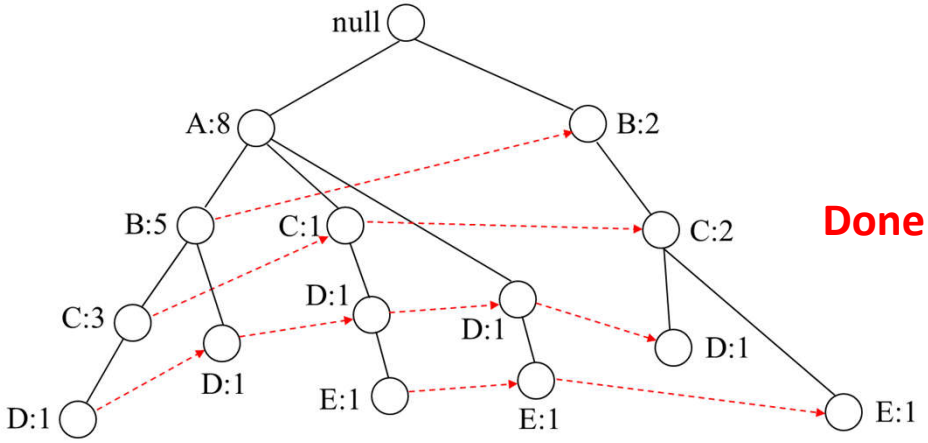
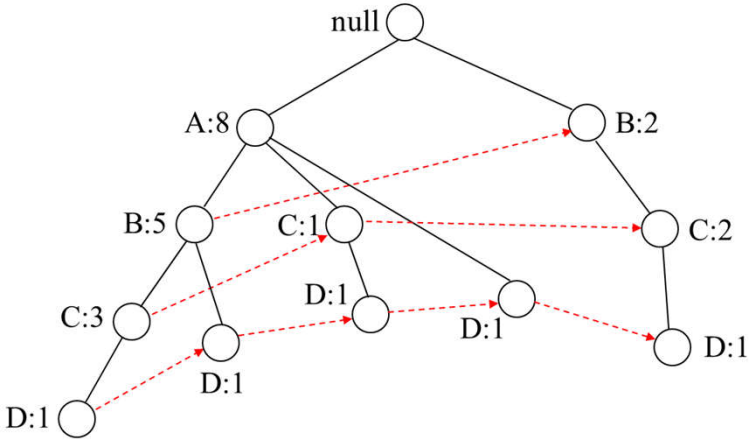
Frequent Itemset Generation in FP-Growth Algorithm

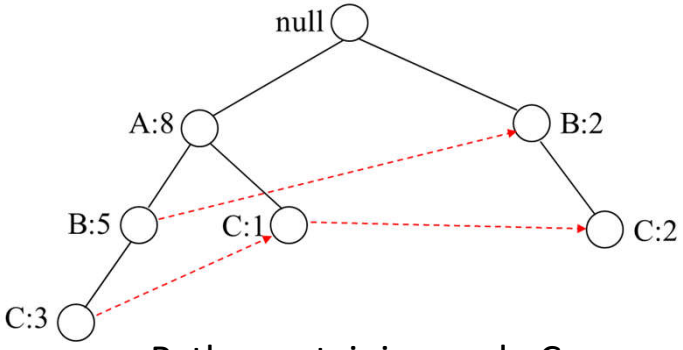


Done!

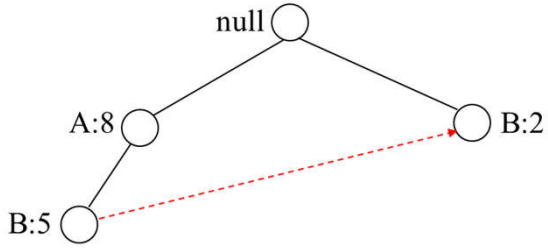
Paths containing node E



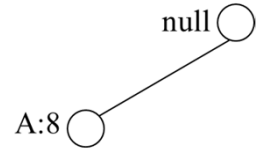
Paths containing node D



Paths containing node C



Paths containing node B



Paths containing node A

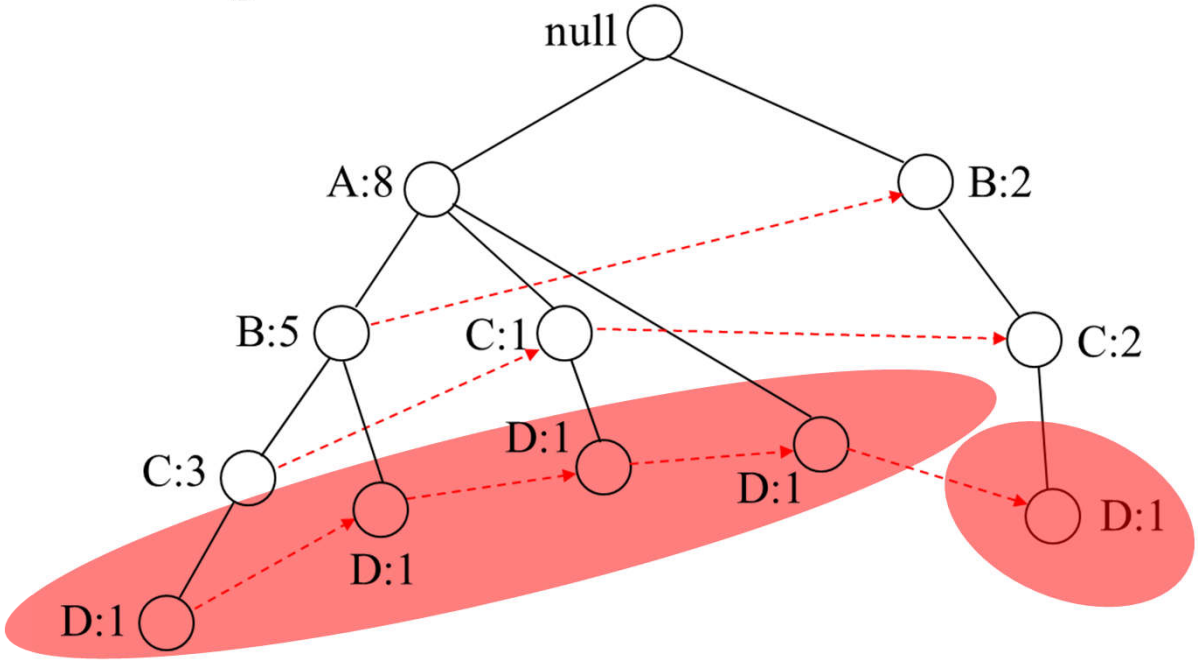
Frequent Itemset Generation in FP-Growth Algorithm

Suffix	Support Count
E	{E}, {D,E}, {A,D,E}, {C,E}, {A,E}
D	
C	
B	
A	

Frequent Itemset Generation in FP-Growth Algorithm

- Gather all the paths containing node D.
- Find support count for D by adding the support counts associate with node D from the prefix paths.

Prefix paths ending in D:

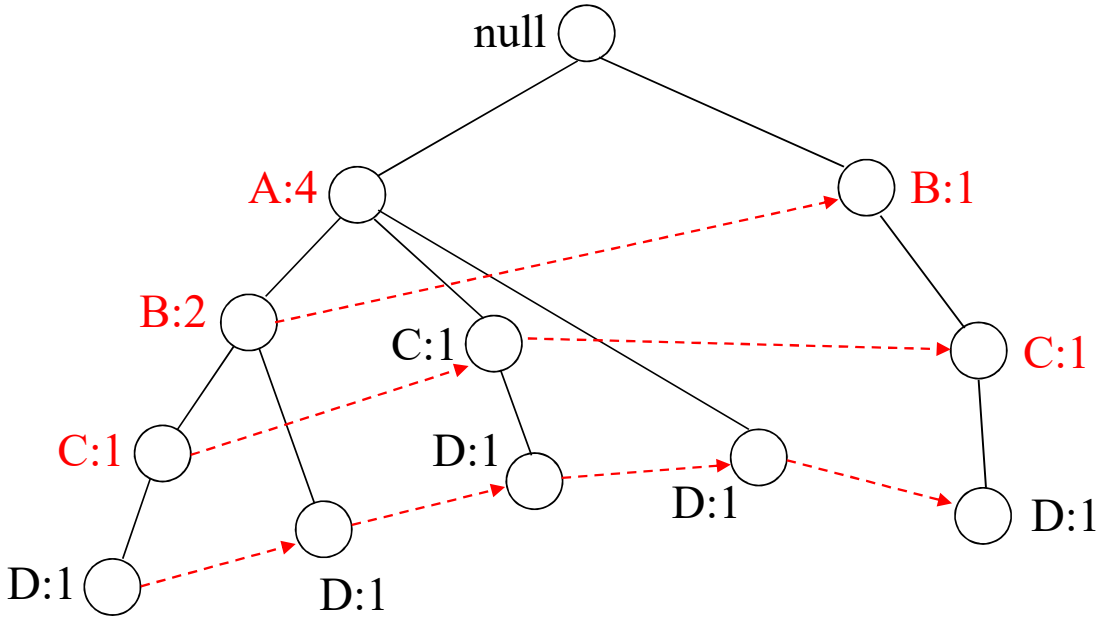


Support count of D = 4

Frequent Itemset Generation in FP-Growth Algorithm

- Itemset {D} is frequent, then convert the prefix paths ending in D into a **conditional FP-tree** in order to solve the subproblems of finding frequent itemsets ending in CD, BD, and AD.

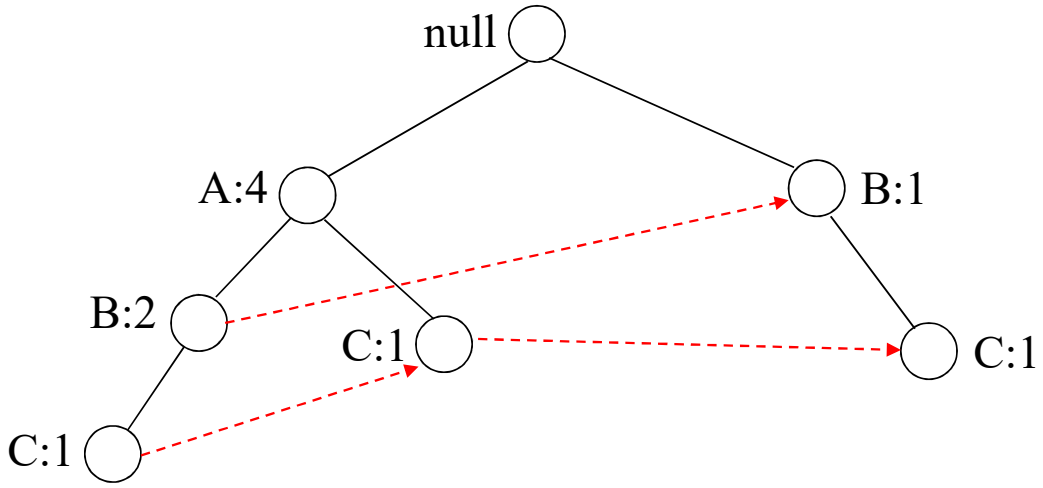
    1) Update the support count along the prefix paths because some of the counts include transactions that do not contain item D.



Frequent Itemset Generation in FP-Growth Algorithm

2) Remove the nodes D from the prefix paths and calculate support count of each item.

3) Remove nodes of items whose support counts are less than minimum support count.



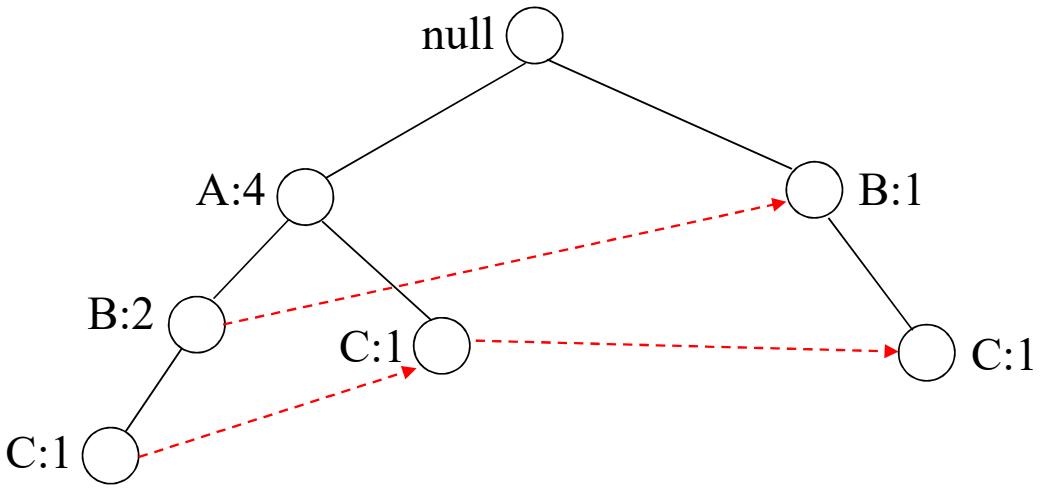
Conditional FP-tree for D

Item	Support Count
A	4
B	3
C	3

Frequent Itemset Generation in FP-Growth Algorithm

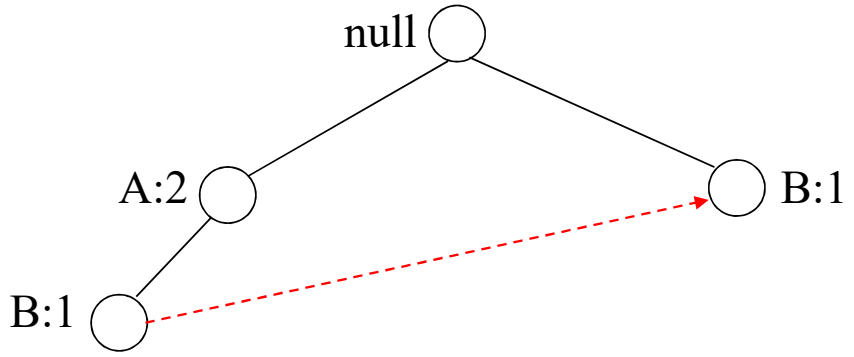
- Solve the subproblems of finding frequent itemsets ending in CD, BD, and AD using the conditional FP-tree for D.

Prefix paths ending in CD



- Support count of CD = 3
- {C,D} is a frequent itemset.
- Construct the conditional FP-tree for CD

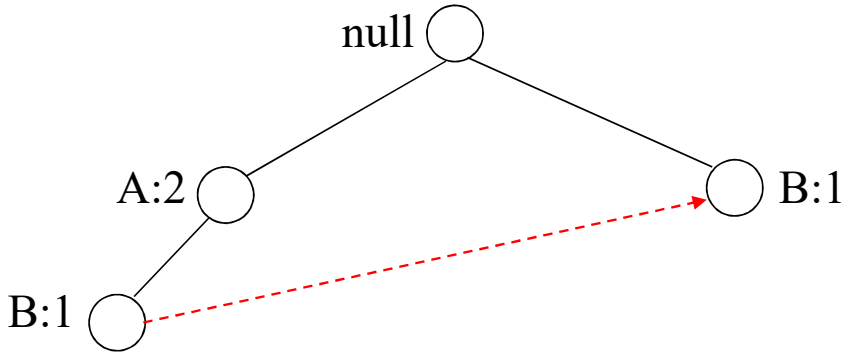
Conditional FP-tree for CD



Frequent Itemset Generation in FP-Growth Algorithm

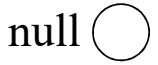
- Solve the sub-subproblems of finding frequent itemsets ending in BCD and ACD using the conditional FP-tree for CD.

Prefix paths ending in BCD



- Support count of BCD = 2
- {B,C,D} is a frequent itemset.
- Construct the conditional FP-tree for BCD

Conditional FP-tree for BCD

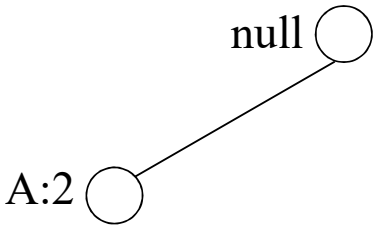


Since the conditional FP-tree contains only null node, then the algorithm proceed to the next sub-subproblem ending in ACD.

Frequent Itemset Generation in FP-Growth Algorithm

- Once we are done with BCD, the algorithm moves on to the next sub-subproblem, which is to generate frequent itemsets ending in ACD.

Prefix paths ending in ACD



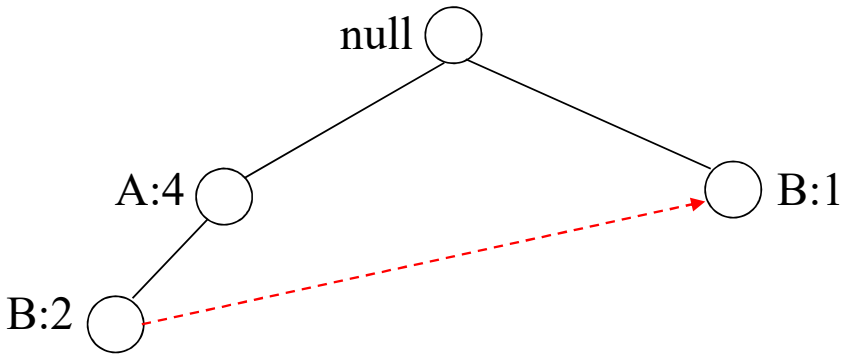
- Support count of ACD = 2
- {A,C,D} is a frequent itemset.

Done with finding frequent itemsets ending in CD!

Frequent Itemset Generation in FP-Growth Algorithm

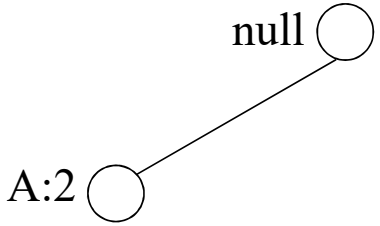
- Once we are done with CD, the algorithm moves on to the next subproblem, which is to generate frequent itemsets ending in BD.

Prefix paths ending in BD



- Support count of BD = 3
- {B,D} is a frequent itemset.
- Construct the conditional FP-tree for BD

Conditional FP-tree for BD

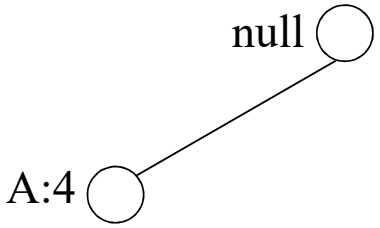


Since the conditional FP-tree contains only one item, A, whose support count is equal to minimum support count, then the algorithm extracts the frequent itemset {A,B,D}.

Frequent Itemset Generation in FP-Growth Algorithm

- Once we are done with BD, the algorithm moves on to the next subproblem, which is to generate frequent itemsets ending in AD.

Prefix paths ending in AD



- Support count of AD = 4
- {A,D} is a frequent itemset.

Done with finding frequent itemsets ending in D!

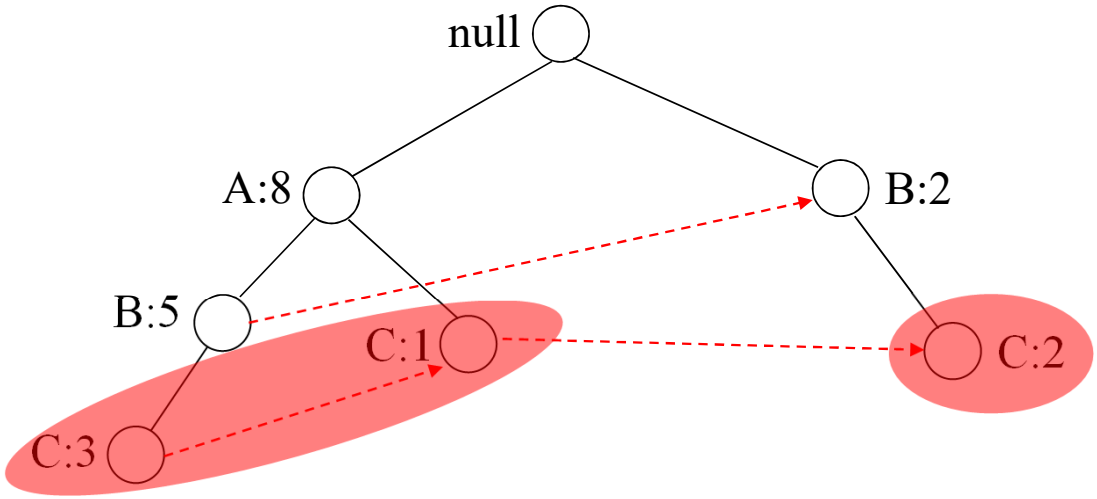
Frequent Itemset Generation in FP-Growth Algorithm

Suffix	Support Count
E	{E}, {D,E}, {A,D,E}, {C,E}, {A,E}
D	{C}, {C,D}, {B,C,D}, {A,C,D}, {B,D}, {A,B,D}, {A,D}
C	
B	
A	

Frequent Itemset Generation in FP-Growth Algorithm

- Gather all the paths containing node C.
- Find support count for C by adding the support counts associate with node C from the prefix paths.

Prefix paths ending in C:

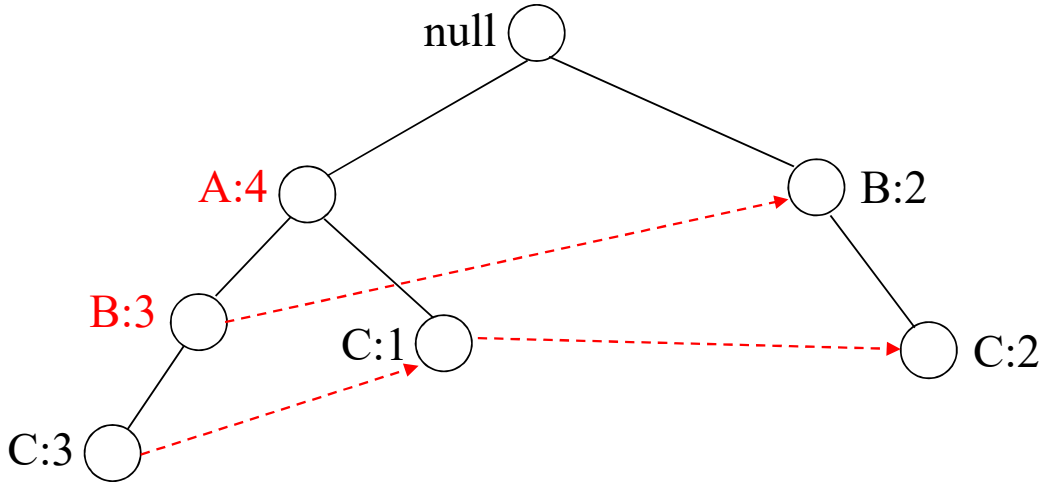


Support count of C = 6

Frequent Itemset Generation in FP-Growth Algorithm

- Itemset {C} is frequent, then convert the prefix paths ending in C into a **conditional FP-tree** in order to solve the subproblems of finding frequent itemsets ending in BC and AC.

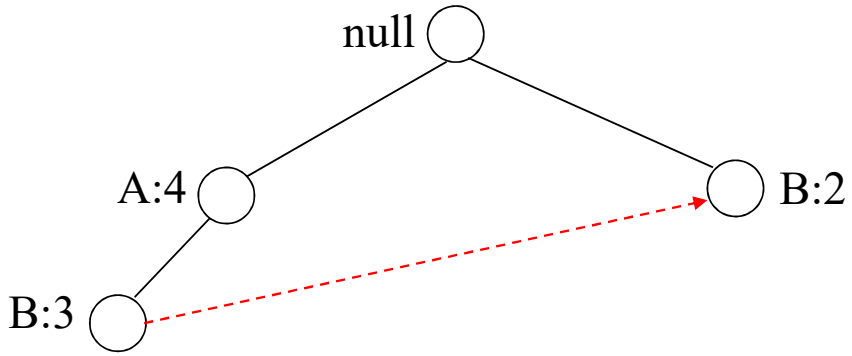
    1) Update the support count along the prefix paths because some of the counts include transactions that do not contain item C.



Frequent Itemset Generation in FP-Growth Algorithm

2) Remove the nodes C from the prefix paths and calculate support count of each item.

3) Remove nodes of items whose support counts are less than minimum support count.



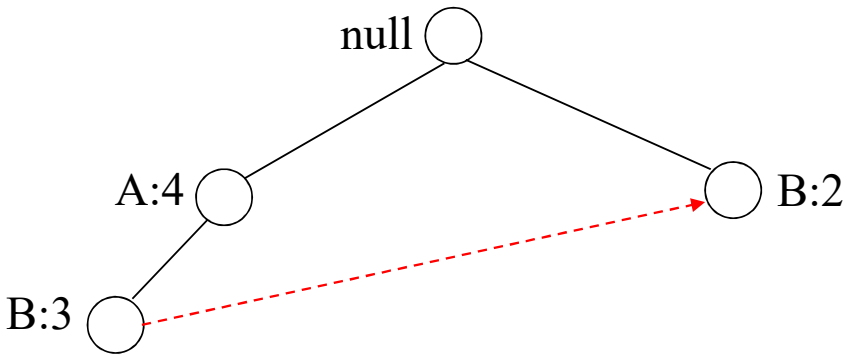
Conditional FP-tree for C

Item	Support Count
A	4
B	5

Frequent Itemset Generation in FP-Growth Algorithm

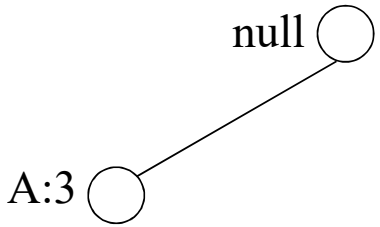
- Solve the subproblems of finding frequent itemsets ending in BC and AC using the conditional FP-tree for C.

Prefix paths ending in BC



- Support count of BC = 5
- {B,C} is a frequent itemset.
- Construct the conditional FP-tree for BC

Conditional FP-tree for BC

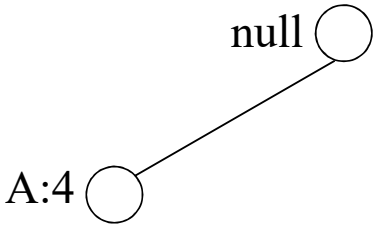


Since the conditional FP-tree contains only one item, A, whose support count is greater than minimum support count, then the algorithm extracts the frequent itemset {A,B,C}.

Frequent Itemset Generation in FP-Growth Algorithm

- Once we are done with BC, the algorithm moves on to the next subproblem, which is to generate frequent itemsets ending in AC.

Prefix paths ending in AC



- Support count of AC = 4
- {A,C} is a frequent itemset.

Done with finding frequent itemsets ending in C!

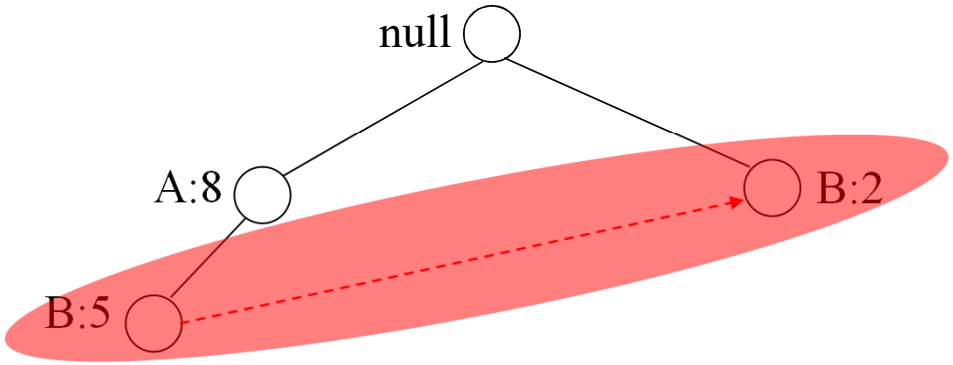
Frequent Itemset Generation in FP-Growth Algorithm

Suffix	Support Count
E	{E}, {D,E}, {A,D,E}, {C,E}, {A,E}
D	{C}, {C,D}, {B,C,D}, {A,C,D}, {B,D}, {A,B,D}, {A,D}
C	{C}, {B,C}, {A,B,C}, {A,C}
B	
A	

Frequent Itemset Generation in FP-Growth Algorithm

- Gather all the paths containing node B.
- Find support count for B by adding the support counts associate with node B from the prefix paths.

Prefix paths ending in B:

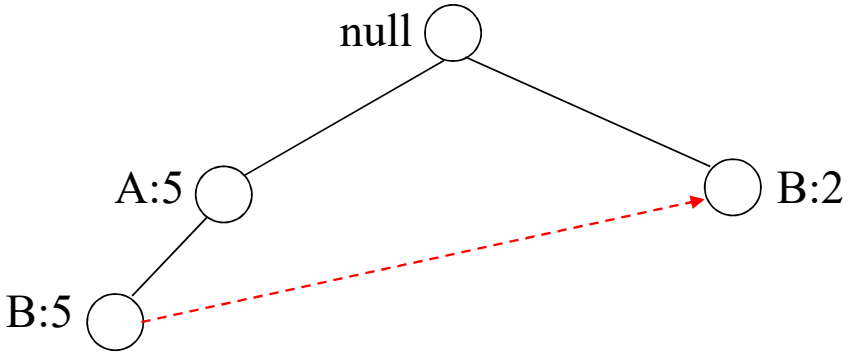


Support count of B = 7

Frequent Itemset Generation in FP-Growth Algorithm

- Itemset {B} is frequent, then convert the prefix paths ending in B into a **conditional FP-tree** in order to solve the subproblems of finding frequent itemsets ending in AB.

    1) Update the support count along the prefix paths because some of the counts include transactions that do not contain item B.

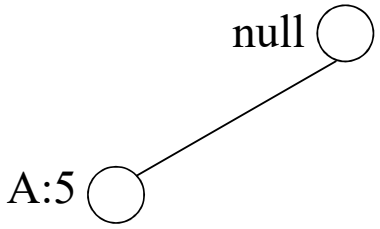


Frequent Itemset Generation in FP-Growth Algorithm

2) Remove the nodes B from the prefix paths and calculate support count of each item.

3) Remove nodes of items whose support counts are less than minimum support count.

Conditional FP-tree for B



Item	Support Count
A	5

Since the conditional FP-tree contains only one item, A, whose support count is greater than minimum support count, then the algorithm extracts the frequent itemset {A,B}.

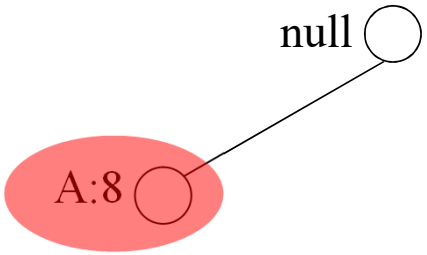
Frequent Itemset Generation in FP-Growth Algorithm

Suffix	Support Count
E	{E}, {D,E}, {A,D,E}, {C,E}, {A,E}
D	{C}, {C,D}, {B,C,D}, {A,C,D}, {B,D}, {A,B,D}, {A,D}
C	{C}, {B,C}, {A,B,C}, {A,C}
B	{B}, {A,B}
A	

Frequent Itemset Generation in FP-Growth Algorithm

- Gather all the paths containing node A.
- Find support count for A by adding the support counts associate with node A from the prefix paths.

Prefix paths ending in A:



Support count of A = 8

Since the prefix paths ending in A, whose support count is greater than minimum support count, then the algorithm extracts the frequent itemset {A}.

Frequent Itemset Generation in FP-Growth Algorithm

Suffix	Support Count
E	{E}, {D,E}, {A,D,E}, {C,E}, {A,E}
D	{D}, {C,D}, {B,C,D}, {A,C,D}, {B,D}, {A,B,D}, {A,D}
C	{C}, {B,C}, {A,B,C}, {A,C}
B	{B}, {A,B}
A	{A}