GUJARAT TECHNOLOGICAL UNIVERSITY
MCA - SEMESTER– IV• EXAMINATION – WINTER 2020

Subject Code:4649306   **Date:05/01/2021**
Subject Name:Big Data Tools
Time:02:00 PM to 04:00 PM   **Total Marks: 56**
Instructions:
1. **Attempt any FOUR questions out of EIGHT questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1	**(a)**	Define terms
		    1. JobTracker
		    2. RDD
		    3. SerDe
		    4. RCFile
		    5. NewSQL
		    6. YARN
		    7. Piggy Bank
			07

	(b)	How does Hadoop support Big Data? Explain Hadoop ecosystem with suitable diagram.	**07**
Q.2	**(a)**	Explain different types of NoSQL databases and also differentiate Relational Vs NoSQL databases.	**07**
	(b)	What is Big Data? Explain Big Data in terms of Volume, Velocity and Variety.	**07**
Q.3	**(a)**	Illustrate and explain phases of MapReduce.	**07**
	(b)	Explain CRUD operations in MongoDB with an example.	**07**
Q.4	**(a)**	What is the role of HDFS in Hadoop cluster? Illustrate and discuss HDFS architecture.	**07**
	(b)	Explain Replication and Sharding in MongoDB with suitable diagram.	**07**
Q.5	**(a)**	How does HIVE support Big Data? Draw and discuss HIVE architecture.	**07**
	(b)	Explain execution modes and data types in Pig.	**07**
Q.6	**(a)**	What is HiveQL? Explain DDL and DML commands in HIVE with proper example.	**07**
	(b)	How Spark uses Hadoop? Explain features of Spark and MLlib in brief.	**07**
Q.7	**(a)**	How Pig is useful in Big Data analytics? Explain relational operators in Pig.	**07**
	(b)	Illustrate and explain components of Spark in detail.	**07**
Q.8	**(a)**	Write a short note on classification and characteristics of data.	**07**
	(b)	Write a program to demonstrate the job of MapReduce.	**07**

****