

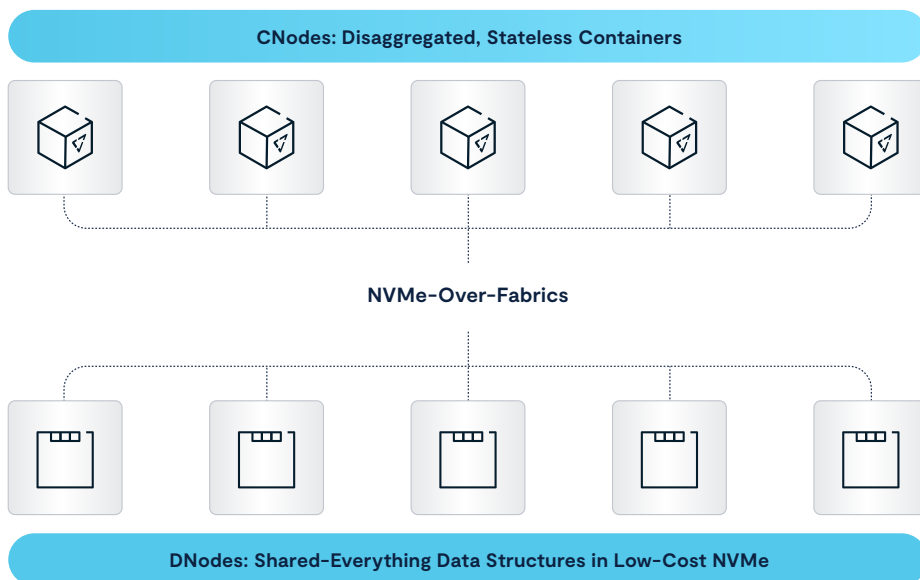
VAST DASE Architecture with Cisco Nexus 9000 Series



The Infrastructure Reality Check

Here's what nobody wants to say: legacy infrastructure wasn't built for AI. It was built for incremental improvements to old problems. VAST's Disaggregated And Shared Everything (DASE) architecture doesn't iterate on the past. It reimagines the entire relationship between compute and data. Any C-Node accesses any data across any D-Node. No pinning. No locality constraints. Just radical parallelism at exabyte scale.

Cisco Nexus 9000 provides the lossless fabric DASE requires to make this real. The VAST AI Operating System turns this into something practical: LLM training, RAG pipelines, inference serving, all running on a converged platform. Nexus 9000 interconnects C-Nodes and D-Nodes via NVMe-oF as the foundational layer of Cisco Secure AI Factory with NVIDIA and Cisco AI PODs.



Accelerate AI initiatives with infrastructure that keeps GPUs productive, scales seamlessly as demands grow, and reduces operational complexity.



Performance

Delivers lossless, low-latency NVMe-oF so GPUs stay busy, not waiting.



Scalability

NVIDIA ERA-compliant fabric scales from small clusters to tens of thousands of GPUs.



Simplicity

Unifies networking and storage access, simplifying operations across AI pipelines.

Why Legacy Thinking Fails

The old model multiplies complexity: separate fabrics for compute and storage, data movement that idles GPUs, packet loss that cascades into RDMA timeouts. You end up with expensive GPUs doing the one thing they shouldn't: waiting.

Cisco Nexus 9000 collapses this. Converged frontend and storage networks handle incast patterns and mixed traffic without separate infrastructure. Proactive congestion prevention uses flowlet switching, per-packet spraying, and ECMP simultaneously, each method selected for the traffic type it serves best. Policy-driven Dynamic Load Balancing increases path utilization and delivers predictable throughput. This isn't evolutionary. It's necessary.

Why Cisco Nexus 9000 Powers VAST DASE Architecture



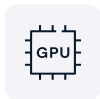
Lossless RDMA

Non-Negotiable for DASE: Zero packet loss isn't optional. VAST DASE depends on it. Nexus 9000 delivers Data Center Bridging with Priority Flow Control, Explicit Congestion Notification, and deep buffers for lossless RoCEv2. At full utilization of high-speed RDMA links, the fabric eliminates retransmissions that stall GPUs and waste cycles you already paid for.



Performance Without Compromise

Ultra-low latency NVMe-oF between C-Nodes and D-Nodes. Cisco Intelligent Packet Flow with Dynamic Load Balancing and packet spraying outperforms standard ECMP in AI collectives. Independent MLPerf validation proves shorter job completion times with LLM training on NVIDIA GPUs.



NVIDIA Integration That Actually Works

NVIDIA GPUDirect Storage enables direct GPU-to-VAST transfers. No CPU bottleneck. Tuned for NVIDIA NCCL collective operations across AI Enterprise and NIM. This isn't aspirational. It's tested and validated.



Scale That Matches Reality

Silicon One and Cloudscale ASICs support DASE from starter clusters to tens of thousands of GPUs. Hardware rerouting converges rapidly during failures while maintaining lossless fabric integrity. Combined with VAST's data resilience, there's no single point of failure.



Open Ethernet Without Lock-In

Comprehensive Ultra Ethernet Consortium 1.0 support for full fabric deployments with packet trimming, credit-based flow control, and zero-copy NVMe-oF. Open standards enable cost-effective scaling with existing data center investments and future-proof compatibility as the ecosystem evolves.



Data Efficiency by Design

VAST DASE combines global namespace, intelligent reduction, and efficient NVMe-oF transport. More useful data reaches GPUs with fewer copies and less cross-cluster noise. Effective throughput improves because the architecture eliminates waste.



Programmability and Visibility

P4 runtime programmability on Silicon One means custom AI pipelines without hardware respins. Software upgrades protect investment as speeds evolve. Real-time telemetry with In-Band Network Telemetry and microburst detection optimizes C-Node to D-Node flows through Nexus Dashboard, cutting job rescheduling.

What This Means



GPUs stay productive with no idle time waiting for data



Scale to exabytes with consistent microsecond response



CPU overhead drops via RDMA offload



Universal data access through disaggregated DASE and NVMe-oF

About Cisco

Cisco (NASDAQ: CSCO) is the worldwide leader in IT that helps companies seize the opportunities of tomorrow by proving that amazing things can happen when you connect the previously unconnected. For more information, please visit www.cisco.com

About VAST Data

VAST Data is the data platform company built for the AI era. As the new standard for enterprise AI infrastructure, organizations trust the VAST Data Platform to serve their most data-intensive computing needs. VAST Data empowers enterprises to unlock their data's full potential by providing simple, scalable, and architected AI infrastructure to power deep learning and GPU-accelerated data centers and clouds. Launched in 2019, VAST Data is the fastest-growing data infrastructure company in history. For more information, please visit www.vastdata.com

Cisco Nexus Switch Portfolio for AI Infrastructure

Switch Model	ASIC	Bandwidth	Buffer	Ports	Breakout Configurations
Nexus 9364E-SG2-Q	Silicon One G200	51.2 Tbps	256MB	64x 800GbE QSFP-DD	8x100G, 2x400G
Nexus 9364E-SG2-O	Silicon One G200	51.2 Tbps	256MB	64x 800GbE OSFP	
Nexus 9364D-GX2A	Cloudscale	25.6 Tbps	120MB	64x 400GbE QSFP-DD	4x10G, 4x25G, 4x50G, 4x100G, 2x200G
Nexus 9332D-GX2B	Cloudscale	12.8 Tbps	120MB	32x 400GbE QSFP-DD	

Ready to build beyond?

Contact Cisco to discover how Nexus 9000 brings compute directly to VAST DASE data for scalable AI acceleration.

Contact us

vastdata.com/demo
hello@vastdata.com

