① Explain OLAP Server Architecture.

ROLAP, MOLAP, HOLAP..

ROLAP → ROLAP servers are placed between relational backend servers and client from front-end tools. It uses relational or extended DBMS to store and manage warehouse data. ROLAP has basically 3 main components:

Database server, ROLAP server and front end tools.

Advantages → Data can be stored efficiently. ROLAP is used for handle the large amount of data.

Disadvantages → Performance of ROLAP can be slow. Limited by SQL functionalities.

MOLAP → MOLAP does not uses relational database to storage. It stores in optimized multidimensional array storage. The storage utilization may be low with multidimensional data stores. Many MOLAP servers handle dense and sparse data sets by using two levels of data storage representation.

MOLAP has 3 components:

Database server, MOLAP server and front-end tools.

Advantages → MOLAP is basically used for complex calculation. MOLAP is optimal for operation such as slice and dice.

Disadvantages → MOLAP can't handle large amount of data. In MOLAP, requires additional investment.

HOLAP → HOLAP is a combination of both ROLAP and MOLAP. It offers functionalities of both ROLAP and as well as MOLAP like faster computation and higher scalability of ROLAP.

The aggregations are stored separately in MOLAP store. Its server allows storing the large data volumes of detailed information.

Advantages → HOLAP provides the functionalities of both MOLAP and ROLAP.

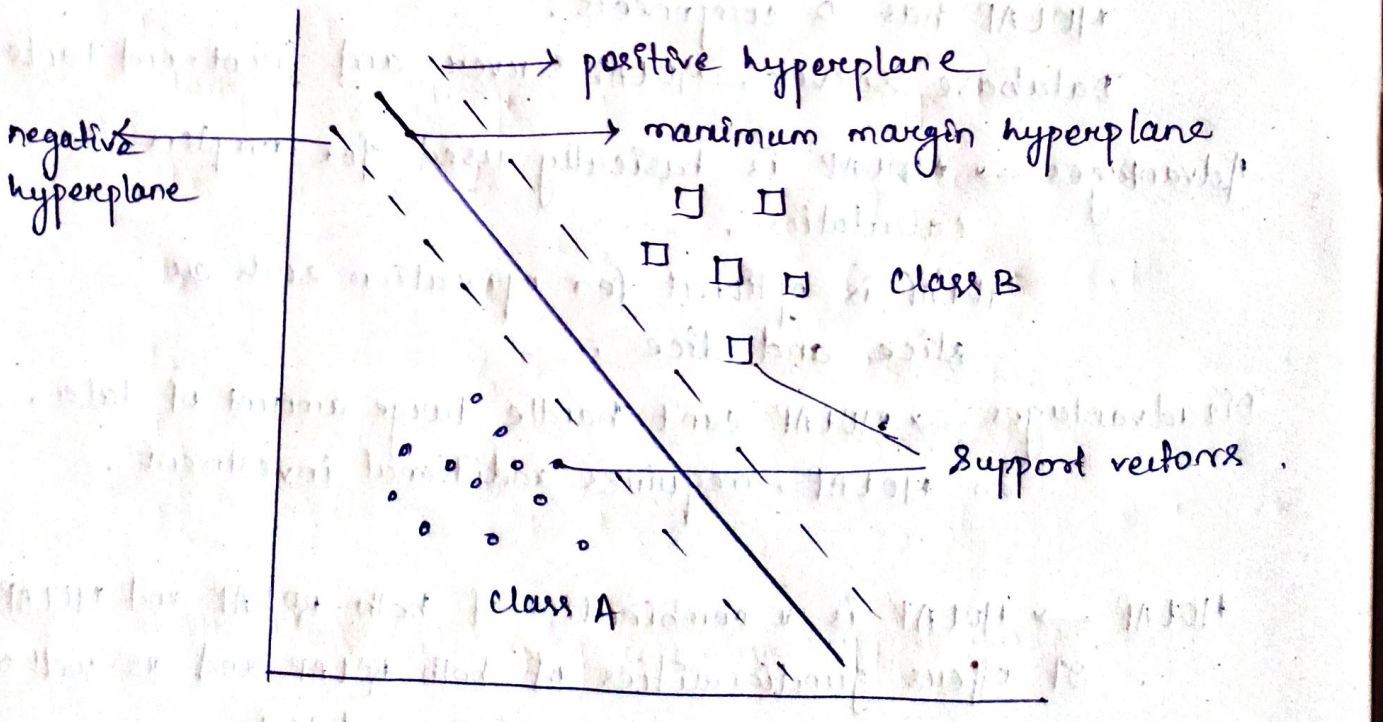
HOLAP provides fast access at all levels of aggregation.

Disadvantages → HOLAP architecture is very complex to understand because it supports both MOLAP and ROLAP.

(2) Write a short note on Support vector Machine.

→ Support vector machine (SVM) is one of the most popular supervised ~~too~~ learning algorithm which is used for classification as well as regression problem.

→ The goal of the SVM algorithm is to create the best line or hyperplane that can separate n-dimensional space in to classes so that we can easily put new point in the correct category in the future.



→ SVM choose the extreme points / vectors that help in creating the ~~extra~~ hyperline. These extreme cases are called as Support vectors and hence algorithm is termed as support vector machine.

(3) Back Propagation Algorithm.

→ It is an algorithm that back propagates the errors from output nodes to the input nodes. Therefore it is simply refer to as back propagation of errors.

→ It is used in various applications of neural network in data mining like character recognition, signature, verification etc.

Algorithm —

Step-1 - Input x arrived through the preconnected path.

Step-2 The input is modeled using true weights. weights are usually choosen randomly.

Step-3 Calculate the output of each neuron from the input layer to the hidden layer to the output layer.

Step-4 Calculate the errors in outputs.

Step-5 From the output layer go back to the hidden layer to adjust the weights to reduce errors.

Step-6 Repeat the process until the desired output is achieved.

Advantages →

- It is simple, fast and easy to program.
- Only no of input are required, no need of any other parameter.
- It is flexible and efficient.
- no need for users to learn any special functions.

disadvantages —

(1) noisy data leads to inaccurate results.

(2) Performance is highly dependent on input data.

(3) Spending too much time in training.

(4) For the following given transaction dataset generate rules using apriorce algorithm. consider values has support = 22 % and confidence = 70 %

Transaction ID	Item sets
1	I_1, I_2, I_5
2	I_2, I_4
3	I_2, I_3
4	I_1, I_2, I_4
5	I_1, I_3
6	I_2, I_3
7	I_1, I_3
8	I_1, I_2, I_3, I_5
9	I_1, I_2, I_3

Item	frequency	support
I_1	6	$6/9 \times 100 = 66.6$
I_2	7	$7/9 \times 100 = 77.7$
I_3	6	66.6
I_4	2	22.2
I_5	2	22.2

All items satisfying minimum support count.

Item	frequency	Support
$\{I_1, I_2\}$	4	$4/9 \times 100 = 44$
$\{I_1, I_3\}$	4	44
$\{I_1, I_4\}$	1	11 (x)
$\{I_1, I_5\}$	2	22
$\{I_2, I_3\}$	4	44
$\{I_2, I_4\}$	2	22
$\{I_2, I_5\}$	2	22
$\{I_3, I_4\}$	0	0 (x)
$\{I_3, I_5\}$	1	11 (x)
$\{I_4, I_5\}$	0	0 (x)

Item	frequency	support
$\{I_1, I_2, I_3\}$	2	22
$\{I_1, I_2, I_5\}$	2	22
$\{I_1, I_2, I_4\}$	1	11 (x)
$\{I_1, I_3, I_5\}$	1	11 (x)
$\{I_2, I_3, I_4\}$	0	0 (x)
$\{I_2, I_3, I_5\}$	1	11 (x)
$\{I_2, I_4, I_5\}$	0	0 (x)

Items	frequency	support
$\{I_1, I_2, I_3\}$	2	22%
$\{I_1, I_2, I_5\}$	2	22%

Items	frequency	support
$\{I_1, I_2, I_3, I_5\}$	1	11.11 %

It does not satisfy min^n support count.

(1) $\{I_1, I_2, I_3\}$

(2) $\{I_1, I_2, I_5\}$

$\{I_1, I_2, I_3\}$

⇓ non-empty sets

Confidence

R_1 $\{I_1\} \longrightarrow \{I_2, I_3\} \longrightarrow \frac{2/9}{6/9} \times 100 = 33.33$ (X)

R_2 $\{I_2\} \longrightarrow \{I_1, I_3\} \longrightarrow \frac{2/9}{7/9} \times 100 = 28.57$ (X)

R_3 $\{I_3\} \longrightarrow \{I_1, I_2\} \longrightarrow \frac{2/9}{6/9} \times 100 = 33.33$ (X)

R_4 $\{I_1, I_2\} \longrightarrow \{I_3\} \longrightarrow \frac{2/9}{4/9} \times 100 = 50$ (X)

R_5 $\{I_1, I_3\} \longrightarrow \{I_2\} \longrightarrow \frac{2/9}{4/9} \times 100 = 50$ (X)

R_6 $\{I_2, I_3\} \longrightarrow \{I_1\} \longrightarrow \frac{2/9}{4/9} \times 100 = 50$ (X)

R_7 $\{I_1, I_2, I_3\}$ 0

$\{I_1, I_2, I_5\}$

⇓ non-empty sets.

R_8 $\{I_1\} \longrightarrow \{I_2, I_5\} \longrightarrow \frac{2/9}{6/9} \times 100 = 33.33$ (X)

R_9 $\{I_2\} \longrightarrow \{I_1, I_5\} \longrightarrow \frac{2/9}{7/9} \times 100 = 28.57$ (X)

R_{10} $\{I_5\} \longrightarrow \{I_1, I_2\} \longrightarrow \frac{2/9}{2/9} \times 100 = 100$ (✓)

R_{11} $\{I_1, I_2\} \longrightarrow \{I_5\} \longrightarrow \frac{2/9}{4/9} \times 100 = 50$ (X)

R_{12} $\{I_1, I_5\} \longrightarrow \{I_2\} \longrightarrow \frac{2/9}{2/9} \times 100 = 100$ (✓)

R_{13} $\{I_2, I_5\} \longrightarrow \{I_1\} \longrightarrow \frac{2/9}{2/9} \times 100 = 100\ (\checkmark)$

R_{14} $\{I_1, I_2, I_5\} \longrightarrow 0 \longrightarrow 0$.

R_{10}, R_{12}, R_{13} satisfy the minimum confidence threshold. So there are strong association rules.

(5) Generate FP tree for following transaction dataset.
minimum support = 30%.

Tid	Dataset
1	E, A, D, B
2	D, A, C, E, B
3	C, A, B, E
4	B, A, D
5	D
6	D, B
7	A, D, E
8	B, C

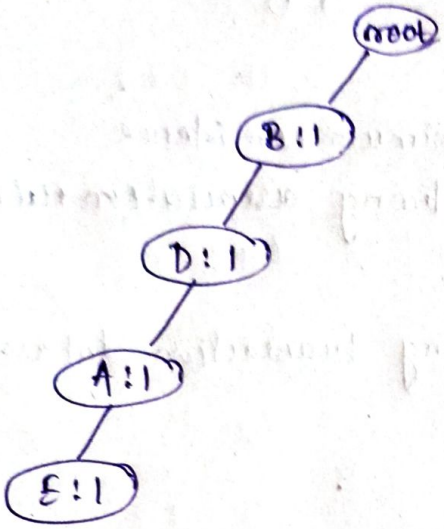
Item set	frequency	support count
A	5	5/8 - 62
B	6	6/8 - 75
C	3	3/8 - 37
D	6	6/8 - 75
E	4	4/8 - 50%.

Item set
B
D
A
E
C

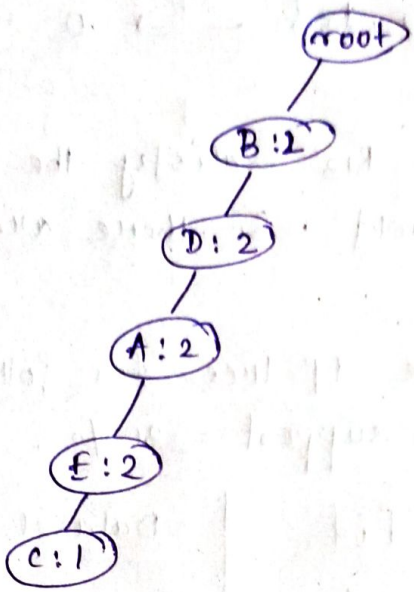
Itemset	Frequency	Support count
B	6	6/8 = 75
D	6	6/8 = 75
A	5	5/8 = 62
E	4	4/8 = 50

C 3 3/8 = 37

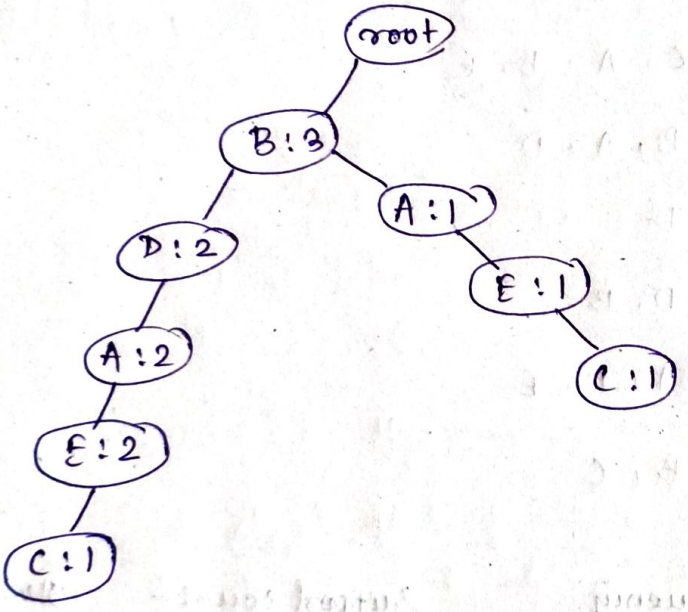
1. EADB → BDAE



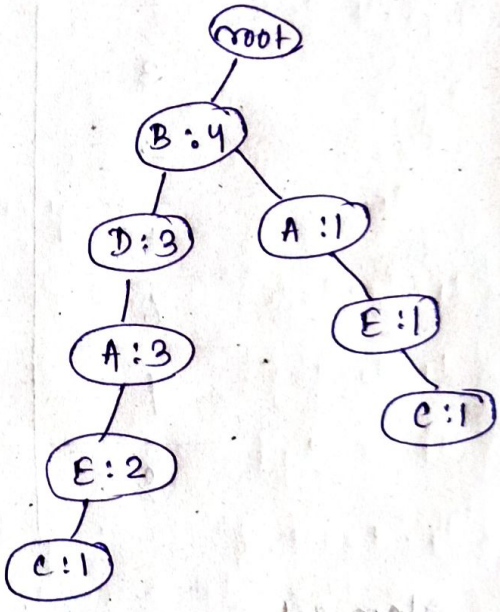
2. DACEB → BDAEC



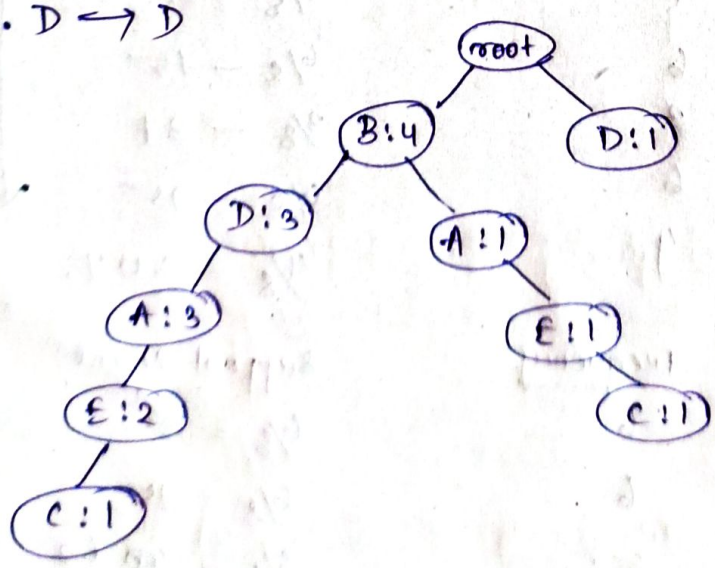
3. CABE → BAEC



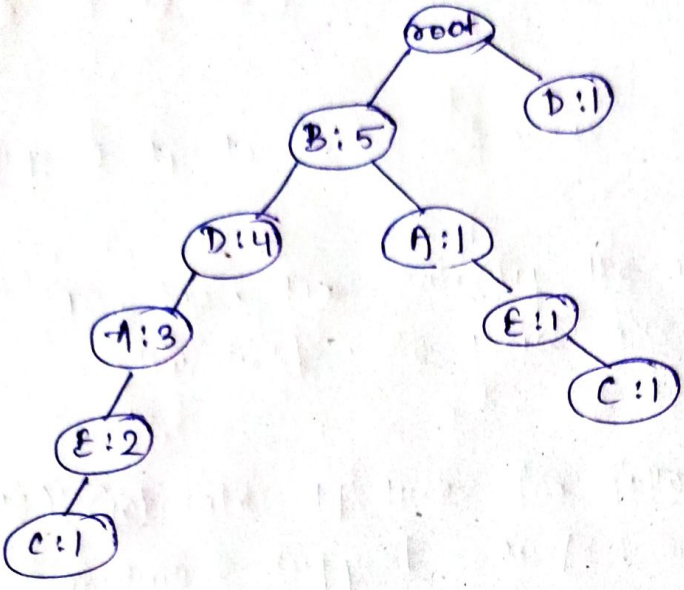
4. BAD → BDA



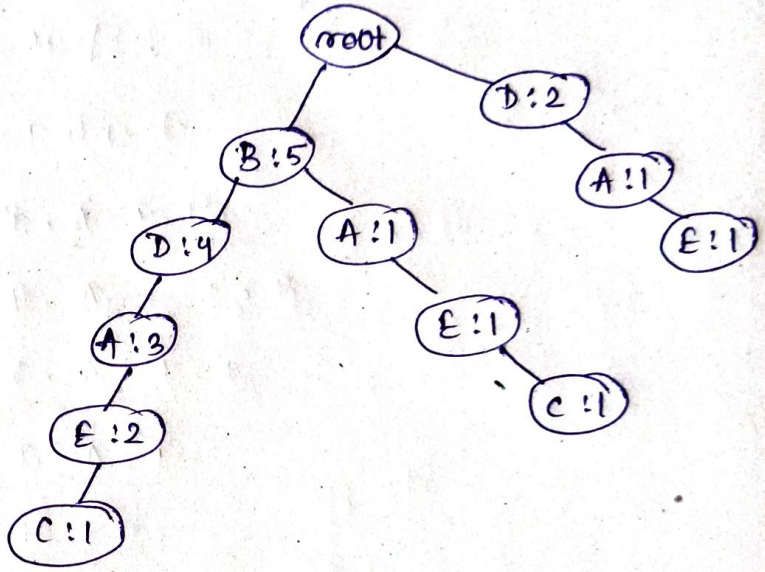
5. D → D



6. D B ⟶ B D



7. A D E ⟶ D A E



8. B C ⟶ B C