Year: B. Tech IV (Semester VII)

Subject Name: Big Data Analytics **Subject Code:** BTCO14709
Type of course: Professional Elective - V
Prerequisite: Database Management System, Data Mining

Rationale: In this subject, students will gain knowledge of Big Data analytics. The subject serves as an introductory course for students who are expecting to explore Big Data storage, processing, analysis, and application issues in both workplaces and research environments.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	0	3	60	25	15	00	00	100

Contents:

Unit No	Contents	Hrs
1	Introduction to BIG DATA : Introduction to Big Data, Types of Data, V's of Big Data, Challenges with Big Data, Big Data Analytics, Classification of Analytics, Importance and applications of Big Data Analytics	5
2	Hadoop Ecosystem : Introduction, Hadoop Architecture, Basics of MapReduce and HDFS, Architecture of HDFS, Concept of Blocks in HDFS, Name Node, Data Node, Secondary Name Node, File read and File write operation, Task Tracker, Job Tracker (Hadoop 1. X), YARN (Hadoop 2. X)	12
3	MapReduce : MapReduce Architecture: Mapper, Reducer, Combiner, Partitioner, Shuffling and Sorting, MapReduce-based examples - Word Count, Matrix Multiplication	8
4	PIG and HIVE : Pig - Architecture, Data Types -scalar and complex, Pig Operators, Hive - Services and Architecture, Comparing Hive to traditional Databases, Hive Data Types, Hive File Format, Basics of Hive Query Language	6
5	NoSQL, MongoDB, and Cassandra : NoSQL, Types of No SQL databases, SQL Vs No SQL, why No SQL, Introduction to MongoDB, Data Types in MongoDB, Apache Cassandra, Features of Cassandra, CRUD operations	10
6	SPARK : Introduction to Data Analysis with Spark Architecture, RDD and Operations, Driver Node, Worker Nodes, Cluster Managers, Working with Key/Value Pairs	4

Suggested Specification table with Marks (Theory): (For B.Tech only)

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
15	20	10	10	5	0

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Reference Books:

Sr No	Title of book /article	Author(s)	Publisher and details like ISBN	Year of publication	Publication Edition
1	Big Data and Analytics	Seema Acharya, Subhashini Chellappan	Wiley India	2019	2nd
2	Hadoop: The Definitive Guide	Tom White	O'Reilly	2012	3rd
3	Learning Spark	Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee	O'Reilly	2020	2nd
4	Understanding Big Data	Paul C. Zikopoulos, Chris Eaton, Dirk deRoos, Thomas Deutsch, George Lapis	McGraw Hill	2017	-

Course Outcomes (CO):

Sr. No.	CO statement	Marks % weightage
1	Discuss the concepts of Big Data Analytics and its importance	10

2	Demonstrate map reduce programming model using Hadoop architecture	40
3	Utilize SPARK framework for performing data analytics	10
4	Compare and contrast Pig and Hive with relational database	15
5	Practice various data manipulation operations with NoSQL, MongoDB and Cassandra	25

List of Open learning website:

1. https://onlinecourses.nptel.ac.in/noc20_cs92/preview

List of Open Source Software:

1. Hadoop
2. Spark
3. Cassandra
4. MongoDB
5. Tabula
6. PowerBi