

CS8091 BIG DATA ANALYTICS

DETAILED SYLLABUS

OBJECTIVES:

- To know the fundamental concepts of big data and analytics.
- To explore tools and practices for working with big data
- To learn about stream computing.
- To know about the research that requires the integration of large amounts of data.

UNIT I INTRODUCTION TO BIG DATA

Evolution of Big data - Best Practices for Big data Analytics - Big data characteristics - Validating – The Promotion of the Value of Big Data - Big Data Use Cases- Characteristics of Big Data Applications - Perception and Quantification of Value -Understanding Big Data Storage - A General Overview of High-Performance Architecture - HDFS - MapReduce and YARN - Map Reduce Programming Model

UNIT II CLUSTERING AND CLASSIFICATION

Advanced Analytical Theory and Methods: Overview of Clustering - K-means - Use Cases - Overview of the Method - Determining the Number of Clusters - Diagnostics - Reasons to Choose and Cautions - Classification: Decision Trees - Overview of a Decision Tree - The General Algorithm - Decision Tree Algorithms - Evaluating a Decision Tree - Decision Trees in R - Naïve Bayes - Bayes' Theorem - Naïve Bayes Classifier.

UNIT III ASSOCIATION AND RECOMMENDATION SYSTEM

Advanced Analytical Theory and Methods: Association Rules - Overview - Apriori Algorithm – Evaluation of Candidate Rules - Applications of Association Rules - Finding Association & finding similarity - Recommendation System: Collaborative Recommendation- Content Based Recommendation - Knowledge Based Recommendation- Hybrid Recommendation Approaches.

UNIT IV STREAM MEMORY

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. Using Graph Analytics for Big Data: Graph Analytics

UNIT V NOSQL DATA MANAGEMENT FOR BIG DATA AND VISUALIZATION

NoSQL Databases: Schema-less Models: Increasing Flexibility for Data Manipulation-Key Value Stores- Document Stores - Tabular Stores - Object Data Stores - Graph Databases Hive - Sharding –HBase – Analyzing big data with twitter - Big data for E-Commerce Big data for blogs - Review of Basic Data Analytic Methods using R.

TEXT BOOKS:

1. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012.

2. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", Morgan Kaufmann/Elsevier Publishers, 2013.