

CS/B.TECH/CSE(New)/SEM-7/CS-704C/2013-14

2013

DATA WAREHOUSING AND DATA MINING

Time Allotted : 3 Hours

Full Marks : 70

The figures in the margin indicate full marks.

*Candidates are required to give their answers in their own words
as far as practicable.*

GROUP – A

(Multiple Choice Type Question)

1. Choose the correct alternatives for the following: 10 x 1 = 10
 - i) A Data Warehouse should be
 - a) subject oriented
 - b) integrated
 - c) time variant
 - d) All of these.
 - ii) In Type of schema Fact tables are normalized.
 - a) Snow flake schema
 - b) Star schema
 - c) Fact constellation schema
 - d) None of these.
 - iii) The 'Dice' operation is concerned with
 - a) multiple runs of slice
 - b) slice on more than one dimension
 - c) selecting certain cells of more than one dimension
 - d) two consecutive slice operations in two more dimensions.

- iv) Which of the following indexing techniques are used for data warehousing?
- a) Hashing on primary keys
 - b) Indexing on foreign keys of fact table
 - c) Bitmap indexing and Join indexing
 - d) Both (a) and (b).
- v) Which frequent pattern mining technique mines without candidate generation?
- a) Partitioning
 - b) Apriori
 - c) FP-growth
 - d) Dynamic intensive counting.
- vi) In type of architecture information delivery can be from both centralized data warehouse and dependent data marts.
- a) Federated
 - b) Hub and spoke
 - c) Independent data marts
 - d) data mart bus.
- vii) is an example of hierarchical clustering algorithm.
- a) CLARANS
 - b) C4.5
 - c) BIRCH
 - d) ROCK
- viii) The clustering technique k-means is based on
- a) Root mean square
 - b) Hamming distance
 - c) Centroid
 - d) Euclidean distance
- ix) A data cube structure is used for
- a) aggregation
 - b) navigation

- c) pre-computation of summarized data
 - d) all of these.
- x) The mining activity which mines web log records to discover user access patterns of web pages is
- a) web content mining
 - b) web usage mining
 - c) web structure mining
 - d) web search mining.

GROUP – B

(Short Answer Type Questions)

Answer any *three* of the following. 3 x 5 = 15

2. What is the difference between OLAP and OLTP operation?
What types of OLAP operations are possible in a data warehouse?
3. What types of requirement dictate the granularity or levels of detail in a data warehouse? List two aspects of information delivery that are directly impacted by business requirements. 3+2
4. What is the purpose of data mart? What are the different architecture types of data mart? Mention the purpose of each. 3+2
5. What are the major steps and requirements of Extraction, Transformation and Loading?
6. What are the different types of meta data? Why is meta data vital for end users? 3+2

GROUP – C

(Long Answer Type Questions)

Answer any *three* of the following. 3 x 15 = 45

7. a) Explain the three tier data warehouse architecture.

- b) Compare ROLAP with MOLAP servers. Which one do you consider more appropriate for data warehouse applications and why?
 - c) What are dimension hierarchies? Give two examples.
 - d) Explain the difference between top-down and bottom-up approaches for building a data warehouse? Which one is preferable and why? 3+5+3+4
- 8.
- a) How can business dimensions be useful for defining requirements in a business data warehouse? In which situation JAD methodology may be useful for collecting requirements?
 - b) Explain the schemas which are used in a data warehouse design with suitable examples. Which schema design is more popular? Draw a schema design for a university admission system. Consider appropriate dimensions and fact table as required. 3+3+4+5
- 9.
- a) How can you use the web as a data source for data warehouse? Why is data security a major concern for a web enabled data warehouse?
 - b) Consider the five transactions given below. If minimum support is 0.2 and minimum confidence is 50% determine the frequent itemsets and association rules using apriori algorithm.

Transaction id	Items
T10	Computer, Anti-Virus
T11	Computer Anti-Virus, Pen drive
T12	Data card, anti-virus
T13	Computer, datacard, web camera
T14	Pen drive, web camera

- c) What are the demerits of Apriori algorithm. (4+4)+5+2
- 10.
- a) Mention three reasons why data quality is critical in a data warehouse.
 - b) Name some data mining application areas. Write down the difference between data mining and OLAP.

- c) Write an algorithm for Decision tree construction and mention the criterion for splitting an attribute. 3+6+6
11. Write short notes on any *three* of the following: 3 x 5
- a) Knowledge discovery in databases (KDD)
 - b) Data mining techniques
 - c) Data warehouse software and hardware tools
 - d) Web enabled data warehouse.

=====