Reg. No.

Question Paper Code : **51356**

**B.E./B. Tech. DEGREE EXAMINATION, MAY/JUNE 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.   List the characteristics of a data warehouse.

2.   State why data partitioning is a key requirement for effective parallel execution of database operations.

3.   What is a reporting tool ?

4.   Give examples for managed query tools.

5.   List out the data mining functionalities.

6.   Why data cleaning routines are needed ?

7.   Define frequent itemset.

8. Give examples for binary and multidimensional association rules.

9. Classify hierarchical clustering methods.

10. What is 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.

12. (a) (i) Diagrammatically illustrate and discuss the architecture of MOLAP and ROLAP. **(12)**

    (ii) Compare MOLAP and ROLAP. **(4)**

    **OR**

    (b) List and discuss the features of Cognos Impromptu. **(16)**

13. (a) Write short notes on :

    (i) Classification of data mining systems **(8)**

    (ii) Data mining task primitives. **(8)**

    **OR**

    (b) Discuss the various pre-processing tasks. **(16)**

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

| Trans ID | Items Purchased |
|----------|-----------------|
| 101 | Litchi, Hill Banana, Straw Berry |
| 102 | Litchi, Passion Fruit |
| 103 | Passion Fruit, Tomato |
| 104 | Litchi, Hill banana, Straw Berry |
| 105 | Pears, Straw Berry |
| 106 | Pears |
| 107 | Pears, Passion Fruit |
| 108 | Litchi, Hill Banana, Water Melon, Straw Berry |
| 109 | Water Melon, Tomato |
| 110 | Litchi, Hill Banana |

Use 0.3 for the minimum support value. (16)

**OR**

(b) Explain the working of the naive Bayesian classifier with an example. (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 the density-based local outlier detection. (16)