

Connecting Data Centers and Cloud in the AI Era



James Massett

Director, Americas Internet
& Mass-scale Infrastructure



Rick Gomez

I&MI AE
ASP DC Colo



Zac Campbell

Leader
Automation Assurance



Abhishek Jamwal

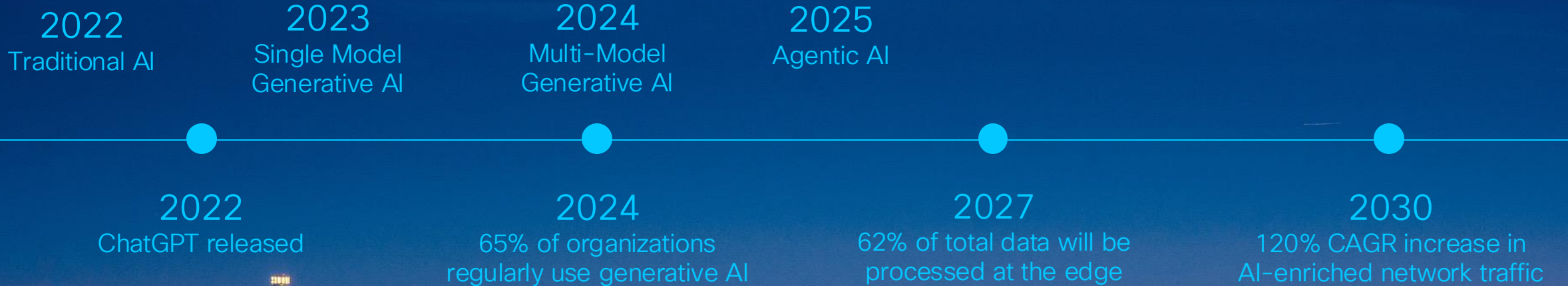
Technical Leader
Automation Assurance



Agenda

1. Market evolution in the AI Era
2. Cisco strategy
3. Cisco innovations
4. Connecting data centers
5. Automation and assurance
6. One Cisco
7. Q&A and next steps

AI is re-shaping network connectivity



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

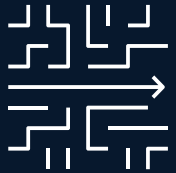
— Cisco Report:
Race to an Agentic Future

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

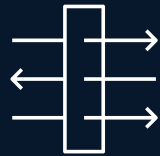
— Gartner

The Fundamentals of Connectivity

The *objectives* have not changed...



Simplified and
Autonomous Operations



Increased Service
Resiliency



Lowest TCO
and Carbon Footprint



New Monetization
Opportunities



Unparalleled
Customer Experience

... but the architectural imperatives and eco-system have

The Rise of AI

New Players in the
Market

Highly
Sophisticated Bad
Actors

AI/GenAI use cases across industries



Knowledgebase copilots

AI assistants



Content & code generation

Text | Images | Video | Code



Reporting & data analytics

Summarize texts |
Generate visualization



Language translation

Multilingual real-time
communication



Virtual agent & chatbots

Specialized domain
specific chatbots



Detection & prediction

Forecasts |
Anomalies | Insights

What Makes AI Traffic Different



Every AI request is unique – no content is cached



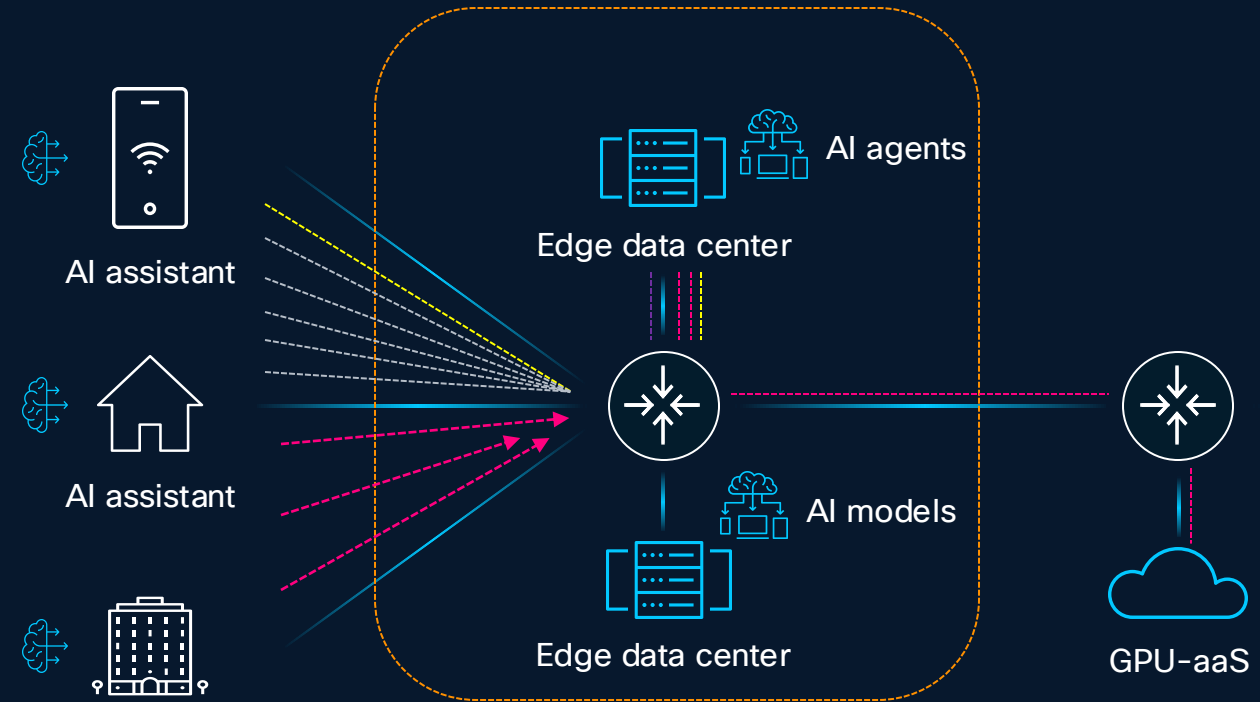
AI requests generate a high volume of tokens



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



AI drives a disproportional increase in upstream traffic

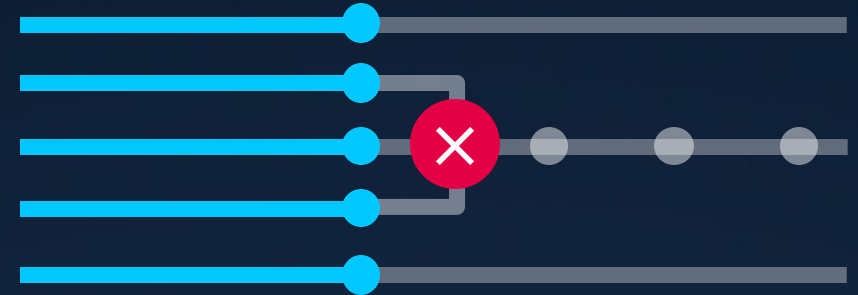


More traffic, Upstream tilt. Resiliency is everything.

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



Downtime no longer leaves users waiting – it brings mission-critical digital workers to a halt



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

AI Traffic Is Rising – The Real Wave Is Still Ahead

Agentic workflows will power the next wave of demand

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

1448%

Increase in tokens processed by AI models in the last 12 months¹

75%+

Of inference-driven data creation and processing at the edge by 2030²

10x

Bandwidth required upstream compared to downstream

36x

Increase in AI traffic as early as 2023-2024¹

1: Openrouter.ai/rankings
2: KPMG

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

The network becomes an LLM Fabric

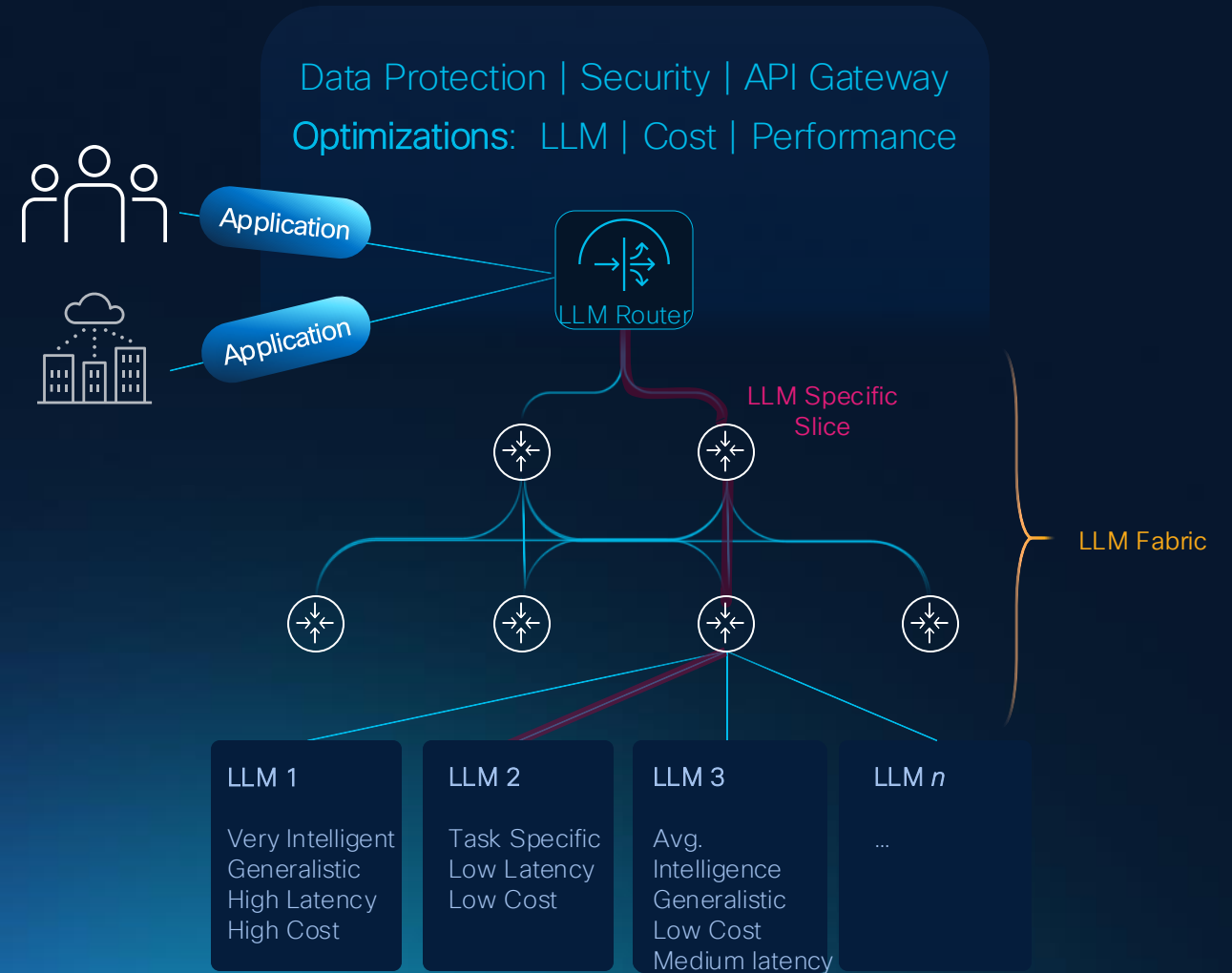
Cloud and Hybrid

Regulation and Data Sovereignty

Multi-LLM Challenges

Stringent SLA Requirements and Visibility

Distribution of Agents and LLMs



Cisco 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



Simplify networking for AI connectivity

Flexibly deliver services from locations closer to end-user demand



Service optimization

End-to-end service convergence and network slicing

Reduce architecture complexity and cost



IP and optical convergence

Routing platform for network and layer convergence

Simplify networks with systems built from a unified silicon architecture

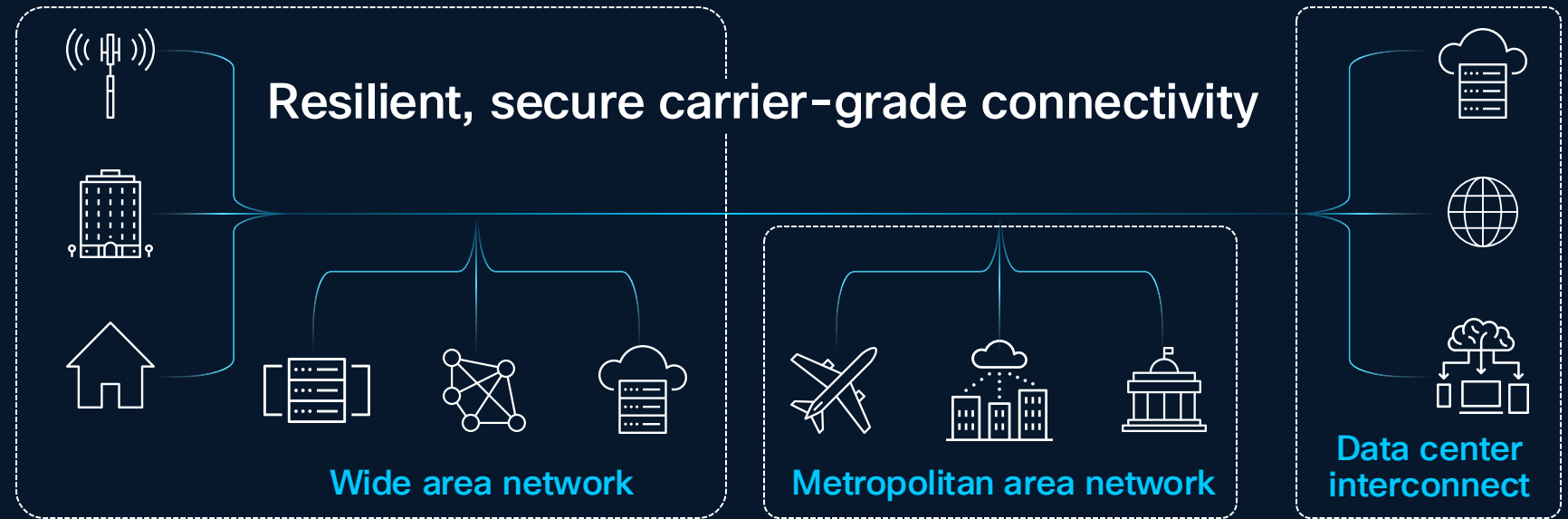


Programmable silicon

Consistent experience and capabilities across devices

Common AI connectivity needs across all sectors

- ▶ **Scale network capacity**
Distribute 100G 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

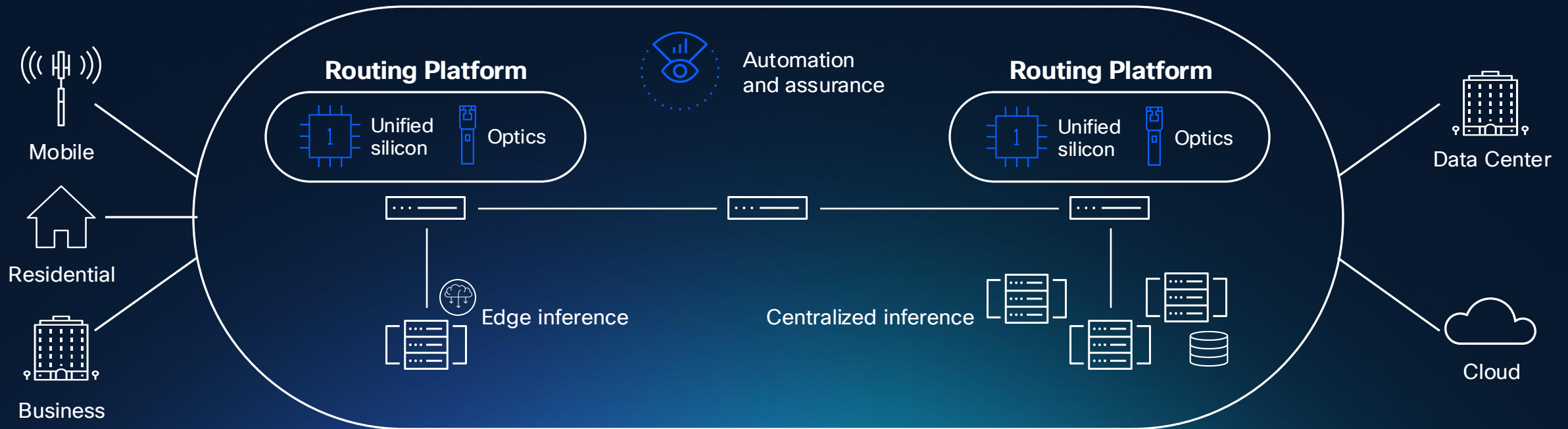


Telcos

Cisco Agile Services Networking

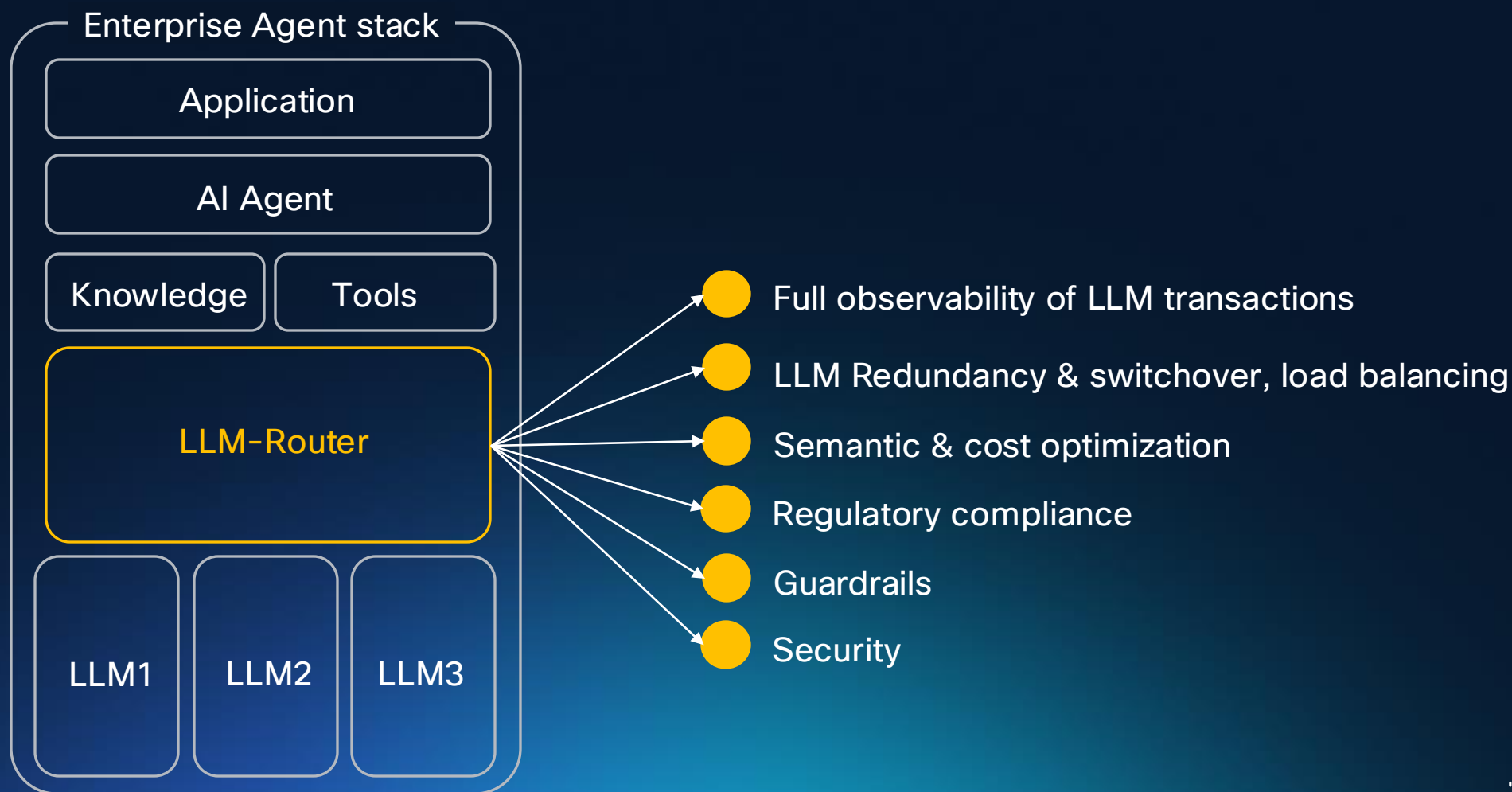
Predictable, assured, and secure network performance for AI connectivity

End-to-end architecture | Any services, anywhere



LLM Routers

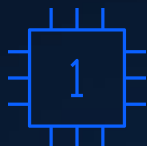
Inference Resiliency and Control



Cisco Innovation

Cisco Agile Services Networking innovations for AI connectivity

Simplified and faster
network deployment



Unified silicon

Flexible service
delivery closer to
end-user demand



Routing platforms

High performance,
scalability and
reliability



Optics

Streamline and scale
operations



Automation and
assurance



Security everywhere

Cisco Silicon One

Scalable and programmable solutions for 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



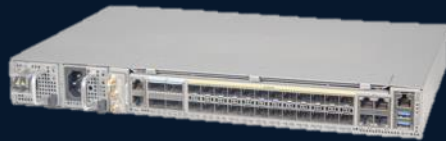
Deep visibility and analytics



Cisco 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

Cisco Optical Systems for Enterprise and Public Sector

Cisco Optical Network Controller
Cisco Optical Site Manager
Cisco Optical Network Planner

3rd party
Waves /
Spectrum

Coherent plugs



400G DCOs 1.2T C&L-Band

Transponders/Xponders



2.4T C&L-Band 1.2T C&L-Band CFP2/QDD 400G LCs

L0 cards



Compact Modular Shelf L0 and L1 cards A/D, ROADM, AMP LCs

(RON) Metro P2P Open Line System Pluggable



QSFPP-DD OLS

Metro DCI Open Line System



NCS 1001

C & L-Band Integrated Open Line System



NCS 1010

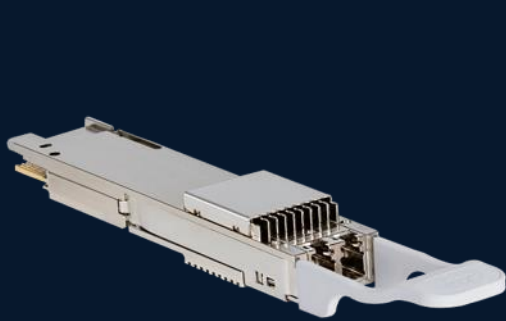
Modular Open Line System



NCS 2000

Cisco Routed Optical Networking

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

Secure Networking Innovations



Agile Services Networking



Simplified routing
infrastructure



IP / Optical
convergence



AI for Automation
and Assurance



Security fused into
the network



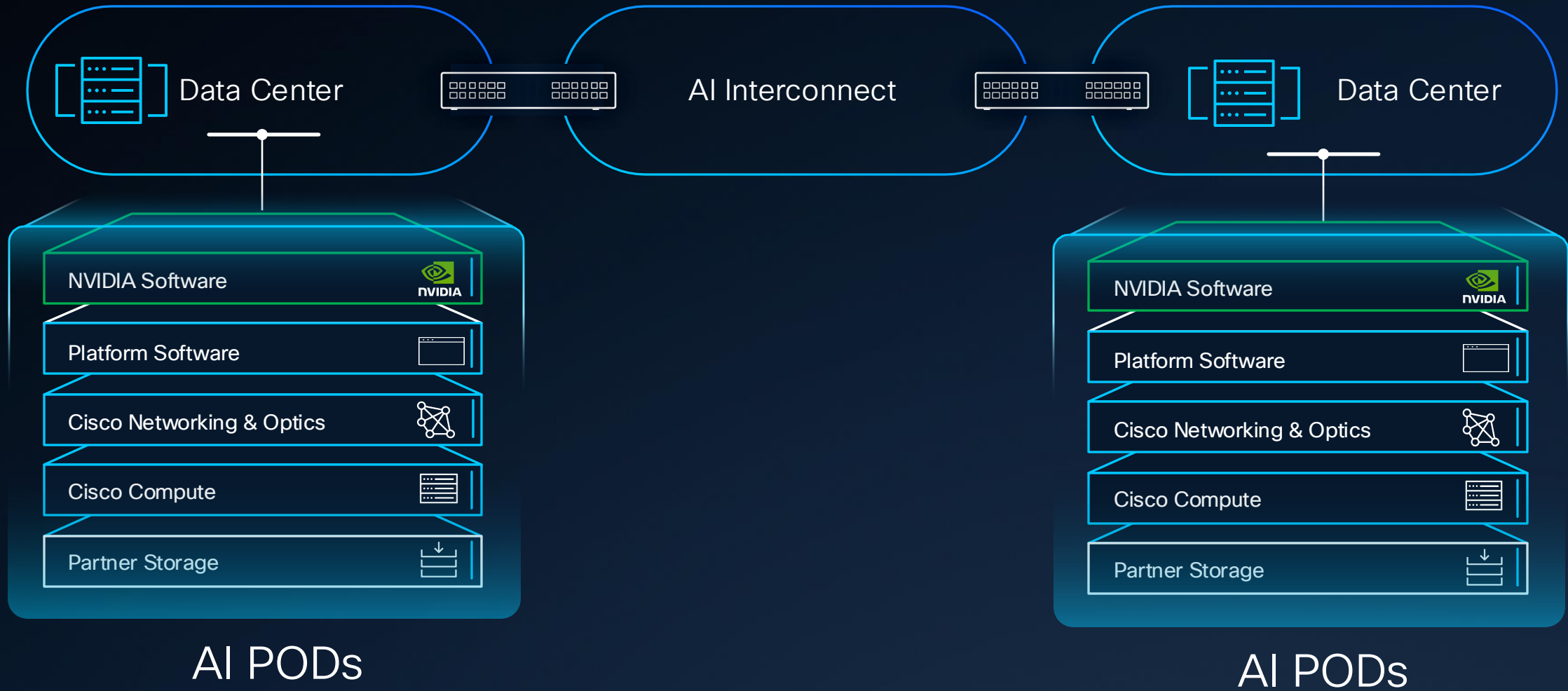
Secure AI
factory

AI READY
CONNECTIVITY

AI READY
SERVICES

AI-ready DCI

AI-Ready Data Center Interconnect



Device options for connecting data centers

Catalyst Switch



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

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
- Silicon One
- 19.2 Tbps
- Multisite connectivity
- Coherent optics to 400G
- Pluggable OLS
- Cisco Network Controller
- Segment Routing



Simplified Data Center Interconnect with Routed Optical Networking



46%
lower TCO

73%
less power

91%
less weight

Low Latency 400G DCI – now on Nexus up to 120km



Nexus capable of adding Digital Coherent Optics (check [TMGmatrix](#) for latest):

Nexus Product Family		Product ID	OS Type	DCO PID			
				QDD-400G-ZR-S	QDD-400G-ZRP-S	DP04QSDD-HE0	ONS-QDD-OLS
N9300	N9K-C9316D-GX	ACI		1	1		
		NX-OS		1	1	1	
	N9K-C9332D-GX2B	ACI		1	1		
		NX-OS		1	1	1	1
	N9K-C9332D-H2R	ACI		1	1		
		NX-OS		1	1		
	N9K-C93400LD-H1	ACI		1	1		
		NX-OS		1	1		
	N9K-C9348D-GX2A	ACI		1	1		
		NX-OS		1	1	1	1
	N9K-C93600CD-GX	ACI		1	1		
		NX-OS		1	1	1	
N9500	N9K-C9364D-GX2A	ACI		1	1		
		NX-OS		1	1	1	1
	N9K-X9400-8D	ACI		1	1		
		NX-OS		1	1	1	1
	N9K-X9716D-GX	ACI		1	1		
		NX-OS		1	1		
N9800	N9K-X9836DM-A	NX-OS		1	1	1	
		NX-OS		1	1	1	

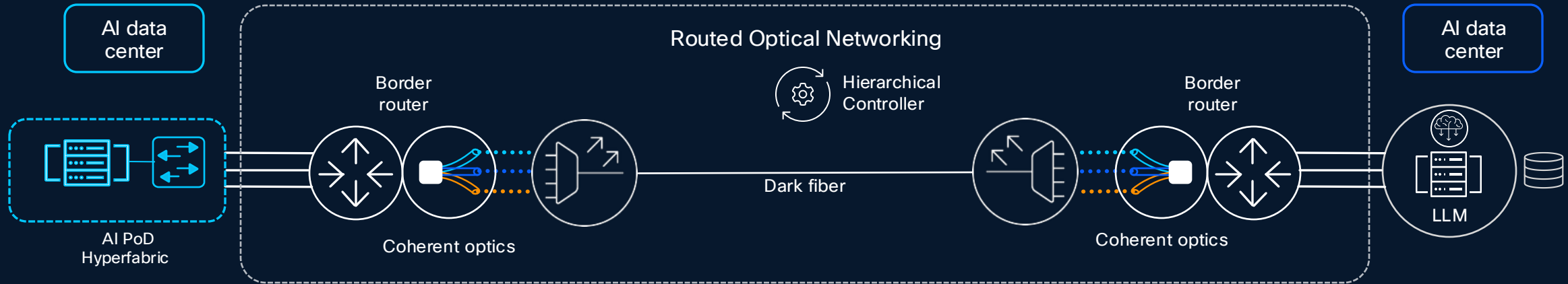
1. Low Latency DCI recipe:

- Nexus w Advanced License
- 2x Bundle: QDD400GZRH-2-BUN
- 10G PCA probe optic
- Dark Fiber
- 24x7 NOC

2. Where to start:

- Assessment
- Dark Fiber & Wavelength

AI-Ready Data Center Interconnect



Key benefits

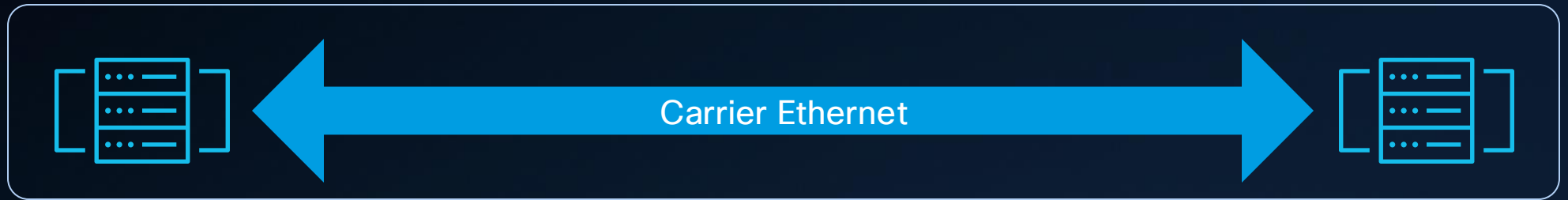
- Cisco Routed Optical Networking replaces power-hungry transponders with small pluggable optics (converge IP and optical layers)
- Suitable for any scale: from 100s of Gbps to 10s of Tbps to meet growing demands of AI
- Broad platform compatibility (e.g., Cisco 8000 and Nexus)
- Delivers up to 66% TCO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

Why Cisco

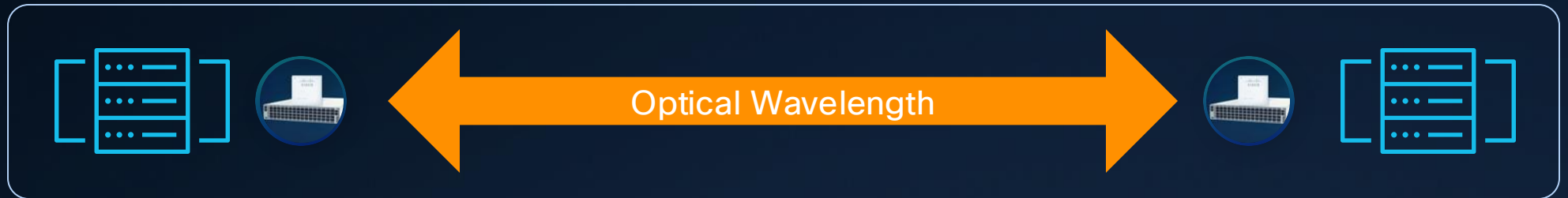
- Industry leader in optics: largest number of ports shipped to the market and the most complete portfolio, with 100G, 400G, 800G, for different reaches and use cases
- Highest level of quality assurance on optics in the industry
- End-to-end solution: Routers, optics, optical, management/automation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

ACG TCO study on common DCI designs

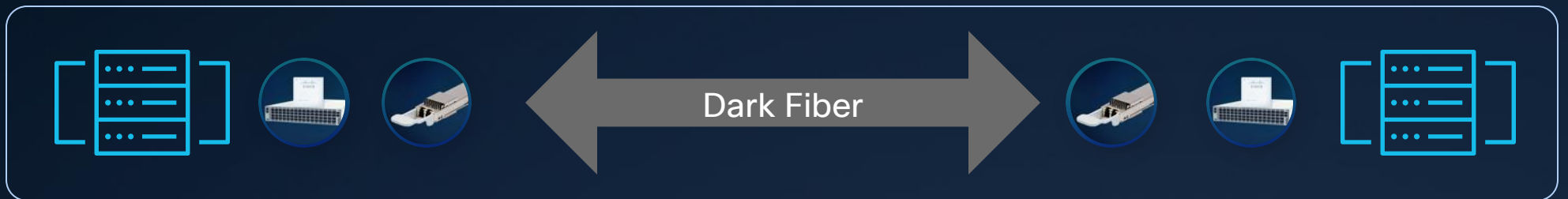
1



2

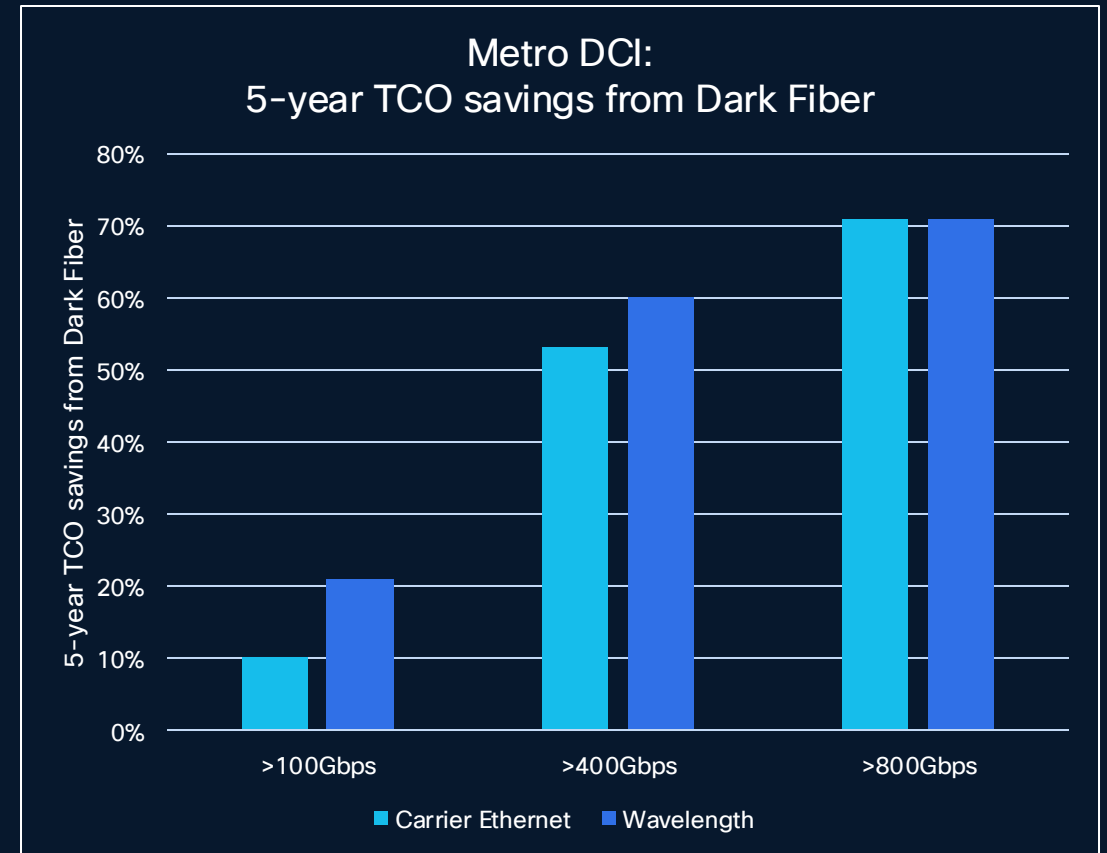


3



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.

Where to Start ?



Cisco sponsored assessments

- Free Customer Assessments
 - Security
 - Collaboration
 - Networking
 - AI
- Partner leads, includes Cisco Experts
- No public sector

Four assessment opportunities

Security

Customer Assessment Incentive for Security provides rebates to Cisco partners who conduct security assessments that help identify and mitigate security vulnerabilities, thereby improving overall cybersecurity posture.

[View security assessment benefits >](#)

Networking

Customer Assessment Incentive for Networking provides a framework and personalized roadmap to help organizations assess and evolve their network to support strategic digital initiatives and business goals.

[Explore network assessment benefits >](#)

Collaboration

Customer Assessment Incentive for Collaboration rewards partners for identifying opportunities to implement hybrid work solutions, refresh conference rooms, and reimagine a customers' workspace.

[See collaboration assessment benefits >](#)

Artificial Intelligence (AI)

Customer Assessment for AI provides the ability to identify customers' AI consumption models, infrastructure strategies, security needs, data and model governance requirements.

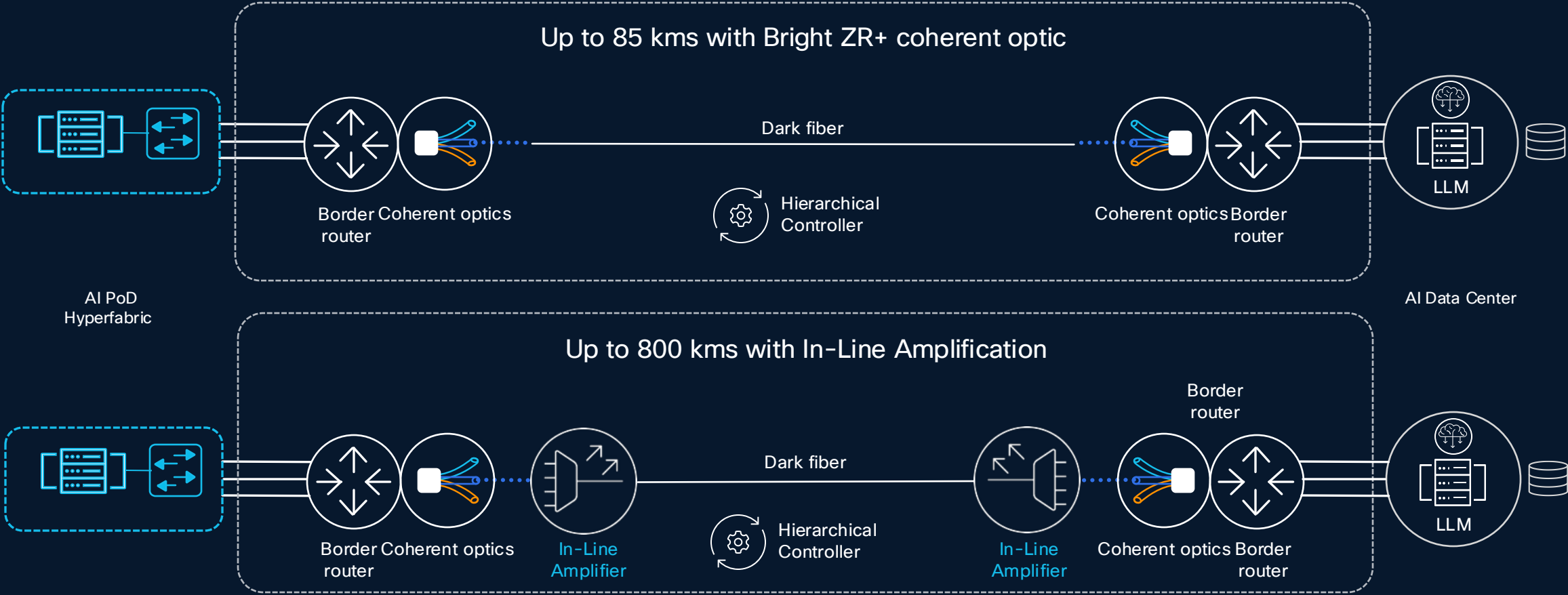
[Access AI assessment benefits >](#)

ebook: <https://ebooks.cisco.com/story/customerassessmentincentive/page/1>

Building Blocks for Data Center Connectivity

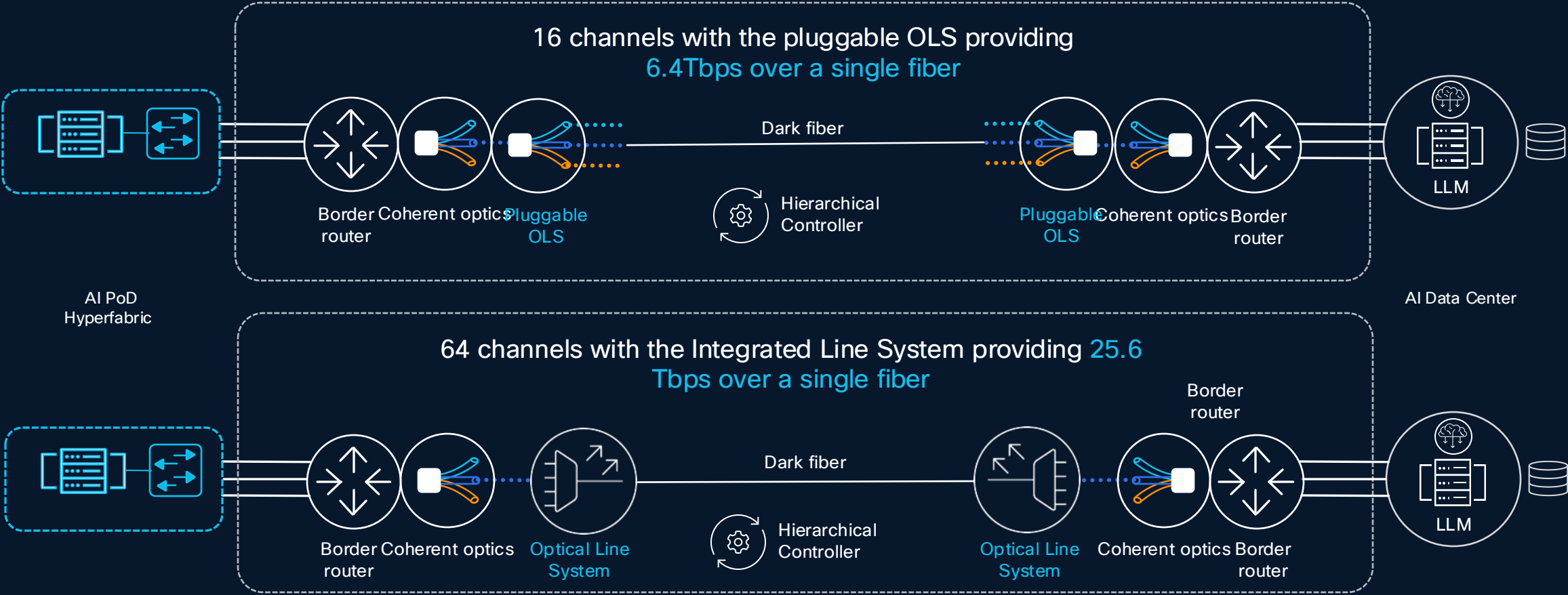
AI-Ready Data Center Interconnect

Single Span, Single 400G Wave



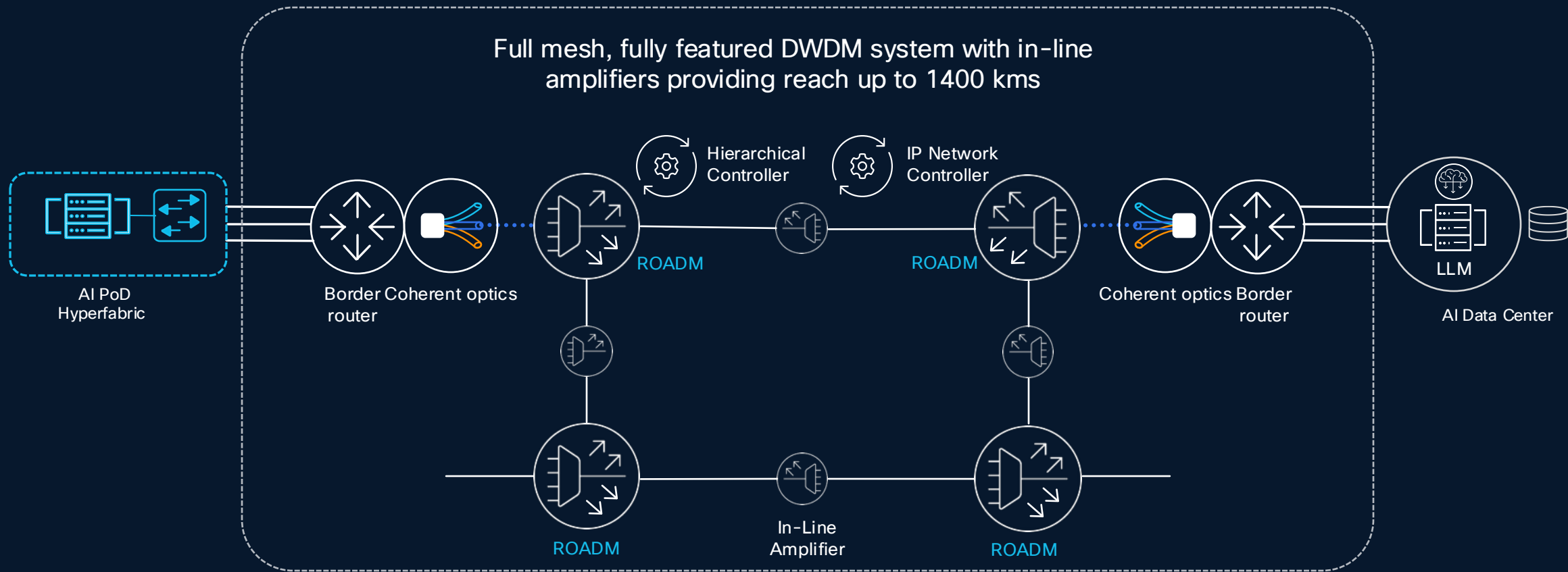
AI-Ready Data Center Interconnect

Single Span, Multiple 400G Waves



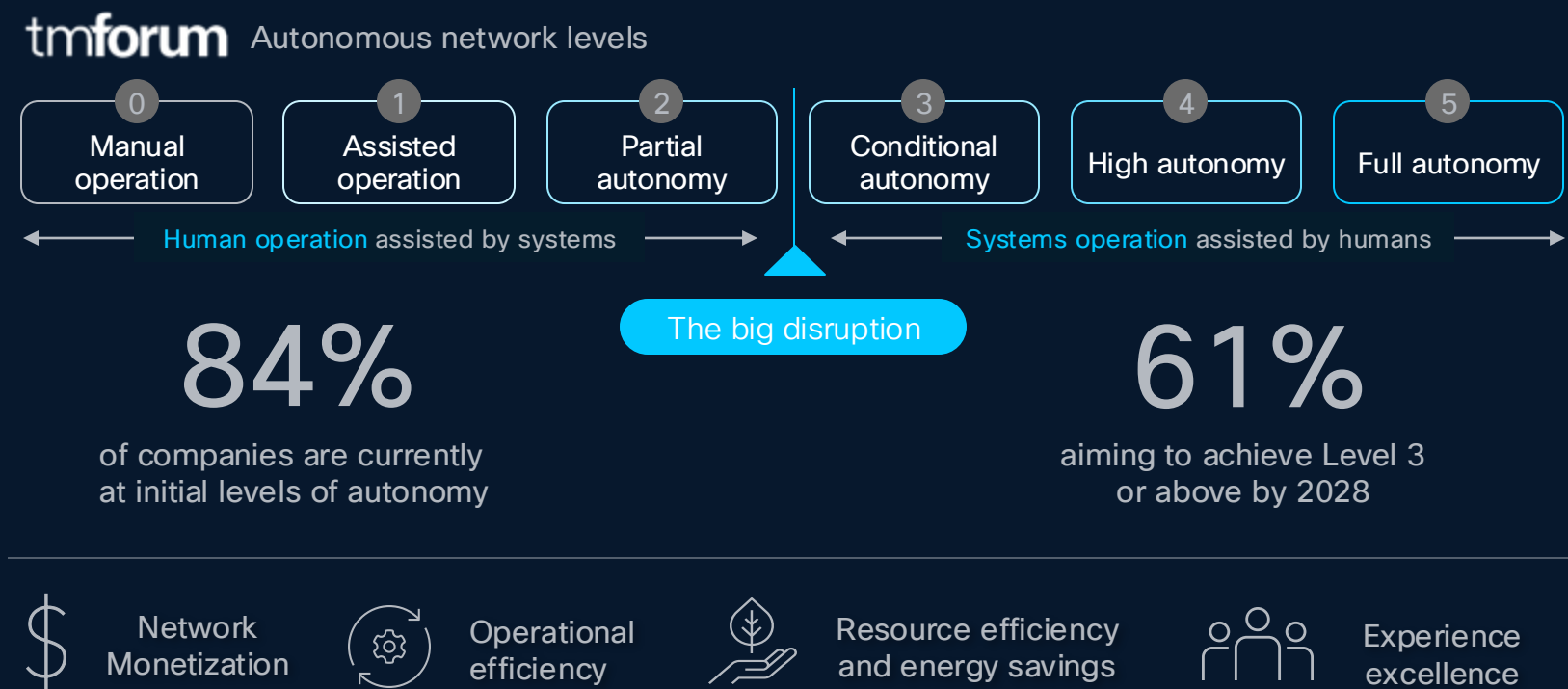
AI-Ready Wide Area Network

Multiple Spans, Multiple 400G Waves



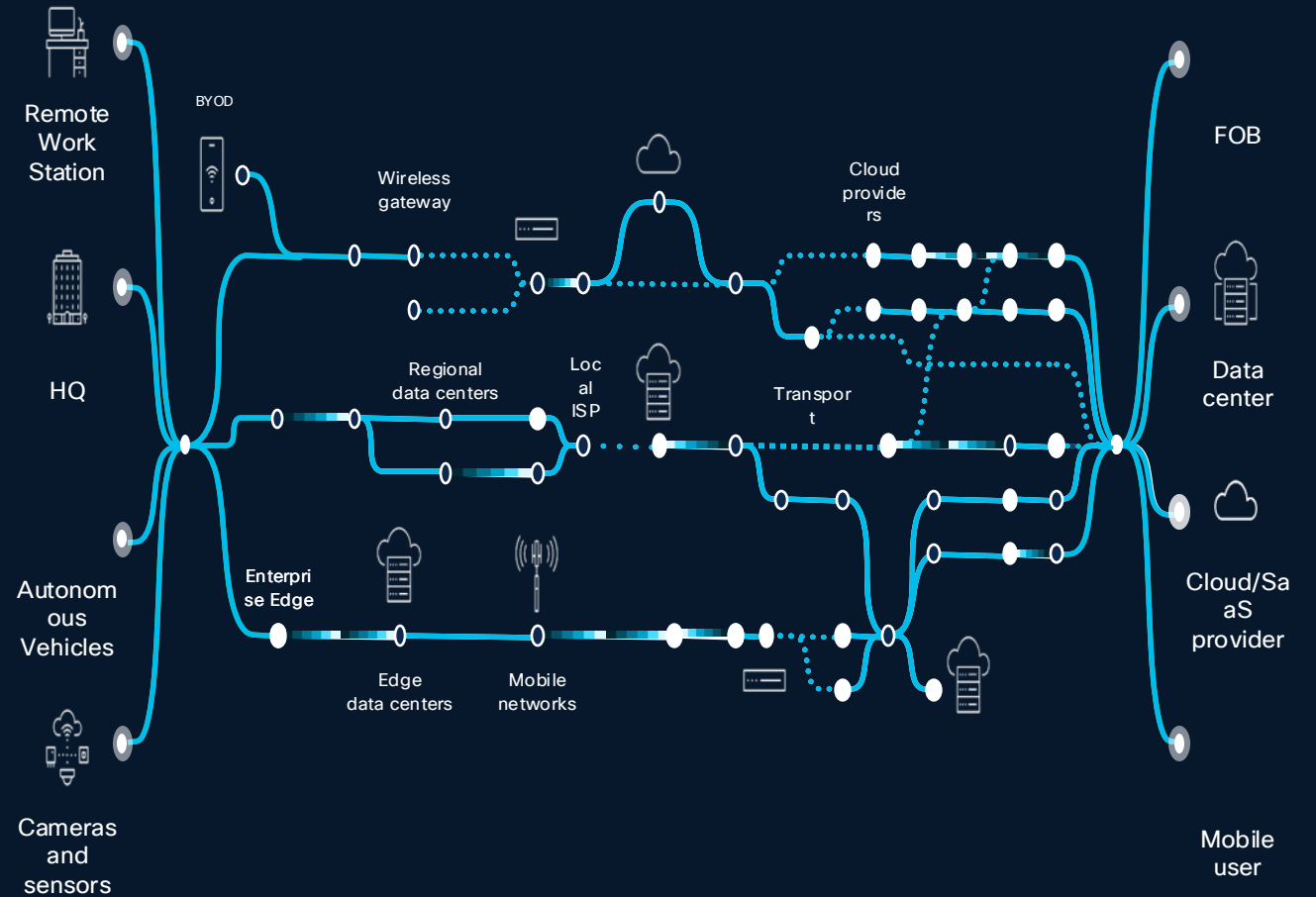
Automation and Assurance

Journey and drivers for Autonomous Networking



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**

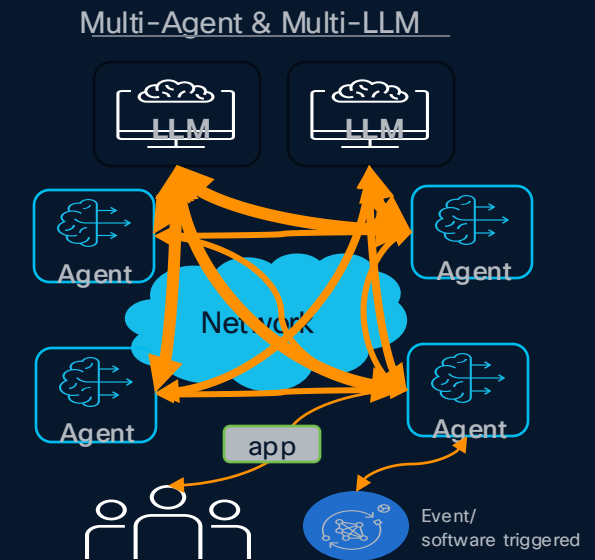
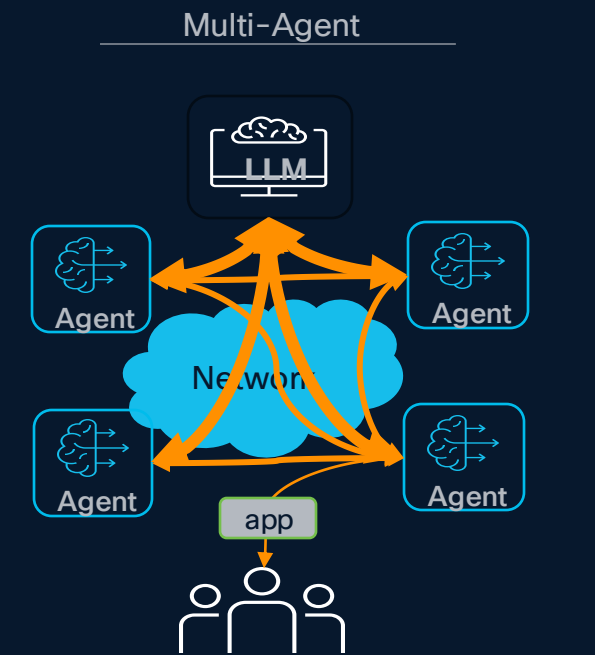
- 1% packet loss rate can reduce effective GPU computing time to less than 5%.

- **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**



Automation & Assurance Portfolio



Domain SMEs



NOC / SOC, End Customer



Crosswork Hierarchical Controller

Multi-layer, multi-domain IP and Optical hierarchical controller



Optical Network Controller

Optical Controller



Crosswork Network Controller

Multivendor IP SDN Controller



Provider Connectivity Assurance

Multivendor Service Assurance

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

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

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

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

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



Crosswork Workflow

Low-code automation workflow



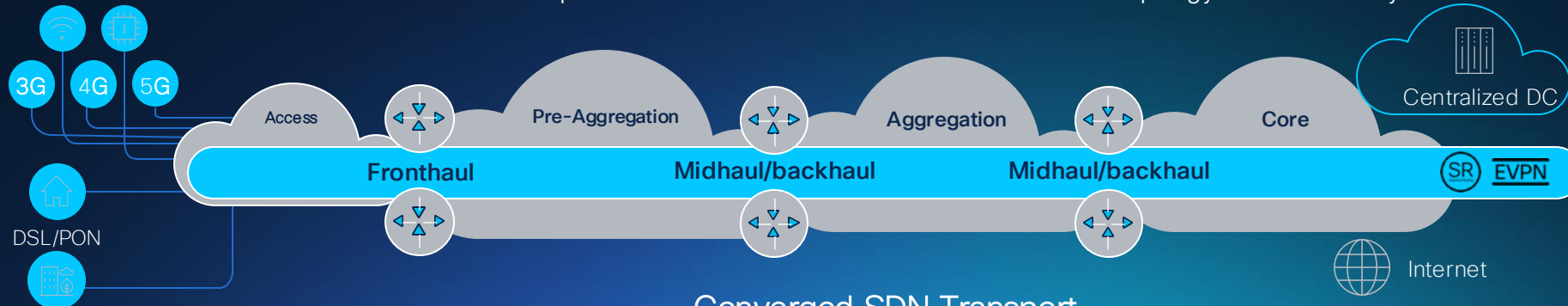
NSO

Service Orchestration



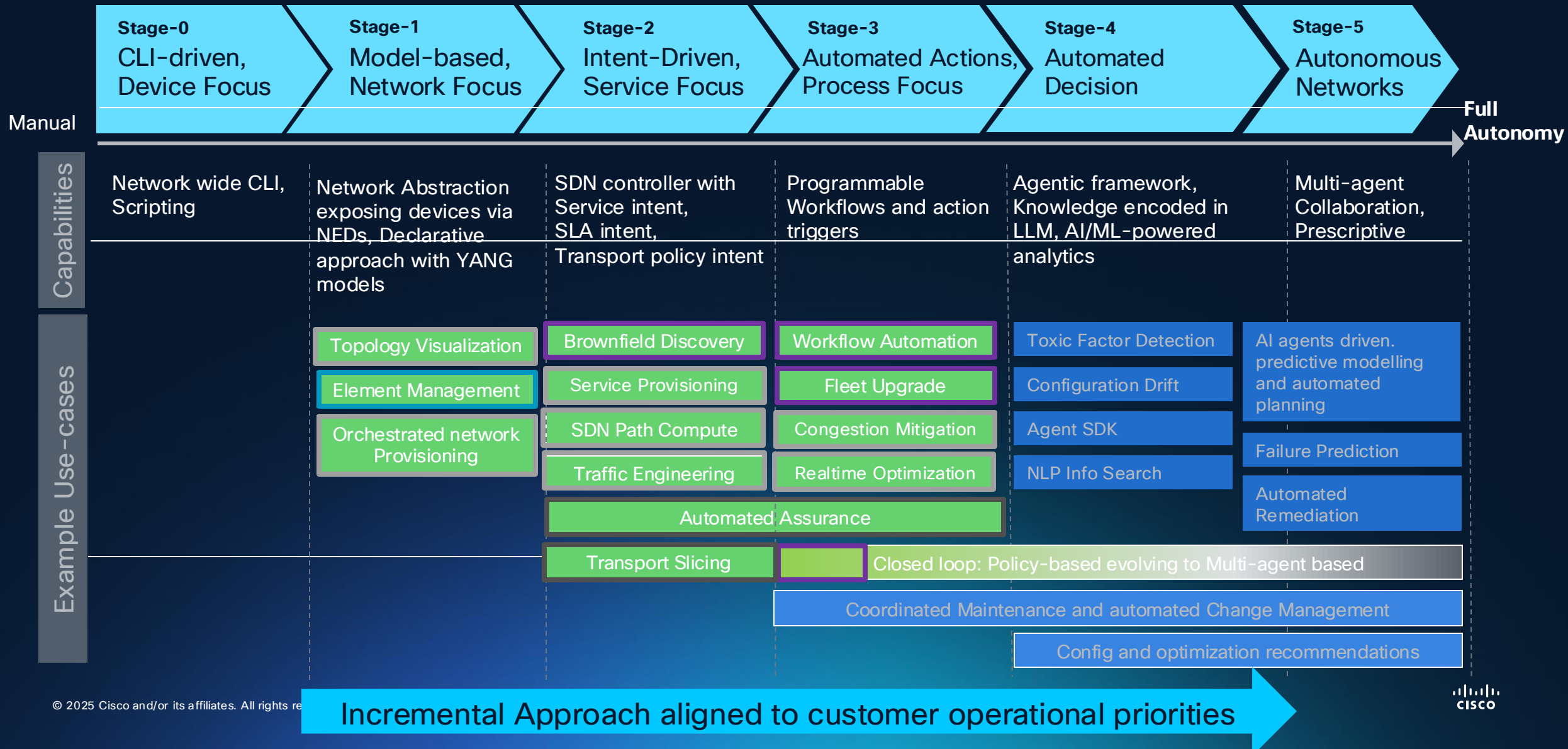
Crosswork Planning

Capacity planning tool

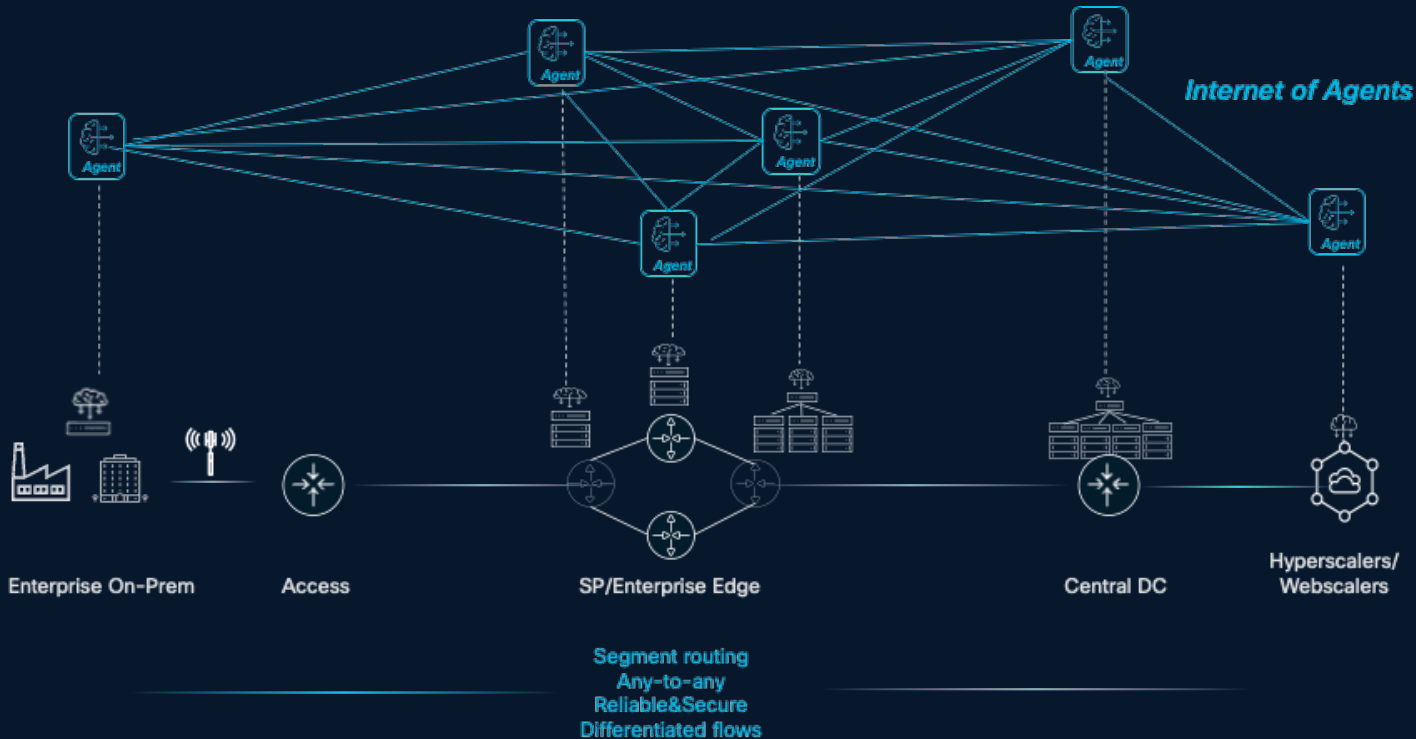


Converged SDN Transport

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

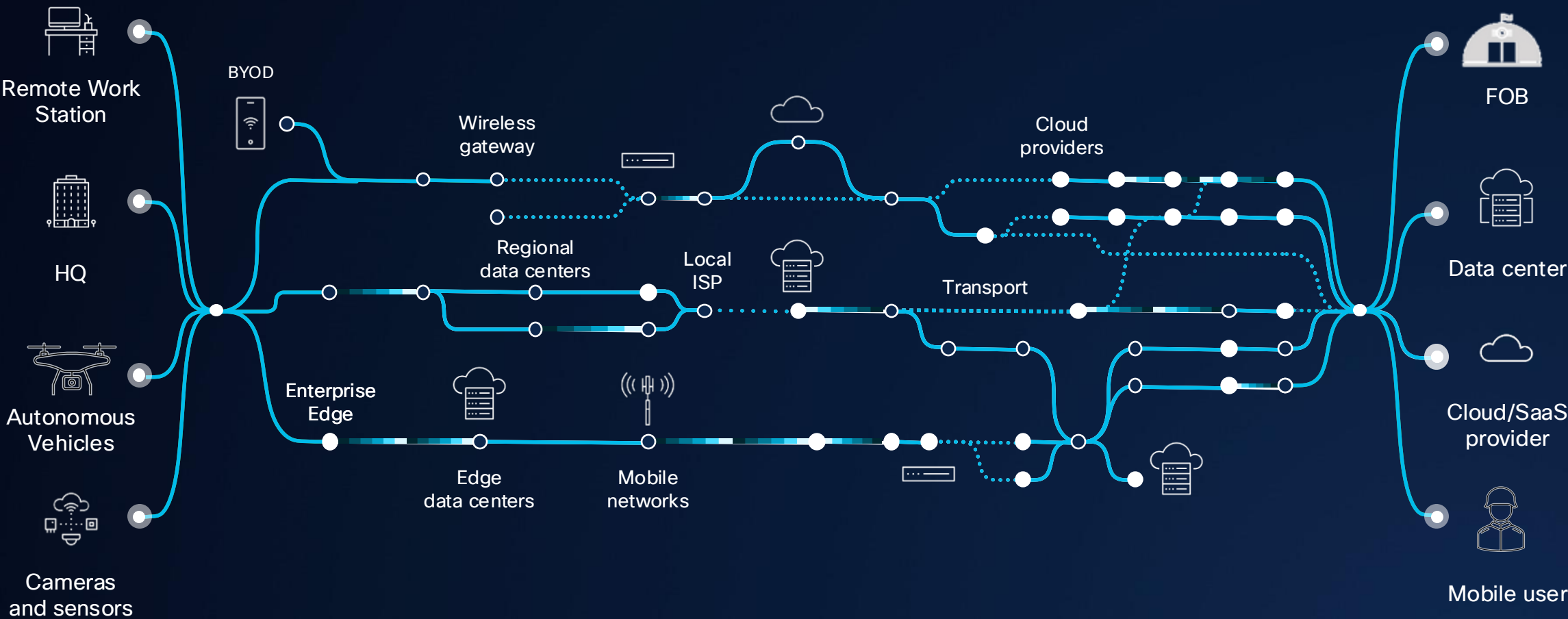
Visibility – comprehend new protocols and behavior

Traffic differentiation and assurance – QoS & path SLA

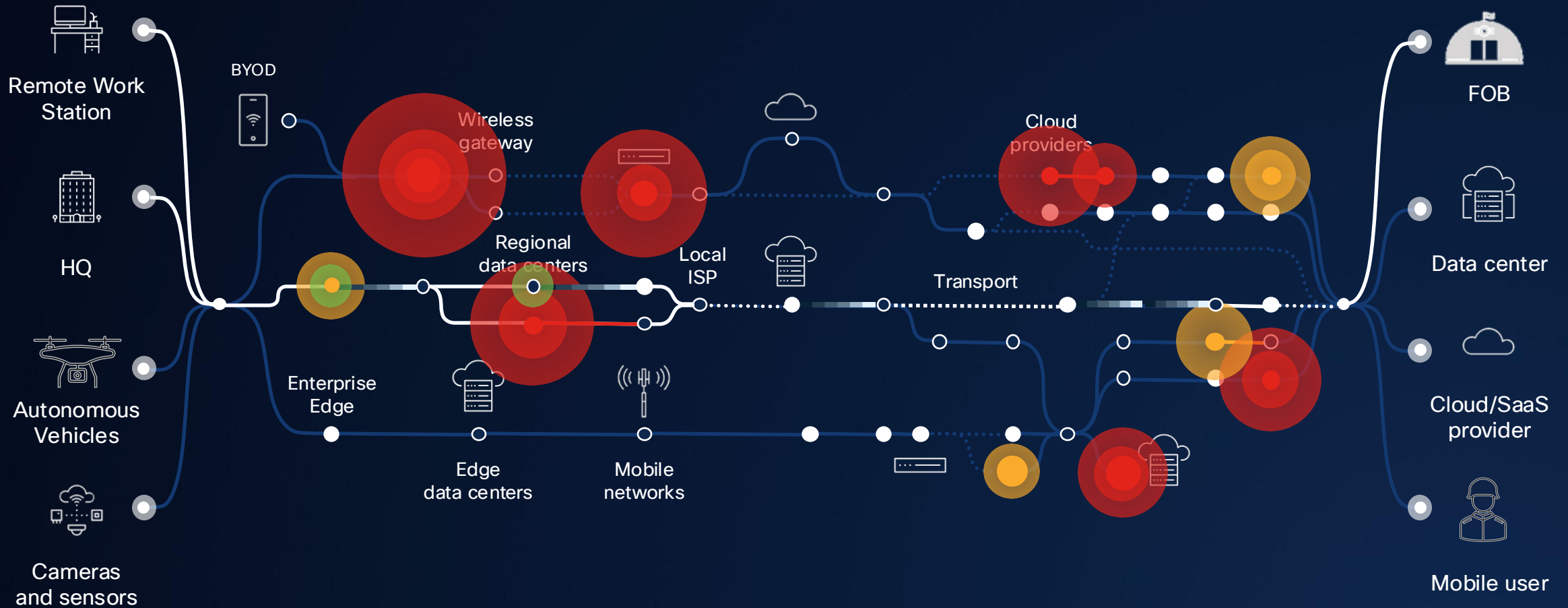
Capacity expansion – more BW, more flows, less latency

High resilience with fast failovers, multi-paths

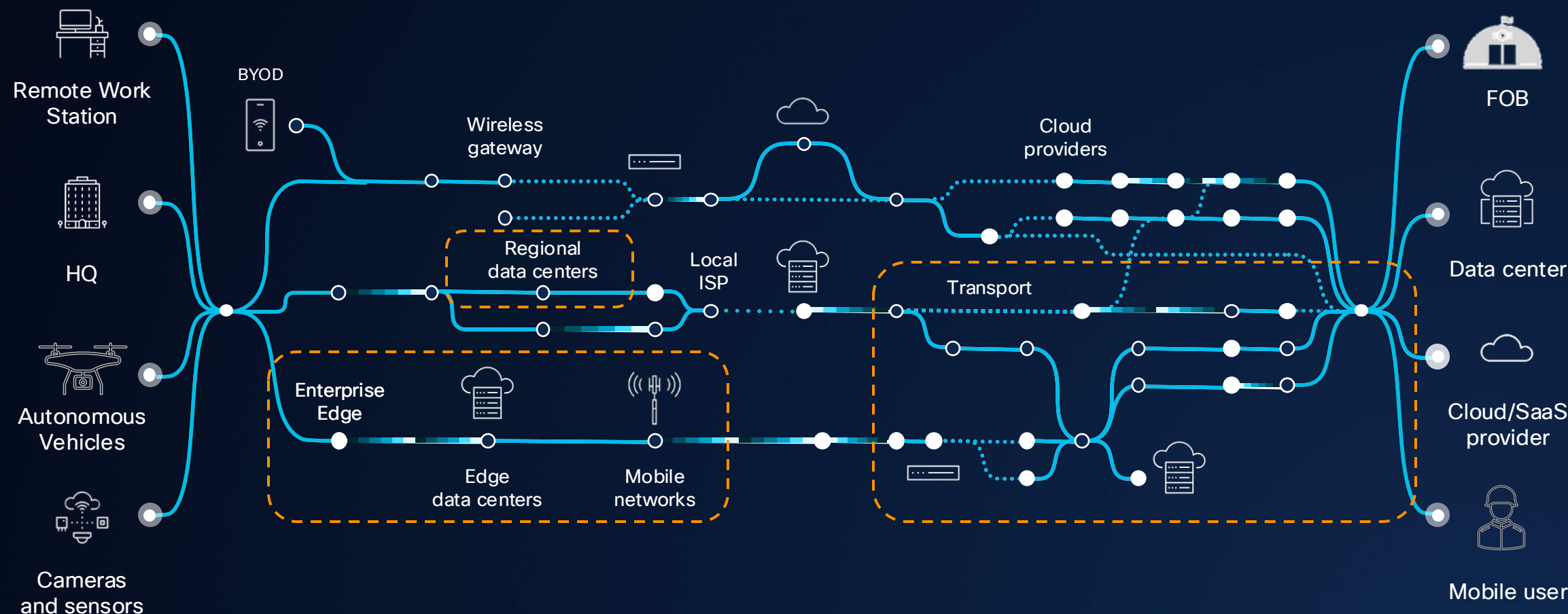
The Global Area Network



Many things could disrupt the global area network

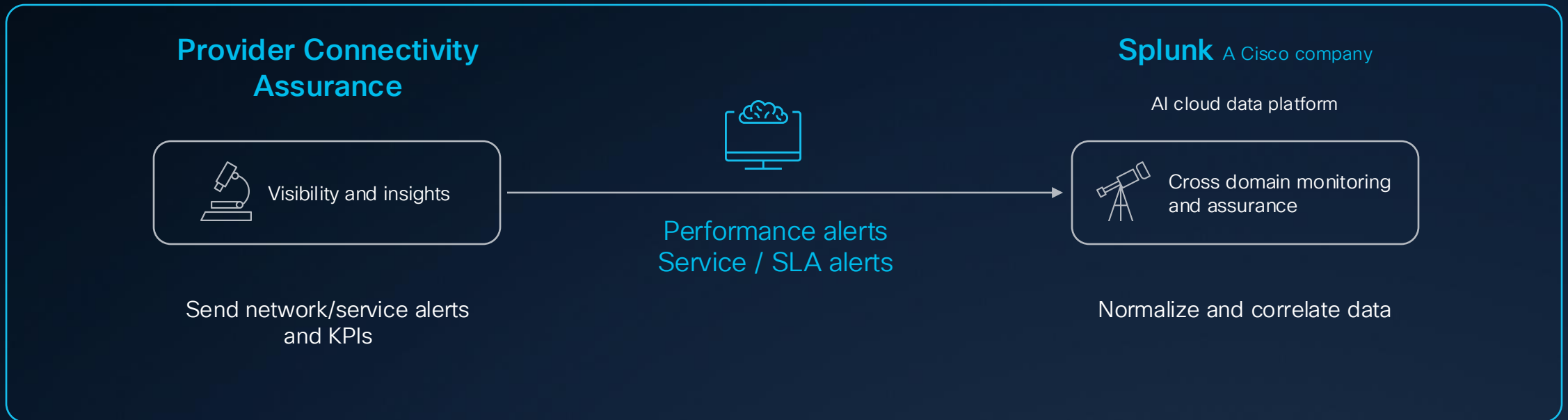


Assure critical networks with Connectivity Assurance



Cisco Provider Connectivity Assurance

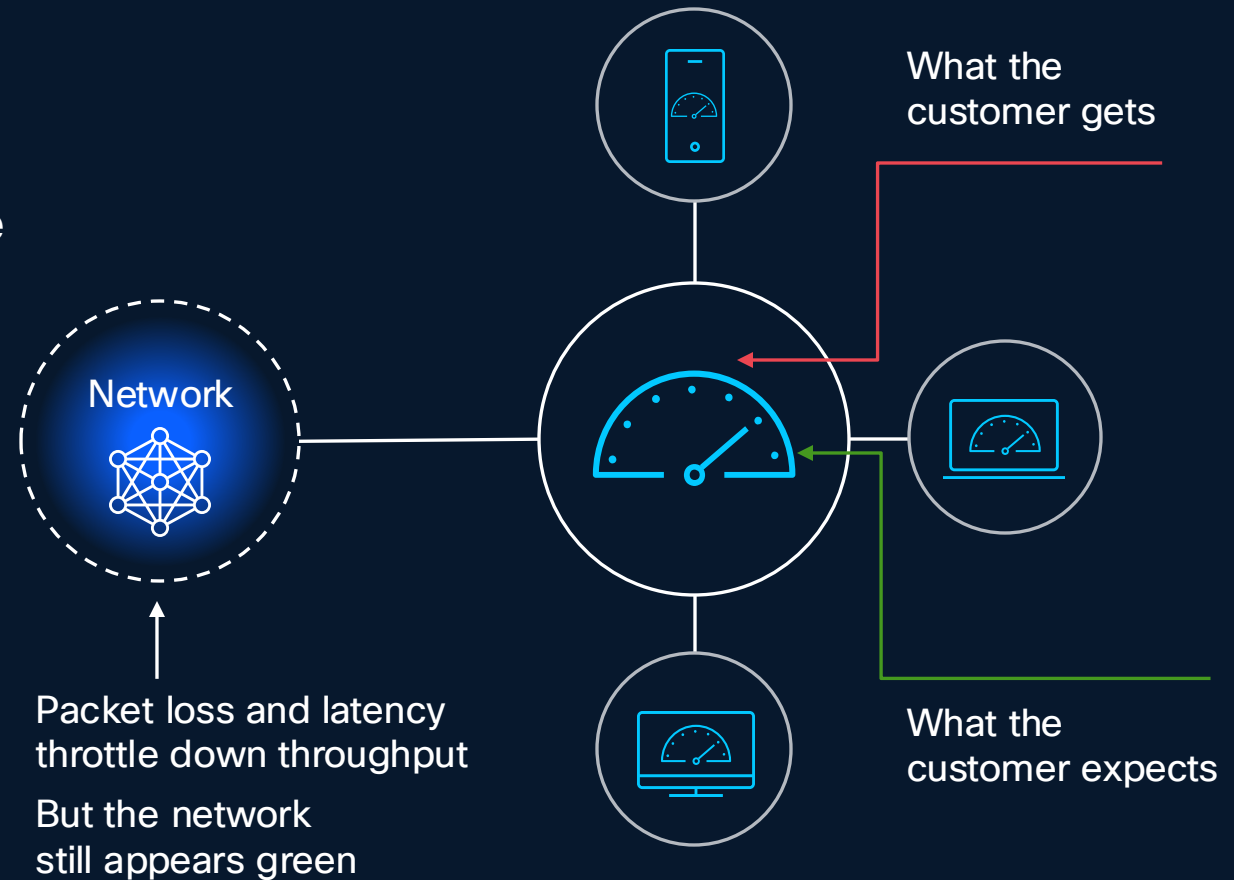
- ✓ Simplify operations with a unified view of customer experience
- ✓ Identify, prioritize and predict customer-impacting events
- ✓ Drive automated decisions to accelerate resolution



Most tools are ineffective at detecting today's QoE issues

Without deep visibility into micro events:

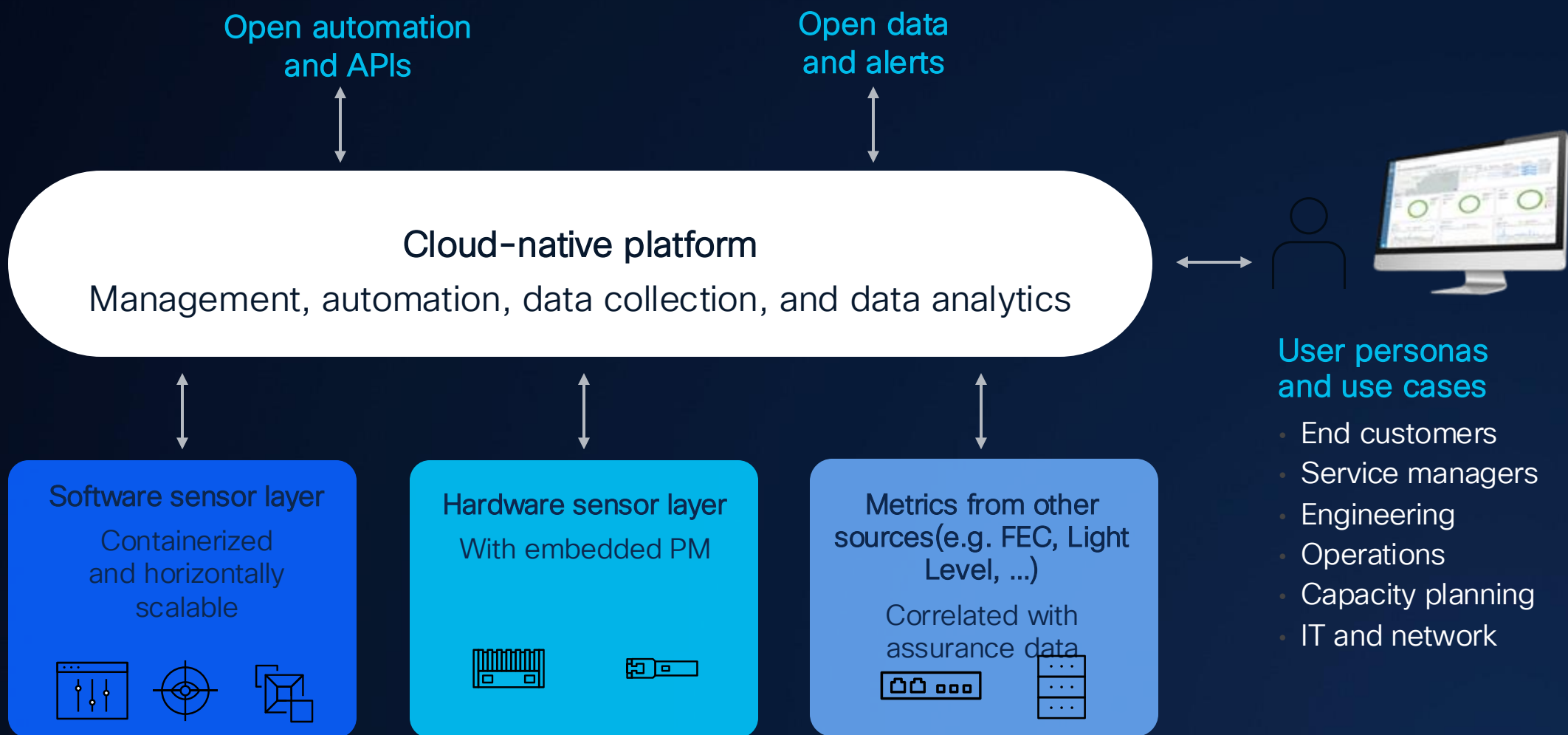
- 0.53% **packet loss** leads to a 50% decrease in data throughput
- 5 ms **delay** leads to a 10% decrease in data throughput
- 10 ms **jitter** leads to a 10% decrease in data throughput



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

Provider Connectivity Assurance solution overview

High-level view



Network-wide service performance visibility

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

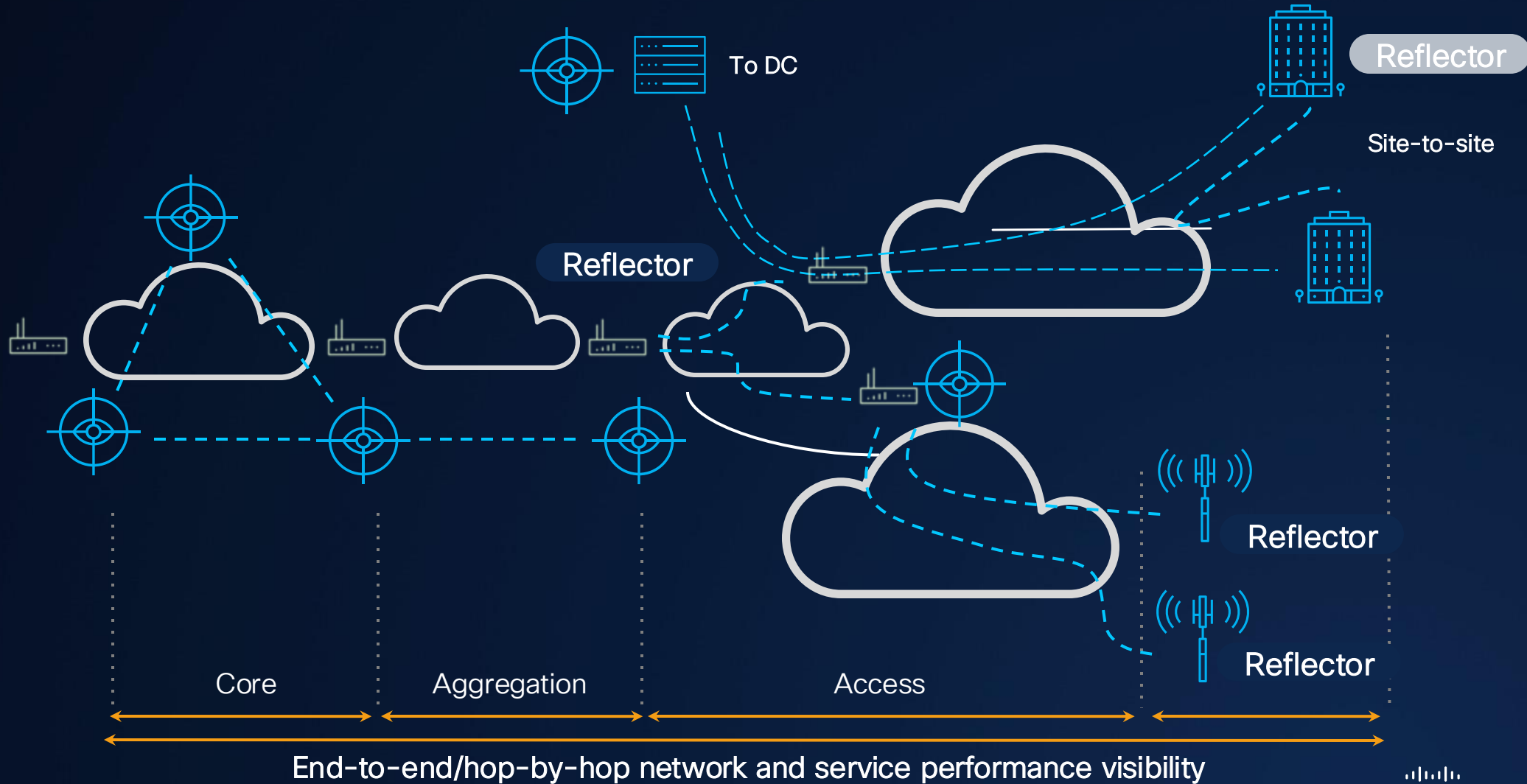
Provider Connectivity Assurance Reporting and Analytics



Assurance Sensors

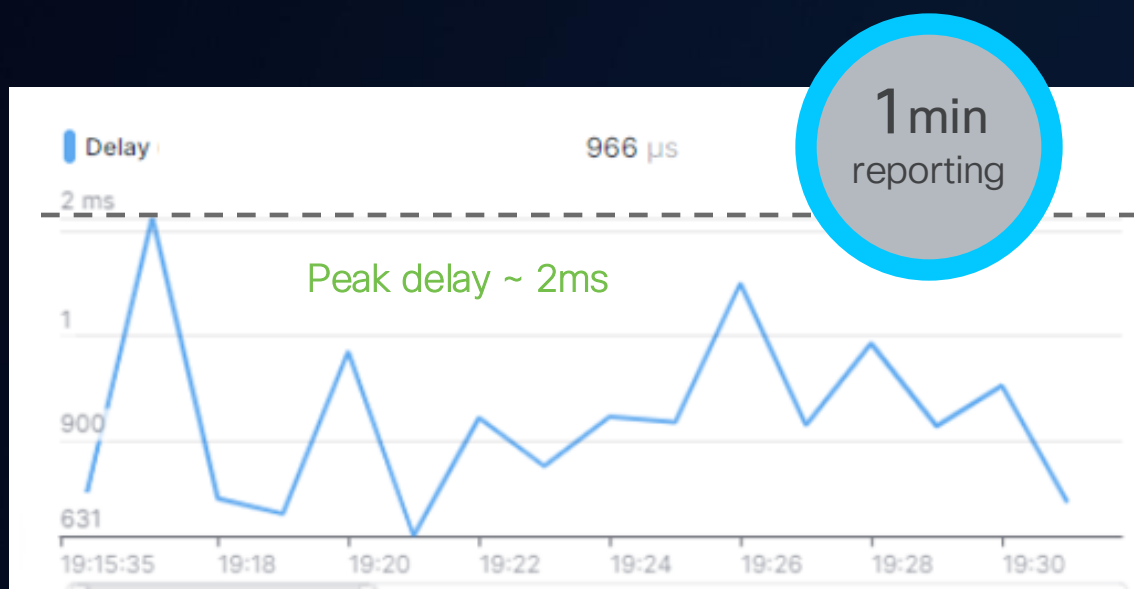


Continuous Active PM (TWAMP, Y,1731 etc.)



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.

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

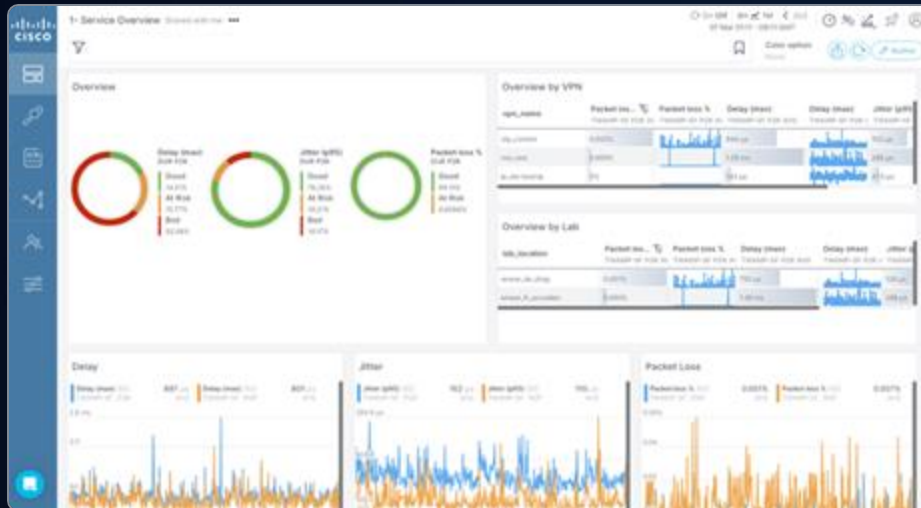
- RFC2544 generation and reflection
- Y.1564 generation and reflection

Metrics from other sources

Assurance reporting and analytics

From the macro view to micro details

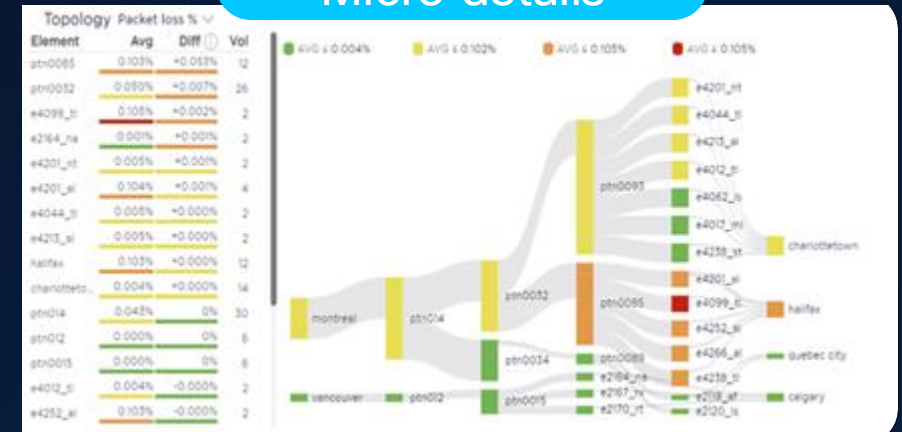
Macro view



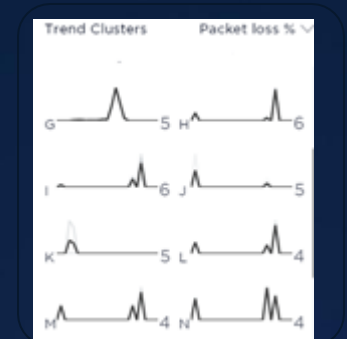
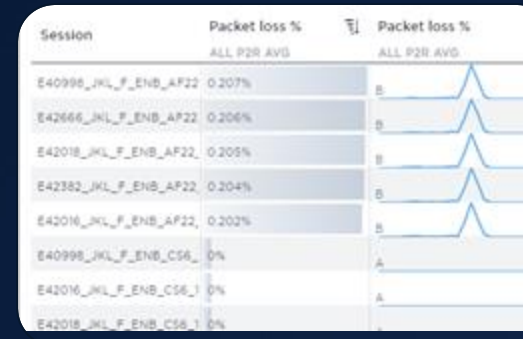
Analysis

Drill-down
and
clustering

Micro details



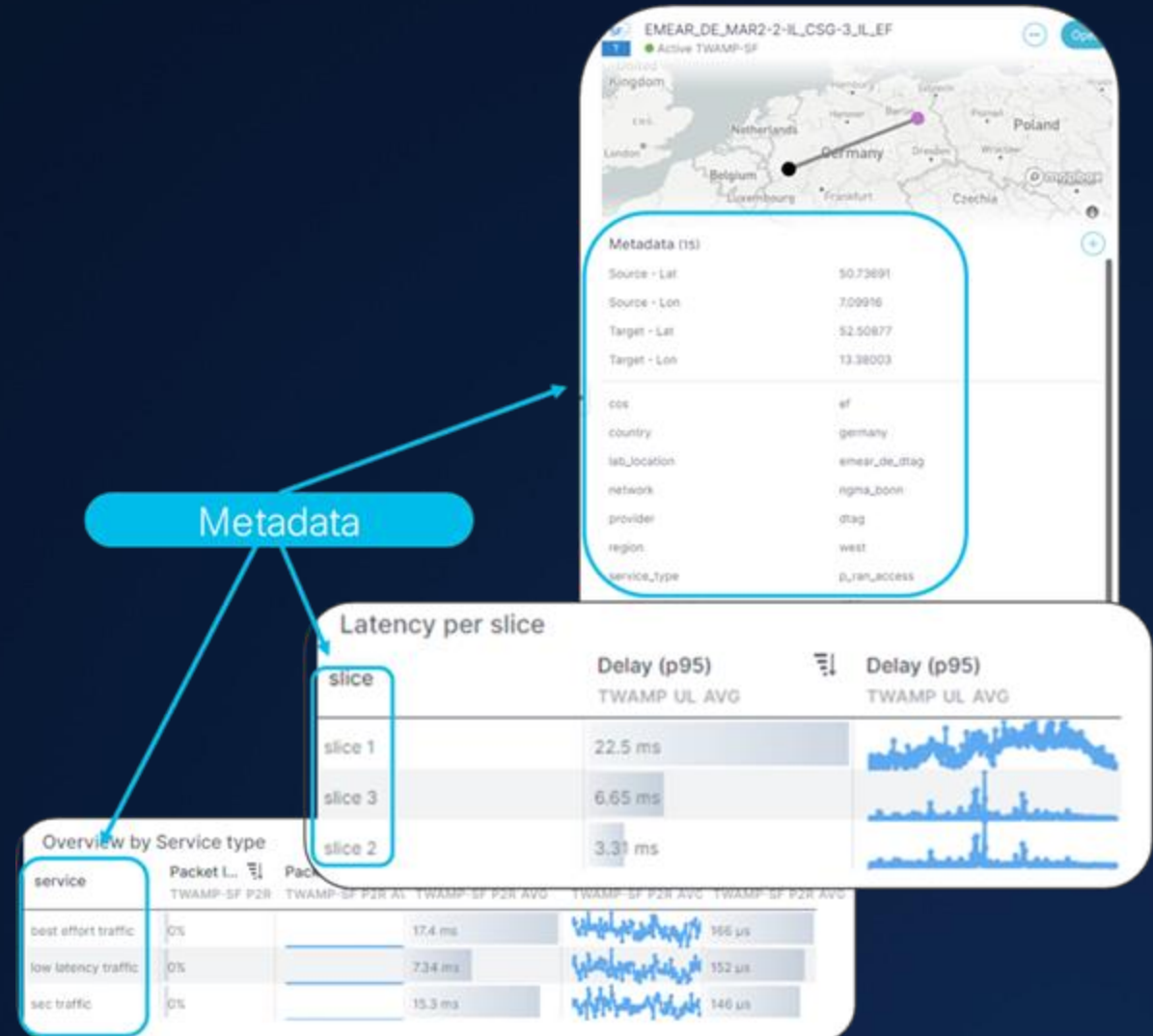
- Analytics and correlation capabilities
- Metadata enrichment makes data meaningful
- Quickly visualize network and pinpoint problems
- Increase efficacy of operations teams/



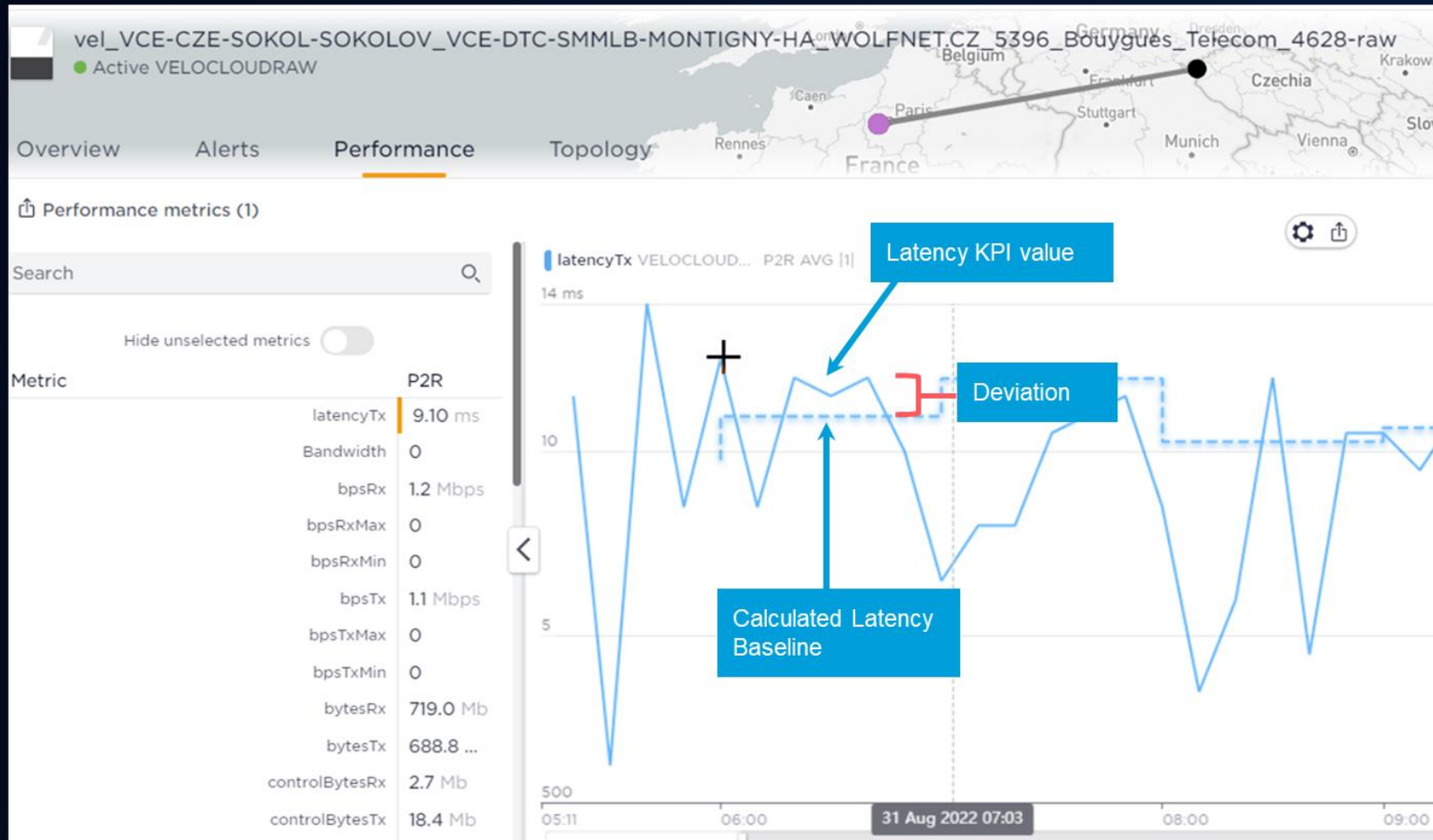
Assurance reporting and analytics

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 Baseline
- 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)

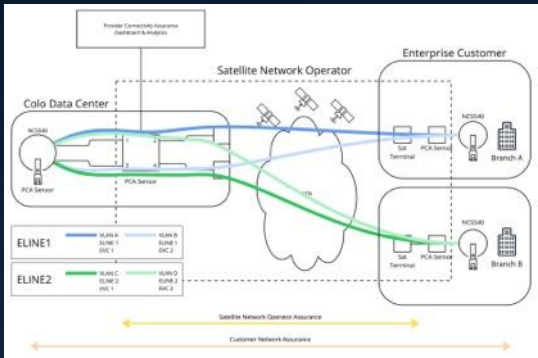
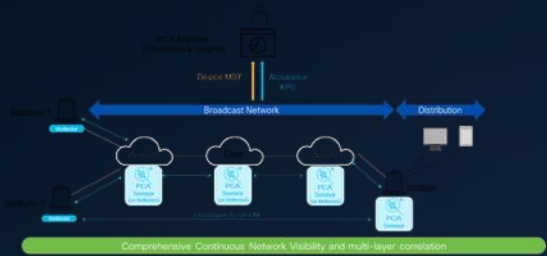
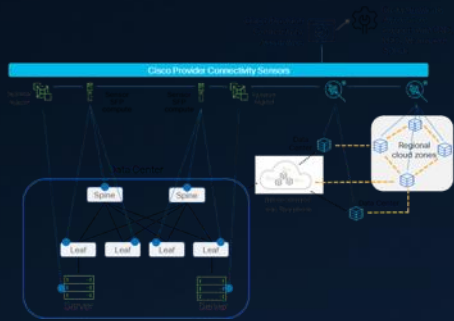
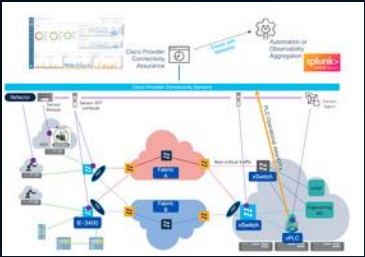
Media / Live broadcasting
infra

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

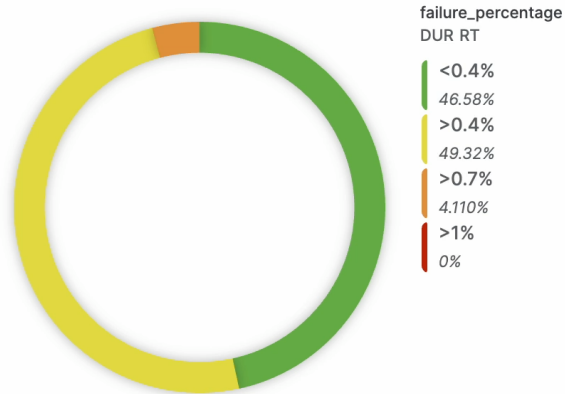
Assurance of non-terrestrial
networks (LEO)

1- AI LLM User Experience Shared By (Abhishek Jamwal)

Color option Metric

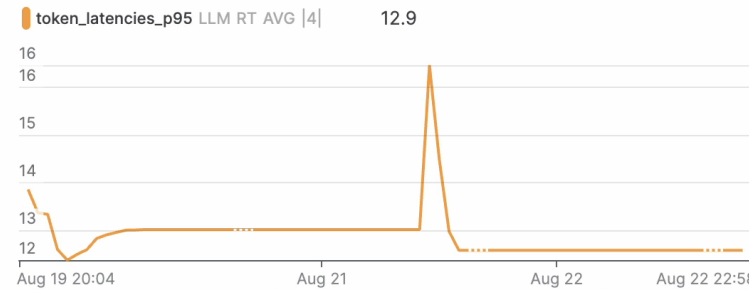
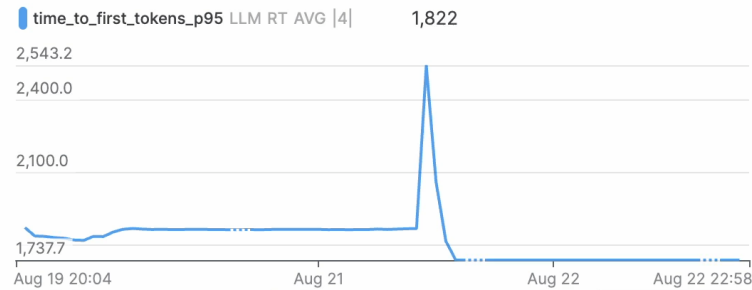
LLM Experience - Public or Enterprise LLM

Failure Percentage

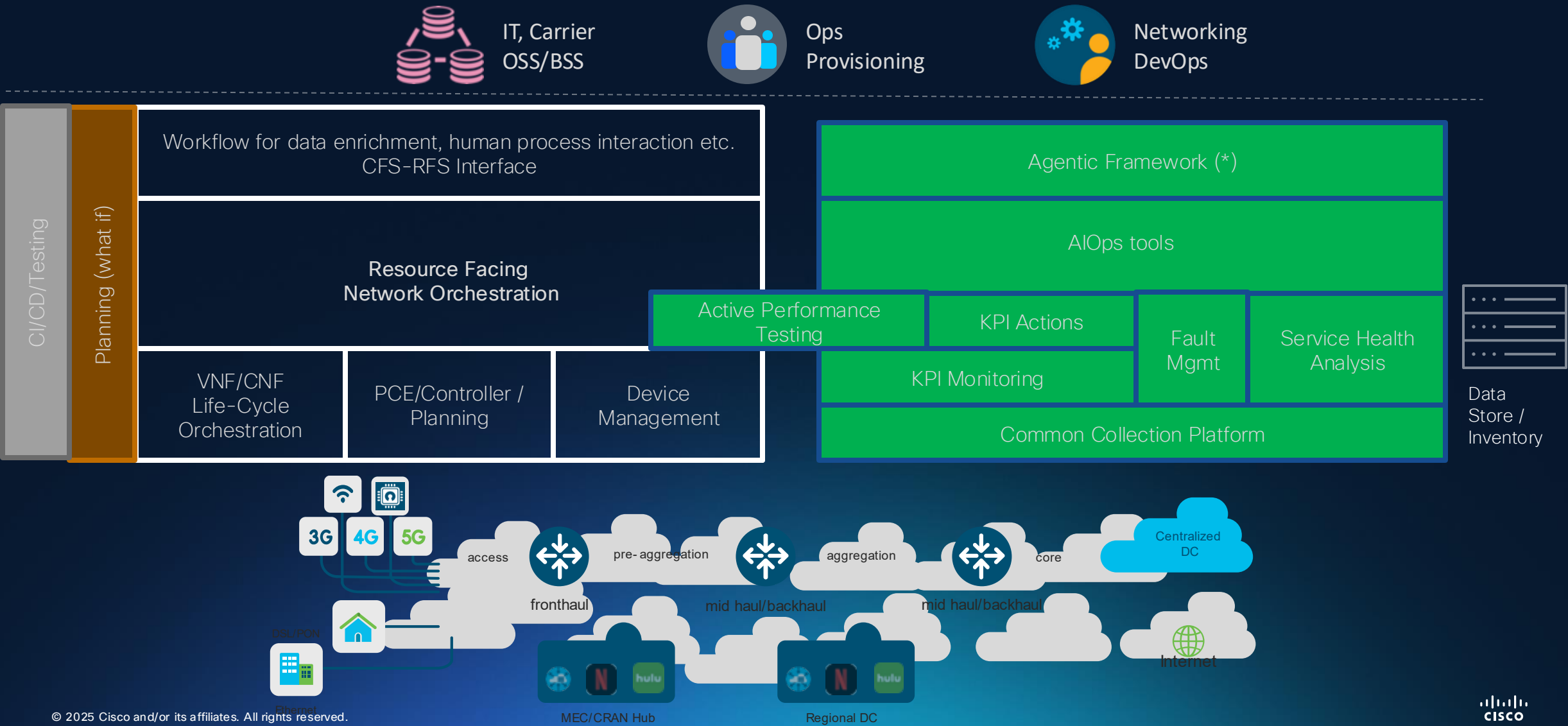


LLM_Name	trafficClass	failure_percentage LLM RT AVG	total_input_tokens LLM RT AVG	total_output_tokens LLM RT AVG	total_responses LLM RT AVG
gpt-4o	best effort	0.466	584,087	209,165	735
gpt-4o	priority_ef	0.397	600,648	215,967	756

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



Cisco is executing in these objectives



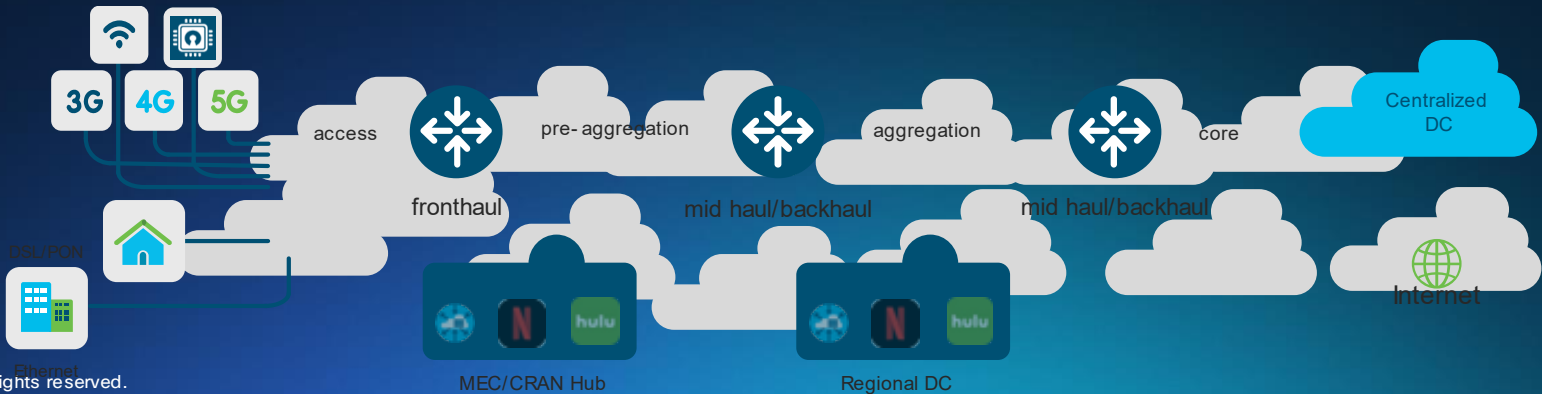
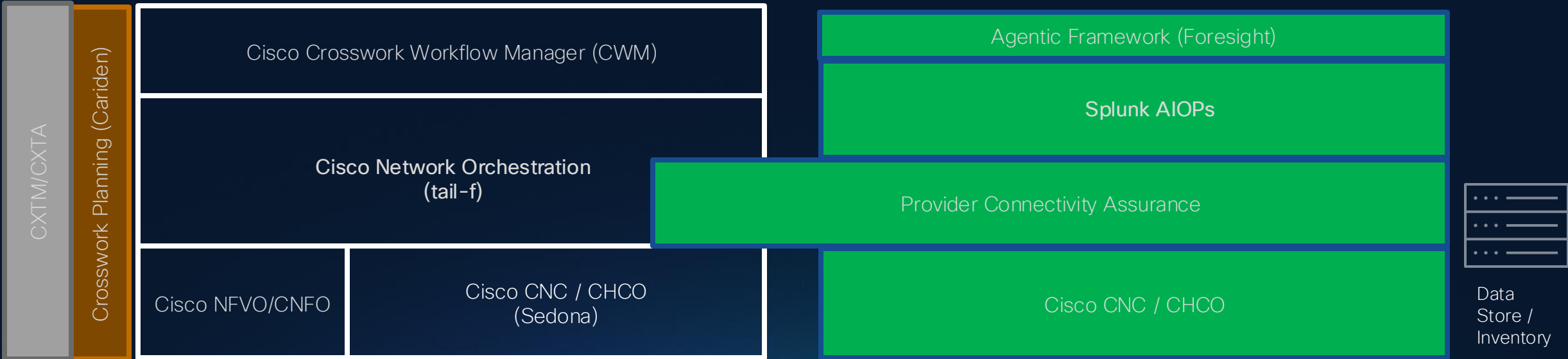
IT, Carrier
OSS/BSS



Ops
Provisioning



Networking
DevOps



One Cisco

Our AI Connectivity Vision

Simplified Networking

Simplification, resiliency,
and convergence

Autonomics with Data Intelligence

Digitized and secured
operations

Service- and AI- Ready Connectivity

Enhanced business outcomes
with intelligent service delivery

Outcomes enabled by Agile Services Networking

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

Assure experiences with AI-powered automation, observability and security

Optimize business application from an intelligent service delivery architecture

Q&A

Next Step: Is your DCI network AI ready?

- Research indicates **8/10** companies are not prepared for AI traffic growth
- **Where to Start?** Use Cisco sponsored network assessment to review existing network and propose options to get you AI Ready
- Assessments have resulted in savings of **40%** WAN costs
- Key assessment partners include:
 - **DSI Partner: Complete Communications**
 - Dark fiber and wavelength finder
 - Power, space and cooling assessments
 - **Gold Partner: Kore-Tek**
 - On-site multi-vendor Optical and Routing review
 - Network Assurance, Security
- Contact your **Cisco Partner or Cisco Account Executive**

Thank you



Cisco Agile Services Networking

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



Cisco 8000

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



Cisco Crosswork Network Automation

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



Cisco Provider Connectivity Assurance

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



Cisco Routed Optical Networking

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



Blogs

Visit blogs.cisco.com/sp360