

Cisco Partner Architect Exchange 2025

AI Ready Data Centers

Alistair Crawford | alicrawf@cisco.com
Technical Solutions Architect

Michael Petrinovic | mipetrin@cisco.com
Technical Solutions Architect



Agenda

- 01 Building for the AI Era
- 02 DC Networking
- 03 Fabric Computing
- 04 Ubiquitous Security
- 05 Digital Resilience
- 06 Wrap up

Building for the AI Era



Security

Enterprise Applications

Artificial Intelligence

Frameworks and Tooling

Application Platforms

Compute and Data

Networking

Observability



Data center



Edge

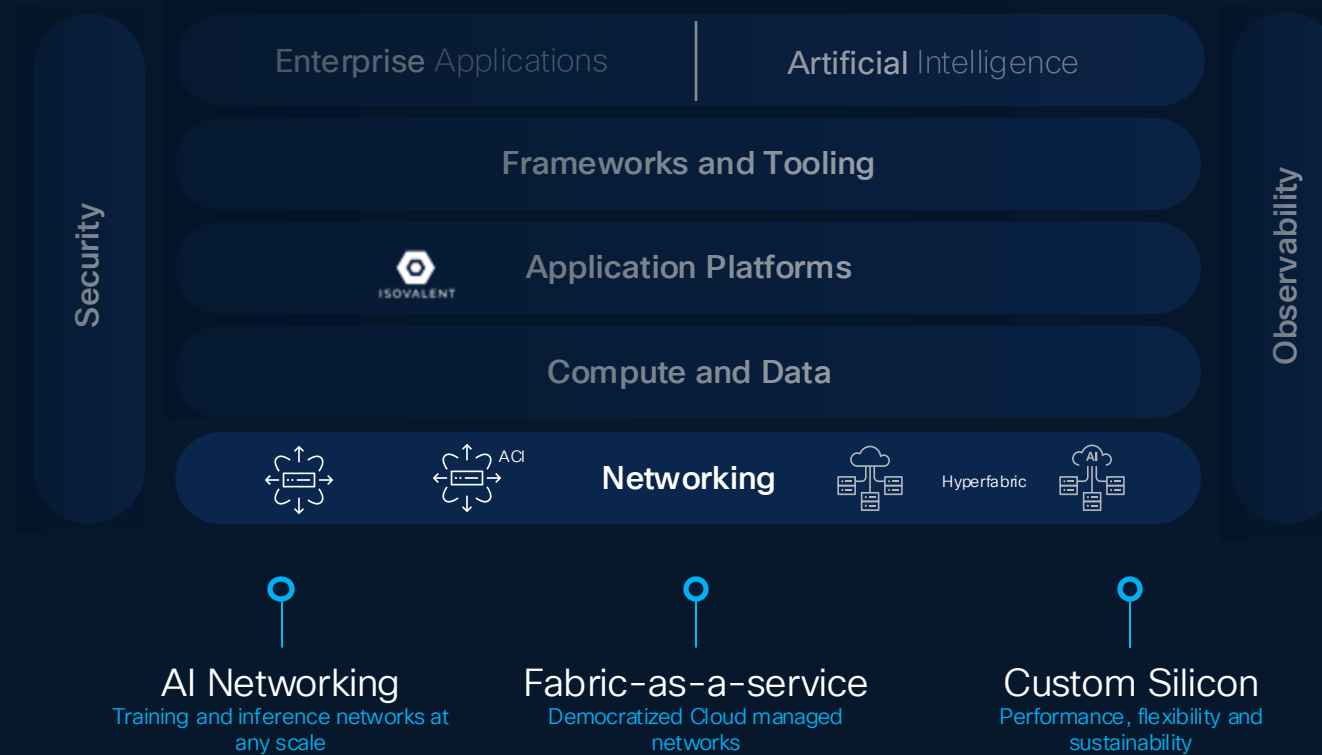


Colocation



Public cloud

Data Center Networking



NEXUS DASHBOARD

Physical or virtual cluster



Data plane | Policy and control plane | Management plane

NX-OS

ACI

Classic Ethernet

VXLAN

IP Media Fabrics

AI/ML Fabrics

Provision, Secure,
Manage and Analyze
your networks

Standards-based
ACI VXLAN EVPN
Gateway

Microsegmentation
for VXLAN fabrics

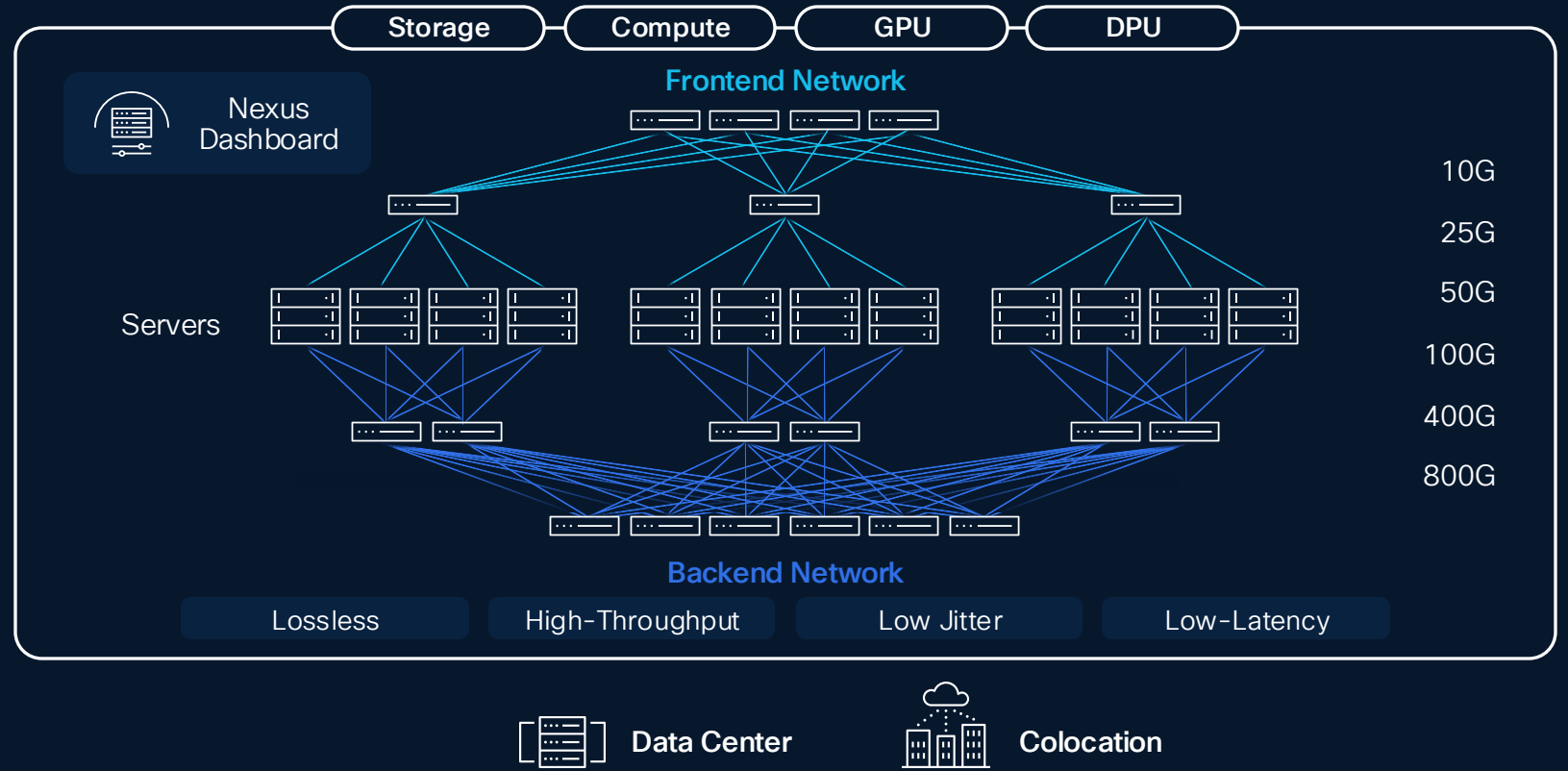
Interconnecting
NX-OS VXLAN and
ACI

Unified Architecture

Powering AI Fabrics

Optimized AI/ML Fabrics with Nexus

Silicon, Systems, Software, Operations



OUTCOMES

Best performing

AI/ML networks, front-end and back-end

Managed by Nexus Dashboard

AI fabric templates, AI analytics, telemetry, congestion scores

Cisco Intelligent Packet Flow

Advanced Load Balancing features

Intelligent buffering, low latency, RoCEv2

Lower TCO

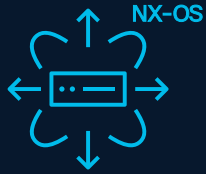
Validated designs for network and ecosystem partners

Data Centre Networking Portfolio



Nexus Dashboard

On-Premises Delivered



Powered by Nexus 9000 Series

Day 2 Ops

- Visibility
- Analytics
- Troubleshooting
- Compliance



Nexus Hyperfabric

Cloud Delivered



Powered by Cisco 6000 Series

Delivering networking at speed from the cloud



Design



Order



Deploy



Validate



Monitor



Upgrade



Collaborate





Cisco Silicon (S1 and Cloud Scale) & Cisco Optics fully validated as part of Spectrum-X
The only 3rd party switch validated as part of the reference architecture

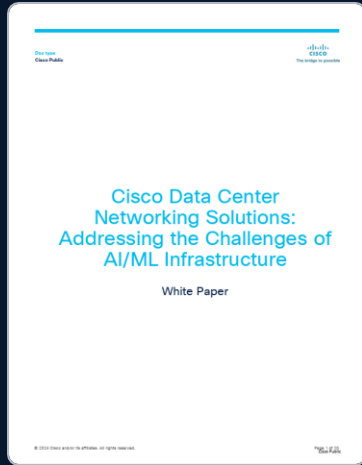
Support packet spraying and adaptive routing capabilities in Cisco Silicon Nexus Switches

Co-develop future Nexus based switches based on Spectrum-X ASIC

Get it right from the start: AI / ML design blueprints



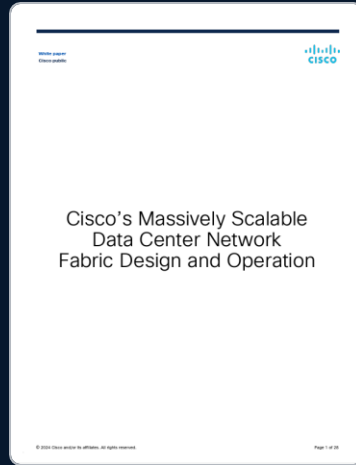
Understand the basics



Cisco Nexus 9000 Series

Solving inter-GPU networking challenges for scalable AI / ML performance

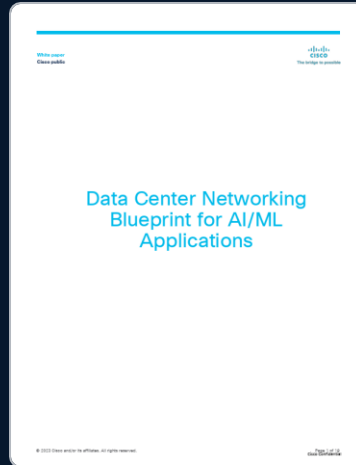
Design the network



Cisco Nexus 9000 Series

Designing scalable, automated Layer 3 IP fabrics for hyperscale data centers

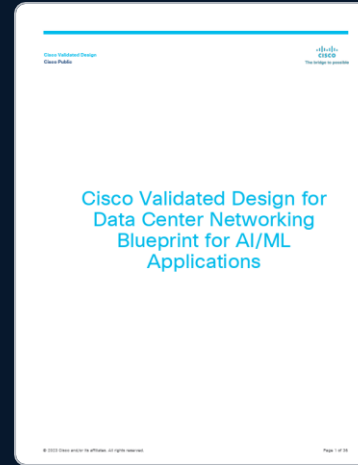
Configure the network



Cisco Nexus 9000 Series

Best practices for building for high-performance, lossless Ethernet networks for AI / ML workloads

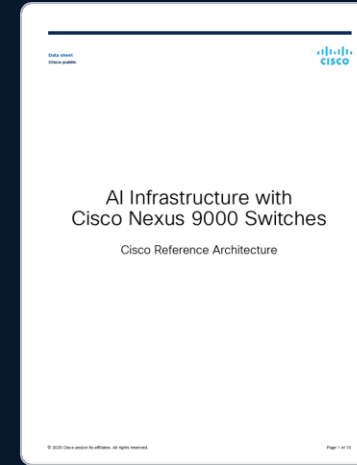
Validated design



Cisco Nexus 9000 Series

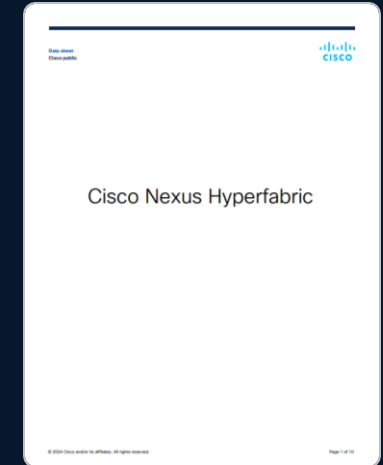
Validated configurations for building scalable, lossless Ethernet fabrics for AI / ML workloads

Enterprise Reference Architecture (ERA)



Cisco Nexus 9000 Series

Cisco UCS C885A
NVIDIA HGX H200
NVIDIA Spectrum-X



Cisco Hyperfabric

Cisco UCS C885A
NVIDIA HGX H200
NVIDIA Spectrum-X

Available:

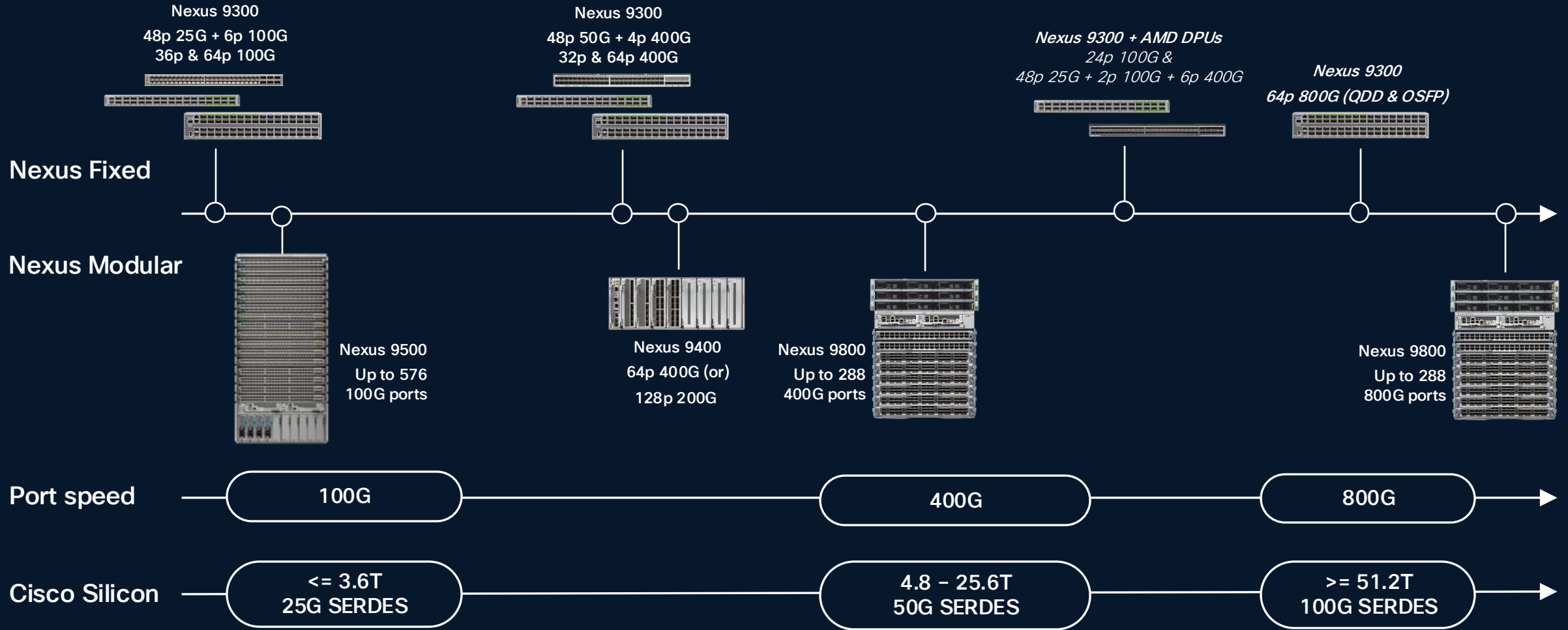
[Cisco Nexus 9000 switches for AI Clusters Performance Validation Insights \(Available Now\)](#)

CVD Cisco Nexus 9000 switches with NVIDIA DGX for AI Clusters (Q4FY25)

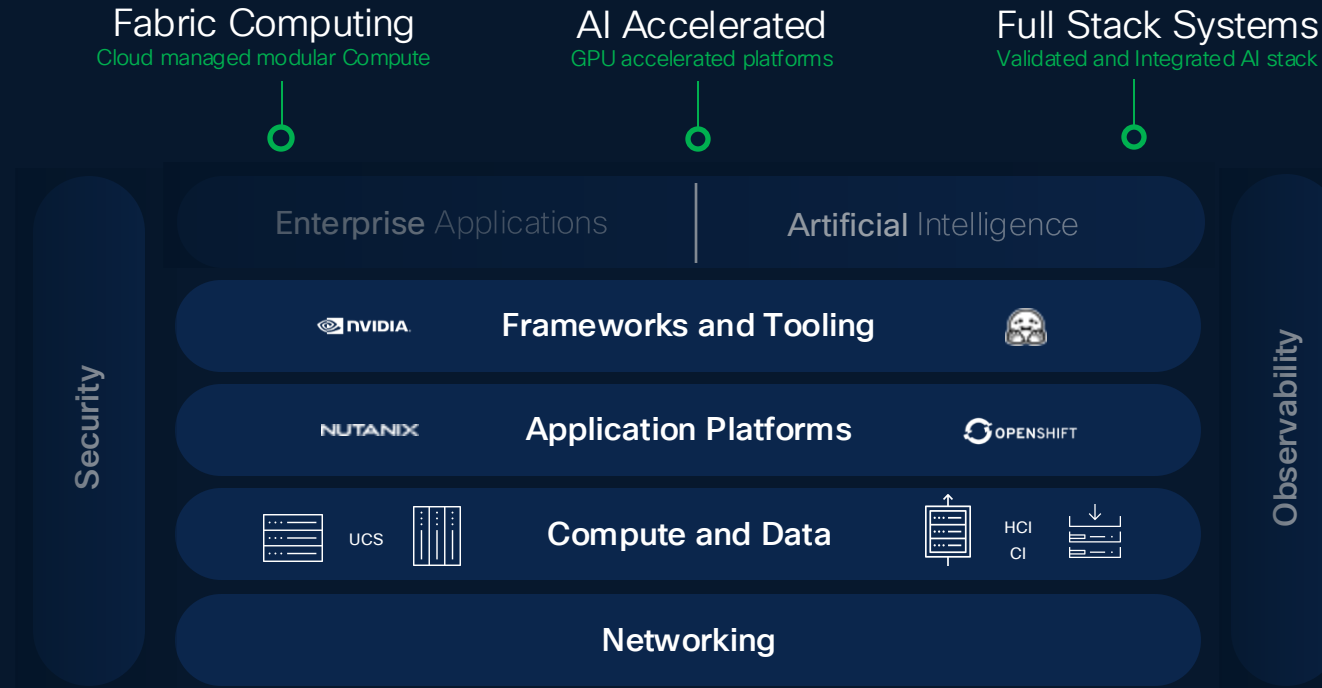
Cisco provides consistent network ops for all GPU vendors - NVIDIA, AMD, Intel



Cisco Nexus 9000 Series Switches



Fabric Computing

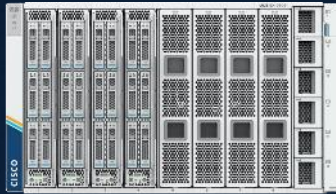


Cisco UCS Compute Portfolio

MAINSTREAM ENTERPRISE SERVERS

UCS X-Series
X9508 Chassis

IFM Module



UCS X-Series Direct



UCS X210c M7



UCS X210c M8



UCS X410c M7



UCS B200 M6



UCS X215c M8



UCS C240 M8E3S
36 EDSFF E3.S1T



UCS C240 M8SX
28 HDD/SDD/NVMe



UCS C240 M8L
16 LFF + 4 SFF



UCS C240 M7SN
28 NVMe



UCS C240 M6S
14 SSD/HDD Media drive



UCS C240 M6N
14 NVMe Media Drive



UCS C220 M8E3S
16 EDSFF E3.S1T



UCS C220 M8S
10 HDD/SSD/NVMe



UCS C220 M7N
10 NVMe



UCS C245 M8SX
28 HDD/SDD



UCS C225 M8S
10 HDD/SSD



UCS C225 M8N
10 NVMe

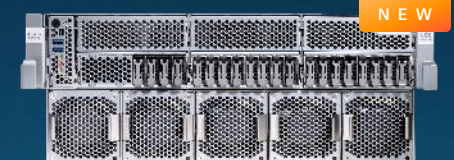


AI SERVERS

UCS C885A M8
8RU Dense GPU Server



UCS C845A M8
4RU MGX Server





Unified Solutions

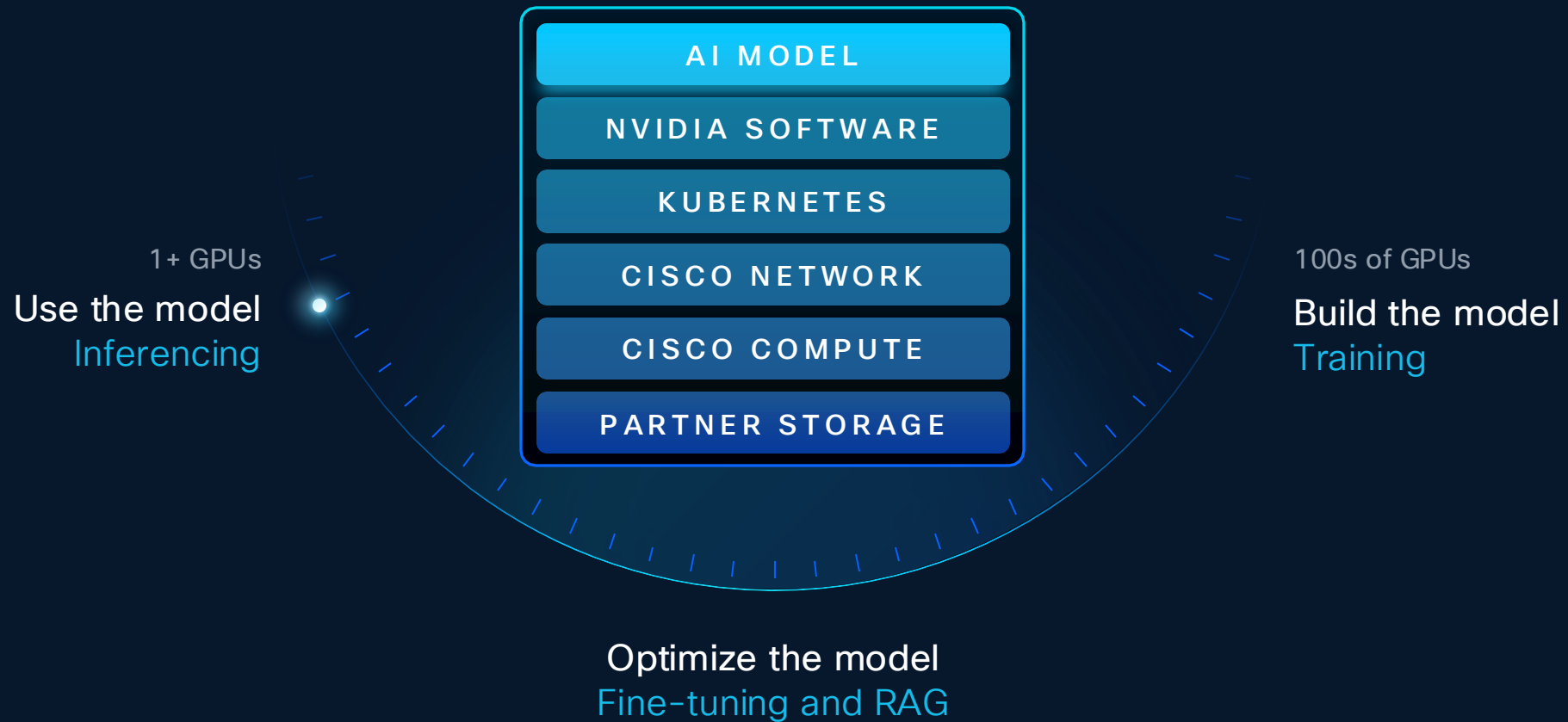
The power of partnering with industry-leading tools



NEW

Introducing **AI PODs** for entire AI lifecycle

Scalable architecture, support any AI workload

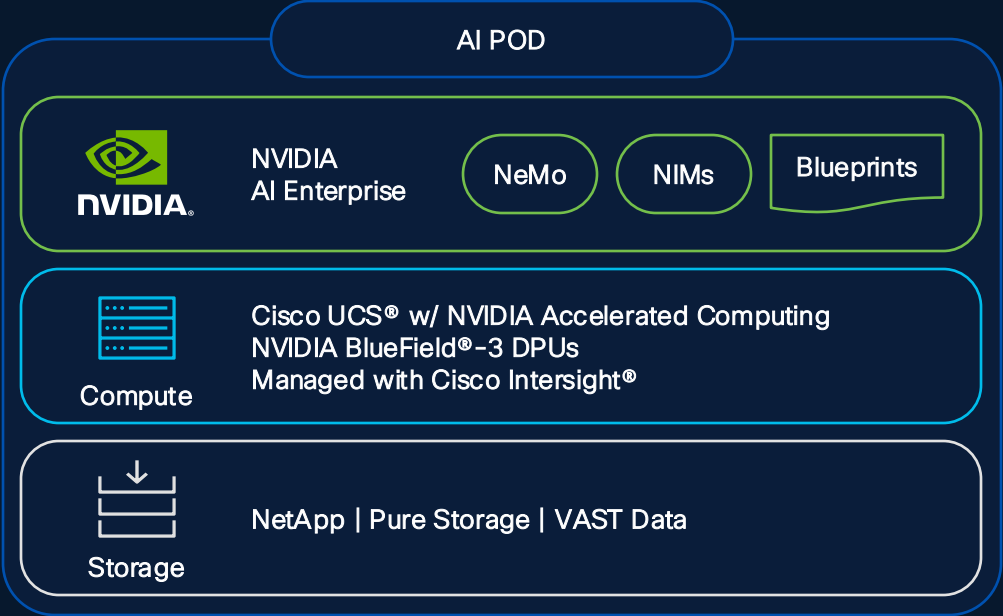


Customizable
Full Stack

Based on Validated
Designs

Meet customers
where they are

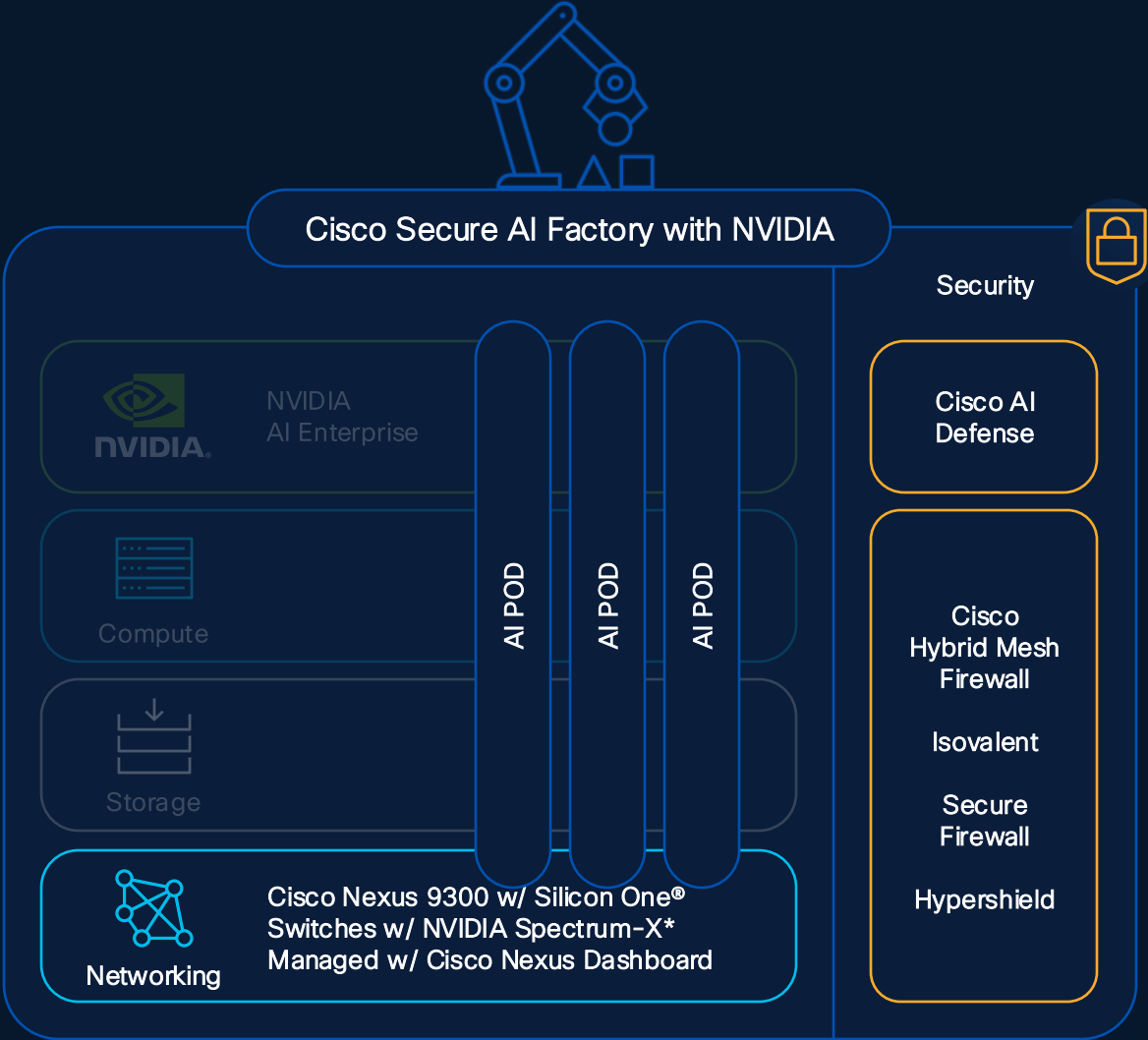
Flexibility of choice: AI Pod to Secure AI Factory



Atomic unit of Secure AI Factory with NVIDIA

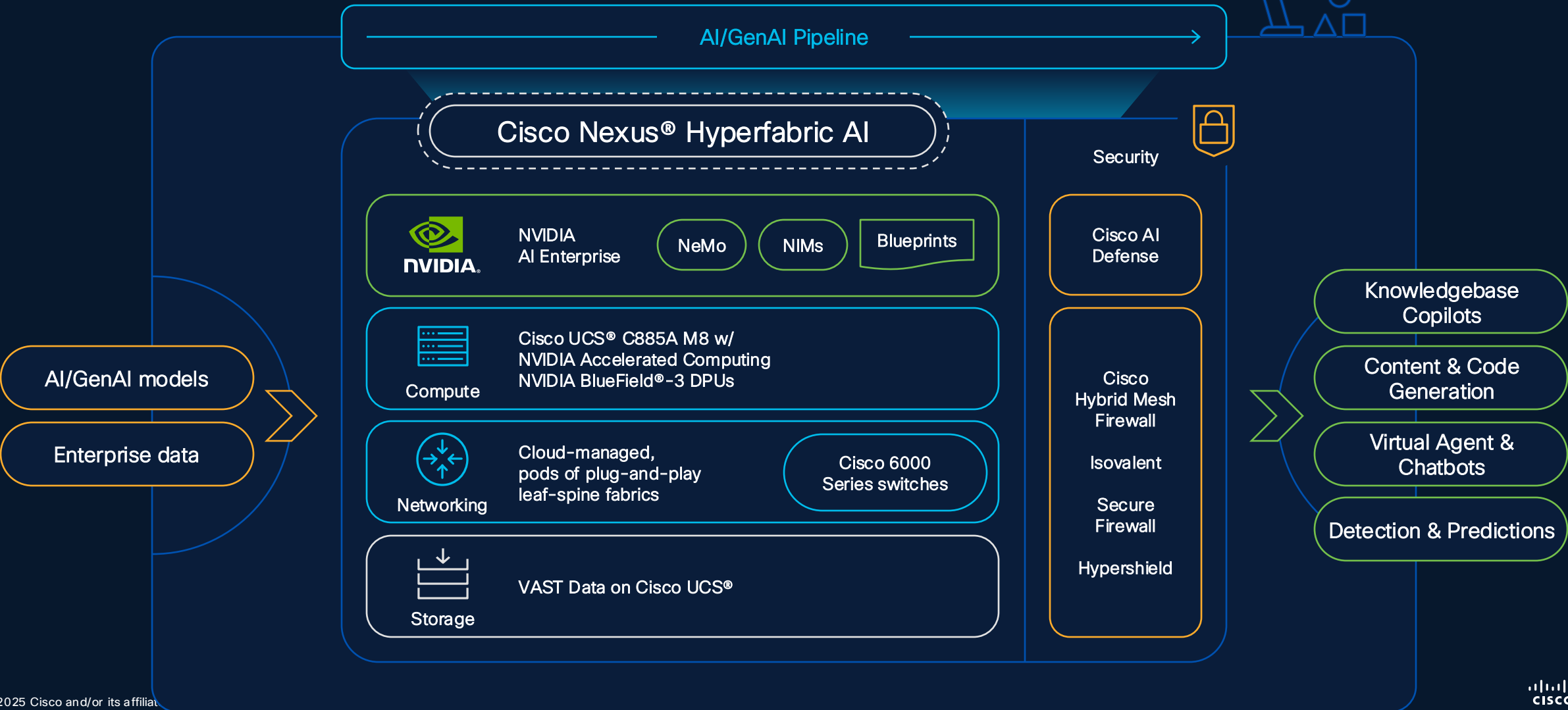


Scale the Secure AI Factory to meet your business needs



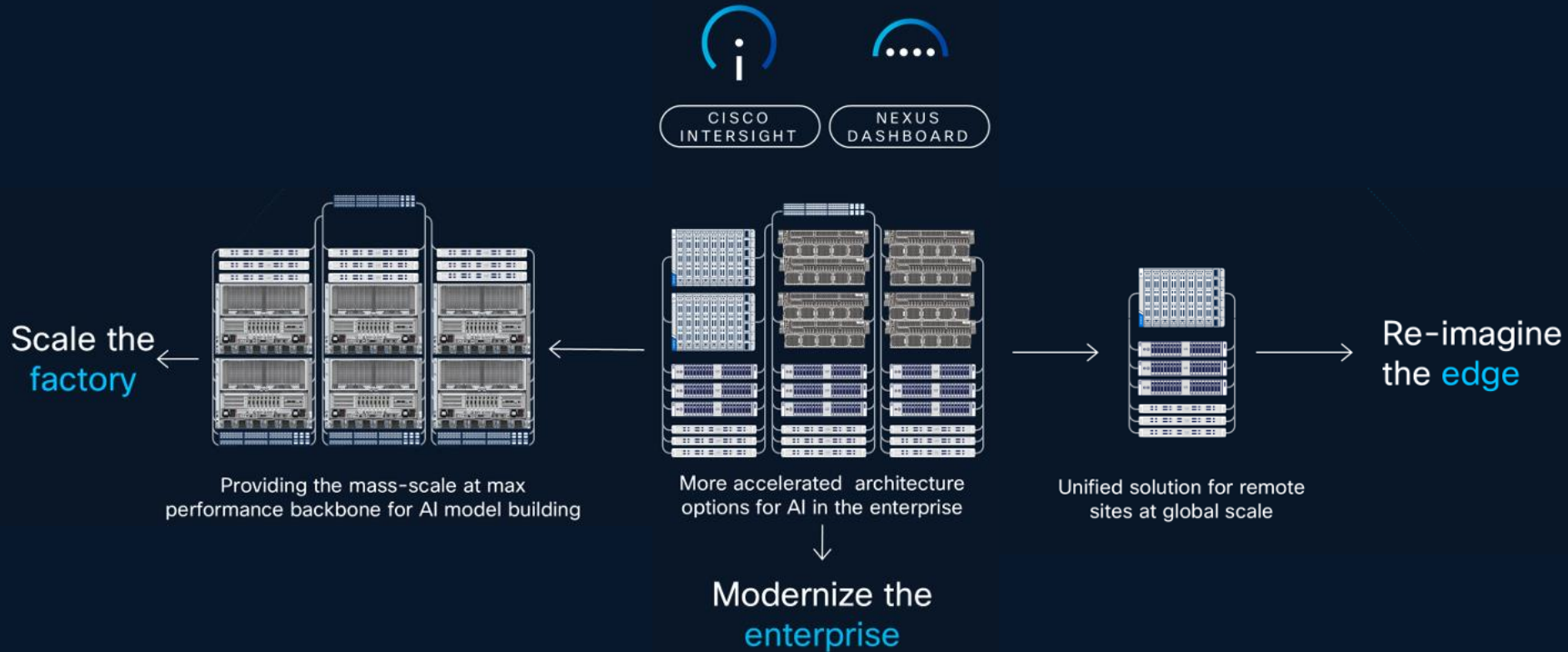
Vertically Integrated System

Cisco Secure AI Factory with Nvidia



Cisco AI Pods - Expansion

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

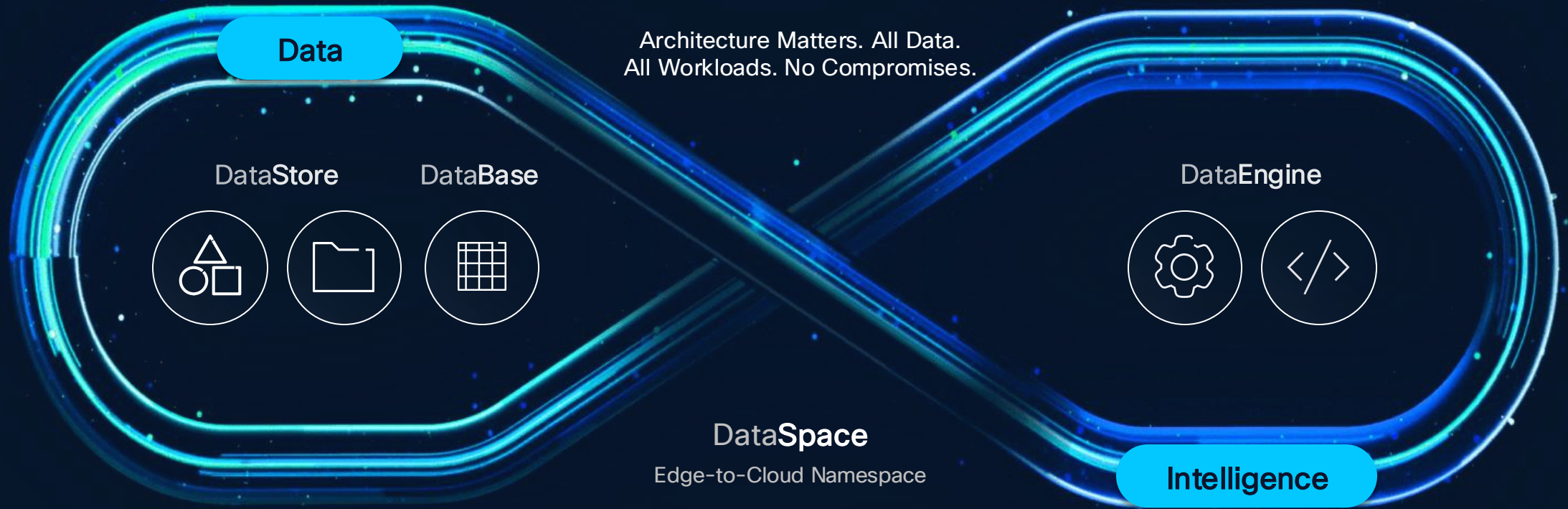


VAST data

The data platform for the AI era

VAST Data Platform

Architecture Matters. All Data.
All Workloads. No Compromises.



Unified data architecture



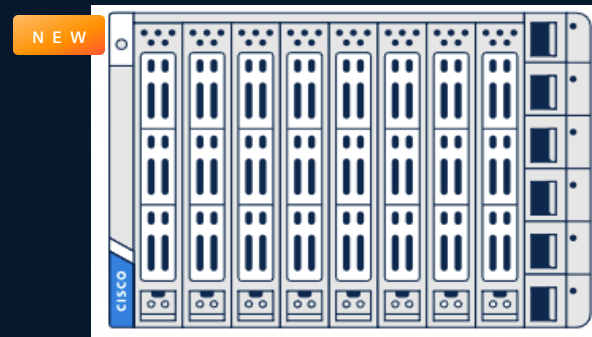
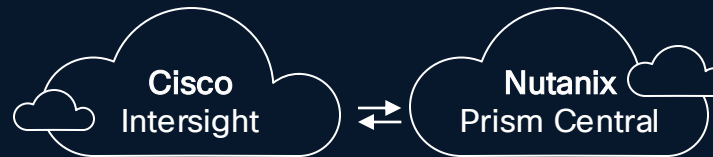
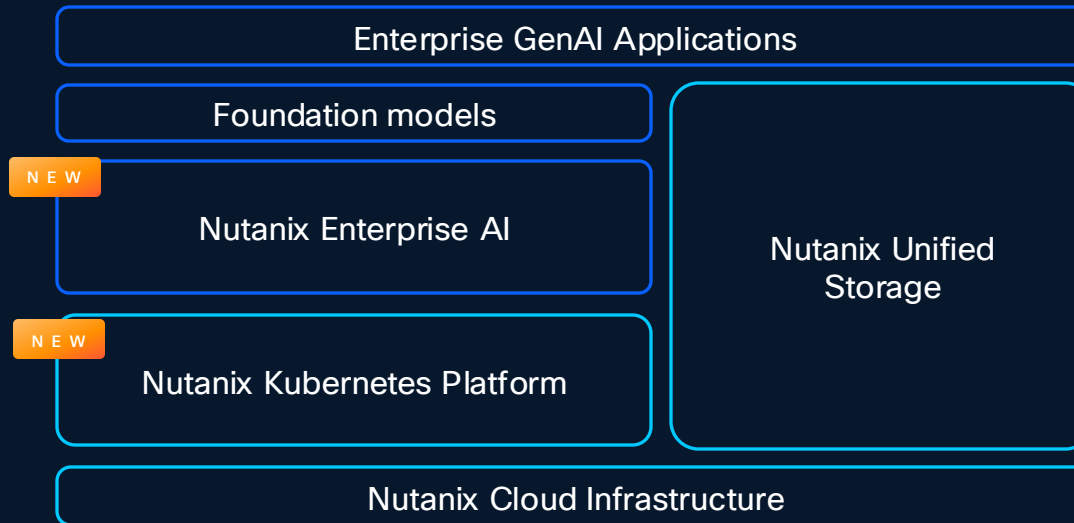
Real-time performance



Enterprise-grade
security & compliance

Nutanix GPT in a Box 2.0 on Cisco

On-premise GenAI solution on Cisco with Nutanix CVD



Cisco UCS

KEY BENEFITS

HCI Turnkey solution on Cisco with GPU enabled nodes

Generate secure LLM Inference end points with Role-based access controls

Accelerated deployment and continuous integration

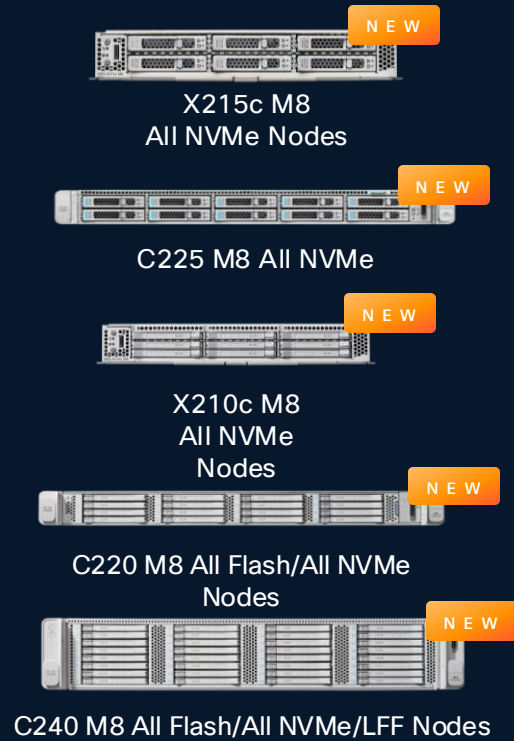
Generative AI stack

Enterprise grade AI and Kubernetes platforms

Software stack offers

- Hugging Face
- NVIDIA NIM
- BYO Model

The data platform for Data Centre Modernisation

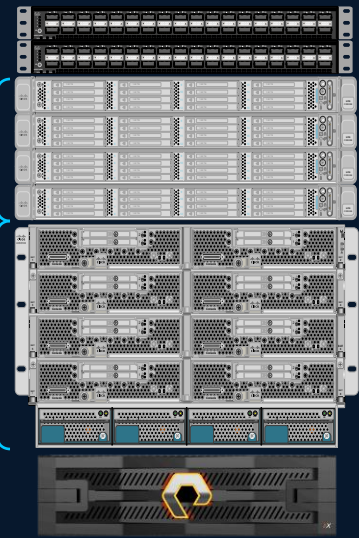


Available Late 2025

Available Late 2025

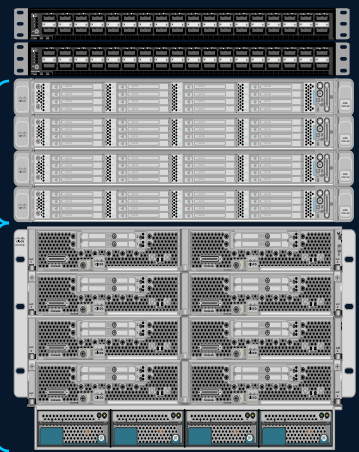
UCS HCI C-Series
UCS HCI X-Series
with Nutanix AHV

Scale with
existing B200
M5/M6 UCS
Blades



New or
existing UCS
HCI w/
Nutanix
(AHV) cluster

Scale with
existing B200
M5/M6 UCS
Blades



Single UCS
HCI with
Nutanix
(AHV)
cluster

Scaled
without
storage
resources

New Hardware Paradigms

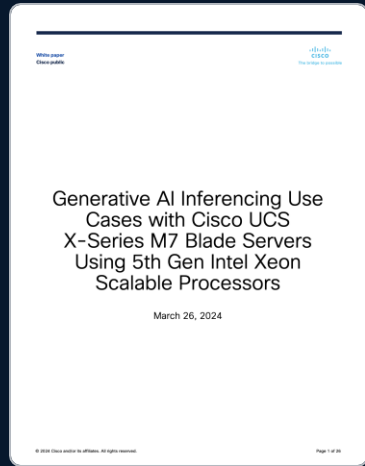
Flash Stack with Nutanix AHV

Compute Node Scaling

Get it right from the start: AI / ML design blueprints



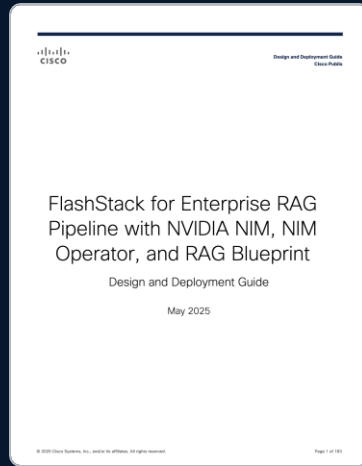
AI with Intel CPU's



Cisco UCS-X Series

Understanding AI use cases addressed using Intel CPU's

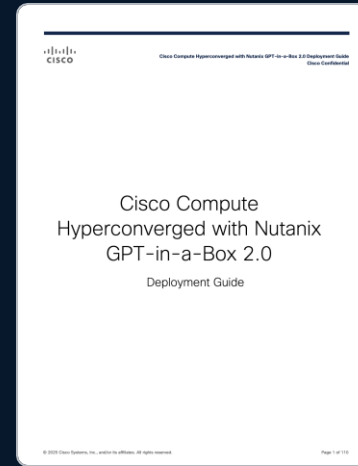
Retrieval-Augmented Generation (RAG) Blueprints



Cisco UCS-X and Pure

FlashStack for Enterprise RAG Pipeline with NVIDIA NIM, NIM Operator, and RAG Blueprint

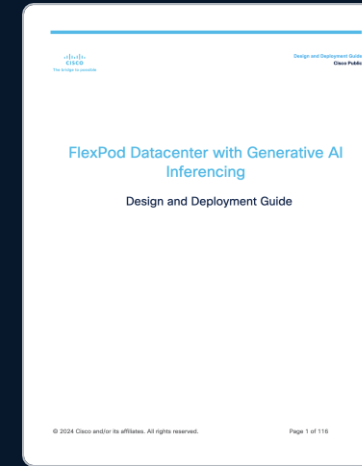
GPT in a Box 2.0



Cisco HCI with Nutanix

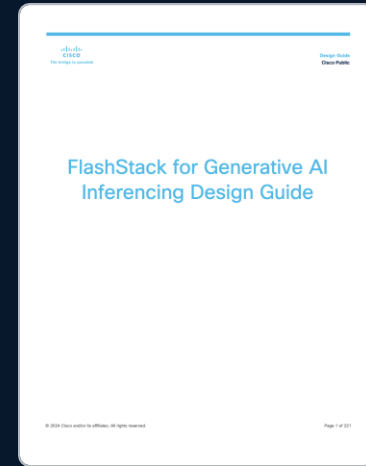
Validated configurations for building a Full Stack AI solution on Nutanix

Enterprise Reference Architecture (ERA)



Cisco UCS-X and Netapp

FlexPod with Red Hat OpenShift and NVIDIA AI Enterprise



Cisco UCS-X and Pure

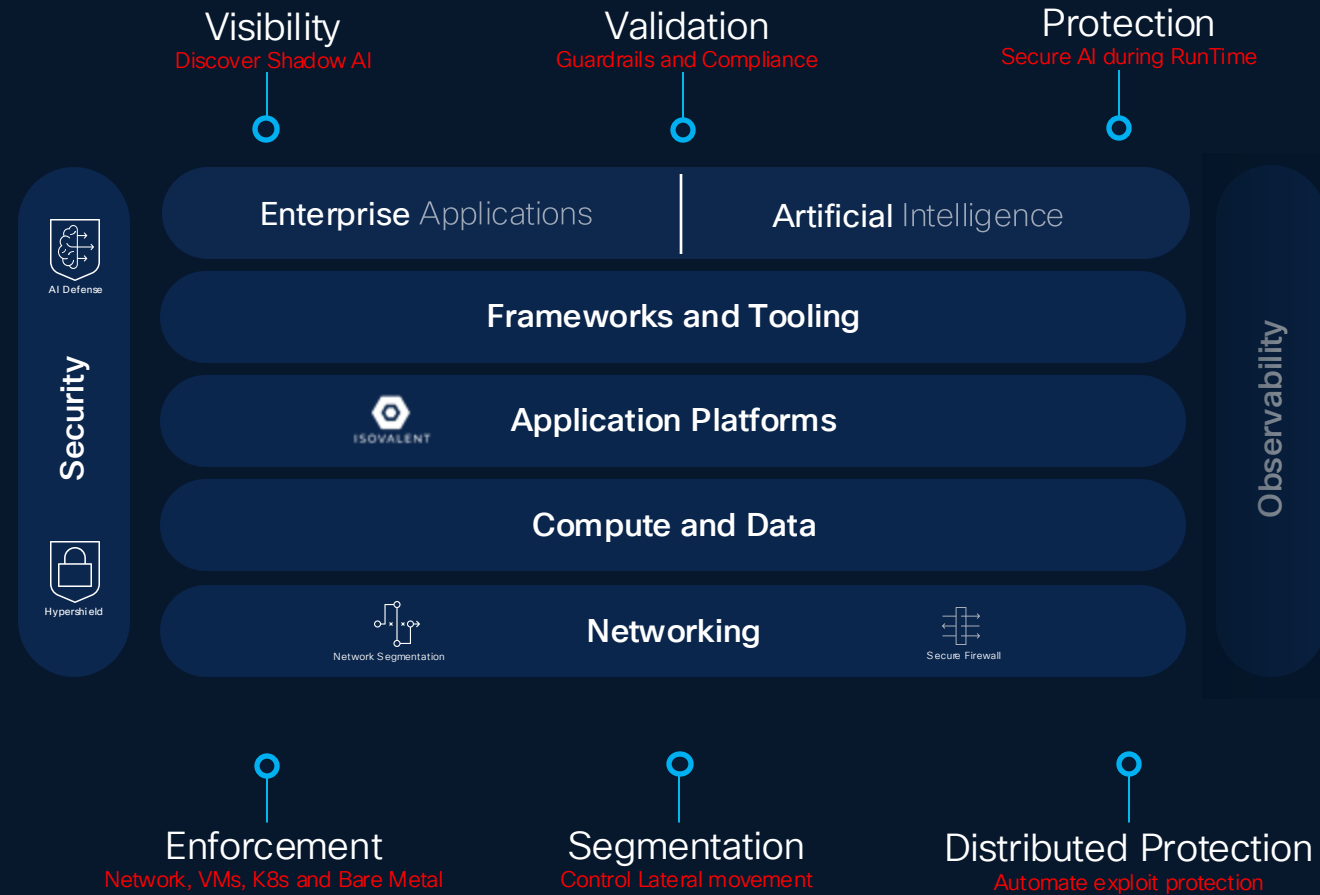
FlashStack with Red Hat OpenShift and NVIDIA AI Enterprise

Available:

Design guides for AI-ready infrastructure (Available Now)



Ubiquitous Security



MACROSEGMENTATION

MICROSEGMENTATION



Dev

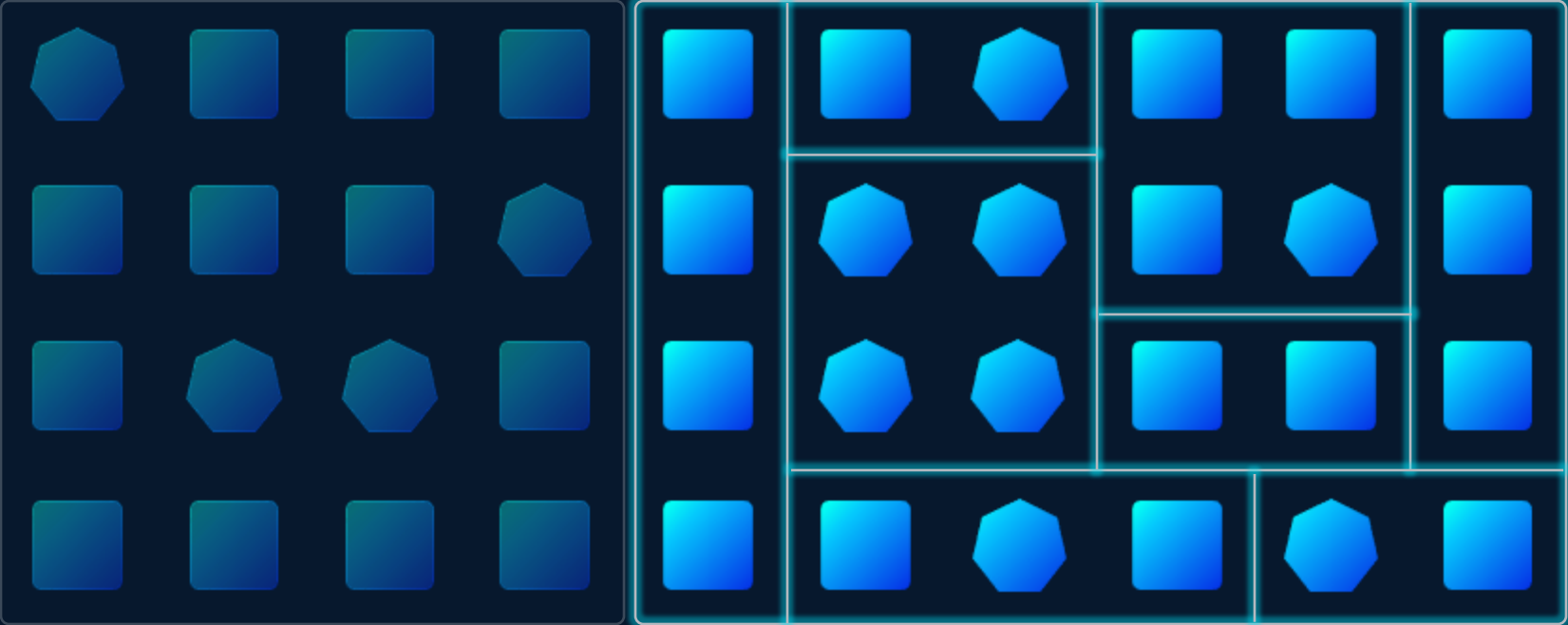
Prod

Network Constructs

VLANs and VRF as boundaries for segmentation

MACROSEGMENTATION

MICROSEGMENTATION



Dev

Prod

Network + Security Constructs

Security Groups
ACI (EPG/ESG) and NXOS (VXLAN-GPO)

MACROSEGMENTATION

MICROSEGMENTATION



Dev

Prod



Traffic steering and redirection

Send traffic to firewalls, load balancers and other L4-L7 services or service chains by using Policy-Based Redirection (PBR)



Monitor and failover L4-L7 services

Detect node health changes and failover as needed, allowing load-share and selective redirection



Smart Switch

Smart Switch



Power

Software licenses RENEW ALL Optics

Support contracts RENEW ALL

Rack Space



Introducing Cisco N9300 Smart Switches

NEW



Network + Security
in one switch



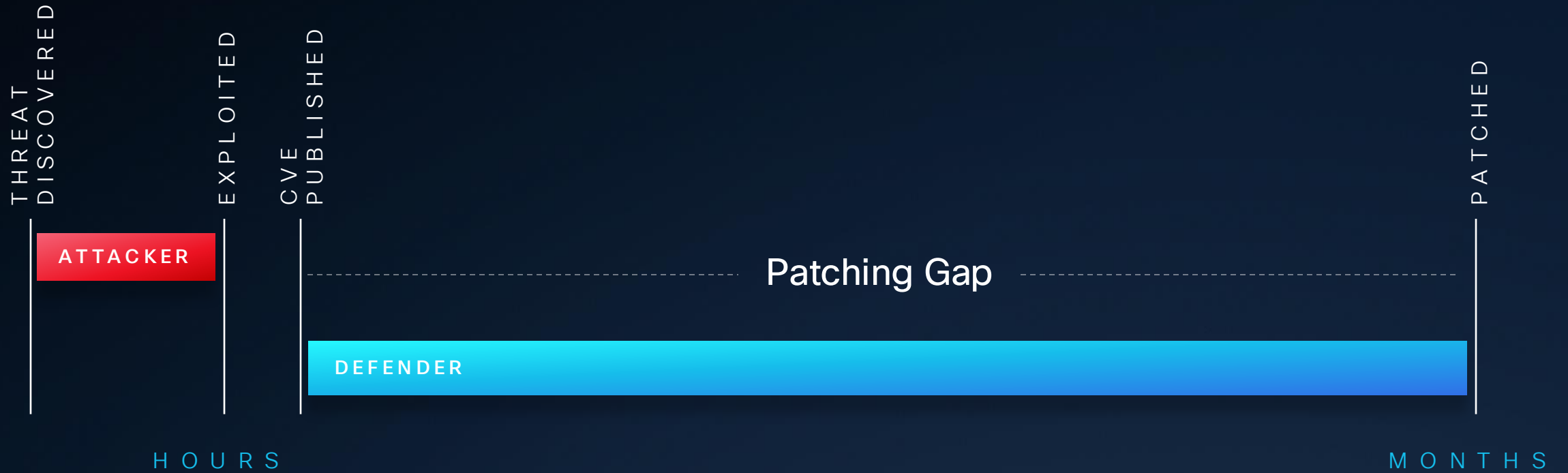
Separate workflows and
separate data flows for
networking and security



Up to 84% TCO savings

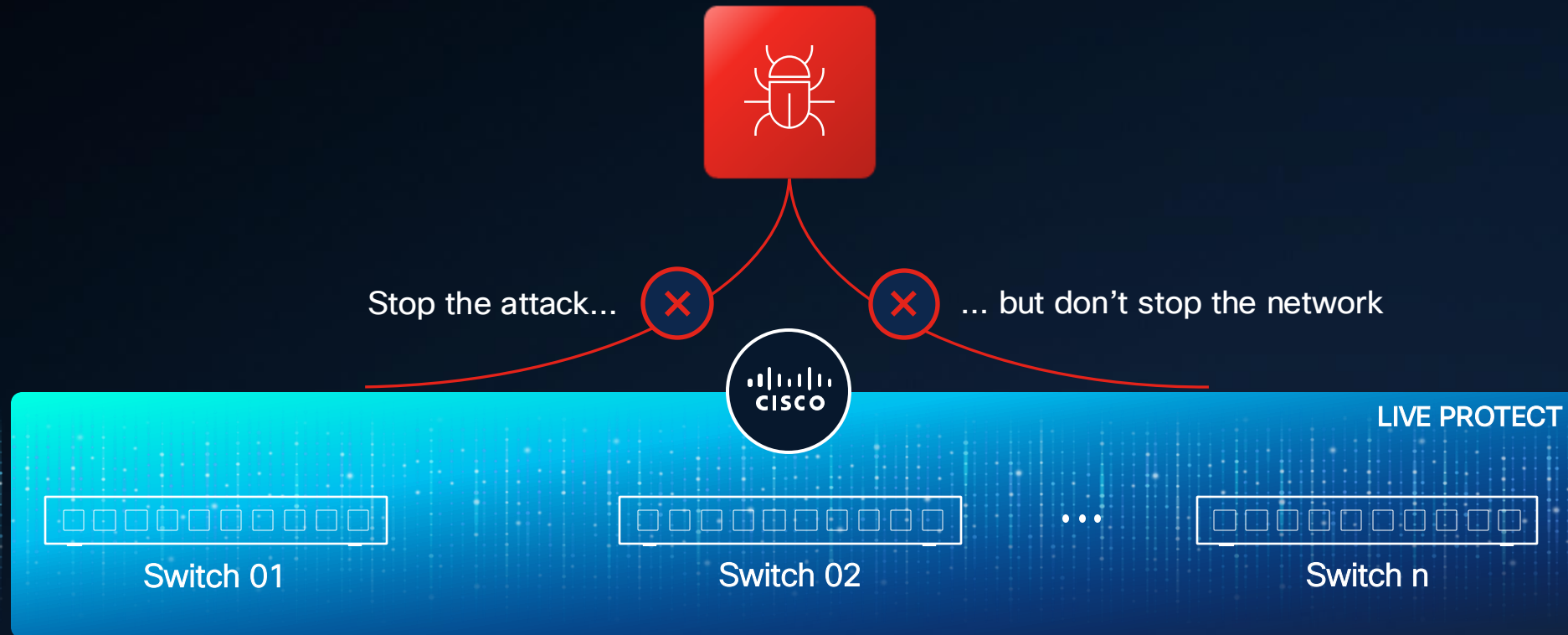


Patching is hard



Live Protect

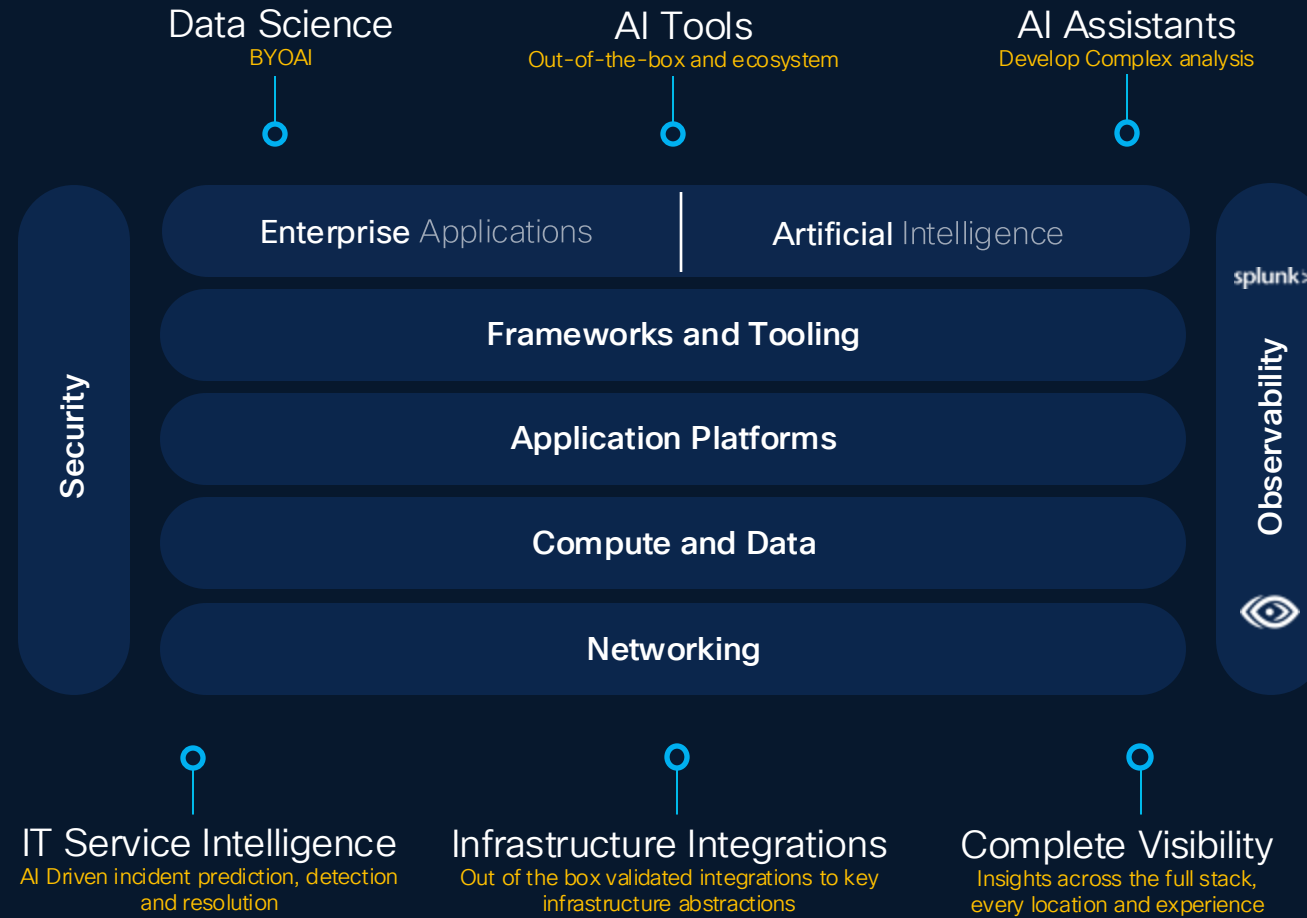
Vulnerability shielding for Cisco **all** Networking devices



Based on Isovalent's ability to intervene and block exploits with kernel-level compensating controls

Available for NX-OS September 2025

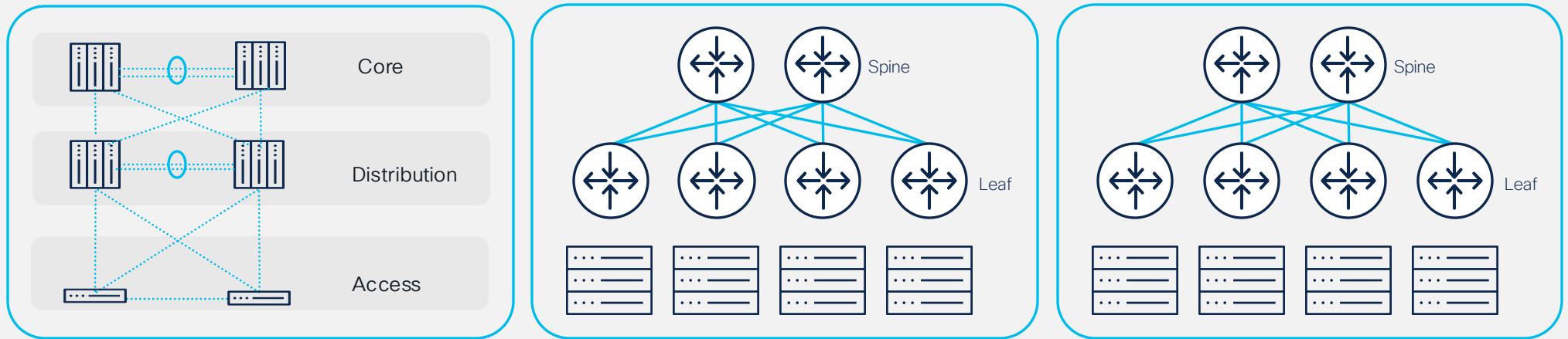
Digital Resilience



Unified Operations



Standardization with Nexus Dashboard



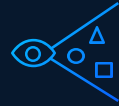
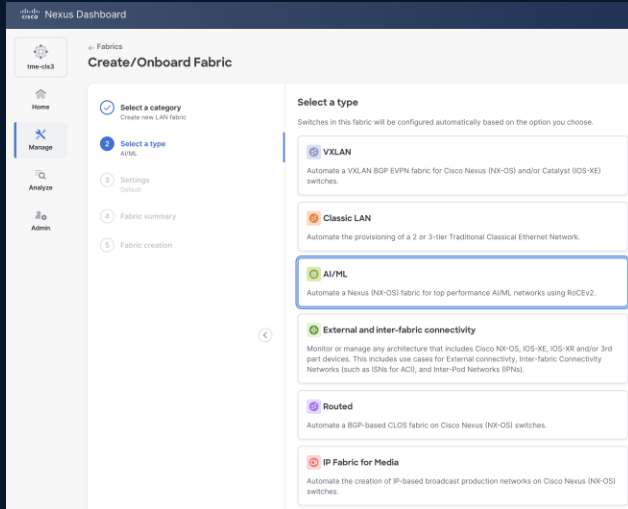
Existing Enterprise DC Pod (Front End)
ACI, Classic LAN, VXLAN/EVPN, L3

New AI/ML Pod (Back End)
NXOS - L3, VXLAN

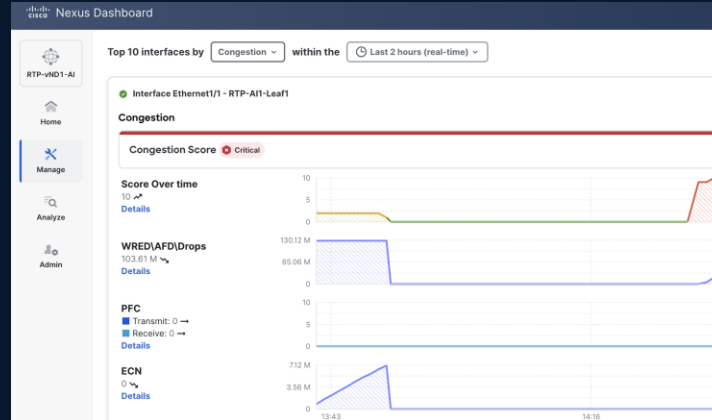
Operational Readiness for all your networks



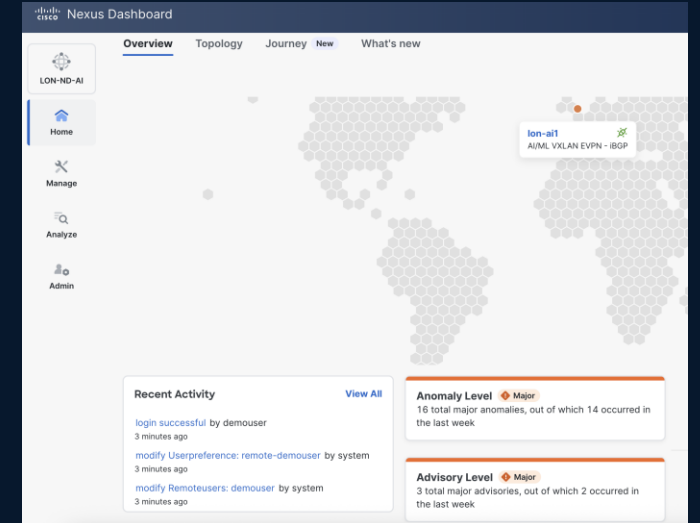
Quickly deploy any Data Center network fabrics from traditional LAN, VXLAN/EVPN, AI/ML, Media, Packet Broker, SAN environments



Real-time telemetry, traffic analytics, anomaly and advisories to proactively resolve network issues and ease troubleshooting



Unifies front-end and back-end network operations for consistent, simplified network management



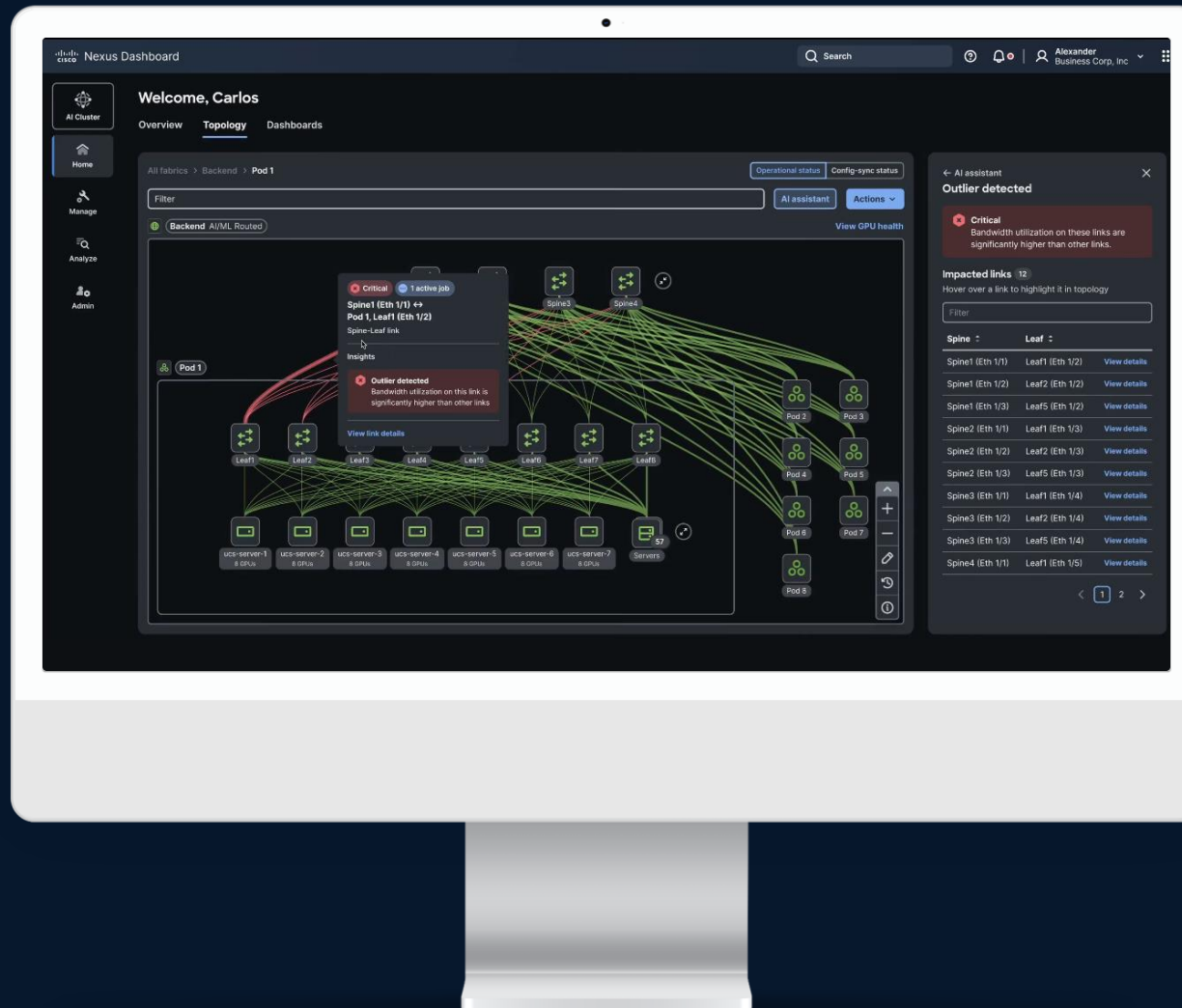
AI Job Observability

End-to-end visibility

Monitor AI workloads across the entire stack. Track network, NICs, GPUs, and distributed compute nodes.

Topology-aware visualization

Correlate network elements with hardware components. See connections between infrastructure and AI job health.



Real-time metrics

Track throughput, latency, and GPU utilization. Get actionable insights into distributed AI workloads.

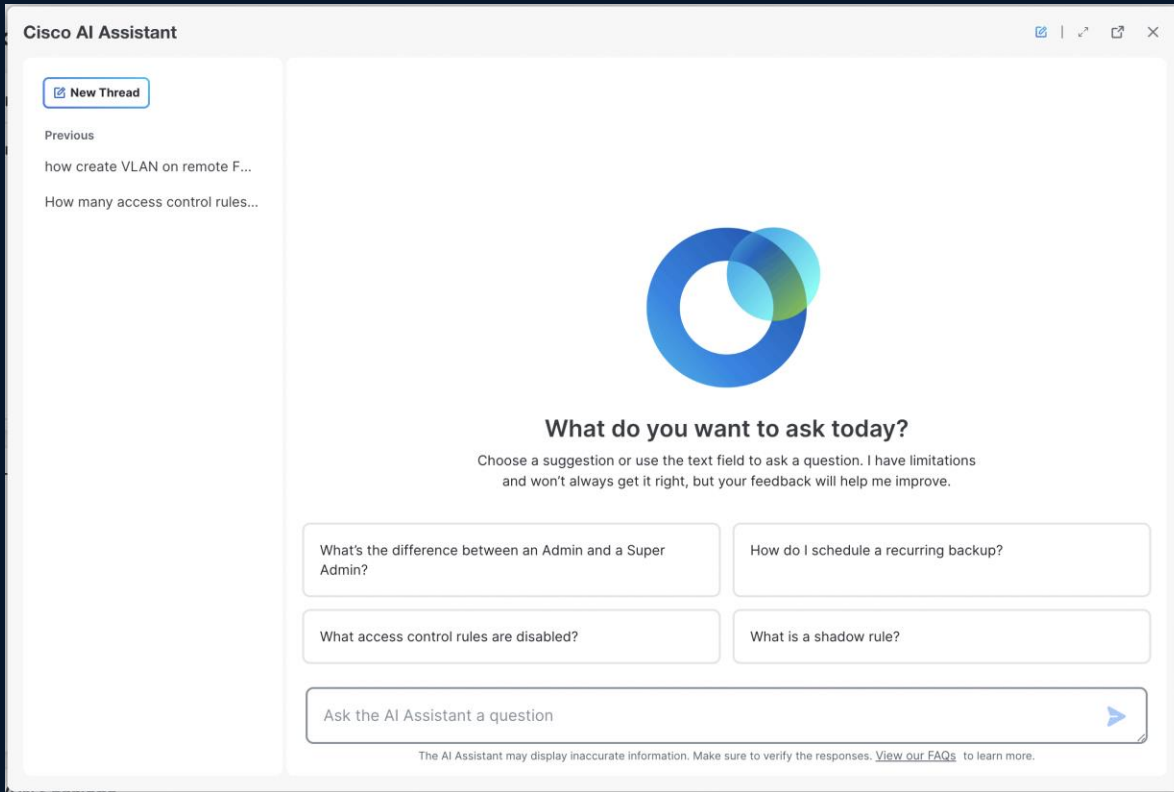
Proactive troubleshooting

Detect anomalies against historical baselines. Accelerate root-cause analysis with automated recommendations.

Nexus Dashboard

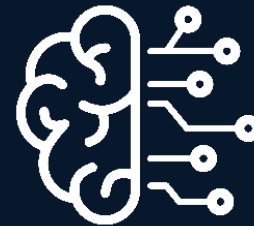
Planning

Cisco AI Assistant



Interact with Nexus Dashboard and other Cisco technologies

GenAI



Ask questions

Run commands

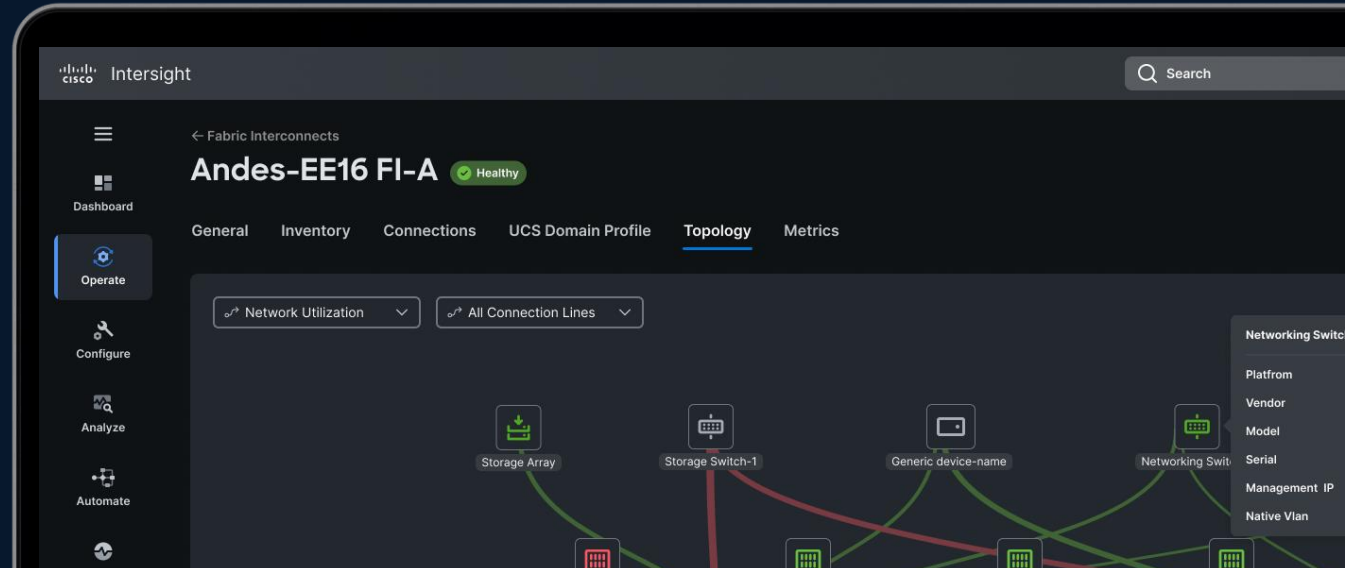
Display paths

Get help from TAC

Generate configurations

Benefits

Unified Intelligence with Intersight



Next-gen metrics

Uncover hidden trends, optimize performance, and supercharge operational awareness

AVAILABLE NOW

Enhanced topology view

Gain a holistic perspective of UCS domains for clarity, context, and control

AVAILABLE NOW

Complete visibility and assurance

Gain deeper visibility into server health and performance, and ensure high-speed networking with proactive monitoring

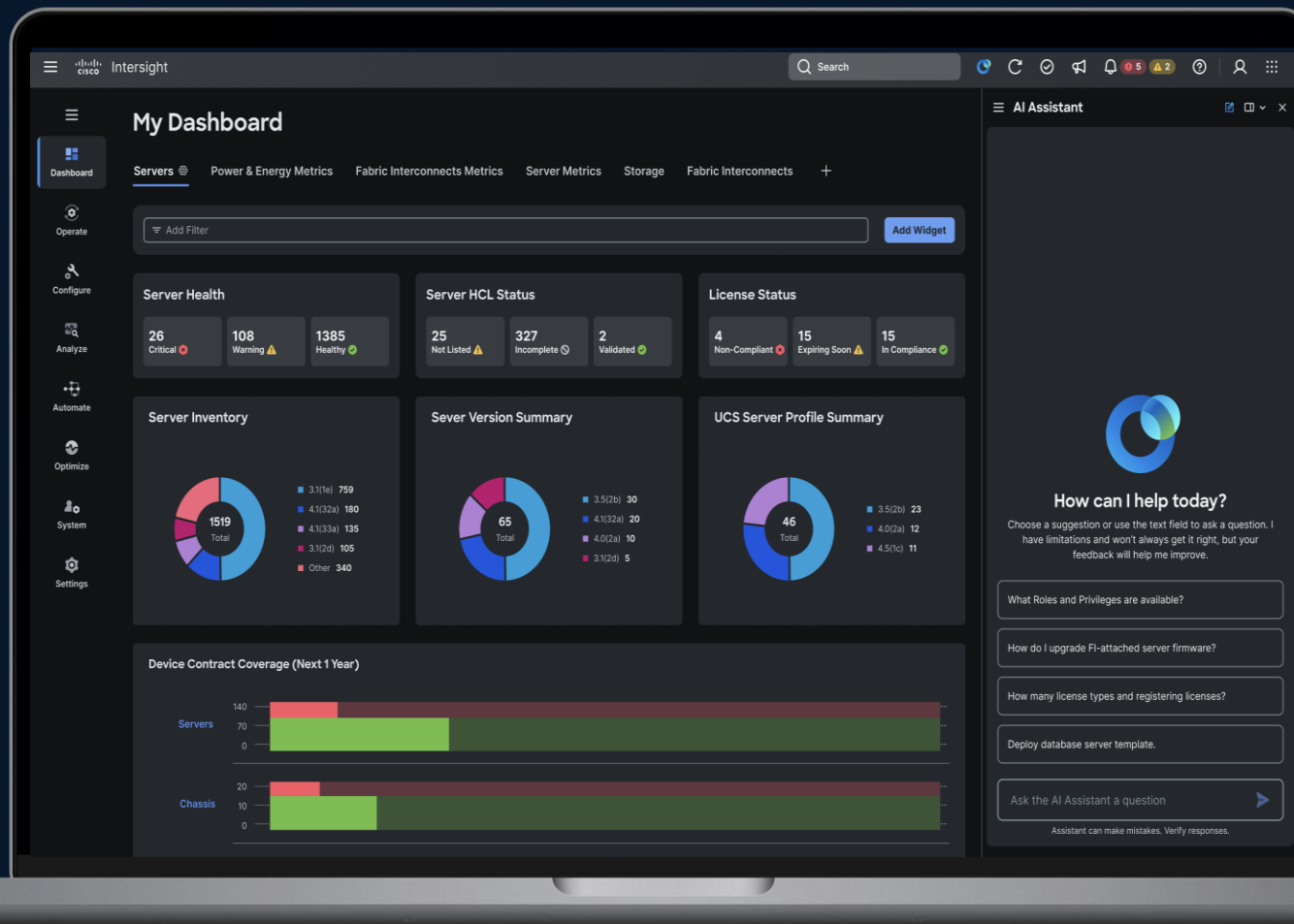
COMING SOON

Unified Intelligence with Intersight AI Assistant

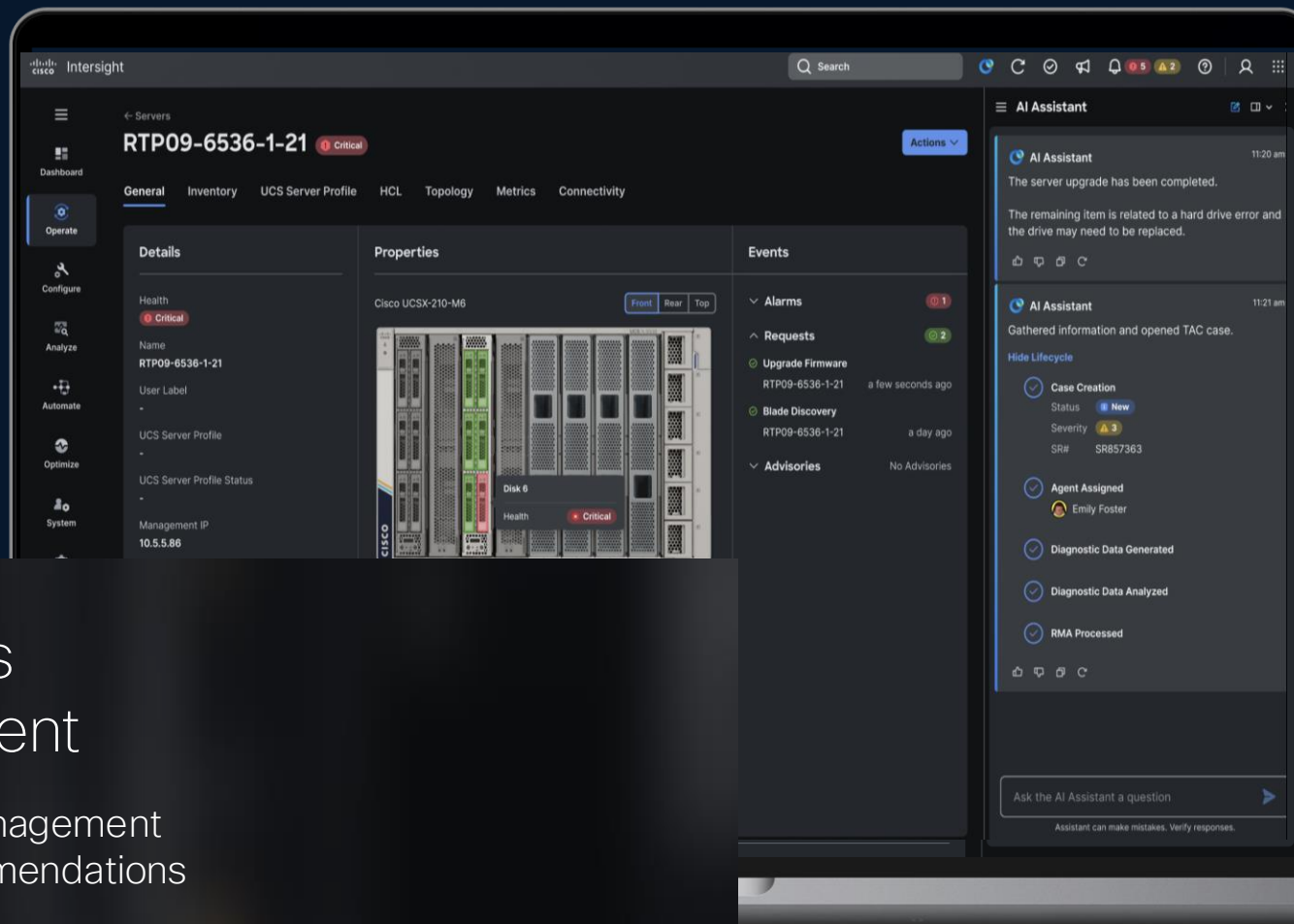
AI Assistant GA

TAC Agent Integration
CVE Insights and Explanation
Contract Compliance

Target Q4 CY25



Unified Intelligence with Intersight AI Assistant



AI Assistant GA

TAC Agent Integration
CVE Insights and Explanation
Contract Compliance

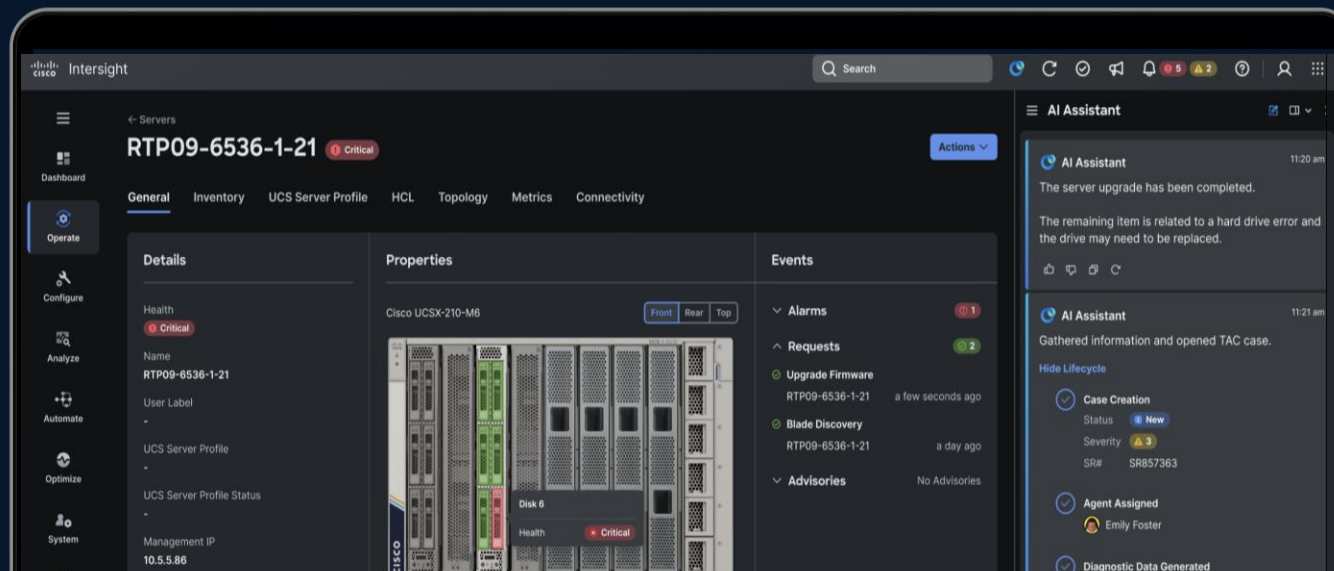
Target Q4 CY25

Continuous Enhancement

TAC Lifecycle Management
Firmware recommendations

Target H1 CY26

Unified Intelligence with Intersight AI Assistant



AI Assistant GA

TAC Agent Integration
CVE Insights and Explanation
Contract Compliance

Target Q4 CY25

Continuous Enhancement

TAC Lifecycle Management
Firmware recommendations

Target H1 CY26

Agent Driven Action

Firmware
Profiles

Planning

Wrap Up

Cisco AI-Ready Data Center Building Blocks

Cloud

Protected

Nexus 9000



Data Center Networking

Nexus 9000 | NXOS | ACI
Silicon ONE | Nexus Dashboard

Hyperfabric



Cloud managed
Network fabric

Nexus 6000 | SONIC
Silicon ONE

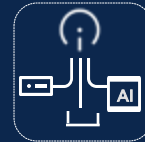
Unified Computing
System



Cloud managed
compute fabric

Modular | Rack
GPU Accelerated | Intersight

AI Pod



Validated hardware and
software AI system

Compute | Network | Storage
OpenShift | NVAIE

Hyperfabric AI



Cloud managed
integrated AI System

Nexus 600 | SONIC | Silicon
ONE | NVIDIA GPU/DPU | VAST

Secure AI
Factory



Secure and integrated AI
full stack system

AI Defense | Hybrid Mesh Firewall
Hyperfabric AI | AI Pod

AI READY INFRASTRUCTURE

AI FULL STACK SYSTEMS

Digital

Resilience

Thank you



