

目的: 找到 frequent itemset

①

如果一个 itemset 是 frequent, 那么其 subset 必为 frequent

c, d, e 为 frequent 则 c, d, e 组成的子节点也为 frequent

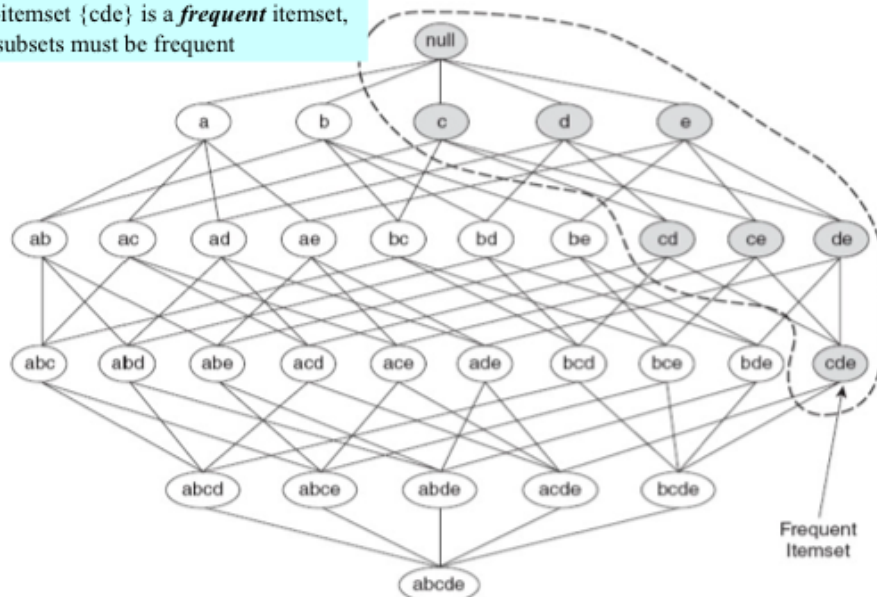
②

如果一个 itemset 是 infrequent, 那么包含其的必为 infrequent

ab 为 infrequent 则 ab 引出的所有子节点都为 infrequent

Illustration 1: *Apriori* principle

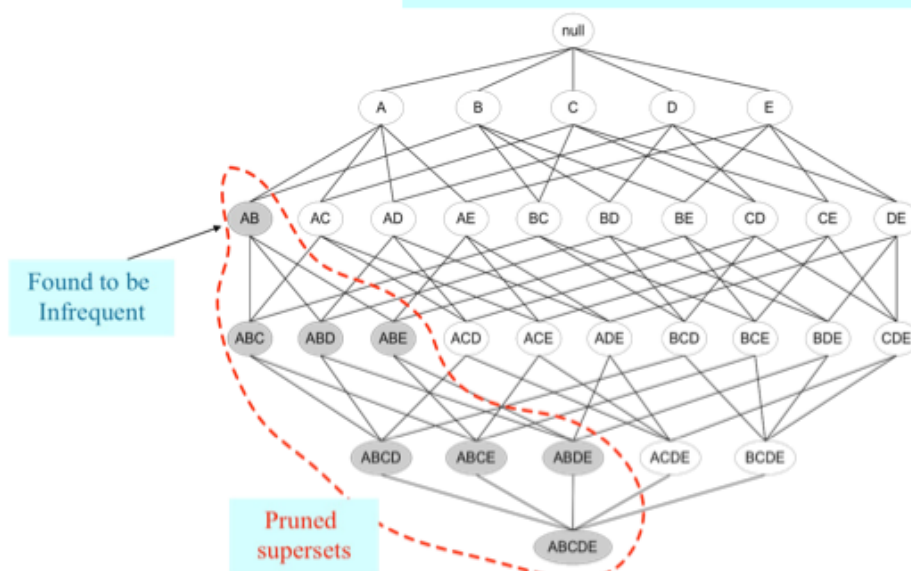
If a 3-itemset {cde} is a **frequent** itemset, all its subsets must be frequent



38

Illustration 2: *Apriori* : Support σ -based Pruning

If {AB} is an infrequent itemset, then all its supersets are infrequent. This strategy can be used to reduce the number of candidate itemsets.

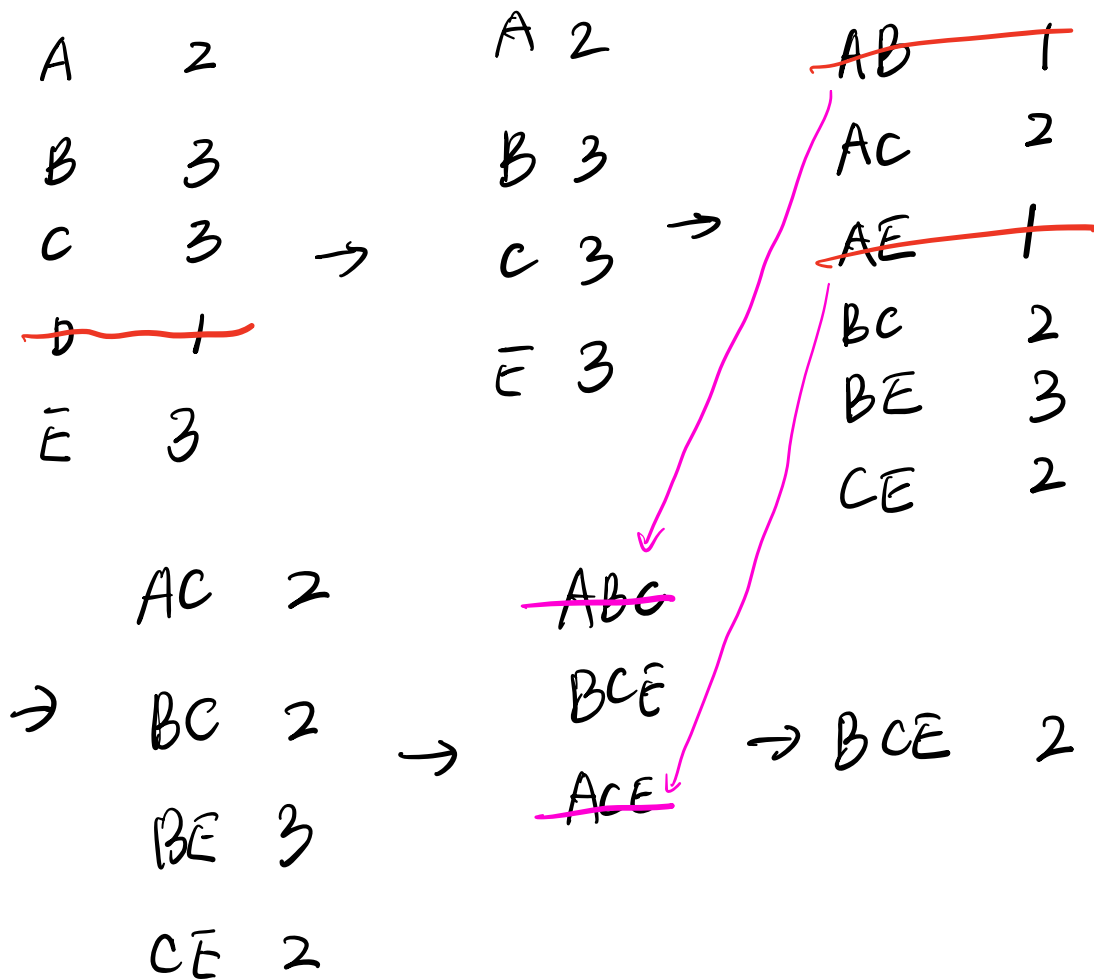


39

Database TDB

$minsup = 2$

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E



Example 1(a)

Consider the following transaction dataset, where each letter represents an item.

TID	Items Bought
1	N, E, W
2	N, O, W
3	W, E
4	O, N, E
5	O, W, N
6	G, O

- (a) Assume that the minimum support is **30%**. Use the *Apriori* algorithm to find all the frequent itemsets in the dataset.
- (b) Give two association rules which have a minimum confidence of 60%.

(a) $\text{minsup} = 6 \times 30\% = 1.8 \rightarrow 2$

N	4		N	4		NE	2		NE	2
E	3		E	3		NW	3		NW	3
W	4	→	W	4	→	NO	3	→	NO	3
O	4		O	4		EW	2		EW	2
G	1					EO	1		WO	2
						WO	2			

→ ~~NEW~~ 1
→ ~~NEO~~ 1
→ ~~NWO~~ 2
→ ~~EWO~~ 2
→ NWO 2

(b)

$$N W \rightarrow O \quad \text{conf} = \frac{2}{3}$$

$$N D \rightarrow W \quad \text{conf} = \frac{2}{3}$$