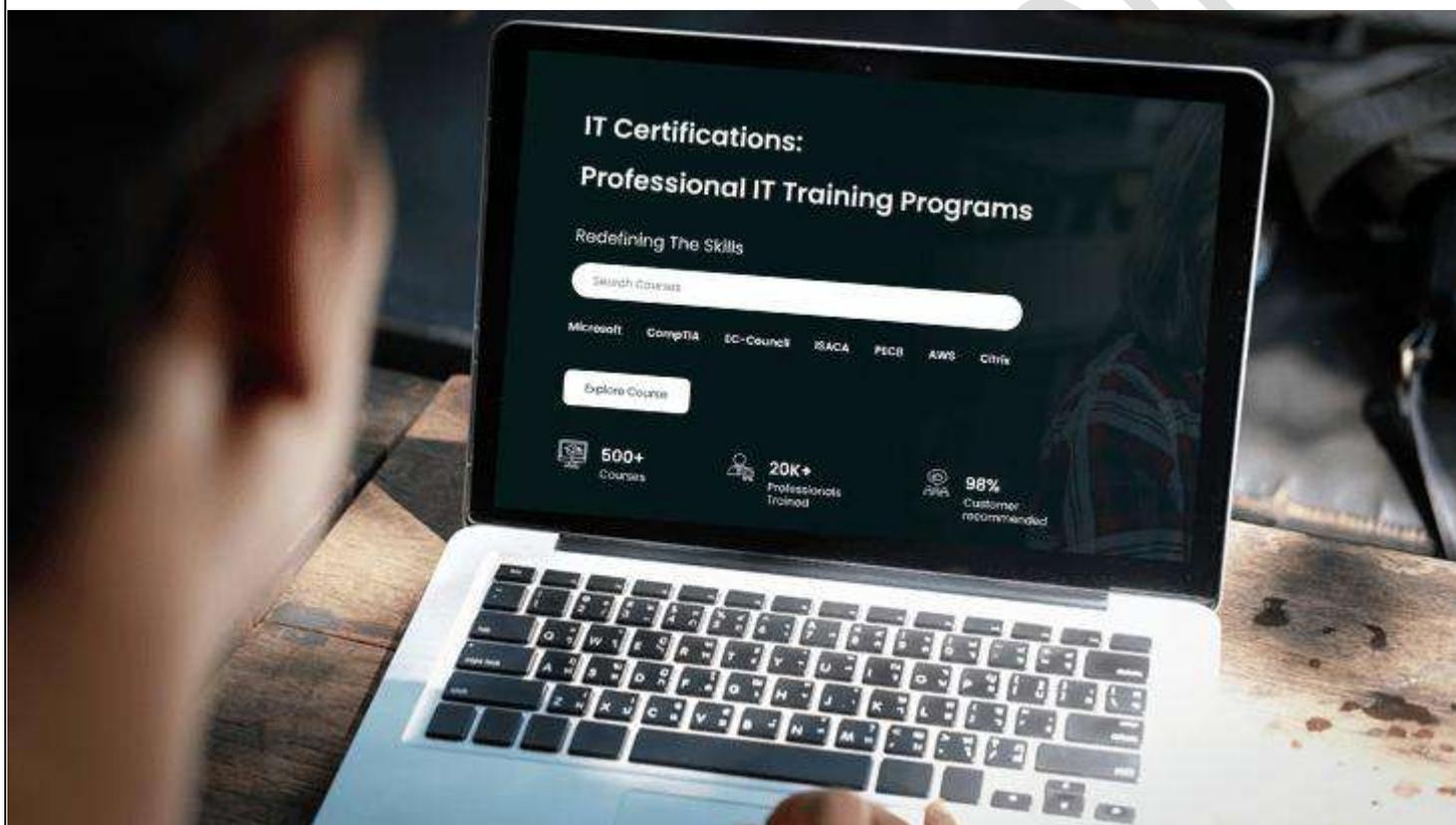




Redefining The Skills



## **DP-203: DATA ENGINEERING ON MICROSOFT AZURE TRAINING**

**Duration: 4 Days**

### **Course Description**

DP-203: Data Engineering on Microsoft Azure Training is a program solely intended for data engineers. The basic need of this course consists of the willing professional to have subject matter expertise. They should have strong hold in processing languages such as SQL, Python, and Scala. They should know how to integrate, transform, and consolidate the data.

Completing this program will help professionals learn about the pipelines of data, Azure Data Lake Storage Gen2, Azure Stream Analytics, Azure Databricks, and Azure Synapse Analytics. Further, learners will also gain practical experience in labs and implement their understanding into Serverless SQL pools, and Azure Synapse Apache Spark Pools.

This training is designed based on the objectives of the course variant DP-203T00-A.

### **Who should attend this course?**

- The primary audience for this course is data professionals, data architects, and business intelligence professionals who want to gain knowledge about data engineering and building analytical solutions by leveraging data platform technologies that exist on Microsoft Azure.
- The secondary audience for this course is professionals working as data analysts and scientists. Basically they are working with analytical solutions built with Microsoft Azure.
- Given below are professionals who can use Data Engineering on Microsoft Azure Training to upskill their current positions:
  - Data Scientist
  - Data Engineers
  - Cloud Solution Architects
  - DevOps Engineers
  - Database Administrators
  - Technical Team Leads
  - IT professionals
  - Data Architects

### **What you will learn**

- Explore compute and storage options for data engineering workloads in Azure
- Design and implement the serving layer
- Understand data engineering considerations
- Run interactive queries using serverless SQL pools
- Explore, transform, and load data into the Data Warehouse using Apache Sparkgreen-tickPerform data Exploration and Transformation in Azure Databricks
- 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
- Optimize Query Performance with Dedicated SQL Pools in Azure Synapse
- Analyze and Optimize Data Warehouse Storage
- 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
- Build reports using Power BI integration with Azure Synapse Analytics
- Perform Integrated Machine Learning Processes in Azure Synapse Analytics
- Implement a data lake house analytics solution in Azure Databricks

## Prerequisites

- AZ-900T00: Microsoft Azure Fundamentals
- DP-900T00: Microsoft Azure Data Fundamentals
- Before starting this program, one should have completed Microsoft Azure Data Fundamentals Certification or need to be certified in an equivalent certification.

## Curriculum

### Module 1: Get started with data engineering on Azure

- Introduction to data engineering on Azure
  - Identify common data engineering tasks
  - Describe common data engineering concepts
  - Identify Azure services for data engineering
- 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
- 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 2: Build data analytics solutions using Azure Synapse serverless SQL pools

- 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
- 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.
- Create a lake database in Azure Synapse Analytics
  - Understand lake database concepts and components
  - Describe database templates in Azure Synapse Analytics
  - Create a lake database
- Secure data and manage users in Azure Synapse serverless SQL pools
  - Choose an authentication method in Azure Synapse serverless SQL pools
  - Manage users in Azure Synapse serverless SQL pools
  - Manage user permissions in Azure Synapse serverless SQL pools

### Module 3: Perform data engineering with Azure Synapse Apache Spark Pools

- 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.
- 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
- 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 4: Transfer and transform data with Azure Synapse Analytics pipelines**

- 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.
- 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 5: Implement a Data Analytics Solution with Azure Synapse Analytics**

- 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.
- 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
- 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.
- 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.
- 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.
- 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 6: Work with Data Warehouses using Azure Synapse Analytics**

- 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.
- 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
- Manage and monitor data warehouse activities in Azure Synapse Analytics
  - Scale compute resources in Azure Synapse Analytics
  - Pause compute in Azure Synapse Analytics
  - Manage workloads in Azure Synapse Analytics
  - Use Azure Advisor to review recommendations
  - Use Dynamic Management Views to identify and troubleshoot query performance
- Secure a data warehouse in Azure Synapse Analytics
  - Understand network security options for Azure Synapse Analytics
  - Configure Conditional Access
  - Configure Authentication
  - Manage authorization through column and row level security
  - Manage sensitive data with Dynamic Data masking
  - Implement encryption in Azure Synapse Analytics

### **Module 7: Work with Hybrid Transactional and Analytical Processing Solutions using Azure Synapse Analytics**

- Plan hybrid transactional and analytical processing using Azure Synapse Analytics
  - Describe Hybrid Transactional / Analytical Processing patterns.
  - Identify Azure Synapse Link services for HTAP.
- 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.
- 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.

### **Module 8: Implement a Data Streaming Solution with Azure Stream Analytics**

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

- 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 9: Govern data across an enterprise**

- Introduction to Microsoft Purview
  - Evaluate whether Microsoft Purview is appropriate for your data discovery and governance needs.
  - Describe how the features of Microsoft Purview work to provide data discovery and governance.
- Discover trusted data using Microsoft Purview
  - Browse, search, and manage data catalog assets.
  - Use data catalog assets with Power BI.
  - Use Microsoft Purview in Azure Synapse Studio.
- Catalog data artifacts by using Microsoft Purview
  - Describe asset classification in Microsoft Purview.
- Manage Power BI assets by using Microsoft Purview
  - Register and scan a Power BI tenant.
  - Use the search and browse functions to find data assets.
  - Describe the schema details and data lineage tracing of Power BI data assets.
- 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.

---

*For any query Contact Us – Microtek Learning*

---