

Ethernet Fabrics for AI Clusters

CISCO Live !

Building high performance, scalable and non-blocking Ethernet fabrics to power the future

Christopher Stokes
Principal Engineer

Laxmi Vellanki
Technical Leader

Cisco Webex App

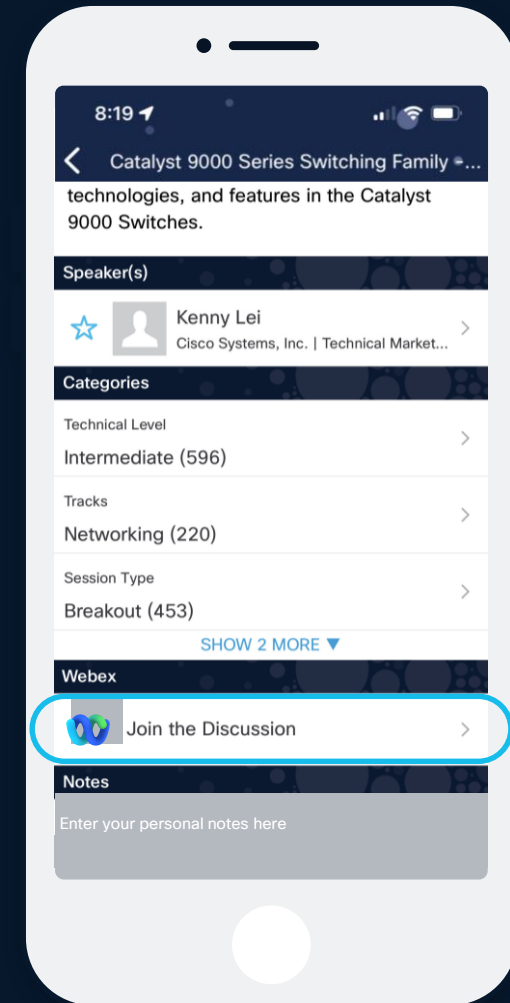
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 13, 2025.



About your Speaker

- Christopher Stokes
- Principal Engineer
- Based out of Austin, TX
- Cisco employee since 2007
- 25 years in IT industry
- Focus on DC Networking

How I spend my free time:

- Family
- Fishing
- “Fixing” Things 😊



About your Speaker

- Laxmi S Vellanki
- Network Engineer
- Cisco employee since 2008
- 17+ years in IT industry
- Focus on DC Networking

How I spend my free time:

- Gardening
- Playing Volleyball
- Being a Mom



Agenda

- 01 Importance of AI to Cisco
- 02 The Business Priority
- 03 Requirements
- 04 Architecture and Design
- 05 Build Implementations
- 06 Lessons Learned
- 07 What's Next?

Why is AI Important?

- Changing the world
- It's everywhere
- New market opportunities
- Improving our products
- Empower our services



Business Priority

Ask from Leadership

Build AI-Ready Data Centers leveraging Cisco platforms and Ethernet technology to support Cisco product and business innovation, reference architectures and IT operational excellence

- **AI Infrastructure Validation:**
 - Deploy clusters to validate design using Cisco's technology stack
- **AI Cluster Deployments:**
 - **256 NVIDIA H100 GPUs:** Performance benchmarking for high-demand workloads
 - **512 NVIDIA H200 GPUs:** Enable scalable AI for business applications
 - **96 NVIDIA H200 GPUs in UCS C885 servers:** Power Cisco's internal cloud workloads
 - **HyperFabric-Enabled Cluster:** Optimize AI performance with ultra-low latency interconnects
- **Support AI Use Cases:**
 - Facilitate engineering and business AI model training
 - Enable **LLM (Large Language Model)** use cases across Cisco products
 - Power AI applications for business and IT automation
- **GPU as a Service:**
 - Provide scalable, on-demand GPU resources to meet diverse AI needs

Evolving AI Infrastructure: From SONiC to Nexus

Cluster 1: SiliconOne/H100

32 x Nvidia DGX (256 GPU)

Technology:

Cisco 8100 with Silicon One ASICs
SONiC Operating System
Nvidia H100 GPUs
NetApp AFF A900 nodes

JUN
'24

Cluster 2: Nexus/H200

64 x Nvidia DGX (512 GPU)

Technology:

Cisco Nexus 9000 with NX-OS
VXLAN DC Fabrics
NDFC and NDI for Fabric Build and Management
Nvidia H200 GPUs
NetApp AFF A900 nodes

JAN
'25

Cluster 3: Nexus/UCS

12 x UCS C885 (96 GPU)

Technology:

Cisco Nexus 9000 with NX-OS
VXLAN DC Fabrics
NDFC and NDI for Fabric Build and Management
Nvidia H200 GPUs
NetApp AFF A900 nodes

APR
'25

Cluster 4: Nexus/H200

32 x Nvidia DGX (256 GPU)

Technology:

Cisco Nexus 9000 with NX-OS
VXLAN DC Fabrics
NDFC and NDI for Fabric Build and Management
Nvidia H200 GPUs
NetApp AFF A900 nodes

MAY
'25

Cluster 5: HyperFabric AI

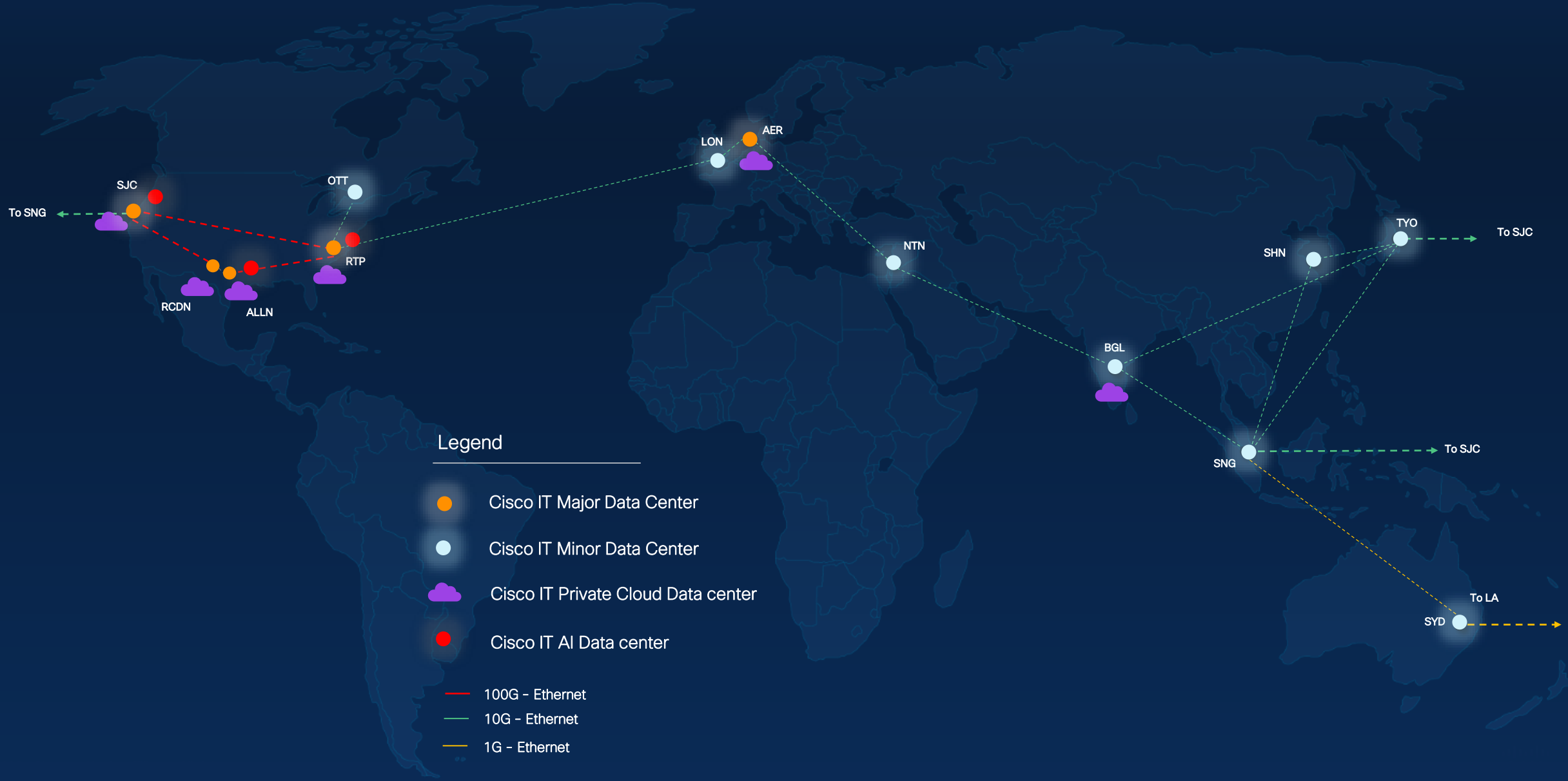
32 x UCS C885 (256 GPU)

Technology:

Silicon One ASICs
VXLAN DC Fabrics
HyperFabric AI controller Build and Management
Nvidia H200 GPUs
VAST Storage

AUG
'25

Cisco IT- DataCenter Presence



AI Cluster Fabric Requirements

Where We Began

Learning and Upskilling the Team

- Internal/External Research
- DevNet
- Training VODs/Classes
- Procured Lab Test Devices
- Partnerships with Cisco and NVIDIA Engineering Leadership
- Divide and Conquer

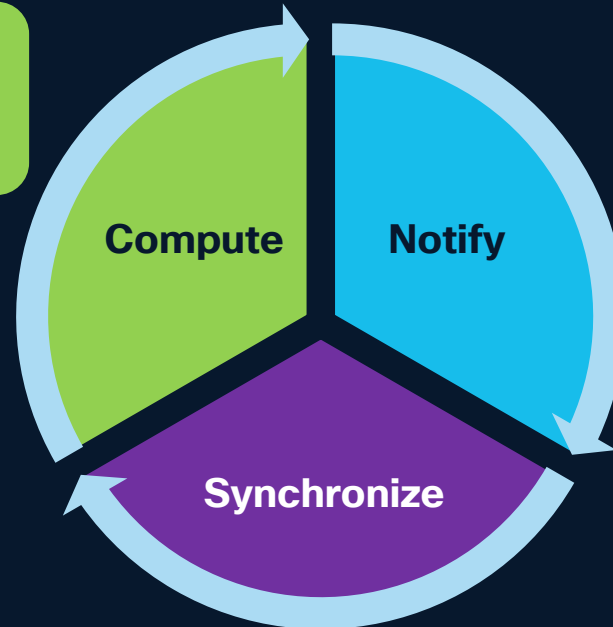
Unique Challenges of Building AI Fabrics

- Low entropy -> inefficient load balancing
- Stringent latency requirements for some AI jobs
- Need for lossless fabric
 - Training jobs could run for long periods
- Very high bandwidth flows
 - Large amount of data sets being moved around
- Synchronized and bursty traffic causes transient oversubscription
 - Links can get saturated in microseconds
- Versatility challenge
 - Different traffic patterns, workloads and communication algorithms
- Security requirements
 - Massive volumes of data, need protection for the entire model/AI application, runtime protection and input validation

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)

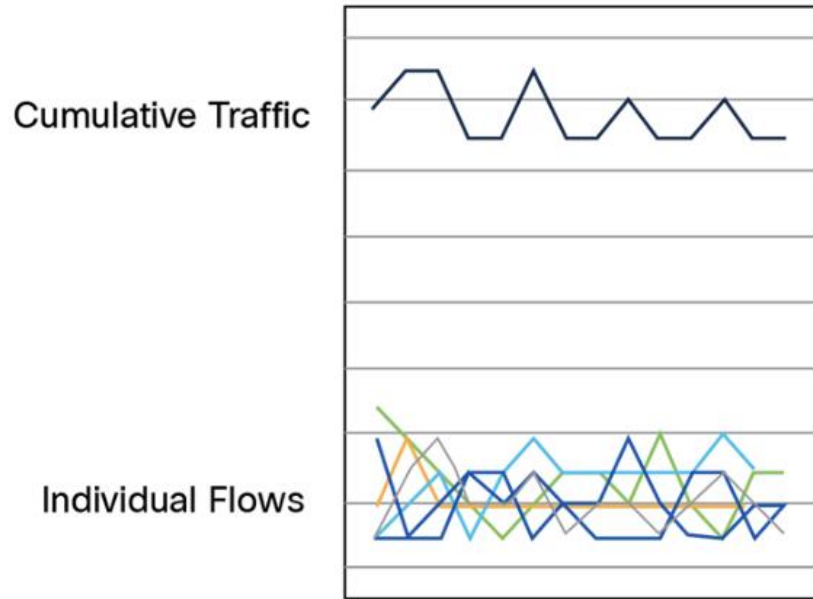
Wait for all GPUs to complete

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

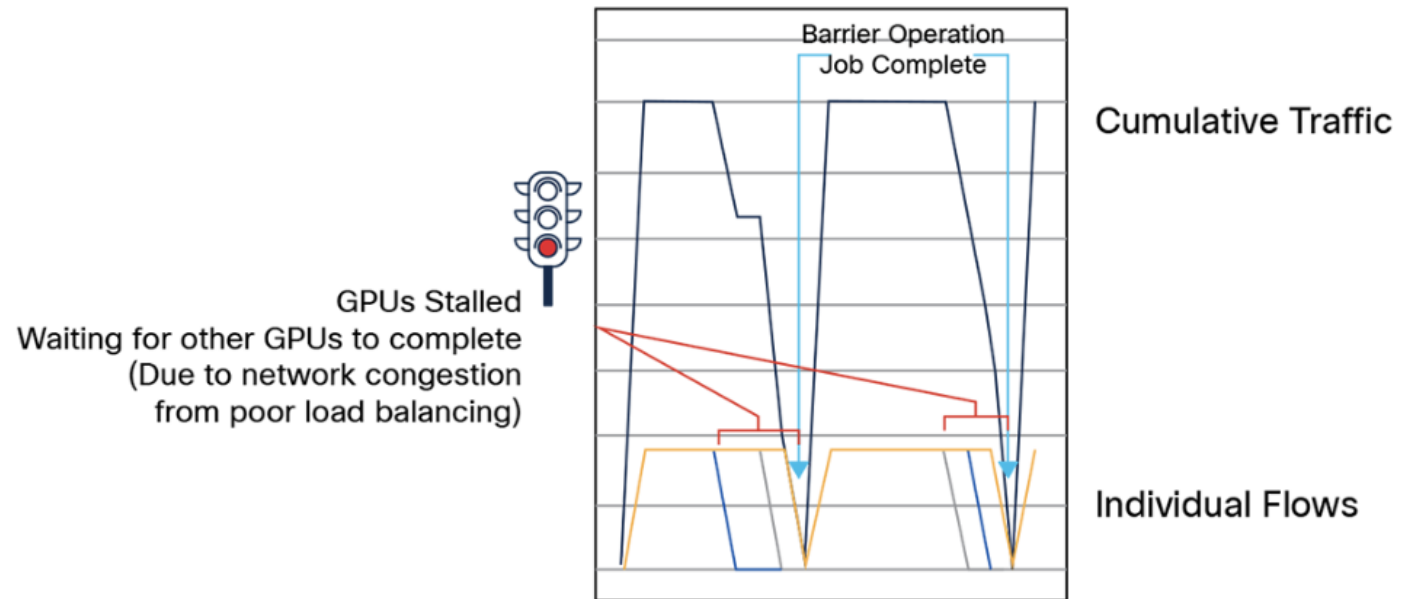
The network can become the bottleneck

Traditional DC Traffic Pattern



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

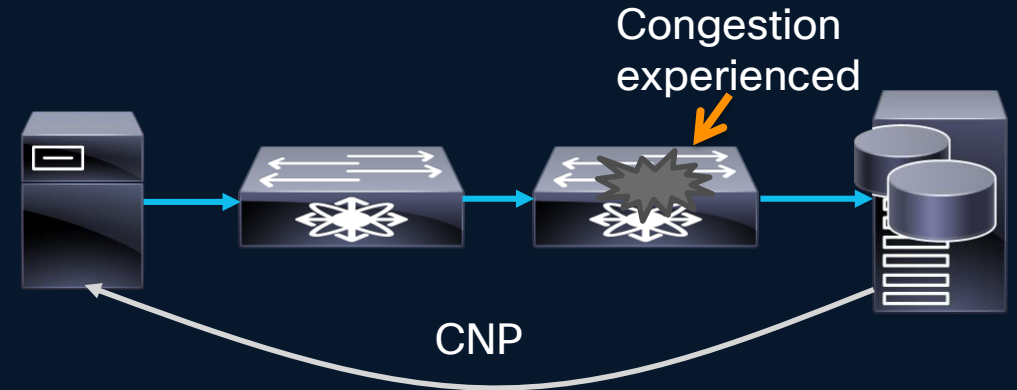
AI (All-to-all Collective) Traffic Pattern



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

Explicit Congestion Notification (ECN)

- **Purpose:** Provides congestion notification in IP networks
- **Function:** Enables end-to-end congestion notification between two endpoints
- **Mechanism:**
 - Utilizes 2 LSB of the Type of Service field in the IP header
 - Upon congestion, it triggers the transmitting device to reduce the transmission rate using a Congestion Notification Packet (CNP) without halting traffic

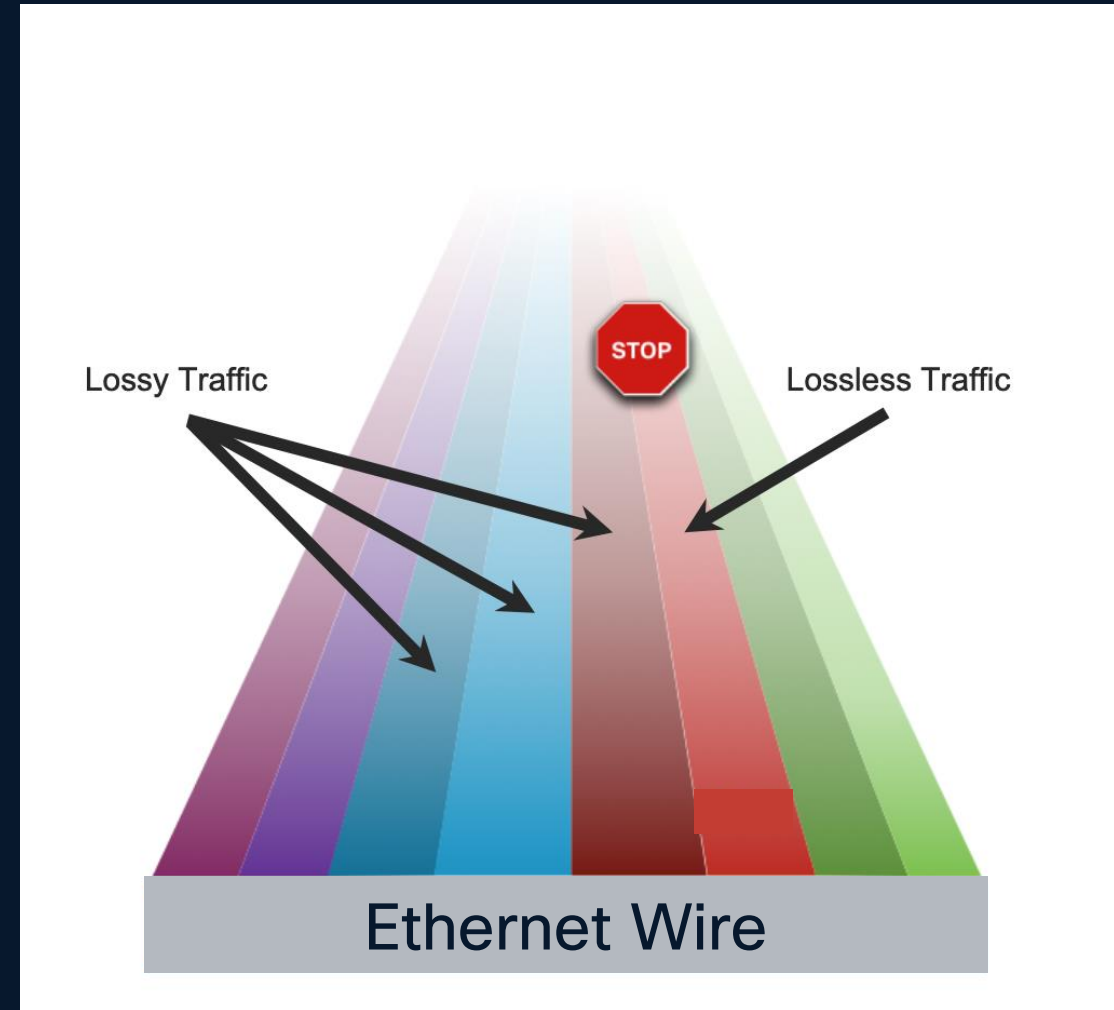


ECN	ECN Behavior
00	Non ECN Capable
10	ECN Capable Transport (0)
01	ECN Capable Transport (1)
11	Congestion Encountered

Priority Flow Control (PFC)

Flow Control Mechanism – 802.1Qbb

- **Also Known As:** "Lossless Ethernet"
- **Function:** Enables Flow Control on a Per-Priority basis
- **Alternate Name:** Per-Priority-Pause
- **Key Benefits:**
 - Allows both lossless and lossy priorities on the same wire.
 - Ensures traffic can operate over a lossless priority independently.
 - Other traffic on different priorities continues to transmit, relying on upper layer protocols for retransmission.



AI Ethernet Fabric LB Options

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

Foundational Principles for our AI Clusters



Focus on application performance



Intelligent buffering, low latency, telemetry / visibility



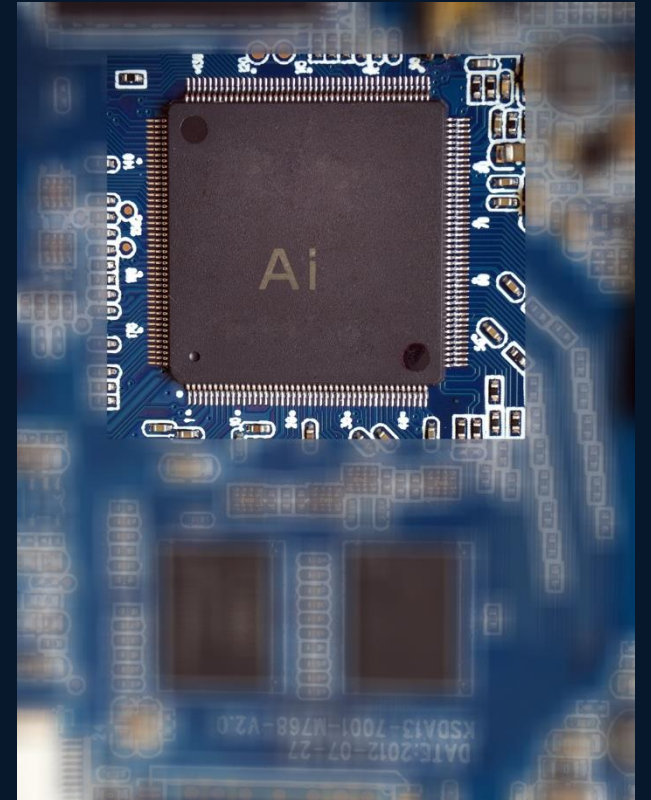
Dynamic congestion avoidance for various workloads



Dedicated front-end / back-end networks. Non-blocking fabric



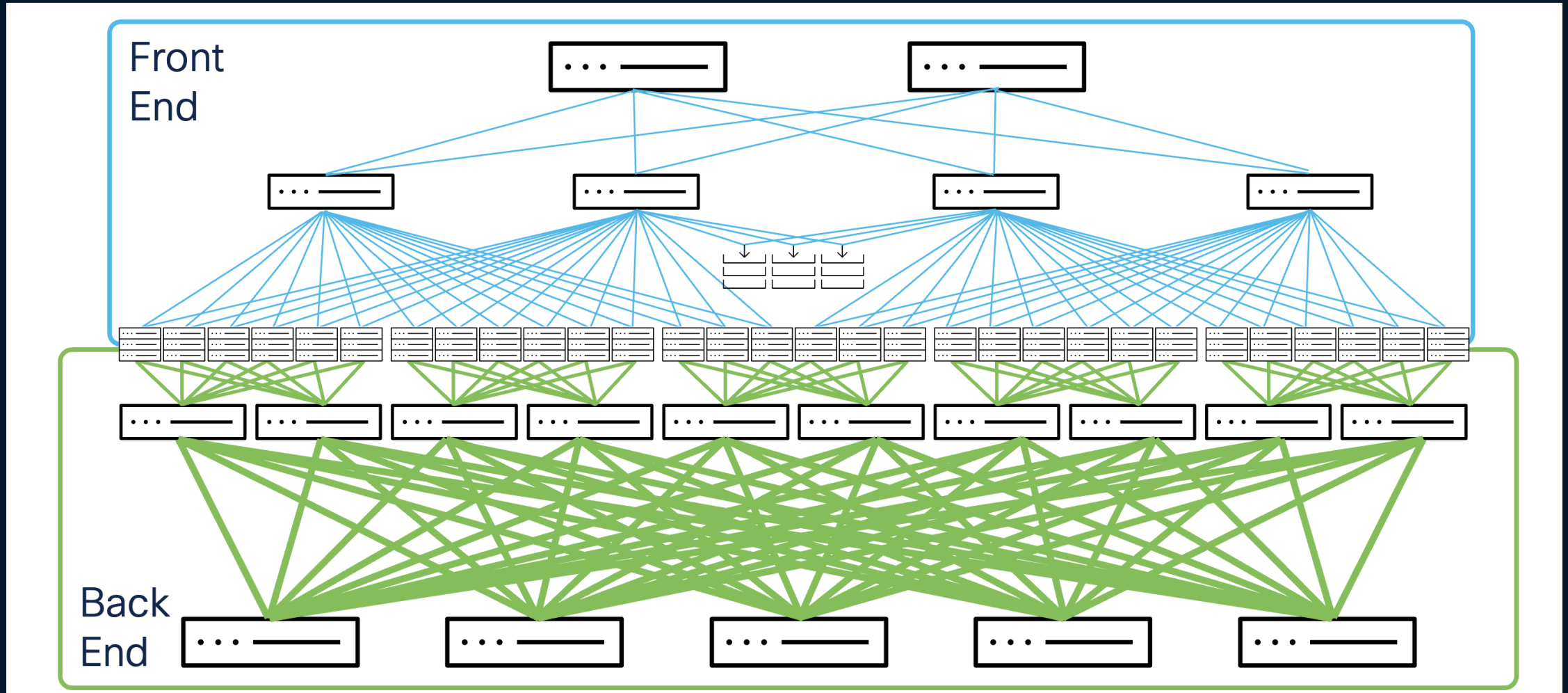
Leverage Automation for Day 0 to 2 operations



AI Fabric Architecture and Design

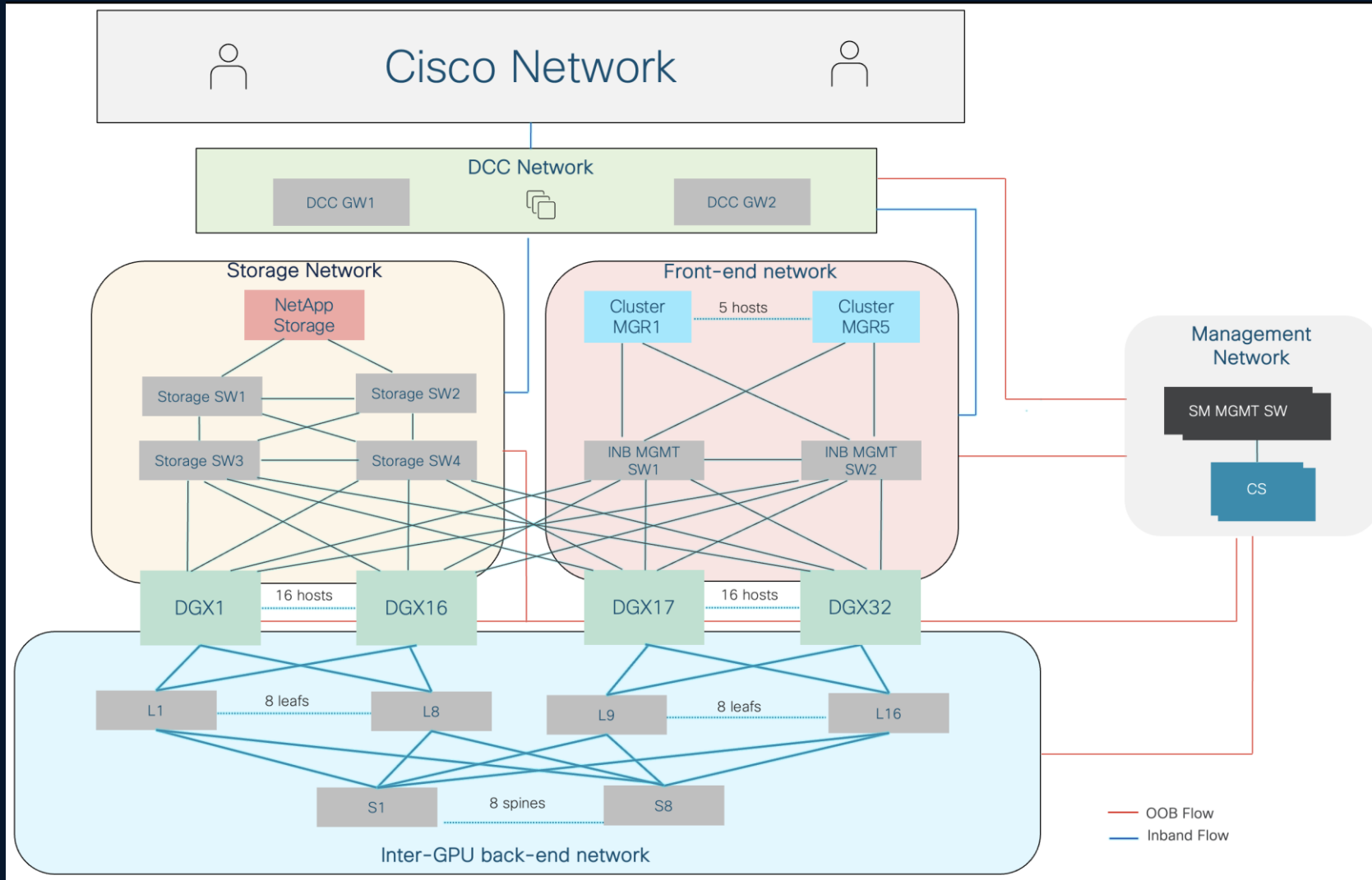
AI Cluster Fabrics

Scalable Dedicated Environments



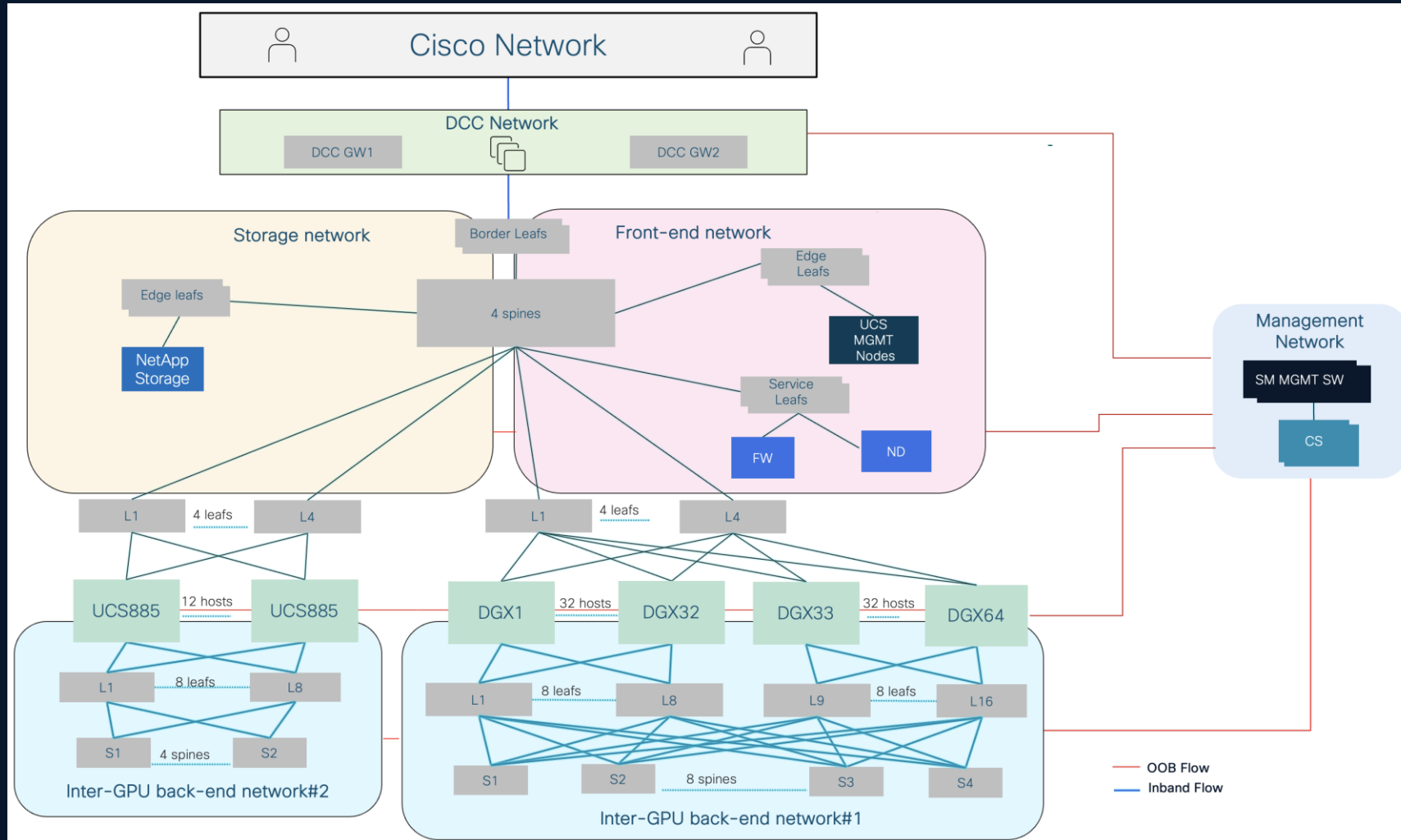
AI Fabric Design

Design #1



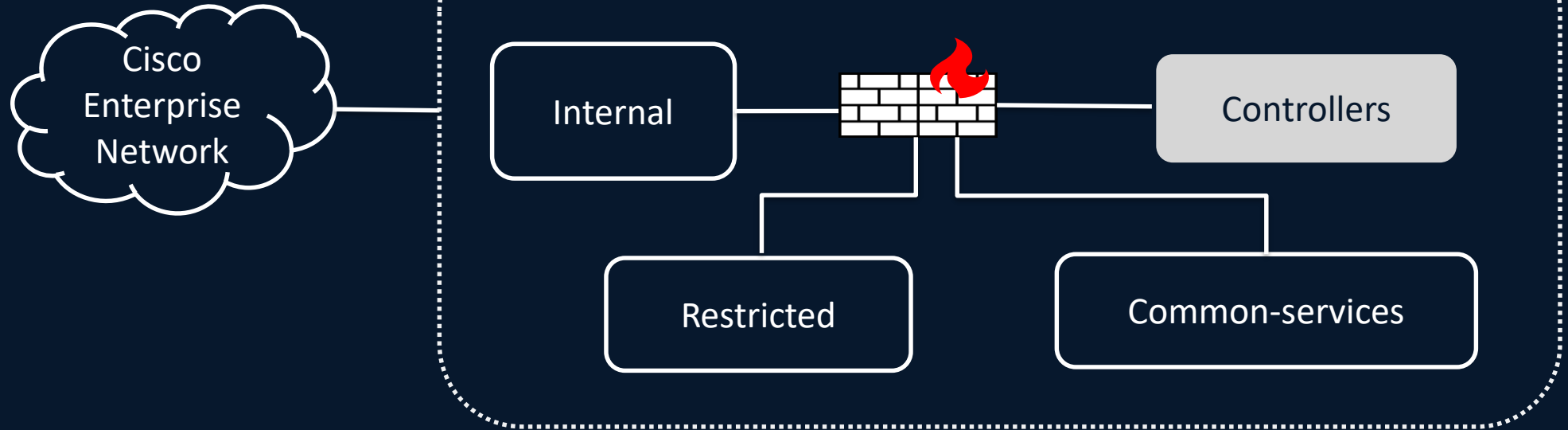
AI Fabric Design

Design #2 – Converged Front End



AI Fabric Design

Logical Topology



VRF Name	Purpose	Examples	FW Role
Controllers-vrf	AI fabric orchestration	NDI, NDFC	Enforces admin-only access and restricts management protocols and sources
Common-services-vrf	Shared infra services	Storage, Cluster-mgmt, Client access tools	Blocks unauthorized access; allows only approved service traffic
Restricted-vrf	Workload isolation based on Data Classification	Compute nodes used for Model training & inference jobs	Segregates restricted data; enforces granular north-south policies to prevent lateral movement

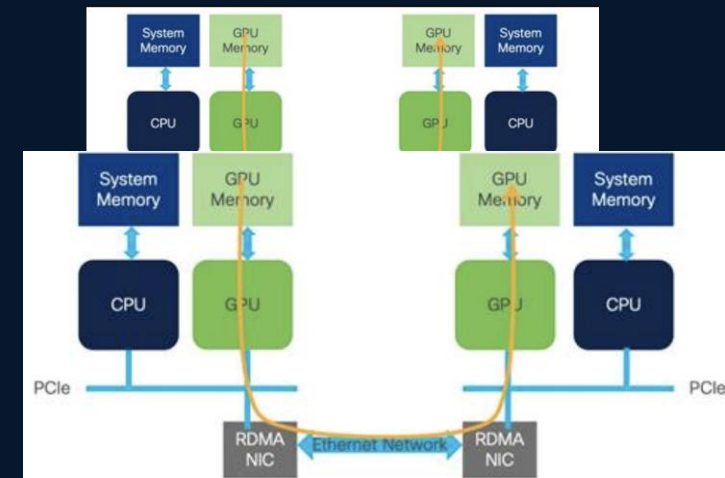
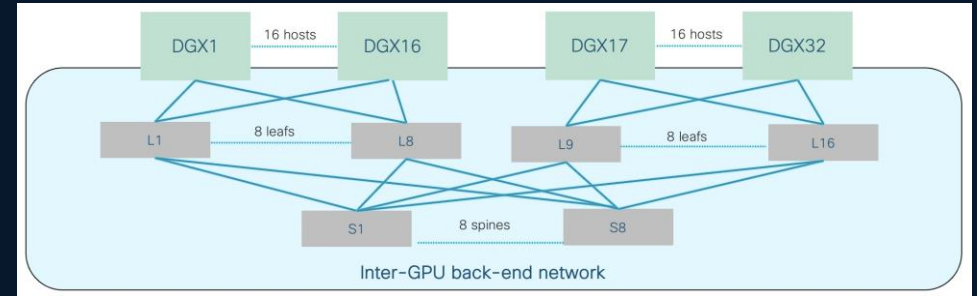
AI Fabric building blocks

Inter-GPU back-end network

Objective - High-speed data transfer between GPUs for distributed AI training workloads

Network Highlights

- ROCEv2 Support (GPU Direct RDMA)
- 400Gig non-blocking bandwidth per GPU
- Scalable Units with RAIL optimized design
- QoS with ECN and PFC
 - DSF



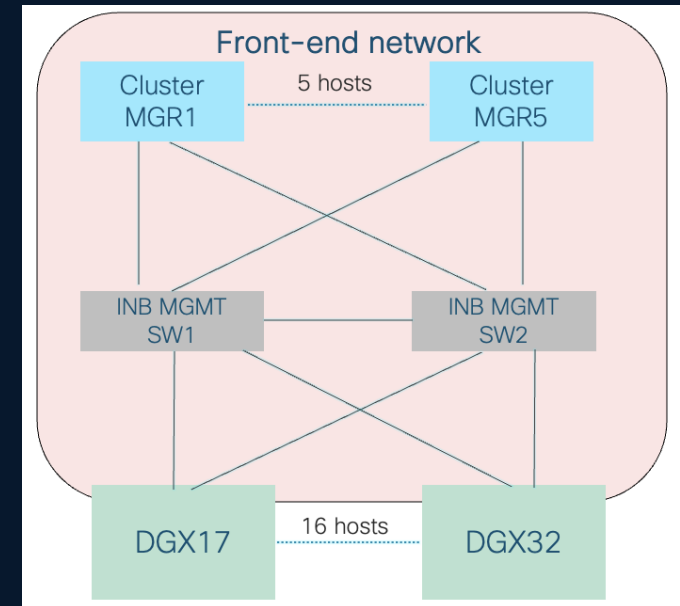
AI Fabric building blocks

Front-end Network

Objective - Enable clients to access the AI cluster, stage data, and schedule jobs on GPUs

Network Highlights

- Mirrors traditional datacenter network architecture
- Hosts UCS Cluster Manager for AI/ML job orchestration and fabric controllers
- High-speed 100G Ethernet connectivity
- Front-end switches connected to the broader Cisco enterprise network



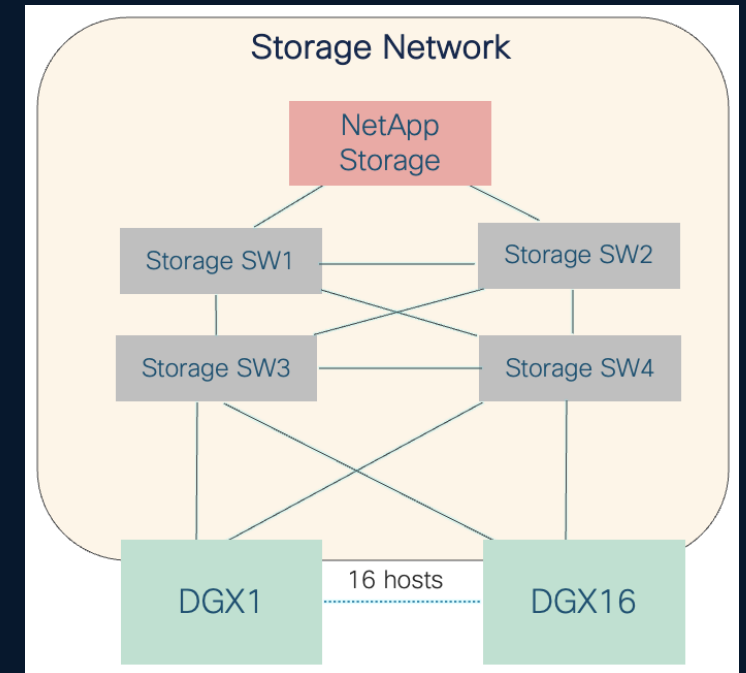
AI Fabric building blocks

Storage Network

Objective – Ensure high-throughput delivery of massive data volumes from large-scale storage infrastructure to GPUs with speed and reliability

Network Highlights

- Fabric for GPU-to-storage server communication
- High-speed 100G Ethernet connectivity
- Enables client and DGX access to periodic data snapshots



AI Fabric Data Center

Entire DC Layout

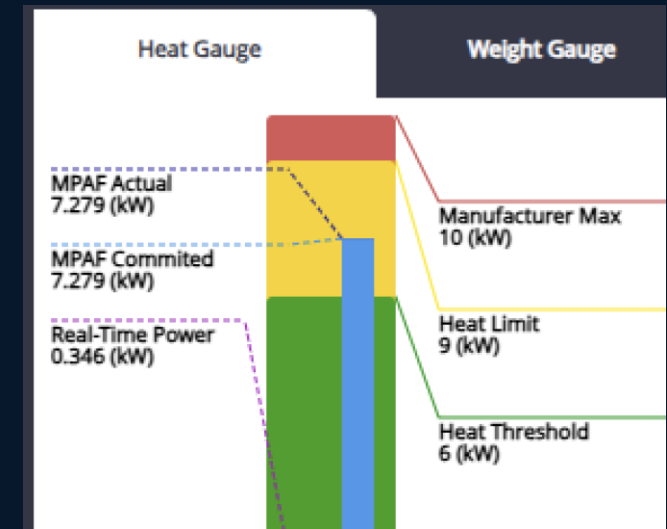
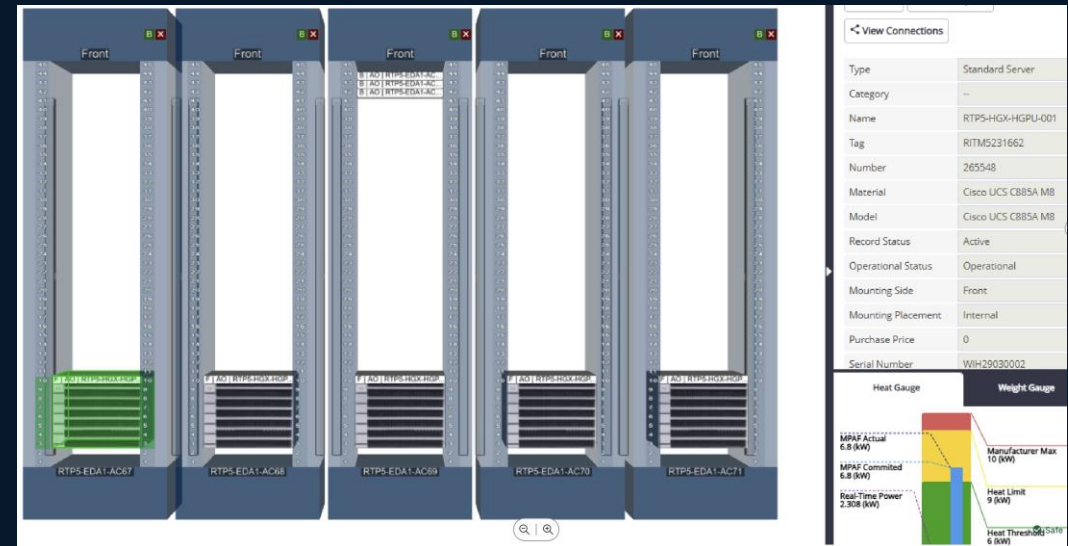
- Utilized existing cabinets, power, and cooling
- 32 DGX Units distributed for high power and cooling
- Direct connections with long patch cords to save time and cost



AI Fabric Data Center

Compute cabinets

- Distributed DGX hardware (~10 kW per cabinet)
 - Newer DCs optimized for 25kW (2 Nvidia) per cabinet
- Maintained N+1 power and cooling redundancy
- Reduced cost and deployment time



AI Fabric Data Center

Front-end network Cabinets

Network Row Optimization

- Centralized network and storage hardware
- Reduced cabling and patching complexity
- Ensured failover and power redundancy



AI Fabric Capacity planning

Inter-GPU back-end network

Total compute ports needed

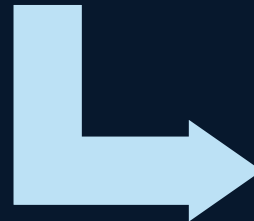
DGX nodes: 64
Ports per node: 8
Total downlink ports needed: 512
(64×8=512)



Total number of leafs needed

Downlinks per leaf: 32
Uplinks per leaf: 32
Number of leafs needed:

512 DGX ports
----- = 16 leafs
32 leaf-dgx ports



Total number of spines needed

Number of spines needed:

Total leaf-spine ports 512
----- = 8 Spines
64 port per spine switch

AI Fabric Capacity planning

All AI fabrics

	GPU Nodes	Inter-GPU back-end network	Front-end Network	Storage Network
Sonic AI Cluster	256	8 leafs, 4 spines	Layer2 Network 2 - Cisco 8102-64H	Layer2 Network 4 - Cisco 8102-32FH
Nexus AI Cluster (ALLN)	256	8 leafs, 4 spines	12 Leafs, 4 Spines	
Nexus AI Cluster (RTP)	512	16 leafs, 8 spines	12 Leafs, 4 Spines*	
Nexus AI Cluster (RTP)	96	8 leafs, 4 spines	12 Leafs, 4 Spines*	
Hyperfabric AI Cluster	256	8 leafs, 4 spines	TBD	TBD

*Converged VXLAN front end network

Sonic AI Fabric Build

Sonic AI Fabric Build

Inter-GPU back-end network

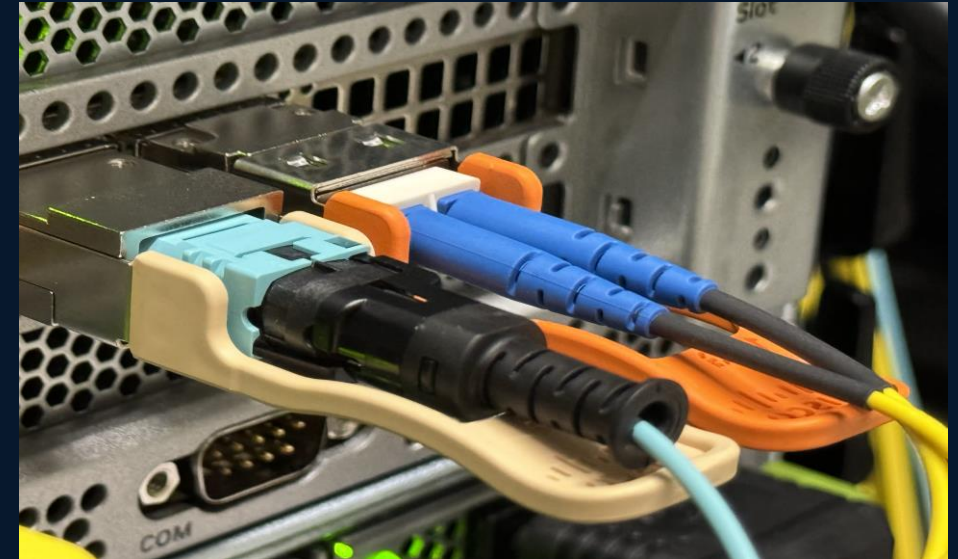
- High-speed data transfer between GPUs
- RAIL optimized design with L3 connections
- iBGP with Route Reflectors
- 2x400Gig LAG Leaf to Spine connectivity
- 400Gig non-blocking bandwidth per GPU
- Cisco 8101-32FH-O
 - Then 8122-64EH-O
- QoS with ECN and PFC
 - Then Distributed Scheduled Fabric



Sonic AI Fabric Build

Front-end network

- Dedicated client access to the network
- Layer 2 between the DGX and UCS Managers
- Inband MGMT switches connected to Cisco Network
- Cisco 8102-64H-O switch, 100Gig ports
- 5 Management UCS hosts
- Simple Linux Utility for Resource Management (SLURM)



Sonic AI Fabric Build

Storage network

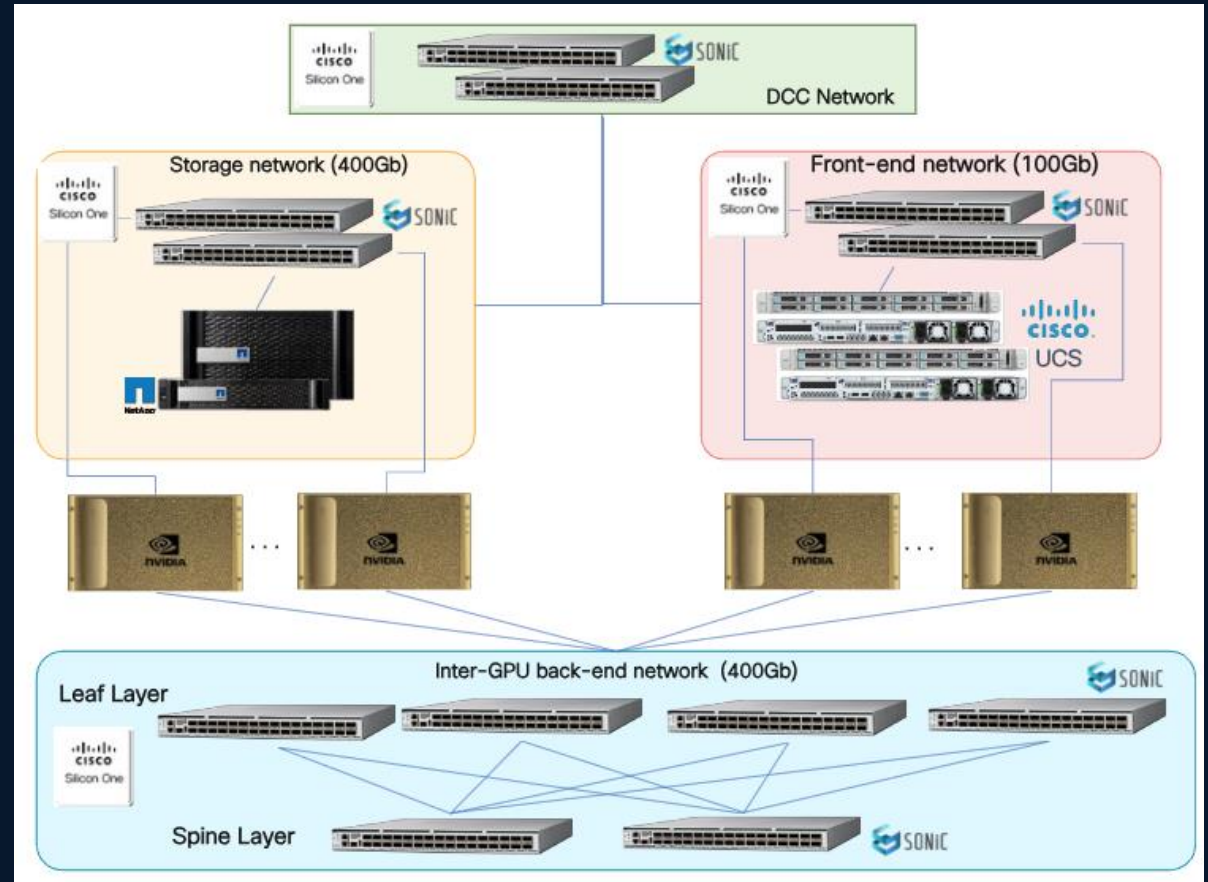
- Dedicated client access to data snapshots
- Netapp Storage, A900 series
- Layer 2 network, 400Gig Links between storage switches
- Two 100Gig links to NetApp and DGX
- Cisco 8101-32FH-O



Sonic AI Fabric Build

Overview

- First fabric built on Silicon One ASICs
- Utilizing SONiC as the NOS
- Ethernet-based architecture
- 32 Nvidia compute nodes (8 GPUs each)
- 12.8T throughput per back-end leaf(8101)
 - 51.2T throughput per back-end leaf (8122)
- 32 PetaFLOPS AI performance per node
- 4 NetApp AFF A900 nodes



Cisco Network Gear for Sonic AI Fabric



8101-32FH-O / 8122-64EHF-O



8102-64H-O



93108TC-FX3H

ASIC	Silicon One	Silicon One	Cloud Scale
Rack Units	1RU / 2RU	2RU	1RU
NOS	SONiC	SONiC	NX-OS
Ports	32xQDD / 64xOSFP800	64xQSFP28	48x10GBASE-T
Total Throughput	12.8 Tbps / 51.2 Tbps	6.4 Tbps	1.8 Tbps
Use Case	Inter-GPU back-end network Storage Network	Front-end network	Management network
Port Speed	400Gbps / 800Gbps	100 Gb/s	1/10Gb/s

Optics for Sonic AI Fabric



QDD-400G-DR4-S



QSFP-100G-SR4-S



QSFP-100G-SM-SR



MMS4X00-NS-FLT

Vendor	Cisco	Cisco	Cisco	Nvidia
Speed	400G	100G	100G	800G Duplex 400G-DR4
Fibre Mode	Single Mode	Multimode	Single Mode	Single Mode
Max Cable Distance	500m	70m	2000m	100m
Fibre Connector Type	MTP/MPO-12	MTP/MPO-12	Duplex LC	Duplex MPO-12
Use Case	Compute Fabric (Leafs/Spines)	Storage and Inband Mgmt	DCC	GPU Node

Cisco 8000 Network Configuration

- Native sonic cli

- Modify the config_db JSON File

```
admin@svlngen4-sonic-8k-1:~$ sudo config interface ip add Ethernet16 172.20.12.1/30

admin@svlngen4-sonic-8k-1:~$ sudo config interface ip add Ethernet24 192.168.20.1/24

admin@svlngen4-sonic-8k-1:~$ sudo config save -y
admin@svlngen4-sonic-8k-1:~$
```

```
admin@svlngen4-sonic-8k-1:~$ show ip interfaces
```

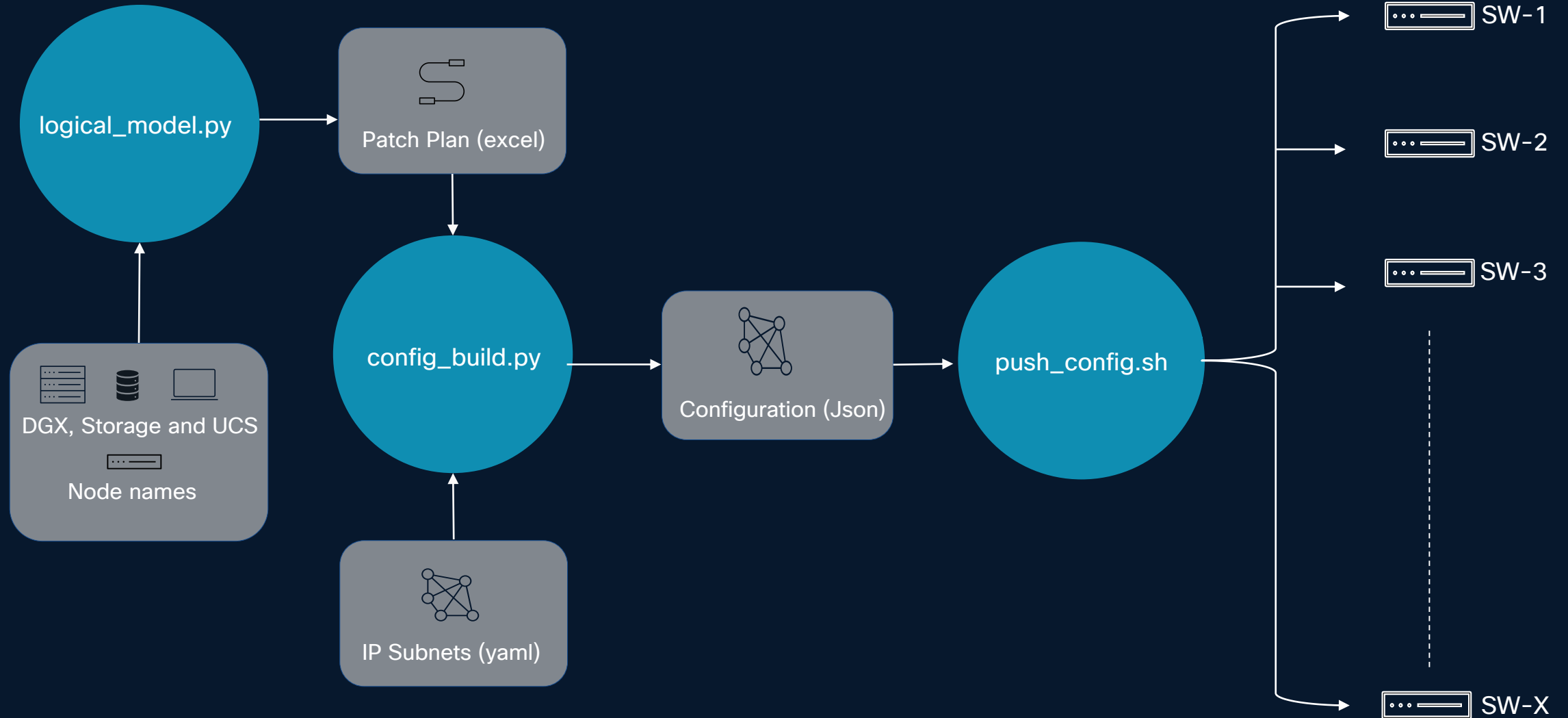
Interface	Master	IPv4 address/mask	Admin/Oper	BGP Neighbor	Neighbor IP
Ethernet16		172.20.12.1/30	up/up	8101-2	('default', '172.20.12.2')
Ethernet24		192.168.20.1/24	up/up	8101-2-vlan20	('default', '192.168.20.2')

```
admin@svlngen4-sonic-8k-1:/etc/sonic$ pwd
/etc/sonic
admin@svlngen4-sonic-8k-1:/etc/sonic$ ls | grep config_db.json
config_db.json

admin@svlngen4-sonic-8k-1:/etc/sonic$ grep -A 5 INTERFACE config_db.json
"INTERFACE": {
    "Ethernet16": {},
    "Ethernet16|172.20.12.1/30": {},
    "Ethernet24": {},
    "Ethernet24|192.168.20.1/24": {}
},

admin@svlngen4-sonic-8k-1:~$ sonic-cfggen -j /etc/sonic/config_db.json
admin@svlngen4-sonic-8k-1:~$
admin@svlngen4-sonic-8k-1:~$ sudo config save -y
Running command: /usr/local/bin/sonic-cfggen -d --print-data > /etc/sonic/config_db.json
admin@svlngen4-sonic-8k-1:~$
```

Cisco 8000 Automated Provisioning



QoS Configuration

```
"QUEUE": {  
  "Ethernet0|0": {  
    "scheduler": "scheduler.0"  
  },  
  "Ethernet0|1": {  
    "scheduler": "scheduler.0"  
  },  
  "Ethernet0|2": {  
    "scheduler": "scheduler.0"  
  },  
  "Ethernet0|3": {  
    "scheduler": "scheduler.1",  
    "wred_profile": "CISCO_LOSSLESS"  
  },  
  "Ethernet0|4": {  
    "scheduler": "scheduler.1",  
    "wred_profile": "CISCO_LOSSLESS"  
  },  
}
```

```
"WRED_PROFILE": {  
  "CISCO": {  
    "ecn": "ecn_green_yellow",  
    "green_drop_probability": "5",  
    "green_max_threshold": "4194304",  
    "green_min_threshold": "1048576",  
    "wred_green_enable": "true",  
    "wred_yellow_enable": "true",  
    "yellow_drop_probability": "0",  
    "yellow_max_threshold": "6144000",  
    "yellow_min_threshold": "0"  
  },  
  "CISCO_LOSSLESS": {  
    "ecn": "ecn_green_yellow",  
    "green_drop_probability": "5",  
    "green_max_threshold": "4000000",  
    "green_min_threshold": "3000000",  
    "wred_green_enable": "true",  
    "wred_yellow_enable": "true",  
    "yellow_drop_probability": "0",  
    "yellow_max_threshold": "6144000",  
    "yellow_min_threshold": "0"  
  },  
}
```

```
"TC_TO_PRIORITY_GROUP_MAP": {  
  "CISCO": {  
    "0": "0",  
    "1": "0",  
    "2": "0",  
    "3": "3",  
    "4": "4",  
    "5": "0",  
    "6": "0",  
    "7": "7"  
  },  
}
```

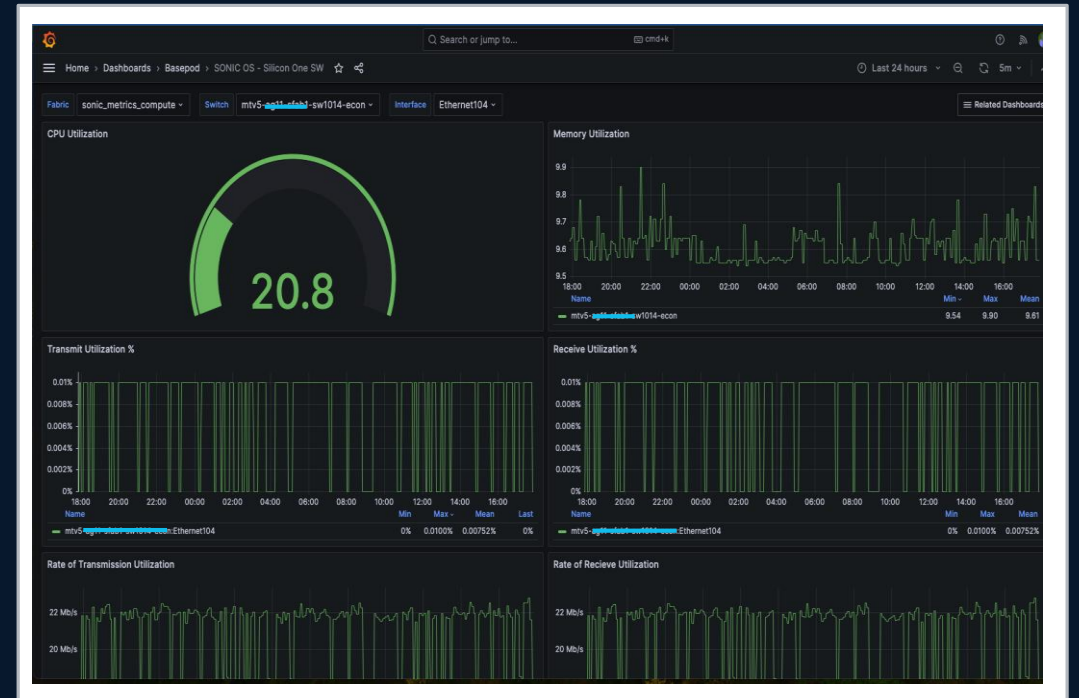
```
"TC_TO_QUEUE_MAP": {  
  "CISCO": {  
    "0": "0",  
    "1": "1",  
    "2": "2",  
    "3": "3",  
    "4": "4",  
    "5": "5",  
    "6": "6",  
    "7": "7"  
  },  
}
```

```
"PORT_QOS_MAP": {  
  "Ethernet0": {  
    "dscp_to_tc_map": "CISCO",  
    "pfc_enable": "3,4",  
    "pfc_to_queue_map": "CISCO",  
    "pfcwd_sw_enable": "3,4",  
    "tc_to_pg_map": "CISCO",  
    "tc_to_queue_map": "CISCO"  
  },  
}
```

```
"Ethernet100": {  
  "dscp_to_tc_map": "CISCO",  
  "pfc_enable": "3,4",  
  "pfc_to_queue_map": "CISCO",  
  "pfcwd_sw_enable": "3,4",  
  "tc_to_pg_map": "CISCO",  
  "tc_to_queue_map": "CISCO"  
},
```

```
"Ethernet104": {  
  "dscp_to_tc_map": "CISCO",  
  "pfc_enable": "3,4",  
  "pfc_to_queue_map": "CISCO",  
  "pfcwd_sw_enable": "3,4",  
  "tc_to_pg_map": "CISCO",  
  "tc_to_queue_map": "CISCO"  
},
```

SONiC Day 2 Ops – Grafana Dashboards

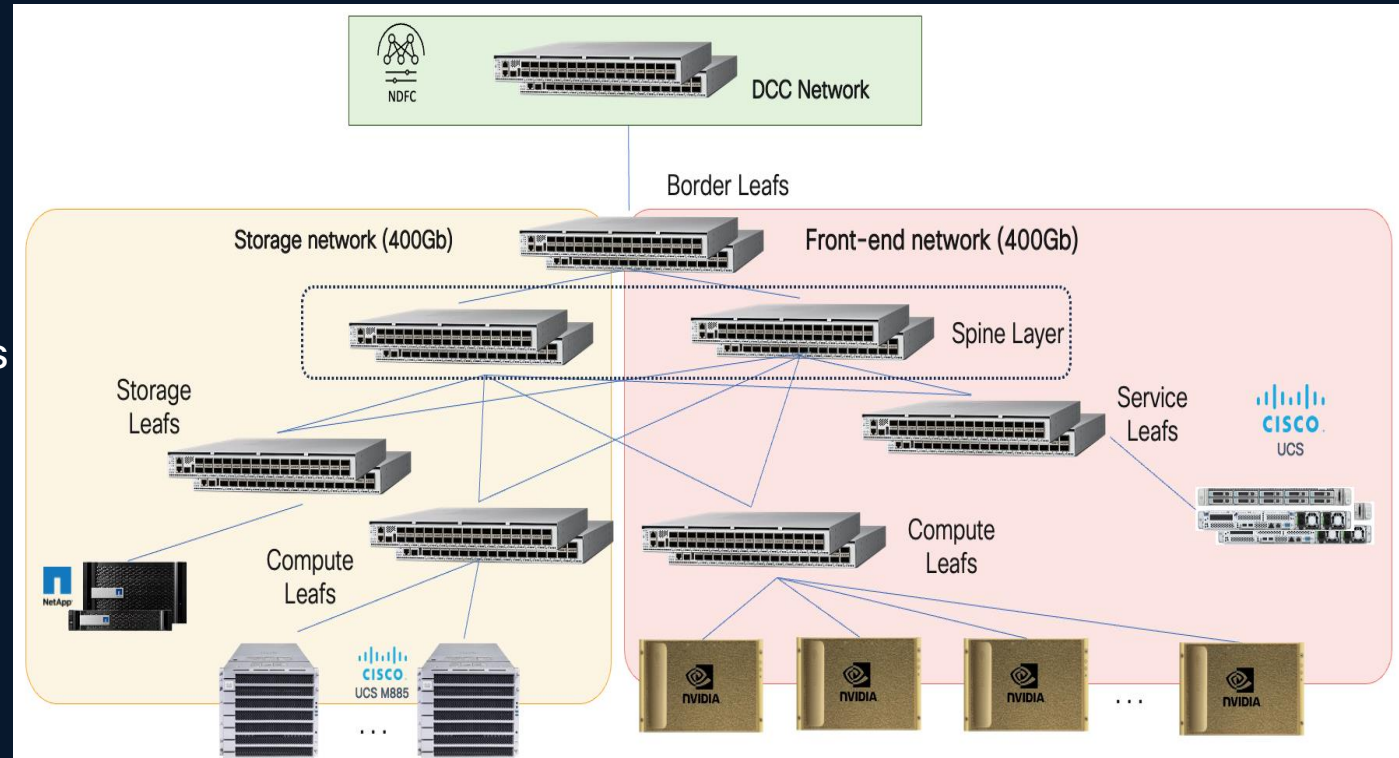


Nexus AI Fabric Build

Collapsed Front-end and Storage Network

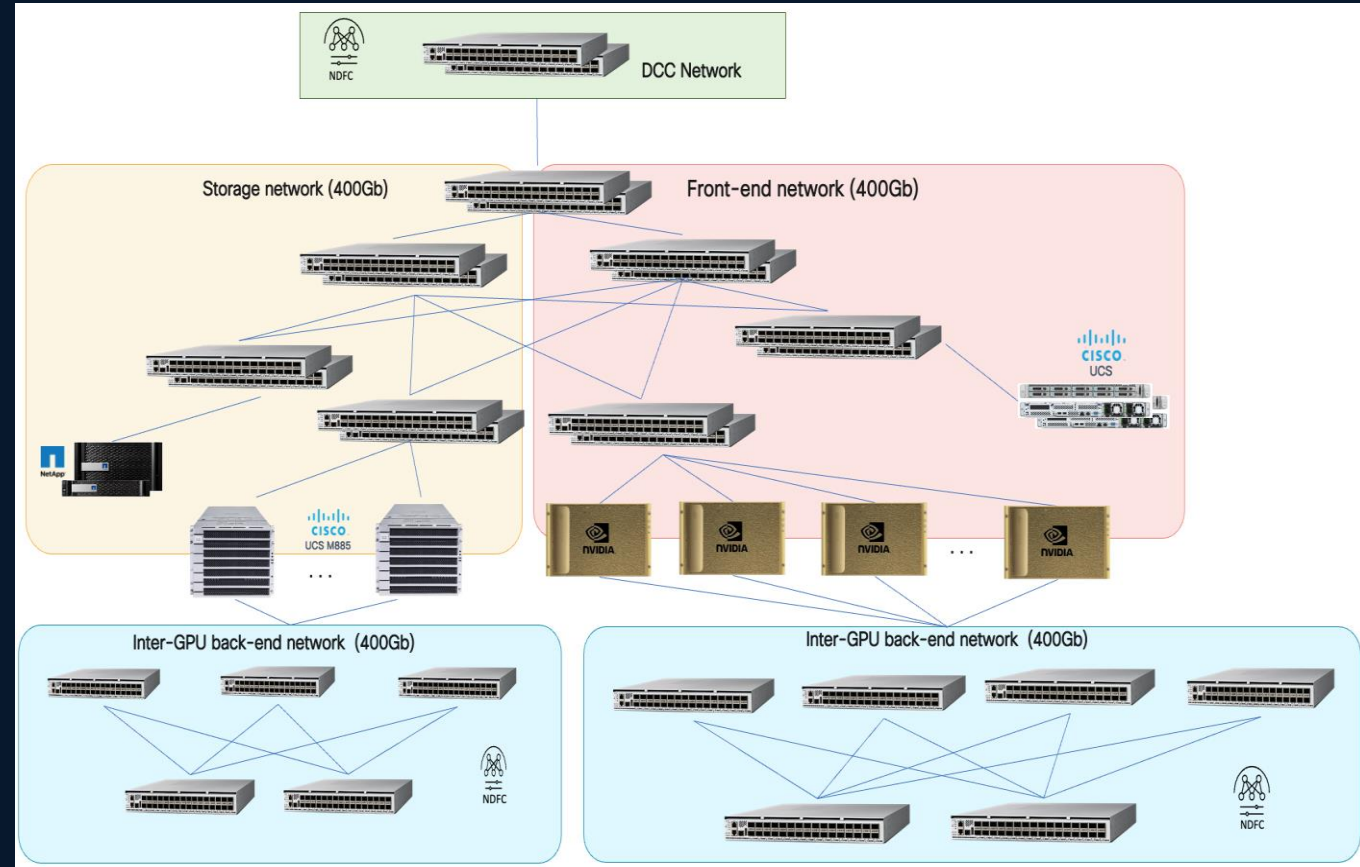
New Design on Nexus

- Client and Data Access
- 8 Management UCS Hosts
- NetApp Storage, A900 series
- VXLAN over BGP, 400Gig Leaves and Spines
- 100Gig links to NetApp, DGX and UCS
- Nexus N9K-C9364D-GX2A
- Nexus N9K-C93600CD-GX

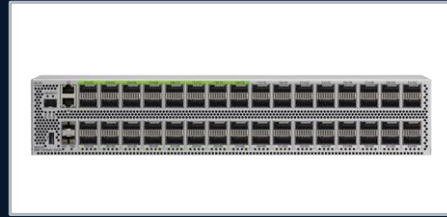


Nexus AI Fabric Build Overview

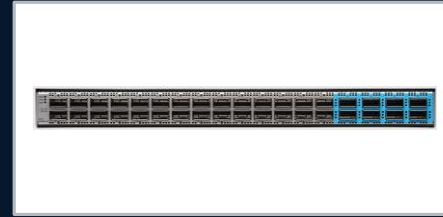
- Utilizing NX-OS as the NOS
- Our first VXLAN DC Fabrics
- NDFC and NDI for Fabric Build & Management
- Ethernet-based architecture
- 400G non-blocking compute fabric
- 64 Nvidia compute nodes (8 GPUs each)
- 51.2T throughput per Back-End leaf
- 32 PetaFLOPS AI performance per node
- 4 NetApp AFF A900 nodes



Cisco Network Gear for Nexus AI Fabric



N9K-C9364D-GX2A



N9K-C93600CD-GX



N9K-C93108TC-FX3H

ASIC	Cloud Scale	Cloud Scale	Cloud Scale
Rack Units	2RU	1RU	1RU
NOS	NX-OS	NX-OS	NX-OS
Ports	64xQSFP-DD	28xQSFP28, 8xQSFP-DD	48x10GBASE-T
Total Throughput	25.6 Tbps	12 Tbps	1.8 Tbps
Use Case	Inter-GPU back-end network Front-end network (Spines)	Front-End network Leafs Management network GWs	Management network
Port Speed	400Gb/s	100 Gb/s	1/10Gb/s

Optics for Nexus AI Fabric



QDD-400G-FR4-S



QSFP-100G-DR-S



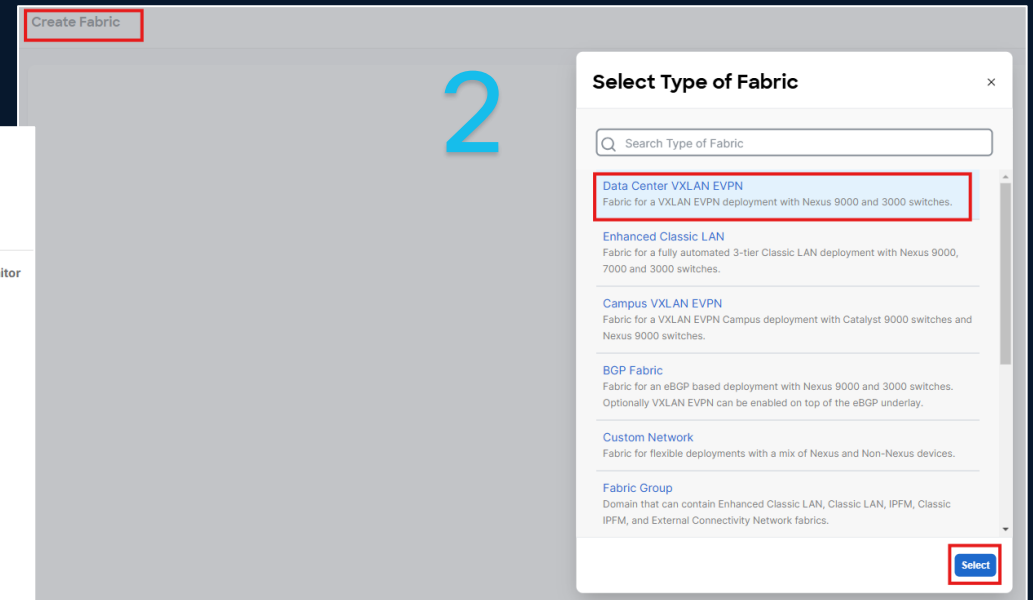
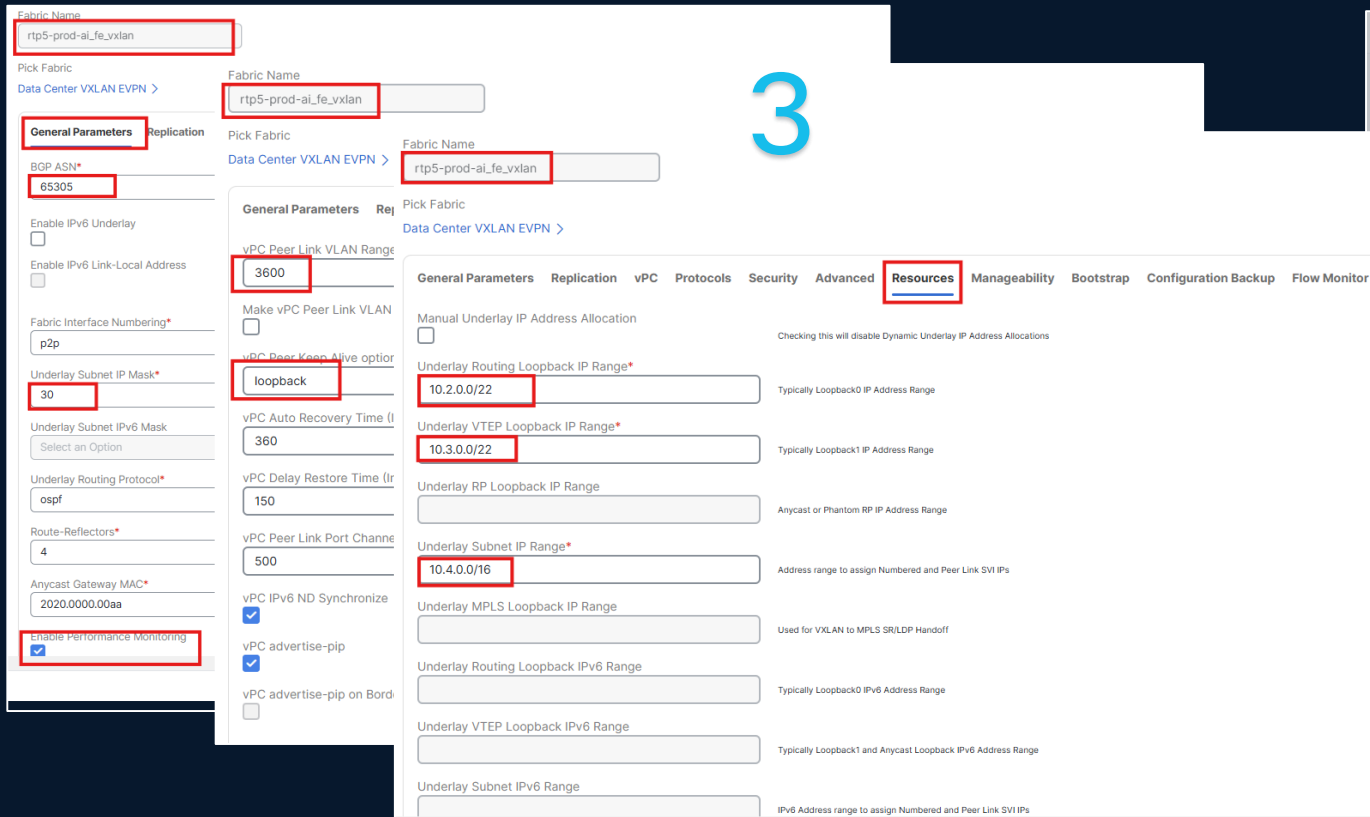
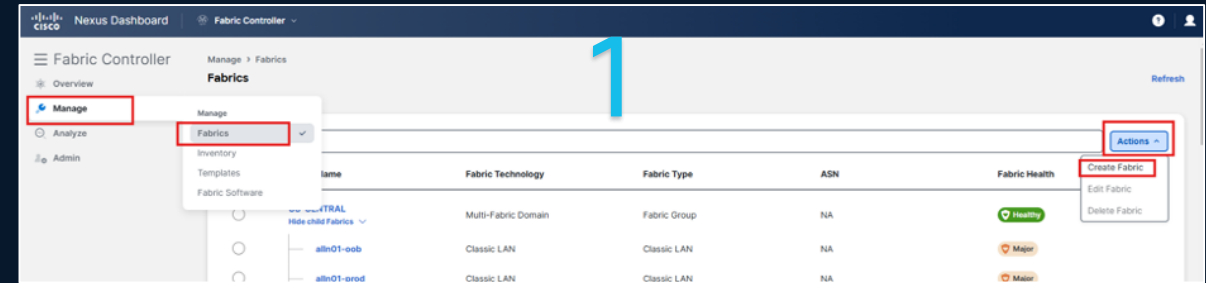
QSFP-100G-FR-S

Vendor	Cisco	Cisco	Cisco
Speed	400G	100G	100G
Fibre Mode	Single Mode	Single Mode	Single Mode
Max Cable Distance	2000m	500m	2000m
Fibre Connector Type	Duplex LC	Duplex LC	Duplex LC
Use Case	Leaf to Spine links	UCS Mgmt Nodes	Nvidia Storage and In-Band Links

Nexus 9300 – NDFC Fabric Setup

Fabric Setup

1. Run Fabric Creation Wizard
2. Select Fabric Template
3. Define Fabric Parameters



Nexus 9300 Setup

- POAP and Configure Network
 1. PreProvision Switches
 2. Discovery phase
 3. Boot up with config
 4. Provision overlay (VXLAN EVPN)

Host Name	IP Address	Role	serialNumber	Model	Software
rtp5- fefab1-sw1042	10.115.	leaf	FDO2833034E	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw1041	10.115.	leaf	FDO2833037Q	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw1032	10.115.	leaf	FDO2833036J	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw1072	10.115.	leaf	FDO2833031Z	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw3822	10.115.	border	FDO2833033E	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw3904	10.115.	border spine	FDO283507P1	N9K-C9364D-GX2A	10.3(6)
rtp5- fefab1-sw1031	10.115.	leaf	FDO28330VYS	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw1071	10.115.	leaf	FDO28300FZF	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw3821	10.115.	border	FDO28330313	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw3903	10.115.	border spine	FDO283507NX	N9K-C9364D-GX2A	10.3(6)
rtp5- fefab1-sw1022	10.115.	leaf	FDO28330335	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw1062	10.115.	leaf	FDO2833031M	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw3812	10.115.	border	FDO28330338	N9K-C93600CD-GX	10.3(6)
rtp5- fefab1-sw3902	10.115.	border spine	FDO283400UY	N9K-C9364D-GX2A	10.3(6)

Deploy Configuration - rtp5-prod-af, fe, ixlan

Config Preview

Switch Name	IP Address	Role	Serial Number	Fabric Status	Pending Config	Status Description	Progress	Replay Details
rtp5- fefab1-sw3812	10.115.	border	FDO28330338	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw3904	10.115.	border spine	FDO283507P1	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1031	10.115.	border spine	FDO283507VY	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1001	10.115.	leaf	FDO2833037L	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw3821	10.115.	leaf	FDO28330313	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1071	10.115.	leaf	FDO28300FZF	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1032	10.115.	leaf	FDO2833036J	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1042	10.115.	leaf	FDO2833034E	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1072	10.115.	leaf	FDO2833031Z	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw1062	10.115.	leaf	FDO2833031M	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay
rtp5- fefab1-sw3812	10.115.	leaf	FDO28330338	In Sync	0 Lines	In Sync	<div style="width: 100%;"></div>	Replay

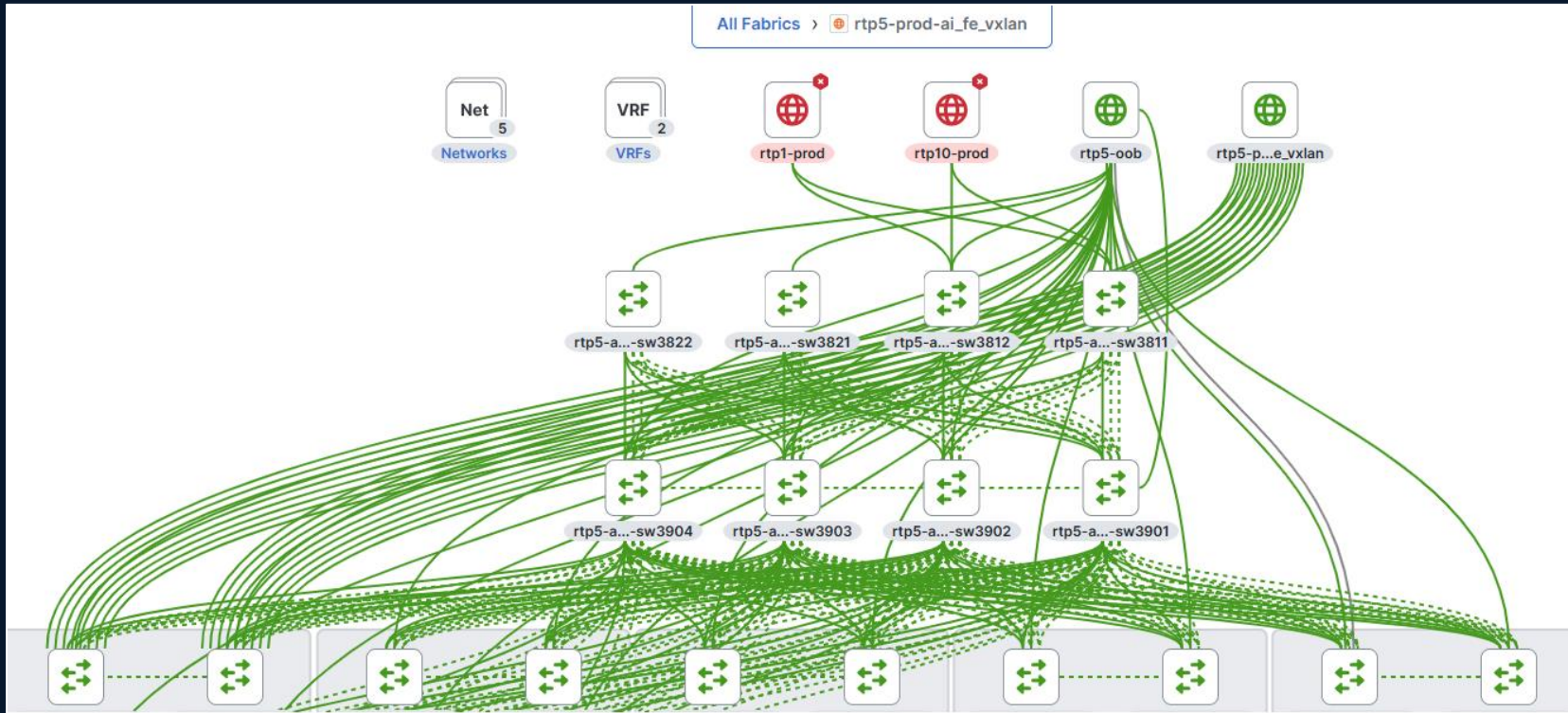
Overview **Switches** Links Interfaces Interface Groups Policies Networks VRFs Services Security Event Analysis History Resources Metrics

Filter by attributes

Switch	IP Address	Role	Serial Number	Mode	Config Status	Oper Status	Discovery Status	Model
<input type="checkbox"/>	rtp5- fefab1-sw1042	Leaf	FDO2833034E	Normal	In Sync	Minor	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw1041	Leaf	FDO2833037Q	Normal	In Sync	Healthy	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw1032	Leaf	FDO2833036J	Normal	In Sync	Minor	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw1072	Leaf	FDO2833031Z	Normal	In Sync	Minor	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw3822	Border	FDO2833033E	Normal	In Sync	Major	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw3904	Border Spine	FDO283507P1	Normal	In Sync	Minor	OK	N9K-C9364D-GX2A
<input type="checkbox"/>	rtp5- fefab1-sw1031	Leaf	FDO28330VYS	Normal	In Sync	Healthy	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw1071	Leaf	FDO28300FZF	Normal	In Sync	Minor	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw3821	Border	FDO28330313	Normal	In Sync	Major	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw3903	Border Spine	FDO283507NX	Normal	In Sync	Healthy	OK	N9K-C9364D-GX2A
<input type="checkbox"/>	rtp5- fefab1-sw1022	Leaf	FDO28330335	Normal	In Sync	Healthy	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw1062	Leaf	FDO2833031M	Normal	In Sync	Major	OK	N9K-C93600CD-GX
<input type="checkbox"/>	rtp5- fefab1-sw3812	Border	FDO28330338	Normal	In Sync	Healthy	OK	N9K-C93600CD-GX

Interface and QoS setup

- Define VPC Pairs
- Define Interface configurations (Interface Wizard)
- Define QoS Policy (standard templates)
- Deploy policy



Nexus 9300 – NDFC Fabric QoS Setup

Edit Template : AI_Fabric_QOS_400G_950_WRED

Template Properties

2 Template Content

```
AI_Fabric_QOS_400G_950_WRED
1  ##template variables
2
3  #   Copyright (c) 2023 by Cisco Systems, Inc.
4  #   All rights reserved.
5
6  ##
7  ##template content
8
9  class-map type qos match-any ROCEv2
10 match dscp 26
11 class-map type qos match-any CNP
12 match dscp 48
13
14 policy-map type qos QOS_CLASSIFICATION
15 class ROCEv2
16 | set qos-group 3
17 class CNP
18 | set qos-group 7
19 class class-default
20 | set qos-group 0
21
22 policy-map type queuing QOS_EGRESS_PORT
23 class type queuing c-out-8q-q6
24 | bandwidth remaining percent 0
25 class type queuing c-out-8q-q5
26 | bandwidth remaining percent 0
27 class type queuing c-out-8q-q4
28 | bandwidth remaining percent 0
29 class type queuing c-out-8q-q3
30 | bandwidth remaining percent 50
```

Edit Template : AI_Fabric_QOS_400G_950_WRED

Help Validate No Errors

Template Properties

2 Template Content

```
AI_Fabric_QOS_400G_950_WRED
22 policy-map type queuing QOS_EGRESS_PORT
23 class type queuing c-out-8q-q6
24 | bandwidth remaining percent 0
25 class type queuing c-out-8q-q5
26 | bandwidth remaining percent 0
27 class type queuing c-out-8q-q4
28 | bandwidth remaining percent 0
29 class type queuing c-out-8q-q3
30 | bandwidth remaining percent 50
31 | random-detect minimum-threshold 950 kbytes maximum-threshold 3000 kbytes drop-probability 7 weight 0 ecn
32 class type queuing c-out-8q-q2
33 | bandwidth remaining percent 0
34 class type queuing c-out-8q-q1
35 | bandwidth remaining percent 0
36 class type queuing c-out-8q-q-default
37 | bandwidth remaining percent 50
38 class type queuing c-out-8q-q7
39 | priority level 1
40
41 policy-map type network-qos qos_network
42 class type network-qos c-8q-nq3
43 | pause pfc-cos 3
44 | mtu 4200
45 class type network-qos c-8q-nq-default
46 | mtu 9216
47
48 priority-flow-control watch-dog-interval on
49
50 system qos
51 | service-policy type network-qos qos_network
52 | service-policy type queuing output QOS_EGRESS_PORT
```

Nexus Day 2 Ops – NDFC Inventory and Events

Fabric Controller - Manage > Fabric Software - **Fabric Software**

Overview Images Image Policies Devices History

Images 3 NDK Images

Policies 5 NDK

Fabric	Current Version(s)	Policy	Status	Fabric Type	Switches
aer01-ooB	9.3(10)	None	None Prepare	Classic LAN	4
aer01-prod	7.3(12)N1(1) and 2 more	None	None Prepare	Classic LAN	16
alt01-ooB	7.3(12)N1(1)	None	None Prepare	Classic LAN	12

43 items found

Fabric Controller - Manage > Fabric Software - **Fabric Software**

Overview Images Image Policies Devices History

Images

Platform	Bits	Image Name	Image Type	Image Sub Type	NOS Version	Image Version	Size (Bytes)	Reference Count	Image Present	Checksum
NK/NK	64bit	nxos64-cs.10.3.5.M.bin	image	nxos	10.3.5	10.3.5	2,411,492,352	0	Present	e262c5409bc95c78e10d0f0ba2998b
NK/NK	64bit	nxos64-cs.10.3.6.M.bin	image	nxos	10.3.6	10.3.6	2,437,178,880	3	Present	0a4eb079aaae0e44648099017e21c9
NK/NK	64bit	nxos64-cs.10.5.2.F.bin	image	nxos	10.5.2	10.5.2	2,371,266,048	2	Present	6675c7ab87582918815ae9595a043a

Fabric Controller - Manage > Inventory - **Inventory**

Switches Interfaces

Switch	IP Address	Role	Serial Number	Fabric Name	Mode	Config Status	Oper Status	Discovery Status	Model
CL-AI-BE-FABRIC-LEAF01	10.100.1.1	leaf	FGD2804000V	CL-AI-Fabric-VXLAN-EVPN	Leaf	ok	Up	Ready	N9K-C9304T-03TX
CL-AI-BE-FABRIC-LEAF02	10.100.1.2	leaf	FGD2804000V	CL-AI-Fabric-VXLAN-EVPN	Leaf	ok	Up	Ready	N9K-C9304T-03TX
CL-AI-BE-FABRIC-LEAF03	10.100.1.3	leaf	FGD2804000V	CL-AI-Fabric-VXLAN-EVPN	Leaf	ok	Up	Ready	N9K-C9304T-03TX
CL-AI-BE-FABRIC-LEAF04	10.100.1.4	leaf	FGD2804000V	CL-AI-Fabric-VXLAN-EVPN	Leaf	ok	Up	Ready	N9K-C9304T-03TX
CL-AI-BE-FABRIC-SPINE01	10.100.2.1	spine	FGD2804000V	CL-AI-Fabric-VXLAN-EVPN	Spine	ok	Up	Ready	N9K-C9304D-03TX
CL-AI-BE-FABRIC-SPINE02	10.100.2.2	spine	FGD2804000V	CL-AI-Fabric-VXLAN-EVPN	Spine	ok	Up	Ready	N9K-C9304D-03TX

Fabric Controller - Manage > Interfaces - **Interfaces**

Interface Groups

Fabric Name	Device Name	Interface	Admin Status	Oper Status	Reason	Policies	Overlay Network	Sync
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	mgmt0	Up	Up	ok	int_mgmt	NA	In
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	Vlan1	Down	Down	Administratively down	NA	NA	Out
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	Vlan2000	Up	Up	ok	NA	MyVRF_50000	Out
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	Loopback0	Up	Up	ok	int_fabric_loopback_1_1	NA	In
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	Loopback1	Up	Up	ok	int_fabric_loopback_1_1	NA	In
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	Ethernet1/1	Up	Up	ok	int_trunk_host	NA	In
CL-AI-Fabric-VXLAN-EVPN	CL-AI-BE-FABRIC-LEAF01	Ethernet1/2	Up	Up	ok	int_trunk_host	NA	In

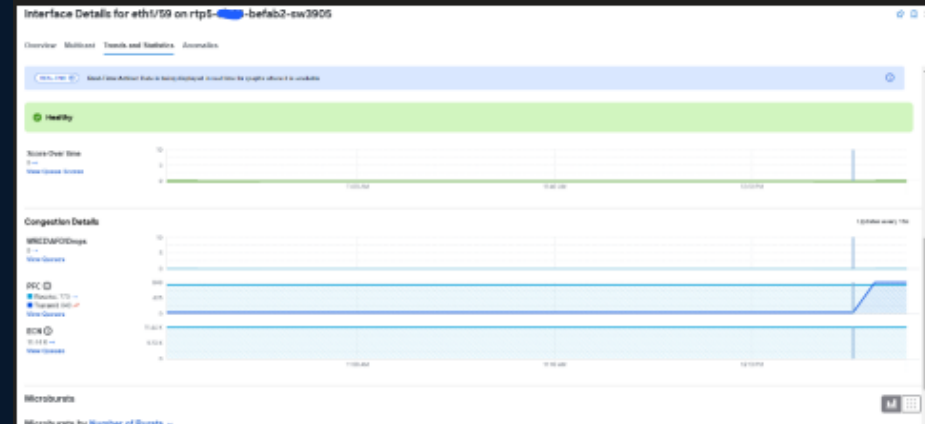
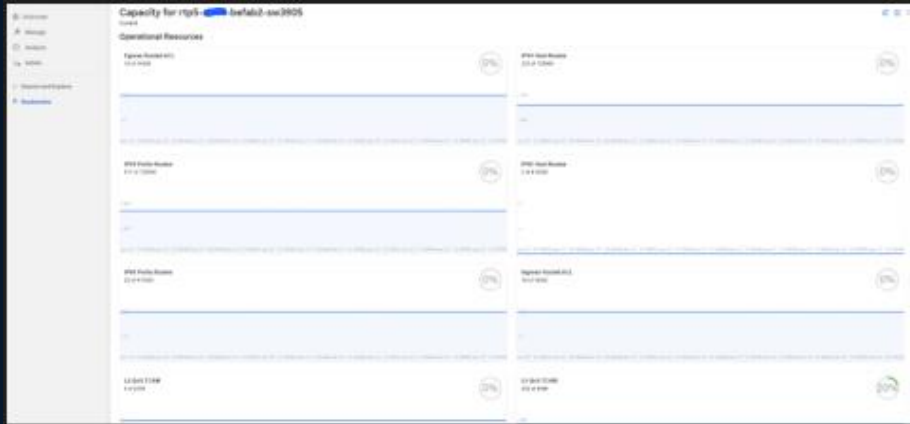
Fabric Overview - CL-AI-Fabric-VXLAN-EVPN

Overview Switches Links Interfaces Interface Groups Policies Networks VRF's Services Security **Event Analytics** History Resources Metrics

Alerts Cleared Alerts Events Recent Tasks

Severity	Source	Name	Category	Creation Time	Policy	Message
Minor	10.100.1.1	CL-AI-BE-FABRIC-LEAF01	INTERFACE STATUS	Jun 21 2025, 10:14:10	discovery	Interface ethernet1/20 is unconfigured state-up-down Link not connected
Minor	10.100.1.2	CL-AI-BE-FABRIC-LEAF02	INTERFACE STATUS	Jun 21 2025, 10:14:04	discovery	Interface ethernet1/20 is unconfigured state-up-down Link not connected
Warning	CL-AI-FABRIC-VXLAN-EVPN	FABRIC-VXLAN-EVPN-CL-AI-Fabric-VXLAN-EVPN	EXTERNAL	Jun 21 2025, 10:54:31	int_fabric errors	VRF 100 Subnet 177.174.174.0/24 duplicate with fabric-subnet-100-177.174.174.0

Nexus Day 2 Ops – NDI Performance Monitor



Monitoring Dashboard for rtp5-befab2-sw3905

Overview: Connectivity Anomalies Anomalies

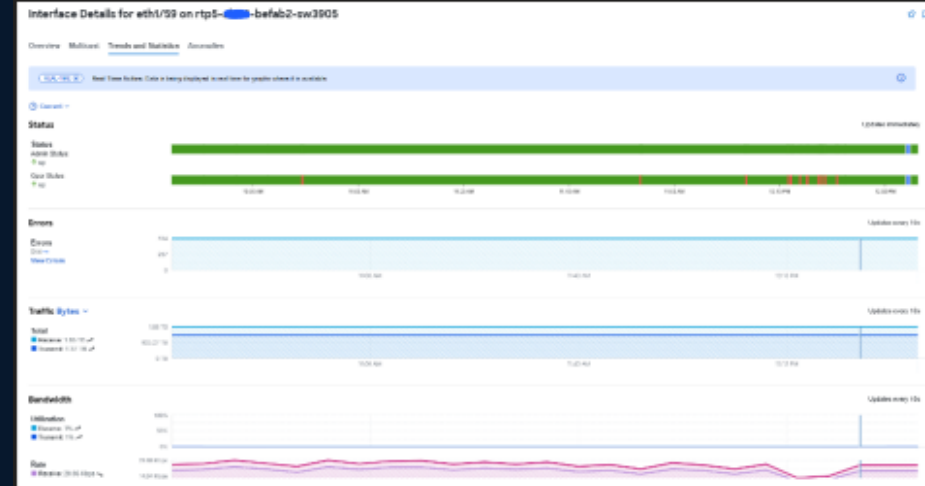
Interfaces: L3 Neighbors Endpoints L3 Overview Multicast Status

File

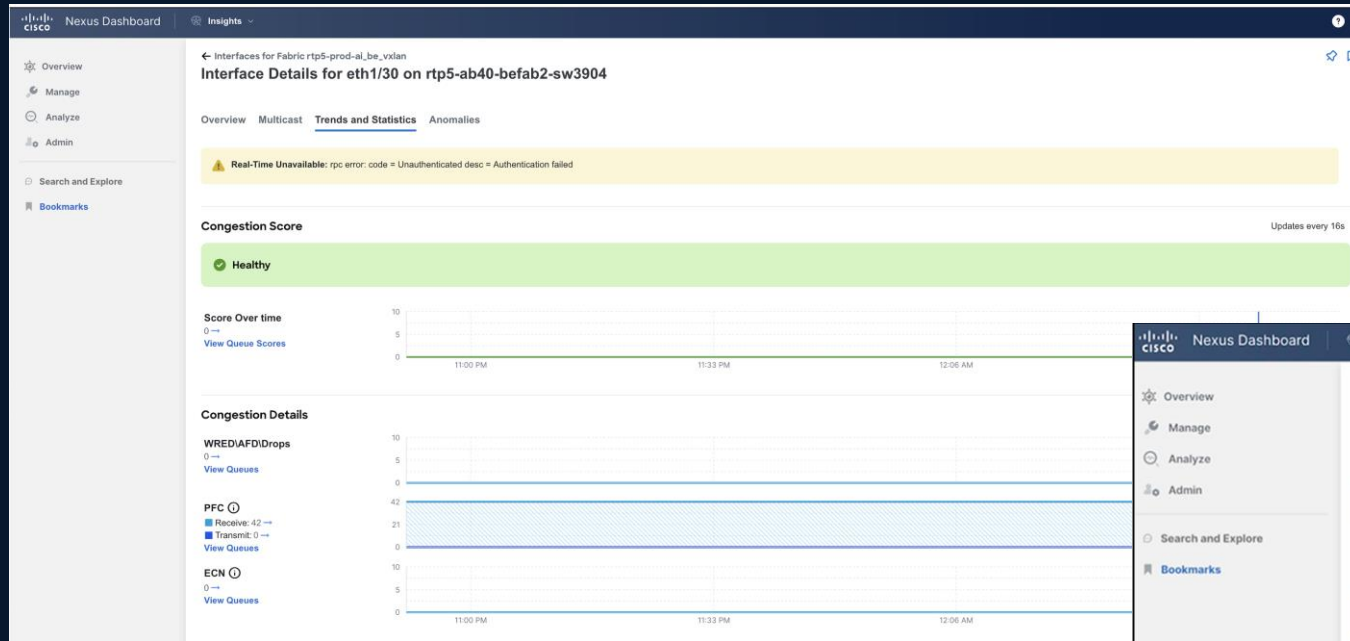
Security Level: 66

Admin Status: Operational Status: Type

Interface	Operational Level	Operational Status	Type	Admin Status	Operational Status	Source Utilization	Source Bytes	Source Packets	Source Rate	Received Bytes	Received Packets	Received Rate	Received Utilization	Errors
eth1/59	Operational	Operational	Physical	Operational	Operational	1.82%	23841012	2012 Pkts	1.26 T/s	24902887	26.78 Pkts	0	0%	0
eth1/58	Operational	Operational	Physical	Operational	Operational	0.11%	9196470	11.88 Pkts	1.01 T/s	1080540	11.87 Pkts	0	0%	0
eth1/57	Operational	Operational	Physical	Operational	Operational	1.24%	22819106	23.05 Pkts	402.01 G/s	16100049	26.49 Pkts	0	0%	0
eth1/56	Operational	Operational	Physical	Operational	Operational	1.91%	26704008	23.78 Pkts	1.14 T/s	11080871	11.71 Pkts	0	0%	0
eth1/55	Operational	Operational	Physical	Operational	Operational	1.22%	22681021	21.27 Pkts	1.10 T/s	23027042	21.04 Pkts	0	0%	0
eth1/54	Operational	Operational	Physical	Operational	Operational	1.17%	24102084	24.18 Pkts	1.34 T/s	11028000	10.53 Pkts	0	0%	0



Nexus Day 2 Ops – NDI Performance Monitor



Use Cases

A Sample of Current Usage

Team	Use Case	Business Outcomes
AI Defense	Custom model training to improve AI Defense capabilities	Protects against the safety and security risks introduced by the development and deployment of AI applications
Sales	16 (Sales)+15 (marketing) use cases, e.g.Sales AI agent	One stop for all sales queries, client status before sales client visits etc
Webex Audio & Video	Improving codec development for noise cancellation and lower bandwidth data prediction	Improved Webex features – gesture recognition, background, face landmarks, etc
HR	LLM training – create a bot or assistant that utilizes natural language processing and task-oriented frameworks to provide personalized and specific information to employees.	Elevating overall employee experience for HR services by building an ask P & C agent
CX Engineering	Parse/collate data and provide sales teams with high probability leads targeting existing customers	Increase renewals rate; reduce renewal attrition

Lessons Learned

AI presents challenges for IT teams



INFRASTRUCTURE DEMANDS

Rapid growth in data volume and variety

Unfamiliar application stacks and new, complex infrastructure patterns

Insufficient IT automation and observability

Greater cybersecurity threats

New operational silos

Shortage of technical expertise

Disorienting AI hype

High entry cost and lock-in issues

Cisco IT AI Fabric Key Takeaways

Principles	Requirements	Solutions
Iterative Approach	Achieve simplicity	Use a needs-based approach Crawl Walk Run
Fabric Design	Non-Blocking and Scalable 2 Tier Fabric	CLOS Fabric Design Nexus for Enterprise SONiC for Hyperscalers
Lossless Fabric	Fabric with lower latency and effective congestion management	Priority Flow Control Explicit Congestion Notification Ultra Ethernet
Load Balancing	Need efficient load balancing despite low entropy	Distributed Scheduled Fabric
Visibility	Operator needs visibility to app perf, network utilization and adverse events	Need well integrated application and network monitoring tools; telemetry

Future Plans

Hyperfabric AI

Simplified cloud managed AI Infrastructure

More AI clusters

Investing in more AI clusters leveraging Cisco UCS 885

SiliconOne Nexus

Deploy N9K-C9364E-SG2 based fabric

AI Defense

Protect AI models and runtime

Need for Speed

Deliver 800G AI Fabric

Complete Your Session Evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to win 1 of 5 full conference passes to Cisco Live 2026.



Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.



Level up and earn exclusive prizes!



Complete your surveys in the Cisco Live mobile app.

Continue your education



Visit the Cisco Showcase for related demos



Book your one-on-one Meet the Engineer meeting



Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs



Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact us at: chstokes@cisco.com, lvellank@cisco.com

Thank you

CISCO Live !

