

# Data Analytics



PAPER ID-310908

Roll No:

**BTECH**  
**(SEM V) THEORY EXAMINATION 2024-25**  
**DATA ANALYTICS**

**TIME: 3 HRS****M.MARKS: 70**

Note: Attempt all Sections. In case of any missing data; choose suitably.

**SECTION A****1. Attempt all questions in brief. 2 x 07 = 14**

Q no.	Question	CO	Level
a.	Differentiate between Predictive and Prescriptive Data Analytics.	1	K2
b.	Define the term data lake, data base and data warehouse.	1	K1
c.	Explain the concept of Outliers.	2	K2
d.	Describe the concept of Lasso Regression.	2	K2
e.	Differentiate between Stream Processing and Traditional Data Processing.	3	K2
f.	Write the two limitations of K-Mean.	4	K1
g.	Discuss the various categories of clustering techniques.	5	K2

**SECTION B****2. Attempt any three of the following: 07 x 3 = 21**

a.	Explain the different categories of data analytics with examples.	1	K2
b.	Explore PCA. Given data = {4, 8, 13, 7; 11, 4, 5, 14}. Compute the principal component using PCA algorithm. Also use PCA to reduce dimension from 2 to 1.	2	K3
c.	Explore the term- Market Basket Analysis. Is it supervised or unsupervised? Determine how would a company use market basket analysis to improve its marketing strategies?	3	K3
d.	Differentiate between CLIQUE and ProCLUS clustering	4	K4
e.	Differentiate between NoSQL database and a Relational database. Identify when one should use a NoSQL database instead of a relational database with a suitable example.	5	K4

**SECTION C****3. Attempt any one part of the following: 07 x 1 = 07**

a.	Differentiate between Structured data, Semi-structured data and Unstructured Data.	1	K2
b.	Describe Big Data and its characteristics.	1	K2

**4. Attempt any one part of the following: 07 x 1 = 07**

a.	Differentiate between Neural Network and Artificial Neural Network.	2	K2
b.	A= {(10,0.2), (20,0.4), (25,0.7), (30,0.9), (40,1), (50,0.4)} B= {(10,0.4), (20,0.1), (25,0.9), (30,0.2), (40,0.6), (50, 0.6)} Apply Union, Intersection, Complement, Bold Union and Bold Intersection operations on above listed Fuzzy Sets.	2	K3

**5. Attempt any one part of the following: 07 x 1 = 07**

a.	Explain and apply Flajolet-Martin algorithm on the following stream of data to identify unique elements in the stream. S=1,3,2,1,2,3,4,3,1,2,3,1 Given: h(x)=(6x+1) mod 5	3	K3
----	---	---	----



PAPER ID-310908

Printed Page: 2 of 2  
Subject Code: BCS052

Roll No: \_\_\_\_\_

**BTECH**  
**(SEM V) THEORY EXAMINATION 2024-25**  
**DATA ANALYTICS**

**TIME: 3 HRS****M.MARKS: 70**

b.	Discuss the Concept of filtering in Data Stream Processing. Explain Bloom Filtering in detail.	3	K2
----	--	---	----

**6. Attempt any one part of the following:****07 x 1 = 07**

a.	<p>Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9). Initial cluster centers are A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as</p> $P(a, b) =  x_2 - x_1  +  y_2 - y_1 $ <p>Use K-Means Algorithm to find the three cluster centers after implanting all eight points.</p>	4	K3														
b.	<p>The database has 6 transactions. Assume Support threshold=50%, Confidence= 60%</p> <table border="1" data-bbox="298 1062 965 1327"> <thead> <tr> <th>TID</th> <th>Items Bought</th> </tr> </thead> <tbody> <tr> <td>10</td> <td>Beer, Nuts, Diaper</td> </tr> <tr> <td>20</td> <td>Beer, Coffee, Diaper</td> </tr> <tr> <td>30</td> <td>Beer, Diaper, Eggs</td> </tr> <tr> <td>40</td> <td>Nuts, Eggs, Milk</td> </tr> <tr> <td>50</td> <td>Nuts, Coffee, Diaper, Eggs, Milk</td> </tr> <tr> <td>60</td> <td>Beer, Nuts, Diaper</td> </tr> </tbody> </table> <p>i) Use Apriori algorithm to find all frequent itemsets.  ii) Show all the strong association rules (with support and confidence)</p>	TID	Items Bought	10	Beer, Nuts, Diaper	20	Beer, Coffee, Diaper	30	Beer, Diaper, Eggs	40	Nuts, Eggs, Milk	50	Nuts, Coffee, Diaper, Eggs, Milk	60	Beer, Nuts, Diaper	4	K3
TID	Items Bought																
10	Beer, Nuts, Diaper																
20	Beer, Coffee, Diaper																
30	Beer, Diaper, Eggs																
40	Nuts, Eggs, Milk																
50	Nuts, Coffee, Diaper, Eggs, Milk																
60	Beer, Nuts, Diaper																

**7. Attempt any one part of the following:****07 x 1 = 07**

a.	Brief about the main components of MapReduce. <a href="https://www.aktuonline.com">https://www.aktuonline.com</a>	5	K2
b.	Draw the architecture of HIVE with its features.	5	K2

<https://www.aktuonline.com>

Whatsapp @ 9300930012

Send your old paper &amp; get 10/-

अपने पुराने पेपर्स भेजे और 10 रुपये पायें,

Paytm or Google Pay से

**B.Tech.**  
**(SEM V) THEORY EXAMINATION 2022-23**  
**DATA ANALYTICS**

**Time: 3 Hours****Total Marks: 100**

**Note:** Attempt all Sections. If require any missing data; then choose suitably.

**SECTION A**

**1. Attempt all questions in brief. 2 x 10 = 20**

- (a) What are the common open source tools for the model building phase?
- (b) What is decision tree?
- (c) What is learning rate?
- (d) What is rule induction?
- (e) How can you deal with uncertainty?
- (f) What is DSMS?
- (g) Write names of two visualization tools.
- (h) Explain the principle behind Hierarchical clustering technique.
- (i) Differentiate between Pig and SQL.
- (j) Write a short note on R graphical user interface.

**SECTION B**

**2. Attempt any three of the following: 10x3 = 30**

- (a) Compare and contrast traditional analytics structure to modern analytics architecture.
- (b) Explain multivariate analysis and Bayesian network.
- (c) Explain Datar-Gionis-Indyk-Motwani (DGIM) algorithm for counting oneness in a window.
- (d) Why PCY algorithm is preferred over Apriori algorithm?
- (e) How RDBS is different from NoSQL?

**SECTION C**

**3. Attempt any one part of the following: 10x1 = 10**

- (a) Explain Apache Hadoop , KNIME & Open refine in detail.
- (b) Explain various phases of Data Analytics Life Cycle.

**4. Attempt any one part of the following: 10x1 = 10**

- (a) Differentiate between Crisp logic and Fuzzy logic.
- (b) What are the different kernel methods of Data Analytics?

**5. Attempt any one part of the following:** **10x1 = 10**

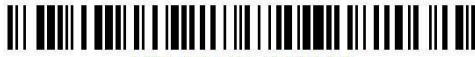
- (a) Explain Bernoulli sampling with its algorithm.
- (b) What are the different components of a general stream processing model? List few sources of streaming data .

**6. Attempt any one part of the following:** **10x1 = 10**

- (a) What is Prediction error ? With the help of suitable example explain prediction error in classification and regression.
- (b) Explain SON algorithm to find all or most frequent item sets using at most two passes.

**7. Attempt any one part of the following:** **10x1 = 10**

- (a) Draw and discuss the architecture of Hive in detail.
- (b) What are the approaches to integrate the human in data exploration process to realize different types of approaches to visual data mining?

Roll No: 

--	--	--	--	--	--	--	--	--	--

**B TECH**  
**(SEM-V) THEORY EXAMINATION 2020-21**  
**DATA ANALYTICS**

**Time: 3 Hours****Total Marks: 100****Note:** 1. Attempt all Sections. If require any missing data; then choose suitably.**SECTION A****1. Attempt all questions in brief.****2 x 10 = 20**

Q no.	Question	Marks	CO
a.	What are the different types of data?	2	1
b.	Explain decision tree.	2	1
c.	Give the full form of RTAP.	2	3
d.	List various phases of data analytics lifecycle.	2	1
e.	Explain the role of Name Node in Hadoop.	2	5
f.	Discuss heartbeat in HDFS.	2	5
g.	Differentiate between an RDBMS and Hadoop.	2	5
h.	Write names of two visualization tools.	2	4
i.	How can you deal with uncertainty?	2	3
j.	Data sampling is very crucial for data analytics. Justify the statement.	2	3

**SECTION B****2. Attempt any three of the following:**

Q no.	Question	Marks	CO
a.	Explain K-Means algorithms. When would you use k means? State whether the statement “K-Means has an assumption each cluster has a roughly equal number of observations” is true or false. Justify your answer	10	4
b.	Illustrate and explain the steps involved in Bayesian data analysis.	10	2
c.	Suppose that A, B, C, D, E and F are all items. For a particular support threshold, the maximal frequent item sets are {A, B, C} and {D, E}. What is the negative border?	10	1
d.	Discuss any two techniques used for multivariate analysis.	10	2
e.	Design and explain the architecture of data stream model.	10	3

**SECTION C****3. Attempt any one part of the following:**

Q no.	Question	Marks	CO
a.	Describe the architecture of HIVE with its features.	10	5
b.	Brief about the main components of MapReduce	10	5

**4. Attempt any one part of the following:**

Q no.	Question	Marks	CO
a.	Describe any two data sampling techniques.	10	1
b.	Explain any one algorithm to count number of distinct elements in a Data stream.	10	3

Roll No:        **5. Attempt any *one* part of the following:**

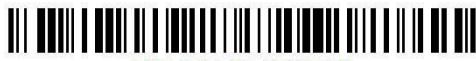
Q no.	Question	Marks	CO
a.	Brief about the working of CLIQUE algorithm.	10	4
b.	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Initial cluster centers are A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x <sub>1</sub> , y <sub>1</sub> ) and b = (x <sub>2</sub> , y <sub>2</sub> ) is defined as $P(a, b) =  x_2 - x_1  +  y_2 - y_1 $ Use K-Means Algorithm to find the three cluster centers after the second iteration	10	4

**6. Attempt any *one* part of the following:**

Q no.	Question	Marks	CO
a.	What is prediction error? State and explain the prediction error in regression and classification with suitable example.	10	4
b.	Given data = {2, 3, 4, 5, 6, 7; 1, 5, 3, 6, 7, 8}. Compute the principal component using PCA Algorithm.	10	2

**7. Attempt any *one* part of the following:**

Q no.	Question	Marks	CO
a.	Develop and explain the data analytics life cycle	10	1
b.	Distinguish between supervised and unsupervised learning with example.	10	1

Roll No: 

--	--	--	--	--	--	--	--	--	--	--	--

**BTECH**  
**(SEM VII) THEORY EXAMINATION 2023-24**  
**DATA ANALYTICS FOT IOT**

**TIME: 3 HRS****M.MARKS: 100**

**Note:** 1. Attempt all Sections. If require any missing data; then choose suitably.

**SECTION A**

**1. Attempt all questions in brief. 2 x 10 = 20**

Q no.	Question	Marks
a.	Write the needs of IoT for futuristic Growth.	2
b.	Explain the working of actuator in IoT.	2
c.	Define Protocols with example.	2
d.	Discuss the use of wireless devices in IoT.	2
e.	Explain the concept of Cloud.	2
f.	Write a short notes on Data Processing.	2
g.	State the need of explore the Data in IoT.	2
h.	How different kind of Data used in IoT Network?	2
i.	Define Machine Learning concept.	2
j.	Write the use of Data Science to manipulate the Data.	2

**SECTION B**

**2. Attempt any three of the following: 10 x 3=30**

a.	Draw the Architecture if Internet of Things and its applications.	10
b.	Illustrate the use of Networking Connecting Protocols in IOT in brief.	10
c.	Explain the process of data management in IoT with Big data technologies.	10
d.	Discuss the steps involved in Data Visualization process in detail.	10
e.	Describe the Feature engineering concept with vast data of IoT.	10

**SECTION C**

**3. Attempt any one part of the following: 10 x 1= 10**

a.	Define Analytics. Explain the use of data analytics in Internet of Things network.	10
b.	Write the various challenges faced in the process of the IoT Analytics.	10

**4. Attempt any one part of the following: 10 x 1= 10**

a.	Describe the process of Analyzing data to infer protocols and device characteristics.	10
b.	Discuss about the IoT Networking Data Messaging protocols in detail.	10

**5. Attempt any one part of the following: 10 x 1= 10**

a.	Define Elastic Analytics. Explain its concept in data analytics with IoT architecture.	10
b.	Describe the process of Designing steps in Data Processing for analyzing the Data.	10



PAPER ID-310340

Printed Page: 2 of 2

Subject Code: KOT077

Roll No:

**BTECH  
(SEM VII) THEORY EXAMINATION 2023-24  
DATA ANALYTICS FOT IOT**

**TIME: 3 HRS**

**M.MARKS: 100**

**6. Attempt any *one* part of the following: 10 x 1= 10**

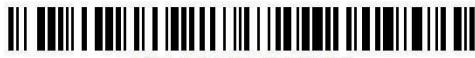
a.	Illustrate the different kind of technologies used to understand the Quality of Data.	10
b.	Explain the process of Statistical analysis with proper example.	10

**7. Attempt any *one* part of the following: 10 x 1= 10**

a.	Demonstrate the use of deep learning methods with IoT data with example.	10
b.	Define Validation Process. Explain the Validation process in IoT with Machine Learning.	10

QP24DP1\_290  
/18-01-2024 13:42:20 / 117.55.242.131

Data Analytics  
Data Warehousing  
&  
**BIG DATA**



Roll No: \_\_\_\_\_

**MCA**  
**(SEM III) THEORY EXAMINATION 2023-24**  
**BIG DATA**

**TIME: 3HRS****M.MARKS: 100**

**Note:** 1. Attempt all Sections. If require any missing data; then choose suitably.

**SECTION A**

**1. Attempt all questions in brief. 2 x 10 = 20**

Q no.	Question	Marks	CO
a.	Define the term Big Data.	2	1
b.	Which type of data can be used for the Big Data?	2	1
c.	How does Apache Hadoop help process the data?	2	2
d.	List the 2 components of Hadoop.	2	2
e.	Write down the default block size of HDFS.	2	3
f.	Discuss two compression formats in Hadoop.	2	3
g.	Which component was introduced newly in Hadoop 2.0?	2	4
h.	List the responsibilities division of Job Tracker in Hadoop 2.0.	2	4
i.	Discuss the execution modes of the Pig	2	5
j.	Write 2 differences between Hbase and RDBMS.	2	5

**SECTION B**

**2. Attempt any three of the following: 10 x 3 = 30**

a.	Discuss the various types of analytics used in Big Data.	10	1
b.	Describe the Hadoop Ecosystem.	10	2
c.	What do you mean by scaling out? Discuss the various reasons for scaling out.	10	3
d.	Compare the NoSQL databases with Relational databases. Describe the various types of NoSQL databases.	10	4
e.	Explain the architecture of Pig with the help of a neat diagram.	10	5

**SECTION C**

**3. Attempt any one part of the following: 10 x 1 = 10**

a.	Discuss the 5 Vs used in Big Data with the help of an example.	10	1
b.	Illustrate the use of cloud computing to manage the data.	10	1

**4. Attempt any one part of the following: 10 x 1 = 10**

a.	Discuss the role of Hadoop Pipes in the map-reduce process.	10	2
b.	Illustrate the map-reduce architecture with the help of a neat diagram.	10	2

**5. Attempt any one part of the following: 10 x 1 = 10**

a.	Illustrate the rack awareness algorithm for writing the data in HDFS.	10	3
b.	Illustrate the process of reading a block in Hadoop.	10	3

**6. Attempt any one part of the following: 10 x 1 = 10**

a.	Explain the components of features and components of Apache Spark.	10	4
b.	Discuss the role of MapReduce in Hadoop. Give the comparison between MapReduce V1 and V2.	10	4

**7. Attempt any one part of the following: 10 x 1 = 10**

a.	Describe the Hive Architecture with the help of a neat diagram.	10	5
b.	Discuss the requirements and architecture of zookeeper.	10	5

Paper Id: 

2	3	2	3	4	2
---	---	---	---	---	---

Roll No. 

--	--	--	--	--	--	--	--	--	--	--	--	--	--

**MCA**  
**(SEM III) THEORY EXAMINATION 2022-23**  
**BIG DATA**

**Time: 3 Hours****Total Marks: 100****Note:** Attempt all Sections. If you require any missing data, then choose suitably.**SECTION A**

**1. Attempt all questions in brief.  $2 \times 10 = 20$**

- a) Define types of digital data.
- b) What is big data privacy and ethics?
- c) Write history of hadoop.
- d) What is real world Map Reduce?
- e) What is HDFS concept?
- f) What is Avro and file-based data structures?
- g) Describe Hadoop ecosystem components.
- h) State the need of NoSQL databases?
- i) Describe the comparison of Pig with databases.
- j) What is a zookeeper?

**SECTION B**

**2. Attempt any three of the following:  $10 \times 3 = 30$**

- a) Explain the concept big data with its architecture, and technological components.
- b) Define anatomy of map reduce job run.
- c) How does hdfs store, read, and write files.
- d) Differentiate in following in reference with proper examples-
  - (i) SQL and NOSQL
  - (ii) MapReduce and YARN
  - (iii)HBase and HDFS
  - (iv)Spark and SCALA
- e) Describe applications on big data using Pig, Hive, and HBase.

**SECTION C**

**3. Attempt any one part of the following:  $10 \times 1 = 10$**

- a) Analyze nature of data, and challenges of convolutional systems over intelligent data analysis? Also differentiate between analysis and reporting.
- b) Describe 5 v's of data and big data features in detail.

**4. Attempt any one part of the following:  $10 \times 1 = 10$**

- a) Describe Apache Hadoop, Hadoop streaming, and Hadoop pipes.
- b) Explain how does map reduce works. Also explain unit tests with MR unit, test data and local tests.

- 5. Attempt any one part of the following: 10x1 = 10**
- a) Explain the procedure of setting up a Hadoop cluster, cluster specification, cluster setup and installation.
  - b) Define data ingest with flume and scoop. Also, explain the concept of Hadoop in the cloud.
- 6. Attempt any one part of the following: 10x1 = 10**
- a) What are spark applications and how to install spark.
  - b) Describe inheritance, functions and closures concept of scale.
- 7. Attempt any one part of the following: 10x1 = 10**
- a) What do you mean by apache hive architecture? Explain hive installation.
  - b) What do you mean by infosphere, BigInsights and big sheets?

QP23DP1\_029  
| 20-02-2023 13:26:33 | 125.21.249.98

**MCA**  
**(SEM V) THEORY EXAMINATION 2019-20**  
**BIG DATA**

**Time: 3 Hours****Total Marks: 70****Note:** 1. Attempt all Sections. If require any missing data; then choose suitably.**SECTION A**

1. **Attempt all questions in brief.** **2 x 7 = 14**

a.	Name different sources of Big Data.
b.	What is the need for analysis of Big Data?
c.	Define Replication.
d.	Define version stamping.
e.	Define Rack Awareness.
f.	Define Workflow in Map Reduce.
g.	What is the role of heartbeat in slave node?

**SECTION B**

2. **Attempt any three of the following:** **7 x 3 = 21**

a.	What is Big Data? Define three Vs of Big Data.
b.	Define NOSQL. Explain any one advantage of NOSQL.
c.	What is Streaming? How it is different from Hadoop Pipes?
d.	Explain different types of job scheduling in YARN.
e.	What is the use of HIVE? Explain the architecture of Hadoop with the help of neat diagram.

**SECTION C**

3. **Attempt any one part of the following:** **7 x 1 = 7**

(a)	What is Big Data Analytics? Explain types of Big Data Analytics.
(b)	Define Crowd Sourcing Analytics. Explain the benefits of Crowd Sourcing Analytics.

4. **Attempt any one part of the following:** **7 x 1 = 7**

(a)	Discuss Sharding with the help of an example. What are the advantages of sharding.
(b)	Explain any two types of distribution models.

5. **Attempt any one part of the following:** **7 x 1 = 7**

(a)	Explain the concept of Blocks, Master Node, Slave Node in Hadoop File System.
(b)	Explain the Avro file based data structures.

6. **Attempt any one part of the following:** **7 x 1 = 7**

(a)	Write the detailed execution of MAP Reduce phase with the help of any example.
(b)	What is the role of YARN? Explain the architecture of YARN.

7. **Attempt any one part of the following:** **7 x 1 = 7**

(a)	How Cassandra is integrated in Hadoop? Explain the Cassandra Write.
(b)	Differentiate between HBASE and RDBMS.

**MCA**  
**(SEM V) THEORY EXAMINATION 2018-19**  
**BIG DATA**

**Time: 3 Hours****Total Marks: 70**

**Note:** 1. Attempt all Sections. If require any missing data; then choose suitably.  
 2. Any special paper specific instruction.

**SECTION A****1. Attempt all questions in brief****2 x 7 = 14**

- a. What is Big data? Discuss it in terms of volume and velocity.
- b. What are the advantages of Hadoop?
- c. Explain Metastore in HIVE.
- d. Discuss big data in healthcare and medicine.
- e. Define Serialization in Hadoop.
- f. What is Zookeeper?
- g. What is the necessity of PIG Latin?

**SECTION B****2. Attempt any three of the following:****7 x 3 = 21**

- a. What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities.
- b. Write a short note on NoSQL databases. List the differences between NoSQL and relational databases?
- c. Draw and explain HDFS Architecture. Explain the function of NameNode and DataNode. What is a Secondary Namenode? Is it a substitute to the Namenode?
- d. Explain "Shuffle & Sort" phase and "Reducer Phase" in MapReduce.
- e. What are views in HIVE? What is the difference between internal and external tables in HIVE?

**SECTION C****3. Attempt any one part of the following:****7 x 1 = 7**

- (a) What are structures, unstructured and semi-structured data? Explain with examples.
- (b) What are the different modes in which Hadoop can be installed and what is the use of each mode from application and developer point of view?

**4. Attempt any one part of the following:****7 x 1 = 7**

- (a) Define the role of combiner and partitioner in a map reduces application.
- (b) Explain Master slave and peer-peer replication in detail.

**5. Attempt any one part of the following:****7 x 1 = 7**

- (a) How does HDFS ensure data Integrity in a Hadoop Cluster?
- (b) Explain Avro file-based data structures in detail.

**6. Attempt any one part of the following:** http://www.aktuonline.com**7 x 1 = 7**

- (a) Discuss Hadoop YARN in detail with failures in classic MapReduce.
- (b) Explain briefly about Input format and Output format in MapReduce.

**7. Attempt any one part of the following:****7 x 1 = 7**

- (a) What are the components of Pig Execution Environment?
- (b) Explain Storage mechanism in HBase. Write a query to create a table in HBase.

(Following Paper ID and Roll No. to be filled in your Answer Books)

Paper ID : 2012307

Roll No.

<input type="text"/>									
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

MCA

Regular Theory Examination (Odd Sem-V), 2016-17

### BIG DATA

*Time : 3 Hours*

*Max. Marks : 100*

#### SECTION-A

1. Attempt all parts. All parts carry equal marks. Write answer of each part in short.  $(10 \times 2 = 20)$

- a) What is Bigdata? Why do you need to analyze big data?
- b) Write down any four industry examples for Big Data
- c) Compare and Contrast No SQL Relational Databases.
- d) Write down the disadvantages of aggregate Oriented Database, How to overcome that?
- e) Define Data Locality Optimization,
- f) State the purpose of Hadoop pipes
- g) What is map reduce?
- h) List down the entities of YARN.
- i) How Cassandra integrated with Hadoop?
- j) List down the tools related with Hadoop.

**NMCAE - 44**

**SECTION - B**

**Note: Attempt any 5 questions from this section(5×10=50)**

2. Discuss about the three dimensions of Big data.
3. Explain Master slave and peer-peer replication in detail.
4. Describe about graph database and schemaless databases
5. Discuss the design of Hadoop Distributed File System(HDFS) and concept in detail
6. Show on how a client read and write data in HDFS, Give an example code.
7. Explain briefly about Input format and Output format in Detail.
8. Write in detail abut Hbase data model and pig data model in detail.
9. Discuss the queries involved in Hive data definition .

**SECTION - C**

**Note: Attempt any 2 questions from this section(2×15=30)**

10. Explain Crowd sourcing analytics and inter and trans firewall analytics.
11. How does Hadoop system analyse data. Explain in your answer with an example code.
12. Discuss different types and formats of Map Reduce with examples.

**ବେଳେବେଳେ**



**Roll No:** \_\_\_\_\_

MCA

# **(SEM III) THEORY EXAMINATION 2023-24**

## **DATA WAREHOUSING & DATA MINING**

**TIME: 3 HRS**

M.MARKS: 100

**Note:** Attempt all Sections. If you require any missing data, then choose suitably.

## SECTION A

### **1. Attempt *all* questions in brief.**

$$2 \times 10 = 20$$

Qno	Questions	Marks	CO
a.	How Data Warehouse is different from Database?	2	1
b.	How Data Mining is different from Data Warehousing?	2	1
c.	Why is Data Transformation required?	2	2
d.	Discuss how regression is related to classification	2	3
e.	What do mean by support and confidence in association rule mining?	2	3
f.	Define Slice and Dice operation.	2	3
g.	What is data generalization?	2	4
h.	What is data Aggregation?	2	5
i.	What do you mean by temporal mining?	2	5
j.	Explain the concept of concept hierarchies in data mining.	2	2

## **SECTION B**

**2. Attempt any *three* of the following:**

$$\checkmark 10^*3 = 30$$

a.	Write detail about the 3-tier architecture and the implementation of data warehouse with a neat diagram.	10	1												
b.	Differentiate between the three main types of data usage: Information Processing, Analytical Processing and Data Mining. Examine the motivation behind OLAP mining.	10	2												
c.	<p>Find all frequent item sets for the given training set using Aprori Algorithm.</p> <table> <thead> <tr> <th>TID</th> <th>ITEMS BOUGHT</th> </tr> </thead> <tbody> <tr> <td>T100</td> <td>(M, O, N, K, E, Y)</td> </tr> <tr> <td>T200</td> <td>(D, O, N, K, E, Y)</td> </tr> <tr> <td>T300</td> <td>(M, A, K, E)</td> </tr> <tr> <td>T400</td> <td>(M, U, C, K, Y)</td> </tr> <tr> <td>T500</td> <td>(C, O, O, K, E, Y)</td> </tr> </tbody> </table>	TID	ITEMS BOUGHT	T100	(M, O, N, K, E, Y)	T200	(D, O, N, K, E, Y)	T300	(M, A, K, E)	T400	(M, U, C, K, Y)	T500	(C, O, O, K, E, Y)	10	3
TID	ITEMS BOUGHT														
T100	(M, O, N, K, E, Y)														
T200	(D, O, N, K, E, Y)														
T300	(M, A, K, E)														
T400	(M, U, C, K, Y)														
T500	(C, O, O, K, E, Y)														
d.	Examine how classification tasks performed by SVM.	10	4												
e.	Describe various OLAP operations with example.	10	5												

## SECTION C

**3. Attempt any *one* part of the following:**

$$10 * 1 = 10$$

a.	Briefly discuss the three schemas for multidimensional databases that consists of 4 dimensions (Time, Branch, Item, Location). Include the appropriate measures required for the schemas.	10	2
b.	Discuss why do we need to pre-process the data. Mention the various tasks to be accomplished as part of pre-processing. Elaborate each.	10	2



PAPER ID-311475

Printed Page: 2 of 2

Subject Code: KCA012

Roll No: \_\_\_\_\_

**MCA**  
**(SEM III) THEORY EXAMINATION 2023-24**  
**DATA WAREHOUSING & DATA MINING**

**TIME: 3 HRS****M.MARKS: 100****4. Attempt any *one* part of the following: 10 \*1 = 10**

a.	Draw the Data Warehouse architecture and discuss its various components.	10	1
b.	Discuss the various applications of Data Warehousing?	10	1

**5. Attempt any *one* part of the following: 10\*1 = 10**

a.	How classification is different from clustering? Discuss Decision Tree based classification technique in detail.	10	3
b.	What do you mean by Data Extraction? Discuss the tools used for Data Extraction.	10	3

**6. Attempt any *one* part of the following: 10\*1 = 10**

a.	How Bayes Theorem can be used for classification task? Explain	10	4
b.	Write short notes on i) CURE ii) DNSCAN	10	4

**7. Attempt any *one* part of the following: 10\*1 = 10**

a.	Discuss in detail the security issues in Data Warehousing.	10	5
b.	Write Notes on i) Web Mining ii) Spatial Mining	10	5

**MCA**  
**(SEM III) THEORY EXAMINATION 2022-23**  
**DATA WAREHOUSING & DATA MINING**

**Time: 3 Hours****Total Marks: 100**

**Note:** Attempt all Sections. If require any missing data; then choose suitably.

**SECTION A**

- 1. Attempt all questions in brief. **2 x 10 = 20****

- a. Write down the components of data warehouse.
- b. What is meta data?
- c. What is SMP?
- d. What do you mean by workload matrix?
- e. What are the steps involved in data transformation?
- f. What is binning technique?
- g. Briefly describe genetic algorithm.
- h. Explain market basket analysis.
- i. What is MOLAP?
- j. Define web mining.

**SECTION B**

- 2. Attempt any three of the following: **10 x 3 = 30****

- a. Define data warehouse. What are the key features of data warehouse?
- b. What is data warehouse planning? Explain its activities.
- c. Describe the steps involved in data mining when viewed as a process of knowledge discovery.
- d. Describe classification. Also discuss Naive Bayesian classification.
- e. Describe various OLAP operations in multidimensional data model.

**SECTION C**

- 3. Attempt any one part of the following: **10 x 1 = 10****

- a. Write short note on :
  - (i) Shared-disk architecture
  - (ii) Distributed memory architecture
- b. Differentiate between OLAP and OLTP.

- 4. Attempt any one part of the following: **10 x 1 = 10****

- a. Discuss master slave processing with the help of diagram.
- b. Explain two primary categories of parallel hardware used for data warehousing.

- 5. Attempt any one part of the following:                   10 x 1 = 10**
- What do you mean by cleaning of the data? Explain the important types of data cleaning.
  - What is decision tree? Explain the classification by decision tree induction.
- 6. Attempt any one part of the following:                   10 x 1 = 10**
- What is clustering? Briefly explain k-means clustering algorithm.
  - Describe neural network. How the neural network is useful in classification.
- 7. Attempt any one part of the following:                   10 x 1 = 10**
- Write a short note on the following:
    - Tuning the data warehouse
    - Testing the data warehouse
  - List various algorithms used in web mining. Briefly discuss any one of them.

QP23DP1\_029  
| 13-02-2023 13:46:01 | 125.21.249.98



**Roll No:** \_\_\_\_\_

**MCA**  
**(SEM III) THEORY EXAMINATION 2021-22**  
**DATA WAREHOUSING & DATA MINING**

*Time: 3 Hours*

Total Marks: 100

**Note:** 1. Attempt all Sections. If require any missing data; then choose suitably.

## SECTION A

- 1. Attempt all questions in brief.** **2 x 10 = 20**

Qno.	Question	Marks	CO
a.	List down the Key Features of Data Warehouse.	2	1
b.	Define Data Mart.	2	1
c.	What is SMP?	2	2
d.	What is Master-slave processing?	2	2
e.	Briefly explain Association rules.	2	3
f.	Define Data cleaning.	2	3
g.	What is data Generalization?	2	4
h.	List down the various measures of Central Tendency.	2	4
i.	What are the characteristics of OLAP?	2	5
j.	Define Web Mining.	2	5

## SECTION B

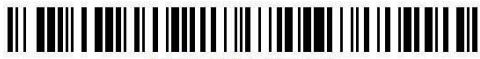
- 2. Attempt any *three* of the following:**

Qno.	Question	Marks	CO
a.	What do you mean by Data Warehouse? Discuss the key features with suitable examples.	10	1
b.	What is data Warehouse implementation? What are the key considerations that are taken while finalizing the physical warehouse schema design?	10	2
c.	Describe the steps involved in data mining when viewed as a process of knowledge discovery.	10	3
d.	Describe Classification. Briefly Outline the major ideas of Bayesian Classification.	10	4
e.	Describe various OLAP operations with an example.	10	5

### **SECTION C**

- 3. Attempt any *one* part of the following:**

Qno.	Question	Marks	CO
a.	Write a note on Shared disk and Shared nothing architecture.	10	1
b.	What is multidimensional data model? How do we convert tables & spreadsheets to data cubes.	10	1



PAPER ID-410750

Roll No: 

--	--	--	--	--	--	--	--	--	--

**4. Attempt any one part of the following:**

Qno.	Question	Marks	CO
a.	What do you mean by Workload matrix?	10	2
b.	Explain Attribute Relevance Analysis with an example.	10	2

**5. Attempt any one part of the following:**

Qno.	Question	Marks	CO
a.	Describe the basic architecture of data mining.	10	3
b.	Why data preprocessing is required in Data Mining. Explain.	10	3

**6. Attempt any one part of the following:**

Qno.	Question	Marks	CO
a.	What is clustering? List types of clustering. Explain any one approach for clustering.	10	4
b.	Describe CLIQUE algorithm.	10	4

**7. Attempt any one part of the following:**

Qno.	Question	Marks	CO
a.	Write short-notes on: (i) Crawlers (ii) Harvest System	10	5
b.	Explain & Define Markov Model.	10	5

**MCA**  
**(SEM IV) THEORY EXAMINATION 2017-18**  
**DATA WAREHOUSING & MINING**

**Time: 3 Hours**

**Total Marks: 100**

**Note:** Attempt all Sections. If require any missing data; then choose suitably.

**SECTION A**

- 1. Attempt all questions in brief. 2 x 10 = 20**
- a. Define Data warehouse.
  - b. What is Meta Data.
  - c. Explain the term Crossover and Mutation in Genetic Algorithm.
  - d. List Various OLAP tools.
  - e. Why data cleaning routines are needed?
  - f. What is Data discretization?
  - g. What is Z-score normalization?
  - h. What do you mean by dimensionality Reduction.
  - i. Define Wavelet Transforms
  - j. Differentiate between Supervised & Unsupervised Learning.

**SECTION B**

- 2. Attempt any three of the following: 10 x 3 = 30**
- a. Explain the concept of Client/Server Computing Model. Explain the various generation of Client/ Server Model in Detail.
  - b. Explain the Data Warehouse Concept. Explain the two-tier data warehouse architecture and its advantages.
  - c. Describe ETL tools.
  - d. What do you mean by decision Tree? Describe ID3 algorithm for decision tree.
  - e. What is Web Mining? What are the classification factors of Web Mining?

**SECTION C**

- 3. Attempt any one part of the following: 10 x 1 = 10**
- (a) Draw the Data warehouse Architecture and briefly explain its various components.
  - (b) Why Data Warehouse is maintained separately from Database? Differentiate between OLTP & OLAP.
- 4. Attempt any one part of the following: 10 x 1 = 10**
- (a) Enumerate the steps involved in mapping the data warehouse to a multiprocessor architecture.
  - (b) Summarize the smoothing techniques followed in data cleaning process.
- 5. Attempt any one part of the following: 10 x 1 = 10**
- (a) Discuss, how Statistics is useful in Data Mining.

- (b) Discuss various OLAP operations. Explain how query performance can be improved by cascading operations.

**6. Attempt any one part of the following:** **10 x 1 = 10**

- (a) What do you mean by neural network? Explain multilayer Feed -Forward neural network.
- (b) Consider five points {X1, X2, X3, X4, X5} with the following coordinates as a two dimensional sample for clustering: X1=(0,2.25); X2=(0,0.25); X3=(1.25,0); X4=(4.5,0); X5=(4.5,2.5); Illustrate the K-means partitioning algorithm (clustering algorithm) using the above data set.

**7. Attempt any one part of the following:** **10 x 1 = 10**

- (a) What is Data Visualization? Explain various data visualization techniques. Explain Q-Q Plot.
- (b) Discuss the architecture of Multimedia Data Mining.

**MCA**  
**(SEM IV) THEORY EXAMINATION 2017-18**  
**DATA WAREHOUSING AND MINING**

**Time: 3 Hours****Total Marks: 70**

**Note:** Attempt all Sections. If require any missing data; then choose suitably.

**SECTION A**

- 1. Attempt all questions in brief. 2 x 7 = 14**
- Define Data Warehouse.
  - What is Client Server Computing?
  - List various functionalities of Data Mining.
  - Differentiate between Inliers and Outliers.
  - How to handle missing values for data cleaning?
  - Define Bayes Theorem.
  - Define Numerosity Reduction.

**SECTION B**

- 2. Attempt any three of the following: 7 x 3 = 21**
- Explain Multidimensional Data Model with an example.
  - Explain 2-tier client server architecture. What are its characteristics.
  - What are different stages of Knowledge Discovery (KDD). Explain with the help of diagram.
  - What do you mean by Association Rules? For what purposes these are being used? Explain with an example.
  - List and explain various types of OLAP servers.

**SECTION C**

- 3. Attempt any one part of the following: 7 x 1 = 7**
- Name four distinguishing characteristics of Data Warehouse architecture. Describe each component briefly.
  - Differentiate between Star Schema & Snow Flake Schema.
- 4. Attempt any one part of the following: 7 x 1 = 7**
- Describe various phases of data transformation. List the important types of transformation engines and explain any one of them.
  - Why business organizations consider data warehousing as a critical need? Write and explain various phases involved in building of data warehouse.
- 5. Attempt any one part of the following: 7 x 1 = 7**
- Explain ETL process in detail.
  - Give a brief description of the following:
    - 1) Binning
    - 2) Regression

- 6. Attempt any one part of the following:** **7 x 1 = 7**
- (a) What do you mean by neural network? Explain multilayer Feed-Forward neural network.
  - (b) What do you mean by decision tree? Describe ID3 algorithm of decision tree.
- 7. Attempt any one part of the following:** **7 x 1 = 7**
- (a) What is Web mining? What are the classification factors of web mining ?
  - (b) Write short notes on the following:
    - 1) Slice & Dice Operation
    - 2) Testing of Data Warehouse

M.C.A.

**THEORY EXAMINATION (SEM-IV) 2016-17**  
**DATA WAREHOUSING & MINING**

**Time : 3 Hours****Max. Marks : 100****Note : Be precise in your answer. In case of numerical problem assume data wherever not provided.****SECTION – A**

- 1. Explain the following:** **10 x 2 = 20**

- (a) State the attributes of data warehouse.
- (b) Enumerate the difference between RISC and CISC.
- (c) List out the categories of query and reporting tools.
- (d) Mention few access tools to handle all possible data warehouse needs.
- (e) Provide the high level view of the modeling process.
- (f) Provide the difference between multidimensional and multi relational OLAP.
- (g) What is the need of a decision tree?
- (h) Diagrammatically represent the centric view of data mining process.
- (i) What is image mining? How it differs from image processing?
- (j) Justify why concept hierarchies are useful in data mining.

**SECTION – B**

- 2. Attempt any five of the following questions:** **5 x 10 = 50**

- (a) Diagrammatically illustrate and discuss the architecture of a data warehouse.
- (b) List and discuss the steps involved in mapping the data warehouse to a multiprocessor architecture.
- (c) Explain the potential performance problems with star schema. Give examples.
- (d) How to extract, cleanup, transform and migrate data in data warehouse?
- (e) Discuss the Apriori algorithm for discovering frequent item sets. Apply the Apriori algorithm to the following data set:

Trans ID	Items Purchased
101	strawberry, litchi, oranges
102	strawberry, butter fruit
103	butter fruit, vanilla
104	strawberry, litchi, oranges
105	banana, oranges
106	banana
107	banana, butter fruit
108	strawberry, litchi, apple, oranges
109	apple, vanilla
110	strawberry, litchi

The set of items is {strawberry, litchi, apple, oranges, vanilla, banana, butter fruit}. Use 0.3 for the minimum support value.

- (f) Consider the following multi-feature cube query: Grouping by all subsets of {item, region, month}, find the minimum shelf life in 2014 for each group, and the fraction of the total sales due to tuples whose price is less than \$100, and whose shelf life is within 25% of the minimum shelf life, and within 50% of the minimum shelf life.
  - (i) Draw a multi-feature cube graph for the query.
  - (ii) Express the query in extended SQL.

- (iii) Is this a distributive multi-feature cube? Justify.
- (g) Summarize the issues involved in mining and extracting information from multimedia databases.
- (h) Enumerate some of the technology integration issues and uses of digital libraries.

### SECTION – C

**Attempt any two of the following questions:**

**$2 \times 15 = 30$**

3. (a) Design multi-dimensional data model for hospital data warehouse, consist three dimensions time, doctor, and patient and the two measures count and charge, where charge is a fee that a doctor charges patients for a visit.
  - (i) Enumerate three classes of schema that are popularly used for modeling data warehouses.
  - (ii) Draw a schema diagram for the above data warehouse using all of the schema classes.
  - (b) How to reduce the size of the fact table? Explain with example.
4. (a) How data mining system can be integrated with a data warehouse?
   
 (b) Suppose that the data for analysis include the attribute ‘age’. The age values for the data tuples are: 13, 15, 16, 16, 19, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70.
  - (i) Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.
  - (ii) How would you determine outliers in the data?
5. Describe the architecture and functionality of MM-DBMS.

Printed Pages: 4

CA407/NMCAE-013/MCAE12

(Following Paper ID and Roll No. to be filled in your  
Answer Books)

Paper ID : 214407

Roll No.

M.C.A.

Theory Examination (Semester-IV) 2015-16

**DATA WAREHOUSING & MINING**

*Time : 3 Hours*

*Max. Marks : 100*

**Note: Attempt questions from all Sections as per directions.**

**Section-A**

**Attempt all parts of this section. Answer in brief. ( $10 \times 2 = 20$ )**

1. (a) What do you mean by Data cleaning?
- (b) Explain the use of Neural network in Data Mining.
- (c) What is clustering? How clustering is different from classification?
- (d) Describe the basic operations performed on data cube.

(1)

P.T.O.

- (e) What is data-mart? Explain its role in Data Warehousing.
- (f) What are the one, two and three dimensional data? Explain.
- (g) Discuss various types of web mining techniques.
- (h) What are outliers? How outliers analysis can be done?
- (i) Write the differences between Data Warehouse and Database System.
- (j) How is Data Mining different from KDD?

### **Section-B**

#### **2. Attempt any five questions from this section.**

$(10 \times 5 = 50)$

- (a) What is OLAP? Differentiate among OLAP, ROLAP and MOLAP servers.
- (b) How is nearest neighbour prediction algorithm used in Data Mining?

- (c) Determine a Clustering method suitable for finding clusters in large database. Explain its working.
- (d) Draw and explain the three-tier architecture of Data Warehouse model.
- (e) What is a Decision tree? Explain the classification by decision tree induction. Describe the tree pruning.
- (f) Explain the mining multidimensional association rules from relational databases and Data Warehouses.
- (g) What do you mean by neural network? Explain multilayer Feed-Forward neural network. Differentiate between Feed-Forward and Feedback system.
- (h) How frequent item set are generated using Apriori algorithm? Explain the method.

### **Section-C**

**Note : Attempt any two questions from this section.**

**$(15 \times 2 = 30)$**

- 3. (i) How ID3 algorithm create a decision on given database? Assume a database to explain your answer.

- (ii) How association rules can be formed on hierachal data? Discuss various approaches.
4. (i) What is hierachal clustering method? Explain its advantages over non-hierachal clustering method.
- (ii) Explain K-means clustering algorithm.
5. Write short notes on any three of the following:
- (i) Data Mining Interfaces
- (ii) Data Visualization
- (iii) Testing of Data Warehouses
- (iv) Data Transformation Tools

**Printed Pages : 3****CA407****(Following Paper ID and Roll No. to be filled in your Answer Book)****PAPER ID : 214407****Roll No.**

<input type="text"/>							
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

**M. C. A.****(SEM. IV) THEORY EXAMINATION, 2014-15****DATA WAREHOUSING AND DATAMINING**Time : **3** Hours][Total Marks : **100**

**Note :** (1) Attempt all questions.  
 (2) All questions carry equal marks.

- 1** Attempt any FOUR questions. **5x4=20**
- (a) How can you link data mining with DBMS?
  - (b) Discuss the role of data mining in data warehousing.
  - (c) What do you mean by cleaning of the data?
  - (d) Write and describe important types of difficulties in data mining process.
  - (e) Write short note on Dimensionality reduction.
  - (f) Explain Histogram. The following data are a list of prices of commonly sold items at a company. The number have been sorted 1, 1, 5, 5, 5, 8, 8, 10, 10, 15, 15, 15, 20, 20, 20, 20. Make a histogram for price using singleton buckets.

**2** Attempt any FOUR question : **5x4=20**

- (a) How frequent item set are generated using Apriori algorithm? Explain the method.
- (b) What do you mean by association rules, for what purposes it is being used ? Explain with example.
- (c) What are the main purposes of statistics, used in data mining?
- (d) Discuss why analytical characterization and attribute, relevance analysis are needed and how these can be performed?
- (e) Define the term data generalization and analytical characterization with example.
- (f) Describe Mining single association rules from Transactional Databases.

**3** Attempt any TWO questions. **10x2=20**

- (a) What do you mean by neural network? Explain multilayer Feed-Forward neural network. Differentiate between Feed-Forward and Feedback system.
- (b) What do you mean by decision tree? Describe ID3 algorithm of the decision tree. Why it is unsuitable for data mining applications ?
- (c) Write short notes on the following
  - (i) Bayesian classification
  - (ii) Back propagation Algorithm

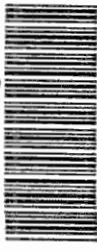
**4** Attempt any TWO questions. **10x2=20**

- (a) (i) Explain Snow - Flake schema with an example.
- (ii) Explain Fact Constellations with an example.
- (b) Give E.F. Codd's 12 guidelines for OLAP.
- (c) Explain briefly the following
  - (i) Concept hierarchy
  - (ii) 3 Tier Architecture.

**5** Attempt any TWO questions. **10x2=20**

- (a) Define and describe the basic similarities and differences among ROLAP, MOLAP, and HOLAP.
  - (b) Explain OLAP functions and tools in brief. What are the main features of OLAP servers?
  - (c) Write short notes on :
    - (i) Backup and Recovery
    - (ii) Tuning Data Warehouse.
-

Printed Pages : 3



\* 2144 \*

419

M.C.A.

**(SEM-IV) THEORY EXAMINATION 2014-15**

**DATA WAREHOUSING & MINING**

*Total Marks : 100*

**Note:** Attempt all questions. All questions carry equal marks

$$5 \times 4 = 20$$

- a) What is Data Warehouse? Explain the two tier architecture of Data Warehouse.
  - b) Explain Data Marts with its advantages.
  - c) Draw a three tiered second generation client/server computing model. What are the main functions of server?
  - d) What do you mean by data cleaning? Describe important types of data cleaning approaches.

2144317

11

Contd

- e) Explain Clustering and Regression with example. 5x4=20

2. Attempt any four parts

  - Explain Data Cube aggregation model.
  - Describe various phases of data transformation. List the important types of transformation engines and explain any one of them.
  - Explain these terms.
    - MOLAP
    - RLOP
  - What is Z-score normalization?
  - What is the role of Statistics in Data Mining.

3. Attempt any two parts : 10x2=20

  - Why business organizations consider data warehousing as a critical need? Write and explain various phases involved in building of data warehouse.
  - What is decision tree? Explain it with an example. Describe the ID3 decision tree algorithm in brief.
  - Write and briefly explain twelve guidelines for OLAP selection. —X—



(Following Paper ID and Roll No. to be filled in your Answer Book)

PAPER ID : 214431

Roll No.

**MCA**  
**(SEM. IV) THEORY EXAM. 2014-15**  
**DATA WAREHOUSING AND MINING**

Time : 3 Hours]

[Total Marks : 100

**Note :** Attempt the questions as indicated.

**Q1.** Attempt any *four* questions from the following :  $5 \times 4 = 20$

- (a) How is a data warehouse different from a database? How they are similar?
- (b) Explain the different schemas for Multidimensional Databases.
- (c) What is Data Mart? List the advantages of Data Mart.

- (d) What are the areas in which data warehouses are used in present and in future?
- (e) Explain the client server computing model. Also list the advantages of client server computing model.
- (f) Explain Distributed DBMS and its advantages.

**Q2.** Attempt any *four* questions from the following :  $5 \times 4 = 20$

- (a) Explain 3-Tier Data warehouse architecture with clean figure.
- (b) List OLAP operations in Multidimensional Data model and explain with an example.
- (c) Define Association Rule Mining. How market analysis forms the Association rules? Discuss the basic concepts.
- (d) How does clustering differ from classification?
- (e) What are the fields in which clustering techniques are used?
- (f) Define Artificial Intelligence and its areas of application.

**Q3.** Attempt any *two* questions from the following :  $10 \times 2 = 20$

- (a) Explain Bayesian classification and rule based classification. Give an example for any one classification and explain in detail.
- (b) Differentiate between OLAP and OLTP.
- (c) Describe neural network techniques for data mining. What are the major difficulties in using these techniques?

**Q4.** Attempt any *two* questions from the following :  $10 \times 2 = 20$

- (a) What are the various issues in data mining? Explain.
- (b) What is decision tree? Explain how classification is done using decision tree induction.
- (c) What is KDD process? Explain each of its steps in detail.

**Q5.** Attempt any two of the following :  $10 \times 2 = 20$

- (a) Discuss the genetic algorithms as data mining techniques. What are the main difficulties in using these techniques?

(b) Write the differences between the Nearest Neighbour Data mining techniques and clustering.

(c) Write short notes on any *three* the following

- i) Web Data Mining
  - ii) Multi Media database
  - iii) Data Visualization
  - iv) Application of Data Mining
-

(Following Paper ID and Roll No. to be filled in your Answer Book)

**PAPER ID : 1452****Roll No.**

<input type="text"/>									
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

**M.C.A.****(SEMESTER-IV) THEORY EXAMINATION, 2011-12****DATA WAREHOUSING & MINING****Time : 3 Hours ]****[ Total Marks : 100****Note :** Attempt all Sections as directed.**SECTION – A****Attempt all parts.** **$10 \times 2 = 20$** **1. Short answer type :**

- (a) Define Data-Mart.
- (b) Write the Inmon definition of Data warehouse.
- (c) Explain meta data and its importance.
- (d) Explain the term crossover, mutation in Genetic algorithm.
- (e) What do you mean by data processing and data cleaning ?
- (f) Discuss for what purpose statistics is used in data mining.
- (g) Is web data analysis and web mining different ? Comment.
- (h) Define Classification and association techniques used to mine the data.
- (i) What is Bottom-up design approach to build a data warehouse ?
- (j) Define the fact table and dimension table of data warehouse.

**SECTION – B****2. Attempt any three questions.** **$3 \times 10 = 30$** 

- (a) What do you mean by client – server model ? What is the need of client – server architecture ? Also, explain 2-tier and 3-tier architecture in detail.
- (b) Draw the data warehouse architecture and briefly explain its various components.
- (c) Explain the following :
  - (i) Data ETL tools.
  - (ii) Reporting & query tools and applications.
- (d) What is a decision tree, show by example ? Describe ID3 algorithm of the decision tree.
- (e) What is data visualization ? How can data visualization help in decision making ?

## **SECTION – C**

**Attempt all questions.**

- 3. Attempt any two parts.  $2 \times 5 = 10$**
- (a) Compare and contrast OLTP system and data warehouse system.
  - (b) Explain multidimensional model of a data warehouse with help of an example.
  - (c) Explain the following :
    - (i) Shared – Disk architecture
    - (ii) Shared – Nothing architecture
- 4. Attempt any two parts.  $2 \times 5 = 10$**
- (a) What is data warehousing ? What can a data warehouse do ? List five common data warehousing applications.
  - (b) Explain with help of suitable example/diagram the three basic conceptual DBMS schemas.
  - (c) What are the different types of database parallelism and what are the data partitioning techniques ?
- 5. Attempt any one part.  $1 \times 10 = 10$**
- (a) Briefly explain and give Dr. E. F. Codd 12 guidelines and requirement as the basis for selecting OLAP systems.
  - (b) What is an OLAP ? Explain MOLAP, ROLAP and Managed Query Environment (MQE), also give architecture for each.
- 6. Attempt any one part.  $1 \times 10 = 10$**
- (a) Define data mining. List and draw the KDD process and explain the steps of the process. Also give application of data mining.
  - (b) Explain the following algorithm of clustering.
    - (i) K – Means
    - (ii) Nearest Neighbour
- 7. Write short notes on any two of the following :  $2 \times 5 = 10$**
- (a) Multimedia Data-Mining
  - (b) Web Data-Mining
  - (c) Neural Networks

(Following Paper ID and Roll No. to be filled in your Answer Book)

PAPER ID : 1452

Roll No.

--	--	--	--	--	--	--	--	--	--	--	--	--	--

## MCA

(SEM IV) EVEN SEMESTER THEORY EXAMINATION,  
2009-2010

### DATA WAREHOUSING & MINING

Time : 3 Hours

Total Marks : 100

Note : (i) Attempt ALL the questions.

(ii) All questions carry equal marks.

1. Attempt any four parts of the following : (4x5=20)

- (a) What do you mean by data mining ? Differentiate between dataware house and data mining.
- (b) What are different sources of information ? Explain the term data, Information and Knowledge with suitable example.
- (c) What is data cleaning ? What are the different data cleaning approaches explain ?
- (d) Explain in brief wavelet transformation.
- (e) Explain data cube aggregation model.
- (f) Explain the concept of statistics in data mining application in brief.

2. Attempt any two of the following : (2x10=20)
- (a) Explain the Dataware House Concept. Explain the two tier dataware house architecture and its advantages.
  - (b) Explain the concept of Client/Server Computing Model. Explain the various generation of Client/Server model in details.
  - (c) Explain the concept of Distributed Dataware houses and What are the advantages of Distributed Dataware house ?
3. Attempt any two parts of the following : (2x10=20)
- (a) Explain OLAP model and its advantages. Explain the various OLAP schema and operations.
  - (b) What are the different types of reports ? Explain each one.
  - (c) Explain Data Marts. What are the different types of Data Marts ? Give advantages of data marts over dataware houses.
4. Attempt any two parts of the following : (2x10=20)
- (a) How are Antecedent and Consequent identified in sequence association rule mining explain ?
  - (b) Compare and contrast the production rules obtained from decision trees with the association rules.
  - (c) What is Clustering and its attributes explain ? Explain K-Mean Clustering Algorithm.

5. Attempt any two parts of the following : (2x10=20)
- (a) What is Data Visualisation and explain various data visualisation techniques. What is a Q-Q Plot ? What data assumptions can you check with it ?
  - (b) What is Web mining ? What are the classification factors of web mining ? What are the access limitations in web mining, Explain it.
  - (c) What is Meta data ? What are the major components of Meta data ? What is the role of Meta data in Dataware house ?

- o O o -