



Customer

Leading AI Model Developer

Use Case

Frontier Large Language
Model Training

Contact Us

P: +1 (650) 757-8400

W: <https://hammerspace.com>

E: contact@hammerspace.com



CASE STUDY

Challenge

A leading AI model company faced a critical challenge as they prepared to train a new frontier foundation model. Their compute environment consisted of a large-scale GPU cluster with thousands of GPUs. Each GPU server contained 31 - 61 TB of local NVMe storage depending on the specific model, and prior to Hammerspace this storage was unused and was a sunk cost.

They had run out of capacity and had no storage budget to spend, and meeting an aggressive time-to-market deadline for their next AI product launch was an imperative.



Solution

To overcome the storage bottleneck and keep their training schedule on track, the company deployed Hammerspace Tier 0 in their production environment. They brought approximately 20 PBs of Tier 0 storage capacity online in days, turning that idle GPU server local storage into a large pool of highly performant shared storage.

It is worth highlighting that no client software or agent needed to be installed on any of their GPU servers, they only needed to make Hammerspace aware of those storage volumes so they could be managed, protected, and served as shared storage as part of the Hammerspace file system.

The company automated data protection using client-side mirroring, orchestrated by Hammerspace policies within a unified global namespace that spanned the entire GPU cluster.

Data was accessed using standard file and object protocols.

Outcomes

Brought High-Speed Tier 0 Capacity Online Quickly

Easily “turned on” approximately 20 PBs of high-speed Tier 0 capacity in days, compared to the weeks or months it would take to deploy external storage.

Met Critical Launch Deadline

Maintained model training velocity and delivered on time.

Significant Cost Avoidance

Avoided approximately \$7 million in capital expenditures by avoiding the need for additional shared storage systems.

Zero Additional Power Draw

Achieved storage expansion without increasing datacenter power consumption.

