



DP-3012

Implementing a Data Analytics Solution with Azure Synapse Analytics



Información general.

Este es un curso diseñado para formar a los alumnos sobre los grupos de Spark dedicados y sin servidor de SQL y proporcionar instrucciones de limpieza de datos y el proceso ELT mediante canalizaciones de Synapse, que es muy similar a los que están familiarizados con Azure Data Factory (ADF) para mover datos a la base de datos del grupo dedicado de Synapse.

Duración.

1 Día.

Perfil del público.

El público debe estar familiarizado con los cuadernos que usan diferentes lenguajes y un motor de Spark, como Databricks, Jupyter Notebooks, cuadernos de Zeppelin y mucho más. También deben tener cierta experiencia con las herramientas de SQL, Python y Azure, como Data Factory.

Examen.

Este curso no tiene un examen asociado.

Temario.

Módulo 1: Introducción a Azure Synapse Analytics.

Obtenga información sobre las características y funcionalidades de Azure Synapse Analytics: una plataforma basada en la nube para el procesamiento y el análisis de macrodatos.

- Qué es Azure y cómo funciona Azure Synapse Analytics.
- Cuándo usar Azure Synapse Analytics.

Ejercicio: exploración de Azure Synapse Analytics.

En este módulo aprenderá a:

- Identificación de los problemas empresariales que Azure Synapse Analytics soluciona.
- Descripción de las funcionalidades principales de Azure Synapse Analytics.
- Determinación de cuándo usar Azure Synapse Analytics.

Módulo 2: Uso de un grupo de SQL sin servidor de Azure Synapse para consultar archivos en un lago de datos.

Con un grupo de SQL sin servidor de Azure Synapse, puede aprovechar sus aptitudes de SQL para explorar y analizar datos en archivos, sin la necesidad de cargar los datos en una base de datos relacional.

- Descripción de las capacidades y los casos de uso de los grupos de SQL sin servidor de Azure Synapse.
- Consulta de archivos mediante un grupo de SQL sin servidor.
- Creación de objetos de base de datos externos.

Ejercicio: Consulta de archivos mediante un grupo de SQL sin servidor.

Al finalizar este módulo podrá hacer lo siguiente:

- Identificación de las capacidades y los y casos de uso de los grupos de SQL sin servidor en Azure Synapse Analytics.
- Consulta de archivos CSV, JSON y Parquet mediante un grupo de SQL sin servidor.
- Creación de objetos de base de datos externos en un grupo de SQL sin servidor.



Módulo 3: Análisis de datos con Apache Spark en Azure Synapse Analytics.

Apache Spark es una de las tecnologías principales para el análisis de datos a gran escala. Aprenda a usar Spark en Azure Synapse Analytics para analizar y visualizar datos en un lago de datos.

- Introducción a Apache Spark.
- Uso de Spark en Azure Synapse Analytics.
- Análisis de datos con Spark.
- Visualización de datos con Spark.

Ejercicio: Análisis de datos con Spark.

Tras finalizar este módulo, podrá:

- Identificar las características y las capacidades principales de Apache Spark.
- Configurar un grupo de Spark en Azure Synapse Analytics.
- Ejecutar código para cargar, analizar y visualizar datos en un cuaderno de Spark.

Módulo 4: Uso de Delta Lake en Azure Synapse Analytics.

Delta Lake es un área de almacenamiento relacional de código abierto para Spark que puede usar para implementar una arquitectura de lago de datos en Azure Synapse Analytics.

- Descripción de Delta Lake.
- Creación de tablas de Delta Lake.
- Creación de tablas de catálogo.
- Uso de Delta Lake con datos de streaming.
- Uso de Delta Lake en un grupo de SQL.

Ejercicio: Uso de Delta Lake en Azure Synapse Analytics.

En este módulo aprenderá a:

- Describir las características y funcionalidades principales de Delta Lake.
- Crear y usar tablas de Delta Lake en un grupo de Spark de Synapse Analytics.
- Crear tablas de catálogo de Spark para datos de Delta Lake.
- Usar tablas de Delta Lake para transmitir datos.
- Consultar tablas de Delta Lake desde un grupo de SQL de Synapse Analytics.

Módulo 5: Análisis de datos en un almacenamiento de datos relacional.

Los almacenes de datos relacionales son un elemento principal de la mayoría de las soluciones empresariales de Business Intelligence (BI) y se usan como base para los modelos de datos, los informes y el análisis.

- Diseñar un esquema de almacenamiento de datos.
- Crear tablas de almacenamiento de datos.
- Carga de tablas de almacenamiento de datos.
- Consultar un almacenamiento de datos.

Ejercicio: Exploración de un almacenamiento de datos.

En este módulo aprenderá a:

- Diseñar un esquema para un almacenamiento de datos relacional.
- Crear tablas de hechos, dimensiones y almacenamiento provisional.
- Usar SQL para cargar datos en tablas de almacenamiento de datos.
- Usar SQL para consultar tablas de almacenamiento de datos relacionales.

Módulo 6: Creación de una canalización de datos en Azure Synapse Analytics.

Las canalizaciones son la esencia de una solución de análisis de datos. Las canalizaciones son la esencia de una solución de análisis de datos.

- Descripción de las canalizaciones en Azure Synapse Analytics.
- Creación de una canalización en Azure Synapse Studio.
- Definición de flujos de datos.
- Ejecución de una canalización.

Ejercicio: Creación de una canalización de datos en Azure Synapse Analytics.

En este módulo, aprenderá a lo siguiente:

- Describir los conceptos básicos de las canalizaciones.
- Crear una canalización en Azure Synapse Studio.
- Implementar una actividad de flujo de datos en una canalización.
- Iniciar y supervisar ejecuciones de canalización.

