Relevant for Models: Q3200-RA and Q3400-RA

This manual describes the installation and basic use of the NVIDIA XDR InfiniBand switch systems based on the NVIDIA Quantum™-3 switch ASIC. This manual is intended for IT managers and system administrators.

Ordering Information

System Model	NVIDIA SKU	Description
Q3200-RA	920-9B34F-00RX-FS0	Quantum-3 based Two-Adjoining XDR InfiniBand Switches, Q3200-RA, 2U, with 36 XDR Ports over 18 OSFP cages per Switch, 4 Power Supplies (Power Cords Not Included), Standard Depth, Managed, C2P Airflow, Rail Kit
Q3400-RA	920-9B36F-00RX-8S0	NVIDIA Quantum-3 based XDR InfiniBand Switch, , Q3400-RA, 4U, 144 XDR Ports over 72 OSFP Cages, 8 Power Supplies (Power Cords Not Included), Standard Depth, Managed, C2P Airflow, Rail Kit

Related Documentation

Document	Description
InfiniBand Architecture Specification Volume 1 Release 1.5	The InfiniBand Trade Association (IBTA) InfiniBand® Specification at https://www.infinibandta.org.
NVIDIA NVOS User Manual for InfiniBand Switches	This document contains information regarding the configuration and management of the NVOS® software - https://docs.nvidia.com/networking/software/switch-software/index.html#infiniband-nvos
Hands-on workshops	Visit https://academy.nvidia.com/en/infiniband-customized-training/.
On-site/remote services	For any tailor-made service, contact: nbu-services-sales@nvidia.com.

Revision History

A list of the changes made to this document are provided in Document Revision History.

Introduction

The NVIDIA Quantum-3 family of fixed-configuration switches revolutionizes the performance, scalability, and efficiency of high-performance computing and AI infrastructures, enabling faster and more effective AI processing and computation. These switches are available in both 4U and 2U systems.

The NVIDIA Quantum-X800 Q3400-RA 4U switch, leverages 200Gb/s-per-lane serializer/deserializer (SerDes) technology. It features 144 ports at 800Gb/s distributed across 72 Octal Small Form-Factor Pluggable (OSFP) cages. The switch's high radix supports a two-level fat-tree topology capable of connecting up-to 10,368 ConnectX-8 network interface cards (NICs) with minimal latency and optimal job locality. The Q3400 is air-cooled.

For integration with existing NDR or HDR infrastructures, the NVIDIA Quantum-X800 Q3200 2U air-cooled configuration switch is ideal. This system houses two independent switches within a single enclosure, together providing 72 ports at 800Gb/s (over 36 OSFP cages).

Both the Q3400 and Q3200 switches include a dedicated InfiniBand in-band management port specifically for NVIDIA Unified Fabric Manager (UFM®) management, separated on the front panel from the other ports. This separation allows the full set of standard ports to be used for data network connectivity, simplifying port allocation and streamlining topology design.

Additionally, NVIDIA Quantum-X800 switches feature optional router capabilities, facilitating the expansion of InfiniBand clusters to support a large scale of nodes located across multiple sites.

The 4U switch systems are ideal for HPC and AI workloads, leveraging the cutting-edge NVIDIA Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)™, alongside enhanced adaptive routing and performance isolation technologies. With this powerful combination, they deliver unparalleled performance for handling the most demanding tasks in HPC and AI domains. The NVIDIA Quantum-3 switches are available with internal management capabilities.

As a rack-mounted InfiniBand solution, the NVIDIA Quantum-3 switches provide maximum flexibility for various network topologies. They support a wide range of configurations, including Fat Tree, Dragonfly+, multi-dimensional Torus, and more. In these setups, the 4U systems act as spines, while the 2U systems serve as leaves, complementing the overall network architecture.