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

Question Paper Code : 51324

B.E./B.Tech. DEGREE EXAMINATION, MAY/JUNE 2014.

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)

(Common to PTCS 2032 – Data warehousing and Data Mining for B.E. (Part-Time) Sixth Semester – Computer Science and Engineering – Regulation 2009)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is the need for back end process in data warehouse design?
2. What are the advantages dimensional modelings?
3. List out the two different types of reporting tools.
4. Define OLAP.
5. What is Legacy database?
6. What is descriptive and predictive data mining?
7. How is prediction differing from classification?
8. How do you choose best split while constructing a decision tree?
9. What is a STING?
10. Define Wave Cluster?

PART B — (5 × 16 = 80 marks)

11. (a) (i) Explain the mapping of data warehouse to multiprocessor architecture. (10)

(ii) Discuss about data warehouse meta data. (6)

Or

(b) With a neat diagram describe the various stages of building a data warehouse. (16)

12. (a) (i) Explain the data model which is suitable for data warehouse with an example. (8)

(ii) Write the difference between multi-dimensional OLAP and multi-relational OLAP. (8)

Or

(b) Explain the different types of OLAP tools. (16)

13. (a) What is the use of data mining task? What are the basic types of data mining tasks? Explain with examples. (16)

Or

(b) Explain various methods of data cleaning in detail. (16)

14. (a) (i) Write and explain the algorithm for mining frequent itemsets without candidate generation. (8)

(ii) A database has nine transactions let min_sup = 30%. (8)

TID	List of items_IDs
1	a, b, e
2	b, d
3	b, c
4	a, b, d
5	a, c
6	b, c
7	a, c
8	a, b, c, e
9	a, b, c

Find all frequent itemsets using the above algorithm.

Or

(b) With an example explain various attribution selection measures in classification. (16)

15. (a) (i) Explain the different types of data used in cluster analysis. (10)

(ii) Discuss the use of outlier analysis. (6)

Or

(b) (i) Write the difference between CLARA and CLARANS. (6)

(ii) Explain how data mining is used for retail industry. (10)

———————