

Cisco and Hyve AI Networking Solution



Cisco and Hyve AI Networking solution overview

AI workloads present unique networking challenges that traditional data-center architectures cannot adequately address. The synchronized, bursty nature of AI traffic patterns, combined with the need for massive east/west data movement between GPU clusters, requires purpose-built networking solutions.

The Cisco and Hyve AI Networking Solution delivers a comprehensive approach to these challenges, combining advanced switching hardware, intelligent software capabilities, and unified management platforms. This integrated rack solution enables organizations to minimize the network footprint and maximize space and power for AI clusters while building scalable,

high-performance AI infrastructure that accelerates time-to-value while maintaining operational efficiency and robust security.

This joint switching solution has been specially designed for building AI training/inference clusters and edge aggregation racks that require a high-bandwidth, low-latency fabric.

Trends and challenges

Artificial intelligence is fundamentally transforming data-center networking requirements and driving unprecedented demand for high-bandwidth, low-latency, and lossless network performance. The exponential growth of AI workloads—from training massive language models to executing real-time inference and generative AI applications—necessitates a new class of data-center networking infrastructure specifically designed for accelerated computing environments.

Unlike traditional networking architectures that have served enterprises for decades, AI networking requires specialized infrastructure optimized for GPU clusters running multibillion to trillion parameter-scale models. These workloads

demand innovative software architectures, distributed networking solutions, and specialized processors including SmartNICs, IPUs, and DPUs to deliver the lightning-fast data movement, ultra-low latency, and near-zero packet loss essential for AI success.

The performance demanded for connecting chips within supercomputers, interconnecting servers in AI clusters, or linking those clusters to the network edge is unparalleled. With, training and now Inference taking center stage, AI networking demands higher capacity (scaling to 400 Gbps and 800 Gbps and beyond), higher throughput, lower latency, high reliability, faster access to storage, and optimized clustering.

Market trends and technology evolution

- **Transition to higher network speeds:** The industry is experiencing rapid adoption of 800G and early deployment of 1.6T networking technologies. Organizations are upgrading from 400G to 800G fabrics, with research and development clusters beginning to implement 1.6T solutions. This evolution may simplify some architectural challenges with flatter networks while dramatically increasing throughput density, enabling more efficient data-center designs.
- **Widespread adoption of RoCEv2:** Remote direct memory access over Converged Ethernet version 2 (RoCEv2) has become

the standard for AI networking, enabling direct memory access between GPU devices through an Ethernet infrastructure. This technology significantly reduces CPU utilization and latency, making it ideally suited for AI-workload communication patterns.

- **Implementation of lossless Ethernet:** Modern AI networks require lossless Ethernet capabilities implemented through advanced switching technologies and topologies. These include sophisticated flow-control mechanisms, enhanced congestion management, optimized hashing algorithms, intelligent buffering strategies, and comprehensive flow telemetry—capabilities that extend far beyond traditional switching infrastructure.
- **Ultra Ethernet Consortium leadership:** The Ultra Ethernet Consortium (UEC), established in July 2023, was created to scale up Ethernet and make it more suited for AI and HPC workloads. It represents a collaborative industry effort to enhance Ethernet capabilities for AI and high-performance computing workloads. This initiative focuses on developing standards and technologies that address the unique requirements of modern AI infrastructure.

Key benefits

- **Ultra-high performance connectivity:** delivers 800G networking capabilities with advanced Cisco® Silicon One® and Cisco Cloud Scale ASIC technology, providing the bandwidth and low-latency performance essential for demanding AI workloads.
- **Intelligent traffic management suite:** features Cisco Intelligent Packet Flow technology with dynamic load balancing, congestion awareness, and real-time fault detection to optimize network performance and reduce job completion times.
- **Lossless network architecture:** vital for critical AI workloads in data centers, ensuring seamless data flow, reduced latency, and optimal performance. It enables precise training and real-time inference, preventing disruptions that could compromise AI accuracy and efficiency during critical AI operations.
- **Unified management and analytics:** provides centralized network orchestration through Cisco Nexus® Dashboard with built-in AI fabric templates, end-to-end RoCEv2 visibility, and advanced congestion analytics for simplified operations.
- **Future-ready investment protection:** is Ultra Ethernet Consortium (UEC) standards-ready and provides vendor-agnostic compatibility with leading AI ecosystem partners including NVIDIA, AMD, Intel®, VAST, and WEKA.
- **Deployment-ready rack solution:** Cisco and Hyve AI Networking Solutions arrive pre-installed, cabled, labeled, and fully tested. With dock-to-deployment, setup time is reduced from weeks to days. Each rack undergoes rigorous thermal and workload stress testing to ensure it is deployment-ready and operational upon arrival.

How it works

High-performance switching and rack solution

Cisco Nexus 9364E-SG2 series switches

The Cisco Nexus 9364E-SG2 series switches are the flagship AI networking platform and feature 64 ports of 800 Gigabit in a 2-Rack-Unit (2RU) form factor, powered by a custom Silicon One G200 ASIC. This switch provides 51.2 Tbps of switching capacity with a 256MB fully shared packet buffer and delivers exceptional performance for large-scale AI training and inference workloads.

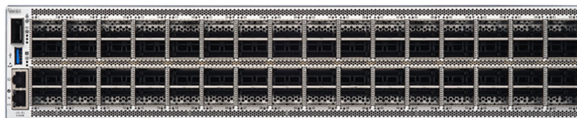


Figure 1. Cisco Nexus 9364E-SG2-O switch

Cisco Nexus 9300-GX2 Series fixed switches

Cisco Nexus 9300-GX2 Series Fixed Switches are designed for organizations that are beginning their AI infrastructure journey. These switches provide 400G connectivity with 120MB shared-memory architecture.

Available in multiple port configurations, they offer flexible deployment options while maintaining compatibility with existing data-center investments.

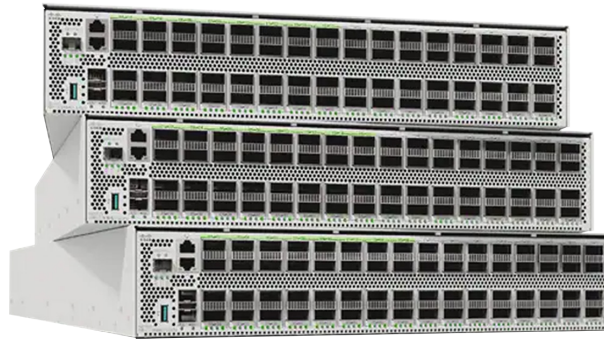


Figure 2. Cisco Nexus 9300-GX2 Series Fixed Switches

Advanced software capabilities

Intelligent packet flow technology

Cisco's comprehensive traffic management suite dynamically responds to real-time network conditions through live telemetry, congestion awareness, and fault detection. The system employs multiple load-balancing strategies including flowlet-based dynamic load balancing, per-packet distribution, weighted cost multipath routing, and policy-based traffic prioritization.

Lossless fabric implementation

The solution provides lossless network architecture enabling optimal performance and reliability for critical AI workloads in data centers.

Unified management platform

Cisco Nexus Dashboard integration

The centralized management platform provides built-in templates for AI fabric provisioning, comprehensive RoCEv2 visibility, and advanced congestion analytics. Future enhancements will include job-level insights, topology-aware visualization, and NIC observability capabilities.

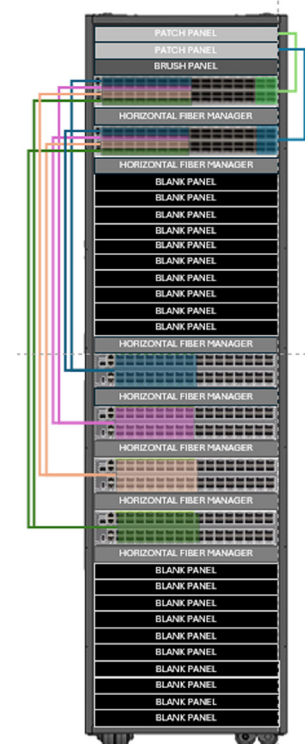


Figure 3. Hyve integrated rack with Cisco Nexus fabric



Primary use cases

- **AI model training:** large-scale distributed training environments requiring high-bandwidth, low-latency communication between GPU clusters for gradient synchronization and parameter updates.
- **Inference and real-time AI:** production AI environments serving real-time inference requests with stringent latency requirements and high throughput demands.
- **Generative AI workloads:** support for large language models and generative AI applications that require massive data movement and consistent network performance.
- **High-performance computing:** traditional HPC workloads that benefit from lossless, high-bandwidth networking for scientific computing and research applications.
- **AI-driven analytics:** data-intensive analytics workloads that require efficient movement of large datasets between storage and compute resources.

Services

The Cisco and Hyve AI Networking Solution provides comprehensive services to ensure successful AI networking deployments. Our expert teams deliver design consultation, implementation planning, performance optimization, and ongoing support services tailored to AI infrastructure requirements.

Our collaborative approach includes assessment of existing infrastructure, development of migration strategies, implementation of best practices, and provision of training programs to maximize the value of AI networking investments.

The Cisco and Hyve advantage

The partnership between Cisco and Hyve Solutions delivers unique value through complete control of the technology stack—from custom silicon development to software innovation and management platforms. This integrated approach ensures superior reliability, accelerated innovation cycles, and comprehensive accountability throughout the solution lifecycle.

Hyve is a global leader in the design and deployment of AI digital infrastructure, delivering purpose-built network rack solutions engineered for modern data centers, colocation facilities, and rack-level AI deployments.

Through direct collaboration with leading technology providers such as Cisco, Hyve combines deep technical expertise in network rack integration with the ability to manage complex, large-scale infrastructure projects with precision. Our dedicated integration and test facilities, strategically located worldwide, ensure every deployment meets the highest standards of quality, reliability, and performance.

By pairing best-in-class networking technologies with Hyve's proven design-to-deployment capabilities, organizations gain a trusted partner that simplifies delivery, accelerates scale, and maximizes the value of their infrastructure investments.

Together, Cisco and Hyve provide customers with confidence in long-term technology investments while protecting against obsolescence through continuous innovation and standards leadership.

Cisco Capital

Financing to help you achieve your objectives

Cisco Capital can help you acquire the technology you need to achieve your objectives and stay competitive. We can help you reduce CapEx. Accelerate your growth. Optimize your investment dollars and ROI. Cisco Capital financing gives you flexibility in acquiring hardware, software, services, and complementary third-party equipment. And there's just one predictable payment. Cisco Capital is available in more than 100 countries. [Learn more.](#)

Learn more

Organizations ready to transform their AI networking infrastructure can benefit from immediate action. The rapidly evolving AI landscape demands networking solutions that can scale with growing computational requirements while maintaining consistent performance and operational simplicity.

Contact your Cisco representative or authorized Hyve Solutions partner to schedule a comprehensive assessment of your AI networking requirements. Our teams will evaluate your current infrastructure, develop a tailored deployment strategy, and provide implementation guidance to accelerate your AI initiatives.

For additional technical resources and detailed implementation guidance, visit the [Cisco AI Networking](#) resource center or [Hyve Solutions](#) engage with our technical specialists for personalized consultation services.

As a US-led global ODM, Hyve Solutions is a leader in the design to worldwide deployment of hyperscale digital infrastructures. In partnership with customers, Hyve leverages deep-seated industry experience to design and deliver purpose-built server, storage, and networking solutions optimized for AI workloads and emerging technologies, meeting data center demands for today and beyond.

©2025 Hyve Solutions Corporation, a wholly owned subsidiary of TD SYNEX Corporation. All rights reserved. Hyve Solutions, and all other Hyve Solutions product and service names and slogans are trademarks or registered trademarks of TD SYNEX Corporation. Reg. U.S. Pat. & Tm Off. Other names and marks are the property of their respective owners.

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R) C22-5471602-00 10/25

