IV/IV B. TECH. FIRST SEMESTER
BIG DATA CONCEPTS
(Required)

Course Code: CS 7T1 **Credits: 3**
Lecture:3 period/week **Internal assessment: 30 Marks**
Tutorial: 1period/week **Semester end examination: 70 Marks**

Prerequisites: Data Structures, File Structures, DBMS, DMDW

Course Objectives:

1. Understand the history of Hadoop and the associated computing techniques.
2. Analyze the Weather Dataset with Unix Tools and Hadoop Tools.
3. Analyze the Hadoop Distributed File system.
4. Evaluate Map Reduce Application development and working process.
5. Analyze the types and formats of Map Reduce.
6. Analyze the Features of Map Reduce.

Course Outcomes:
At the end of this course student will:

CO1) Analyze the data with Hadoop framework

CO2) Explain HDFS concepts, interfaces, and basic file system operations

CO3) Understand the fundamentals of i/o in hadoop

CO4) Develop and implement Map reduce applications on hadoop

CO5) Explore Map reduce types and input formats and output formats

Syllabus:

UNIT 1

Introduction to Hadoop: Data, Data types, Storage and Analysis, Relational Database Management System, Grid Computing, Volunteer Computing, A Brief History of Hadoop, Apache Hadoop and the Hadoop Ecosystems.

Map Reduce: A Weather Dataset: Data Format, Analyzing the data with Unix Tools, Analysing the Data with Hadoop: MapReduce, Java MapReduce, Scaling Out: Data Flow, Combiner Function,s Running a Distributed Map Reduce Job,

UNIT 2

The Hadoop Distributed Filesystem: The Design of HDFS, HDFS Concepts, The Command_Line Interface, Hadoop Filesystems, The Java Interface, Data Flow, Data Ingest with Flume and Sqoop, Parallel Copying with distcp and Hadoop Archieves.

UNIT 3

Developing a Map Reduce Application: The Configuration API: Setting up the Development Environment, Writing a Unit Test with MRUnit, Running Locally on Test Data, Running on a cluster, Tuning a Job, Map Reduce Workflows.

UNIT 4

How Map Reduce Works: Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution.

UNIT 5

Map Reduce Types and Formats: Map Reduce Types, Input Format: Input Splits and Records, Text Input, Binary Input, Multiple Inputs, Database Input and Output, Output Formats: Text Output, Binary Output, Multiple Outputs, Lazy Output, Database Output.

Learning Resource

Text Books

Hadoop: The Definitive Guide, Tom White, 3rd Edition (2012), O'Reilly(SPD).

References

Hadoop Essentials: A Quantitative Approach, Henry H. Liu, 1st Edition (2012), PerfMath Publishers.