

Tema 8: Organización Secuencial Indexada

- 8.1 Introducción
 - 8.2 Estructura de la organización secuencial
 - 8.3 Ocupación del fichero secuencial
 - 8.4 Operaciones con los ficheros secuenciales
 - 8.5 Resumen
-

Contenidos extraídos del libro:

Ficheros. Organizaciones clásicas para el almacenamiento de la información

Autores: *Irene Luque Ruiz, Juan Antonio Romero del Castillo y Miguel Ángel Gómez-Nieto*
Editorial: Servicio de Publicaciones de la Universidad de Córdoba, 1998.
ISBN 84-7801-468-3

8.1 Introducción

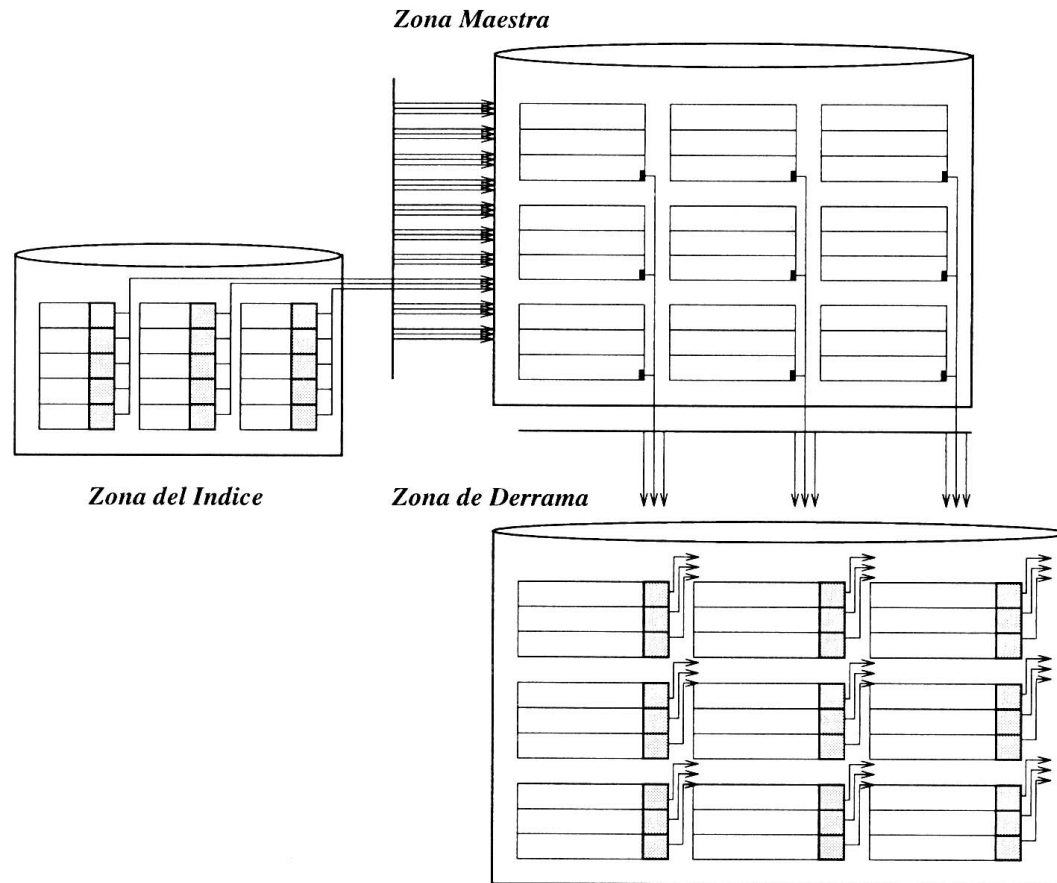
El fichero secuencial indexado:

- Las organizaciones anteriores son eficientes en accesos ordenados, pero no en accesos a un determinado registro.
- Necesitan de fuertes reorganizaciones.
- La organización secuencial indexada intentará paliar estos inconvenientes.
- Características:
 - Registros ordenados por una clave
 - Concepto de índice para acceder rápidamente a registros concretos
 - Dispone de: zona maestra + zona de índices + zona de derrama

A veces se da a esta organización el nombre ISAM: **I**ndexed **S**equential **A**ccess **M**ethod, inicialmente desarrollada por IBM.

8.1 Introducción

Las tres zonas del fichero secuencial indexado



8.2 Estructura

El índice

- **Definición:** estructura organizada para mejorar el acceso a la información en distintas organizaciones de ficheros
- Cada entrada del índice está compuesta por dos items: argumento + función
- **Argumento:**
 - ➔ Es un atributo o conjunto de atributos por el que los registros se encuentran ordenados.
 - ➔ Los predicados de búsqueda lo utilizarán como base de sus peticiones
- **Función:**
 - ➔ Es un indicador del registro en la zona maestra asociado con el argumento.
 - ➔ Generalmente se trata de un puntero.

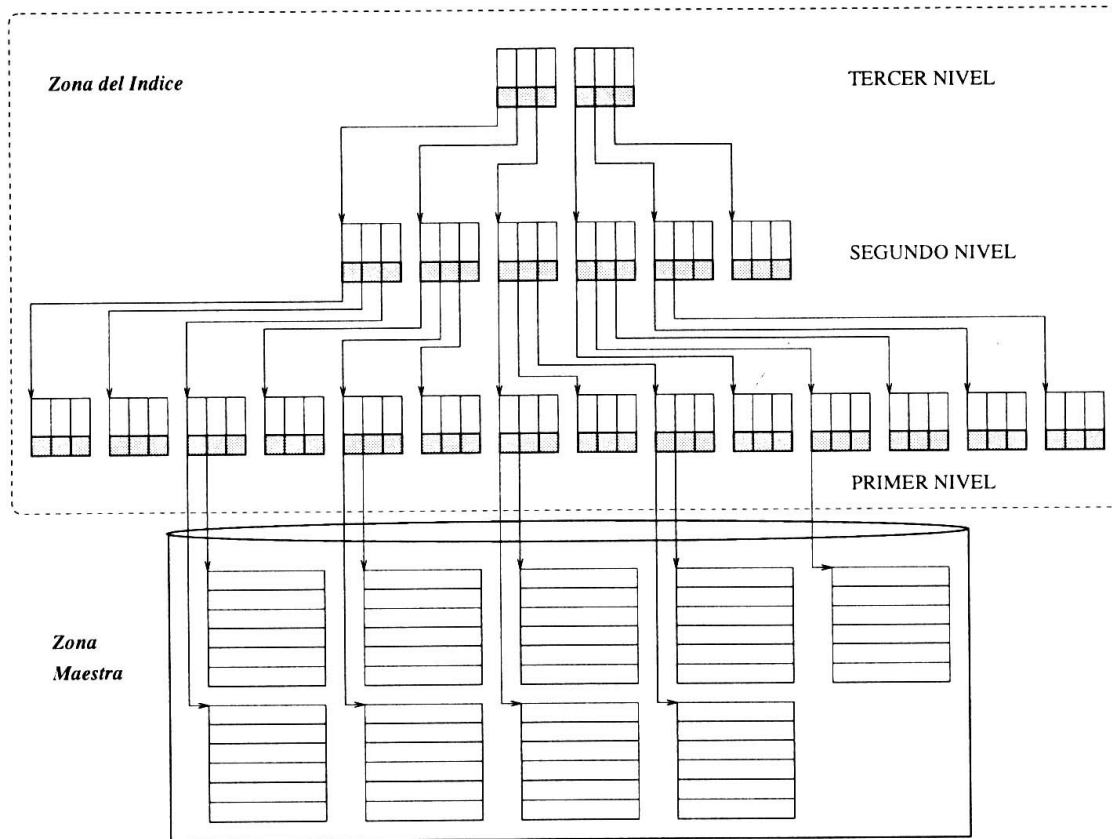
8.2 Estructura

Características del índice

- Se organiza en *árbol multivía* con dos tipos de niveles.
- El primer nivel (hojas del árbol) tiene tantas entradas como bloques tenga la zona maestra.
- Por ello los registros se encuentran ordenados en la zona maestra.
- Más bien se encuentran ordenados dentro de los bloques, los bloques no tienen por que estar ordenados entre sí.
- Si el índice ocupa más de un bloque, el primer nivel del índice es, a su vez, indexado añadiendo un segundo nivel al árbol.
- El proceso anterior se repite hasta que el último nivel ocupa solo un bloque

8.2 Estructura

El índice



8.2 Estructura

Razón de salida del índice

- Definición: número de registros o índices a los que se puede acceder con una entrada de un índice.

$$F_{B_I} = \left\lfloor \frac{B}{V + P} \right\rfloor$$

B : tamaño de un bloque

V : tamaño del argumento del índice

P : tamaño del puntero

concepto parecido al factor de bloqueo: F_B

8.2 Estructura

Ejemplo práctico: *El índice en la organización ISAM*

Consideremos un caso práctico en el que se desea organizar la información mediante una organización secuencial indexada y donde los valores de los parámetros de interés son los siguientes:

Tamaño de bloque $B=2048$ bytes

Tamaño de registro $R=128$ bytes

Tamaño del argumento del índice $V=26$ bytes

Tamaño del puntero del índice $P=6$ bytes

Número de registros en el fichero $r=1.600.000$

- *Suponer que no existe zona de derrama*
- *Suponer un emblocamiento fijo*

8.2 Estructura

Ejemplo práctico: *El índice en la organización ISAM*

$$F_B = \left\lfloor \frac{B}{R} \right\rfloor = \left\lfloor \frac{2048}{128} \right\rfloor = 16$$

$$b = \left\lceil \frac{r}{F_B} \right\rceil = \left\lceil \frac{1.600.000}{16} \right\rceil = 100.000 \text{ bloques}$$

$$I_M = b \times B = 100.000 \times 2048 = 204.800.000 \text{ bytes}$$

8.2 Estructura

Ejemplo práctico: *El índice en la organización ISAM*

Ocupación del índice:

- En el primer nivel harán falta tantas entradas como bloques en la zona maestra:

$$E_1 = b = 100.000 \text{ entradas}$$

- Cada entrada de índice almacena el argumento V y el puntero P:

$$I_1 = E_1 \times (V + P) = 1000.000 \times (26 + 6) = 3.200.000 \text{ bytes}$$

- El número de bloques del primer nivel, por tanto es:

$$b_1 = \left\lceil \frac{I_1}{B} \right\rceil = \left\lceil \frac{3.200.000}{2048} \right\rceil = 1563 \text{ bloques}$$

8.2 Estructura

Ejemplo práctico: *El índice en la organización ISAM*

Análogamente para el resto de niveles:

$$E_2 = b_1 = 1563 \text{ entradas}$$

$$\Gamma_2 = E_2 \times (V + P) = 1563 \times (26 + 6) = 50.016 \text{ bytes}$$

$$b_2 = \left\lceil \frac{\Gamma_2}{B} \right\rceil = \left\lceil \frac{50.016}{2048} \right\rceil = 25 \text{ bloques}$$

$$E_3 = b_2 = 25 \text{ entradas}$$

$$\Gamma_3 = E_3 \times (V + P) = 25 \times (26 + 6) = 800 \text{ bytes}$$

8.2 Estructura

Ejemplo práctico: *El índice en la organización ISAM*

- Razón de salida de los niveles 2 y 3 (señala a bloques de índices de niveles inferiores):

$$F_{B_I} = \left\lfloor \frac{B}{(V+P)} \right\rfloor = \left\lfloor \frac{2048}{(26+6)} \right\rfloor = 64$$

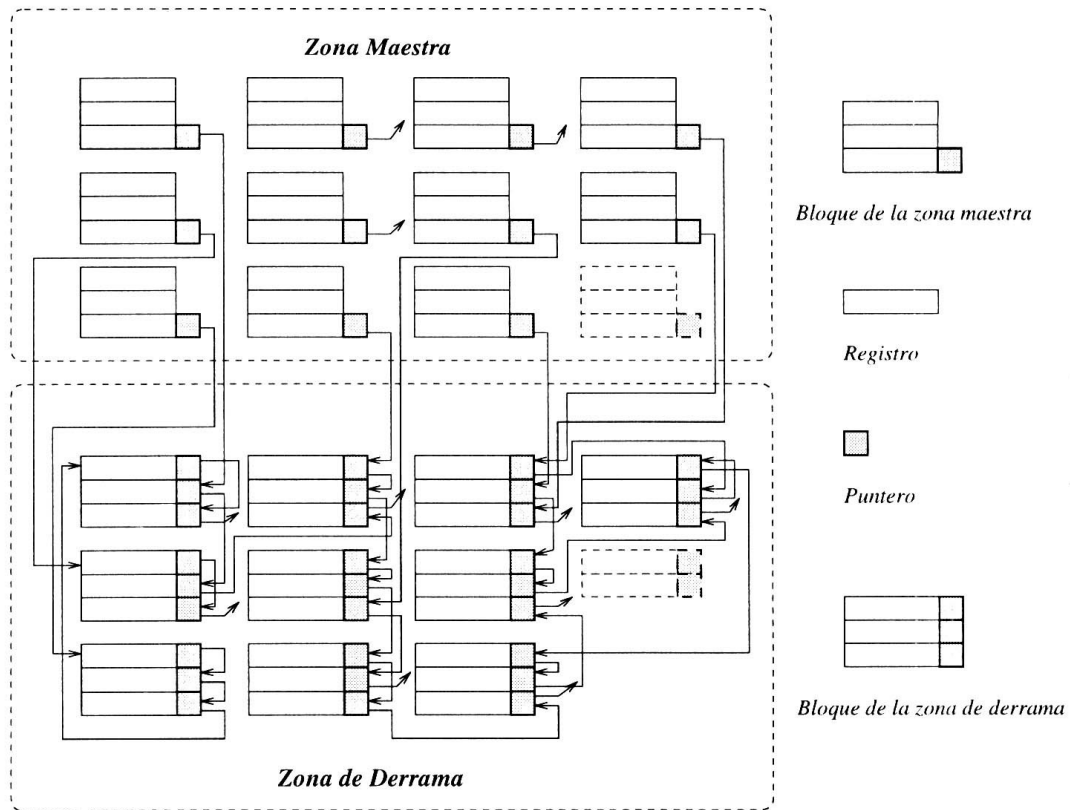
- Razón de salida del primer nivel (señala a bloques de registros en la zona maestra):

$$F_{B_I} = \left\lfloor \frac{B}{R} \right\rfloor = \left\lfloor \frac{2048}{128} \right\rfloor = 16$$

Generalmente no se necesitan más de tres niveles en el índice

8.2 Estructura

Zona maestra y zona de derrama



8.3 Ocupación

- Tamaño del registro de índice (o entrada de índice)

$$R_I = V + P$$

- Número de entradas del primer nivel del índice igual al número de bloques de la zona maestra:

$$E_{I_1} = b_M = \frac{r}{F_B} = \frac{r}{\left\lceil \frac{B-P}{R_M} \right\rceil}$$

- La ocupación del primer nivel del índice:

$$\Gamma_{I_1} = E_{I_1} \times R_I$$

8.3 Ocupación

- Número de entradas del segundo nivel:

$$E_{I_2} = b_{I_1} = \left\lceil \frac{E_{I_1}}{F_{B_I}} \right\rceil$$

$$F_{B_I} = \frac{B}{V + P}$$

- La ocupación del segundo nivel del índice:

$$\Gamma_{I_2} = E_{I_2} \times R_I$$

$$F_B = \left\lceil \frac{B - P}{R_M} \right\rceil$$

8.3 Ocupación

$$E_{I_j} = b_{I_{j-1}} = \left\lceil \frac{E_{I_{j-1}}}{F_{B_I}} \right\rceil$$

$$b_{I_j} = \left\lceil \frac{E_{I_j}}{F_{B_I}} \right\rceil = E_{I_{j+1}}$$

$$\begin{aligned} \Gamma_I &= (b_{I_1} + b_{I_2} + \dots + b_{I_{h-1}} + 1) \times B = \\ &= \Gamma_{I_1} + \Gamma_{I_2} + \dots + \Gamma_{I_{h-1}} + B \end{aligned}$$

8.3 Ocupación

- Ocupación total:

$$\Gamma = r_M \times R_M + r_D \times R_D + \Gamma_I$$

- Siendo el número de bloques de la zona de derrama:

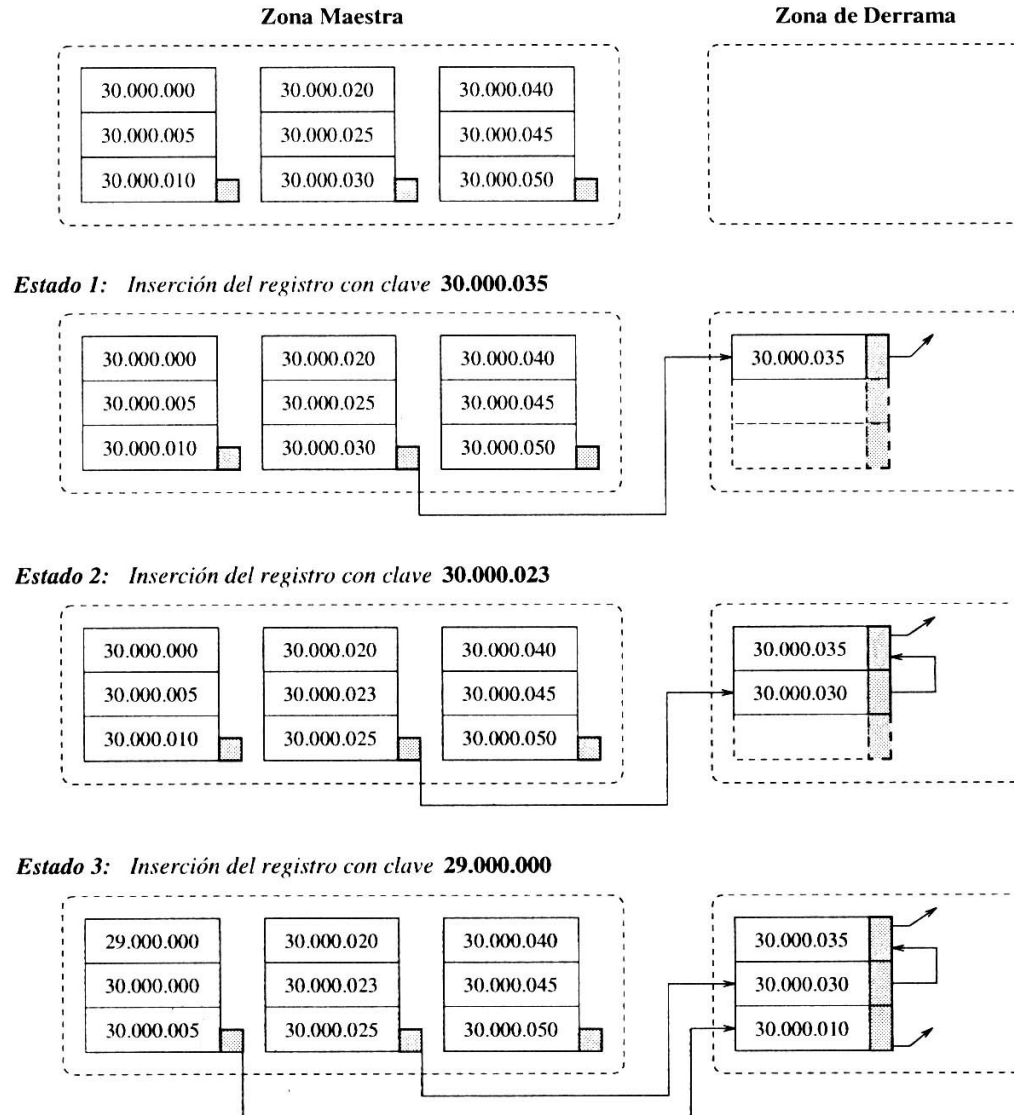
$$b_D = \frac{r_D}{\left\lceil \frac{B}{R_M + P} \right\rceil}$$

- Una medida de la ocupación total por registro sería:

$$\Gamma_R = \frac{\Gamma}{r_M + r_D} \text{ bytes}$$

8.4 Operaciones

Inserción: Normalmente los bloques de la Z. Maestra están llenos después de una reorganización:



8.4 Operaciones

Inserción:

El proceso incluye:

1. Evaluación del índice para ver el bloque de la zona maestra que le corresponde.
2. Si el bloque de la zona maestra está completo:
 - 1) Se localiza el bloque en la zona de derrama.
 - 2) Se reestablecen los punteros para seguir manteniendo el orden en la cadena.
3. Si la clave es anterior a la del primer registro hay que reestructurar la entrada del índice correspondiente y el primer bloque de la zona maestra.

8.4 Operaciones

Inserción: T_I

- Si hay que insertar en la zona de derrama

$$T_I = T_L + T_{RE} + t_r + T_{tB} + T_{RE}$$

T_L : localizar bloque donde irá el nuevo registro

T_{RE} : reescribir ese bloque con el nuevo orden

t_r : ir a derrama (en el mismo cilindro)

T_{tB} : traer el bloque de derrama a memoria

T_{RE} : reescribir el bloque de derrama

- Si se realiza en la zona maestra

$$T_I = T_L + T_{RE}$$

8.4 Operaciones

Lectura en la zona maestra: T_L

$$T_{LM} = t_l + (h - 1)(t_r + T_{tB}) + t_r + T_{tB} + T_P$$

1. Ir al cilindro donde está el índice
2. Recorrer (h-1) niveles hasta llegar al primer nivel (la raíz se supone que está siempre en memoria desde que se abre el fichero)
3. Ir con el puntero a la zona maestra
4. Traer el bloque de la zona maestra a memoria
5. Tiempo de procesamiento

8.4 Operaciones

Lectura en la zona de derrama: T_L

Acceso a la cabeza de la cadena en la zona de derrama

$$T_{LD1} = t_l + t_r + T_{tB} + T_P$$

Recorrido de la cadena en busca del registro

$$T_{LD2} = \frac{C}{2} (t_r + T_{tB}) + T_P$$

8.4 Operaciones

Tiempo Total de Lectura: T_L

$$T_L = T_{LM} + P_D (T_{LD1} + T_{LD2})$$

P_D : probabilidad de que sea necesario ir a la zona de derrama

8.4 Operaciones

Lectura Consecutiva: T_{LC}

Casos posibles:

- Si el último registro accedido está en el bloque de memoria y el siguiente registro también, se lee directamente.
- Si el último registro accedido era el último del bloque en memoria y no existe cadena de derrama, acceder al siguiente bloque en la zona maestra a través del índice.
- Si el último registro accedido era el último del bloque en memoria y existe cadena de derrama, buscar en dicha cadena el registro.
- Si el último registro accedido está en derrama y el siguiente también (en el mismo o en otro bloque), leer directamente o traer siguiente bloque de derrama.
- Existen muchos casos. El término promedio es:

$$T_{LC} \approx \left(\frac{1 - P_D}{F_B} + P_D \right) (t_r + T_{tB})$$

8.4 Operaciones

Actualización: T_A

- Si no se modifica la clave:

$$T_A = T_L + T_{RE}$$

- Si se modifica la clave:

$$T_A = T_L + T_{RE} + T_I$$

8.4 Operaciones

Lectura Exhaustiva (total): T_{LT}

Supone leer sistemáticamente todos los datos sin importar el orden

$$T_{LT} = b_M T_{tB} + b_D T_{tB}$$

8.4 Operaciones

Lectura Ordenada: T_{LO}

- Leer todos los datos en orden
- En este caso ordenados por la clave del fichero
- Basamos esta operación en la Lectura Consecutiva

$$T_{LO} = T_L + (r_M + r_D - 1)T_{LC}$$

8.4 Operaciones

Reorganización: T_{RO}

- Las operaciones de actualización y borrado generan registros marcados para borrado
- Volatilidad de los datos: cuando los datos almacenados dejan de ser útiles debido a una operación (borrado, actualización, etc.)
- La reorganización debe restituir el fichero a su estado óptimo:
 - ❑ Liberación de espacio no utilizado de los registros marcados para borrado
 - ❑ Eliminación de la zona de derrama e integración de sus registros en la zona maestra

8.4 Operaciones

Reorganización: T_{RO}

El proceso incluye:

1. Eliminar registros no significantes y el índice
2. La zona de derrama debe quedar vacía
3. Todos los registros quedan ordenados: ordenar zona de derrama y leer en paralelo junto con la secuencial y crear un único fichero con todos los registros ordenados
4. Borrar ficheros originales y renombrar el nuevo
5. Crear un nuevo índice con la zona maestra reorganizada

8.4 Operaciones

Reorganización: T_{RO}

$$T_{RO} = T_{LT} + T_C(r_M + r_D - r_B) + (r_M + r_D - r_B)T_{tR} + \frac{\Gamma_I}{t_t}$$

T_{LT} : leer todo el fichero en busca de marcados para borrado

T_C : ordenación de todo el fichero

r_B : número de registros marcado para borrado

$\frac{\Gamma_I}{t_t}$: estimación del tiempo empleado en crear el nuevo índice

8.5 Resumen

- Se usa cuando se necesita la información ordenada en base a una clave y a la vez accesos rápidos por la misma clave.
- Al estar los registros parcialmente ordenados en la zona maestra también se favorece la lectura secuencial ordenada.
- Su inconveniente es la zona de derrama que enlentece el resto de operaciones.
- Además, solo puede existir una clave.

Fin