School of Computing Science and Engineering

LAB - 1 Exercises

Course Code	:	CSE6017 – Mining Massive Dataset	Date	:	03/02/2021
Lab Experiment	:	K-Means Algorithm in MapReduce and Spark Environment	Slots	:	L45+L46
Instructors	:	Dr. Ramesh Ragala			

Objective:

To implement k-means clustering (llyod approach) in MapReduce and Apache Spark Environments.

Assume your own dataset to test the K-means algorithm based on MapReduce and Spark. Test the performance of the k-means in both MapReduce and Spark framework