Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL

A Constituent unit of MAHE, Manipal

VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGINEERING)

END SEMESTER EXAMINATIONS, MAY 2024

SUBJECT: BIG DATA ANALYTICS [CSE 4059]

REVISED CREDIT SYSTEM
(--/05/2024)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL FIVE** questions.
- Missing data may be suitably assumed.

QNo.	Questions	Marks	CO/ CLO	AHEP LO	Blooms Taxonomy Level
1A.	Formulate Cassandra queries to perform the following operations on the column family **'user_details'.** The schema for this column family is as follows: {user_id text primary key, FirstName text, LastName text, email set <text>}. **Operation 1:** Remove an element 'manav@gmail.com' from the set email using subtraction operator for user_id='Manav21'. **Operation 2:** Alter the user_details table and add a column 'places_visited' of type list with data type text. **Operation 3:** Append 'Turkey' to the end of the list of places_visited' where user_id='Scot23'. **Operation 4:** Retrieve email address of 'Albert Samuel' from this set	**4**	**2**	**2,5**	**5**
1B.	Using the same schema mentioned in 1A, formulate CQL queries to perform the following operations. **Operation 1:** Remove an element from the index position of '5' from places_visited in the 'user_details' table with user_id = 'Max21'. **Operation 2:** Alter the user_detials table and add a map column 'date_of_visit' of type timestamp and text. **Operation 3:** Delete an element from the map 'date_of_visit' with the timestamp '2024-04-24' for user_id ='Sarah24'. **Operation 4:** Retrieve the first 2 rows from this table	**4**	**2**	**2,5**	**5**
1C.	Formulate MongoDB queries to create a collection named **Employees** and insert the data given in table 1C into it. Next perform the following operations by writing queries in MongoDB Query Language : **Operation 1**: Update the Department of John Attkins with Id : 1 to IT from Finance. Note: If there is an existing document it	**2**	**2**	**2,5**	**5**

<table>
<tr><td></td><td>should update it. If there is no existing document it will insert it Operation 2: Remove the field ‘Department’ with value ‘IT’ in the document (Id :1) of ‘Employees’
Table 1C
<table><tr><td>Id</td><td>EmpName</td><td>Department</td><td>Position</td></tr><tr><td>1</td><td>John Attkins</td><td>Finance</td><td>Manager</td></tr></table></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2A.</td><td>Consider the collection named ‘colours’ shown in Table 2A.Each document in this collection is a ‘colors’ array. Formulate MongoDB Queries to perform the following operations on this array. Operation 1 : Find the documents with _id:1 from the ‘colours’ collection and display two elements from the array ‘colors’ starting with the element at the second index position. Operation 2: Update the document with _id:4 and replace the element present in the second index position in the ‘Colors’ array with ‘black’. Operation 3: Update the document with ‘_id: 4’ by popping an element from the list of elements present in the array ‘colors’ The element popped is the one from the beginning of the array. Operation 4: Find documents from the collection where ‘blue’ is present in the 2nd index position of the colors array. Operation 5: Find those documents from the colours collection where the size of the array is 3
Table 2A
<table><tr><td>_id:1</td><td>colors: [‘black’,’white’,’grey’,’beige’]</td></tr><tr><td>_id:2</td><td>colors:[‘pink’,’blue’,’purple’,’orange’]</td></tr><tr><td>_id:3</td><td>colors:[‘green’,’ cyan’,’burgundy’,’ violet’,’grey’]</td></tr><tr><td>_id:4</td><td>colors:[‘brown’,’magenta’,’purple’,’turquoise’]</td></tr><tr><td>_id:5</td><td>colors:[‘maroon’,‘peach’,’indigo’]</td></tr></table></td><td>5</td><td>2</td><td>2,5</td><td>5</td></tr>
<tr><td>2B.</td><td>Describe the 3 classes of Digital Data</td><td>3</td><td>1</td><td>1,4</td><td>6</td></tr>
<tr><td>2C.</td><td>Summarize the two main parts of Hadoop 1.0</td><td>2</td><td>3</td><td>1,4</td><td>5</td></tr>
<tr><td>3A.</td><td>Write a MapReduce program to count the occurrence of similar words across 50 files</td><td>5</td><td>3</td><td>5</td><td>3</td></tr>
<tr><td>3B.</td><td>Summarize any 6 differences between HBase and Hadoop/HDFS</td><td>3</td><td>3</td><td>1,4</td><td>5</td></tr>
<tr><td>3C.</td><td>Summarize the functions of MapReduce Daemons</td><td>2</td><td>3</td><td>1,4</td><td>5</td></tr>
<tr><td>4A.</td><td>With the help of an example illustrate the how the summary function in spark performs aggregate operations on a DataFrame</td><td>4</td><td>4</td><td>5</td><td>4</td></tr>
<tr><td>4B.</td><td>Outline the advantages of using Scala in Data Science</td><td>3</td><td>4</td><td>4</td><td>4</td></tr>
<tr><td>4C.</td><td>Describe any 6 features of Hadoop system</td><td>3</td><td>3</td><td>1,4</td><td>6</td></tr>
<tr><td>5A.</td><td>Define RDD. With the help of an example explain how it can be created</td><td>4</td><td>4</td><td>4</td><td>5</td></tr>
</table>

5B.	With the help of an example describe the different mechanisms using which caching is carried out in Spark	**4**	**4**	**4**	**6**
5C.	With the help of an example describe the working of the countByValue method in RDD API	**2**	**4**	**4**	**6**

Abbreviations:

M – Marks

CO/CLO– Course Outcome (NBA)/Course Learning Outcome (IET).

CLO – Course Learning Outcome as per AHEP 4

BT - Blooms Taxonomy Level