B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

August 2022 Semester End Main Examinations

Programme: B.E   Semester: VI
Branch: Information Science and Engineering   Duration: 3 hrs.
Course Code: 20IS6PEBDA   Max Marks: 100
Course: Big Data Analytics   Date: 22.08.2022

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

UNIT - I

1	a)	Comment of the classification of digital data	05
	b)	Mention different techniques on how to deal with unstructured data	05
	c)	Give the anatomy of file read and write in Hadoop environment.	10

UNIT - II

2	a)	What is HDFS? List all the components of HDFS and explain any four components.	10
	b)	Describe the various features of Hadoop YARN administration. Compare Hadoop1.X and YARN	10

OR

3	a)	Explain with neat diagrams the MapReduce Programming workflow and its data analysis task	10
	b)	With suitable example, explain any five HDFS commands.	10

UNIT - III

4	a)	Cassandra adheres to availability and partition tolerant as stated by CAP theorem. Justify	10
	b)	Explain the purpose of using MemTables and SSTables in Cassandra	10

OR

5	a)	With a neat diagram explain the hive architecture	10
	b)	With examples explain the file formats supported in Hive	10

UNIT - IV

6	a)	With neat diagram explain the anatomy of Spark Job Run,Job Submission and DAG construction.	10
	b)	List the steps required to import data from various data store/databases using sqoop script	10

UNIT - V

7 a) With neat diagram explain the working of ZooKeeper in Hadoo   **10**

    b) Explain the anatomy of read and write in ZooKeeper   **10**

******