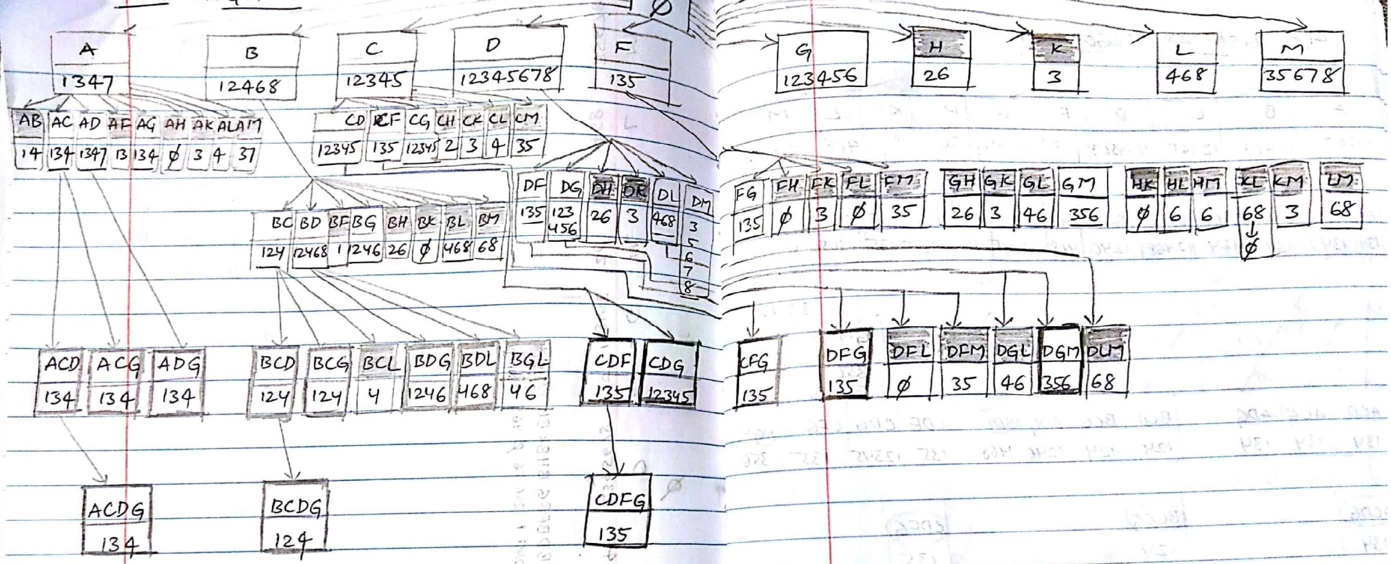
4(a) **ECLAT ALGORITHM:**

Vertical Data Layout

| A | B | C | D | F | G | H | K | L | M |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 4 | 3 |
| 3 | 2 | 2 | 2 | 3 | 2 | 6 | | 6 | 5 |
| 4 | 4 | 3 | 3 | 5 | 3 | | | 8 | 6 |
| 7 | 6 | 4 | 4 | | 4 | | | | 7 |
| | 8 | 5 | 5 | | 5 | | | | 8 |
| | | | 6 | | 6 | | | | |
| | | | 7 | | | | | | |
| | | | 8 | | | | | | |

→ ECLAT ALGORITHM.




Frequent Itemsets :

ACDG , BCDG , CDFG

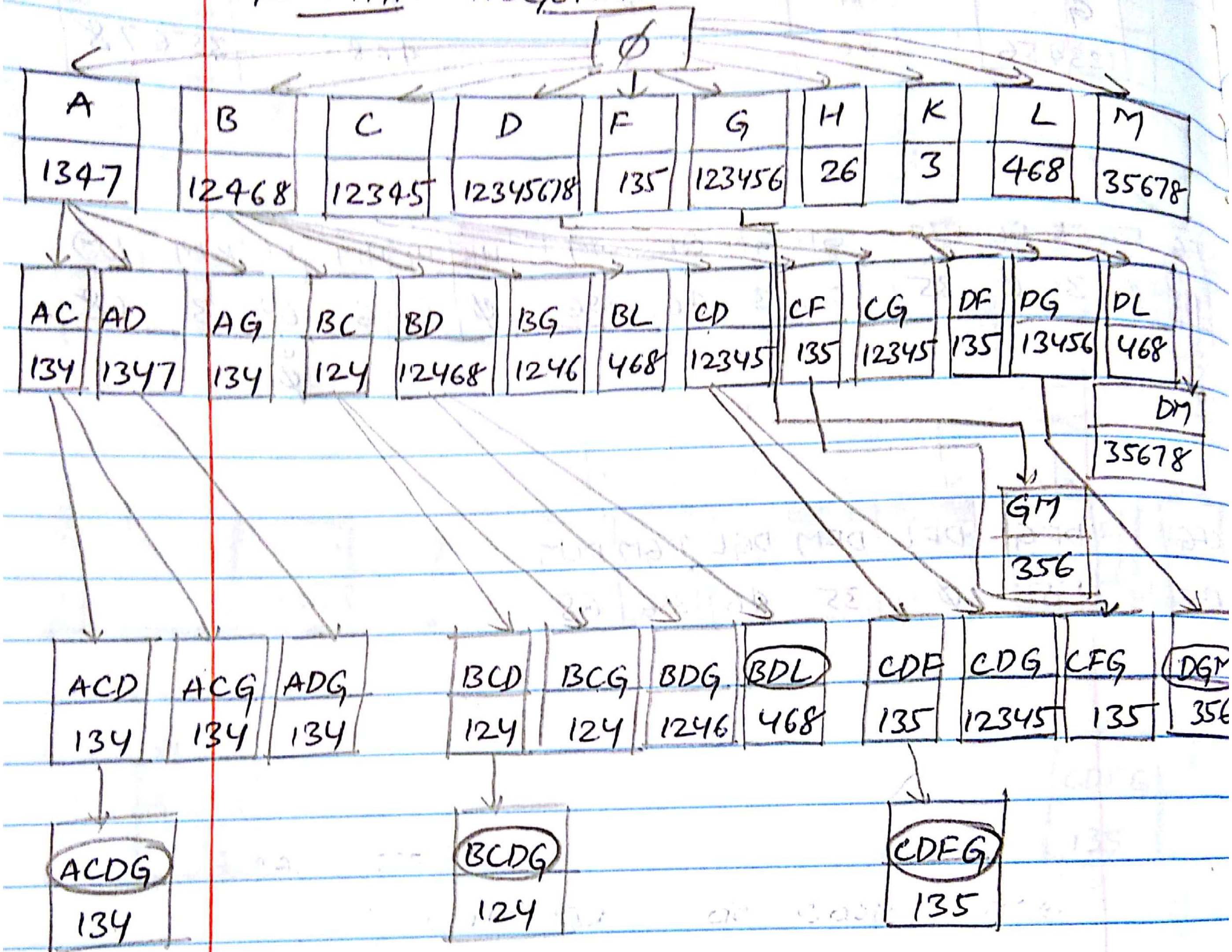
ACD, ACG, ADG

BCD, BCG, BDG, BDL

CDF, CDG, CFG

DFG, DGM

4(b) GENMAX ALGORITHM:



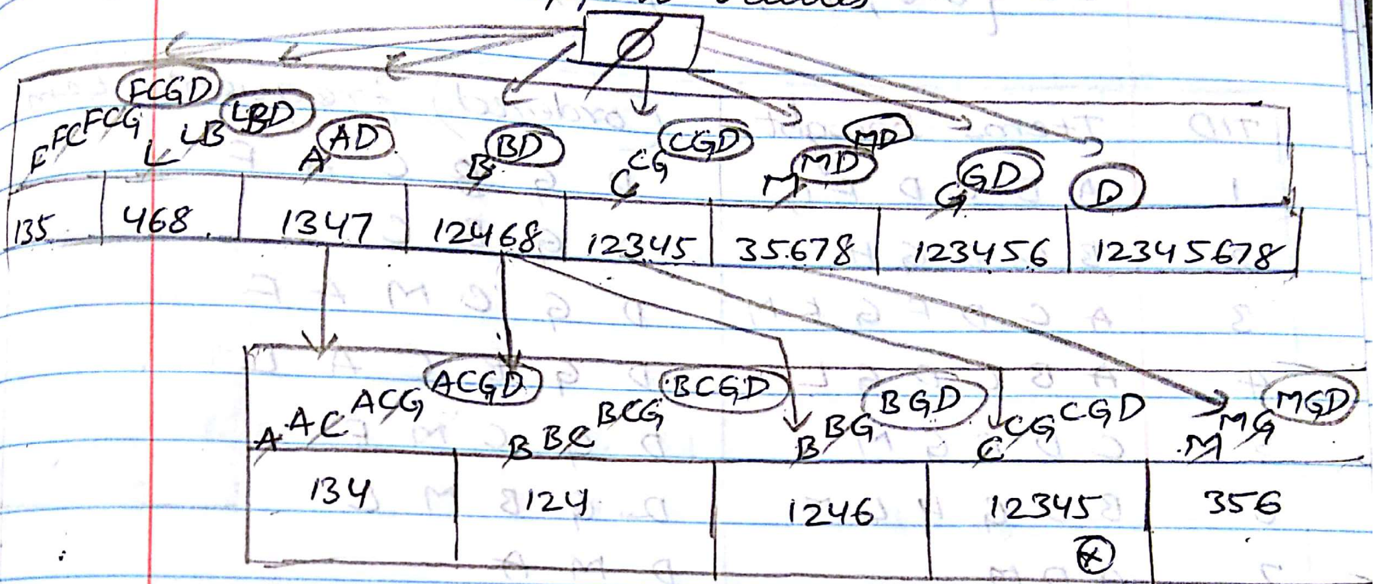

* Maximal Frequent Itemsets:

ACDG, BCDG, CDFG

BDL, DGM

(c) CHARM's ALGORITHM:

Arranging the values in increasing order of their support values




* Closed Itemsets using charm's algorithm

FCGD LBD AD, BD, CGD MD GD D*

ACGD BCGD BGD MGD

(d) Ordering the itemset in descending order

L = {D:8, G:6, M:5, C:5, B:5, A:4, L:3, F:3}

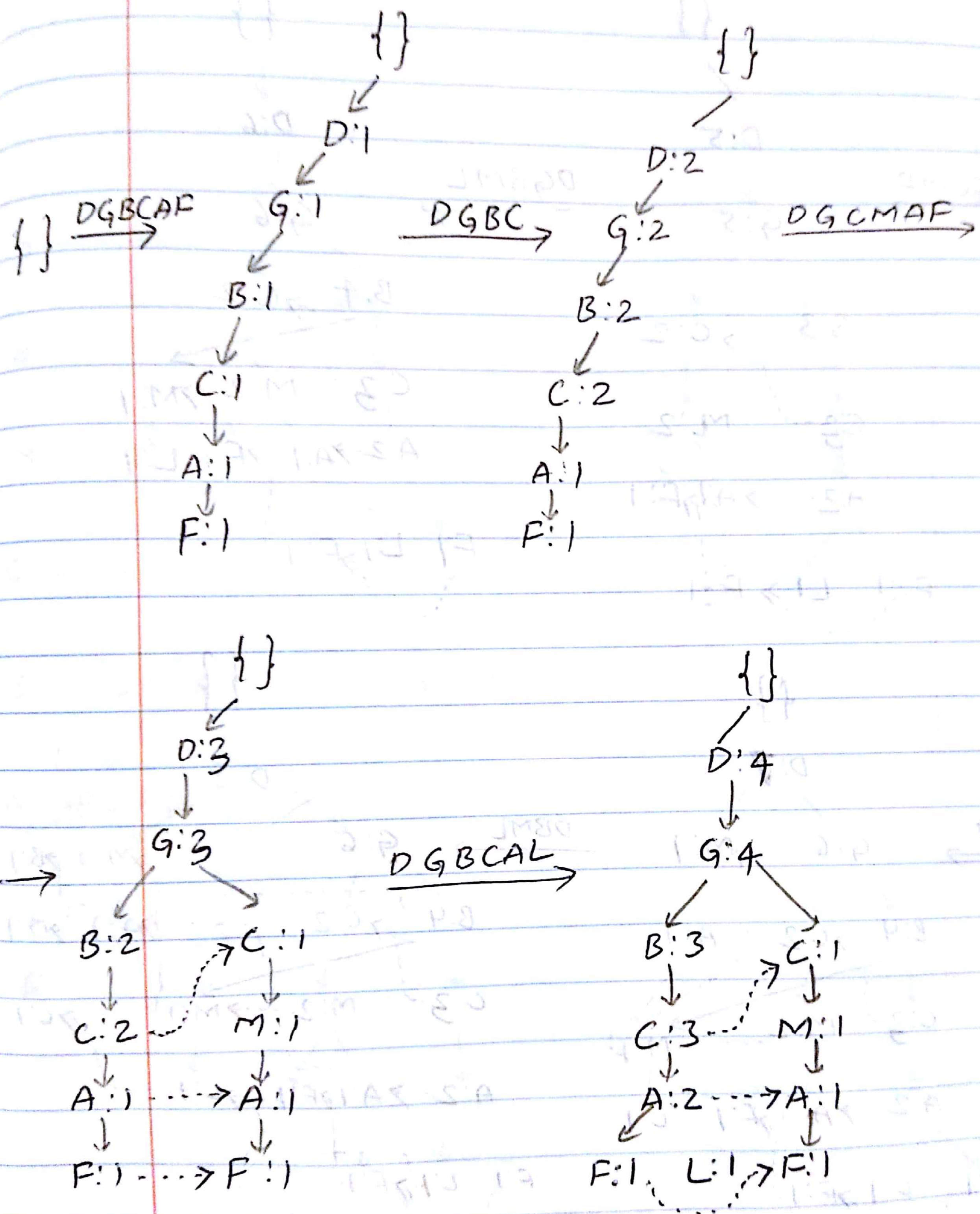
| TID | Items bought | (ordered) frequent items |
|---|---|---|
| 1 | A B C D F G | D G B C A F |
| 2 | B C D G H | D G B C |
| 3 | A C D F G K M | D G C M A F |
| 4 | A B C D G L | D G B C A L |
| 5 | C D F G M | D G C M F |
| 6 | B D G H L M | D G B M L |
| 7 | A D M | D M A |
| 8 | B D L M | D B M L |

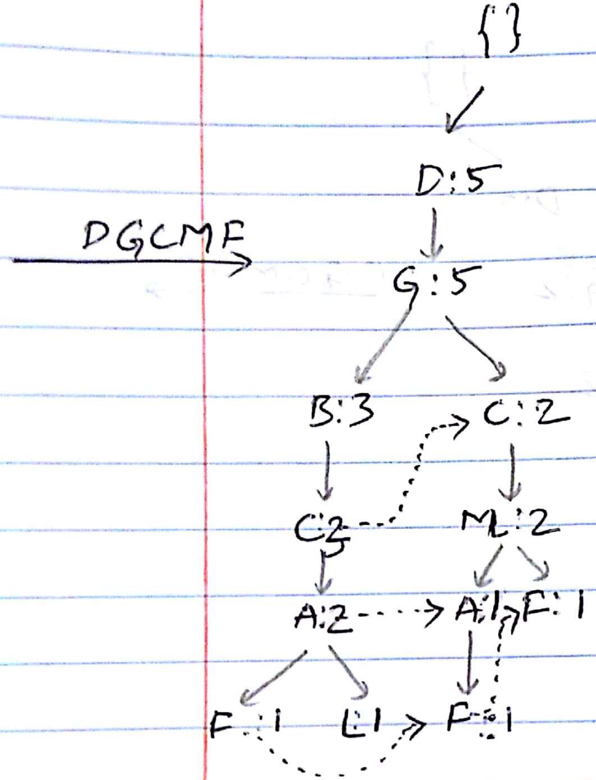
* MAIN-STEPS:

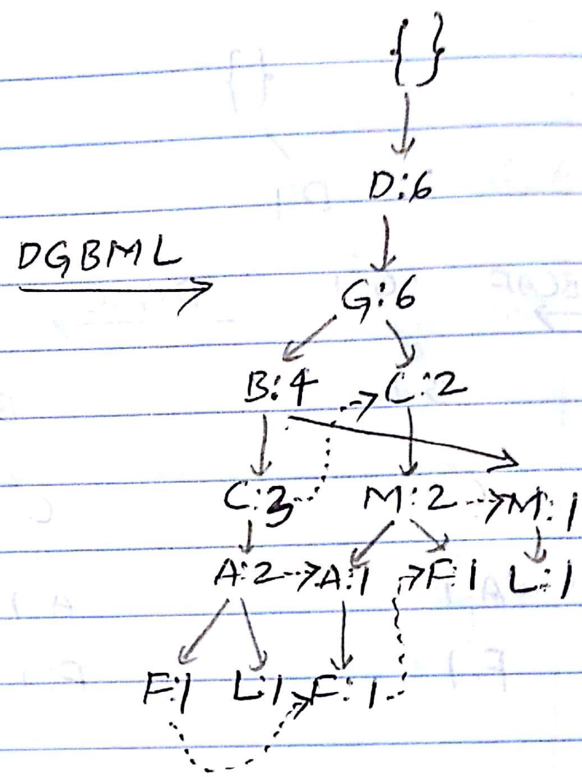
1. order the itemset in decreasing order of count
2. start building the tree by starting with each Itemset (ordered in the decreasing order)
3. Proceed through increasing the transactions item count TID 1 to 8 and design the tree structure
4. Obtain the final FP-Tree and conditional FP-Tree.

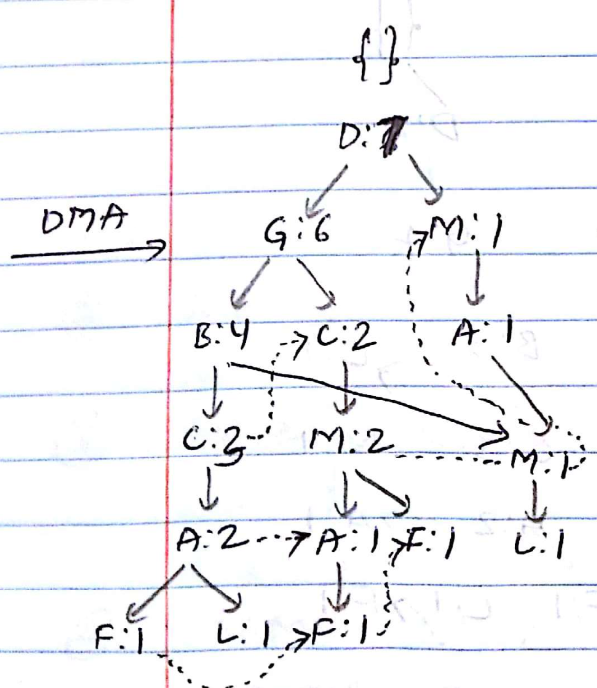
Scanned by CamScanner

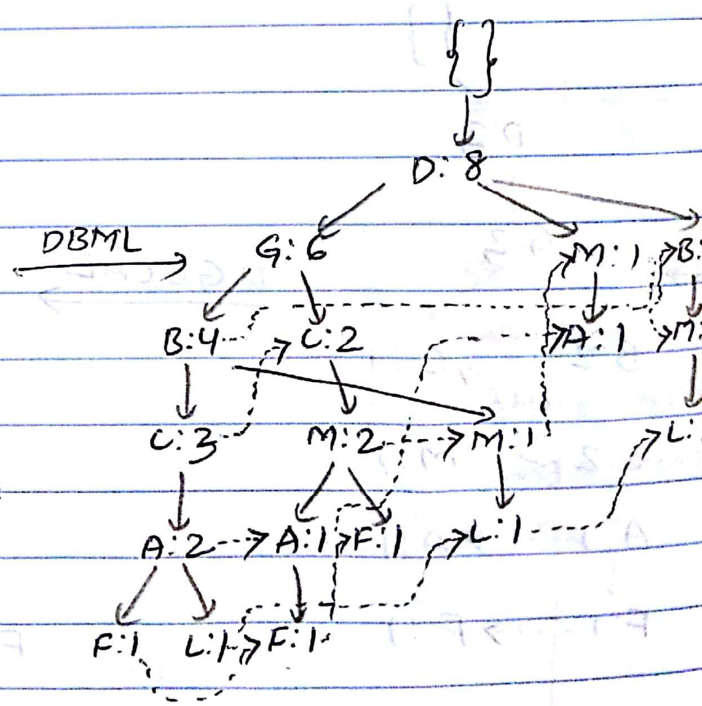
# FP TREE :


{}
DGBCAF
D:1
G:1
B:1
C:1
A:1
F:1
DGBC
D:2
G:2
B:2
C:2
A:1
F:1
DGCMAF
D:3
G:3
B:2
C:1
C:2
M:1
A:1
A:1
F:1
F:1
DGBCAL
D:4
G:4
B:3
C:1
C:3
M:1
A:2
A:1
F:1
L:1
F:1

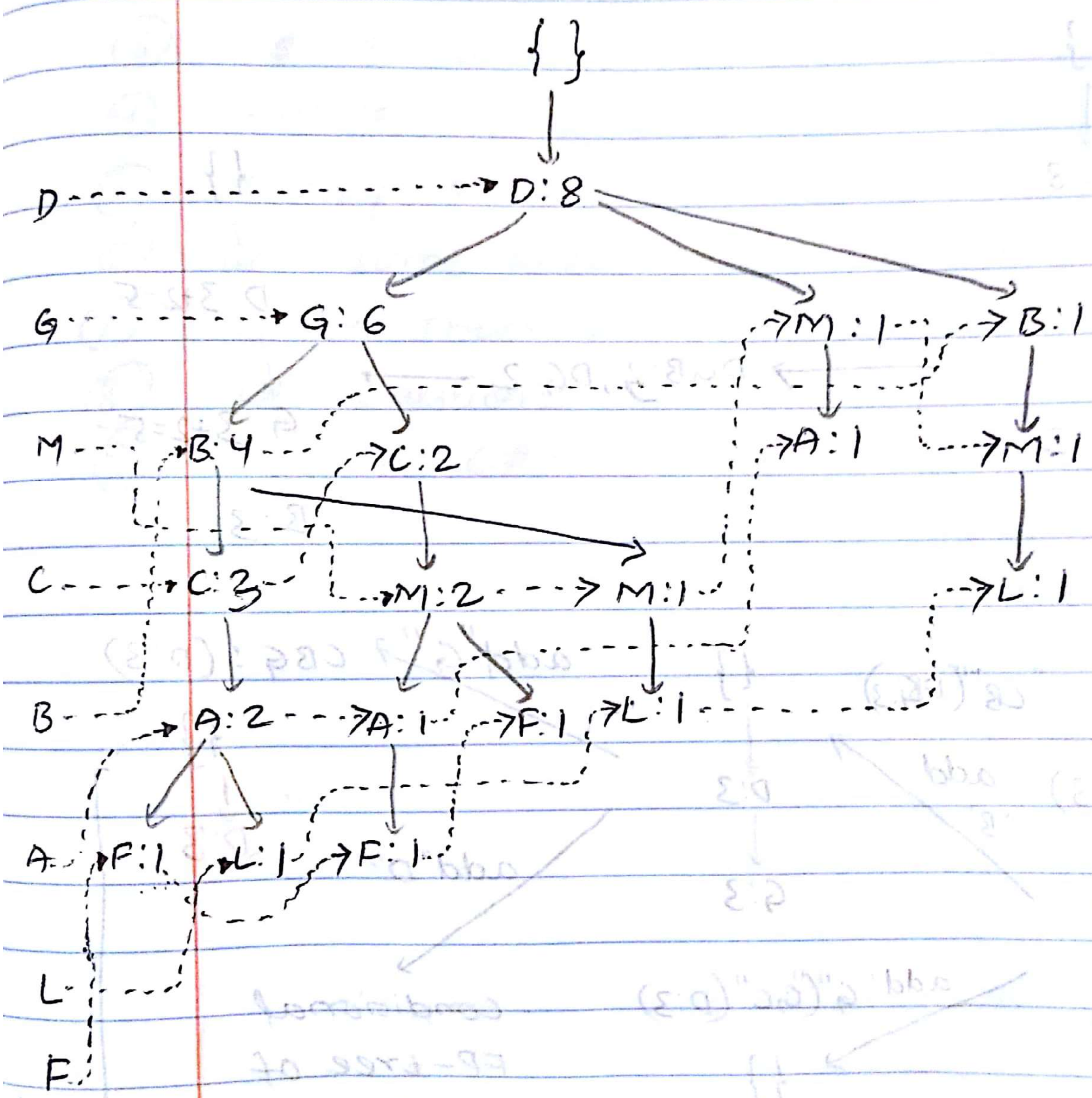

DGCMF
{}
D:5
G:5
B:3
C:2
C:3
M:2
A:2
A:1
F:1
F:1
L:1
F:1


DGBML
{}
D:6
G:6
B:4
C:2
C:3
M:2
M:1
A:2
A:1
F:1
L:1
F:1
L:1
F:1


DMA
{}
D:7
G:6
M:1
B:4
C:2
A:1
C:3
M:2
M:1
A:2
A:1
F:1
L:1
F:1
L:1
F:1


DBML
{}
D:8
G:6
M:1
B:1
B:4
C:2
A:1
M:1
C:3
M:2
M:1
L:1
A:2
A:1
F:1
L:1
F:1
L:1
F:1

* Final FP Tree.



## Conditional FP Tree

A : DM:1 , DGCM:1, DGBC: 2

B : DG : 4 , D:1

C : DGB:3 , DG:2 ⇒ {DG : 5}

D : {}

G : D:6

L : DGBM:1 , DBM:1 ⇒ {DBM:2}

M : DGC:2 , DGB:1 , D:1 , DB:1 ⇒ D:5

F : DGBCA:1, DGCMAF:1 , DGCM:1

{DGC:3}

C : DGB:3 , DG:2

{DG:5}

Frequent itemsets that contain C

DGCF:3, DGC:5

All Frequent itemsets

C , CD , CG, CF, CDG, CDF, CFG, DCFG

{}

↓

D:5

↓

G:5

↓

C:5

↓

F:3

* Conditional FP-Tree of "C":
(DGF)

{}
↓
D:5
↓
G:3
↓
F:3

add F → conditional FP Tree "CF"
(DG)

{}
↓
D:5
↓
G:5

add "G" → conditional FP Tree "CFG"

{}
↓
D:5

add "D" → conditional FP Tree "CFD"

{} 3

conditional Add "D" → Conditional FP Tree "CFGD" 3

add G → conditional FP Tree "CG"
(DF)

{}

D:5
↓
F:3

add D → conditional FP tree "CD"
(GF)

{} → 3

D:5

5.

| | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|
| Pa | 0 | 0 | 1 | 1 | 1 | 0 |
| Pb | 1 | 1 | 0 | 0 | 1 | 1 |
| Pc | 0 | 1 | 1 | 1 | 0 | 0 |
| Pd | 1 | 0 | 1 | 1 | 1 | 1 |
| Pe | 1 | 1 | 0 | 0 | 0 | 1 |
| Pf | 0 | 1 | 0 | 0 | 1 | 1 |
| Pg | 1 | 1 | 1 | 1 | 1 | 1 |
| Ph | 1 | 1 | 0 | 1 | 1 | 1 |
| Pj | 0 | 0 | 1 | 1 | 1 | 0 |

MAXIMAL ITEMSETS:

(i) ADGJ
(ii) BDHG.
(iii) EBHG.

CLOSED ITEMSETS

(i) ADGJ — support 3 (3)
(ii) EBGH — support 3 (3)
(iii) BGH — support 4 (4)
(iv) BFGH — support 3 (3)

Scanned by CamScanner

Supporting sets for Rows

(i) ADGJ

| | | |
|---|---|---|
| A | 001110 | 345 |
| D | 101111 | 13456 |
| G | 111111 | 123456 |
| J | 001110 | 345 |

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|
| $P_A$ | 0 | 0 | 1 | 1 | 1 | 0 |
| $P_D$ | 1 | 0 | 1 | 1 | 1 | 1 |
| $P_G$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $P_J$ | 0 | 0 | 1 | 1 | 1 | 0 |

→ Maximal frequent itemset

ADGJ is present in transactions 345
(support = 3)

(ii)

| BDHG | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|
| $P_B$ | 1 | 1 | 0 | 0 | 1 | 1 |
| $P_D$ | 1 | 0 | 1 | 1 | 1 | 1 |
| $P_H$ | 1 | 1 | 0 | 1 | 1 | 1 |
| $P_J$ | 0 | 0 | 1 | 1 | 1 | 0 |
| $P_G$ | 1 | 1 | 1 | 1 | 1 | 1 |

writing $S_1$ at last in BDHG

| | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_1$ |
|---|---|---|---|---|---|---|
| $P_B$ | 1 | 0 | 0 | 1 | 1 | 1 |
| $P_D$ | 0 | 1 | 1 | 1 | 1 | 1 |
| $P_H$ | 1 | 0 | 1 | 1 | 1 | 1 |
| $P_G$ | 1 | 1 | 1 | 1 | 1 | 1 |

BDHG is present in transactions 156
(support = 3)

Closed Itemsets

(i) ADGJ:

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|
| $P_d$ | 0 | 0 | 1 | 1 | 1 | 0 |
| $P_G$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $P_J$ | 0 | 0 | 1 | 1 | 1 | 0 |
| $P_A$ | 0 | 0 | 1 | 1 | 1 | 0 |

DGJ is closed frequent in transactions 345
(Support = 3)

(ii) EBGH

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|
| $P_B$ | 1 | 1 | 0 | 0 | 1 | 1 |
| $P_G$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $P_H$ | 1 | 1 | 0 | 1 | 1 | 1 |
| $P_E$ | 1 | 1 | 0 | 0 | 1 | 1 |

EBGH is closed frequent in transactions
156
(support = 3)

* **Explanation:**

Maximal Itemsets: ADGJ

(i) consider the set of items ADGJ
If we check the immediate superset of ADGJ for Example ACDGJ which has association support 2(S3 S4) < 3, Hence ACDGJ is infrequent. Similarly if we check ABDGJ, ADEGJ, ADGHJ... all are infrequent itemsets.
Hence by the definition of Maximal itemset, we infer that ADGJ is a maximal itemset

(ii) BDGH → consider the itemset BDGH. As in the above case, we try to append any other item with the above itemset and the Resultant Itemset becomes infrequent. Hence BDGH is a maximal itemset. It is a maximal itemset formed by joining any elements of BDGH.

## Closed Itemsets:

(i) ADGJ :

Closed Itemset Definition: Closed Itemset is an Itemset whose immediate supersets has support not equal to the Itemset.

If we consider the supersets of DGJ For Example BDAGJ it has a suppot of 1 (<3). Therefore, by verifying all such similar sets (or) supersets of ADGJ, we have the support ≠3, hence ADGJ is a closed Itemset

(ii) EBGH :

With the similar procedure as above if we verify the supersets of EBGH like EABGH, CEBGH, DEBGH, FEBGH they have support of. 0,1,2,2. respectively which is not equal to 3. Since the supersets of EBGH are having support #3(≠3) we conclude that EBGH is a closed itemset.

Scanned by CamScanner

(b) consider the example of closed set ADGJ, we have already seen that the support of superset of ADGJ is not equal to ADGJ. By Interpretation, since ADGJ is associated in transactions 345. Lets check the other items with these transactions

| | $S_3$ | $S_4$ | $S_5$ |
|---|---|---|---|
| A | 1 | 1 | 1 |
| D | 1 | 1 | 1 |
| G | 1 | 1 | 1 |
| J | 1 | 1 | 1 |
| B | 0 | 0 | 1 |
| C | 1 | 1 | 0 |
| E | 0 | 0 | 0 |
| F | 0 | 0 | 1 |
| H | 0 | 1 | 1 |

A, D, G, J → 345

B, C, E, F, H $\neq$ 1 1 1

closure of a set can be interpreted by projecting the items onto transaction and back

A → A

| | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ |
|---|---|---|---|---|---|---|---|---|
| ~~A~~ | 0 | 0 | 1 | 1 | 1 | 0 | | |
| B | 1 | 1 | 0 | 0 | 1 | 1 | | |
| C | 0 | 1 | 1 | 1 | 0 | 0 | | |
| ~~D~~ | 1 | 0 | 1 | 1 | 1 | 1 | | |
| E | 1 | 1 | 0 | 0 | 0 | 1 | | |
| F | 0 | 1 | 0 | 0 | 1 | 1 | | |
| ~~G~~ | 1 | 1 | 1 | 1 | 1 | 1 | | |
| H | 1 | 1 | 0 | 1 | 1 | 1 | | |
| ~~J~~ | 0 | 0 | 1 | 1 | 1 | 0 | | |

closure(A) = ADGJ

* Potential use of closed Itemsets :

1. They are useful in Removing some Redundant association Rules.
2. They are a subset of frequent itemsets and so they are a compact Representation.
3. They include the Necessary Information that helps us determine the support of their subsets.

→ Maximal and closed Itemsets

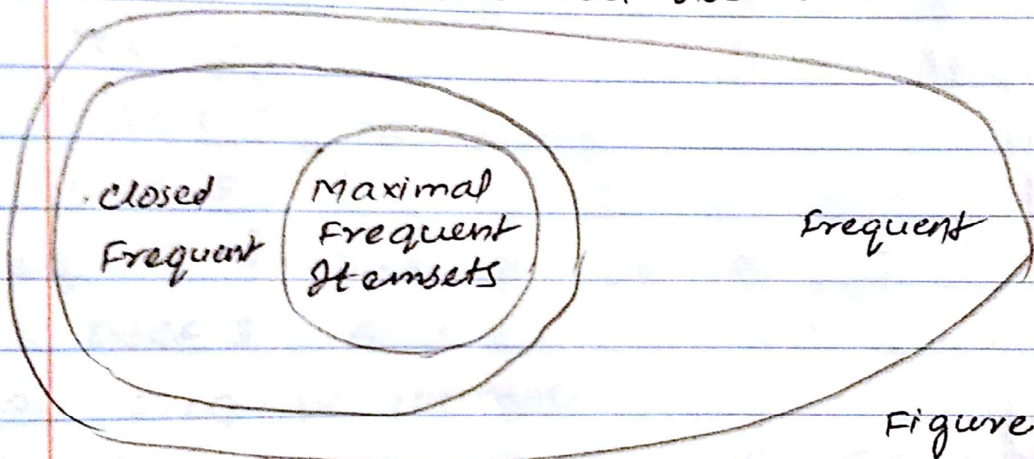



Figure (b).2

From the Figure (b).1 we can conclude that when we try to append an element (or) check its superset, the support of the newly formed itemsets (or) elements will not be equal to previous support.

From the Figure (b).2 we can say that all Maximal Frequent Itemsets will also be closed Frequent, because if the superset of itemsets are infrequent, the superset of items would have support not equal to prior support.

## (c) Maximal Frequent Itemsets:

consider a maximal frequent itemset BDGH. BDGH is present in Transactions 156. If we check any superset of BDGH we have found that it is infrequent.

Maximal Frequent itemsets refer to the itemset formed by the given items, where the items formed by the subset of the prior are infrequent not maximal. In other words, it is the itemset with maximal frequency possible with combinations of its subset being infrequent not maximal.

Example

BDHG

| | S1 | S5 | S6 |
|---|---|---|---|
| B | 1 | 1 | 1 |
| D | 1 | 1 | 1 |
| H | 1 | 1 | 1 |
| G | 1 | 1 | 1 |

If we try to add another item to BDHG, then the set has support < 3.

| Addition | Support | Conclusion |
|---|---|---|
| ABDHG | 0+1=1 | Infrequent |
| CBDHG | 0 | Infrequent |
| EBDHG | 2 | Infrequent |
| FBDHG | 2 | Infrequent |
| JBDHG | 1 | Infrequent |

→ Maximal Frequent set can also be interpreted by drawing a largest square (or) rectangles combining Items and transactions together.

* Potential use of Maximal Frequent Itemsets

① They are valuable because they provide a compact representation of frequent itemsets.

② In actuality they form of smallest representation of frequent itemsets and so they are the most practical to use when space is an issue.

* Limitations:

① They do not contain the support of their subsets and so an additional pass needs to be made of the data in order to find this information.

→ Maximal Frequent itemset is the largest frequent itemset that can be formed by the elements present it. when we try to append another element to maximal frequent set, then the resultant set would be infrequent.

Interpretation in terms of plants A B C D E F G H J and states S1, S2, S3, S4, S5, S6 using maximal and closed :

→ Itemset ADGJ is a maximal Itemset from which we can conclude that plants A, D, G, J is the highest set of plants found (or) associated together with which the plants A, D, G, J, given the plants from A to J.

→ Also, by knowing all the maximal Itemsets we can find the smallest representation of all the plants associated together.

→ Itemset BGH is a closed Itemset with transactions $S_1 S_2 S_5 6$ from which we can conclude that the plants B, G, H are found in the states $S_1$, $S_2$, $S_5$, $S6$ and by appendi another plant it will not be found in all the 4 states. That means that the set of BGH plants is the only associated plants found in all these 4 states together.

| Maximal Frequent Itemset | closed Frequent Itemset |
| --- | --- |
| ① It is a frequent itemset for which none of its immediate supersets are frequent. | ① It is a frequent itemset that is a superset of the maximal frequent itemset<br>→ A closed frequent itemset is a frequent itemset that is both closed and its support is greater than equal to minsup<br>→ An itemset X is closed in a dataset if there exists no superset that has the same support as X in the dataset |
| ② Identification<br>i) Examine frequent itemsets that appear at bottom between infrequent & frequent itemsets<br>ii) Identify all immediate supersets<br>iii) If there exists at least 1 superset that is frequent, the itemset is not maximal frequent, otherwise none of its items of immediate supersets are frequent, the itemset is maximal frequent. | ② Identification<br>i) First identify all frequent itemsets<br>2(ii) From this group find the closed by checking see if there exists a superset that was same support as frequent one.<br>(iii) Finally of these itemsets you have found to be closed, crosscheck that their support is greater than minsup. |