

The logo for the SNIA Persistent Memory Summit. It features the SNIA logo (a small square with a dot) to the left of the text. The text is arranged in two lines: "PERSISTENT MEMORY" in a dark purple font, and "SUMMIT" in a large, blue, stylized font. The "PM" in "PERSISTENT MEMORY" is also in a large, blue, stylized font. The background of the logo is a light blue gradient with faint binary code and circuit patterns.

SNIA PERSISTENT MEMORY
PM SUMMIT

JANUARY 24, 2018 | SAN JOSE, CA

Workload Performance Comparison

NVMe SSD, NVDIMM Block IO & NVDIMM Mmap

Eden Kim, Calypso Systems, Inc.

Agenda

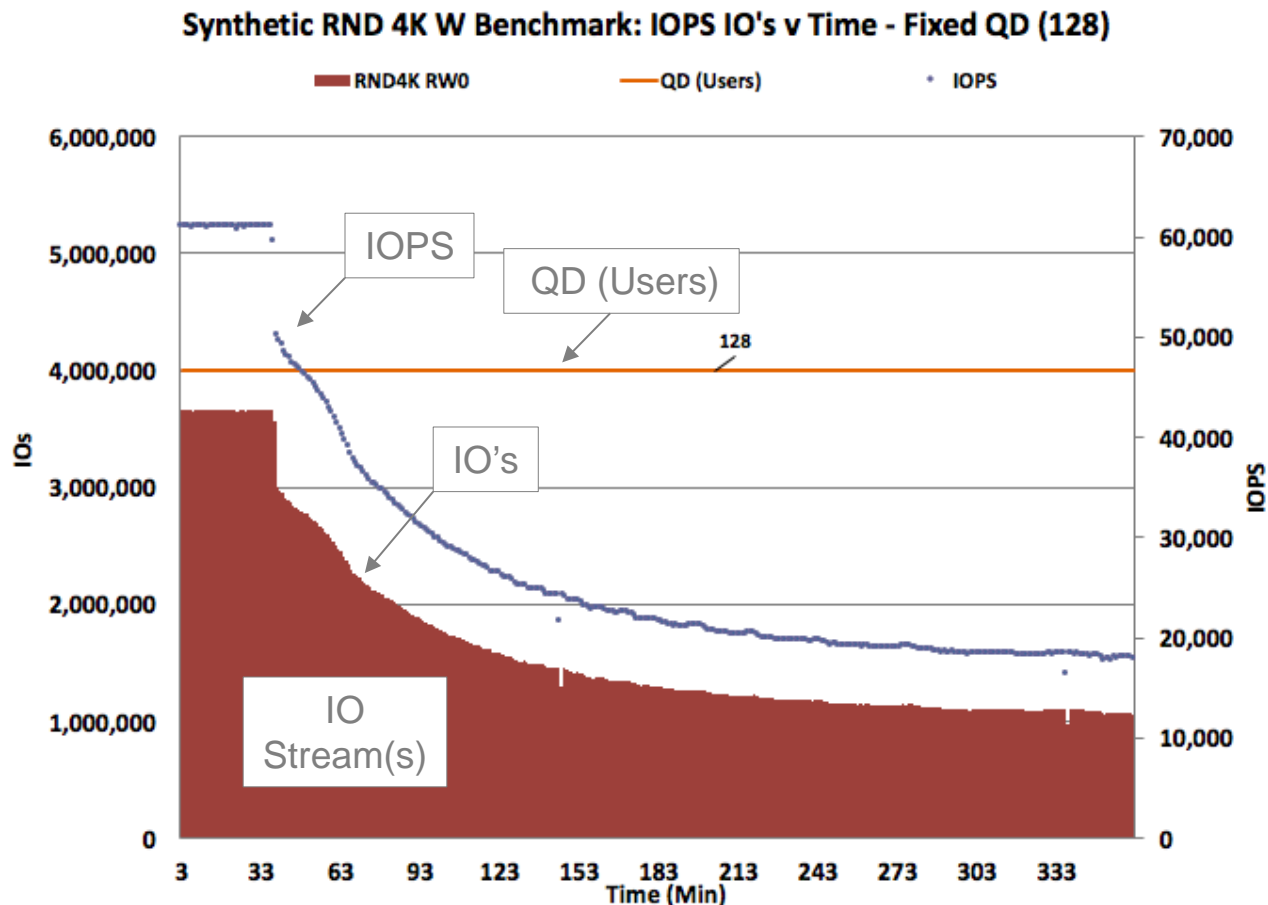
- **Part 1: Workload Comparisons:** Synthetic Benchmarks & Real World Workloads
- **Part 2:** NVMe SSD, NVDIMM Block IO & NVDIMM Mmap
- **Part 3: NVDIMM DAX:** memcopy; non temporal writes; msync
- **Conclusions**

Part 1: Workload Comparisons

- Synthetic Benchmark: RND 4K Writes
- Synthetic SQL Workload: RND 8K 65:35 RW Mix
- Real World SQL Workload: 6 IO Stream Replay & Fixed Composite

RND 4K W Monotonic Benchmark

- WSAT (Write Saturation) Test
- Single IO Stream
 - 100% RND 4K W
- Fixed QD (Users) - 128

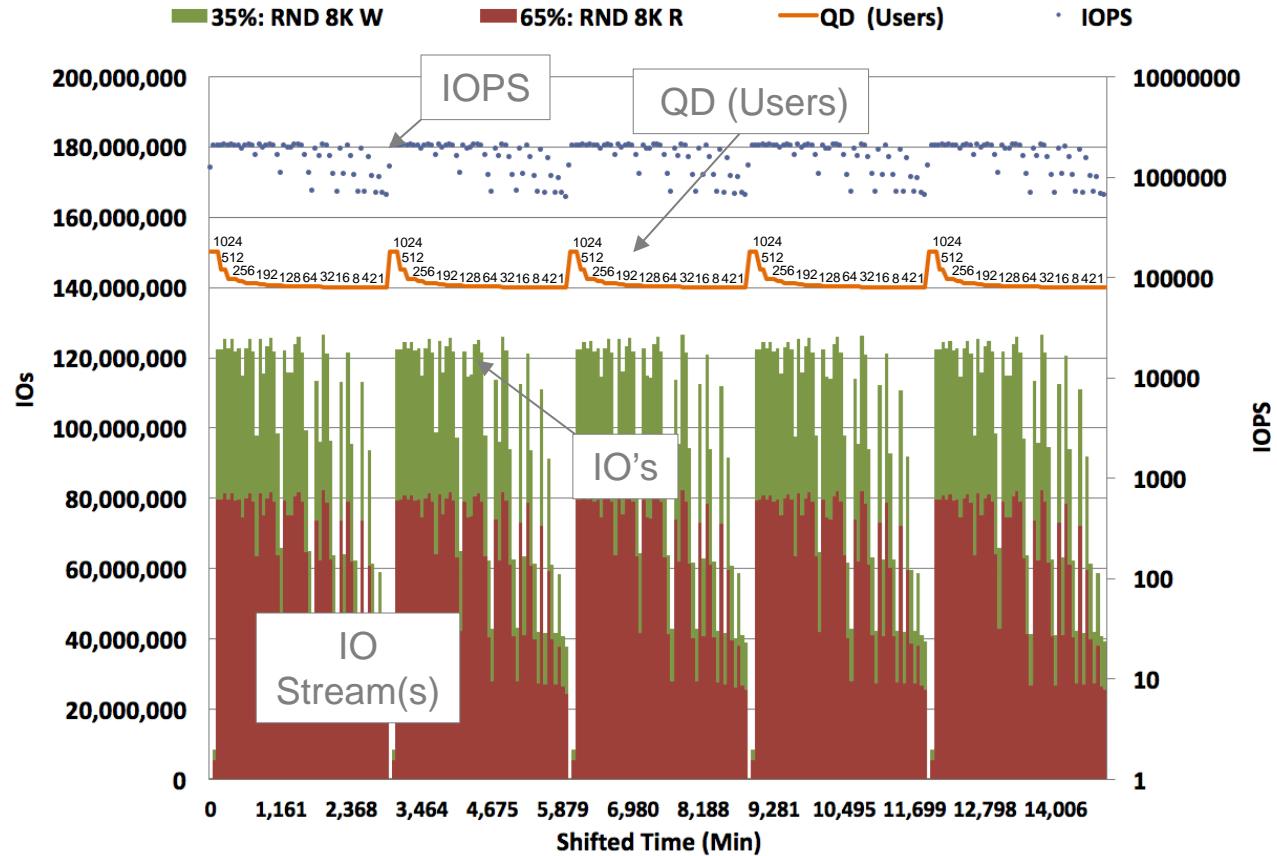




Synthetic SQL Workload: RND 8K 65:35 RW - QD 1 to 1024

Synthetic SQL Server Workload

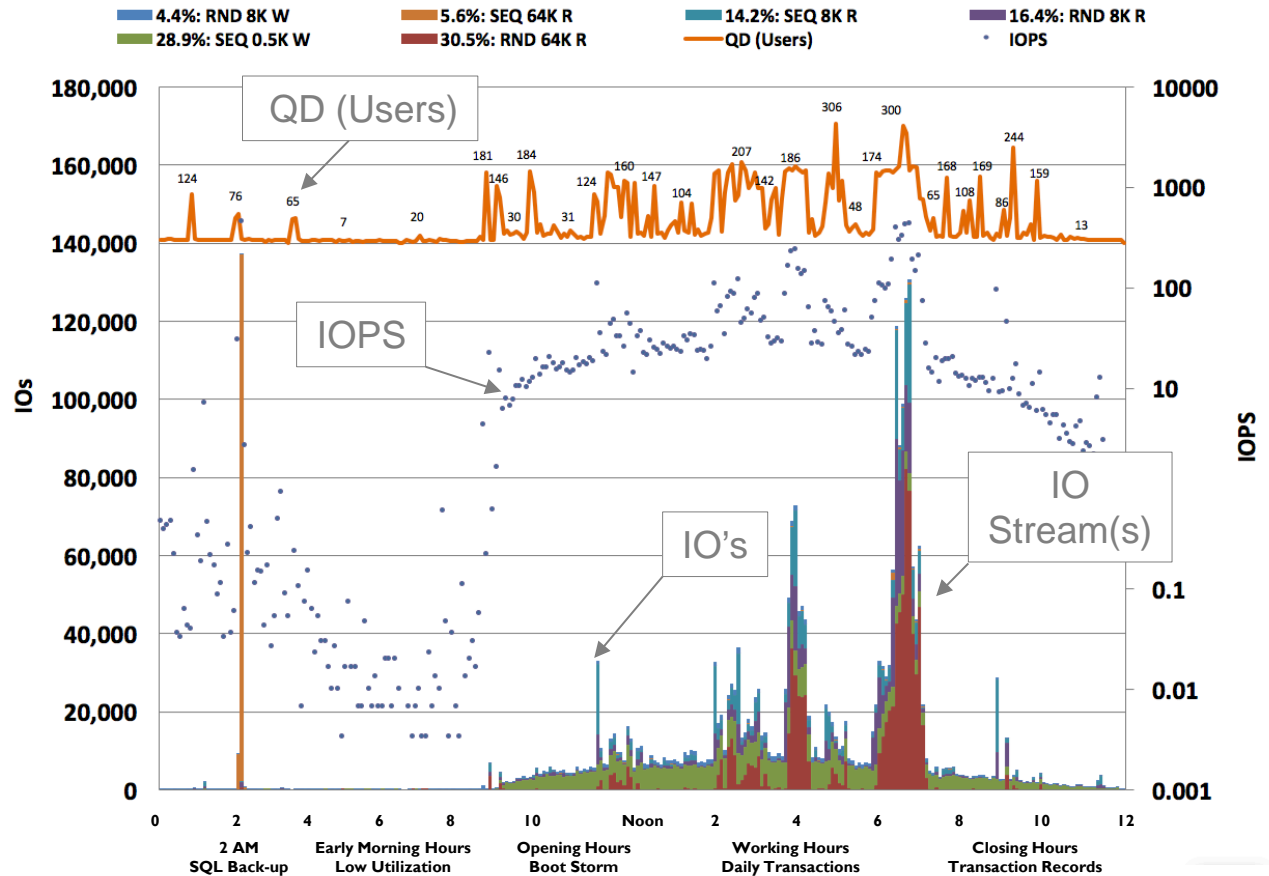
- DIRTH Test
- Two IO Streams
 - 65% RND 8K R
 - 35% RND 8K W
- QD (User) Range 1-1024



Real World SQL from Retail Web Portal - IO Streams, Queue Depths & IO's

Real World SQL Server Workload

- Replay Native Test
- Six IO Streams
 - 30.5% RND 64K R
 - 28.9% SEQ 0.5K W
 - 16.4% RND 8K R
 - 14.2% SEQ 8K R
 - 5.6% SEQ 64K R
 - 4.4% RND 8K W
- QD (User) Range 3 - 306



Part 2: NVMe SSD, NVDIMM Block IO & Mmap

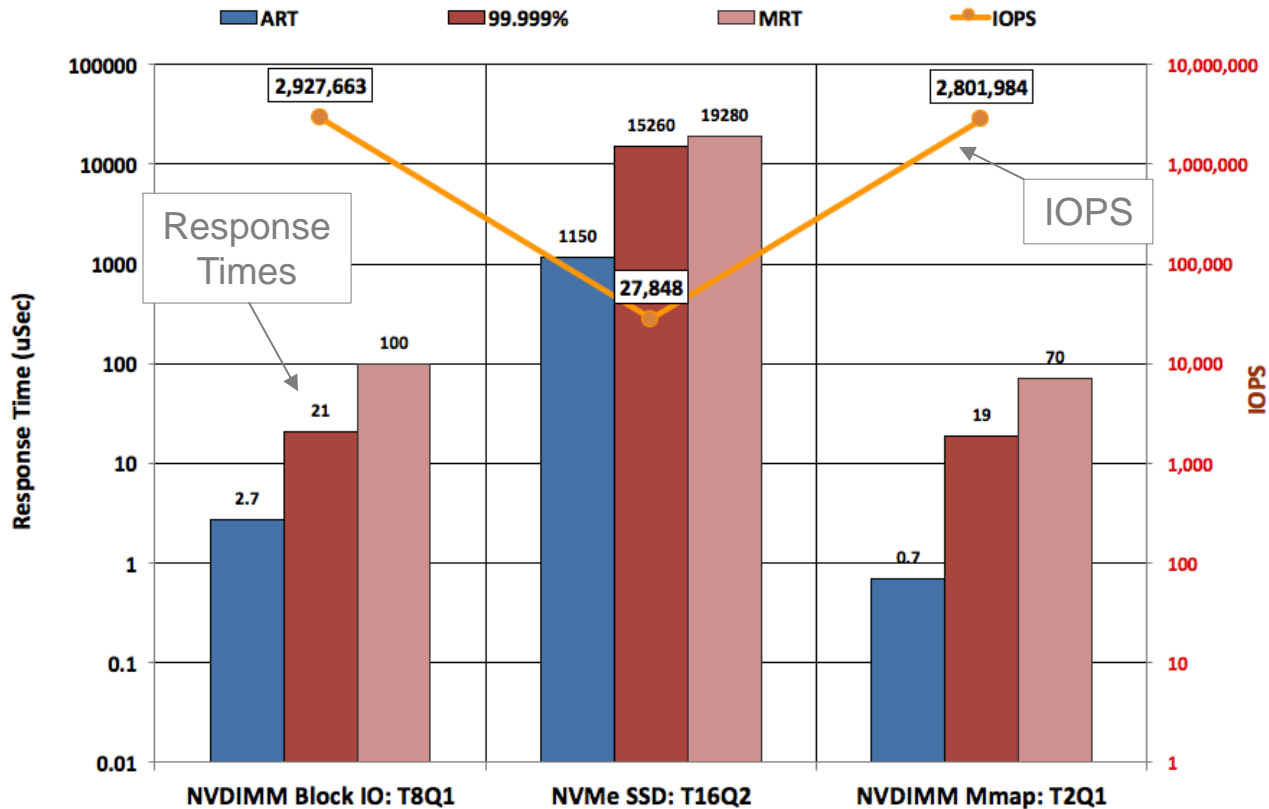
- A. Synthetic Benchmark: RND 4K Writes
- B. Synthetic SQL Workload: RND 8K 65:35 RW Mix
- C. Real World SQL Workload: 6 IO Stream Fixed Composite
- D. Real World SQL Workload: 6 IO Stream Replay with Native QDs

A: RND 4K W – Monotonic Workload, Multiple T/Q Multiple Cores

Storage	Workload RND 4K W DIRTH Synthetic Benchmark	Settings (Multiple Cores)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVMe SSD	RND 4K 100% W	Block IO T2/Q16 WCD	27,848	1,150 uSec	1,526,000 uSec
NVDIMM Block IO	RND 4K 100% W	Block IO T8/Q1 WCD	2,927,663	2.7 uSec	21 uSec
NVDIMM Mmap	RND 4K 100% W	DAX FS T2/Q1 Mmap (no Flush)	2,801,984	0.7 uSec	19 uSec

Test Platform: Supermicro X11DR1 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

RND 4K W Benchmark - DIRTH IOPS & Response Times

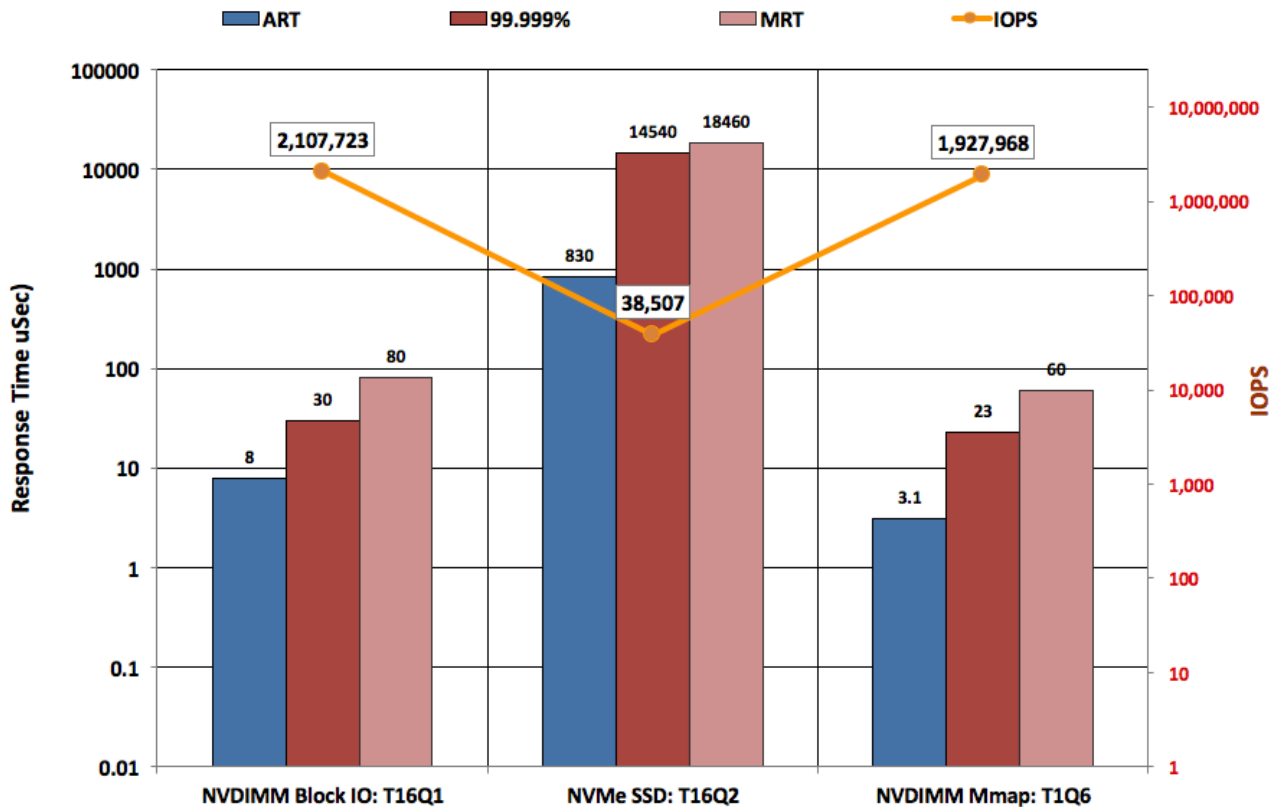


B: Synthetic SQL Workload – Monotonic RND 8K RW65

Storage	Workload RND 8K RW65 DIRTH Synthetic SQL	Settings (Multiple Cores)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVMe SSD	RND 8K 65:35 RW	Block IO T2/Q16 WCD	38,507	830 uSec	14,540 uSec
NVDIMM Block IO	RND 8K 65:35 RW	Block IO T8/Q1 WCD	2,107,723	8 uSec	30 uSec
NVDIMM Mmap	RND 8K 65:35 RW	DAX FS T2/Q1 Mmap (no Flush)	1,927,968	3.1 uSec	23 uSec

Test Platform: Supermicro X11DRI 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

Synthetic SQL Workload - DIRTH RND 8K RW65 IOPS & Response Times

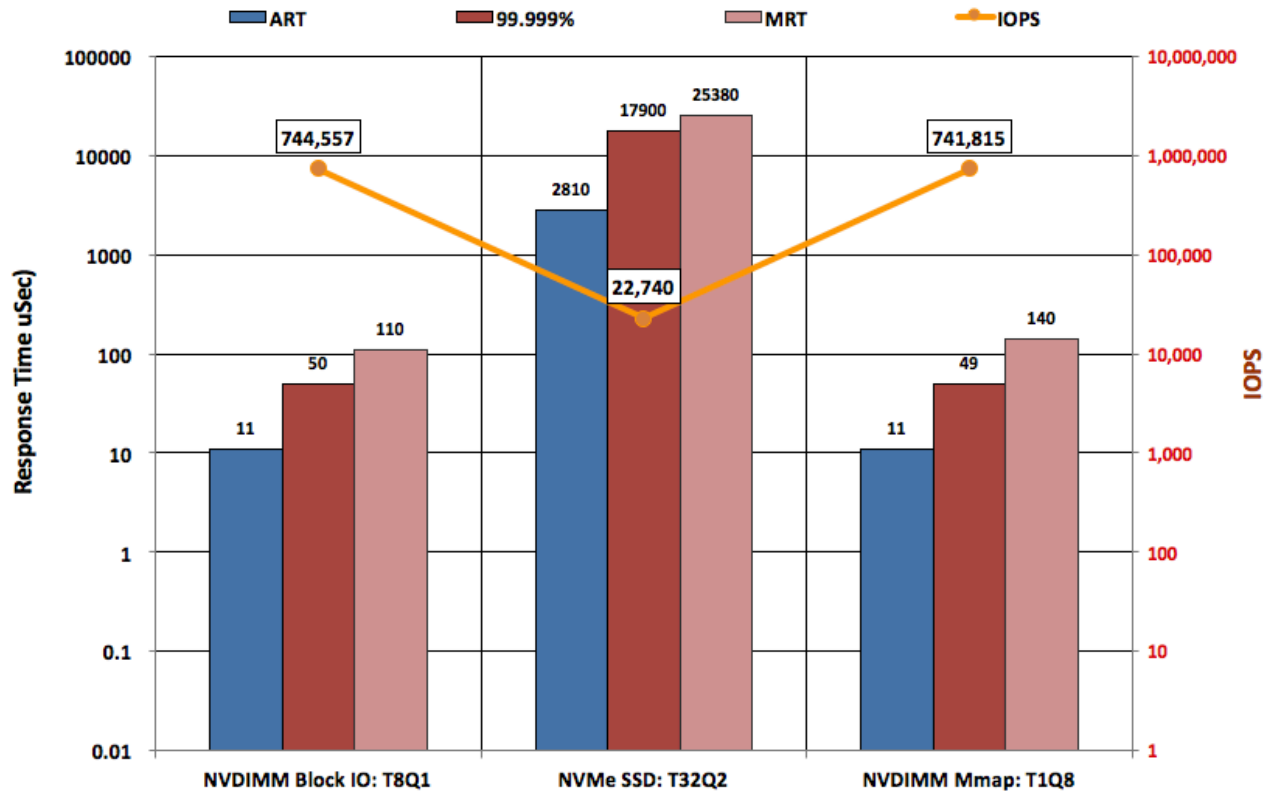


C: Real World SQL – Fixed 6 IO Stream Composite, QD 1-1024

Storage	Workload Real World SQL DIRTH	Settings (Multiple Cores)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVMe SSD	Fixed Composite % of: RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	Block IO T32/Q2 WCD	22,740	2,810 uSec	17,900 uSec
NVDIMM Block IO	Fixed Composite % of: RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	Block IO T8/Q1 WCD	744,557	11 uSec	50 uSec
NVDIMM Mmap	Fixed Composite % of: RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	DAX FS T1/Q8 Mmap (no Flush)	741,815	11 uSec	49 uSec

Test Platform: Supermicro X11DR1 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

Real World SQL - DIRT 6 IO Stream Composite

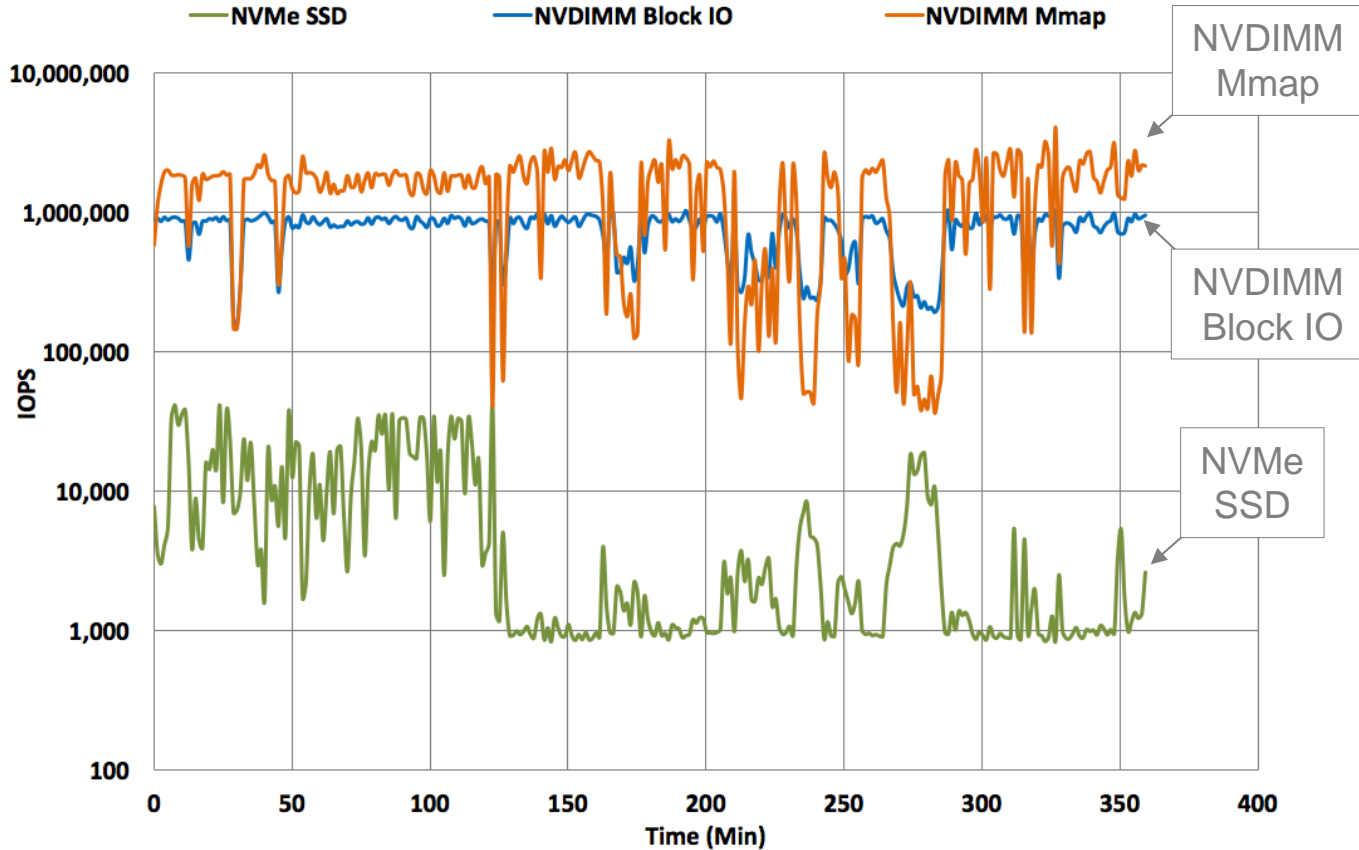


D: Real World SQL Replay – Changing IO Stream Combos & QDs

Storage	Workload Real World SQL Native Replay	Settings (Multiple Cores)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVMe SSD	Replay Capture Sequence RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	Block IO TI/QI WCD	7,217	1,910 uSec	15,020 uSec
NVDIMM Block IO	Replay Capture Sequence RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	Block IO TI/QI WCD	749,952	15 uSec	140 uSec
NVDIMM Mmap	Replay Capture Sequence RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	DAX FS TI/QI Mmap (no Flush)	1,927,968	1 uSec	2,860 uSec

Test Platform: Supermicro XI1DRI 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

Real World SQL Server - Replay Test - Native QD



Changing IO Stream Combinations & QD (Users)

Part 3: NVDIMM – Single Core T1Q1: mmap, memcpy, msync

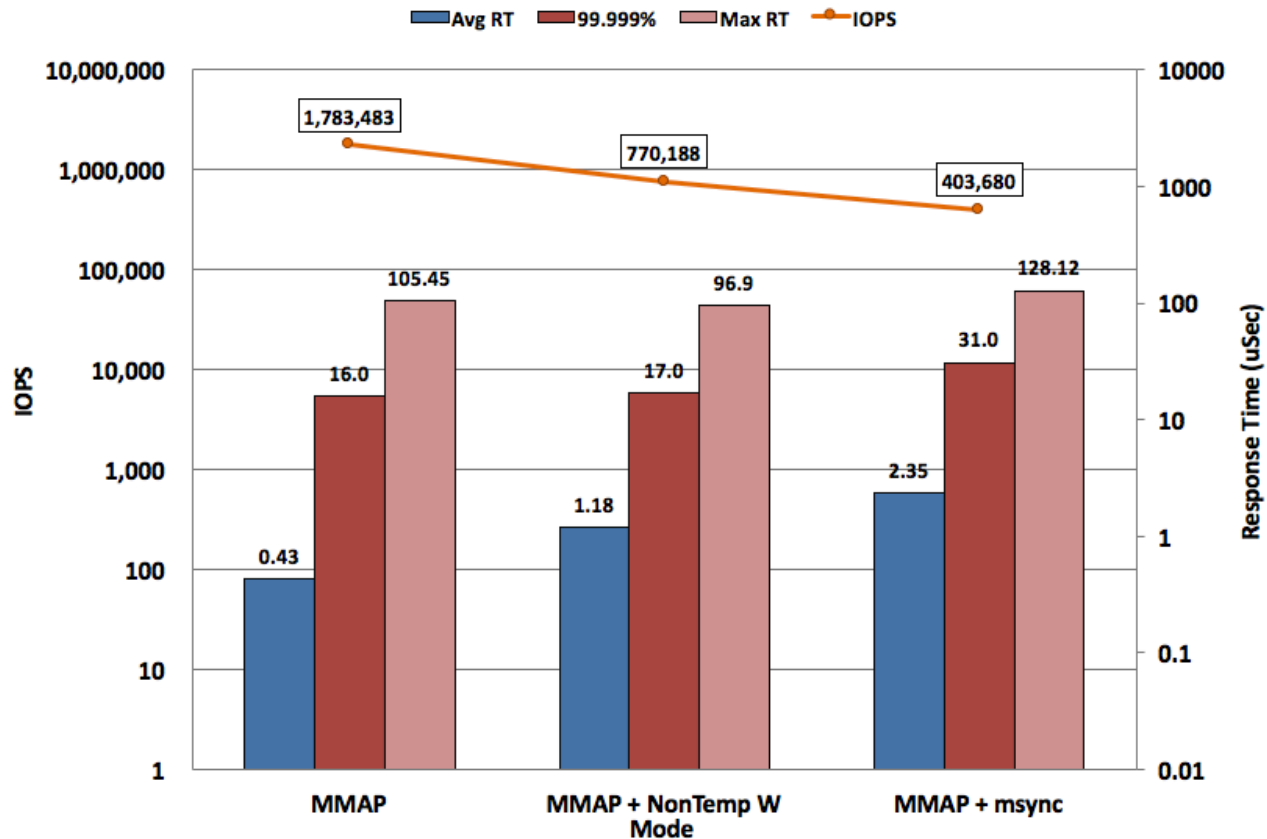
- A. Synthetic Benchmark: RND 4K Writes
- B. Synthetic SQL Workload: RND 8K 65:35 RW Mix
- C. Real World SQL Workload: 6 IO Stream Fixed Composite

A: NVDIMM Mmap –RND 4K W – Single Core / variable Flush settings

Storage	Workload RND 4K W Benchmark	Settings (Single Core)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVDIMM Mmap	RND 4K 100% W	TI/QI Memcopy – (No Flush)	1,783,483	0.43 uSec	16 uSec
NVDIMM Non Temp W	RND 4K 100% W	TI/QI Non Temporal W (implicit Flush)	770,188	1.18 uSec	17 uSec
NVDIMM Msync	RND 4K 100% W	TI/QI Msync (explicit Flush)	403,680	2.35 uSec	31 uSec

Test Platform: Supermicro XI IDRI 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

NVDIMM Mmap - RND 4K W - Single Core, T1Q1

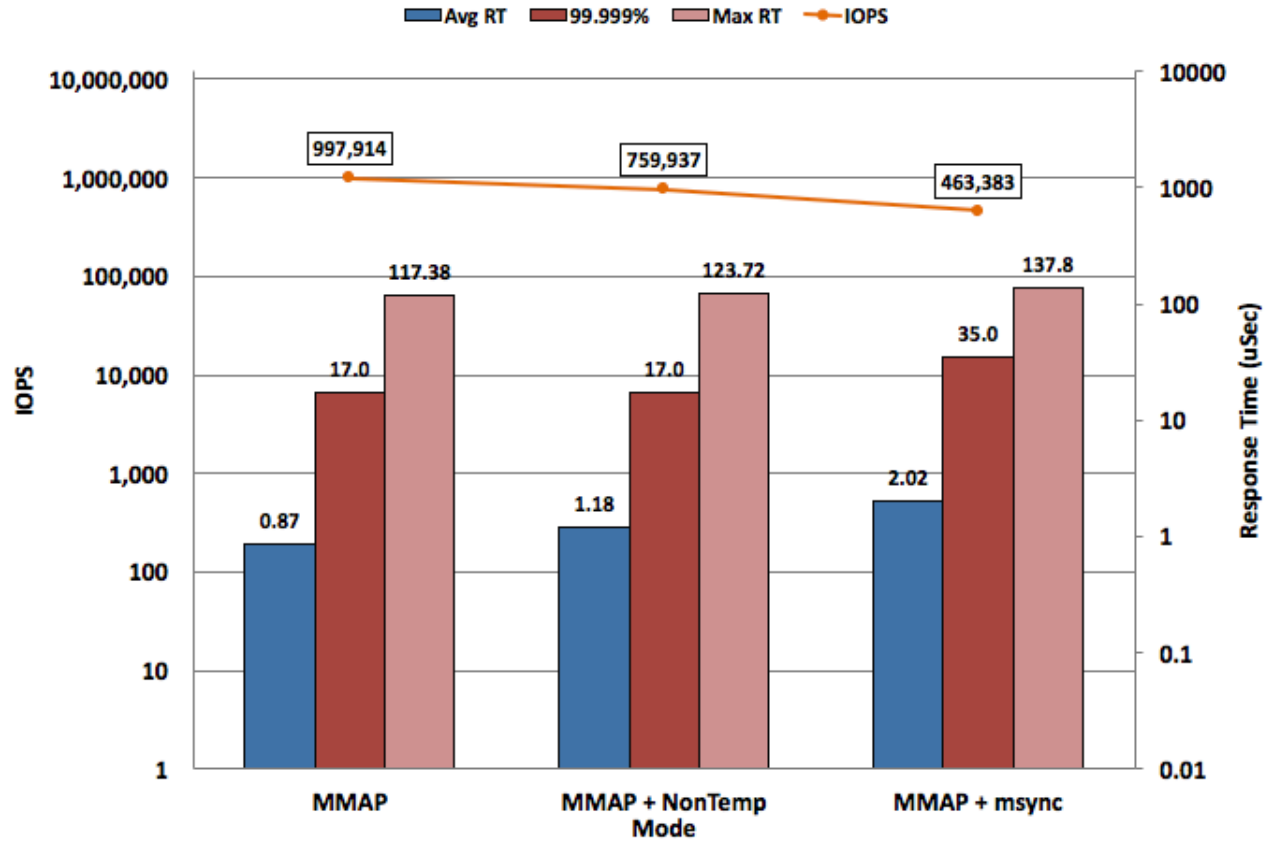


B: NVDIMM Mmap – RND 8K RW65 – Single Core / variable Flush settings

Storage	Workload Synthetic SQL	Settings (Single Core)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVDIMM Mmap	RND 8K 65:35 RW	TI/QI Memcopy – (No Flush)	997,914	0.87 uSec	17 uSec
NVDIMM Non Temp W	RND 8K 65:35 RW	TI/QI Non Temporal W (implicit Flush)	759,937	1.18 uSec	17 uSec
NVDIMM Msync	RND 8K 65:35 RW	TI/QI Msync (explicit Flush)	463,383	2.02 uSec	35 uSec

Test Platform: Supermicro XI IDRI 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

NVDIMM Mmap - RND 8K RW65 - Single Core, T1Q1

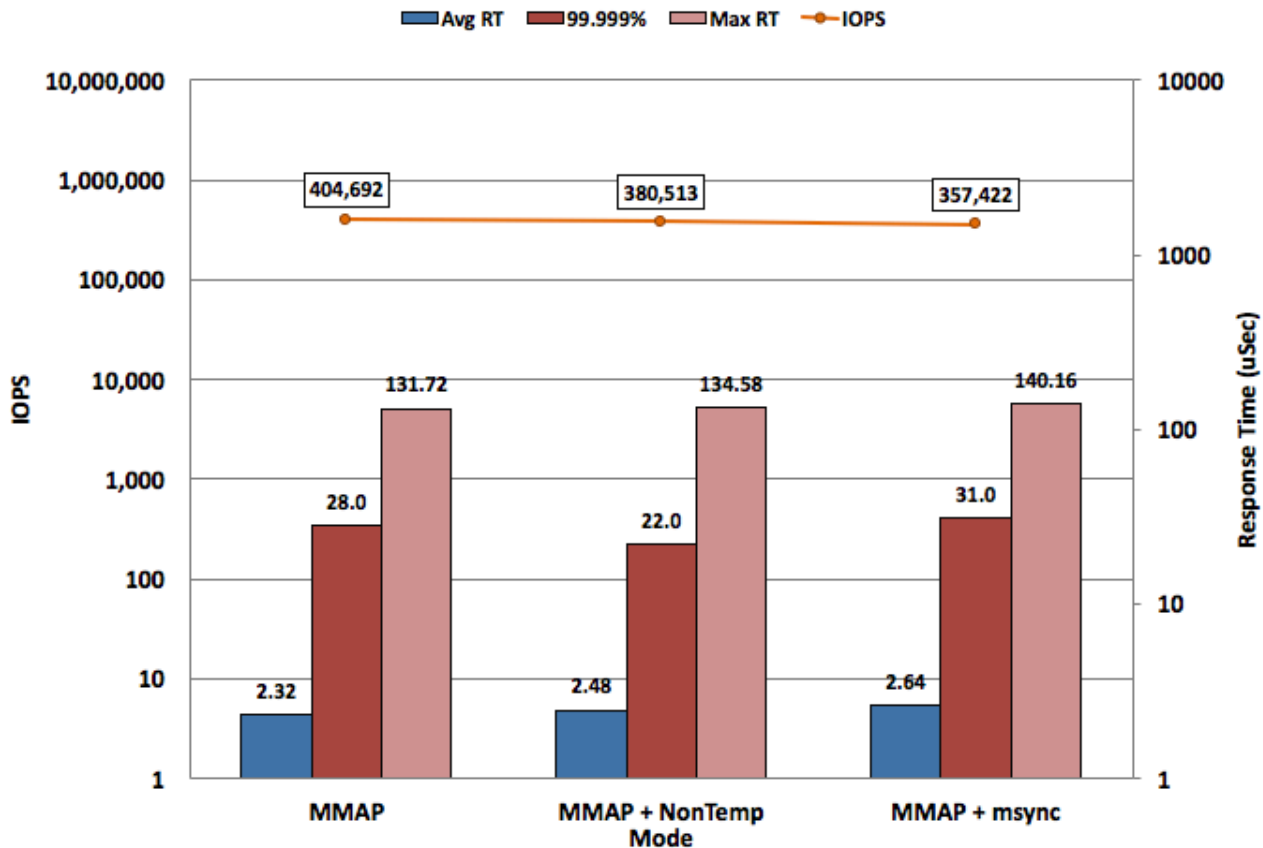


C: NVDIMM Mmap – Real World SQL – Single Core / variable Flush settings

Storage	Workload Real World SQL Native Replay	Settings (Single Core)	IOPS	Average Response Time uSec	5 9s Response Time uSec
NVDIMM Mmap	Fixed Composite % of: RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	TI/QI Memcopy (No Flush)	404,692	2.32 uSec	28 uSec
NVDIMM Non Temp W	Fixed Composite % of: RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	TI/QI Non Temporal W (implicit Flush)	380,513	2.48 uSec	22 uSec
NVDIMM Msync	Fixed Composite % of: RND 64K R SEQ 0.5K W RND 8K R SEQ 8K R SEQ 64K R RND 8K W	TI/QI Msync (explicit Flush)	357,422	2.64 uSec	31 uSec

Test Platform: Supermicro X11DR1 16GB DDR4 2133 Mhz RDIMM RAM, Intel XEON 8160 2.1 Ghz 24 core, 16 GB DDR4 JEDEC NVDIMM-N
 Software: Ubuntu 16.04.3 LTS Linux 4.10.0-28; DAX File System
 Test Software: Calypso CTS 7.0 fe 1.26.25 be 1.9.317

NVDIMM Mmap - Real World SQL Composite - Single Core, T1Q1



Conclusions

- Workload definition has an impact on performance comparisons
- NVDIMM Block IO is faster than NVMe SSD by orders of magnitude
- NVDIMM Block IO and Mmap load/store w/o flush are substantially equivalent in IOPS but Mmap has better Response Times
- What is the Impact of Mmap with Flush?
 - Using Mmap with No Flush – data may not be persistent unless there is Flush on Fail
 - Using Mmap with Non Temp W – slower but flushes on demand (explicit command)
 - Using Mmap with Msync – flushes every write using file system software