Vertical Apriori Algorithm
Equivalence class Transformation

* Both Apriori and FP-growth use horizontal data format

* ECLAT mines frequent itemset using the Vertical data format

* It is a depth first search based algorithm

* In this each item is stored together with its T-ID (Transaction ID)

* It uses intersection based approach to compute the support of an itemset.

Example

TID	LIST of items
T1	I_1, I_2, I_5
T2	I_2, I_4
T3	I_2, I_3
T4	I_1, I_2, I_4
T5	I_1, I_3
T6	I_2, I_3
T7	I_1, I_3
T8	I_1, I_2, I_3, I_5
T9	I_1, I_2, I_3

Min support count = 2
Confidence = 70%

Generate association rule using Vertical Apriori Algorithm.

DB in horizontal data format

Soln
Step 1

Database in Vertical data format.

Itemset	List of items
I_1	$T_1, T_4, T_5, T_7, T_8, T_9$
I_2	$T_1, T_2, T_3, T_4, T_6, T_8, T_9$
I_3	$T_3, T_5, T_6, T_7, T_8, T_9$
I_4	T_2, T_4
I_5	T_1, T_8

Step 2 — Itemset generated by intersection of 1 itemset

Itemset	List of Items
I_1	$T_1, I_4, T_5, T_9, T_8, T_9$
I_2	$T_1, T_2, T_3, T_4, T_6, T_8, T_9$
I_3	$T_3, T_5, T_6, T_9, T_8, T_9$
I_4	T_2, T_4
I_5	~~T_8, T_9~~ T_1, T_8

Frequent - 2 itemsets generated.

Itemset	List of items
I_1, I_2	T_1, T_4, T_8, T_9
I_1, I_3	T_5, T_9, T_8, T_9
I_1, I_4	T_4
I_1, I_5	T_1, T_8
I_2, I_3	T_3, T_6, T_8, T_9
I_2, I_4	T_2, T_4
I_2, I_5	T_1, T_8
I_3, I_4	$--$
I_3, I_5	T_8
I_4, I_5	$\cdots$

Itemset	List of Items	Items with min support count = 2
I_1, I_2	T_1, T_4, T_8, T_9	
I_1, I_3	T_5, T_9, T_8, T_9	
I_1, I_5	T_1, T_8	
I_2, I_3	T_3, T_6, T_8, T_9	
I_2, I_4	T_2, T_4	
I_2, I_5	T_1, T_8	

Step 3 : Frequent 3-itemset generated by intersection of 2-itemset

Itemset	List of items
I₁, I₂, I₃	T₈, T₉
I₁, I₂, I₅	T₁, T₈
I₁, I₃, I₅	T₈
I₂, I₃, I₄	—
I₂, I₃, I₅	T₈
I₂, I₄, I₅	—

Itemset with min support count = 2

Itemset	List of items
I₁, I₂, I₃	T₈, T₉
I₁, I₂, I₅	T₁, T₈

Step 4 : Generate frequent 4-items

Itemset	List of items
I₁, I₂, I₃, I₅	T₈

Itemset with min support count < 2

So stop the algorithm. Rules can be formed from the following three itemsets.

Itemset	List of Items
I₁, I₂, I₃	T₈, T₉
I₁, I₂, I₅	T₁, T₈

we can expand any rule   Confidence = 70%.
Ex I_1, I_2, I_5

Association Rule	Confidence	Confidence %
$I_1 \wedge I_2 \rightarrow I_5$	$c(I_1, I_2, I_5)/c(I_1, I_2)$ $= 2/4$	50%
✓ $I_1 \wedge I_5 \rightarrow I_2$	$c(I_1, I_2, I_5)/c(I_1, I_5)$ $= 2/2$	100%
✓ $I_2 \wedge I_5 \rightarrow I_1$	$c(I_1, I_2, I_5)/c(I_2, I_5)$ $= 2/2$	100%
$I_1 \rightarrow I_2 \wedge I_5$	$c(I_1, I_2, I_5)/c(I_1)$ $= 2/6$	33%
$I_2 \rightarrow I_1 \wedge I_5$	$c(I_1, I_2, I_5)/c(I_2)$ $= 2/7$	39%
✓ $I_5 \rightarrow I_1 \wedge I_2$	$c(I_1, I_2, I_5)/c(I_5)$ $= 2/2$	100%