COURSE OVERVIEW

Course Name:
(DP-203) Data Engineering on Microsoft Azure

COURSE DURATION: 4 Days

Gauteng:
3rd Floor, 34 Whitely Road
Melrose Arch
Johannesburg
2196

Gauteng:
192 on Bram
192 Bram Fischer Drive
Ferndale, Randburg
Johannesburg
2160

Cape Town:
3rd Floor, Thomas Pattullo Building
19 Jan Smuts St
Cape Town
8000

Durban:
9 Mountview Close
Broadlands
Mount Edgecombe
Durban
4302

087 941 5764
sales@impactful.co.za
impactful.co.za

INTRODUCTION

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.

DELIVERY METHOD

Our courses have flexible delivery options:

- In-person classroom training at the Impactful training facilities
  - Johannesburg, Durban, Cape Town
- Virtual instructor-led training
- Nationally: on-site at the client

IMPACTFUL
Powered by LRMG

INTENDED 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.

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

COURSE CONTENT

Introduction to data engineering on Azure

Microsoft Azure provides a comprehensive platform for data engineering; but what is data engineering? Complete this module to find out.

Introduction to Azure Data Lake Storage Gen2

Data lakes are a core element of data analytics architectures. Azure Data Lake Storage Gen2 provides a scalable, secure, cloud-based solution for data lake storage.

Introduction to Azure Synapse Analytics

Learn about the features and capabilities of Azure Synapse Analytics - a cloud-based platform for big data processing and analysis.

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

With Azure Synapse serverless SQL pool, you can leverage your SQL skills to explore and analyze data in files, without the need to load the data into a relational database.

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

By using a serverless SQL pool in Azure Synapse Analytics, you can use the ubiquitous SQL language to transform data in files in a data lake.

Create a lake database in Azure Synapse Analytics

Why choose between working with files in a data lake or a relational database schema? With lake databases in Azure Synapse Analytics, you can combine the benefits of both.

Analyze data with Apache Spark in Azure Synapse Analytics

Apache Spark is a core technology for large-scale data analytics. Learn how to use Spark in Azure Synapse Analytics to analyze and visualize data in a data lake.

Transform data with Spark in Azure Synapse Analytics

Data engineers commonly need to transform large volumes of data. Apache Spark pools in Azure Synapse Analytics provide a distributed processing platform that they can use to accomplish this goal.

Use Delta Lake in Azure Synapse Analytics

Delta Lake is an open source relational storage area for Spark that you can use to implement a data lakehouse architecture in Azure Synapse Analytics.

Analyze data in a relational data warehouse

Relational data warehouses are a core element of most enterprise Business Intelligence (BI) solutions, and are used as the basis for data models, reports, and analysis.

Load data into a relational data warehouse

A core responsibility for a data engineer is to implement a data ingestion solution that loads new data into a relational data warehouse.

Build a data pipeline in Azure Synapse Analytics

Pipelines are the lifeblood of a data analytics solution. Learn how to use Azure Synapse Analytics pipelines to build integrated data solutions that extract, transform, and load data across diverse systems.

Use Spark Notebooks in an Azure Synapse Pipeline

Apache Spark provides data engineers with a scalable, distributed data processing platform, which can be integrated into an Azure Synapse Analytics pipeline.

Plan hybrid transactional and analytical processing using Azure Synapse Analytics

Learn how hybrid transactional / analytical processing (HTAP) can help you perform operational analytics with Azure Synapse Analytics.

Implement Azure Synapse Link with Azure Cosmos DB

Azure Synapse Link for Azure Cosmos DB enables HTAP integration between operational data in Azure Cosmos DB and Azure Synapse Analytics runtimes for Spark and SQL.

Implement Azure Synapse Link for SQL

Azure Synapse Link for SQL enables low-latency synchronization of operational data in a relational database to Azure Synapse Analytics.

Get started with Azure Stream Analytics

Azure Stream Analytics enables you to process real-time data streams and integrate the data they contain into applications and analytical solutions.

Ingest streaming data using Azure Stream Analytics and Azure Synapse Analytics

Azure Stream Analytics provides a real-time data processing engine that you can use to ingest streaming event data into Azure Synapse Analytics for further analysis and reporting.

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

By combining the stream processing capabilities of Azure Stream Analytics and the data visualization capabilities of Microsoft Power BI, you can create real-time data dashboards.

Introduction to Microsoft Purview

In this module, you'll evaluate whether Microsoft Purview is the right choice for your data discovery and governance needs.

Integrate Microsoft Purview and Azure Synapse Analytics

Learn how to integrate Microsoft Purview with Azure Synapse Analytics to improve data discoverability and lineage tracking.

Explore Azure Databricks

Azure Databricks is a cloud service that provides a scalable platform for data analytics using Apache Spark.

Use Apache Spark in Azure Databricks

Azure Databricks is built on Apache Spark and enables data engineers and analysts to run Spark jobs to transform, analyze and visualize data at scale.

Run Azure Databricks Notebooks with Azure Data Factory

Using pipelines in Azure Data Factory to run notebooks in Azure Databricks enables you to automate data engineering processes at cloud scale.