MS-DP3012: IMPLEMENTING A DATA ANALYTICS SOLUTION WITH AZURE SYNAPSE ANALYTICS

DURATION	LEVEL	TECHNOLOGY	DELIVERY METHOD	TRAINING CREDITS
1 Day	Intermediate	Azure	Instructor-led	NA

INTRODUCTION

This is a single-day Instructor Lead Course designed to give the learners instruction on the SQL dedicated and serverless Spark pools and provide instruction of data wrangling and the ELT process using Synapse Pipelines which is very similar to those familiar with Azure Data Factory (ADF) to move data into the Synapse dedicated pool database.

AUDIENCE PROFILE

The Audience should have familiarity with notebooks that use different languages and a Spark engine, such as Databricks, Jupyter Notebooks, Zeppelin notebooks, and more. They should also have some experience with SQL, Python, and Azure tools, such as Data Factory.

PREREQUISITES

This course does not have any formal prerequisites; however, it is beneficial if participants have familiarity with:

– Notebooks that use different languages and a Spark engine, such as Databricks, Jupyter Notebooks, or Zeppelin Notebooks.
– Experience with SQL, Python, and Azure tools like Data Factory.

COURSE OBJECTIVES

After completion of this course, you will be able to:

– Understand and use SQL dedicated and serverless Spark pools.
– Perform data wrangling and the ELT process using Synapse Pipelines, similar to Azure Data Factory (ADF).
– Move data into the Synapse dedicated pool database effectively.

COURSE CONTENT

Module 1: 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.
Lessons

– Introduction.
– What are Azure Synapse Analytics?
– How Azure Synapse Analytics works.
– When to use Azure Synapse Analytics.
– Exercise - Explore Azure Synapse Analytics.
– Knowledge check.
– Summary.
By the end of this module, you'll be able to:

– 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: 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.
Lessons

– Introduction.
– Understand Azure Synapse serverless SQL pool capabilities and use cases.

– Query files using a serverless SQL pool.
– Create external database objects.
– Exercise - Query files using a serverless SQL pool.
– Knowledge check.
– Summary.
By the end of this module, you'll be able to:

– 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 3: Analyse 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 analyse and visualize data in a data lake.

Lessons

– Introduction.
– Get to know Apache Spark.
– Use Spark in Azure Synapse Analytics.
– Analyse data with Spark.
– Visualize data with Spark.
– Exercise - Analyse data with Spark.
– Knowledge check.
– Summary.

By the end of this module, you'll be able to:

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

Module 4: 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.

Lessons

– Introduction.
– Understand Delta Lake.
– Create Delta Lake tables.
– Create catalog tables.
– Use Delta Lake with streaming data.
– Use Delta Lake in a SQL pool.
– Exercise - Use Delta Lake in Azure Synapse Analytics.
– Knowledge check.
– Summary.

By the end of this module, you'll be able to:

– 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 5: Analyse 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.

Lessons

– Introduction.
– Design a data warehouse schema.
– Create data warehouse tables.
– Load data warehouse tables.
– Query a data warehouse.
– Exercise - Explore a data warehouse.
– Knowledge check.
– Summary.

By the end of this module, you'll be able to:

– 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 6: 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.

Lessons

– Introduction.
– Understand pipelines in Azure Synapse Analytics.
– Create a pipeline in Azure Synapse Studio.
– Define data flows.
– Run a pipeline.
– Exercise - Build a data pipeline in Azure Synapse Analytics.
– Knowledge check.
– Summary.

In this module, you'll practice how to:

– 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 7: 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.

Lessons

– Introduction.
– Transform data files with the CREATE EXTERNAL TABLE AS SELECT statement.
– Encapsulate data transformations in a stored procedure.
– Include a data transformation stored procedure in a pipeline.
– Exercise - Transform files using a serverless SQL pool.
– Knowledge check.
– Summary.

After completing this module, you'll be able to:

– 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 8: 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.

Lessons

– Introduction.
– Understand lake database concepts.
– Explore database templates.
– Create a lake database.
– Use a lake database.
– Exercise - Analyze data in a lake database.
– Knowledge check.
– Summary.

After completing this module, you'll be able to:

– Understand lake database concepts and components.
– Describe database templates in Azure Synapse Analytics.
– Create a lake database.

Module 9: Secure data and manage users in Azure Synapse serverless SQL pools
Learn how you can set up security when using Azure Synapse serverless SQL pools

Lessons

– Introduction.
– 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.
– Knowledge check.
– Summary.

After completing this module, you'll be able to:

- 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 10: 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.
Lessons
- Introduction.
- Modify and save dataframes.
- Partition data files.
- Transform data with SQL.
- Exercise: Transform data with Spark in Azure Synapse Analytics.
- Knowledge check.
- Summary.
After completing this module, you'll be able to:
- Use Apache Spark to modify and save data frames.
- Partition data files for improved performance and scalability.
- Transform data with SQL.

Module 11: 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.
Lessons
- Introduction.
- Load staging tables.
- Load dimension tables.
- Load time dimension tables.
- Load slowly changing dimensions.
- Load fact tables.
- Perform post load optimization.
- Exercise - load data into a relational data warehouse.
- Knowledge check.
- Summary.

After completing this module, you'll be able to:
- 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: Manage and monitor data warehouse activities in Azure Synapse Analytics
Learn how to manage and monitor Azure Synapse Analytics.
Lessons
- Introduction.
- 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.
- Knowledge check.
- Summary.
After completing this module, you'll be able to:
- 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.

Module 13: Secure a data warehouse in Azure Synapse Analytics
Learn how to approach and implement security to protect your data with Azure Synapse Analytics.
Lessons
- Introduction.

- Understand network security options for Azure Synapse Analytics.
- Configure Conditional Access.
- Configure authentication.
- Manage authorization through column and row level security.
- Exercise - Manage authorization through column and row level security.
- Manage sensitive data with Dynamic Data Masking.
- Implement encryption in Azure Synapse Analytics.
- Knowledge check.
- Summary.
After completing this module, you'll be able to:
- 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 14: 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.
Lessons
- Introduction.
- Understand Synapse Notebooks and Pipelines.
- Use a Synapse notebook activity in a pipeline.
- Use parameters in a notebook.
- Exercise - Use an Apache Spark notebook in a pipeline.
- Knowledge check.
- Summary.
After completing this module, you'll be able to:
- Describe notebook and pipeline integration.
- Use a Synapse notebook activity in a pipeline.
- Use parameters with a notebook activity.

ASSOCIATED CERTIFICATIONS & EXAM

There is no Associated Certification or Exam for this course.