8CS4-01: Big Data Analytics

Credit: 3 **Max. Marks: 150(IA:30, ETE:120)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:**Objective, scope and outcome of the course.	01
2	**Introduction to Big Data:** Big data features and challenges, Problems with Traditional Large-Scale System , Sources of Big Data, 3 V's of Big Data, Types of Data. Working with Big Data: Google File System. Hadoop Distributed File System (HDFS) - Building blocks of Hadoop (Namenode. Data node. Secondary Namenode. Job Tracker. Task Tracker), Introducing and Configuring Hadoop cluster (Local. Pseudo-distributed mode, Fully Distributed mode). Configuring XML files.	10
3	**Writing MapReduce Programs:** A Weather Dataset. Understanding Hadoop API for MapReduce Framework (Old and New). Basic programs of Hadoop MapReduce: Driver code. Mapper code, Reducer code. Record Reader, Combiner, Partitioner.	08
4	**Hadoop I/O:** The Writable Interface. Writable Comparable and comparators. Writable Classes: Writable wrappers for Java primitives. Text. Bytes Writable. Null Writable, Object Writable and Generic Writable. Writable collections. Implementing a Custom Writable: Implementing a Raw Comparator for speed, Custom comparators.	08
5	**Pig:**Hadoop Programming Made Easier Admiring the Pig Architecture, Going with the Pig Latin Application Flow. Working through the ABCs of Pig Latin. Evaluating Local and Distributed Modes of Running Pig Scripts, Checking out the Pig Script Interfaces, Scripting with Pig Latin.	07
6	**Applying Structure to Hadoop Data with Hive:** Saying Hello to Hive, Seeing How the Hive is Put Together, Getting Started with Apache Hive. Examining the Hive Clients. Working with Hive Data Types. Creating and Managing Databases and Tables, Seeing How the Hive Data Manipulation Language Works, Querying and Analyzing Data.	06
	Total	40

8CS4-21: Big Data Analytics Lab

Credit: 2 **Max. Marks: 50(IA:30, ETE:20)**
0L+0T+2P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Implement the following Data structures in Java i) Linked Lists ii) Stacks iii) Queues iv) Set v) Map
2	Perform setting up and Installing Hadoop in its three operating modes: Standalone, Pseudodistributed, Fully distributed.
3	Implement the following file management tasks in Hadoop: • Adding files and directories • Retrieving files • Deleting files Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.
4	Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
5	Write a Map Reduce program that mines weather data. Weather sensors collecting data everyhour at many locations across the globe gather a large volume of log data, which is a goodcandidate for analysis with MapReduce, since it is semi structured and record-oriented.
6	Implement Matrix Multiplication with Hadoop Map Reduce
7	Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.
8	Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
9	Solve some real life big data problems.