Data Engineering with Databricks Training

Duration: **2 Day**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

Overview

Data Engineering with Databricks provides data professionals from a range of disciplines with a comprehensive introduction to the Databricks Lakehouse Platform components that support bringing ETL pipelines into production.

Throughout the course, participants use SQL and Python to define and schedule pipelines that incrementally process newly arriving data from multiple sources. These pipelines are designed to support analytics applications and dashboards running in the lakehouse environment.

The course includes hands-on work with key Databricks technologies, including Databricks Data Science and Engineering Workspace, Databricks SQL, Delta Live Tables, Databricks Repos, Databricks Task Orchestration, and Unity Catalog.

The training also supports preparation for the **Databricks Certified Data Engineer Associate** certification exam.

Prerequisites

Participants should have:

- Beginner-level familiarity with basic cloud concepts such as virtual machines, object storage, and identity management,
- The ability to complete basic code development tasks, including creating compute instances, running code in notebooks, performing standard notebook operations, and importing repositories from Git,
- Intermediate SQL knowledge, including commands such as CREATE, SELECT, INSERT, UPDATE, DELETE, GROUP BY, and JOIN,
- Intermediate experience with SQL concepts such as aggregate functions, filters, sorting, indexes, tables, and views,
- Basic knowledge of Python programming, the Jupyter Notebook interface, and PySpark fundamentals.

Participants who do not yet meet one or more of these prerequisites may benefit from completing introductory training in SQL querying and Python-based data handling.

Who Should Attend

This course is designed for:

- Data Engineers who want to strengthen their Databricks and Delta Lake expertise,
- Data Analysts looking to develop their knowledge of data pipelines and transformation,
- Cloud Engineers and Developers working with big data frameworks,
- Professionals preparing for the Databricks Associate Data Engineering certification.

Data Engineering with Databricks Training

Duration: **2 Day**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

What You Will Learn

By the end of this course, participants will be able to:

- Use the Databricks Lakehouse Platform to perform core responsibilities involved in data pipeline development,
- Build production data pipelines with SQL and Python to extract, transform, and load data into lakehouse tables and views,
- Simplify data ingestion and incremental change propagation using Databricks-native features and syntax, including Delta Live Tables,
- Orchestrate production pipelines that deliver up-to-date results for ad hoc analytics and dashboarding.

Outline

Day 1

Delta Lake

- Explore the fundamentals of Delta Lake and its role in data engineering workflows

Relational Entities on Databricks

- Understand relational entities within the Databricks environment

ETL with Spark SQL

- Use Spark SQL to perform extract, transform, and load operations

Just Enough Python for Spark SQL

- Apply the essential Python knowledge required to support Spark SQL workflows

Incremental Data Processing with Structured Streaming and Auto Loader

- Process newly arriving data incrementally
- Work with Structured Streaming
- Use Auto Loader to manage data ingestion

Day 2

Medallion Architecture in the Data Lakehouse

- Understand how Medallion Architecture is applied within a lakehouse environment

Delta Live Tables

- Build and manage data pipelines using Delta Live Tables

Data Engineering with Databricks Training

Duration: **2 Day**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

Task Orchestration with Databricks Jobs

- Use Databricks Jobs to orchestrate data tasks and pipeline workflows

Databricks SQL

- Use Databricks SQL for analytics queries and data operations

Managing Permissions in the Lakehouse

- Manage access to data and platform resources within the lakehouse

Productionizing Dashboards and Queries on Databricks SQL

- Prepare dashboards and queries for production use
- Deliver analytical results in a reliable and repeatable way