Tribhuvan University

Institute of Science and Technology

MODEL QUESTION

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

Attempt any two Questions **(10 x 2 = 20)**

1. Suppose that a data warehouse for Big University consists of four dimensions: student, course, semester, and instructor, and two measures count and avg-grade. When at the lowest conceptual level (eg. for a given student, course, semester, and instructor combination), the avg-grade measure stores the actual course grade of the student. At higher conceptual levels, avg-grade stores the average grade for the given combination.
   a) Draw a snowflake schema diagram for the data warehouse.
   b) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (eg. roll-up from semester to year) should one performing in order to list the average grade of CS courses for each Big University Student.
   c) If each dimension has five levels (including all), such as "student < major < status < university < all", how many cuboids will this cube contain (including the base and apex cuboids)?

2. A= {A1, A2, A3, A4, A5, A6}, Assume rho = 35%. Use A priori algorithm to get the desired solution.

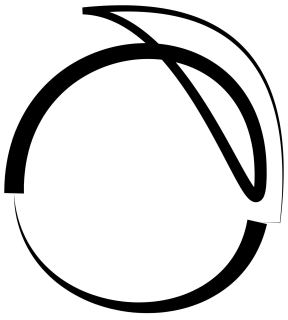
A1	A2	A3	A4	A5	A6
0	0	0	1	1	1
0	1	1	1	0	0
1	0	0	1	1	1
1	1	0	1	0	0
1	0	1	0	1	1
0	1	1	1	0	1
0	0	0	1	1	0
0	1	0	1	0	1
1	0	0	1	0	0
1	1	1	1	1	1

3. What kind of data processing do we need before applying datamining algorithm to any data set? Explain binning method to handle noisy data with example.

Group - B

Attempt any eight Questions. (Question no. 13 is compulsory) **(8 x 5 = 40)**

4. Explain the use of frequent item set generation process.
5. Differentiate between data marts and data cubes.
6. Explain OLAP operations with example.
7. List the drawbacks of ID3 algorithm with over-fitting and its remedy techniques.
8. Write the algorithm for k-means clustering. Compare it with k-nearest neighbor algorithm.
9. What is text mining? Explain the text indexing techniques.
10. Describe genetic algorithm using as a problem solving technique in data mining.
11. What do you mean by WWW mining? Explain WWW mining techniques.
12. What is DMQL? How do you define Star Schema using DMQL?
13. Write short notes (Any Two)
    a) Test Database Mining
    b) Back propagation Algorithm
    c) Regression
    d) HOLAP



Tribhuvan University

Institute of Science and Technology

2069

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

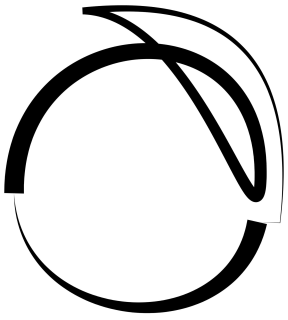
Attempt any two Questions **(10 x 2 = 20)**

1. Explain the architecture of Data mining system with block diagram.
2. Define clustering. Explain with example of the partitioning and hierarchical clustering methods.
3. Explain the architecture and implementation of data warehouse with example.

Group - B

Attempt any eight Questions. (Question no. 13 is compulsory) **(8 x 5 = 40)**

4. What do you mean by knowledge discovery in database (KDD)?
5. Explain the application of data warehouse and data mining.
6. Differentiate between OLAP and OLTP.
7. Explain the data mining techniques.
8. Explain the Aprion Algorithm.
9. Explain the K-Mediod Algorithm.
10. Define the spatial database and its features.
11. What is data cube? Explain with example.
12. Explain the data mining languages.
13. Write short notes (Any Two)
    a) OLAP queries
    b) Snow flakes
    c) K-mean
    d) Mining text databases



Tribhuvan University

Institute of Science and Technology

2071

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

Attempt any two Questions **(10 x 2 = 20)**

1. What do you mean by representative object based clustering technique? Explain in detail with example.

2. Explain the various data mining task primitives in detail.

3. Explain the architecture of data mining system with schematic diagram.

Group - B

Attempt any eight Questions. **(8 x 5 = 40)**

4. What are the basic stages of KDD?

5. Differentiate between DBMS and Data Warehouse.

6. Explain the distributed and virtual data warehouse.

7. Explain the data cube with example.

8. What are the data warehouse back and tools? Explain.

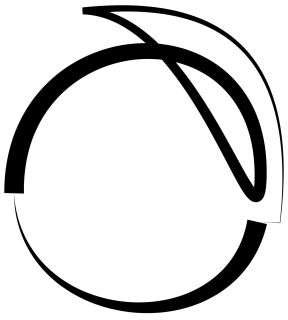
9. Explain the data mining tasks performed on a text database.

10. Define the spatial database and its features.

11. Differentiate between OLTP and OLAP.

12. Explain the Aprion Algorithm.

13. Write short notes (Any Two)
    a) Stars
    b) HOLAP
    c) Data Specification
    d) Mining and world wide web (WWW)



Tribhuvan University

Institute of Science and Technology

2071 (II)

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

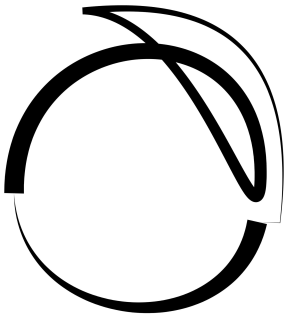
Attempt any two Questions **(10 x 2 = 20)**

1. What are the key steps in knowledge discovery in databases? Explain.
2. Explain the functionalities and classification of data mining system with example.
3. Explain about the architecture and implementation of data warehouse with example.

Group - B

Attempt any eight Questions. **(8 x 5 = 40)**

4. What are the stages of knowledge discovery in database (KDD)?
5. List down the functionality of meta data.
6. Differentiate between OLAP and OLTP
7. Explain the multidimensional data model.
8. List down the data mining tools.
9. Write down the two measures of association rule.
10. What is the objective of K-means algorithm?
11. Explain the application of spatial databases.
12. Explain the methods of mining multimedia database.
13. Write short notes (Any Two)
    a) MOLAP
    b) Data cubes
    c) Snowflakes
    d) Regression



Tribhuvan University

Institute of Science and Technology

2072

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

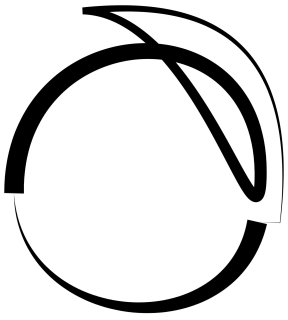
Attempt any two Questions **(10 x 2 = 20)**

1. Differentiate between Data-Warehouse and Data-mining. Explain the stages of knowledge discovery in database with example.
2. What do you mean by clustering? Explain the K-Mean and K-Mediod algorithm with example.
3. Explain the data warehouse architecture. Differentiate between distributed and virtual data warehouse.

Group - B

Attempt any eight Questions. **(8 x 5 = 40)**

4. Explain the multidimensional data model with example.
5. Differentiate between OLTP and OLAP.
6. Explain the tuning and testing of Data Warehouse.
7. Differentiate between KDD and Data Mining.
8. Explain the data mining query language with example.
9. What are the advantages and disadvantages of association rules?
10. What are the types of Regression? Explain.
11. Explain the Aprion Algorithm.
12. Explain the application of mining used in WWW.
13. Write short notes (Any Two)
    a) HOLAP
    b) Hierarchy specification
    c) Spatial Database



Tribhuvan University

Institute of Science and Technology

2073

✡

Bachelor Level/ Fourth Year/ Seventh Semester/ Science — Full Marks: 60
Computer Science and Information Technology (CSc. 451) — Pass Marks: 24
(Data Warehousing and Data Mining) — Time: 3 hours.

Candidates are required to give their answers in their own words as for as practicable.

Group - A

Attempt any two Questions **(10 x 2 = 20)**

1. Differentiate between Data-Warehouse and Data-mining..
2. Explain the DBMS vs. Data Warehouse.
3. Explain the K-mean and K-Mediod Algorithm with example.

Group - B

Attempt any eight Questions. **(8 x 5 = 40)**

4. Differentiate between Data marks and Meta data.
5. What do you mean by virtual data warehouse.
6. Explain the tuning and testing of Data Warehouse.
7. Differentiate between KDD and Data Mining.
8. Explain the data mining query language.
9. Explain the Aprion Algorithm.
10. Explain the types of Regression.
11. Explain the association rules with advantages and disadvantages.
12. Explain mining text databases.
13. Write short notes (Any Two)
    a) Data cubes
    b) HOLAP
    c) Spatial Database