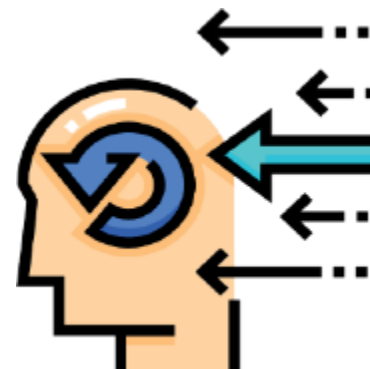




Módulo 7

Sesión N° 4



ACTIVIDAD:



Uso de Spark SQL para Consultas de Sensores de una Represa Eléctrica

- Objetivo: El objetivo de esta actividad es aprender a utilizar Spark SQL para realizar consultas sobre datos provenientes de sensores instalados en una represa eléctrica. A través de esta actividad, los estudiantes aprenderán a trabajar con datos estructurados en Spark, cómo registrar un DataFrame como una vista temporal, y cómo ejecutar consultas SQL para realizar análisis de los datos.



Instrucciones:

Descripción de la Data:

La data que utilizaremos será de sensores de una represa eléctrica, donde se incluyen las siguientes columnas:

- Fecha: Fecha y hora del registro de la medición.
- SensorID: Identificador único del sensor.
- Temperatura: Temperatura del agua medida en grados Celsius.
- Presión: Nivel de presión en la represa (en Pa).
- NivelAgua: Nivel del agua en la represa (en metros).
- Caudal: Caudal de agua que pasa por la turbina (en litros por segundo).
- Estado: Estado operativo del sensor (activo/inactivo).



```
# Crear la sesión de Spark
spark = SparkSession.builder.appName("RepresaElectricidadSQL").getOrCreate()

# Generar datos ficticios para los sensores
data = []
for i in range(100):
    sensor_id = f"S{i+1}"
    fecha = datetime.now() - timedelta(days=i)
    temperatura = round(uniform(15.0, 30.0), 2) # Temperatura entre 15 y 30 grados
    nivel_agua = round(uniform(5.0, 20.0), 2) # Nivel de agua entre 5 y 20 metros
    caudal = randint(100, 500) # Caudal entre 100 y 500 litros por segundo
    presion = randint(900, 1500) # Presión entre 900 y 1500 Pascales
    estado = "Activo" if randint(0, 1) == 0 else "Inactivo" # Estado aleatorio

    data.append((sensor_id, fecha, temperatura, nivel_agua, caudal, presion, estado))

# Definir los nombres de las columnas
columns = ["Sensor_ID", "Fecha", "Temperatura", "Nivel_Agua", "Caudal", "Presion", "Estado"]

# Crear el DataFrame
df_sensores = spark.createDataFrame(data, columns)

# Mostrar el DataFrame
df_sensores.show(5)
return go(f, seed, [])
```



Realizar Consultas Básicas usando SQL

- Consultar los sensores activos que tienen un nivel de agua superior a 10 metros.
- Obtener el promedio de la temperatura de todos los sensores.
- Mostrar los sensores con el caudal más alto.
- Filtrar los sensores inactivos y mostrar la fecha y la presión de las lecturas.
- Calcular el total de caudal acumulado por sensor.

► Modalidad grupal.

► Tiempo estimado: 60 minutos.

