

# DP-203: Data Engineering on Microsoft Azure

*Summary: Duration – 5 full days, Level - 300 (MS Scale), Delivery method - Virtual Instructor-led class or In-class*

In this course, you will learn how to implement and manage data engineering workloads on Microsoft Azure, using Azure services such as Azure Synapse Analytics, Azure Data Lake Storage Gen2, Azure Stream Analytics, Azure Databricks, and others. The course focuses on common data engineering tasks such as orchestrating data transfer and transformation pipelines, working with data files in a data lake, creating and loading relational data warehouses, capturing and aggregating streams of real-time data, and tracking data assets and lineage.

## **Important notice:**

This course is originally a 4- day official course from Microsoft. SQL Master Academy has added 1 day extra with content and to provide more time for examples, exercises, demonstrations and discussions.

## **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 includes data analysts and data scientists who work with analytical solutions built on Microsoft Azure. To attend this class, you should already have fundamental knowledge of data workloads and technologies, equivalent to the [Microsoft Azure Data Fundamentals](#) certification.

## **AFTER THE TRAINING ATTENDEES WILL BE ABLE TO:**

- Identify common data engineering tasks, Describe common data engineering concepts, Identify Azure services for data engineering
- Describe the key features and benefits of Azure Data Lake Storage Gen2, Enable Azure Data Lake Storage Gen2 in an Azure Storage account, Compare Azure Data Lake Storage Gen2 and Azure Blob storage, Describe where Azure Data Lake Storage Gen2 fits in the stages of analytical processing, Describe how Azure data Lake Storage Gen2 is used in common analytical workloads
- Identify the business problems that Azure Synapse Analytics addresses, Describe core capabilities of Azure Synapse Analytics, Determine when to use Azure Synapse Analytics.

- Use Azure Synapse serverless SQL pool to query files in a data lake, Use Azure Synapse serverless SQL pools to transform data in a data lake, Create a lake database in Azure Synapse Analytics, Secure data and manage users in Azure Synapse serverless SQL pools
- Analyze data with Apache Spark in Azure Synapse Analytics, Transform data with Spark in Azure Synapse Analytics, Use Delta Lake in Azure Synapse Analytic
- Analyze data in a relational data warehouse, Load data into a relational data warehouse, Manage and monitor data warehouse activities in Azure Synapse Analytics, Analyze and optimize data warehouse storage in Azure Synapse Analytics, Secure a data warehouse in Azure Synapse Analytics
- Build a data pipeline in Azure Synapse Analytics, Use Spark Notebooks in an Azure Synapse Pipeline
- Plan hybrid transactional and analytical processing using Azure Synapse Analytics, Implement Azure Synapse Link with Azure Cosmos DB, Implement Azure Synapse Link for SQL
- Process real-time data streams and integrate the data they contain into applications and analytical solutions with Azure Stream Analytics. Ingest streaming data using Azure Stream Analytics and Azure Synapse Analytics, Visualize real-time data with Azure Stream Analytics and Power BI.
- Evaluate whether Microsoft Purview is the right choice for your data discovery and governance needs, Discover trusted data using Microsoft Purview, Catalog data artifacts by using Microsoft Purview, Manage Power BI assets by using Microsoft Purview, Integrate Microsoft Purview and Azure Synapse Analytics
- Provision an Azure Databricks workspace, Identify core workloads and personas for Azure Databricks, Describe key concepts of an Azure Databricks solution. 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

## TOPICS:

**Module 1:** Get started with data engineering on Azure

**Module 2:** Analyze data with Azure Synapse Analytics serverless SQL pools

**Module 3:** Perform data engineering with Azure Synapse Apache Spark Pools

**Module 4:** Work with data warehouses using Azure Synapse Analytics

**Module 5:** Transfer and transform data with Azure Synapse Analytics pipelines

**Module 6:** Work with hybrid transactional and analytical processing (HTAP) Solutions using Azure Synapse Analytics

**Module 7:** Implement a data streaming solution with Azure Stream Analytics

**Module 8:** Govern data across an enterprise

**Module 9:** Data engineering with Azure Databricks