Data Cube Computation Practice

Nikki Fitzherbert 13848336

30/07/2020

Start-Up Task

Assume a base suboid of four dimensions (**student, course, semester, lecturer**), contains only two base cells:

1. (a1, d2, d3, d4)
2. (d1, d2, c3, d4), where a1 $\neq$ d1 and c3 $\neq$ d3.

The measure of the cube is **tally**.

Question 1

How many non-empty cuboids will a full data cube contain?

If there are no hierarchies associated with each dimension, then the number of cuboids for a n-dimensional cuboids is the same as the number of non-empty cuboids in a full data cube. Therefore, the number of non-empty cuboids in a full data cube is $2^n = 2^4 = 16$, where n is the number of dimensions.

Question 2

How many closed cells are in the full cube?

There are three (3) closed cells in the full cube - the two base cells and (*, d2, *, d4)

Recall and review quiz

Consider a data cube with four dimensions (**subject, lecturer, time** and **student**), and the measure **count**. Given the following four cells:

a = (CP5806, *, *, *, 480)
b = (CP5806, *, 2018, *, 120)
c = (CP5806, Ickjai Lee, 2018, *, 10)
d = (CP5806, Ickjai Lee, 2018, Michael Jordan, 2)

Question 1

Cell b is a _____?

2-D cell as exactly two (2) out of the four dimensions are not * (aggregated).

Question 2

which is the base cell?

d, which is a 4-D cell as it doesn't contain any aggregation.

Question 3

The cell **a** is _____ of the cell **c**.

Cell **a** is a 1-D cell and cell **c** is a 3-D cell, which means that cell **a** is higher up in the lattice of cuboids. Therefore, cell **a** is an **ancestor** of cell **c**. It cannot be a parent as there is a layer in the lattice separating the two cells.

Question 4

Which cell is the parent of cell **c**?

b, as it is a 2-D cell and **c** is a 3-D cell.

Question 5

If the minimum support for iceberg cube materialisation is set to 10, then is cell **c** materialised?

Yes, as the usual condition would be $having\ count(*) >= min\_sup$

Question 6

Is the cell **c** a closed cell?

No.

A cell is a closed cell if there is no descendent such that a *-value can be replaced by a non-* value and the descendent has the same measure value.

Now consider a base cuboid of 10 dimensions that contains only three base cells:

1. (a1, d2, d3, d4,. . . , d9, d10)
2. (d1, b2, d3, d4,. . . , d9, d10)
3. (d1, d2, c3, d4,. . . , d9, d10), where a1 $\neq$ d1, b2 $\neq$ d2 and c3 $\neq$ d3.

The measure of the cube is **count**.

Question 7

How many non-empty cuboids will a full data cube contain?

If there are no hierarchies associated with each dimension, then the number of cuboids for a n-dimensional cuboids is the same as the number of non-empty cuboids in a full data cube. Therefore, the number of non-empty cuboids in a full data cube is $2^n = 2^{10} = 1024$, where n is the number of dimensions.

Question 8

How many non-empty aggregate cells will an iceberg cube contain if the condition of the iceberg cube is $count \geq 2$?

Aggregate cells are any cell from a non-base cuboid. Those cells aggregate over one or more dimensions, which are denoted by * in the cell notation. Each cell generates $2n - 1$ non-empty aggregate cells, so in this case there are $3 \times (2 \times 10 - 1) = 3(2 \times 10) - 3) = 3069$ non-aggregate cells.

$2^{n-1} * (m + 1)^{n+1}$ cells in a lattice of cuboids