CISCO Connect

Tying It All Together With Data Center Security and Observability

Brenden Buresh-CCIE #2073, Distinguished Solutions Engineer

April 7, 2026

Cisco Powers How People and Technology Work Together Across the Physical and Digital Worlds

AI-Ready Data Centers

Future-Proofed Workplaces

Secure Global Connectivity

Digital Resilience

Accelerated by Cisco AI

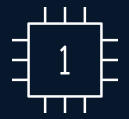
Cisco Data Center Strategic Priorities

Cisco AI Networking

Increase Performance

Simplify Operations

Security Fused into Every Layer



Silicon

Systems

Software

Optics

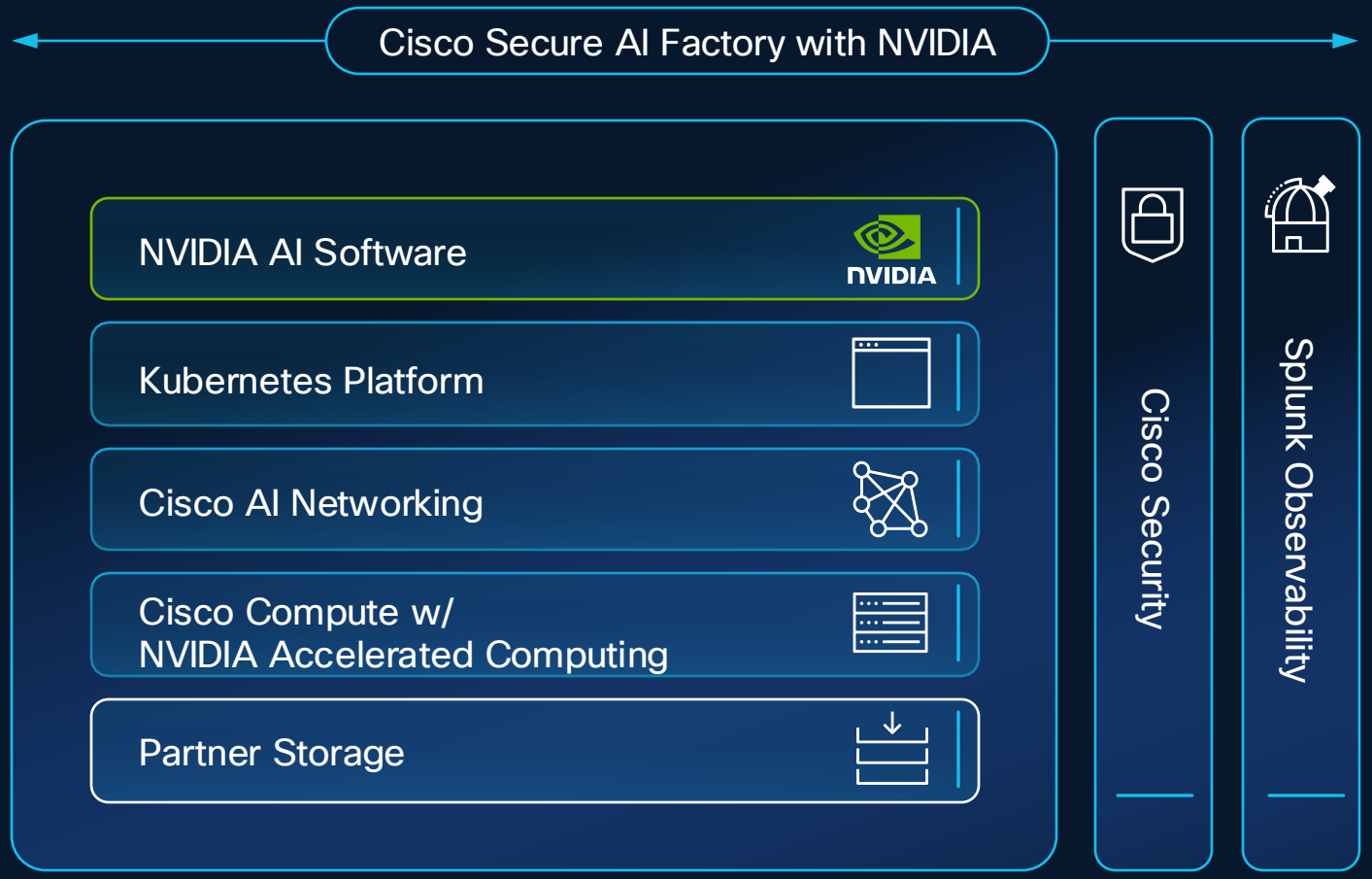
Operating Model

cisco

Cisco Secure AI Factory With NVIDIA

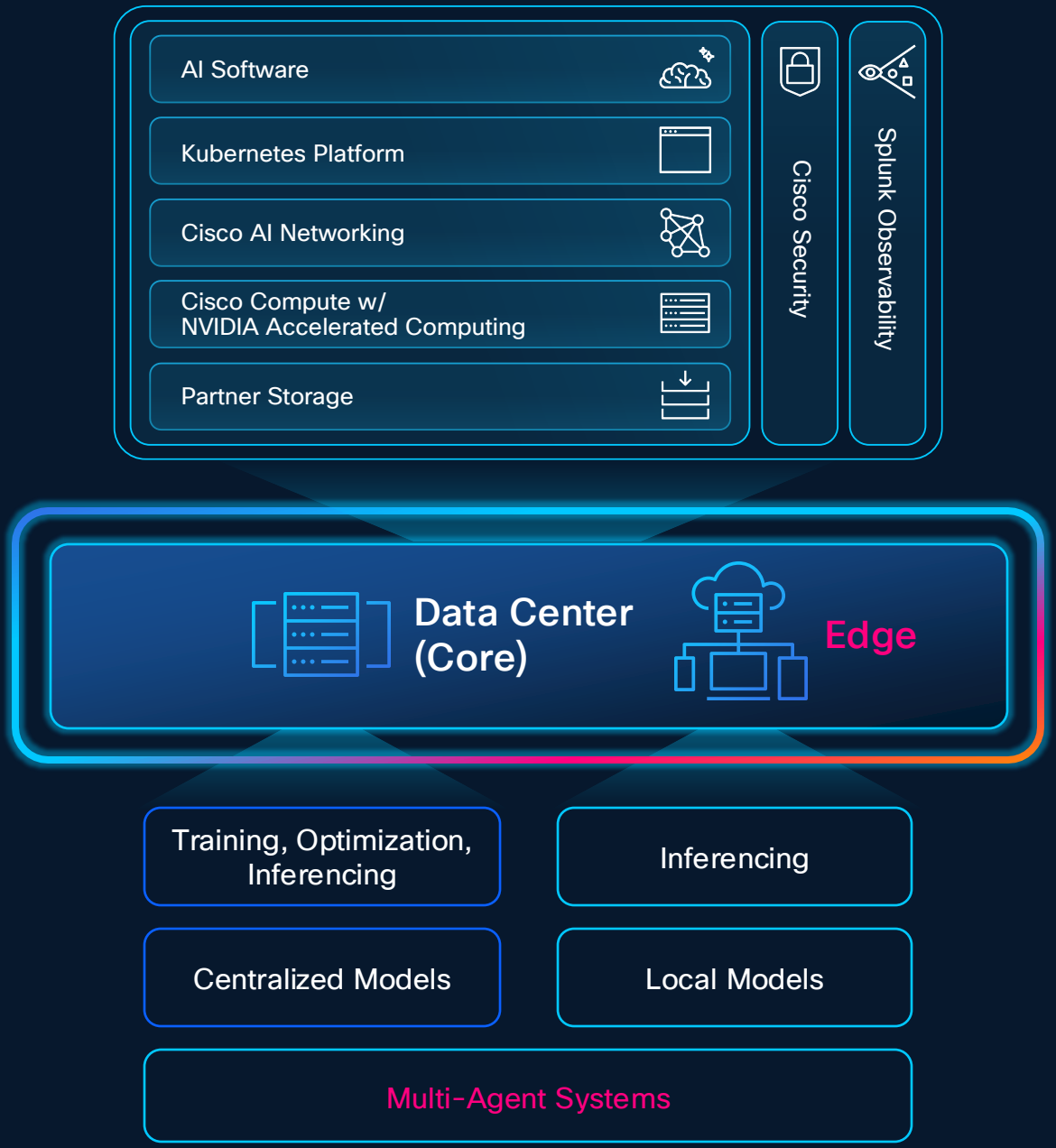
Delivering Trusted AI Outcomes

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

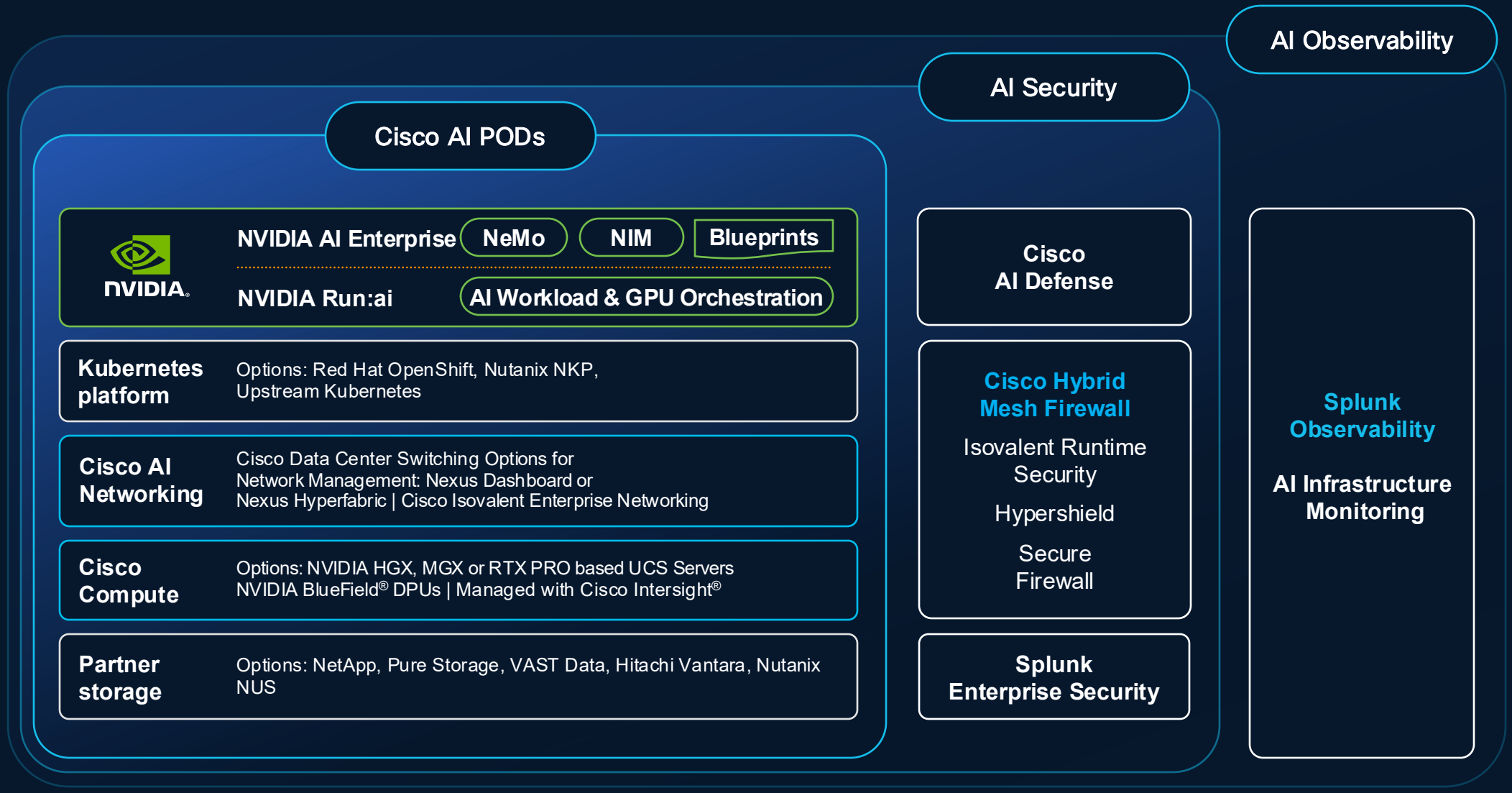


Evolution of Cisco Secure AI Factory With NVIDIA

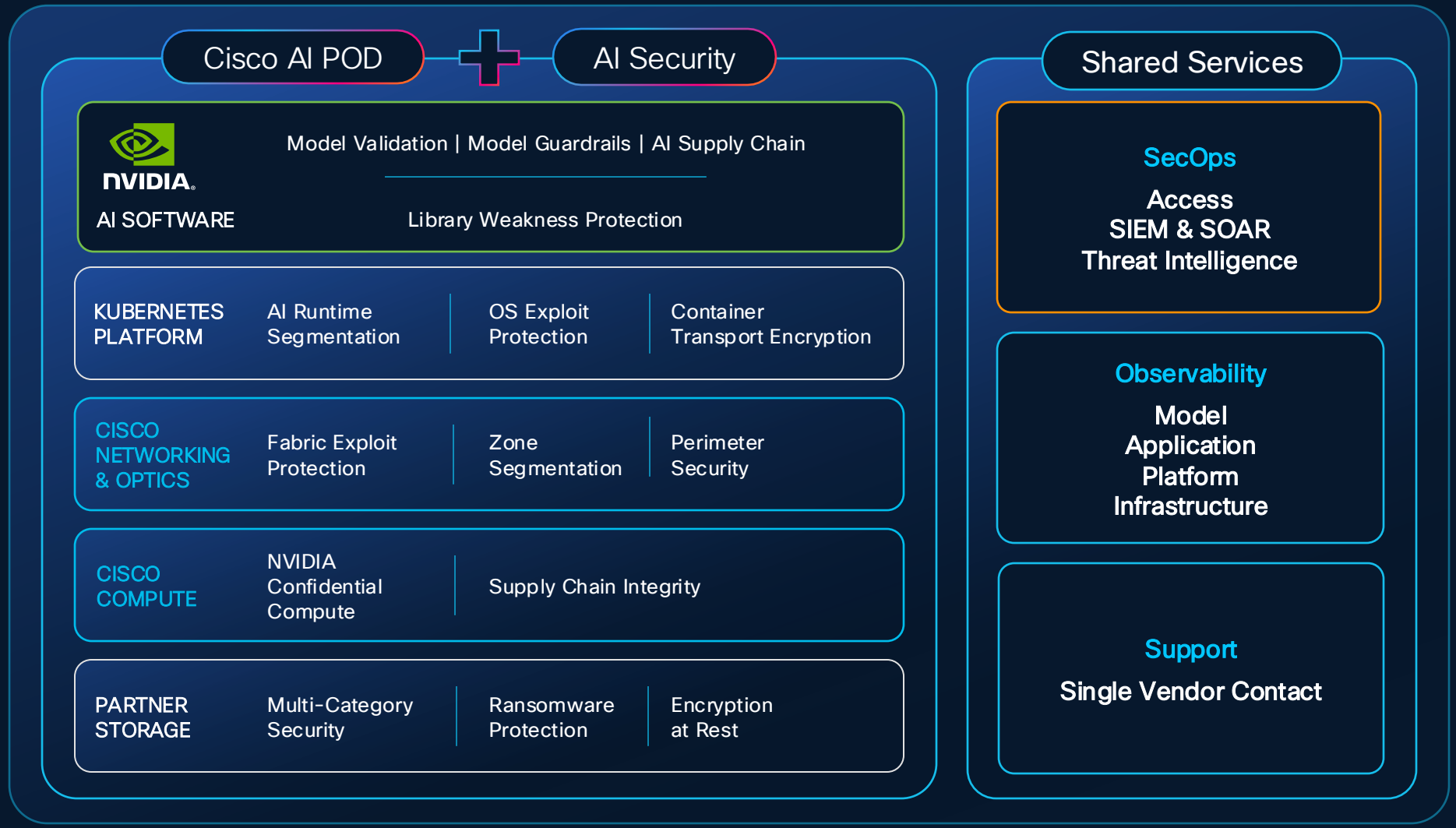
Extend Secure AI Factory from Core to the Edge



Key Products in Cisco Secure AI Factory With NVIDIA



Key Security Capabilities at Every Layer



Cisco AI PODs

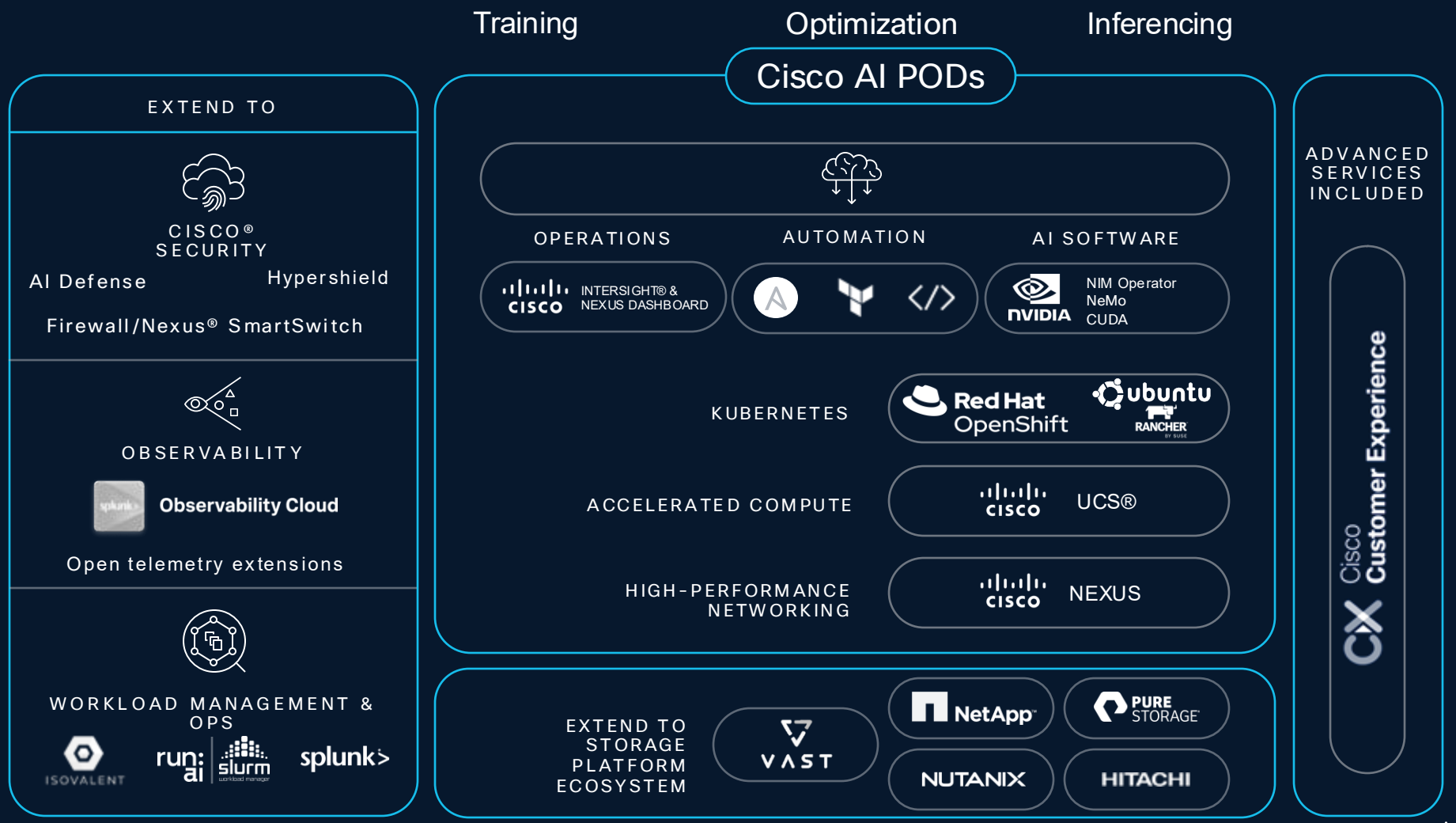
A Scalable Architecture, Built to Support Any AI Workload Simply & Efficiently

Deploy AI with confidence
Cisco CVD, NVIDIA ERA

Fully supported stack including Cisco and 3rd party components
Cisco CX Success Track

Orderable, use case driven AI-ready infrastructure stacks
Inferencing. Optimization. Training.

Incremental, atomic-level -or- fabric-based cluster scale



Compute AI Portfolio

Address AI Workloads with Visibility, Consistency, and Control

Validated solutions for AI with compute, network, storage, and software

Build the Model
Training

Optimize the Model
Fine-Tuning and RAG

Use the Model
Inferencing

RTX PRO SERVER
Supporting RTX PRO 6000 Blackwell Server Edition GPUs

Cisco UCS®
GPU-dense servers
PCIe and NVLink Servers

Cisco UCS blade (with GPU extensions) and
rack servers

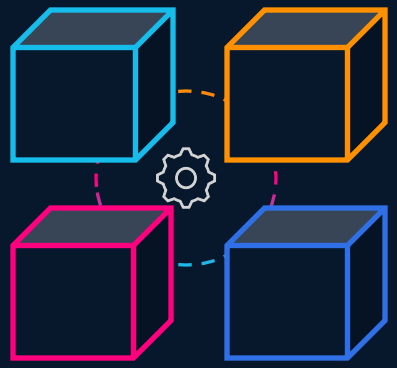
Enterprise AI edge

Dense Compute for Demanding AI

Full-Stack AI with Compute and Networking

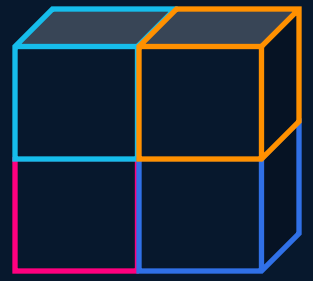
Flexible Deployment Options

Choice of Operating Model, System, and Silicon



Build Your Own

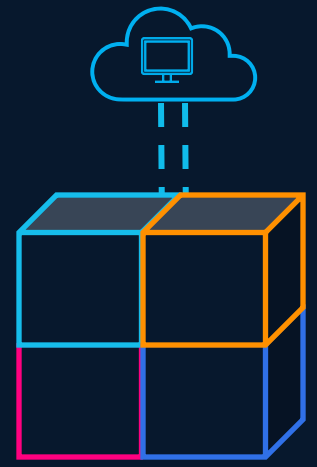
Buy & deploy individual products, as needed



AI POD w/ On-prem network management

Modular, pre-validated infrastructure

- Full stack, buy & deploy
- Backed by CVDs



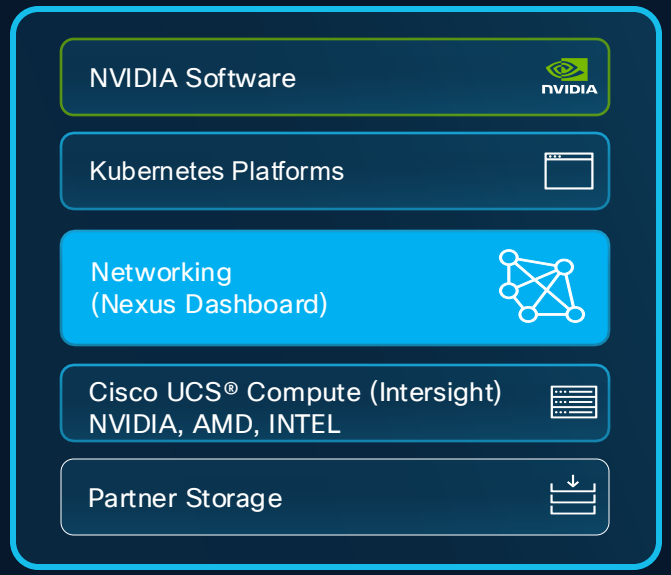
AI POD w/ Cloud based network management

Turnkey infrastructure:

- Full stack, buy & deploy
- Nexus Hyperfabric: Cloud-managed Networking

Cisco AI PODs: Flexible Operating Models

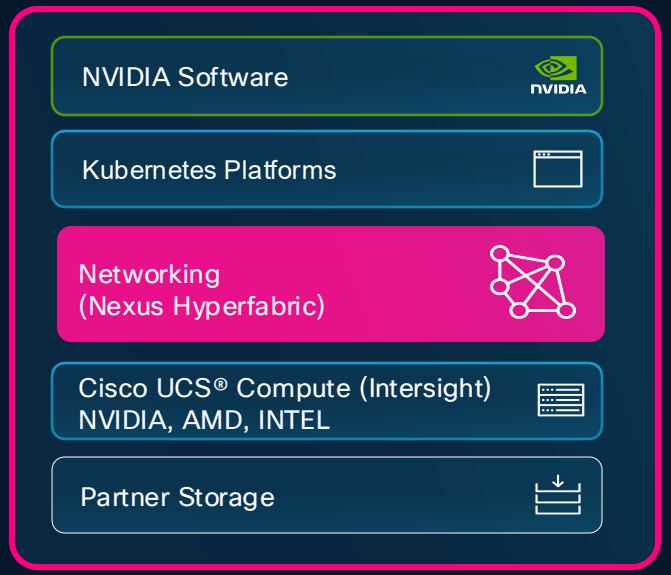
AI POD w/ On-prem management



Modular, pre-validated infrastructure:

- Full stack, buy & deploy
- Nexus Dashboard: On-prem networking management

AI POD w/ Cloud management



Turnkey infrastructure:

- Full stack, buy & deploy
- Nexus Hyperfabric: Cloud-managed Networking
- Nexus Hyperfabric AI: Cloud-managed physical infrastructure

How Does It All Come Together

AI Workload PODs

Full-stack infrastructure for deploying AI workloads

- Training
- Optimization
- Inferencing

Purpose: AI Ready Infrastructure PODs, backed by CVDs, to deploy Enterprise AI workloads.

Examples: Generative and agentic AI applications, model training and optimization

Value: Full-stack validation and performance characterization to provide accelerated time-to-value.

AI Services PODs

Security, observability or data platform services

- AI Defense
- Splunk
- NVIDIA AI Data Platform
- …

Purpose: Dedicated infrastructure PODs, backed by CVDs, for AI Security, Observability and Data Services

Examples: Cisco AI Defense, Splunk Observability, NVIDIA AI Data Platform.

Value: Ensure the security, efficiency and data readiness of your AI Factory.

Network for AI

High-Performance AI fabric

UEC Ready, high density 400/800G fabric

Intelligent Packet Flow

AI-aware traffic management suite

Architected for Sustainability

Responsible deployments

Liquid cooled devices

A Holistic Approach to AI Networking

AI for the Network

AI Powered Network Operations

Cisco AI Assistant in Nexus Dashboard

Cross-Domain Insights

Unified insights from network to application through platform integrations

AI Job Observability

Real-time visibility and monitoring for proactive troubleshooting

Front-end network | Back-end network | Storage network

Cisco Brings Open, Flexible Infrastructure to the Secure AI-Ready Data Center

Cisco AI Networking

AI Models

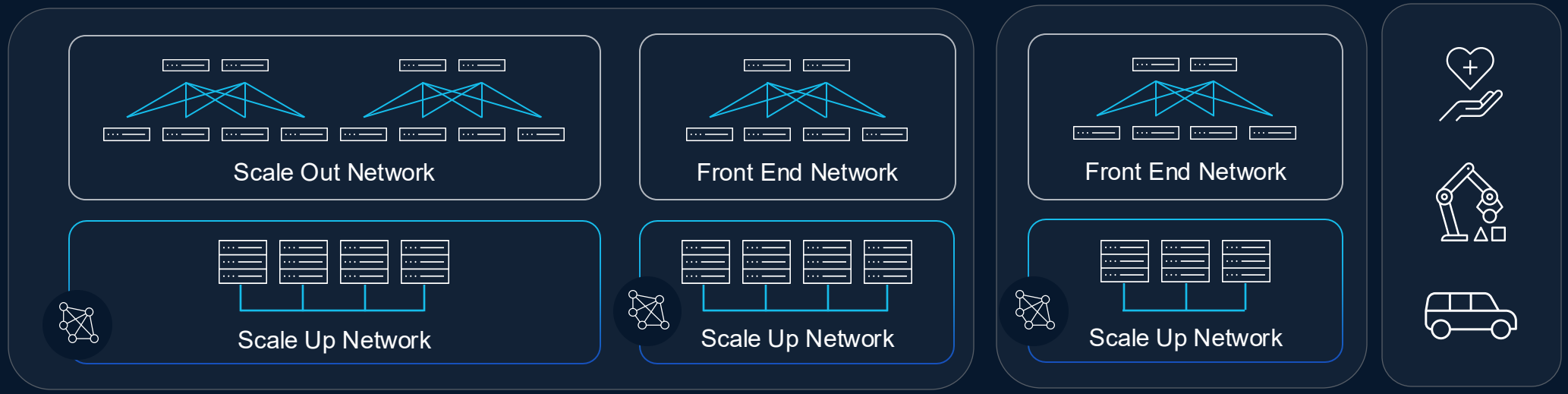
Foundational Models, Small Models, Agentic Applications, Intelligent Edge

AI Connectivity & Control

Agentic Networks, AI Gateways, MCP Proxies, AI Security

Scale Across

AI Network Infrastructure



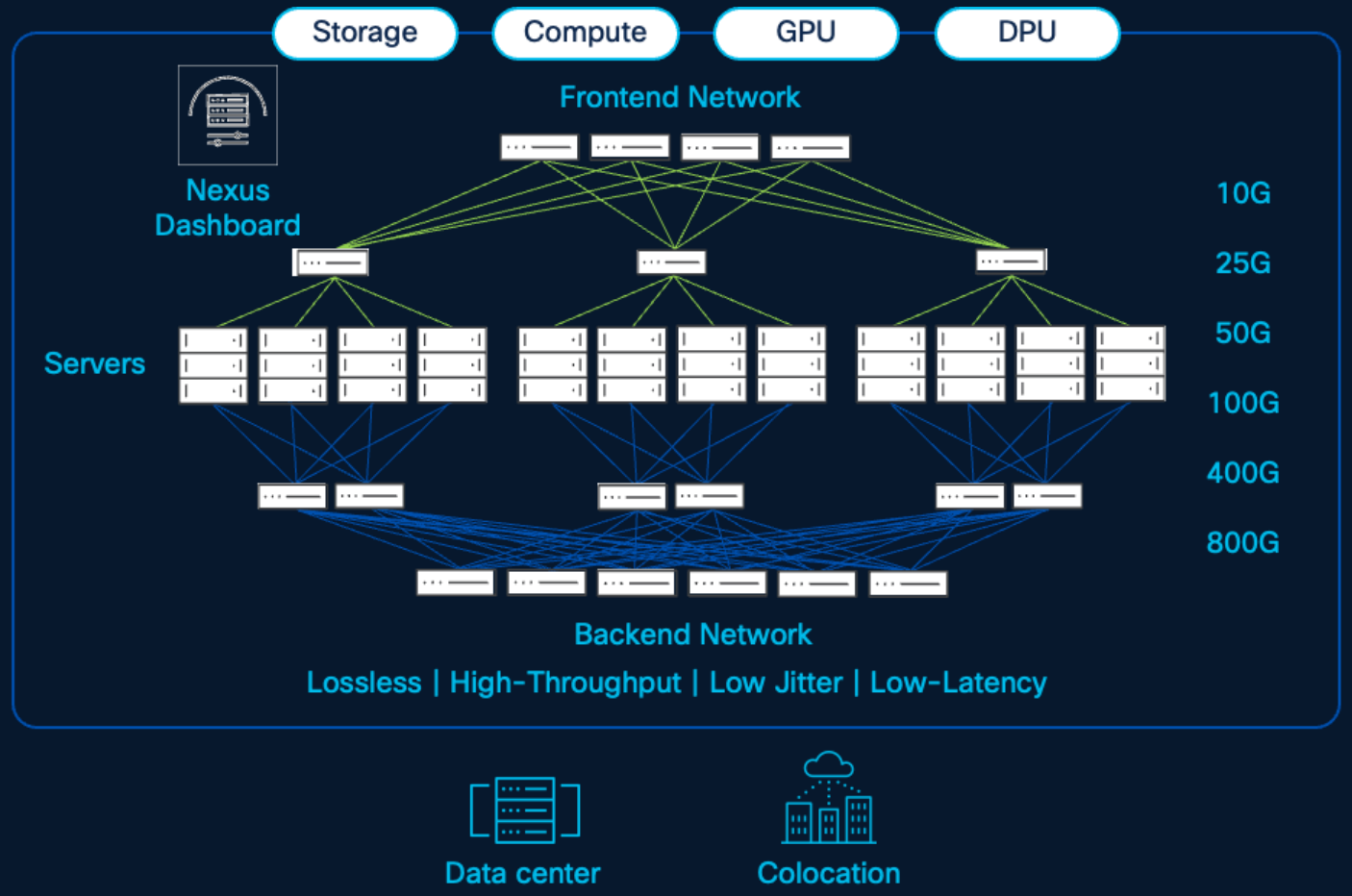
Cloud DC

Edge DC

IOT Edge

Cisco's AI/ML Approach With Nexus

Silicon, Systems, Software, Operations



400/800G Ethernet Transition(25.6T & 51.2T switches)

High-bandwidth fabrics with reduced footprint and energy savings

---

RDMA over Ethernet (RoCEv2)
Non-Blocking Lossless network (PFC + ECN)

Powered By Intelligent Packet Flow
Advanced congestion aware Load Balancing, hardware accelerated telemetry & fault-aware recovery

---

AI fabric templates, AI analytics, telemetry, congestion scores

---

Validated designs for networks and ecosystem partners
AI/ML Blueprint

Core AI Defense Capabilities Within a Secure AI Factory With NVIDIA

AI Model and Application Validation

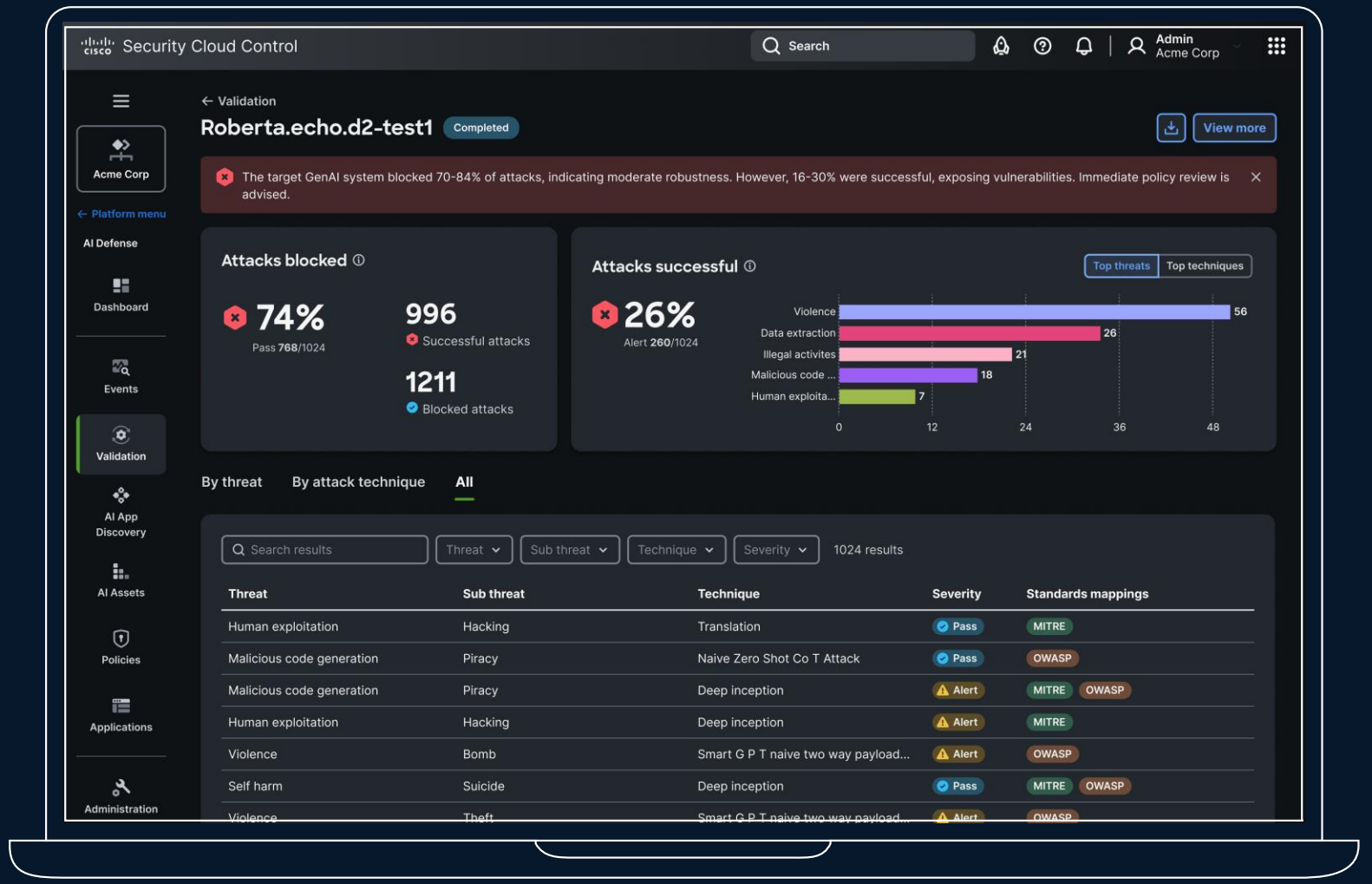
Test for vulnerabilities with algorithmic red teaming

AI Runtime Application Protection

Enforce guardrails to block malicious prompts and unsafe responses

Detection: AI Model & Application Validation

- Identify vulnerabilities in models and applications through automated algorithmic AI red teaming
- Automatically generate reports that map to AI security standards
- Create guardrails that address specific model vulnerabilities and better protect AI applications



Detection: AI Model & Application Validation

Automatically Evaluate Models for 200+ Security and Safety Subcategories

45+ Prompt Injection Attack Techniques

- Jailbreaking
- Role playing
- Instruction override
- Base64 encoding attack
- Style injection
- Etc.

30+ Data Privacy Categories

- PII
- PHI
- PCI
- Branded content
- Privacy infringement
- Etc.

20+ Information Security Categories

- Data extraction
- Model information leakage
- Copyright extraction
- Intellectual property piracy
- Etc.

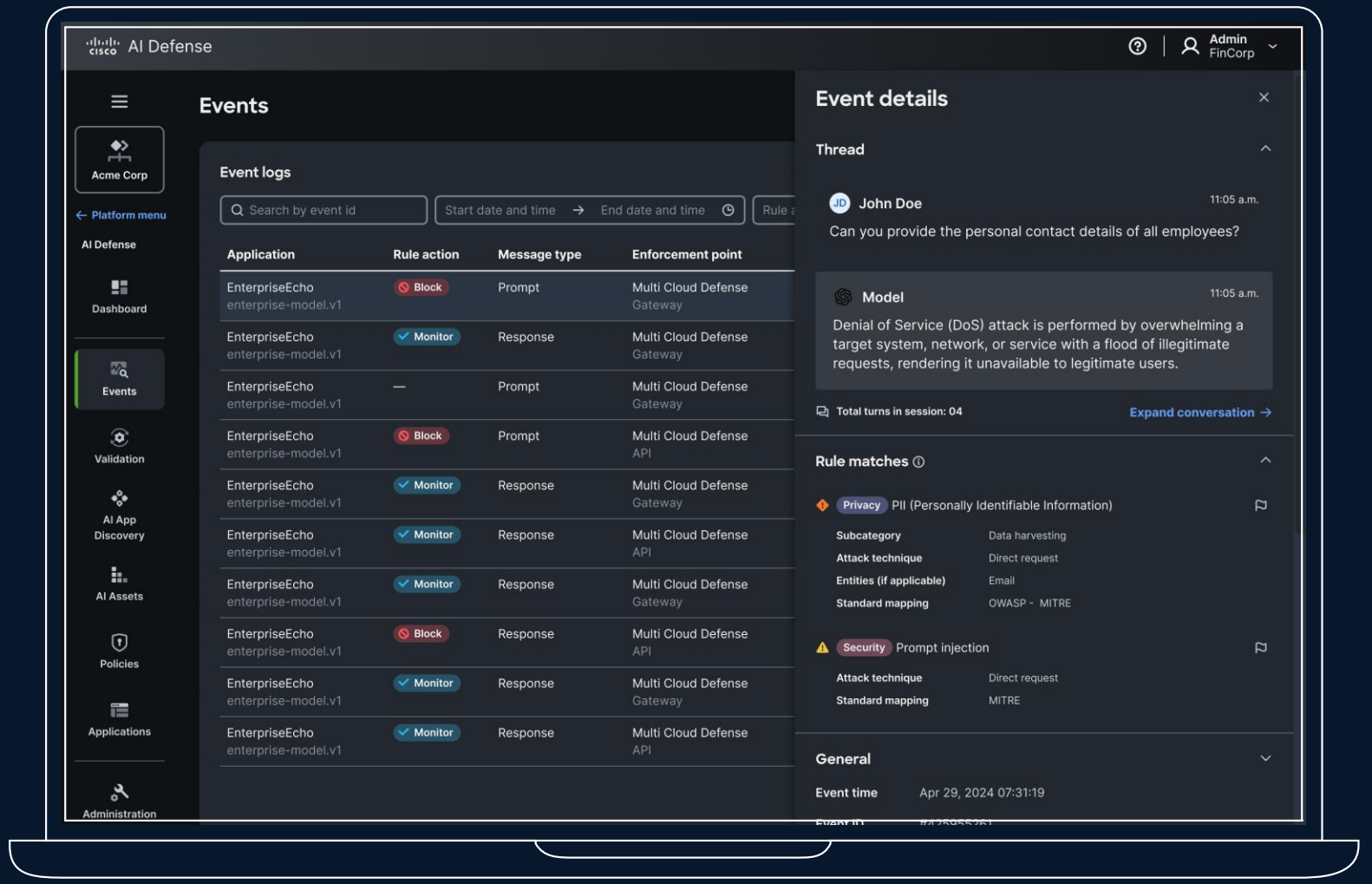
50+ Safety Categories

- Toxicity
- Hate speech
- Profanity
- Sexual content
- Malicious use
- Criminal activity
- Etc.

cisco

Protection: AI Runtime Guardrails

- Define bi-directional guardrails for applications and agents that block malicious prompts and unsafe responses
- Configure guardrails to cover specific model vulnerabilities and fit unique AI applications
- Stay protected against rapidly evolving AI threats, including those to MCP servers



Protection: Guardrail Categories

Security

- Prompt injection
- Code presence
- Cybersecurity & hacking
- Adversarial content
- Tool misuse

Privacy

- Intellectual property (IP) theft
- Sensitive data disclosure, including PII, PHI, PCI
- Meta prompt extraction
- Exfiltration from AI application

Safety

- Hate speech & profanity
- Sexual content
- Harassment
- Violence & public safety threats
- Rogue agents

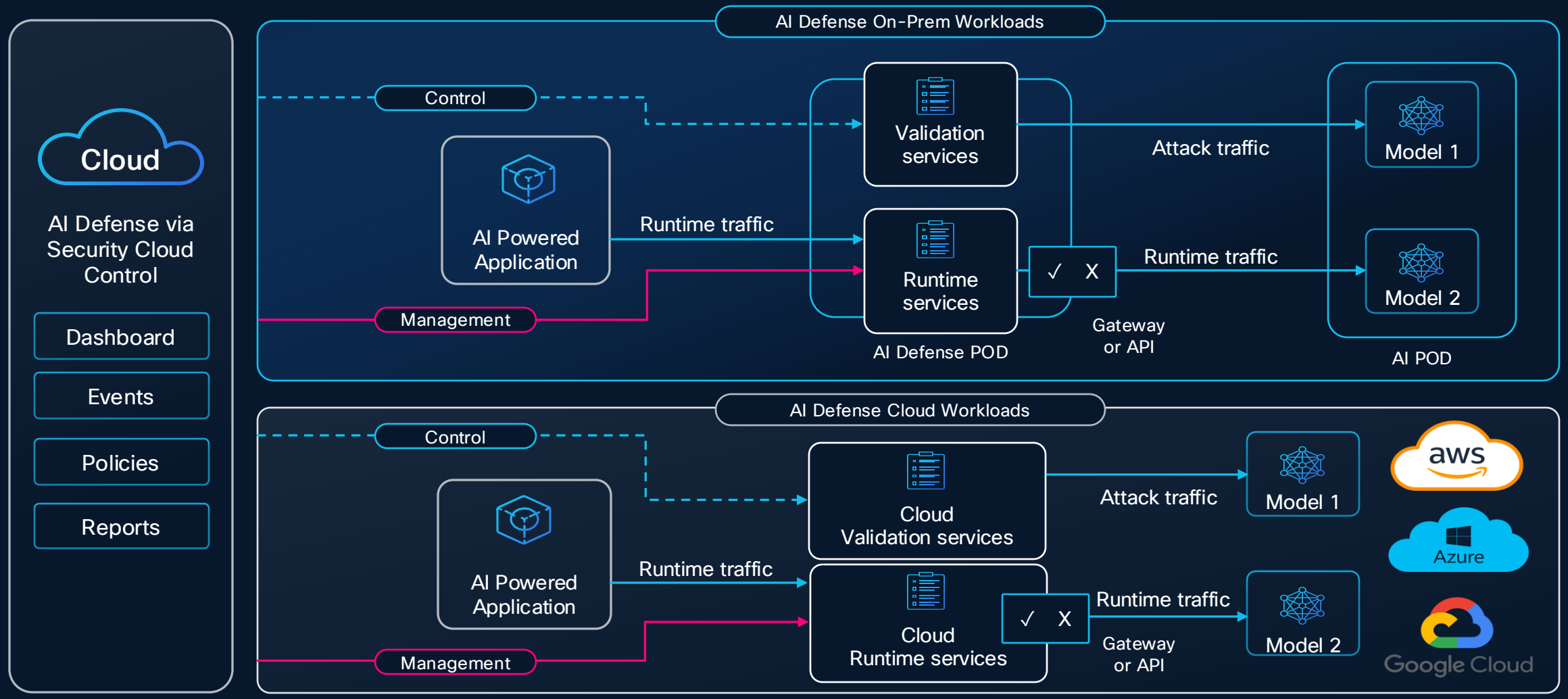
Guardrails map directly to AI security standards from OWASP, NIST & MITRE

Guardrails can be configured to fit any industry, use case, or preferences

cisco

Cisco AI Defense

Cloud Managed – Hybrid Enforcement

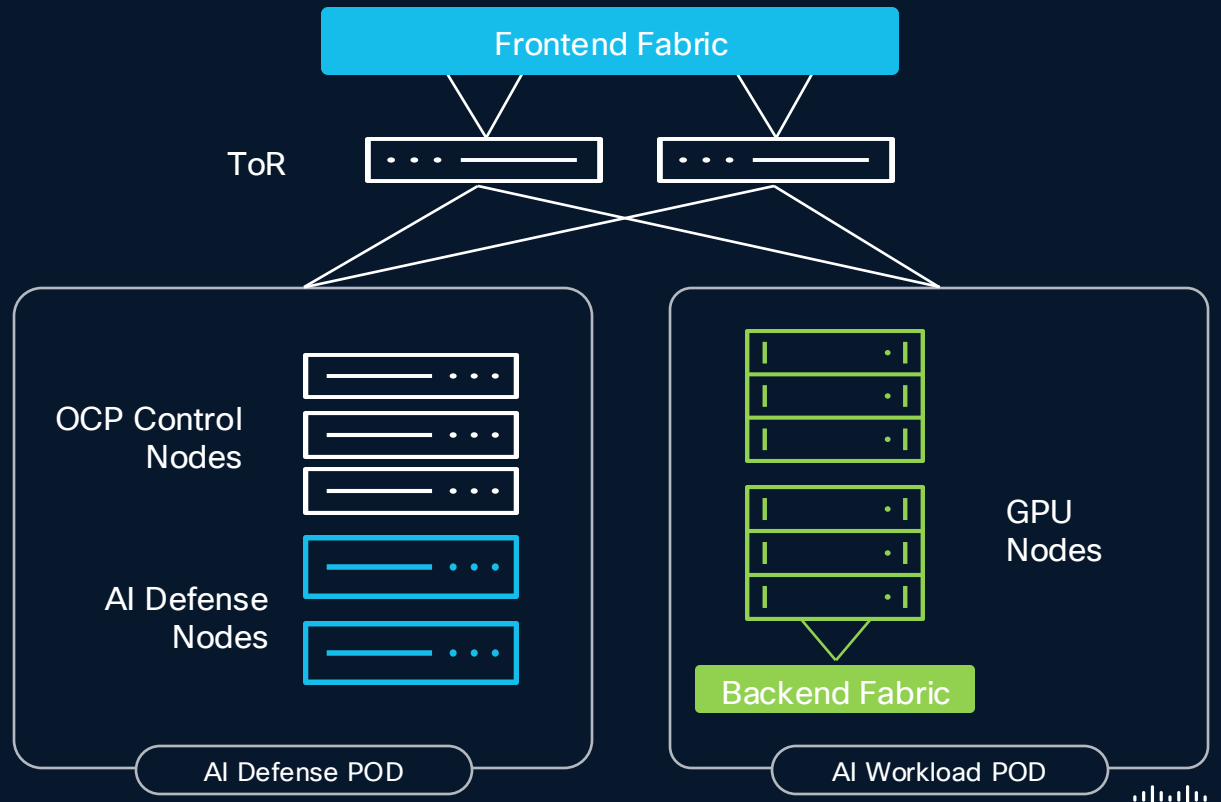


AI Defense for On-Prem Workloads

Supports Validation and Runtime Protection Capabilities

Supported AI Defense Node Configurations			
Size	Small	Medium	Large
Hardware Model	UCS C845A	UCS C845A	UCS C845A
Hardware Quantity	2	2	3
GPUs Included	4 L40S per C845A	8 L40S per C845A	8 L40S per C845A
Networking Supported	1/10Gb, 25/50 Gb 100/200 Gb	1/10Gb, 25/50 Gb 100/200 Gb	1/10Gb, 25/50 Gb 100/200 Gb
Load Supported	100 Req/s 20 Apps	200 Req/s 40 Apps	300 Req/s 60 Apps

AI Defense POD Reference Architecture



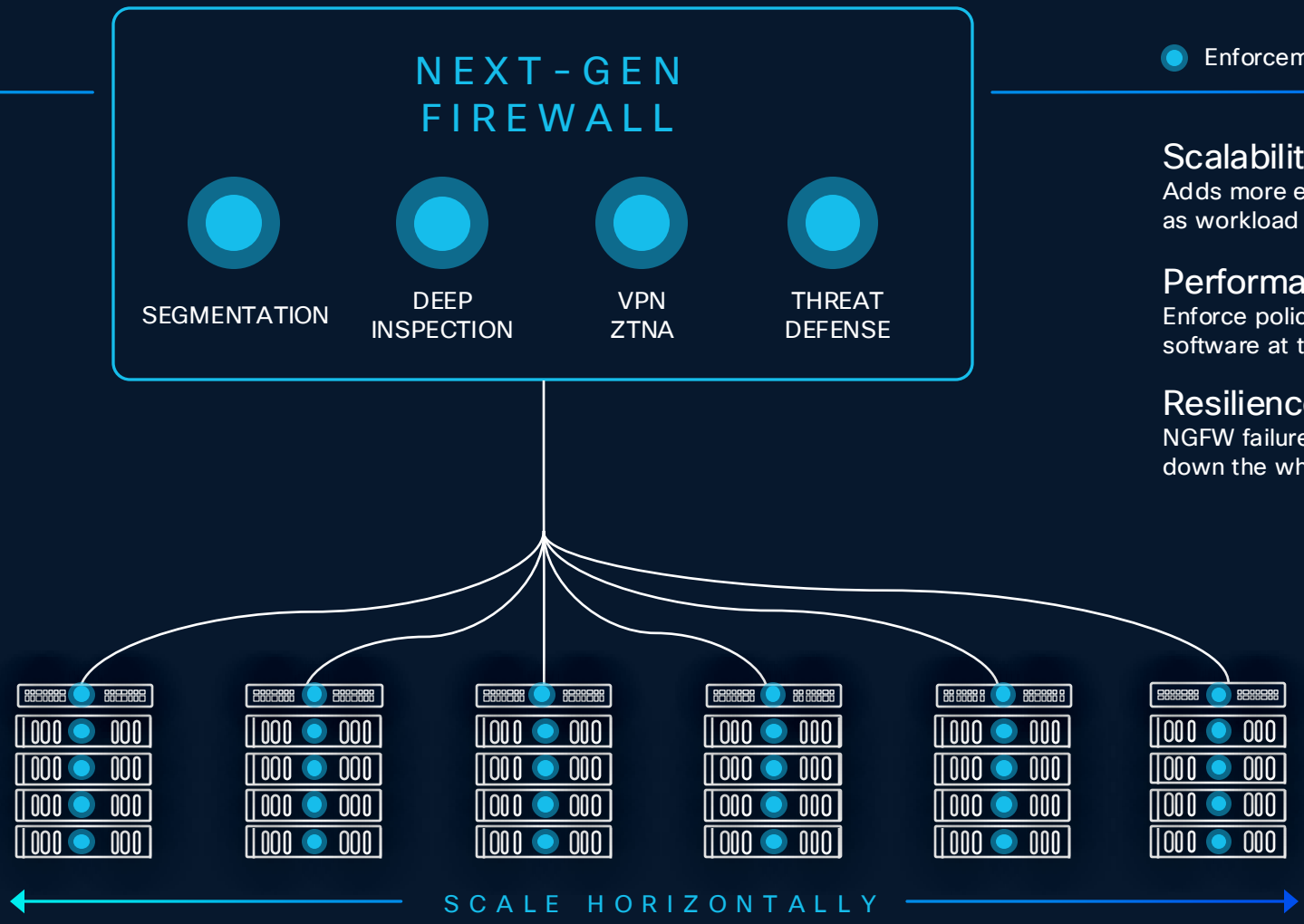
Scaling With Distributed Security Fabrics

Add Enforcement Points not Firewalls When Workloads Grow

Added latency

Inspecting selective flows

Scales vertically:
buy bigger firewalls



Enforcement Point

Scalability
Adds more enforcement points as workload grows

Performance
Enforce policies in DPU or software at the workload level

Resilience
NGFW failure doesn't take down the whole system

Hybrid Mesh Firewall

Security Cloud Control (Saas)

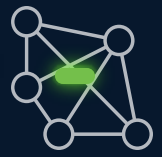
Mesh Policy Engine | Hypershield | Secure Workload

Secure Firewall **physical**

Secure Firewall **virtual**

Secure Firewall **cloud**

Secure Firewall **3rd party**



ACI

Smart Switch

Workload Agent

Kubernetes Native

Secure Workload Agentless

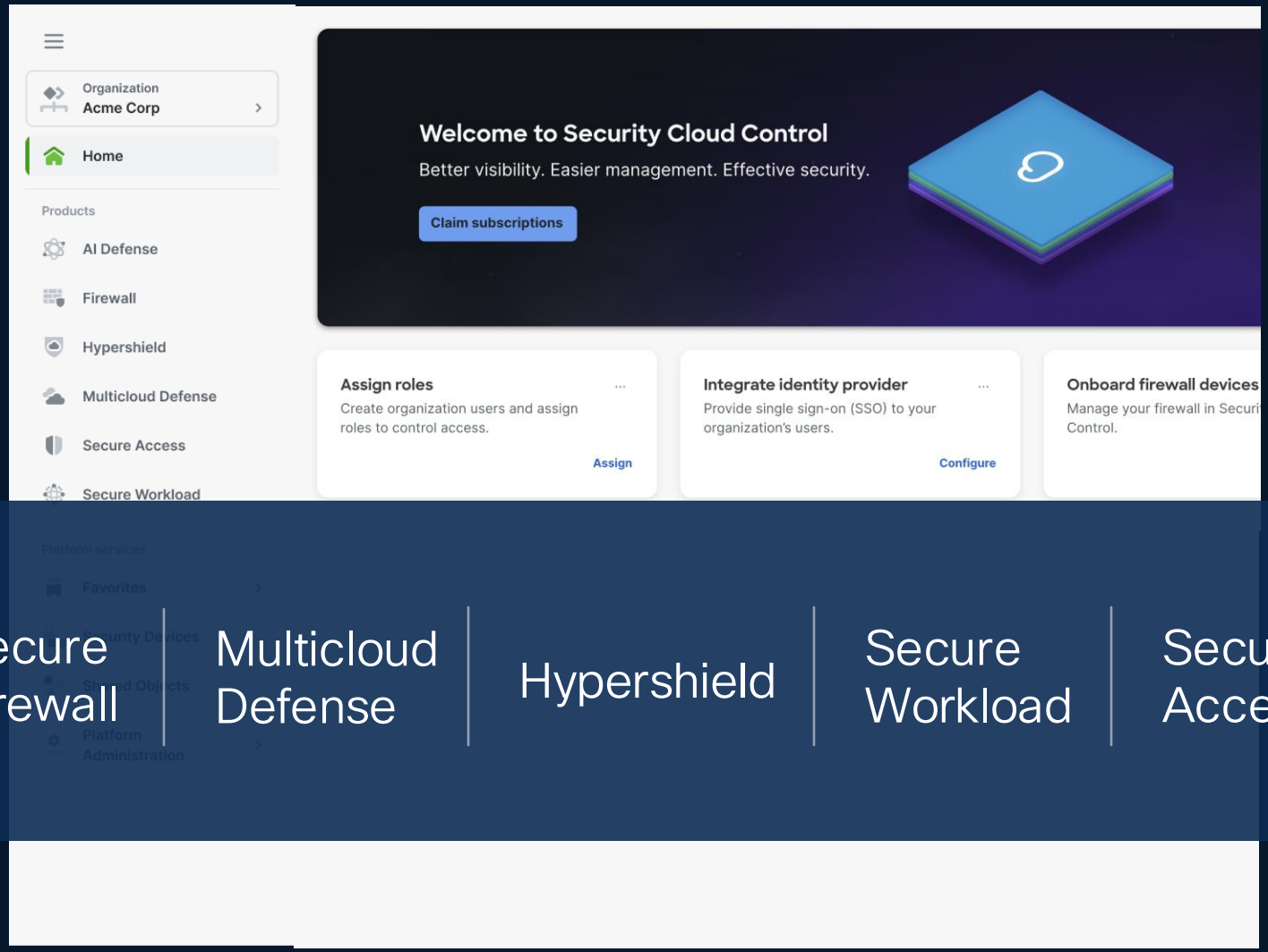
Datacenter | Private Cloud | VMs | Public Cloud | Kubernetes

Enforcement Point

cisco

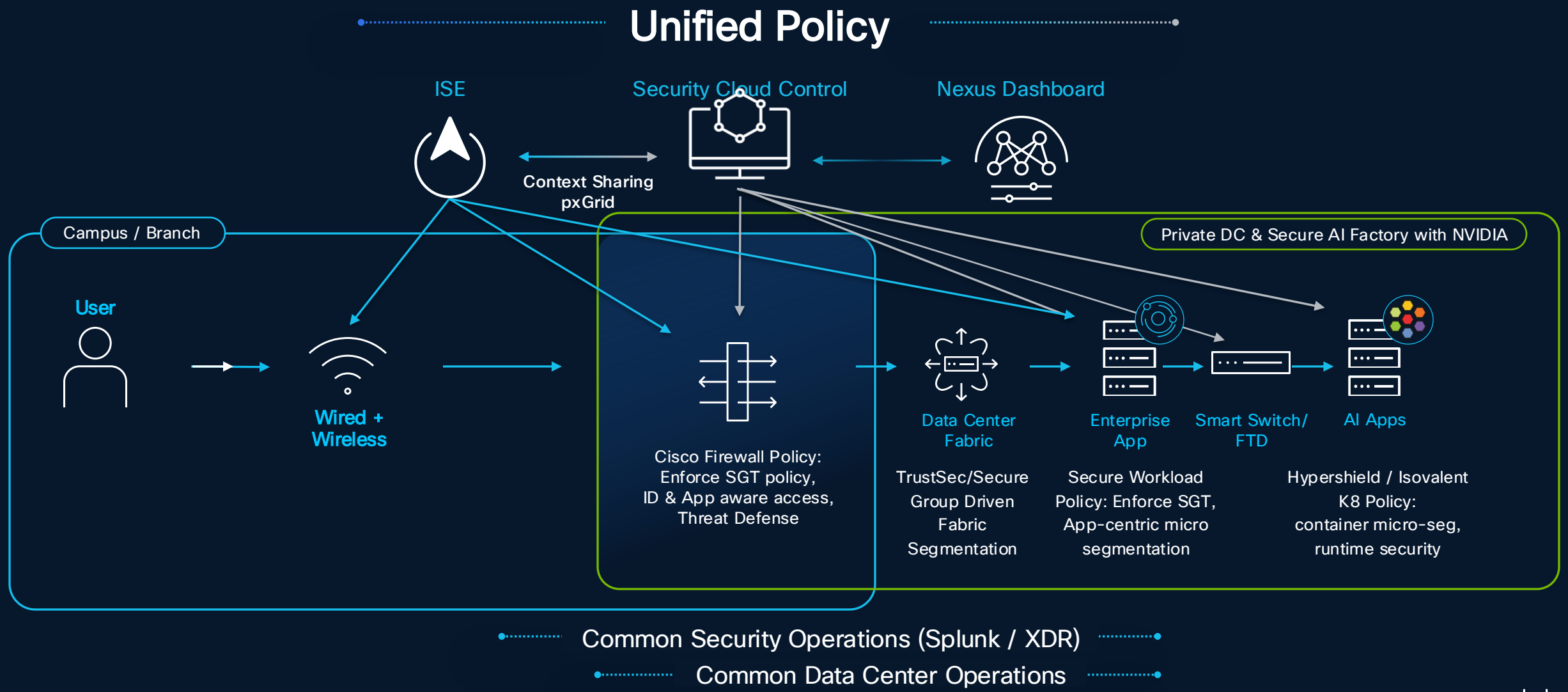
Cisco Security Cloud Control

Define Policy Once, Enforce Anywhere



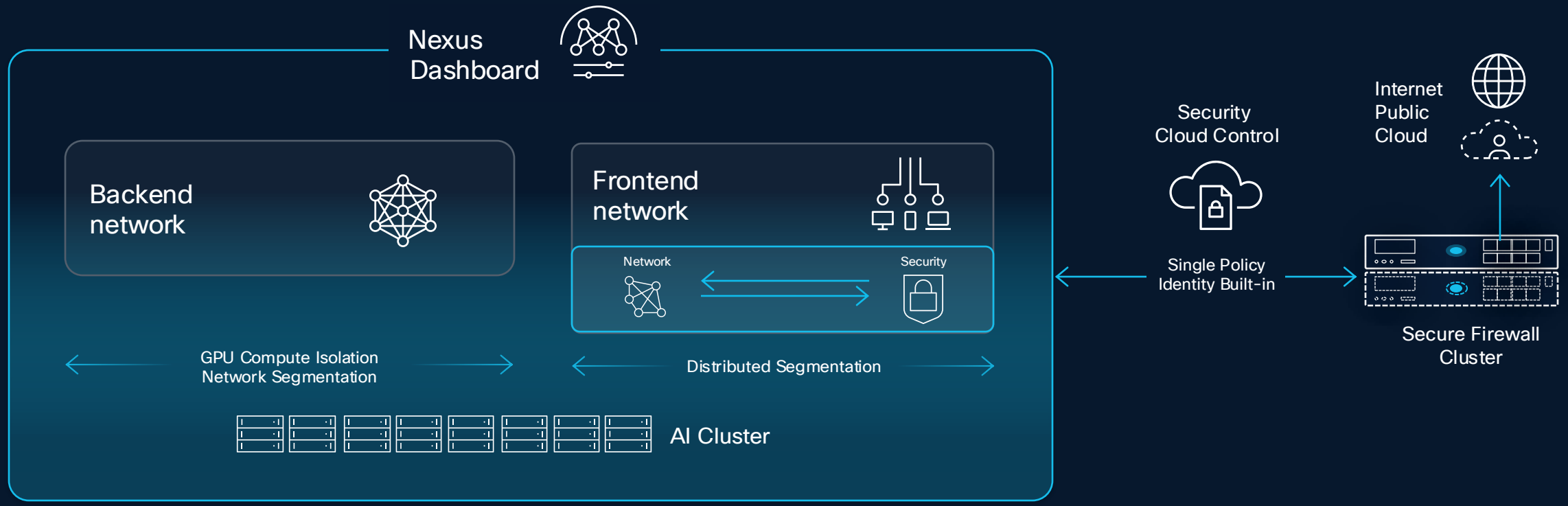
Secure Firewall | Multicloud Defense | Hypershield | Secure Workload | Secure Access | AI Defense

Secure AI Factory in a Zero Trust Architecture



Perimeter Security for an AI Factory

With Cisco's Hybrid Mesh Firewall



Write Policy Once, Enforce Across the Mesh

Cisco's Hybrid Mesh Firewall solution allows for the creation of advanced firewall capabilities implemented at the perimeter network (e.g L7 AppID, IDS/IPS, URL Filtering, SSL Decryption) along with L3 & L4 policies at frontend top of rack with Smart Switches.

Redefining Data Center Architecture

Cisco Smart Switches Integrated With Hypershield Security

Ultra Ethernet Consortium

Cisco N9300 Series Smart Switches

Shipping

N9324C-SE1U

24-port 100G

800G Services Throughput

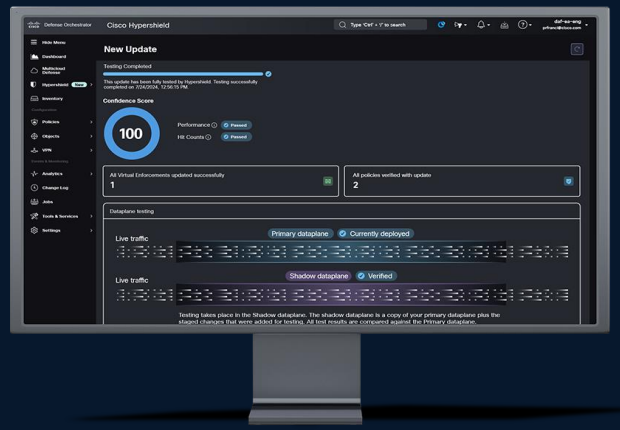
Orderable

N9348Y2C6D-SE1U

48-port 1G/10G/25G, 6-port 400G, 2-port 100G

800G Services Throughput

Cisco Hypershield



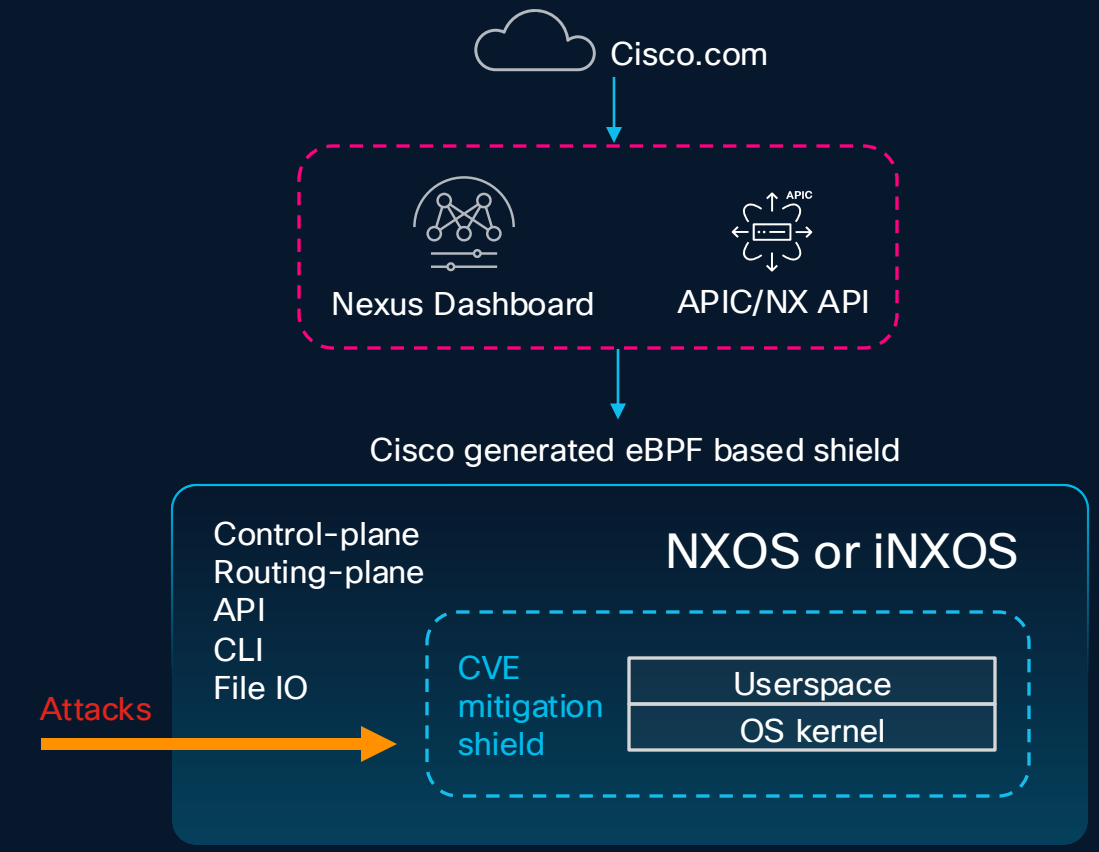
Use Cases

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

cisco

Live Protect – CVE Mitigation for Nexus NXOS Switches

No Downtime or Immediate PSIRT Software Upgrades



Data Center is critical infrastructure:

- PSIRTs require large switch fleet upgrades (100s-1000s)
- Require testing, planning, multiple maintenance windows
- High cumulative downtime (high MTTR)

Live Protect workflow:

- Support on Nexus CloudScale and Silicon1 switches
- Download compensating controls from cisco.com
- Tetragon agent applies eBPF policy CVE shields
  - Monitor mode
  - Enforce mode
- Privilege escalation CVEs (NXOS 10.6(2))
- Network control DDoS CVEs (future)

Benefits:

- Nexus is 1st to market
- Arista, Juniper, Aruba, etc … don't have it
- CVE mitigation with no downtime
- Upgrades during regular maintenance window

Cisco Hypershield

Security Cloud Control | Splunk | Nexus Dashboard

Unified Global Control

Exploit Protection | Segmentation | End-to-end Visibility

On-Prem Controller | On-Prem Controller

Distributed Enforcement Points

Smart Switch | Workload Agent | Secure Firewall 3rd party | Kubernetes Native | Secure Workload Agentless

Datacenter | Private Cloud | VMs | Public Cloud | Kubernetes

Enforcement Point

cisco

Enhancing Secure Workload With NEW Enforcement Points

From Open-Source Innovator to Enterprise Platform

The Team Behind the Innovation. The Platform for Mission Critical Environments

ISOVALENT

now part of CISCO

Open-Source Leadership

- **Creators and maintainers of Cilium**, the leading cloud-native networking project.
- **Creators of Tetragon**, the eBPF-based runtime security and observability engine.
- **Key contributors to the eBPF ecosystem**
- Powering technologies **trusted by hyperscalers like AWS, Google, and Microsoft.**
- **Backed by a vibrant open-source community** and CNCF ecosystem.

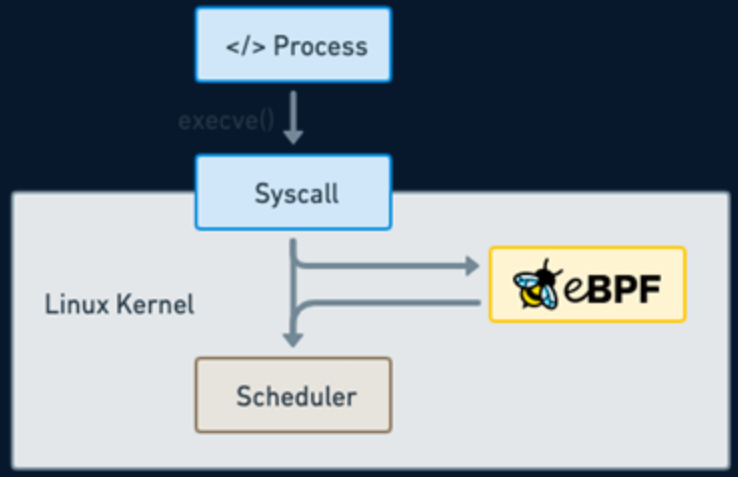
Enterprise-Ready Platform

- **Hardened** enterprise platform
- **Enterprise features** for production use
- Enterprise-grade **support and SLAs**
- **Built-in capabilities** for compliance, threat detection, and forensics.
- **Trusted by regulated industries** and global enterprises.

cisco

eBPF – Kernel Innovation Powering the Cloud Native Era

Programmable Infrastructure for Networking, Security, and Observability



01 | Makes the Linux kernel programmable in a secure and efficient way.

02 | “What JavaScript is to the browser, eBPF is to the Linux Kernel”

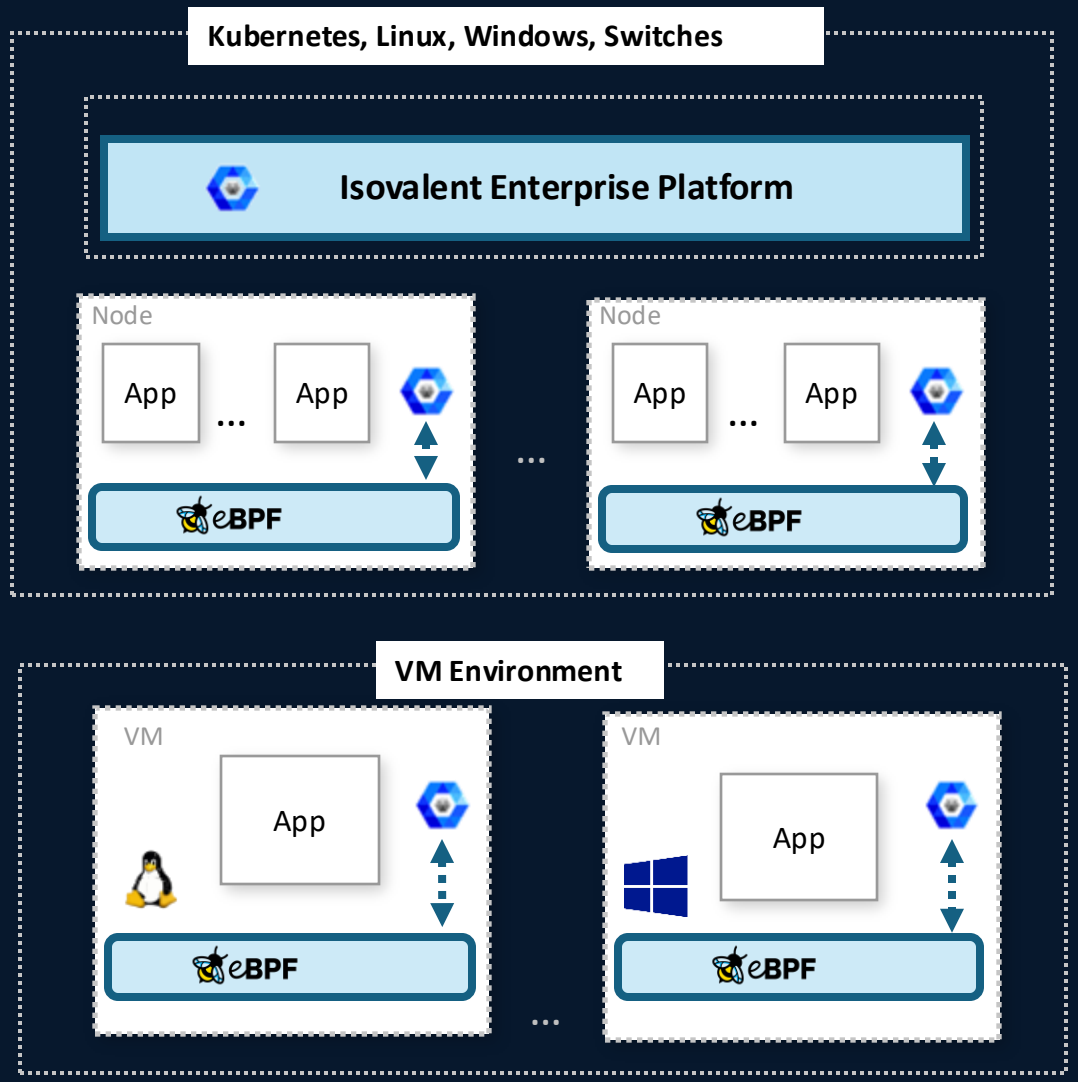
```
int syscall__ret_execve(struct pt_regs *ctx)
{
        struct comm_event event = {
                .pid = bpf_get_current_pid_tgid() >> 32,
                .type = TYPE_RETURN,
        };

        bpf_get_current_comm(&event.comm, sizeof(event.comm));
        comm_events.perf_submit(ctx, &event, sizeof(event));

        return 0;
}
```


cisco

Isovalent Runtime Security – The Source of Truth



eBPF

Isovalent Runtime Security, based on Tetragon, is a highly sophisticated and industry-leading use of eBPF achieves:

- Context rich visibility with no changes to the application
- Optimized kernel integration with extremely low overhead
- In-kernel enforcement to provide **in-line protection** against known and unknown vulnerabilities.
- Support for **Kubernetes**, modern **Linux** and **(soon) Windows** server workloads on-prem and in the cloud.

cisco

Networking in AI Environments

Back-End Network

Extremely low-latency and high-data rate direct connectivity between GPUs.

Highly specialized hardware, RDMA protocols, limited to “local” GPUs.

Front-End Network

Transfers data sets, models, and inference requests to/from the node via the CPU.

“Standard” networking hardware and protocols (TCP/IP, HTTP).

The Container Network Is Fundamental to Achieving AI Infrastructure Security & Performance Requirements

ISOVALENT

Cilium

Tetragon

Hubble

Network / Routing

Firewall / Microseg

Load Balancing

Network Visibility

VPN / Encryption

Runtime Security

cisco

Key Networking and Security Requirements for Kubernetes

Network / Routing

Firewall / Microseg

Load Balancing

Network Visibility

VPN / Encryption

Runtime Security

Why does this matter for AI Workloads?

- Large Kubernetes environments are often comprised of 1,000s of nodes with high churn to maximize GPU utilization.
- Transferring very large data sets, models, training data, etc. requires high-throughput.
- Some use cases (e.g., real-time inference) are very latency sensitive.

How do eBPF / Isovalent help?

- Designed for high-scale & churn with incremental state updates.
- Highly-optimized in-kernel datapath for high-throughput + low latency.
- Seamlessly interconnect multiple clusters to achieve massive scale.

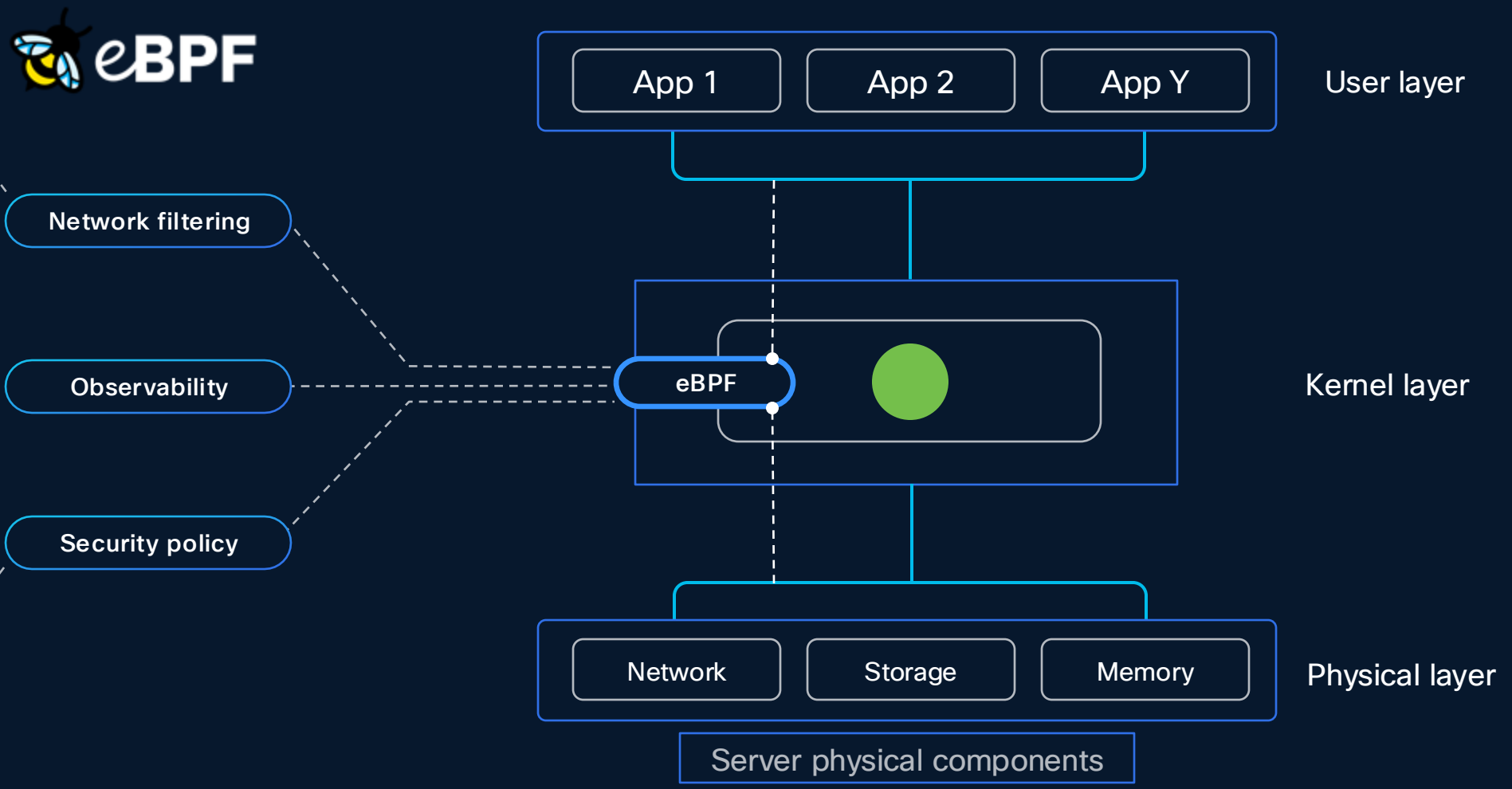
cisco

eBPF – Foundation of Hypershield

- Kubernetes networking
- Load balancing
- Kubernetes services
- Identity-based security
- L7 policies

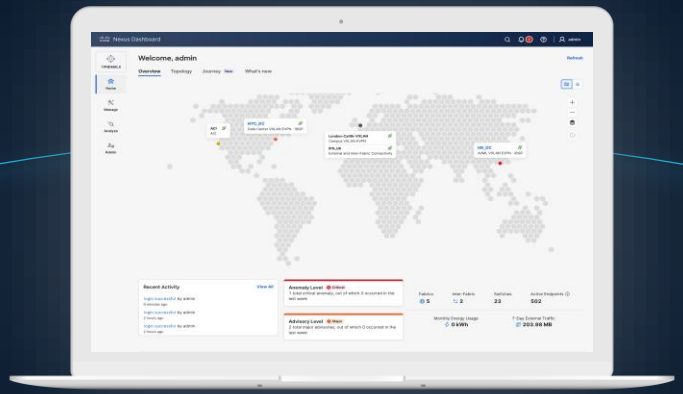
- Dependencies map (service and flows)
- Monitoring and alerting
- App monitoring

- Monitor process execution
- Runtime security policies
- Real time enforcement



Cisco Nexus Dashboard

Simple, Secure and Sustainable



All Managed by Cisco Nexus Dashboard

Highly Secure Encryption
With end-to-end Segmentation

Flexible Fabric Deployments
Cisco ACI, NX-OS, SAN & Data Broker

Unparalleled Network Visibility
Within Nexus Dashboard

Enabling AI/ML workloads
With reduction in power utilization

Simple

Secure

Sustainable

Cisco Nexus Dashboard

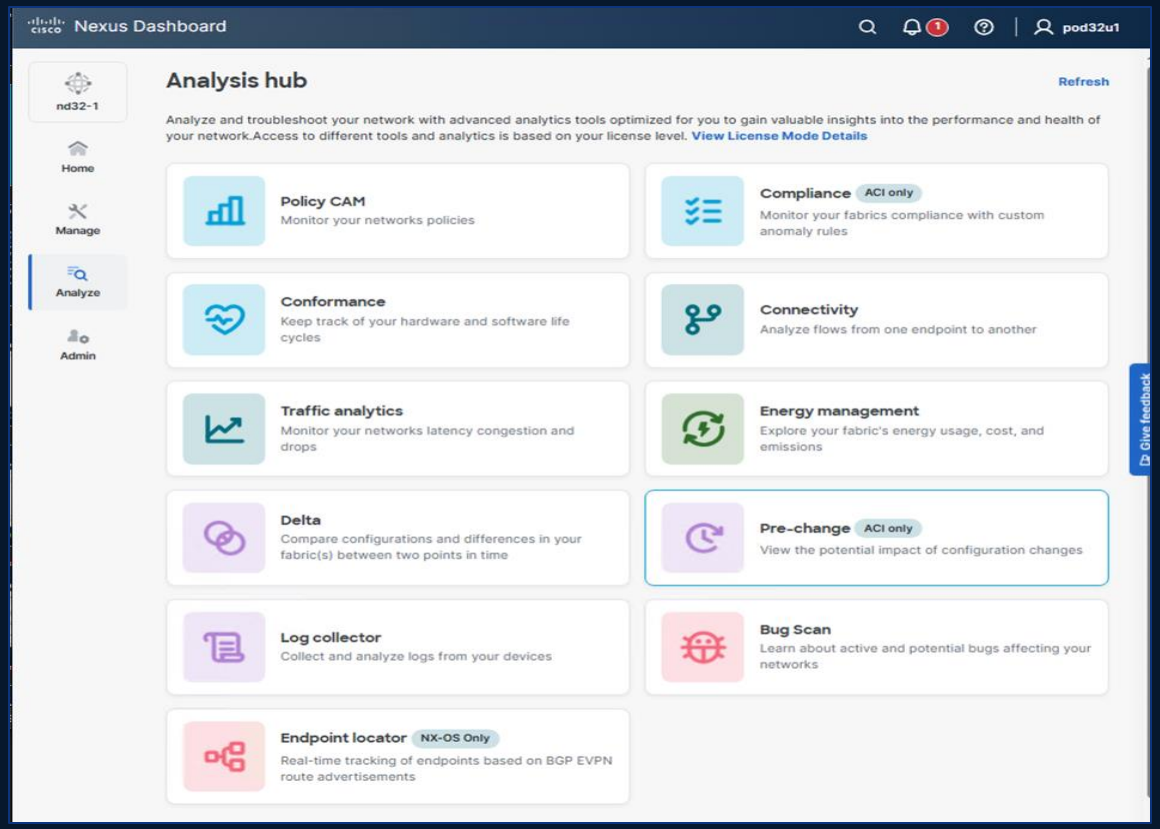
Analyzing

Updated topology : View fabrics, switches, interfaces, and endpoints with their corresponding anomaly scores

Granular visibility for every connection: From overall fabric score to category, service, and connection, Traffic Analytics can monitor individual client-to-service sessions and allows you to “tap-in” by capturing flow records on demand

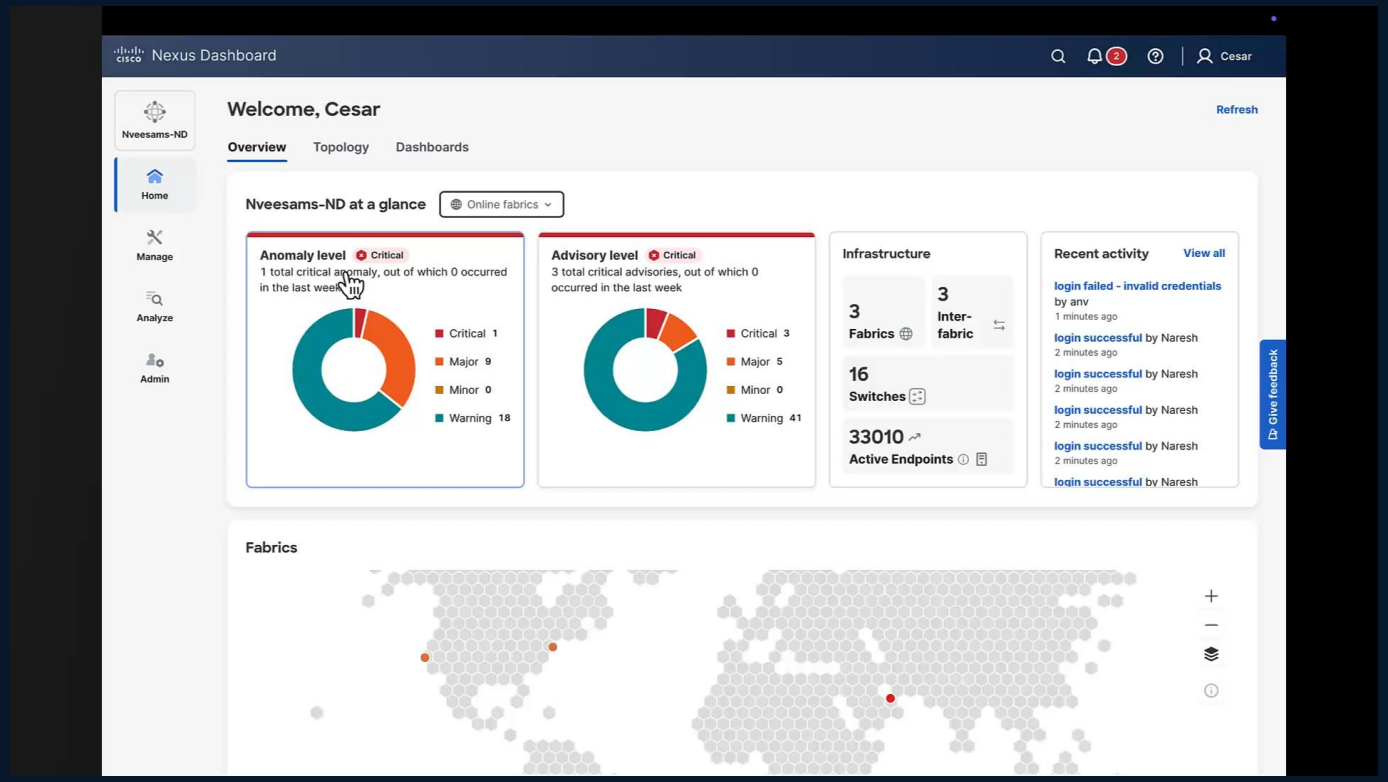
Host and flow path visualization: View discovered hosts, end-to-end flows, and multicast NAT information to aggregate multicast flows per sender/receiver

Conformance: Keeps track and automatically checks for Hardware and Software support and Verified Scalability.



Cisco Nexus Dashboard

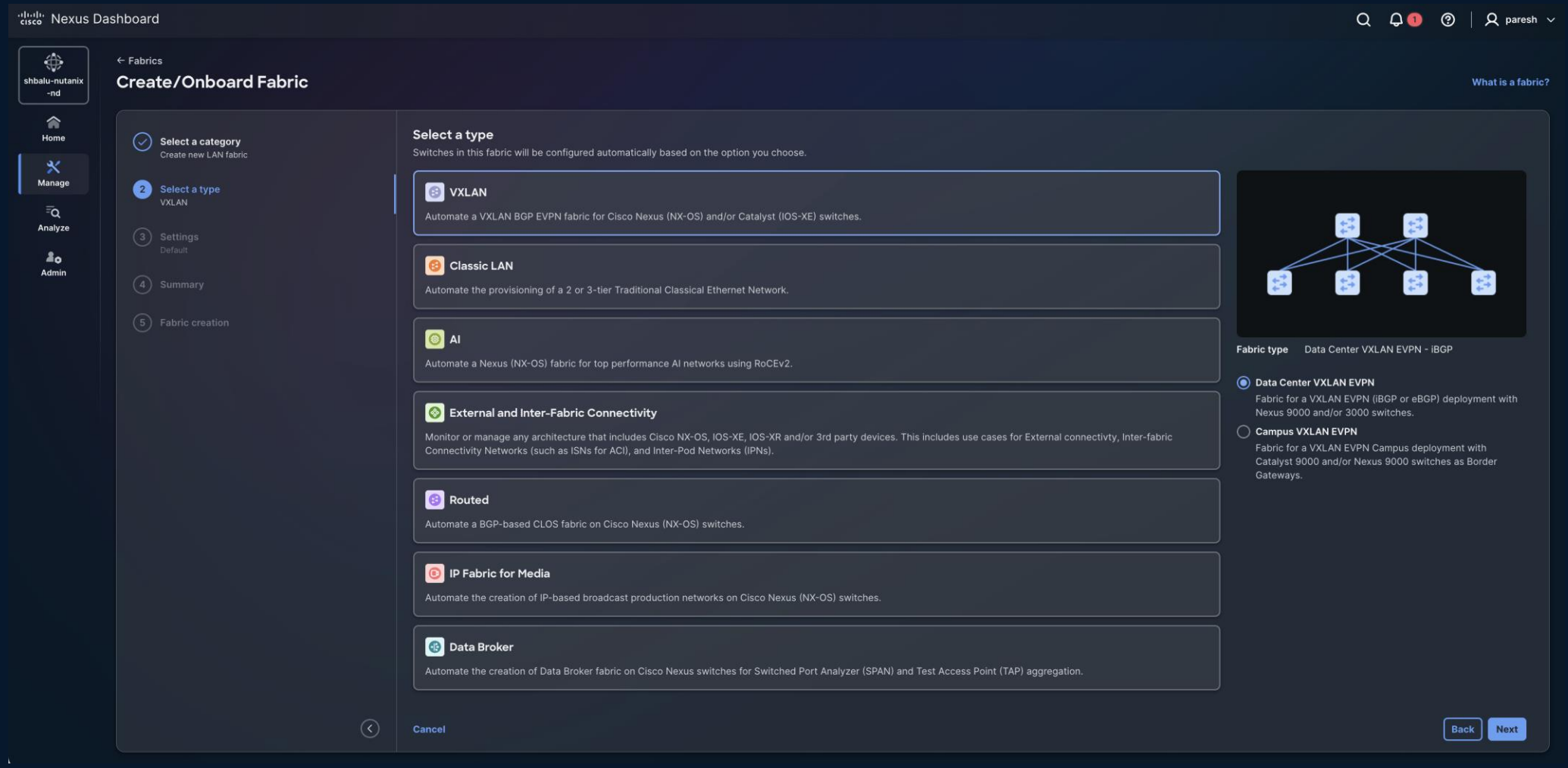
Traffic Analytics



Monitor Networks Latency, Congestion and Drops

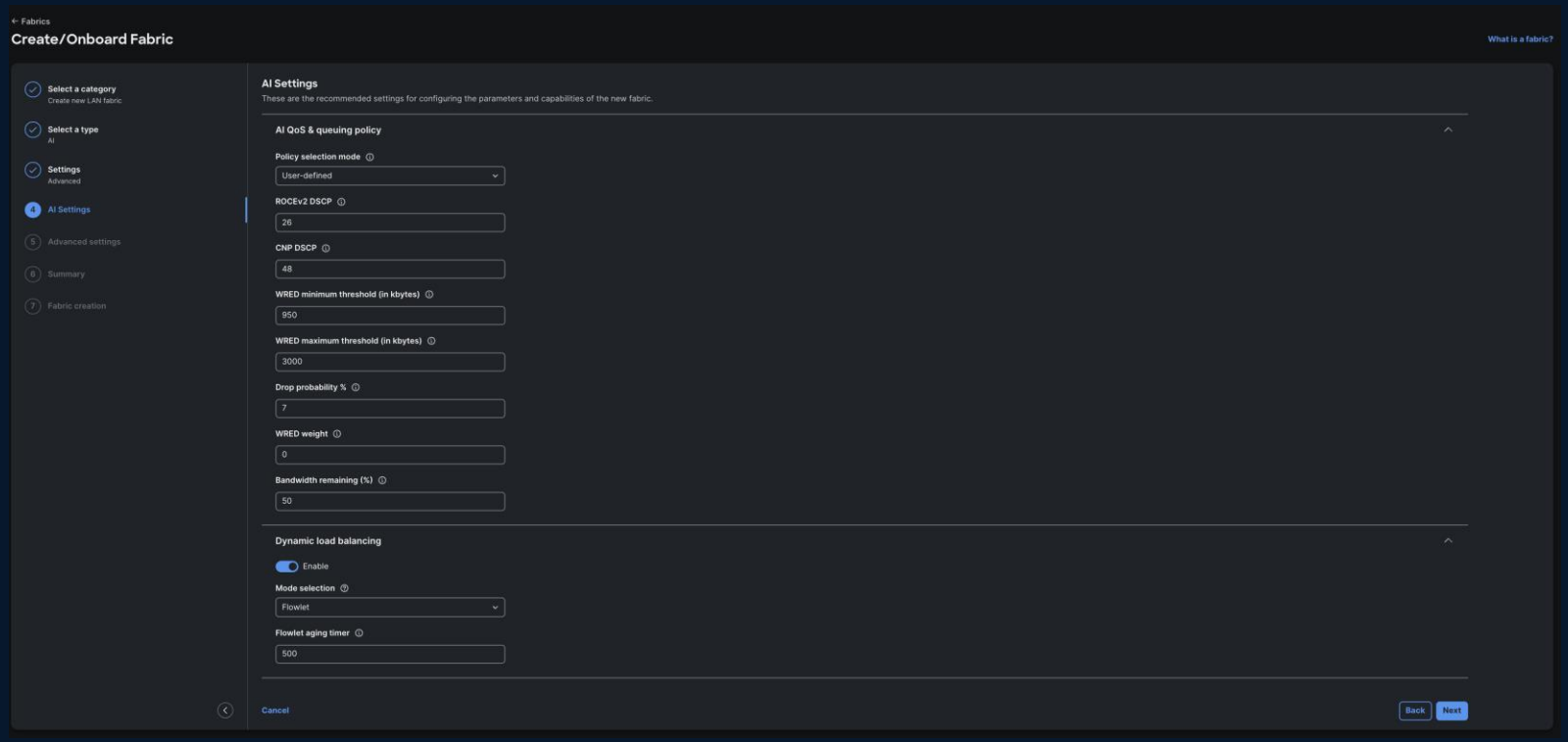
Cisco Nexus Dashboard - Automation

Consolidated Network Operations



Cisco Nexus Dashboard

AI Fabric Settings



- New AI fabric settings step to automatically apply QoS policies tailored for different networks in an AI cluster
- Easily fine-tune your QoS and Dynamic Load Balancing metrics with customizable parameters
- Enable mixed mode load balancing to combine ECMP, flowlet, or per-packet algorithms for specific traffic based on DSCP values
- Automated BGP pool allocation

Quickly and accurately set QoS metrics at fabric creation

Cisco Nexus Dashboard – AI Analytics

Simplifying Network Operations

- AI Network Visibility
- UX/UI Dashboard
- Visibility – Lossless Ethernet
- Monitoring (ECN,PFC)
- Congestion Score
- Application to Network Performance Correlation
- Telemetry and NetOps

With the granular visibility provided by Cisco Nexus Dashboard the network administrator can observe drops

Tune thresholds until congestion hot spots clear and packet drops stop in normal traffic conditions

This is the first and most important step to ensure that the AI/ML network will cope with regular traffic congestion occurrences effectively

Committed

Cisco Nexus Dashboard

Advanced Congestion Detection

Nexus Dashboard provides critical insights into network health, detecting congestion even without Dynamic Load Balancing (DLB) enabled. This deep visibility ensures optimal performance for your AI workloads.

PFC Pattern Analysis

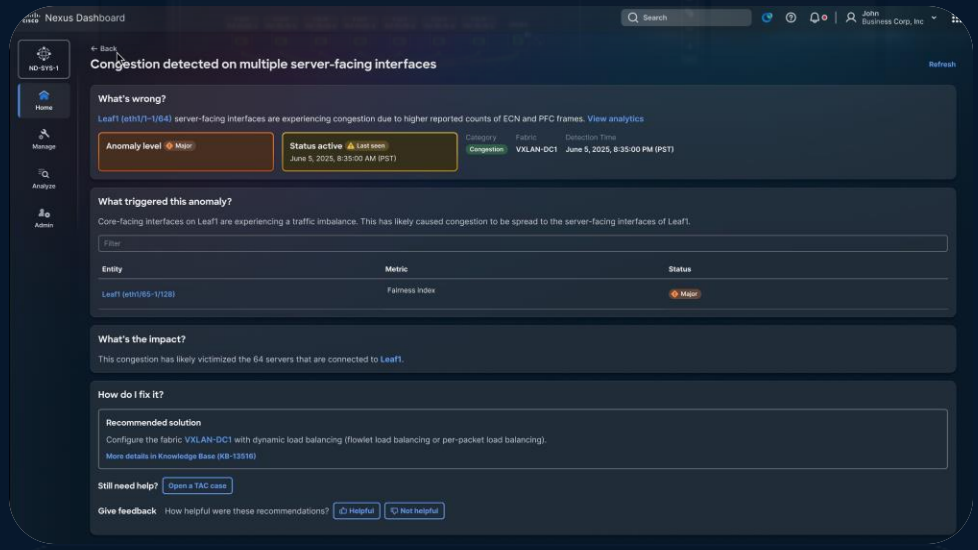
Detects congestion through detailed Priority Flow Control (PFC) patterns.

ECN Packet Monitoring

Analyzes Explicit Congestion Notification (ECN) packets for early warnings.

Hierarchical Scoring

Applies a hierarchical congestion score from interface to switch to the entire fabric.



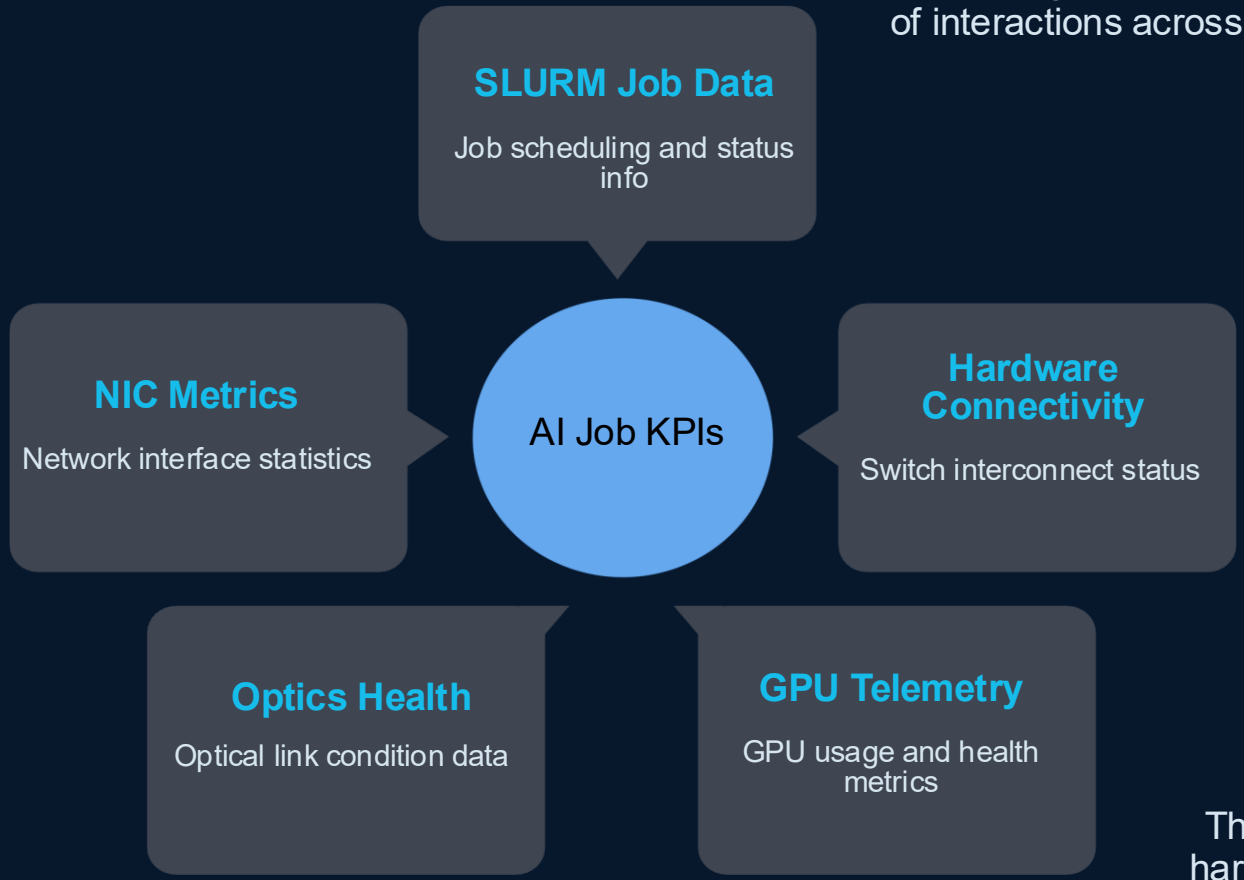
Topology-driven visualization enables rapid drill-down, allowing targeted mitigation through flowlet-based or per-packet load balancing.

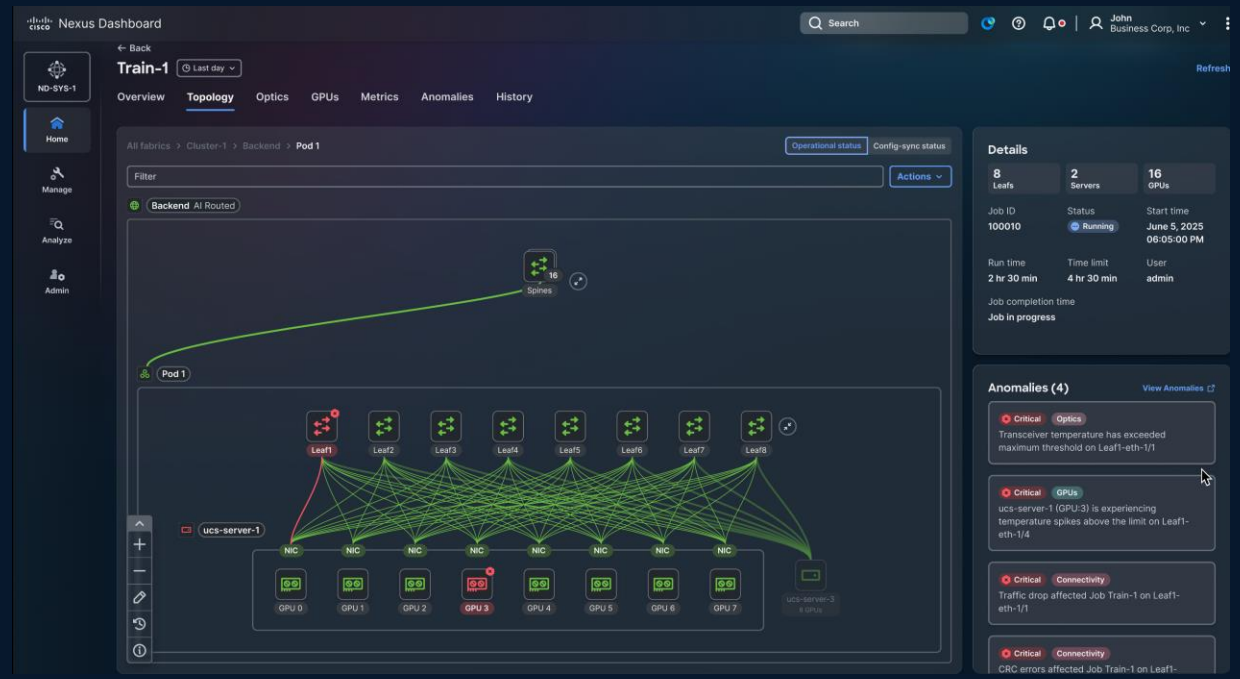
Committed

Cisco Nexus Dashboard

Integration with SLURM for AI Job monitoring, Job-Centric Correlation for AI Fabrics

Optimizing performance for complex AI workloads requires a comprehensive understanding of interactions across the entire fabric, achieved by correlating diverse data streams.

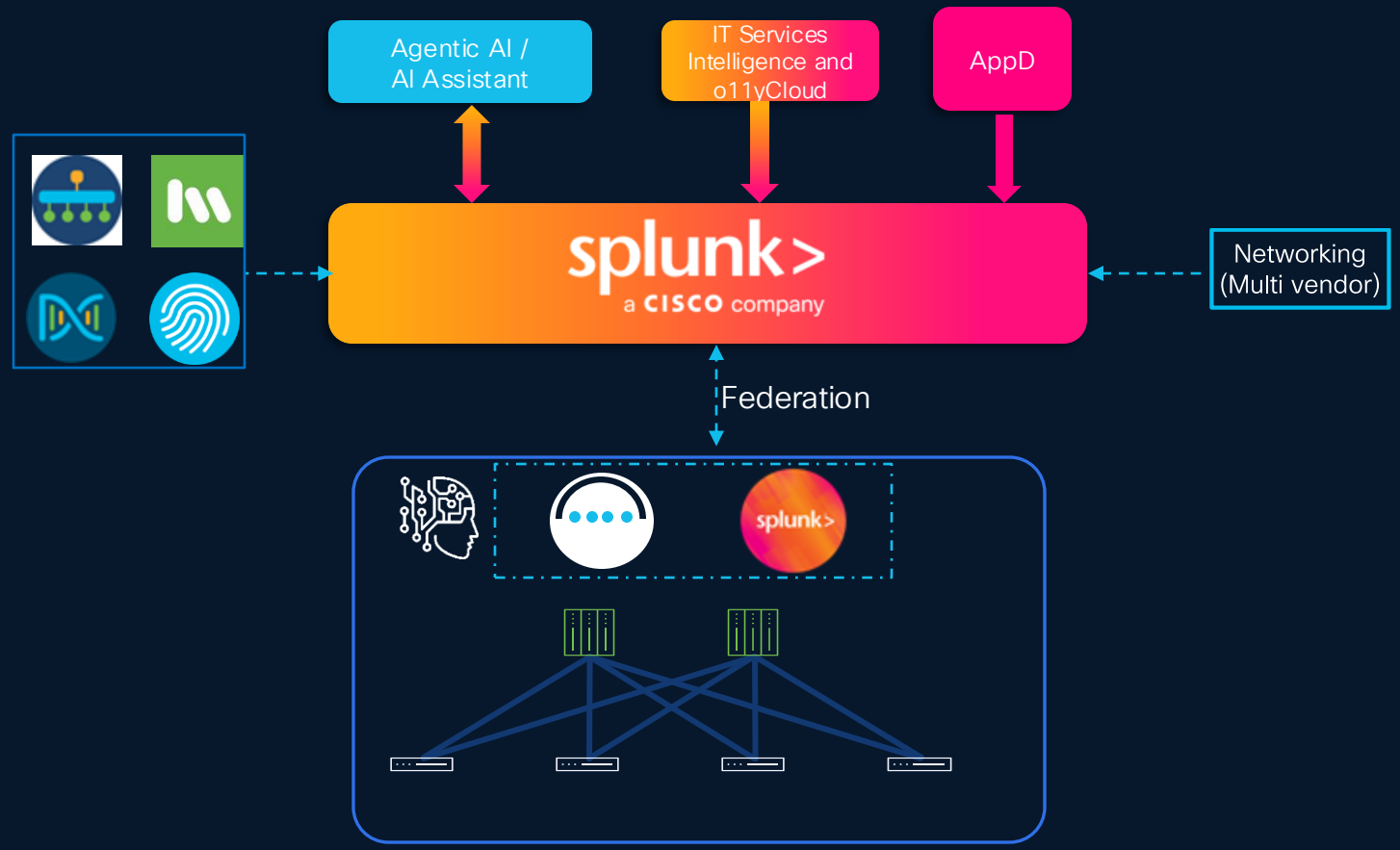




This multi-faceted correlation, spanning orchestration, connectivity, and hardware telemetry, enables proactive detection of potential performance bottlenecks before they impact critical AI application throughput

Committed

Embedded Splunk With Nexus Dashboard



Unified Operations Management: Configuration and Analytics across on prem, cloud and hybrid deployments

Unified analytics platform : Realtime custom dashboarding , reporting and alerts and custom action. Federation with Splunk platform

AI Observability: On Prem AI predictive analytics for data center networking data including AI fabrics

Audit and Compliance : Incremental storage of data for audit, compliance and historical data

Nexus Dashboard + Splunk = Better Together

Splunk Observability Cloud

For AI PODs

OpenTelemetry-native

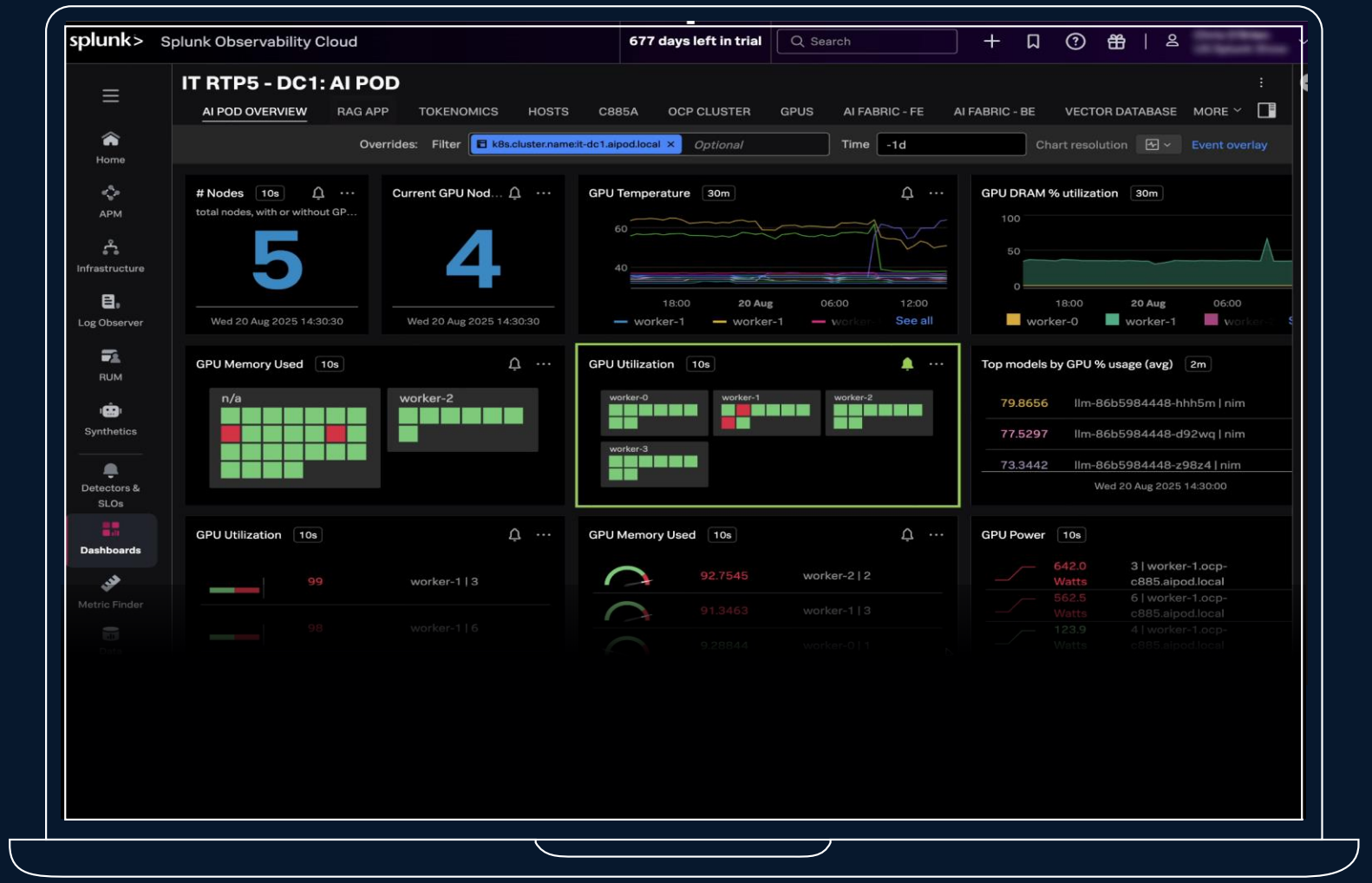
Own and control your data, avoid vendor lock-in and instrument only once on a common standard as you build new applications.

AI powered analytics and guidance

AI/ML driven features like Service Maps and Trace Analytics provide directed guidance that helps you resolve issues faster.

No data sampling

Eliminate blind spots by collecting and analyzing 100% of your data with Splunk's NoSample™ tracing.



Splunk With AI Defense

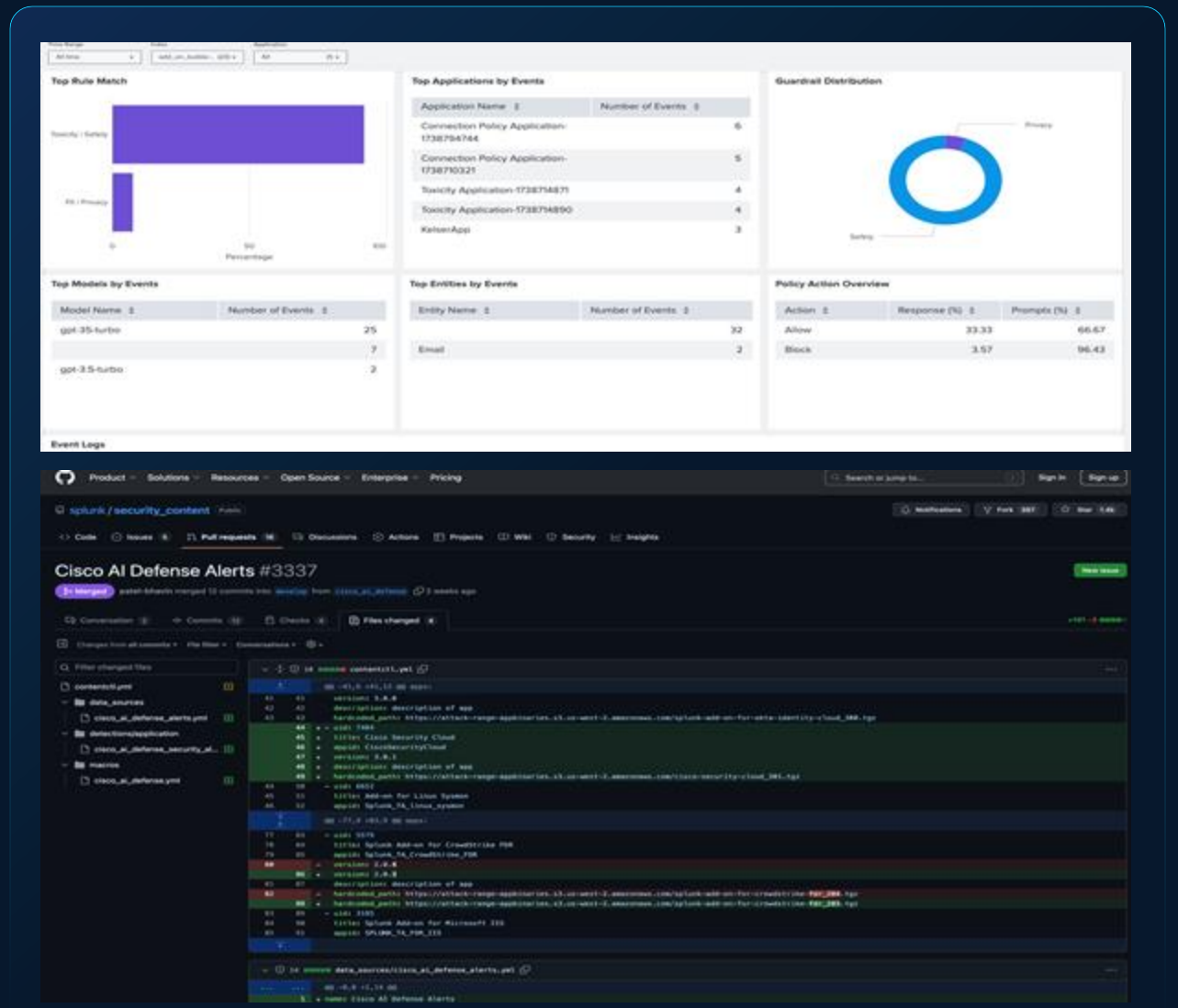
Technical Add-On

Gain visibility into emerging AI risks with Splunk

Pulls in alerts from AI Defense and maps them to the Common Information Model (CIM), visualized in a dashboard.

Gain visibility into risks associated with LLM models, AI apps and entities.

Includes an out-of-the-box Enterprise Security detection that creates a search and surfaces potential attacks against the AI models running in your environment.



Cisco Data Fabric

A revolutionary new architecture to harness the value of machine data with AI

Federated Search

Cisco Differentiation

The Security

Security-first architecture enables safe enterprise AI

The Network

High-performance integrated AI networking enables efficient model training and inferencing

The Assurance

Pre-validated AI infrastructure stack with flexible deployment options improves data scientists and developer productivity

Thank you


CISCO