

Implementing a Data Analytics Solution with Azure Databricks

Course DP-3011

1 Days

Instructor-led, Hands-on

Introduction

This is a single day Instructor-led course designed to teach how to harness the power of Apache Spark and powerful clusters running on the Azure Databricks platform to run large data engineering workloads in the cloud.

At Course Completion

Upon successful completion of this course, students will have the following skills and knowledge:

- Explore Azure Databricks
- Use Apache Spark in Azure Databricks
- Use Delta Lake in Azure Databricks
- Use SQL Warehouses in Azure Databricks
- Run Azure Databricks Notebooks with Azure Data Factory

Prerequisites

Successful students start this course with familiarity with notebooks using diverse languages and a Spark engine. Experience with SQL, Python, and Azure tools such as Data Factory.

Outline

Module 1: Explore Azure Databricks

- Get started with Azure Databricks
- Identify Azure Databricks workloads
- Understand key concepts

Module 2: 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 3: Use Delta Lake in Azure Databricks

- Get Started with Delta Lake
- Create Delta Lake tables
- Create and query catalog tables
- Use Delta Lake for streaming data

Module 4: Use SQL Warehouses in Azure Databricks

- Get started with SQL Warehouses
- Create databases and tables
- Create queries and dashboards

Module 5: Run Azure Databricks Notebooks with Azure Data Factory

- Understand Azure Databricks notebooks and pipelines
- Create a linked service for Azure Databricks
- Use a Notebook activity in a pipeline
- Use parameters in a notebook