

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.

