3.Big Data Analytics

L-T-P 3-0-0 Cr. –3

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 Hodoop 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 DataManipulation-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, HDFSperformance 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, SparkMining 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 timeAnalytics Platform (RTAP) applications, Case Studies, Real Time Sentiment Analysis- StockMarket 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'Reily 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 withTools, Techniques, NoSQL, and Graph", 2013.