Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 60356

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2016.

Seventh Semester

Computer Science and Engineering

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

(Common to Sixth Semester Information Technology)

(Regulations 2008/2010)

(Common to PTCS 2032/10144 CSE 32 – Data Warehousing and Data Mining for B.E. (Part-Time) Sixth Semester – Computer Science and Engineering – Regulations 2009/2010)

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. State the needs of a Multidimensional data model.

4. What is a data cube?

5. Differentiate between data characterization and discrimination.

6. Give the need for data pro-processing.

7. List the two interesting measures of an association rule.

8. What is decision tree induction?

9. Let $x_1 = (1, 2)$ and $x_2 = (3,5)$ represent two points. Calculate the Manhattan distance between the two points.

10. How outliers may be detected by clustering?

PART B — (5 × 16 = 80 marks)

11. (a) Draw any two multi-dimensional schemas suitable for representing weather data and give their advantages and disadvantages. (16)

Or

(b) Explain the multi-tier architecture suitable for evolving a data warehouse with suitable diagram. (16)

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) With diagrammatic illustration discuss data mining as a confluence of multiple disciplines. (8)

(ii) List and discuss the data mining task primitives. (8)

Or

(b) Discuss the following schemes used for integration of a data mining system with a database or data warehouse system :

(i) No coupling (4)

(ii) Loose coupling (4)

(iii) Semi tight coupling (4)

(iv) Tight coupling. (4)

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

Trans ID	Items purchased
101	Mulberry, Raseberry, Cherry
102	Mulberry, Papaya
103	Papaya, Mango
104	Mulberry, Rasberry, Cherry
105	Passion Fruit, Cherry
106	Passion Fruit
107	Passion Fruit, Papaya
108	Mulberry, Rasberry, Guava, Cherry
109	Guava, Mango
110	Mulberry, Rasberry

Use 0.3 for the minimum support value. (16)

Or

(b) State Baye's theorem of posterior probability and explain the working of a Bayesian classifier with an example. (16)

15. (a) (i) How agglomerative hierarchical clustering works? Explain with an example. (8)

(ii) How divisible hierarchical clustering works? Explain with an example. (8)

Or

(b) Consider five points $\{x_1, x_2, x_3, x_4, x_5\}$ with the following coordinates as a two dimensional sample for clustering :

$x_1 = (0,2)$, $x_2 = (1,0)$, $x_3 = (2,1)$, $x_1 = (4,1)$ and $(x_5) = (5,3)$

Illustrate the K-means algorithm on the above data set. The required number of clusters is two and initially, clusters are formed from random distribution of samples : $C_1\{x_1, x_2; x_4\}$ and $C_1\{x_3, x_5\}$. (16)
