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



**End Semester Examination –Nov/Dec– 2017**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **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. | Give the definition of Big Data according to Gartner. Discuss the challenges of big data in detail. | CO1 | 10 |
| b. | How to deal with unstructured data? Illustrate the techniques that are used to find patterns in or interpret unstructured data. | CO1 | 10 |
| (OR) | | | | |
| 2. | a. | Elucidate the Characteristics of big data with appropriate examples and statistics. Share your experience as a customer on an e-commerce site. Comment on the big data that gets created on a typical e-commerce site. | CO1 | 10 |
| b. | Why is big data analytics important? Compare and contrast Analytics 1.0, Analytics 2.0 and Analytics 3.0. | CO1 | 5 |
|  | c. | Discuss the various applications and sources of big data. | CO1 | 5 |
|  |  |  |  |  |
| 3. | a. | Elaborate the three categories of digital data, its sources and its features. | CO1 | 10 |
| b. | State and explain the CAP theorem in detail. | CO1 | 5 |
| c. | Mention the responsibilities of a “Data Scientist” | CO1 | 5 |
| (OR) | | | | |
| 4. | a . | Write the difference between SQL and NoSQL databases. | CO1 | 6 |
| b. | Why NoSQL? Discuss the advantages of NoSQL. | CO1 | 4 |
|  | c. | List the types of NoSQL data bases and give examples for each. |  | 10 |
|  |  |  |  |  |
| 5. | a. | Consider the following collection called “Employee”and give the MongoDB commands for the following questions  Employee(e-id,e-name, basic\_pay,Dept)   1. Create a collection called “Employee” and insert 5 documents 2. Display the list of collections in the current database. 3. Find the document wherein the “e-name” has the “Rani” 4. Find the document where the basic\_pay of the employee is greater than Rs. 10,000. 5. Find the documents from the “Employee” collection where e\_name starts with “a”. 6. Sort the documents from the “Employee” collection and display from the fourth document. | CO3 | 12 |
| b. | Enumerate few features of MongoDB and list the difference between MongoDB and SQL. | CO3 | 8 |
| (OR) | | | | |
| 6. | a. | Distinguish between Hadoop 1.X and Hadoop 2.X with a neat diagram. | CO2 | 5 |
| b. | List the components of the Hadoop Eco System. Describe each component in detail. | CO2 | 15 |
|  |  |  |  |  |
| **7.** | a. | Write the Cassandra command to design a table / column family to support the following requirements:   1. Store the basic information about students such as Student Roll number, Student Name, Date of Birth and Address 2. Store the subject preferences of each student. There should be minimum of two subject preferences and maximum of four. The order of preferences as given by the student must be preserved. 3. Store the hobbies of each student. There should be minimum of two hobbies and maximum of four. The hobbies as given by the student should be arranged in alphabetical order 4. Import data from the “D:\Students.csv” into the table “student1” 5. Update the name of S-rollno=1, to ‘John' 6. Display the details of the studentswhose DOB is greater than 10-10-2007 7. Alter the schema of the table “Student” to add a column Mail\_ids | CO1 | 14 |
| b. | Explain the replication strategy in Cassandra. | CO1 | 6 |
| (OR) | | | | |
| 8. | a. | Explain the functions of Master Server and Region Server in Hbase with a neat diagram. Also express the role of Zookeeper in Hbase | CO3 | 15 |
| b. | Describe compaction in Hbase | CO3 | 5 |
|  | |  |  |  |
|  | | **Compulsory:** |  |  |
| 9. | a. | Write short notes on the following independent entities of MapReduce Job Run:   1. Job Tracker 2. Task Tracker 3. HDFS | CO2 | 12 |
| b. | Discuss the types of failures in MapReduce and its causes. | CO2 | 8 |

ALL THE BEST