

Roll No. ....

**MCA-11 (Master of Computer Application)  
6<sup>th</sup> Semester Examination-2015**

**MCA-22A**

**Data Mining and Data Warehousing**

**Time : 3 Hours**

**Maximum Marks : 60**

**Note : This paper is of sixty (60) marks divided into three (03) sections A, B, and C. Attempt the questions contained in these sections according to the detailed instructions given therein.**

**Section - A**

**(Long Answer Type Questions)**

**Note : Section 'A' contains four (04) long-answer-type questions of fifteen (15) marks each. Learners are required to answer any two (02) questions only. (2×15=30)**

1. Explain various methods of data cleaning in detail.
2. What is Cluster Analysis? Explain different types data in cluster analysis.
3. Describe the data warehouse architecture with neat diagram.
4. Discuss the various OLAP operations in the Multidimensional Data Model.

## Section - B

### (Short Answer Type Questions)

**Note : Section 'B' contains eight (08) short-answer-type questions of five (05) marks each. Learners are required to answer any four (04) questions only. (4×5=20)**

1. Differentiate classification and clustering.
2. What is data mining? What are the advantages of data mining?
3. Define strong association rule.
4. How is a data warehouse different from a database?
5. List the primitives that specify a data mining task.
6. What are the technologies used in data mining?
7. What is data warehouse? What are the goals of a data warehouse?
8. Discuss major issues of data mining.

## Section - C

### (Objective Type Questions)

**Note : Section 'C' contains ten (10) objective-type questions of one (01) mark each. All the questions of this section are compulsory. (10×1=10)**

**Select the correct option :**

1. 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

2. A data warehouse is which of the following?
  - (a) Can be updated by end users.
  - (b) Contains numerous naming conventions and formats.
  - (c) Organized around important subject areas.
  - (d) Contains only current data.
3. 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 the above.
4. Which of the following features usually applies to data in a data warehouse?
  - (a) Data are often deleted
  - (b) Most applications consist of transactions
  - (c) Data are rarely deleted
  - (d) Relatively few records are processed by applications
5. Which of the following statements is true?
  - (a) The data warehouse consists of data marts and operational data
  - (b) The data warehouse is used as a source for the operational data
  - (c) The operational data are used as a source for the data warehouse
  - (d) All of the above
6. The following technology is not well-suited for data mining:
  - (a) Expert system technology
  - (b) Data visualization

- (c) Technology limited to specific data types such as numeric data types
  - (d) Parallel architecture
7. Data cleaning is :
- (a) Large collection of data mostly stored in a computer system
  - (b) The removal of noise errors and incorrect input from a database
  - (c) The systematic description of the syntactic structure of a specific database. It describes the structure of the attributes the tables and foreign key relationships.
  - (d) None of the above
8. OLAP stands for \_\_\_\_\_ .
- (a) Online Analytical Processing.
  - (b) Online Linear Analytical Processing.
  - (c) Online Animated Process.
  - (d) Online Analytical Problem.
9. Query tools and data mining tools are \_\_\_\_\_ .
- (a) Same
  - (b) Different
  - (c) Complementary
  - (d) Standard
10. A data warehouse is an integrated collection of data because \_\_\_\_\_ .
- (a) It is a collection of data of different data types.
  - (b) It is a collection of data derived from multiple source.
  - (c) It is a relational database.
  - (d) It contains summarized data.