RED HAT®
STORAGE

RED HAT CEPH STORAGE AND INTEL CACHE ACCELERATION SOFTWARE

Accelerating object storage with the Intel SSD Data Center family

SOLUTION OVERVIEW

Achieve flash-accelerated object storage performance at lower costs than proprietary all-flash array solutions.

Use the Intel SSD Data Center family and Intel CAS to intelligently prioritize I/O for caching.

Confidently deploy software-defined Red Hat Ceph Storage for large object-storage workloads backed by extensive Red Hat and Intel testing.

INTRODUCTION

To manage massive data growth, organizations are increasingly choosing object storage solutions, allowing them to scale storage flexibly while controlling costs. Ceph is a popular solution, letting organizations deploy cost-effective industry-standard hardware as a part of proven software-defined storage infrastructure. With this shift, the storage media itself has become a key consideration. Traditional hard disk drives (HDDs) are affordable, but often lack the desired input/output (I/O) performance for demanding workloads, such as storing large numbers of objects. Proprietary all-flash arrays offer performance, but can be cost-prohibitive for large-scale deployments.

Red Hat® Ceph Storage combined with Intel® Solid State Drives (SSDs) Data Center family and Intel® Cache Acceleration Software (CAS) has emerged as a compelling option. Organizations can use Intel CAS to selectively classify key portions of a given I/O workload for acceleration with the high-performance and low latency of flash storage. The performance difference can be remarkable. When testing large object-count storage workloads, Red Hat saw performance improvements of up to 400% for small-object writes when using Intel SSDs and Intel CAS.[1] The solution is also cost-effective, achieving strong performance results with only 2-4 SSDs per system.

INTEL SSDS FOR OPTIMIZING CEPH OBJECT STORAGE WORKLOADS

Flash memory has long been recognized for its ability to accelerate both I/O operations per second (IOPS) and throughput in software-defined storage technologies like Ceph. The Intel SSD Data Center family is optimized for performance, reliability, and endurance, making it an ideal match for Red Hat Ceph Storage and object storage workloads. High-performance, low-latency Intel SSDs can serve multiple purposes and boost performance in Ceph Storage deployments in a number of ways:

- **Ceph object storage daemon (OSD) write journals.** Ceph OSDs store objects on a local filesystem and provide access over the network. Writes to Ceph Storage OSDs are completed only when they are written to the OSD's journal, creating a potential HDD bottleneck. Placing OSD journals on Intel SSDs can accelerate writes as well as overall throughput.[1]
- **Ceph Bucket index pool.** Using flash media to store the Ceph bucket index pool can dramatically boost overall object write operations per second (OPS), since flash media is significantly faster than spinning media. This configuration choice also removes the index update load from HDDs. Because the bucket index pool does not store actual object data, it does not require a large amount of storage capacity. As such, utilizing the excess capacity on flash devices employed as OSD write journals is a cost-effective way to boost small-object performance.
- **Improving HDD seek times with Intel CAS.** Ceph relies heavily on the stability and performance of the underlying filesystem. As clusters grow to hundreds of millions of objects, filesystem metadata I/O can become a bottleneck, hindered by slow HDD seek times. Intel CAS can be used to selectively cache extended filesystem (XFS) metadata, effectively masking the performance of slower HDDs. These efficiencies can dramatically increase throughput while reducing latency.

1 redhat.com/en/resources/ceph-storage-on-qct-servers-reference-architecture

Based on pioneering work by Intel Labs, Intel CAS provides performance optimization based on I/O classification—allowing selective acceleration of particular data or metadata.

INTEL CAS AND RED HAT CEPH STORAGE

Intel CAS increases storage performance via intelligent caching and is designed to work with high-performance Intel SSDs.[2] Unlike inefficient conventional caching techniques that rely solely on data temperature, Intel CAS employs a classification-based solution that intelligently prioritizes I/O. This unique capability allows organizations to further optimize their performance based on I/O types (e.g., data versus metadata), size, and additional parameters.

The advantage of Intel's approach is that logical block-level storage volumes can be configured with multiple performance requirements in mind. For example, the filesystem journal could receive a different class of service than regular file data, allowing workloads to be better tuned for specific applications. Efficiencies can be increased incrementally as higher-performing Intel non-volatile memory express (NVMe) SSDs are used for data caching. Intel CAS is depicted logically in Figure 1, in the context of Ceph OSDs.

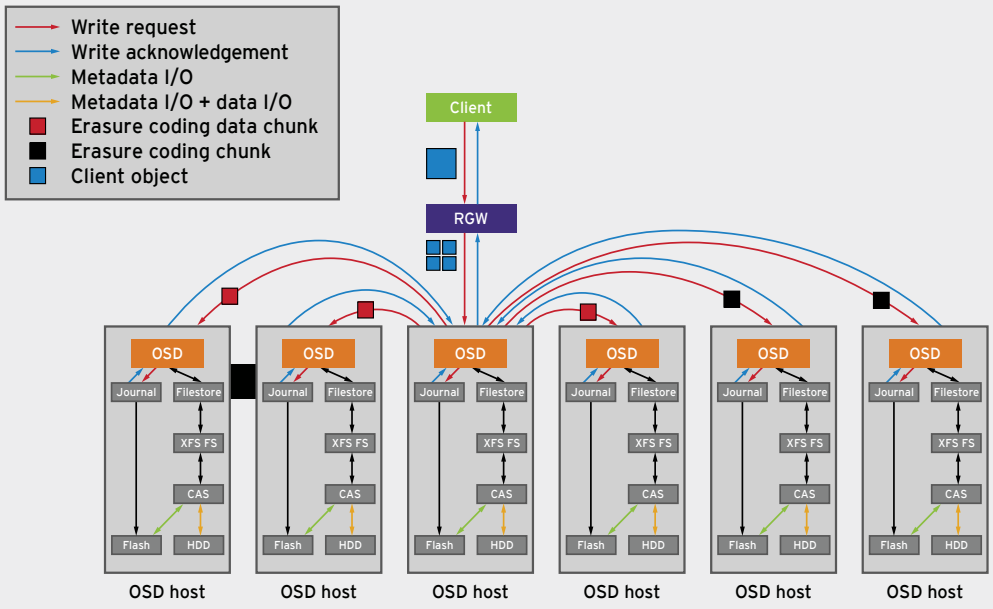

Figure 1. Intel CAS helps accelerate Ceph object stoarge workloads via the RADOS gateway (RGW) interface.

Intel CAS also features several enhancements to improve deployment and usability for Red Hat Ceph Storage installations. The software uses a small default memory footprint that can be further reduced by using a feature called selective cache line size, offering potential cost savings benefits with higher-density storage servers. Caching mode can be changed on the fly, with immediate effect. Other features, such as in-flight upgrades, allow Intel CAS to be upgraded without stopping applications or rebooting servers, yielding further operational efficiencies.

2 *Intel Cache Acceleration Software is licensed on a perpetual basis per SSD and includes one year of support at the time of license. To learn more, visit intel.com/CAS.*

In Red Hat testing, Intel CAS provided up to 400% better performance for small-object (64KB) writes, while providing better latency than other approaches.[1]

INTEL CAS ACCELERATES LARGE OBJECT-COUNT WORKLOADS

Object storage cluster performance can vary significantly based on total cluster object count. To characterize this important dynamic, Red Hat conducted hundreds of hours of testing to fill and test a sample cluster with over 100 million objects.[1] Testing then recorded the highest measured throughput and latency with these higher object counts.

The study revealed that filesystem metadata lookups become increasingly burdensome to performance as clusters grow to support object counts in the tens to hundreds of millions of objects. With all other variables held constant, Red Hat tested a variety of filesystem configurations to evaluate the effect of caching filesystem metadata on small-object (64KB) reads and writes as compared to tuning filestore split/merge values. Workloads were comprised of 130 million objects, and testing was run for 49 hours to evaluate longer-term performance trends. Red Hat sought to determine which of the following configurations provided the best performance and lowest latency:

- Default Ceph OSD filestore.
- Tuned Ceph OSD filestore (with higher split/merge threshold values).
- Default Ceph OSD filestore with Intel CAS for filesystem metadata caching.
- Tuned Ceph OSD filestore with Intel CAS for filesystem metadata caching.

The default Ceph OSD filestore with Intel CAS provided the optimal overall configuration, holding performance steady even as the cluster grew to 130 million objects. As expected, the results were most pronounced for workloads comprised of small-object (64KB) reads and writes. With approximately 60 million objects in the cluster, the Intel CAS configuration improved client write operations per second by over 400% compared to the default filestore configuration (2,400 versus 475 operations per second—see Figure 2).

As shown in Figure 3, client write OPS latency held steady as the cluster grew to 100+ million objects, and Intel CAS provided a 100% better (lower) latency result compared to the default Ceph OSD filestore configuration. Together, these results demonstrate that while split/merge tuning can improve performance over the default filestore configuration, using Intel CAS alone for filesystem metadata caching provides the best results in terms of both performance and latency.

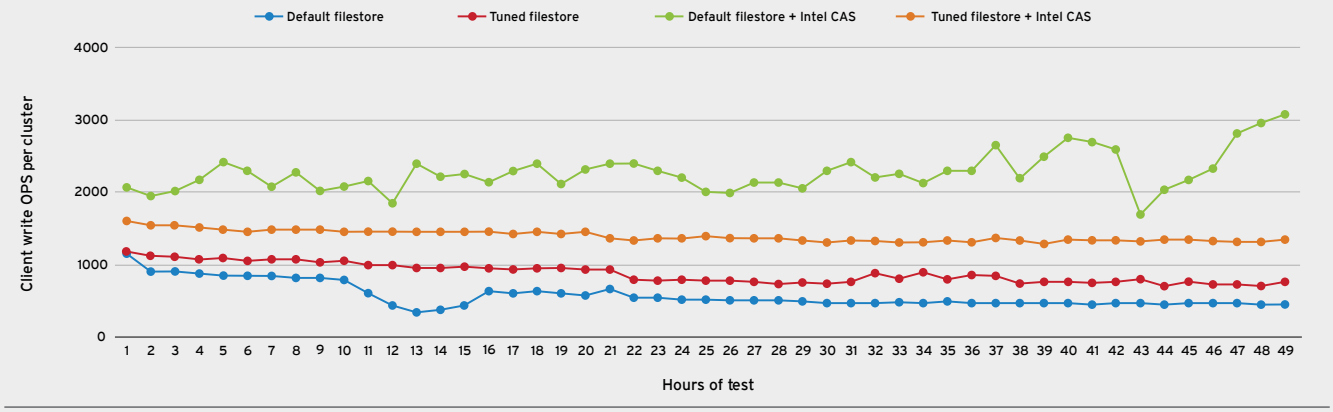

Figure 2. Write OPS per cluster for various filesystem configurations on a six-node cluster, 64KB objects (higher is better).

Learn more

For full details on Red Hat object storage testing that utilized Intel CAS and Intel SSDs for the datacenter, see the Red Hat Ceph Storage: Scalable Object Storage on QCT Servers reference architecture at:

redhat.com/en/resources/ceph-storage-on-qct-servers-reference-architecture

For more information on Intel SSDs for the data center and Intel CAS, visit:

intel.com/go/ssd

intel.com/cas

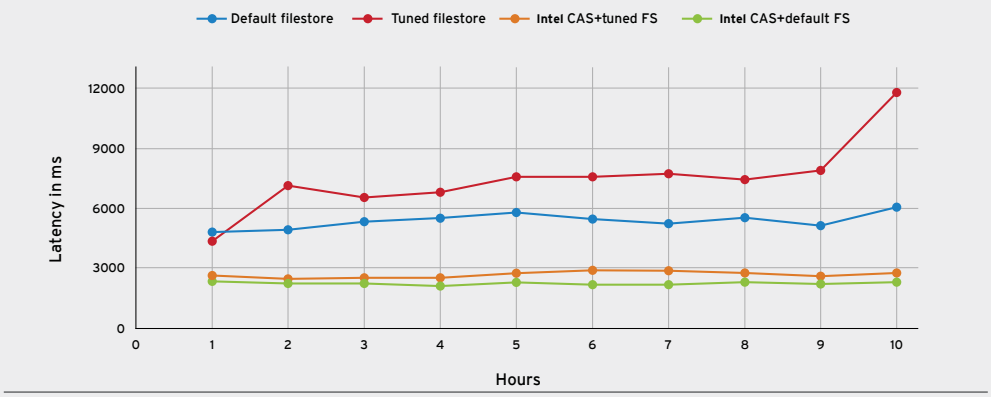

Figure 3. 99th percentile write latency during the first 10 hours of testing, six-node cluster, 64K object size (lower is better).

CONCLUSION

The combination of Intel SSDs for the datacenter and Intel CAS presents an opportunity to optimize and accelerate object storage performance on software-defined Red Hat Ceph Storage clusters. With sufficient throughput and capacity, SSDs can be effectively shared across different optimization methodologies, allowing flash-accelerated Ceph OSD journaling, cached Ceph bucket indexes, and filesystem acceleration via Intel CAS, with only a few SSDs per server. Together, these optimizations help deliver flash-enabled object storage performance without the significant costs of proprietary all-flash arrays.

ABOUT RED HAT

Red Hat is the world's leading provider of open source software solutions, using a community-powered approach to provide reliable and high-performing cloud, Linux, middleware, storage, and virtualization technologies. Red Hat also offers award-winning support, training, and consulting services. As a connective hub in a global network of enterprises, partners, and open source communities, Red Hat helps create relevant, innovative technologies that liberate resources for growth and prepare customers for the future of IT.

facebook.com/redhatinc
@redhatnews
linkedin.com/company/red-hat

NORTH AMERICA
1 888 REDHAT1

EUROPE, MIDDLE EAST, AND AFRICA
00800 7334 2835
europe@redhat.com

ASIA PACIFIC
+65 6490 4200
apac@redhat.com

LATIN AMERICA
+54 11 4329 7300
info-latam@redhat.com

redhat.com
#F9880_0218