

LIB
8/6/13 AN

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code : 21284

B.E./B.Tech. DEGREE EXAMINATION, MAY/JUNE 2013.

Seventh Semester

Computer Science and Engineering

CS 2032 / CS 701 / 10144 CSE 32 – DATA WAREHOUSING AND DATA MINING

(Common to Sixth Semester Information Technology)

(Regulation 2008 / 2010)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is a data mart?
2. State why one of the biggest challenges when designing a data warehouse is the data placement and distribution strategy.
3. What are production reporting tools? Give examples.
4. Define a data cube.
5. State the need for data cleaning.
6. What is pattern evaluation?
7. What is market basket analysis?
8. State the need for pruning phase in decision tree construction.
9. Classify hierarchical clustering methods.
10. What is an outlier?

PART B — (5 × 16 = 80 marks)

11. (a) (i) Explain the role played by sourcing, acquisition, cleanup and transformation tools in building a data warehouse. (8)
- (ii) What is meta data? Classify meta data and explain the same. (8)

Or

- (b) (i) What is a multi dimensional data model? Explain star schema with an example and diagrammatic illustration. (8)
- (ii) Explain the potential performance problems with star schema. Give examples. (8)
12. (a) (i) Perform a comparative study between MOLAP and ROLAP. (8)
- (ii) Explain with diagrammatic illustration managed query environment (MQE) architecture. (8)

Or

- (b) Explain the features of the reporting and query tool COGNOS IMPROMPTU. (16)
13. (a) (i) Explain with diagrammatic illustration data mining as a step in the process of knowledge discovery. (12)
- (ii) What is evolution analysis? Give example. (4)

Or

- (b) (i) Explain with diagrammatic illustration data mining as a confluence of multiple disciplines. (8)
- (ii) Explain with diagrammatic illustration the primitives for specifying a data mining task. (8)

14. (a) Apply the Apriori algorithm for discovering frequent item sets to the following data set:

Trans ID	Items Purchased
101	Kiwi, Grapes, Star fruit
102	Kiwi, Gooseberry
103	Gooseberry, Pear
104	Kiwi, Grapes, Star fruit
105	Lemon, Star fruit
106	Lemon
107	Lemon, Gooseberry
108	Kiwi, Grapes, Mango, Star fruit
109	Mango, Pear
110	Kiwi, Grapes, Star fruit

Use 0.3 for the minimum support value. Illustrate each step of the Apriori algorithm. (16)

Or

- (b) What is classification? Explain with an example Bayesian classification. (16)
15. (a) Consider five points $\{X_1, X_2, X_3, X_4, X_5\}$ with the following coordinates as a two dimensional sample for clustering :

$$X_1 = (0, 2.25); X_2 = (0, 0.25); X_3 = (1.25, 0) X_4 = (4.5, 0); X_5 = (4.5, 2.5)$$

Illustrate the K-means partitioning algorithm (clustering algorithm) using the above data set. (16)

Or

- (b) Explain with an example density-based local outlier detection. (16)