

CSS309: BIG DATA ANALYTICS

Hours/Week: 4

Credits: 4

I.A. Marks: 30

Exam. Marks: 70

Course Learning Objectives: Students will try to learn,

- To optimize business decisions and create competitive advantage with Big Data analytics
- To explore the fundamental concepts of big data analytics.
- To learn to analyze the big data using intelligent techniques.
- To understand the various search methods and visualization techniques

Course Outcomes: After completing the course, the students will be able to,

- CO1: Implement statistical analysis techniques for solving practical problems.
- CO2: Perform statistical analysis on variety of data.
- CO3: Practically realize the working experiments of Python using Hadoop.
- CO4: Perform appropriate statistical tests using R and visualize the outcome.
- CO5: Understands the applications using Map Reduce Concepts.
- CO6: Develop Big Data Solutions using Hadoop Eco System.
- CO7: Manage Job Execution in Hadoop Environment.

UNIT- I

12 Hrs.

Introduction to Big Data Platform – Challenges of Conventional Systems - **Intelligent data analysis** – **Nature of Data - Analytic Processes and Tools** - Analysis vs Reporting - Modern Data Analytic Tools Statistical Concepts: Sampling Distributions - **Re-Sampling - Statistical Inference** - Prediction Error.

UNIT- II

12 Hrs.

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.**

UNIT-III

12 Hrs.

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.

UNIT- IV

12 Hrs.

Setting up a Hadoop Cluster - Cluster specification - Cluster Setup and Installation – **Hadoop Configuration-Security in Hadoop** - Administering Hadoop – HDFS - Monitoring-Maintenance Hadoop benchmarks- Hadoop in the cloud. Applications on **Big Data Using Pig and Hive –** Data processing operators in **Pig – Hive services** – HiveQL – Querying Data in Hive - fundamentals of HBase and **Zoo Keeper** - IBM InfoSphereBigInsights and Streams. Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications

REFERENCE BOOKS:

1. Tom White " Hadoop: The Definitive Guide" Third Edit on, O'reily Media, 2012.
2. Seema Acharya, SubhasiniChellappan, "Big Data Analytics" Wiley 2015.
3. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
4. Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press (2013)
5. Tom Plunkett, Mark Hornick, "Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop", McGraw-Hill/Osborne Media (2013), Oracle press.
6. AnandRajaraman and Jeffrey David Ulman, "Mining of Massive Datasets", Cambridge University Press, 2012.
7. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley & sons, 2012.
8. Glen J. Myat, "Making Sense of Data", John Wiley & Sons, 2007
9. Pete Warden, "Big Data Glossary", O'Reily, 2011.
10. Michael Mineli, Michele Chambers, AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
11. ArvindSathi, "BigDataAnalytics: Disruptive Technologies for Changing the Game", MC Press, 2012
12. Paul Zikopoulos ,Dirk DeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ", Tata McGraw Hill Publications, 2012.

