HPC Data Infrastructure Wasn't Built for HPC + AI

SOLUTION BRIEF

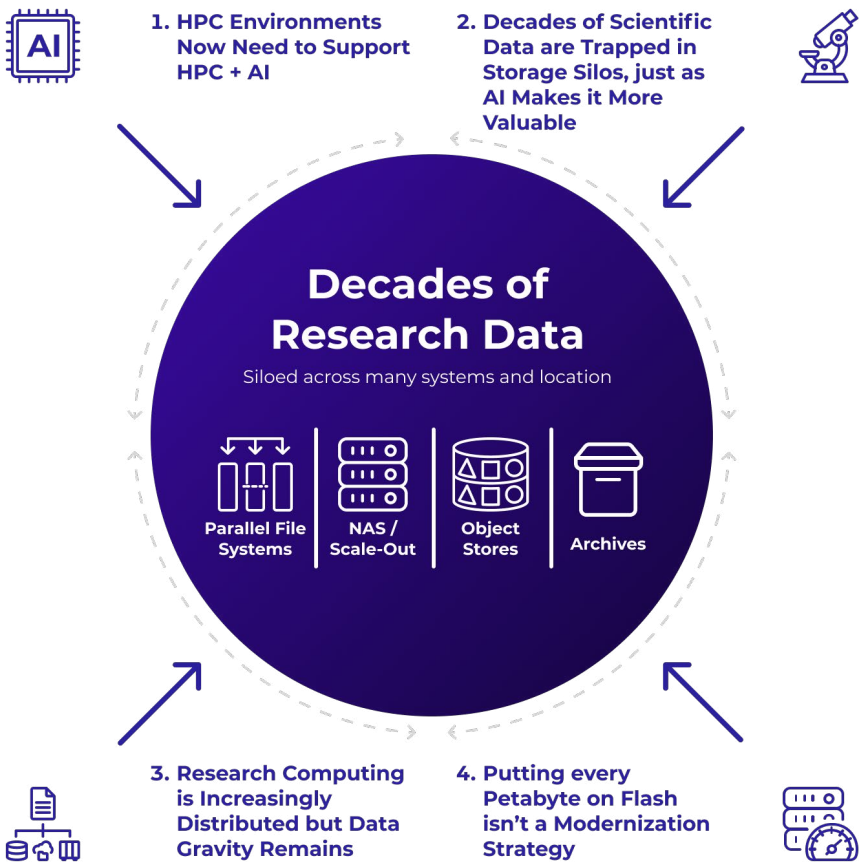
Why Research Computing Leaders Need to Rethink the Data Architecture Behind Modern HPC

For decades, research computing organizations have built infrastructure to support demanding modeling, simulation, and scientific computing workloads. Over time, these environments have accumulated multiple generations of parallel file systems, NAS-based home directories, object storage, tape archives, cloud storage and other specialized infrastructure containing petabytes of valuable scientific data.

Now another major transition is underway.

Research organizations are adding AI workloads to run alongside traditional HPC applications. Existing scientific data is becoming more valuable as an input to AI and new methods of discovery. Compute and data are increasingly distributed across clusters, sites, collaborators, and selectively into the cloud.

The challenge isn't simply delivering more storage performance. It's about operationalizing research data for HPC + AI without abandoning decades of scientific data, replacing everything that already exists, or putting every petabyte on flash.



Four Factors Reshaping HPC Data Infrastructure

1. HPC environments now need to support HPC + AI

Modeling, simulation, and scientific computing aren't going away. Research computing organizations are being asked to support emerging AI workloads for training and tuning models, inference, RAG and agentic AI alongside traditional HPC workloads.

Adding another storage environment for each new workload or compute cluster may solve an immediate requirement, but it also creates another silo. Over time, separate storage systems for scratch, performance, project data, archive, and new AI pipelines create increasingly fragmented workflows and data copy sprawl as data is continually moved between them.

The challenge is increasingly to support HPC and AI together, rather than creating another storage silo for each new workload and then trying to identify, vectorize and orchestrate data to the desired models and AI environments

2. Decades of scientific data are trapped in home directories and storage silos - and AI is making it more valuable

Research institutions have accumulated enormous datasets across years or decades of scientific work. AI is changing the value equation: historical datasets can become inputs to new models, analyses, and methods of discovery.

But that data is often siloed across generations of parallel file systems, home directories, object stores, and tape archives. Making it useful to new compute can require staging, copying, or migrating massive datasets between environments.

The more valuable existing research data becomes, the greater the cost of leaving it trapped in storage silos.

3. Putting every petabyte on flash isn't a modernization strategy

Flash is critical where the highest performance is required, but research environments contain data with very different performance, capacity, and access requirements.

Keeping every dataset permanently on premium flash is economically difficult, particularly as SSD pricing, availability, and procurement constraints put additional pressure on infrastructure budgets. At the same time, high-performance NVMe already deployed inside compute nodes may remain stranded as isolated node-local scratch rather than contributing to the broader data infrastructure.

Modernization requires balancing performance, capacity, and cost across the data lifecycle, rather than treating each as a separate storage decision.

4. **Research computing is increasingly distributed, but data gravity remains**

   Research increasingly spans multiple clusters, sites, facilities, collaborators, and cloud resources. Compute can increasingly run wherever resources are available, but data remains tied to storage infrastructure and location.

   Moving a workload to another cluster or location can mean staging another copy of its data, maintaining another storage environment, or creating another workflow for researchers and administrators. The result is more data movement, more copies, and greater operational complexity.

   As compute becomes more distributed, research data needs to become less dependent on where it is physically stored.

Modernizing the Data Layer

The question isn't simply:

"**What storage should we buy for the next cluster?**"

The bigger question is:

"**How should we modernize our data infrastructure so existing research data, existing storage, and new HPC + AI infrastructure can work together?**"

The next generation of research computing requires a different approach to the data layer, one that can deliver high-performance access where required, make existing scientific data available to new compute, span distributed environments, and balance performance, capacity, and cost.

Modernizing HPC isn't about replacing what you have or buying more storage. It's about modernizing the data layer that connects existing research data with the next generation of HPC and AI.