Be AI-Ready

Build Scalable, Resilient Data Center Networks

Davis Creemer
SE – Cloud & AI Infra
dcreemer@cisco.com

Alan Gehami
AE – Infrastructure Architecture
agehami@cisco.com

Agenda

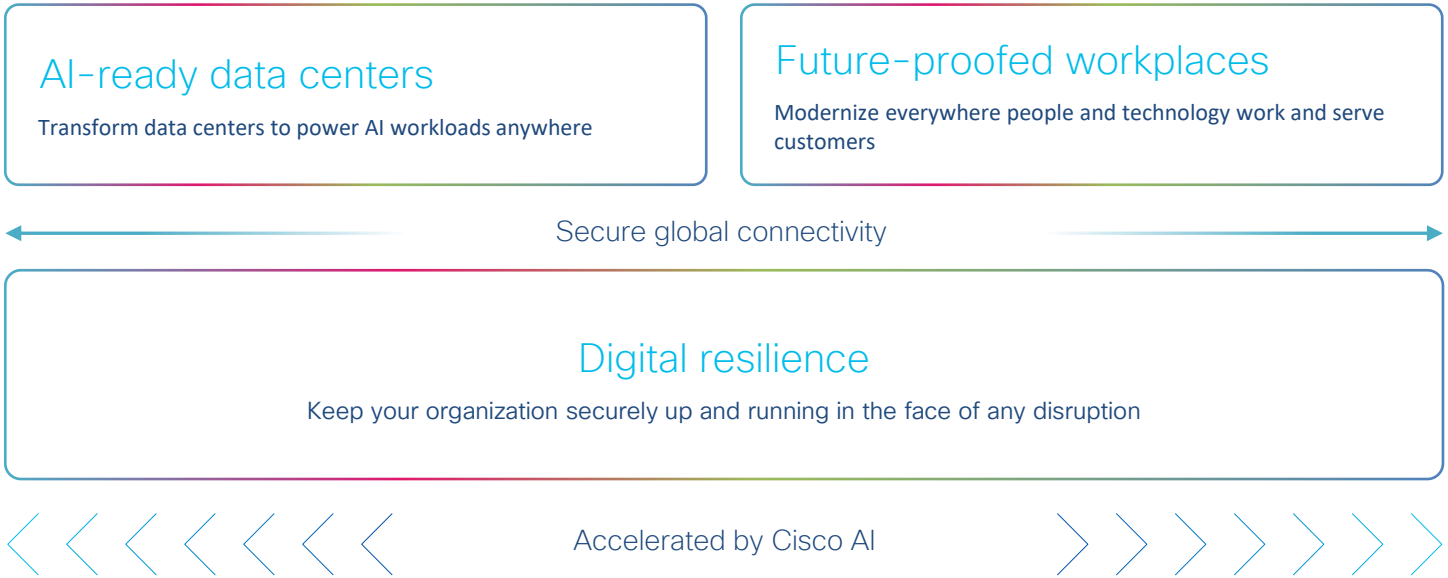
- Introduction
- AI/ML Networking
- Nexus in the AI-Ready Data Center
- Fabric Options for the AI-Ready Data Center
- Cisco Agile Data Center Interconnects

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Introduction

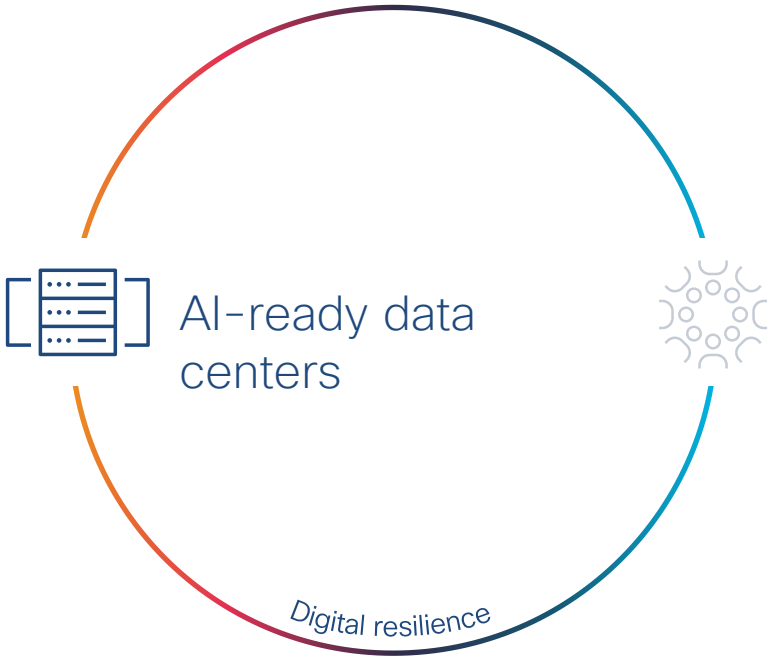


Cisco powers how people and technology work together across the physical and digital worlds



Transform data centers to power AI workloads anywhere

Public and private clouds, edge, on-premises



Comprehensive infrastructure

Power AI with networking, compute, and storage in fully-integrated, scalable, and modular systems for all workloads

Seamless operations and observability

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

Security from ground to cloud

Protect hyper-distributed workloads by infusing security everywhere

In the 90s, before the internet, we had to read books

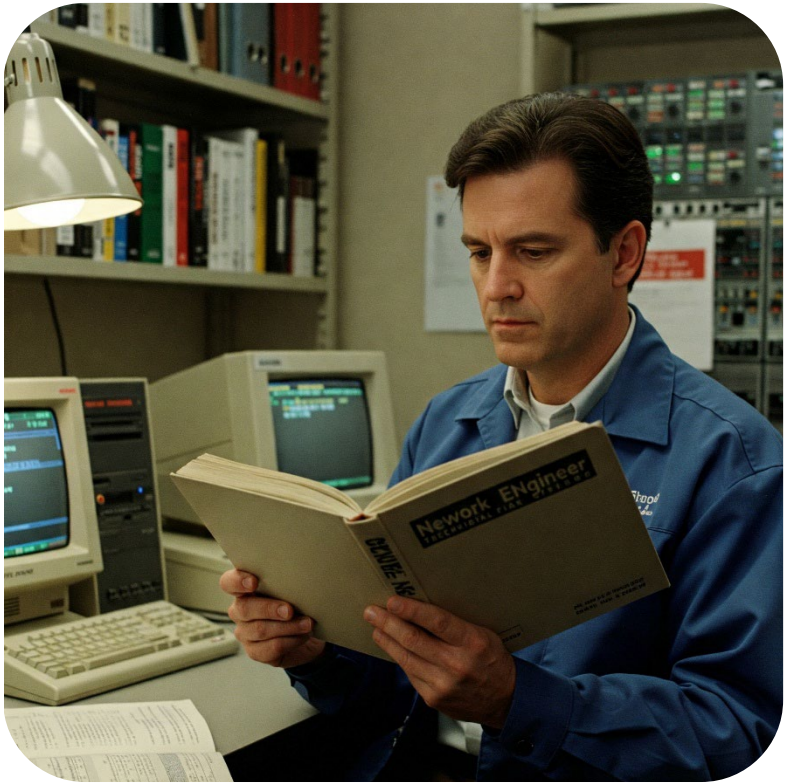


image generated by AI

The internet and search engines have made a fundamental change to how we work



image generated by AI

AI will bring another fundamental change of similar, if not greater, magnitude

Prompt

”Explain RAG to me”

Response

Okay, let’s break down RAG – Retrieval-Augmented Generation. It’s a really hot topic in the world of AI, particularly with large language models (LLMs) like ChatGPT. Here’s a simplified explanation:

1. The Problem with Traditional LLMs:

* **LLMs are Trained on Massive Datasets:** Models like ChatGPT are trained on *huge* amounts of text data from the internet.
* **They Can Hallucinate:** Because they’re predicting the next word based on patterns, they can sometimes generate incorrect, misleading, or completely fabricated information – we call this “hallucinating.” They don’t truly *understand* the information they’re presenting.
* **Limited Context:** LLMs have a limited “memory” of the conversation. They often struggle to maintain context over …

AI Networking



Every organization's AI approach and needs are different

The model
Training

Optimize the model
Fine-tuning and RAG

Use the model
Inferencing



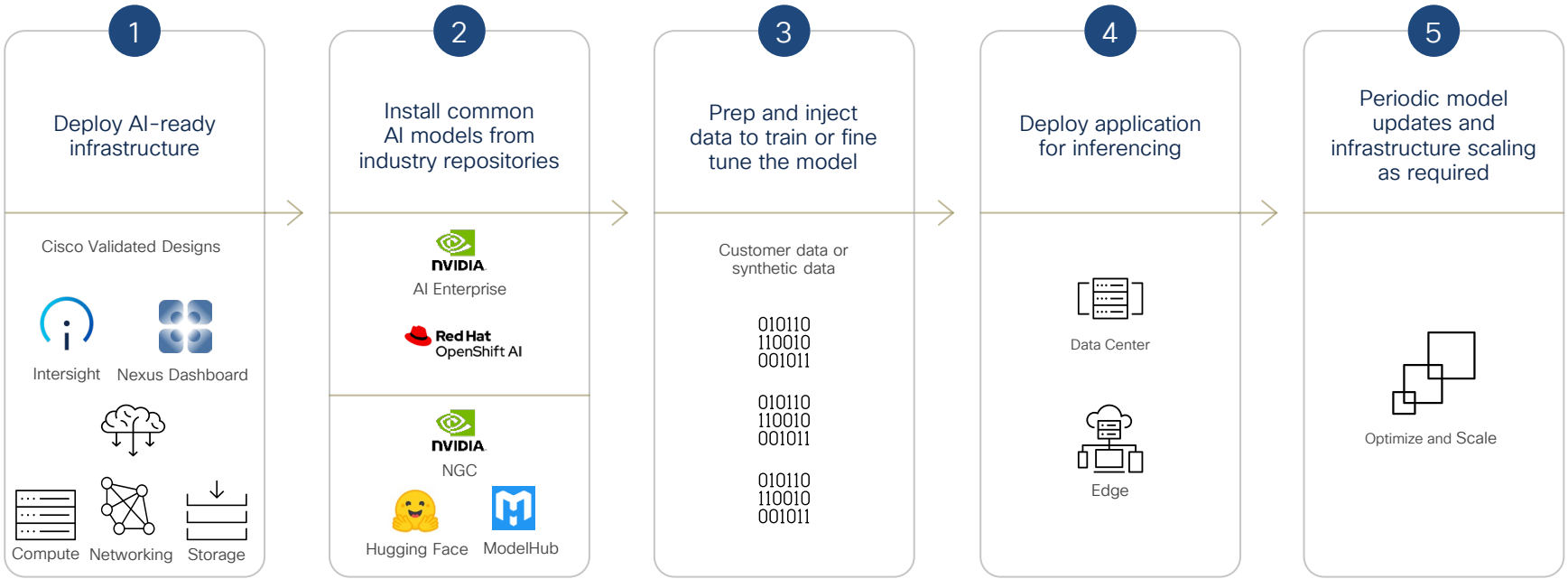




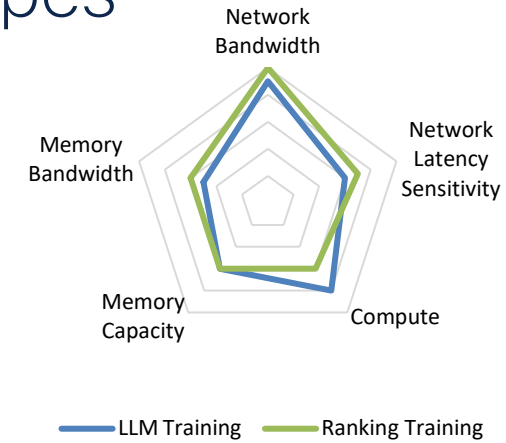


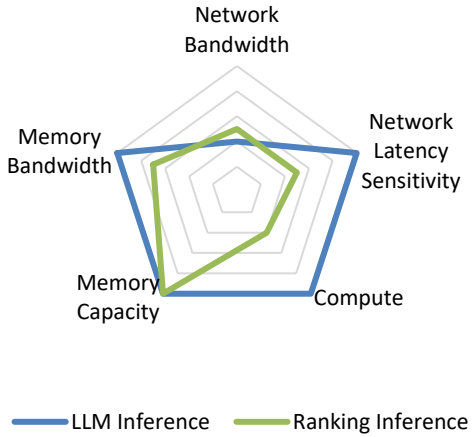

cisco Connect

A helicopter view of an AI Deployment Journey



AI Workload Types

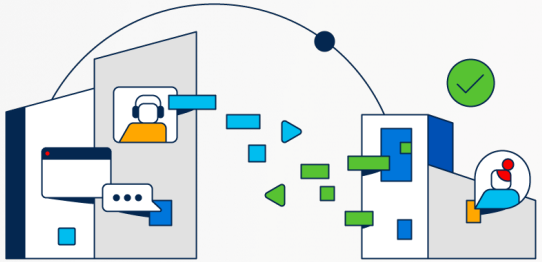




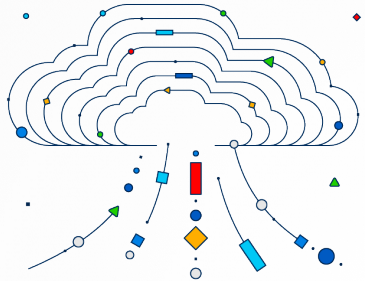
	Training	Inference
Node to Node Bandwidth	High	Low
Key Metric	Training time of a model	High Availability and Latency
Operational Mode	Model training is offline	Usually online, requires real time response
Infrastructure requirement	Large network with many GPU/CPU hosts	Smaller network with mid size of CPU/GPU hosts

Where do you run your AI/ML training cluster?

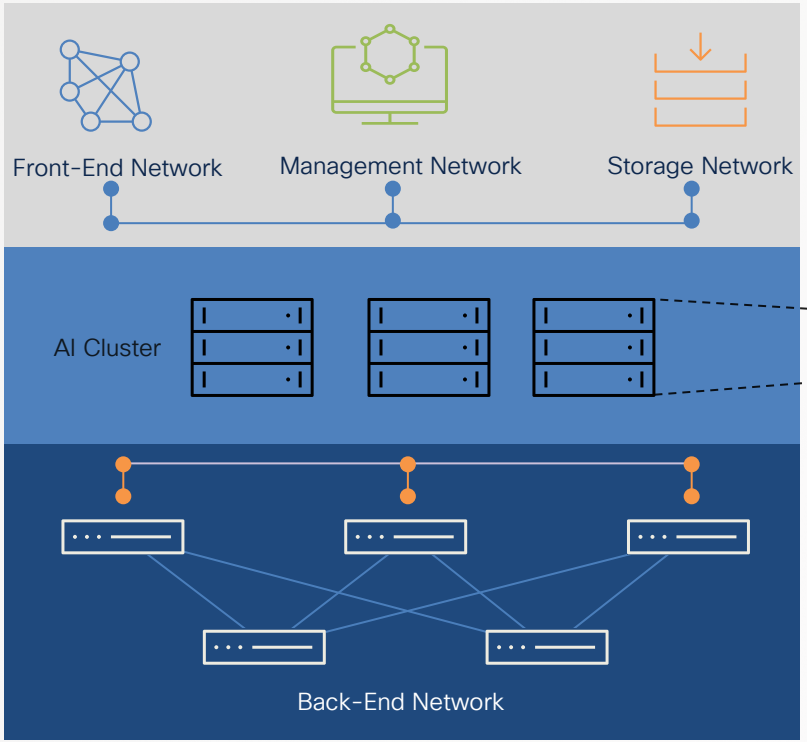
- On Prem
  - Always available for enterprise to use
  - Flexibility for large enterprise to leverage same cluster for different functions
  - Data is stored on prem



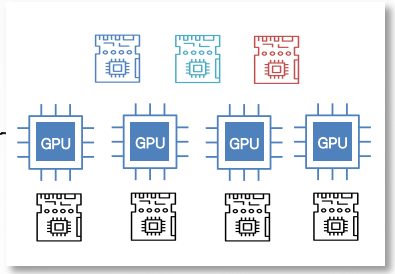
- In Public Cloud
  - Provides flexibility, pay for what you need
  - Cost will grow with more data and training
  - Challenge: Cost of egress data from the cloud



AI Infrastructure Building Blocks



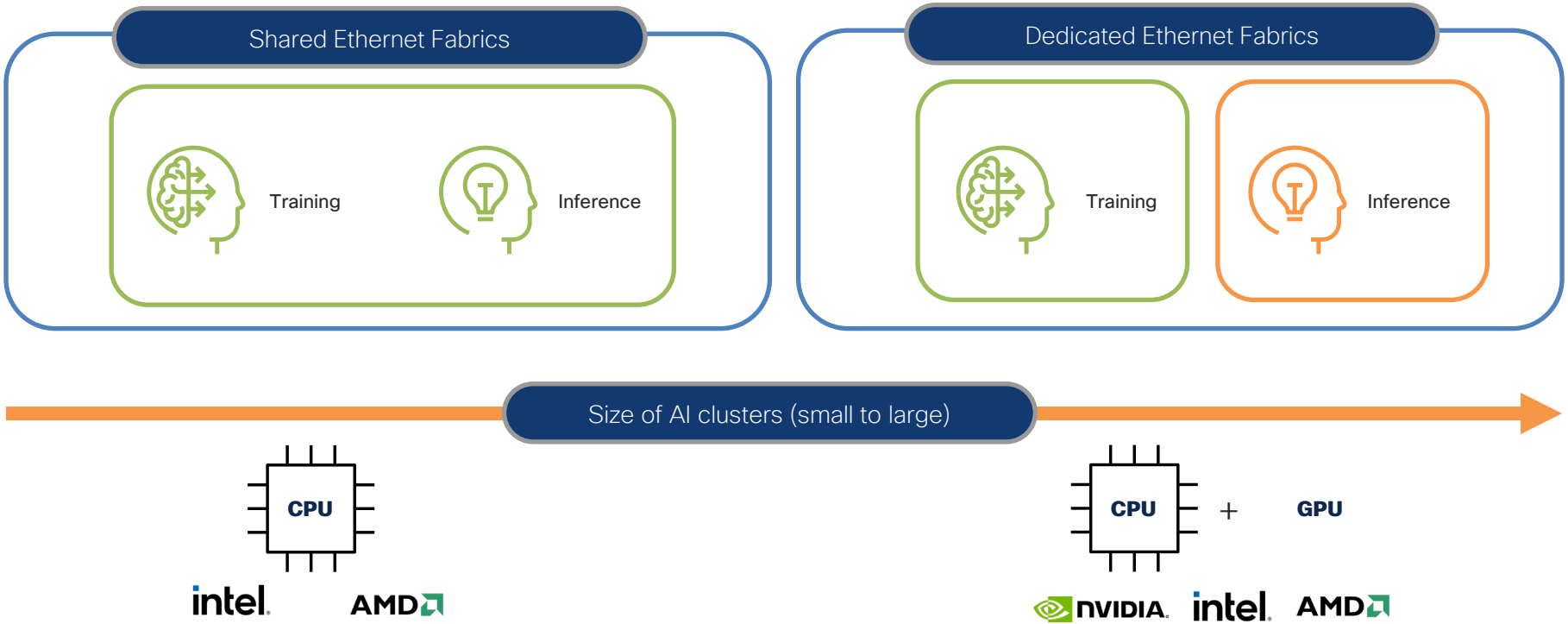
Network connecting hosts to the world
Used to manage hosts
Shared or dedicated storage network



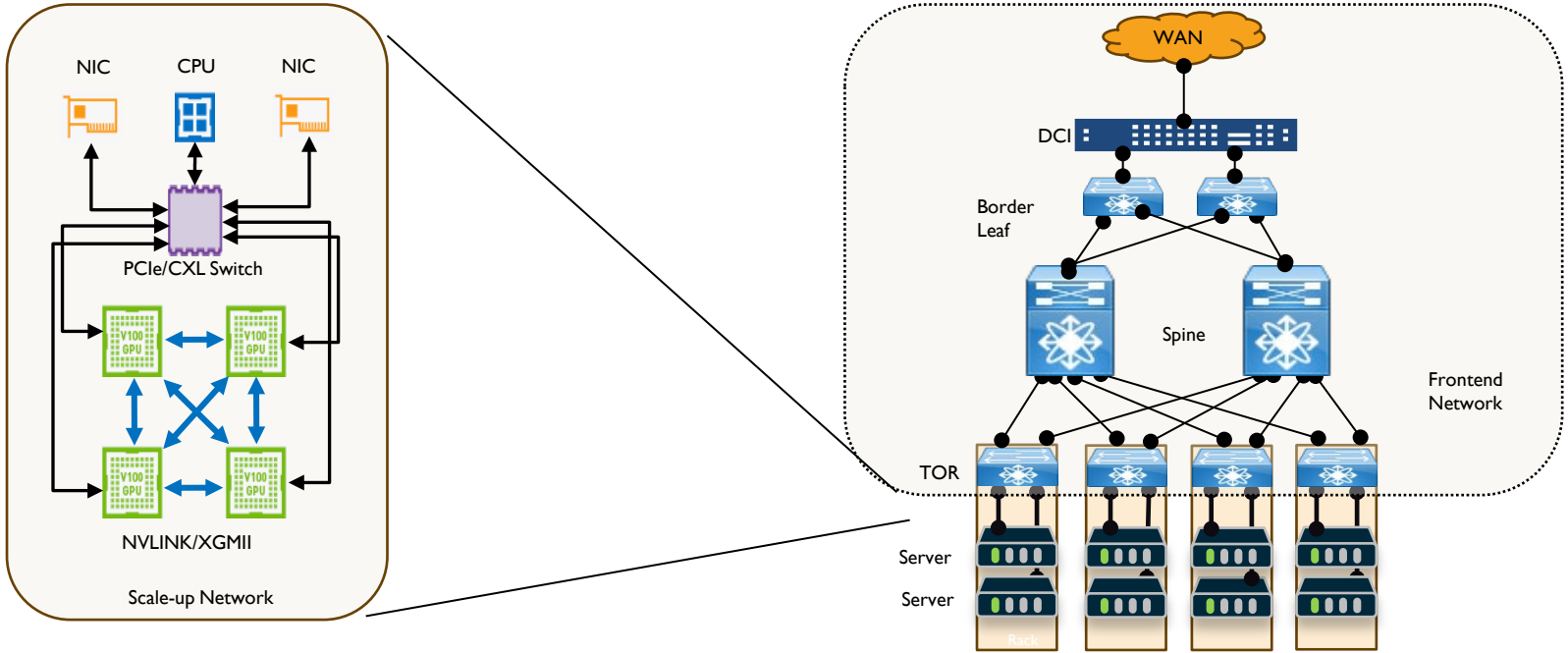
AI Node

Network to interconnect GPUs
Only contains GPU to GPU communication
Lossless Network
Non-blocking / High Speed

Deployment Options for Ethernet Fabrics

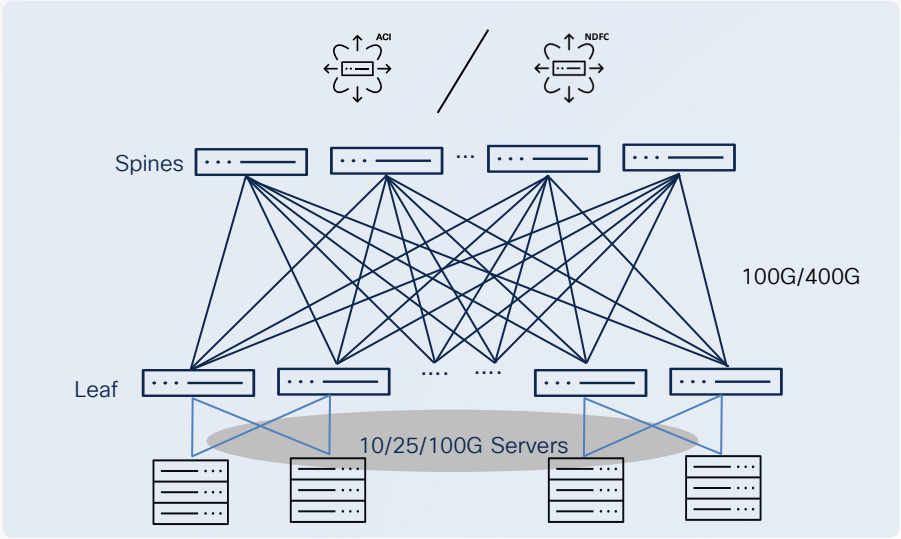


Scale Up w/in a Workload
Scale Out – Between Workloads



ACI or NXoS - as the Front-End network

Repurposing existing infrastructure for AI workloads

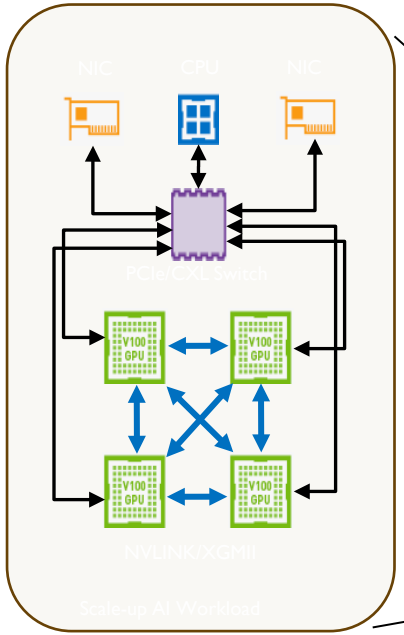


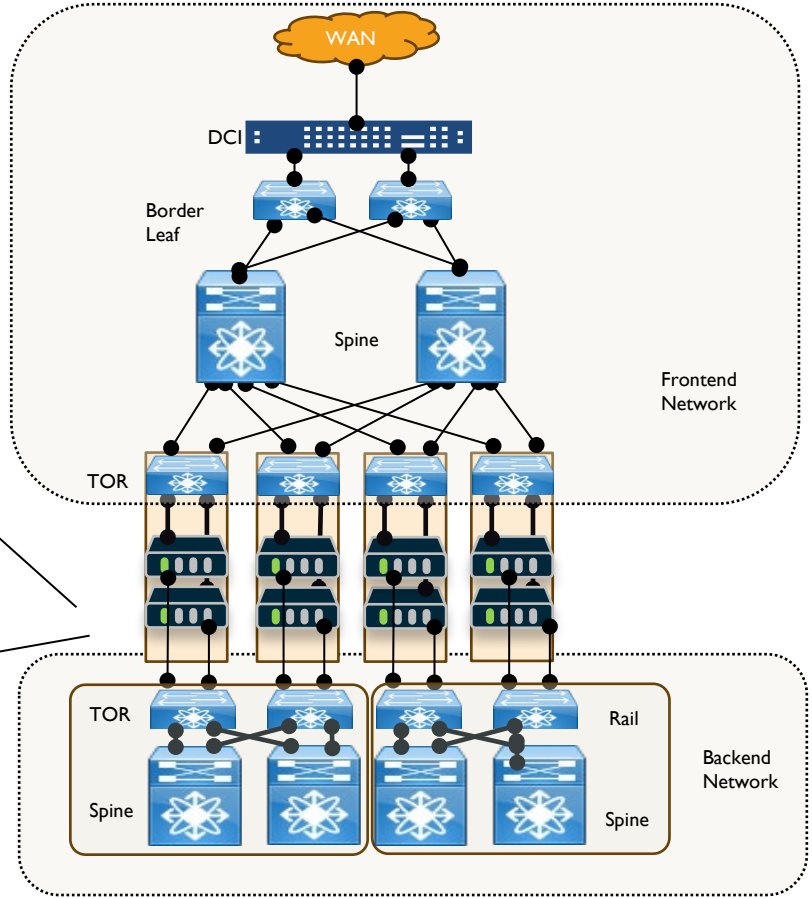
Management Network

Storage Network

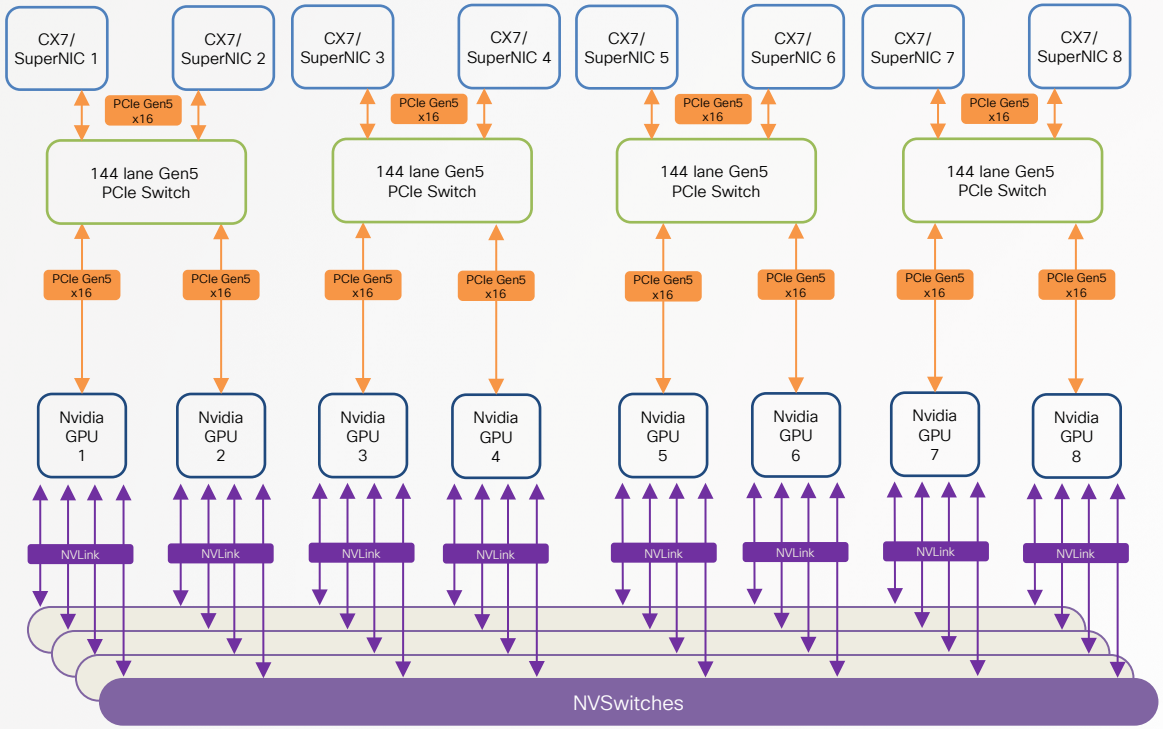
Front-End Network

Scale Up w/in a Workload
Scale Out – Between Workloads





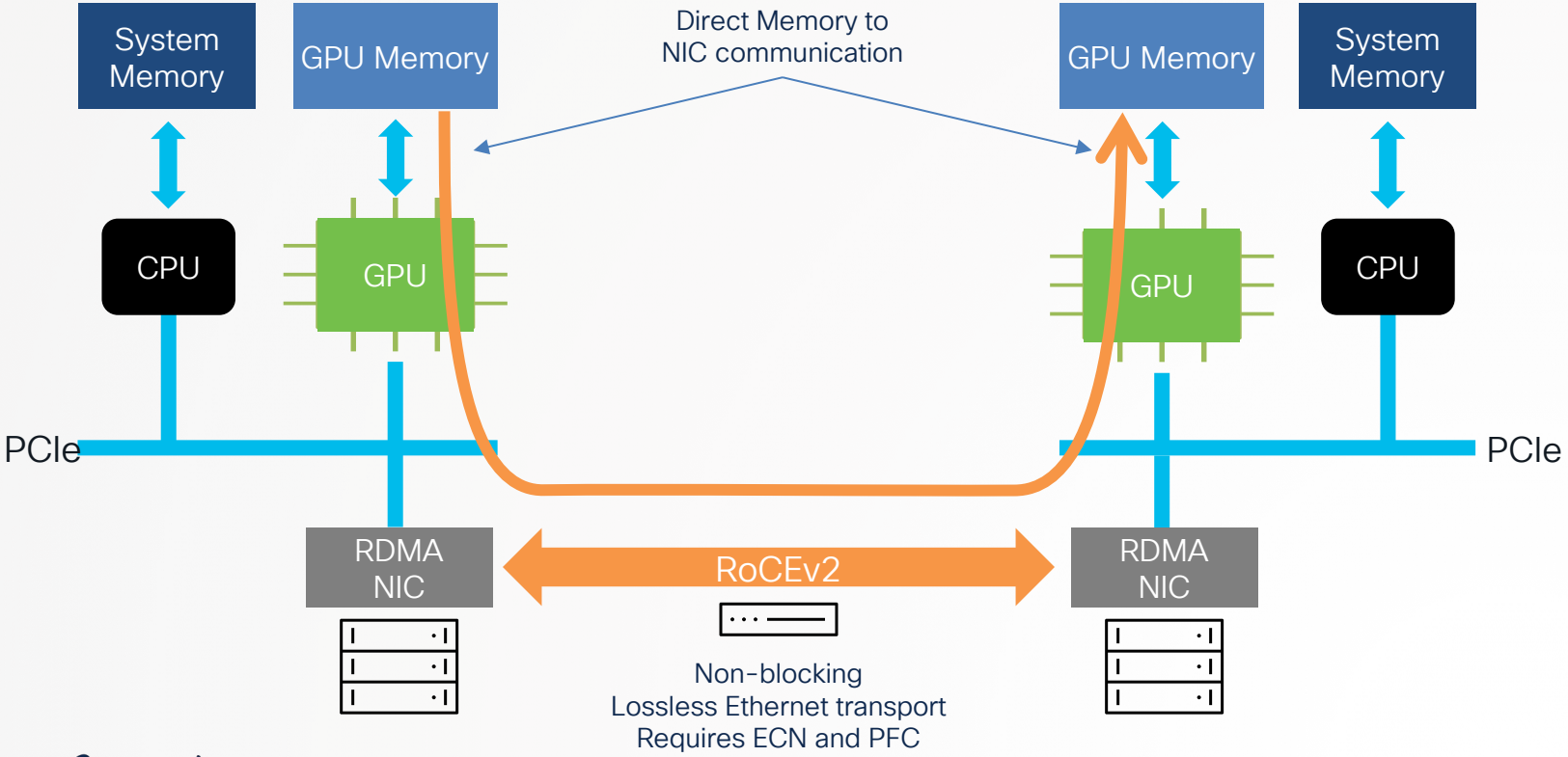
UCS C885A – Nvidia GPU Connectivity



- 8x Nvidia H100/H200 SXM5 Tensor Core GPUs
- Each H100/H200 GPU has multiple NVLink ports and connects to all four NVSwitches
- 4 x fully non-blocking NVSwitches that connect all 8 GPUs
- NVLink bidirectional speed of 900GB/s between any pair of GPUs in the same node
- Each H100/H200 GPU also has a dedicated NIC/SuperNIC connected via PCIe Gen5 x16 for GPU-to-GPU connectivity across nodes

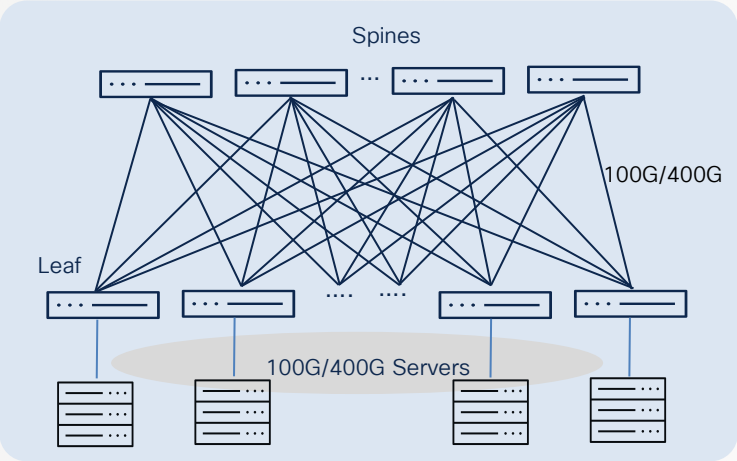
Remote Data Memory Access (RDMA)

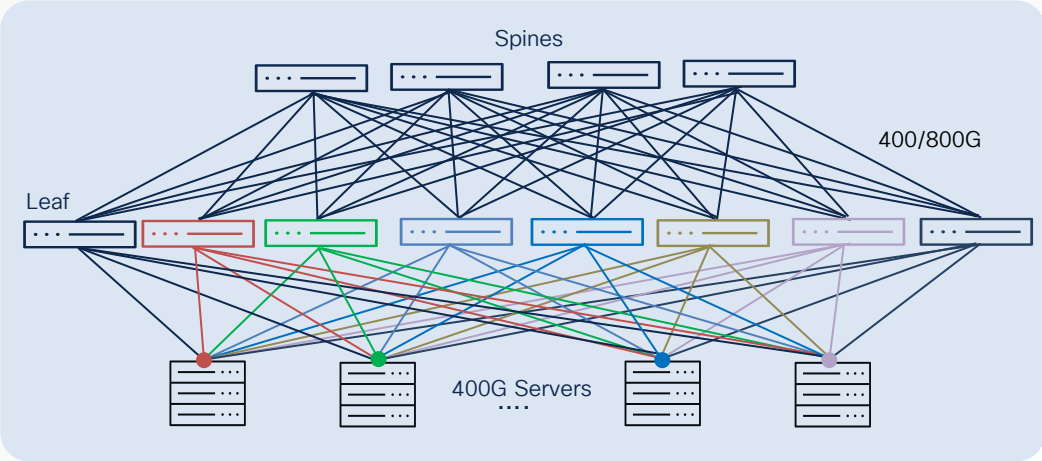
RDMA over Converged Ethernet (RoCE)



Back-End Network Options, but not limited to

Classic or Rail-Optimized Designs with Nexus 9300 Series





Classic Design

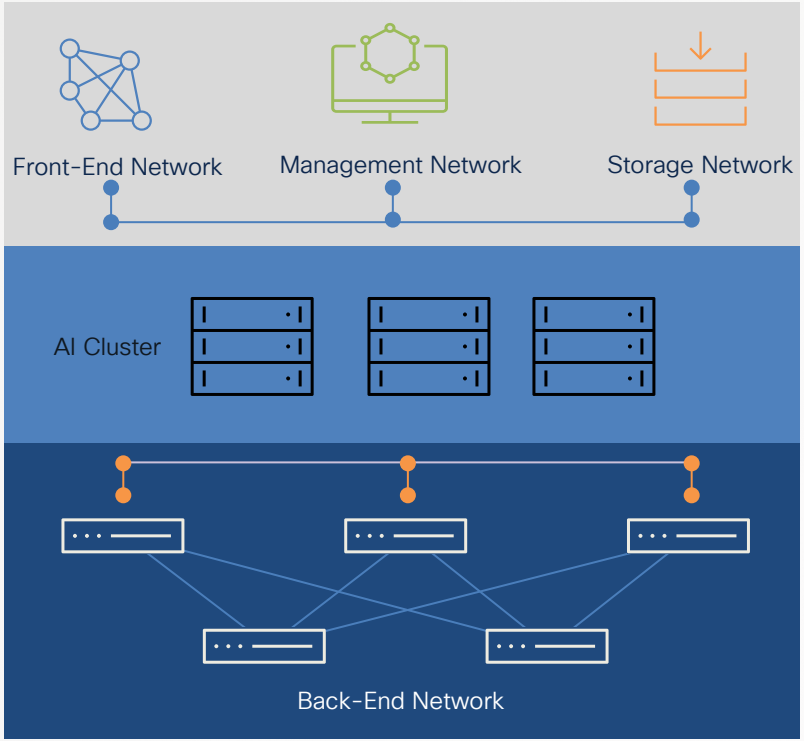
Hosts mapped per Rack
Non-blocking / 1:1 Oversubscription
Up to 64-port 400G or 800G options for Leaf/Spine

Rail-Optimized Design

8 GPU:NIC pair mapped per Rail/Switch
Non-blocking / 1:1 Oversubscription
Up to 64-port 400G or 800G options for Leaf/Spine

These design choices strictly depend on Collective Communication Library (CCL) optimizations to improve overall performance.

Where RoCEv2 traffic is mostly seen…



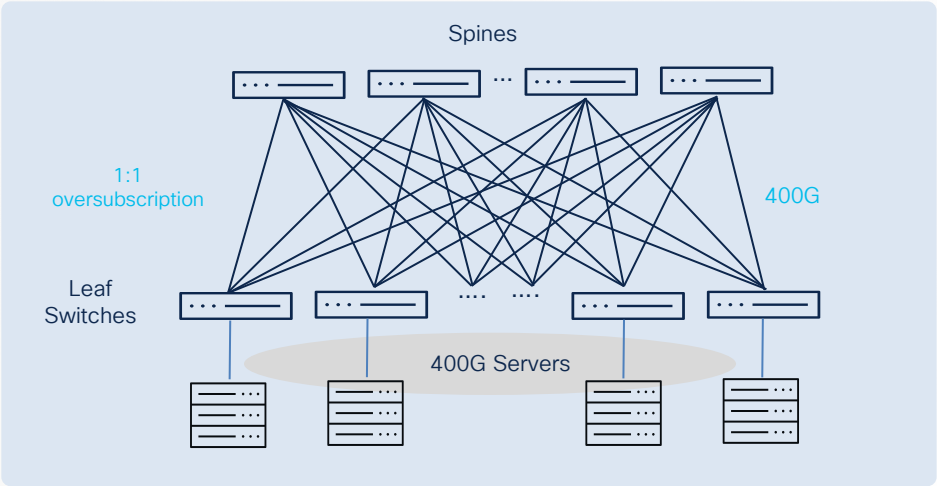
RoCEv2

ECN + PFC configuration required for storage traffic (dedicated or shared network) and for GPU traffic in the Back-End network

RoCEv2

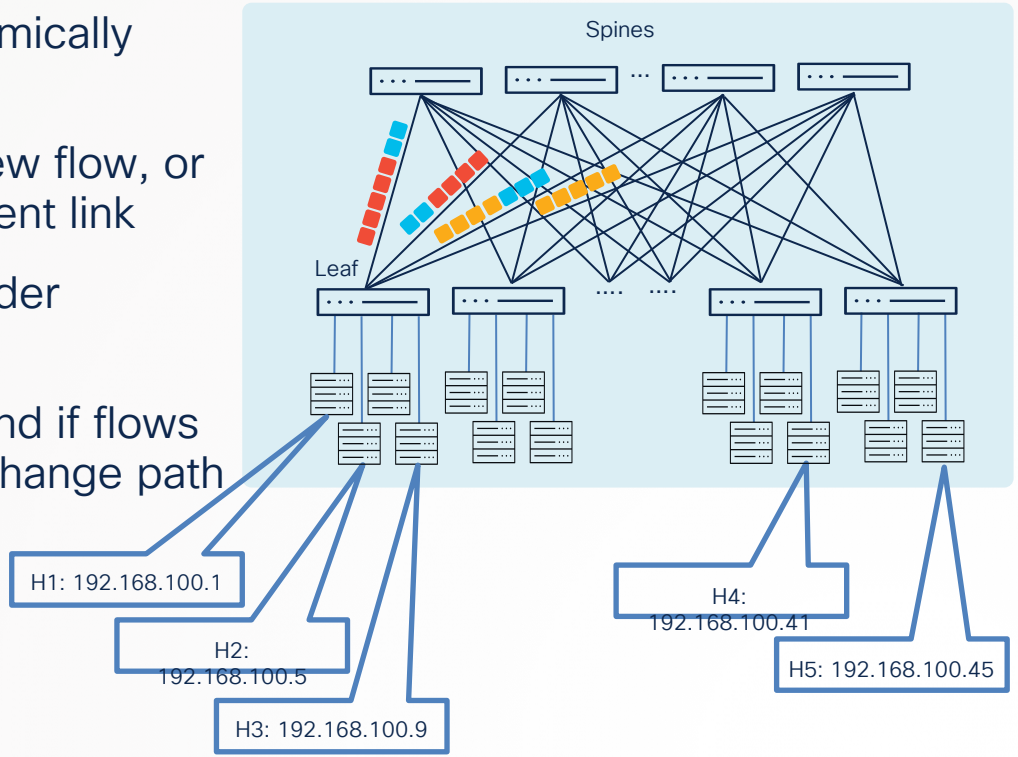
Non-blocking Network

- Allows host to talk to other hosts at full bandwidth
- Leaf: Same bandwidth to the host as to the spine
- Reduces need for congestion management to increase performance



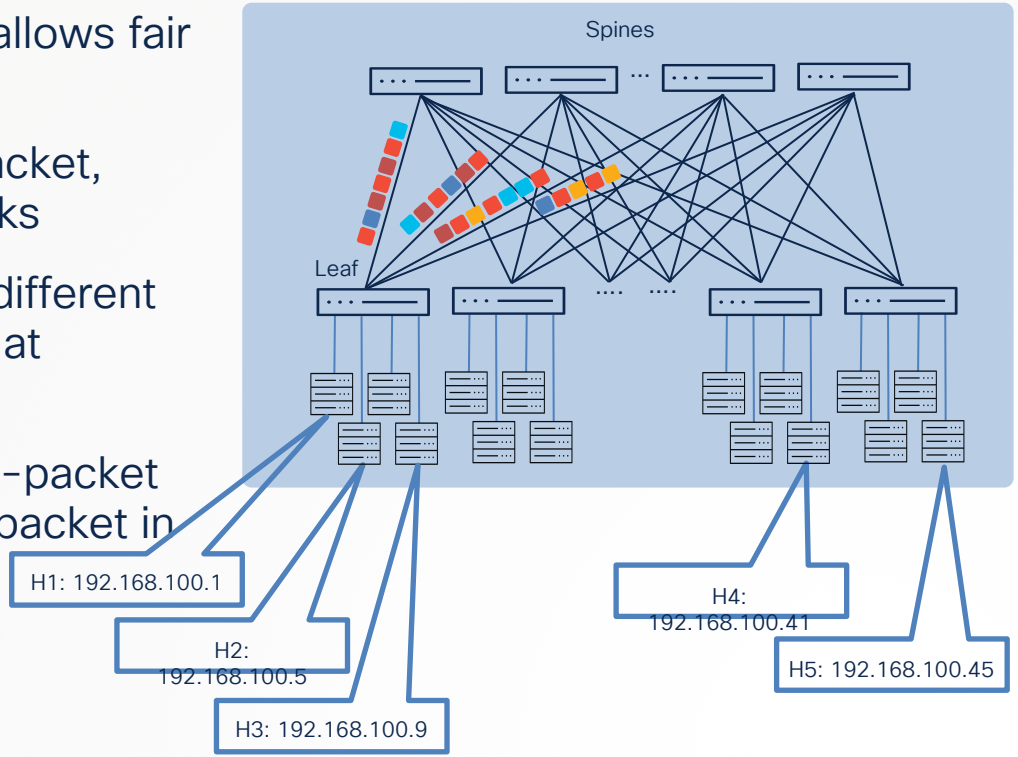
Dynamic Load Balancing – Flowlet Mode

- Flowlets maintain the link, or dynamically change the link
- Least utilized link is chosen, for new flow, or at times flows are moved to different link
- Aging time is used avoid out of order packets; it should be 2 x RTT
- AI workloads are balanced fairly and if flows are bursty, they can dynamically change path



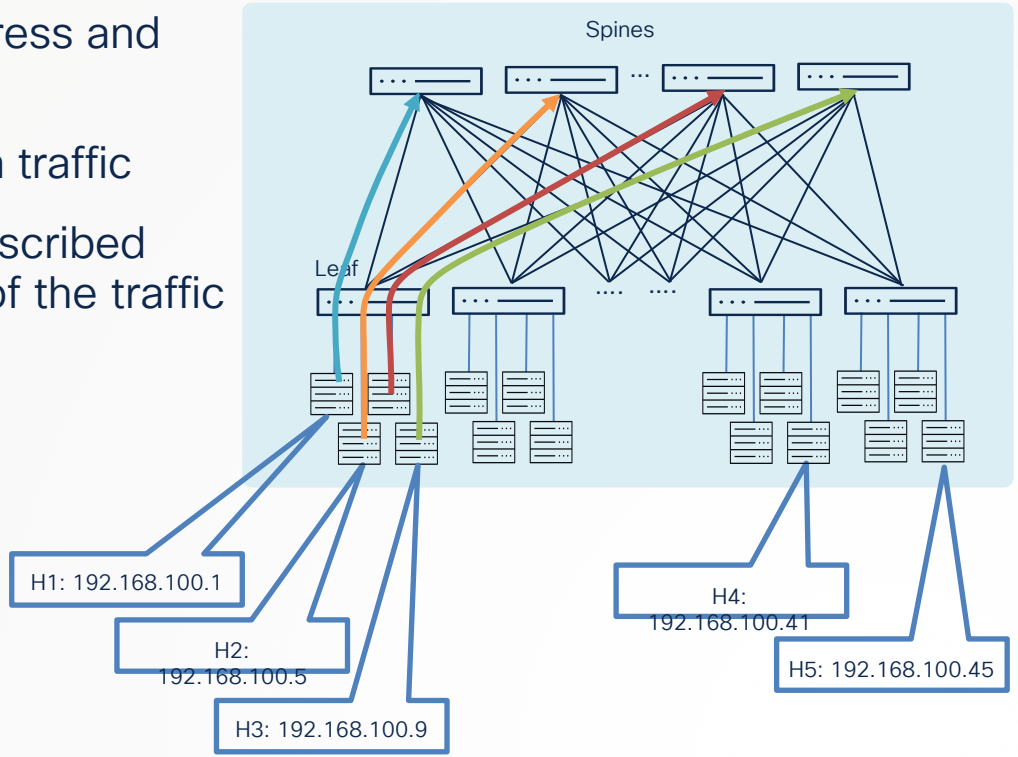
Dynamic Load Balancing – Per-packet Mode

- Per packet mode load balancing, allows fair distribution of traffic
- Least utilized link is chosen per packet, delivering uniform usage of the links
- As packet of same flow, will take different path, may happen that they arrive at destination endpoints out of order
- Expected in deployments with per-packet load balancing for receiver to put packet in order
  - Host needs to do packet reordering



Dynamic Load Balancing – Static Pinning

- Allows static pinning between ingress and egress ports
- Valid only for multipath destination traffic
- Steady bit rate flows, or undersubscribed network, to allow fair distribution of the traffic
- Manual configuration, or API



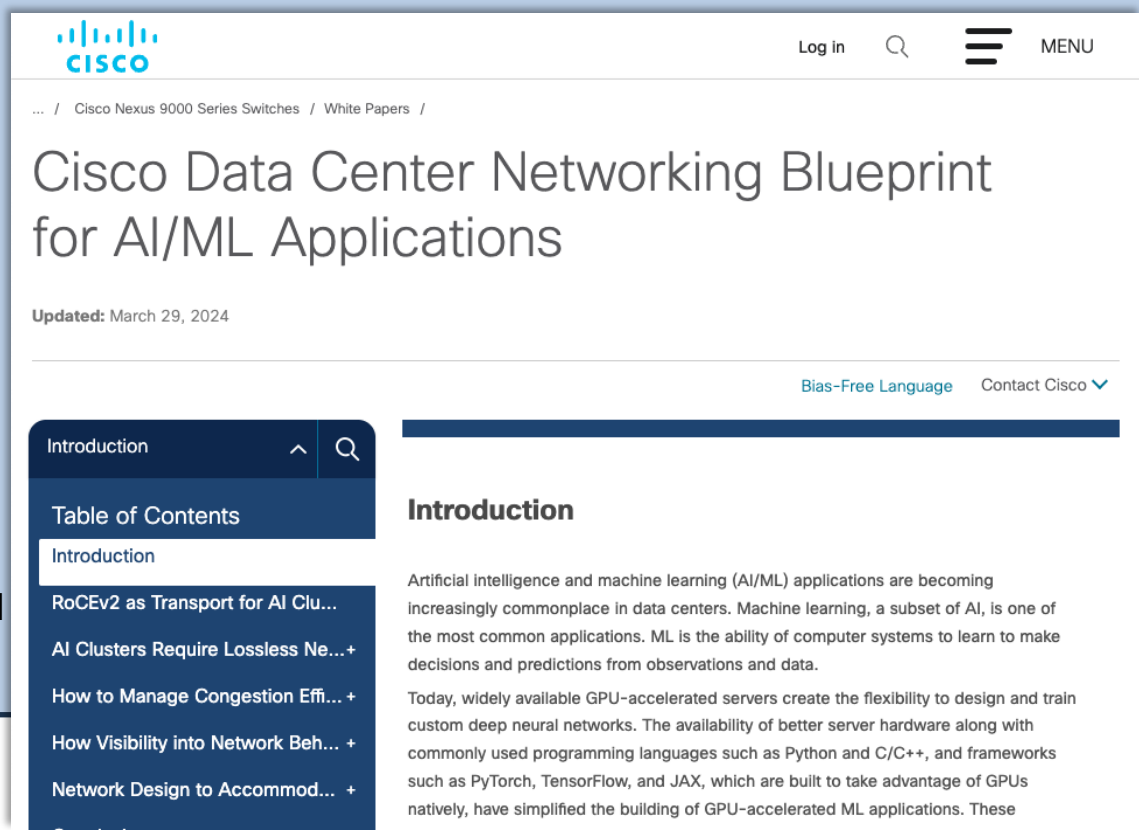
Nexus for AI Ready Data Centers



AI/ML Data Center Blueprint for Networking

Addressing needs of the most demanding datacenter production networks

- Performance for AI/ML networks
- Intelligent buffering, low latency, telemetry/visibility, RoCEv2
- Easily build lossless fabrics
- Efficient application performance
- Lower TCO
- Validated designs for network and ecosystem partners



cisco Connect

Introducing The First 51.2T Switch

Cisco Nexus 9364E-SG2 - 64p 800G Fixed Switch



Compact 2RU 51.2T Switch

G200 ASIC (5nm) | 112G SerDes | 256MB packet buffer

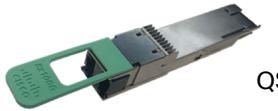
64 800G capable ports | Up to 128 line-rate 400G ports (2x400G breakout)

QSFP-DD800 ports backward compatible with QSFP-DD, QSFP28, QSFP+

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

Cisco NXOS spine and AI/ML leaf capable

Introducing Cisco QSFP-DD800 modules



QSFP-DD 8x100G FR

QSFP-DD 2x400G FR4



Increased density	Double port bandwidth for Single Mode Fiber links up to 2km
Investment protection	Reuse existing cabling infrastructure: Dual Duplex LC and Dual MPO-12 SMF connectors
Backwards compatibility	Connect existing pluggable transceivers: QSFP+, QSFP28, QSFP56, QSFP112, QSFP-DD
Flexible design support	800G port to port | Breakout to 400G or 100G ports
Improved sustainability	Supports over 30W of power dissipation and riding heatsink in host platform
Standards compliant	QSFP-DD 800 MSA, IEEE 400GBASE-FR4, 100GBASE-FR1

Cisco Nexus 9300 Series – 400G Fixed Switches

N9K-C9364D-GX2A



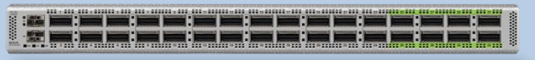
- Single 25.6T ASIC
- 2 RU 64 400G ports
- MACsec on 16 ports

N9K-C9348D-GX2A



- Single 25.6T ASIC
- 2 RU 48 400G ports
- MACsec on all 48 ports

N9K-C9332D-H2R
N9K-C9332D-GX2B



- Single 12.8T ASIC
- 1 RU 32 400G ports
- MACsec on all 32 ports (H2R)
- MACsec on 8 ports (GX2B)

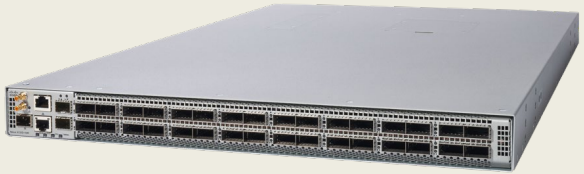
Cisco Nexus 9300 Series – 1RU 32p 400G Switches

N9K-C9332D-GX2B



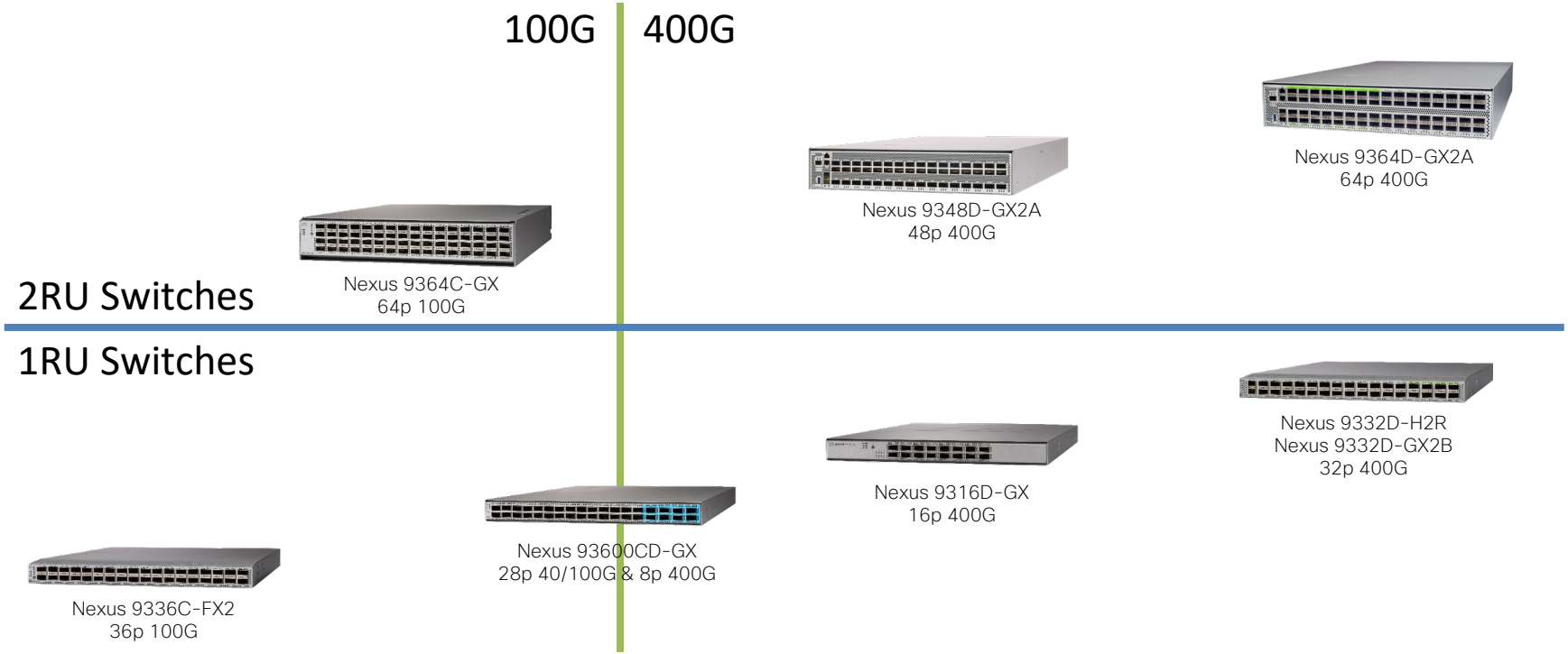
- 32-port 400G 1RU switch
- 120MB on-die packet buffer
- Line-rate MACsec on 8 ports
- ZR and ZR+ support
- 4x 100G, 4x 25G, and 4x 10G breakout
- 6K ingress and 3K egress TCAM
- Port-side intake and port-side exhaust
- ACI Spine, ACI Leaf, and NXOS

N9K-C9332D-H2R



- 32-port 400G 1RU switch
- 80MB on-die packet buffer + 8GB HBM packet buffer
- Line-rate MACsec on all ports
- ZR and ZR+ support
- 4x 100G, 4x 25G, and 4x 10G breakout
- Sync-E, IEEE 1588, PTP, Class-C Timing
- 14K shared TCAM
- Port-side intake
- NXOS (Shipping)
- ACI Spine and Leaf (Roadmap)

Cisco Nexus 9300 Series – 100G/400G Fixed Switches



Cisco Nexus 9000 Series – 10G/25G/50G Fixed Switches

RJ45 | SFP

2RU Switches

1RU Switches



Nexus 93216TC-FX2
96p 1/10GG + 12p 40/100G



Nexus 93360YC-FX2
96p 10/25G + 12p 40/100G



Nexus 93240YC-FX2 (1.2RU)
48p 10/25G + 12p 40/100G



Nexus 93108TC-FX3P/Nexus 93108TC-FX3
48p 1/10GT + 6p 40/100G



Nexus 93180YC-FX3
48p 10/25G + 6p 40/100G



Nexus 93400LD-H1
48p 50G + 4p 400G



Nexus 92348GC-FX3
48p 100M/1GT + 4p 1/10/25G + 2p 40/100G
(No VXLAN)



Nexus 9348GC-FX3/Nexus 9348GC-FX3PH
48p 100M/1GT + 4p 1/10/25G + 2p 40/100G

Cisco Nexus 9000 Series – Modular Switch Evolution



Nexus 9408

Optimized for
100G/200G/400G

Compact



Nexus 9500

Optimized for 1G through 400G

4-slot, 8-slot, 16-slot chassis

800G Ready
High Port Density



Nexus 9800

Very high density 400G & 100G ports

800G ready

4-slot and 8-slot chassis

Customer Challenges

Mounting pressure to rethink
data center architecture for the AI era

Complex, distributed environments

Service-specific devices drive architectural complexity and higher operational costs

Data Center Modernization

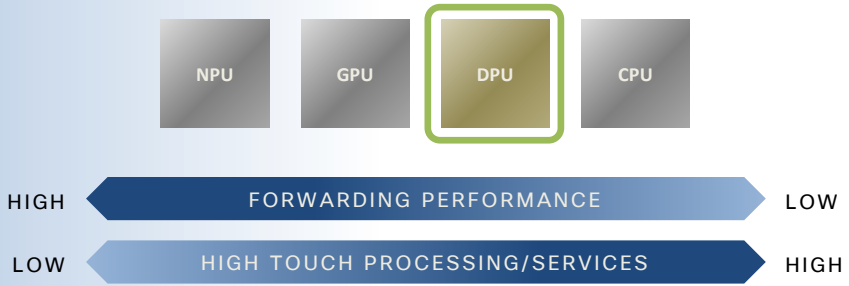
Faster switches aren't enough to meet the needs of modern data centers

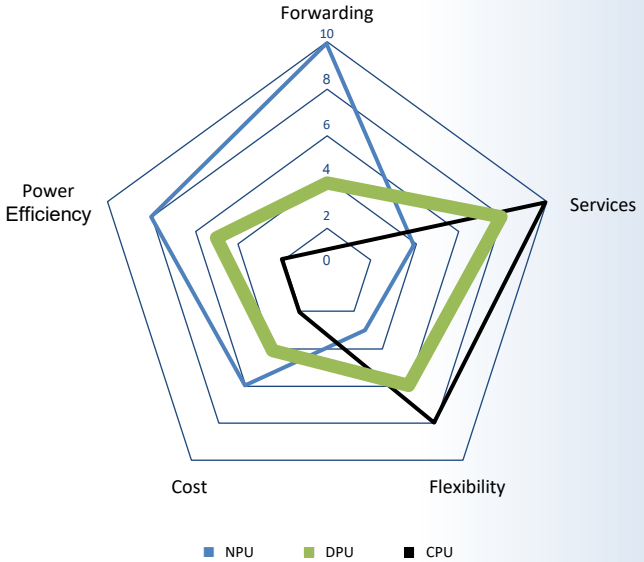
Security

Inconsistent enforcement due to lack of granular placement control

Data Processing Units

A game changer to deliver network services at scale





Higher scores are better

Cisco and AMD – Better Together

Unmatched flexibility, performance, and efficiency



- 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



- 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

Large-Scale NAT | IPSEC Encryption | Distributed Firewall | Event-Based Telemetry | DoS Protection

Cisco N9300 Series Smart Switches

Best of breed platforms for data center services

Now Shipping



N9324C-SE1U

24-port 100G

- Cloud Edge, Zone-Based segmentation, DCI, Top-of-Rack
- 800G Services Throughput
- Silicon One E100 ASIC + AMD DPUs

Target General availability July 2025



N9348Y2C6D-SE1U

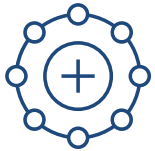
48-port 25G, 6-port 400G, 2-port 100G

- DC Top-of-Rack
- 800G Services Throughput
- Silicon One E100 + AMD DPUs

Fabric Options



Complexity is the new normal



Greater application complexity

750M new applications

by 2025



Accelerating pace of operations

Infrastructure and apps are deployed

56% faster



Surging data from AI workloads

117M TB of data

from AI-lifecycle workload



Mandate to do more with less

Cloud costs up 20%

versus last year



Rising demands for sustainability

In a typical office, data centers consume

<50x power



Expanding attack surface

125% rise

in cyber threats, costing $10T in 2025

ESG Research Report: Observability from Code to Cloud, May 2023; 2023 Global Hybrid Trends Report, 451 Research; IDC, 750 Million New Logical Applications: More Background, December 2021; Cybersecurity Ventures, October 2022; IDC Semiannual Enterprise Infrastructure Tracker: Workloads, 2022H2

Data Center Networking
Unified Approach

Unified Architecture

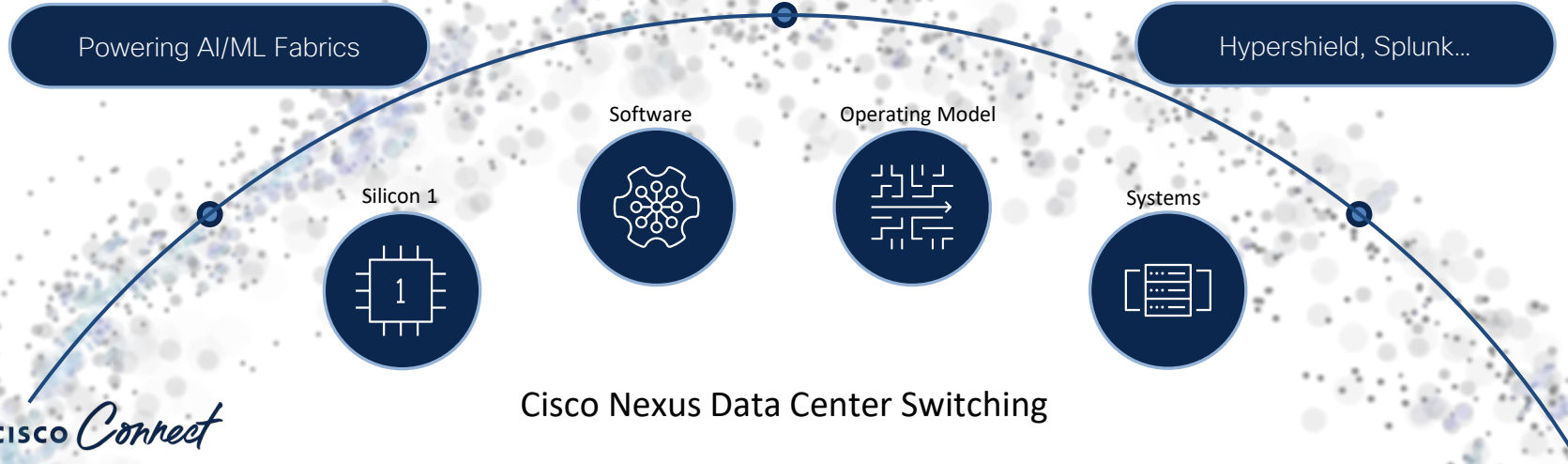
- Flexible Fabric Deployments
- Powering AI/ML Fabrics

Unified Operations

- Nexus One

Unified Intelligence

- Nexus Dashboard
- Hypershield, Splunk...



Cisco Nexus Data Center Switching

Nexus Dashboard



Cisco Nexus Dashboard

Data Center Networking Management

Provision

Quickly Deploy any Data Center Network Fabrics
Deploy traditional LAN, EVPN VXLAN, AI/LM and SAN fabrics

Manage

Multisite fabric and device management in a single view
Device upgrades, Change control with roll-back, traffic steering, ect…

Analyze

Identify, troubleshoot and resolve issues quickly
Traffic Analytics, Delta and pre-change analysis

Segmentation

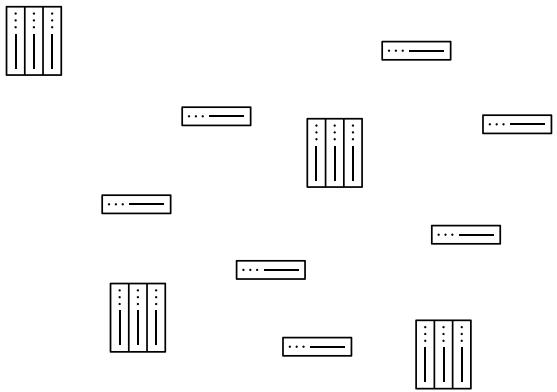
Unify data center networks across both ACI and NX-OS
Leveraging the capabilities of Cisco Nexus One

Simplified Network Management
for modern data center
architectures



Provision a new fabric in minutes

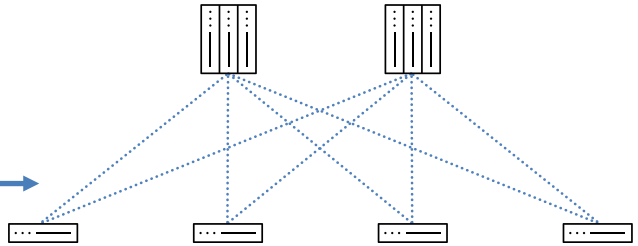
Un-provisioned switches



Nexus Dashboard

Fabric Builder

Cisco best practice implemented



Fast, automated process

Benefits

Accelerate fabric deployments

Automated consistency

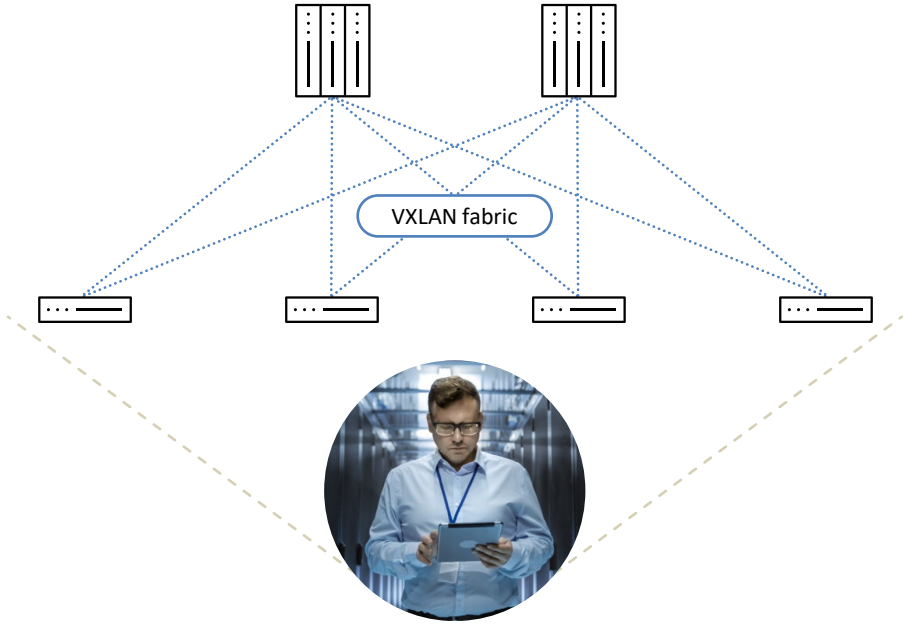
Minimize risk

Automate VXLAN-EVPN deployments

Cisco best practice templates for VXLAN EVPN templates

Fabric builder

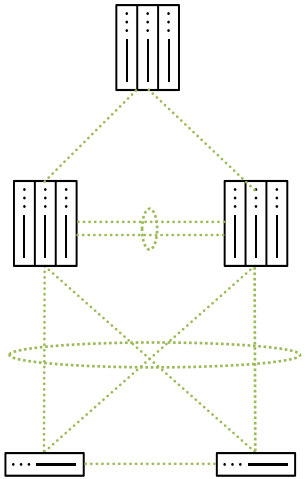
Support for both brownfield and greenfield deployments



Benefit

Simplify deployment time, reduce chances of errors

Enhanced Classic LAN



Classic LAN Fabric

Fully automated fabric – Enhanced Classic LAN

Support for greenfield and brownfield deployments

Provisioning of 3tier architecture
L2/L3 Networks and VRFs

VRF-Lite Between Agg and Core

Benefits

Best Practice Templates

Simplified workflows

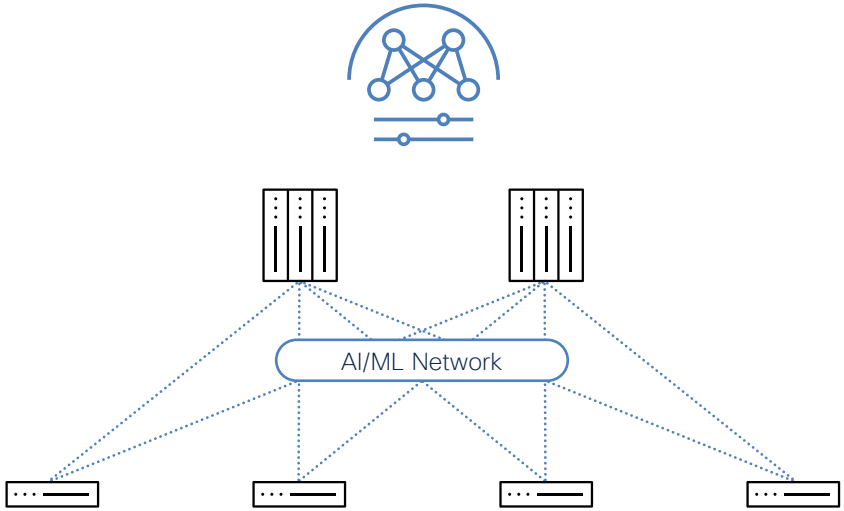
Flexibility based on customer needs

Automated of AI/ML networks with ROCEv2

Provision routed (eBGP)
or VXLAN (iBGP) AI Network

Automatically create QoS policies
for AI/ML workload

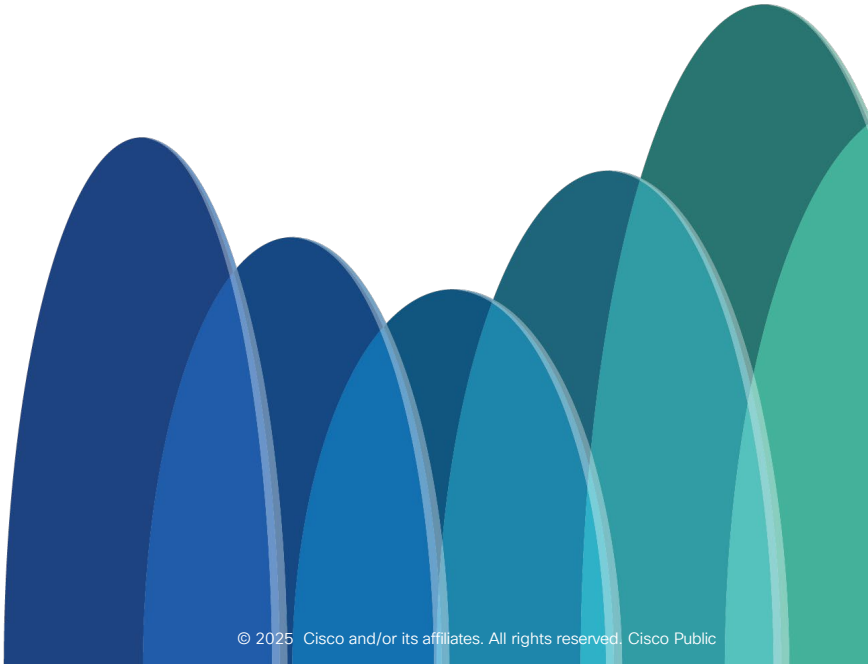
Single Point of automation and
provisioning for AI/ML deployments



Benefits

Unified and Seamless Management across various platforms

Demo time!



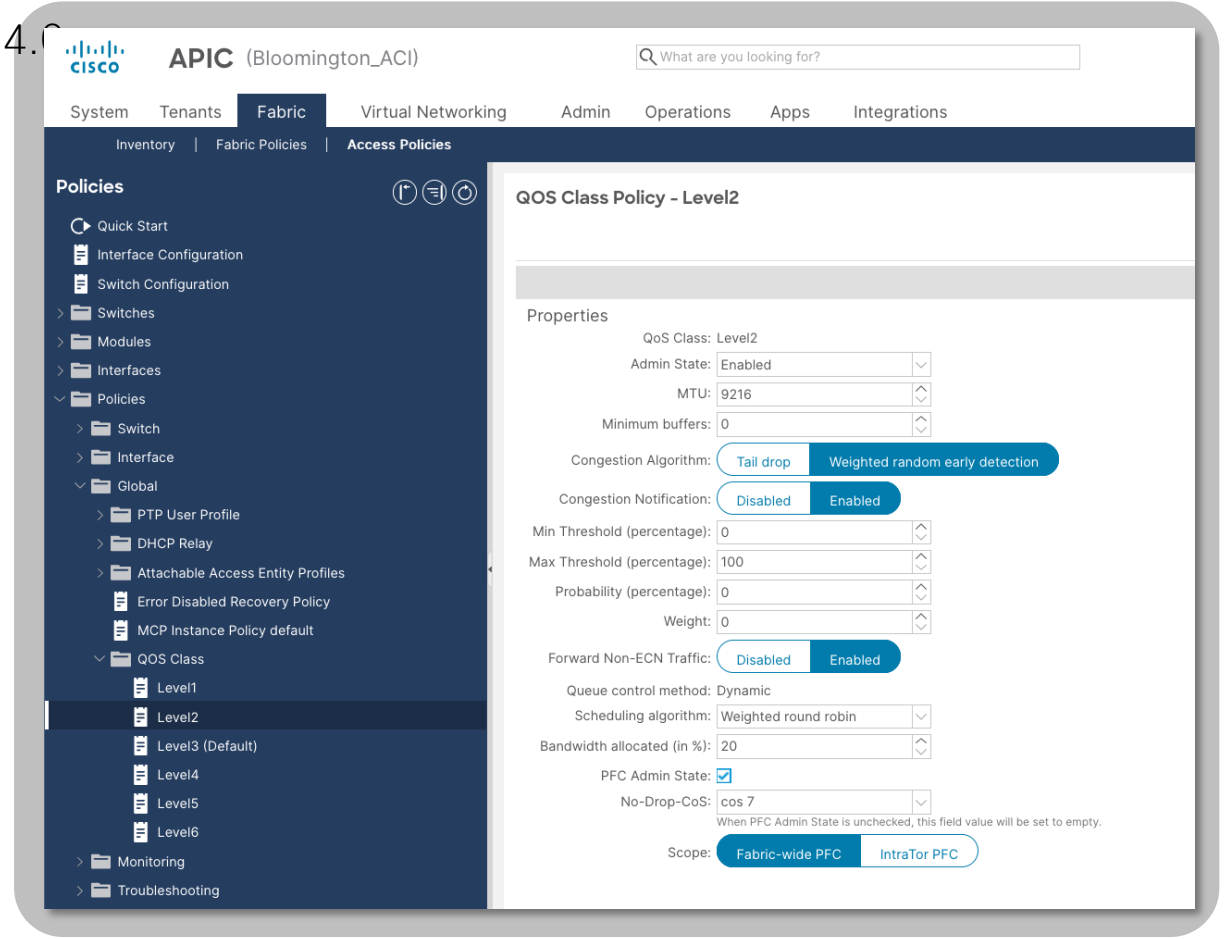
© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Application Centric Infrastructure

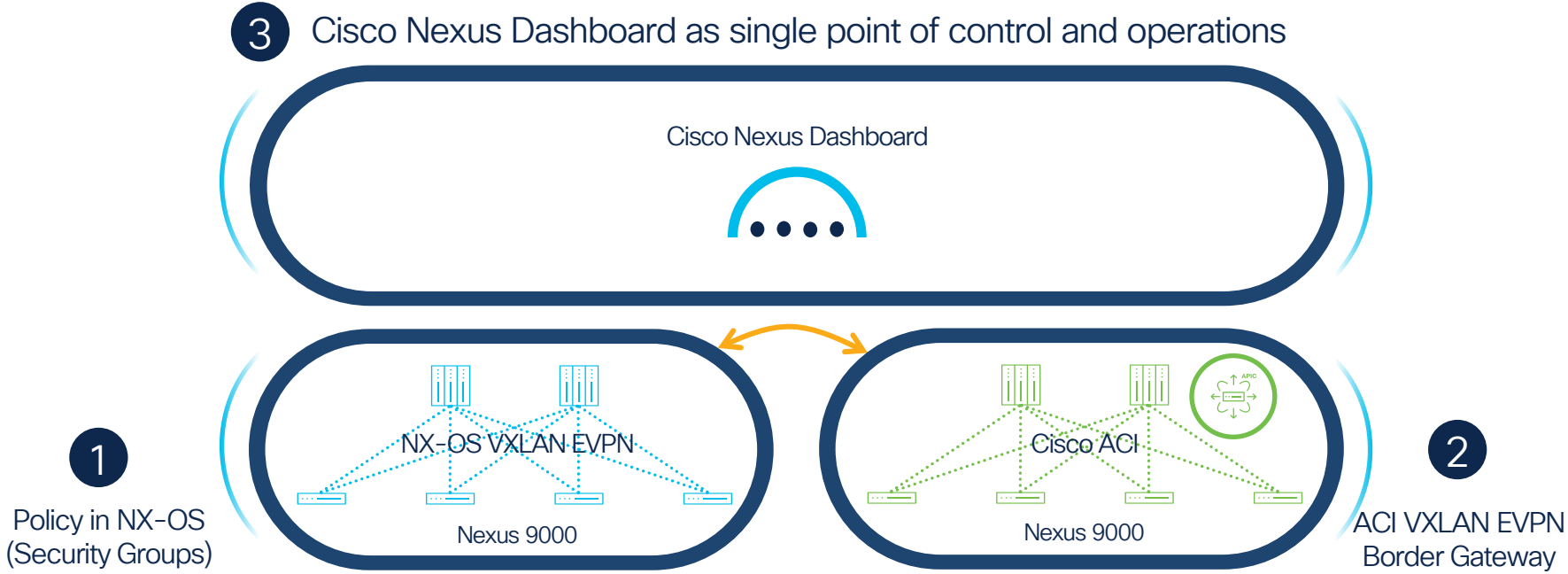


RoCEv2 in ACI

Supported in ACI since version 4.0



Nexus ONE Fabric Experience
Architectural and Functional Components



Different fabric architectures

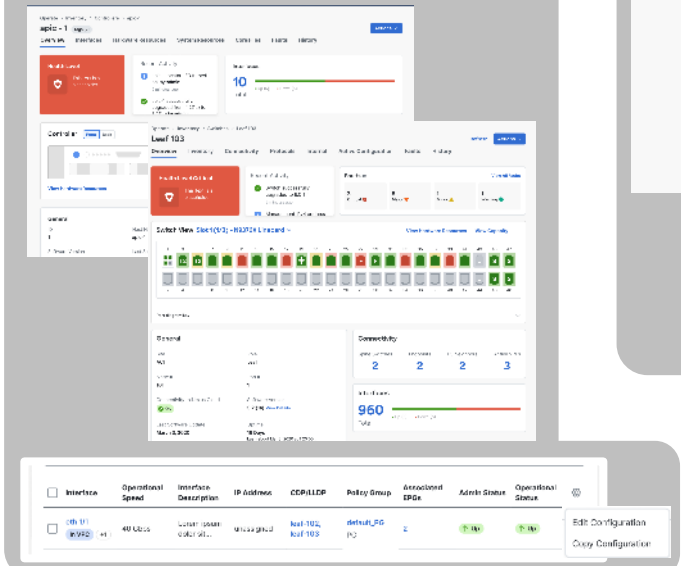
Same outcome with common experience

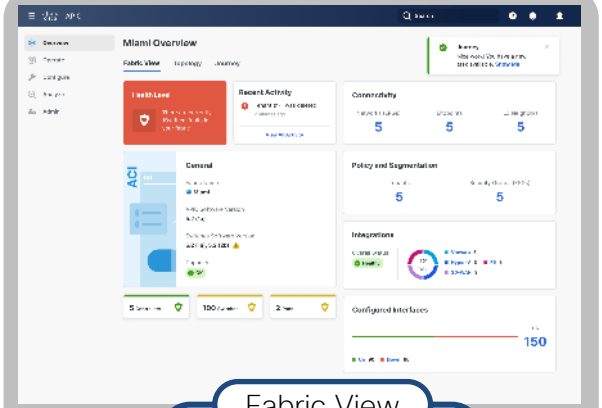
ACI simplification

A modern, faster UI for 2024 and beyond

Find the information you need with ease
Edit configurations faster (Copy + Paste, Clone)
Jointly visualize Policy Groups and EPGs configured on an interface

Switches and Controllers



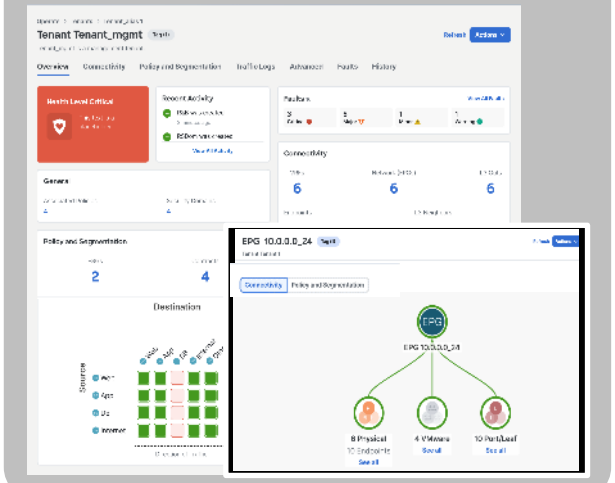


Fabric View

Quick access to main items
In-context information
Reduce jumping between sections

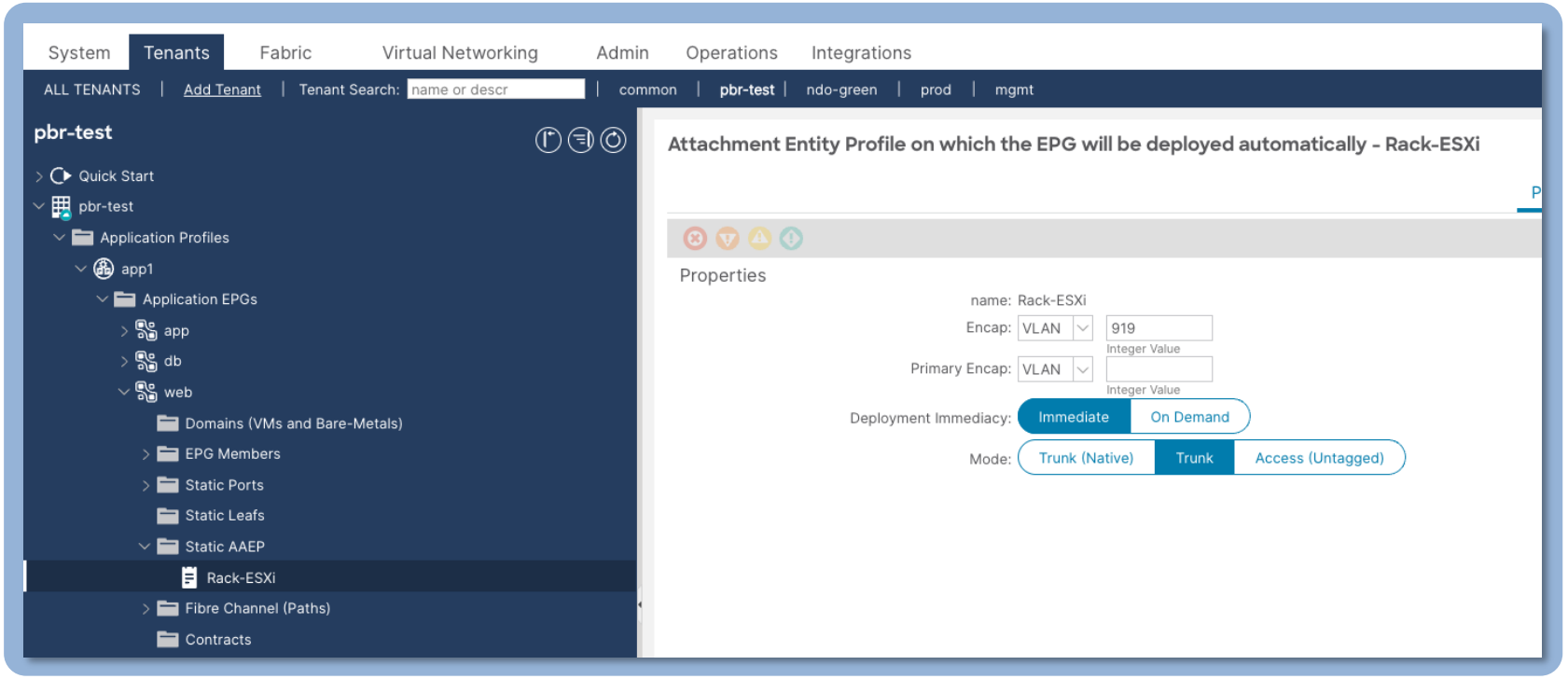
Quickly identify who can talk to whom and how
Associate contracts and L4-L7 devices with ease
In-context security policy and/or domain association for every EPG/ESG

Tenants, EPGs & contracts



cisco Connect

Static AAEP under the EPG



Available in ACI 6.1(3f)

Cisco ACI Areas of Investment



Hardware Platforms
Nexus 9800 Spines



APL Certifications &
Quantum Crypto



Common policy: Campus and
Data Center Segmentation
CDA Integration



Multi-Site ESG,
VzAny+PBR & Topology



VXLAN-EVPN
Gateway with Policy



New UI:
Fast & workflow-driven



Virtualization
(Nutanix, RedHat, etc ...)



New Use Cases:
Autonomous remote leaf
group AI/ML fabrics



Hardened Upgrades
Mixed version support



Networking as Code
(Ansible, Terraform, PyACI)



Container networking:
Floating L3out, Observability,
Automation & CNIs



Insights: Complete
data/control-plane
dependencies

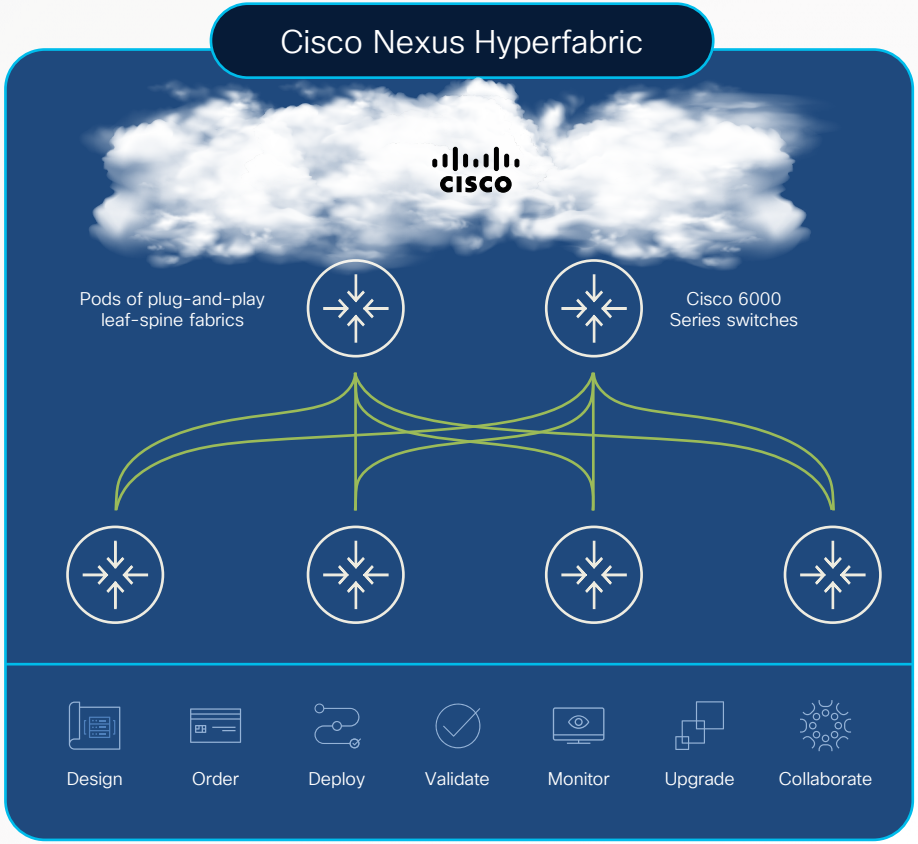
ACI roadmap for 2025 and beyond

HyperFabric



Cisco Nexus Hyperfabric

- Design, deploy and operate on-premises fabrics located anywhere
- Easy enough for IT generalists, application and DevOps teams
- Outcome driven by a purpose-built vertical stack



cisco Connect

Nexus Hyperfabric Components

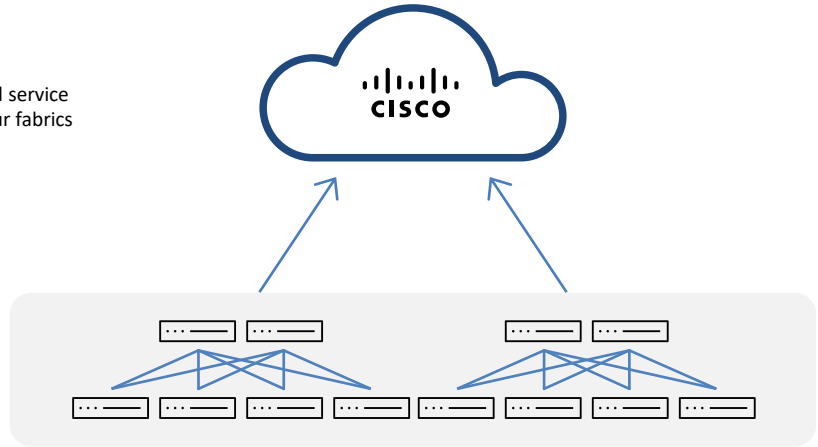
Cloud Controller

- Scalable, globally distributed multi-tenant cloud service
- Design, plan, control, upgrade, and monitor your fabrics
- Browser, API, and mobile access



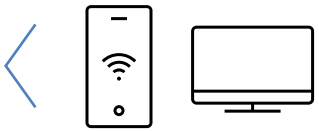
Cloud-managed 6000 Switch

- Boot-strapped from cloud
- Full visibility and control from the cloud



High-performance Fabrics

- Initially thousands of 10/25/100/400 GbE host ports
- EVPN/VXLAN, layer 2 VLANs, IPv4/IPv6 routing
- Mesh and spine leaf fabrics

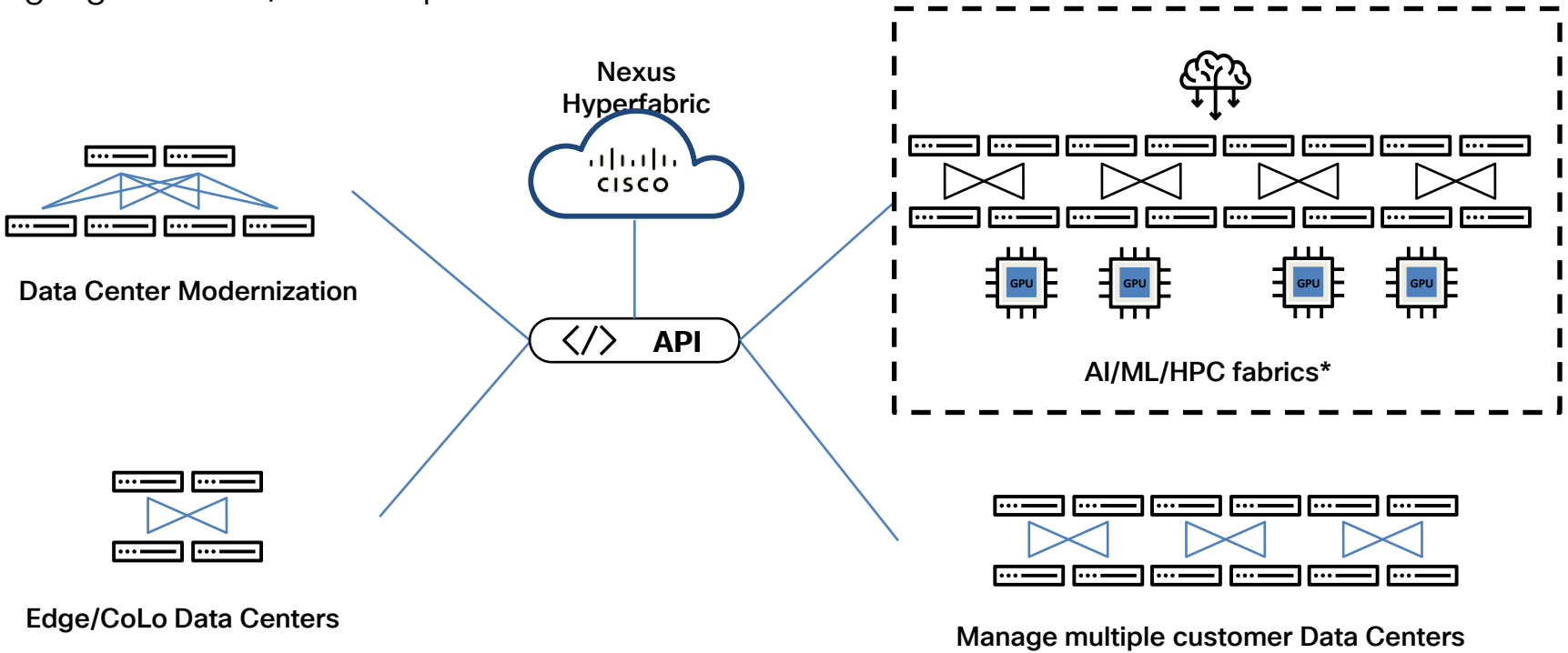


Helping Hands App

- Step-by-step deployment tasks
- Registration and cabling
- Real-time validation

Use Cases

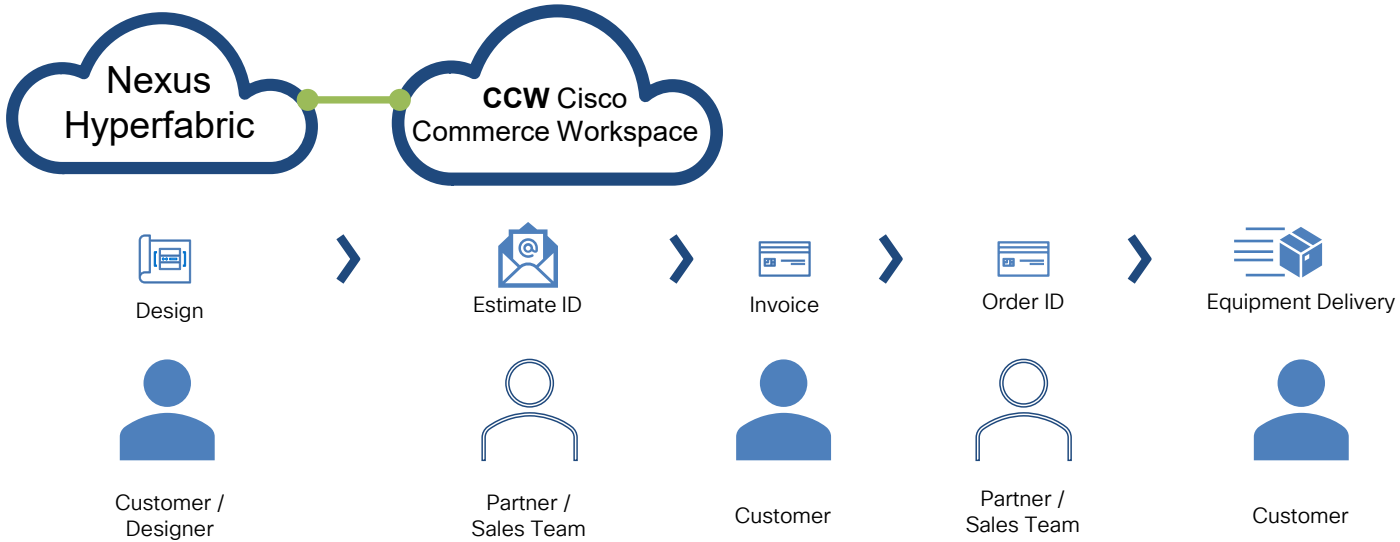
Single global GUI / API endpoint for all owned fabrics



*Nexus Hyperfabric AI – Mid 2025

Design to BOM

Integrated design to BOM feature
reduces errors and shortens transaction time



Flexible Architectures

Full-mesh spine-less fabrics

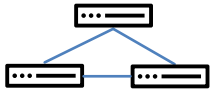
A Fabric of One



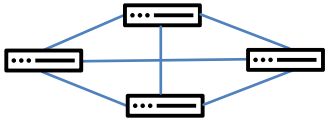
2-switch fabric
120 host ports



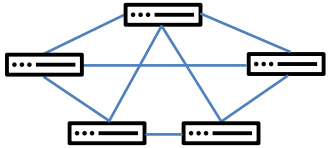
3-switch fabric



4-switch fabric
240 host ports

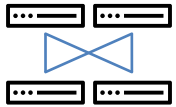


5-switch fabric

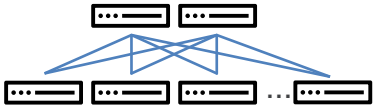


Leaf-spine fabrics

2 spine, 2 leaf



2 or 4-way spine, 2-32 leaf
Nearly 2000 host ports



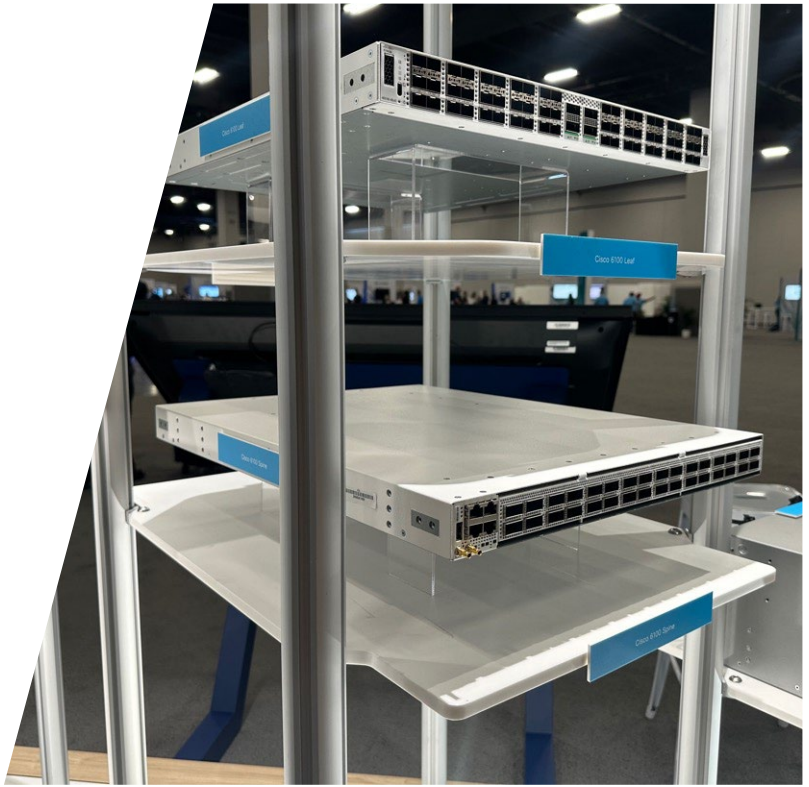
Cisco 6000 series switches

Leaf: HF6100-60L4D

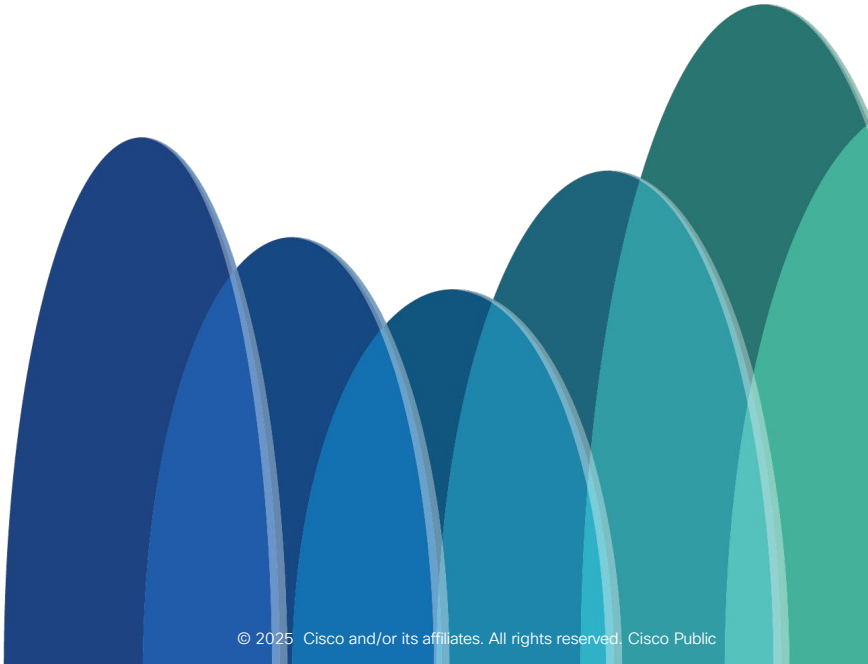
- 4x 100/400GbE QSFP56-DD (16x 100G breakout)
- 60x 10/25/50GbE SFP56

Spine/Leaf: HF6100-32D

- 32x 100/400GbE QSFP56-DD
- 128x 100GbE via 400:100 breakout



Demo time!



© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Cisco Agile Data Center Interconnect (DCI)



"As the backbone of the digital economy, data centers are experiencing unprecedented growth in traffic driven by cloud adoption, AI, and edge computing. To meet these demands, we need to build out massive, high-capacity bandwidth between data centers, ensuring seamless connectivity and performance for businesses and consumers alike. This investment is critical to enabling the future of innovation."— **Sundar Pichai, CEO of Alphabet and Google**



Agile Data Center Interconnect (DCI)

- What is DCI
- DCI Options
- Leveraging dark fiber – Grey Optics
- Coherent (DWDM) optics based
- DCI Scale
- DCI Summary

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

What is DCI

- Data Center Interconnect
  - Connecting two data centers geographically separated
    - Typically, 10km-100km (6miles-60miles)
    - Typically, point-to-point
    - Leased dark fiber or Carrier Services
  - Connecting a data center to a cloud colocation (colo) provider
    - Equinix, Digital Realty, CoreSite, QTS, Zayo, … others
    - High speed direct connect/access to: AWS, Azure, Google, IBM, Oracle … others
  - WEB Scalers use DCI to create regional clusters of multiple DC
    - Switching fabric extended within cluster and requires high bandwidth
  - Enterprise similar to web scalers, but typically lower scale

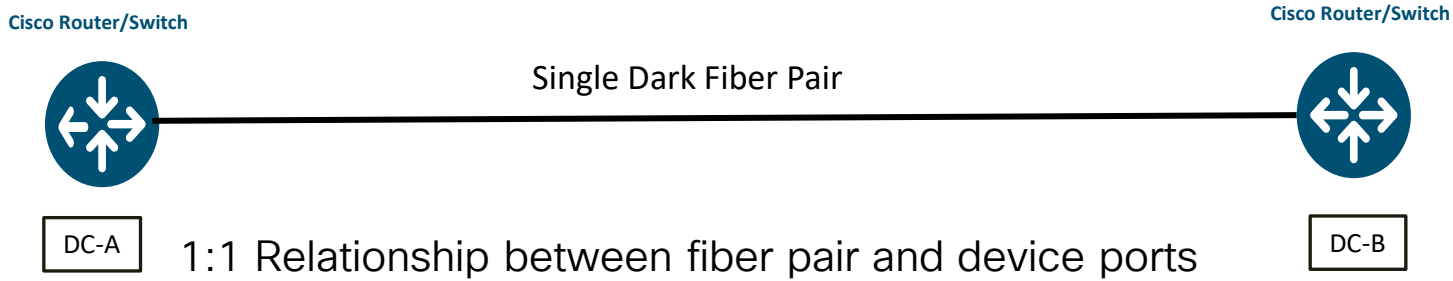
© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

DCI Options

- Three (3) primary common options
  - Dark fiber build – very costly & takes a long time, unlimited bandwidth
  - Lease dark fiber – not cheap, but provides bandwidth leverage options
  - Managed carrier lit service – lowest cost, but bandwidth limited
- Managed carrier lit service
  - Shared Ethernet or MPLS based services for up to 100G circuits
  - Wavelength services at 100G growing to 400G
    - Similar to leasing dark fiber with limited bandwidth
    - Cost effective for Enterprises needing long reach (300-3000 miles)
    - Simple pt-pt handoffs and fully managed, minimal infrastructure cost

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Leveraging Dark Fiber – Grey Optics



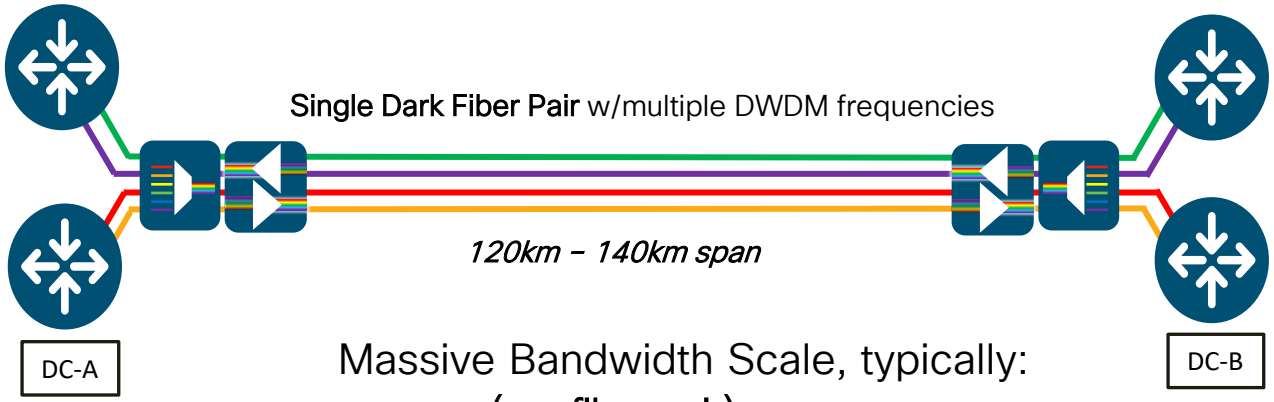
1:1 Relationship between fiber pair and device ports

Distance limited to 40km for 400G and 70km for 100G

Expensive to scale if leasing fiber ($MRC per pair)

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Leveraging Dark Fiber – Coherent Optics



Massive Bandwidth Scale, typically:

(per fiber pair)

- 88x100G – 8.8Tbs
- 55x400G – 22Tbs
- 45x800G – 36Tbs

Cisco Switch/Router



100G, 400G, 800G Coherent



DWDM Passive Filter



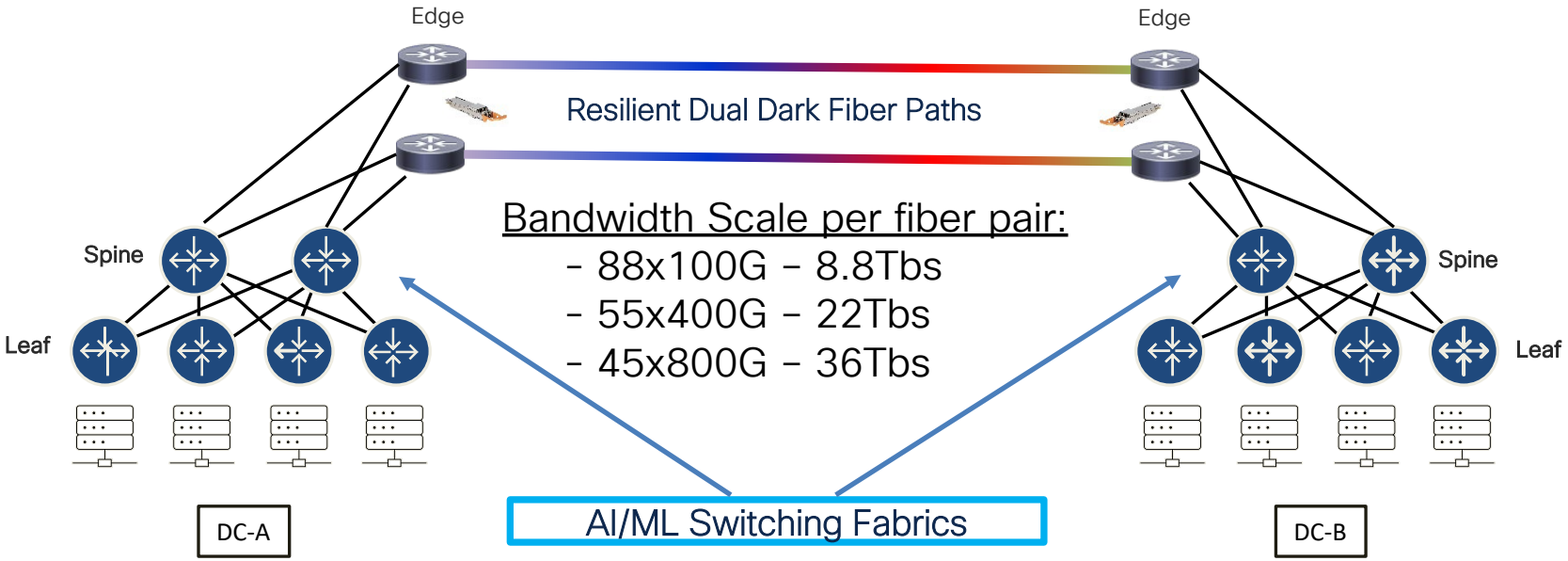
NCS1001 Amplifier



© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

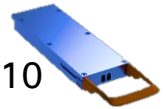
Simple AI Fabric Extension w/Coherent Optics

Coherent 100/400/800G Optics Plug Directly into Edge Router/Switch Ports



© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

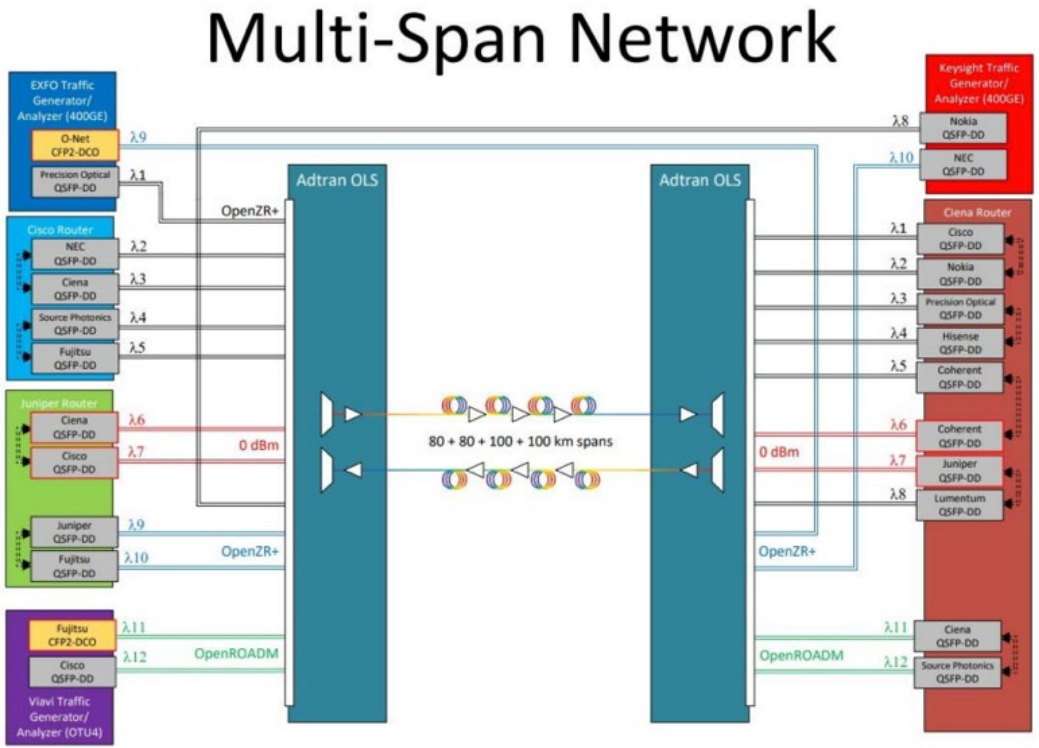
Cisco Acacia Coherent Optics/Modules

	2021	2022	2023	2024	2025
MSA Pluggables Power-optimized	400G(60GB+) 400ZR/ZR+ OpenROADM BiDi 	Bright 400ZR+ 		800G (120GB+) 800ZR/ZR+ 400G ULH 	1.6T (240GB+) 1600LR/ZR/ZR+ 
ASICs	Greylock 400G ZR/Metro/LH — 7nm	Jannu 1.2T DCI/Metro/LH Single-core — 5nm	Delphi 800G LR/ZR/ZR+ — 4nm	NN-1 1.6T ZR/LR — 2/3nm	NN-2 2.4/1.6T DCI/Metro/LH — 2nm
Modules		1.2T (128GB+) C-band  CIM 8	1.2T (128GB+) L-band  CIM 8		2.4T (256GB+) CIM10 

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Cisco's Commitment to Interoperability Testing

Participant in multi-vendor interoperability demos @ EANTC, OFC, ECOC



© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

DCI Summary

- A simple means for DC AI-Fabric Extension
  - Scales to 36Tbs per pt-pt dark fiber span
  - With a reach of ~140km between DC/Colocation
- Cisco Acacia coherent optics keeps it simple
  - Plugs directly into a Router or Switch
  - Simple filters and amplifiers provide leverage & reach
- Coherent optics are high quality
  - Extensive testing with all Cisco platforms
  - Extensive testing with 3rd party platform and optics

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Stay Connected

- Davis Creemer
  dcreemer@cisco.com

- Alan Gehami
  agehami@cisco.com

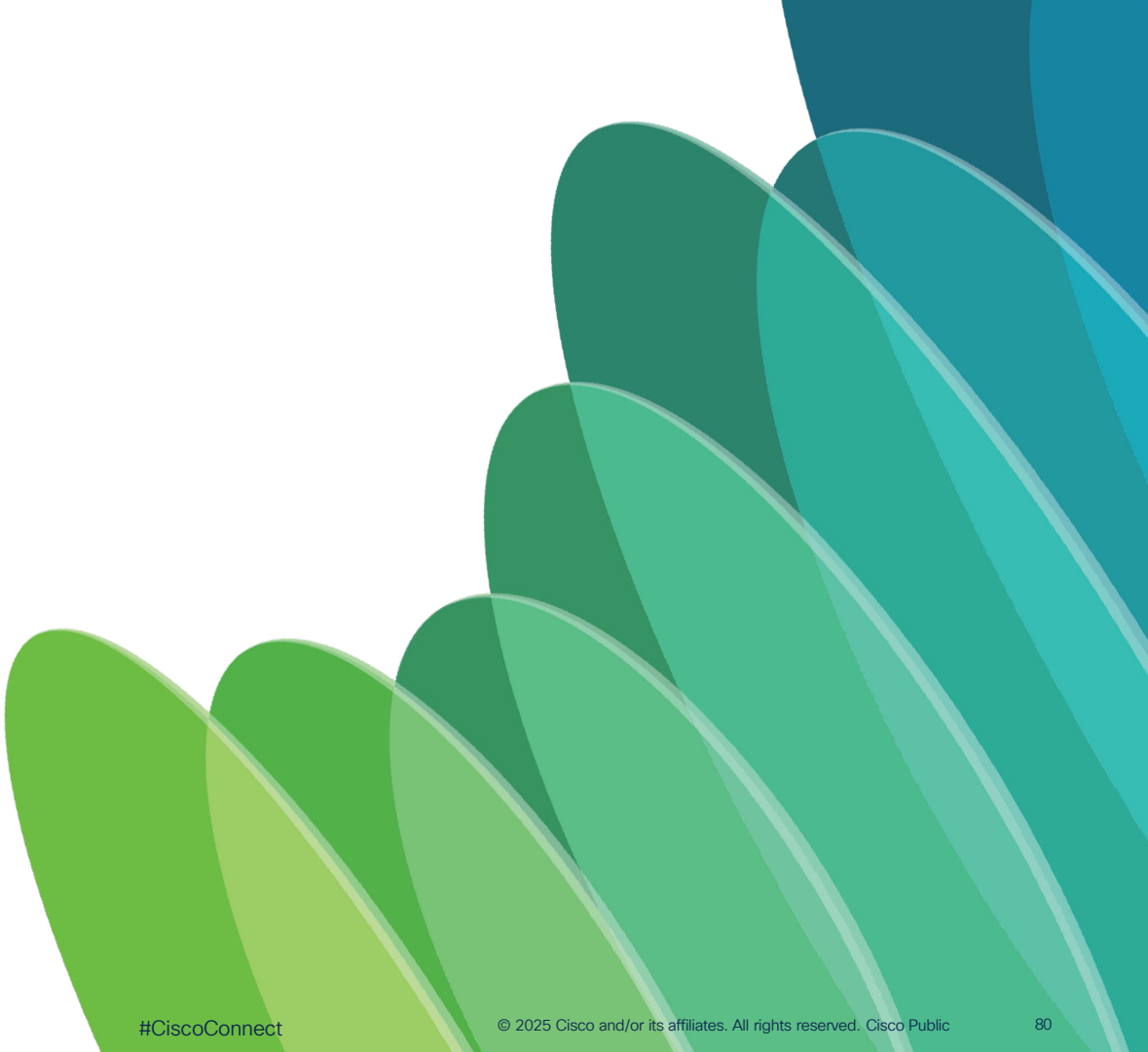
- Today's presentations will be posted in 2 weeks

- Continue the Conversation

Contact me via email

cisco Connect

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public



Thank you

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public