

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

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

Autonomous Institute Affiliated to VTU

September / October 2022 Supplementary Examinations

Programme: B.E

Branch: Information Science and Engineering

Course Code: 20IS6PEBDA

Course: Big Data Analytics

Semester: VI

Duration: 3 hrs.

Max Marks: 100

Date: 10.10.2022

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) Define Big data. Explain characteristics of Big Data. | 05 |
| | b) Identify the category of data for the following <ul style="list-style-type: none"> i. The receipt given to you at the petrol pump ii. The billing of the items purchased at the supermarket iii. The CCTV footage iv. The consumer complaints and feedback v. The word document | 05 |
| | c) Explain the techniques to interpret unstructured data. | 05 |
| | d) Exemplify challenges of Big Data. | 05 |

UNIT - II

- | | | |
|---|--|----|
| 2 | a) Differentiate between RDBMS and Hadoop. | 05 |
| | b) Describe the Hadoop Components with a neat diagram | 08 |
| | c) Give the anatomy of file write in Hadoop environment. | 07 |

OR

- | | | |
|---|---|----|
| 3 | a) Explicate the key aspects of Hadoop. | 05 |
| | b) Implement a mapreduce code to count the number of words in the given document. | 10 |
| | c) Sketch and explain the architecture of YARN. | 05 |

UNIT - III

- | | | |
|---|--|----|
| 4 | a) Perform the following DB operations using Cassandra. | 10 |
| | <ol style="list-style-type: none"> 1. Create a key space by name Employee 2. Create a column family by name Employee-Info with attributes Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name 3. Insert the values into the table in batch 4. Update Employee name and Department of Emp-ID 121 5. Sort the details of Employee records based on salary | |

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.

6. Alter the schema of the table Employee_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.

7. Update the altered table to add project names.

8 Create a TTL of 15 seconds to display the values of Employees

- b) Create a managed table and external table in Hive with name train having fields like id, name source, destination, ticket price with comments "train details" fields terminated by "\t" lines terminated by "\n" and stored in TEXT FILE. 05
- c) Create tables in hive for voter with appropriate attributes and to store it in textfile, sequence file, rcf file , avro, orc, parquet file formats. 05

OR

- 5 a) Perform the following DB operations using Cassandra. 10

- i. Create a key space by name Library
 - ii. Create a column family by name Library-Info with attributes Stud_Id Primary Key, Counter_value of type Counter, Stud_Name, Book-Name, Book-Id, Date_of_issue
 - iii. Insert the values into the table in batch
 - iv. Display the details of the table created and increase the value of the counter
 - v. Write a query to show that a student with id 112 has taken a book "BDA" 2 times.
 - vi. Export the created column to a csv file
 - vii. Import a given csv dataset from local file system into Cassandra column family
- b) i) Create a managed table and external table in Hive with table name doctor having fields such as id, name, fees and area of specialization with comments "doctor details" fields terminated by "\t" lines terminated by "\n" and stored in TEXT FILE. 10
- ii) Write hive commands to perform the following
 - change the name of the table
 - change the name of field from id to doc_id
 - change the data type of fees from float to double
 - drop the column areaofspecialization
 - add column experience

UNIT - IV

- 6 a) Explain different ways to create spark RDD with code snippets. 07
- b) Explain 4 key concepts in Anatomy of a spark job run. 08

- c) i) Write commands in sqoop to import data from MYSQL to HDFS specified directory. **05**

ii) sqoop import \
--connect jdbc:mysql://localhost/userdb \
--username root \
--table student_add \
--m 1 \
--where "city ='sec-bad'" \
--target-dir /wherequery.

Predict the output of the following command applied for student_add table with fields id, hno, street, city.

UNIT - V

- 7 a) Exemplify zookeeper architecture with a neat block diagram. **08**
- b) Describe benefits of zoo keeper. **07**
- c) Explain the limitations of flume. **05**

SUPPLEMENTARY EXAMS 2022

