

# Data Engineering on Microsoft Azure (4 Days)

## Overview

In this course, students learn about 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.

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.

## COURSE BENEFITS

- 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.
- Integrate Data from Notebooks with Azure Data Factory or Azure Synapse Pipelines.
- Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link.
- Perform end-to-end security with Azure Synapse Analytics.
- Perform real-time Stream Processing with Stream Analytics.
- Create a Stream Processing Solution with Event Hubs and Azure Databricks.

## COURSE OUTLINE

### **Explore compute and storage options for data engineering workloads**

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

### **Run interactive queries using Azure Synapse Analytics 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

### **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

### **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

### **Ingest and load data into the data warehouse**

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

### **Transform data with Azure Data Factory or Azure Synapse Pipelines**

- Data integration with Azure Data Factory or Azure Synapse Pipelines
- Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines

### **Orchestrate data movement and transformation in Azure Synapse Pipelines**

- Orchestrate data movement and transformation in Azure Data Factory

**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

**Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link**

Design hybrid transactional and analytical processing using Azure Synapse Analytics  
Configure Azure Synapse Link with Azure Cosmos DB  
Query Azure Cosmos DB with Apache Spark pools  
Query Azure Cosmos DB with serverless SQL pools

**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  
Ingest data streams with Azure Stream Analytics

**Create a Stream Processing Solution with Event Hubs and Azure Databricks**

Process streaming data with Azure Databricks structured streaming