Solution Brief

Intel® QLC 3D NAND SSDs
Intel® Optane™ Persistent Memory
Distributed Asynchronous Object Storage (DAOS)

Achieve High-Performance Storage with DAOS

Meet the performance requirements of HPC with Intel® QLC 3D NAND SSDs and Intel® Optane™ technology.

The ways in which organizations use data in high-performance computing (HPC) are changing fast

Exponential data growth is driven, in part, by new analytics capabilities and the convergence of artificial intelligence (AI) and HPC. Those advances, in turn, drive the creation of even more data that must be stored and made instantly available when needed. These changes have resulted in data input/output (I/O) patterns that are increasingly complex, with a mix of reads and writes and wide variability in block sizes. That complexity can result in increased latency that limits HPC capabilities.

With the increasing prevalence of real-time inputs, write performance has become a critical bottleneck that can lead to slow or dropped data feeds. But the increased use of shared data-storage pipelines means any storage clusters that are serving HPC activities might also be concurrently serving other workloads. Now, with more high-speed networks being deployed—100 Gb/second fabric is common, and the adoption of 200 Gb/second fabric is increasing—those responsible for data center budgets don't want to see these significant investments underutilized. Traditional IT storage systems have been designed for rotating media and POSIX I/O, representing a key performance limitation. And legacy storage systems cannot evolve to provide enough read and write throughput to support these new data models and high concurrency environments.

DAOS delivers world-class results

The DAOS architecture makes data-access times possible that can be several orders of magnitude faster than existing storage systems—from milliseconds (ms) to microseconds (µs). In fact, recent IO-500 benchmarking results placed DAOS-based solutions among the top scores, beating some of the best supercomputers in the world.[1]

Meeting HPC storage performance requirements

Distributed Asynchronous Object Storage (DAOS) is the foundation of the Intel exascale storage stack, overcoming the limitations of traditional distributed storage. DAOS is an open source software-defined scale-out object store that is designed to use low-latency, high-message-rate user-space communications that bypass the operating system. DAOS uses multiple tiers of storage that can support any combination of NVM Express (NVMe) solid state drive (SSD) types.

Unlike traditional storage stacks that were primarily designed for rotating media, DAOS is architected from the ground up to make use of new NVM technologies. It is also extremely lightweight because it operates in the user space from end to end. DAOS offers a shift away from an I/O model designed for block-based, high-latency storage to a model that inherently supports fine-grained data access and unlocks the performance of next-generation storage technologies.

Existing distributed storage systems use high-latency peer-to-peer communication, whereas DAOS is designed to use low-latency, high-message-rate user-space communications that completely bypass the operating system. Most storage systems today are designed for block I/O, with all operations going through the Linux kernel using a block interface. While progress has been made in

optimizing access to the block device—coalescing, buffering, and aggregation, for example—those optimizations are not relevant for the next-generation storage devices that Intel is targeting and will add unnecessary overhead. DAOS, on the other hand, is designed to optimize access to Intel® Optane™ persistent memory (PMem) and NVMe SSDs with Persistent Memory Development Kit (PMDK) and Storage Performance Development Kit (SPDK) libraries. That eliminates the unnecessary overhead associated with traditional storage stacks.

Optimizing Intel Optane technology and Intel 3D NAND SSDs

When built with a combination of Intel Optane PMem, Intel Optane SSDs, and Intel 3D NAND storage, DAOS delivers high-bandwidth, low-latency, and high input/output operations per second (IOPS) containers to HPC applications. That enables next-generation data-centric workflows that support simulation, data analytics, and AI. The addition of support for Intel QLC 3D NAND SSDs now gives DAOS implementations the ability to massively scale storage in a cost-effective and operationally efficient manner.

The HPC solution relies on the following Intel memory-storage technologies:

Intel Optane PMem: This revolutionary technology brings big and persistent memory to DAOS solutions, making use of these capabilities to provide byte-addressable persistent storage that supports remote persistent memory (RPMem) access based on traditional remote direct memory access (RDMA). Intel Optane PMem capacity is used to manage metadata operations and absorb a large amount of small block (<64K) host writes to optimize the write pattern before de-staging to quad-level cell (QLC) NAND SSDs. When Intel Optane PMem reaches a set fill level, it de-stages, writing data to the Intel QLC 3D NAND SSDs. This approach gives applications write latencies similar to Intel Optane technology—measured in 10s of microseconds[2]—for small I/O and NAND-like read latencies once the data is de-staged. Persistence means that the data being held in the device is not at risk of being lost before it is de-staged.

Intel Optane SSDs: Upcoming versions of DAOS will include data-layout capabilities that enable the differentiation of data tiers on NVMe based on their characteristics. This opens up new possibilities to differentiate across a range of NVMe media classes such as Intel Optane SSDs and NAND-based SSDs that are available both today and in the future.

Intel QLC 3D NAND SSDs: With writes shaped by DAOS in Intel Optane PMem, QLC write performance and endurance is now optimized, enabling the solution to take advantage of the high-capacity and operational efficiency delivered by the devices in use. Reads are delivered directly from QLC NAND SSDs at low latency, with read-optimized performance that saturates the PCIe 4.0 bus. The combination of DAOS, Intel Optane technology, and Intel QLC 3D NAND SSDs delivers efficient performance per dollar by enabling cost effective saturation of the network at low latency on both the read and write sides in HPC disaggregated storage.

Evaluating DAOS performance

To assess the performance capabilities of DAOS running on its most advanced storage devices, Intel built a reference storage platform consisting of an Intel® Xeon® Gold processor, a 100 gigabit Ethernet (GbE) network interface card, Intel Optane PMem, and Intel QLC 3D NAND SSDs. Intel measured read and write bandwidth and read and write latencies using Flexible I/O (FIO) to evaluate how performance changed as changes were made in the platform.

Low read tail latencies with DAOS 1.0 and Intel QLC 3D NAND SSDs

Depending on when the host issues a read command, reads can be from Intel Optane PMem or Intel QLC 3D NAND SSDs. If the data has not been de-staged to NAND SSDs, the read is from Intel Optane PMem. If it has been de-staged, the read is from the NAND SSDs. DAOS 1.0 read latency with Intel Optane PMem yields sub-40 us latency at four nines (P99.99), and just over 50 us latency at five nines (P99.999). When reading from NAND SSDs, the most common use case is a combination of software (DAOS 1.0) and QLC NAND (including Intel SSD D5-P4326 and Intel SSD D5-P5316) that delivers consistent read tail latency of up to five nines between 200 and 300 µs.

When built with a combination of Intel Optane PMem, Intel Optane SSDs, and Intel 3D NAND storage, DAOS delivers high-bandwidth, low-latency, and high IOPS containers to HPC applications. This enables next-generation data-centric workflows that support simulation, data analytics, and AI. The addition of support for Intel QLC 3D NAND SSDs now gives DAOS implementations the ability to massively scale storage in a cost-effective and operationally efficient manner.

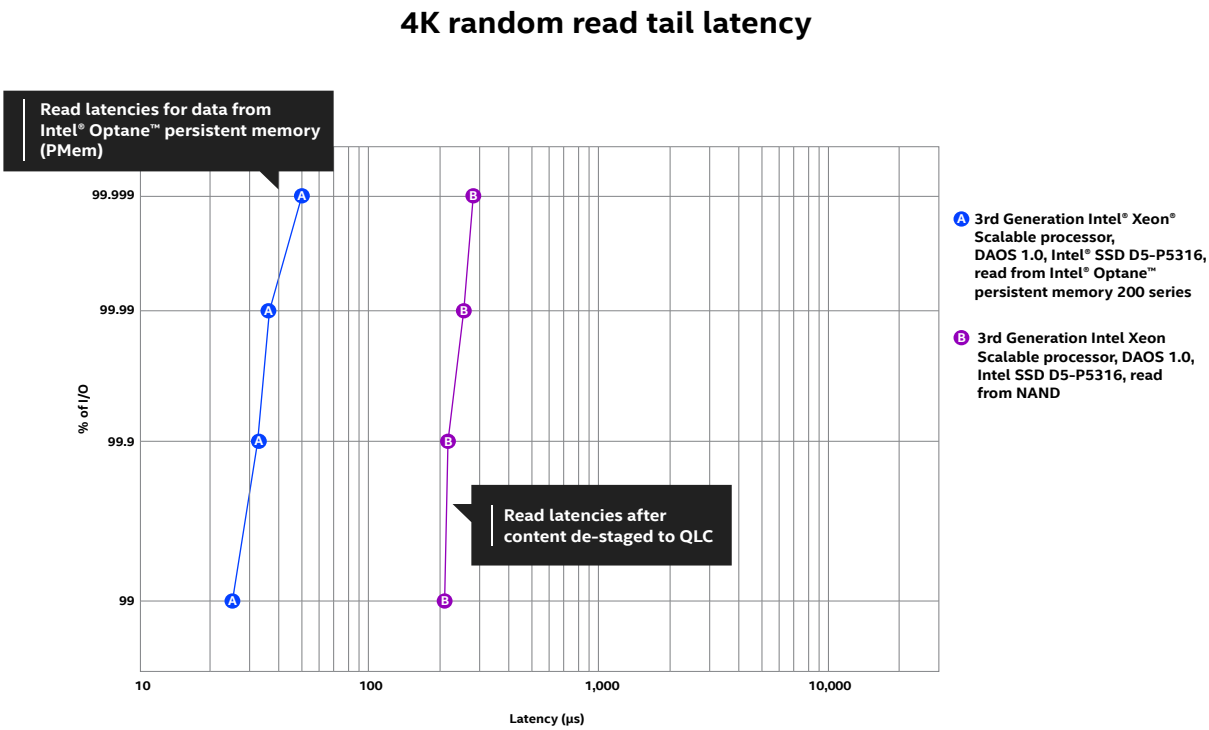

Figure 1. 4K random read latency for DAOS 1.0 reading from Intel Optane PMem and an Intel SSD D5-P5316[3]

Microsecond write latencies with DAOS and Intel Optane PMem

A combination of persistence and large capacity enables DAOS to direct all small host writes to Intel Optane PMem. Figure 2 shows the low latencies possible with DAOS 1.0 and the increase in responsiveness compared to a non-DAOS environment, where host writes are directed to NAND. Low write latencies are crucial to optimizing costly compute resources in intense HPC environments.

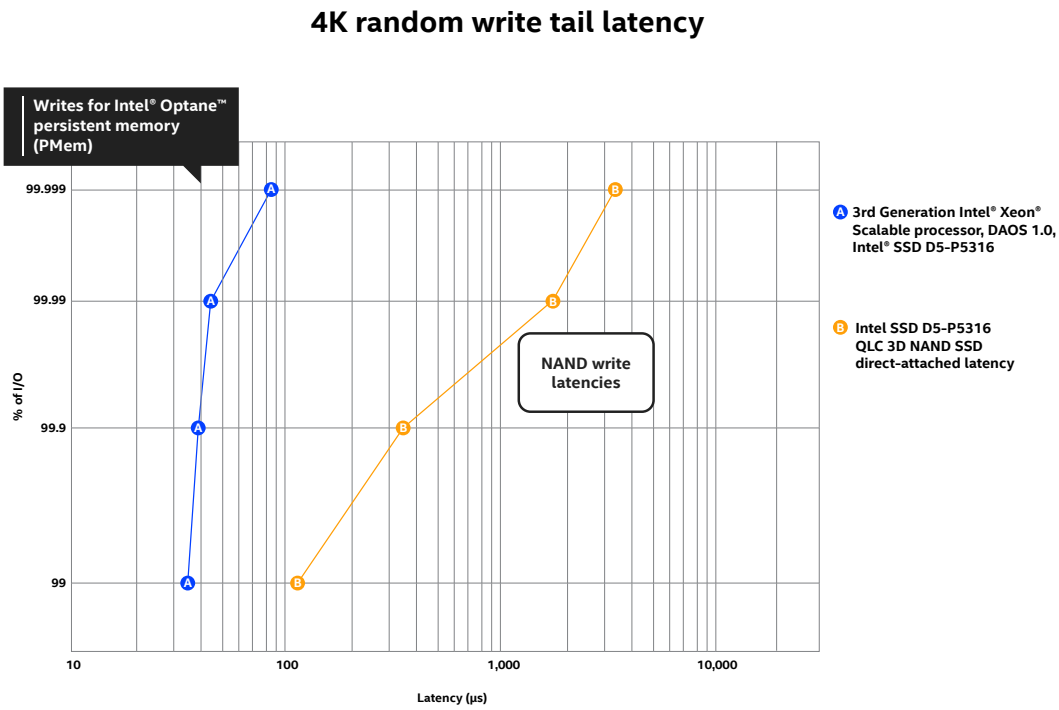

Figure 2. 4K random write tail latency for DAOS 1.0 writing to Intel Optane PMem and Intel SSD D5-P5316[3]

Low read latencies, even in the presence of write pressure

Bulk reads are taken directly from 3D NAND capacity storage, and often those reads will be in the presence of write pressure. Writes can come from many sources—de-staging from Intel Optane PMem, running compression and deduplication algorithms, and drive recovery, for example. Figure 3 shows NAND read latencies when write pressure is successively ratcheted up from zero to 2,500 MB/s. A "keep out zone" is overlaid onto the graph in Figure 3 to illustrate how DAOS and NAND can meet theoretical service-level agreements (SLAs), delivering three nines (P99.9) read tail latency of less than 1 ms and five nines (P99.999) of less than 5 ms, even in the presence of writes. Tail latency becomes increasingly important as distributed storage is scaled out.

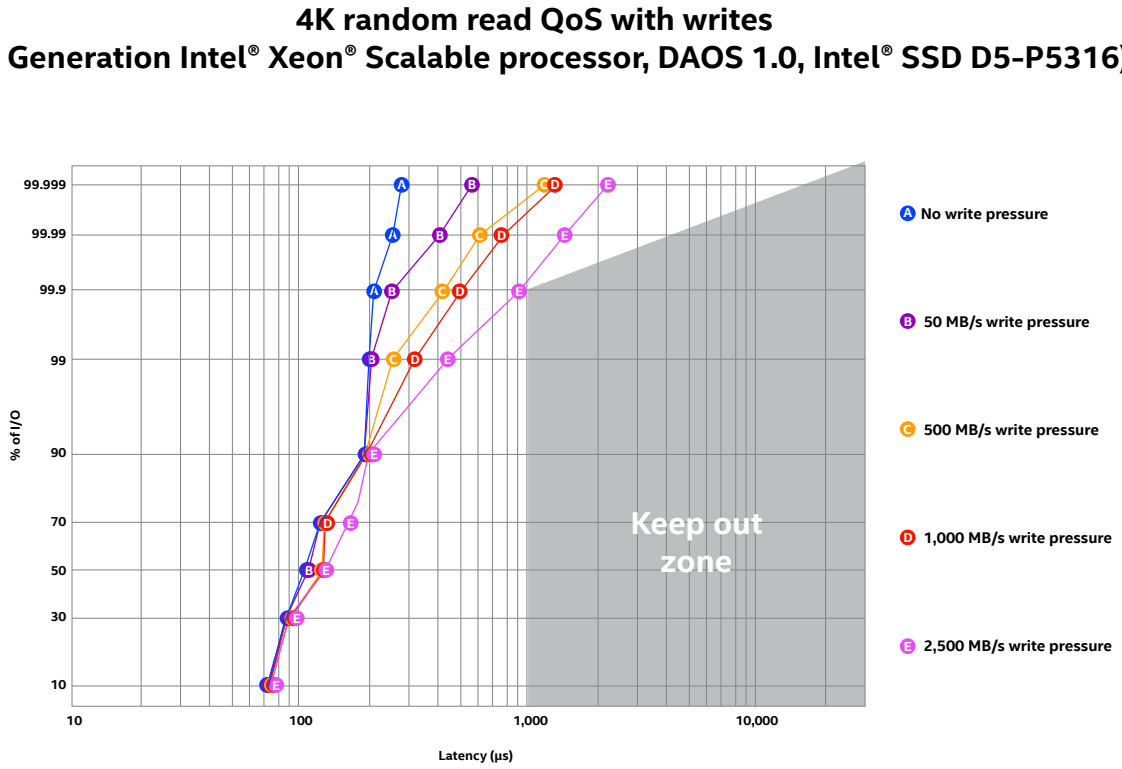

Figure 3. 4K random read QoS with writes for DAOS 1.0 reading from Intel SSD-P5316[3]

Conclusions

HPC is changing fast. Exponential data growth, the convergence of workloads like AI, and system upgrades in response to these changes are stressing traditional approaches to disaggregated storage. DAOS breaks through the limitations of traditional block I/O-and operating system–based storage, tapping into the innovative capabilities offered by Intel Optane technology and Intel QLC 3D NAND SSDs to deliver high-bandwidth, low-latency read and write performance for HPC storage systems.

By utilizing the large capacity and persistence of Intel Optane PMem, DAOS absorbs and shapes all host writes, making class-leading space and operationally efficient capacity storage in Intel QLC 3D NAND SSDs possible. Intel QLC 3D NAND SSDs' performance optimizations deliver read performance that saturates the PCIe 4.0 bus at low latency, delivering read performance levels that open the door for capacity storage scaling that can keep pace with evolving HPC storage needs.

Visit these pages to learn more about Intel storage innovations:

- About Intel 3D NAND SSDs
- About Intel Optane Persistent Memory
- Intel® SSD D5-P5316 product brief
- "QLC NAND Technology Is Ready for Mainstream Use in the Data Center"
- "QLC NAND SSDs Are Optimal for Modern Workloads"

[1] Per IO500 SC20 list, https://io500.org/.

[2] See Figure 2, 4K Random Write Tail Latency.

[3] Results based on Intel testing from April through May of 2021. **Test configurations included:** 1 x NVIDIA Mellanox ConnectX-5 100 GbE ethernet adapter card, 1 x Intel Xeon Platinum 8368Q processor (2.60 GHz), 8 x 128 GB Intel Optane PMem 200 series, 6 x 15.36TB Intel SSD D5-P5316 series. The test configuration workloads were structured as follows: All data collection was executed via FIO, I/O transfer size was 4K with a queue depth (QD) of 1, and write pressure used two single-socket clients.

Printed in USA 0621/JR/PRW/PDF Please Recycle 347309-001US