COURSE CONTENT

BIG DATA AND BUSINESS ANALYTICS LABORATORY							

VII SEMESTER: IT

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
AITC29	Core	0	0	3	1.5	30	70	100

Contact Classes: Nil	Tutorial Classes: Nil	Practical Classes: 36	Total Classes: 36

Prerequisites: Database Management Systems

I. COURSE OVERVIEW:

Fundamental principles of Big Data Analytics and its role in making better decisions and predictions in the organization. The course also covers the Technology, Infrastructure and Applications of Big Data. Concepts of data identification, data cleansing and integration are also addressed. Software requirements of Big Data are addressed and case studies of Big Data Applications are discussed

II. COURSE OBJECTIVES:

The students will try to learn:
  I.   The implications and challenges of Big Data Analytics.
  II. The technologies used to store, manage, and analyze big data in a Hadoop ecosystem.
  III. The hypothesis on the optimized business decisions in solving complex real-world problems.

III. COURSE SYLLABUS:

Expt. 1: INSTALL VMWARE
Installation of VMWare to setup the Hadoop environment and its ecosystems

Expt. 2: HADOOP MODES
Perform setting up and Installing Hadoop in its three operating modes.
i.   Standalone.
ii. Pseudo distributed.
iii. Fully distributed.

Expt. 3: USING LINUX OPERATING SYSTEM
Implementing the basic commands of LINUX Operating System –File/Directory creation, deletion, update operations.
a. Create a directory in HDFS at given path(s).
b. List the contents of a directory.
c. Upload and download a file in HDFS.
d. See contents of a file
e. Copy a file from source to destination
f. Copy a file from/To Local file system to HDFS
g. Move file from source to destination.
h. Remove a file or directory in HDFS.

Expt. 4: FILE MANAGEMENT IN HADOOP
Implement the following file management tasks in Hadoop:
a. Copy a file from/To Local file system to HDFS
b. Move file from source to destination.
c. Remove a file or directory in HDFS.
d. Display the aggregate length of a file.

Expt. 5: MAPREDUCE PROGRAM 1
Run a basic wordcount MapReduce program to understand MapReduce Paradigm.

Expt. 6: **MAPREDUCE PROGRAM 2**
Write a MapReduce program that mines weather data. Hint: Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with MapReduce, since it is semistructured and record-oriented

Expt. 7: **MAPREDUCE PROGRAM 3**
Implement matrix multiplication with Hadoop MapReduce.

Expt. 8: PIG LATIN LANGUAGE - PIG
Installation of PIG.

Expt. 9: **PIG COMMANDS**
Write Pig Latin scripts sort, group, join, project, and filter your data.

Expt. 10: **PIG LATIN MODES, PROGRAMS**
   a. Run the PigLatin Scripts to find WordCount
   b. Run the PigLatin Scripts to find a max temp for each and everyyear.

Expt. 11: **HIVE**
Installation of HIVE.

Expt. 12: **HIVE OPERATIONS**
Use Hive to create, alter, and drop databases, tables, views, functions, and indexes.

IV. REFERENCE BOOKS:
1. Jay Liebowitz, "Big Data and Business Analytics Laboratory", CRC Press.
2. Black Book Big Data, Dreamtech publications, 1st Edition, 2017
3. Tom White, "Hadoop: The Definitive Guide, O'Reilly, 3rd Edition, 2012

V. WEB REFERENCES:
   1. https://www.iiitnr.ac.in/
   2. https://ibse.iitm.ac.in/
   3. https://www.bits-pilani.ac.in/