SCHEME OF COURSE WORK

Department of Information Technology

Course Details:

COURSE TITLE	BIG DATA AND HADOOP		
COURSE CODE	13CT1136	L T P C	4 0 0 3
PROGRAM	B.TECH		
SPECIALIZATION	CSE, IT		
SEMESTER	VIII		
PRE-REQUISITES	DBMS, OOPs through Java		
COURSES TO WHICH IT IS A PRE-REQUISITE	Data Analytics		

Course Outcomes (COs):

1	Explain big data and Apache Hadoop Eco system.
2	Explain design concepts of HDFS (Hadoop Distributed File system).
3	Analyse data using UNIX tools.
4	Apply Hadoop concepts.
5	Discuss Hive Shell.

Course Outcomes (CO) versus **Program Outcomes (PO)**

Course Outcome	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	S	S	S	M					S		S	S
CO2	S	S	M								S	
CO3	S	S	M	S					S			
CO4	S	S	M	S								
CO5	S	S	S	M							M	

S - Strongly correlated, *M* - *Moderately* correlated, *Blank* - *No correlation*

Assessment Methods	Assignment / Quiz / Mid-Test

Teaching- Learning & Evaluation

Week	Topic/ Contents	Course Outcomes	Sample questions	Teaching learning strategy	Assessment method & schedule
1	Introduction to Big Data. What is Big Data. Why Big Data is Important and Meet Hadoop. Data. Data Storage and Analysis. Comparison with other systems, Grid Computing. A brief history of Hadoop. Apache hadoop and the Hadoop EcoSystem. Linux refresher, VMWare Installation of Hadoop.	CO1	1. Define the term Big Data? Explain its Characteristics? 2. Illustrate the various sources of Big Data?	Lecture	Assignment-1, Test- 1 Quiz-1
2	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 write. Replica placement and Coherency Model. Parallel copying with distcp, Keeping an HDFS cluster balanced	CO2	1. Explain HDFS Design and Architecture? 2. Give a detail about HDFS Concepts and the process of Interacting HDFS using command line and Java APIs?	Lecture	Assignment-1, Test- 1 Quiz-1
3	Introduction to Analyzing data with unix tools. Analyzing data with hadoop. Java MapReduce classes (new API). Data flow, combiner functions. Running a distributed MapReduce Job.	CO3	1. Explain various unix tools to analyze data. 2. What are the steps to run a distributed MapReduce job.	Lecture	Assignment-1, Test- 1 Quiz-1
4	Test-1				
Week	**Topic/ Contents**	**Course Outcomes**	**Sample questions**	**Teaching learning**	**Assessment method &**

				strategy	schedule
5	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. The MapReduce WebUI	CO3	1. Explain the steps to configure API. 2. Write a unit test with MR Unit.	Lecture Programming	Assignment-2, Test- 2 Quiz-2
6	Classic Mapreduce. Job submission, Job Initialization, Task Assignment, Task execution. Progress and status updates, Job Completion. Shuffle and sort on Map and reducer side. Configuration Tuning, MapReduce Types. Input formats. Output formats, Sorting. Map side and Reduce side joins.	CO4	1. Illustrate phases in map reduce. 2. Explain Map reduce output formats	Lecture	Assignment-2, Quiz-2, Test-2
7	The Hive Shell. Hive services. Hive clients. The meta store. Comparison with traditional databases. HiveQl. Hbasics. Concepts, Implementation Java and Mapreduce clients. Loading data, web queries	CO5	1. Explain Hive Services. 2. Compare and contrast Relational Databases and NoSQL Databases?	Lecture Programming	Assignment-2, Test- 2, Quiz-2
8	Test-2				