

AMIETE – IT (OLD SCHEME)

Code: AT19
Time: 3 Hours

Subject: DATA WAREHOUSING AND DATA MINING
Max. Marks: 100

DECEMBER 2010

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 half an hour 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)

- Which of the following is not true about a data warehouse?
(A) Can be updated by users.
(B) Contains numerous naming conventions and formats.
(C) Organized around important subject areas.
(D) Contains only current data.
- Which of the following is an operational system?
(A) A system that is used to run the business in real time and is based on historical data.
(B) A system that is used to run the business in real time and is based on current data.
(C) A system that is used to support decision making and is based on current data
(D) A system that is used to support decision making and is based on historical data
- The active 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
- Data transformation includes which of the following?
(A) A process to change data from a detailed level to a summary level
(B) A process to change data from a summary level to detailed level.
(C) Joining data from source into various sources of data.
(D) Separating data from one source into various sources of data.
- A goal of data mining includes which of the following?
(A) To explain some observed event or condition
(B) To confirm that data exists
(C) To analyze data for expected relationships
(D) To create a new data warehouse

- f. Which of the following types of tables better describes a snowflake schema?
 (A) Fact (B) Dimension
 (C) Helper (D) All of the above
- g. Which one of the following terms does not describe OLAP
 (A) The gathering of input information
 (B) processing input information
 (C) Updating existing information to reflect to the gathered and processed information
 (D) None of the above
- h. _____ includes normalization and aggregation as data preprocessing procedures.
 (A) Data transformation (B) Data cleansing
 (C) Data reduction (D) Data integration
- i. Fact tables are which of the following?
 (A) Completely demoralized (B) partially demoralized
 (C) Completely normalized (D) partially normalized
- 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. Compare and contrast the system development life cycle for data warehouse with classical SDLC. (8)
 b. Discuss the four levels of data in the architected dataware environment? (8)
- Q.3** a. Differentiate between
 (i) OLAP and OLTP.
 (ii) Data warehouse and Data mart (4+4)
 b. Explain snowflake schema with the help of examples. What are the advantages and disadvantages of snowflake schema? (8)
- Q.4** 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.5** a. Discuss the problem related to use and storage of unstructured data in data warehouse? What are the methods to capture and store external information? (8)
 b. Explain metadata in context of External data? (8)

- Q.6** a. What are the strategies of data reduction?
b. What is data preprocessing? What are various steps involved in data preprocessing? (8)
- Q.7** a. What all the technology changes are considered while bringing the system-of-record data into the data warehouse? (8)
b. Write the significance of feedback loop for the success of data warehouse implementation? (8)
- Q.8** a. Discuss the criteria used to compare and evaluation of the classification and prediction methods. (8)
b. What are the typical requirement of clustering in data mining? (8)
- Q.9** a. What is association rules and how it helps in MSQL? (8)
b. Differentiate between data mining & statistics in context of Structured Query Machine Learning. (8)