BIG DATA ANALYTICS LABORATORY

VII Semester: IT

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
AITB16	Core	-	-	3	1.5	30	70	100
Contact Classes: Nil	**Tutorial Classes: Nil**	**Practical Classes: 45**				**Total Classes: 45**		

I. COURSE OVERVIEW:

Big data and Business Analytics Laboratory demonstrates distributed computing environment. It includes hands on experience on installation process of VMWare, LINUX commands, HDFS file man- agement, MapReduce functions, Pig and Hive operations. This experience can be used to develop big data applications such as Web click stream analysis, Recommendation systems, Sentiment analysis etc.

II. OBJECTIVES:

The course should enable the students to:

I	The steps involved in creating distributed environment.
II	The platform for creating and run big data Map Reduce programs on Hadoop.
III	Fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.
IV	How to solve complex real-world problems in for decision support.

III. COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	**Demonstrate** distributed environment and its ecosystem with the help of VMWare and Linux commands. .	understand
CO 2	**Make use of** hadoop distributed file management modes forhandling big data in business analytics.	Apply
CO 3	**Analyze** the Big Data using Map-reduce programming in Hadoopframework. Big data in business analytics.	Analyze
CO 4	**Apply** Hive commands for reading, writing and managing largedatasets in hdfs.	Apply
CO 5	**Implement** the Pig Latin scripts in two different modes to perform a particular operation on the data that exists in the HDFS.	Apply
CO 6	**Analyze** adequate perspectives of big data analytics in various applications like recommender systems, social media applicationsetc.	Analyze

IV. SYLLABUS:

LIST OF EXPERIMENTS

Week-1	**INSTALL VMWARE**

Installation of VMWare to setup the Hadoop environment and its ecosystems.

Week-2	**HADOOP MODES**

a. Perform setting up and Installing Hadoop in its three operating modes.
   i. Standalone.
   ii. Pseudo distributed.
   iii. Fully distributed.
b. Use web based tools to monitor your Hadoop setup.

Week-3	**USING LINUX OPERATING SYSTEM**
Implementing the basic commands of LINUX Operating System – File/Directory creation,deletion,update operations.	
Week-4	**FILE MANAGEMENT IN HADOOP**
Implement the following file management tasks in Hadoop: i. Adding files and directories ii. Retrieving files iii. Deleting files Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies theminto HDFS using one of the above command line utilities.	
Week-5	**MAPREDUCE PROGRAM 1**
Run a basic word count Map Reduce program to understand Map Reduce Paradigm.	
Week-6	**MAPREDUCE PROGRAM 2**
Write a Map Reduce program that mines weather data. Hint: Weather sensors collecting data every hour at many locations across the globe gather alarge volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented.	
Week-7	**MAPREDUCE PROGRAM 3**
Implement matrix multiplication with Hadoop Map Reduce.	
Week-8	**MAPREDUCE PROGRAM 4**
Write a Map Reduce program that makes the dataset to be compressed.	
Week-9	**MAPREDUCE PROGRAM 5**
Write a Map Reduce program to run sorting techniques to the relevant data.	
Week-10	**PIG LATIN LANGUAGE - PIG**
Installation of PIG.	
Week-11	**PIG COMMANDS**
Write Pig Latin scripts sort, group, join, project, and filter yourdata.	
Week-12	**PIG LATIN MODES**
Implement the Pig Latin scripts in two different modes: Local mode and HDFS mode and run the different scripts and UDF's.	
Week-13	**PIG PROGRAM**
Run the Pig Latin Scripts to find a max temp for each and every year.	
Week-14	**HIVE**
Installation of HIVE.	
Week-15	**HIVE OPERATIONS**
Use Hive to create, alter, and drop databases, tables, views, functions, and indexes.	

Reference Books:
1. Jay Liebowitz, “Big Data And Business Analytics Laboratory”, CRC Press.
Web References:
1. Hadoop : http://hadoop.apache.org/ 2. Hive: https://cwiki.apache.org/confluence/display/Hive/Home 3. Pig latin: http://pig.apache.org/docs/r0.7.0/tutorial.html
SOFTWARE AND HARDWARE REQUIREMENTS FOR 36 STUDENTS: **HARDWARE:** Desktop Computers with 4 GB RAM36 nos. **SOFTWARE:** VMWare, HADOOP.