National University of Computer and Emerging Sciences

Laboratory Manual-07

for

Fundamentals of Big Data Lab

Course Instructor: Iqra Safdar
Lab Instructors: Mr. Muhammad Mazarib; Mr Muhammad Aiss Shahid
Section: BDS-4B
Date: 15-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.

<u>**Exercises:**</u>

1. Complete the documentation of MRJOB library (upto map-reduce and apache hadoop heading) from above link thoroughly. Also cover Single-StepJobs writing for mappers and reducers. Try to resolve all your queries.
   MRJOB documentation https://mrjob.readthedocs.io/en/latest/
2. Run and understand the all given examples from the following link.
   https://www.geeksforgeeks.org/hadoop-mrjob-python-library-for-mapreduce-with-example/
3. Write a map-reduce code to count Words in a file using MRJob library. Write a mapper function as well as reducer function.
4. Display How many words start with each alphabet in a text file by making map and reduce functions separately.
5. Count all words of length 5 from a text file and display them on a screen using mapper and reducers.

   Bonus Question
6. Write a map-reduce code to Invert Index of each word in a text file using MRJob library

Inverted Index Example

■ This was the original Google's usecase
■ Generate an inverted index of words from a given set of files

- **Map:**
  ▫ parses a document and emits <word, docId> pairs
- **Reduce:**
  ▫ takes all pairs for a given word, sorts the docId values, and emits a <word,list(docId)> pair