



BIG DATA
ACADEMY

Ejercicios Spark

FORMADOR: ALONSO MELGAREJO
alonsoraulmgs@gmail.com

EJERCICIOS SPARK

1. [EJERCICIO 1] SPARK SQL

Con las tablas:

- main_UNIVERSAL.PERSONA
- main_UNIVERSAL.EMPRESA
- main_UNIVERSAL.TRANSAccion

Construir la tabla "userXXX_UNIVERSAL.TRANSAccion_ENRIQUECIDA", donde "userXXX" es el nombre de tu usuario. La tabla contiene los siguientes campos

CAMPO	TIPO	DESCRIPCIÓN
ID_PERSONA	INT	ID de la persona que realizó la transacción
NOMBRE_PERSONA	STRING	Nombre de la persona que realizó la transacción
EDAD_PERSONA	INT	Edad de la persona que realizó la transacción
SALARIO_PERSONA	DOUBLE	Salario de la persona que realizó la transacción
TRABAJO_PERSONA	STRING	Nombre de la empresa donde trabaja la persona
MONTO_TRANSAccion	DOUBLE	Monto de la transacción realizada
FECHA_TRANSAccion	STRING	Fecha de la transacción realizada
EMPRESA_TRANSAccion	STRING	Nombre de la empresa donde se realizó la transacción

2. [EJERCICIO 2] SPARK SQL

Con la tabla anterior, generar estos 4 reportes en dataframes:

- **DATAFRAME: "dfReporte1"**
Filtrar y almacenar las transacciones realizadas por mayores de 50 años.

- **DATAFRAME: "dfReporte2"**
Transacciones realizadas por mayores de 25 años, cuyo nombre empiece con la letra "a".
- **DATAFRAME: "dfReporte3"**
Transacciones realizadas cuyo monto de transacción sea al menos el 25% del salario de la persona que la realizó.
- **DATAFRAME: "dfReporte4"**
Transacciones realizadas en la misma empresa donde la persona trabaja.

3. [EJERCICIO 3] SPARK SQL

Con el DATAFRAME "dfReporte1" generar el reporte "dfReporte5" que agrupe las transacciones para cada persona y muestre los siguientes campos:

CAMPO	TIPO	DESCRIPCIÓN
NOMBRE_PERSONA	STRING	Nombre de la persona
SUMA_TRANSACCIONES	DOUBLE	Suma de todos los montos de las transacciones realizadas por la persona
MAYOR_TRANSACCION	DOUBLE	Mayor monto gastado de todas las transacciones realizadas por la persona
MENOR_TRANSACCION	DOUBLE	Menor monto gastado de todas las transacciones realizadas por la persona
PROMEDIO_TRANSACCIONES	DOUBLE	Promedio de montos gastado de todas las transacciones realizadas por la persona

4. [EJERCICIO 4] SPARK DATAFRAMES

Implementar los ejercicios 3, 4 y 5 utilizando SPARK DATAFRAMES