

Time: 3 Hours

JUNE 2014

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 table contains foreign keys and numeric fact values?
- (A) Fact Table (B) Dimension Table
(C) Factless fact table (D) None of these
- b. Which technique allows more data to be loaded into a single block?
- (A) Compaction (B) Indexing
(C) Partitioning (D) Clustering
- c. The generic two-level data warehouse architecture includes which of the following?
- (A) At least one data mart
(B) Data that can be extracted from numerous internal and external sources
(C) Near real-time updates
(D) All of these
- d. Which of the following data mining technique is used for optimization?
- (A) Artificial Neural Networks (B) If then rule induction
(C) Genetic algorithms (D) Decision tree
- e. Adding value to the data to give it more meaning is called _____.
- (A) Data cleansing (B) Data profiling
(C) Data integration (D) Data enrichment
- f. Which of the following schema supports the normalization in dimensional modeling?
- (A) Star schema (B) Snow-Flake schema
(C) Fact-Constellation (D) None of these

Code: AT78

Subject: DATA MINING & WAREHOUSING

- g. Data Warehouse is defined as subject oriented, integrated, time variant and _____.
- (A) Application oriented (B) Information oriented
(C) Analysis oriented (D) All of these
- h. Data scrubbing is which of the following?
- (A) A process to reject data from the data warehouse and to create the necessary indexes
(B) A process to load the data in the data warehouse and to create the necessary indexes
(C) A process to upgrade the quality of data after it is moved into a data warehouse
(D) A process to upgrade the quality of data before it is moved into a data warehouse
- i. A multifield transformation does which of the following?
- (A) Converts data from one field into multiple fields
(B) Converts data from multiple fields into one field
(C) Converts data from multiple fields into multiple fields
(D) All of these
- j. Which operation performs data aggregation by climbing up a dimensional hierarchy?
- (A) Slice (B) Dice
(C) Roll-up (D) Drill-down

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

- Q.2** a. With the help of suitable figure explain the architecture of a typical data mining system. **(8)**
- b. List and describe the five primitives for specifying a data mining task. **(8)**
- Q.3** a. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
- (i) What is the mean of the data? What is the median?
- (ii) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal etc.)
- (iii) What is the midrange of the data?
- (iv) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
- (v) Give the five-number summary of the data. **(5×2 = 10)**

- b. Why is data cleansing and data transformation functions considered to be a vital task in the integration process. Explain these functions and also mention where these functions take place. (3+3)
- Q.4** a. Explain the architecture of a data warehouse. Also explain the single-tier and three tier architectures of a data warehouse. (8)
- b. You are the senior analyst responsible for selecting the tools in your data warehouse. Make a list of the tools you will provide for use by the developers and the end users of your data warehouse. Describe the features of these tools. (2+6)
- Q.5** a. Write a short note on fact less fact table. Draw a star schema representing a factless fact table of a patient visiting a hospital. (3+3)
- b. Give a description on the methods used in data cube implementation. Explain with examples. (6)
- c. Give reasons why feeding data into the OLAP system directly from the source system is not preferred. Explain. (4)
- Q.6** a. Explain the working of decision tree algorithm with the help of an example. Write down the advantages and limitations of the technique. (8)
- b. Discuss the different criteria based on which frequent pattern mining can be classified. (8)
- Q.7** a. Prediction is frequently referred to as the forecasting of missing numerical values. Justify the statement with an example. (6)
- b. Give the underlying principle of neural networks. State its advantages and shortcomings. What are the applications of this technique in today's scenario? (6)
- c. With the help of example, explain post-pruning. (4)
- Q.8** a. What is wave cluster? Why is wavelet transformation useful for clustering?(8)
- b. Classify and explain hierarchical clustering techniques. Also differentiate between the approaches. (8)
- Q.9** a. How can data mining be used for customer retention, target marketing and CRM? Explain with examples. (6)
- b. Why is data mining and warehousing applied in biomedical area? Explain. (6)
- c. Give a brief note on time series analysis and its relevance in data mining. (4)