



NVIDIA-Certified Professional: AI Networking Exam Study Guide



NVIDIA-Certified Professional: AI Networking Exam Study Guide

Contents

AI Data Center Design and Optimization:	2
Exam Weight 5%	
NVIDIA Spectrum Networking:	3
Exam Weight 30%	
NVIDIA InfiniBand Networking:	5
Exam Weight 30%	
Kubernetes Integration:	6
Exam Weight 5%	
Troubleshooting Tools:	7
Exam Weight 20%	
Automation and Configuration:	8
Exam Weight 10%	

This study guide provides an overview of each topic covered on the NVIDIA AI Networking certification exam, recommended training, and suggested reading to prepare for the exam. Information about NVIDIA certifications can be found [here](#).

Job Description

NVIDIA AI networking engineers play a critical role in building and optimizing AI infrastructure powered by NVIDIA's advanced networking technologies. They configure high-performance networks to ensure seamless data flow and low-latency communication for AI and machine learning workloads. By fine-tuning protocols and integrating NVIDIA products with servers, GPUs, and storage systems, they enable efficient deployment in data centers. Troubleshooting issues and collaborating with cross-functional teams are key aspects of their work.

Job Responsibilities

1. Configure high-performance networking solutions for AI and machine learning workloads.
2. Optimize network performance to ensure maximum throughput and minimal latency for AI computations.
3. Implement and fine-tune network protocols to enhance data transfer speeds and efficiency.
4. Integrate NVIDIA networking products with existing AI infrastructure, including servers, GPUs, and storage systems.
5. Deploy networking solutions in data centers to ensure seamless connectivity between AI components.
6. Diagnose and resolve networking issues impacting AI workloads to maintain optimal system performance.
7. Provide technical support and guidance to teams managing AI infrastructure.
8. Collaborate with data scientists, researchers, and IT professionals to understand networking requirements and challenges..

Recommended Qualifications and Experience

1. Bachelor's degree in computer science, software engineering, AI, or a related field
2. Expertise in NVIDIA GPU and DPU technologies, networking, and AI infrastructure.

Certification Topics and References

AI Data Center Design and Optimization: Exam Weight 5%

Test takers will be assessed on their ability to analyze and design robust AI data center architectures, select appropriate network technologies (Ethernet, InfiniBand), differentiate storage options for AI throughput, and optimize topologies for performance, scalability, and redundancy in NVIDIA-powered environments.

- 1.1 Describe an AI factory networking architecture and its components (e.g., GPUs, NVIDIA® BlueField®, scalable units, switches).
 - 1.2 Describe rail-optimized topologies for high-performance AI workloads.
 - 1.3 Describe GPU-to-GPU communications
-

Recommended Training (Optional)

- > [InfiniBand Essentials | NVIDIA Academy](#)

Suggested Readings

- > [Aurelien Degremont, Nathan Dauchy LUG'24 - May 7-8, 2024](#)
- > [BlueField-3 Administrator Quick-Start Guide | NVIDIA Docs](#)
- > [NVIDIA DGX SuperPOD™: AI Infrastructure for Enterprise Deployments](#)
- > [Doubling all2all Performance With NVIDIA Collective Communication Library 2.12 | NVIDIA Technical Blog](#)
- > [Host-Side Interface Configuration | NVIDIA Docs](#)
- > [Key Components of the DGX SuperPOD | NVIDIA Docs](#)
- > [Logging | NVIDIA Docs](#)
- > [Modes of Operation | NVIDIA Docs](#)
- > [Networking for Data Centers and the Era of AI | NVIDIA Technical Blog](#)
- > [NVIDIA BlueField Networking Platform](#)
- > [NVIDIA BlueField Reset and Reboot Procedure | NVIDIA Docs](#)
- > [NVIDIA DGX SuperPOD: Scalable Infrastructure for AI Leadership](#)
- > [NVIDIA GB200 NVL72 Delivers Trillion-Parameter LLM Training and Real-Time Inference | NVIDIA Technical Blog](#)
- > [NVIDIA Spectrum-X™ Whitepaper](#)
- > [NVIDIA Unveils Reference Architecture for AI Cloud Providers | NVIDIA Blog](#)
- > [NVIDIA NVLink™ and NVSwitch™: Fastest HPC Data Center Platform](#)
- > [Overview of NCCL | NCCL 2.28.3 Documentation](#)
- > [Rail-Optimized Topology Validation | NVIDIA Docs](#)
- > [Rail-Optimised Networking: How NVIDIA Is Rethinking AI Network Design in the Data Centre | Vespertec](#)
- > [SONiC Wiki](#)
- > [Turbocharging Generative AI Workloads With NVIDIA Spectrum-X Networking Platform | NVIDIA Technical Blog](#)

NVIDIA Spectrum Networking: Exam Weight 30%

Tasks focus on configuring, optimizing, securing, and troubleshooting Spectrum-X switches. The test taker must demonstrate skills using tools such as NVIDIA NetQ™, CloudAI Benchmark, and DPUs, define advanced quality of service (QoS) and telemetry, monitor and tune network performance, and resolve complex real-world connectivity and microsegmentation scenarios in Ethernet-based AI fabrics.

- 2.1 Configure NVIDIA Spectrum-X switches for RDMA over Converged Ethernet (RoCE) to enable high-speed, low-latency communication.
 - 2.2 Enable and verify QoS, explicit congestion notification (ECN), and priority flow control (PFC), advanced features like adaptive routing, and telemetry
 - 2.3 Configure multi-tenancy Border Gateway Protocol Ethernet VPN (BGP-EVPN) to isolate tenant workloads.
 - 2.4 Use NVIDIA Air to simulate network environments and identify potential issues.
 - 2.5 Diagnose congestion or packet loss using in-band telemetry and NVIDIA What Just Happened® (WJH) services.
 - 2.6 Use NetQ for real-time network monitoring, including congestion detection and latency measurements.
 - 2.7 Install NVIDIA DOCA™.
 - 2.8 Configure NVIDIA SuperNIC™ functionality for advanced packet processing and congestion control.
-

Recommended Training (Optional)

- > [NVIDIA Cumulus® Linux Essentials | NVIDIA Academy](#)

Suggested Readings

- > [Adaptive Routing: Cumulus NetQ 4.8 | NVIDIA Docs](#)
- > [AI Overview and Fabric Requirements | Dell Technologies](#)
- > [Automating Data Center Networks With NVIDIA NVUE and Ansible | NVIDIA Technical Blog](#)
- > [Basic Configuration: Cumulus Linux 4.4 | NVIDIA Docs](#)
- > [BGP Weighted Equal Cost Multipath: Cumulus Linux 5.7 | NVIDIA Docs](#)
- > [CloudAI Benchmark Framework | GitHub](#)
- > [Configure and Monitor Threshold-Crossing Events | NVIDIA Docs](#)
- > [Cumulus Linux Administration | NVIDIA Academy](#)
- > [Custom Topology | NVIDIA Air](#)
- > [Developer Guide | NVIDIA Docs](#)
- > [Diagnosing Network Issues Faster With NVIDIA WJH | NVIDIA Technical Blog](#)
- > [DOCA Installation Guide for Linux | NVIDIA Docs](#)
- > [DOCA Profiles | NVIDIA Docs](#)
- > [Ethernet SuperNICs: Networking for Next Wave of AI](#)
- > [EVPN Multihoming: Cumulus Linux 5.10 | NVIDIA Docs](#)
- > [Flow Analysis: Cumulus NetQ 4.4 | NVIDIA Docs](#)
- > [Get Started With NVIDIA DOCA](#)
- > [Layer 3 Extensions With VRF Route Leaking | NVIDIA Docs](#)
- > [Maximize Network Automation Efficiency With Digital Twins on NVIDIA Air | NVIDIA Technical Blog](#)
- > [Network Operations and Cumulus NetQ](#)
- > [Networking for the Era of AI: The Network Defines the Data Center Whitepaper](#)
- > [NVIDIA Air](#)
- > [NVIDIA ConnectX-8 SuperNIC User Manual | NVIDIA Docs](#)
- > [NVIDIA DOCA Profiles | NVIDIA Docs](#)
- > [NVIDIA Launches Accelerated Ethernet Platform for Hyperscale Generative AI](#)

- > [NVIDIA Spectrum-X Network Platform Architecture Whitepaper](#)
- > [NVIDIA Spectrum-X Network Platform Architecture: The First Ethernet Network Designed to Accelerate AI Workloads Whitepaper](#)
- > [NVIDIA Spectrum-X Whitepaper](#)
- > [QoS" NVUE 5.x | NVIDIA Docs](#)
- > [RDMA over Converged Ethernet \(RoCE\): Cumulus Linux 4.4 | NVIDIA Docs](#)
- > [Resource Diagnostics Using cl-resource-query: Cumulus Linux 4.4 | NVIDIA Docs](#)
- > [Set Up Your KVM Virtual Machine for a Single On-Premises Server: Cumulus NetQ 4.4 | NVIDIA Docs](#)
- > [Spectrum-X Datasheet](#)
- > [Troubleshooting | NCCL 2.28.3 Documentation](#)
- > [Turbocharging Generative AI Workloads With NVIDIA Spectrum-X Networking Platform | NVIDIA Technical Blog](#)
- > [Using VXLAN Routing With EVPN Through Asymmetric or Symmetric Models | NVIDIA Technical Blog](#)
- > [Virtual Routing and Forwarding \(VRF\): Cumulus Linux 4.4 | NVIDIA Docs](#)
- > [What Is a SuperNIC? | NVIDIA Blog](#)
- > [What Is ISER? | NVIDIA Enterprise Support Portal](#)
- > [What Just Happened \(WJH\): Cumulus Linux 4.4 | NVIDIA Docs](#)

NVIDIA InfiniBand Networking: Exam Weight 30%

Tasks cover setting up and maintaining InfiniBand switches, configuring reliable, multi-tenant networks, managing subnet managers and fabric partitioning, monitoring performance with tools like NVIDIA Unified Fabric Manager (UFM®), resolving congestion or topology issues, and validating advanced, rail-optimized AI data center designs for scalability and low latency.

- 3.1 Perform initial configuration and provisioning, including high availability (HA).
- 3.2 Configure partition keys (PKeys) to ensure secure multi-tenancy in InfiniBand networks.
- 3.3 Configure QoS and adaptive routing to dynamically adjust paths based on congestion.
- 3.4 Use UFM to monitor InfiniBand link status and bandwidth utilization.

Recommended Training (Optional)

- > [InfiniBand Essentials | NVIDIA Academy](#)
- > [InfiniBand Network Administration | NVIDIA Academy](#)

Suggested Readings

- > [Appendix: Partitioning | NVIDIA Docs](#)
- > [Configure HA With SSH Trust | NVIDIA Docs](#)
- > [How to Configure Adaptive Routing and Self-Healing Networking \(New\) | NVIDIA Enterprise Support Portal](#)
- > [IB Router | NVIDIA Docs](#)
- > [InfiniBand | NVIDIA Docs](#)
- > [InfiniBand Fabric Managed by UFM | NVIDIA Docs](#)
- > [InfiniBand Fabric Utilities | NVIDIA Docs](#)
- > [InfiniBand Network | NVIDIA Docs](#)
- > [InfiniBand QoS | NVIDIA Docs](#)
- > [InfiniBand-Related Troubleshooting | NVIDIA Docs](#)
- > [InfiniBand Switching | NVIDIA Docs](#)
- > [Initial Configuration | NVIDIA Docs](#)
- > [Installation Notes | NVIDIA Docs](#)
- > [Introduction to InfiniBand Whitepaper](#)
- > [Managing Data Centers Securely and Intelligently With NVIDIA UFM Cyber-AI | NVIDIA Technical Blog](#)
- > [Memory Efficiency, Faster Initialization, and Cost Estimation With NVIDIA Collective Communications Library 2.22 | NVIDIA Technical Blog](#)
- > [Network Traffic Map | NVIDIA Docs](#)
- > [NVIDIA MLNX-OS User Manual v3.10.4100 \(LTS\) | NVIDIA Docs](#)
- > [NVIDIA SM | NVIDIA Docs](#)
- > [NVIDIA UFM Enterprise User Manual v6.17.0 | NVIDIA Docs](#)
- > [NVIDIA Unified Fabric Manager \(UFM\) Datasheet](#)
- > [Operations Procedures | NVIDIA Docs](#)
- > [Overview | NVIDIA Docs](#)
- > [Overview: Generative AI in the Enterprise With NVIDIA Spectrum-X Networking Platform | Dell Technologies Info Hub](#)
- > [Subnet Manager: NVIDIA MLNX-OS User Manual v3.11.1014 | NVIDIA Docs](#)
- > [Subnet Manager: NVIDIA MLNX-OS User Manual v3.11.1004 | NVIDIA Docs](#)
- > [Telemetry | NVIDIA Docs](#)
- > [Transport Modes | NVIDIA Docs](#)
- > [Upgrading Operating System Software | NVIDIA Docs](#)
- > [Using mlxconfig | NVIDIA Docs](#)
- > [Using NVIDIA Scalable Hierarchical Aggregation Reduction Protocol \(SHARP\)™ With NVIDIA NCCL | NVIDIA Docs](#)

Kubernetes Integration: Exam Weight 5%

The test taker will be assessed on deploying and managing the NVIDIA Network Operator and device plug-ins to enable RDMA/InfiniBand within Kubernetes clusters, validating GPU resources, and ensuring seamless integration with AI workloads. Tasks include troubleshooting cluster-network interfaces, deploying drivers, and scheduling resources on heterogeneous infrastructure.

4.1 Deploy the NVIDIA Network Operator to manage RDMA interfaces and InfiniBand networks within Kubernetes clusters.

4.2 Verify NVIDIA Network Operator functionality.

Suggested Readings

- > [Chapter 5. NVIDIA GPUDirect Remote Direct Memory Access \(RDMA\), Hardware Accelerators: OpenShift Container Platform 4.19 | Red Hat Documentation](#)
- > [Getting Started With Kubernetes | NVIDIA Docs](#)
- > [Getting Started With Red Hat OpenShift | NVIDIA Docs](#)
- > [NVIDIA Network Operator | GitHub](#)
- > [Network Operator | NVIDIA Docs](#)
- > [Network Operator: NVIDIA DGX BasePOD™ on RHEL | NVIDIA Docs](#)
- > [NVIDIA Network Operator | NVIDIA Docs](#)
- > [NVIDIA Network Operator Deployment Guide With Kubernetes | NVIDIA Docs](#)
- > [NVIDIA Network Operator v25.7.0 | NVIDIA Docs](#)

Troubleshooting Tools: Exam Weight 20%

Assessment tasks include using diagnostic command-line tools and telemetry (e.g., `cl-resource-query`, WJH, `ib_write_lat`, NetQ, UFM) for root-cause analysis of latency, congestion, or connectivity issues in AI networks. The candidate must quickly identify faults in complex environments and recommend effective, production-ready remediation strategies.

- 5.1 Use tools like `cl-resource-query` to check resource allocation in Spectrum-X environments.
- 5.2 Use WJH services for real-time event analysis.
- 5.3 Verify low-latency interconnects between GPUs, CPUs, and storage systems.
- 5.4 Use UFM system health to diagnose InfiniBand issues
- 5.5 Use commands like `ib_write_lat`, `ib_write_bw`, `ibping`, `ibstat`, `ibdiagnet`, `ibnodes`, and `iblinkinfo` to diagnose connectivity issues.

Suggested Readings

- > [Appendix: Diagnostic Utilities | NVIDIA Docs](#)
- > [Benchmark NVIDIA GPUDirect® RDMA With InfiniBand Write Bandwidth | Oracle Help Center](#)
- > [Doubling all2all Performance With NVIDIA Collective Communication Library 2.12 | NVIDIA Technical Blog](#)
- > [Forwarding Table Size and Profiles: Cumulus Linux 5.7 | NVIDIA Docs](#)
- > [linux-rdma/perftest: Infiniband Verbs Performance Tests | GitHub](#)
- > [Ib_write_lat | NVIDIA Enterprise Support Portal](#)
- > [ibnetdiscover Command: Sun Datacenter InfiniBand Switch 72 | Oracle Help Center](#)
- > [InfiniBand Fabric Diagnostic Utilities | NVIDIA Docs](#)
- > [InfiniBand Fabric Utilities: MLNX_OFED Documentation Rev 5.0-2.1.8.0 | NVIDIA Docs](#)
- > [InfiniBand Fabric Utilities: MLNX_OFED Documentation Rev 4.6-1.0.1.1 | NVIDIA Docs](#)
- > [IP over InfiniBand \(IPoIB\) | NVIDIA Docs](#)
- > [Linux InfiniBand Drivers](#)
- > [Network Operator: NVIDIA DGX BasePOD on RHEL | NVIDIA Docs](#)
- > [NVIDIA Collective Communications Library \(NCCL\)](#)
- > [NVIDIA MLNX_OFED Documentation v23.10-1.1.9.0 LTS | NVIDIA Docs](#)
- > [NVIDIA Unified Fabric Manager \(UFM\)](#)
- > [Resource Diagnostics: Cumulus Linux 5.12 | NVIDIA Docs](#)
- > [Resource Diagnostics Using cl-resource-query: Cumulus Linux 4.4 | NVIDIA Docs](#)
- > [SRP - SCSI RDMA Protocol | NVIDIA Docs](#)
- > [Switches: Cumulus NetQ 4.12 | NVIDIA Docs](#)
- > [System Health | NVIDIA Docs](#)
- > [Troubleshooting | NVIDIA Docs](#)
- > [UFM Health Tab | NVIDIA Docs](#)
- > [UFM Server Health Monitoring | NVIDIA Docs](#)
- > [UFM System Dump Tab | NVIDIA Docs](#)
- > [What Just Happened \(WJH\): Cumulus Linux 4.4 | NVIDIA Docs](#)
- > [What Just Happened \(WJH\): Cumulus Linux 5.14 | NVIDIA Docs](#)

Automation and Configuration: Exam Weight 10%

Assessment tasks focus on automating Spectrum-X and InfiniBand switch deployments using tools like NVUE templates and custom Ansible playbooks, testing for zero-touch operations, and maintaining configuration integrity across multi-vendor, multi-device AI networking environments, including automating RoCE, virtual local area network (VLAN), and high-availability settings.

6.1 Manage Spectrum-X switch configurations through NVUE templates.

6.2 Write Ansible playbooks to automate network setup tasks like VLAN creation or RoCE configuration.

Recommended Training (Optional)

> [Cumulus Linux Essentials | NVIDIA Academy](#)

Suggested Readings

- > [Automating Data Center Networks With NVIDIA NVUE and Ansible | NVIDIA Technical Blog](#)
- > [Automation With Cumulus Linux | NVIDIA Docs](#)
- > [Command-Line Reference | Open Network Install Environment \(ONIE\) Documentation](#)
- > [Interface: NVUE 5.x | NVIDIA Docs](#)
- > [Introduction to Automation and Ansible | NVIDIA Docs](#)
- > [NVIDIA User Experience \(NVUE\) Cheat Sheet | NVIDIA Docs](#)
- > [NVUE Automation With Ansible | NVIDIA Docs](#)
- > [NVUE CLI: Cumulus Linux 5.4 | NVIDIA Docs](#)
- > [NVUE Command Reference | NVIDIA Docs](#)
- > [Running the Playbooks | NVIDIA Docs](#)
- > [Upgrading Cumulus Linux | NVIDIA Docs](#)