

Lab 2- Assignment

All students must take clear screenshots of each step, including starting the cluster, loading the dataset, running DataFrame operations, executing SQL queries, applying UDFs, and viewing the final results for every task and sub-task. At the end of the file, you must include the complete code for each task, and then upload everything in a PDF file.

Task 1:

Write more than five sentences, open both terminals at the same time, take one picture each time you add a sentence, capture five pictures for every sentence, ensure all photos are clear and saved properly, and review everything once you finish.

When you receive an sentences from socket, put your name before each word in sentences , and put last name after that(use UDF)

Input: **Sweden is a good country**

Output look like:

azad**Sweden**shokrollahi,azad**i**shokrollahi,azad**a**shokrollahi,azad**good**shokrollahi,azad**country**shokrollahi

Task 2

Average word length from all the words received from socket

Use SQL

Write more than ten words, open both terminals at the same time, take one picture each time you add a word, capture five pictures for every word, ensure all photos are clear and saved properly, and review everything once you finish.

Task 3:

Use SQL to filter only the even numbers from the streaming data.

Write more than ten numbers, open both terminals at the same time, take one picture each time you add a number, capture five pictures for every number, ensure all photos are clear and saved properly, and review everything once you finish.