Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VI (NEW) EXAMINATION – SUMMER 2024

Subject Code: 3161607 Date:22-05-2024

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.**

Q.1	**(a)**	1) Define the following terms: a) Big Data b) Machine Learning c) Heartbeat	**03**
	(b)	Explain 4Vs property of Big Data.	**04**
	(c)	a) Discuss the challenges of convectional system. b) Traditional Vs Big Data Business approach.	**03** **04**
Q.2	**(a)**	Distinguish Horizontal and Vertical scaling with suitable example.	**03**
	(b)	Explain HDFS operations in detail.	**04**
	(c)	Discuss the big data case study of "Walmart"	**07**
		OR	
	(c)	Discuss the big data case study of "Uber"	**07**
Q.3	**(a)**	Write the steps to setup the Hadoop Cluster.	**03**
	(b)	Differentiate: Apache pig Vs Map Reduce.	**04**
	(c)	Explain Hadoop components with diagram.	**07**
		OR	
Q.3	**(a)**	Define Zookeeper. Enlist and discuss the benefits of it.	**03**
	(b)	Explain SPARK unified stack.	**04**
	(c)	Justify "Spark is faster than MapReduce".	**07**
Q.4	**(a)**	Explain streaming data architecture.	**03**
	(b)	List out the features of HIVE. Explain the architecture of HIVE.	**04**
	(c)	What is NoSQL? List out the features of NoSQL. Explain types of NoSQL databases in brief.	**07**
		OR	
Q.4	**(a)**	Enlist the benefits of using streaming data architecture.	**03**
	(b)	Explain job scheduling of capacity scheduler in Map Reduce.	**04**
	(c)	Differentiate SQL and NoSQL. Enlist the industry applications of NoSQL.	**07**
Q.5	**(a)**	Discuss Machine Learning with **MLlib** in SPARK.	**03**
	(b)	Explain the terms: a) Stream Data b) RTAP c) In-memory computing d) Block report	**04**
	(c)	Explain Map-Reduce operation for "word count" problem.	**07**
		OR	
Q.5	**(a)**	Define graph analytics. Enlist and explain types of graph analytics.	**03**
	(b)	Explain metastore in Hive.	**04**
	(c)	Explain real time "stock market prediction" using streaming data mining.	**07**

****