

AI-200T00-Develop AI cloud solutions on Azure

Overview

Course Duration: 5 Days

About This Course

This course teaches developers how to create, monitor, and troubleshoot AI solutions on Microsoft Azure. Students will learn how to implement Azure compute and containerization patterns to host applications, build serverless APIs with Azure Functions, and integrate services using event-driven and message-based architectures such as Azure Service Bus and Event Grid. The course also covers working with Azure data services that support AI workloads, including designing and querying solutions with Cosmos DB for NoSQL, Azure Database for PostgreSQL with pgvector, and Azure Managed Redis for caching, streaming, and vector search. By the end of the course, developers will be able to connect services, orchestrate AI workflows, and build secure, scalable, and observable AI-driven applications on Azure.

Audience Profile

This course is designed for developers who build backend and AI-driven applications on Azure and need practical skills in containerized compute, data services for AI, event-driven workflows, and application security and monitoring.

Course Outline

Module 1: Implement container application hosting on Azure

This learning path guides you through core container hosting workflows on Azure for AI applications and backend services. You use Azure Container Registry to store and organize images, build images in the cloud with ACR Tasks, and apply tagging and versioning practices that support reliable deployments. From there, you deploy custom containers to Azure App Service, configure runtime behavior such as ports, startup commands, and persistent storage, and externalize environment-specific configuration using application settings.

- Store and manage containers in Azure Container Registry
- Deploy containers to Azure App Service

Module 2: Deploy and manage apps on Azure Container Apps

This learning path guides you through the complete lifecycle of running containerized applications on Azure Container Apps. You start by deploying container apps to environments, configuring runtime settings with environment variables and secrets, and setting up registry authentication. You then manage the day-two lifecycle by updating images, managing revisions, monitoring logs, and configuring health probes. Finally, you learn to configure automatic horizontal scaling using HTTP rules, KEDA scalers, and traffic management to optimize performance and cost.

- Deploy containers to Azure Container Apps
- Manage containers in Azure Container Apps
- Scale containers in Azure Container Apps

Module 3: Deploy and monitor applications on Azure Kubernetes Service

This learning path guides you through the complete lifecycle of running applications on Azure Kubernetes Service. You start by creating deployment manifests and exposing applications with Kubernetes Services. You then externalize configuration using ConfigMaps, secure sensitive settings with Secrets, and attach persistent storage for stateful workloads. Finally, you learn to monitor application health using logs and metrics, troubleshoot pod and Service issues, and verify connectivity paths to ensure reliable access to your applications.

- Deploy applications to Azure Kubernetes Service
- Configure applications on Azure Kubernetes Service
- Monitor and troubleshoot applications on Azure Kubernetes Service

Module 4: Develop AI solutions with Azure Cosmos DB for NoSQL

This learning path guides you through developing AI solutions using Azure Cosmos DB for NoSQL. You start by building a data foundation with the Cosmos DB resource model, SDK integration, CRUD operations, and SQL queries to retrieve document data for AI applications.

You then implement vector search capabilities to store embeddings, execute similarity queries using the VectorDistance function, combine vector search with metadata filters and hybrid search, and use the change feed to keep embeddings synchronized.

Finally, you optimize query performance by analyzing query patterns, configuring range and composite indexes, selecting vector index types, and choosing consistency levels that balance freshness with cost efficiency.

- Build queries for Azure Cosmos DB for NoSQL
- Implement vector search on Azure Cosmos DB for NoSQL
- Optimize query performance for Azure Cosmos DB for NoSQL

Module 5: Develop AI solutions with Azure Database for PostgreSQL

This learning path guides you through developing AI solutions using Azure Database for PostgreSQL. You start by building a data foundation with schema design, efficient SQL queries, and secure Python integration using Microsoft Entra authentication.

You then implement vector search using the pgvector extension to store embeddings, execute similarity searches with different distance metrics, and build retrieval patterns that integrate with RAG pipelines for semantic search and recommendations.

Finally, you optimize vector search performance by tuning PostgreSQL and pgvector configuration, selecting appropriate vector indexes, designing efficient data layouts, scaling for high-volume workloads, and implementing connection pooling for AI applications.

- Build and query with Azure Database for PostgreSQL
- Implement vector search with Azure Database for PostgreSQL
- Optimize vector search in Azure Database for PostgreSQL

Module 6: Enhance AI solutions with Azure Managed Redis

Learn how to use Azure Managed Redis to enhance your AI solutions, including caching strategies, data operations, event messaging, and vector storage.

- Implement data operations in Azure Managed Redis
- Implement event messaging with Azure Managed Redis
- Implement vector storage in Azure Managed Redis

Module 7: Integrate backend services for AI solutions

This learning path teaches you how to build and integrate backend services that support AI solutions on Azure. You start by using Azure Service Bus to decouple AI application components, queue inference requests, and process messages reliably with queues, topics, and dead-letter queues. You then build event-driven workflows with Azure Event Grid to route events between services with low latency, configure delivery policies, and publish custom events from AI applications. Finally, you create serverless AI backends with Azure Functions, including inference endpoints, event processors, and secure integrations with other Azure services.

- Queue and process AI operations with Azure Service Bus

- Develop event-driven AI workflows with Azure Event Grid
- Build serverless AI backends with Azure Functions

Module 8: Manage application secrets and configuration for AI solutions

This learning path teaches you how to securely manage secrets and centralize configuration for AI solutions on Azure. You start by using Azure Key Vault to store, organize, and retrieve secrets with managed identity authentication, handle secret versioning and rotation for zero-downtime credential updates, and implement caching strategies that reduce API calls while maintaining credential freshness. You then use Azure App Configuration to centralize application settings, organize key-value pairs with labels for environment-specific variants, manage feature flags for controlled rollouts, and reference Key Vault secrets so the application retrieves configuration and secrets through a single path.

- Manage application secrets with Azure Key Vault
- Manage application settings with Azure App Configuration

Module 9: Observe and troubleshoot apps on Azure

This learning path teaches you how to gain end-to-end observability into distributed AI applications on Azure. You start by instrumenting applications with OpenTelemetry to capture distributed traces, create custom spans, and export telemetry to Azure Monitor Application Insights. You then analyze the collected telemetry by writing KQL queries, exploring error patterns and performance trends, building dashboards and workbooks for operational visibility, and configuring alerts to detect failures and anomalies.

- Instrument an app with OpenTelemetry
- Analyze app telemetry with logs and metrics

Prerequisites

- Programming experience with languages such as Python, JavaScript, or C#.
- Basic understanding of Azure services and cloud computing concepts.
- Familiarity with container concepts.
- Familiarity with command-line tools including Azure CLI.
- Familiarity with containerization and Kubernetes fundamentals.
- Access to Azure Kubernetes Service and related tools for practical exercises.
- Basic understanding of Azure services and cloud computing concepts.
- Familiarity with JSON document structures.
- Understanding of machine learning concepts including embeddings and similarity search.
- Familiarity with relational databases and SQL fundamentals.
- Understanding of machine learning concepts including embeddings and similarity search.
- Knowledge of caching concepts and data structures.
- Familiarity with AI embeddings and vector similarity concepts.
- Basic understanding of message queuing patterns and when to use them for decoupling services.
- Access to Azure Managed Redis and related tools for hands-on practice.
- Familiarity with distributed systems concepts and event-driven architectures.
- Familiarity with REST APIs and asynchronous messaging patterns.
- Familiarity with distributed application architectures.