

GUJARAT TECHNOLOGICAL UNIVERSITY

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

Subject Code:3161607

Date:02-12-2024

Subject Name:Big Data Analytics

Time:02:30 PM TO 05:00 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) Define Big Data. Explain various types of Big Data.	03
	(b) Explain the terms in Hadoop. i) Scaling out ii) Hadoop Streaming	04
	(c) i) Enlist and explain Big Data characteristics. ii) Discuss two scheduling policies of YARN.	03 04
Q.2	(a) Explain Hadoop in the cloud.	03
	(b) Enlist and explain four ways that NoSQL systems handle big data problems.	04
	(c) Define HDFS. Draw HDFS architecture and explain its components.	07
	OR	
	(c) Draw the architectural diagram for Physical Organization of Computer Nodes. Explain Map-Reduce framework in detail.	07
Q.3	(a) Enlist the difference between NoSQL and RDBMS.	03
	(b) Explain Key-Value Stores and Column Family (Bigtable) stores as NoSQL data architecture patterns.	04
	(c) Explain the followings in details. i) Decaying Window ii) Define RTAP. Describe RTAP applications.	07
	OR	
Q.3	(a) Enlist the difference between master-slave and peer-to-peer distribution models.	03
	(b) Explain Graph Stores and Document stores as NoSQL data architecture patterns.	04
	(c) Explain Stream data model and its architecture in detail with a neat diagram.	07
Q.4	(a) Explain interactive Spark with PySpark in detail.	03
	(b) Explain HBase in detail.	04
	(c) Explain real time “stock market prediction” using streaming data mining.	07
	OR	
Q.4	(a) How Implicit filtering is differing from explicit filtering.	03
	(b) Enlist and explain the key features of pig.	04
	(c) Explain real time “Sentiment Analysis” using streaming data mining.	07

- Q.5** (a) Explain the following commands of HDFS: **03**
i) copyFromLocal ii) setrep iii) checksum
(b) Explain Spark components in detail. **04**
(c) Write short note on Zookeeper. **07**

OR

- Q.5** (a) Explain the reasons why Spark is more suitable for streaming data analytics. **03**
(b) Discuss Machine Learning with MLlib in Spark. **04**
(c) Explain working of Hive with proper steps and diagram. **07**
