



DP-601T00

Implementing a Lakehouse with Microsoft Fabric



Información general.

Este curso está diseñado para crear sus habilidades fundamentales en la ingeniería de datos en Microsoft Fabric, centrándose en el concepto de almacén de lago. En este curso se explorarán las eficaces funcionalidades de Apache Spark para el procesamiento de datos distribuidos y las técnicas esenciales para la eficiencia en la administración de datos, el control de versiones y la confiabilidad trabajando con tablas de Delta Lake.

En este curso también se explorará la ingesta y orquestación de datos mediante canalizaciones de Dataflows Gen2 y Data Factory. En este curso se incluye una combinación de conferencias y ejercicios prácticos que le prepararán para trabajar con almacenes de lagos en Microsoft Fabric.

Duración.

1 Día.

Perfil del público.

La audiencia principal de este curso son profesionales de datos que están familiarizados con el modelado, la extracción y el análisis de datos. Está diseñado para profesionales interesados en obtener conocimientos sobre la arquitectura del almacén de lagos, la plataforma Microsoft Fabric y el proceso de habilitación del análisis de un extremo a otro mediante estas tecnologías.

Examen.

Este curso no tiene un examen asociado.

Temario.

Módulo 1: Introducción al análisis de un extremo a otro con Microsoft Fabric.

Descubra cómo Microsoft Fabric puede satisfacer las necesidades de análisis de su empresa en una plataforma. Obtenga información sobre Microsoft Fabric y cómo funciona e identifique cómo puede usarlo para sus necesidades de análisis.

- Exploración del análisis de un extremo a otro con Microsoft Fabric.
- Equipos de datos y Microsoft Fabric.
- Habilitación y uso de Microsoft Fabric.

En este módulo aprenderá a:

- Análisis de un extremo a otro en Microsoft Fabric

Módulo 2: Introducción a los almacenes de lago en Microsoft Fabric.

Los almacenes de lago combinan la flexibilidad del almacenamiento en lagos de datos con el análisis de almacenamiento de datos. Microsoft Fabric ofrece una solución de almacén de lago para un análisis exhaustivo en una sola plataforma SaaS.

- Exploración del almacén de lago de Microsoft Fabric.
- Trabajar con almacenes de lago de Microsoft Fabric.
- Exploración y transformación de datos en un almacén de lago.

Ejercicio: Creación e ingesta de datos con un almacén de lago de Microsoft Fabric.



En este módulo aprenderá a:

- Descripción de las características y las capacidades principales de los almacenes de lago en Microsoft Fabric.
- Creación de un almacén de lago.
- Ingesta de datos en archivos y tablas de un almacén de lago.
- Consulta de tablas de almacén de lago con SQL.

Módulo 3: Uso de Apache Spark en Microsoft Fabric.

Apache Spark es una de las tecnologías principales para el análisis de datos a gran escala. Microsoft Fabric proporciona compatibilidad con clústeres de Spark, lo que le permite analizar y procesar datos en un almacén de lago a escala.

- Preparación para usar Apache Spark.
- Ejecución de código de Spark.
- Trabajo con datos en un DataFrame de Spark.
- Trabajo con datos con Spark SQL.
- Visualización de datos en un cuaderno de Spark.

Ejercicio: Análisis de datos con Apache Spark.**En este módulo aprenderá a:**

- Configurar Spark en un área de trabajo de Microsoft Fabric.
- Identificar escenarios adecuados para cuadernos y trabajos de Spark.
- Usar DataFrames de Spark para analizar y transformar datos.
- Usar Spark SQL para consultar datos en tablas y vistas.
- Visualizar datos en un cuaderno de Spark.

Módulo 4: Uso de tablas de Delta Lake en Microsoft Fabric.

Las tablas de un almacén de lago de Microsoft Fabric se basan en el formato de almacenamiento de Delta Lake que se usa habitualmente en Apache Spark. Mediante las funcionalidades mejoradas de las tablas delta, puede crear soluciones de análisis avanzadas.

- Descripción de Delta Lake.
- Creación de tablas delta.
- Uso de tablas delta en Spark.
- Uso de tablas delta con datos de streaming.

Ejercicio: Uso de tablas delta en Apache Spark.**En este módulo aprenderá a:**

- Descripción de tablas delta y de Delta Lake en Microsoft Fabric.
- Creación y administración de tablas delta mediante Spark.
- Uso de Spark para consultar y transformar datos en tablas delta.
- Uso de tablas delta con Spark Structured Streaming.

Módulo 5: Ingesta de datos con flujos de datos Gen2 en Microsoft Fabric.

La ingesta de datos es fundamental en los análisis. Microsoft Fabric Data Factory ofrece flujos de datos para crear visualmente la ingesta y transformación de datos en varios pasos con Power Query Online.

- Descripción de los flujos de datos Gen2 en Microsoft Fabric.
- Exploración de flujos de datos Gen2 en Microsoft Fabric.
- Integración de flujos de datos Gen2 y canalizaciones en Microsoft Fabric.

Ejercicio: Creación y uso de un flujo de datos Gen2 en Microsoft Fabric.**En este módulo aprenderá a:**

- Descripción de las funcionalidades de flujo de datos en Microsoft Fabric.
- Creación de soluciones de flujo de datos para ingerir y transformar datos.
- Inclusión de un flujo de datos en una canalización.

Módulo 6: Uso de canalizaciones de Data Factory en Microsoft Fabric.

Microsoft Fabric tiene funcionalidades de Data Factory, incluida la capacidad de crear canalizaciones que organizan tareas de ingesta y transformación de datos.

- Descripción de las canalizaciones.
- Uso de la actividad Copiar datos.
- Uso de plantillas de canalización.
- Ejecución y supervisión de canalizaciones.





DP-601T00

Implementing a Lakehouse with Microsoft Fabric

A

Ejercicio: Ingesta de datos con una canalización.

En este módulo aprenderá a:

- Descripción de funcionalidades de canalización en Microsoft Fabric.
- Uso de la actividad Copiar datos en una canalización.
- Creación de canalizaciones basadas en plantillas predefinidas.
- Ejecución y supervisión de canalizaciones.

