

Data Cube vs Nano Cube

Jorge Luis Huanca Mamani

Febrero 2021

1 Introducción

A medida que la computación toma mas protagonismo en la vida cotidiana, nacen problemas que requieren plantearse nuevas estructuras para dar una solución a dichos problemas, y el ámbito de las bases de datos y estudio de grandes volúmenes de datos no es esquivo, a lo largo de este ensayo se señalará el problema común que se tenía en el pasado, que era como se puede agrupar datos con características específicas, tambien mostraremos la solución que se obtuvo usando una estructura Data Cube y como este fue contrastado con la aparición de una nueva estructura como lo es el Nano Cube, tambien se mostrará una comparativa de ambas estructuras.

2 Desarrollo

Supongamos que se quiere modelar una base de datos referente a los dispositivos móviles representada por la siguiente tabla:

Devise	Spacial I	Spacial II
Android	01	0011
Android	01	0111
iPhone	01	0010
iPhone	01	0110
iPhone	10	1001
iPhone	10	1101
iPhone	10	1000
iPhone	10	1100
Android	11	1011
Android	11	1111
Android	11	1010
Android	11	1110
Android	00	0001
Android	00	0101
iPhone	00	0000
iPhone	00	0100
ALL	ALL	ALL

Si analizamos la tabla anterior a detalle se puede observar que en cada columna de dicha tabla hay una cantidad de elementos diferentes que se repiten, en la primera columna tenemos 3 elementos que son Android, iPhone y ALL, en la segunda columna tenemos 5 elementos 00, 01, 10, 11. ALL y en la tercera columna tenemos 17 elementos diferentes, esta pequeña observación nos da una idea de la cantidad de elementos del data cube que estaría representada por $3 \times 5 \times 17 = 255$, dando como resultado una super tabla con 255 elementos que tendrá todas las combinaciones posibles.

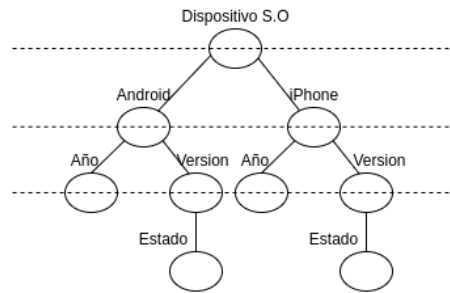
Devise	Spacial I	Spacial II
Android	01	0011
Android	01	0111
Android	ALL	---
Android	11	1011
Android	11	1111
Android	11	1010
Android	11	1110
Android	ALL	---
Android	00	0001
Android	00	0101
Android	ALL	---
iPhone	01	0010
iPhone	01	0110
iPhone	ALL	---
iPhone	10	1001
iPhone	10	1101
iPhone	10	1000
iPhone	10	1100
iPhone	ALL	---
iPhone	00	0000
iPhone	00	0100
iPhone	ALL	---



La tabla anterior muestra el inicio de la tabla con las posibles combinaciones, se debe tener en cuenta que esta tabla continua hasta llegar a la cantidad de 255 filas, dicha tabla de 255 filas es el data cube.

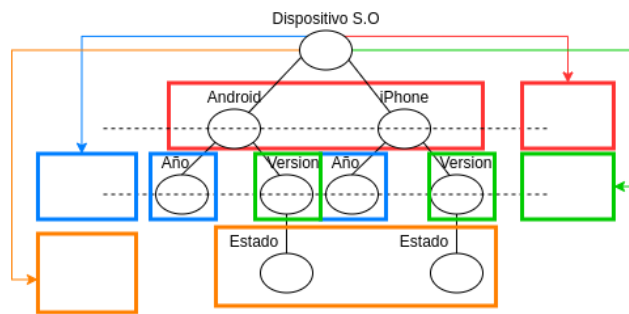
Ahora bien para poder construir un nano cube es necesario determinar una dimension espacial, categórica y una temporal para este ejemplo plantearemos dos espaciales y una categórica.

Para la dimensión categórica es necesario usar un arbol anidado de altura uno como se muestra en la siguiente imagen.



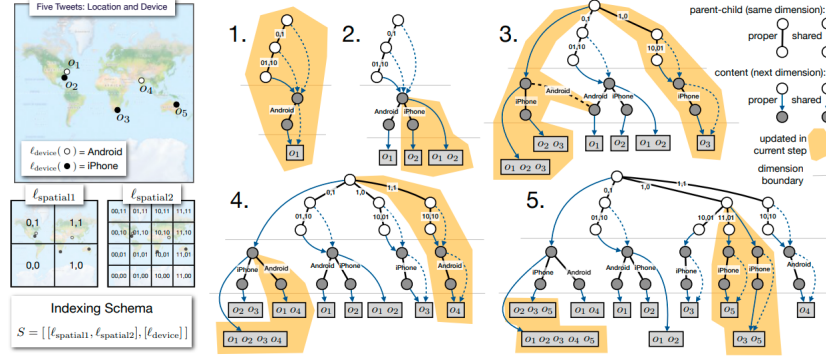
Como se puede observar este arbol anidado esta dividida en 3 niveles, cada nivel describe un arbol de altura 1 y este a su vez describe una serie de características.

A continuación se muestra el mismo arbol con una pequeña modificación.



Al ver el grafico anterior se puede notar que hay nodos agrupados por colores y estos estan siendo referenciados al nodo raiz, y esto se hace para poder acceder a una característica en concreto sin tener que hacer un recorrido desde la raíz hasta dicha característica, manteniendo una distancia de 1.

A continuación se muestra la inserción de los datos en un quadtree, la generalización de la optimización de la dimensión categórica del ejemplo propuesto.



3 Conclusiones

Como se pudo observar, en el data cube calcular la cantidad de combinaciones es costoso, también almacenar la tabla resultante es más pesado mientras que en el nano cube se hace un poco más de operaciones pero el espacio es más reducido.

En los nano cubes se pueden insertar datos de diferentes dimensiones como lo es el espacial, temporal y categórico mientras que en el data cube no.

En el nano cubes en su inicio y al momento de insertar datos tiene un crecimiento exponencial, pero mientras más datos se pone dicho crecimiento empieza a normalizarse mientras que en el data cube el crecimiento se mantienen de inicio a fin.

Ambos requiere de una función de agregación.