



NVIDIA-Certified Professional: AI Networking 认证考试学习指南



NVIDIA-Certified Professional: AI Networking 认证考试学习指南

目录

AI 数据中心设计和优化: 考试权重 5%	2
NVIDIA Spectrum 网络: 考试权重 30%	3
NVIDIA InfiniBand 网络: 考试权重 30%	5
Kubernetes 集成: 考试权重 5%	7
故障排除工具: 考试权重 20%	8
自动化和配置: 考试权重 10%	10

针对 NVIDIA-Certified Professional: AI Networking (NCP-AIN) 认证考试，本学习指南包含认证所涵盖的各个技术主题的介绍，并推荐了相关培训课程和阅读资料。

查看所有 NVIDIA 认证，[请单击此处](#)。

考试适用人群

在构建和优化采用 NVIDIA 高级网络技术的 AI 基础架构方面，NVIDIA AI 网络工程师发挥着关键作用。负责为 AI 和机器学习工作负载配置高性能网络，确保无缝数据流和低延迟通信。通过优化协议并将 NVIDIA 产品与服务器、GPU 和存储系统相集成，在数据中心实现了高效部署。故障排除和与跨职能团队协作。

相关工作职责示例

1. 为 AI 和机器学习工作负载配置高性能网络解决方案
2. 优化网络性能，确保 AI 计算实现最大吞吐量和最低延迟
3. 实施和调优网络协议，提高数据传输速度和效率
4. 将 NVIDIA 网络产品与现有的 AI 基础架构 (包括服务器、GPU 和存储系统) 相集成
5. 在数据中心部署网络解决方案，确保 AI 组件之间的无缝连接
6. 诊断和解决影响 AI 工作负载的网络问题，保持系统性能最佳状态
7. 为管理 AI 基础架构的团队提供技术支持和指导
8. 与数据科学家、研究人员和 IT 专业人员协作，了解网络需求和挑战

建议具备的知识和经验

1. 具备 NVIDIA GPU/DPU 技术、网络和 AI 基础架构专业知识
2. 理工科背景，如计算机科学、软件工程、AI 等

认证主题和参考资料

AI 数据中心设计和优化：考试权重 5%

分析和设计稳健型 AI 数据中心架构、选择适当的网络技术 (以太网、InfiniBand)、区分支持 AI 吞吐量的存储选项，以及在 NVIDIA 驱动的环境中优化拓扑以实现性能、可扩展性和冗余。

1.1 描述 AI 工厂网络架构及其组件 (例如 GPU、NVIDIA® BlueField®、可扩展单元、交换机)。

1.2 描述面向高性能 AI 工作负载的轨道优化拓扑。

1.3 描述 GPU 之间的通信

培训推荐（可选）

> InfiniBand 必备知识 ([了解详情](#))

阅读内容推荐

- > [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 网络：考试权重 30%

任务侧重于配置、优化、保护 Spectrum-X 交换机并排除故障。使用 NVIDIA NetQ™、CloudAI Benchmark 和 DPU 等工具展示技能，定义高级 QoS 和遥测，监测和调优网络性能，并在以太网架构的 AI Fabric 中，解决复杂的真实场景连接需求及微分段问题。

2.1 配置用于 RDMA over Converged Ethernet (RoCE) 的 NVIDIA Spectrum-X 交换机，以实现高速、低延迟通信。

2.2 启用并验证 QoS、显式拥塞通知 (ECN)、优先级流量控制 (PFC)、自适应路由和遥测等高级特性

2.3 配置多租户边界网关协议以太网 VPN (BGP-EVPN)，以隔离租户工作负载。

2.4 使用 NVIDIA Air 模拟网络环境并识别潜在问题。

2.5 使用带内遥测和 NVIDIA WJH (故障快照) 服务诊断拥塞或数据包丢失。

2.6 使用 NetQ 进行实时网络监控，包括拥塞检测和延迟测量。

2.7 安装 NVIDIA DOCA™。

2.8 配置 NVIDIA SuperNIC™ 功能，用于高级数据包处理和拥塞控制。

培训推荐（可选）

- > Spectrum-X 网络平台管理 ([了解详情](#))
- > NVIDIA Cumulus® Linux 必备知识 ([了解详情](#))

阅读内容推荐

- > [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 网络：考试权重 30%

设置和维护 InfiniBand 交换机、配置可靠的多租户网络、管理子网管理器 and 网络分区、使用 NVIDIA Unified Fabric Manager (UFM®) 等工具监测性能、解决拥塞或拓扑问题，以及验证高级轨道优化 AI 数据中心设计，实现可扩展性和低延迟。

3.1 执行初始配置和预配，包括高可用性 (HA)。

3.2 在 InfiniBand 网络中配置分区键 (PKey)，确保安全的多租户环境。

3.3 配置 QoS 和自适应路由，根据拥塞动态调整路径。

3.4 使用 UFM 监控 InfiniBand 链路状态和带宽利用。

培训推荐（可选）

> InfiniBand 必备知识 ([了解详情](#))

> InfiniBand 网络管理 ([了解详情](#))

阅读内容推荐

> [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 集成：考试权重 5%

部署和管理 NVIDIA Network Operator 和设备插件以在 Kubernetes 集群中启用 RDMA/InfiniBand，验证 GPU 资源，以及确保与 AI 工作负载的无缝集成。排除集群网络接口故障、部署驱动程序以及在异构基础架构上调度资源。

4.1 部署 NVIDIA Network Operator 以管理 Kubernetes 集群内的 RDMA 接口和 InfiniBand 网络。

4.2 验证 NVIDIA Network Operator 功能。

阅读内容推荐

- > [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](#)

故障排除工具：考试权重 20%

使用诊断命令行工具和遥测 (例如, cl-resource-query、WJH (故障快照)、ib_write_lat、NetQ、UFM) 对 AI 网络中的延迟、拥塞或连接问题进行根本原因分析。快速识别复杂环境中的故障, 并推荐有效且可直接用于生产环境的修复策略。

5.1 使用 cl-resource-query 等工具检查 Spectrum-X 环境中的资源分配。

5.2 使用 WJH (故障快照) 服务进行实时事件分析。

5.3 验证 GPU、CPU 与存储系统之间的低延迟互连。

5.4 使用 UFM 系统运行状况诊断 InfiniBand 问题。

5.5 使用 ib_write_lat、ib_write_bw、ibping、ibstat、ibdiagnet、ibnodes 和 iblinkinfo 等命令诊断连接问题。

培训推荐 (可选)

> MLXlink 和 MLXcables 调试工具 ([了解详情](#))

阅读内容推荐

- > [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](#)

自动化和配置：考试权重 10%

使用 NVUE 模板和自定义 Ansible Playbook 等工具实现 Spectrum-X 和 InfiniBand 交换机部署自动化，零接触操作测试，以及在多网络厂商、多设备 AI 网络环境中保持配置完整性，包括自动完成 RoCE、虚拟局域网 (VLAN) 和高可用性设置。

6.1 通过 NVUE 模板管理 Spectrum-X 交换机配置。

6.2 写入 Ansible Playbook，以自动完成网络设置任务，如创建 VLAN 或配置 RoCE。

培训推荐（可选）

- > 面向网络工程师的 Ansible 必备知识 ([了解详情](#))
- > Cumulus Linux 必备知识 ([了解详情](#))

阅读内容推荐

- > [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](#)

遇到问题？

问题咨询，请发邮件至 dlichina@nvidia.com