FlexPod®

Secure AI Factory with FlexPod AI

Deploy secure, enterprise AI workloads with confidence using the validated infrastructure from NetApp, Cisco, and NVIDIA.

In recent years, generative AI has seen exponential growth with models such as large language models (LLMs) and foundation models leading the charge. Many organizations look to utilize capabilities such as natural language understanding, image generation, and code completion to improve productivity, reduce operating costs, and accelerate innovation and time-to-market. However, operationalizing and deploying AI models at enterprise scale remains challenging due to unique infrastructure and security constraints that many organizations struggle to address, including:

- **Massive compute demands** that require specialized GPUs such as NVIDIA RTX 6000 Pro GPUs, to support increased parallel processing capabilities.
- **Potential for new** AI security risks that require dedicated strategies to prevent risks such as data breaches and model poisoning.
- **Simplified, scalable data infrastructure** for compute, network, and storage that scale with models without bottlenecks.
- **Avoiding vendor lock-in** that can confine enterprises to inflexible terms, affecting agility and cost efficiency.

The need for security in AI

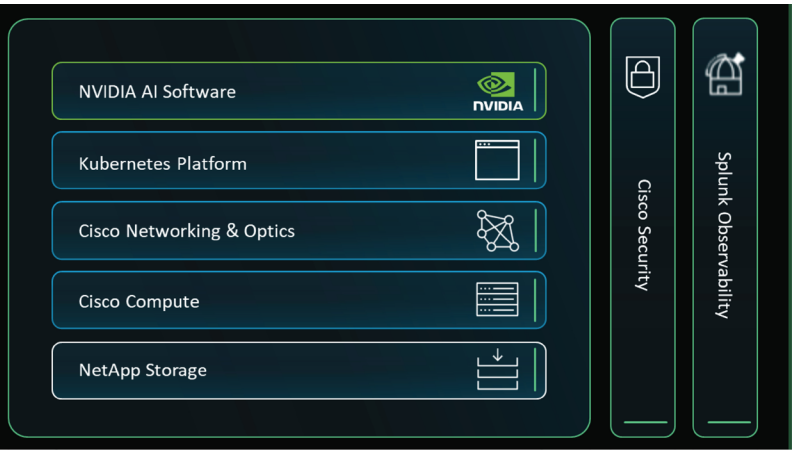
AI systems now sit at the heart of mission-critical operations across all industries including finance, healthcare, and government. Unlike traditional IT systems, AI introduces new attack surfaces across data pipelines, models, and inference endpoints, which require dedicated security strategies. Their ability to process sensitive data, make real-time data-driven decisions, and influence large-scale outcomes makes them highly attractive targets for cyber threats.

What is it?

Reference architecture

Validated solutions and turnkey offerings

Differentiated with Security and Observability



Key cyber risks AI presents include:

- **Data breaches and ransomware attacks** that can be financially damaging and lead to loss of competitive advantage.
- **Model poisoning,** which can alter model behavior, introduce bias, or enable system sabotage.
- **Prompt injection and adversarial attacks,** which can produce harmful outputs, as well as misclassify or misinterpret critical information.
- **Model theft and reverse engineering,** which can introduce vulnerabilities into deployed AI systems.

Point solutions and DIY alternatives are not sufficient to mitigate these ever-evolving risks and leave enterprises vulnerable to cyber threats. To reduce risk, speed up time to deployment, and ensure end-to-end protection, security must be built into the foundation of enterprise operations.

Introduction to FlexPod AI with NetApp® AFX, NetApp® AI Data Engine (AIDE), and Secure AI Factory

Secure AI Factory takes advantage of the combined strengths of three technology leaders:

- **Cisco** delivers enterprise compute through Cisco UCS platforms including C240 servers, and ultralow latency, high-bandwidth data flows through Cisco Nexus networking.
- **NVIDIA** provides the AI acceleration foundation with NVIDIA RTX 6000 Pro GPUs and AI Enterprise software for accelerated model training, fine-tuning, and inference.
- **NetApp** delivers AI-optimized, disaggregated storage and data management with high-performance NetApp® AFX systems and NetApp® AIDE.

With this collaboration, every layer—from the GPU compute nodes to the network fabric, and the storage and data pipelines—is validated and secure to accelerate deployment, reduce risk, and unlock actionable insights faster.

FlexPod AI introduction

With security as a foundational element, FlexPod AI with NetApp® AFX, AIDE, and Secure AI Factory tightly integrates controls across compute, networking, and storage within an enterprise AI infrastructure. It provides a standardized, validated, and secure architecture designed to accelerate AI deployment while reducing risk.

At the core of Secure AI Factory lies FlexPod® AI, a validated converged infrastructure solution that serves as the backbone for building and securely scaling enterprise-grade AI capabilities in an organization's data center. Co-developed by Cisco and

Streamline AI Deployment

Simple

Start small and expand with a pre- configured, pre-validated, turnkey solution that is fast, easy to deploy, and cost effective.

Secure Enterprise Data

Secure

Retain peace of mind knowing your data and investments are both protected.

Scalable Solution for Growth

Scalable

Deploy a high-performance solution with the flexibility to scale alongside the needs of your business.

NetApp, FlexPod AI is purpose-built to meet the unique challenges of modern AI workloads with the high performance, reliability, and flexibility needed to unlock the full potential of AI. With its robust architecture, FlexPod AI seamlessly integrates with the Secure AI Factory framework, enabling organizations to accelerate AI adoption while maintaining strict security, governance, and operational consistency.

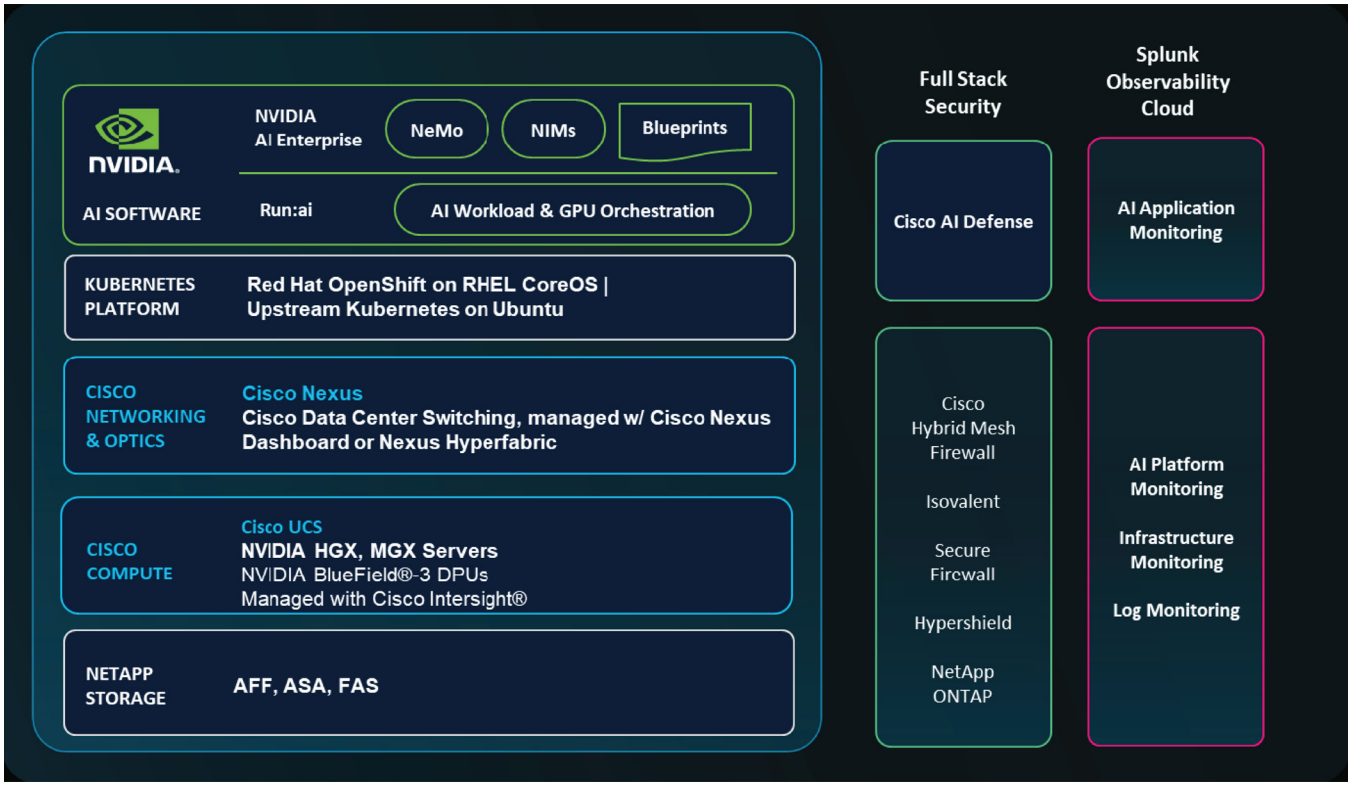
With NVIDIA GPUs and NVIDIA AI Enterprise software, FlexPod AI becomes a turnkey solution that not only accelerates deployment but also delivers the scalability and efficiency needed to power AI innovation at any scale. By anchoring Secure AI Factory on FlexPod AI, organizations gain the solid foundation needed to power AI workloads from data preparation and model training to inference and deployment, all within a secure, enterprise-ready environment.

FlexPod AI and Secure AI Factory with NVIDIA

Designed to address the unique security challenges of AI workloads, FlexPod AI with Secure AI Factory provides the secure foundation needed to accelerate AI deployments across environments—from data and workloads to infrastructure.

FlexPod AI with Secure AI Factory enables organizations to accelerate AI adoption through a robust, secure, and performance-optimized infrastructure. This comprehensive solution overcomes the unique challenges of AI deployment at scale, enabling enterprises to innovate rapidly while maintaining rigorous security and compliance standards.

To learn more on accelerating your AI journey, explore FlexPod resources or contact your NetApp or Cisco representative today to evaluate your AI-readiness.



©2026 NetApp, Inc. All rights reserved. No portions of this document may be reproduced without prior written consent of NetApp, Inc. Specifications are subject to change without notice. NETAPP, the NETAPP logo, and the marks listed at http://www.netapp.com/TM are trademarks of NetApp, Inc. Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. DS-4507-0526