

Implementing Cisco Data Center AI Infrastructure v1.0 (300-640)

Exam Description: Implementing Cisco Data Center AI Infrastructure v1.0 (DCAI 300-640) is a 90-minute examination associated with the CCNP Data Center Certification. This exam certifies a candidate's knowledge in the design, implementation, monitoring, and troubleshooting of AI infrastructure, including network, compute, storage, and orchestration solutions. The courses, Operate and Troubleshoot AI Solutions on Cisco Infrastructure (DCAIAOT) and AI Solutions on Cisco Infrastructure Essentials (DCAIE), help candidates prepare for this exam.

The following topics are general guidelines for the content likely to be included on the exam. However, other related topics may also appear on any specific delivery of the exam. To better reflect the contents of the exam and for clarity purposes, the guidelines below may change at any time without notice.

- 20% 1.0 AI Fundamentals and Applications**
 - 1.1 Describe AI/ML workload types
 - 1.1.a RAG
 - 1.1.b Training
 - 1.1.c Inference
 - 1.1.d Generative AI
 - 1.2 Describe the AI lifecycle
 - 1.3 Describe AI use cases
 - 1.4 Describe the types of AI infrastructure
 - 1.4.a Cloud
 - 1.4.b Hybrid
 - 1.4.c On-premises
 - 1.4.d Edge AI
 - 1.5 Describe the components used for AI environments
 - 1.5.a Network
 - 1.5.b Compute and GPUs deployment (NVLink)
 - 1.5.c Virtualization and containerization
 - 1.5.d Orchestration
 - 1.5.e Monitoring
 - 1.5.f Storage such as SAN, Fibre Channel, NVMe, Block and File
 - 1.6 Describe Cisco AI solutions
 - 1.6.a AI PODs
 - 1.6.b AI Canvas
 - 1.6.c Hyperfabric AI

- 30%** **2.0 AI Infrastructure Components and Architecture**
 - 2.1 Evaluate network deployment based on AI workload requirements such as bandwidth, latency, redundancy, scalability, and security
 - 2.2 Evaluate compute deployment based on AI workload requirements such as CPU resources, GPU resources and connectivity, memory, virtualization support, scalability, redundancy, and workload types
 - 2.3 Evaluate storage deployment based on AI workload requirements such as capacity, performance, redundancy and availability, and scalability
 - 2.4 Evaluate power, efficiency, and sustainability based on AI workload requirements such as power and cooling, power usage effectiveness, and renewable energy
 - 2.5 Evaluate hybrid AI deployment with cloud integration such as secure connectivity, data synchronization, and workload mobility

- 30%** **3.0 AI Infrastructure Deployment and Data Management**
 - 3.1 Configure high-performance networks to support AI workloads using Cisco Data Center
 - 3.1.a Congestion control mechanisms (PFC, ECN, ETS)
 - 3.1.b RDMA over Converged Ethernet (RoCE, RoCEv2)
 - 3.1.c Quality of service (QoS)
 - 3.1.d Load distribution

 - 3.2 Configure high-performance compute and storage to support AI workloads using Cisco UCS
 - 3.2.a Domain profiles
 - 3.2.b Power policy
 - 3.2.c Storage policies
 - 3.2.d LAN connectivity and vNIC policies
 - 3.2.e QoS policies and system classes
 - 3.2.f NTP policy

 - 3.3 Deploy AI-ready fabrics using Cisco orchestration tools
 - 3.3.a Nexus Dashboard
 - 3.3.b APIC
 - 3.3.c Hyperfabric
 - 3.3.d Intersight

- 20%** **4.0 AI Infrastructure Operations and Troubleshooting**
 - 4.1 Implement benchmarks to evaluate AI infrastructure performance
 - 4.2 Implement monitoring of AI data center infrastructures using Cisco solutions such as Nexus Dashboard and Intersight
 - 4.3 Monitor AI infrastructure using system messages and management tools to ensure reliability, scalability and performance
 - 4.3.a Operational telemetry
 - 4.3.b System health
 - 4.3.c Alerts
 - 4.3.d Log correlation

 - 4.4 Troubleshoot AI infrastructure using system messages and management tools