

National
University of Computer and Emerging Sciences



Laboratory Manual-07

for

Fundamentals of Big Data Lab

Course Instructor: Iqra Safdar
Lab Instructors: Rida Mahmood, Mr. Muhammad Mazarib
Section: BDS-4A
Date: 14-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.

Exercises:

1. Install MRJOB Library by following the instructions in the link below.
2. 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/>

3. Run and understand the all given examples from the following link.
<https://www.geeksforgeeks.org/hadoop-mrjob-python-library-for-mapreduce-with-example/>
4. Write a map-reduce code to count Words in a file using MRJob library. Write a mapper function as well as reducer function.
5. Display How many words start with each alphabet in a text file by making map and reduce functions separately.
6. Count all words of length 5 from a text file and display them on a screen using mapper and reducers.