

CE3-R3: DATA WAREHOUSING AND MINING

NOTE:

1. Answer question 1 and any FOUR questions from 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 do you understand by spatial database and spatial data mining? Can we construct a spatial data warehouse?
- b) Why fact constellation schema data warehouse model is not suitable for data marts? Which process model is suited for designing data marts? Justify.
- c) There are two database recovery strategies: rollback and rollforward. Explain how the two strategies work and what recovery situations each strategy suite?
- d) What is classification? What are the classification tasks? List various classification techniques.
- e) "Clustering is called unsupervised classification." State true or false and Justify.
- f) What is the role of support and confidence in association rule mining? Briefly explain with examples.
- g) Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. What are the limitations of data mining?

(7x4)

2.

- a) Data mining systems can be categorized according to various criteria. How do we categorize data mining systems?
- b) What are the various OLAP operations performed to mine data from data warehouse? Briefly explain each of them.
- c) Spatial data cube is a multidimensional database where both dimensions and measures may contain spatial components. What are the methods for computation of spatial data cube?

(8+6+4)

3.

- a) Apriori algorithm is used to find the frequent itemsets from candidate dataset. Explain the major two steps of the algorithm.
- b) What is meant by authoritative web page? How can a search engine automatically identify authoritative web pages for the topic? How can we use hub pages to find authoritative pages?
- c) What is cluster analysis? What are the requirements and general applications of clustering in data mining?

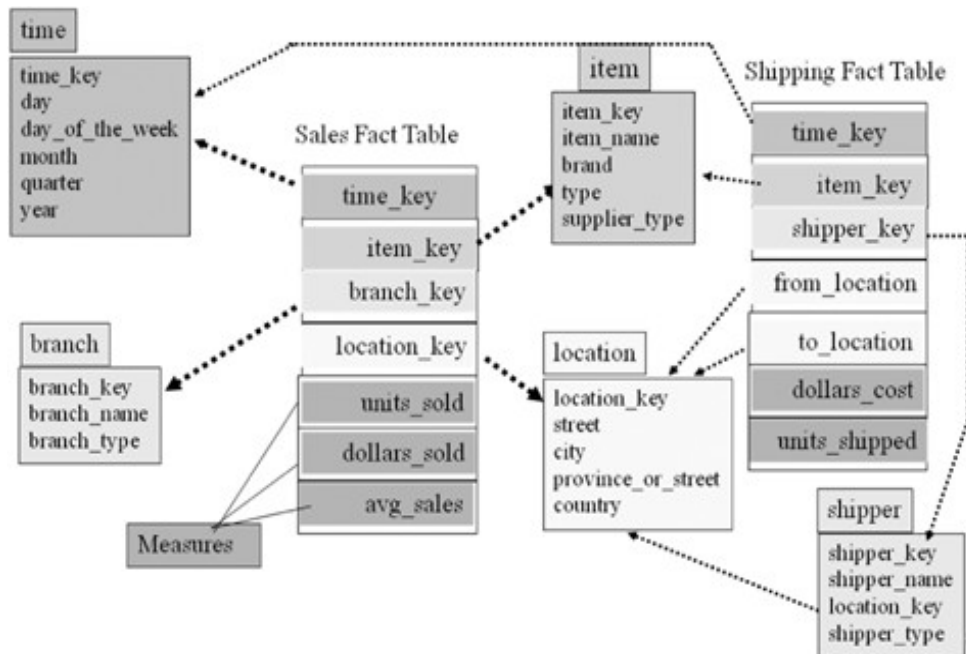
(6+6+6)

4.

- a) Classification constructs a model based on the training set and the values (class labels) in a classifying attribute then it in classifying new data. Explain the two approaches to avoid overfitting of the training data in classification? Explain each in brief.
- b) "Data mining is the core of knowledge data discovery process." Briefly explain its steps.
- c) How is the data warehouse different from operational data? Describe the comparison of data warehouse and operational data.

5.

- a) What are Bayesian classifiers? Briefly explain Baye's theorem. Also explain how Naive Bayesian classifier works?
- b) Following is the fact constellation schema. Define the schema in DMQL.



- c) Data mining is a young discipline with wide and diverse applications and there is still a nontrivial gap between general principles of data mining and domain-specific, effective data mining tools for particular applications. How to choose a data mining system? Explain.

(6+6+6)

6.

- a) Why is it required to preprocess the data before creating the data warehouse? Explain in brief.
- b) What is ARCS (Association Rule Clustering System)? How does ARCS work? What are its limitations?
- c) What is a decision tree? What are the two phases for generating the decision tree?
- d) Clustering is unsupervised classification with no predefined classes. What are the major clustering approaches?

(4+6+4+4)

7.

- a) What is spatial classification and spatial trend analysis? Briefly explain with examples.
- b) What are the problems in web linkage structure while mining the World Wide Web?
- c) Interactive, exploratory mining giga-bytes of data always demand the use of constraints. What kind of constraints can be used?
- d) What kind of requirements demands the use of Online Analytical Mining (OLAM)? Briefly explain OLAM architecture.

(4+4+4+6)