GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Artificial Intelligence and Data Science

Subject Code: ME02095021

Course/Subject Name: Big Data Analytics

WEF Academic Year	2024-25
Semester	2
Category of the Course	PCC

Prerequisite:	Basic knowledge of database, network, data structure and algorithms.
Rationale:	• The course will explore the different methods and techniques for data analytics. • The course will focus to conceptualize the different method for data storage. • The course will also cover the different Hadoop services.

Course Outcome:

After completion of the Course, Students will be able to:

No	Course Outcomes	RBT Level*
01	To understand the basic concepts of big data and Hadoop.	UN
02	To apply the analytics algorithms to analyze large-scale distributed databases.	AP
03	To formulate data storage problems for real time applications.	AP
04	To analyse data analytics models & techniques.	AN
05	To evaluate the different real time applications by using Hadoop services.	EL

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

Teaching and Examination Scheme:

Teaching Scheme (in hours)			Total Credits (L+T+PR/2)	Assessment Pattern and Marks				Total Marks
				Theory		Tutorial/Practical		
L	T	PR	C	ESE (E)	PA/CA (M)	PA/CA (I)	ESE (V)	
03	00	02	04	70	30	20	30	150

Course Content:

Unit	Course Content	No of Hours	% of Weightage
1	**Unit-I: Introduction to Big Data and Hadoop** Introduction– distributed file system–Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Big Data – Apache Hadoop & Hadoop Eco System, Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce –Data Serialization	05	10
2	**Unit-II: HDFS, HIVE AND HIVEQL, HBASE** HDFS-Overview, Installation and Shell, Java API; Hive Architecture and Installation, Comparison with Traditional Database, HiveQL Querying Data, Sorting And Aggregating, Joins & Sub queries, HBase concepts, Advanced Usage, Schema Design, Advance Indexing, PIG, Zookeeper, how it helps in monitoring a cluster	09	25
3	**Unit-III: Data Analysis with Spark Shell** Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution	07	15
4	**Unit-IV: NoSQL** What is it?Where It is Used Types of NoSQL databases, Why NoSQL? Advantages of NoSQL, Use of NoSQL in Industry, SQL vs NoSQL, NewSQL, Database for the Modern Web	08	15

5	**Unit-V: MongoDB, Spark SQL and GraphX** Introduction to MongoDB key features, Core Server tools, MongoDB through the JavaScript's Shell, Creating and Querying through Indexes, Document-Oriented, principles of schema design, Constructing queries on Databases, collections and Documents, MongoDB Query Language, SQL Context – Importing and Saving data – Data frames – using SQL – GraphX overview – Creating Graph – Graph Algorithms	10	25
6	**Unit-VI : Current Trends** The latest technology and algorithms in data analytics for real time applications.	06	10
	TOTAL	**45**	**100**

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
10	**20**	**20**	**20**	**20**	**10**

Where R: Remember; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (as per Revised Bloom's Taxonomy)

Reference/Suggested Learning Resources:

(a) Books:

1. Big Data and Analytics by Seema Acharya, Subhashini Chhellappan Willey, Second edition
2. Professional Hadoop Solutions by Boris lublinsky, Kevin t. Smith, Alexey Yakubovich Wiley, ISBN-13: 978-1118611937, ISBN-10: 1118611934, 2015.
3. Understanding Big data by Chris Eaton and Paul C. Zikopoulos McGraw Hill, 2012
4. HADOOP: The definitive Guide by Tom White O'Reilly, 2012
5. Big Data Analytics with R and Hadoop by Vignesh Prajapati ,Packet Publishing, 2013
6. Mongo DB in Action 3.0 by Kyle Banker ,Peter Bakkum , Shaun Verch , Doug Garrett by Manning. ISBN-10-ISBN-1617291609
7. MongoDB: The Definitive Guide by Kristina Chodorow, 2nd Edition, , O'Reilly Media. ISBN- 1449344682

(b) Open source software and website

- Course-related online MOOCs on NPTEL/SWAYAM platform.
- Recently Published papers/articles in reputed journals.

Suggested Course Practical List:

- The practical work will be carried out based on the content covered during the academic sessions**.**

List of Laboratory/Learning Resources Required: Programming development environment (open source is encouraged) related to the course content.

Suggested Project List: The subject teacher has to assign the relevant project work to the students in individual/team.

Suggested Activities for Students: The subject teacher has to assign the outcome based activities to the students in individual/team.

****