



**B.Tech - Odd Sem : End Semester Exam**  
**Academic Year:2020-2021**  
**18EC4072 - Data Sciences & Big-Data**  
**Set No: 4**

| Time: |   | Max.Marks: 100 |         |         |     |        |         |
|-------|---|----------------|---------|---------|-----|--------|---------|
| S.NO  | Answer All Questions  | Choice         | Options | Marks   | CO  | CO BTL | COI BTL |
| 1.    | Consider this example: Suppose the National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and fullsize cars. It collects a sample of three for each of the treatments (cars types). Using the hypothetical data provided below, test whether the mean pressure applied to the driver's head during a crash test is equal for each types of car. Use $\alpha = 5\%$ . Table ANOVA.1 Compact cars Midsize cars Full-size cars 643 469 484 655 427 456 702 525 402 | choice Q-2     |         | 10Marks | CO1 | 2      | 1       |
| 2.    | Suppose the average number of lions seen on a 1-day safari is 5. What is the probability that tourists will see fewer than four lions on the next 1-day safari?   |                |         | 10Marks | CO1 | 2      | 1       |
| 3.    | Write question either No.3 or Q.No .4   | choice Q-4     |         | 15Marks | CO1 | 2      | 2       |
| 3.A.  | Explain Big Data and Data Science.  |                |         | 5Marks  | CO1 | 2      | 2       |
| 3.B.  | Explain A/B testing with a suitable example.  |                |         | 5Marks  | CO1 | 2      | 2       |
| 3.C.  | Briefly explain on Power and sample size  |                |         | 5Marks  | CO1 | 2      | 2       |
| 4.    | Write question either No.3 or Q.No .4   |                |         | 15Marks | CO1 | 2      | 2       |
| 4.A.  | Explain how linear regression can be used for prediction.   |                |         | 5Marks  | CO1 | 2      | 2       |
| 4.B.  | Consider the following data set: 14, 23, 9, 12, 21, 8, 8 Find mean,median, mode and Standard Deviation  |                |         | 5Marks  | CO1 | 2      | 2       |
| 4.C.  | Explain the importance of ANOVA   |                |         | 5Marks  | CO1 | 2      | 2       |
| 5.    | Illustrate in detail Big data storage Technologies  | choice Q-6     |         | 10Marks | CO2 | 2      | 1       |
| 6.    | Explain how to sample K items from a stream   |                |         | 10Marks | CO2 | 2      | 1       |
| 7.    | Answer either Q.No.7 or Q.No.8  | choice Q-8     |         | 15Marks | CO2 | 2      | 1       |
| 7.A.  | Explain the strategy for combining the estimates of m, the number of distinct elements  |                |         | 5Marks  | CO2 | 2      | 2       |
| 7.B.  | Explain the characteristics of Big data   |                |         | 5Marks  | CO2 | 2      | 2       |
| 7.C.  | Brief the different Issues in Stream processing   |                |         | 5Marks  | CO2 | 2      | 2       |
| 8.    | Answer either Q.No.7 or Q.No.8  |                |         | 15Marks | CO2 | 2      | 2       |
| 8.A.  | Explain any one algorithm to count the number of distinct elements in a Datastream  |                |         | 5Marks  | CO2 | 2      | 2       |
| 8.B.  | Elaborate how many tuples from the stream can be stored as sample.  |                |         | 5Marks  | CO2 | 2      | 2       |
| 8.C.  | Explain importance of distance metrics  |                |         | 5Marks  | CO2 | 2      | 2       |

|       |   |                |  |         |     |   |   |
|-------|---|----------------|--|---------|-----|---|---|
| 9.    | Define HDFS. Discuss the HDFS architecture and HDFS commands in brief.  | choice<br>Q-10 |  | 10Marks | CO3 | 2 | 1 |
| 10.   | Explain the Map-Reduce framework in detail.   |                |  | 10Marks | CO3 | 2 | 1 |
| 11.   | Answer either Q.No.11 or Q.No.12  | choice<br>Q-12 |  | 15Marks | CO3 | 2 | 2 |
| 11.A. | Explain the Storage mechanism in HBase. Compare Row oriented and Column Oriented database structures                  |                |  | 5Marks  | CO3 | 2 | 2 |
| 11.B. | Compute the time required to transfer a 1GB data file from a database to instance(RAM) over the network with 100MBPS. |                |  | 5Marks  | CO3 | 2 | 2 |
| 11.C. | Give a brief note the Map Reduce.   |                |  | 5Marks  | CO3 | 2 | 1 |
| 12.   | Answer either Q.No.11 or Q.No.12  |                |  | 15Marks | CO3 | 2 | 2 |
| 12.A. | Explain how Apache Spark solvesmap reduce problems using Resilient Distributed Dataset.                               |                |  | 5Marks  | CO3 | 2 | 2 |
| 12.B. | Explain the importance of estimation of moments   |                |  | 5Marks  | CO3 | 2 | 2 |
| 12.C. | Elaborate the terms DataNode, Master node, Slave node, Namenode and Mapper  |                |  | 5Marks  | CO3 | 2 | 2 |
| 13.   | Explain various data insertion techniques in HIVE with example.   | choice<br>Q-14 |  | 10Marks | CO4 | 3 | 1 |
| 14.   | How Big Data Analytics can be useful in the development of smart cities.  |                |  | 10Marks | CO4 | 3 | 1 |
| 15.   | Answer either Q.No. 15 or Q.No16  | choice<br>Q-16 |  | 15Marks | CO4 | 3 | 2 |
| 15.A. | Explain Data Analysis of RAPID system.  |                |  | 5Marks  | CO4 | 3 | 2 |
| 15.B. | Contrast Tidy text with other Data structures.  |                |  | 5Marks  | CO4 | 3 | 2 |
| 15.C. | Give a short note on Hive   |                |  | 5Marks  | CO4 | 3 | 1 |
| 16.   | Answer either Q.No.15 or Q.No.16  |                |  | 15Marks | CO4 | 3 | 2 |
| 16.A. | Discuss how Pig data model will help in an effective data flow  |                |  | 5Marks  | CO4 | 3 | 2 |
| 16.B. | Explain the impact of impact of data analysis in stock market prediction.   |                |  | 5Marks  | CO4 | 3 | 2 |
| 16.C. | Explain how to perform social media sentiment analysis.   |                |  | 5Marks  | CO4 | 3 | 2 |

[object HTMLDivElement]