



VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

## School of Computing Science and Engineering

### LAB - 7 Exercises

<b>Course Code</b>	<b>:</b>	<b>CSE3025 - Large Scale Data Processing</b>	<b>Date</b>	<b>:</b>	<b>18/09/2019</b>
<b>Lab Experiment</b>	<b>:</b>	<b>Practice of MapReduce Programming using Partitioner concept</b>	<b>Slots</b>	<b>:</b>	<b>L15+L16</b>
<b>Instructors</b>	<b>:</b>	<b>Dr. Bharadwaja Kumar and Prof. Ramesh Ragala</b>			

Objective:

1. To understand the detailed processing of MapReduce Framework using partitioner

Problem- 1:

Consider a student dataset with attributes student\_name and marks respectively. Create your own dataset based on the attributed discussed. Use the two reducers in MapReduce framework to group all the students whose scored mark is greater than 60 using partitioner concept. Please use write individual java files to mapper, reducer, partitioner and driver program.



**VIT<sup>®</sup>**

**Vellore Institute of Technology**

(Deemed to be University under section 3 of UGC Act, 1956)

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/CSE3025-Large-Scale-Data-Processing-/tree  
/master/Lab/Week%20-%207](https://github.com/rameshragala/CSE3025-Large-Scale-Data-Processing-/tree/master/Lab/Week%20-%207)