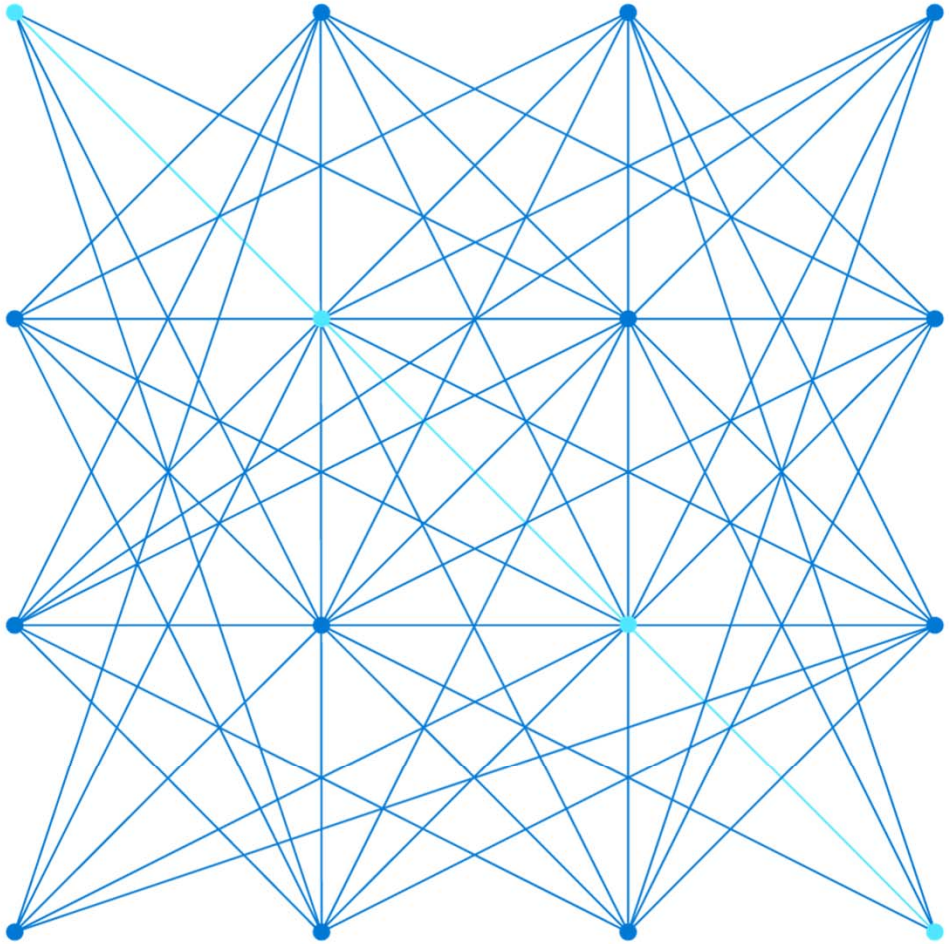
Microsoft Azure

DP-203T00: Data Engineering in Azure



Agenda

About this course

Audience

Course agenda

Prerequisites

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 agenda

Module 0
Data Engineering in Azure course overview

Lesson 01 – About this course

Lesson 02 – Course agenda

Lesson 03 – Audience

Lesson 04 – Prerequisites

Lesson 05 – Course lab setup

Module 1
Explore compute and storage options for data engineering workloads

Lesson 01 – Introduction to Azure Synapse Analytics

Lesson 02 – Describe Azure Databricks

Lesson 03 – Introduction to Azure Data Lake storage

Lesson 04 – Describe Delta Lake architecture

Lesson 05 – Work with data streams by using Azure Stream Analytics

Course agenda *(continued #1)*

Module 2
Run interactive queries using Azure Synapse Analytics serverless SQL pool

Lesson 01 – Explore Azure Synapse serverless SQL pools capabilities

Lesson 02 – Query data in the lake using Azure Synapse serverless SQL pools

Lesson 03 – Create metadata objects in Azure Synapse serverless SQL pools

Lesson 04 – Secure data and manage users in Azure Synapse serverless SQL pools

Module 3
Data exploration and transformation in Azure Databricks

Lesson 01 – Describe Azure Databricks

Lesson 02 – Read and write data in Azure Databricks

Lesson 03 – Work with DataFrames in Azure Databricks

Lesson 04 – Work with DataFrames advanced methods in Azure Databricks

Course agenda *(continued #2)*

Module 4
Explore, transform, and load data into the data warehouse using Azure Synapse Analytics Apache Spark

Lesson 01 – Understand big data engineering with Apache Spark in Azure Synapse Analytics

Lesson 02 – Ingest data with Apache Spark notebooks in Azure Synapse Analytics

Lesson 03 – Transform data with DataFrames in Apache Spark Pools in Azure Synapse Analytics

Lesson 04 – Integrate SQL and Apache Spark pools in Azure Synapse Analytics

Module 5
Ingest and load data into the data warehouse

Lesson 01 – Use data loading best practices in Azure Synapse Analytics

Lesson 02 – Petabyte-scale ingestion with Azure Data Factory or Azure Synapse Pipelines

Course agenda *(continued #3)*

Module 6
Transform data with Azure Data Factory or Azure Synapse Pipelines

Lesson 01 – Data integration with Azure Data Factory or Azure Synapse Pipelines

Lesson 02 – Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines

Module 7
Orchestrate data movement and transformation in Azure Data Factory or Azure Synapse Pipelines

Lesson 01 – Orchestrate data movement and transformation in Azure Data Factory or Azure Synapse Pipelines

Course agenda *(continued #4)*

Module 8
End-to-end security with Azure Synapse Analytics

Lesson 01 – Secure a data warehouse in Azure Synapse Analytics

Lesson 02 – Configure and manage secrets in Azure Key Vault

Lesson 03 – Implement compliance controls for sensitive data

Module 9
Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link

Lesson 01 – Design hybrid transactional and analytical processing using Azure Synapse Analytics

Lesson 02 – Configure Azure Synapse Link with Azure Cosmos DB

Lesson 03 – Query Azure Cosmos DB with Apache Spark for Azure Synapse Analytics

Lesson 04 – Query Azure Cosmos DB with SQL serverless for Azure Synapse Analytics

Course agenda *(continued #5)*

Module 10
Real-time stream processing with Azure Stream Analytics

Lesson 01 – Enable reliable messaging for Big Data applications using Azure Event Hubs

Lesson 02 – Work with data streams by using Azure Stream Analytics

Lesson 03 – Ingest data streams with Azure Stream Analytics

Module 11
Create a stream processing solution with Event Hubs and Azure Databricks

Lesson 01 – Process streaming data with Azure Databricks structured streaming

Audience

Primary audience:

The audience for this course are data engineers, data professionals, data architects, and business intelligence professionals who want to learn about the data platform technologies that exist on Microsoft Azure that can be used to perform data engineering and storage for analytical solutions.

Secondary audience:

The secondary audience for this course are individuals who develop applications that deliver content from the data platform technologies that exist on Microsoft Azure.

Prerequisites

In addition to their professional experience, students who take this training should have technical knowledge equivalent to the following course:

Data fundamentals

Course lab setup