



VIT[®]

Vellore Institute of Technology

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

School of Computing Science and Engineering

LAB - 5 Exercises

Course Code	:	CSE3025 - Large Scale Data Processing	Date	:	28/08/2019
Lab Experiment	:	MapReduce Programming Exercise on Movie Lens Dataset	Slots	:	L15+L16
Instructors	:	Dr. Bharadwaja Kumar and Prof. Ramesh Ragala			

Objective:

1. To understand the detailed processing of MapReduce Framework

Problem- 1: Write the MapReduce Program to determine the average ratings of movies. The input dataset consists of a series of lines, each containing a user number(USERID), movie number (MOVIEID), rating, and a timestamp.

The Dataset (**MovieLenDataset.base**) is in my github page:

<https://github.com/rameshragala/CSE3025-Large-Scale-Data-Processing-/blob/master/Lab/Week%20-%205/MovieLenDataset.base>

Problem - 2: Develop the python MapReduce programs to the above problems also.