

Seat No.: \_\_\_\_\_

Enrolment No. \_\_\_\_\_

## **GUJARAT TECHNOLOGICAL UNIVERSITY**

**BE - SEMESTER-VI (NEW) EXAMINATION – SUMMER 2023**

**Subject Code:3161607**

**Date:12-07-2023**

**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
<b>Q.1</b>	(a) Enlist and explain the Big Data characteristics in detail.	<b>03</b>
	(b) Enlist and explain the types of Big Data.	<b>04</b>
	(c) Differentiate Traditional Vs. Big Data business approach.	<b>07</b>
<b>Q.2</b>	(a) List out 4 V's and Explain in detail.	<b>03</b>
	(b) Write a MapReduce code for Word Count.	<b>04</b>
	(c) Explain Map-Reduce framework in detail. Draw the architectural diagram for Physical Organization of Computer Nodes.	<b>07</b>
<b>OR</b>		
	(c) Define HDFS. Discuss the HDFS architecture in detail.	<b>07</b>
<b>Q.3</b>	(a) Enlist the steps to set up the Hadoop cluster.	<b>03</b>
	(b) Explain the terms below: i) Scaling out ii) Hadoop stream iii) Big data iv) Task tracker	<b>04</b>
	(c) Explain working of Hive with proper steps and diagram.	<b>07</b>
<b>OR</b>		
<b>Q.3</b>	(a) Explain Metastore in Hive.	<b>03</b>
	(b) Differentiate: Apache pig Vs Map Reduce.	<b>04</b>
	(c) What are the components of Spark? Also state the features of Spark.	<b>07</b>
<b>Q.4</b>	(a) What is Zookeeper? List the benefits of it.	<b>03</b>
	(b) Enlist NoSQL data architecture pattern. Explain column family (Bigtable) stores in detail.	<b>04</b>
	(c) i) List the differences between NoSQL and relational databases.	<b>03</b>
	ii) Explain the distribution models master-slave vs. peer-to-peer.	<b>04</b>
<b>OR</b>		
<b>Q.4</b>	(a) Explain the following commands of HDFS: i) copyFromLocal ii) setrep iii) checksum	<b>03</b>
	(b) Explain the data processing operators in PIG.	<b>04</b>
	(c) Enlist the benefits of data stream processing.	<b>07</b>

- Q.5** (a) Explain in detail the data stream processing. **03**  
(b) Draw and discuss the Architecture of Hive. **04**  
(c) What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset? **07**

**OR**

- Q.5** (a) Write short note on Graph analytics. **03**  
(b) Write short note on Real-time sentiment analysis. **04**  
(c) Discuss the concepts of regions in HBase and storing Big Data with HBase. **07**

\*\*\*\*\*