

- Un registro va a estar en un bloque de datos

(puede que no se llene de info y tener fragmentación interna)

Ejercicio 1

- Archivo r almacenar:

- N° de carnet: 8 bytes, todos diferentes, comienza por 1
- Nombre de alumno: 40 bytes
- Código de carrera: 2 bytes (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
- Edad: 8 bytes
- Índice académico: 2 bytes

$$V(\text{campo}) =$$

N° de valores diferentes
de un campo

- Número de registros = $M_R = 100\ 000$

- Cada bloque tamaño = $B = 512$ bytes

- Ocupación 65%

- Longitud puntero a bloque = 6 bytes

- Longitud puntero a registro = 2 bytes

- Cada bloque 12 bytes de control

• Tamaño útil de bloque = $B_{util} = (512 - 12) \cdot 0.65 = 325$ bytes

• Longitud del registro = $L_R = \sum C_{cp} = 8 + 40 + 2 + 8 + 2 = 60$ bytes

• Factor de bloque = N° de registros completos que entran en un bloque =

$$= f_R = \left\lfloor \frac{B_{util}}{L_R} \right\rfloor = \left\lfloor \frac{325}{60} \right\rfloor = 5.41 = 5 \text{ Registros/Bloque}$$

• Tamaño de bloque estimado = $b_R = \left\lceil \frac{M_R}{f_R} \right\rceil = \left\lceil \frac{100\ 000}{5} \right\rceil = 20\ 000$ bytes

→ Nº Registros estimados y coste

• $\sigma_{\text{carmet}} = 2345(n)$

$N_{Bc} = N_V \cdot \frac{M_B}{V(\text{carmet})} = 1 \cdot \frac{100,000}{100,000} = 1 \text{ Registro}$

coste

- como no hay indice adicional para buscar (índice, ordenado, ...)

↳ Lectura secuencial



coste = $B_R = 20,000 \text{ bloques}$

(*) Montículo = se inserta segun va llegando) Secuencia Poco a poco

• $\sigma_{\text{codigo}} = 2(n)$

$M_{Bc} = 1 \cdot \frac{100,000}{10} = 10,000 \text{ Registros}$

Coste = 20.000 Bloques

• Secuencial (ordenado por un campo)

↳ Busqueda ↳ secuencial (coste) = B_R

↳ Búsqueda (coste) = $\lceil \log_2(B_R) \rceil + \lceil \frac{N_{Bc}}{f_R} \rceil - 1$

Ej:

archivo ordenado por carnet

$\sigma_{\text{carmet}} = 2345(n)$

$N_{Bc} = N_V \cdot \frac{M_B}{V(\text{carmet})} = 1 \cdot \frac{100,000}{100,000} = 1 \text{ Registro}$

¿coste?

↳ Secuencial = $B_R = 20,000 \text{ bloques}$

↳ Búsqueda = $\lceil \log_2(B_R) \rceil + \lceil \frac{N_{Bc}}{f_R} \rceil - 1 = \lceil \log_2(20,000) \rceil + \lceil \frac{1}{5} \rceil - 1 = 15 \text{ Bloques}$

$\sigma_{\text{codigo_carrera}} = 2(n)$

Solo búsqueda secuencial ya que no está ordenado por ese campo

sí σ_m está ordenado por ese campo \Rightarrow puedo binaria

- Ej
- Organizado de forma secuencial por cod_carmet,
 - Se añade puntero 12 bytes para formar lista enlazada

$$L_B = \sum L_{C_i} = 8 + 40 + 2 + 8 + 2 + 12 = 72 \text{ bytes}$$

$$B_{util} = 325 \text{ bytes}$$

$$f_R = \left\lfloor \frac{B_{util}}{L_R} \right\rfloor = \left\lfloor \frac{325}{72} \right\rfloor = \left\lfloor 4.5 \right\rfloor = 4 \text{ registros/Bloque}$$

$$B_R = \left\lceil \frac{M_R}{f_R} \right\rceil = \left\lceil \frac{100000}{4} \right\rceil = 25000 \text{ bloques}$$

$$\sigma_{carmet} = 2345(n)$$

$$N_{RC} = f \text{ Registro}$$

$$\text{coste } \xrightarrow{\text{Secuencial}} = 25000 \text{ Bloques}$$

$\xrightarrow{\text{Binario}}$ No puedo (porque está ordenado por cod_carmet)

$$\sigma_{cod_carmet} = 2(n)$$

$$N_{RC} = 10000 \text{ Registros}$$

$$\text{coste } \xrightarrow{\text{Secuencial}} = 25000 \text{ Bloques}$$

$\xrightarrow{\text{Binario}}$

$$b = \left\lceil \log_2(25000) \right\rceil + \left\lceil \frac{10000}{4} \right\rceil - 1 = 2514 \text{ Bloques}$$

B+

ej

- Organizado por carmet

Nodo raíz / Intermedio

Número de punteros a bloque = N

calculando

$$\log_3 10 = \frac{\log 10}{\log 3}$$

$$(N \cdot L_{p_{Bloque}} + (N-1) \cdot L_{carmet}) \leq B_{util}$$

$$(N \cdot 6 + (N-1) \cdot 8) \leq 325$$

$$(14N \leq 333) \Rightarrow N \leq \frac{333}{14} \Rightarrow N = 23 \text{ punteros en bloque}$$

Nodo hoja

$$N_h \cdot L_A + L_{B \text{ Bloque}} \leq B \text{ bloques}$$

$$N_h \cdot 60 + 6 \leq 325$$

$$N_h \leq \frac{319}{60} \Rightarrow N_h \leq 5'3 \Rightarrow N_h = 5 \text{ Registros completos}$$

¿Cuántos niveles tiene el árbol?

¿Nodos hoja?

$$\left\lceil \frac{M_R}{N_h} \right\rceil = \left\lceil \frac{100\ 000}{5} \right\rceil = 20\ 000 \text{ bloques} \quad (1)$$

$$\text{Nodos internos 1} = \left\lceil \frac{20\ 000}{23} \right\rceil = 870 \text{ bloques} \quad (2)$$

$$\text{Nodos internos 2} = \left\lceil \frac{870}{23} \right\rceil = 38 \text{ bloques} \quad (3)$$

$$\text{Nodos internos 3} = \left\lceil \frac{38}{23} \right\rceil = 2 \text{ bloques} \quad (4)$$

$$\text{Nodos internos 4} = \left\lceil \frac{2}{23} \right\rceil = 1 \text{ bloque} \quad (5)$$

El árbol ocupa $\sum = 20\ 971$ Bloques, 5 niveles

$$O_{carnet} = 2345(n)$$

$$M_{RC} = I$$

$$B^+ \text{ coste} = 1 + 1 + 1 + 1 + 1 = 5 \text{ bloques} \quad (\text{un bloque por cada nivel})$$

$$O_{cod_carrera} = 2(n)$$

Impleca búsqueda secuencial

$$B^+ \text{ coste} = 1 + 1 + 1 + 1 + 20\ 000 = 20\ 004 \text{ Bloques}$$

Particionamiento (hash) [N particiones o cajones (buckets)]

ej:

$$\text{Hash carnet} \bmod 100 \quad (0 - 99)$$

Número de cajones $N = 100$ cajones

$$N^{\circ} \text{ registros por cajón} = N_c = \frac{M_R}{N} = \frac{100\ 000}{100} = 1000 \text{ Registros}$$

$$L_R = \sum l_i = 60 \text{ bytes}$$

$$f_B = \left\lfloor \frac{B_{util}}{L_R} \right\rfloor = \left\lfloor \frac{325}{60} \right\rfloor = 5 \text{ Registros/Bloque}$$

$$\text{Tamaño de cada cajón} = b_c = \left\lceil \frac{N_c}{f_B} \right\rceil = \left\lceil \frac{1000}{5} \right\rceil = 200 \text{ bloques}$$

$$b_R = N \cdot b_c = 100 \cdot 200 = 20000 \text{ bloques}$$

$$O_{carnet} = 2345(n)$$

$$h(\text{carnet}) = \text{carnet} \bmod 100 = 45 \Rightarrow \text{Lee cajón } 45 \Rightarrow \text{Coste } 200 \text{ bloques}$$

$$O_{cod_camara} = 2(n)$$

$$\text{Lectura secuencial} \Rightarrow 20000 \text{ bloques}$$

Índice

- Coste = $C_{índice} + C_{datos}$
- ordenado respecto al campo de búsqueda \Rightarrow Primario
 - No ordenado por el campo de búsqueda \Rightarrow Secundario (siempre denso)
- El valor no se repite \Rightarrow Campo clave
 - Se repite el valor \Rightarrow Campo No clave
- $N_{RI} = N^o$ de registros en el índice (Nunca duplicados)

Registro índice

$$L_{RI} = L_{PB} + L_{campo}$$

O'

$$L_{RI} = L_{PR} + L_{campo}$$

Clasificación:

- P + CC $\Rightarrow N_{RI} = V(A)$
- P + CNC $\Rightarrow N_{RI} = V(A)$ (solo se apunta al primero)
- S + CC $\Rightarrow N_{RI} = V(A)$
- S + CNC \Rightarrow Cajones de punteros $\Rightarrow N_{RI} = V(A)$ (se apunta al cajón)

Coste

- Índice
 - Secuencial
 - Hash
 - B+
- Datos
 - Primario = $\left\lceil \frac{N_{RC}}{p_R} \right\rceil$
 - Secundario = N_{BC}

$$M_R = 100\ 000$$

ordenado por carnet

1) Índice multível secuencial de 3 mi^o Vértex
(ordenado)

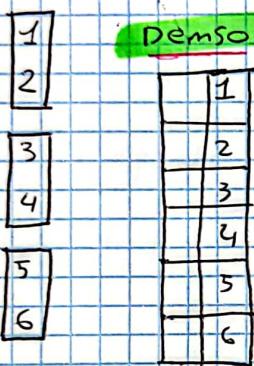
• Primario + Campo clave

• N_{B_I} → Denso (Apunta a todos los valores diferentes)
(solo cuando es primario)

$$N_{B_I} = \sqrt{M_R} (\text{carnet}) = 100\ 000 \text{ registros índice}$$

• Disperso (Apunta a ciertos valores)

$$N_{B_I} = \min \{ b_R, \sqrt{M_R} \}$$



Denso



Disperso

Secundario sólo → Denso

Denso

$$L_{B_I} = L_{C_{\text{carnet}}} + L_{P_{\text{registro}}} = 8 + 7 = 15 \text{ bytes}$$

$$B_{utl} = 325 \text{ bytes}$$

$$f_{B_I} = \left\lfloor \frac{B_{utl}}{L_{B_I}} \right\rfloor = \left\lfloor \frac{325}{15} \right\rfloor = 21 \text{ Registros/Bloque}$$

1º Nivel

$$b_{\text{nivel}_1} = \left\lceil \frac{M_R}{f_{B_I}} \right\rceil = \left\lceil \frac{100\ 000}{21} \right\rceil = 4762 \text{ Bloques}$$

2º Nivel, ademas

$$L_{B_2} = L_{P_{\text{Bloque}}} + L_{C_{\text{carnet}}} = 6 + 8 = 14 \text{ bytes}$$

$$f_B = \left\lfloor \frac{B_{utl}}{L_{B_2}} \right\rfloor = \left\lfloor \frac{325}{14} \right\rfloor = 23 \text{ Registros/Bloque}$$

$$b_{\text{nivel}_2} = \left\lceil \frac{4762}{23} \right\rceil = 208 \text{ Bloques de nivel 2}$$

$$b_{\text{nivel}_3} = \left\lceil \frac{208}{23} \right\rceil = 10 \text{ Bloques nivel 3}$$

$$\text{El índice ocupa} = 10 + 208 + 4762 = 4980 \text{ Bloques}$$

Disperso

$$N_{RI} = N^{\circ} \text{ Reg. Índice} = \min \{ N(\text{carmet}), b_R \} = \min \{ 100\ 000, 20\ 000 \} = 20\ 000 \text{ Reg. Índice}$$

$$L_R = 60$$

$$P_R = \left\lfloor \frac{\text{Bultip}}{L_R} \right\rfloor = \left\lfloor \frac{325}{60} \right\rfloor = 5 \text{ Reg./Bloque}$$

$$b_2 = \left\lceil \frac{M_R}{P_R} \right\rceil = \left\lceil \frac{100\ 000}{5} \right\rceil = 20\ 000 \text{ Bloques}$$

$$L_{RI} = L_{c_{\text{carmet}}} + L_{p_{\text{Reg}}} = 8 + 7 = 15 \text{ bytes}$$

$$P_{BI} = \left\lfloor \frac{325}{15} \right\rfloor = 21 \text{ Reg./Bloque}$$

$$b_{R1} = \left\lceil \frac{20\ 000}{21} \right\rceil = 953 \text{ Bloques Nivel 1}$$

$$L_{R2} = L_{p_{\text{Bloque}}} + L_{c_{\text{carmet}}} = 6 + 8 = 14 \text{ bytes}$$

$$P_{B2} = \left\lfloor \frac{325}{14} \right\rfloor = 23 \text{ Reg./Bloque}$$

$$b_{R2} = \left\lceil \frac{953}{23} \right\rceil = 42 \text{ Bloques Nivel 2}$$

$$b_{R3} = \left\lceil \frac{42}{23} \right\rceil = 2 \text{ Bloques Nivel 3}$$

$$\text{El índice ocupa} = 2 + 42 + 953 = 997 \text{ bloques}.$$

$$\sigma_{\text{carmet}} = 2345(n)$$

$$N_{RC} = NV \cdot \frac{M_R}{V(\text{carmet})} = 1 \text{ Registro}$$

coste

- Secuencial = $b_A = 20000$ bloques

- Binaria = $\lceil \log_2(b_A) \rceil + \lceil \frac{N_{RC}}{p_A} \rceil - 1 = 15$ bloques

- Índice

- Denso

$$\underbrace{\lceil \log_2(10) \rceil}_{\text{Nivel 3}} + 1 + 1 + 1 = 7 \text{ bloques}$$

↓
Datos

- Disperso

$$\underbrace{\lceil \log_2(2) \rceil}_{\text{Nivel 3}} + 1 + 1 + 1 = 4 \text{ Bloques}$$

↓
Datos

$$\sigma_{\text{cod_camera}} = 2(n)$$

$$N_{RC} = NV \cdot \frac{M_R}{V(\text{cod_camera})} = 10000 \text{ Registros}$$

coste

- Secuencial = 20000 Bloques

- No está ordenado por este campo

↳
No puedo buscar por índice

2) Índice multimedivel secuencial sobre cod_carrera

secundario + campo no clave \Rightarrow cajones de punteros

siempre Denso

$$N_{RZ} = V(\text{cod_carrera}) = 10$$

$$L_{RI} = L_{\text{cod_carrera}} + L_{P_{\text{Regue}}} = 2 + 6 = 8 \text{ bytes}$$

$$f_R = \left\lfloor \frac{\text{Butil}}{L_{RZ}} \right\rfloor = \left\lfloor \frac{325}{8} \right\rfloor = 40 \text{ Reg/Bloque}$$

$$b_{\text{Nivel1}} = \left\lceil \frac{N_{RZ}}{f_R} \right\rceil = \left\lceil \frac{10}{40} \right\rceil = 1 \text{ Bloque}$$

$$\text{Cajones de punteros} = N = V(\text{cod_carrera}) = 10 \text{ cajones}$$

$$\text{Nº punteros registrados por cajón} = \frac{M_R}{N} = \frac{100\ 000}{10} = 10\ 000 \text{ punteros a registros}$$

$$L_{P_{\text{Reg}}} = 4 \text{ bytes}$$

$$f_{\text{Bcajón}} = \left\lfloor \frac{\text{Butil}}{L_{P_{\text{Reg}}}} \right\rfloor = \left\lfloor \frac{325}{4} \right\rfloor = 46 \text{ punteros a Reg/Bloque.}$$

$$\text{Nº bloques por cajón} = b_{\text{cajón}} = \left\lceil \frac{N}{f_{\text{Bcajón}}} \right\rceil = \left\lceil \frac{10\ 000}{46} \right\rceil = 218 \text{ bloques}$$

$$\text{tamaño índice} = 1 + 10 + 218 = 218 \text{ bytes}$$

$$O_{\text{carmet}} = 2345(n)$$

$$N_{RI} = 1$$

Coste

- Secuencial = 20 000 bloques
- Binaria = $\lceil \log_2(20000) \rceil + \lceil \frac{1}{5} \rceil - 1 = 15$ bloques

$$O_{\text{cod-carment}} = 2(n)$$

- Secuencial = 20 000 bloques

- Índice

$$C = C_{\text{Índice}} + C_{\text{datos}} = 1 + \underbrace{218}_{\text{Reg por cajón}} + \underbrace{20000}_{\text{Reg por cajón}} = 20219 \text{ Bloques}$$

3) B+ sobre carment

Primario + Campo clave

$$N_{RI} = \sqrt{(\text{carment})} = 100 000$$

Nodos Raíz / Intermedio

$$N \cdot L_p_{\text{Bloque}} + (N-1) L_{\text{carment}} \leq B_{\text{utip}}$$

$$N \cdot 6 + (N-1) \cdot 8 \leq 325$$

$$N = 23 \text{ punteros a bloque}$$

Nodo hoja

$$N_h (L_{\text{carment}} + L_{\text{Reg}}) + L_{\text{P Bloque}} \leq B_{\text{utip}}$$

$$15 N_h \leq 319$$

$N_h = 21$ punteros a registros o valores de campo

$$\text{hojas} = \left\lceil \frac{N_{BI}}{N_h} \right\rceil = \left\lceil \frac{100 000}{21} \right\rceil = 4762 \text{ bloques}$$

$$\text{Nodos Inter } 1 = \left\lceil \frac{4762}{23} \right\rceil = 208 \text{ bloques}$$

$$\text{Nodos Inter } 2 = \left\lceil \frac{208}{23} \right\rceil = 9 \text{ bloques}$$

$$\text{Nodos Inter } 3 = \left\lceil \frac{9}{23} \right\rceil = 1 \text{ bloque}$$

$$\text{Ocupa} = 1 + 9 + 208 + 4762 = 4980 \text{ Bloques (4 niveles)}$$

$\sigma_{\text{carmet}} = 2345 (\text{n})$

$$N_{RC} = 1$$

coste

$$\underbrace{1+1+1+1}_{\text{Índice}} + \underbrace{1}_{\text{Datos}} = 5 \text{ bloques}$$

Índices Asociativos (Hash)

• Estáticos: Se crean todos los cajones al inicio

• Dinámicos: Se crean algunos cajones

$$N^{\circ} \text{ de cajones en este momento} = N = 2^8$$

~~cjw~~

Índice hash sobre carnet

$$\text{Función de asociación } h(\text{carmet}) = \text{número bin } 8 \text{ dígitos} \quad 2^8 = 256 \text{ cajones}$$

Archivo ordenado por carnet

Primario + CC

$$N_{RI} = \sqrt{(\text{carmet})} = 100 \text{ 000}$$

Hash \longrightarrow Siempre denso

$$N^{\circ} \text{ Registros por cajón} = N_{RC} = \frac{N_{RI}}{256} = \frac{100000}{256} = \underline{391 \text{ Registros Índice por cajón}}$$

¿Cuánto ocupa un cajón?

$$L_{BI} = L_{\text{carmet}} + L_{PR} = 8 + 7 = 15 \text{ bytes}$$

$$f_{bR} = \left\lceil \frac{B_{\text{tip}}}{L_{BI}} \right\rceil = \left\lceil \frac{325}{15} \right\rceil = \underline{21 \text{ Reg}} \quad \text{Bloque}$$

$$b_R = \left\lceil \frac{N_{AC}}{f_{bR}} \right\rceil = \left\lceil \frac{391}{21} \right\rceil = 19 \text{ Bloques por cajón}$$

$$\text{Indices ocupados} = 256 \cdot 19 = 4864 \text{ Bloques}$$

$$\sigma_{carmet} = 2345(n)$$

$$N_{RC} = I \text{ Reg}$$

¿Coste?

- Secuencial = 20 000 bloques
- binaria = 15 bloques
- Índices = $19 + 1 = 20$ bloques

$$\sigma_{cod_camera} = 2(n)$$

- Secuencial = 20 000 bloques

6) Índice hash sobre cod_camera

Función de asociación mod 5

$$\text{mod } 5 \Rightarrow (0-4) \Rightarrow 5 \text{ cajones}$$

Secundaria + CNC \Rightarrow cajones de punteros

II

$$\text{Demso} \Rightarrow N_{RI} = \sqrt{\sigma_{cod_camera}} = 10 \text{ Reg de tipo índice}$$

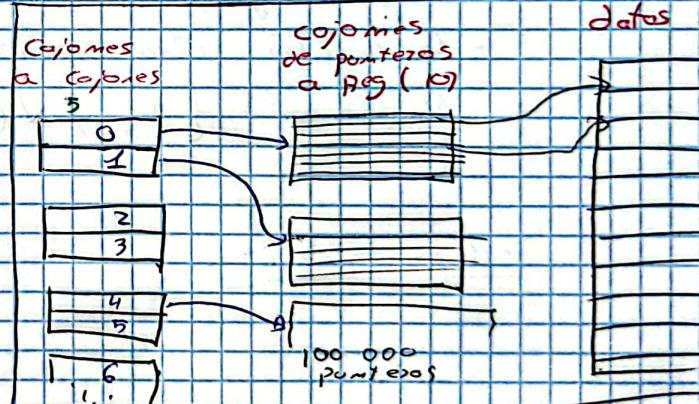
$$N^o \text{ Reg por cajón} = \frac{10}{5} = 2 \text{ Reg} \quad \text{Índices por cajón}$$

$$L_{RI} = L_{cod_camera} + L_{p_B} = 2+6 = 8 \text{ bytes}$$

$$f_{begin} = \left\lceil \frac{B_{RI}}{L_{RI}} \right\rceil = \left\lceil \frac{325}{8} \right\rceil = 40 \text{ Reg/Bloque}$$

$$b_c = \left\lceil \frac{2}{f_{begin}} \right\rceil = \left\lceil \frac{2}{40} \right\rceil = 1 \text{ Bloque}$$

- cajones de punteros a Reg



$$N^o \text{ cajones de punteros} = \sqrt{\sigma_{cod_camera}} = 10 \text{ cajones}$$

$$f_{begin} = \frac{100000}{10} = 10000 \text{ punteros a Reg por cajón}$$

$$f_b = \left\lceil \frac{325}{16} \right\rceil = 46 \text{ punteros a Reg/Bloque}$$

$$b_{c妖} = \left\lceil \frac{100000}{46} \right\rceil = 218 \text{ bloques}$$

$$\text{Ocupa} = 5 \cdot 1 + 10 \cdot 218 = 2185 \text{ bloques}$$

$$O_{camera} = 2345(n)$$

$$N_{AC} = I \cdot Reg$$

coste

- secuencial = 20000 Bloques

- Binaria = 15 Bloques

$$O_{cod_camera} = 2(n)$$

$$N_{AC} = 10\ 000 \cdot Reg$$

• coste

- Secuencial = 20 000 Bloques

- Indices

$$I + 218 + 10000 = 10\ 219 \text{ Bloques}$$

↑
datos
↓ cajón punteros
↓ cajón hash

cjw

$$(kod_camera = 2) \vee (cod_camera = 4)$$

$$= 2 \cdot (I + 218 + 10000)$$

Indices multiclave

Rejilla

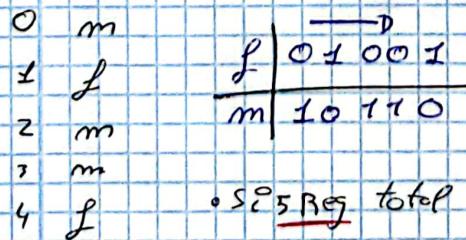
(hundir la rejilla)

- La rejilla tiene punteros a bloques, que son cajones

$$N_{BI} = m^{\otimes n} \{ N_A, \sqrt{(campo 1)} \cdot \sqrt{(campo 2)} \}$$

cajón de bits

Mapa de bits



• 595 Reg total \Rightarrow mapa tiene 5 bits

$$\sigma_{\text{corner}} = 2345 \quad \checkmark \quad \sigma_{\text{cod-camera}} = 2$$

$$N_{RC} = 1 \cdot \frac{100000}{100000 \cdot 10} = 0'1 = 1 \text{ Reg}$$

$$\sigma_{\text{corner}} = 2345 \quad \checkmark \quad \sigma_{\text{cod-camera}} = 9$$

$$N_{RC} = 1 \cdot \frac{100000}{100000} + \frac{100000}{10} = 10001 \text{ Reg}$$

Unidad 2

TI saldo de todo lo que entra, solo me quedo con saldo (Select)

Saldo < 2500 (Where)

Desde "from" hacia la derecha y lo último es select

M: N° de bloques en memoria.

$$\text{Coste total} = \sum_{i=1}^n c_i \text{ [bloques]}$$

- Lo que me entra en memoria, tengo que materializarlo y guardarlo en disco

Álgebra Relacional

- Proyección $\Pi_A(r)$ Select distinct A from r
- Selección $\sigma_{\text{condición}}(r)$ Select * from r Where condición
- Producto cartesiano $r \times s$ Select * from r, s
- Unión $r \cup s$ Select * from r Union select * from s
- Diferencia $r - s$ Select * from r Except select * from s
- Intersección $r \cap s$ Select * from r Intersect select * from s

- Natural Join $r \bowtie s$ Select * from r Natural Join s

- Join $r \bowtie_{\text{condición}} s$ Select * from r Inner Join s On condición

- Outer Join $r \bowtie_{\text{condición}} s$ Select * from r Full Outer Join s

- Agregado a $\sigma_{\text{count}(B)}(r)$ Select A, count(B) from r group by A

$r \times s$

A	A
1	1
2	2
3	4

$r - s$

$= 9 \text{ tuplas}$

$\boxed{2}$ \oplus

$\sqcup R \sqcup S$

1, 1
3, 3

$r \cap s$

1
3

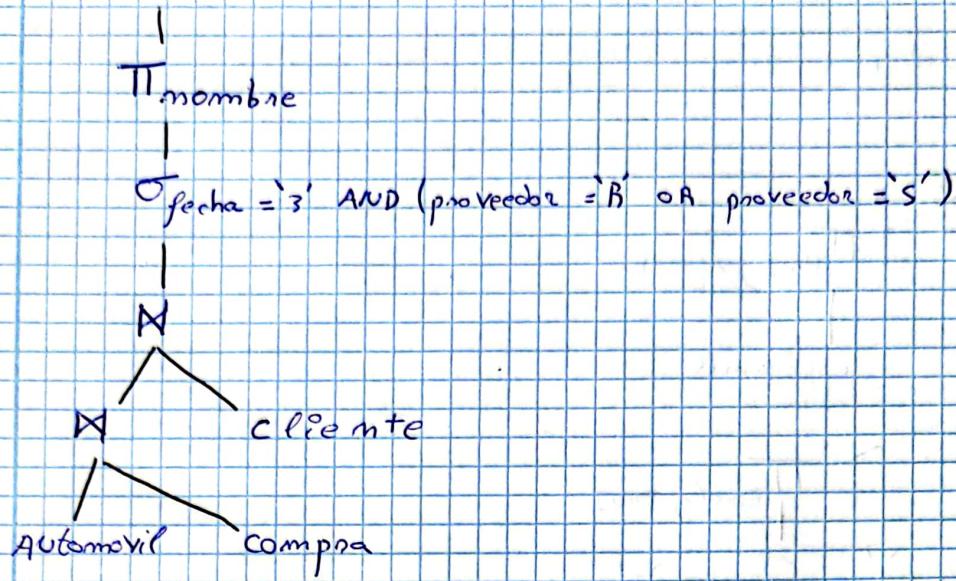
Union

1
2
3
1
2
4

Natural Join, lo mismo pero la condición lo hace solo

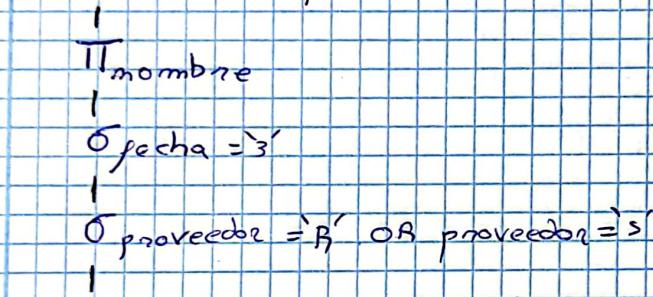
• Full outer join, los que no coinciden, salen a null

Pjón

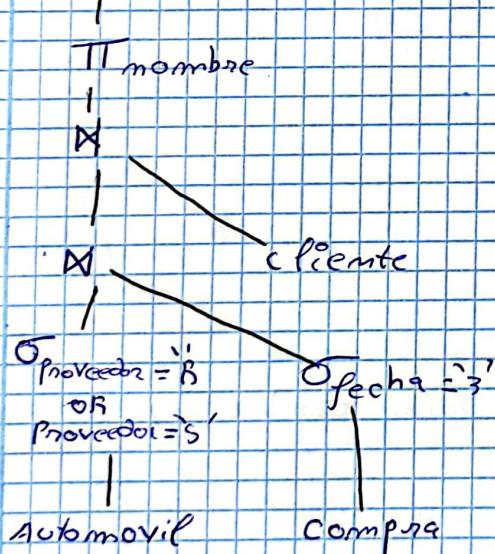


- Aplico optimización eunística

① Hecho tijera por AND



② bajo selección a su tupla

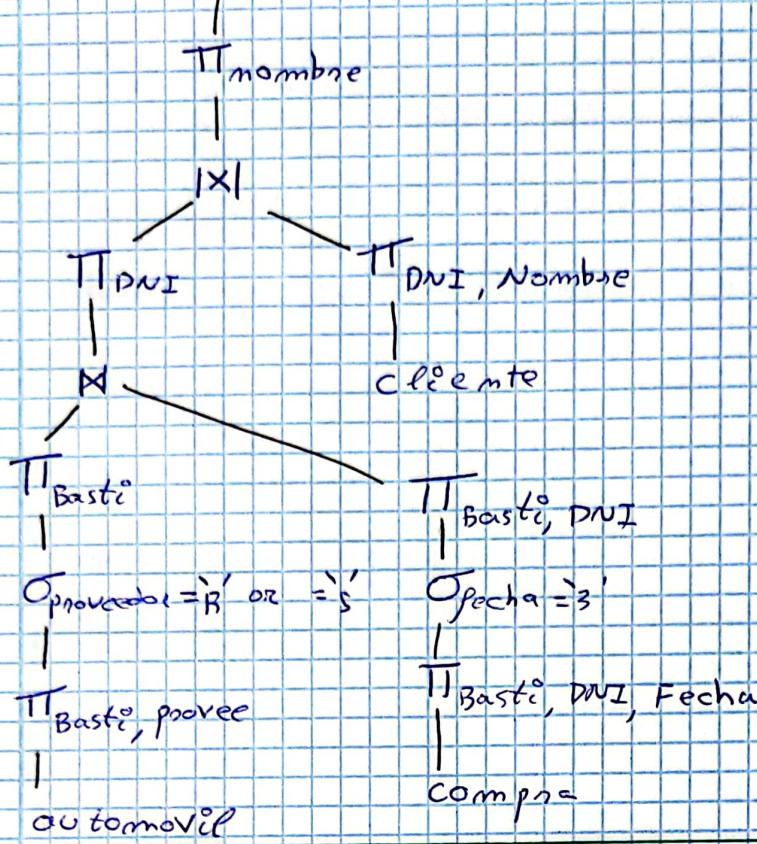


③ 1º las de menos tuplas (las + sencillas)

④ Producto cartesiano

⑤ Crear proyecciones

⑥ Encuadrar y Materia Pizar



Tamaño de cada operación

π_A entran N_R tuplas
salen N_R tuplas

σ_{com} entran N_R tuplas

$$\text{Salida} = \frac{N_R \cdot N_V}{V(\text{campo})}$$

$R \times S$ (Producto cartesiano)

$$\text{Salida} = N_R \cdot N_S \text{ tuplas}$$

$R \bowtie S$ (Natural Join) y $R \bowtie_{\text{con}} S$ (Join condicional)

$$\text{Salida} = \frac{N_R \cdot N_S}{\max(V(A, B), V(A, S))}$$

$R \cup S$

$$\text{Salida} = N_R + N_S$$

$R \cap S$ Salida = $\min(N_R, N_S)$

$R - S$ Salida = N_R

$$A \setminus_{\text{common}(B)}^{\text{(R)}} \text{salida} = V(A, R)$$

B \setminus_{S} (outer JOIN)

$$\text{salida} N_A + N_B + N_S$$

$$\frac{N_S = 25}{\text{c } V(A, B) = \min(N_S, V(A, B)) = \min(25, 50) = 25}$$

$$V(A, R) = 50$$

$$N_R = 1000$$

$$\frac{1}{T_1 \text{ nombre}} N_R = \frac{400 \cdot 50000}{\max(400, 50000)} = 400 \text{ Registros}$$

$$V(\text{Nombre}) = \min(N_R, V(\text{Nombre}, \text{cliente})) = \\ = \min(400, 50000) = 400$$

$$T_{\text{Apellido}} = \frac{400 \cdot 10000}{\max(400, 10000)} = 400 \text{ Reg}$$

$$V(\text{Apellido}) = \min(N_R, V(\text{Apellido})) = 400$$

$$N_R = 50000 \\ V(\text{DNI}) = 50000 \\ V(\text{Nombre}) = 50000$$

$T_1 \text{ DNI, Nombre}$
ciente

$T_1 \text{ Baste}^o$

$$\left\{ \begin{array}{l} N_R = \frac{2 \cdot 100000}{500} = 400 \text{ Reg} \\ V(\text{Prove}) = 2 \\ V(\text{Baste}^o) = 400 \end{array} \right.$$

$$N_R = \frac{0'2 \cdot 50000}{3} = 10000 \text{ cuplos} \\ V(\text{DNI}) = 10000 * \min() \\ V(\text{Baste}^o) = 10000 * \min() \\ V(\text{Fecha}) = 1'1$$

Prove = 'R' OR prove = 'S'

Fecha = '3'

$$N_R = 100000$$

$$V(\text{Baste}^o) = 100000 *$$

$$V(\text{Prove}) = 500 *$$

$T_1 \text{ Prove, Baste}^o$

$$N_R = 100000$$

$$V(\text{Baste}^o) = 100000$$

$$V(\text{Prove}) = 500$$

Automovil

$$N_R = 50000$$

$$V(\text{DNI}) = 50000$$

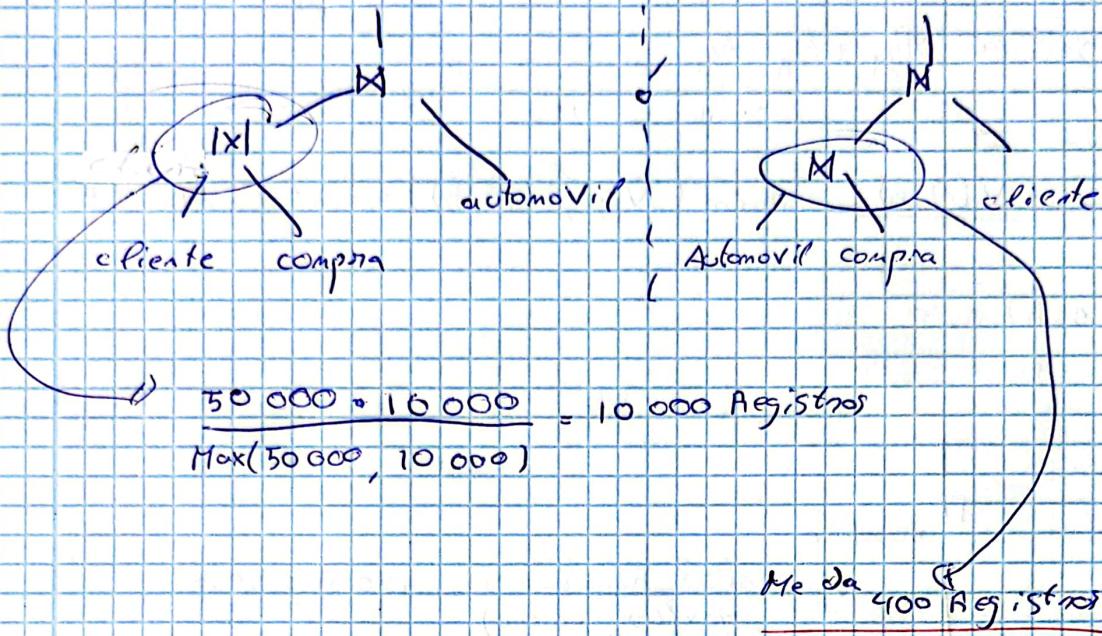
$$V(\text{Baste}^o) = 50000$$

$$V(\text{Fecha}) = 3 \rightarrow 0'2 \\ \text{---} \\ 0'3 \\ \text{---} \\ 0'5$$

compra

3º Regla

Debo mirar si es + óptimo hacer un JOIN u otra



Me quedo con el que me de menos

- Coste: Nº bloques leídos del disco
- **Materialización**: Escribir a disco cuando no entra en memoria una operación
(order by, group by, Join, U, Π, -)
- **Encauzar**: No usar el disco \Rightarrow coste = 0.

Operaciones que no tienen que esperar la llegada de todas las tuplas
 (Π, σ)

6º Regla

- Encauzar todo lo que pueda
- Materializar solo lo imprescindible

Rama Automóvil

- Lectura secuencial

$$L_A = 140$$

$$f_R = \left\lceil \frac{1024}{140} \right\rceil = 8 \text{ Reg/Blog}$$

$$b_R = \left\lceil \frac{100000}{8} \right\rceil = \underline{14286 \text{ Bloques}} = \text{coste}$$

- Binaria, si la tabla estuviese ordenada por campo proveedor

$$\text{coste} = Nv \cdot \left(\lceil \log_2 b_R \rceil + \lceil \frac{N_{RC}}{f_R} \rceil - 1 \right)$$

El enunciado no dice que esté ordenado \Rightarrow No lo puedo usar

- Índices, tipo B+ 4 niveles sobre proveedor

B+ 4 niveles

- Principio (solo por Venta)

$$N_{RC} = NY \cdot \frac{M_A}{V(\text{campo})} = 2 \cdot \frac{100000}{500} = 200$$

$$\text{coste} = \left(4 + \frac{200}{f_R} \right) \cdot 2$$

Secundario

$$\text{coste} = (4 + 200) \cdot 2 = \underline{408 \text{ Bloques}}$$

Aplico índices pero una T implica lectura secuencial \Rightarrow lo gasto

~~T~~

~~O~~

(solo lo dejo con lectura secu)

~~T~~
~~I~~
Automóvil

(después de leer se pierden los índices y no podremos usar Bin o índices)

Rama Compra

- secuencial coste = $\lceil \frac{50000}{8} \rceil = \underline{6250 \text{ Bloques}}$
- Binaria }
- Índice } No puedo aplicar

Join automóvil y comp79

~~M=5~~

$$N_R = 400 \quad b_{\text{R}} = 8$$

$$N_S = 10000 \quad b_S = 100$$

TI Basico

$$N_R = 400$$

$$L_R = 20$$

$$f_R = \left\lceil \frac{1024}{20} \right\rceil = 51 \text{ Reg/Bloq}$$

$$b_R = \left\lceil \frac{400}{51} \right\rceil = 8 \text{ bloques}$$

TI Basic, DNI

$$N_S = 10000$$

$$L_R = 20 + 20 = 40$$

$$f_R = 25$$

$$b_R = 400 \text{ bloques}$$

Algoritmos

- Bucle anidado por bloque (siempre se puede aplicar)

$$\text{coste} = \left\lceil \frac{b_R}{M-2} \right\rceil \cdot b_S + b_R = \left\lceil \frac{8}{3} \right\rceil \cdot 8 + 400 =$$

* b_R siempre tiene que ser el menor

* Si las 2 entradas NO vienen de disco

II

Encauzo la Rama + grande
Materializo la Rama + pequeña

$$M-2 = 5-2 = 3! > \min\{8, 400\}$$

$$\text{coste} = \left\lceil \frac{b_R}{M-2} \right\rceil \cdot b_S \Rightarrow \left\lceil \frac{400}{3} \right\rceil \cdot 8 = 1072 \text{ Bloq} + 8 \text{ de Materialización}$$

~~ej~~

$$\text{Si } M=10$$

$$M-2 = 8 > \min\{8, 400\}$$

II

No necesito encauzar la Rama 8

II

8 a Ram y encauzo 400

II

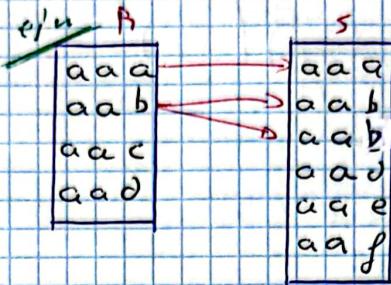
$$\text{coste} = 0$$

- Reunión por mezcla

- Si están las 2 ordenadas por el campo común (en disco)

U

$$\text{Coste} = b_R + b_S$$



- Si las 2 NO vienen de disco

$$\text{Coste} = 0 \text{ ya que encuadraría las 2}$$

- Bucle Ampliado Indexado

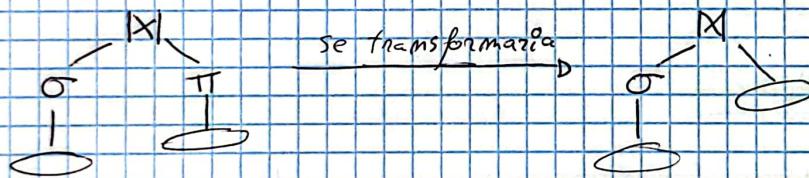
$$C = N_R \cdot c(C_{\text{índice}} + C_{\text{datos}}) + b_R$$

|
 X |
 / \
 R S
 Índice

- Encuadro 1 rama, la que no tiene índice

$$C = N_R \cdot c(C_{\text{índice}} + C_{\text{datos}})$$

- Si fuese



- Reunión por Asociación (Hash).

1	1
2	2
3	3
4	4
	5
	5

$$N = N^{\circ} \text{ de cajones}$$

$$C = 3(b_R + b_S) \quad \text{sob si } M \geq N+1$$



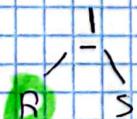
- Si los 2 NO vienen de disco, puedo encuadrarlos 2



$$C = 2(b_R + b_S)$$



- + - ∩ Jocim



- Encuadro R , siempre la de la Izquierda

- Ordenación Mezcla externa

$b_R > M \Rightarrow$ Materializado antes de ordenar

$$C = b_R \left(2 \cdot \lceil \log_{M-1} \left(\frac{b_R}{M} \right) \rceil + 1 \right)$$

+

$$\text{orden} \Rightarrow \text{coste} = b_R + C + b_R$$

+

Tema -3-

• ACID (Atomicidad, consistencia, aislamiento, durabilidad)

{ o todo o nada

{ bloques

↳ sistema de Recuperación

{ Diario sistema

(Recovery)

• Transacciones

- conjunto de instrucciones individuales que se ejecutan como una unidad

- Planificación

{ concurrente

{ serie

secuencia → conflicto

↳ Vista

• Grafo de precedencia:

{ - tantos nodos como transacciones
- 1 conflicto, 1 arco entre nodos

R → W

W → R

W → W



t_1, t_2

t_2, t_1



t_1, t_2 planificación serie

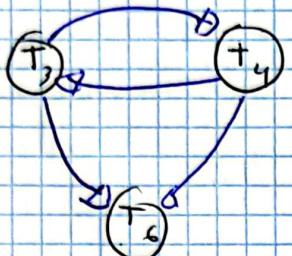
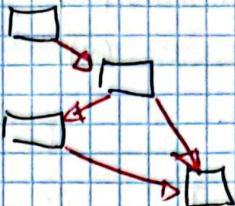
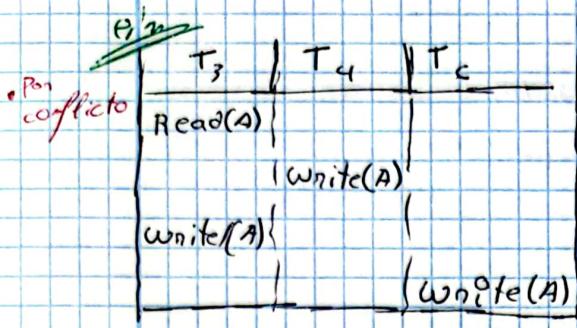
• Si hay ciclo no es secuenciable

• Vista:

1.- Ver quienes leen el valor inicial de la DB

2.- Si alguien lee un valor escrito por otra transac., debe ocurrir lo mismo

3.- Quién escribe el valor final de la DB



Hay ciclo \Rightarrow No es secuenciable con el ciclo

• Vista

Vista A

- 1º T_3 lee el valor inicial
- 2º No hay nadie que haga write y luego haga read
- 3º T_c escribe el valor final

T_3, T_4, T_c

Es secuenciable en vista

- Si es secuenciable en conflicto, por teoría \Rightarrow es en vista
- Planiificación recuperable

• Recuperable

R
W
R
C
C

• No Recuperable

R
W
C
C

- Si no hay Roll back en cascada \Rightarrow puedo afirmar que es Recuperable

• Si hay Roll back

W
R
C
C

• No hay Roll back

W
C
R
C

Concurrencia

- Bloqueos
 - Compartido (c) Leer datos
 - Exclusivo (x) escribir datos

Petición

		C	X
		T	F
sistema actual	C	T	F
	X	F	F

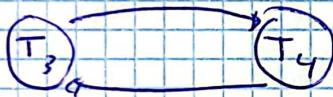
T se concede

F, no se concede

ej:

$$A \rightarrow T_4(c, 5) \rightarrow T_3(x, 8)$$

$$B \rightarrow T_3(x, i) \rightarrow T_4(c, 7)$$



- Si hay ciclo, el sistema está en interbloqueo

- Quito el que tenga menos permisos concedidos

$$\left. \begin{array}{l} T_3 = 1 \\ T_4 = 1 \end{array} \right\} \Rightarrow \text{el que quería}$$

T_3
Bloqueo X(B)

Escríb(B)

T_4

Bloquear C(A)

Bloquear C(B)

Bloquear X(A)

ej:

T_4 espera por T_3



- Fase de crecimiento Solo pide bloques
- Fase de decrecimiento "quita"

- Antes del commit

- Puedo producir Rollback en cascada (No es seguro)

- Despues del commit

- No hay Rollback en cascada \Rightarrow Planificación Recuperable

- Estricto (se mantiene X hasta después del commit)

- Rígido (suelto todos después del commit)

- Refinado Permite cambios de bloqueo

ejer 10

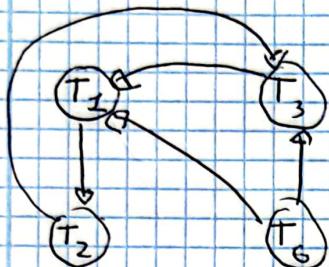
A $\rightarrow T_1(x, 1) \xrightarrow{\circ} T_3(x, 7) \xrightarrow{\circ} T_6(x, 8)$

B $\rightarrow T_1(s, 3) \xrightarrow{\circ} T_5(s, 5) \xrightarrow{\circ} T_4(s, 9)$

C $\rightarrow T_4(s, 4)$

D $\rightarrow T_3(s, 6) \xrightarrow{\circ} T_2(x, 10)$

E $\rightarrow T_2(x, 2) \xrightarrow{\circ} T_1(s, 11)$



• Hay interbloqueo T_1, T_2, T_3

• T_4, T_5 no están en el grafo \rightarrow

sí estuviesen y no sacan flechas

¿cuál mato?

- El que tenga menos permisos concedidos

$$T_1 = 2$$

$$\begin{array}{l} T_2 = 1 \\ T_3 = 1 \end{array} \left. \begin{array}{l} \\ \end{array} \right\} \text{el que fallece}$$

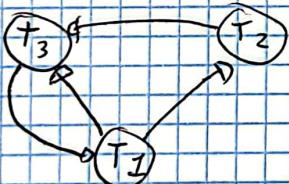
ejer

A $\rightarrow T_3(x, 2) \xrightarrow{\circ} T_2(x, 7) \xrightarrow{\circ} T_1(c, 12)$

B $\rightarrow T_1(x, 5) \xrightarrow{\circ} T_3(x, 23)$

C $\rightarrow T_3(c, 2)$

• Interbloqueo T_3, T_1



$$T_3 = 2$$

$$T_1 = 1 \Rightarrow \text{Lo mato}$$

• Ejecución T_3, T_2 , T_1 desecharlo

Recuperación

- Modificación diferida:

Los datos modificados por una transacción salen a la B.D.
cuando se hace commit
(Redo Log)

ejm

<u>Redo Log</u>	<u>Memoria global</u>
$\langle T_1, \text{start} \rangle$	
$\langle T_1, A, 10 \rangle$	
$\langle T_1, \text{commit} \rangle$	$A = 10$

Checkpoint \Rightarrow Pasa lo que esté en memoria a disco

- Modificación inmediata:

Según se produce el write, se escribe en memoria
(Undo/Redo Log)

Tenemos 2 listas

- { - Undo \rightarrow sin commit
- Redo \rightarrow con commit

Primer deshago de $-a-$ y luego Rehago de $-a-$

ejm

<u>Undo/Redo Log</u>	<u>Memoria global</u>
$\langle T_1, \text{start} \rangle$	$A = 0$ $B = 10$
$\langle T_1, A, 10 \rangle$	$A = 10$
$\langle T_1, \text{commit} \rangle$	
$\langle T_2, \text{start} \rangle$	
$\langle T_2, B, 30 \rangle$	$B = 30$

Si se cae

$$\text{Undo} = \{ T_2 \}$$

$$\text{Redo} = \{ T_2 \}$$

$A = 10$
 $B = 10$

Puntos:

- Posibilidades

1º ¿Existe en checkpoint?

- NO \Rightarrow Modo Normal

$\frac{y}{y}$ Valores iniciales y Valores de M. global

- SI \Rightarrow 1º Los últimos antes del start

y luego

Los de salida a Memoria después de start

Recuperaciones

¿Existe en checkpoint?

NO

• Valores iniciales y valores en M. global

Si

• Los últimos modificados antes del start

• Los que hayan sido a M. global días después start

3) Multicamino (3 niveles)

• Índice sobre carnet

Primario + campo clave

$$N_{R1} = V(\text{carnet}) = 100\ 000$$

Disperso

$$N_R = \min \{V(\text{carnet}), b_R\} = \{100\ 000, 20\ 000\} = 20\ 000 \text{ Reg. índice}$$

$$L_R = 60 \text{ Bytes}$$

$$f_R = \left\lfloor \frac{B_{\text{util}}}{L_R} \right\rfloor = \left\lfloor \frac{325}{60} \right\rfloor = 5 \text{ Reg./Bloque}$$

$$b_R = \left\lceil \frac{M_R}{f_R} \right\rceil = \left\lceil \frac{100\ 000}{5} \right\rceil = 20\ 000 \text{ bloques}$$

$$L_{R1} = L_{\text{carnet}} + L_{\text{Reg}} = 8 + 7 = 15 \text{ Reg.}$$

$$f_{R1} = \left\lfloor \frac{325}{15} \right\rfloor = 21 \text{ Reg./Bloques}$$

$$b_{R1} = \left\lceil \frac{20\ 000}{21} \right\rceil = 953 \text{ bloques nivel 1}$$

$$L_{R2} = L_{\text{carnet}} + L_{\text{Reg/Bloque}} = 8 + 6 = 14 \text{ Bytes}$$

$$f_{R2} = \left\lfloor \frac{325}{14} \right\rfloor = 23 \text{ Reg./Bloques}$$

$$b_{R2} = \left\lceil \frac{953}{23} \right\rceil = 42 \text{ Bloques nivel 2}$$

$$b_{R3} = \left\lceil \frac{42}{23} \right\rceil = 2 \text{ Bloques nivel 3}$$

$$O_{\text{carnet}} = 2345(2)$$

$$N_{RC} = 1 \text{ Reg.}$$

• Secuencial = 20 000 Bloques

• Binaria = $\lceil \log_2(20\ 000) \rceil + \lceil \frac{1}{5} \rceil - 1 = 15 \text{ bloques}$

• Índice

$$\cdot \text{Disperso} = \lceil \log_2(2) \rceil + 1 + 1 + 1 = 4 \text{ Bloques}$$

$$\sigma_{cod_camara} = 2(n)$$

$$N_{RC} = 10\ 000 \text{ registros}$$

- Secuencial = 20 000 bloques

- No está ordenado por este campo \Rightarrow No puedo buscar por índice

2) Multivíel sobre cod_carrera

Secundario + CNC

$$N_{RI} = V(\text{cod_carrera}) = 10$$

$$L_{RI} = L_{cod_carrera} + L_{P_B \text{ Bloque}} = 2+6 = 8 \text{ bytes}$$

$$f_R = \left\lceil \frac{325}{8} \right\rceil = 40 \text{ Reg/Bloque}$$

$$b_{RI} = \left\lceil \frac{N_{RI}}{f_R} \right\rceil = \left\lceil \frac{10}{40} \right\rceil = 1 \text{ Bloque}$$

Cajones de punteros = $N = V(\text{cod_carrera}) = 10$ cajones

$$N^{\circ} \text{ punteros por cajón} = \frac{M_B}{N} = \frac{100\ 000}{10} = 10\ 000 \text{ punteros a registros}$$

$$L_{PAG} = 7 \text{ bytes}$$

$$f_{Pajón} = \left\lceil \frac{\text{Bufif}}{L_{PAG}} \right\rceil = \left\lceil \frac{325}{7} \right\rceil = 46 \text{ Punteros Reg/Bloque}$$

$$N^{\circ} \text{ bloques por cajón} = b_{cajón} = \left\lceil \frac{N}{f_{Pajón}} \right\rceil = \left\lceil \frac{10\ 000}{46} \right\rceil = 218 \text{ Bloques}$$

$$\sigma_{camet} = 2745(n)$$

$$N_{RC} = 1$$

- Secuencial = 20 000 bloques

- Binaria = 15 bloques

$$\sigma_{cod_camara} = 2(n)$$

- Secuencial = 20 000 bloques

$$\underbrace{1}_{1 \text{ cajón}} + \underbrace{218}_{\text{Reg por cajón}} + \underbrace{10\ 000}_{\text{Reg por cajón}} = 10\ 219 \text{ Bloques}$$

3) Primario + Campo Clave
 $N_{RI} = \sqrt{(\text{carmet})} = 100\ 000$

$N = 23$ punteros a bloque

$N_h = 21$ punteros a registros o valores de campo

$H_{R,0,5} = 47\ 62$ Bloq.

$N_1 = 208$ bloq

$N_2 = 9$ bloq

$N_3 = 1$ bloq

$O_{\text{carmet}} = 2345(n)$

$N_{RC} = 1$ Reg

coste

$$\underbrace{1+1+1+1}_{\text{Indices}} + \underbrace{1}_{\text{Datos}} = 5 \text{ bloques}$$

$O_{\text{cod_camera}} = 2(n)$

$$N_{RC} = NV \cdot \frac{M_R}{\sqrt{(\text{cod_camera})}} = 1 \cdot \frac{100000}{10} = 10000 \text{ Registros}$$

coste

$$\underbrace{1+1+1+1}_{\text{Indices}} + \underbrace{4762}_{\text{Datos}} = 4766 \text{ Bloques}$$