

**Data Engineering on Microsoft Azure (DP-203)**

**Course Number:** MOC-DP-203
**Duration:** 4 days

**Overview**

This Microsoft official DP-203 course, Data Engineering on Microsoft Azure, teaches attendees the patterns and practices of real-time analytical solutions using Azure. Participants learn how to interactively explore data stored in files in a data lake, transform data, monitor and analyze the performance of analytical systems, and much more. This course prepares students for the [DP-203 exam](https://docs.microsoft.com/en-us/learn/certifications/exams/DP-203) for which every attendee receives a voucher.

**Prerequisites**

All students must have knowledge of cloud computing, core data concepts, and professional experience with data solutions. Attendees should have taken [AZ-900: Azure Fundamentals](file:////training/microsoft-azure-fundamentals) and [DP-900: Microsoft Azure Data Fundamentals](file:////training/microsoft-azure-data-fundamentals), or have equivalent experience.

**Materials**

All Microsoft training students receive Microsoft official courseware.

For all Microsoft Official Courses taught in their entirety that have a corresponding certification exam, an exam voucher is included for each participant.

**Software Needed on Each Student PC**

Attendees will not need to install any software on their computers for this class. The class will be conducted in a remote environment that Accelebrate will provide; students will only need a local computer with a web browser and a stable Internet connection. Any recent version of Microsoft Edge, Mozilla Firefox, or Google Chrome will work well.

**Objectives**

* Explore Azure Synapse Analytics
* Describe key concepts of an Azure Databricks solution
* Create an Azure Storage Account by using the portal
* Get started with Azure Stream Analytics
* Query files using a serverless SQL pool
* Analyze data in a lake database
* Set up security when using Azure Synapse serverless SQL pools
* Use Spark in Azure Databricks
* Use Delta Lake in Azure Databricks
* Analyze data with Spark
* Integrate SQL and Apache Spark pools in Azure Synapse Analytics
* Implement workload management
* Ingest data between various data stores using Azure Data Factory
* Examine Azure Data Factory and the core components that enable large-scale data ingestion solutions in the cloud
* Perform common data transformation and cleansing activities within Azure Data Factory without using code
* Integrate a Notebook within Azure Synapse Pipelines
* Understand how hybrid transactional/analytical processing (HTAP) can perform operational analytics with Azure Synapse Analytics
* Configure an Azure Cosmos DB Account to use Azure Synapse Link
* Approach and implement security to protect your data with Azure Synapse Analytics
* Store secrets in Azure Key Vault
* Enable Microsoft Defender for SQL and Data Classification
* Connect sending and receiving applications with Event Hubs to handle extremely high loads without losing data

**Outline**

* Introduction to Azure Synapse Analytics
	+ What is Azure Synapse Analytics
	+ How Azure Synapse Analytics works
	+ When to use Azure Synapse Analytics
	+ Explore Azure Synapse Analytics
* Explore Azure Databricks
	+ Get started with Azure Databricks
	+ Identify Azure Databricks workloads
	+ Understand key concepts
	+ Explore Azure Databricks
* Introduction to Azure Data Lake Storage
	+ Understand Azure Data Lake Storage Gen
	+ Create an Azure Storage Account by using the portal
	+ Compare Azure Data Lake Store to Azure Blob storage
	+ Understand the stages for processing big data by using Azure Data Lake Store
	+ Examine uses for Azure Data Lake Storage Gen
* Get started with Azure Stream Analytics
	+ Understand data streams
	+ Understand event processing
	+ Explore Azure Stream Analytics
* Use Azure Synapse Serverless SQL Pool to Query Files in a Data Lake
	+ Understand Azure Synapse serverless SQL pool capabilities and use cases
	+ Query files using a serverless SQL pool
	+ Create external database objects
	+ Query files using a serverless SQL pool
* Create a Lake Database in Azure Synapse Analytics
	+ Understand lake database concepts
	+ Explore database templates
	+ Create a lake database
	+ Use a lake database
	+ Analyze data in a lake database
* Secure Data and Manage users in Azure Synapse Serverless SQL Pools
	+ Choose an authentication method in Azure Synapse serverless SQL pools3 min
	+ Manage users in Azure Synapse serverless SQL pools3 min
	+ Manage user permissions in Azure Synapse serverless SQL pools
* Use Apache Spark in Azure Databricks
	+ Get to know Spark
	+ Create a Spark cluster
	+ Use Spark in notebooks
	+ Use Spark to work with data files
	+ Visualize data
	+ Use Spark in Azure Databricks
* Use Delta Lake in Azure Databricks
	+ Get Started with Delta Lake
	+ Create Delta Lake tables
	+ Create and query catalog tables
	+ Use Delta Lake for streaming data
	+ Use Delta Lake in Azure Databricks
* Analyze Data with Apache Spark in Azure Synapse Analytics
	+ Get to know Apache Spark
	+ Use Spark in Azure Synapse Analytics
	+ Analyze data with Spark
	+ Visualize data with Spark
	+ Analyze data with Spark
* Integrate SQL and Apache Spark Pools in Azure Synapse Analytics
	+ Describe the integration methods between SQL and spark pools in Azure Synapse Analytics
	+ Understand the use-cases for SQL and spark pools integration
	+ Authenticate in Azure Synapse Analytics
	+ Transfer data between SQL and spark pool in Azure Synapse Analytics
	+ Authenticate between spark and SQL pool in Azure Synapse Analytics
	+ Integrate SQL and spark pools in Azure Synapse Analytics
	+ Externalize the use of spark pools within Azure Synapse Workspace
	+ Transfer data outside the synapse workspace using the PySpark connector
* Use Data Loading Best Practices in Azure Synapse Analytics
	+ Understand data load design goals
	+ Explain load methods into Azure Synapse Analytics
	+ Manage source data files
	+ Manage singleton updates
	+ Set up dedicated data load accounts
	+ Implement workload management
	+ Simplify ingestion with the Copy Activity
* Petabyte-Scale Ingestion with Azure Data Factory or Azure Synapse Pipeline
	+ List the data factory ingestion methods
	+ Describe data factory connectors
	+ Use the data factory copy activity
	+ Manage the self-hosted integration runtime
	+ Set up the Azure integration runtime
* Integrate Data with Azure Data Factory or Azure Synapse Pipeline
	+ Understand Azure Data Factory
	+ Describe data integration patterns
	+ Explain the data factory process
	+ Understand Azure Data Factory components
	+ Azure Data Factory security
	+ Set up Azure Data Factory
	+ Create linked services
	+ Create datasets
	+ Create data factory activities and pipelines
	+ Manage integration runtimes
* Perform Code-Free Transformation at Scale with Azure Data Factory or Azure Synapse Pipeline
	+ Explain Azure Data Factory transformation methods
	+ Describe Azure Data Factory transformation types
	+ Author an Azure Data Factory mapping data flow
	+ Debug mapping data flow
	+ Use Azure Data Factory wrangling data
	+ Use compute transformations within Azure Data Factory
	+ Integrate SQL server integration services packages within Azure Data Factory
* Orchestrate Data Movement and Transformation in Azure Data Factory or Azure Synapse Pipeline
	+ Understand data factory control flow
	+ Work with data factory pipelines
	+ Debug data factory pipelines
	+ Add parameters to data factory components
	+ Integrate a Notebook within Azure Synapse Pipelines
	+ Execute data factory packages
* Plan Hybrid Transactional and Analytical Processing using Azure Synapse Analytics
	+ Understand hybrid transactional and analytical processing patterns
	+ Describe Azure Synapse Link
* Implement Azure Synapse Link with Azure Cosmos DB
	+ Enable Cosmos DB account to use Azure Synapse Link
	+ Create an analytical store enabled container
	+ Create a linked service for Cosmos DB
	+ Query Cosmos DB data with Spark
	+ Query Cosmos DB with Synapse SQL
	+ Implement Azure Synapse Link for Cosmos DB
* 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 authorization through column and row level security
	+ Manage sensitive data with Dynamic Data Masking
	+ Implement encryption in Azure Synapse Analytics
* Configure and Manage Secrets in Azure Key Vault
	+ Guidelines for using Azure Key Vault
	+ Manage access to secrets, certificates, and keys
	+ Store secrets in Azure Key Vault
	+ Manage certificates
* Implement Compliance Controls for Sensitive Data
	+ Explore data classification
	+ Explore server and database audit
	+ Implement Dynamic Data Masking
	+ Implement Row Level Security
	+ Understand Microsoft Defender for SQL
	+ Explore Azure SQL Database Ledger
	+ Implement Azure Purview
	+ Enable Microsoft Defender for SQL and Data Classification
* Enable Reliable Messaging for Big Data Applications using Azure Event Hubs
	+ Create an Event Hub using the Azure CLI
	+ Use the Azure CLI to Create an Event Hub
	+ Configure applications to send or receive messages through an Event Hub
	+ Configure applications to send or receive messages through an Event Hub
	+ Evaluate the performance of the deployed Event Hub using the Azure portal
	+ Evaluate the performance of the deployed Event Hub using the Azure portal
* Conclusion