

Base de Datos

DC. FCEyN
2025 1c C4

Emilio Platzer
tute@dc.uba.ar

Slides de Mg. Cecilia Ana Ruz

$\exists \neg \rightarrow \Rightarrow \forall$

Optimización - Parte física

Las diapositivas están basadas en las de Cecilia Ana Rus.

Temas

- Organización física de los registros
- Ejemplo de organización física
- Tipos de índices
- Organización física de los índices
- Costo de búsqueda
- Cálculo T'

ORGANIZACIÓN FÍSICA DE LOS REGISTROS

Existen distintos tipos de almacenamiento

Organización en disco

- Los registros son conjuntos de campos

- Las tablas son conjuntos de registros

- (Hay casos con almacenamiento directo al disco físico saltando al SO)

Los registros pueden ser de longitud fija o variable

- Por simplicidad se asume que son de longitud fija.

- Esto no es cierto, por los campos NULL, los campos varchar y los campos BLOB

LONGITUD DE REGISTRO

En este ejemplo cada registro tiene 3 campos , un id que ocupa 9 posiciones, un nombre de 20 y una fecha de 10

Esta “tabla” tiene 3 registros.

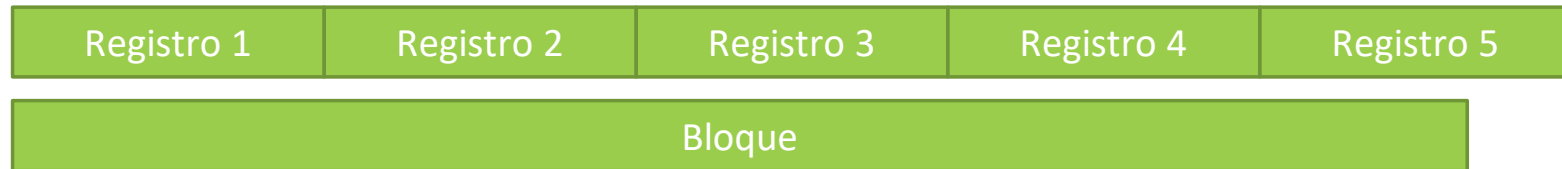
Cual es la longitud del registro?

- Formalmente, si $R (A_1, A_2, \dots A_n)$
- $L_R (\text{longitud de un registro de } R) = \sum \text{long}(A_i)$

	1234567890123456789012345678901234567890	
1	JUAN PEREZ	15/12/2020
2	MARA LOPEZ	
3	J	14/11/2010

FACTOR DE BLOQUEO

Mide cuantos registros entran en un bloque



Cuantos registros entran en este bloque?

FACTOR DE BLOQUEO

- Mide cuantos registros entran en un bloque



- Cuantos registros entran en este bloque?

- $FB_R = \lfloor B/L_R \rfloor$ (redondeado por defecto)

- Cuantos bloques ocupa una tabla?

- $B_R = \lceil T_R / FB_R \rceil$ (redondeado por exceso)

ÍNDICE

Organización física de los registros

Ejemplo de organización física

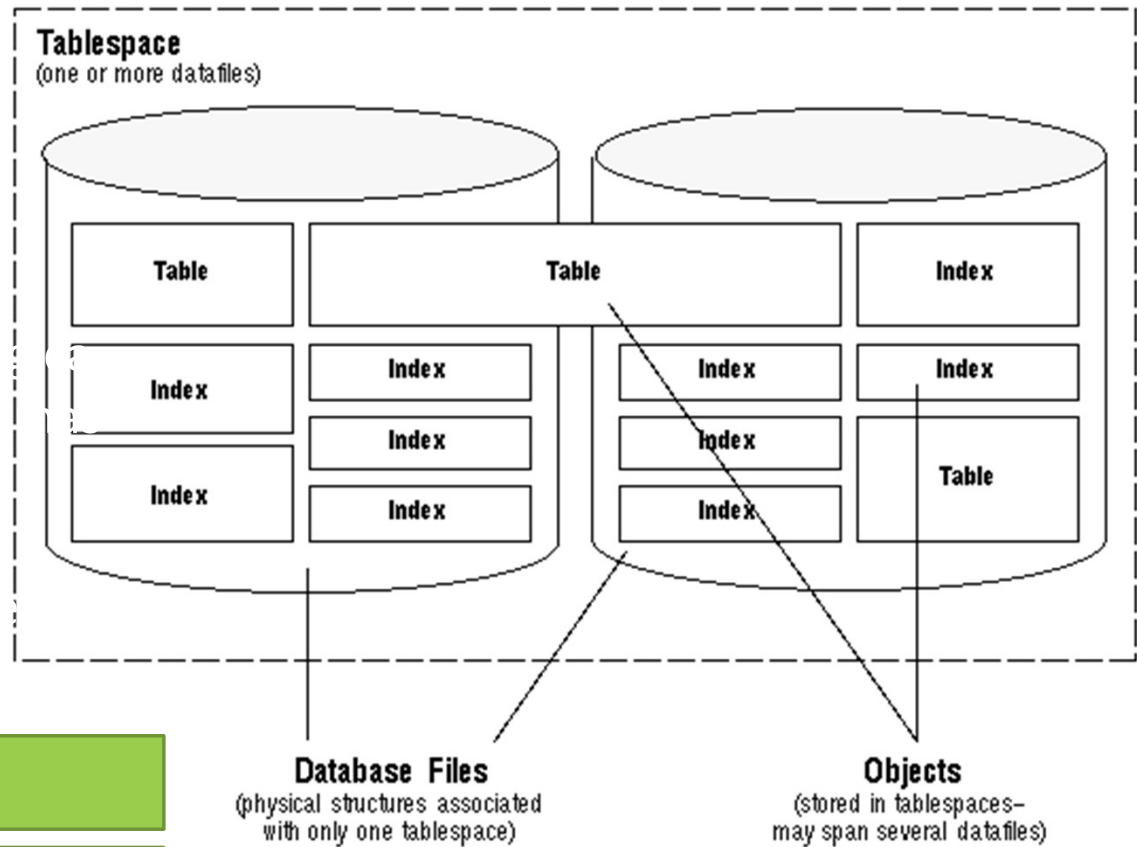
Tipos de índices

Organización física de los índices

Costo de búsqueda

Cálculo T'

ORACLE



Bloque

R1

R2

R3

pctfree

ÍNDICE

Organización física de los registros

Ejemplo de organización física

Tipos de índices

Organización física de los índices

Costo de búsqueda

Calculo T'

TIPOS DE ÍNDICES

Clustered (Mismo ordenamiento que el archivo)

 Primario (sobre valores únicos) .

 No primario (sobre valores NO únicos)

Non clustered (distinto ordenamiento que el archivo)

Cuantos índices Clustered puede tener un archivo?

ÍNDICE CLUSTERED PRIMARIO

DATOS

P	Id	Nombre	Depto
P1	1	Juan	A
	2	Pedro	B
	3	Maria	B
P2	4	Andres	C
	5	Carla	A
	6	Santiago	C
P3	9	Laura	B
	11	Leticia	C

No existe en la
tabla

INDICE

KEY	
1	P1
4	P2
9	P3

✓ Una entrada por cada bloque de datos

✓ Es NO denso (Un índice es denso cuando tiene una entrada por cada valor de la clave)

INDICE CLUSTERED NO PRIMARIO

DATOS

P	Id	Nombre	Depto
P1	1	Juan	A
	5	Carla	A
P2	2	Pedro	B
	3	Maria	B
	7	Laura	B
P3	4	Andres	B
P4	6	Santiago	C
	8	Leticia	C

No existe en la
tabla

INDICE

KEY	
A	P1
B	P2
C	P4

✓ Una entrada por cada valor
de la clave

✓ Cada vez que cambia la
clave se inicia un nuevo
bloque

✓ Es DENSO

ÍNDICE NON CLUSTERED

DATOS

P	Id	Nombre	Depto
P1	1	Juan	A
	2	Pedro	B
	3	Maria	B
P2	4	Andres	C
	5	Carla	A
	6	Santiago	C
P3	7	Laura	B
	8	Leticia	C

No existe en la
tabla

ÍNDICE

KEY	
A	PI1
B	PI2
C	PI3

Diagram illustrating the mapping from index keys to data pages:

- Key A (PI1) points to P1;P2
- Key B (PI2) points to P1;P3
- Key C (PI3) points to P2;P3

ÍNDICE

Organización física de los registros

Ejemplo de organización física

Tipos de índices

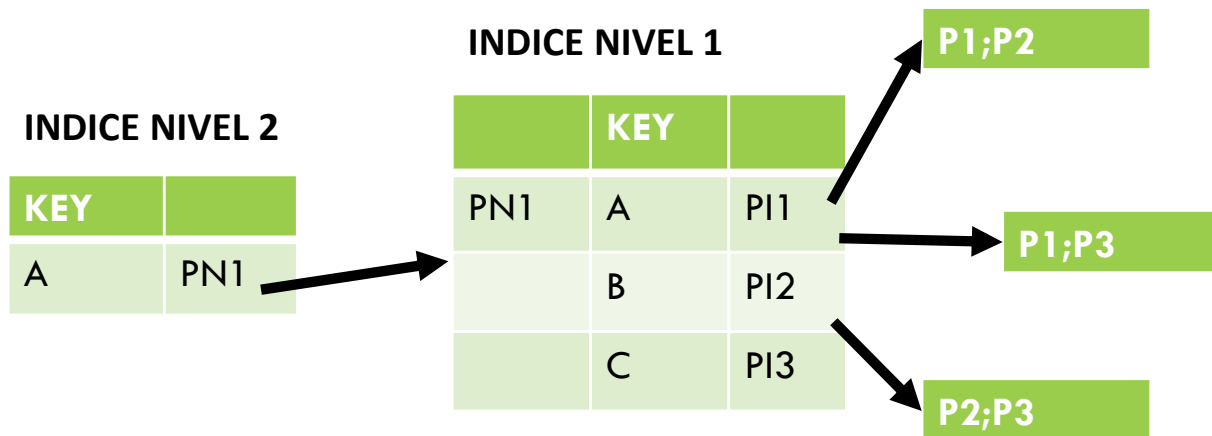
Organización física de los índices

Costo de búsqueda

Calculo T'

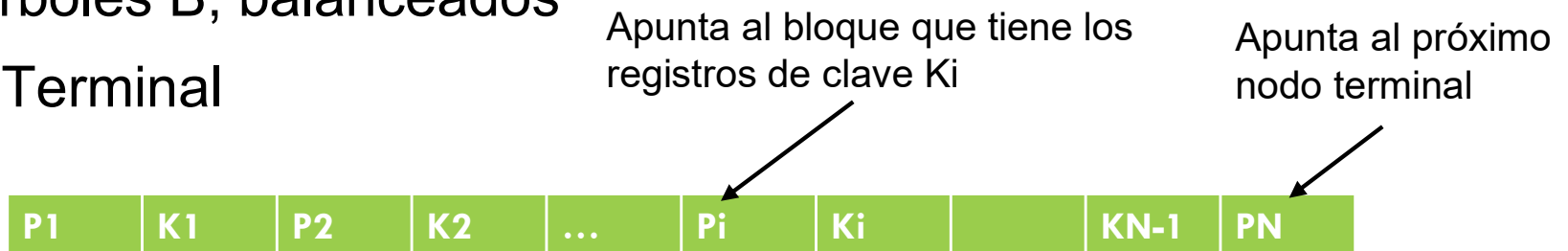
ÍNDICES MULTINIVEL

- Se construyen índices sobre los índices, de que tipo son esos nuevos ÍNDICES desde el nivel 2 en adelante?
- Tomemos el ejemplo del índice non clustered

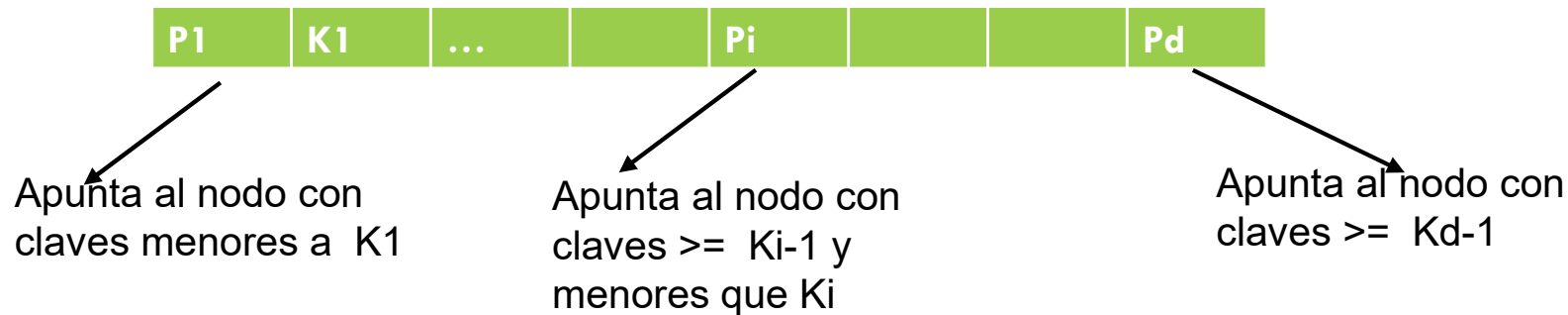


ARBOLES B+

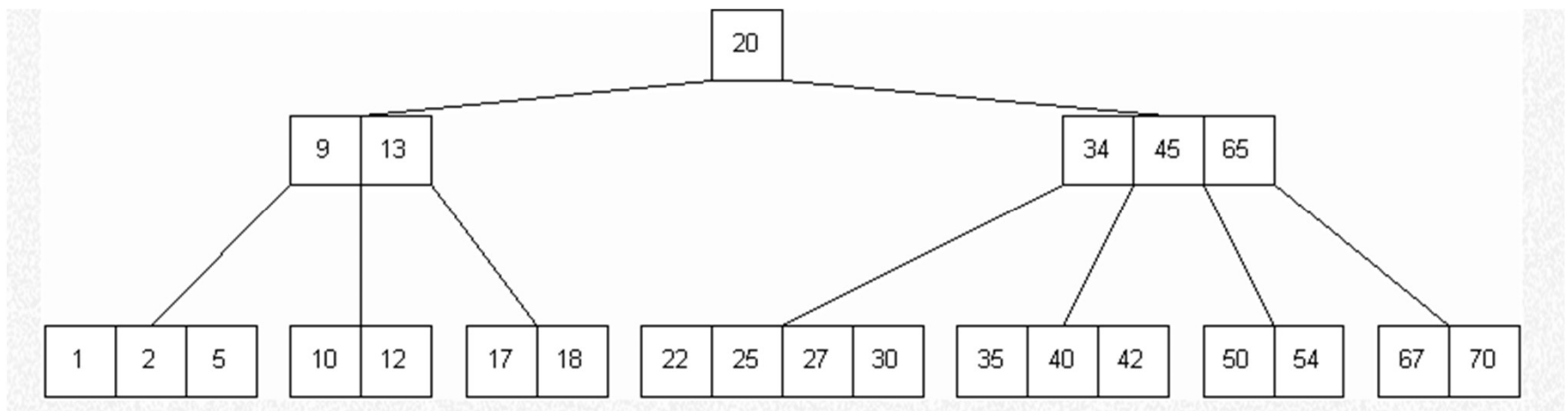
Son arboles B, balanceados
Nodo Terminal



Nodo Intermedio

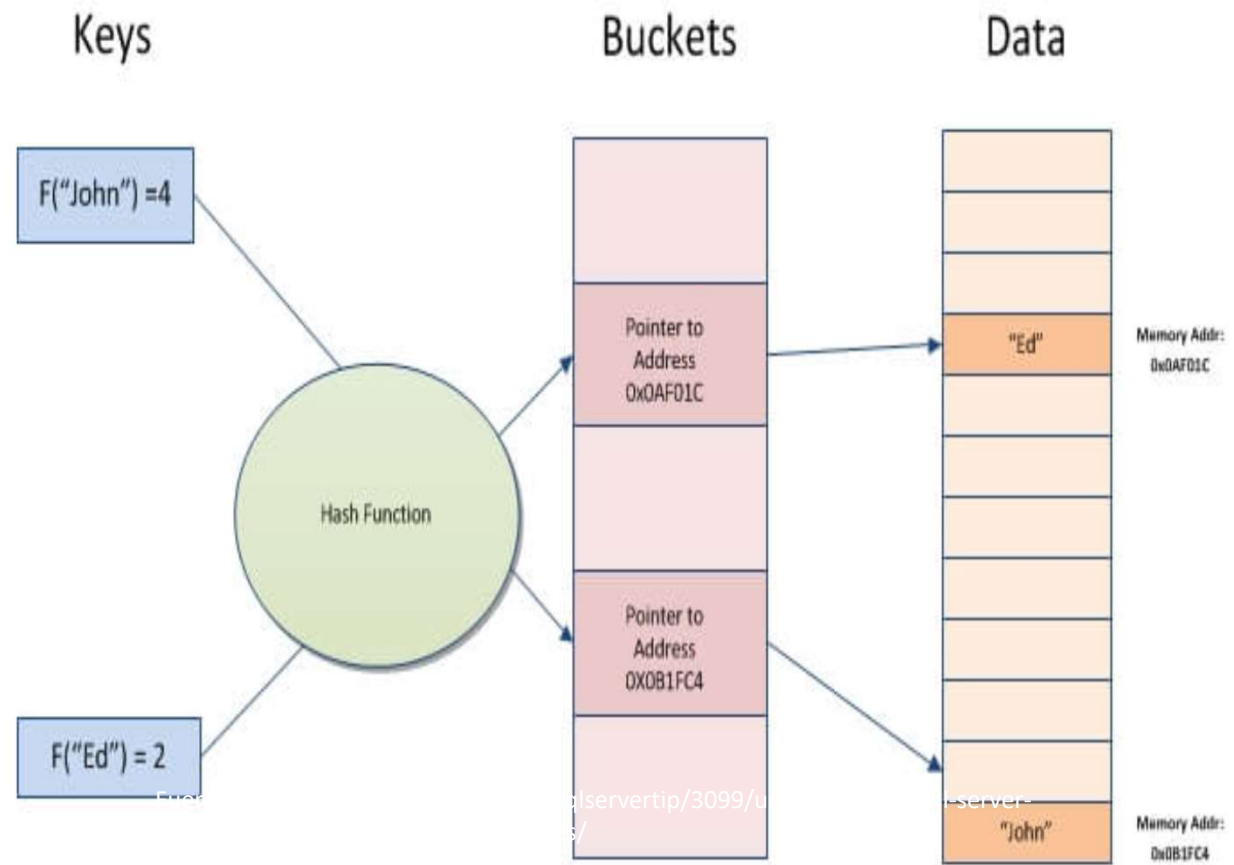


EJEMPLO DE ÁRBOL B+



ÍNDICE HASH

- Se organizan en base a funciones de Hash
- Puede haber varios buckets para cada key
- Es clave que la función de hash distribuya los valores en forma balanceada



ÍNDICE

Organización física de los registros

Ejemplo de organización física

Tipos de índices

Organización física de los índices

Costo de búsqueda

Calculo T'

COSTO DE BÚSQUEDA

TIPO ARCHIVO	$A = a$	$A > a$
SIN INDICE (HEAP FILE)		
SORTED FILE		
INDICE B+ CLUSTERED		
INDICE B+ UNCLUSTERED		
HASH INDEX		

COSTO DE BÚSQUEDA

TIPO ARCHIVO	$A = a$	$A > a$
SIN INDICE (HEAP FILE)	B_R	$=$
SORTED FILE	$\log_2(B_R) + \lceil T'/FB_R \rceil$	$=$
INDICE B+ CLUSTERED	$X + \lceil T'/FB_R \rceil$	$=$
INDICE B+ UNCLUSTERED	$X + T'$	$=$
HASH INDEX	$MB \times B + T'$	B_r

- A : atributo por el cual se busca.
- a : valor específico del atributo.
- BR : número de bloques en el archivo de datos.
- T' : número de tuplas que cumplen la condición.
- FBR : número de tuplas por bloque.
- X : altura del índice B+.
- MB : número de páginas del índice hash.
- B : profundidad del árbol hash.

ÍNDICE

Organización física de los registros

Ejemplo de organización física

Tipos de índices

Organización física de los índices

Costo de búsqueda

Calculo T'

CÁLCULO DEL T'

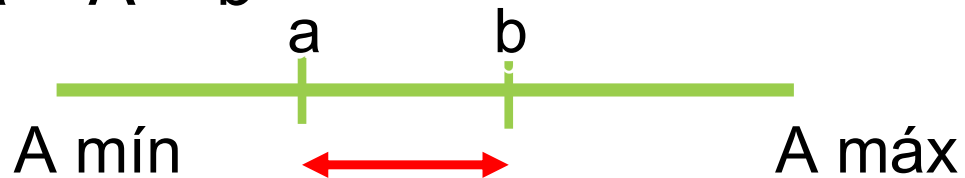
- T': cuántos registros cumplen una determinada condición
- Se asume distribución uniforme. En todos los casos se redondea por exceso.
- $A = a$, $T' = \lceil T_R \rceil \lceil I_{R.A} \rceil$
- $(A = a \text{ y } B = b)$, $T' = \lceil T_R \rceil (\lceil I_{R.A} \rceil * \lceil I_{R.B} \rceil)$
- $A > a$



- $T' = \lceil \lceil T_R \rceil * \text{dist}(A_{\text{max}}, a) / \text{dist}(A_{\text{max}}, A_{\text{min}}) \rceil$

CÁLCULO DEL T' (CONT.)

- $a < A < b$



- $T' = \lceil |T_R| \rceil * \text{dist}(b, a) / \text{dist}(A_{\text{máx}}, A_{\text{mín}}) \rceil$

