



DP-3011

Implement a Data Analytics Solution with Azure Databricks



Información general.

Obtenga información sobre cómo aprovechar todas las ventajas de Apache Spark y los eficaces clústeres que se ejecutan en la plataforma de Azure Databricks para ejecutar grandes cargas de trabajo de ingeniería de datos en la nube.

Duración.

1 Día.

Examen.

Este curso no tiene un examen asociado.

Temario.

Módulo 1: Explorar Azure Databricks.

Azure Databricks es un servicio en la nube que proporciona una plataforma escalable para el análisis de datos mediante Apache Spark.

- Introducción a Azure Databricks.
- Identificación de las cargas de trabajo de Azure Databricks.
- Descripción de los conceptos clave.

Ejercicio: Explorar Azure Databricks.

Módulo 2: Uso de Apache Spark en Azure Databricks.

Azure Databricks se basa en Apache Spark y permite a los ingenieros y analistas de datos ejecutar trabajos de Spark para transformar, analizar y visualizar datos a escala.

- Descubra Spark.
- Creación de un clúster de Spark.
- Uso de Spark en cuadernos.
- Uso de Spark para trabajar con archivos de datos.
- Visualización de datos.

Ejercicio: Uso de Spark en Azure Databricks.

Módulo 3: Uso de Delta Lake en Azure Databricks.

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 Databricks.

- Primeros pasos con Delta Lake.
- Creación de tablas de Delta Lake.
- Creación y consulta de tablas de catálogo.
- Uso de Delta Lake para transmitir datos.

Ejercicio: Uso de Delta Lake en Azure Databricks.

Módulo 4: Uso de almacenes SQL en Azure Databricks.

Azure Databricks proporciona almacenes SQL que permiten que los analistas trabajen con datos a través de consultas SQL relacionales conocidas.

- Introducción a los almacenes SQL.
- Creación de bases de datos y tablas.
- Creación de consultas y paneles.





DP-3011

Implement a Data Analytics Solution with Azure Databricks

A

Ejercicio: Uso de un almacén SQL en Azure Databricks.

Módulo 5: Ejecución de cuadernos de Azure Databricks con Azure Data Factory.

El uso de canalizaciones en Azure Data Factory para ejecutar cuadernos en Azure Databricks permite automatizar procesos de ingeniería de datos a escala de la nube.

- Descripción de los cuadernos y las canalizaciones de Azure Databricks.
- Creación de un servicio vinculado para Azure Databricks.
- Uso de una actividad de cuaderno en una canalización.
- Uso de parámetros en un cuaderno.

Ejercicio: Ejecución de un cuaderno de Azure Databricks con Azure Data Factory.

