

DP 3011- Implementing a Data Analytics Solution with Azure Databricks Training



Days: 1

Audience/Prerequisites: Before taking this course, learners should already be comfortable with the fundamentals of Python and SQL. This includes being able to write simple Python scripts and work with common data structures, as well as writing SQL queries to filter, join, and aggregate data. A basic understanding of common file formats such as CSV, JSON, or Parquet will also help when working with datasets. In addition, familiarity with the Azure portal and core services like Azure Storage is important, along with a general awareness of data concepts such as batch versus streaming processing and structured versus unstructured data. While not mandatory, prior exposure to big data frameworks like Spark, and experience working with Jupyter notebooks, can make the transition to Databricks smoother.

Description: This course explores how to use Databricks and Apache Spark on Azure to take data projects from exploration to production. You'll learn how to ingest, transform, and analyze large-scale datasets with Spark DataFrames, Spark SQL, and PySpark, while also building confidence in managing distributed data processing. Along the way, you'll get hands-on with the Databricks workspace—navigating clusters and creating and optimizing Delta tables. You'll also dive into data engineering practices, including designing ETL pipelines, handling schema evolution, and enforcing data quality. The course then moves into orchestration, showing you how to automate and manage workloads with Lakeflow Jobs and pipelines. To round things out, you'll explore governance and security capabilities such as Unity Catalog and Purview integration, ensuring you can work with data in a secure, well-managed, and production-ready environment.

OUTLINE:

LESSON 1 - EXPLORE AZURE DATABRICKS

- Get started with Azure Databricks
- Identify Azure Databricks workloads
- Understand key concepts
- Data governance using Unity Catalog and Microsoft Purview
- Module assessment

LESSON 2 - PERFORM DATA ANALYSIS WITH AZURE DATABRICKS

- Ingest data with Azure Databricks
- Data exploration tools in Azure Databricks
- Data analysis using DataFrame APIs
- Module assessment

LESSON 3 - USE APACHE SPARK IN AZURE DATABRICKS

- Get to know Spark
- Create a Spark cluster
- Use Spark in notebooks
- Use Spark to work with data files
- Visualize data
- Module assessment

LESSON 4 - MANAGE DATA WITH DELTA LAKE

- Get started with Delta Lake
- Create Delta tables
- Implement schema enforcement
- Data versioning and time travel in Delta Lake
- Data integrity with Delta Lake
- Module assessment

LESSON 5 - BUILD LAKEFLOW DECLARATIVE PIPELINES

- Explore Lakeflow Declarative Pipelines
- Data ingestion and integration
- Real-time processing
- Module assessment

LESSON 6 - DEPLOY WORKLOADS WITH LAKEFLOW JOBS

- What are Lakeflow Jobs?
- Understand key components of Lakeflow Jobs
- Explore the benefits of Lakeflow Jobs
- Deploy workloads using Lakeflow Jobs
- Module assessment