

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA INTEGRATED– SEMESTER VIII- EXAMINATION –WINTER-2023

Subject Code:2688601**Date: 02/12/2023****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. Use of simple calculators and non-programmable scientific calculators are permitted.

- Q.1 (a) Answer the following: 03**
 1. List the applications of big data.
 2. Define NewSQL.
 3. Define RDD.
- (b) Discuss the key drivers of big data. 04**
(c) Discuss hadoop ecosystem in brief. 07
- Q.2 (a) Differentiate following: 07**
 1. SQL vs NoSQL 2. Pig vs Hive
- (b) What is distributed file system? Discuss architecture of Hadoop Distributed File System. 07**
- OR**
- (b) Discuss the classification of big data. 07**
- Q.3 (a) What is MapReduce? Explain Mapper and Reducer phases in brief. 07**
(b) Discuss hive architecture in detail. 07
- OR**
- Q.3 (a) Define HQL. Discuss DDL and DML statements with reference to HQL. 07**
(b) Write and discuss the process of Extract, Transform and Load in PIG. 07
- Q.4 (a) Write short note on HBase. 07**
(b) What is apache spark? List and explain apache spark components. 07
- OR**
- Q.4 (a) List and explain types of NoSQL databases. 07**
(b) What is zookeeper? Explain characteristics of zookeeper. 07
- Q.5 (a) Discuss key features of MongoDB. 07**
(b) What is NOSQL? List the advantages and uses of NoSQL. 07
- OR**
- Q.5 (a) Explain following terms with reference to MongoDB. 07**
 1. Document 2. Collection 3. Object Id
(b) Explain aggregate function in MongoDB with suitable example. 07
