

COURSE: DP-203: Data Engineering on Microsoft Azure

In this course, the student will learn about the data engineering as it pertains to working with batch and real-time analytical solutions using Azure data platform technologies.

Students will begin by understanding the core compute and storage technologies that are used to build an analytical solution. The students will learn how to interactively explore data stored in files in a data lake. They will learn the various ingestion techniques that can be used to load data using the Apache Spark capability found in Azure Synapse Analytics or Azure Databricks, or how to ingest using Azure Data Factory or Azure Synapse pipelines. The students will also learn the various ways they can transform the data using the same technologies that is used to ingest data. They will understand the importance of implementing security to ensure that the data is protected at rest or in transit. The student will then show how to create a real-time analytical system to create real-time analytical solutions.

Summary

<i>Duration:</i>	<i>5 days</i>
<i>Level:</i>	<i>300</i>
<i>Delivery method:</i>	<i>Virtual Instructor-led class</i>
<i>Language:</i>	<i>English or Bulgarian</i>

* The difficulty level is consistent with the widely accepted scale of technical difficulty of training on Microsoft Corp

AUDIENCE:

The primary audience for this course is data professionals, data architects, and business intelligence professionals who want to learn about data engineering and building analytical solutions using data platform technologies that exist on Microsoft Azure.

The secondary audience for this course data analysts and data scientists who work with analytical solutions built on Microsoft Azure.

Successful students start this course with knowledge of cloud computing and core data concepts and professional experience with data solutions.

Specifically completing:

AZ-900 - Azure Fundamentals /DP-900 - Microsoft Azure Data Fundamentals

AFTER THE TRAINING ATTENDEES WILL BE ABLE TO:

- Explore compute and storage options for data engineering workloads in Azure
- Run interactive queries using serverless SQL pools
- Perform data Exploration and Transformation in Azure Databricks
- Explore, transform, and load data into the Data Warehouse using Apache Spark
- Ingest and load Data into the Data Warehouse
- Transform Data with Azure Data Factory or Azure Synapse Pipelines

TOPICS:

Module 1: Explore compute and storage options for data engineering workloads

- Introduction to Azure Synapse Analytics
- Describe Azure Databricks
- Describe Azure Databricks Delta Lake architecture
- Introduction to Azure Data Lake storage
- Work with data streams by using Azure
- Stream Analytics

Module 2: Run interactive queries using serverless SQL pools

- Explore Azure Synapse serverless SQL pools capabilities
- Query data in the lake using Azure Synapse serverless SQL pools
- Create metadata objects in Azure Synapse serverless SQL pools
- Secure data and manage users in Azure Synapse serverless SQL pools

Module 3: Data Exploration and Transformation in Azure Databricks

- Describe Azure Databricks
- Read and write data in Azure Databricks
- Work with DataFrames in Azure Databricks
- Work with DataFrames advanced methods in Azure Databricks

Module 4: Explore, transform , and load data into the Data Warehouse using Apache Spark

- Understand big data engineering with Apache Spark in Azure Synapse Analytics
- Ingest data with Apache Spark notebooks in Azure Synapse Analytics
- Transform data with DataFrames in Apache Spark Pools in Azure Synapse Analytics
- Integrate SQL and Apache Spark pools in Azure Synapse Analytics

Module 5: Ingest and load Data into the Data Warehouse

- Use loading best practices in Azure Synapse Analytics
- Petabyte-scale ingestion with Azure Data Factory

Module 6: Transform Data with Azure Data Factory or Azure Synapse Pipelines

- Data integration with Azure Data Factory
- Code-free transformation at scale with Azure Data Factory

Module 7: Integrate Data from Notebooks with Azure Data Factory or Azure Synapse Pipelines

- Orchestrating data movement and transformation in Azure Data Factory
- Run Azure Databricks Notebooks with Azure Data Factory

Module 8: End-to end security with Azure Synapse Analytics

- Secure a data warehouse in Azure Synapse Analytics
- Configure and manage secrets in Azure Key Vault
- Implement compliance controls for sensitive data

Module 9: Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link



- Plan hybrid transactional and analytical processing using Azure Synapse Analytics
- Implement Azure Synapse Link with Azure Cosmos DB
- Query Cosmos DB data with Spark
- Query Cosmos DB with Synapse SQL

Module 10: Real-time Stream Processing with Stream Analytics

- Enable reliable messaging for Big Data applications using Azure Event Hubs
- Work with data streams by using Azure Stream Analytics
- Transform data by using Azure Stream Analytics

Module 11: Create a Stream Processing Solution with Event Hubs and Azure Databricks

Process Streaming data with Azure Databricks structured streaming

Certification

This training will help you prepare for exam [DP-203: Data Engineering on Microsoft Azure](#).

By passing this exam you will earn the **Microsoft Certified: Azure Data Engineer Associate certification**.