

Course Objectives

- The aim of this course is to provide depth knowledge about Big data Technologies and tools used for Big data.
- The students will learn to implement and work on tools to handle large volume of data in parallel and distributed environments. Retrieval and analysis of unstructured data are done using NOSQL databases.

Course Outcomes

CO1: Understand fundamental concepts of Big Data and its technologies

CO2: Apply concepts of MapReduce framework for optimization

CO3: Analyze appropriate NoSQL database techniques for storing and processing large volumes of structured and unstructured data

CO4: Apply data analytics solutions using Hadoop ecosystems

CO5: Explore modern reporting tools for Machine learning

CO-PO Mapping

PO/ PSO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO														
CO1	3	2	1							1			3	2
CO2	3	3	2	2	3	2	2	2	2	2			3	2
CO3	3	3	2	2	3	2	2	2	2	2			3	2
CO4	3	3	2	2	3	2	2	2	2	2			3	2
CO5	2			2	3		2	3	2	2			3	2

Syllabus

Unit 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.

Unit 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.

Unit 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.

Text Book(s)

Seema Acharya, Subhashini Chellappan, “Big Data and Analytics”, Wiley Publication, 2015.

Reference(s)

Hurwitz JS, Nugent A, Halper F, Kaufman M. Big data for dummies. John Wiley & Sons; 2013.

Tom White, “Hadoop: The Definitive Guide”, O’Reilly Publications, 2011.

Kyle Banker, “Mongo DB in Action”, Manning Publications Company, 2012.

Russell Bradberry, Eric Blow, “Practical Cassandra A developers Approach “, Pearson Education, 2014.

Evaluation Pattern

Assessment	Internal	External
Periodical 1	10	
Periodical 2	10	
*Continuous Assessment (Theory) (CAT)	10	
Continuous Assessment (Lab) (CAL)	40	
End Semester		30

*CA – Can be Quizzes, Assignment, Projects, and Reports.