|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | **Reg. No.:** | |  | | | |
| **Name :** | |  | | | |
|  | | | | | | | | | | |
| **Final Assessment Model Question Paper** | | | | | | | | | | |
| Programme | | | : | **M.Tech. CSE with Specialization in BDA and CS** | | Semester | | : | **FALL 2020-21** | |
| Course Title | | | : | **Big Data Frameworks** | | Course Code | | : | **CSE6001** | |
| Faculty | | | : | **Prof. Ramesh Ragala** | | Slot | | : | **D1** | |
| Class Nbr | | : | **CH2020211002020** | |
| Time | | | : | **Three hours** | | Max. Marks | | : | **100** | |
| **Answer any ALL questions** | | | | | | | | | | |
| **Q.No** | **Question Description** | | | | | | | | | **Marks** |
| 1. | Recently, Bandhan Bank has opened 720 branches with ten million employees. The bank management wants to maintain ERP system for their employees. So they are planning to store employee data in distributed manner. Suggest a suitable distributed system, which should have high fault-tolerance and explain the working procedure of it, with neat architecture. | | | | | | | | | **[10]** |
| 2. | Illustrate the working procedure of MapReduce framework with your own example. Explain map and reduce steps diagrammatically. | | | | | | | | | **[10]** |
| 3. | A sample medical store dataset is as follows:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | S.No | CustomerName | Medicine | Gender | Amount Spent in Rupess | | 1 | Putin | Avil | M | 1024 | | 2 | Hillary | Paracet | F | 2048 | | 3 | Broadman | Avil | M | 1024 | | 4 | MariyaHings | Metacin | F | 516 | | 5 | Hopkins | Avil | M | 256 |   Develop a MapReduce code using Python programming language to calculate the total amount spent on each type of medicine of medical store dataset | | | | | | | | | **[10]** |
| 4. | Develop a standalone Scala application in Apache Spark framework to calculate the total amount spent on each type of medicine. Refer the medical store dataset, which is specified in the above question. | | | | | | | | | **[10]** |
| 5. | With MapReduce framework, the keys are sorted but the values associated with each key are not. To sort the values, we need to write a code which is referred as a secondary sort. The input to this program is a bunch of employee details with attributes such as empID, DoB, FirstName, LastName, Gender, JoinDate and DeptID. Write a MapReduce code to print DeptID, empID, FirstName and LastName of all employees. The output should be in ascending order of DeptID and further sorted by empID in descending order of employees in the same department. | | | | | | | | | **[20]** |
| 6. | List down the drawbacks of MapReduce programming model. Illustrate one drawback of MapReduce programming with suitable example. | | | | | | | | | **[10]** |
| 7. | Illustrate the need of streaming analytics in big data era and explain the role of Spark Streaming with your own example. | | | | | | | | | **[10]** |
| 8. | Where do you use Serialization and De-serialization in Hadoop? Discuss the necessity for developing new serialization methods in Hadoop instead of using the built-in Java serialization methods. | | | | | | | | | **[5]** |
| 9. | VIT management has two files namely students.txt and dues.txt. The file student.txt contains the details of the students currently studying in the college. The fields in student.txt are separated by space and it includes name, roll number, father name and address. The file dues.txt contains the list of roll numbers of students who did not pay the current semester fee. Write a MapReduce code to produce list of students with roll number and address using distributed cache mechanism. | | | | | | | | | **[15]** |
|  |  | Total Marks | | | | | | | | **[100]** |
| ⇔⇔⇔ | | | | | | | | | | |