

Professional Course Examination, November/December 2019

(5th Semester)

BACHELOR OF COMPUTER APPLICATIONS

Course No. : BCA-5E3

(Data Mining and Warehousing)

Full Marks : 75

Time : 3 hours

(PART : A—OBJECTIVE)

(Marks : 25)

The figures in the margin indicate full marks for the questions

SECTION—A

(Marks : 15)

Tick (✓) the correct answer in the brackets provided :

1×10=10

1. The technique in which existing heterogeneous segments are reshuffled, relocated into homogeneous segments is

- (a) clustering () (b) aggregation ()
(c) segmentation () (d) partitioning ()

2. Classification rules are extracted from

- (a) root node () (b) decision tree ()
(c) siblings () (d) branches ()

3. Web content mining describes the discovery of useful information from the _____ contents.

- (a) text () (b) web ()
(c) page () (d) level ()

4. In web mining, _____ is used to know which URLs tend to be requested together.
- (a) clustering ()
 - (b) sequential analysis ()
 - (c) associations ()
 - (d) classification ()
5. A data warehouse is
- (a) updated by end users ()
 - (b) contains numerous naming conventions and formats ()
 - (c) organized around important subject areas ()
 - (d) contains only current data ()
6. _____ databases are owned by particular departments or business groups.
- (a) Informational ()
 - (b) Operational ()
 - (c) Both informational and operational ()
 - (d) Flat ()
7. The important aspect of the data warehouse environment is that data found within the data warehouse is
- (a) subject-oriented ()
 - (b) time-variant ()
 - (c) integrated ()
 - (d) All of the above ()
8. _____ defines the structure of the data held in operational databases and used by operational applications.
- (a) Operational metadata ()
 - (b) Data warehouse metadata ()
 - (c) User-level metadata ()
 - (d) Data mining metadata ()
9. Fact tables are
- (a) completely demoralized ()
 - (b) partially demoralized ()
 - (c) completely normalized ()
 - (d) partially normalized ()
10. The star schema is composed of _____ fact table.
- (a) one ()
 - (b) two ()
 - (c) three ()
 - (d) four ()

State whether the following statements are *True (T)* or *False (F)* by putting a Tick (✓) mark in the bracket provided : 1×5=5

1. The output of KDD is useful information.

(T / F)

2. The link web propose a measure of standing a node based on path counting.

(T / F)

3. Data warehouse architecture is based on RDBMS.

(T / F)

4. The data is stored, retrieved and updated in OLTP.

(T / F)

5. The dimension tables describe the keys.

(T / F)

SECTION—B

(Marks : 10)

Answer the following questions : 2×5=10

1. What are the requirements of cluster analysis?

2. What is the difference between a static web page and a dynamic web page?

3. What are the scopes of data warehouse?

4. Define data mart.

5. What is a multidimensional database?

(PART : B—DESCRIPTIVE)

(Marks : 50)

The figures in the margin indicate full marks for the questions

1. (a) Describe the issues and challenges in the implementation of data mining systems. 6
(b) What are the steps of knowledge discovery in databases? 4

OR

- (c) Explain the association rules of data mining systems. 10
2. (a) Describe the three categories of web mining techniques. 10

OR

- (b) Explain the different activities of text mining. 6
(c) What are the steps of a web personalization process? 4
3. (a) Illustrate the different types of data warehouse. 10

OR

- (b) Write short notes on the following : 5+5=10
(i) Online Transaction Processing (OLTP)
(ii) Online Analytical Processing (OLAP)
4. (a) Discuss the data warehouse architecture with diagram. 10

OR

- (b) With the help of a neat sketch, explain the various components in a data warehousing system. 10
5. (a) What is star schema? What are the component tables? 2+2=4
(b) What are the advantages and disadvantages of Snowflake? 3+3=6

OR

- (c) What are aggregate fact tables? Why are they needed? Give an example. 3+4+3=10
