

DP-203

Data Engineering on Microsoft Azure

Durata: 4 gg

Descrizione

In questo corso illustreremo:

- i modelli e le pratiche di data engineering per quanto riguarda il lavoro con soluzioni analitiche batch e in tempo reale utilizzando le tecnologie della piattaforma dati Azure.
- le tecnologie di calcolo e di archiviazione di base che vengono utilizzate per costruire una soluzione analitica.
- come progettare un livello di servizio analitico con considerazioni di data engineering per lavorare con i file di origine.
- come esplorare interattivamente i dati memorizzati nei file in un data lake
- le varie tecniche di data ingestion che possono essere utilizzate per caricare i dati utilizzando la capacità di Apache Spark presente in Azure Synapse Analytics o Azure Databricks, o utilizzando Azure Data Factory o Azure Synapse pipeline.
- i modi in cui possono trasformare i dati utilizzando le stesse tecnologie che vengono utilizzate per importare i dati.
- come monitorare e analizzare le prestazioni del sistema analitico in modo da poter ottimizzare le prestazioni dei data loads, o le query che vengono emesse contro i sistemi.
- come implementare la sicurezza per assicurare che i dati siano protetti at rest o in transit.
- come i dati in un sistema analitico possono essere utilizzati per creare dashboard o costruire modelli predittivi in Azure Synapse Analytics.

A chi è rivolto?

Questo corso è rivolto a data professionals, data architects e professionisti della business intelligence che vogliono imparare data engineering e la costruzione analytical solutions utilizzando le tecnologie della piattaforma dati che esistono su Microsoft Azure. Inoltre è rivolto a data analysts e data scientists che lavorano con soluzioni analitiche costruite su Microsoft Azure.

Prerequisiti

Per partecipare con profitto a questo corso gli allievi dovrebbero aver partecipato ai seguenti corsi oppure possedere competenze equivalenti:

AZ-900 – Azure Fundamentals

Contenuti

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: 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
- Module 3: Data engineering considerations for source files
- Design a Modern Data Warehouse using Azure Synapse Analytics
- Secure a data warehouse in Azure Synapse Analytics

Module 3: Data engineering considerations for source files

-
- Design a Modern Data Warehouse using Azure Synapse Analytics
- Secure a data warehouse in Azure Synapse Analytics

Module 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

Module 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

Module 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

Module 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

Module 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

Module 9: Orchestrate data movement and transformation in Azure Synapse Pipelines

- Orchestrate data movement and transformation in Azure Data Factory

Module 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

Module 11: Analyze and Optimize Data Warehouse Storage

- Analyze and optimize data warehouse storage in Azure Synapse Analytics

Module 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

Module 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

Module 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

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

- Process streaming data with Azure Databricks structured streaming

Module 16: Build reports using Power BI integration with Azure Synapse Analytics

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

Module 17: Perform Integrated Machine Learning Processes in Azure Synapse Analytics

- Use the integrated machine learning process in Azure Synapse Analytics