COURSE HANDOUT

B. Tech (CSE) – 8th Semester

Course Title : **BIGDATA ANALYTICS** Dated: 27-11-2018

Course Code : CSE 4424 Academic Year 2017-18

Course Structure : 3-1-0-4

Course coordinator : Mr. M.Balajee

Instructor(s) : Dr S S Gantayat

Course Objective:
The course content enables students to:

a. To introduce the fundamental concepts of BIG Data
b. To introduce various analytical techniques to crunch massive data
c. To have a knowhow about applications which uses Big Data
d. To know the Architectural components to handle Big Data.
e. To have a model to handle massive data using Hadoop Map Reduce.

Course Outcomes:
At the end of the course students are able to:

i. Identify the need for big data analytics for a domain.
ii. Apply big data analytics for a given problem.
iii. Suggest areas to apply big data to increase business outcome.
iv. Use Hadoop, Map Reduce Framework handle massive data

Text Books:

1. Big Data Analytics: Disruptive Technologies for Changing the Game, Dr. Arvind Sathi, MC Press online.
2. Hadoop: The Definitive Guide, Tom White, O'Reilly Media / Yahoo Press, 2012

Reference Books:

1. Frank J. Ohlhorst,Big Data Analytics: Turning Big Data into Big Money, Wiley and SAS Business Series, 2012.

2. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", Wiley and SAS Business Series, 2012.

3. Paul Zikopoulos, Chris Eaton, Paul Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGraw Hill, 2011.

Syllabus

UNIT I 10+3 Hrs

Introduction to Big Data: Analytics – Nuances of big data – Value – Issues – Case for Big data – Big data options Team challenge – Big data sources – Acquisition – Nuts and Bolts of Big data. Features of Big Data -Security, Compliance, auditing and protection - Evolution of Big data – Best Practices for Big data Analytics - Big data characteristics - Volume, Veracity, Velocity, Variety.

UNIT II 11+4 Hrs

Applications of Big Data & Data Analysis: Drivers for big data – Automation – Monetization- Applications of Big Data.- Social Media Command Center-Product knowledge hub-infrastructure and knowledge hub-Product selection, Design and Engineering- Location Based services- Online Advertizing- Improved Risk management. Analytic data sets – Analytic methods –analytic tools – Cognos – Micro strategy - Pentaho.

UNIT-III 11+4 Hrs

Architectural components: Massively Parallel Processing Platforms (MPP) - Unstructured data analytics and reporting-Context sensitive and domain specific searches- categories and ontology-focus on specific time slice-big data and single view of customer-Data privacy protection- Real time adaptive analytics and decision engine.

UNIT IV 13+4 Hrs

Hadoop Framework: Big data implementation-Revolutionary, Evolutionary and Hybrid Approaches- Overview of HadoopRDBMS (vs) HADOOP- IBM for Big Data – Map Reduce Framework and Architecture. Hadoop Distributed file systems –Features of HDFS- Developing Map reduce – Analyzing big data with twitter.

Course Plan:

No. Lecture	Learning objectives	Topic(s) to be covered	Chapter in the textbook/reference
	Unit-I		
1	Overview of Big data analytics	Introduction to BDA	…
2	Learn big data basics	Analytics – Nuances of big data – Value	R1 ch:1
3	Understand bigdata case studies	Case for Big data	R1 ch:3
4	Learn data sources in big data	Big data sources	R1 ch:3
5	Understand data acquisition methodologies	Big data Acquisition	R1 ch:5
6	Learn nuts and bolts of big data	Nuts and Bolts of Big data.	R1 ch:6
7	Understand features of big data	Features of Big Data	R1 ch:7
8	Learn security issues in big data	Security	R1 ch:7
9	Learn big data compliances	Big Data Compliance	R1 ch:7
10	Understand the protection mechanisms	auditing and protection	R1 ch:8
11	Learn about challenges in big data	Big data options Team challenge	R1 ch:3
	Unit-II		
12	Understand the applications of big data	Applications of Big Data & Data	T1 Ch:2.1

		Analysis: Drivers for big data	
13	Understand the applications of big data	Automation – Monetization	T1 Ch:2.2
14	Understand the applications of big data	Applications of Big Data.	T1 Ch:3
15	Learn about social media center	Social Media Command Center	T1 Ch:3
16	Understand about knowledge hub	infrastructure and knowledge hub	T1 Ch:3
17	Learn designing of big data	Product selection, Design and Engineering	T1 Ch:3
18	Understand location based services	Location Based services	T1 Ch:3
19	Learn about advertizing in online	Online Advertizing	T1 Ch:3
20	Understand about risk management	Improved Risk management.	T1 Ch:3
21	Learn about analytic tools	analytic tools – Cognos – Micro strategy	T1 Ch:3
22	Understand about pentaho	Analytic methods Pentaho.	T1 Ch:3
	Unit-III		
23	Understand the architectural components in big data	Architectural components	T1 Ch:4.1
24	Learn about MPP	Massively Parallel Processing Platforms (MPP)	T1 Ch:4.1
25	Learn about unstructured data	Unstructured data analytics and reporting	T1 Ch:4.2
26	Learn about unstructured data	Unstructured data analytics and reporting	T1 Ch:4.2
27	Understand the domain specific searches	Context sensitive and domain specific searches	T1 Ch:4.2
28	Learn the ontological perspective on big data	categories and ontology-focus on specific time slice	T1 Ch:4.2
29	Learn the ontological perspective on big data	categories and ontology-focus on specific time slice	T1 Ch:4.2
30	Understand the customer perspective in big data	big data and single view of customer	T1 Ch:4.3
31	Understand the data privacy in big data	Data privacy protection	T1 Ch:4.4
32	Understand the data privacy in big data	Data privacy protection	T1 Ch:4.4
33	Learn about decision engine	Real time adaptive analytics and decision engine.	T1 Ch:4.5
34	Learn about decision engine	Real time adaptive analytics and decision engine.	T1 Ch:4.5
	Unit-IV		
35	Understand Hadoop framework	Hadoop Framework	T1 Ch:6
36	Understand Hadoop framework	Hadoop Framework	T1 Ch:6
37	Learn big data implementation	Big data implementation	T1 Ch:6
38	Understand about evolutionary and hybrid approaches	Revolutionary, Evolutionary and Hybrid Approaches	T1 Ch:6
39	Understand about evolutionary and hybrid approaches	Revolutionary, Evolutionary and Hybrid Approaches	T1 Ch:6

40	Understand Hadoop RDBMS	Overview of Hadoop RDBMS (vs) HADOOP	T2 ch:1
41	Understand Hadoop RDBMS	Overview of Hadoop RDBMS (vs) HADOOP	T2 ch:1
42	Understand map reduce technique	IBM for Big Data – Map Reduce Framework and Architecture	T2 ch:1
43	Learn about distributed file systems	Hadoop Distributed file systems	T2 ch:3
44	Learn about distributed file systems	Hadoop Distributed file systems	T2 ch:3
45	Understand Features of HDFS, Big Data vs. Twitter	Features of HDFS- Developing Map reduce , Analyzing big data with twitter	T2 ch:3

Tutorials:

Tutorial No.	Topic(s) to be covered
	Unit-I
1	Tutorial-1
2	Tutorial-2
3	Tutorial-3
	Unit-II
4	Tutorial-4
5	Tutorial-5
6	Tutorial-6
7	Tutorial-7
	Unit-III
8	Tutorial-8
9	Tutorial-9
10	Tutorial-10
11	Tutorial-11
	Unit-IV
12	Tutorial-12
13	Tutorial-13
14	Tutorial-14
15	Tutorial-15

Evaluation Scheme:

Component	Duration (minutes)	Marks	% of weightage	Date & Time	Venue
Sessional Test – 1	90	20	20%	01.01.2018- 06.01.2018 3:15-4:45 P.M	Block-5
Sessional Test – 2	90	20		12.02.2018-17.02.2018 3:15-4:45 P.M	Block-5
Sessional Test – 3	90	20		26.03.2018-31.03.2018 9:00-10:30 A.M	Block-5
Comprehensive quiz examination	20	10	10%	26.03.2018-31.03.2018	CA Lab.
External Examination	**180**	**70**	**70%**	09.04.2018-21.04.2018	Block-5

Chamber Consultation Hour: 4.00PM

Venue: CSE Staff Room

Notices: CSE Main notice board

Signature of the Instructor **Signature of the course-coordinator**