

AI Networking for Inference at Scale

VP, Product Management, Data Center Networking



The background features a dark blue gradient with a bright blue beam of light entering from the left. This beam illuminates a complex, fractal-like structure of glowing particles in shades of blue, purple, and orange, resembling a nebula or a complex data visualization. The particles are dense and interconnected, with some areas appearing more brightly lit than others.

We're now in the age of inferencing.



Scalable

Secure

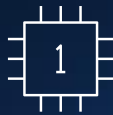
Simple

Cisco AI Networking

Deliver Resilient, High-Performance Infrastructure

Fuse Security into Every Layer

Unify Operations & Experiences



Silicon



Systems



Software



Optics



Operating Model

Cisco Nexus One

FULLY INTEGRATED STACK

Nexus Dashboard

On premises

Nexus Hyperfabric

Cloud management

SILICON

Cisco Silicon One
NVIDIA Spectrum-X Ethernet

SYSTEMS

Cisco N9000
Systems

OPTICS

Cisco Optics

SOFTWARE

Cisco NX-OS
Cisco ACI, SONiC

SECURITY AND OBSERVABILITY

Introducing

Cisco Silicon One G300

102.4 Tbps

AI Network Switch

200 Gbps

Integrated Serdes

1.6T Ethernet

Port Switch

512 ports

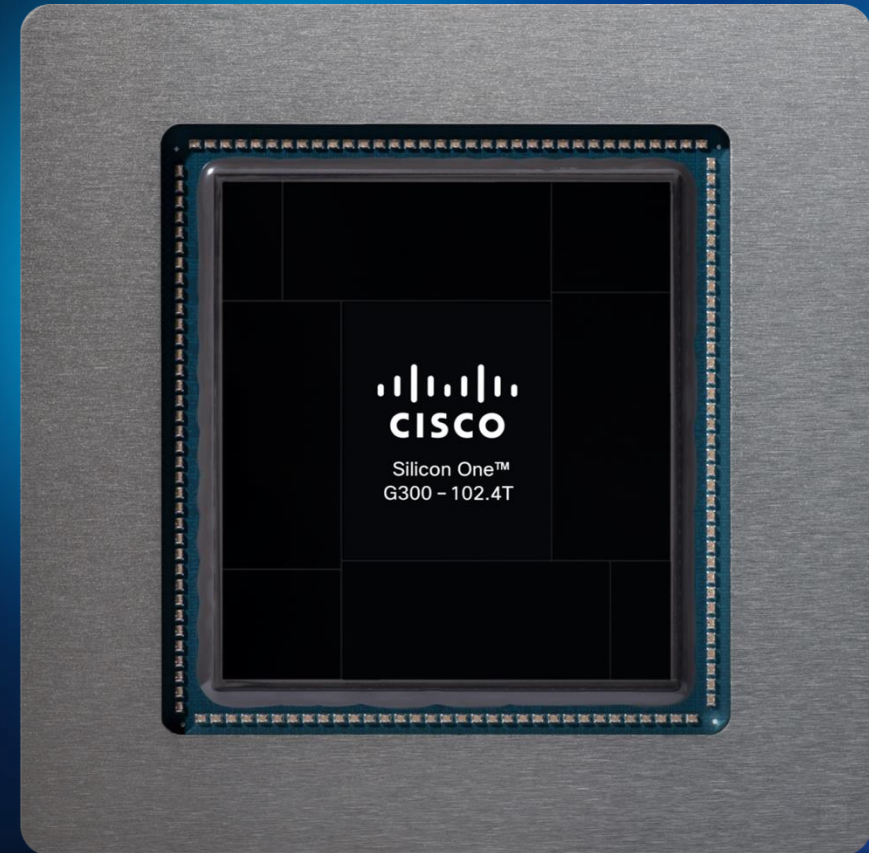
High Radix Scaling

256MB

Fully-Shared
Packet Memory

Optimizing TCO

for Scale AI
Deployments

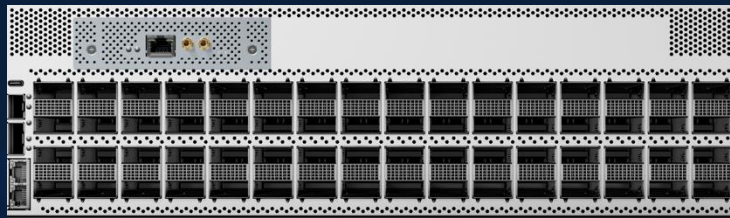


Introducing H2'26

Cisco Advanced 102.4T Systems Portfolio

Seamless deployment flexibility for optimized AI performance

Cisco N9364F-SG3 Cisco 8133
Air Cooled



Features

3RU, 64p 1.6T OSFP
EIA 19"

Redundant Pluggable Power

102.4Tbps

Cisco N9364F-SG3 Cisco 8132
Fully Liquid Cooled



Features

20U, 64p 1.6T OSFP
ORv3N 21" blind mate

Redundant Integrated Power
Advanced Leak Detection

102.4Tbps

Hyperscale | Neocloud | Service Providers | Sovereign Cloud | Enterprise

Cisco expands Silicon One P200 Across the Portfolio

With New Cisco N9000 and Expanded Cisco 8000 Systems

Cisco N9364E-SP2R and Cisco 8223 51.2T Fixed Systems



51.2T, 3RU,
64p 800G Fixed
OSFP800 and
QSFP-DD800




NX-OS


ACI


IOS-XR

Cisco N9836E-SP2R and Cisco 88-LC2 28.8T Modular Line Cards



28.8T, 36p 800G
OSFP800



Cisco 800G ZR/ZR+ optics



1000km+
reach

Industry Leading
Performance and Scalability

Smart Adaptive Processing

Quantum Safe
Line-Rate Encryption

Scale-Across | DCI | Universal Spine | Core & Peer Routing

Reliable connectivity for AI data centres

Cisco OSFP 1.6T Pluggable Optics

High-performance and reliability

Cisco Silicon Photonics technology

Advanced DSP-based diagnostics and monitoring



High bandwidth AI scale-out

1.6T switch-to-NIC connectivity and high-density fabrics for massive AI cluster builds (2x800G, 4x400G, and 8x200G links)

Diversified supply chain

Seamless AI cluster scaling with diversified supply chain and interoperability across Cisco and 3rd party DSPs

Efficiency and flexibility

Optics for power-efficiency and deployment flexibility in liquid cooled and air-cooled platforms

Cisco 800G Linear Pluggable Optics (LPO) Solutions

Powered by **Cisco Silicon Photonics**

AI Scale-out power efficiency and reliability for 800G Networks



Cisco L800G OSFP LPO

50% less power per module

30% less power for integrated systems

Shift power consuming optics DSP functionality to Cisco Silicon One platform

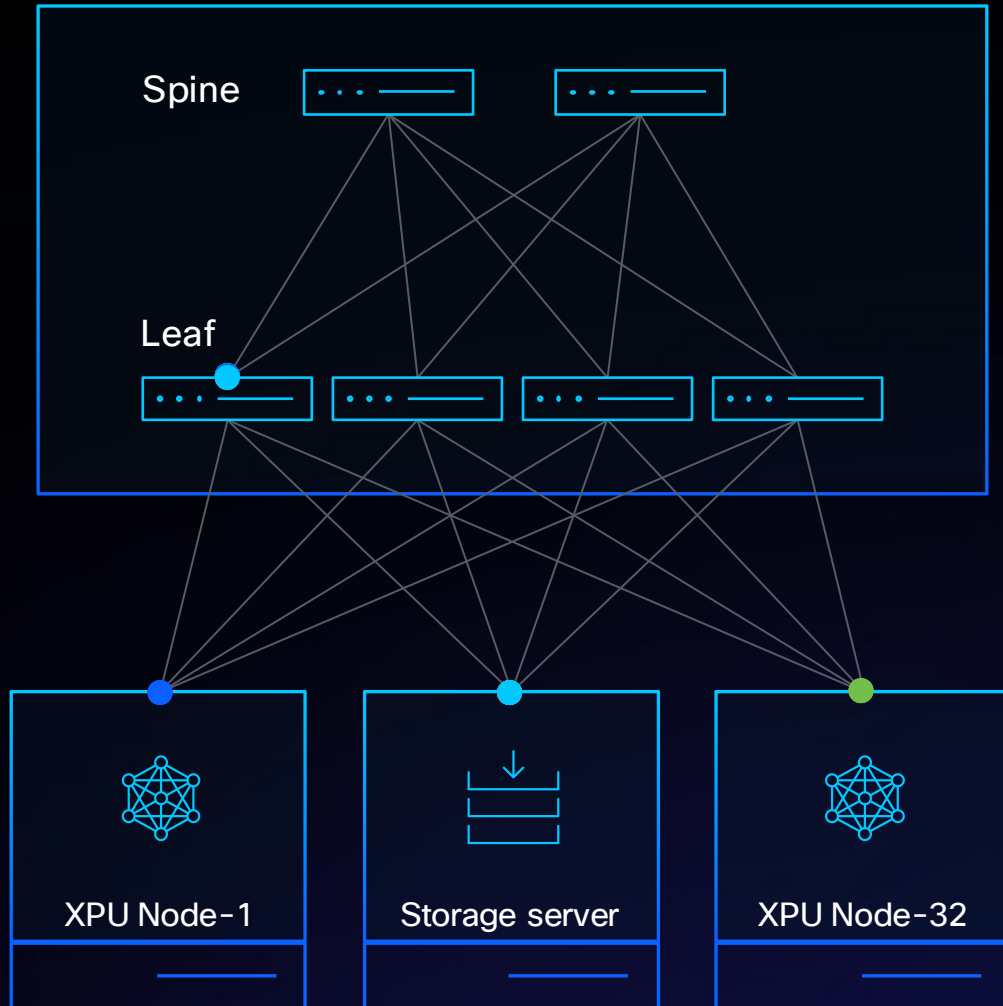
High signal integrity driven by Cisco Silicon Photonics and Cisco Silicon One

High reliability with lower thermal stress

Increase port density without PDU upgrades

Intelligent Packet Flow for Nexus One

Advanced Load-Balancing Capabilities for AI Workloads



Cisco is the only vendor who can do per flowlet, per-packet, a mix of both, or dynamically switch between.

38%

NCCCL performance gains

Meeting you where you are
with choice, flexibility, and partnership.

NVIDIA

AMD

VAST

NetApp

intel

GitHub

Red Hat

NUTANIX

vmware
by Broadcom

CLOUDERA

groq

GitLab

HashiCorp

PURESTORAGE

HITACHI

servicenow

accenture

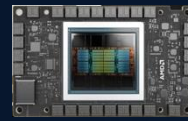
IBM

Cisco and AMD Partnership

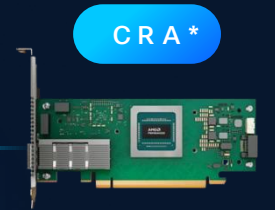
Cisco Nexus switches are included in AMD reference architectures for AI cluster design

Smart Switch enables networking and security services in a compact all-in-one solution

Cisco Nexus 9364E-SG2



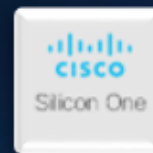
AMD Instinct MI300 accelerators



AMD Pensando Pollara 400 AI NIC

CRA*

Cisco Nexus Smart Switch



Cisco Silicon One E100



AMD Pensando DPUs

UNIFIED WITH NEXUS ONE

*CRA: Cisco Cloud Reference Architecture

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco and NVIDIA Partnership

NVIDIA Enterprise Reference Architecture (ERA) for <1K GPUs

Based on NVIDIA Cloud Partner (NCP) for Cisco's Cloud Reference Architecture (CRA) for >1K GPUs

*CRA: Cisco Cloud Reference Architecture

*NCP RA: NVIDIA Cloud Partner Reference Architecture Compliant

*ERA: NVIDIA Enterprise Reference Architecture (ERA)

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco N9300 Series Switches



NVIDIA BlueField



Powered by Cisco Silicon with NVIDIA Spectrum-X integration



NVIDIA BlueField

ERA*

CRA*

NEW

G300

Cisco N9100 Series Switches



NVIDIA BlueField



Powered by NVIDIA Spectrum-X Ethernet switch Silicon



NVIDIA BlueField

NCP RA*

NEW

Spectrum-6

UNIFIED WITH NEXUS ONE

Cisco Reference Architectures

ENTERPRISES

Cisco Enterprise Reference Architecture (ERA)

NVIDIA ERA Compliant

Cisco AI Servers with < 1K GPUs

Cisco Silicon One

Cisco N9300 Series Switches (51.2T, 102.4T) Spectrum-X License

NEOCLOUDS
SOVEREIGN CLOUDS

Cisco Cloud Reference Architecture (CRA)

Cisco CRA

Cisco AI Servers from ~1K to 32K GPUs

Cisco Silicon One

Cisco N9300 Series Switches (51.2T, 102.4T) Spectrum-X License

NVIDIA NCP RA Compliant

NVIDIA Spectrum-X Ethernet Silicon

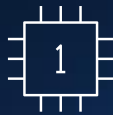
Cisco N9100 Series Switches (51.2T, 102.4T)

Cisco AI Networking

Deliver Resilient, High-Performance Infrastructure

Fuse Security into Every Layer

Unify Operations & Experiences



Silicon



Systems



Software



Optics



Operating Model

83%

of network engineers have
security as part of their remit

ZK Research, 2026

OUR STRATEGY

Turn every switch port into an enforcement point

Cisco N9300 Smart Switch



Nexus Networking

VLAN / VRF / VXLAN
Network context

Hypershield Security

800Gbps of stateful performance
Instant protection for new workloads

OUR STRATEGY

Detection only matters if enforcement happens in real-time

Live Protect

Vulnerability shielding for Cisco AI networking devices



66%

of organizations hosting generative AI models use Kubernetes for some or all inference workloads.

-Cloud Native Computing Foundation, 2026

APPS

K8s NODES



FABRIC



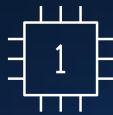
TRADITIONAL NETWORK

Cisco AI Networking

Deliver Resilient, High-Performance Infrastructure

Fuse Security into Every Layer

Unify Operations & Experiences



Silicon



Systems



Software



Optics



Operating Model

Native Splunk

Nexus One

splunk> DATA LAKE



Nexus Dashboard admin

Analysis hub Refresh

Analyze and troubleshoot your network with advanced analytics tools optimized for you to gain valuable insights into the performance and health of your network. Access to different tools and analytics is based on your license level. [View License Mode Details](#)

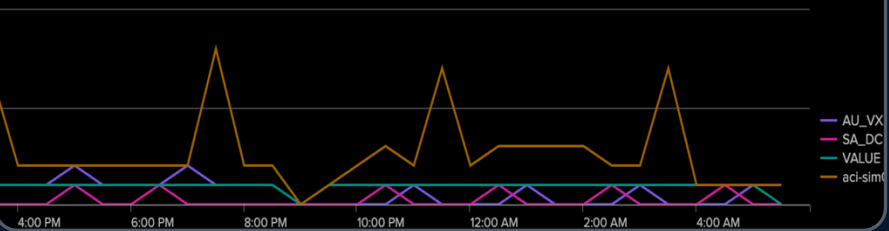
- Policy CAM**: Monitor your networks policies
- Compliance** ACI only: Monitor your fabrics compliance with custom anomaly rules
- Conformance**: Keep track of your hardware and software life cycles
- Connectivity**: Analyze flows from one endpoint to another
- Traffic analytics**: Monitor your networks latency congestion and drops
- Energy management**: Explore your fabric's energy usage, cost, and emissions
- Delta**: Compare configurations and differences in your fabric(s) between two points in time
- Pre-change** ACI only: View the potential impact of configuration changes
- Log collector**: Collect and analyze logs from your devices
- Bug Scan**: Learn about active and potential bugs affecting your networks
- Endpoint locator** NX-OS Only: Real-time tracking of endpoints based on BGP EVPN route advertisements
- Splunk**: Launch Splunk

AUDIT LOGS

Unique Actions by User

username	Action	Count
admin	Login failed - invalid credentials	4
	Login successful. roles: super-admin	
	Modify Userpreference: local-admin	
	Starting upgrade of release: 4.2.1	

Fabric over time

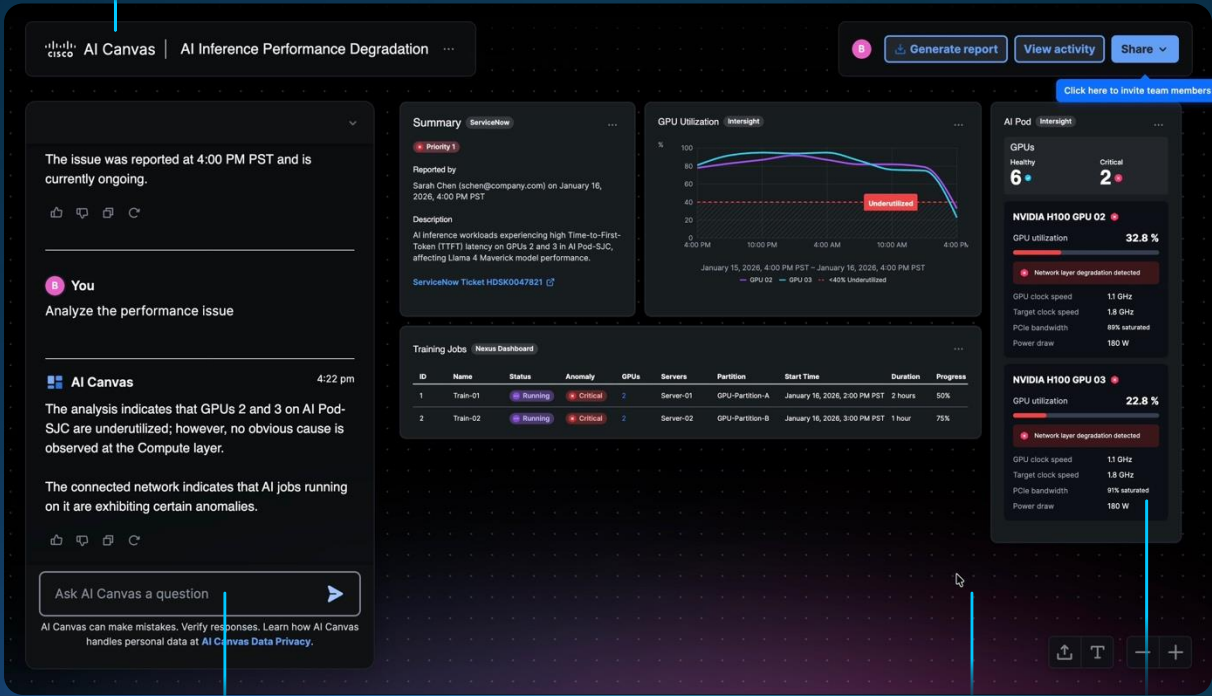


CUSTOM DASHBOARD

Cisco Deep Network Model

Domain-Specific LLM - 20% higher precision on networking tasks

Generative UI Interface



AI Assistant

MCP SERVER

Shared Workspace
Users

Cisco Nexus One

OUR STRATEGY

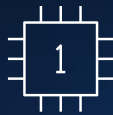
Today's operations requires AI agents, not just AI assistance

Cisco AI Networking

Deliver Resilient, High-Performance Infrastructure

Fuse Security into Every Layer

Unify Operations & Experiences



Silicon



Systems



Software



Optics



Operating Model

CISCO Connect

Thank you

