MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

IT-704C

DATA WAREHOUSING AND DATA MINING

Time Allotted: 3 Hours | Full Marks: 70

The questions are of equal value.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable.
All symbols are of usual significance.

GROUP A
(Multiple Choice Type Questions)

1. Answer *all* questions. 10×1 = 10

(i) Consider the 3-tier architecture of the data warehouse. The OLAP engine corresponds to

(A) the first layer (B) the second layer
(C) the third layer (D) none of these

(ii) A data warehouse is said to contain a "time-varying" collection of data because

(A) its contents vary automatically with time
(B) its life-span is very limited
(C) it contains historical data
(D) its content has explicit time-stamp

(iii) Dimension data within a Warehouse exhibits one of the following properties

(A) dimension data consists of the minor part of the warehouse
(B) the aggregated info is actually dimension data
(C) it consists historical data
(D) dimension data is the info that is used to analyze the elemental transaction

(iv) K-means is based on

(A) Euclidean distance (B) RMS
(C) Hamming distance (D) None of these

(v) A data warehouse is an "integrated" collection of data because

(A) it is a collection of data of different types
(B) it is a relational database
(C) it is a collection of data derived from multiple sources
(D) it contains summarized data

(vi) Which one if not a data mining task?

(A) indexing (B) prediction (C) clustering (D) regression

(vii) Which is not a schema for multidimensional database?

(A) Star (B) Snowflake
(C) Fact constellation (D) Transaction

(viii) In order to populate the data warehouse which of the following set of operations are appropriate?

(A) refresh and load (B) create and edit
(C) insert and delete (D) query and update

(ix) ROLAP is preferred over MOLAP when

(A) A data warehouse and relational database are inseparable
(B) The data warehouse is in relational tables, but no slice and dice operation are required
(C) The multidimensional model does not support query optimization
(D) A data warehouse contains many fact tables and many dimension tables

(x) The slice operation deals with

(A) selecting all but one dimension of the data cube
(B) merging the cells along one dimension
(C) merging cells of all but one dimension
(D) selecting the cells of any one dimension of the data cube

GROUP B
(Short Answer Type Questions)

Answer any *three* questions. 3×5 = 15

2. (a) What is metadata? 2
   (b) What is the typical content of metadata of a data warehouse? 3

3. (a) What is a Data Mart? 2
   (b) State the differences between Data Mart and Data Warehouse. 3

4. Explain temporal data mining with an example. How is it different from spatial data mining? 5

5. Differentiate between Hierarchical clustering and Partitioning clustering technique. 5

6. (a) What do you mean by multidimensional data model? 2
   (b) How it can be represented by data cube? 3

GROUP C
(Long Answer Type Questions)

Answer any *three* questions. 3×15 = 45

7. (a) Define a frequent set. Show that every subset of any item set must contain either a frequent set or a border set. 6
   (b) Define support and confidence. 2

HTTP://WWW.MAKAUT.COM

CS/B.Tech/IT/Odd/Sem-7th/IT-704C/2015-16

(c) Find all the frequent sets using A Priori algorithm of the following database: 7

	A1	A2	A3	A4	A5	A6	A7	A8	A9
1	1	0	0	0	1	1	0	1	0
2	0	1	0	1	0	0	0	1	0
3	0	0	0	1	1	0	1	0	0
4	0	1	1	0	0	0	0	0	0
5	0	0	0	0	1	1	1	0	0
6	0	1	1	1	0	0	0	0	0
7	0	1	0	0	0	1	1	0	1
8	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	1	0
10	0	0	1	0	1	0	1	0	0
11	0	0	0	0	1	1	0	1	0
12	0	1	0	1	0	1	1	0	0
13	1	0	1	0	1	0	1	0	0
14	0	1	1	0	0	0	0	0	1
15	0	0	1	0	1	0	1	0	0

Assume θ = 20%.

8. (a) Define Decision Tree. 2
   (b) What are the advantages and disadvantages of the decision tree approach over other approaches for data mining? 4+5
   (c) Discuss briefly the tree construction principle. 4

9. (a) Write down the design steps for a typical data warehouse? 6
   (b) Explain the flowchart for KDD process. 5
   (c) Explain Roll-up and Drill-down process with a suitable example. 4

10. Explain Star schema model with a suitable example. Mention its advantage and disadvantages. What is Factless Fact Table? 7+5+3

11. Write short notes on any *three* of the following: 3×5
   (a) HOLAP
   (b) Web mining
   (c) Text Mining
   (d) Data warehouse Schema
   (e) Essbase Arbor

HTTP://WWW.MAKAUT.COM

HTTP://WWW.MAKAUT.COM