



Department of Information Technology	LP: IT16703
B.E/B.Tech/M.E/M.Tech : Information Technology Regulation:2016	Rev. No: 01
PG Specialisation : NA	Date: 22/06/2020
Sub. Code / Sub. Name : IT16703 – BIG DATA ANALYTICS	
Unit : I	

Unit Syllabus:

UNIT I INTRODUCTION TO BIG DATA

9

Introduction to Big Data Analytics – Challenges and limitations of big data analytics- Conventional Systems - Nature of Data, Evolution of Analytic Scalability - Intelligent data analysis- Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Re-Sampling - Statistical Inference - Prediction Error.

Objective:

Session No *	Topics to be covered	Ref	Teaching Aids
1.	Introduction to Big Data Platform	R1- Ch.1(Pg.3-25)	PPT
2.	Challenges of conventional Systems	R1- Ch.2 (Pg.29-50)	PPT
3.	Nature of Data	R1- Ch.4 (Pg.87-117)	PPT
4.	Evolution of Analytic scalability	R1- Ch.5 (Pg.121-151)	PPT
5.	Analytic processes and tools	R1- Ch.6 (Pg.153-175)	PPT
6.	Analysis vs reporting, Modern data analytic tools	R1- Ch.7 (Pg.179-184)	PPT
7.	Statistical concepts: Sampling distributions	T1- Ch.1 (Pg.12-14)	PPT
8.	Re-Sampling, Statistical inference	T1- Ch.2 (Pg.17-32)	PPT
9.	Prediction error.	T1- Ch.2 (Pg.33-68)	PPT

Content beyond syllabus covered (if any):

* Session duration: 50 minutes



Sub. Code / Sub. Name: IT16703 – BIG DATA ANALYTICS
Unit : II

UNIT II MINING DATA STREAMS**9**

Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications - Case Studies - Real Time Sentiment Analysis, Stock Market Predictions.

Objective:

	Topics to be covered	Ref	Teaching Aids
10.	Introduction to Streams Concepts - Stream data model and architecture	T2- Ch.4 (Pg.108-112)	PPT
11.	Stream Computing, Sampling data in a stream	T2- Ch.4 (Pg.112-115)	PPT
12.	Filtering streams	T2- Ch.4 (Pg.115-118)	PPT
13.	Counting distinct elements in a stream -Estimating moments	T2- Ch.4 (Pg.118-122)	PPT
14.	Counting oneness in a window	T2- Ch.4 (Pg.122-133)	PPT
15.	Decaying window - Real time Analytics Platform(RTAP) applications	T2- Ch.4 (Pg.133-136)	PPT
16.	Case studies - RTAP applications - Smart Traffic	Internet	PPT
17.	Real time sentiment analysis	Internet	PPT
18.	Stock market predictions	Internet	PPT

Content beyond syllabus covered (if any):



Sub. Code / Sub. Name: IT16703 – BIG DATA ANALYTICS
Unit : III

Unit Syllabus :

UNIT III INTRODUCTION TO BIG DATA ANALYTICS & R PROGRAMMING

9

Analyzing, Visualization and Exploring the Data, Statistics for Model Building and Evaluation, Introduction to R and RStudio, Basic analysis in R, Intermediate R, Intermediate analysis in R, Advanced Analytics - K-means clustering, Association rules-Speedup, Linear Regression, Logistic Regression, Naïve Bayes, Decision Trees, Time Series Analysis, Text Analysis.

Objective:

Session No *	Topics to be covered	Ref	Teaching Aids
19.	Analyzing, Visualization & Exploring the Data	Internet	PPT
20.	Statistics for Model Building & Evaluation	Internet	PPT
21.	Introduction to R and RStudio, Basic Analysis in R	Internet	PPT
22.	Intermediate in R, Intermediate analysis in R	Internet	PPT
23.	Advanced Analytics – K-means clustering	Internet	PPT
24.	Association rules-Speedup	Internet	PPT
25.	Linear Regression, Logistic Regression	http://faculty.cas.usf.edu/mbrannick/regression/Part3/Reg2.html	PPT
26.	Naïve Bayes, Decision Trees	Internet	PPT
27.	Time Series Analysis & Text Analysis	Internet	PPT

Content beyond syllabus covered (if any):

Random Forest Algorithm , KNN algorithm



Sub. Code / Sub. Name: IT16703 – BIG DATA ANALYTICS
Unit : IV

Unit Syllabus :**UNIT IV HADOOP**

9

History of Hadoop- The Hadoop Distributed File System – Components of Hadoop, Analyzing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce Features.

Objective:

Session No *	Topics to be covered	Ref	Teaching Aids
28.	History of Hadoop-The Hadoop Distributed File System	R7-Ch.1; Pg.No:1-13	PPT
29.	Components of Hadoop	R7-Ch.4; Pg.No:83-144	PPT
30.	Analyzing the Data with Hadoop-Scaling-Hadoop Streaming	R7-Ch.2; Pg.No:17-44	PPT
31.	Design of HDFS-Java Interfaces to HDFS Basics	R7-Ch.3; Pg.No:45-82	PPT
32.	Developing a Map Reduce Application-How Map Reduce works	R7-Ch.5; Pg.No:145-186	PPT
33.	Anatomy of a Map Reduce Job run-Failures	R7-Ch.6; Pg.No:187-203	PPT
34.	Job Scheduling-Shuffle and Sort	R7-Ch.6; Pg.No:204-212	PPT
35.	Task execution	R7-Ch.6; Pg.No:212-220	PPT
36.	Map Reduce Types and Formats- Map Reduce Features.	R7-Ch.7,8; Pg.No:221-294	PPT

Content beyond syllabus covered (if any):

* Session duration: 50 mins



Sub. Code / Sub. Name: IT16703 – BIG DATA ANALYTICS

Unit : V

Unit Syllabus :

UNIT V FRAMEWORKS

9

Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive - fundamentals of HBase and ZooKeeper - IBM InfoSphere BigInsights and Streams. Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications.

Objective:

Session No *	Topics to be covered	Ref	Teaching Aids
37.	Applications on Big Data Using Pig and Hive	https://www.tutorials-point.com/hive/hive_i	PPT
38.	Data processing operators in Pig	R7-Ch.11; Pg.No:397-407	PPT
39.	Hive services	R7-Ch.12; Pg.No:417-419	PPT
40.	HiveQL – Querying Data in Hive	R7-Ch.12; Pg.No:422-427	PPT
41.	Fundamentals of HBase and ZooKeeper	R7-Ch.13; Pg.No:457-458	PPT
42.	IBM InfoSphere BigInsights and Streams	Internet	PPT
43.	Visualizations - Visual data analysis techniques	Internet	PPT
44.	Visualizations- interaction techniques	Internet	PPT
45.	Visualizations - Systems and applications	Internet	PPT

Content beyond syllabus covered (if any):

* Session duration: 50 mins



Sub Code / Sub Name: IT16703 – BIG DATA ANALYTICS

Course Outcome 1:

Ability to identify the characteristics of datasets and compare the trivial data and big data for various applications.

Course Outcome 2:

Interpret business models and scientific computing paradigms, and apply software tools for big data analytics.

Course Outcome 3:

Ability to understand and apply scaling up machine learning techniques and associated computing techniques and technologies.

Course Outcome 4:

Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies like hadoop and mapreduce.

Course Outcome 5:

Ability to understand of how to manage Big Data

Text Books:

1. Vignesh Prajapati, —Big Data Analytics with R and Hadoopl, Packt Publishing, 2014.
2. Stephan Kudyba, —Big Data, Mining, and Analytics: Components of Strategic Decision Makingl, First Edition, CRC Press, 2014.

References:

1. Michael Minelli, Michele Chambers, Ambiga Dhiraj,|Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businessesl, Wiley Publications, 2013.
2. Dr. Mark Gardener, —Beginning R: The Statistical Programming Language| (Wrox), 2013
3. Anand Rajaraman and Jeffrey David Ullman,—Mining of Massive Datasetsl, Cambridge University Press, 2012.
4. Bill Franks,|Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analyticsl, John Wiley & sons, 2012.
5. Pete Warden, —Big Data Glossaryl, O_Reilly, 2011.
6. Zikopoulos, Paul, Chris Eaton, —Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Datal, Tata McGraw Hill Publications, 2011.
7. Tom White,|Hadoop: The Definitive Guidel, Third Edition, O_reilly Media, 2012.
8. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, —Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Datal, McGrawHill Publishing, 2012.

	Prepared by	Approved by
Signature		
Name	Ms.A.Indumathi, Ms.P.Meenakshi	Dr.V.Vidhya
Designation	Assistant Professor (IT)	Professor & HOD/IT i/c
Date	22/06/2020	22/06/2020
Remarks *:		

* If the same lesson plan is followed in the subsequent semester/year it should be mentioned and signed by the Faculty and the HOD