Data Engineering on Microsoft Azure

Duration: 4 Days Course Code: M-DP203 Delivery Method: Company Event

Overview:

Exclusive - In this course, the student will learn how to implement and manage data engineering workloads on Microsoft Azure, using Azure services such as Azure Synapse Analytics, Azure Data Lake Storage Gen2, Azure Stream Analytics, Azure Databricks, and others.
The course focuses on common data engineering tasks such as orchestrating data transfer and transformation pipelines, working with data files in a data lake, creating and loading relational data warehouses, capturing and aggregating streams of real-time data, and tracking data assets and lineage.

Company Events

These events can be delivered exclusively for your company at our locations or yours, specifically for your delegates and your needs. The Company Events can be tailored or standard course deliveries.

Target 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 includes data analysts and data scientists who work with analytical solutions built on Microsoft Azure.

Objectives:

- Students will learn:
- Introduction to data engineering on Azure
- Introduction to Azure Data Lake Storage Gen2
- Introduction to Azure Synapse Analytics
- Use Azure Synapse serverless SQL pool to query files in a data lake
- Use Azure Synapse serverless SQL pools to transform data in a data lake
- Create a lake database in Azure Synapse Analytics
- Analyze data with Apache Spark in Azure Synapse Analytics
- Transform data with Spark in Azure Synapse Analytics
- Use Delta Lake in Azure Synapse Analytics
- Analyze data in a relational data warehouse
- Load data into a relational data warehouse
- Build a data pipeline in Azure Synapse Analytics
- Use Spark Notebooks in an Azure Synapse Pipeline
- Plan hybrid transactional and analytical processing using Azure Synapse Analytics
- Implement Azure Synapse Link with Azure Cosmos DB
- Implement Azure Synapse Link for SQL
- Get started with Azure Stream Analytics
- Ingest streaming data using Azure Stream Analytics and Azure Synapse Analytics
- Visualize real-time data with Azure Stream Analytics and Power BI
- Introduction to Microsoft Purview
- Integrate Microsoft Purview and Azure Synapse Analytics
- Explore Azure Databricks
- Use Apache Spark in Azure Databricks
- Run Azure Databricks Notebooks with Azure Data Factory

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
- M-DP900 - Microsoft Azure Data Fundamentals
- M-AZ900 - Microsoft Azure Fundamentals

Content:

Module 1 : Introduction to data engineering on Azure

- Identify common data engineering tasks
- Describe common data engineering concepts
- Identify Azure services for data engineering

Module 2 : Introduction to Azure Data Lake Storage Gen2

- Describe the key features and benefits of Azure Data Lake Storage Gen2
- Enable Azure Data Lake Storage Gen2 in an Azure Storage account
- Compare Azure Data Lake Storage Gen2 and Azure Blob storage
- Describe where Azure Data Lake Storage Gen2 fits in the stages of analytical processing
- Describe how Azure data Lake Storage Gen2 is used in common analytical workloads

Module 3 : Introduction to Azure Synapse Analytics

- Identify the business problems that Azure Synapse Analytics addresses.
- Describe core capabilities of Azure Synapse Analytics.
- Determine when to use Azure Synapse Analytics.

Module 4 : Use Azure Synapse serverless SQL pool to query files in a data lake

- Identify capabilities and use cases for serverless SQL pools in Azure Synapse Analytics
- Query CSV, JSON, and Parquet files using a serverless SQL pool
- Create external database objects in a serverless SQL pool

Module 5 : Use Azure Synapse serverless SQL pools to transform data in a data lake

- Use a CREATE EXTERNAL TABLE AS SELECT (CETAS) statement to transform data.
- Encapsulate a CETAS statement in a stored procedure.
- Include a data transformation stored procedure in a pipeline.

Module 6 : Create a lake database in Azure Synapse Analytics

- Understand lake database concepts and components
- Describe database templates in Azure Synapse Analytics

Module 9 : Use Delta Lake in Azure Synapse Analytics

- Describe core features and capabilities of Delta Lake.
- Create and use Delta Lake tables in a Synapse Analytics Spark pool.
- Create Spark catalog tables for Delta Lake data.
- Use Delta Lake tables for streaming data.
- Query Delta Lake tables from a Synapse Analytics SQL pool.

Module 10 : Analyze data in a relational data warehouse

- Design a schema for a relational data warehouse.
- Create fact, dimension, and staging tables.
- Use SQL to load data into data warehouse tables.
- Use SQL to query relational data warehouse tables.

Module 11 : Load data into a relational data warehouse

- Load staging tables in a data warehouse
- Load dimension tables in a data warehouse
- Load time dimensions in a data warehouse
- Load slowly-changing dimensions in a data warehouse
- Load fact tables in a data warehouse
- Perform post-load optimizations in a data warehouse

Module 12 : Build a data pipeline in Azure Synapse Analytics

- Describe core concepts for Azure Synapse Analytics pipelines.
- Create a pipeline in Azure Synapse Studio.
- Implement a data flow activity in a pipeline.
- Initiate and monitor pipeline runs.

Module 13 : Use Spark Notebooks in an Azure Synapse Pipeline

- Describe notebook and pipeline integration.
- Use a Synapse notebook activity in a pipeline.
- Use parameters with a notebook activity.

Module 14 : Plan hybrid transactional and analytical processing using Azure Synapse Analytics

Module 17 : Get started with Azure Stream Analytics

- Understand data streams.
- Understand event processing.
- Understand window functions.
- Get started with Azure Stream Analytics.

Module 18 : Ingest streaming data using Azure Stream Analytics and Azure Synapse Analytics

- Describe common stream ingestion scenarios for Azure Synapse Analytics.
- Configure inputs and outputs for an Azure Stream Analytics job.
- Define a query to ingest real-time data into Azure Synapse Analytics.
- Run a job to ingest real-time data, and consume that data in Azure Synapse Analytics.

Module 19 : Visualize real-time data with Azure Stream Analytics and Power BI

- Configure a Stream Analytics output for Power BI.
- Use a Stream Analytics query to write data to Power BI.
- Create a real-time data visualization in Power BI.

Module 20 : Introduction to Microsoft Purview

- Evaluate whether Microsoft Purview is appropriate for data discovery and governance needs.
- Describe how the features of Microsoft Purview work to provide data discovery and governance.

Module 21 : Integrate Microsoft Purview and Azure Synapse Analytics

- Catalog Azure Synapse Analytics database assets in Microsoft Purview.
- Configure Microsoft Purview integration in Azure Synapse Analytics.
- Search the Microsoft Purview catalog from Synapse Studio.
- Track data lineage in Azure Synapse Analytics pipelines activities.

Module 22 : Explore Azure Databricks

- Provision an Azure Databricks workspace.
- Identify core workloads and personas for Azure Databricks.
- Describe key concepts of an Azure Databricks solution.

Module 23 : Use Apache Spark in Azure Databricks

- Create a lake database

Module 7 : Analyze data with Apache Spark in Azure Synapse Analytics

- Identify core features and capabilities of Apache Spark.
- Configure a Spark pool in Azure Synapse Analytics.
- Run code to load, analyze, and visualize data in a Spark notebook.

Module 8 : Transform data with Spark in Azure Synapse Analytics

- Use Apache Spark to modify and save dataframes
- Partition data files for improved performance and scalability.
- Transform data with SQL

- Describe Hybrid Transactional / Analytical Processing patterns.
- Identify Azure Synapse Link services for HTAP.

Module 15 : Implement Azure Synapse Link with Azure Cosmos DB

- Configure an Azure Cosmos DB Account to use Azure Synapse Link.
- Create an analytical store enabled container.
- Create a linked service for Azure Cosmos DB.
- Analyze linked data using Spark.
- Analyze linked data using Synapse SQL.

Module 16 : Implement Azure Synapse Link for SQL

- Understand key concepts and capabilities of Azure Synapse Link for SQL.
- Configure Azure Synapse Link for Azure SQL Database.
- Configure Azure Synapse Link for Microsoft SQL Server.

- Describe key elements of the Apache Spark architecture.
- Create and configure a Spark cluster.
- Describe use cases for Spark.
- Use Spark to process and analyze data stored in files.
- Use Spark to visualize data.

Module 24 : Run Azure Databricks Notebooks with Azure Data Factory

- Describe how Azure Databricks notebooks can be run in a pipeline.
- Create an Azure Data Factory linked service for Azure Databricks.
- Use a Notebook activity in a pipeline.
- Pass parameters to a notebook.

Further Information:

For More information, or to book your course, please call us on 00 971 4 446 4987

training@globalknowledge.ae

www.globalknowledge.com/en-ae/

Global Knowledge, Dubai Knowledge Village, Block 2A,First Floor, Office F68, Dubai, UAE