

FlexPod with Cisco AI PODs: Infrastructure for AI Training and Fine-Tuning

Solution overview

Is your organization struggling to balance the rapid innovations required for AI with the need for a stable, secure, and manageable data-center environment? Building AI infrastructure from scratch is complex, risky, and slow. FlexPod AI with Cisco AI PODs changes that by providing a modular, pre-validated infrastructure solution specifically optimized for AI training, fine-tuning, and high-throughput inferencing. By combining the converged architecture of FlexPod Datacenter with advanced GPU-accelerated compute, you gain a high-performance platform that integrates seamlessly into your existing IT environment. Whether you are working with Large Language Models (LLMs) or real-time analytics, this solution provides the scalability and reliability necessary to turn your AI goals into business results.

What it does

FlexPod with Cisco AI PODs delivers a full-stack infrastructure tailored for the unique demands of modern AI workloads. Unlike fragmented DIY approaches, our solution offers:

- **Validated performance:** Every component—from Cisco UCS® servers and Cisco Nexus switches to NetApp storage—is tested and documented to ensure predictable performance
- **Integrated management:** Centralize control of your compute and network fabrics through Cisco Intersight and Nexus Dashboard, reducing the burden on your IT staff
- **Flexible scalability:** Start with the right-sized configuration and scale your GPU clusters as your AI adoption grows, avoiding costly upfront over-provisioning
- **Advanced AI software stack:** Seamlessly run NVIDIA AI Enterprise and Red Hat OpenShift on a foundation designed for high-speed networking and persistent storage

Benefits

- **Accelerate AI lifecycle:** Dramatically shorten the time from infrastructure deployment to AI model training and fine-tuning
- **Simplify management:** Unify your AI infrastructure management through Cisco Intersight® and Cisco Nexus® Dashboard for better operational visibility
- **Scale with confidence:** Leverage a modular, pre-validated architecture that grows with your business needs, from pilot projects to large-scale production
- **Reduce operational complexity:** Streamline deployment with validated designs and automation playbooks, minimizing risk and maximizing performance

“FlexPod with Cisco AI POD provides a complete, pre-validated solution to meet the full spectrum of AI infrastructure requirements, allowing enterprises to operationalize AI initiatives with confidence and reduced time-to-value.”

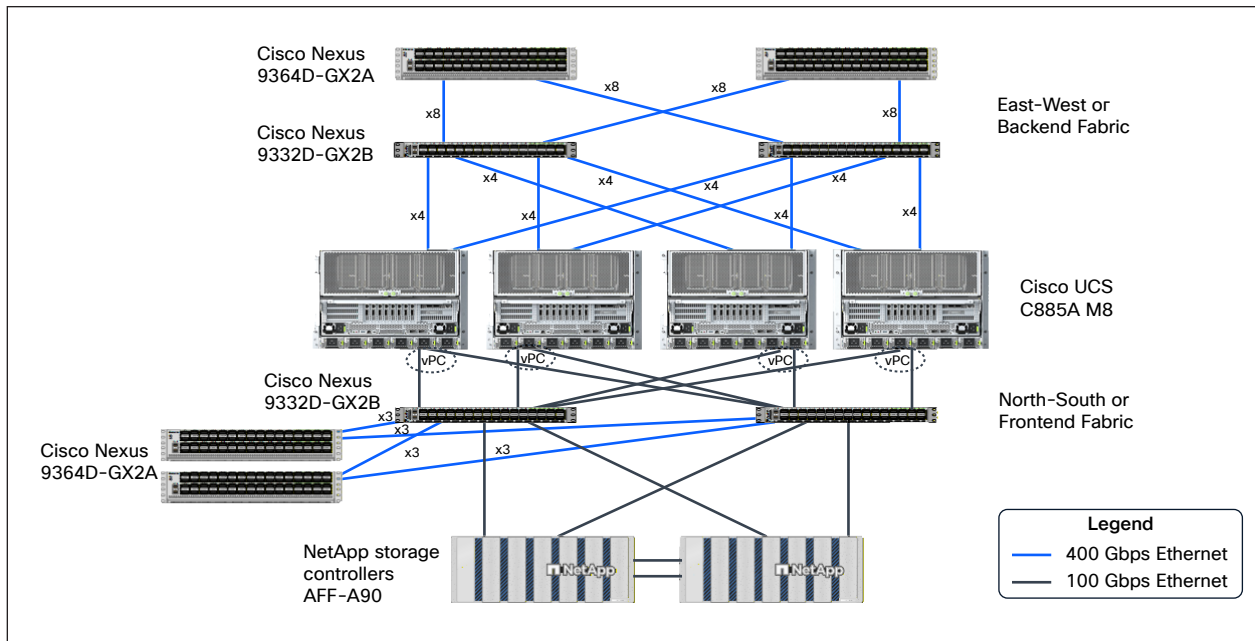


Figure 1. Simplified architecture diagram of FlexPod AI



Learn more

Start scaling your AI initiatives with confidence. Access the full technical specifications and step-by-step deployment guidance in our latest Cisco Validated Design (CVD): [FlexPod with Cisco AI PODs: Infrastructure for AI Training and Fine-Tuning Deployment Guide](#). Get the assurance you need for your enterprise AI deployment by visiting: [Cisco Validated](#).