Solution Brief

Modernized Storage Architecture
Data Center

Accelerate I/O: Replace HBA Write Buffers with Fast, Persistent Media and Open Source Software Tools

Larger cache volumes with storage class memory can speed up I/O by up to thousands of times by replacing a RAID host bus adapter (HBA) write buffer, using open source caching technologies.

Architecture Benefits

- **High performance.** Storage caches have a higher capacity than the HBA's onboard DRAM, resulting in higher overall performance.[1]
- **Less complexity.** Eliminating both the HBA card and the battery backup removes single-point-of-failure components and simplifies the hardware environment.
- **Redundant.** Integrated and hybrid RAID solutions can replace the RAID capability of the HBA to provide a volume redundancy with the HDDs as well as a redundant RAID1 caching layer to remove any single points of failure.
- **Persistence.** An HBA's write buffer uses DRAM; sudden loss of system power can corrupt the filesystem (in write-back mode). To prevent this, a battery backup or flash add-on module is required to protect cached data. In contrast, storage device can provide persistent caching, eliminating the need for battery backup.

Summary

As the amount of data every enterprise processes every day grows enormously, CPU performance advances have helped enterprises keep up. But storage performance can often be a bottleneck, preventing full utilization of powerful CPU. Enterprises must also consider cost efficiency and data protection when designing their storage systems.

Traditional SATA-based hard disk drive (HDD) storage systems used for low cost deployments are traditionally aggregated into larger pools by using a host bus adapter (HBA) card in either RAID or JBOD mode. These HBAs use a DRAM-based write buffer to increase the storage performance of the large pool of disks, but the caching capacity is so small that it barely makes a dent in modern applications' storage performance. In addition, HBA cards increase system complexity because they require a battery backup to protect the write buffer. The bottom line is that while legacy HBAs and write buffers adequately served slower storage media on previous-generation processors, new innovations in processing and networking demand similar innovation to improve storage performance and caching functionality.

Even though DRAM is faster than any storage media, a larger amount of fast, persistent media combined with open source caching tools can modernize this HDD based storage stack to provide a higher performing storage solution. By increasing the caching capacity from a <8GB to 100s of GBs, the heavy IO from today's applications can more effectively be absorbed. In addition, by using hot-swappable storage devices, the cache can be easily replaced in case of failure, making a more simple and reliable solution

Solution architecture and value: modernizing storage caching

To demonstrate the potential of this modernized storage architecture, two open source software tools can be leveraged to virtualize the functions of the HBA. First Linux mdadm/mdraid can handle the RAID functionality. In this case, Intel® Virtual RAID on CPU (Intel® VROC) is used as a fully supported version of mdraid. Second, Open Cache Acceleration Software is a new intelligent caching tool available for integration into a variety of environments.

Then a storage class memory device is needed. For example, Intel® Optane™ SSDs offer breakthrough performance and persistence that is excellent for this use case. They are designed for low-latency, high-bandwidth storage performance, particularly on write operations. These characteristics make Intel Optane SSDs optimal for a write buffer device. These devices replace the tiny amount of DRAM on the HBA for caching storage I/O. The far greater capacity of the Intel Optane

SSD (up to 1.5TB) can speed up caching performance for HDD arrays. Further, because the Intel Optane SSD is persistent, a battery backup is not needed.

By running a simple 70/30 Read/Write FIO workload, the benefits of a larger, persistent cache can be seen.

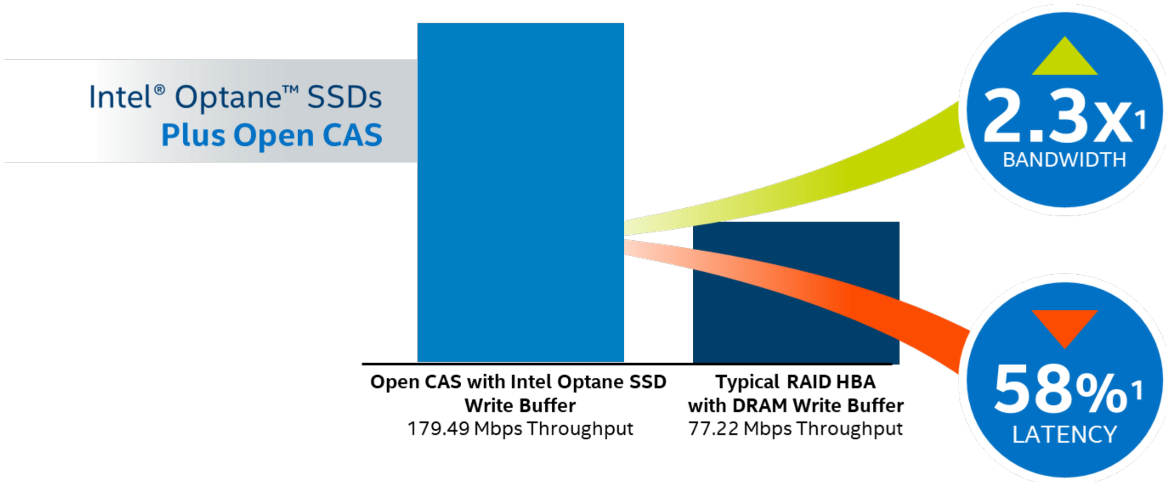

Figure 1. Using a larger, persistent storage cache, like Intel Optane SSDs, creates a more responsive and higher bandwidth storage subsystem than using smaller DRAM cache on an HBA.

Open CAS Linux for Non-VROC-Compatible Systems

For enterprises with servers that already contain an HBA, they can still obtain the benefits of SCM caching by installing Open CAS Linux, in front of their HBA. In this scenario, enterprises can continue to use all their existing hardware—including the RAID HBA or SAS Expander HBA—to connect to HDDs or an external HDD enclosure. However, the RAID HBA's write buffer should be disabled and replaced by a high-capacity Intel Optane SSD to serve as the write buffer. Because the Intel Optane SSD is far larger than the limited RAM on the HBA, overall storage performance improves dramatically. Open CAS Linux can be tuned for even higher performance.

Conclusion

In the past, when storage media and processors were both slower, legacy HBAs and write buffers kept up with storage performance requirements. However, recent innovations in process and networking have outpaced storage performance and caching functionality. Open source tools can be used to modernize the storage stack, and combined with next generations storage media, can modernize the architecture of entire data centers. While the benefits are significant for enterprises using SATA-based HDDs and SSDs, when these enterprises upgrade to quad-level cell (QLC) NVMe-based SSDs, they will still be able to take advantage of this new architecture.

Learn More

- Intel® Optane™ Solid State Drives
- Open CAS Linux
- Media Aware Storage Framework

[1] Intel Tested. May 20, 2020.System configuration: SuperMicro X11DPU, 2x Intel® Xeon ® Gold 6252 processor, 24 cores @ 2.1 GHz, DRAM = 192 GB, BIOS Release = 02/21/2020, BIOS Version = 3.3. OS: Red Hat Enterprise Linux (RHEL) v7.8, kernel = 3.10.0-1127.el7.x86_64, mdadm = v4.1 - 2018-10-01, Intel® VROC Pre-OS version 6.2.0.1034. Storage: 5x 4 TB Western Digital SATA HDDs (Model: WD4000FYYZ) connected to internal backplane, 2x 375 GB Intel® Optane™ SSD DC P4800X Series (Model: SSDPE21K375GA01) connected to internal backplane. BIOS setting: PackageC-State (Auto), Turbo (Enabled), SpeedStepP-States (Enabled), HardwareP-States (Disabled), AutonomousCoreC- State (Disabled), CPU-C6Report (Auto), EnhancedHaltStateC1E (Enabled), SoftwareControlledT-States (Enabled). RAID Settings: Intel VROC (SATA RAID) 5-Disk RAID5, chunk size=4K, group_thread_cnt=4. Cache Settings (when enabled): Intel VROC with integrated caching for RHEL7.8, Write-Only Mode, 64K Cache Line Size. Workload Generator: FIO 3.19, RANDOM: Workers = 4, I/O depth = 16, No Filesystem, CPU not affinitized.

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

Performance results are based on testing as of the date noted in the configuration details and may not reflect all publicly available updates. See configuration disclosure for details. No product or component can be absolutely secure.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.Your costs and results may vary.Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

121/RT/HC 345539-001