Data Engineering with Databricks

Course ID #: 7000-1005-ZZ-Z
Hours: 16

Course Content

Course Description:

Data professionals from all walks of life will benefit from this comprehensive introduction to the components of the Databricks Lakehouse Platform that directly support putting ETL pipelines into production. You will leverage SQL and Python to define and schedule pipelines that incrementally process new data from a variety of data sources to power analytic applications and dashboards in the Lakehouse. This course offers hands-on instruction in Databricks Data Science & Engineering Workspace, Databricks SQL, Delta Live Tables, Databricks Repos, Databricks Task Orchestration, and the Unity Catalog.

Prerequisites:

Before attempting this course, please ensure that you meet these prerequisites. If you don't, you might have trouble following along with course content.

Note that this course provides context for how to perform the actions within it using both Spark SQL and PySpark. If you are interested in one over the other, feel free to focus on that content (and skip the videos that are not relevant to you).

If you plan to follow along in Spark SQL, here are your pre-requisites:

- Beginner familiarity with cloud computing concepts (virtual machines, object storage, etc.)
- Production experience working with data warehouses and data lakes
- Intermediate experience with basic SQL concepts (select, filter, groupby, join, etc)

If you plan on following along in both Spark SQL and PySpark, please see these additional pre-requisites:

- Beginner programming experience with Python (syntax, conditions, loops, functions)
- Beginner programming experience with the Spark DataFrame API (PySpark)
- Configure DataFrameReader and DataFrameWriter to read and write data (PySpark)
- Express query transformations using DataFrame methods and Column expressions (PySpark)
- Navigate the Spark documentation to identify built-in functions for various transformations and data types (PySpark)

Data Engineering with Databricks

Course ID #: 7000-1005-ZZ-Z

Hours: 16

Topics:

Lesson 1: Get Started with Databricks Data Science and Data Engineering Workspace

Lesson 2: Transform Data with Spark

Lesson 3: Manage Data with Delta Lake

Lesson 4: Build Data Pipelines with Delta Live Tables

Lesson 5: Deploy Workloads with Databricks Workflows

Lesson 6: Manage Data Access with Unity Catalog

Register for this class by visiting us at:
www.tcworkshop.com or calling us at 800-639-3535