

# Secure AI Factory Network: Scalable, Lossless Ethernet Fabrics with Cisco Silicon One, Nexus Platforms, and Hyperfabric AI

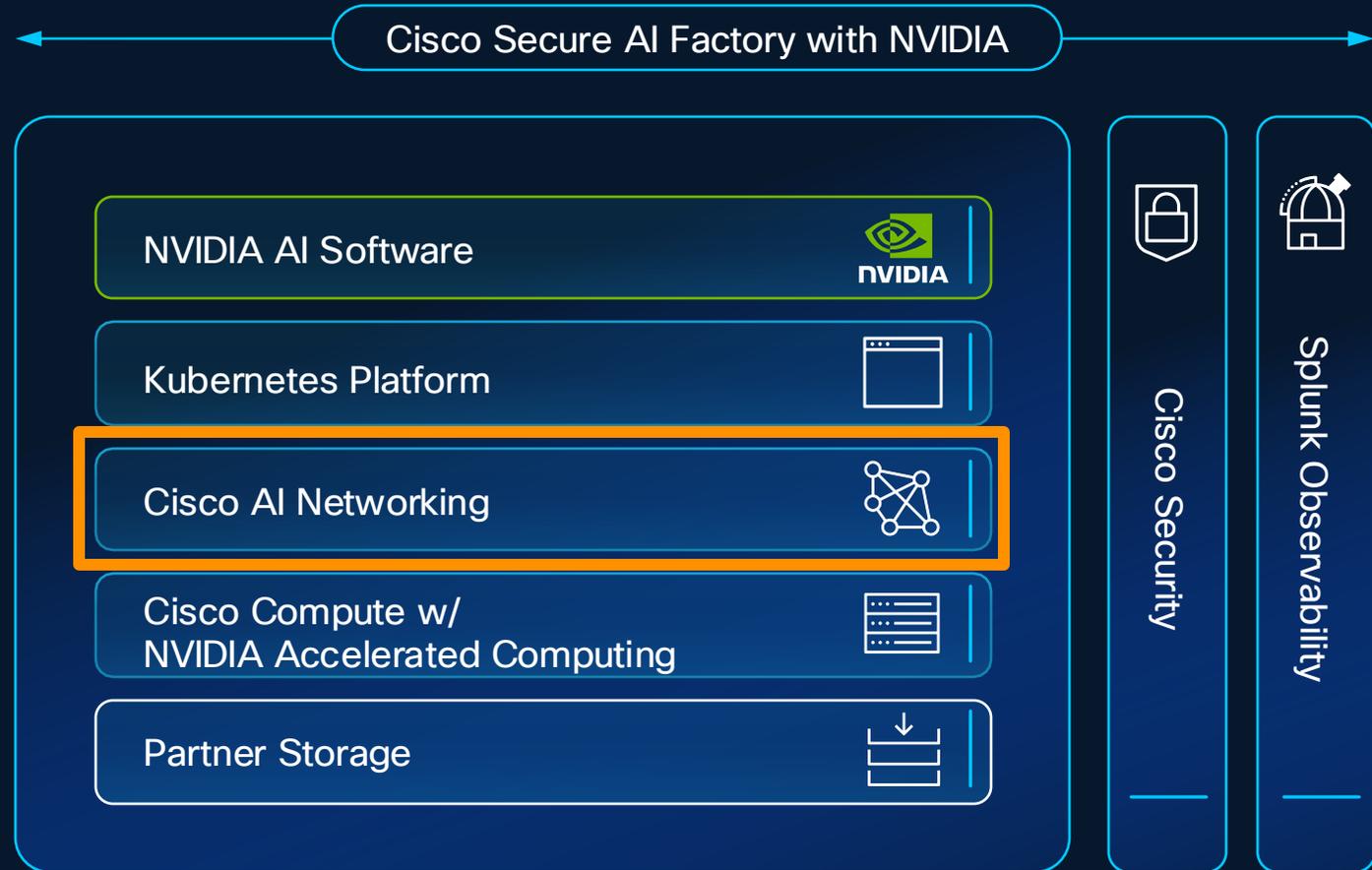


Raul Arias, DCN SE | Cloud and AI Infrastructure

# Cisco Secure AI Factory with NVIDIA

What is it?

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



# Agenda

## 1. AI Ready DC Network **Infrastructure**

- AI fabric design requirements
- Hardware innovation – Cisco silicon, platforms, optics
- Software innovation – NXOS feature / function for AI fabrics

## 2. AI Ready DC Network **Security**

- Hardware integrated security – Cisco Smart Switches
- Software integrated security – Cisco Live Protect

## 3. AI Ready DC Network **Operation**

- Unified Nexus Dashboard
- Operational integration(s)

# AI Ready DC Network Infrastructure

# As Enterprises Adopt AI There Are Unique Network Challenges



Parallel GPU workloads  
create massive, bursty  
east-west traffic



AI jobs require  
predictable, lossless  
packet delivery



Conventional  
congestion control  
insufficient for AI scale

# Cisco Data Center Networking – AI Workload



## Cisco Nexus Hyperfabric

Cisco cloud-managed, fully integrated

Easily design, order, deploy, monitor and upgrade fabrics

Purpose built vertical stack

Cisco 6000 series switches



## Cisco Nexus (Dashboard)

Private cloud managed, flexible solution

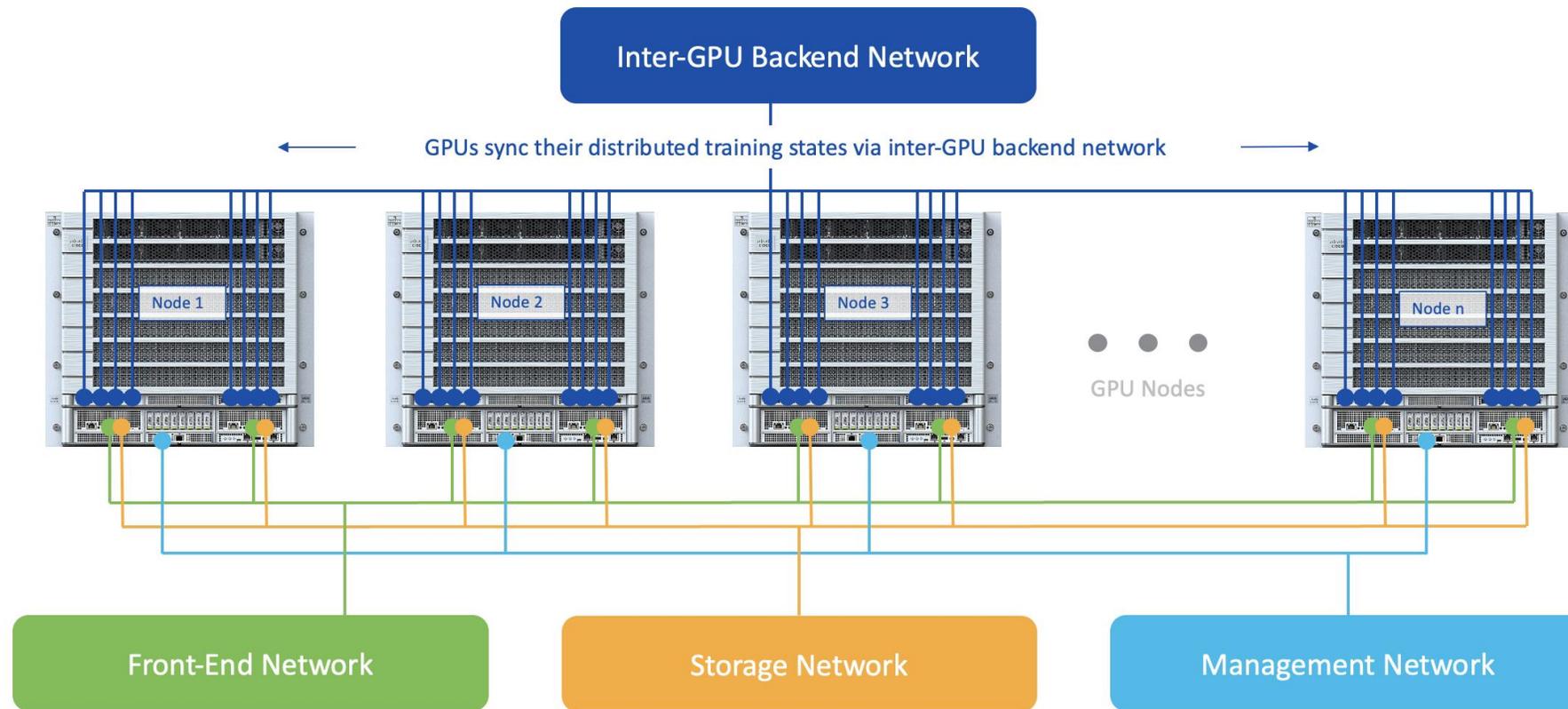
Versatile data center solution

Nexus Dashboard for simplified ops

Nexus 9000 series switches

# Functional Networks

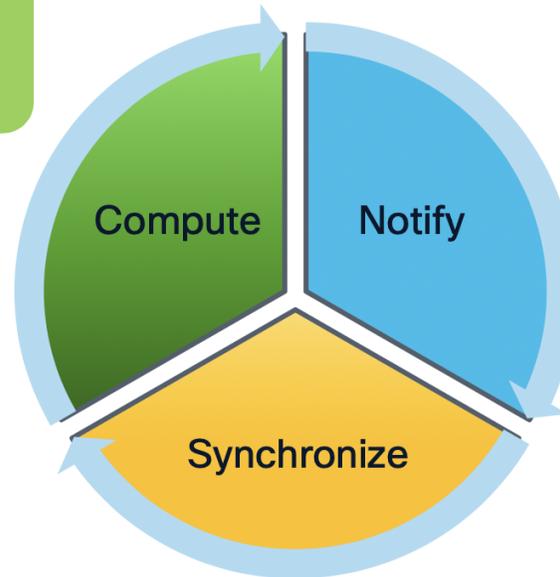
## UCS C880A M8 Networking Blueprint



# The AI/ML Workload Cycle

## GPU Execute Instructions

High Bandwidth capable GPUs can saturate network links



## Send results of computation

Different collective communication patterns  
All Reduce (Aggregate/reduce everyone's data and send to everyone)

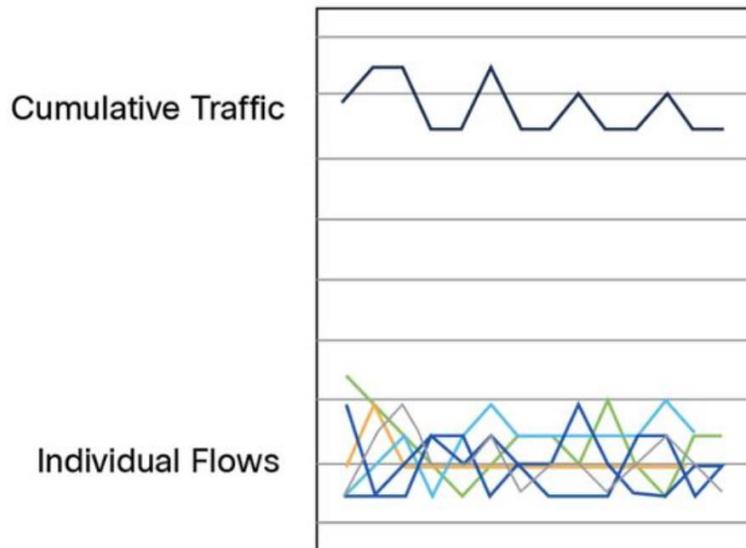
## Wait for all GPUs to complete

Synchronizes all GPUs  
Compute stalls, waiting for the slowest path  
**Job Completion Time (JCT) influenced by the worst-case tail latency**

# Your AI/ML Training is only as fast as the slowest GPU

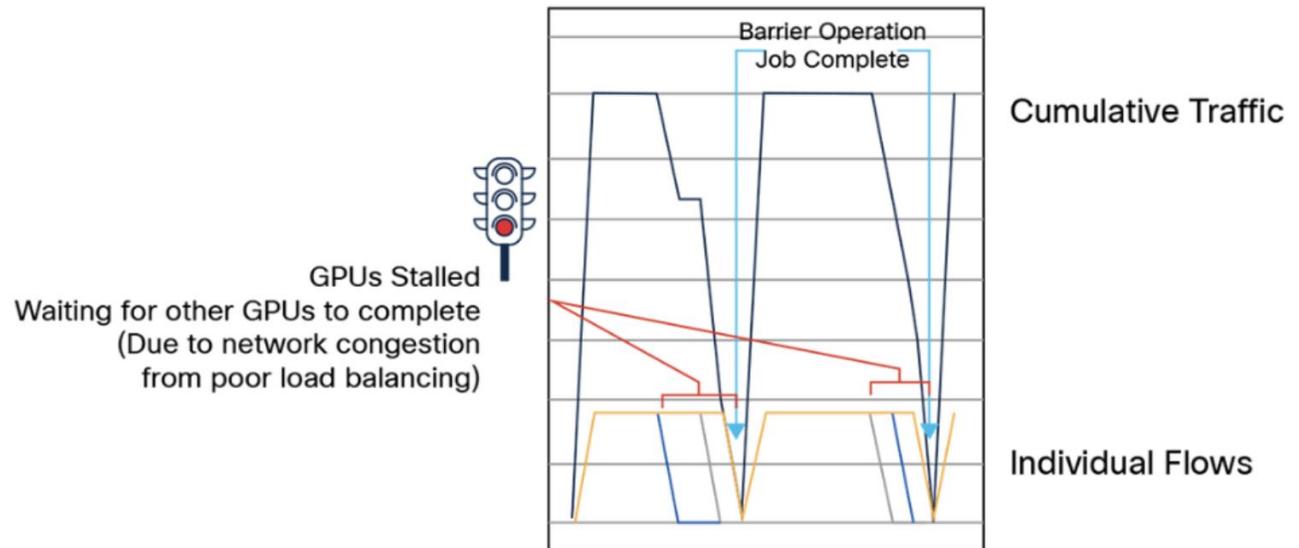
*The network can become the bottleneck*

Traditional DC Traffic Pattern



Many asynchronous small BW flows  
Chaotic pattern averages out  
to consistent load

AI (All-to-all Collective) Traffic Pattern



Few synchronous high BW flows  
Synchronization magnifies long tail  
latency and bad load balancing decisions

# Oversubscription and Non-Blocking

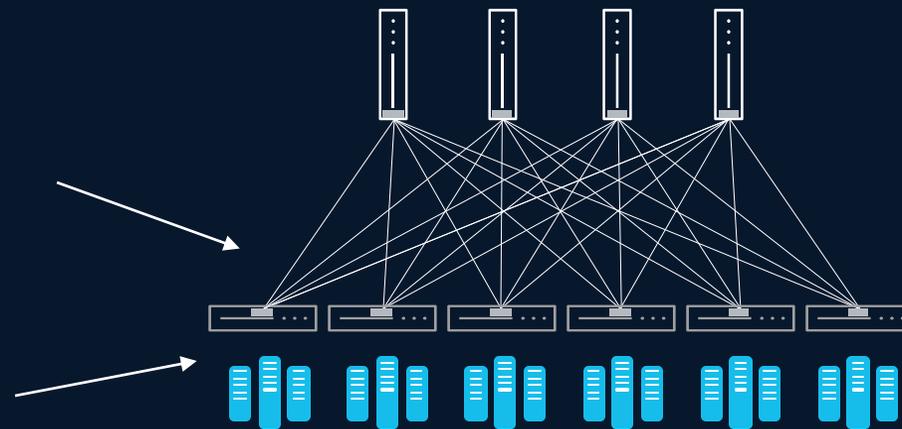
*Oversubscription happens when an intermediate network device (switch) or link doesn't have enough capacity (bandwidth) to allow line rate communication between the two devices*

## Oversubscription Ratio

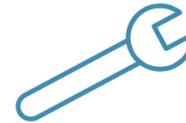
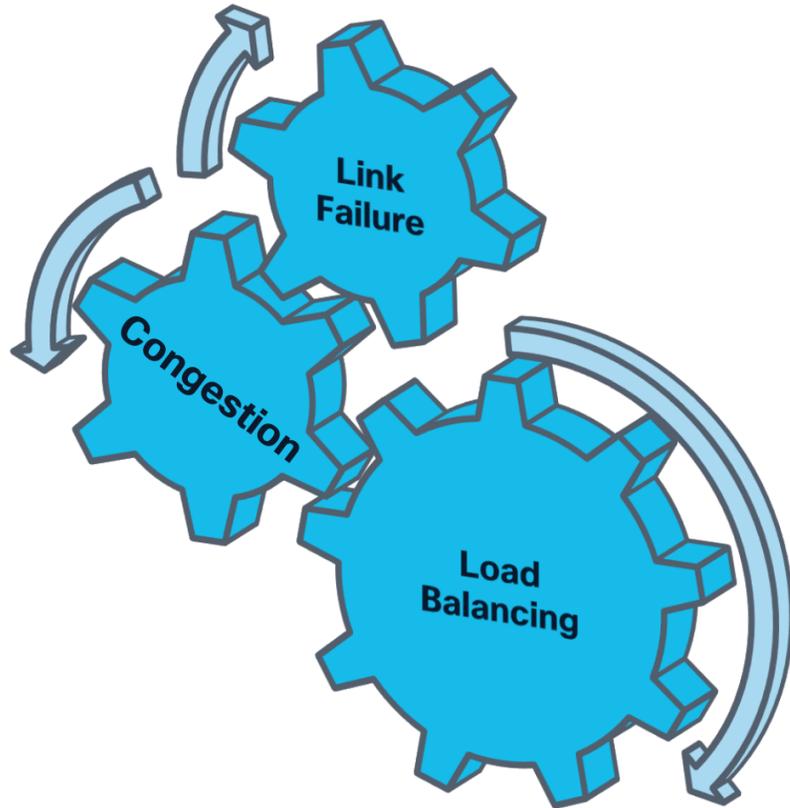
$$[\text{Number of Downlink Ports} * \text{Speed}] \div [\text{Number of Uplink to Spine} * \text{Speed}] : 1$$

**Uplink Ports**  
4x100Gbps = 400 Gbps

**Downlink Ports**  
48x25Gbps = 1.2Tbps



# Minimizing Job Completion Time is *the AI Challenge*

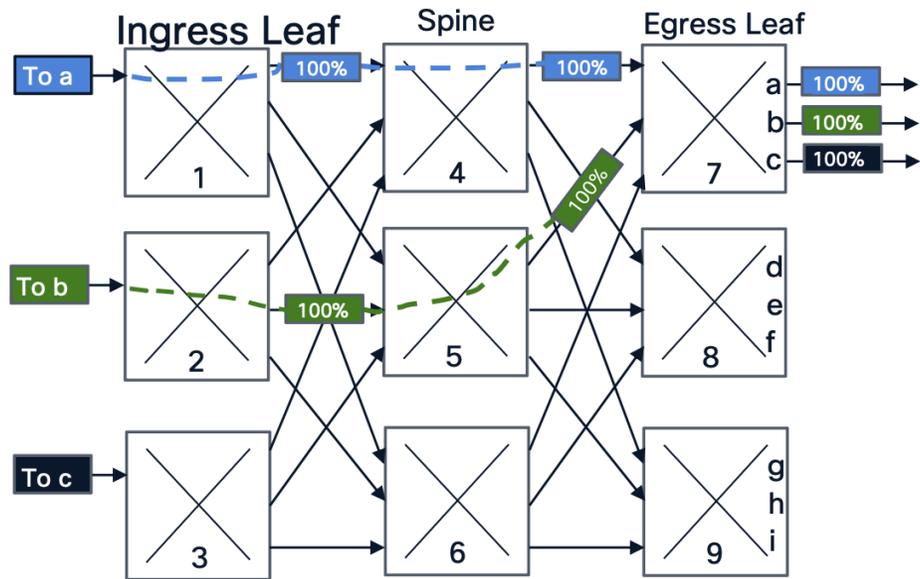


## Wrenches in the works

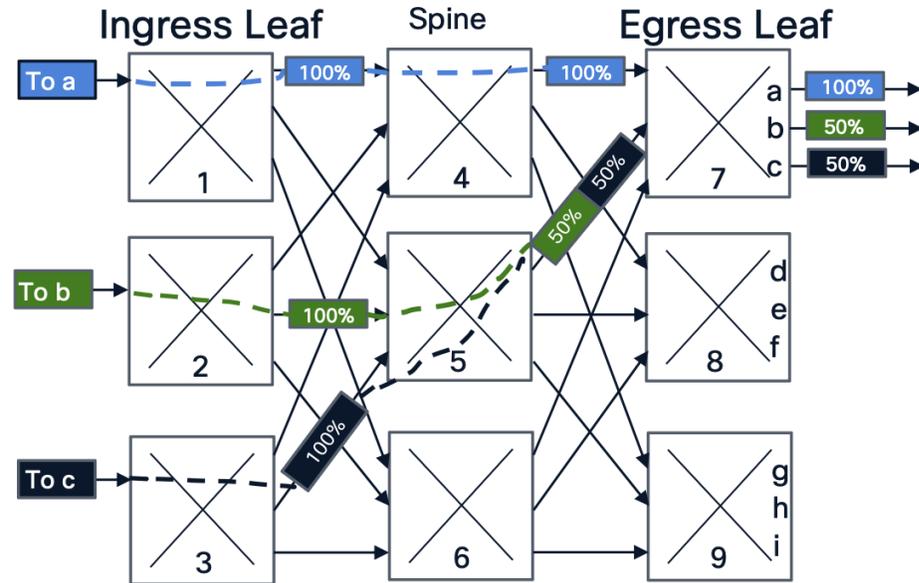
- Underutilized fabric links
- Head of Line blocking
- Incast Congestion
- Link failures and black holing

# Load Balancing Basics - ECMP

Uncongested Links

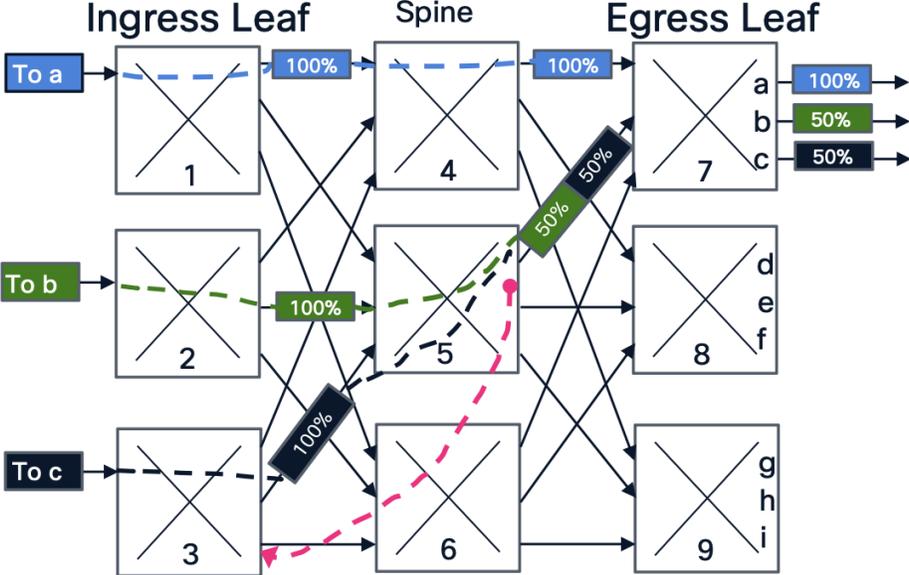


Congested Links

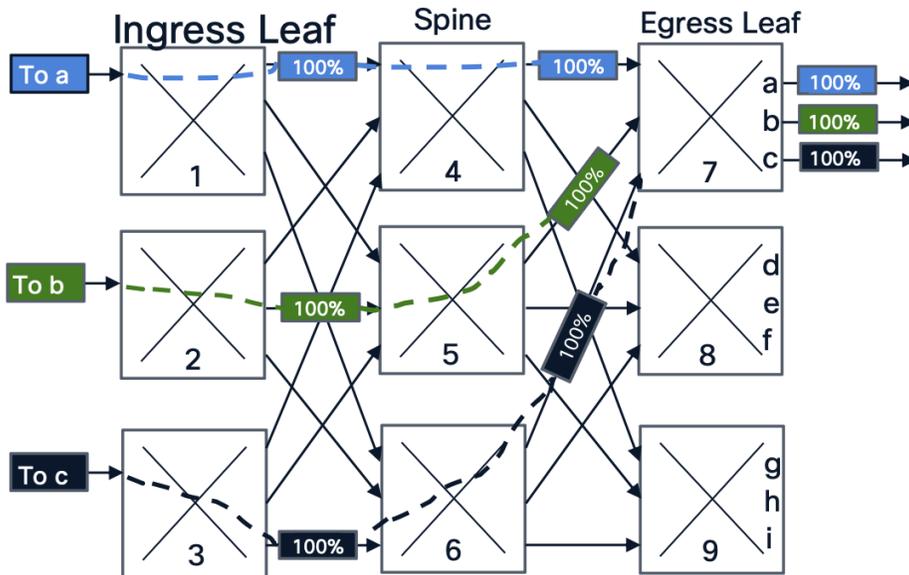


# Load Balancing Basics – Enhanced Ethernet

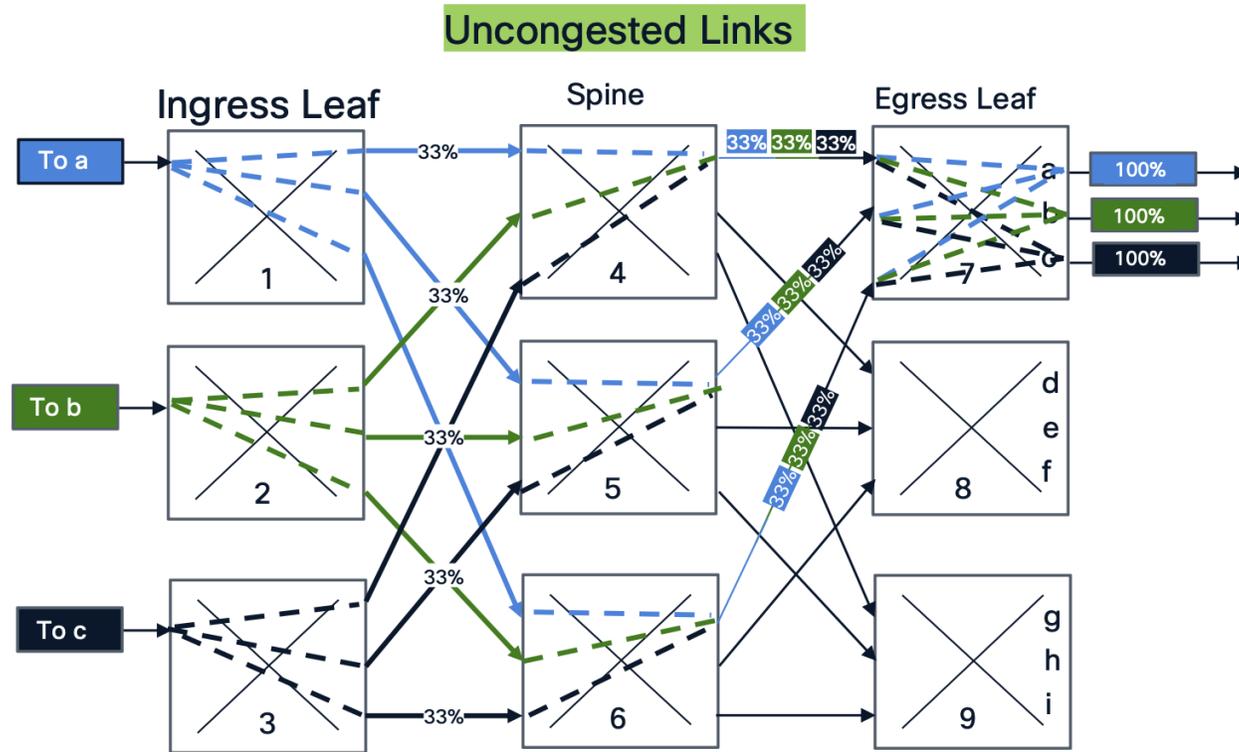
Congested Links



Uncongested Links



# Load Balancing Basics – Scheduled Ethernet



# AI Ethernet Fabric Options

	1 Ethernet	2 <i>Enhanced</i> Ethernet		3 <i>Scheduled</i> Ethernet
<b>Load Balance</b>	Stateless ECMP	Stateful Flow/Flowlet	Spray & Re-order in SmartNIC	Spray & Re-order in leaf
<b>Fabric Congestion Management</b>	Congestion Reaction with ECN/PFC	Congestion Reaction with congestion score to adjust distribution		Congestion Avoidance
<b>Link Failure</b>	Software	Hardware		
<b>Job Completion Time</b>	Good	Better		<i>Best</i>
<b>Coupling between NIC and Fabric</b>	No		Yes	No
<b>Place in Network</b>	Frontend & Backend, Training & Inference			
<b>Fabric Infrastructure</b>	Leaf/Spine or Modular Chassis			Modular Chassis

# ULTRA ETHERNET VISION

Deliver an Ethernet based open, interoperable, high performance, full-communications stack architecture to meet the growing network demands of AI & HPC at scale

**THE NEW ERA  
NEEDS A  
NEW  
NETWORK**

*Ultra***Ethernet**

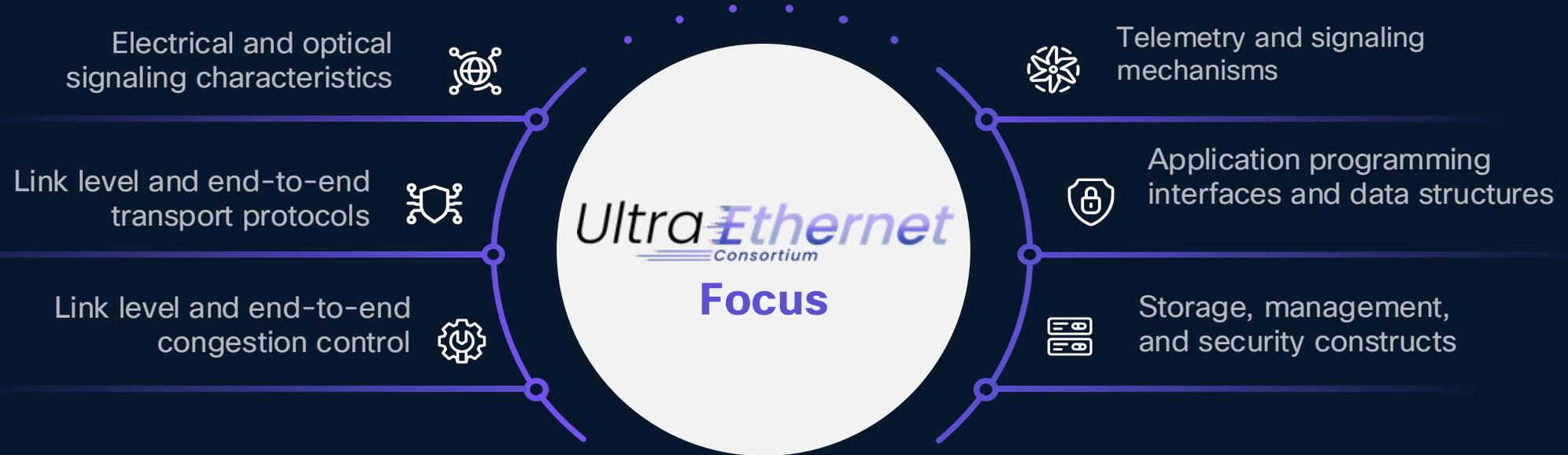
As **performant** as a  
supercomputing interconnect

As **ubiquitous** and  
**cost-effective** as Ethernet

As **scalable** as a cloud data  
center

# UEC Seeks to Bring Open Standards to AI Networks

*Open* specifications, APIs, source code for optimal performance of AI and HPC workloads at scale.



ARISTA



# AI Ethernet Fabric Options

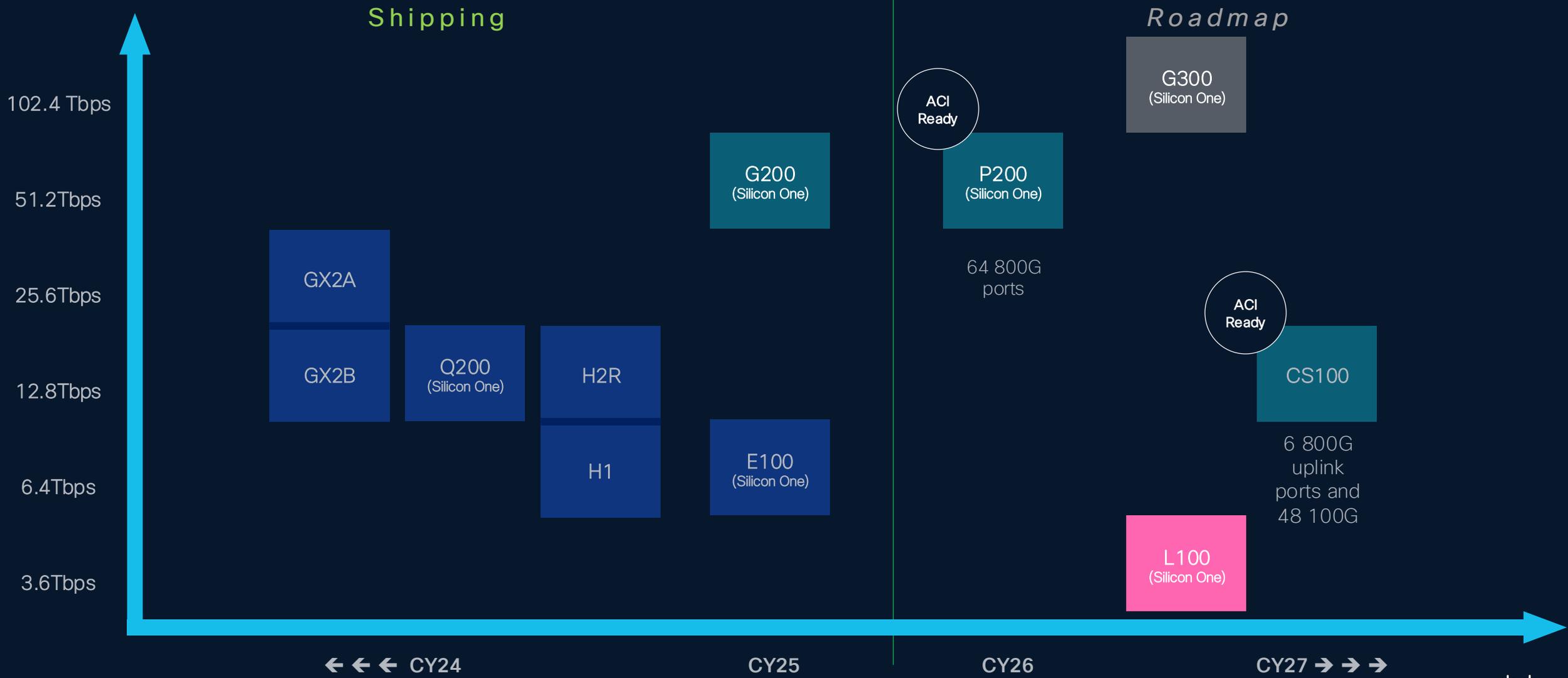
	1 Ethernet	2 Enhanced Ethernet	2 <sup>u</sup> <i>Ultra Ethernet</i>	3 Scheduled Ethernet	
<b>Load Balance</b>	Stateless ECMP	Stateful Flow/Flowlet	Spray & Re-order in SmartNIC	Endpoint Controlled adaptive packet spraying	Spray & Re-order in leaf
<b>Fabric Congestion Management</b>	Congestion Reaction with ECN/PFC	Congestion Reaction with congestion score to adjust distribution		Network influenced Congestion Management	Congestion Avoidance
<b>Link Failure</b>	Software	Hardware			
<b>Job Completion Time</b>	Good	Better		Even Better	<i>Best</i>
<b>Coupling between NIC and Fabric</b>	No		Yes	No	
<b>Place in Network</b>	Frontend, Backend			Backend	Frontend, Backend
<b>Fabric Architecture</b>	Leaf/Spine or Modular Chassis			Modular Chassis	

*Effectiveness IS dependent on Traffic Characteristics*

*Effectiveness IS NOT dependent on Traffic Characteristics*

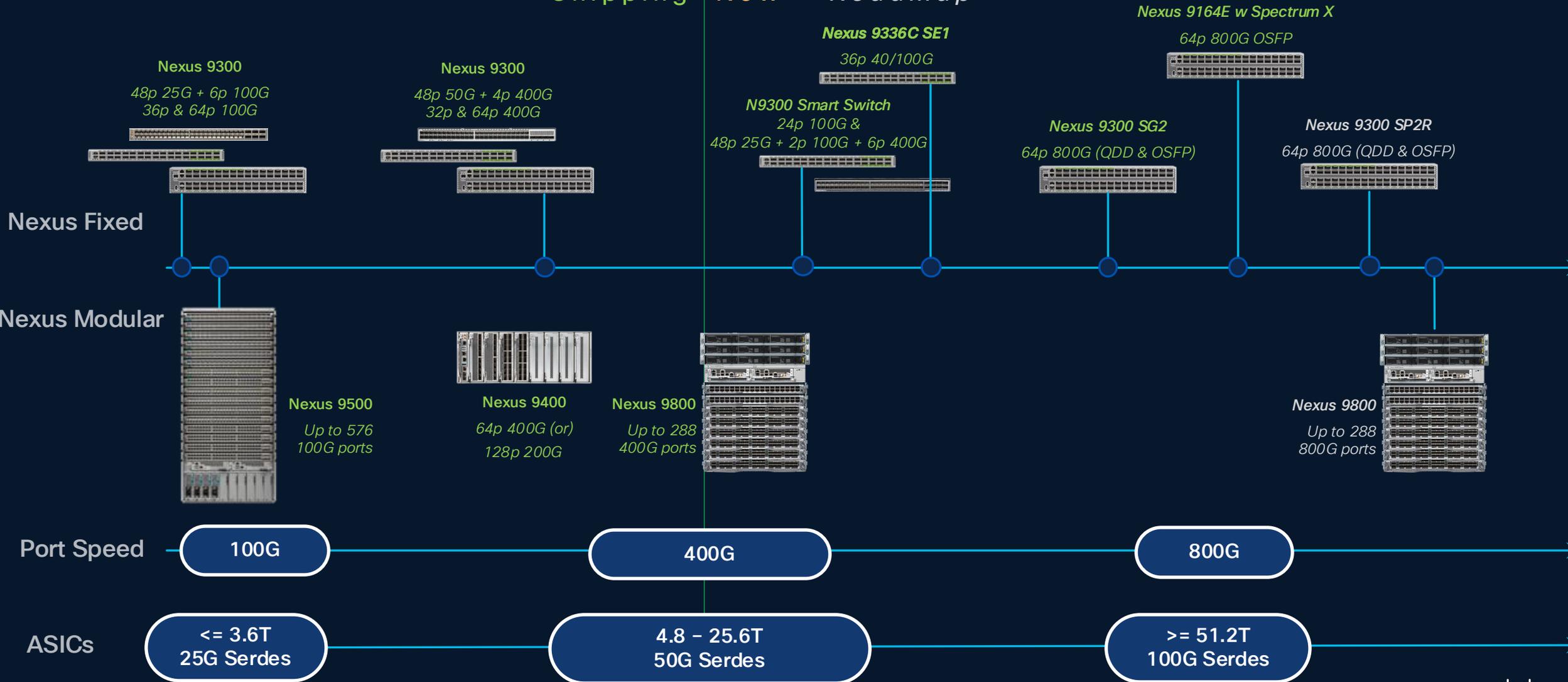
# Cisco Nexus 9000 ASICs

3nm; 200G SERDES
3 or 5nm; 100G SERDES
7nm; 50G SERDES
5nm 25G SERDES



# Cisco Nexus 9000 Series Switches

Shipping *New + Roadmap*



# Cisco Nexus 9300 64 Port 800G Switch

512-wide radix

Fully shared packet buffer

Advanced load balancing

Low Latency

Compact 2RU 51.2T Switch

G200 ASIC (5nm) | 100G SerDes | 256MB packet buffer

64 800G ports | Up to 128 line-rate 400G ports (2x400G breakout)

Choice of QSFP-DD or OSFP ports

Multi Core x86 CPU | 32GB RAM | 128GB SSD

Cisco NXOS spine and AI/ML spine/leaf capable

Nvidia FGLB license available



N9364E-SG2-Q or N9364E-SG2-O

27C | 50% load | 64x DAC

995 W

27C | 100% load | 64x DAC

1,186 W

27C | 100% load | 64x 16W optics

2,332 W

40C | 100% load | 64x 16W optics

2,567 W

Ultra Ethernet  
Consortium

READY



Cisco Silicon (S1 and Cloud Scale) & Cisco Optics fully validated as part of **Spectrum-X**

**The only 3rd party switch validated as part of the reference architecture**

Support packet spraying and adaptive routing capabilities in Cisco Silicon Nexus Switches

Co-develop Cisco N9164E switch based on **Spectrum-X ASIC**

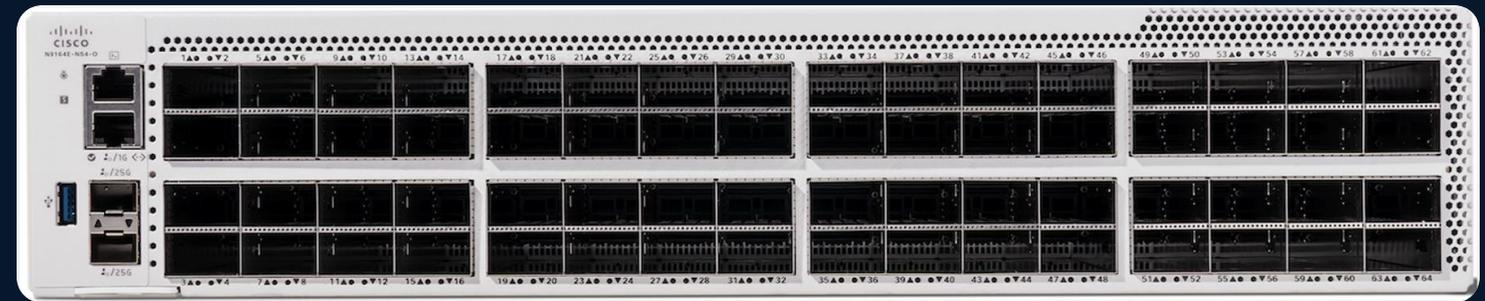
Orderable!

# Cisco Nexus 9164E 64 Port 800G Switch

NCP Reference Architecture (RA) Compliant

Advanced load balancing

Low Latency



Compact 2RU 51.2T Switch

Nvidia Spectrum-X ASIC

64 800G ports | Up to 128 line-rate 400G ports (2x400G breakout)

Low-Latency, Lossless Ethernet

Supported by Nexus Dashboard

Supports NX-OS or SONiC

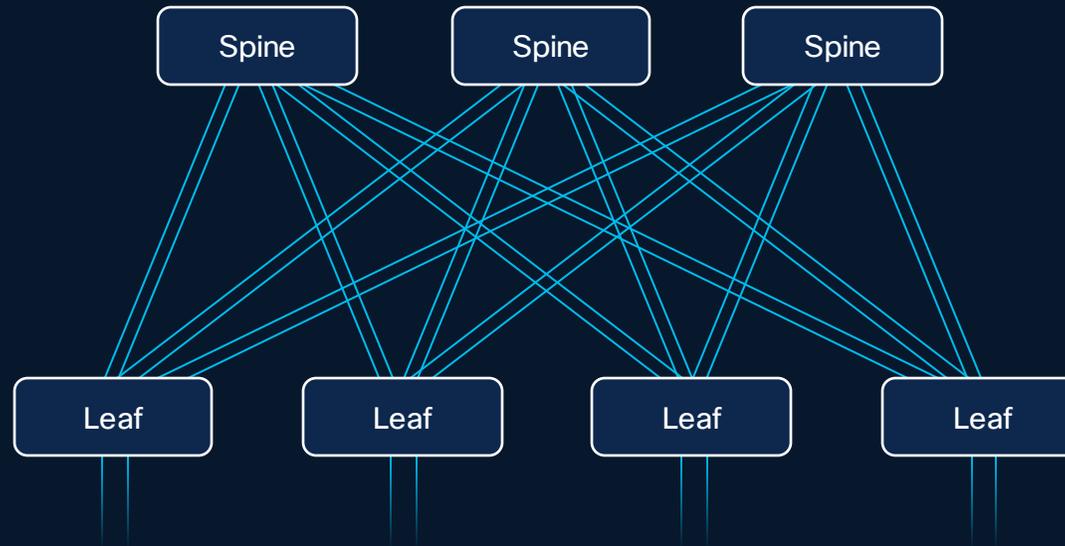
Ultra Ethernet Consortium

READY

# Powering AI Fabrics with Cisco Intelligent Packet Flow

Intelligent Load Balancing for every path and every flow

Ultra Ethernet  
Consortium  
**READY**



**NEW'ish**

## FEATURES

Dynamic Load Balancing (DLB) with Load and Congestion Awareness

Weighted Cost Multi Path Load Balancing (h-WCMP) with Dynamic Load Balancing

## Advanced load balancing

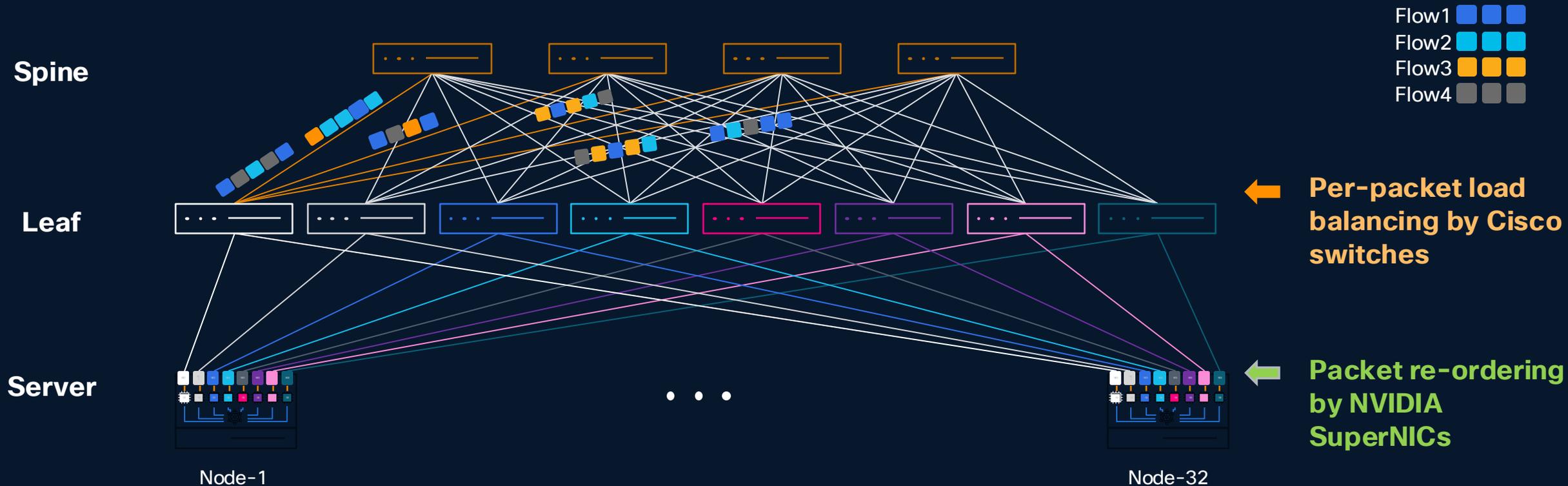
Per Packet and Selective Packet Spray (Ex: RDMA vs. non-RDMA) + **NVIDIA Spectrum-X integration\***

Policy Based Flowlet Load Balancing (DSCP, ACL...) + **RoCEv2 Header Filter\***

ECMP Static Pinning

High Entropy ECMP

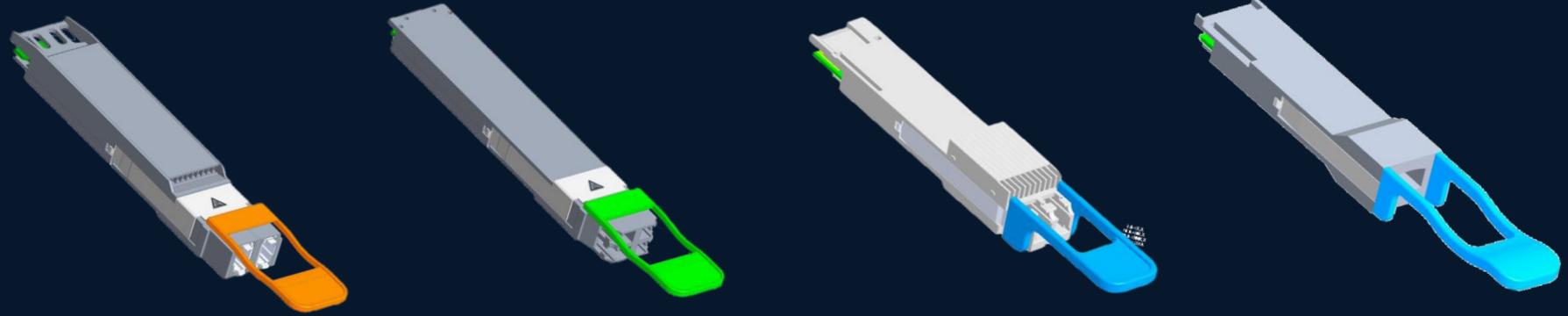
# NEW Per-Packet Load Balancing with **NVIDIA Spectrum-X** Integration



## Cisco Per-Packet Load Balancing with **NVIDIA Spectrum-X** integration

- Traffic is forwarded on a **per packet basis**, every packet will be hashed on different output port
- Expected that receiver will put packet in order (**server NICs do the packet reordering**)
- **Turned 'ON' on-NIC** by connecting to Spectrum-X capable switches
- **Provide best utilization of network links, leading to more efficient and reliable data transmission**

# Pluggable Optical Modules Options



OSFP

OSFP-RHS

QSFP-DD

QSFP

Host high-speed pins	8	8	8	4
Module capacity (Gbit)	400, 800, 1600	400, 800, 1600	400, 800, 1600	40, 100, 200, 400
Cooling	Integrated heatsink	Riding heatsink	Riding heatsink	Riding heatsink
High-speed IO Compatibility (Gbps)	50-100-200	50-100-200	50-100-200	1-25-50-100
Port compatibility	OSFP	OSFP-RHS	QSFP-DD & QSFP	
Management	CMIS			

System and ASIC design dictate which modules are supported and speed backwards compatibility

# Cisco Data Center Networking – AI Workload



## Cisco Nexus Hyperfabric

Cisco cloud-managed, fully integrated

Easily design, order, deploy, monitor and upgrade fabrics

Purpose built vertical stack

Cisco 6000 series switches



## Cisco Nexus (Dashboard)

Private cloud managed, flexible solution

Versatile data center solution

Nexus Dashboard for simplified ops

Nexus 9000 series switches

# Cisco Nexus Hyperfabric

It's the same Hyperfabric service,  
extended for AI



NOW ORDERABLE

# Cisco Nexus Hyperfabric

High-performance Ethernet

Cloud-managed operations

Unified stack including NVAIE

AI-native operational model

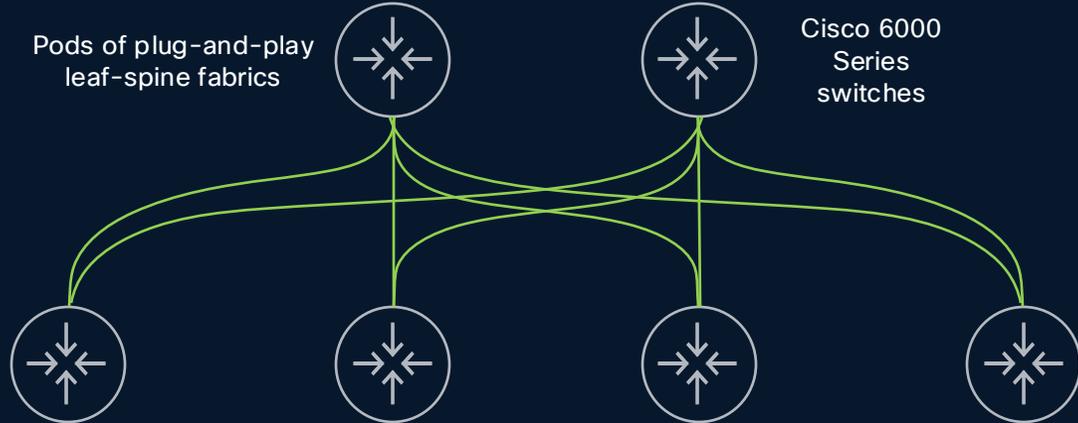
Democratize AI infrastructure

Visibility into full stack AI

## Cisco Nexus Hyperfabric



### On-prem AI infrastructure

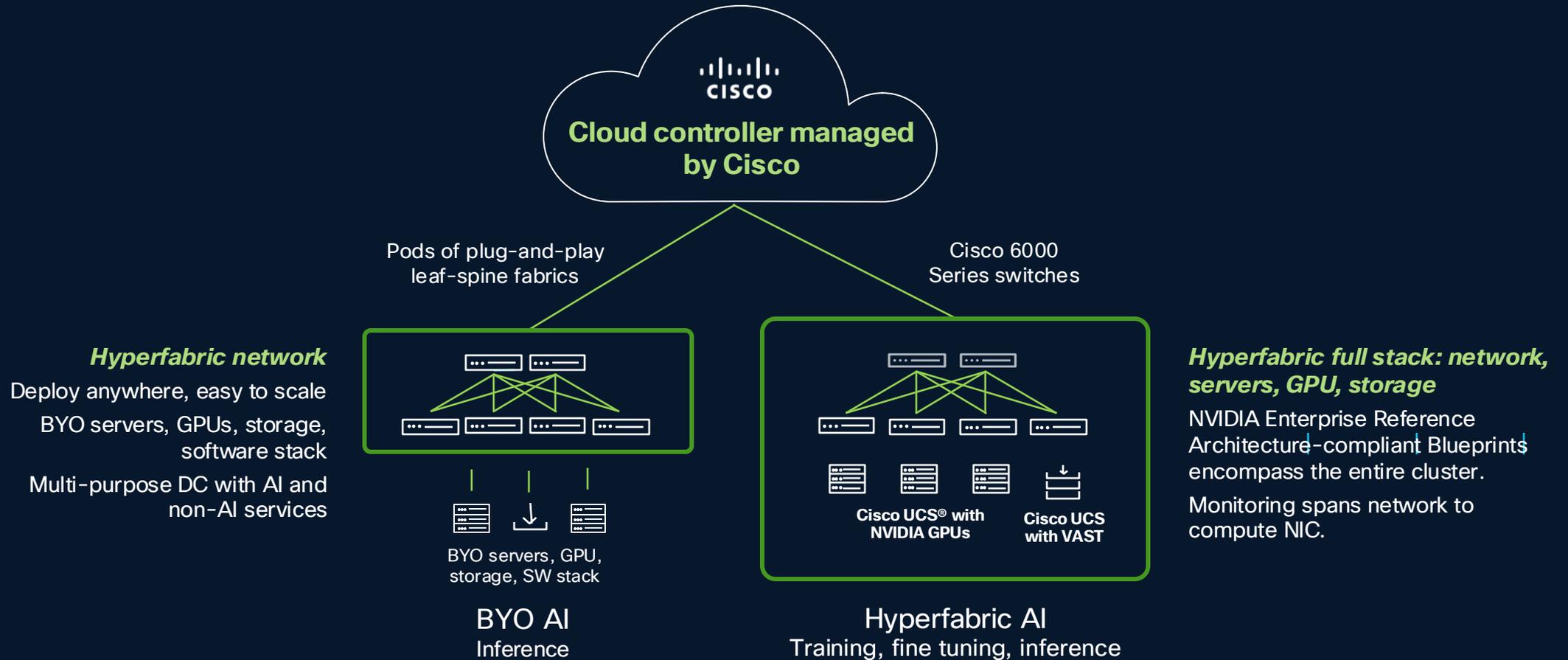


NVIDIA GPU

NVIDIA DPU / NIC BlueField-3



# AI Solutions with Cisco Nexus Hyperfabric



**Easy to design, deploy, operate, and scale High-performance Ethernet networks**

# AI Ready DC Network Security

# Cisco Hybrid Mesh Firewall in Data Center

## SECURITY CLOUD CONTROL



Write policy once, enforce across the mesh

# Nexus Smart Switch

Unmatched Flexibility, Performance, and Efficiency

Cisco  
Smart Switches

## Networking



- Rich NX-OS Features and Services
- High-speed connectivity and scalable performance
- Optimized for latency and power efficiency



Routing  
Switching



EVPN/MPLS/  
VXLAN/SR



Rich  
Telemetry



Line-rate  
Encryption



Power  
Efficiency



Cisco Nexus 9300 Services Accelerated Switch

## Services



- Software-defined Stateful Services
- Programmable at all layers: add new services without HW change
- Scale-out services with wire-rate performance
- Power down DPU complex when not used



Hypershield  
Security



IPSEC  
Encryption



Large-Scale  
NAT



Event-Based  
Telemetry



DoS  
Protection

Future Use Cases

# 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

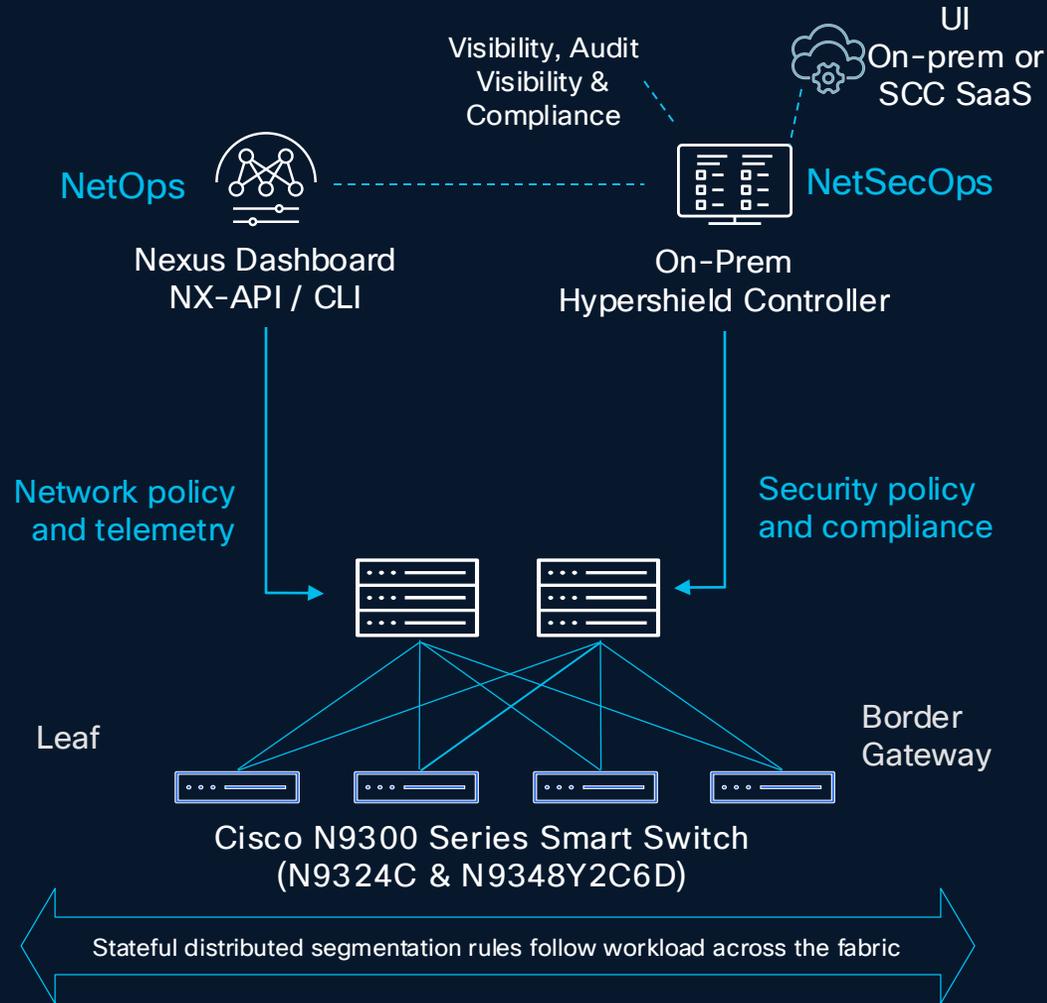
Top of Rack segmentation and enforcement

Cloud Edge

Zone-based segmentation

# Smart Switch “Networking & Security” Use Case

Top of Rack L4 Segmentation – November GA



## Security Infused in Data Center Fabric

**Version:** NXOS 10.6(2), Hypershield 1.2

**Smart Switches:** N9348Y2C6D-SE1U, N9324C-SE1U

**Fabric:** VXLAN-EVPN, VXLAN-multi-site, BGP fabric, brownfield

**Segmentation:** VRF/VLAN + CIDR rules, stateful/stateless, 100K rules, 800G throughput (final scale based on benchmarking)

**Policy:** CRD schema, policy validation and canary rollout/rollback

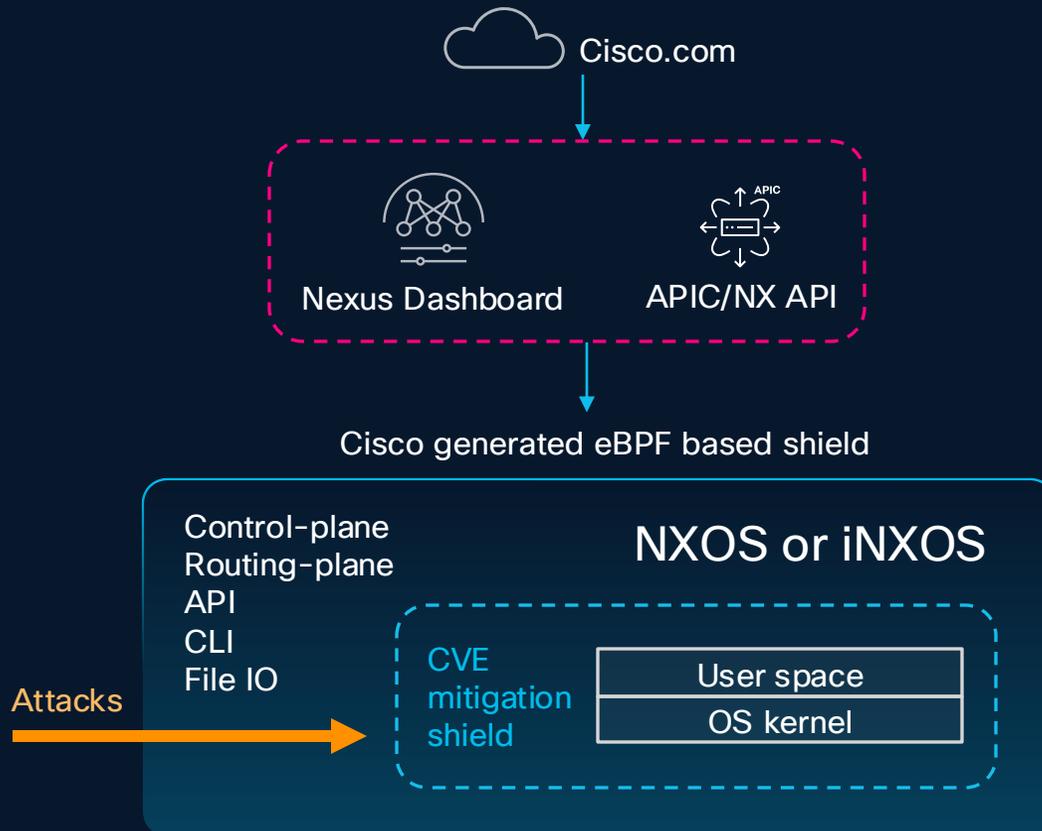
**Hypershield:** Air-gap ready on-prem controller\* and optional Security Cloud Control SaaS

**Upgrade:** NXOS CLI for DPU load, SMU for Hypershield agent

**Observability:** Nexus Dashboard, Splunk, Prometheus/Grafana

# 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)

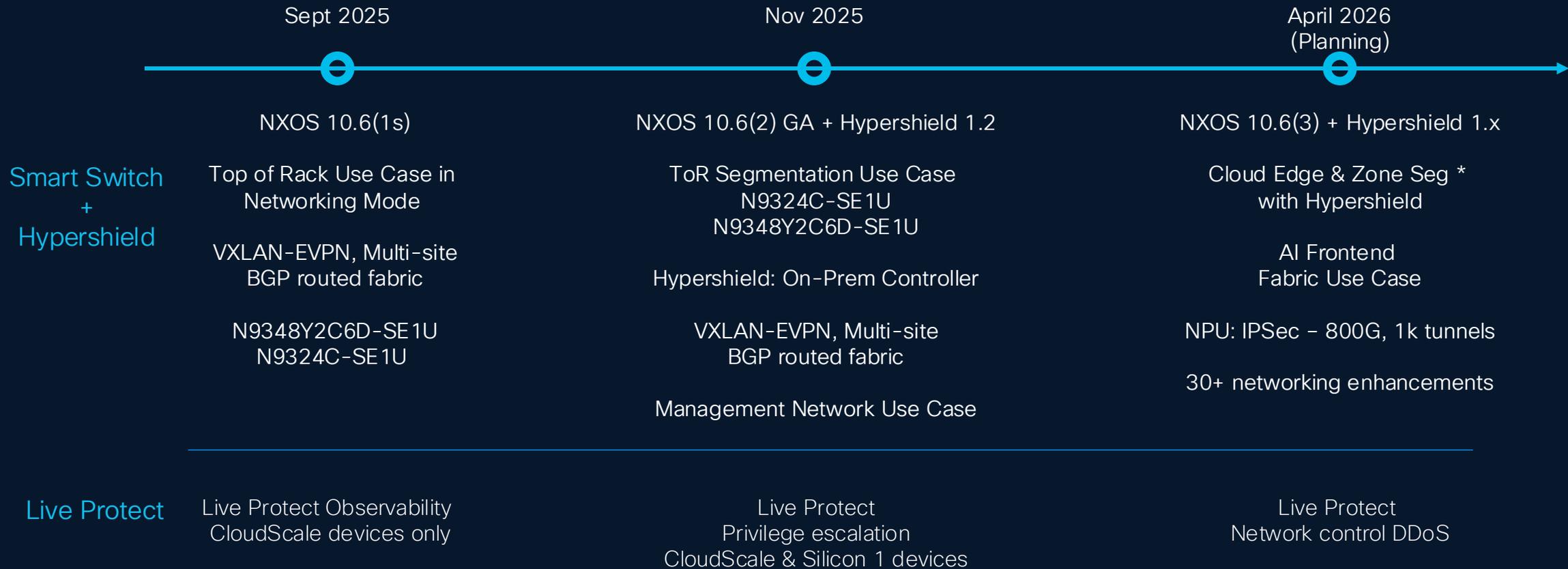
## Live Protect workflow:

- 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 1<sup>st</sup> to market
- Arista, Juniper, Aruba, etc ... don't have it
- CVE mitigation with no downtime
- Upgrades during regular maintenance window

# Roadmap



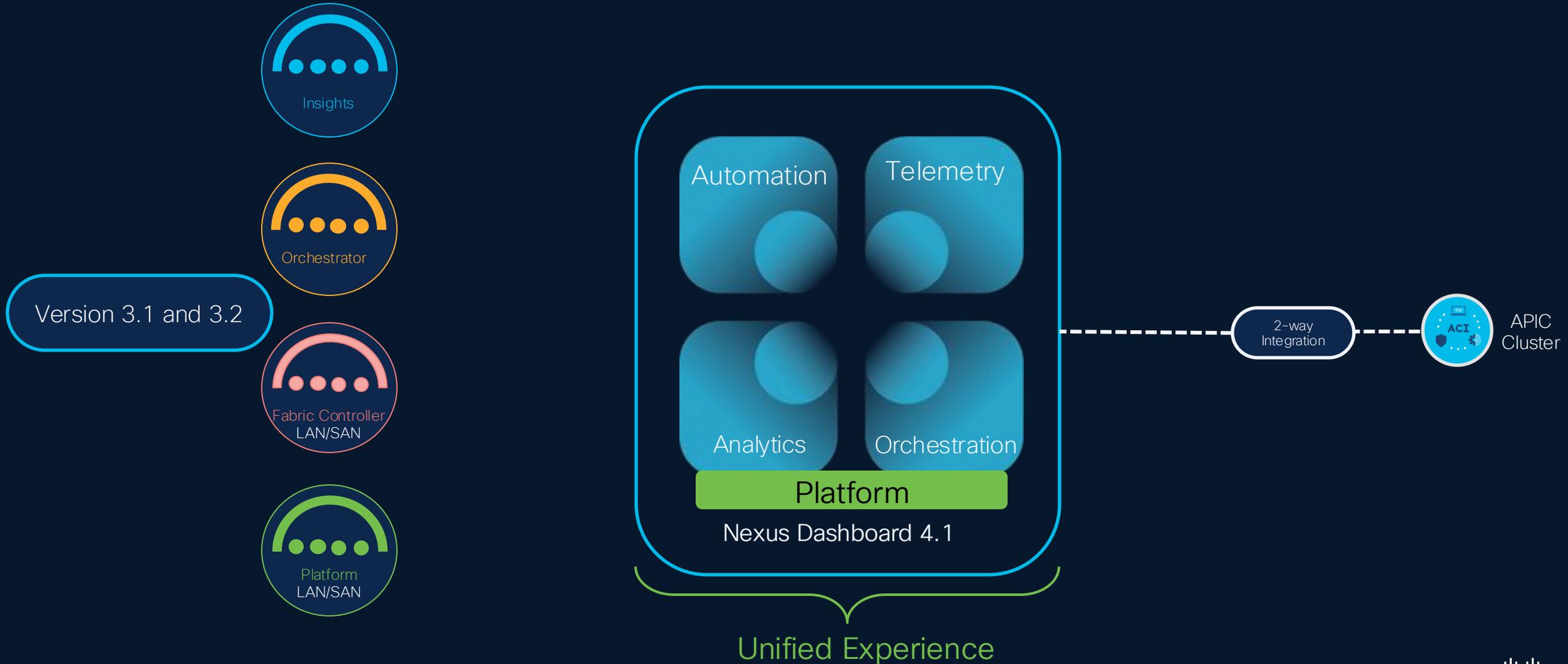
# AI Ready DC Network Operation

# Unified Cisco Nexus Dashboard

A consistent user experience  
for open networking



# Unify Your Operational Experience





# NEXUS DASHBOARD

Physical or virtual cluster

Data plane | Policy and control plane | Management plane

SAN

NX-OS

ACI

Campus

WAN  
Transport

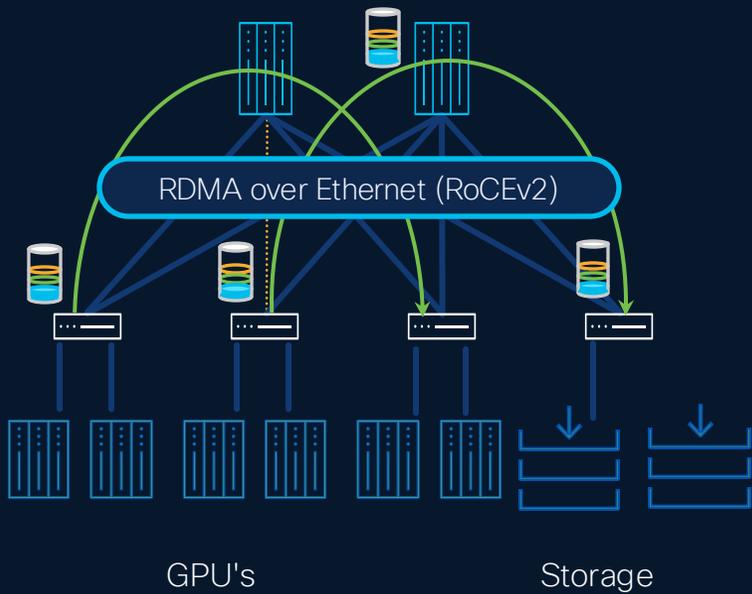
Classic Ethernet | VXLAN | IP Media Fabrics | AI/ML Fabrics

Virtual, K8  
3rd party Networks

# Cisco Nexus Dashboard

Automate your AI/ML network configurations

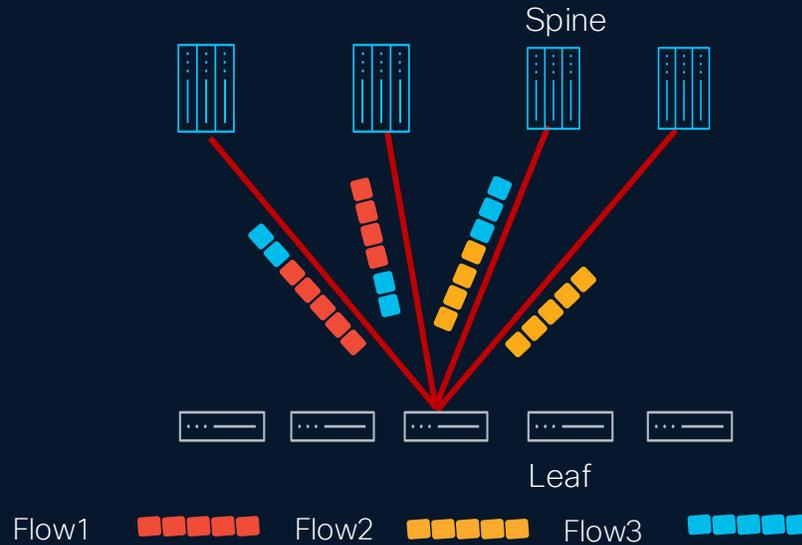
Shipping



Manage network congestion with Lossless Network (PFC + ECN)

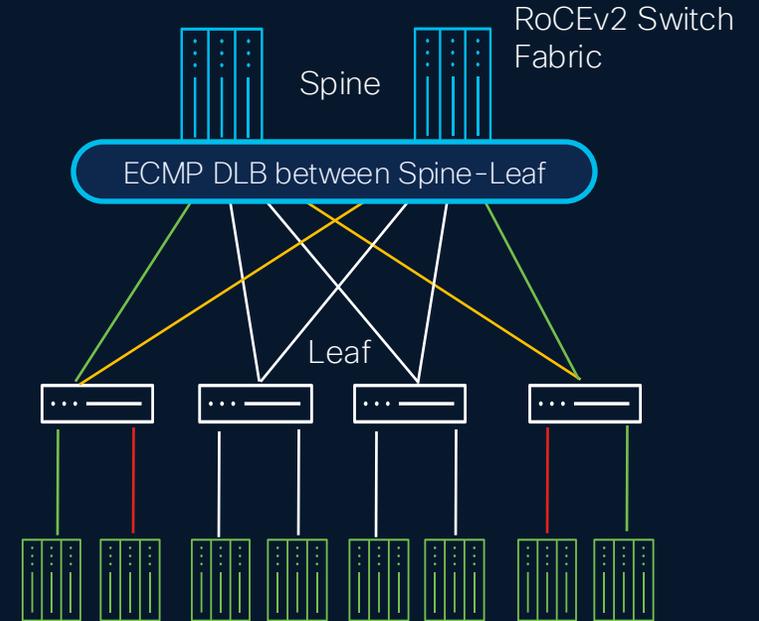
RoCEv2 support on Nexus 9800

## ECMP DLB between Spine-Leaf



Load balance flows/flowlets/Packets based on link utilization

Better hashing results in AI fabrics with uniform flow size and header information



Traffic efficiency through pinning rules  
Map traffic from each downlink to the desired uplink

Allows efficient selection of Spines for communication between leaf and spines

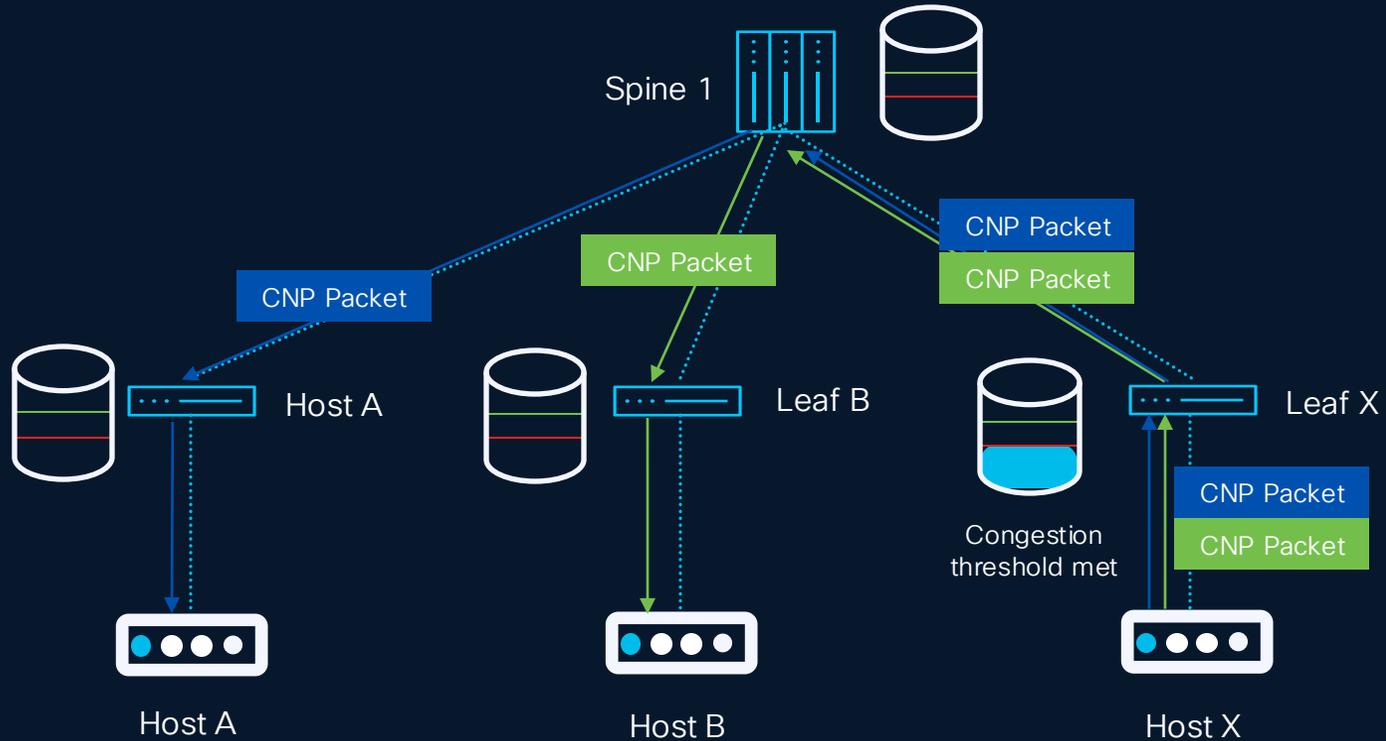
Benefit

Improved Congestion Management

# Cisco Nexus Dashboard

## AI/ML Fabric Settings

Committed



QoS metrics easily accessible with a new tab in fabric settings

Automated BGP pool allocation

Easily enable PFC/ECN and fine-tune thresholds

Enable Dynamic Load Balancing and set parameters

### Benefits

Quickly and accurately set QoS metrics at fabric creation

# Cisco Nexus Dashboard

Committed

## 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

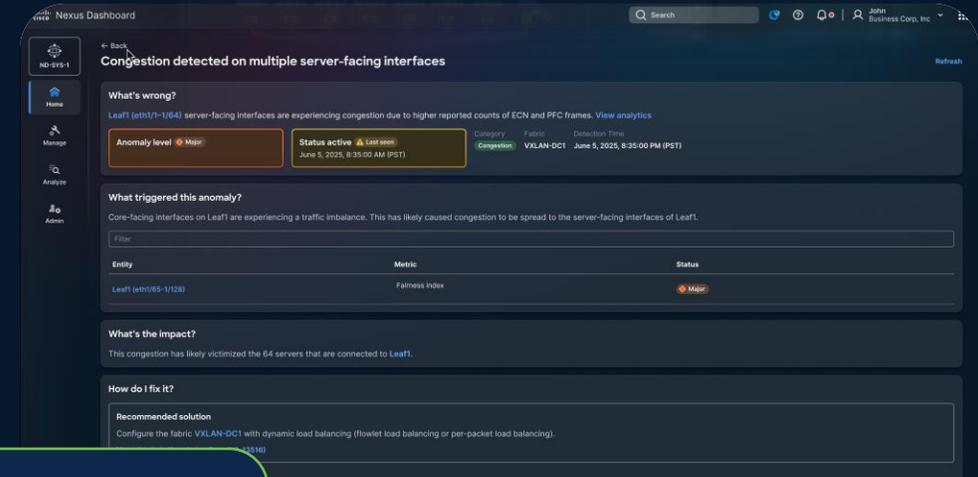
Detects congestion through detailed Priority Flow Control (PFC) patterns.

### 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.



## Benefits

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

# 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

## Security from ground to cloud

Protect distributed workloads by infusing security everywhere

## Seamless operations and observability

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

# Resources



**Cisco Data Center Networking**

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



**Cisco Nexus Dashboard**

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



**Cisco Nexus Hyperfabric**

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



**Cisco Secure AI Factory with NVIDIA**

Visit [Cisco.com](https://www.cisco.com)



**Blogs**

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



**Online community**

Visit the [Data Center and Cloud online community](https://www.cisco.com)



**CISCO** Connect

**Thank you**

