

Indices: Introducción

Fuente

Oracle Database Express Edition

Database System Concepts, Silberschatz, Korth, Sudarshan

Fundamentos de Sistemas de Bases de Datos, Elmasri R., Navathe S.



Conceptos básicos

- Estructura de datos que mejora la velocidad de las operaciones de consulta
- Permite rápido acceso a registros de una tabla en una base de datos
- **Organización de información**
- Acelera la recuperación de datos en las tablas de una



Tipos básicos de índices

a) Índices ordenados

Basados en una disposición ordenada de los valores.

b) Índices asociativos

Se basan en la distribución uniforme de los valores a través de una serie de cajones (*buckets*). El valor asignado a cada cajón está determinado por una función de asociación.

(hash function)



Indices Ordenados

Los registros se almacenan ordenados por el valor de la clave de búsqueda.

Indices primarios: Es un archivo ordenado secuencialmente y cuya clave de búsqueda especifica el orden secuencial del archivo.

La clave de búsqueda de un índice primario suele ser la clave primaria, aunque no necesariamente.



Indices Ordenados

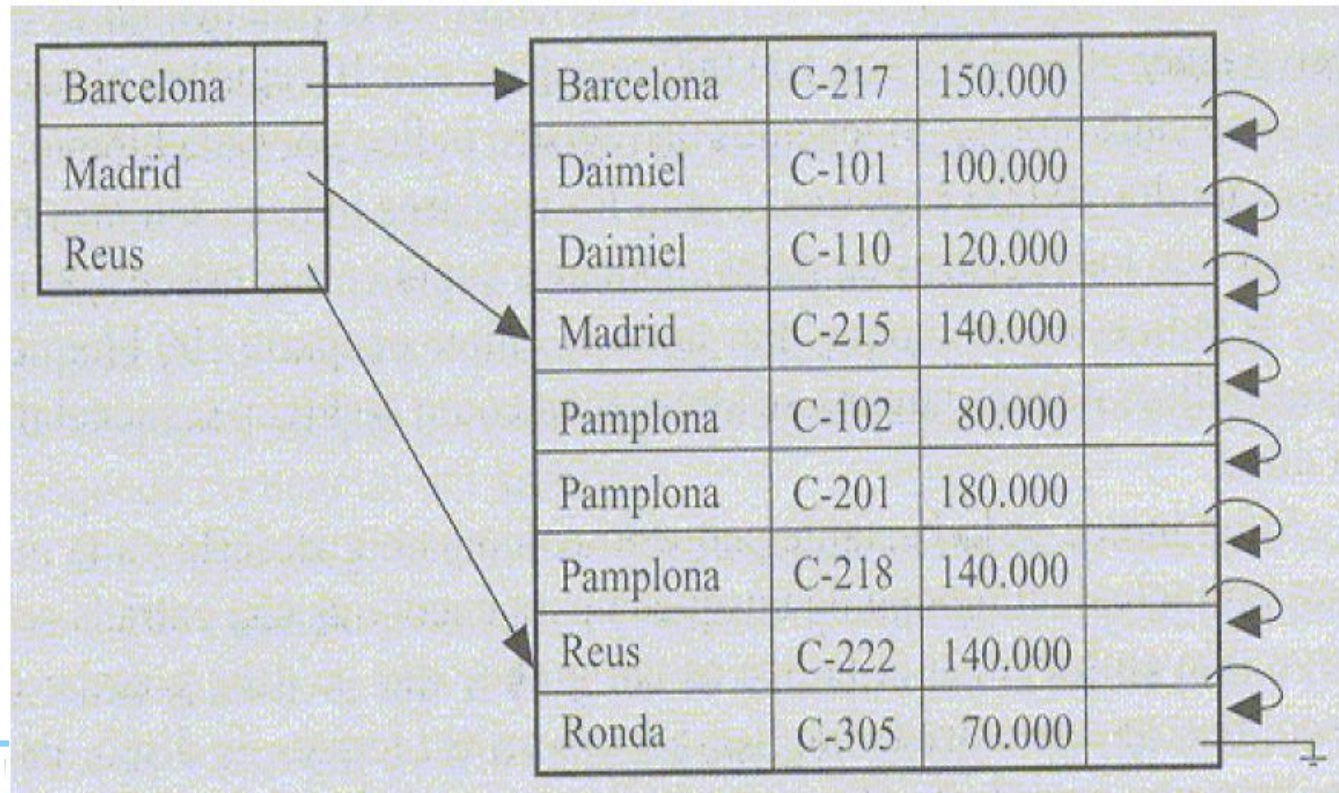
Índice secundario: es un índice cuya clave de búsqueda especifica un orden distinto del orden secuencial del archivo.

Archivo Secuencial Indexado: archivos ordenados secuencialmente con índice primario.



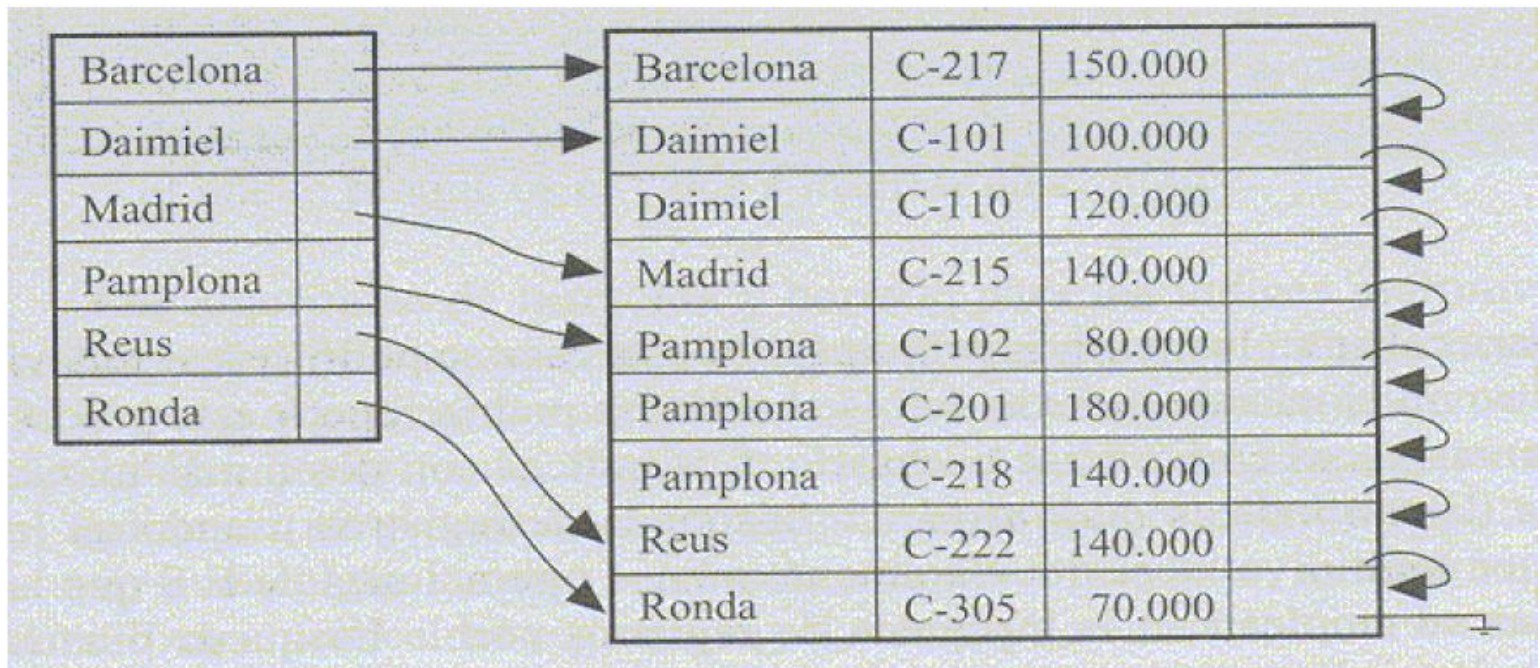
Índice disperso

Sólo se crea un registro índice para algunos de los valores de la clave de búsqueda



Indice denso

Aparece un registro índice para cada valor de la clave de búsqueda en el archivo



Indices B +

Es una estructura de datos que toma la forma de un árbol equilibrado. (Balanceado)

Los caminos de la raíz a cada hoja del árbol son de la misma longitud.

Cada nodo interno tiene entre $n/2$ y n hijos, donde n es un número fijo para cada árbol.

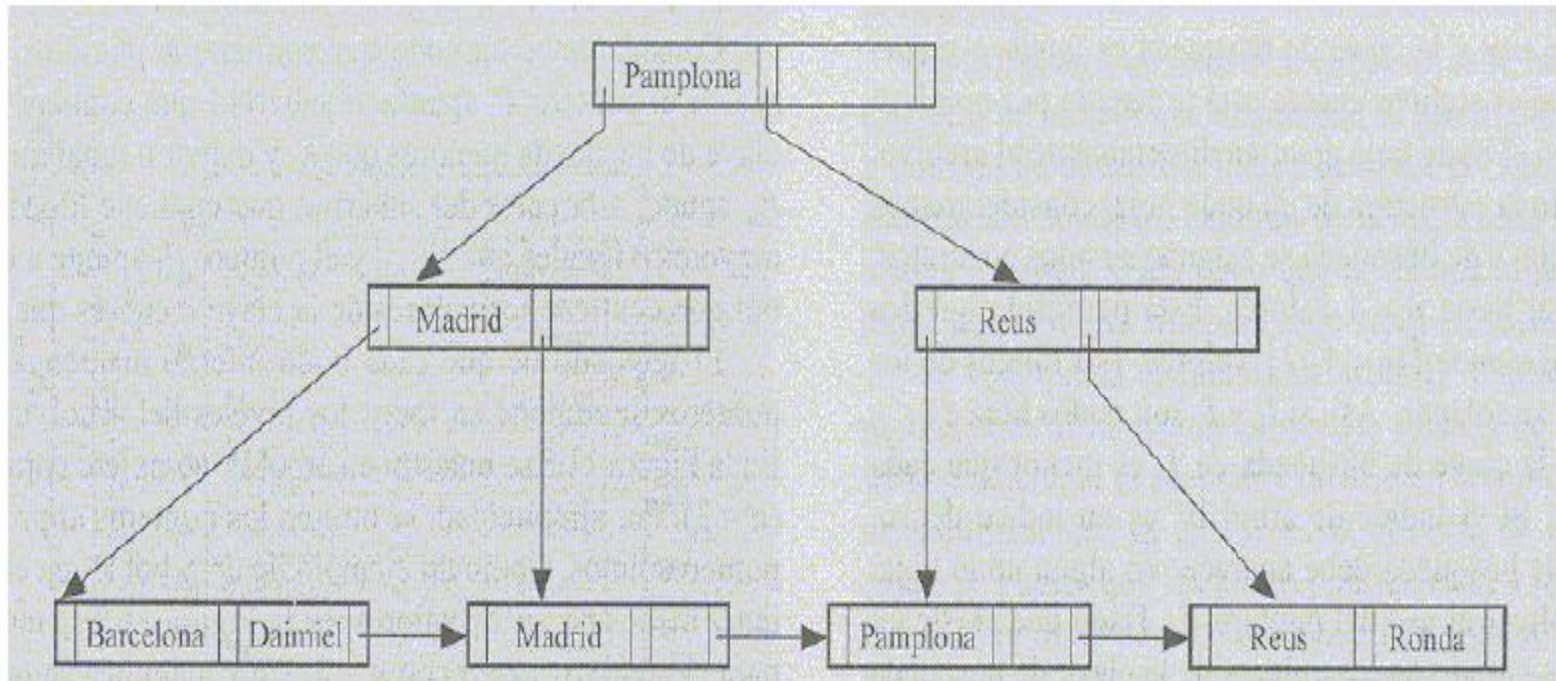
Puede contener hasta $n-1$ claves de búsqueda y n punteros

Cada hoja puede guardar hasta $n-1$ valores.



Indices B +

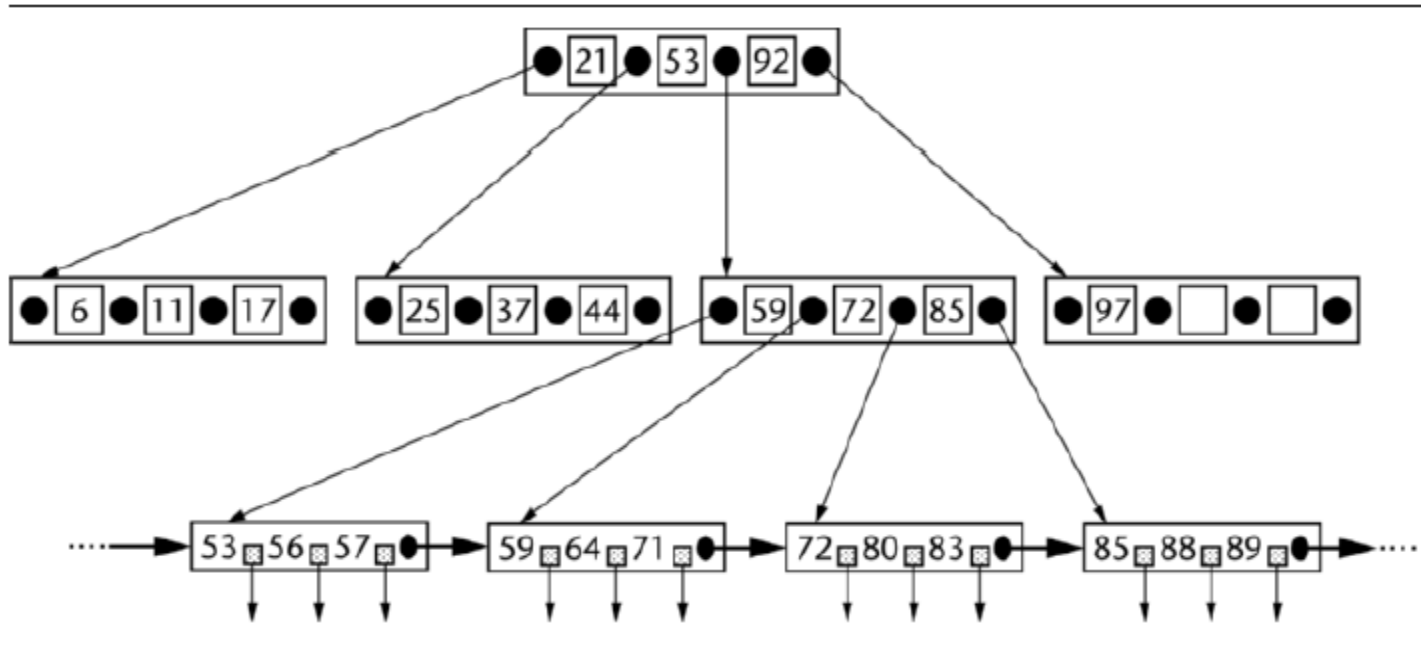
□ Indices B⁺



Árbol B+, con $n = 3$



Indices B +



Árbol B+, con $n = 4$



Indices B +

❑ Consultas en Indices B⁺

Recorrer un camino en el árbol desde la raíz hasta un nodo hoja .

Si hay K valores de la clave de búsqueda, este camino no será más largo que:

$$\text{Log}_{n/2} (k)$$



Indices B +

□ Ejemplo

Un nodo = tamaño de un bloque de disco = 4BK

Clave de búsqueda del tamaño de 32 bytes

Puntero a disco de 8 bytes

n está alrededor de 100

$K = 1000.000$ de valores de la clave de búsqueda

$\log_{50} (1000.000) = 4$ accesos a disco



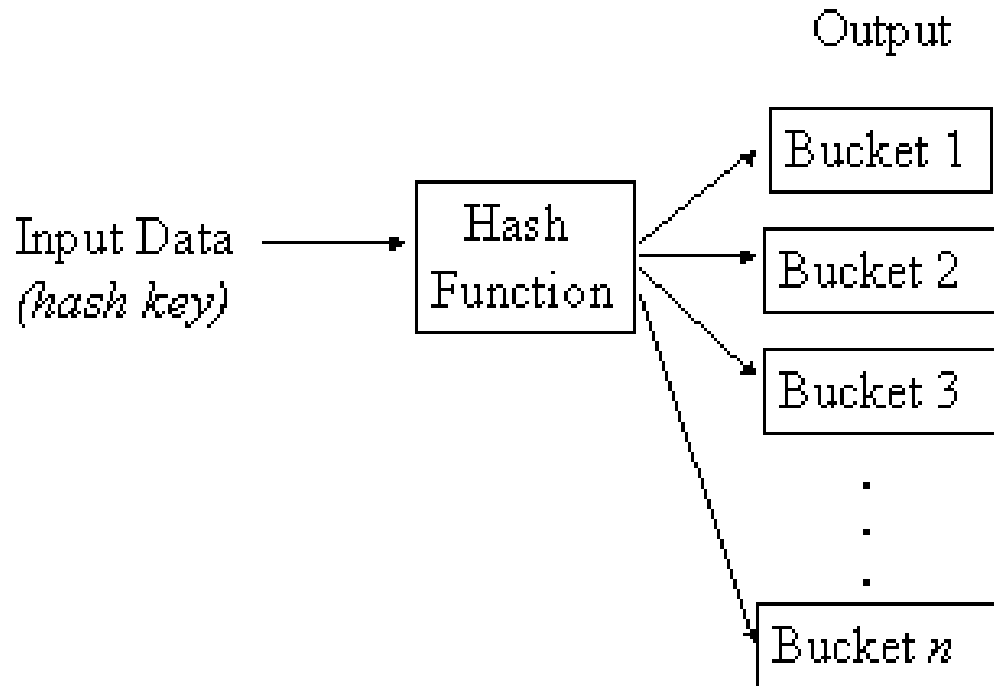
Indices B +

```
CREATE INDEX nombre_indice  
ON nombre_tabla  
USING B-TREE;
```



Indices Asociativos

❑ Funciones de asociación



Indices Asociativos

❑ Funciones de asociación

Distribución Uniforme: Se asigna a cada cajón el mismo numero de valores de la clave de búsqueda dentro del conjunto de todos los posibles valores de la misma.



Indices Asociativos

❑ Funciones de asociación

Distribución Aleatoria:

En el caso promedio cada cajón tendrá asignado casi el mismo número de valores, independientemente de la distribución real de los valores de la clave de búsqueda. El valor de la asociación no está correlacionado con ningún ordenamiento de los valores.



Indices Asociativos

❑ Funciones de asociación

- La peor función de asociación asigna todas las claves de búsqueda al mismo cajón; esto haría el tiempo de acceso proporcional al número de claves en el fichero.
- Una función de asociación ideal es uniforme, se asigna a cada cajón el mismo número de claves de búsqueda.
- Es aleatoria sí cada cajón tendrá el mismo número de registros asignados independientemente de la distribución actual de claves de búsqueda.



Indices Asociativos

Cajón 0

Cajón 1

C-215	
C-305	

Cajón 2

C-101	
C-110	

Cajón 3

C-217	
C-102	

C-201	

Cajón 4

C-218	

Cajón 5

Cajón 6

C-222	

Barcelona	C-217	150.000
Daimiel	C-101	100.000
Daimiel	C-110	120.000
Madrid	C-215	140.000
Pamplona	C-102	80.000
Pamplona	C-201	180.000
Pamplona	C-218	140.000
Reus	C-222	140.000
Ronda	C-305	70.000



Indices Asociativos

```
CREATE INDEX nombre_indice  
ON nombre_tabla  
USING hash;
```



Cómo elegir un índice?

- ❑ Tipos de acceso
- ❑ Tiempo de acceso
- ❑ Tiempo de borrado
- ❑ Espacio adicional requerido



Indices basados en los tipos de consulta

Las Consultas con un Operador de Igualdad

```
Select  $A_1, A_2, \dots, A_n$   
FROM  $R$  where  $A_i = c$ 
```

Usar un índice asociativo

Si se usara un índice ordenado el tiempo es proporcional al logaritmo para el número de valores para A_i en R . En una estructura asociativa es una constante que no depende del tamaño de la base de datos.

SILBERSCHATZ (2002)



Indices basados en los tipos de consulta

Las Consultas para un rango de datos

```
SELECT   $A_1, A_2, \dots, A_n$   
FROM    R  
WHERE     $A_i > c1$    and    $A_i < c2$ 
```

Usar un índice ordenado

En este tipo de consulta es mejor usar un índice ordenado ya que primero se hace una búsqueda del valor $c1$, luego se sigue la cadena de punteros en el índice en orden para leer el siguiente bloque y se continúa de esta manera hasta encontrar a $c2$. SILBERSCHATZ (2002)



Indices basados en los tipos de consulta

Ejemplo

Unidad de medida: Milisegundos (ms)

Número de registros por tabla: 570.000 aproximadamente



Indices basados en los tipos de consulta

Ejemplo

Consulta No. 1	Sin índice	PostgreSQL		Oracle
		B-tree	Hash table	B-tree
SELECT tituloprincipal FROM documentodigital WHERE codigo=12547;	457.3 0 ms	0.394 ms	0.097 ms	0.357 Ms



Indices basados en los tipos de consulta

Ejemplo

Consulta	Sin índice s	PostgreSQL		Oracle
		B-tree	Hash table	B-tree
SELECT * FROM documentodigital WHERE fechacreacion BETWEEN '2006-01-01 ' AND '2006-12-31 ' AND dcd_codigo <= 33 ORDER BY dcd_codigo;	543 .27 ms	57.33 ms	554.03 ms	55.27 ms



Indices basados en los tipos de consulta

Consulta	B-tree	Hash table
SELECT codigo FROM documentodigital WHERE tituloprincipal LIKE 'Bases de datos%'	0.256 ms	590.33 ms

```
Bitmap Heap Scan on material (cost=40.22..6474.52 rows=2870 width=4) (actual time=0.073  
  Filter: (mat_titulo_principal ~~ 'Bases de datos% '::text)  
    -> Bitmap Index Scan on name (cost=0.00..40.22 rows=2870 width=0) (actual time=0.050  
        Index Cond: ((mat_titulo_principal ~>= 'Bases de datos' '::text) AND (mat_titulo_  
xt))  
Total runtime: 0.256 ms
```



Indices basados en los tipos de consulta

Consulta	B-tree	Hash table
SELECT codigo FROM documentodigital WHERE tituloprincipal LIKE '%Bases de datos%'	0.256 ms	590.33 ms

```
Seq Scan on material (cost=0.00..17732.41 rows=1 width=4) (actual time=11.639..1082.3  
  Filter: (mat_titulo_principal ~~ '%Bases de datos% '::text)  
Total runtime: 1082.472 ms
```

