

DWM Question bank

Short Questions (for 2 Marks)

1. What are facts?
2. What are dimension table?
3. Define data cube?
4. Define factless fact table?
5. Define Data Warehouse?
6. What are the Benefits of Data Warehousing?
7. Explain the Features of a Data Warehouse?
8. What are the various types of Hierarchical Clustering
9. What is a dendrogram
10. What are Crawlers?
11. Define web mining?
12. Differentiate between data mining and web mining?
13. What is usage mining?
14. What is a strong association rule?
15. Define the term "Data warehouse".
16. Define the term "Data mart".
17. What is dendrogram with respect to clustering algorithms?
18. State drawbacks of Apriori algorithm

Long Question (for 10 Marks)

1. Explain three tier data warehouse architecture with suitable block diagram?
2. A dimension table is wide and fact table is deep explain? What is star schema and snowflakes schema? what are its advantages and disadvantages?
3. Draw a star schema for sales occupancy with dimension like time hotel.
4. Draw a star schema and snowflake schema for super market chain
5. i. Design star schema for "Hotel Occupancy" considering dimensions like Time, Hotel, Room, etc. (10 marks)

ii. Calculate the maximum number of base fact table records for the values given below: Time period: 5 years Hotels: 150 Rooms: 750 rooms in each Hotel (about 400 occupied in each hotel daily). (5 marks)

6. Explain star schema, snowflake schema, and factless fact table with example.
7. For Hotel occupancy considering dimensions like time, hotel etc. Design star schema
8. Explain the different steps in ETL process with suitable diagram
9. Explain ETL of data warehousing In detail?
10. Explain different OLAP operations?
11. Explain the different architectural options for OLAP.
12. Compare OLTP and OLAP systems
13. The college wants to record the Marks for the courses completed by students using the dimensions: I) Course, II) Student, III) Time & a measure Aggregate marks
Create a Cube and perform following OLAP operations:
b) Rollup ii) Drill down iii) Slice iv) Dice v) Pivot.
14. Explain slice-and-dice , rollup, pivot, drill down with examples
15. Consider a data warehouse for a hospital, where there are three dimensions 1) doctor, (2) patient and (3) time, and two measures: (1) count and (2) charge, where charge is the fee that the doctor charges a patient for a visit. Using the above example describe the following OLAP operations: slice, dice, roll up, drill down and pivot.
16. All electronics company have sales department Sales consider three dimensions
(b) Time (ii) Product (iii) Store.

The schema contain a central fact table sales with two measures.

(b) dollars-cost and (ii) units-sold

Using the above example describe the following OLAP operations

(i) Dice (ii) Slice (iii) Roll-up (iv) Drill-down

17. Consider the following transactions

Transaction_Id	Items
A	1, 3, 4, 6
B	2, 3, 5, 7
C	1, 2, 3, 5, 8
D	2, 5, 9, 10
E	1,4

Apply the Apriori Algorithm with minimum support of 30% and minimum

confidence of 75% and find all frequent item sets and strong association rules

17. Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are: A1 (2, 10), A2 (2, 5), A3 (8, 4), B1 (5, 8), B2 (7,5), B3 (6, 4), C1 (1, 2), C2 (4, 9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only (i) The three cluster centers after the first round of execution.

18. Differentiate between simple linkage, average linkage, and complete linkage algorithms. Use simple linkage, average linkage and complete linkage algorithm to find the clusters from the following dataset.

X	4	8	15	24	24
Y	4	4	8	4	12

19. Explain K-Means clustering algorithm? Apply K-Means algorithms for the following data set with two cluster. Data Set = {1, 2, 6, 7, 8, 10, 15, 17, 20}

20. Discuss Association Rule Mining and Apriori Algorithm. Apply AR Mining to find all frequent item sets and association rules for the following dataset:

Minimum Support Count = 2

Minimum Confidence = 70%

Transaction_ID	Items
100	1, 2, 5
200	2, 4
300	2, 3

400	1, 2, 4
500	1, 3
600	1, 3
700	1, 3, 2, 5
800	1, 3
900	1, 2, 3

19. Given a transaction dataset above, let min-sup-60% and min-conf 80%;
- Find all frequent itemsets using the Apriori algorithm.
 - List all the strong association rules.

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, K, I, E}

21. Apply FP growth Mining to find all frequent item sets and association rules for the following dataset:
- Minimum Support Count = 2
- Minimum Confidence = 70%

TID	Items
1	A, B
2	B, C, D
3	A, C, D, E
4	A, D, E
5	A, B, C

22. Frequent pattern mining algorithms consider 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 itemset efficiently consider multiple occurrences of items? Generate a Frequent Pattern Tree for the following transaction with 30% minimum support:

Transaction ID	Items
T1	E,A,D,B
T2	D,A,C,E,B
T3	C,A,B,E
T4	B,A,D
T5	D
T6	D,B
T7	A,D,E
T8	B,C

23. Construct dendrogram using single and average link approach for the following adjacency matrix. Show all the steps

	A	B	C	D	E	F
A	0					
B	6	0				
C	1	4	0			
D	3	2	7	0		
E	1	6	4	5	0	
F	2	6	1	5	4	0

24. Explain page rank algorithm in detail
 25 Explain different Updates in Dimension Table

Long Question (for 05 Marks)

- (a) Explain the data staging area and related functions with respected to a typical data warehouse architecture.

(b) Use complete linkage algorithm to form clusters from the following dataset

X	4	8	15	24	24
Y	4	4	8	4	12