

Temario



2.1 Estructuras Básicas de Archivos

2.2 Técnicas de Indexación

Técnicas de Indexación

- Índice: archivo usado para agilizar la recuperación de los registros, en respuesta a ciertas condiciones de búsqueda.
- Un índice...
 - Es redundante pues la información que guarda se encuentra presente en el archivo de datos asociado.
 - Agrega un mayor uso del espacio en disco...
 - ...que se compensa con posibilidad de encontrar los datos más rápidamente.

Técnicas de Indexación

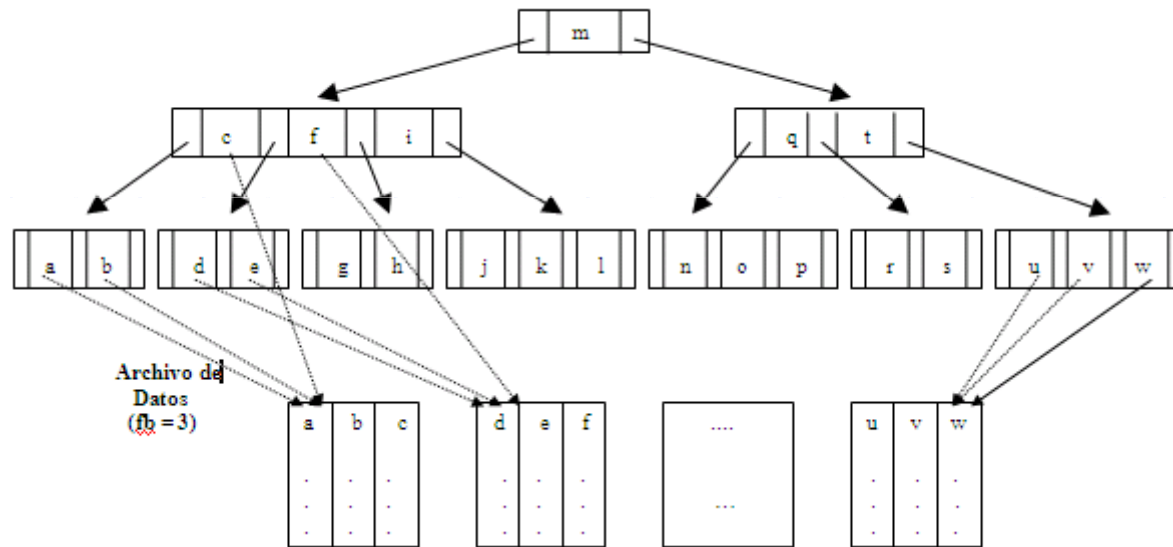
Tipos de Índices

- **Criterio:** Campo de Indexación.
 - Índice Primario (Principal): especificado sobre el campo clave de ordenamiento de un archivo ordenado de registros, o sobre el campo sobre el cual se construyó el archivo.
 - Índice de Grupos: construido sobre un campo de ordenamiento que abarca varios registros con el mismo valor, dentro de un archivo ordenado de registros.
 - Índice Secundario: especificado sobre un campo que no es usado para ordenar un archivo

Técnicas de Indexación

Tipos de Índices

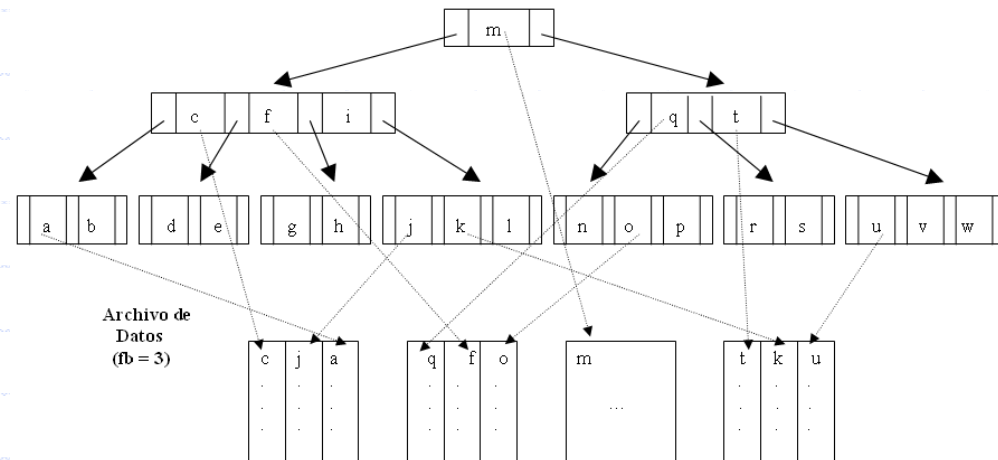
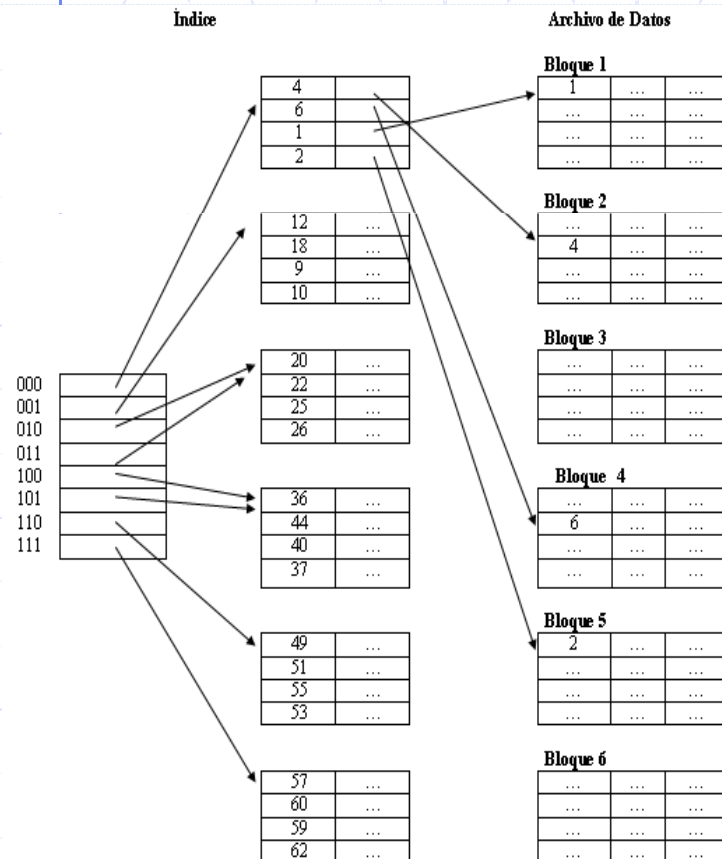
- Ejemplo de Índice Primario...



Técnicas de Indexación

Tipos de Índices

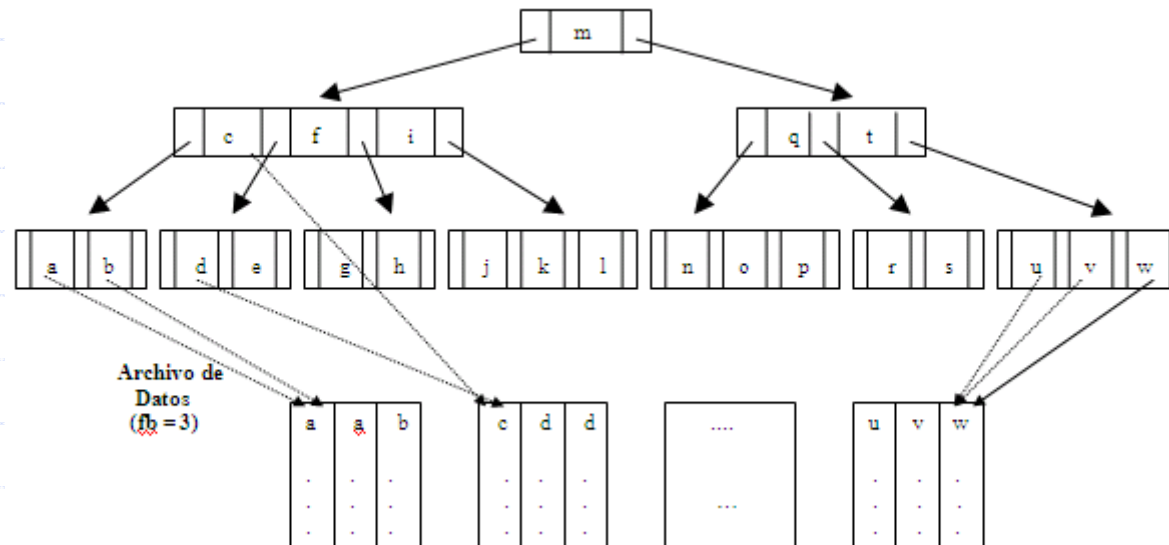
- Ejemplos de Índice Secundario...



Técnicas de Indexación

Tipos de Índices

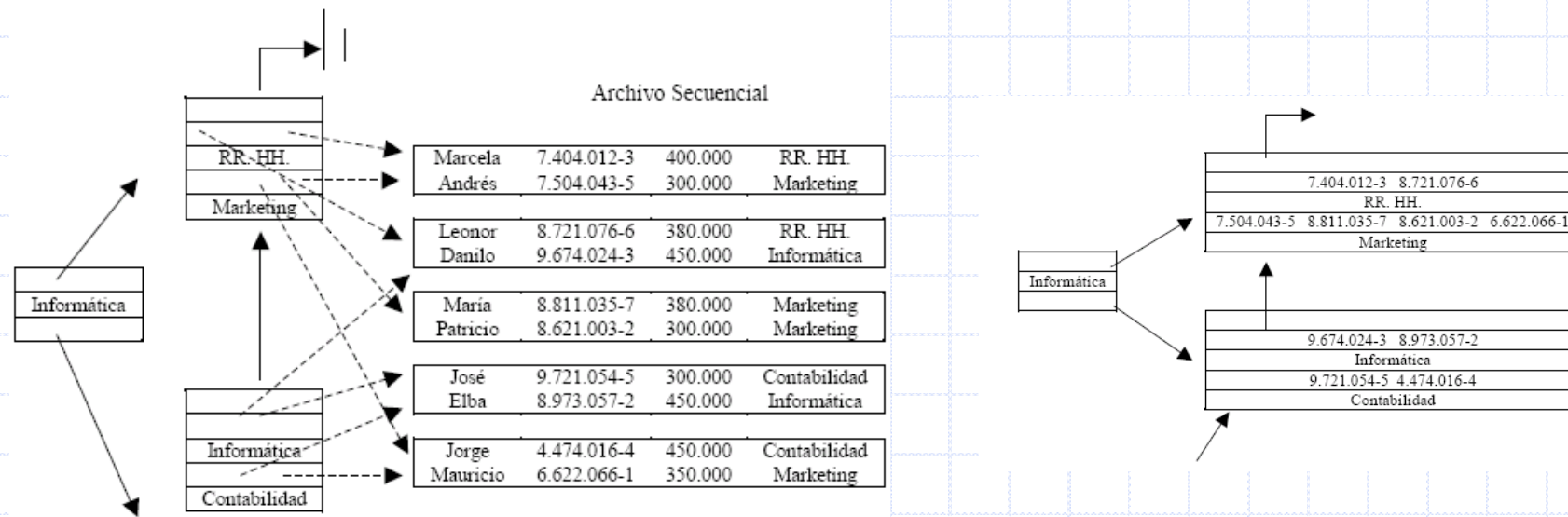
- **Criterio:** Número de Referencias.
 - Índice Denso: si tiene una entrada por cada registro.
 - Índice Disperso o no Denso: en caso contrario, tal como....



Técnicas de Indexación

Tipos de Índices

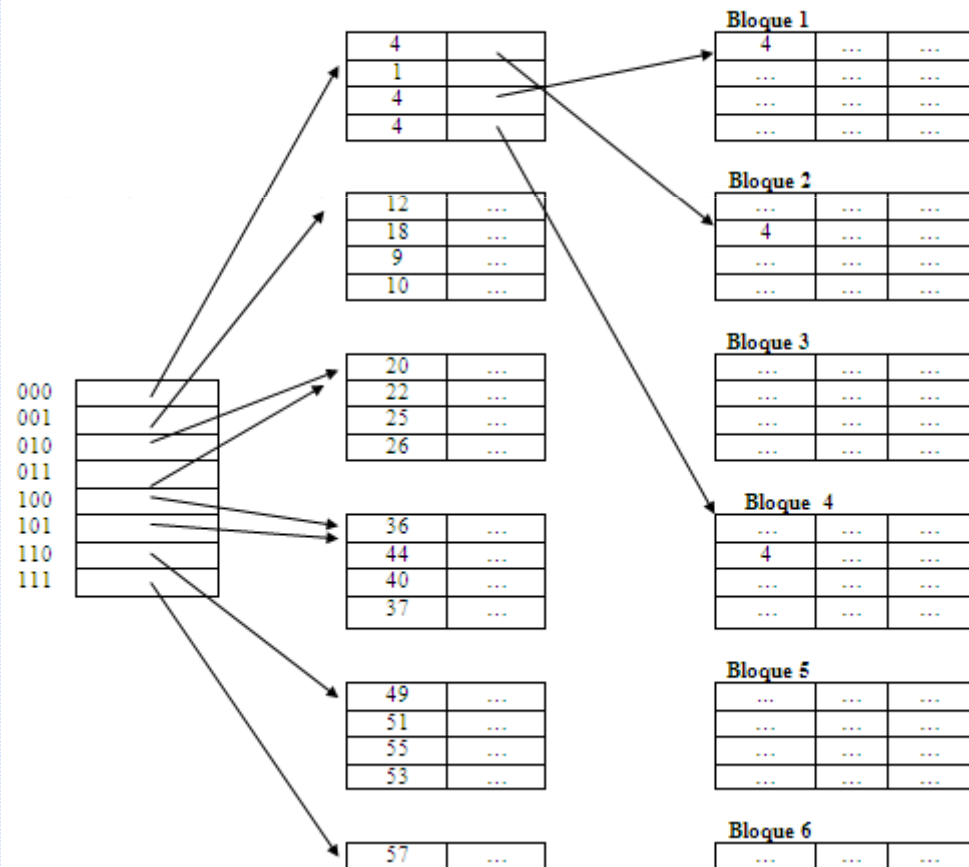
- **Criterio:** Tipo de Referencias.
 - Índice Físico: si la entrada contiene una referencia física (puntero) al archivo de datos.
 - Índice Lógico: si cada entrada del índice tiene como referencia el valor de otro campo.



Técnicas de Indexación

Tipos de Índices

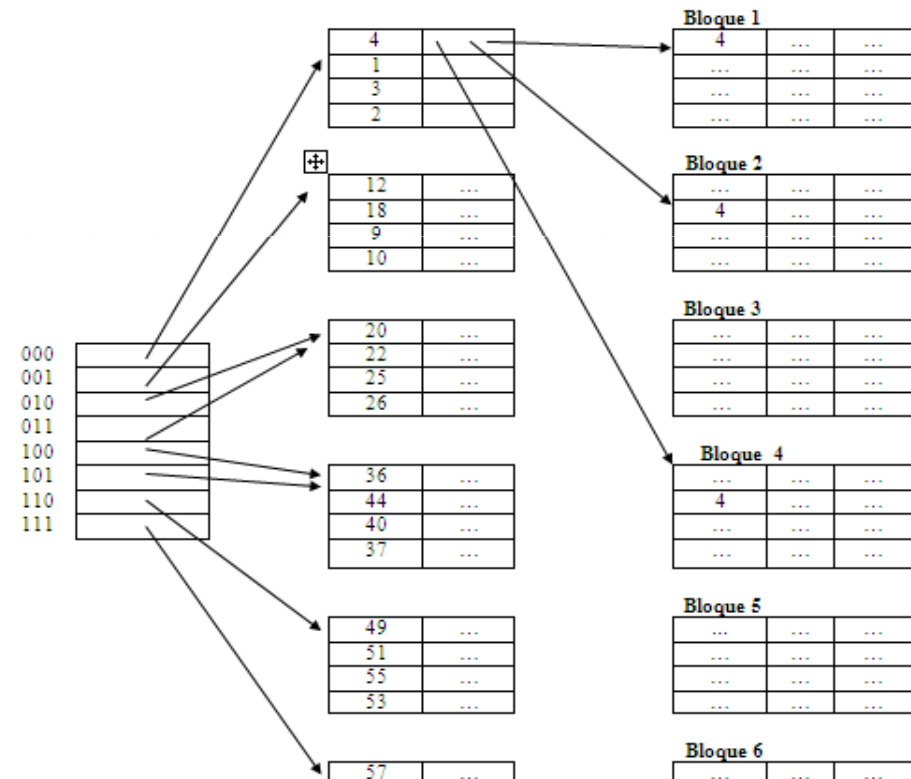
- **Criterio:** Estructura de las Referencias.
 - Tener una entrada de largo fijo, con un puntero al archivo de datos, ya sea como puntero a un bloque de disco o como puntero a un registro.



Técnicas de Indexación

Tipos de Índices

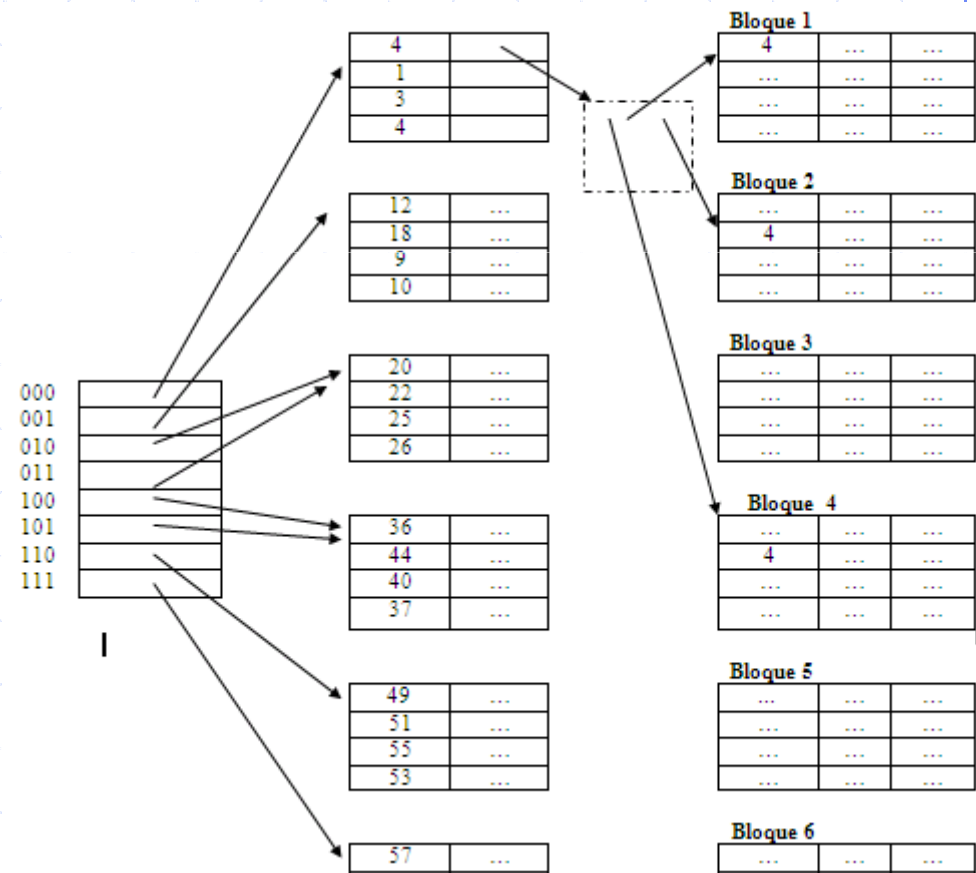
- **Criterio:** Estructura de las Referencias.
 - Tener entradas de largo variable, con un campo repetitivo para almacenar un puntero a cada bloque que contiene un registro. A este esquema se le acostumbra llamar *lista invertida*.



Técnicas de Indexación

Tipos de Índices

- **Criterio:** Estructura de las Referencias.
 - Una única entrada de largo fijo por cada valor del campo de indexación, con un nivel extra de indirectación para direccionar a los registros asociados al valor de dicha entrada.



Técnicas de Indexación

Tipos de Índices

- **Criterio:** Estructura de las Referencias.
 - Entradas de largo fijo y tener un *bitmap* asociado a cada una, con un bit por cada bloque del archivo de datos; el bitmap guarda un valor 1 en los bits de los bloques que contiene un registro con el valor de la entrada, y un valor 0 en caso contrario.

Técnicas de Indexación

Tipos de Índices

- Ejemplos de Índice Bitmap...

Columna a Indexar	B ₈	B ₇	B ₆	B ₅	B ₄	B ₃	B ₂	B ₁	B ₀
3	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	1	0	0
8	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	1
7	0	1	0	0	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0

Referencias del Índice

1	0011
2	0011
3	1101
4	1000
5	1100
6	0110

Archivo

...	4
...	5
...	3

...	6
...	5
...	3

...	2
...	1
...	6

...	3
...	1
...	2

Técnicas de Indexación

- Una de las formas de organizar jerárquicamente los datos de un archivo más recomendable resulta ser la estructura de árbol B y sus variaciones.
- Por las mismas ventajas descritas en la unidad anterior, este tipo de árbol puede ser usado también como estructura para índices.
- Similar situación se presenta con las técnicas *hashing*.

Técnicas de Indexación

Índices Dinámicos

- Índice de Tipo **Árbol B**:
 - Cumple las mismas propiedades vista en la unidad, cuando se estudió como estructura de archivo.
 - Un nodo B de orden M usado como índice se vería del siguiente modo...

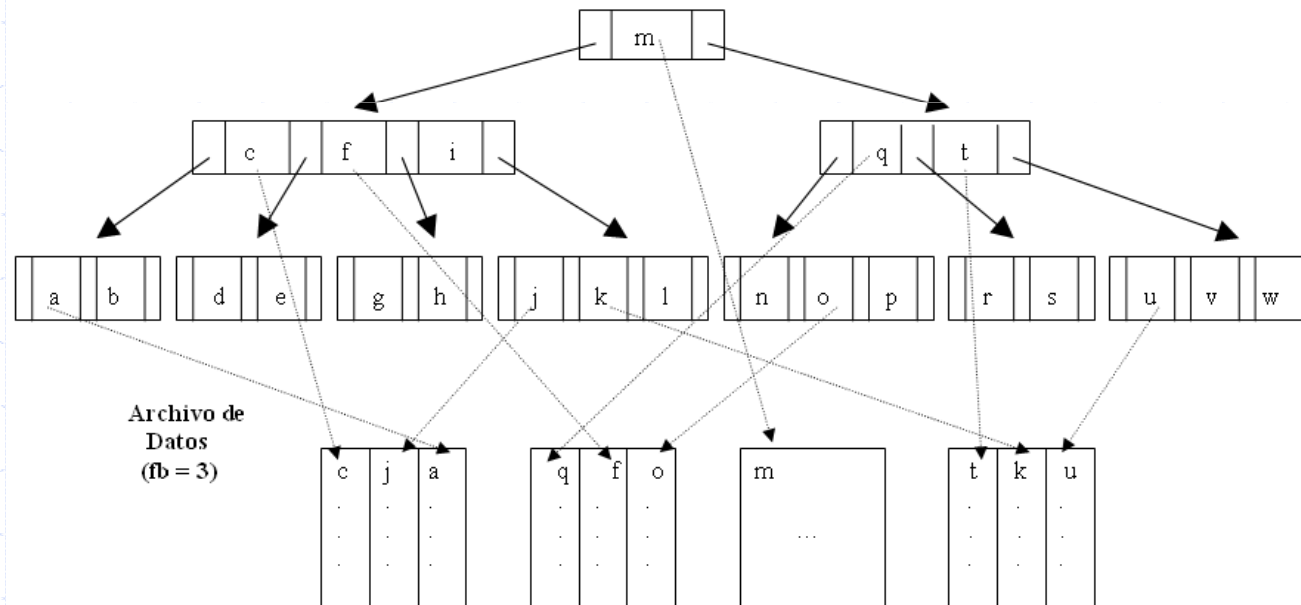
Puntero ₀	Clave de Registro ₁	Puntero ₂	Clave de Registro ₂	...	Puntero _{M-1}	Clave de Registro _{M-1}	Puntero _M
	Puntero Registro ₁		Puntero Registro ₂			Puntero Registro _{M-1}	

donde el puntero al registro_i representa la dirección al bloque donde está el registro correspondiente, o a la posición que éste ocupa dentro de un determinado bloque.

Técnicas de Indexación

Índices Dinámicos

- Ejemplo de Índice de tipo Árbol B (en este caso, índice secundario, de orden 5).



Técnicas de Indexación

Índices Dinámicos

- Variaciones del Árbol B:
 - Árbol B*: cada nodo debe estar lleno en al menos $2/3$ de su capacidad de almacenamiento.
 - Árbol B virtual: la raíz y otros nodos intermedios son mantenidos en la memoria principal, tras su primera lectura, evitando posteriores idas al disco.
 - Árbol B+...

Técnicas de Indexación

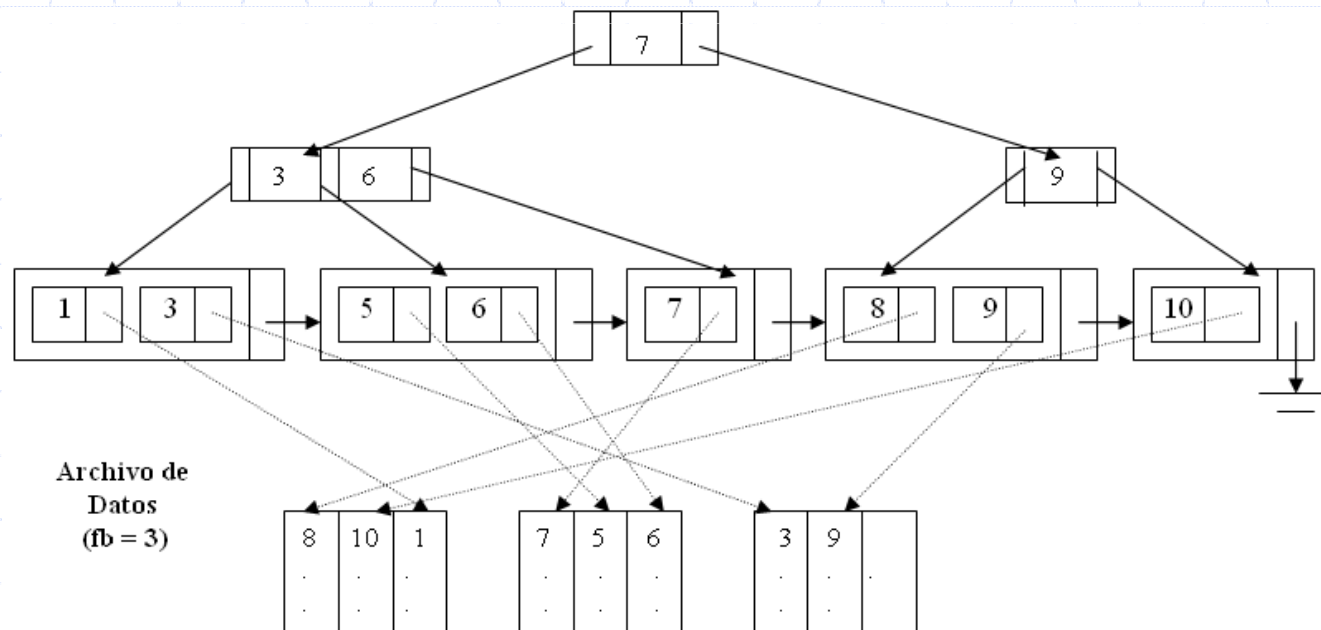
Índices Dinámicos

- Índice de tipo Árbol B+:
 - Sólo en las hojas se encuentran las referencias a los datos.
 - En los nodos internos se almacenan las claves que corresponden al mayor valor que se puede alcanzar a través del sub-árbol direccionado por el puntero izquierdo de dicha clave.
- Se puede ver como un árbol B con un nivel adicional (hojas), cuyos nodos componentes se encuentran enlazados entre sí, formando una lista enlazada.
- Lo anterior permite, además de la búsqueda binaria desde la raíz, un recorrido lineal de los nodos hojas para acceder los datos de forma ordenada en base al atributo de indexación.

Técnicas de Indexación

Índices Dinámicos

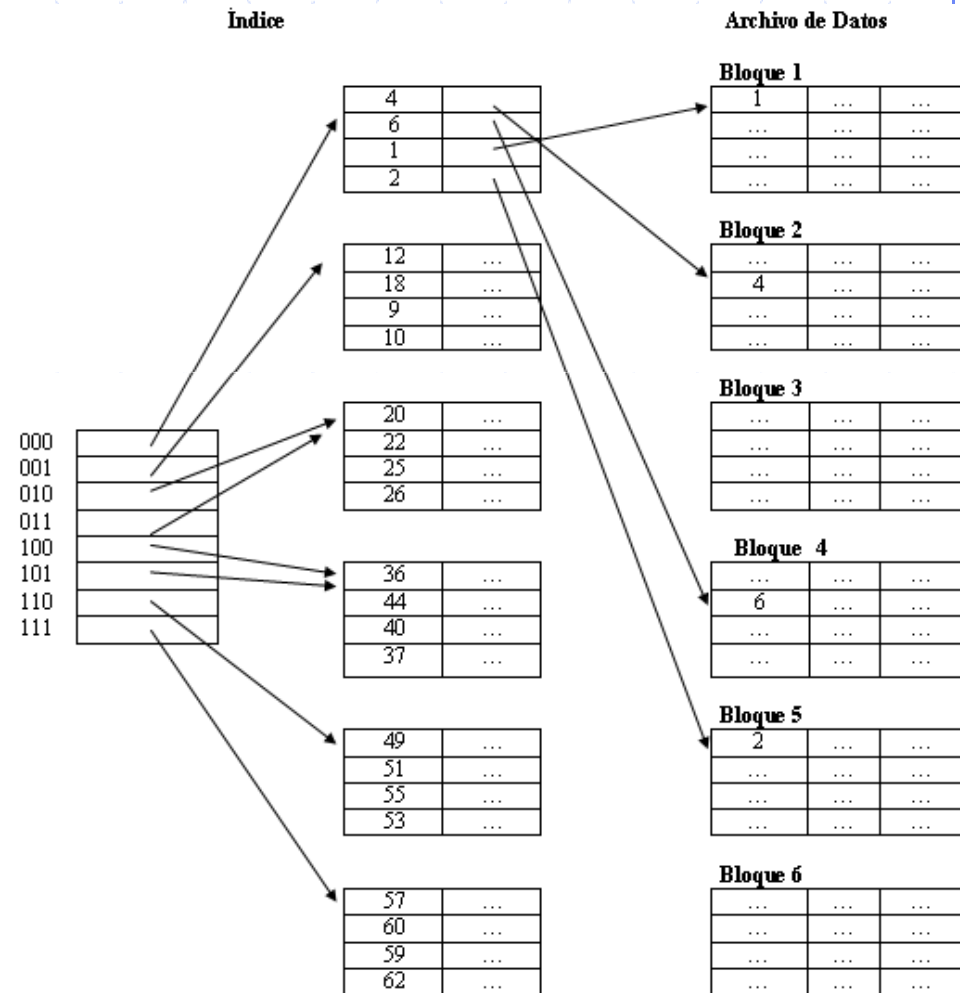
- Ejemplo de Índice de tipo Árbol B+ (en este caso, índice secundario, nodo interno de orden 5, nodo hoja de orden 2).



Técnicas de Indexación

Índices Dinámicos

- Todas las técnicas hashing vistas en la unidad anterior se pueden adaptar para ser utilizadas para construir índices.



Técnicas de Indexación

Índices Bitmap

- Los bitmaps combinados con las estructuras de datos anteriores, son una alternativa a las listas invertidas.
- Cada entrada del índice, maneja un bitmap en el cual cada bit representa un bloque de disco ocupado por el archivo de datos.

Referencias del Índice

1	0011
2	0011
3	1101
4	1000
5	1100
6	0110

Ar

Página 1

...	4	...
...	5	...
...	3	...

Página 2

...	6	...
...	5	...
...	3	...

Página 3

...	2	...
...	1	...
...	6	...

Página 4

...	3	...
...	1	...
...	2	...

Técnicas de Indexación

Índices Bitmap

- **Esquema de Codificación:** abarca dos posibilidades....por igualdad o por rangos. Este esquema se relaciona, con una representación que se preocupa por el tipo de consultas a responder.
- **Esquema de Descomposición:** en este caso, la representación se centra en el almacenamiento de los bits, pudiéndose manejar diversos sub-componentes para guardar un conjunto de valores mediante un bitmap.

Técnicas de Indexación

Índices Bitmap

- **Índice Valor-Lista de un Componente.**
 - Es más bien un conjunto de bitmap, uno por cada valor del dominio del atributo indexado.
 - Cada bitmap tiene una cantidad de bits equivalente al número de registros del archivo que se esté indexando.
 - Para determinar el contenido de cada bitmap:
 - Suponer que se tiene aquél asociado a un valor v dentro del dominio del atributo indexado.
 - El i -ésimo bit de dicho bitmap valdrá uno si el i -ésimo registro del archivo tiene el valor v en el campo de indexación.
 - En caso contrario, el bit valdrá cero.

Técnicas de Indexación

Índices Bitmap

- Ejemplo de Valor-Lista de un Componente, sobre un atributo con dominio [0..8].

Columna a
Indexar

	B ₈	B ₇	B ₆	B ₅	B ₄	B ₃	B ₂	B ₁	B ₀
3	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	1	0	0
8	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	1
7	0	1	0	0	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0

Técnicas de Indexación

Índices Bitmap

- **Índice Valor-Lista de N Componentes.**
 - En el caso anterior, existe el inconveniente de que el número de bitmaps es igual al tamaño del dominio del atributo de indexación.
 - Para disminuir la cantidad de memoria requerida, se puede utilizar una representación en alguna base numérica que requiera menos bits.

Técnicas de Indexación

Índices Bitmap

- Ejemplo de Valo-Lista de 2 Componentes, en base $\langle 3, 3 \rangle$, sobre el mismo dominio anterior.

		Segundo Componente			Primer Componente		
		B ₂	B ₁	B ₀	B ₂	B ₁	B ₀
3	$1 \times 3 + 0$	0	1	0	0	0	1
2	$0 \times 3 + 2$	0	0	1	1	0	0
1	$0 \times 3 + 1$	0	0	1	0	1	0
2	$0 \times 3 + 2$	0	0	1	1	0	0
8	$2 \times 3 + 2$	1	0	0	1	0	0
2	$0 \times 3 + 2$	0	0	1	1	0	0
2	$0 \times 3 + 2$	0	0	1	1	0	0
0	$0 \times 3 + 0$	0	0	1	0	0	1
7	$2 \times 3 + 1$	1	0	0	0	1	0
5	$1 \times 3 + 2$	0	1	0	1	0	0

Técnicas de Indexación

Índices Bitmap

- **Índice Codificado por Rangos de un Componente.**
 - En este esquema también hay b bitmaps, pero la representación del valor v tiene un valor cero en los v bitmaps “más a la derecha”, y valores uno en los restantes bitmaps.

Columna a Indexar		B ₇	B ₆	B ₅	B ₄	B ₃	B ₂	B ₁	B ₀
3		1	1	1	1	1	0	0	0
2		1	1	1	1	1	1	0	0
1		1	1	1	1	1	1	1	0
2		1	1	1	1	1	1	0	0
8		0	0	0	0	0	0	0	0
2		1	1	1	1	1	1	0	0
2		1	1	1	1	1	1	0	0
0		1	1	1	1	1	1	1	1
7		1	0	0	0	0	0	0	0
5		1	1	1	0	0	0	0	0

Técnicas de Indexación

Índices Bitmap

- Ejemplo de Índice Bitmap, codificado por rangos, de dos componentes (base $\langle 3,3 \rangle$), sobre un atributo con dominio $[0..8]$.

Columna a Indexar	Segundo Componente		Primer Componente	
	B ₁	B ₀	B ₁	B ₀
3	1	0	1	1
2	1	1	0	0
1	1	1	1	0
2	1	1	0	0
8	0	0	0	0
2	1	1	0	0
2	1	1	0	0
0	1	1	1	1
7	0	0	1	0
5	1	0	0	0

Técnicas de Indexación

Índices Bitmap

- **Índice Bit-Sliced.**

- Un índice *bit-sliced* de un atributo es una proyección de los bits que componen los valores de dicho atributo para el tipo llamado binario; o es un valor-lista con tantos componentes como dígitos tenga el dominio del atributo en cuestión.

Técnicas de Indexación

Índices Bitmap

- Índice *Bit-Sliced* Binario: para un atributo numérico (entero), de dos bytes de representación interna.

[illegible]

- Índice *Bit-Sliced* en Base 10: para un atributo numérico (entero), de tres dígitos.

Archivo	Índice Bit-Sliced																
	Tercer Componente							Segundo Componente					Primer Componente				
A	B ₉	...	B ₈	B ₃	B ₂	B ₁	B ₀	B ₉	...	B ₂	B ₁	B ₀	B ₉	...	B ₂	B ₁	B ₀
201	0	...	0	0	1	0	0	0	...	0	0	1	0	...	0	1	0
...

Técnicas de Indexación

Índices Bitmap

- **Índices Bitmap Decodificados:**
 - Este esquema hace uso de funciones de codificación, o mapping, sobre el dominio del atributo y construye un índice bit-sliced binario sobre el resultado de la función.
 - Ejemplo:

Valor	Resultado
1	0000
2	0001
5	0010
9	0011

Notar que la idea de fondo de este esquema es poder traducir el dominio no continuo de un atributo a una representación que consuma menos memoria (utilizar una representación de N bits para un dominio de valores de M bits, donde $N < M$).