

MS-DP3011: IMPLEMENTING A DATA ANALYTICS SOLUTION WITH AZURE DATABRICKS



DURATION	LEVEL	TECHNOLOGY	DELIVERY METHOD	TRAINING CREDITS
1 Day	Intermediate	Azure Data	Instructor-led	NA

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.

AUDIENCE PROFILE

The primary audience includes:

- Data Scientists: Individuals who build and deploy machine learning models using large datasets.
- Machine Learning Engineers: Professionals who focus on the engineering aspects of machine learning, including model deployment and monitoring.
- Data Engineers: Those responsible for preparing and managing data for machine learning projects.
- AI Engineers: Engineers who design and implement AI solutions using machine learning models.
- Developers: Software developers who integrate machine learning models into applications.
- IT Professionals: Individuals who support the infrastructure and deployment of machine learning models.

PREREQUISITES

Before attending this course, delegates must have:

- Basic understanding of data analytics concepts and techniques.
- Familiarity with cloud computing concepts and Azure services.

COURSE OBJECTIVES

After completion of this course, you will be able to:

- Provision an Azure Databricks Workspace: Learn to set up and configure an Azure Databricks workspace and cluster.
- Use Apache Spark in Azure Databricks: Understand how to use Apache Spark for data processing and analysis.
- Prepare Data for Machine Learning: Learn techniques for data ingestion, preparation, and preprocessing.
- Train Machine Learning Models: Use Azure Databricks to train machine learning models with various frameworks like Scikit-Learn, PyTorch, and TensorFlow2.
- Track Experiments with MLflow: Utilize MLflow to log parameters, metrics, and manage the lifecycle of machine learning models.
- Tune Hyperparameters: Optimize model performance by tuning hyperparameters using libraries like Hyperopt.
- Use AutoML: Simplify the process of building machine learning models with Azure Databricks' AutoML capabilities.

COURSE CONTENT

Module 1: Explore Azure Databricks

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

Lessons

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

- Knowledge check.
- Summary.

By the end of this module, you'll be able to:

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

Module 2: Perform data analysis with Azure Databricks

Learn how to perform data analysis using Azure Databricks. Explore

various data ingestion methods and how to integrate data from sources like Azure Data Lake and Azure SQL Database. This module guides you through using collaborative notebooks to perform exploratory data analysis (EDA), so you can visualize, manipulate, and examine data to uncover patterns, anomalies, and correlations.

Lessons

- Introduction.
- Ingest data with Azure Databricks.
- Data exploration tools in Azure Databricks.

- Data analysis using DataFrame APIs.
 - Exercise - Explore data with Azure Databricks.
 - Knowledge check.
 - Summary.
- By the end of this module, you'll be able to:
- Ingest data using Azure Databricks.
 - Using the different data exploration tools in Azure Databricks.
 - Analyze data with DataFrame APIs.

Module 3: 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.

Lessons

- Introduction.
- 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.
- Knowledge check.
- Summary.

By the end of this module, you'll be able 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.

Module 4: Manage data with Delta Lake

Delta Lake is a data management solution in Azure Databricks

providing features including ACID transactions, schema enforcement, and time travel ensuring data consistency, integrity, and versioning capabilities.

Lessons

- Introduction.
- Get Started with Delta Lake.
- Manage ACID transactions.
- Implement schema enforcement.
- Data versioning and time travel in Delta Lake.
- Data integrity with Delta Lake.
- Exercise – Use Delta Lake in Azure Databricks.
- Knowledge check.
- Summary.

In this module, you learn:

- What Delta Lake is.
- How to manage ACID transactions using Delta Lake.
- How to use schema versioning and time travel in Delta Lake.
- How to maintain data integrity with Delta Lake.

Module 5: Build data pipelines with Delta Live Tables

Building data pipelines with Delta Live Tables enables real-time, scalable, and reliable data processing using Delta Lake's advanced features in Azure Databricks

Lessons

- Introduction.
- Explore Delta Live Tables.
- Data ingestion and integration.
- Real-time processing.
- Exercise - Create a data pipeline with Delta Live Tables.
- Knowledge check.
- Summary.

By the end of this module, you'll be able to:

- Describe Delta Live Tables.
- Ingest data into Delta Live Tables.
- Use Data Pipelines for real time data processing.

Module 6: Deploy workloads with Azure Databricks Workflows

Deploying workloads with Azure Databricks Workflows involves orchestrating and automating complex data processing pipelines, machine learning workflows, and analytics tasks. In this module, you learn how to deploy workloads with Databricks Workflows.

Lessons

- Introduction.
- What are Azure Databricks Workflows?
- Understand key components of Azure Databricks Workflows.
- Explore the benefits of Azure Databricks Workflows.
- Deploy workloads using Azure Databricks Workflows.
- Exercise - Create an Azure Databricks Workflow.
- Knowledge check.
- Summary.

In this module, you learn:

- What Azure Databricks Workflows are.
- The key components and benefits of Azure Databricks Workflows.
- How to deploy workloads using Azure Databricks Workflows.

ASSOCIATED CERTIFICATIONS & EXAM

There is no Associated Certification or Exam for this course.