GANPAT UNIVERSITY

FACULTY OF ENGINEERING & TECHNOLOGY

U. V. Patel College of Engineering

Computer Engineering / Information Technology / CE(AI)

VII – SEMESTER

2CEIT702: BIG DATA ANALYTICS (SYLLABUS) (Nov-Dec 2024)

Unit	Content
1.	**Introduction:** Distributed file system–Big Data and its importance, Vs of Big Data, Drivers for Big data, big data applications.
2.	**Big Data Enabling Technologies:** Big Data – Apache Hadoop & Hadoop EcoSystem, Hadoop, Distributed File system, YARN, Hadoop Mapreduce programming model.
3.	**Big Data Storage Platforms**: CAP Theorem, Apache Cassandra, Apache Zookeeper, ACID Properties, BASE Properties
4.	**Apache Spark and Kafka:** Data analysis with Spark, Introduction to Big Data Streaming Platforms , Big Data Pipelines for Real-Time computing, Introduction to Spark Streaming, Kafka, Streaming Ecosystem, SparkSQL, Kafka Introduction.
5.	**Big Data Application:** Big data in Machine learning (Introduction, Relation of AI,ML,DL,DS and Big Data), ML Process, Types of ML, Linear Regression, K-Means Algorithm)

NOTE: Subject Materials and Lecture PPT's are available on Moodle (LMS)