

Reg. No. :

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

Question Paper Code: 41255

B.E. / B.Tech. DEGREE EXAMINATION, NOV 2016

Fifth Semester

Computer Science and Engineering

14UCS505 - DATA WAREHOUSING AND DATA MINING

(Regulation 2014)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 1 = 10 Marks)

- The process of removing the deficiencies and loopholes in the data is called as
 - Aggregation of data
 - Extracting of data
 - Cleaning up of data
 - Loading of data
- Data warehouse bus matrix is a combination of
 - Dimensions and data marts
 - Dimensions and facts
 - Facts and data marts
 - Dimensions and detailed facts
- Which of the following should not be considered for each dimension attribute?
 - Attribute name
 - Rapid changing dimension policy
 - Attribute definition
 - Sample data
- Which of the following tools a business intelligence system will have
 - OLAP tool
 - Data mining tool
 - Reporting tool
 - Both (a) and (b)
- Which of the following is/are the data mining tasks?
 - Regression
 - Classification
 - Clustering
 - All the above

6. What are the criteria used for comparing classification and prediction

(a) Predictive accuracy	(b) Robustness
(c) Scalability and interoperability	(d) All the above
7. Which of the following is/are outlier detection approaches?

(a) Statistical approach	(b) distance based approach
(c) Both (a) and (b)	(d) None of these
8. The another name of agglomerative approach is

(a) Bottom-up approach	(b) Top- down approach
(c) Both (a) and (b)	(d) None of these
9. What are the requirements of cluster analysis?

(a) Scalability	(b) High dimensionality
(c) Both (a) and (b)	(d) None of these
10. Which of the following is/are applications of data mining?

(a) Financial Data Analysis	(b) Retail Industry
(c) Telecommunication Industry	(d) All the above

PART - B (5 x 2 = 10 Marks)

11. Define slice and dice operation.
12. What is dimensionality reduction?
13. List the techniques to improve the efficiency of Apriori algorithm.
14. What is tree pruning?
15. Differentiate between row scalability and column scalability issues.

PART - C (5 x 16 = 80 Marks)

16. (a) Explain in detail about the architecture and implementation of the data warehouse with neat sketch. (16)

Or

- (b) Explain in detail about metadata. Classify metadata and explain the same. (16)

17. (a) (i) List and discuss the basic features that are provided by reporting and query tools used for business analysis. (10)
- (ii) Describe multidimensional data model in detail. (6)

Or

- (b) Discuss Multidimensional Online Analytical Processing (MOLAP) and Multi Relational Online Analytical Processing (ROLAP) with relevant example. (16)
18. (a) (i) Explain in detail about the data mining task primitives. (10)
- (ii) Discuss the various issues that have to be addressed during data integration. (6)

Or

- (b) How data mining systems are classified? Discuss each classification with an example. (16)
19. (a) Explain in detail about constraint based association mining. (16)

Or

- (b) Write short notes on the following:
- (i) Comparison of classification and prediction methods. (8)
- (ii) Issues in classification and prediction methods. (8)
20. (a) (i) Explain with an example density based clustering methods. (8)
- (ii) Explain in detail about K-means partitioning algorithm. (8)

Or

- (b) Explain outlier analysis in detail with an example. Discuss the use of outlier analysis. (16)
-

