

Hammerspace for Artificial Intelligence

DATASHEET

All industries are emerging into a new data cycle powered by the rise of massive data processing that is required to fuel innovation in all types of enterprises. Organizations are urgently assessing their data architectures to leverage AI to grow revenue streams and improve operational efficiencies. AI architects are looking for solutions that will help them design for flexibility for the future while providing the key building blocks for today.

One of the biggest challenges facing organizations is putting distributed unstructured data sets to work in their AI strategies while simultaneously delivering the performance and scale not found in traditional Enterprise solutions.

Hammerspace helps organizations by creating a single global data platform powered by its parallel file system coupled with automated data orchestration. It integrates Enterprise standard data services to ensure governance and data protection goals are met. It offers the flexibility to load data sets from any existing storage silo and the flexibility to add new data sets or data sources with the click of a button in the future. No matter where the AI model is located, local to the data or in a remote cloud or SaaS tool, Hammerspace makes the data easy to access, analyze, process, and move when needed.

Hammerspace powers training of some of the largest LLM in the world with 600 nodes, 60 Tbit/sec performance, 16,000 GPUs per site.

The Hammerspace Difference for AI

1. High-Performance

Most AI projects will become business critical in both decision making and cost containment. It is critical that the data pipeline is designed to use all available compute power and can make data available to the cloud models such as those found in Databricks and Snowflake.

Hammerspace high-performance data pipelines power some of the fastest compute farms in the world, some exceeding 60,000 GPUs in a single cluster. It has the performance optimizations to stream data at nearly line rate when loading data, it returns fast results when parameters are regularly checkpointed, and read requests are simultaneously delivered at the speed applications require.



2. Multiple Data Sources

AI models limited to using data from a single storage silo will be at a major disadvantage compared with those that can access a wide range of data sets stored across multiple storage platforms and likely multiple geographic locations. The AI data pipeline needs to be generated from data created and stored in edge devices, data centers, and the cloud.

To effectively access all this data, Hammerspace unifies siloed storage types as well as orchestrates data across multiple geographic locations. Doing so enables systems to place, present, and preserve data for access by the AI and ML models wherever the data is and wherever the models are run.

3. Burst to Remote GPU Clusters

GPU resources are critical to successful AI architectures. It is often desirable to take advantage of multiple GPU clusters to have the computational resources available when needed without the constant overhead of maintaining maximum capacity. This adaptability is especially crucial in a field like AI, where workloads and requirements can be highly variable.

4. Massive Scale

AI models must be trained with large quantities of data to be most accurate. The more data that is available to them, the more accurate the results will be from the outset.

Hammerspace integrates existing data sets, cloud instances, and any new infrastructure into a unified global data environment that will scale as AI workloads evolve.

5. Enterprise Standard

Enterprise standard data interfaces and data services play a critical role in efficacy, scalability, and reliability of AI systems while also aligning with organizational objectives for compliance, governance, and efficiency. Hammerspace helps organizations meet regulatory requirements by automating the processes and data services which oversee data access and protection to ensure data privacy, security, and appropriate usage is enforced.

To speed the integration of models with data, Hammerspace presents industry standard NFS as the interface to data makes it easy for applications to connect to data without being rewritten for object. And using NFSv4.2, performance can scale to the limits of any hyperscale environment.

6. Data Governance

Data governance plays a crucial role in AI and becomes even more complex when the data sources are distributed across many different silos and locations.



Hammerspace helps ensure data quality by unifying multiple data sources into a single namespace and metadata control plane. This helps reduce biases and ensures the accuracy, completeness, reliability, and timeliness of data to drive accuracy in model predictions and data being trustworthy for decision-making.

Parallel File System Performance and Scale, Fueled by ALL Data, Generated Anywhere

Hammerspace makes data a live, globally shared resource, that is no longer localized or trapped within specific storage systems or specific cloud data services.

In the AI Era, the volume of data, and the ways it can be used, will grow exponentially. High-performance local read/write access to data that is orchestrated globally in real time in a global data environment will become indispensable and ubiquitous.

With Hammerspace, organizations can readily leverage any data with any AI model, no matter where each is located.

