Hammerspace for High Performance Computing

Run traditional simulation, AI, and GPU workloads on a single high-performance data platform spanning node-local NVMe, SSD, and HDD storage - across any data center or cloud

SOLUTION BRIEF

The Need to Modernize Research Data Infrastructure for HPC and AI Workloads

Classical HPC architectures were designed around centralized supercomputers and tightly coupled simulation workflows. Modern research computing environments must now span multiple data centers, clouds, storage systems, GPU clusters, and increasingly diverse AI pipelines operating against massive volumes of distributed unstructured data.

What has not changed is the need to keep expensive GPU and CPU resources fully utilized despite increasingly distributed data, mixed workflows, and infrastructure fragmentation. This requires a data platform capable of delivering:

- High-throughput, low-latency data access.
- High-bandwidth streaming reads and writes.
- Fast metadata operations.
- Shared access across diverse workloads and storage tiers.
- Automated data orchestration across performance, capacity, and archive tiers.

Traditional alternatives create operational challenges:

- Legacy HPC parallel file systems are operationally complex, requiring proprietary clients to achieve the highest-throughput and lowest-latency.
- Scale-out NAS architectures require significantly more infrastructure to achieve HPC-class performance.
- Multiple storage silos for scratch, performance, archive, and AI pipelines create operational overhead, fragmented workflows, and large-scale data copy sprawl that slows research and AI initiatives alike.

AI workloads are exposing the operational limits of fragmented HPC storage architectures, where data movement, isolated performance tiers, and siloed infrastructure increasingly constrain GPU utilization and workflow scalability.

Solution Benefits

Unify siloed storage systems into a global namespace.

Achieve high-throughput, low-latency performance without a proprietary client.

Run traditional Mod/Sim and emerging AI workloads on a single platform.

Use node-local NVMe as a tier of high-performance shared storage.

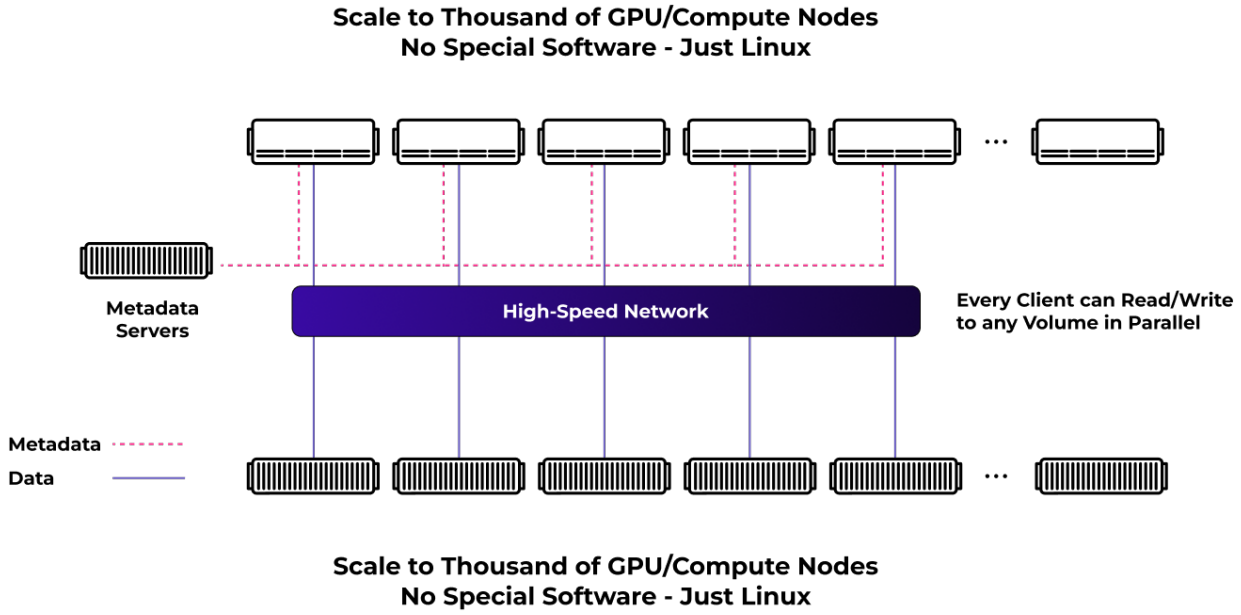
Enable hybrid-cloud computing with data orchestration.

HAMMERSPACE

A Standards-Based Parallel File System Architecture for Mixed HPC and AI Workloads

Hammerspace delivers HPC-class performance through a standards-based parallel file system architecture designed for modern research computing and AI infrastructure.

Unlike traditional HPC parallel file systems that depend on proprietary client software and tightly coupled storage architectures, Hammerspace uses Linux-native NFSv4.2 and pNFS capabilities already included in modern enterprise and HPC Linux distributions running in virtually every data center today.



This architecture:

- Delivers high-throughput, low-latency parallel performance for HPC and AI workloads.
- Eliminates the need for proprietary client software on every compute node.
- Scales linearly to thousands of compute nodes and storage nodes.
- Supports flexible networking options including standard Ethernet, RoCE, and Infiniband.
- Allows heterogeneous file (NFS/SMB) and object (S3) storage systems to participate in a unified global data platform.
- Spans multiple sites and clouds under a unified global namespace.
- Enables policy-driven orchestration and dynamic placement of data across performance, capacity, and cloud tiers.

By separating the metadata control plane from the data plane, Hammerspace enables direct parallel data access between clients and storage while maintaining centralized visibility, orchestration, automation, and policy control across distributed and otherwise incompatible infrastructure.

The result is a high-performance, unified data platform that simplifies operations, eliminates fragmented storage silos, and enables AI and HPC workloads to operate across distributed infrastructure as a single coordinated environment.

One Data Platform from Node-Local NVMe to Archive

The fastest storage in any HPC environment is the node-local NVMe already installed inside GPU and CPU servers. Historically, this storage has been trapped as isolated scratch space or caching layers, creating fragmented islands of performance, underutilized flash capacity, and unnecessary dependence on external high-performance storage infrastructure.

Hammerspace transforms this server-local NVMe into a shared high-performance tier within a global file system, enabling ultra-low-latency shared data access to hot datasets, checkpointing workloads, and AI pipelines directly across GPU and CPU clusters. This becomes increasingly important as AI workflows require repeated access to active datasets across training, inference, simulation, and analytics pipelines.

Because Tier 0 is part of the Hammerspace global namespace, data can be dynamically orchestrated across node-local NVMe, flash, capacity, archive, and cloud tiers without disrupting data access, workflows, or visibility.

Proven in Real HPC Environments

Vanderbilt ACCRE

The Vanderbilt Advanced Computing Center for Research and Education (ACCRE) selected Hammerspace to modernize a large-scale research computing environment supporting 750 compute nodes, 80 GPU nodes, hundreds of research projects, supporting both traditional simulation environments and rapidly growing AI and GPU-driven research workloads.

ACCRE unified node-local NVMe, commodity flash storage servers, and a multi-petabyte disk-based archive under a single global namespace to achieve:

- 48% reduction in storage costs
- Elimination of fragmented storage silos
- Dynamic provisioning of storage resources to research teams
- Simplified operations across performance, capacity, and archive tiers
- Shared high-performance infrastructure for both AI and traditional HPC workflows
- Reduced dependence on dedicated parallel storage silos

Typical Use Cases

Scratch & Simulation

High-performance data placement directly adjacent to GPU and CPU resources using node-local NVMe. Ideal for:

- Checkpointing and restart
- Hot training and inference datasets
- Ultra-low-latency workloads

Project & Shared Research Workspaces

Persistent, high-performance storage for the bulk of HPC data:

- Shared team workspaces for research groups and institutes
- Multi-tenant AI and data science environments
- Unified access to multi-vendor storage pools
- Heterogeneous on-premises and cloud storage infrastructure unified under a single global namespace

Home Directories, Collaboration & Visualization

Enterprise-grade data services for user environments:

- Researcher home directories
- Departmental shares and visualization pipelines
- Mixed protocol access (NFS, SMB, S3) enabling researchers, AI pipelines, visualization tools, and enterprise workflows to operate against the same shared datasets.

Archive & AI Reuse of Historical Data

Modernized long-term retention that doesn't strand data on tape:

- Global visibility into archived datasets
- Policy-based recall and re-placement on flash or performance tiers
- Enables historical research and simulation datasets to be reactivated for AI training, inference, analytics, and retrieval workflows without large-scale migration projects.

The Data Foundation for Next-Gen HPC & AI

Modern research computing environments require more than high-performance storage infrastructure. They require a unified data layer capable of orchestrating distributed data, performance tiers, and workflows across heterogeneous environments without creating new silos.

Hammerspace enables organizations to unify performance, capacity, archive, and cloud storage into a single coordinated data platform that simplifies operations, maximizes infrastructure efficiency, and accelerates both scientific discovery and AI innovation.