



# Implementing a Data Analytics Solution with Azure Databricks

## Course DP-3011: 1 day; Intermediate; Instructor-Led

### Introduction

Learn 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.

### Prerequisites

- None

### Course Outline

#### Module 1: Explore Azure Databricks

Azure Databricks is a cloud service that provides a scalable platform for data analytics using Apache Spark.

In this module, you'll learn how to:

- Provision an Azure Databricks workspace.
- Identify core workloads and personas for Azure Databricks.
- Describe key concepts of an Azure Databricks solution.

#### Prerequisites

- Before starting this module, you should have a fundamental knowledge of data analytics concepts.

#### Lessons

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

#### Module 2: Use Apache Spark in Azure Databricks

Azure Databricks is built on Apache Spark and enables data engineers and analysts to run Spark jobs to transform, analyze and visualize data at scale.

In this module, you'll learn how to:

- Describe key elements of the Apache Spark architecture.
- Create and configure a Spark cluster.
- Describe use cases for Spark.
- Use Spark to process and analyze data stored in files.
- Use Spark to visualize data.

#### Prerequisites

- Before starting this module, you should have a basic knowledge of Azure Databricks.

#### Lessons

- Get to know Spark
- Create a Spark cluster
- Use Spark in notebooks
- Use Spark to work with data files
- Visualize data
- Exercise - Use Spark in Azure Databricks

### Module 3: Use Delta Lake in Azure Databricks

Delta Lake is an open source relational storage area for Spark that you can use to implement a data lakehouse architecture in Azure Databricks.

In this module, you'll learn how to:

- Describe core features and capabilities of Delta Lake.
- Create and use Delta Lake tables in Azure Databricks.
- Create Spark catalog tables for Delta Lake data.
- Use Delta Lake tables for streaming data.

#### Prerequisites

- Before starting this module, you should know how to use Apache Spark in Azure Databricks. Consider completing the Use Apache Spark in Azure Databricks module before this one.

#### Lessons

- Get Started with Delta Lake
- Create Delta Lake tables
- Create and query catalog tables
- Use Delta Lake for streaming data
- Exercise - Use Delta Lake in Azure Databricks

### Module 4: Use SQL Warehouses in Azure Databricks

Azure Databricks provides SQL Warehouses that enable data analysts to work with data using familiar relational SQL queries.

In this module, you'll learn how to:

- Create and configure SQL Warehouses in Azure Databricks.
- Create databases and tables.
- Create queries and dashboards.

#### Prerequisites

- Before starting this module, you should have a basic knowledge of Azure Databricks.

#### Lessons

- Get started with SQL Warehouses
- Create databases and tables
- Create queries and dashboards
- Exercise - Use a SQL Warehouse in Azure Databricks

### Module 5: Run Azure Databricks Notebooks with Azure Data Factory

Using pipelines in Azure Data Factory to run notebooks in Azure Databricks enables you to automate data engineering processes at cloud scale.

In this module, you'll learn how to:

- Describe how Azure Databricks notebooks can be run in a pipeline.
- Create an Azure Data Factory linked service for Azure Databricks.
- Use a Notebook activity in a pipeline.
- Pass parameters to a notebook.

#### Prerequisites

- Before starting this module, you should have a basic knowledge of Azure Databricks.

#### Lessons

- 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
- Exercise - Run an Azure Databricks Notebook with Azure Data Factory