DHARMSINH DESAI UNIVERSITY, NADIAD

FACULTY OF TECHNOLOGY

B.TECH. SEMESTER VII [CE]

SUBJECT: (CE720) BIG DATA ANALYTICS

[

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Examination** | | **: Third Sessional** | **Seat No** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_** | |
| **Date** | | **: 12/10/2019** | **Day** | **: Saturday** | |
| **Time** | | **: 1:45 p.m. TO 3:00 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.** | | |
| (a) | Mention purposes of parameter and data adapter in the development of Jasper Report. | **[2]** |
| (b) | What is Data Science? Diagrammatically justify that Data Science is multi-disciplinary. | **[2]** |
| (c) | What do you understand by predictive statistics and prescriptive statistics? Where do they fit with respect to second school of thought about Data Analytics? | **[2]** |
| (d) | List out spark deployment alternatives. | **[2]** |
| (e) | State true or false and justify your answer: Spark SQL is more fast compared to applying core RDD operations directly. | **[2]** |
| (f) | List out all programming languages supported by spark currently. | **[2]** |
|  | | | |
| **Q.2** | Attempt ***Any Two*** from the following questions. | |  |
| (a) | Compare mapreduce and spark. Explain with figure why spark is very fast compared to mapreduce. | **[6]** |
| (b) | Explain spark ecosystem. | **[6]** |
| (c) | Answer following questions:   1. Which object of spark is needed to run any spark application? 2. List out two types of RDD operations. Also list out their sub types, if exist. 3. Which object of spark is needed to run any spark streaming application? 4. Can we run spark streaming application with executor having one CPU core? 5. What is RDD in spark? 6. Differentiate RDD and Dataframe. | **[6]** |
|  | | | |
| **Q.3** | (a) | Develop a case study solution for a domain of Medical Emergency Management by government, using Big Data and Analytics. Describe your solution using following steps:   1. Problem Statement 2. Challenges (minimum 3) 3. Leveraging Technologies (minimum 3) 4. Impacts | **[6]** |
|  | (b) | Explain below topics with respect to Hive:   1. Managed Tables 2. SerDe 3. RCFile | **[6]** |
| **OR** | | | |
| **Q.3** | (a) | What is the primary challenge in mining rapidly arriving data streams? Provide ways queries get asked and techniques to perform analytics. Name any four applications of such kind. | **[6]** |
| (b) | Provide description of below headings with respect to Pig (2 topics each):   1. Main Components 2. Complex Data Types 3. Relational Operators | **[6]** |