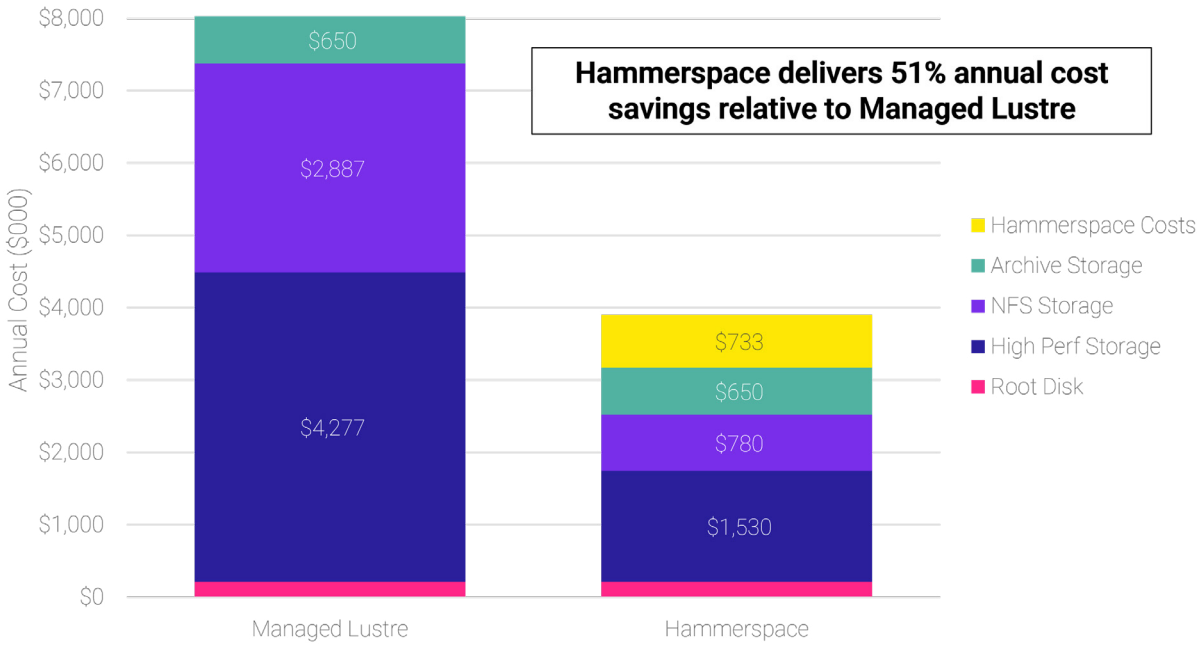
Comparing Storage Costs for HPC and AI Workloads in the Cloud

EXECUTIVE BRIEF

Summary

This executive brief analyzes the costs for running a high performance storage environment in the cloud, such as what would be used for a typical AI or HPC workload. The analysis specifically compares a multi-petabyte storage environment based on Managed Lustre to a similar environment based on Hammerspace running on standard cloud VMs.

The analysis shows that **Hammerspace delivers 57% annual cost savings compared to Managed Lustre**, as shown in the chart below.



The main factors that drive this savings are:

- **Complexity and High Costs of Lustre:** The high costs of tuning, deploying, and administering a complex file system like Lustre in the cloud translate to high managed services costs. Hammerspace, by contrast, runs on standard cloud instances and is built on a standards-based architecture that provides equivalent performance without the complexities of Lustre. In other words, Hammerspace delivers the ease-of-use of a fully managed cloud service at the cost point of standard cloud infrastructure services.
- **Ability to "Right-Size" Environment with Hammerspace:** Managed services can be rigid in their ability to provision capacity independent of performance. Excess capacity is often required to achieve a specific performance goal. Hammerspace, by contrast, is easy to provision, configure, and manage in the cloud.

Because it can scale up and down elastically, it enables IT infrastructure teams to “right size” their infrastructure with the ease-of-use of a fully managed cloud service, but at the cost point of standard cloud infrastructure services. This delivers significant cloud cost savings, as shown in this analysis.

- **Hammerspace Automates Data Movement Between Tiers:** Hammerspace automates the movement of data between storage tiers, and because all tiers are part of a single namespace, files remain visible to users and applications regardless of the tier on which they reside, and even as they move between tiers. Contrast that to the bulk multi-copy transform (file <> object) for data processing in a non-Hammerspace environment. There are multiple atomic operations - in other words file copies - that need to occur, and all of this adds cost and complexity, adds administrative burden, and results in a poor experience for data users.

The Growing Use of Hybrid- and Multi-Cloud Accelerated Computing

AI and HPC workloads require a large amount of compute resources and low-latency access to large data sets. Compute resources are not always co-located with the data, and cloud computing offers a fast and elastic option for ‘burst’ compute capacity. In fact, NVIDIA survey results from earlier this year indicate that 49% of their customers plan to run AI projects both in the cloud and on-premises in 2025. So there is a strong and growing demand for hybrid- and multi-cloud accelerated computing, and a need for high performance cloud storage systems to feed those compute clusters.

Options for High Performance File Storage in the Cloud

In order to provide fast access to data to the cloud GPUs, it is critical to have fast storage. There are several options for providing fast file storage in the cloud, but one of the more popular is to use Lustre, the open source file system that is available as a managed service from all of the leading cloud providers.

Hammerspace offers a viable alternative to Lustre, with a standards-based parallel file system architecture that provides low-latency, high-throughput data access to GPU clusters, both on-prem and in the cloud. Unlike Lustre, which requires a proprietary file system client, Hammerspace uses standard NFS connectivity - no proprietary client required.

Hammerspace is also a global file system, creating a Global Namespace that spans multiple sites, multiple cloud regions, and multiple cloud providers. Hammerspace Data Orchestration services automate the flow of data within that Global Namespace, so customers can seamlessly and efficiently bring their large datasets to the compute resources that need them, whether they are local or cloud-based.

Unlike Lustre, Hammerspace was designed for the enterprise, and includes the reliability, availability, and serviceability (RAS) features that are required by global enterprises, while also providing a rich set of enterprise data services for data orchestration, protection, and intelligent tiering.

Cost Comparison Analysis and Assumptions

The goal of this analysis is to compare the costs of running a typical high performance storage environment in the cloud using Managed Lustre to the costs of running a similar environment in the cloud using Hammerspace. The costs of Root Disk and Archival Storage are shown for completeness, but costs have been kept the same for the purposes of this analysis.

The table below summarizes the comparison and results.

Storage Class	Cloud Managed Services			Hammerspace on Cloud Infrastructure		
	Cloud Service	Capacity (TB)	Annual Cost	Cloud Service	Capacity (TB)	Annual Cost
Root Disk - 500GB SSD	Virtual Disk		$212,626	Virtual Disk		$212,626
High Performance Shared	Managed Lustre	1504	$4,277,110	Cloud VM with 32 vCPUs (QTY 10)	1504	$1,530,000
NFS-Mounted Shared	Managed File Service	1504	$2,887,499	Cloud VM with 16 vCPUs (QTY 7)	2240	$779,520
Archival / Object	Cloud Object Storage	3008	$649,688	Cloud Object Storage	3008	$649,688
				Hammerspace Software Components		
				Hammerspace Licensing		$675,200
				Anvil (2x VMs with 32 vCPUs)		$31,676
				DSX Nodes (2x VMs with 16 vCPUs)		$26,280
	Total	**6016**	**$8,026,923**		**6752**	**$3,904,989**

Assumptions

- Both environments use a total of three storage classes within the storage environment. These classes are defined below, and the specific services used for each class are shown in the table above.
  - **High Performance Shared Storage:** This is high computation shared storage, often referred to as the "scratch space" in HPC lexicon. This is used for computation, often "check-pointed" used in loosely and tightly coupled HPC workloads.
  - **NFS-Shared Storage:** This is often referred to as home directories or pre-staging. The data preparation and collection of data that the HPC cluster will compute against is stored here. Also, the results of the compute jobs are written here after all calculations are finished.
  - **Archival Storage:** Used for archival resubmission, often used to share / publish results between institutions. (Note: Although included for completeness, the costs of archival storage are the same in both environments.)
  - For completeness, the costs of **Root Disk** are also included in the analysis, but the figures are the same for both environments.
- The total storage capacity compared is just over 6 Petabytes, with the assumption that:
  - 25% of total storage capacity is High Performance Shared Storage
  - 25% of capacity is NFS-Shared Storage
  - 50% of capacity is Archival storage
- Both environments were designed to achieve similar performance results
- The Hammerspace software components include the Hammerspace software licensing costs, as well as the services required to run the Hammerspace infrastructure in the cloud - two Anvil metadata servers, and two data services or DSX nodes to manage data orchestration.
- Costs are based on cloud commercial costs with no discounts applied

Conclusion

It can take months and even years to deploy a new HPC cluster. Now with Hammerspace organizations can spin up a cluster in the cloud at a fraction of the cost of a traditional HPC file system like Lustre.

Hammerspace makes the movement of files in and out of the cloud a breeze. Declarative policies, called Objectives, ensure data is moved efficiently into and out of the cloud, and resides in the appropriate storage tier on the cloud.

By making it easier to use elastic cloud resources for burst computing capacity, organizations can “right size” their infrastructure with the ease-of-use of a fully managed cloud service, but at the cost point of standard cloud infrastructure services. This delivers significant cloud cost savings, as shown in this analysis.

Hammerspace also delivers unique value by enabling infrastructure teams to consolidate their unstructured data across data centers and multiple cloud regions, unifying new and existing data into a Global Namespace that spans sites, clouds, and storage.

To learn more, go to **https://hammerspace.com/get-started**