

SOLUTION BRIEF

High Performance Computing Solutions on Pure Storage FlashBlade//S

Experience storage optimized for the challenges of HPC.

High performance computing (HPC) performs parallel operations to solve complex problems at high speed using clusters of individual computers or hosts. The applications that use HPC clusters require compute platforms with high CPU and GPU counts and a corresponding storage platform that has been designed to process, manage, and store data at equally high speeds with linear scalability. Many HPC-based workloads like simulation, modeling, analytics, machine learning, rendering, and genome sequencings are accessed as files over network file system (NFS) and server message block (SMB) protocols from shared storage for concurrent access and collaboration among workflows and users.

Storage Challenges of HPC Applications

Applications in HPC environments generate workloads with different characteristics and diverse IO profiles. Some HPC applications generate massive amounts of metadata; others require high throughput. These applications may have sequential, random, large, or small block read and write characteristics. Legacy HPC storage architectures are typically built on IBM General Parallel File Systems (GPFS) and open-source Lustre Filesystems. Legacy storage used for HPC typically employs parallel filesystems where multiple servers in a cluster access large files concurrently. Traditional parallel file systems are complicated and require constant tuning to support different workload profiles. This is often solved with different storage silos, but that is expensive and requires different skills to support each one. Coordination and efficient load balancing between systems is another limitation.



Speed

Highly parallel HPC jobs demand faster processing times enabled by multi-dimensional performance and high bandwidth storage.



Scalability

Disparate HPC workloads that have varied IOPs, bandwidth, and latency requirements need storage that scales seamlessly.



Simplicity

HPC requires storage that is not only fast, but simple to manage and protects massive volumes of disparate data.

The optimal storage platform for HPC workloads provides the multi-dimensional performance and bandwidth required for multiple HPC applications. It is built with modular compute and capacity, and a parallel architecture that is all-flash, scale-out, and highly available. It supports both file and object protocols as well as all data types regardless of size and data profile. Lastly, it can scale performance or capacity as data and application requirements grow and can be upgraded with new features without interruption in service.

FlashBlade//S: Storage Built for HPC Workloads

[FlashBlade//S™](#) from Pure Storage® is a symmetric, distributed, and scale-out data platform architected to load balance client requests from an HPC cluster and process data at scale. FlashBlade//S can scale with respect to performance and capacity using configurable blades with compute and DirectFlash® Modules (DFM) for high file and object count environments. FlashBlade//S is an ideal storage platform for the unstructured data workloads in HPC workloads that require variable block sizes to handle a high number of IOPs from small files, as well as high throughput for large file sizes with low latencies.



Figure 1. Pure Storage FlashBlade//S

Architected for Speed, Scalability, Simplicity, and Sustainability

FlashBlade//S is a modern data platform that speeds HPC workloads built with:

- **Proven Purity//FB - Scale-out storage software:** The heart of FlashBlade®, Purity enables scale-out, enterprise-grade data storage, services, and management. It can handle billions of files and objects and delivers unmatched performance for HPC workloads.
- **All-QLC flash - Unified fast file and object (UFFO) storage:** Built using DirectFlash Modules and all-QLC architecture, FlashBlade//S is the ideal foundation for HPC workloads. It is designed to easily support the most demanding unstructured data workloads, without compromising on system performance or efficiency
- **Disaggregated architecture - Scale compute and storage separately:** Designed with a unique modular architecture that allows you to easily increase capacity or performance without downtime, FlashBlade//S is a customizable platform that gives you the ability to tailor your configuration for specific HPC requirements. It provides the flexibility to easily adapt to your data growth projections and evolving storage needs.
- **Evergreen innovation – Future-proof and sustainable:** An Evergreen subscription makes FlashBlade//S the last scale-out storage platform you'll ever need. Get the latest generation blade upgrades every three years (Ever Modern) and new software features that are continually added with premium-level proactive and predictive support included.

FlashBlade//S Top 10 Values for HPC Workloads

- **Speed:** FlashBlade//S allows more HPC jobs to run concurrently over industry standard protocols like NFSv3/v4.1 with faster turnaround time, thus optimizing the license costs of the HPC applications to drive faster time to market (TTM).
- **Scale:** Horizontal scaling by adding more compute nodes to the HPC cluster does not require adding storage in the same ratio. Compute and storage can be disaggregated while scaling performance and capacity. Data reduction allows us to store more data for less storage.
- **Simplicity:** FlashBlade//S can create file systems and export them quickly to mount or be mapped to Linux and Windows clients, respectively. Using RESTful APIs and provisioning tools can further simplify storage consumption on demand for HPC workloads.
- **Sustainability:** The modern architecture of FlashBlade//S delivers leading power and space efficiency. Apart from deeper compression capabilities, which enable significant reduction in the storage footprint, FlashBlade//S provides 67% more [energy and emission savings](#) compared to competitive, legacy all-flash systems and 60% savings versus competitive upstart all-flash systems.
- **Data mobility:** Data has gravity. However, it does not have to be stored in silos for various HPC workloads without freedom to be moved. Multiple workflow pipelines, like simulation and machine learning, can share and collaborate from data stored on FlashBlade//S as an ideal consolidated data platform with support for industry-standard protocols including NFS and SMB. Data from heterogenous storage endpoints can be moved easily to FlashBlade and also from FlashBlade to the cloud for additional flexibility.
- **Data availability:** Access to common configuration and project files over NFS or SMB are easy to recover from a process or HPC node failure, thus providing resiliency and higher uptime.
- **Data protection:** Array-level snapshots allow immediate file system rollback and backup/restore for any kind of disaster. File-system replication to DR sites enables faster recovery from site failures and long-term data retention.
- **Data security:** FlashBlade provides centralized control with distributed access, encryption of data at rest, Krb5 support for user authentication, and NFSv4.1/SMB ACL for authorization. Additionally, Purity's SafeMode™ feature secures the critical data you need to recover from a cyberattack so you can restart your business services quickly, without succumbing to attacker demands.
- **Hybrid cloud connectivity:** Public cloud providers have enabled customers to scale CPUs and GPUs for HPC clusters beyond datacenter boundaries. FlashBlade provides on-demand data storage for HPC clusters with a hybrid cloud architecture that provides low latency, high bandwidth, and scalable performance, flexible capacity configurations, and maintains data security and sovereignty for HPC application data.
- **Observability:** Managing and reporting from multiple FlashBlade endpoints can be configured and viewed from a single pane of management using standard open-source tools with long term retention of historical data.

HPC Customer Use Case

A key priority for the [Nanyang Technology University \(NTU\)](#) High Performance Computing Center was the ability to run multiple analyses concurrently. “If we looked at one genome a week, it would take us 20 years to cover 1,000 species, so it was important for us to do things in parallel and moving to FlashBlade has enabled us to significantly speed up the process,” said project lead Assistant Professor Jarkko Salojarvi. With FlashBlade enabling them to run up to four jobs in parallel, the team has since crossed the project’s halfway mark, completing more than 550 assemblies in just 18 months, compared to less than 100 assemblies within the same time span if examined one by one.

HPC Industry Solutions

Pure Storage has enabled customers across a broad range of industries with HPC workloads requiring parallel processing at scale to achieve faster time to insight and accelerated time to market for their product and services. Most of the customers in the following industries have one characteristic in common - a high volume of data that needs to be quickly processed, stored, managed, and protected throughout project or production lifecycles.

- [Research centers](#) and [laboratories](#): HPC is used for complex problem solving by data scientists in the research and scientific communities.
- [Semiconductor design](#): Large scalable CPU-based compute farms are used in the design and manufacturing industries for small and durable silicon on chip (SoC) (also referred to as electronic design automation).
- [Automotive](#): HPC environments are used in automotive design for modeling and simulating during the design and product development process.
- [Genome sequencing](#): Large scale GPUs and CPUs are used in HPC environments to sequence and analyze large volumes of genome data.
- [Financial service institutions](#): Large HPC environments are used to solve complex financial concerns including fraud detection, risk analysis, smart trading, etc.
- [Oil and gas](#): HPC is used for 3D simulations and seismic processes during the upstream discovery while evaluating complex and dangerous oil reservoirs and deposits.
- [Media and entertainment](#): High-speed image modeling with animation and fast editing requires HPC environments in large render farms.

Additional Resources:

- Learn more about [Unified Fast File and Object Storage with FlashBlade//S](#).
- Discover how to bring [innovation and productivity to life](#) with high performance computing.
- Explore more [high performance computing blogs](#).

purestorage.com

800.379.PURE

