

(PART-B : Descriptive)

Time : 2 hrs. 40 min.

Marks : 50

[Answer question no.1 & any four (4) from the rest]

1. What is Multi dimensional Data model? Define different OLAP operations with examples. 4+6=10
2. What is Data Mining? Define different stages of KDD. Compare data mining and KDD. 4+4+2=10
3. Write about Association rule of mining. Write about the Apriori Algorithm. 5+5=10
4. Define Warehouse schema. Describe the different types of schemas. 3+7=10
5. Write about the different clustering methods. Describe k-means algorithm for partitioning. 6+4=10
6. Write about FP-tree growth algorithm. Explain with suitable example. 5+5=10
7. What is classification? Define general approach to classification. Describe decision tree induction. 3+3+4=10
8. Write short notes on *any two*: 5+5=10
 - a) Meta Data.
 - b) Data Mart.
 - c) Web Mining.
 - d) Complex Data Mining.

== *** ==

MASTER OF COMPUTER APPLICATION
FOURTH SEMESTER (Repeat)
DATA MINING & WAREHOUSING
MCA-402

(Use separate answer scripts for Objective & Descriptive)

Duration : 3 hrs.

Full Marks : 70

(PART-A : Objective)

Time : 20 min.

Marks : 20

Choose the correct answer from the following:

1x20=20

1. The full form of KDD is.....
 - a. Knowledge Database
 - b. Knowledge Discovery Database
 - c. Knowledge Data House
 - d. Knowledge Data Definition
2. Classification is:
 - a. A subdivision of a set of examples into a number of classes.
 - b. A measure of the accuracy, of the classification of a concept that is given by a certain theory.
 - c. The task of assigning a classification to a set of examples.
 - d. None of these.
3. The output of KDD is.....
 - a. Data
 - b. Information
 - c. Query
 - d. Useful information
4. Bayesian classifiers is:
 - a. Ordinary queue A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.
 - b. Any mechanism employed by a learning system to constrain the search space of a hypothesis.
 - c. An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.
 - d. None of these.
5. The extract process is which of the following?
 - a. Capturing all of the data contained in various operational systems
 - b. Capturing a subset of the data contained in various operational systems
 - c. Capturing all of the data contained in various decision support systems
 - d. Capturing a subset of the data contained in various decision support systems
6. Fact tables are which of the following?
 - a. Completely denormalized
 - b. Partially denormalized
 - c. Completely normalized
 - d. Partially normalized
7. A star schema has what type of relationship between a dimension and fact table?
 - a. Many-to-many
 - b. Many-to-one
 - c. One-to-one
 - d. Array implementation of linked list One-to-one.

8. Data transformation includes which of the following?
- A process to change data from a detailed level to a summary level.
 - A process to change data from a summary level to a detailed level.
 - Joining data from one source into various sources of data.
 - Separating data from one source into various sources of data.
9. Cluster is:
- group of similar objects that differ significantly from other objects.
 - operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm.
 - symbolic representation of facts or ideas from which information can potentially be extracted.
 - none of these.
10. The data is stored, retrieved and updated in.....
- OLAP
 - OLTP
 - SMTP
 - FTP
11. Multi-dimensional knowledge is:
- a class of learning algorithms that tries to derive a program from examples.
 - a table with n independent attributes can be seen as an n-dimensional space.
 - a prediction made using an extremely simple method, such as always predicting the same output.
 - none of these.
12. Prediction is:
- the result of the application of a theory or a rule in a specific case.
 - one of several possible entries within a database table that is chosen by the designer as the primary means of accessing the data in the table.
 - discipline in statistics that studies ways to find the most interesting projections of multi-dimensional spaces.
 - none of these.
13.is a subject-oriented, integrated, time-variant, nonvolatile collection or data in support of management decisions.
- Data Warehousing
 - Data Mining
 - Document Mining
 - Text Mining
14. What is/are the different types of Meta data?
- Administrative
 - Business
 - Operational
- Only (I) above
 - Both (II) and (III) above
 - Both (I) and (III) above
 - All (I), (II) and (III) above
15. The exposes the information being captured, stored, and managed by operational systems.
- Top-down view
 - Data warehouse view
 - Data source view
 - Business query view
16. The Data Warehouse is.....
- Read only
 - Write only
 - Read write only
 - None

17. The apriori property means:
- if a set cannot pass a test, all of its supersets will fail the same test as well.
 - to improve the efficiency the level-wise generation of frequent item sets.
 - if a set can pass a test, all of its supersets will fail the same test as well.
 - to decrease the efficiency the level-wise generation of frequent item sets.
18. Which of the following is not a data mining functionality?
- Characterization and discrimination
 - Classification and regression
 - Selection and interpretation
 - Clustering and analysis
19. An operational system is.....
- used to run the business in real time and is based on historical data.
 - used to run the business in real time and is based on current data.
 - used to support decision making and is based on current data.
 - used to support decision making and is based on historical data.
20.is the heart of the warehouse.
- Data mining database servers
 - Data warehouse database servers
 - Data mart database servers
 - Relational data base servers