

# Compute Infrastructure for the AI Era



Dan DeBusschere – Solutions Engineer

Perry Carraway – Account Executive, Architecture

# AI Data Center Track

## 1) Compute & Containers



## 2) Networking: Inter-



## 3) Connecting: Intra-



## 4) Security & Observability



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

## AI-ready data centers

Transform data centers to power AI workloads anywhere

## Future-proofed workplaces

Modernize everywhere people and technology work and serve customers

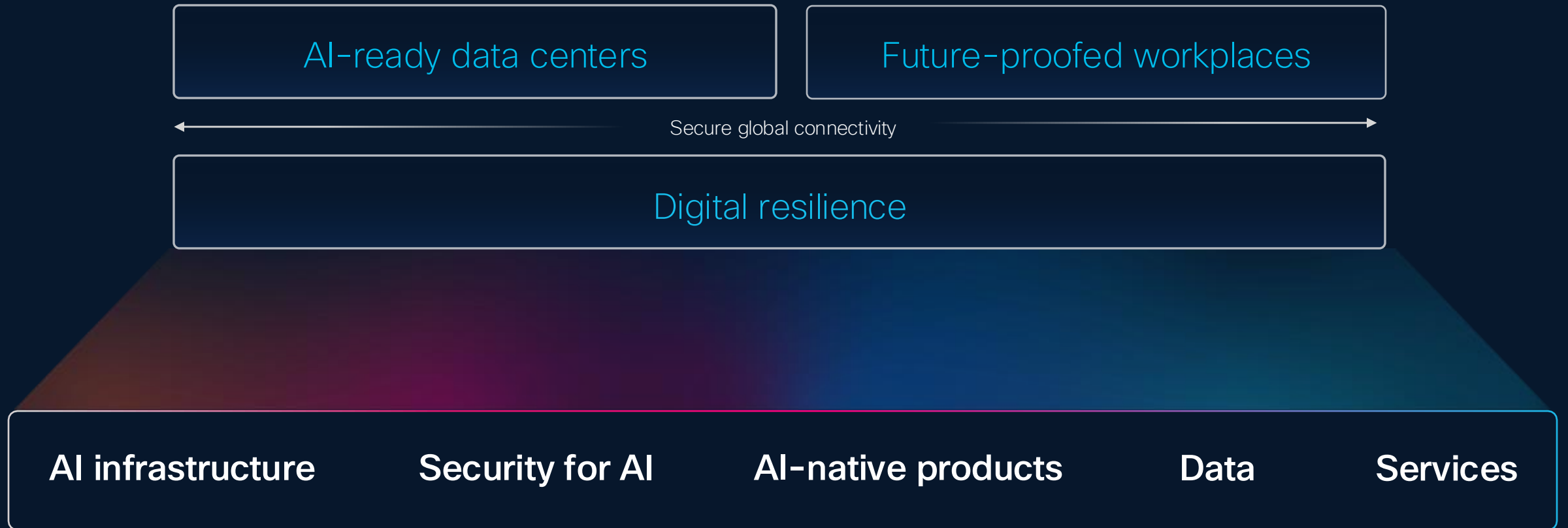
Secure global connectivity

## Digital resilience

Keep the organization securely up and running in the face of any disruption

Accelerated by Cisco AI

# Cisco AI: Accelerating outcomes



# Transform data centers to power AI workloads anywhere



## Robust, flexible infrastructure

Power AI and traditional workloads with networking, silicon, and compute in fully-integrated and scalable systems

## Seamless operations and observability

Remove silos with unified management, observability, and assurance for traditional and AI workloads, across all environments

## Security from ground to cloud

Protect distributed workloads by infusing security everywhere

Our customers tell us

# Standing up enterprise-grade AI infrastructure is hard. Organizations face complexity, fragmentation, and security concerns.



Too many tools,  
not enough integration



Uncertain ROI to build  
AI-ready environments



Security risks across  
the AI lifecycle

# The AI-ready data center



Powers all  
workloads



Scales for  
growth



Unifies  
management

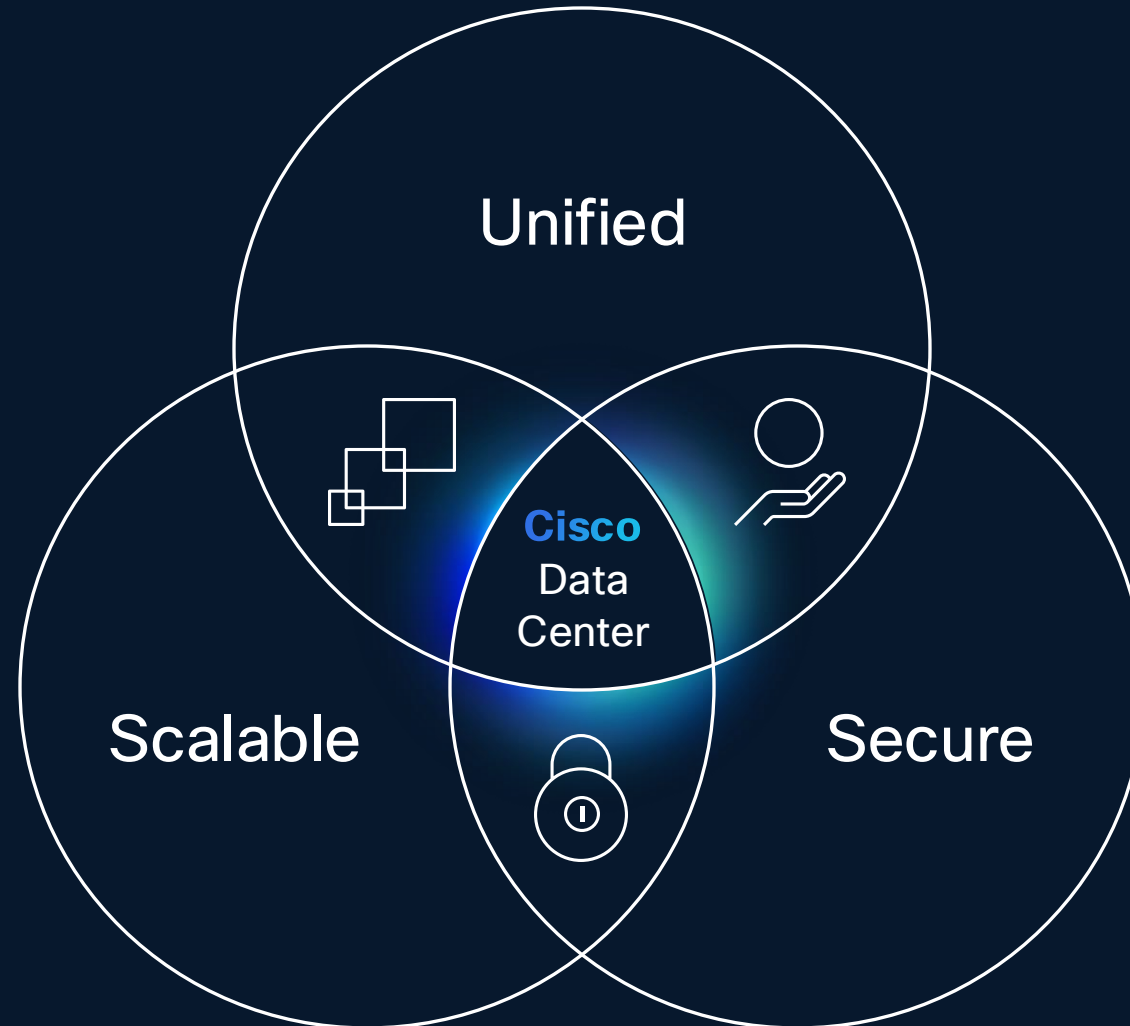


Secures the  
entire stack



Keeps  
resilient

Whether you're modernizing traditional workloads,  
or scaling for AI, Cisco's approach is...



# Our approach



Unified  
architecture



Unified  
operations



Unified  
solutions

Cisco Networking



Cisco Networking



Cisco Networking



Cisco Silicon One



Unified Architecture

# The foundation for enterprise workloads

Cisco Unified Compute



Cisco Unified Compute



Cisco Unified Compute



Cisco Optics



Unified Architecture

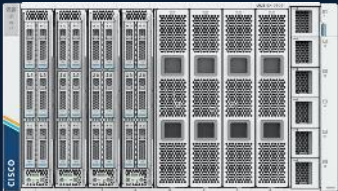
We're future-proofing  
**your infrastructure  
for AI**

Compute for  
**every workload**

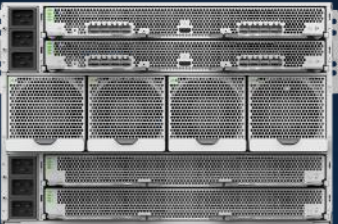
# Cisco UCS Compute Portfolio

## MAINSTREAM ENTERPRISE SERVERS

UCS X-Series  
X9508 Chassis  
IFM Module



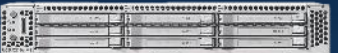
UCS X-Series Direct



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

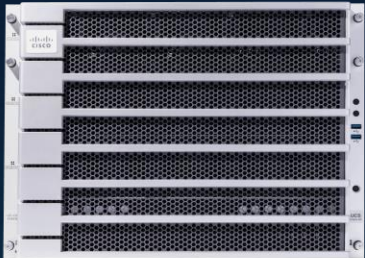


## Edge Platforms

Check [cisco.com](https://cisco.com) on/after November 4<sup>th</sup> for a groundbreaking edge platform announcement

## AI SERVERS

UCS C885A M8  
8RU Dense GPU Server



UCS C845A M8  
4RU MGX Server



Reliability when it matters most

# Cisco tops list for server reliability

According to a survey by the **Information Technology Intelligence Consulting Corp (ITIC)**  
reported by Tech Channel

**TechChannel<sup>®</sup>**

**Server Reliability  
Survey: IBM, Lenovo,  
Cisco Top ITIC's List**

Shipping Now

# 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 4<sup>th</sup> Gen/5<sup>th</sup> Gen  
EPYC Processors

Available as an option in Nexus  
Hyperfabric AI

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  
Nexus Hyperfabric AI

## High performance in a compact form factor

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

with NVIDIA RTX PRO 6000 Blackwell GPUs  
[orderable now]

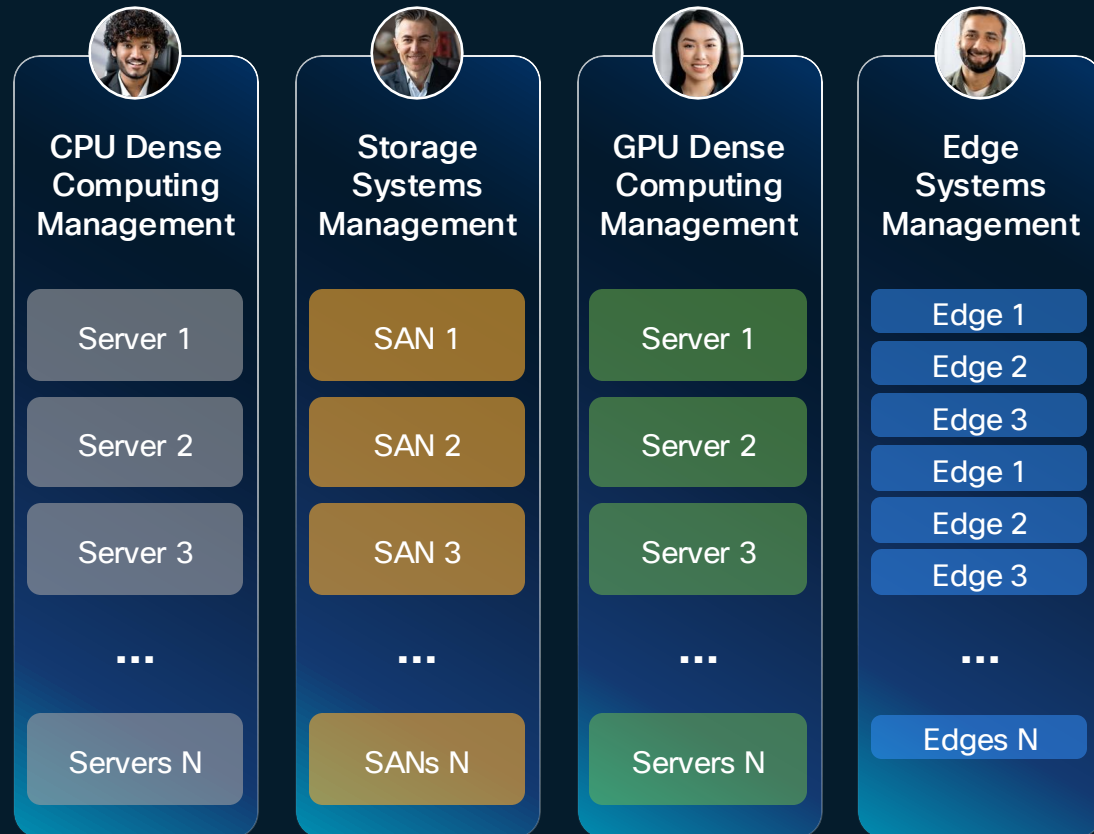
Unified Operations

We're making AI  
**easier than ever**  
**to operate**

# Centralized infrastructure management

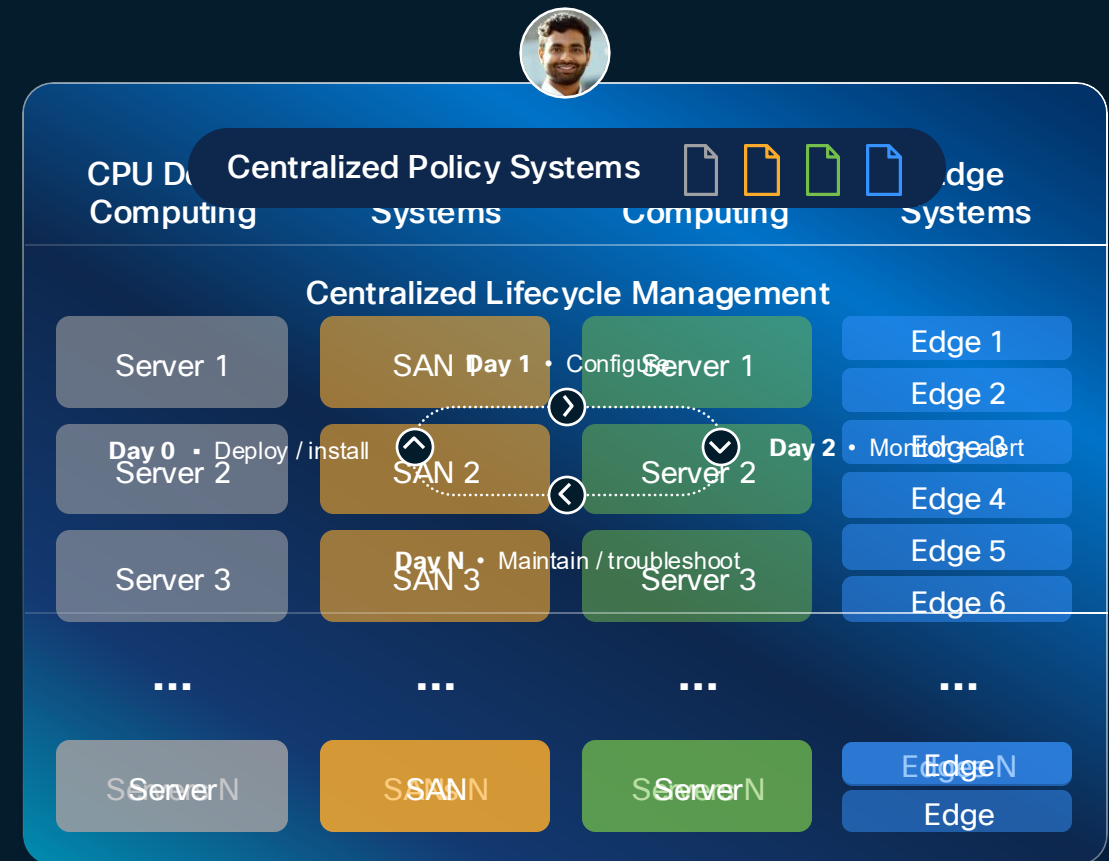
## Conventional approach

Fragmented and Siloed Systems Management



## Cisco Computing System

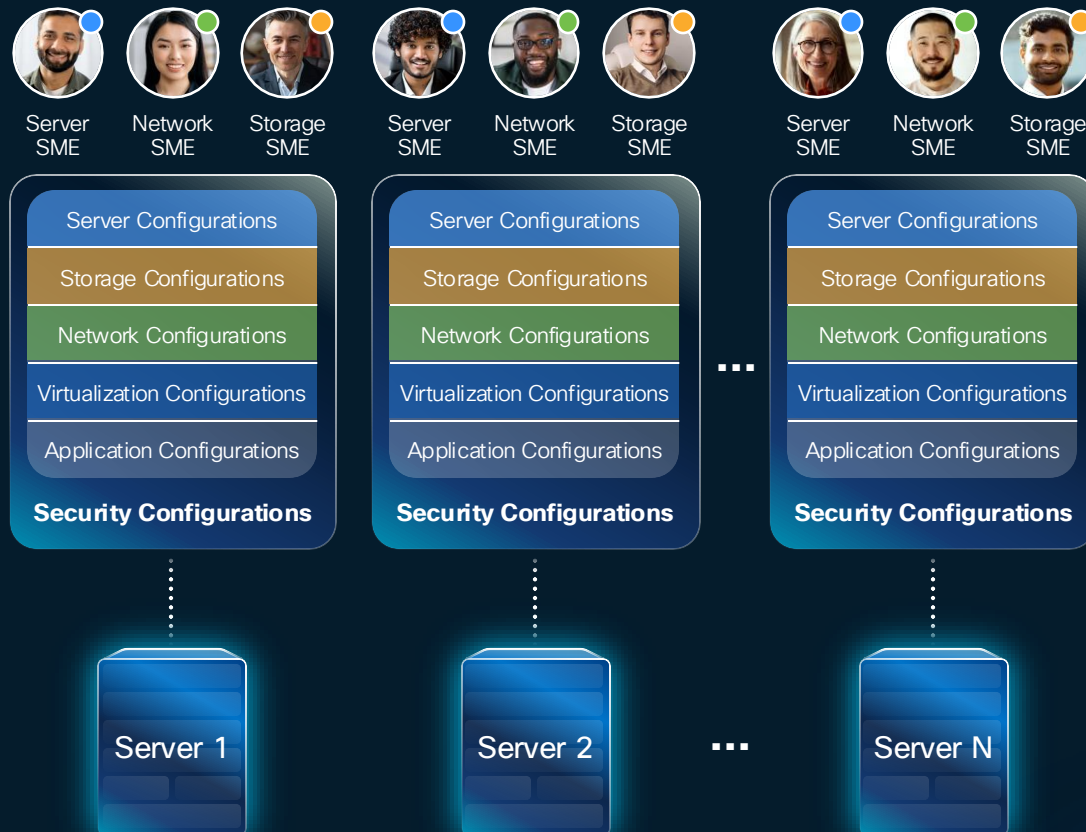
Centralized visibility, policy, automation across entire footprint



# Cisco software-defined computing

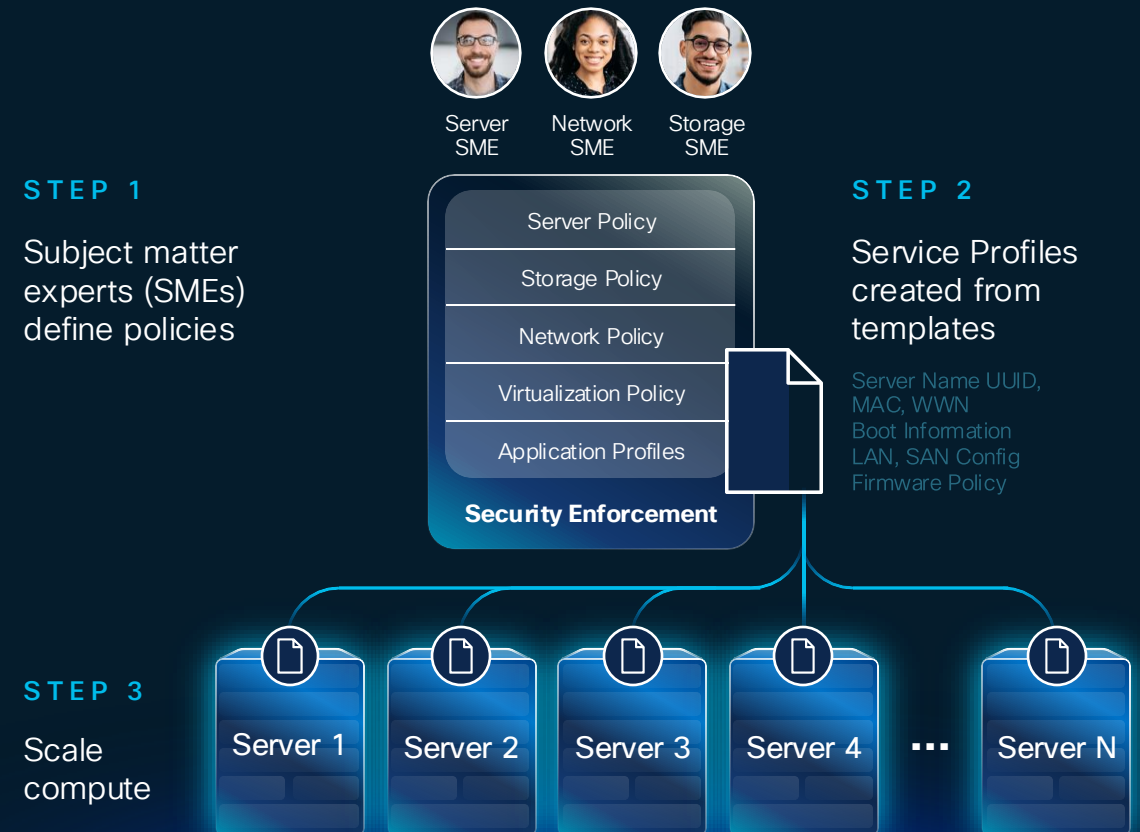
## Conventional Approach

Massive complexity and time to scale compute



## Cisco Computing System

State-less to State-full compute in minutes



# What if you could...



Deploy and configure infrastructure for any type of workload (traditional or AI) in minutes, anywhere?



Monitor and track energy consumption while controlling power management across your server fleet?



Operate and monitor your on-premises and edge environments from one dashboard?



Address potential hardware, OS, or software issues before they impact users, or minimize them when they occur?



Continuously identify and close security gaps?



Give your staff the ability to accomplish more without creating additional burdens?

# You can do all of this with...

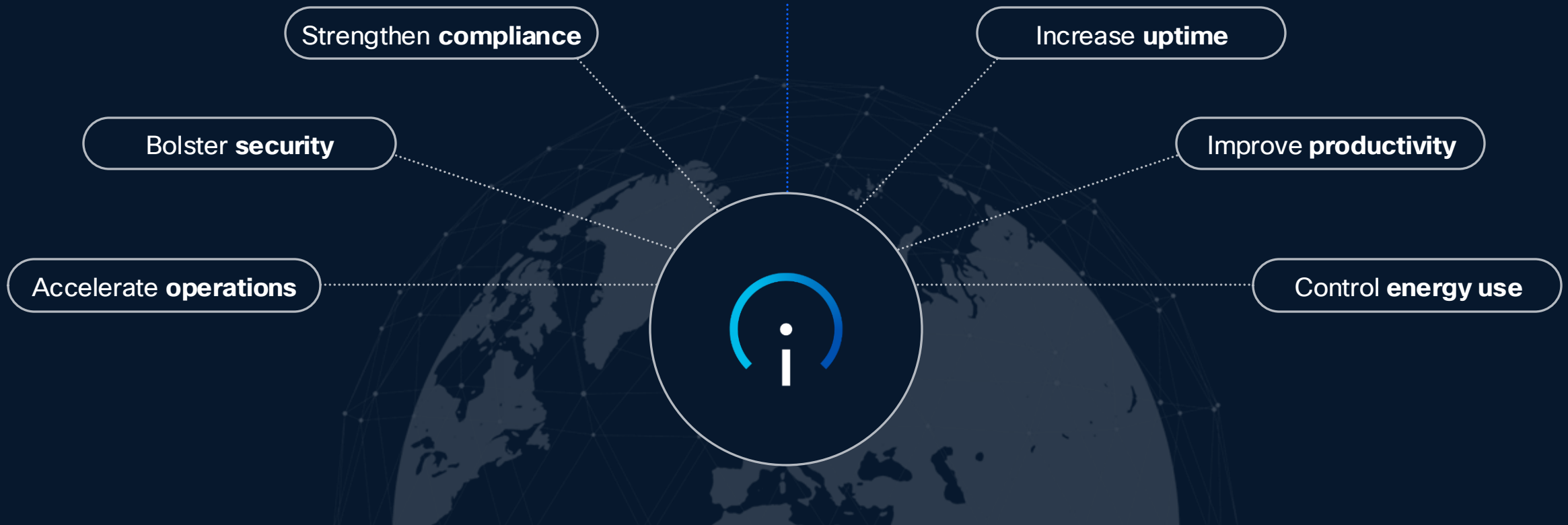


## Cisco Intersight®

See, control, and automate your Cisco UCS®, converged, and hyperconverged infrastructure throughout its lifecycle—wherever it is—from one place

# Intersight - Benefits

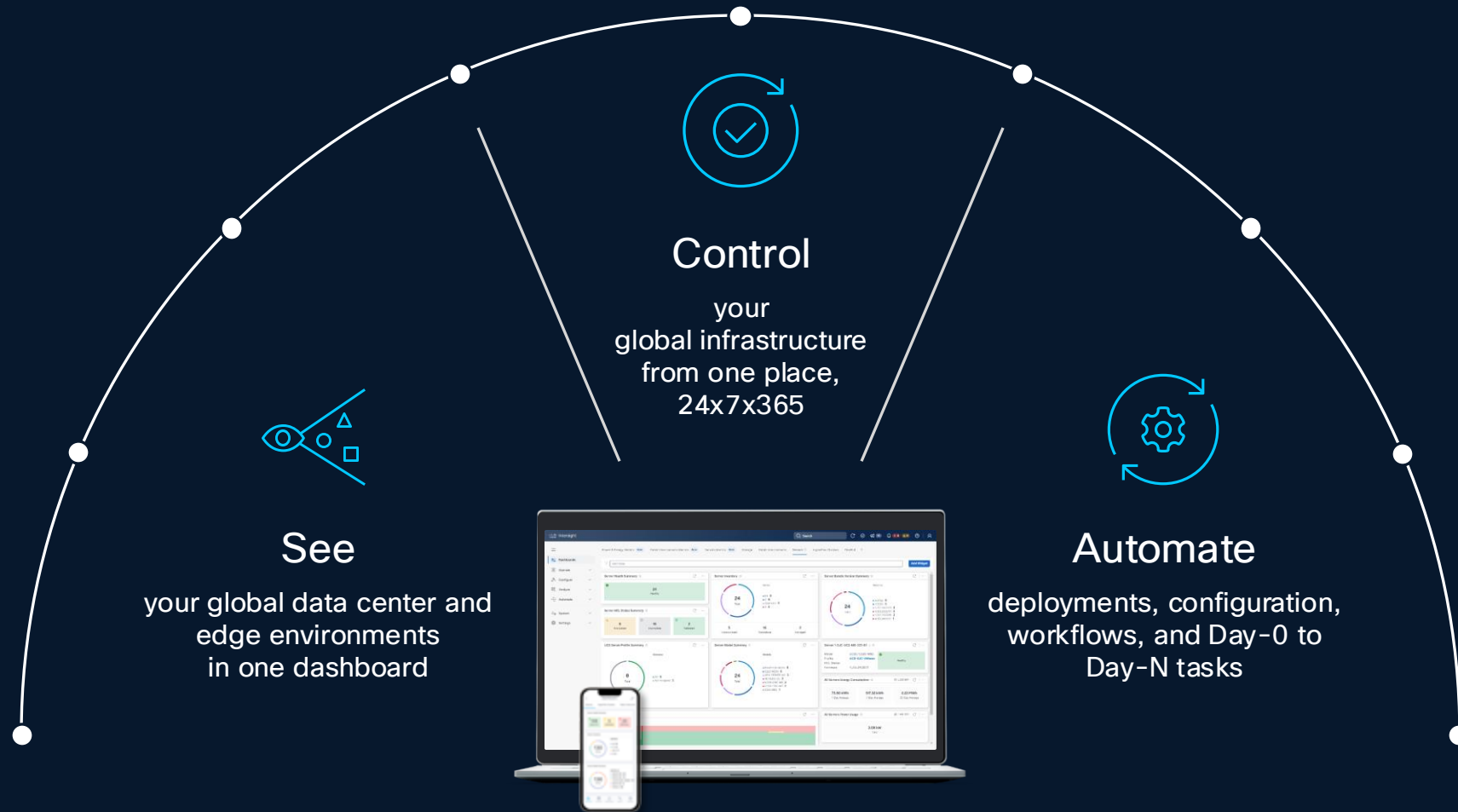
Cisco Intersight®



IT operations. Simplified.

# Intersight - IT operations. Simplified.

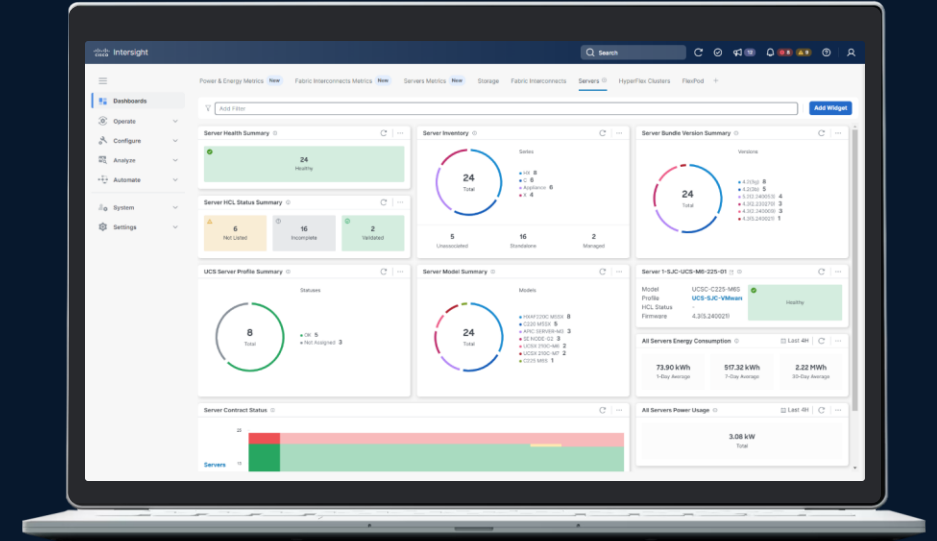
Cisco Intersight®



See

# Intersight – View in one place

- Your global Cisco® Unified Computing System (UCS), converged, and hyperconverged infrastructure
- Servers, virtual machines, storage, and networking
- Real-time physical and virtual device status, UCS domain connectivity issues, security advisories, field notices, and hardware compatibility status
- Selected historical data (up to two years)



The customizable dashboard uses widgets for a single view of the health and inventory status of managed devices.

“[Cisco] Intersight® allowed us to quickly find a problem that would have taken us months to troubleshoot without the visibility that Intersight provides.”

Director of IT and Communications, regional medical center

“With Cisco Intersight, I can now see information that I did not have before. It is one of the first places that we go to when we have performance issues or if a tech upgrade is requested.”

Director of IT, state governmental agency

# Intersight – Comprehensive infrastructure support



## Cisco UCS®

Configure, deploy, operate, and maintain C-Series and X-Series rack and blade servers, anytime and anywhere



## Cisco® converged infrastructure with UCS X-Series

See converged infrastructure inventory and incorporate into orchestrated workflows



## Cisco Compute Hyperconverged with Nutanix

See and control your entire hyperconverged infrastructure fleet in one place—spanning clusters globally



## AI-ready infrastructure

Cisco AI PODs: Deploy, monitor, and support AI inferencing solutions on-premises and at the edge

Cisco Intersight® provides the most comprehensive set of infrastructure management capabilities for any UCS server form factor and most generations in one place.

Other vendors either provide less functionality, cannot cover their full server suite, or require you to use multiple tools.

# Intersight – Ecosystem integration

## Storage plugins



- See storage capacity, controllers, configuration, snapshots, and replication
- Automate storage management and orchestration
- Use storage task library to create and execute workflows

## Virtualization\*



- See virtual machine inventory and supporting infrastructure across on-premises/edge environments
- Standardize management of virtual infrastructure
- Efficiently provision and simplify management of daily operations

*\*Feature set varies by hypervisor*

## ServiceNow plugins



### Service Graph Connector for Cisco Intersight®

Ingest data from Cisco Intersight-connected devices into ServiceNow to resolve issues faster

### Cisco Intersight: Incident Management Integration

Automated Incident Generation: Raises incidents in ServiceNow for alarms and advisories from Cisco Intersight

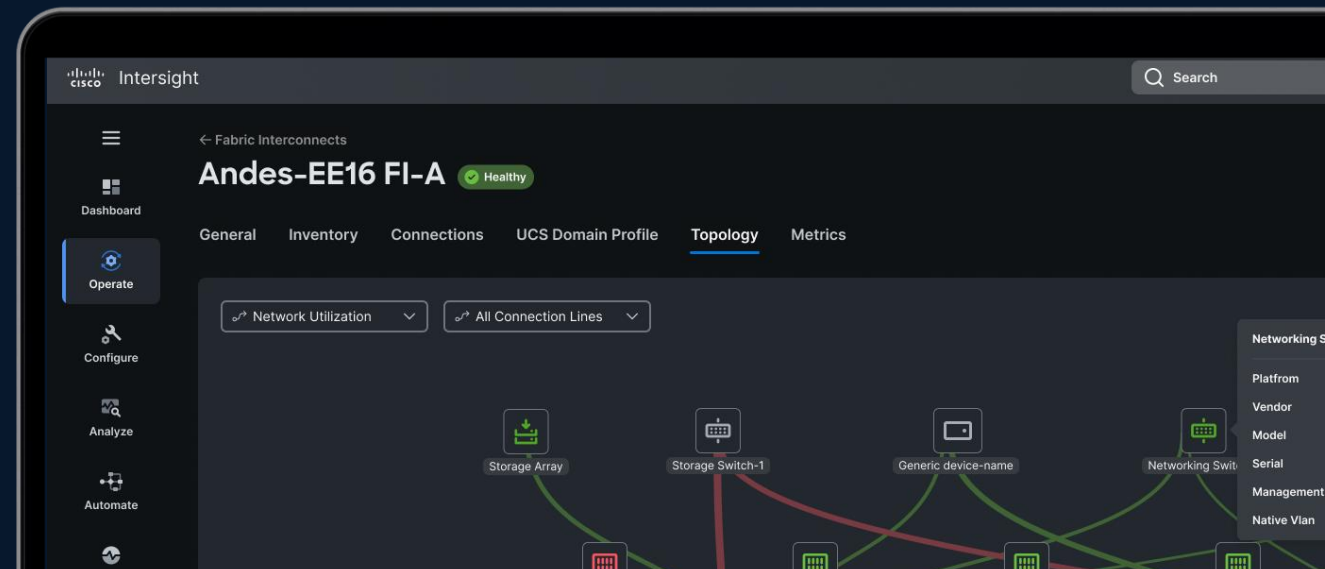
## Cisco Intersight Add-on for Splunk



- Monitor your Cisco UCS infrastructure from Splunk
- Quickly diagnose and act on potential issues
- Leverage Splunk's analytics capabilities to identify potential issues in Intersight-managed UCS infrastructure
- Unify data from compute and other platforms in Splunk to gain a comprehensive understanding of IT operations
- Strengthen security posture by integrating Intersight's infrastructure insights with Splunk Enterprise's advanced threat detection and compliance monitoring

NEW

# Unified intelligence with Cisco Intersight



## Next-gen metrics

Uncover hidden trends, optimize performance, and supercharge operational awareness

## Enhanced topology view

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

## Complete visibility and assurance

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

Unified solutions

**We're expanding our  
solutions to meet  
your needs**

GitHub

Red Hat

NUTANIX

NVIDIA

intel

AMD

vmware  
by Broadcom

CLOUDERA

Bitbucket



Jenkins

NetApp

# Unified Solutions

## The power of partnering with industry-leading tools

servicenow

RANCHER  
BY SUSE

GitLab

HashiCorp

PURESTORAGE

HITACHI

VAST

python

rubrik

Qumulo

ANSIBLE

Unified Solutions

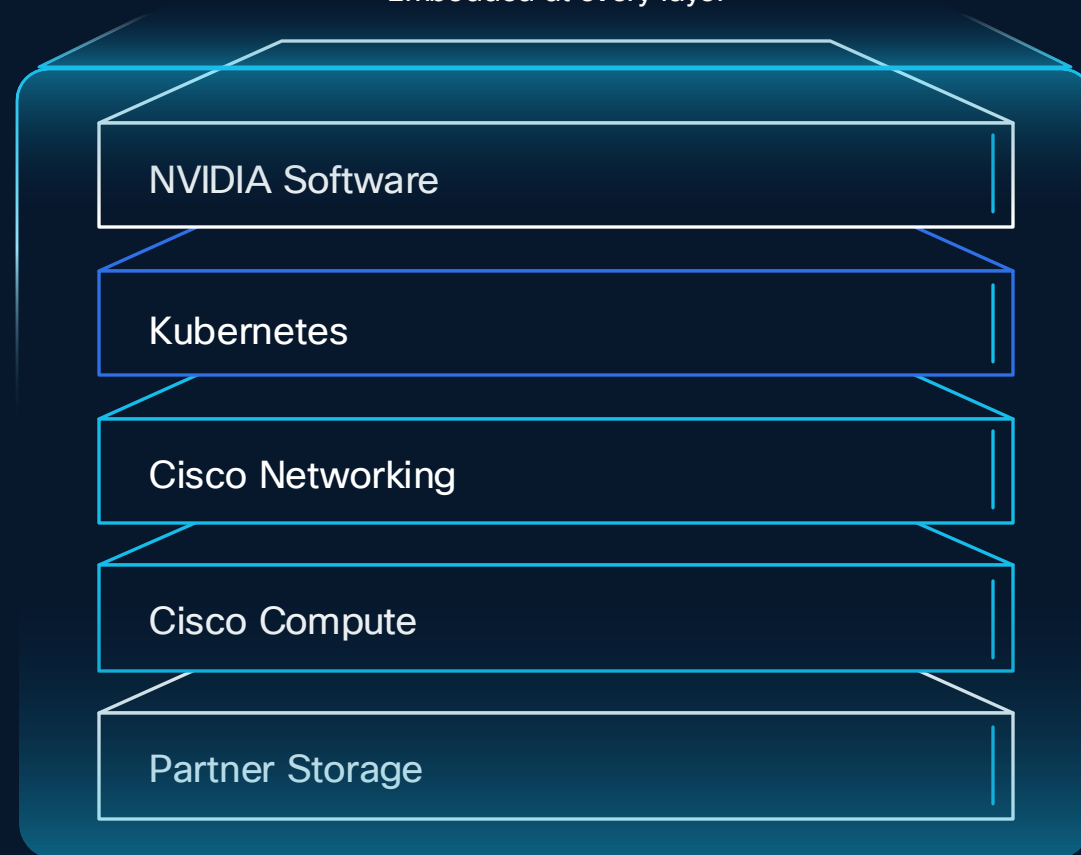
# Cisco Secure AI Factory with NVIDIA

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

Nexus Hyperfabric AI

AI PODs

Cisco Security  
Embedded at every layer



NEW

# Cisco **AI PODs** for entire AI lifecycle

Scalable architecture, support any AI workload

GitHub

Red Hat

NUTANIX

NVIDIA

intel

AMD

vmware  
by Broadcom

CLOUDERA

Bitbucket



Jenkins

NetApp

# Unified Solutions

## The power of partnering with industry-leading tools

servicenow

RANCHER  
BY SUSE

GitLab

HashiCorp

PURESTORAGE

HITACHI

VAST

python

rubrik

Qumulo

ANSIBLE

# Cisco Compute Hyperconverged With Nutanix

A better together partnership where 1+1=11

## Strategy

Complete roadmap leveraging the strengths of both companies in service of our customers

## Solutions

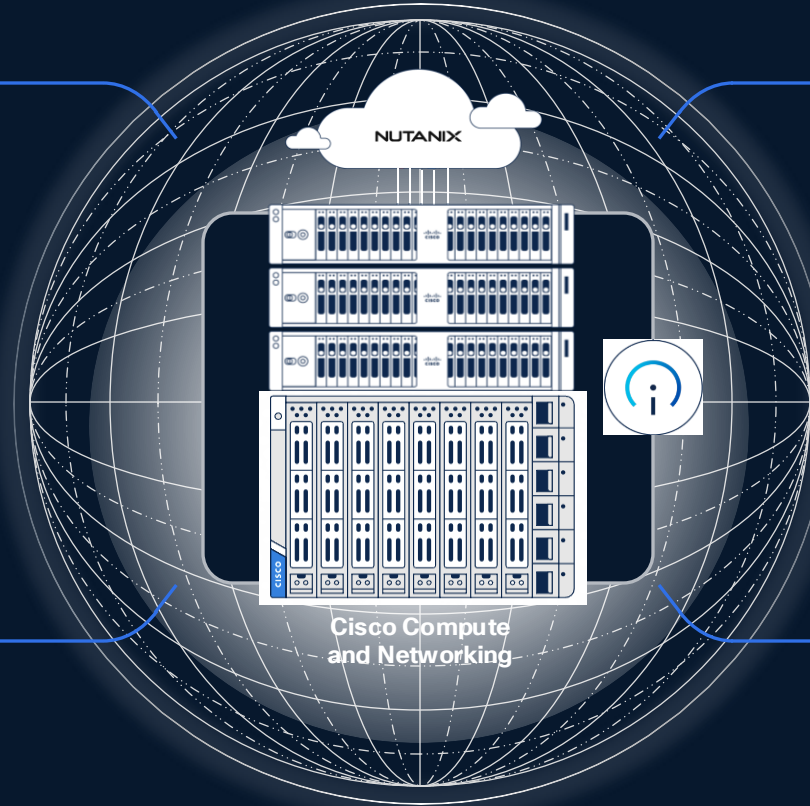
Built, managed, and supported holistically for a seamless end-to-end experience

## Go to Market

A buying process focused on simplifying the seller and customer experience

## Support

A joint enhanced support model built for information-sharing to ensure our customers' success



Operate at scale

More choice and flexibility

End-to-end security

Comprehensive support

Easy to buy

GitHub

Red Hat

NUTANIX

NVIDIA

intel

AMD

vmware  
by Broadcom

CLOUDERA

Bitbucket



Jenkins

NetApp

# Unified Solutions

## The power of partnering with industry-leading tools

servicenow

RANCHER  
BY SUSE

GitLab

HashiCorp

PURESTORAGE

HITACHI

VAST

python

rubrik

Qumulo

ANSIBLE

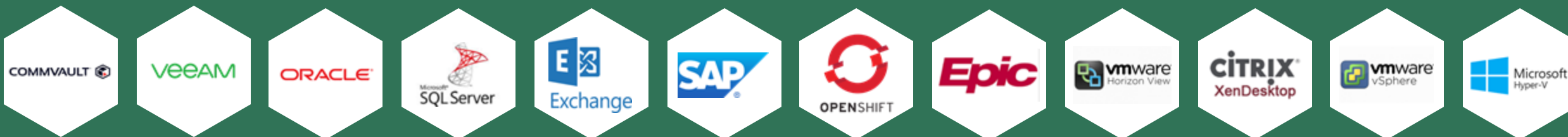


61+  
CVDs



- Full Stack solution tested and validated by Cisco and Pure Storage
- Designed, implemented, and supported by Pure partners
- Storage, Network, Compute: Best of breed components, flexible design.

FlashStack Solutions Are Available For:



GitHub

Red Hat

NUTANIX

NVIDIA

intel

AMD

vmware  
by Broadcom

CLOUDERA

Bitbucket



Jenkins

NetApp

# Unified Solutions

## The power of partnering with industry-leading tools

servicenow

RANCHER  
BY SUSE

GitLab

HashiCorp

PURESTORAGE

HITACHI

VAST

python

rubrik

Qumulo

ANSIBLE

# FlexPod – Customer Benefits



Virtualization

Cisco UCS Servers

Cisco Nexus / MDS  
Networking

NetApp Storage

vmware

Microsoft

Red Hat

CISCO



Bare  
Metal

NetApp™



## Greater Consolidation

FlexPod allows enables network consolidation and reduces unnecessary costs on standby racks and storage by half that saves millions of dollars.

## Faster Time to Value

Cisco Validated Designs (reference architectures) helped reduce by 20% the time required for senior IT staff to manage network architecture.



## Simplified Deployment

FlexPod is designed to simplify and accelerate the deployment of virtualized data center infrastructure.



## Scalability

FlexPod can scale up or down to meet changing business requirements and support growth.



## AI-Ready

Pre-validated reference designs for Gen AI, RAG, and MLOps



## Improved Performance and Scale

The combination of Cisco UCS and NetApp storage provides verified, superior performance; scale and reliability for all your applications.



## Support

FlexPod is backed by Cisco and NetApp, providing customers with a single point of contact for support and service.

GitHub

Red Hat

NUTANIX

NVIDIA

intel

AMD

vmware  
by Broadcom

CLOUDERA

Bitbucket



NetApp

# Unified Solutions

## The power of partnering with industry-leading tools

servicenow

RANCHER  
BY SUSE

GitLab

HashiCorp

PURESTORAGE

HITACHI

VAST

python

rubrik

Qumulo

ANSIBLE

# VAST data

The data platform for the AI era

## VAST Data Platform

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

Data

DataStore

DataBase



DataEngine



DataSpace

Edge-to-Cloud Namespace

Intelligence



Unified data architecture



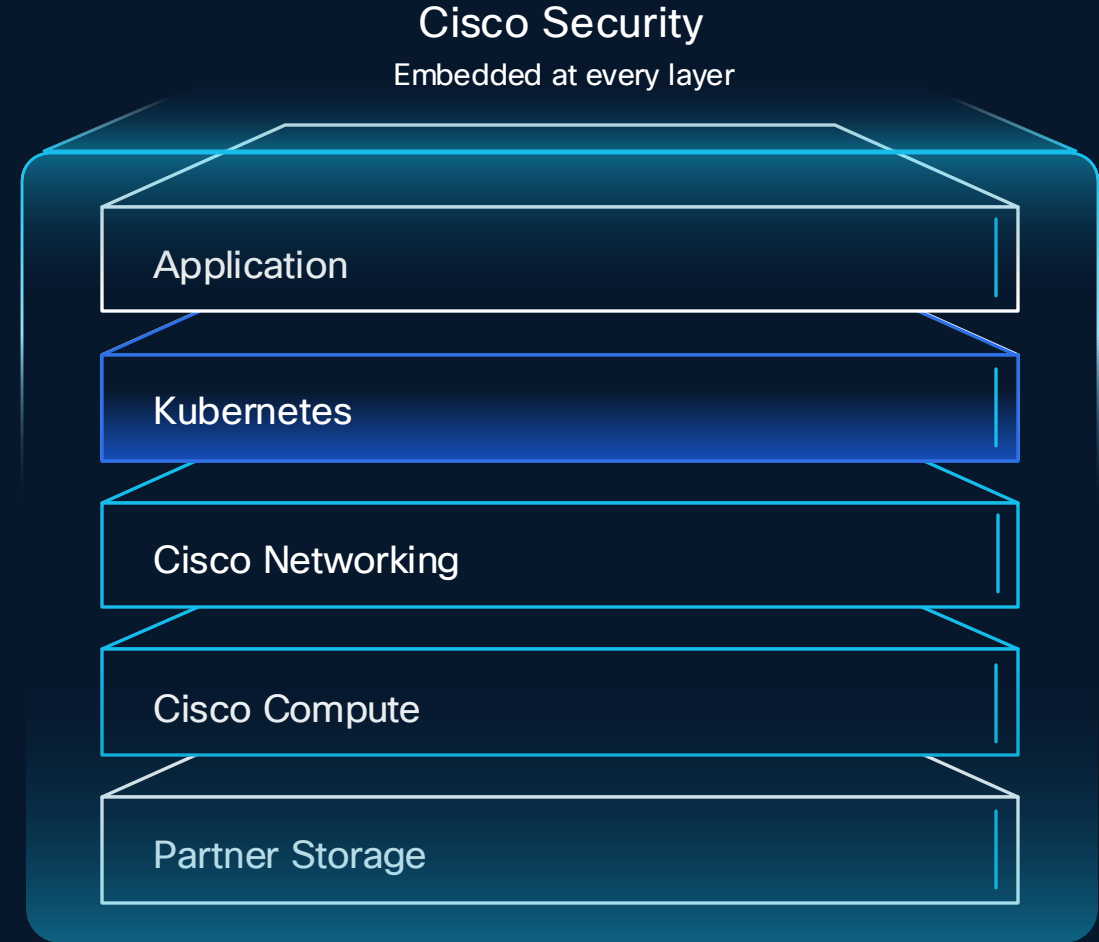
Real-time performance



Enterprise-grade  
security & compliance

Cisco is fusing **Security**  
into every layer of the stack

**It's not just about AI:  
Cisco offers full-stack  
readiness for  
every workload**



# Cisco Security & Observability

Embedded at every layer

Application

Virtualization

Modern Apps

AI/ML

## Kubernetes



Virtual  
Networking



Firewall /  
Microsegmentation

## Isovalent



Load  
Balancing



Network  
Visibility

Enterprise  
Platform



eBPF



VPN/  
Encryption



Runtime  
Security



cilium

Cisco Networking

Nexus

Nexus Dashboard

Cisco Compute

UCS

Intersight

Partner Storage



Cisco and Isovalent

# What Isovalent brings to UCS Customers



Extends the value of the Cisco portfolio into the platform layer



Enables workload portability

# What this means for your teams...



Splunk Enterprise Security  
Sec Ops



Hypershield  
Network Sec



Splunk Observability  
Developers



Nexus, ACI, Isovalent LB  
Network Ops

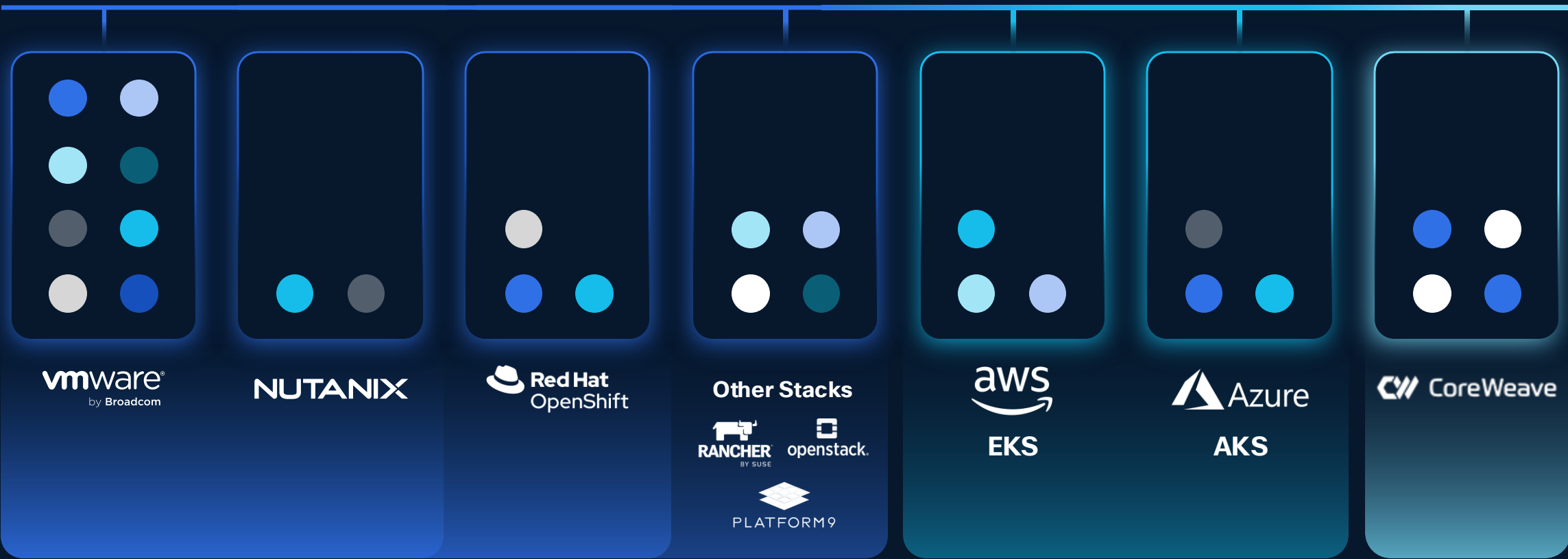
## Server Ops

Intersight, Cisco Validated Designs

# The future of Cisco and Isovalent



# Connect & Secure Workloads Anywhere



Private Cloud

Public Clouds

AI Clouds

Why we believe this is a game-changer

# Envisioning a world where you can migrate with confidence and modernize without disruption

## Security

without  
Silos

## Migration

without  
Disruption

## Visibility

without  
Complexity

## Choice

without  
Compromise

Full or partial  
hypervisor migrations

Multi-tenant  
environments

AI and traditional  
workloads

East-west traffic  
visibility and control



# Cisco Data Centers

## Unified | Scalable | Secure





Software to  
unlock productivity



Security for AI,  
AI for security



Infrastructure  
to power AI



Data to drive insights  
and context



Services to accelerate  
the value of your  
investments

**Cisco is bringing  
these together to make  
your enterprise AI journey easier**

# Resources to learn more



## Cisco Compute

View on [cisco.com/go/ucs](https://cisco.com/go/ucs)



## AI-Ready Infrastructure

View on [cisco.com](https://cisco.com)



## Isovalent Enterprise Platform

Visit [Isovalent.com](https://Isovalent.com) (now part of Cisco)



## Cisco Compute YouTube channel

View on [youtube.com](https://youtube.com)



## Blogs

Visit [blogs.cisco.com/datacenter](https://blogs.cisco.com/datacenter)



## Online community

Visit the [Data Center and Cloud online community](https://Data Center and Cloud online community)



**Thank you**

