

Assignment 4

Frequent Itemsets

$$(2) \text{ Support } (B, C) = 3$$

$$\text{Support } (B, A) = 1$$

$$\text{Support } (B, H) = 2$$

$$\text{Support } (B, J) = 2$$

$$\text{Support } (C, A) = 0$$

$$\text{Support } (C, H) = 3$$

$$\text{Support } (C, J) = 2$$

$$\text{Support } (A, H) = 1$$

$$\text{Support } (A, J) = 1$$

$$\text{Support } (H, J) = 2$$

Given, support threshold = 2

$$\text{frequent itemsets} = \{(B, C), (B, H), (B, J), (C, H), (C, J), (H, J)\}$$

$$(3) \quad S = 10,000$$

$$\text{items} = 1,000,000$$

$$\text{Prob of bucket to be frequent} = \frac{1,000,000}{\# \text{ buckets}}$$

No of freq. pairs that map to a bucket

$$\left(P \times \frac{1,000,000}{\# \text{ buckets}} \right)$$

During pass 1, we have at least

$$\frac{S - 4MB}{4} \approx \frac{S}{4} \text{ buckets in hash table}$$

$$\text{So for pass 2, we need } P \times \frac{12,000,000}{\# \text{ buckets}}$$

$$\text{So, } S^2 = 49548387 \times P$$

$$\therefore P \leq \frac{S^2}{49548387}$$

④ Set of items = $\{A, B, C, D, E, F, G, H\}$

Frequent itemsets = $\{A, B\}, \{A, C\}, \{A, D\},$
 $\{B, C\}, \{E\}, \{F\}$

Singleton sets in negative border $\{G\}, \{H\}$

Double sets in negative border $\{A, E\}, \{A, F\},$
 $\{B, D\}, \{B, E\}, \{B, F\}, \{C, D\}, \{C, E\},$
 $\{C, F\}, \{D, E\}, \{D, F\}, \{E, F\}$

Triple sets in negative border $\{A, B, C\}$

Total sets in negative border are 14.