7th Semester	RIT7D002	Bigdata Analytics	L-T-P 3-0-0	3 CREDITS

Module-1

Introduction to Big Data: Types of Digital Data-Characteristics of Data – Evolution of Big Data - Definition of Big Data - Challenges with Big Data - 3Vs of Big Data - Non Definitional traits of Big Data - Business Intelligence vs. Big Data - Data warehouse and Hadoop environment - Coexistence. Big Data Analytics: Classification of analytics - Data Science - Terminologies in Big Data - CAP Theorem - BASE Concept. NoSQL: Types of Databases – Advantages – NewSQL - SQL vs. NOSQL vs NewSQL. Introduction to Hadoop: Features – Advantages – Versions - Overview of Hadoop Eco systems - Hadoop distributions - Hadoop vs. SQL – RDBMS vs. Hadoop - Hadoop Components – Architecture – HDFS - Map Reduce: Mapper – Reducer – Combiner – Partitioner – Searching – Sorting - Compression. Hadoop 2 (YARN): Architecture - Interacting with Hadoop Eco systems.

Module-2

No SQL databases: Mongo DB: Introduction – Features - Data types - Mongo DB Query language - CRUD operations – Arrays - Functions: Count – Sort – Limit – Skip – Aggregate - Map Reduce. Cursors – Indexes - Mongo Import – Mongo Export. Cassandra: Introduction – Features - Data types – CQLSH - Key spaces - CRUD operations – Collections – Counter – TTL - Alter commands - Import and Export - Querying System tables.

Module-3

Hadoop Eco systems: Hive – Architecture - data type - File format – HQL – SerDe - User defined functions - Pig: Features – Anatomy - Pig on Hadoop - Pig Philosophy - Pig Latin overview - Data types - Running pig - Execution modes of Pig - HDFS commands - Relational operators - Eval Functions - Complex data type - Piggy Bank - User defined Functions - Parameter substitution - Diagnostic operator. Jasper Report: Introduction - Connecting to Mongo DB - Connecting to Cassandra - Introduction to Machine learning: Linear Regression – Clustering - Collaborative filtering - Association rule mining - Decision tree.

Books:
1. Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley Publication, 2015.
2. Judith Hurwitz, Alan Nugent, Dr. Fern Halper, Marcia Kaufman, "Big Data for Dummies", John Wiley & Sons, Inc., 2013.
3. Tom White, "Hadoop: The Definitive Guide", O'Reilly Publications, 2011.
4. Kyle Banker, "Mongo DB in Action", Manning Publications Company, 2012.
5. Russell Bradberry, Eric Blow, "Practical Cassandra A developers Approach", Pearson Education, 2014.