Data Engineering on Microsoft Azure (DP-203)

Description

This **Data Engineering on Microsoft Azure (DP-203)** course will introduce you to data engineering models and practices within the context of real-time and batch analytical solutions using Azure data platform technologies.

You will learn about the fundamental computing and storage technologies used to build an analytical solution and then explore how to design analytical service layers, focusing on data engineering considerations for working with source files.

Through this Data Engineering on Microsoft Azure course, you will learn to interactively explore data stored in files in a data lake, the different ingestion techniques that can be used to load data using Apache Spark functionality in Azure Synapse Analytics or Azure Databricks, or how to ingest data using Azure Data Factory or Azure Synapse pipelines.

You will also cover the various ways to transform data using the same technologies used for data acquisition and learn how to monitor and analyze the performance of analytical systems to optimize the performance of data loads or queries on the systems. Finally, you will understand the importance of implementing security to ensure data protection at rest or in transit.

Course Content
Module 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

Module 2: 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

Module 3: 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

Module 4: 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

Module 5: Ingest and load data into the data warehouse

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

Module 6: 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

Module 7: Orchestrate data movement and transformation in Azure Synapse Pipelines

- Orchestrate data movement and transformation in Azure Data Factory

Module 8: 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

Module 9: 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

Module 10: 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

Module 11: Create a Stream Processing Solution with Event Hubs and Azure Databricks

- Process streaming data with Azure Databricks structured streaming

Lab / Exercises
Official Microsoft Labs

- Lab 1: Explore compute and storage options for data engineering workloads
- Lab 2: Run interactive queries using serverless SQL pools

- Lab 3: Data Exploration and Transformation in Azure Databricks
- Lab 4: Explore, transform, and load data into the Data Warehouse using Apache Spark
- Lab 5: Ingest and load Data into the Data Warehouse
- Lab 6: Transform Data with Azure Data Factory or Azure Synapse Pipelines
- Lab 7: Orchestrate data movement and transformation in Azure Synapse Pipelines
- Lab 8: End-to-end security with Azure Synapse Analytics
- Lab 9: Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link
- Lab 10: Real-time Stream Processing with Stream Analytics
- Lab 11: Create a Stream Processing Solution with Event Hubs and Azure Databricks

Documentation

- Access to Microsoft Learn (online learning content)

Exam

- This course prepares you to the DP-203 : Data Engineering on Microsoft Azure exam. If you wish to take this exam, please contact our secretariat who will let you know the cost of the exam and will take care of all the necessary administrative procedures for you.

Participant profiles

- 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
- Data analysts and data scientists who work with analytical solutions built on Microsoft Azure

Prerequisites

- Knowledge of cloud computing and core data concepts
- Professional experience with data solutions
- Have followed the courses: Microsoft Azure Fundamentals (AZ-900) and Microsoft Azure Data Fundamentals (DP-900)

Objectives

- 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

Description
Data Engineering on Microsoft Azure (DP-203)
Niveau
Intermédiaire
Classroom Registration Price (CHF)
3200
Virtual Classroom Registration Price (CHF)

3000

Duration (in Days)

4

Reference

DP-203T00