ROLL NO. _____

Code: AT19   **Subject: DATA WAREHOUSING AND DATA M**

AMIETE – ET (OLD SCHEME)

Time: 3 Hours   **JUNE 2012**   **Max. Marks: 100**

PLEASE WRITE YOUR ROLL NO. AT THE SPACE PROVIDED ON EACH PAGE IMMEDIATELY AFTER RECEIVING THE QUESTION PAPER.

NOTE: There are 9 Questions in all.
- **Question 1 is compulsory and carries 20 marks. Answer to Q.1 must be written in the space provided for it in the answer book supplied and nowhere else.**
- **The answer sheet for the Q.1 will be collected by the invigilator after 45 Minutes of the commencement of the examination.**
- **Out of the remaining EIGHT Questions answer any FIVE Questions. Each question carries 16 marks.**
- **Any required data not explicitly given, may be suitably assumed and stated.**

Q.1   **Choose the correct or the best alternative in the following:**   **(2×10)**

a.   Which of the following is not a Data Mining Functionality?

   (A) Association analysis   **(B)** Cluster Analysis
   (C) Classification & Prediction   **(D)** Randomization

b.   Which of the following is not an OLAP server?

   (A) ROLAP   **(B)** DOLAP
   (C) MOLAP   **(D)** HOLAP

c.   Which of the following is a data smoothing technique?

   (A) Histogram   **(B)** Regression
   (C) Correlation **(D)** Induction

d.   Which of the following is not a basis for classification for an association rule?

   (A) types of values   **(B)** dimensions of data involved
   (C) volume of data involved **(D)** levels of abstractions involved

e.   In which of the following areas the data mining technique cannot be applied?

   (A) Medical-diagnosis   **(B)** Prediction in share market
   (C) Retail Industry   **(D)** all of these

Code: AT19   **Subject: DATA WAREHOUSING AND DATA MI**

f.   The generic two-level data warehouse architecture includes which of the following?

(A) At least one data mart.
(B) Data that can extracted from numerous internal and external sources.
(C) Near real-time updates.
(D) All of the above

g.   In the data warehouse architecture, the _____ component interleaves with and connects other components.

(A) Metadata   (B) Data cube
(C) Data mart   (D) none of the above

h.   A star schema has what type of relationship between a dimension and fact table?

(A) Many-to-many   (B) One-to-one
(C) One-to-many   (D) All of the above

i.   Which of the following is the extract process?

(A) Capturing all of the data contained in various operational systems
(B) Capturing a subset all of the data contained in various operational system
(C) Capturing all of the data contained in various decision support system
(D) Capturing a subset of the data contained in various decision support system

j.   _____ stores multidimensional aggregate information.

(A) Data cube   (B) Data Mart
(C) Both (A) & (B)   (D) None of the above

Answer any FIVE Questions out of EIGHT Questions.
Each question carries 16 marks.

Q.2   a. Define a data warehouse elaborating its key features. How do the organizations benefits from it?   **(9)**

b. What are the major features that differentiate OLTP from OLAP?   **(7)**

Q.3   a. What are the features of external/unstructured data that pose problems while storing it in the data warehouse? Describe an effective technique for handling unstructured data.   **(9)**

b. Mean and standard deviation of the values for an attribute income are given as \$54,000 and \$16,000 respectively. Using z-score normalization transform a value of \$73,600 for income.   **(7)**

Q.4 a. Use diagram to explain the path of migration from corporate data model to a DSS.   **(8)**

b. Discuss in brief three data warehouse models-enterprise warehouse, data mart and the virtual warehouse.   **(8)**

Q.5 a. Discuss the list of technological challenges includes in data migration methodology.   **(8)**

b. Discuss in brief four levels in architected environment.   **(8)**

Q.6 a. Briefly describe the three problems with naturally evolving architecture.   **(6)**

b. Explain data transformation with following:
   (i) Smoothing   (ii) Aggregation
   (iii) Generalization   (iv) Normalization
   (v) Attribute construction   **(10)**

Q.7 a. What do you mean by association rule mining? Explain constraint-based association mining.   **(8)**

b. What is the relationship between the data model and external data?   **(8)**

Q.8 a. Explain how the data warehouse forms a basis for EIS.   **(6)**

b. Explain drill-down analysis and event mapping in context of EIS.   **(10)**

Q.9   Write a short note on (Any **FOUR**):

(i) Clustering
(ii) Feedback loop Technique
(iii) Archiving External Data
(iv) Data Cube Technology
(v) Criterion for comparing classification methods.   **(4×4)**