Sample Questions:

Module 1:

1. What is DWH? Explain DWH characteristics.
2. What are the advantages and applications of DWH?
3. Why is the ER model not suitable for DWH?What are the steps in dimensional modeling?
4. Define dimension, fact , fact table and dimension table with example.
5. Difference between star and snowflake schema.
6. Design star and snowflake schema for given system.
7. Difference between OLTP and OLAP.
8. What are different OLAP operations?Explain with example.
9. Problems on writing a sequence of OLAP operations for the given query.
10. Explain steps of KDD
11. State any 2 decision making activities for which organizations are using data in DWH.
12. What is concept hierarchy, partial and total order concept hierarchy? Ex[plain with an example.
13. What is data mining? State applications of data mining.
14. What are the different types of patterns that can be mined?

Module 2:

1. What are the different types of attributes? Explain with examples
2. Problems on basic statistical descriptions of data like finding mean, median, midrange standard deviation, variance,modes for given data.Drawing q-q plot and boxplot for given data.
3. What is a five number summary of data?
4. How can we compute dissimilarity between two binary attributes?
5. What is Euclidean distance, Manhattan distance, Minkowski distance? Problems on computing these distances between given objects.
6. What is cosine similarity?problems on finding similarity between given documents.
7. Problems based on finding dissimilarity matrices between nominal,binary and ordinal attributes .
8. Explain in brief the major tasks in data preprocessing.
9. What are the different ways to handle missing data?
10. What are the different ways to handle noisy data?
11. Problems on correlation analysis for categorical(Chi square test) and numerical data.
12. What are the different data transformation strategies?
13. Problems on min max ,z score and decimal scaling normalization.
14. State different data reduction strategies.

Module 3:

Classification:

Supervised and unsupervised learning

What is classification? classification applications

classification model building phases

Classification algorithms:

Explain the Decision tree-building process with an example.

Decision Tree algorithm

Entropy,Information Gain,Gain Ratio and Gini Index

Feature selection measures in building Decision Tree/splitting attribute selection measure.

Different Metrics used for Evaluating Classifier Performance

Confusion matrix:
Decision Tree Pruning

Problems on Decision Tree:

Naive Bayes Algorithm

Problems on Naive Bayes algorithm

State Bayes theorem. How can it be applied for data classification? b) With example explain Bayesian belief network.

Based on the following data determine the gender of a person having height 6 ft., weight 130 lbs. and foot size 8 in. (use Naive Bayes algorithm).

person	height (feet)	weight (lbs)	foot size (inches)
male	6.00	180	10
male	6.00	180	10
male	5.50	170	8
male	6.00	170	10
female	5.00	130	8
female	5.50	150	6
female	5.00	130	6
female	6.00	150	8