HW1

Kaichen Zhang

2017-09-26

4 THE LIFT MEASURE

assuming antecedent are consequent are independent

$$
\begin{aligned}
lift(X \Rightarrow Y) &= \frac{c(X \Rightarrow Y)}{s(Y)} \\
&= \frac{\sigma(X \cup Y)}{\sigma(X)S(Y)} \\
&= \frac{\sigma(X)\sigma(Y)}{\sigma(X)\sigma(Y)} \\
&= 1
\end{aligned}
$$

5a) Candidate 1-item set

Item	Count
{ bananas }	3
{ carrots }	6
{ figs }	4
{ apples }	5
{ donots }	5
{ eggs }	4

Min support count = 0.3 * 10 (item sets given in the problem) = 3, so frequency sets are the same

Frequent 1-item set

Item	Count
bananas	3
carrots	6
figs	4
apples	5
donots	5
eggs	4

Candidate 2-item set

Items	Count
{ bananas, carrots }	3
{ bananas, figs }	2
{ bananas, apples }	1
{ bananas, donuts }	1
{ bananas, eggs }	0
{ carrots, figs }	3
{ carrots, apples }	2
{ carrots, donuts }	2
{ carrots, eggs }	3
{ figs, apples }	1
{ figs, donuts }	3
{ figs, eggs }	1

Frequent 2-item set

Items	Count
{ bananas, carrots }	3
{ carrots, figs }	3
{ carrots, eggs }	3
{ figs, donuts }	3

No candidate 3-item set or frequent 3-item set

5b) Maximal frequent set:

Item	Count
bananas	3
carrots	6
figs	4
apples	5
donots	5
eggs	4

Items	Count
{ bananas, carrots }	3
{ carrots, figs }	3
{ carrots, eggs }	3
{ figs, donuts }	3

Maximal frequent set: the immediate super set can contain all its elements:

1) Apples
2) { bananas, carrots }
3) { carrots, figs }
4) { carrots, eggs }
5) { figs, donuts }

5c) c({bananas}=>{carrots}} = 3/3 = 1

6a) Construct FP-tree

TID	Items
1	{apples}
2	{apples, carrots}
3	{apples, carrots, donuts}
4	{apples, bananas, eggs}
5	{apples, bananas, carrots, donuts}
6	{bananas, donuts, eggs}
7	{bananas, carrots, figs}
8	{apples, bananas, carrots}
9	{apples, bananas, donuts, eggs}
10	{apples, carrots, eggs}

Item	Count
apples	8
bananas	6
carrots	6
donuts	4
eggs	4
figs	1

Min support count = 0.2 * 10 (item sets given in the problem) = 2, so figs are eliminated

Filtered and sorted

TID	Items
1	{apples}
2	{apples, carrots}
3	{apples, carrots, donuts}
4	{apples, bananas, eggs}
5	{apples, bananas, carrots, donuts}
6	{bananas, donuts, eggs}
7	{bananas, carrots }
8	{apples, bananas, carrots}
9	{apples, bananas, donuts, eggs}
10	{apples, carrots, eggs}

a = apples, b = bananas, c = carrots, d = donuts, e = eggs

TID	Items
1	{a}
2	{a, c}
3	{a, c, d}
4	{a, b, e}
5	{a, b, c, d}
6	{b, d, e}
7	{b, c }
8	{a, b, c}
9	{a, b, d, e}
10	{a, c, e}

Transactions*
- a

Header table
a: 8
b: 6
c: 6
d: 4
e: 4

T1

∅

a:1
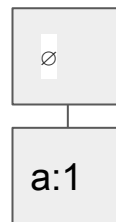
Transactions*
- a
- a, c

Header table
a: 8
b: 6
c: 6
d: 4
e: 4

∅

a:2

c:1
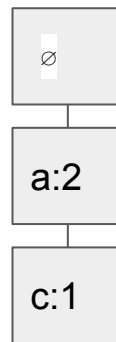
Transactions*
- a
- a, c
- a, c, d

Header table
a: 8
b: 6
c: 6
d: 4
e: 4

∅

a:3

c:2

d:1
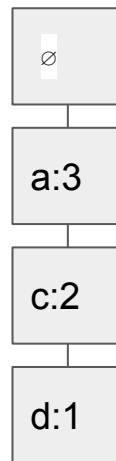
Transactions*
- a
- a, c
- a, c, d
- a, b, e

Header table
a: 8
b: 6
c: 6
d: 4
e: 4
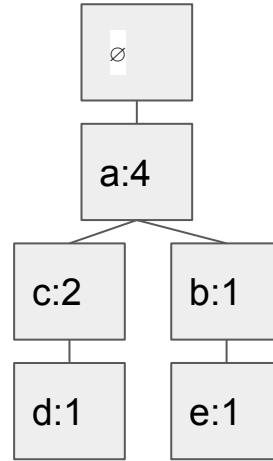
Transactions*
- a
- a, c
- a, c, d
- a, b, e
- a, b, c, d

Header table
a: 8
b: 6
c: 6
d: 4
e: 4
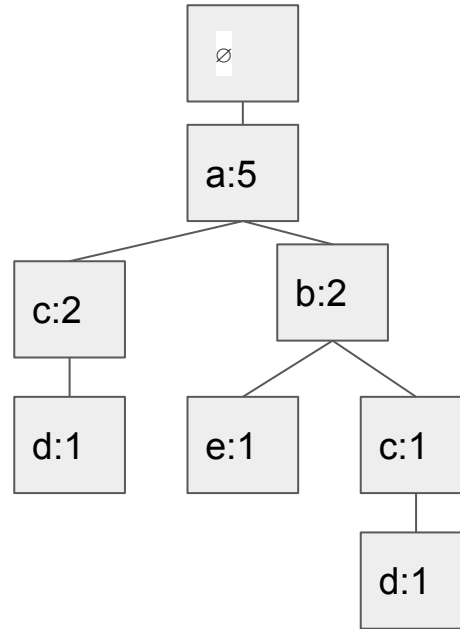
Transactions*
- a
- a, c
- a, c, d
- a, b, e
- a, b, c, d
- b, d, e

Header table
a: 8
b: 6
c: 6
d: 4
e: 4
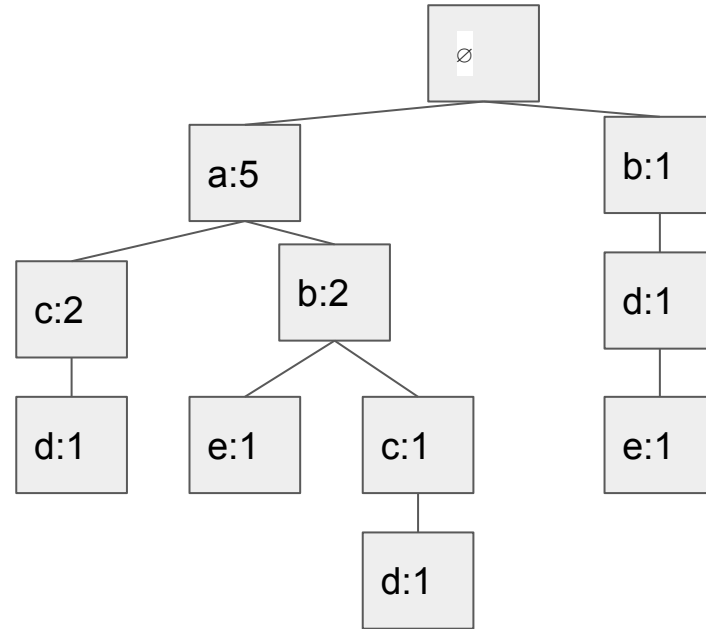
Transactions*
- a
- a, c
- a, c, d
- a, b, e
- a, b, c, d
- b, d, e
- b, c

Header table
a: 8
b: 6
c: 6
d: 4
e: 4

```
                          ∅
              /                      \
           a:5                        b:2
          /     \                    /     \
        c:2     b:2               d:1       c:1
         |     /    \             |
        d:1  e:1    c:1          e:1
                     |
                    d:1
```
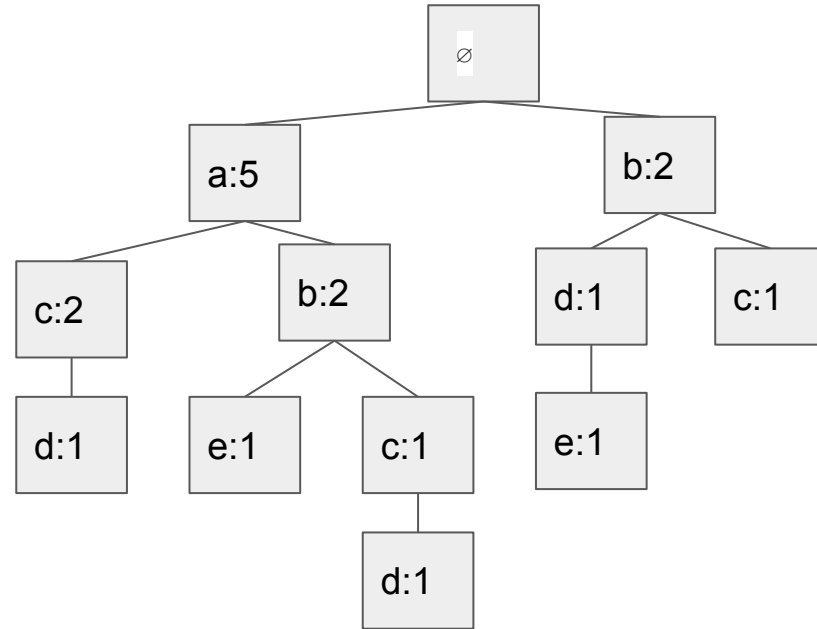
Transactions*
- a
- a, c
- a, c, d
- a, b, e
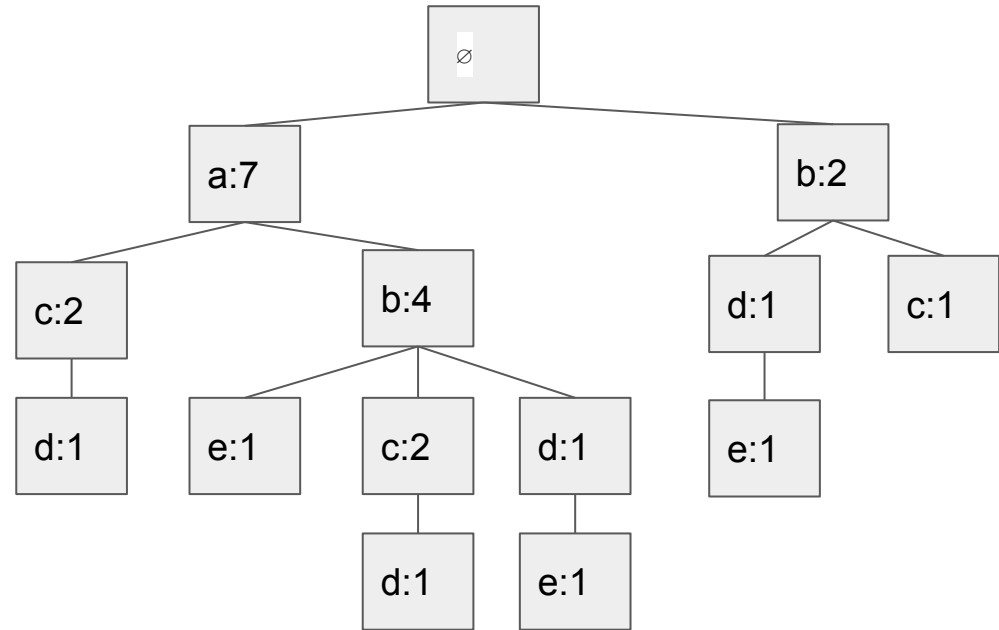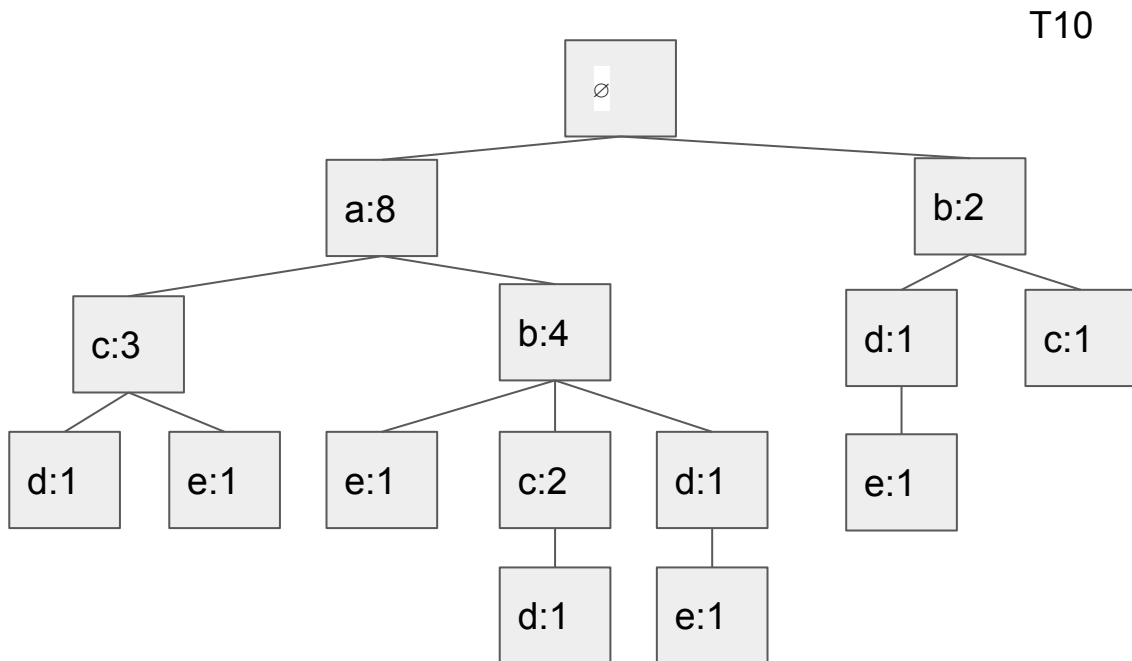- a, b, c, d
- b, d, e
- b, c
- b, b, c

Header table	
a: 8	
b: 6	
c: 6	
d: 4	
e: 4	
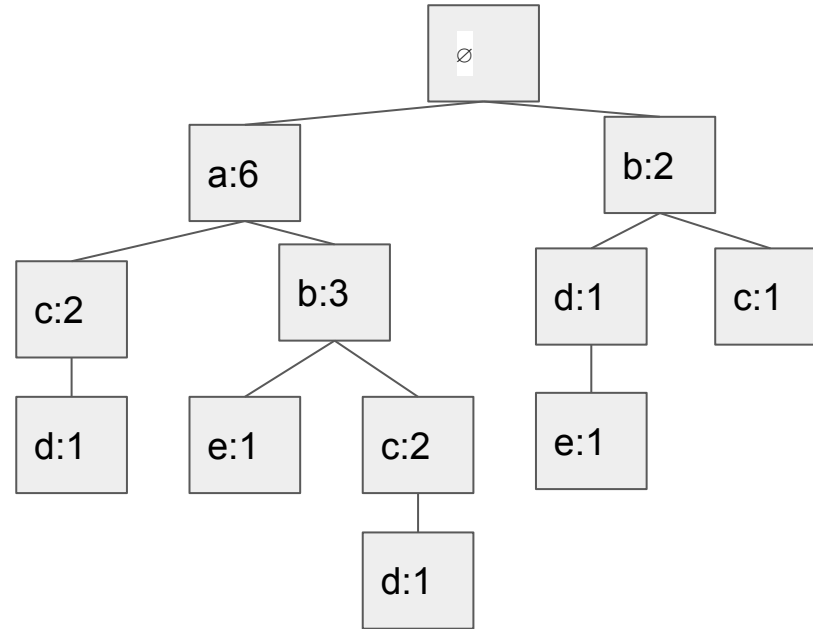
Transactions*
- a
- a, c
- a, c, d
- a, b, e
- a, b, c, d
- b, d, e
- b, c
- a, b, c
- a, b, d, e

Header table	
a: 8	
b: 6	
c: 6	
d: 4	
e: 4	

∅

a:7   b:2

c:2   b:4   d:1   c:1

d:1   e:1   c:2   d:1   e:1

d:1   e:1

Transactions*
- a
- a, c
- a, c, d
- a, b, e
- a, b, c, d
- b, d, e
- b, c
- a, b, c
- a, b, d, e
- a, c, e

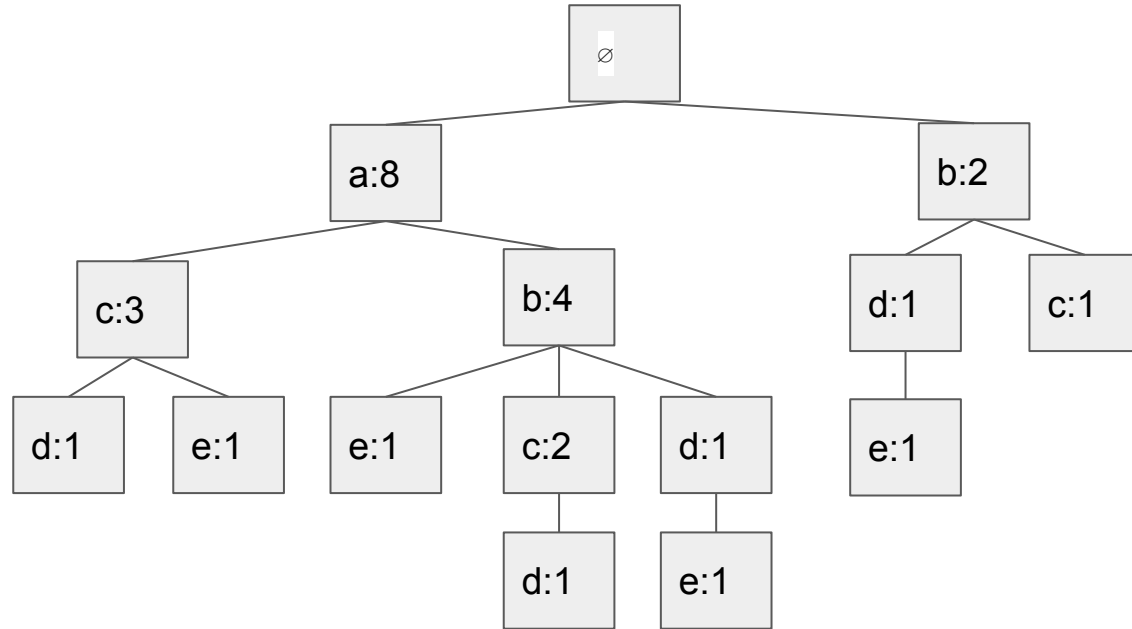| Header table	
a: 8	
b: 6	
c: 6	
d: 4	
e: 4	

T10

6b) Finding frequent itemsets containing donuts

Transactions*
- a
- a, c
- a, c, d
- a, b, e
- a, b, c, d
- b, d, e
- b, c
- a, b, c
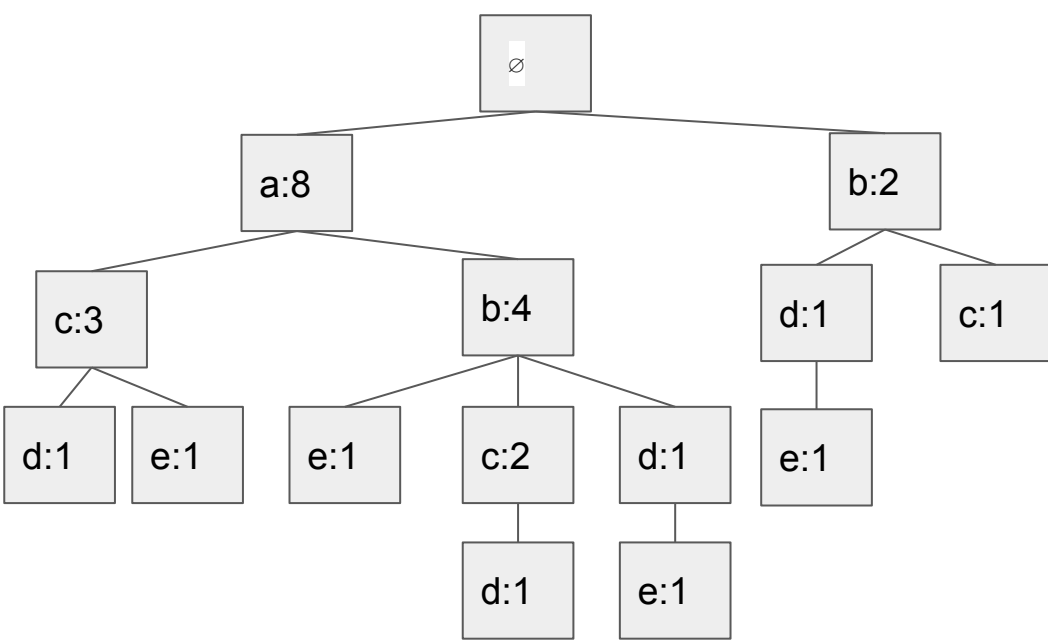- a, b, d, e
- a, c, e

S_{min} = 2

Header table
a: 8
b: 6
c: 6
d: 4
e: 4
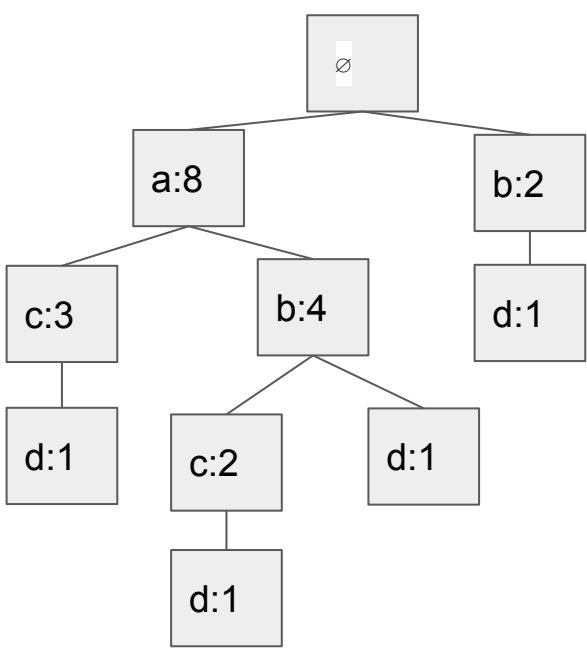
Ending in d

Header table

a: 8	
b: 6	
c: 6	
d: 4	
e: 4	

S_{min} = 2

Conditional d
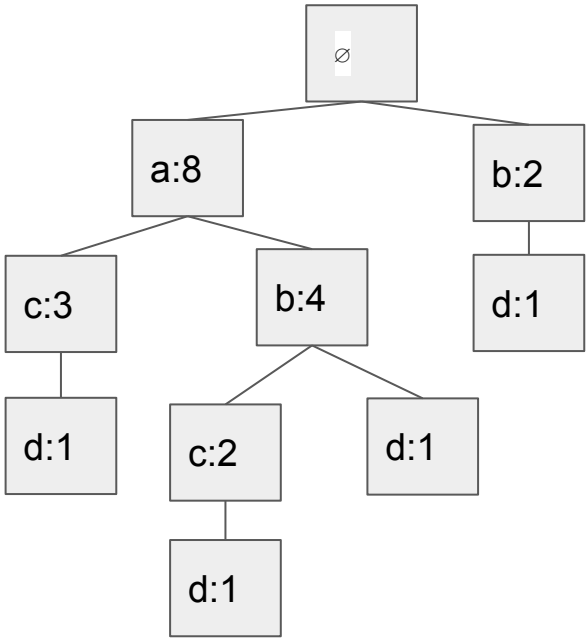
Header table

a: 8	
b: 6	
c: 6	
d: 4	
e: 4	
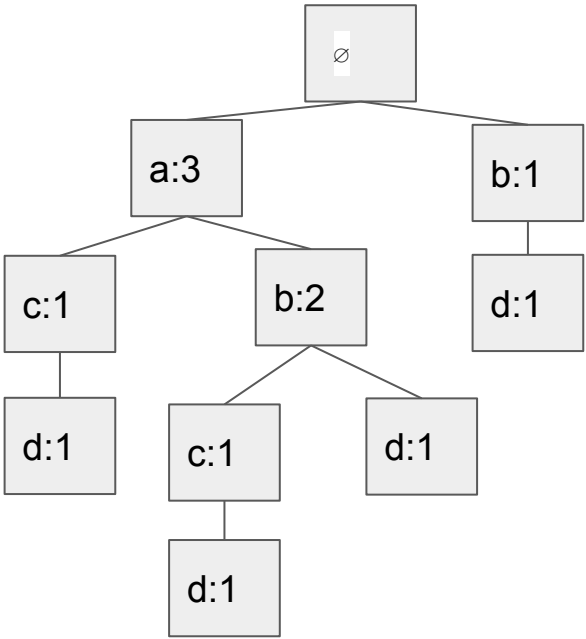
Because we only care about the branching ending in d

S_{min} = 2

Conditional d

Header table
a: 8
b: 6
c: 6
d: 4
e: 4

∅

a:3

b:1

c:1

b:2

c:1
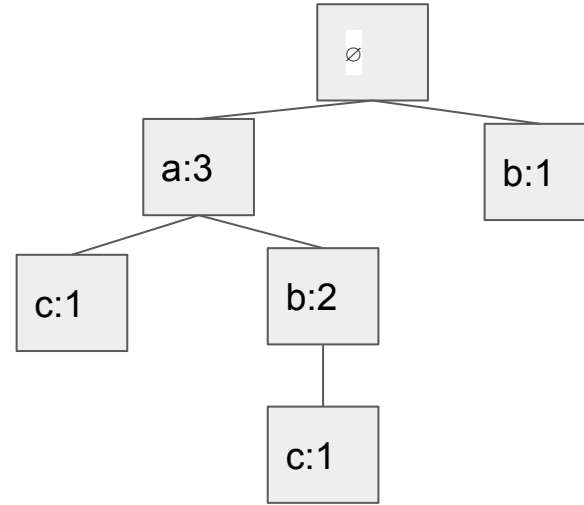
Frequent itemsets:
Ending in c: remove the b on the right branch,
remove c, so cd, then a = 1, b = 1, no more
Ending in b: remove c, so bd, a = 1 no more
Ending in a: remove c, b, so ad
Result: cd, bd, ad

$S_{min} = 2$

7b) 15.6s 7c) 9.7s
7d) FPGrowth found 2 rules

1. [i11=1, i218=1, i148=1]: 50098 ==> [i6=1]: 49866   <conf:(1)> lift:(1.64) lev:(0.02) conv:(84.4)
2. [i11=1, i148=1]: 55759 ==> [i6=1]: 55230   <conf:(0.99)> lift:(1.63) lev:(0.02) conv:(41.3)

7 e) Much shorter
21:44:25: Started weka.associations.FPGrowth
21:44:26: Finished weka.associations.FPGrowth

21:44:27: Started weka.associations.FPGrowth
21:44:28: Finished weka.associations.FPGrowth

21:44:29: Started weka.associations.FPGrowth
21:44:29: Finished weka.associations.FPGrowth

21:44:29: Started weka.associations.FPGrowth
21:44:30: Finished weka.associations.FPGrowth

21:44:31: Started weka.associations.FPGrowth
21:44:31: Finished weka.associations.FPGrowth