

From Data Center to Edge:

Accelerating AI Inferencing Where It Matters Most

Danny McGinniss

VP, Product Management, Cisco Compute

cisco Live !

Everyone's talking about value.

82%

Feeling the heat on ROI in the last 6 months

69%

Making AI a top IT budget priority



But not everyone is ready to act

Only

13%

Are fully AI-ready



Questions your teams are asking...

Is our infrastructure
truly ready to scale
with AI demands?

Is our infrastructure
aligned to the
actual AI models
we need to run?

Are we prepared to
manage AI across
hybrid and edge
environments?

Can we access,
govern, and
trust our data for
AI use cases?

Can we secure
sensitive data
across the
AI lifecycle?

Are we building
to support new
AI without having
to rip and replace
in a year?

Are our AI tools
and platforms
integrated—or
creating friction?

Can we control
costs while proving
ROI on AI
investments?

Do we have the
talent—and
buy-in—to
operationalize AI?

Industries are transforming

Retail



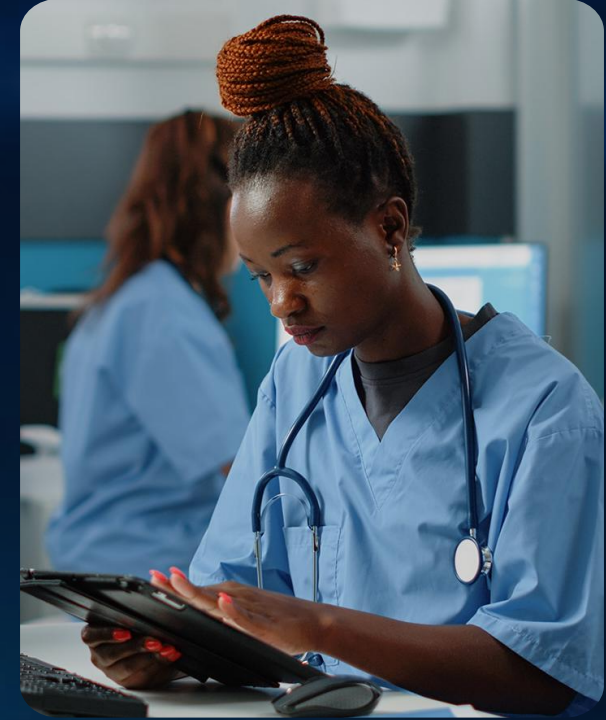
Operational efficiency
Loss prevention
Store of the future

Manufacturing



Streamlined operations
Risk management
Quality control

Healthcare



Improved patient outcomes
Data privacy
Enhanced patient care

Driving greater demand

for edge solutions

Retail

NCR V^oYIX

WaitTime™

PreciTaste®

Quividi

Manufacturing

Rockwell Automation

AVEVA Schneider Electric

EPIC iO

ultralytics

axxonsoft

Healthcare

Medical Informatics

chooch

Cross Vertical AI

nvidia
METROPOLIS

intel
openVINO™

Red Hat AI
Inference Server

Arcee^{AI}

ONSTAK

VAIDIO®



More than ever
Business and IT
must move as one

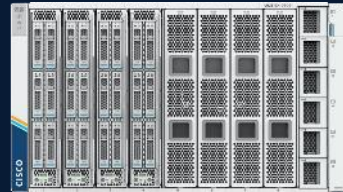
The core and the edge are no longer separate worlds

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

Unified infrastructure

Unified fabric across any workload, anywhere

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

New



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/SSD

New



UCS C225 M8S
10 HDD/SSD

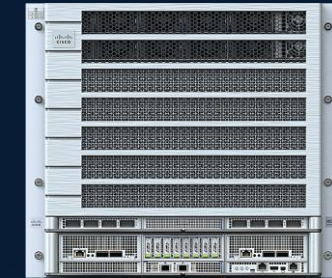
New



UCS C225 M8N
10 NVMe

AI Servers

New



UCS C885A M8
UCS C880A M8
8RU Dense GPU Server

...and even more
HGX follow ons

Coming

New



UCS C845A M8
4RU MGX Server

Edge

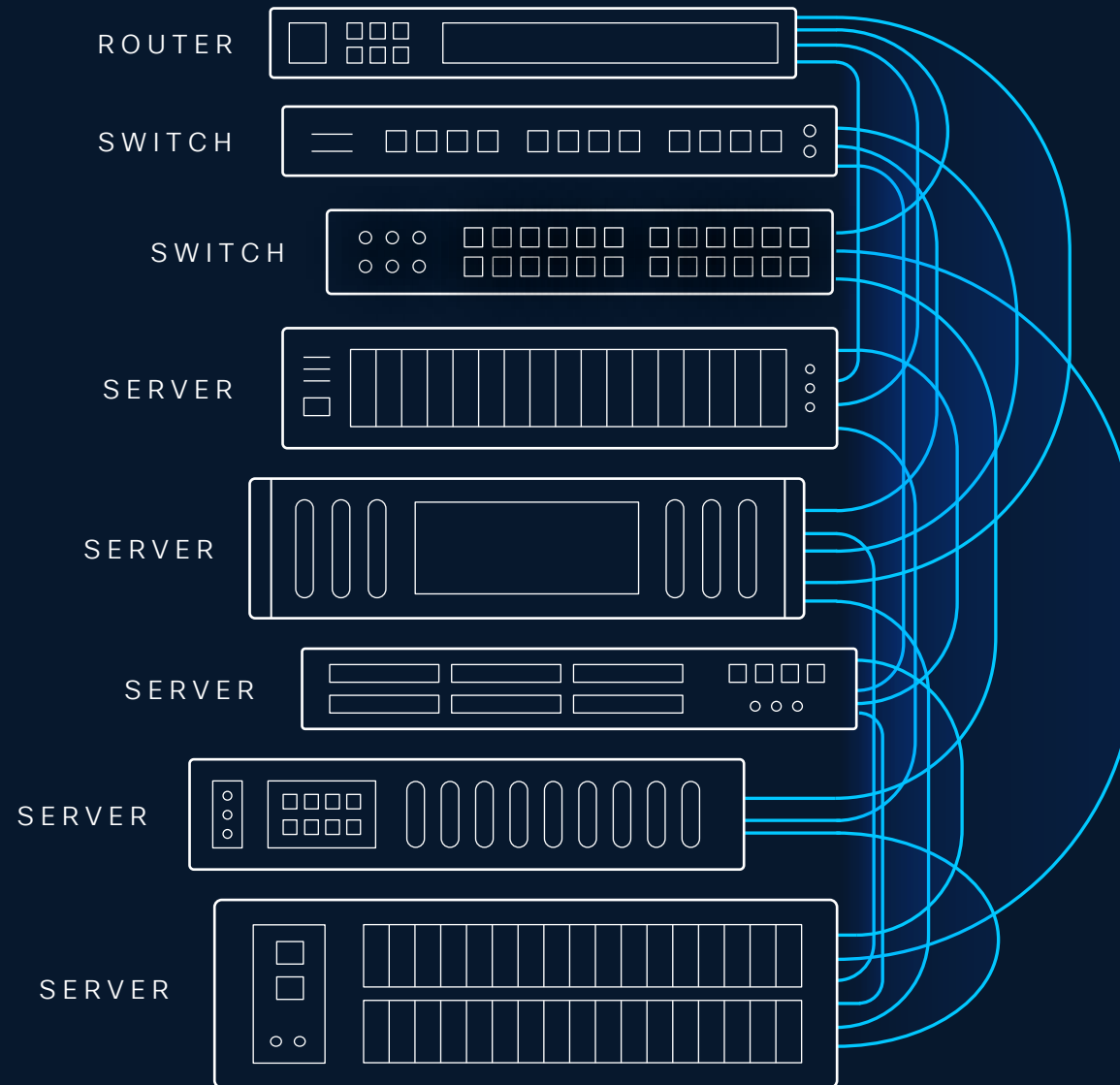
Coming



UCS XE9305 Chassis
UCS XE130c M8
Compute Nodes

The enterprise
is ready

but the
infrastructure
is lagging



25 cables

Introducing

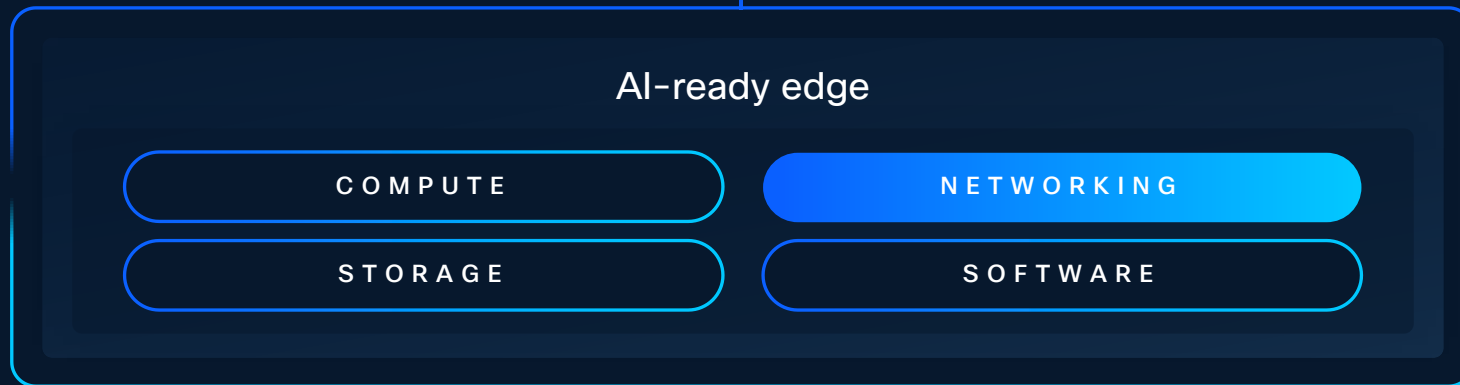
Cisco Unified Edge



Introducing Cisco Unified Edge

Not a server, a platform.

Unified operations from core to edge



Security fused into the network

Modular chassis

Interchangeable networking
and compute nodes

5 s



Node personalization

Networking

Enhanced
PoE support



Redundancy

2 x 25

SWITCHES

Failover protection



Performance

250G max / 50G per node



I N T E R N A L I N F E R E N C I N G A N D D A T A T R A N S F E R

• • • • •

Performance

10G to 100G

NORTHBOUND WAN

100G

EXTERNAL PERFORMANCE



Performance



AI-READY GPUs



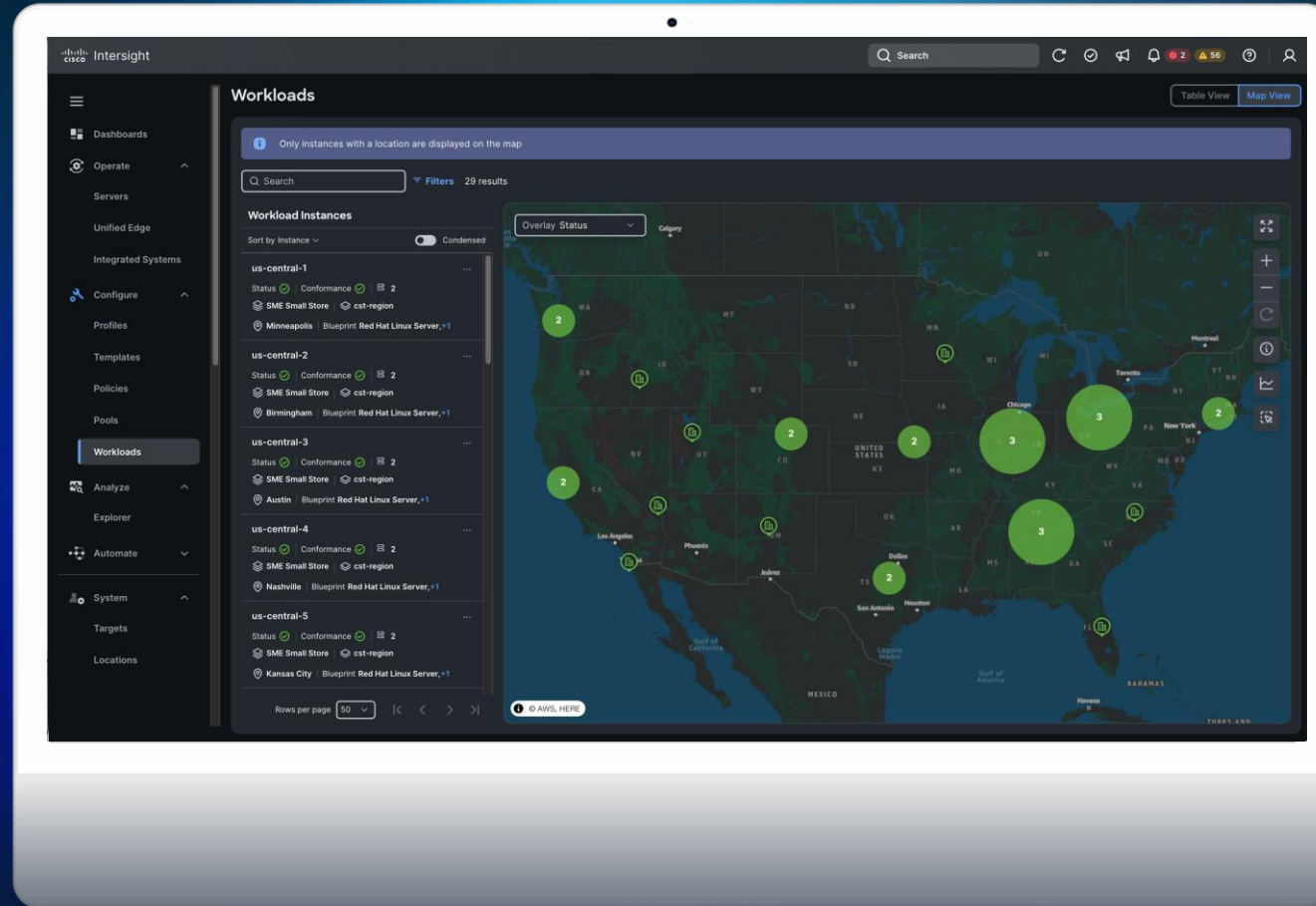
INTEL XEON 6 SOC
PROCESSOR





A growing ecosystem of partnerships building services on top of Cisco Unified Edge

Simplified operations from core to edge



DAY 0

DAY 1

DAY N

**Zero-touch
provisioning**

**Deployment
at scale**

**Fleet-wide
operations**



ZERO TOUCH
PROVISIONING

9:41

<

Targets

Name

sjc02-terminalA-gate4

Status

✓

Claimed

Model

model-name

Serial

N37HHE

Target Type

Unified Edge

Vendor

Cisco Systems, Inc.

Claim Date/Time

Feb 26, 2025 1:03 PM

✓

ECMC A Details

>

✓

ECMC B Details

>

Location

ACME NY 21-2

>

🏠

Targets

📄

Logs

⚙️

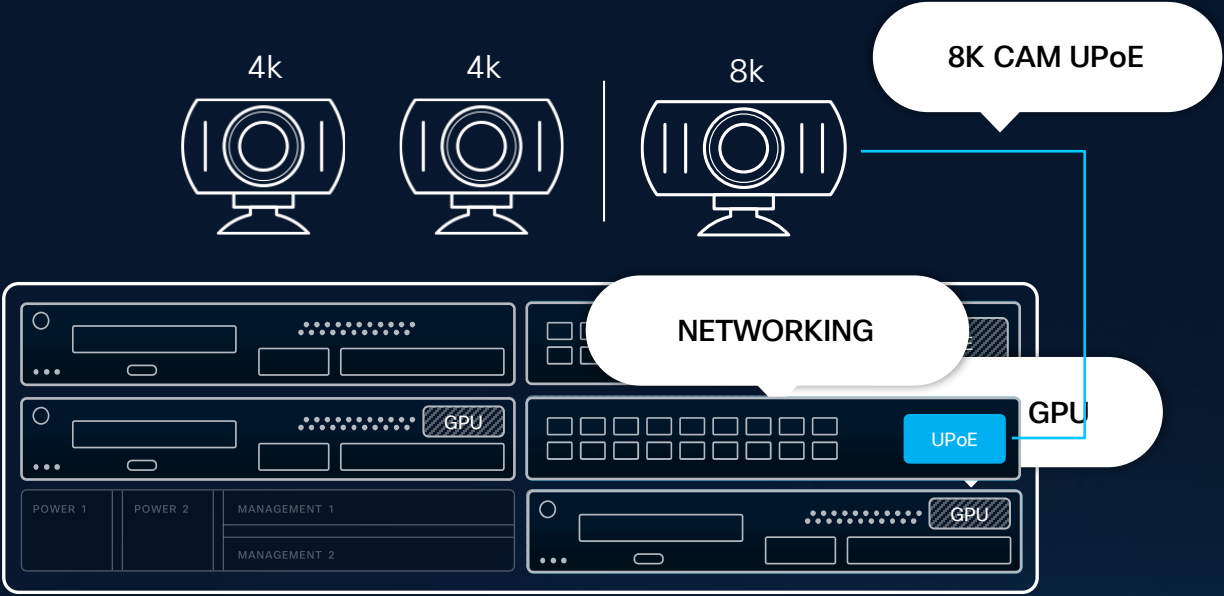
Settings



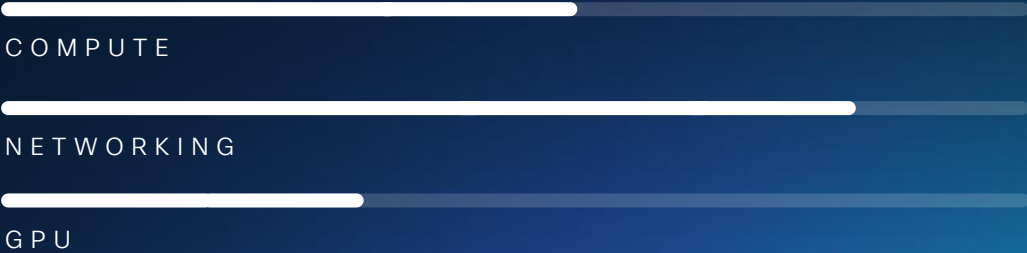


DEPLOYMENT
AT SCALE

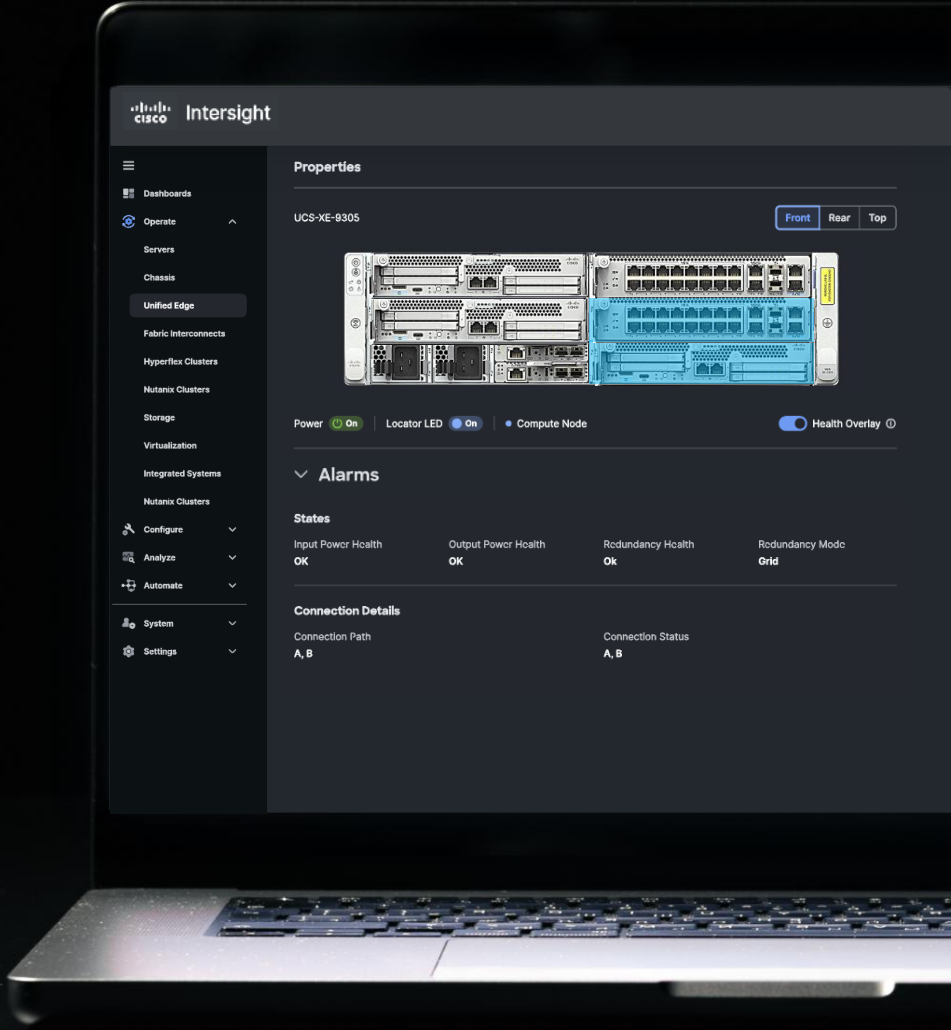
MANUFACTURING SITE : NEW JERSEY



Capacity



CENTRAL IT : NEW YORK



The image features a dark blue, abstract background with flowing, ethereal light patterns. In the upper right, a metallic, reflective object, possibly a car's headlight or a futuristic device, is partially visible. The text "FLEET-WIDE OPERATIONS" is printed in white, uppercase letters on the metallic surface. The overall aesthetic is high-tech and modern.

FLEET-WIDE
OPERATIONS

Workloads

Table View

Map View

* All  My American +

 Search

Instances

Sort by Name ▾

NJ Site 01

Status  Conform

 Manufacturing S

Actions ▾

Workload Definition

Predictive Maintenance Blueprint: [Data Analytics Model](#)

Defect Detection Blueprint: [Computer Vision Model](#)

Robotics and Safety Blueprint: [Large Vision Model](#)

Would you like to deploy across 3 new locations?

Clone and deploy

Workloads

Table View Map View

* All  My American +

Q Search

Filters

Actions

Instances

Sort by Name

Condensed

NJ Site 01

Status  Conformance  

Manufacturing Site Manufacturing Site NJ

TX Site 01

Status  Conformance  

Manufacturing Site Manufacturing Site TX

CA Site 01

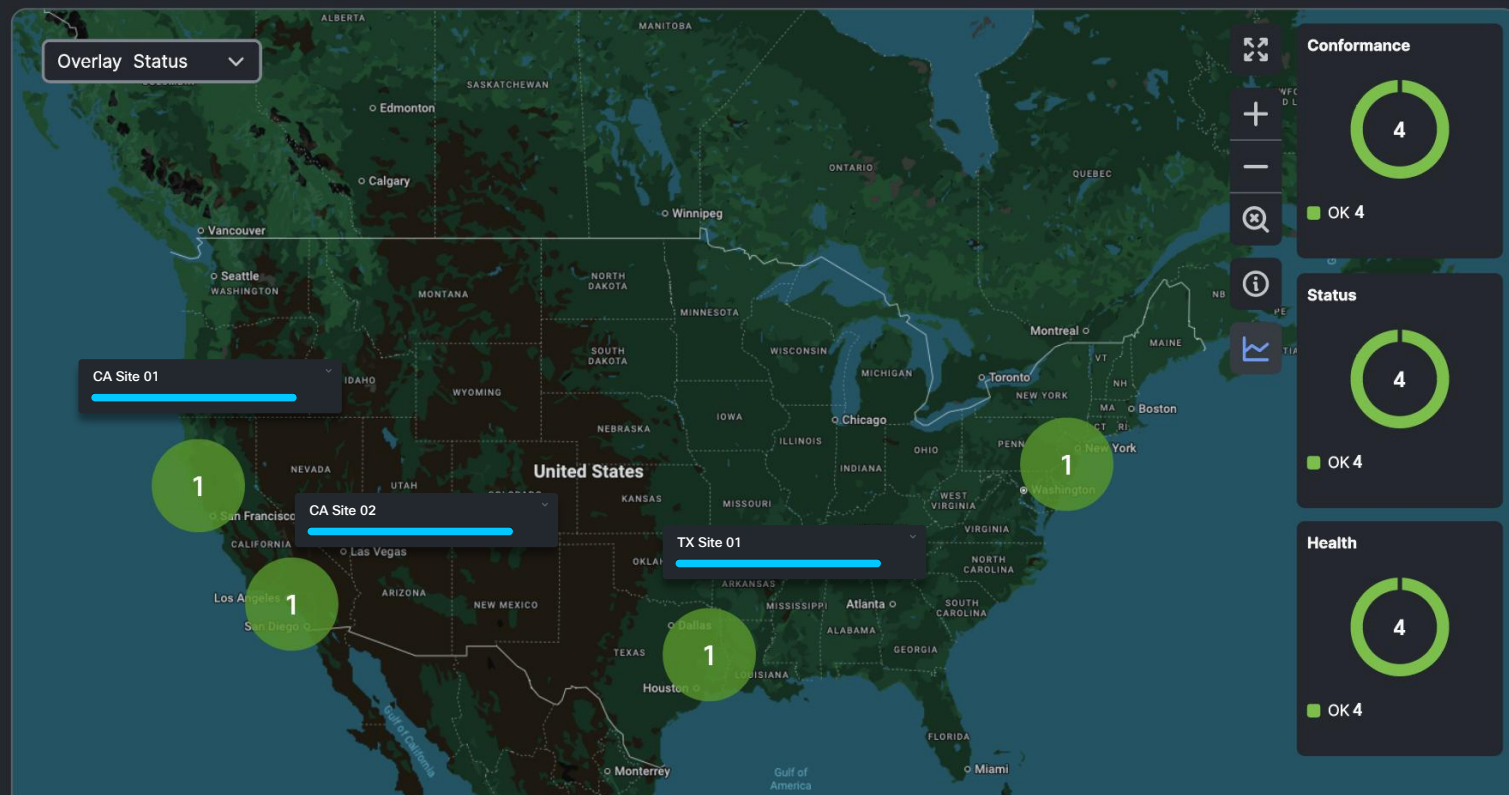
Status  Conformance  

Manufacturing Site Manufacturing Site CA

CA Site 02

Status  Conformance  

Manufacturing Site Manufacturing Site CA



Security at every layer to protect the edge

Security operations and visibility

Built in

HARDWARE-BASED
PROTECTION



Add on

SECURE AI

ZERO TRUST
ACCESS

SEGMENTATION

Only Cisco can

Converge network, compute, storage, and security 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 center



Proven track
record in enterprise
infrastructure

