

AI Infrastructure and Operations – Professional Training

Prepare for NCP-AII and NCP-AIO Certifications

OUTLINE

Training Overview

In today's AI-driven era, the ability to deploy and manage AI clusters efficiently and effectively is critical for organizations seeking to leverage the power of artificial intelligence. This specialized course is designed to equip professionals with the skills and knowledge necessary to optimize efficiency, reliability, and scalability in deploying AI environments within data centers.

Covering various infrastructure and operation aspects like compute platforms, and storage, the course focuses on key AI components such as GPUs, CPUs, and DPUs.

Participants will gain practical insights into provisioning and managing AI data centers, implementing AI workloads, and utilizing AI virtualization techniques. Through a blend of hands-on training and theoretical learning, participants develop the expertise needed to effectively deploy and manage AI infrastructure, ensuring optimal performance and continuous operation of AI data centers by course end.

NCP-AII and NCP-AIO Certifications

The **NCP-AI Infrastructure (NCP-AII)** certification is an intermediate-level credential that validates a candidate's ability to deploy, configure, and validate advanced NVIDIA AI infrastructure.

The **NCP-AI Operations (NCP-AIO)** certification is an intermediate-level credential that validates a candidate's ability to monitor, troubleshoot, and optimize AI infrastructure by NVIDIA.

This multi-day workshop is recommended for training for the **NCP-AII** and **NCP-AIO** certifications.

Training Delivery Method

Instructor-led remote training sessions via NVIDIA Academy platform. Hands-on lab exercises focused on the data center infrastructure and operations.

Target Audience

The course is designed for administrators, DevOps professionals, and IT-related roles who want to gain the knowledge and skills necessary to deploy and operate AI data centers.

Training Duration

Remote | 6 sessions of 5 hours

Prerequisites

- Knowledge of core networking concepts and principles, including the TCP/IP model, Ethernet standards, basics of routing and switching, common network topologies, and IP addressing schemes.
- Hands-on experience in Linux-like systems administration, such as managing users and permissions, installing software packages, configuring network settings, and troubleshooting common issues in a Linux environment.
- Basic understanding of server hardware components and their roles in a data center environment. This includes knowledge of CPUs, memory, storage devices, and networking interfaces commonly found in servers.
- Knowledge of storage concepts and principles, including different file systems and their characteristics.
- Basic understanding of virtualization technologies, including virtual machines (VMs) and containers. You should be familiar with VM creation, management, and the role of hypervisors in virtualized environments.
- Basic understanding of artificial intelligence (AI) concepts and terminology. This may include knowledge of topics such as machine learning, deep learning, neural networks, and common AI applications.
- Before attending the course, we recommend completing the [AI Infrastructure and Operation Fundamentals](#) self-paced course. This course will provide the foundations for this training.

Training Outline

AI in the Data Center Overview

- AI Overview
- Data Center Architecture for AI Workloads

Compute Platforms for AI

- AI Compute Platforms Overview
- Scaling AI Compute
- NVIDIA System Management (NVSM)
- GPU Monitoring with DCGM
- Hardware Installation and Replacement
- NCCL Overview
- AI Cluster Validation
- **Practice 1:** GPU Driver Installation
- **Practice 2:** Data Center Server and GPU Management Tools

Networking for AI

- AI Data Centers Networks

BlueField Networking Platform for AI

- BlueField Overview
- BlueField Bring-up
- BlueField Interfaces
- Running Services on the BlueField DPU – 45 min
- **Practice 1:** BlueField Initial Configuration
- **Practice 2:** Inspect The DPU
- **Practice 3:** Running the BlueMan Service – 15 min

Storage for AI

- Storage for AI Overview
- Storage Configuration
- **Practice 1:** Testing Storage Performance

AI Data Center Management

- AI Data Center Management Overview
- Base Command Manager (BCM) Overview
 - Head Node Installation
 - Bringing Up the Cluster
 - Managing Nodes and Software Images
 - Base Command Manager Monitoring
 - Base Command Manager User Management
- Scheduling AI Workloads with Slurm
- AI Cluster Orchestration with Kubernetes – 60 min
- Run:ai Administration – 90 min
 - Run:ai Overview
 - Run:ai Building Blocks
 - Run:ai Manageability and Visibility
 - Run:ai Workloads
- **Practice 1:** Base Command Manager Administration
- **Practice 2:** Slurm Installation
- **Practice 3:** Kubernetes Installation – 30 min

AI Virtualization

- GPU Partitioning
- Configuring MIG - Demo

NVIDIA AI Software

- NVIDIA AI Software Overview – 60 min
- Using AI Applications with Containers
- **Practice 1:** - Installing and Running AI Software