Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA – SEMESTER-IV EXAMINATION –SUMMER-2020

Subject Code:4649306 Date:07-11-2020

Subject Name:Big Data Tools (BDT)

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.**

Que. 1 Do as directed: [1 mark for each correct answer] 14

1. Match the followings:

1. NLP	a. Uses the methods at the interaction of AI & DBs
2. Text Analysis	b. Comprehend Human or natural language input.
3. UIMA	c. Text Mining
4. Data Mining	d. Content Analysis

2. Data collected using CCTV footage will consider under _____ category.
3. _____ is the large data repository which stores the data in its native format until it is needed.
4. Big data is the High Volume, High Velocity and High Verity information assets that demands _____, _____ form of information processing for enhanced _____ and _____.
5. In _____ architecture multiple processors have their own private memory.
6. Hadoop is _____ based flat architecture.
7. Name Node use _____ to store file system namespace.
8. YARN is responsible for _____.
9. Global Resource Manager distribute _____ among applications.
10. HDFS has _____ / _____ architecture.
11. HDFS is built using _____ language.
12. A _____ contain the list of all blocks on a Data Node.
13. MogoDB documents are represented as _____.
14. Database stores data in MogoDB is called _____.

Que. 2 (a) What is Hadoop? Explain Features and Key Advantages of Hadoop. 07

Que. 2 (b) Explain the objective and syntaxes of MONGOIMPORT and MONGOEXPORT commands. Also write proper commands for performing following tasks: 07

1. Write a command to import the data of Student.csv which is stored at “D:/MongoData” in to the STUDENT collection of STUDENTDATA database.
2. Write a command to export the data of EMPLOYEE collection which is created in EMPLOYEEDATA database in to the text file and save it at “D:/MongoData/Employee”.

OR

Que. 2 (b)	Write a complete syntax for aggregate in MongoDB. Write a MongoDB query of aggregate to get following output. Collection ORDER (CustID, ItemID, OrderDate, DispatchDate, OrderQty) 1. Display total number of quantities ordered by each customer for any items. 2. Display total number of quantities ordered by each customer for any items on 8/29/2018 order by CID. 3. Count Item wise total number or quantities ordered by each customer on 8/29/2019.	07
Que. 3 (a)	What is an Unstructured Data? Explain different sources for acquiring Unstructured Data also explain the different ways of dealing with it.	07
Que. 3 (b)	List out and give an overview of Hadoop ecosystems.	07
	OR	
Que. 3 (a)	What is MapReduce? Explain major phases of MapReduce framework.	07
Que. 3 (b)	What is NoSQL? Explain the advantages of NoSQL.	07
Que. 4 (a)	Explain CAP theorem.	07
Que. 4 (b)	Differentiate following things: 1. HADOOP v/s SQL 2. PIG v/s HIVE	04 03
	OR	
Que. 4 (a)	What is mrjob? Explain advantages of mrjob for writing MapReduce applications.	07
Que. 4 (b)	Differentiate following things: 1. BI v/s Big Data 2. MongoDB v/s RDBMS	04 03
Que. 5 (a)	Write a Java code of User Define Function in Hive which convert value of fields into Upper case. Write a command to convert this Java Program into a JAR. Also write a CREATE statement to register this UDF into the HIVE.	07
Que. 5 (b)	What is RDD? How to creates RDD in python explain it with example.	07
	OR	
Que. 5 (a)	Write an example of word count using PIG for the content written in line.txt.	07
Que. 5 (b)	List out all RDD operations. Explain Transformation operation in detail.	07

****