Roll No. ...210313006

Total Pages : 4

312304

December 2022

BCA (OS)-III SEMESTER

Data Warehouse and Data Mining (BCA-DS-204)

Time : 3 Hours] [Max. Marks : 75

Instructions :

1. *It is compulsory to answer all the questions (1.5 marks each) of Part-A in short.*
2. *Answer any four questions from Part-B in detail.*
3. *Different sub-parts of a question are to be attempted adjacent to each other.*
4. *Assume data wherever required.*

PART-A

1. (a) Differentiate OLAP systems with typical OLTP systems. (1.5)

   (b) What is metadata repository in data warehousing? (1.5)

   (c) What is meant by concept hierarchy? Explain its need. (1.5)

   (d) What do you mean by Bitmap indexing? (1,5)

(e) Describe various methods for data cube materialization. (1.5)

(f) Differentiate between ROLLUP and DRILLDOWN operations of data warehouse. (1.5)

(g) What is meant by Data Marts? What are its types? (1.5)

(h) How we can find center and radius of a cluster? (1.5)

(i) What is the difference between supervised and unsupervised learning? (1.5)

(j) Why data preprocessing is an important issue for both data warehousing and data mining? (1.5)

PART-B

2. (a) Explain three tier data warehouse architecture with the help of an explanatory diagram. (10)

(b) What is the difference between ROLAP, MOLAP and HOLAP servers? (5)

3. (a) Describe in detail the concepts behind clustering. Also explain why k-medoids algorithm is better than k-means algorithm? (10)

(b) Describe various steps of KDD in detail. (5)

4. (a) Suppose that a Data Warehouse for a Big-University consists of four dimensions student, course, semester and instructor and two measures count and avg_grade. When at the lowest conceptual level (e.g. for a given student, course, semester and instructor combination), the avg_grade measure stores the actual course grade of the student. At higher conceptual levels, avg_grade stores average grade for the given combination.

(i) Draw the schema diagram for the above data warehouse using snowflake schema class.

(ii) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations should one perform in order to list the average grade of CS courses for each Big_University student. (10)

(b) How tuning and testing of data warehouse is performed? (5)

5. (a) What are Decision trees? How they assist in classifying data? Explain with the help of suitable example. (10)

(b) How genetic algorithm approach assists in the process of classification? (5)

6. A database has four transactions. Let *min_sup*=2 and *min_conf*=85% :

TID	Items bought
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

(a) Find all the frequent itemsets using A-priori algorithm.

(b) List all the strong association rules satisfying the *min_sup* and *min_conf*. The rules should match the following metarule, where X is a variable representing customers and items denotes variable representing items :

$\forall_x \in$ transaction, buys(X, item$_1$)^ buys(X, item$_2$)

$\Rightarrow$ buys(X, item$_3$). (15)

7. Write short note on the following (any *three*) :

(a) Mining spatial databases.

(b) Data Mining Query Language.

(c) Time-Series Data mining.

(d) Data Warehouse back-end tools. (15)