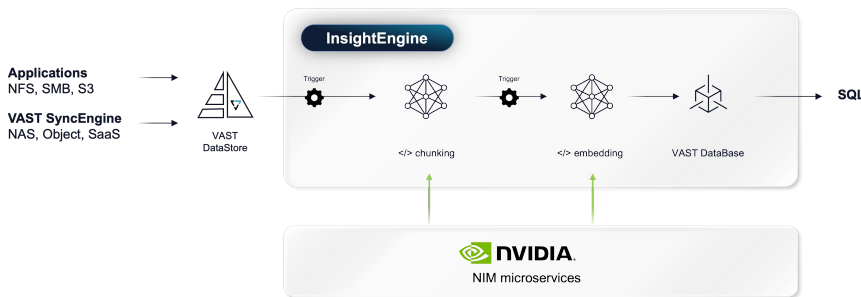


Supermicro Unified AI Data Platform



Building AI Systems at Scale

Enterprise AI stalls when compute, storage, databases, event pipelines, and data services are assembled as separate layers that require handoffs before models can act on business information. The Supermicro AI Data Platform with VAST AI OS turns that fragmented design into a system architecture, combining modular Supermicro building blocks with one software environment so organizations can stand up production AI faster, cut integration drag and keep governed data continuously available for new AI use at scale.



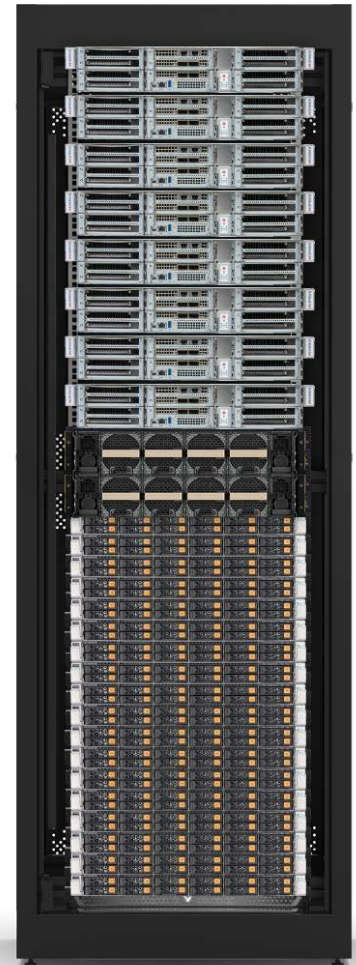
Within Supermicro Data Center Building Block Solutions, VAST becomes the software system that makes the hardware act as one platform. VAST DataBase, VAST InsightEngine, and DataEngine's event-driven architecture run with managed Kubernetes across it, while Supermicro CNode-X supplies GPU acceleration for vector retrieval, real-time SQL, and NVIDIA NIM microservices as a service. Supermicro EBox adds resilient all-flash infrastructure and new dense data services, giving enterprises a repeatable rack-scale design they can deploy and scale without rebuilding the AI stack.

Aligned to NVIDIA AI Data Platform reference design, the architecture keeps data in place and turns Supermicro racks into a governed execution environment for retrieval, analytics, inference, and agents. Integration points highlighted in the press release & blog include NVIDIA cuVS for vector search, Sirius for GPU-accelerated SQL, NIM services running inside the platform and support for BlueField and Spectrum-X architectures. The result is a unified system where Supermicro building blocks, VAST software, and NVIDIA acceleration operate together for enterprise AI factories.

The VAST + Supermicro Alliance

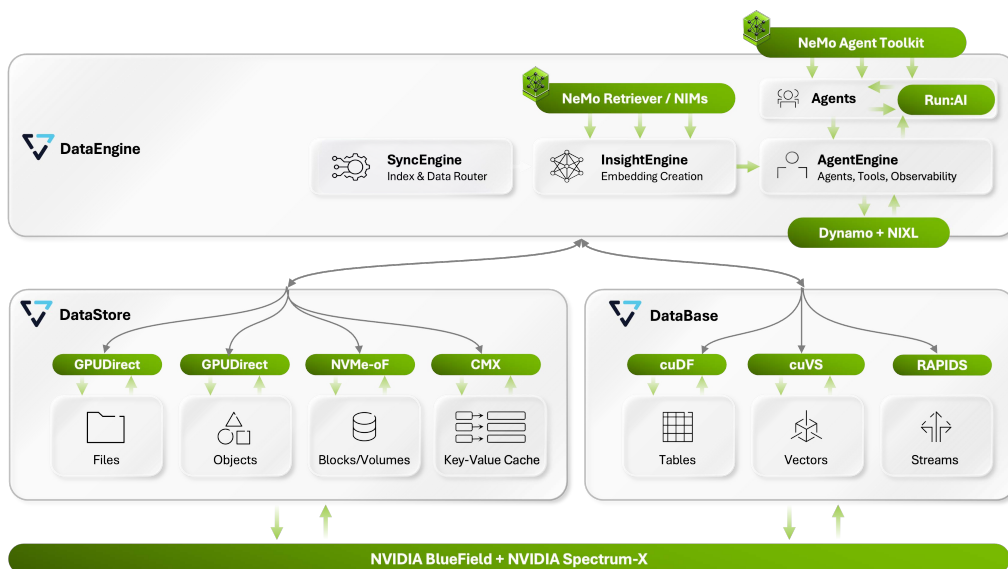
Supermicro and VAST pair infrastructure delivery with a software-defined AI data platform built for continuous use. Supermicro contributes Data Center Building Block Solutions, validated rack designs, liquid-cooled and air-cooled options, manufacturing depth, and the execution discipline to move from configuration to deployment quickly. VAST supplies the software layer that unifies data services, vectors, databases, event-driven pipelines, InsightEngine workflows, and AI services. Together, they give enterprises a repeatable way to deploy AI factories with less bespoke integration, faster scale-out, and a cleaner path to production outcomes.

Supermicro EBox | **Supermicro CNode-X**
A+ Server AS -1116CS-TN | SYS-212GB-FNR | NVIDIA RTX PRO 6000



The VAST AI OS

VAST AI OS gives the Supermicro AI Data Platform a shared software foundation for enterprise AI. Built on VAST's Disaggregated Shared-Everything architecture, it unifies resilient data services on EBox with CNode-X acceleration for VAST DataBase, VAST InsightEngine, and DataEngine. Managed Kubernetes orchestrates NVIDIA NIM microservices as a service close to the data, while NVIDIA cuVS accelerates vector retrieval and Sirius accelerates SQL execution. The platform also combines a native vector database, event-driven pipelines, real-time metadata and search, and high-performance file, object, table, and stream services in one operating environment. That means organizations can ingest, enrich, retrieve, reason over, and serve enterprise data without shuttling it across fragmented systems, giving developers and operators one platform for retrieval, analytics, inference, and agentic workflows at scale. It also creates a single control point for data movement, indexing, governance, and service delivery, so teams can scale AI factories without multiplying tools, copies, or operational handoffs.



Example Use-Cases

Enterprise RAG

Use InsightEngine and NVIDIA NIM microservices to chunk, embed, index, govern, and retrieve enterprise content in place at scale.

Real-Time Analytics

Use DataEngine events and GPU-accelerated SQL to process logs, streams, and operational data in real time as new data arrives daily.

Multimodal AI

Search and analyze documents, images, logs, and enterprise content with unified vector, metadata, and data services in real time.

AI Factory

Support AI agents with live context, vector retrieval, event triggers, and shared NIM-powered services running across the platform.

Conclusion

Supermicro and VAST deliver a unified AI data platform that turns validated building blocks into a production system for enterprise AI. Supermicro brings rack-scale design choice, manufacturing depth, integration discipline, and deployment speed through Data Center Building Block Solutions. VAST brings durable data services, a vector database, InsightEngine workflows, DataEngine automation, and NIM-powered AI services. Together, with NVIDIA integrations including cuDF, cuVS, Sirius, BlueField, and Spectrum-X Ethernet support, they help enterprises reduce integration risk, activate governed data faster, and scale AI from pilots to live services.

Turn data into faster AI outcomes with a unified platform built to scale.

Key Benefits



Unified AI Data Platform

Turns Supermicro building blocks into shared data, database, vector, eventing, and AI services for enterprise AI use cases.



In-Place Acceleration

Uses cuVS, Sirius, and NIM-as-a-Service on CNode-X to accelerate retrieval, SQL, and inference directly where data resides



Enterprise Durability

EBox delivers unified all-flash data services with high availability, nondisruptive operations, plus governance and scale.



Scale With Less Sprawl

Scales across Supermicro rack designs with managed Kubernetes and shared platform services instead of fragmented AI silos.

About VAST Data

VAST is the AI Operating System company, delivering a unified software platform for storage, database, event-driven pipelines, vector search, and AI services at scale. By simplifying how data is stored, indexed, governed, and activated, VAST helps organizations move beyond data management toward data-driven intelligence.