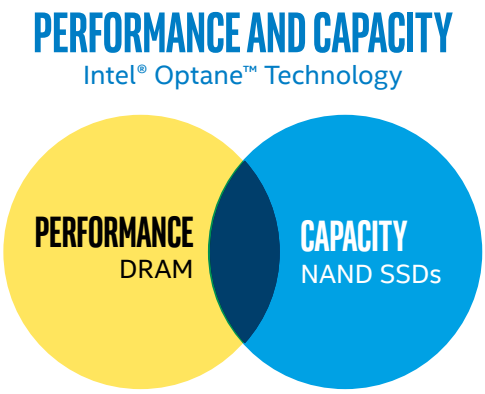
SOLUTION BRIEF

Software-defined Storage
Intel® Optane™ SSDs

# Intel® Optane™ SSDs Boost Red Hat Ceph Storage Cluster Performance

## Using Intel® Optane™ Solid State Drives with Red Hat Ceph Storage can help reduce latency and deliver higher IOPS for faster and more efficient performance on an all-flash cluster[1]




### Performance of DRAM with the Capacity of NAND SSDs

- The unique architecture of Intel® Optane™ DC SSDs enables them to achieve writes at the byte or page level for fast and predictable performance with no need for garbage collection.
- Intel Optane DC SSDs have much higher write endurance compared to Intel® 3D NAND SSDs.[4]
- Over the entire cluster, Intel 3D NAND SSDs and Intel Optane DC SSDs help provide excellent cost savings and price-per-performance.

## Executive Summary

There's no escaping the data explosion. Businesses of all sizes and across all industries must contend with data that continues to grow both in volume and scope from an increasing array of digital sources like artificial intelligence (AI), machine learning (ML), and big data analytics. How organizations manage their data and their ability to extract meaningful insights are powerful competitive differentiators. It is necessary to re-imagine data center storage architectures to keep up with the expanding, overwhelming amount of data.

Today's data-intensive workloads have become more performance-sensitive. Even minimal increases in latency or reductions in throughput can have dramatic effects. Businesses that use Red Hat Ceph Storage with BlueStore are looking for ways to accelerate their clusters, cut latency, and increase IOPS. A recent innovation—Intel® Optane™ technology—can help.

Adding a single Intel® Optane™ DC SSD per node to the cluster for RocksDB and WAL partitions with the option of adding one Intel Optane DC SSD for caching can significantly boost Red Hat Ceph Storage performance by improving the cluster's P99 latency and IOPS.[2] Additionally, organizations can use Intel® Cache Acceleration Software (Intel® CAS) available for Intel® SSDs to increase storage performance by caching frequently accessed data and/or selected I/O classes.

Red Hat performed testing to determine the extent of performance increases possible by coupling Red Hat Ceph Storage with Intel Optane DC SSDs and Intel® Xeon® Scalable processors. They tested a five-node Red Hat Ceph Storage cluster that used NVMe-based Intel® 3D NAND SSDs for bulk storage and an Intel Optane DC SSD for metadata (RocksDB) and the WAL. This configuration delivered up to 13.82 percent lower P99 latency and up to 9.55 percent higher IOPS compared to a similar configuration that did not include the Intel Optane DC SSD, depending on workload.[3]

## Boosting Red Hat Ceph Storage Cluster Performance

Red Hat Ceph Storage is growing in popularity and it's no wonder. It's open source, software-defined, enterprise-class, and unified—all in a single platform. However, organizations realize they need to squeeze out every ounce of performance as they strive to achieve the lowest possible latency and fastest throughput to handle the increasing demands of current—and future—workloads. Performance was recently identified as a critical requirement by 63 percent of Ceph users.[5]

## Red Hat Ceph Block Storage: Compelling Use Cases

While Red Hat Ceph Storage was initially used primarily for object storage-based solutions, it's now also deployed for block-based storage. A growing number of users—including businesses and government agencies—are taking advantage of Red Hat Ceph Storage's block storage capabilities for a growing variety of use cases. Block storage can be invaluable in situations where low latency and maximum performance are important. Some of the most compelling current use cases include cloud infrastructure, data analytics, media repositories, and back and restore systems. Adding Intel® Optane™ DC SSDs to an existing HDD cluster can raise the performance level of these uses even higher. In addition, users can use Intel Optane SSDs to accelerate even all-flash clusters configured with NVMe-based Intel® 3D NAND SSDs.

## The Right Intel® SSDs Suit Varying Objectives

Intel® SSDs can provide significant business value when added to Red Hat Ceph Storage. In particular, NVMe-based Intel 3D NAND SSDs are an excellent choice for bulk storage, while Intel Optane DC SSDs are most beneficial when they're used for the metadata tier (RocksDB) and write-ahead log (WAL), and for caching object storage daemons (OSDs).

**CEPH USERS WANT PERFORMANCE**[5]

63 percent of Ceph users identified performance as a top need going forward.

## Spotlight on QLC NAND Technology

Another potential approach to configuring Red Hat Ceph Storage clusters would be to use Intel® QLC 3D NAND SSDs for bulk storage and an Intel® Optane™ SSD for metadata. A recent storage innovation, QLC NAND provides an alternative to its predecessor, TLC NAND. Compared to TLC NAND drives, QLC NAND drives have 33 percent higher storage density, with four 0 or 1 states in each flash cell.[6] Deciding which generation is best for any particular application depends on what's considered most important—density or endurance. QLC is denser, so it can help reduce the cost of storage. However, QLC has limited write endurance, so it's best-suited for use cases where read operations far outpace the number of write operations.[7]

You may be able to achieve the following objectives by adding the right Intel SSD to your Red Hat Ceph Storage cluster:

- **Boost throughput.** Using Intel Optane SSDs for RocksDB and the WAL on Red Hat Ceph Storage clusters can increase IOPS per node and lower P99 latency.
- **Optimize performance.** Intel Optane SSDs can also be used as the cache for a TLC NAND flash array.

On a five-node Red Hat Ceph Storage cluster with an all-flash NVMe-based capacity tier, adding a single Intel® Optane™ SSD DC P4800X for RocksDB/WAL/cache reduced P99 latency by up to 13.82 percent and increased IOPS by up to 9.55 percent compared to the five-node cluster without an Intel Optane SSD (see Figure 1).[8]

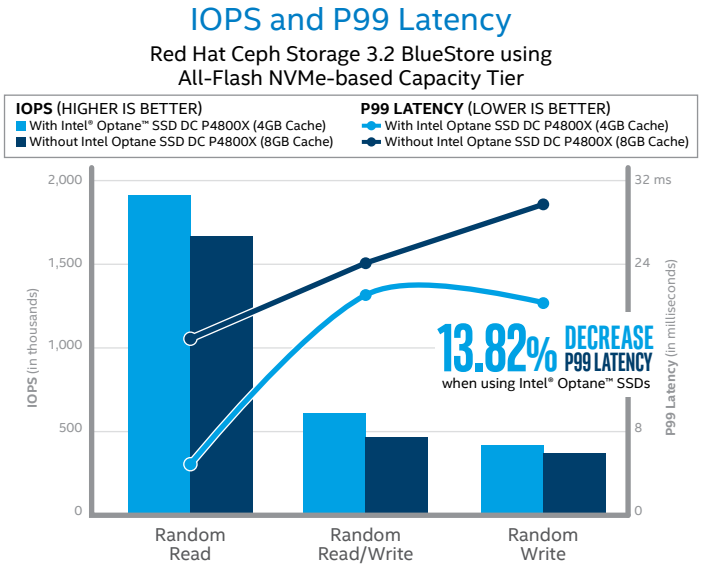



**Figure 1.** Using Intel® Optane™ SSDs with Red Hat Ceph Storage 3.2 BlueStore on all-flash achieves up to a 9.55 percent increase In IOPS and up to a 13.82 percent decrease in P99 latency, compared to a similar configuration without Intel Optane SSDs.[9] (Source: ceph.com/community/part-4-rhcs-3-2-bluestore-advanced-performance-investigation)

## Solution Architecture: Optimizing Red Hat Ceph Storage Clusters

Intel Optane DC SSDs combine Intel® Optane™ memory media (non-volatile), an advanced system memory controller, interface hardware, and firmware. These SSDs are Intel's latest non-NAND-based class of storage.

A typical Red Hat Ceph Storage node uses the same media for all hot and cold data (see Figure 2). Using Intel Optane SSDs for RocksDB, WAL, and optional OSD caching can potentially improve cost efficiency and cluster performance.

Another benefit of Intel Optane SSDs is node reduction. Intel Optane SSDs can help reduce your node count to less than half required by an all-flash solution. Intel Optane SSDs can maintain up to 63x lower latency than a 3D NAND at high pressure due to consistently lower read latency under increasing write pressure.[10] In addition, because they have an endurance rating of up to 60 device writes per day,[11] fewer Intel Optane SSDs are needed, compared to 3D NAND SSDs, to handle the soaring demands of Ceph, metadata, and caching.

It's simple to implement a cache using Intel Optane SSD DC P4800X. Intel® Cache Acceleration Software (Intel® CAS), enterprise-quality software that runs on Linux or Windows platforms, has been optimized for Intel® SSD Data Center Family performance. By caching frequently-accessed data and/or selected I/O classes, Intel CAS can accelerate storage performance.

## Intelligent Caching Speeds Performance

Now, there's a unique way to cache based on I/O classification. Open Cache Acceleration Software (Open CAS) gives you the ability to cache even the "hottest" data—like metadata. Open CAS is an open source project encompassing block caching software libraries, adapters, tools, and more. The main goal of this cache acceleration software is to accelerate a backend block device(s) by utilizing a higher performance device(s).

Open CAS Framework (OCF) is at the core of Open CAS. Open CAS provides adapter implementations for Linux operating systems and for Storage Performance Development Kit (SPDK) applications. Besides OCF, Open CAS also includes Open CAS Linux and SPDK OCF Block Device. Some of the important features and capabilities of Open CAS include:

- Boosts performance with a very small cache
- Offers the ability to cache small random blocks or files with I/O classification
- Is validated on common Linux distributions and kernels and Windows Server
- Offers several caching modes and cache-cleaning policies, which allows tuning to the user's workload
- Includes support for atomic writes and TRIM operations
- Can be used with Intel® Optane™ DC SSDs

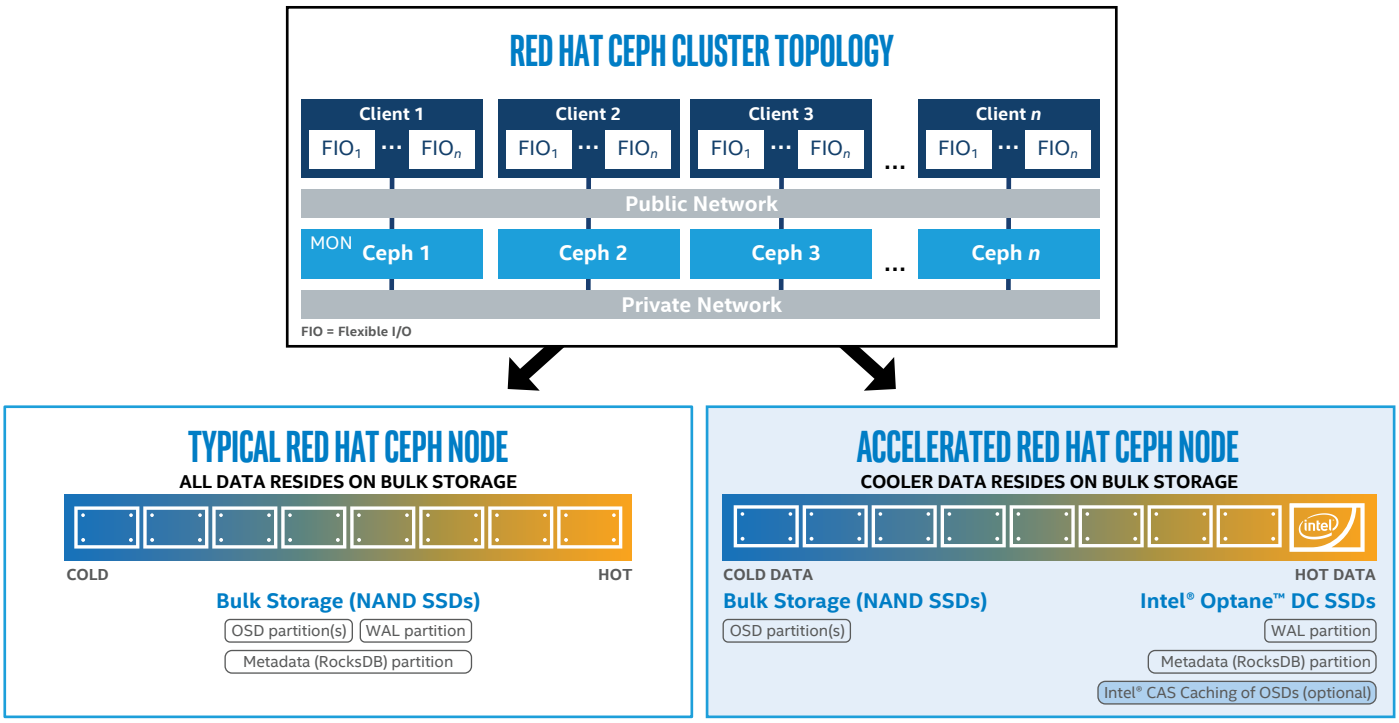



**Figure 2.** Intel® Optane™ SSDs can improve the performance of all-flash Red Hat Ceph Storage clusters.

## Conclusion

You can improve performance in Red Hat Ceph Storage deployments by adding Intel Optane SSDs, especially if you use the SSDs for RocksDB, WAL, and optional OSD caching. Even a small number of Optane SSDs as an accelerator can boost performance of all-flash clusters. Adding Intel Optane SSDs can help to significantly reduce latency and increase throughput for more efficient data storage.

Find the right solution for your organization. Contact your Intel representative or visit **intel.com/optane**.

## Learn More

**Webinar**

- How to Improve vSAN and Ceph Performance While Reducing Cost with Intel® Optane™ DC SSDs

**Websites**

- Red Hat Ceph Storage
- Intel® Optane™ SSD Data Center P4800X Series
- Intel® Xeon® Scalable processors
- Open Cache Acceleration Software

**White Paper**

- Red Hat Ceph Storage on Servers with Intel Processors and SSDs
- Using Intel® Optane™ Technology with Ceph to Build High-Performance OLTP Solutions
- Using Intel® Optane™ Technology with Ceph to Build High-Performance Cloud Storage Solutions on Intel® Xeon® Scalable Processors
- Use Intel® Optane™ Technology and Intel® 3D NAND SSDs to Build High-Performance Cloud Storage Solutions

**Solution Provided By:**

Red Hat intel

[1] Testing by Red Hat as of July 2019. Performance results are based on testing as of July 2019 and may not reflect all publicly available security updates. No product or component can be absolutely secure. **Common Configuration for 5-node Red Hat Ceph Storage (RHCS) cluster:** Chassis = Cisco UCS C220-M5SN Rack Server; CPU = 2x Intel® Xeon® Platinum 8180 processor (28 cores, 56 hyper-threaded cores, 2.50 GHz, 196 GB memory); NIC = Cisco UCS VIC 1387 2 port x 40 GbE; Ceph Pool Placement Groups = 4096; software = Red Hat Enterprise Linux (RHEL) 7.6, Linux Kernel 3.10, RHCS 3.2 (12.2.8-52). **Common Configuration for 7 Client Nodes:** Chassis = Cisco UCS B200 M4 blade servers; CPU = 2x Intel® Xeon® CPU E5-2640 v4 (2.40 GHz, 528 GB memory); NIC = Cisco UCS VIC 1387 2 port x 10 GbE; software = Red Hat OpenStack Platform 10, RHEL 7.6, Linux Kernel 3.10. **Workload:** Pbench-FIO 3.3, block size = 8K, 84 x RBD volumes, I/O depth = 32. **Baseline Configuration:** 7x Intel® SSD DC P4500 4.0 TB with Ceph data and RocksDB/WAL co-located. IOPS Results: Random read = 1,034,000 IOPS; Random write = 516,000 IOPS; Random read/write = 371,900 IOPS. **P99 latency results:** Random read = 16.56 ms; Random write = 14.48 ms; Random read/write = 26.91 ms. **Test Configuration:** 7x Intel® SSD DC P4500 4.0 TB for Ceph data and 1x Intel® Optane™ SSD DC P4800X 375 GB for RocksDB/WAL and OSD caching; **Results:** Random read = 1,076,000 IOPS; Random write = 565,300 IOPS; Random read/write = 398,800 IOPS. **P99 latency results:** Random read = 16.18 ms; Random write = 13.9 ms; Random read/write = 23.19 ms.

[2] See endnote 1.

[3] See endnote 1.

[4] Intel: Endurance ratings. intel.com/content/www/us/en/products/memory-storage/solid-state-drives/data-center-ssds/optane-dc-p4800x-series/p4800x-750gb-2-5-inch.html

[5] Ceph, July 2018, "Ceph User Survey 2018," (page 56). ceph.com/wp-content/uploads/2018/07/Ceph-User-Survey-2018-Slides.pdf

[6] Intel, 33 percent more bits per cell. TLC contains 3 bits per cell and QLC contains 4 bits per cell. Calculated as (4-3)/3 = 33% more bits per cell.

[7] ComputerWeekly.com, May 2018, "Storage 101: The final flash generation? QLC vs MLC, TLC, SLC." computerweekly.com/feature/Storage-101-The-final-flash-generation-QLC-vs-MLC-TLC-SLC

[8] See endnote 1.

[9] See endnote 1.

[10] Tested by Intel. Average read latency measured at queue depth 1 during 4K random write workload. Measured using FIO 3.1. **Common Configuration:** Intel® 2U Server System; OS = CentOS 7.5, kernel 4.17.6-1.el7.x86_64; CPU = 2x Intel® Xeon® Gold 6154 processor @ 3.0 GHz (18 cores), RAM = 256 GB DDR4 @ 2666 MHz; Intel® Microcode = 0x2000043; System BIOS = 00.01.0013; ME Firmware = 04.00.04.294; BMC Firmware = 1.43.91f76955; FRUSDR = 1.43. **Test Configuration:** Intel® Optane™ SSD DC P4800X 375 GB compared to Intel® SSD DC P4600 1.6 TB. SSDs tested were commercially available at time of test. The benchmark results may need to be revised as additional testing is conducted. Performance results are based on testing as of July 2019 and may not reflect all publicly available security updates. No product can be absolutely secure.

[11] See endnote 4.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

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.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit intel.com/benchmarks.

Intel does not control or audit third-party data. You should review this content, consult other sources, and confirm whether referenced data are accurate.

Intel, the Intel logo, Optane, and Xeon are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

Other names and brands may be claimed as the property of others. © Intel Corporation 1019/JGAL/KC/PDF 340757-001US