

## BIG DATA ANALYSIS (3-0-0)

### Objective:

- To familiarize students with big data analysis as a tool for analysing large complex dataset.
- To learn to use various techniques for mining data stream.
- Understand the applications using Map Reduce Concepts
- Provide hands on Hadoop Eco System
- To introduce programming tools PIG & HIVE in Hadoop echo system

### MODULE – I

Introduction To Big Data, Data Storage and Analysis - Characteristics of Big Data – Big Data Analytics - Typical Analytical Architecture – Requirement for new analytical architecture – Challenges in Big Data Analytics – Need of big data frameworks

### MODULE – II

NoSQL Database: NoSQL Databases - Schema less Models, Increasing Flexibility for Data Manipulation-Key Value Stores, Document Stores, Tabular Stores, Object Data Stores – GraphDatabases, Big data for twitter, Big data for E-Commerce blogs.

### MODULE – III

Big Data: Evolution of Big data, Best Practices for Big data Analytics - Big data characteristics - Big Data Use Cases, Characteristics of Big Data Applications, Big Data Modelling, HDFS performance and tuning, Map reduce algorithm, Hadoop Eco system Pig : Introduction to PIG, Execution Modes of Pig, Grunt, Pig Latin, User Defined Functions, Data Processing operators.

### MODULE – IV

Hive: Hive Shell, Hive Services, HiveQL, Tables, Querying Data and User Defined Functions. Hbase :HBasics, Concepts, Clients, Example, Spark Mining Data Streams: Introduction to Streams Concepts, Stream Data Model and Architecture -Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream –Real time Analytics Platform (RTAP) applications, Case Studies, Real Time Sentiment Analysis- Stock Market Predictions.

### Outcomes:

At the end of the course the students will be able to:

- Process data in Big Data platform and explore the big data analytics techniques for business applications
- Analyse Map Reduce technologies in big data analytics
- Develop Big Data solutions using Hadoop Eco System
- Design efficient algorithms for stream data mining on big data platform

### Books Recommended:

1. Jure Leskovec, AnandRajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2014.
2. Tom White ,Hadoop: The Definitive Guide, 4th edition O'Reilly Publications, 2015
3. Judith Hurwitz, Alan Nugent, Dr. Fern Halper, and Marcia Kaufman, "Big data for dummies" A wiley brand publications.
4. Holden Harau, "Learning Spark: Lightning-Fast Big Data Analysis", O-Reilly Publications
5. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", 2013.
6. BartBaesens, "Analytics in a Big Data World: The Essential Guide to Data Science and itsApplications", Wiley Publishers, 2015.

7. Kim H. Pries and Robert Dunnigan, "Big Data Analytics: A Practical Guide for Managers "CRC Press, 2015.
8. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", Wiley publishers, 2015.
9. Dietmar Jannach, Markus Zanker, Alexander Felfernig and Gerhard Friedrich "Recommender Systems: An Introduction", Cambridge University Press, 2010.
10. Jimmy Lin, Chris Dyer and Graeme Hirst, "Data-Intensive Text Processing with MapReduce", Synthesis Lectures on Human Language Technologies, Vol. 3, No. 1, Pages 1-177, Morgan Claypool publishers, 2010.