DP-3011: Implement a Data Analytics Solution with Azure Databricks

Course Code: DP-3011
Duration: 1 day
Instructor-led Training (ILT) | Virtual Instructor-led Training (VILT)

OVERVIEW

Implement a Data Analytics Solution with Azure Databricks

Learn how to harness the power of Apache Spark and powerful clusters running on the Azure Databricks platform to run data analytics workloads in a data lakehouse.

SKILLS COVERED

- 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

WHO SHOULD ATTEND?

- It Professionals

PREREQUISITES

There are no prerequisites required to attend this course.

MODULES

Module 1: Explore Azure Databricks

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

Learning objectives
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. Consider completing Azure Data Fundamentals certification before starting this module.

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.

Learning objectives

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. Consider completing the Explore Azure Databricks module before this one.

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.

Learning objectives

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.

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.

Learning objectives

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. Consider completing the Explore Azure Databricks module before this one.

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.

Learning objectives

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. Consider completing the Explore Azure Databricks module before this one.

END OF PAGE