Databricks Lakehouse Fundamentals

Duration: 16 hours

Course Objectives:

- Understand the evolution and significance of Databricks.

- Learn the fundamentals of the Databricks Lakehouse architecture.

- Explore security, governance, and best practices in Databricks.

- Gain hands-on experience with Databricks products and features.

- Understand supported workloads and use cases.

Module 1: The Databricks Story

Topics:

- Founding of Databricks and its relation to Apache Spark

- Evolution from big data to the Lakehouse paradigm

- Industry adoption and key customers

- Strategic partnerships (Azure, AWS, Google Cloud)

Demo/Lab:

- Walkthrough of the Databricks platform UI and history page

Module 2: Data Lakehouses and Data Intelligence

Topics:

- Traditional data architectures (Data Warehouses vs. Data Lakes)

- What is a Data Lakehouse? (combining lakes & warehouses)

- Benefits of a Data Lakehouse over traditional architectures

- Role of AI and data intelligence in a Lakehouse

Demo/Lab:

- Compare query speeds on a Data Warehouse, Data Lake, and a Lakehouse

---

Module 3: Intro to Data Lakehouse

Topics:

- Why organizations need a Lakehouse

- Key principles of a Lakehouse

- Real-world use cases of Data Lakehouse

Demo/Lab:

- Compare a traditional ETL workflow vs. a Lakehouse-based workflow

---

Module 4: Intro to Databricks Lakehouse Platform

Topics:

- Overview of Databricks' ecosystem

- How it integrates with cloud platforms

- Unified data analytics and AI on the Lakehouse

Demo/Lab:

- Explore Databricks Workspace and UI

Module 5: Intro to Databricks Lakehouse Platform Architecture and Security Fundamentals

Topics:

- Architecture deep dive: Compute, storage, and security layers

- Databricks Data Governance (Unity Catalog, access controls)

- Compliance and best practices

Demo/Lab:

- Configure workspace security settings

- Implement row-level security with Unity Catalog

Module 6: Intro to Supported Workloads on the Databricks Lakehouse Platform

Topics:

- SQL analytics and BI

- Data engineering and ETL
- Machine learning and AI workloads
- Real-time streaming analytics

Demo/Lab:

- Execute a real-time data streaming pipeline in Databricks

Module 7: Databricks Architecture

Topics:

- Components of Databricks architecture
  - Compute and Clusters
  - Storage layers
  - Delta Lake and its role in the architecture
- How Databricks abstracts cloud storage (Azure, AWS, GCP)

Demo/Lab:

- Hands-on session creating a Databricks cluster
- Run a simple Spark job

Module 8: Databricks Security and Governance

Topics:

- Identity and Access Management (IAM) in Databricks
- Role-based access control (RBAC)
- Data security and encryption mechanisms
- Unity Catalog for data governance
- Compliance with industry standards

Demo/Lab:

- Set up user roles and permissions in Databricks
- Create and use Unity Catalog for governance

Module 9: Databricks Products and Features

Topics:

- Databricks SQL: Running SQL queries on Databricks

- Databricks Data Science & Engineering

- Databricks Machine Learning

- Integration with BI tools (Power BI, Tableau, etc.)

Demo/Lab:

- Explore Databricks SQL and run basic queries

- Create a simple ML pipeline using Databricks ML

Final Project (Optional but Recommended)

- Build an end-to-end Data Lakehouse solution in Databricks

- Ingest, transform, and analyze data using Delta Lake

- Secure and govern data using Unity Catalog