GANPAT UNIVERSITY									
FACULTY OF ENGINEERING & TECHNOLOGY									
Programme	Bachelor of Technology					Branch/Spec.	Computer Science & Engineering (BDA)		
Semester	VII					Version	1.0.0.0		
Effective from Academic Year			2022–23			Effective for the batch Admitted in			June 2019
Subject code	2CSE711		Subject Name			ADVANCED 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	1	0	4	Theory	40	60	100
Hours	3	0	2	0	5	Practical	30	20	50

Pre-requisites:

Data Science & Analytics, Big Data Analytics, Database Management System, HDFS and MapReduce

Learning Outcome:

After successful completion of the course students will be able to:

- Understand several key big data technologies used for storage, analysis and manipulation of data.
- Recognize the key concepts of MongoDB, Scala and Spark
- Solve business intelligence related queries using Data Visualization tool, i.e., tableau, PowerBI
- Apply Advance concepts related to big data in projects/real life scenarios

Theory syllabus

Unit	Content	Hrs
1	**NOSQL Databases** Types and importance of NoSQL Databases, MongoDB CRUD operations, Aggregation Framework, indexes	15
2	**Processing stream data** **SCALA:** What is Scala? Basic Operations, variable types, control structure, foreach loop, functions, procedures, array, higher order functions, Class in Scala, getters and setters, constructor, singletons, traits **SPARK:** Spark Components & its Architecture, Spark Deployment Modes, Spark Resilient Distributed Dataset (RDDs), RDD operations, transformations and actions, data loading and saving, Key-Value Pair RDDs, RDD Persistence, SPARK SQL, dataframes and datasets, JSON and Parquet file formats	20
3	**Data Visualization tools and techniques** **Tableau:**	10

	Data type, file type, tool type, show me menu, Type of data source supported by, how to connect different data sources, edit metadata, filter fields, filter data source, type of charts, filter data, data joining, data blending, extract data, adding filter data, apply filter on chart and data, number functions, string functions. **Power BI:** Components of Power BI, designing tables and reports, preparing dashboards	

Practical content
Practicals will be based on: ● MongoDB CRUD Operations ● MongoDB Aggregation Framework ● Tableau: Preparing Dashboards and Stories ● Spark: Multiple practicals on Scala and Spark and SPARK SQL (dealing with different types of data like mysql, csv, json, parquet, etc)

Text Books	
1	Spark: The Definitive Guide by Bill Chambers and Matei Zaharia
2	MongoDB: The Definitive Guide by Shannon Bradshaw , Eoin Brazil

Reference Books	
1	Scala in depth by Joshua D. Suereth
2	Tableau For Dummies by Molly Monsey and Paul Sochan
3	Introducing Microsoft Power BI, Alberto Ferrari and Marco Russo

Course Outcomes:

Cos	Description
CO1	Understand several key big data technologies used for storage, analysis and manipulation of data.
CO2	Recognize the key concepts of MongoDB, Scala and Spark
CO3	Solve business intelligence related queries using Data Visualization tool like tableau, PowerBI
CO4	Apply Advance concepts related to big data in projects/real life scenarios

Mapping of CO and PO:

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