

Purple = P1
Orange = P2
Green = P3
Blue = P4

Data warehouse & Mining D.W.M

MODULE-1

1. What is Metadata? Why do we need metadata when search engines like Google seem so effective? OR Meta data with example. OR Role of metadata.
2. Suppose that a data warehouse for DB-University consist of four dimensions students, course, semester, and instructor, and two measures count and avg-grade. At the lowest conceptual level (e.g. 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. i. Draw a snowflake schema diagram for the data warehouse. ii. Starting with the basic cuboid[student,course,semester,instructor], what specifies OLAP operations (e.g. roll-up from semester to year) should you perform in order to list the average Grade of CS courses for each DB-University student.
3. A simple example from the stock market involving only discrete ranges has profit as categorical attribute, with values (up, down) and the training data is: Apply decision tree algorithm and show the generated rules. OR Define Classification. Discuss the issues in Classification. A simple example from the stock market involving only discrete ranges has profit as categorical attribute,with values {Up, Down} and the training data is : Apply decision tree algorithm and show the generated rules.
4. Information requirements are recorded for "hotel occupancy" considering dimensions like Hotel, Room and Time. Few Facts are vacant rooms,occupied rooms, number of occupants,etc.
5. Discuss Data Warehouse design strategies in detail?
6. Explain Data Warehouse Architecture in detail.
7. A manufacturing company has a huge sales network To control the sales, it is divided into regions. Each region has multiple zones. Each zone has different cities . Each sales person is allocated different cities. The objective is to track sales figure at different granularity levels of region , sales person and time.Convert the star schema and snowflake schema.

8. Differentiate top-down and bottom-up approaches for building data warehouse. Discuss the merits and limitations of each approach
9. Explain Updates to dimension tables in detail.

MODULE-2

1. The college wants to record the Marks for the courses completed by students using the dimensions: I)Course, II)Student, III)Time and a measure Aggregate marks. Create a Cube and perform following OLAP operations: i) Rollup ii)Drill down iii)Slice iv) Dice
2. Discuss various OLAP models and their architecture.
3. Explain ETL of data warehousing in detail.
4. Discuss the process of extraction, transformation and loading with a neat and labelled diagram.

MODULE-3

1. Write a short note on 1) Data preprocessing OR Discuss different steps involved in Data Pre-processing. 2) DBSCAN Data visualization
2. Applications of data mining (Minimum two in detail)
3. Answer the following questions for this problem: i) Design the star schema. ii) Can you convert this star schema to a snowflake schema? If yes, justify and draw the snowflake schema.
4. Explain Data mining as a step in KDD. Illustrate the architecture of a typical data mining system. OR Explain Data mining as a step in KDD. Give the architecture of typical data mining System.
5. Discuss the types of attributes and data visualization for data exploration.
6. Discuss: i) Architecture of a typical data mining system. ii) Application and major issues in Data Mining
7. In the real world, tuples with missing values for some attributes are a common occurrence. Describe Various methods for handling this problem.
8. Briefly outline with examples, how to compute the dissimilarity between objects described (10) by the following: i)Nominal attributes ii)Asymmetric binary attributes.

9. Data quality can be assessed in terms of several issues, including accuracy, completeness, and consistency. For each of the above three issues, discuss how data quality assessment can depend on the intended use of the data, giving examples. Propose two other dimensions of data quality.
10. Present an example where data mining is crucial to the success of a business. What data mining functionalities does this business need (e.g., think of the kinds of patterns that could be mined)? Can such patterns be generated alternatively by data query processing or simple Statistical analysis.

MODULE-4

1. Frequent pattern mining algorithms considers only distinct items in a transaction. However, multiple occurrences of an item in the same shopping basket, such as four cakes and three jugs of milk, can be important in transactional data analysis. How can one mine frequent itemsets efficiently considering multiple occurrences of items? Generate Frequent Pattern Tree.
2. Numerical on Find clusters using k-means clustering algorithm.
3. Write a short note on 1) FP Tree 2) Hierarchical clustering
4. Numerical on Naive Bayes classifier algorithm. OR Why naive Bayesian classification is called "naive" ? Briefly outline the major ideas of Naive Bayesian classification.
5. Differentiate between simple linkage, average linkage and complete linkage algorithms. Use complete linkage algorithm to find the clusters from the given dataset.

MODULE-5

1. Numerical on Find all frequent patterns using Apriori Algorithm. ii) bList strong association rules
2. Consider the data given below. Create adjacency matrix. Apply single link algorithm to cluster the given data set and draw the dendrogram.
3. Multidimensional and multilevel association mining.

MODULE-6

1. What is spatial data structure? Outline their importance in GIS.
2. With respect to web mining, is it possible to detect visual objects using meta-objects?