



NVIDIA Spectrum-X Ethernet Platform

Ethernet purpose-built for AI.



The NVIDIA Spectrum-X™ Ethernet platform is designed specifically to improve the performance and efficiency of Ethernet-based AI infrastructure of AI factories and clouds. This breakthrough technology achieves 1.6x better AI networking performance compared to off-the-shelf (OTS) Ethernet, along with consistent, predictable performance in multi-tenant environments.

Spectrum-X Ethernet is built on network innovations powered by the tight coupling of the NVIDIA Spectrum-X Ethernet switch and NVIDIA® SuperNIC™ adapters. By optimizing the network for massive transformer-based generative AI models, Spectrum-X Ethernet reduces runtimes and delivers faster time to insight.

Enhancing Ethernet for AI

AI workloads require a low-latency network with high, effective bandwidth that's also compatible with collective libraries such as the NVIDIA Collective Communications Library (NCCL). In AI cloud environments where multiple jobs run simultaneously, performance isolation is critical to prevent interference among tenants running on the same compute infrastructure. To meet the requirements for high-performance networking and cloud multi-tenancy, Spectrum-X Ethernet leverages remote direct-memory access (RDMA) over Converged Ethernet (RoCE) extensions. Built with standards-based Ethernet, these extensions improve effective bandwidth and ensure performance isolation between workloads.

Spectrum-X Ethernet Innovations

The Spectrum-X Ethernet platform combines switches, SuperNICs, and software to form an NCCL-optimized network fabric built for AI cluster performance at any scale. The Spectrum-X Ethernet platform includes a suite of end-to-end innovations.

> **Adaptive Routing and Congestion Control:** Adaptive routing avoids congestion by dynamically routing large AI flows away from congestion points. The Spectrum-X Ethernet switch employs fine-grained load balancing to reroute active flows, while SuperNICs handle out-of-order packets, placing them correctly in destination memory. Additionally, Spectrum-X Ethernet inherently eliminates cascading performance issues from a lost link, limiting bandwidth loss to only that single link.

Simultaneously, RoCE congestion control collects in-band network telemetry. SuperNIC algorithms use this telemetry and deep learning models for data metering, optimizing settings for multi-job, multi-tenant systems.

Key Components

NVIDIA Spectrum-X Ethernet Networking Platform Components

- > NVIDIA Spectrum-X Ethernet switches (SN5610/SN5600/SN5600D/SN6600/SN6800/SN6810)
- > NVIDIA Spectrum-X Ethernet SuperNICs (BlueField-3, ConnectX®-8, ConnectX-9)
- > NVIDIA LinkX® cables and transceivers

NVIDIA Innovations Enabling Spectrum-X Ethernet and Spectrum-XGS Ethernet

- > 200G SerDes
- > Adaptive routing
- > Programmable congestion control
- > Performance isolation
- > Hardware-based accelerated multiplane architecture
- > End-to-end high-frequency telemetry
- > Scale-across precision latency management
- > Enhanced AI fabric security
- > Power-efficient, co-packaged optics platforms

- **Multiplane Topologies:** Multiplane is the breakthrough that allows Ethernet to scale from thousands to hundreds of thousands of GPUs in a flat two-tier topology. Effective scaling requires overcoming challenges due to switch radix limits, oversubscription, and component reliability. Spectrum-X multiplane solves this with hardware-coordinated SuperNIC routing. In a Spectrum-X multiplane network, each SuperNIC (which corresponds to a GPU) is split into two or more ports, which each connects into a separate network plane. Each network plane features its own distinct leaf/spine/optional superspine layer. The SuperNIC intelligently routes network traffic to different planes, load balancing packet-by-packet in hardware for optimal performance.

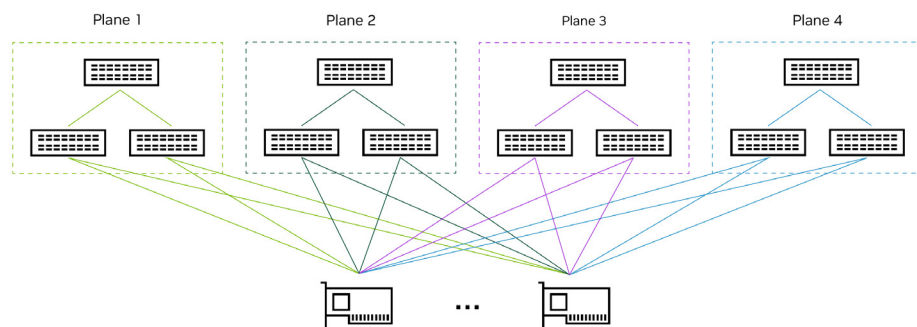


Figure 1: Spectrum-X Multiplane Topology

For NVIDIA Vera Rubin AI factories connected with Spectrum-X Ethernet featuring Spectrum-6 and ConnectX-9 (1.6 T per GPU, with eight planes), this increases the number of GPUs connected in two levels from 2K to up to 128K—an increase of up to 64x more scale.

In addition, Spectrum-X multiplane improves resiliency by offering multiple independent paths between GPUs. Leveraging high-frequency telemetry from the switch and integrating with NCCL for a complete view of the plane topologies, the SuperNIC is able to route traffic away from planes with problematic link outages or congestion issues.

- **NVIDIA Full-Stack Integration and Extreme Co-Design:** NVIDIA's extreme co-design approach tightly integrates the full software- and hardware-accelerated computing stack—including Spectrum-X Ethernet.

Spectrum-X Ethernet is engineered through deep, full-stack integration across the entire NVIDIA Accelerated Computing platform. From the latest acceleration libraries to network visibility and simulation tools, every layer of the stack is co-designed with the networking fabric to optimize performance from the GPU through the SuperNIC and across every switch hop.

This tight alignment ensures that collective communication, data movement, congestion control, and telemetry are all optimized holistically rather than in isolated subsystems.

- **Rack-Scale Ethernet:** Spectrum-X Ethernet is fully integrated into the NVIDIA MGX™ ETL rack-scale architecture, delivering high-performance Ethernet networking for ultra-dense compute racks and expanding MGX liquid-cooling and reliability innovations to a wider variety of compute options. The power-efficient, copper-based cable cartridge connects liquid-cooled Spectrum-X Ethernet switch trays to up to 256 GPUs in a single rack, while front-panel optical connectivity enables scale-out to larger GPU domains.

► **Network Visibility:** With end-to-end network visibility powered by NVIDIA NetQ™, operators gain flow-level insight across switches (Spectrum-X Ethernet and NVLink), GPUs, SuperNICs, and switch ports, including detailed per-hop behavior on RoCE queues. This integrated telemetry enables precise performance monitoring, rapid troubleshooting, and deterministic optimization of multi-tenant AI workloads running across the fabric.

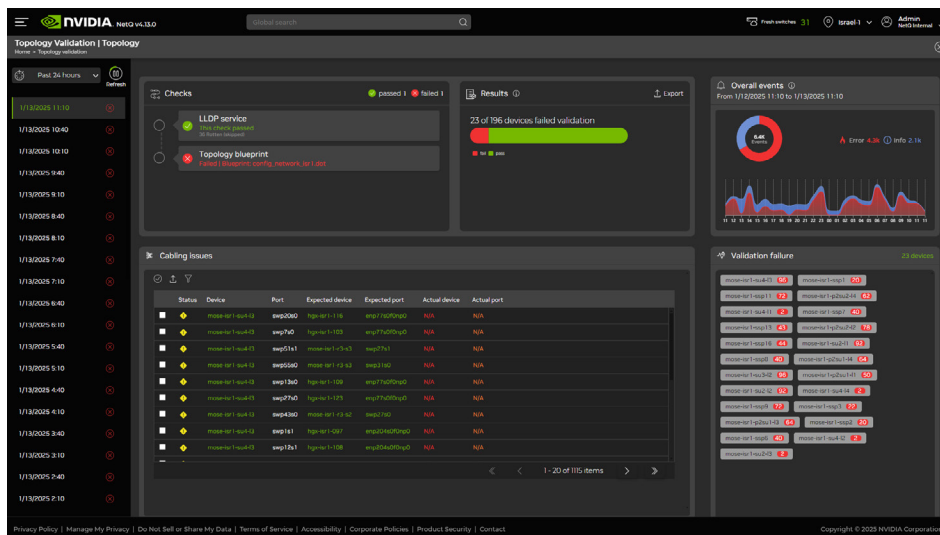


Figure 2: Network visibility shown via NVIDIA NetQ

► **Network Simulation:** The NVIDIA DSX Air platform simplifies Spectrum-X Ethernet deployments by enabling digital twins of the entire network to design, test, and validate network provisioning, automation, and security policies. NVIDIA DSX Air allows day-zero operations before hardware is deployed by simulating and automating changes ahead of production. It accelerates deployment of Spectrum-X Ethernet infrastructure, reducing time to first token while ensuring cloud-scale efficiency and reliability.

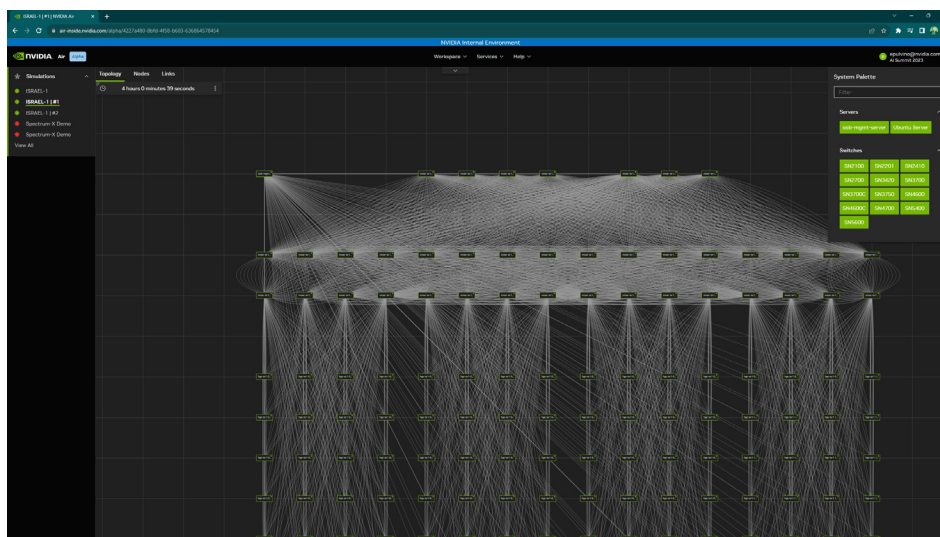


Figure 3: Network simulation with NVIDIA DSX Air

► **End-to-End AI Fabric Security:** Spectrum-X Ethernet provides comprehensive, end-to-end security for the AI fabric, ensuring data integrity, confidentiality, and compliance without sacrificing performance. Its switches and SuperNICs are built with secure boot, hardware roots of trust, encrypted firmware, and countermeasures against physical attacks. Multi-tenant environments benefit

from VXLAN BGP EVPN with L3EVPN, delivering secure segmentation and tenant isolation. NVIDIA's patented 802.1X provisioning further simplifies secure onboarding. For data-in-transit protection, Spectrum-X SuperNICs offload and accelerate IPsec and PSP, enabling transparent, line-rate encryption across the AI fabric. Together, these capabilities establish Spectrum-X Ethernet as a resilient, scalable networking platform purpose-built to safeguard the world's largest AI workloads.

- **Gigascale-Ready With NVIDIA Spectrum-XGS Ethernet:** As AI factories grow, physical footprints often force data centers to span multiple buildings or campuses. Connecting these facilities with off-the-shelf Ethernet introduces latency and jitter.

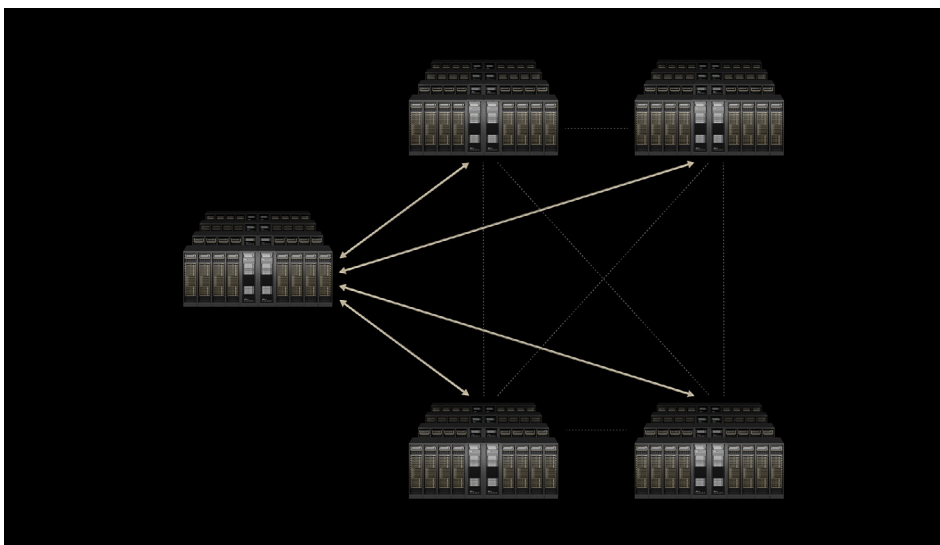


Figure 4: Scale-across networking with NVIDIA Spectrum-XGS

NVIDIA Spectrum-XGS Ethernet provides high-performance scale-across networking, enabling multiple data centers to operate as a single, massive AI factory. Spectrum-XGS Ethernet introduces a modified congestion control algorithm that accounts for distance to efficiently balance traffic. Combined with precise latency management, it allows facilities separated by hundreds of kilometers to function as a unified gigascale cluster. With advanced load balancing and topology-aware congestion control, NVIDIA Spectrum-XGS Ethernet accelerates multi-site NCCL collectives by up to 1.9x, enabling high-efficiency scale-across training and inference.

All of these innovations can be explored further in the [NVIDIA Spectrum-X Ethernet technical whitepaper](#).

Ready to Get Started?

Learn more about the NVIDIA Spectrum-X Ethernet networking platform at: nvidia.com/spectrumx