DP-203: Data Engineering on Microsoft Azure

Overview

In this DP-203T00 Data Engineering on Microsoft Azure 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.

Duration

4 Days

Who Should Attend

The primary audience for this Azure certification 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. Understanding hybrid network connectivity methods, such as VPN.
- Understanding resilience and disaster recovery, including high availability and restore operations.

DP-203: Data Engineering on Microsoft Azure

Course Topics

Module 1	**Introduction to Azure Synapse Analytics** • Learn about the features and capabilities of Azure Synapse Analytics – a cloud-based platform for big data processing and analysis.
Module 2	**Explore Azure Databricks** • Azure Databricks is a cloud service that provides a scalable platform for data analytics using Apache Spark.
Module 3	**Introduction to Azure Data Lake storage** • Learn how Azure Data Lake Storage provides a cloud storage service that is highly available, secure, durable, scalable, and redundant and brings new efficiencies to processing big data analytics workloads.
Module 4	**Work with data streams by using Azure Stream Analytics** • Explore how Azure Stream Analytics integrates with your applications or Internet of Things (IoT) devices to gain insights with real-time streaming data. Learn how to consume and analyze data streams and derive actionable results.

DP-203: Data Engineering on Microsoft Azure

Module 5	**Use Azure Synapse serverless SQL pool to query files in a data lake** • With Azure Synapse serverless SQL pool, you can leverage your SQL skills to explore and analyze data in files, without the need to load the data into a relational database.
Module 6	**Create a lake database in Azure Synapse Analytics** • Why choose between working with files in a data lake or a relational database schema? With lake databases in Azure Synapse Analytics, you can combine the benefits of both.
Module 7	**Secure data and manage users in Azure Synapse serverless SQL pools** • Learn how you can set up security when using Azure Synapse serverless SQL pools
Module 8	**Use Apache Spark in Azure Databricks** • Azure Databricks is built on Apache Spark and enables data engineers and analysts to run Spark jobs to transform, analyze and visualize data at scale.
Module 9	**Use Delta Lake in Azure Databricks** • Delta Lake is an open source relational storage area for Spark that you can use to implement a data lakehouse architecture in Azure Databricks.
Module 10	**Analyze data with Apache Spark in Azure Synapse Analytics** • Apache Spark is a core technology for large-scale data analytics. Learn how to use Spark in Azure Synapse Analytics to analyze and visualize data in a data lake.
Module 11	**Integrate SQL and Apache Spark pools in Azure Synapse Analytics** • Learn how to integrate SQL and Apache Spark pools in Azure Synapse Analytics.
Module 12	**Use data loading best practices in Azure Synapse Analytics** • Learn the best practices you need to adopt to load data into a data warehouse in Azure Synapse Analytics.

DP-203: Data Engineering on Microsoft Azure

Module 13	**Petabyte-scale ingestion with Azure Data Factory or Azure Synapse Pipeline** • In this module, you will learn the various methods that can be used to ingest data between various data stores using Azure Data Factory.
Module 14	**Integrate data with Azure Data Factory or Azure Synapse Pipeline** • In this module, you will examine Azure Data Factory and the core components that enable you to create large scale data ingestion solutions in the cloud
Module 15	**Perform code-free transformation at scale with Azure Data Factory or Azure Synapse Pipeline** • In this module, you will learn how to perform common data transformation and cleansing activities within Azure Data Factory without using code.
Module 16	**Orchestrate data movement and transformation in Azure Data Factory or Azure Synapse Pipeline** • In this module, you will learn how Azure Data Factory can orchestrate large scale data movement by using other Azure Data Platform and Machine Learning technologies.
Module 17	**Plan hybrid transactional and analytical processing using Azure Synapse Analytics** • Learn how hybrid transactional / analytical processing (HTAP) can help you perform operational analytics with Azure Synapse Analytics.
Module 18	**Implement Azure Synapse Link with Azure Cosmos DB** • Azure Synapse Link for Azure Cosmos DB enables HTAP integration between operational data in Azure Cosmos DB and Azure Synapse Analytics runtimes for Spark and SQL.
Module 19	**Secure a data warehouse in Azure Synapse Analytics** • Learn how to approach and implement security to protect your data with Azure Synapse Analytics.

DP-203: Data Engineering on Microsoft Azure

Module 20	**Configure and manage secrets in Azure Key Vault** • Storing and handling secrets, encryption keys, and certificates directly is risky, and every usage introduces the possibility of unintentional data exposure. Azure Key Vault provides a secure storage area for managing all your app secrets so you can properly encrypt your data in transit or while it's being stored.
Module 21	**Implement compliance controls for sensitive data** • Explore data classification capabilities and degrees of confidentiality. Implement security options to maintain private data safe, including Azure SQL auditing, Microsoft Defender for SQL, row-level security, Dynamic Data Masking and Azure SQL Database Ledger.
Module 22	**Enable reliable messaging for Big Data applications using Azure Event Hubs** • Connect sending and receiving applications with Event Hubs so you can handle extremely high loads without losing data.