

Case Study



Intel® High Performance Computing
Intel® Xeon® Scalable Processors
Intel® NVMe Technology

Flatiron Institute Taps Ceph for Extreme Scale, Reliability, and Speed

Flatiron Institute's HPC infrastructure, built from open-source and Intel technologies, accelerates breakthrough science for hundreds of researchers

Solution Summary:

- 3rd Generation Intel® Xeon® Scalable processors
- Intel NVMe technology



Executive summary

[The Flatiron Institute's](#) researchers perform advanced science in fields like genomics, quantum physics, astronomy, and neuroscience. As data-intensive research grew in complexity, Flatiron needed a flexible, performant, and exceptionally scalable high performance computing (HPC) system to accommodate the needs of hundreds of scientists. Flatiron explored many HPC architecture options and ultimately adopted a novel solution using open-source Ceph as their primary storage system. Ceph, utilizing 3rd Generation Intel® Xeon® Scalable processors, Intel NVMe technology, and more, gave scientists the compute power and fast storage to manipulate the enormous data sets involved in their breakthrough research.

Challenge

The Flatiron Institute initially supported a handful of scientists performing advanced data analysis using a combination of servers and desktop computing systems. Over time, the institute grew to hundreds of researchers performing workloads involving highly complex calculations, modeling, and analytics. The diverse scientific projects at Flatiron involve data sets that require reliable and performant storage in the petabyte range. Therefore, Flatiron needed to offer its



Using data from powerful telescopes, Chris Hayward of the Flatiron Institute's Center for Computational Astrophysics and collaborators developed a simulation to physically model and visualize galaxy cluster SPT2349-56 and predict how it will change in the future. Simulations such as this one require millions of CPU hours and produce 10s of TBs of data. (Image courtesy of Simons Foundation)

Case Study | Flatiron Institute Taps Ceph for Extreme Scale, Reliability, and Speed

researchers much more powerful computing resources with exceptional scalability and storage speed to accommodate ever-growing needs for storage capacity.

Solution

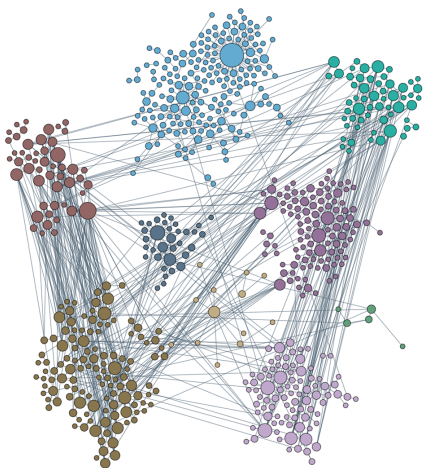
After much consideration, the Flatiron Institute took a non-traditional approach to their HPC storage deployment and chose open-source Ceph Software-Defined Storage as the best solution for their needs. Ceph offers file (CephFS), block (RBD) and object (RGW) interfaces to data and can run on a variety of hardware. Many top university scientists have chosen Ceph for their HPC Storage needs for several years, but Flatiron Institute's HPC deployment pushes CephFS to its limits as a primary filesystem.

After evaluating its processor options, the institute chose 3rd Gen Intel Xeon scalable processors for the new Ceph building blocks. Because the CPUs provide flexibility for general-purpose usage, Flatiron found that the processors offered uniquely beneficial connectivity among the PCI bus, controller, and other elements to accelerate the system's read and write capability.

Compared to previous-generation Intel Xeon Scalable processors, the newest CPUs provide more CPU cores and PCI Express (PCIe) Gen4 capability, which increases the

"When past systems could not meet our scientists' growing demands, we lost valuable research time. With Ceph and Intel products behind our HPC system, we have the scale, performance, and reliability to enable breakthrough science."

— Ian Fisk, Ph.D., Scientific Computing Core Co-Director, Flatiron Institute



HumanBase is software developed at the Flatiron Institute's CCB which uses machine learning to integrate data from more than 40,000 genomic experiments and 14,000 scientific publications to uncover genes' tissue specific function and roles in disease, inter-relationships between genes, and the functional impacts of genetic variants. This gene network generated by HumanBase is for one kidney cell type in diabetic kidney patients. Each circle represents one gene expressed in that cell type, color-coded according to the biological processes it activates. The information shown here requires approximately 2TB of storage. (Image courtesy of Simons Foundation)

Spotlight on the Flatiron Institute

The Flatiron Institute is an internal division of the Simons Foundation that supports a community of researchers. The institute focuses its research on five disciplines: astrophysics, biology, mathematics, neuroscience, and quantum physics. Flatiron empowers large-scale data analysis, theory, modeling, and simulation with HPC and modern computational tools to advance humanity's scientific understanding. [Learn more here.](#)

storage and network bandwidth per PCIe slot by a factor of two. Currently, the Ceph system uses Intel Xeon Scalable processors in over 100 dual-socket servers.

By combining Ceph's prowess, Intel CPUs, and SSDs, Flatiron's HPC system can read and write quickly to its over 4,000 drives for storage. The deployment also makes the most of the SSD's non-volatile memory express (NVMe) capability for high bandwidth and low latency I/O speed. Additionally, the HPC system benefits from 100 Gigabit network cards that support the fast information transfer required for researchers' massive data sets. The mix of Ceph and Intel components deliver an over 100 Gigabyte/sec I/O throughput for fast message passing to meet the demands of highly complex simulations.

Results

The combination of these elements offers Flatiron the building blocks for an HPC system that addresses researchers' scalability and performance needs with better uptime, availability, and assurance of data integrity than other HPC solutions the institute team evaluated.

Flatiron's HPC architecture is designed with futureproofing in mind. As scientists grow in number and their workloads involve larger data sets, the institute can expand the system easily. Because Ceph is not restricted to specific servers or drives, Flatiron can choose the hardware that best meets their needs. Plus, even during an upgrade process, Ceph does not require downtime.

"Ceph and Intel products provide us an unbeatable combination for HPC. With the latest product iterations in place, we've seen two-to-three times faster networking performance which delivers the results our researchers need faster than ever before."

—Andras Pataki, Ph.D., Senior Data Scientist, Scientific Computing Core, Flatiron Institute

Currently, Flatiron's researchers focus their efforts on five scientific disciplines: astrophysics, biology, mathematics, neuroscience, and quantum physics. For this reason, the institute's HPC architecture must offer the flexibility to handle diverse projects without compromising performance. For example, genomic research begins with a large amount of input data. In contrast, astronomy research generates the most data during simulations of galaxies or black holes.

Given the complexity of these workloads, the journaling ratio per NVMe in Flatiron's implementation is higher than a typical

Spotlight on Ceph

Ceph is a Linux Foundation project involving individuals from businesses, governments, and academic environments who work to develop and market the project. As an open source distributed storage system, Ceph offers scalable and reliable storage that supports block, object, and file storage in a single unified system. Learn more here.

Intel became a significant Ceph Foundation supporter early on. Today, Intel is one of the top three contributors to Ceph and remains committed to the open source community to make Ceph faster and easier to use, deploy, and manage. Intel remains focused on three critical development areas:

- **Enterprise readiness:** Intel engineers added Erasure Coding to improve system storage efficiency, among other advancements. They also made key contributions to BlueStore and critical enterprise storage features like support for compression, encryption, and consistency groups. Further streamlining derives from techniques that accelerate compression and encryption using CPU offloading.
- **Performance:** Since 2018, Intel developers have worked to improve Ceph's performance and maximize system



resources. Intel contributed client and server-side block and object caching, which improves average and tail latency performance of Ceph by leveraging fast storage and memory technologies based on Intel Optane SSD and Persistent Memory.

- **Manageability:** Intel created Virtual Storage Manager, more commonly known in the open source community as Ceph Dashboard. Another vital contribution is "Rook," a cloud-native storage orchestrator for Ceph.
- Intel's further investments for Ceph will target optimizations for newer generations of NVMe, CXL, Intel Optane Persistent Memory, accelerators, and high performance, low-latency storage use cases:
- Intel is a lead contributor to the Crimson OSD project, aiming to improve Ceph CPU efficiency and performance when using fast networking devices and newer storage/memory technologies like ZNS SSDs and persistent memory.
- Intel's Distributed Asynchronous Object Storage (DAOS) technology will help optimize storage by reducing bottlenecks involved with memory-intensive workloads and enabling a software-defined object store built for large-scale, distributed NVM.

HPC system. The Flatiron team undertook an extensive testing and validation process of commercially available memory technology. They first evaluated the performance of individual hardware components with Ceph, including CPUs, NVMe devices and networking technology. They then tested the combined group of components working together as a whole.

After evaluation, Flatiron initially chose the Intel P3700 SSDs. Over the years they've continued to adopt various Intel NVMe technology including the Intel SSD P4610 and the Intel Optane SSD P4800X. The most recent choice was the Intel SSD P5600. The combination of the 6.4 TB NVMe drives and the third generation of Intel Xeon Scalable processors provided optimized performance for Flatiron's demanding Ceph environment.¹

Over time, Flatiron deployed various drives featuring Intel NVMe technology into their production Ceph environment. In comparison to the P3700 drives originally used, the current P4800X drives offer an order of magnitude better performance in IOPS and latency. In addition, Flatiron is finding the latest PCIe Gen4 Intel NVMe P5600 devices in their 3rd Gen Intel Xeon Scalable processor-based servers meets their needs better than alternative platform designs.²

"For us, it's all about scale. Data sets can grow exponentially as researchers take on increasingly complex projects. With Ceph and Intel solutions in place, we can grow our storage capacity easily without compromising performance or uptime."

— Ian Fisk, Ph.D., Scientific Computing Core Co-Director, Flatiron Institute

However, it's not only about speed. Flatiron chose Intel's drives for their endurance. Over 100 of the original P3700 drives still support the system. Of those still used in Ceph nodes, only one P3700 needed replacement in the last five years.

Lessons

Start small, learn, and grow: By deploying a smaller-scale HPC system first, Flatiron's team determined how best to administer Ceph, chase down bottlenecks, and optimize the system for scientific workloads. This "start small" process helped them grow the system to its present size with fewer technical challenges.

Data Redundancy: While open-source systems can run on any hardware, they won't necessarily run optimally. By adopting best practices and test methodology for hardware selection and possible failure scenarios, the Flatiron team maximized their system's reliability, speed, and uptime.

Technical Components of the Solution

- 3rd Gen Intel® Xeon® Scalable processors
- Intel SSDs P3700, P4610 and P5600
- Intel Optane™ SSD P4800X

Additional Information:

Learn more about [Intel HPC solutions](#)

Explore the ways [3rd Gen Intel® Xeon® Scalable processors](#) advance HPC computing

Find out how [Intel Optane™ technology](#) can benefit you

Read more about the [Flatiron Institute](#)

Learn more about the [Ceph Foundation](#)



¹ Flatiron runs a series of benchmarks trying to reproduce workloads similar to CEPH with commonly used benchmarking tools such as Flexible I/O (FIO)

² <https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/optane-ssd-dc-p4800x-p4801x-brief.html>

Notices and Disclaimers

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

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

For workloads and configurations visit www.Intel.com/PerformanceIndex. Results may vary.

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.