

AI Infrastructure and Operations Fundamentals

Duration: 5 Days

Day 1: AI Infrastructure Foundations

Module 1: Introduction to AI Infrastructure

- Understanding AI Infrastructure and Enterprise AI Ecosystem
- AI vs Traditional IT Infrastructure
- AI/ML Lifecycle: Data Ingestion, Data Preparation, Training, Validation, Deployment, Inference, Monitoring
- Core Components of AI Infrastructure
- Enterprise AI Reference Architecture
- Infrastructure Requirements for AI Workloads
- Scalability, Availability and Resilience Principles

Hands-on Lab: Design an enterprise AI infrastructure and map the AI/ML lifecycle.

Module 2: AI Deployment Environments

- On-Premises AI Infrastructure
- Cloud AI Infrastructure Concepts
- Hybrid AI Architecture
- Choosing the Right Deployment Model
- High Availability and Disaster Recovery
- Infrastructure Design Best Practices

Hands-on Lab: Compare on-premises, cloud and hybrid architectures.

Day 2: Compute, Storage and Data Management

Module 3: Compute Infrastructure

- CPU, GPU and AI Accelerators
- High Performance Computing Concepts
- GPU Resource Allocation
- Workload Scheduling
- Compute Sizing and Capacity Planning
- Resource Utilization and Optimization

Hands-on Lab: Configure compute resources and analyze utilization.

Module 4: Storage and Data Pipelines

- Object, Block and Distributed Storage
- Large Dataset Management
- Data Ingestion
- Data Preprocessing Pipelines
- Pipeline Orchestration
- Data Governance
- Data Quality
- Integration with Analytics Platforms

Hands-on Lab: Build a sample data ingestion and preprocessing pipeline.

Day 3: AI Workload Deployment

Module 5: Model Deployment

- Model Packaging
- Model Registry and Versioning
- Deployment Strategies
- Rollback and Release Management
- Managing Model Lifecycle

Hands-on Lab: Package, version and deploy an AI model.

Module 6: Containerization and Orchestration

- Docker Fundamentals
- Container Images and Registries
- Kubernetes Architecture
- Pods, Deployments and Services
- Scaling AI Applications
- Production Deployment Best Practices

Hands-on Lab: Containerize an AI application using Docker and deploy on Kubernetes.

Day 4: AI Operations (MLOps & AIOps)

Module 7: MLOps Fundamentals

- Introduction to MLOps
- CI/CD for AI Models
- Training Pipelines
- Deployment Automation
- Workflow Automation

- Model Validation

Hands-on Lab: Create a basic AI CI/CD workflow.

Module 8: Monitoring and AIOps

- Monitoring AI Systems
- Logging and Metrics
- Alerting
- Performance Dashboards
- AIOps Concepts
- Predictive Operations
- Automated Incident Detection

Hands-on Lab: Configure monitoring dashboards and alerts.

Day 5: Performance, Security and Operations

Module 9: Performance Optimization

- GPU Optimization
- Horizontal and Vertical Scaling
- Load Balancing Concepts
- Infrastructure Performance Tuning
- Cloud and Hybrid Cost Optimization

Hands-on Lab: Optimize AI workload performance and scaling.

Module 10: Security, Compliance and Troubleshooting

- Securing AI Infrastructure
- Identity and Access Management (IAM) Concepts
- Authentication and Authorization
- Least Privilege Principle
- Secrets Management
- Data Protection
- Compliance and Governance
- Performance Monitoring
- Capacity Planning
- Incident Management
- Root Cause Analysis

Hands-on Lab: Secure an AI application and troubleshoot operational issues.