CS412: An Introduction to Data Warehousing and Data Mining **Fall 2015**

Sample Question

Mingkun *Due Date: N/A*

Note: This is a sample question with the solution. It aims at helping you better understand the **Bottom-Up Computation** method.

Bottom-Up Computation

1. Consider a 2-D data array containing **two** dimensions A and B. The data contained in the array is as follows:

$$\begin{array}{ll} (a_0, b_0) : 1 & (a_0, b_1) : 1 \\ (a_1, b_0) : 1 & (a_1, b_1) : 1 \\ (a_2, b_0) : 1 & (a_2, b_1) : 1 \end{array}$$

   Now, we want to construct the iceberg cube for dimension A and B. If we set $mini\_support = 2$ with the order of $A \rightarrow B$, how many cells would be considered/computed? Draw the trace tree and you should report the number of cells to be considered/computed. For these cells, you should also list each cell with its count and report whether the cell is expansible in the **BUC** process.

Solution:

Trace Tree in Figure 1.

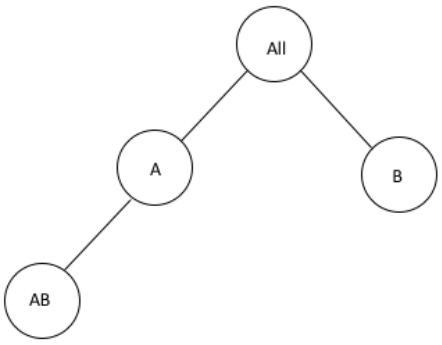

Figure 1: The trace tree.

If we follow the order of $A \rightarrow B$, the cells with their information (count, expansibility) that need to be computed are listed as follows:

All $(*, *)$: 6 - expansion

A $(a_0, *)$: 2 - expansion

A $(a_1, *)$: 2 - expansion

A $(a_2, *)$: 2 - expansion

AB (a_0, b_0) : 1

AB (a_0, b_1) : 1

AB (a_1, b_0) : 1

AB (a_1, b_1) : 1

AB (a_2, b_0) : 1

AB (a_2, b_1) : 1

B $(*, b_0)$: 3

B $(*, b_1)$: 3

Thus, there are totally 12 cells which would have to be computed.