# CONSUMER DECISION TREE

THE FRESH QUERY FASHION OF EBAY CUSTOMER

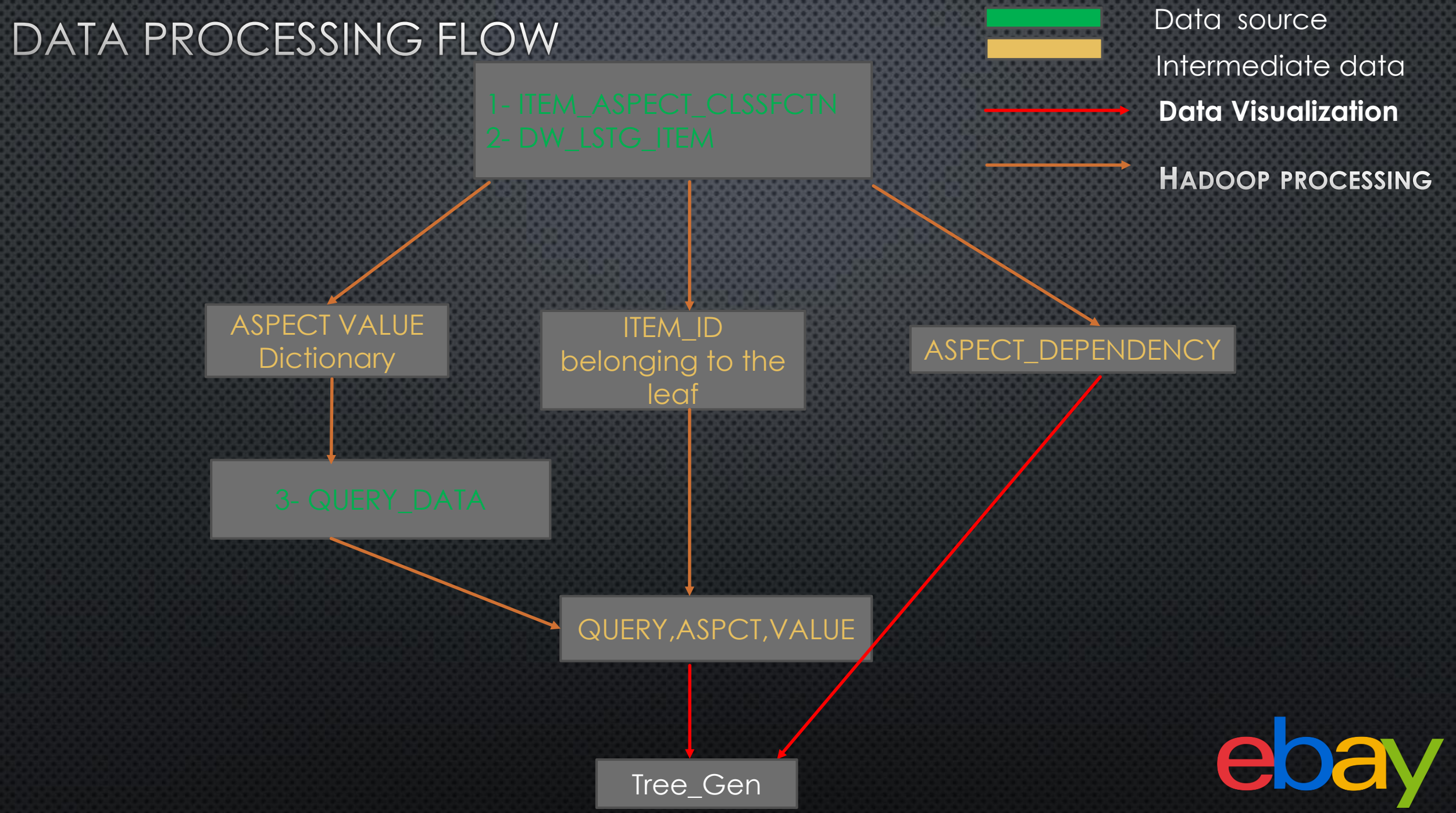
# AGENDA

1. Motivation
2. Data Process Outline
3. Implementation
    1. Tree generation
    2. handling aspect dependencies
    3. Query Generalization
    4. Threshold for filtering
4. Demo
5. Conclusion

# MOTIVATION

- OVER 328,104,000 QUERY DATA CREATED PER DAY.
- GENERALIZING THE QUERY DATA AND VISUALIZING IT IN A HUMAN FRIENDLY WAY IS THE GOAL OF THIS PROJECT

# DATA PROCESSING FLOW



# IMPLEMENTATION DETAIL

- TIME PERIOD : QUARTER WORTH OF DATA FOR DW_LSTG_ITEM AND QUERY DATA
- REFRESH FREQUENCY : MONTHLY
- INPUT : USER QUERY HISTORY

```
cat pico
eel tote
tote flag
bag michel
cabas tote
yellow ugg
bag sequins
hobo bianka
jemco purse
paloma  bag
purse waist
```

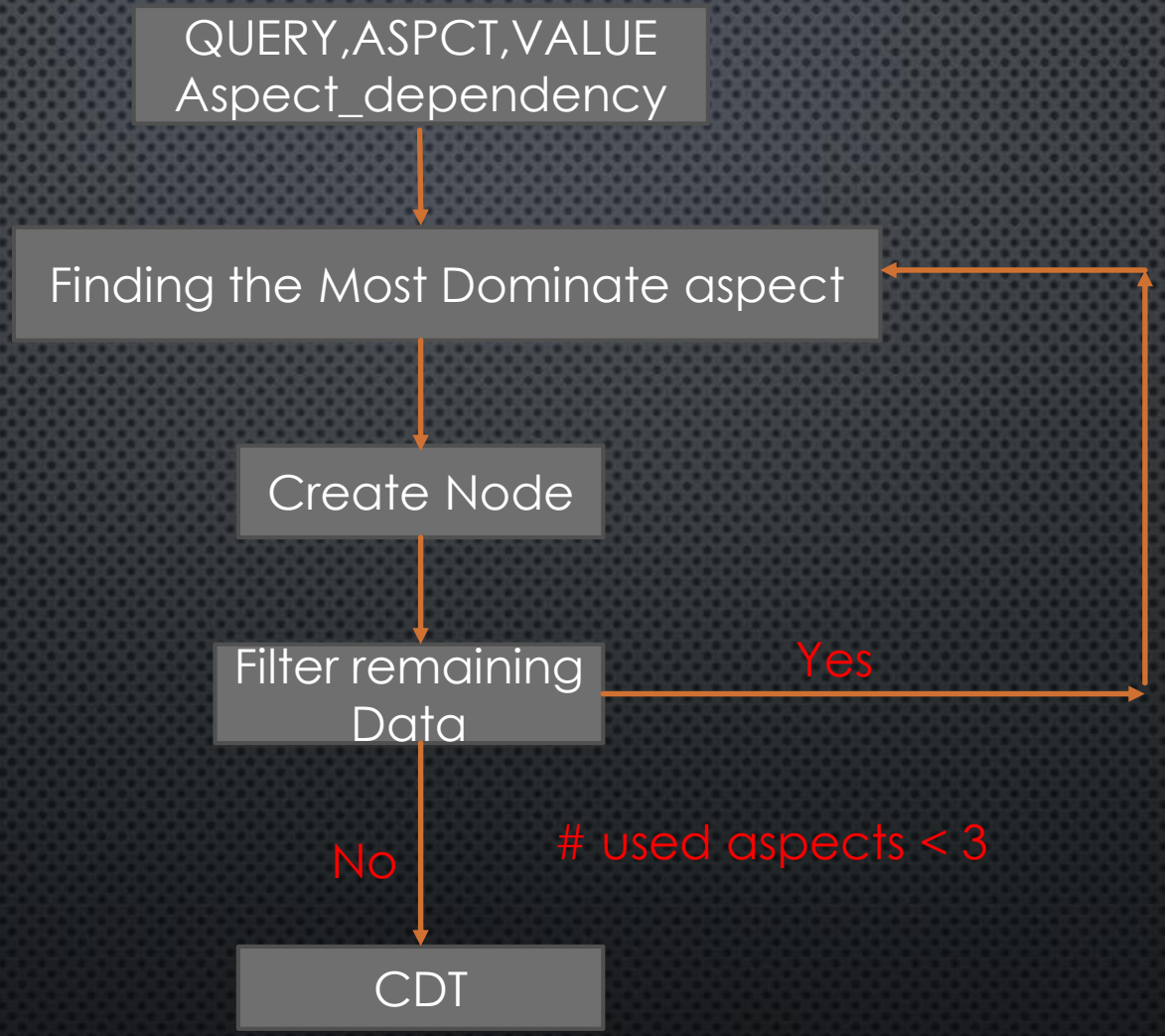
- OUTPUT : ASPECT_DEPENDENCE

| | | |
|---|---|---|
| Size | Type | 1.0404271863295997 |
| Size | Color | 3.5172675299805585 |
| Size | Style | 2.628315362991928 |
| Size | Theme | 1.9308485776904087 |
| Size | Gender | 0.003839656554871644 |
| Size | Closure | 0.9981675148281981 |
| Size | Pattern | 1.753928912429129 |

QUERY_FREQUENCY_ASPECT_VALUE

| | | | | | |
|---|---|---|---|---|---|
| cat pico | 1 | Theme | cat | Theme | cat |
| eel tote | 1 | Type | tote | Type | tote |
| tote flag | 1 | Type | tote | Type | tote |
| bag michel | 1 | Type | bag | Type | bag |
| cabas tote | 11 | Type | tote | Type | tote |
| yellow ugg | 1 | Color | yellow | Color | yellow |
| bag sequins | 1 | Type | bag | Type | bag |
| hobo bianka | 2 | Style | hobo | Style | hobo |
| jemco purse | 4 | Type | purse | Type | purse |
| paloma  bag | 1 | Type | bag | Type | bag |
| purse waist | 2 | Type | purse | Type | purse |

# TREE GENERATION FLOW CHART



# RIGHT ASPECT TO SPLIT THE ROOT

Aspect dependency between model and Brand
The Dominant aspect (i.e. Brand) is chosen as the root

```
Model   RAM     0.6343940650191069
Model   Type    0.8155021060078818
Model   Brand   0.1408226430218756
Model   Color   1.8182620655389206
```

One Brand has many models, yet one model only belongs to one Brand.

So we choose the Brand as the appropriate aspect.

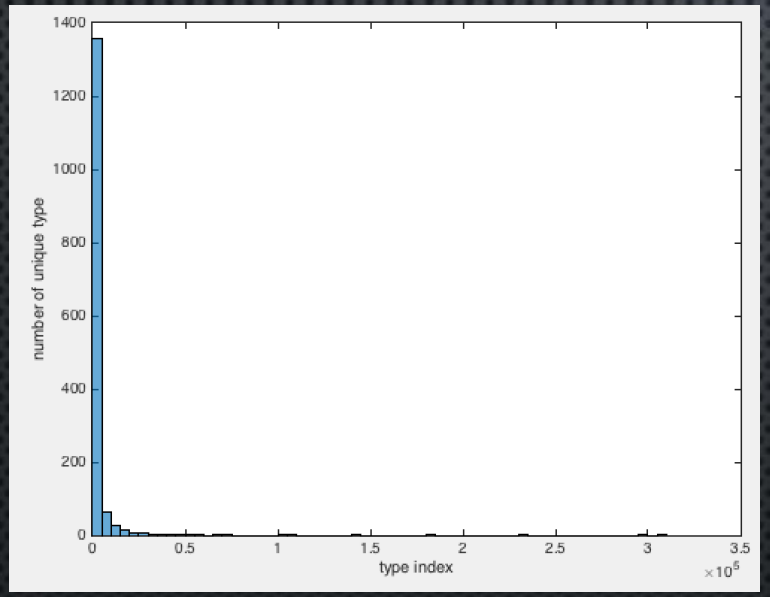
# DIGIT AND ALPHABET

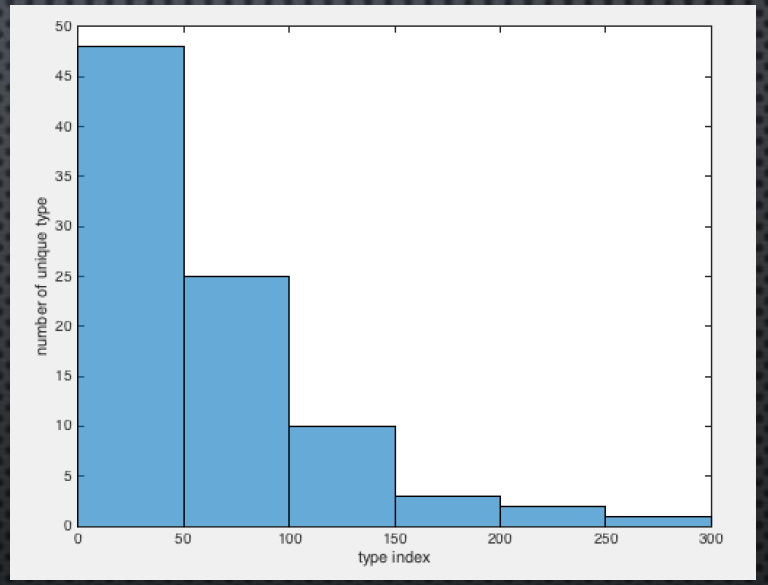
1. AIR FORCE 1 (BRAND) → AIR FORCE ONE (PRODUCT LINE)


2. 1 PLUS -> ONE PLUS

# THRESHOLD FILTERING

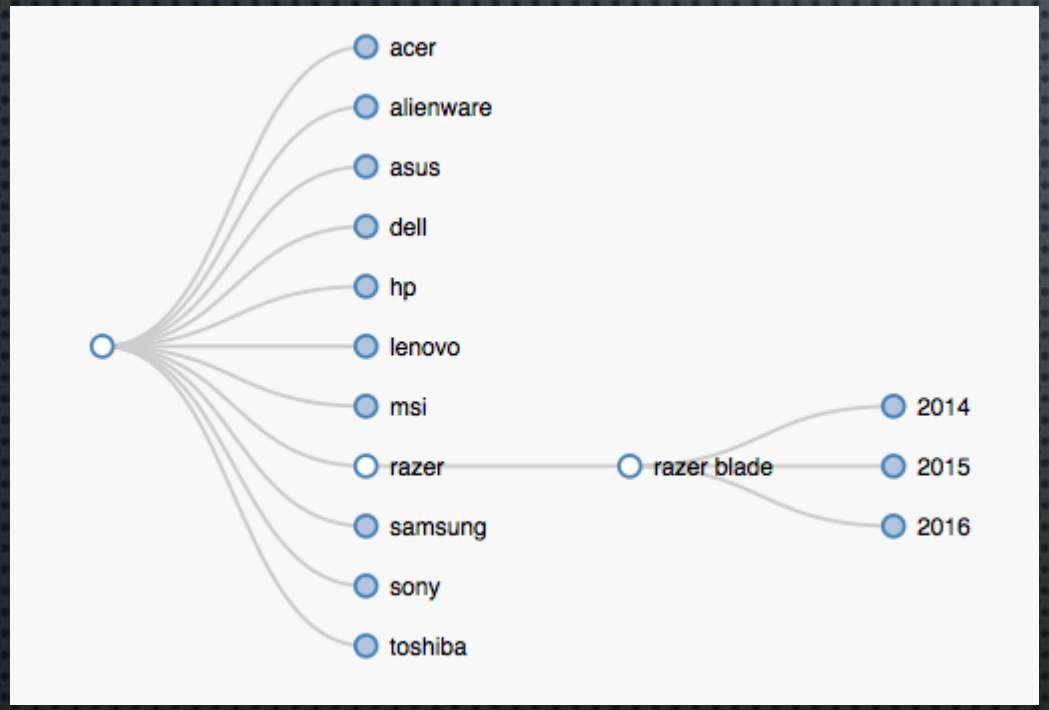
- QUESTION : HOW TO SET THE THRESHOLD FILTERING AUTOMATICALLY, DEPENDING ON THE VARIOUS SIZE OF DATASET?
- OBSERVATION : HUGE DATASET(CELL PHONE) SMALL DATASET( KITCHEN & DINNING BAR)



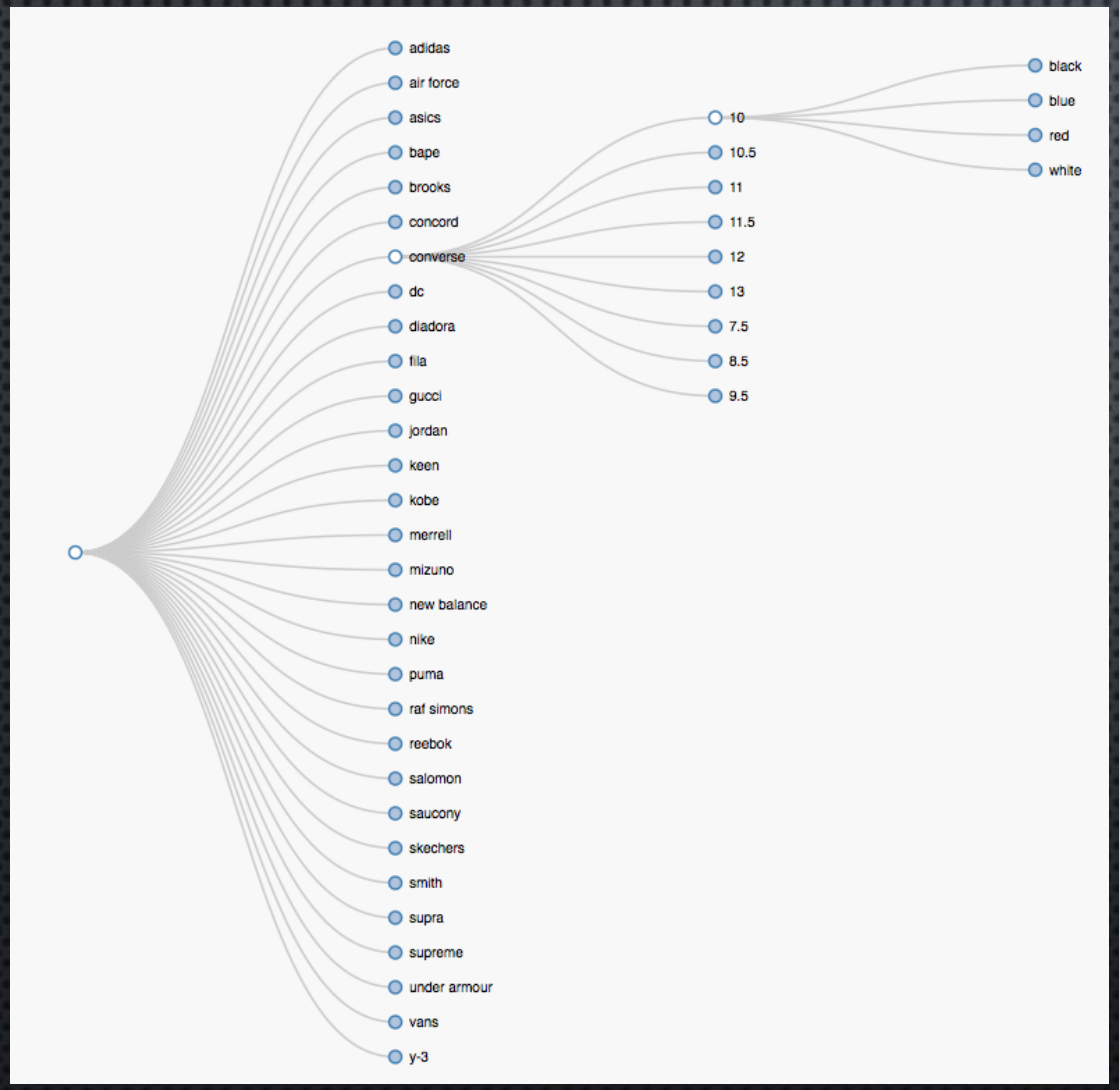

- PARETO PRINCIPLE (**80–20 RULE**): CDT 80% INFORMATION ARE CREATED BY 20% DISTINCT TYPE QUERY HISTORY.

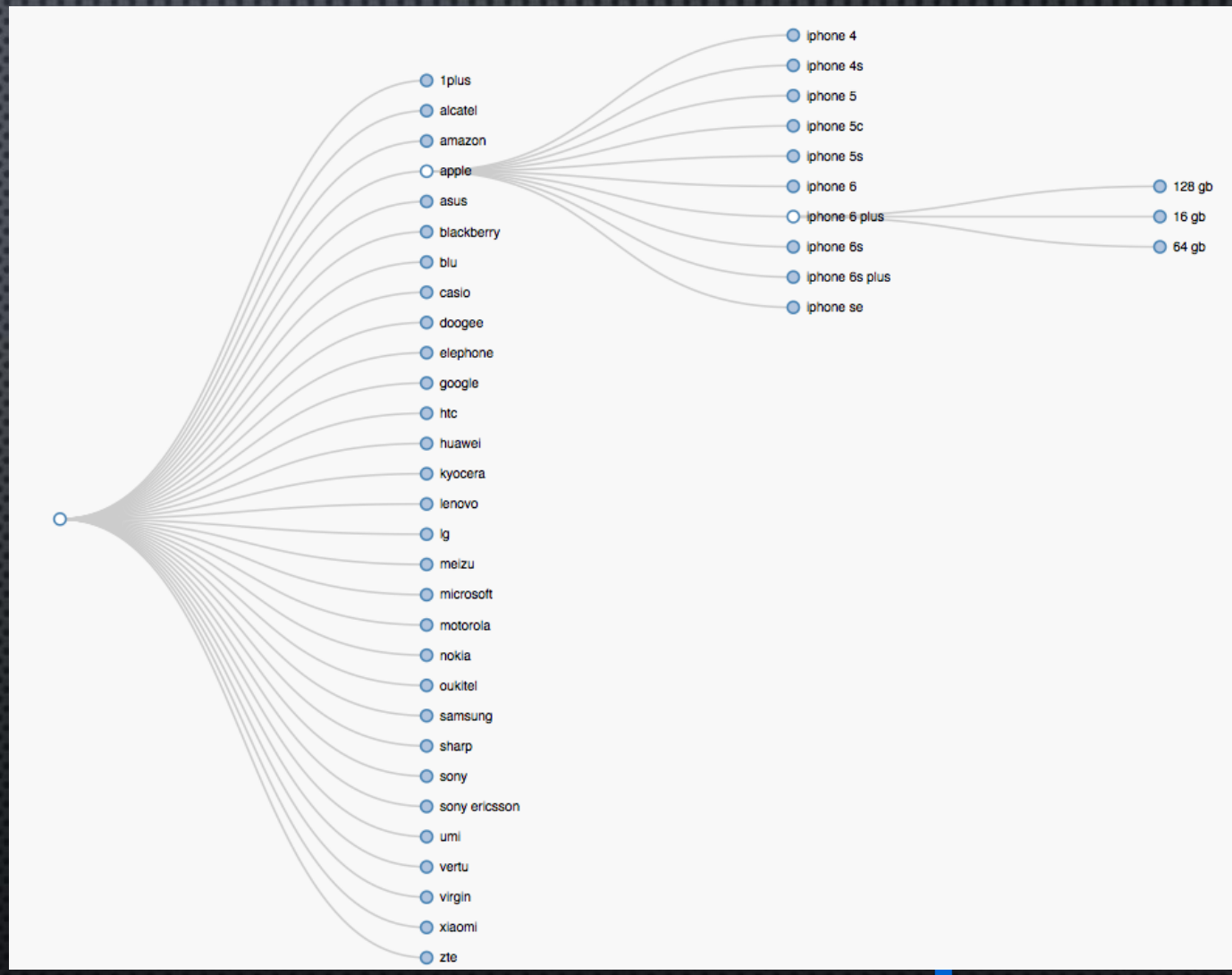
# THRESHOLD FILTERING




- Sorting the query history reversely
- Distinct the query frequency
- Obtaining the minimum value of top 20% query's frequency
- How many different queries' frequency > minimum value
- $N^3 \approx$ total
- Threshold ≈ N

# DEMO


adidas
air force
asics
bape
brooks
concord
converse
dc
diadora
fila
gucci
jordan
keen
kobe
merrell
mizuno
new balance
nike
puma
raf simons
reebok
salomon
saucony
skechers
smith
supra
supreme
under armour
vans
y-3
10
10.5
11
11.5
12
13
7.5
8.5
9.5
black
blue
red
white


1plus
alcatel
amazon
apple
asus
blackberry
blu
casio
doogee
elephone
google
htc
huawei
kyocera
lenovo
lg
meizu
microsoft
motorola
nokia
oukitel
samsung
sharp
sony
sony ericsson
umi
vertu
virgin
xiaomi
zte
iphone 4
iphone 4s
iphone 5
iphone 5c
iphone 5s
iphone 6
iphone 6 plus
iphone 6s
iphone 6s plus
iphone se
128 gb
16 gb
64 gb

# CONTRIBUTION

- USE THE FREQUENCY AND ENTROPY AS BASIC CRITERIA TO IMPLEMENT THE CONSUMER DECISION TREE
- RIGHT ASPECT TO SPLIT THE ROOT
- SIMILAR ITEMS AGGREGATION
- AUTOMATICALLY SETTING THRESHOLD

# THANKS TO GRO


ebay