Hammerspace MLPerf® Storage v1.0 Benchmark Results

Delivering the best price/performance storage for GPU computing

TECHNICAL BRIEF

Summary

The MLCommons MLPerf Storage benchmark is intended to demonstrate the performance of various storage systems for simulated machine learning workloads, so that technical buyers and decision makers have some criteria when evaluating storage system performance for machine learning, deep learning, and other forms of GPU computing.

This year, Hammerspace submitted results for the MLPerf v1.0 Storage Benchmark for the first time. This technical brief summarizes the results of that testing, including:

- Background on MLCommons and the MLPerf Storage Benchmark
- A summary of Hammerspace's results relative to other vendors, including the test setups used for the benchmark
- A discussion on the advantages of Hammerspace standards-based parallel file system architecture compared to scale-out NAS and HPC parallel file systems

The results prove the price/performance advantage of Hammerspace for high-throughput, low-latency file and object storage, both on-prem and on-cloud.

Hammerspace Price/Performance Advantage

Hammerspace is the only vendor that demonstrated HPC (high-performance computing) levels of performance with the standard networking and interfaces of Enterprise storage.
A bit of background: Parallel file systems are needed to achieve the extreme scalability and performance necessary to support HPC workloads, such as AI/DL and Mod-Sim (modeling/ simulation). Historically parallel file systems such as Lustre, BeeGFS, and others have been deployed for these use cases, which have predominantly been in specialized HPC environments. But the problem is these file systems are famously complex, very fragile, subject to downtime, and require expensive networking such as InfiniBand to perform at the level required. Most importantly, these solutions require specialized client software for every server accessing the storage, which requires that applications be designed to operate with the proprietary interfaces of the system.

While classic HPC research environments have adapted to this reality, the expansion of AI has driven HPC-class workloads into the broader enterprise and hyperscaler data center, where such limitations are not acceptable. The added cost, operational complexity, and specialized talent required to keep classic HPC file systems operational is not practical in enterprise data environments, especially since the deployment and optimization of these HPC file systems can take weeks or even months to fine tune.

These complexities have caused many organizations to assess scale-out NAS as an alternative to HPC parallel file systems. Scale-out NAS solutions are easier to use, include enterprise reliability features, support standard Ethernet networking, and are well understood by conventional IT staff in enterprise data centers. Unfortunately, scale-out NAS cannot deliver the performance needed for high-performance use cases at scale.

- Scale-out NAS systems require 2x the number of servers and 2x the number of network ports when compared with parallel file systems, which create a direct data path between clients and storage. Scale-out NAS storage solutions also struggle to deliver performance at scale, which is why they have not submitted any results to the MLPerf benchmark.
- Traditional HPC parallel file systems like Lustre require proprietary client software that adds complexity and ongoing administrative burden and costs. They require custom hardware, and exotic and expensive networking technology like InfiniBand and Slingshot.

With its Hyperscale NAS architecture, Hammerspace brings together the best of these technologies while overcoming the negative challenges of each, to deliver the best price/performance storage for GPU computing in AI, ML, HPC and deep learning. Hammerspace delivers the best price/performance by using:

- **50% fewer servers and 50% fewer network ports** than scale-out NAS architectures such as Dell PowerScale, Qumulo, and VAST
- **Standard Ethernet connectivity**, eliminating the need for a specialized second network, such as InfiniBand, which is used in the MLPerf benchmarks published by other parallel file system vendors including DDN, HPE and WEKA
- **Existing Linux client servers** to connect to the file system without the need for any specialized client software
- **Existing applications** natively designed to interface via NFS, SMB, or S3 can do so, without the need to redesign for a parallel file system interface

By reducing the number of servers and switches, this also reduces the amount of power and cooling required. While this can save costs, most importantly it also frees up wattage for the compute environment.

Because Hammerspace software supports any server hardware and storage solutions from any vendor, including in the cloud, buyers are free to purchase hardware from any source, including OCP servers such as those being used by Meta in their AI Research Supercluster. **Meta chose Hammerspace** as their high-performance solution for provisioning their Llama 2 and Llama 3 LLM training pipelines, because only Hammerspace demonstrated the linear scalability to achieve over 12TB/sec over standard networking, feeding data between Meta's existing 1,000-node NVMe storage cluster and a 3,000-node GPU cluster with 24,000 GPUs in total. No other vendor came close.

About MLCommons® and the MLPerf® Storage Benchmark

"MLCommons is an Artificial Intelligence engineering consortium, built on a philosophy of open collaboration to improve AI systems. Through our collective engineering efforts with industry and academia we continually measure and improve the accuracy, safety, speed, and efficiency of AI technologies–helping companies and universities around the world build better AI systems that will benefit society." (from **https://mlcommons.org/about-us**)

The MLPerf Storage Benchmark Suite consists of several simulated AI workloads. Hammerspace chose to run two of them:

- **3D-Unet:** A visual ML workload segmenting 3D medical imagery. This is a bandwidth-intensive test that opens large files in small batches and reads them sequentially.
- **ResNet-50:** A deep learning convolutional neural network that excels at image classification – detecting objects within images and classifying them accordingly. This test involves concurrently reading many smaller (~100KB) samples from within a large number (>1000) of larger (>100MB) files. Compared to U-Net3D this workload consists of smaller, more random I/O.

For each test, the goal is to demonstrate the maximum number of simulated GPUs ("Accelerators" in MLPerf terminology) that the storage system can simultaneously supply with data, keeping the utilization of every simulated GPU at 90% or higher. Total throughput is also reported in average MB/s.

There are two divisions defined by MLCommons for testing, Open and Closed. Closed-division rules establish a level playing field by requiring the use of a defined set of benchmark tuning parameters and options. This is intended to enable apples-to-apples comparisons between storage systems. Open division submissions have more flexibility for tuning both the benchmark and storage system configuration, and results are not directly comparable. The Open division is designed to allow vendors to showcase unique approaches or features that provide benefits when running AI/ML workloads.

MLPerf Storage benchmarks are unique in that results submitted during a defined time window that successfully complete the MLCommons review process are considered "verified" results. Results submitted outside the defined window are not subject to review and are considered "unverified."

Hammerspace completed two rounds of testing. The first round (Test 1 in this document), performed using cloud infrastructure, was in the Closed division and was verified by the MLCommons Association. The second round (Test 2 in this document), performed on-premises in the Hammerspace performance test lab, was in the Open division and is not verified by the MLCommons Association.

Results Summary

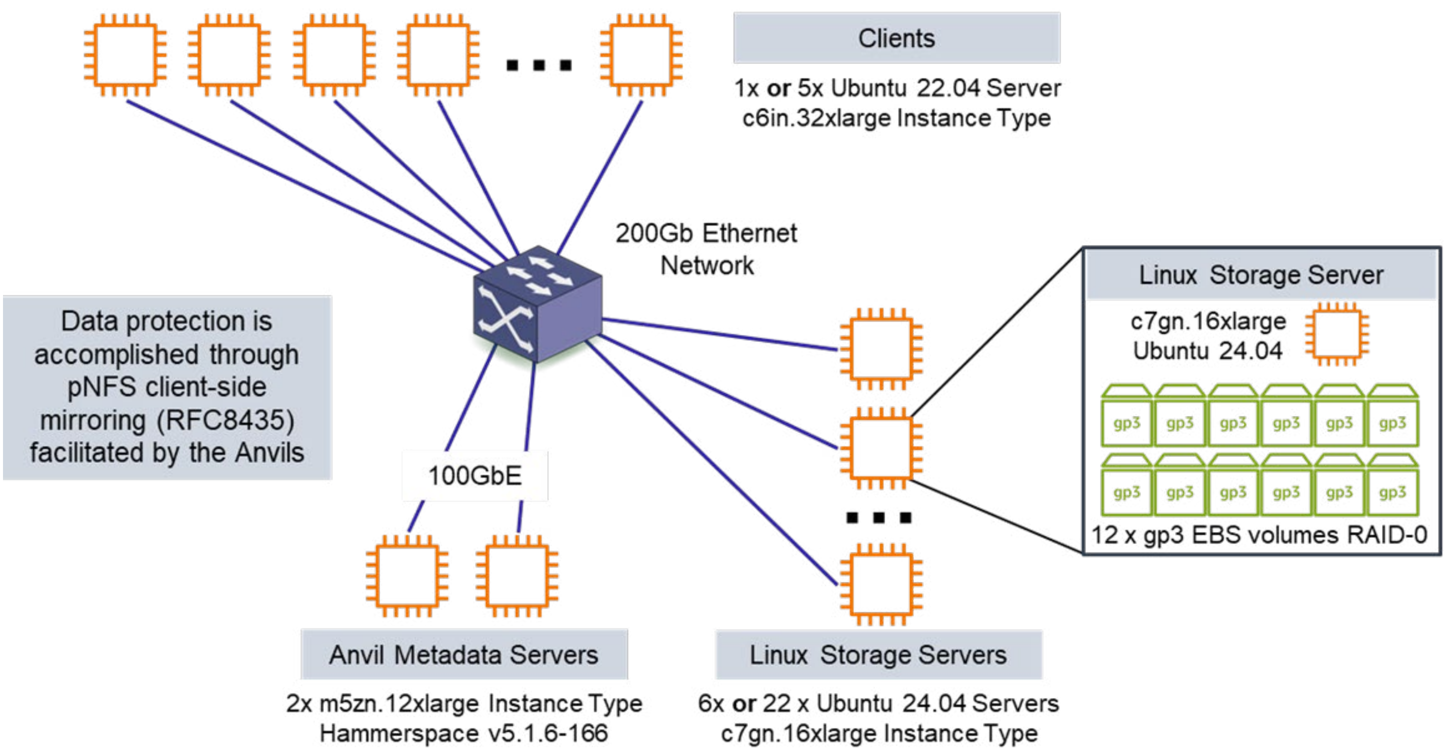
- **Hammerspace Demonstrated Excellent Performance Results:** Hammerspace delivered excellent performance results as measured by the number of simulated GPUs, and throughput that could be driven from a Hammerspace storage system.
- **Hammerspace Unlocks GPU-Local NVMe Storage:** Tier 0 turns the typically unused NVMe storage inside GPU servers into a new tier of ultra-fast shared storage. Because Tier 0 is part of the Hammerspace parallel global file system, data residing there benefits from the full complement of Hammerspace data services, including data protection and data orchestration. Unlike parallel file systems that use proprietary client software, Tier 0 has virtually zero impact on CPU utilization.
- **Hammerspace Does Not Require a Proprietary File System Client:** Unlike other HPC parallel file systems that submitted results to MLPerf, including Lustre and Weka, Hammerspace does not require a proprietary file system client. Instead, Hammerspace uses capabilities built into the NFSv4.2 client to deliver low-latency, high-throughput parallel file system performance without the complexities of proprietary client software. In addition, pNFSv4.2 can connect to any storage from any vendor via standard NFSv3.
- **Hammerspace Does Not Require Specialized Networking Technologies:** Unlike the other HPC parallel systems which submitted results based on InfiniBand or HPE Slingshot networking, Hammerspace results are based on standard Ethernet networking.
- **Hammerspace Delivers Performance Both On-Cloud and On-Premises:** Notably, Hammerspace was the only parallel file system vendor with verified results based on a cloud-native test environment, demonstrating the flexibility of Hammerspace to act as a high-performance file system whether deployed on physical hardware or cloud environments.

Test 1: Public Cloud Hyperscale NAS Configuration

The results of this test in the Closed division are verified by MLCommons Association as shown on the [MLPerf Storage Results page](#).

Test System Configuration

Testing was performed using Amazon Web Services (AWS) public cloud infrastructure. Cloud was selected for this MLPerf submission to show the performance that can be achieved in standard cloud instances without specialized hardware designs.



The Hammerspace 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 LSS nodes serve test data using associated solid-state EBS volumes storage devices. Single-client tests used six LSS nodes, 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.

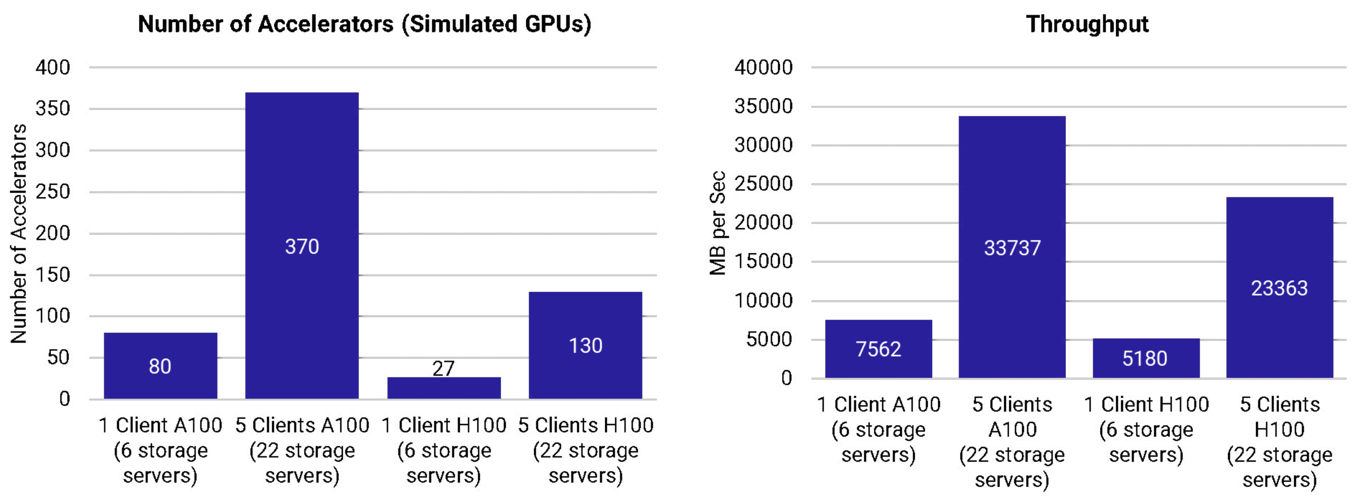
All client systems mounted a Hammerspace share using standard pNFSv4.2. This is the [Hyperscale NAS](#) architecture.

Clients and storage servers 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), 200GbE was not necessary.

Hammerspace Results

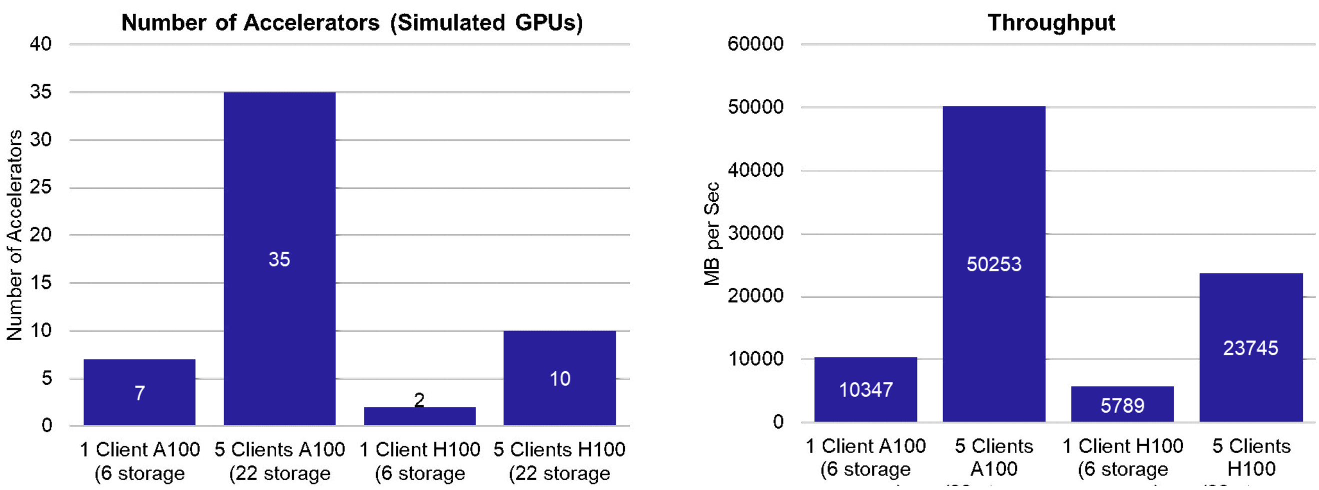
Hammerspace test results are shown in the figures below.

ResNet-50 Results



In the ResNet-50 image classification workload simulation, a Hammerspace system with 22 flash-based Linux storage servers (LSSs) drove 370 simulated A100 GPUs and 130 simulated H100 GPUs to > 90% utilization, delivering 33.7GB/s and 23.3GB/s aggregate read performance, respectively.

3D-Unet Results



With the 3D-Unet simulated image segmentation workload, this system drove 35 simulated A100s and 10 simulated H100s, delivering 50.3GB/s and 23.7GB/s, respectively.

A smaller configuration with six LSSs performed admirably as well, demonstrating the scalability of the system. It supported 80 simulated A100s at 7.6GB/s and 27 simulated H100s at 5.2GB/s in the ResNet-50 test, and 7 simulated A100s and 2 simulated H100s at 10.3GB/s and 5.8GB/s, respectively, in the 3D-Unet test.

Both Hammerspace configurations used standard Ethernet networking and standard pNFSv4.2, requiring no special client-side software or agents. Neither was tested to its limits.

Hammerspace Results Relative to Other Vendors

To compare Hammerspace results relative to other parallel file system vendors, we chose to compare the maximum number of accelerators delivered by each of these vendors, independent of the number of clients used.

We chose this methodology because in the MLPerf Storage benchmark, clients ("host nodes" in MLPerf terminology) exist to generate the load placed on the storage system under test. They are not necessarily representative of real-world clients running an application. For example, a single MLPerf Storage client system may simulate dozens or even hundreds of GPUs, far more than could be attached to any computer in the real world. It is more useful to normalize based on the total number of accelerators supported by the storage system, independent of client count and type.

Comparing results head-to-head is difficult, because each vendor used different numbers and types of clients, as well as different numbers and types of storage systems, as shown in the comparison table below.

3D-Unet - H100 Simulated Accelerator Results

Vendor and Systems	DDN AI400X2 Turbo	Hammerspace	HPE Cray C500	HPE Cray ClusterStor E1000	Weka 8-node WEKAPod
Max # of Accelerators	36	10	3	15	13
Number of Clients	18	5	3	5	1
Number of Storage Servers	1	22	1	1	8

ResNet-50 H100 Simulated Accelerator Results

Vendor and Systems	DDN AI400X2 Turbo	Hammerspace	HPE Cray C500	HPE Cray ClusterStor E1000	Weka 8-node WEKAPod
Max # of Accelerators	512	130	20	34	74
Number of Clients	32	5	10	1	1
Number of Storage Servers	1	22	1	1	8

Test Configuration

Vendor and Systems	DDN AI400X2 Turbo	Hammerspace	HPE Cray C500	HPE Cray ClusterStor E1000	Weka 8-node WEKAPod
Client Description	Bare metal nodes 1x CPU, 92GB RAM	AWS C6in.32xlarge 128x Virtual CPUs, 256GB Ram	Single CPU socket 131GB memory	Single CPU socket 131GB memory	NVIDIA HGX H100 Dual socket, 2048GB RAM
Network Type	Infiniband HDR100 links (one per client)	200GbE Virtual Network	Infiniband	HPE Slingshot	400Gb/s NDR Infiniband
Storage Server Description	DDN AI400X2 Turbo w 24x Phison NVMe Drives, 14TB each	c7gn.16xlarge, AWS Graviton3 64x vCPUs, 128GB 12x 500GB GP3 EBS Volume, Raid0	Cray C500	Cray ClusterStor E1000 w 2x AFAs, 24x 7.68TB drives	1U WekaPOD Servers AMD 48-core, 384GB RAM 14x 3.84TB TLC NVMe SSDs

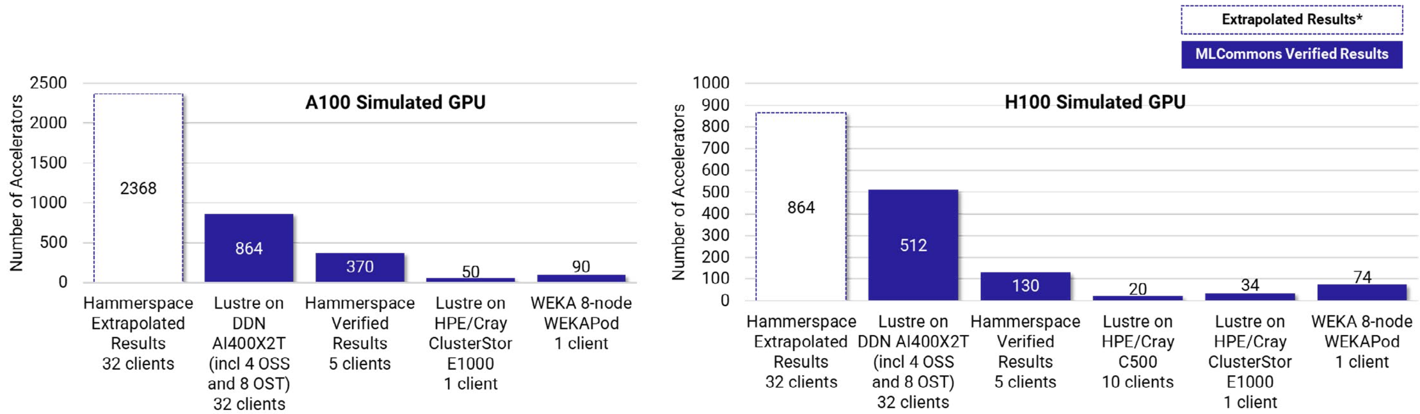
The notable differences are summarized in the chart below, which highlights that Hammerspace was the only parallel file system vendor that used cloud infrastructure for this testing, used standard Ethernet, and standard NFS connectivity for this submission.

	Hammerspace	DDN	HPE	Weka
Infrastructure	**Cloud**	**Physical Hardware**	**Physical Hardware**	**Physical Hardware**
Networking	**Ethernet**	**Infiniband**	**Infiniband**	**Infiniband**
Client Connectivity	**Standard NFS**	**Lustre Client**	**Lustre Client**	**Weka Client**

Based on the methodology described above, and noting the differences in both client type and client count, etc. – the charts below show Hammerspace results relative to other parallel file system vendors.

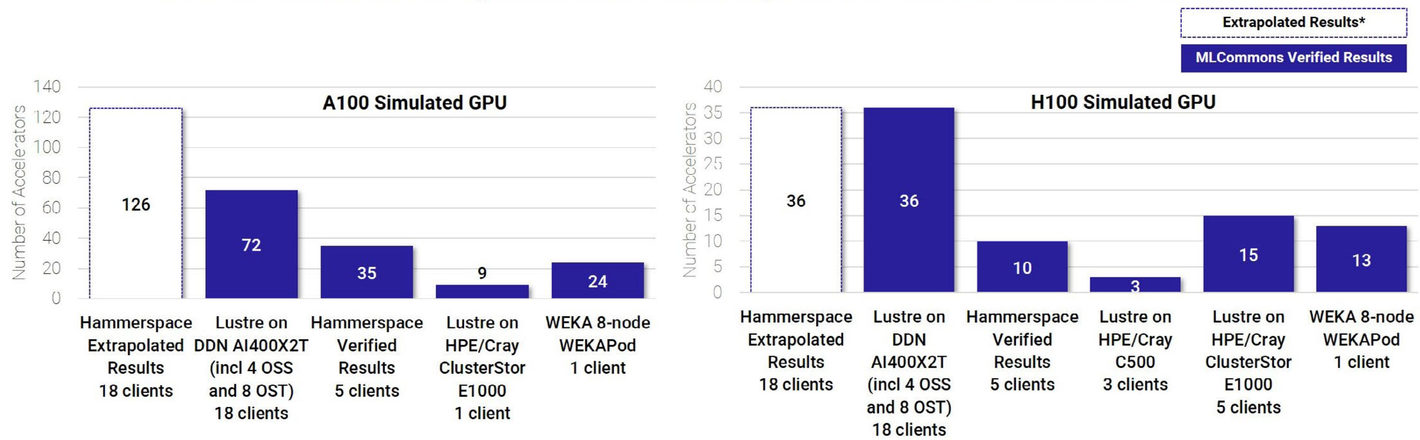
ResNet-50 Results

NOTE: These charts include both single-client and multi-client results, based on the client counts show beneath each vendor.



U-Net3D Results

NOTE: These charts include both single-client and multi-client results, based on the client counts show beneath each vendor.



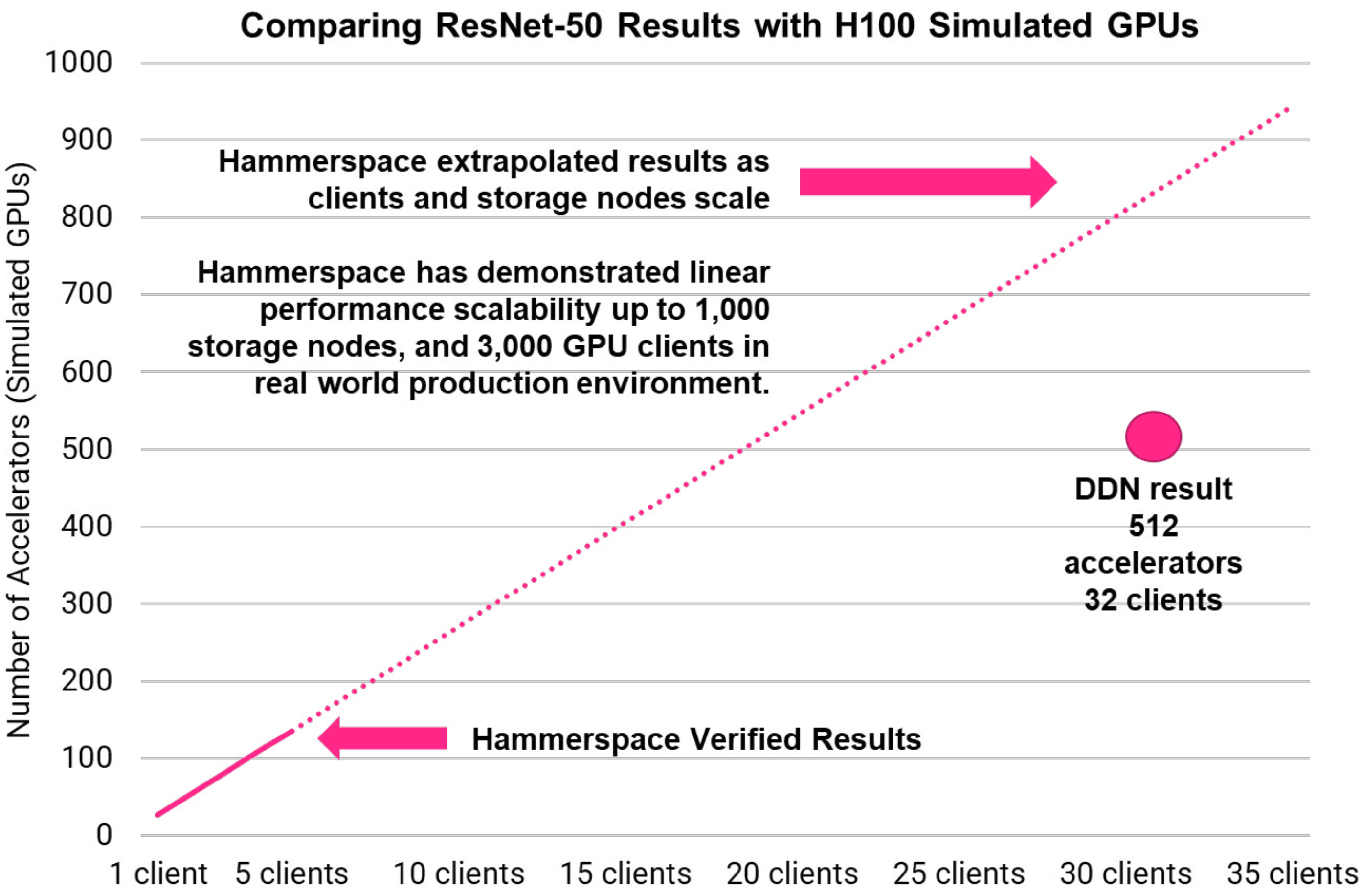
** The extrapolated results show the expected result at the same number of clients as submitted by DDN. These extrapolated results are based on Hammerspace internal analysis, would require a different AWS instance configuration, and have not been verified by MLCommons.*

Explanation of Verified Results vs Extrapolated Results

To provide prospective buyers with a somewhat normalized result, the charts above display both the MLCommons verified results that were submitted as part of the v1.0 benchmark, as well as extrapolated results based on Hammerspace internal analysis.

The extrapolated results show expected Hammerspace results based on the same number of clients used by DDN for their verified, submitted results. The intent is to communicate Hammerspace's linear performance scalability, based on these assumptions:

- Hammerspace submitted both single-client results and results based on five clients, which showed linearly scalable performance in that range.
- Further, Hammerspace has demonstrated linear performance scalability in real world production environments up to 1,000 storage nodes and 3,000 GPU clients with 24,000 GPUs.
- Because Hammerspace ran the MLPerf Storage benchmark on AWS Cloud infrastructure, scaling both the number of clients and the number of storage servers can be done simply using elastic cloud compute and storage resources.
- By adding more clients and storage nodes, Hammerspace would be able to deliver results that surpass the leading results, as shown in the diagram below.



Test 2: Hyperscale NAS with Tier 0

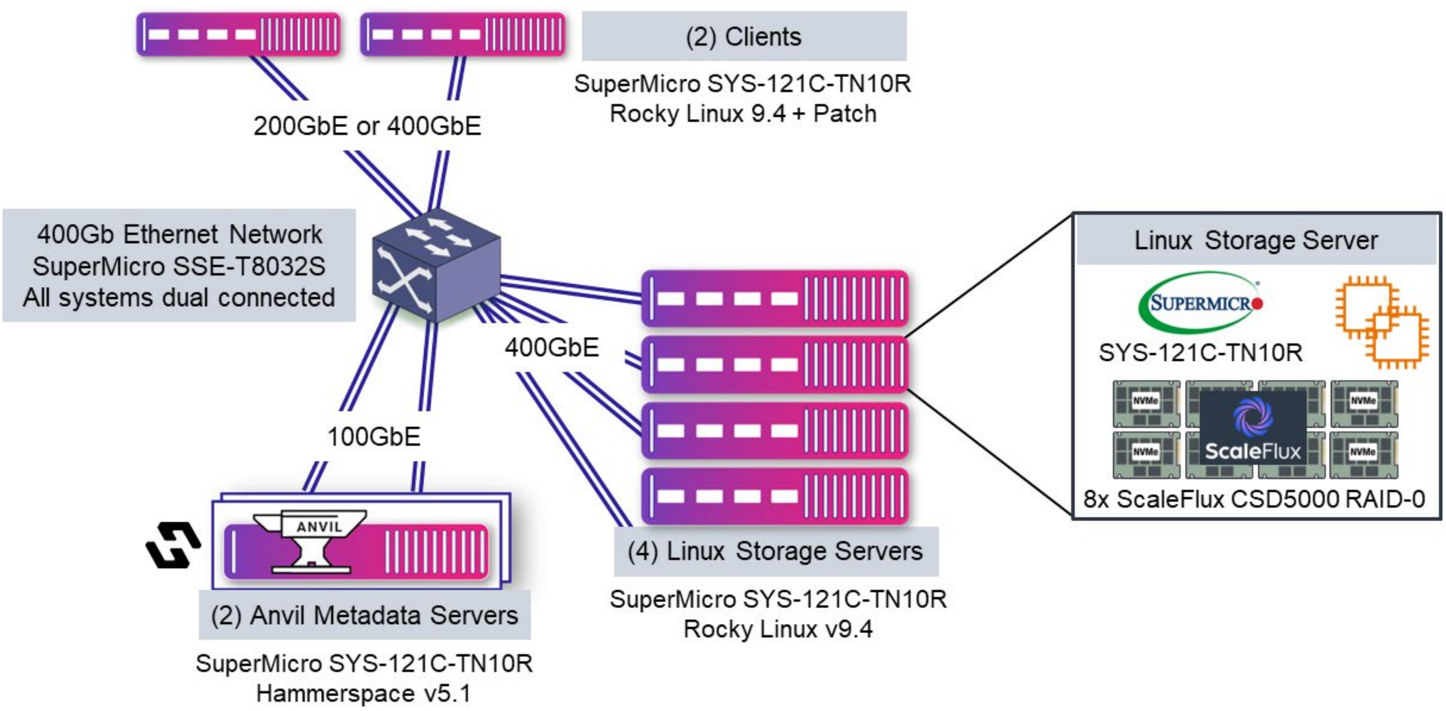
The results of this test in the Open division are not verified by MLCommons Association.

Testing was performed in the Hammerspace performance test lab on physical servers. To highlight the positive performance benefits of Tier 0, two configurations were used. The first (2a) was a typical Hyperscale NAS configuration using dedicated Linux Storage Servers (LSS). Two runs were completed with this configuration, one using 200GbE client connectivity, and one using 400GbE client connectivity. The second configuration (2b) used the storage within the clients themselves, the hallmark of Tier 0.

With Tier 0, the NVMe drives in each client machine are promoted from islands of difficult-to-use local storage to become a new tier of ultra-fast storage that is part of the Hammerspace parallel global file system. An NFS protocol bypass (**LOCALIO**), engineered by Hammerspace, and contributed upstream to standard Linux provides performance acceleration. Because Hammerspace leverages the software already built into the Linux kernel, it consumes only a tiny amount of CPU utilization, leaving more server resources for running business workloads.

To learn more about the benefits and applications of this new tier of ultra-fast shared storage, see the **Tier 0** page on the Hammerspace web site.

Test System – Configuration 2a



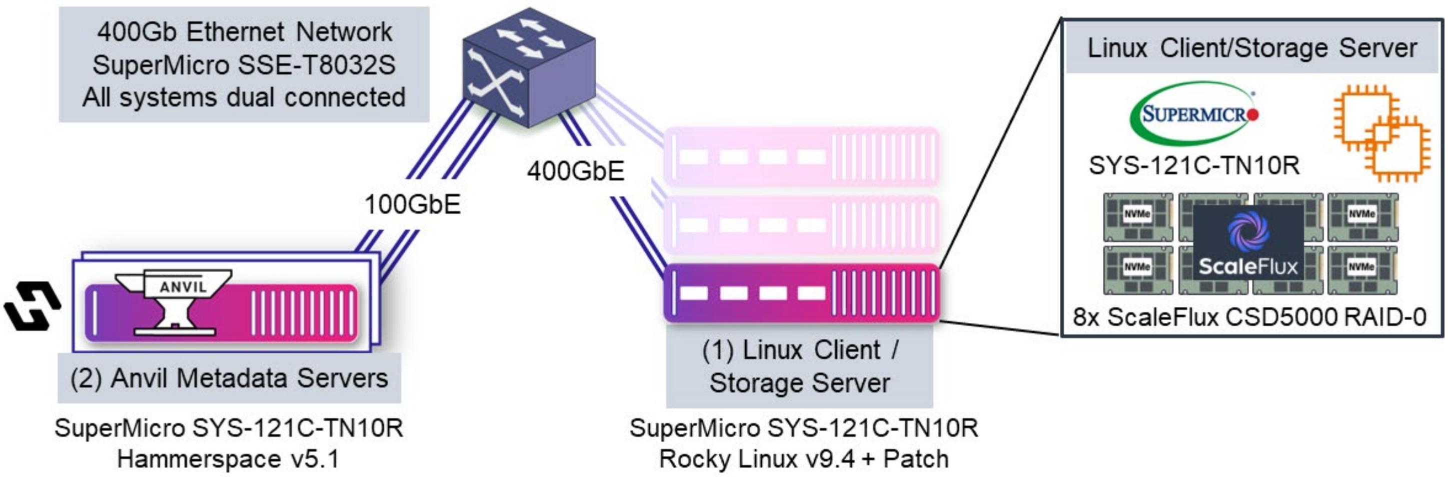
Configuration 2a is the Hyperscale NAS configuration using external shared storage. This is not the Tier 0 configuration - this test is used to show the same Hammerspace software and storage server hardware running the benchmark as a baseline for comparison with the Tier 0 results. This configuration consists of two redundant Anvil metadata servers in an active/passive configuration, four LSSs, and two clients. The Anvil servers are responsible for metadata operations and cluster coordination tasks, while the LSSs serve test data using internal ScaleFlux CSD5000 Computational Storage NVMe drives. No computational storage features of the ScaleFlux drives were used in this test.

It's important to emphasize that the LSSs are just standard Linux servers exporting NFSv3, with no added software.

All client systems mounted a Hammerspace share using standard pNFSv4.2. This is the **Hyperscale NAS** architecture. Like the LSSs, the clients are standard Linux servers. Unlike other parallel file systems, Hammerspace does not require the installation of any special software on clients to achieve peak performance.

Clients and storage servers were connected to the network using 2x 200GbE or 2x 400GbE interfaces each. Anvil nodes were each connected via 2x 100GbE. Since Anvils are only involved in metadata communication (no data flows through them), 100GbE was sufficient.

Test System – Configuration 2b



Configuration 2b is used to demonstrate Tier 0 performance. Two redundant active/passive Anvil metadata servers are responsible for metadata operations and cluster coordination tasks. The client in this test has two roles. It runs the benchmark code as usual, but is also the storage server, serving the test data from internal ScaleFlux CSD5000 Computational Storage NVMe drives. No computational storage features of the ScaleFlux drives were used in this test.

It's important to emphasize that the combination client/storage server machine is just a standard Linux server with no added software. The internal storage is exported via NFSv3 and mounted using pNFSv4.2. Though the file system metadata path traverses the network to the Anvils, the data path remains entirely within with client/storage server host, providing the client with direct access to the local filesystem using the Tier 0 NFS protocol bypass. This direct data path boosts throughput and reduces latency.

The client/storage server was connected to the network using 2x 400GbE interfaces. Anvil nodes were each connected via 2x 100GbE links. Since Anvils are only involved in metadata communication (no data flows through them), 100GbE was sufficient.

The two additional faded client/storage server machines shown in the diagram reflect the configuration that will generate the extrapolated three-client results shown on the graphs below.

Hardware Details

The same hardware was used for both configurations described above.

Hammerspace Anvil Metadata Servers (2)

Component	Manufacturer	Model	Details
Chassis	SuperMicro	SYS-121C-TN10R	1U Dual Socket
CPU (x2)	Intel	Xeon Gold 6542Y	60MB Cache, 2.90 GHz, 24-Core
Memory (x16)	Micron	MEM-DR564L-CL02-ER56	64GB DDR5-5600 ECC RDIMM 1TB Total RAM per server
Boot Drive (x2)	Micron	7450 MTFDKBA960TFR	960GB NVMe PCIe4 M.2 3D TLC
NIC (x2)	NVIDIA	ConnectX-7 Dx	Dual 400GbE QSFP56 PCIe4 x16
Storage (x4)	ScaleFlux	CSD5000	7.68TB NVMe PCIe4 U.2

Linux Storage Servers (4) and Clients (2)

Component	Manufacturer	Model	Details
Chassis	SuperMicro	SYS-121C-TN10R	1U Dual Socket
CPU (x2)	Intel	Xeon Gold 6542Y	60MB Cache, 2.90 GHz, 24-Core
Memory (x16)	Micron	MEM-DR564L-CL02-ER56	64GB DDR5-5600 ECC RDIMM 1TB Total RAM per server
Boot Drive (x2)	Micron	7450 MTFDKBA960TFR	960GB NVMe PCIe4 M.2 3D TLC
NIC (x2)	NVIDIA	ConnectX-7 Dx	Dual 400GbE QSFP56 PCIe4 x16
Storage (x8)	ScaleFlux	CSD5000	7.68TB NVMe PCIe5 x4 U.2

Network Switch

Manufacturer	Model	Details
SuperMicro	SSE-T8032S	Twin-port 2x 400GbE Open Networking Switch, SONiC OS

Software Details

The same software was used for both configurations described above.

Hammerspace Anvil Metadata Servers

Anvil nodes ran Hammerspace v5.1. Hammerspace is packaged as a single installation that includes the Linux operating system, application code, and all dependencies.

Linux Storage Servers

Linux Storage Servers ran Rocky Linux v9.4 with no additional patches.

Clients and Tier 0 Combined Client / Storage Server

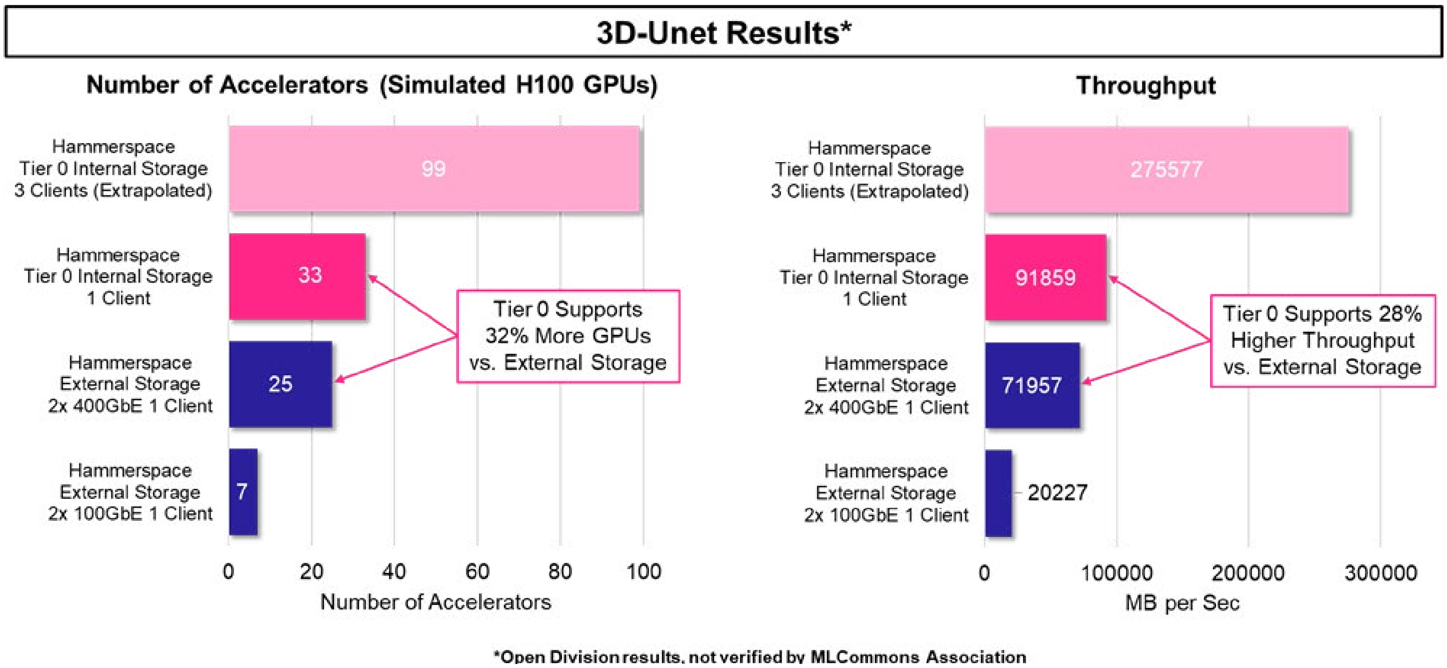
These machines ran Rocky Linux v9.4 with the addition of one upstream **kernel patch** to enable RDMA to honor the NFS nconnect mount option.

MLPerf Benchmark Code

Modified MLPerf benchmark code was used that enabled bypassing the client-side page cache. This is not required for Tier 0, it represents additional tuning to improve performance.

Hammerspace Results

The graph below summarizes the results of test two.



32% more GPUs were supported, and 28% higher throughput was observed when running the 3D-Unet test using Tier 0 storage internal to the GPU server vs. external shared storage. This was true even when the GPU server was connected to the external storage with two 400GbE interfaces.

Tier 0 provides meaningfully higher performance compared to networked external storage, even when the GPU server has high-bandwidth connections to the external storage, and especially if network bandwidth is constrained, as was the case with the client connected via 2x 100GbE.

Key Observations

The results of this test show clear benefits to using Tier 0. There are several observations worth highlighting.

Tier 0 eliminates network bandwidth constraints

It is common knowledge that high performance storage is required to keep GPUs busy processing data. The test results show that high-speed networking is also critical. Based on the jump from 7 to 25 GPUs between the 2x 100GbE and 2x 400GbE clients (and a commensurate jump in throughput), it's obvious that the 100GbE interfaces were a serious bottleneck.

The only thing better than a fast network is no network. Eliminating the network and using storage local to the GPUs provides the best possible performance. As shown in the graph, using Tier 0 local storage enabled the client to support 32% more simulated H100 GPUs than it could when accessing storage over 2x 400GbE, with 28% higher aggregate throughput.

In practical terms, this means for GPU servers with unused local NVMe storage and sub-optimal networking, Hammerspace software can provide the performance benefits of a major network infrastructure upgrade without the cost or headaches of replacing existing NICs and switches.

Tier 0 performance is linearly scalable

Tier 0 enables GPU servers to work on data that is stored locally. Hammerspace data orchestration is used to deliver data to Tier 0 storage, protect that data, and offload checkpoint files and computation results to other tiers of storage. Because processing is local, performance scales linearly as more GPU servers with Tier 0 storage are added to the cluster.

On the graphs below, the dashed outline shows this linear scaling potential as an extrapolated result. If three clients with Tier 0 storage were used for this test, the number of supported GPUs and aggregate throughput would triple vs. the single-client results.

Tier 0 provides Capex and Opex benefits

With Tier 0, Hammerspace brings existing GPU-local NVMe storage into a global shared file system, eliminating the obstacles that otherwise make it impractical to use. This has numerous financial benefits.

- **External Storage Offset:** Using local NVMe storage reduces the amount of external high-performance storage needed, along with the associated network, power, and cooling expenses.
- **Operational Time Savings:** Hammerspace software enables existing storage to be used in minutes, freeing up time that would otherwise be spent installing and configuring external storage and networking hardware.
- **CPU Efficiency:** Unlike traditional parallel file systems that require resource-hungry proprietary client software, Tier 0 operates with virtually zero CPU overhead. This preserves more server resources for running business workloads.
- **Increased GPU Efficiency:** Tier 0 reduces checkpointing durations from minutes to seconds. This unlocks significant additional GPU compute capacity, enabling jobs to be completed more quickly without investing in additional hardware. A full analysis of the checkpointing use case is here: https://hammerspace.com/whitepaper-tier-0-checkpointing/

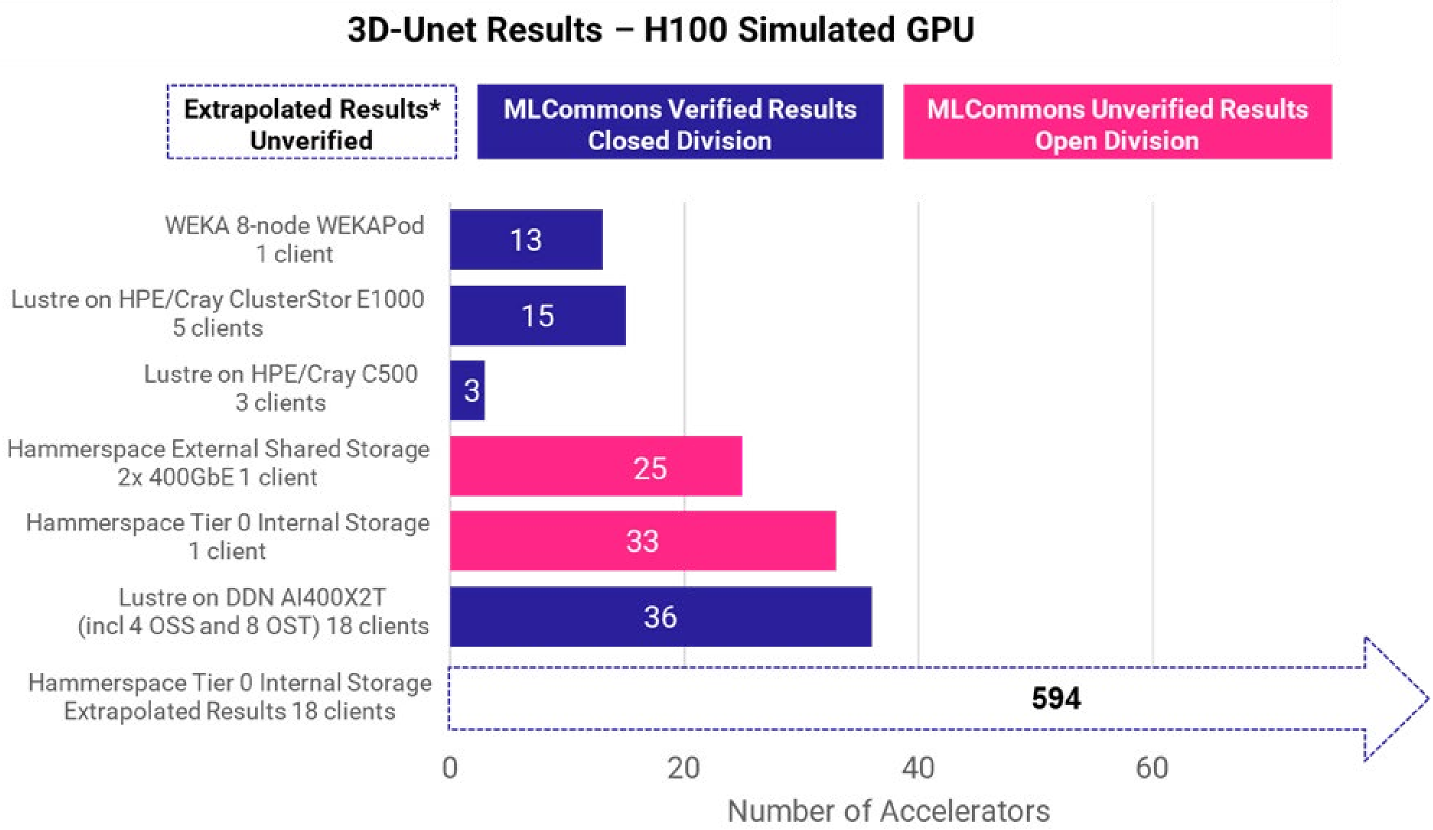
Hammerspace Results Relative to Other Vendors

As with test 1, we chose to compare the maximum number of accelerators delivered by each of these vendors, independent of the number of clients used.

We chose this methodology because in the MLPerf Storage benchmark, clients (“host nodes” in MLPerf terminology) exist to generate the load placed on the storage system under test. They are not necessarily representative of real-world clients running an application. For example, a single MLPerf Storage client system may simulate dozens or even hundreds of GPUs, far more than could be attached to any computer in the real world. It is more useful to normalize based on the total number of accelerators supported by the storage system, independent of client count and type.

Comparing results head-to-head is difficult, because each vendor used different numbers and types of clients, as well as different numbers and types of storage systems, as shown in the comparison tables in the Test 1 section of this document.

Based on the methodology described above, and noting the differences in both client type and client count, etc. – the chart below shows Hammerspace results relative to other parallel file system vendors.



Extrapolated results show expected performance at the same client count submitted by DDN and are based on Hammerspace internal analysis, not verified by MLCommons. Both single-client and multi-client results are shown.

The single Hammerspace Tier 0 client supported 33 simulated H100 GPUs, using only 1U of storage (within the 1U client) and a total of 3U of rack space (2x 1U for the Anvils, 1U for the client/storage server). This shows that Tier 0 configurations are very space- and power-efficient.

To provide prospective buyers with a somewhat normalized result, the chart above displays Hammerspace's Open division, unverified results for Tier 0, along with extrapolated results for Tier 0 with a larger client count, and MLCommons Closed division, verified results for the other vendors.

The extrapolation shows expected Hammerspace results based on the same number of clients used by DDN for their verified results. The intent is to communicate Hammerspace's linear performance scalability. This extrapolation is valid because with Tier 0, every client works on data housed on its local NVMe drives, therefore there are no performance dependencies between clients.

Hammerspace Architecture

The Hammerspace Data Platform has three key technologies all in a single software solution, with a single all-inclusive license:

1. Multiprotocol Global Namespace
2. Parallel Global File System
3. Automated Data Orchestration

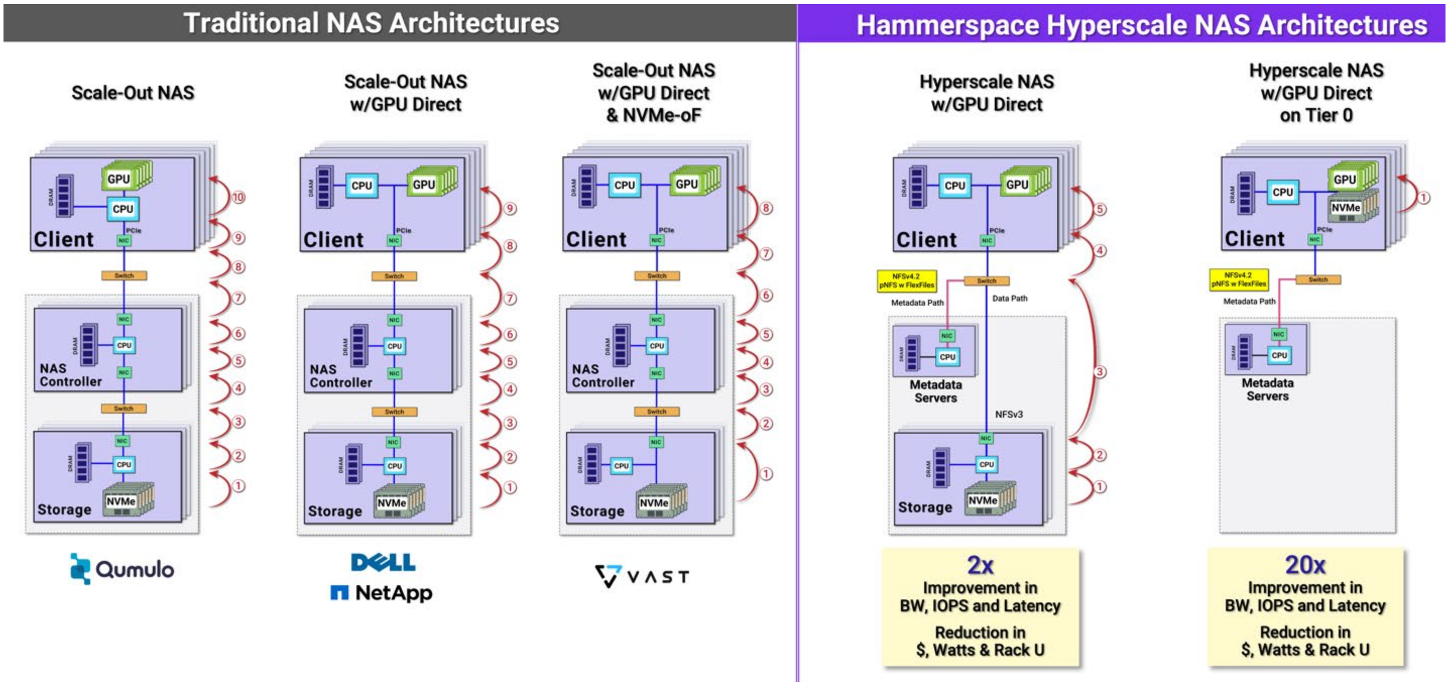
The high-performance storage portion of Hammerspace technology, which combines our parallel file system and global namespace, is delivered as a **Hyperscale NAS**. Hyperscale NAS uses standard Ethernet and requires no proprietary client like a traditional enterprise-class scale-out NAS, but couples this with the parallel file system performance and linear scalability found in industry-leading parallel file systems in the HPC space.

	SCALE-OUT NAS	LEGACY PARALLEL FILE SYSTEMS	HYPERSCALE NAS
SAFE Reliability, Availability, Serviceability	✔	✘	✔
EASY Standards-Based, Plug-N-Play	✔	✘	✔
FAST HPC-Class Performance	✘	✔	✔
AFFORDABLE Cost Effective at Scale	✘	✔	✔

Hammerspace has invested heavily into enhancing the standard NFS protocol to include a fast, feature rich, parallel file system client that is built into the Linux kernel of all commercial distributions.

Advantages of Hammerspace Hyperscale NAS Architecture

It is notable that no scale-out NAS vendor submitted results as part of the MLPerf Storage Benchmark. Well known NAS vendors like Dell, NetApp, Qumulo, and VAST Data are absent. Why wouldn't these companies submit results? Most likely it is because there are too many performance bottlenecks in the I/O paths of scale-out NAS architectures to perform well in these benchmarks.



A traditional Scale-out NAS design requires a NAS controller in the storage system. There are two key areas where this causes issues:

1. Performance, even at small scale

a. The eight or nine hops a single bit needs to take for a read or write operation in scale-out NAS architectures introduces latency

b. The data and the metadata in those systems are sharing the same network path, both trying to squeeze in to a fixed amount of network capacity (kind of like a traffic jam in Los Angeles – only so many cars can fit on the highway to go straight (Data path), and, when you add in cars merging to get to exits or the HOV lane, additional friction is created (metadata path)

c. If anything breaks or goes offline in the data path, the environments become fragile and slow, or go down completely, because they don't have the client-side intelligence and the telemetry feedback loop to automatically reroute around blockages which is part of the pNFS v4.2 standard.

2. Performance at large scale

Scale-out NAS architectures face performance thresholds at scale due to several key limitations inherent in their design:

1. **Metadata Bottleneck:** File system metadata is typically managed by a limited set of central controller servers. As the number of files and the volume of data grow, these controllers can become bottlenecks, slowing down file access and system performance that are also bottlenecked through the same controller.
2. **Network Overhead:** Since scale-out NAS architectures distribute data across multiple nodes, they rely heavily on network infrastructure to connect those nodes to each other behind the centralized controller. As the system scales, the internal network traffic and cache contention (for data retrieval, replication, synchronization, etc.) increases, and standard Ethernet or other networking technologies may not provide sufficient bandwidth, creating a performance ceiling.
3. **Concurrency Limits:** Scale-out NAS systems often face challenges with concurrent access, particularly when many users or applications are accessing the same data. Locking mechanisms to ensure data consistency (such as file locks) can cause delays and limit scalability, as more clients try to access or modify the same files simultaneously.
4. **Data Distribution Overheads:** As the system scales, managing where data is stored across the storage nodes becomes increasingly complex. Some scale-out NAS systems distribute data using techniques like hashing or striping across multiple nodes, which adds overhead as the system tries to keep track of where each piece of data resides.
5. **Protocol Inefficiencies:** Traditional client-side NAS protocols like NFSv3 or SMB, while widely used, are not optimized for massive scalability. They have a serious performance problem of excessive chattiness between the client and the server, because there is no intelligence in the client to retain state. This leads to overhead in terms of session management, data transfer, and security features that may limit throughput as the system scales.
6. **Non-Linear Scaling of I/O:** While additional storage nodes increase capacity, they can't linearly scale to improve input/output (I/O) performance. Data movement between nodes, replication overhead, and inter-node communication can cause I/O bottlenecks, limiting scalability of scale-out NAS architectures from all vendors.
7. **Latency Amplification:** As the number of storage nodes increases, latency from factors like network hops, inter-node communication, and coordination overhead grows, impacting performance.

Hammerspace Delivers High Performance On-Premises and In the Cloud

An NVIDIA survey to their customer base in 2024 indicated that 49% of their customers plan to run AI projects **both on-premises and in-cloud**. In other words, about half of NVIDIA's customers will require high-performance file and object storage that can run on-premises (when GPU clusters are local), and in the cloud (to access GPU resources in the public cloud). Hammerspace MLPerf results demonstrate excellent performance in a 100% cloud-based environment, showcasing the ability for enterprises to achieve low-latency, high-throughput file storage no matter where it runs.

And because Hammerspace is a global file system that can span sites and multiple clouds, with data orchestration services that automate the flow of data, it means that customers can bring their data to the compute resources as needed, whether those compute resources are local or cloud based.