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

Subject Code:3170722   **Date:06-12-2023**
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 various challenges of conventional system.	03
	(b)	What is big data? Explain four V's of big data.	04
	(c)	Discuss structured, unstructured and semi-structured big data types. Give proper examples.	07
Q.2	(a)	Explain name node, data node and secondary node.	03
	(b)	Explain NoSQL business drivers.	04
	(c)	Write a short note on Hadoop ecosystem.	07

OR

	(c)	Explain working of following phases of Map Reduce with one common example. (i) Map Phase (ii) Combiner Phase (iii) Shuffle and Sort Phase (iv) Reducer Phase	07
Q.3	(a)	Explain following commands of HDFS with syntax and at least one example of each. (i) copyFromLocal (ii) mv (iii) cat	03
	(b)	Write basic wordcount program using PySPARK.	04
	(c)	What is transformation and actions in Apache Spark? Discuss various commands available for this activity in Apache Spark?	07

OR

Q.3	(a)	Discuss various benefits of stream processing.	03
	(b)	What is RDD? Explain role of RDD in Spark.	04
	(c)	Describe different types of NoSQL databases with the help of example.	07
Q.4	(a)	Write differences between NoSQL and SQL.	03
	(b)	Define Hadoop. List and explain advantages of Hadoop.	04
	(c)	Explain Bloom's Filter with the help of example.	07

OR

Q.4	(a)	Explain NoSQL base property.	03
	(b)	Differentiate between Pig vs Hive.	04
	(c)	Discuss Flajolet-Martin Algorithm to count Distinct elements in a stream with proper example.	07
Q.5	(a)	Describe important features of Apache Spark.	03
	(b)	Write a short note: Stock Market Predictions	04
	(c)	Explain working of Hive with proper steps and diagram.	07

OR

Q.5	(a)	Explain any three HiveQL DDL command with its syntax and example.	03
	(b)	How samples are collected from data stream?	04
	(c)	Discuss four ways that NoSQL system handles big data problems.	07

****