Data Engineering on Microsoft Azure (DP-203)

Modality: On Demand

Duration: 5 Hours

About this Course:

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.

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 Synpase Analytics
- Perform Integrated Machine Learning Processes in Azure Synapse Analytics

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.

Prerequisites:

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

Course Outline:

- Course Introduction
- Module 1: Explore Compute and Storage Options for Data Engineering Workloads
- Module 2: Run Interactive Queries Using Azure Synapse Analytics Serverless SQL Pools
- Module 3: Data Exploration and Transformation in Azure Databricks
- Module 4: Explore, Transform, and Load Data into the Data Warehouse using Apache Spark
- Module 5: Ingest and Load Data into the Data Warehouse
- Module 6: Transform Data with Azure Data Factory or Azure Synapse Pipelines
- Module 7: Integrate Data from Notebooks with Azure Data Factory or Azure Synapse Pipelines
- Module 8: End-to-end Security with Azure Synapse Analytics
- Module 9: Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link
- Module 10: Real-time Stream Processing with Stream Analytics
- Module 11: Create a Stream Processing Solution with Event Hubs and Azure Databricks