U.S.N.										

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

Autonomous Institute Affiliated to VTU

September / October 2023 Supplementary Examinations

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

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

UNIT - I

1 a) Differentiate between SQL and NoSQL. **6**

b) Demonstrate various key terminologies used in Big Data Analytics 3.0. **10**

c) Explain the characteristics of Big data. **4**

UNIT - II

2 a) Using CQL write queries for the following: **10**

i) Create a Keyspace Company and Create the Column Family Employee (Emp_no , Name, Salary, Department, Designation, Marital_status, Hobbies (Set)).
ii) Insert required row to the Column Family.
iii) Display Name and Department of employees whose designation is "Manager" and salary is greater than 85,000 in decreasing order.
iv) Add new field "Previous_experience" with Value "15" to the Emp_no 111 of "Employee" Column Family.
v) Import an existing csv file into the current column family.

b) Write Cassandra queries for the following: **10**

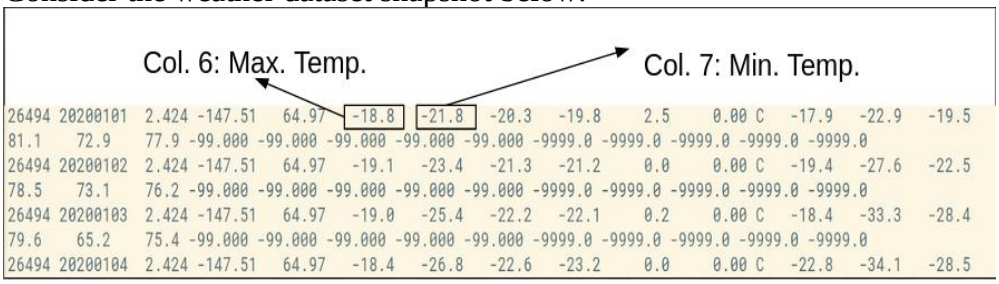
(i) Create a keyspace by name Employee. Create a column family by name Employee-Info with attributes Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name.
(ii) Insert the values into the table in batch.
(iii) Update Employee name whose Department is "financial".
(iv) Alter the schema of the table Employee_Info to add a column Projects which stores a set of Projects done by the corresponding Employee. Update the altered table to add project names.
(v) Display Emp_Name and the projects handled by the Employee.
(vi) Create a TTL of 30 seconds to display the values of Employees.

UNIT - III

3 a) What are the core components of Hadoop? Explain each of its components in brief. **6**

b) Explain Hadoop Distributed File System. **6**

c) Consider the weather dataset snapshot below. **8**




```
26494 20200101  2.424 -147.51   64.97   -18.8   -21.8   -20.3   -19.8     2.5     0.00 C   -17.9   -22.9   -19.5
81.1    72.9    77.9 -99.000 -99.000 -99.000 -99.000 -99.000 -9999.0 -9999.0 -9999.0 -9999.0 -9999.0
26494 20200102  2.424 -147.51   64.97   -19.1   -23.4   -21.3   -21.2     0.0     0.00 C   -19.4   -27.6   -22.5
78.5    73.1    76.2 -99.000 -99.000 -99.000 -99.000 -99.000 -9999.0 -9999.0 -9999.0 -9999.0 -9999.0
26494 20200103  2.424 -147.51   64.97   -19.0   -25.4   -22.2   -22.1     0.2     0.00 C   -18.4   -33.3   -28.4
79.6    65.2    75.4 -99.000 -99.000 -99.000 -99.000 -99.000 -9999.0 -9999.0 -9999.0 -9999.0 -9999.0
26494 20200104  2.424 -147.51   64.97   -18.4   -26.8   -22.6   -23.2     0.0     0.00 C   -22.8   -34.1   -28.5
```

Write a MapReduce program to find the hot and cold days assuming that if maximum-temperature >35^{o}c it is a hot day and if minimum-temperature <10^{o}c it is a cold day.

OR

4 a) Show the chores of Mapper, Combiner, Partitioner , Shuffle and Sort and Reducer for the Word count problem by considering the following words. **5**

BMS, College, of, Engineering, BMS, College, of, Architecture, BMS, Women, College.

b) Write down the steps on the request to Mapreduce and the types of process in Mapreduce. **10**

c) Explain Sorting technique in MapReduce with an example. **5**

UNIT - IV

5 a) Demonstrate the three main capabilities provided by Spark SQL with a neat diagram. **4**

b) Write Spark SQL queries for the following by making necessary imports and creating appropriate contexts. **6**

i) Create a SchemaRDD by loading a JSON file which contains Employee details.

ii) Select Employee name, Salary in the increasing order of salary of 10 employees.

iii) Access the first column from the SchemaRDD created above

c) Explain any five transformations and actions used on RDDs with example. **10**

OR

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

- Create an arraybuffer with 5 vehicle brand names.
- Access the 3rd element from the arraybuffer
- Add (“BMW”,”MG”) to the end of the arraybuffer.
- Sort the array buffer in ascending order.
- Remove last two elements from the arraybuffer.

b) Write major difference between Spark Cache() and Persist(). Write a Scala program to store dataset into memory and to remove dataset from memory. **5**

c) Discuss in detail about maps and tuples in Scala. Consider a map that stores the country and corresponding capital city. Write a Scala function def getCapital(country: String="India", cmap: Map[String, String]) that accepts the map and country name as input and returns the corresponding capital city as output if the country exists in the map. By default the function has to return the capital city of India. **10**

UNIT - V

7 a) Demonstrate a model for Recommendation Systems based on a utility matrix of preferences with a suitable example. **10**

b) Compare Content based filtering and Collaborative filtering. Explain Content Based filtering with an example problem. **10**

******