GANPAT UNIVERSITY

FACULTY OF COMPUTER APPLICATIONS

Programme	Master of Computer Applications	**Branch/Spec.**	Computer Application
Semester	II	**Version**	1.0.0.2
Effective from Academic Year	2022-23	**Effective for the batch Admitted in**	June 2022
Subject Code	P12A4BD1	**Subject Name**	Big Data Analytics-I

Teaching scheme						**Examination scheme (Marks)**			
(Per week)	**Lecture (DT)**		**Practical (Lab.)**		**Total**		**CE**	**SEE**	**Total**
	L	TU	P	TW					
Credit	2	0	2	0	4	Theory	40	60	100
Hours	2	0	4	0	6	Practical	20	30	50

Objective:

- This course provides an overview of Apache Hadoop, HDFS and Map Reduce concepts.
- Understand the different technology trends, Big Data market and different job roles in Big Data.
- This course provides the knowledge of Hadoop installation on your machine and understand the complex architectures of Hadoop and its components.
- learn how to inject data into Hadoop for build, maintain, scalable and distributed systems.
- Provides the overview, how MapReduce can be utilize and implement to analyze big data sets.

Pre-requisites:

- The student should have a basic knowledge of Object-Oriented Programming and High-Level Programming Language concepts.
- Basic knowledge of Linux operating system command

Course Outcomes :

- **1 = Slight (Low); 2 = Moderate (Medium); 3 = Substantial (High); "-" = No Correlation**

Name of CO	Description
CO1	Explain Big Data concepts, evolution, architecture, and analytics terminology including BI and KPI components.
CO2	Classify Big Data characteristics, data types, and analysis techniques to select suitable analytical methods.
CO3	Demonstrate Hadoop ecosystem components by installing Hadoop and executing basic Hadoop commands.
CO4	Implement HDFS and MapReduce frameworks to store, manage, and process large-scale datasets.

Mapping of CO and PO								
Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	2	2	1	1	–	–	–	2
CO2	2	3	2	2	–	–	–	2

CO3	3	2	2	3	1	–	–	2
CO4	3	3	3	3	1	–	–	2

Content:

Unit	SECTION-I	Hrs
01	**Introduction to Big Data:** Introduction to Big Data, Evolution of Big Data, Use of Big Data in the Market, Big Data Architecture, Concepts and Terminology: Datasets, Data Analysis, Data Analytics, Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, Prescriptive Analytics, Business Intelligence (BI), Key Performance Indicators (KPI).	7
02	**Big Data Characteristics & Analysis Techniques:** **Big Data Characteristics:** Volume, Velocity, Variety, Veracity, Value, Different Types of Digital Data: Structured Data, Semi-Structured data, Unstructured Data, Difference between Structured, Semi-structured and Unstructured data. **Big Data Analysis Techniques:** Quantitative analysis, Qualitative analysis, Data mining, Statistical analysis, Semantic analysis, Visual analysis	8
	SECTION-II	
03	**Introducing Hadoop:** Introduction to Hadoop and its Features, The Hadoop Ecosystem (Core Components), Different Hadoop Distributions, Different types of Hadoop Vendors, Hadoop Installation and Basic Hadoop Command, and Developing Enterprise Applications with Hadoop.	7
04	**Big Data Storage:** Introduction Hadoop Distributed File System, HDFS Architecture, HDFS core components: Name Node, Data Node, Secondary Name Node, Blocks Storage, Data Replication and Rack Awareness, Using HDFS Files, Managing HDFS Command from Command Line, Anatomy of File Read and Write on HDFS. Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce, Data Serialization **Analysis using Map Reduce:** Introduction Map Reduce, MapReduce Distribute Processing, Introduction to Yarn and Resource Manager.	8

Practical Content:

List of programs specified by the subject teacher based on above mentioned topics.

Text Books:

1	Hadoop: The Definitive Guide by Tom White, O'Reilly Media,4th Edition,2015
2	Big Data Simplified by Sourabh Mukherjee, Amit Kumar Das, Pearson Publications, 1st Edition 2019.

Reference Books:

1	Big Data Fundamentals: Concepts, Drivers & Techniques by Thomas; Khattak, Wajid; Buhler, Paul Erl, Pearson publication, 1st Edition 2015.
2	Professional Hadoop Solutions by Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich, Wiley India Pvt. Ltd., ISBN: 9788126551071, 2015
3	Big Data and Analytics by Seema Acharya, Subhashini Chellappan, Wiley India Pvt. Ltd.,2 nd Edition 2019

MOOC/Certification Courses:

1	https://nptel.ac.in/
2	https://www.edx.org/course/big-data-hadoop-and-spark-basics?index=product
3	https://www.udemy.com/course/big_data/

4.	https://careerfoundry.com/en/blog/data-analytics/free-data-analytics-courses/

Question Paper Scheme:

University Examination Duration: 3 Hours

Note for Examiner: -

(I) Questions 1 and 4 are compulsory with no options.

(II) Internal options should be given in questions 2, 3, 5 and 6.

SECTION – I

Q.1 –8 Marks

Q.2 –11 Marks

Q.3 –11 Marks

SECTION - II

Q.4 –8 Marks

Q.5 –11 Marks

Q.6 –11 Marks