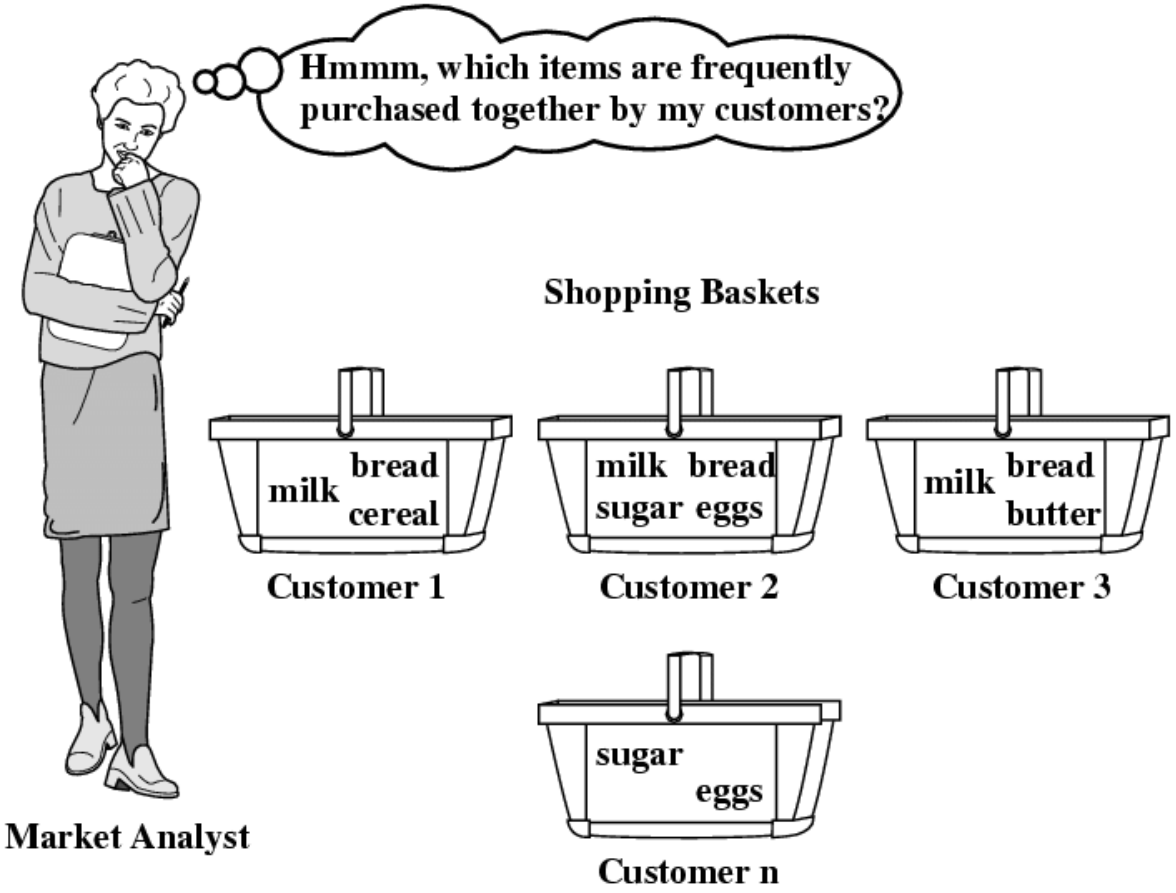
關聯規則探勘(購物籃分析)



TID	Transaction
T_1	{Milk, Bread, Cereal}
T_2	{Milk, Bread, Sugar, Eggs}
T_3	{Milk, Bread, Butter}
T_4	{Sugar, Eggs}

Milk $\Rightarrow$ Bread [75%, 100%]

Motivation

- **Purpose of Association Rule Mining**
  - Finding associations, correlations, or causal structures among item sets in databases
- **Application**
  - Market Basket Analysis
- **Example**
  - {啤酒} → {尿布}

Related Definitions about Association Rule Mining

- **Item**
  - e.g., A, B, C, D, E, F
- **Itemset**
  - e.g., {ABC}
- **Contain**
  - e.g., {AB} is contained in {ABC}
- **Superset**
  - e.g., {ABC} is a superset of {AB}
- **Subset**
  - e.g., {AB} is a subset of {ABC}
  - e.g., The complete set of the subsets of {ABC} is {{A}, {B}, {C}, {AB}, {AC}, {BC}, {ABC}}

Transaction Database	
TID	Transaction
1	ABC
2	AC
3	AD
4	BEF

Related Definitions about Association Rule Mining

- **Tidset of an itemset**
  - *Tidset*(**{**AC**}**) = **{**1, 2**}**
- **Support count of an itemset**
  - e.g., *SC*({AC}) = |*Tidset*(**{**AC**})**| = 2
- **Support of an itemset**
  - e.g., *SP*({AC}) = 2/4 = 50%
- **Frequent itemset**
  - An itemset *X* is called *frequent itemset* iff *SP*(*X*) is no less than a user-specified minimum support threshold δ ($0\% \leq \delta \leq 100\%$).
  - e.g., if δ = 50%, {AC} is a frequent itemset but {ABC} is not

Transaction Database	
TID	Transaction
1	ABC
2	AC
3	AD
4	BEF

Related Definitions about Association Rule Mining

- **Rule**
  - R: $X \rightarrow Y$
    - X and Y are frequent itemset
    - $X \cap Y = \varnothing$
    - X is called *antecedent* (*Left hand side*; *LHS*)
    - Y is called *consequent* (*Right hand side*; *RHS*)
  - e.g., {A}→{C}

Transaction Database	
TID	Transaction
1	ABC
2	AC
3	AD
4	BEF

- **Support of a rule**
  - $SP(R) = SP(X \cup Y)$
  - e.g., $SP(\{A\} \rightarrow \{C\}) = SP(\{A\} \cup \{C\}) = 2/4$

Related Definitions about Association Rule Mining

- **Confidence of a rule**
  - $CF(R) = SP(X \cup Y) / SP(X)$
  - e.g., $CF(\{A\} \rightarrow \{C\}) =$
    $SP(\{A\} \cup \{C\}) / SP(\{A\}) = (2/4) / (3/4) = 2/3$
  - $CF(R) = SC(X \cup Y) / SC(X)$
  - e.g., $CF(\{A\} \rightarrow \{C\}) = SC(\{AC\}) / SC(\{A\}) = 2/3$

Transaction Database	
TID	Transaction
1	ABC
2	AC
3	AD
4	BEF

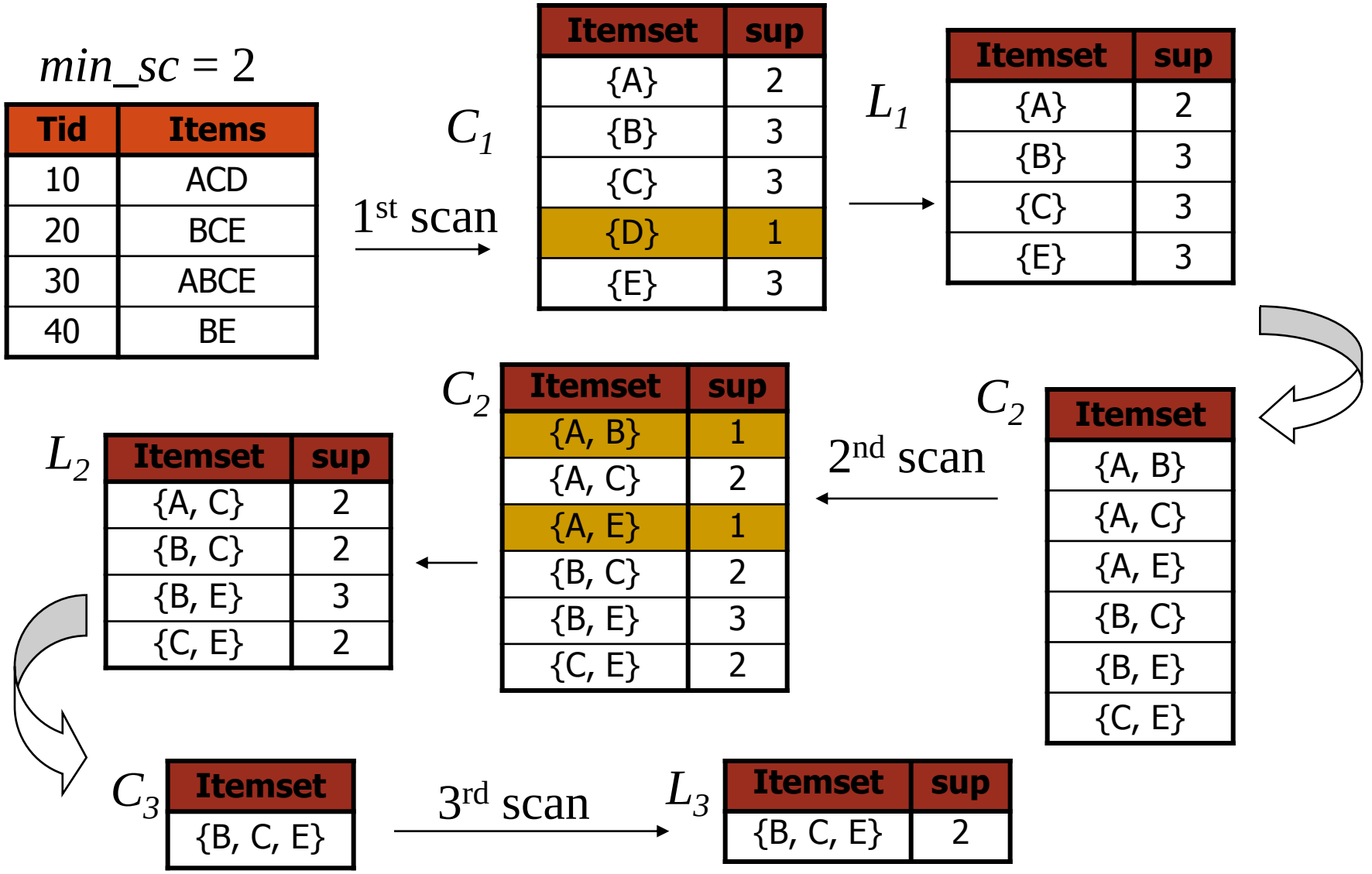
Related Definitions about Association Rule Mining

- **Association Rule**
  - A rule *R* is called *association rule* iff its support and confidence are no less than user-specified minimum support *min_sup* and minimum confidence *min_conf* thresholds, respectively.

Two Main Steps in Association Rule Mining

- Association rule mining consists of two main steps:
  - Frequent itemset mining
  - Rule generation

The Apriori Algorithm - An Example



Downward Closure Property

- **Downward Closure Property**
  - Any superset of an infrequent itemset are infrequent
  - Any subset of a frequent itemset are frequent

Transaction Database	
TID	Transaction
1	ABC
2	AC
3	AD
4	BEF

Rule Generation

- If {ABC} is frequent
  - A → BC sup count(ABC) /sup count(A)
  - sup(ABC) / sup(A)