

GUJARAT TECHNOLOGICAL UNIVERSITY**MCA - SEMESTER– IV EXAMINATION – WINTER 2019****Subject Code: 4649306****Date: 18-11-19****Subject Name: Big Data Tools****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.

- Q.1 (a)** Explain Following. **07**
1. List out Diagnostic Operator
 2. REGEX_EXTRACT_ALL
 3. COGROUP
 4. SQL Vs. HQL
 5. _____Default Delimiter of LOAD
 6. Example of DBPROPERTIES()
 7. _____Default Block Size of HDFS in Hadoop 2.x
- (b)** Explain 4V's of Big Data with Example. **07**
- Q.2 (a)** What is Hadoop? Describe Hadoop Eco System with details. **07**
- (b)** Explain Replication in Hadoop with Example. **07**
- OR**
- (b)** Explain How Hadoop 1.x is different from Hadoop 2.x in details. **07**
- Q.3 (a)** Explain YARN architecture. **07**
- (b)** What is NoSQL? Explain MongoDB Sharding with figure. **07**
- OR**
- Q.3 (a)** What is MapReduce? Explain its Architecture. **07**
- (b)** Write HDFS command for following Query. **07**
1. Create a file "Sample" in a local file system and export it to the HDFS File System
 2. Write the HDFS command for copying a "Sample" file from HDFS to local File System.
 3. Write HDFS command to display complete list of directories and files of HDFS
 4. Write HDFS command for copying an existing "Sample" file in a "Test" HDFS directory to some another HDFS directory

- Q.4 (a)** Write a Query for Implementation of Word Count Example in Pig and HIVE. **07**
(b) How to implement piggybank.jar in pig? Explain with Example. **07**

OR

- Q.4 (a)** Explain Complex Data Type in Pig with Example. **07**
(b) What is HIVE? Explain its Architecture. **07**

- Q.5 (a)** Explain HIVE Vs. PIG Vs. RDBMS **07**
(b) Create a StudentMaster database with a collection called “Student” containing documents with some or all of the following fields: StudentRollNo, StudentName, Grade, Hobbies, and DOJ. Perform the following operations on the database: **07**

1. Retrieve Student Name and Grade of student who is having _id column is 1
2. Find those documents where the Hobbies is set neither to ‘Chess’ nor is set to ‘Dancing’.
3. Find total the number of documents where Grade is ‘VII’.
4. Find all documents in proper format. (Without _Id field)
5. Find those documents where the Grade is not set to ‘VII’.

OR

- Q.5 (a)** What is Spark? Explain RDD in details. **07**
(b) Write a pig script to load and store “Student data”.(Student file contain Roll no, Name, Marks and GPA) and perform following query in Pig **07**

1. Partition a relation based on the GPA’s acquired by students
2. Join two relation namely Student and department (Rno,DeptNo,DeptName) based on the values contain in the roll no column
3. To calculate the average marks for each student
4. Filter all the students who are having GPA>5
5. Remove duplicates tuple of Student list
