RCETCSAC04 - Big data Analytics

Course outcomes:

- Identify Big Data and its Business Implications.
- List the components of Hadoop and Hadoop Eco-System
- Access and Process Data on Distributed File System
- Manage Job Execution in Hadoop Environment
- Develop Big Data Solutions using Hadoop Eco System

Syllabus:

UNIT I

Introduction To Big Data And Analytics

Classification of Digital Data, Structured and Unstructured Data –Introduction to Big Data: Characteristics – Evolution – Definition - Challenges with Big DataOther Characteristics of Data - Why Big Data - Traditional Business Intelligence versus Big Data - Data Warehouse and Hadoop Environment Big Data Analytics: Classification of Analytics – Challenges - Big Data Analytics important - Data Science - Data Scientist - Terminologies used in Big Data Environments - Basically Available Soft State Eventual Consistency - Top Analytics Tools

UNIT II

Introduction To Technology Landscape

NoSQL, Comparison of SQL and NoSQL, Hadoop -RDBMS Versus Hadoop - Distributed Computing Challenges – Hadoop Overview - Hadoop Distributed File System - Processing Data with Hadoop - Managing Resources and Applications with Hadoop YARN - Interacting with HadoopEcosystem

UNIT III

Introduction To Mongodb and Mapreduce Programming

MongoDB: Why Mongo DB - Terms used in RDBMS and Mongo DB – Data Types - MongoDB Query Language MapReduce: Mapper – Reducer – Combiner – Partitioner – Searching – Sorting – Compression

Reference Text Books:

1. The Data Warehouse Lifecycle Toolkit, Kimball et al., Wiley 1998
2. Hadoop in Practice by Alex Holmes, MANNING Publ.
3. Multidimensional Databases and Data Warehousing, Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen, Morgan & Claypool Publishers, 2010
4. Hadoop in Action by Chuck Lam, MANNING Publ.