| **Examination** | **: B.Tech Semester - VII** | **Seat No.** | **:** |
| --- | --- | --- | --- |
| **Date** | **: 04/08/2022** | **Day** | **: Thursday** |
| **Time** | **: 11:00 a.m. to 12:15 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. |

| **Q.1** | **Do as directed.** | | **[12]** |
| --- | --- | --- | --- |
| **CO4** | (a) | Out of mentioned, which are falling under horizontal scaling approach? Justify your answer. (I) supercomputer (II) private cluster (III) cloud | [2] |
| **CO1** | (b) | Analyze box plot in relevance to five number textual/numeric summary. | [2] |
| **CO4** | (c) | Describe CAP theorem w.r.to (with respect to) Distributed Systems. | [2] |
| **CO1** | (d) | Provide specialities of SQL, NoSQL and NewSQL. | [2] |
| **CO4** | (e) | Compare and contrast Client-server with Master-worker architecture. | [2] |
| **CO4** | (f) | What is the full form of BASE w.r.to (with respect to) NoSQL? | [1] |
| **CO1** | (g) | List categories of NoSQL databases. | [1] |
| **Q.2** | Attempt ***Any TWO*** from the following questions. | | **[12]** |
| **CO2** | (a) | Evaluate below given plots using basic statistical descriptions of data, as part of “know your data” exercise:   | Plot.I of dataset I | Plot.II of dataset II | Plot.III of dataset III | Plot.IV of dataset IV | | --- | --- | --- | --- | |  |  |  |  | | [6] |
| **CO2** | (b) | Imagine that you are at the university library. You see a few students browsing through the library catalog on kiosks. You observe the librarians busy at work issuing and returning books. You see a few students fill up the feedback form on the services offered by the library. Quite a few students are learning using the e-learning content. Think for a while on the different types of data being generated in this scenario. Support your answer with logic. | [6] |
| **CO2** | (c) | Knowing various types of attributes, support how to handle peculiarities of data in the process of knowledge discovery from data (KDD). | [6] |
| **Q.3** | Attempt ***Any ONE*** from the following questions.. | | **[12]** |
| **CO5** | (a) | As in any university, students are required to bookkeep their data, imagine that the university has housed a long run private infrastructure using hadoop distributed file system. Create a help document showing any four CLI operations and GUI interaction/usage by students, how to. | [6] |
| **CO6** | (b) | Outline/diagramatize how the map-reduce framework works using example of counting books based on trim size into ‘Mass market paperback books’ (4.25 x 6.87”), Novels (5 x 8”), Hardbacks (6 x 9”) and Textbooks (from 6 x 9” plus all the way up to 8.5 x 11.”). | [6] |
|  | **OR** | |  |
| **CO5** | (a) | If you may have many computer systems (commodity hardware) with networking support, recommend a solution to handle big data characteristics. Summarize steps for installation of software recommended. | [6] |
| **CO6** | (b) | Develop a program to take advantage of hadoop map-reduce for counting words’ frequencies sorted alphabetically by word, if possible. Make sure to write documenting comments to explain the working of your functions/api. | [6] |