MC5502 BIG DATA ANALYTICS

DETAILED SYLLABUS

OBJECTIVES:

- 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.
- To learn to use various techniques for mining data stream.
- To understand the applications using Map Reduce Concepts

UNIT I INTRODUCTION TO BIG DATA

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 MINING DATA STREAMS

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 HADOOP ENVIRONMENT

History of Hadoop- The Hadoop Distributed File System – Components of Hadoop Analyzing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Hadoop filesystems-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 - Setting up a Hadoop Cluster - Cluster specification - Cluster Setup and Installation – Hadoop Configuration-Security in Hadoop

UNIT IV DATA ANALYSIS SYTEMS AND VISUALIZATION

Link Analysis – PageRank - Efficient Computation of PageRank- Topic-Sensitive PageRank – Link Spam- Recommendation Systems- A Model for Recommendation Systems- Content Based Recommendations - Collaborative Filtering- Dimensionality Reduction- Visualizations - Visual data analysis techniques-interaction techniques- Systems and applications.

UNIT V FRAMEWORKS AND APPLICATIONS

IBM for Big Data –Framework - Hive – Sharding – NoSQL Databases –Mango DB-Casandra Hbase – Impala – Analyzing big data with twitter – Big data for Ecommerce – Big data for blogs.

REFERENCES:

1. AnandRajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, CambridgeUniversity Press, 2014

2. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, “Understanding BigData: Analytics for Enterprise Class Hadoop and Streaming Data”, McGrawHill Publishing, 2012

3. Da Ruan, Guoquing Chen, Etienne E. Kerre, Geert Wets, Intelligent Data Mining, Springer, 2007

4. Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012

5. Glenn J. Myatt, “Making Sense of Data”, John Wiley & Sons, 2007

6. Jiawei Han, Micheline Kamber “Data Mining Concepts and Techniques”, Second Edition, Elsevier, Reprinted 2008

7. Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer, 2007

8. Michael Minelli, Michele Chambers, AmbigaDhiraj, Big Data, BigAnalytics: Emerging Business Intelligence and Analytic Trends for Today's

Businesses, WileyPublications,2013

9. Paul Zikopoulos, DirkdeRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, Harness the Power of Big Data the IBM Big Data Platform, Tata McGraw HillPublications, 2012

10. Pete Warden, “Big Data Glossary”, O"Reilly, 2011

11. Tom White “Hadoop: The Definitive Guide” Fourth Edition, O"reilly Media, 2015

12. Zikopoulos, Paul, Chris Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoopand Streaming Data, Tata McGraw Hill Publications, 2011.