

GUJARAT TECHNOLOGICAL UNIVERSITY**BE – SEMESTER- VII EXAMINATION-SUMMER 2023****Subject Code: 2171607****Date: 27/06/2023****Subject Name: Big Data Analytics****Time: 10:30 AM TO 01:00 PM****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

		MARKS
Q.1	(a) Define big data. Mention few applications of big data analytics.	03
	(b) Explain 4 'v' characteristics of big data with suitable example.	04
	(c) Explain following important daemons of HDFS and MapReduce: DataNode, NameNode and Secondary NameNode, JobTracker and TaskTracker	07
Q.2	(a) What is Hadoop eco-system? Mention the components of Hadoop eco-system.	03
	(b) Differentiate RDBMS and Hadoop.	04
	(c) Elaborate on following characteristics of Hadoop: Distributed, fault-tolerant, massive-storage, faster-processing, scalability	07
OR		
Q.3	(c) What is NoSQL? Explain types of NoSQL databases in brief.	07
	(a) Mention the uses of NoSQL databases.	03
	(b) What is Hive? Briefly explain data units supported by Hive.	04
	(c) Explain following phases of Map and Reduce in brief: Map, Combiner, partitioner, shuffle and sort, Reduce	07
OR		
Q.3	(a) Explain use of data serialization in hadoop.	03
	(b) Differentiate followings:	04
	(i) NameNode vs. DataNode (ii) Hive vs. RDBMS	
	(c) What is Hive? Explain its architecture and its working with suitable diagrams.	07
Q.4	(a) What is NewSQL? How it differs from NoSQL?	03
	(b) What is Zookeeper? How it helps in monitoring a cluster?	04
	(c) What is the importance of HBase? Explain the data model supported by HBase. Also differentiate Row-oriented database vs. Column-oriented database.	07
OR		
Q.4	(a) How Spark is faster than MapReduce? Explain in brief.	03
	(b) What is MongoDB? Explain the important features of MongoDB.	04
	(c) Explain working of MapReduce with suitable example.	07
Q.5	(a) What is Pig? How it is different from MapReduce?	03
	(b) Why do we need Pig? Explain the architecture of Pig.	04
	(c) Explain Resilient Distributed Dataset with reference to Spark.	07

OR

- Q.5** (a) Draw architecture of spark and label it's components. **03**
(b) Elaborate basic CRUD operations in MongoDB with suitable example of each. **04**
(c) Explain the workflow of Apache Zookeeper with suitable diagram. **07**
