Modernizing the Data Foundation for High-Performance Computing

How Federal Research Organizations Can Accelerate Scientific Discovery, Improve Research Productivity, and Prepare for the Convergence of HPC and AI

Executive Summary

United States federal research and scientific-focused agencies are at a critical infrastructure crossroads. For decades, High-Performance Computing (HPC) investments have focused heavily on maximizing raw compute performance, optimizing throughput, and accelerating time to insight for predictable modeling and simulation workloads.

Today, these same environments face a dual pressure: managing unprecedented data growth while simultaneously integrating completely different data access patterns introduced by artificial intelligence (AI) and machine learning (ML). While compute architectures have advanced rapidly, legacy storage and data management architectures have become limiting constraints on overall system utilization and researcher productivity.

This paper outlines a modernized approach to HPC data infrastructure. By combining the enterprise-class resilience and performance of Hitachi Virtual Storage Platform (VSP) One with advanced data orchestration layers, Hitachi Federal enables agencies to break down data silos, maximize throughput at scale, and support the coexistence of traditional HPC and modern AI workloads, all while preserving and extending existing infrastructure investments.

The Evolving Landscape of Federal Research Computing

Federal agencies, including the Department of Energy (DOE), NASA, NIH, NIST, NOAA, and DoD research laboratories, operate some of the world's most demanding computing environments. Historically, data flows within these systems were relatively predictable, moving linearly through generation, processing, analysis, and archiving within isolated infrastructure clusters.

Modern research and scientific missions have evolved beyond this localized model. Agencies now encounter highly complex requirements defined by:

- **The Fusion of Simulation and AI:** Traditional modeling and simulation applications now run alongside AI training, large-scale inferencing, digital twin environments, and advanced analytics.
- **Explosive Data Scale:** Datasets across genomics, climate forecasting, aerospace engineering, and geospatial intelligence continue to expand exponentially in size and structural complexity.
- **Distributed Collaborative Teams:** Research increasingly requires seamless collaboration across multiple sites, distinct institutions, and hybrid on-premises/cloud environments.

As these workloads scale, data gravity intensifies. Moving massive datasets between fragmented storage environments creates severe operational friction, turning the data management architecture into the ultimate bottleneck.

Critical Bottlenecks in Legacy HPC Storage Architectures

When infrastructure performance degrades, organizations frequently focus on expanding compute nodes or adding localized storage capacity. However, many modern HPC challenges stem from architectural fragmentation rather than storage speed alone. Federal research IT leaders consistently report five primary points of friction:

1. Storage Performance and Metadata Constraints

Traditional storage platforms frequently hit limits when managing high-throughput I/O operations and dense metadata tracking. As job scales increase, metadata bottlenecks cause long-running compute jobs to stall, lowering the return on expensive supercomputing cluster investments.

2. Fragmented Data Silos

Research data routinely sits isolated across multiple storage platforms, legacy NAS devices, object repositories, and cloud tiers. Researchers spend significant percentages of their active hours locating, staging, and preparing data for processing rather than conducting core analysis.

3. Inefficient Cross-Site Collaboration

Sharing massive datasets across geographically distributed teams frequently requires manual copying and bulk data transfers. This results in massive capacity waste due to duplicate datasets, severe version control challenges, and critical synchronization delays.

4. Divergent Access Patterns (HPC vs. AI)

Traditional HPC workloads typically execute sequential write and read operations on large files. Conversely, AI and machine learning workflows ingest millions of small, random files simultaneously across highly distributed nodes. Forcing these opposing access patterns onto a single legacy storage platform causes extreme resource competition and severe performance degradation.

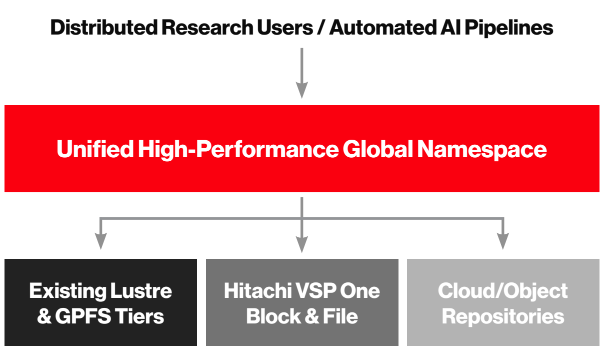
5. Disruptive Modernization Cycles

Replacing entrenched, high-performance file systems like Lustre or IBM Storage Scale (GPFS) during an infrastructure refresh introduces extreme migration risks, high costs, and potential downtime for mission-critical scientific programs.

The Hitachi Architecture for HPC Modernization

To overcome these barriers without forcing complex, wholesale infrastructure replacements, Hitachi Federal delivers a unified data foundation engineered for multi-site research environments.

The solution combines the absolute hardware reliability of the **Hitachi Virtual Storage Platform (VSP) One** block and file storage platforms with the intelligent data orchestration capabilities of a high-performance **Parallel File System** layer. By decoupling data access from physical hardware, this architecture establishes an enterprise-class global namespace that spans local, remote, and cloud storage tiers.



High-Performance Parallel File Access at Scale

The architecture leverages a software-defined, distributed parallel file system layer that natively aggregates high-speed flash storage across distributed nodes. By executing parallel data paths directly between compute nodes and storage media, the platform delivers the massive throughput and ultra-low latency required to saturate GPU-accelerated clusters and advanced HPC nodes.

High-Throughput Metadata Orchestration

Instead of centralizing metadata management onto a single controller, a common cause of bottlenecks in legacy NAS platforms, the architecture distributes metadata processing dynamically across the environment. This ensures that as datasets grow into billions of files, directory transversals, metadata updates, and lookups scale linearly without performance degradation.

Transparent Multi-Site Data Mobility

Through automated, policy-driven live data mobility, data is orchestrated across different storage tiers, locations, and cloud environments seamlessly. Researchers access files globally through standard protocols, while the orchestration layer handles backend placement based on active usage, performance requirements, or storage economics, all without manual user intervention.

Product Integration: Delivering a Resilient, Scalable Foundation

Hitachi VSP One Block and File Storage

The baseline physical layer of this architecture utilizes the Hitachi Virtual Storage Platform (VSP) One. Engineered for mission-critical availability and high-density performance, VSP One provides:

- **Uncompromising Resiliency:** Delivers complete data availability and enterprise data protection services, ensuring that massive simulation or clinical research datasets remain permanently accessible without unplanned downtime.
- **NVMe-Driven Throughput:** Utilizes optimized flash media to provide high-IOPS, low-latency performance to eliminate storage tier bottlenecks beneath demanding applications.
- **Seamless Scalability:** Scales capacity and throughput dynamically to absorb massive, unstructured data growth without forcing administrators to reconfigure existing application clusters.

Coexistence with Specialized Solutions: Hammerspace and Hitachi Content Software for File (HCSF)

Hitachi Federal implements an open, flexible integration strategy rather than enforcing a rigid, single-vendor stack. Depending on whether an agency's bottleneck is multi-site data isolation or localized cluster performance, Hitachi integrates distinct, specialized software layers over the robust VSP One infrastructure foundation:

- **For Global Data Orchestration (Hammerspace):** For environments requiring advanced cross-site collaboration across heterogeneous storage vendors, Hitachi integrates Hammerspace. This layers a global namespace directly above existing storage arrays, bringing automated multi-site replication, live data mobility, and granular file placement to distributed research teams without moving the underlying data.
- **For Ultra-High Performance File Access (HCSF powered by WEKA):** For environments pushing the absolute limits of raw throughput inside the supercomputing or GPU cluster,

Hitachi offers Hitachi Content Software for File (HCSF) powered by WEKA. HCSF is a software-defined, POSIX-compliant parallel file system that delivers maximum IOPS and ultra-low latency directly into high-density compute environments. It allows agencies to effortlessly handle the blended, random-access workloads of modern AI/ML pipelines alongside traditional sequential HPC modeling and simulation.

By leveraging these tailored software layers, Hitachi Federal ensures that data is seamlessly orchestrated across the entire enterprise while guaranteeing maximum storage performance at the compute tier.

Strategic Value to Federal Research Agencies

Co-Existence and Protection of Investment

The primary differentiator of the Hitachi Federal approach is the ability to **complement, rather than replace**, current investments. Agencies do not need to tear out working Lustre, GPFS, or DDN deployments. The Hitachi architecture layers over or integrates alongside existing environments, acting as a high-performance orchestration layer that connects historical data repositories to modern processing nodes.

Maximized Research Productivity

By presenting all files through a single, unified data access layer, researchers spend zero time moving or staging files manually. Data is automatically positioned on the appropriate tier when a job is scheduled, allowing scientists to focus entirely on analysis and operational insights.

Built-In Readiness for Future AI Integration

Implementing a software-defined, highly parallel data layer ensures that your infrastructure can support traditional block-and-file simulation models alongside modern, random-access AI/ML pipelines simultaneously, preventing infrastructure obsolescence.

Customer Spotlight: Supporting Next-Generation Space-Based Threat Detection

A U.S. defense agency required a platform capable of supporting a medium-earth orbit missile tracking system that would ingest and process data from SBIRS and next-generation OPIR satellites.

Hitachi Federal partnered with a leading systems integrator to architect and deliver a data platform capable of ingesting, storing, and processing telemetry, multimedia, and mission-critical data at scale.

Outcomes

- Future-ready scalable infrastructure foundation
- Continuous access to mission-critical data
- High-performance support for advanced analytics
- Real-time processing for intelligence generation
- Expandable architecture for future mission requirements

Conclusion

Modernizing federal HPC data infrastructure is no longer simply about buying faster hardware or expanding raw capacity. True modernization requires an intelligent data architecture capable of eliminating bottlenecks, unifying data silos, and accelerating access across highly distributed research ecosystems.

Hitachi Federal combines the performance, scalability, and absolute resilience of Hitachi VSP One with advanced parallel file access and orchestration solutions to deliver a mission-ready data foundation. By extending existing infrastructure investments and simplifying administrative complexity, we ensure that federal research and scientific agencies maximize the true velocity of their high performance missions.

Continue Your HPC Modernization Journey

Explore additional resources, solution overviews, customer success stories, and insights on building a modern data foundation for HPC and AI.

Visit the HPC Resource Center

Hitachi Federal

Corporate Headquarters
11950 Democracy Drive, Suite 200
Reston, VA 20190 USA
hitachifederal.com

HF-PAL-OV-ModernizingDataFoundationForHPC-14Jul26-B