CISCO Connect

Full Stack AI-Ready Data Center

Built for AI, Ready for What's Next

Vinay Thamatam
AI solutions Engineer



October 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



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

Artificial intelligence outcomes span every industry

Manufacturing

Predictive maintenance

Quality control

Demand forecasting



Public sector

Smart cities

Security and safety

Services improvement





Retail

Personalization

Inventory optimization

Sales forecasting



Financial services

Fraud detection

Risk assessment

Trading



Healthcare

Diagnosis

Drive-thru optimization

Patient support

Education

Learning & teaching experiences

Smart & secure facilities



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

AI adoption is high.
Scaling to production is still hard.

88%
Enterprises use AI today

<40%
Have scaled beyond pilots

70–80%
AI projects still fail

What breaks when AI becomes a production data center workload?

Infrastructure
GPU, storage, power, cooling, and fabric

Platform & data path
Kubernetes, pipelines, placement, dependencies

Security
Trust zones, policy, risk and governance

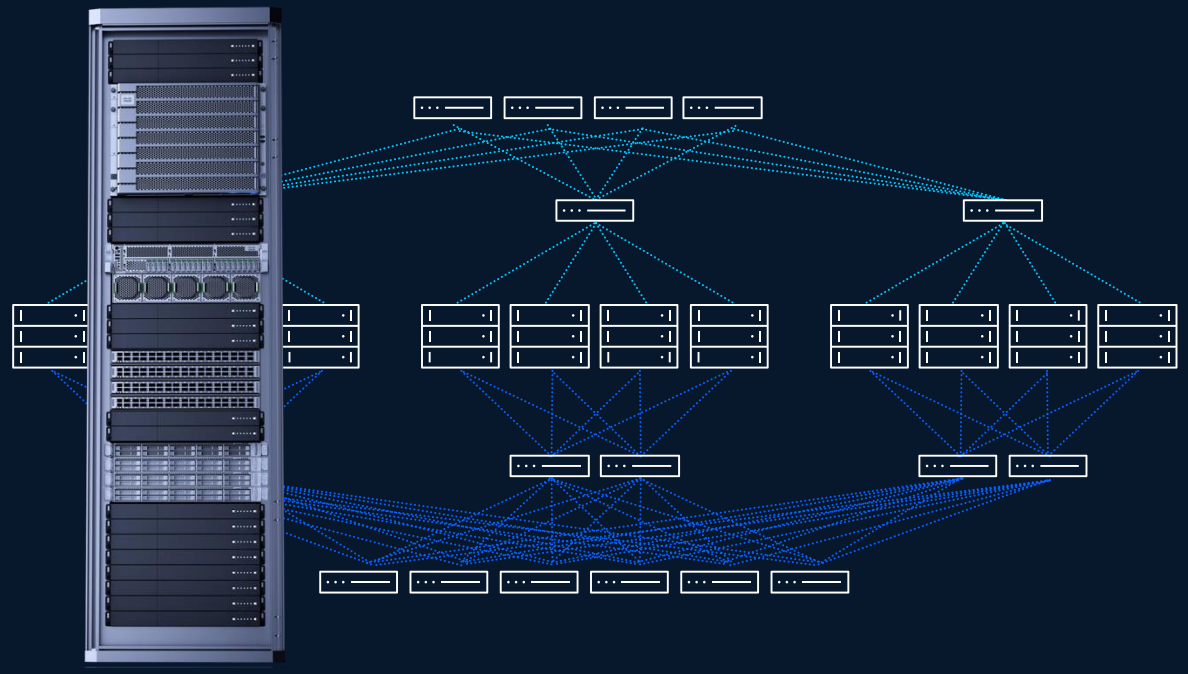
Operations
Visibility, ownership, and lifecycle

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

Sources: Deloitte State of AI 2026, McKinsey, NVIDIA State of AI 2026, Cloudera/HBR 2026

Full-stack design changes how production AI gets built

...so production AI is designed as a system, not assembled as parts



Validated architecture

Capacity, fabric, security, operations

Faster integration

Fewer handoffs, fewer surprises

Day-one operations

Visibility, lifecycle, change control

Trust & accountability

Policy, ownership, governance

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

An end-to-end approach to AI



© 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.

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.

cisco

The network isn’t just connectivity.
It’s what lets models learn and think.

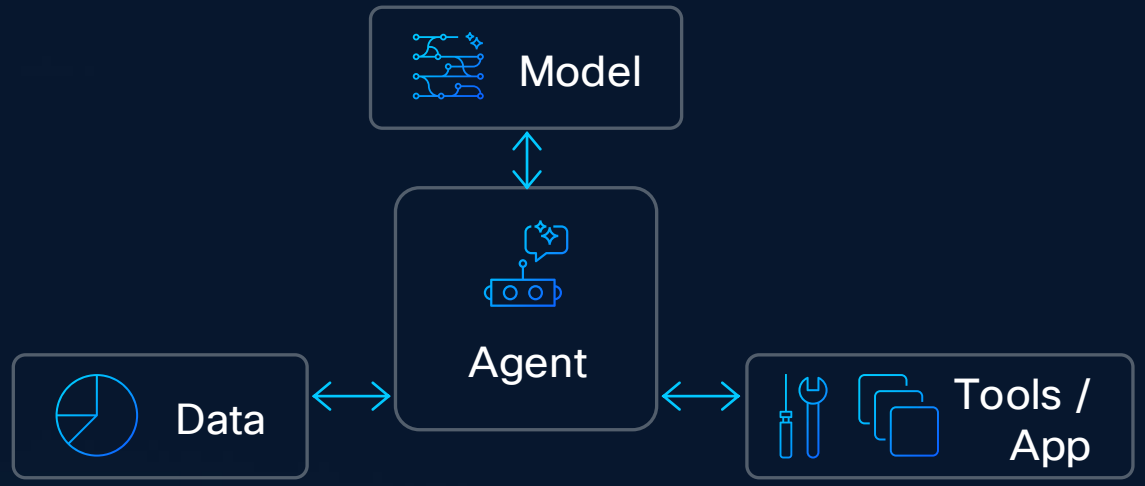
Agents add dynamic performance loops to applications

Kubernetes distributed apps. AI adds dynamic loops across models, data, tools, and accelerators.

Modern applications



Agentic applications



Why AI fabric matters

More flows, changing paths
↓
Adapt to dynamic traffic

More data per answer
↓
Deliver more bandwidth

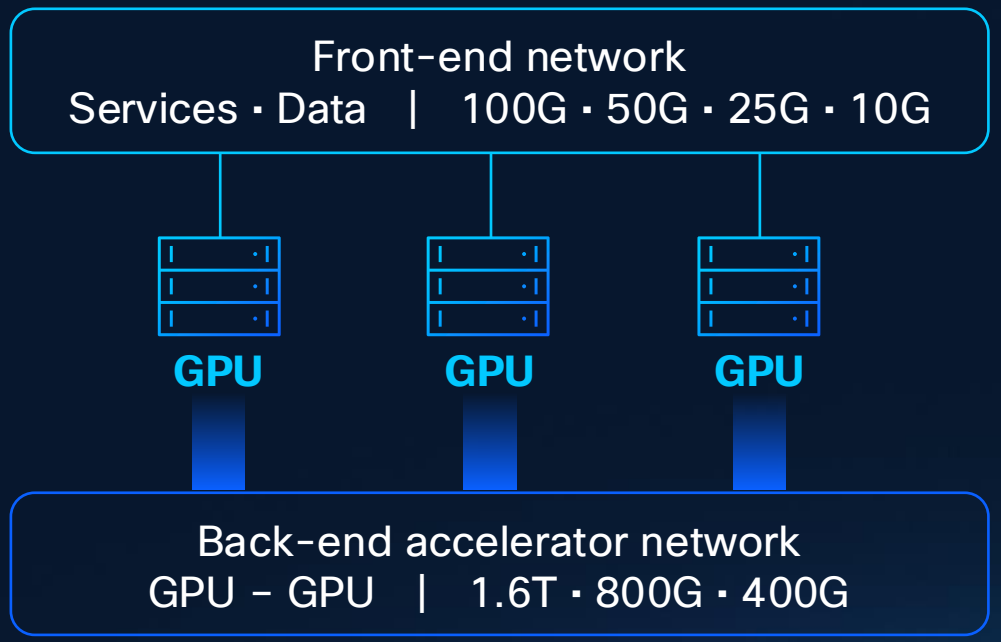
Repeated network round trips
↓
Keep latency low and predictable

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

Cisco helps teams move AI work–and run the infrastructure behind it

Move AI work

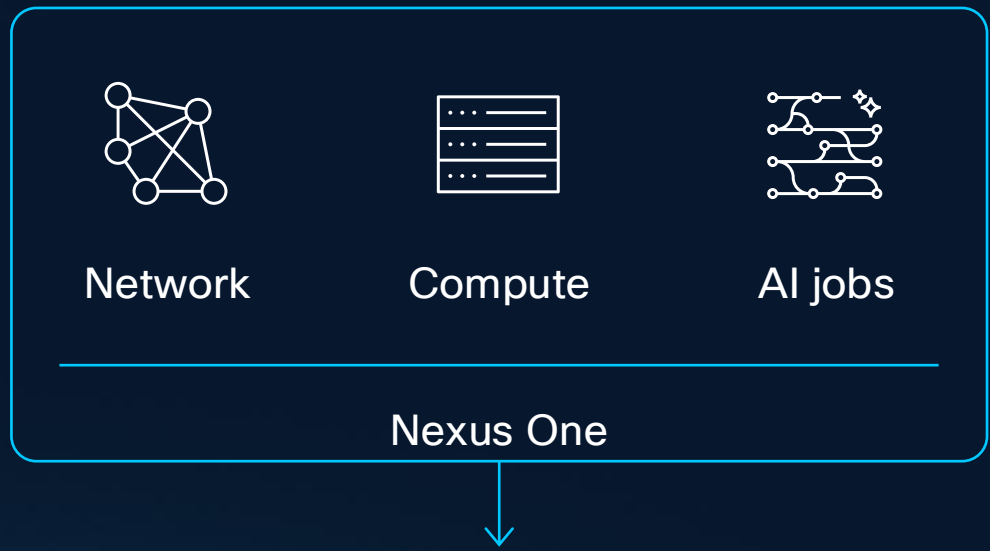
Connect AI services and data



Keep accelerators doing useful work

Run as one system

See the impact. Take action.

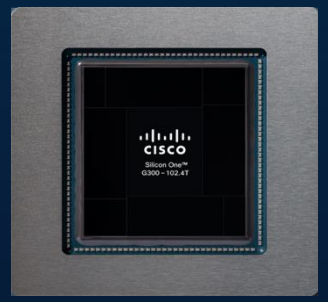


Operate with system-wide context

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

Recent AI networking innovations driving AI outcomes

SILICON



High-Performance AI Scale

Silicon One G300
102.4 Tbps

Built for performance, efficiency, and scale

SYSTEMS



Predictable Scale-Out Performance

Cisco N9000 | Cisco 8000
Cisco N9100 with NVIDIA Spectrum-X

Designed for consistent throughput at massive scale

OPTICS



High-speed AI Fabric Connectivity

1.6T and 800G optics for dense AI clusters

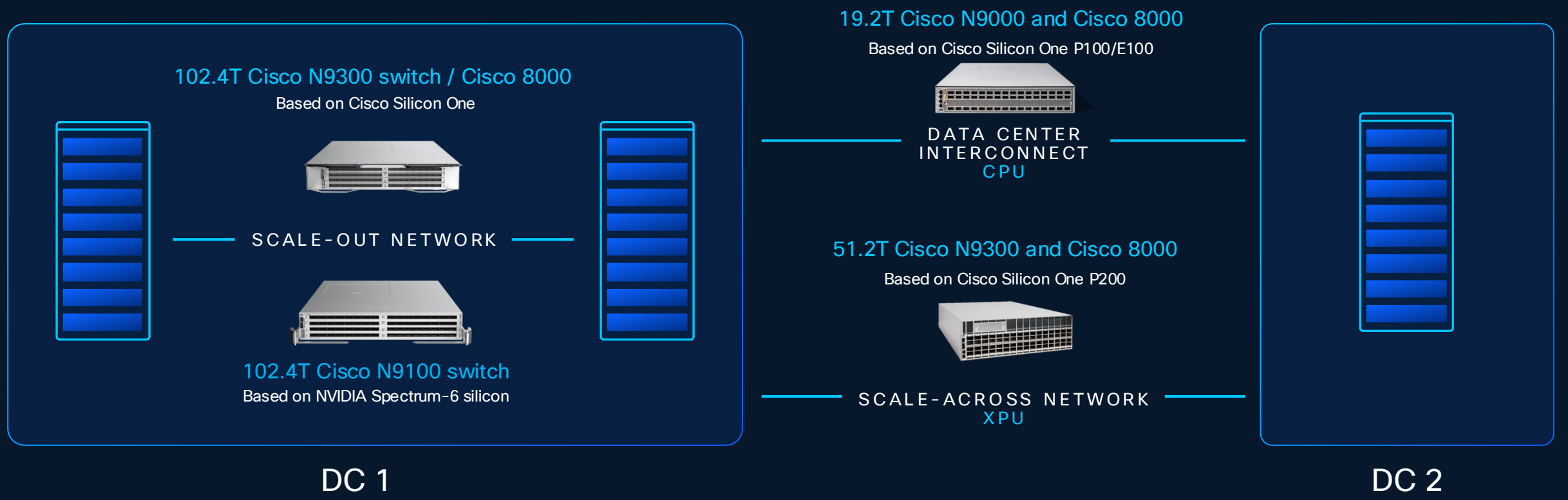
Built for signal integrity, low latency, low power, and uptime

SCALE-OUT | HYPERSCALE | ENTERPRISE | NEOCLOUD | SOVEREIGN CLOUD | SERVICE PROVIDERS

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

Scale-out, DCI and scale-across networking

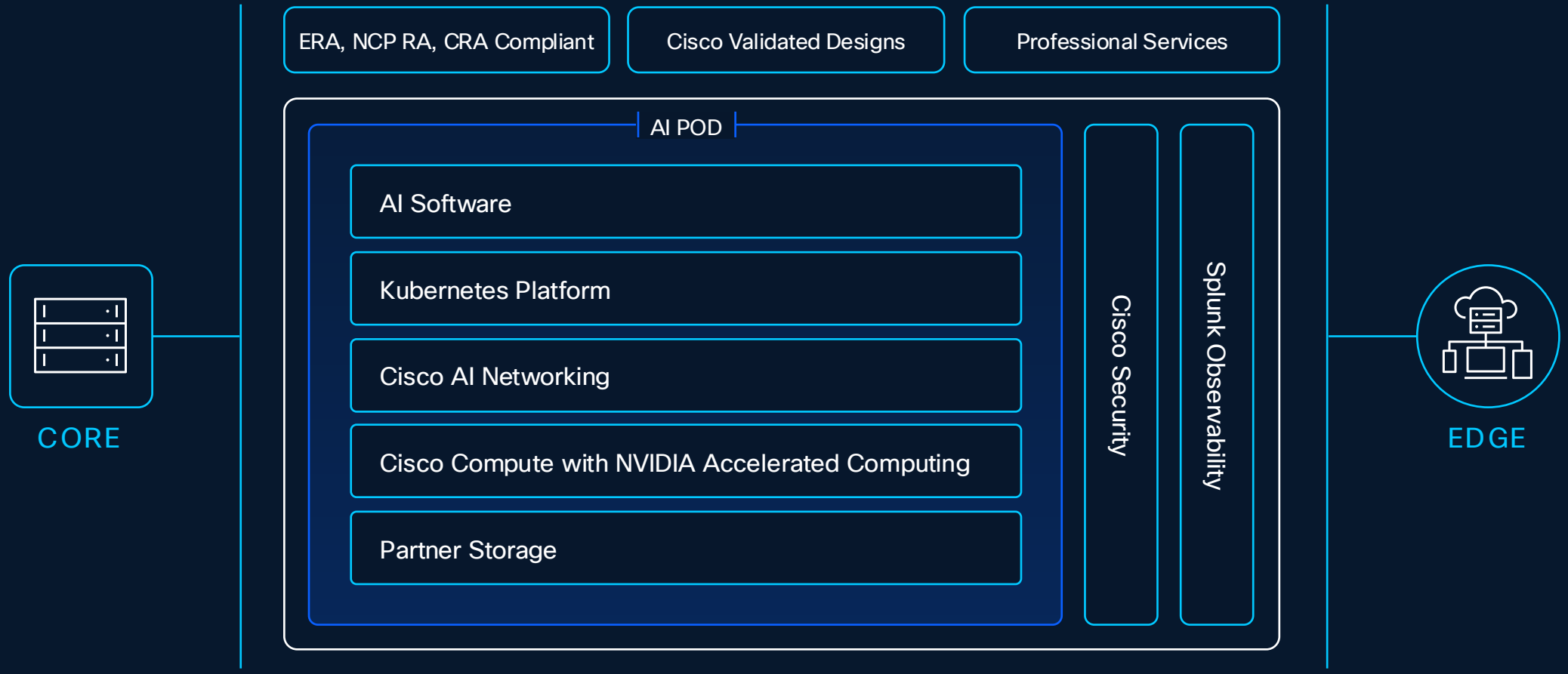
Validated architecture choices for AI scaling



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

Cisco Secure AI Factory with NVIDIA

Co-engineered reference architecture with NVIDIA



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

Unified approach to enterprise and AI workloads

LARGE SCALE MODEL TRAINING, INFERENCE, AGENTIC AI

MODEL TRAINING

MODEL OPTIMIZATION, INFERENCE, AGENTIC AI

INFERENCE, AGENTIC AI

Rack-Scale Systems

NEW



Supermicro NVIDIA Vera Rubin NVL72

NEW



Supermicro NVIDIA Grace Blackwell 300 NVL72

Dense GPU Servers

NEW



Supermicro w/ NVIDIA HGX Rubin NVL8 liquid-cooled

NEW



Supermicro w/ NVIDIA HGX B300 NVL8 liquid-cooled

NEW



Supermicro w/ NVIDIA HGX B300 NVL8 air-cooled

PCIe GPU Servers

NEW



Supermicro MGX NVIDIA RTX PRO



Cisco UCS® C845A M8 NVIDIA RTX PRO & AMD Instinct MI



Cisco UCS® X-Series Modular Servers NVIDIA RTX PRO



Cisco UCS® C245 M8 NVIDIA RTX PRO



Cisco UCS® C240 M8 NVIDIA RTX PRO

Edge PCIe GPU Platform



Cisco Unified Edge NVIDIA RTX PRO

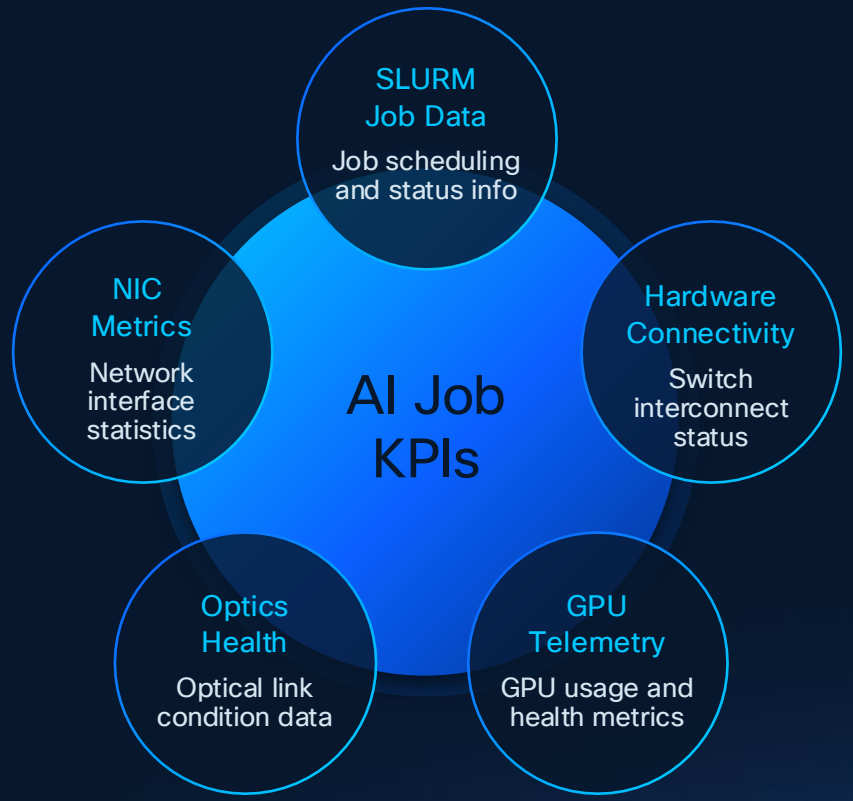
CORE — EDGE

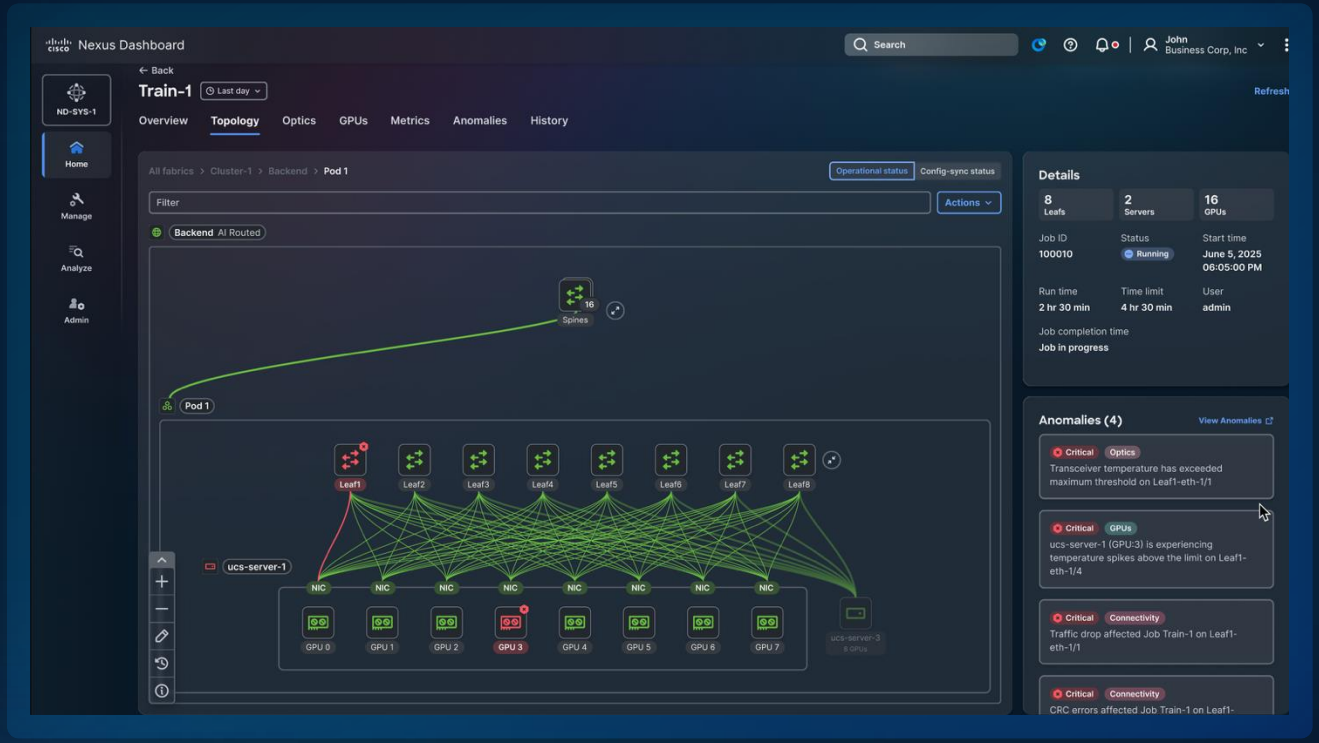
Cisco Cloud Control | Cisco Intersight – Simplified operations and fleet management

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

Job centric AI fabric visibility in Cisco Cloud Control with Cisco Nexus Dashboard

Correlate scheduler, NIC, GPU, optics, and fabric telemetry.

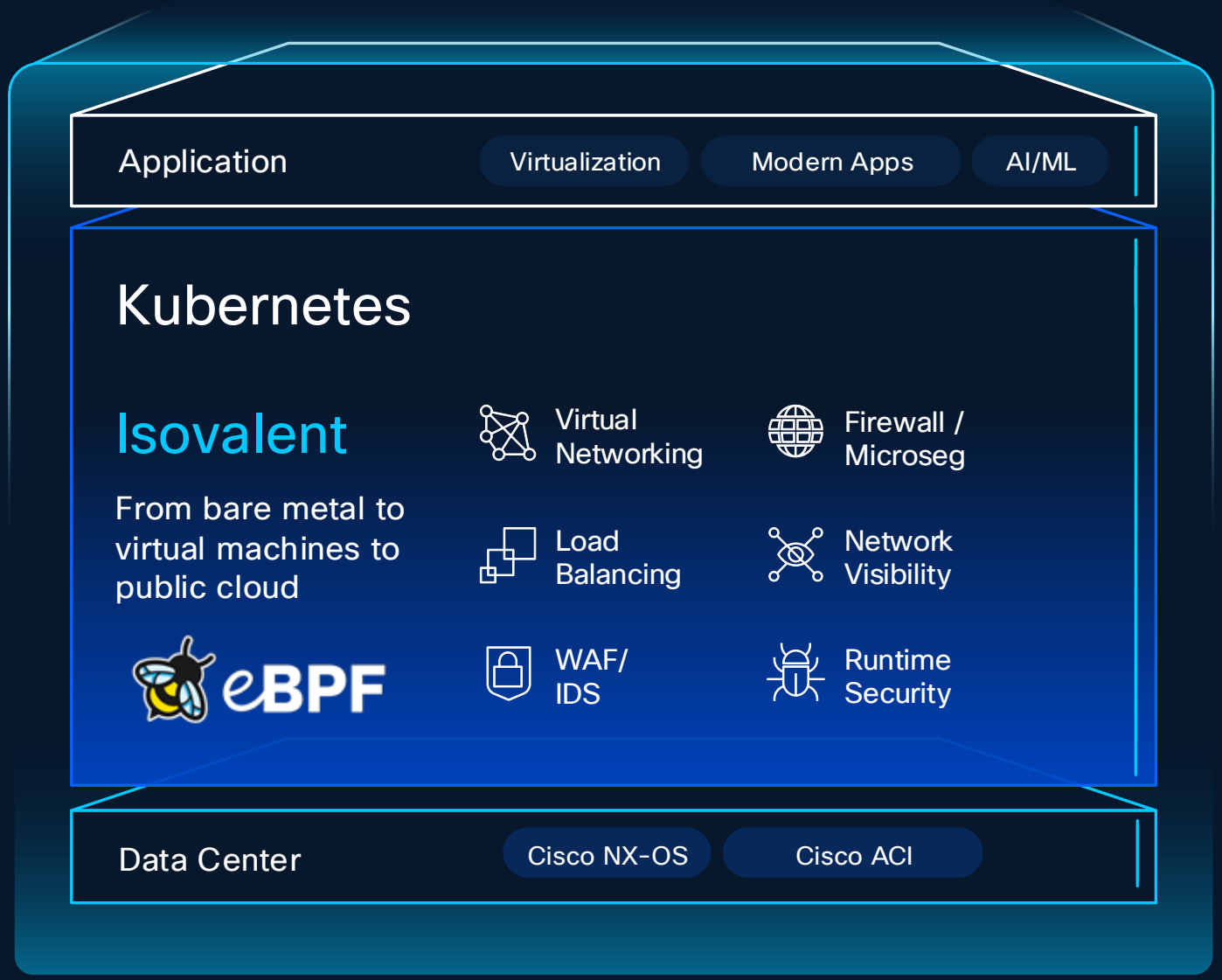




Detect bottlenecks before they affect AI job throughput.

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

Where does Isovalent operate?



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

Cisco AI-Ready Data Centers

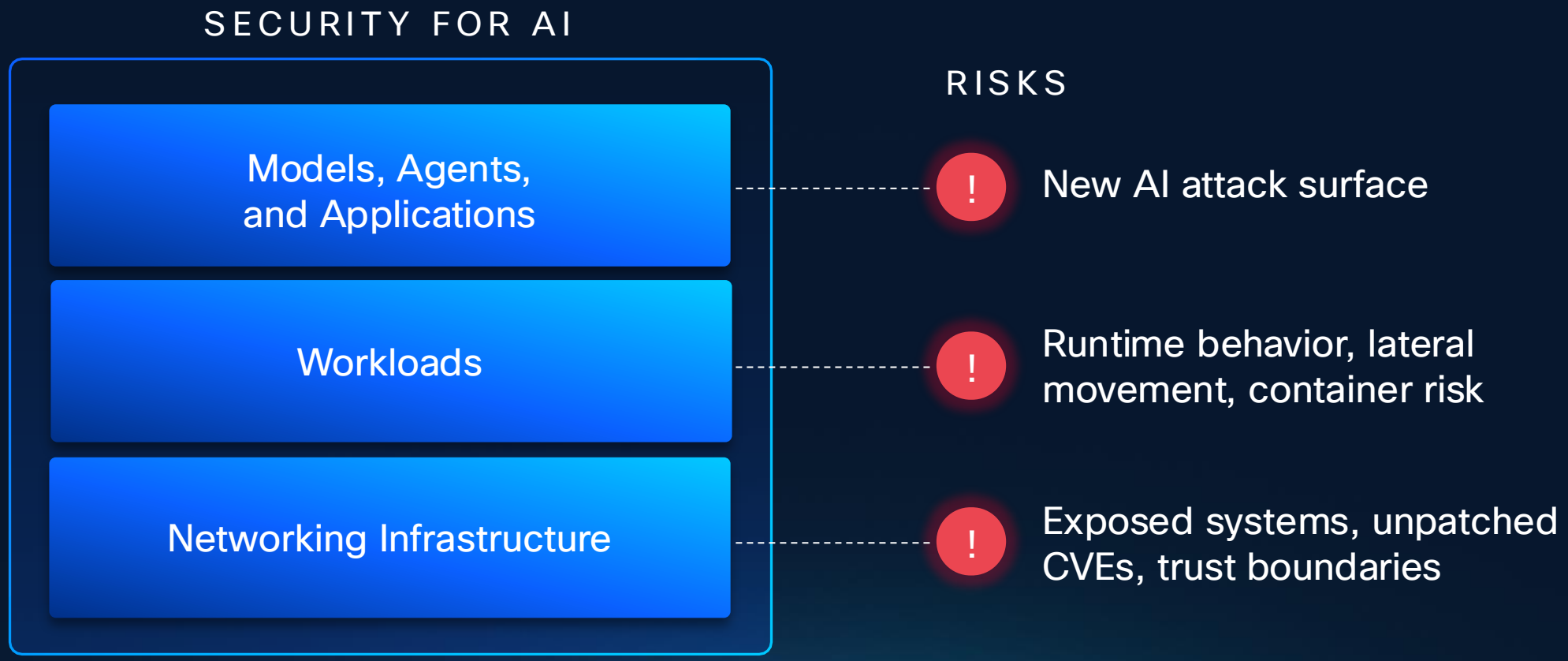
Full-stack AI infrastructure, engineered as one system

Distributed Security fused into every layer

AgenticOps aligning speed with precision

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

Secure the AI era: new stack, faster attackers



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

Frontier AI is changing the rules. Permanently.

Scale

1000s

of critical vulns identified
in weeks, not years

Chained exploits

3→1

medium CVEs chained into
one critical breach path

Shrinking window

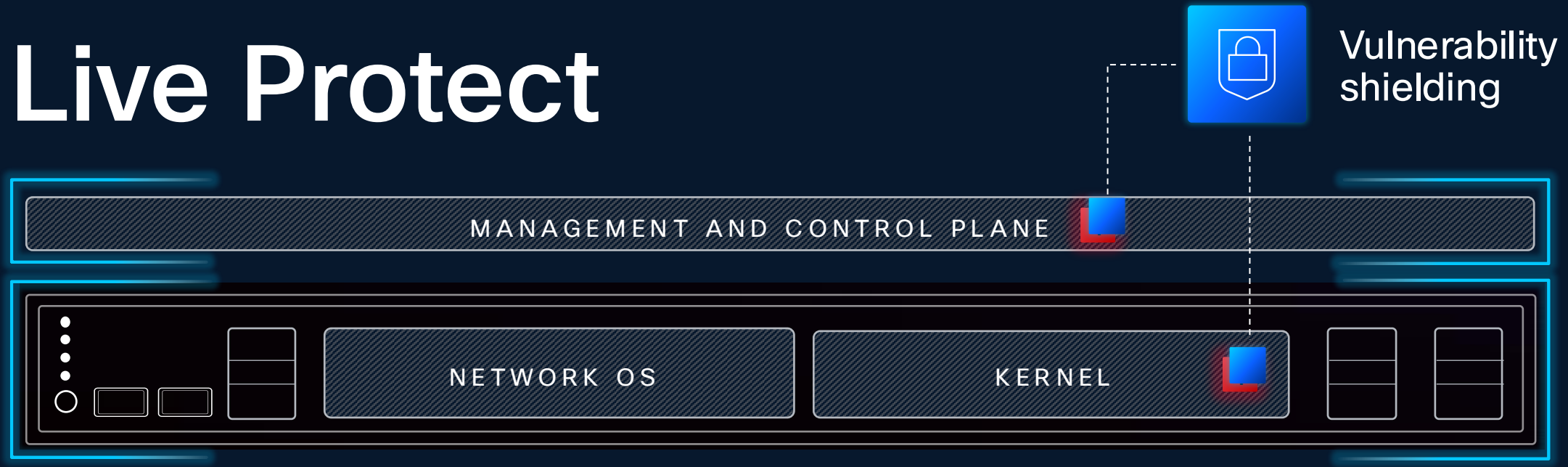
Hours

from patch release to
adversarial reverse-engineering

The imperative: Architect structural resilience that withstands the unknown.

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

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.

Cisco Hybrid Mesh Firewall

Distributed security at the speed of change

Define policy once, enforce everywhere



Cisco Firewalls

Physical | Virtual | Cloud | FWaaS



3rd Party Firewalls



NVIDIA Bluefield DPU Firewall on AI Servers



Smart Switches

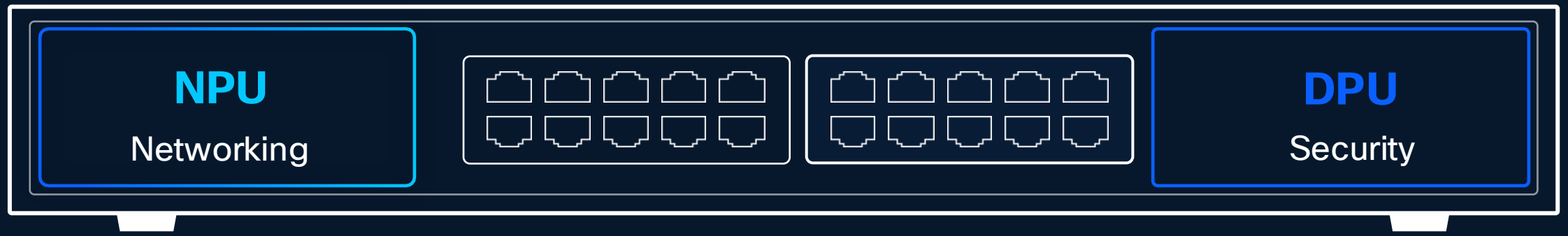


Workload Agents

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

Threat containment at massive scale built into your network

AVAILABLE NOW



CISCO N9300 SMART SWITCH & HYPERSHIELD

Hardware-accelerated **segmentation** on every port

Single appliance, dual personality

Greater visibility, fewer dropped packets

More easily scale security as traffic increases

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

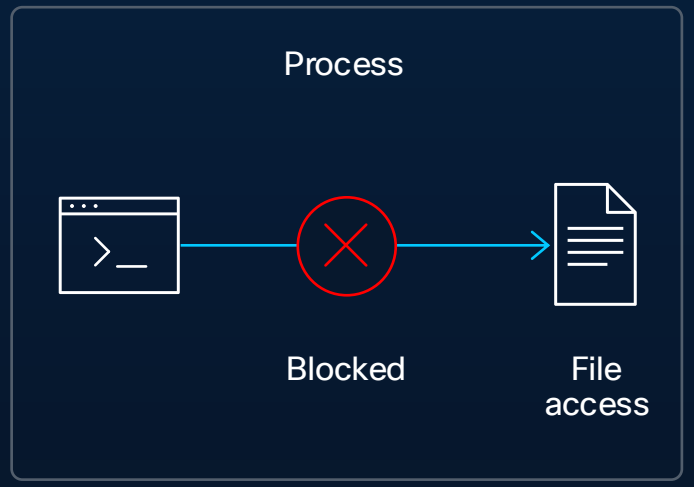
Cisco Isovalent

Stop threats from process to network

Process actions

Control what processes do

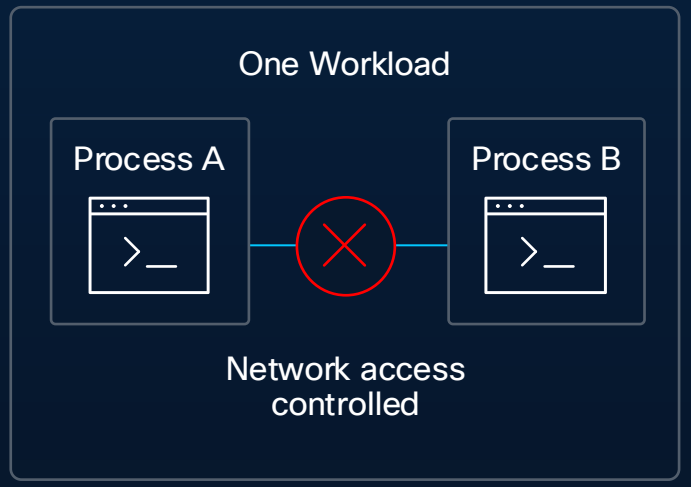
Enforce specific runtime actions



Process connections

Control traffic between processes

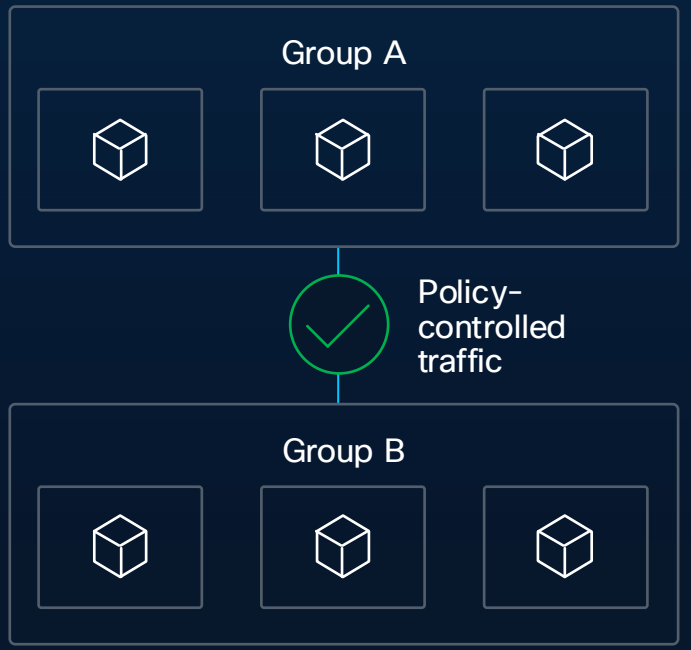
Process-aware network enforcement



Workload groups

Control traffic between groups

Segmentation across groups or zones



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

Cisco AI Defense

SECURING CUSTOM AGENTS AND AI APPS



Discover

New

Adaptive red teaming and guardrails



Detect

New

Protecting the full agent supply chain



Protect

New

Ecosystem to work with any cloud, any GPU

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

cisco

Cisco AI Ready Data Center

Full-stack AI infrastructure, engineered as one system

Distributed Security fused into every layer

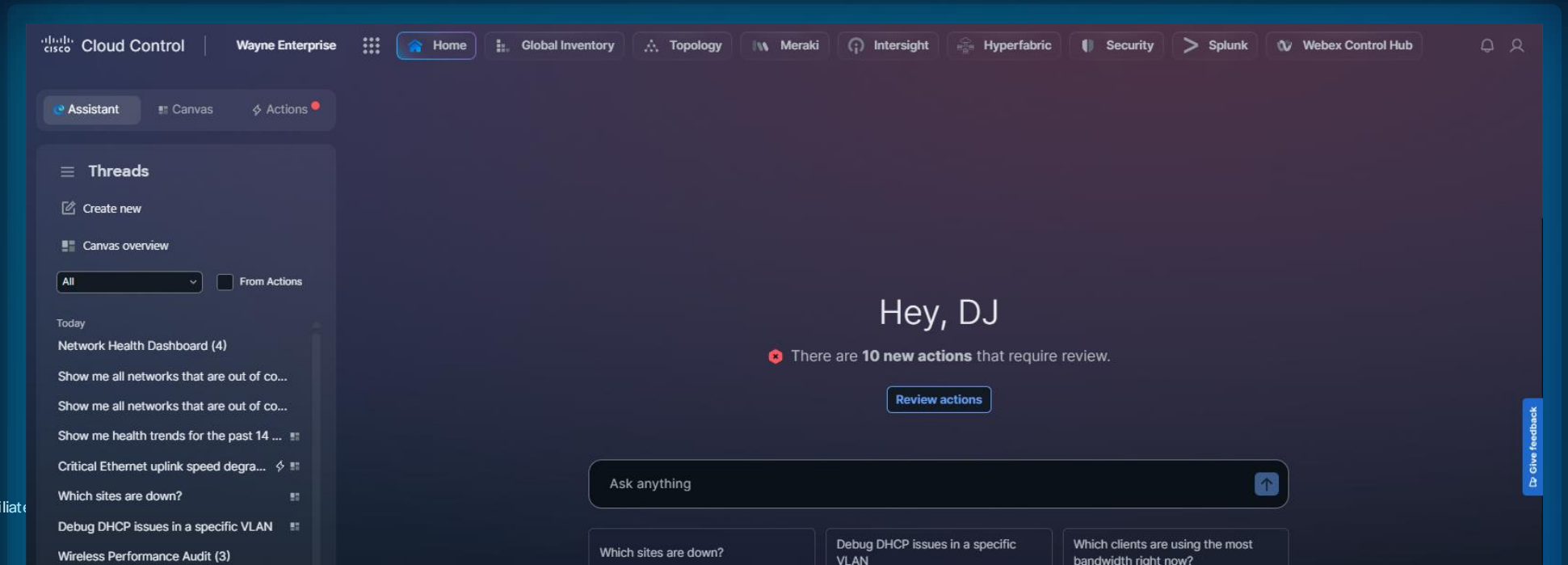
AgenticOps aligning speed with precision

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

Cisco Cloud Control

One platform. Every domain. The foundation everything else is built on.

ThousandEyes · Security Cloud Control · Splunk · Nexus Dashboard · Nexus Hyperfabric · Catalyst SD-WAN · Intersight · & more

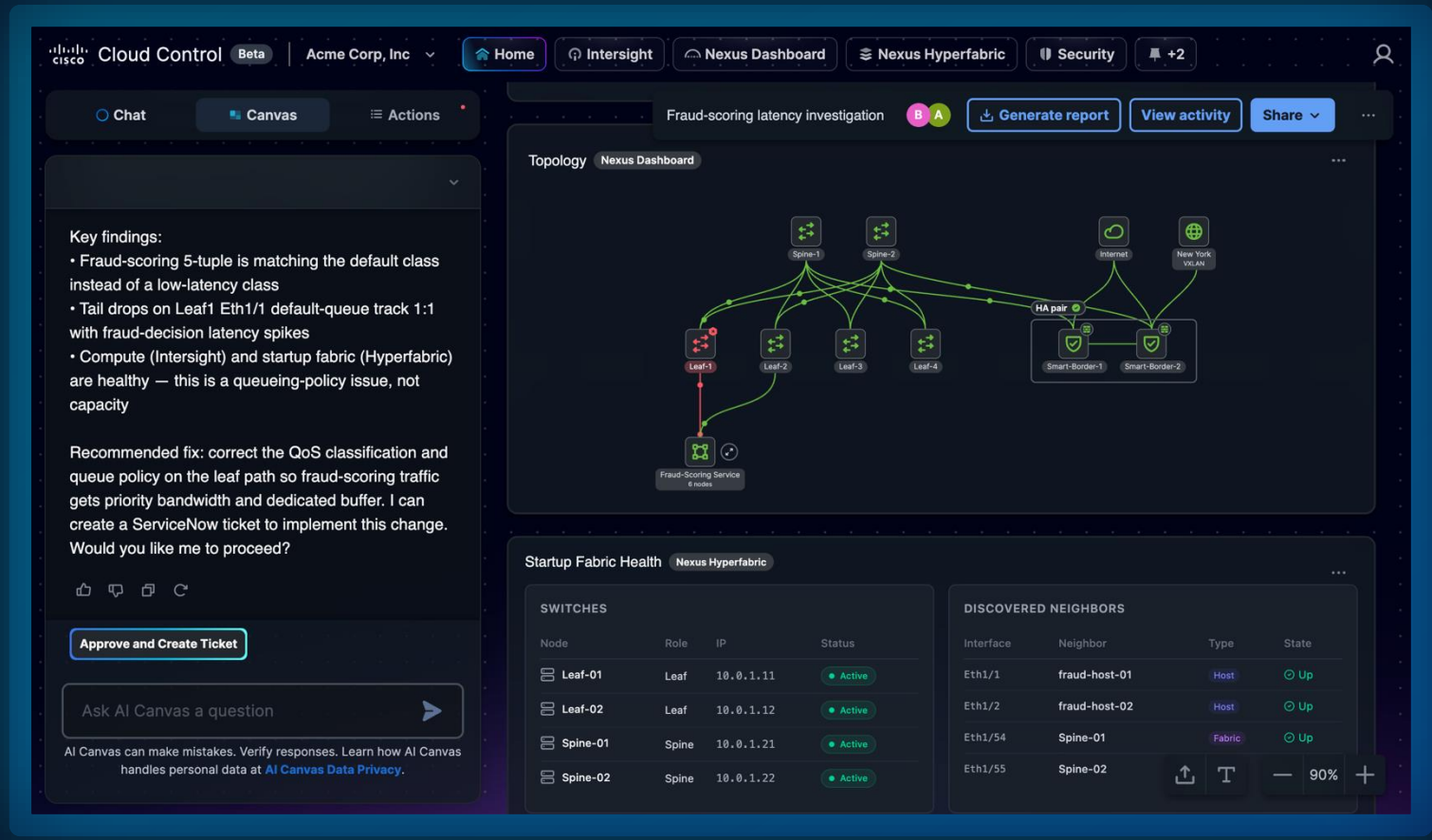


© 2026 Cisco and/or its affiliat

Cisco Cloud Control

Extend insights to the data center

Connect insights across networking, compute, and security for faster troubleshooting and continuous management



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

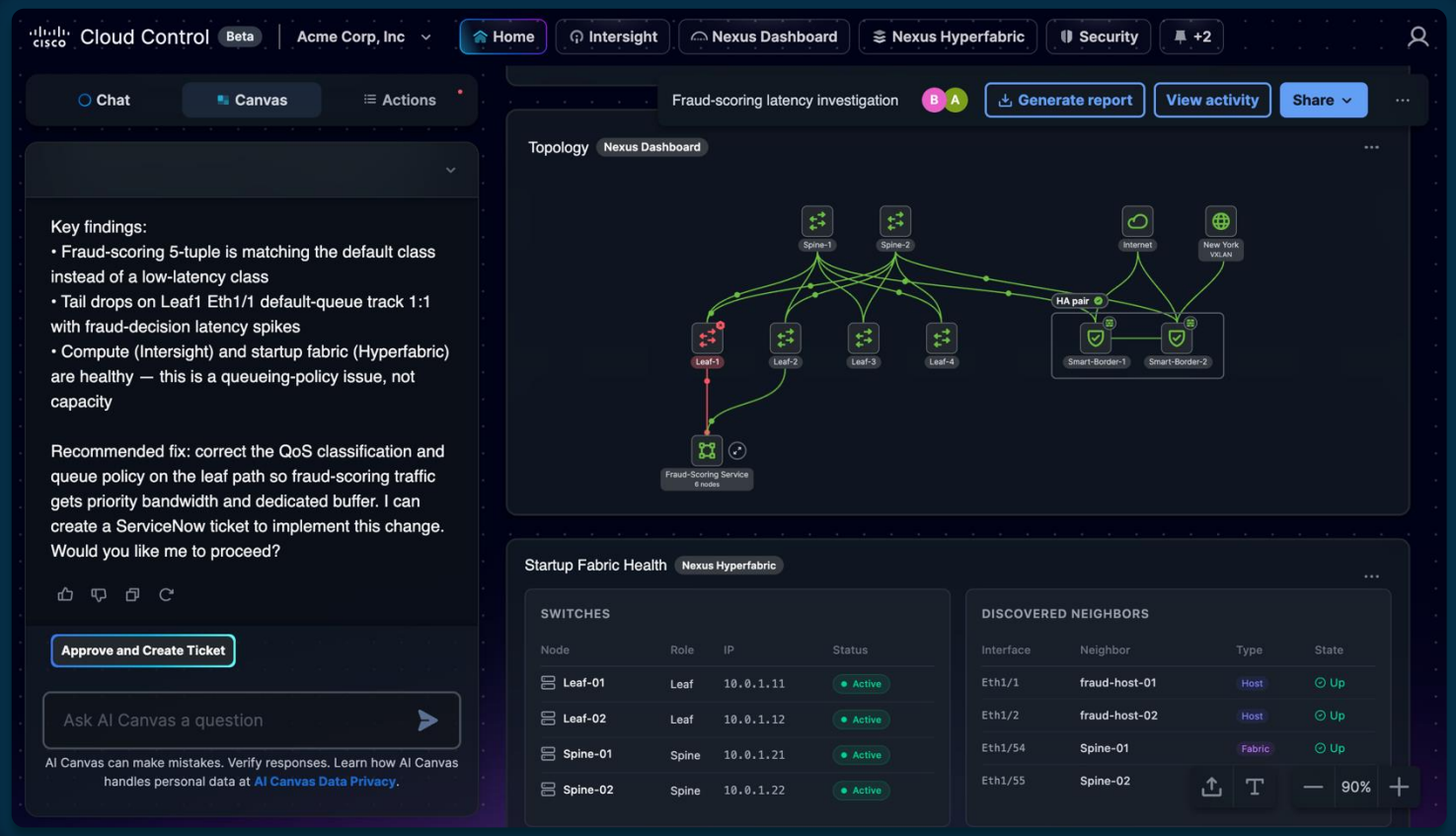
- VPN tunnel health
- Traffic flow & routing context
- Guided remediation

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

Cisco Cloud Control

Nexus One and Isovalent

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

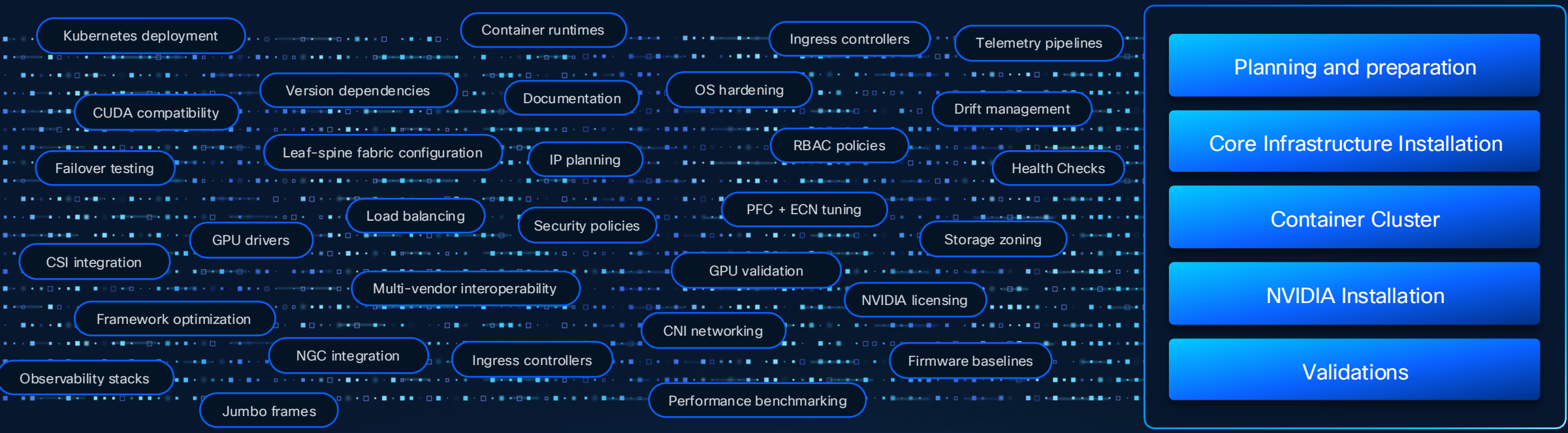
Isovalent

- Formulation/auditing of Segmentation Policy
- Vulnerability Management

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

BEFORE

Manual, Component-Based Assembly Across Systems, Tools, Teams and Sites



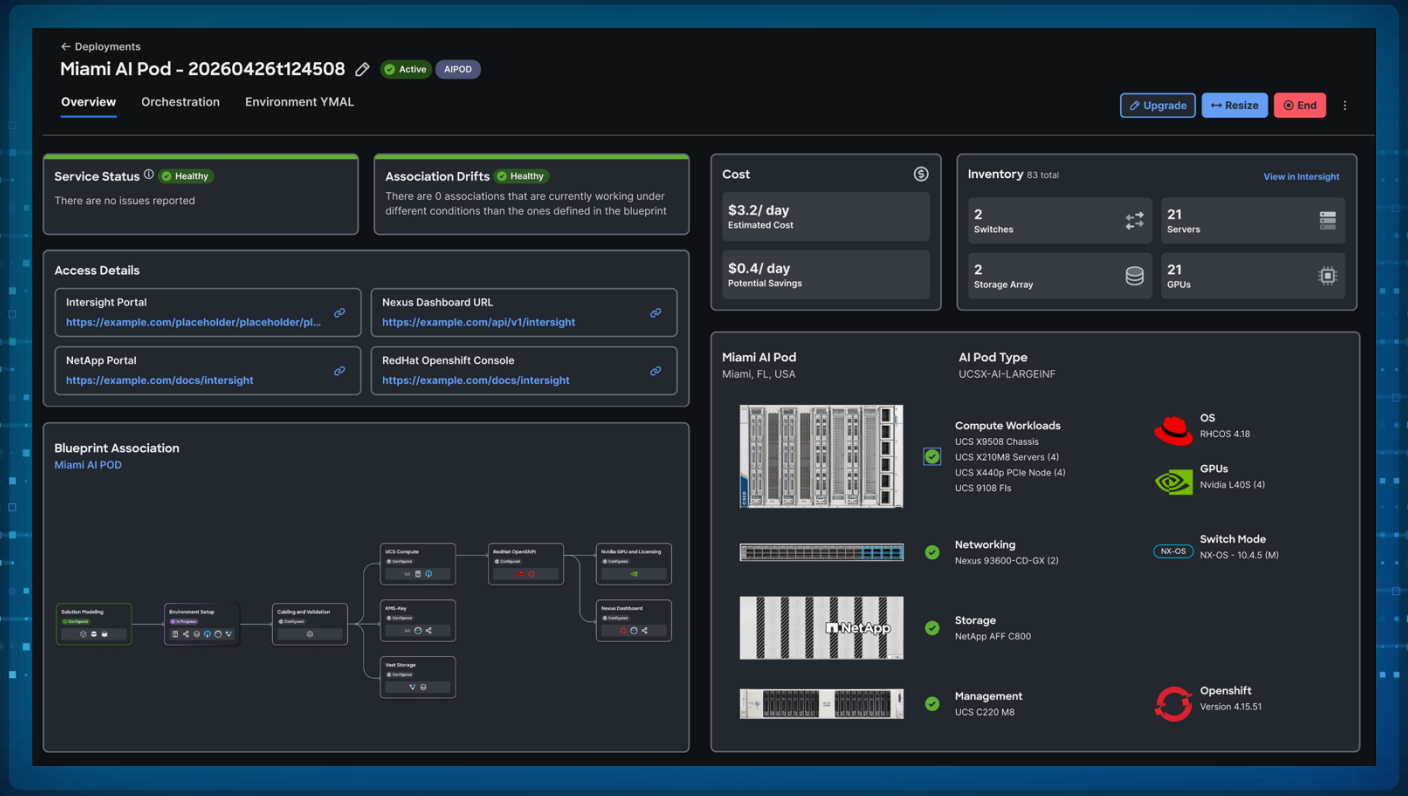
25-40	300-700+	40-100+	High	WEEKS
Tool configurations	Infrastructure actions	Handoffs & approvals	Operational Risk	Deployment Time

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

NOW

Stack Automation by Quali (Part of Cisco Cloud Control)

Intelligent, Automated Deployment from Rack to App



Automated	Automated	Automated	Near-Zero	HOURS
Tool configurations	Infrastructure actions	Handoffs & approvals	Operational Risk	Deployment Time

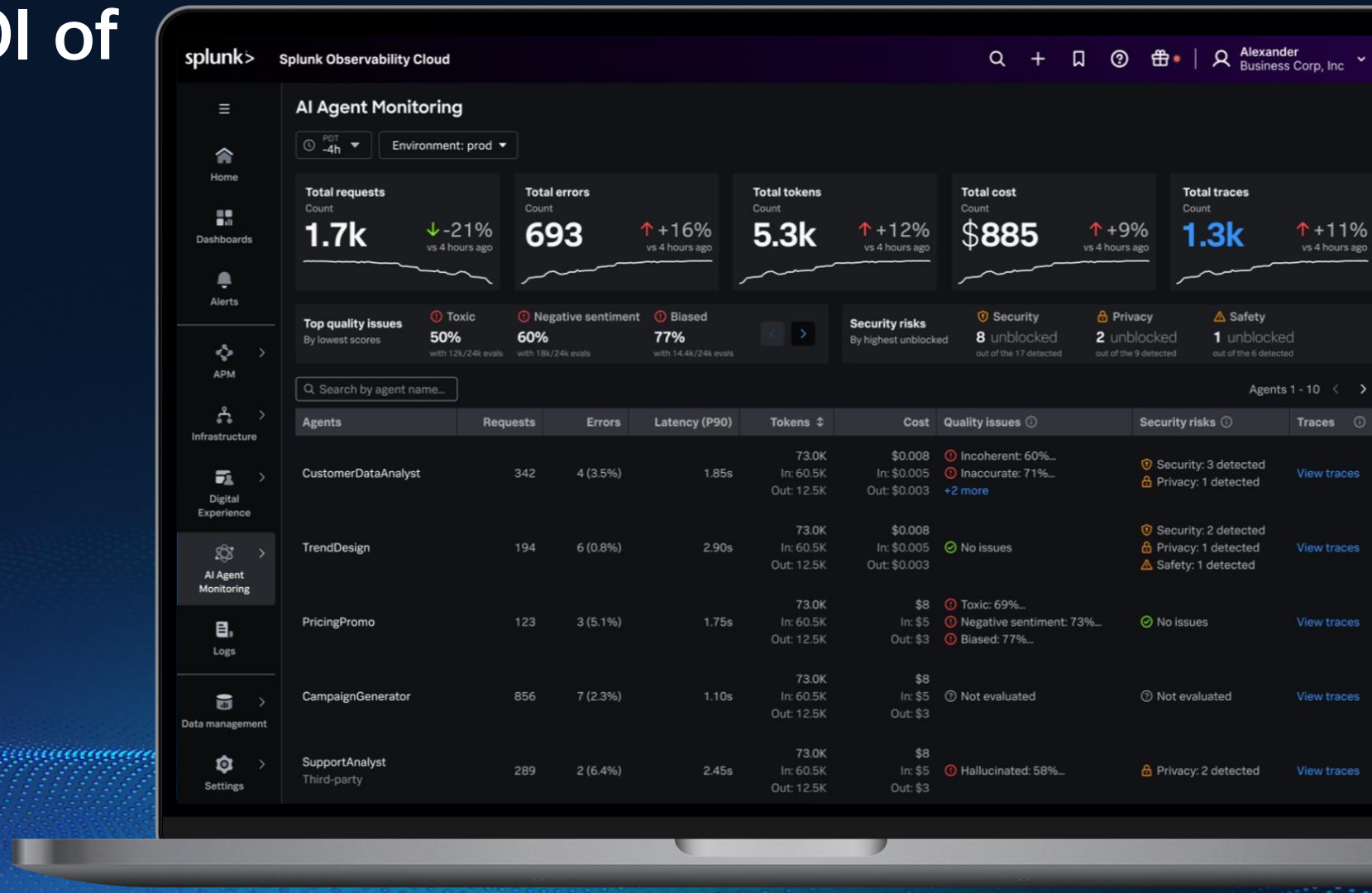
© 2026 Cisco and/or its affiliates. All rights reserved.

Optimize the performance and ROI of your AI stack

Monitoring the performance, quality, security, and cost with Splunk Observability

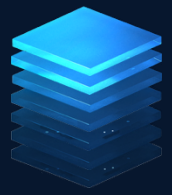
AI agent monitoring

AI infrastructure monitoring



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

Working with you to improve our full-stack approach



Integrated, Secure AI Infrastructure

01

Cisco Validated Designs

Tested, validated configurations

02

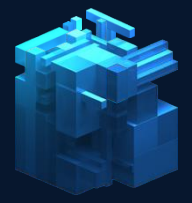
Cisco AI PODs

Pre-packaged solutions

03

Cisco Secure AI Factory with NVIDIA

A foundation for deploying AI workloads from core to edge



AI Solutions Ecosystem

04

Cisco Compatible AI Solutions

An ecosystem of real-world, AI-based outcomes deployable at scale



Automated Full Stack Deployment

05

Stack Automation by Quali

Automated deployment from rack to app in hours, not weeks

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

Next steps

Pick one AI use case to pressure-test

1

Map compute, fabric, security, and operations owners

2

Choose the next technical discussion with your Cisco team

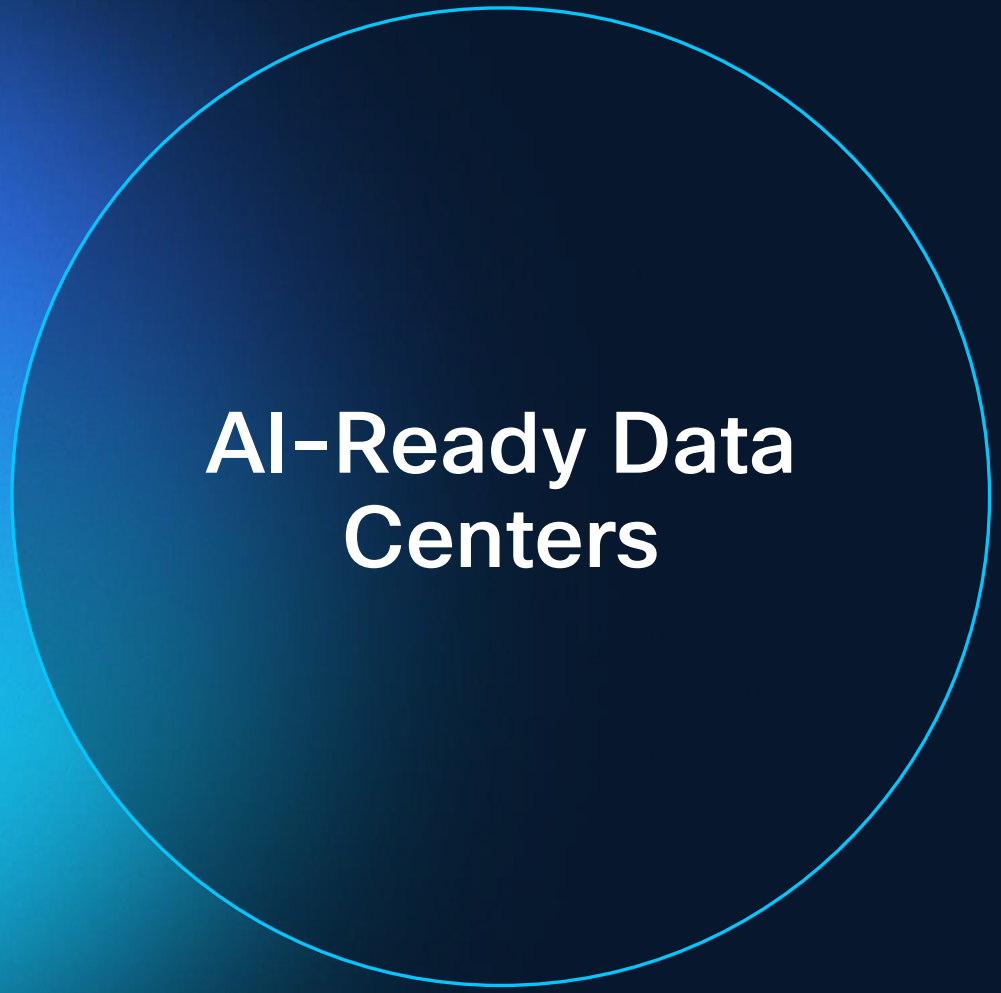
3

AI POD / SAIF architecture, fabric visibility, AI security, or day-2 ops

Use the regional follow-up to align the right owners and pick the next technical discussion.

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

Our products and solutions



Scale networking for AI

- Silicon One
- Nexus One
- N9000 Series
- Cisco 6000
- Cisco Optics
- Cisco 8000

Run full-stack infrastructure efficiently

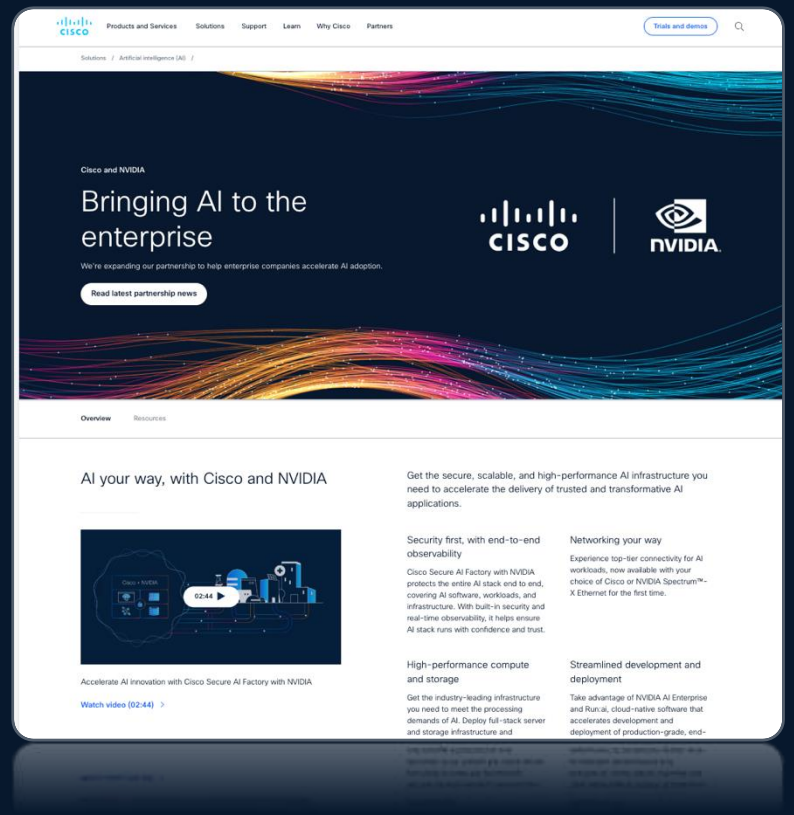
- AI PODs
- Unified Edge
- UCS-X
- Intersight
- UCS AI Servers
- NVIDIA, AMD, and Intel silicon
- Secure AI Factory with NVIDIA
- Red Hat AI
- Cisco Cloud Control
- Splunk
- Storage partners

Protect east-west traffic and workloads

- Smart Switches
- Secure Firewall
- Secure Workload
- Hypershield
- Isovalent Enterprise
- Live Protect
- AI Defense
- Identify services engine
- Cisco Multicloud Defense

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

Resources



SecureAIFactory.com

FAQ

Recent Cisco Validated Designs

FlexPod with Cisco AI POD: Infrastructure for AI Training and Fine-Tuning Deployment Guide

Cisco AI POD for Enterprise Training and Fine-Tuning with Everpure Deployment Guide

Cisco AI Defense on AI PODs

Cisco AI POD for Enterprise Training and Fine-Tuning Design Guide

FlashStack for Enterprise RAG Pipeline with NVIDIA NIM

FlexPod for Accelerated RAG Pipeline with NVIDIA NIM

NVIDIA Reference Architecture Compliant Cisco RAs

Cisco Enterprise RA with Cisco Nexus 9000 switches, compliant with NVIDIA ERA

Cisco Nexus Hyperfabric Enterprise Reference Architecture (RA), compliant with NVIDIA Enterprise RA

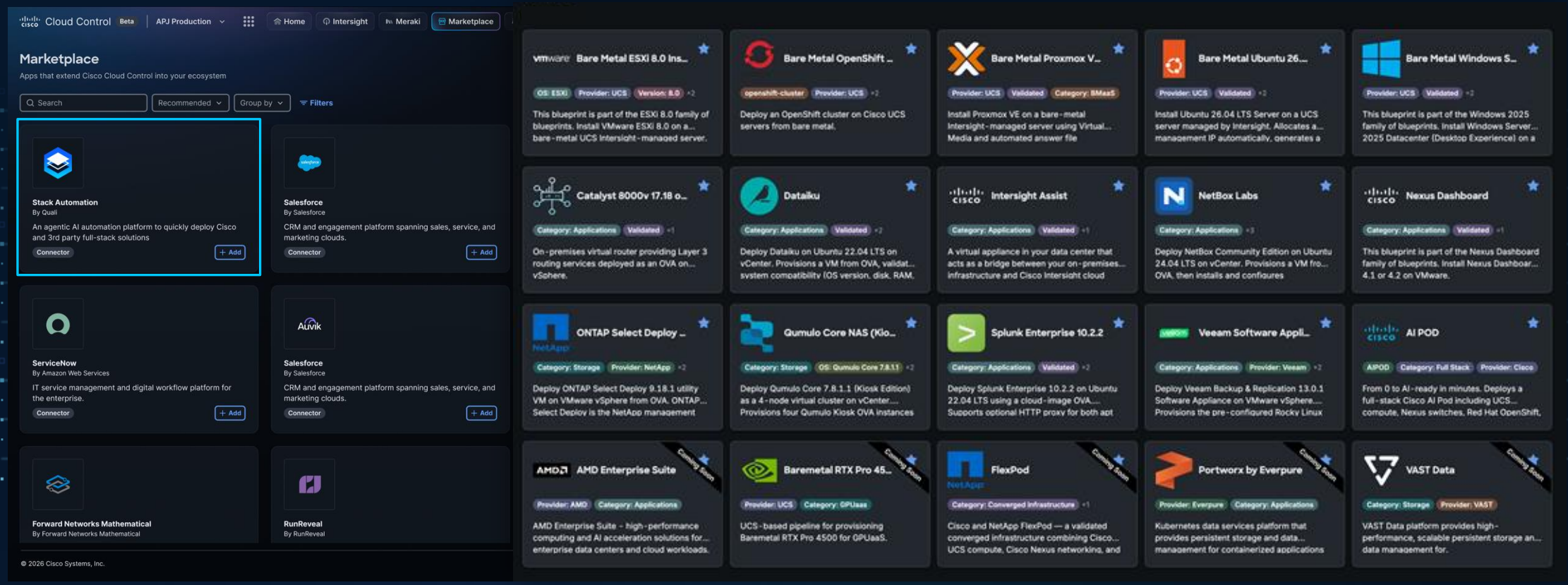
Cisco Cloud RA with Cisco Nexus 9000 switches, complaint with NVIDIA NCP RA

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

NOW

Stack Automation by Quali

Click-To-Deploy Solutions Hub, Cisco Cloud Control Marketplace-Ready



Blueprints embedded with policies, best practices, validated designs, and agentic intelligence

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

Cisco IT deploys AI-ready data center in weeks

See the full story

Challenges

Critical business need with short build timeline

Existing data center facilities weren't built to power modern AI

Extensive technical requirements including ethernet-based lossless fabric

Operations for new environment needed to be automated

Solutions

Cisco 8100 Series embedded with Silicon One

Cisco N9000 Series running NX-OS

Cisco Nexus Dashboard

SONiC running on Cisco 8000 series switches

Results

Speed of deployment: AI Cluster was designed and deployed in 3 months, 80% faster than alternative rebuild

Enhanced AI capabilities: AI Cluster supports more than 25 use cases with more added each week

Designed to support future needs: Maximized agility to power any workload, traditional or AI, at massive scale, no matter where the data lives

"Not only were we able to support the needs of the business today, but we're designing how our data centers need to evolve for the future.

The modularity and flexibility of Cisco's infrastructure gives us confidence that we can keep scaling with the business."

Jon Woolwine
Director, Technical Systems Engineering, Cisco IT

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

Cisco Unified Computing System

Unified approach to Enterprise and AI workloads

Validated solutions for AI with compute, network, storage, and software



GPU Accelerated

GPU Optimized

Unified Edge

© 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

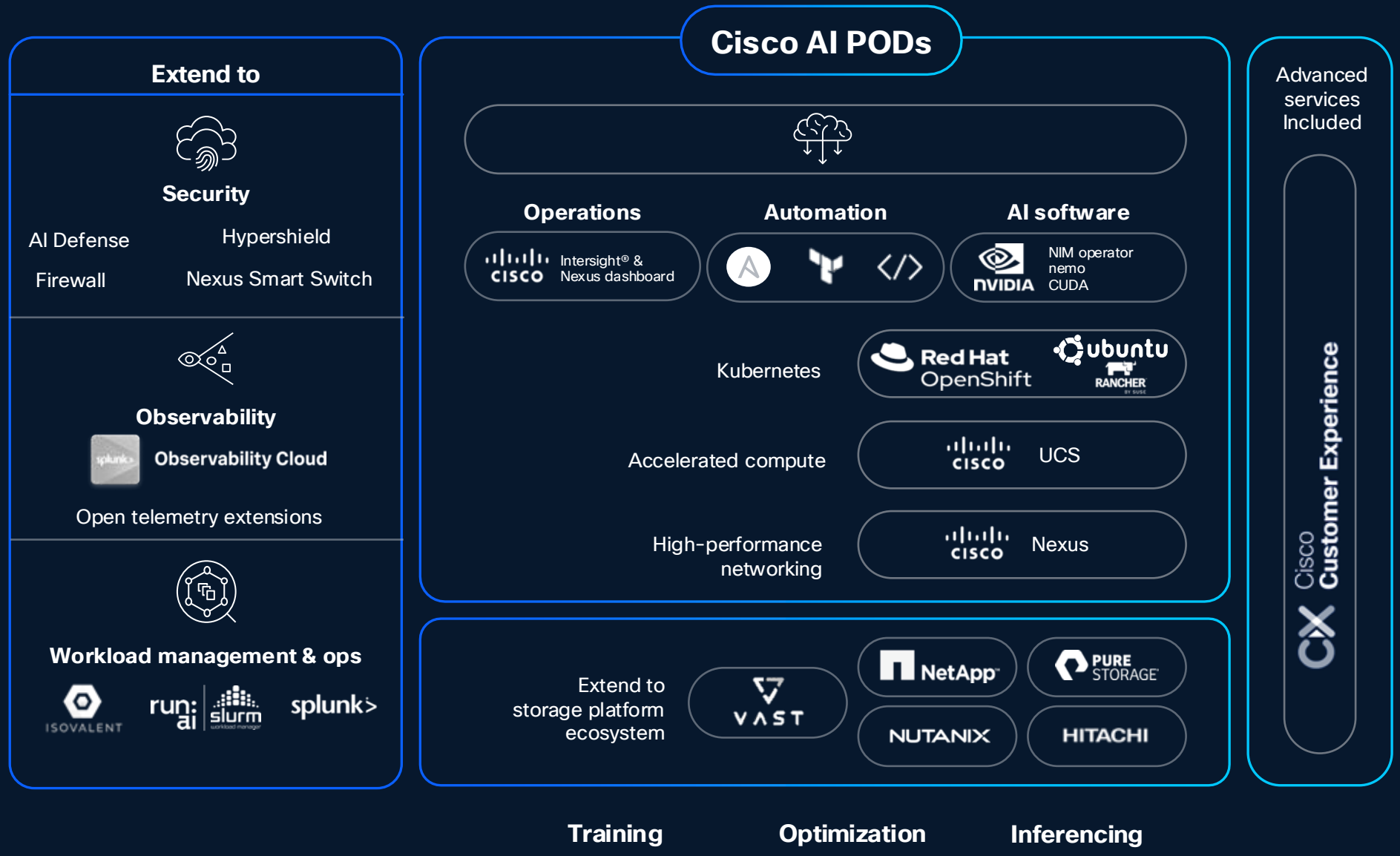
Fabric
- NCPRA compliant
- Rack to fabric liquid cooling

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

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

Cisco AI PODs

A scalable architecture, built to support any AI workload simply & efficiently



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

Innovating in the Fabric:

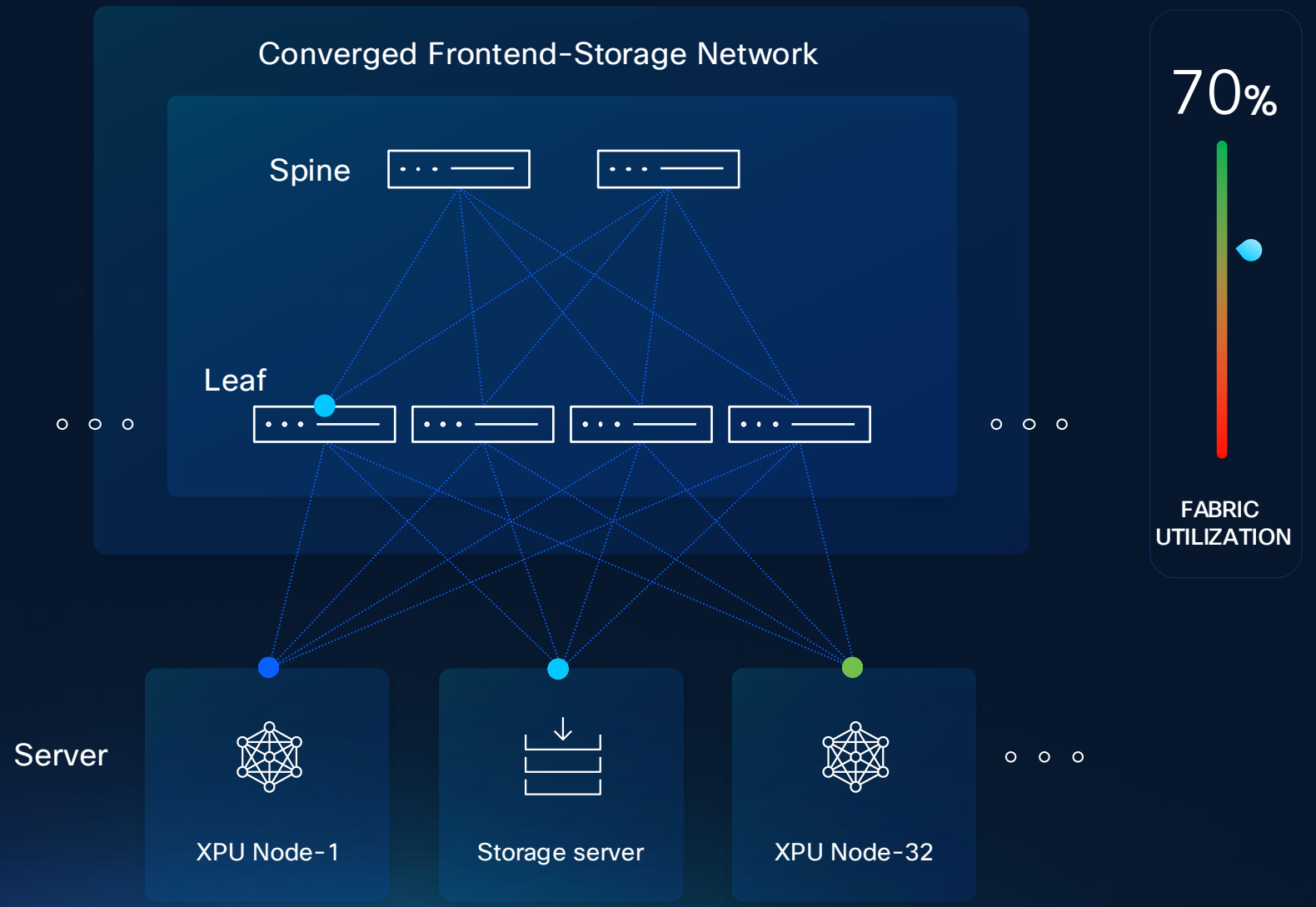
Cisco Intelligent Packet Flow

Per flowlet, per-packet, a mix of both, or dynamically switch between

Flow1*
Flow2*
Flow3X
Flow4

* DLB enabled flow

X QoS Policy flow



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