VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

# School of Computer Science and Engineering

## LAB - 8 Exercises

| | | | | | |
|---|---|---|---|---|---|
| **Course Code** | **:** | **CSE3025 – Large Scale Data Processing** | **Date** | **:** | **09/04/2021** |
| **Lab Experiment** | **:** | **Practice of MapReduce Programming using Partitioner and combiner concepts** | **Slots** | **:** | **L29+L30** |
| **Instructors** | **:** | **Dr. Ramesh Ragala** | | | |

Objective:

1. To understand the detailed processing of MapReduce Framework using pratitioner and combiner concepts

**Problem- 1:**

Consider a Data10.csv dataset with attributes such as S.No., Emp.Id, Name, School and Count. Use the ten reducers in MapReduce framework to compute the total number of duties contributed from each school using partitioner concept. Hint: number of schools = number of reducers. Please use write individual java files to mapper, reducer, partitioner and driver program.

Dataset link:

https://github.com/rameshragala/Large-Scale-Data-Processing-Winter-2020-2021-/tree/main/Lab/Week-8

**Problem - 2:**

The baby-name-dataset.csv consists of name, gender, age and score as attributes. The user want to find the maximum score for each gender in the age groups 0-18, 18-30,

30-40, and 40-50 , then develop a MapReduce program using partition data according to some logic into four groups and each group data will processed on same reducer.

Dataset Name: baby-names.csv

Please get the dataset from

https://github.com/rameshragala/Large-Scale-Data-Processing-Winter-2020-2021-/tree/main/Lab/Week-8

## Problem - 3:

Write a MapReduce application using Combiner to find the number of records for each exchange of a given stock mark dataset. The attributes of the stock market dataset are id, exchange, stockname, sector, country, date, open, high, low, close, volume, adj_close.

Please use the following link to download dataset:

https://github.com/protechskills/stock-data