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



**UNIVERSITY**

(Karunya Institute of Technology & Sciences)

(Declared as Deemed-to-be University under Sec.3 of the UGC Act, 1956)

**Supplementary Examination – June – 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. | Which category will you place your word document in? Discuss the techniques that are used to find patterns or interpret unstructured data. | CO1 | 10 |
| b. | Assume that you are a data scientist. A banking sector hires you to analyze their data. How will you analyze? Will you analyze the data stored or analyze a sample? Will you provide predictive or prescriptive analytics? Which will benefit the most? Explain your answers with clear justification. | CO1 | 10 |
| (OR) | | | | |
| 2. | a. | Why is big data analytics important? Describe the various challenges with big data. | CO1 | 10 |
| b. | Explain the characteristics of big data and its applications in detail. | CO1 | 10 |
| 3. | a. | What are the properties mentioned in CAP theorem? Explain CAP theorem with an appropriate scenario. | CO1 | 5 |
| b. | Illustrate the different types of NoSQL database with suitable examples. State the need for NoSQL databases and its advantages. | CO1 | 10 |
| c. | What do you mean by BASE? Where it is used and how it is achieved? | CO1 | 5 |
| (OR) | | | | |
| 4. | a . | Write the difference between, “limit” and “skip” in MongoDB. Give examples. | CO1 | 4 |
| b. | Consider the following collection called Customer in MongoDB  Customer(c-id,c-name, aac-bal. acc-type)   1. Give the syntax for removing the document from the collection ‘Customer’ which contains the c-name as ”Smith”. 2. Write the command to add a new field called “address” with a value “Coimbatore” to the document with c-id=3 of ‘Customers’ 3. Give the syntax for sorting the documents from the ‘Customer’ collection in the descending order of acc-bal 4. Write the command to display the last 2 documents from the ‘Customer’ collection. 5. What does the following command do?   db.Customer.find().skip(4). limit(4)   1. What is the significance of \_id in MongoDB? | CO1 | 12 |
| c. | Describe the process of sharding in MongoDB. | CO1 | 4 |
| 5. | a. | Demonstrate the MapReduce programming architecture with the word count example. | CO2 | 15 |
| b. | Distinguish between Hadoop 1.X and Hadoop 2.X with a neat diagram | CO2 | 5 |
| (OR) | | | | |
| 6. | a. | Draw a neat diagram to depict the Hadoop components. Describe the Hadoop Eco system and its functionalities in detail | CO2 | 10 |
| b. | Illustrate the working of Flume and Sqoop with their architecture diagram. | CO2 | 10 |
| **7.** | a. | Consider the following schema and write the Cassandra command for the given queries. Employee(employee\_id,employee\_name,employee\_address,city,state, pstal\_code)   1. Create a keyspace called “University” with replication factor- 2 and simple strategy and create a “employee” table with the above mentioned fields, insert the necessary data 2. Insert a row into the employee table with new fields age and DOB. 3. Update the pstal\_code of employee\_id=1, to ‘641067' 4. Alter the employee table by adding a set called “skills” and update the values for “skills” 5. Alter the employee table by adding a map “subject\_allocation” and update the values for “subject\_allocation” 6. Display the details of customer living the city Coimbatore. | CO1 | 12 |
| b. | Explain the Gossip protocol used in Cassandra. | CO1 | 8 |
| (OR) | | | | |
| 8. | a. | Discuss the data model of Hbase in detail with example. | CO1 | 6 |
| b. | Elucidate the Hbase architecture with a neat diagram. | CO1 | 14 |
|  | | **Compulsory:** |  |  |
| 9. | a. | Write short notes on the followingindependent entities of MapReduce Job Run:   * Job Tracker * Task Tracker * HDFS | CO2 | 12 |
| b. | How status updates are propagated through the MapReduce system? Demonstrate with a neat diagram, necessary classes and methods. | CO2 | 8 |

ALL THE BEST