Roll No. | | | | | | | | | | |

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. / B.Tech. END SEMESTER EXAMINATIONS, APRIL / MAY 2024

MINOR DEGREE ON DATA SCIENCE

SIXTH SEMESTER

CSM511: BIG DATA ANALYTICS

(Regulations 2019

Time: 3 Hours | Answer ALL Questions | Max. Marks: 100

CO 1	To understand big data.
CO 2	To learn and use NoSQL big data management.
CO 3	To learn MapReduce analytics using Hadoop and related tools.
CO 4	To work with MapReduce applications
CO 5	To understand the usage of Hadoop related tools for Big Data Analytics.

BL – Bloom's Taxonomy Levels

(L1 - Remembering, L2 - Understanding, L3 - Applying, L4 - Analysing, L5 - Evaluating, L6 - Creating)

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

Q. No	Questions	Marks	CO	BL
1	Outline a primary key with an example.	2	CO1	L1
2	What is a foreign key? Give example.	2	CO1	L2
3	Define a schemaless database.	2	CO2	L3
4	What is a collection in MongoDB?	2	CO2	L1
5	Write a note on MapReduce.	2	CO3	L1
6	Define a distributed system.	2	CO3	L1
7	What is Hadoop?	2	CO4	L2
8	Outline the advantages of replication.	2	CO4	L2
9	What is Hbase?	2	CO5	L1
10	Write a note on Hive.	2	CO5	L1

PART- B (5 x 13 = 65 Marks)

(Answer all Questions)

Q. No	Questions	Marks	CO	BL
11 (a)	What is big data? Outline the characteristics of big data.	13	CO1	L1
	OR			
11 (b)	What is crowd sourcing? Outline crowd sourcing analytics with an example.	13	CO1	L1
12 (a)	What is a NoSQL database? Outline the features of NoSQL databases.	13	CO2	L3
	OR			
12 (b)	Illustrate CRUD operations using MongoDB for the following Relational Schema: EMPLOYEE (**ENO**, NAME, DESIGNATION, BASIC_PAY, BCODE) BRANCH (**BCODE**, BNAME)	13	CO2	L3
13 (a)	Outline the steps to calculate how many times each word is repeated across *'n'* sentences using the MapReduce framework with an example *(n = 6)*.	13	CO3	L3
	OR			
13 (b)	Outline the Hadoop YARN architecture with a diagram.	13	CO3	L3
14 (a)	Present an outline of Hadoop Distributed File system.	13	CO4	L2
	OR			
14 (b)	Outline the Cassandra data model with an example.	13	CO4	L2
15 (a)	Outline the pig data model with an example.	13	CO5	L2
	OR			
15 (b)	Outline the features of HiveQL with an example.	13	CO5	L2

PART- C (1 x 15 = 15 Marks)

(Q.No.16 is compulsory)

<table>
<tr><th>Q. No</th><th>Question</th><th>Marks</th><th>CO</th><th>BL</th></tr>
<tr><td>16.</td><td>Consider the relation EMP and DEPT presented below:</td><td></td><td></td><td></td></tr>
<tr><td></td><td>
EMP
<table>
<tr><th><u>ENO</u></th><th>NAME</th><th>DNO</th></tr>
<tr><td>24001</td><td>Mary</td><td>NULL</td></tr>
<tr><td>24002</td><td>Lilly</td><td>NULL</td></tr>
<tr><td>24003</td><td>Jane</td><td>NULL</td></tr>
<tr><td>24004</td><td>Helen</td><td>PH</td></tr>
<tr><td>24005</td><td>Sam</td><td>PH</td></tr>
</table>
DEPT
<table>
<tr><th><u>DNO</u></th><th>DNAME</th></tr>
<tr><td>ZO</td><td>ZOOLOGY</td></tr>
<tr><td>BT</td><td>BOTANY</td></tr>
<tr><td>PH</td><td>PHYSICS</td></tr>
<tr><td>CH</td><td>CHEMISTRY</td></tr>
<tr><td>MT</td><td>MATHEMATICS</td></tr>
</table>
</td><td></td><td></td><td></td></tr>
<tr><td></td><td>Perform Cartesian product, equi join, left outer join, right outer join and full outer join between relation EMP and relation DEPT.</td><td>15</td><td>CO1</td><td>L3</td></tr>
</table>