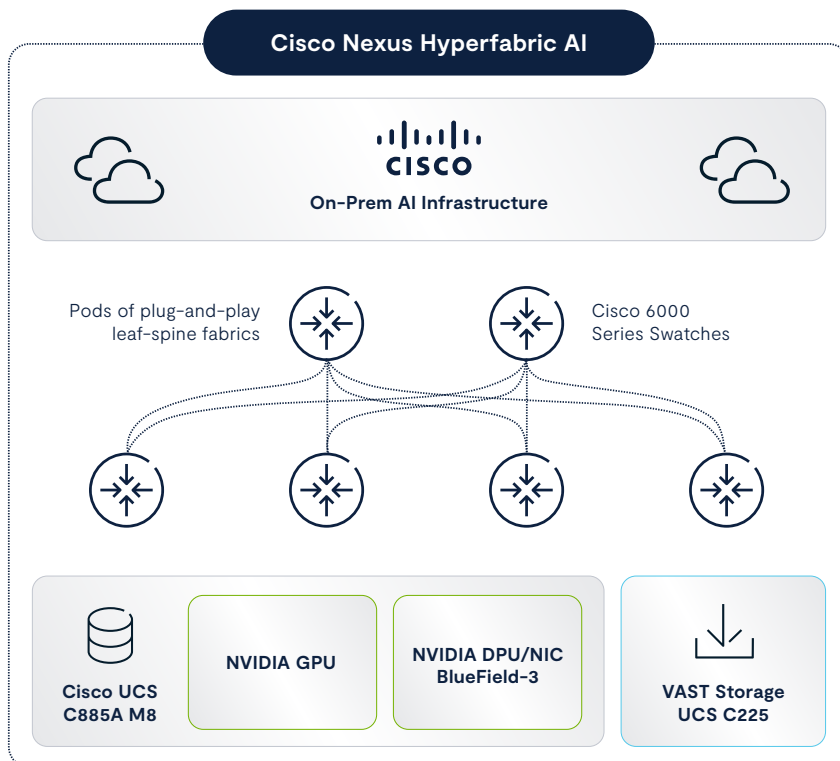


Scaling AI Workloads with VAST Data and Cisco



Enterprise AI Adoption Creates Opportunity

AI is fundamentally transforming the demands placed on IT infrastructure, necessitating a shift from traditional, complex systems to more agile, scalable, and efficient solutions. Addressing challenges such as the complex implementation and management of diverse AI components, significant upfront costs, infrastructure limitations, and a lack of technical expertise requires a new approach. The difficulty in scaling infrastructure to meet the burgeoning needs of AI is often compounded by outdated data center capabilities regarding power and cooling, highlighting the urgent need for modernization. Furthermore, inefficient data movement and processing within AI workloads can create significant bottlenecks, impeding the ability to unlock timely insights from vast datasets.



Why VAST + Cisco?

AI workloads can demand complexity from any infrastructure environment, but Cisco and VAST deliver **simplicity, flexibility, and scalability** in any situation, enabling organizations to flourish in a data-driven, AI-focused world.

Simplicity

Cisco Nexus Hyperfabric AI, in partnership with VAST Data, allows effortless cloud-managed AI infrastructure to deliver fast time to value.

Flexibility

Meeting the needs of data-hungry algorithms and services is a critical job that can be easily managed using Cisco Nexus Hyperfabric for AI's SaaS consumption models integrated with VAST Data from day one.

Scalability

Cisco and VAST Data offer a scalable solution that can scale at multiple layers (GPU, Storage, Compute), ensuring management, visibility, and operational efficiency coupled to performance-optimized compute, storage, and networking.

Generative and agentic AI, in particular, demand unprecedented scale and performance for rapid data access, compelling enterprise IT buyers to make substantial investments in Ethernet networking technologies. According to IDC¹, the Ethernet switching market has experienced significant growth due to the increasing number of enterprise AI deployments, with the overall market expanding by 20.1% in 2023 to reach \$44.2 billion. By leveraging the capabilities of NVIDIA and VAST Data, Cisco integrates compute resources, storage, and data management with Ethernet-powered AI network fabrics. This integration provides innovative connectivity solutions characterized by unparalleled programmability, performance, flexibility, and efficiency, effectively addressing the substantial requirements of contemporary AI/ML platforms.

First Class Citizens

VAST and Cisco are helping enterprises build AI data centers with NVIDIA GPU servers and Cisco networking from design to implementation to management. They also provide assertion-based end-to-end visibility into compute, networking, storage, and data management, allowing organizations to seamlessly build and scale their AI operations through their Cisco Nexus Hyperfabric for AI solution.

The strategic alliance between Cisco and VAST Data directly addresses these critical infrastructure challenges by simplifying the deployment and management of AI. By integrating Cisco's networking and compute expertise and NVIDIA's AI solutions, including their GPUs and Bluefield 3 DPUs, with VAST Data's innovative storage solutions, the partnership delivers a unified, AI-ready infrastructure offering, such as Cisco Nexus Hyperfabric for AI. This collaboration aims to simplify complex implementation processes and management of AI infrastructure components and the ongoing lifecycle of the cluster. This joint approach is designed to provide resilient, scalable, and high-performance infrastructure solutions that can better handle the demands of AI workloads, ensuring seamless integration from compute to storage and making AI adoption more accessible for enterprises looking to derive value from their data.

Transformative Solutions for the Modern AI Era

The increasing use of Generative AI leads to a growing demand for GPUs and storage capacity. Modern AI applications require large clusters, often consisting of thousands of GPUs and storage devices, with the potential to scale to tens of thousands as demands increase. With GPU speeds doubling every other year, it's essential to prevent compute and network bottlenecks through scalable network design. As application teams enhance compute capacity, network teams must carefully evaluate the interconnect to ensure optimal performance. Networks that balance performance with simplicity, observability, and scalability are positioned to lead the AI revolution.

Similarly, AI's demand for storage to provide multi-protocol, scalable, and performance-oriented access to structured and unstructured data calls for reimagining storage solutions. Creating an independently scalable, secure, and multi-tenant solution utilizing the performance and ubiquity of Ethernet shows that current and future challenges can be easily addressed using proven technologies.

Cisco and Vast Data represent the ideal combination of simplicity, observability, and scalability through proven technologies like Ethernet and NVMe over Fabrics. By utilizing their combined resources, customers can be confident that they are ready to face any challenge AI presents.

¹ <https://www.idc.com/getdoc.jsp?containerId=prUS51948824>

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



Scan here to learn more.