JNTUA COLLEGE OF ENGINEERING (AUTONOMUS): PULIVENDULA
Department of Computer Science & Engineering
B.Tech IV Year II Semester

Subject: Big Data Analytics

Lesson plan

I.

Course Title	:	**Big data Analytics**			
Course Structure	:	Lectures	Tutorials	Practical's	Credits
		4	1	0	4
Course Coordinator	:	Mr.K.Balachandra Reddy, Asst.prof(Adhoc)			
Team of Instructors	:	Mr. G.Murali (HOD)			

Course Overview

Learning and designing about the big data and also hadoop using map reduce programs and also done by eco-systems

II. Prerequisite(s):

Level	Credits	Periods / Week	Prerequisites
UG	3	4	Data Mining and Data Warehousing

III. Assessment:

FORMATIVE ASSESMENT	
Mid Semester Test I (Theory) for 20 Marks in first two units is conducted at the end of 9th week. Mid Semester Test II (Theory) for 20 Marks in last three units is conducted at the end of the course work.	20 Marks
Multiple Choice Test I for 10 Marks in first two and halfunits is conducted at along with Theory exam. Multiple Choice Test II for 10 Marks in last two and halfunits is conducted at along with Theory exam. Note: After evaluating these tests 30 marks are calculated as 80% from best marks and 20% from other.	10 Marks
Total (Formative)	30 Marks
SUMMATIVE ASSESMENT	
End Semester Examination in all units is conducted for 70 Marks	70 marks
Grand Total	100 Marks

IV. Course objectives:

At the end of the course, participants should be able to:

1. Master the concepts of HDFS and Map Reduce framework.
2. Understand Hadoop Architecture.
3. Setup Hadoop Cluster and write Complex Map Reduce programs.
4. Perform Data Analytics using Hive.
5. Implement HBase and Map Reduce Integration.
6. Implement best Practices for Hadoop Development.
7. They will understood about R analytics Based on big data.

V. Course Outcomes:

- Student will be able to know where the Big data is used and its importance.
- Students will be able to know how the Big data will be handled and its problems.
- Students will be able to learn the importance of Hadoop.
- Students will be able to know about Doug Cutting and how the Hadoop came into existence.
- Student will be able to know about HDFS, MapReduce and Hadoop releases.
- Students will be able to know how to write a program in Hadoop
- Students will be able to know how Map and Reduce done in Hadoop
- Students will be able to know how to view information about jobs in web browser

VI. Program outcomes:

a. An ability to apply knowledge of Big data, literature survey, module specifications, and computer science and engineering theory in the modeling and design of computer-based systems to real-world problems (fundamental engineering analysis skills).

b. An ability to design and conduct experiments, as well as to analyze and interpret the software information (information retrieval skills).

c. An ability to design , implement, and evaluate a computer-based system, process, component, module or program

to meet desired needs, within realistic constraints such as economic, environmental, social, political, health and safety, manufacturability, and sustainability (Creative Skills) requirements.

d. An ability to function effectively on multi-disciplinary teams (team work).

e. An ability to analyze a problem, identify, formulate and use the appropriate computing and engineering requirements for obtaining its solution (engineering problem solving skills).

f. An understanding of professional, ethical, legal, security and social issues and responsibilities (professional integrity)

g. An ability to communicate effectively both in writing and orally (speaking / writing skills) with customers (stakeholders).

h. The broad education necessary to analyze the local and global impact of computing and engineering solutions on individuals, organizations, and society (engineering impact assessment skills).

i. Recognition of the need for, and an ability to engage in continuing professional development and life-long learning (continuing education awareness).

j. A Knowledge of contemporary issues (social awareness).

k. An ability to use current techniques, skills, and tools necessary for computing and engineering practice (practical engineering analysis skills).

VII. Syllabus:

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. Grid Computing. A brief history of Hadoop. Apache hadoop and the Hadoop EcoSystem. 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. The MapReduce WebUI.

UNIT-IV : 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. 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. Comparison with traditional databases. Hive Ql. Hbasics. Concepts. Implementation. Java and Map reduce clients. Loading data, web queries.

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

IX. Course Plan:

The course plan is meant as a guideline. There may probably be changes.

Lecture No	Topic	Course Outcomes	Reference
2/12/2019	**Introduction to Big Data:-** What is big data Why Big Data is important	Student will be able to know where the Big data is used and its importance.	T1:2.1, T2:.2.3,R1:1.9
4/12/2019	**Meet Hadoop**:- Data Data storage and Analysis	Students will be able to know how the Big data will be handled and its problems.	T1:2.3,T2:2.8
5/12/2019	Comparison with Other systems Grid computing	Students will be able to learn the importance of Hadoop.	T1:2.5,T2:1.25
6/12/2019	A Brief History of Hadoop	Students will be able to know about Doug Cutting and how the Hadoop came into existence.	T1:3.1,T1:3.2,3.4, R1:2.6
9/12/2019	Apache Hadoop and the Hadoop Ecosystem	Student will be able to know about HDFS, MapReduce and Hadoop releases.	T1:3.3,3.4, R1:3.2, R1:4.13
11/12/2019	VM Ware Installation of Hadoop	Students will be able to install Hadoop VM Ware.	T1:3.6,T2:5.5,5.6
12/12/2019	**The Hadoop Distributed File System :-** The Design of HDFS	Students will be able to know about HDFS .	T1:4.6,T2:5.7
13/12/2019	HDFS Concepts	Students will be able to know about Hadoop and HDFS architecture.	T2:6.5, W2.1, R1:2.5, R1:7.4
16/12/2019	The Command –Line Interface to HDFS	Students will be able to know about the commands used to interact with HDFS	T1:8.6 T2:8.7
18/12/2019	Hadoop File Systems	Students will be able to know about file systems and java implementations in Hadoop	T2:10.2, R1: 4.5, 4.6,
19/12/2019	Interfaces	Students will be able to learn how the other file	T2:10.3,10.4, R1:6.4, 6.5, 6.6, R1:4.7, 4.9

		systems will interact	
20/12/2019	Java Interface to Hadoop	Students will be able to know how to interface Hadoop by using Java	T2: 10.5,10.6, R1:6.7.1, R1:4.11,4.12,4.13
23/12/2019	Anatomy of a file	Students will be able to know how client read data from HDFS.	T1:9.5, T2:4.2, R1:7.1.1,
26/12/2019	Anatomy of file write	Students will be able to know how client write data to HDFS	T2:4.3, R1:7.3
27/12/2019	Replica Placement and Coherency Model	Students will be able to know how data will be replicated in data center in racks.	T1:7.9,T2:4.4
30/12/2019	Parallel copying with distcp	Students will be able to know about distcp program which is used to copy large amount of data to and from HDFS.	T2:4.5,4.6
1/01/2020	Keeping an HDFS Cluster Balanced	Students will be able to know how HDFS clusters will be balanced	T2:7.1, R1:7.4, R1:5.4
3/01/2020	**MapReduce :-** Analyzing data with Unix tools	Students will be able to know how to analyze data with Unix	T2:7.2
6/01/2020	Analyzing data with Hadoop	Students will be able to know how to analyze the data with Hadoop	T2:7.3,7.4, R1:7.5, R1:5.5
8/01/2020	Java MapReduce classes	Students will be able to know how to write a program in Hadoop	T2:12.1,12.2, R1:5.2, 5.3, 5.4, R1:5.3
9/01/2020	Data Flow	Students will be able to know how data flow in data centers	T2:12.3,12.4, R1:5.9, R1: 6.7,7
10/01/2020	Combiner Functions, Running a distributed MapReduce job	Students will be able to know how Map and Reduce done in Hadoop	T2:12.5,12.6, w4.1, w4.2, R1: 5.5, 5.6, 5.7, R1:18.4,
19/02/2020	**Developing a MapReduce Application**:- Configuration API	Students will be able to know writing a MapReduce program	T2:11.2, R1:10.1,10.2, R1:11, R6:8
20/02/2020	Setting up the development environment, Managing configuration	Students will be able to know in which environment program runs.	T2:11.3,11.4,R1:10.3
24/02/2020	Writing a unit test with MRUnit, Running a job in local job runner, Running on a cluster	Students will be able to know how to write a program and execution of program	T1:11.5,11.6, R1:4.4
26/02/2020	Launching a job	Students will be able to know about driver used to	T1:9.1,9.2, R1:12.1,8.1.2,

		launch a job	R1:12.5, R1:2.23, R1:10.6
27/02/2020	The MapReduce WebUI	Students will be able to know how to view information about jobs in web browser	T1:9.3,9.4, R`1:8.2,8.3,8.4
28/02/2020	**How MapReduce Works:-** **Classic MapReduce** Job submission, Job initialization, Task assignment, Task execution, Progress and status updates	Students will be able to know how to write advanced MapReduce programs.	T1:13.1,13.2, R1:12.2, 12.4, 12.7, R1:13.4
02/03/2020	Shuffle and Sort on Map and Reduce side	Students will be able to know how to Shuffle and sorting procedure on map and Reduce side.	T1:13.3,R1:13.8
04/03/2020	Configuration Tuning	Students will be able to know how to tune shuffle on Map and Reduce side	T2:13.4,R1:14.7
05/03/2020	**MapReduce Types and Formats**:- MapReduce types	Students will be able to learn MapReduce Models in detail	T1:4.1, T1:4.2
06/03/2020	Input and Output formats	Students will be able to learn how to process different types of data formats by using Hadoop.	T1:4.3,T2:11.6
09/03/2020	**MapReduce Features**:- Sorting Map and Reduce side	Students will be able to know about different Sorting-Partial sort, Total sort, Secondary sort.	T1:4.4,T2:6.8
11/03/2020	Joins	Students will be able to know about how to perform Joins on Map and Reduce side.	T1:11.1,T1:3.2,3.4, R1:2.6
12/03/2020	**Hive**:- The Hive Shell	Students will be able to know how Hive works.	T1:3.3,3.4, R2:10.2, R1:8.13
13/03/2020	Running Hive ,Hive services, Metastore	Students will be able to know how to run Hive along with commands used to run Hive.	T2:5.5,5.6
16/03/2020	Comparison with	Students will be able to	T2:5.7

	traditional databases, Hive Ql	know importance of Hive	
18/03/2020	**HBase:-** HBasics concepts	Students will be able to know how HBase is used when real –time read/write random access performed on large datasets.	T2:8.5, W2.1, R1:2.5, R1:7.4
19/03/2020	Implementation	Students will be able to know how to implement HBase.	T2:8.6 T2:8.7
20/03/2020	Clients:- Java, MapReduce	Students will be able to know different types of clients, interacting with HBase.	T2:10.2, R1: 4.5, 4.6,
23/03/2020	Loading Data, Web Queries	Students will be able to know how to load data in HBase.	T2:10.3,10.4, R1:6.4, 6.5, 6.6, R1:4.7, 4.9

Mapping course outcomes leading to the achievement of the program outcomes

Course Outcomes	**Program Outcomes**												
	A	**b**	**c**	**D**	**E**	**f**	**g**	**h**	**i**	**J**	**k**	**l**	**m**
1	S	H											
2			H		S								
3			H										
4				H									S
5					H								S
6					H								
7			H										
8											H		

S = Supportive **H = Highly Related**

Justification of CO's –PO's Mapping Table:

By mapping CO-1 to the PO's A & B which are related to the course CO1: The student is able to analyze and Implement Problems

By mapping CO-2 to the PO's C & E, which are related to the course CO2: The student is able to analyze the problem and solutions using various software architecture analyzing approaches.

By mapping CO-3 to the PO's C which are related to the course CO3: The student is able to understand the purpose of different software engineering project teams.

By mapping CO-4 to the PO's D & M which are related to the course CO4: The student is able to understand the Purpose of different software engineering architecture plans.

By mapping CO-5 to the PO's E & M which are related to the course CO5: The student is able to understand the Purpose of different architecture description languages.

By mapping CO-6 to the PO's E which are related to the course CO6: The student is able to understand the concept of architecture implementation and various goals of analysis.

By mapping CO-7 to the PO's C which are related to the course CO7: The student is able to different conceptual and technical skills in the analysis and design.

By mapping CO-8 to the PO's K which are related to the course CO8: The student is able to understand the software engineering CASE tools.