SOLUTION BRIEF

Yahoo* Accelerates Scale-out Storage Performance with Intel® CAS 3.0

Pathfinding program with Intel leads to breakthrough solution

"By moving to Ceph*, we were able to reduce our storage costs and increase scalability. Adding Intel® CAS 3.0 on Intel® SSD Data Center Family for PCIe* enabled us to increase performance exponentially and use Ceph to process our warm as well as hot data."

– Ruiping Sun, Senior Principal Architect, Yahoo*[1]

Introduction

Escalating storage costs and the demand for rapid scalability are driving many enterprises to adopt open-source, scale-out storage solutions such as Ceph* and Swift*. While these solutions deliver numerous benefits, they also present performance challenges. Through a pathfinding program with Yahoo*, Intel has developed a compelling solution that enables enterprises to overcome many of those challenges.

The combination of Intel® Cache Acceleration Software (CAS) 3.0 and Intel® SSD Data Center Family for PCIe* has enabled Yahoo to drive a better user experience — doubling throughput, decreasing latency up to 75 percent, reducing error recovery time by 70 percent, and cutting the performance impact during recovery by 50 percent in its Ceph environment, all without need to migrate data or make major changes to the storage infrastructure.[2]

This high-performing scale-out storage solution—derived from groundbreaking work by Intel® Labs and the Intel® Non-Volatile Memory Solutions Group (NSG)—is broadly applicable to other enterprises and the entire software-defined storage ecosystem.

Storage Growth Presents Challenges

The digital universe is doubling in size every two years, with growth expected to reach 44 zettabytes by 2020.[3] To keep pace with this exponential data growth, enterprises are devoting increasing amounts of their IT budgets to data storage, which is contributing to the 2 percent to 4 percent annual increase in IT spending.[4]

Yahoo's challenges exemplify this dilemma. Every day, over 1 billion people use Yahoo Mail*, Flickr*, Tumblr*, and other free Yahoo applications. Quickly and efficiently processing the data for those applications poses a significant challenge, as the company must store exponentially more data each year while also containing—and even reducing—storage costs.

The solution for many enterprises worldwide, including Yahoo, has been to move away from traditional NAS and SAN solutions toward software-defined, scale-out storage solutions. In 2014 Yahoo moved its storage requirements to Ceph, which led to numerous benefits typical of scale-out storage solutions, including greater scalability and a 50 percent reduction in capital costs.[1]

Intel® Cache Acceleration Software (CAS) 3.0

Intel CAS 3.0 provides a unique, intelligent way to look deep into the kernel, perform detailed analytics, label I/O based on predetermined classifications, and prioritize the order in which items will be evicted from the cache. Other benefits of Intel CAS include:

- Transparent to users and applications
- Performance levels nearly as high as if SSDs replaced all hard drives
- Extra performance at critical times to key applications
- Several caching modes and capabilities for specific work loads and environment optimization
- An Intel CAS software license is included with Intel® SSD Data Center Family for PCIe* .

However, open-source, scale-out storage solutions such as Ceph and Swift can also present performance challenges. In particular, Yahoo found that Ceph performed well with cold (rarely accessed) data, such as year-old email attachments. But throughput and latency posed significant problems when attempting to process hot, IOPS-centric data, such as same-day email attachments.

To keep storage costs down and provide customers with the instant access they expect from its industry-leading applications, Yahoo needed to find a way to process its warm and hot data faster and more efficiently.

Pathfinding Program Leads to Breakthrough Solution

Intel Labs is the research arm of Intel that pushes the boundaries of possibility with far-reaching technological innovations. Since 2008, Intel Labs has been researching how to improve caching in enterprise storage systems, leading to advancements including the invention of Differentiated Storage Services (DSS).[5] In late 2014, Yahoo executives learned of DSS at an Intel® Developer Forum and soon entered a pathfinding engagement to explore the potential of the technology to solve some of the performance challenges the company was experiencing with Ceph.

The potential benefits of DSS can be seen by taking a closer look at erasure coding, which Yahoo uses in its Ceph environment to reduce storage costs and boost utilization of storage. Erasure coding breaks data into fragments that are stored across distributed storage environments. Overall performance and customer experience depends on rapidly searching through thousands of inode blocks, which store information about files and directories, and performing several file system metadata accesses to find the fragments and reconstruct each object as quickly as possible. Performance is bottlenecked by the slowest disk I/O.

DSS includes an I/O classification feature that can speed up erasure coding by providing granular insight into the type of I/Os occurring in the cache. For instance, DSS can indicate if a block contains file system metadata (e.g., super block, inode, directory entry) or is part of a regular file. Labeling I/O in this manner enables DSS to analyze storage traffic and prioritize different items in the cache, discarding them as necessary in reverse-priority order.

Over the course of several months, Intel Labs worked closely with Yahoo to optimize DSS to address Yahoo's specific storage challenges. Eventually, with the help of the Intel NSG, the DSS technology was incorporated into what became Intel CAS 3.0. When combined with Intel® SSD Data Center Family for PCIe, Intel CAS 3.0 allows Yahoo to prioritize and selectively cache its data, leading to dynamic performance improvements in erasure coding and other storage processes.

Intel CAS 3.0 is specifically optimized around the NVMe storage interface standard available with Intel SSD Data Center Family for PCIe drives. Internal testing has shown that the NVMe-designed provide much better through-put and lower latency compared to both hard drives and SATA SSDs.

Yahoo Achieves Massive Improvements in Storage Performance

By combining Intel CAS 3.0 with Intel® SSD DC P3600 1.6TB for NVMe, Yahoo has doubled throughput and reduced latency up to 75 percent in its Ceph environment.[2] Yahoo achieved the performance improvement by adding just 50GB on the Intel SSDs with Intel CAS software for every 8TB of HDD. In addition, the performance improvements took place without need for a major data migration or changes to the underlying storage infrastructure, and the solution is cost-optimized with room to scale as needed.

In addition, Intel CAS 3.0 has helped Yahoo reduce recovery time from unexpected data corruption, hard drive, or other system failures by 70 percent. The performance impact during recovery has also been cut in half,[2] resulting in significantly improved customer experience during the recovery window.

Read Requests Latency

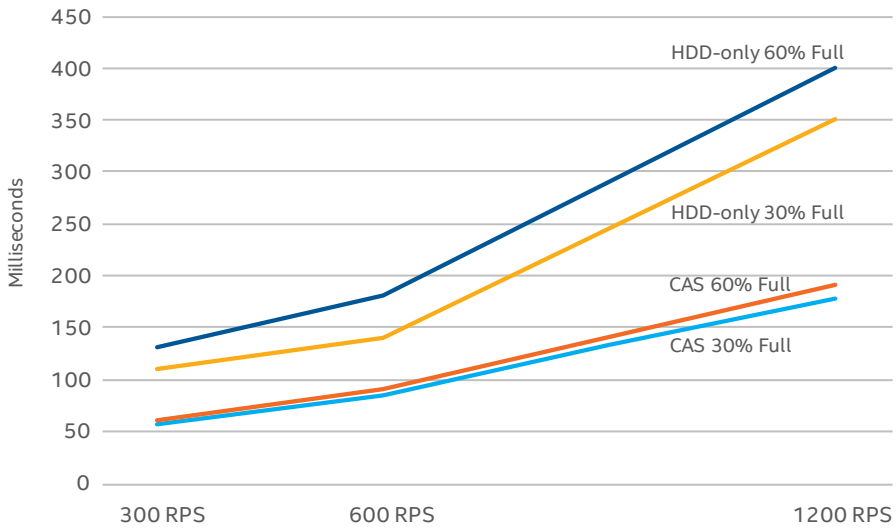

Figure 1. Read latency in systems enhanced by Intel CAS is about half the latency of an HDD-only solution.[2]

Write Requests Latency

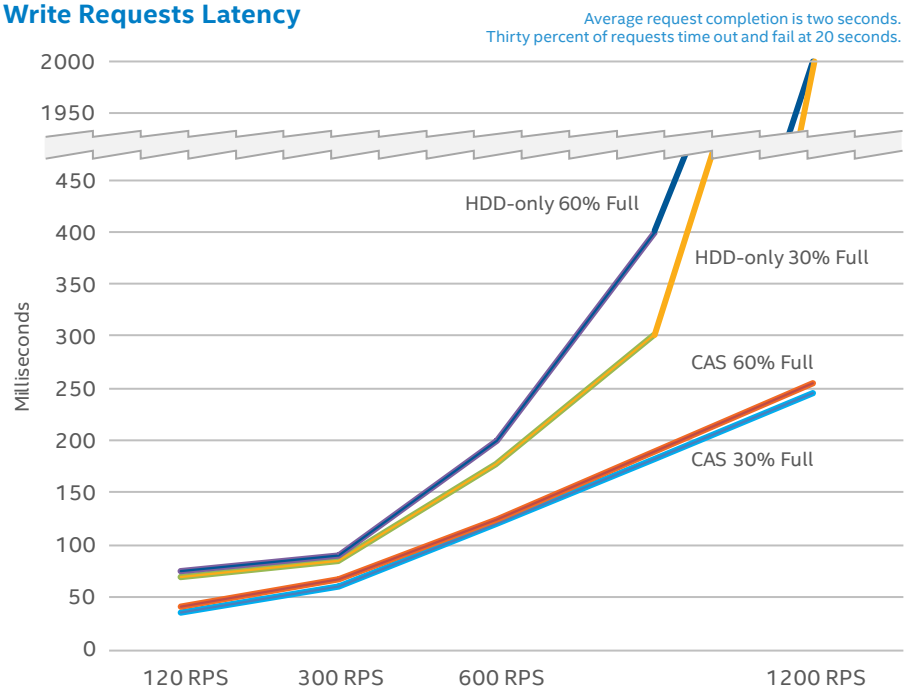

Figure 2. In a system enhanced by Intel CAS, all write requests complete, and latency does not increase even as system storage capacity fills.[2,6]

Benefits of Non-Volatile Memory Express* (NVMe)

Intel led the industry in creating NVMe, a standardized, high-performance software interface available for PCI Express* (PCIe) SSDs. NVMe was architected from the ground up to overcome the performance limitations of SAS and SATA SSDs and make SSDs more efficient, scalable, and manageable. Features include a streamlined protocol and more efficient queuing mechanism to scale for multi-core CPUs and deliver low clock cycles per I/O. Intel® CAS is designed and optimized to take full advantage of the Intel® SSD Data Center Family for PCIe* hardware capabilities.

The solution deployed by Yahoo was designed for Ceph but is broadly applicable to bare-metal storage systems and software-defined storage systems (e.g., Swift and Lustre*). The entire solution can be ordered via a single part number from Intel to simplify license distribution.

Conclusion

Enterprises are experiencing exponential increases in demand for cloud storage, driving many to adopt open-source, scale-out storage solutions that can help reduce costs and improve scalability. As shown through the pathfinding program with Yahoo, Intel CAS 3.0 on SSD Data Center Family for NVMe can extend the benefits of scale-out storage solutions like Ceph, delivering dramatic improvements in performance and recoverability, without need for a costly infrastructure overhaul.

For more information:

www.intel.com/cas
www.intel.com/nvme
www.intel.com/ssd

[1] www.youtube.com/watch?v=vtIIbxO4Zlk

[2] Based on internal Yahoo measurements, 2015. Using a 600 OSD Ceph cluster, 3PB storage, erasure coding 8+3, (Ten) 8T SATA disk, 1MB object size, Intel CAS 3.0, Intel P3600 1.6 TB NVMe SSD, single reads and writes. Testing done with and without Intel CAS/SSD components.
Each OSD node in the cluster contained: SERVER: HP ProLiant DL180 G6 ySPEC 39.5; CPU: 2x Xeon X5650 2.67GHz (HT enabled, total 12 cores, 24 threads); CHIPSET: Intel 5520 IOH-36D B3 (Tylersburg); RAM: 48GB 1333MHz DDR3 (12 x 4GB PC3-10600 Samsung DDR3-1333 ECC Registered CL9 2Rx4); HDD: (10)8TB 7200 RPM SATA HDDs; NETWORK: (2)HP NC362i/Intel 82576 Gigabit, (2) Intel 82599EB 10Gbe; OS: RHEL 6.5, kernel 3.10.0-123.4.4.el7. The Intel CAS 3.0 configuration added an SSD to each node in the system: SSD: (1)1.6TB Intel P3600 SSD (10GB journal per OSD, 1.5TB cache).

[3] www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

[4] Gartner (June 2014). www.gartner.com/newsroom/id/2783517

[5] www.sigops.org/sosp/sosp11/current/2011-Cascais/printable/05-mesnier.pdf

[6] Write latency ranges from 40% lower in lightly loaded systems to over 90% lower in heavily loaded systems, for an average of about 75%.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase.

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.

Printed in USA ♻ Please Recycle 0316/tc/rt 334106-001US