SNIA PERSISTENT MEMORY PM SUMMIT

JANUARY 23, 2020 | SANTA CLARA, CA

Real World Workloads, AI & Storage Optimization

Eden Kim, CEO – Calypso Systems, Inc.

Yafei Yang, CEO – DapuStor, Inc.

I. Real World Workloads

IO Capture, AI Training Loops, Performance Comparison

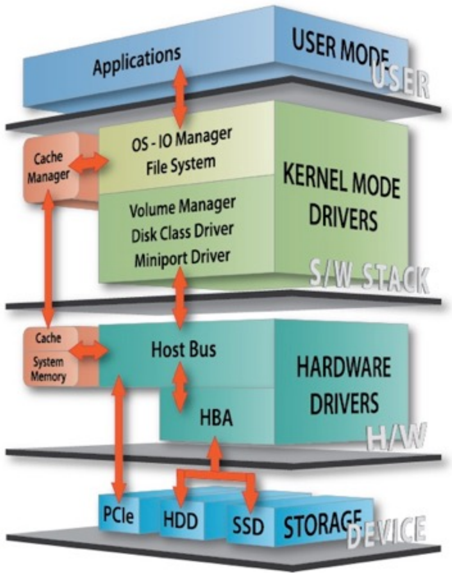
Eden Kim, CEO – Calypso Systems, Inc.

What Are Real World Workloads?

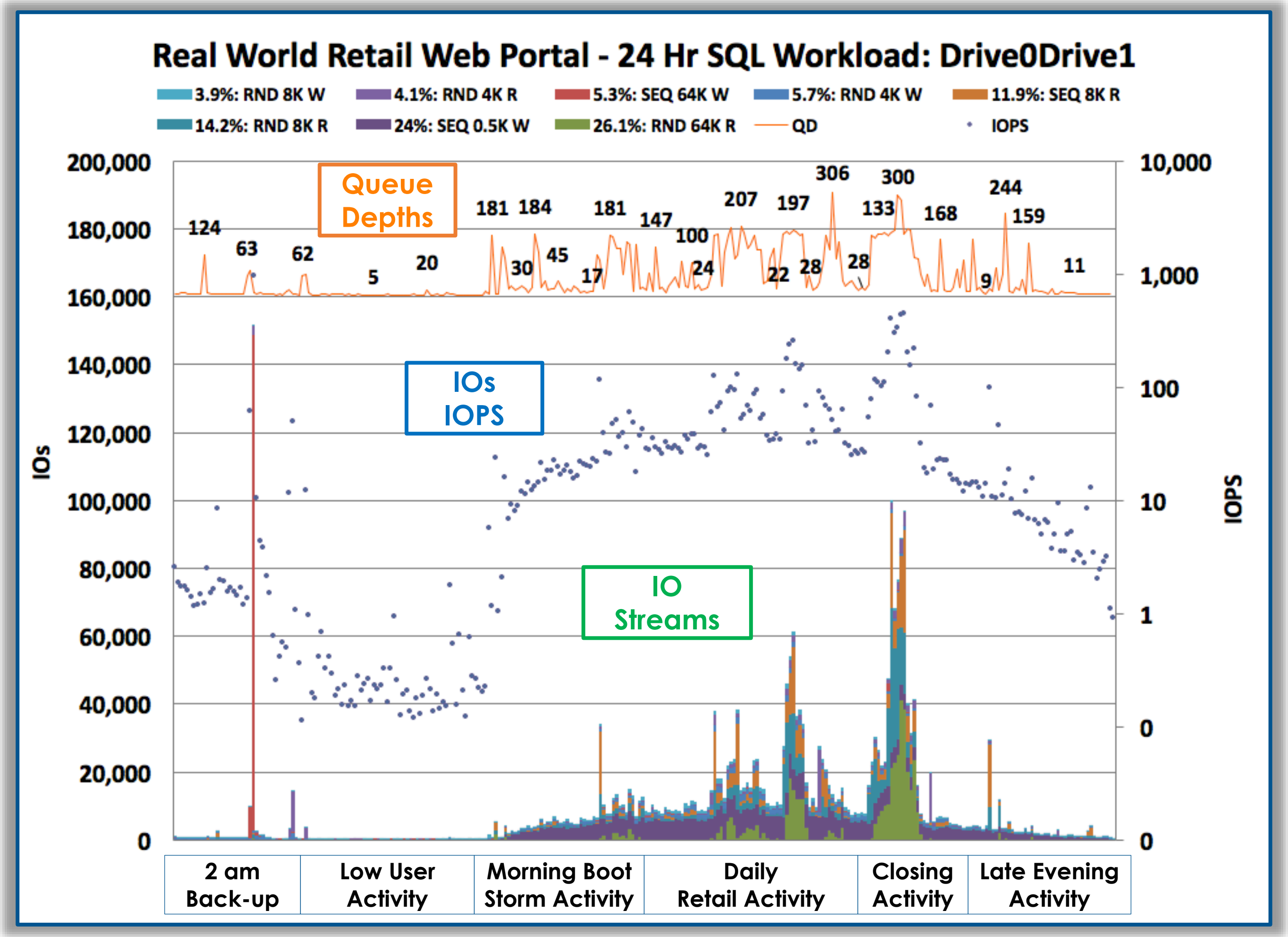
SNIA PERSISTENT MEMORY PM SUMMIT
JANUARY 23, 2020 | SANTA CLARA, CA

IO Streams that Occur During Application Use

Real World Workloads Change as they Traverse the Software/Hardware Stack from point to point



IOs are Affected at Each Layer of SW/HW Abstraction



Real World Workloads are Constantly Changing Combinations of IO Streams & QDs

Σ Cumulative Workload

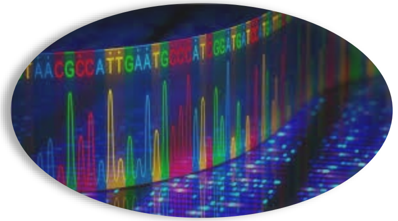
RND 64K R	18.5%	842,361
SEQ 0.5K W	17.0%	775,127
RND 8K R	10.0%	456,175
SEQ 8K R	8.4%	382,972
RND 4K W	4.0%	182,251
SEQ 64K W	3.7%	169,571
SEQ 64K R	3.4%	155,798
RND 4K R	2.92%	132,777
RND 8K W	2.78%	126,550
SEQ 4K R	2.20%	100,309

Total IOs of 5,086 streams: 4,551,062
RW mix: 66% R : 34% W
Selected 9 streams: 3,223,582 (71%) E

Changing Combinations of Thousands of IO Streams

Real World Workload Use Cases

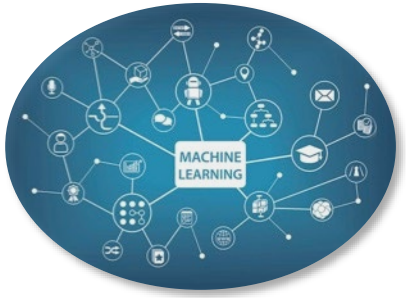
Artificial Intelligence



Genetics/Genomics

3D Animations

5G Edge Servers



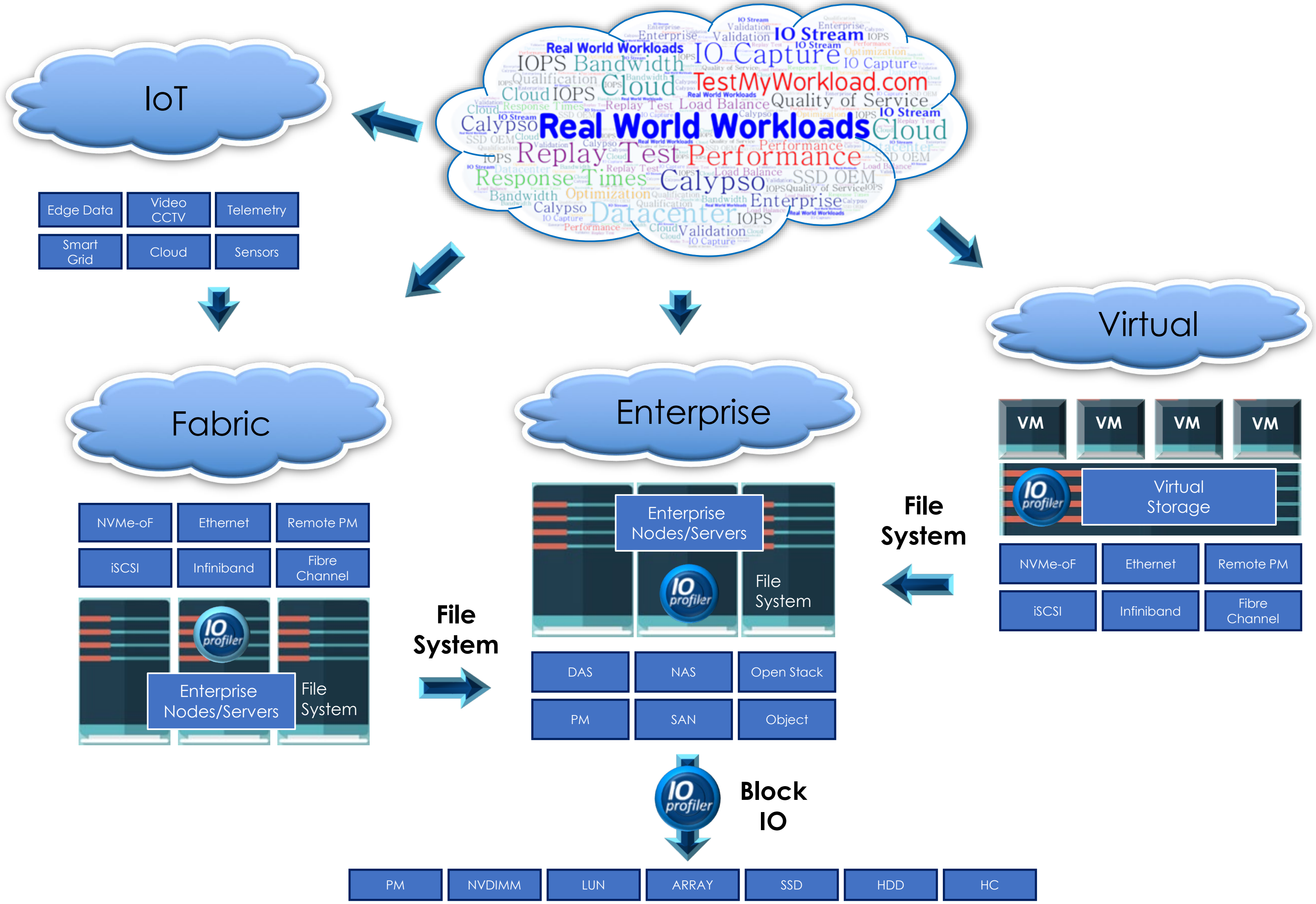
Machine Learning

Cloud Storage

Big Data

Web Portals

How & Where to Capture Workloads

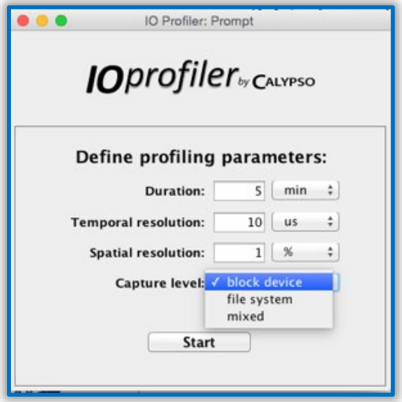


IO Profile (IPF) capture tools monitor storage server IOs at the **File System, Block IO or Virtual Layer**

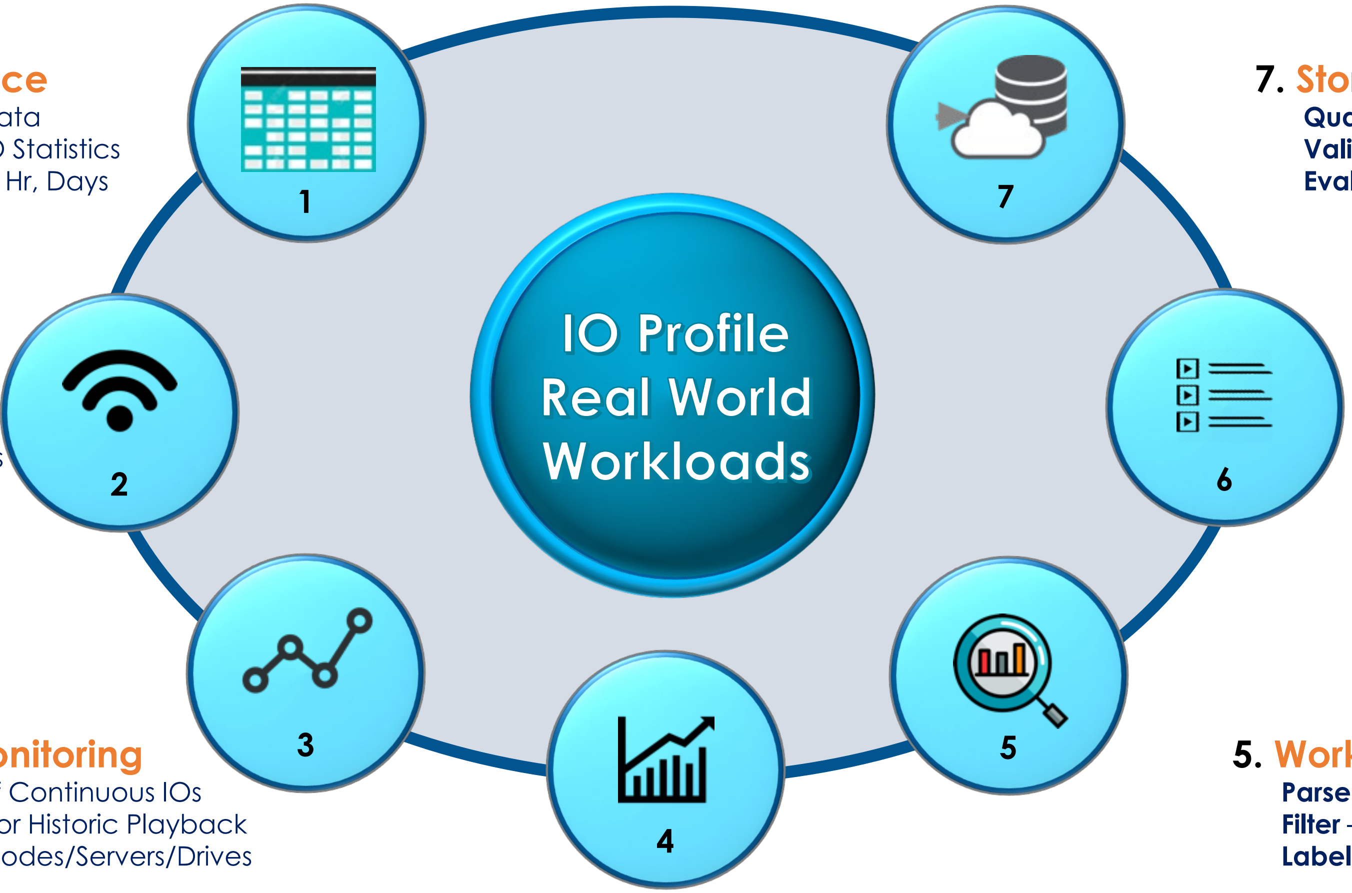
IO traffic to all logical storage recognized by the OS is captured, including Direct Attached, Remote, Fabric and/or Virtual Storage

IO Profile captures save IO Stream metrics in **tables of statistics** – no actual or personal data is saved

Free IO Profile capture tools are at www.TestMyWorkload.com



Workload Capture, Curation, Optimization



1. **IO Profile v IO Trace**
   - **Secure** - No personal data
   - **Step Data** – Tables of IO Statistics
   - **Granularity** – uS, S, Min, Hr, Days

2. **Edge Transmission**
   - **Portability** – Small file size
   - **Real Time Telemetry** – Streaming Steps
   - **Phone Home** – Automated Delivery

3. **Continuous Monitoring**
   - **View** – Database of Continuous IOs
   - **Replay** – Real Time or Historic Playback
   - **Monitor** – Multiple Nodes/Servers/Drives

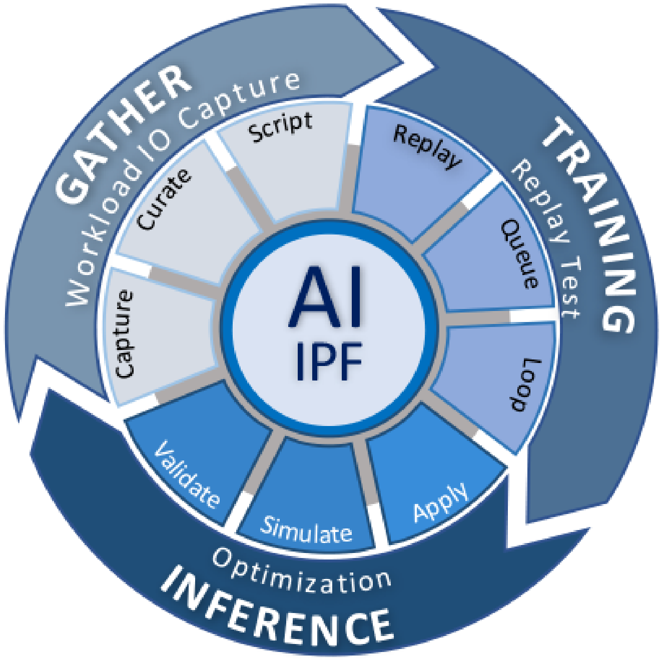
4. **Key Performance Indicators**
   - **IO Streams** – IO Stream Combinations & QDs
   - **Metrics** – PIDs, Performance, IO Bursts, IO Sequentiality
   - **Spikes** – TRIMs, Bursts, QDs, Boot Storms

5. **Workload Curation**
   - **Parse** – Time Ranges, PIDs, Storage LUNs
   - **Filter** – PIDs, Performance, LBA Ranges
   - **Label** – Structured Data, Applications

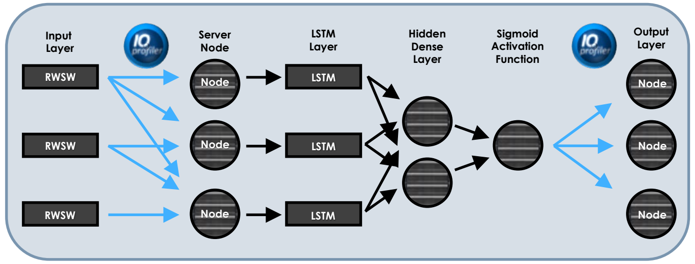
6. **Replay Training**
   - **Loop** – AI Training Loops
   - **Validate** – Continuous Data Check
   - **Infer** – Optimized Performance

7. **Storage Optimization**
   - **Qualification** – Storage Benchmark Performance
   - **Validation** – Load Balance, Bottleneck Check
   - **Evaluation** – Infrastructure Storage Selection

AI & IPF Workloads: Gather, Training, Inference



GATHER



Capture – Edge Data IO Input Layer
Curate – Key Performance Indicators
Script – Create Replay Script

TRAINING

Replay – Replay Test
Queue – Multiple Test List
Loop – Repeat Replay Test

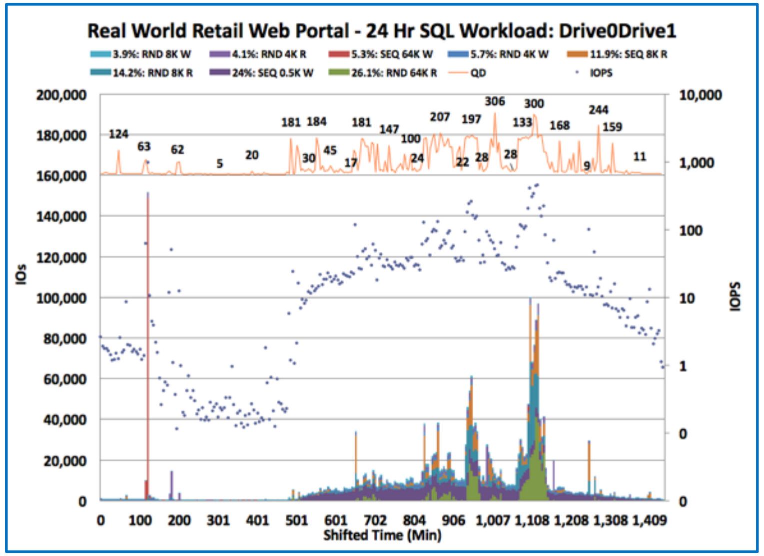
INFERENCE

Apply – Run LSTM RNN AI Algorithms
Simulate – Optimize Performance Gains
Validate - Continuous Delivery & Integration
Monitor with Real Time Telemetry

Comparing Workload IO Stream Maps

SNIA PERSISTENT MEMORY PM SUMMIT
JANUARY 23, 2020 | SANTA CLARA, CA

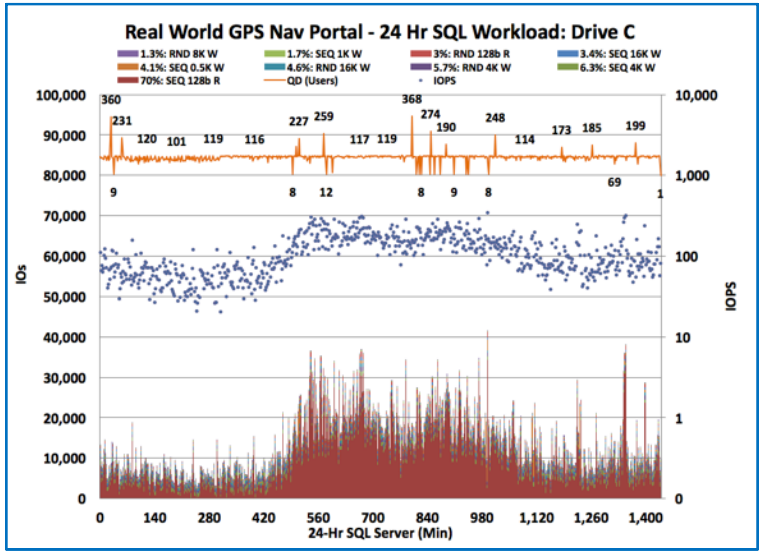
1. Retail Web Portal:
Drive0 File System/Boot;Drive1 Storage



9 / 5,038 IO Streams = 71%
66:34 RW – Mixed IO Streams
Retail Store Events

Ave QD = 22
Median QD = 19
Max QD = 306

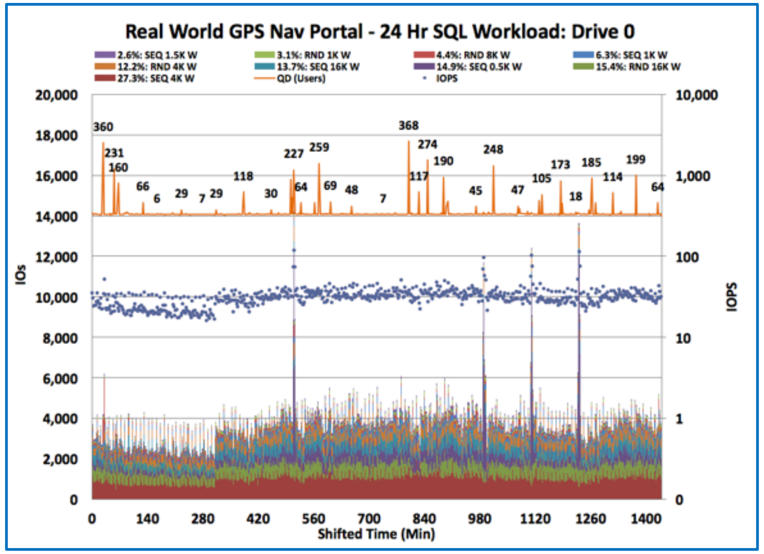
2. GPS Nav Portal:
Drive C - File System/Boot



9 / 7,169 IO Streams = 86%
67:33 RW – SEQ 128b Dominant
SEQ 128b R = 70% IOs

Ave QD = 114
Median QD = 115
Max QD = 368

3. GPS Nav Portal:
Drive 0 - Storage



9 / 1,033 IO Streams = 78%
96% W – Uniform Mix IO Streams
SEQ 0.5K W – 12K IOPS Spikes

Ave QD = 15
Median QD = 8
Max QD = 368

Key Performance Indicators include:

IO Stream Consistency — IO/RT Spikes — IOPS / Throughput — IO Bursts/Clusters — IO Ranges — RT Quality of Service

Transfer Sizes — RW Mixes — QD Ranges — TRIMs — LBA Ranges — Dedupe Ratio

Labels — Metadata — Process IDs — Encryption — Compression Ratio

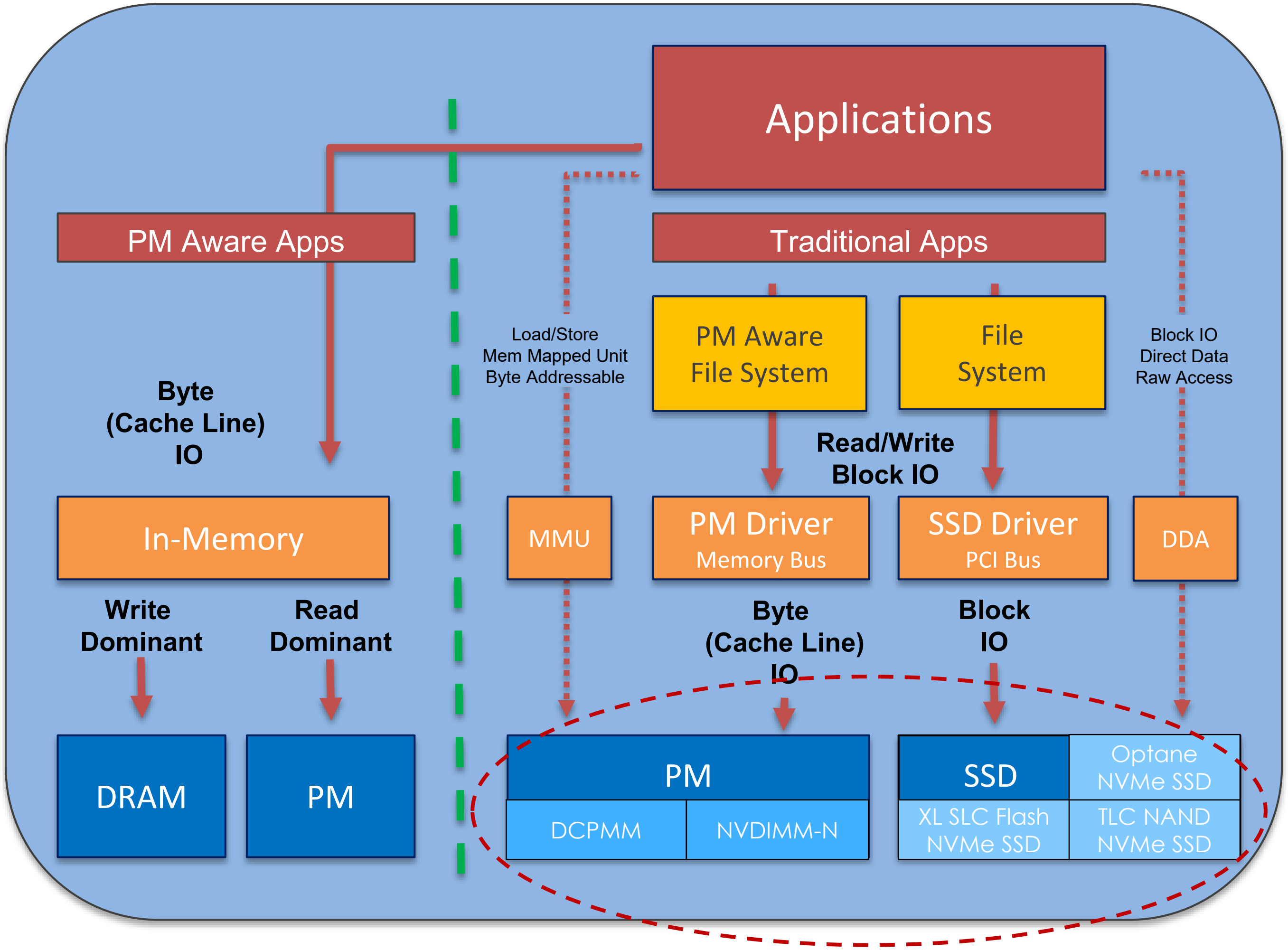
Reference Workloads: Synthetic & Real World

PERSISTENT MEMORY PM SUMMIT
JANUARY 23, 2020 | SANTA CLARA, CA

Reference Workloads								
Application	R/W Mix	General Description	IO Streams	Min / Median / Max QD	IO Step Resolution & Duration	Drives	Main Block Sizes	Access
Synthetic Database	65% R	Synthetic DB/OLTP RND 8K 65% R Performance Saturation over DI Range	2	Range: 1-1024	60s/360 step/6 hr	1	8 K	100% RND
Retail Web Portal	64% R	2,000 Outlet Retail Web Portal; WinServer 2012R; SQL Server; Retail Activities	5,088	1 - 23 - 306	300s/288 step/24 hr	2	64K, 0.5K 8K, 4K	68% RND
GPS Nav Portal	96% W	GPS Navigation Portal; 53 GB SAS HDD; Win 2008; NTFS; SQL Server; GPS Activity	1,033	5 - 8 - 368	120s/720 step/24 hr	1	4K, 0.5K 16K, 1K	49% RND
DC VDI Storage	75% R	Datacenter VDI Storage; 6 Fiber Channel SSDs; 28 GB total; Aggregated single capture	1,223	47 - 62 - 79	300s/158 steps/13 hr	6	4K; 32K 128K; 8K	37% RND
VOD Edge Server	60% R	VOD Edge Server Storage; Single 500GB SCSI SSD; Windows OS	1,754	1 - 3 - 42	3s/34,364 step/24 hr	1	4K, 64K, 128K, 8K	48% RND

Note: Reference Real World Workloads are posted at **www.TestMyWorkload.com**

Performance Comparison



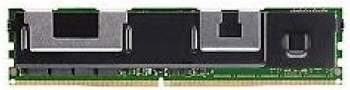
Test Sample Pool:

Intel Optane SSD, DapuStor XL SLC Flash SSD,
Huawei TLC NVMe SSD,
Intel Optane DCPMM, SMART Modular NVDIMM-N

XL SLC Flash NVMe SSD - DapuStor

3D X Point Optane NVMe SSD - Intel

TLC NAND NVMe SSD - Huawei



Optane 3D X Point PM – Intel DCPMM

NVDIMM-N PM – SMART Modular

Test Plan, Workloads & General Observations

Performance Depends on the Workload

PERSISTENT MEMORY PM SUMMIT
JANUARY 23, 2020 | SANTA CLARA, CA

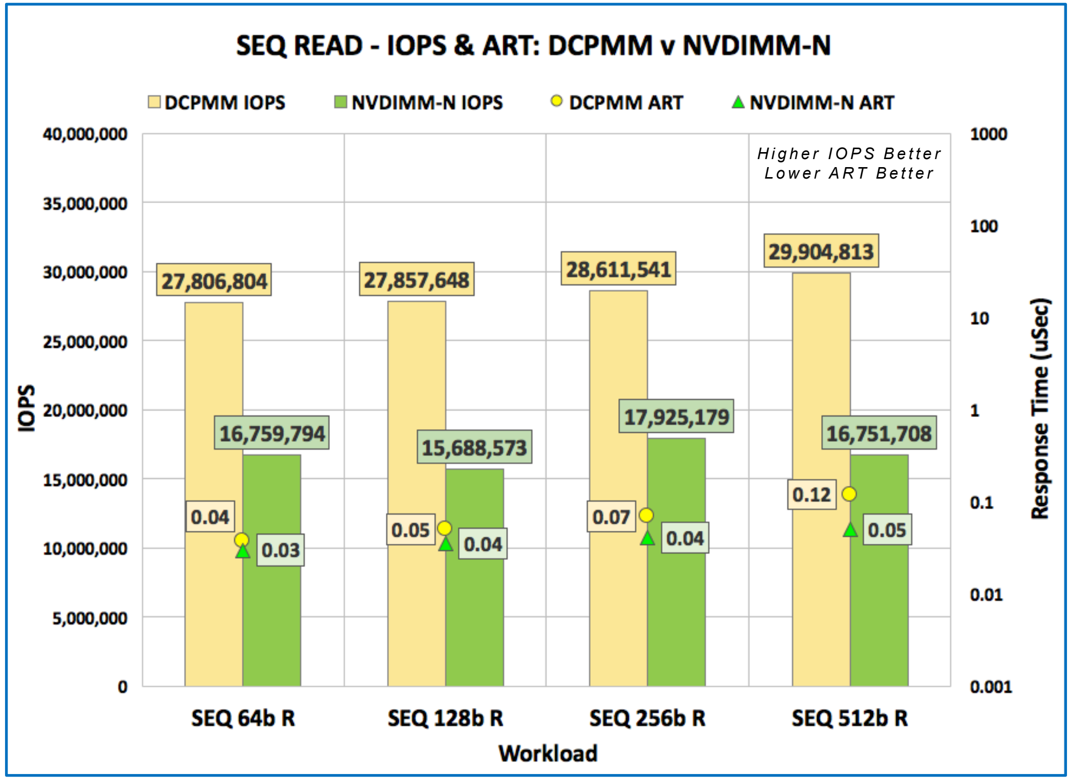
Test Phase	Description	Workloads	Persistent Memory [1]		NVMe SSD [2]		
			DCPMM	NVDIMM-N	XL SLC Flash [3]	3D X Pt	TLC 3D NAND
Synthetic Corner Case Benchmark	SNIA PM PTS [4] Draft	RND/SEQ R: 64b,128b,256b,512b IOPS/ART	Faster SEQ R	Slower SEQ R			
		RND/SEQ W: 64b,128b,256b,512b IOPS/ART	Slower RND W	Faster RND W			
	SNIA SSD PTS [5] v2.0.1	IOPS: RND R/W: 0.5K – 1024K IOPS			1st RND 4K R IOPS	1st RND 4K W IOPS	2nd RND 4K RW IOPS
		Bandwidth: SEQ 128K R/W MB/s			2nd SEQ 128K R	2nd SEQ 128K W	1st SEQ 128K R/W
	XL Flash [6] 4K R Latency	Latency: RND/SEQ R/W: 0.5K, 2K, 4K ART (uS)			2nd RND 4K R LAT	1st RND 4K R LAT	3rd RND 4K R LAT

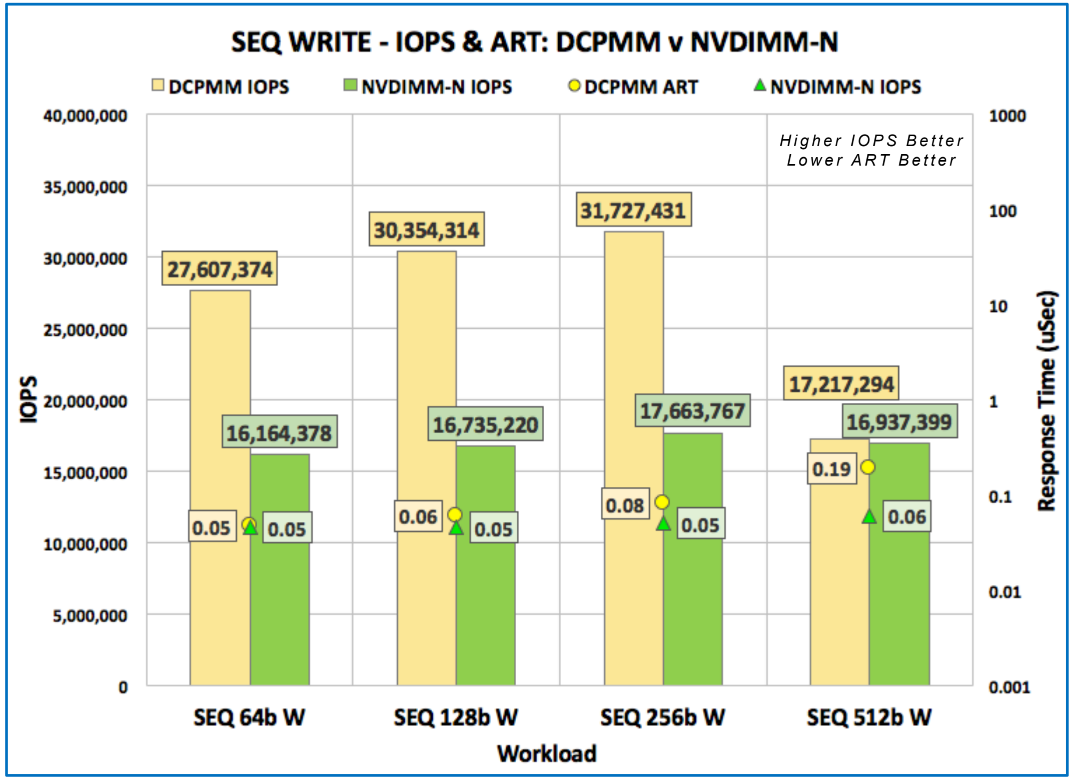
Test Phase	Description	Workload	Blk Sizes	Access	RW Mix	DCPMM IOPS	DCPMM ART	NVDIMM-N IOPS	NVDIMM-N ART	XL SLC Flash	3D X Pt IOPS	3D X Pt ART	TLC 3D NAND IOPS	TLC 3D NAND ART
Synthetic Demand Intensity (DIRTH) [7]	Fixed IO Streams over Range of Demand Intensity (QDs)	DB OLTP	8K	RND	65% R	2	2	1	1		1	1	2	2
		DC VDI Storage	4K; 32K 128K; 8K	52% SEQ	75% R	1	2	2	1		2	1	1	2
		Retail Web	64K, 0.5K 8K, 4K	68% RND	66% R	1	2	2	1		2	1	1	2
		GPS Nav	4K, 0.5K 16K, 1K	51% SEQ	96% W	2	2	1	1		1	1	2	2
		VOD Edge	4K, 64K, 128K, 8K	52% SEQ	60% R	2	2	1	1		1	1	2	2
Real World Workload [7] Replay	Sequence of IO Streams & QDs from the Original IO Capture	DC VDI Storage	4K; 32K 128K; 8K	52% SEQ	75% R	1	2	2	1		2	2	1	1
		Retail Web	64K, 0.5K 8K, 4K	68% RND	66% R	2	2	1	1		2	2	1	1
		GPS Nav	4K, 0.5K 16K, 1K	51% SEQ	96% W	2	2	1	1		2	2	1	1
		VOD Edge	4K, 64K, 128K, 8K	52% SEQ	60% R	2	2	1	1		1	1	2	2

[1] PM DAX FS & Mmap; Intel PM RTP - x6 256GB Interleaved DCPMM; SMART NVDIMM-N RTP - x1 16GB NVDIMM-N; [2] Intel 3D X Pt 750GB Optane AIC NVMe SSD, DapuStor 1600GB AIC XL SLC Flash NVMe SSD, Huawei TLC V3 3200GB NVMe SSD;
[3] XL Flash Pre Production – no Synthetic or Real World tests; [4] SNIA PM PTS Draft 0.3; [5] SNIA PTS v2.0.1; [6 7] SNIA RWSW PTS v1.0.7; Ordinal Rank: DCPMM v NVDIMM-N & 3D X Pt v TLC NVMe; Reference Workloads posted at www.TestMyWorkload.com

PM SEQ R/W Benchmark – Small Block Corner Case: DCPMM v NVDIMM-N

SEQ R / SEQ W: 64b, 128b, 256b, 512b

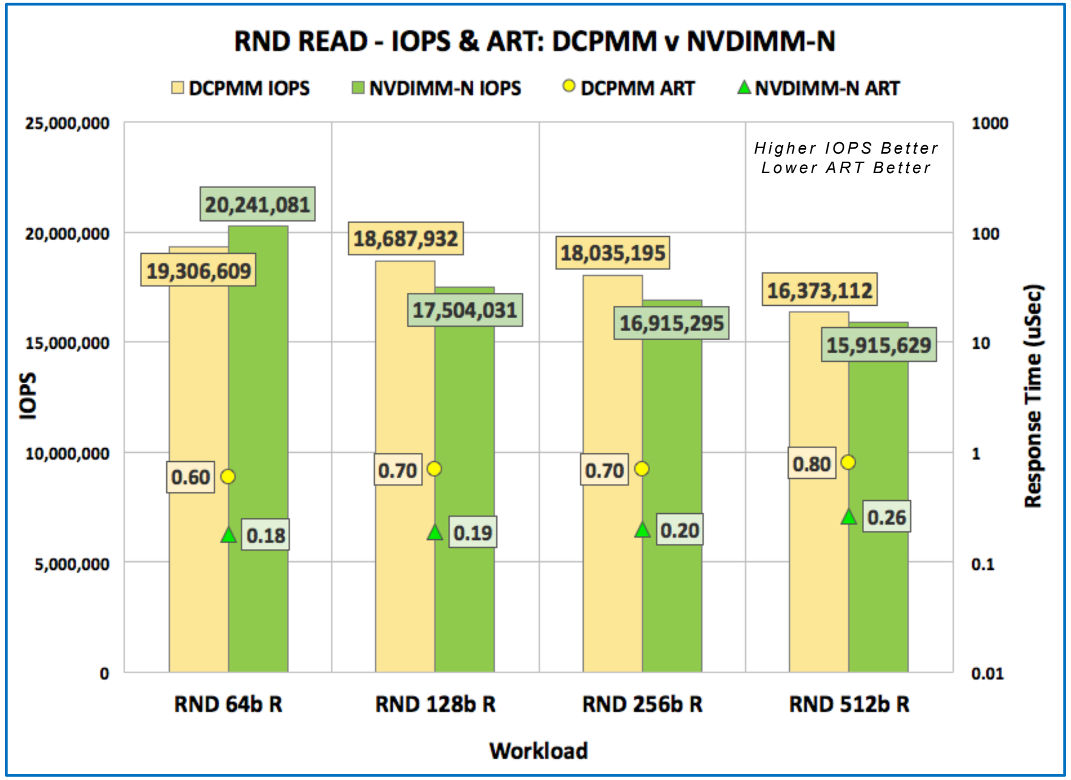


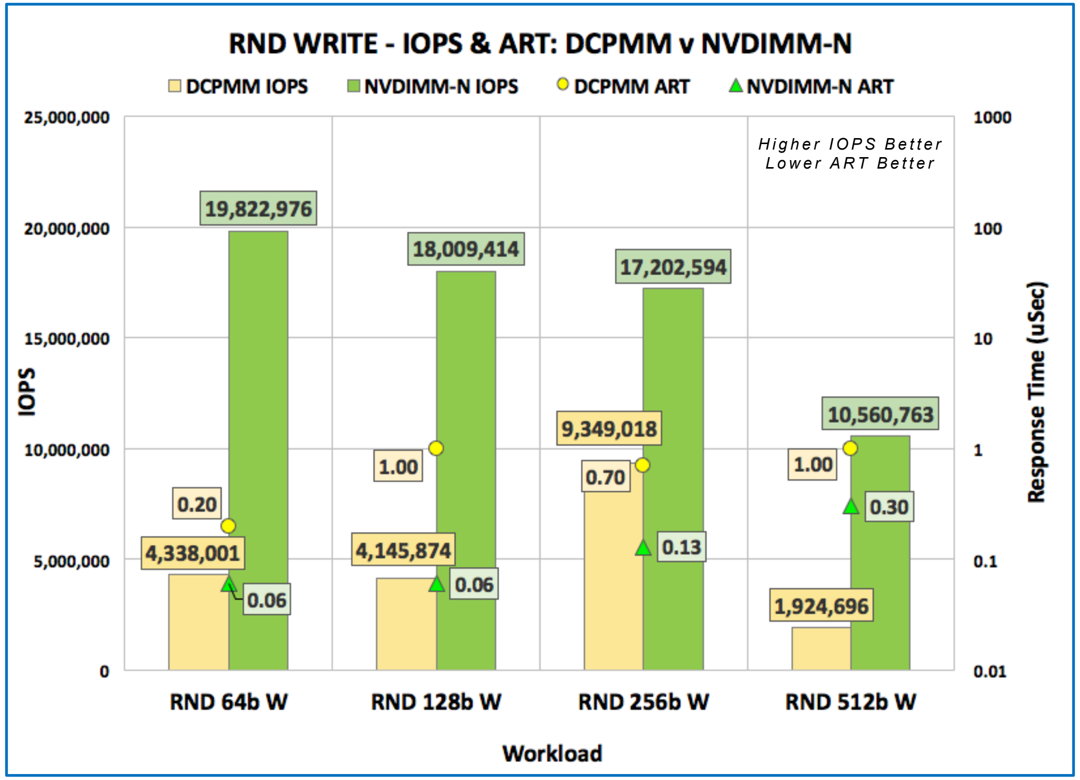


ORDINAL RANK	RND 64b R	RND 128b R	RND 256b R	RND 512b R
IOPS – DCPMM	1	1	1	1
IOPS - NVDIMM-N	2	2	2	2
ART - DCPMM	2	2	2	2
ART - NVDIMM-N	1	1	1	1

ORDINAL RANK	RND 64b W	RND 128b W	RND 256b W	RND 512b W
IOPS – DCPMM	1	1	1	1
IOPS -NVDIMM-N	2	2	2	2
ART - DCPMM	1	2	2	2
ART - NVDIMM-N	1	1	1	1

PM RND R/W Benchmark – Small Block Corner Case: DCPMM v NVDIMM-N RND R / RND W: 64b, 128b, 256b, 512b



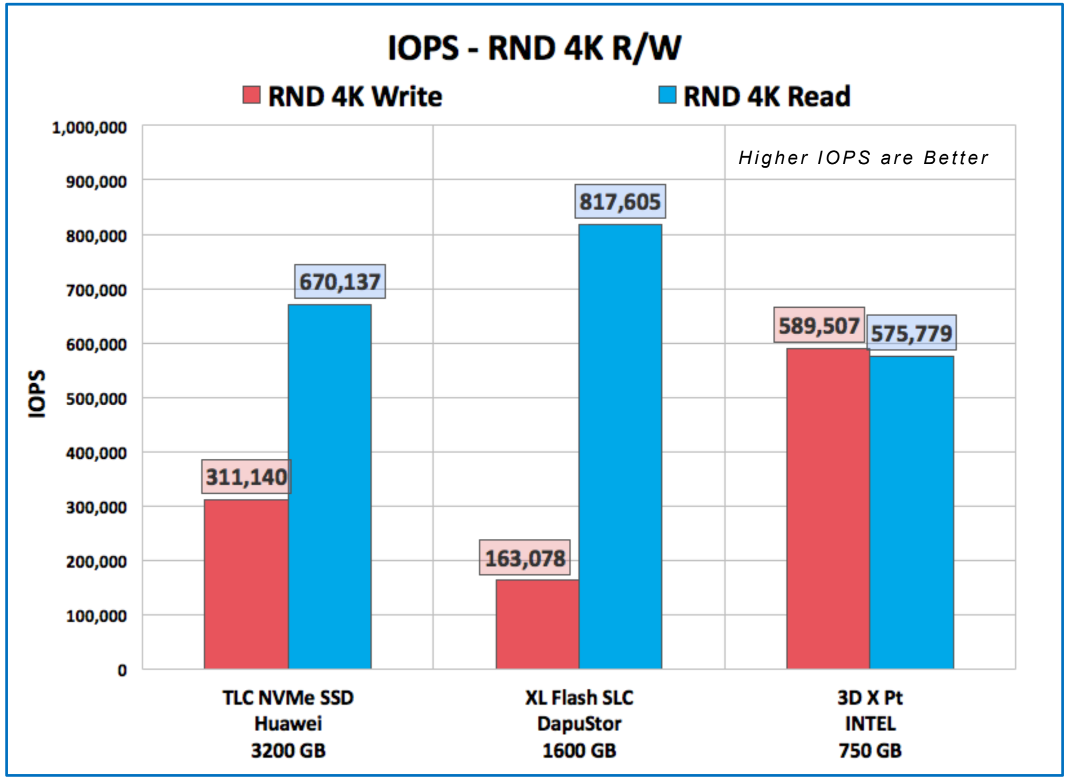


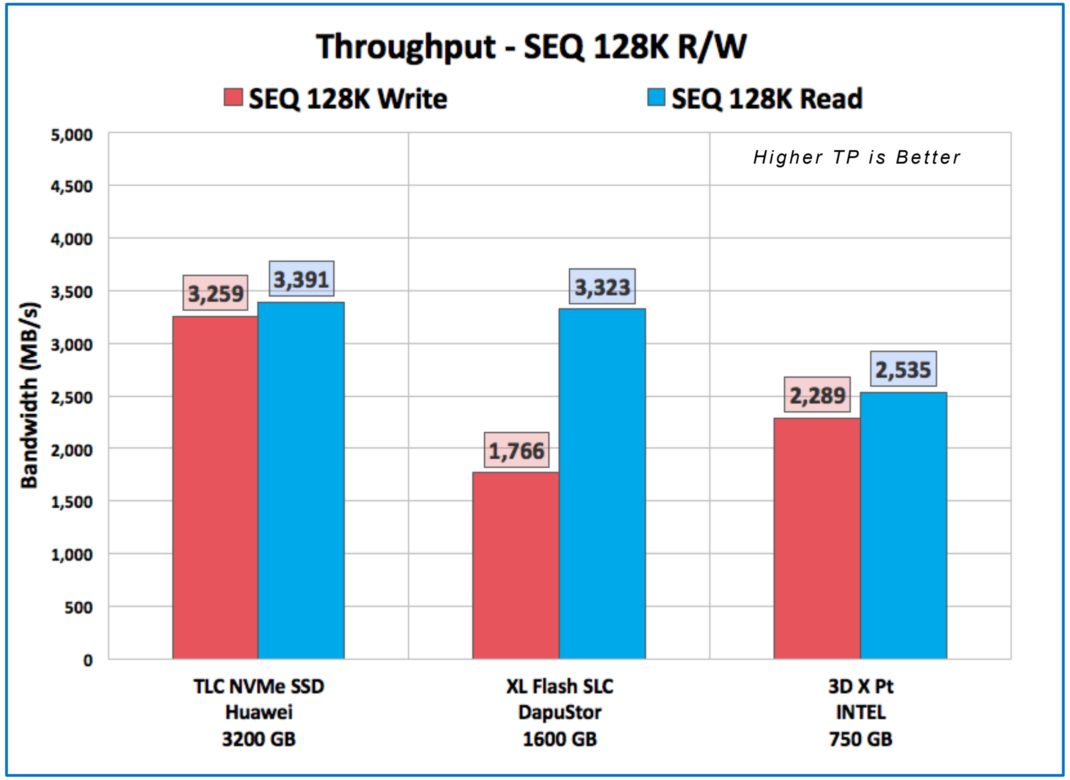
ORDINAL RANK	RND 64b R	RND 128b R	RND 256b R	RND 512b R
IOPS – DCPMM	2	1	1	1
IOPS - NVDIMM-N	1	2	2	2
ART - DCPMM	2	2	2	2
ART - NVDIMM-N	1	1	1	1

ORDINAL RANK	RND 64b W	RND 128b W	RND 256b W	RND 512b W
IOPS – DCPMM	2	2	2	2
IOPS -NVDIMM-N	1	1	1	1
ART - DCPMM	2	2	2	2
ART - NVDIMM-N	1	1	1	1

SSD PTS: IOPS/TP – TLC NVMe SSD v XL SLC Flash v 3D X Pt Optane

RND 4K IOPS (Mixed BS/RW Loop) & SEQ 128K Bandwidth



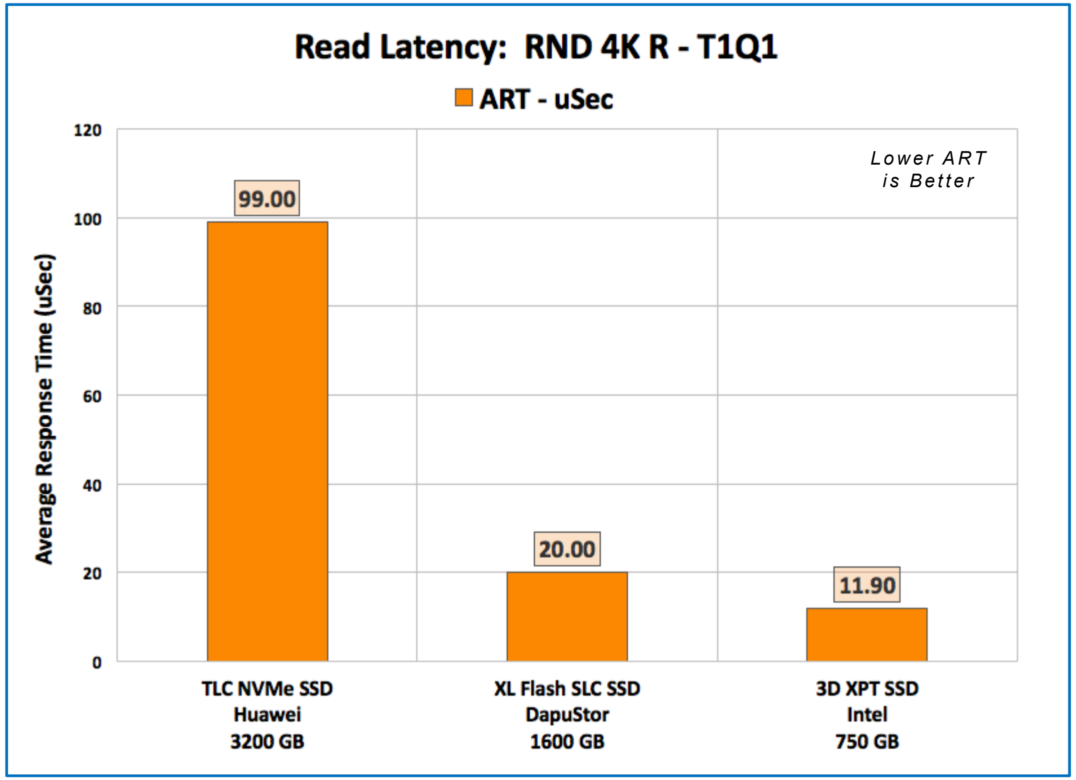


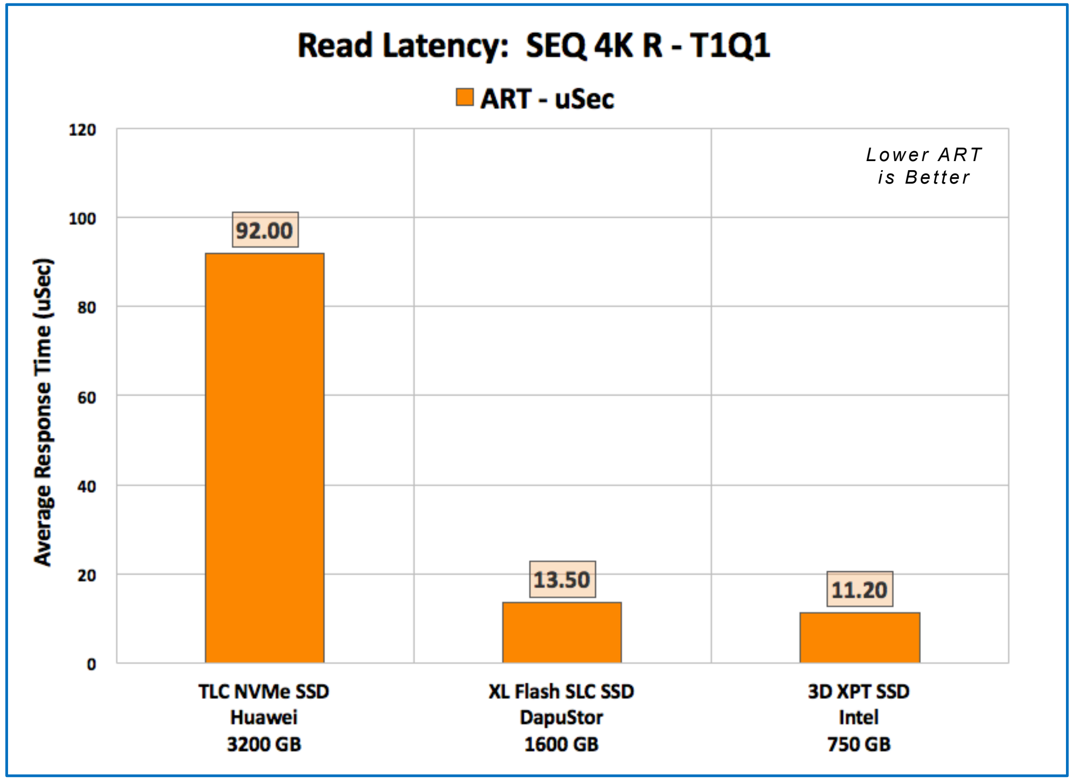
ORDINAL RANK	RND 4K W IOPS	RND 4K R IOPS
TLC NVMe SSD	2	2
XL Flash SLC SSD	3	1
3D X Pt Optane SSD	1	3

ORDINAL RANK	SEQ 128K W MB/s	SEQ 128K R MB/s
TLC NVMe SSD	1	1
XL Flash SLC SSD	3	2
3D X Pt Optane SSD	2	3

SSD PTS: Read LAT – TLC NVMe SSD v XL SLC Flash v 3D X Pt Optane

PTS Latency Test - RND/SEQ 4K Read, T1Q1



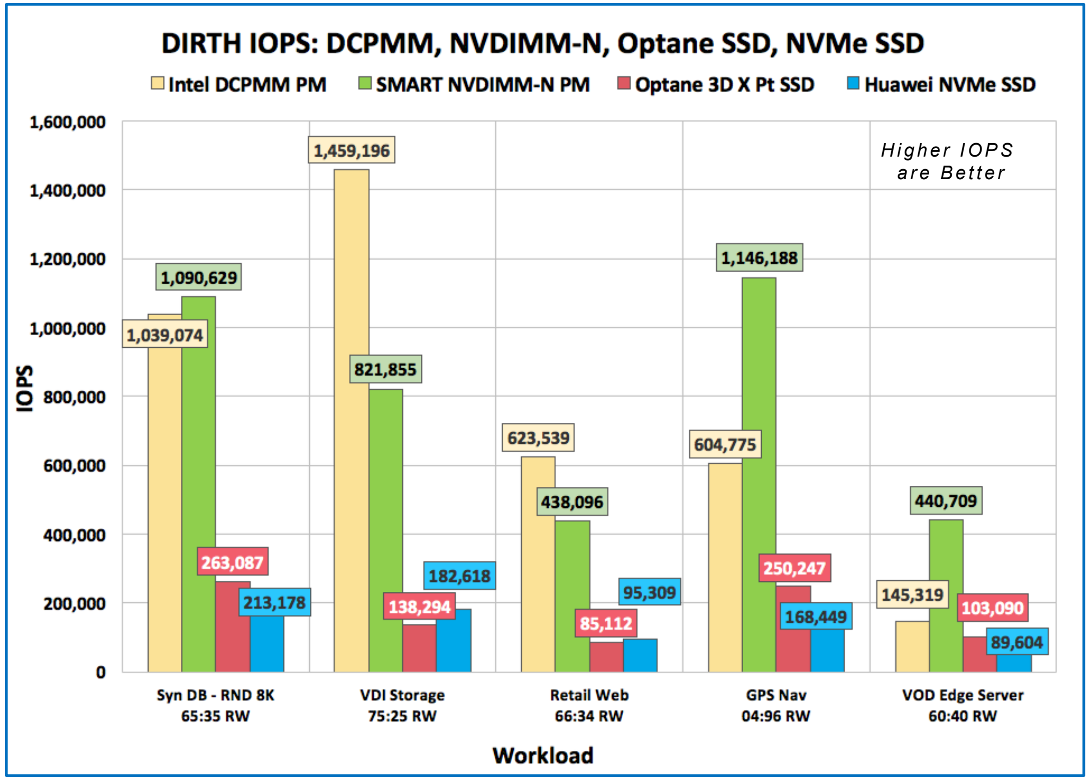


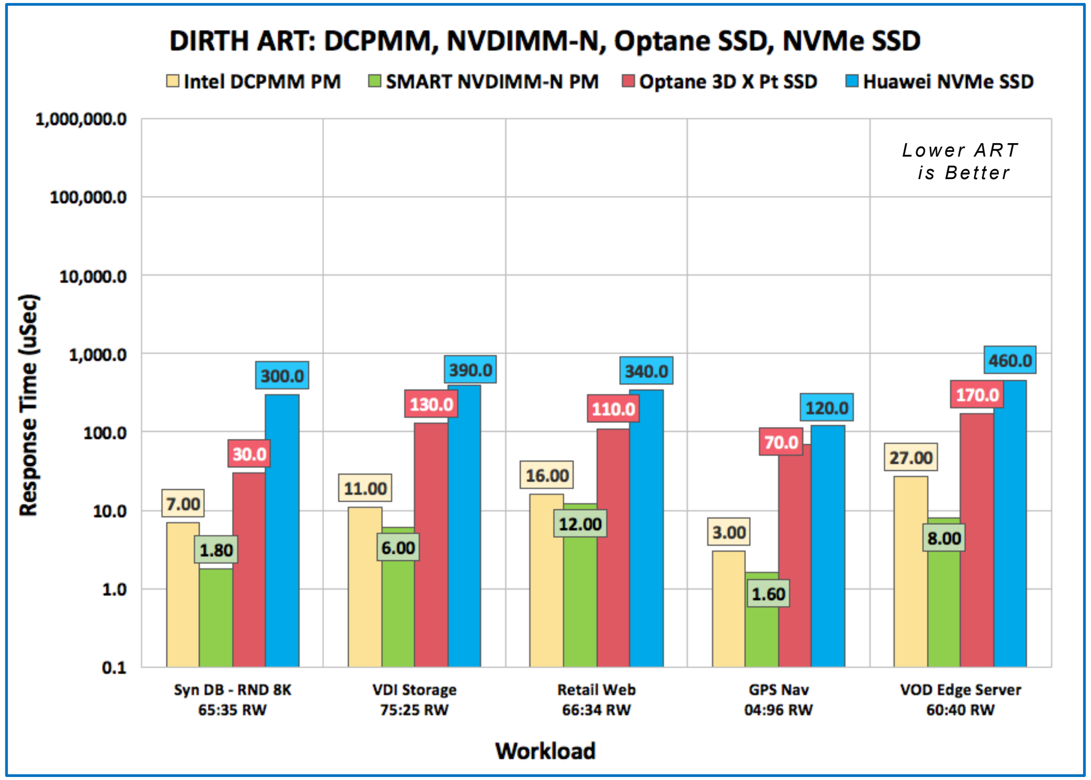
ORDINAL RANK	RND 4K R Ave RT
TLC NVMe SSD	3
XL Flash SLC SSD	2
3D X Pt Optane SSD	1

ORDINAL RANK	SEQ 4K R Ave RT
TLC NVMe SSD	3
XL Flash SLC SSD	2
3D X Pt Optane SSD	1

Synthetic DIRTH – Fixed IO Stream Combinations over a Range of Demand Intensity

OLTP/DB; DC VDI Storage; Retail Web Portal; GPS Nav Portal; VOD Edge Server





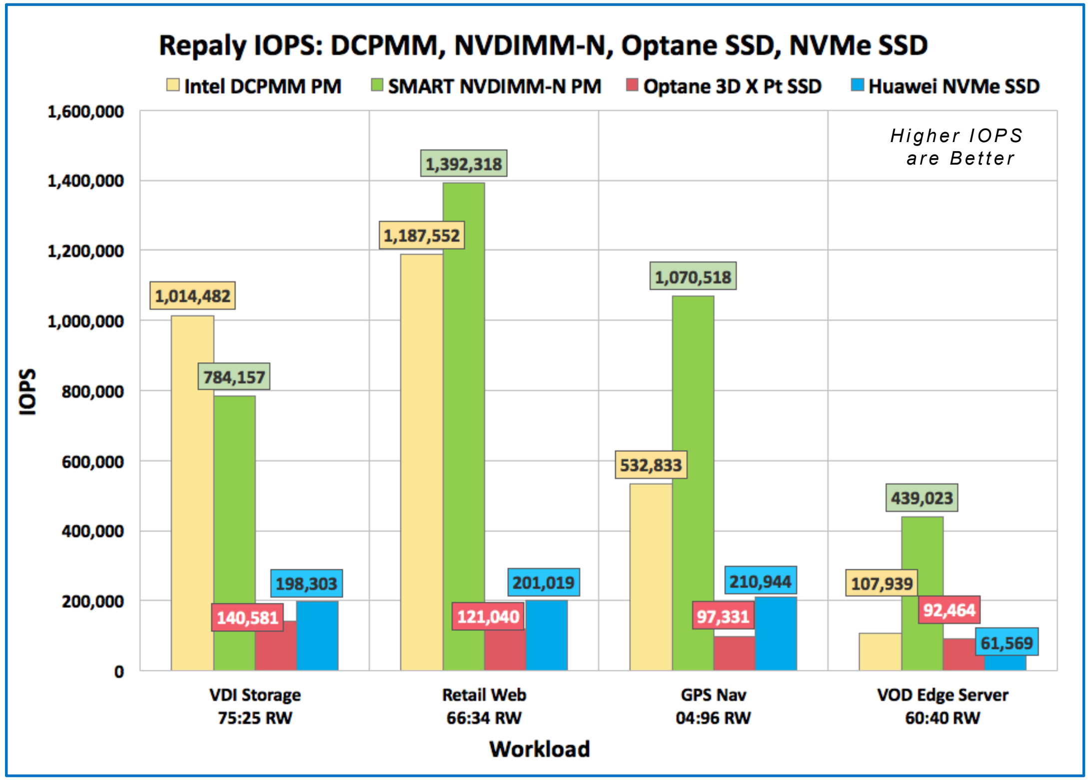
ORD RANK - IOPS	Synthetic DB	VDI Storage	Retail Web Portal	GPS Nav Portal	VOD Edge Server
DCPMM - PM	2	1	1	2	2
NVDIMM-N - PM	1	2	2	1	1
3D XPT NVMe - SSD	1	2	2	1	1
TLC NVMe - SSD	2	1	1	2	2

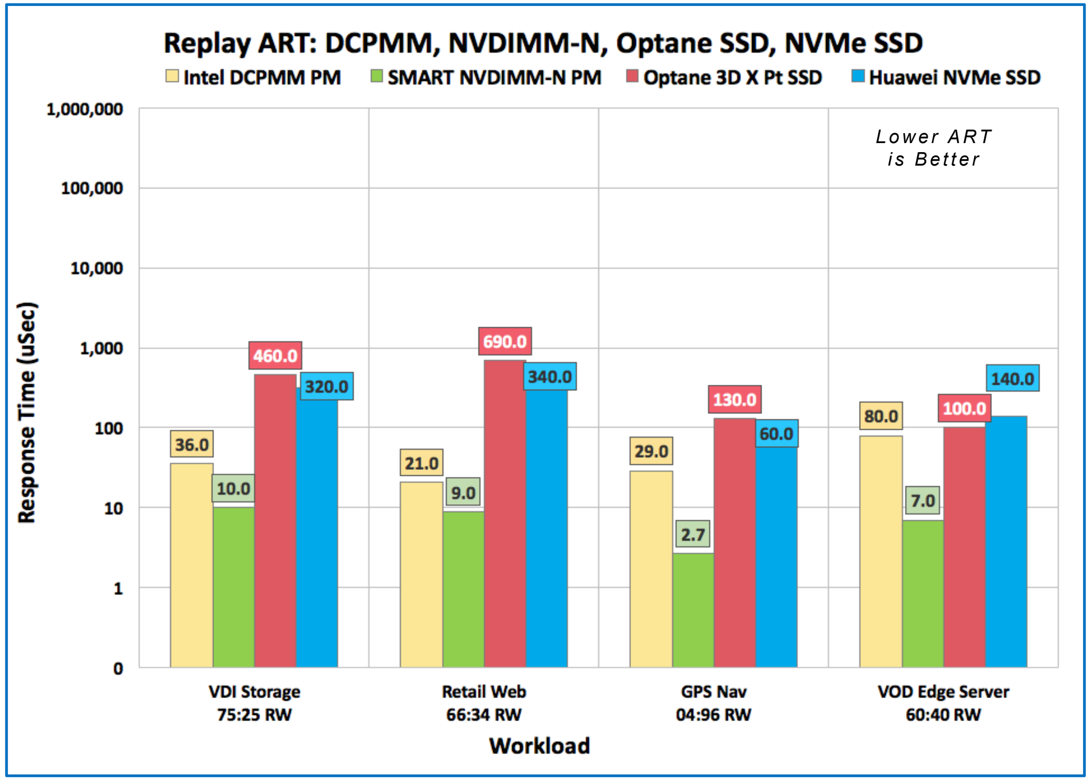
ORD RANK - ART	Synthetic DB	VDI Storage	Retail Web Portal	GPS Nav Portal	VOD Edge Server
DCPMM - PM	2	2	2	2	2
NVDIMM-N - PM	1	1	1	1	1
3D XPT NVMe - SSD	1	1	1	1	1
TLC NVMe - SSD	2	2	2	2	2

Real World REPLAY – Sequence of IO Stream Combinations and QDs from the Original IO Capture

DC VDI Storage; Retail Web Portal; GPS Nav Portal; VOD Edge Server

PERSISTENT MEMORY PM SUMMIT
SNIA
JANUARY 23, 2020 | SANTA CLARA, CA





ORD RANK - IOPS	VDI Storage	Retail Web Portal	GPS Nav Portal	VOD Edge Server
DCPMM - PM	1	2	2	2
NVDIMM-N - PM	2	1	1	1
3D XPT NVMe - SSD	2	2	2	1
TLC NVMe - SSD	1	1	1	2

ORD RANK - ART	VDI Storage	Retail Web Portal	GPS Nav Portal	VOD Edge Server
DCPMM - PM	2	2	2	2
NVDIMM-N - PM	1	1	1	1
3D XPT NVMe - SSD	2	2	2	1
TLC NVMe - SSD	1	1	1	2

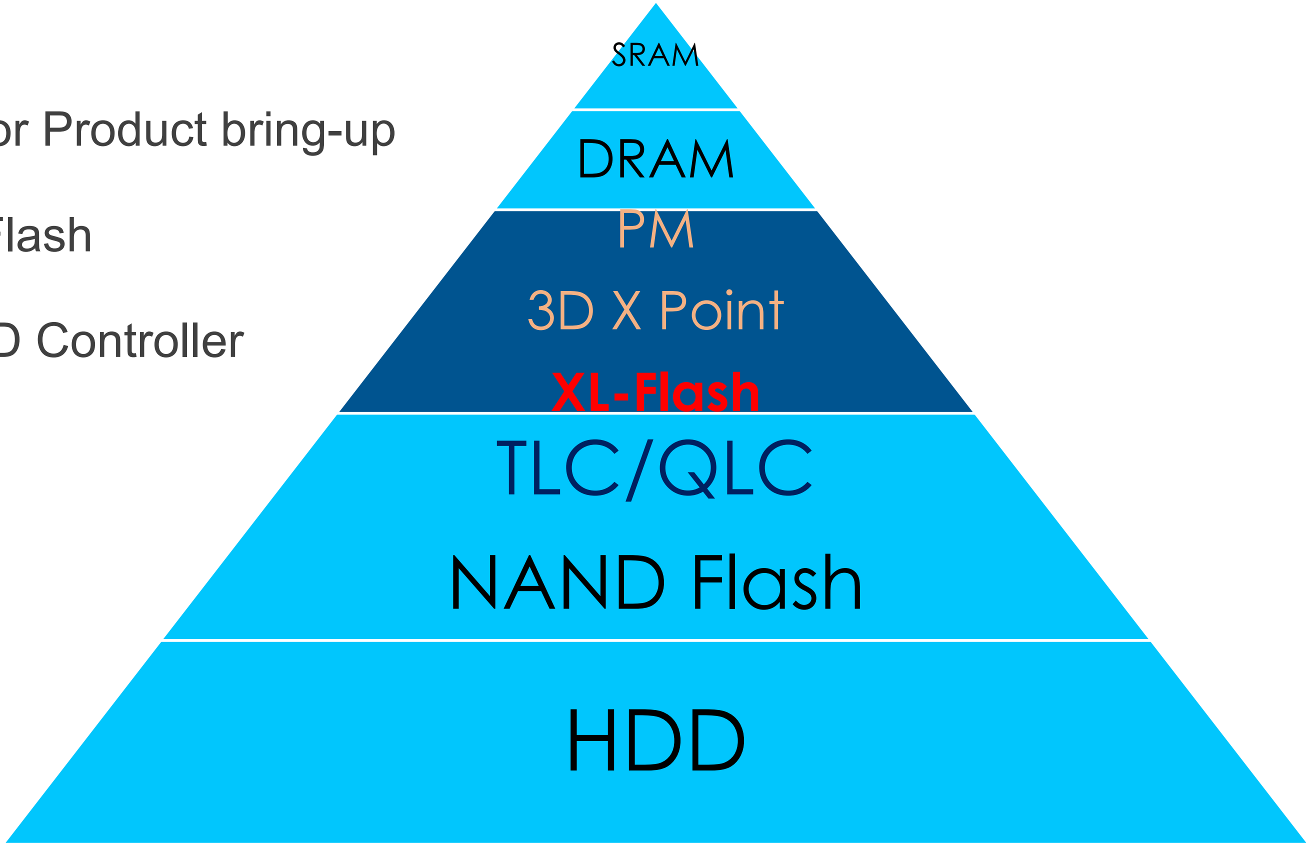
II. Performance Tuning w/ Real World Workloads

Optimization using Long Short Term Memory (LSTM) and Recurrent Neural Networks

Yafei Yang, CEO – DapuStor

XL-Flash Characteristics

- SLC 3D NAND flash memory, focusing on low latency
- Lower Latency, tR < 5 us; 25-45 uS Round Trip RT
- Proven Flash Technology, more open eco-system for Product bring-up
- More QoS aware design vs traditional TLC NAND Flash
- Compatible packages, porting easily to existing SSD Controller
- High Endurance



PM
Latency


NVDIMM-N PM


Optane SSD


DCPMM PM


XL-Flash SSD


TLC 3D NAND SSD

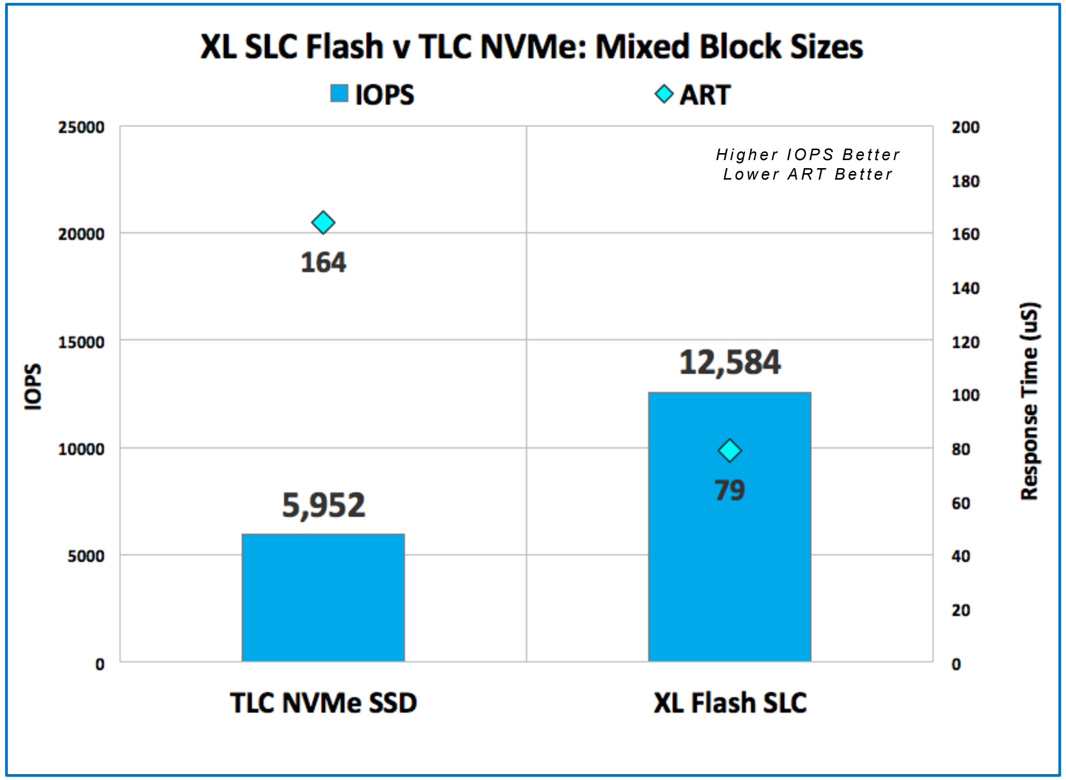
CPU/ SRAM	DRAM	Gap - PM	NAND MLC/TLC/QLC	HDD	Tape

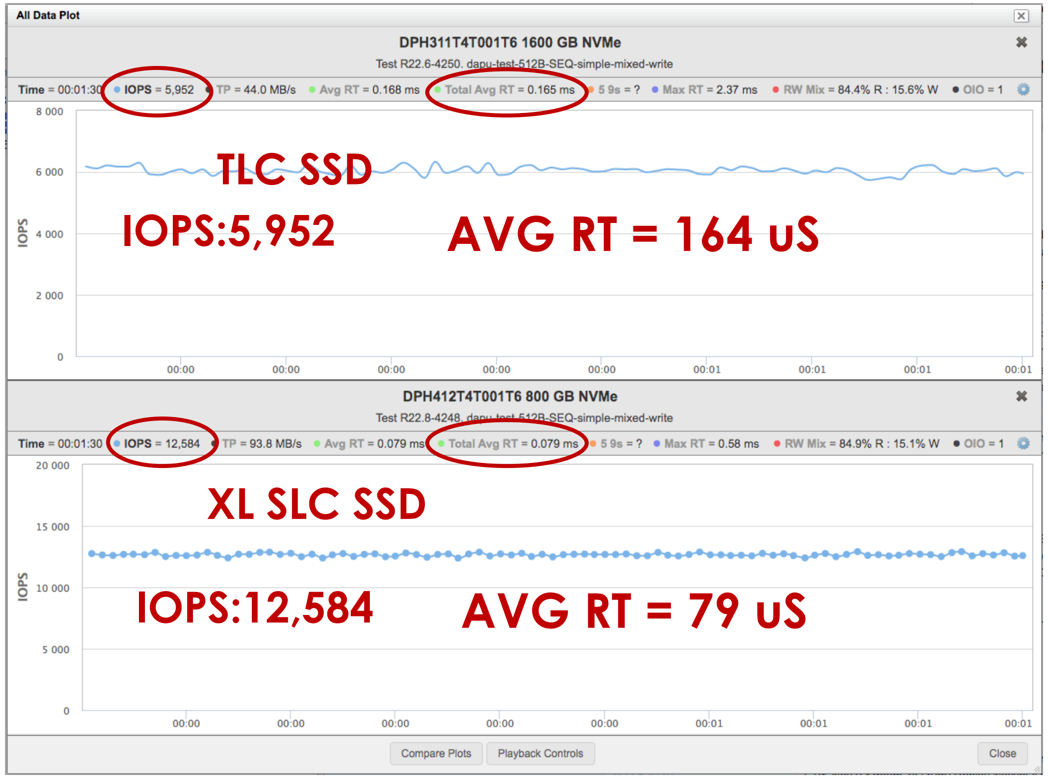
Latency: 1ns | 10ns | 100ns | 1μs | 10μs | 100μs | 1ms | 10ms | 100ms | 1s

Higher Capacity →

← Higher Cost

XL vs TLC Performance

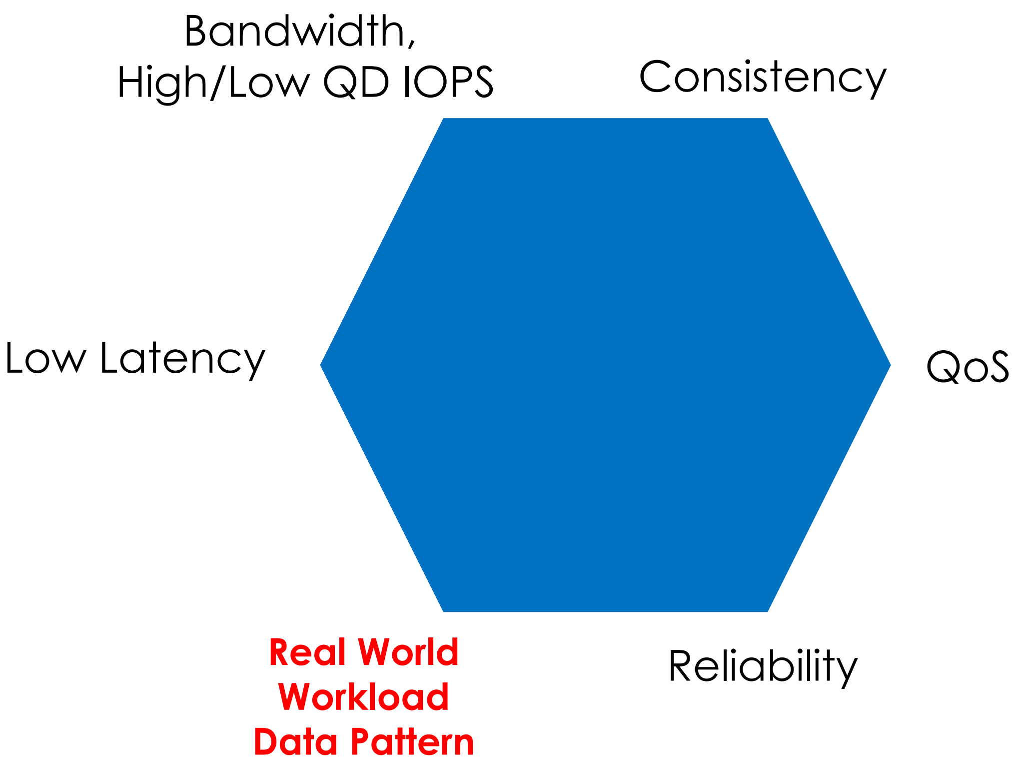




- **ART: 164 uS TLC to 79 uS SLC XL-Flash**
- **IOPS: 5,953 TLC to 12,584 SLC XL-Flash**

Real World Workloads

Challenge for Performance Tuning



Performance Tradeoff

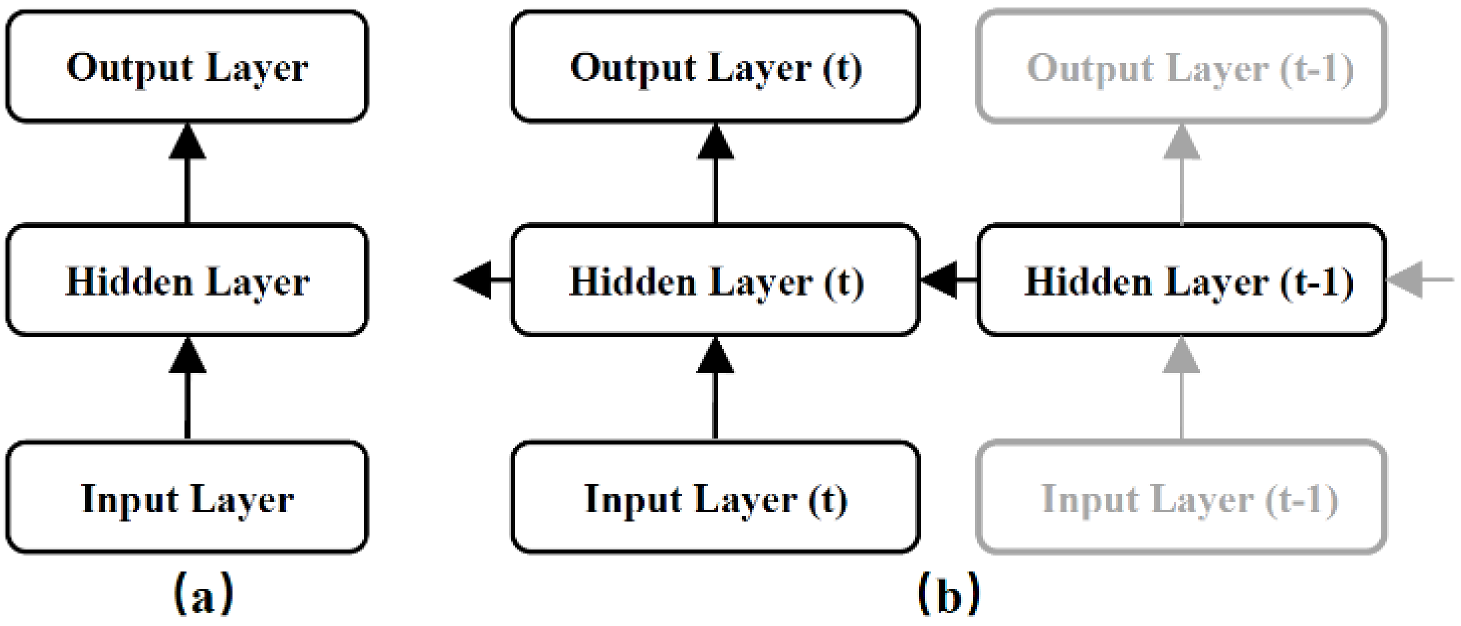
1. High QD IOPS vs low Latency
2. Benchmark pattern vs real workload
3. Reliability vs Performance test case

Challenge

1. Embedded system, making tools to trouble shooting
2. E2E system, tuning scheduler\dram\flash\algorithms
3. Real workload hard to get from the user cases

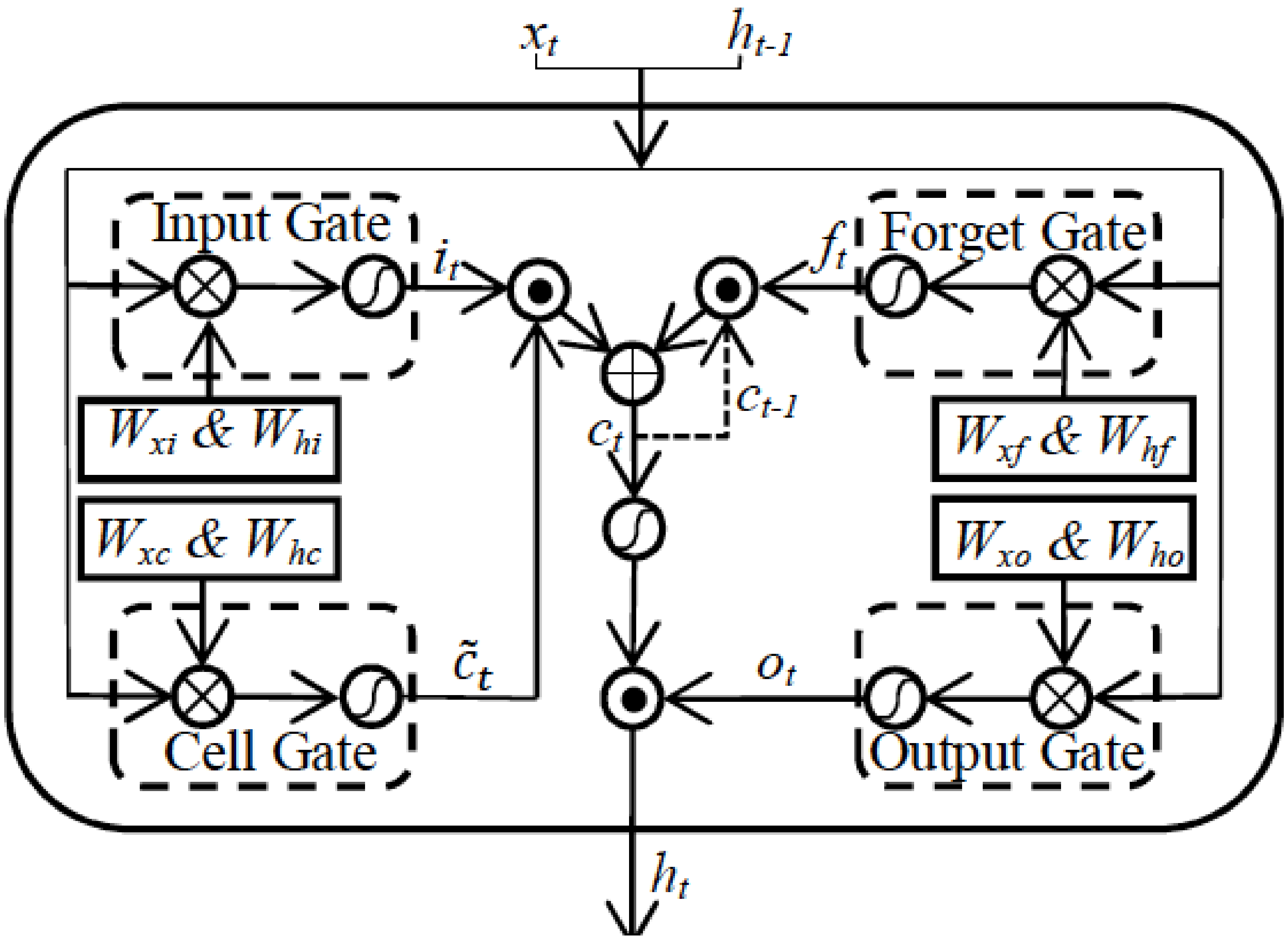
Predicting I/O Patterns Using LSTM

- LSTM applied in various applications, such as weather forecast, language modeling and speech recognition
- Storage I/O prediction is a time series problem
- Major challenge of using LSTM for the storage I/O prediction is the computation complexity

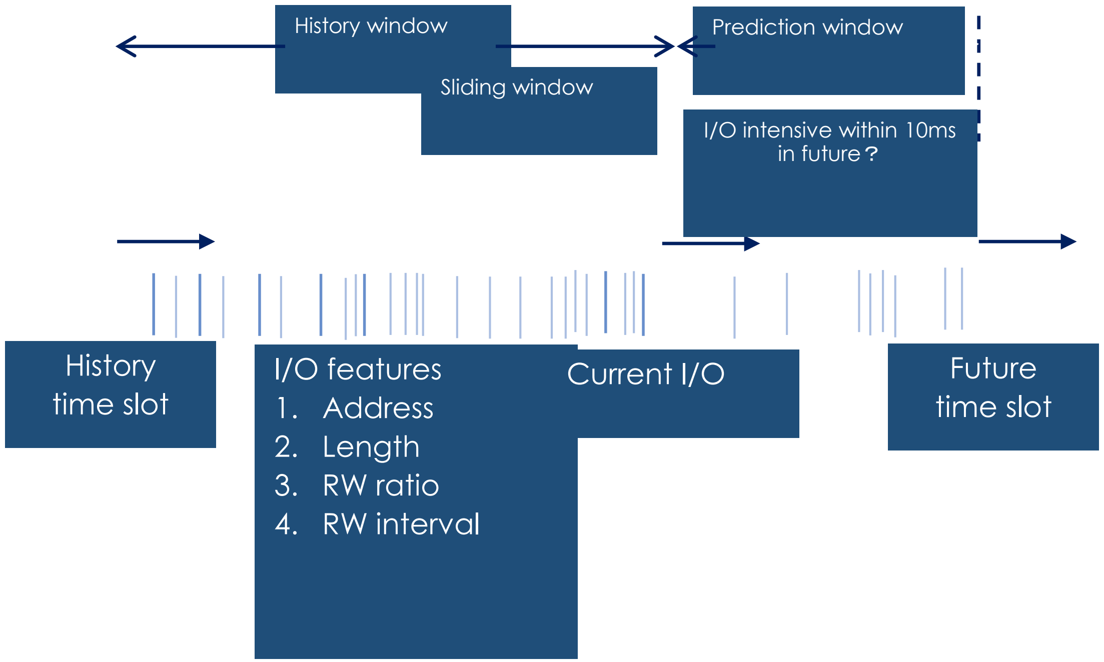


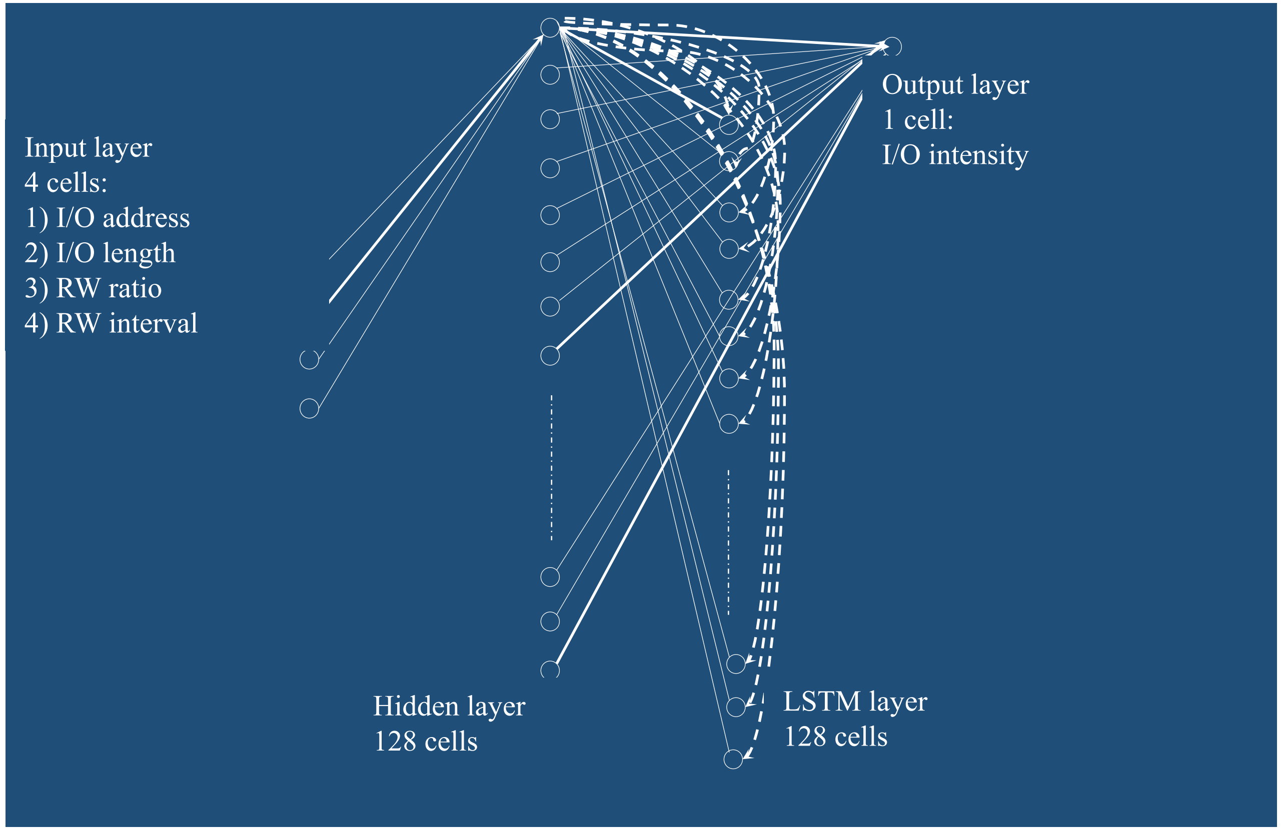
LSTM (RNN) time series neural network

Architecture of a LSTM Cell

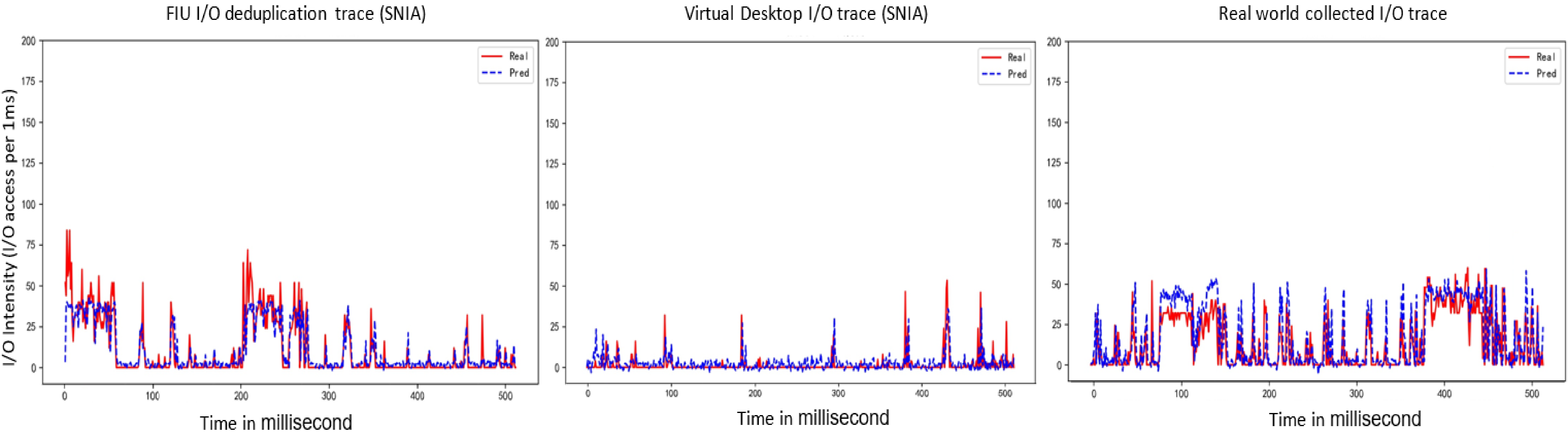


Input & Output of the Model



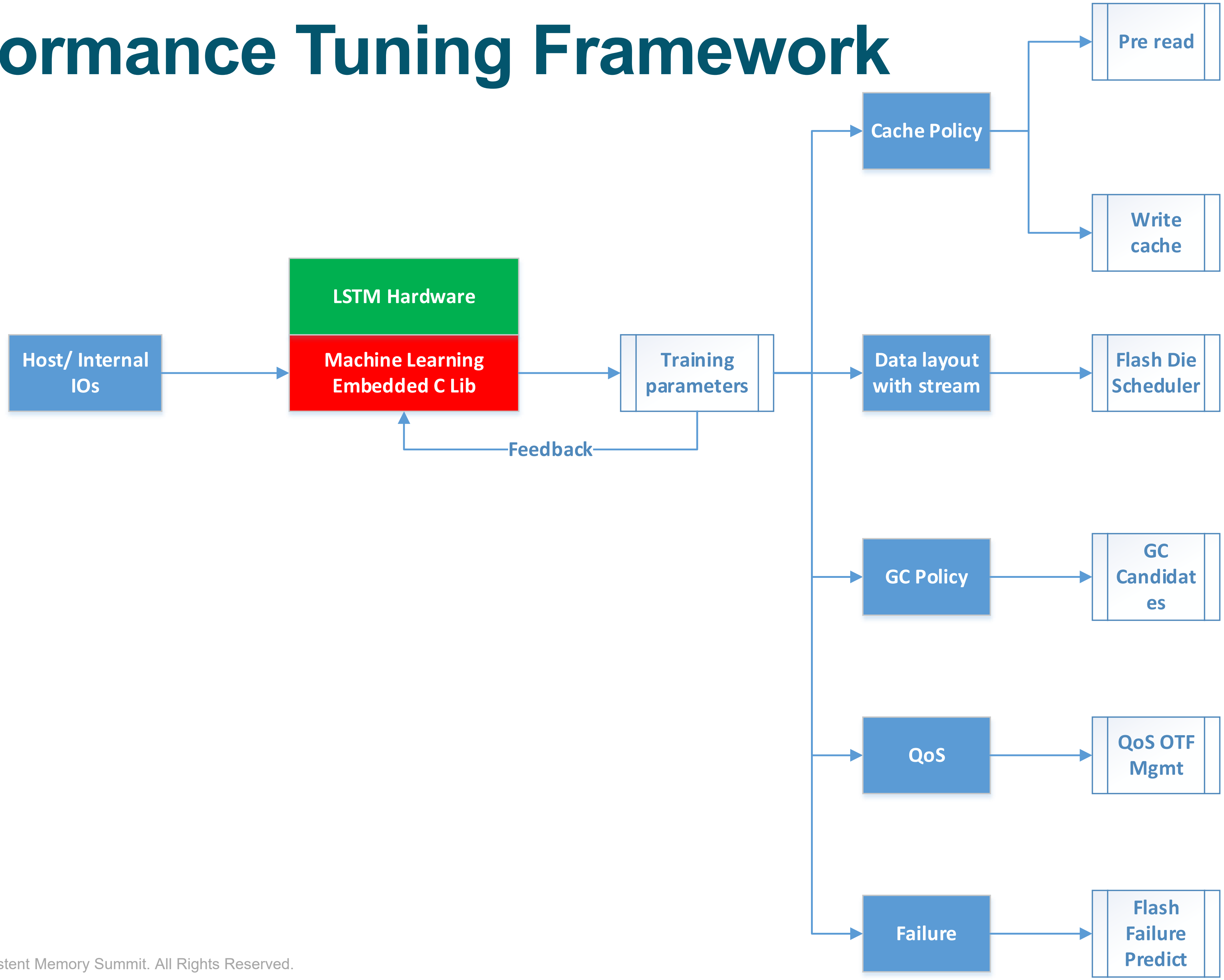


IO Prediction Results



The prediction result of storage I/O traces (configuration: history =1K, predict period=1ms)

Performance Tuning Framework



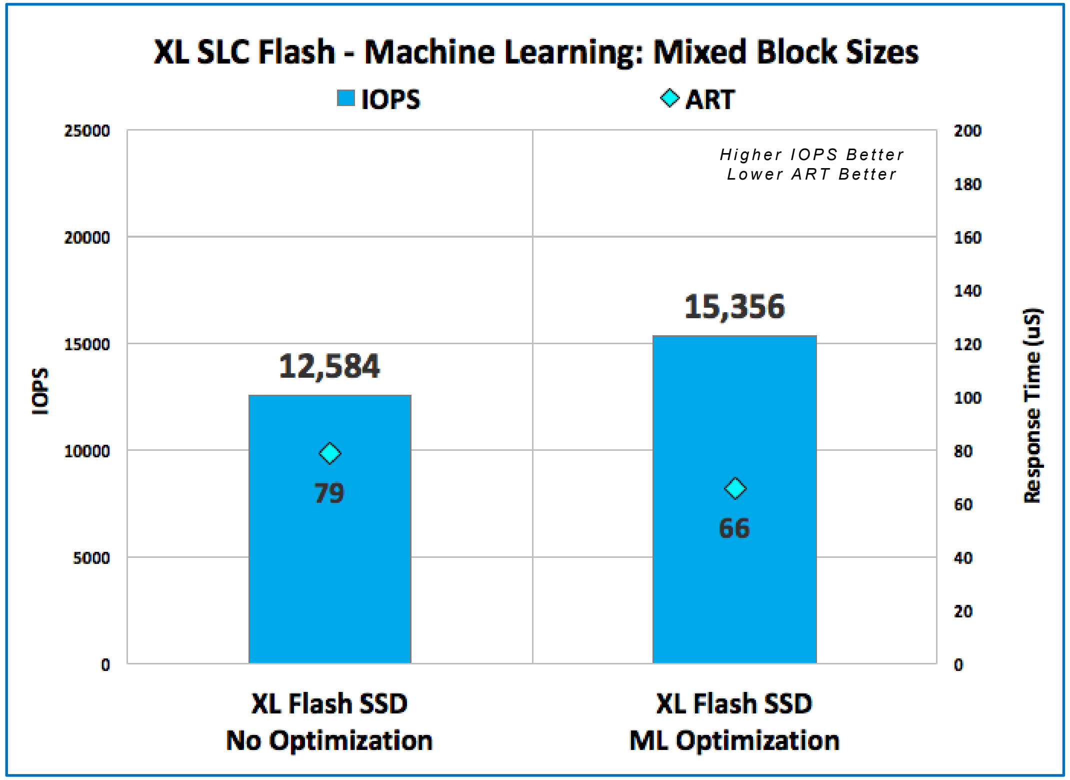
Performance Tuning Comparison

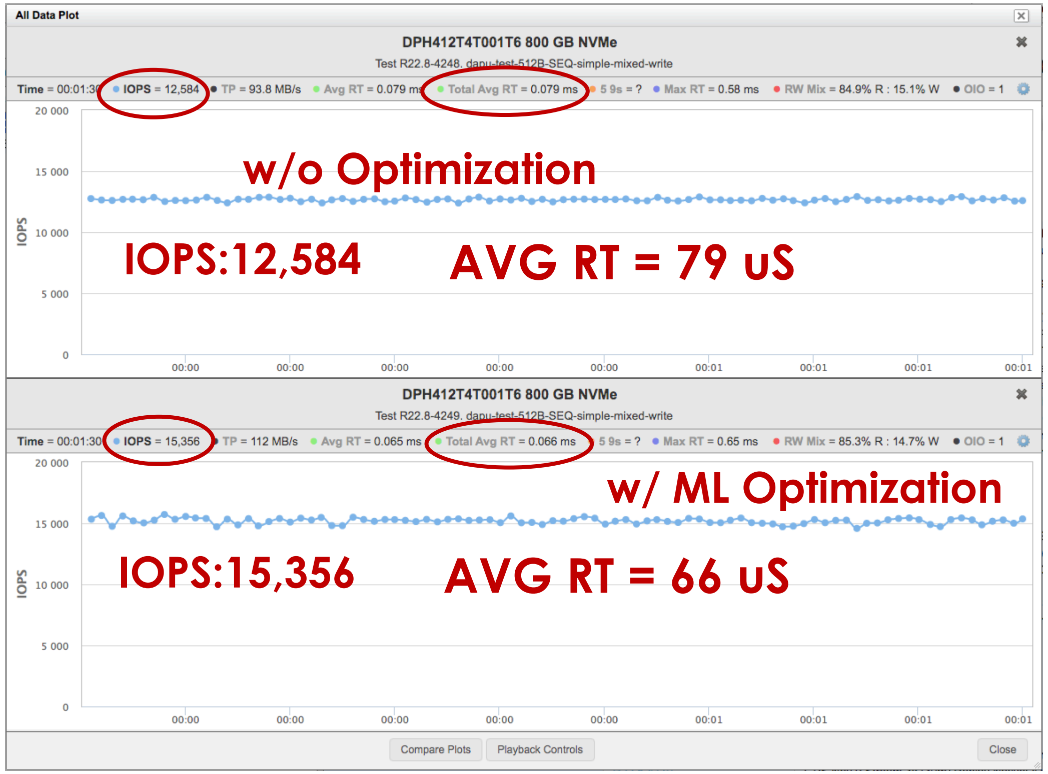
Hardware vs Software

	Hardware	Software
Data	float	int
Response time	ns	us
Predict window	ms	s
Predict performance	> 95%	~80%

Performance Tuning Results

XL-Flash w/ Machine Learning





Up to 20% percent improved with Machine Learning

- Smart IO Tag detection using machine learning
- Smart Cache Design using machine learning

Thank you

Please visit www.snia.org/pmsummit for presentations