Building the AI Internet: Cisco's Foundational Infrastructure for AI Everywhere

Kevin Wollenweber
SVP/GM, Data Centre and Internet Infrastructure

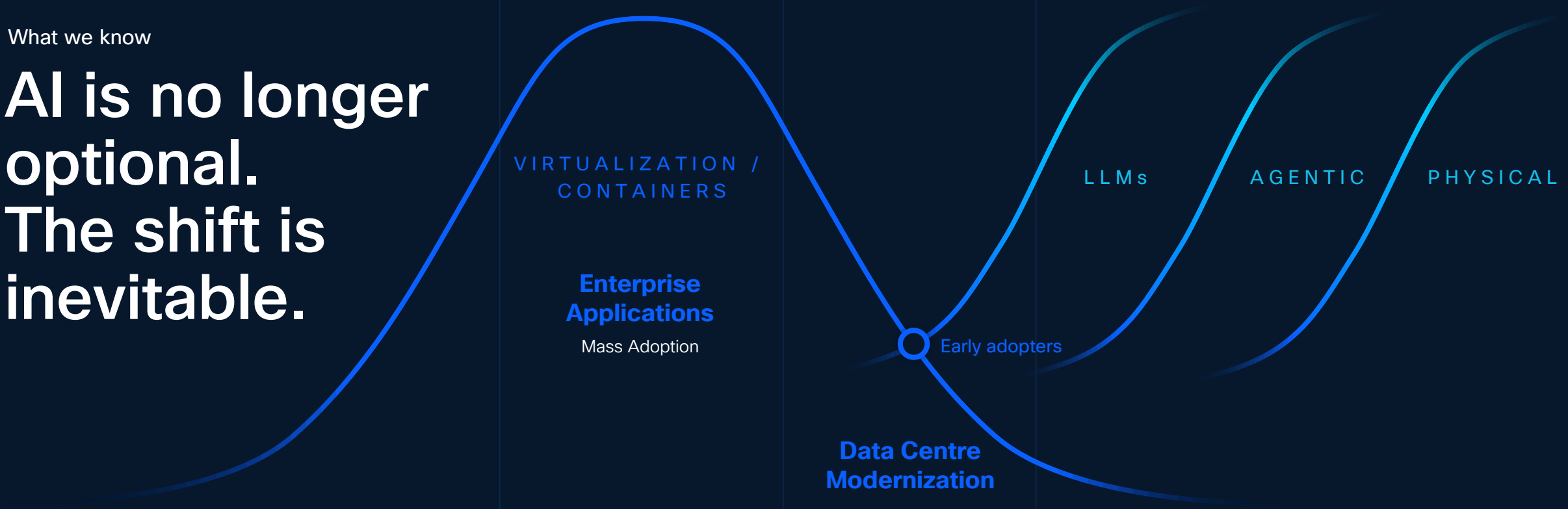
Danny McGinniss
VP, Product Management, Cisco Compute

Murali Gandluru
VP, Data Centre Networking

Session ID: KDDDCN-1000

What we know

AI is no longer optional. The shift is inevitable.



Data Centre

2010

AI-Ready Data Centre

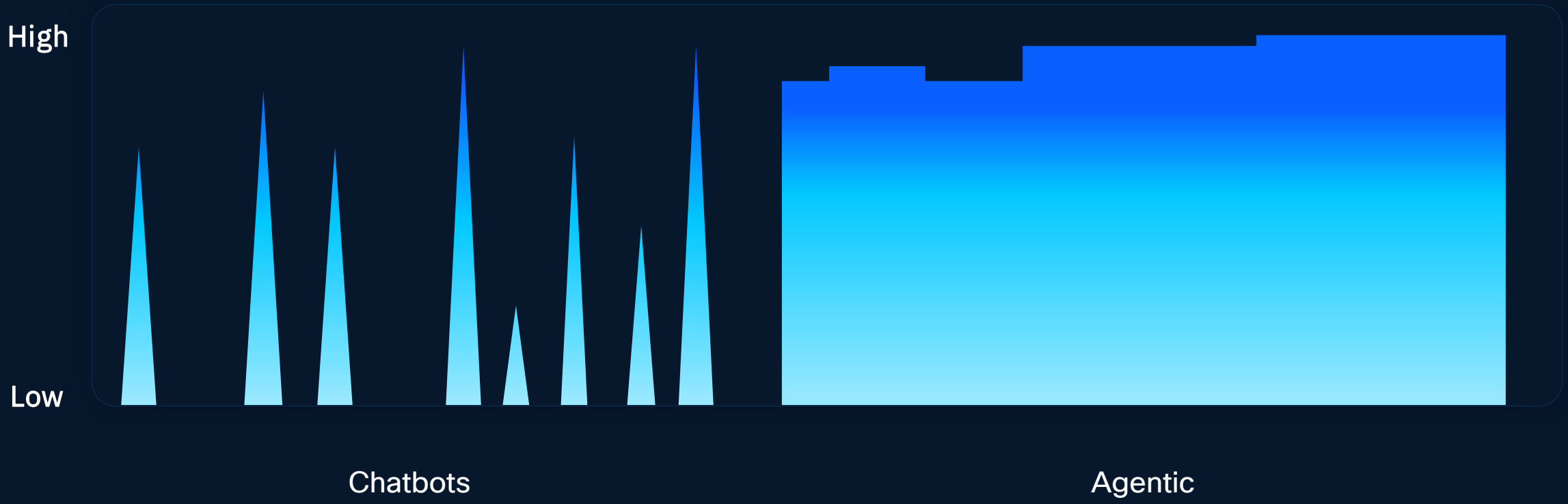
2025

AI Data Centre

2030

Inference Demand

POWER | COMPUTE | NETWORKING



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

Implications on Infrastructure

1

Surge of AI Workloads

***Evolution** of infrastructure*

2

A New Operational Reality

***Simplicity** at scale*

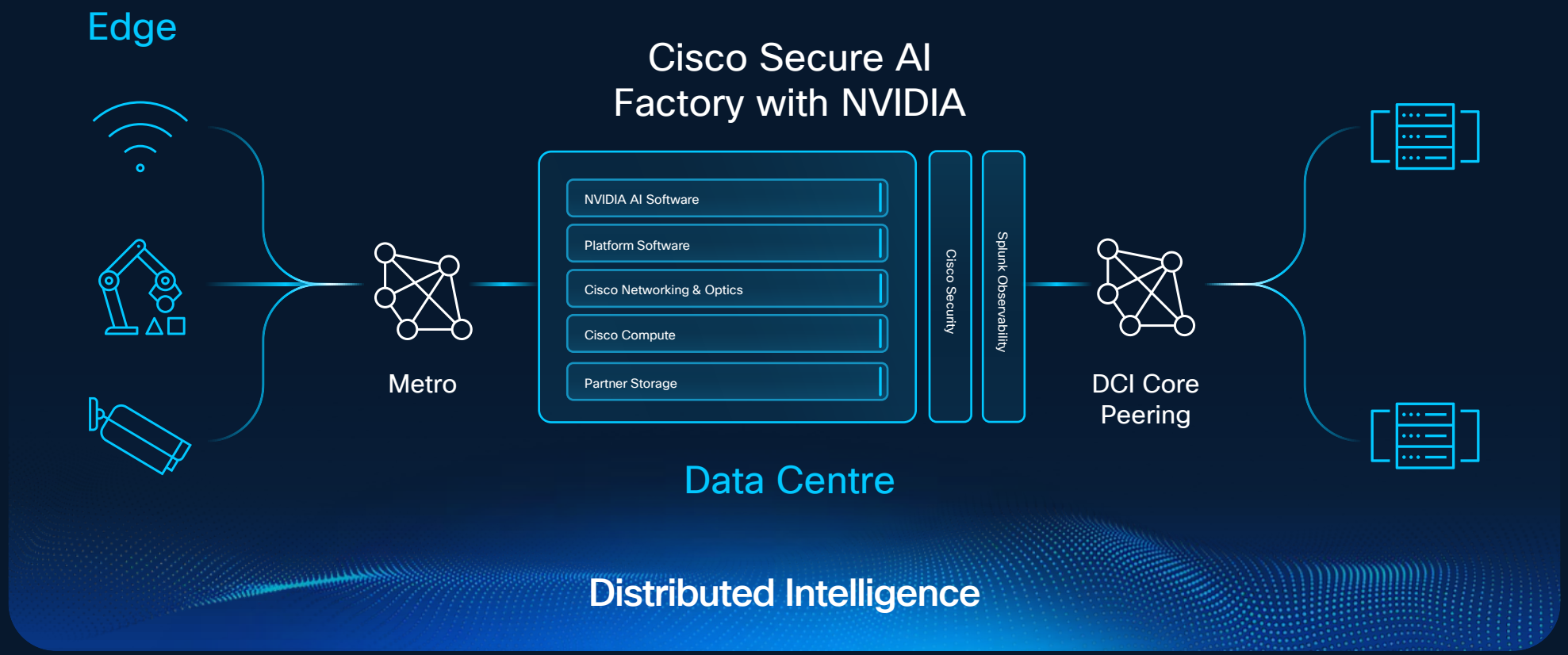
3

More Data at the Edge

***Intelligence** everywhere*

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

Cisco uniquely delivers AI-ready infrastructure from core to edge



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

Enabling the Platform Advantage

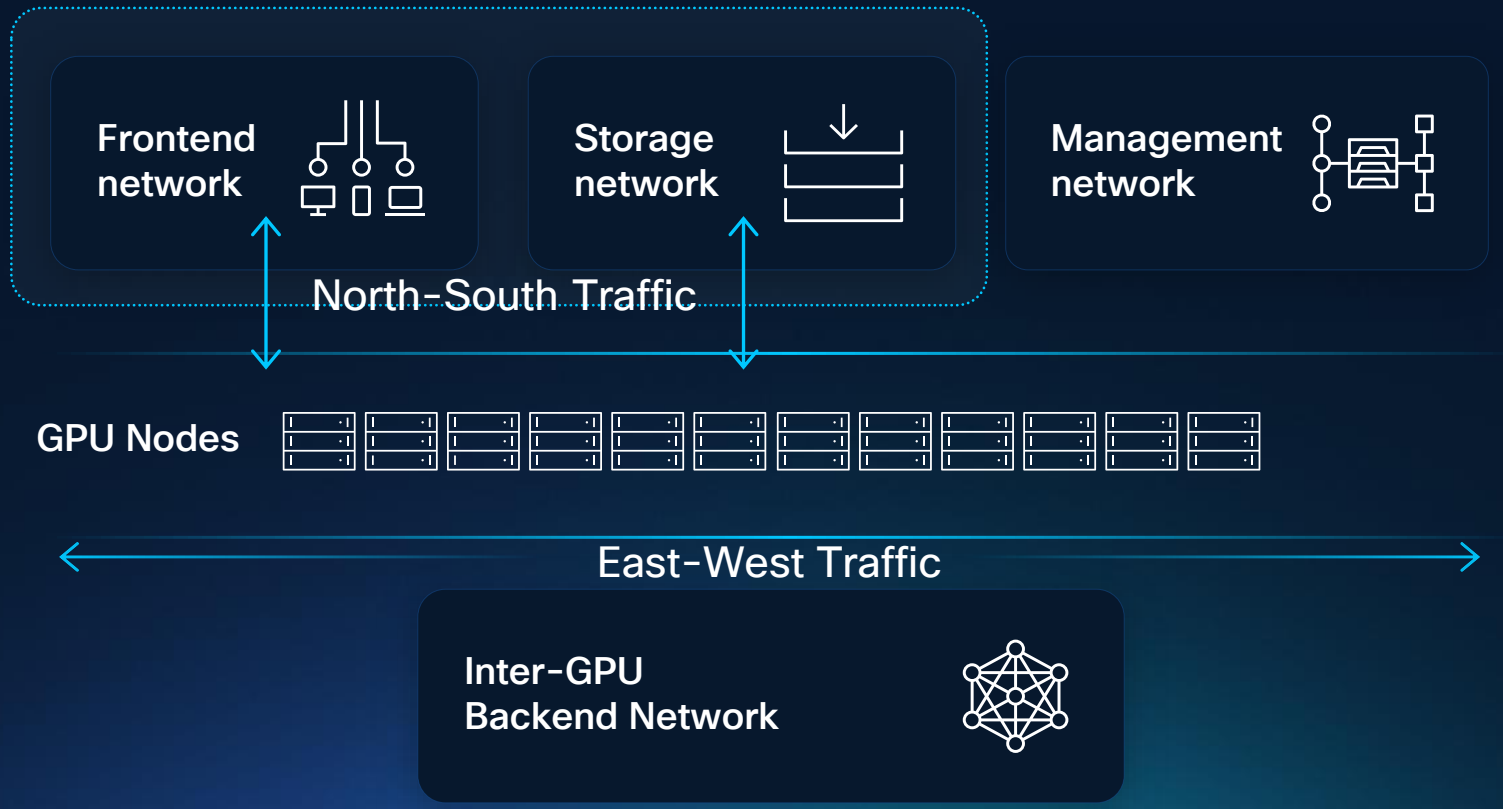
Across silicon, systems, software, and optics

Fabric for the AI era

Murali Gandluru

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

There are new demands on the network fabric



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

Cisco Nexus is purpose-built for the AI era

Frontend network

Storage network

Management network

GPU Nodes

Inter-GPU Backend Network

Customizable and turnkey ready solutions

High performance 400G / 800G Nexus 9000 platforms

Managed with Nexus Dashboard and Hyperfabric

Deployed in Enterprise AI, Sovereign AI and Neo Clouds

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

Nexus Dashboard

Physical or virtual cluster

Data plane | Policy and control plane | Management plane

SAN

NX-OS

ACI

Classic Ethernet | VXLAN | IP Media Fabrics | AI/ML Fabrics

Campus

WAN Transport

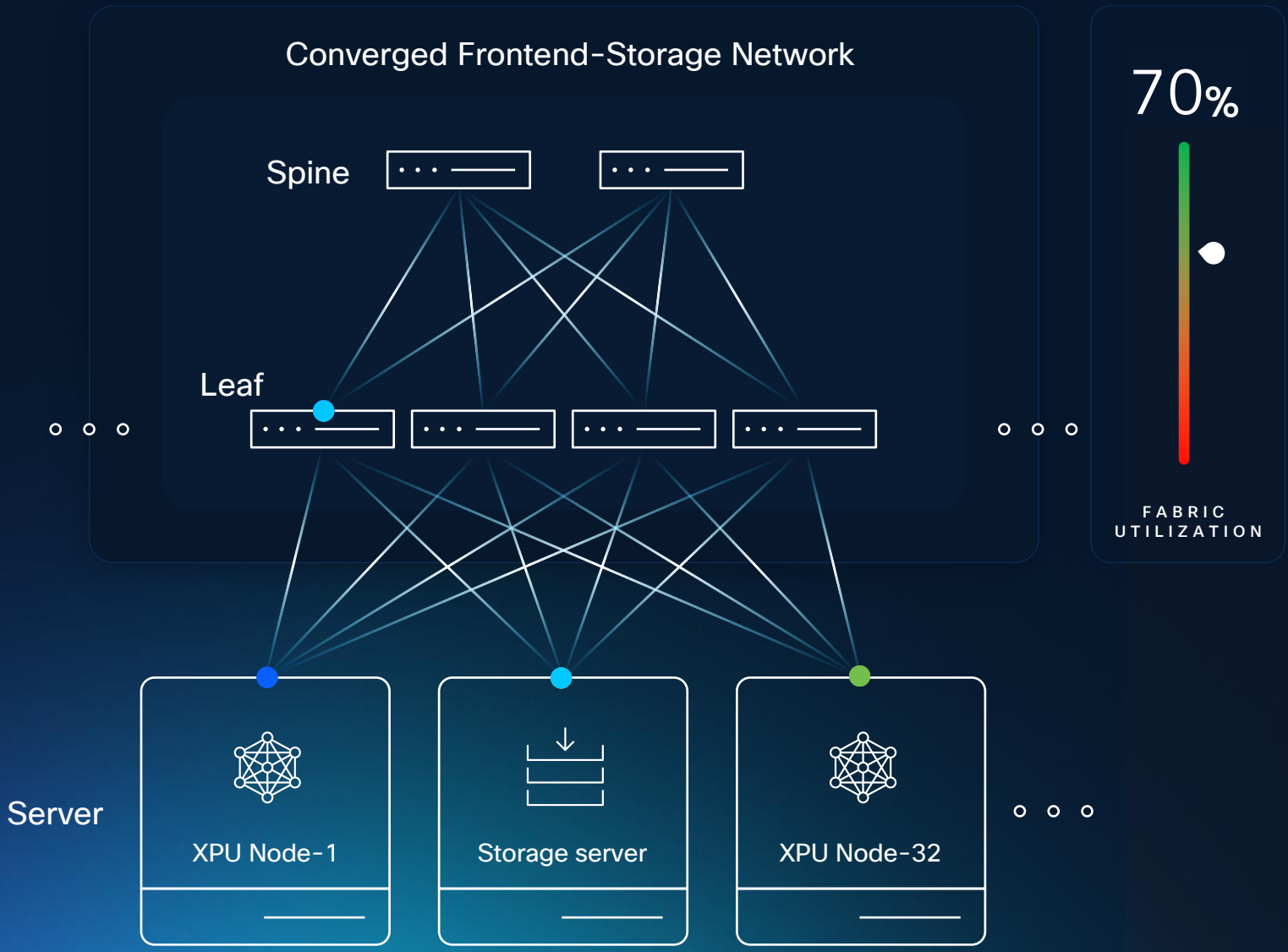
Virtual, K8
3rd party Networks

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

cisco

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

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

A new operating model

Now orderable for AI

Nexus Hyperfabric

Effortless, cloud-managed AI infrastructure with fast time to value

DEPLOY MANAGE SCALE

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

cisco

Fusing security into the network

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

Cisco Smart Switches integrated with Hypershield security

Cisco N9300 Series

Smart Switches

Shipping

N9324C-SE1U

24-port 100G

Shipping

N9348Y2C6D-SE1U

48-port 25G, 6-port 400G, 2-port 100G

Cisco Hypershield

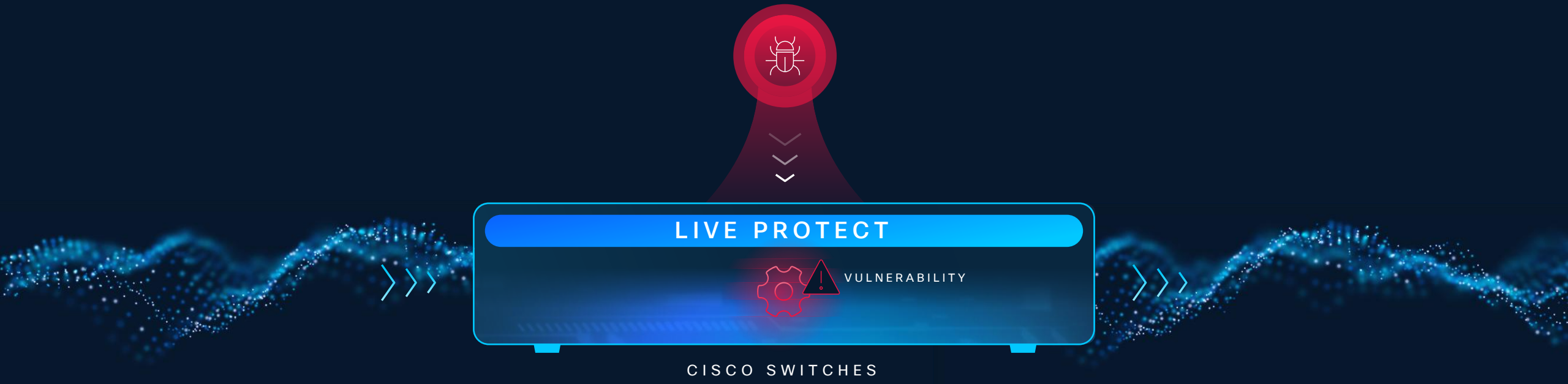
Use Cases

- Top of Rack segmentation and enforcement
- Cloud Edge
- Zone-based segmentation

Live Protect

Vulnerability shielding for Cisco AI networking devices

Stop the attack... but don't stop the network.



Apply policy to live system | No PSIRT patching or upgrade | No downtime or switch reboot | Continuous re-validation

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

CISCO
NVIDIA

Cisco Secure AI Factory with NVIDIA

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

Build Open, High-Performance AI networks

Based on NVIDIA Cloud Partner (NCP) design principles and Cisco's Cloud Reference Architecture (CRA)

Cisco N9300 Series Switches

NVIDIA BlueField

Powered by Cisco Silicon with NVIDIA Spectrum-X integration

CRA*

NVIDIA BlueField

Cisco 9100 Series Switches

NVIDIA BlueField

Powered by NVIDIA Spectrum-X Ethernet Switch Silicon

NCP*

NVIDIA BlueField

Built for Cloud Service Providers and High-Performance Computing Centres deploying **1K+ GPU clusters**

UNIFIED WITH NEXUS DASHBOARD

*CRA: Cisco Cloud Reference Architecture

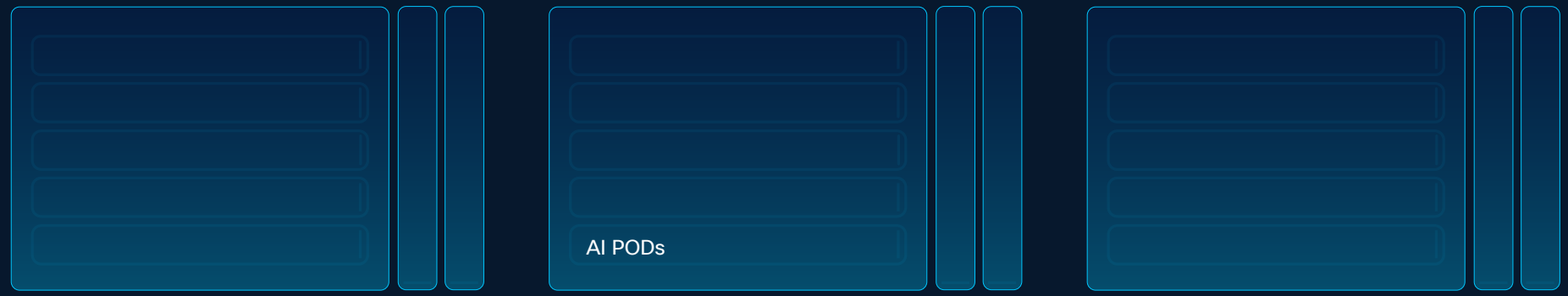
*NCP: NVIDIA Cloud Partner Reference Architecture Compliant

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

AI PODs for Cisco Secure AI Factory with NVIDIA use cases

Services

AI Defense | Splunk | NVIDIA AI Data Platform | ...



Inferencing → Optimization → Training

Workloads

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

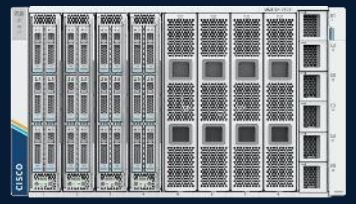
Cisco Confidential

Where data and AI meet

Danny McGinniss

Unified fabric across any workload, anywhere

Blade



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

PCIe Gen5 node
PCIe Gen5 switch module

Rack

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

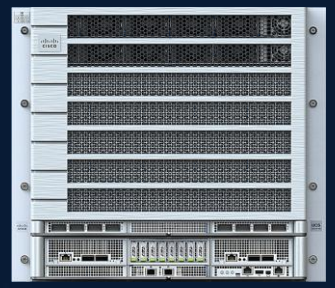
UCS C220 M7N
10 NVMe

UCS C245 M8SX
28 HDD/SDD

UCS C225 M8S
10 HDD/SSD

UCS C225 M8N
10 NVMe

AI Servers



UCS C885A M8
UCS C880A M8
8RU Dense GPU Server

...and **even more**
HGX follow ons

UCS C845A M8
4RU MGX Server

Edge

UCS XE9305 Chassis
UCS XE130c M8
Compute Nodes

The core and the edge are no longer separate worlds

To compete in the AI era, enterprises need an infrastructure strategy that **spans both**

Introducing

Cisco Unified Edge

A platform, not a server.

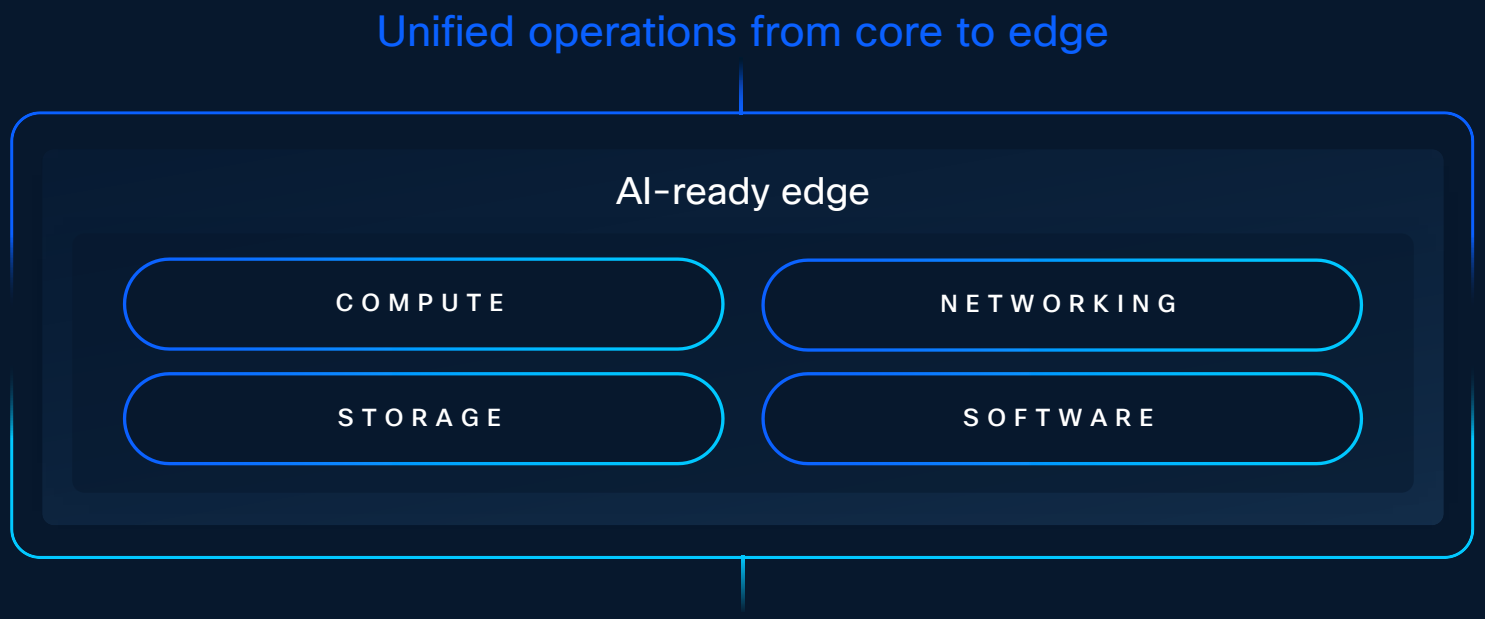
Unified operations from core to edge

Security fused into the network

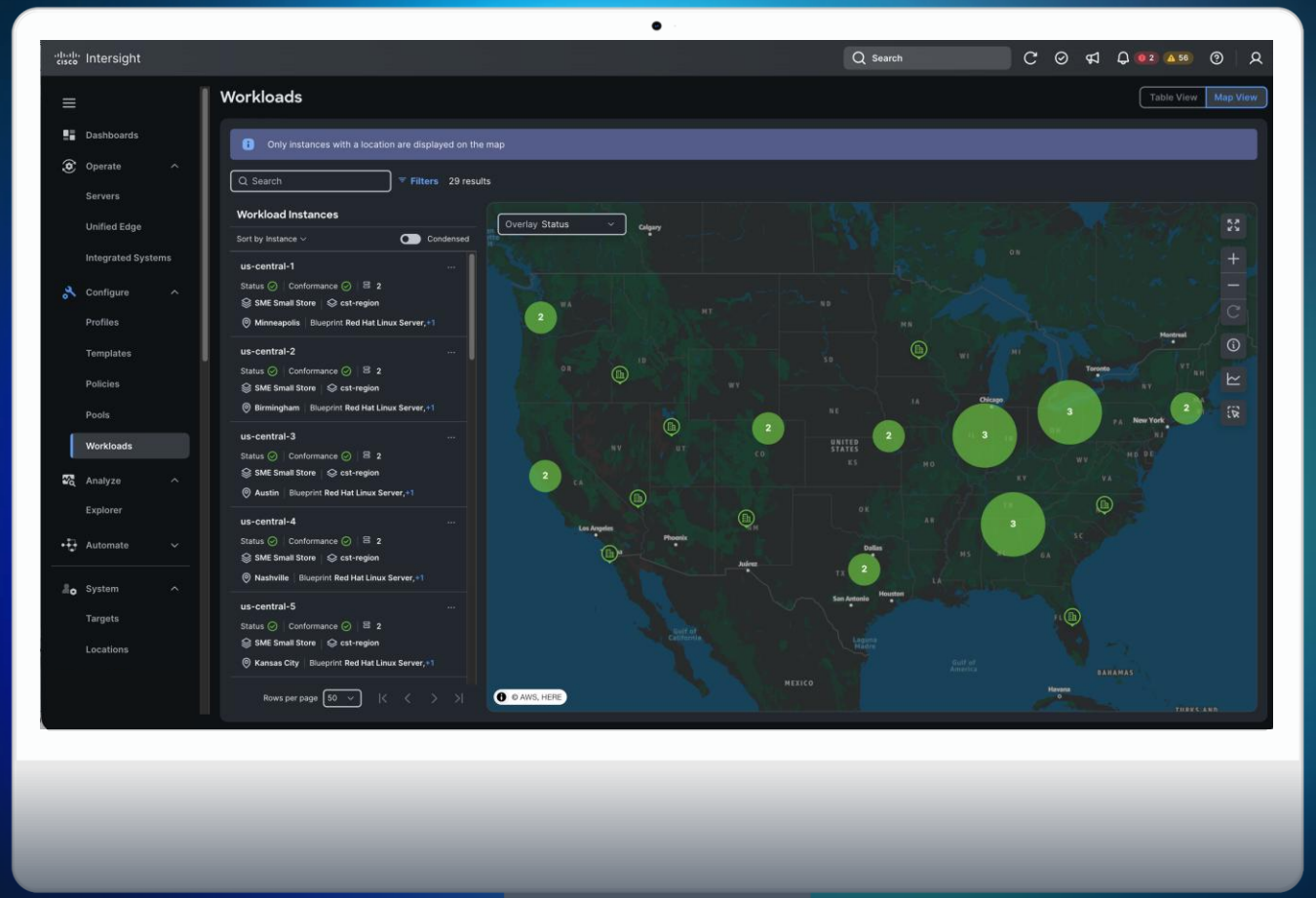
Introducing

Cisco Unified Edge

A platform, not a server.



Simplified operations that scales from core to edge



DAY 0

Zero-touch provisioning

DAY 1

Deployment at scale

DAY N

Fleet-wide operations

An ecosystem of services tailored to the edge

Sample Ecosystem Partnerships in Development

Security at every layer

Add-on

Security and Observability

Built-in

Trusted Infrastructure

Physical Security

Add-on

Trusted Access

Segmentation

Add-on

AI Model Security

Application Protection

How Cisco can help you simplify AI deployments from the edge up.

Deep networking expertise combined with compute innovation.

Comprehensive SaaS management for global fleet operations.

Built-in, multi-layered security from device to data centre.

Proven track record in enterprise infrastructure.

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

Cisco Confidential

Thank you