

### CE3-R3: DATA WAREHOUSING AND MINING

**NOTE:**

1. Answer question 1 and any FOUR from questions 2 to 7.
2. Parts of the same question should be answered together and in the same sequence.

**Time: 3 Hours**

**Total Marks: 100**

**1.**

- a) What is meant by pattern? How is a data warehouse different from a database?
- b) Can we construct a data cube for multimedia data analysis? Justify your answer with an example.
- c) What is noisy data? Explain briefly methodologies to remove noise from data.
- d) What do you understand by spatial database and spatial data mining? Can we construct a spatial data warehouse?
- e) Why should you put your data warehouse on a different system than your OLTP system?
- f) Update-driven approach is chosen rather than query driven approach while integrating multiple heterogeneous information sources. Justify whether true or false.
- g) Differentiate data query and knowledge query.

**(7x4)**

**2.**

- a) Star schema or a snowflake schema is used to model data warehouse. What are the similarities and the differences of the two models, and then state their advantages and disadvantages.
- b) What are the steps involved in KDD (Knowledge discovery in databases)? What is the use of the knowledge base?
- c) What are the factors involved while choosing data mining system? Briefly explain the importance of each of them.

**(6+6+6)**

**3.**

- a) What is the purpose of Apriori Algorithm? How to generate association rules from frequent item sets? Explain.
- b) What are multidimensional association rules? Explain in brief.
- c) List out the OLAP operations in multidimensional data model? What is roll-up and drill-down operation?

**(8+6+4)**

**4.**

- a) What are OLAP services and its advantages of OLAP over two-dimensional reports? What is OLAP functionality?
- b) What do understand by principal component partitioning algorithm? Explain the algorithm in detail.

**(8+10)**

**5.**

- a) What does hierarchical clustering mean? In what way it is different from partition-based methods. Discuss the functionality of Chameleon's clustering method with an example.
- b) Why there is a need of data ware house? List some data warehousing tools.

**(12+6)**

**6.**

- a) What are Bayesian classifiers? Explain briefly Baye's theorem. Also explain how Naïve Bayesian classifier works?
- b) When can we say the association rules are interesting? Explain Association rule in mathematical notations.

**(10+8)**

**7.**

- a) What kind of requirements demands the use of Online Analytical Mining (OLAM)? Briefly explain OLAM architecture?
- b) What are the problems in web linkage structure while mining the World Wide Web?
- c) Discuss the problems that will occur in a continuous-media system if data is delivered either too slowly or too fast?

**(8+4+6)**