Datasheet

![NVIDIA logo]

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 powering AI factories and clouds. This breakthrough technology achieves 1.6x better AI networking performance, along with consistent, predictable performance in multi-tenant environments. Spectrum-X is built on network innovations powered by the tight coupling of the NVIDIA Spectrum-X Ethernet switch and NVIDIA® SuperNIC™ adapters. Spectrum-X network optimizations reduce runtimes of massive transformer-based generative AI models and deliver faster time to insight.

Enhancing Ethernet for AI

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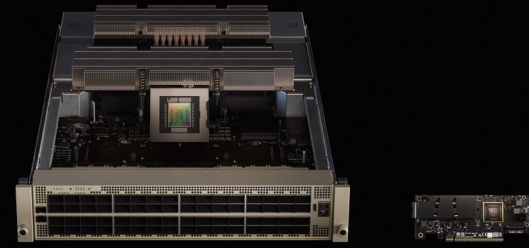
Key Features

NVIDIA Spectrum-X Ethernet Networking Platform Components

> NVIDIA Spectrum-4 Ethernet switches (SN5610/SN5600/SN5600D)

> NVIDIA ConnectX®-8 SuperNICs

> NVIDIA BlueField®-3 SuperNICs

> NVIDIA LinkX® cables and transceivers

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

> RoCE adaptive routing

> Programmable congestion control

> Performance isolation

> End-to-end high-frequency telemetry

> Precision latency management

> AI fabric security

RoCE Adaptive Routing for AI: All-Reduce

All-Reduce

Average Bandwidth

Collective Size

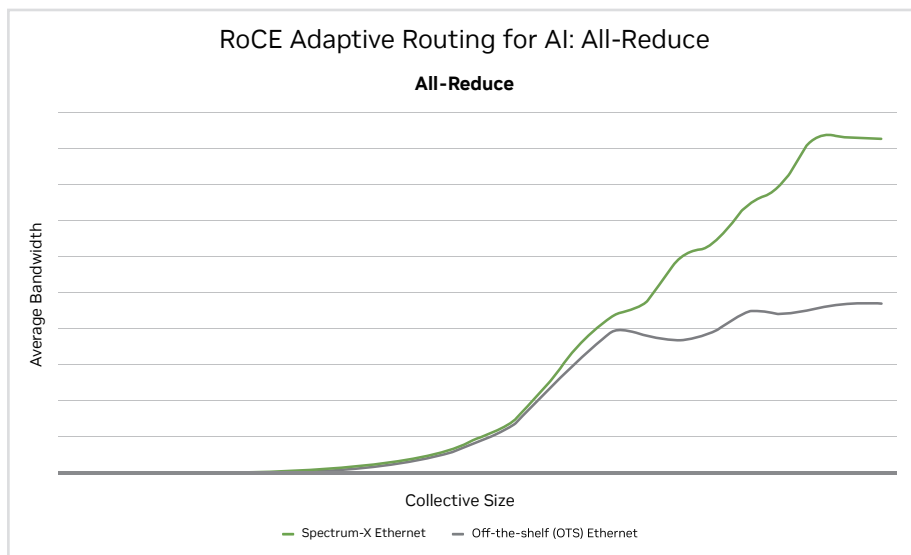— Spectrum-X Ethernet   — Off-the-shelf (OTS) Ethernet

Figure 1. RoCE Adaptive Routing for AI. All-ReduceThe Key Benefits of NVIDIA Spectrum-X Ethernet

The Key Benefits of NVIDIA Spectrum-X Ethernet

> **Improved AI Performance:** Spectrum-X Ethernet enhances AI networking performance by 1.6x, accelerating processing, analysis, and execution of AI workloads and, in turn, the development and deployment of AI solutions.

> **Accelerated NCCL Multi-Site Performance:** NVIDIA Spectrum-XGS Ethernet load balancing and topology-aware congestion control accelerates multi-site NCCL collectives by 1.9x, enabling the high performance needed for training and inference workloads to scale-across data centers.

> **Standard Ethernet Connectivity:** While enhancing Ethernet functionality, Spectrum-X is fully standards-based and is completely interoperable with Ethernet-based stacks.

> **Increased Power Efficiency:** By improving performance, NVIDIA Spectrum-X Ethernet contributes to a more power-efficient AI environment. This leads to reduced power consumption and lower operational costs for AI clusters.

> **Enhanced Multi-Tenant Performance:** Performance isolation in multi-tenant environments ensures that each tenant's workloads perform optimally and consistently, resulting in higher customer satisfaction and improved service quality.

> **Better AI Fabric Visibility:** Visibility into the flows as they traverse multiple switches makes it possible to identify performance bottlenecks and is a key part of a modern, automated fabric-validation solution.

> **Higher AI Scalability:** Scale to 256x 200G ports in a single hop or thousands of 800G endpoints in a two-tier leaf/spine topology supports the expansion of the AI cloud while maintaining high levels of performance, making it an ideal solution for organizations with evolving AI infrastructure needs.

> **Faster Network Setup:** The automated, end-to-end configuration of advanced networking functionality is fully tuned for AI workloads.

> **Higher Resiliency:** With higher resiliency, the cascading performance issues that occur with a lost link are eliminated, limiting the loss in bandwidth to that single link.

> **End-to-End AI Fabric Security:** Spectrum-X Ethernet delivers built-in protection across switches, SuperNICs, and cables with hardware roots of trust, secure boot, firmware and physical attack defenses, tenant isolation with VXLAN BGP EVPN, and offloaded IPsec/PSP encryption—ensuring confidential, compliant, and resilient GPU-to-GPU communication at AI scale.

> **Scale-Out and Scale-Across AI Networking:** Spectrum-XGSEthernettechnology enables a shared network architecture for high-performance connectivity within the data center as well as between multiple data centers.

Increased Power Efficiency

Power capping has become a common practice in data centers due to the growing demand for computing resources and the need to control energy costs. Overall, improving AI performance per watt is essential to achieve greater computational efficiency and deliver insights faster, while staying within power budgets. This is particularly important in applications such as deep learning, where training models can be computationally intensive and require a large amount of power. In addition to its raw performance benefits, Spectrum-X Ethernet improves power efficiency, delivering superior performance per watt over off-the-shelf Ethernet.

Spectrum-X Technology Innovations

Spectrum-X Ethernet switches and SuperNICs work in tight coordination to form an **NCCL-optimized network fabric** built to optimize AI cluster performance with a suite of end-to-end innovations:

> **RoCE adaptive routing** avoids congestion by dynamically routing large AI flows away from congestion points. This approach improves network resource utilization, leaf/spine efficiency, and performance. The Spectrum-X Ethernet switch employs fine-grained load balancing to reroute active flows and eliminate congestion. Additionally, the Spectrum-X SuperNICs work in tandem to handle out-of-order packets, placing packets in the correct order in the destination memory. RoCE adaptive routing supports profiles for efficient provisioning and automation.

> **RoCE congestion control** collects network performance data with in-band network telemetry. The Spectrum-X SuperNICs use the collected switch telemetry data to optimize network data rates. SuperNIC algorithms use deep learning models for data metering, optimizing settings for multi-job, multi-tenant systems.

> **End-to-end visibility with NVIDIA NetQ™** traces flow-level performance from the GPU to the SuperNIC and maps the path and per-hop behavior across switch ports and RoCE queues.

> **NVIDIA full-stack integration** includes NVIDIA NetQ, NCCL, Nsight™, NVIDIA Blackwell and NVIDIA Hopper Platforms, Spectrum-4 switches, ConnectX-8 and BlueField-3 SuperNICs, SDKs, and software—all configurable with NVIDIA Base Command™ Manager for simple, production-ready deployment and faster time to AI.

> **Very low latency** is critical for AI and machine learning workloads that require real-time processing. The Spectrum-4 switch consistently low latency 400/800 GbE switching, ensuring ultra-low application latency and jitter.

> **NCCL-optimized** switch behavior thresholds enable the optimization of buffer and congestion thresholds for AI workloads. This ensures synchronized collective operations, reducing the likelihood of congestion and packet loss.

Spectrum-X Ethernet offers a groundbreaking solution for organizations building Ethernet-based, multi-tenant AI infrastructure. Spectrum-X Ethernet enhances performance and energy efficiency of AI across various applications, resulting in higher predictability and consistency. This leads to faster time to market and a stronger competitive advantage.

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, BlueField-3 and ConnectX-8 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.

Achieving Giga-Scale AI With NVIDIA Spectrum-XGS Ethernet

As AI demand continues to grow, it requires larger AI factories—but power and physical footprints place limits on data centers within a building or campus. Connecting multiple data centers together with off-the-shelf Ethernet networking introduces long-tail latency, increased jitter, and unpredictable, inconsistent performance. To solve this, NVIDIA Spectrum-XGS Ethernet technology provides high-performance scale-across networking for AI, enabling multiple data centers separated by long distances to operate as a single, massive data center at giga-scale.

Spectrum-XGS Ethernet is an innovative technology integrated into the Spectrum-X Ethernet platform and is based on the same hardware combination of Spectrum-X Ethernet switches and ConnectX-8 SuperNICs. To meet the unique requirements of high-performance scale-across networking, Spectrum-XGS Ethernet introduces a modified congestion control algorithm that takes into account the distance between data center facilities to efficiently balance traffic. Combined with end-to-end telemetry and precise latency management for deterministic performance, Spectrum-XGS Ethernet enables data centers in different buildings or even campuses separated by hundreds of kilometers to function as a single AI factory. Spectrum-XGS Ethernet is critical for scaling infrastructure to accommodate the largest-scale training and inference workloads.

NVIDIA Spectrum-X Ethernet Networking Platform

The Spectrum-X Ethernet networking platform consists of the following components.

The Spectrum-4 Ethernet switch for smart-leaf, spine, and super-spine designs offers 64 ports of 800GbE in a dense 2U form factor, playing a vital role in NVIDIA Spectrum-X Ethernet deployments. The Spectrum-4 Ethernet switch models (SN5600, SN5600D, SN5610) support both standard leaf and spine designs with top-of-rack switches and rail-optimized, end-of-row topologies. These switches offer diverse connectivity in combinations of 10–800GbE and boast an industry-leading total throughput of 51.2 terabits per second (Tb/s).

The ConnectX-8 SuperNIC delivers the highest network performance with enhanced I/O programmability, while also providing performance isolation—essential for efficiently managing multi-tenant generative AI factory environments. As an Ethernet SuperNIC, ConnectX-8 provides a total of 800 Gb/s data throughput (via two ports of 400 Gb/s). ConnectX-8 is PCI Express Gen6-based, offering up to 48 lanes for use cases such as PCIe switching, is backwards-compatible with PCIe Gen 5/4/3, and supports advanced fabric enhancement features like Quality of Service (QoS) and congestion control, among others.

The BlueField-3 SuperNIC delivers up to 400GbE RoCE network connectivity between GPU servers and enables NVIDIA GPUDirect® RoCE for optimizing peak AI workload efficiency. The BlueField-3 SuperNIC accelerates AI networking and performs low-compute tasks. It's best suited for the B3140H model, providing an HHHL form-factor and a sub-75 watt power envelope, making it an ideal choice for NVIDIA Spectrum-X Ethernet deployments.

LinkX transceivers and cables provide a robust selection of direct attach copper (DAC) cables, active copper cables (ACCs), active optical cables (AOCs), and optics needed to connect octal small form-factor pluggable (OSFP)- and quad small form-factor pluggable (QSFP)-based fabrics. Spectrum-X Ethernet switches and SuperNICs, combined with the LinkX portfolio, enable the most power-efficient AI fabric on the market.

Ready to Get Started?

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

サーヴァンツ インターナショナル 株式会社

160-0023　東京都新宿区西新宿 6-20-7
コンシェリア西新宿タワーズウエスト　4 F
電話:03-4455-7531 FAX 03-3346-5234
www.servants.co.jp

NVIDIA