| U.S.N. | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|

# B.M.S. College of Engineering, Bengaluru-560019

**Autonomous Institute Affiliated to VTU**

## August 2022 Semester End Main Examinations

**Programme: B.E.** **Semester: VI**
**Branch: Computer Science and Engineering** **Duration: 3 hrs.**
**Course Code: 20CS6PEBDA** **Max Marks: 100**
**Course: Big Data Analytics** **Date: 22.08.2022**

**Instructions:** 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

**Important Note:** Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.



### UNIT - I

1 a) Define Big data. Explain 3V's of Big Data. **5**

b) Identify the following techniques/Systems: **5**

i) The technology that helps to query that resides in a computer's random-access memory (RAM) rather than data stored on physical disks.

ii) A single common main memory that is shared by two or more identical processors.

iii) A database System that is loosely coupled and composed by individual machines which can run their individual application.

iv) A coordinated processing of a program by multiple processors, each working on different parts of the program and using its own operating system and memory.

v) A database System that is tightly coupled which has access to a common memory.

c) Analyse the below scenarios and identify which among the CAP theorem suits for each scenario: **6**

i) Imagine there is a very popular mobile operator in your city and you are its customer because of the amazing plans it offers. Besides that, they also provide an amazing customer care service where you can call anytime and get your queries and concerns answered quickly and efficiently. Whenever a customer calls them, the mobile operator is able to connect them to one of their customer care operators. The customer is able to elicit any information required by her/him about his accounts like balance, usage, or other information.

ii) Now, you have recently shifted to a new house in the city and you want to update your address registered with the mobile operator. You decide to call the customer care operator and update it with them. When you call, you connect with an operator. This operator makes the relevant changes in the system. But once you have put down the phone, you realize you told them the correct street name but the old house number (old habits die hard!).

So you frantically call the customer care again. This time when you call, you connect with a different customer care operator but they are able to access your records as well and know that you have recently updated your address. They make the relevant changes in the house number and the rest of the address is the same as the one you told the last operator.

iii) Recently you have noticed that your current mobile plan does not suit you. You do not access that much mobile data any longer because you have good wi-fi facilities at home and at the office, and you hardly step outside anywhere. Therefore, you want to update your mobile plan. So you decide to call the customer care once again.

On connecting with the operator this time, they tell you that they have not been able to update their records due to some issues. So the information lying with the operator might not be up to date, therefore they cannot update the information.

d) List and explain the types of NoSQL databases with an example for each. **4**

## UNIT - II

2 a) How Consistency can be achieved in Cassandra? Explain. **6**

b) Create a keyspace "XYZ" having "PAGE_VIEW_COUNT" as a column family with columns "No_OF_VIEWS", "URL_NAME", "PAGE_NAME". Use Counters to increment the "No_OF_VIEWS" when the URL is viewed. Add a record for PAGE_VIEW_COUNT column family. **6**

c) Using CQL create a database for Public_Health_Centre: **8**
i) Create a database for Public_Health_Centre and Create a Public_Health_Centre(_idPHC,Doctor_Name, Highest_Qualification, Patient_id, ailment, visit_date,medicine_prescribed) Collection.
ii) Insert required documents to the collection.
iii) Display Doctor_name and Patient_id whose visited PHC on "20/10/2017" in desc order.
iv) Add new field "Location" with Value "Shivagiri" to the document (_id:2) of "Public_Health_Centre" Document.

## UNIT - III

3 a) Illustrate the working of HDFS daemons. **6**

b) Write a MapReduce program to sort the data of a Student by Student name. **10**

| Input Data |
|---|
| 1001, John, 45 |
| 1002,Jack,39 |
| 1003, Alex,44 |
| 1004,Smith,38 |

c) Write the HDFS commands for the following: **4**
i) To get the list of complete directories and files of HDFS
ii) To remove a directory from HDFS
ii) To copy a file from HDFS to local file system using get
iv) To display the contents of a HDFS file on console

**OR**

4 a) With a neat diagram explain the steps involved in YARN architecture. **6**

b) **10**

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Transaction_date | Product | Price | Payment_ | Name | City | State | Country | Account_Created | Last_Login | Latitude | Longitude |
| 2 | 01-02-2009 06:17 | Product1 | 1200 | Mastercar | carolina | Basildon | England | United Kir | 01-02-2009 06:00 | 01-02-2009 06:08 | 51.5 | -1.11667 |
| 3 | 01-02-2009 04:53 | Product1 | 1200 | Visa | Betina | Parkville | MO | United Sta | 01-02-2009 04:42 | 01-02-2009 07:49 | 39.195 | -94.6819 |
| 4 | 01-02-2009 13:08 | Product1 | 1200 | Mastercar | Federica e | Astoria | OR | United Sta | 01-01-2009 16:21 | 01-03-2009 12:32 | 46.18806 | -123.83 |
| 5 | 01-03-2009 14:44 | Product1 | 1200 | Visa | Gouya | Echuca | Victoria | Australia | 9/25/05 21:13 | 01-03-2009 14:22 | -36.1333 | 144.75 |
| 6 | 01-04-2009 12:56 | Product2 | 3600 | Visa | Gerd W | Cahaba He | AL | United Sta | 11/15/08 15:47 | 01-04-2009 12:45 | 33.52056 | -86.8025 |
| 7 | 01-04-2009 13:19 | Product1 | 1200 | Visa | LAURENCE | Mickleton | NJ | United Sta | 9/24/08 15:19 | 01-04-2009 13:04 | 39.79 | -75.2381 |
| 8 | 01-04-2009 20:11 | Product1 | 1200 | Mastercar | Fleur | Peoria | IL | United Sta | 01-03-2009 09:38 | 01-04-2009 19:45 | 40.69361 | -89.5889 |
| 9 | 01-02-2009 20:09 | Product1 | 1200 | Mastercar | adam | Martin | TN | United Sta | 01-02-2009 17:43 | 01-04-2009 20:01 | 36.34333 | -88.8503 |
| 10 | 01-04-2009 13:17 | Product1 | 1200 | Mastercar | Renee Elis | Tel Aviv | Tel Aviv | Israel | 01-04-2009 13:03 | 01-04-2009 22:10 | 32.06667 | 34.76667 |

Consider the above file SalesJan2009.csv. Write a MapReduce program to find out number of products sold in each country.

c) Differentiate between RDBMS and Hadoop. **4**

**UNIT - IV**

5 a) Illustrate the Spark components with a neat diagram. **8**

b) Write a code in Python to load the data into Parquet and save the contents of a Schema RDD to Parquet. **6**

c) Write a code in Python for splitting lines into multiple words in Java. **6**

**OR**

6 a) Define Transformations in Spark. Explain filter() and union() transformations with an example for each. **6**

b) Write a Scala program to perform the following: **5**

i) Create an arraybuffer with 5 subjects names.
ii) Access the 4th element from the arraybuffer.
iii) Add ("BDA","CNS") to the end of the array buffer.
iv) Sort the array buffer in ascending order.
v) Remove the last two elements from the arraybuffer.

c) Predict the output of the following program: **4**

```
object MapOperations
{
def main(args: Array[String])
{
 val student = Map(12 -> "Reena", 13 -> "Micheal", 14 -> "Peter")
val marks: Map[String, Int] = Map()
 println("Keys : " + student.keys)
 println("Values : " + student.values)
println("Check if student is empty : " + student.isEmpty)
println("Check if marks is empty : " + marks.isEmpty)
 }
}
```

d) Differentiate between map() and flatmap() with an example. **5**

## UNIT - V

7 a) Explain how decision tree is used in classification algorithms. **4**

b) Identify the below recommendation systems and explain the paradigms used in each model: **6**

i) If user A likes Apple, Banana, and Mango while user B likes Apple, Banana, and Jackfruit, they have similar interests. So, it is highly likely that A would like Jackfruit and B would enjoy Mango.

ii) If a user likes to watch movies such as Iron Man, the recommender system recommends movies of the superhero genre or films describing Tony Stark.

c) Explain how collaborative filtering can be employed in developing a movie recommender system. **10**

******