Azure Databricks Platform Architect

Duration: 24 hours

Course Objectives

- Gain an understanding of Azure Databricks architecture and its key components.
- Learn platform administration fundamentals, including workspace management, compute resources, and security.
- Explore Unity Catalog for data governance and access control.
- Understand how to manage compute resources and data access efficiently.
- Get hands-on experience in administering an Azure Databricks environment.

Course Contents

Module 1: Azure Databricks

Topics:

- Introduction to Azure Databricks
- Key features and benefits
- Databricks vs. other cloud-based analytics platforms
- Integration with Azure services (Azure Data Lake, Synapse, Key Vault, etc.)
- Pricing and cluster cost management

Demo/Lab:

- Deploy an Azure Databricks workspace
- Explore the Azure Databricks portal and workspace UI

Module 2: Platform Administration Fundamentals

Topics:

- Azure Databricks account and workspace structure
- Managing users, groups, and roles in Azure Databricks
- Managing workspace settings and permissions
- Best practices for multi-user environments

- High availability and disaster recovery considerations

Demo/Lab:

- Set up user roles and permissions
- Configure workspace settings and security policies

Module 3: Databricks Overview

Submodule 3.1: Databricks Infrastructure

Topics:

- Underlying cloud infrastructure and networking
- Cluster types: Standard, High Concurrency, and SQL Pro
- Auto-scaling and performance optimization

Demo/Lab:

- Create and configure an autoscaling cluster

Submodule 3.2: Databricks Data Intelligence Platform

Topics:

- What is Databricks Data Intelligence?
- AI-powered data management and analytics
- Integration with ML models and AI workloads

Demo/Lab:

- Run a simple AI-based data transformation in Databricks

Submodule 3.3: Unity Catalog Overview

Topics:

- What is Unity Catalog?
- Benefits of Unity Catalog for governance
- Multi-cloud and cross-workspace data access

Demo/Lab:

- Enable Unity Catalog and configure a metastore

Submodule 3.4: Databricks Workspace Walkthrough

Topics:

- Navigating the Databricks workspace
- Managing notebooks, jobs, and clusters
- Working with Databricks Repos

Demo/Lab:

- Explore Databricks UI and create a notebook

Module 4: Databricks Platform Administration

Submodule 4.1: Data Governance in Unity Catalog

Topics:

- Importance of data governance
- How Unity Catalog simplifies governance
- Managing lineage and audit logs

Demo/Lab:

- Set up and explore lineage tracking in Unity Catalog

Submodule 4.2: Managing Principles in Unity Catalog

Topics:

- Defining principles (users, groups, service principals)
- Assigning roles and permissions
- Least privilege access model

Demo/Lab:

- Create and assign Unity Catalog roles

Submodule 4.3: Managing Unity Catalog Metastores

Topics:

- What is a metastore?
- Centralized vs. workspace-specific metastores
- Best practices for managing metastores

Demo/Lab:

- Set up and manage a Unity Catalog metastore

Submodule 4.4: Compute Resources and Unity Catalog

Topics:

- Compute resources in Azure Databricks
- Linking compute clusters with Unity Catalog
- Performance tuning and cluster policies

Demo/Lab:

- Configure a cluster with Unity Catalog

Submodule 4.5: Data Access Control in Unity Catalog

Topics:

- Row-level and column-level security
- Fine-grained access control in Unity Catalog
- Managing data access for different workloads

Demo/Lab:

- Implement row-level and column-level security in Unity Catalog

---

Final Project (Optional but Recommended)

- Set up an Azure Databricks workspace with Unity Catalog
- Configure governance policies, access control, and compute resources
- Run a sample data processing workflow with governance policies applied

---