IV B.Tech II Semester

15ACS83-BIG DATA ANALYTICS

L T P C
3 1 0 3

Course Objectives:

The objectives of this subject are to:

1. *Introduce students the concept and challenge of big data (3 V's: volume, velocity, and variety).*
2. *Teach students in applying skills and tools to manage and analyze the big data*
3. *The students should acquire knowledge on how to design BI solutions for different BI targets and users.*

UNIT-I

Introduction to Big Data. What is Big Data? Why Big Data is Important. Meet Hadoop. Data.Data Storage and Analysis.Comparison with other systems.GridComputing ,Cloud Computing .A brief history of Hadoop.Apachehadoop and the HadoopEcoSystem.Linux refresher; VMWare Installation of Hadoop.

UNIT-II

The design of HDFS.HDFS concepts.Command line interface to HDFS.Hadoop File systems.Interfaces.Java Interface to Hadoop. Anatomy of a file read. Anatomy of a file writes. Replica placement and Coherency Model. Parallel copying with distcp, keeping an HDFS cluster balanced.

UNIT-III

Introduction. Analyzing data with unix tools. Analyzing data with hadoop. Java MapReduce classes (new API). Data flow, combiner functions, Running a distributed MapReduce Job. Configuration API.Setting up the development environment.Managing configuration .Writing a unit test with MRUnit.Running a job in local job runner.Running on a cluster.Launching a job.TheMapReduceWebUI.

UNIT-IV

Classic Mapreduce.Jobsubmission.JobInitialization.Task Assignment. Task execution .Progress and status updates. Job Completion. Shuffle and sort on Map and reducer side. Configuration tuning. Map Reduce Types. Input formats. Output formats, Sorting. Map side and Reduce side joins.

UNIT-V

The Hive Shell. Hive services. Hive clients. The meta store, Hive Architecture And Installation , Comparison with traditional databases. Hive Ql-Querying Data,Sorting And Aggregating . Hbasics.Concepts.Implementation. Java and Map reduce clients. Loading data, web queries, Include Pig tool .
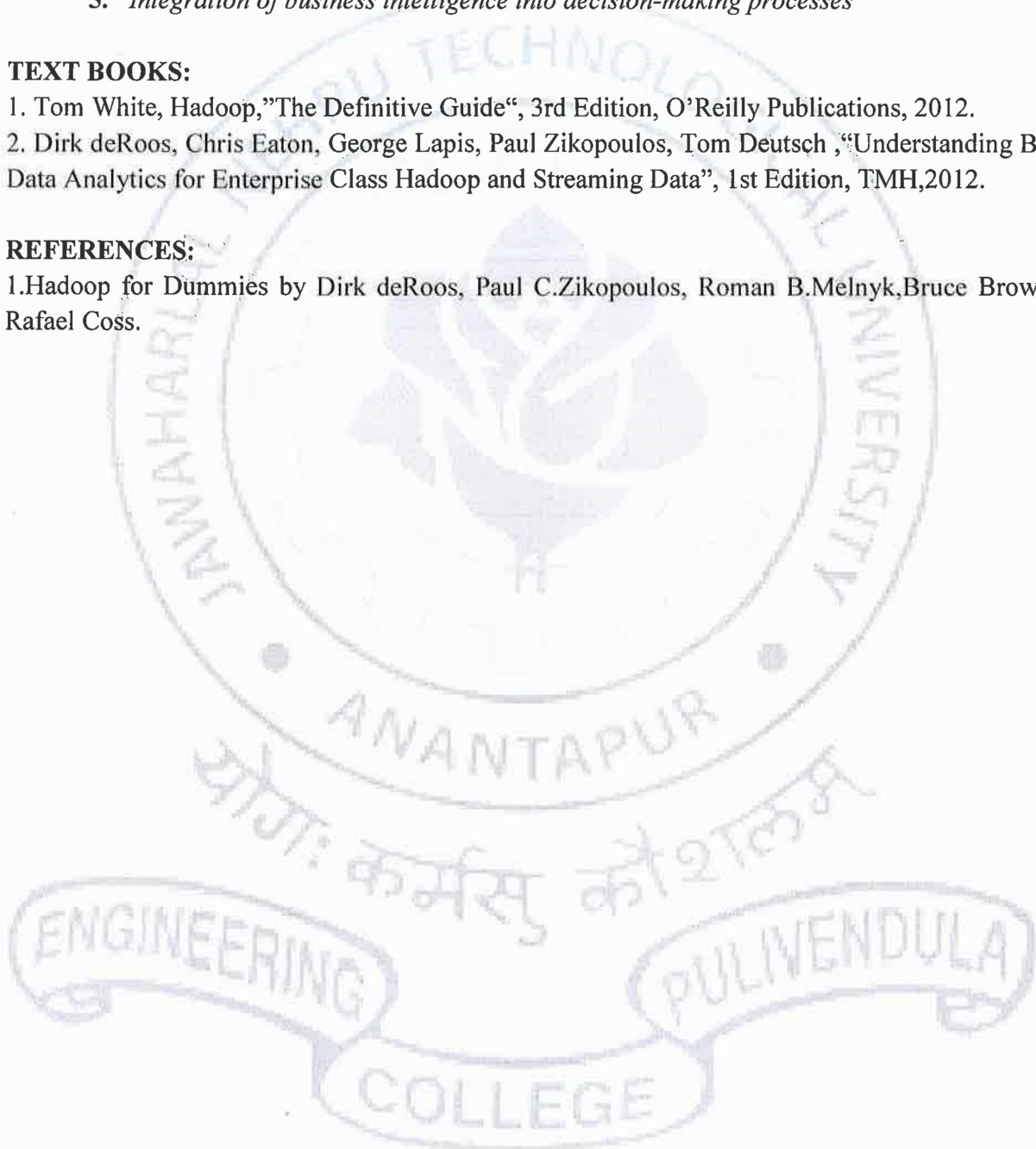
Course Outcomes:
1. *Organizational and individual decision-making*
2. *Key concepts and current practices of business intelligence*
3. *The individual, organizational and societal impacts of BI systems*
4. *Analytical techniques used in business intelligence systems*
5. *Integration of business intelligence into decision-making processes*

TEXT BOOKS:

1. Tom White, Hadoop,"The Definitive Guide", 3rd Edition, O'Reilly Publications, 2012.
2. Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch ,"Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data", 1st Edition, TMH,2012.

REFERENCES:

1.Hadoop for Dummies by Dirk deRoos, Paul C.Zikopoulos, Roman B.Melnyk,Bruce Brown, Rafael Coss.