White paper
Cisco public

Why Cisco Nexus 9000 with Nexus Dashboard for AI Networking

July 2025

Contents

Introduction

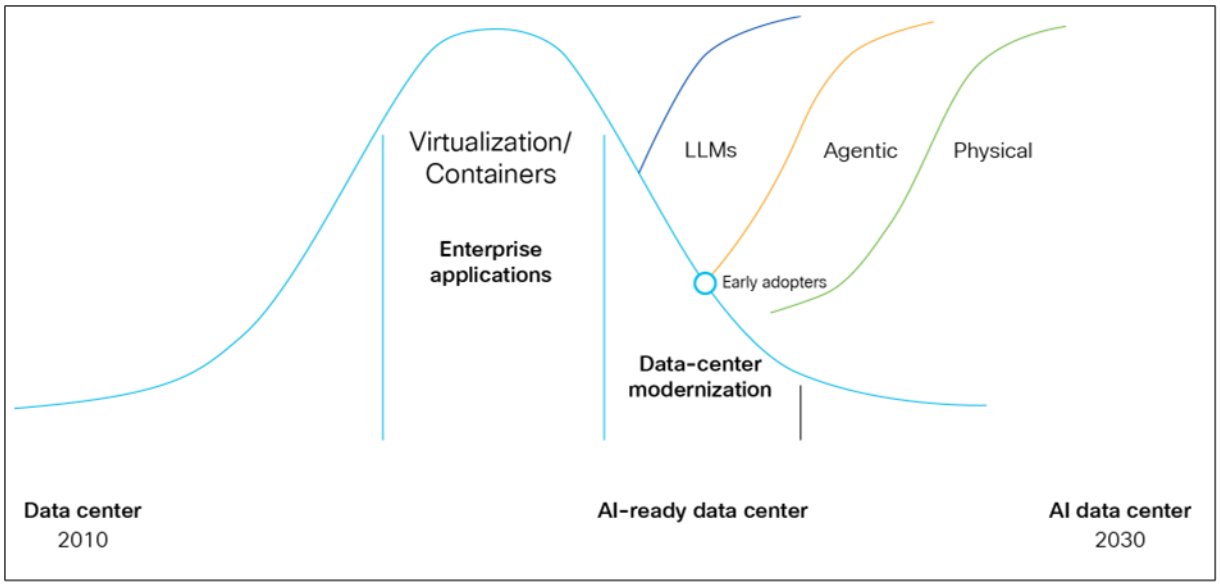

Figure 1.
Big shifts redefining the data center

The evolution of AI can be viewed through distinct phases, beginning with **Large Language Models (LLMs)**, which represent the foundational stage focused on understanding, generating, and reasoning with human language primarily within the digital domain. Building upon this, **Agentic AI** emerges as the next phase, where AI systems gain capabilities such as planning, memory, and tool use, enabling them to execute complex, multi-step tasks and automate business functions. **Physical AI** marks the advanced stage, integrating agentic intelligence into embodied forms that can perceive, interact with, and manipulate the real world, extending AI's capabilities beyond digital operations into tangible actions.

As these generative AI technologies rapidly advance and become integral to business operations, organizations will require **highly optimized networking infrastructure** to support the intensive workloads and real-time collaboration these AI systems enable.

Traditional data centers lack the high-speed, low-latency networking and massive computational power required for AI workloads such as model training, fine-tuning, inference, and Retrieval-Augmented Generation (RAG)-based and agentic workloads. AI demands specialized hardware – for example GPUs, Data Processing Units (DPUs), and AI accelerators – as well as scalable architectures, to handle the exponential growth in data and processing needs, which legacy systems can't efficiently deliver.

Modern data centers must deliver high-bandwidth, non-blocking, and lossless connectivity, support distributed AI workloads with minimal latency, and provide advanced monitoring, telemetry, and seamless integration with existing infrastructure. They must also accommodate both RDMA over Converged Ethernet (RoCE) and TCP/IP-based AI frameworks.

Ethernet stands out as the preferred solution for both frontend and backend data center networks. For frontend connections, it offers the high-speed, low-latency performance essential for demanding applications and end-user experiences. In backend environments, Ethernet enables robust bandwidth and efficient server-to-server or GPU-to-GPU communication, supporting complex AI and high-performance computing workloads. Its reliability, scalability, and cost-effectiveness, along with broad hardware compatibility and continuous advancements in speed and security, position Ethernet as a future-ready, adaptable choice for all layers of the modern data center.

Cisco Nexus 9000 Series Switches deliver purpose-built networking solutions designed specifically to address these challenges, providing the foundation for scalable, high-performance AI infrastructures that accelerate time-to-value while maintaining operational efficiency and security. Built on Cisco Cloud Scale and Silicon One® Application-Specific Integrated Circuits (ASICs), these switches provide a comprehensive solution for AI-ready data centers.

Why is Cisco the best choice for customers' AI networking for data-center infrastructure?

- **Fully integrated solution** combining custom silicon, high-performance systems, advanced software, and a unified operating model for building AI data centers.
- **High-speed and low latency switches:** powered by Cisco Silicon One and Cloud Scale ASICs.
- **Features optimized for AI traffic:** includes features such as Priority Flow Control (PFC), Explicit Congestion Notification (ECN), Data-Center Quantized Congestion Notification (DCQCN), etc.
- **Cisco Intelligent Packet Flow:** Cisco's advanced traffic management suite is designed to respond dynamically to real-time network conditions. By incorporating live telemetry, congestion awareness, and fault detection, it ensures efficient and reliable traffic steering across the fabric, helping customers unlock consistent performance and reduced Job Completion Time (JCT).
    - To keep throughput high and workloads balanced, Cisco Intelligent Packet Flow employs a comprehensive set of load balancing strategies such as:
        - **Flowlet-based/dynamic load balancing:** optimizing traffic distribution in real time
        - **Per-packet load balancing / packet spraying:** enabling multipath distribution for improved efficiency
        - **ECMP static pinning:** ensuring deterministic routing for consistent performance
        - **Weighted cost multipath:** combining DLB with real-time link telemetry and path weighting for enhanced routing decisions
        - **Policy-based load balancing:** utilizing ACLs, DSCP tags, or RoCEv2 header fields for traffic prioritization
        - **Packet trimming:** reducing packet sizes instead of dropping them to maintain data integrity
    - Cisco Intelligent Packet Flow extends visibility deep into the fabric with advanced telemetry, empowering operators to observe and optimize traffic in real time: **microburst detection congestion signaling, tail timestamping, and In-Band Network Telemetry (INT)**.
    - In large-scale AI environments, failures such as link degradation or switch outages can introduce performance bottlenecks. It mitigates these risks with **real-time fault detection, traffic rerouting around degraded paths, fast convergence, and failure isolation**.

- **Flexible connectivity options:** With support for 400G and 800G in both QSFP-DD and OSFP form factors, **Cisco Optics** ensure broad compatibility.
- **Unified management:** Cisco Nexus Dashboard provides built-in templates for AI fabric provisioning and delivers end-to-end RoCEv2 visibility and congestion analytics. Ongoing enhancements will bring job-level insights, topology-aware visualization, and NIC observability, advancing operational intelligence across AI fabrics.
- **Ultra Ethernet-ready:** Cisco is one of the key steering members for Ultra Ethernet Consortium (UEC), and Cisco Nexus 9000 switches are Ultra Ethernet-ready to support the mandatory network-side requirements.
- **Vendor-agnostic:** Cisco offers vendor-agnostic support with validated designs across leading AI ecosystem partners including NVIDIA, AMD, Intel®, VAST, WEKA, and others.

Performance benchmark data with Cisco Nexus 9000 Series Switches

Performance benchmarking is essential for understanding, measuring, and optimizing the efficiency of systems, applications, or infrastructure. It provides a baseline to evaluate how well a solution performs under specific workloads or scenarios, enabling organizations to identify bottlenecks, inefficiencies, or underutilized resources.

Cisco has successfully validated the Cisco Nexus 9000 Series Switches and the Cisco Nexus Dashboard platform for RoCEv2-based AI clusters, demonstrating their ability to meet the performance demands of distributed AI workloads. Comprehensive testing, including RDMA/IB performance, NCCL benchmarks, evaluations of multiple load-balancing strategies, and model training showcase exceptional performance and operational simplicity. Read the detailed performance validation insights for Nexus 9300-GX2As deployed in Cisco IT clusters here. To get insights on uncompromised Ethernet performance and benchmarking for AI/ML fabrics with Cisco G200-based switches, read here.

Foundations of Cisco AI Networking with Nexus 9000 and Nexus Dashboard

Cisco is the only vendor in the industry to provide a fully integrated solution combining silicon, systems, optics, software (OS), and unified management for building AI-ready data centers.

Customer requirements are currently evolving to the point where they need a vendor that controls its own fate, meaning the vendor owns and manages its entire technology stack, from hardware to software, without relying heavily on third-party dependencies. This level of control ensures greater reliability, innovation, and accountability, and Cisco is that vendor to provide this differentiation. Customers appreciate Cisco for its clear vision, substantial investments in research and development, and comprehensive nature of its solutions. Our rich ecosystem partnership also provides cutting-edge solutions while protecting our customers' investments These differentiators instill confidence in the company's long-term stability and dedication to delivering innovative, cutting-edge technologies.

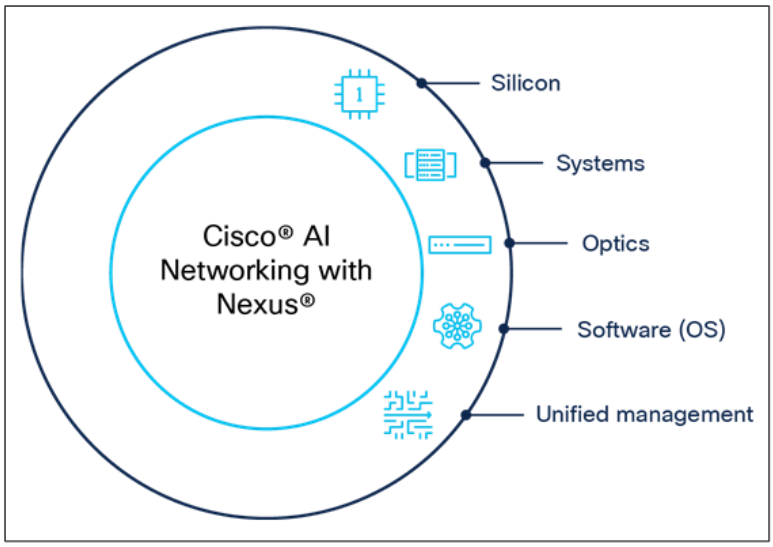

Figure 2.
Foundations of Cisco AI Networking with Nexus 9000 and Nexus Dashboard

Let's examine how each of these pillars – silicon, systems, optics, software (OS), and unified management – help Cisco differentiate for powering AI Infrastructure.

Silicon

Custom silicon is the cornerstone of Cisco's competitive edge, driving unmatched innovation and differentiation. It enables groundbreaking advancements such as flexible forwarding tiles, intelligent buffering, and advanced load-balancing capabilities, setting Cisco apart in the industry. Custom silicon embodies seamless integration across hardware and software. At the same time, it delivers a unified ecosystem for sales, marketing, and support to accelerate the value for customers.

Cisco Nexus 9300-GX2A and Nexus 9300-GX2B switches are based on Cisco's Cloud Scale ASIC, ideal for existing customers who want to get started with building AI infrastructure.

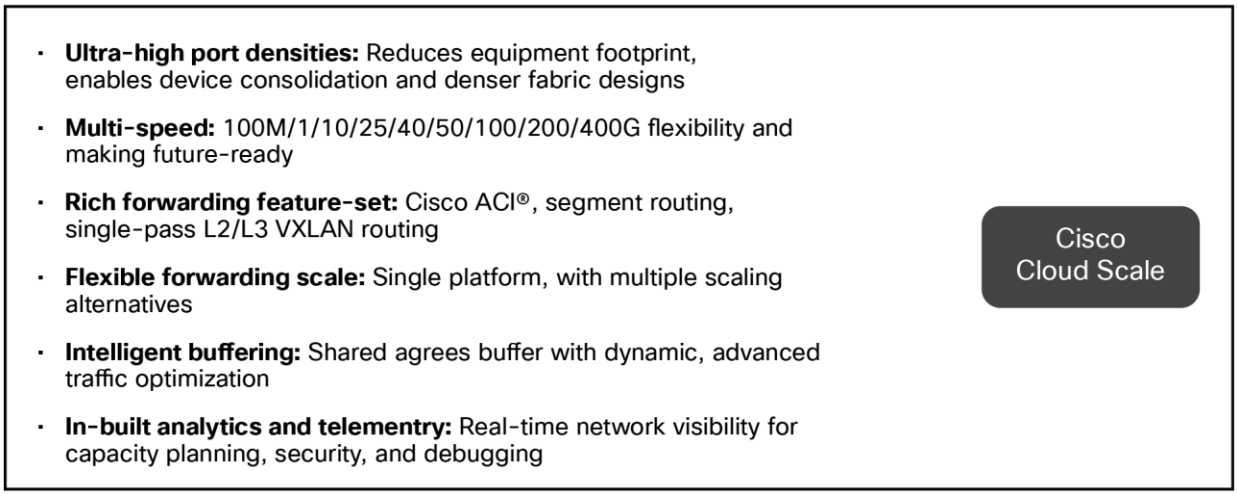

Figure 3.
Cisco Cloud Scale ASIC family

Cisco Nexus 9364E-SG2 switches based on custom Silicon One G200 ASIC are uniquely optimized for all AI workloads and provide flexible 800G connectivity options.

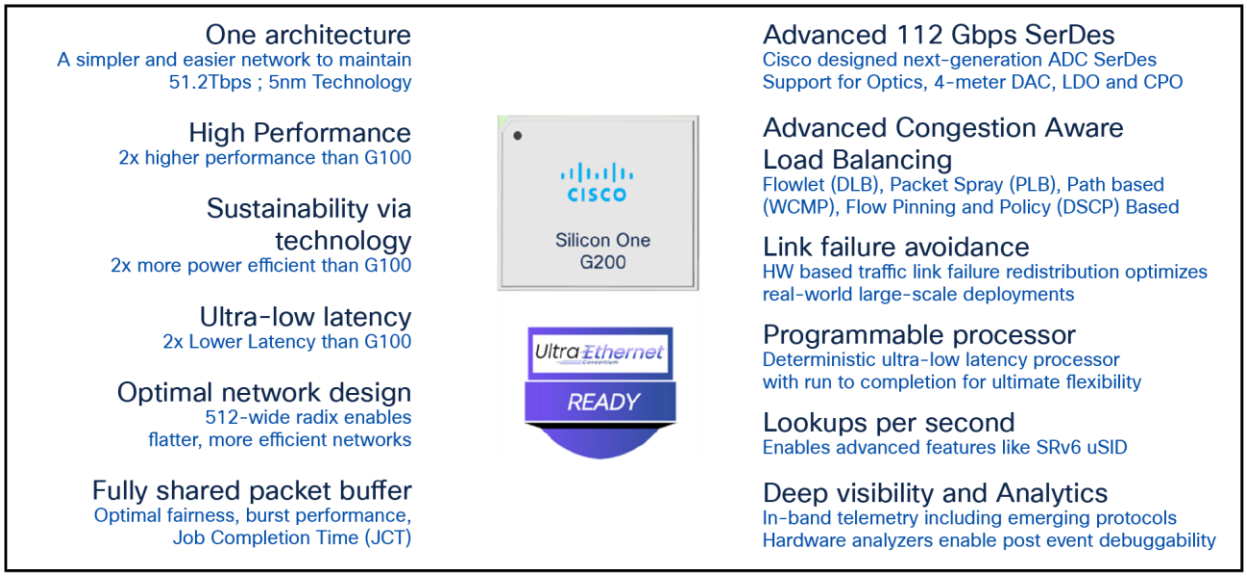

Figure 4.
Cisco Silicon One G200 - uniquely efficient and optimized for AI

By tightly aligning hardware design with software innovation, strategic vision, and serviceability, Cisco's custom silicon transforms products into powerful platforms for competitive differentiation, long-term investment protection, and market leadership.

ASIC family	Ideal use case	Platform	Key differentiator
Silicon One	800G scale-out training, inferencing	9364E-SG2	High radix, shared buffer
Cloud Scale	Entry/expansion	9300-GX2A/B	Legacy customer alignment

Systems

Switching platforms

By leveraging validated reference topologies, users can effortlessly construct both frontend and backend networks with Cisco Nexus 9000 Series Switches to support a diverse range of AI use cases, including training and inference.

Nexus 9364E-SG2 series switches

Figure 5.
Nexus 9364E-SG2-Q and Nexus 9364E-SG2-O

The Cisco Nexus 9364E-SG2 is a 2-Rack-Unit (2RU) fixed-port switch featuring 64 ports of 800 Gigabit Ethernet (GbE). The switch supports both QSFP-DD and OSFP optical module form factors, ensuring compatibility with high-speed data-center optical infrastructures. It supports flexible configurations, including 128 ports of 400GbE or 512 ports of 100/50/25/10GE ports, accommodating diverse AI cluster requirements.

The switch is powered by the Cisco Silicon One G200 chip, a 51.2 Tbps ASIC with a high-radix architecture supporting 512 x 100G Ethernet ports, resulting in significant energy and rack-space savings.

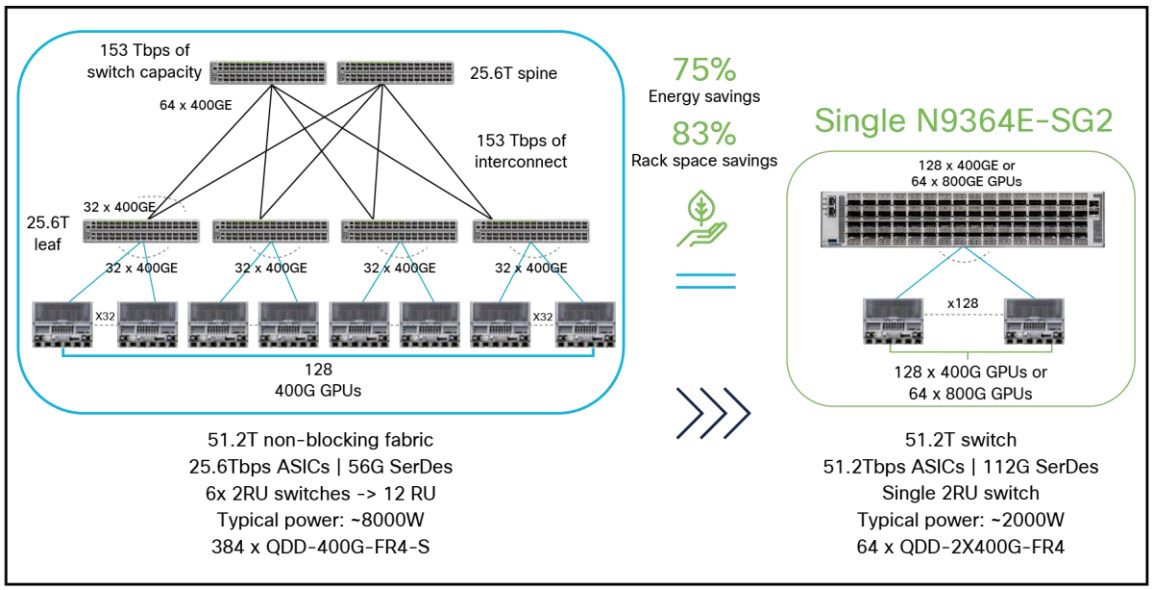

Figure 6.
Savings and benefits of having a higher radix switch

The switch provides a 256MB fully shared packet buffer, enabling consistent traffic performance, enhanced burst absorption, and minimal packet loss. This large buffer capacity is critical for handling the bursty traffic patterns of AI workloads, ensuring low-latency data transfer between GPU nodes.

More details about the switch can be found in the data sheet.

Nexus 9300-GX2 series switches

Figure 7.
Nexus 9364D-GX2A

The Cisco Nexus 9364D-GX2A is a 2-Rack-Unit (2RU) switch that supports 25.6 Tbps of bandwidth and 8.35 bpps across 64 fixed 400G QSFP-DD ports and 2 fixed 1/10G SFP+ ports. QSFP-DD ports also support native 200G (QSFP56), 100G (QSFP28), and 40G (QSFP+). Each port can also support 4x 10G, 4x 25G, 4x 50G, 4x 100G, and 2x 200G breakouts. The first 16 ports, marked in green, are capable of wire-rate MACsec encryption.

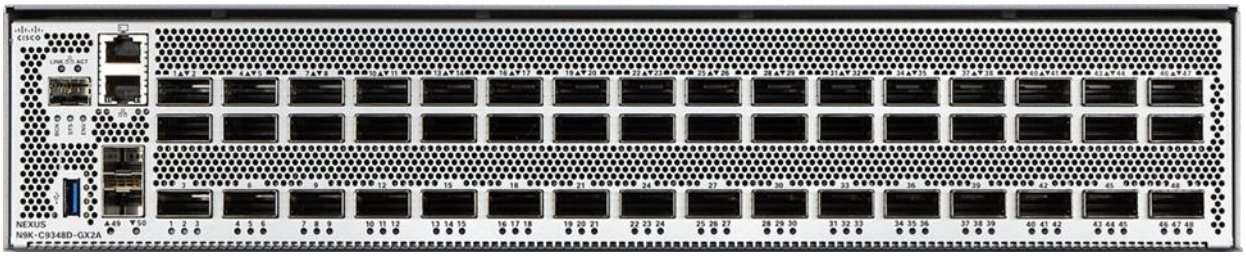

Figure 8.
Nexus 9348D-GX2A

The Cisco Nexus 9348D-GX2A is a 2-Rack-Unit (2RU) switch that supports 19.2 Tbps of bandwidth and 8.35 bpps across 48 fixed 400G QSFP-DD ports and 2 fixed 1/10G SFP+ ports. QSFP-DD ports also support native 200G (QSFP56), 100G (QSFP28), and 40G (QSFP+). Each port can also support 4x 10G, 4x 25G, 4x 50G, 4x 100G, and 2x 200G breakouts. All 48 ports are capable of wire-rate MACsec encryption.

Figure 9.
Nexus 9332D-GX2B

The Cisco Nexus 9332D-GX2B is a compact form-factor 1-Rack-Unit (1RU) switch that supports 12.8 Tbps of bandwidth and 4.17 bpps across 32 fixed 400G QSFP-DD ports and 2 fixed 1/10G SFP+ ports. QSFP-DD ports also support native 200G (QSFP56), 100G (QSFP28), and 40G (QSFP+). Each port can also support 4x 10G, 4x 25G, 4x 50G, 4x 100G, and 2x 200G breakouts. The last 8 ports, marked in green, are capable of wire-rate MACsec encryption.

The Nexus 9300-GX2 series switches implement a 120MB shared-memory egress buffered architecture, enabling consistent traffic performance, enhanced burst absorption, and minimal packet loss. This large buffer capacity is critical for handling the bursty traffic patterns of AI workloads, ensuring low-latency and lossless data transfer between GPU nodes.

More details about the switch can be found in the data sheet.

Optics

Cisco Optics stand as a leader in the optical networking industry, offering an unparalleled breadth of solutions that cater to diverse network needs. With a portfolio spanning speeds from 1G to 800G, Cisco provides high-performance optical transceivers designed for data centers, service providers, and enterprise networks.

Figure 10.
Cisco Optics transceiver module

These solutions are rigorously tested for quality and reliability, ensuring seamless interoperability across Cisco and alternative platforms. Cisco's leadership is further reinforced by its commitment to innovation, as evidenced by its role in driving industry standards and its investments in cutting-edge technologies such as silicon photonics.

Cisco Silicon One includes a unique (serializer/deserializer) component that is used to convert parallel data into serial data and vice versa. Its unique capabilities allow for the potential use of lower-power Linear Pluggable Optics (LPO) or passive Direct-Attach Copper Cables (DAC), thereby reducing power consumption for a sustainable AI cluster.

Cisco's 400G BiDi optics provide the high-bandwidth, low-latency, and scalable networking infrastructure crucial for demanding AI workloads. This technology enables efficient data transfer for the massive datasets generated by AI, optimizes performance by minimizing network bottlenecks, and allows for cost-effective network upgrades to support growing AI compute demands.

By delivering unmatched scalability, sustainability, and support, Cisco Optics empower organizations to meet the growing demands of modern networking with confidence. The Cisco Optics-to-Device Compatibility Matrix provides details on the optical modules available and the minimum software release required for each supported optical module.

Thus, there is indeed a difference between Cisco and alternative transceiver modules.

Software (OS)

Cisco NX-OS is a purpose-built data-center operating system enabling organizations to build AI-ready infrastructure with the unparalleled performance, reliability, and scalability that is needed for the most demanding AI applications. Combined with Cisco's broader data-center-class innovations, it delivers end-end operational excellence. Cisco NX-OS supports multitenancy capability to securely segment and manage multiple AI workloads on shared infrastructure, ensuring isolation between different users or applications.

Lossless fabric for AI applications

Cisco Nexus 9000 Series Switches running NX-OS support a variety of features optimized for AI fabrics. On the one hand, reactive congestion management approaches include Explicit Congestion Notification (ECN) to mark packets and signal congestion without dropping them, Weight Random Early Detection (WRED) to assign drop probabilities for different traffic classes, and Priority Flow Control (PFC) to ensure lossless Ethernet by pausing traffic during congestion. On the other hand, proactive congestion management addresses non-uniform utilization of inter-switch links leading to congestion, by using flexible mechanisms such as Cisco Intelligent Packet Flow tailored to customer-specific requirements.

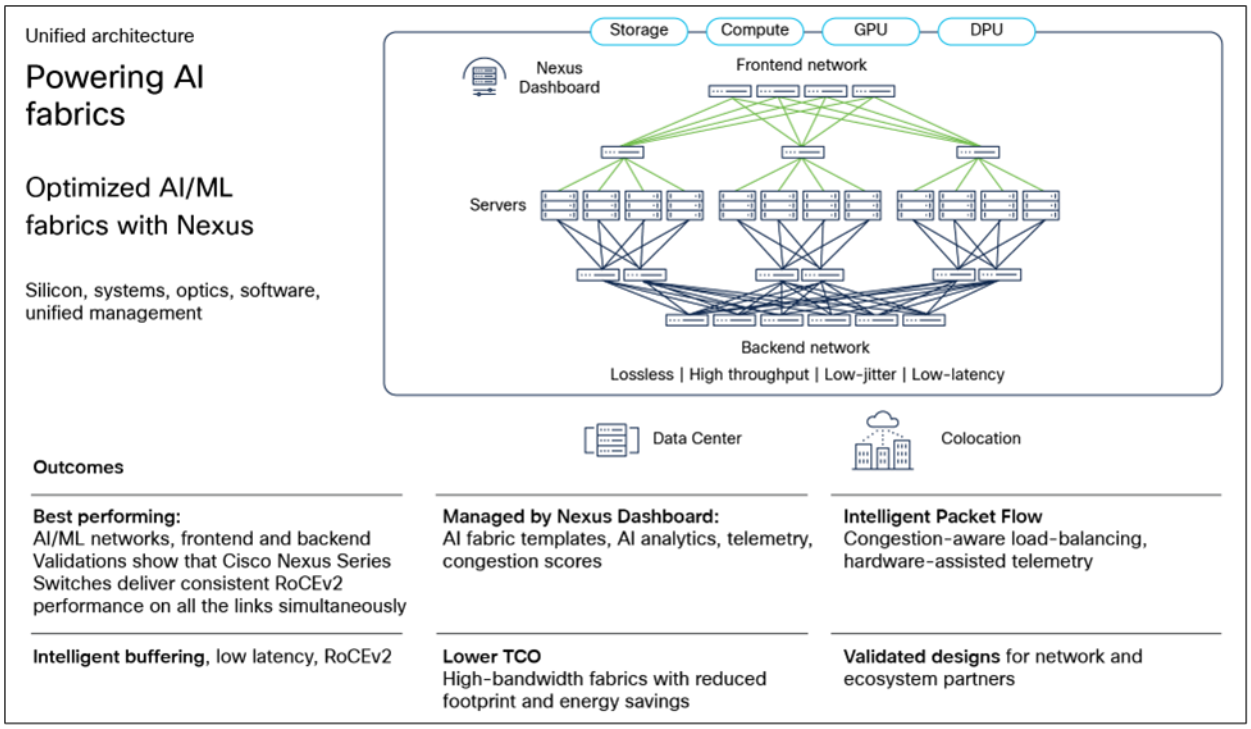

Figure 11.
Cisco Nexus 9000 Switches for networking AI clusters

Let's briefly examine each of the features supported by Cisco Nexus 9000 Series Switches running NX-OS to power AI-networking infrastructure:

RDMA over Converged Ethernet (RoCEv2)

GPUs connected through a network leverage InfiniBand (IB) verb APIs to initiate RDMA operations. During this process, data destined for another GPU is segmented into multiple payloads. Each payload is then encapsulated with Ethernet, Internet Protocol (IP), and User Datagram Protocol (UDP) headers before being transmitted using the network's standard forwarding mechanisms. This method of executing RDMA operations over an Ethernet network with UDP/IP encapsulation is known as RDMA over Converged Ethernet version 2 (RoCEv2). In this way, RoCEv2 enables direct memory access, reducing CPU overhead and latency in GPU-to-GPU communication.

Figure 12.
RoCEv2 frame format with the RoCEv2 IP and UDP header on top of Ethernet

Reactive congestion management approach

Explicit congestion notification

Explicit Congestion Notification (ECN) and Weighted Random Early Detection (WRED) are network congestion management mechanisms essential for AI backend networks. ECN marks packets with a congestion flag, signaling endpoints to reduce their transmission rates, while WRED proactively detects and reacts to congestion in the network by marking traffic that may contribute to it with ECN bits. Combined, ECN and WRED enhance network efficiency for critical communication between AI GPU clusters.

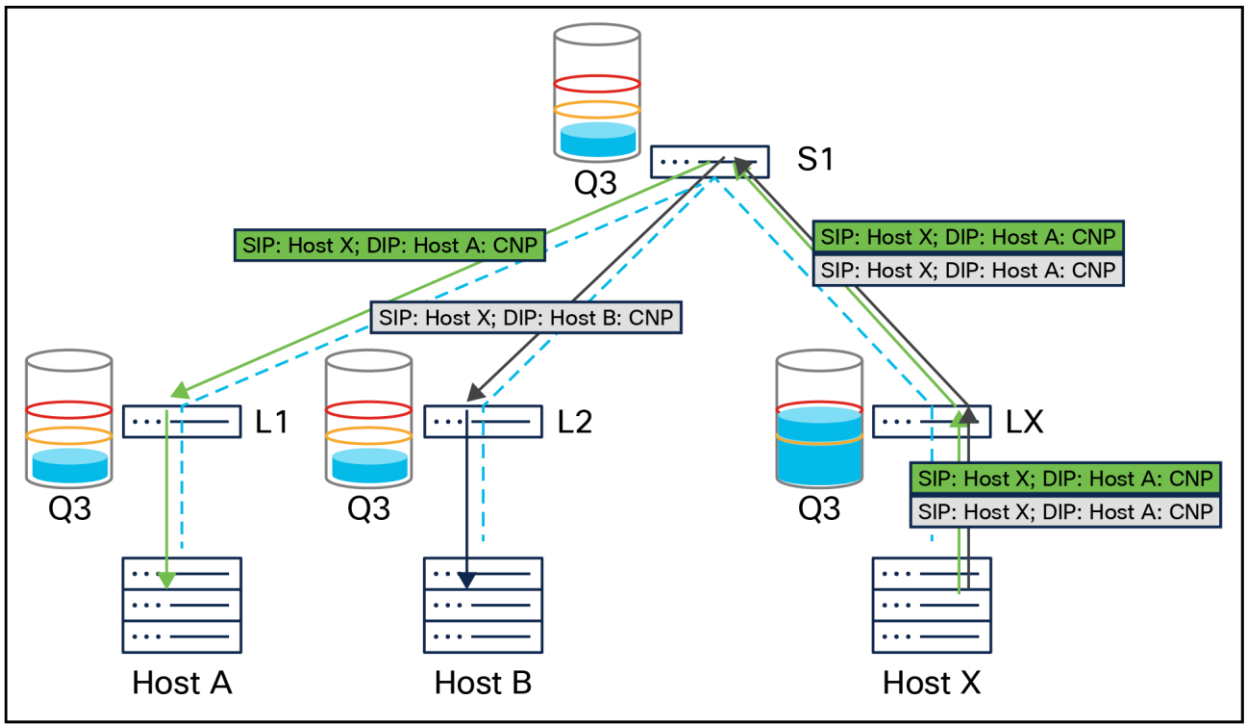

Figure 13.
Host X informs hosts A and B about network congestion by sending them CNP packets

Priority flow control

Priority Flow Control (PFC) ensures lossless transmission for specific traffic classes in AI backend networks by pausing traffic during congestion. PFC operates on individual priorities, allowing selective pausing of traffic classes while permitting others to continue. The PFC watchdog detects and mitigates PFC pause storms, where excessive pause frames stall queues and potentially halt network traffic.

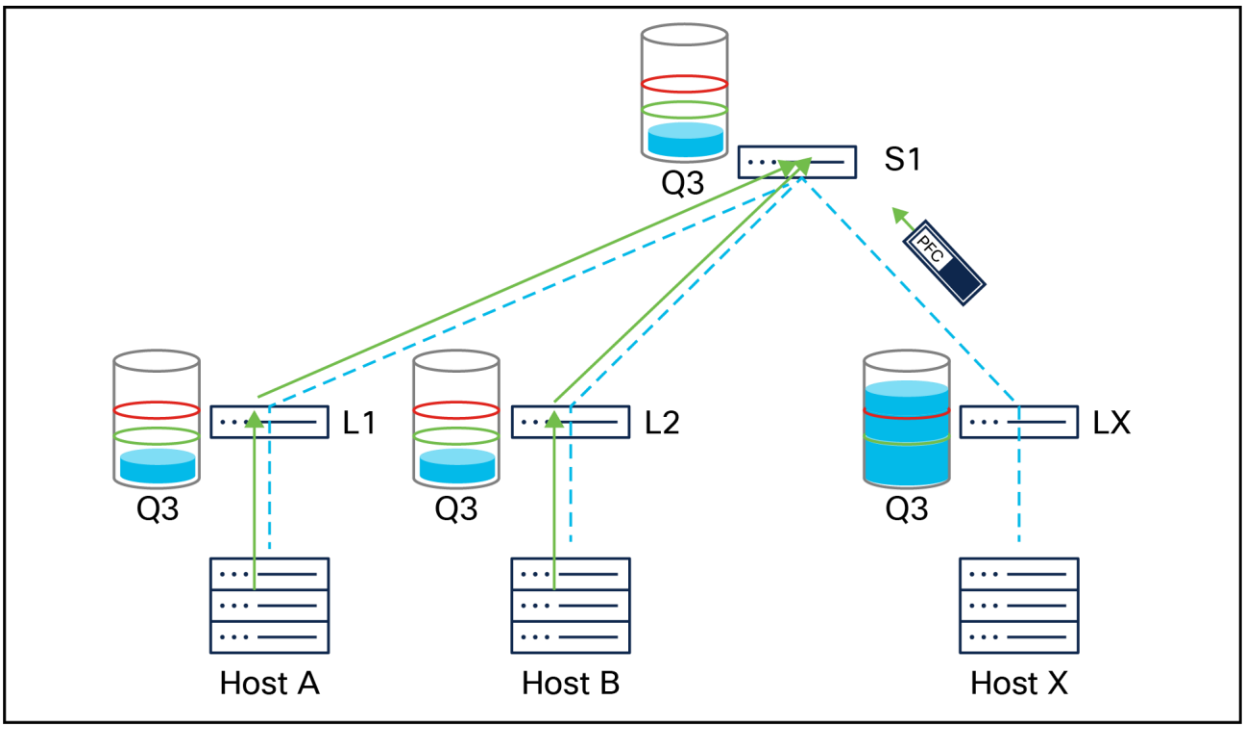

Figure 14.
PFC signaled from leaf X (LX) to spine 1 (S1) , pausing traffic from the spine switch to the leaf switch

Data-center quantized congestion notification

ECN and PFC effectively manage congestion. ECN reacts first to mitigate congestion, while PFC acts as a fail-safe to prevent traffic drops if ECN is insufficient. This collaborative process is known as Data-Center Quantized Congestion Notification (DCQCN) supported by Cisco Nexus 9000 Series Switches. Together, PFC and ECN enable efficient end-to-end congestion management. In cases of minor congestion with moderate buffer usage, WRED with ECN manages congestion seamlessly. For severe congestion or microburst-induced high buffer utilization, PFC takes over.

Approximate fair drop

Cisco Nexus 9000 Series Switches also support a distinctive capability called Approximate Fair Drop (AFD), which helps manage congestion by distinguishing high-bandwidth “elephant flows” from short-lived, low-bandwidth “mice flows.” AFD marks ECN bits (0x11) only for elephant flows, with the number of marked packets proportional to the flow bandwidth (for example, fewer for 1G flows, more for 10G flows). This targeted marking enables end-host algorithms to efficiently slow down the high-bandwidth flows, contributing most to congestion while preserving low-latency performance for smaller flows.

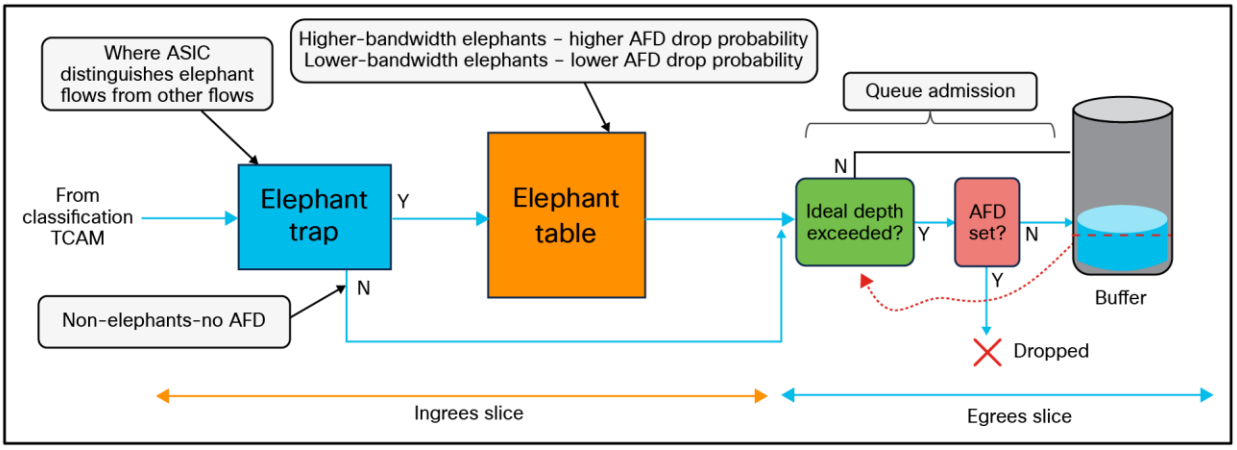

Figure 15.
Approximate Fair Drop (AFD)

Unlike WRED, which marks all traffic equally, AFD provides granular control, avoiding penalties for mice flows. In AI clusters, this ensures that short-lived communications complete faster, as AFD prioritizes their completion without inducing packet drops, while regulating long data transfers to minimize system-wide congestion.

Proactive congestion management approach

Challenges with traditional Equal Cost Multipath (ECMP)

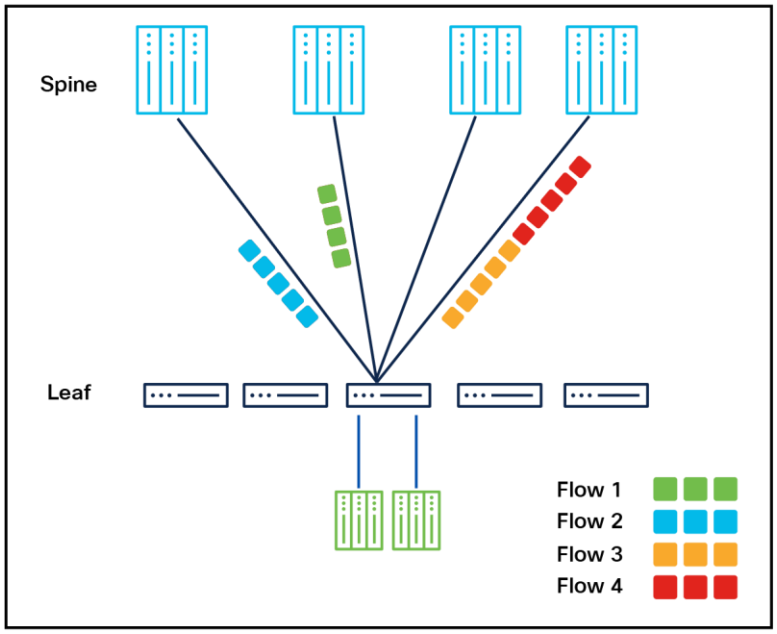

Figure 16.
ECMP-based on 5-tuple

Traditional ECMP distributes packets with the same 5-tuple (source IP, destination IP, source port, destination port, protocol) over the best path, with subsequent frames maintained on the same link, avoiding out-of-order packet delivery. However, this approach can lead to under-utilization or over-utilization of links due to low hashing entropy, and it lacks adaptability to bursty traffic. As a result, it can degrade the overall throughput efficiency and is typically not well-suited for AI fabrics.

For customers looking to address these limitations while still leveraging ECMP, Cisco Nexus 9000 Series Switches support enhancements to traditional ECMP that allow parsing of specific Base Transport Header (BTH) fields in RoCEv2 to improve entropy. These BTH fields include opcode, destination queue pair, and packet sequence number to enable more balanced traffic distribution compared to relying solely on the traditional 5-tuple information.

Cisco Intelligent Packet Flow—advanced traffic management for AI workloads

AI workloads demand high throughput, low latency, and adaptive network behavior. Traditional static approaches such as ECMP fall short in addressing the bursty, synchronized, and east-west-heavy nature of AI Traffic. This is where Cisco Intelligent Packet Flow comes into picture. It is a comprehensive traffic management framework designed to meet these challenges. It brings adaptive intelligence through real-time telemetry, congestion awareness, and fault detection dynamically steering traffic to optimize flow completion and reduce Job Completion Time (JCT).

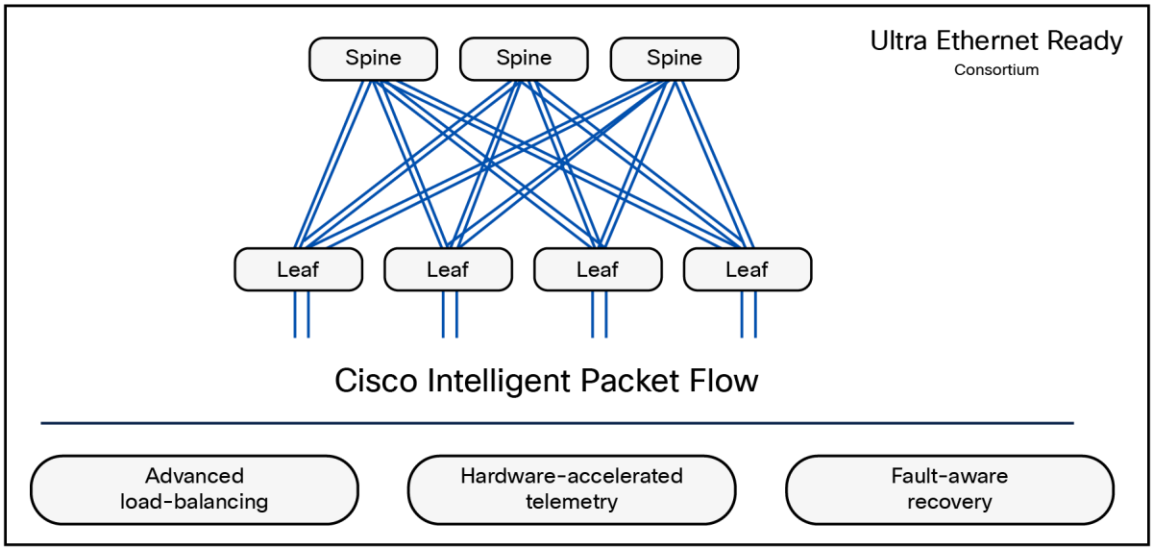

Figure 17.
Cisco Intelligent Packet Flow—advanced load balancing for AI workloads

Cisco Nexus 9000 Series Switches support multiple load balancing modes tailored for AI/ML fabrics:

- Dynamic Load Balancing (DLB) - flowlet
- Per-packet load balancing
- Weighted cost multipath + DLB
- Policy-based dynamic load balancing
- ECMP flow pinning
- Packet trimming

Dynamic load balancing (flowlet-based)

Customer use-case: Optimizing traffic distribution in real-time

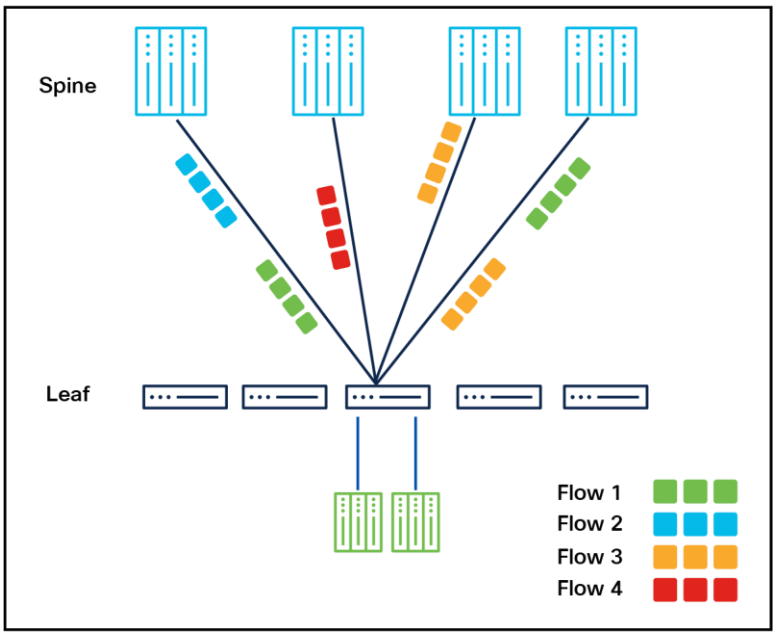

Figure 18.
Dynamic load balancing (flowlet-based)

Dynamic Load Balancing (DLB) distributes incoming traffic across all available Equal-Cost Multipath (ECMP) links by considering real-time link transmission (TX) utilization local to the switch. To handle flowlets—bursts of packets from a flow separated by gaps large enough to allow independent routing—the switch ASIC maintains a flowlet table that maps flowlets (identified by flowhash) to their output ports. When a packet arrives, the flowlet table is checked; if there is a hit, the corresponding output port is used without triggering a new link selection. If no entry exists, the DLB logic selects a randomized link with the least TX utilization and adds it to the flowlet table, which retains entries for a configurable aging time. In cases of hash collisions, where a new flow's flowhash matches an existing entry but the current port is unavailable in the ECMP paths, the system defaults to regular ECMP hashing for port selection. This mechanism ensures efficient traffic distribution, minimizes packet reordering, and maintains robust handling of edge cases like collisions.

Per-packet load balancing

Customer use-case: Enabling multipath distribution for improved efficiency.

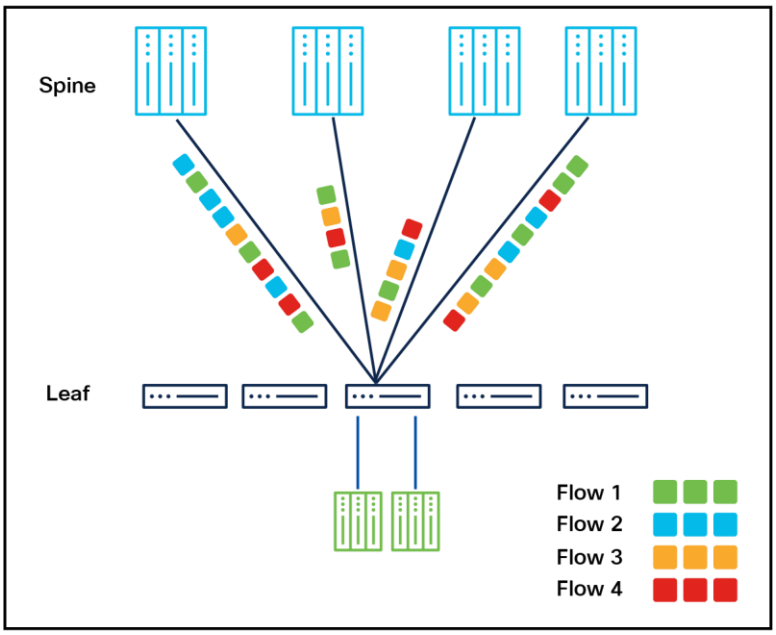

Figure 19.
Per-packet load balancing

Congestion-aware per-packet load balancing is an advanced traffic distribution mechanism designed to optimize network performance by leveraging real-time transmission (TX) link utilization metrics. This approach dynamically selects the randomized ECMP link with the lowest TX utilization for each individual packet, ensuring even traffic distribution and maximizing link utilization. Unlike flow-based load balancing, this mode performs a new port selection for every packet, which allows for rapid adjustments in traffic patterns and minimizes congestion across the network. The method is particularly effective in reducing burstiness, thereby enhancing overall network stability and performance. To enable this functionality, endpoints such as GPU NICs must support out-of-order packet reassembly, as packets may arrive at the destination on different paths. Key benefits of this approach include improved link utilization, higher overall network throughput, fast convergence during link failures, and reduced congestion. However, the dependency on endpoint capabilities for packet reordering is a critical consideration when deploying this mode.

ECMP static pinning

Customer use-case: Ensuring deterministic routing for consistent performance.

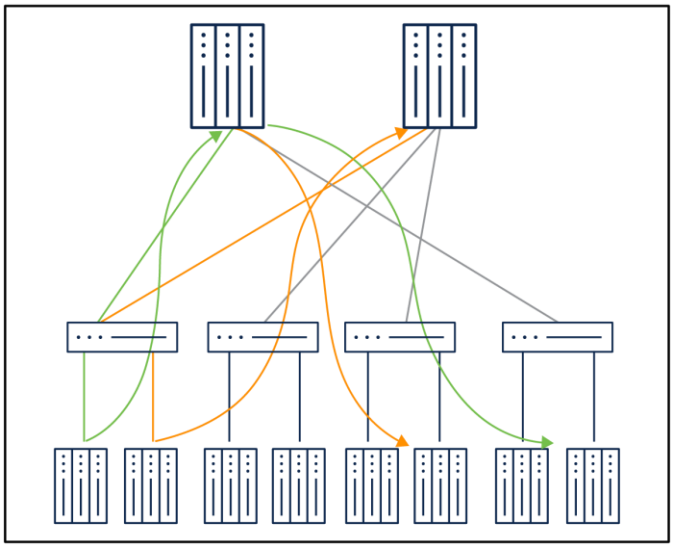

Figure 20.
ECMP static pinning

ECMP static pinning is a traffic-forwarding mechanism that enables precise control by pinning a source interface to a specific destination interface. This approach gives users the ability to explicitly manage traffic distribution on a per-switch basis, ensuring predictable and consistent forwarding behavior. In the event that the pinned port becomes unavailable, the system automatically selects the next port with the lowest link utilization, maintaining network efficiency and continuity. By eliminating over-subscription and providing deterministic routing, ECMP static pinning enhances the reliability and performance of the network. It is especially suited for use cases requiring low-latency communication.

Weighted Cost Multipath (WCMP) + Dynamic Load Balancing (DLB)

Customer use-case: Combining DLB with real-time link telemetry and path weighting for enhanced routing decisions.

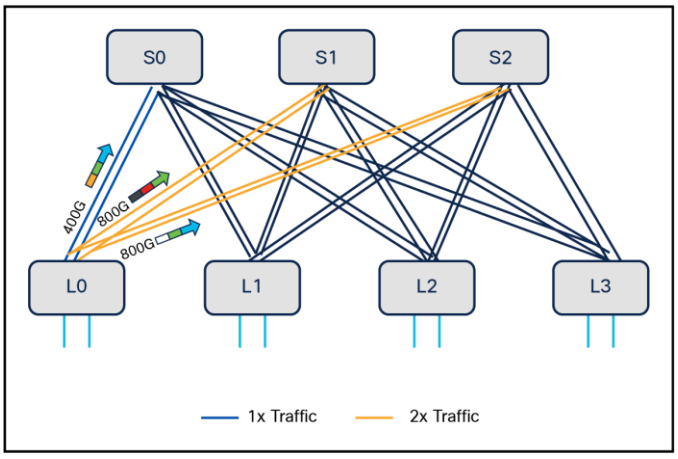

Figure 21.
Weighted Cost Multipath (WCMP) load balancing

Weighted Cost Multipath (WCMP) is an advanced traffic-distribution mechanism that allocates network traffic based on the relative bandwidth of available links. By considering the capacity of local links, WCMP ensures that traffic is load balanced proportionally to the bandwidth of each path, optimizing resource utilization. This is achieved using unique weights and ranges, which determine the distribution of traffic across multiple paths in the network. WCMP enhances capacity adoption by effectively balancing traffic loads, ensuring that no single path becomes overutilized while others remain underutilized. It leverages External Border Gateway Protocol (eBGP) to allocate bandwidth fairly per flow, further improving the efficiency and fairness of traffic distribution. This approach is particularly beneficial for networks with diverse link capacities, because it maximizes throughput while minimizing congestion.

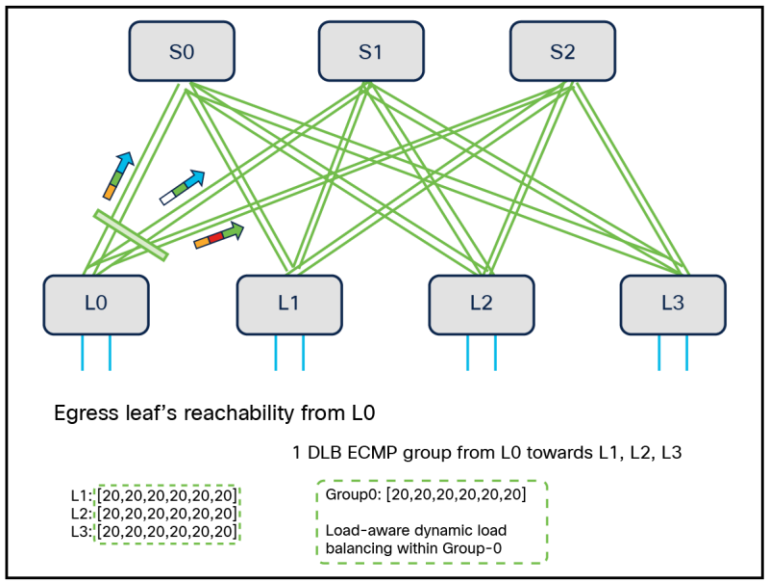

Figure 22.
Steady state reachability

In steady state, when the network is symmetric, DLB achieves the optimal bandwidth utilization within single Group0, as shown in the figure.

In case of a link failure, with WCMP+ DLB enabled, the cumulative weights advertised are adjusted. DLB is performed for the traffic toward only the members of the unaffected group. The hierarchical WCMP provides the ability to pick the spines with full bandwidth toward a leaf and the rest of the spines with partial bandwidth toward the leaf having link failure.

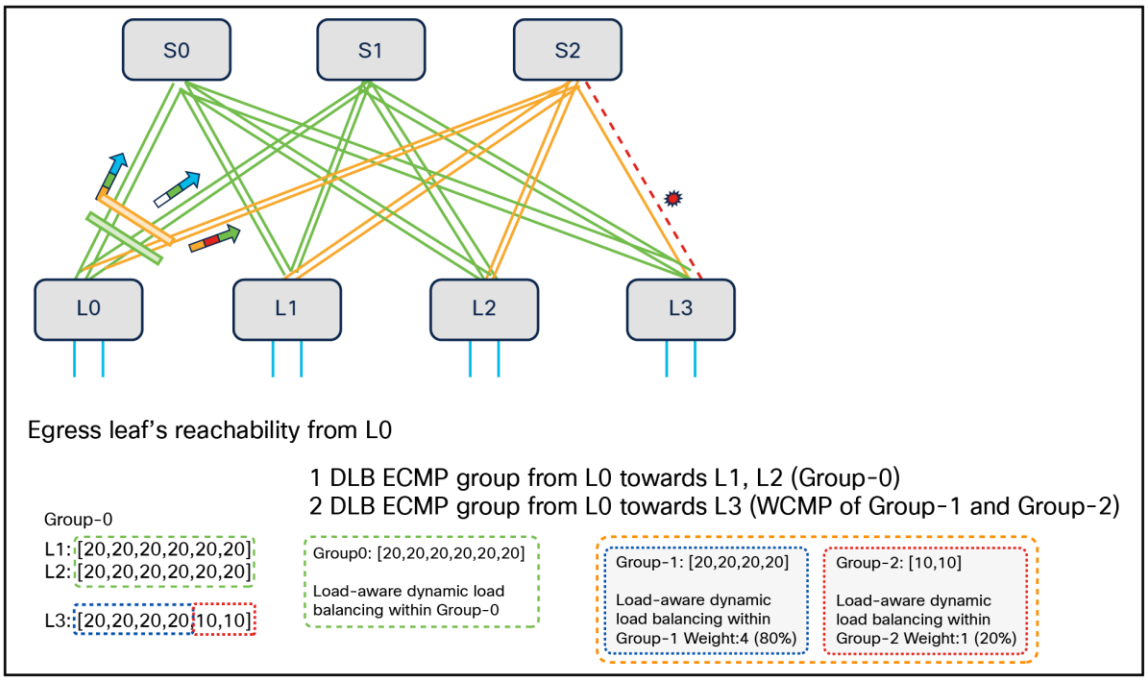

Figure 23.
Egress leaf's reachability from L0 when link fails from L3 to S2

In Figure 23, where there is a link failure between S2 and L3, DLB ECMP Group0 from L0 toward L1 and L2, marked in green, remains unaffected. DLB ECMP groups from L0 toward L3 form a WCMP group (marked in orange) of Group1 (marked in blue) and Group2 (marked in red), where Group1 gets 80 percent of the traffic and Group2 gets 20 percent of the traffic. This is a differentiating capability with Cisco, where a customer can achieve minimal impact in case of link failures even if it must satisfy the requirements of having multiple links from leaf to spine and maintaining uniform traffic distribution.

Policy-driven dynamic load balancing

Customer use-case: Utilizing ACLs, DSCP tags, or RoCEv2 header fields for traffic prioritization.

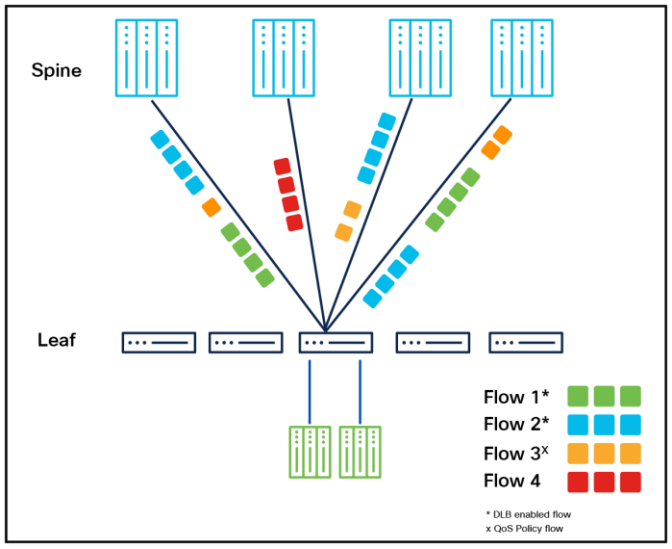

Figure 24.
Policy-based dynamic load balancing

This capability introduces a highly flexible traffic management approach that allows dynamic switching between flowlet-based and per-packet load balancing. By default, flowlet-based dynamic load balancing (DLB) is applied to DLB traffic, whereas regular traffic follows ECMP-based forwarding. A QoS policy can override the default DLB behavior, using match criteria such as DSCP or ACL-based rules to enable per-packet load balancing for specific traffic flows. With the policy override enabled, new port selection occurs at the flowlet boundary for regular DLB traffic and at the per-packet boundary for policy-matched traffic. This flexibility enables coexistence with regular ECMP traffic while optimizing network performance for high-priority or congestion-sensitive flows. For per-packet flows, GPU NICs or endpoints must support out-of-order packet reassembly to ensure proper delivery. This approach provides the most adaptable DLB option, supporting up to 256 ECMP groups and enabling scalable, congestion-aware traffic distribution tailored to specific application requirements.

Packet trimming

Customer use-case: Reducing packet sizes instead of dropping them to maintain data integrity.

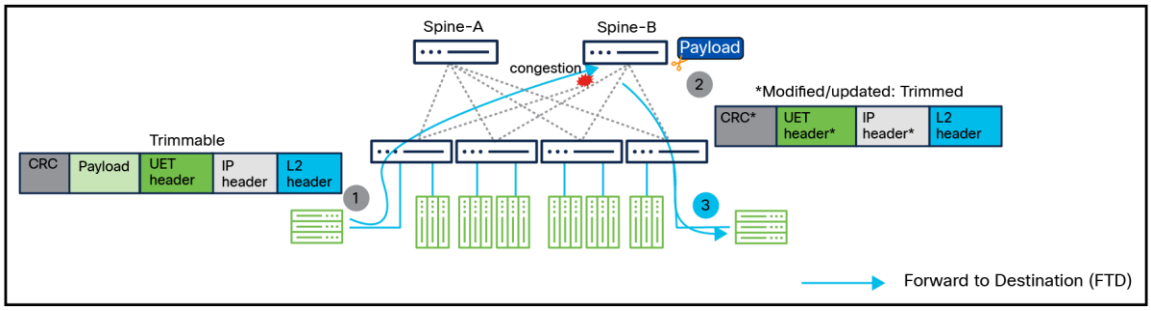

Figure 25.
Packet trimming based on UEC 1.0 specification

Packet trimming is a technique that, instead of discarding packets during switch-buffer congestion, trims them to a small size and forwards them with higher priority to the destination. This functionality is applied exclusively to valid IP packets that fail buffer admission checks, ensuring that only eligible traffic is processed. Packet trimming occurs at the point of buffer admission, typically after ingress processing has validated the L2 frame and L3 header and a valid egress port is identified through next-hop lookup; however, its implementation may vary depending on the switch architecture. The feature is particularly effective for transports capable of interpreting trimmed packets to trigger fast retransmissions of the original data, minimizing latency and improving congestion recovery. Packet trimming relies on Differentiated Service Codepoints (DSCPs) configured by network operators, with at least two categories: TRIMMABLE, for packets eligible for trimming, and TRIMMED, for packets already processed. Only packets marked with TRIMMABLE DSCPs can undergo trimming, and each category must include at least one DSCP value (for example, {1,2,3} for TRIMMABLE and {4,5,6} for TRIMMED). By leveraging these DSCP categories, it ensures selective and controlled trimming, preserving critical network performance while reducing retransmission overhead. This mechanism significantly improves packet-loss detection and recovery for congestion-sensitive applications.

Cisco Intelligent Packet Flow—hardware-accelerated telemetry

Cisco Intelligent Packet Flow delivers advanced telemetry features, including microburst detection, congestion signaling, tail timestamping, and In-Band Network Telemetry (INT) for granular flow visibility. These capabilities provide network operators with real-time insights into network behavior, allowing them to proactively manage congestion and optimize performance, even under rapidly changing AI workload demands.

Cisco Intelligent Packet Flow—fault-aware recovery

For large-scale AI environments, Cisco enhances network reliability through real-time fault detection, intelligent rerouting around degraded paths, and rapid convergence to prevent performance bottlenecks. This autonomous recovery capability enables consistent, resilient, and efficient operation across AI training, inference, and data movement workloads.

These latest enhancements, along with previous innovations, establish Cisco Nexus 9000 Series switches as the intelligent foundation for next-generation AI and UEC-ready networks, precisely balancing every flow and path. With Cisco Intelligent Packet Flow, networks can adapt in real time to AI workloads, and ongoing advancements will ensure they keep pace with the evolving needs of modern data centers.

Unified management

Cisco Nexus Dashboard, included with every Cisco Nexus 9000 switch tiered licensing purchase, serves as a centralized hub that unifies network configurations and visibility across multiple switches and data centers. For AI fabric operations, it acts as the central command center, enabling everything from AI fabric automation to continuous analytics, all within a few clicks.

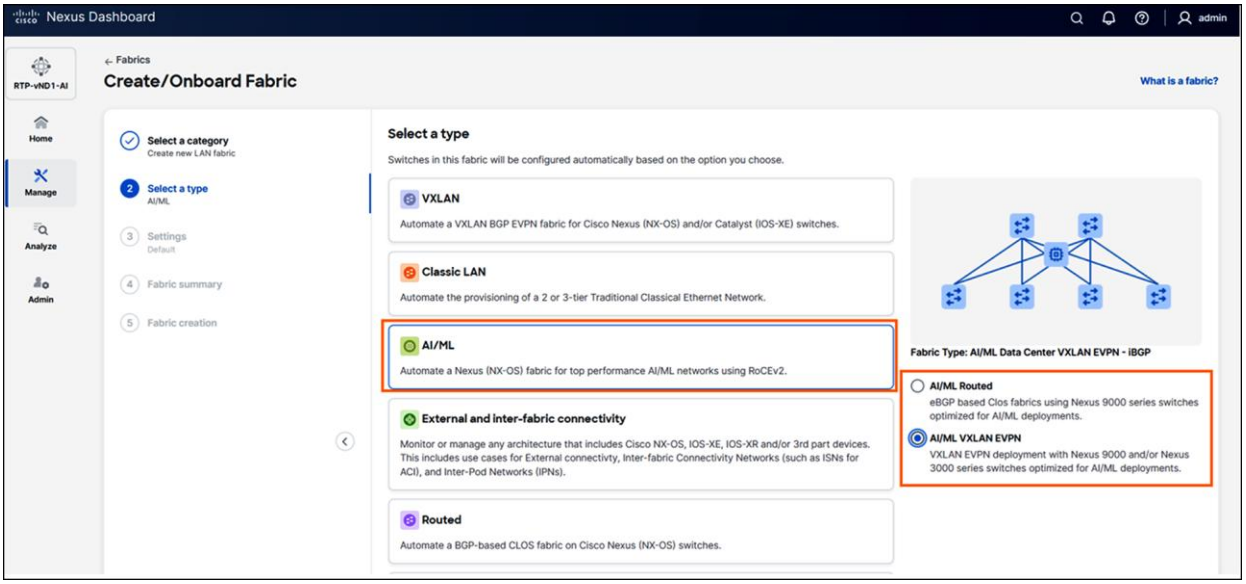

Figure 26.
AI fabric workflow on Cisco Nexus Dashboard

Key capabilities, such as congestion scoring, PFC/ECN statistics, and microburst detection, empower organizations to proactively identify and address performance bottlenecks for their AI backend infrastructure.

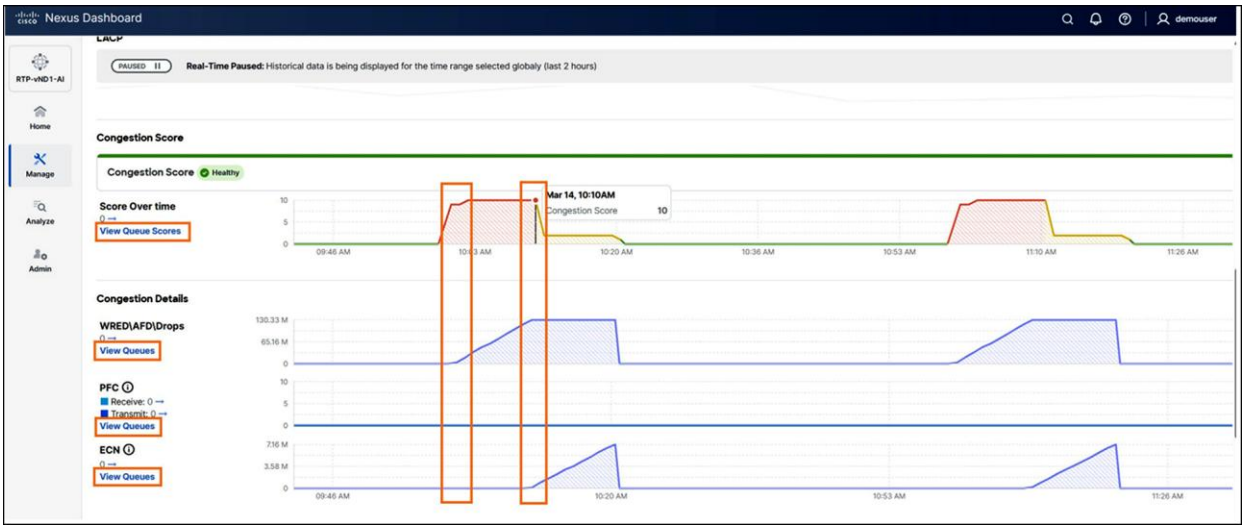

Figure 27.
Congestion Score and Congestion Details on Cisco Nexus Dashboard

Advanced features, such as anomaly detection, event correlation, and suggested remediation, ensure that networks are not only resilient but also self-healing, minimizing downtime and accelerating issue resolution.

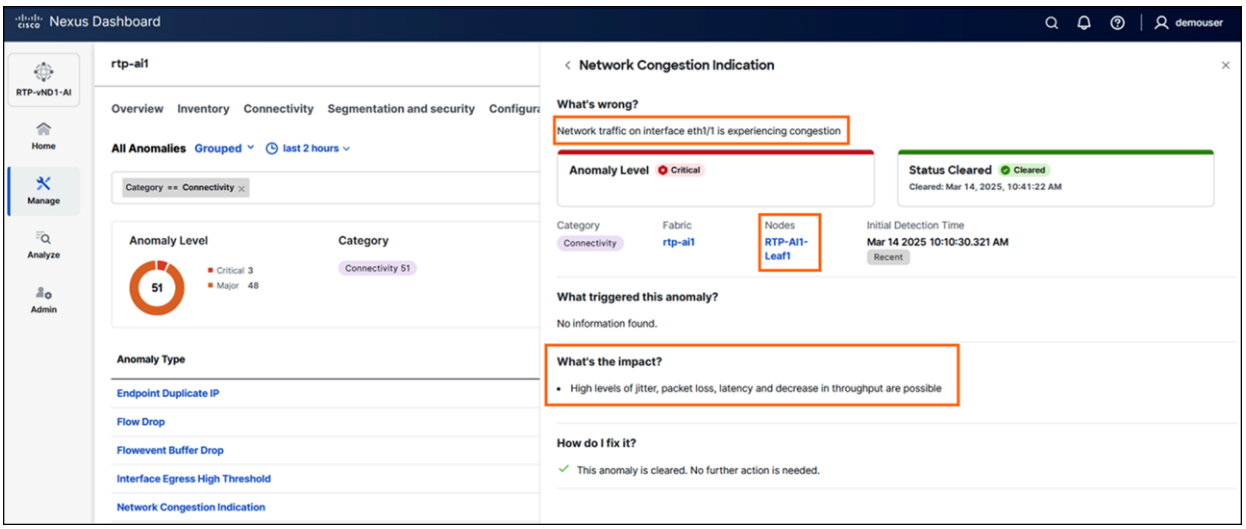

Figure 28.
Anomaly detection on Cisco Nexus Dashboard

AI job observability enables customers to get end-to-end visibility into AI workloads across the entire stack. It allows customers to monitor the network, NICs, GPUs, and distributed compute nodes in real time. On top of it, you can get capabilities such as topology-aware visualization, real-time metrics, and proactive troubleshooting for AI jobs, which results in providing comprehensive, actionable insights into both infrastructure and AI job health which is coming soon on Nexus Dashboard.

Figure 29.
Visibility into AI jobs, GPUs, and NICs on Cisco Nexus Dashboard (coming soon)

With the growing demand for AI networking infrastructure, optimizing energy efficiency is crucial, and Cisco Nexus Dashboard delivers actionable insights into network performance, energy consumption, costs, and emissions, along with recommendations for better resource allocation and system optimization to maximize uptime and reduce environmental impact.

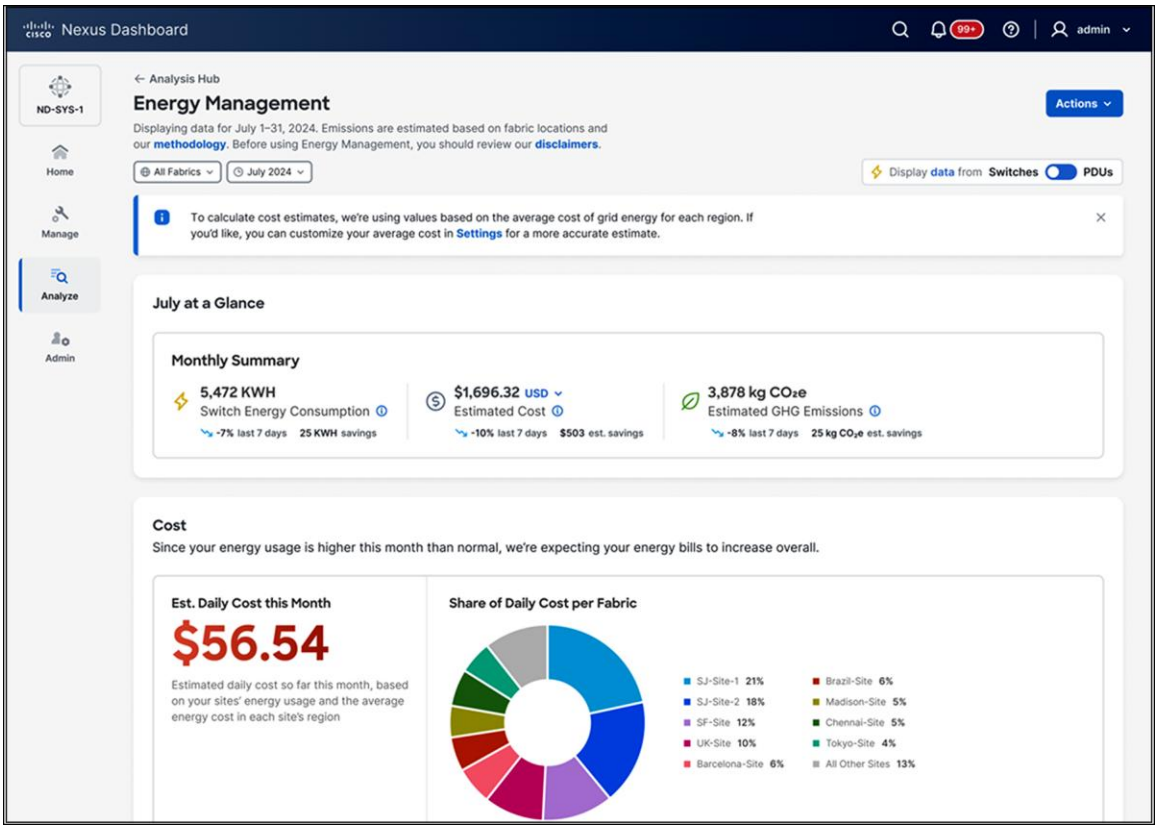

Figure 30.
Sustainability insights on Cisco Nexus Dashboard

Built to handle the high demands of AI workloads, Cisco Nexus Dashboard transforms network operations into a seamless, data-driven experience, unlocking the full potential of AI fabrics.

Summary

Cisco's unparalleled expertise encompasses the fundamental pillars of networking: silicon, systems, optics, software (OS), and unified management. These innovations empower Cisco to provide comprehensive end-to-end solutions that redefine performance, efficiency, and scalability. With a deep understanding of systems, Cisco seamlessly integrates hardware and software to create intelligent and adaptable platforms for AI networking infrastructure, and that are validated to provide exceptional performance. Our investment in the development of custom silicon drives cutting-edge capabilities and long-term investment protection for customers. Furthermore, Cisco's leadership in optics facilitates ultra-fast and high-bandwidth connectivity, meeting the increasing demands of AI workloads. Cisco Nexus Dashboard provides a single command center to orchestrate AI fabrics and deliver end-end analytics for faster time to issue resolution. By partnering with a vendor that controls their own destiny, customers can be rest assured that they will remain agile, resilient, and capable of supporting future growth and challenges. These disciplines collectively position Cisco as a technology leader, delivering holistic AI infrastructure solutions that establish the benchmark for the future of networking.

Glossary

- RDMA – Remote Direct Memory Access
- RoCE – RDMA over Converged Ethernet
- QoS – Quality of Service
- ECN – Explicit Congestion Notification
- PFC – Priority Flow Control
- DCQCN – Data-Center Quantized Congestion Notification
- ECMP – Equal Cost Multipath
- DLB – Dynamic Load Balancing
- WCMP – Weight Cost Multipath
- BGP – Border Gateway Protocol
- JCT – Job Completion Time
- ACL – Access Control List
- VRF – Virtual Routing and Forwarding
- VXLAN – Virtual extensible Local Area Network
- UEC – Ultra Ethernet Consortium

References

White papers:

- Cisco Nexus 9000 Series Switches for AI Clusters—Performance Validation Insights
- Cisco Data Center Networking Solutions: Addressing the Challenges of AI Infrastructure
- Cisco Data Center Networking Blueprint for AI Applications
- Cisco Validated Design for Data Center Networking Blueprint for AI Applications
- Cisco Massively Scalable Data Center Network Fabric Design and Operation White Paper

Enterprise reference architecture:

- AI Infrastructure with Cisco Nexus 9000 Switches Data Sheet
- NVIDIA Certified Cisco Nexus Hyperfabric AI Enterprise Reference Architecture

Cisco Blogs:

- Uncompromised Ethernet: Performance and Benchmarking for AI Fabric
- Redefine Your Data Center with New Cisco N9300 Series Smart Switch Innovations
- Embracing the AI Era: Cisco Secure AI Factory with NVIDIA

Solution briefs:

- Cisco Secure AI Factory with NVIDIA Solution Overview
- VAST Cisco Nexus 9000 Switch Support
- Lenovo Integrated AI Solutions: Solution Overview
- AI Infrastructure for the Agentic Era

dCloud demo:

- **Cisco AI-Ready Data Center with Cisco Nexus 9000 and Nexus Dashboard (requires Cisco account)** https://dcloud2-sjc.cisco.com/content/instantdemo/cisco_ai_ready_with_nexus

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **https://www.cisco.com/go/offices**.

Printed in USA C11-5186166-01 07/25