

BD Avanzada

Tema 2: Indexación y Asociación

Indexación y Asociación

Objetivo General:

- Al completar este tema el estudiante estará en la capacidad de manejar los conceptos sobre Indexación y Asociación.

Objetivos Específicos: Al completar este tema el estudiante deberá estar en la capacidad de discutir sobre los siguientes conceptos:

- Conocer la importancia del uso de índices en una base de datos.
- Crear índices usando SQL.

Indexación y Asociación

- **Tipos básicos de índices**
 - **Índices ordenados.** Están basados en una disposición ordenada de los valores
 - **Índices asociativos** (*hash indexes*). Están basados en una distribución uniforme de los valores a través de una serie de cajones (*buckets*)
 - El valor asignado a cada cajón está determinado por una función, llamada **función de asociación** (*hash function*)

Indexación y Asociación

- **Criterios valoración técnicas de indexación:**
 - **Tipos de acceso**
 - Tipos de acceso que se soportan eficazmente.
 - Podrían incluir la búsqueda de: (1) registros con un valor concreto en un atributo o (2) registros cuyos atributos contengan valores en un rango especificado
 - **Tiempo de acceso**
 - Tiempo que se tarda en buscar un determinado elemento de datos, o conjunto de elementos, usando la técnica en cuestión

Indexación y Asociación

- **Criterios valoración técnicas de indexación (Cont.):**
 - **Tiempo de inserción**
 - Tiempo empleado en insertar un nuevo elemento de datos
 - Este valor incluye el tiempo utilizado en buscar el lugar apropiado donde insertar el nuevo elemento de datos, así como el tiempo empleado en actualizar la estructura del índice
 - **Tiempo de borrado.** Tiempo empleado en borrar un elemento de datos
 - **Espacio adicional requerido.** Espacio adicional ocupado por la estructura del índice
- **Claves de búsqueda,** atributos o conjunto de atributos usados para buscar en un archivo

Indexación y Asociación

- **Índices ordenados**
 - Los registros en el archivo indexado pueden estar almacenados siguiendo un orden
- **Índice primario**
 - Índice cuya clave de búsqueda especifica el orden secuencial del archivo
 - También se llaman **índices con agrupación**
- **Índices secundarios o índices sin agrupación**
 - Índices cuyas claves de búsqueda especifican un orden diferente del orden secuencial del archivo
- **Archivos secuenciales indexados**
 - Archivos con índice primario según una clave de búsqueda

Indexación y Asociación

- **Registro índice o entrada del índice**
 - Consiste en un valor de la clave de búsqueda y punteros a uno o más registros con ese valor de la clave de búsqueda
- **Clases de índices ordenados**
 - **Índice denso.** Aparece un registro índice por cada valor de la clave de búsqueda en el archivo
 - **Índice disperso.** Sólo se crea un registro índice para algunos de los valores
- **Índices multinivel**
 - Índices con dos o más niveles

Indexación y Asociación

- **Los índices secundarios**

- Deben ser densos, con una entrada en el índice por cada valor de la clave de búsqueda, y un puntero a cada registro del archivo

- **Un índice primario**

- Puede ser disperso, almacenando sólo algunos de los valores de la clave de búsqueda, ya que siempre es posible encontrar registros con valores de la clave de búsqueda intermedios mediante un acceso secuencial a parte del archivo, como se describió antes

Indexación y Asociación

- **Archivos de índices de árbol B+**
 - Es la más extendida de las estructuras de índices que mantienen su eficiencia a pesar de la inserción y borrado de datos
 - Toma la forma de un **árbol equilibrado** donde los caminos de la raíz a cada hoja del árbol son de la misma longitud
 - Implica una degradación del rendimiento al insertar y al borrar, además de un espacio extra
 - El número de punteros de un nodo se llama *grado de salida* del nodo

Indexación y Asociación

- **Archivos de índices de árbol B+ (Cont.)**
 - La «B» en árbol B+ proviene del inglés *balanced* (equilibrado). Es esta propiedad de equilibrio de los árboles B+ la que asegura un buen rendimiento para las búsquedas, inserciones y borrados
 - En la **organización de archivo con árboles B+**, los nodos hoja del árbol almacenan registros, en lugar de almacenar punteros a registros

Indexación y Asociación

- **Técnica de asociación**
 - Permite evitar el acceso a la estructura de índice
 - Proporciona una forma de construir índices
- **Organización de archivos por asociación**
 - Se obtiene la dirección del bloque de disco que contiene el registro deseado mediante el cálculo directo de una función sobre el valor de la clave búsqueda del registro
- **Cajón (*bucket*)**
 - Unidad de almacenamiento que puede guardar uno o más registros

Indexación y Asociación

- **Función de asociación**
 - Al asignar los valores de las claves de búsqueda a los cajones se debe cumplir lo siguiente:
 - **Distribución *uniforme***. Cada cajón tiene asignado el mismo número de valores de la clave de búsqueda dentro del conjunto de *todos* los valores posibles de la clave de búsqueda
 - **Distribución *aleatoria***. Cada cajón tendrá casi el mismo número de valores asignados a él, sin tener en cuenta la distribución actual de los valores de la clave de búsqueda
- **Desbordamiento de cajones**
 - Si el cajón no tiene suficiente espacio

Indexación y Asociación

- **Razones por desbordamientos de cajones:**
 - **Cajones insuficientes**
 - **Atasco.** Algunos cajones tienen asignados más registros que otros, por lo que un cajón se podría desbordar incluso cuando otros cajones tienen todavía espacio
- **Cadena de desbordamiento**
 - El tratamiento del desbordamiento utilizando una lista enlazada

Indexación y Asociación

- **Asociación abierta**
 - Se fija el conjunto de cajones y no hay cadenas de desbordamiento
- **Ensayo lineal**
 - Política en la que se utiliza el siguiente cajón (en orden cíclico) que tenga espacio

Indexación y Asociación

- **Índice asociativo** (*hash index*)
 - Organiza las claves de búsqueda, con sus punteros asociados, dentro de una estructura de archivo asociativo
 - Se construye:
 1. Aplicando una función de asociación sobre la clave de búsqueda para identificar un cajón
 2. Almacenando la clave y los punteros asociados en el cajón

Indexación y Asociación

- **Asociación dinámica**
 - Permiten modificar la función de asociación dinámicamente para acomodarse al aumento o disminución de la base de datos
- **Asociación lineal**
 - Evita el nivel extra de indirección asociado con la asociación extensible y el posible costo de más cajones de desbordamiento

Indexación y Asociación

- **Comparación de la indexación ordenada y la asociación**
 - *Indexación ordenada*, a menos que se sepa de antemano que las consultas sobre un rango de valores van a ser poco frecuentes, en cuyo caso se utiliza la asociación
- *Las organizaciones asociativas* son útiles para archivos temporales creados durante el procesamiento de consultas, siempre que se realicen búsquedas basadas en un valor de la clave, pero no consultas de rangos

Indexación y Asociación

- **Definición de índices en SQL**

- Un índice se crea mediante la orden **create index**, la cual tiene la forma

```
create index <nombre-índice> on <nombre-relación>  
(<lista-atributos>)
```

- El nombre de índice especificado con el índice se necesita para hacer posible la eliminación (*drop*) de índices. La orden **drop index** tiene la forma

```
drop index <nombre-índice>
```