NVIDIA Announces New Switches Optimized for Trillion-Parameter GPU Computing and AI Infrastructure

- *NVIDIA Quantum-X800 InfiniBand for Highest-Performance AI-Dedicated Infrastructure*
- *NVIDIA Spectrum-X800 Ethernet for AI-Optimized Networking in Every Data Center*
- *NVIDIA Software Distributes Computing Across Blackwell, New Switches and BlueField-3 SuperNICs to Boost AI, Data Processing, HPC and Cloud Workloads*

GTC—NVIDIA today announced a new wave of networking switches, the X800 series, designed for massive-scale AI.

The world's first networking platforms capable of end-to-end 800Gb/s throughput, NVIDIA Quantum-X800 InfiniBand and NVIDIA Spectrum™-X800 Ethernet push the boundaries of networking performance for computing and AI workloads. They feature software that further accelerates AI, cloud, data processing and HPC applications in every type of data center, including those that incorporate the newly released NVIDIA Blackwell architecture-based product lineup.

"NVIDIA Networking is central to the scalability of our AI supercomputing infrastructure," said Gilad Shainer, senior vice president of Networking at NVIDIA. "NVIDIA X800 switches are end-to-end networking platforms that enable us to achieve trillion-parameter-scale generative AI essential for new AI infrastructures."

Initial adopters of Quantum InfiniBand and Spectrum-X Ethernet include Microsoft Azure and Oracle Cloud Infrastructure.

"AI is a powerful tool to turn data into knowledge. Behind this transformation is the evolution of data centers into high-performance AI engines with increased demands for networking infrastructure," said Nidhi Chappell, Vice President of AI Infrastructure at Microsoft Azure. "With new integrations of NVIDIA networking solutions, Microsoft Azure will continue to build the infrastructure that pushes the boundaries of cloud AI."

CoreWeave is also among early adopters.

Next Standard for Extreme Performance

The Quantum-X800 platform sets a new standard in delivering the highest performance for AI-dedicated Infrastructure. It includes the NVIDIA Quantum Q3400 switch and the NVIDIA ConnectX®-8 SuperNIC™, which together achieve an industry-leading end-to-end throughput of 800Gb/s. This is 5x higher bandwidth capacity and a 9x increase of 14.4Tflops of In-Network Computing with NVIDIA's Scalable Hierarchical Aggregation and Reduction Protocol (SHARPv4) compared to the previous generation.

The Spectrum-X800 platform delivers optimized networking performance for AI cloud and enterprise infrastructure. Utilizing the Spectrum SN5600 800Gb/s switch and the NVIDIA BlueField®-3 SuperNIC, the Spectrum-X800 platform provides advanced feature sets crucial for multi-tenant generative AI clouds and large enterprises.

Spectrum-X800 optimizes network performance, facilitating faster processing, analysis, and execution of AI workloads, thereby expediting the development, deployment, and time to market of AI solutions. Designed specifically for multi-tenant environments, Spectrum-X800 ensures performance isolation for each tenant's AI workloads to maintain optimal and consistent performance levels, enhancing customer satisfaction and service quality.

NVIDIA Software Support

NVIDIA provides a comprehensive suite of network acceleration libraries, software development kits and management software to optimize performance for trillion-parameter AI models.

This includes NVIDIA Collective Communications Library (NCCL), which extends GPU parallel computing tasks to the Quantum-X800 network fabric, taking advantage of its powerful In-Network Computing capabilities with SHARPv4 supporting FP8, supercharging performance for large model training and generative AI.

NVIDIA's full-stack software approach provides advanced programmability, making data center networks more flexible, reliable and responsive, ultimately increasing overall operational efficiency and supporting the needs of modern applications and services.

Ecosystem Momentum

Next year, Quantum-X800 and Spectrum-X800 will be available from a wide range of leading infrastructure and system vendors around the world, including Aivres, DDN, Dell Technologies, Eviden, Hitachi Vantara, Hewlett Packard Enterprise, Lenovo, Supermicro and VAST Data.

About NVIDIA

Since its founding in 1993, NVIDIA (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company's

invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing infrastructure company with data-center-scale offerings that are reshaping industry. More information at https://nvidianews.nvidia.com/.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance, features, and availability of NVIDIA's products and technologies, including NVIDIA Quantum-X800 InfiniBand, NVIDIA Spectrum-X800 Ethernet, NVIDIA Blackwell architecture-based products, NVIDIA Quantum Q3400 switch, NVIDIA ConnectX-8 SuperNIC, NVIDIA's Scalable Hierarchical Aggregation and Reduction Protocol (SHARPv4), Spectrum SN5600 800Gb/s switch, NVIDIA BlueField-3 SuperNIC, and NVIDIA Collective Communications Library (NCCL); NVIDIA X800 switches enabling us to achieve trillion-parameter-scale generative AI essential for new AI infrastructures; NVIDIA's full-stack software approach making data center networks more flexible, reliable and responsive, ultimately increasing overall operational efficiency and supporting the needs of modern applications and services; and third parties' use and adoption of NVIDIA's products and technologies, and the benefits thereof are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

Alex Shapiro
Enterprise Networking
1-415-608-5044
ashapiro@nvidia.com