GANPAT UNIVERSITY									
FACULTY OF ENGINEERING & TECHNOLOGY									
Programme	Bachelor of Technology					Branch/Spec.	Computer Science & Engineering (BDA)		
Semester	V					Version	1.0.0.1		
Effective from Academic Year			2022-23			Effective for the batch Admitted in			June 2020
Subject Code	2CSE506		Subject Name			BIG DATA ANALYTICS			
Teaching scheme						Examination scheme (Marks)			
(Per week)	Lecture(DT)		Practical(Lab.)		Total		CE	SEE	Total
	L	TU	P	TW					
Credit	3	0	2	0	5	Theory	40	60	100
Hours	3	0	4	0	7	Practical	60	40	100

Pre-requisites:

Database Management System, JAVA/Python Programming Language

Learning Outcomes:

Upon Completion of the course, the students will be able to

- Understand big data phenomena and its applications
- Understand distributed file system concepts and algorithms
- Implement distributed storage and big data processing concepts using open source tools like Hadoop
- Apply Hadoop ecosystem tools to implement big data related case studies.

Theory syllabus

Unit	Content	Hrs
1	**Distributed file system fundamentals** Introduction to distributed file system, difference between normal file system and distributed file system, benefits and requirement of distributed file system, distributed file system algorithms	4
2	**Data Science Basics** Introduction to big data and data analytics and its importance, Characteristics of Big Data, drivers of big data, big data case studies, data science pipeline	5

3	**Hadoop Architecture (HDFS, MapReduce, YARN)** introduction to Hadoop Distributed File System (HDFS), HDFS commands, HDFS Architecture, HDFS read and write operations, MapReduce Framework, Map Reduce phases, Failover mechanism, Introduction to YARN, YARN Architecture, Use cases	**9**
4	**Hadoop Ecosystem** Storing and Querying Data using PIg and HIVE, HBase, Slider and Knox, Sqoop	**20**
5	**Hadoop Administration** Key areas of Hadoop Administration, Creating and configuring Hadoop cluster, Apache Ambari, Zookeeper, Security and Governance.	**4**
6	**Enterprise Analytics Tool** Introduction to IBM Watson Studio, Analyzing data with Watson Studio, Big Data tools available in watson studio	**3**

Practical List
The Practicals will be based on implementing various tasks using Hadoop ecosystem tools - HDFS commands, Pig Latin, Hive, HBASE, Sqoop, Watson Studio Big Data tools

Text Books	
1	Hadoop: The Definitive Guide, By Tom White

Reference Books	
1	Big Data and Analytics, by Subhashini Chellappan Seema Acharya
2	Big Data Analytics with Hadoop 3 by Sridhar Alla
3	Harness the Power of Big Data The IBM Big Data Platform by Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan

Course Outcomes	
COs	Description
CO1	Understand big data phenomena and its applications
CO2	Understand distributed file system concepts and algorithms
CO3	Implement distributed storage and big data processing concepts using open source tools like Hadoop
CO4	Apply Hadoop ecosystem tools to implement big data related case studies.

Mapping of CO and PO:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	2	0	0	0	0	0	1	0	2	0
CO2	3	2	2	0	0	0	0	0	1	0	2	0
CO3	3	2	2	0	0	0	0	0	2	0	1	0
CO4	2	2	0	0	0	0	0	0	1	0	0	0