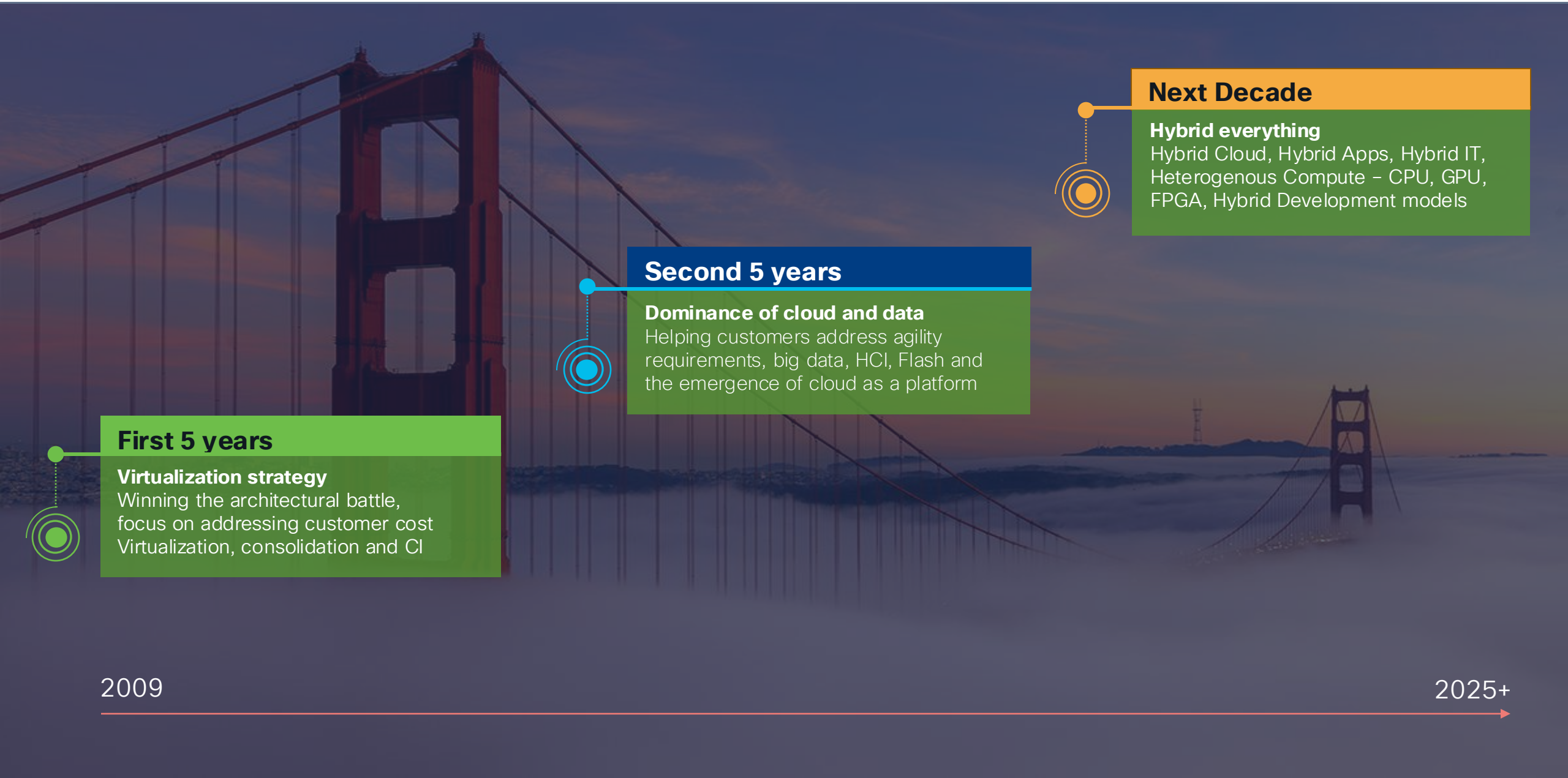
Cisco Compute: Built to Power What's Next

December 2025

Mahesh Natarajan
Sr. Director, Product Management, Cisco Compute


Next Decade
Hybrid everything
Hybrid Cloud, Hybrid Apps, Hybrid IT, Heterogenous Compute – CPU, GPU, FPGA, Hybrid Development models
Second 5 years
Dominance of cloud and data
Helping customers address agility requirements, big data, HCI, Flash and the emergence of cloud as a platform
First 5 years
Virtualization strategy
Winning the architectural battle, focus on addressing customer cost Virtualization, consolidation and CI
2009
2025+

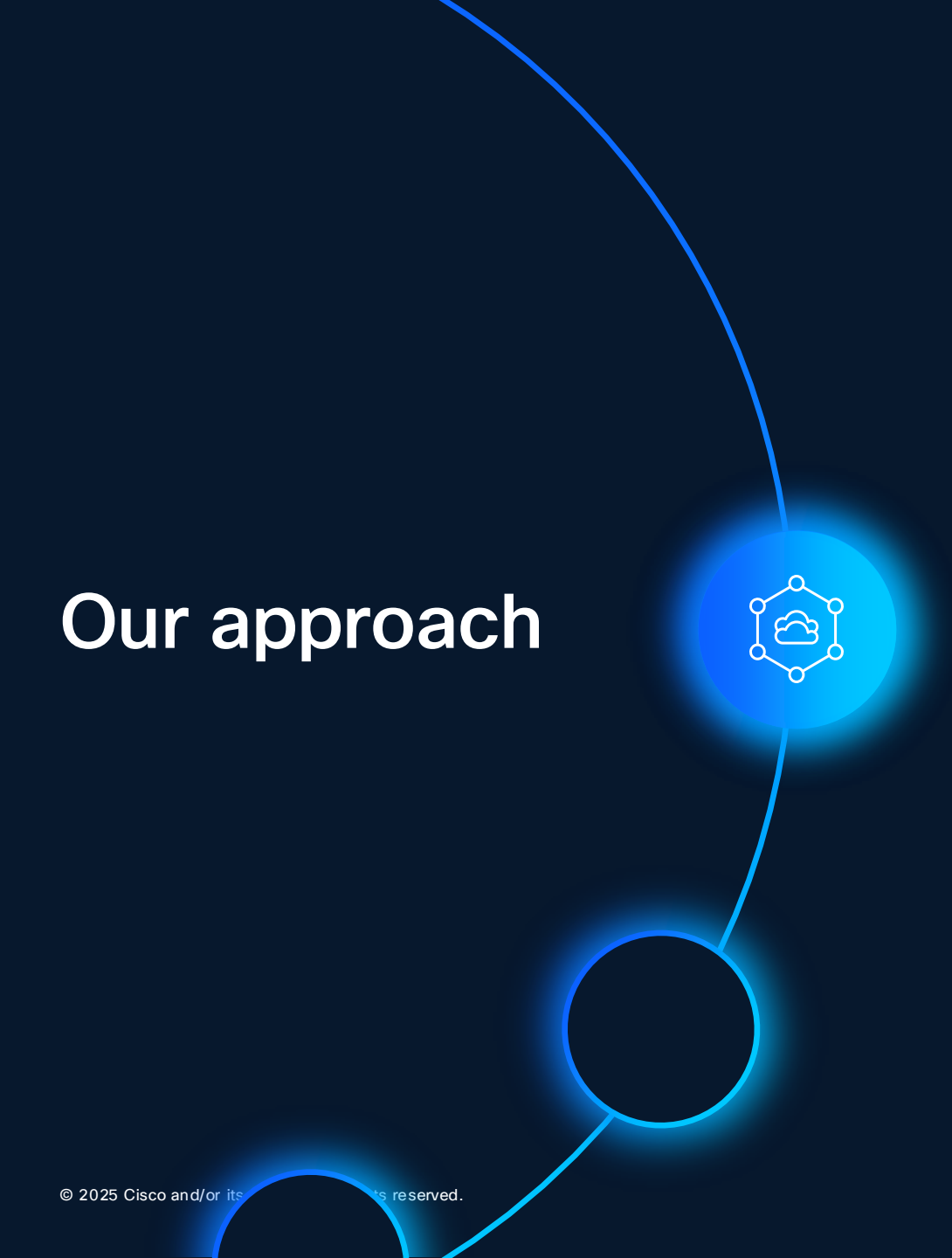
Our approach

Future-ready infrastructure

A unified operating model

Integrated solutions

Our approach



Future-ready infrastructure

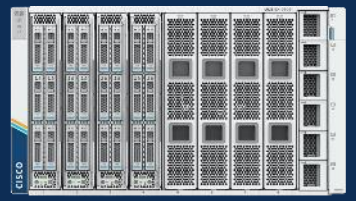
All-in on innovation, anchored by a strong portfolio

Cisco UCS Compute Portfolio

MAINSTREAM ENTERPRISE SERVERS

UCS X-Series X9508 Chassis

IFM Module



UCS X-Series Direct

NEW

UCS X210c M7

UCS X210c M8

NEW

UCS X410c M7

UCS B200 M6

UCS X215c M8

NEW

UCS C240 M8E3S
36 EDSFF E3.S 1T

NEW

UCS C240 M8SX
28 HDD/SSD/NVMe

NEW

UCS C240 M8L
16 LFF + 4 SFF

NEW

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.S 1T

NEW

UCS C220 M8S
10 HDD/SSD/NVMe

NEW

UCS C220 M7N
10 NVMe

UCS C245 M8SX
28 HDD/SSD

NEW

UCS C225 M8S
10 HDD/SSD

NEW

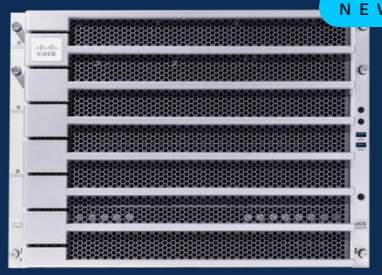
UCS C225 M8N
10 NVMe

NEW

AI SERVERS

UCS C885A M8
8RU Dense GPU Server

NEW



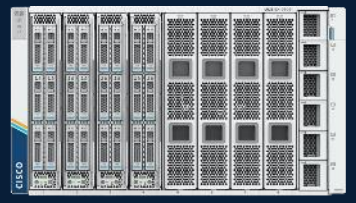
UCS C845A M8
4RU MGX Server

NEW

Infrastructure

Cisco UCS Compute Portfolio

Blade



UCS X-Series
X9508 Chassis

IFM Module

New

UCS X-Series Direct

UCS X210c M7

New

UCS X210c M8

UCS X410c M7

UCS B200 M6

New

UCS X215c M8

Coming

GoldenEye
PCIe Gen5 node
PCIe Gen5 switch module

Rack

New

UCS C240 M8E3S
36 EDSFF E3.S1T

New

UCS C240 M8SX
28 HDD/SDD/NVMe

New

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

New

UCS C220 M8E3S
16 EDSFF E3.S1T

New

UCS C220 M8S
10 HDD/SSD/NVMe

UCS C220 M7N
10 NVMe

New

UCS C245 M8SX
28 HDD/SDD

New

UCS C225 M8S
10 HDD/SSD

New

UCS C225 M8N
10 NVMe

AI Servers

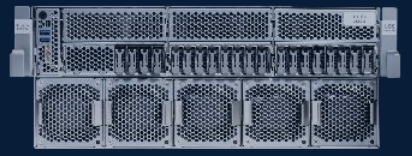
New

UCS C885A M8
8RU Dense GPU Server

...and even more HGX follow ons

Coming

New



UCS C845A M8
4RU MGX Server

Edge

Avatar
UCS XE9305 Chassis
UCS XE130c M8
Compute Nodes

Introducing Cisco UCS E-Box for VAST

Optimized for Enterprises and Neo Clouds

VAST

UCS C225 M8

Elemental building block of a VAST cluster

1 RU server

8x 15TB, 122TB raw | 100TB usable capacity

New! 30TB and 60TB also supported with 1.9TB SCM

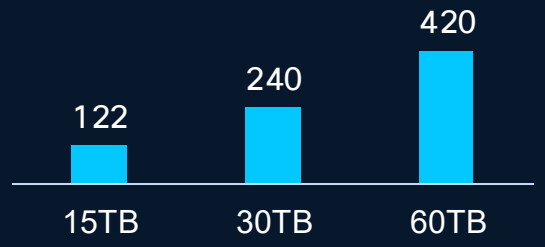
960GB SCM

384GB DDR5 memory

AMD 9454P 2.7GHz 290W 48 cores

2x CX-7 | 2x B3220L, 1x X710 (management)

Raw Capacity per Node



Evolution of data center infrastructure

Integrated architecture

cisco

Unified Architecture for

Modern Workloads

NAS Consolidation

Backup Target

Training

Inference, RAG

Splunk SmartStore

Data Lake, Hadoop FS

Media & Entertainment

Single point purchasing

Single support model

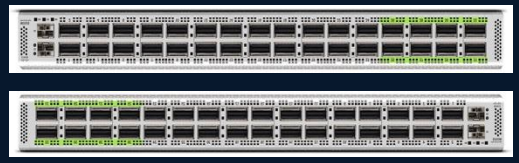
cisco

Data Protection Is Your Competitive Superpower

PBBA “Islands”

23% Less Power
4x Less Floor Space
4x Less Rack (U)
1.4x Backup Performance
32x Recovery Performance

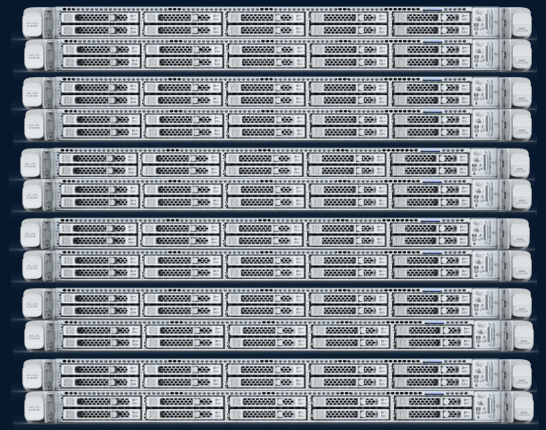
Global Deduplication Pool
No Management Overhead



VAST

DataStore

Scalable Multi-Protocol backup Target



cisco

UCS –Rackmount Servers Differentiation

It’s not a server, it’s a system

Faster Apps, Better User Experience

Less Cabling, Simpler Physical Management

Consistent, Predictable Deployment

Use Your Automation Tools Your Way

Respond And Remediate Instantly Pass Audits

Visibility And Management Of Everything, Everywhere

#1 in Integrated & Certified Validated Designs

Flashstack (Pure Storage)

Flexpod (NetApp)

Cisco + Nutanix

AI Optimized Workloads (RedHat, NVIDIA)

UCS by the Numbers

+55K Customers

+250 CVDs

+200 World Record performance benchmarks

+1M Intersight Connected Devices

10+ years top tier reliability

High-density GPU servers

For data-intensive use cases like model training and deep learning

UCS Accelerated | Cisco UCS C885A

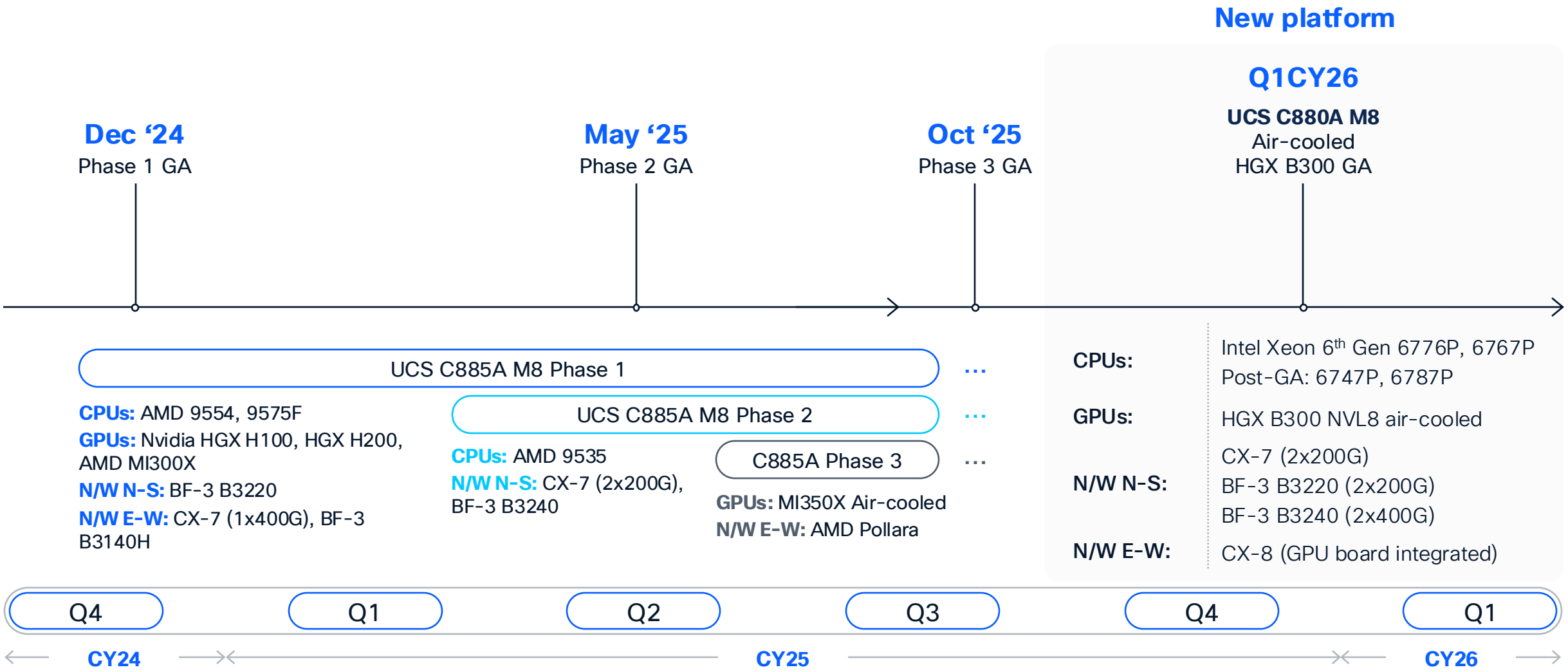
NVIDIA HGX™ reference design

Supporting 8 NVIDIA HGX™ H100 or H200 GPUs and NVIDIA AI Enterprise software

And 2 AMD 4th Gen/5th Gen EPYC Processors

Available as an option in Hyperfabric AI

Cisco dense GPU server roadmap



 Note: Roadmaps are tentative and can change.

UCS C880A M8

Massive performance for AI at scale

Designed for compute-intensive workloads, the server accelerates LLM training, deep learning, fine-tuning, and HPC, delivering unparalleled processing power and scalability

UCS accelerated

Dense-GPU server supporting:
8 NVIDIA HGX B300 NVL8 GPUs

2 6th Gen Intel Xeon (Granite Rapids) Processors

Key target verticals

- Financial services
- Manufacturing
- Service providers (neo-cloud)
- Healthcare and life sciences
- Automotive

UCS C880A use cases

Artificial intelligence

- GenAI–LLM training and fine-tuning, e.g., text-to-image, AI-powered content creation
- Deep learning model training, e.g., NLP, computer vision, speech recognition
- Reinforcement learning for robotics and automation
- AI inference and deployment for real-time applications

HPC and scientific research

- Large-scale simulations—physics, climate modeling, astrophysics
- Molecular dynamics and genomics research
- AI-driven drug discovery and medical training

Data science and big data analytics

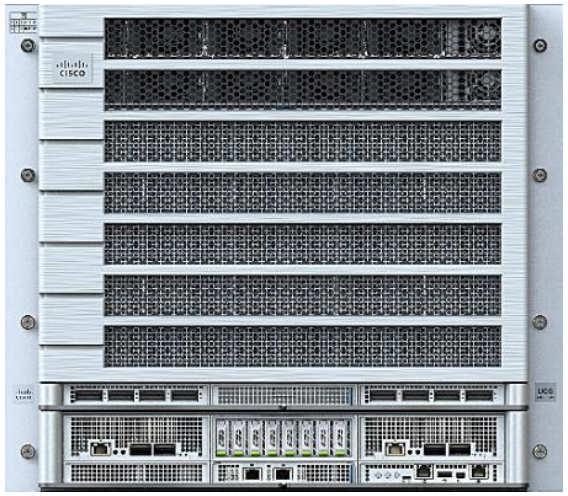
- AI-driven predictive analytics for business intelligence
- Real-time data processing: GPU-accelerated big data processing (RAPIDS, Apache Spark)
- Accelerated Extract, Transform, Load (ETL) processing

Unparalleled performance

High-speed GPU interconnect

Scalability

UCS C880A Dense GPU Server Specifications



Product specifications

Form factor	• HGX 10RU 19” Rack Server
Compute + memory	• 2x 6th Gen Intel Xeon CPUs (Select SKUs for AI and HPC workloads) • Up to 32x DDR5 RDIMMS
Storage	• 2x M.2 SATA Boot Drives with HW RAID Controller (Boot) • Up to 8x PCIe Gen5 x4 E1.S NVMe SSDs (Data)
GPUs	• 8x NVIDIA HGX B300 NVL8 air-cooled GPUs
Network cards	• **E-W:** Integrated ConnectX-8 • **N-S:** 4x PCIe Gen5 x16 FHHL slots, 1x OCP TSFF Gen5 x8
Cooling	• 20 Hot swappable FANs
Physical I/O	• 1 USB 3.0 A, 1 mDP, 1 ID Button, 1 System Power Button, 1 USB 2.0 A UART (for debugging), 1 RJ45 (OOB mgmt.), 1 System ID LED/button
Power supply	• 12x 50V 3.2kW (N+N redundancy)

Shipping Now

Flexible, modular AI servers

"Start small and scale up" with AI

UCS Accelerated | Cisco UCS C845A

NVIDIA MGX™ reference design

With NVIDIA H100, H200, L40S, AMD MI210 GPUs

Included as an option in Cisco Hyperfabric

High performance in a compact form factor

Enhanced power delivery, fewer PCBs, and better cable routing for optimal airflow and thermal management

Orderable this month
with NVIDIA RTX PRO 6000 Blackwell GPUs

cisco

2nd Gen X580p PCIe Node and X9516 X-Fabric

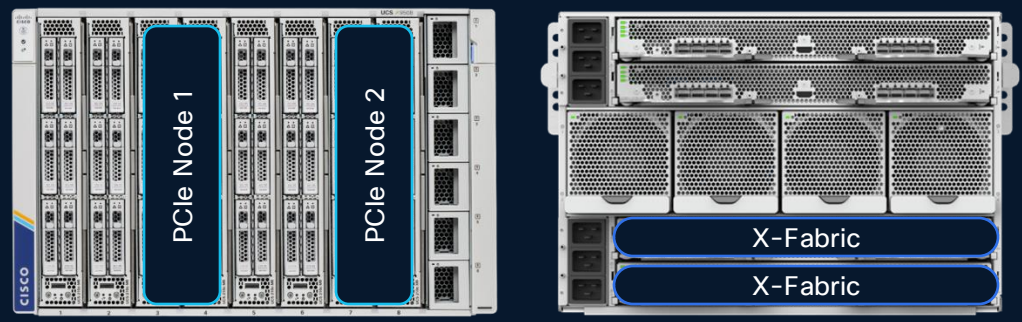
Cloud-operated, composable infrastructure for AI and traditional workloads

Solution for customer
who needs higher GPU density

Supports wide range of workloads

Intersight managed solution

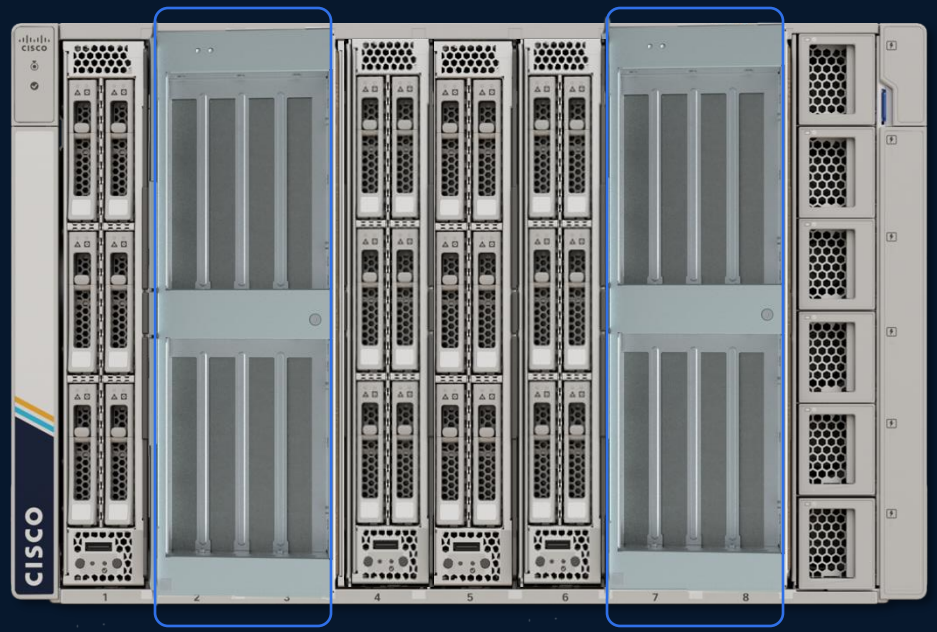
Competitive differentiation
with X-Fabric and X-Series



UCS X-Fabric Technology with PCIe Node

- PCIe Switching with PCIe Gen 5 connectivity
- 4x FHFL or HHHL GPUs per PCIe node
- Intra-host GPU interconnect with NVLink
- Intersight policy-based Management
- Inter-host scaling with RDMA over AI Fabric

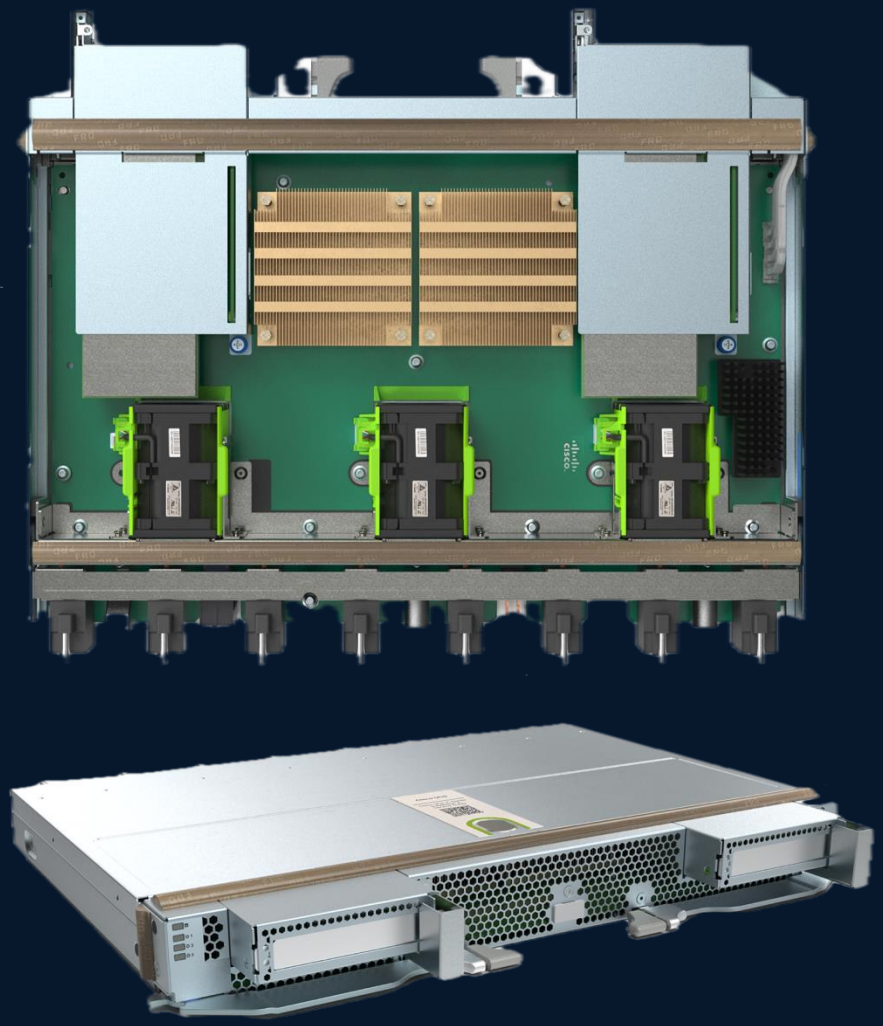
UCS X580p PCIe Node



- Double wide PCIe node for 4x FHFL GPU and PCIe G5 GPU support
  - Nvidia H200-NVL, RTX PRO 6000 & L40S
- Support multiple vendors: Nvidia, AMD*/Intel*
- NVLink bridge support
- Support up to 600W FHFL GPU
- Managed PCIe node with BMC support
- Policy based GPU management
- Ability to share GPUs across two Compute nodes

* AMD & Intel GPUs support will be post FCS

UCS X9516 X-Fabric



- PCIe Gen5 Switching
- 2x CEM Slots to support HHHL NIC cards
  - ConnectX-7 (2x 200GB & 1x 400G)
- Managed XFM Modules with BMC support
- GPU Direct Support over RDMA
- GPU Backend (East-West Traffic) network support

Introducing Cisco Unified Edge

AI-ready edge

Compute | Storage | Networking | Software | SaaS Management | Analytics | Security

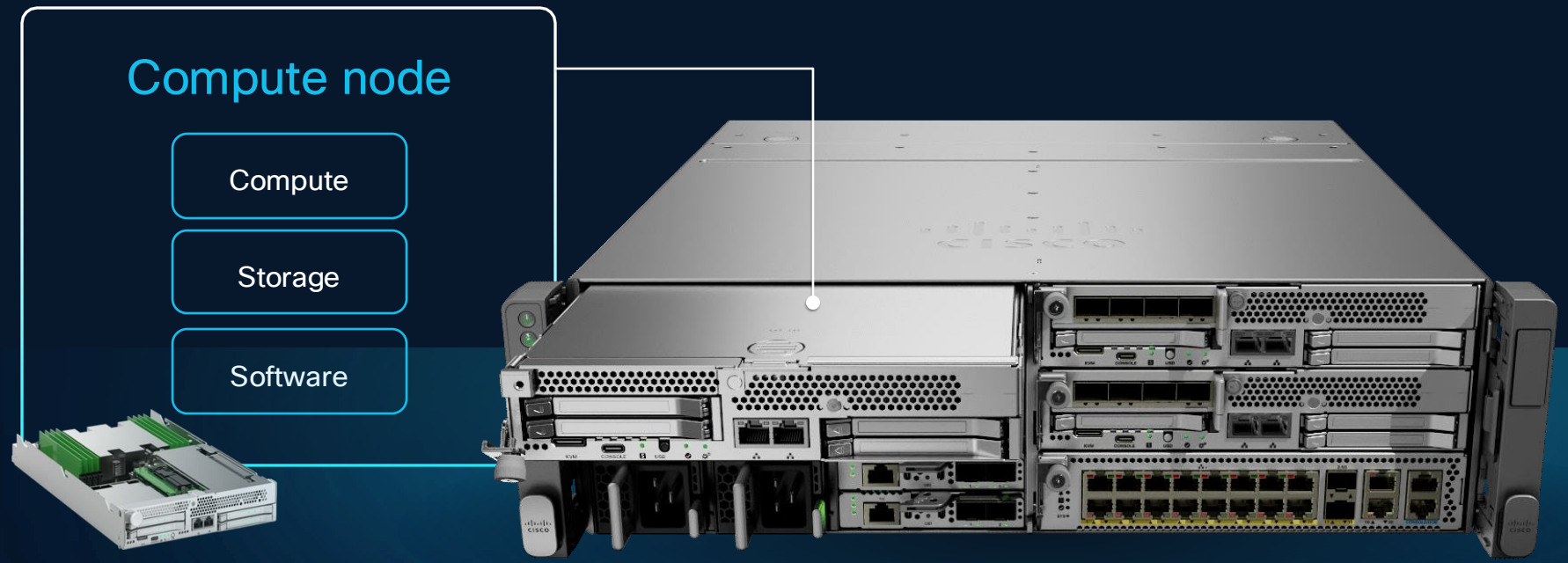
NUTANIX | Red Hat | vmware by Broadcom | intel. | cisco | RANCHER BY SUSE

cisco

Cisco Unified Edge: Future-Ready Performance

AI-ready edge

Integrates compute, networking, storage, and security



NUTANIX · Red Hat · vmware by Broadcom · intel. · cisco · RANCHER BY SUSE

cisco

Cisco Unified Edge

AI-ready edge

Fully validated, full-stack system that integrates advanced network, compute, storage and security

Compute node

- Compute
- Storage
- Software

Networking node

- Network
- Security

NUTANIX · Red Hat · vmware by Broadcom · intel · cisco · RANCHER BY SUSE

cisco

Our approach

A unified operating model

100% confidence—Intersight stands in a league of its own

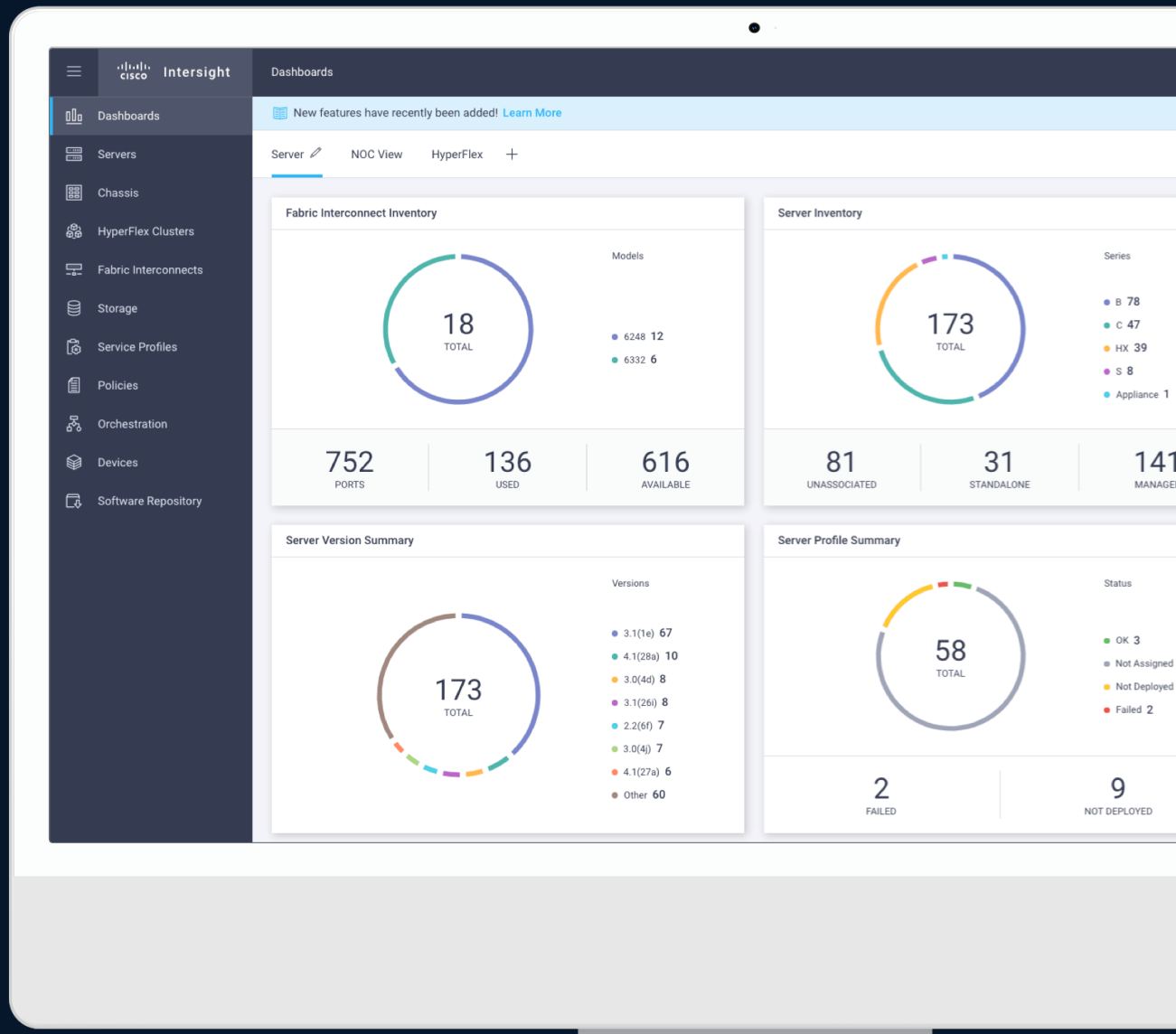
Cisco Intersight

Simpler, smarter, more agile computing

Flexible SaaS or on-prem management

Across Data Center, AI, and Edge

Every feature available via APIs



Unified operating model

Integrate, extend, and orchestrate

Intersight API

Seamless integration with popular tools like Ansible and Terraform

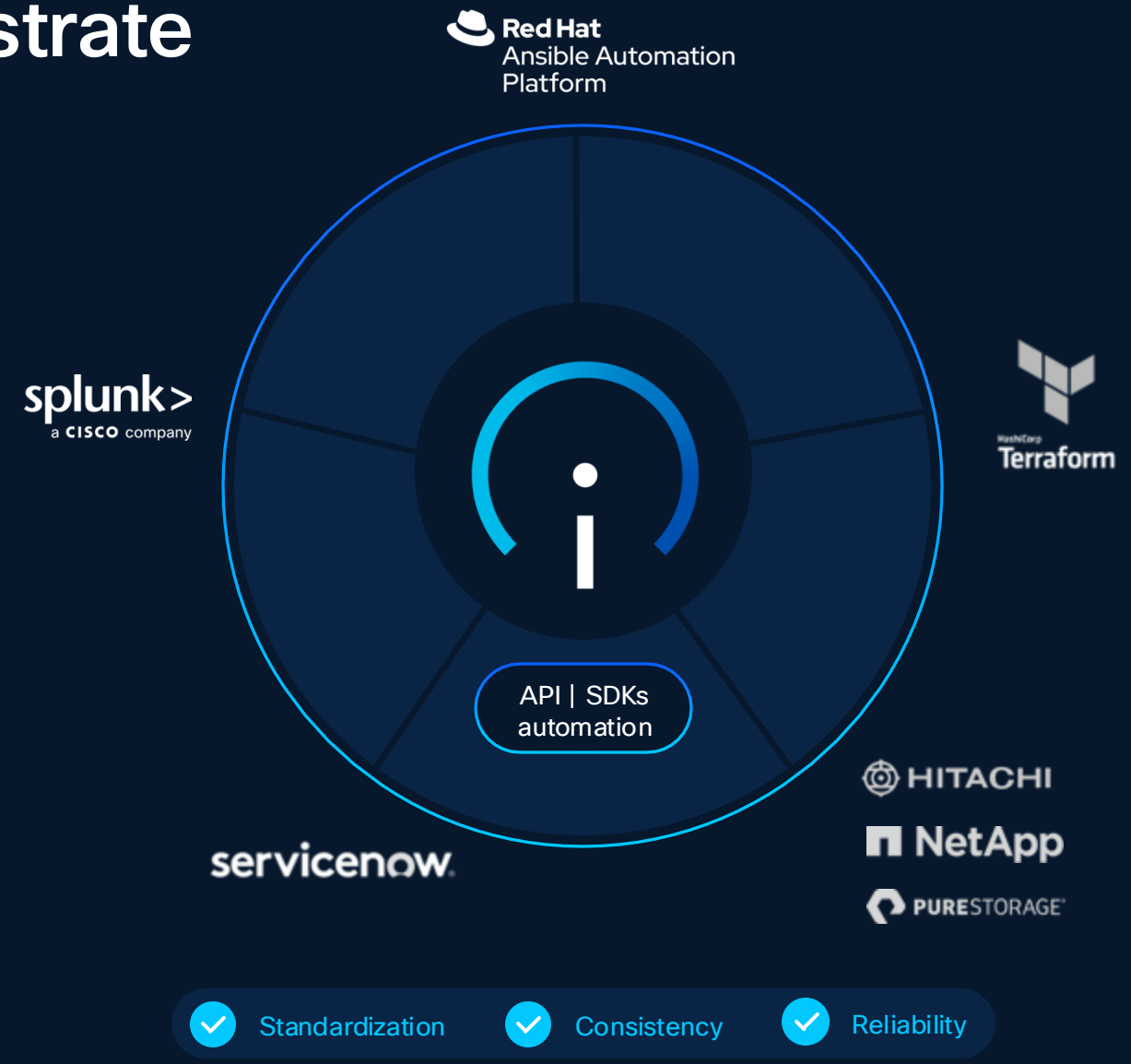
Standardize deployments and configuration management

Consistent and repeatable operations across environments

Third-party integration

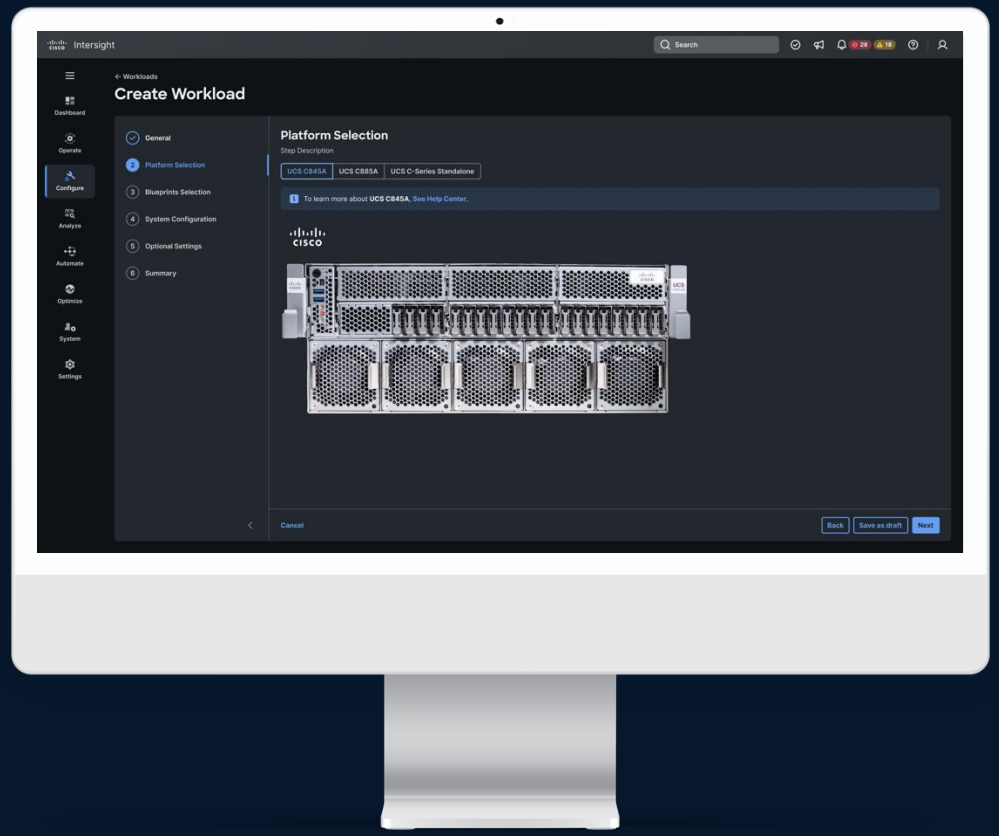
Storage plugins: automate management and orchestration

ServiceNow integration for automated incident management



Powered by AI, for AI

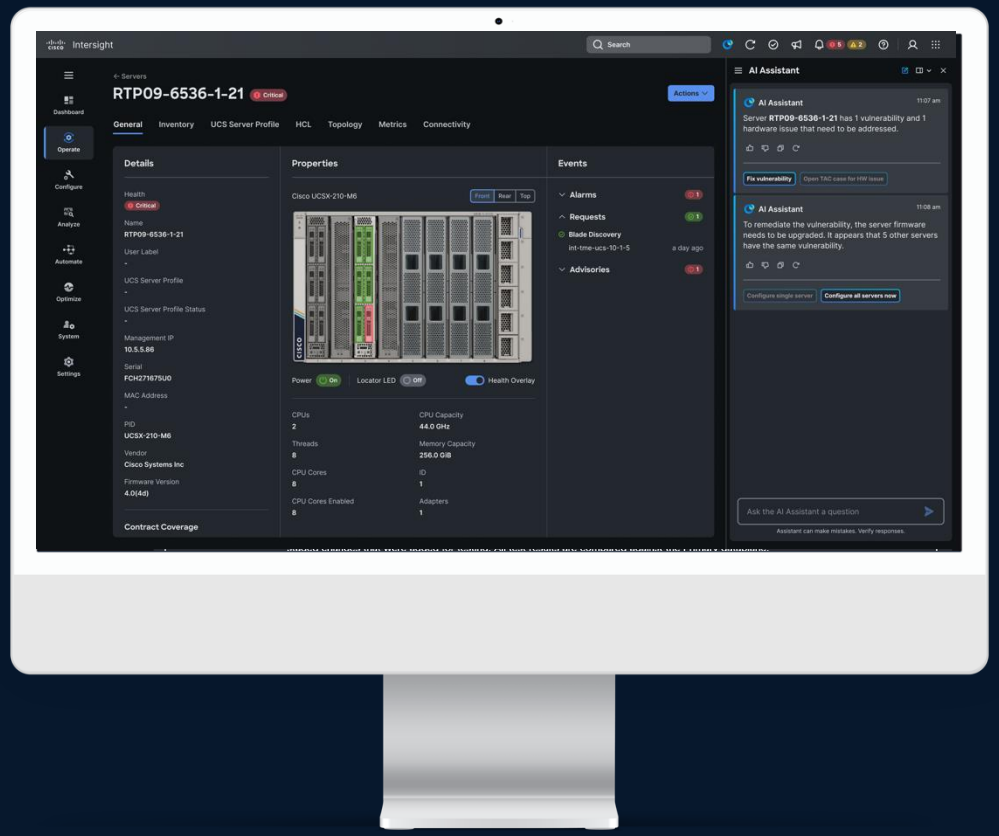
GPU Support



AI outside

Coming

AI Assistant



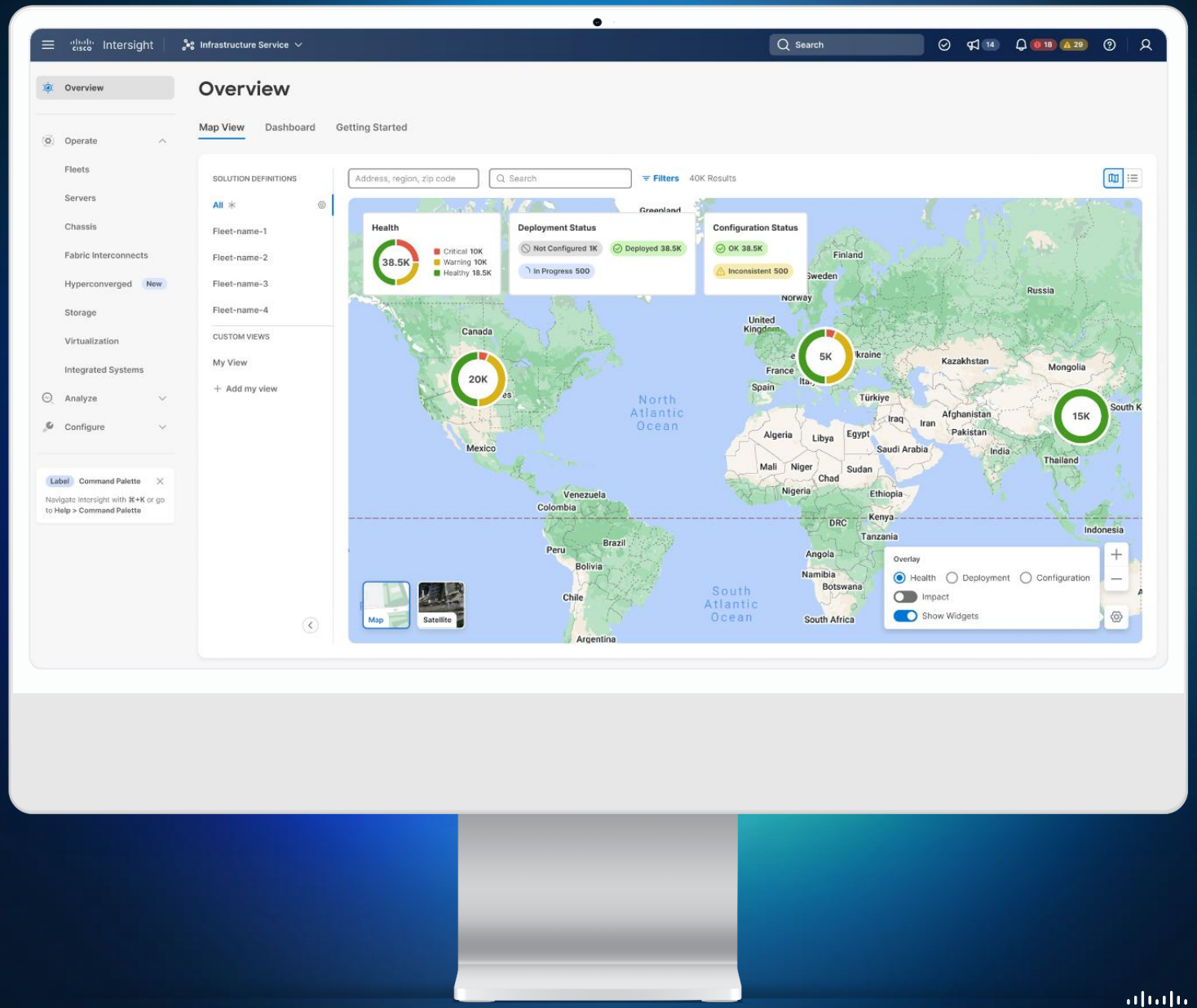
AI inside

Intersight Global Fleet Operations

Cisco Intersight

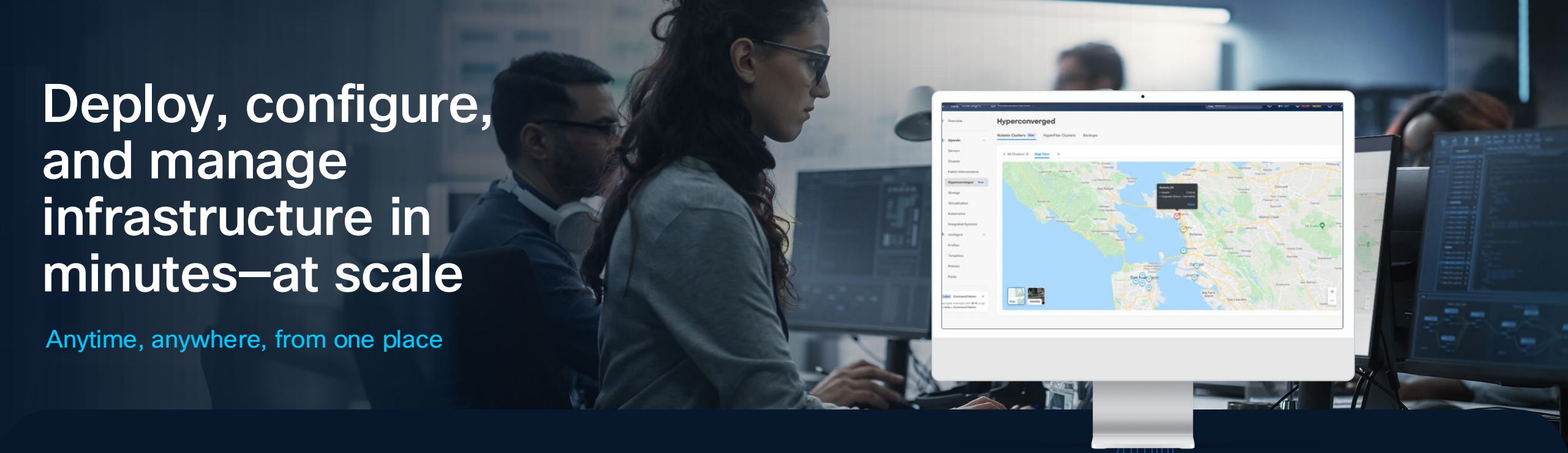
The only system with a true SaaS operating platform

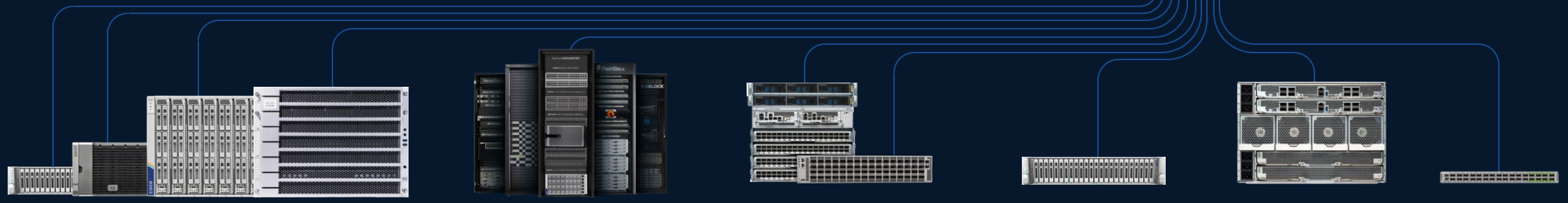
1. Intersight delivers end-to-end visibility, operations and support
2. Simplify to eliminate complexity with one tool to manage the entire environment
3. Remove the burden of patching, updating and securing system management tools
4. Continuous feature delivery - always have access to the latest features, functions and updates



Deploy, configure, and manage infrastructure in minutes—at scale

Anytime, anywhere, from one place





Compute | Storage | Networking | Compute | Compute | Networking

Data Centers | Edge

Our approach

Integrated solutions

Full-stack. Full-solution. Full-growth.

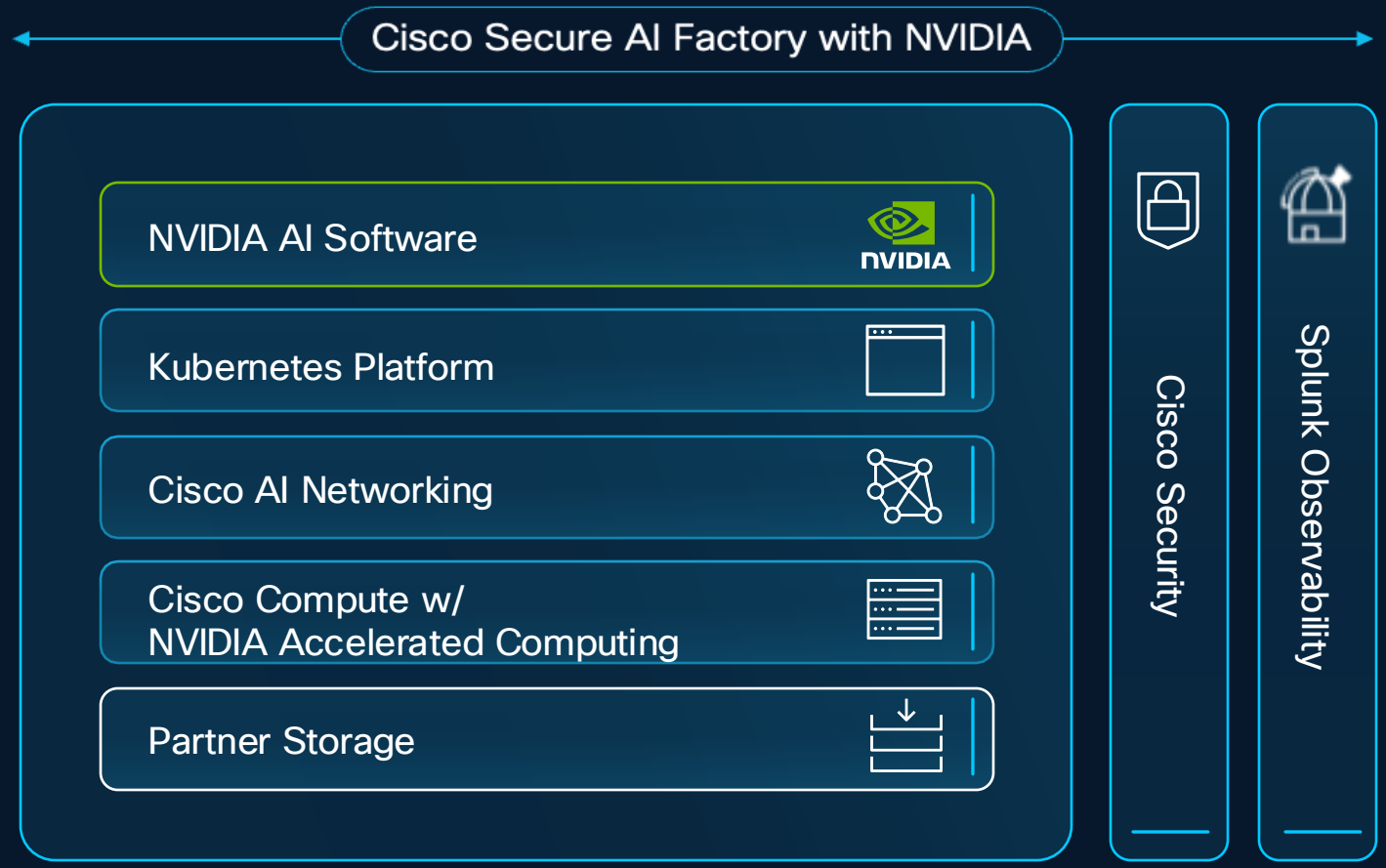
The power of partnering with industry-leading tools

cisco

Cisco Secure AI Factory with NVIDIA

What is it?

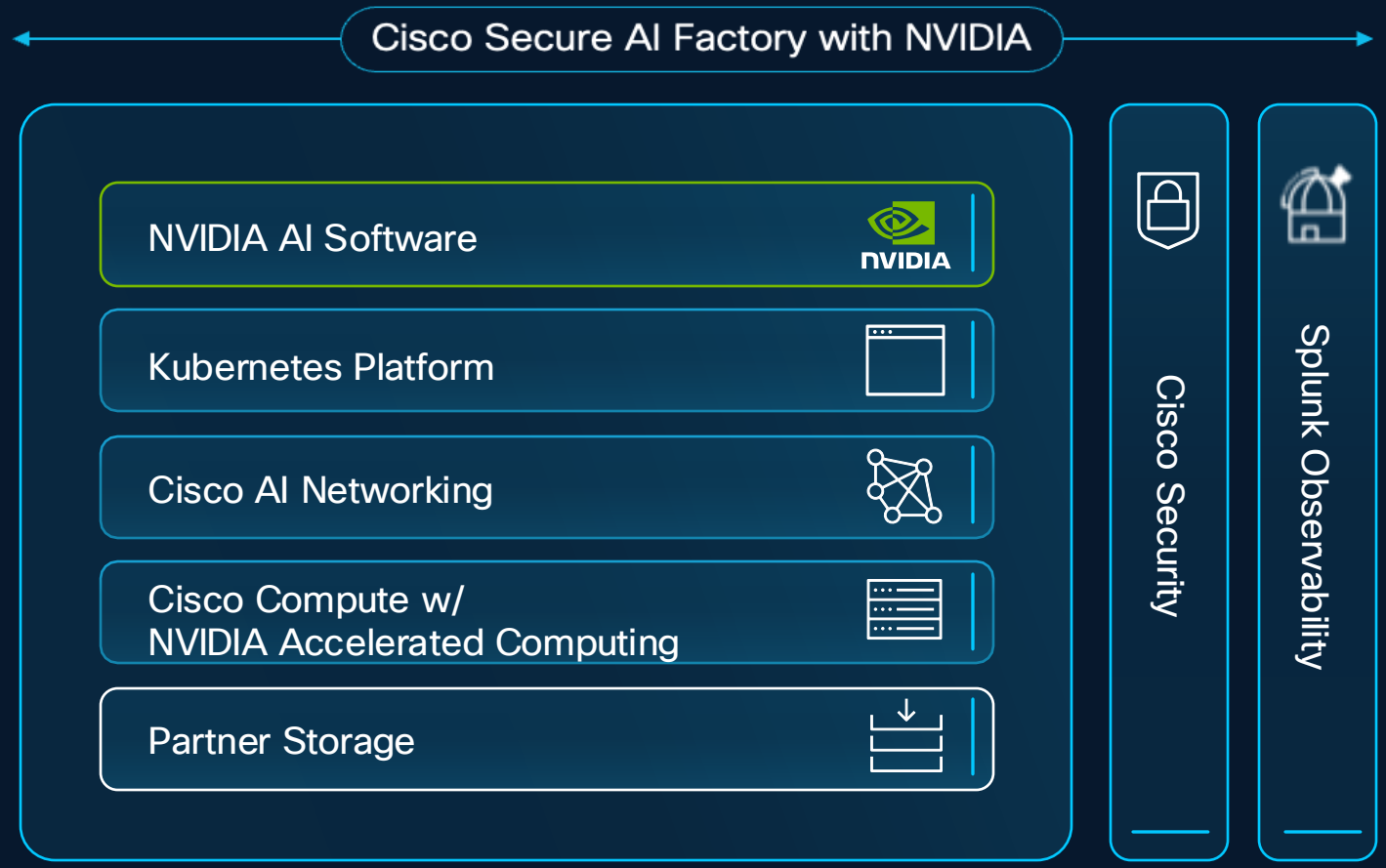
A modular reference design that combines high-performance infrastructure with full-stack security and observability



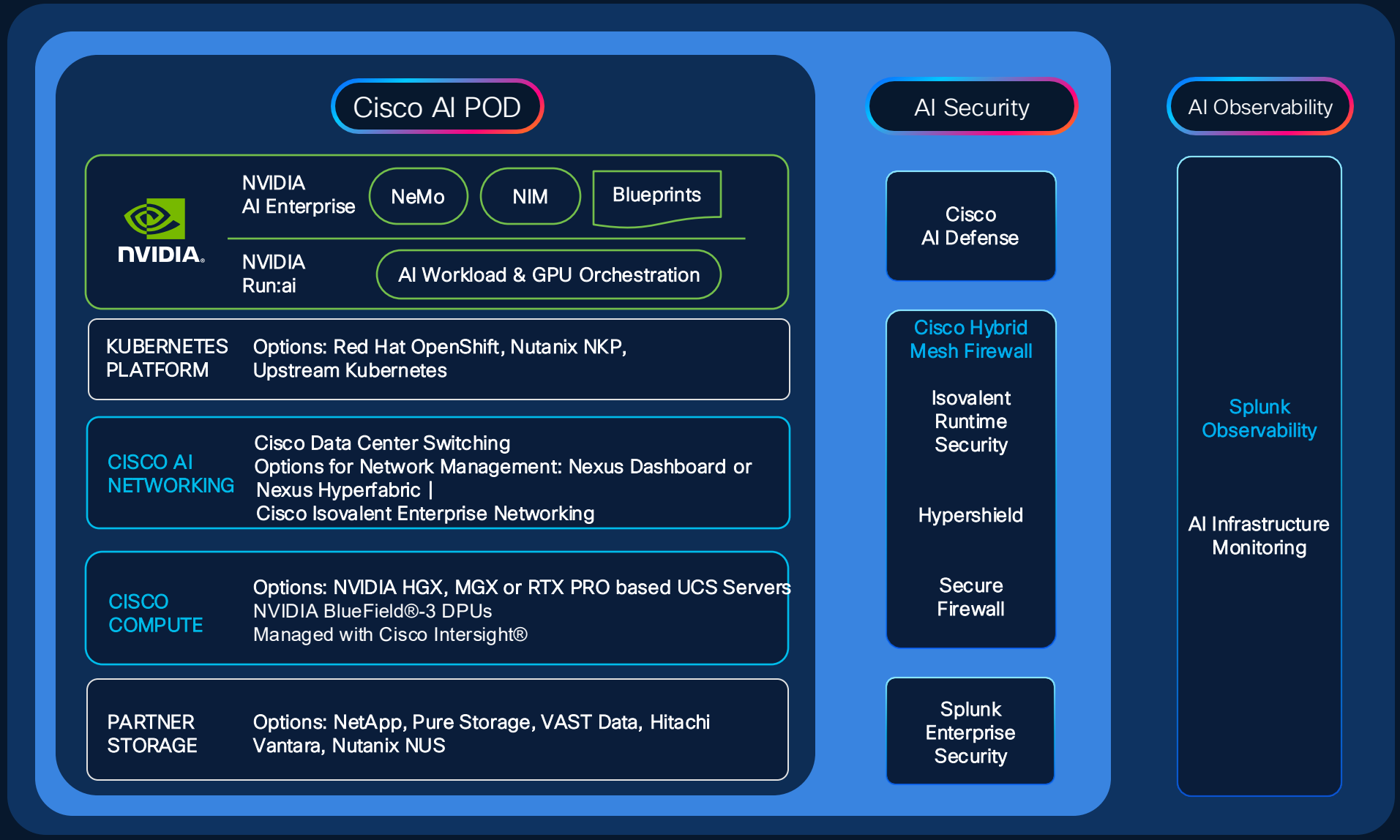
Cisco Secure AI Factory with NVIDIA

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

- Backed by an expanding set of Cisco Validated Designs that combine core AI Infrastructure, security and observability
- Curated Solution Catalog with T-shirt sizing for simplified ordering and predictable results
- Meets customers where they are with flexible options at key layers of stack (K8's, networking, compute, storage)
- Cisco supported

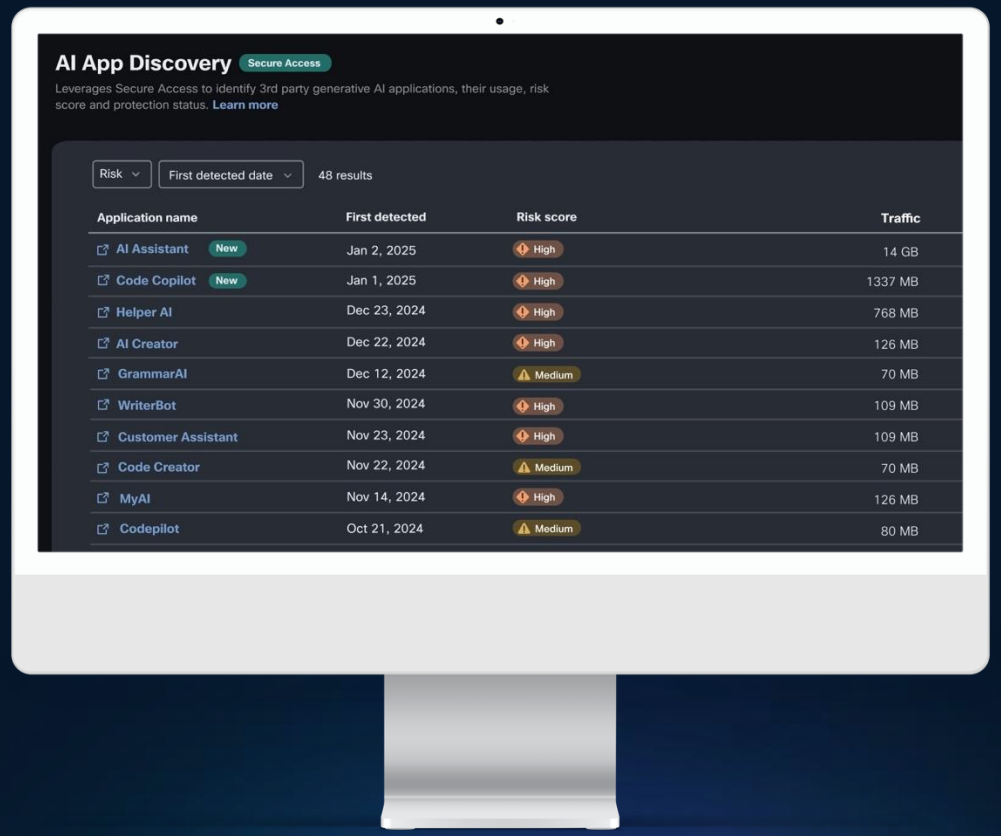


Product view: Cisco Secure AI Factory with NVIDIA

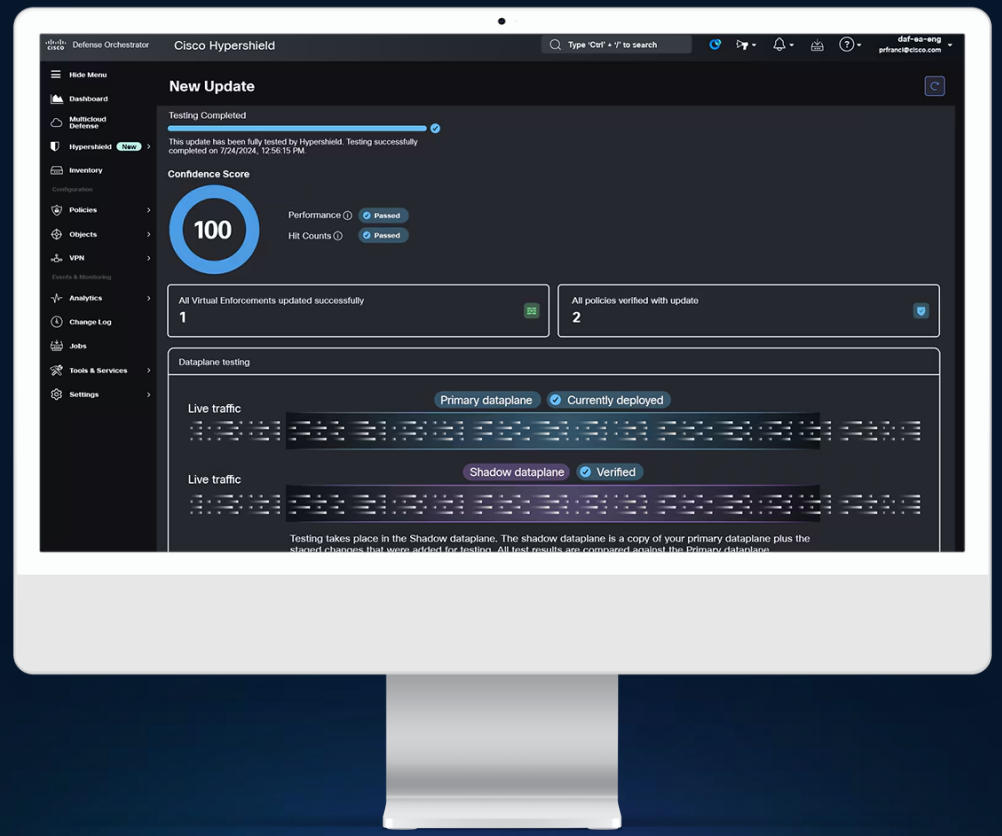


Security at every layer of the stack

Cisco AI Defense



Cisco Hypershield



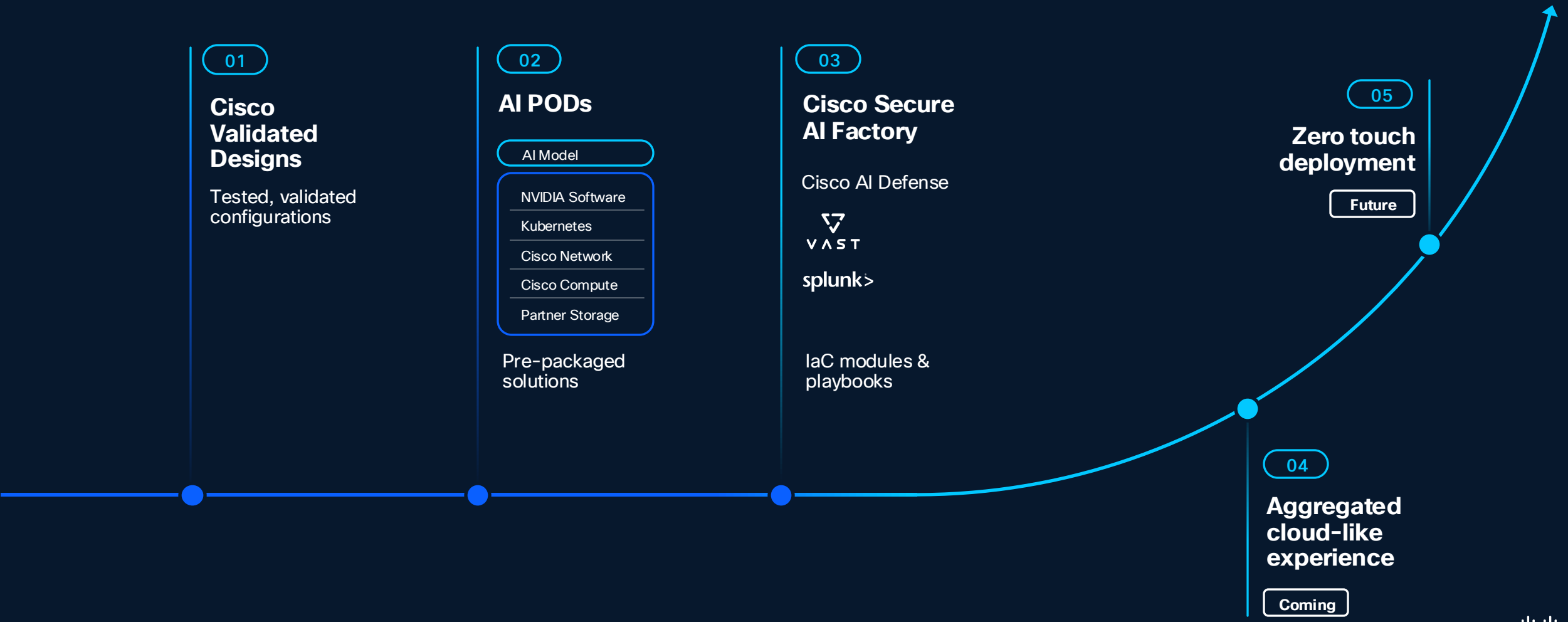
Secure what AI is doing

Secure where AI is running

Simpler solutions, by design

UNDER NDA

Helping customers do more with less



What we've delivered and what's next

Delivered

- Smart Switch
- AI POD for Inferencing
- Nexus 800G Switch
- AI POD for the entire AI lifecycle
- Nexus ONE
- UCS C885A HGX Server
- UCS C845A MGX Server

What's next

- NVIDIA AI Data Platform with VAST
- Splunk Dashboard for AI POD
- UCS with 2U RTX PRO
- Hyperfabric AI
- AI Defense on AI PODs
- Cisco Unified Edge

Infrastructure to power AI

Services to accelerate the value of your investments

Security for AI, AI for security

Data to drive insights and context

Software to unlock productivity

Cisco is bringing these together to make your data center journey easier

Questions?