

CC42A – BASES DE DATOS
Profesores: Claudio Gutiérrez, Gonzalo Navarro
Auxiliar: Mauricio Monsalve

Auxiliar 6

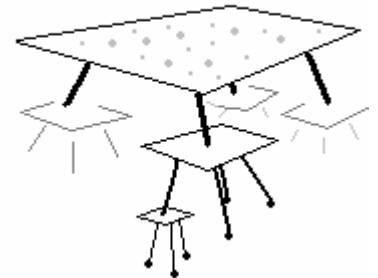
- Más sobre índices
- Bases de datos actuales
- ***ETAPA 4***

Más sobre índices

Por supuesto que existen muchos más métodos de indexación que B-Tree y los hashing. Existen muchas estructuras de datos que aprovechar.

R-Trees:

- Tienen la particularidad de ser como los B-Trees pero con ramas multidimensionales.
- R-Tree se utiliza para colocar información en dos dimensiones. Originalmente es una estructura de datos que puede almacenar información sobre puntos del plano cartesiano (\mathbb{R}^2). Cada nodo apunta a un subconjunto de la estructura, sucesivamente, hasta llegar a las hojas que poseen los puntos tales, sea aquí la información.
- Posee profundidad variable, todo depende de la densidad de “puntos” en el área del nodo padre.
- El desglosamiento de la información es mayor que B-Tree, pudiéndose aquí hacer distinción sobre dos tipos de propiedad a la vez (las dos dimensiones).
- En general el rendimiento es proporcional a B-Tree. Tiene el mismo orden asintótico. Pero es importante saber que entre sus muchas variantes hay una que tiene excelentes órdenes en el peor caso. Esta variante posee prioridad y se ha publicado este año.



Bitmaps:

- Método especialmente rápido y preciso en la búsqueda. Son realmente *mapas de bits* almacenando información sobre propiedades de las columnas de una tabla, por fila.
- Realiza operaciones lógicas AND y OR como operaciones de bits sobre el bitmap
- Lidia con la ineficiencia de B-Tree en lo que respecta al retorno de muchas filas. En este caso bitmaps ofrece una eficiencia enorme.
- Hay grandes problemas con la actualización. Como son mapas de bits hay que escoger propiedades sobre búsquedas recurrentes. Y al momento de actualizar la alternativa es reindexar toda la tabla en el bitmap. Esta alternativa, por ende, se debe ocupar en bases de datos relativamente estáticas.
- La reducción de datos es enorme. En una tabla con 4 millones de columnas, sólo se requieren 500kb para saber cuál es la de ingenieros, por ejemplo.

Bases de datos actuales

- Casi todas las bases de datos comerciales (Sybase, IBM <DB2>, Oracle, Tandem <SQL/MP>) soportan índices B+Tree, o un derivado cercano.
- Por temas de rendimiento, los índices se han optimizado para consultas AND y OR.
- Bitmaps han sido implementados en Oracle, Sybase, Red Brick, etc.

Etapa 4 – PRETAREA

- La etapa 4 consta de la creación de consultas SQL. Estas consultas tienen que ser complicadas en el sentido en que tienen que tener joins complejos. Los joins deben ser entre tablas distintas (tres o cuatro ojalá). En la entrega se debe especificar la consulta SQL, su significado y la relevancia en el universo del discurso de la base de datos.
- De la etapa 4 ustedes deberán escoger la consulta más complicada para usarla en la etapa 5.
- La finalidad de la etapa 5 es comparar el rendimiento entre una base de datos sin índices, con índices B-Tree y con índices Hashing. Se deberá iterar aquella consulta costosa miles de veces para distintos tamaños de la base de datos (distintos tamaños, de hasta 10.000 filas por tabla). Esta tarea es MUY costosa en tiempo, el muestreo puede tardar varios días continuados, así que es bueno prepararse para ella.