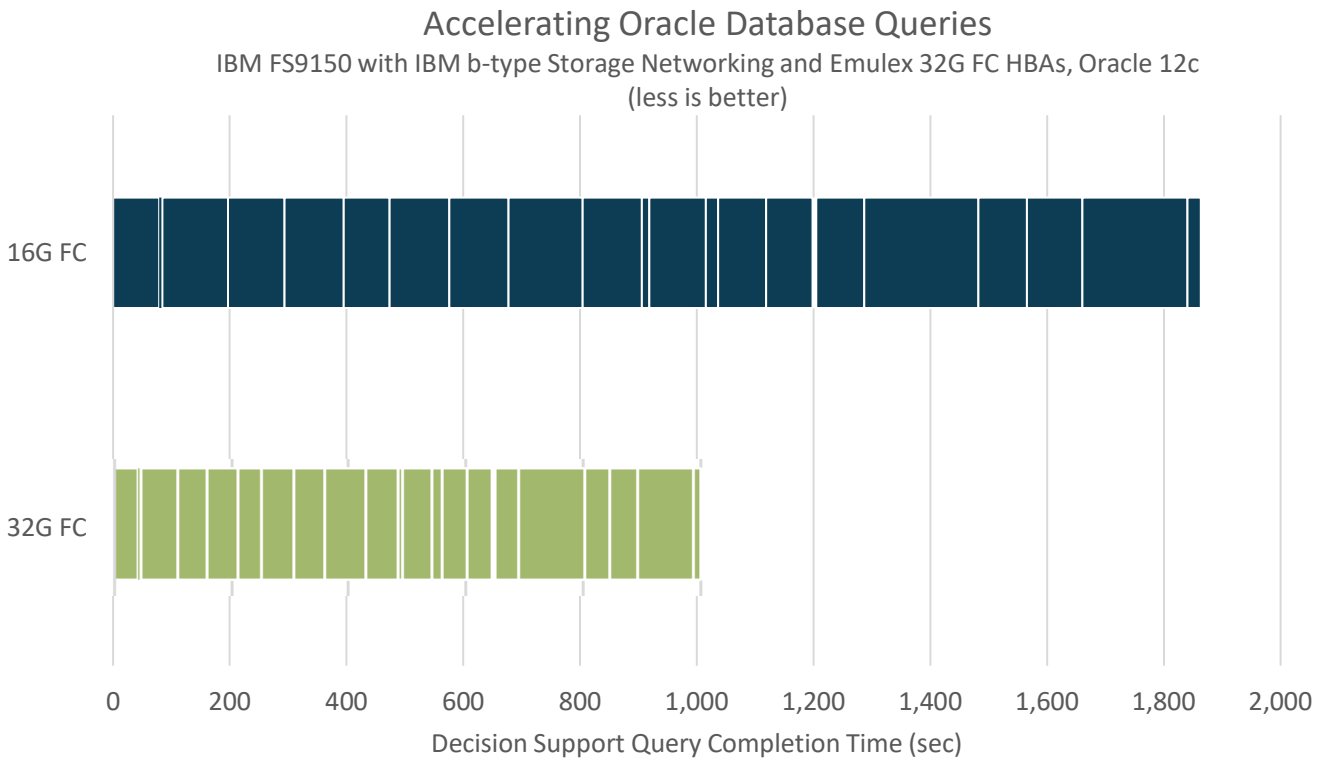[2] For more configuration details, see the Appendix.

Decision Support Systems (DSS) Performance

The ESG validation began with a series of performance benchmark tests that were designed to quantify the benefits of a 32G end-to-end NVMe storage network upgrade. Throughput-intensive application workloads that ingest and process large data sets benefit the best from a storage network bandwidth and protocol upgrade. Examples include artificial intelligence, machine learning, video editing, and analytics.

An analytics workload was tested with an Oracle 12c database configured with a 32KB block size for this throughput-intensive test.[3] The HammerDB workload generator emulated a decision support system that examines large volumes of data, executes queries with a high degree of complexity, and gives answers to critical business questions.

The results shown in Figure 4 illustrate how end-to-end 32G FC reduced the elapsed completion time for a series of 22 queries by 46%, from 1,863 seconds with 16G Fibre Channel to 999 seconds with 32G Fibre Channel.
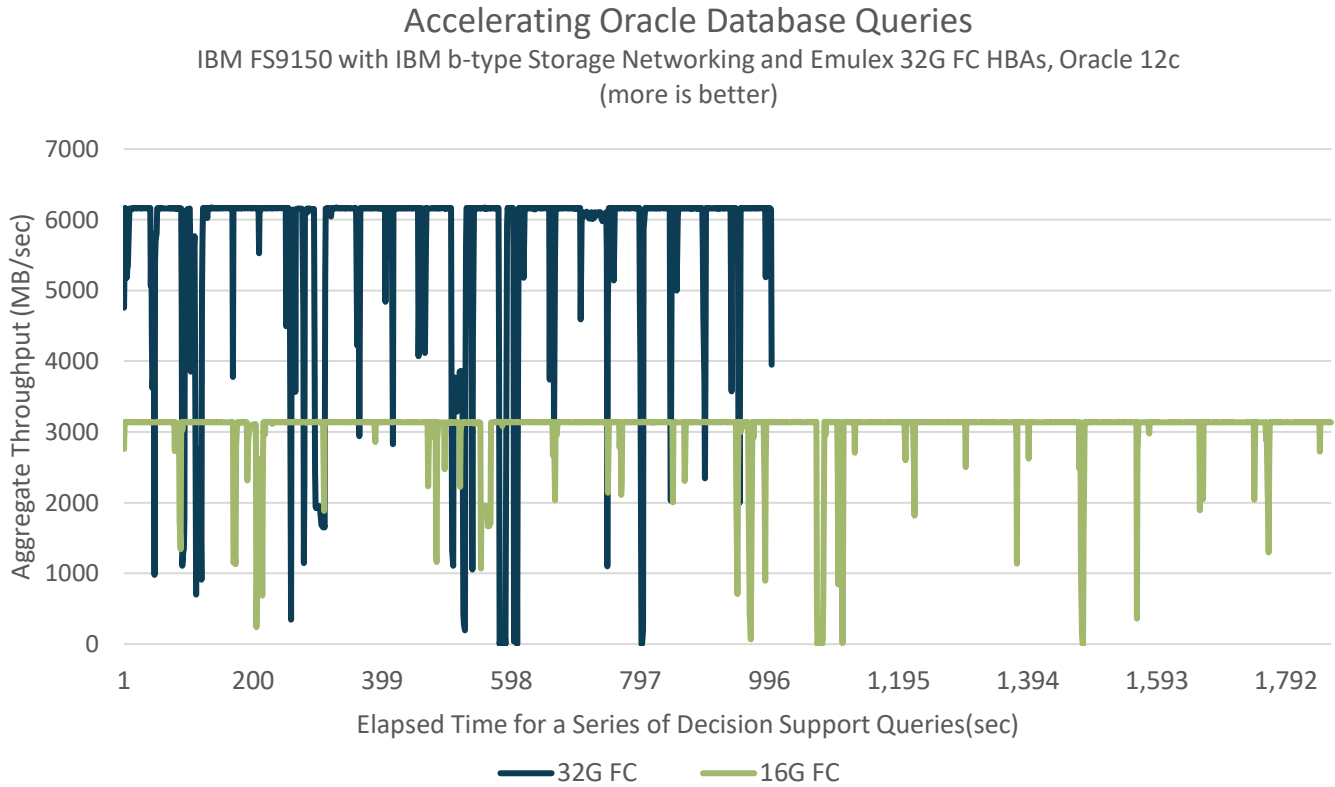
Figure 4. Oracle Query Completion Time

Accelerating Oracle Database Queries
IBM FS9150 with IBM b-type Storage Networking and Emulex 32G FC HBAs, Oracle 12c
(less is better)

Source: Enterprise Strategy Group

---

[3] For more configuration details including Oracle database settings, see the Appendix.

Figure 5 illustrates the aggregate throughput over time while the series of database query tests were running. The dips in the graph indicate where HammerDB recorded the results and restarted each of the 22 queries. The peaks illustrate how the 32G FC storage network sustained rates of 6.16 GB/sec, about twice as fast as the 16G FC network (3.14 GB/sec).

Figure 5. Oracle Query Throughput

Accelerating Oracle Database Queries

IBM FS9150 with IBM b-type Storage Networking and Emulex 32G FC HBAs, Oracle 12c

(more is better)

Source: Enterprise Strategy Group

> ## ⓘ Why This Matters
>
> Real-time analytics, AI, and ML are increasing business insight and agility and driving greater revenue and customer satisfaction. Faster analytics mean faster answers and better business decisions.
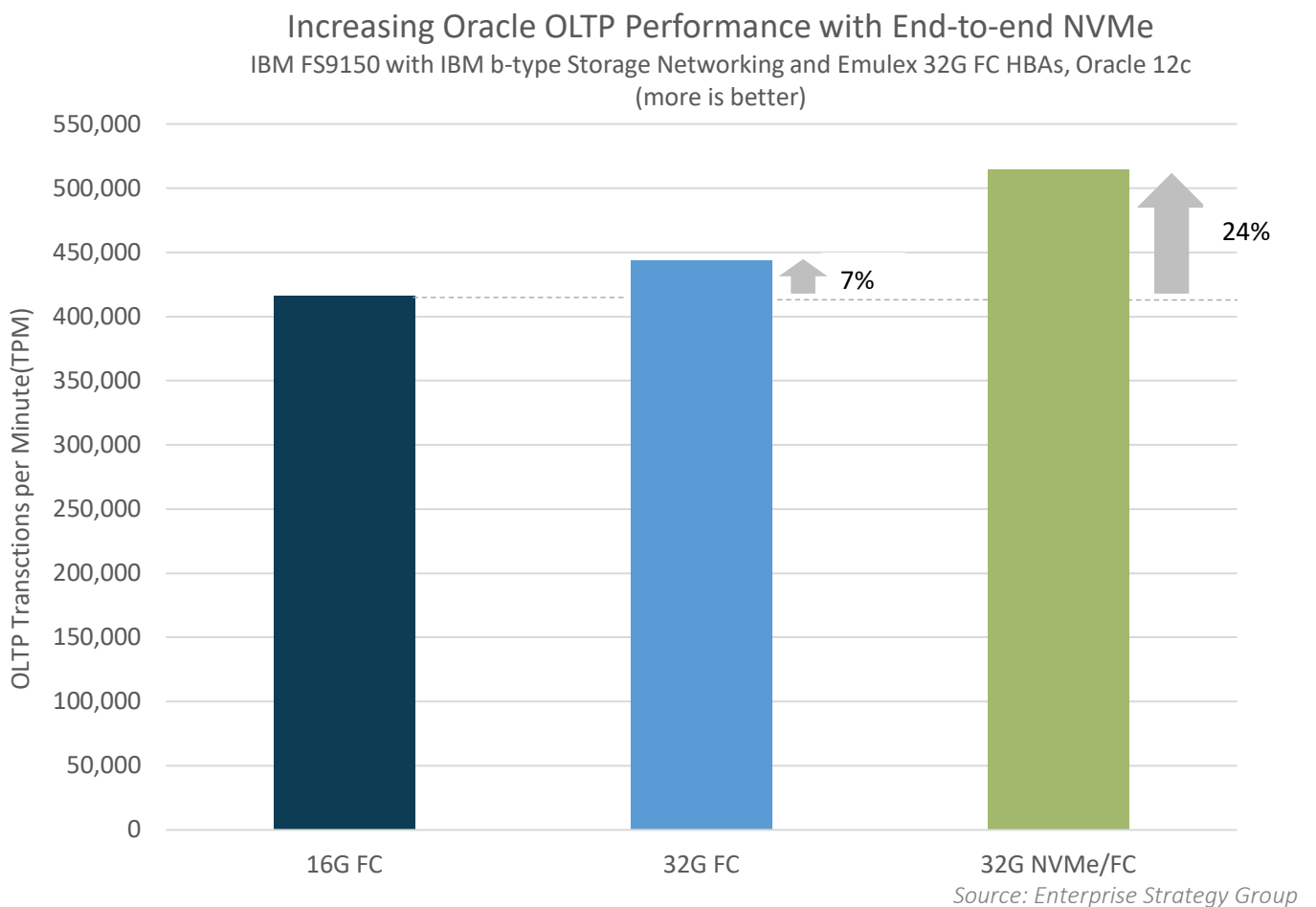>
> ESG validated that the IBM 9150 solution with the 32G IBM b-type storage networking and Emulex 32G FC technology accelerated decision support query time by 46% and increased throughput by 96%. By utilizing the full bandwidth of end-to-end 32G Fibre Channel, organizations can optimize revenue and increase customer satisfaction.

Online Transaction Processing (OLTP) Performance

The ESG validation included a series of performance tests that quantified the benefits associated with the lower latency of IBM b-type storage networking and Emulex Fibre Channel technology and the efficiency of an end-to-end NVMe over Fabric storage network. Response time-sensitive workloads benefit the most from the lower latency and improved efficiency of end-to-end NVMe. Examples of multi-user application workloads that benefit from lower I/O latency include e-commerce, software-as-a-service (SaaS), and online transaction processing (OLTP) systems.

An OLTP workload was tested with an Oracle 12c database and HammerDB. The HammerDB workload generator emulated the activity of users in a typical online brokerage firm as they generate trades, perform account inquiries, and execute market research. The Oracle 12c database was configured with an 8KB database block size.[4] The results are summarized in Figure 6 and Table 1. (Note: This test was designed with a goal of quantifying protocol performance benefits, not the maximum performance capabilities of the FS9150.)

Figure 6. Oracle OLTP Transactions Per Minute

Increasing Oracle OLTP Performance with End-to-end NVMe
IBM FS9150 with IBM b-type Storage Networking and Emulex 32G FC HBAs, Oracle 12c
(more is better)

Source: Enterprise Strategy Group

---

[4] For more configuration details including Oracle database settings, see the Appendix.

Table 1. Oracle OLTP Performance Results

Test	TPM	NOPM	IOPS
16G FC	416,032	140,167	87,802
32G FC	443,398	148,779	91,655
32G NVMe/FC	514,921	177,365	108,988

Source: Enterprise Strategy Group

What the Numbers Mean

- Transactions per minute (TPM) is the number of Oracle 12c database transactions recorded by the HammerDB utility.

- New-order transactions per minute (NOPM) is a count of the new warehouse orders that were processed as the system executed four other transactions types (payment, order-status, delivery, and stock-level).

- Higher TPM and NOPM results indicate the database infrastructure can support a larger number of concurrent users who are getting more work done and processing more orders.

- I/O statistics recorded by the Linux dstat utility (IOPS, MB/sec read, and MB/sec written) illustrate how I/O efficiency improvements translate into application-level performance benefits for response time-sensitive applications.

- Transaction results captured by the HammerDB utility and I/O statistics captured with the Linux dstat utility quantify the benefits of the improved efficiency of Emulex FC technology and end-to-end NVMe:

  o Upgrading from 16G FC to 32G FC improved TPM and NOPM by 7% and 6% respectively.

  o OLTP testing on a 32G end-to-end NVMe storage network with the exact same hardware boosted IOPS to 24%, TPM to 24%, and NOPM to 27%.

ℹ Why This Matters

For many organizations, response time-sensitive workloads such as e-commerce, SaaS, and OLTP are critical revenue drivers. The faster these processes can be executed, the greater the impact to the business. More transactions per minute lead to higher revenue and greater customer satisfaction.

ESG validated that the IBM 9150 end-to-end NVMe solution with IBM b-type storage networking delivered 24% more Oracle OLTP transactions per minute and 27% more new orders per minute.

Efficiency Analysis

The improved efficiency of IBM b-type Fibre Channel hardware and the NVMe over Fabric protocol reduces the CPU and I/O overhead of application servers. Performance statistics were captured with the Linux iostat utility during the Oracle OLTP performance test with a goal of quantifying the improved server efficiency. The results are shown in Table 2 and Figure 7.

Table 2. Oracle OLTP CPU Utilization Analysis

Test	% system	% user	% iowait	% idle	TPM per % system
16G FC	14.0%	26.3%	24.5%	35.1%	6,258
32G FC	10.8%	24.6%	25.6%	39.0%	8,487
32G NVMe/FC	5.1%	30.6%	18.1%	46.2%	21,370

Source: Enterprise Strategy Group

Figure 7. Oracle Server CPU Efficiency

Reducing Oracle CPU utilization with End-to-End NVMe
IBM FS9150 with IBM b-type Storage Networking and Emulex 32G FC HBAs, Oracle 12c
(less is better)
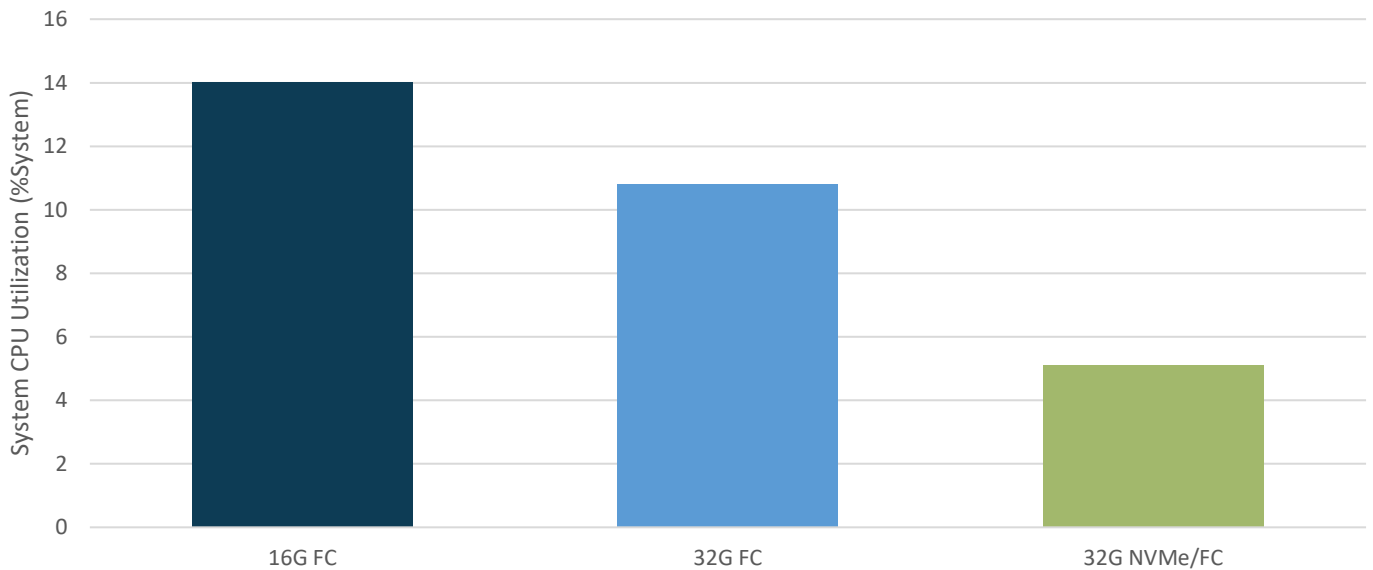
Source: Enterprise Strategy Group

What the Numbers Mean

- The kernel level system CPU utilization (% system) statistics quantify the increased hardware efficiency of IBM b-type Fibre Channel technology (23% improvement, from 14% with 16G FC to 10.8% with 32G FC).

- The % system CPU utilization benefits are magnified dramatically due to the lower latency and increased efficiency of the NVMe over Fabric protocol (64% improvement, from 14% with 16G FC to 5.1% with 32G NVMe/FC).

- Kernel CPU utilization improvements (% system) result in more CPU for applications (% user), less time waiting for I/Os to complete (% iowait), and less CPU idle time (% idle).

- Correlating the application-level transactions per minute with kernel CPU utilization provides a useful measure of the I/O CPU efficiency benefits of end-to-end NVMe for Oracle applications, in this case up to 3.4X more TPM per % system.

- Kernel CPU utilization reduction is only one of the potential efficiency benefits of an IBM end-to-end NVMe storage infrastructure. Additional cost-saving and return-on-investment benefits include:

    o   Leveraging existing Fibre Channel infrastructure, processes, tools, expertise, and vendor relationships.

    o   Reducing capital equipment and licensing costs using a future-proof upgrade path from Fibre Channel to FC over NVMe instead of building a new Ethernet storage network with NVMe over Fabric (e.g., RoCE).

    o   Increasing the consolidation and utilization of storage and server infrastructure.

    o   Lowering the costs of power, cooling, and floor space in the data center.

    o   Reducing the number of server CPU cores to meet the needs of business-critical applications and the software licensing costs that are tied to CPU core counts (e.g., Oracle database licenses).

ℹ️ Why This Matters

Organizations could get the fastest performance by running their applications in memory, but the cost of that would be prohibitive. Therefore, IT is continually trying to balance performance with cost and efficiency, striving to service more users and add more business value with less infrastructure and cost.

ESG validated that the IBM 9150 NVMe-oF solution with IBM b-type storage networking required up to 64% less kernel-level I/O CPU utilization with up to 26% less I/O wait time; we also validated up to 3.4X more I/O CPU efficiency for Oracle databases.

If the infrastructure associated with business-critical Oracle workloads is struggling to meet the performance needs of the business, an end-to-end NVMe strategy with IBM can provide significant CapEx and OpEx savings and immediate ROI. Leveraging existing FC components and networks makes NVMe/FC an immediately viable alternative for many organizations, enabling them to increase performance without ripping and replacing infrastructure.

Auto Connect and Asymmetric Namespace Access (ANA)

The Fibre Channel over NVMe specification includes powerful new capabilities compared with a traditional Fibre Channel storage network:

- **NVMe Auto Connect eliminates the need to run a host rescan utility to find new devices.** ESG confirmed this by loading the native NVMe fabric driver after IBM FlashSystem devices had been configured. As expected, the IBM FlashSystem subsystem (nvme list-subsys) and its block devices (lsblk) appeared automatically after the driver was loaded.

- **Asymmetric Namespace Access (ANA) adds multipath load balancing and failover into native NVMe drivers.** ESG testing with a heavy Oracle OLTP workload and the native NVMe driver built into SUSE Linux E15 confirmed that failover and failback is built in and automated (see Figure 8).

Figure 8. ANA Multipath Failover
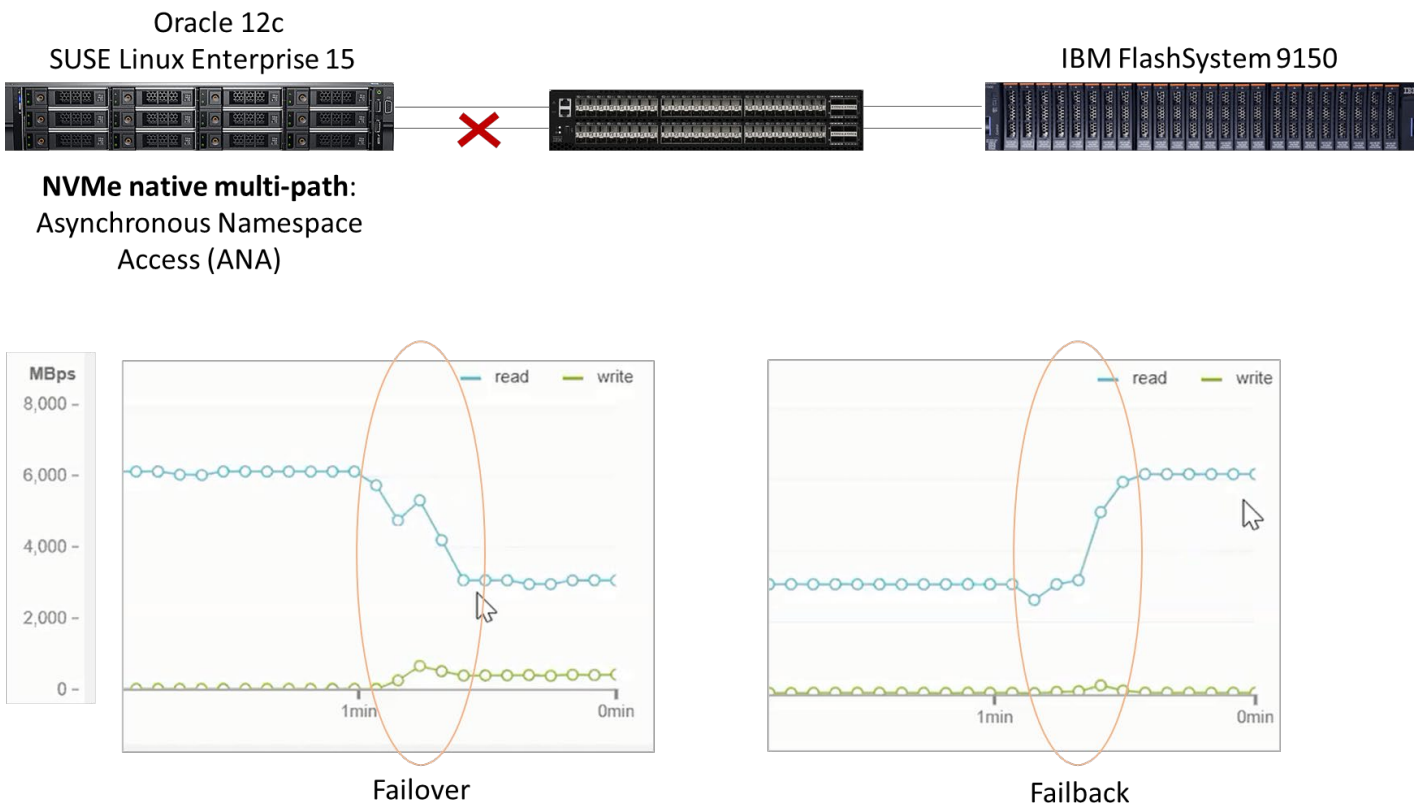
Source: Enterprise Strategy Group

ℹ **Why This Matters**

Business-driving applications demand high availability and high performance. Managing and maintaining infrastructure is costly and can interfere with business processes. A key business goal is for IT staff to spend less time keeping their applications running without interruption. ESG validated that the Auto Connect and ANA features enabled multipath failover with the native NVMe driver in SLES-15 and IBM end-to-end NVMe. This supports high application availability without administrator intervention.

The Bigger Truth

Improving drive performance makes a significant difference in application performance. But improvements to only one component of the data path often reveal bottlenecks in other areas. Moving from disk to flash provided about a 10X performance improvement, and the advent of NVMe drives increased performance again—but then host connectivity became a bottleneck. The data path prevented applications from achieving the full performance potential of NVMe drives.

NVMe-oF provides *end-to-end* NVMe that brings another 10X improvement. This speeds connectivity from the host through the switch to the flash array, enabling another order of magnitude performance improvement.

This solution—the IBM FlashSystem with IBM b-type 32G storage networking and Emulex 32G FC technology— demonstrates numerous benefits of end-to-end NVMe. Higher performance is the most obvious, but in addition it enables workload consolidation, which leads to lower costs. The NVMe protocol is more efficient than traditional SCSI, and NVMe-oF allows NVMe native devices to operate on existing Fibre Channel infrastructure, making it simple and cost-effective to add into a current FC environment. And, by lowering the CPU utilization on the host, it can significantly reduce the processing and memory needs of servers, freeing up resources to run applications, and saving on power, cooling, and licensing costs for applications such as Oracle.

ESG validated that an IBM 9150 end-to-end 32G NVMe solution with IBM b-type storage networking:

- Reduced Oracle DSS query time by 46% and increases throughput by 96%.

- Delivered 24% more Oracle OLTP transactions per minute and 27% more new orders per minute.

- Improved Oracle server I/O CPU utilization by 64% with 3.4X more I/O CPU efficiency.

- Works with the native NVMe ANA multipath and failover support that's built into SUSE Linux E15.

Competition drives enterprises today to ever-higher levels of performance for databases, analytics, and transaction-oriented workloads. But IT must always make a tradeoff between performance and cost, even with technology innovations like NVMe. The advantage of the solution we tested is that these tradeoffs are diminished. This solution delivers high performance and low latency with frictionless data movement from the host through the network, all the way to back-end storage, but it also provides for greater efficiency—and lower costs—at both the storage and server CPU level. While native NVMe drivers are currently supported on only two flavors of Linux, support for additional operating systems and hypervisors is on the horizon. If your organization is looking to increase performance while reining in costs with a future-proof NVMe infrastructure, the IBM FlashSystem with IBM b-type storage networking and Emulex 32G FC technology is worth a hard look.

Appendix

Configuration Details

Component	Configuration Details
NVMe storage system	IBM FlashSystem 9150, 24X9.6TB NVMe FlashCore Modules, 1.5TB cache Code 8.3.0.0, DRAID6 with a stripe width of 16, 1 spare, 16x100 GB LUNs, Emulex LPe35004
Storage switch	IBM b-type SAN128B-6
Host bus adapters	Emulex LPe35002 FW: 12.2.299.23 Emulex LPFC OOB Driver version 12.2.299.13
Server	Dual socket Intel Xeon Gold 6146, 48 cores, 64GB RAM, 18GB swap, Power mode: performance, Boot mode: BIOS
Operating system	SUSE Linux Enterprise Service 15 (x86_64), kernel 4.12.14-345.ge903367-default
NVMe CLI	Version 1 (updated for auto connect)
Workload generator	HammerDB version 3.0
Database	Oracle 12c, XFS file system, asynchronous and direct file system I/O type, Dedicated server connection type
DSS DB config	300GB dataset size, 32KB database block size, 24GB SGA, 6GB PGA
OLTP DB config	500GB dataset size, 8KB database block size, 50GB SGA, 6GB PGA

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

© 2019 by The Enterprise Strategy Group, Inc. All Rights Reserved.