Printed Pages : 2 Roll No. | | | | | | | | | | | NMCAE13

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.

(iii) Is this a distributive multi-feature cube? Justify.

(g) Summarize the issues involved in mining and extracting information from multimedia databases.

(h) Enumerate some of the technology integration issues and uses of digital libraries.

SECTION – C

Attempt any two of the following questions: **2 x 15 = 30**

3. **(a)** Design multi-dimensional data model for hospital data warehouse, consist three dimensions time, doctor, and patient and the two measures count and charge, where charge is a fee that a doctor charges patients for a visit.

(i) Enumerate three classes of schema that are popularly used for modeling data warehouses.

(ii) Draw a schema diagram for the above data warehouse using all of the schema classes.

(b) How to reduce the size of the fact table? Explain with example.

4. **(a)** How data mining system can be integrated with a data warehouse?

(b) Suppose that the data for analysis include the attribute 'age'. The age values for the data tuples are: 13, 15, 16, 16, 19, 20, 21, 22, 22, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(i) Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.

(ii) How would you determine outliers in the data?

5. Describe the architecture and functionality of MM-DBMS.