U.S.N.										

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

Autonomous Institute Affiliated to VTU

September / October 2022 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 **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.

SUPPLEMENTARY EXAMS 2022

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 Analytics with its classification. List and explain the challenges associated with big data.		8
	b)	"Distributed data bases relaxes ACID properties". Analyze the properties supported by distributed systems with a scenario. Reflect databases that follow one of the three possible combinations with a neat diagram.		8
	c)	Differentiate Between SQL and NoSQL.		4

UNIT - II

2	a)	Demonstrate various features of Cassandra in detail.		10
	b)	Using CQL write queries for the following: i) Create a Keyspace Hospital and Create the Column Family Doctor(ID, Name, Reg_no, Salary, Department, Designation, Specializations, VisitingHospitals) assuming appropriate data type. ii) Insert required row to the Column Family. iii) Display Name and Department whose designation is "SeniorSurgeon" and salary is greater than 1,00,000 in decreasing order. iv) Create a table to add patient_name and disease .Insert values which will be valid for 30 days. v) Import an existing csv file into the current column family.		10

UNIT - III

3	a)	Explain Hadoop Architecture along with its working.		10
	b)	Consider a Product CSV file with following information- Product name, price, payment mode, city, country of client etc. Develop a Map Reduce program to find number of products sold in each country.		10

OR

4	a)	With a block diagram, discuss the framework that run under YARN.		6
	b)	Show the chores of Mapper, combine partitioner, shuffle and sort and reduce		6

for the word count problem by considering the following words

BMS, Medical, College, Women, College, College, of, Engineering, MS.

c) Consider a movie booking app through which a person is searching for the movie "Jack". Write a map reduce program demonstrating the searching process and explain the steps involved in running the application. **8**

UNIT - IV

5 a) Demonstrate the three main capabilities provided by Spark SQL with a neat diagram. **10**
Write Spark SQL queries for the following by making necessary imports and creating appropriate contexts.
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 Schema RDD created above

b) Describe the overview of Spark along with its limitations **10**

OR

6 a) Elaborate the types of operations supported by RDDs. **10**

b) Summarize the functions of Spark SQL. **04**

c) Differentiate between Data Science Task and Data Processing Applications with respect to Spark context. **06**

UNIT - V

7 a) Write short notes on : **10**
i) Content based Recommendation
ii) Collaborative Filtering System

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

******

SUPPLEMENTARY EXAMS 2022