Training overview

AI Solutions on Cisco Infrastructure Essentials (DCAIE)

Description

The **AI Solutions on Cisco Infrastructure Essentials (DCAIE)** training covers the essentials of deploying, migrating, and operating AI solutions on Cisco data center infrastructure. You'll be introduced to key AI workloads and elements, as well as foundational architecture, design, and security practices critical to successful delivery and maintenance of AI solutions on Cisco infrastructure.

This training, along with the Operate and Troubleshoot AI Solutions on Cisco Infrastructure (DCAIAOT) training, prepares you for the Implementing Cisco Data Center AI Infrastructure (300-640 DCAI) v1.0 exam. If passed, you earn the Cisco Certified Specialist - Data Center AI Infrastructure certification and fulfill the concentration exam requirement for the Cisco Certified Network Professional (CCNP) Data Center certification. This training also earns 34 Continuing Education (CE) credits toward recertification.

How you'll benefit

This training will help you:

- Gain the knowledge you need to deploy, migrate, and operate AI solutions on Cisco data center infrastructure
- Qualify for professional-level job data center roles
- Prepare for the 300-640 DCAI v1.0 exam
- Earn 34 CE credits toward recertification

Who should enroll

- Network Designers
- Network Administrators
- Storage Administrators
- Network Engineers
- Systems Engineers
- Data Center Engineers
- Consulting Systems Engineers
- Technical Solutions Architects

- Cisco Integrators/Partners
- Field Engineers
- Server Administrators
- Network Managers
- Program Managers Project
- Managers

Technology areas

- Data Center
- Artificial Intelligence (AI)

Objectives

- Describe key concepts in artificial intelligence, focusing on traditional AI, machine learning, and deep learning techniques and their applications
- Describe generative AI, its challenges, and future trends, while examining the nuances between traditional and modern AI methodologies
- Explain how AI enhances network management and security through intelligent automation, predictive analytics, and anomaly detection
- Describe the key concepts, architecture, and basic management principles of AI-ML clusters, as well as describe the process of acquiring, fine-tuning, optimizing and using pre-trained ML models
- Use the capabilities of Jupyter Lab and Generative AI to automate network operations, write Python code, and leverage AI models for enhanced productivity
- Describe the essential components and considerations for setting up robust AI infrastructure
- Evaluate and implement effective workload placement strategies and ensure interoperability within AI systems
- Explore compliance standards, policies, and governance frameworks relevant to AI systems
- Describe sustainable AI infrastructure practices, focusing on environmental and economic sustainability
- Guide AI infrastructure decisions to optimize efficiency and cost
- Describe key network challenges from the perspective of AI/ML application requirements
- Describe the role of optical and copper technologies in enabling AI/ML data center workloads
- Describe network connectivity models and network designs
- Describe important Layer 2 and Layer 3 protocols for AI and fog computing for Distributed AI processing
- Migrate AI workloads to dedicated AI network
- Explain the mechanisms and operations of RDMA and RoCE protocols
- Understand the architecture and features of high-performance Ethernet fabrics
- Explain the network mechanisms and QoS tools needed for building high-performance, lossless RoCE networks
- Describe ECN and PFC mechanisms, introduce Cisco Nexus Dashboard Insights for congestion monitoring, explore how different stages of AI/ML applications impact data center infrastructure, and vice versa
  Introduce the basic steps, challenges, and techniques regarding the data preparation process
- Use Cisco Nexus Dashboard Insights for monitoring AI/ML traffic flows
- Describe the importance of AI-specific hardware in reducing training times and supporting the advanced
- processing requirements of AI tasks
  Understand the computer hardware required to run AI/ML solutions
- Understand existing AI/ML solutions
- Describe virtual infrastructure options and their considerations when deploying
-

- Explain data storage strategies, storage protocols, and software-defined storage
- Use NDFC to configure a fabric optimized for AI/ML workloads
- Use locally hosted GPT models with RAG for network engineering tasks

Prerequisites

There are no prerequisites for this training. This is an essentials training that progresses from beginner to intermediate content. Familiarity with Cisco data center networking and computing solutions is a plus but not a requirement. However, the knowledge and skills you are recommended to have before attending this training are:

- Cisco UCS compute architecture and operations
- Cisco Nexus switch portfolio and features
- Data Center core technologies

These skills can be found in the following Cisco Learning Offerings:

- Introducing Cisco Unified Computing Systems (DCIUCS)
- Implementing Cisco NX-OS Switches and Fabrics in the Data Center (DCNX)
- Implementing Cisco Data Center Core Technologies (DCCOR)

Outline

- Fundamentals of AI
- Generative AI
- AI Use Cases
- AI-ML Clusters and Models
- AI Toolset Mastery - Jupyter Notebook
- AI Infrastructure
- AI Workload Placements and Interoperability
- AI Policies
- AI Sustainability
- AI Infrastructure Design
- Key Network Challenges and Requirements for AI Workloads
- AI Transport
- Connectivity Models
- AI Network
- Architecture Migration to AI/ML Network
- Application-Level Protocols
- High Throughput Converged Fabrics
- Building Lossless Fabrics
- Congestive Visibility
- Data Preparation for AI
- AI/ML Workload Data Performance
- AI-Enabling Hardware
- Compute Resources
- Compute Resource Solutions
- Virtual Resources
- Storage Resources

- Setting Up AI Cluster
- Deploy and Use Open Source GPT Models for RAG

Lab Outline

- AI Toolset—Jupyter Notebook
- AI/ML Workload Data Performance
- Setting Up AI Cluster
- Deploy and Use Open Source GPT Models for RAG

What to expect on the exam

Implementing Cisco Data Center AI Infrastructure (300-640 DCAI) v1.0 is a 90-minute exam associated with the Cisco Certified Specialist - Data Center AI Infrastructure certification and satisfies the concentration exam requirement for the CCNP® Data Center certification.

This exam tests your knowledge and skills related to design, implementation, monitoring, and troubleshooting of AI infrastructure, including:

- Network
- Compute
- Storage
- Orchestration solutions