Data warehousing & Mining   BE comp

Sem VIII

Q.P. Code : 733600   CBGS

(3 Hours)   [Total Marks : 80

Note: 1. Question No.1 is compulsory
2. Attempt any **Three** questions out of remaining questions
3. Assume suitable data wherever necessary and state them clearly

Q1 a) For a Super market chain, consider the following dimensions namely product, store, [10] time and promotion. The schema contains a central fact table for sales.
   i. Design star schema for the above application.
   ii. Calculate the maximum number of base fact table records for warehouse with the following values given below:
   • Time period – 5 years
   • Store – 300 stores reporting daily sales
   • Product – 40,000 products in each store (about 4000 sell in each store daily)

b) Discuss: [10]
   i. The steps in KDD process
   ii. The architecture of a typical DM system

Q2 a) We would like to view sales data of a company with respect to three dimensions namely [10] Location, Item and Time. Represent the sales data in the form of a 3-D data cube for the above and Perform Roll up, Drill down, Slice and Dice OLAP operations on the above data cube and Illustrate.

b) A simple example from the stock market involving only discrete ranges has profit as [10] categorical attribute, with values {Up, Down} and the training data set is given below.

Age	Competition	Type	Profit
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

Apply decision tree algorithm and show the generated rules.

Q3 a) Illustrate the architecture of a typical DW system. Differentiate DW and Data Mart. [10]

b) Discuss different steps involved in Data Preprocessing. [10]

Q4 a) Discuss various OLAP Models. [10]

b) Explain K-Means clustering algorithm? Apply K-Means algorithms for the following [10] data set with two clusters. Data Set = {1, 2, 6, 7, 8, 10, 15, 17, 20}

TURN OVER

Q5 a) Describe the steps of ETL process.   [10]

  b) Discuss Association Rule Mining and Apriori Algorithm. Apply AR Mining to find all   [10]
      frequent item sets and association rules for the following dataset:
      Minimum Support Count = 2
      Minimum Confidence = 70%

Transaction ID	Items
100	1, 2, 5
200	2, 4
300	2, 3
400	1, 2, 4
500	1, 3
600	1, 3
700	1, 3, 2, 5
800	1, 3
900	1, 2, 3

Q6 Write Short notes on any four of the following: [20]

  i.   Updates to Dimension tables

  ii.   Metrics for Evaluating Classifier Performance

  iii. FP tree

  iv.   Multilevel & Multidimensional Association Rule

  v.   Operational Vs. Decision Support System

_____