

CP5293 BIG DATA ANALYTICS

DETAILED SYLLABUS

OBJECTIVES:

- To understand the competitive advantages of big data analytics
- To understand the big data frameworks
- To learn data analysis methods
- To learn stream computing
- To gain knowledge on Hadoop related tools such as HBase, Cassandra, Pig, and Hive for big data analytics

UNIT I INTRODUCTION TO BIG DATA

Big Data – Definition, Characteristic Features – Big Data Applications - Big Data vs Traditional Data - Risks of Big Data - Structure of Big Data - Challenges of Conventional Systems – Web Data – Evolution of Analytic Scalability - Evolution of Analytic Processes, Tools and methods - Analysis vs Reporting - Modern Data Analytic Tools.

UNIT II HADOOP FRAMEWORK

Distributed File Systems - Large-Scale File System Organization – HDFS concepts – MapReduce Execution, Algorithms using MapReduce, Matrix-Vector Multiplication – Hadoop YARN

UNIT III DATA ANALYSIS

Statistical Methods: Regression modelling, Multivariate Analysis - Classification: SVM & Kernel Methods - Rule Mining - Cluster Analysis, Types of Data in Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density Based Methods, Grid Based Methods, Model Based Clustering Methods, Clustering High Dimensional Data - Predictive Analytics – Data analysis using R.

UNIT IV MINING DATA STREAMS

Streams: Concepts – Stream Data Model and Architecture - Sampling data in a stream – Mining Data Streams and Mining Time-series data - Real Time Analytics Platform (RTAP) Applications - Case Studies - Real Time Sentiment Analysis, Stock Market Predictions.

UNIT V BIG DATA FRAMEWORKS

Introduction to NoSQL – Aggregate Data Models – Hbase: Data Model and Implementations – Hbase Clients – Examples – Cassandra: Data Model – Examples – Cassandra Clients – Hadoop Integration. Pig – Grunt – Pig Data Model – Pig Latin – developing and testing Pig Latin scripts. Hive – Data Types and File Formats – HiveQL Data Definition – HiveQL Data Manipulation – HiveQL Queries

REFERENCES:

1. Bill Franks, —Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced AnalyticsII, Wiley and SAS Business Series, 2012.
2. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", 2013.

3. Michael Berthold, David J. Hand, —Intelligent Data AnalysisII, Springer, Second Edition, 2007.
4. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
5. P. J. Sadalage and M. Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence", Addison-Wesley Professional, 2012.
6. Richard Cotton, "Learning R – A Step-by-step Function Guide to Data Analysis, O'Reilly Media, 2013.