Reg. No. \_\_\_\_\_\_\_\_\_\_\_\_

****

**End Semester Examination –Nov / Dec– 2019**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Code :** | **15CS3007** | **Duration :** | **3hrs** |
| **Sub. Name :** | **BIG DATA PLATFORMS** | **Max. Marks :** | **100** |

**ANSWER ALL QUESTIONS (5 x 20 = 100 Marks)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q. No.** | **Sub Div.** | **Questions** | **Course**  **Outcome** | **Marks** |
| 1. | a. | Discuss about the different charecteristics of big data with suitable examples. | CO1 | 10 |
| b. | Distinguish between traditional Business Intelligence (BI) environment and big data environment. | CO1 | 5 |
| c. | What are the various phases in big data analytics? Describe the challenges in each phase in detail. |  | 5 |
| **(OR)** | | | | |
| 2. | a. | How do you classify the digital data? Explain each category in detail. | CO1 | 10 |
| b. | Give a brief note on data science and list the responsibilities of a data scientist. | CO1 | 5 |
| c. | Differentiate between parallel and distributed systems. | CO1 | 5 |
|  |  |  |  |  |
| 3. | a. | State the Brewer’s theorem and discuss about Basically Available Soft State Eventual Consistency (BASE). | CO1 | 8 |
| b. | Compare and Contrast SQL, NoSQL and NewSQL. Elucidate the different categories, advantages, limitations and vendors of NoSQL databases in detail. | CO1 | 12 |
| **(OR)** | | | | |
| 4. | a . | With neat diagram, illustrate the MapReduce programming workflow word count example. | CO2 | 15 |
| b. | What is the difference between replication and sharding? | CO1 | 5 |
|  |  |  |  |  |
| 5. | a. | Consider the following collection called “Bank” and write the MongoDB commands for the following questions.  Bank (c-id,c-name, acc\_no, branch, balance)  i) Create a collection called “Bank” and insert 5 documents  ii) Display the list of collections in the current database.  iii) Insert a document into the Bank collection with new field “age” using  save command and observe the difference between save and insert.  iv) Display the details of customer whose age is >40.  v) Create an array called “phone\_nos” in the Bank collection and insert  values  vi) Sort the documents based on the c\_name.  vii) Display the name of the customer starts with ‘J’without displaying \_id  field. | CO3 | 14 |
| b. | Demonstrate the aggregate function in MongoDB with suitable examples. | CO3 | 6 |
| **(OR)** | | | | |
| 6. | a. | Compare and contrast Hadoop and SQL. List the key advantages of Hadoop. | CO2 | 5 |
| b. | What are the key aspects of Hadoop? Explain the Hadoop components with a neat architecture diagram. | CO2 | 15 |
|  |  |  |  |  |
| **7.** | a. | Explain the Hbase architecture with a neat diagram. | CO3 | 15 |
| b. | Give a brief note on minor compaction and major compaction. | CO3 | 5 |
| **(OR)** | | | | |
| 8. | a. | Consider the following schema and write the Cassandra command for the given queries. Customer (c\_id,c\_name,c\_address,city,state, postal\_code)   1. Create a keyspace called “Company” with replication factor- 2 and simple strategy and create a “customer” table with the above mentioned fields, insert the necessary data. 2. Insert a row into the customer table with new fields age and DOB. 3. Update the postal\_code of c\_id=1, to ‘641028' 4. Alter the customer table by adding a set called “skills” and update the values for “skills” 5. Alter the customer table by adding a map “to\_do” and update the values for “to\_do” 6. Display the details of customer living in the city Chennai. | CO1 | 12 |
| b. | Elucidate the features of Cassandra. | CO1 | 8 |
|  | | **Compulsory:** |  |  |
| 9. | a. | Give a detailed note on map-side joins and reduce- side joins. | CO2 | 10 |
| b. | How status updates are propagated through the Map Reduce system? Demonstrate with a neat diagram, necessary classes and methods. | CO2 | 10 |