STORAGE DEVELOPER CONFERENCE
SDC 21
BY Developers FOR Developers
Virtual Conference
September 28-29, 2021
A SNIA Event

Scalable Storage Performance for High Density Applications

Naveen Krishnamurthy
Jayamohan Kallickal

Agenda

- NVMe and NVMe-oF – Brief Introduction
- NVMe specification updates
- Business Cases for NVMe/NVMe-oF
- Potential Architectural References
- NVMe and Hybrid Cloud
- Performance Results

NVMe-Intro

NVMe Technology Is Everywhere:

Solid State Disks

Disaggregated Storage Architectures

Cell Phones

Servers

Mobile Computing

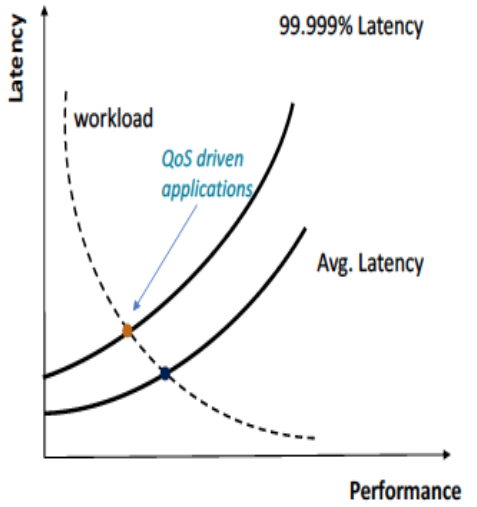
Storage Arrays

Data Centers

NVMe® Technology Differentiation

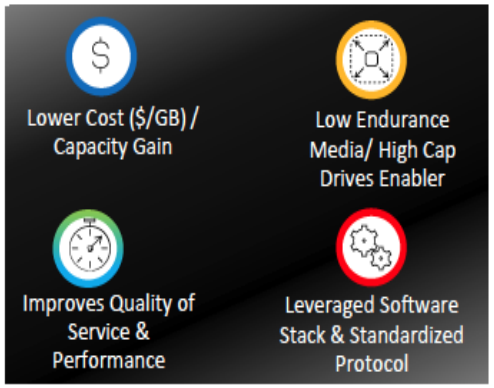
Business benefits

Low Latency



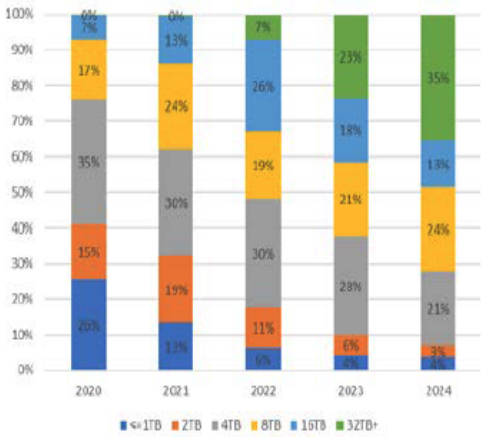
- Traditional applications: average latency of SSDs is a performance criterion for making purchase decisions
- QoS-driven applications (e.g. Web apps): Five 9's latency is critical

Zone Named Spaces



- ZNS reduces write amplification and enables QLC adoption
- Eliminates GC, reduces long tail latencies/ consistency
- 50% (or higher) average latency improvement

High-Density



- eSSD Capacities moving up to 4TB and above, expected increases with QLC NAND adoption
- Expanding use cases to include cooler storage

Flash Memory Summit

nvm EXPRESS®

NVMe® over Fabrics Specification

Business benefits

LOW LATENCY

NVMe-oF™ technology can deliver latencies on par with NVMe® SSDs inside servers

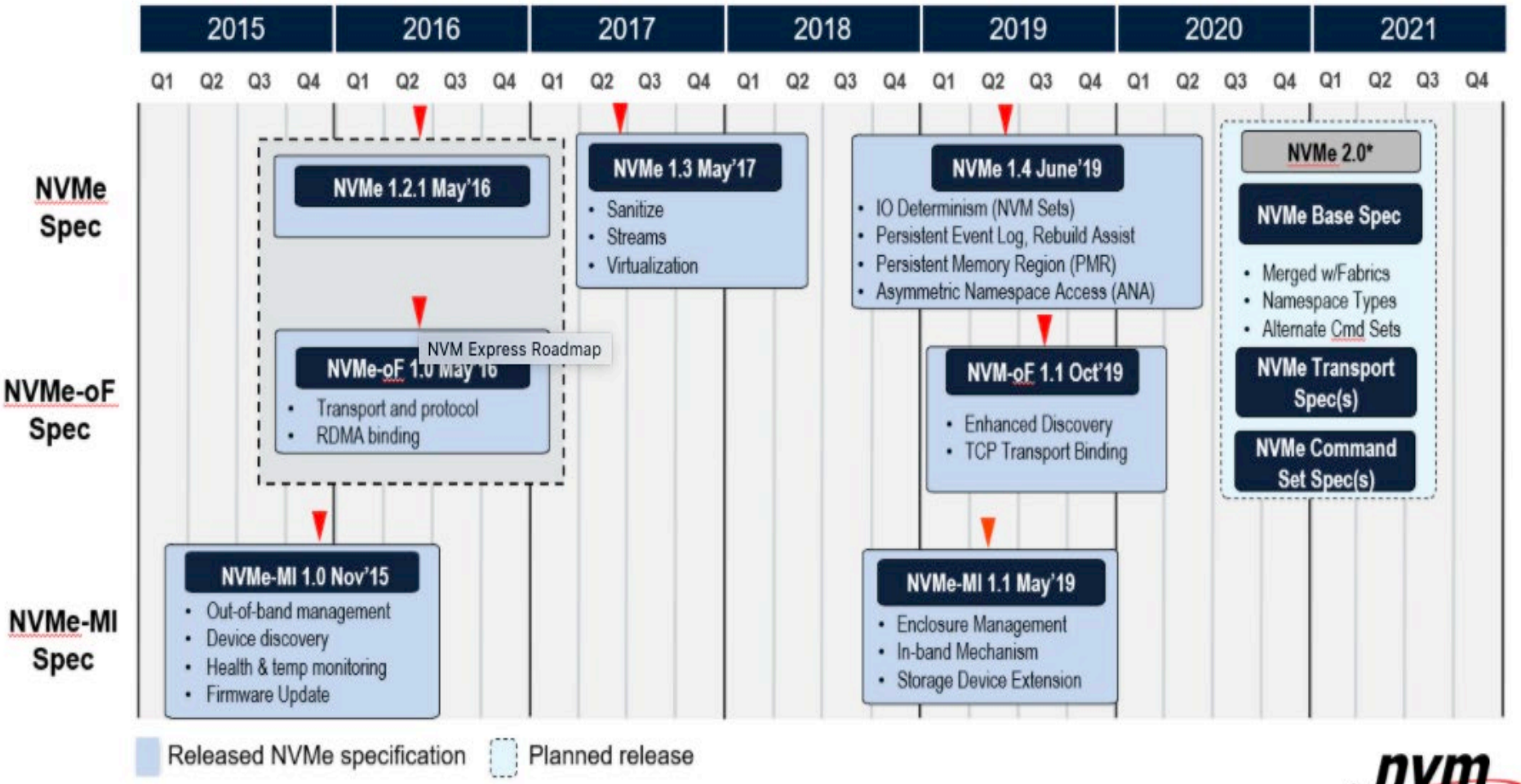
HIGH-PERFORMANCE SHARING

NVMe-oF attached SSDs can be shared amongst many of application servers resulting in higher utilization and lower TCO

DATA ACCESS & MOBILITY

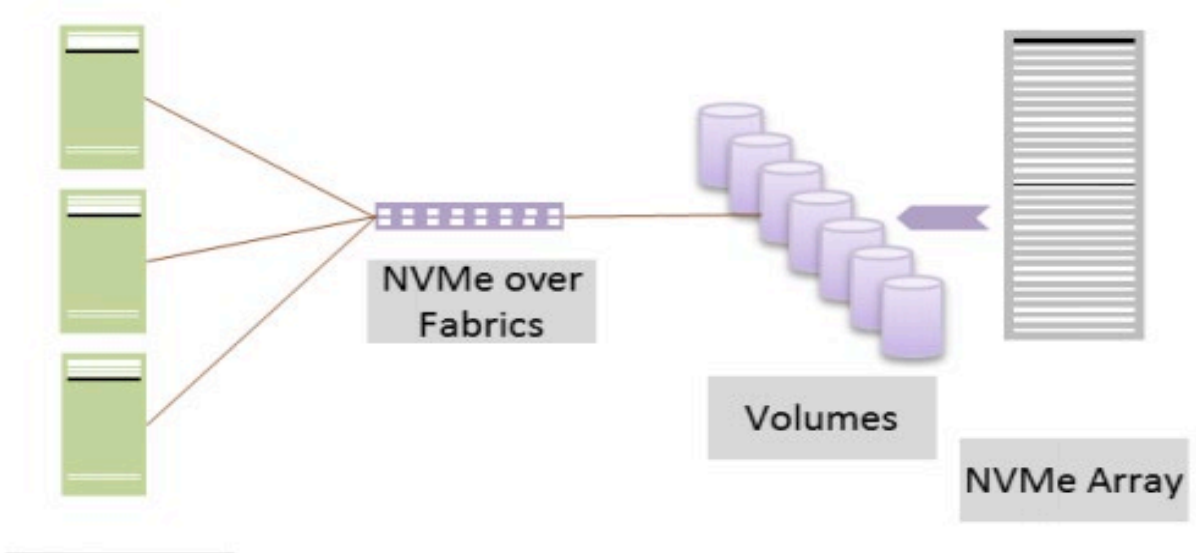
Fabric-attached data enables Cloud-like dynamic access and workload mobility

NVM Express Technology Specification Roadmap



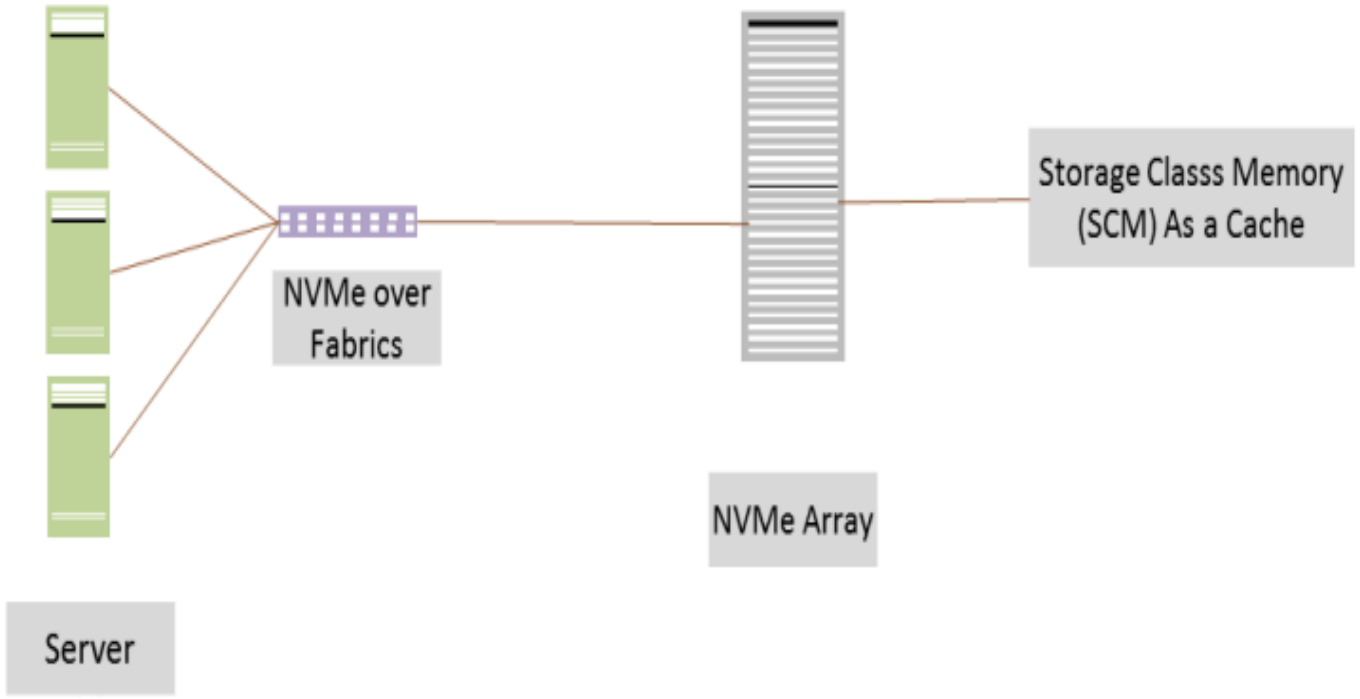
Business Cases

Use Case 1: Faster Flash



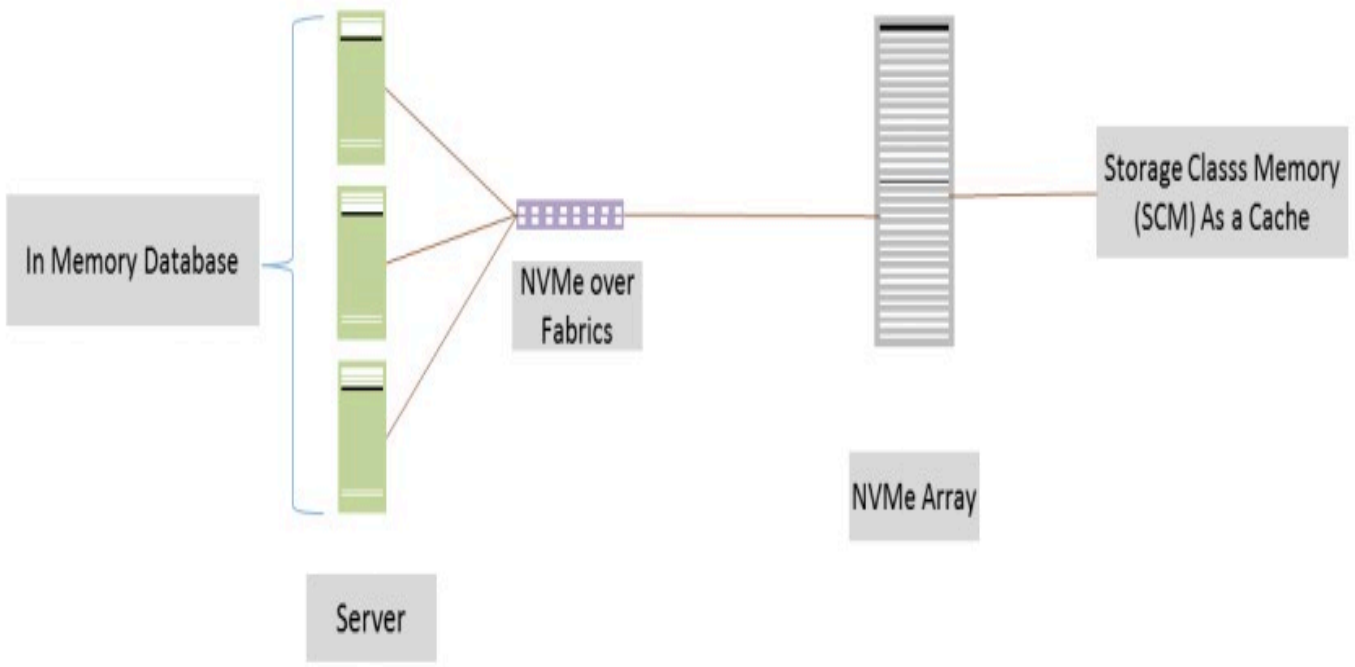
- Achieving lower latency with NVMe over fabrics compared with legacy SCSI
- Very low server CPU utilization
- Scalable to 100’s of drives thus making higher capacity available to applications o 25GE/32GFC 100GE/128GFC bandwidth to support 32G (PCIe) and faster NVMe drives

Use Case 2: Storage Class Memory as a Storage array Cache



- Improved performance:
- Benefit from low latency media
- Caching/fast storage removes PCIe latency
- Benefits from improved performance, higher bandwidth and lower latency

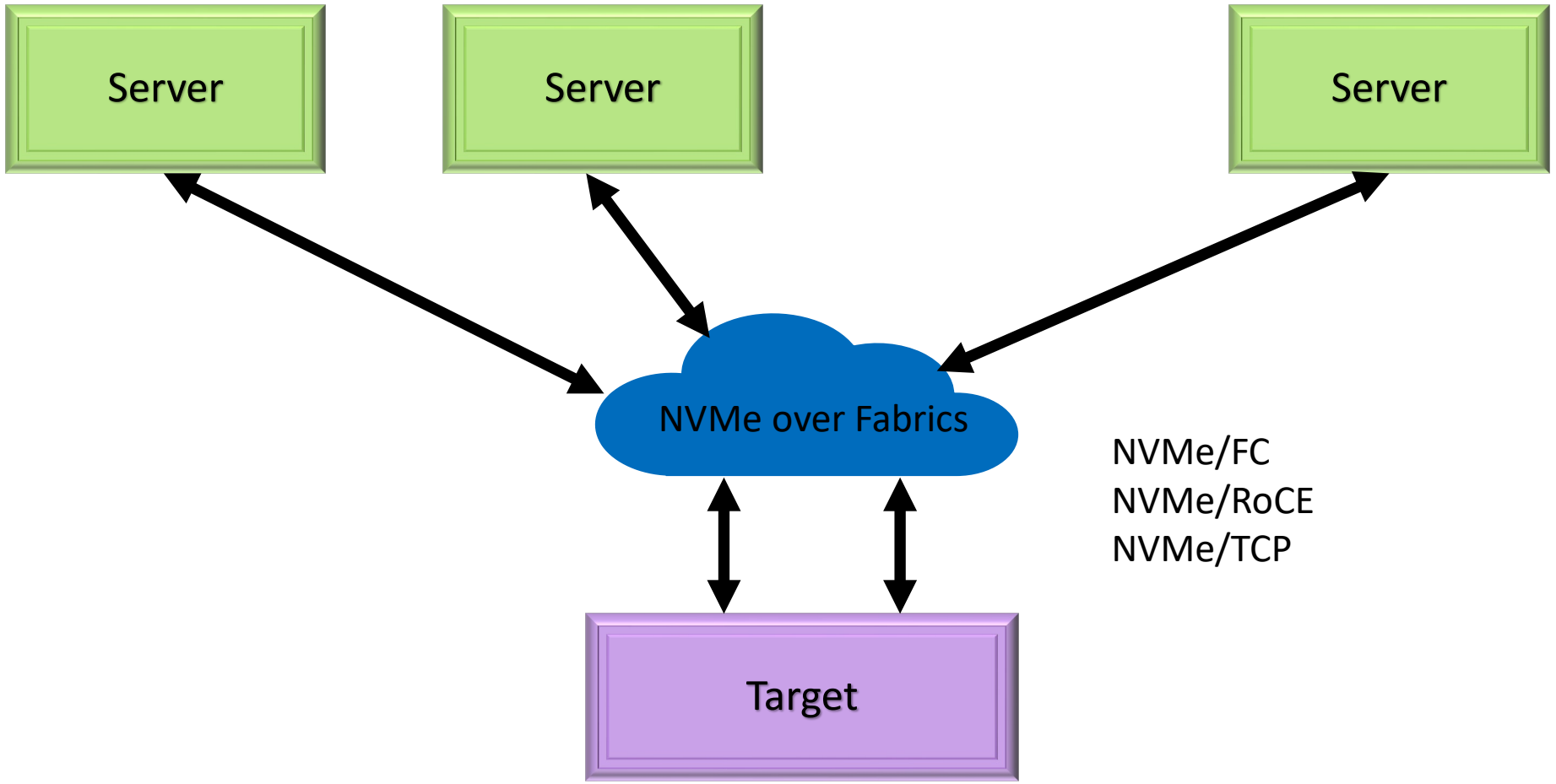
Use Case 3: In-memory databases for Data Management



- New Storage Tier: In-Memory Databases
- Eliminates datacenter siloes
- Eliminates stranded storage
- Enables
  - Snapshots
  - Data Tiering (HANA)
  - Data Availability
  - Workload Migration

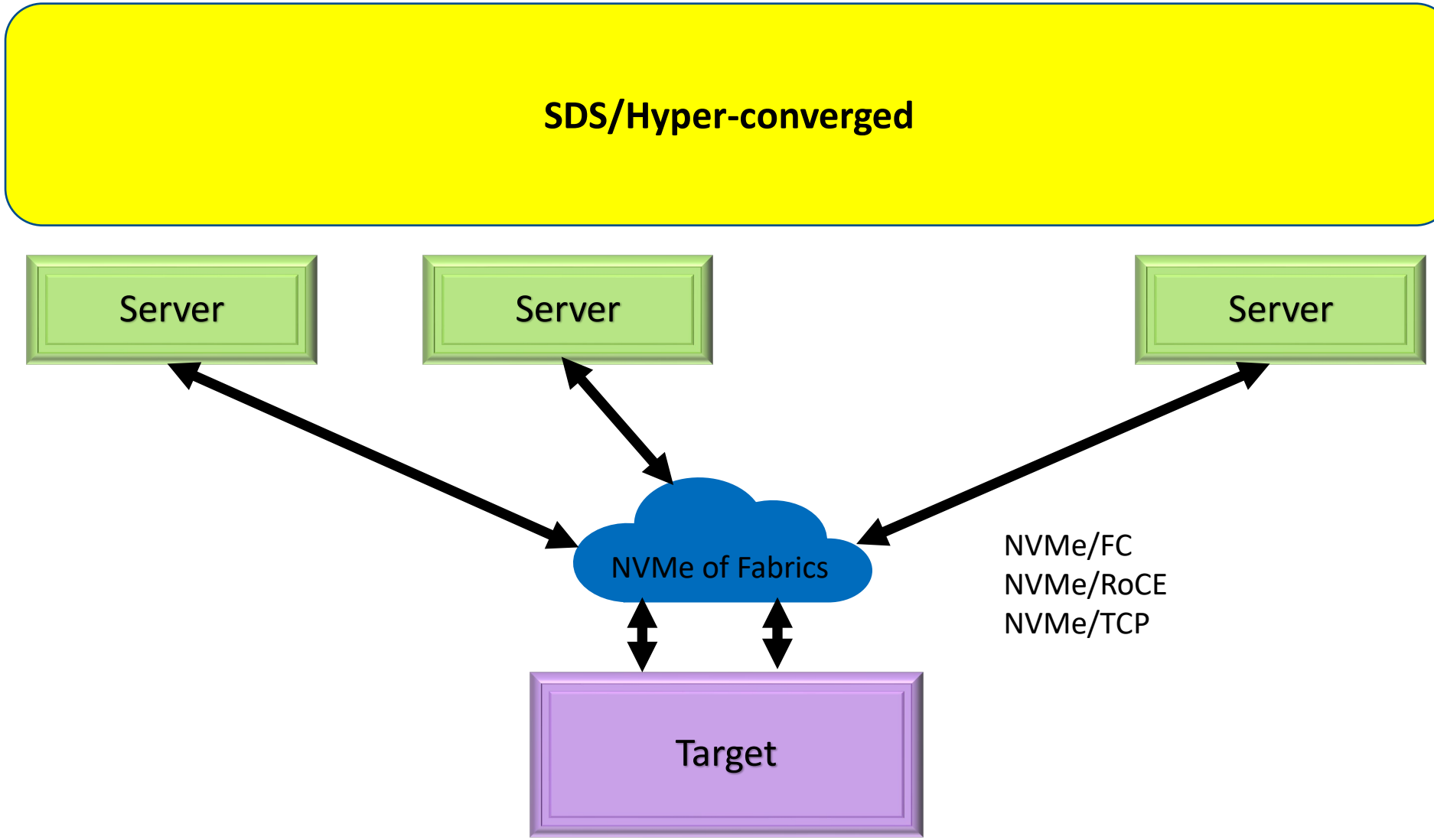
Potential Architectures for NVMe-oF Usage

SAN



NVMe/FC
NVMe/RoCE
NVMe/TCP

SDS/Hyper-converged

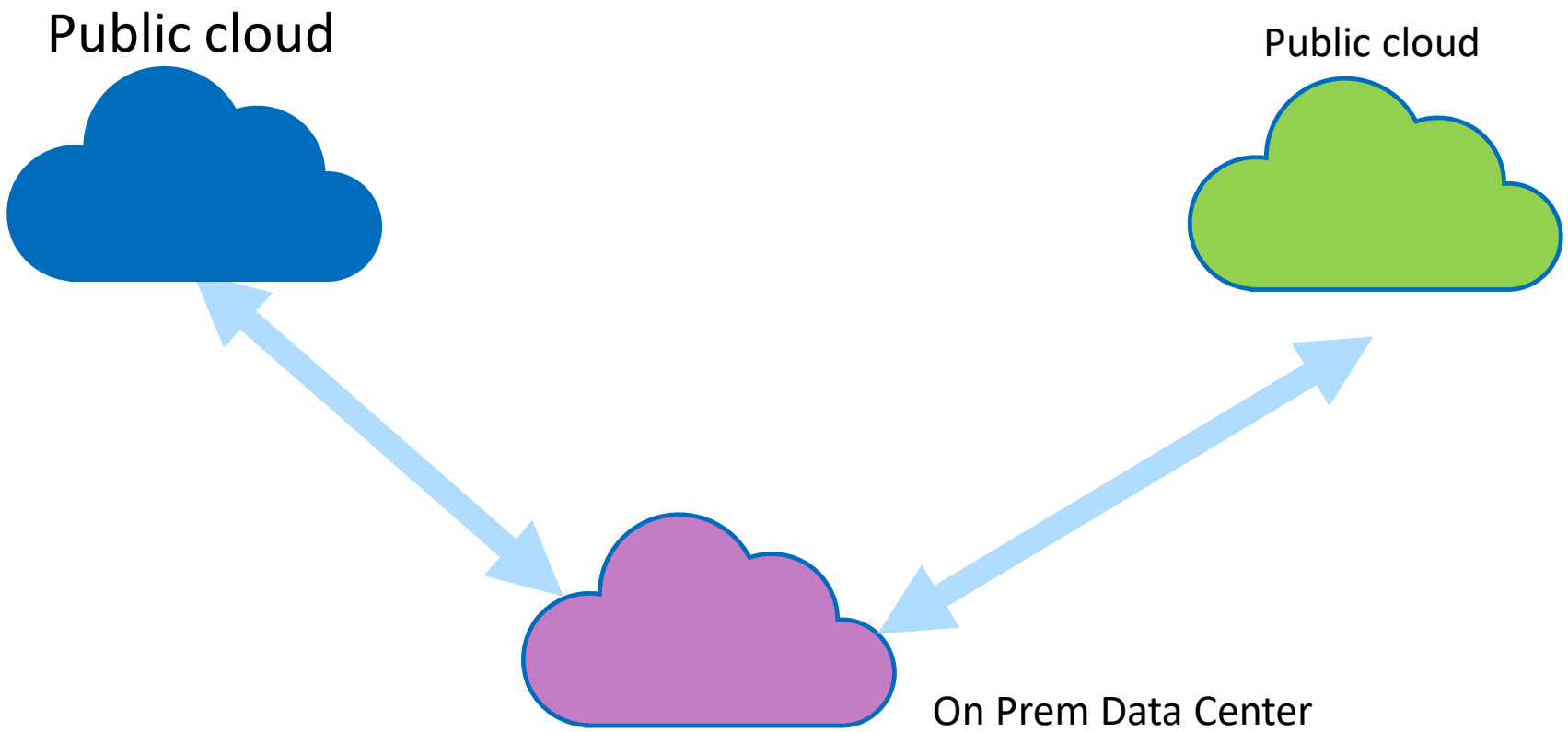


Why Hybrid

- The cost of Cloud, aTrillion Dollar paradox
  - Sarah Wong and Martin Casado, A16Z
  - *You're crazy if you don't start in the cloud; you're crazy if you stay on it.*
  - 50 Public SW companies lose 100B in Market cap due to cloud
    - Broader Public Companies market cap is suppressed by >500B
    - Dropbox saved about $75M over two years
    - Dropbox increased gross margins from 33% to 67% between 2015 to 2017
  - *If you're operating at scale, the cost of cloud can at least double your infrastructure bill.*

- **The honeymoon is over for cloud computing**
  - **Aug 12, 2021, Business Insider**

Hybrid Cloud



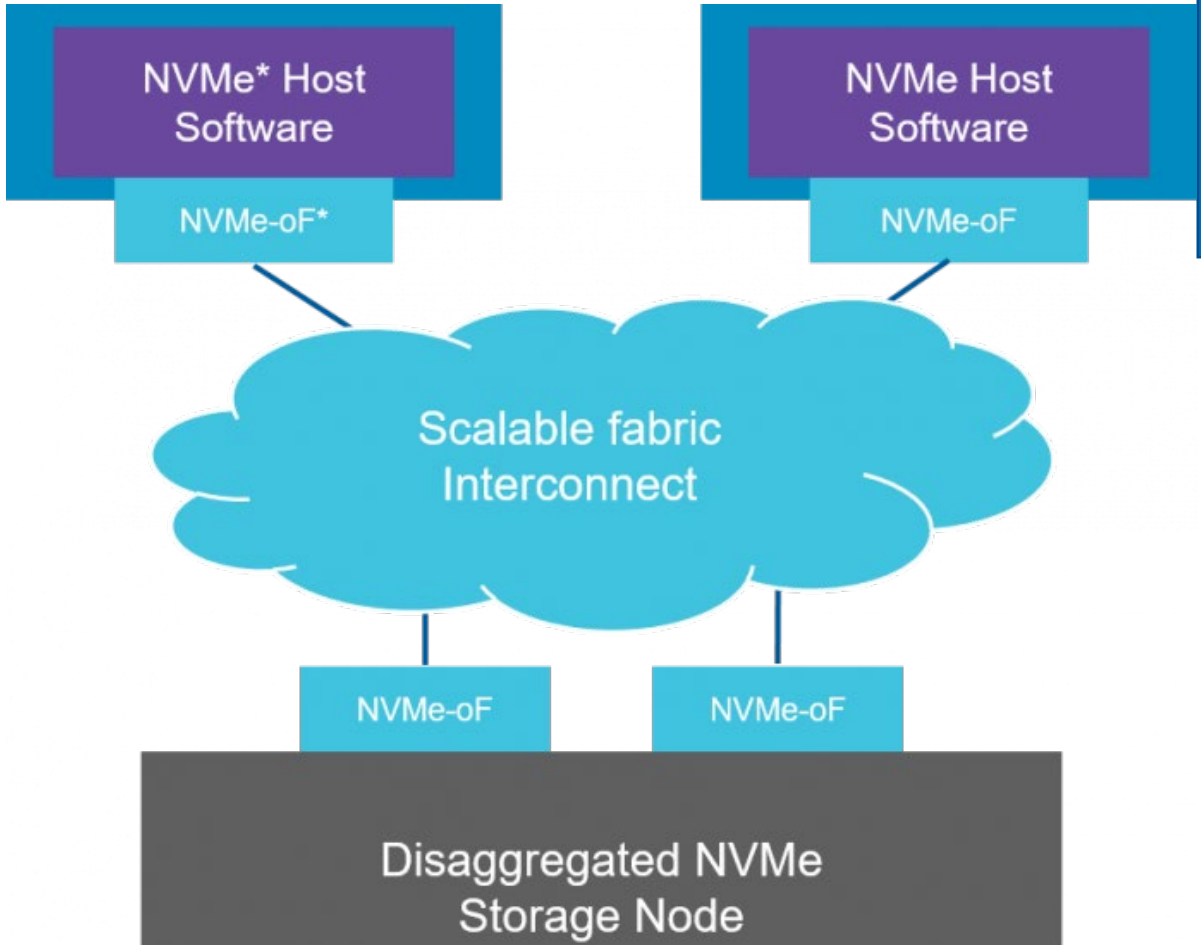
Hybrid Cloud

- Not everything belongs in a public cloud, which is why so many forward-thinking companies are choosing a hybrid mixture of cloud services.
- Hybrid clouds offer the benefits of both public and private clouds and take advantage of existing architecture in a datacenter.
- A hybrid cloud architecture includes these characteristics:
  - Your on-premises data center, private and public cloud resources, and workloads are tied together under common data management while staying distinct.
  - You can connect existing systems running on traditional architectures that run business-critical applications or contain sensitive data that might not be suited for the public cloud.

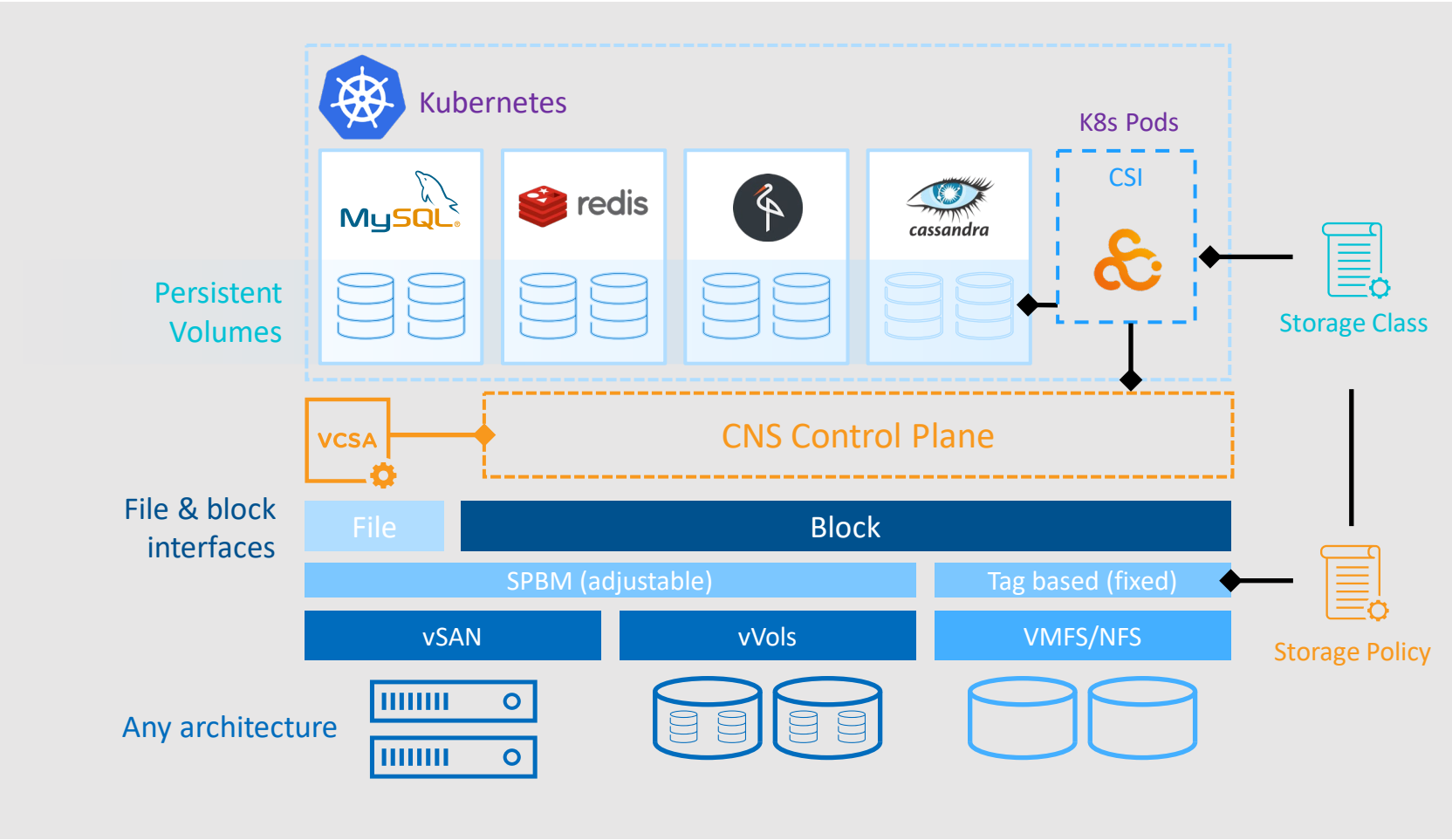
NVMe-oF as a Cloud Storage Fabric

- One of the most important factors for any cloud storage service, is the speed of data access, which depends heavily on the performance of the backend storage devices.

- The following image shows a modern cloud deployment model, which typically includes a set of compute nodes, which are disaggregated, as well as some disaggregated NVMe storage nodes. We can dynamically create nodes out of compute and storage resources, using these fabric interconnects.



VMware Cloud Native Storage



- Based on CSI standard for container storage
- Offers block and file-based Persistent Volumes
- De-facto storage platform for vSphere with Tanzu
- Storage intensive apps benefit from NVMe/NVMe-oF

NVMe-TCP for Edge Storage

- A recent International Data Corporation (IDC) Data Age 2025 report predicts that by 2025 6 billion consumers interact with data every day, or 75% of the world's population.
  - This will skyrocket demands for storage at the edge. The best way to handle this massive data overload is with solutions that separate storage – which is growing exponentially – from compute, which is not growing as fast. The fastest and most efficient way to make that possible leverages NVMe-oF
- With the standardization of NVMe-TCP, new production-grade NVMe/TCP storage solutions are providing the highest performance without any constraints or requirements to the client-side infrastructure.
  - NVMe-TCP solutions are especially important for edge deployments, where adding network constraints can be impossible. Thanks to the protocol, new storage solutions are helping organizations disaggregate storage over any IP network, cluster several proximate edge locations into a high-availability (HA) storage pools, have stateless edge instances seamlessly utilizing storage at the aggregation layer.

Performance

vSphere NVMe-oF Solution Performance

BROADCOM

Tolly. Tolly Test Report, April 2020

NVMe-FC vs SCSI-FC

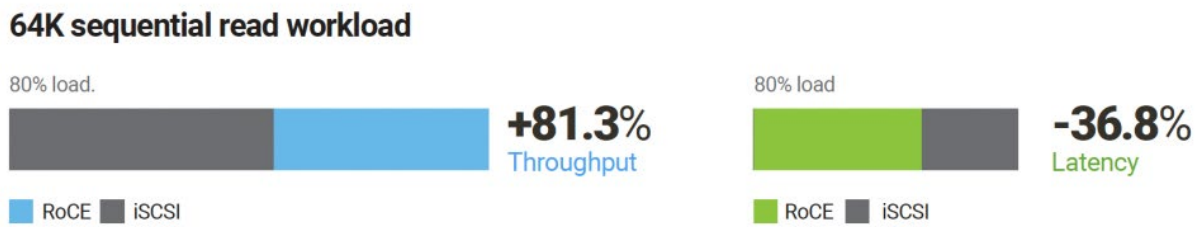
- **2.4x** higher transactions compared to SCSI for MS SQL Server 2017
- **2.1x** higher transactions compared to SCSI for Oracle 19c

Emulex Gen 7 LPe35002 32GFC HBAs running NVMe/FC against the performance of the same adapters running in SCSI mode on NetApp AFF 800. Brocade G620 32GFC Switch.

NVIDIA

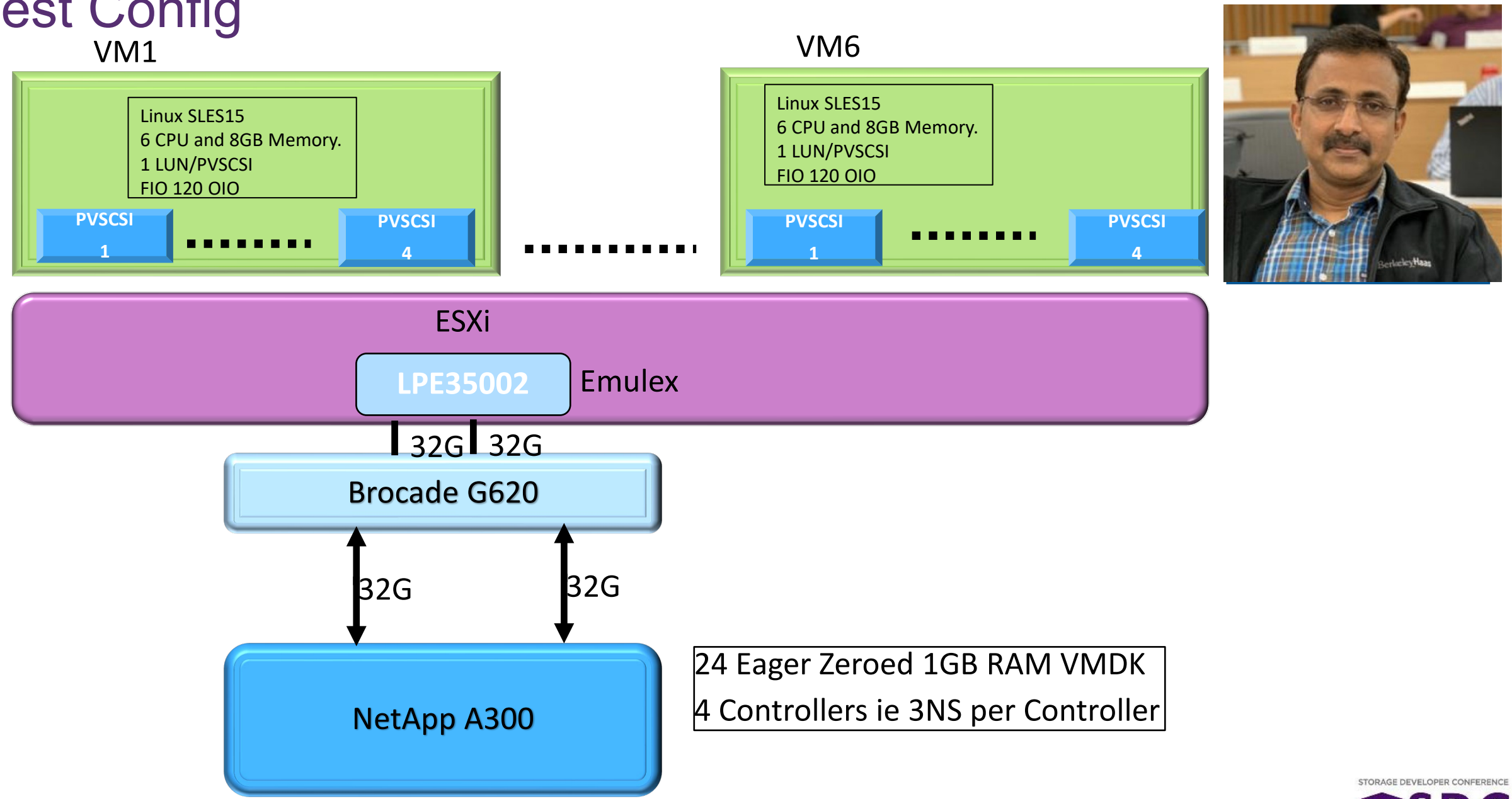
StorageReview Enterprise Lab, August 2020

NVMe-RDMA vs iSCSI

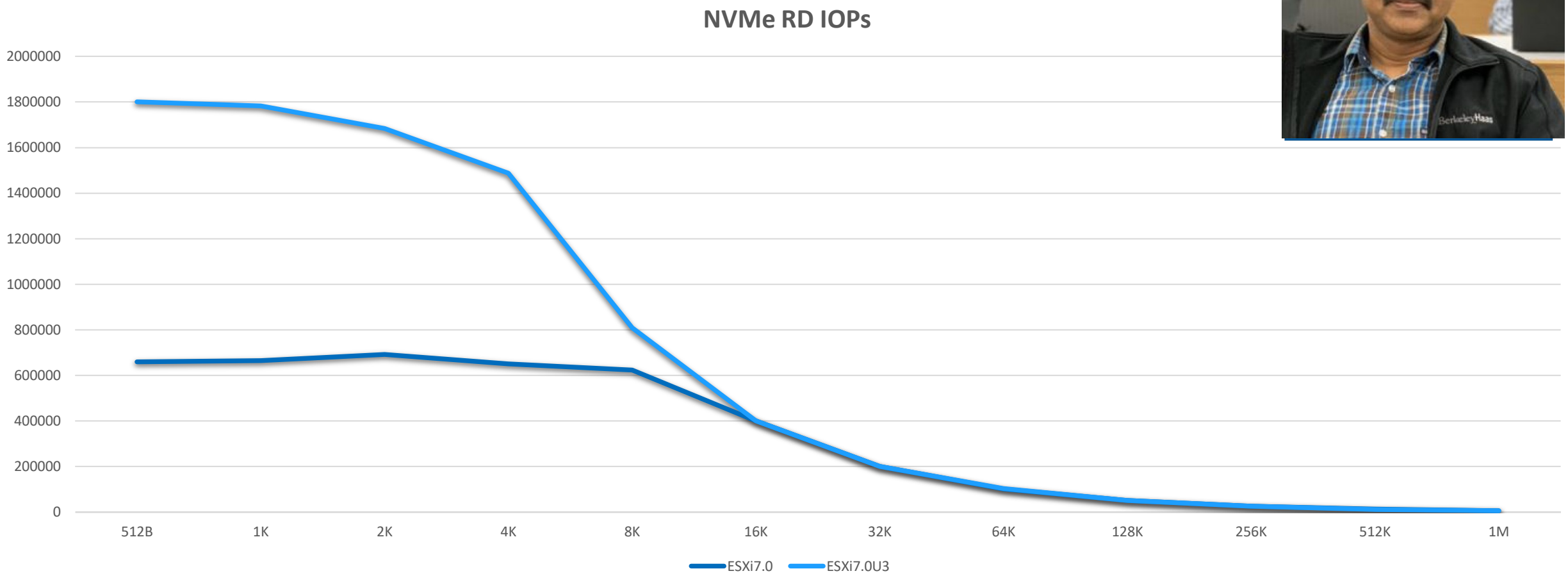


PCIe Gen 3.0 ConnectX-5 cards at 25GbE in a pair of Dell R720 servers over a network connected via a pair of NVIDIA's Spectrum SN2010 switches to the Pure Storage Flash Array.

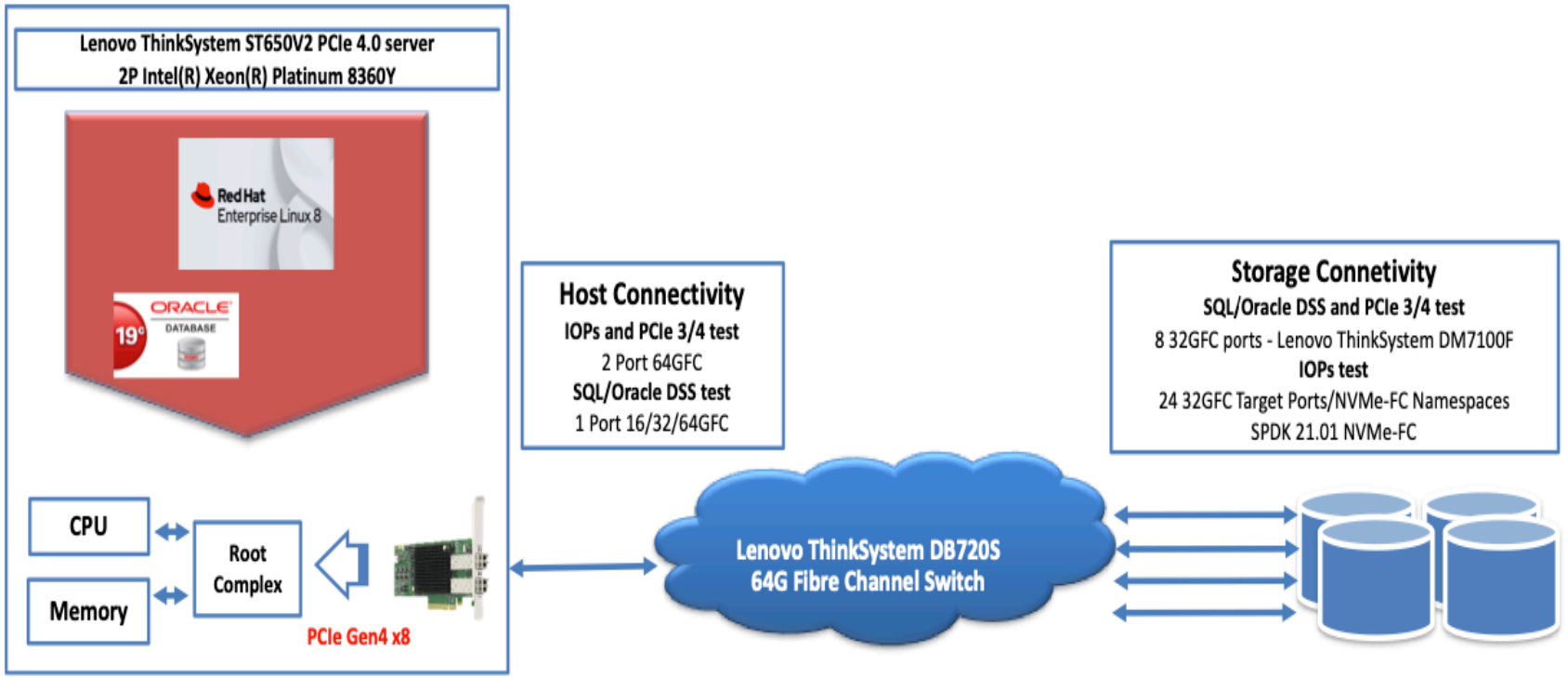
Test Config



Improvement- RD IOPs



Composite Test Bed Topology

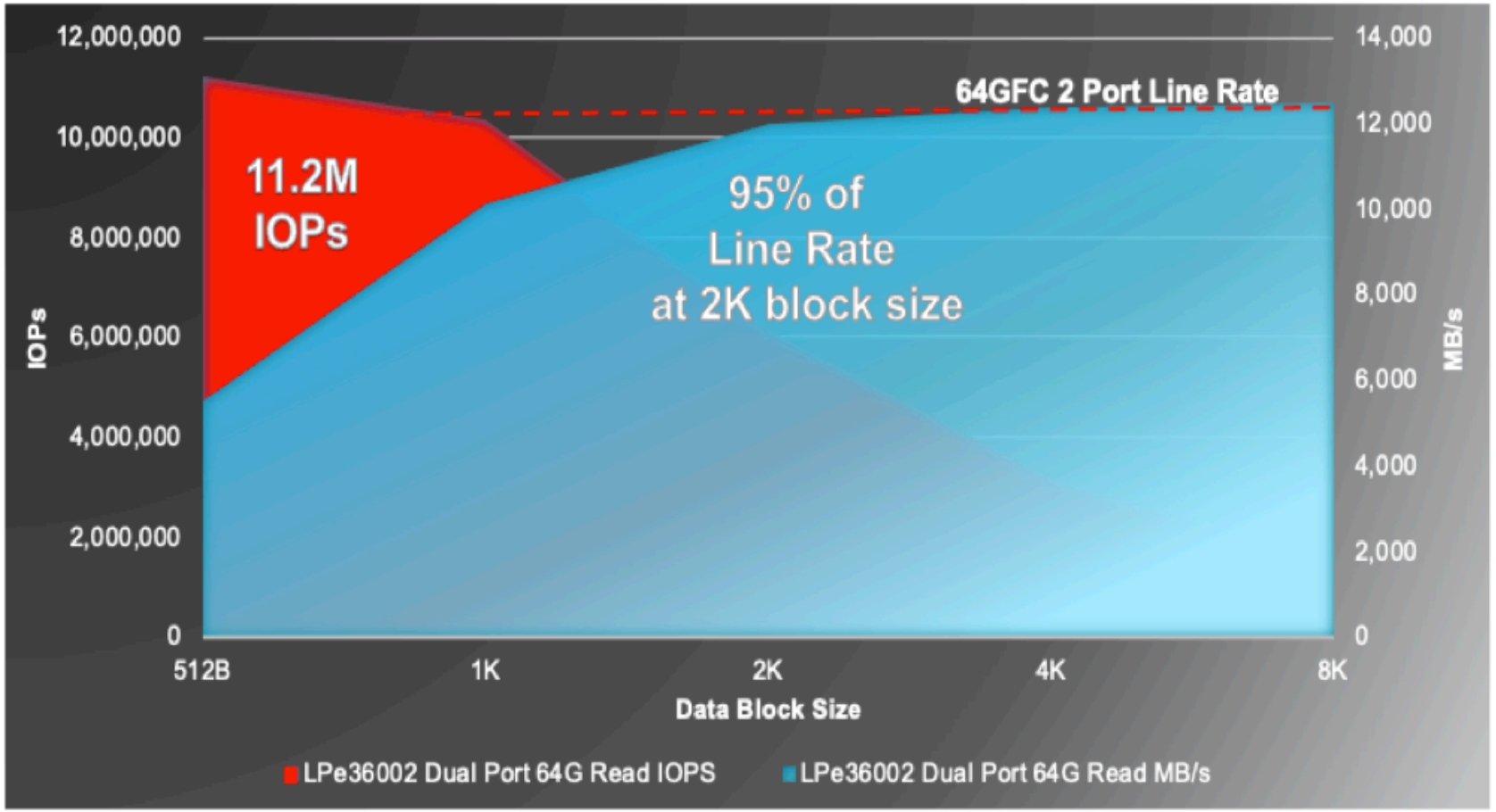


Source: Tolly, April 2021

Figure 4

ThinkSystem LPe36002 64G HBA Dual-Port Aggregate IOPS & Throughput

(as reported by Medusa Labs Test Tool Suite v7.4)



Notes: Lenovo ThinkSystem ST650V2 with PCIe 4.0 running Red Hat Linux Server 8.3 to 24 open source NVMe-FC SPDK targets.

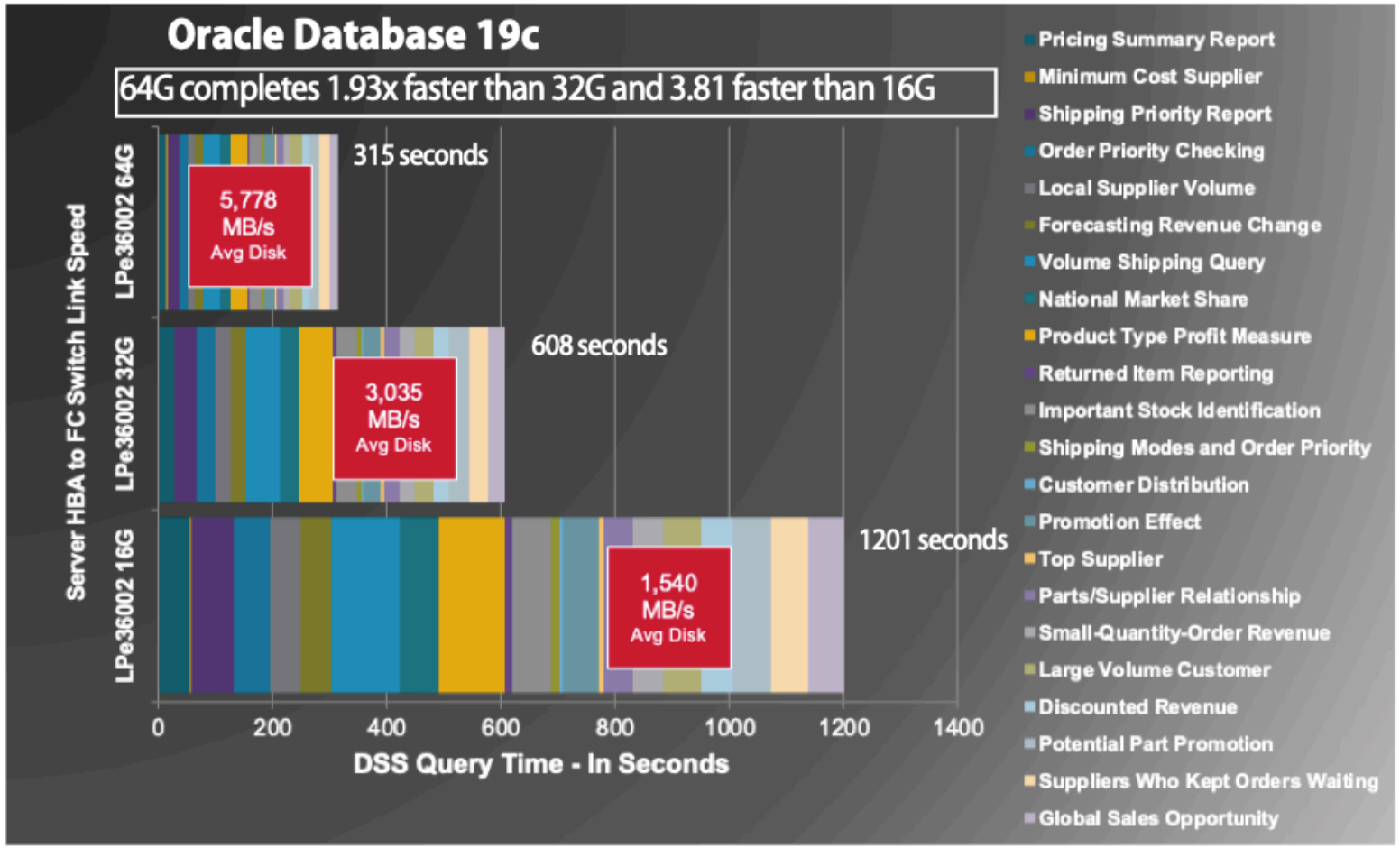
Source: Tolly, April 2021

Figure 1

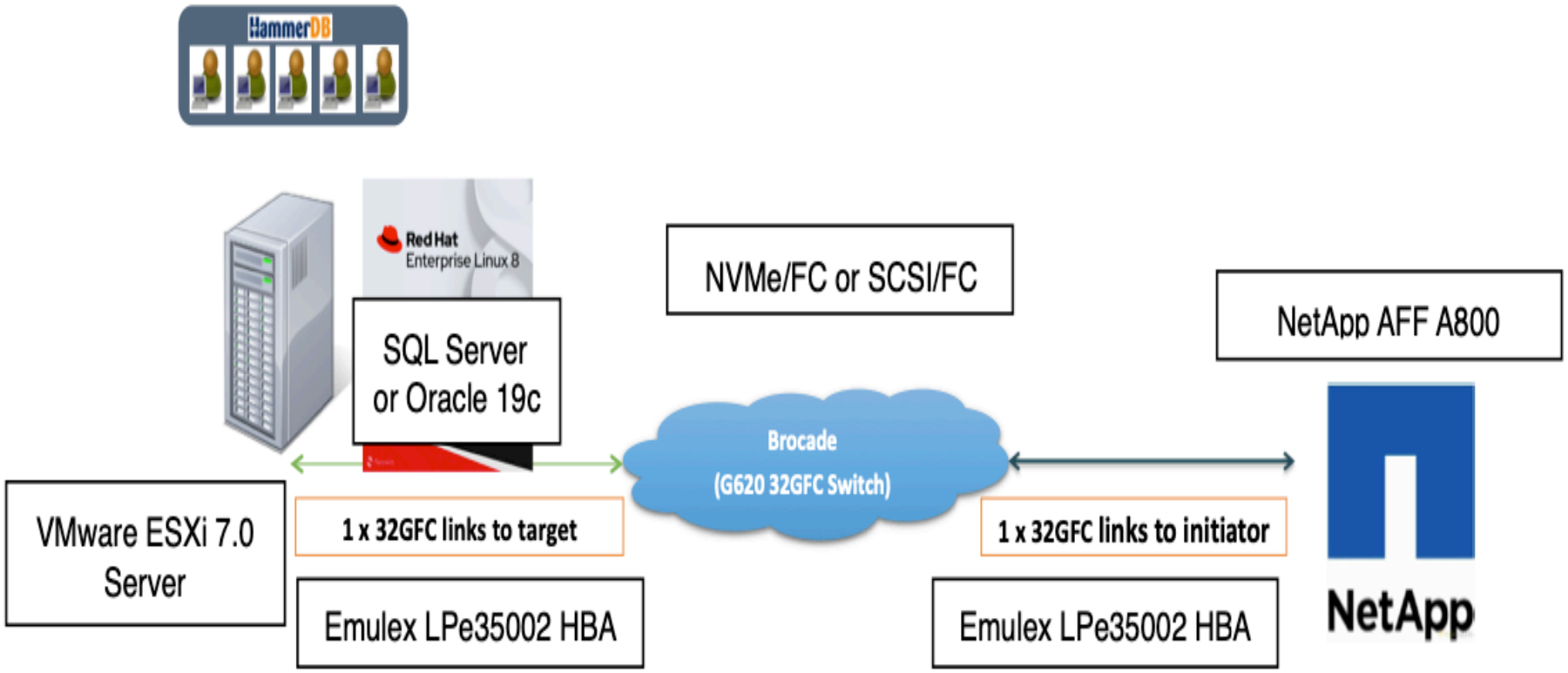
ThinkSystem LPe36002 64G HBA DSS Warehousing Performance: 64G vs. 32G vs. 16G

(as reported by HammerDB v3.3)

(Shorter runtime and higher throughput better)



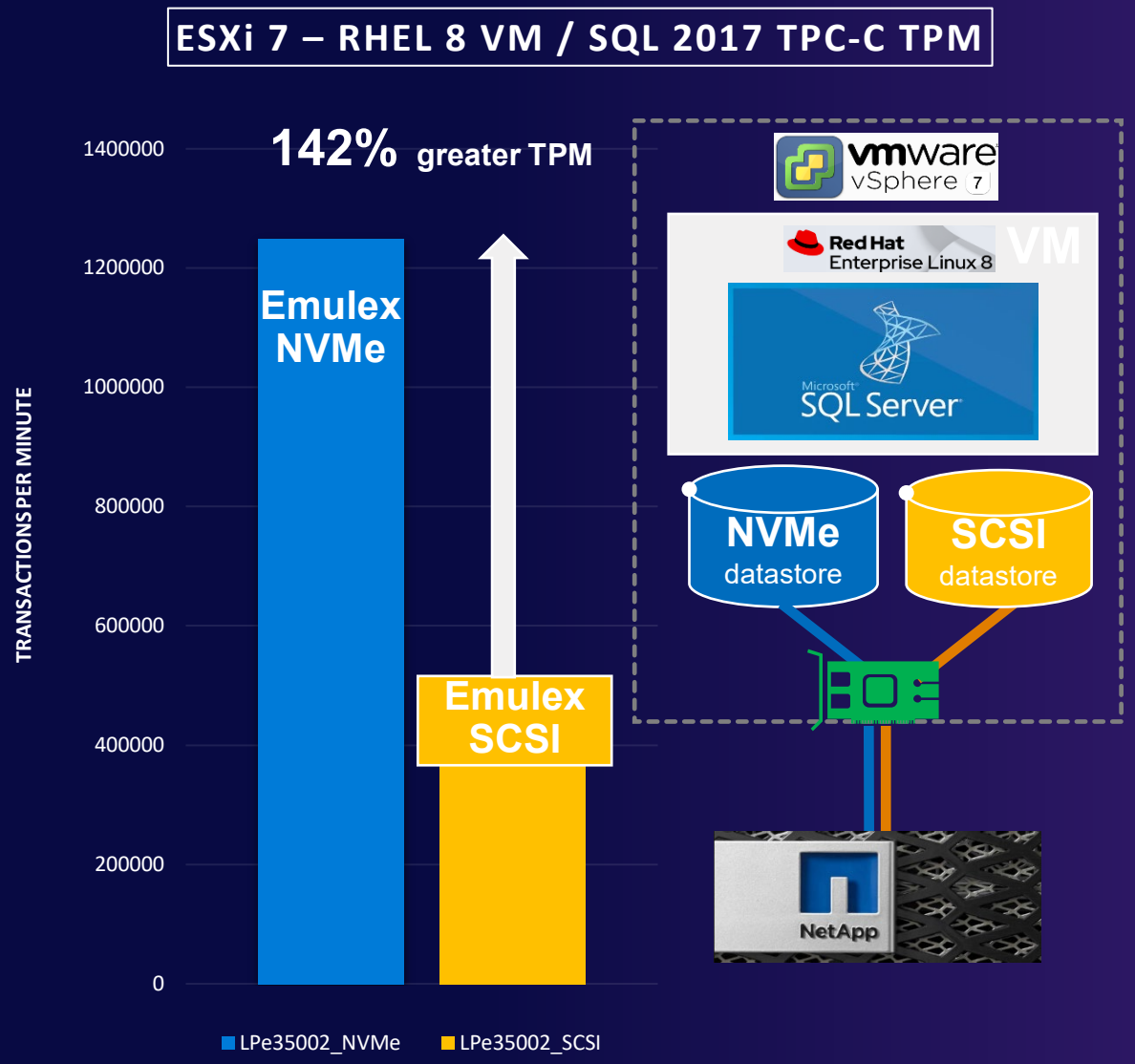
Composite Test Bed Topology

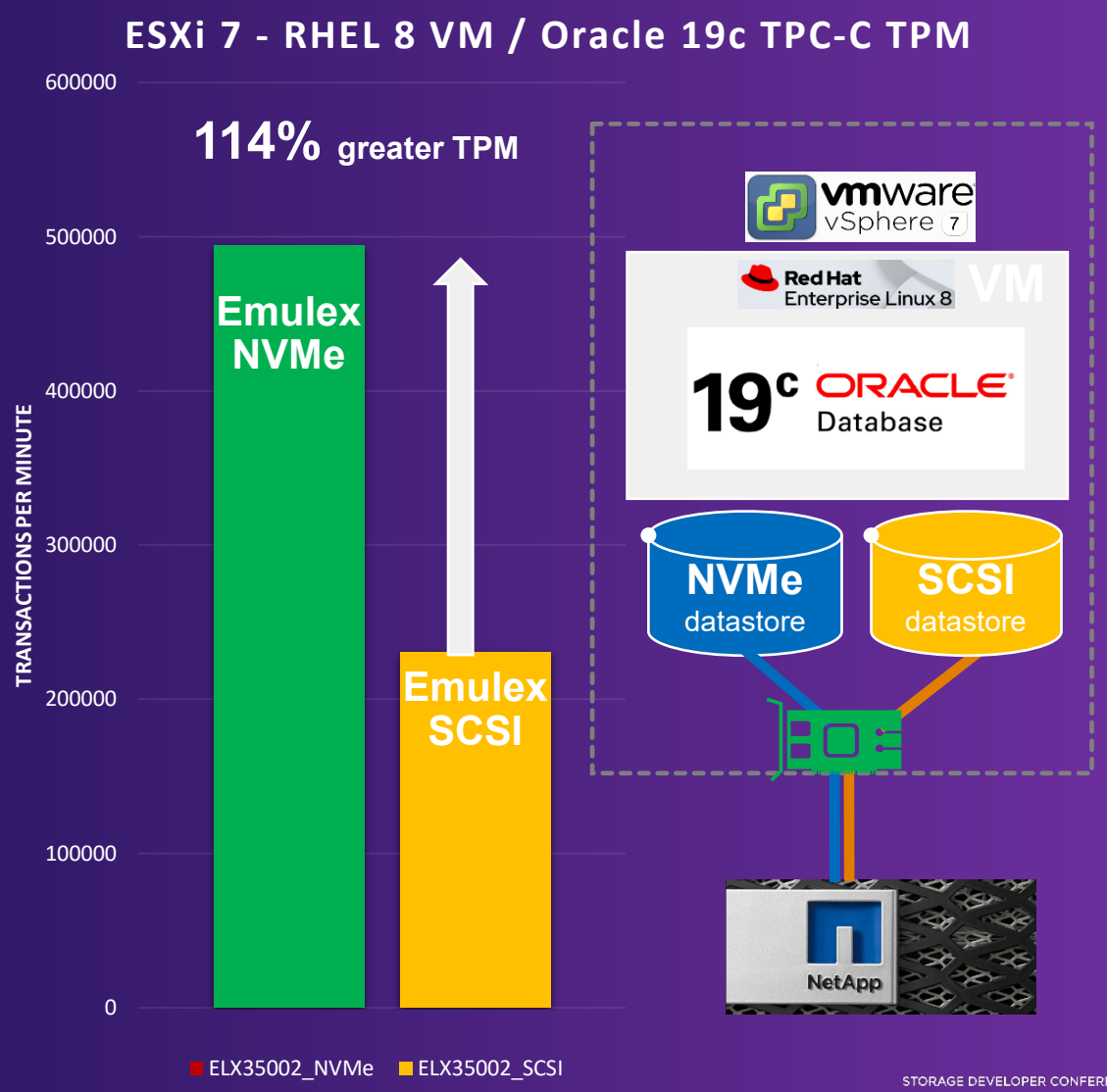


Source: Tolly, April 2020

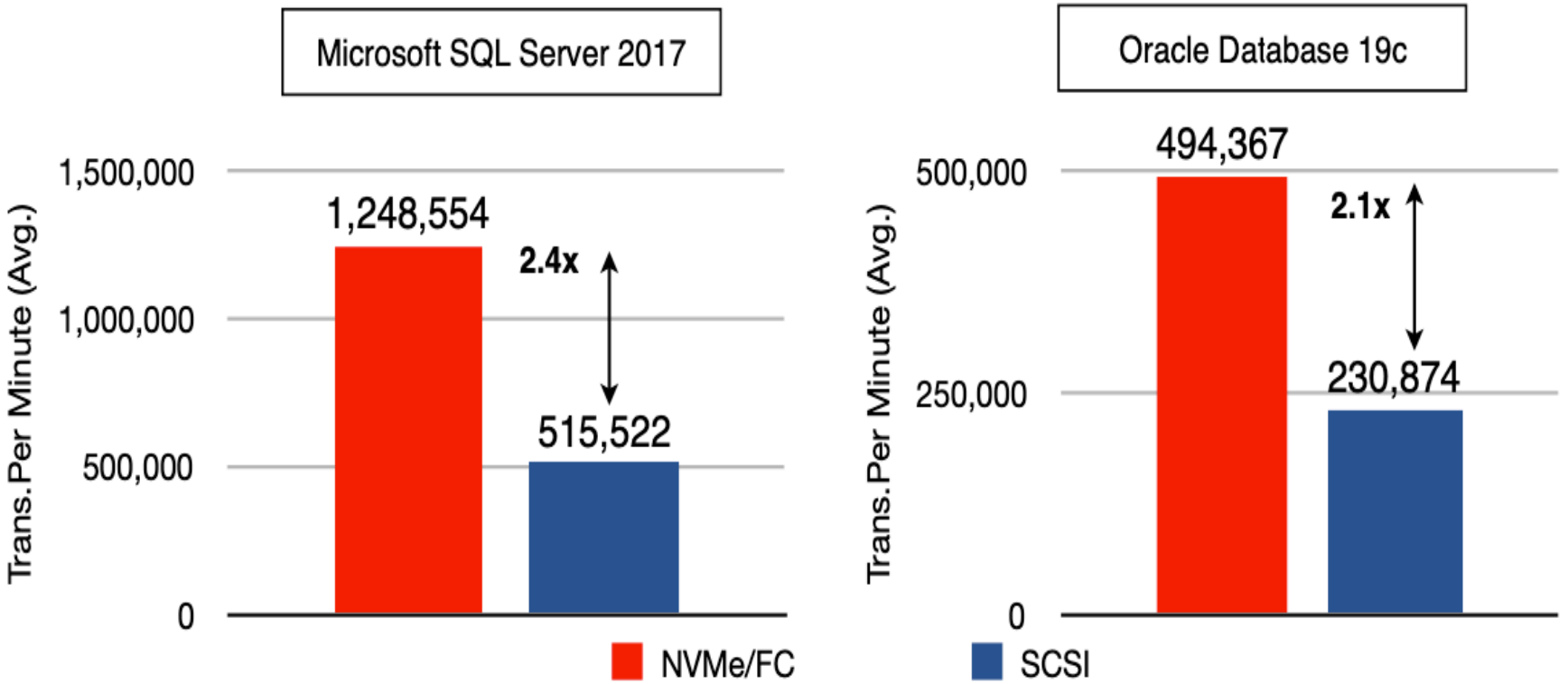
Figure 2

Increased Performance for Virtualized Enterprise Applications





32G Fibre Channel HBA OLTP Database Workloads
NVMe/FC vs. SCSI: VMware ESXi 7.0 Server
(as reported by HammerDB)



Notes: HammerDB TPC-C load. RHEL8 VM environment. Better of two runs. All flash NVMe array target.

Source: Tolly, April 2020

Figure 1

Please take a moment to rate this session.

Your feedback is important to us.

