Connecting Data Centers and Clouds in the AI Era

Kevin Baranowski
Sales Director, Internet & Mass-Scale Infrastructure

Leigh Wade
Solutions Engineer, Internet & Mass-Scale Infrastructure

Aaron Trompeter
Solutions Engineer, Automation & Assurance

11th December 2025

Agenda

1. Market needs – 5 - 6slides
2. Cisco strategy – 5 -5 slides
3. Cisco innovation – 40 – 11slides
4. Use Cases: DCI & WAN – 10 - 10slides
5. Building blocks – 10 - 3slides
6. Q&A and next steps - 5

AI is re-shaping network connectivity

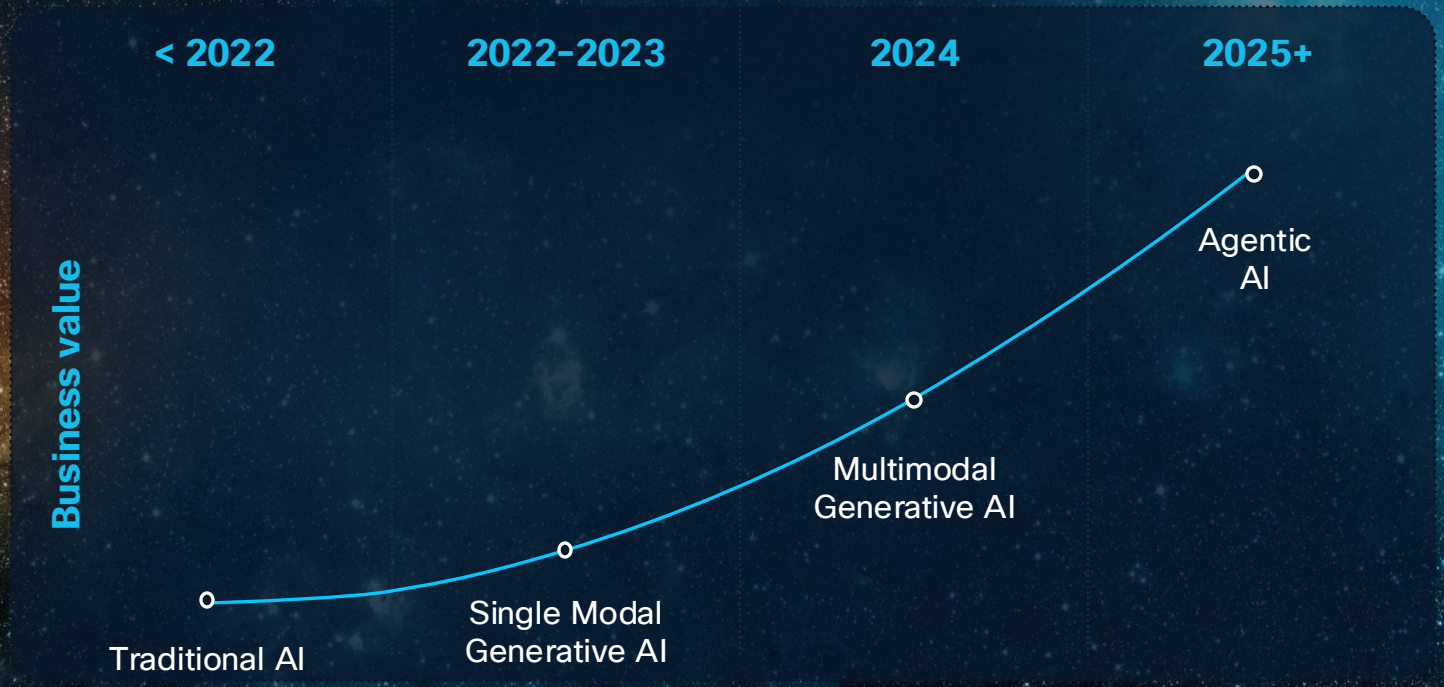
2022
ChatGPT released

2024
65% of organizations regularly use generative AI

2027
62% of total data will be processed at the edge

2030
120% CAGR increase in AI-enriched network traffic

The Rise of Agentic AI



"33% of enterprise software applications will include agentic AI by 2028, up from less than 1% in 2024"

– Gartner

"By 2028, 68% of all customer service and support interactions with technology vendors are expected to be handled by agentic AI."

– Cisco Report:
Race to an Agentic Future

AI and GenAI use cases across industries

Knowledgebase copilots

AI assistants

Content & code generation

Text | Images | Video | Code

Security

Physical | Virtual

Language translation

Multilingual real-time communication

Virtual agent & chatbots

Specialized domain specific chatbots

Detection, prediction & reporting

Analytics | Forecasts | Anomalies | Insights

cisco

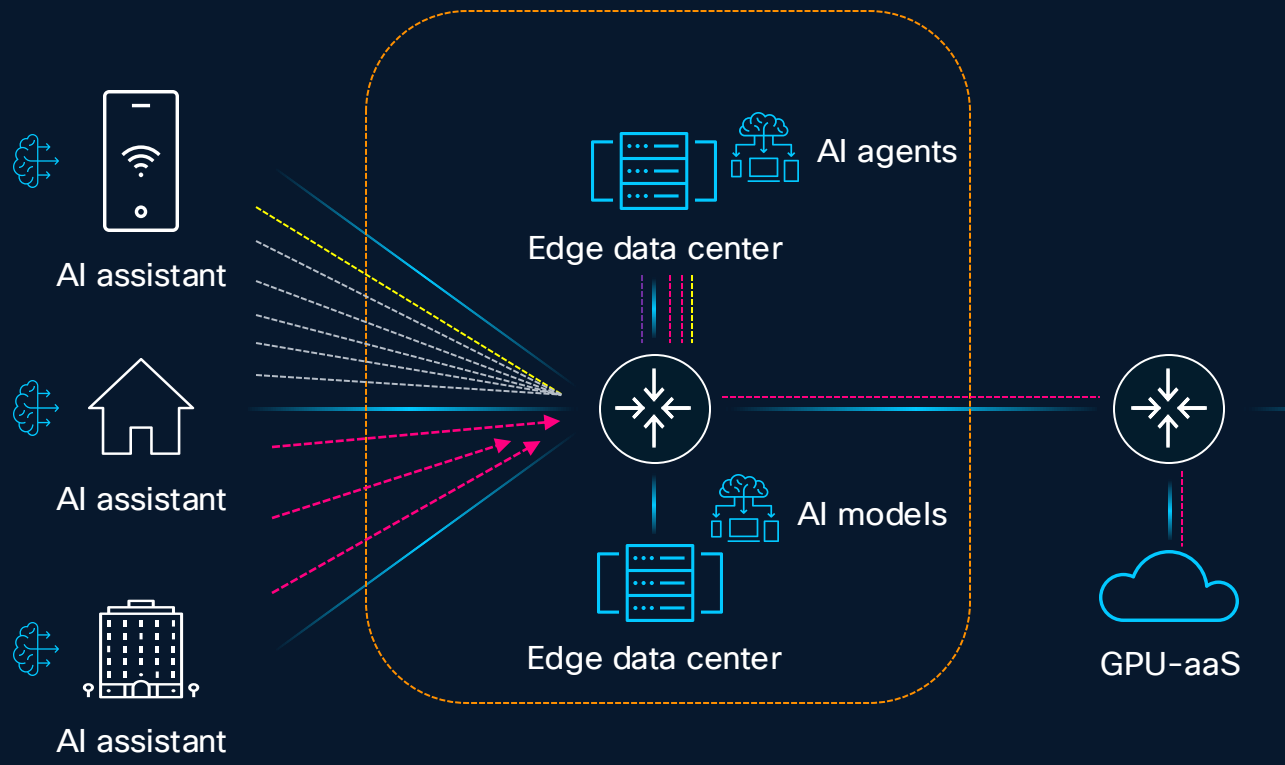
What Makes AI Traffic Different

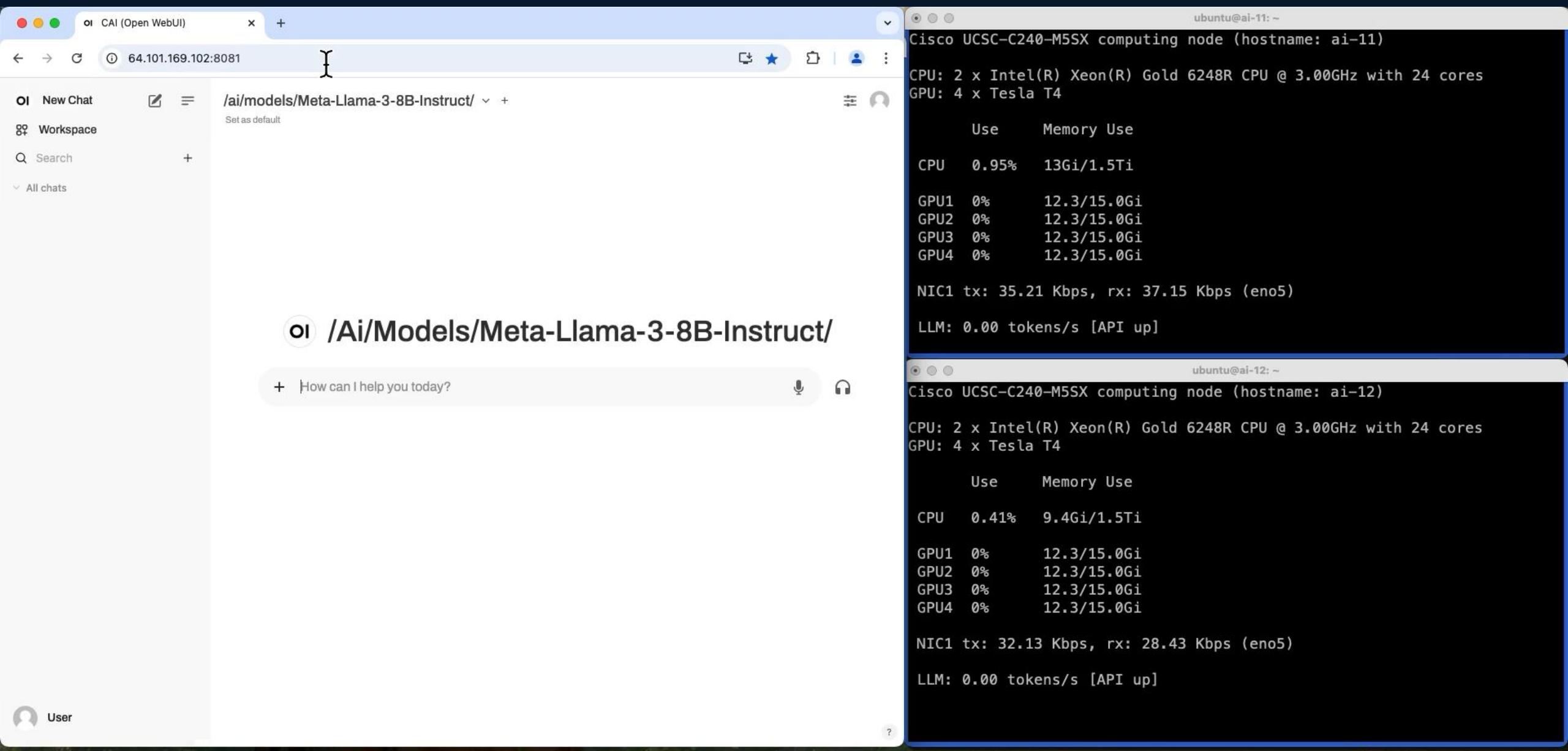
Every AI request is unique – no content is cached

AI requests generate a high volume of tokens

AI inferencing is needed at the edge, leading to new network requirements

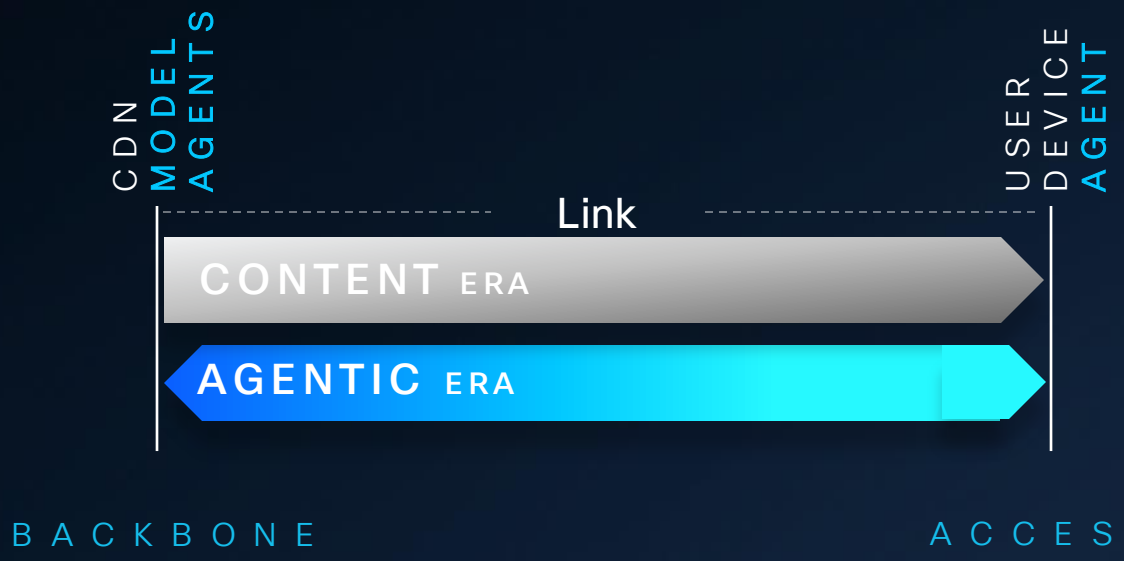
AI drives a disproportional increase in upstream traffic



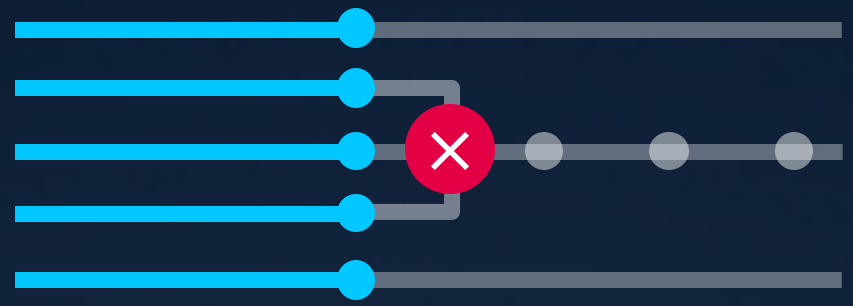

CAI (Open WebUI)
64.101.169.102:8081
New Chat
Workspace
Search
All chats
/ai/models/Meta-Llama-3-8B-Instruct/
Set as default
/Ai/Models/Meta-Llama-3-8B-Instruct/
How can I help you today?
User
ubuntu@ai-11: ~
Cisco UCSC-C240-M5SX computing node (hostname: ai-11)
CPU: 2 x Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz with 24 cores
GPU: 4 x Tesla T4
Use Memory Use
CPU 0.95% 13Gi/1.5Ti
GPU1 0% 12.3/15.0Gi
GPU2 0% 12.3/15.0Gi
GPU3 0% 12.3/15.0Gi
GPU4 0% 12.3/15.0Gi
NIC1 tx: 35.21 Kbps, rx: 37.15 Kbps (eno5)
LLM: 0.00 tokens/s [API up]
ubuntu@ai-12: ~
Cisco UCSC-C240-M5SX computing node (hostname: ai-12)
CPU: 2 x Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz with 24 cores
GPU: 4 x Tesla T4
Use Memory Use
CPU 0.41% 9.4Gi/1.5Ti
GPU1 0% 12.3/15.0Gi
GPU2 0% 12.3/15.0Gi
GPU3 0% 12.3/15.0Gi
GPU4 0% 12.3/15.0Gi
NIC1 tx: 32.13 Kbps, rx: 28.43 Kbps (eno5)
LLM: 0.00 tokens/s [API up]

More traffic, Upstream tilt. Resiliency is everything.

AI agents operating at machine speed increase total traffic and tilt direction upstream.



Downtime no longer leaves users waiting – it brings digital workers to a halt.



One inference failure halts every downstream step.
1% packet loss means 40-60% performance degradation.

cisco

The network becomes an LLM Fabric

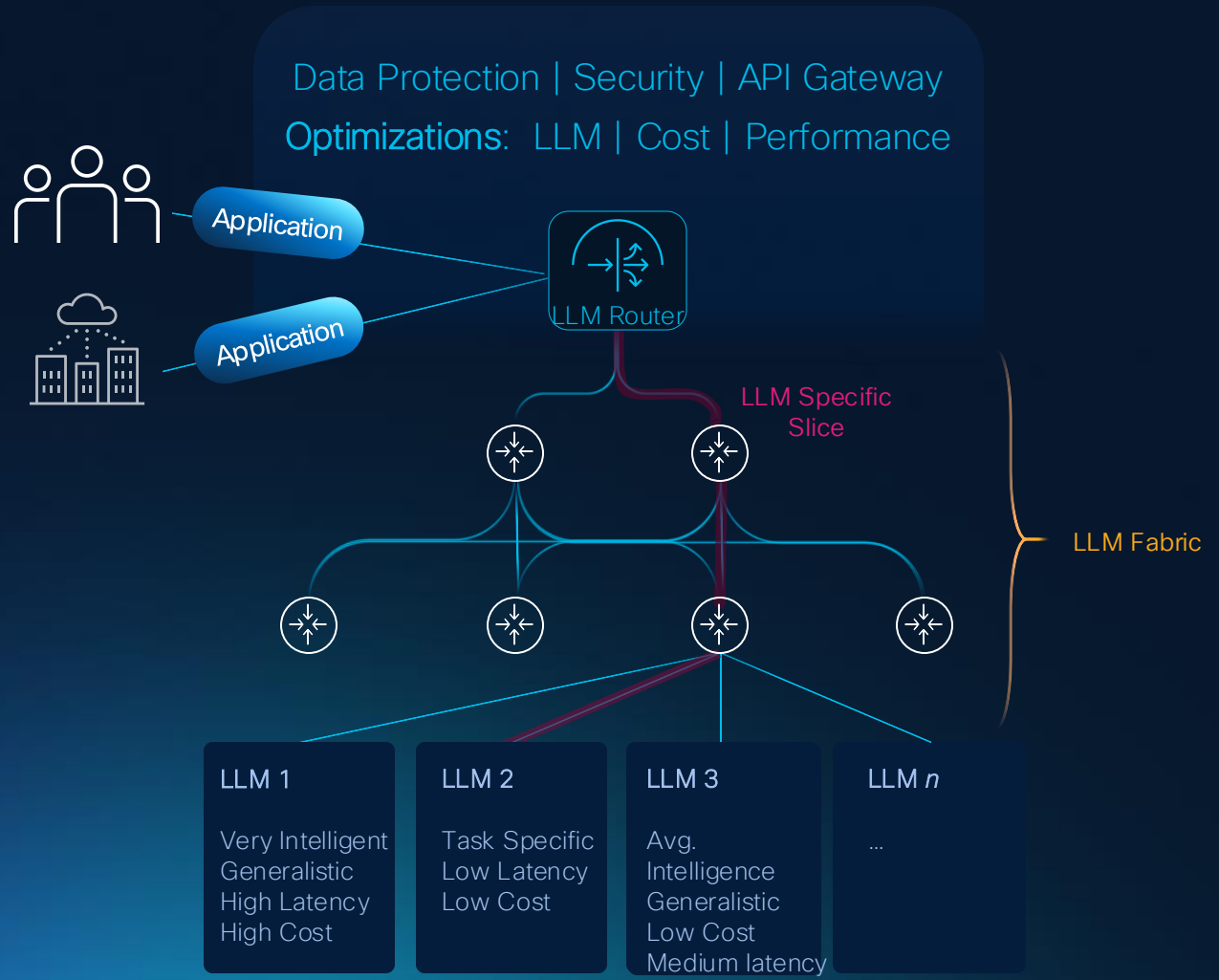
Cloud and Hybrid

Regulation and Data Sovereignty

Multi-LLM Challenges

Stringent SLA Requirements and Visibility

Distribution of Agents and LLMs



Cisco Strategy

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

AI-Ready Data Centers

Future-Proofed Workplaces

SECURE GLOBAL CONNECTIVITY

Digital Resilience

Accelerated by Cisco AI

Challenges in Secure Global Connectivity

Cost

How can I afford the increase in high-bandwidth low-latency connections?

Operations

How do I manage new AI workloads without more networking engineers?

Performance

How do I assure and secure the performance of mission-critical applications?

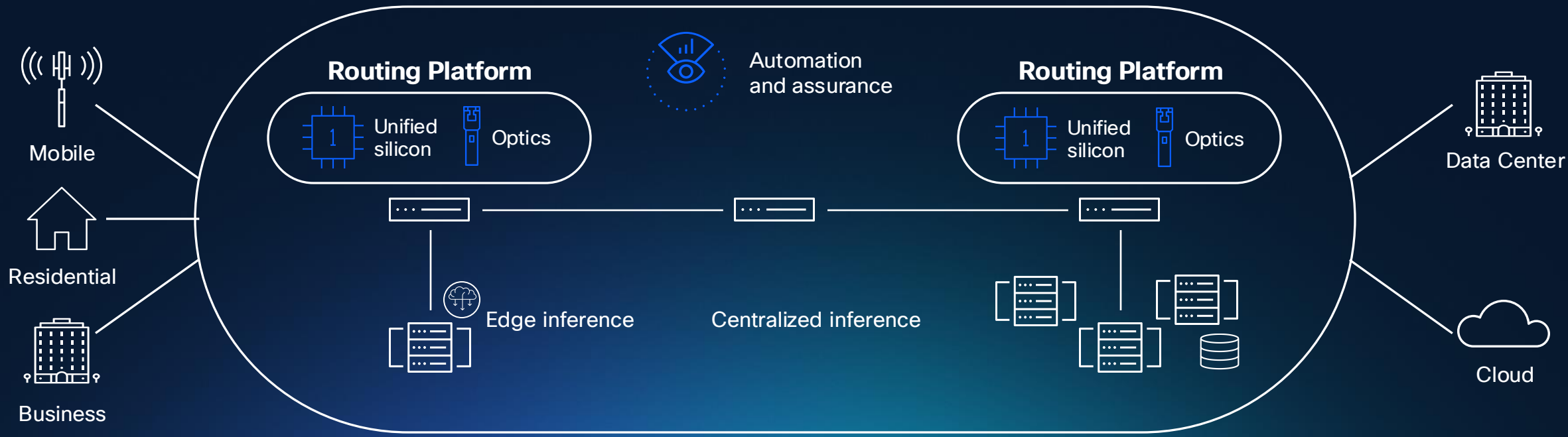
Our response

Agile Services Networking

Cisco Agile Services Networking

Predictable, assured, and secure network performance for AI connectivity

End-to-end architecture | Any services, anywhere

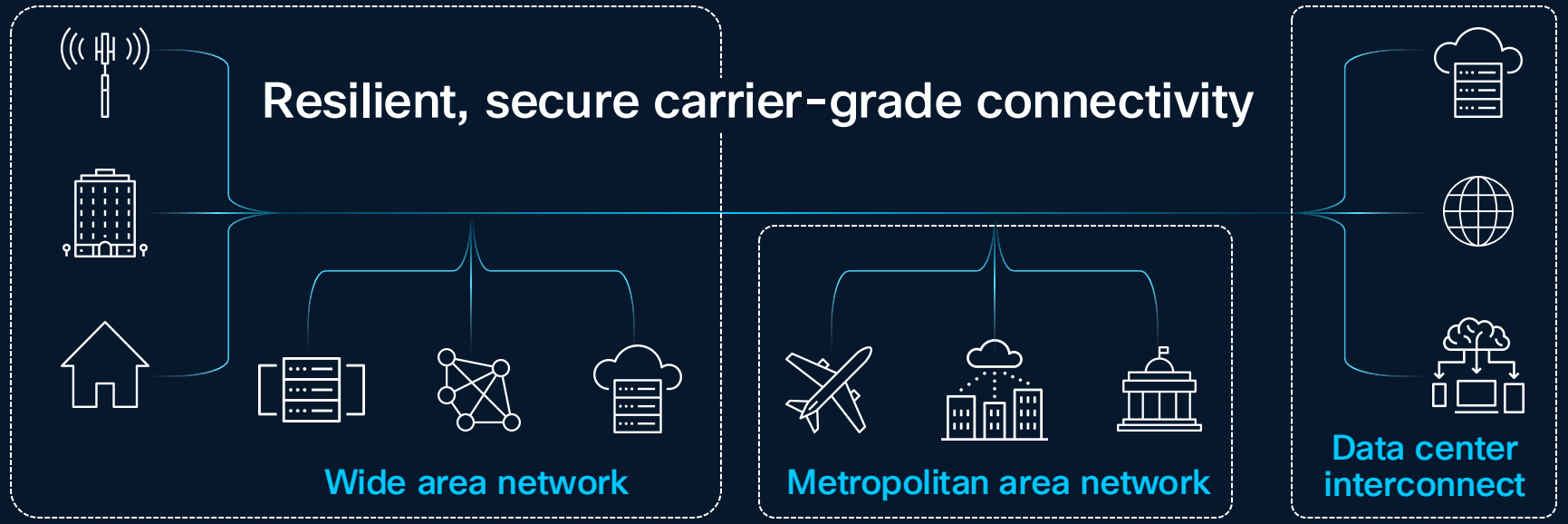


Common AI connectivity needs across all sectors

Scale network capacity
Distribute100G to 800G data rates closer to user demand

Improve quality of service
Migrate from traditional MPLS to segment routing and SRv6

Protect assets
Increase visibility and security across the network



Manufacturing | Automotive | Energy | Finance | Education | Cities | Retail | Government | Healthcare | Telcos

Cisco Innovation

Agile Services Networking

Common Silicon

IP/Optical Convergence

End-to-End Network Programmability

Unified Observability and Assurance

Embedded Security & Trust

Common Silicon One family set

Secure, scalable & controlled

Consistent building blocks

Optical extended into the data center

Coherent pluggable optics

Routed Optical Networking

Seamless E2E service delivery

Seamless DC-to-SR/SRv6

Seamless E2E SR/SRv6

Cross-domain service insights

Observability/Visibility

Data platform & correlation

Pervasive, built-in security

Trustworthy platforms

Quantum safe

cisco

Cisco Silicon One

Scalable and programmable NPUs for front-end, back-end, Core, Edge, Metro and Access networks. Consistent control, sustainability, security, and manageability from the silicon level.

- One architecture
- High capacity and scale
- Service convergence
- Optimal network design
- Fully shared packet buffer
- Flexible interfaces
- Traffic management and load balancing
- Comprehensive security
- Programmable services
- Deep visibility and analytics

Cisco Silicon One | Not one chip, One Architecture

End-to-end Coverage

1 Architecture
The Language of block communication

5 Series
Assemble market optimized blocks

0 Compromises
Common language, Market optimized blocks

Easier to validate, deploy, innovate and troubleshoot

AI Optimized Data Center Switch
Hyperscale Front End
AI Scale-Up
AI Scale-Out
AI Scale-Across

AI Optimized Scale-Across Router
AI Secure Scale-Across
Wide Area Networks (WAN)
Data Center Interconnect (DCI)

Feature Rich Switch
Enterprise Data Center Leaf
Enterprise Campus core

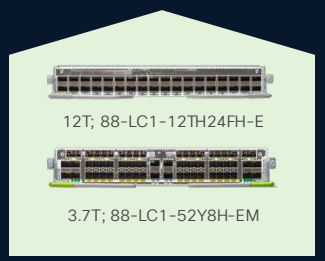
Feature Rich Router
Agile Services Edge
Enterprise Campus Core

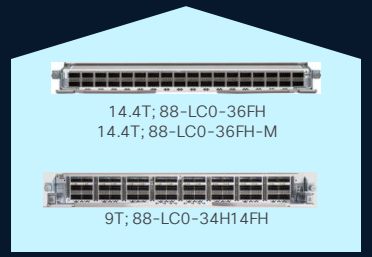
Feature Rich Access
Converged Access
Enterprise Access

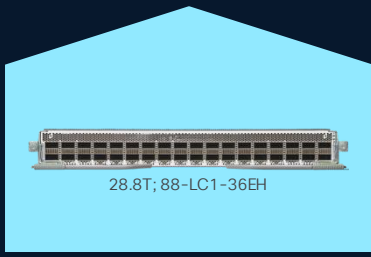
Cisco 8000 Routing Portfolio

One Family from Access to Core with IOS-XR

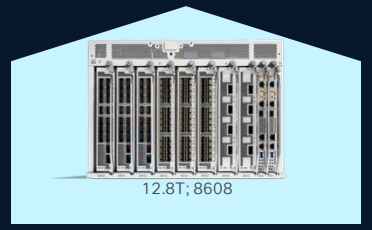
Distributed



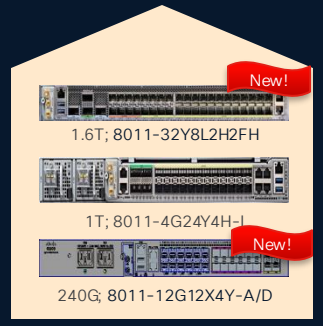


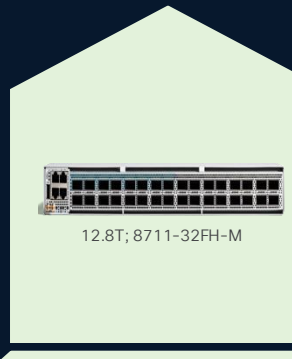


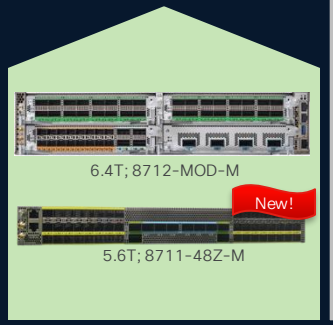
Centralized



Fixed





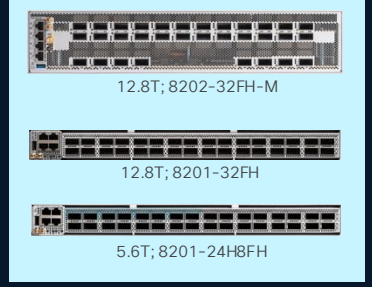


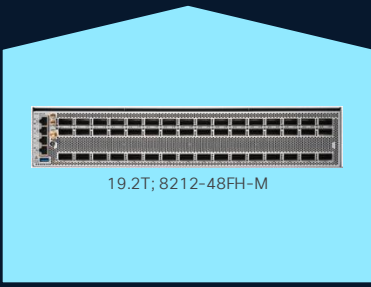
Cisco 8000 Series

Powered By

Silicon One NPU

Cisco IOS® XR





A100 3rd Gen NPU

P100 3rd Gen NPU

K100 3rd Gen NPU

Q200 2nd Gen NPU

P100 3rd Gen NPU

Access

Aggregation & Edge

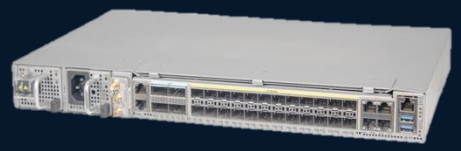
Core & Peering

Powered by Cisco IOS XR

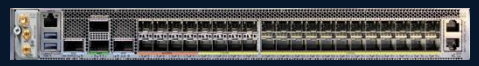
Cisco Routing Platforms for Enterprise and Public Sector

Access routers

Silicon One A100



8011-4G24Y4H-I
1T 1RU



8011-32Y8L2H2FH
1.6T 1RU

8011-12G12X4Y-A/D
240G 1RU

Services scale and feature set for access and aggregation: Class C Timing, MACsec and IPsec encryption, and QoS features

Edge Routing

Silicon One K100

8712-MOD-M
6.4T 2RU

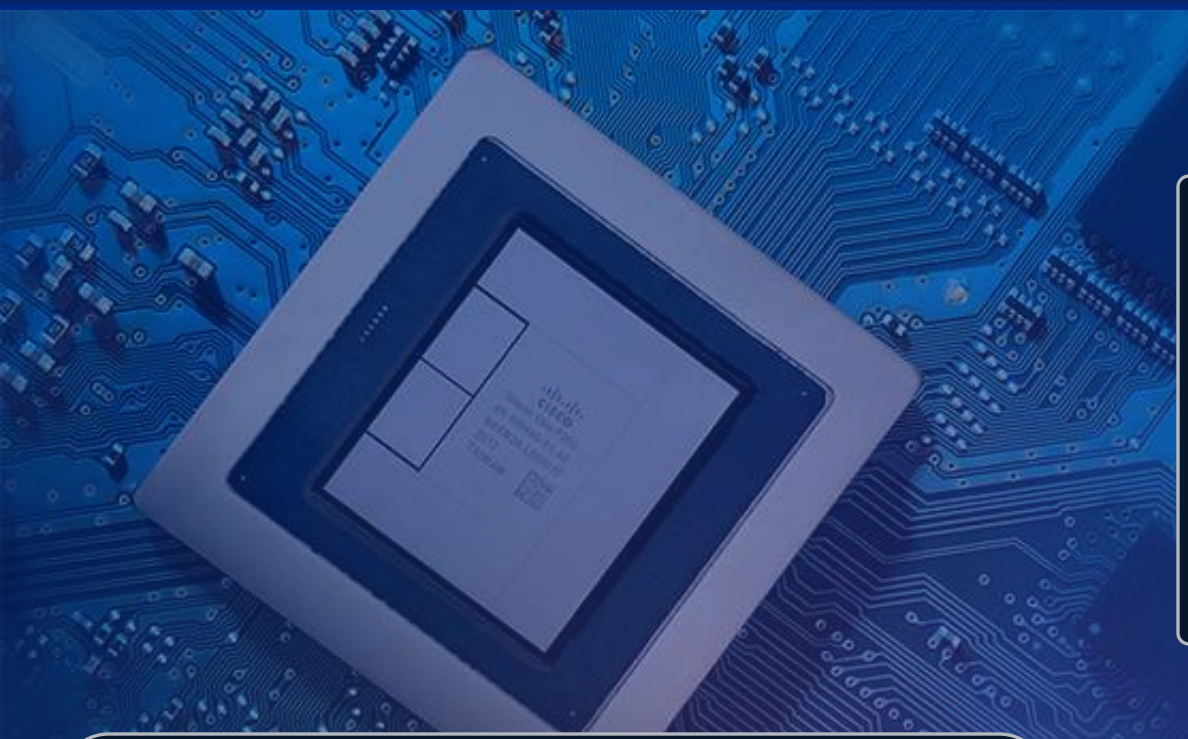
8404-SYS
4.8T 4RU

8711-48Z-M
5.6T 1RU

Combines the efficiency and simplicity of a single processor with fixed and centralized form factors

Innovation in AI Scale-Across

Cisco 8223 Powered by Cisco Silicon One P200



Cisco 8223-64EF

51.2T, 3RU, 64 x 800G (OSFP800)

Cisco 8223-64E

51.2T, 3RU, 64 x 800G (QSFP-DD800)

P200- 51.2T Router Chip

Power Efficient. Scalable.
Programmable. Secure.

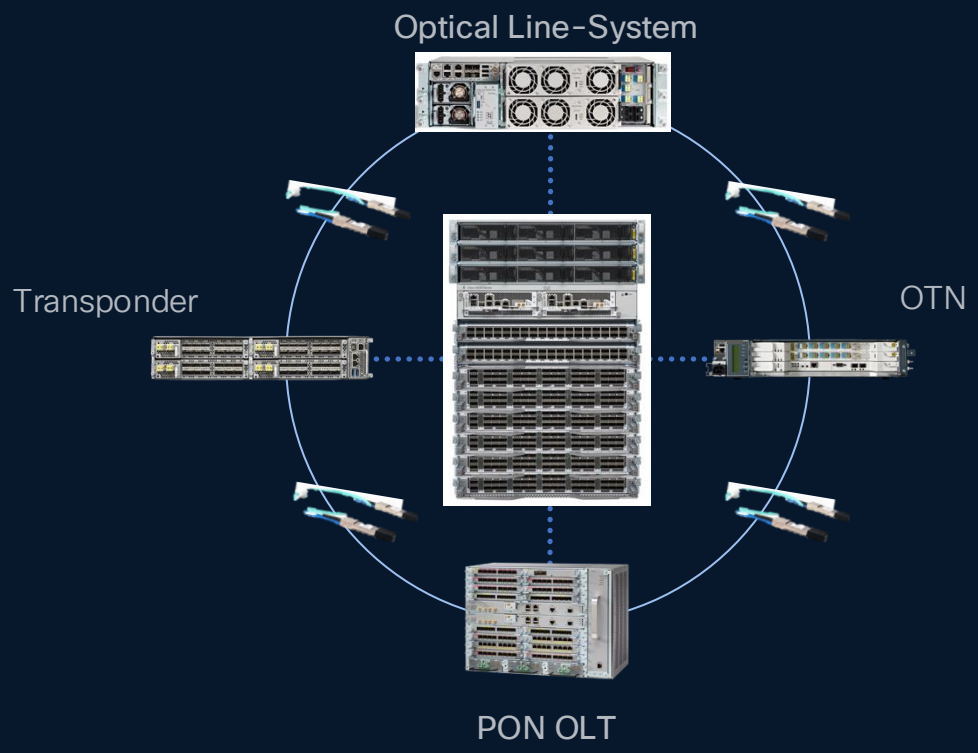
8223 Routing Systems

Industry's most scalable, efficient, secure 51.2T Routing System for Scale-Across

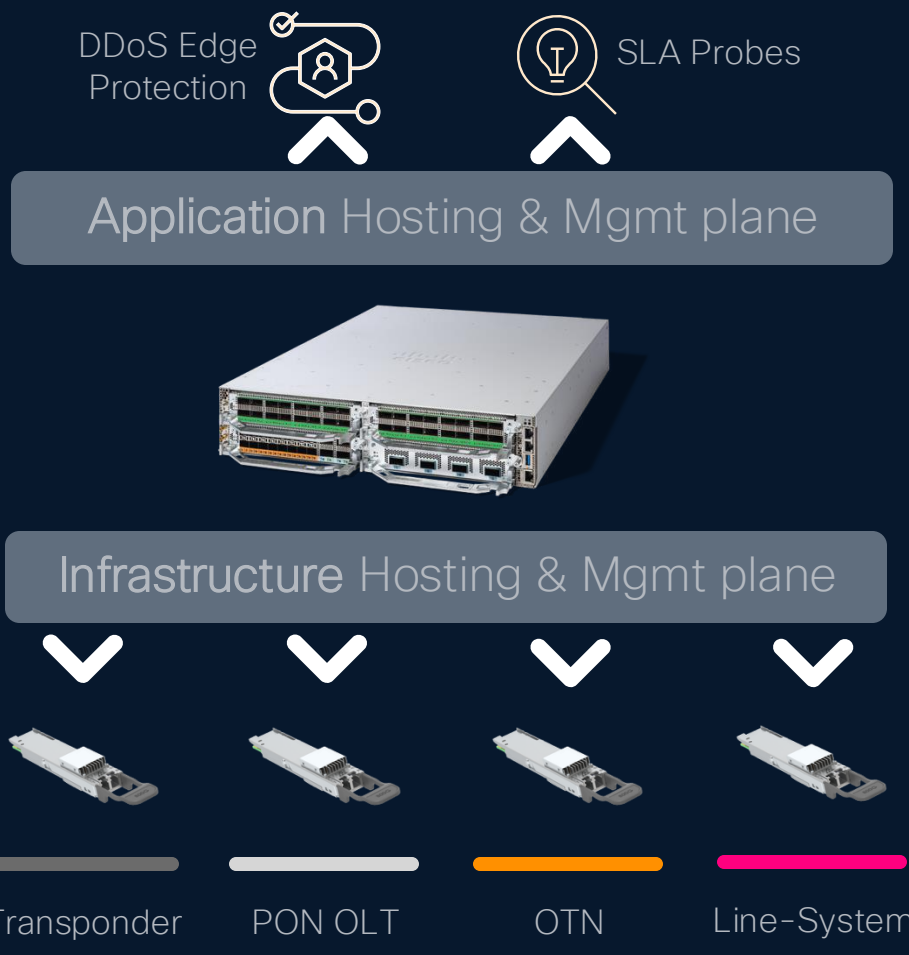
cisco

Innovation in Optical Systems

(Redundant) Classic infrastructure



Cisco Agile Services Networking



Innovation in Digital Coherent Optics

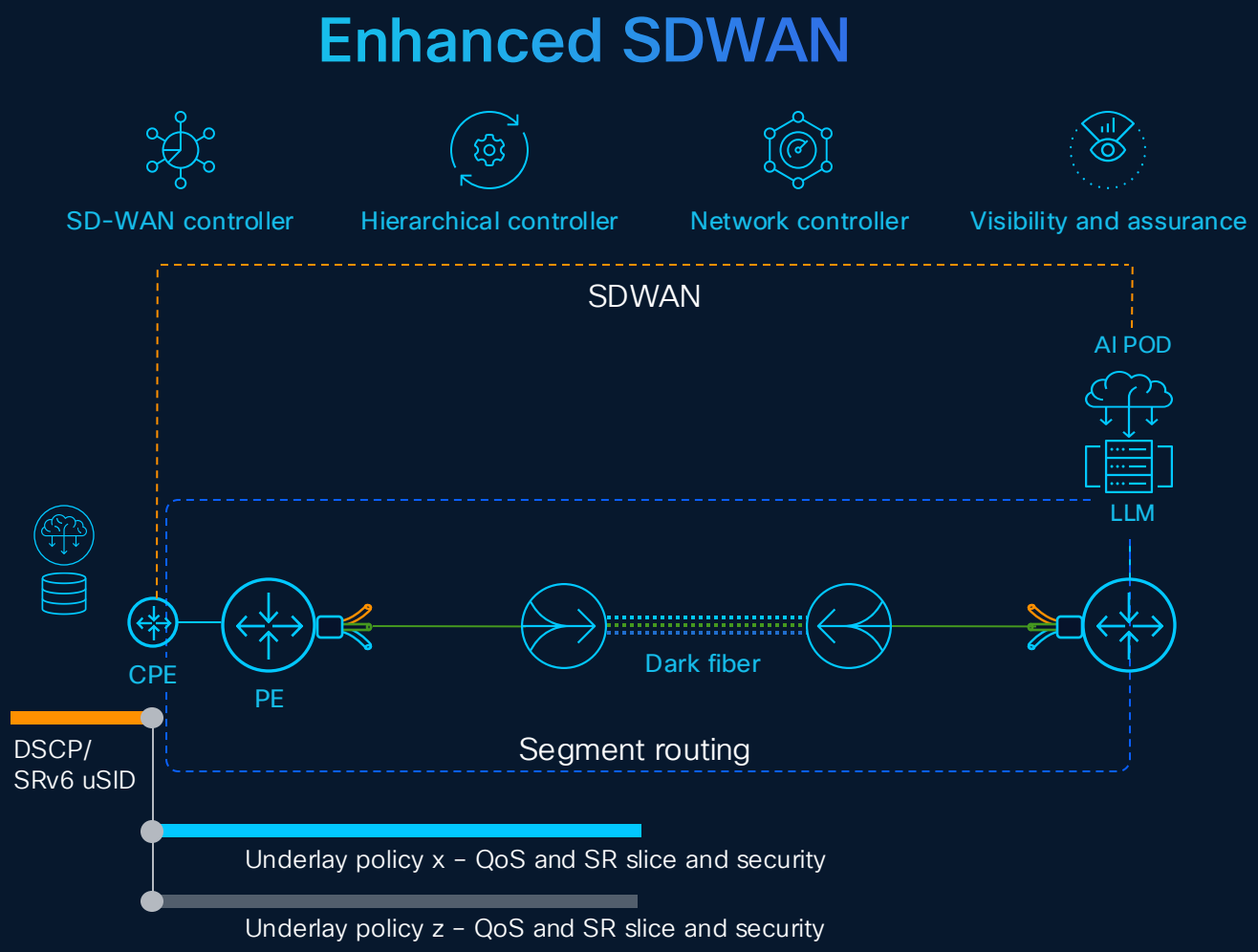
Address 100% of network applications with Cisco coherent pluggable optics

100G QSFP28 ..... 800G ZR/ZR+ 400G ZR/ZR+ ..... 400G ULH

Campus | DCI | Metro | Regional | Ultra Long Haul

Innovation in AI Application Performance

SLA Enforcement and Performance KPIs through overlay/underlay integration



SD-WAN overlay integrated with SR underlay

- SD-WAN service SLA mapped to SR slice, QoS and security policy
- Integrated service visibility - Network and SD-WAN data correlated for an integrated overlay/underlay service view

Integrated service view

- PCA collects network and SDWAN controller data for an integrated overlay/underlay service view
- Insights for closed-loop assurance
- Monitoring of AI transport KPIs

cisco

Automation & Assurance Innovations

Domain SMEs

NOC / SOC, End Customer

Crosswork Hierarchical Controller
Multi-layer, multi-domain IP and Optical hierarchical controller

New!

- IPv6/SRv6 support
- Northbound notifications
- CONC, EPNM, Nokia, Ciena, Huawei optical adapters

Cisco Crosswork Cloud

- **Trust Insights** — Track integrity of infrastructure
- **Network Insights** — Analyze and identify the source of routing anomalies
- **Traffic Analysis** — Optimize network traffic at peering points

Optical Network Controller
Optical Controller

New!

- Single-VM Optical EMS + Controller
- Optical Assurance & Orchestration

Crosswork Network Controller
Multivendor IP SDN Controller

New!

- Single-VM CNC and KVM support
- Geo-redundancy
- Multi-vendor EMS
- Operations automation via CWM

Provider Connectivity Assurance
Multivendor Service Assurance

New!

- Digital Experience Assurance
- Telemetry consumption
- Predicting Performance
- Topology and ECMP analytics

New!

Crosswork Workflow
Low-code automation workflow

- New CNC Platform
- Adaptative Forms

New!

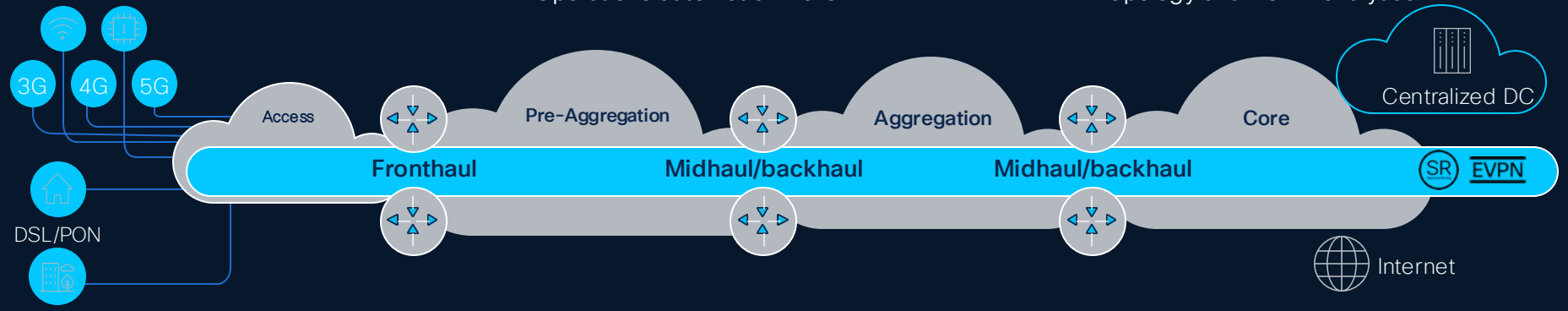
NSO
Service Orchestration

- AI Doc Chatbot
- Brownfield Scenarios
- NSO for Controllers
- Enhanced compliance reports

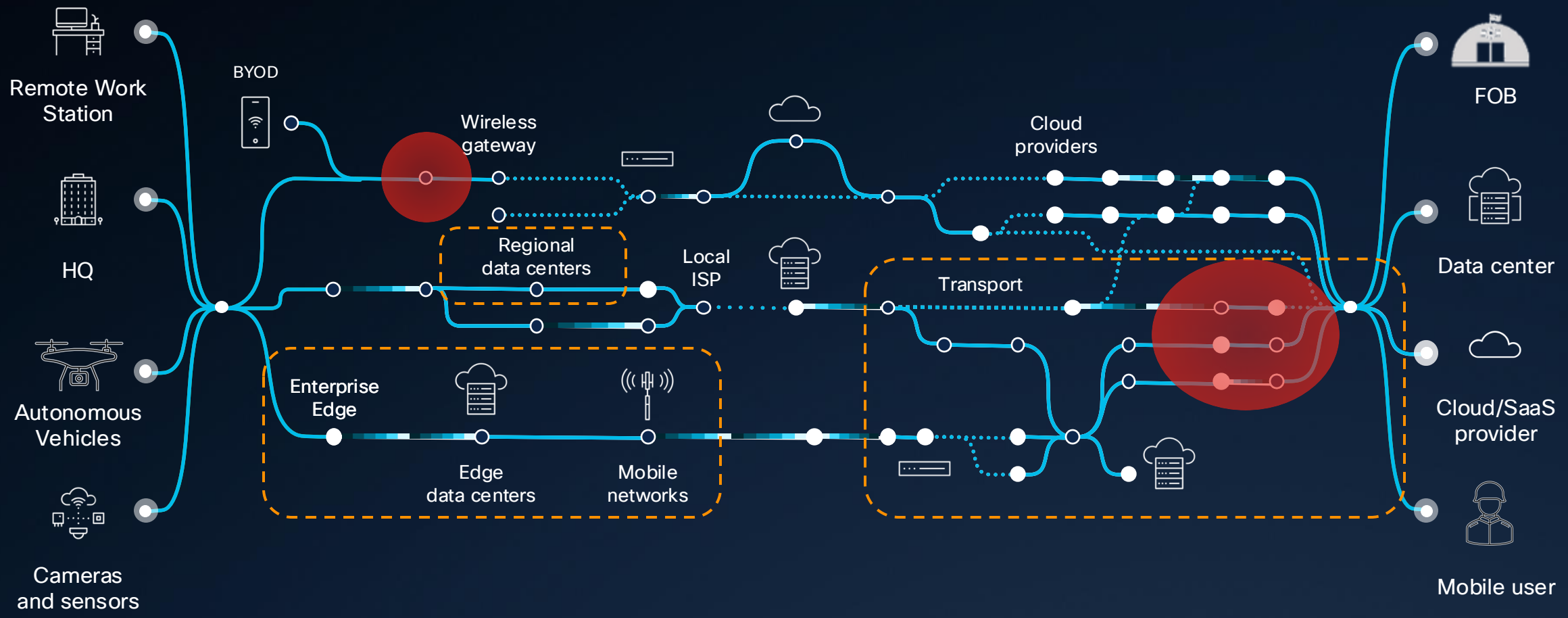
New!

Crosswork Planning
Capacity planning tool

- Predictive-AI capacity planning
- Crosswork Planning Chatbot



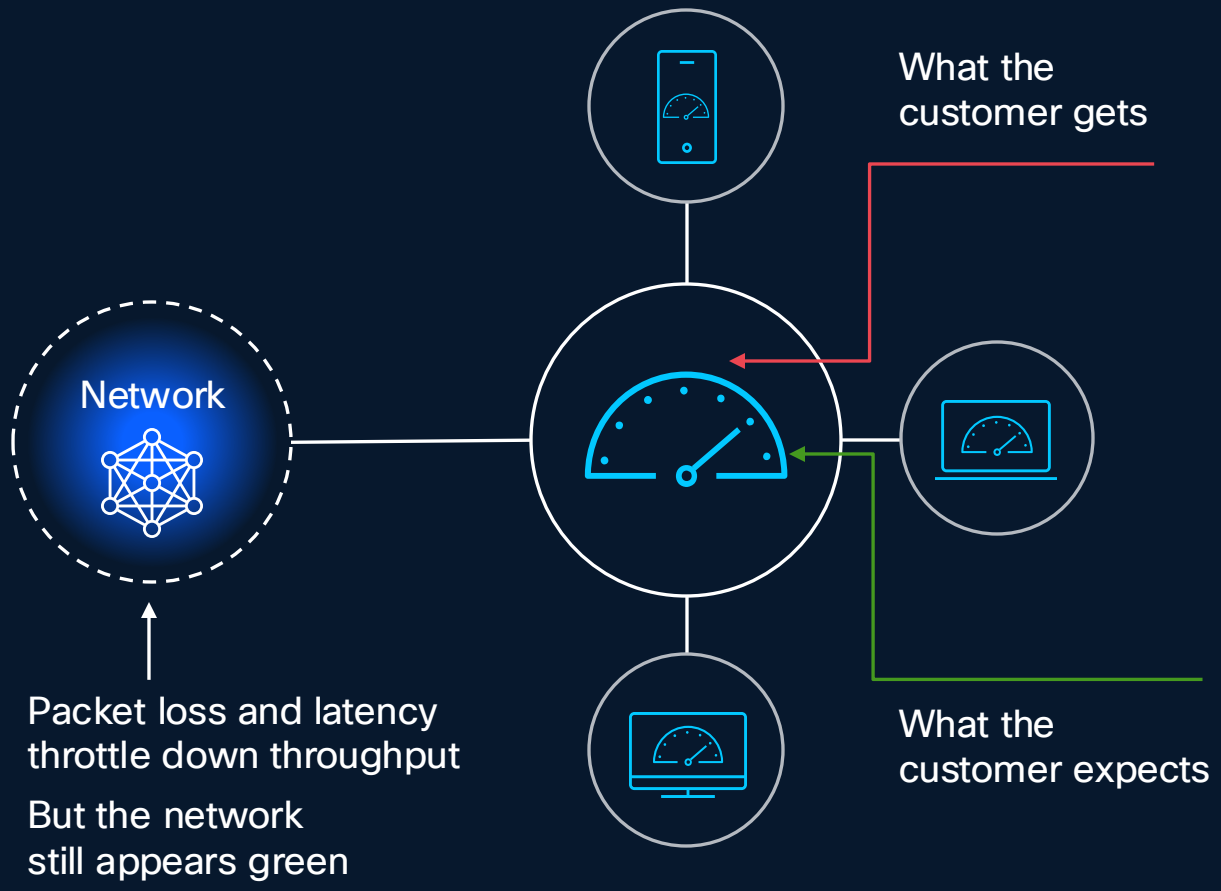
Assure critical networks with Connectivity Assurance



Most tools are ineffective at detecting today's QoE issues

Without deep visibility into micro events:

- 0.53% packet loss leads to a 50% decrease in data throughput
- 5 ms delay leads to a 10% decrease in data throughput
- 10 ms jitter leads to a 10% decrease in data throughput

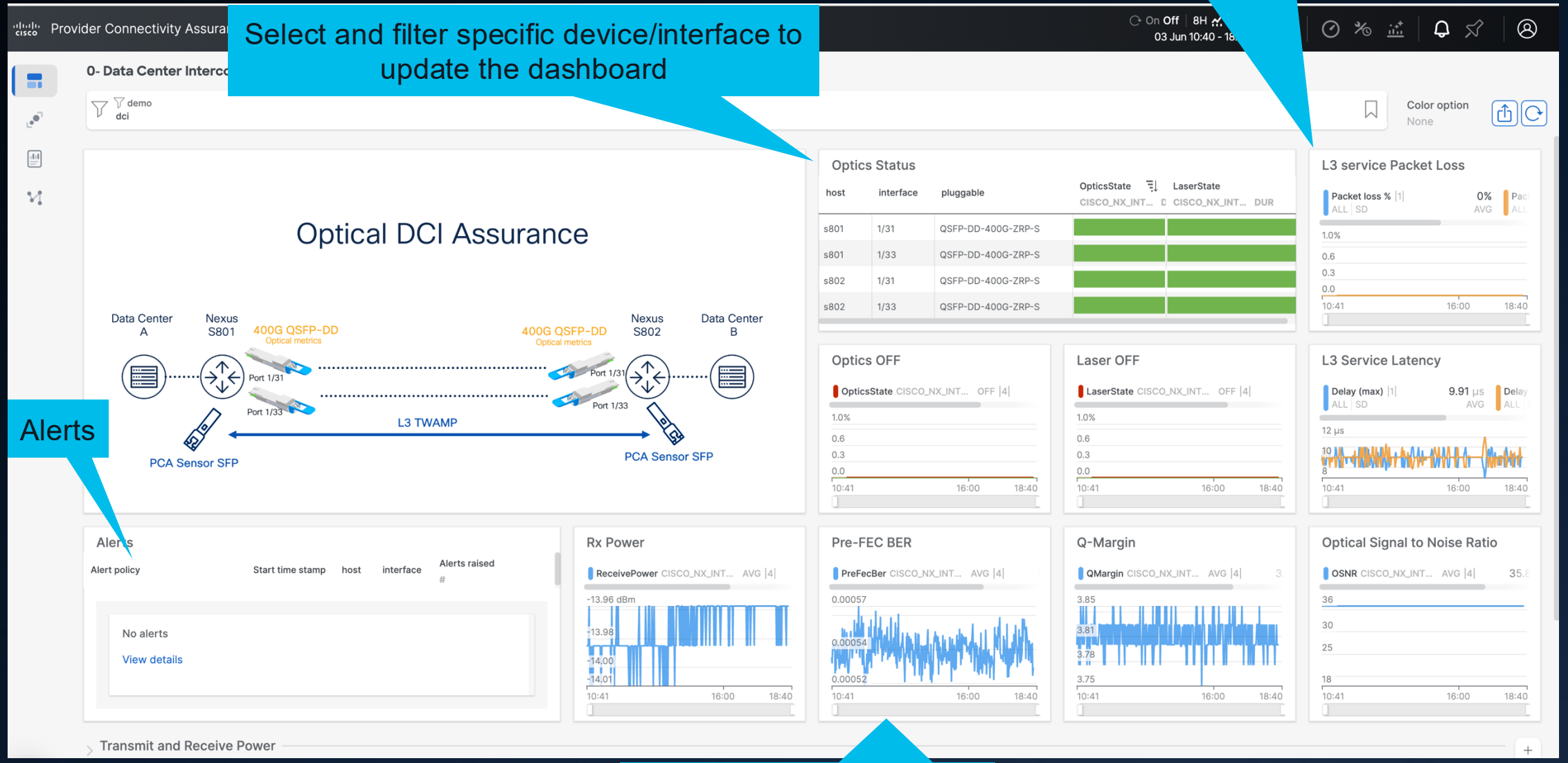


The quality of the network has a major influence on user quality of experience (QoE)

PCA for DCI assurance

L3 Packet Loss and Latency

Select and filter specific device/interface to update the dashboard

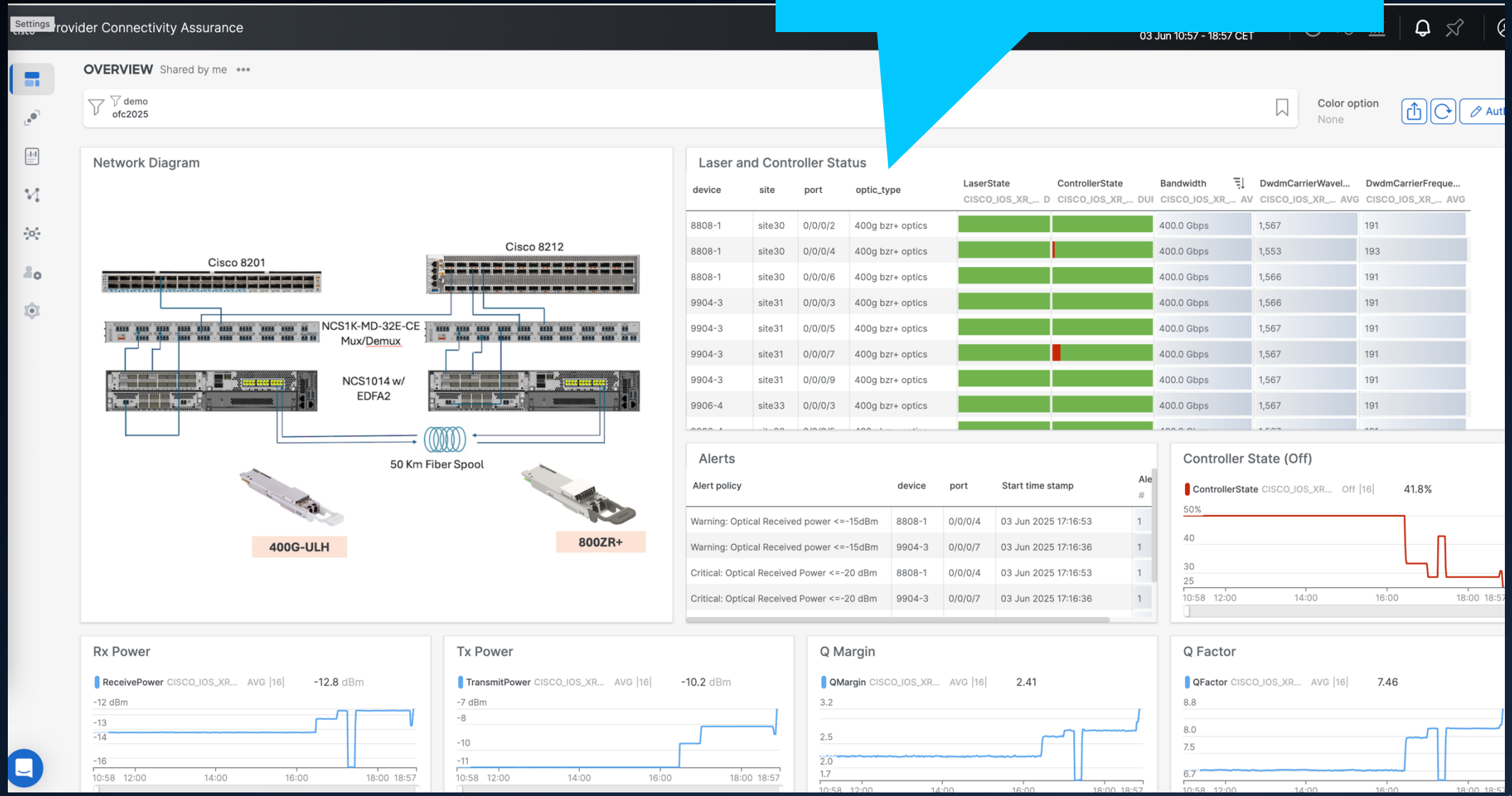


Alerts

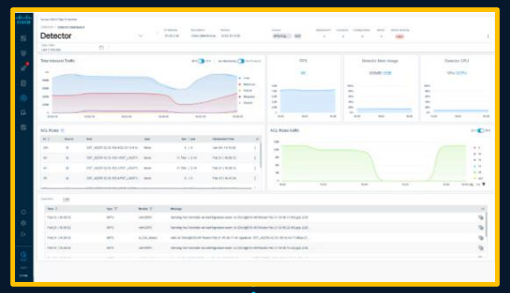
Various optical metrics

PCA Dashboard

KPIs from different optic types and devices
- Similar to DCI dashboard -



Secure Networking Innovations



1

Establish trust at the physical layer

Verify that HW and SW are unaltered, even in untrusted domains

2

Protect from threats in transport

Use routers as your first line of defense against DDoS attacks

3

Protect from emerging threats

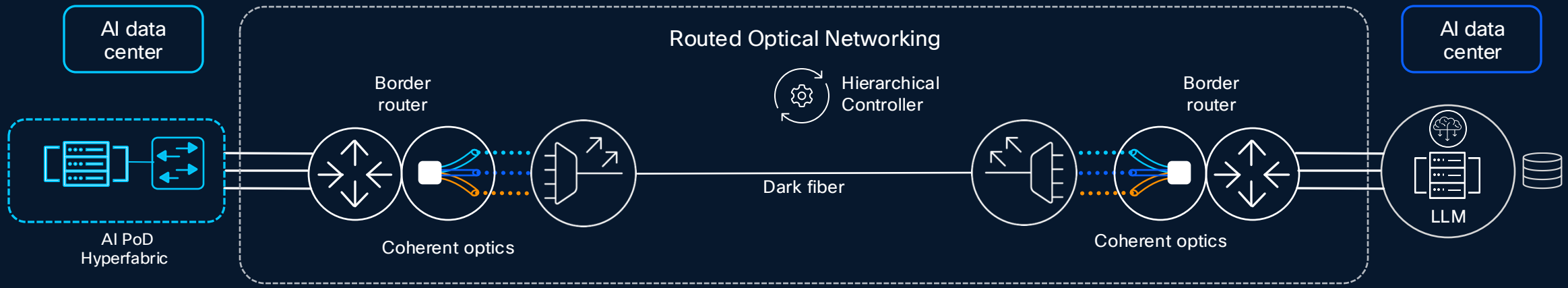
Implementing secure key distribution for post-quantum threats

cisco

Use Cases

AI-ready DCI	Scalable high capacity and high-performance data center interconnect
AI-ready WAN	Reliable, secure, programmable, power, and cost-effective networking for SLA-based networking

AI-Ready data center interconnect



Key benefits

- Cisco Routed Optical Networking replaces power-hungry transponders with small pluggable optics (converge IP and optical layers)
- Suitable for any scale: from 100s of Gbps to 10s of Tbps to meet growing demands of AI
- Broad platform compatibility (e.g., Cisco 8000 and Nexus)
- Delivers up to 66% TCO savings, 95% power savings and requires no extra space

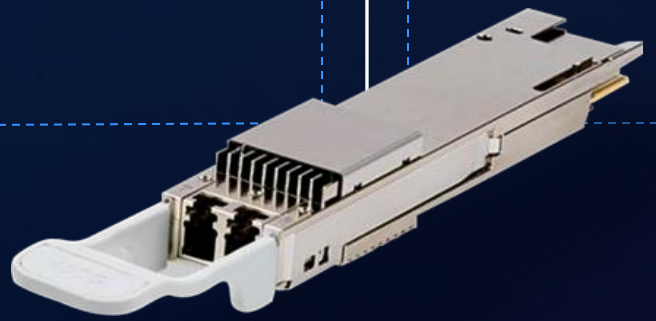
Why Cisco

- Industry leader in optics: largest number of ports shipped to the market and the most complete portfolio, with 100G, 400G, 800G, for different reaches and use cases
- Highest level of quality assurance on optics in the industry
- End-to-end solution: Routers, optics, optical, management/automation
- 350+ customers for Routed Optical Networking and optical customers–including the most demanding hyperscalers in the world

Combining layers with IP + Optical

92%
Power reduction

66%
CAPEX reduction



TRANSPONDER

OPTICAL TRANSPORT NETWORK

Device options for connecting data centers

Catalyst SDWAN

- IOS-XE
- Merchant Silicon
- 26.6 Tbps
- Point to Point
- Gray optics to 400G
- No coherent optics
- Application aware

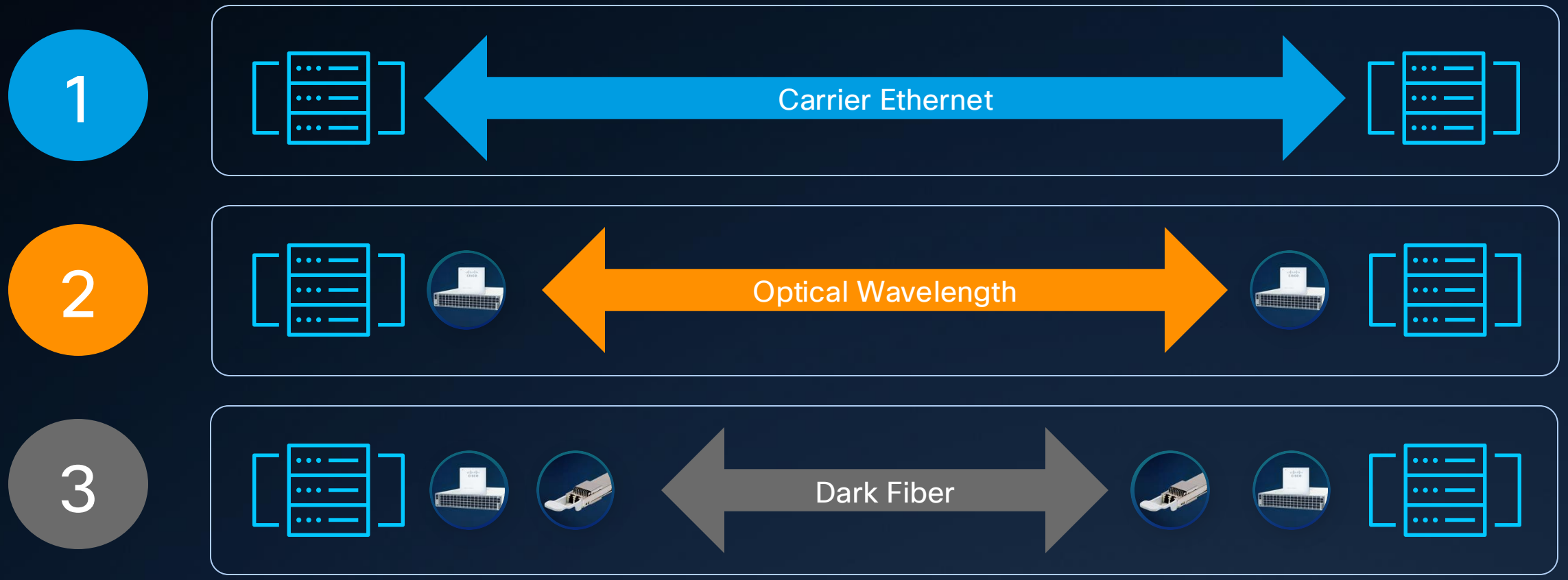
Nexus Switch

- NX-OS
- Silicon One
- 51.2 Tbps
- Point to Point, Low latency
- Coherent optics to 400G (80kms and up to 1,200 kms with amplification)
- Nexus dashboard

XR Router

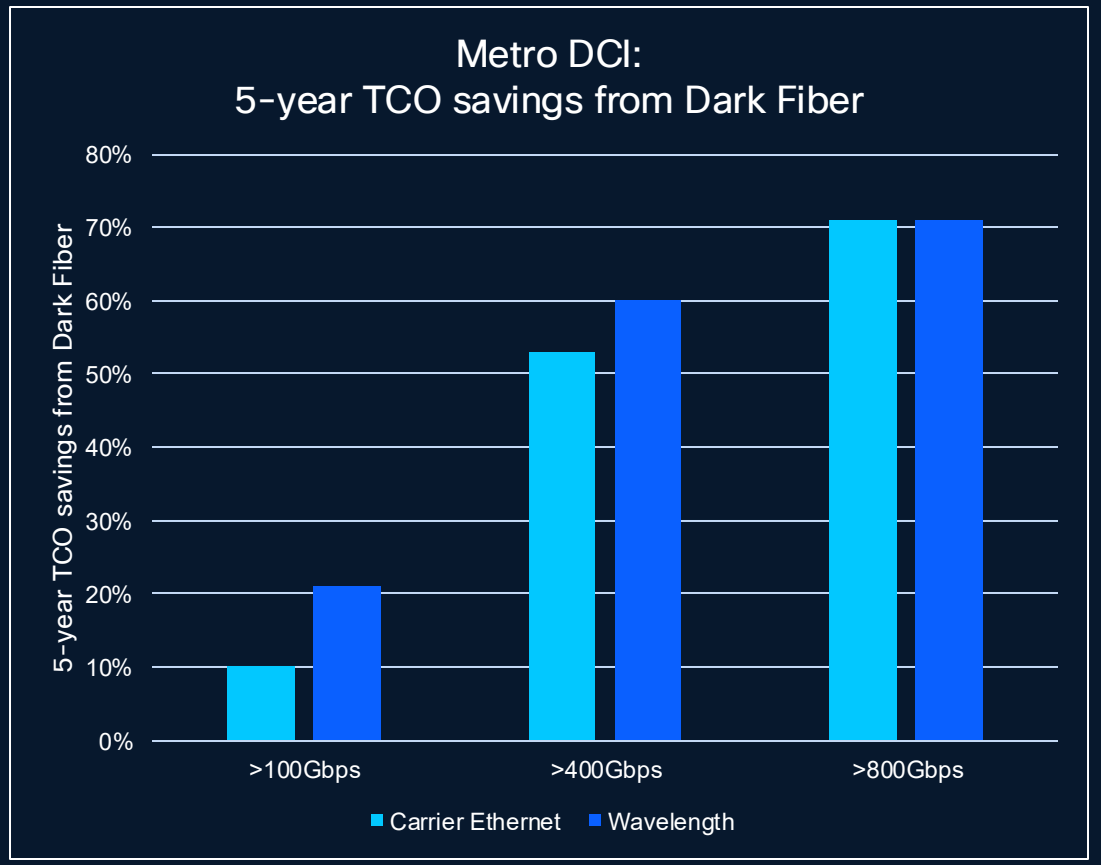
- IOS-XR or SONiC
- Silicon One
- 51.2 Tbps
- Multisite connectivity
- Coherent optics to 400G
- Pluggable OLS
- Cisco Network Controller
- Segment Routing

ACG TCO study on common DCI designs



Key finding from ACG's TCO study

- Dark fiber provides the lowest cost of ownership at 100Gbps:
  - 10% lower than Carrier Ethernet
  - 21% lower than Wavelength services
- Savings with the dark fiber design increase at 400Gbps:
  - 53% lower than Carrier Ethernet
  - 60% lower than Wavelength
- Dark fiber provides the most sustainable upgrade path for growing AI traffic



Download the full ACG TCO study

A customer’s five-year journey

40G

Leased SP circuits fail to provide sufficient bandwidth to operate the business. Made decision to build own DCI.

200G

Required additional bandwidth for hybrid cloud and upgraded existing hardware to 200G and collapsed IP and DWDM.

The start | Phase 1 | Phase 2 | Phase 3

100G

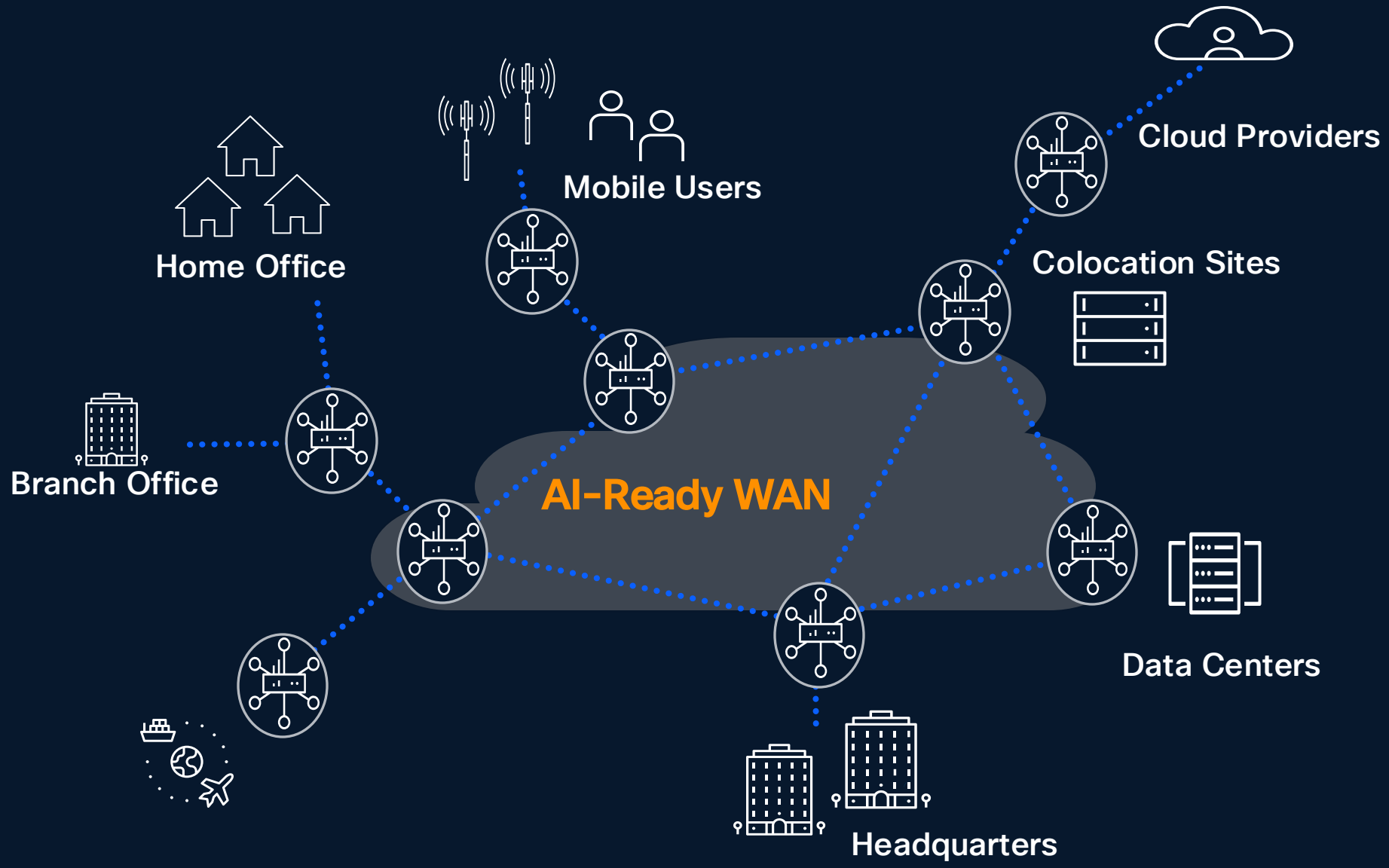
Built a private network with dark fiber. Outsourced installation and NON. Reduced annual operating expense by $4M.

400G

Upgraded the infrastructure to 400G routed optical networking with digital coherent optics. Removed legacy transponders.

cisco

Use Case: AI-Ready WAN



Evolve your network to SRv6 services

Network simplification

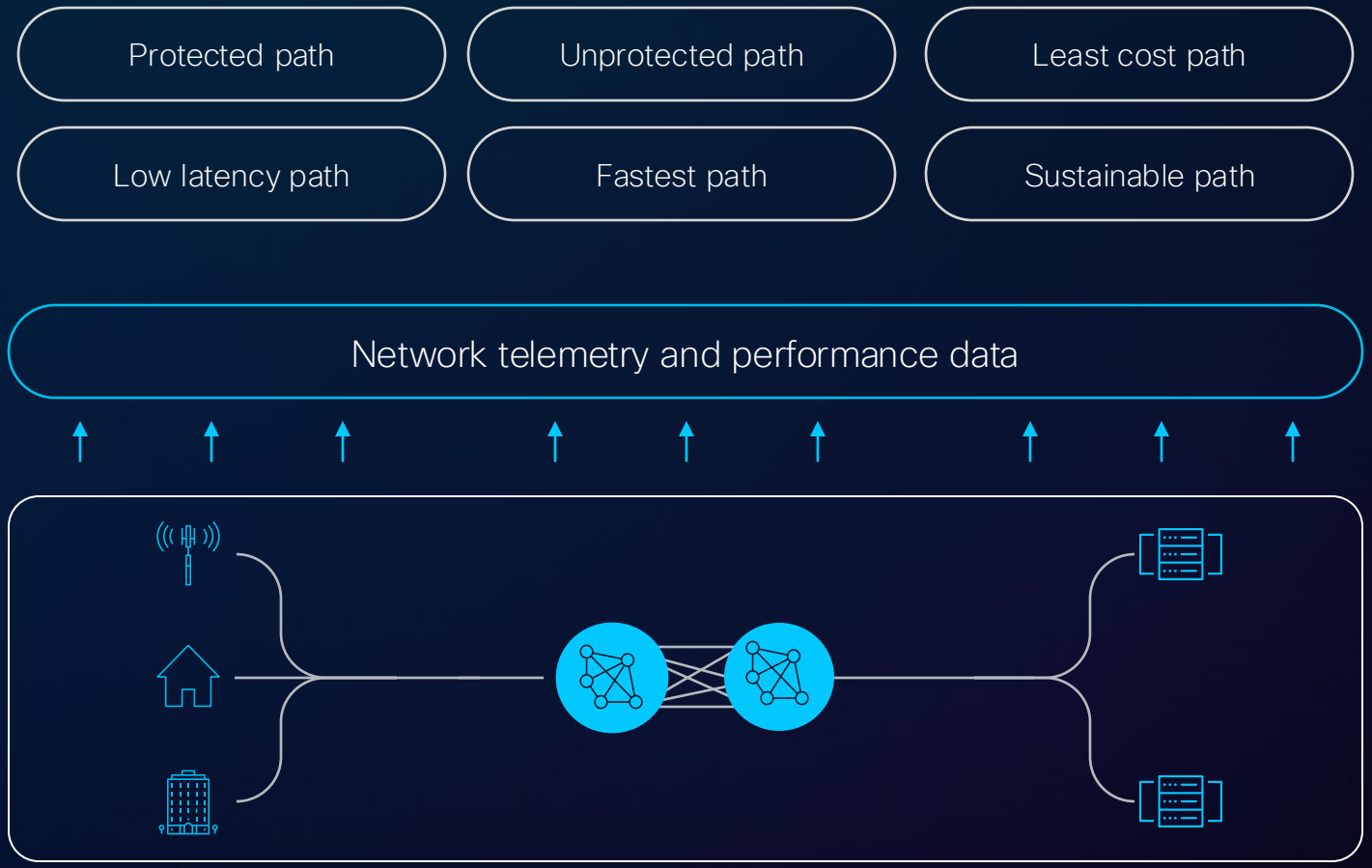
- Automated traffic engineering
- Policy-based routing
- On-demand service provisioning
- Agile to changing network needs

Performance optimization

- Granular and stateless traffic engineering
- Efficient resource utilization
- Reduce latency
- Flexible and scalable design

Intelligent service delivery with intent-based networking

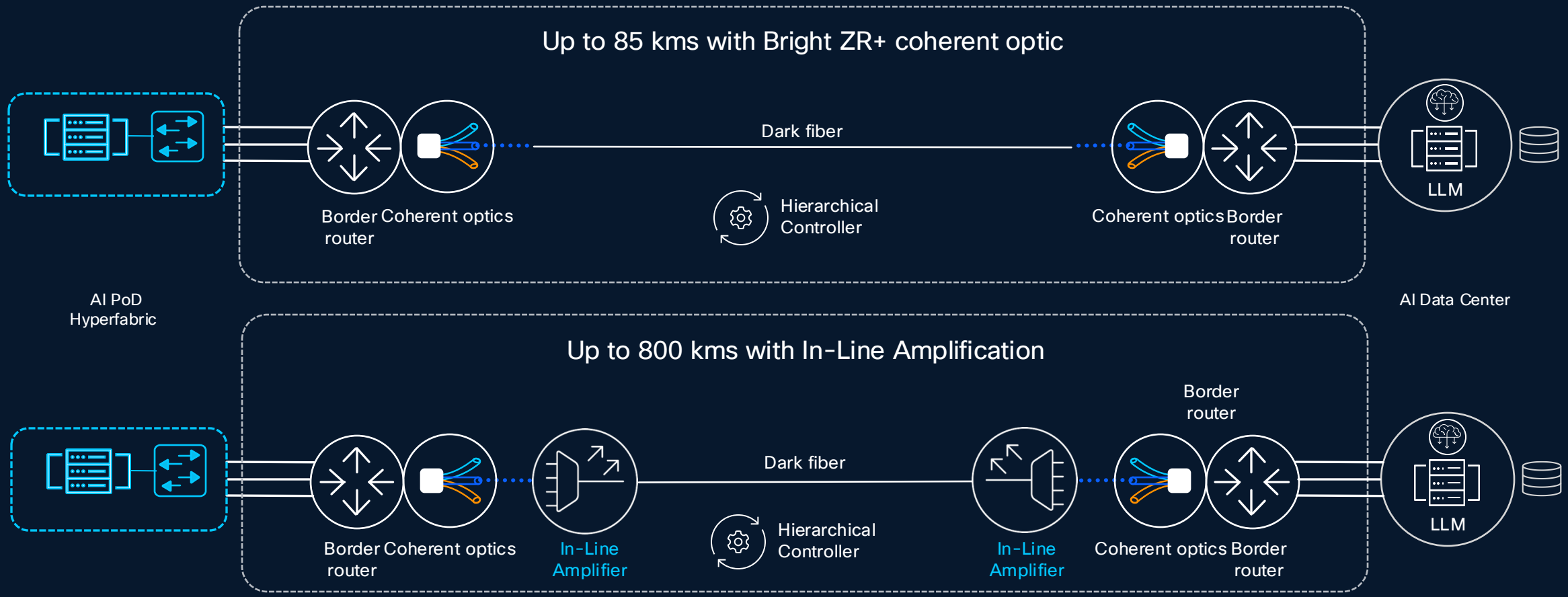
- Cost optimization for service delivery
- Differentiated SLAs consumed as a service
- Program network characteristics based on business intent



Building Blocks for Data Center Connectivity

AI-Ready Data Center Interconnect

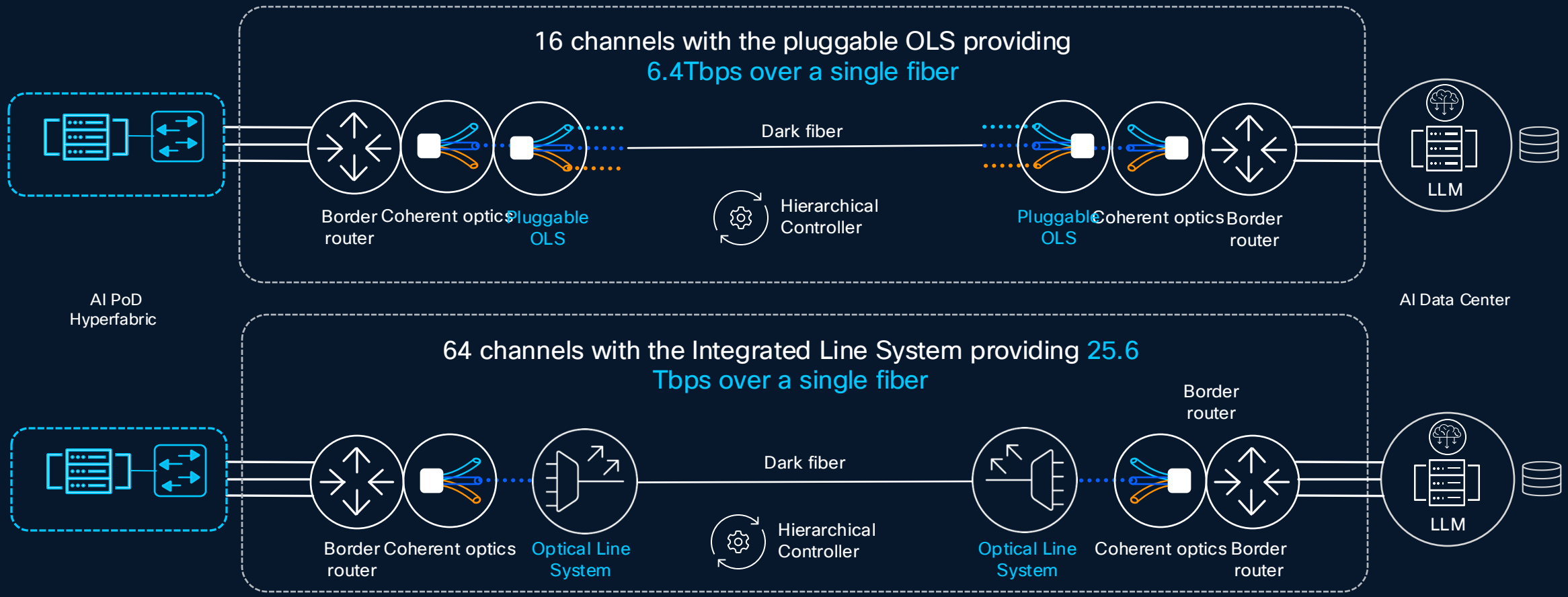
Single Span, Single 400G Wave



NB: Distance depends on fiber quality and network design. Calculations based on 0.25dB loss/km

AI-Ready Data Center Interconnect

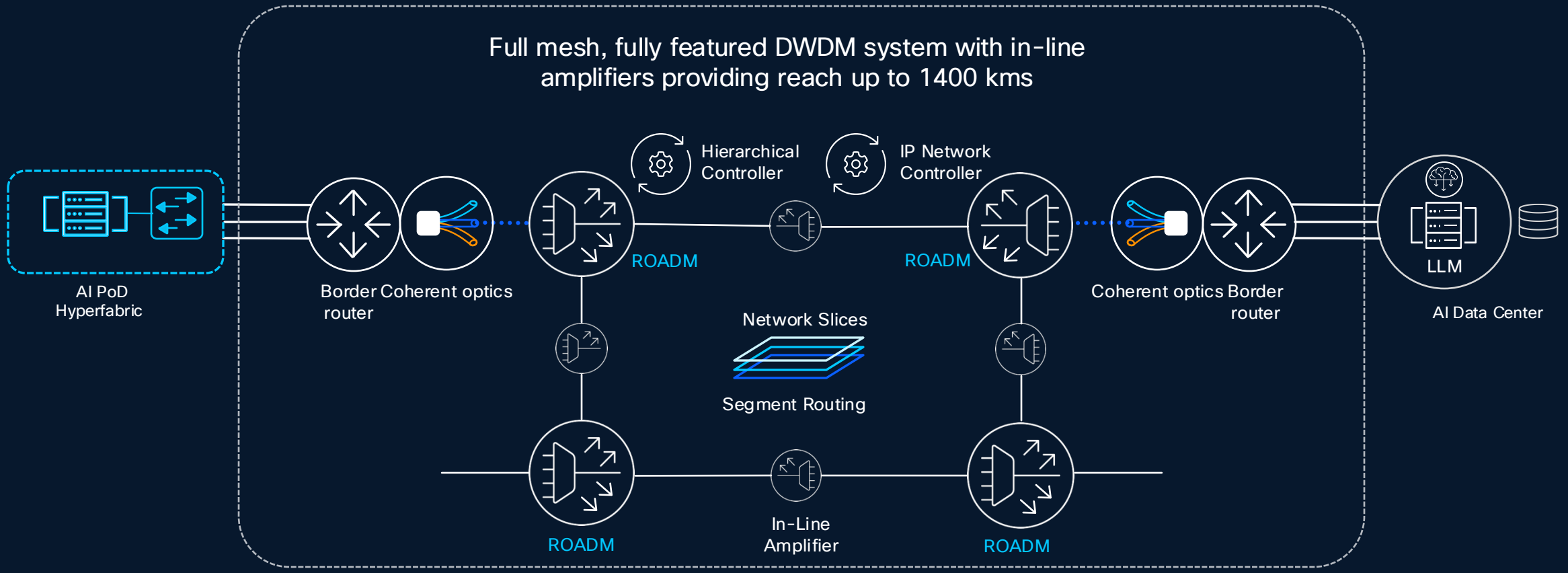
Single Span, Multiple 400G Waves



NB: Distance depends on fiber quality and network design. Calculations based on 0.25dB loss/km

AI-Ready Wide Area Network

Multiple Spans, Multiple 400G Waves



NB: Distance depends on fiber quality and network design. Calculations based on 0.25dB loss/km

One Cisco

Infrastructure
to power AI

Services to accelerate
the value of AI

Security for AI,
AI for security

Data to drive insights
and context

Software to
unlock productivity

Only Cisco brings together **networking, compute, security, and observability** to power AI-ready connectivity

Resources to learn more

Cisco Agile Services Networking
View on Cisco.com

Cisco 8000
View on Cisco.com

Cisco Crosswork Network Automation
View on Cisco.com

Cisco Provider Connectivity Assurance
Visit Cisco.com

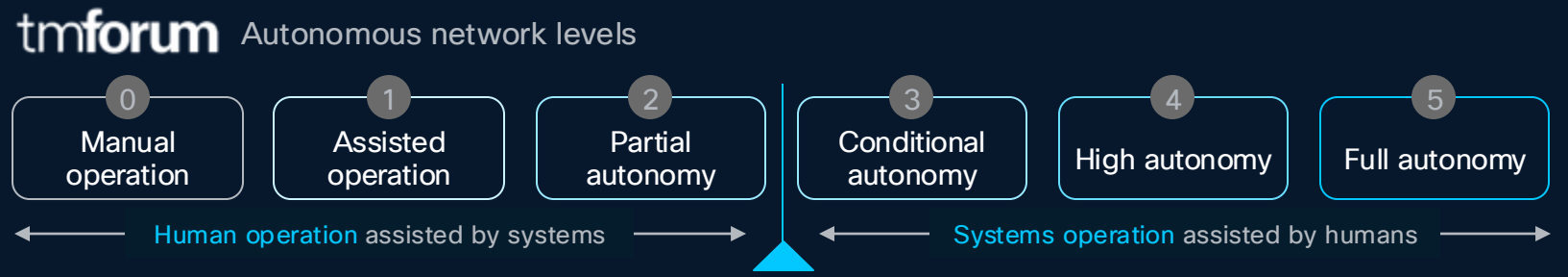
Cisco Routed Optical Networking
Visit Cisco.com

Blogs
Visit blogs.cisco.com/sp360


CISCO

Full Automation & Assurance Deck

Journey and drivers for Autonomous Networking



84%

of companies are currently at initial levels of autonomy

61%

aiming to achieve Level 3 or above by 2028

Network Monetization

Operational efficiency

Resource efficiency and energy savings

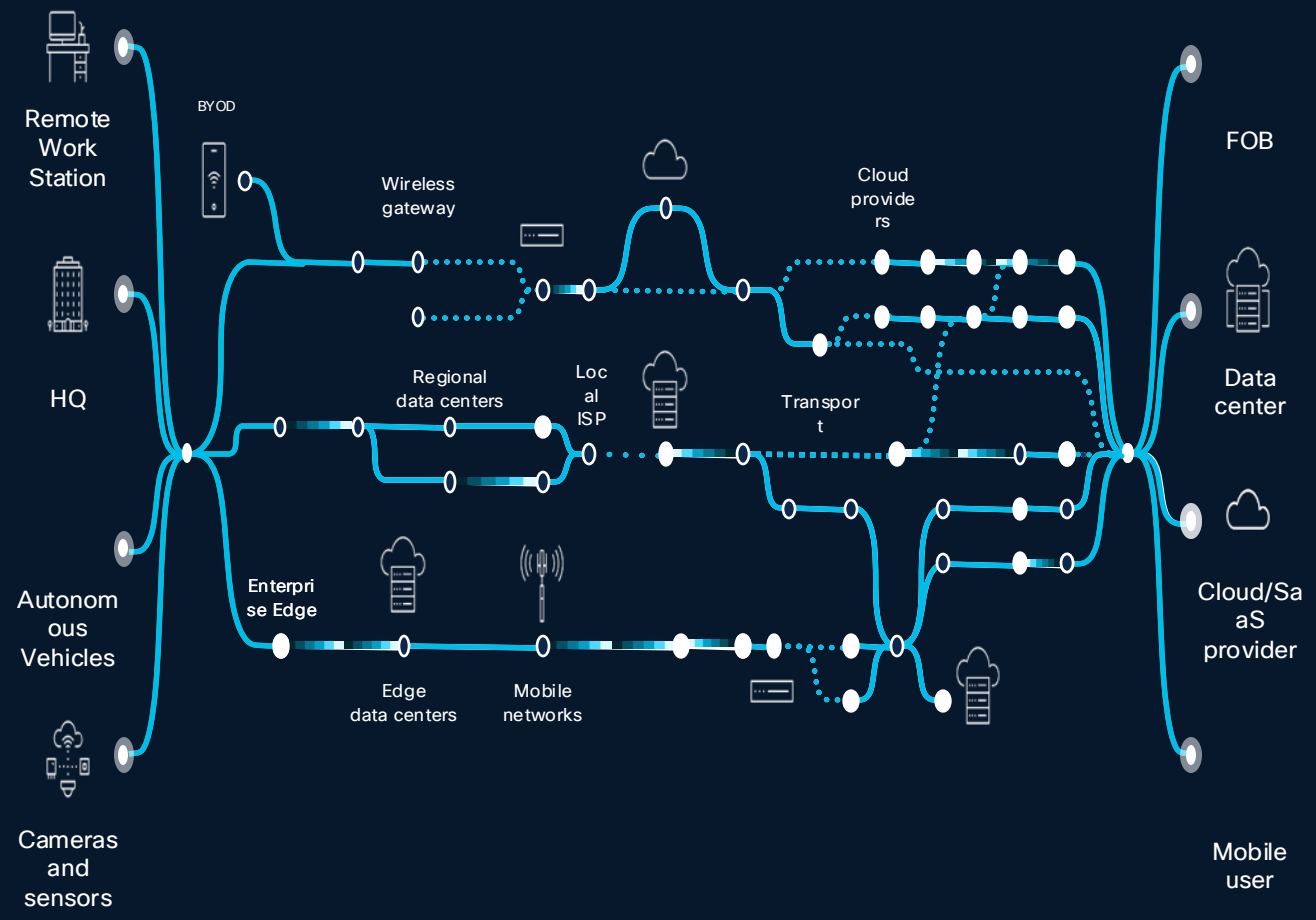
Experience excellence

Source: Cap Gemini "Capgemini Research Institute, Autonomous Telco Networks Survey, December 2023–January 2024

Source: https://inform.tmforum.org/research-and-analysis/reports/autonomous-networks-exploring-the-evolution-from-level-0-to-level-5

Today’s Network

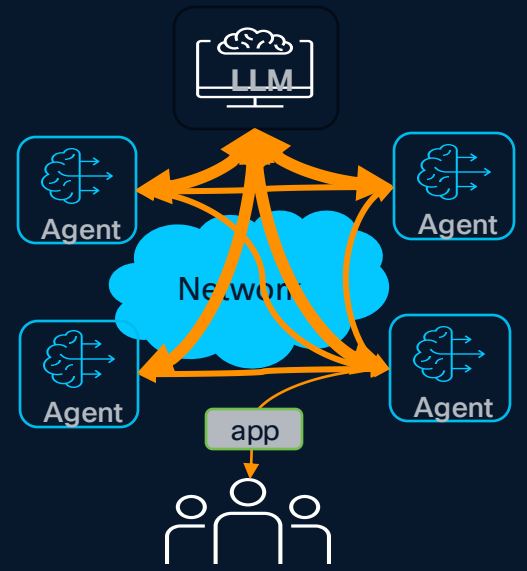
- Hierarchical Design
- Static Configuration and Management
- Reactive Monitoring
- Generalized Quality of Service (QoS)
- Redundancy for Uptime
- Client-Server Traffic Patterns



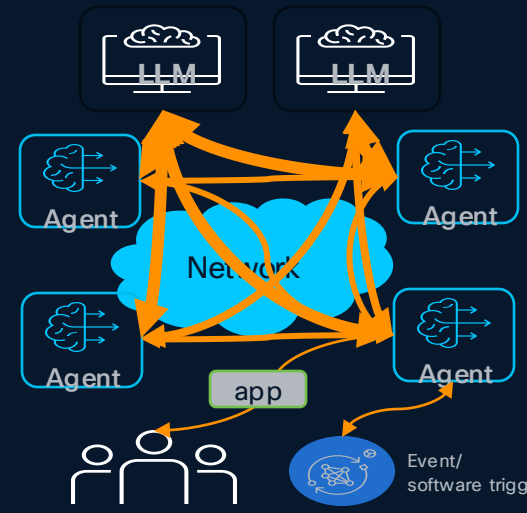
Tomorrow's Needs in the AI era

- **Ultra-High Bandwidth**
  - Training large models require massive data transmission per compute engine
- **Ultra-Low Latency and Jitter**
  - Half the time spent processing AI workloads can occur in the network.
- **Lossless Connectivity**
  - 1% packet loss rate can reduce effective GPU computing time to less than 5%.
- **Dynamic Adaptability and AI-driven Automation (AI-Driven)**
- **Proactive and Predictive Monitoring and Self-Healing**
- **User-Centric granular QoS**
- **Architectural Flexibility for diverse AI workloads**

Multi-Agent



Multi-Agent & Multi-LLM



Automation & Assurance Portfolio

Domain SMEs

NOC / SOC, End Customer

Cisco Crosswork Cloud

- **Trust Insights** — Track integrity of infrastructure
- **Network Insights** — Analyze and identify the source of routing anomalies
- **Traffic Analysis** — Optimize network traffic at peering points

Crosswork Hierarchical Controller
Multi-layer, multi-domain IP and Optical hierarchical controller

- IPv6/SRv6 support
- Northbound notifications
- CONC, EPNM, Nokia, Ciena, Huawei optical adapters

Optical Network Controller
Optical Controller

- Single-VM Optical EMS + Controller
- Optical Assurance & Orchestration

Crosswork Network Controller
Multivendor IP SDN Controller

- Single-VM CNC and KVM support
- Geo-redundancy
- Multi-vendor EMS
- Operations automation via CWM

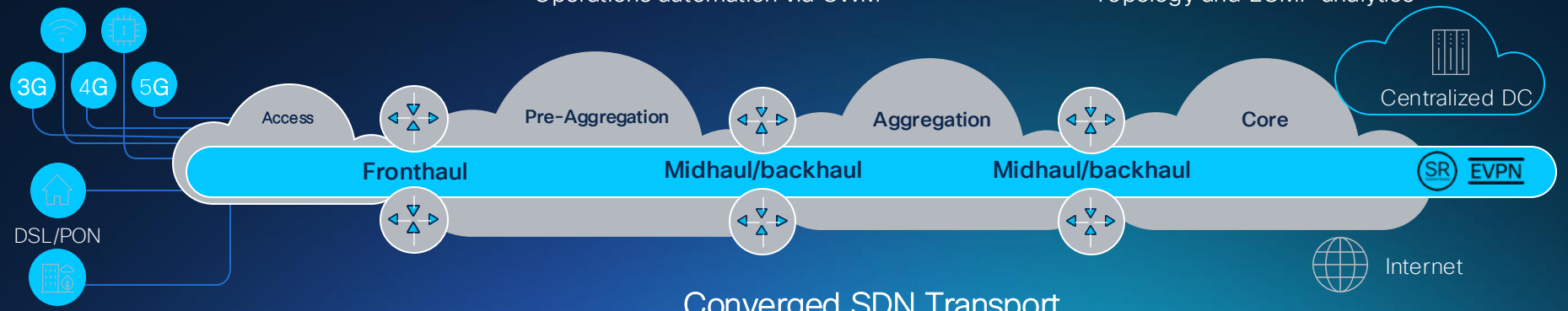
Provider Connectivity Assurance
Multivendor Service Assurance

- Digital Experience Assurance
- Telemetry consumption
- Predicting Performance
- Topology and ECMP analytics

Crosswork Workflow
Low-code automation workflow

NSO
Service Orchestration

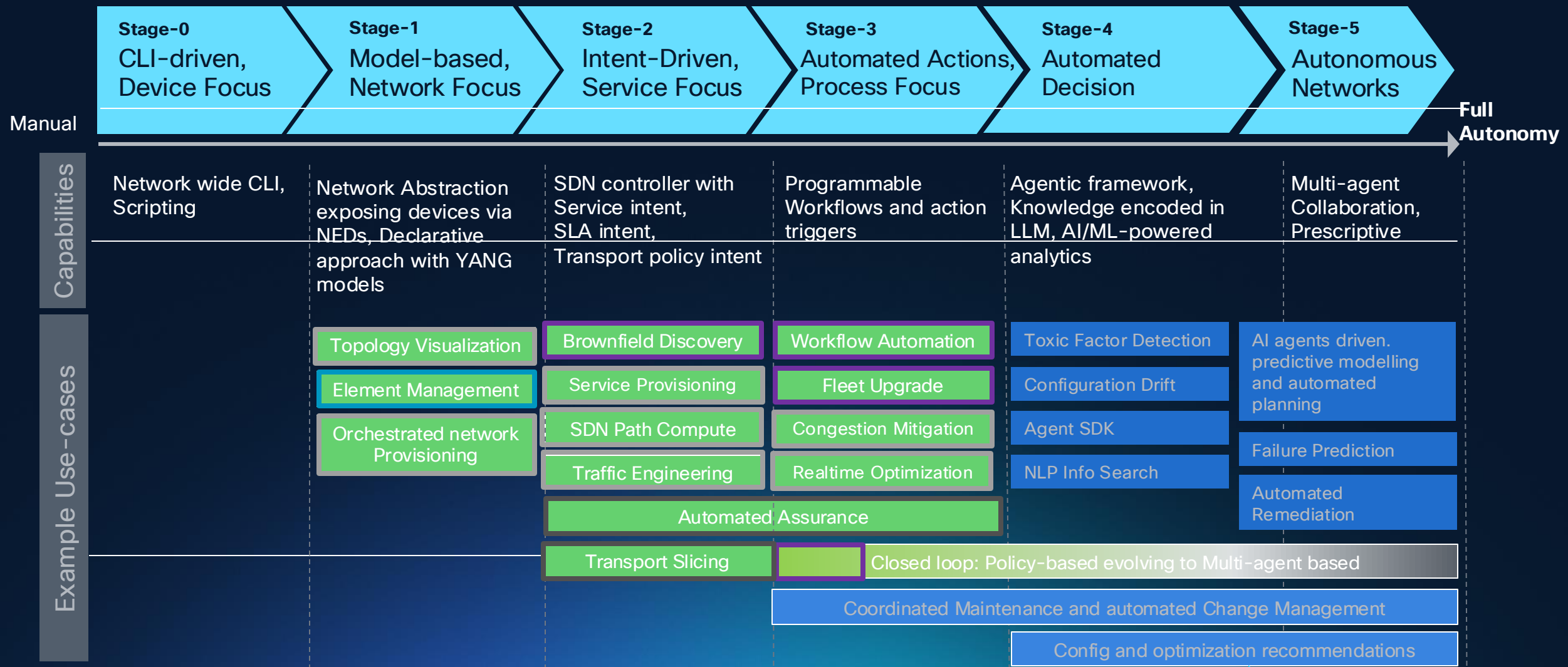
Crosswork Planning
Capacity planning tool



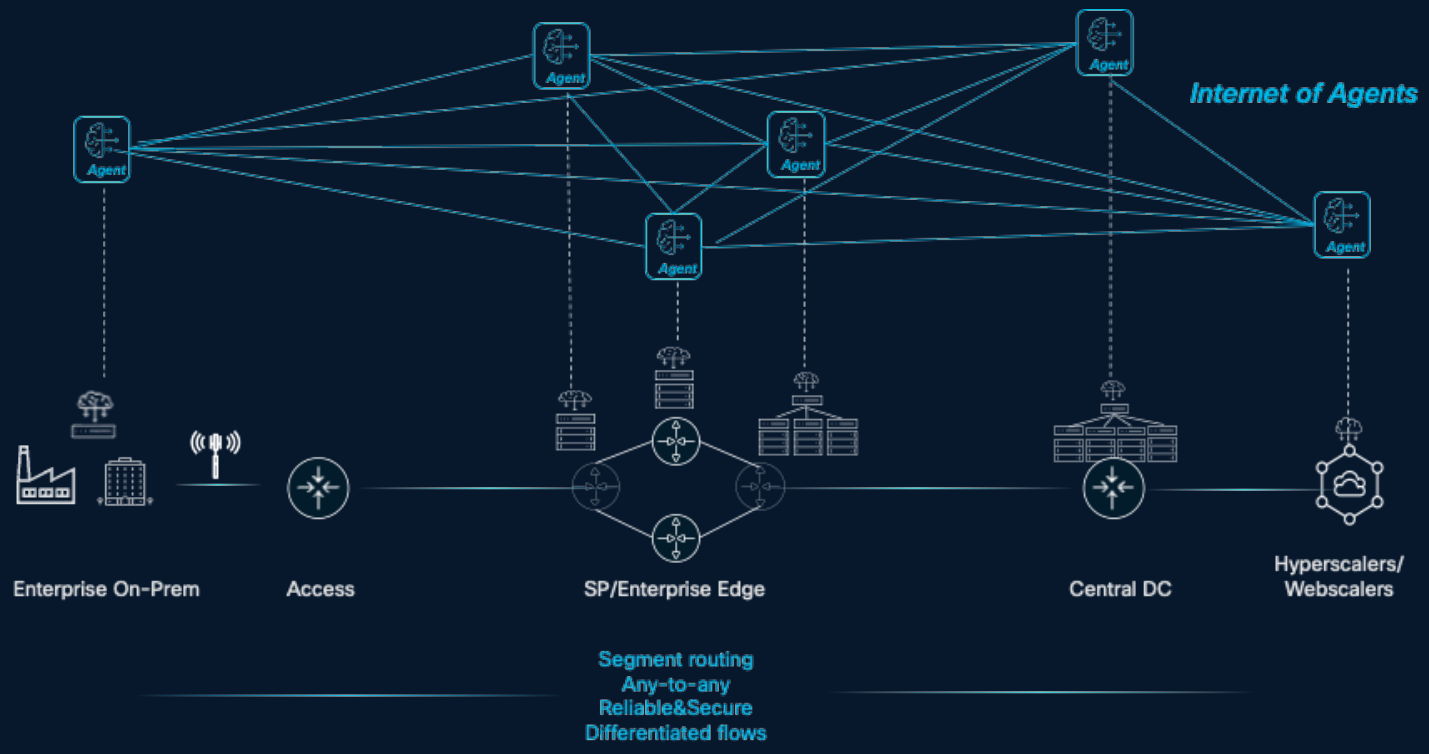
Converged SDN Transport

Delivered | Roadmap

Delivering on the Journey for Autonomous Networking



AI Agents: Implications in Transport Network



Challenge

AI agents routing into network routing

New protocols – MCP, A2A, LLM APIs

Real-time dynamic coordination/interaction

Transport Requirements

Any-to-any connectivity at scale, across all network segments including access

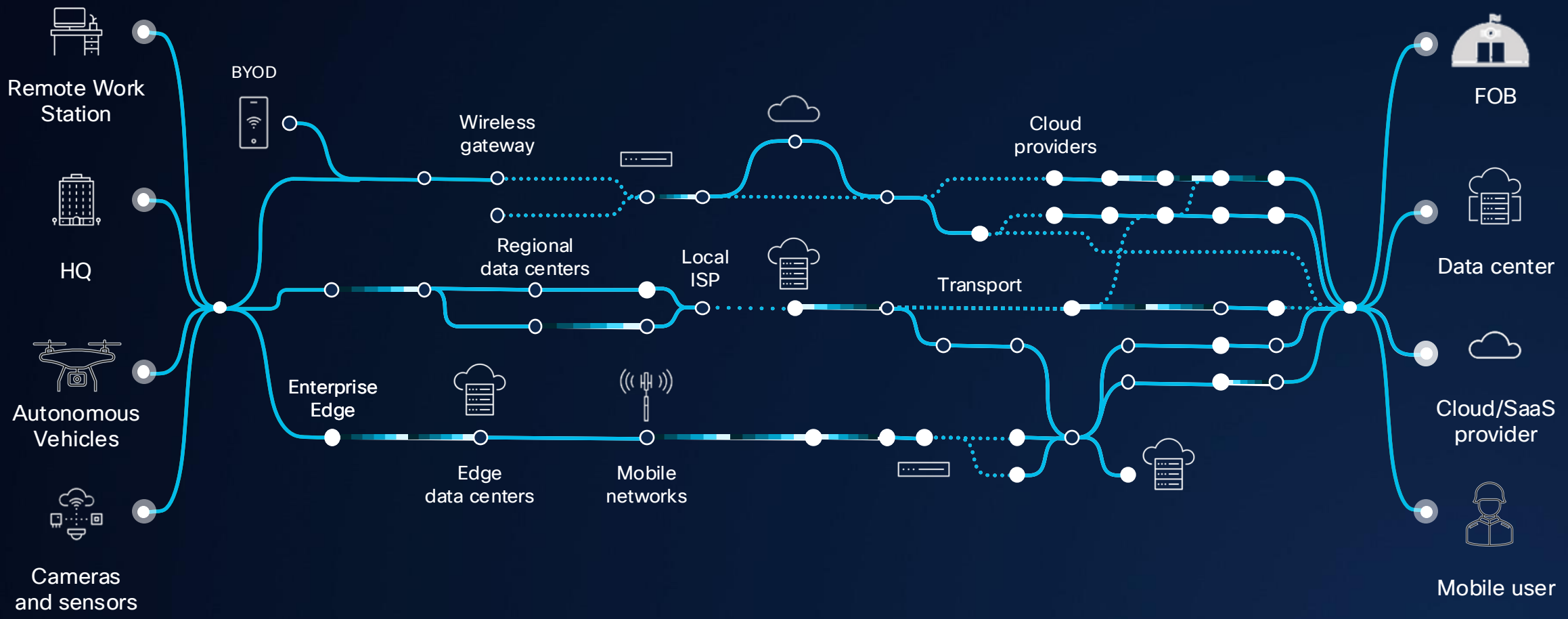
Visibility – comprehend new protocols and behavior

Traffic differentiation and assurance - QoS & path SLA

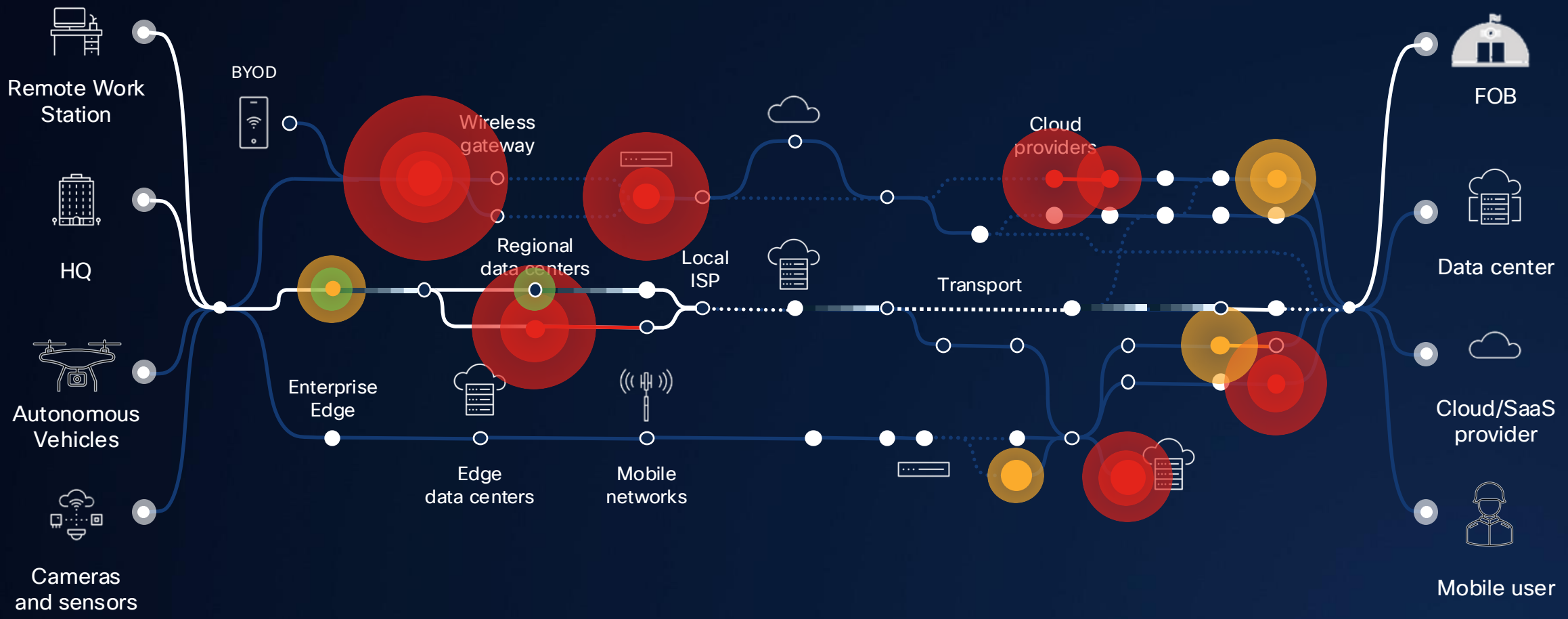
Capacity expansion – more BW, more flows, less latency

High resilience with fast failovers, multi-paths

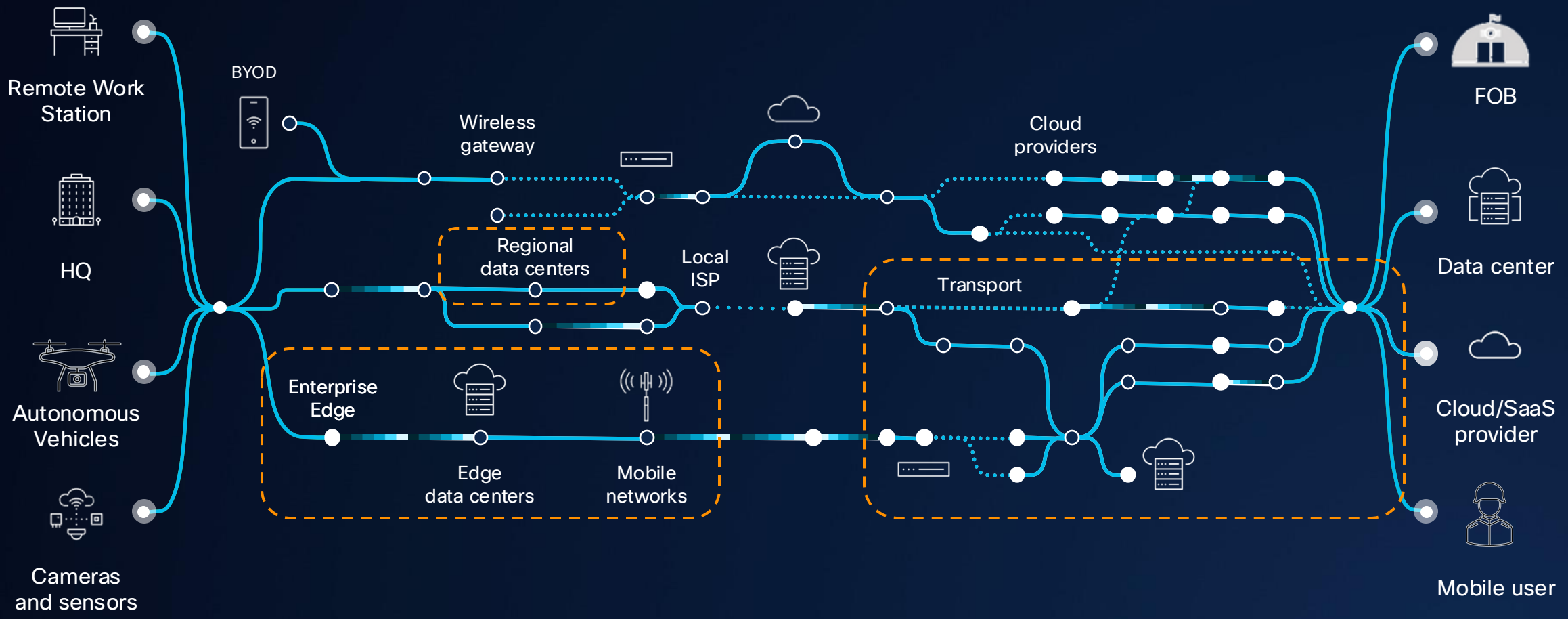
The Global Area Network



Many things could disrupt the global area network

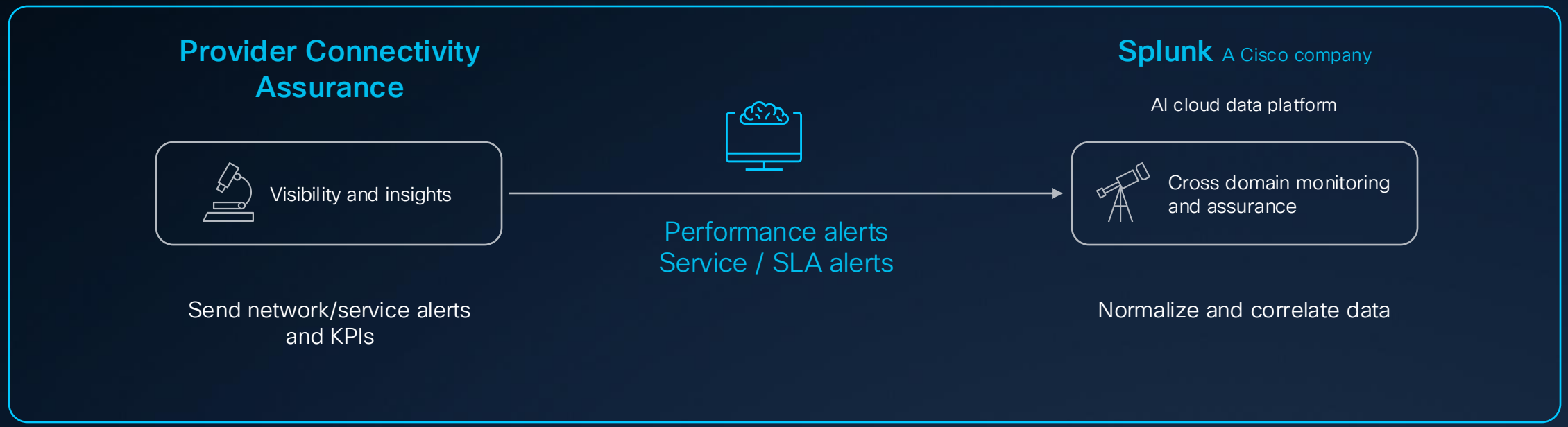


Assure critical networks with Connectivity Assurance



Cisco Provider Connectivity Assurance

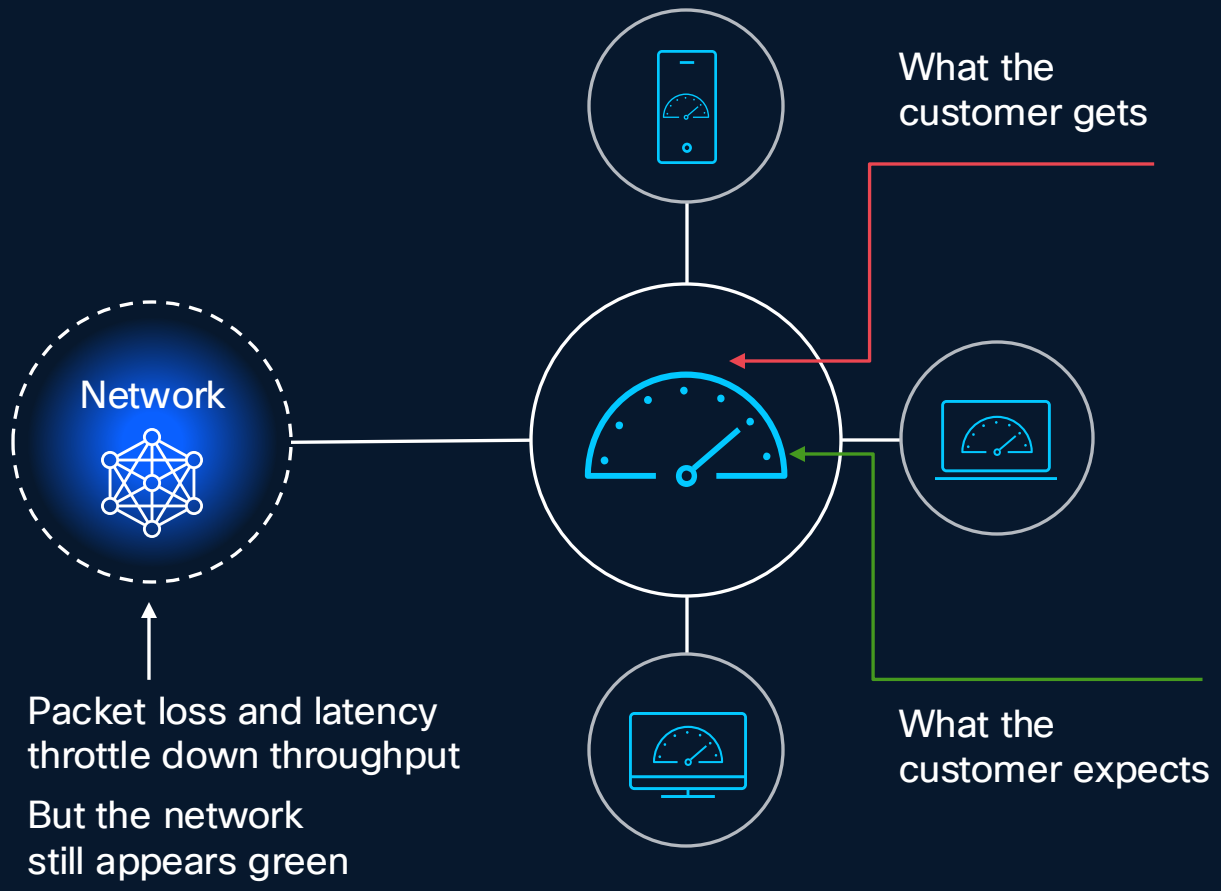
- Simplify operations with a unified view of customer experience
- Identify, prioritize and predict customer-impacting events
- Drive automated decisions to accelerate resolution



Most tools are ineffective at detecting today's QoE issues

Without deep visibility into micro events:

- 0.53% packet loss leads to a 50% decrease in data throughput
- 5 ms delay leads to a 10% decrease in data throughput
- 10 ms jitter leads to a 10% decrease in data throughput



The quality of the network has a major influence on user quality of experience (QoE)

Provider Connectivity Assurance solution overview

High-level view

Open automation and APIs

Open data and alerts

Cloud-native platform

Management, automation, data collection, and data analytics

User personas and use cases

- End customers
- Service managers
- Engineering
- Operations
- Capacity planning
- IT and network

Software sensor layer

Containerized and horizontally scalable

Hardware sensor layer

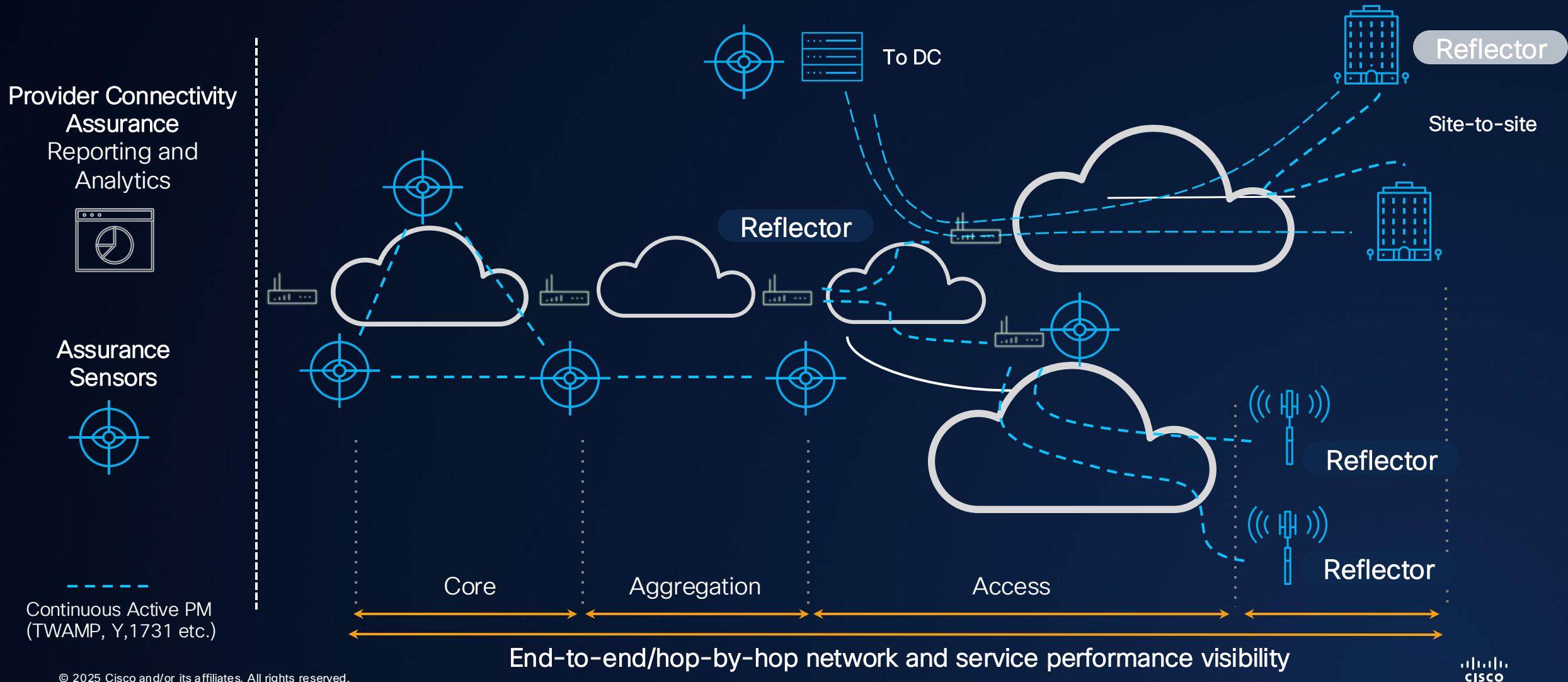
With embedded PM

Metrics from other sources(e.g. FEC, Light Level, ...)

Correlated with assurance data

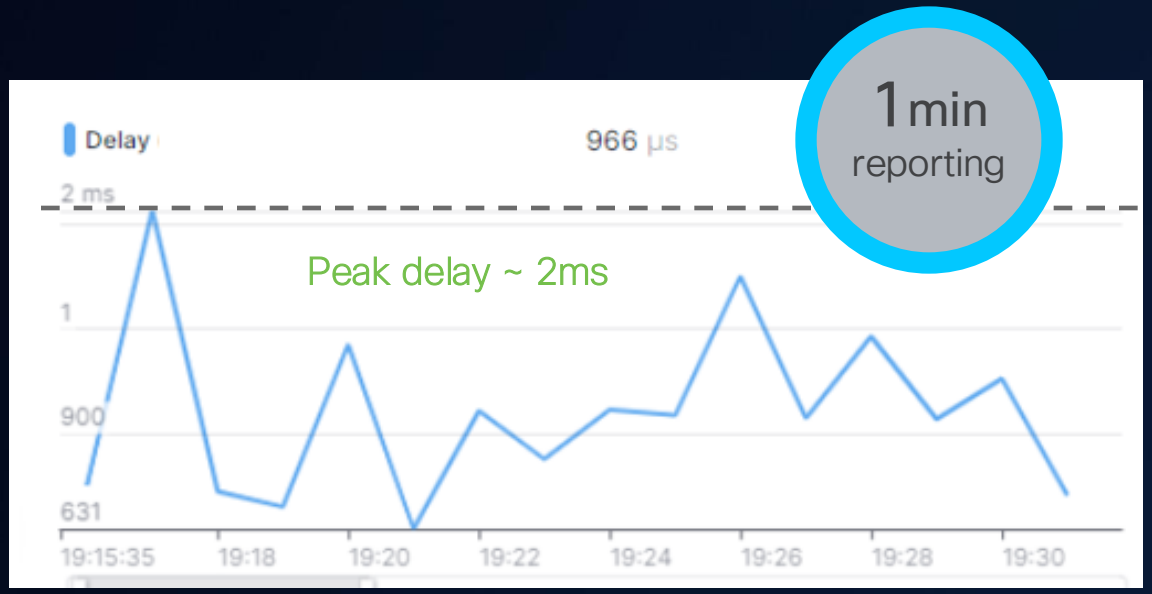
Network-wide service performance visibility

- Deploy in any physical or virtual infrastructure
- Continuous active/synthetic monitoring
- Passive real time per packet analysis



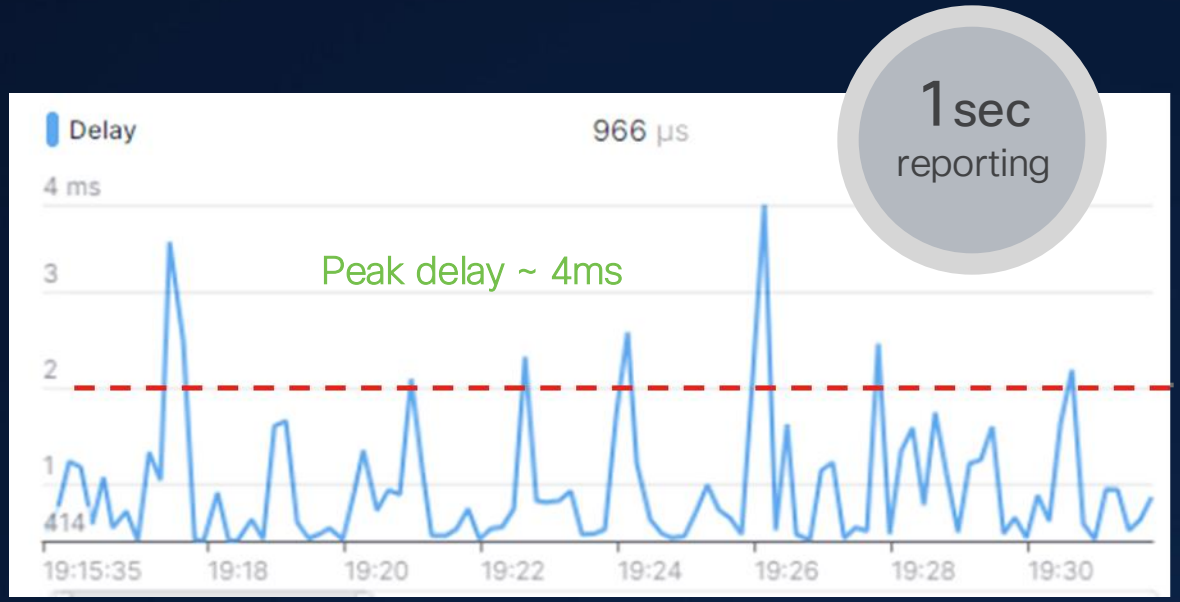
PCA observability

The power of high-resolution monitoring



Perception ✗

Latency is below threshold.
Service performing according to SLA/SLO.



Reality ✓

Multiple latency threshold crossing alerts.
Service affected by latency issues.

cisco

Enriched service insights from extended KPIs

Over 50 relevant, actionable, near real-time metrics for SLAs

Active/synthetic PM

One-way delay, PDV, and IPDV (jitter)
- Min/max/average
- Median (p50)
- Percentile 25/75/95/96/98/99
- Standard deviation

One-way packet statistics
- Packets lost (number and %)
- Loss bursts
- Longest loss burst
- Shortest loss burst
- Reordered packets (number and %)
- Packets duplicated (number and %)

One-way packet field and QoS metrics
- IP TOS max (DSCP diffserv)
- IP TOS min
- TTL max/min
- VLAN Pbit max/min
- ETH-OAM MEG level max/min
- MOS
- R-value

Meta metrics
- Session ID
- Interval sequence number
- Interval timestamp (UTC)
- Interval length (Report interval)
- Up or downlink direction

Bandwidth metering

Throughput metrics
(in-line or out-of-line mode)
- Min Throughput – **Per Flow**
- Average Throughput – **Per Flow**
- Max Throughput – **Per Flow**

Service activation testing

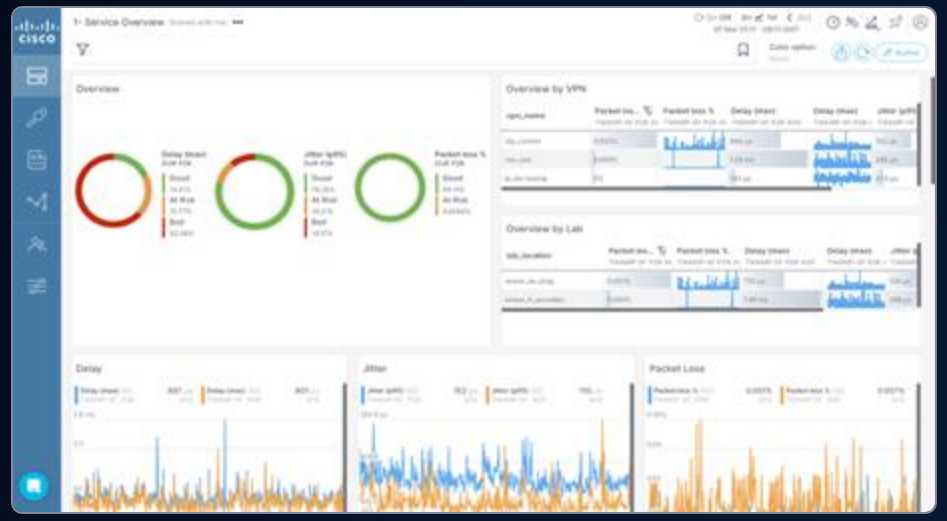
Throughput validation – circuit readiness
- RFC2544 generation and reflection
- Y.1564 generation and reflection

Metrics from other sources

Assurance reporting and analytics

From the macro view to micro details

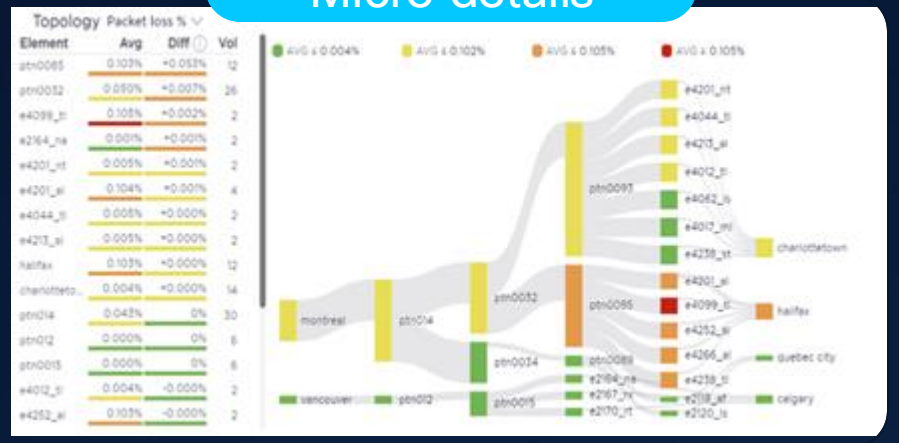
Macro view

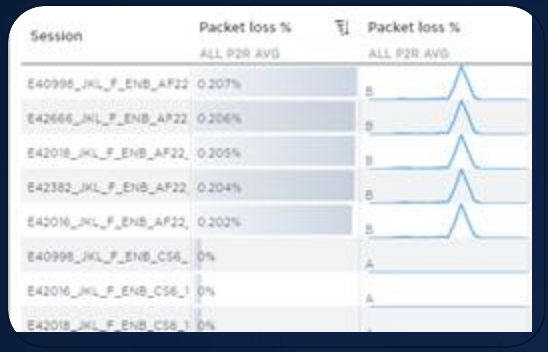


Analysis

Drill-down and clustering

Micro details





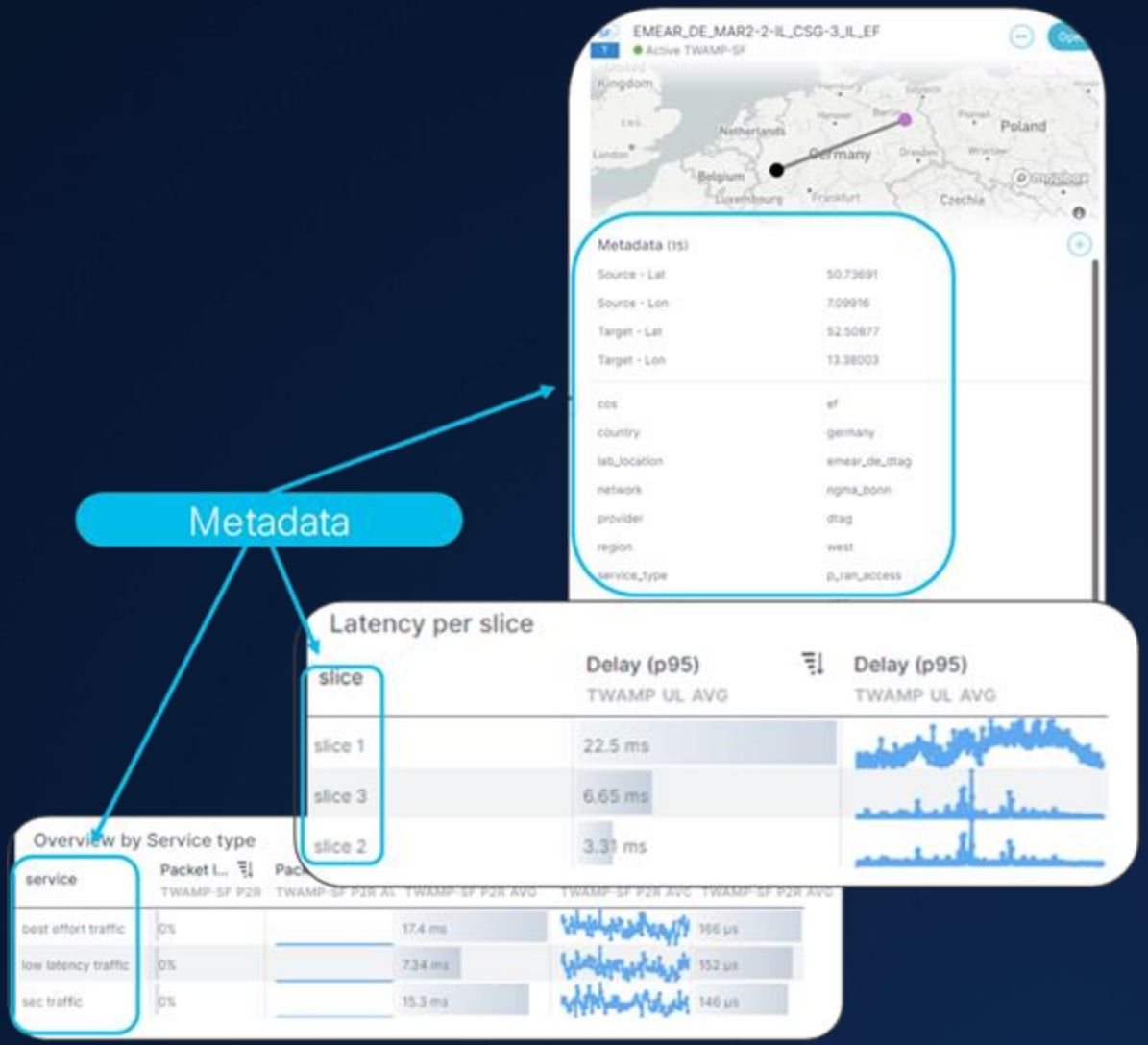
- Analytics and correlation capabilities
- Metadata enrichment makes data meaningful
- Quickly visualize network and pinpoint problems
- Increase efficacy of operations teams/

cisco

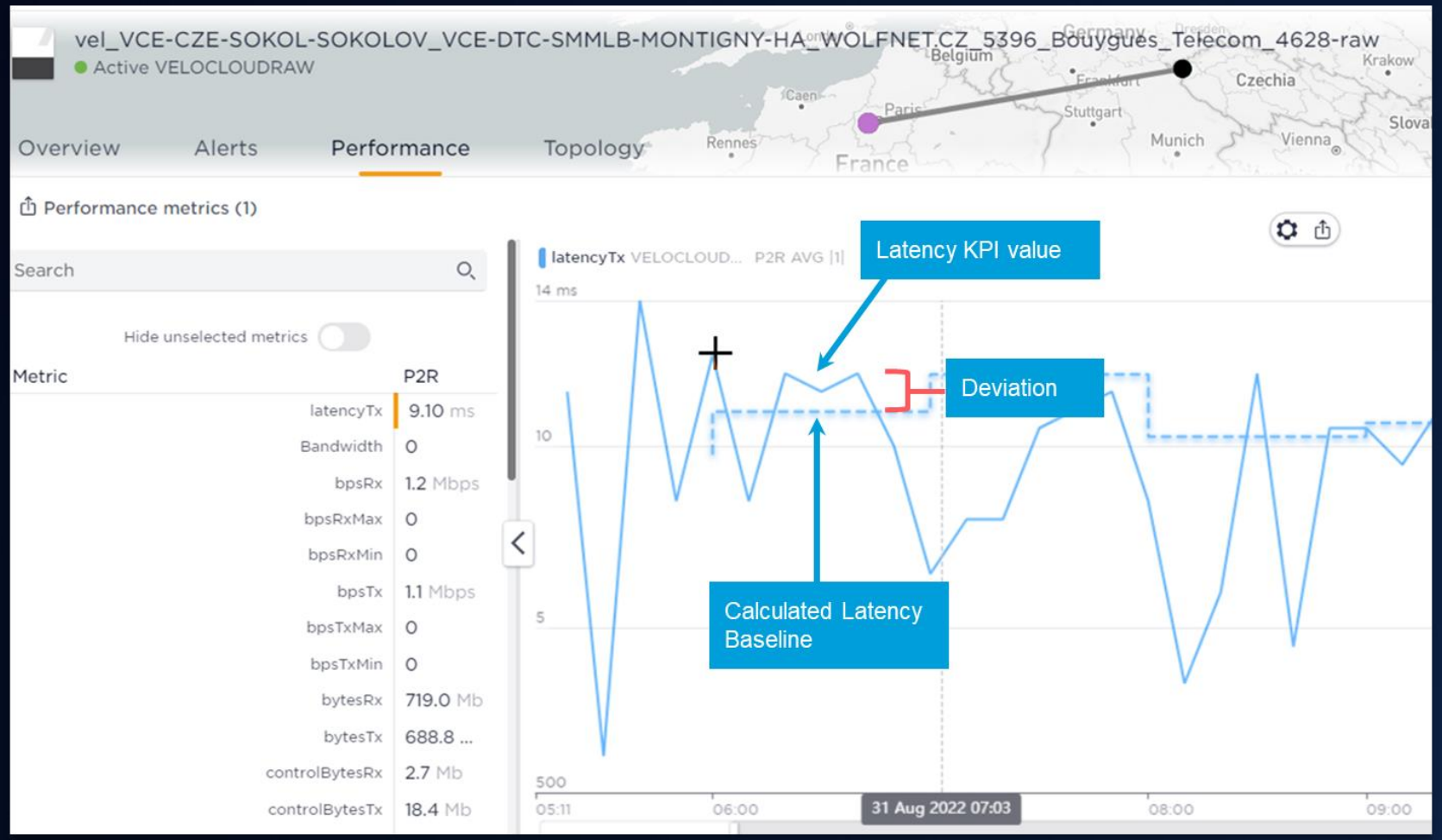
Assurance reporting and analytics

Leveraging metadata

- All Cisco assurance and non-Cisco data can be enriched with metadata
- Metadata is user-defined contextual information
  - Examples: site, region, class of service, geo coordinates, topology, etc.
- How it is used:
  - Filtering, grouping, aggregating relevant data
  - Correlate and find commonalities for root cause analysis and troubleshooting
  - Adds flexibility to dashboards to suit multiple usage scenarios



Move From Problem Reporting to Automating the Experience



Machine Learning

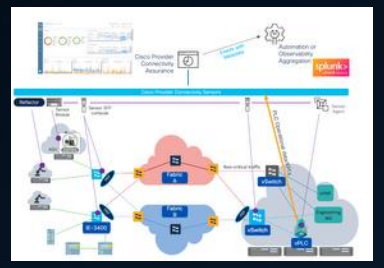
- Time of Day, Day of week Baselining
- Ability to measure deviation from normal service performance baseline

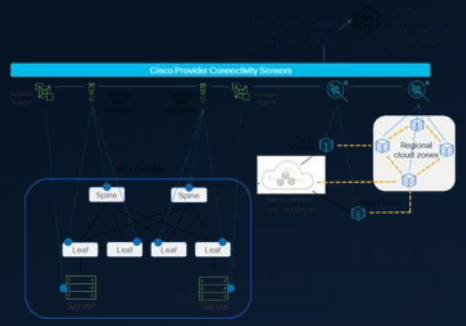
Alerting

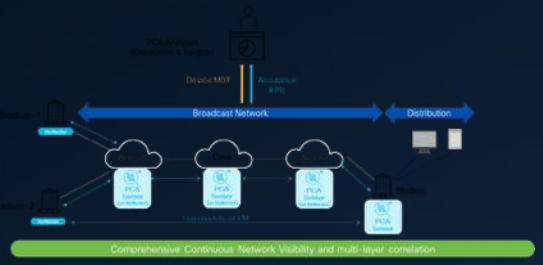
- Remove eyes from the glass
- Proactive monitoring
- Ties into AT&T systems and work flows

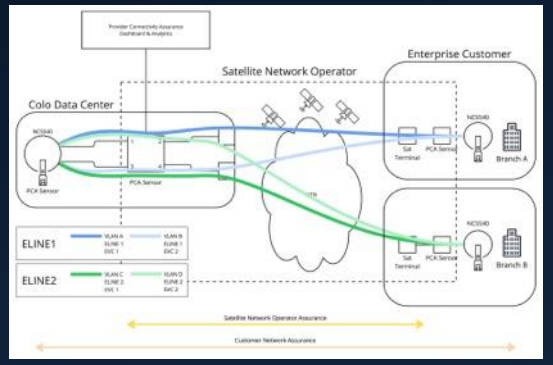
Assurance Use Cases

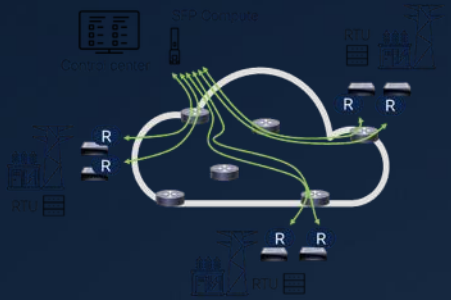
Use Case Name
Smart connected factories
Public Sector (SLED)
Media / Live broadcasting infra
Utilities (IoT)
Critical Network SLA management
Critical Network / B2B Customer portals
Service differentiation and new revenue enablement











Use Case Name
Hyperscaler / Webscaler Network Assurance
Data center network connectivity and service assurance
Unified visibility of multi-layer performance
Per-service Carbon Intensity and Power as Traffic Engineering inputs
How the solution can help achieve sustainability goals
Assurance of non-terrestrial networks (LEO)

1- AI LLM User Experience

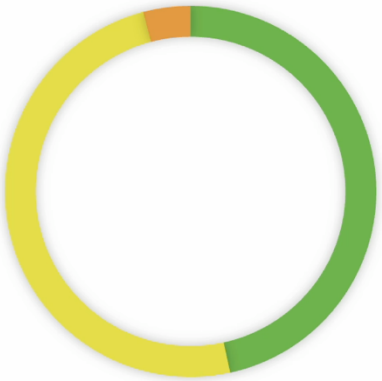
Shared By (Abhishek Jamwal)

Color option
Metric

Author

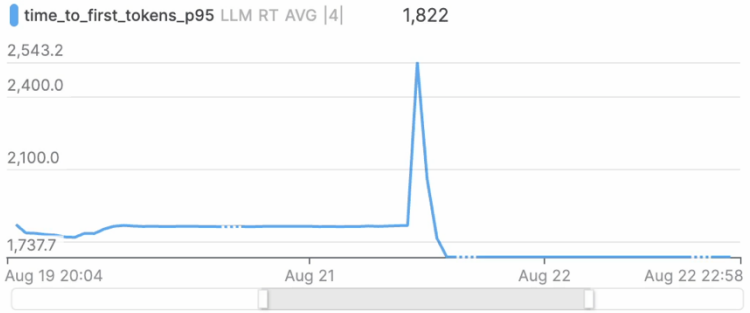
LLM Experience - Public or Enterprise LLM

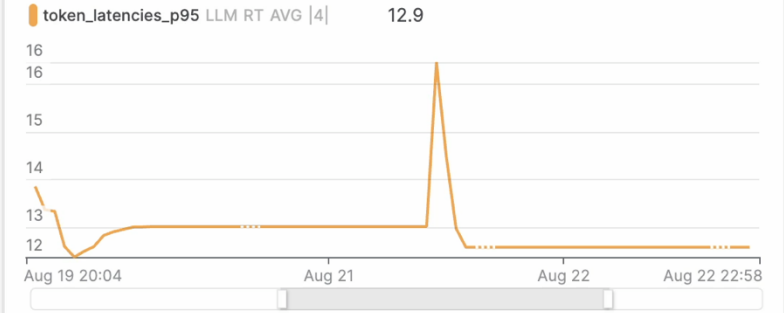
Failure Percentage

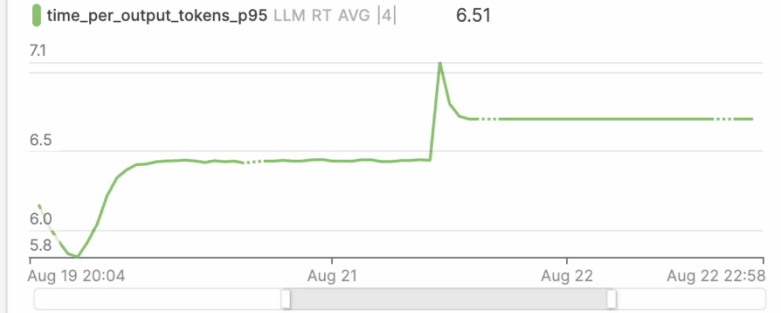


LLM_Name	trafficClass	failure_percentage LLM RT AVG	total_input_tokens LLM RT AVG	total_output_tokens LLM RT AVG	total_responses LLM RT AVG
gpt-4o	best effort	0.466	584,087	209,165	735
gpt-4o	priority_ef	0.397	600,648	215,967	756

LLM_Name	trafficClass	time_to_first_tokens_p95 LLM RT AVG	time_per_output_tokens_p95 LLM RT AVG	token_latencies_p95 LLM RT AVG	token_latencies_p99 LLM RT AVG
gpt-4o	priority_ef	1,869	6.81	13.3	23.9
gpt-4o	best effort	1,764	6.37	12.5	28.1







What do you need to run your network

IT, Carrier OSS/BSS

Ops Provisioning

Networking DevOps

CI/CD/Testing

Planning (what if)

Workflow for data enrichment, human process interaction etc.
CFS-RFS Interface

Resource Facing Network Orchestration

VNF/CNF Life-Cycle Orchestration

PCE/Controller / Planning

Device Management

Agentic Framework (*)

AIOps tools

Active Performance Testing

KPI Actions

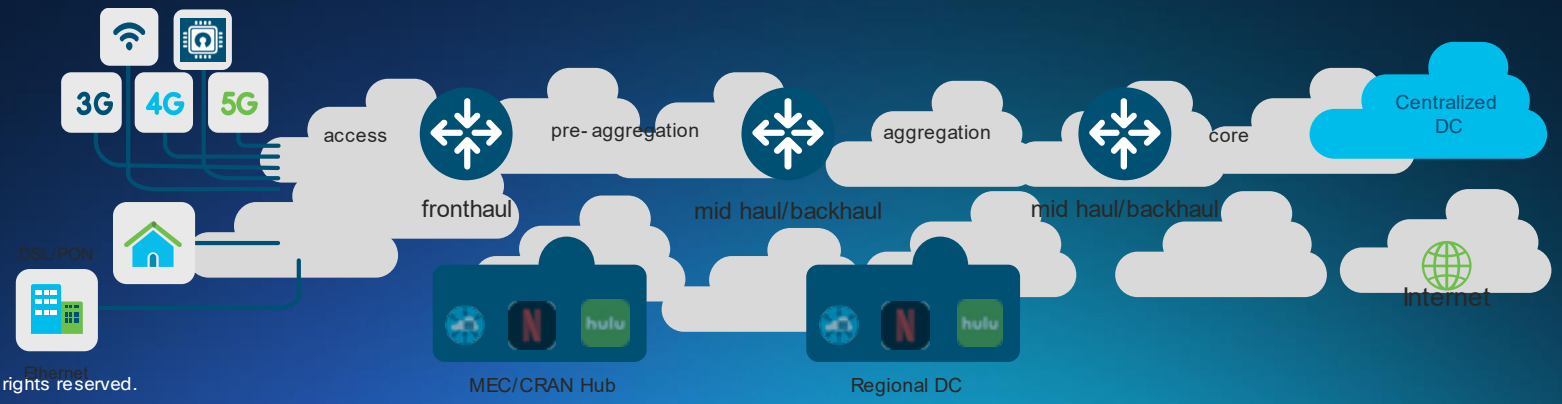
KPI Monitoring

Fault Mgmt

Service Health Analysis

Common Collection Platform

Data Store / Inventory



cisco

Cisco is executing in these objectives

IT, Carrier OSS/BSS

Ops Provisioning

Networking DevOps

CXTM/CXTA

Crosswork Planning (Cariden)

Cisco Crosswork Workflow Manager (CWM)

Cisco Network Orchestration (tail-f)

Cisco NFVO/CNFO

Cisco CNC / CHCO (Sedona)

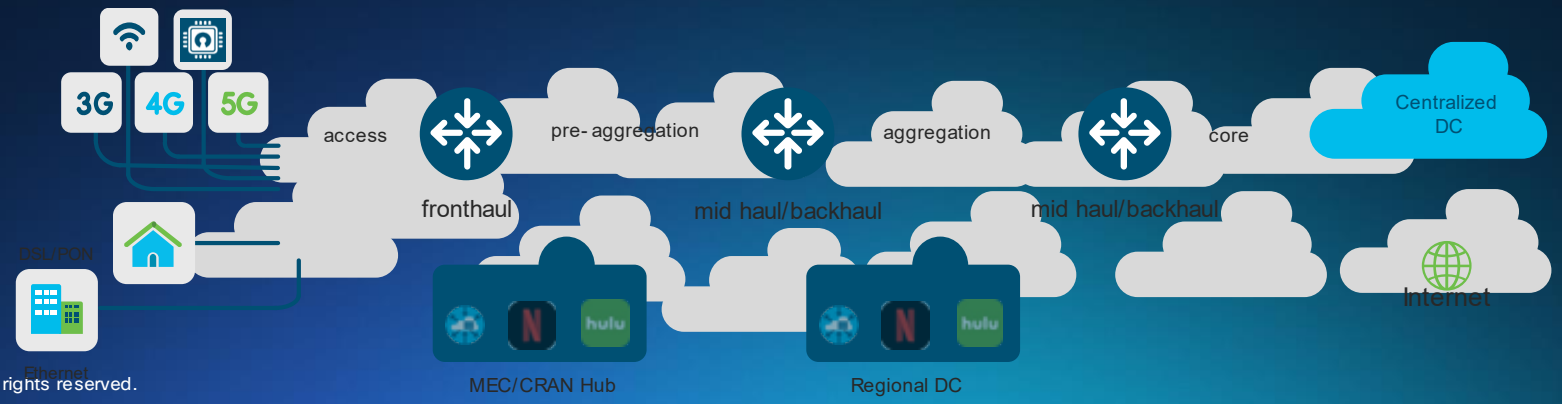
Agentic Framework (Foresight)

Splunk AIOPs

Provider Connectivity Assurance

Cisco CNC / CHCO

Data Store / Inventory



Extra

Enterprises want service providers' expertise to get ready for AI

51%

of enterprises see SPs as a potential strategic partner for AI adoption

62%

of enterprises rate stronger data security and governance as the key AI-driven change in WAN network requirements

57%

of enterprises see higher reliability and uptime as major WAN requirements for AI

52%

of enterprises see AI-optimized data transport and routing as crucial capabilities for SPs to become key players in the AI value chain

STL Partners – Cisco: Enterprise AI Survey, preliminary results, Aug 2025

cisco

AI Traffic Is Rising – The Real Wave Is Still Ahead

Agentic workflows will power the next wave of demand

LAYER	DRIVER	IMPACT
Adoption X Capability X Interaction	Commoditization	↑ Lower cost & demand growth
	Better LLMs	↑ Addressable tasks
	Multi-modal	↑ Payload size
	Reasoning	↑ Tokens & Compute
	AI Agents	↑ Transactions per request
	Multi-agent	↑ Requests per session

1448%
Increase in tokens processed by AI models in the last 12 months[1]

75%+
Of inference-driven data creation and processing at the edge by 2030[2]

10x
Bandwidth required upstream compared to downstream

36x
Increase in AI traffic as early as 2023-2024[1]

1: Openrouter.ai/rankings 2: KPMG

cisco

Cisco Agile Services Networking for AI-Interconnect

Simple

Reduce cost and increase efficiency with a converged infrastructure for flexible service delivery

Resilient

Assure experiences and mitigate risk with AI-powered automation, observability and security

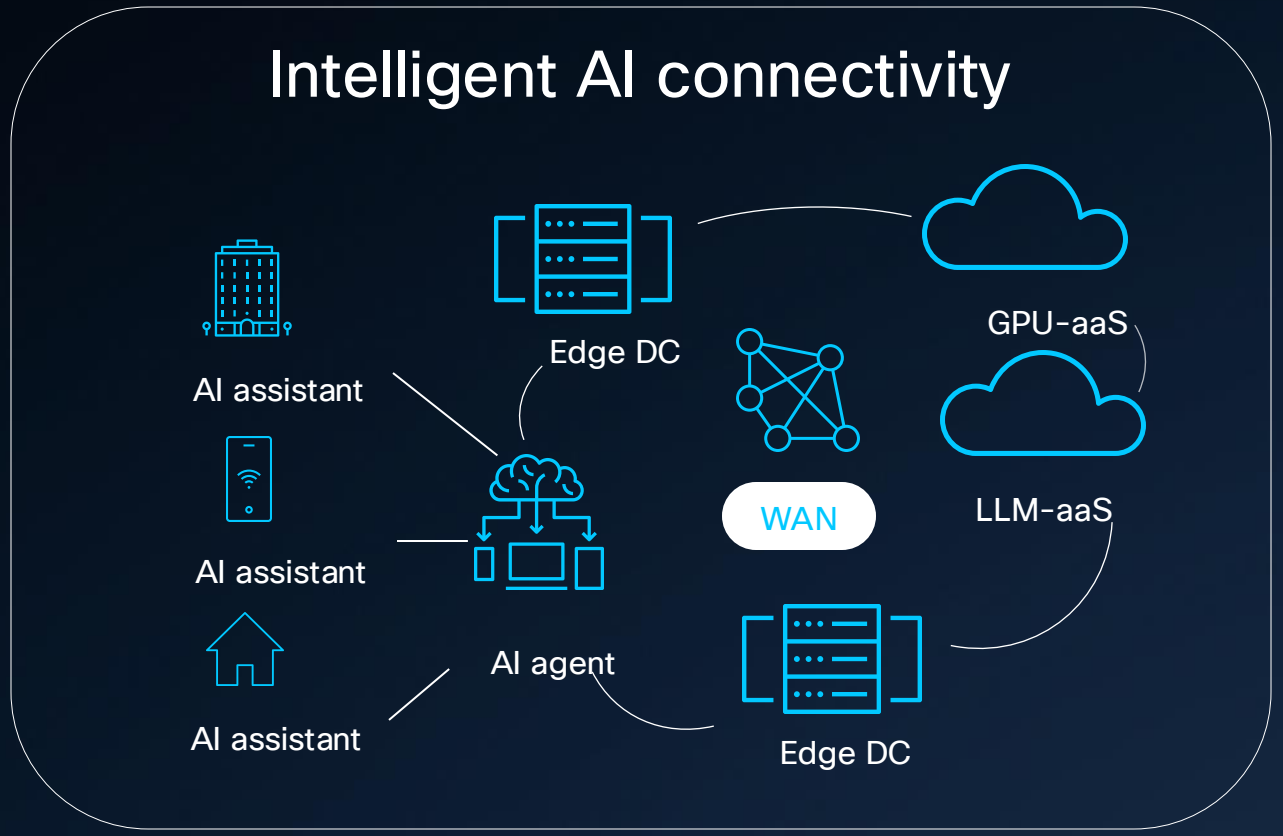
Intelligent

Deliver network and services capabilities from an architecture optimized for intelligent services

Cisco innovations

Unified silicon architecture, optics, software, automation, security, data / intelligence platform

AI-Ready Wide Area Network

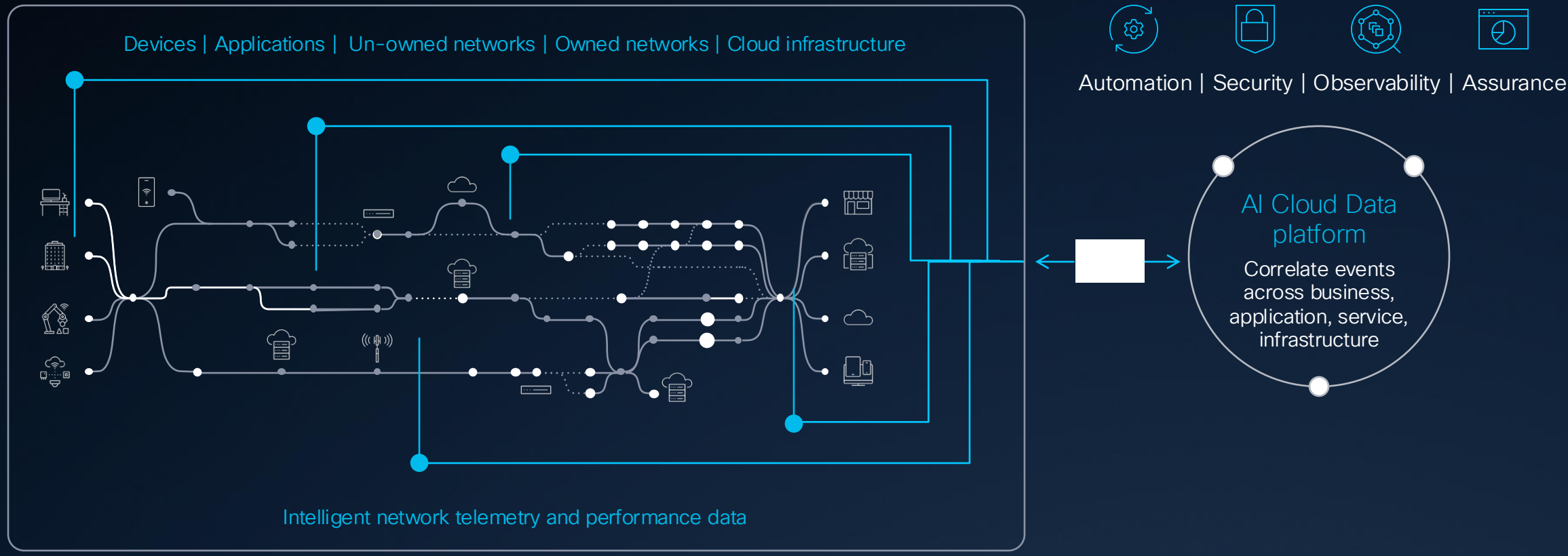


Segment Routing
- Fewer protocols
- Simpler network to deploy, operate and manage

Embedded Security
- Trusted and protected network
- Quantum keys distribution

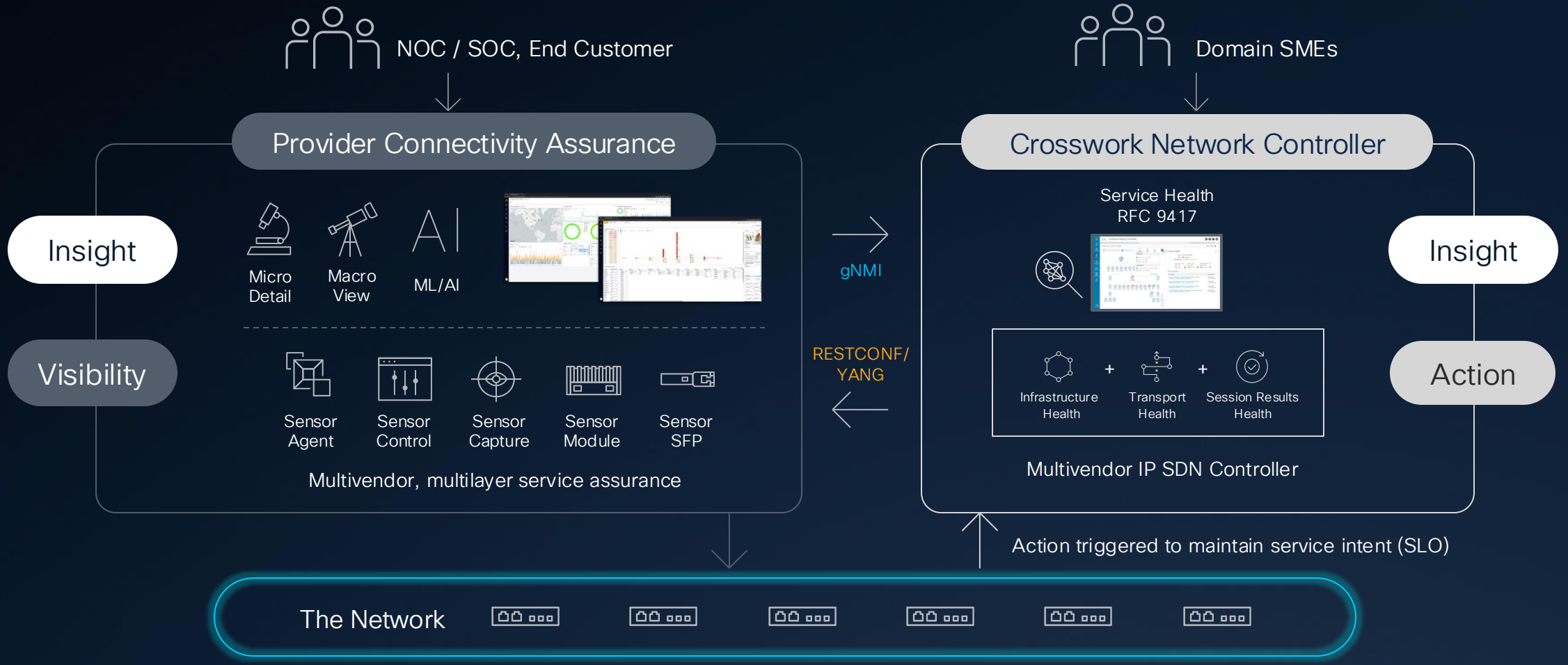
Routed Optical Networking
- Fewer layers in the network
- Reduced footprint

Assure experiences with resilient networking and services



AI enabled Automation and Assurance

Proactive operations from insights to action



Cisco 8000 – 800G/800GbE roadmap

P100

3rd Gen (shipping)

- Supports 800G port with 2x 400GE or 8x 100GE break-out for dense 400GE / 100GE
- Not compatible with ETC or IEEE standards

P200

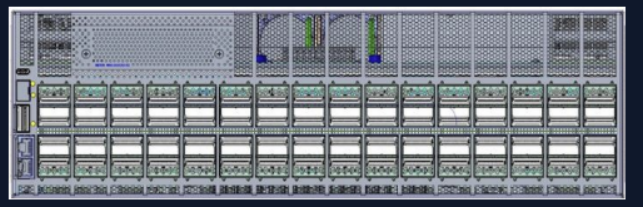
Q4CY26 – Q1CY27

- Support 800GE MAC
- Compatible with 800GE IEEE 802.df standard and ETC standards

SLUGGER (8212-48FH-M)

2RU, 19.2T;
24x QDD-800, 24x QDD

MUSTANG (8223-64EF-M)



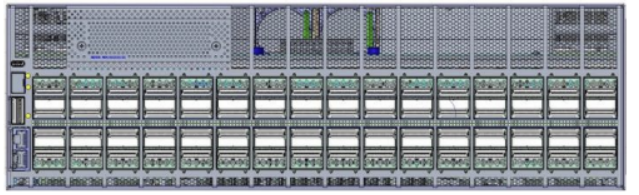
3RU - 51.2T; 64x OSFP-800

SATURN (8711-32FH-M)

1RU, 12.8T;
16x Q800 + 16x QDD

Pending EC

TITAN (8223-64E-M)



3RU - 51.2T; 64x QSFP-DD800