

University of Mumbai

B.E Fourth Year 2013 - 2014 May

Semester 8 (BE Fourth Year)

Data Warehousing and Mining

[www.shaalaa.com](http://www.shaalaa.com)

**TIME - 3 Hrs**

**Marks - 100**

- Note:**
1. Question 1 is compulsory
  2. Answer any 4 out of the remaining questions.
  3. Answers to sub questions must be written together

**Q1.** A bank wants to develop a data warehouse for effective decision-making about their loan schemes. The bank provides loans to customers for various purposes like House Building Loan, Car Loan, Educational Loan, Personal Loan, etc. The whole country is categorized into a number of regions, namely, North, South, East and West. Each region consists of a set of states. Loan is disbursed to customers at interest rates that change from time to time. Also, at any given point of time, the different types of loans have different rates. The data warehouse should record an entry for each disbursement of loan to customer.

- a) Design an information package diagram for the application. (05)
- b) Design a star schema for the data warehouse clearly identifying the fact table(s), Dimensional table(s), their attributes and measures. (05)
- c) Describe an algorithm the bank can use to cluster its potential customers, based on their attributes. (05)
- d) Describe how data warehousing and mining help the bank increase its productivity. (05)

**Q2.** Define the following terms by giving examples

- (a) Fact Constellation
- (b) Snowflake Schema
- (c) Aggregate Fact tables
- (d) Snapshot and Transaction Tables (5 X 4 = 20)

**Q 3.(a)** Consider an online travel agency that helps customers to plan and schedule their holidays. The agency maintains all past history in a data warehouse. Describe the different classes of users who could access this data warehouse and design the information delivery framework for this data warehouse. (10)

(b) Describe the working of the K-Means clustering algorithm with the help of the a sample dataset. (10)

**Q 4.(a)** Consider a data warehouse for a hospital, where there are three dimensions:

(1) Doctor (2) Patient (3) Time; and two measures: (1) Count & (2) Fees;

For this example create a OLAP cube and describe the following OLAP operations:

- (1) Slice (2) Dice (3) Rollup (4) Drill Down (5) Pivot (10)

(b) Consider the following transaction database:

TID	Items
01	A, B, C, D
02	A, B, C, D, E, G
03	A, C, G, H, K
04	B, C, D, E, K
05	D, E, F, H, L
06	A, B, C, D, L
07	B, I, E, K, L
08	A, B, D, E, K
09	A, E, F, H, L
10	B, C, D, F

Apply the Apriori algorithm with minimum support of 30% and minimum confidence of 70%, and find all the association rules in the data set. (10)

Q 5.(a)

Transaction	Income	Credit	Decision
1	Very High	Excellent	AUTHORIZE
2	High	Good	AUTHORIZE
3	Medium	Excellent	AUTHORIZE
4	High	Good	AUTHORIZE
5	Very High	Good	AUTHORIZE
6	Medium	Excellent	AUTHORIZE
7	High	Bad	REQUEST ID
8	Medium	Bad	REQUEST ID
9	High	Bad	REJECT
10	Low	Bad	CALL POLICE

Using the above table illustrate any one classification technique. Further indicate how we can classify a new transaction, with (Income = Medium and Credit=Good). (10)

(b) (a) Describe clearly the different steps of the ETL (Extract - Transform - Load) cycle in Data Warehousing (10)

Q6.(a) Give a brief description of web mining (10)

(b) Explain clearly the role of Meta data in a data warehouse. (10)

Q7. Write detailed notes on:-

(a) Data Warehouse Architecture

(b) Hierarchical Clustering methods.

(10 X 2 = 20)