

Reg. No. :

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

**Question Paper Code : 21356**

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2015.

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. List the characteristics of a data warehouse.
2. State why data partitioning is a key requirement for effective parallel execution of database operations.
3. State the needs of a Multidimensional data model.
4. What is a data cube?
5. Why data preprocessing is an important issue for both data warehousing and data mining?
6. Define a pattern.
7. What are eager learners? Give examples.
8. Write about correlation analysis.
9. State any two applications of data mining.
10. Define an outlier.

PART B — (5 × 16 = 80 marks)

11. (a) (i) Explain with diagrammatic illustration the relationship between operational data, a data warehouse and data marts. (6)
- (ii) "A data warehouse can be modeled by either a star schema or a snowflake schema". With relevant examples discuss the two types of schema. (10)

Or

- (b) What is a data warehouse? Explain the steps in building a data warehouse. (16)
12. (a) Highlight the features of Cognos Impromptu business analysis tool. (16)

Or

- (b) Explain with an example and diagrammatic illustrations the following OLAP operations in a multidimensional data model:
- (i) Roll-up (5)
- (ii) Drill-down (5)
- (iii) Slice and Dice (6)
13. (a) (i) List and explain the primitives for specifying a data mining task. (12)
- (ii) What is a concept hierarchy? Explain the same with an example. (4)

Or

- (b) (i) Explain with diagrammatic illustration the steps involved in the process of knowledge discovery from databases. (8)
- (ii) Explain with diagrammatic illustration how data mining acts as a confluence of multiple disciplines. (8)
14. (a) Apply the Apriori algorithm for discovering frequent item sets to the following data set:

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

Use 0.3 for the minimum support value. (16)

Or

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

15. (a) Illustrate the K-means partitioning algorithm (clustering algorithm) using an example. (16)

Or

- (b) What is hierarchical clustering? With an example discuss dendrogram representation for hierarchical clustering of data objects. (16)
-