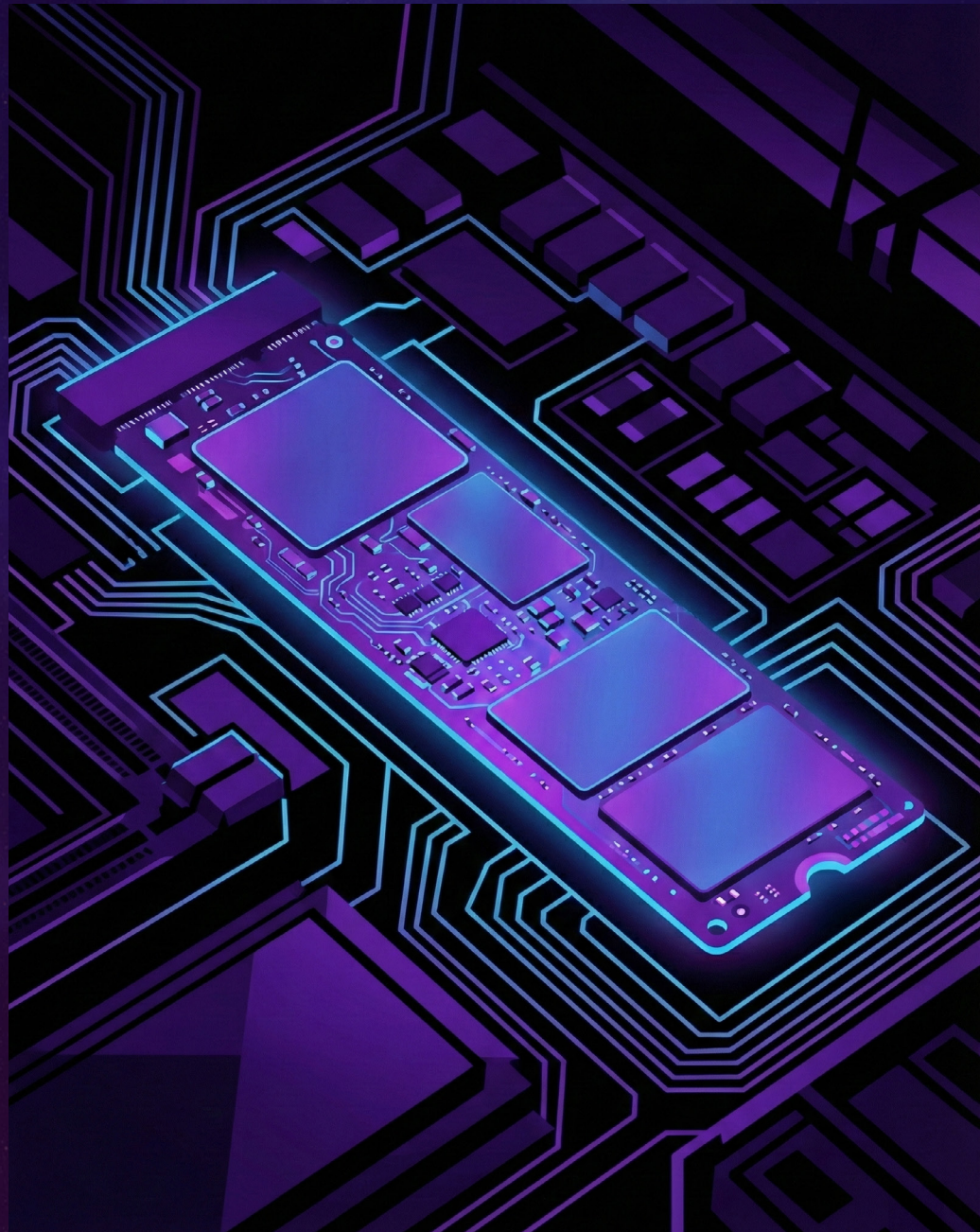


THE SSD SUPPLY CRISIS OF 2026:

A Strategic Infrastructure Survival Guide

Prepared for IT Infrastructure
Leaders and Storage Architects

 **HAMMERSPACE**
www.HAMMERSPACE.com



Executive Summary

The SSD supply crisis of 2026 reflects a structural reallocation of global semiconductor capacity toward AI infrastructure. Hyperscaler and AI demand are consuming NAND flash, DRAM, and manufacturing capacity, leaving enterprise buyers exposed to persistent price volatility, extended lead times, and constrained availability.

What IT Infrastructure Leaders are Experiencing:

- **200% price increase** in NAND flash in 2H'2025¹
- **46% price increase** on HDDs since September 2025²
- **20-week lead times** on high-capacity enterprise SSD³
- **52-week lead times** on high-capacity nearline HDDs⁴

To drive AI initiatives and other critical business initiatives, **adopting a flexible “open” data architecture mitigates supply chain risk.** This Survival Guide outlines the structural drivers behind the supply crisis and provides a practical checklist to reduce supply risk with Hammerspace:

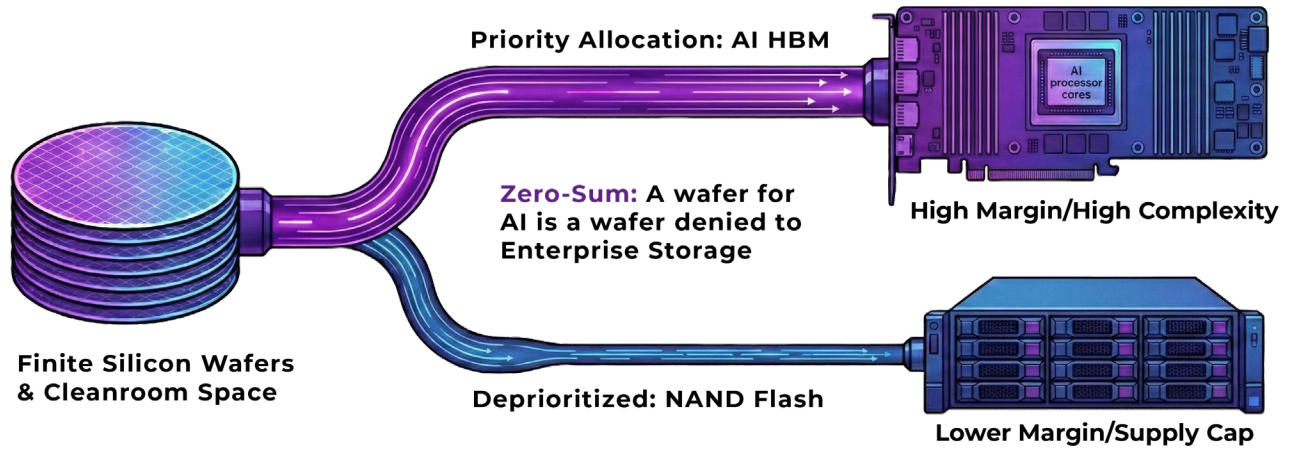
1. **Use Server Flash Capacity You Already Own:** Turn GPU/CPU server-local flash storage into a tier of high-performance shared storage with Enterprise-class reliability, managed and protected by Hammerspace with no drivers or agents required.
2. **Utilize HDD Storage:** Hammerspace supports any drive type including HDD in addition to SSDs. You can leverage existing storage systems from 3rd party vendors or new storage systems built with HDDs, SSD or hybrid storage systems.
3. **Leverage Available Capacity:** Leverage existing storage that you already own. Hammerspace is able to uniquely store data on your existing storage systems that support NFSv3 or S3 protocols – SSD and/or HDD.
4. **Use Cloud Storage as Part of a Unified Global Namespace:** Include cloud storage capacity in a global namespace with policy-based data movement to/from cloud.
5. **Multi-Site Copy Reduction:** Eliminate multi-site data copying by orchestrating data in a global namespace, reducing effective storage capacity used across sites.

The Reality Today: 200%+ Price Increases, 20-52 Week Lead Times

The current SSD supply crisis is being driven by a reallocation of silicon capacity toward AI infrastructure, particularly high-bandwidth memory (HBM) and high-performance compute platforms. Enterprise NAND flash, which competes for the same wafers, cleanroom capacity, and manufacturing tools, has been deprioritized as manufacturers concentrate on higher-margin AI components.



This shortage is not a cycle, and instead reflects a reallocation of finite silicon supply and fab capacity.



For any organization that needs to expand storage capacity to meet the demands of AI, HPC, or other unstructured data processing, the implications are immediate. Pricing volatility has replaced predictability, with buyers experiencing rapid and sustained price increases alongside extended lead times that now stretch from several months to more than a year.

Hyperscaler and AI demand have overwhelmed flash and HDD supply, resulting in higher prices, longer lead times, and extreme pricing volatility.

200%	Price increase on NAND flash in 2H'2025 ¹
46%	Price increase on HDDs since September 2025 ²
20 weeks	Lead times for high-capacity enterprise SSDs ³
52 weeks	Lead times for high-capacity nearline HDDs ⁴

Inventory buffers that once absorbed short-term shocks have largely disappeared, exposing buyers to spot-market dynamics and, in some cases, daily or weekly pricing. The shortage is also cascading across adjacent components, including high-capacity HDDs, driving substitution toward enterprise SSDs and further tightening flash availability. Organizations that depend on proprietary, vendor-locked storage platforms face the highest risk, as expansion is constrained by the availability of specific drive modules and firmware-qualified components.

In summary, the 2026 SSD supply crisis is a structural reality of the AI era, and requires a different strategic approach that prioritizes architectural flexibility over hardware dependence.




The Storage Scarcity Solution: An Open, Flexible Data Architecture

You can't plan your way out of this with procurement alone. The organizations that move fastest are shifting focus from buying capacity to using existing capacity more effectively across SSDs, HDDs, NAS, object storage, and cloud, and designing architectures that can flex when hardware isn't immediately available. **That shift from hardware dependence to architectural flexibility is the real survival strategy.**

Traditional storage-centric designs tend to bind performance, capacity, and data services to a specific appliance platform and a narrow hardware compatibility list. When pricing spikes or allocation tightens, these designs force organizations into expensive overprovisioning, delayed projects, or disruptive re-platforming.

An open architecture takes the opposite approach: it decouples the data layer from the underlying storage and treats storage media as interchangeable. That flexibility matters in 2026. It allows teams to add capacity from whatever is available (TLC, QLC, HDD), incorporate existing NAS and object stores, activate server-local NVMe already deployed in CPU/GPU clusters, and extend to cloud for burst capacity or lifecycle placement - all while maintaining a consistent global namespace and policy-driven data movement. The table below compares these two architectural approaches.

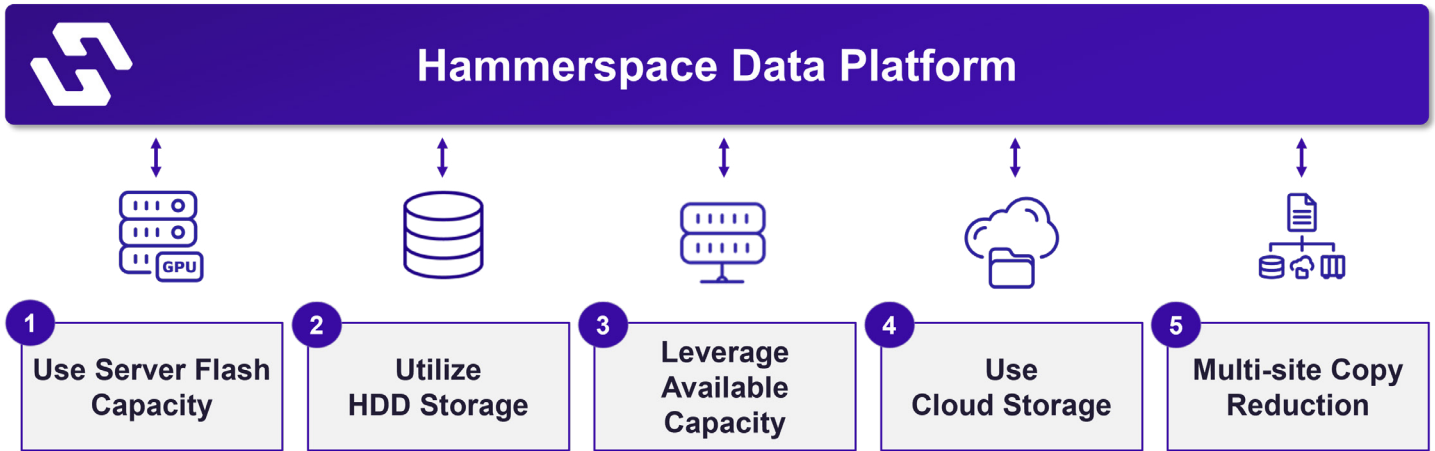
	Legacy Storage-Centric Architecture	Hammerspace "Open" Architecture 
Hardware Compatibility	Limited: Typically single vendor or small HCL	Broad: Aggregate capacity from any x86 storage server into a global file system
Storage Media Supported	Rigid: Most modern data platforms limited to flash-only	Flexible: Use TLC flash, QLC flash, HDD, and even tape
Support for Third-party Storage	None	Use existing NAS from all leading vendors. Use any S3-compatible object store
Use GPU/CPU Server Flash Storage	Not Supported: Server-local storage is siloed	Use the server-local flash storage you already own as part of your shared storage
Unify on-prem and cloud storage	Not Supported: Inflexible replication or data copying	Unify on-prem and cloud storage into a single global namespace with automated data movement



Five Practical Ways to Address the SSD Supply Crisis

Hammerspace helps IT infrastructure leaders reduce exposure to SSD and HDD supply constraints while continuing to execute on AI readiness, infrastructure modernization, and performance-critical initiatives. Each approach focuses on architectural flexibility rather than hardware dependency.

Access all data via a Global Namespace using standard file and object protocols.
Parallel file system architecture delivers performance and scale for AI/HPC



1. Use Server Flash Capacity You Already Own

The Situation: One of the fastest ways to reduce dependence on new SSD purchases is to use the flash capacity you already own in your CPU/GPU cluster. Today's CPU and GPU servers often contain large amounts of local NVMe storage. **In a large compute cluster, this can represent petabytes of flash capacity already deployed but largely underutilized, simply because it is siloed inside individual servers.**

The Solution: Hammerspace enables organizations to aggregate this local flash across their compute clusters and use it as a Tier 0 of high-performance shared storage. Critically, this is not isolated storage. It is part of a shared global file system that can include the Tier 0 flash storage inside a compute cluster, Tier 1 external networked flash storage, disk-based Tier 2 storage, and archival storage.

This means you can unlock ultra-low-latency and high-throughput shared storage using SSDs you already have in place and activate this capacity in hours instead of weeks or months.

Your Action: Assess your compute cluster, either CPUs or GPUs, to determine how much local flash storage is available in each server and across the cluster, and how it is being used today.

A Hammerspace Tier 0 architecture turns CPU/GPU server-local flash storage into a tier of high-performance shared storage

The Situation:

Most compute servers come with local flash storage that is now being prioritized from a supply standpoint – this could be Petabytes of flash in a large compute cluster.

The Solution:

A Hammerspace Tier 0 architecture turns this GPU/CPU server-local storage into a tier of high-performance shared storage with Enterprise-class reliability managed and protected by Hammerspace

The Benefit:

Unlock ultra-low-latency, high-throughput shared storage. Zero new hardware spend, activate capacity in hours.



30 – 60 TB+ local flash storage per server

Case Study: A leading AI model company activated 20 Petabytes of high-speed NVMe flash storage in days using Hammerspace. This flash storage was already in place in their GPU cluster.

[READ THE CASE STUDY](#)



2. Utilize HDD Storage

The Situation: Not all data needs to live on SSDs. Rigid all-flash architectures force organizations to over-provision expensive SSD capacity simply to accommodate data growth, but the reality is that most data does not require the performance of flash and could be tiered to HDD storage to free up flash capacity.

The Solution: Hammerspace allows organizations to use SSDs for the highest performance workloads, then automatically place less active data on HDD-based storage, including NAS, object storage, and cloud storage. All data is part of a shared global file system regardless of what media it is stored on, and Hammerspace automates the movement of data between flash and disk based on policy.

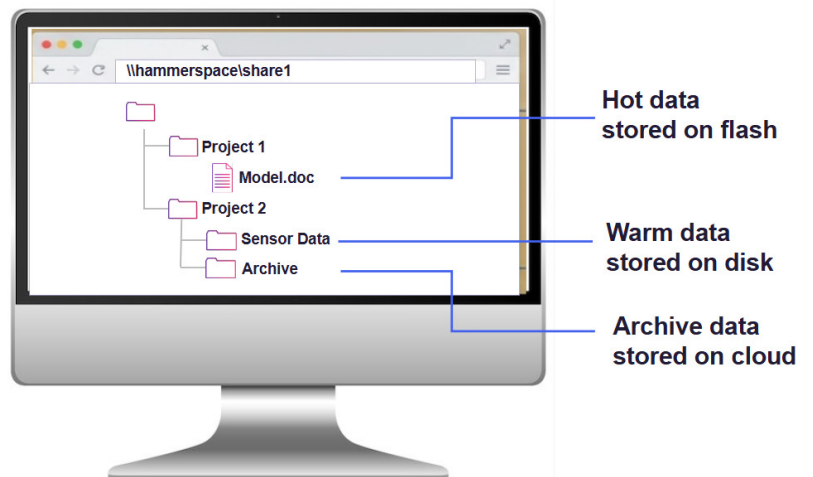
And because Hammerspace is built on a parallel file system architecture, organizations can achieve high-throughput, low-latency performance at scale even with disk-based storage systems.

Your Action: Analyze the data and workloads stored on flash today. Estimate how much of the data is inactive – a typical number is greater than 50% - and estimate how much flash capacity that would free up.

Use Flash AND HDD storage in one file system with policy-based tiering, while delivering the performance and scale for AI/HPC workloads

- Tier to NAS, S3 object storage, cloud storage
- File-level granularity
- Files remain visible and accessible
- No links or stubs

The Benefits: Reduce need for SSDs, reduce costs, deliver performance and scale for AI/HPC

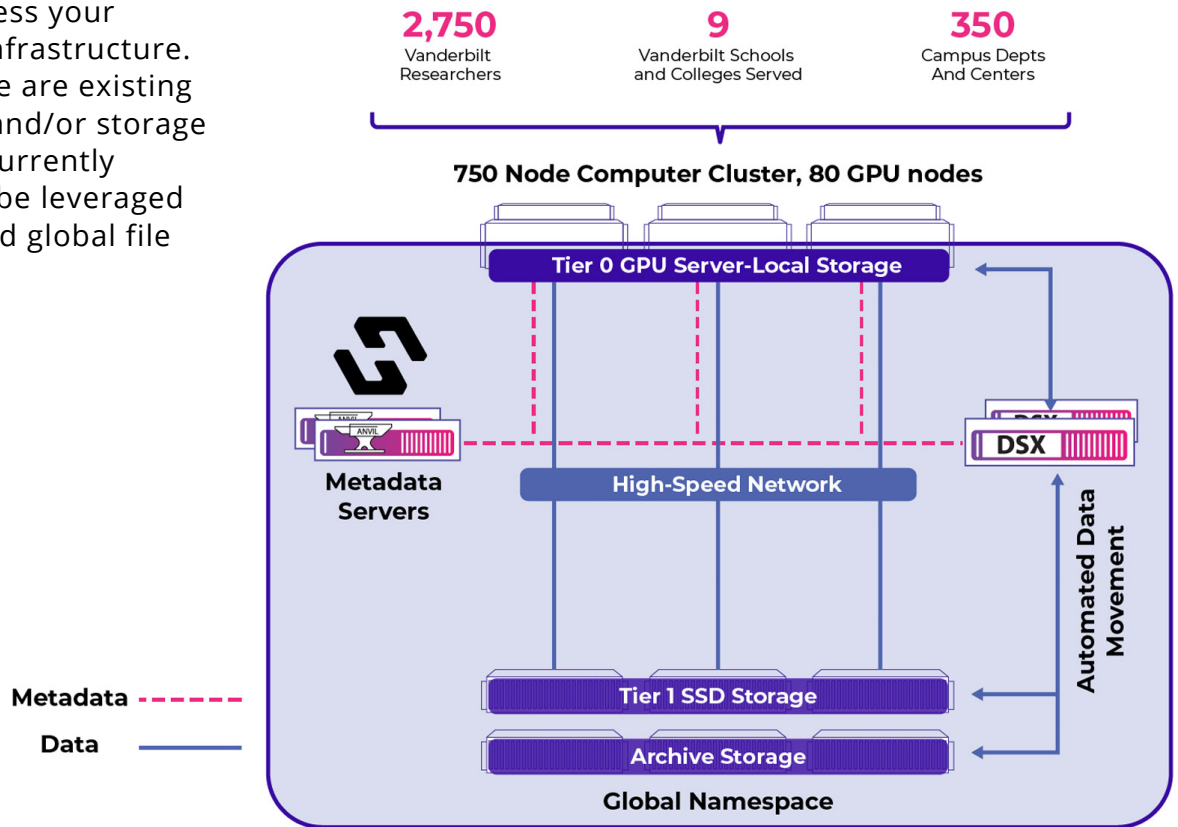


3. Leverage Available Capacity

The Situation: Most enterprises already operate a mix of storage servers, NAS, and object storage systems across vendors, generations, and performance tiers.

The Solution: Hammerspace enables you to leverage existing storage that you already own. Hammerspace is able to uniquely store data on your existing storage systems that support NFSv3 or S3 protocols – SSD and/or HDD. For example, use x86 flash storage servers for tier 1 storage, disk-based servers for tier 2 storage, and existing NAS or object storage either SSD or HDD) for additional tier 2 capacity. All of these storage systems can be added into the Hammerspace global file system to logically aggregate existing capacity while the data itself stays put.

Your Action: Assess your current storage infrastructure. Determine if there are existing storage systems and/or storage servers that are currently siloed that could be leveraged as part of a shared global file system.



Case Study: Vanderbilt ACCRE aggregated 10 PBs of capacity from their compute cluster, storage servers, and existing disk-based storage – reducing storage costs by 48% while meeting performance and scale requirements for high-performance computing.

[READ THE CASE STUDY](#)



For NAS systems, Hammerspace offers the additional capability of “assimilating” the metadata from those NAS systems, while the data itself stays in place. Using this process, millions of files on existing NAS are visible and accessible through Hammerspace in minutes, without time-consuming data migrations.

This means you can extend the life of those NAS systems during this supply-constrained period.

Problem:

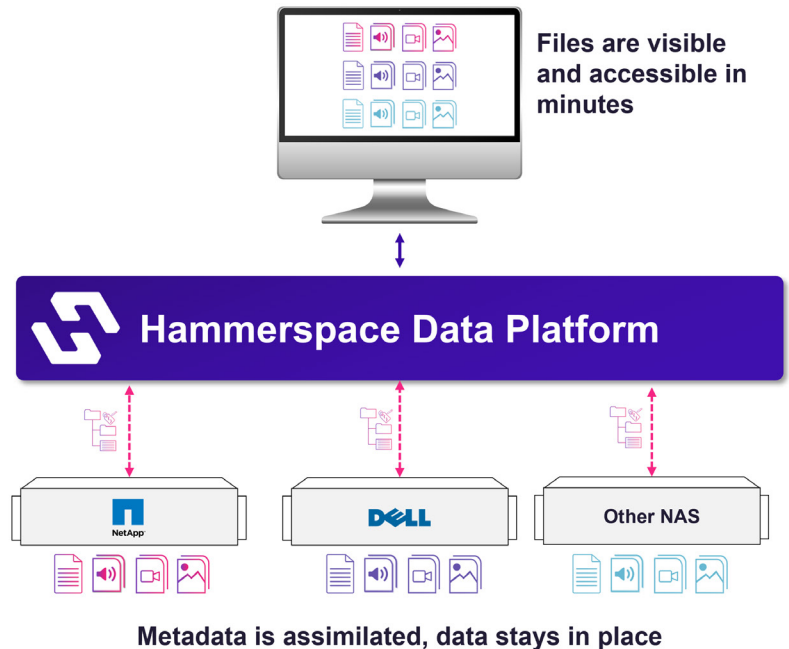
Migrating Petabytes of data takes months.

Solution:

Hammerspace assimilates metadata from existing NAS storage in minutes, making files visible and accessible instantly – without moving data.

Benefits:

Keep using existing NAS assets during supply crunch and extend their useful life. Migrate data to new storage over time, transparently to users.



Case Study: A Federal government agency that produces video and image content had aging Dell PowerScale systems. They deployed Hammerspace as the primary file system, assimilated the data, and kept those NAS systems in-place.

- ✓ Automated tiering between flash PowerScale system and disk-based PowerScale system
- ✓ Able to shift PowerScale systems to third-party maintenance
- ✓ Saved \$500K over 3 years



4. Use Cloud Storage with a Unified Global Namespace

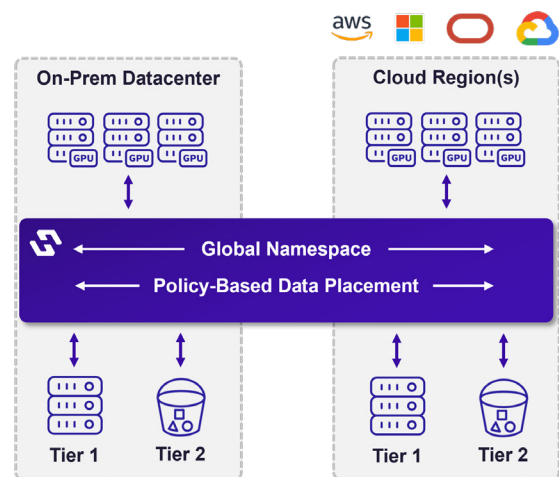
The Situation: Although enterprises are facing severe supply constraints on storage capacity, the hyperscalers have supply available of both compute and storage. But legacy storage architectures do not allow organizations to seamlessly extend into one cloud, let alone move workloads to different cloud regions or cloud providers that happen to have available capacity.

The Solution: Create a Global Namespace with Hammerspace spans sites, clouds and storage systems with policy-based data movement. With Hammerspace, you can automate the flow of data between on-prem and cloud storage, so cloud becomes an extension of your on-prem environment rather than a destination. This can apply to running persistent high-performance file workloads in the cloud, bursting to the cloud to leverage elastic cloud compute resources, or tiering to cloud object storage for long term archiving and protection.

Your Action: Summarize what cloud provider(s) you are using today. Check with your account teams to determine if they have available capacity. Using the data analysis above, determine how much of your data could be moved to the cloud.

Create a Global Namespace that spans sites, clouds, and storage systems – with policy-based data placement – to enable true hybrid-cloud agility

- **Run-in-Cloud:** Deploy Hammerspace in the cloud for high-performance file workloads
- **Burst-to-Cloud:** Automate data movement between on-prem and cloud regions
- **Tier-to-Cloud:** Tier inactive data to cloud object storage to free up on-prem capacity



Case Study: A global leader in digital payments deployed a 4 PB Hammerspace global file system that spanned their on-premises data center and a cloud region.

- ✓ Gained agility to leverage elastic cloud compute resources
- ✓ Simplified user workflow and improved productivity of 3,000 workers
- ✓ Reduced storage costs by \$3-5M



5. Multi-Site Copy Reduction

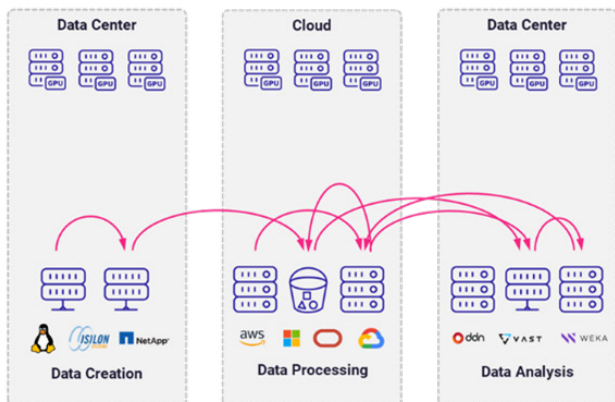
The Situation: In multi-site and hybrid-cloud environments, teams often resort to copying entire datasets between locations so applications and users can access local data. Over time, this creates “data copy sprawl” – multiple replicas of the same data across sites – which drives up effective capacity consumed and increases costs.

The Solution: Instead of duplicating data everywhere, Hammerspace orchestrates data within a global namespace, placing data where and when it’s needed using plain-language, administrator-defined policies. This eliminates manual data copying and lowers the effective storage capacity required per site.

Your Action: Audit how many copies of the same datasets exist across your data centers and clouds today. If you find that replication policies, manual copying, or “just-in-case” duplication are driving a significant portion of your storage footprint, that’s a strong signal that Hammerspace could reduce your storage capacity pressure.

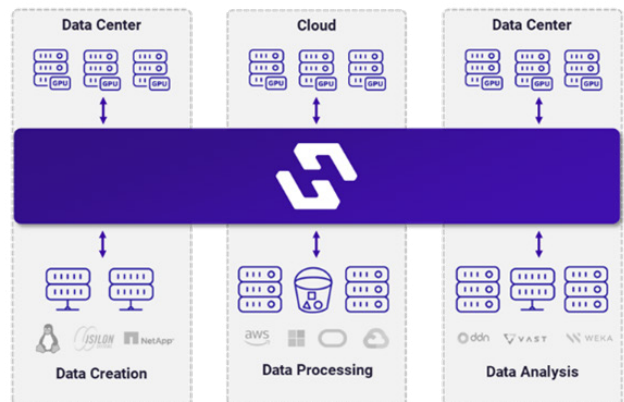
Without Hammerspace:

Entire datasets are copied between sites, leading to excess capacity and operational complexity



With Hammerspace:

Data orchestrated in a global namespace, reducing effective storage capacity used across sites



Case Study: Global creative agency BUCK eliminated manual data copying and enabled global collaboration at scale with Hammerspace

[READ THE CASE STUDY](#)



Vendor Analysis

If you need more storage capacity in 2026 to drive your initiatives forward, it is critical to select a data platform that gives you the flexibility and agility to navigate today's environment.

The matrix below focuses on the capabilities that matter most during this supply crisis: support for both flash and HDD tiers, the ability to use commodity server hardware, support for CPU/GPU server-local flash as shared storage, the ability to incorporate existing third-party NAS systems, and a true global namespace that spans sites and clouds.

The Hammerspace Data Platform delivers unmatched flexibility to mitigate supply risk AND deliver the performance and scale for AI and HPC

	Competitor X	Competitor Y	Competitor Z	Hammerspace
Use Flash Storage All modern data platforms support flash, but Vast Data and Weka are both "flash-only"	✓	✓	✓	✓
Use HDD Storage Build high-performance storage using a combination of flash AND HDD	✗	✗	✓	✓
Use CPU/GPU server flash storage Use the flash storage you already own inside your compute cluster as a tier of shared storage	✗	✓	✗	✓
Use Third-party NAS Reuse and extend the life of existing NAS storage. no need to rip and replace	✗	✗	✗	✓
Use Cloud Storage Seamlessly Use cloud storage seamlessly and dynamically with a Global Namespace	✓	✗	✗	✓
Multi-Site Copy Reduction Eliminate redundant data copies across systems and locations with Data Orchestration	✗	✗	✗	✓



Survival Guide Checklist

Use this checklist to reduce storage supply risk while keeping AI readiness and infrastructure modernization initiatives on track. The goal is to create a data architecture that can flex with pricing, lead times, and component availability - without forcing application changes or disruptive migrations.

Step 1: Assess Overall Risk and Exposure in 2026

- What are your top 5 initiatives for the year?
- How many of these rely on purchasing and deploying additional storage capacity?
- If that capacity is not available for 3 months, 6 months, 12 months - what would that mean for those projects?
- Which initiatives have regulatory, contractual, or revenue impact if delayed?
- What assumptions in current roadmaps rely on historical storage price declines?

Outcome:

Identify which initiatives are most vulnerable to storage pricing, lead time, or availability risk.

Step 2: Assess Your Storage Infrastructure

- If you have a compute cluster, how much flash capacity is available across that cluster? How is that local flash storage being used today, if at all?
- How many independent storage islands exist today?
- How difficult is it to re-balance capacity across systems?
- Are storage systems tightly coupled to specific workloads or applications?
- Do you have file servers used for specific applications that could be repurposed?
- Any NVMe or flash storage servers that could be re-allocated for critical purposes?
- Do you have aging NAS storage systems?
- How much capacity is available in each system?
- How much of that data could be tiered to an object storage system or cloud object storage - assuming no disruption to your user's workflows?

Outcome:

Identify siloed and/or stranded capacity that could be aggregated into a shared global file system.



Step 3: Assess Your Data

- How much of your data is hot, and needs to be on flash?
- How much of your data is less frequently accessed and could be tiered to less expensive storage to free up flash capacity?
- How often is data temperature actually measured vs. assumed?
- How much flash is consumed by:
 - Snapshots and replicas
 - Temporary / intermediate / checkpoint data
 - "Just in case" performance over-allocation
- Are tiering policies static or dynamic?
- How quickly can data be re-promoted to flash if needed?

Outcome:

Identify which datasets could be placed on HDD rather than flash storage.

Step 4: Assess Hybrid-Cloud Readiness

The hyperscale cloud providers have available cloud compute and storage capacity.

- Of your new capacity needs, how much of that capacity could be deployed in the cloud?
- Can cloud object storage be used as a pressure-release valve, not a permanent destination?
- What are the barriers to making that happen?
- Are egress costs modeled under sustained usage, not just short-term bursts?

Outcome:

Identify datasets and workloads (both compute and storage) that could be moved to the cloud to free up on-premises compute and storage capacity.

Guidance for Procurement Leaders

- Negotiate for partial-ship and mixed-SKU fulfillment instead of single-SKU dependencies
- Plan for weekly or monthly price variability; avoid budget assumptions based on annual price-per-TB decline
- Create a runbook for capacity triage: what data can be tiered down, archived, or moved to cloud object when supply tightens?
- Align business stakeholders on service tiers (hot/warm/cold) and acceptable tradeoffs before scarcity forces decisions
- Qualify multiple vendors for each performance tier, not just capacity tier
- Avoid architectures that require homogeneous drive sizes or generations



Conclusion

The supply crisis of 2026 is not a temporary disruption; it is a structural shift. The assumptions that shaped two decades of infrastructure planning (namely predictable availability and steady price-per-terabyte declines) no longer apply. As silicon capacity is prioritized for AI-optimized components, availability and lead times have become just as critical as budget.

Organizations that succeed in this environment will not attempt to outbuy the shortage. They will out-architect it. By decoupling data from proprietary hardware, fully utilizing existing SSD, HDD, and server-local NVMe capacity, and treating cloud as an extension rather than a destination, infrastructure teams can reduce supply risk while continuing to modernize.

Teams that act early will reduce procurement pressure and regain control over capacity timelines before shortages force reactive decisions.

Don't Try to Outbuy the Shortage, Out-Architect it.

The market is paralyzed by 52-week lead times. You don't have to be.
Schedule a 30-minute review, and we'll help you:

Identify Stranded Capacity
Assess Vendor Dependencies
Map Architectural Options

SCHEDULE MY ARCHITECTURE REVIEW

Sources:

1. <https://www.forbes.com/sites/tomcoughlin/2026/01/02/digital-storage-and-memory-projections-for-2026-part-2>
2. <https://www.tomshardware.com/pc-components/hdds/hard-drive-prices-have-surged-by-an-average-of-46-percent-since-september-iconic-24tb-seagate-barracuda-now-usd500-as-ai-claims-another-victim>
3. <https://www.eetimes.com/how-sk-hynix-redefined-the-memory-market>
4. <https://www.trendforce.com/presscenter/news/20250915-12714.html>

