CISCO Connect

# End-to-End AI Networking

From silicon to workload

Todd Philippi
DataCenter Specialist SE




10/13/2026

# Cisco forward-looking statements

This presentation contains or may contain projections and other forward-looking statements regarding future events or the future financial performance of Cisco, including future operating results. These projections and statements are only predictions. Actual events or results may differ materially from those in the projections or other forward-looking statements. Please see Cisco's filings with the SEC, including its most recent filing on Form 10-K, for a discussion of important risk factors that could cause actual events or results to differ materially from those in the projections or other forward-looking statements.

© 2026 Cisco and/or its affiliates. All rights reserved.

# Cisco powers how people and technology work together across the physical and digital worlds



AI-Ready Data Centers



Future-Proofed Workplaces

Secure Global Connectivity



Digital Resilience

Accelerated by Cisco AI

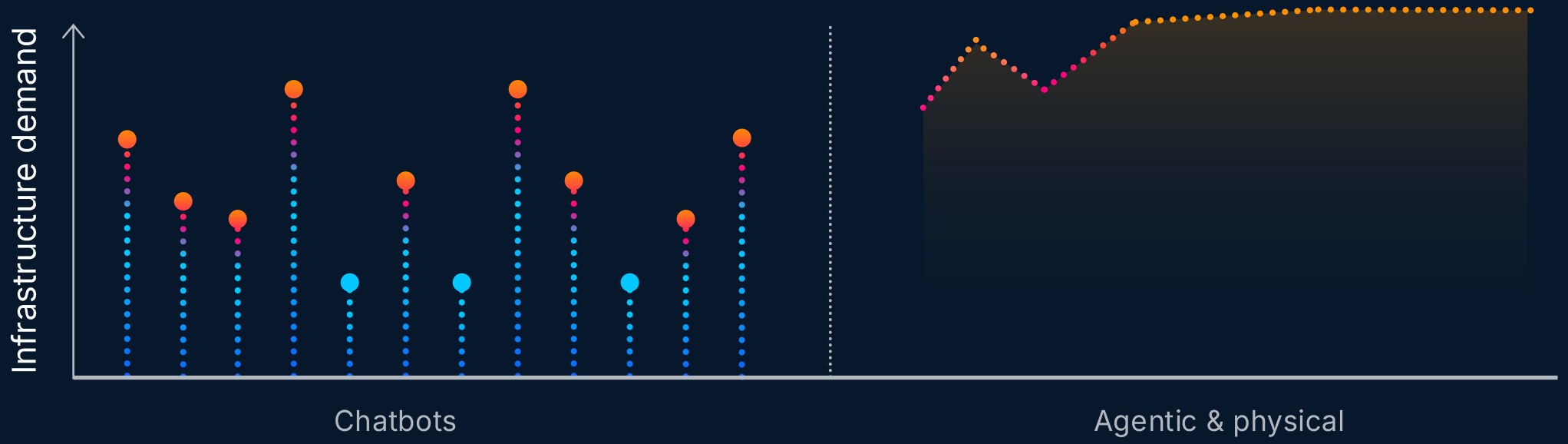
© 2026 Cisco and/or its affiliates. All rights reserved.

# Big shifts are redefining the data center




© 2026 Cisco and/or its affiliates. All rights reserved.

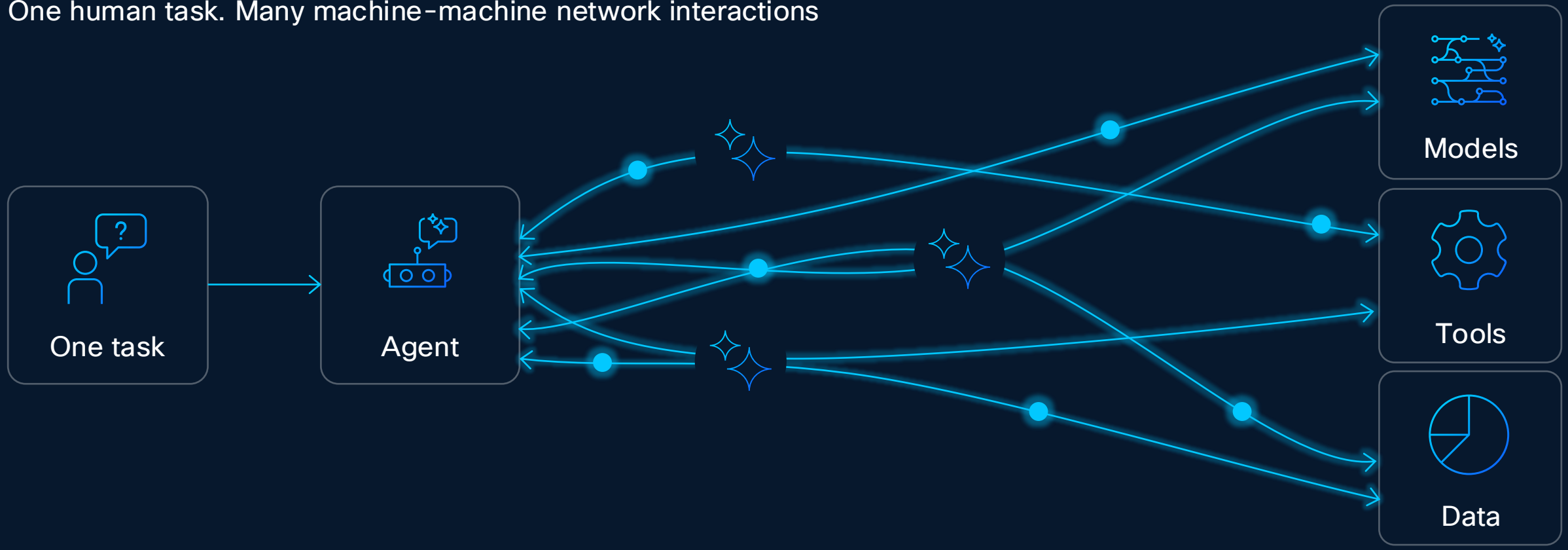
# Infrastructure demands are fundamentally different




© 2026 Cisco and/or its affiliates. All rights reserved.

# Agentic AI changes the role of the network

One human task. Many machine-machine network interactions




## Performance at scale

Latency, congestion, throughput, traffic placement

## Security in the fabric

Policy, segmentation, trust boundaries

## Operations at speed

Telemetry, automation, incident response

© 2026 Cisco and/or its affiliates. All rights reserved.

# Cisco AI-Ready Data Centers

Full-stack AI infrastructure, engineered as one system

Distributed Security fused into every layer

AgenticOps aligning speed with precision



Cisco Cloud Control

© 2026 Cisco and/or its affiliates. All rights reserved.

Rack-Scale Infrastructure

# Cisco’s AI compute portfolio

## Inferencing, Agentic AI at the edge

### NVIDIA PCIe GPU: Edge Platform

Cisco Unified Edge



## Optimization, Inferencing, Agentic AI

### NVIDIA PCIe GPU Servers

Supermicro MGX Server




Cisco UCS® C845A MGX Server



Cisco UCS® X-Series Modular Servers



Cisco UCS® C245 M8SX



Cisco UCS® C240 M8SX



## Model Training

### HGX: Dense GPU Servers

Supermicro NVIDIA HGX Rubin NVL8 liquid-cooled




Supermicro NVIDIA HGX B300 NVL8 liquid-cooled




Supermicro NVIDIA HGX B300 NVL8 air-cooled




## Large scale Inference, Agentic AI, Trillion Parameter Model Training

### Rack-Scale Systems

Supermicro NVIDIA Vera Rubin NVL72




Supermicro NVIDIA Grace Blackwell 300 NVL72

## Storage Server

Cisco UCS® C225 M8N

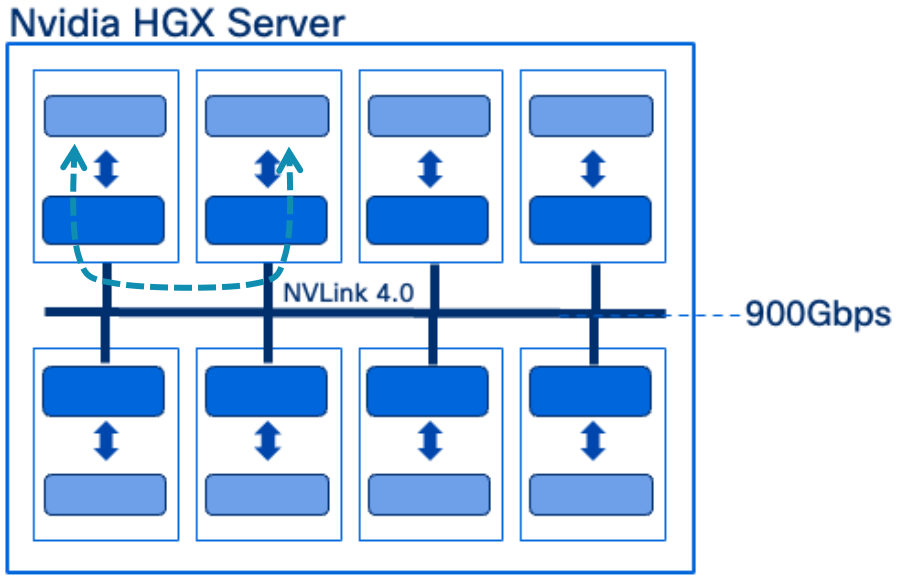


© 2025 Cisco and/or its affiliates. All rights reserved.

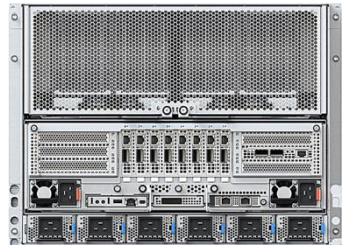
# Inter-GPU Communication Inside A Server

Transfer

- Physical Architecture
  - PCIe6 – 128 GB
  - NVLink v4 – 900GB GPU to GPU
- Memory-centric
  - Direct Memory Access (DMA) between GPUs
  - Unified memory architecture (Nvlink)



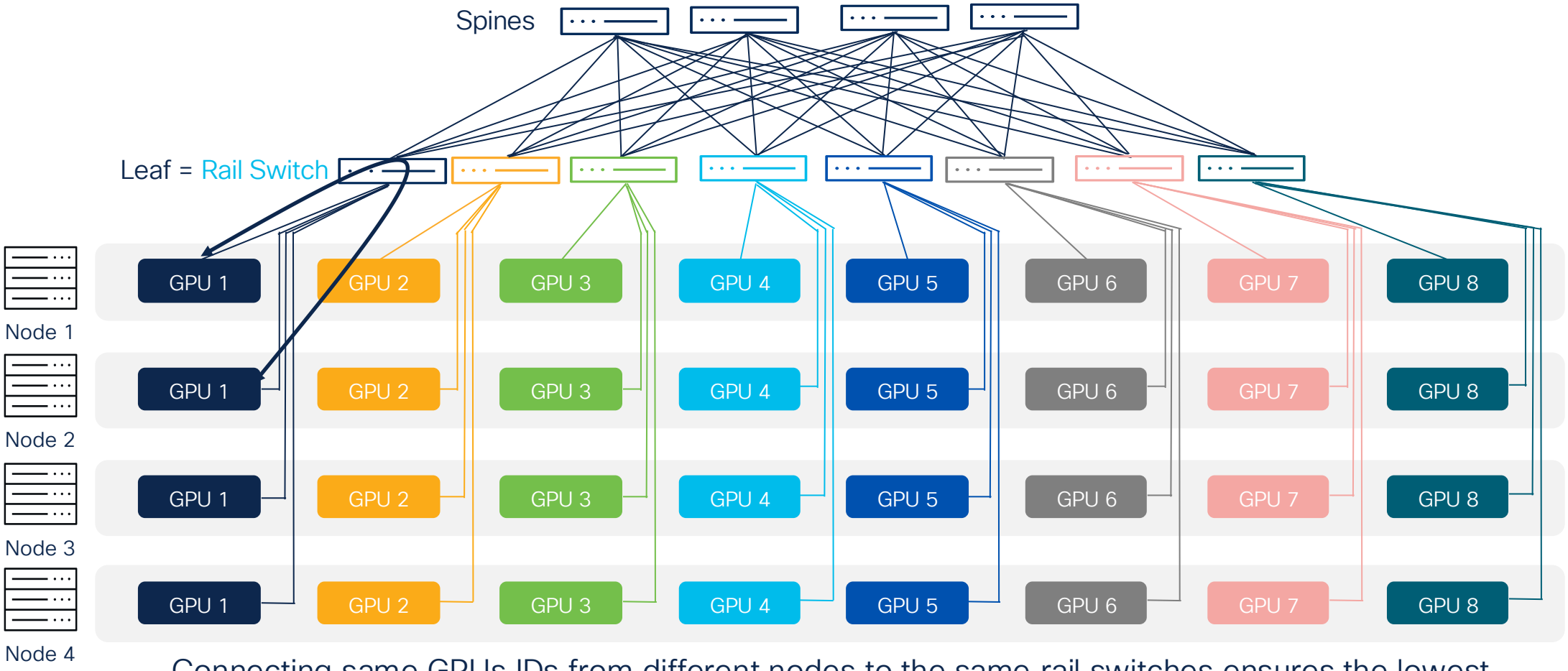

UCS 885A



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

# Back-End Network Options

## Rail-Optimized example with 4 nodes, 8-way GPUs



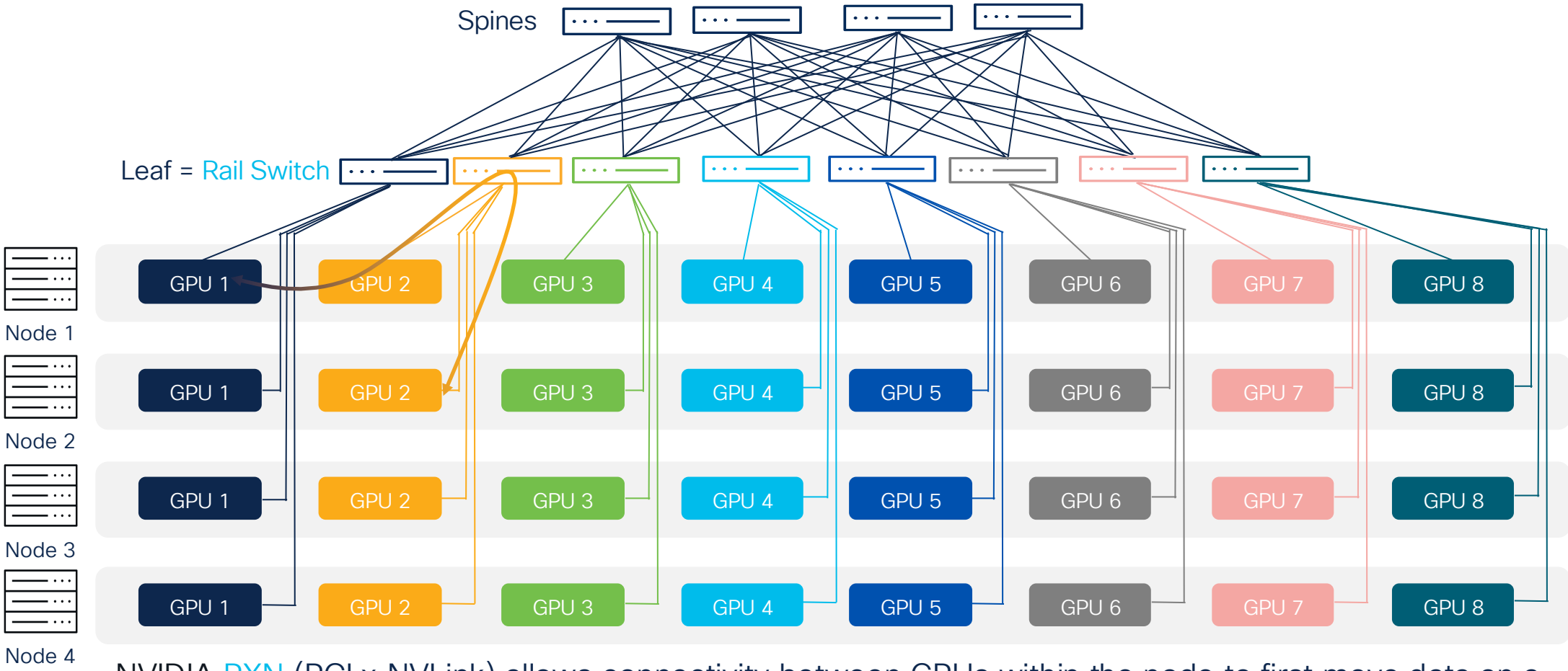

Connecting same GPUs IDs from different nodes to the same rail switches ensures the lowest possible latency across them and helps maximize all2all and all-reduce operations

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

Source: Doubling all2all Performance with NVIDIA Collective Communication Library 2.12

# Back-End Network Options

## Rail-Optimized example with 4 nodes, 8-way GPUs



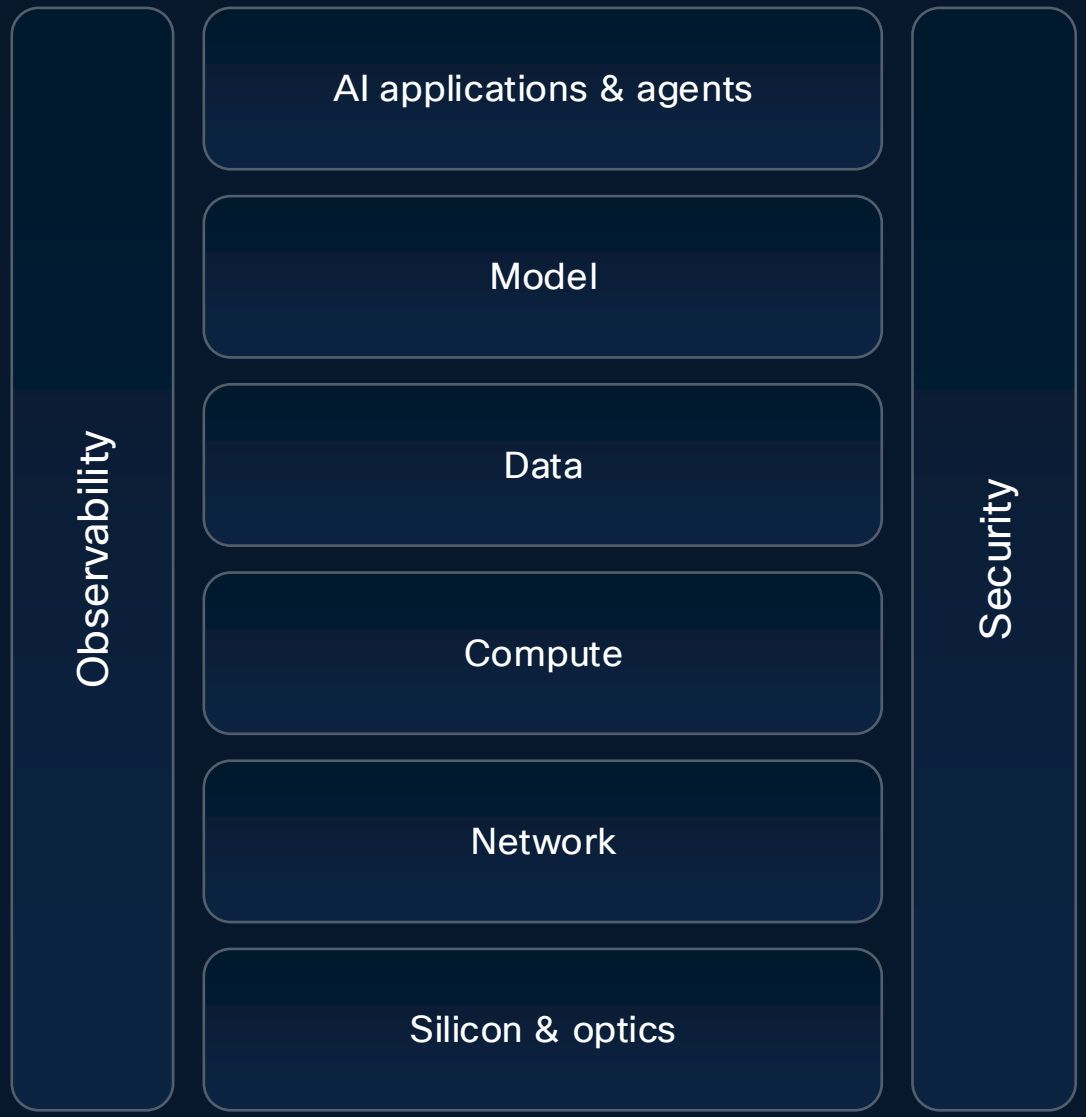

NVIDIA PXN (PCI x NVLink) allows connectivity between GPUs within the node to first move data on a GPU on the same rail as the destination, then send it to the destination without crossing rails.

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

Source: Doubling all2all Performance with NVIDIA Collective Communication Library 2.12

# Vertically integrated platform

## Co-designed full-stack



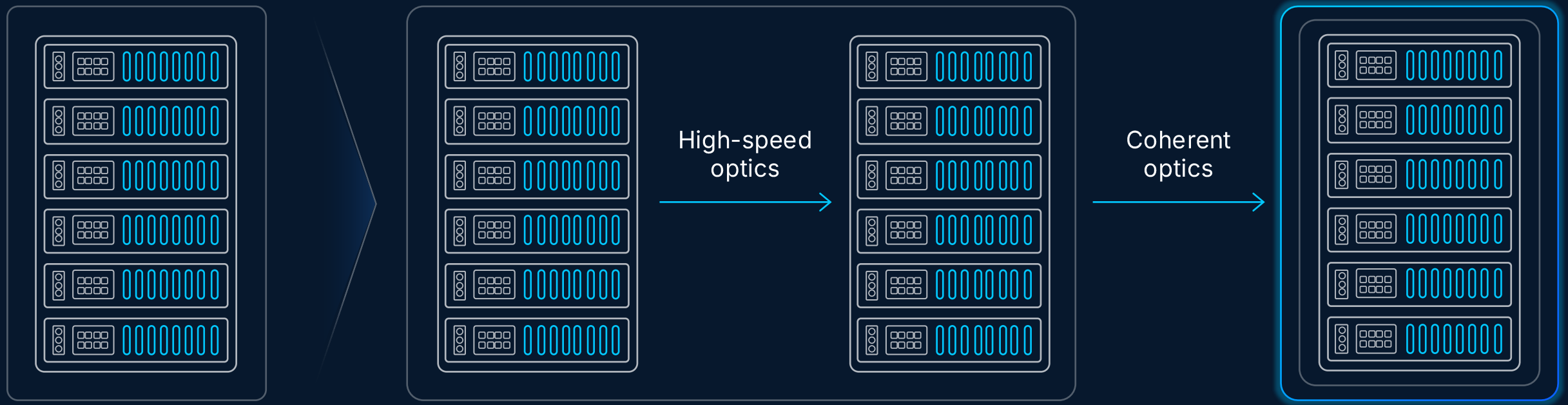

© 2026 Cisco and/or its affiliates. All rights reserved.

# Agentic AI requires unprecedented scale

**Scale up**
within a system

**Scale out**
across systems

**Scale across**
between data centers




© 2026 Cisco and/or its affiliates. All rights reserved.

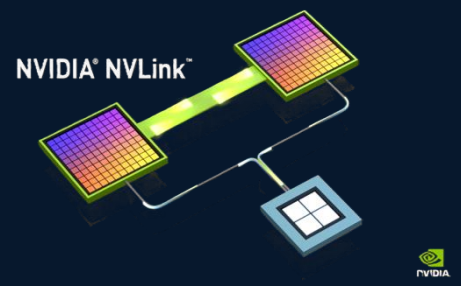
# Cisco Optical/Optics, Silicon, and Systems power up AI ready network connectivity

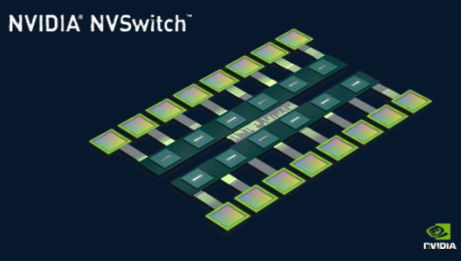
Scale-UP

Scale-out

Scale-Accross

**ESUN | UALoE**












N9364E-SG2
64x300G | 2RU (G200)



N9364E-SP2R
64x800G | 3RU



N9364F-SG3-L
64x1.9T | 2OU (G300)



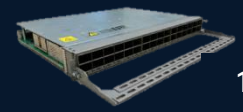
N9364E-SP2R
64x800G | Line Card



**500km+ reach**
100G ZR
**1000km+ reach**
400/800G ZR/ZR+
**3000km+ reach**
400G ULH



3.2T Pluggable Line Card



12.8T OSFP line card



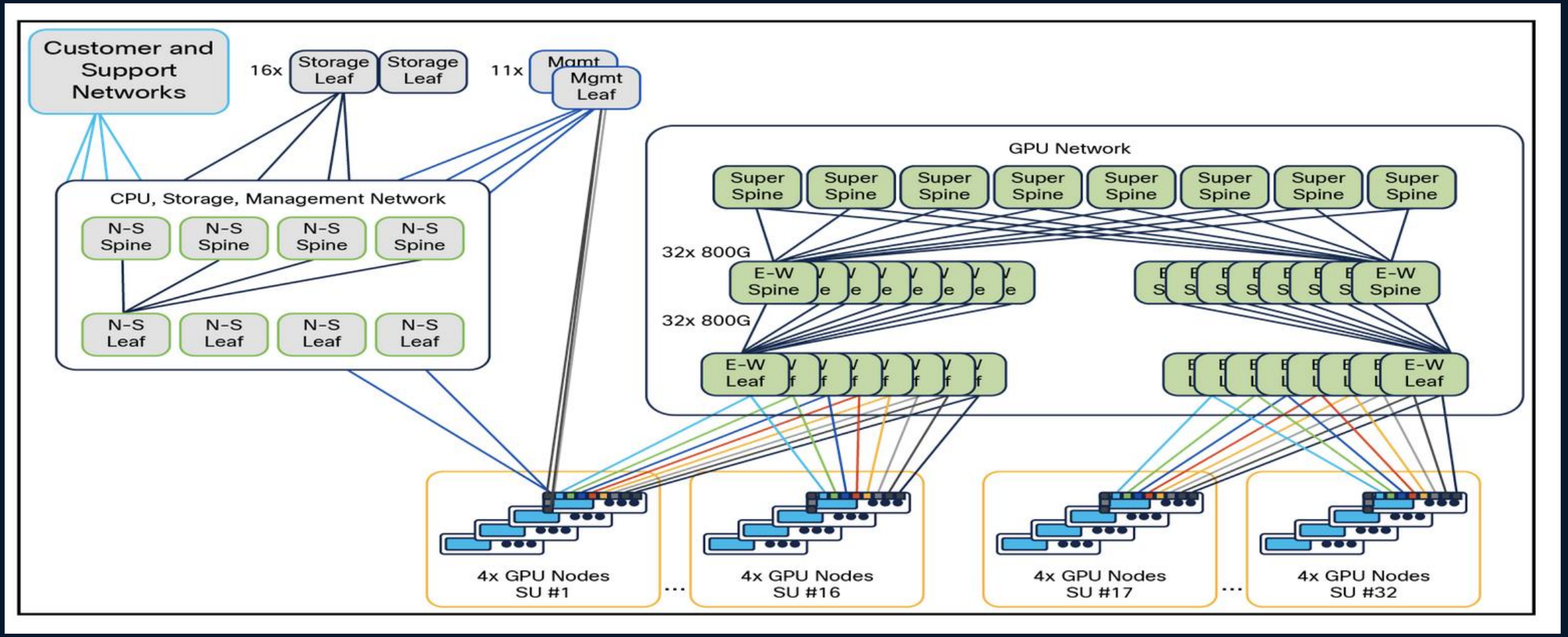
NCS 1014 multi-haul Optical platform
25.6T in 2U

**Scalable**

**Reliable**

**Power-Efficient**

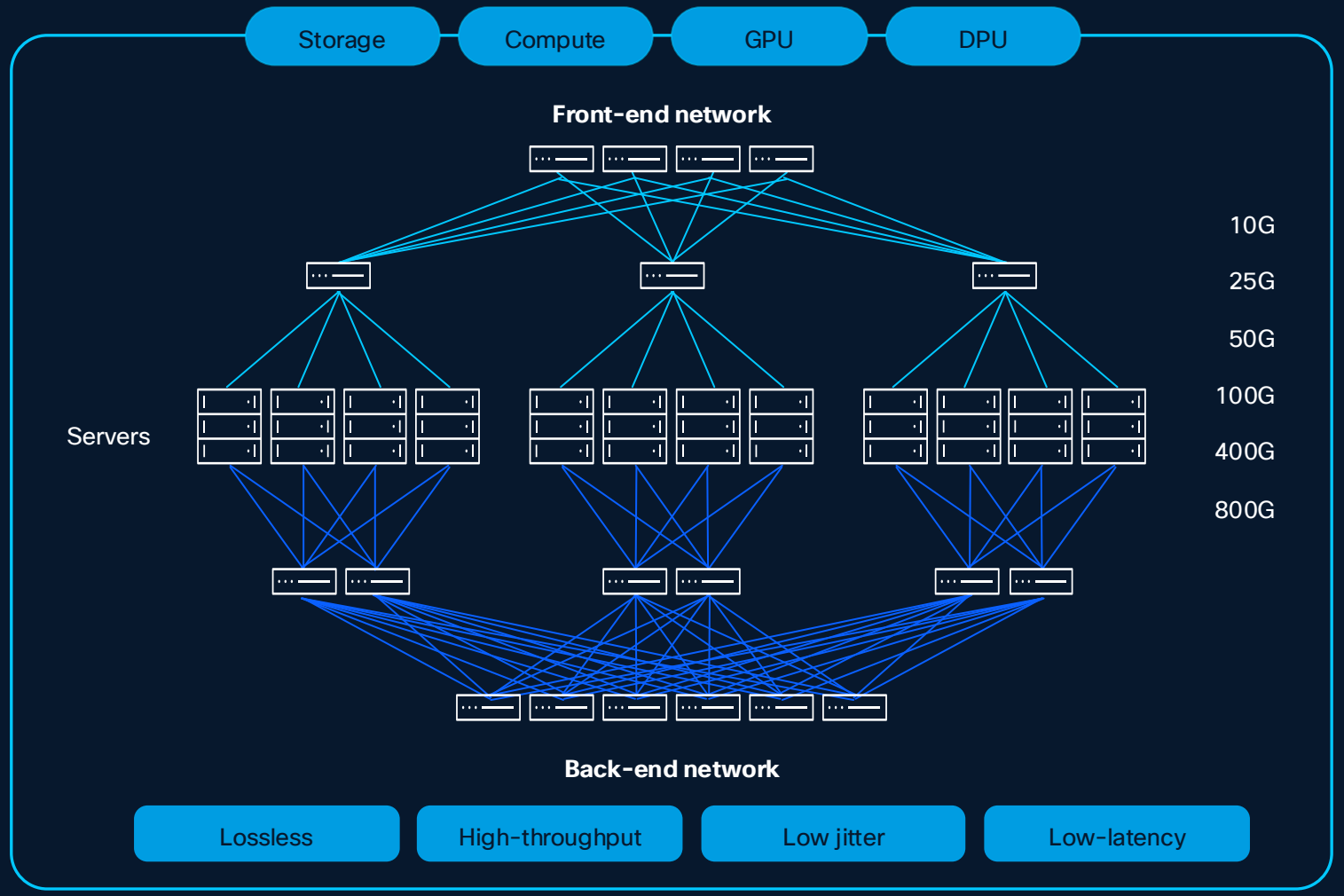
# The Complete View




© 2026 Cisco and/or its affiliates. All rights reserved.

# Cisco's AI/ML approach with Nexus

Platforms, Software, Operations, Partnerships




### Platforms
CloudScale, Silicon One, Spectrum
400G, 800G & 1.6T

### NX-OS /ACI/ IOS-XR/SONIC
Cisco Intelligent Packet Flow, Adaptive Routing, UEC features

### Cisco Nexus Dashboard
AI fabric templates, AI analytics, telemetry, Congestion Scores, Job Monitoring, GPU/NIC stats

### Vendor Agnostic Approach
NVIDIA, AMD, Intel, VAST, Supermicro and others

### Benchmark Performance Testing
ML Commons, IBPerf, NCCL, DeepSeek

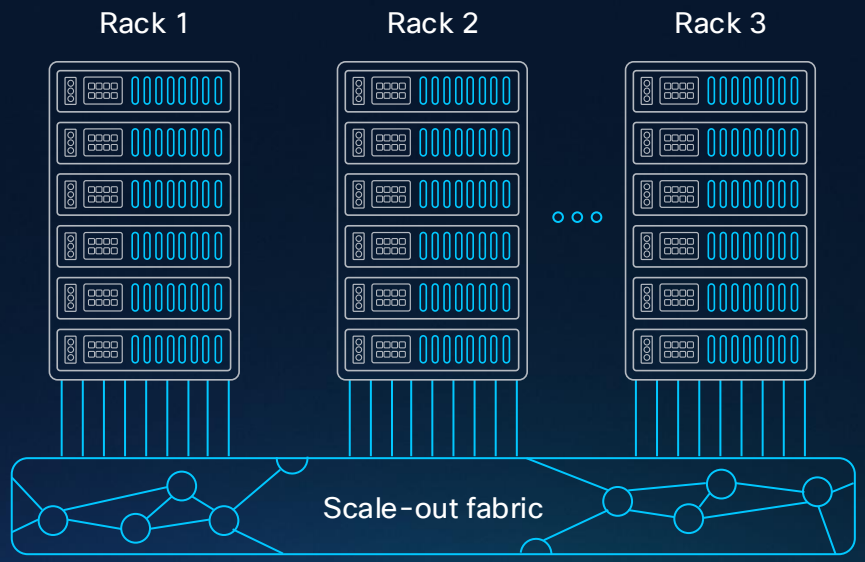
© 2026 Cisco and/or its affiliates. All rights reserved.

# AI performance at scale depends on the network

Scale-out extends AI compute beyond a single rack

## Backend XPU-to-XPU network

For distributed training and inference




## Traffic patterns

Synchronized bursts. Many-to-many exchanges.

## Keep accelerators busy

High bandwidth, low and predictable latency

## Handle traffic bursts

Congestion control, intelligent buffering, adaptive load balancing

## Grow efficiently

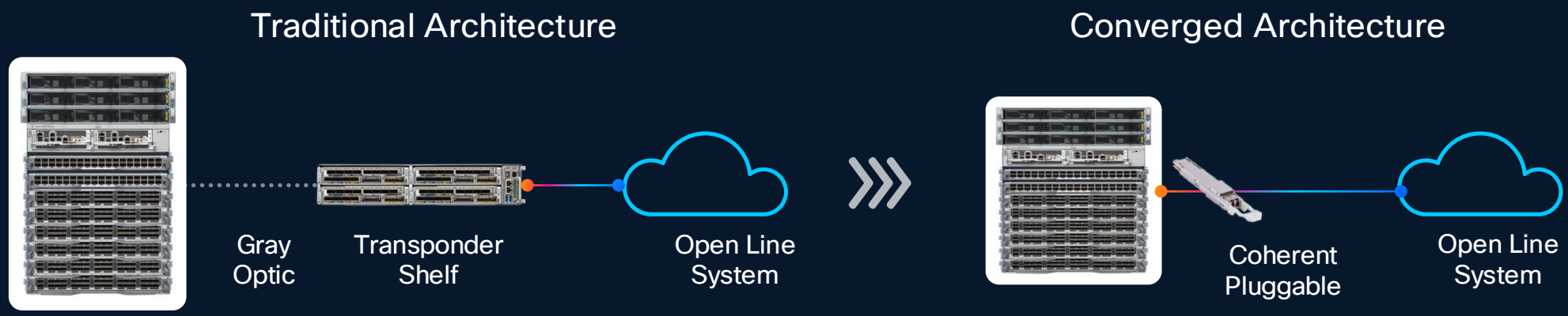
High port density, fewer network tiers, power efficiency

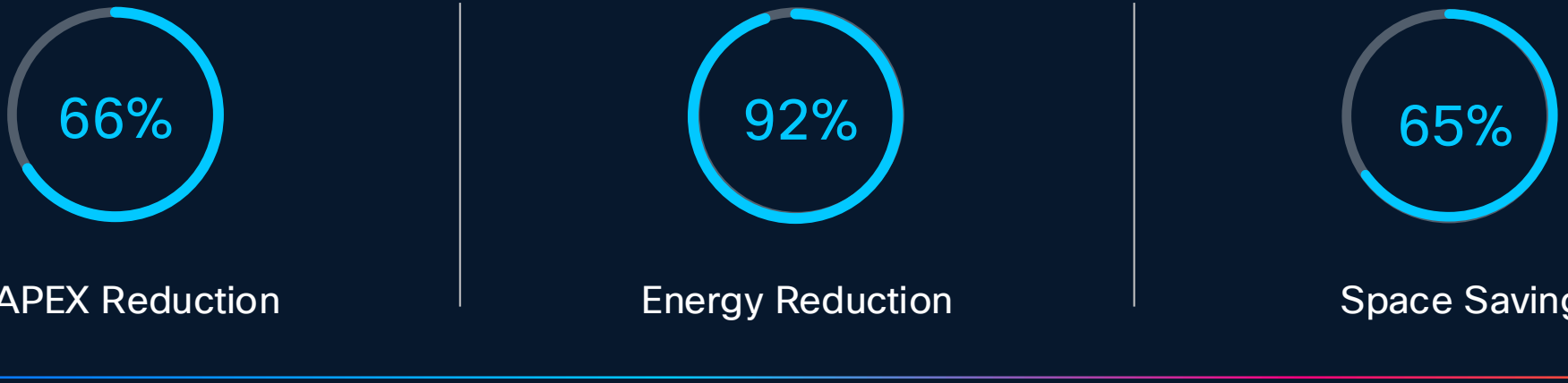
## Maintain performance

Reliable connectivity, rapid fault recovery, granular telemetry

© 2026 Cisco and/or its affiliates. All rights reserved.

# Pluggable optics benefits for DCI and scale-across



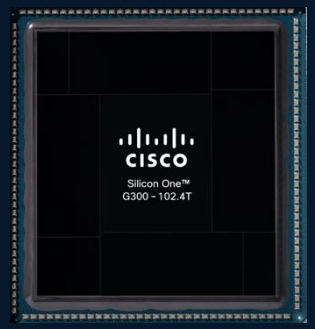




© 2026 Cisco and/or its affiliates. All rights reserved.

*Results from ACG Research*

# Cisco Silicon One

Scalable, programmable unified networking architecture for the AI Ready Data Centers




**G-series**
AI scale-out

Low latency, high-radix fabric optimized for AI training and inference




**P-series**
DCI & AI scale-across

Deep-buffer routing across data centers




**E-series**
Enterprise modernization

Flexible leaf-spine switching with integrated security

## Foundation for AgenticOps

Hardware telemetry powering proactive operations through Cisco Cloud Control

© 2026 Cisco and/or its affiliates. All rights reserved.

# Data center networking hardware portfolio

CISCO CLOUD CONTROL

Modernize enterprise data centers

Build dedicated AI data centers



**Storage networking**

Cisco MDS ASIC

Up to 64G



**Fixed switches**

Silicon One

Cloud Scale

Up to 800G



**Modular switches**

Silicon One

Cloud Scale

Up to 800G



**Smart switches**

Silicon One E100

AMD DPUs

Up to 400G

Scale-out



**AI fabric switches**

Silicon One G-Series

NVIDIA Spectrum-X

NCP reference architecture

Up to 1.6T

Scale-across



**Deep-buffer systems**

Silicon One P-Series*

Up to 800G

ENTERPRISES | SERVICE PROVIDERS | NEOCLOUDS | SOVEREIGN CLOUDS | HYPERSCALERS

© 2026 Cisco and/or its affiliates. All rights reserved.

*Smart switches uplink throughput is 400G, with AMD DPUs service throughput is 800G.
*P-Series for Scale Across, Universal Spine, and DCI

# Cisco N9300: High-density 800G for AI

Efficient, scalable networking for AI clusters

Scale out

## 64 × 800G | 2RU

NEW
Q3CY26



**N9364E-SG2X-O**
OSFP



**N9364E-SG2X-Q**
QSFP-DD

### Lower power

Linear Pluggable Optics support
Reduce power consumption by 30%

### Built to scale

256MB shared buffer, 512 radix
Larger AI clusters. Fewer network layers

### Faster AI jobs

Low latency reduces
job completion time

© 2026 Cisco and/or its affiliates. All rights reserved.

Scale out

# 1.6T AI networking. More deployment choices.

102.4 Tbps | 64 x 1.6T OSFP ports

## Cisco Silicon One G300

### Air Cooled



3RU, 64p 1.6T OSFP
EIA 19"

Redundant
Pluggable Power

Cisco N9364F-SG3-EA | Cisco 8133
Q2 CY27

### Liquid Cooled



2OU, 64p 1.6T OSFP
ORv3 21" blind mate

Integrated power
Leak detection

Cisco N9364F-SG3-RL | Cisco 8132
Q2 CY27

## NVIDIA Spectrum-X

### Liquid Cooled

NCP reference architecture



2RU, 64p 1.6T OSFP
MGX 19”

Integrated power
Leak detection

Cisco N9164F-NS6-ML
Q2 CY27

Operational choice: NX-OS and SONiC

© 2026 Cisco and/or its affiliates. All rights reserved.

Scale out

# G300: More performance. Less infrastructure.

Cisco Silicon One with Intelligent Collective Networking

## Use more network capacity

**33%**

Higher network throughput (1)

Handle more GPU traffic

## Finish AI Jobs sooner

**28%**

Lower job completion time (2)

Free GPUs for the next job

## Reduce network footprint

**1 replaces 6**

Same bandwidth (3)

70% better efficiency (3)

(1) simulated: 33% higher aggregate throughput for static ECMP load balancing compared to reduced buffer solutions

(2) simulated: 28% reduction in JCT for G300 path-aware load balancing compared to per-packet (spray) port selection

© 2026 Cisco and/or its affiliates. All rights reserved.

# Scale-Out Optics: More bandwidth. Less power.

High-speed connectivity for AI backend networks

## 1.6T OSFP



**Scale bandwidth**

High-speed links for AI switches and NICs

Air- and liquid-cooled platforms

Diversified supply chain

## 800G LPO

Linear Pluggable Optics



**Cut power at 800G**

**50%**

Less optics power (1)

**30%**

Less switch power (2)

**Simplify optics management**

Telemetry | Diagnostics | Firmware updates

© 2026 Cisco and/or its affiliates. All rights reserved.

1. Compared with retimed optical modules. 2. On supported Cisco N9000 and 8000 systems using LPO.Source: Cisco G300, systems and optics announcement, February 2026.

# Cisco reference architectures scale from enterprise to clouds

## 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

NVIDIA
NCP RA Compliant

**Cisco AI Servers from ~1K to 32K GPUs**

Cisco Silicon One

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

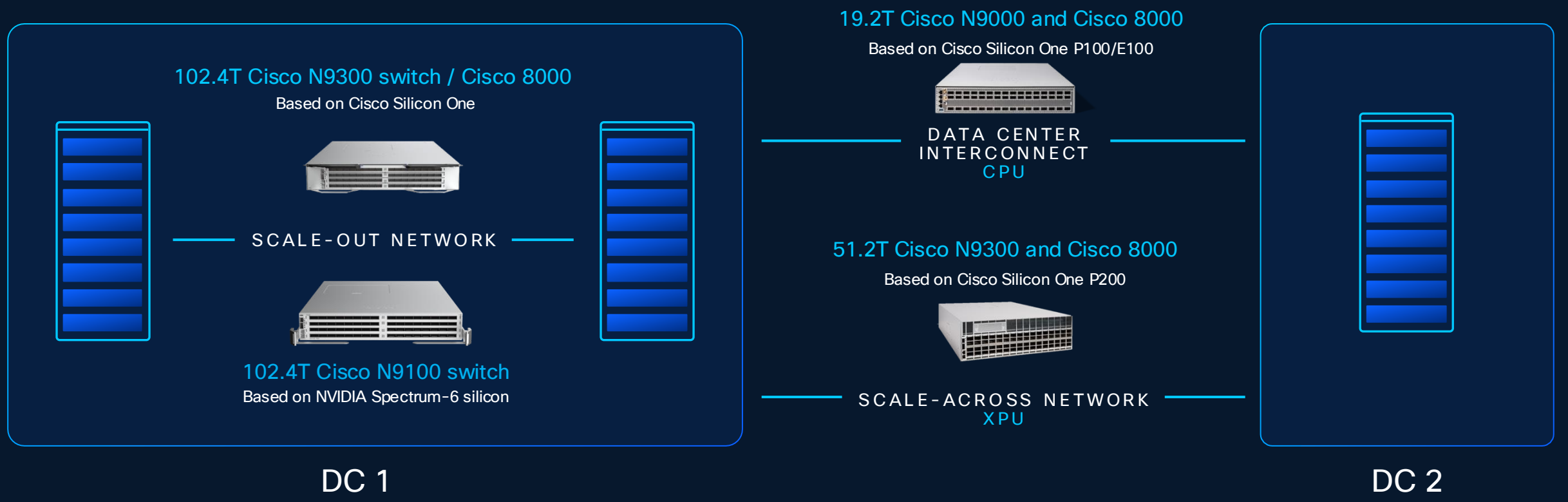
NVIDIA Spectrum-X
Ethernet Silicon

**Cisco N9100**
Series Switches
(51.2T, 102.4T)

© 2026 Cisco and/or its affiliates. All rights reserved.

# Scale-out, DCI and scale-across networking

Validated architecture choices for AI scaling




© 2026 Cisco and/or its affiliates. All rights reserved.

DCI

# Data Center Interconnect

## Cisco N9000 and Cisco 8000 Systems powered by Cisco Silicon One P-Series and E-/K-Series




NX-OS

### Fixed Systems

Up to 3.6T
36x100G




IOS-XR

### Fixed Systems

Up to 6T
100G/400G

### Modular Systems

Up to 9.6T per slot
100G/400G




SONiC IOS-XR NX-OS ACI

### Fixed Systems

Up to 51.2T
64x800G

### Modular Systems

Up to 28.8T per slot
36x800G



### 100G-800G ZR/ZR+

Long-reach coherent optics
800G ZR/ZR+ for 1000km+ reach

**Long-reach, lower-complexity DCI**

- 100G-800G ZR/ZR+ optics
- Fewer transport layers, up to 60% lower TCO

**Predictable transport under load**

- Silicon One programmable packet processing

**Secure, observable**

- Quantum resilient line-rate encryption
- Multi-layer visibility and automation

© 2026 Cisco and/or its affiliates. All rights reserved.

*P100 available on Cisco 8000

cisco

Scale across/DCI

# Scale across networks

## Cisco N9000 and Cisco 8000 Systems powered by Cisco Silicon One P200




NX-OS ACI SONiC IOS-XR









### P200 Silicon One

51.2T deep-buffer, programmable packet processing

### Fixed Systems

51.2T scalable fabric

### Modular System

28.8T line-card scale

### Coherent Optics

800G ZR/ZR+ for 1000km+ reach

### Open Line Systems

NCS 1010

**Pool AI capacity across sites**

- Overcome power, space, and data limits

**Predictable performance for GPU traffic**

- Manage bursts and congestion over distance
- Quantum-resistant line-rate encryption

**Reduce transport cost**

- 800G ZR/ZR+ coherent pluggables
- High-performance interop and reaches

**Scale across fiber pairs**

- C&L Band spectral efficiency
- In-line amplification

© 2026 Cisco and/or its affiliates. All rights reserved.

cisco

# Cisco Nexus One

FULLY INTEGRATED STACK

## Nexus Dashboard
On premises

## Nexus Hyperfabric
Cloud management

| SILICON | SYSTEMS | OPTICS | SOFTWARE |
|---|---|---|---|
| Cisco Silicon One<br>Cisco Cloud Scale<br>NVIDIA Spectrum-X Ethernet | Cisco N9000 Systems | Cisco Optics | Cisco NX-OS<br>Cisco ACI<br>SONiC |

SECURITY AND OBSERVABILITY

© 2026 Cisco and/or its affiliates. All rights reserved.

# AI Infra Orchestrator Integration with Nexus One

## Multi-tenancy in the backend AI Network

**CISCO NEXUS ONE**
**workload-aware fabric services**

**Automated Tenant Onboarding**

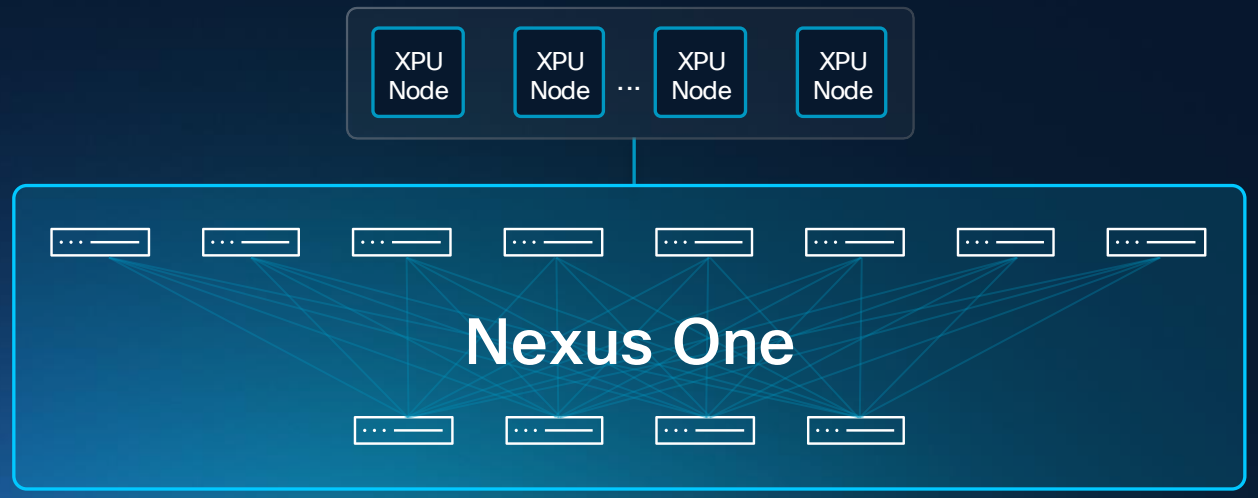
**Multi-Tenant Isolation at Scale**

**GPU Cloud / Tenant & AI Infra Orchestrator**

GPUaaS Enablement Layer

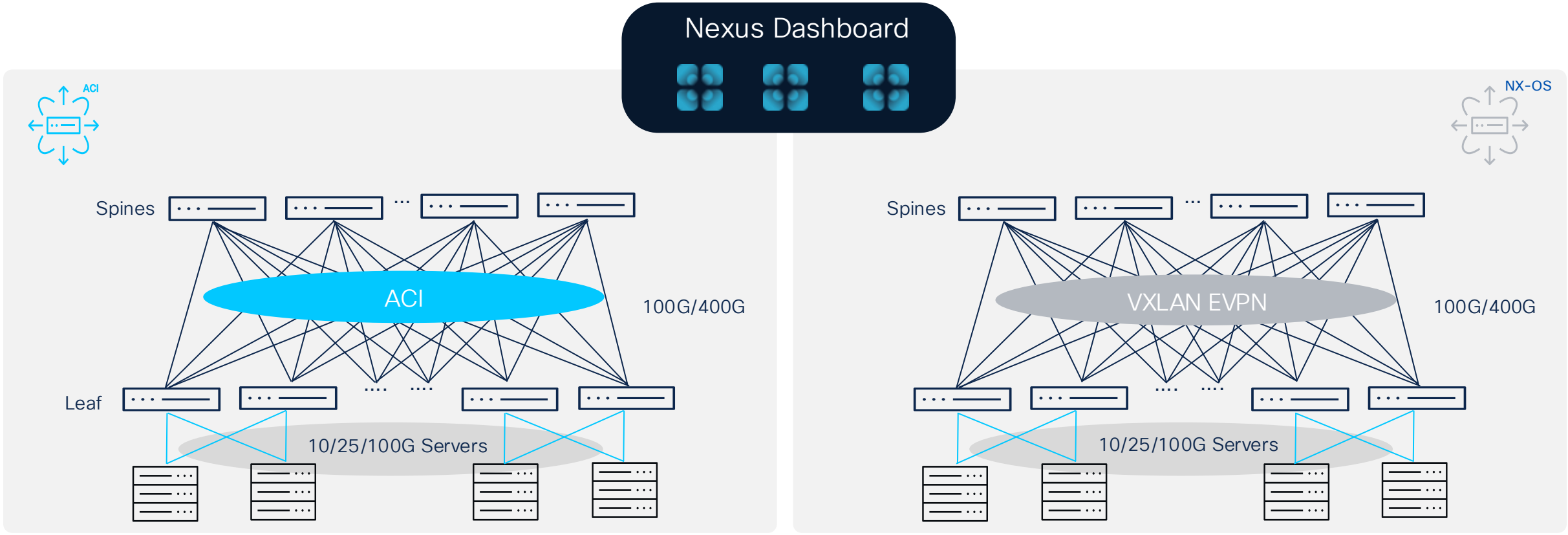
Open-Source Options

Proprietary Operator Platform



# Front-End Network Options

- Building or repurposing existing infrastructure for AI workloads



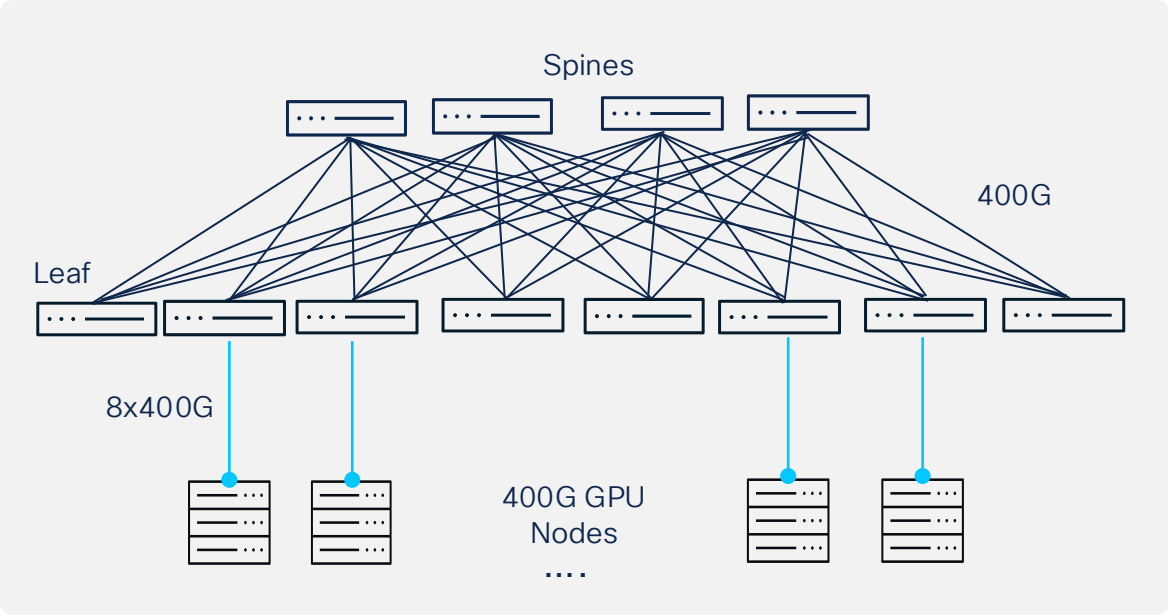

ACI Fabrics

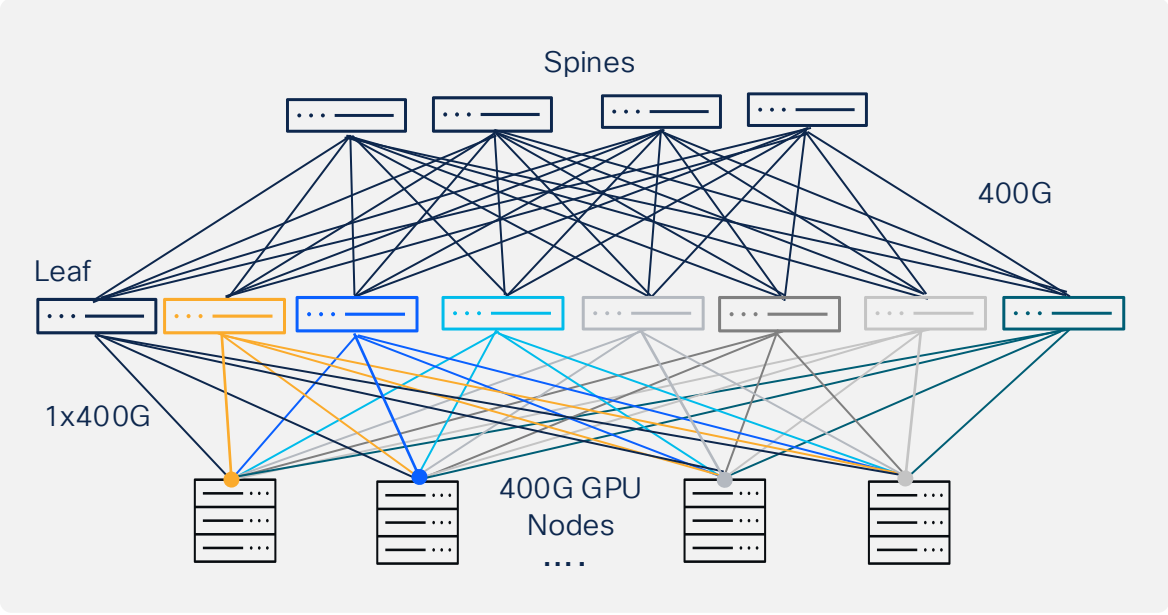
VXLAN EVPN Fabrics

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

# Back-End Network Options, but not limited to

- Classic or Rail-Optimized Architectures with Nexus 9000 Fixed Series




Classic Design
Hosts mapped per Rack/Switch
Non-blocking / 1:1 Oversubscription / BGP Fabric
Up to 64/128-port 400G or 800G options for Leaf/Spine

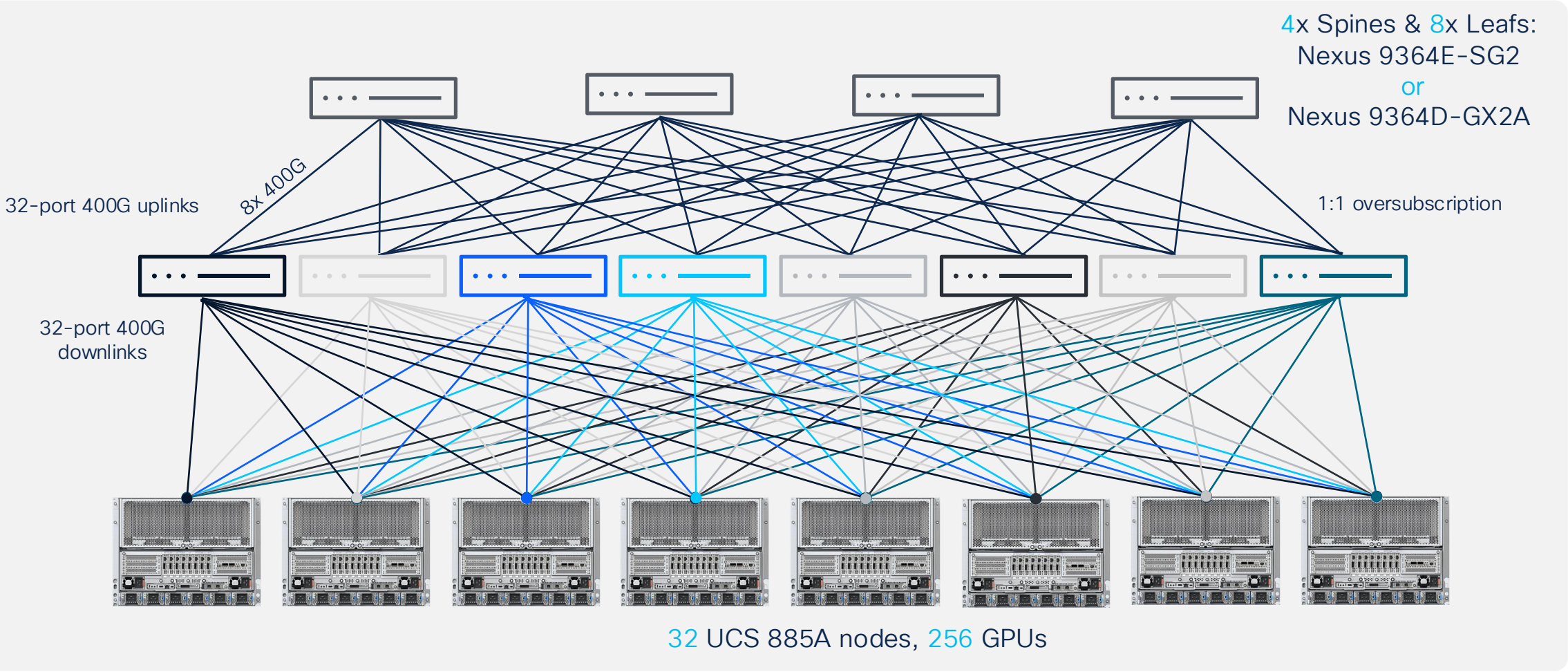
Rail-Optimized Design
8 GPU:NIC pair mapped per Rail/Switch
Non-blocking / 1:1 Oversubscription / BGP Fabric
Up to 64/128-port 400G or 800G options for Leaf/Spine

These design choices strictly depend on Collective Communication Library optimizations to improve overall performance.

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

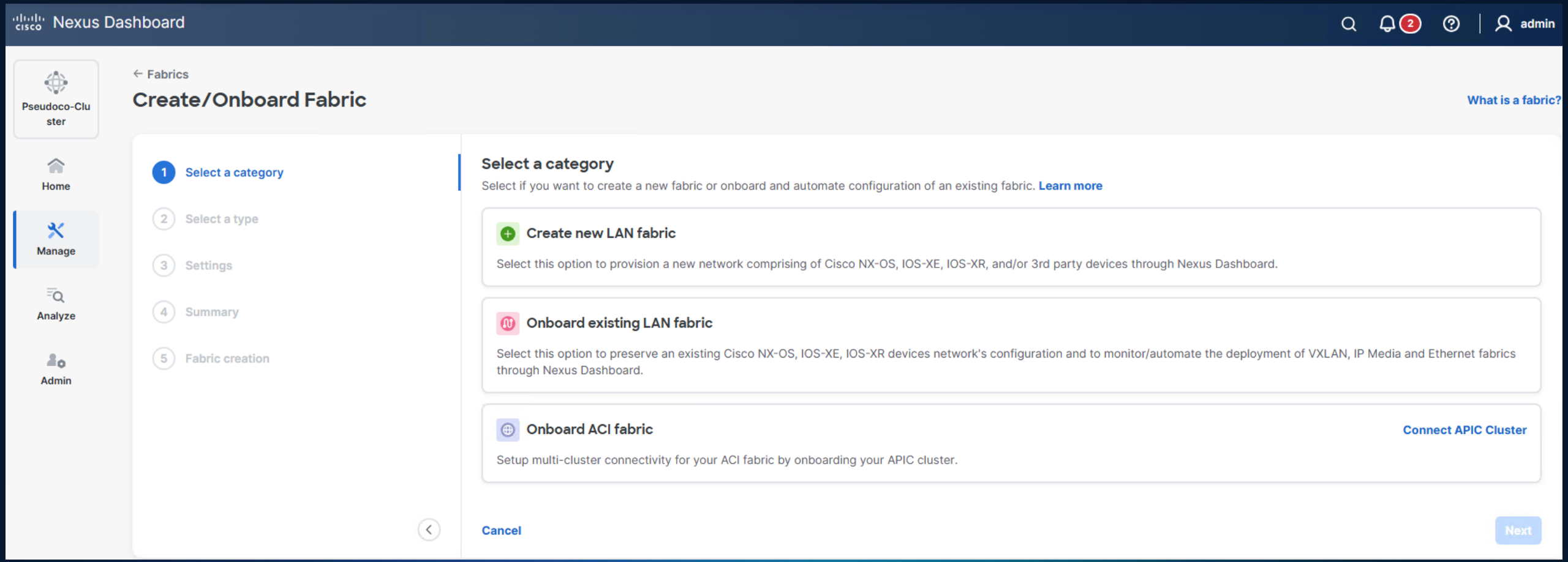
# 256 GPUs Back-End Network

- Spine and Leaf with Rail-Optimized



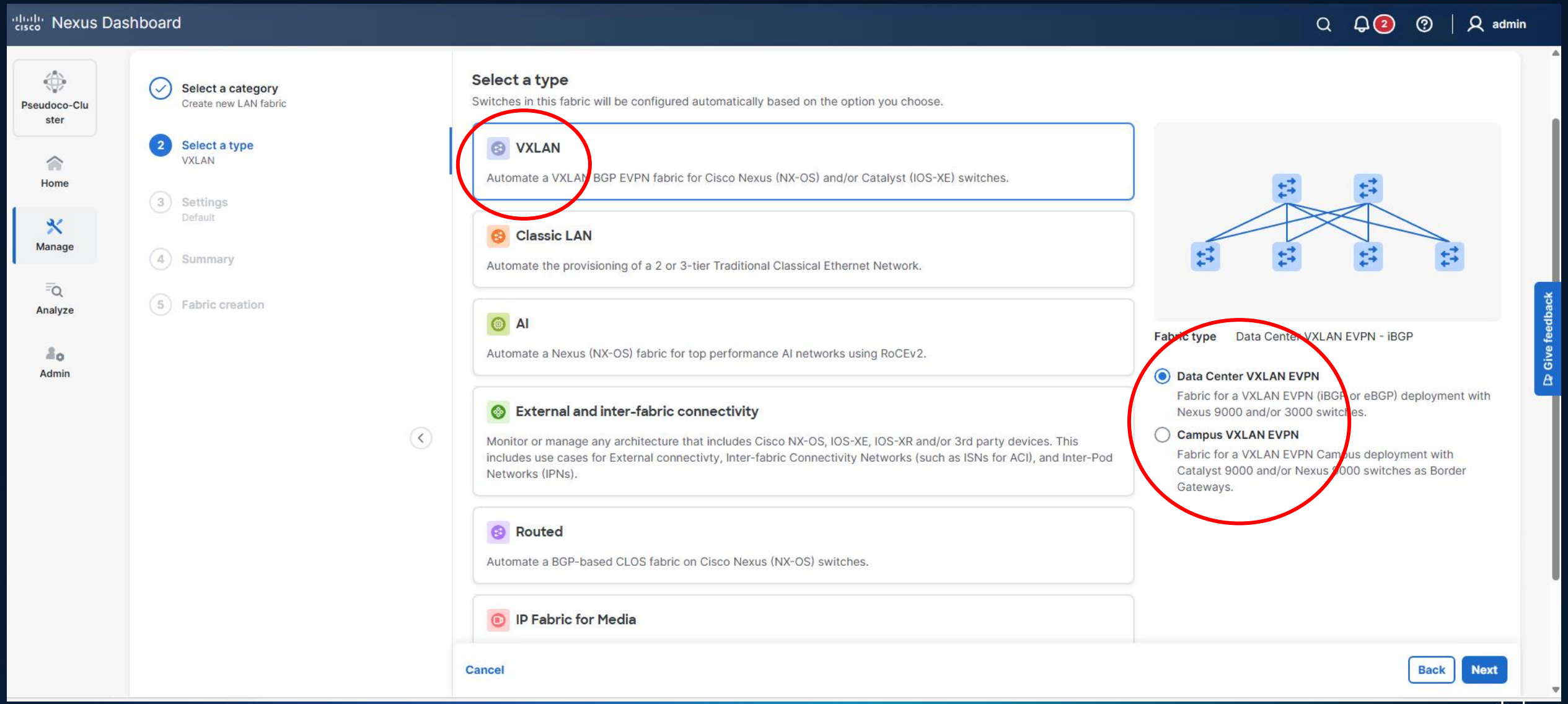

© 2025 Cisco and/or its affiliates. All rights reserved.

# Nexus Dashboard, to configure them all.



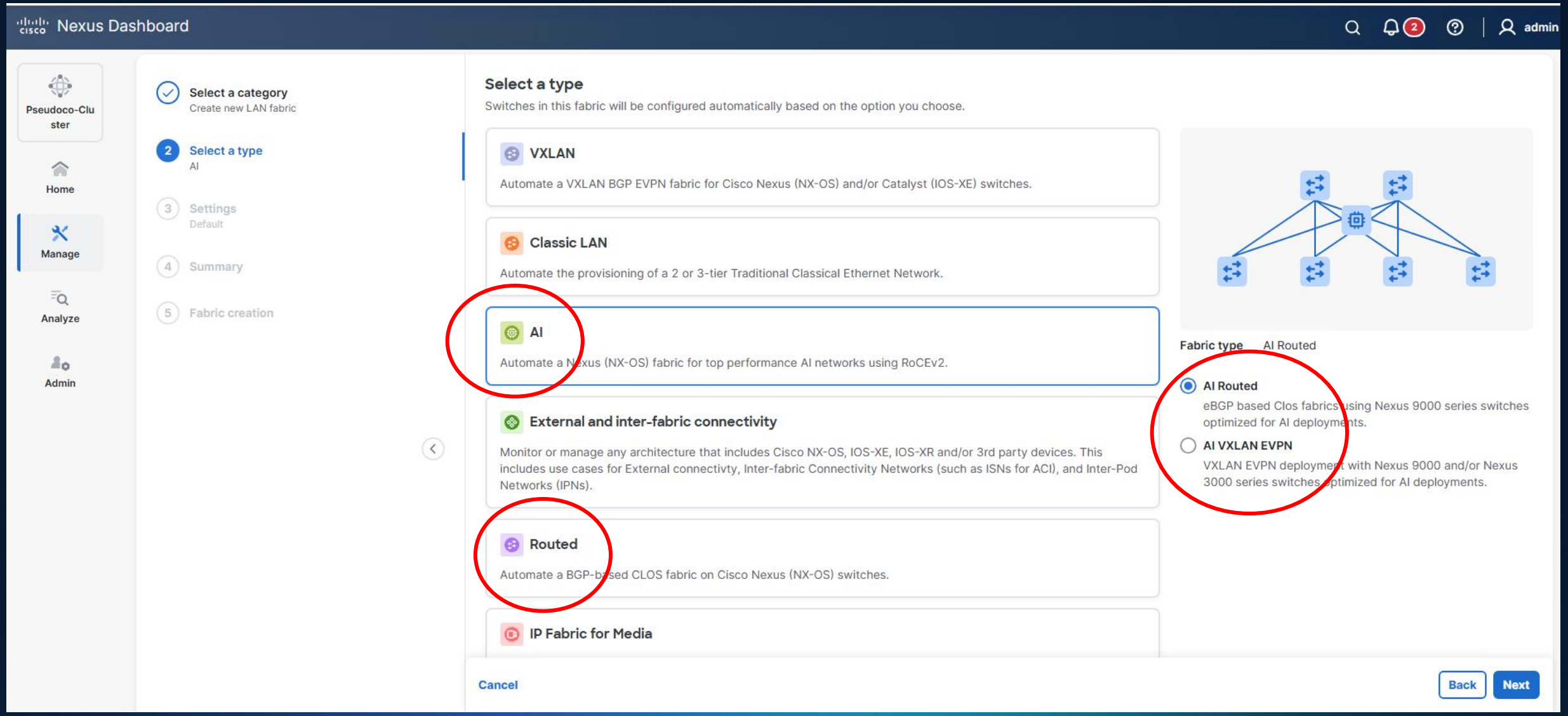

© 2026 Cisco and/or its affiliates. All rights reserved.

# Nexus Dashboard, to configure them all.




© 2026 Cisco and/or its affiliates. All rights reserved.

# Nexus Dashboard, to configure them all.




© 2026 Cisco and/or its affiliates. All rights reserved.

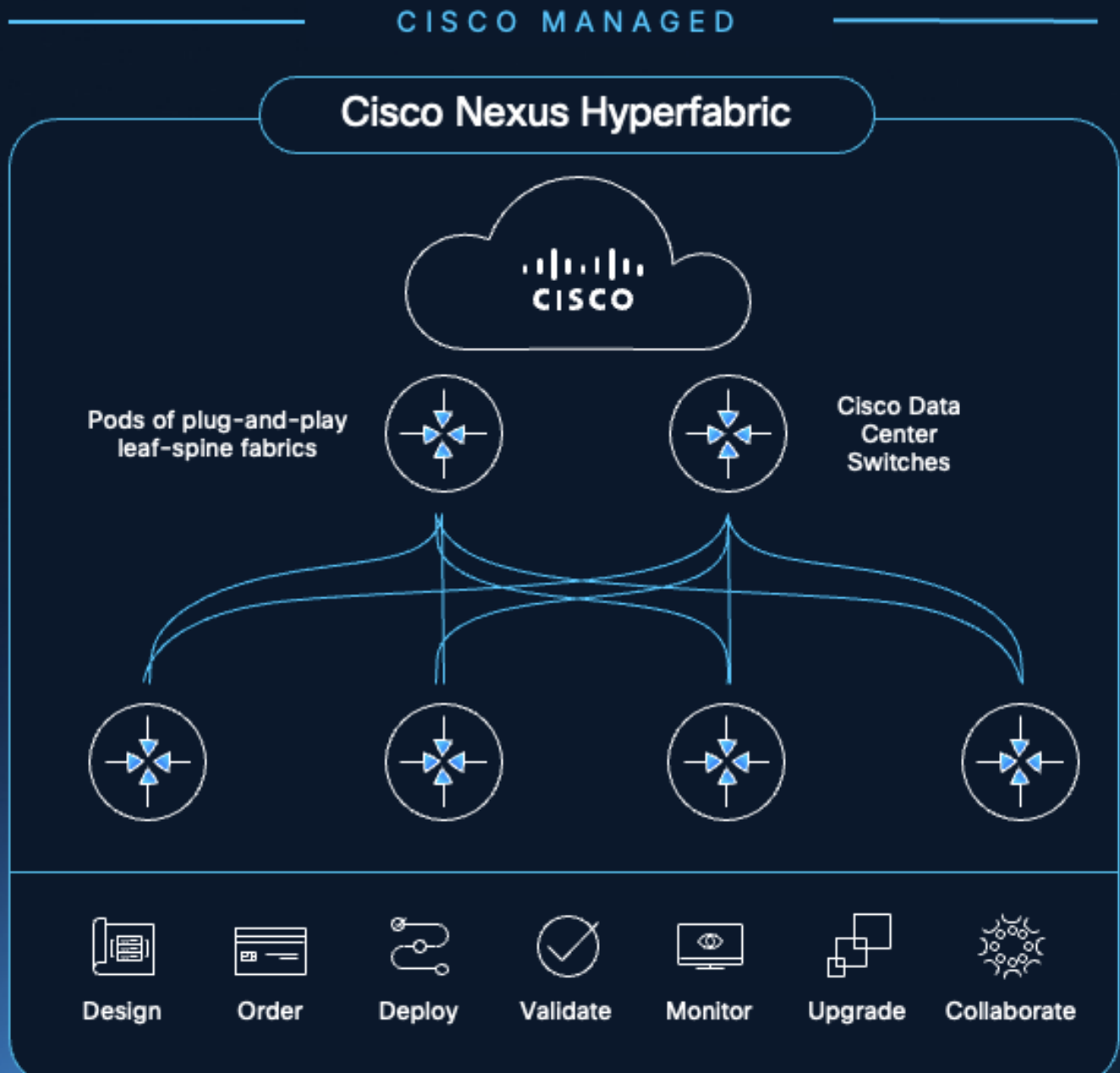
Cisco Cloud Control

# Nexus Hyperfabric

Design, deploy, and operate on-premises fabrics—anywhere, from the cloud.

Streamlined operations for IT generalists, application, and DevOps teams

Outcome driven using purpose-built vertical stack



# Cisco 6000 series switches







## Backend Switch HF6100-64ED

- OSFP form factor
- 64x 100/200/400/800GbE
- Breakouts supported

## Storage Switch HF6100-32D

- QSFP56-DD form factor
- 32x 100/200/400 GbE

## Management Switch HF6100-60L4D

- 4x 100/400GbE QSFP56-DD (16x 100G breakout)
- 60x 10/25/50GbE SFP56

## Benefits

Cloud-managed | Open & extensible | High performance Silicon One | Optimized for AI workloads

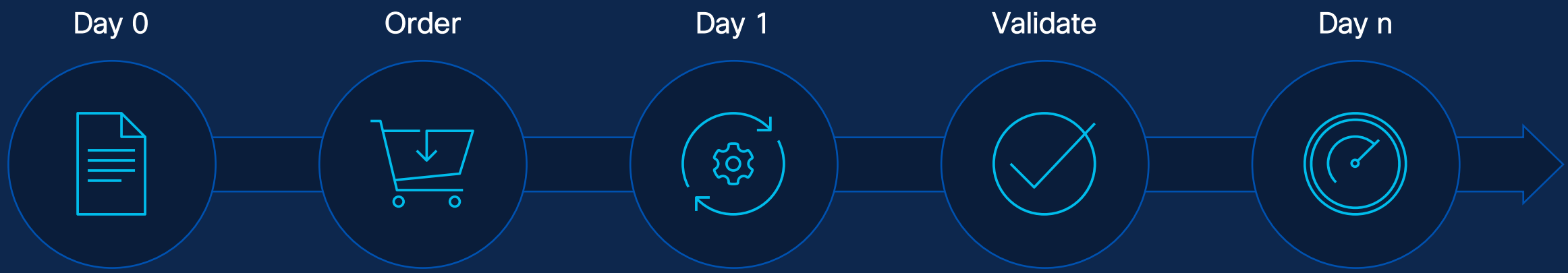
© 2025 Cisco and/or its affiliates. All rights reserved.

# Nexus Hyperfabric AI experience



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

# Complete lifecycle experience




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

# Complete AI stack

## Cloud-managed controller

- Scalable, globally distributed multi-tenant cloud service
- GUI, mobile, and API access
- Helping hands for smart remote hands visibility

## Cloud-managed GPU cluster

- Cisco UCS high density NVIDIA GPU servers with Bluefield 3
- Boots from cloud
- Full visibility and control from the cloud
- Available with the full suite of NVAIE/NIM tools from Nvidia




## Cloud-managed switch

- Cisco 6000 series
- Boots from cloud
- Full visibility and control from the cloud



## Cloud-managed VAST storage cluster

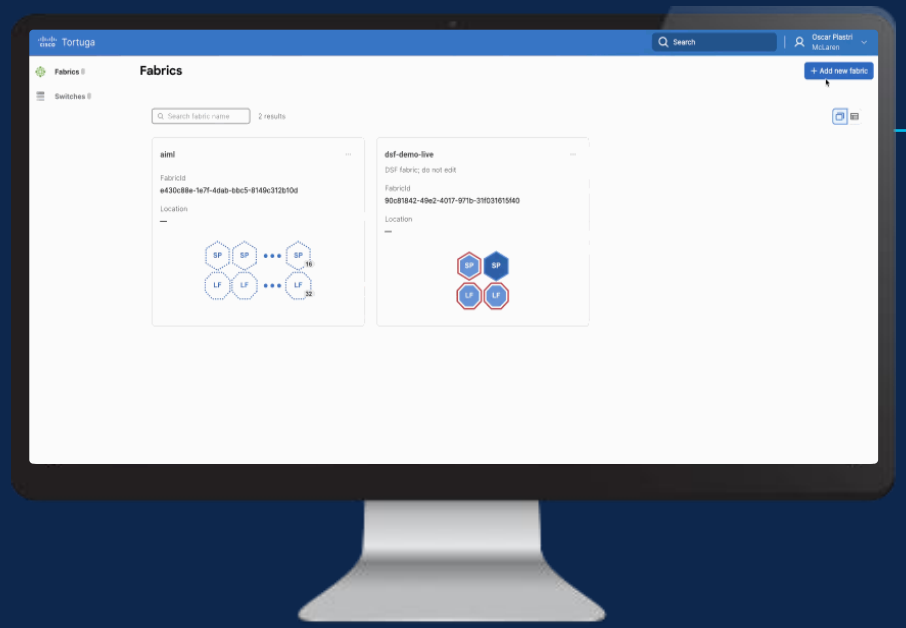
- Cisco UCS servers with VAST software
- Boots from cloud
- Full visibility and control from the cloud

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

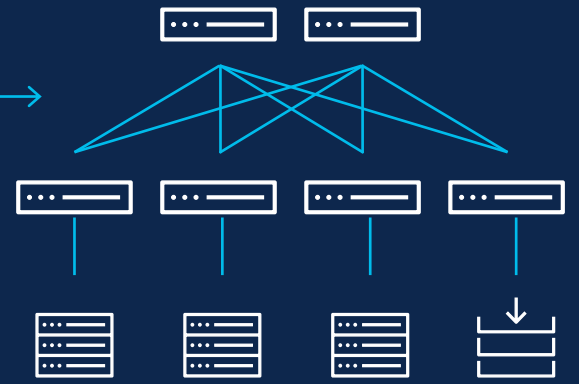
# How it works

API

Plan, deploy, manage




Cisco 6000 series switches



Cisco
UCS C885 M8
servers

Cisco UCS C225 M8
VAST
storage servers

## Plug-and-Play DC Fabrics

- Self discovery / standards-based
- Always-on telemetry
- Assertion-based monitoring

Purpose-built for predictable outcomes optimized for ease of use

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

# Design to BOM

Integrated design to BOM feature
reduces errors and shortens transaction time



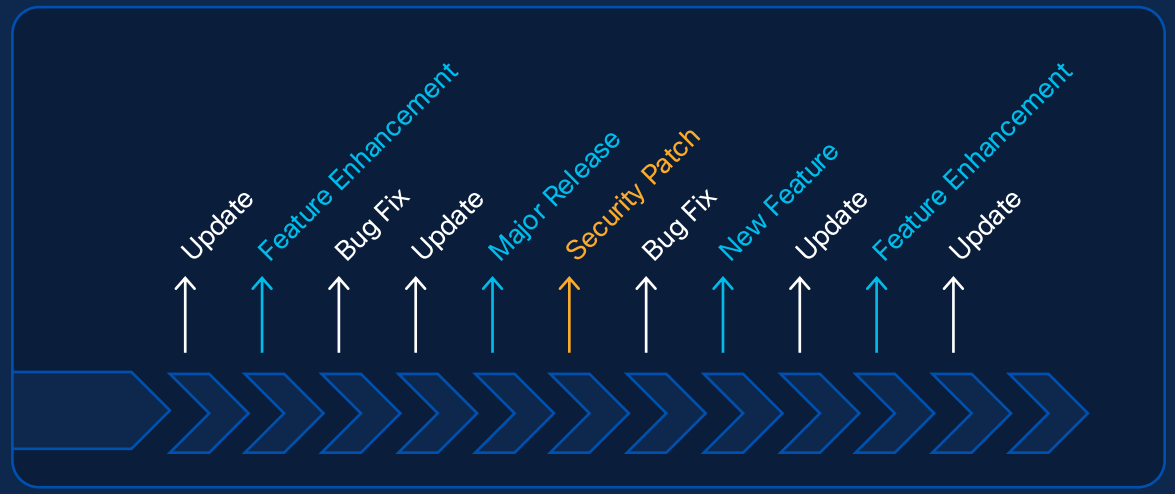

© 2025 Cisco and/or its affiliates. All rights reserved.

# Software lifecycle management

## Cloud SaaS controller

### Continuous delivery model: Always up-to-date

- Continuous delivery of new features and software updates to the production cloud service
- No user testing or software maintenance required




## On-prem compute, switch, storage software

### Cloud-delivered software upgrades: User-driven update schedule

- Schedule firmware updates
- Software rollback support
- Intelligent sequencing of fabric upgrades



# Flexible architectures

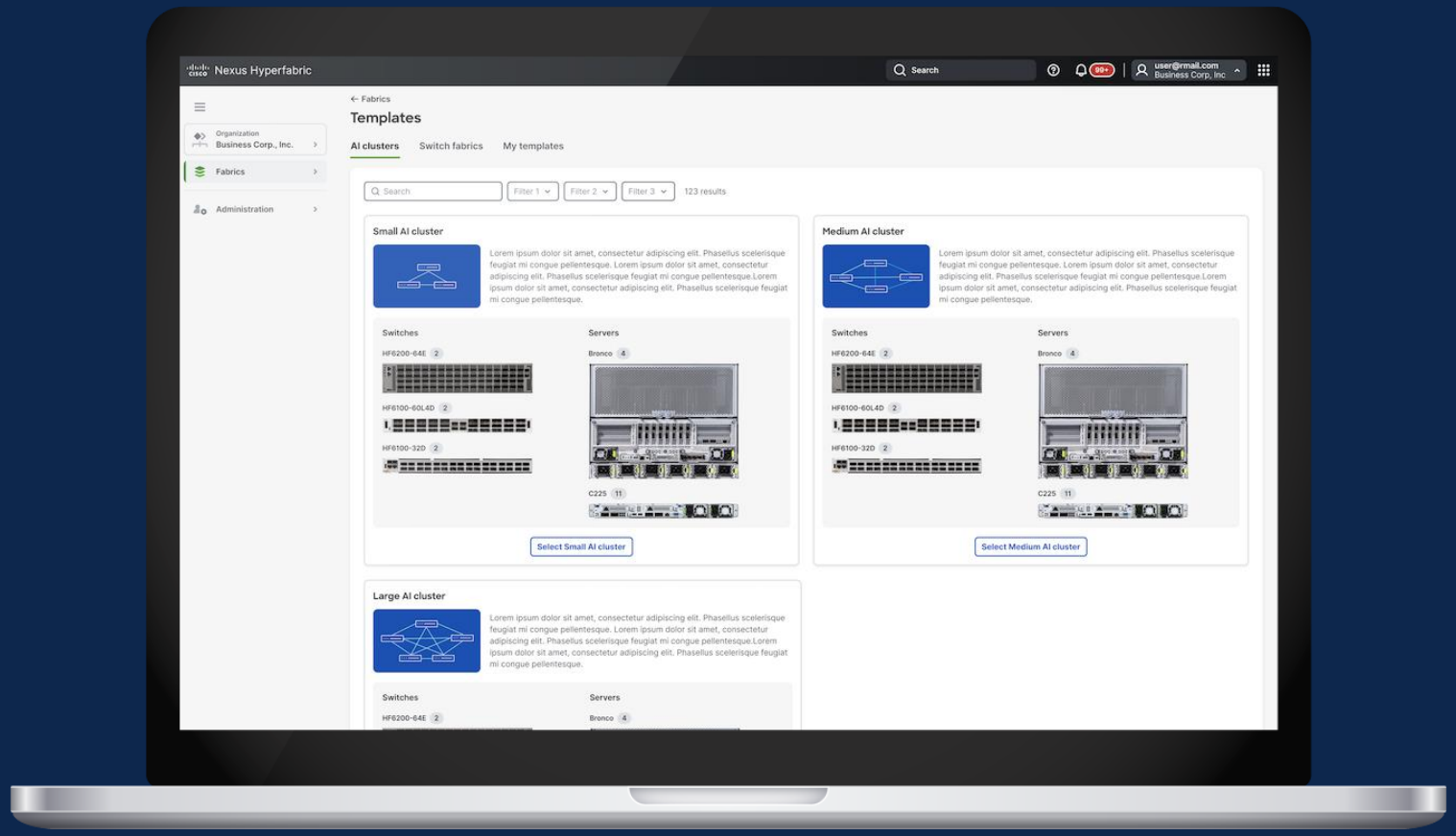
Choose the size your business needs with elasticity for future

Small AI Cluster

Medium AI Cluster

Large AI Cluster

Option to independently scale Network, Compute and Storage as per your need




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

# Cisco UCS GPU servers



## UCS-C885A-M8-CN1

- 8U, 8 NVIDIA H200 GPUs with NVLink
- 2 AMD CPU, 8 E/W B3140H 1x400G, 2 N/W 3240 2X400G (BF3)

Dense GPU server with high-speed interconnect for demanding AI workloads

High performance, high clock speed CPUs for training/learning operations

Cloud managed Intersight, strong NVIDIA partnership and TAC & Customer support

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

# Cisco UCS storage servers



## UCSC-HFA-VAST-1P

- 11 Servers with option to expand w VAST storage sw
- Rack servers with 15.8TB drives w >1PB storage
- 2 BF3 DPUs

High capacity, High performance, low cost, extensible  NVME solid state drives cloud managed Intersight, strong VAST partnership, and TAC & customer support

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

# See what's slowing your AI jobs

AI job observability in Nexus One

Cisco Cloud Control

## Follow the job across the stack

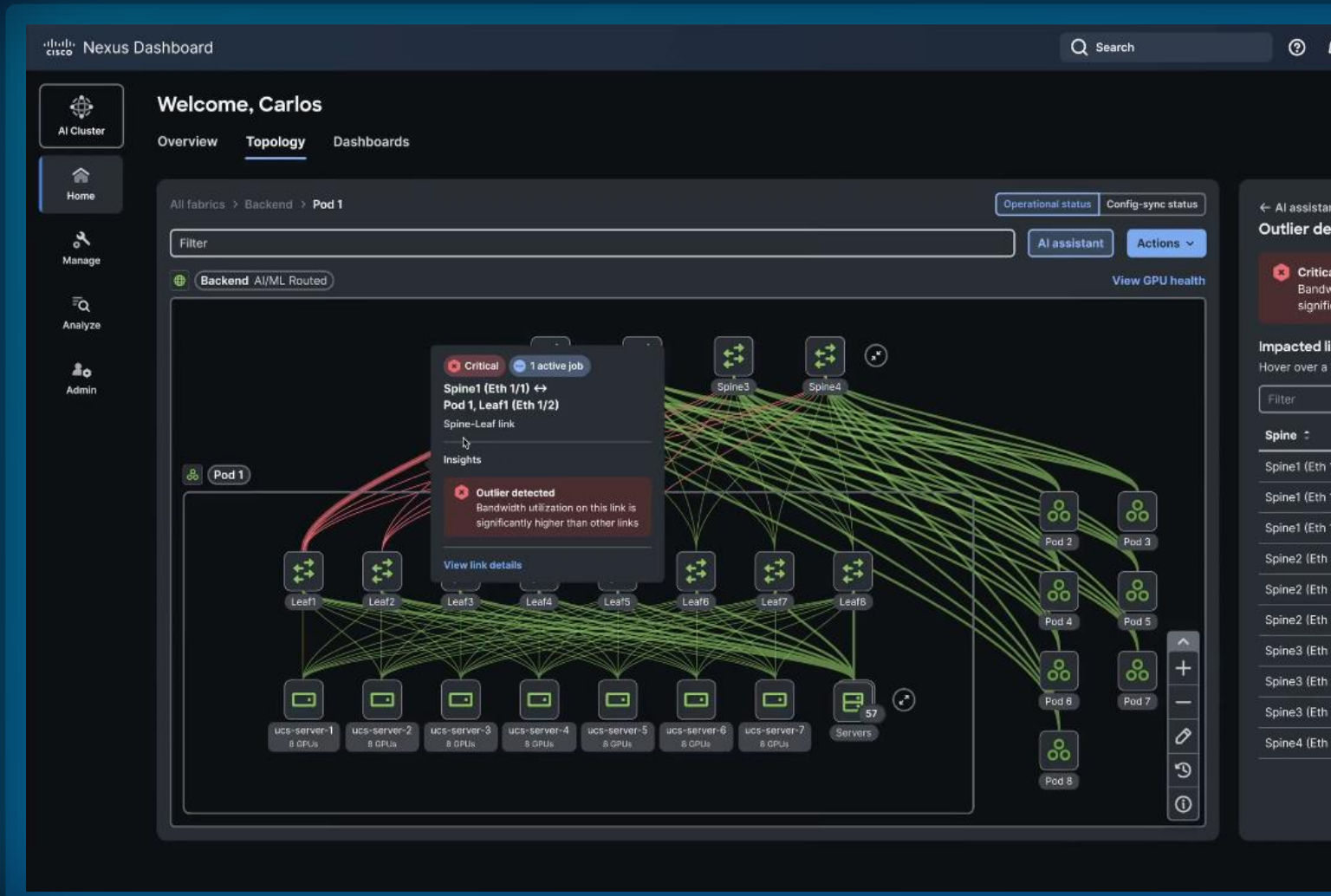
Connect AI Jobs to networks,
NICs, and GPUs

## Pinpoint performance bottlenecks

Correlate throughput, latency,
and GPU utilization.

## Resolve issues faster

Detect anomalies and guide
root-cause analysis




© 2026 Cisco and/or its affiliates. All rights reserved.

# Cisco AI-Ready Data Centers

Full-stack AI infrastructure, engineered as one system

Distributed Security fused into every layer

AgenticOps aligning speed with precision

© 2026 Cisco and/or its affiliates. All rights reserved.

We live in the post-Mythos world. **Every assumption we've had about security & infrastructure must change.**

© 2026 Cisco and/or its affiliates. All rights reserved.

# Secure your networks today with security fused into the network



Resilient infrastructure



Identity



Segmentation & steering



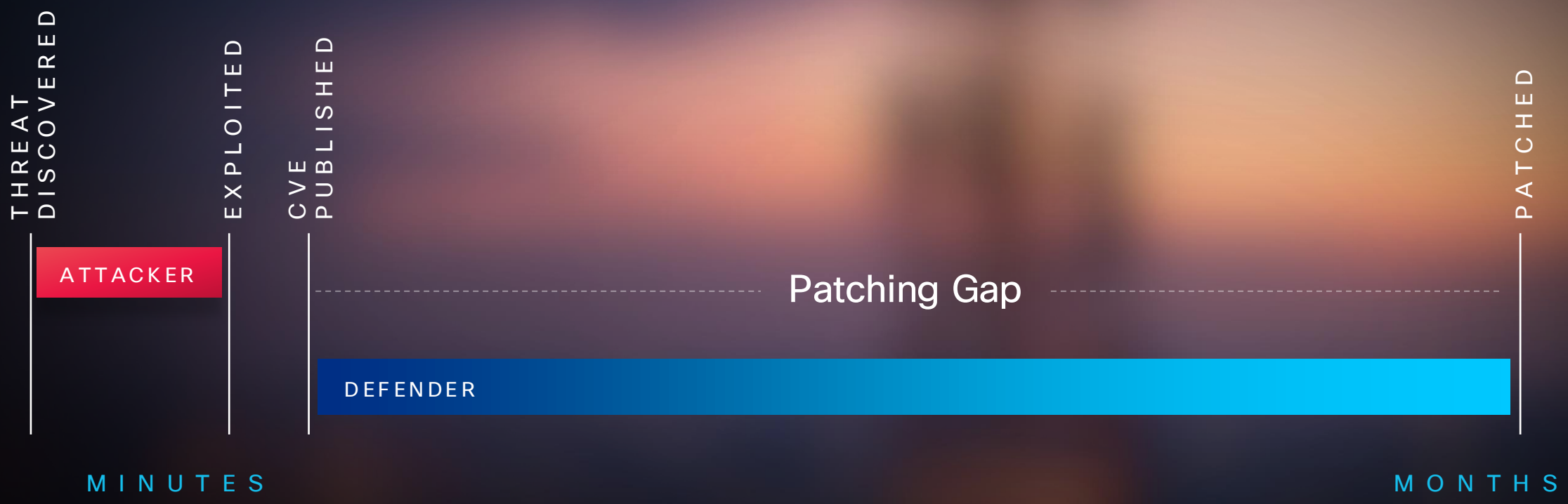
Protection



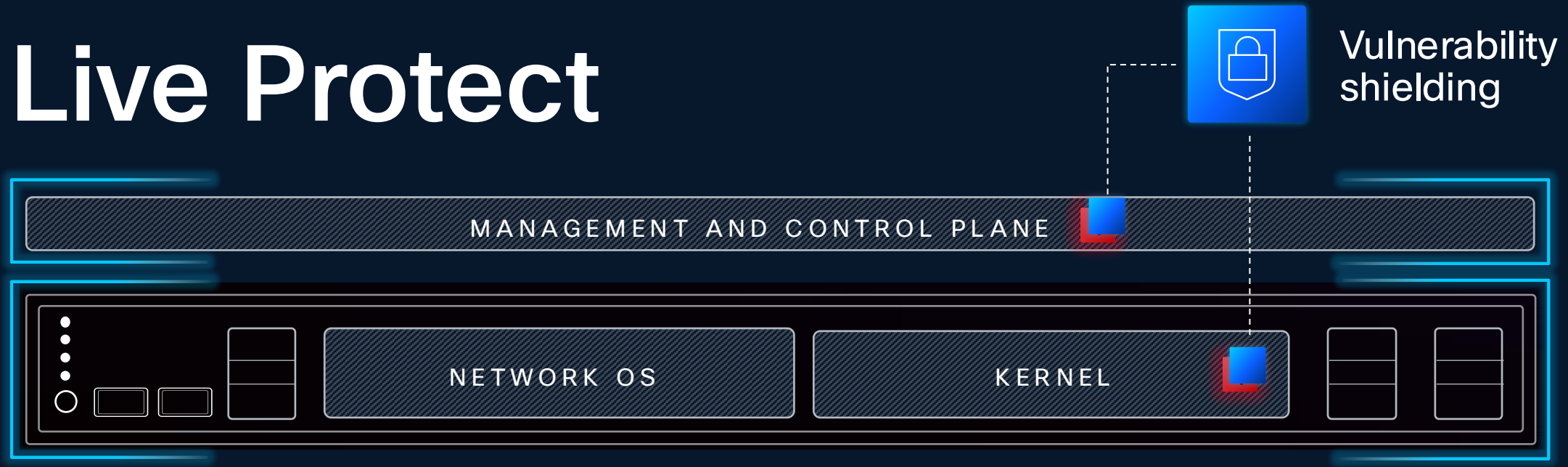
Visibility & insights

© 2026 Cisco and/or its affiliates. All rights reserved.

# AI-speed attacks compress the patching window. Minutes to exploit. Months to patch.



# Live Protect




EXTENDING ACROSS CISCO NETWORKING AND SECURITY

Apply policy to live system

Protect now, patch in planned window

No downtime or switch reboot

Continuous re-validation

© 2026 Cisco and/or its affiliates. All rights reserved.

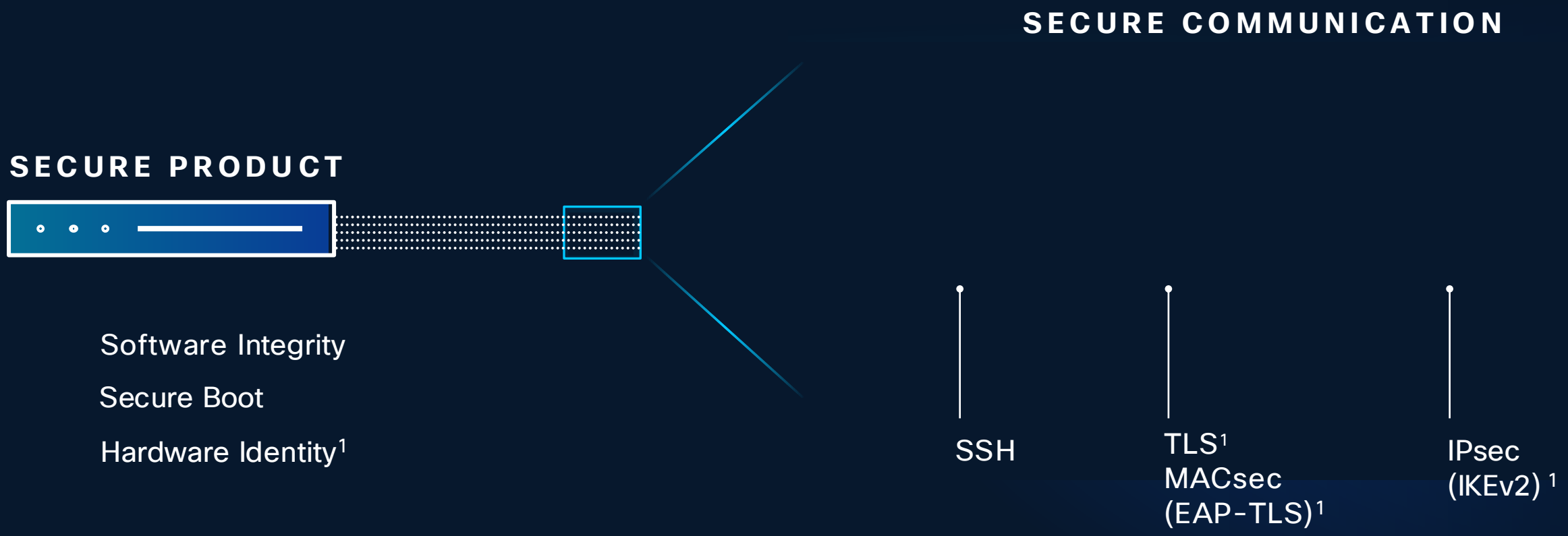
# Quantum-safe data center networks

NEW Q4 CY26

Protect your data from Harvest Now, Decrypt Later

## Post-Quantum Cryptography (PQC) Resilience in Cisco N9000 Series Switches

Future proof investments | Secure sensitive data




[1] Support planned for NX-OS and ACI CY-2027 releases

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

# Cisco AI Ready Data Center

Full-stack AI infrastructure, engineered as one system

Distributed Security fused into every layer

AgenticOps aligning speed with precision

© 2026 Cisco and/or its affiliates. All rights reserved.

# AgenticOps

## Cisco Cloud Control

AI Assistant | AI Canvas | Studio | Cisco dashboards | Third party & APIs



Cross-domain telemetry



Purpose-built models



Trusted agents

© 2026 Cisco and/or its affiliates. All rights reserved.

Cisco Cloud Control

# One console for AgenticOps

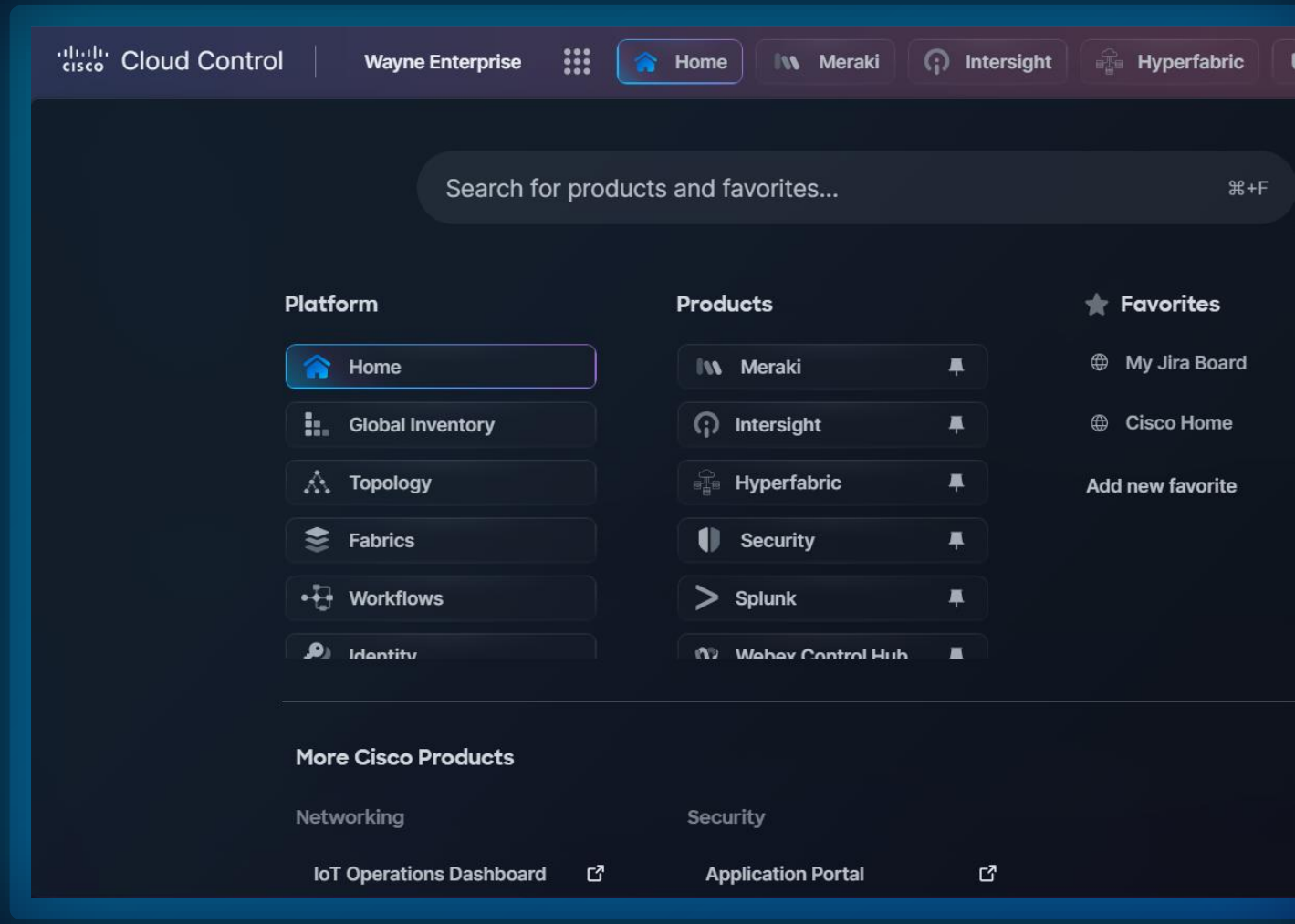
## One login
Access your Cisco products

## One inventory and topology
See assets and connections in real time

## One view for alerts
Correlate alerts across products

## One assistant
Get answers and take action with AI Assistant

## One agentic workspace
Investigate and resolve together in AI Canvas.



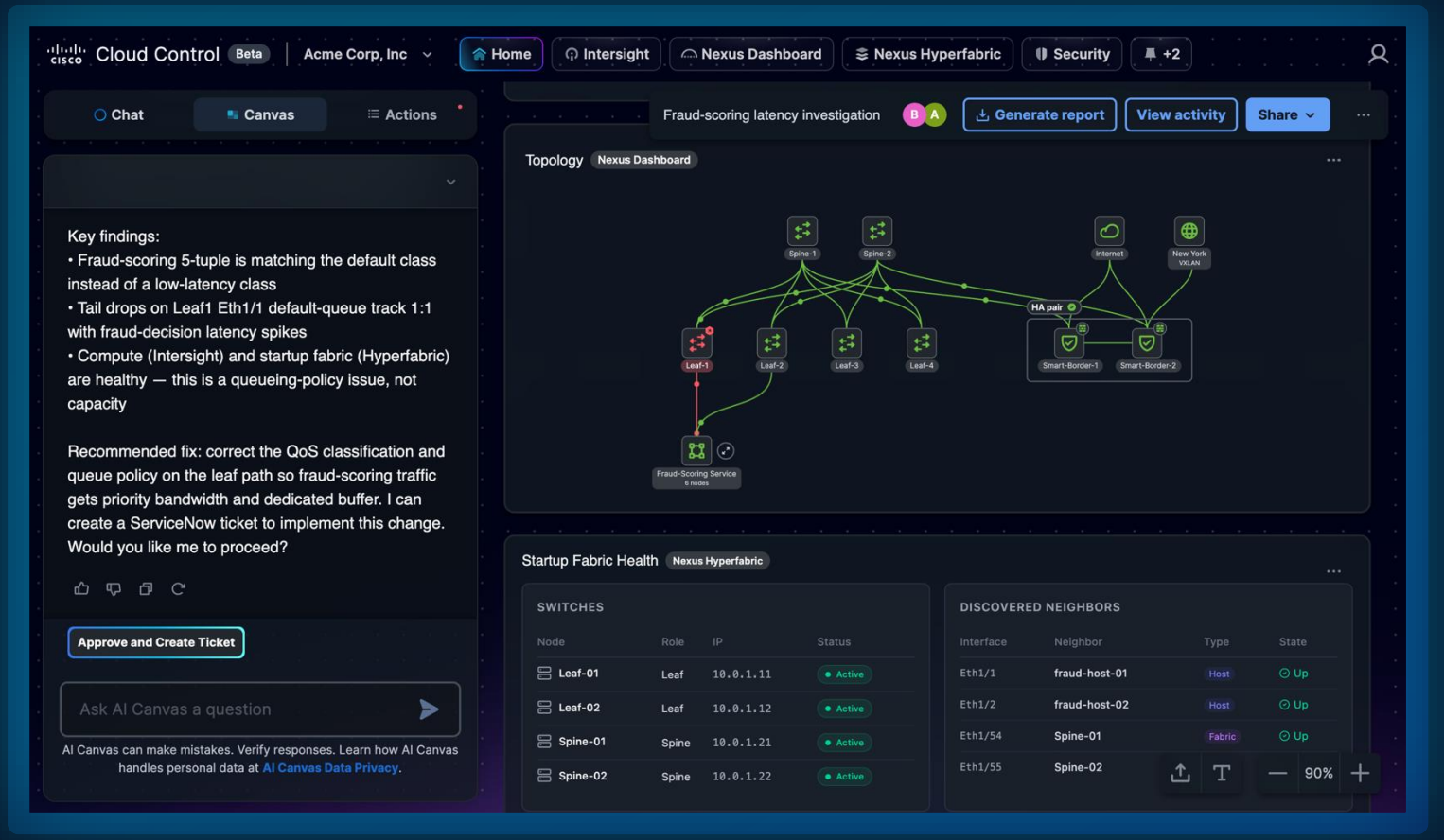

© 2026 Cisco and/or its affiliates. All rights reserved.

NEW

Cisco Cloud Control

# Support for Data Center domains

Leverage cross-domain visibility and telemetry for autonomy — from troubleshooting to continuous management.




## Skills powering deep reasoning agents

### Nexus One

- Fabric Discovery & Topology
- Health & Connectivity
- Anomalies & advisories
- AI Workload Observability
- Network Analytics

### Intersight

- Compute, networking & storage Inventory
- Health and connectivity context
- Change Correlation (audit-record)

### Security Cloud Control

- VPN tunnel health
- Traffic flow & routing context
- Guided remediation

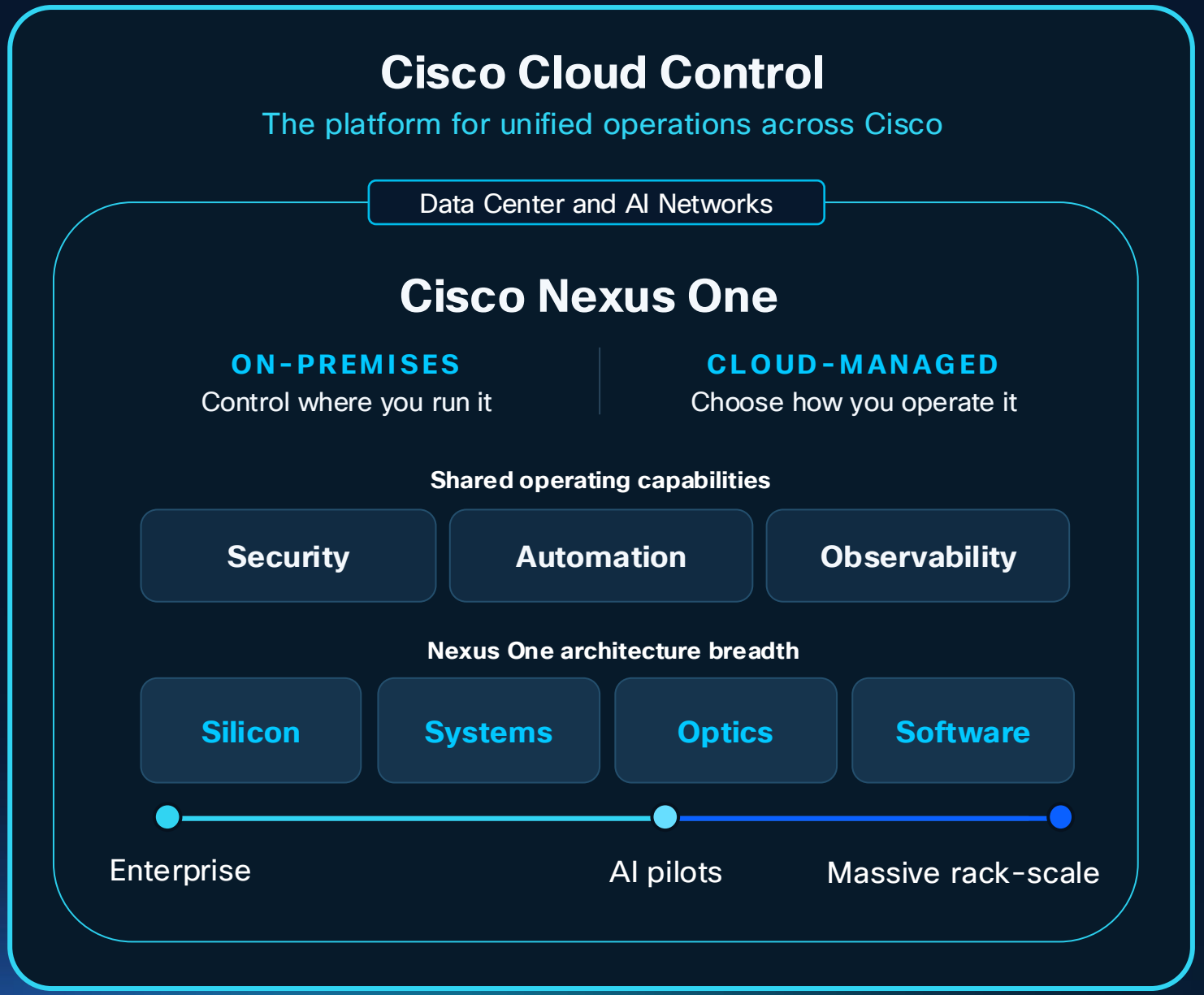
© 2026 Cisco and/or its affiliates. All rights reserved.

CISCO NEXUS ONE

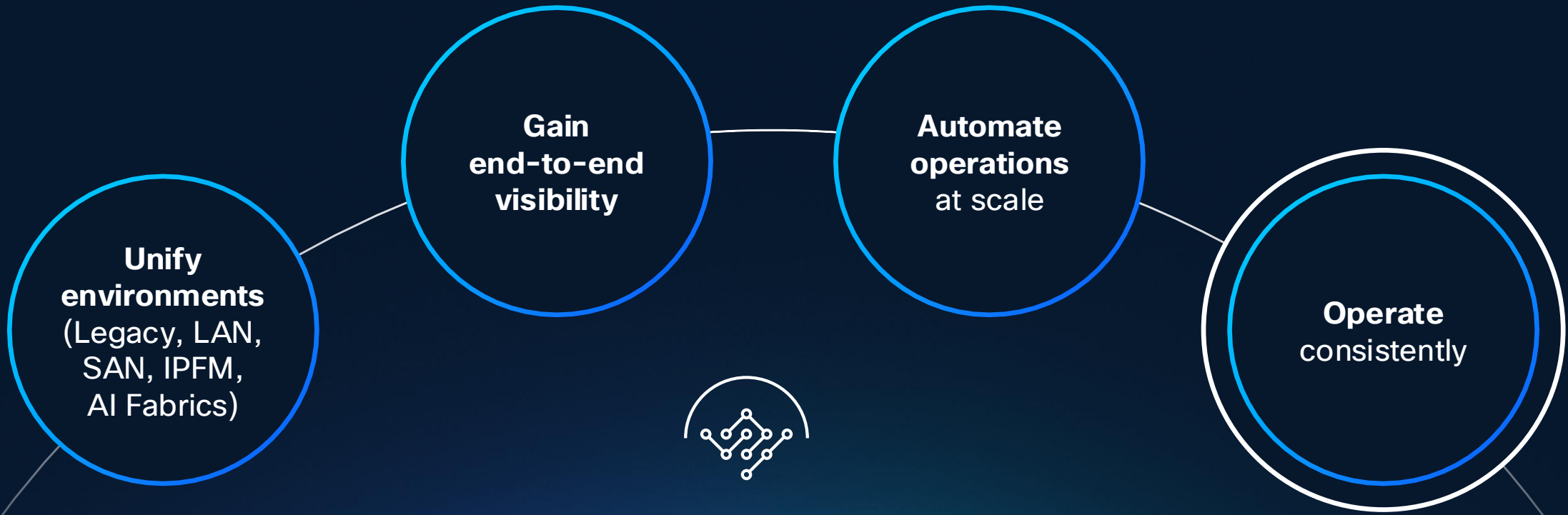
# Your data center network evolves

# Your investments carry forward

# Your choices stay open




© 2026 Cisco and/or its affiliates. All rights reserved.




# Cisco Cloud Control with Nexus Dashboard

On-Premises Cisco Nexus One Controller

© 2026 Cisco and/or its affiliates. All rights reserved.

# Cisco Secure AI Factory with NVIDIA

**A modular reference design** that combines high-performance infrastructure with full-stack security and observability



Accelerate AI Apps Delivery

Training → Optimization → Inferencing

- AI Software w/ NVIDIA AI Enterprise
- Kubernetes Platform
- Cisco AI Networking
- Cisco Compute w/ NVIDIA Accelerated Computing
- Partner Storage

Cisco Security

Splunk Observability

Data center (Core) ⟷ Distributed AI Infrastructure ⟷ Edge Compute

© 2026 Cisco and/or its affiliates. All rights reserved.

# Secure AI Factory with NVIDIA expands to rack-scale

NEW Q4 CY26

Cisco + Supermicro deliver high density liquid cooled fabric and compute

**RACK-SCALE COMPUTE**
**HGX and MGX systems**




Supermicro NVL72 rack scale system with
Cisco Silicon One + NVIDIA Spectrum-X

## One full stack, now extended to rack-scale

### Compute
- Supermicro systems with NVIDIA Accelerated Computing

### Fabric
- NCPRA compliant
- Rack to fabric liquid cooling

### Operate
- Cisco Cloud Control
- Cisco Validated Infrastructure Service (CVIS) validation

© 2026 Cisco and/or its affiliates. All rights reserved.

# Ecosystem

## TECHNOLOGY PARTNERS

NVIDIA
intel
VAST
NUTANIX
Red Hat
AMD
NetApp
ddn
PURESTORAGE
HashiCorp
HITACHI
Qumulo

## ISV (Independent Software Vendor) PARTNERS

WaitTime
axxonsoft
Medical Informatics
Quividi
Rockwell Automation
AI4CMR
EPIC IO
PreciTaste
Red Hat AI Inference Server
AVEVA Schneider Electric
AUTONOMA
NVIDIA METROPOLIS
intel OpenVINO
Arcee AI
VAIDIO
ONSTAK

© 2026 Cisco and/or its affiliates. All rights reserved.

# From Connect to action: Next steps

Want to know more about Cisco End-to-End AI Networking?

Explore additional resources on Cisco AI Networking Webpage:

cisco.com/go/ai-networking

Trials and Demos.
Learn how our products and technologies can support your business and technical needs:
Discover Cloud Control

Talk to a Cisco expert.
We're here to help you make informed decisions every step of the way:
Contact Cisco

© 2026 Cisco and/or its affiliates. All rights reserved.