

Full Stack AI-Ready Data Centers

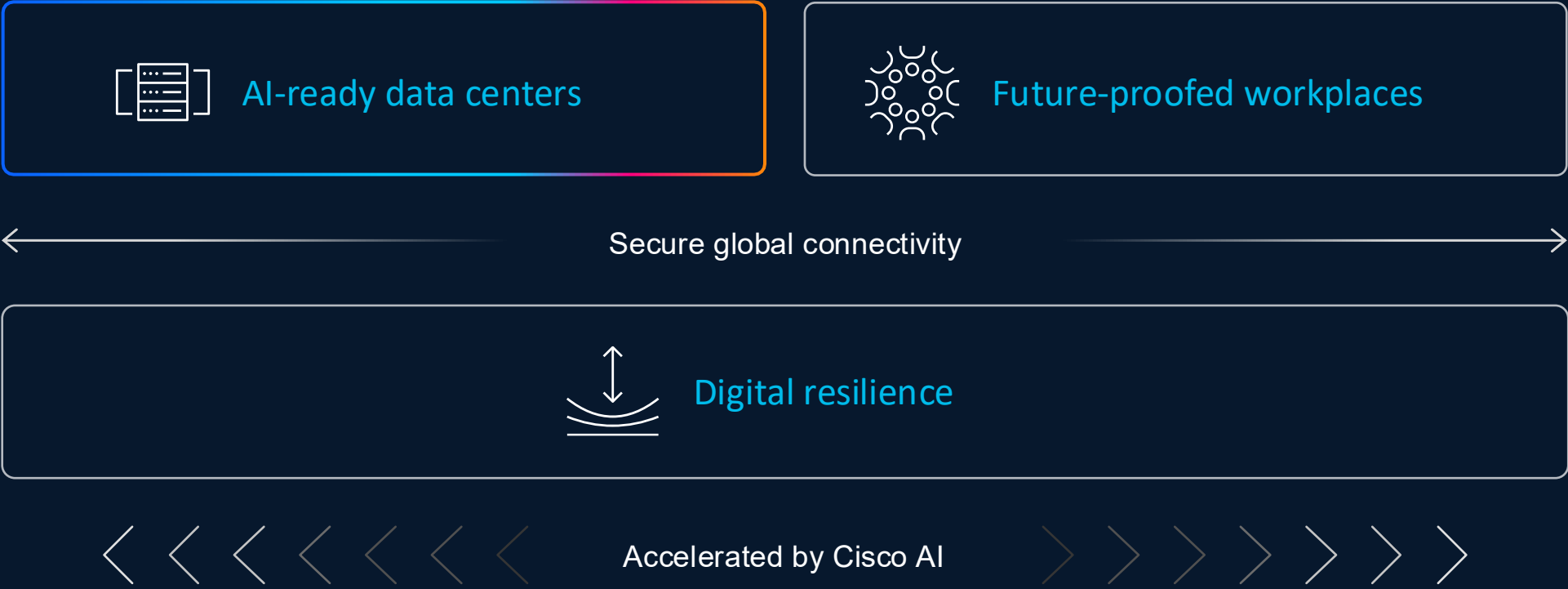
Built for AI, ready for what's next

Kahahn Allen

Observability Specialist, kahallen@cisco.com,



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



AI demand is spreading across 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



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

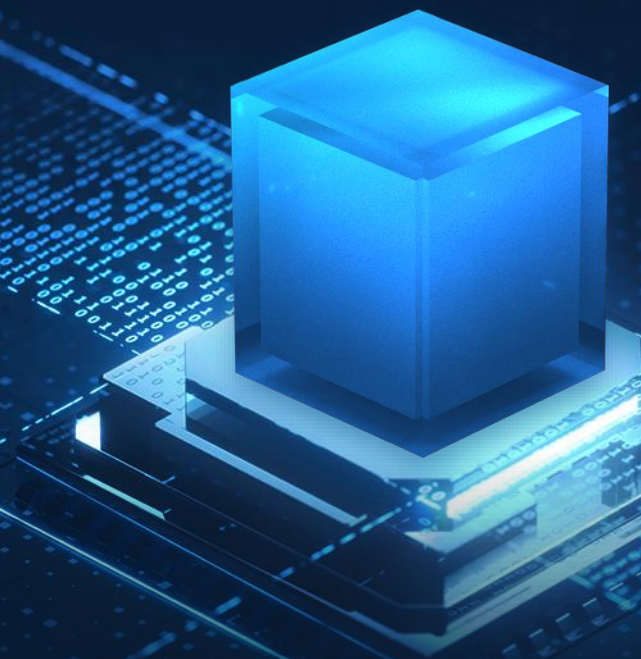
Trust zones, policy, risk and
governance

Operations

Visibility, ownership, lifecycle

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

A full-stack approach to AI

Cisco AI-Ready Data Centers



Cisco AI-Ready Data Centers

Full-stack infrastructure, engineered as
one system

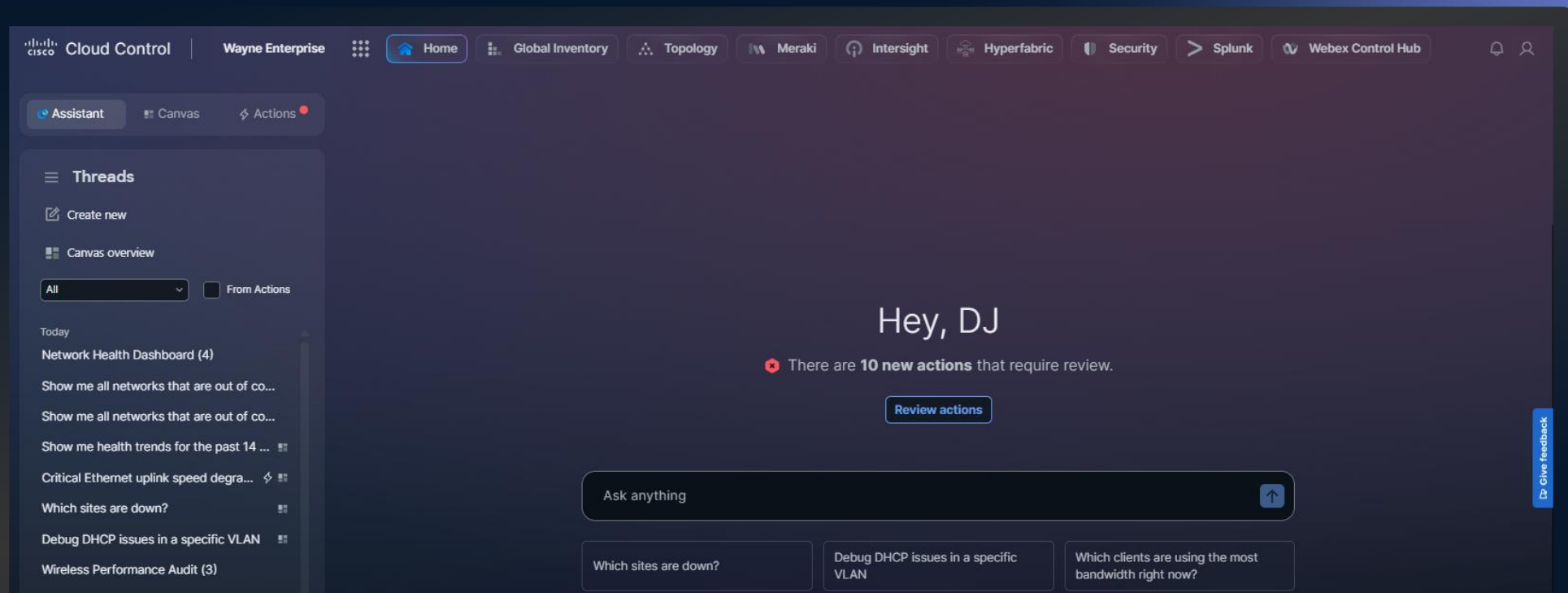
Distributed security
fused into every layer

AgenticOps aligning
speed with precision

Cisco Cloud Control

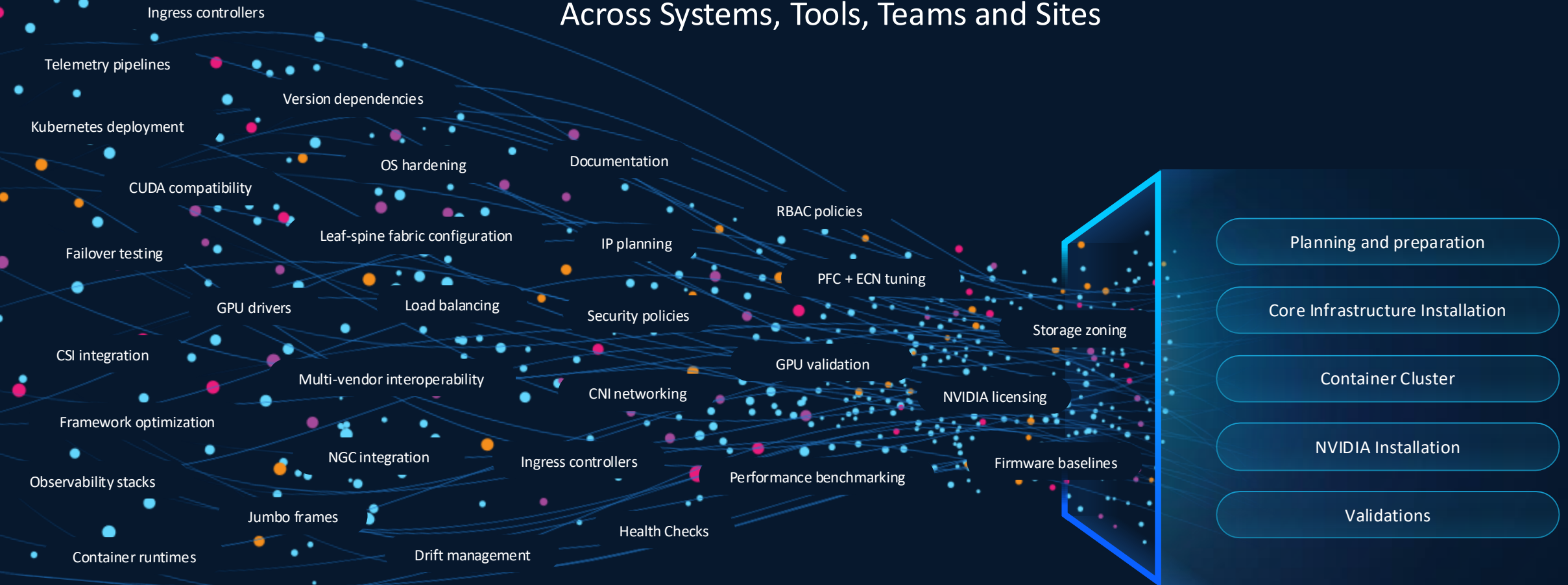
One platform. Every domain. The foundation everything else is built on.

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



BEFORE

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



25-40
Tool configurations

300-700+
Infrastructure actions

40-100+
Handoffs & approvals

High
Operational Risk

WEEKS
Deployment Time

NOW

Stack Automation by Quali

Intelligent, Automated Deployment from Rack to App

← Deployments

Miami AI Pod - 20260426t124508 🔗 Active AIPOD

Overview Orchestration Environment YMAL Upgrade Resize End

Service Status 🔔 Healthy

There are no issues reported

Association Drifts 🔔 Healthy

There are 0 associations that are currently working under different conditions than the ones defined in the blueprint

Access Details

Intersight Portal
<https://example.com/placeholder/placeholder/pl...> 🔗

Nexus Dashboard URL
<https://example.com/api/v1/intersight> 🔗

NetApp Portal
<https://example.com/docs/intersight> 🔗

RedHat Openshift Console
<https://example.com/docs/intersight> 🔗

Blueprint Association
Miami AI POD

```
graph LR; A[Solution Modeling] --> B[Environment Setup]; B --> C[Catalog and Validation]; C --> D[UCS Compute]; C --> E[NetApp]; C --> F[OpenShift]; D --> G[RedHat OpenShift]; E --> H[NetApp]; F --> I[OpenShift]; G --> J[Nvidia GPU and Licensing]; H --> K[Nexus Dashboard]; I --> L[Start Storage];
```

Cost 💰

\$3.2/ day
Estimated Cost

\$0.4/ day
Potential Savings

Inventory 83 total View in Intersight

2 Switches

21 Servers

2 Storage Array

21 GPUs

Miami AI Pod
Miami, FL, USA

AI Pod Type
UCSX-AI-LARGEINF

Compute Workloads
UCS X9508 Chassis
UCS X210M8 Servers (4)
UCS X440p PCIe Node (4)
UCS 9108 Fls

OS
RHCOS 4.18

GPUs
Nvidia L40S (4)

Networking
Nexus 93600-CD-GX (2)

Storage
NetApp AFF C800

Management
UCS C220 M8

Switch Mode
NX-OS - 10.4.5 (M)

OpenShift
Version 4.15.51

Automated
Tool configurations

Automated
Infrastructure actions

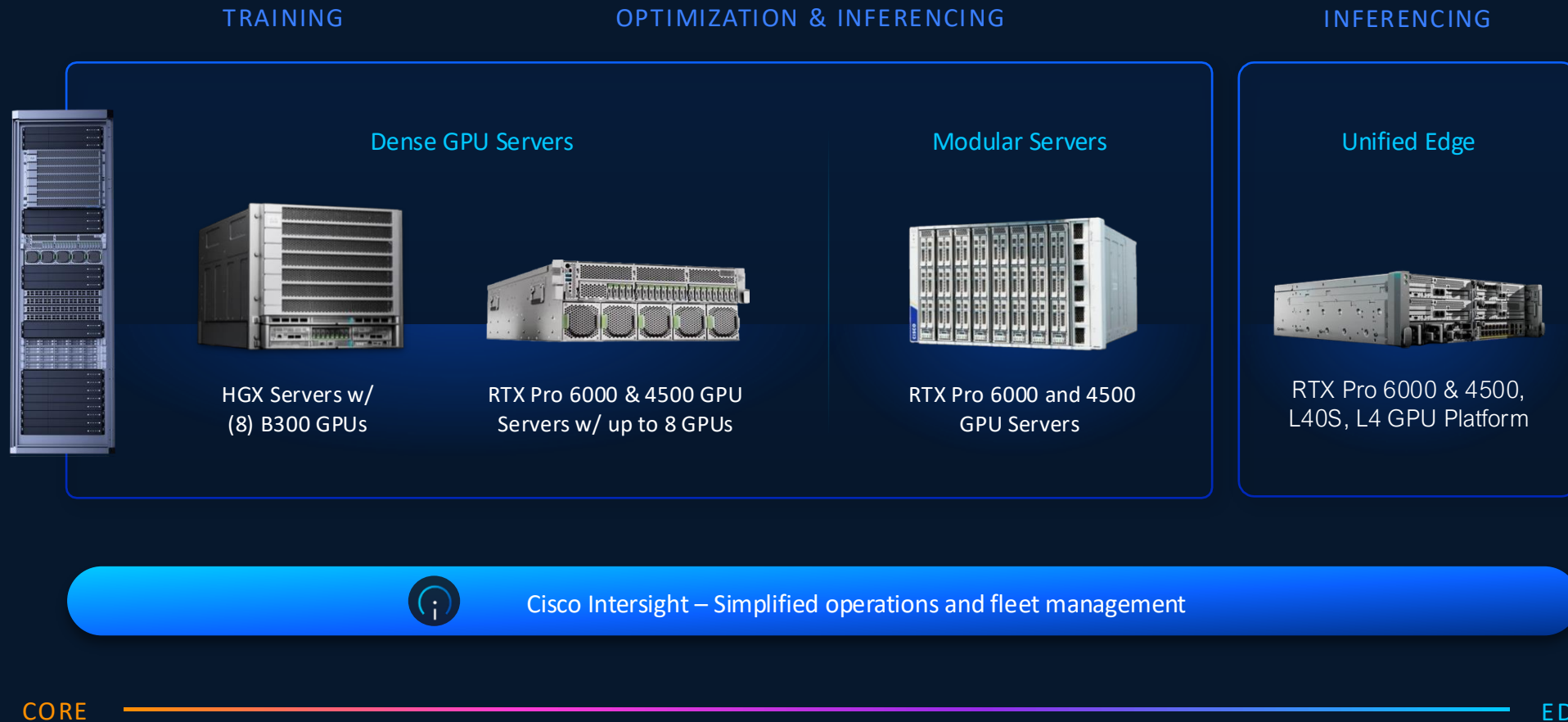
Automated
Handoffs & approvals

Near-Zero
Operational Risk

HOURS
Deployment Time

Unified approach to enterprise and AI workloads

Broad spectrum of Cisco AI servers to meet any requirement



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

Agents add dynamic performance loops to applications

Kubernetes made applications distributed.

AI makes each answer depend on a dynamic loop across models, data, tools, apps, and accelerators.

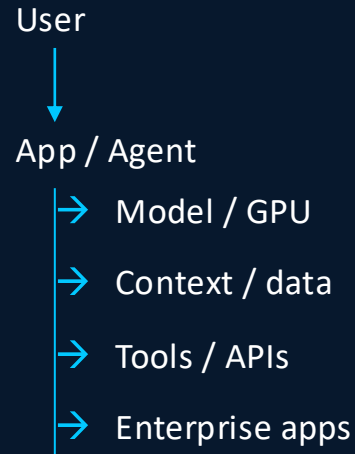
Modern app traffic



Main pressure: scale services reliably

AI-enabled workflow

Dynamic execution loop
retrieve → reason → act → verify



Main pressure: keep every loop fast, predictable, and economical

Why the AI fabric matters

- Paths vary by request
- More data moves per answer
- Loops are latency-sensitive
- GPU wait time is expensive

Cisco AI Network Fabric keeps the AI loop moving

Front-end services feed the AI loop. Back-end fabric keeps accelerators moving.

Cisco fabric building blocks

Build the fabric

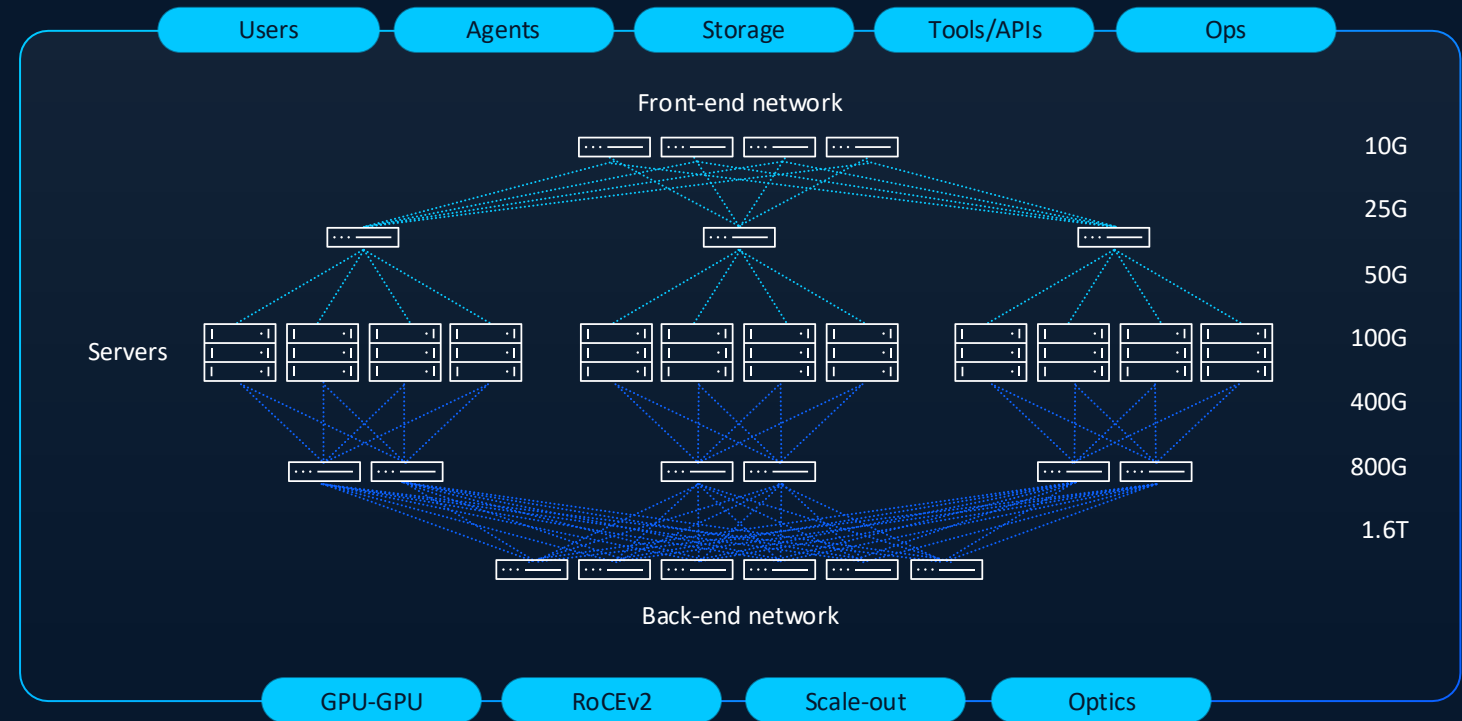
- Cisco N9000
- Cisco Silicon One | NVIDIA Spectrum-X Silicon
- Cisco Optics
- NX-OS | ACI | SONiC

Tune the AI traffic

- RoCEv2 lossless Ethernet
- Intelligent Packet Flow
- Intelligent buffering | deep and shared

Operate it

- Cisco Cloud Control
- Cisco Nexus One
- Automation | telemetry | congestion insight



Lossless

High throughput

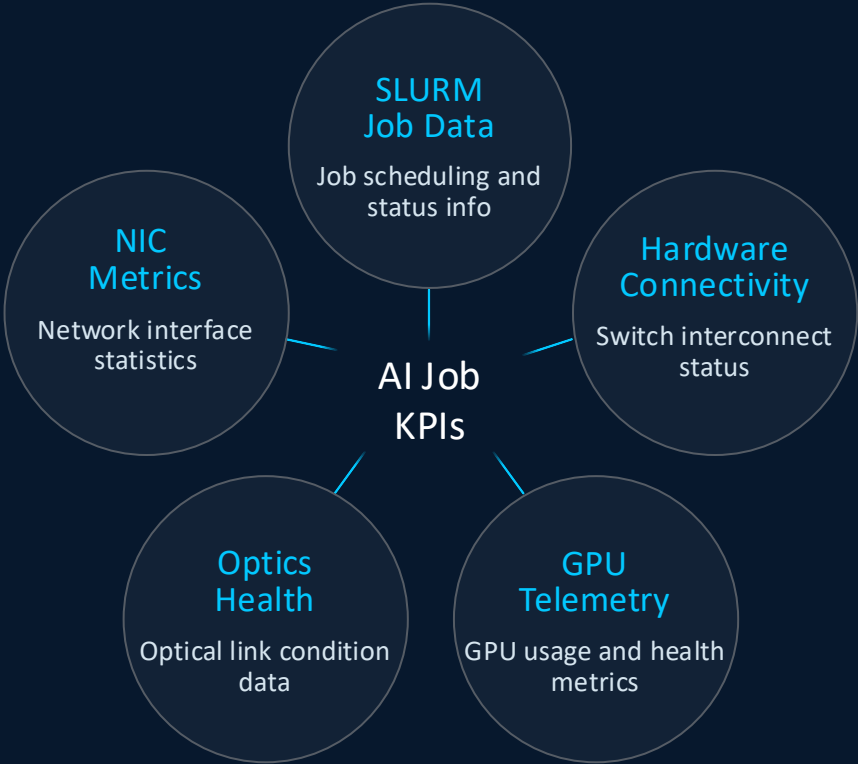
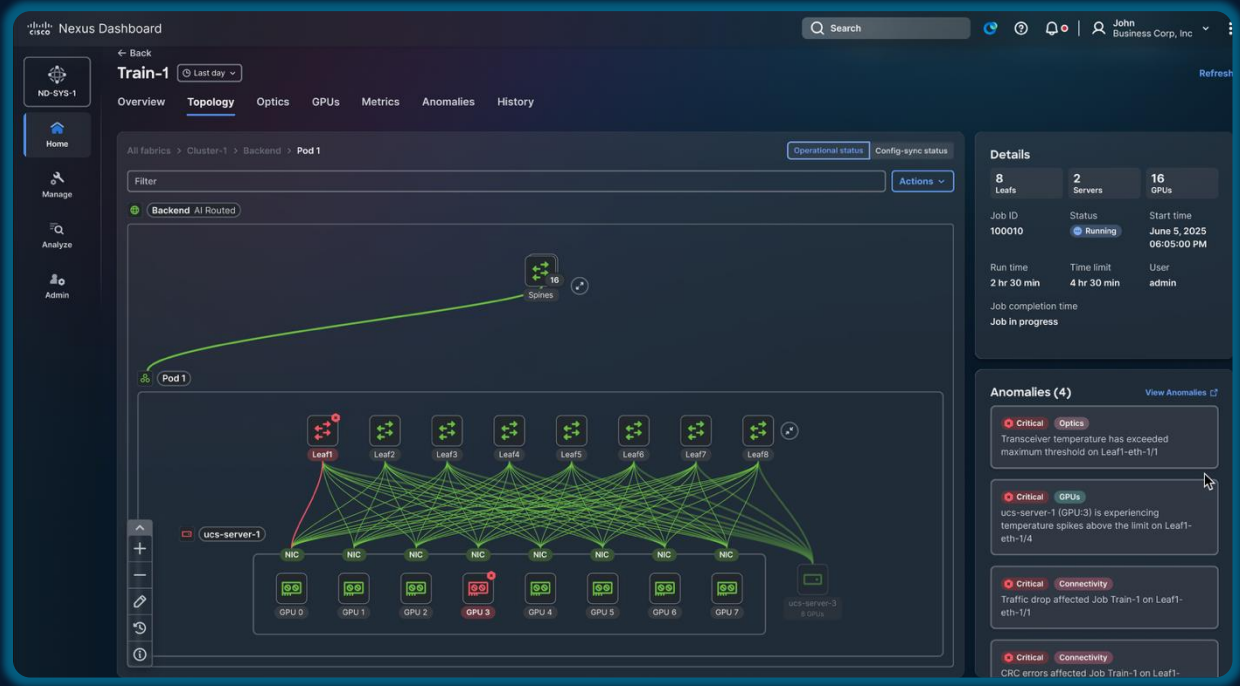
Low latency

Low jitter

Telemetry-ready

Job-centric AI fabric visibility in Cisco Nexus Dashboard

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



Detect likely bottlenecks before they affect AI job throughput.

Recent AI networking innovations

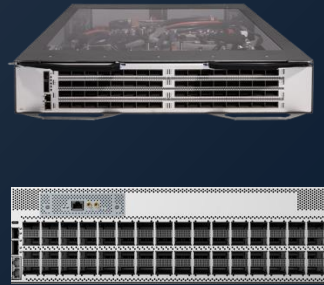
SILICON



Cisco Silicon One G300

102.4 Tbps
Network Switching
Silicon

SYSTEMS



Advanced Scale-Out Systems

Cisco N9000 +
Cisco 8000 Systems

Cisco N9100 Systems
with NVIDIA Spectrum-X
Ethernet

OPTICS

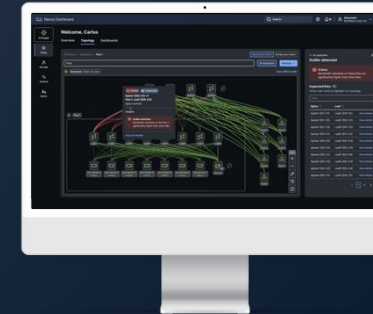


High-Quality Optics

Cisco OSFP 1.6T Pluggable
Optics

Cisco OSFP 800G Linear
Pluggable Optics

SOFTWARE



Enhanced Software Features

AI Job Observability
Native Splunk Integration
AgentOps Capabilities

OPERATING MODEL

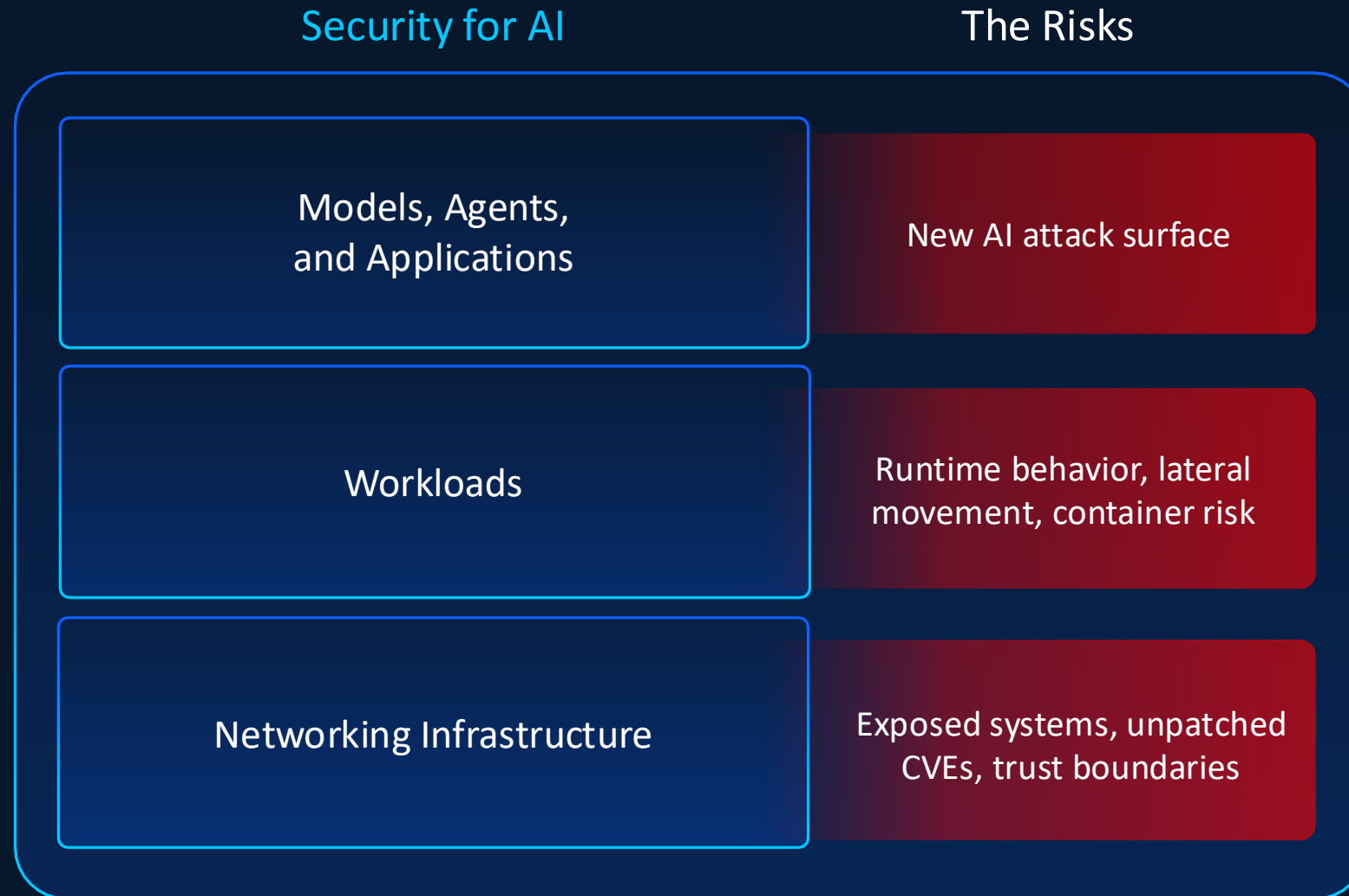
Cisco
Nexus
One

Nexus One

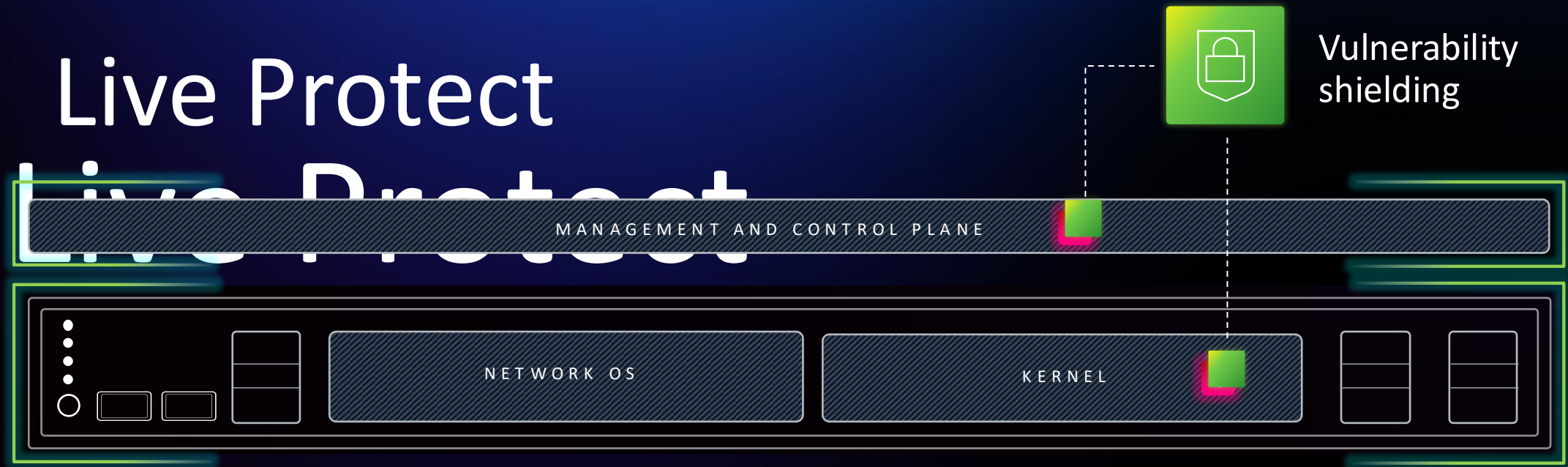
Unified operations across on-
premise and cloud-based
deployments with flexible OS
support

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

Secure the AI era: new stack, faster attackers



Live Protect



EXTENDING ACROSS CISCO NETWORKING AND SECURITY

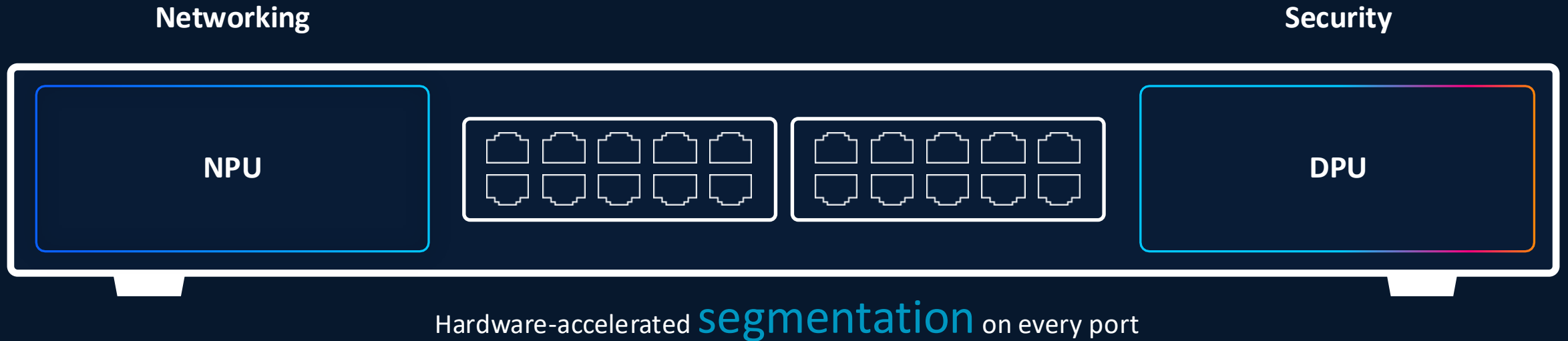
Apply policy to
live system

Protect now,
patch in planned window

No downtime or
switch reboot

Continuous
re-validation

Threat containment at massive scale built into your network



Cisco Hybrid Mesh Firewall

Distributed security at the speed of change



Cisco
Firewalls

Physical | Virtual | Cloud | FWaaS



3rd Party
Firewalls



NVIDIA Bluefield DPU Firewall
on AI Servers



Smart Switches

eBPF



Workload
Agents

Define policy once, enforce everywhere

Cisco AI Defense

Securing Custom Agents and AI Apps



Discover



Detect



Protect

NEW

Adaptive red teaming
and guardrails

NEW

Protecting the full agent
supply chain

NEW

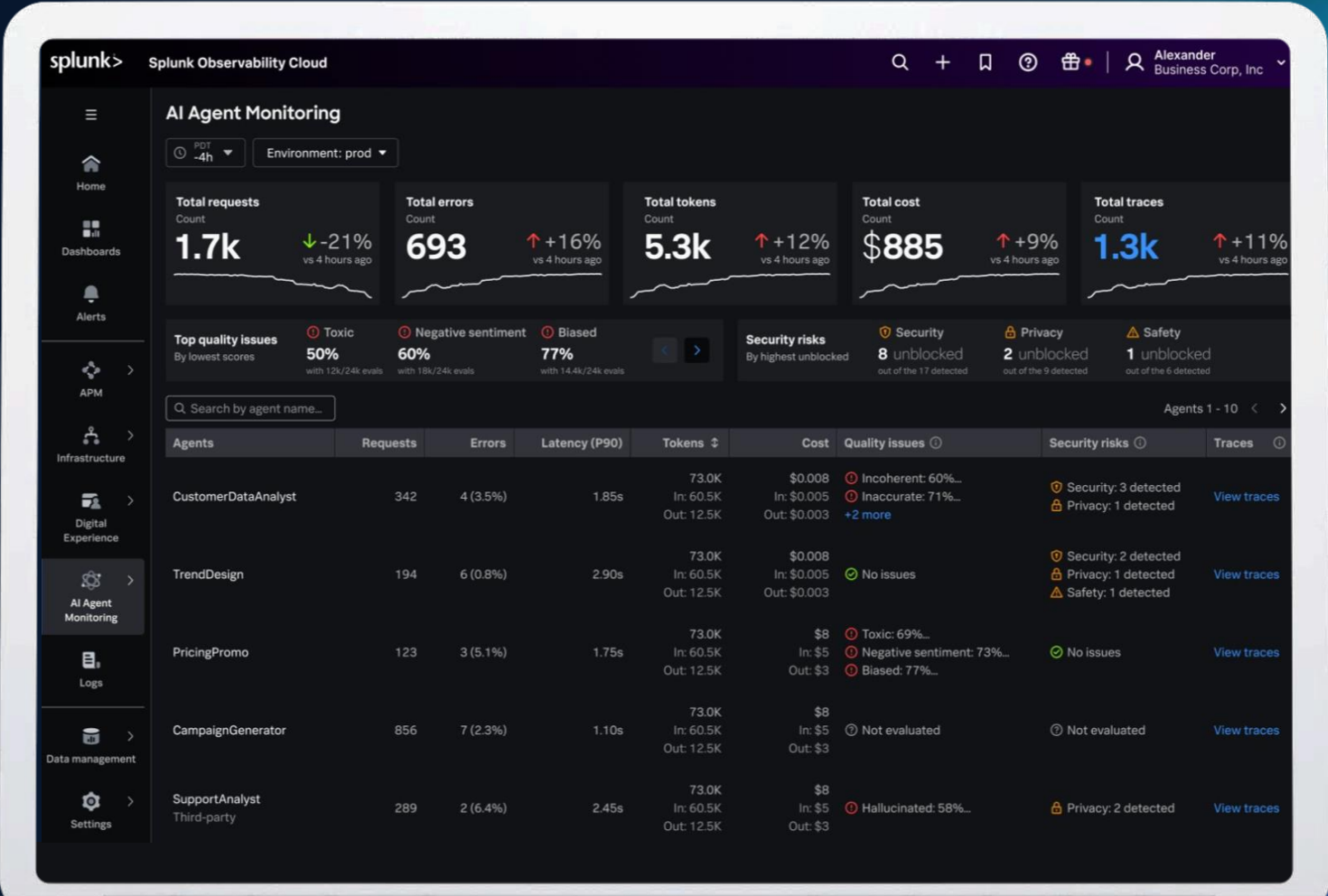
Ecosystem to work with any
cloud, any GPU

Splunk Observability enables unified visibility

Monitoring the performance, quality, security, & cost

AI agent monitoring

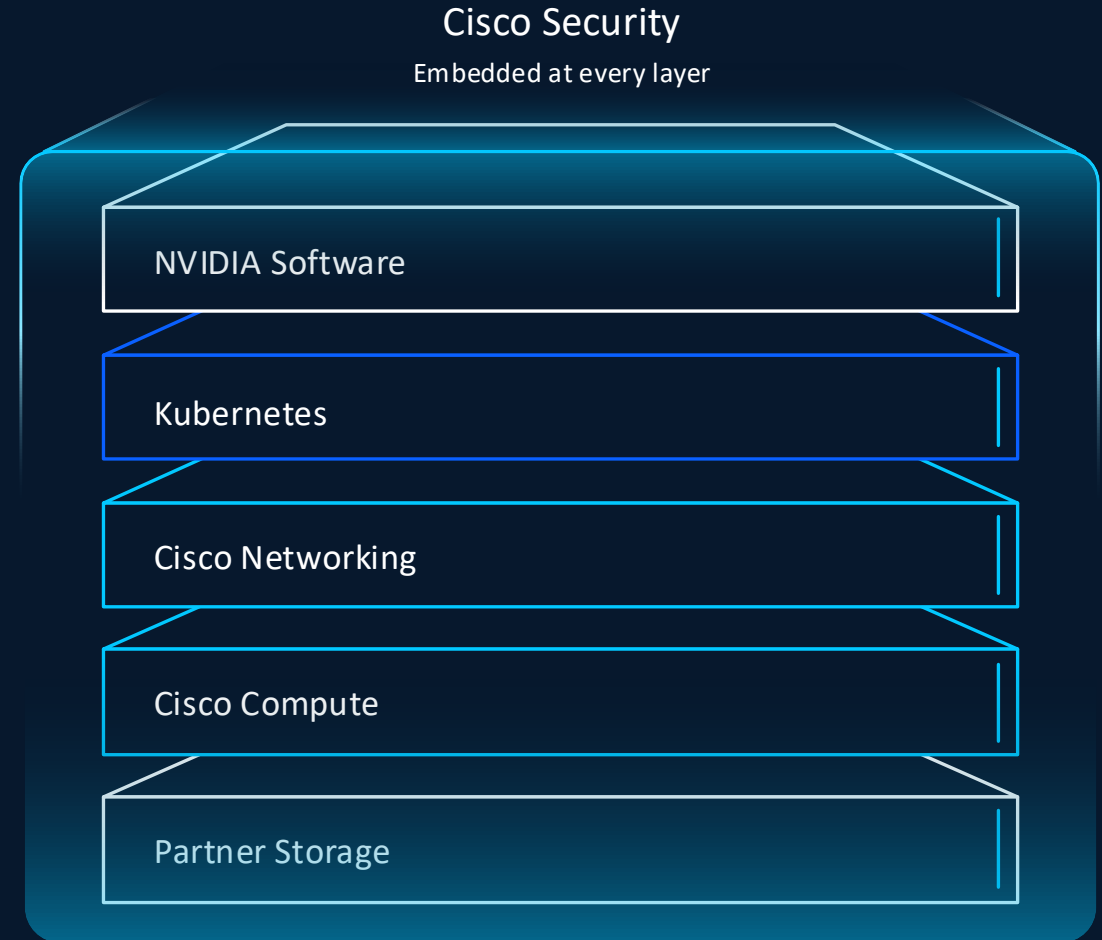
AI infrastructure monitoring



Cisco Secure AI Factory with NVIDIA

A reference design with validated architectures
to **accelerate AI adoption**

- Security-first architecture
- High-performance networking, compute, storage, and AI software
- Flexible deployment options: Vertically integrated or modular system



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

Next steps

Pick one AI use case to pressure-test

- Map compute, fabric, security, and operations owners
- Choose the next technical discussion with your Cisco team
- 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.

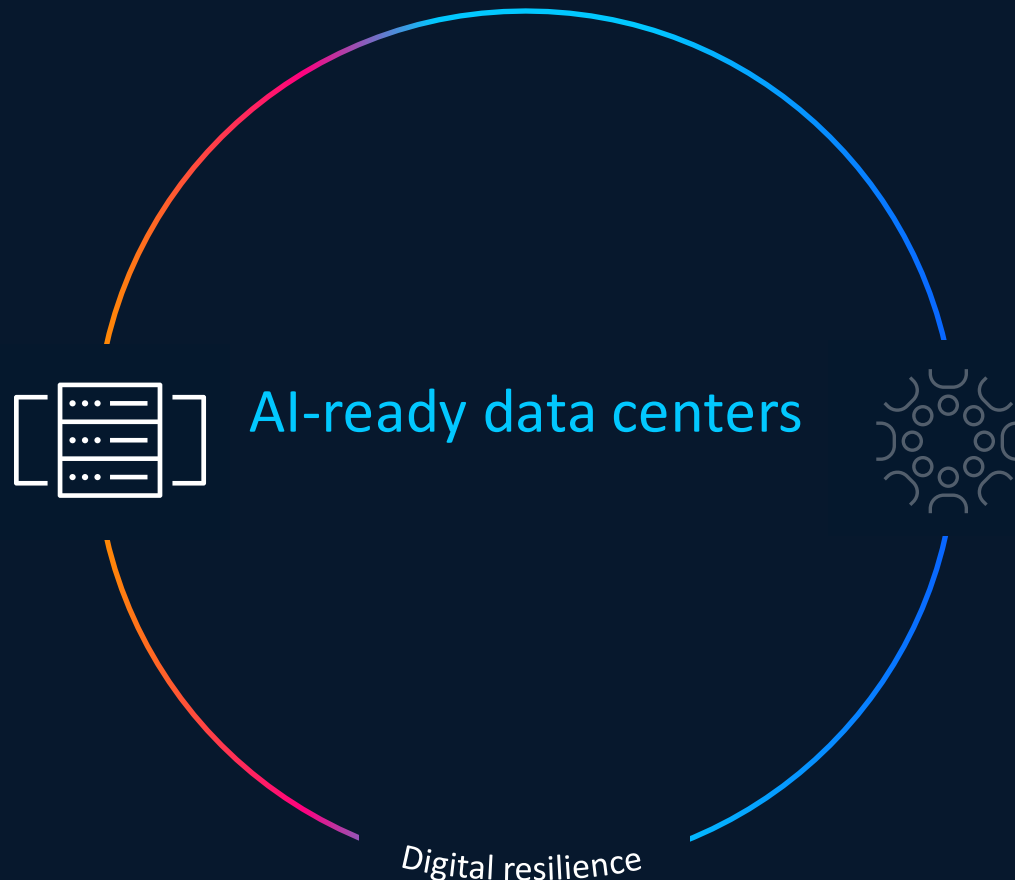
CISCO Connect

Thank you





Cisco delivers the critical components to modernize data centers



Full-stack infrastructure engineered as one system

- Cisco 6000
- Cisco 8000
- Nexus 9000
- Cisco MDS
- Smart Switch
- Silicon One
- Cisco Optics
- NVIDIA/Intel/AMD
- UCS-X
- UCS AI Servers
- Unified Edge
- AI PODs
- Hyperconverged Infrastructure
- Converged Infrastructure
- Virtualization
- Kubernetes
- Frameworks and automation

Distributed security fused into every layer

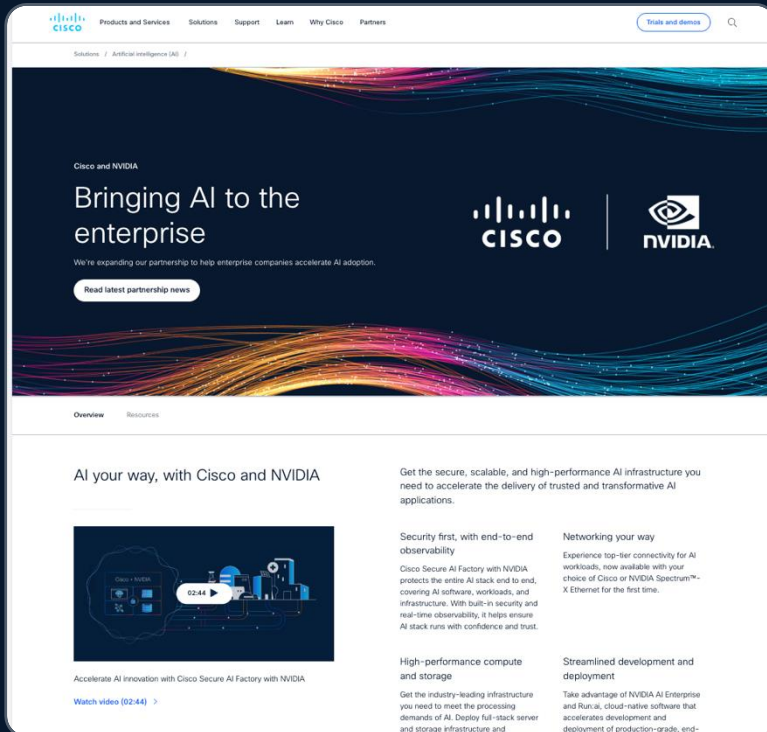
- AI Defense
- Hypershield
- Isovalent Enterprise
- Live Protect
- Secure Firewall
- Secure Workload
- Multicloud Defense
- Network segmentation
- Identify services engine
- Security Cloud Control

Agentic operations align speed with precision

- Cisco Cloud Control
- Intersight
- Nexus One
- ThousandEyes
- Splunk Observability Cloud
- Splunk Enterprise
- NVIDIA AI Enterprise & NIMs

* enhanced with CVDs

Resources



[SecureAIFactory.com](https://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, compliant with NVIDIA NCP RA](#)

NOW

Stack Automation by Quali

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

The screenshot displays the Cisco Cloud Control Marketplace interface. The top navigation bar includes links for Home, Intersight, Meraki, and Marketplace. The Marketplace section is titled "Apps that extend Cisco Cloud Control into your ecosystem" and features a search bar and filters. A grid of application blueprints is shown, each with a logo, name, provider, category, and a brief description. The blueprints are organized into rows and columns. The first row includes "Stack Automation" (highlighted with a red box), "Salesforce", "ServiceNow", "Auvik", "Forward Networks Mathematical", and "RunReveal". The second row includes "vmware Bare Metal ESXi 8.0 Ins...", "Bare Metal OpenShift...", "Bare Metal Proxmox V...", "Bare Metal Ubuntu 26...", and "Bare Metal Windows S...". The third row includes "Catalyst 8000v 17.18 o...", "Dataiku", "Intersight Assist", "NetBox Labs", and "Nexus Dashboard". The fourth row includes "ONTAP Select Deploy...", "Cumulo Core NAS (Kio...", "Splunk Enterprise 10.2.2", "Veeam Software Appl...", and "AI POD". The fifth row includes "AMD Enterprise Suite", "Baremetal RTX Pro 45...", "FlexPod", "Portworx by Everpure", and "VAST Data". Each blueprint card includes a "Connector" button and an "Add" button. The bottom of the interface shows the copyright notice: "© 2026 Cisco Systems, Inc."

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

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

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

Innovating in the Fabric: Cisco Intelligent Packet Flow

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

- Flow1* ● ● ●
- Flow2* ● ● ●
- Flow3^x ● ● ●
- Flow4 ● ● ●

* DLB enabled flow
^x QoS Policy flow

