Lab - 4

Deadline: 23-02-2023 23:59:59

Assignment

1. Write a spark code for executing the Hash example provided in slide 13 on Hashing from Lab 1 Presentation, on the public file: 'gs://bdl-bucket-1/lab4_dataset.csv'. You would have to find the number of user clicks between 0-6, 6-12, 12-18, and 18-24, as was discussed in the first class.

a. Submit the python file with your code.
b. Also, provide the text file containing your output.

2. Provide a brief description of the functionality of the following services:

a. HDFS
b. Hive
c. Pig
d. Yarn