

**Implementing a Data Analytics Solution with Azure Databricks (DP-3011)**

**Course Number:** MOC-DP-3011
**Duration:** 1 day

**Overview**

This live, hands-on, official Microsoft course DP-3011, Data Analytics with Azure Databricks training, teaches attendees how to leverage Apache Spark and powerful clusters on the Azure Databricks platform to run large data engineering workloads in the cloud.

**Prerequisites**

None.

**Materials**

Attendees will not need to install any software on their computers for this class. The class will be conducted in a remote environment that Accelebrate will provide; students will only need a local computer with a web browser and a stable Internet connection. Any recent version of Microsoft Edge, Mozilla Firefox, or Google Chrome will work well.

**Software Needed on Each Student PC**

All Microsoft Azure Databricks training students receive Microsoft official courseware.

For all Microsoft Official Courses taught in their entirety that have a corresponding certification exam, an exam voucher is included for each participant.

**Objectives**

* Explore the core features and capabilities of Azure Databricks, including its architecture, workloads, and key concepts
* Perform data analysis tasks using Azure Databricks, including data ingestion, exploration, and visualization
* Utilize Apache Spark within Azure Databricks to process and analyze large datasets efficiently
* Implement Delta Lake for robust data management, ensuring data integrity, consistency, and versioning capabilities
* Construct data pipelines using Delta Live Tables for real-time, scalable, and reliable data processing
* Orchestrate and automate complex data workflows using Azure Databricks Workflows
* Leverage SQL Warehouses in Azure Databricks to perform SQL-based data analysis and reporting

**Outline**

* Explore Azure Databricks
	+ Introduction to Azure Databricks and its key features
	+ Understanding workloads and core concepts
	+ Data governance using Unity Catalog and Microsoft Purview
* Perform data analysis with Azure Databricks
	+ Data ingestion methods and integration
	+ Exploratory Data Analysis (EDA) using collaborative notebooks
	+ Data analysis with DataFrame APIs
* Use Apache Spark in Azure Databricks
	+ Introduction to Apache Spark
	+ Creating and managing Spark clusters
	+ Using Spark in notebooks for data transformation, analysis, and visualization
* Manage data with Delta Lake
	+ Introduction to Delta Lake and its benefits
	+ ACID transactions, schema enforcement, and time travel
	+ Ensuring data integrity and versioning
* Build data pipelines with Delta Live Tables
	+ Exploring Delta Live Tables for real-time data processing
	+ Data ingestion, integration, and real-time processing
* Deploy workloads with Azure Databricks Workflows
	+ Understanding Azure Databricks Workflows and its components
	+ Benefits of using Workflows for deployment
	+ Deploying workloads using Workflows
* Use SQL Warehouses in Azure Databricks
	+ Introduction to SQL Warehouses
	+ Creating databases, tables, queries, and dashboards
* Run Azure Databricks Notebooks with Azure Data Factory
	+ Understanding notebooks and pipelines
	+ Creating linked services and using Notebook activities