GANPAT UNIVERSITY

FACULTY OF COMPUTER APPLICATIONS

Programme	Master of Computer Applications	**Branch/ Spec.**	Computer Application
Semester	III	**Version**	1.0.0.0
Effective from Academic Year	2023-24	**Effective for the batch Admitted in**	June 2022
Subject Code	P13A5BD2	**Subject Name**	Big Data Analytics-II

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 HBase, Pig, Hive and Spark concepts.
- Students can understand the different technology trends and different job roles in Big Data.
- This course provides the knowledge of software installation on systems and understand the complex architectures and its components. learn how to inject data and analyze data.

Pre-requisites:

- The student should have a basic knowledge of Database Programming and High-Level Programming Language concepts.

Course Outcomes :

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

Name of CO	**Description**
CO1	Explain HBase features & architecture and apply installation, table management, data manipulation, and MapReduce integration.
CO2	Analyze Pig architecture & execution modes and apply operators/functions using Pig Latin and Grunt shell.
CO3	Explain Hive architecture & components and apply DDL/DML commands, functions, views, and indexes for data management.
CO4	Explain Spark architecture and apply RDD creation, transformations, and actions using Spark-

	Scala.
	Mapping of CO and PO

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	2	3	1	-	-	2
CO2	3	3	2	3	1	-	-	2
CO3	3	2	2	3	1	-	-	2
CO4	3	3	2	3	1	-	-	3

Content:

Unit	SECTION-I	Hrs
01	**Apache HBase:** Introduction Apache HBase, Features of HBase, HBase Pros and Cons, Architecture of HBase, HBase Installation, HBase Command, Table Management Commands, Data Manipulation Commands, HBase Operations: Read Data and Write Data Operations, HBase MapReduce Integration	7
02	**Apache Pig:** Introduction to Apache Pig, MapReduce vs Pig, Architecture of Apache Pig, Apache Pig Components: Parser, Optimizer, Compiler, Execution Engine; Job Execution Modes of Pig: Local Mode and Map Reduce Mode; Datatypes, Data Processing Operators: Load and Store, Pig Functions: Evel, Bag and Tuple, String, Math; Grunt Shell: Command, Utility Command;	8
	SECTION-II	
03	**Apache Hive:** Introduction Apache Hive, Features of Apache Hive, How Does Hive Works?, Hive Installation, Apache Hive Architecture & Components, Hive Data Types, Hive Built-In Functions: Collection, Date, Mathematical, Conditional, String; Hive DDL Commands, Hive DML Command, Hive View and Index, Hive Data Model.	7
04	**Introducing Spark:** Introduction Apache Spark Programing Language, Understating Apache Spark Architecture, Apache Spark Ecosystem Components, Use Case of Apache Spark Applications, Apache Spark Installation, Spark Shell Commands to Interact with Spark-Scala, Resilient Distributed Dataset (RDD): Introduction of RDD, Features of Spark RDD, Create RDDs in Apache Spark, RDD Transformation Function: Map, Flat Map, Filter, Union, Intersection, Distinct, Group by Key, Reduce by Key, Sort by Key; RDD Action: Count, Collect, Take, Top, Reduce;	8

Practical Content:

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

Text Books:

1	HBase: The Definitive Guide by Lars George, Publisher : O′Reilly; 1st edition (7 October 2011).
2	Beginning Apache Pig by Balaswamy Vaddeman, Publisher : Apress; 1st ed. edition (16 December 2016).
3	Apache Hive Essentials: Essential techniques to help you process, and get unique insights from, big data, 2nd Edition by Dayong Du
4	Big Data Simplified by Sourabh Mukherjee, Amit Kumar Das,Publisher : Pearson Education; First edition (18 July 2019).
5	Mastering Apache Spark by Mike Frampton,Publisher : Ingram short title (1 January 2015).

Reference Books:	
1	HBase Administration Cookbook by Yifeng Jiang
2	Programming Pig, 2nd Edition by Alan Gates, Daniel Dai
3	Programming Hive: Data Warehouse and Query Language for Hadoop by Dean Wampler, Jason Rutherglen & Edward Capriolo
4	Learning Spark by Matei Zaharia, Patrick Wendell, Andy Konwinski, Holden Karau
5	Advanced Analytics with Spark: Patterns for Learning from Data at Scale by Sandy Ryza, Uri Laserson, Sean Owen and Josh Wills
MOOC/Certification Courses:	
1	https://www.edx.org
2	https://www.udemy.com
3	https://in.coursera.org

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