Real-World Performance of Flash-Based Storage Systems

Session 104-C

Dennis Martin, President, Demartek

Agenda

- About Demartek
- Enterprise Datacenter Environments
- Metrics Basics
- Real-World Workloads
- Performance Results: Various Flash Technologies

Demartek Services Video

Click to view this one minute video
(available in 720p and 1080p)

Demartek YouTube Channel:
http://www.youtube.com/user/Demartek/videos

About Demartek

- Industry Analysis and ISO 17025 accredited test lab
- Lab includes enterprise servers, networking & storage
- We prefer to run real-world applications to test servers and storage solutions
- Demartek is an EPA-recognized test lab for *ENERGY STAR Data Center Storage* testing
- Website: www.demartek.com

Enterprise Datacenter Environments

- Typically support a large number of users and are responsible for many business applications
- Often have specialists for applications, operating environments, networking and storage systems
- Have a large amount of equipment including servers, networking and storage gear
  - Multiple types and generations within each category
- Reliability, Availability and Serviceability (RAS)
- Complex systems working together

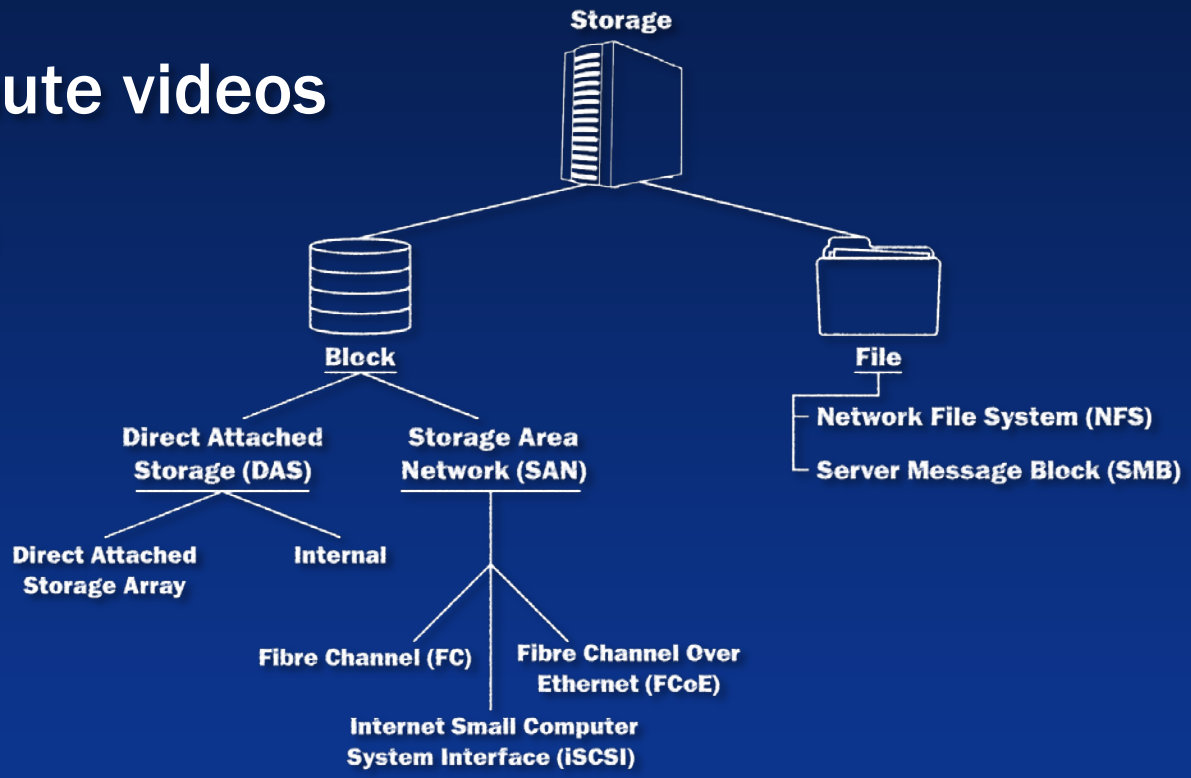
Enterprise Storage Architectures

▸ Flash Can Be Deployed In Any of These

- Direct Attach Storage (DAS)
  - Storage controlled by a single server: inside the server or directly connected to the server ("server-side")
  - ***Block*** storage devices
- Network Attached Storage (NAS)
  - File server that sends/receives ***files*** from network clients
- Storage Area Network (SAN)
  - Delivers shared ***block*** storage over a storage network

Demartek Tutorial Videos

- Short (3-4) minute videos
- Storage Basics



- http://www.demartek.com/Demartek_Tutorial_Video.html

Interface vs. Storage Device Speeds

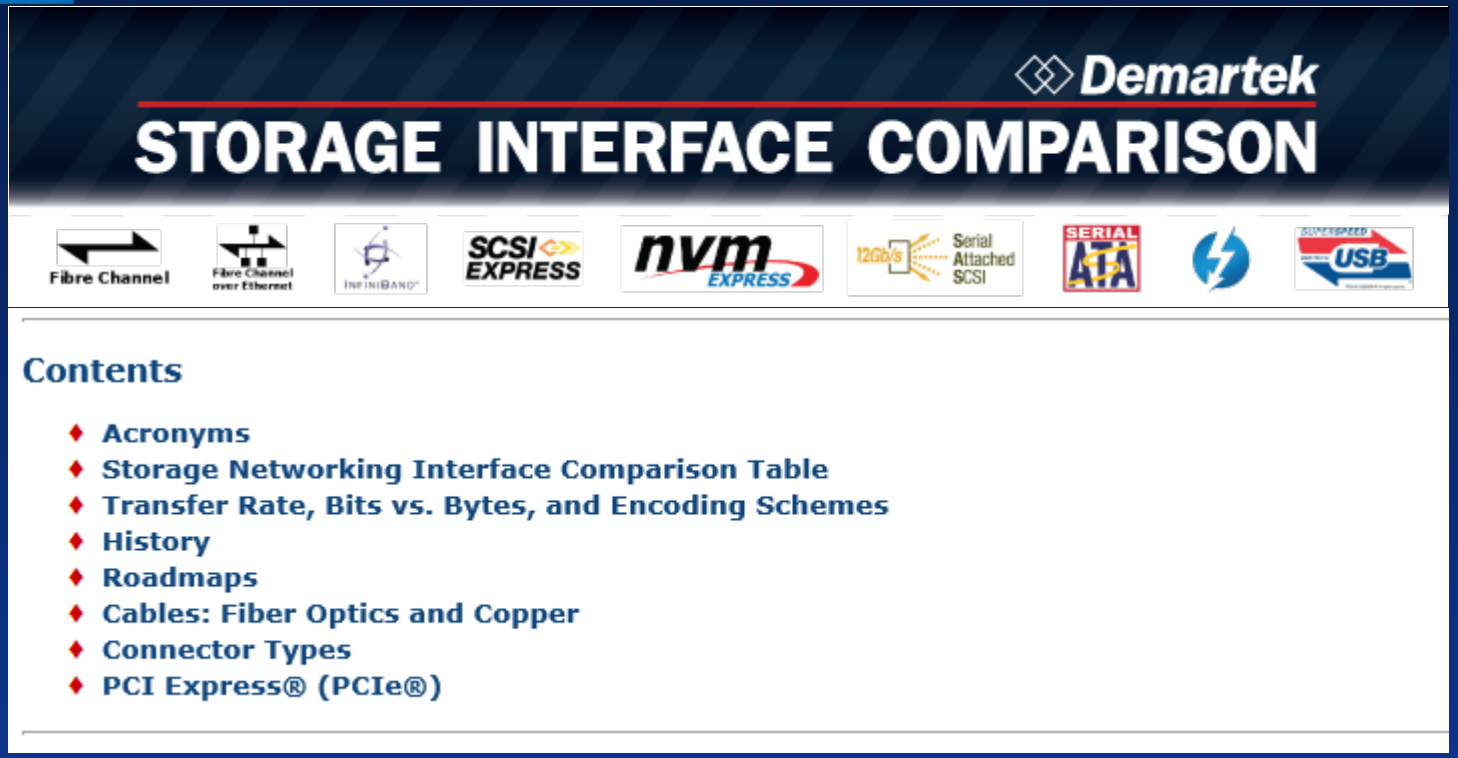
- Interface speeds are generally measured in bits per second, such as megabits per second (Mbps) or gigabits per second (Gbps).
  - Lowercase “b”
  - Applies to Ethernet, Fibre Channel, SATA, etc.
- Storage device and system speeds are generally measured in bytes per second, such as megabytes per second (MBps) or gigabytes per second (GBps).
  - Uppercase “B”
  - Applies to PCIe

Storage Interface Types

▸ Some for devices, others between systems

- Ethernet
- Fibre Channel (FC) and FC over Ethernet (FCoE)
- Infiniband
- PCIe and NVMe
- SAS
- SATA
- Thunderbolt
- USB

Storage Interface Comparison



- Downloadable interactive PDF version now available
- Search engine: “storage interface comparison”
- www.demartek.com/Demartek_Interface_Comparison.html

Key Storage Metrics
▸ IOPS & Bandwidth

- IOPS
  - Number of Input/Output (I/O) requests per second
- Bandwidth
  - Measure of bytes transferred per second (MBps or GBps)
- Read and Write metrics are often reported separately

Key Storage Metrics
▸ Latency

- Latency
  - Response time or round-trip time, generally measured in milliseconds (ms) or microseconds (µs)
  - Sometimes measured as seconds per transfer
  - Time is the numerator, therefore lower latency is faster
- Latency is becoming an increasingly important metric for many real-world applications
- Flash storage provides much lower latency than hard disk or tape technologies

Real-World Workloads

- Use variable levels of compute, memory and Input/Output (I/O) resources as the work progresses
- Typically use multiple block sizes and queue depths for I/O requests, depending on the workload
- Many applications capture their own metrics such as database transactions per second, etc.
- Operating systems can track physical and logical I/O metrics

Real-World Storage Workload Types

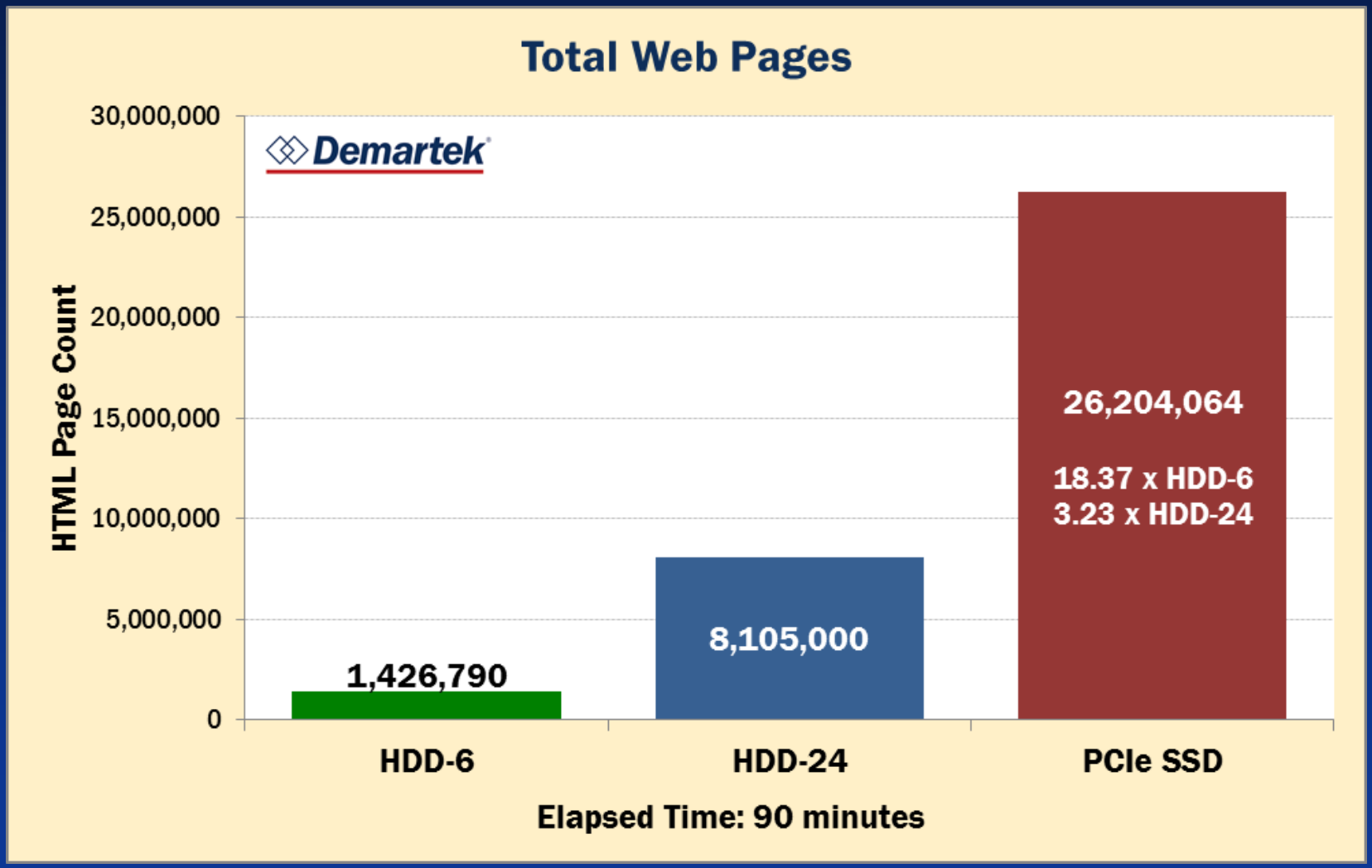
- Transactional (Random)
  - Generally smaller block sizes (4KB, 8KB, 16KB, etc.)
  - Emphasis on the number of I/O's per second (IOPS)
- Streaming (Sequential)
  - Generally larger block sizes (64KB, 256KB, 1MB, etc.)
  - Emphasis on bandwidth or throughput measured in Megabytes per second (MBps)
- Latency is affected differently by different workload types

Performance Results

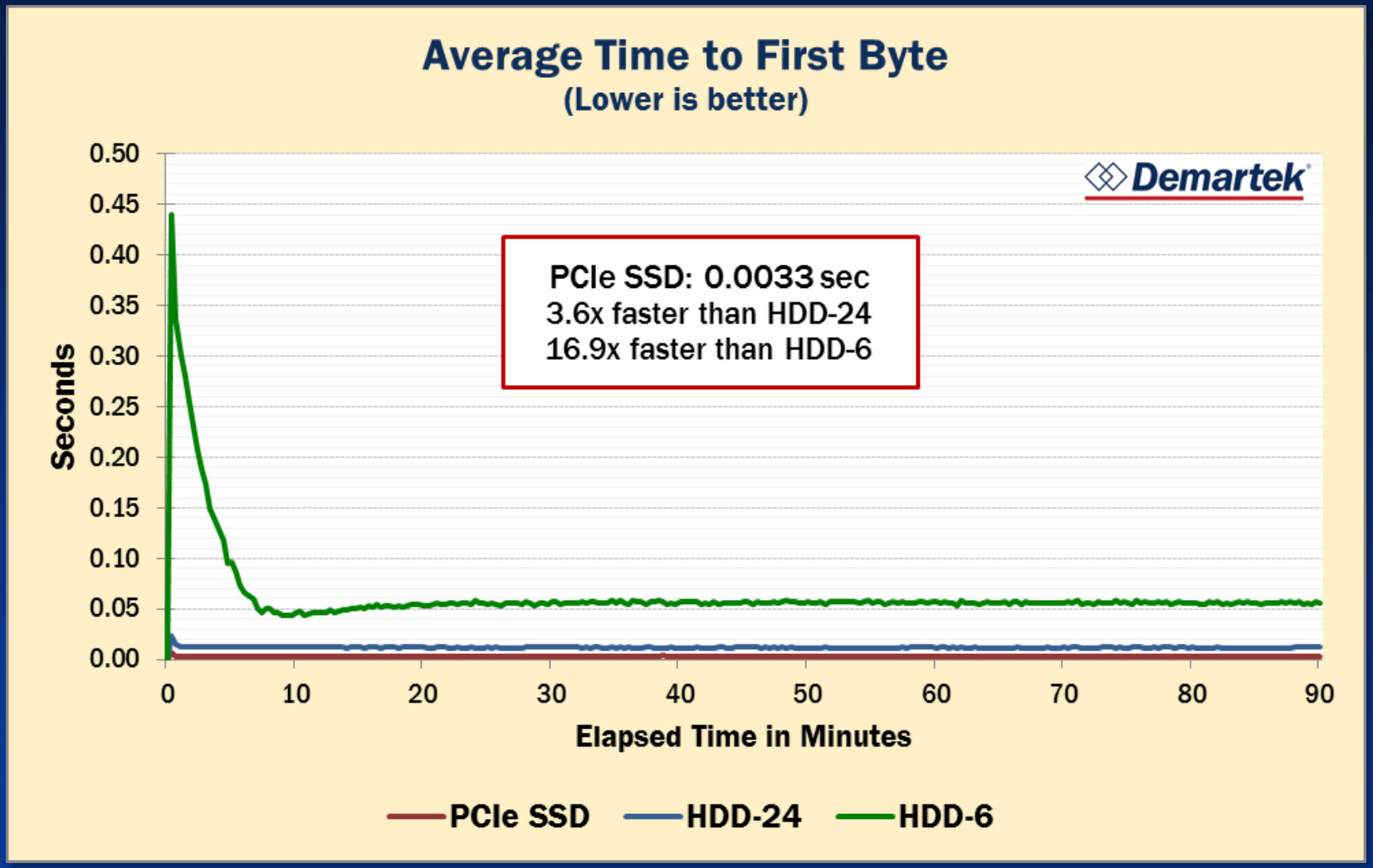
Web Server Test

- Read-intensive web server workload
  - 40GB web server data
  - 1.48 million files
    - 80,000 unique HTML text pages
    - 1.4 million graphic images (JPEG and PNG)
  - Randomly referenced all pages (1 HTML text + 3 images) approximately evenly over a 90-minute test period
- Storage: 6 HDD vs. 24 HDD vs. 1 PCIe SSD
  - HDDs: 73GB 15K RPM SAS, RAID10
  - PCIe SSD: 300GB

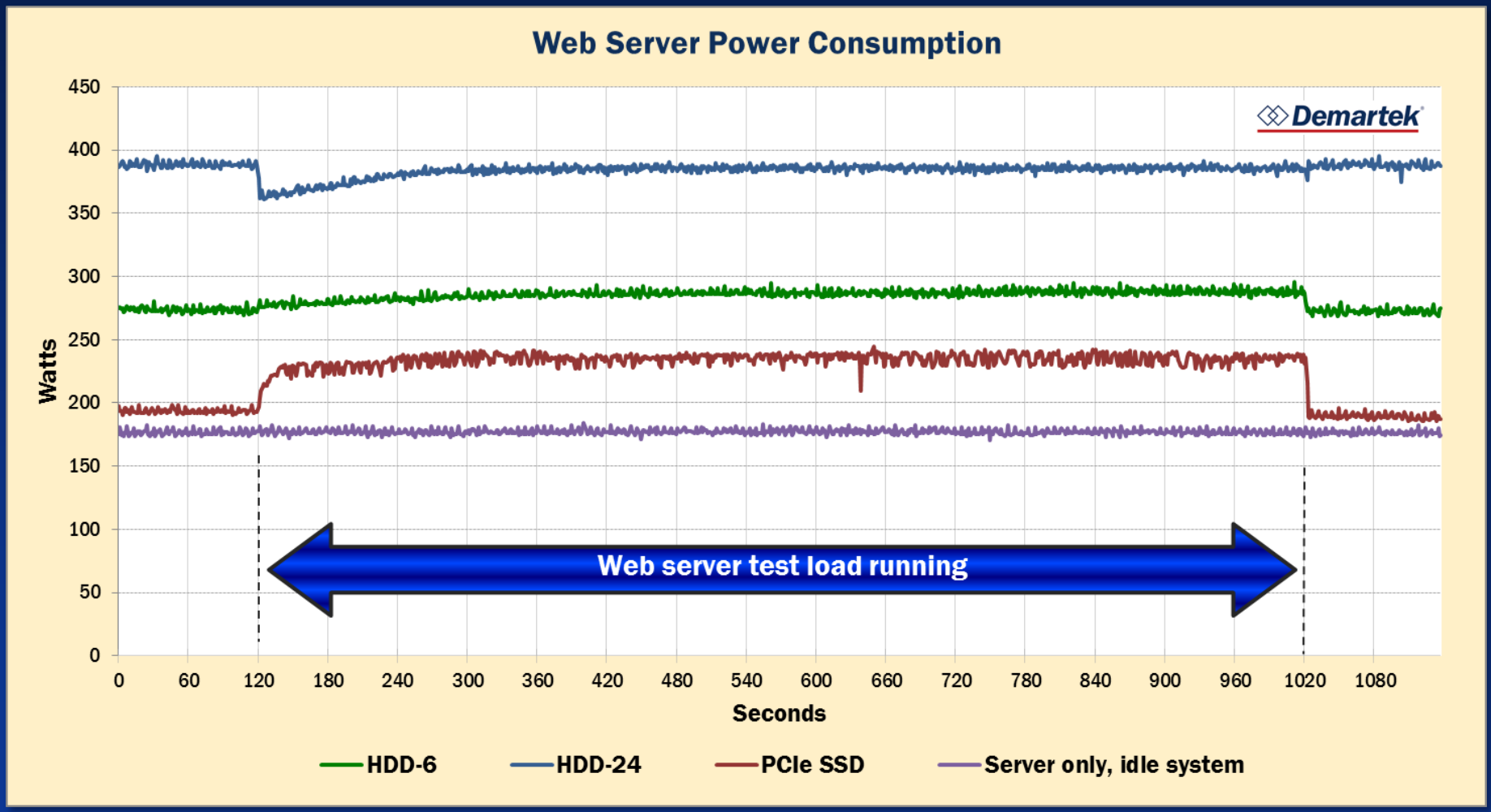
Total Web Pages



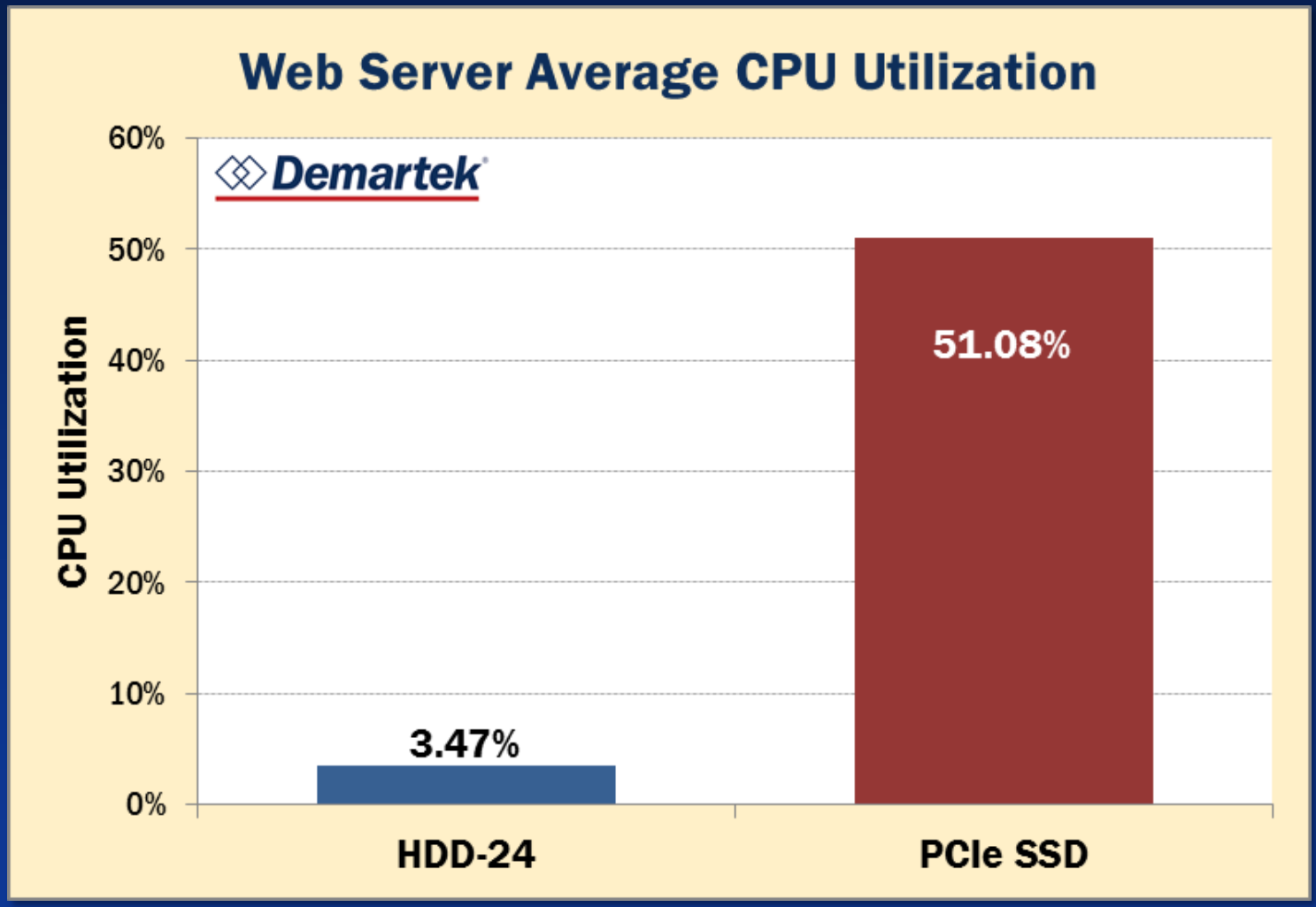
Average Time to First Byte



Web Server Power Consumption

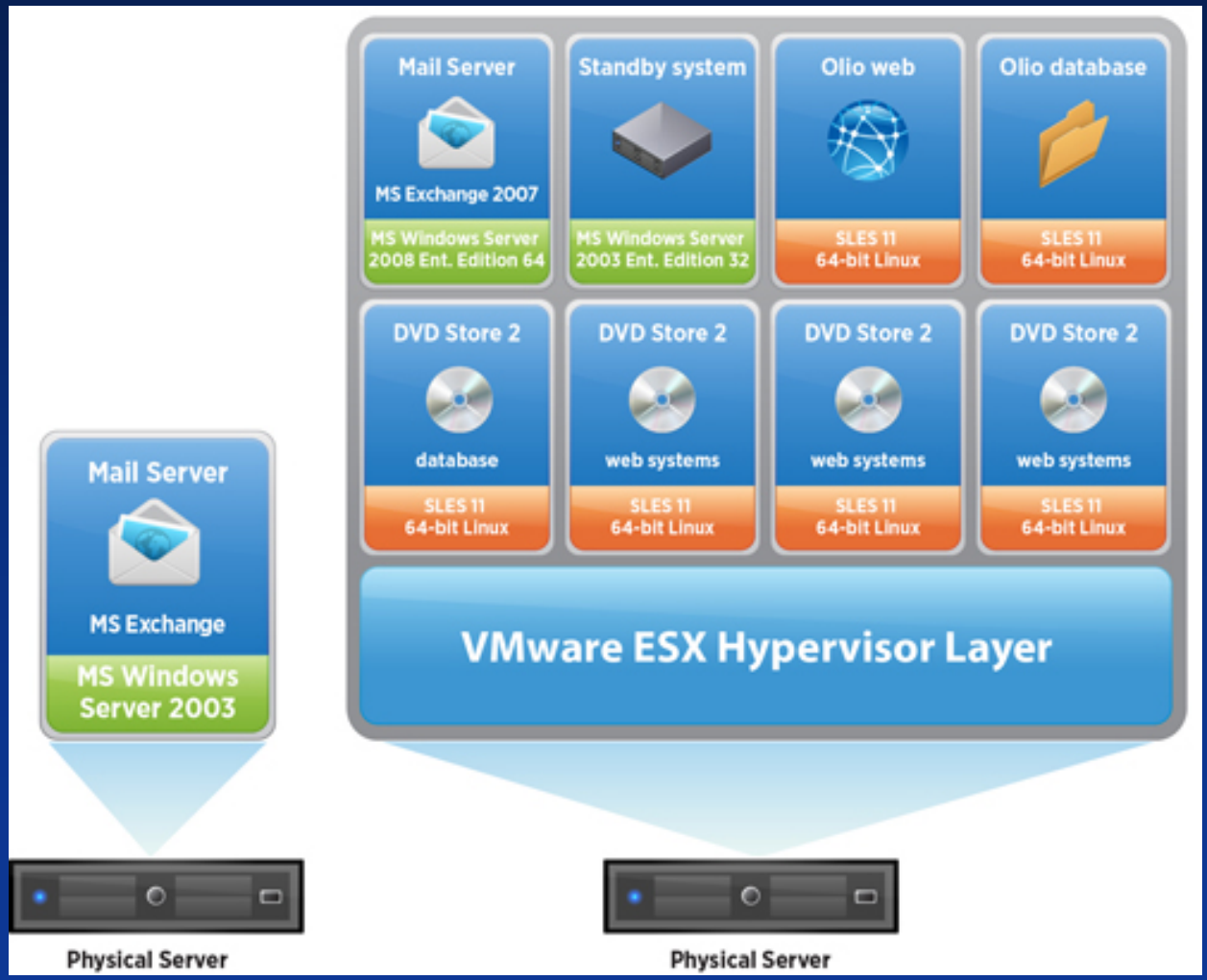


Web Server Average CPU Utilization



VMmark Virtualization Tests

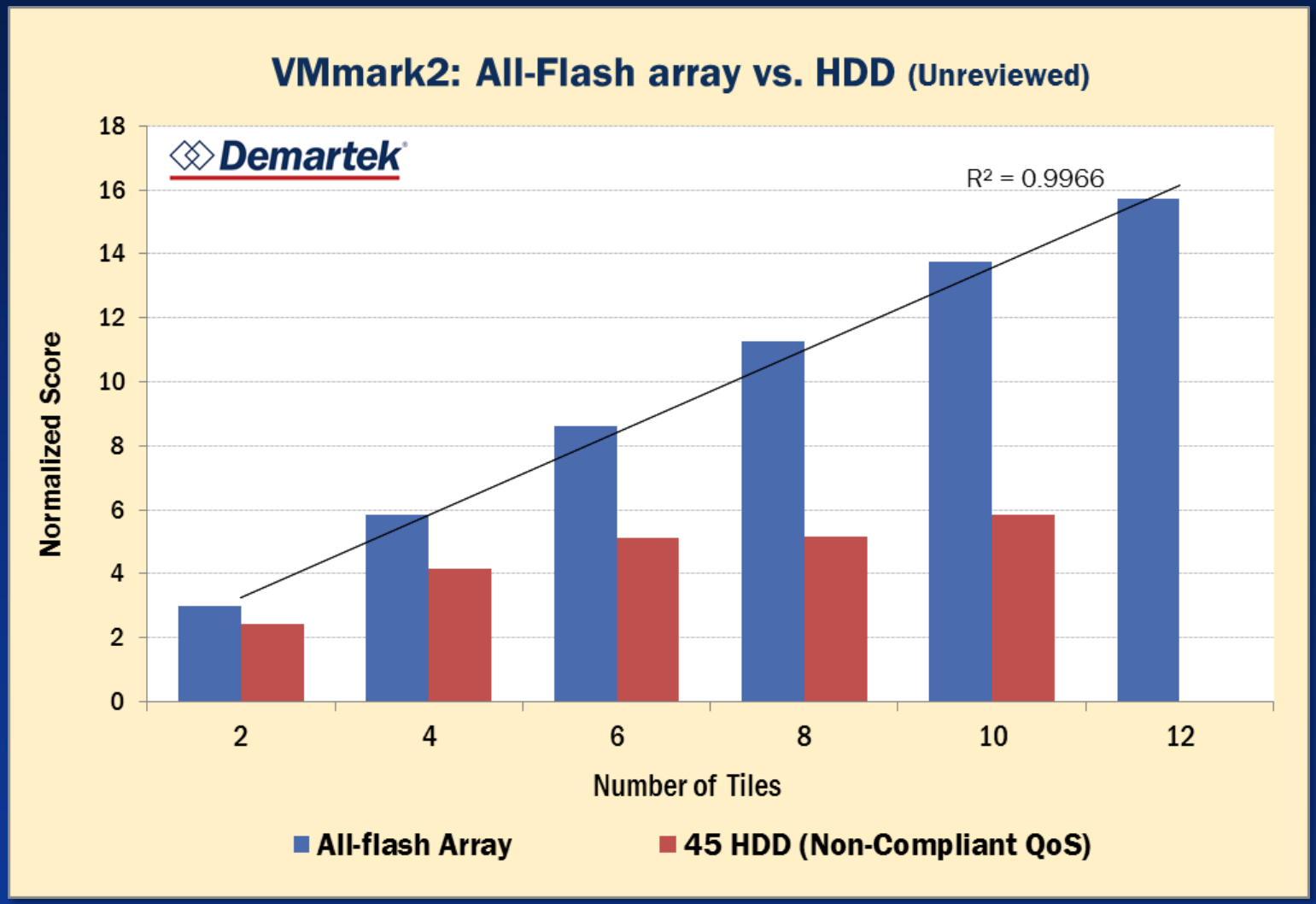
- A VMmark "tile" includes these workloads plus vMotion
- Multiple tiles are configured to stress test the storage



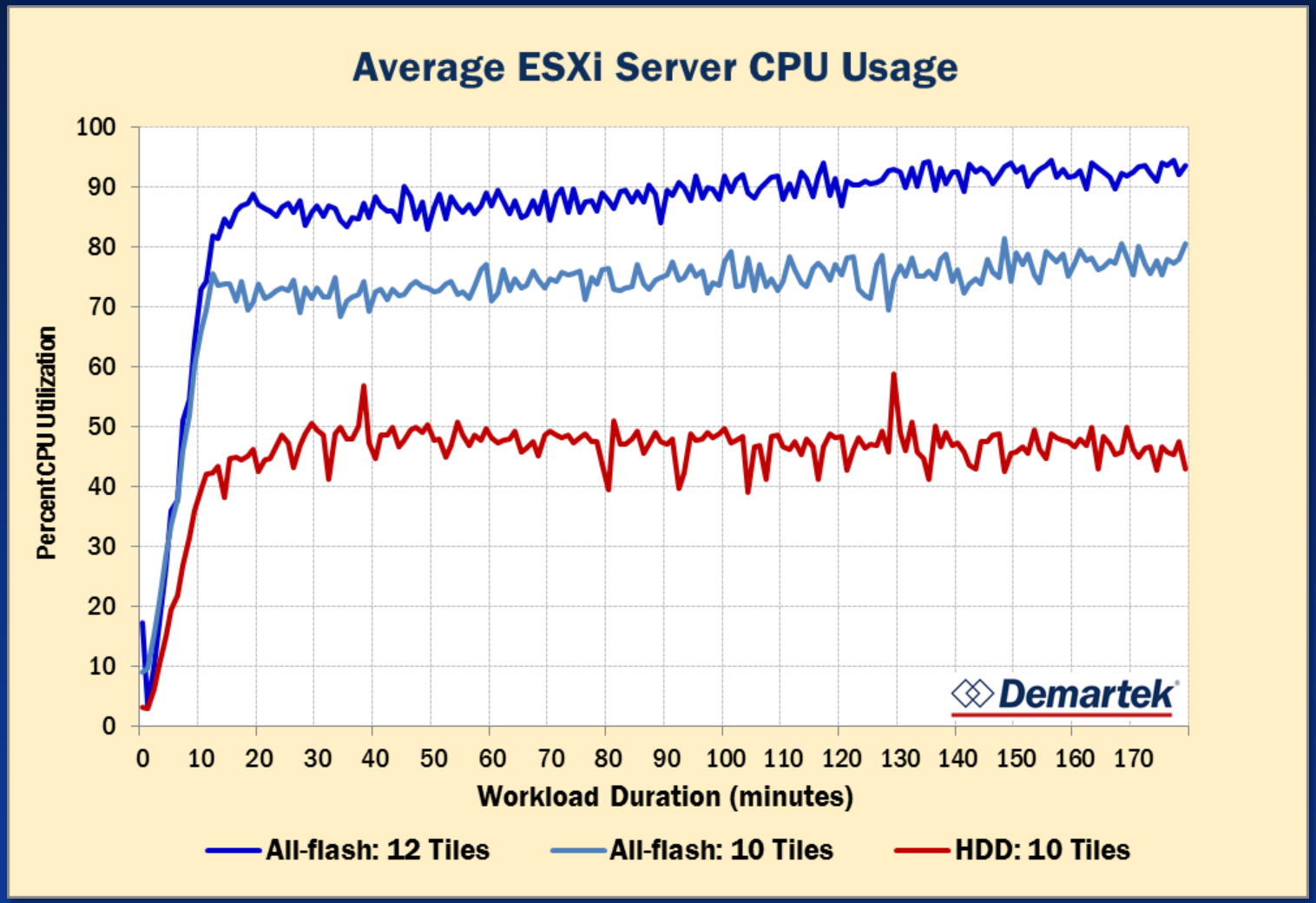
VMmark Configuration

- Storage Infrastructure
  - All-flash array, Fibre Channel SAN attach
  - HDD array (45 HDDs), Fibre Channel SAN attach
  - 16Gb Fibre Channel switch
- Test Cluster Servers
  - Qty. 1: Intel Xeon E5-2690, 2.9 GHz, 16 total cores, 32 logical processors, 192 GB RAM, 16GFC HBAs
  - Qty. 2: Intel Xeon E5-2690 v2, 3.0 GHz, 20 total cores, 40 logical processors, 256 GB RAM, 16GFC HBAs
- Two other servers used for VMmark clients
- 10GbE used for network connections

VMmark2 Scores



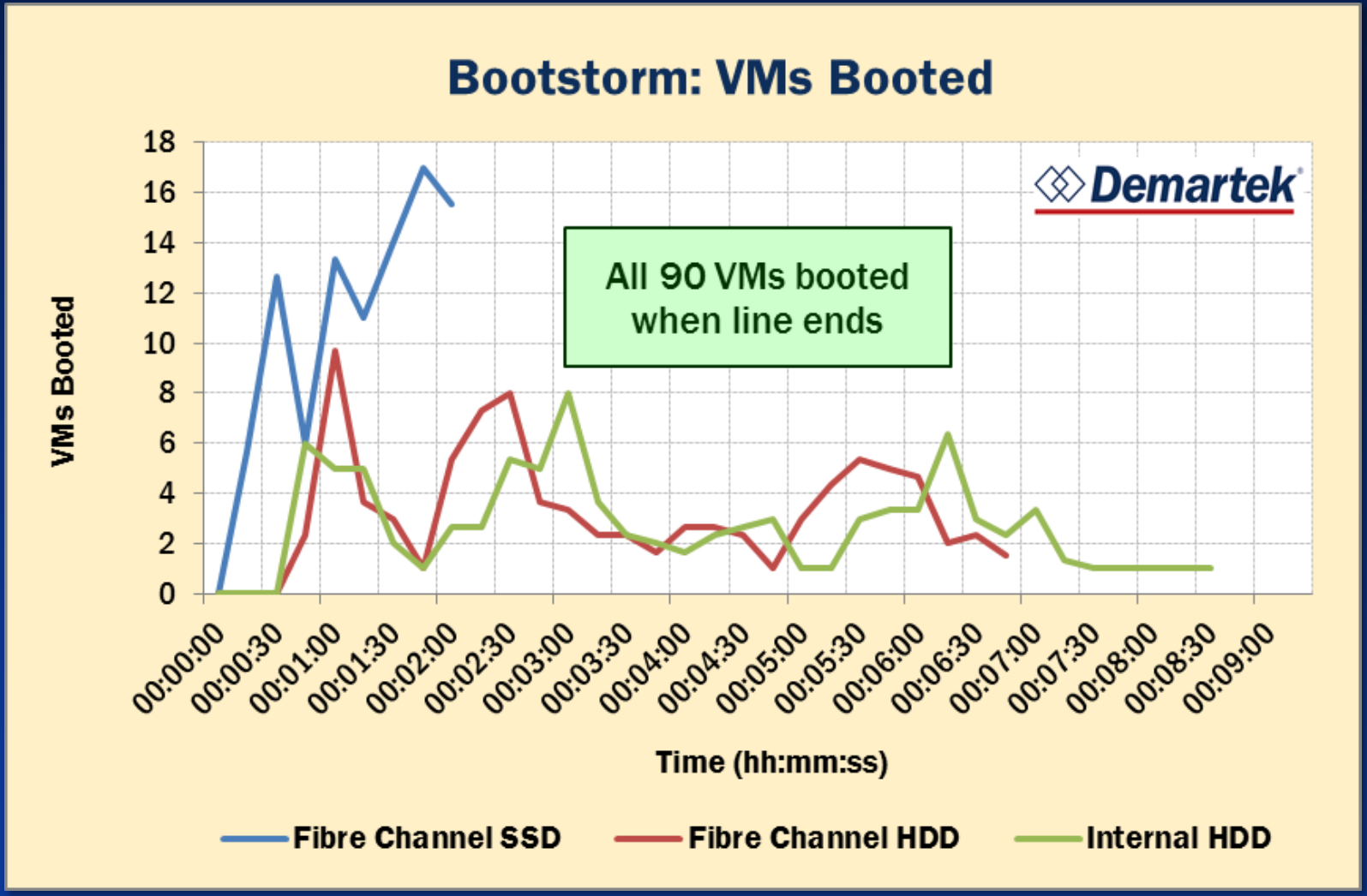
VMmark CPU Utilization



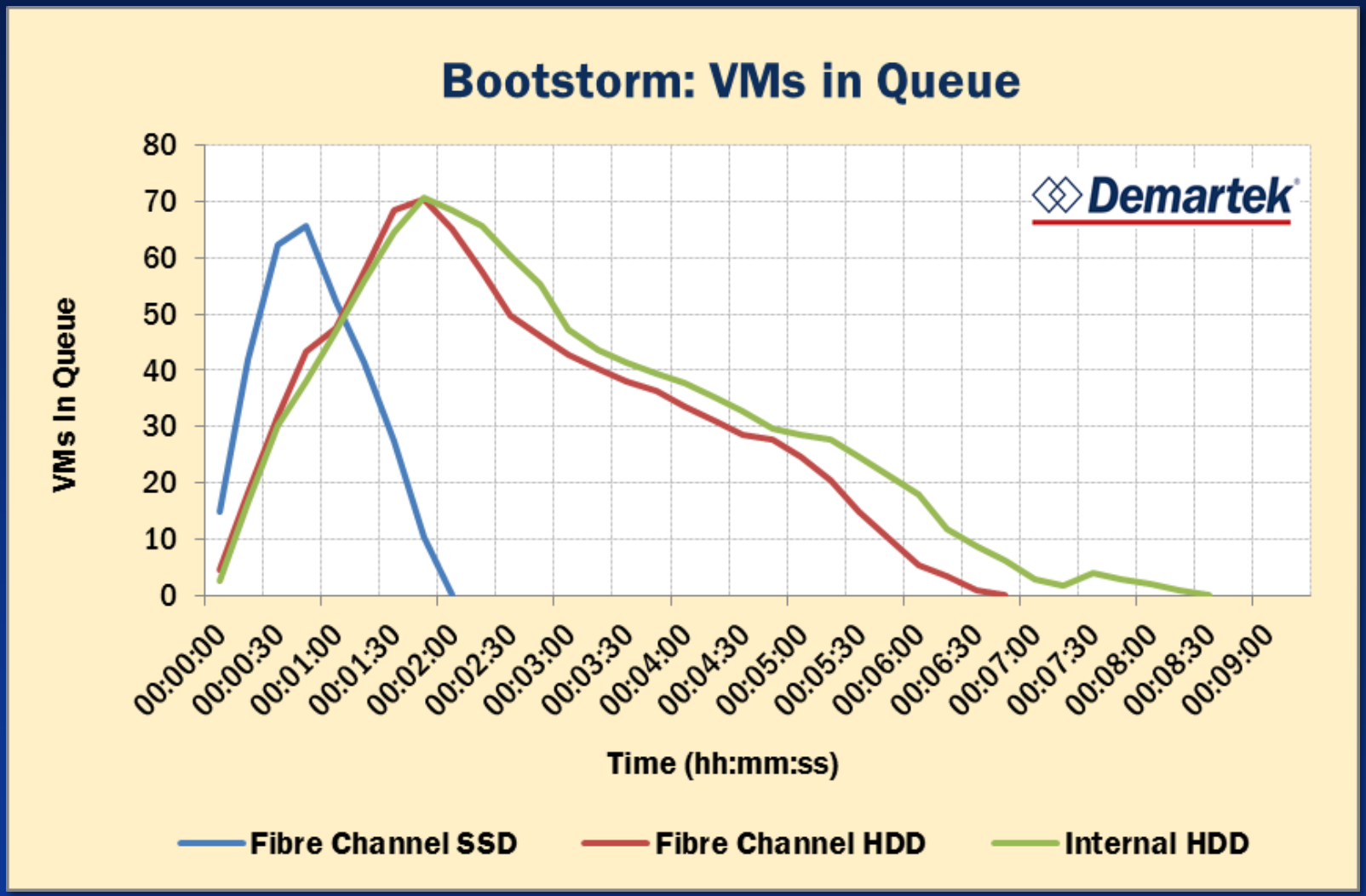
Bootstorm – 90 Virtual Desktops

- Booting 90 desktop virtual machines using one physical server
  - Server: 4x Intel Xeon E5-4650, 2.7 GHz, 32 total cores, 64 logical processors, 256 GB RAM
  - Hypervisor: ESXi 5.1
  - Desktop VMs: Windows 7 Ultimate, 1 vCPU, 2GB RAM
- Use different storage for boot images and VMs
  - Internal HDD: 15x 15K 136GB SAS, RAID0
  - External HDD: 12x 15K 300GB SAS, RAID0, 8Gb FC SAN
  - External SSD: 24x 100GB SSD, RAID0, 8Gb FC SAN

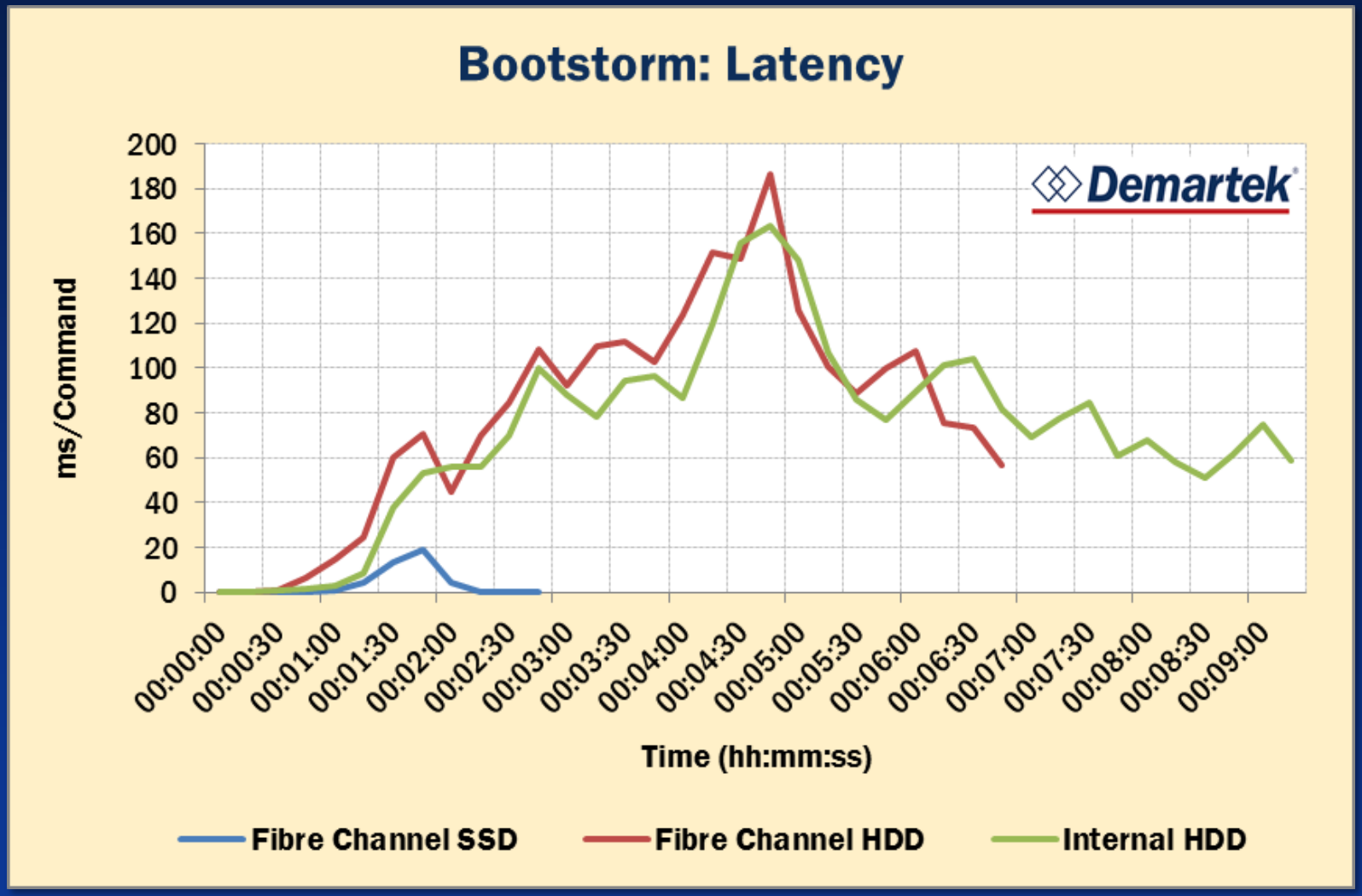
VMs Booted


Bootstorm: VMs Booted
Demartek
All 90 VMs booted when line ends
VMs Booted
0
2
4
6
8
10
12
14
16
18
00:00:00
00:00:30
00:01:00
00:01:30
00:02:00
00:02:30
00:03:00
00:03:30
00:04:00
00:04:30
00:05:00
00:05:30
00:06:00
00:06:30
00:07:00
00:07:30
00:08:00
00:08:30
00:09:00
Time (hh:mm:ss)
Fibre Channel SSD
Fibre Channel HDD
Internal HDD

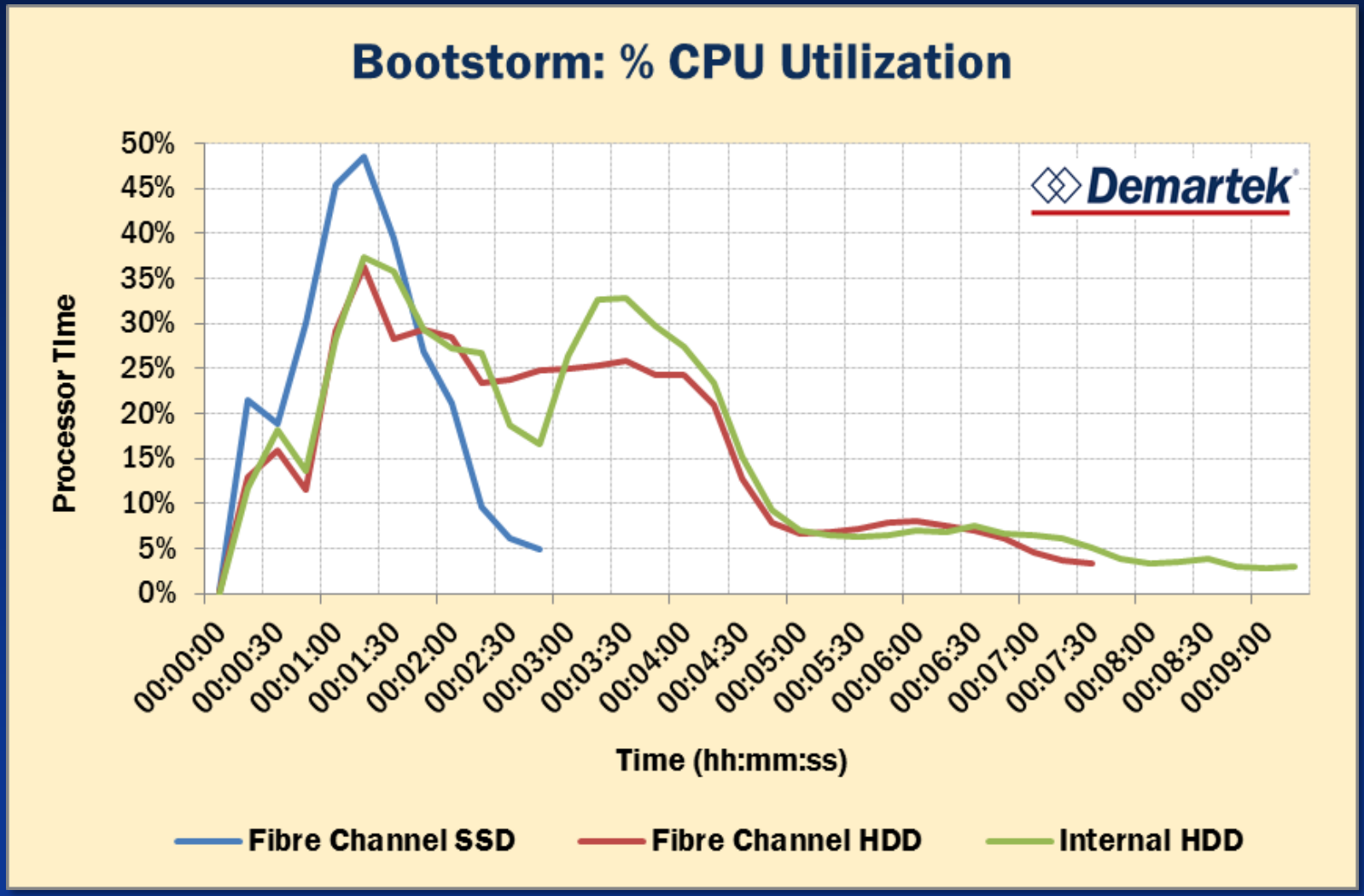
VMs in Queue


Bootstorm: VMs in Queue
Demartek
VMs In Queue
80
70
60
50
40
30
20
10
0
00:00:00
00:00:30
00:01:00
00:01:30
00:02:00
00:02:30
00:03:00
00:03:30
00:04:00
00:04:30
00:05:00
00:05:30
00:06:00
00:06:30
00:07:00
00:07:30
00:08:00
00:08:30
00:09:00
Time (hh:mm:ss)
Fibre Channel SSD
Fibre Channel HDD
Internal HDD

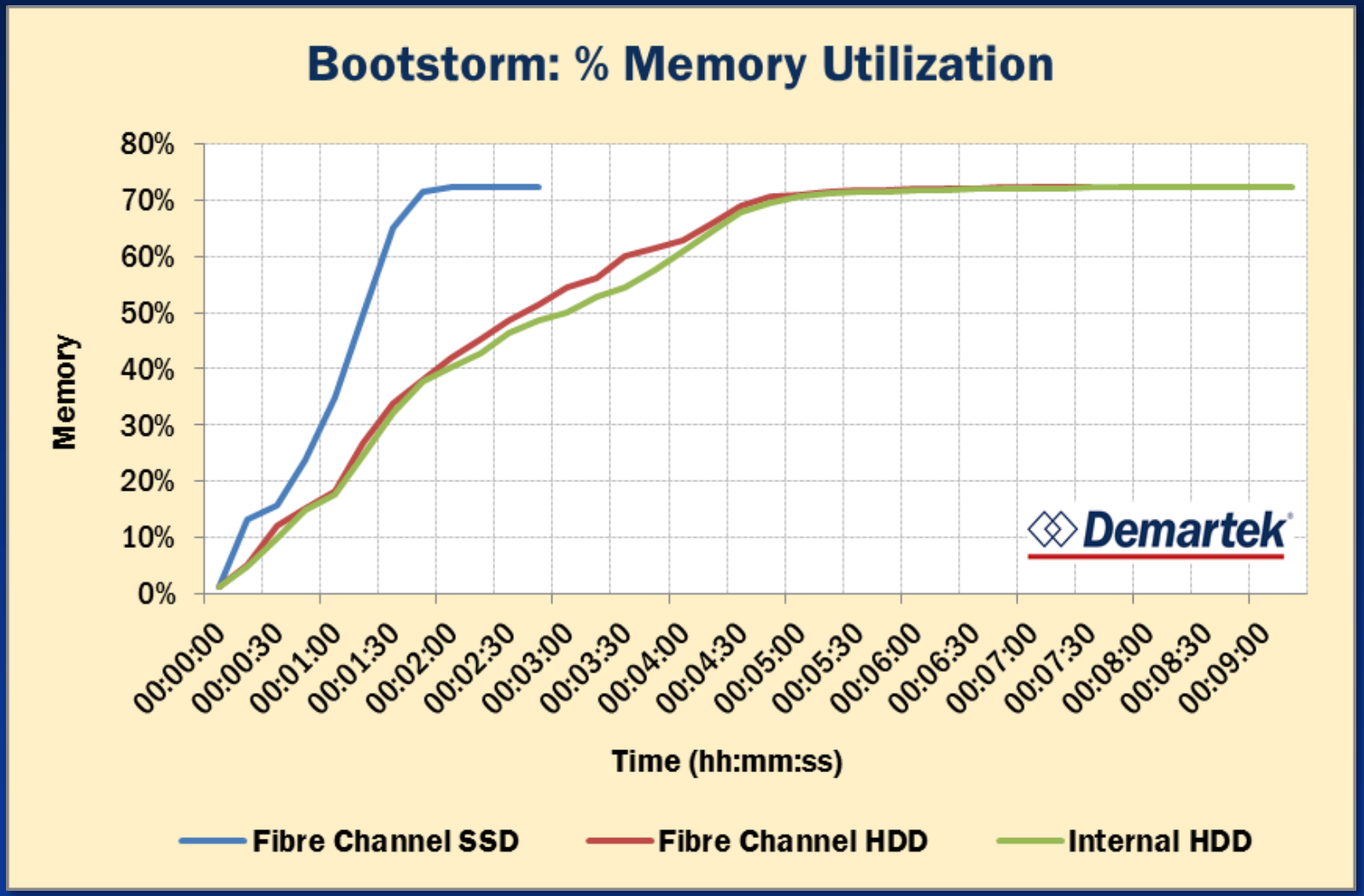
Bootstorm: Latency



Bootstorm: CPU Utilization



Bootstorm: Memory Utilization



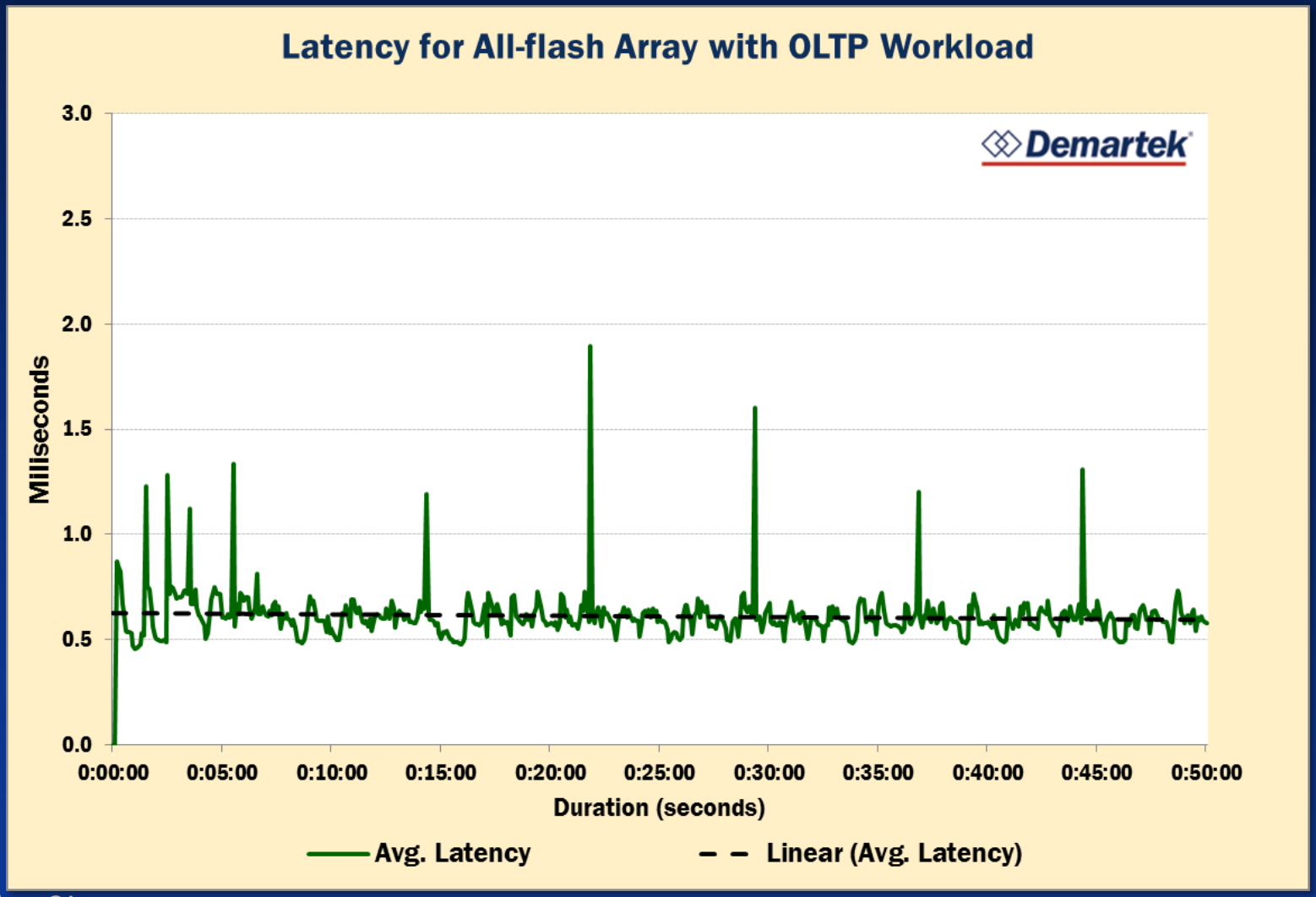
Bootstorm: Other Technical Results

- Internal HDD: RAID controller DRAM cache amounts
- External HDD
  - Read cache (Write-through) vs. Read/Write cache (Write-back)
  - FC HBA queue depth settings
- External SSD: FC HBA queue depth settings
- These data are available in the full report on the Demartek website
  - Search engine: “Demartek bootstorm report”

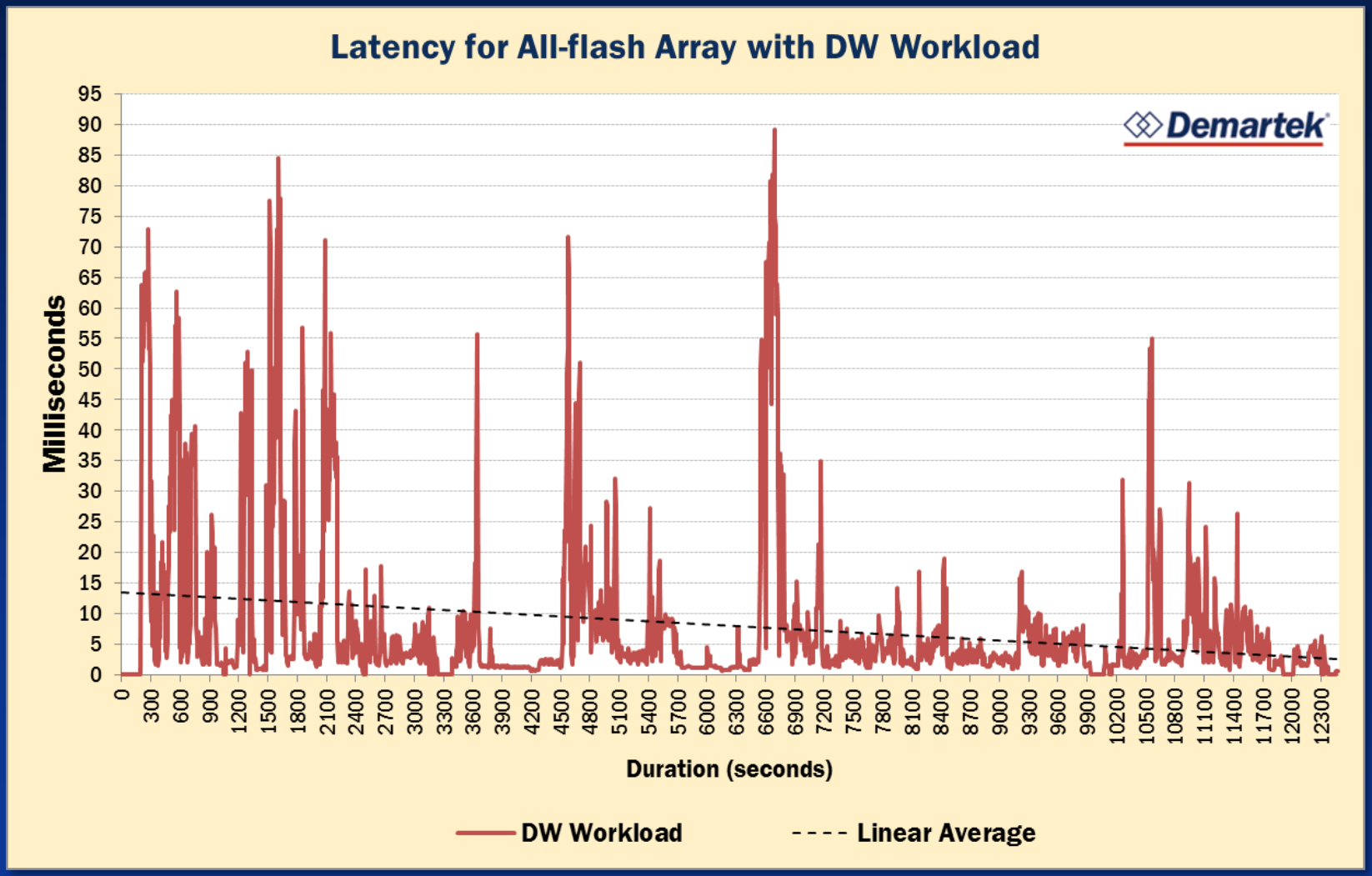
Database Workload Latencies

- Different workloads have different effects on latency, even for all-flash arrays
- Same all-flash array with two different workloads:
  - Microsoft SQL Server Online Brokerage OLTP workload
  - Microsoft SQL Server Data Warehousing (DW) workload

OLTP Workload Latency



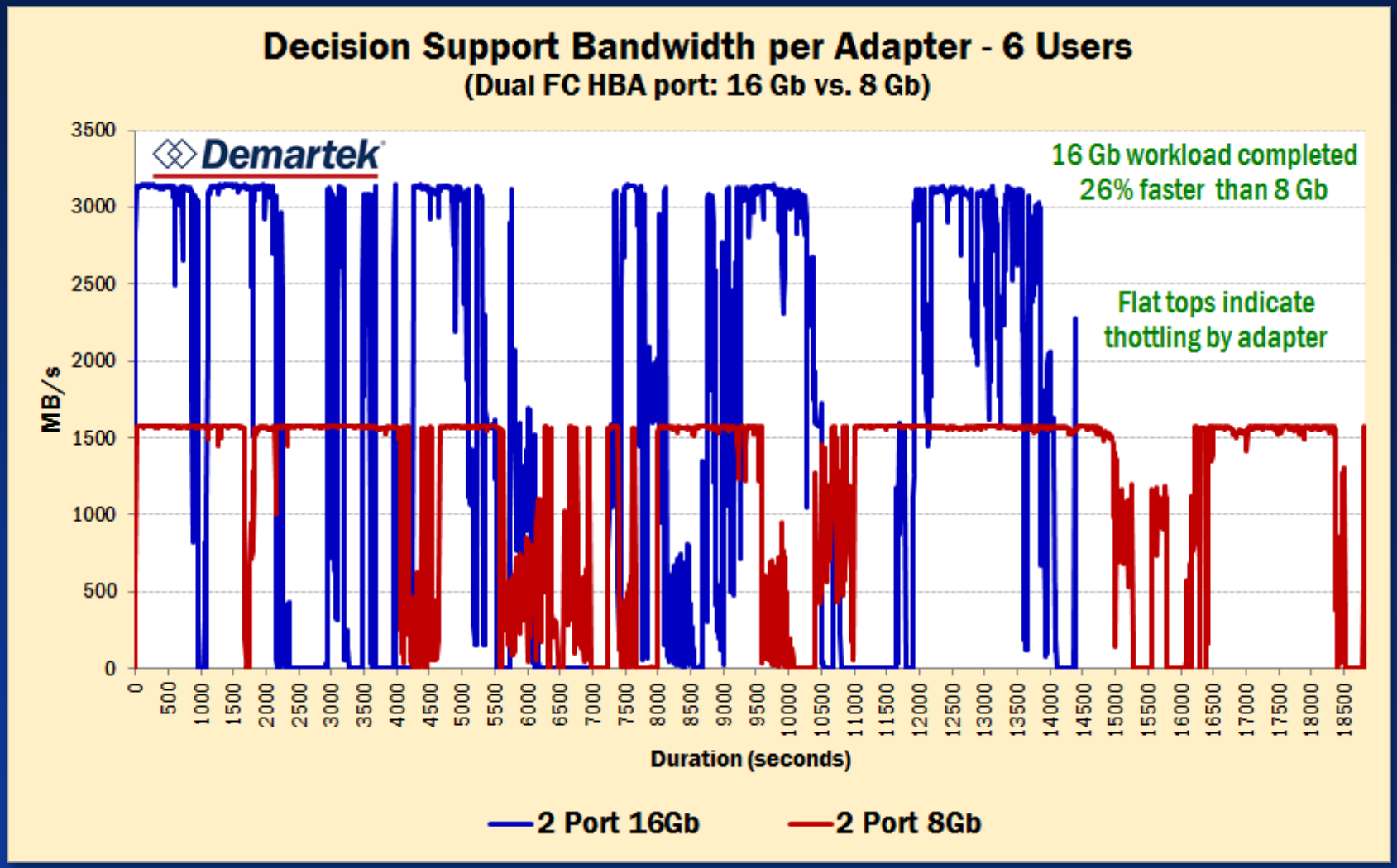
Data Warehousing Latency



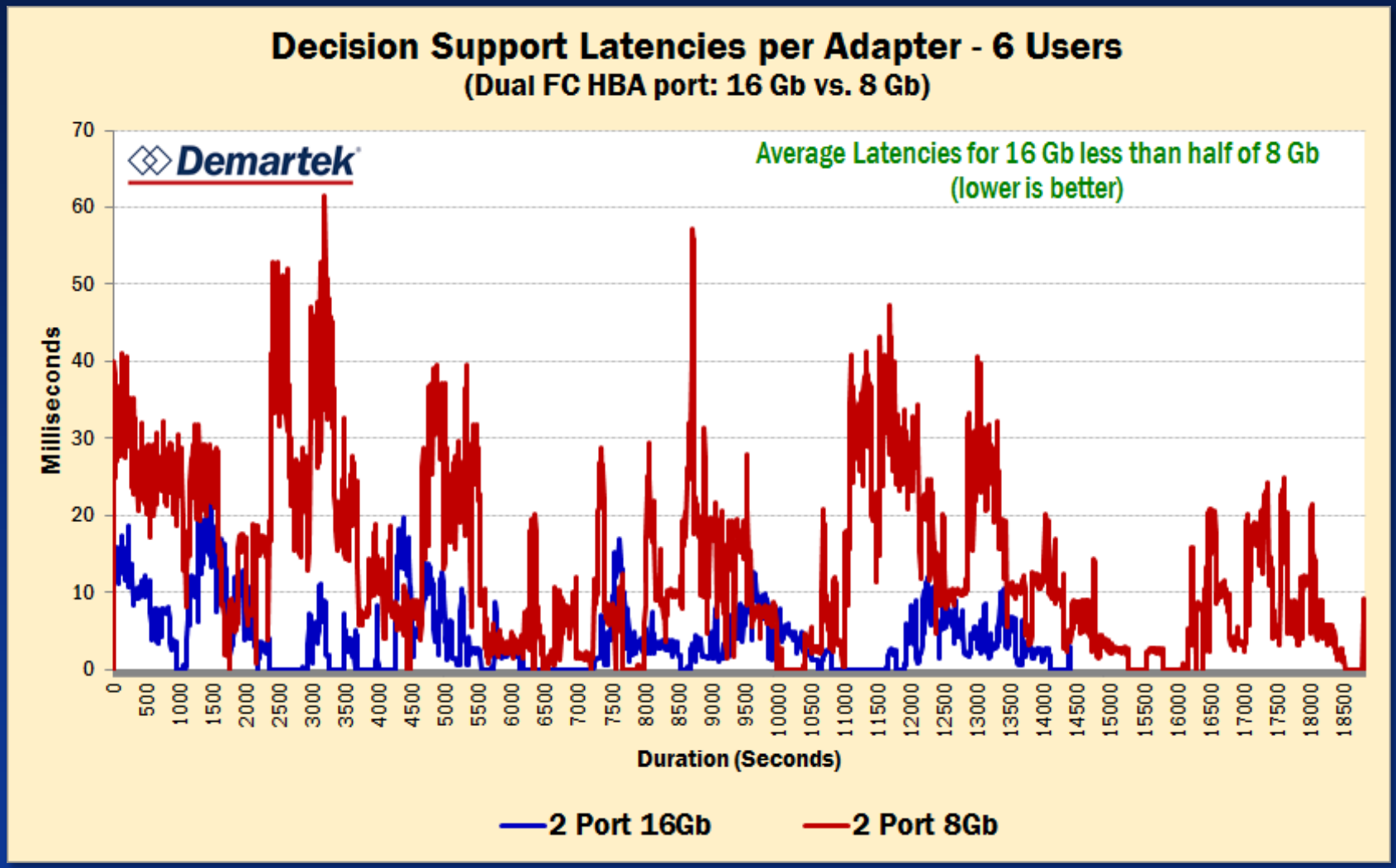
Interface Speed Differences

- All-flash array with decision support database workload
  - Fixed amount of work, faster configuration finishes sooner
- Storage: All-flash array with 4x 8GFC host ports
- Server:
  - 2x Intel Xeon E5-2690, 2.9 GHz, 16 total cores, 32 logical processors, 32GB RAM
  - Dual-port 8GFC HBA – max. bandwidth: 1600 MBps
  - Dual-port 16GFC HBA – max. bandwidth: 3200 MBps

Bandwidth: 16GFC vs. 8GFC



Latency: 16GFC vs. 8GFC



NVM Express (NVMe)

- Scalable host controller interface designed for enterprise and client systems that use PCI Express SSDs
- Designed with Flash memory and technologies coming after Flash memory in mind (non-volatile memory)
- Much faster (lower latency) software stack than existing storage stacks such as SAS and SATA
- Other NVMe sessions here at the Flash Memory Summit
- Additional comments and explanation: www.demartek.com/Demartek_Comments_IDF2013_and_NVMe.html

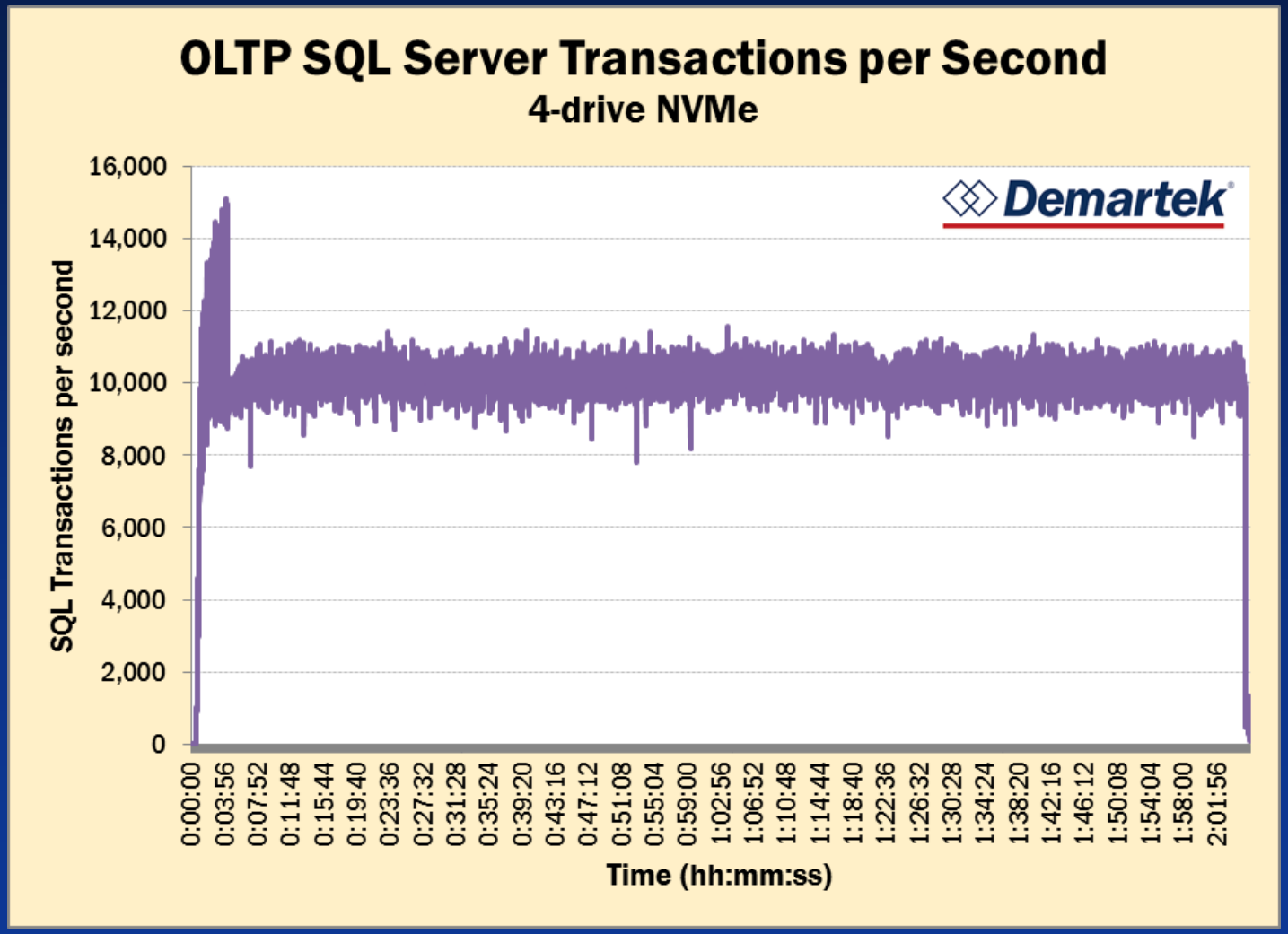
NVM Express (NVMe)

- NVMe SSDs installed in server
  - Two different brands, shipping now
  - No hardware adapter, connects via PCIe (SFF-8639) backplane
- Two database workloads
  - OLTP: Four-drive (2.5-inch) NVMe
  - Data Warehousing: Single-drive (2.5-inch) NVMe
- Currently running additional tests, more results to be published

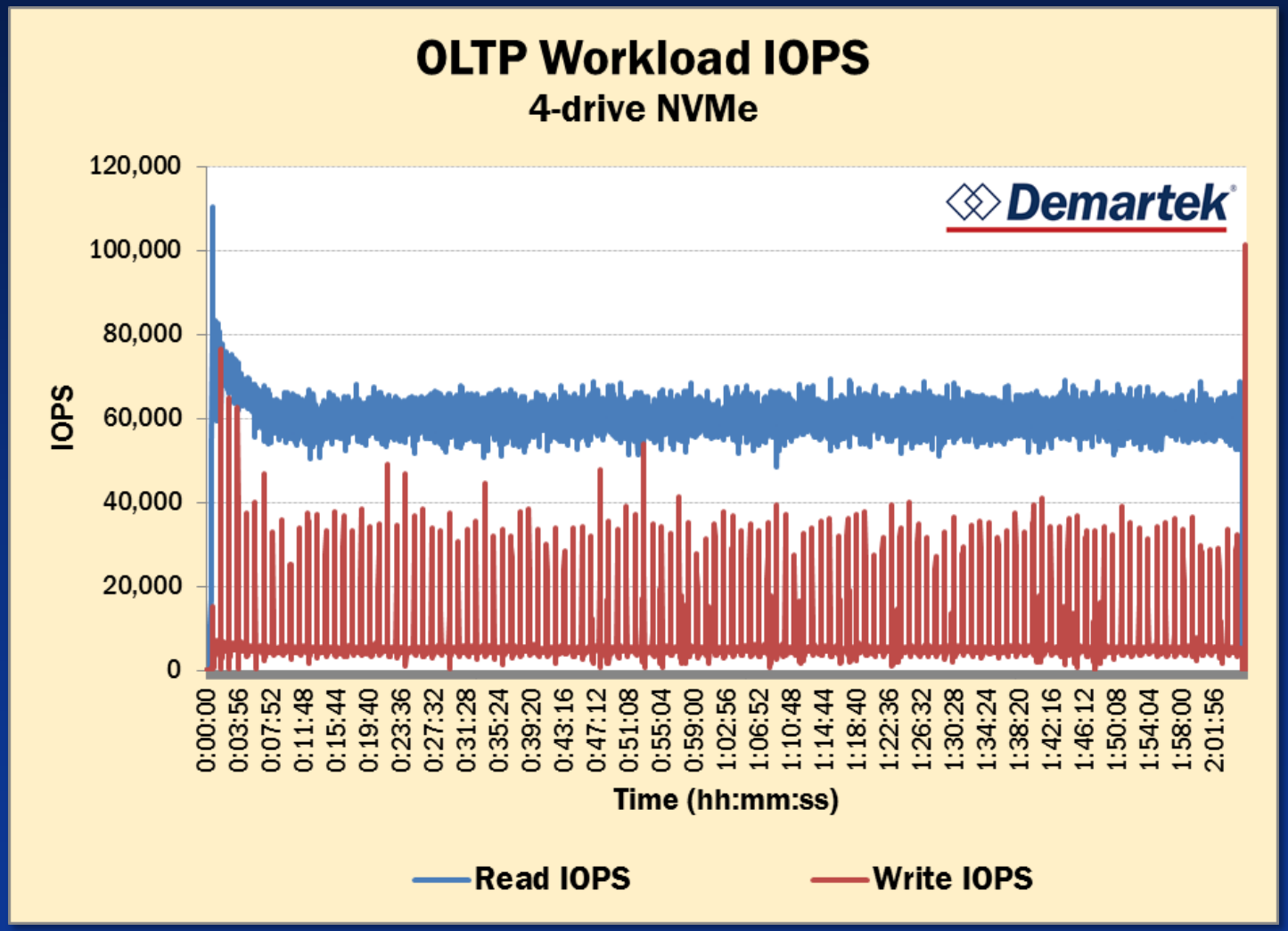
NVMe Configuration

- Server:
  - 4x Intel Xeon E7-4880 v2, 2.5 GHz, 60 cores, 120 logical processors
  - 416 GB RAM
  - SFF-8639 backplanes (NVMe compatible)
  - Windows Server 2012 R2
  - In-box NVMe drivers
  - Microsoft SQL Server 2012
- Four-drive configuration using Windows spanned volume

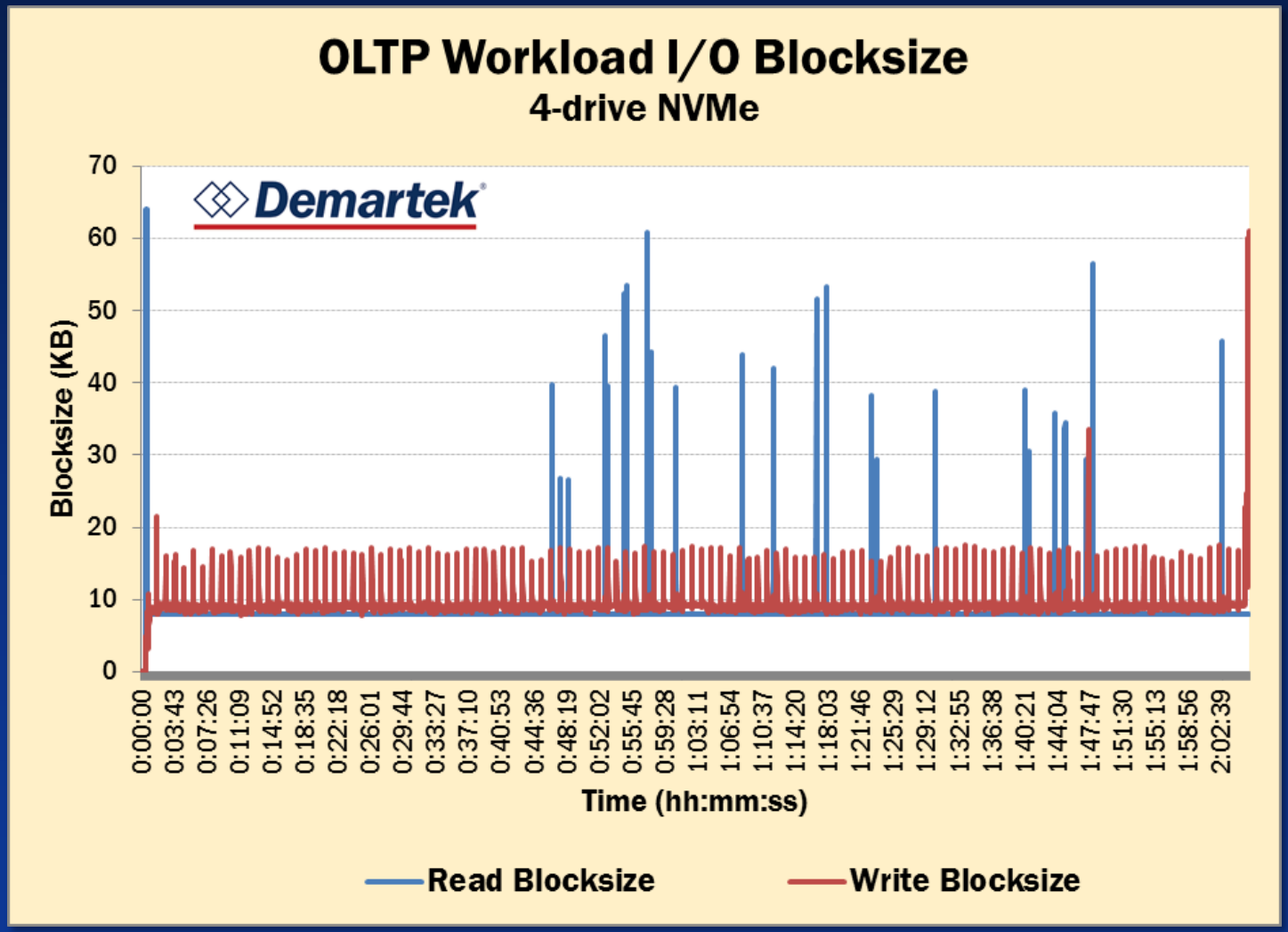
NVMe 4-drive OLTP SQL trans./sec



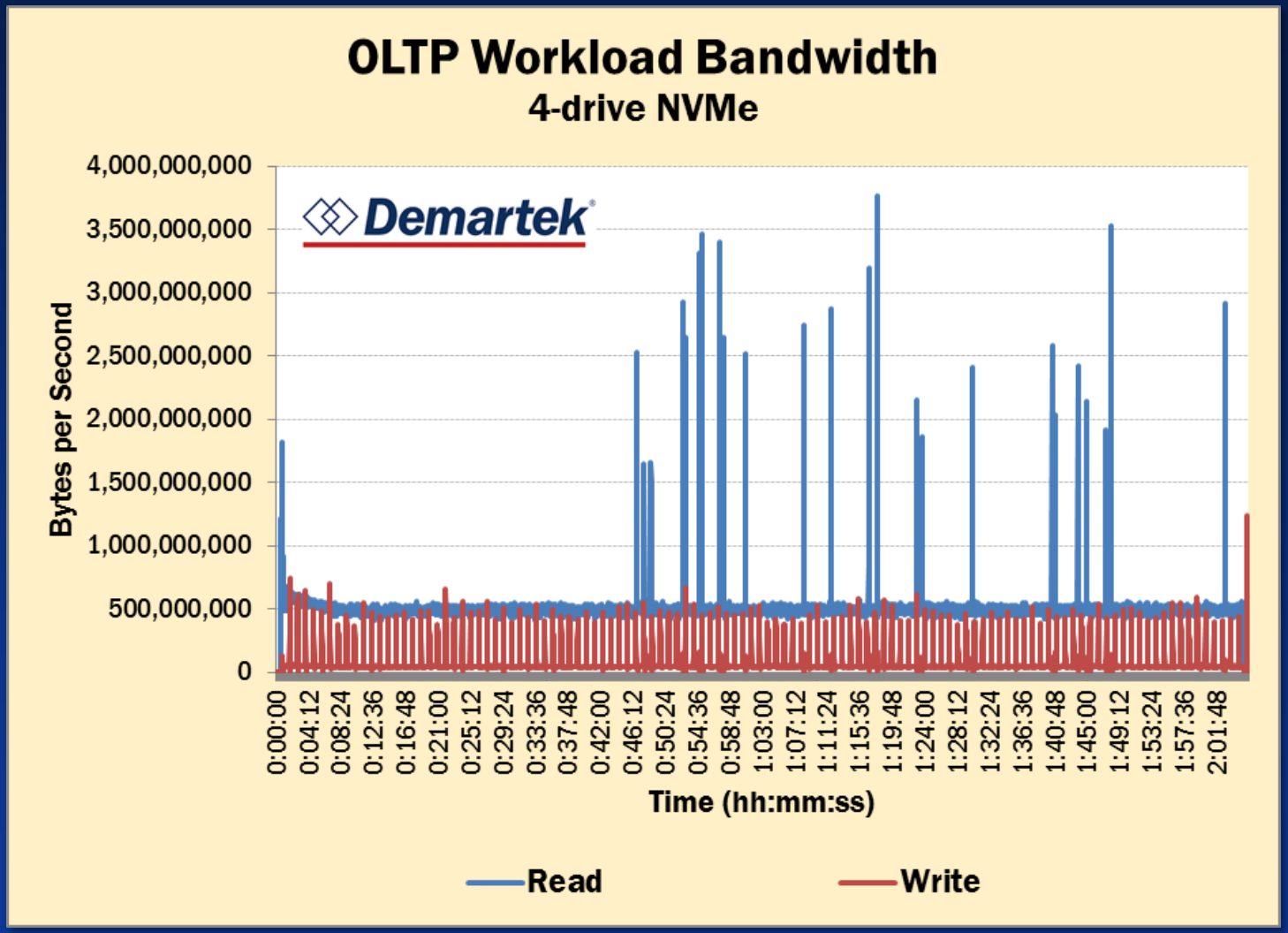
NVMe 4-drive OLTP IOPS



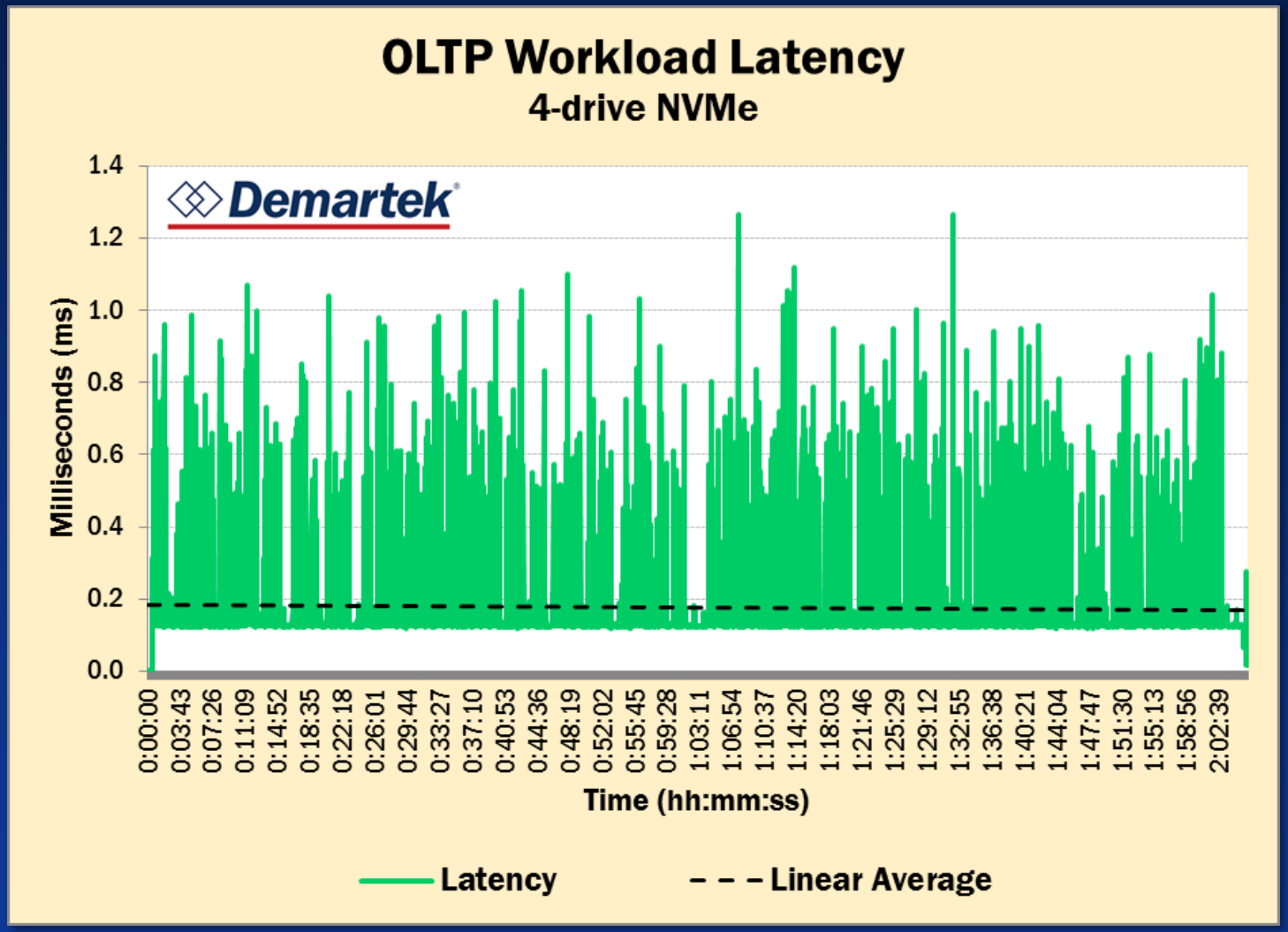
NVMe 4-drive I/O Blocksize



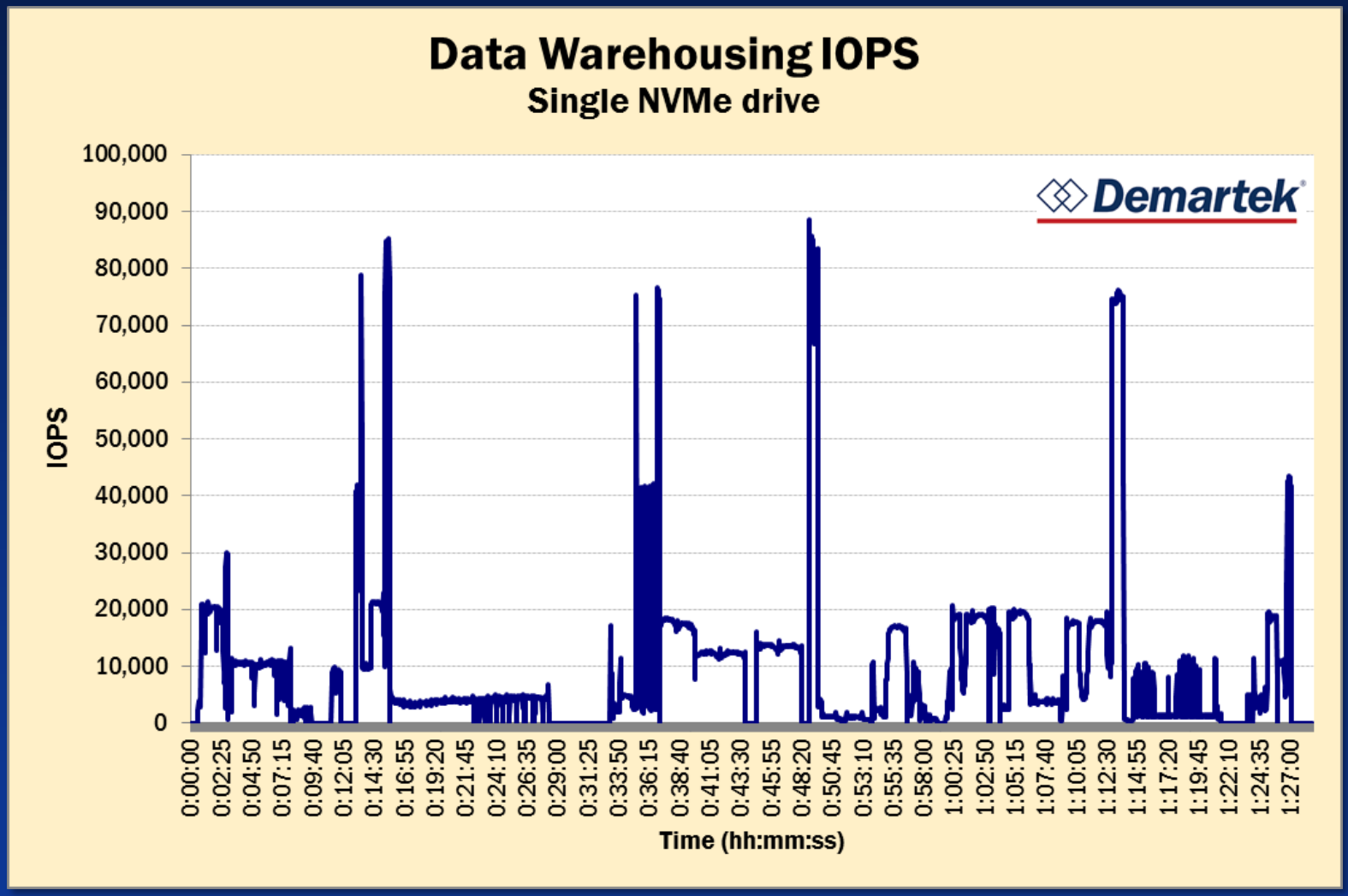
NVMe 4-drive OLTP Bandwidth



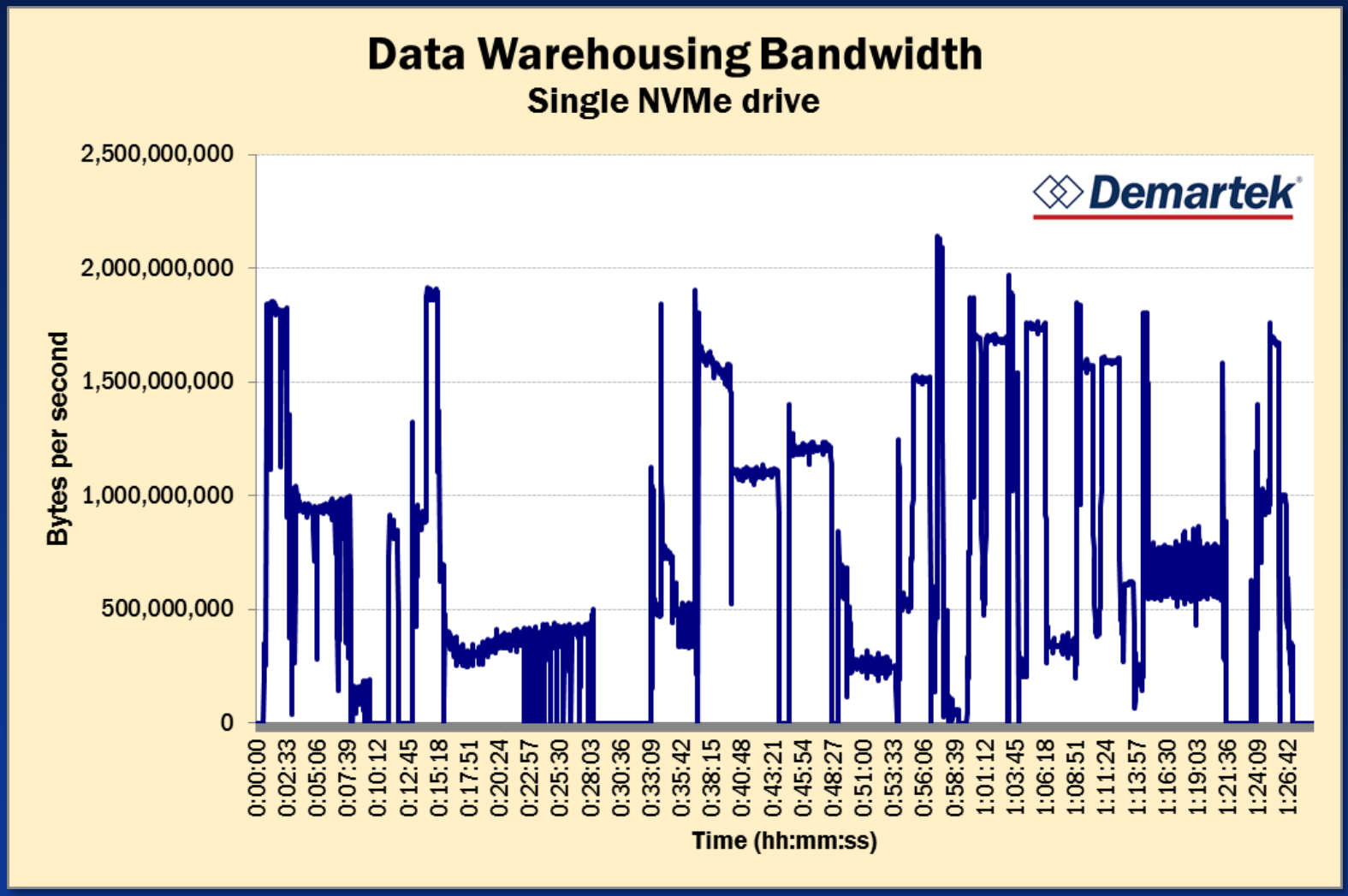
NVMe 4-drive OLTP Latency



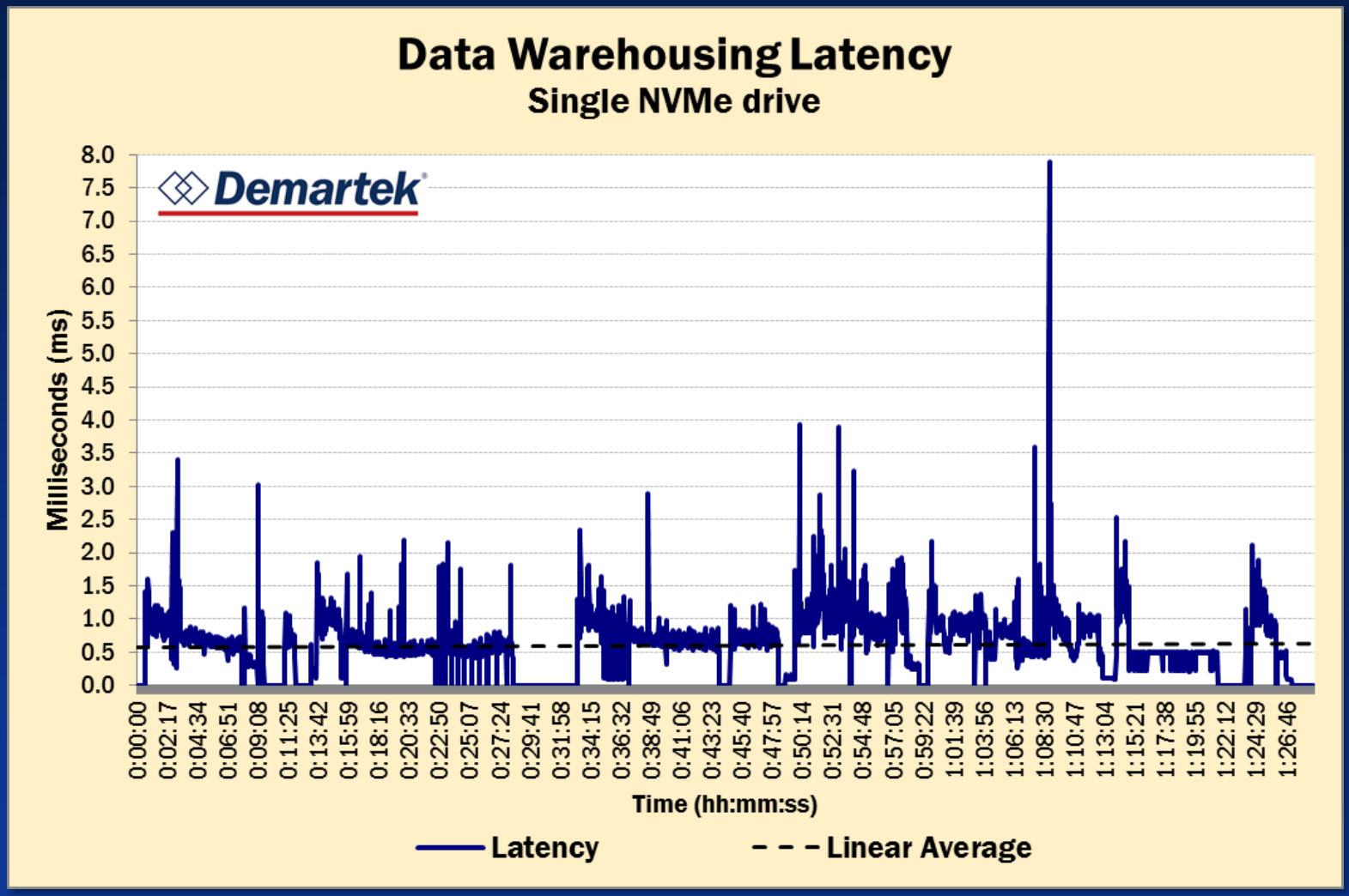
NVMe 1-drive DW IOPS



NVMe 1-drive DW Bandwidth



NVMe 1-drive DW Latency



Flash Storage Happy Side Effects 1

- Flash storage can drive up host CPU utilization
  - This will affect physical-to-virtual machine ratios
  - For some workloads, fewer servers are needed to accomplish the same amount of work when configured with flash storage
- Flash storage will expose new bottlenecks
  - Faster network and storage interfaces, including 10Gb Ethernet, 16Gb Fibre Channel, NVMe, etc. are needed
  - Flash storage and high-speed networks were made for each other

Flash Storage Happy Side Effects 2

- Flash storage matches the performance of HDD storage while consuming less power and rackspace
  - Today's drive form factor enterprise SSDs exceed the capacity of today's enterprise HDDs (10K & 15K RPM)
  - 2.5-inch SSDs are very popular today
  - Expect more dense SSD solutions in the enterprise

Flash Storage Happy Side Effects 3

- SSD marketplace is splitting into write-intensive, mixed read-write and read-intensive devices
- Expect SSDs to become the default choice for boot drives in servers and desktops
  - Will use read-intensive (lower number of write) drives
  - Makes server boot faster and apps run faster
  - It's like getting a new server or desktop and can extend the life of the server or desktop computer
  - We've been doing this since 2010 http://www.demartek.com/Demartek_SSD_production.html

Demartek Free Resources

- Demartek SSD Deployment Guide
  www.demartek.com/Demartek_SSD_Deployment_Guide.html
- Demartek commentary: "Horses, Buggies and SSDs"
  www.demartek.com/Demartek_Horses_Buggies_SSDs_Commentary.html
- Demartek comments on IDF2013 & NVMe
  www.demartek.com/Demartek_Comments_IDF2013_and_NVMe.html
- Demartek Video Library - http://www.demartek.com/Demartek_Video_Library.html
- Demartek FC Zone – www.demartek.com/FC
- Demartek iSCSI Zone – www.demartek.com/iSCSI
- Demartek SSD Zone – www.demartek.com/SSD

Performance reports, Deployment Guides and commentary available for free download.

Free Monthly Newsletter

Demartek publishes a free monthly newsletter, ***Demartek Lab Notes***, highlighting recent reports, articles and commentary.

Look for the newsletter sign-up at:
www.demartek.com/Newsletter

Thank You!

Dennis Martin, President

dennis@demartek.com

www.linkedin.com/in/dennismartin

Demartek

(303) 940-7575

www.demartek.com

http://twitter.com/Demartek

www.youtube.com/Demartek

Skype: Demartek

To learn more about Demartek:

- Download the Aurasma App (Android/iPhone)
- Search and follow "Demartek"
- View image below with viewfinder.

*also on the back of Dennis' business card

Powered by: AURASMA