

Elective II - Data warehousing & mining

Con. 3061-07.

(Library)

[REVISED COURSE]

ND-827

(3 Hours)

[Total Marks : 100]

- N.B 1) Question No.1 is Compulsory.
 2) Attempt any Four out of remaining
 3) Figure to the right indicates full Marks.

Q1. (A) 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. With respect to the above business scenario,

- (1) Design an information package diagram. Clearly explain all aspects of the diagram (5)
 (2) Draw a star schema for the data warehouse clearly identifying the Fact table(s), Dimension table(s), their attributes and measures. (5)

(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	A, D, F, L
08	B, I, E, K, L
09	A, B, D, E, K
10	C, D, H, I, K
11	A, E, F, H, L
12	B, C, D, F
13	A, B, C, D
14	A, D, H, K
15	B, C, D, E, H, L

Apply the Apriori algorithm with minimum support of 30% and minimum confidence of 75%, and find all the association rules in the data set. (10)

Q2. Define the following by giving examples

- (a) Factless Fact Tables
 (b) Snowflake Schema
 (c) Outliers in data mining
 (d) Supervised learning in data mining
 (e) Family of stars

(5 X 4 = 20)

[TURN OVER]

Q3. (a) Consider a data warehouse for a hospital, where there are three dimensions:

(1) Doctor, (2) Patient, and (3) Time, and two measures: (1) Count and (2) Charge, where charge is the fee that the doctor charges a patient for a visit.

Using the above example describe the following OLAP operations

(1) Slice (2) Dice (3) Rollup (4) Drill Down (5) Pivot **(5 X 2 = 10)**

(b) Describe the different clustering algorithms. Discuss the advantages and disadvantages of each. **(10)**

Q4. (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. **(8)**

(b) What is web mining? Illustrate the working of the HITS algorithm by using an Example query on a search engine (example – search for “web mining applications”) **(12)**

Q5. (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) Describe the ETL cycle in a data warehouse **(10)**

Q.6. (a) What are concept hierarchies? Explain with an example. Describe the concept hierarchy using DMQL **(8)**

(b) What is the importance of metadata in a data warehouse? What are the different types of metadata stored in a data warehouse? Illustrate with a simple Customer Sales data warehouse. **(12)**

Q.7. (a) With a neat diagram describe the KDD process **(8)**

(b) Discuss the importance of visualization in a data warehouse and in data mining **(12)**