Cloudian Delivers Scalable AI Infrastructure with Reference Architecture based on Supermicro and NVIDIA

Object storage platform enables seamless performance and capacity growth from terabytes to exabytes for AI model training and inference workloads

SAN MATEO, CA, UNITED STATES, March 18, 2025 /EINPresswire.com/ -- Cloudian today announced a reference architecture for GenAI workloads that integrates Cloudian HyperStore object storage with Supermicro servers based on NVIDIA OVX and NVIDIA HGX. The solution has been shown to deliver throughput over 20 GB/s per node, with performance scaling linearly to meet the demanding I/O requirements of modern AI workloads including model training, inference, and checkpointing.

Object storage is rapidly becoming a cornerstone technology for AI due to its ability to efficiently manage the massive volumes of unstructured data that power modern machine learning and deep learning models. Cloudian delivers full S3 API compatibility, robust data protection, and scalable-performance enabled by integration with NVIDIA GPUDirectStorage, all of which empower organizations to unlock the full potential of AI innovation.

> This architecture combines Supermicro's GPU-optimized systems with Cloudian's enterprise object storage to create a complete solution for AI infrastructure."
>
> *Lawrence Lam, Supermicro*

The reference architecture showcases several important capabilities:

 High-performance flash, leveraging NVIDIA GPUDirect Storage and RDMA networking with linear throughput scaling
 Comprehensive enterprise feature set, including observability, self-healing and repair, multi-tenancy, QoS, dynamic storage provisioning, and data lifecycle management

🗹 Enterprise-grade security certifications with built-in ransomware protection and no custom OS kernel

"The AI storage landscape demands extreme performance, exabyte scalability, and enterprise-grade data management capabilities," said Amit Rawlani, Sr. Director of Alliances at Cloudian. "This solution revolutionizes AI infrastructure by delivering HPC-class performance with the scalability and cost-effectiveness of S3-compatible object storage, all while providing the management features that enterprises require."

"This architecture combines Supermicro's GPU-optimized systems with Cloudian's enterprise object storage to create a complete solution for AI infrastructure," said Lawrence Lam, vice president of Architecture AI and Storage Technology at Supermicro. "Organizations can deploy with confidence knowing the platform will scale seamlessly to meet their evolving needs."

The reference architecture delivers native S3 API compatibility, security, and scalability, supporting configurations from entry-level deployments to massive enterprise AI implementations. This flexibility enables organizations to begin with modest installations and seamlessly expand their infrastructure as AI initiatives grow, while maintaining consistent performance and unified management across the entire environment. The architecture's proven linear scaling ensures predictable growth without compromising system efficiency or administrative simplicity.

Download the reference architecture document here.

About Cloudian
Cloudian is the leader in secure, S3-compatible AI data lake platforms. With enterprise-grade security, limitless scale, and seamless cloud integration, Cloudian delivers plug-and-play interoperability with NVIDIA GPUDirect Storage and with AI tools such as PyTorch, TensorFlow, Kafka, Arrow and more. Our software-defined, on-premises solutions let users simplify and accelerate AI workflows, meet data sovereignty requirements, and cut costs by consolidating information to a single, cloud-like environment. Cloudian's geo-distributed architecture manages and protects object and file data at the edge, core, and in the cloud, for both traditional and modern applications.

Jon Toor
Cloudian
email us here

Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.