



# ASSOCIATION RULE MINING

2143488 BIG DATA AND ARTIFICIAL INTELLIGENCE

DR. JING TANG

# ASSOCIATION RULE MINING (1)

---

- Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in transaction



# ASSOCIATION RULE MINING (2)

- At a supermarket:
  - find the products that the customers



Transaction ID	Items
1	Apple, Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple, Beer, Cereal, Eggs
4	Beer, Eggs



# WHAT DO THESE RULES USE FOR?

---

- If item A and B are bought together more frequently then several steps can be taken to increase the profit. For example:
  - A and B can be **placed together** so that when a customer buys one of the product he doesn't have to go far away to buy the other product.
  - People who buy one of the products can be targeted through **an advertisement campaign** to buy the other.
  - **Collective discounts** can be offered on these products if the customer buys both of them.
  - Both A and B can be **packaged together**.

# FP-GROWTH *(HAN AND YIN 2000)*

---

Steps:

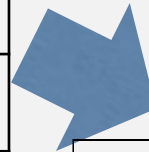
1. Build  (FP-tree)
2. Generate frequent item sets from FP-tree

# FP-GROWTH

## 1. BUILD FP-TREE (1)

- Compute support for each item (size =1)
- **Support** is an indication of how frequently the itemset appears in the dataset.

Transaction ID	Items
1	Apple, Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple, Beer, Cereal, Eggs
4	Beer, Eggs



Items	Transaction ID				Support
	1	2	3	4	
Apple	1	0	1	0	2/4 = 50%
Beer	0	1	1	1	3/4 = 75%
Cereal	1	1	1	0	3/4 = 75%
Diapers	1	0	0	0	1/4 = 25%
Egg	0	1	1	1	3/4 = 75%

# FP-GROWTH

## 1. BUILD FP-TREE (2)

- Sort the sequence of the items in the transaction by **support** (from large to small)

Items	Transaction ID				Support
	1	2	3	4	
Apple	1	0	1	0	2/4 = 50%
Beer	0	1	1	1	3/4 = 75%
Cereal	1	1	1	0	3/4 = 75%
Diapers	1	0	0	0	1/4 = 25%
Egg	0	1	1	1	3/4 = 75%

Transaction ID	Items
1	Apple, Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple, Beer, Cereal, Eggs
4	Beer, Eggs



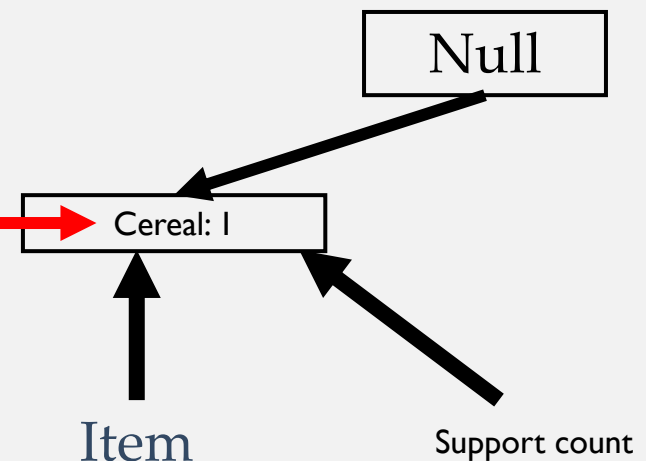
Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

# FP-GROWTH

## 1. BUILD FP-TREE (3)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs



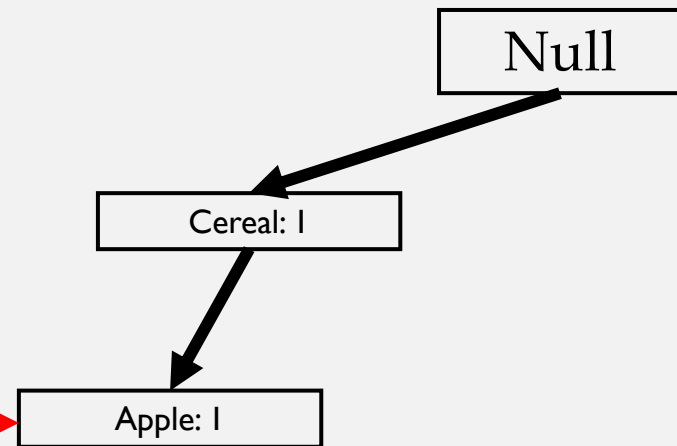


# FP-GROWTH

## 1. BUILD FP-TREE (4)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

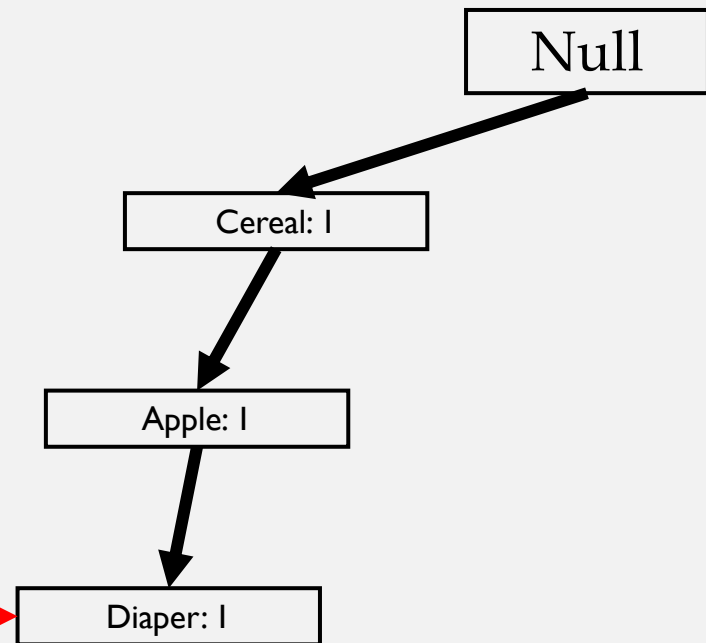


# FP-GROWTH

## 1. BUILD FP-TREE (5)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

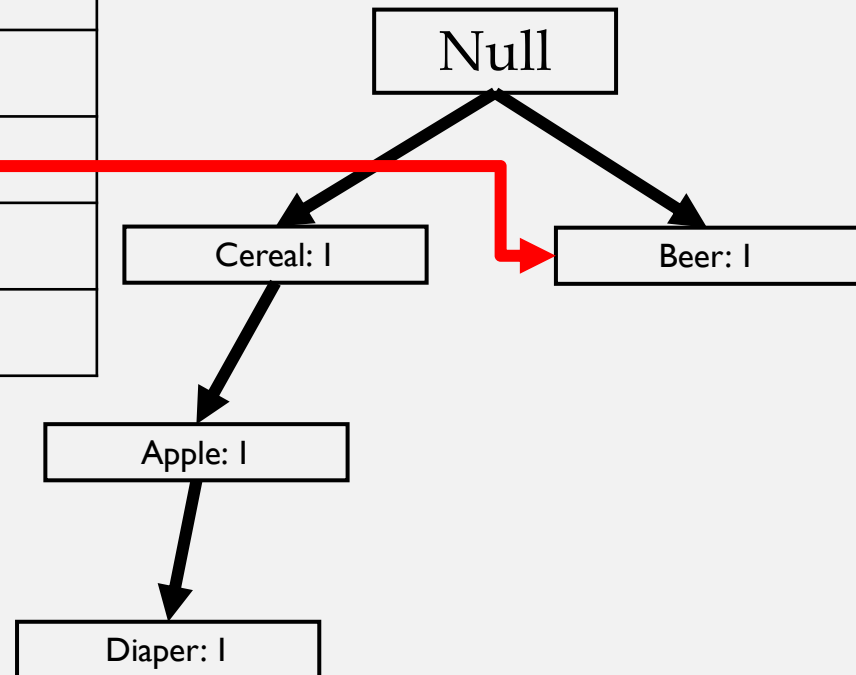


# FP-GROWTH

## 1. BUILD FP-TREE (6)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

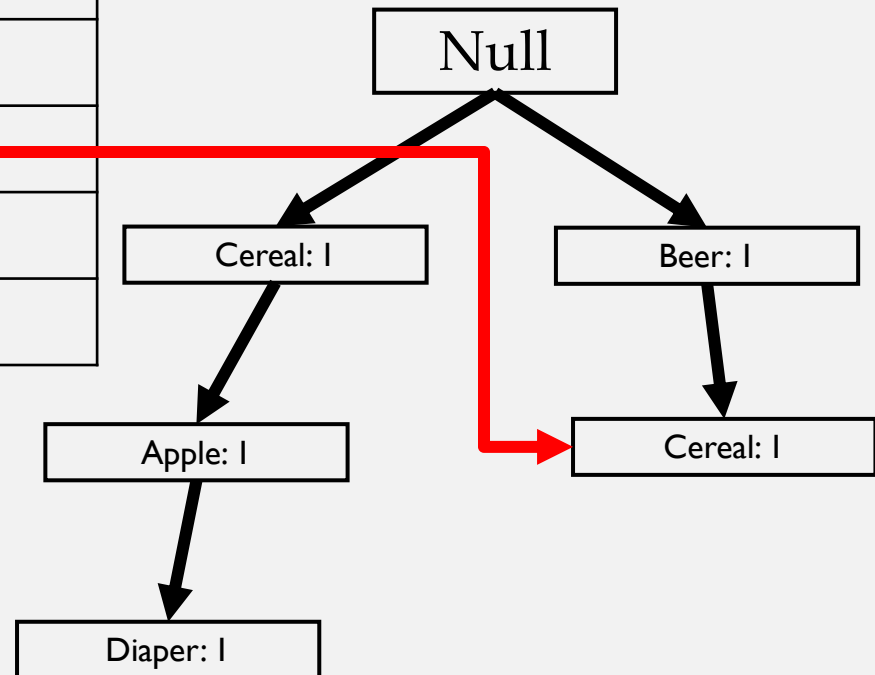


# FP-GROWTH

## 1. BUILD FP-TREE (7)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

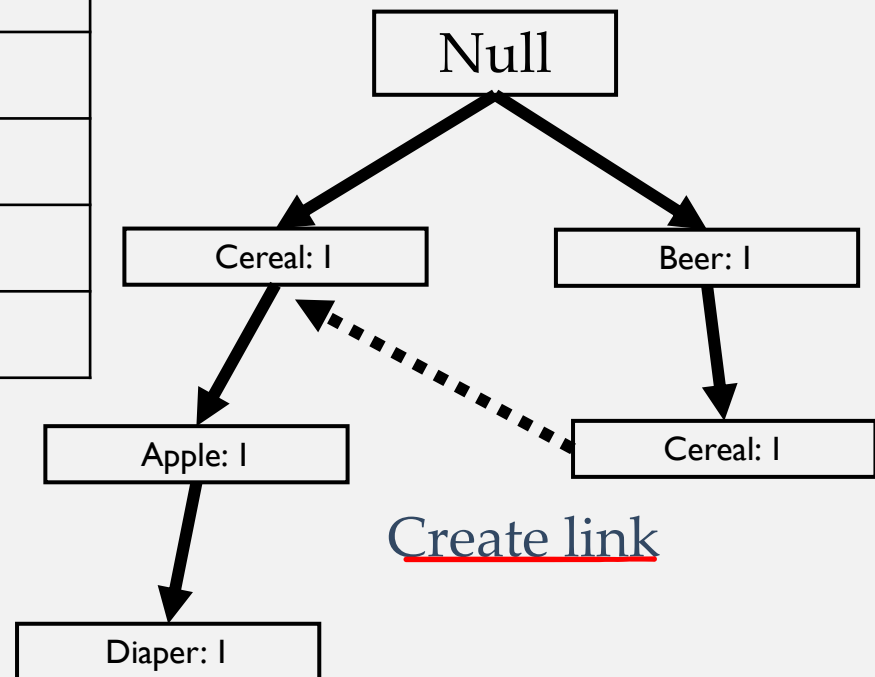


# FP-GROWTH

## 1. BUILD FP-TREE (8)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

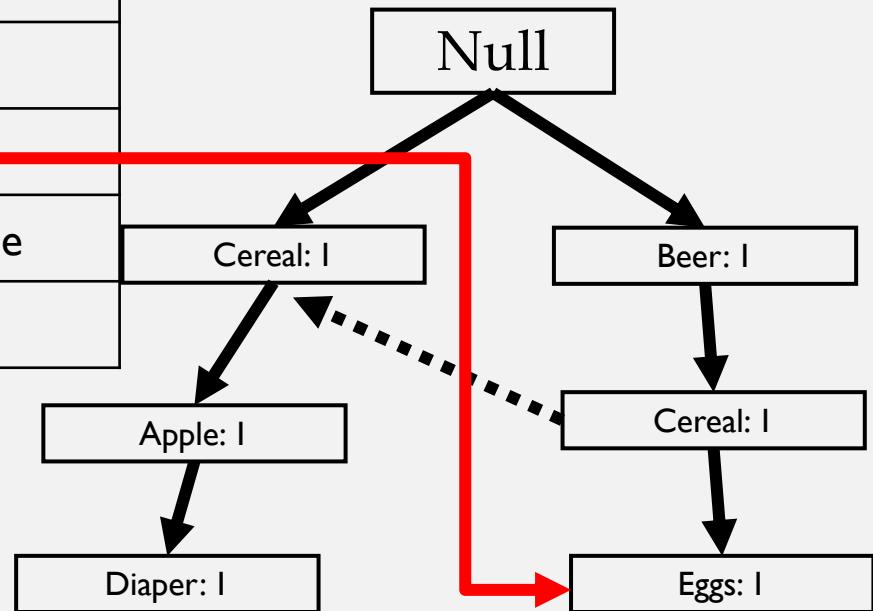


# FP-GROWTH

## 1. BUILD FP-TREE (9)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

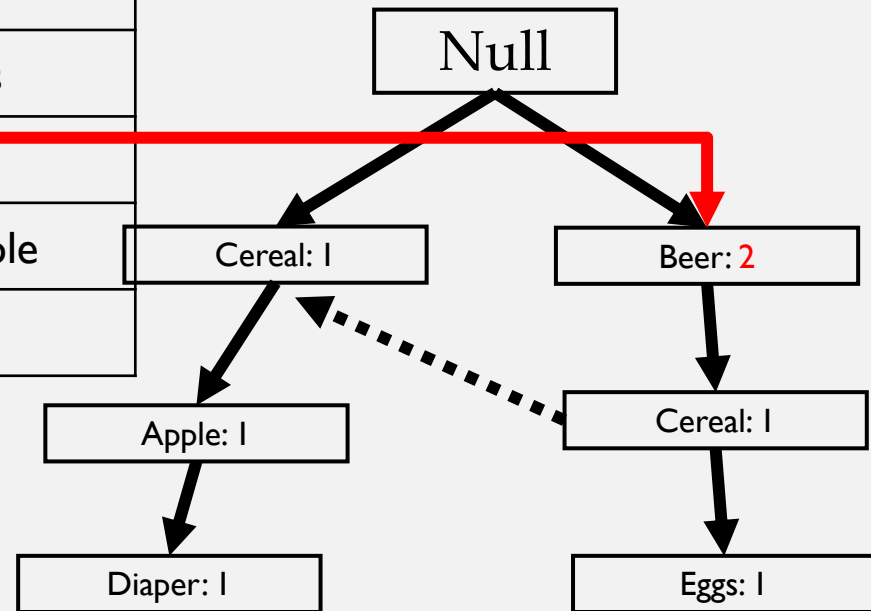


# FP-GROWTH

## 1. BUILD FP-TREE (10)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

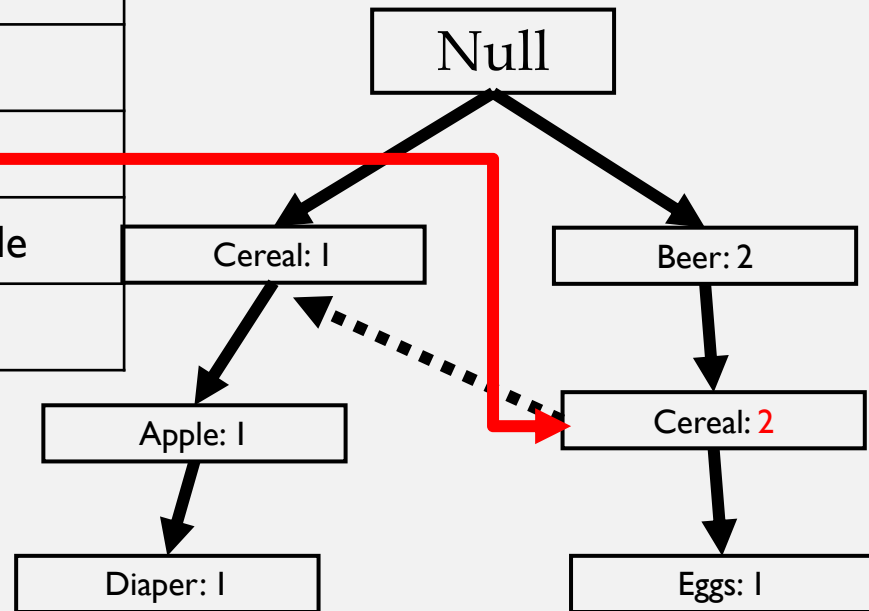


# FP-GROWTH

## 1. BUILD FP-TREE (11)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs



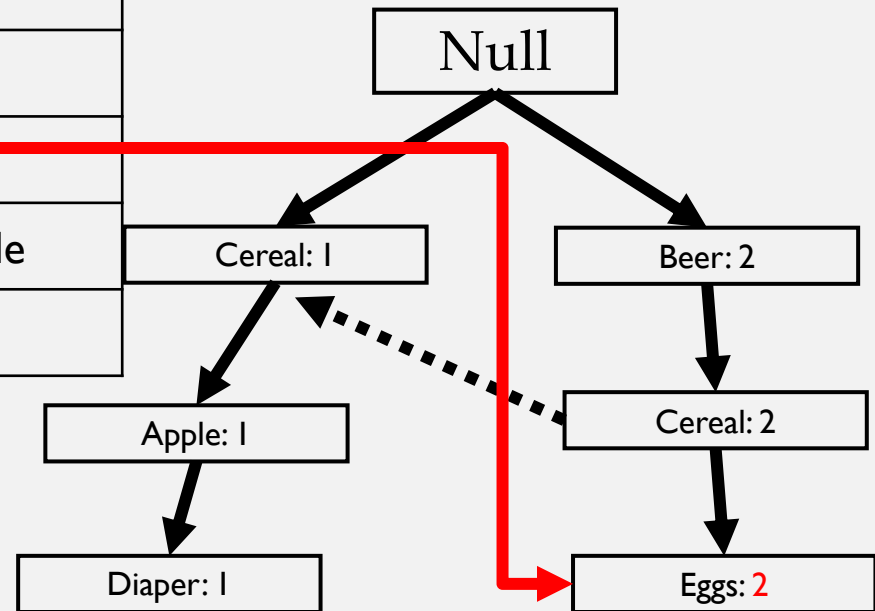


# FP-GROWTH

## 1. BUILD FP-TREE (12)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

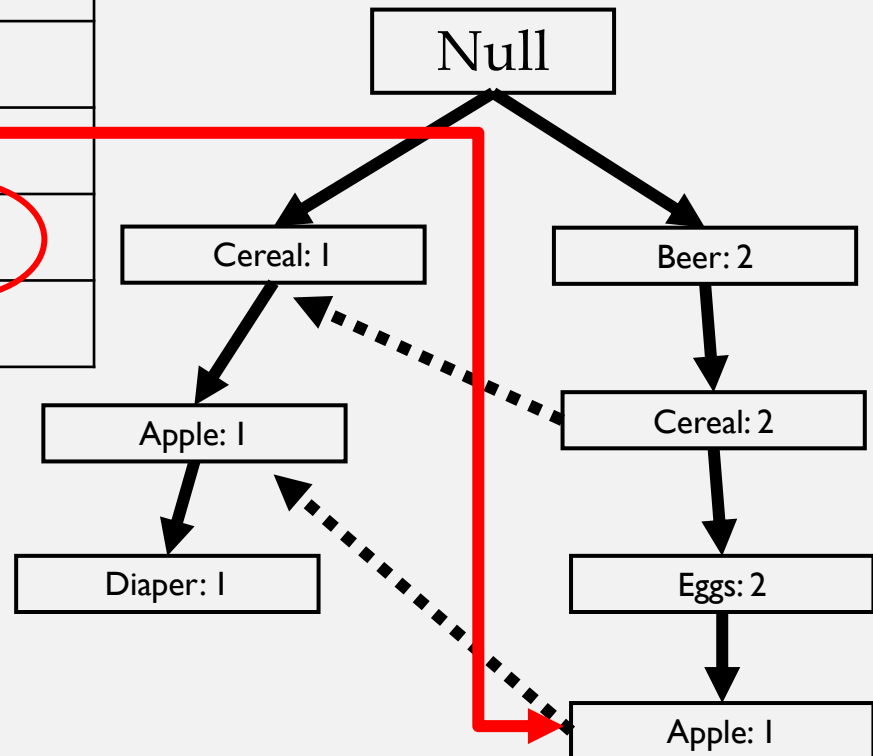


# FP-GROWTH

## 1. BUILD FP-TREE (13)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

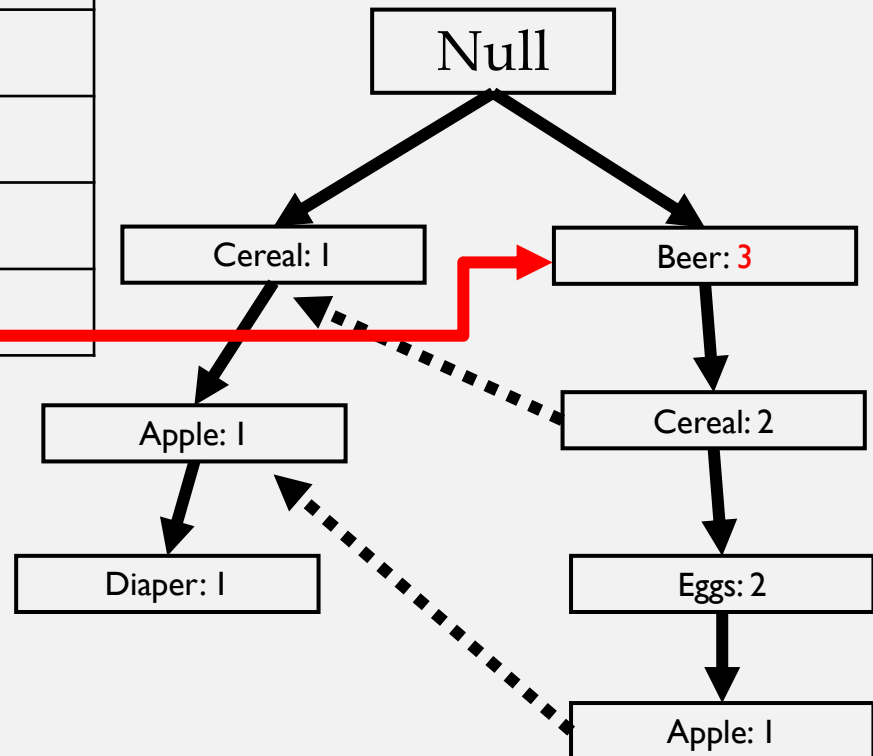


# FP-GROWTH

## 1. BUILD FP-TREE (14)

- Use the items information to build FP-tree

Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs

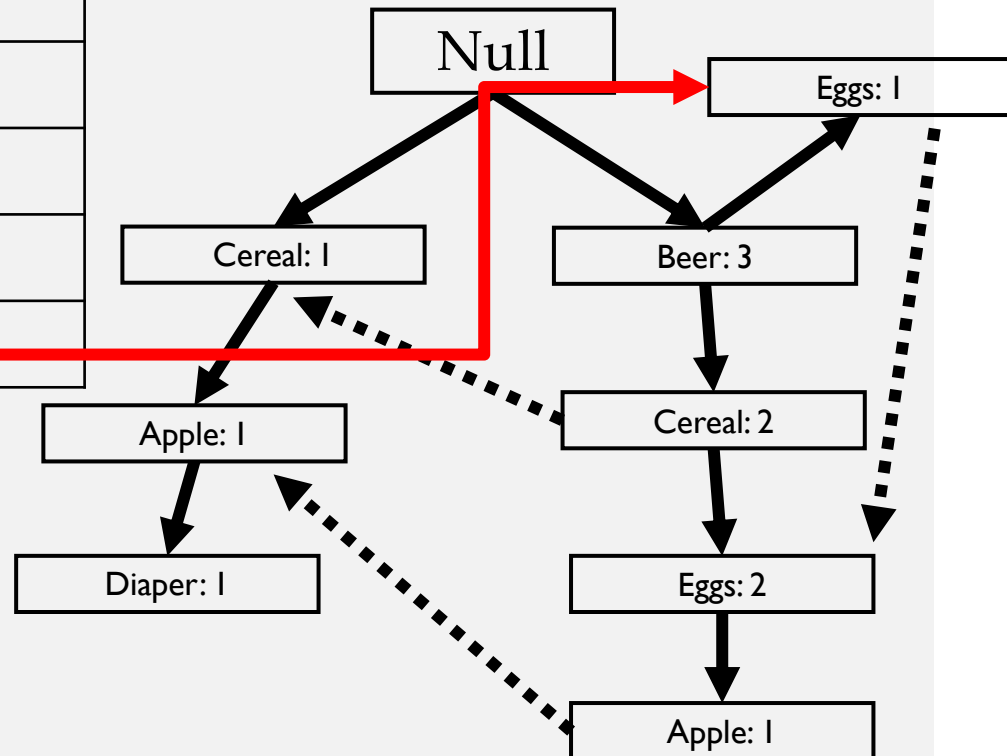


# FP-GROWTH

## 1. BUILD FP-TREE (15)

- Use the items information to build FP-tree

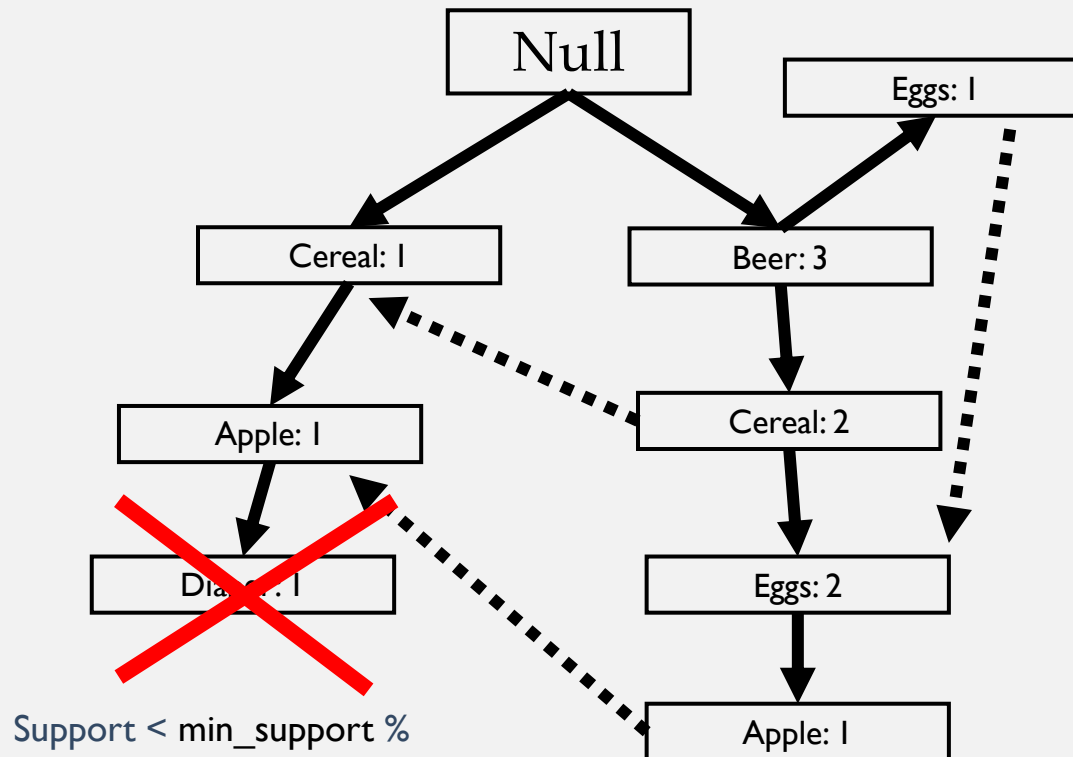
Transaction ID	Items
1	Cereal, Apple, Diapers
2	Beer, Cereal, Eggs
3	Beer, Cereal, Eggs, Apple
4	Beer, Eggs



## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(1)

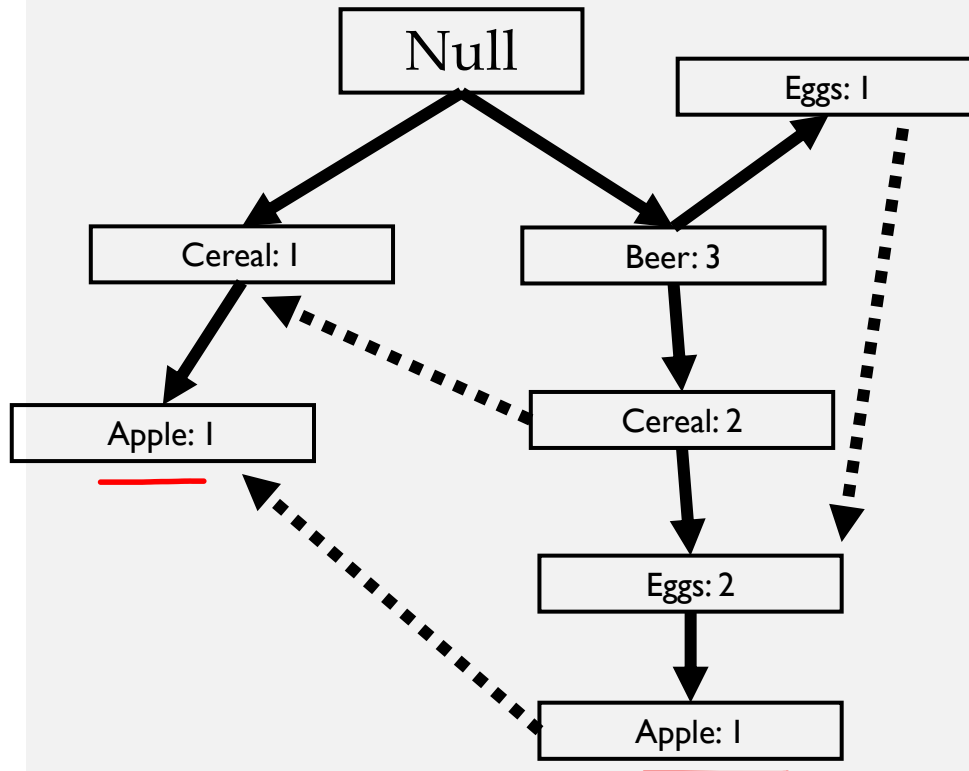
- min\_support = 50% (Defined by user)



## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(2)

- min\_support = 50% (Defined by user)

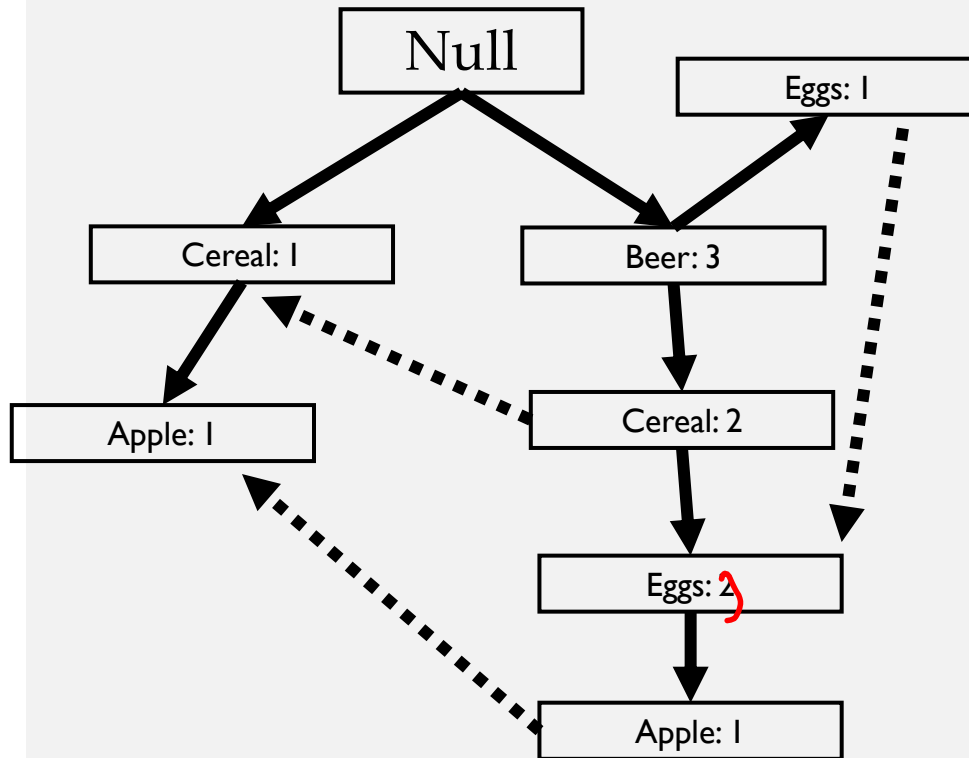
**Apple**

Frequent item set	Support
{Apple}	$2/4 = 50\%$

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(3)

- min\_support = 50% (Defined by user)

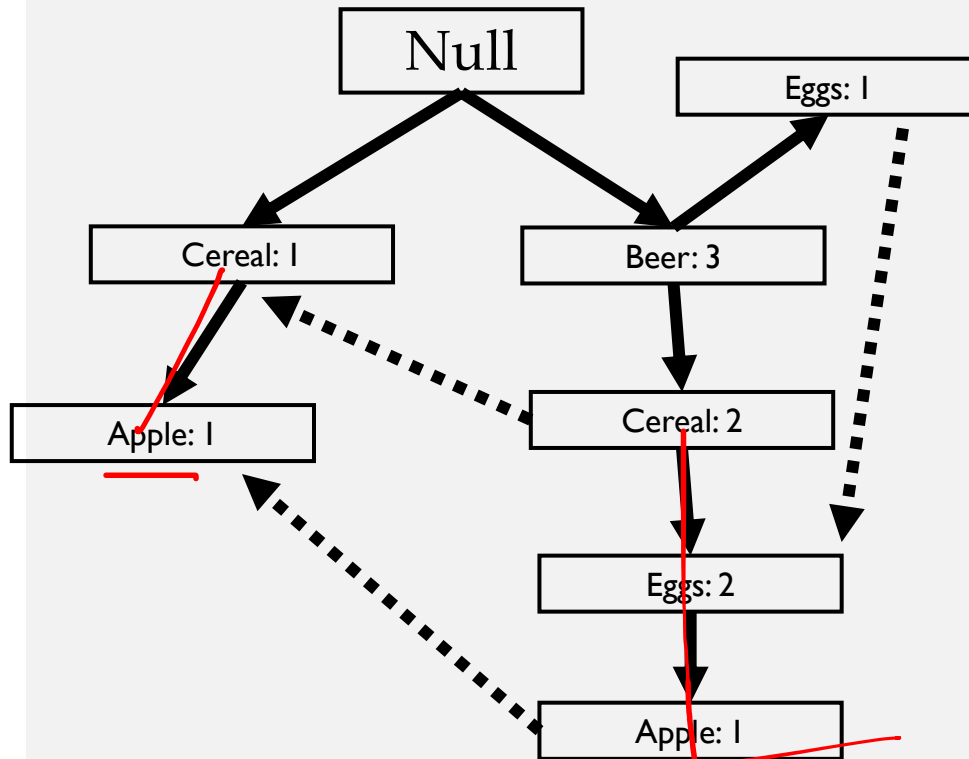
**Apple**

Frequent item set	Support
{Apple}	2/4 = 50%
<del>{Apple, Eggs}</del>	<del>1/4 = 25%</del>

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(4)

- min\_support = 50% (Defined by user)

**Apple**

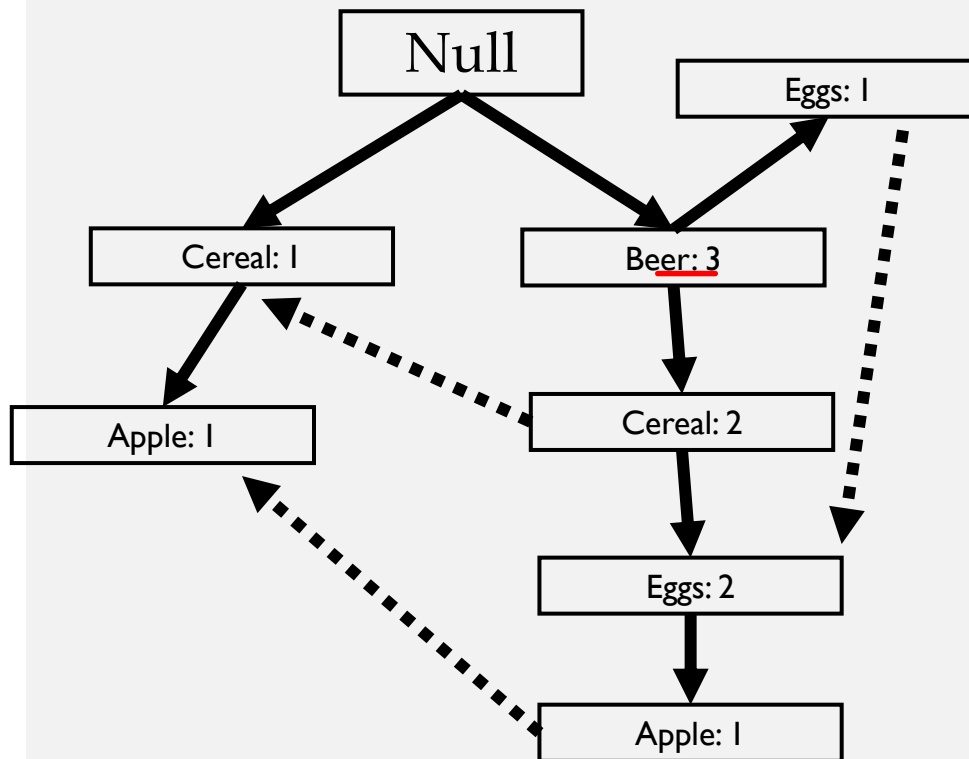
Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%



## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(5)

- min\_support = 50% (Defined by user)

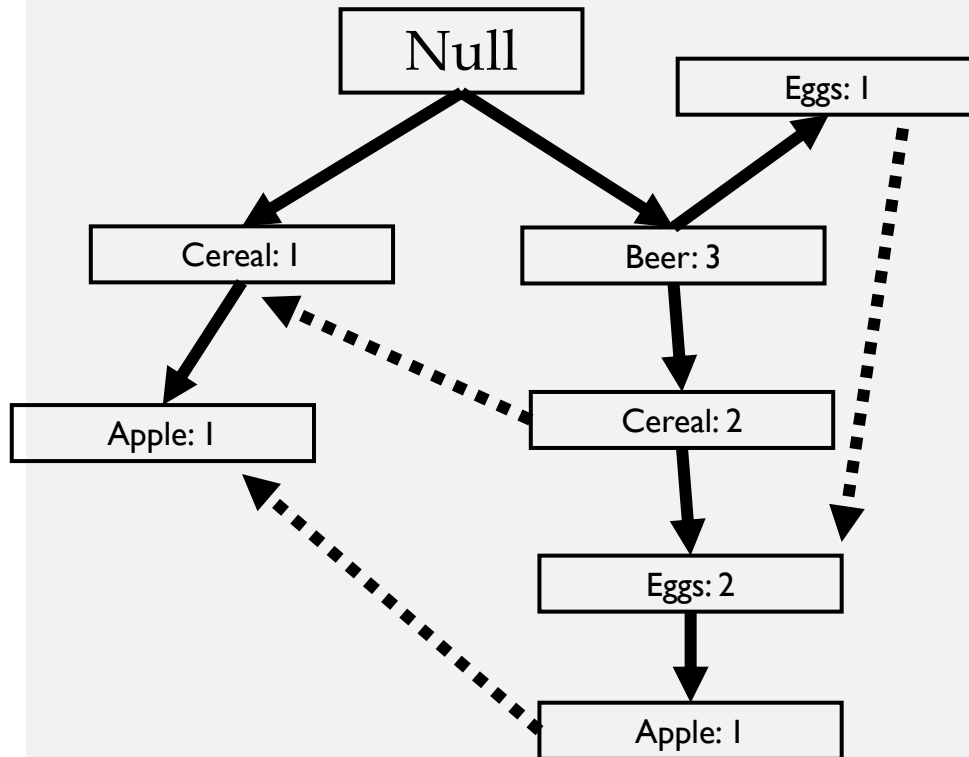
**Apple**

Frequent item set	Support
{Apple}	$2/4 = 50\%$
{Apple, Cereal}	$2/4 = 50\%$
{Apple, Beer}	$1/4 = 25\%$

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(6)

- min\_support = 50% (Defined by user)

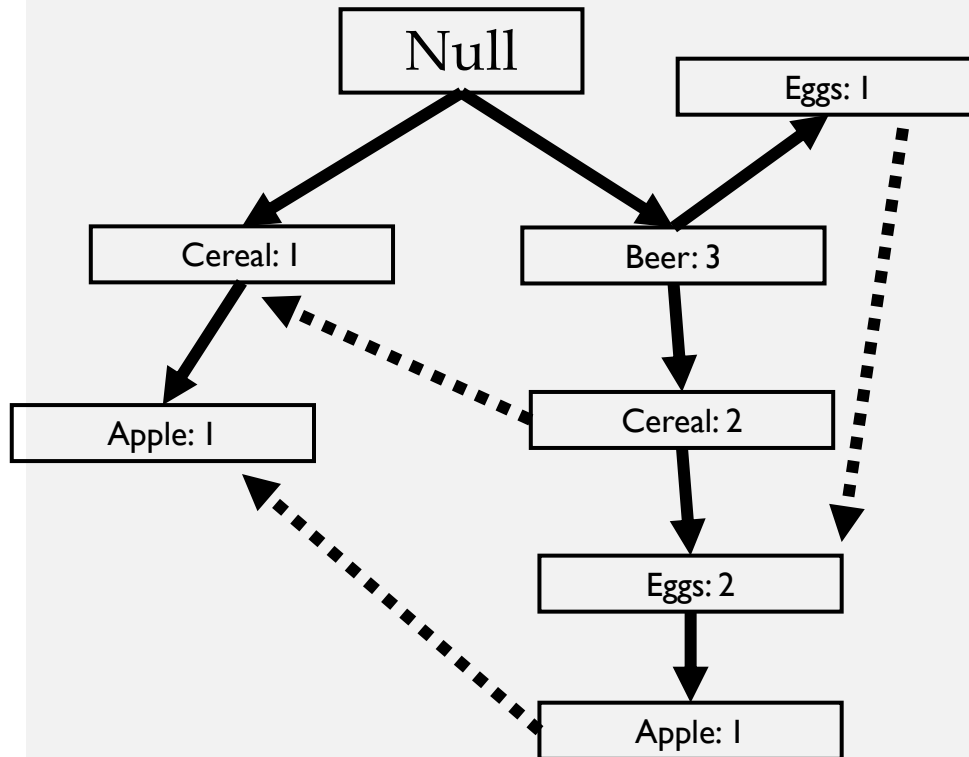
**Apple**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
<del>{Apple, Egg, Cereal}</del>	<del>1/4 = 25%</del>

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(7)

- min\_support = 50% (Defined by user)

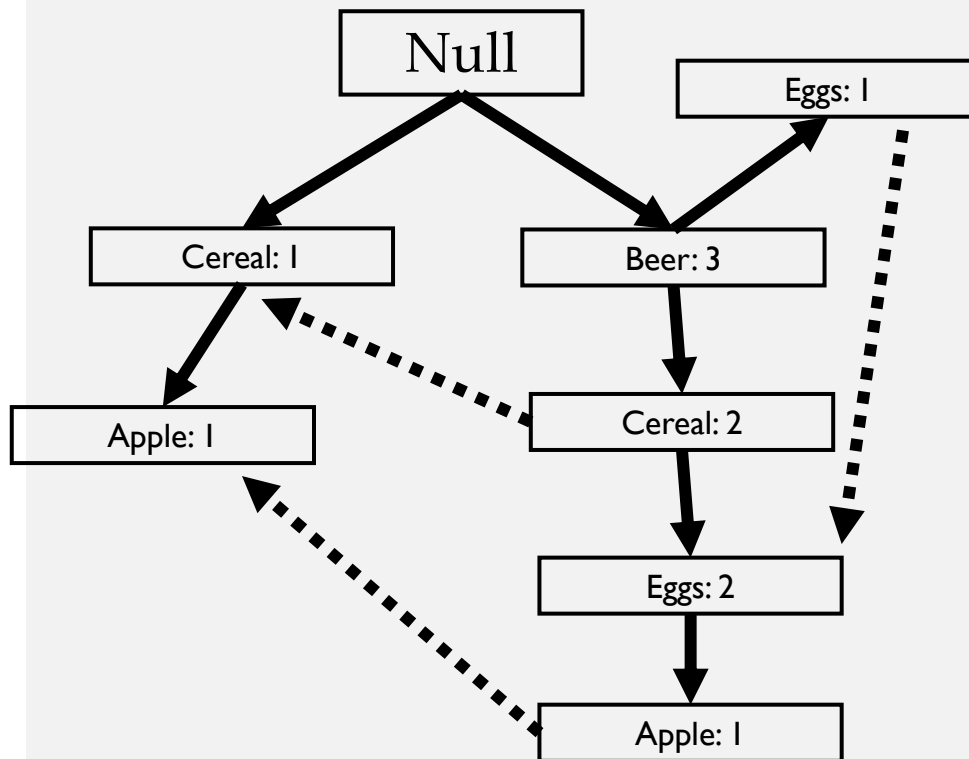
**Apple**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Apple, Egg, Beer}	<del>1/4 = 25%</del>

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(8)

- min\_support = 50% (Defined by user)

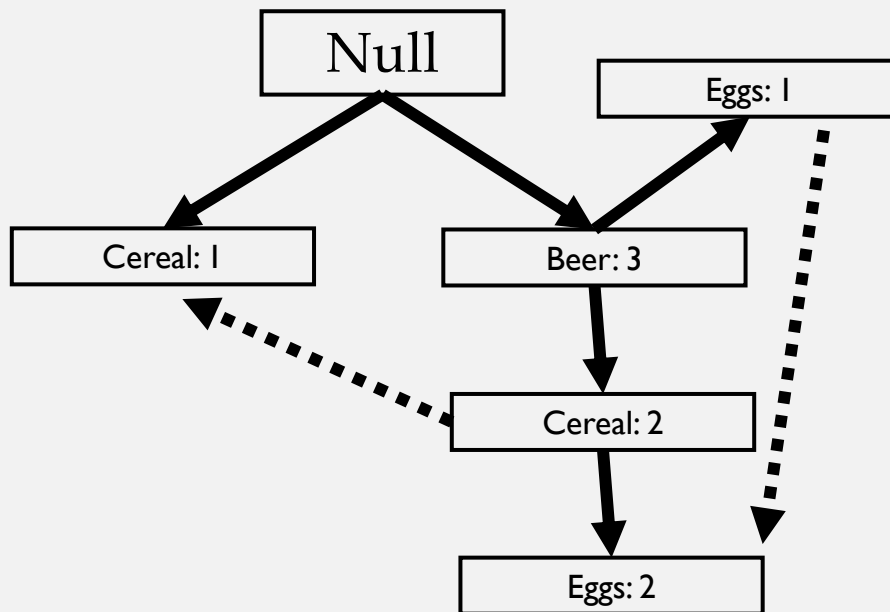
**Apple**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Apple, Cereal, Beer}	<del>1/4 = 25%</del>

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(9)

- min\_support = 50% (Defined by user)

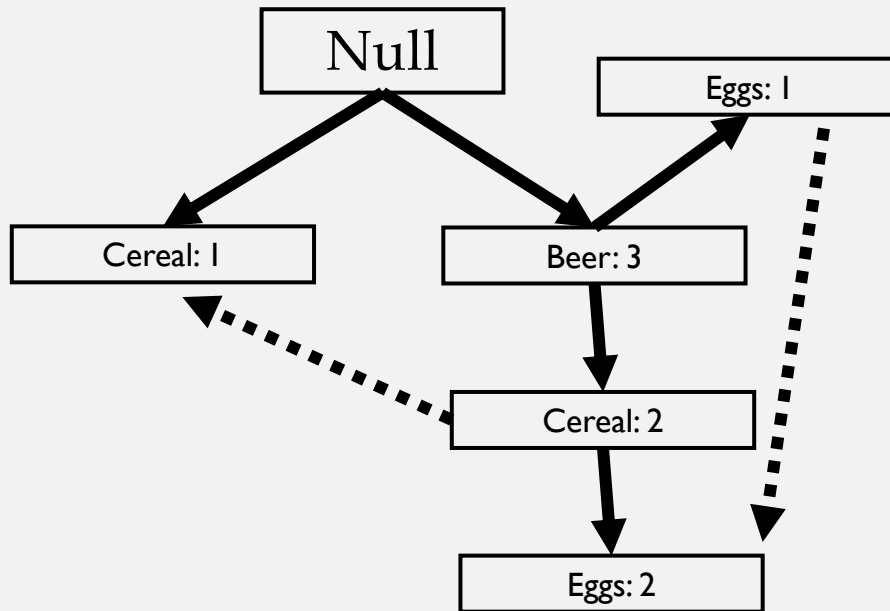
**Eggs**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(10)

- min\_support = 50% (Defined by user)

**Eggs**

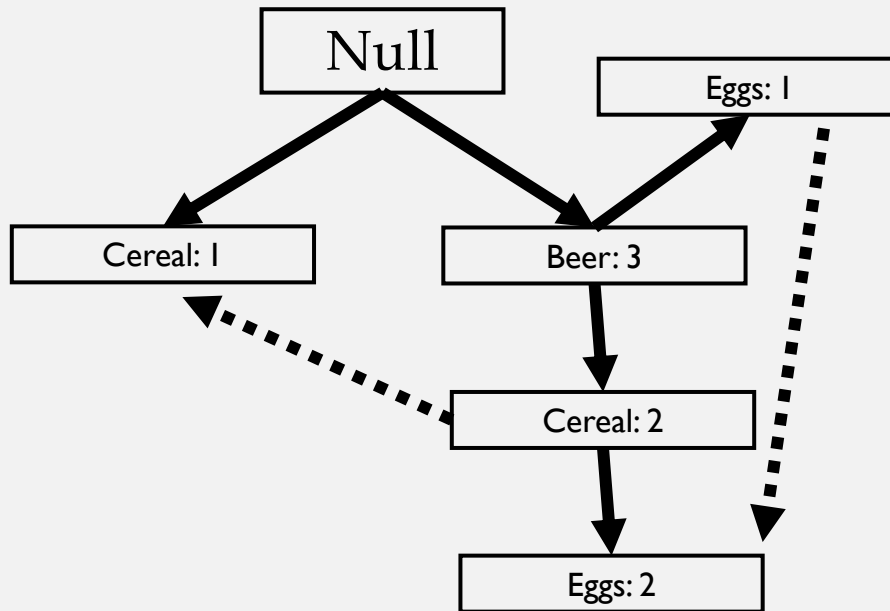
Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%
{Eggs, Cereal}	2/4 = 50%

# FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(11)

- min\_support = 50% (Defined by user)

**Eggs**

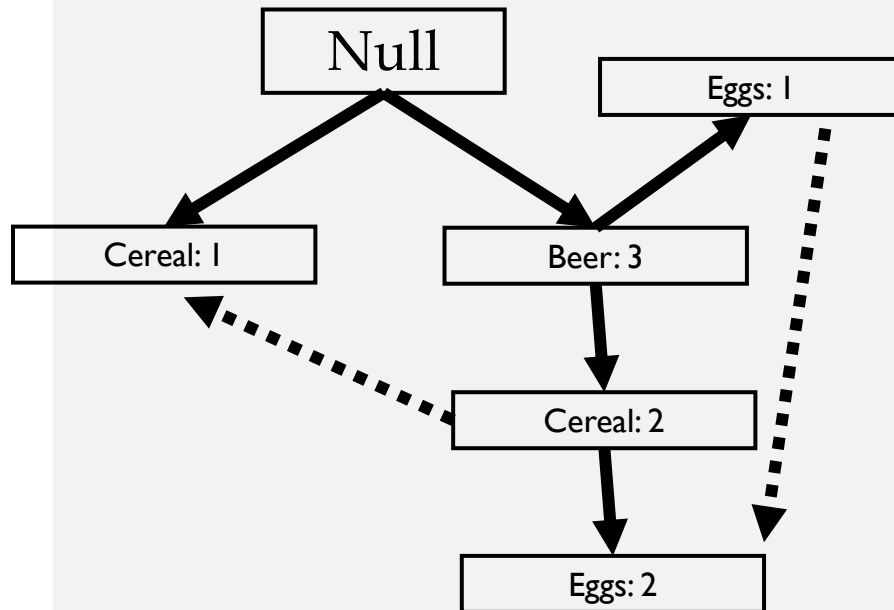


Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%
{Eggs, Cereal}	2/4 = 50%
{Eggs, Beer}	3/4 = 75%

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(12)

- min\_support = 50% (Defined by user)

**Eggs**

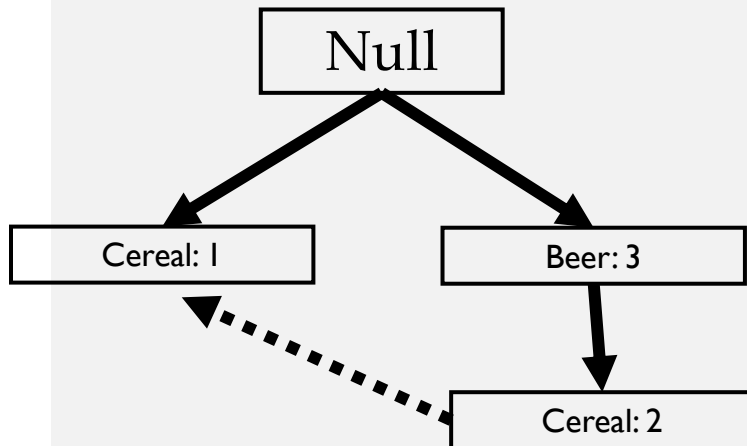
Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%
{Eggs, Cereal}	2/4 = 50%
{Eggs, Beer}	3/4 = 75%
{Eggs, Cereal, Beer}	2/4 = 50%



## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(13)

- min\_support = 50% (Defined by user)

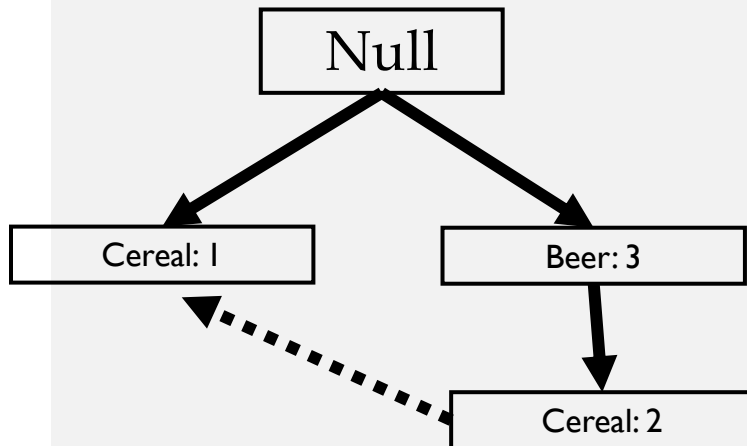
**Cereal**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%
{Eggs, Cereal}	2/4 = 50%
{Eggs, Beer}	3/4 = 75%
{Eggs, Cereal, Beer}	2/4 = 50%
{Cereal}	3/4 = 75%

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(14)

- min\_support = 50% (Defined by user)

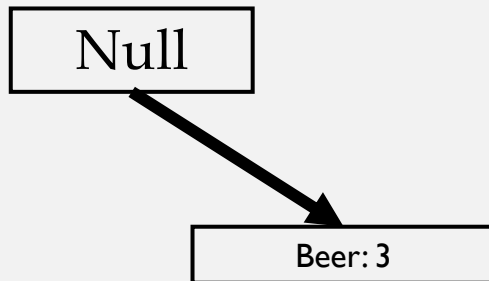
**Cereal**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%
{Eggs, Cereal}	2/4 = 50%
{Eggs, Beer}	3/4 = 75%
{Eggs, Cereal, Beer}	2/4 = 50%
{Cereal}	3/4 = 75%
{Cereal, Beer}	2/4 = 50%

## FP-GROWTH

## 2. GENERATE FREQUENT ITEM SETS FROM FP-TREE(15)

- min\_support = 50% (Defined by user)

**Beer**

Frequent item set	Support
{Apple}	2/4 = 50%
{Apple, Cereal}	2/4 = 50%
{Eggs}	3/4 = 75%
{Eggs, Cereal}	2/4 = 50%
{Eggs, Beer}	3/4 = 75%
{Eggs, Cereal, Beer}	2/4 = 50%
{Cereal}	3/4 = 75%
{Cereal, Beer}	2/4 = 50%
{Beer}	3/4 = 75%

# BUILDING ASSOCIATION RULES

- Determine from the products that customers buy together frequently.
- Structure: LHS  $\rightarrow$  RHS
  - LHS is Left Hand Side of the rule.
  - RHS is Right Hand Side of the rule.

Frequent item set	Support	Size
{Apple, Cereal}	$2/4 = 50\%$	2
{Eggs, Cereal}	$2/4 = 50\%$	2
{Eggs, Beer}	$3/4 = 75\%$	2
{Eggs, Cereal, Beer}	$2/4 = 50\%$	3
{Cereal, Beer}	$2/4 = 50\%$	2

Apple  $\rightarrow$  Cereal

Cereal  $\rightarrow$  Apple

Egg  $\rightarrow$  Cereal

Cereal  $\rightarrow$  Egg

Egg  $\rightarrow$  Beer

Beer  $\rightarrow$  Egg

Eggs, Cereal  $\rightarrow$  Beer

Eggs, Beer  $\rightarrow$  Cereal

Cereal, Beer  $\rightarrow$  Eggs

Cereal  $\rightarrow$  Beer

Beer  $\rightarrow$  Cereal

Appl  
Cereal'

# PERFORMANCE MEASURES (1)

---

- **Confidence** is an indication of the likelihood that RHS is also bought if item LHS is bought.
  - Show the confidence of RHS when LHS occurred before.
  - $Confidence(LHS \rightarrow RHS) = \frac{support(LHS, RHS)}{support(LHS)}$
  - $Confidence(\text{Apple} \rightarrow \text{Cereal}) = \frac{support(\text{Apple}, \text{Cereal})}{support(\text{Apple})}$ 
$$= \frac{2/4}{2/4}$$
$$= 100\%$$

## PERFORMANCE MEASURES (2)

- **Lift** shows correlation between LHS and RHS, which means the increase in the ratio of sale of RHS when LHS is sold
  - If the **lift** is equal to **1**, it means that LHS and RHS are
  - If the **lift** is higher **than 1**, it means that LHS and RHS are  correlated.
  - If the **lift** is **lower than 1**, it means that LHS and RHS are  correlated.
- $$Lift(LHS \rightarrow RHS) = \frac{support(LHS, RHS)}{support(LHS) \times support(RHS)}$$
- $$Lift(Apple \rightarrow Cereal) = \frac{support(Apple, Cereal)}{support(Apple) \times support(Cereal)}$$

$$= \frac{2/4}{2/4 \times 3/4}$$

$$= 1.33$$

# SUMMARY

---

The diagram illustrates the relationship between a rule  $X \Rightarrow Y$  and its three associated metrics. The rule is positioned on the left, with three blue arrows pointing from it to the right. The top arrow points to the Support formula, the middle arrow points to the Confidence formula, and the bottom arrow points to the Lift formula.

$$\begin{aligned} \text{Rule: } X \Rightarrow Y & \begin{cases} \nearrow \text{Support} = \frac{\text{freq}(X, Y)}{N} \\ \longrightarrow \text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)} \\ \searrow \text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)} \end{cases} \end{aligned}$$

## EXAMPLE 1:



Rule	Support	Confidence	Lift
$A \Rightarrow D$			
$C \Rightarrow A$			
$A \Rightarrow C$			
$B \& C \Rightarrow D$			



## EXAMPLE 2: MARKET BASKET ANALYSIS

Row No.	Invoice	product 1	Orders	Sales value
1	131506	Product 20	1	40
2	131506	Product 21	1	80
3	131507	Product 11	1	80
4	131508	Product 19	1	32
5	131509	Product 31	1	9
6	131510	Product 11	1	80
7	131510	Product 20	2	40
8	131510	Product 20	1	40
9	131519	Product 11	1	80
10	131511	Product 11	1	80

ExampleSet (784 examples, 0 special attributes, 4 regular attributes)

# EXAMPLE 2: MARKET BASKET ANALYSIS

## Step 1:

Load transaction data containing a transaction id, a product id and a quantifier. The data denotes how many times a certain product has been purchased as part of a transactions.

### Load Transactions



### Aggregate



### Rename



### Set Role



## Step 2:

Edit, transform & load (ETL) - Aggregate transaction data via concatenation so that the products in a transaction are in one entry, separated by the pipe symbol.

## Step 3:

Using FP-Growth, determine frequent item sets. A frequent item sets denotes that the items (products) in the set have been purchased together frequently, i.e. in a certain ratio of transactions. This ratio is given by the support of the item set.

### FP-Growth



### Create Association ...



## Step 4:

Create association rules which can be used for product recommendations depending on the confidences of the rules.

# EXAMPLE 3: 3 TYPES OF INPUT FOR FP-GROWTH

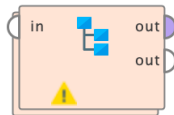
Only "items in dummy coded columns" can be used for "Apply Association rules"

Process

FP-Growth can have different input formats that can be selected via the first parameter.

Look inside the subprocesses to see how data could be transformed into one of the possible input formats.

input format 1

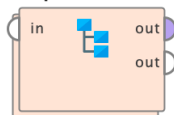


The items are in dummy coded columns:  
The columns represent the products.  
An entry 'true' in a column means that this product is present for this row.

FP-Growth 1



input format 2

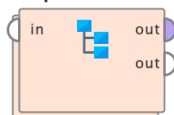


The items are in a list in one column:  
There is only one column with entries representing a list of items.  
The item list contains the items encoded in a csv-like format. In this example they are separated by the pipe symbol.

FP-Growth 2



input format 3



The items are in separate columns:  
The first item of a transaction is in the first column, the second item in the second column and so forth.  
A missing value indicates that there is no item.

FP-Growth 3



re

re

re

re

re

re

## HW 9: WHAT ARE FREQUENTLY SHOPPED TOGETHER IN **GROCERIES MARKET BASKET DATASET?**

---

- The Groceries Market Basket Dataset contains 9835 transactions by customers shopping for groceries. The data contains 169 unique items.
- This is the groceries data with the list of items bought by customers. From the left side is the number of items in a basket then Item 1, 2, 3, etc stands for list of the item. If you think can improve or just to make sure about the data just leave some words in the comment section below.

## H9: SOLUTION STEPS

### 1. Data Preparation & Cleaning

- Select attributes
- Generate ID/ Rename it to be (Order-ID)/ Set its Role to be (Regular)
- **De-Pivot** (item 1,2,3... to be one "item" list)
- Generate "items in dummy coded columns" format for FP-Growth

### 2. Data Visualization & Analysis

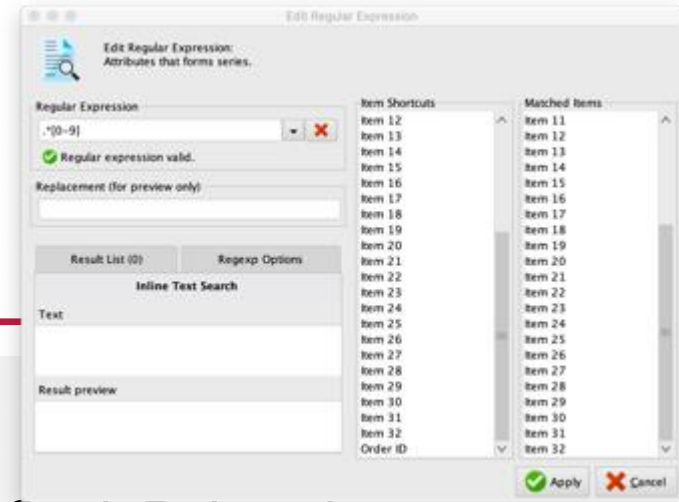
- Split data

### 3. FP-model:

- Choose a proper min\_support (based on the product fraction)

### 4. Generate\apply Association Rules:

- Choose a proper confidence or lift
- Apply association rules



# REFERENCES

---

- Han, J., Pei, J., & Yin, Y. (2000, May). Mining frequent patterns without candidate generation. In ACM sigmoid record (Vol. 29, No. 2, pp. 1-12). ACM.
- Ekasit Pacharawongsakda (2017). *Practical Data Mining with RapidMiner Studio 8*.
- [http://rasbt.github.io/mlxtend/user\\_guide/frequent\\_patterns/fpgrowth/](http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/fpgrowth/) (FP-growth in Python)