

DP-203T00 Data Engineering on Microsoft Azure

Class Length: 4 Days

Overview

In this course, the student will learn about the data engineering patterns and practices 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. They will then explore how to design an analytical serving layers and focus on data engineering considerations for working with source files. 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. The student will spend time on the course learning how to monitor and analyze the performance of analytical system so that they can optimize the performance of data loads, or queries that are issued against the systems. 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 the data in an analytical system can be used to create dashboards, or build predictive models in Azure Synapse Analytics.

Prerequisite Comments

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

Target 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.

Course Objectives

- Explore compute and storage options for data engineering workloads in Azure
- Design and Implement the serving layer
- Understand data engineering considerations
- Run interactive queries using serverless SQL pools
- Explore, transform, and load data into the Data Warehouse using Apache Spark
- Perform data Exploration and Transformation in Azure Databricks
- 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
- Optimize Query Performance with Dedicated SQL Pools in Azure Synapse

Analyze and Optimize Data Warehouse Storage

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

Build reports using Power BI integration with Azure Synapse Analytics

Perform Integrated Machine Learning Processes in Azure Synapse Analytics

Course Outline

1 - 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

2 - Design and implement the serving layer

Design a multidimensional schema to optimize analytical workloads

Code-free transformation at scale with Azure Data Factory

Populate slowly changing dimensions in Azure Synapse Analytics pipelines

3 - Data engineering considerations for source files

Design a Modern Data Warehouse using Azure Synapse Analytics

Secure a data warehouse in Azure Synapse Analytics

4 - 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

5 - 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

6 - 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

7 - Ingest and load data into the data warehouse

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

8 - 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

9 - Orchestrate data movement and transformation in Azure Synapse Pipelines

Orchestrate data movement and transformation in Azure Data Factory

10 - Optimize query performance with dedicated SQL pools in Azure Synapse

Optimize data warehouse query performance in Azure Synapse Analytics
Understand data warehouse developer features of Azure Synapse Analytics

11 - Analyze and Optimize Data Warehouse Storage

Analyze and optimize data warehouse storage in Azure Synapse Analytics

12 - 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

13 - 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

14 - 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

15 - Create a Stream Processing Solution with Event Hubs and Azure Databricks

Process streaming data with Azure Databricks structured streaming

16 - Build reports using Power BI integration with Azure Synapse Analytics

Create reports with Power BI using its integration with Azure Synapse Analytics

17 - Perform Integrated Machine Learning Processes in Azure Synapse Analytics

Use the integrated machine learning process in Azure Synapse Analytics
