III Year Sem II B.Tech	ELECTIVE	L	T	P	C
Code: CS3602	Data Sciences&Big Data Analytics	2	2	0	4

Outcomes

1. Understand big data challenges in different domains including social media, transportation, finance and medicine
2. Analyze scalability and performance of relational model, SQL and emergent systems.
3. Comprehend machine learning and algorithms for data analytics.
4. Build secure big data systems
5. Analyze Map-Reduce programming model for better optimization

Detailed syllabus:

UNIT-I

Overview of Big Data, Stages of analytical evolution, State of the Practice in Analytics, The Data Scientist, Big Data Analytics in Industry Verticals, Data Analytics Lifecycle.

UNIT-II

Regression analysis, Inferential Statistics through hypothesis tests, ANOVA(Analysis of Variance), the t-test and associated confidence interval.

UNIT-III

Linear Algebra-Basics, Dimensionality, Vector Spaces, Dimensionality Reduction using R, Linear and Generalized linear Model in R.

UNIT-IV

The Map task of key-value pairs: Grouping by Key, The Reduce Tasks, Combiners, Coping With Node Failures, Example of Map Reduce: word count of words of documents.

Matrix-Vector Multiplication by Map Reduce , Relational algebra operations by Map Reduce , Matrix-Matrix Multiplication by Map Reduce

UNIT-V

Hadoop, SPARK Ecosystem

UNIT-VI

No SQL Databases, BASE: Introduction and Properties, CAPE Theorem, Mango DB, Cassendra, Graph Data Handling using Neo4j.

Text Books:

1. Bill Franks, Taming the Big Data Tidal Wave, 1st Edition, Wiley, 2012.

3. Anand Rajaraman, Jeffrey D. Ullman, Mining of Massive Datasets Jure Leskovec Stanford Univ.

Reference Books:

1. Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: Springer, 2009

2. Vignesh Prajapati , Big Data Analytics with R and Hadoop , PACKT publishing 2013.

3. Frank J. Ohlhorst, Big Data Analytics,1st Edition, Wiley, 2012.

Video Reference links:

Title	Expert Name	Details of Expert	Web link
1. Data Analytics 2. Big Data and Hadoop	Prof Nandan sudarsanam	Asst prof. in IITM	1. http://nptel.ac.in/courses/110106064/ 2. http://freevideolectures.com/Course/3613/Big-Data-and-Hadoop

Text Reference links:

Title	Expert Name	Details of Expert	Web link
Big data Analytics	Ching-Yung Lin	IBM Chief Scientist	https://www.ee.columbia.edu/~cylin/course/bigdata/