| **Examination** | **: B.Tech Semester VII** | **Seat No.** | **:** |
| --- | --- | --- | --- |
| **Date** | **: 13/10/2022** | **Day** | **: Thursday** |
| **Time** | **: 3:30 p.m. to 4:45 p.m.** | **Max. Marks** | **: 36** |

| **INSTRUCTIONS:** | |
| --- | --- |
| 1. | Figures to the right indicate maximum marks for that question. |
| 2. | The symbols used carry their usual meanings. |
| 3. | Assume suitable data, if required & mention them clearly. |
| 4. | Draw neat sketches wherever necessary. |

Blooms Taxonomy levels : R-Remembering, U- Understanding, A-Applying, N-Analyzing, E- Evaluating, C-Creating

| **Q.1** |  | **Do as directed.** | | **[12]** |
| --- | --- | --- | --- | --- |
| **CO4** | **A** | (a) | If the data is continuously being fed in, how will you perform data analytics? | [2] |
| **CO2** | **N** | (b) | For the example of customers, products and related orders placed from various cities analyze the possibility of a drill down scenario. | [2] |
| **CO1** | **R** | (c) | How does Hive architecture distinguish metadata and data? | [2] |
| **CO2** | **U** | (d) | Describe common steps for report generation as part of Data Analytics. | [2] |
| **CO3** | **C** | (e) | Give an example for which Apache Mahout can be used for. | [2] |
| **CO4** | **R** | (f) | What is SerDe with respect to Hive? | [1] |
| **CO1** | **R** | (g) | List data types supported in Hive. | [1] |
|  |  |  |  |  |
| **Q.2** |  | Attempt ***Any TWO*** from the following: | | **[12]** |
| **CO3** | **E** | (a) | Perform Attribute Oriented Induction for Data Characterization on the following initial working relation and explain your reasoning.   | name | gender | major | birth\_place | birth\_date | residence | phone | gpa | | --- | --- | --- | --- | --- | --- | --- | --- | | [6] |
| **CO3** | **E** | (b) | Device a decision tree for the concept buys\_computer indicating whether an AllElectronics customer is likely to purchase a computer. Assume relevant data samples. AllElectronics:   | RowID | Age | Income | isStudent | CreditRating | Class:buys\_computer | | --- | --- | --- | --- | --- | --- | | [6] |
| **CO3** | **E** | (c) | Suppose a group of 12 sales price records has been sorted as follows:  5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215  Partition above in three bins using clustering technique. Explain your learning. | [6] |
|  |  |  |  |  |
| **Q.3** |  | **Attempt the following:** | | **[12]** |
| **CO1** | **N** | (a) | Explain how spark is polyglot? Illustrate what makes Spark faster compared to core hadoop+mapreduce programming? | [6] |
| **CO1** | **R** | (b) | Describe Spark ecosystem. | [6] |
|  |  | **OR** | |  |
| **Q.3** |  | **Attempt the following:** | | **[12]** |
| **CO1** | **N** | (a) | What is lazy evaluation in Spark? Illustrate how DAG (Directed Acyclic Graph) is used in Spark for runtime performance? | [6] |
| **CO1** | **R** | (b) | Describe Spark architecture. | [6] |