DHARMSINH DESAI UNIVERSITY, NADIAD

FACULTY OF TECHNOLOGY

B. TECH SEMESTER VII

COMPUTER ENGINEERING

SUBJECT: BIG DATA ANALYTICS (CE-720)

| **Examination** | **: Online Exam Section-II** | **Seat No** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| --- | --- | --- | --- |
| **Date** | **: 25/11/2021** | **Day** | **: Thursday** |
| **Time** | **: 1 hour and 30 minutes** | **Max. Marks** | **: 30** |

| **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. | | | |
| **SECTION – II** | | | | |
| **Q.3** | | Attempt the following short questions. (4 X 2) | | **[08]** |
| (a) | Fill in the blanks:  (I) Hive compiles SQL queries into \_\_\_\_\_\_\_\_\_\_\_\_ and then runs the same in \_\_\_\_\_\_\_\_\_\_.  (II) \_\_\_\_, \_\_\_\_\_ and \_\_\_\_\_ are sources/examples of semi structured data which is also referred to as self describing structure. |  |
| (b) | Specify True/False:  (I) JasperReports is an open-source reporting engine completely written in Java.  (II) Pig has a complex data type called ‘bag’ which is nothing but a collection of tuples. |  |
| (c) | Provide the characteristic difference between Data Silos and Data Lakes with any technical analogy. |  |
| (d) | List types of analytics with their purpose in accordance with industrial revolutions. |  |
| **Q.4** | | Attempt ***Any Three*** from the following questions. (3 X 4) | | **[12]** |
| (a) | Explain write operation in Cassandra with figure. |  |
| (b) | Explain the ‘reduce’ phase of MapReduceV2 with all sub phases. |  |
| (c) | Briefly explain Spark Streaming. |  |
| (d) | Briefly explain Spark SQL features. |  |
| (e) | Write a java program/mongoDB commands to perform word count using map-reduce. |  |
| **Q-5** | | Attempt ***Any Two*** from the following questions. (2 X 5) | | **[10]** |
| (a) | (I) Write a short note on key advantages of Hadoop.  (II) Describe CAP theorem. |  |
| (b) | Describe briefly mentioned terms with respect to Apache Hive:  RCFile, Bucketing, SerDe, Metastore, MAP |  |
| (c) | Provide mongoDB commands to achieve below requirements:  (I) Display all databases, start using a database named ‘ecommerce’, create a collection named ‘survey’, display all collections created thus far in the chosen database.  (II) Create documents in ‘survey’ having field and value as mentioned and verify:  doc#1 -> ‘name’:’James’ ‘age’:27 ‘education’:’BSc’ ‘scholarship’:’yes’  doc#2 -> ‘name’:’Harry’ ‘age’:25 ‘salary’: 25000  doc#3 -> ‘name’:’William’ ‘age’:25 ‘salary’: 15000  doc#4 -> ‘name’:’Michel’ ‘age’:20 ‘gender’:’female’ hobbies:’hockey’, ‘painting’, ‘singing’  (III) Display all the documents having salary of ‘age’ <= 25  (IV) Display sorted all the documents in ascending order of age followed by descending order of salary nicely.  (V) Display average salary as AvgSalary group by gender. |  |
|  | **OR** |  |
| **Q-5** | | Attempt ***Any Two*** from the following questions. (2 X 5) | | **[10]** |
| (a) | (I) Write a short note on popular schema less databases.  (II) Describe BASE data system design philosophy. |  |
| (b) | Describe briefly mentioned terms with respect to MongoDB:  Sharding, Upsert, Code, MongoImport, \_id |  |
| (c) | Compare and contrast data sampling, windowing computations with respect to data science and analytics. |  |