SOFTIRON

Storage Wars: Ceph as a secret weapon for HPC

How Ceph can support the explosion of AI/ML data

By Jason Van der Schyff
COO, SoftIron

High-performance computing (HPC) is going mainstream.

With the increased use of artificial intelligence, machine learning and analytics on large datasets, and ever-more sophisticated (and realistic) simulations and modeling, there is growing pressure to find solutions that balance performance and costs. Don't get left behind because your storage is holding you back. If you're looking for a secret weapon in the ongoing battle to manage, grow and optimize your data center, then Ceph is your answer.

Given the large volumes of data now used in most applications, finding the right storage solution is particularly important. High-performance storage can be eye-wateringly expensive; economical storage can impede workflows. And many storage solutions may not easily scale to accommodate future needs.

One overlooked solution that is starting to get renewed interest is Ceph. Ceph is open-source software designed to provide highly scalable object-, block- and file-based storage under a unified system. It enables software-defined storage (SDS) that can be used to provide a massively scalable storage solution for modern workloads like cloud infrastructure, data analytics, and other high-performance computing (HPC) applications.

Why Ceph and Why Now?

As a distributed, scale-out storage framework, Ceph has typically been used for high bandwidth, medium latency types of applications, such as content delivery, archive storage, or block storage for virtualization.

Organizations select Ceph because of its accessibility, interoperability, flexibility, and rich features. In particular, Ceph can do all three types of commonly found storage types (object, block, and file) — flexibility that separates it from most SDS solutions. Its inherent scale-out support allows an organization to build large systems as cost and demand require gradually. Plus, it supports enterprise-grade features such as erasure coding, thin provisioning, cloning, load-balancing, automated tiering between flash and hard drives, and simplified maintenance and debugging.

Even with these credentials, Ceph was not typically considered for HPC in the past. Today however, the reasons it might have been dismissed have since been addressed and the time is right to reconsider Ceph for these environments

Perhaps the main reason Ceph was not considered is its perceived poor performance in an HPC environment. To be sure, years ago, Ceph would not be the best choice for Tier 1 storage plugged directly into a supercomputer. However, Ceph supports distributed storage on commodity hardware allowing organizations to increase read/write performance by adding more devices. As such, it is well-suited for use as Tier 2 storage.

The second reason many have not considered Ceph for HPC in the past is that it was considered hard to use. Most HPC technologies have had a similar fate. Everything used in HPC including multi-core systems, clusters, solid-state disks, high-speed interconnections, and more were first only installed in government labs and academic computing centers. These facilities had the specialist expertise to integrate, optimize and manage the new technologies into their systems to get the most out of them. As HPC went mainstream, the technologies matured and were made easier to use by technology providers. In that way, an IT administrator or network architect could work with the technology and apply it to the organization's business needs.

Ceph has experienced a similar evolution. A large community of open-source Ceph developers are continuously adding improved management features. The Ceph Foundation provides an open, collaborative, and neutral home for stakeholders to coordinate their development and community investments in the Ceph ecosystem. The Foundation is organized as a directed fund under the Linux Foundation. It exists to enable industry members to collaborate and pool resources to support the Ceph project community.

Ceph's use in HPC environments enables scalability, high performance, and timely access to data. It can help support work on large databases such as those found in genomics research or to support collaborative research across geographies.

These were some of the reasons CERN started using Ceph. CERN has been a long-time Ceph user and active community member, running one of the largest production OpenStack clouds backed by Ceph. Backing a range of high energy physics experiments, the Ceph storage service provides object storage for end-user applications, block storage for virtual machines, a POSIX file system for general storage and sharing of files across the network, and an S3 object store. In 2015, CERN ran a Ceph scalability experiment with 30 PB across 7200 object storage daemons (OSDs). To support a more recent experiment, CERN created a larger solution with 10,800 OSDs in total and nearly 65 petabytes of data.

Finding the Right Technology Partner

Many organizations lack the internal expertise for any software-defined storage implementation. While there have been vast improvements in Ceph's ease-of-use, it still brings complexities. For a first project, Ceph typically requires specialists to deploy and manage the software.

One option is to hire new staff with the required expertise. Another is to get in-house staff the training they need to be proficient in Ceph. Both options take time and money. A third option is to team with a technology partner that offers both Ceph and HPC expertise. This is an area where SoftIron can help. SoftIron makes unified data storage solutions for the modern-day enterprise. However, SoftIron is not your traditional storage vendor. It designs and creates

enterprise storage solutions with SDS-dedicated Ceph appliances. SoftIron's solutions optimize Ceph and make it easier to use and adopt within the enterprise.

The heart of its solution line is the **HyperDrive® Storage Node**, a custom-designed, dedicated Ceph appliance, purpose-built for SDS. In contrast to using generic hardware for storage, HyperDrive is custom-built for Ceph, all components integrate with, and exploit Ceph's outstanding functionality. The result is a scale-out storage solution that runs at wire-speed yet at less than 100W per 1U appliance.

To address ease-of-use issues, there is the **SoftIron HyperDrive® Storage Manager**, a powerful, unified, and intuitive system designed to radically simplify the management of Ceph software and storage hardware.

While Ceph is the leading open-source, SDS on the market, it has traditionally challenged organizations due to its' complexity. Storage Manager addresses this challenge managing hardware and simplifying Ceph configuration. The end result is a simple, feature-rich management tool that reduces the overhead of Ceph management.

Given the way organizations use data for HPC applications, one challenge using Ceph has been its inability to support other files shares and protocols. SoftIron built the **HyperDrive® Storage Router** to remove this obstacle and allow everyone to leverage the power of Ceph, no matter who the end-user is.

The Storage Router is a highly intelligent services gateway that enables enterprise-wide adoption of Ceph by consolidating user shares, virtualization, and any other storage technologies onto one scalable, high-performance platform. The result is a highly intelligent platform that not only enables enterprise-wide adoption of Ceph, but it can also accommodate different services modules over time as features or requirements grow.

A recent addition to HyperDrive is a hardware acceleration I/O module that uses programmable silicon to compute Erasure Coding on the fly. Code-named '**Accepherator**, the module solves the traditional "3X replication" data redundancy issue without the expense of high-end CPUs, whilst also sitting inline as a 10GbE Network Interface.

How does this help? Data redundancy refers to the replication or back-ups of data sets that are required to ensure complete restoration of any lost data in the event of drive failures. Data redundancy in a Ceph environment is usually solved by creating three copies of data and spreading it across multiple drives via the CRUSH algorithm. The idea is that if a drive fails, and data is seemingly lost, there are still two additional copies to recover data from. The challenge is that three times replication is expensive, organizations need to buy and maintain three times the useable storage needed to store the data with redundancy.

SoftIron has turned to Erasure Coding (EC), which works by processing data through an algorithm that breaks the data up into chunks and writes a single copy with extra parity bits which can be used to rebuild the data in the event of lost media. Traditionally, EC is also expensive to run because it requires a lot of CPU horsepower to run correctly, which essentially just shifts the cost of data parity and transfers it to the CPUs of the storage appliance.

SoftIron created a dual-purpose EC accelerator + 10GbE SFP network interface to solve this issue, and in so doing has radically reduced the costs and complexity of data redundancy. Offered as an I/O (NIC) option for HyperDrive, it works as an I/O module that computes Erasure Coding on the fly at line rate, removing the load from the CPU while also providing a 10GbE SFP interface.

A Solutions Approach to Ceph for HPC

SoftIron complements these products with services and expertise. The result is that organizations can more quickly realize the benefits of using Ceph for HPC. Key benefits of using SoftIron's Ceph solution for HPC are:

- **Designed, not assembled:** HyperDrive is custom-designed and purpose-built to optimize Ceph.
- **Wire-speed, high-performance:** In a recent a recent head-to-head performance test pitting a cluster of dual-socket Xeon processors against an entry-level HyperDrive, it couldn't match HyperDrive's read/write performance.
- **Built for storage, not compute:** Most storage manufacturers build their appliances around a motherboard and processor, and add software as an afterthought. All HyperDrive components integrate with and exploit Ceph's outstanding functionality. The result is a solution that's optimized for lightning-fast storage performance.
- **Scalable performance:** HyperDrive scales out using distributed storage on commodity hardware. By distributing the load, users get the best possible performance from low-cost hardware. As more HyperDrive boxes are added, read/write performance and scalability increase exponentially.
- **Low power:** HyperDrive runs at less than 100W per 1U appliance. That's 56Tb of all-flash SSD media or 120Tb of traditional spinning media. HyperDrive uses a 64-bit ARM processor rather than a traditional x86 chip because it's a better fit for storage. The result is faster, cooler, more reliable, and uses less power.
- **Made in America:** SoftIron proudly engineers, builds, and assembles its storage appliances at its California production facility, ensuring end-to-end provenance and total peace of mind.
- **Zero risk:** Try before you buy with HyperDrive Test Drive, flexible contracts without lock-in, and pay-as-you go subscription pricing that can be canceled at any time. Additionally, SoftIron supports hardware and software, meaning an organization will never be left in the lurch to mediate between suppliers.

To learn more about SoftIron, Ceph or HyperDrive, visit softiron.com

SOFTIRON

SoftIron® makes the world's finest solutions for the data center.

The company's **HyperDrive®** software-defined storage portfolio is built on Ceph; it's custom-designed and purpose-built for scale-out enterprise storage and runs at wire speed.

HyperCast™ delivers the best density and value for real-time video streaming. SoftIron unlocks greater business value for enterprises by delivering great products without software and hardware lock-in.

softiron.com

@SoftIron

@SoftIron

@SoftIronNews

info@softiron.com