National University of Computer and Emerging Sciences

Laboratory Manual-07

for

Fundamentals of Big Data Lab

Course Instructor: Dr Zareen Alamgir, Iqra Safder
Lab Instructors: Mr. Muhammad Mazarib; Ms Rida Mahmood
Section: BDS-4A
Date: 21-Mar-2023
Semester: Spring 2023

Department of Computer Science

FAST-NU, Lahore, Pakistan

Map-Reduce Concepts:

Do Not use Linux, Use only Jupyter Notebook and Google Collab for lab.

QUESTION:
Consider the weblog of a music streaming application containing the username, date, and the list of songs played by the user on that date. You can assume that the username is unique. The format is as follows.

Username, Date, list of song IDs
Ali, 12-3-20, S1, S2,S3
Fahad 14-04-22, S1, S2, S7,S3
Ali 20-4-22 S1, S2
Sara 20-4-22 S1, S7, S9
Amina 20-4-22 S1

1. Write a map-only job to list the name of users who have played more than 5 songs in a day.
2. Write a map-only job to list the name of users, the date, and the number of songs played on that day.
3. Write a map-reduce job to list the number of songs played by each user up till now. Also, add a combiner.
4. Write a map-reduce to find the average number of songs played daily. Also, add a combiner. For the above data, the average no of songs would be = (3+4+2+3+1)/5
5. Write a map-reduce to find the number of times each song was played. Also, add a combiner.
6. \Write a map-reduce to find the names of the users who have played the maximum number of songs in a day. Also, add a combiner.
7. Write a map-reduce to find the names of the users who have played the maximum number of songs up till now. Also, add a combiner.
8. Write a map-reduce to find the names of the users who have played the maximum number of songs in a day. Also, add a combiner.

NOTE for some tasks above you may need a two-step job.