

GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER-VII (NEW) EXAMINATION – WINTER 2022

Subject Code:2171607**Date:03-01-2023****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.

- Q.1** (a) 1) Define the following terms: **03**
a) Big Data b) Machine Learning c) Heartbeat
- (b) Explain 4Vs property of Big Data. **04**
- (c) a) Explain five P's of Big Data in brief. **03**
b) Explain job scheduling of fair scheduler in Map Reduce. **04**
- Q.2** (a) Explain HDFS operations in detail. **03**
(b) Enlist various applications of Big Data. How it can be used in weather forecasting. **04**
(c) Define Distributed file system. Enlist and explain features of DFS. **07**
- OR**
- (c) a) Define Schema. **01**
b) Explain the importance of Big Data. **04**
c) 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) 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) Define: Term Frequency and Inverse Document Frequency. **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) Explain sharding process of MongoDB. **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 scaling feature of MongoDB. **04**
(c) Explain following for MongoDB. **07**
1) Indexing
2) Aggregation

OR

- Q.5** (a) Explain following in brief with respect to Mongo DB : **03**
1) Collections and documents
2) Indexing and retrieval
- (b) Explain metastore in Hive. **04**
- (c) Explain CRUD operations with suitable example in MongoDB. **07**
