

PS 606.2 E2 BI and ADM

Question Bank

Unit - I

1. What is a data warehouse?
2. Examine the steps involved for the design and construction of data warehouses.
3. With a neat diagram explain the components of data warehouse.
4. What is the difference between OLTP and Data Warehouse?
5. What is data mart? What are the types of data mart? Which schema is suitable for data mart?
6. Explain about the primary components of Business Intelligence
7. Describe the different approaches towards data warehouse development
8. What sort of post processing needs to be done on loaded data? How can you verify data integrity?
9. Mention the characteristics of data warehouse
10. Bring out the roadmap to Business Intelligence

Unit - II

11. What are the steps involved in creating physical model of a data warehouse? Explain with suitable examples
12. List the characteristics of a data warehouse.
13. Differentiate fact table and dimension table. Give example
14. List the major steps involved in the ETL process
15. Describe the type of metadata that is maintained in a data warehouse.
16. Suppose that a data warehouse for big-university consist of the following four dimensions: Student, Course, Semester and Instructor, and two measures count and avg_grade. When at the lowest

conceptual level (Ex. For a given student, course, semester and instructor combination), the avg_grade measure stores the actual course grade of the student. At higher conceptual levels, avg_grade stores the average grade for the given combination.

Draw a snowflake schema diagram for the data warehouse. Starting with the base cuboid [Student, Course, Semester, Instructor], what specific OLAP operations (eg Roll-up from Semester to Year) should one perform in order to list the average grade of CS courses for each big-university students. (5)

If each dimension has five levels (including all) such as Student<major<status<university<all, how many cuboids will this cube contains (including the base and Apex cuboids).

17. With a neat sketch explain the architecture of data warehousing

Unit - III

18. Suppose that the data for analysis includes the attribute age. The age values for the 10 data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70.

- What is the mean of the data? What is the median?
- What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.)
- What is the midrange of the data?
- Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
- Give the five-number summary of the data.
- Show a boxplot of the data.
- How is a quantile-quantile plot different from a quantile plot?

19. The data warehouse for a UNIVERSITY consists of 4 dimensions – STUDENT, COURSE, SEMESTER and TEACHER and 2 measures – COUNT

and AVG-GRADE. Draw a star schema diagram for the UNIVERSITY warehouse. Assume attribute for dimensions

20. Describe different data repositories on which mining can be performed.
21. What is Metadata? Explain categories of Metadata
22. Describe the different classifications of associate rule mining?
23. List the contents of dimension table
24. Detail on data warehouse meta data
25. Discuss the representation, schema and measures of a multi dimensional model in a data warehouse.
26. Demonstrate how Bayesian classification helps in predicting class membership probabilities
27. How Support Vector Machines is applied for the classification of both linear and nonlinear data?

Unit - IV

28. What are the requirements of cluster analysis?
29. Explain decision tree induction algorithm. What are the various characteristics of decision tree induction?
30. Explain with a neat diagram the steps involved in the knowledge discovery in database process
31. Mention Data mining functionality, classification, prediction, clustering & evolution analysis?
32. List any four reasons to perform data preprocessing
33. Differentiate classification and prediction.
34. What is back propagation?
35. List any two applications of data mining.
36. How classification of data mining systems is done? Explain them with example.
37. Discuss the activities of data cleaning with the process associated with it.

38. Demonstrate how Bayesian classification helps in predicting class membership probabilities
39. How Support Vector Machines is applied for the classification of both linear and nonlinear data?
40. Demonstrate the application of data mining for financial analysis
41. What factors lead to the mining of data?
42. What are the various forms of visualizing the discovered patterns in data mining? What is data discretization? Give an example.
43. What is data mining functionality? Explain different types of data mining functionality with examples.
44. Discuss the issues in data mining in detail.
45. Explain how data mining system can be integrated with database/
46. data warehouse system.
47. Describe the different methods for data cleaning.
48. Explain the algorithm for constructing a decision tree from training samples.
49. Discuss in detail about the steps in knowledge discovery in data bases. Explain different techniques in data mining
50. Explain the various data reduction techniques in the preprocessing step of data mining
51. What are the steps involved in data mining process?
52. What is the use of the knowledge base?
53. How do you choose best split while constructing a decision tree?
54. Discuss the major issues in data mining system
55. Discuss the issues to be considered during data integration
56. Describe the different methods for data cleaning
57. Explain the algorithm for constructing a decision tree from training sample
58. Why preprocess the data?
59. Describe the steps involved in knowledge discovery in databases (KDD)
60. Explain Bayesian Classification method
61. Explain how data mining is used for intrusion detection

62. What is meant by outlier? How these outliers are detected using data mining
63. Mention the methods to handle the missing values
64. Discuss in detail about the Bayesian and decision tree classifier
65. Suppose that the data for analysis includes the attribute age. the age values for the data tuples are increasing order 13 16 16 23 23 25 25 25 30 30 30 30 35 35 35 40 40 45 45 45 70
a) How might you determine the outliers in the data?
b) What other methods are there for data smoothing?
66. List and describe the primitives for the data mining task?
67. How will you solve a classification problem using decision trees?
68. What is data mining functionality? Explain different types of data mining functionality with examples.

Unit - V

69. What is an association rule? What metrics are used to measure the strength of an association rule?
70. Discuss the Apriori algorithm in the context of frequent itemset generation, for association rules.
71. What is the need for correlation analysis on association rule? Explain with an example
72. Elaborate the concept of mining multilevel association rules from relational databases and data warehouses with case study.
73. Compare clustering and classification.
74. Discuss the following clustering algorithm using examples :
(i) K-means. (ii) K-medoid.
75. Explain the different partitioning and hierarchical clustering methods in detail.
76. Compare clustering and classification
77. Explain different types data in clustering analysis
78. Define support and confidence in association rule mining
79. Define strong association rule

80. What is clustering? Briefly describe the partitioning and hierarchical clustering methods. Give examples
81. How is association rules mined from large databases?
82. Describe the different classifications of associate rule mining?
83. What are the fields in which clustering techniques are used?
84. What are the major requirements of clustering analysis?
85. Explain mining multidimensional data from transactional databases and relational databases.
86. How can data visualization help in decision-making? Explain the different types of data repositories on which mining can be performed?
87. Define Data Visualization. List two application of data mining.
88. What is Graph Mining? Why is Graph and Graph Mining Important? Explain
89. Consider the transactional data set given below. Let minimum support be 60% and minimum confidence be 75%. Find all frequent item sets using the Apriori method. List all association rules found.

Transaction ID	Items
100	Bread, Eggs, Juice, Cheese
200	Bread, Cheese, Juice
300	Bread, Milk, Yogurt
400	Bread, Juice, Milk
500	Cheese, Juice, Milk

Give any two types of association rules with example. Trace the results of using the Apriori algorithm on the grocery store example with support threshold 2 and confidence threshold 60 %. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated.

Transaction_ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup

T6	HotDogs, Coke, Chips
----	----------------------