ALL – I sincerely apologize for the audio problems encountered throughout the session.

Below is a link to a recent OnDemand CiscoLive session by one of our principal engineers, Parash Gupta, that covers a lot of the same material, going a little farther into how the AI workloads interact with the network

[Why Cisco Nexus 9000 Series Switches for AI/ML Infrastructure Networking - BRKDCN-1226](https://www.ciscolive.com/on-demand/on-demand-library.html?#/session/1762465552174001eSAX)

https://www.ciscolive.com/on-demand/on-demand-library.html?#/session/1762465552174001eSAX

While this does not make up for your time spent, I hope that you will find the content in the OnDemand session valuable.

Data Center Networking for the AI Era

Marc Belle-Isle
DCN TSA | Cloud and AI Infrastructure

February 2026

About This Session

Marc Belle-Isle - DCN TSA

Cloud and AI Infrastructure

-Session Abstract-

AI training and inference clusters generate massive east-west traffic and require predictable, lossless packet delivery. This session covers Ethernet-based fabric design, congestion management, and observability. We'll also explore Cisco's latest innovations in Network Infrastructure (Silicon One, Nexus platforms, Optics, and Hyperfabric), Integrated Security (Live Protect and Smart Switches) and Unified Operations (Nexus Dashboard) for scalable, secure AI-ready networks.

Agenda

1. **Introduction**

   - AI Technology requirements and what "AI Ready" Means

   - Cisco AI fabric operational models

2. **AI Ready DC Network Infrastructure**

   - AI fabric design requirements

   - Hardware innovation – Cisco silicon, platforms, optics

   - Software innovation – NXOS feature / function for AI fabrics

3. **AI Ready DC Network Security**

   - Hardware integrated security – Cisco Smart Switches

   - Software integrated security – Cisco Live Protect

4. **AI Ready DC Network Operation**

   - Unified Nexus Dashboard

   - Operational integration(s)

Artificial Intelligence Outcomes Span Every Industry

Government

- Deliver enhanced citizen services
- Data-driven policy decisions and creation
- Modernizing and streamlining operations
- Optimizing infrastructure management
- AI-powered traffic design and public safety

Manufacturing

- Intelligent quality control
- Proactive machine maintenance
- Digital twin creation
- Supply chain optimization and tracking
- Optimizing production processes

Finance

- Predictive trading algorithms
- Fraud detection and prevention
- Personalized financial advice
- Investment portfolio optimization
- Virtual assistants and seamless transaction experiences
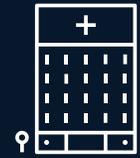
Healthcare

- Medical imaging analysis
- Enhance diagnosis and treatment
- Patient management through predictive analytics
- Improved access to healthcare with remote monitoring tools
- Drug research and development
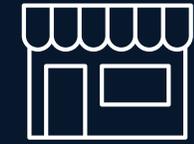
Retail

- Consumer behavior analytics
- Enhanced customer experiences
- Personalized product recommendations
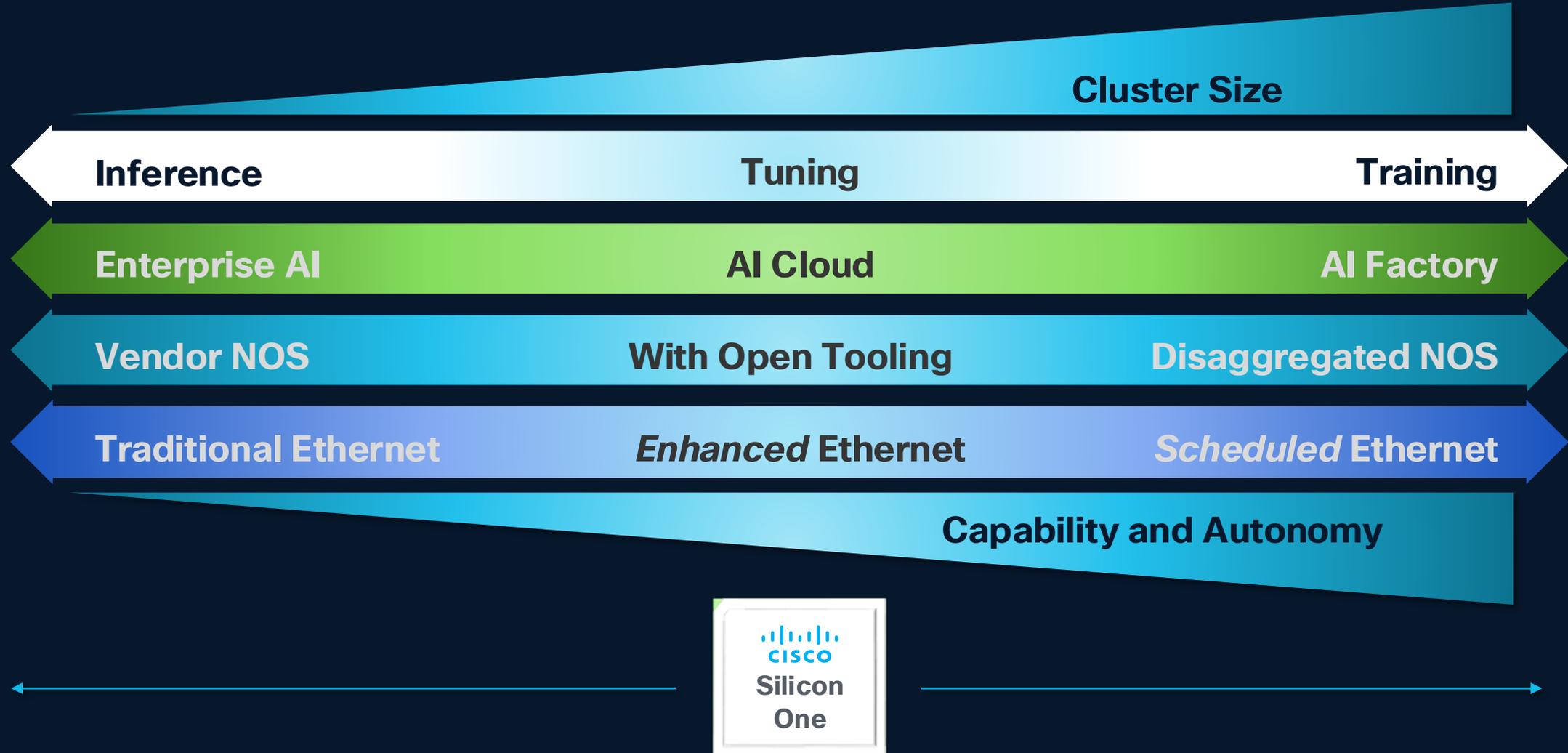- Dynamic virtual shopping experience
- Demand analysis and prediction

Build the model | **Training**

Optimize the model | **Fine-tuning and RAG**
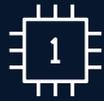
Use the model | **Inferencing**

BRKNWT-2508

CISCO

Where does Ethernet fit across AI Landscape?
An Oversimplified Clustering of AI Technology Requirements

Cluster Size

Inference	Tuning	Training
Enterprise AI	**AI Cloud**	**AI Factory**
Vendor NOS	**With Open Tooling**	**Disaggregated NOS**
Traditional Ethernet	*Enhanced* **Ethernet**	*Scheduled* **Ethernet**

Capability and Autonomy

cisco
**Silicon
One**

Cisco AI Networking

Powering AI at Every Scale, Everywhere

Silicon + **Systems** + **Software** + **Operating model**

1 Accelerated time-to-value through streamlined infrastructure

2 Maximized resource utilization means improves performance while reducing costs

3 Enhanced operational efficiency with simplified management and troubleshooting

4 Future-ready infrastructure that scales with evolving AI requirements

Cisco AI PODs Solution Options

Choose Based on AI Use-Case and Operational Model

Nexus Dashboard
On-premises managed network

Compute — **Cisco UCS**
NVIDIA HGX, MGX or
RTX PRO 6000 Servers

Networking — **Cisco Nexus 9300**
w/ Silicon One
w/ Cloud Scale
w/ NVIDIA Spectrum-X

Storage — **NetApp | Pure Storage |**
VAST Data | Hitachi Vantara

Training, fine-tuning, inference

Nexus 9300 managed with Nexus Dashboard
NVIDIA aligned Enterprise Reference Architecture
Flexible design consistent performance at any scale
Dedicated AI or multi-purpose DC

Nexus Hyperfabric
Cisco-managed network cloud controller

Compute — **Cisco UCS**
NVIDIA HGX, MGX or
RTX PRO 6000 Servers

Networking — **Cisco Nexus Hyperfabric**
w/ Silicon One

Storage — **NetApp | Pure Storage |**
VAST Data | Hitachi Vantara

Inference

Cloud-managed fabric-as-a-service
Deploy anywhere, easy to scale
BYO servers, GPUs, storage, software stack
Multi-purpose DC with AI & non-AI services

Nexus Hyperfabric AI
Cisco-managed network cloud controller

Compute — **Cisco UCS**
NVIDIA HGX (885a with 8x H200)
Future: RTX PRO Server

Networking — **Cisco Nexus Hyperfabric**
w/ Silicon One

Storage — **VAST Data**

Training, fine-tuning, inference

Full stack: network, servers, GPU, storage
NVIDIA Enterprise Reference Architecture compliant
Blueprints encompass the entire cluster
Monitoring spans network to compute NIC

Customizable ←——————————————————————→ *Prescriptive*

CISCO

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

9

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
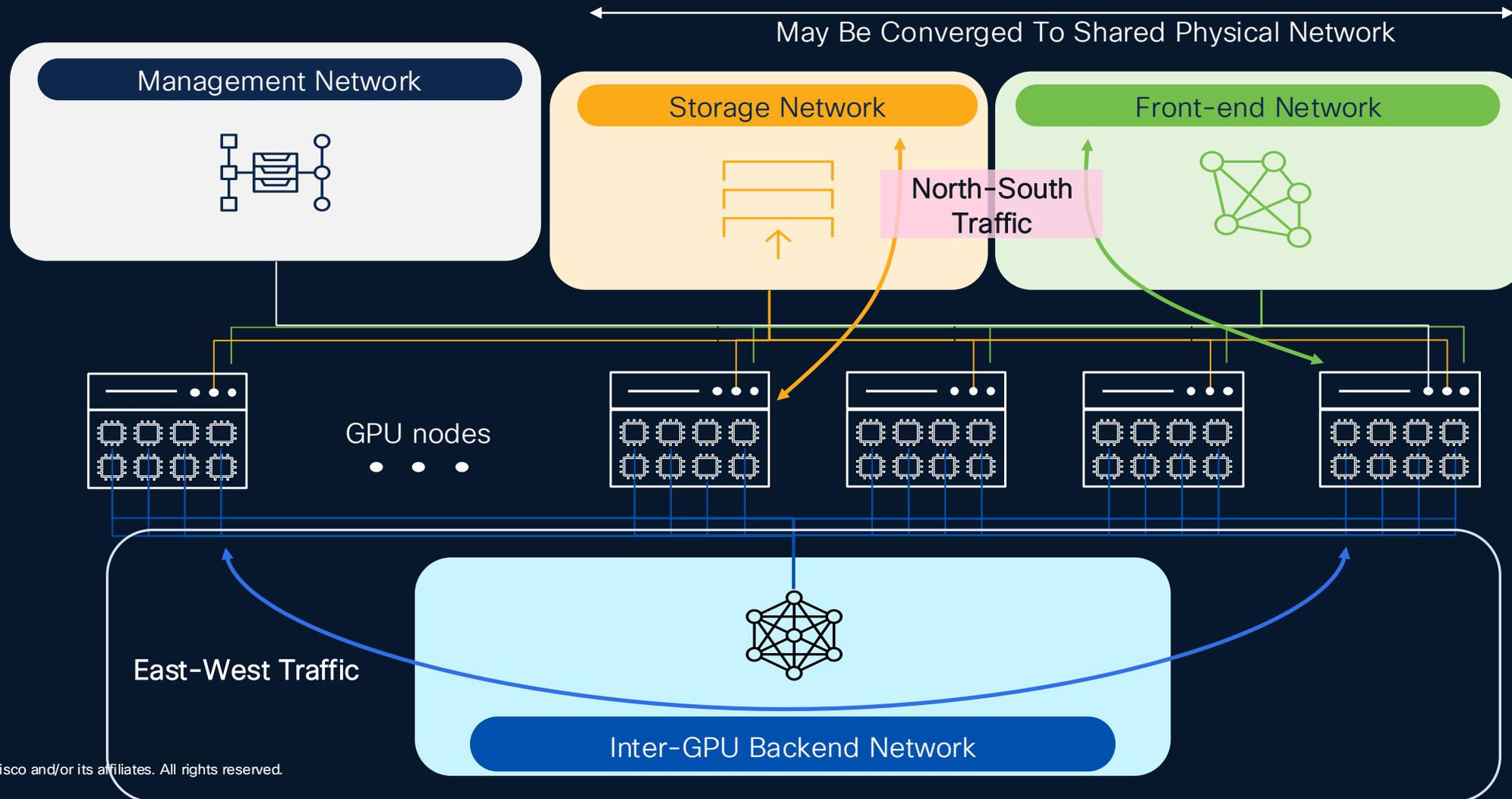
Nexus Dashboard for simplified ops

Nexus 9000 series switches

Functional Networks
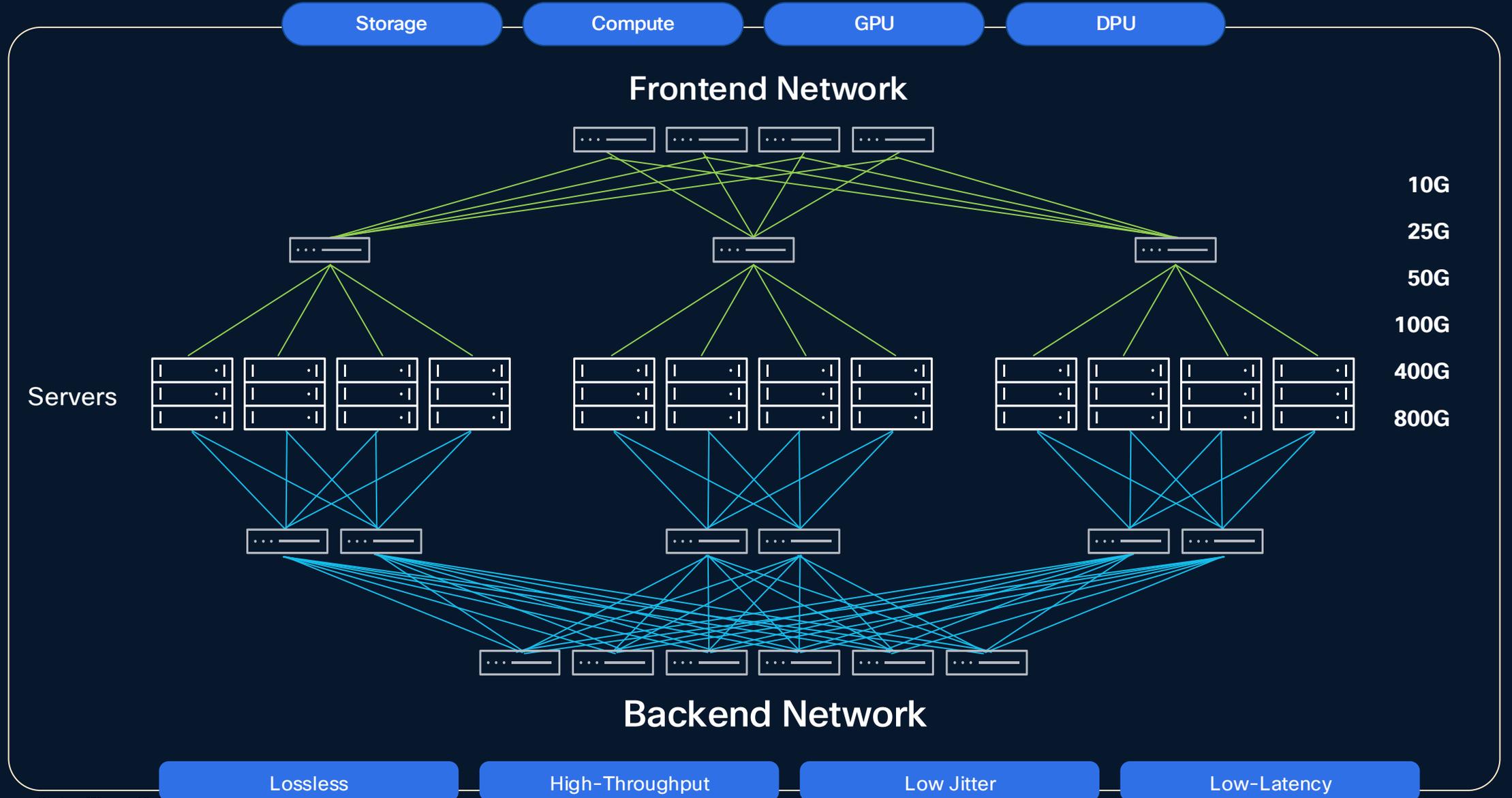## Front End | Storage | Back End (GPU) | Mgmt/OOB

North-South Traffic · East-West Traffic

May Be Converged To Shared Physical Network

Management Network

Storage Network

North-South Traffic

Front-end Network

GPU nodes

East-West Traffic

Inter-GPU Backend Network

**BRKDCN-1226*

CISCO

Cisco AI Networking Infrastructure

Storage   Compute   GPU   DPU

Frontend Network

10G

25G

50G

100G

Servers

400G

800G

Backend Network

Lossless   High-Throughput   Low Jitter   Low-Latency

Network Topology Design

Three-Tier Hierarchical

Inherent scale up topology building blocks
Requires multiple independent protocols configured in
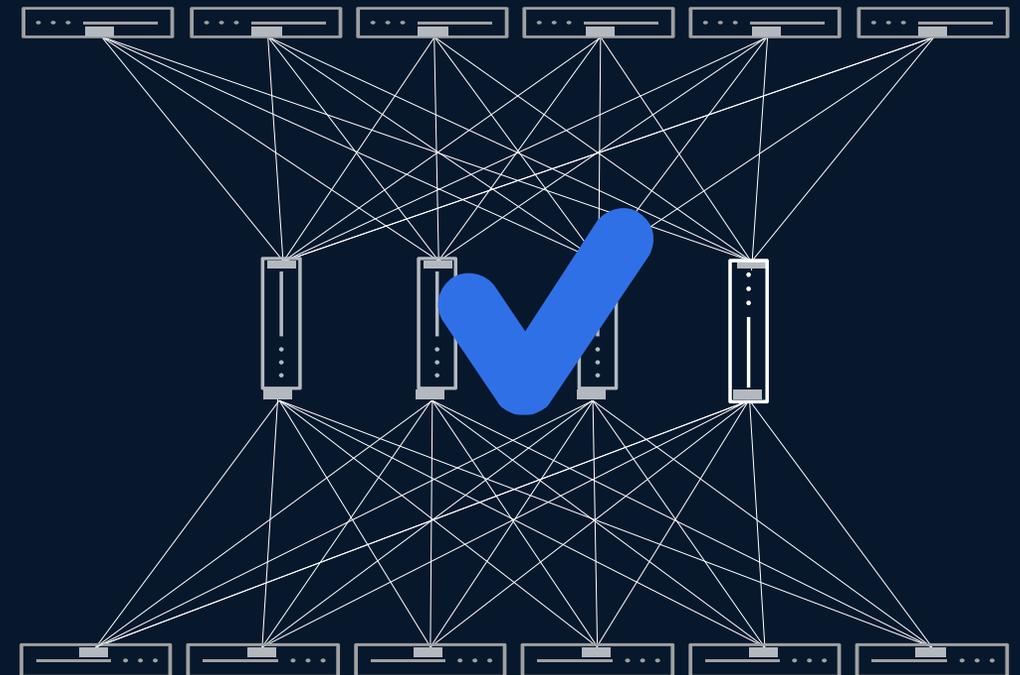concert to be able to use all network links

vPC

FHRP
HSRP|VRRP

STP
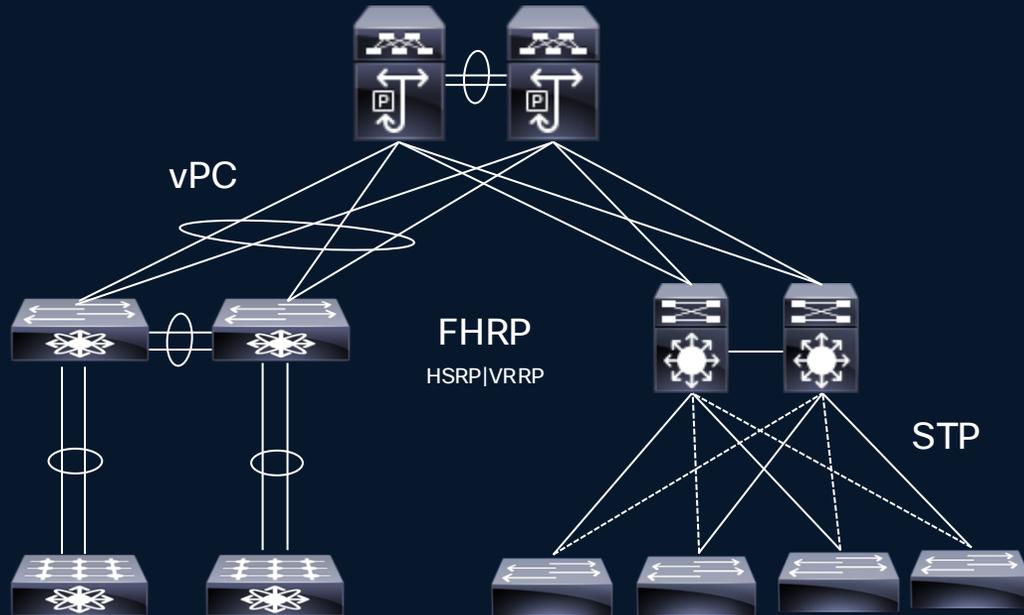
*"I always thought Ethernet forwarding with STP was a kludge, and the
right solution was to do layer 3 forwarding..."*
~Radia Perlman

Clos Fabric

Fully Interconnected 'fabric' with multiple design variations
Inherent modular scale out design (intra and inter-fabric)

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

[Number of Downlink Ports * Speed] ÷ [Number of Uplink to Spine * Speed] : 1

Uplink Ports
4x100Gbps = 400 Gbps

Downlink Ports
48x25Gbps = 1.2Tbps

Oversubscription Ratio

3:1
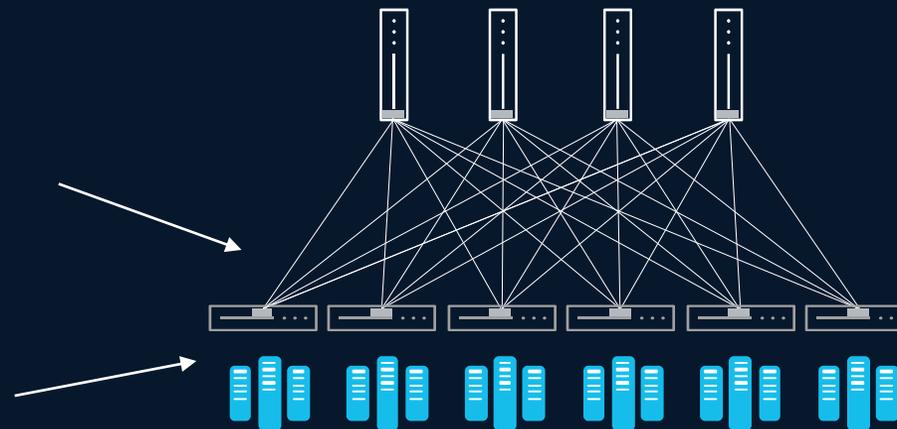
Cisco ASIC Innovations – Nexus 9000 Series

Ultra Ethernet Consortium
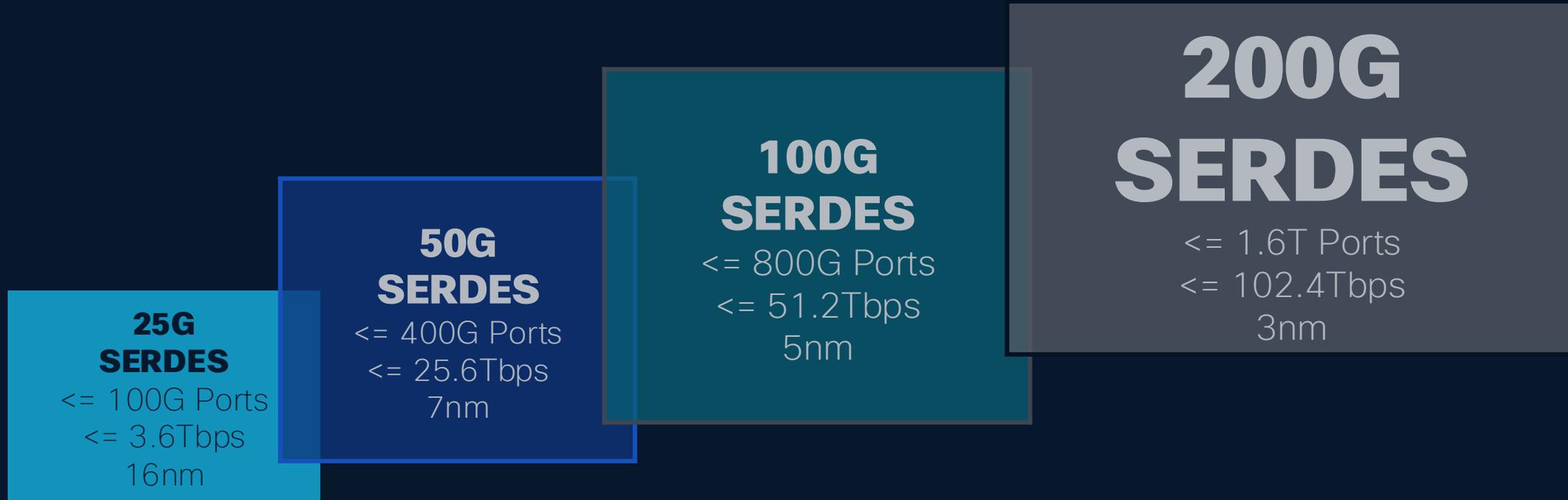
READY

Fully Programmable Multi-Slice Architecture

Features & Performance | Deep Visibility & Observability | Secure Connectivity | Low Latency & Power Efficiency | Modular & Fixed Switches

25G SERDES
<= 100G Ports
<= 3.6Tbps
16nm

50G SERDES
<= 400G Ports
<= 25.6Tbps
7nm

100G SERDES
<= 800G Ports
<= 51.2Tbps
5nm

200G SERDES
<= 1.6T Ports
<= 102.4Tbps
3nm

Cloud Scale | Cisco Silicon One

Session ID: CENDCN-1399

CISCO

Hardware Terminology – Quick Reference
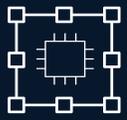
Die / Package

A Die is a block of semiconducting material (silicon) on which an integrated circuit is fabricated

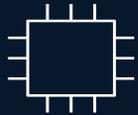Encapsulated / sealed die(s) are referred to as a package

ASIC

Application Specific Integrated Circuit

Sometimes referred to as *'silicon' (Cisco vs Merchant)*

SoC

Switch on Chip / System on Chip

Multiple interconnected dies (chiplets) on a single package – Used in modern ASIC design (ex Cisco CloudScale)

Slice

Self-contained forwarding complex with ingress and egress forwarding pipelines on a single ASIC

Think of them as fully interconnected (no bus/arbitration) line cards within an ASIC

On-Die Buffer and HBM

- Buffer integrated on the silicon die
- High Bandwidth Memory is external buffer memory interconnected to the ASIC

SFP | QSFP | QSFP-DD

- [Quad] Small Form Factor Pluggable – [Double Density]
- Compact (standards based) transceiver

SerDes

- Serializer + Deserializer - Converts serial data to parallel and vice versa
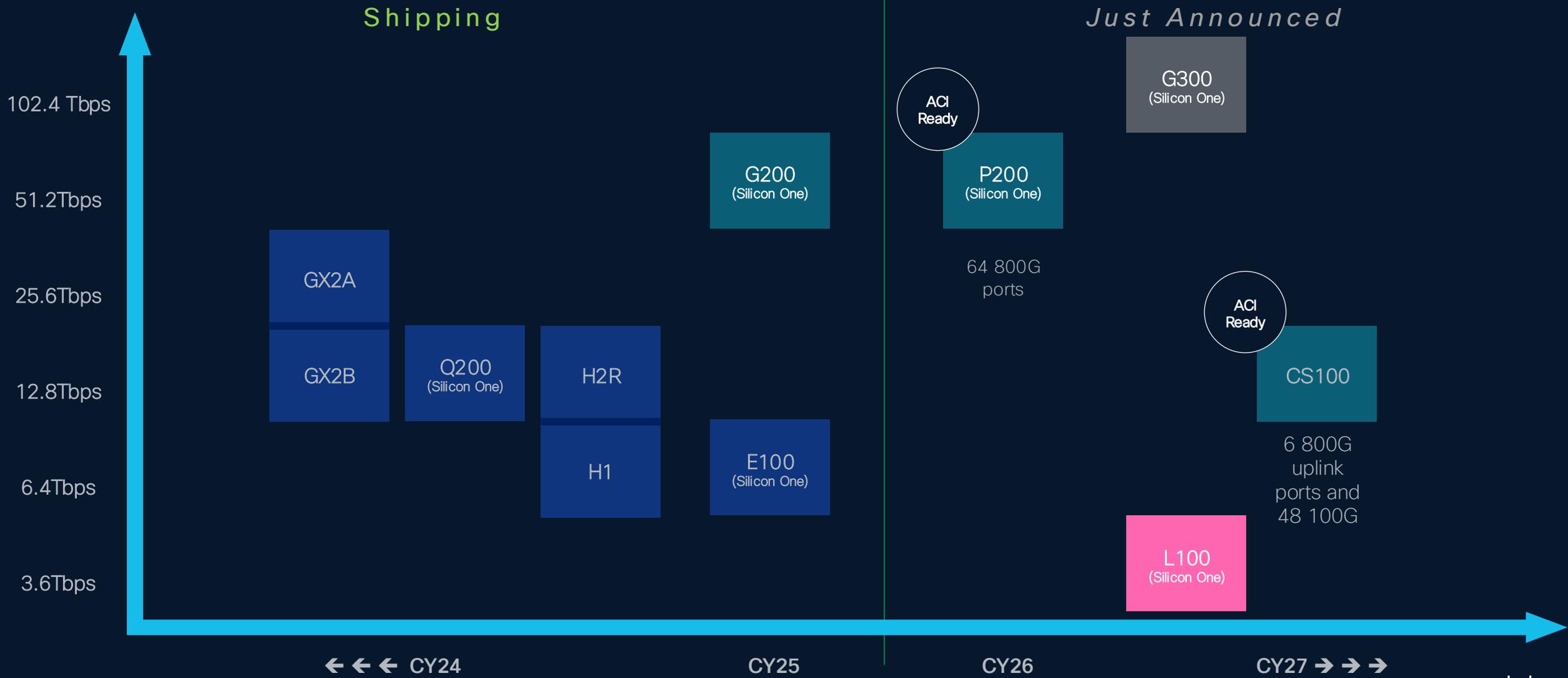- Clocking speed increases result in higher BW

MACSEC | CloudSEC

- MACSEC is line rate encryption of a physical link / port
- CloudSEC is effectively multi-hop or tunneled MACSEC (TEP<->TEP)

Cisco Live OnDemand Library – Search for "Nexus 9000 Architecture"

Cisco Nexus 9000 ASICs

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

Shipping

Just Announced

Y-axis (Tbps): 102.4 Tbps, 51.2 Tbps, 25.6 Tbps, 12.8 Tbps, 6.4 Tbps, 3.6 Tbps

X-axis: ← ← ← CY24, CY25, CY26, CY27 → → →

- GX2A
- GX2B
- Q200 (Silicon One)
- H2R
- H1
- E100 (Silicon One)
- G200 (Silicon One)
- P200 (Silicon One) — ACI Ready — 64 800G ports
- G300 (Silicon One)
- CS100 — ACI Ready — 6 800G uplink ports and 48 100G
- L100 (Silicon One)

Nexus 9K Platform Naming Convention Reference

Nexus 9K Switches adhere to the following naming convention in general

Reference Example – 93180YC-FX3

The Numbers (ex. 9|3|180)

(9)   First number = Switch make

(3)   Second number = Switch model / family

(180)   Third number = Max aggregate throughput in GB **-or-** Port count if all same speed (ex. 9336C-FX2)

The Letters (ex. YC-FX3)

(YC) First letter group designates the switch face plate / port connections (speed and inherent connectivity)

   E: Native 800-Gbps front-panel ports (QSFP-DD)

   D: Native 400-Gbps front-panel ports (QSFP-DD)

   Q: Native 40-Gbps front-panel ports (QSFP)

   Y: Native 25-Gbps front-panel ports (SFP+/SFP28)

   C: Native 100-Gbps front-panel ports (QSFP)

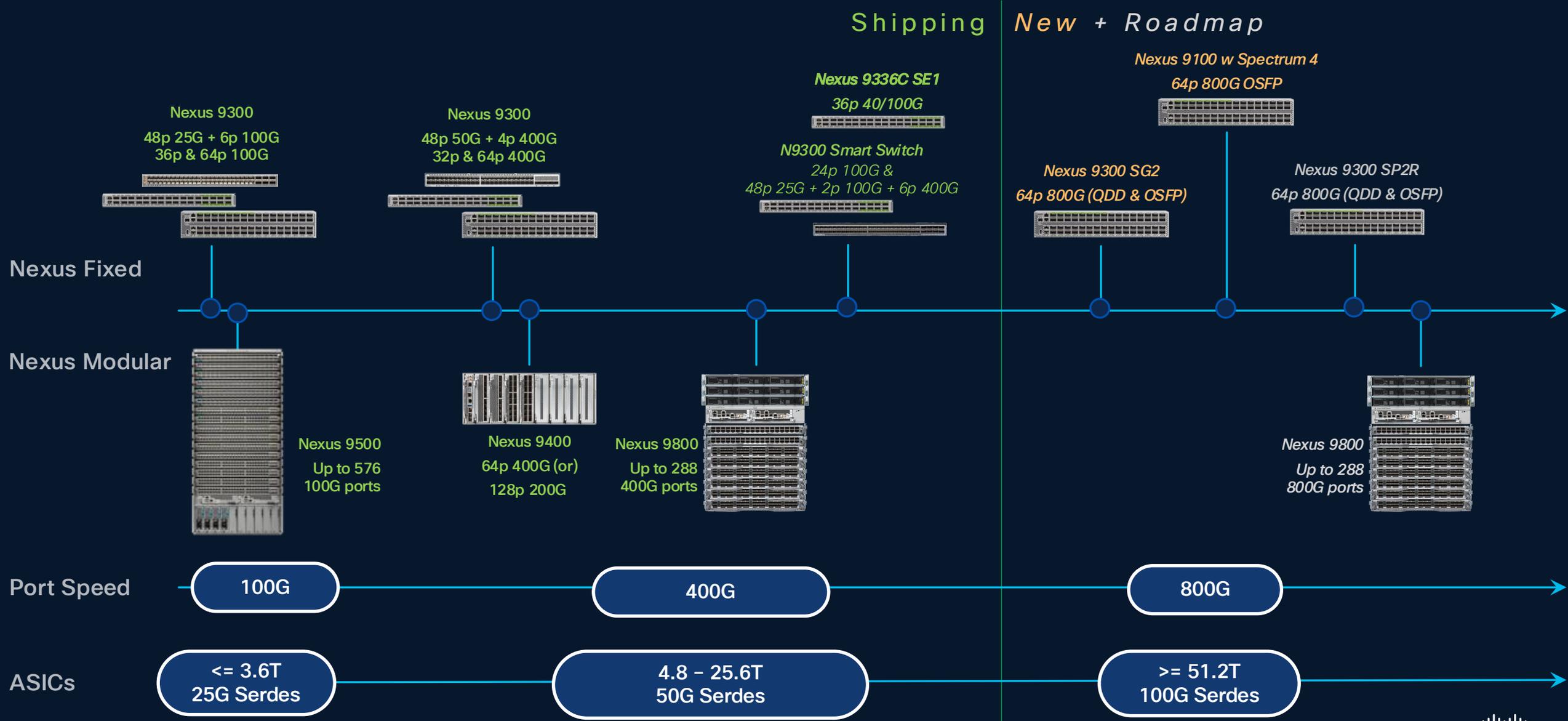   L: Native 50-Gbps front-panel ports (QSFP)

   T: 100M*, 1GT, and 10GT (Base-T – RF45)

   G: 100M, 1G (Base-T – RJ45)

   P: Power Over Ethernet (UPoE) capable

(FX3) Second letter group designates which ASIC is used (EX, FX, FX2, FX3, GX etc..)

Cisco Nexus 9000 Series Switches

Nexus 9300
48p 25G + 6p 100G
36p & 64p 100G

Nexus 9300
48p 50G + 4p 400G
32p & 64p 400G

Nexus 9336C SE1
36p 40/100G

N9300 Smart Switch
24p 100G &
48p 25G + 2p 100G + 6p 400G

Nexus 9100 w Spectrum 4
64p 800G OSFP

Nexus 9300 SG2
64p 800G (QDD & OSFP)

Nexus 9300 SP2R
64p 800G (QDD & OSFP)

Nexus Fixed

Nexus Modular

Nexus 9500
Up to 576
100G ports

Nexus 9400
64p 400G (or)
128p 200G

Nexus 9800
Up to 288
400G ports

Nexus 9800
Up to 288
800G ports

Port Speed

100G	400G	800G

ASICs

<= 3.6T 25G Serdes	4.8 – 25.6T 50G Serdes	>= 51.2T 100G Serdes

Cisco Nexus 9300 64 Port 800G Switch

512-wide radix | **Fully shared packet buffer** | **Advanced load balancing** | **Low Latency**

Nexus 9364E-SG2

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
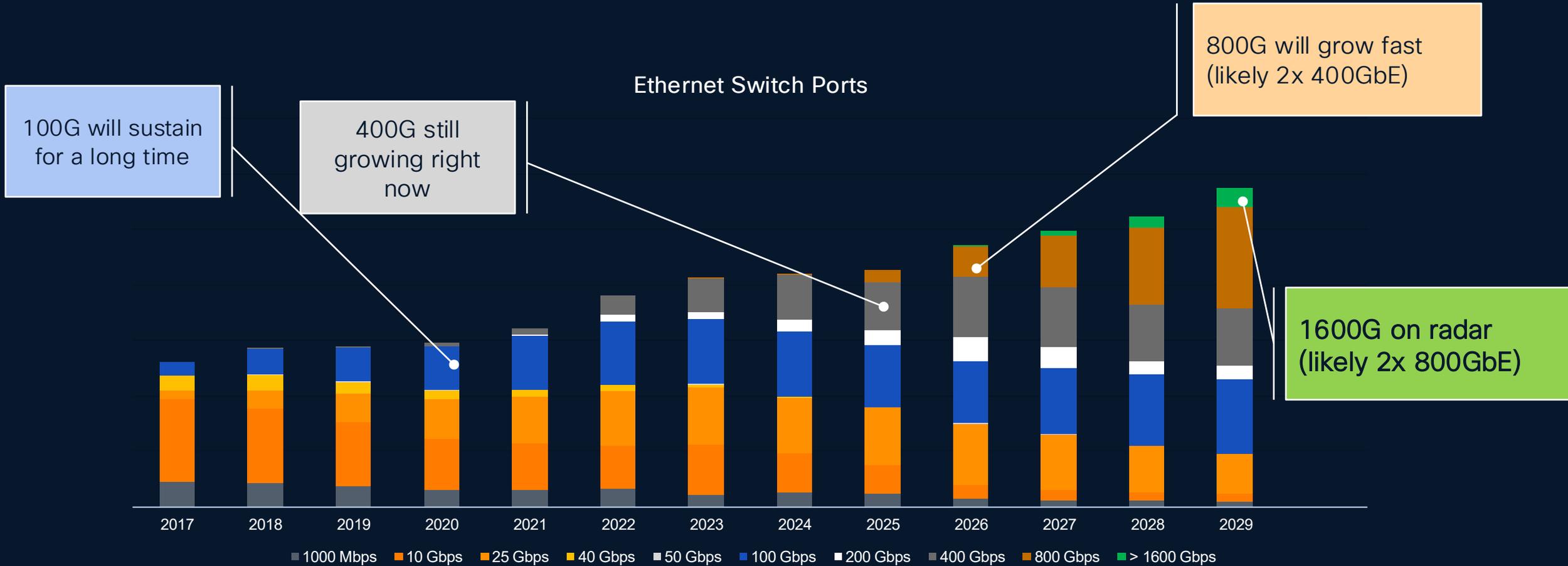
Cisco NXOS spine and AI/ML spine/leaf capable

N9364E-SG2-Q or N9364E-SG2-O

27C | 50% load | 64x DAC	995W
27C | 100% load | 64x DAC	1,186W
27C | 100% load | 64x 16W optics	2,332W
40C | 100% load | 64x 16W optics	2,567W

Ultra Ethernet Consortium
READY

Session ID: CENDCN-1399

Ethernet Data Center Switch Port Speed Transitions

Ethernet Switch Ports

100G will sustain for a long time

400G still growing right now

800G will grow fast (likely 2x 400GbE)

1600G on radar (likely 2x 800GbE)

Years: 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029

Legend: ■ 1000 Mbps ■ 10 Gbps ■ 25 Gbps ■ 40 Gbps ■ 50 Gbps ■ 100 Gbps ■ 200 Gbps ■ 400 Gbps ■ 800 Gbps ■ > 1600 Gbps

- Majority of the highest speed transitions are webscale (top 8) customers
- Webscale will drive the speed transitions quickly to scale
- Rest of the market will leverage that scale in their own time frame

Source: Dell'Oro's Ethernet Switch – Data Center Five Year Forecast Report 2025 – 2029

Ethernet Speed Transitions in AI Networks

Ethernet Ports for Front End Network

| 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 |

■ < 100 Gbps ■ 200 Gbps ■ 400 Gbps ■ 800 Gbps ■ => 1600 Gbps

Ethernet Ports for Back End

| 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 |

■ < 100 Gbps ■ 200 Gbps ■ 400 Gbps ■ 800 Gbps ■ => 1600 Gbps
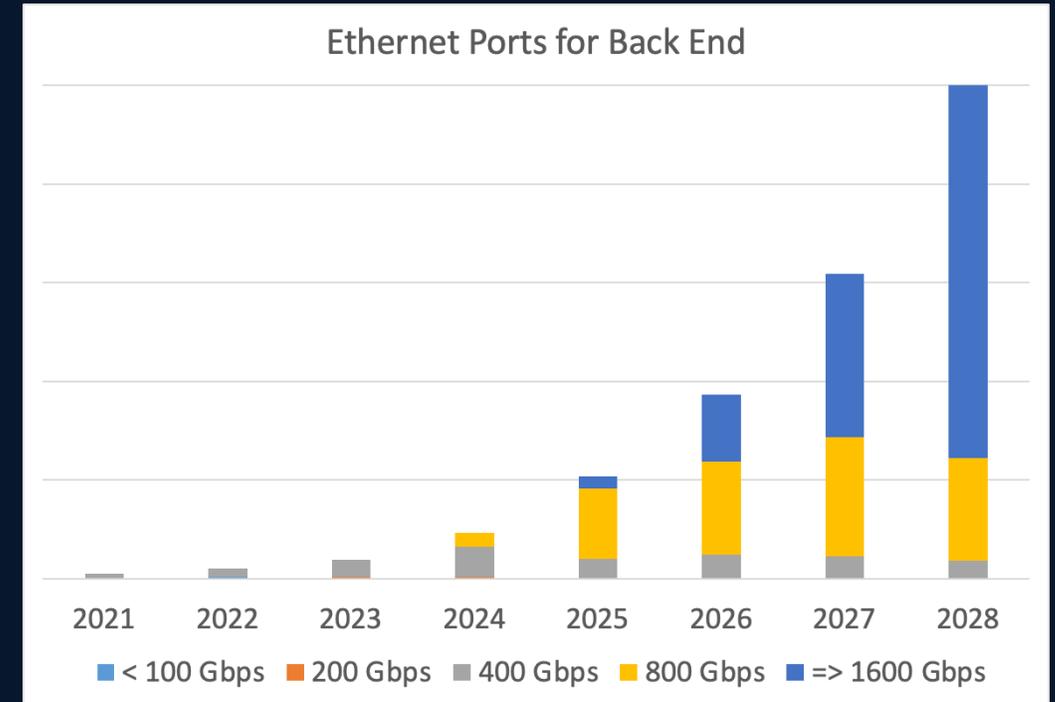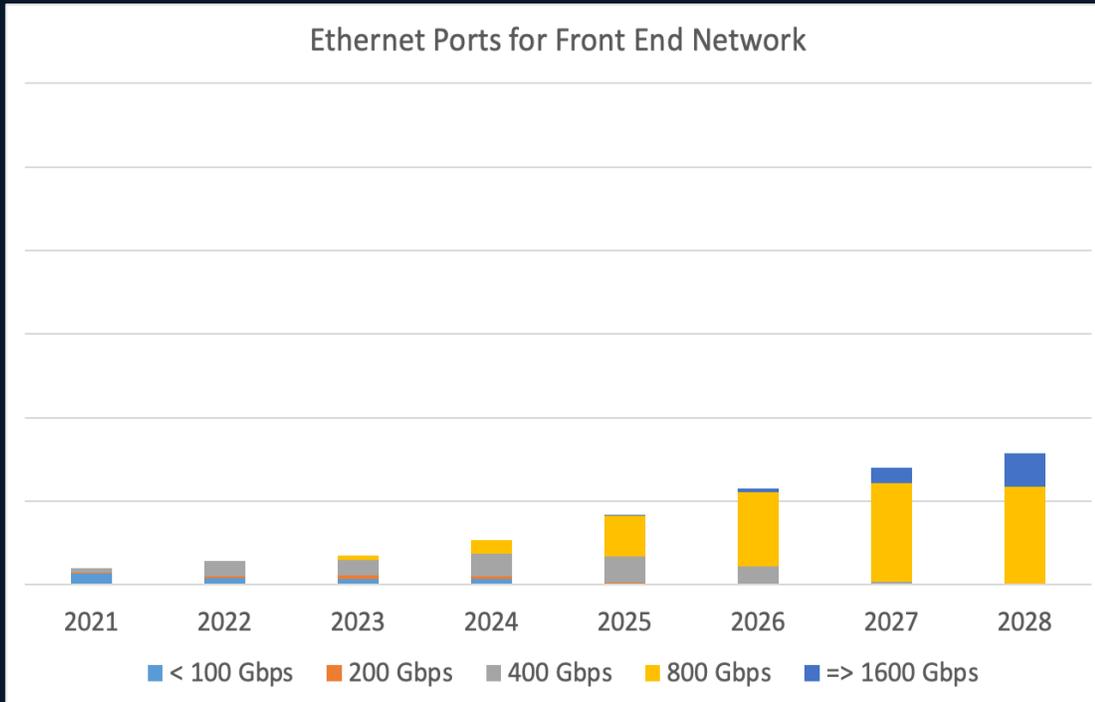
Majority of the switch ports in AI back-end Networks to be 800 Gbps in 2025 and 1600 Gbps in 2027, showing a very fast migration to the highest speeds available in the market.

Dell'Oro: 95N01_Advanced_Research_AI_Networks_For_AI_Workloads_Report_3Q24
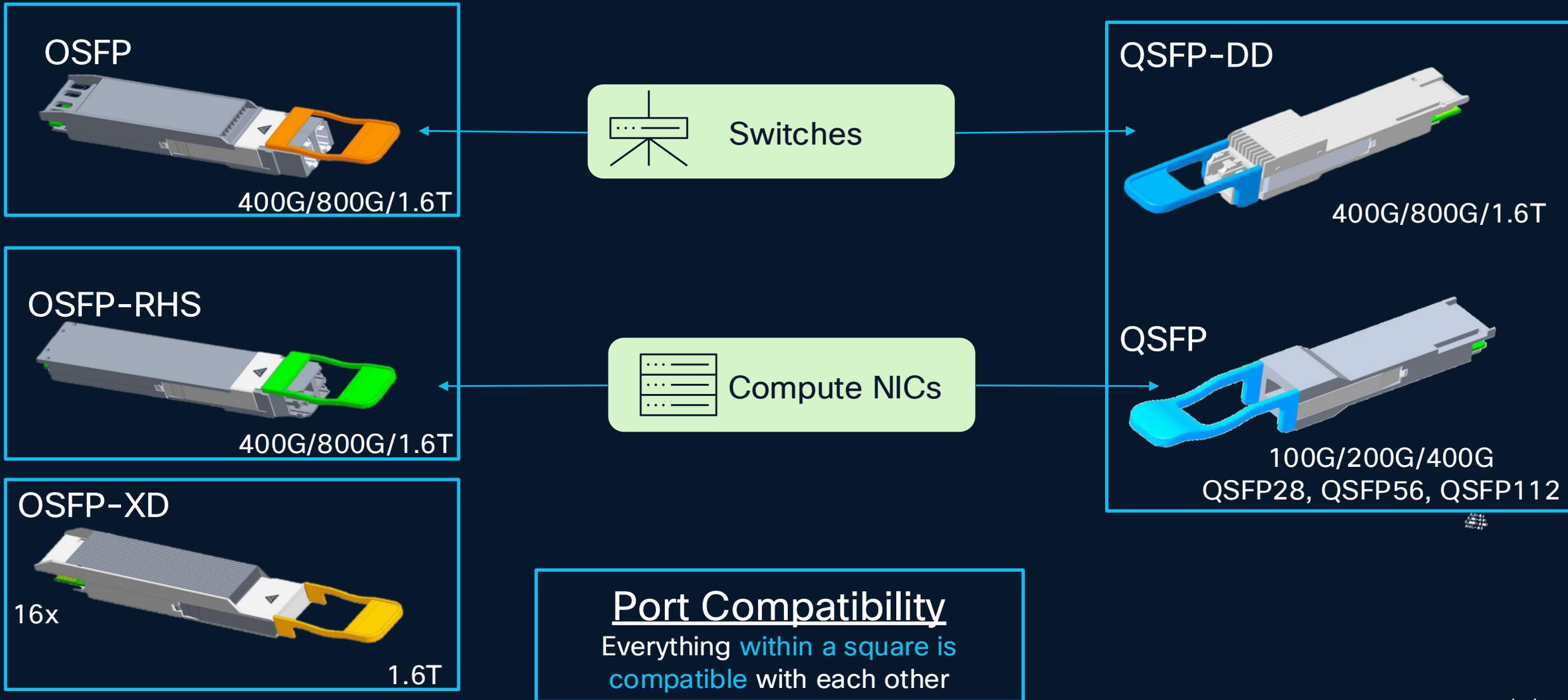
Pluggable Optical Modules Options
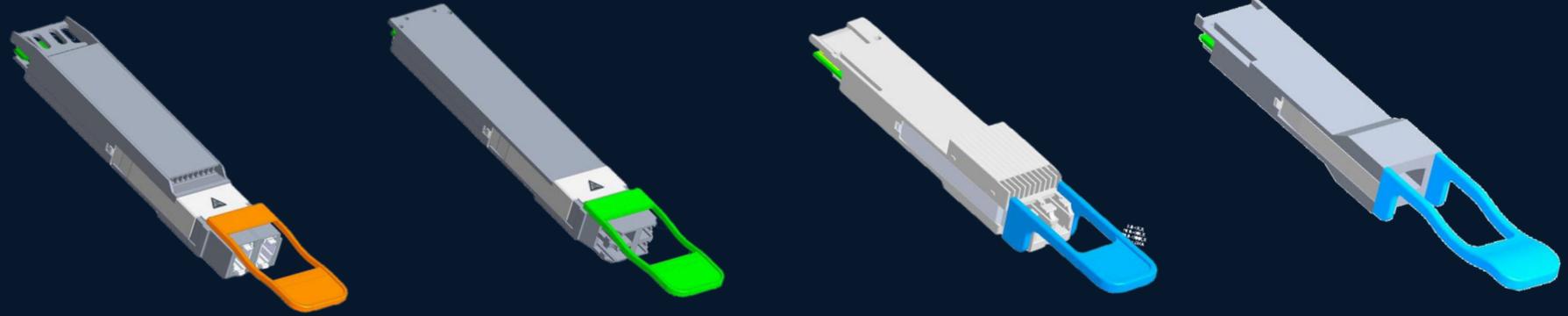
	OSFP	OSFP-RHS	QSFP-DD	QSFP
Host high-speed pins	8	8	8	4
Module capacity (Gbit)	400, 800, 1600	400, 800, 1600	400, 800, 1600	40, 100, 200, 400
Cooling	Integrated heatsink	Riding heatsink	Riding heatsink	Riding heatsink
High-speed IO Compatibility (Gbps)	50-100-200	50-100-200	50-100-200	1−25-50-100
Port compatibility	OSFP	OSFP-RHS	QSFP-DD & QSFP	
Management	CMIS			

System and ASIC design dictate which modules are supported and speed backwards compatibility
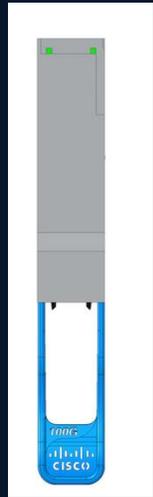
Pluggable form factors – a wealth of options

OSFP

400G/800G/1.6T

OSFP-RHS

400G/800G/1.6T

OSFP-XD

16x

1.6T

Switches

Compute NICs

QSFP-DD

400G/800G/1.6T

QSFP

100G/200G/400G

QSFP28, QSFP56, QSFP112

Port Compatibility
Everything within a square is compatible with each other

QSFP-DD Extends to 1.6T
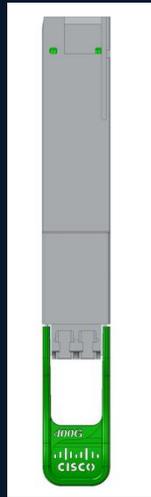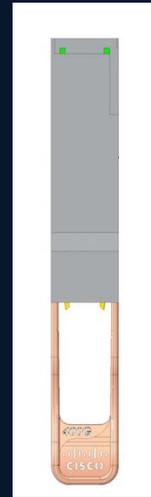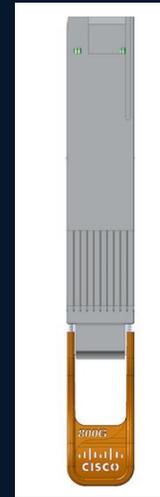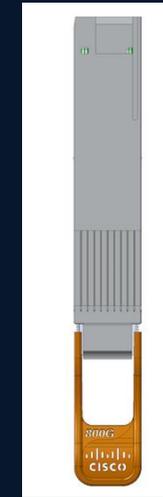
100 Gb/s	400 Gb/s	400 Gb/s	800 Gb/s	1600 Gb/s
25 Gb/s SerDes	50 Gb/s SerDes	100 Gb/s SerDes	100 Gb/s SerDes	200 Gb/s SerDes
QSFP28	**QSFP56-DD**	**QSFP112**	**QSFP-DD800**	**QSFP-DD1600**

← → Complete backwards compatibility

Powerful value proposing to enable seamless network growth and investment protection. Technically superior solution.

Cisco High-Density Fiber Patch Panel

Structured cabling and breakout cassettes

Key features:
Back of the Panel

2RU Panel has 36 Rear MPO12 Ports which enables
1:1 mapping to Cisco's 36 port line cards

Color Coded
Fiber Mapping Diagram: MPO to LC

3 X 8F MPO per Cassette - Angled ports
to reduce bend stress

Rear Access
Door latch

Key features:
Panel Trays

Port Identification Label

Laser Etched Port Numbering

Color Coded
Spreadable Duplex Adapters

Unparalleled accessibility using spreadable adapter technology

Each Patch Panel tray comes equipped with
2 Fiber management sleeves
Fiber Panel can support:
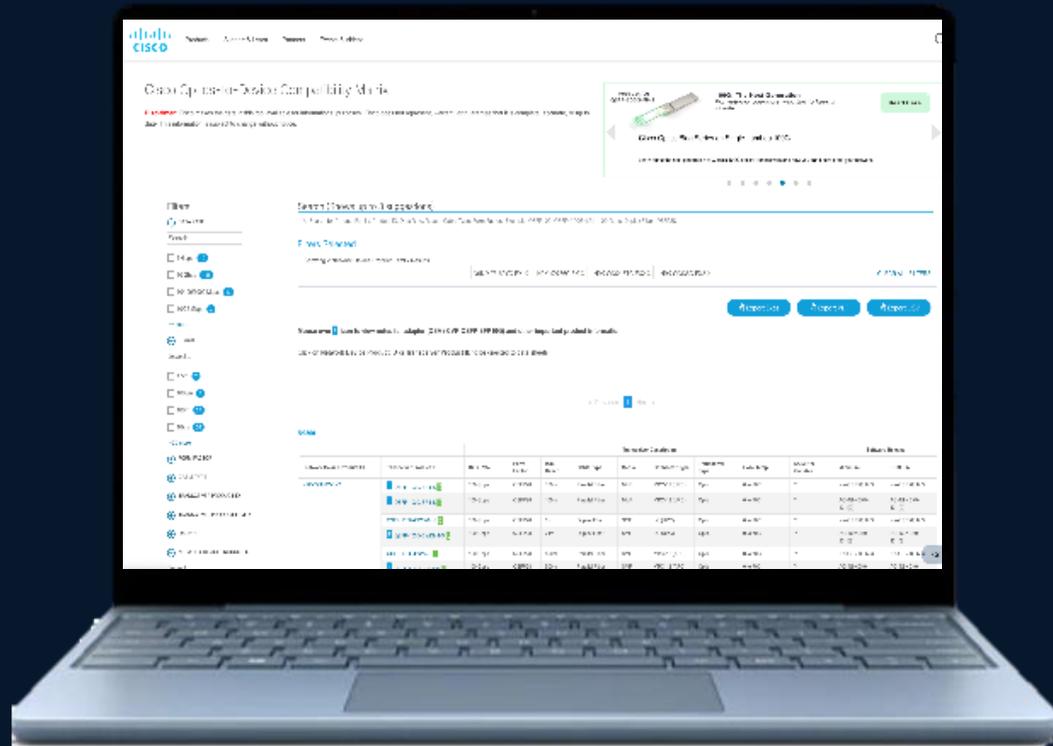1.2mm, 1.6mm, 2.0mm jumpers

1RU == 72 LC Ports (18 MPO)
2RU == 144 LC Ports (36 MPO)
3RU == 216 LC Ports (54 MPO)

64p 400G + Dedicated 2RU Patch Panel
144p 100G LC MMF + 28 400G MPO-12 MMF

TMG Matrix – Authoritative Optics Compatibility



https://tmgmatrix.cisco.com/

- Comprehensive list of Cisco hardware

- Cisco validated compatibility
  - Optics <-> Platform <-> SW release

- Responsive support team

- Continuously maintained to ensure up to date information

- Search by HW platform and/or Optic name

  - Cisco tested and validated Optic compatibility

  - Optics specs & platform software release mapping

  - 1-click to datasheet

Powering AI Fabrics with Cisco Intelligent Packet Flow

Intelligent Load Balancing for every path and every flow

Ultra Ethernet
Consortium
READY

Spine   Spine   Spine

NEW'ish

Leaf   Leaf   Leaf   Leaf
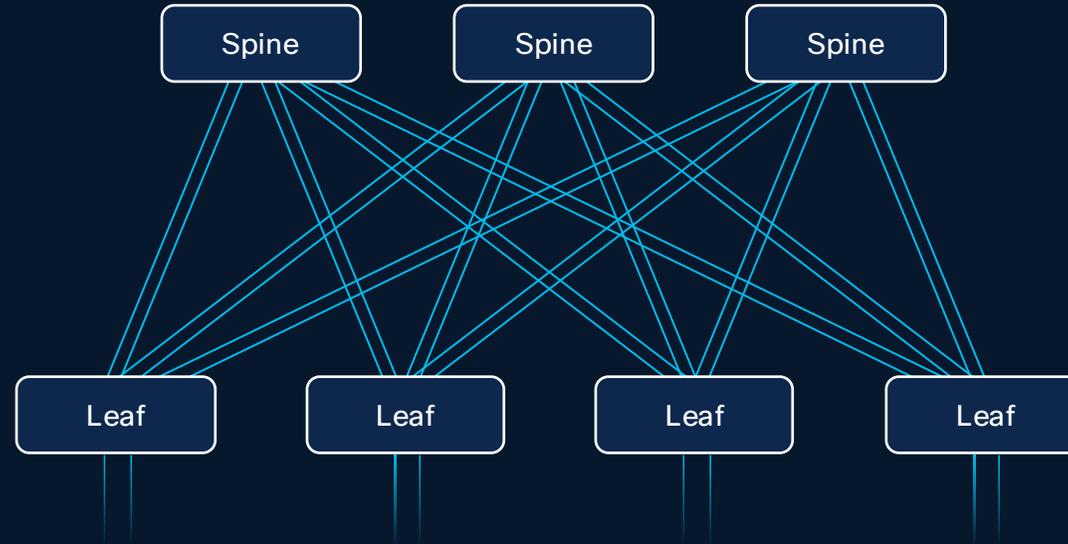
Advanced load balancing

FEATURES

Dynamic Load Balancing (DLB) with Load and Congestion Awareness

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

ECMP Static Pinning

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

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

High Entropy ECMP

*Introduced with NX-OS 10.6.1F

Per-Packet Load Balancing

Flow1
Flow2
Flow3
Flow4

Spine

Leaf

← **Per-packet load balancing by Cisco switches**

Server

Node-1

Node-32
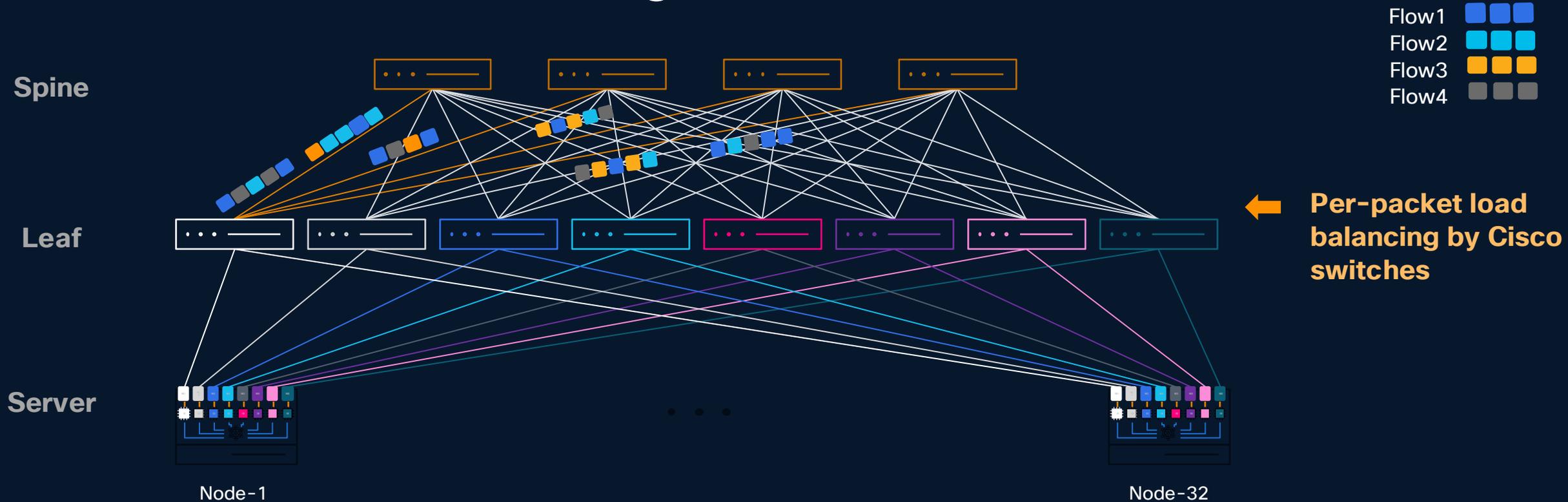
Cisco Per-packet load balancing

- 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**)

Bringing AI to the Enterprise.

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 future Cisco Nexus switches based on Spectrum-X ASIC

Session ID: CENDCN-1399

NEW Per-Packet Load Balancing with **NVIDIA Spectrum-X Integration**
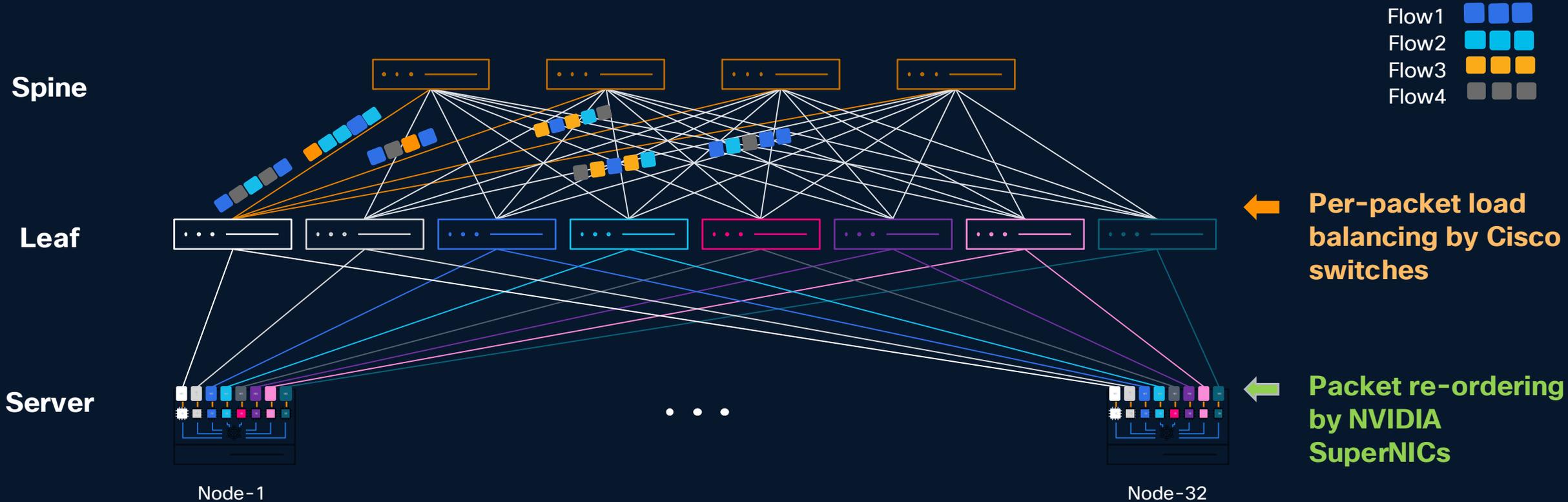
Cisco Per-packet load balancing with NVIDIA Spectrum-X integration

*

*

* **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**

Cisco Nexus 64-port 800G Portfolio

Built for Ethernet-Based AI Clusters That Scale

Silicon Diversity | **NX-OS or SONiC** | **Common Operating Model**

Shipping

Orderable in Nov. '25

Cisco N9364E-SG2

Cisco Silicon One

Cisco Cloud Reference Architecture (CRA)

NVIDIA Spectrum-X Technology Enabled

Cisco N9164E-NS4

NVIDIA Spectrum-X Ethernet Silicon

NVIDIA Cloud Partner (NCP) Reference Architecture

NVIDIA Spectrum-X Technology Built-In

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

Design   Build   Run

ORDERABLE 2H CY25

Cisco Nexus Hyperfabric

High-performance Ethernet

Cloud-managed operations

Unified stack including NVAIE

AI-native operational model

Democratize AI infrastructure
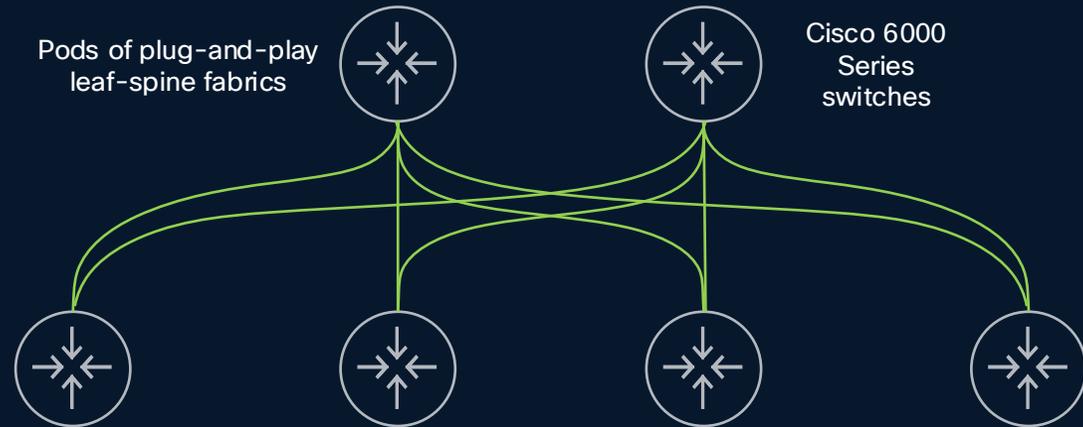
Visibility into full stack AI
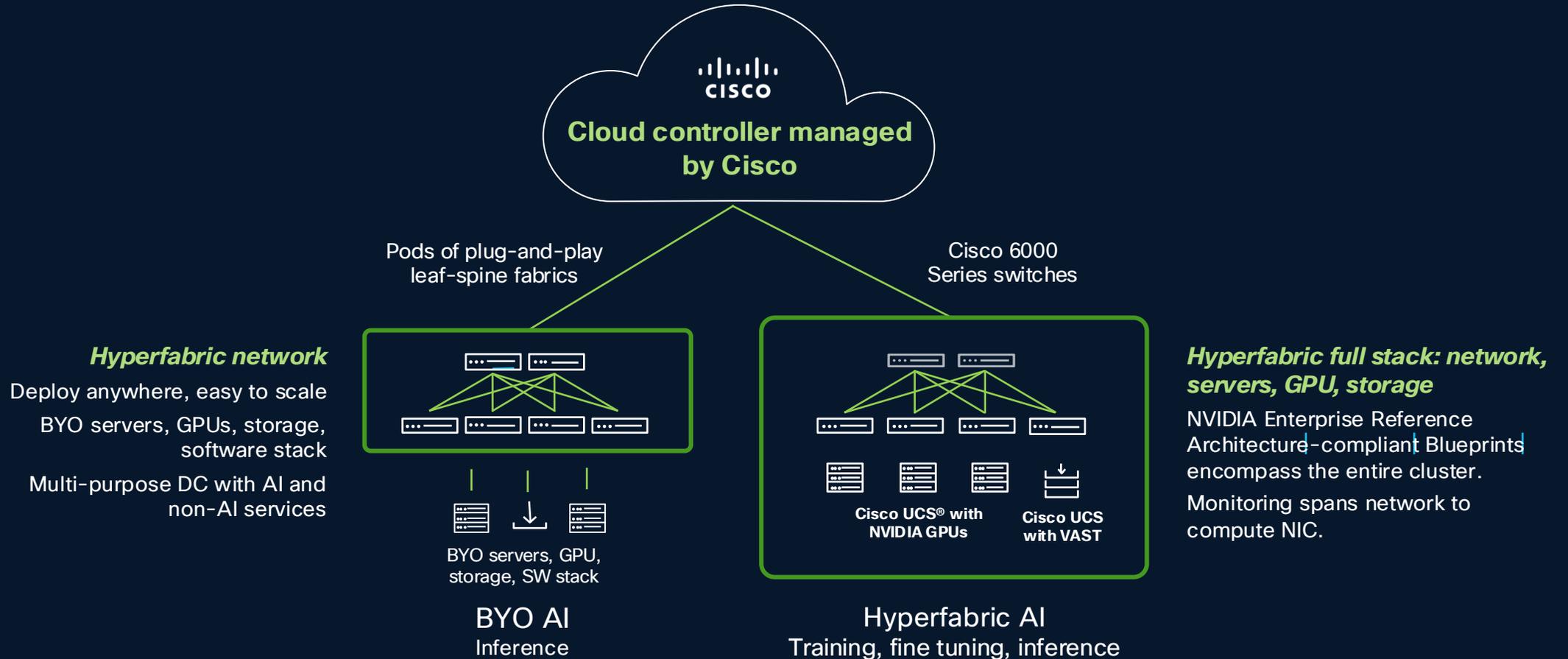
Cisco Nexus Hyperfabric

On-prem AI infrastructure

Pods of plug-and-play leaf-spine fabrics

Cisco 6000 Series switches

Cisco UCS C885A M8

NVIDIA GPU

NVIDIA DPU / NIC BlueField-3

VAST Storage UCS C225

AI solutions with Cisco Nexus Hyperfabric

Cloud controller managed by Cisco

Pods of plug-and-play leaf-spine fabrics

Cisco 6000 Series switches

Hyperfabric network

Deploy anywhere, easy to scale

BYO servers, GPUs, storage, software stack

Multi-purpose DC with AI and non-AI services

BYO servers, GPU, storage, SW stack

BYO AI
Inference

Cisco UCS® with NVIDIA GPUs

Cisco UCS with VAST

Hyperfabric AI
Training, fine tuning, inference

Hyperfabric full stack: network, servers, GPU, storage

NVIDIA Enterprise Reference Architecture-compliant Blueprints encompass the entire cluster.
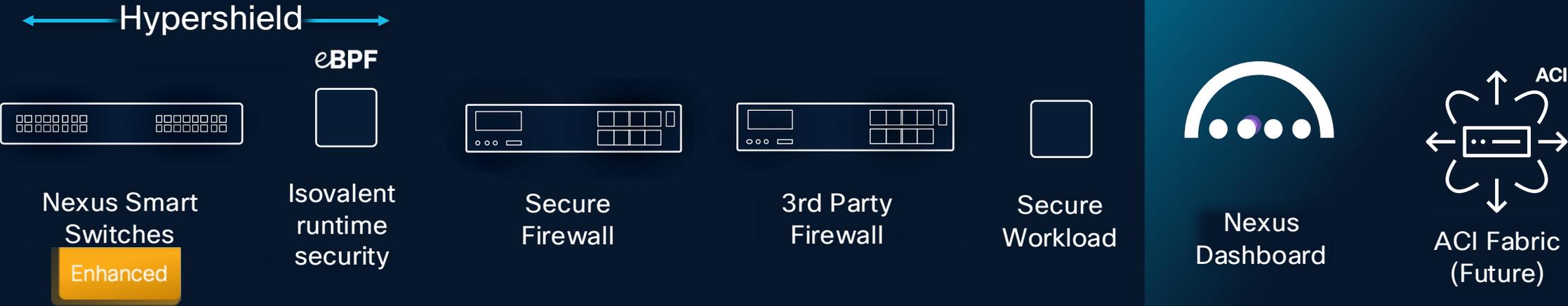
Monitoring spans network to compute NIC.

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

NEW

Hypershield

eBPF

Nexus Smart Switches

Enhanced

Isovalent runtime security

Secure Firewall

3rd Party Firewall

Secure Workload

Nexus Dashboard

ACI

ACI Fabric (Future)

Write policy once, enforce across the mesh

Nexus Smart Switch

Unmatched Flexibility, Performance, and Efficiency

Cisco Smart Switches

Cisco Nexus 9300 Services Accelerated Switch

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 |

Hypershield

- 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

| Distributed 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

Shipping

N9348Y2C6D-SE1U

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

800G Services Throughput
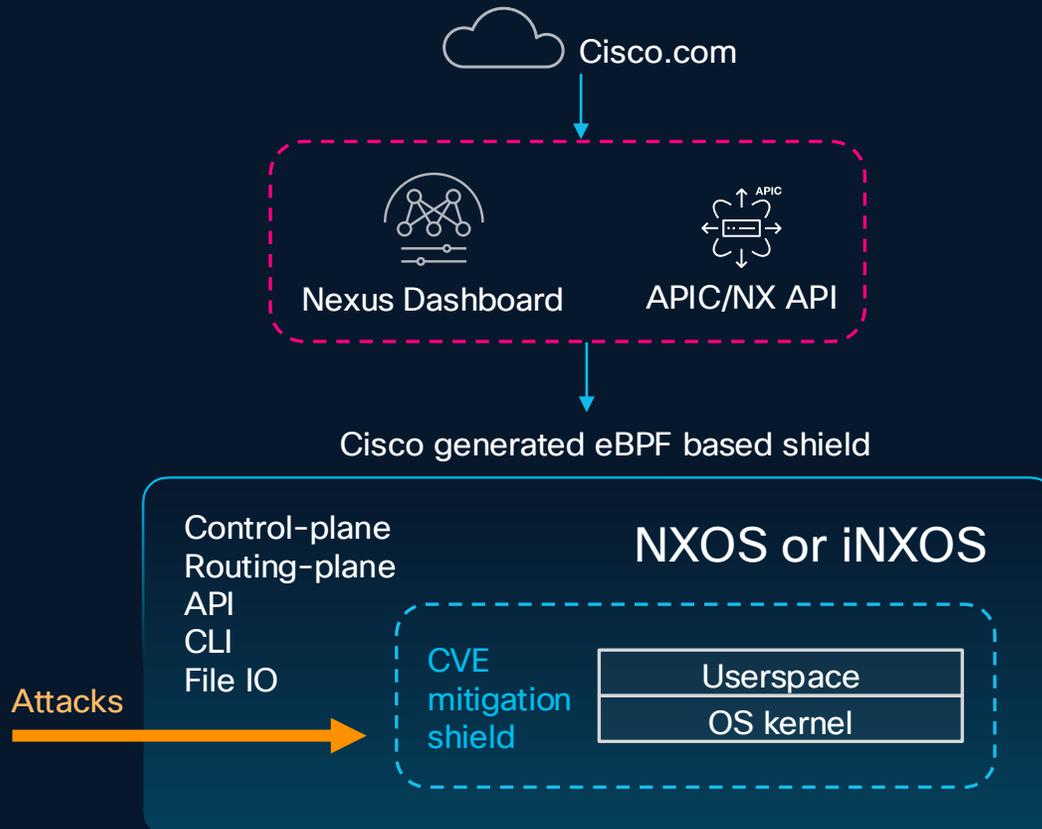
Cisco Hypershield

Use Cases

Top of Rack segmentation and enforcement

Cloud Edge

Zone-based segmentation

Live Protect – CVE Mitigation for Nexus NXOS Switches

No Downtime or Immediate PSIRT Software Upgrades

Cisco.com

Nexus Dashboard   APIC/NX API

Cisco generated eBPF based shield

Control-plane
Routing-plane
API
CLI
File IO

NXOS or iNXOS

Attacks

CVE mitigation shield

| Userspace |
| OS kernel |

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
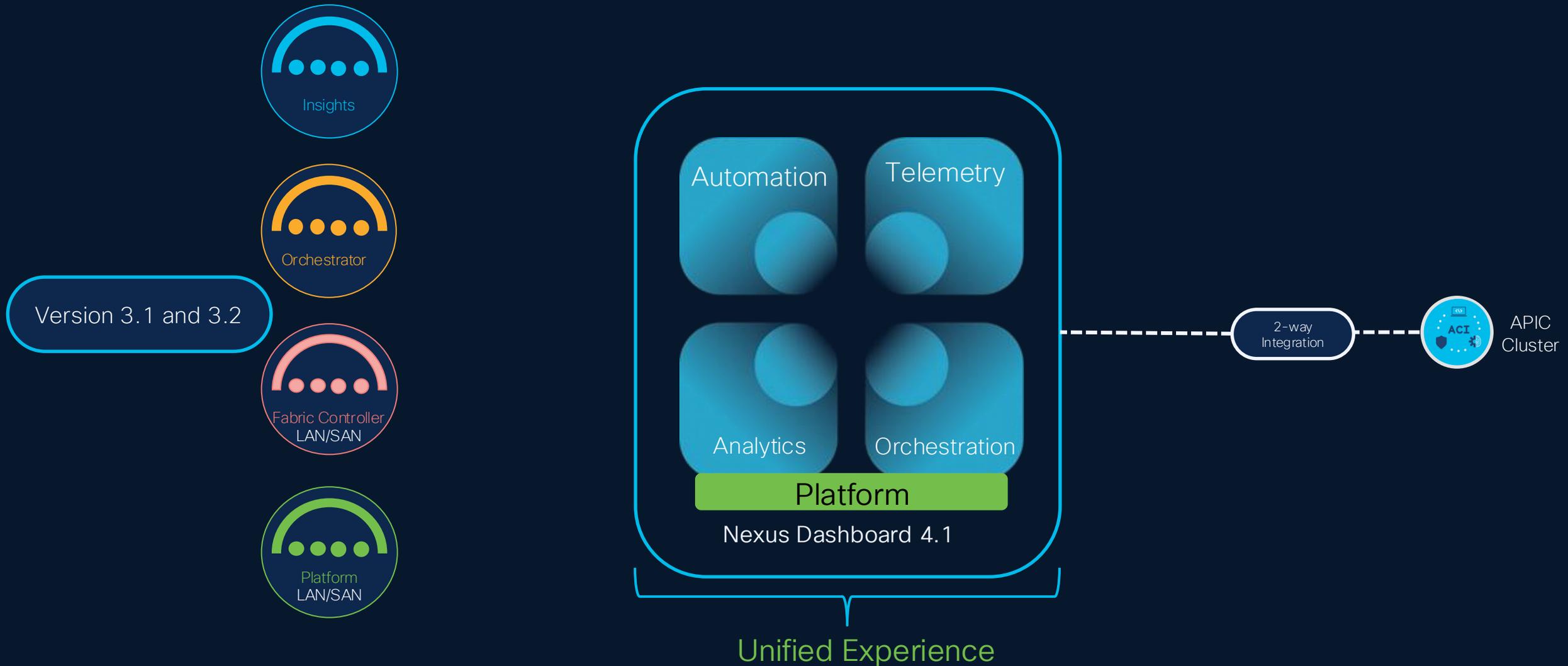- Upgrades during regular maintenance window

AI Ready DC Network
Operation

Unified Cisco Nexus Dashboard

A consistent user experience
for open networking

Unify your Operational Experience

Insights

Orchestrator

Fabric Controller
LAN/SAN

Platform
LAN/SAN

Version 3.1 and 3.2

Automation

Telemetry

Analytics

Orchestration

Platform

Nexus Dashboard 4.1

Unified Experience

2-way Integration

ACI

APIC Cluster

NEXUS DASHBOARD
Physical or virtual cluster

Data plane | Policy and control plane | Management plane

SAN

NX-OS

ACI

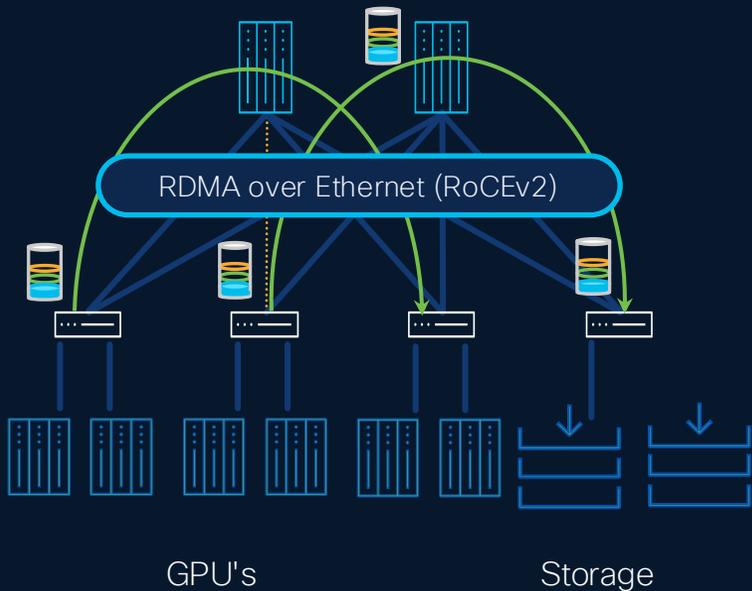Classic Ethernet | VXLAN | IP Media Fabrics | AI/ML Fabrics

Campus

WAN Transport

Virtual, K8
3rd party Networks

Cisco Nexus Dashboard
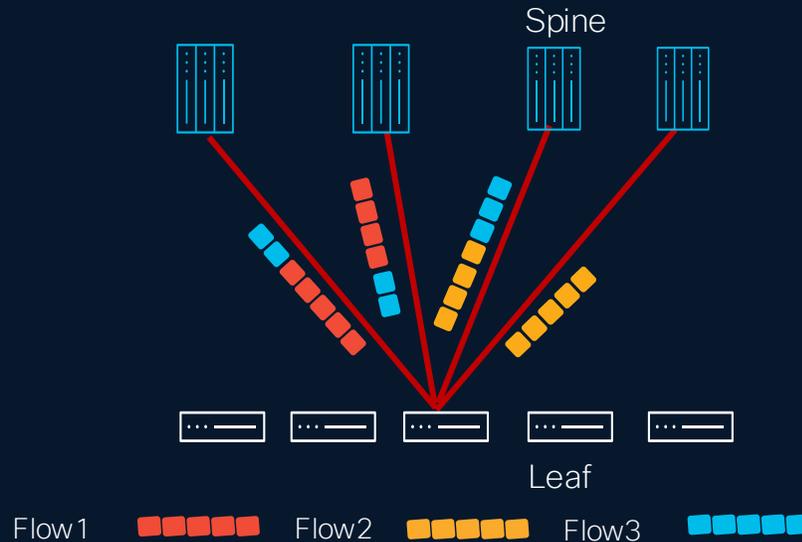## Automate your AI/ML network configurations

ECMP DLB between Spine-Leaf

Spine

RDMA over Ethernet (RoCEv2)

GPU's

Storage

Flow1   Flow2   Flow3

Leaf

RoCEv2 Switch Fabric

Spine

ECMP DLB between Spine-Leaf

Leaf

Manage network congestion
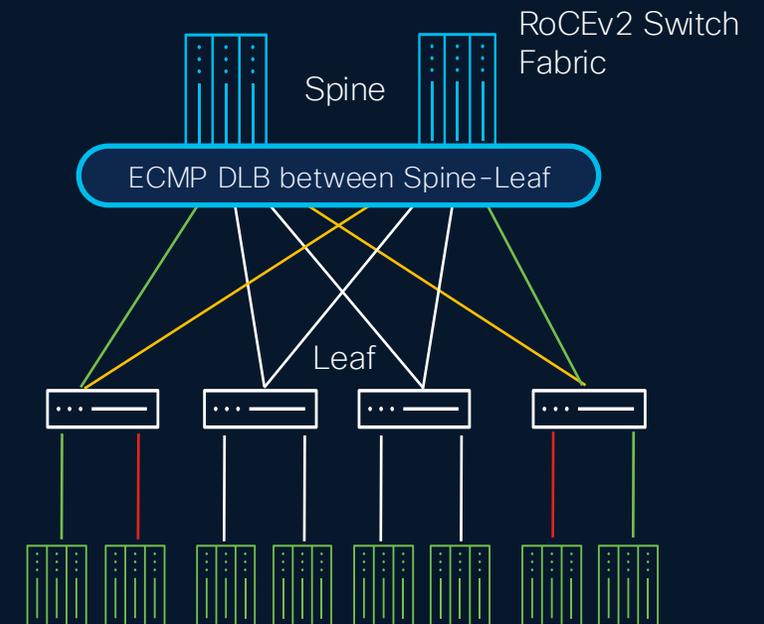with Lossless Network (PFC + ECN)

RoCEv2 support on Nexus 9800

Load balance flows/flowlets/Packets
based on link utilization

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

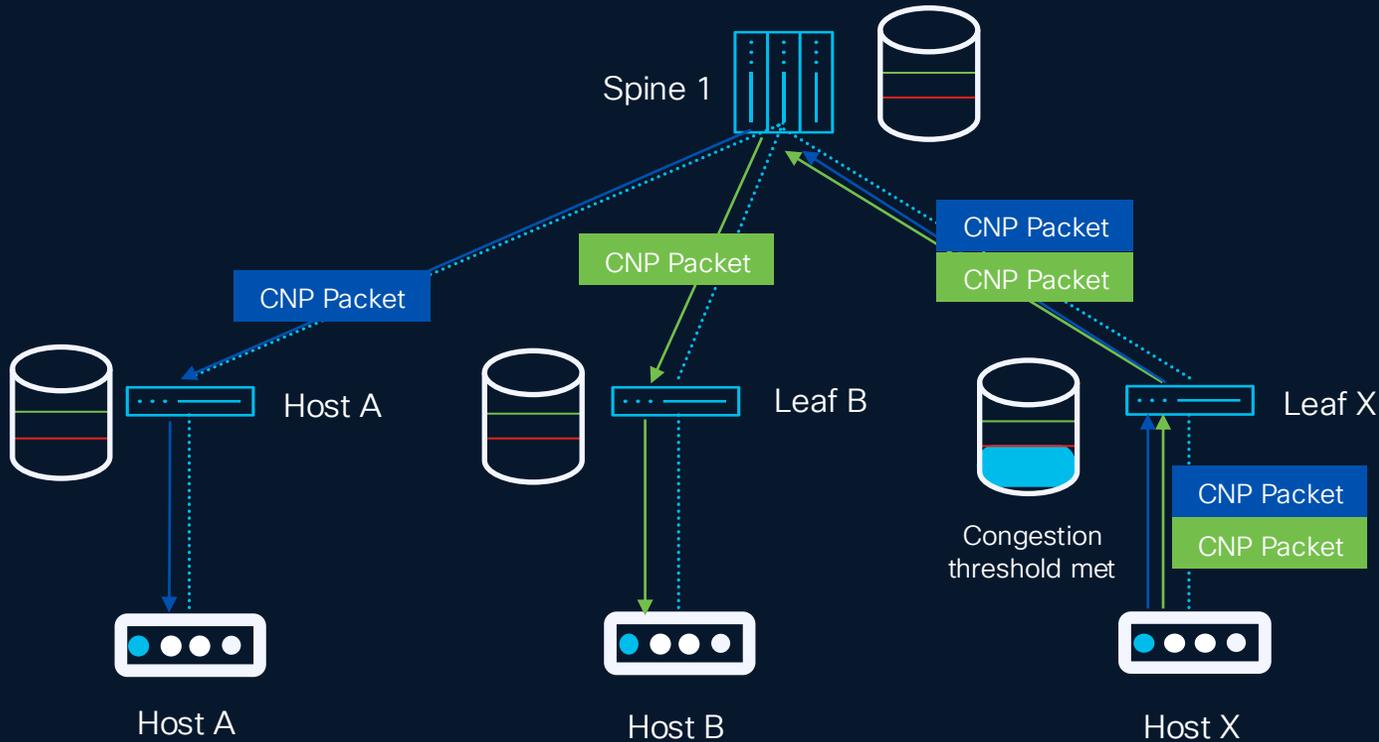**Traffic efficiency through pinning rules**
Map traffic from each downlink to the desired uplink

Allows efficient selection of Spines for communication
between leaf and spines

Benefit

Improved Congestion Management

Cisco Nexus Dashboard
AI/ML Fabric Settings

Spine 1

CNP Packet

CNP Packet

CNP Packet
CNP Packet

CNP Packet

Host A

Leaf B

Leaf X

CNP Packet
CNP Packet

Congestion
threshold met

Host A

Host B

Host X

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

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.
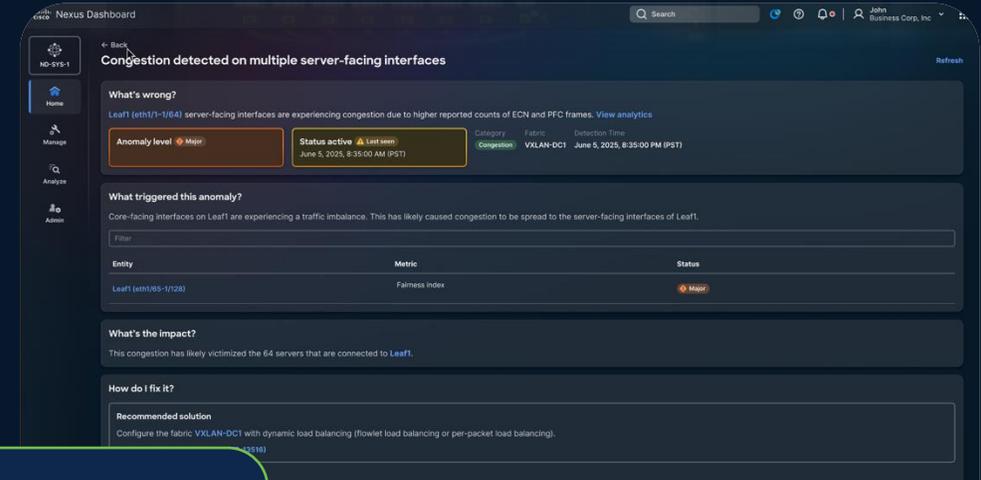
PFC Pattern Analysis

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.

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.

SLURM Job Data

Job scheduling and status info

NIC Metrics

Network interface statistics

AI Job KPIs

Hardware Connectivity

Switch interconnect status

Optics Health

Optical link condition data

GPU Telemetry

GPU usage and health metrics

Benefits

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

Cisco Nexus Dashboard
Optimize experiences with end-to-end visibility

splunk>
a CISCO company

Observability & Security
For networking, IT , security and engineering teams

Cross domain full stack observability with Application, Networking, Security and AI

Airgap / Cloud native including application dependency mapping, IT service intelligence and Enterprise security

Single DCN app for Splunk Enterprise and Splunk Cloud for ND,ACI and N9K

Nexus Dashboard with Splunk Embedded

ACCEDIAN

Assurance
For network and IT teams

Cloud-native AI service assurance for service provider/telco using nexus Dashboard and Provider connectivity assurance

Network performance, enriched connectivity analysis, DC to external services

Accedian SFP recognition, Automation of agent installation using Nexus Dashboard

ThousandEyes

Assurance
For network and IT teams

CloudAI driven closed loop assurance for Airgap, cloud, and hybrid Enterprise networks

DC configuration correlation, active synthetics, proactive recommendations for Inter DC, Intra DC, and Hybrid insights

Automated ThousndEyes agent installation, Internet insights, application performance

Benefits

AI driven assurance and observability: reactive to proactive

Transform data centers to power AI workloads anywhere

AI-ready data centers

Digital resilience

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

CISCO Engage | Tech Day Miami

Resources

Cisco Data Center Networking
View on Cisco.com

Cisco Nexus Dashboard
View on Cisco.com

Cisco Nexus Hyperfabric
View on Cisco.com

Cisco Secure AI Factory with NVIDIA
Visit Cisco.com

Blogs
Visit blogs.cisco.com/datacenter

Online community
Visit the Data Center and Cloud online community