PRESIDENCY UNIVERSITY
BENGALURU

School Of Computer Science and Engineering & Information Science

End-Term Examinations, Aug 2024

Odd Semester: 2023 - 24

Course Code: CSE5011

Course Name: Data Science with Cloud Computing

Department: CSE

Date: 13:08:2024

Time: 9:30 am to 12:30 pm

Max Marks: 100

Weightage: 50%

Instructions:

(i) Read the all questions carefully and answer accordingly.

(ii) Do not write any matter on the question paper other than roll number.

Q.No	Questions	Marks	CO	RBT
1	a. List any five key steps for making Data driven decisions	4	CO1	L1
	b. Infer the features of decision making by using Data driven approach and Probabilistic approach	6	CO1	L2
	c. Consider a start-up company named Duke Tech Solutions. It gets the project from one of its clients for building, and deploying the software. Since it is a start-up, it does not want to buy any resources for the development of software. So, it decides to take services from the cloud as rent. What are the different types of services needed by the company? Explain the services in brief.	10	CO1	L3
OR				
2	a. Recall the types of cloud and brief with suitable example	4	CO1	L1
	b. Discuss the types of Decisions	6	CO1	L2
	c. Outline the common challenges and Misconceptions about Data Driven Decision-Making	10	CO1	L3

Q.No	Questions	Marks	CO	RBT
3	a. How decision making impact the Business?	4	CO2	L1
	b. Classify the significant role of Data Engineer with Data Scientists	6	CO2	L2
	c. Compare and contrast Augmented Analytics and Business Intelligence	10	CO2	L3
OR				
4	a. Define Likelihood with an example	4	CO2	L1
	b. Identify the coherent principles required to build a Data Dashboards and brief its significant variants	6	CO2	L2
	c. Compare with the key principles of Likelihood and Probability with suitable example	10	CO2	L3

5	a. Relate the need for Data Pipeline	4	CO3	L1
	b. Summarize how Dataflow work with a neat sketch	6	CO3	L2
	c. Illustrate on Google cloud Pub/Sub with a neat sketch	10	CO3	L3

OR

6	a. Identify any five use cases of Dataflow model	4	CO3	L1
	b. Discuss on how Google cloud Dataflow works with a neat sketch	6	CO3	L2
	c. Compare and contrast Batch Processing and Stream Processing	10	CO3	L3

7	a. What is MapReduce?	4	CO4	L1
	b. Discuss the steps involved in Pig Architecture with a neat sketch	6	CO4	L2

	c. Assume you have large collection of text documents, and you want to count the frequency of each word across all documents. Use the mapper to process each document and write the pseudocode for the following:
	  a. Count the word frequency
	  b. Word count with stop words removal
		10	CO4	L3

OR

8	a. Mention any four benefits of MapReduce	4	CO4	L1
	b. Write a Map Reduce program that mines the weather data of your choice.	6	CO4	L2

	c. Write the Pig Latin scripts to predict the sales transactions by using sales dataset, and you want to calculate the total sales for each product by using the following operations.
	  a. Group operation
	  b. Sort operation
	  c. Join operation
	  d. Projection operation
	  e. Filter operation
		10	CO4	L3

9	a. Mention any five GCP services with proper illustration	4	CO1	L1
	b. Interpret the significance of cloud sim simulator	6	CO1	L2
	c. Companies have different types of employees, and some of them want to use their devices at work. Here comes the concept of Bring-Your-Own-Device (BYOD). However, what if someone having a Chromebook or MacBook intends to use Windows OS? Is it possible to deliver Windows OS on a Chromebook? Critize how DaaS services work in providing virtual desktops.	10	CO1	L3

OR

10	a. Mention the advantages of Probabilistic approach	4	CO2	L1
	b. Is DaaS in Cloud computing being persistent? Justify	6	CO2	L2
	c. Consider a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0,1,2, or 3. I am allowed to choose 4 balls at random from the bag with replacement. For each possible value of θ, Find the probability of the observed sample $(X1, X2, X3, X4) = (1, 0,1,1)$. For which value of θ is the probability of the observed sample is the largest.	10	CO2	L3