



BIG DATA
ACADEMY

EJERCICIO 1 – SPARK SQL

FORMADOR: ALONSO MELGAREJO
alonsoraulmgs@gmail.com

EJERCICIO SPARK SQL

1. Leer el archivo "persona.data" en un dataframe
2. Leer el directorio "transaccion" en un dataframe
3. Registrar ambos dataframes como una "TempView"
4. Cruzar "persona" y "transaccion" por el campo "id_persona" para obtener el siguiente dataframe:

ID_PERSONA
NOMBRE_PERSONA
EDAD_PERSONA
ID_EMPRESA
MONTO
FECHA

5. Filtrar el dataframe anterior, quedándonos con las transacciones hechas por personas cuya edad sea mayor a 25 años.
6. Filtrar el dataframe anterior, quedándonos con las transacciones realizadas por personas cuya inicial de nombre comienza con la letra "A"
7. Almacenar el resultante anterior en "/FileStore/_spark/output/df3"

BIG DATA
ACADEMY