(intel®)

High-Performance RAID Software for Fast NVMe* Storage Systems with Intel® Optane™ DC Storage

RAIDIX ERA* delivers optimized storage solutions for customers that need speed.

NVM Express* (NVMe*) has been rapidly adopted by enterprise and cloud customers as their primary storage media interface. Intel® Optane™ DC Solid State Drives (SSDs), combined with NVMe, deliver fast write and read performance and high endurance for intensive storage tasks and the most demanding workloads. RAIDIX, a developer of high-performance storage systems, created RAIDIX ERA* software to be optimized for NVMe and Intel Optane DC technologies. Tests performed by RAIDIX show how dramatically RAIDIX ERA outperforms open source software–based systems.

The Rise of NVMe

NVMe is rapidly replacing Serial ATA (SATA) as the preferred storage interface. According to a G2M Research report, more than 60 percent of all-flash arrays (AFA) will utilize NVMe storage media by 2022, and more than 30 percent of all AFAs will use NVMe over Fabrics (NVMe-oF*).[1]

Intel Optane DC SSDs are better able to take advantage of the speed of NVMe than NAND drives. NAND drives, which are block addressable, are inherently slower to write data because of the need to delete blocks before writing. This kind of garbage handling overhead issue has been addressed with Intel Optane DC SSDs, with their new memory media technology that is byte addressable like memory and able to write data without deleting blocks.

Introducing RAIDIX ERA: Optimized for Speed on NVMe Drives

With high-performing Intel Optane DC SSDs on a fast NVMe interface, slow software can become a bottleneck that limits overall system performance. RAIDIX, a high-performance storage solutions company founded in 2009, has addressed this bottleneck by developing RAIDIX ERA software. RAIDIX ERA is presented by the Linux* kernel module and is customized for the most popular Linux distributions.[2] Customers typically receive RAIDIX ERA as part of a complete hardware-plus-software storage solution.

"We expect that NVMe SSDs will represent the majority of drives shipped in the enterprise in 2019."
— Michael Heumann, lead analyst for G2M Research[1]

RAIDIX has proven its ability to deliver large-scale storage solutions to meet client needs, including:[3]

- A post-production media company that needed to process 4K and 8K raw video files in real time

- A city that replaced 2,000 closed-circuit television (CCTV) cameras with high-definition (HD) streaming units and needed to quadruple its storage capacity

- A large railway company that needed to perform rapid overnight backup operations at its headquarters and at 16 regional computer centers

- A large high-performance computing (HPC) infrastructure within a national academic project that needed 65 PB of storage

RAIDIX ERA, the latest software product to emerge from the company's research lab, is a high-performance software RAID especially designed for fast protocols like NVMe and fast storage like Intel Optane DC SSDs. Built around parallel input/output (I/O), lockless architecture, and new methods of erasure coding, RAIDIX ERA delivers high productivity for data-intensive applications and works effectively with both local drives and remote drives shared via NVMe-oF. RAIDIX ERA provides throughput speeds up to 23 GB per second (GBps) and 5,500,000 I/O operations per second (IOPS) for one RAID 6 group.[4]

Test Results: Intel Optane Technology Latency vs. NAND

RAIDIX conducted latency and I/O tests comparing performance of RAIDIX ERA software on an array of four Intel Optane DC SSDs versus an array of 12 NAND SSDs. Testing included both sequential and random write and read operations for each RAID 5 in various combinations of block sizes, queue depths, and write/read percentages. As a result, the system with four Intel Optane DC P4800X drives reached 1,043,000 IOPS, while the system with 12 Intel SSD DC D3700 drives reached only 864,000 IOPS. Figure 2 shows a representative cross-section of the latency for random 70 percent read and 30 percent write operations with 4K blocks and queue depths from one to eight.

The results show lower latency for the Intel Optane DC SSDs, for both read and write operations, and at every I/O depth. The lower I/O depths show the greatest differences in latency.

Test Results: RAIDIX ERA vs. MDRAID vs. ZFS*

RAIDIX conducted tests of latency, speed, and throughput for its ERA software compared to MDRAID and ZFS* on a system using NVMe and Intel Optane DC SSDs.[7] Tests were conducted for random and sequential read and write operations at seven ratios from 100 percent read to 100 percent write and nine block sizes from 4K to 1M.

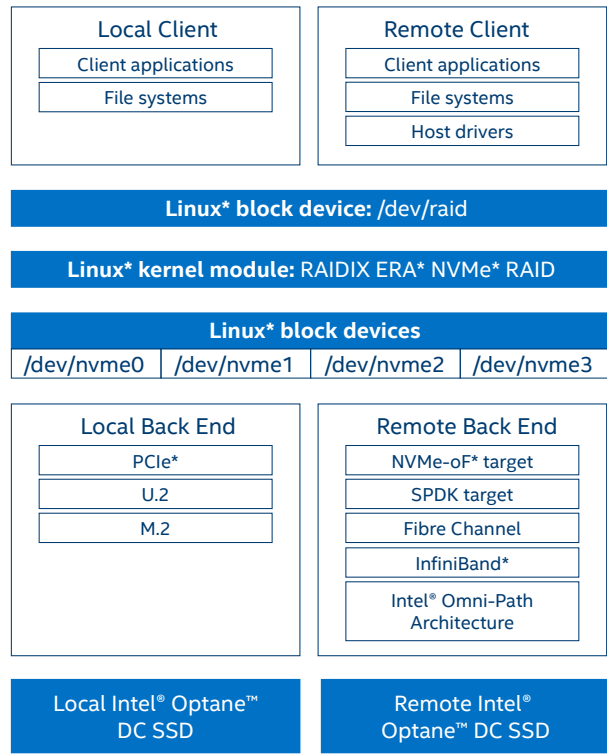Representative results presented here are for random 70 percent read and 30 percent write operations using 4K block size.[8]

Local Client	Remote Client
Client applications	Client applications
File systems	File systems
	Host drivers

Linux* block device: /dev/raid

Linux* kernel module: RAIDIX ERA* NVMe* RAID

Linux* block devices			
/dev/nvme0	/dev/nvme1	/dev/nvme2	/dev/nvme3

Local Back End	Remote Back End
PCIe*	NVMe-oF* target
U.2	SPDK target
M.2	Fibre Channel
	InfiniBand*
	Intel® Omni-Path Architecture

Local Intel® Optane™ DC SSD	Remote Intel® Optane™ DC SSD

Figure 1. RAIDIX ERA* architecture with Intel® Optane™ DC SSDs

Intel® Optane™ Technology: Revolutionizing Memory and Storage

Intel Optane technology is a unique combination of Intel Optane memory media with Intel® technology–based advanced system memory controllers, interface hardware, and proprietary software. Together, these building blocks deliver a game changing technology that breaks through storage bottlenecks.

Intel Optane technology provides an unparalleled combination of high throughput, low latency, high quality of service (QoS), and high endurance to unleash vast system performance potential in a range of products.[5]
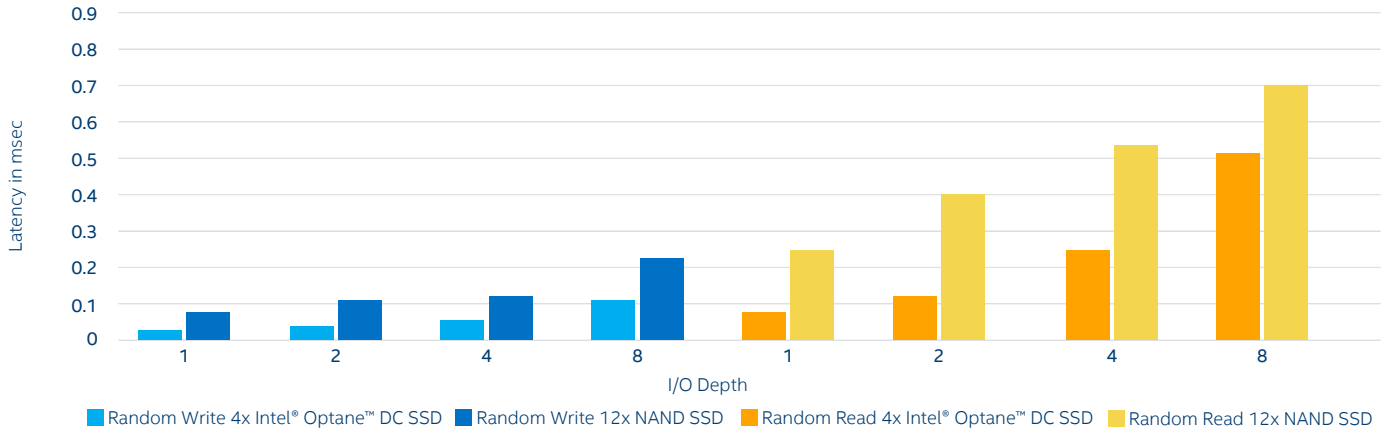
Intel® Optane™ DC SSDs vs. NAND SSDs Latency

Figure 2. RAIDIX ERA* reads and writes faster on Intel® Optane™ DC SSDs than on NAND SSDs[6,7]
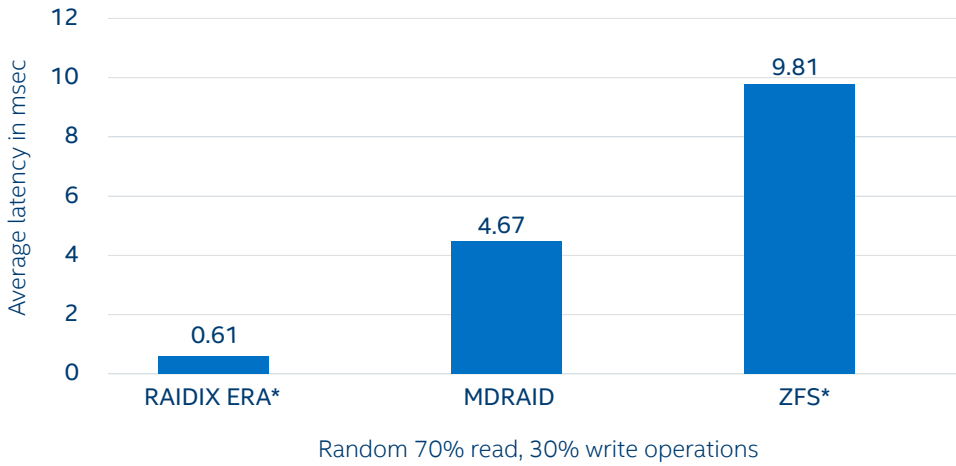
Latency (msec)

Random 70% read, 30% write operations

Figure 3. Random read/write latency is lower for RAIDIX ERA* than for MDRAID or ZFS*[7,8]

Speed (IOPS)

Random 70% read, 30% write operations
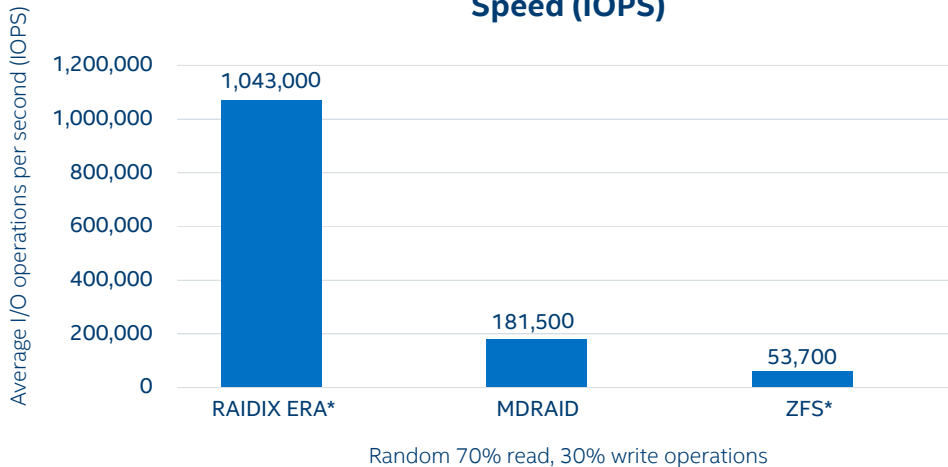
Figure 4. The total I/O operations per second (IOPS) is higher for RAIDIX ERA* than for MDRAID or ZFS*[7,8]
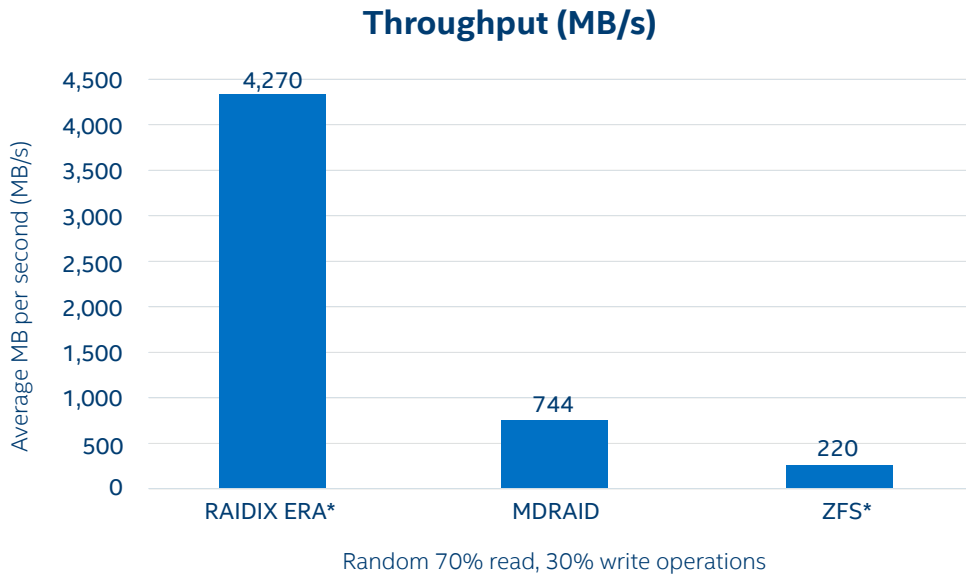
Throughput (MB/s)

Figure 5. Throughput is higher for RAIDIX ERA* than for MDRAID or ZFS*[7,8]

By all three measures, RAIDIX ERA significantly outperformed the open source alternatives. MDRAID performed better than ZFS, but RAIDIX ERA achieved more than 5x advantage over MDRAID.

Faster Together: RAIDIX ERA, NVMe, and Intel Optane Technology

The above results of extensive testing on NVMe systems show conclusively that RAIDIX ERA performs at lower latency on Intel Optane DC SSDs than on NAND SSDs and that, on Intel Optane DC SSDs, RAIDIX ERA software performs much better than MDRAID or ZFS in terms of latency, speed, and throughput. This new software technology is well suited to accelerate storage hardware systems to new levels of high performance.

Learn more about RAIDIX ERA at raidix.com/products/era/.

Learn more about Intel Optane technology at intel.com/optane.

RAIDIX ERA* is suitable for:

- Online analytical processing (OLAP) and online transaction processing (OLTP)

- Real-time editing of 4K/8K/virtual-reality (VR) video content

- High-frequency trading (HFT) and real-time bidding (RTB)

- HPC and technical computing

- Large SQL and NoSQL database processing
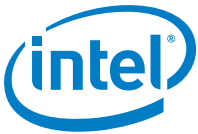
[1] Markets Insider. "G2M Research Releases Fall 2018 NVM Express Market Sizing Report." December 2018.
https://markets.businessinsider.com/news/stocks/g2m-research-releases-fall-2018-nvm-express-market-sizing-report-1027784171.

[2] Ubuntu 16.04 LTS*, Ubuntu 18.04 LTS, SUSE Linux Enterprise Server (SLES) 12 SP3*, and CentOS 7.5* are supported. raidix.com/products/era/.

[3] Read detailed case studies on the RAIDIX* blog: raidix.com/blog/.

[4] Tests were conducted October–December 2018 by RAIDIX, using a Western Digital Serv24* system with 2 x Intel® Xeon® Platinum 8160 processors at 2.10 GHz (Intel® Hyper-Threading Technology enabled, C-states disabled), 256 GB DDR4 RAM, and 24 x Ultrastar DC SN620* (3.2 TB/1.2 drive writes per day [DWPD]).

[5] Intel® technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

[6] Tests were conducted October–December 2018 by RAIDIX. For this test, the following NAND-based server was used for comparison:

NAND-based test server specification:

- Intel® Server System R2224WFTZS + 2 x Intel® Xeon® Gold 6130 processor at 2.10 GHz with Intel® Hyper-Threading Technology enabled, allowing launch of two computing threads from each core
- All NVM Express* (NVMe*) drives were connected through 3 F2U8X25S3PHS backplanes
- 12 Intel® SSD DC D3700 NVMe drives
- 1 x 960 GB Intel® SSD DC S3520 Serial ATA (SATA) boot drive
- 1 x 40Gb Intel® Ethernet Network Connection XL710-QDA2
- 384 GB RAM DDR4-2666

[7] Tests were conducted October–December 2018 by RAIDIX. Each test loop ran until a steady state was reached, with a maximum of 12 rounds.

Test server specification:

- Intel® Server Board S2600WFD + 2 x Intel® Xeon® Gold 6138 processor at 2.00 GHz
- 4 x 375 GB Intel® Optane™ DC P4800X NVM Express* (NVMe*) drives
- 1 x 960 GB Intel® SSD DC S3520 Serial ATA (SATA) boot drive
- 1 x 10Gb Intel® Ethernet Converged Network Adapter X772 with 10GBASE-T dual-port
- 1 x 10Gb Intel® Ethernet Network Connection OCP X527-DA2 dual-port
- 1 x 40Gb Intel Ethernet Network Connection XL710-QDA2
- 384 GB RAM

FIO* configurator details for the tests:

```
[global]
blocksize= [ "4k", "8k", "16k", "32k", "64k", "128k", "256k", "512k", "1m" ]
bwavgtime=100000
direct=1
iodepth= [ "1", "4", "8", "32", "64", "128" ]
ioengine=libaio
numjobs=32
runtime=120
rw= [ "randrw", "rw" ]
rwmixread=[ "0", "10", "30", "50", "70", "90", "100" ]

group_reporting
```

[8] To request the complete test results, contact request@raidix.com.