Name of Subject	Business Intelligence and its Application	Subject Code	DEIT-14501
Batch	2014 and onwards	Class	D3IT A and B

Section A: 2 Marks

1. Define the term 'Data Warehouse'.
2. Write down the applications of data warehousing.
3. When is data mart appropriate?
4. List out the functionality of metadata.
5. What are nine decision in the design of a Data warehousing?
6. List out the two different types of reporting tools.
7. Why data mining is used in all organizations.
8. What are the technical issues to be considered when designing and implementing a data warehouse environment?
9. List out some of the examples of access tools.
10. What are the advantages of data warehousing.
11. Give the difference between the Horizontal and Vertical Parallelism.
12. Draw a neat diagram for the Distributed memory shared disk architecture.
13. Define star schema.
14. What are the reasons to achieve very good performance by SYBASE IQ technology?
15. What are the steps to be followed to store the external source into the data warehouse?
16. Define Legacy data.
17. Draw the standard framework for metadata interchange.
18. List out the five main groups of access tools.
19. Define Data Visualization.
20. What are the various forms of data preprocessing?
21. How is data warehouse different from database? How are they similar?
22. What is data transformation? Give example.
23. With an example explain what is Meta data?
24. What is data mart?
25. Define data.
26. State why the data preprocessing an important issue for data warehousing and data mining.
27. What is the need for discretization in data mining?.
28. What are the various forms of data preprocessing?

29. What is concept Hierarchy? Give an example.
30. What are the various forms of data preprocessing?
31. Mention the various tasks to be accomplished as part of data pre-processing.
32. Define Data Mining.
33. List out any four data mining tools.
34. What do data mining functionalities include?
35. Define patterns.
36. What are the requirements of clustering?
37. What are the applications of spatial data bases?
38. What is text mining?
39. Distinguish between classification and clustering.
40. Define a Spatial database.
41. List out any two various commercial data mining tools.
42. What is the objective function of K-means algorithm?
43. Mention the advantages of Hierarchical clustering.
44. Distinguish between classification and clustering.
45. List the requirements of clustering in data mining.
46. What is web usage mining?
47. What are the requirements of clustering?
48. What are the applications of spatial databases?
49. What is text mining?
50. What is cluster analysis ?
51. What are the two data structures in cluster analysis?
52. What is an outlier? Give example.
53. What is audio data mining?
54. List two application of data mining.

PART-B   5 Marks

1. Enumerate the building blocks of data warehouse. Explain the importance of metadata in a data warehouse environment.
2. Explain various methods of data cleaning in detail
3. Diagrammatically illustrate and discuss the data warehousing architecture with briefly explain components of data warehouse

4. (i) Distinguish between Data warehousing and data mining. (ii)Describe in detail about data extraction, cleanup
5. Write short notes on
6. (i)Transformation (ii)Metadata
7. List and discuss the steps involved in mapping the data warehouse to a multiprocessor architecture.
8. Discuss in detail about Bitmapped Indexing
9. Explain in detail about different Vendor Solutions.
10. Difference between OLAP and OLTP.
11. Classify OLAP tools.
12. What is meant by OLAP?
13. Difference between OLAP & OLTP
14. Define Concept Hierarchy.
15. List out the five categories of decision support tools.
16. Define Cognos Impromptu
17. List out any 5 OLAP guidelines.
18. Distinguish between multidimensional and multi-relational OLAP.
19. Define ROLAP, MOLAP,OLAP,HOLAP.
20. Draw a neat diagram for the web processing model.
21. Draw a neat sketch for three-tired client/server architecture.
22. List out the applications that the organizations uses to build a query and reporting environment for the data warehouse.
23. Distinguish between window painter and data windows painter.
24. Define ADF, SGF and DEF.
25. What is the function of power play administrator?
26. Discuss the typical OLAP operations with an example.
27. List and discuss the basic features that are provided by reporting
28. and query tools used for business analysis.
29. Describe in detail about Business Intelligence
30. Explain about OLAP in detail.
31. With relevant examples discuss multidimensional online analytical processing and multi-relational online analytical processing.
32. Discuss about the OLAP tools and the Internet
33. (i)Explain Multidimensional Data model.
34. (ii)Discuss how computations can be performed efficiently on data cubes.

35. What is meant by market Basket analysis?

36. What is the use of multilevel association rules?

37. What is meant by pruning in a decision tree induction?

38. Write the two measures of Association Rule.

39. With an example explain correlation analysis.

40. Define conditional pattern base.

41. List out the major strength of decision tree method.

42. In classification trees, what are the surrogate splits, and how are they used?

43. The Naïve Bayes' classifier makes what assumptions that motivate its name?

44. What is the frequent item set property?

45. List out the major strength of the decision tree Induction.

46. Write the two measures of association rule.

47. How are association rules mined from large databases?

48. What is tree pruning in decision tree induction?

49. What is the use of multi level association rules?

50. What are the Apriori properties used in the Apriori algorithms?

51. How is predication different from classification?

52. What is a support vector machine?

53. What are the means to improve the performance of association rule mining algorithm?

54. State the advantages of the decision tree approach over other approaches for performing classification.

PART-C   10 marks

1. Decision tree induction is a popular classification method. Taking one typical decision tree induction algorithm , briefly outline the method of decision tree classification.

2. Consider the following training dataset and the original decision tree induction algorithm (ID3). Risk is the class label attribute. The Height values have been already discredited into disjoint ranges. Calculate the information gain if Gender is chosen as the test attribute. Calculate the information gain if Height is chosen as the test attribute. Draw the final decision tree (without any pruning) for the training dataset. Generate all the "IF-THEN rules from the decision tree.

1. Gender Height Risk

2. F (1.5, 1.6) Low

3. M (1.9, 2.0) High

4. F (1.8, 1.9) Medium F (1.8, 1.9) Medium F (1.6, 1.7) Low

5. M (1.8, 1.9) Medium

6. F (1.5, 1.6) Low M (1.6, 1.7) Low M (2.0, 8) High M (2.0, 8) High

7. F (1.7, 1.8) Medium M (1.9, 2.0) Medium F (1.8, 1.9) Medium F (1.7, 1.8) Medium

8. F (1.7, 1.8) Medium
    (a) Given the following transactional database

1 C, B, H
2 B, F, S
3 A, F, G
4 C, B, H
5 B, F, G
6 B, E, O

    (i)   We want to mine all the frequent itemsets in the data using the Apriori algorithm.

9. Assume the minimum support level is 30%. (You need to give the set of frequent item sets in L1, L2,… candidate item sets in C1, C2,…)
    (ii)   Find all the association rules that involve only B, C.H (in either left or right hand side of the rule). The minimum confidence is 70%.

10. Describe the multi-dimensional association rule, giving a suitable example.

    (a)Explain the algorithm for constructing a decision tree from training samples
    (b)Explain Bayes theorem.

11. Develop an algorithm for classification using Bayesian classification. Illustrate the algorithm with a relevant example.

10. Discuss the approaches for mining multi level association rules from the transactional databases. Give relevant example.
11. Write and explain the algorithm for mining frequent item sets without candidate generation. Give relevant example.
12. How is attribute oriented induction implemented? Explain in detail.
13. Discuss in detail about Bayesian classification
14. A database has four transactions. Let min sup=60% and min conf=80%.

TID	DATE	ITEMS_BOUGHT
T100	10/15/07	{K,A,B}
T200	10/15/07	{D,A,C,E,B}

T300	10/19/07	{C,A,B,E}
T400	10/22/07	{B,A,D}

15. Find all frequent item sets using Apriori and FP growth, respectively. Compare the efficiency of the two mining process.
BIRCH and CLARANS are two interesting clustering algorithms that perform effective clustering in large data sets.
   (a) Outline how BIRCH performs clustering in large data sets. (ii) Compare and outline the major differences of the two scalable clustering algorithms BIRCH and CLARANS.

16. Write a short note on web mining taxonomy. Explain the different activities of text mining.

17. Discuss and elaborate the current trends in data mining.

18. Discuss spatial data bases and Text databases

19. What is a multimedia database? Explain the methods of mining multimedia database?

   (a) Explain the following clustering methods in detail.
      (b) BIRCH (b) CURE
   Discuss in detail about any four data mining applications.

20. Write short notes on

      (i)   Partitioning methods (ii) Outlier analysis

21. Describe K means clustering with an example.

22. Describe in detail about Hierarchical methods.

23. With relevant example discuss constraint based cluster analysis.

vidyarthiplus/rss	en_US