

## Remove Storage Bottlenecks in High Performance Computing

### Challenge

- **Storage bottlenecks** can slow down the analysis pipeline and delay results
- **Cost to scale performance** storage systems to keep up with data growth and increasing retention times
- **Need for higher performance** to meet the need for faster turnaround time for analysis
- **High management overhead** to enable collaboration and sharing of data among global, distributed teams
- **Cost effectively keeping data protected and accessible** over increasingly longer retention periods

Rapid advances in the ability to collect a greater quantity and higher quality of information from data generated by sensors, satellites, medical instruments, high definition video devices and computer simulators is increasing the size of data repositories to capacities never seen before. The ability to move seamlessly from raw data to analysis to actionable insights is making it possible for researchers and engineers to transform these vast amounts of data into insights and to deliver better, faster and more cost-effective results with their high performance computing (HPC) environments.

However, for many HPC environments, data intensive workloads can overwhelm data storage systems, creating severe computing bottlenecks and major performance issues. In addition, managing and monitoring these complex storage systems significantly adds to the burden on storage administrators and researchers. Organizations must have the ability to store, process and analyze large volumes of data without impacting the efficiency of their high performance computing environments.

### Solution

Scalable, high performance storage that leverages a parallel distributed file system and ActiveScale™ cloud object storage system: The combined solution provides extreme performance, durability and scalability to meet the intensive storage workload requirements of high performance computing.

As the requirements around data storage performance and capacity keep increasing, researchers, scientists, engineers and storage administrators are increasingly empowered to consider new innovations and approaches for their compute and storage infrastructure.

### Highlights

- **Higher performance** scalable storage that provides predictable and consistent access to large amounts of data
- **Increased data accessibility and flexibility** by aggregating disparate storage into a single global namespace
- **Easy integration** into existing environments with support for standard file and block protocols
- **Storage optimization** by intelligently moving infrequently accessed data to ActiveScale cloud object storage system
- **Increased data durability and integrity** at petabyte scale helps to ensure valuable data is protected long-term
- **Improved IT agility** by quickly provisioning and redeploying storage resources to support new business

### PFS + ActiveScale for HPC

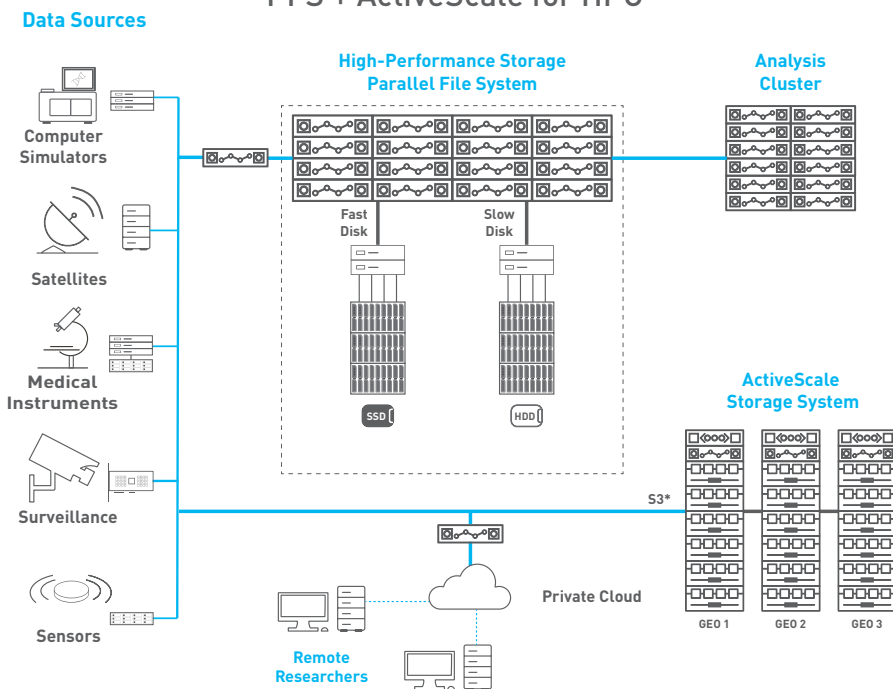


Fig. 1. HPC architecture example

## Accelerating High Performance Computing

Storage and data management bottlenecks are perhaps the most challenging issues in high performance computing. The huge amounts of information that enable their users to gain powerful insight from the analyses they perform can also be a speed bump on the road to enlightenment.

High performance computing environments need the ability to move, store and protect huge amounts of data and enable their users to solve more complex problems faster.

HPC data processing and analysis requires an automated workflow that integrates many data processing and analysis tools. This pipeline typically includes many steps, from initial gathering to realtime processing to data preservation and batch high performance analysis. The sheer breadth of analysis applications required have highly varied compute, memory and I/O requirements. Data access patterns can range from many jobs executing concurrently that need simultaneous access to the same small data files, to applications used to process large, 4D super high-resolution imaging files.

High performance computing (HPC) environments have been dealing with these challenges for years and have had success deploying storage that leverages parallel distributed file systems. Higher performance is achieved by spreading blocks of data from individual files across many file system storage nodes and reading/writing them in parallel.

## Cost Effective Preservation of Information

An often-overlooked strategy is storage optimization, with the objective of reducing cost and improving utilization of the most important and expensive storage tier. Moving data that does not need high performance onto a lower cost, capacity-optimized tier like the ActiveScale cloud object storage systems, helps improve primary storage performance, and reduce costs. ActiveScale can also serve as a private cloud for sharing data internally and with other organizations around the world. And with a single global namespace, finding data is easy and completely unlike the hierarchical limitation of traditional storage architectures.

High performance storage using a parallel distributed file system combined with ActiveScale cloud object storage system, allows HPC organizations to gain control of data growth and analysis workflow. The Quantum solution provides the following benefits:

**Improve storage efficiency** by pooling redundant isolated storage resources under a single global namespace. Free up performance storage tiers by transparently moving infrequently accessed data to ActiveScale using automated lifecycle policies.

**Accelerate workflow and collaboration** to transform enormous amounts of information into insights that contribute to better research results. ActiveScale offers predictable performance, massive scalability and lower management complexity making it easier for teams to collaborate boosting overall productivity.

**Increase data durability and integrity at petabyte scale.** ActiveScale helps ensure valuable data is protected and always available by delivering up to 19 nines durability and site-level fault tolerance in a multi-site configuration. Robust data integrity checks occur automatically and transparently protecting long-term archives; each object can tolerate up to 1000 bit-errors without data loss.

**Simple to Install and manage.** ActiveScale systems are easy to deploy. Simply add power and network connections and it is ready to go. The system self-protects and heals requiring significantly less IT intervention compared to traditional storage systems.

**Easy to integrate.** ActiveScale supports a diverse set of HPC applications and workloads where performance, reliability and availability of data are essential to the business. Native protocol support for NFS, SMB, Object enables seamless integration into existing environments.

## Conclusion

The data management and storage workload challenges faced by HPC organizations are daunting. Sensors, satellites, medical instruments and computer simulators create extreme amounts of data that need to be efficiently processed, analyzed and stored for long periods of time. Storage solutions using high performance parallel file systems combined with ActiveScale cloud object storage system deliver the necessary performance, scale and efficiency needed to address these enormous challenges.



Quantum technology and services help customers capture, create, and share digital content—and preserve and protect it for decades at the lowest cost. Quantum's platforms provide the fastest performance for high-resolution video, images, and industrial IoT, with solutions built for every stage of the data lifecycle, from high-performance ingest to real-time collaboration and analysis and low-cost archiving. Every day the world's leading entertainment companies, sports franchises, research scientists, government agencies, enterprises, and cloud providers are making the world happier, safer, and smarter on Quantum. See how at [www.quantum.com](http://www.quantum.com).