Lab sheet 3 - SparkSQL

Q1. Write a UDF to convert a given text to upper case.

Q2. Execute the following steps and show the output

a. Import a Spark Session into Apache Spark.
b. Create a Spark Session 'spark' using the 'builder()' function.
c. Import the Implicts class into 'spark' Session.
d. Now create a DataFrame 'df' and import data from the 'employee.json' file.

e. Print the schema of 'df' DataFrame.
f. Display the DataFrame 'df'.

(The result is a table of 5 rows of ages and names from our 'employee.json' file.)

Q3.

a. Display the DataFrame after incrementing everyone's age by two years.
b. Filter all the employees above age 30 and display the result.
c. Count the number of people with the same ages. (use the 'groupBy' function for the same.)
d. Creating a temporary view 'employee' of 'df' DataFrame.
e. Perform a 'select' operation on our 'employee' view to display the table into 'sqlDF'.
f. Display the results of 'sqlDF'.

Q4.

a. Create a class 'Employee' to store name and age of an employee.
b. Assigning a Dataset 'caseClassDS' to store the record of Andrew.
c. Display the Dataset 'caseClassDS'.
d. Create a primitive Dataset to demonstrate mapping of DataFrames into Datasets.
e. Assign the above sequence into an array.
f. Display the result.

Reference link:

https://www.edureka.co/blog/spark-sql-tutorial/