

GANPAT UNIVERSITY									
FACULTY OF ENGINEERING & TECHNOLOGY									
Programme	Bachelor of Technology				Branch/Spe c.	Computer Science & Engineering (BDA)			
Semester	VI				Version	1.0.1.1			
Effective from Academic Year		2018-19			Effective for the batch Admitted in			June 2016	
Subject code	2CSE60E13		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 Outcome:									
Upon Completion of the course, the students will be able to									
<ul style="list-style-type: none"> Identify and distinguish big data analytics applications Describe big data analytics tools Explain big data analytics techniques Present cases involving big data analytics in solving practical problems Conduct big data analytics using system tools Suggest appropriate solutions to big data analytics problems 									
Theory syllabus									
Unit	Content								Hrs
1	Introducing Big Data Introduction- Big Data and its importance, 4 Vs, structured, semi structured and structured data, big data problem in future								4
2	Introduction to Hadoop and Hadoop Architecture Features, RDBMS Vs. Hadoop, versions, Introduction to some Hadoop distributions, Hadoop technology stack								5
3	Hadoop Distributed File System (HDFS) Introduction, HDFS Daemons, Different methods to HDFS access								5
4	Map Reduce Introduction, Phases involved in Map Reduce, Architecture, Working of Map Reduce using Example Program								5
5	HBase Limitations of Hadoop, Introduction to HBase, HBase Vs. RDBMS, Features, HBase components, General commands								5
6	Introduction, Overview, and History of NoSQL Databases Definition of the Four Types of NoSQL Database, Why NoSQL? The Value of Relational Databases, Getting at Persistent Data, Concurrency, Integration, Impedance Mismatch, Application and Integration Databases, Attack of the Clusters, The Emergence of NoSQL, Key Points								4

7	Comparison of relational databases to new NoSQL stores, MongoDB, Cassandra, HBASE, Neo4j use and deployment, Application, RDBMS approach, Challenges NoSQL approach, Key-Value and Document Data Models, Column-Family Stores, Aggregate-Oriented Databases Replication and sharding, MapReduce on databases. Distribution Models, Single Server, Sharding, Master-Slave Replication, Peer-to-Peer Replication, Combining Sharding and Replication	4
8	NoSQL Key/Value databases using MongoDB, Document Databases, What Features, Consistency, Transactions, Availability, Query Features, Scaling	4
9	Column- oriented NoSQL databases using Apache HBASE, Column-oriented NoSQL databases using Apache Cassandra, Architecture of HBASE, Features, Consistency, Transactions, Availability, Query Features, Scaling	3
10	NoSQL Key/Value databases using Riak, Key-Value Databases, What Is a Key-Value Store, Key-Value Store Features, Consistency, Transactions, Query Features, Structure of Data	3
11	Graph NoSQL databases using Neo4J ,NoSQL database development tools and programming languages, Graph Databases	3

Self-Study Topics

Hadoop Incubator Projects

Explore projects available on <http://incubator.apache.org/projects/>

Practical content

- The labs are to be designed so that the through understating of various types of regression modeling is carried out by students.
- The labs will emphasize on practical aspects of clustering, rule mining, hurdles dusting big data analytics.
- The lab will focus on active learning of practical concepts so as to achieve the course outcome with effectiveness.

Practical List

- Set up Hadoop and its development environment
- Execute various HDFS commands
- Implement Read and Write Operations in HDFS
- Count the number of words in a text file and print the output
- For each word from a set of files list all the file names that the word is a part of.
- Introduction to MongoDB and its Installation on Windows & Linux
- Description of mongo Shell, Create database and show database
- Commands for MongoDB and To study operations in MongoDB – Insert, Query, Update, Delete and Projection
- Where Clause equivalent in MongoDB
- To study operations in MongoDB – AND in MongoDB, OR in MongoDB, Limit Records and Sort Records. To study operations in MongoDB – Indexing, Advanced Indexing, Aggregation and Map Reduce.
- Practice with ' macdonalds ' collection data for document oriented database. Import restaurants collection and apply some queries to get specified output.
- Column oriented databases study, queries and practices

Text Books	
1	AnandRajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 2011. Ron Bekkerman, Mikhail Bilenko and John Langford, Scaling up Machine Learning: Parallel and Distributed Approaches, Cambridge University Press, 2011.
Reference Books	
1	Tom White, Hadoop: The Definitive Guide, O'Reilly Media, Third Edition, 2012.
2	Bill Franks, Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Wiley, 2012.
3	Michael Minelli, Michele Chambers, and AmbigaDhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley, 2013.
4	Frank J. Ohlhorst, Big Data Analytics: Turning Big Data into Big Money, Wiley, 2012.
5	ArvindSathi, Big Data Analytics: Disruptive Technologies for Changing the Game, MC Press, 2012