

CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY
FACULTY OF TECHNOLOGY AND ENGINEERING
DEPSTAR

Subject Name: Data Science and Analytics
Subject Code: CS442

Semester: VII
Academic year: 2021-22

Practical List

	AIM	Hours	CO	PO	PEO
1.	To install Hadoop framework, configure it and setup a single node cluster. Use web based tools to monitor your Hadoop setup.	4	1,2	1,3,4,5	1,2
2.	To implement file management tasks in Hadoop HDFS and perform Hadoop commands.	2	1,2	1,2,3,5	1,2
3.	To implement basic functions and commands in R Programming. To build WordCloud, a text mining method using R for easy to understand and better visualization than a data table.	2	1,2	1,2,3,5	1,2
4.	A. To implement a word count application using the MapReduce programming model. B. To implement program that count the occurrences of word based on the length.	4	1,2	1,3,4,5	1,2
5.	A. To design and implement MapReduce algorithms to take a very large file of integers and produce as output: a) The largest integer b) The average of all the integers. c) The same set of integers, but with each integer appearing only once. d) The count of the number of distinct integers in the input. B. To design an application to find mutual friend using map reduce.	2	1,2,3,5	1,2,3,5	1,2
6.	To implement basic CRUD operations (create, read, update, delete) in MongoDB and Cassandra.	2	1,2	2,3	1,2
7.	To develop a MapReduce application and implement a program that analyzes weather data.	2	1,2,3,4	1,2,3,5,11	1,2
8.	To Install and Run Hive. Use Hive to create, alter, and drop databases, tables, views, functions, and indexes. To create HDFS tables and load them in Hive and implement joining of tables in Hive.	2	1,2,3	2,5	1,2
9.	To install and run Pig and then write Pig Latin scripts to sort, group, join, project, and filter your data.	2	1,2,3	1,2	1,2
10.	To install, deploy & configure Apache Spark Cluster. To Select the fields from the dataset using Spark SQL. To explore Spark shell and read from HDFS.	4	1,2,5	1,2,3,5	1,2
11.	To perform Sentiment Analysis using Twitter data, Scala and Spark	4	1,2,5	1,2,3,5	1,2
12.	To perform Graph Path and Connectivity analytics and implement basic queries after loading data using Neo4j	2	1,2,5	2,5,11	1,2,5
13.	To perform case study of the following platforms for solving any big data analytic problem of your choice. (1) Amazon web services,(2) Microsoft Azure, (3)Google App engine	2	1,2,5	2,5,11	1,2,5