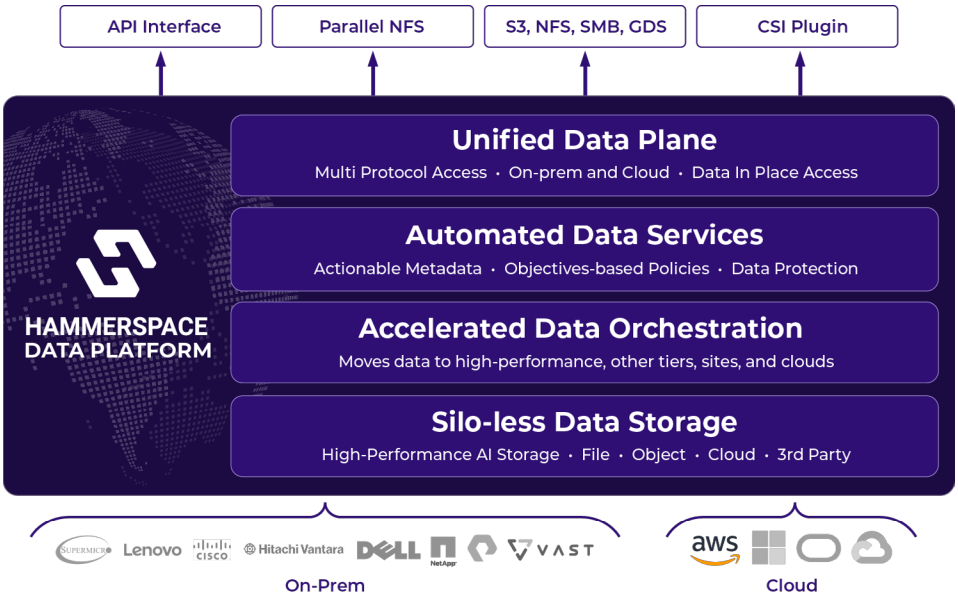
Hammerspace and AWS

SOLUTION GUIDE

Unify, Orchestrate and Accelerate Your Data in the Cloud

Hammerspace helps AWS Cloud customers unlock the full value of their data by breaking down silos across on-premises, cloud providers, and cloud regions. Together with AWS, we make it easy to move data where it's needed, speed up AI and HPC projects, and simplify hybrid-cloud operations. The result is faster time to value, lower infrastructure complexity, and greater consumption of AWS compute and storage resources.

Hammerspace is available in the AWS Marketplace and runs on standard AWS instances.



Solution Benefits:

- Unify Distributed, Siloed Data with a Global Namespace
- Migrate Data from On-Prem and Other Clouds to AWS
- Run Demanding AI/HPC Workloads in the Cloud
- Simplify Hybrid-Cloud Infrastructure
- Streamline Cloud Data Pipelines and Workflows
- Drive AWS Compute and Storage Consumption

Deploying Hammerspace on AWS

Hammerspace software is deployed on standard AWS infrastructure. There are two components of the Hammerspace software architecture:

> **Metadata Servers**, · Metadata Servers, called Anvils, host the metadata database, act as the control plane, and provide routing information to the clients. There are two Anvil metadata servers per cluster, in other words per site, for high-availability.
>
> **Data Services nodes**, or DSX nodes, are used to scale out client access across NFSv3, SMB, and S3 protocols, can act as a data store in the cloud and on-prem, and can also act as a data mover for NAS-to-NAS and NAS-to-Object data mobility. There are 2 to 60 DSX nodes per cluster depending on the performance and scale requirements.Hammerspace can also add storage volumes from any Linux storage server and make those volumes part of the Hammerspace file system. For example, you can add high-performance storage by using Google VMs with NVMe storage.

Hammerspace can also add storage volumes from any Linux storage server and make those volumes part of the Hammerspace file system. For example, you can add high-performance storage by using AWS instances with NVMe storage.

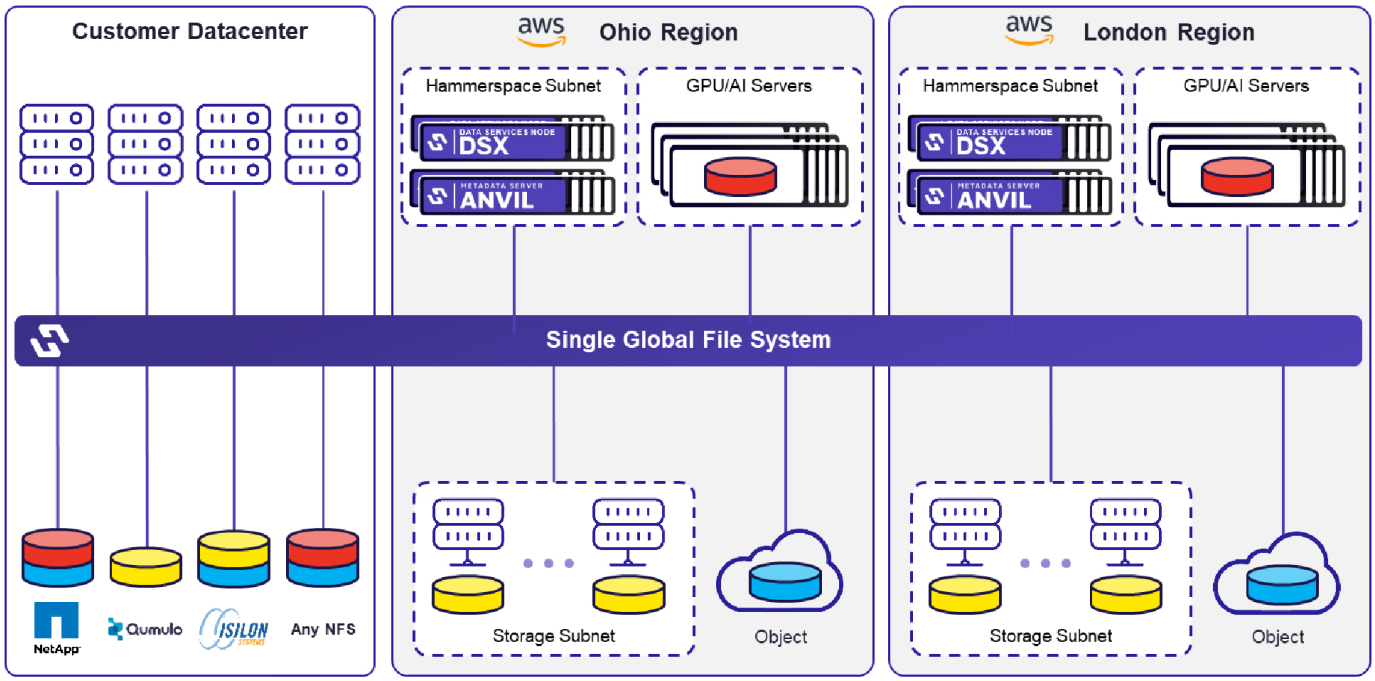
Hammerspace allows you to add the NVMe storage volumes that are local to AWS compute instances and aggregate these volumes together to create a tier of ultra-fast, low-latency NVMe storage that is called “Tier 0.”

With Tier 0, files stored on these local NVMe volumes can be protected by mirroring data on different nodes, and data can also be tiered off to external storage - for example AWS S3 - for long term storage. No matter where the files are stored physically within the Hammerspace global namespace, the files and objects remain visible and accessible via NFS, SMB, and S3.

For additional information, read the Getting Started Guide, and the Hammerspace AWS Marketplace Deployment Guide.

Example Hybrid-Cloud Deployment

An example hybrid-cloud deployment is shown in the figure below, which shows one customer data center and two AWS cloud regions connected with a Hammerspace global file system.



In this example, the Hammerspace metadata servers are deployed on (2) AWS instances, and the DSX nodes are also deployed on (2) AWS instances. Each of the AWS cloud regions have GPU VMs that are connected to the Hammerspace Data Platform.

The local drives in these VMs are used as a Tier 0 layer for the applications running on those instances. A Tier 1 storage layer is also being used, which are Linux VMs with internal NVMe drives in them. And finally, AWS S3 is used as a tiering bucket as well as the shared object store used to transport data between sites.

Performance Benchmarking on AWS

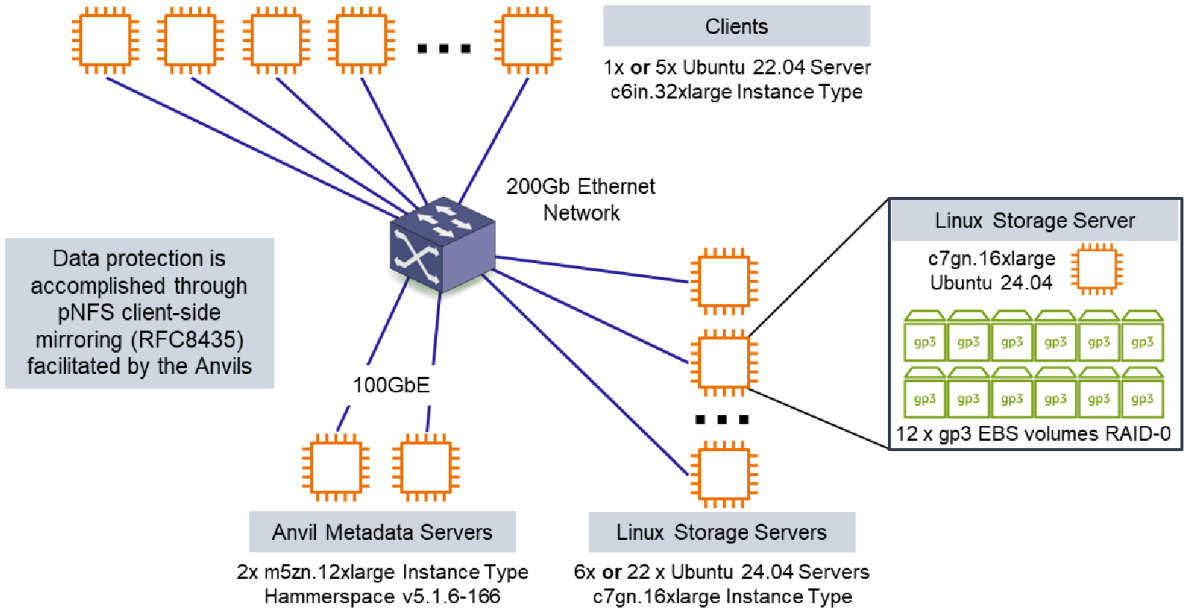
In order to demonstrate Hammerspace performance for AI and machine learning workloads on AWS, Hammerspace conducted MLPerf Storage Benchmark testing on the AWS cloud. The full results and analysis can be read in this technical brief.

Test Environment

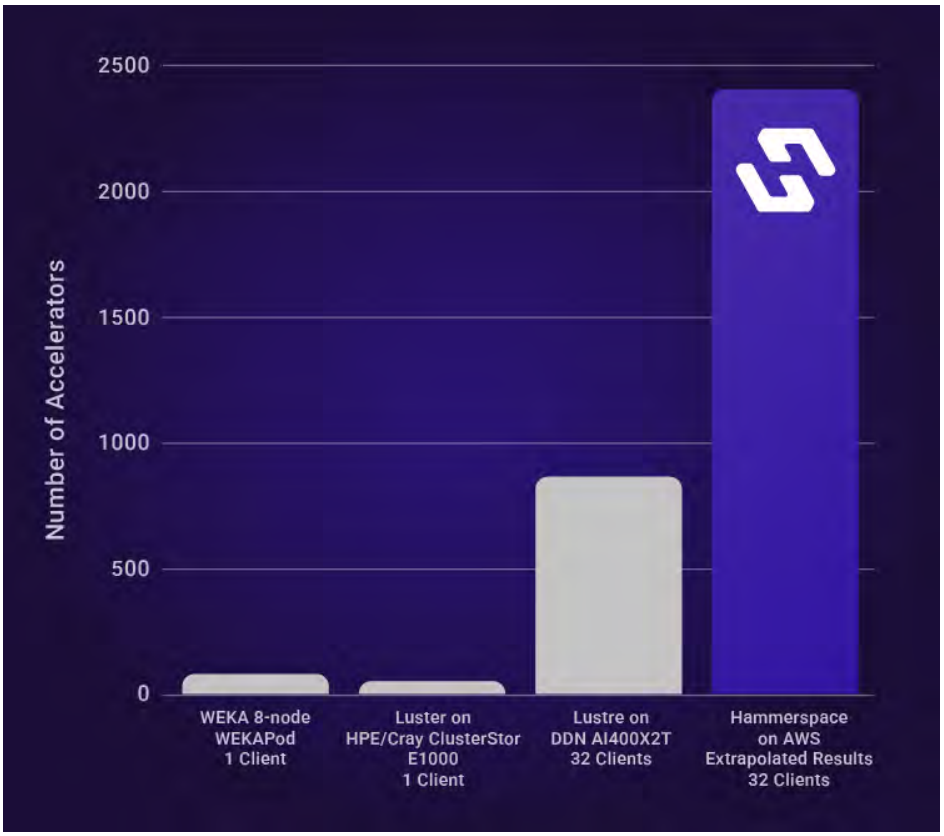
The system consisted of two redundant Anvil metadata servers in an active/passive configuration and up to 22 Linux storage server (LSS) nodes. The Anvil servers are responsible for metadata operations and cluster coordination tasks, while the LSSs serve test data using associated solid-state EBS volumes.

Single-client tests used six LSSs, and multiple-client tests used all 22. It's important to emphasize that the LSS nodes are just standard Linux servers exporting NFSv3, with no added software. All client systems mounted a Hammerspace share using standard pNFSv4.2 with Flexible Files layouts.

Clients and LSSs were connected to the network using 200GbE interfaces. Anvil nodes were connected via 100GbE. Since Anvils are only involved in metadata communication (no data flows through them), they did not require 200GbE.



Test Results



Hammerspace delivered approximately **3x the performance*** of the closest competitor – on standard AWS infrastructure, without any specialized networking, and without the use of a proprietary file system client.

As measured by the number of 'accelerators' or simulated A100 GPUs supplied with data at >90% utilization, using extrapolated results based on the MLCommons MLPerf RESNET-50 benchmark