Datasheet

# NVIDIA Quantum-2 QM9700 Series

Scaling out data centers with 400G InfiniBand smart switches.

## Accelerate Research and Product Innovation with Enhanced Data and In-Network Computing

As high-performance computing (HPC) and artificial intelligence (AI) applications increase in complexity, the demand for advanced high-speed networking is critical for extreme-scale systems. NVIDIA Quantum-2, a premier switch platform, excels in power and density, offering 400 gigabits per second (Gb/s) of InfiniBand throughput. This high networking performance is essential for AI developers and scientific researchers taking on the world's most challenging problems.

## Advanced Computing Needs Advanced Networking

The NVIDIA Quantum-2-based QM9700 and QM9790 switch systems deliver 64 ports of 400Gb/s InfiniBand per port, packed into a 1U standard chassis. Each switch carries an impressive 51.2 terabits per second (Tb/s) of aggregated bidirectional throughput and a landmark capacity exceeding 66.5 billion packets per second (BPPS). Supporting NVIDIA's advanced 400Gb/s interconnect technology, NVIDIA Quantum-2 is a high-speed, extremely low-latency, and scalable networking solution. Key features incorporated include state-of-the-art technologies such as remote direct-memory access (RDMA), adaptive routing, and NVIDIA® Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)™.

Unlike other networking solutions, NVIDIA InfiniBand incorporates self-healing network capabilities, quality of service, enhanced virtual lane mapping, and advanced congestion control, maximizing overall application throughput. As ideal rack-mounted InfiniBand solutions, the QM9700 and QM9790 400Gb/s InfiniBand fixed-configuration switches offer exceptional flexibility, with support for various topologies like Fat Tree, SlimFly, DragonFly+, multi-dimensional Torus, and others. They maintain backward compatibility with previous generations and support a broad software ecosystem.

## The Era of Data-Driven Computing

Today's complex research demands ultra-fast processing for high-resolution simulations, massive datasets, and highly parallelized algorithms with real-time information exchanges. The QM9700 400Gb/s InfiniBand switches, enhancing NVIDIA In-Network Computing technologies, incorporate the third generation of NVIDIA SHARP technology, SHARPv3. This innovation allows virtually unlimited scalability for small and large data aggregation across the network. SHARPv3 delivers AI acceleration that's 32X higher than its predecessor, significantly enhancing the performance of complex computational applications as data moves through the data center network. It actively participates in runtime processes and reduces the volume of data needed across the network.

## Streamlining Network Design and Topologies

By implementing NVIDIA port-split technology, the QM9700 and QM9790 switches provide a double-density radix for 200Gb/s data speeds, reducing the cost of network design and network topologies. Supporting up to 128 ports of 200Gb/s, NVIDIA delivers a highly dense top-of-rack (TOR) switch, offering significant space efficiency and enhanced data throughput. The QM9700 family of switches enables small to medium-sized deployments to scale with a two-level Fat Tree topology while reducing power, latency, and space requirements.

> "NVIDIA SHARPv3 delivers AI acceleration that's 32X higher than its predecessor, significantly enhancing the performance of complex computational applications as data moves through the data center network."

## Integrated Router Capabilities

NVIDIA Quantum-2 InfiniBand switches with optional router capabilities support the scale-out of InfiniBand clusters to a very large number of nodes. This capability significantly surpasses the limits of the previous generation to sustain the peak performance and reliability demands of research, simulations, AI, and data processing for cloud applications.

## Enhanced Management

The internally managed QM9700 switch features an on-board subnet manager that enables simple, out-of-the-box bringup for up to 2,000 nodes. Running the NVIDIA MLNX-OS® software package, the subnet manager delivers full chassis management through command-line interface (CLI), web-based user (WebUI), Simple Network Management Protocol (SNMP), or JavaScript Object Notation (JSON) interfaces. The externally managed QM9790 switch can utilize the advanced NVIDIA Unified Fabric Manager (UFM®) feature sets to empower data center operators to efficiently provision, monitor, manage, preventatively troubleshoot, and maintain the modern data center fabric, realizing higher utilization and reducing overall opex.

## Energy Efficiency

Several NVIDIA Quantum-2 switches grouped in a few racks and connected with short-reach cables create a virtual modular switch. NVIDIA's system uses copper instead of optical cables for better energy efficiency, cutting power usage from 30 watts to 0–3 watts per OSFP port. This reduction not only cuts latency and boosts reliability but also lowers the total cost of ownership. The layout includes InfiniBand switches linked by copper cables in a Fat Tree topology, at the top level, or as an 'open' tree topology when used at the middle levels (e.g., between levels 1 and 2 in a 3-level Fat Tree).

| System Specifications | |
|---|---|
| **Performance** | 400Gb/s per port |
| **Switch radix** | 64 400Gb/s non-blocking ports with aggregate data throughput up to 51.2Tb/s |
| **Connectors and cabling** | 32 octal small form-factor pluggable (OSFP) connectors; passive or active copper or active fiber cable; optical module |
| **Power supply** | > 1+1 redundant and hot-swappable power<br>> Input range: 200–240Vac; US min: 2 phases of<br>> 100–110v—in total, at least 208v<br>> 80 Gold+ and ENERGY STAR certified |
| **Cooling** | Front-to-rear or rear-to-front<br>Cooling option: 6+1 hot-swappable fan unit |
| **Management ports** | > 1x USB 3.0 x1<br>> 1x USB for I2C channel 1x RJ45<br>> 1x RJ45<br>> 1x RJ45 (UART) |
| **CPU** | x86 Coffee Lake i3 |
| **System memory** | Single 8GB, 2,666 mega transfers per second (MT/s), DDR4 SO-DIMM |
| **Storage** | M.2 SSD SATA 16GB 2242 FF |
| **Software** | MLNX-OS |
| **System weight** | 14.5kg |
| **System dimensions** | > Height: 1.7in (43.6 mm)<br>> Width: 17.0in (438 mm)<br>> Depth: 26.0in (660.4 mm**)** |
| **Rack mount** | 1U rack mount |
| **Environmental conditions** | Temperature:<br>> Operational:<br>Forward air flow: 0–35°C<br>Reverse air flow: 0–40°C<br>Non-operational: -40–70°C<br>> Humidity:<br>Operating 10–85% non-condensing<br>Non-operating 10–90% non-condensing<br>> Altitude:<br>Up to 3,050m |
| **EMC (emissions)** | CE, FCC, VCCI, ICES, and RCM |
| **Product safety/compliant/certified** | RoHS, CB, cTUVus, CE, and CU |
| **Warranty** | One year |

## Enterprise Support and Services

A minimum one-year **Enterprise Business-Standard Support** entitlement is required when purchasing NVIDIA Quantum-2 InfiniBand switches.

- Enterprise Support provides access to NVIDIA experts, the NVIDIA Enterprise Support Portal, advanced return material authorization (RMA), and more.
- Add-on services—including installation, configuration, a technical account manager, a four-hour on-site engineer, expedited RMA, media retention, and more—are available.

For details, visit the **NVIDIA Enterprise Support and Services User Guide**.

## Product Specifications

### Switches Portfolio

For details on the NVIDIA Quantum-2 InfiniBand series switches, along with their ordering part numbers, please refer to the **QM9700/QM9790 1U NDR Switch Systems User Manual**.

### Transceivers and Cables

For details on NVIDIA cables and transceivers, visit the **Interconnect documentation hub**.

## Ready to Get Started?

To learn more about NVIDIA Quantum-2 QM9700 switches, visit: **www.nvidia.com/en-eu/networking/quantum2**