**CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY**

**FACULTY OF TECHNOLOGY AND ENGINEERING**

DEPSTAR

**Subject Name:** Data Science and Analytics **Semester:** VII

**Subject Code:** CS442 **Academic year:** 2020-21

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **AIM** | | **LO** | **PO** | **PEO** |
|  | To install Hadoop framework, configure it and setup a single node cluster. Use web based tools to monitor your Hadoop setup. | 2 | 1,2 | 1,3,  4,5 |
|  | To implement file management tasks in Hadoop HDFS like adding, retrieving and deleting files. | 2 | 1,2 | 1,2, 3,5 |
|  | 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 |
|  | To implement a word count application using the MapReduce API. | 2 | 1,2 | 1,3,  4,5 |
|  | 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. | 2 | 1,2,3,5 | 1,2, 3,5 |
|  | To implement basic CRUD operations (create, read, update, delete) in MongoDB and Cassandra | 2 | 1,2 | 2,3 |
|  | To develop a MapReduce application and implement a program that mines weather data. | 2 | 1,2,3,4 | 1,2, 3,5,11 |
|  | To Install and Run Hive then 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 |
|  | 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 |
|  | 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. | 2 | 1,2,5 | 1,2, 3,5 |
|  | To perform PYspark shell exploration and reading and writing in HDFS. To perform Clustering and Regression using MLlib. To perform Sentiment Analysis and use Scala. | 4 | 1,2,5 | 1,2, 3,5 |
|  | To perform Graph analytics and visualization using Neo4j, D3.JS | 2 | 1,2,5 | 2,  5,11 |
|  | To use following platforms for solving any big data analytic problem of your choice. (1) Amazon web services,(2) Micosoft Azure, (3)Google App engine | 4 | 1,2,5 | 2,  5,11 |

**Practical List**