GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VII (OLD) EXAMINATION – WINTER 2023

Subject Code:2171607   **Date:12-12-2023**
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)	Explain five P's of Big Data in brief.	03
	(b)	Explain job scheduling of fair scheduler in Map Reduce.	04
	(c)	1) Define the following terms: a) Name Node b) Data Node c) Heartbeat	03
		2) Explain HDFS operations in detail.	04
Q.2	(a)	Explain Big Data in terms of Volume, Velocity and Veracity.	03
	(b)	Enlist various applications of Big Data. How it can be used to improve the productivity in **agriculture**.	04
	(c)	Explain Hadoop components with diagram.	07
		OR	
	(c)	1) Define data serialization.	01
		2) With proper examples discuss and differentiate structured, unstructured and semi-structured data.	04
		3) Make a note on how type of data affects data serialization.	02
Q.3	(a)	Explain storage mechanism in HBase.	03
	(b)	Differentiate: Apache pig Vs Map Reduce.	04
	(c)	Justify "Spark is faster than MapReduce".	07
		OR	
Q.3	(a)	Define Zookeeper. Enlist and discuss the benefits of it.	03
	(b)	Explain the architecture of HIVE. List out the features of HIVE.	04
	(c)	What is RDD? Explain about transformations and actions in the context of RDDs. State and explain RDD operations in brief.	07
Q.4	(a)	Discuss Machine Learning with **MLlib** in SPARK.	03
	(b)	Explain SPARK unified stack.	04
	(c)	Differentiate SQL and NoSQL. Enlist the industry applications of NoSQL.	07
		OR	
Q.4	(a)	Explain sharding process of MongoDB.	03
	(b)	Explain job scheduling of capacity scheduler in Map Reduce.	04
	(c)	What is NoSQL? List out the features of NoSQL. Explain types of NoSQL databases in brief.	07
Q.5	(a)	Define: Term Frequency and Inverse Document Frequency.	03
	(b)	Explain metastore in Hive.	04
	(c)	Explain following for MongoDB. 1) Indexing 2) Aggregation	07

OR

Q.5 **(a)** Explain scaling feature of MongoDB. **03**

(b) Explain following in brief with respect to Mongo DB : **04**
1) Collections and documents
2) Indexing and retrieval

(c) Explain CRUD operations with suitable example in MongoDB. **07**

****