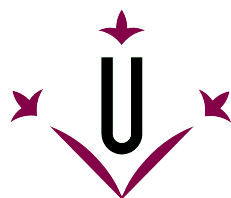

AMPLIACIÓ DE BASES DE DADES I ENGINYERIA DEL PROGRAMARI

Activitat 1: Exercici d'Optimització

Maria Florencia Martínez Malaret
Jordi Rafael Lazo Florensa
Pere Rollón Baiges

21 de març 2021

Grau en Enginyeria Informàtica



Universitat de Lleida
Escola Politècnica Superior

1 Estratègia A

```
select o.NSS, v.any_matriculació, c.capacitat
from vehicle v, camió c, conductor o
where o.anys_experiència > 15 SELECT
and c.capacitat > 3000 SELECT
and o.matricula = c.matricula JOIN
and c.matricula = v.matricula; JOIN
```

Figura 1: Estructura de la sentència SQL.

1.1 Selecció conductor ($\text{anys_experiència} > 15$)

Dades necessàries:

- $\text{card}(\text{conductor}) = 80$
- $b(\text{conductor}) = 16$
- 40% del conductors tenen més de 15 anys d'experiència.

1. Índex clúster/primari? No, perquè no hi ha índex primari a conductor per *anys_d'experiència*.
2. Índex no clúster/secundari? No, perquè no hi ha cap índex secundari.
3. Intersecció d'índexs? No, perquè només té un índex primari a conductor per *NSS* que forma part de la selecció.
4. Seqüencial? Si (sempre aplicable).

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}(R)}{b} \right\rceil = \left\lceil \frac{80}{16} \right\rceil = 5$$

$$\text{Cost}(\text{selecció conductor}) = 5$$

5. Guardar selecció conductor

$$\text{card}'(R) = \text{sl}(\text{condition}) * \text{card}(R)$$

$$\text{card}'(\text{selecció conductor}) = \text{sl}(\text{anys_experiència} > 15) * \text{card}(\text{conductor}) = 40\% * 80 = 32$$

$$\text{Blocs de la relació: } B'(\text{selecció conductor}) = \left\lceil \frac{\text{card}'(R)}{b} \right\rceil = \frac{32}{16} = 2$$

$$\text{Cost}(\text{guardar selecció conductor}) = 2$$

1.2 Selecció de camió (capacitat > 3000)

Dades necessàries:

- $v = (\text{capacitat}) > 3000$
- $\text{MAX}(\text{capacitat}) = 44.000$
- $\text{MIN}(\text{capacitat}) = 1.000$
- $\text{card}(\text{camió}) = 85$
- $b(\text{camió}) = 11$

1. Índex clúster/primari? No, perquè no hi ha índex primari a camió per *capacitat*.
2. Índex no clúster/secundari? Si.

Conté un índex secundari a camió per capacitat ($b_i=100$)

$$\text{Cost} = sl(A_i \text{ op}_i \text{ val}) * \text{card}(R)$$

Selectivitat de seleccions:

$$sl(A > v) = \frac{\text{Max}(A) - v}{\text{Max}(A) - \text{Min}(A)} \longrightarrow sl(\text{capacitat} > 3000) = \frac{44000 - 3000}{44000 - 1000} = \frac{41}{43}$$

$$\text{Cost} = \frac{41}{43} * 85 = \frac{3485}{43} = 81.046 \approx 82$$

3. Intersecció d' índexs? No, perquè només es té una condició.
4. Seqüencial? Si (sempre aplicable).

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}'(R)}{b} \right\rceil = \left\lceil \frac{85}{11} \right\rceil = 7.72 \approx 8$$

$$\text{Cost}(\text{selecció camió}) = 8$$

5. Guardar selecció camió

$$\text{card}'(R) = sl(\text{condition}) * \text{card}(R)$$

$$\text{card}'(\text{selecció camió}) = sl(\text{capacitat} > 3000) * \text{card}(\text{camió}) = \frac{41}{43} * 85 = 81.04 \approx 82$$

$$\text{Blocs de la relació: } B'(\text{selecció camió}) = \left\lceil \frac{\text{card}'(R)}{b} \right\rceil = \frac{82}{11} = 7.45 \approx 8$$

$$\text{Cost}(\text{guardar selecció camió}) = 8$$

1.3 Join entre els resultats de 1.1 i 1.2

- Bucles imbricats (s'escolleix la relació conductor per el bucle extern que te menys implicats)

$$B(\text{conductor}) = 2$$

$$B(\text{camió}) = 8$$

$$\text{Cost total} = 2 + 2 * 8 = 18$$

- Ordenació-fusió:

Camió \rightarrow ordenat per *matrícula* (si s'aplica la selecció conductor (seqüencial)).

Conductor \rightarrow ordenat per *NSS*, no per *matrícula* (per tant s'ha d'ordenar).

$$\text{Cost ordenar conductor: } B * \log_2(B) = 2 * \log_2(2) = 2$$

$$\text{Cost fusió: } B'(\text{conductor}) + B'(\text{camió}) = 2 + 8 = 10$$

Cost total: $2 + 10 = 12$

- Ordenació-fusió amb índex? No, perquè no es tenen índex ja que després de realitzar la selecció sempre es perden els índexs.
- Bucle amb índex? No, perquè no es tenen índex ja que després de realitzar la selecció sempre es perden els índexs.
- Bucle amb hash? No, perquè la taula con conté cap hash.
- Join-índex? No.

1.4 Resultats intermedis entre 1.1 i 1.2

La cardinalitat del *join* depèn de la matrícula que es *foreign key* a conductor s'escolleix els resultats intermedis de fer la selecció de conductor.

$$\text{card}(\text{selecció conductor}) = \text{sl}(\text{anys_experiència} > 15) * \text{card}(\text{conductor}) = 40\% * 80 = 32$$

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}(\text{join})}{b(\text{join})} \right\rceil = \frac{32}{6} = 5.33 \approx 6$$

$$b(\text{join}) = \frac{1}{\frac{1}{16} + \frac{1}{11}} = \frac{1}{\frac{27}{176}} = 6.51 \approx 6$$

1.5 Join del resultat de 1.3 amb vehicle

- Bucles imbricats (s'escolleix els resultats intermedis del join per el bucle extern)

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}(\text{vehicles})}{b} \right\rceil = \left\lceil \frac{100}{10} \right\rceil = 10$$

$$B(\text{join (entre la selecció de camió i la selecció de conductor)}) = 6$$

$$\text{Cost total} = 6 + 6 * 10 = 66$$

- Ordenació-fusió:

Vehicle \longrightarrow ordenat per matrícula.

Els resultats intermedis de 1.1 i 1.2 \longrightarrow estan ordenats per matrícula.

– Cost ordenar: 0

– Cost fusió: $B(join) + B(vehicle) = 6 + 10 = 16$

Cost total: 16+0 = 16

- Ordenació-fusió amb índex? No.
- Bucle amb índex? No.
- Bucle amb hash? No.
- Join-índex? No.

TOTAL ESTRATÈGIA A: 5+2+8+8+12+6+16 = 57

2 Estratègia B

2.1 Join camió-vehicle per matrícula (seleccionant camió)

- Bucles imbricats (s'escolleix la relació camió per el bucle extern que te menys implicats)

$$\text{Blocs de la relació camió: } B = \left\lceil \frac{\text{card}(\text{camió})}{b} \right\rceil = \left\lceil \frac{85}{11} \right\rceil = 8$$

$$\text{Blocs de la relació vehicle: } B = \left\lceil \frac{\text{card}(\text{vechicles})}{b} \right\rceil = \left\lceil \frac{100}{10} \right\rceil = 10$$

$$\text{Cost total} = 8 + 8 * 10 = 88$$

- Ordenació-fusió:

Camió \longrightarrow ordenat per matrícula.

Vehicle \longrightarrow ordenat per matrícula.

– Cost ordenar: 0

– Cost fusió: $B(vehicle) + B(camió) = 10 + 8 = 18$

Cost total: 0+18 = 18

- Ordenació-fusió amb índex? Si, perquè al no haver fet anteriorment cap selecció es mantenen els índexs.

La cardinalitat del *join* depèn de la *matrícula* que es *foreign key* a la relació *camió*.

$$\text{cost} = B_{ir} + B_{is} + 2 * \text{card}(\text{join}(R, S))$$

$$b_i \text{ camió per matrícula} = 100$$

$$b_i \text{ vehicle per matrícula} = 100$$

$$B_{ir} = \left\lfloor \frac{\text{card}(\text{camió})}{b_{\text{camió}}} \right\rfloor = \left\lfloor \frac{85}{100} \right\rfloor = 0,85 \approx 1$$

$$B_{is} = \left\lfloor \frac{\text{card}(\text{vehicle})}{b_{\text{vehicle}}} \right\rfloor = \left\lfloor \frac{100}{100} \right\rfloor = 1$$

$$\text{Cost total} = 1 + 1 + 2 * 85 = 172$$

- Bucle amb índex? Sí, perquè com que no hem fet anteriorment cap selecció es mantenen els índexs.

Bucle sobre camió, abans d'utilitzar l'índex podem seleccionar els camions. Recorregut seqüencial.

$$\text{Blocs de la relació: B (selecció camió)} = \left\lfloor \frac{\text{card}(R)}{b} \right\rfloor = \frac{85}{11} = 7.72 \approx 8$$

Usar índex de vehicle per matrícula.

Quants camions passaran la selecció?:

$$\text{card}(\text{selecció camió}) = \text{sl}(\text{capacitat} > 3000) * \text{card}(\text{camió}) = \frac{41}{43} * 85 = 81.04 \approx 82$$

Per cada accés a l'índex trobarem només un camió, 82 accessos a l'índex.

La cardinalitat del join depèn de l'atribut pel que fem el join i de la taula on és *foreign key*. $\text{card}(\text{join}) = \text{card}(\text{selecció camió}) = 82$

$$\text{cost} = B_s + \text{card}(S) + \text{card}(\text{join}(R, S))$$

$$\text{Cost total} = 8 + 82 + 82 = 172$$

- Bucle amb hash? No, perquè la taula con conté cap hash.
- Join-índex? No.

2.2 Resultats intermedis del join

La cardinalitat del *join* depèn de la matrícula que es *foreign key* a camió s'escolleix els resultats intermedis de fer la selecció de camió.

$$\text{card}(\text{selecció camió}) = \text{sl}(\text{capacitat} > 3000) * \text{card}(\text{camió}) = \frac{41}{43} * 85 = 81.04 \approx 82$$

$$\text{Blocs de la relació: B} = \left\lfloor \frac{\text{card}(\text{join})}{b(\text{join})} \right\rfloor = \frac{82}{5} = 16.4 \approx 17$$

$$b(\text{join}) = \frac{1}{\frac{1}{10} + \frac{1}{11}} = \frac{1}{\frac{21}{110}} = 5,23 \approx 5$$

2.3 Join del resultat de 1 amb conductor (seleccionant conductor)

- Bucles imbricats (s'escolleix la relació camió per el bucle extern que te menys implicats)

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}(\text{conductor})}{b} \right\rceil = \left\lceil \frac{80}{16} \right\rceil = 5$$

$$B(\text{join}) = 7$$

$$\text{Cost total} = 5 + 5 * 7 = 40$$

- Ordenació-fusió:

Els resultats del join 2.1 \rightarrow ordenats per matrícula.

Conductor \rightarrow ordenat per NSS, no per matrícula

$$- \text{Cost ordenar: } B * \log_2(B) = 5 * \log_2(5) = 11,60 \approx 12$$

$$- \text{Cost fusió: } B(\text{join}) + B(\text{conductor}) = 17 + 5 = 22$$

$$\boxed{\text{Cost total: } 12 + 22 = 34}$$

- Ordenació-fusió amb índex? No, ja que el *join* del resultat 1 té índex per matrícula i conductor té índex per NSS.
- Bucle amb índex? No ja que conductor no té índex per matrícula que és l'atribut del *join*.
- Bucle amb hash? No, perquè la taula con conté cap hash.
- Join-índex? No.

$$\boxed{\text{TOTAL ESTRATÈGIA B} = 18 + 17 + 34 = 69}$$

3 Estratègia C

- Definiu una estratègia d'execució diferent, i avalueu-la.
 1. Selecció conductor
 2. Join conductor-camió (seleccionant camió)
 3. Join del resultat 2 amb vehicle

3.1 Selecció conductor (anys_experiència > 15)

Valors pre-calculats a l'estratègia A.

1. Índex clúster/primari? No.
2. Índex no clúster/secundari? No.
3. Intersecció d'índexs? No, perquè Només té un índex primari a conductor per NSS que forma part de la selecció.

4. Seqüencial? Si (sempre aplicable).

$$\text{Blocs de la relació: } B = \left\lfloor \frac{\text{card}(R)}{b} \right\rfloor = \left\lfloor \frac{80}{16} \right\rfloor = 5$$

$$\text{Cost}(\text{selecció conductor}) = 5$$

5. Guardar selecció conductor

$$\text{card}'(R) = \text{sl}(\text{condition}) * \text{card}(R)$$

$$\text{card}(\text{selecció conductor}) = \text{sl}(\text{anys_experiència} > 15) * \text{card}(\text{conductor}) = 40\% * 80 = 32$$

$$\text{Blocs de la relació: } B'(\text{selecció conductor}) = \left\lfloor \frac{\text{card}'(R)}{b} \right\rfloor = \frac{32}{16} = 2$$

$$\text{Cost}(\text{guardar selecció conductor}) = 2$$

3.2 Join conductor-camió (seleccionant camió)

- Bucles imbricats? Sí, agafem la relació de conductor' (resultats de guardar selecció) pel bucle extern.

$$\text{Blocs de la relació: } B(\text{selecció camió}) = \left\lfloor \frac{\text{card}(R)}{b} \right\rfloor = \frac{85}{11} = 7,72 \approx 8$$

$$\text{Cost total} = 2 + 2 * 8 = 18$$

- Ordenació fusió?

Camió \rightarrow ordenat per matrícula

Conductor \rightarrow ordenat per NSS

$$\text{— Cost ordenar: } B * \log_2(B) = 2 * \log_2(2) = 2$$

$$\text{— Cost fusió: } B(\text{selecció conductor}) + B(\text{camió}) = 2 + 8 = 10$$

$$\boxed{\text{Cost total: } 2 + 10 = 12}$$

- Bucle amb índex: Sí, bucle amb selecció conductor i accedim al índex de camió per matrícula.

$$\text{cost} = B(\text{selecció conductor}) + \text{card}(\text{selecció conductor}) + \text{card}(\text{join}(R, S))$$

La cardinalitat del join depèn de l'atribut pel que fem el join i de la taula on és *foreign key*. $\text{card}(\text{join}) = \text{card}(\text{selecció conductor}) = 8$

$$\text{Cost total} = 2 + 32 + 8 = 42$$

- Ordenació fusió amb índex? No, ja que necessitem índexos pels atributs del join, en aquest cas matrícula i conductor no té índex per matrícula.
- Bucle amb Hash? No, no tenim cap hash.
- Join-índex? No.

3.3 Resultats intermedis del join

La cardinalitat del join depen de la matrícula que es *foreign key* a conductor s'escolleix els resultats intermedis de fer la selecció de conductor.

$$\text{card}(\text{selecció conductor}) = \text{sl}(\text{anys_experiència} > 15) * \text{card}(\text{conductor}) = 40\% * 80 = 32$$

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}(\text{join})}{b(\text{join})} \right\rceil = \frac{32}{6} = 5.33 \approx 6$$

$$b(\text{join}) = \frac{1}{\frac{1}{16} + \frac{1}{11}} = \frac{1}{\frac{27}{176}} \approx 6$$

3.4 Join del resultat 3.2 amb vehicle

- Bucles imbricats (s'escolleix els resultats intermedis del join per el bucle extern)

$$\text{Blocs de la relació: } B = \left\lceil \frac{\text{card}(\text{vechicles})}{b} \right\rceil = \left\lceil \frac{100}{10} \right\rceil = 10$$

$$\text{Cost total} = 6 + 6 * 10 = 66$$

- Ordenació-fusió:

Vehicle \longrightarrow ordenat per matrícula

Els resusltats de 3.1 i 3.2 \longrightarrow estan ordenats per matrícula

– Cost ordenació: 0

– Cost fusió: $B(\text{join}) + B(\text{vehicle}) = 6 + 10 = 16$

$$\boxed{\text{Cost total: } 16 + 0 = 16}$$

- Ordenació-fusió amb índex? No.
- Bucle amb índex? No.
- Bucle amb hash? No.
- Join-índex? No.

$$\boxed{\text{TOTAL ESTRATÈGIA C: } 5 + 2 + 12 + 6 + 16 = 41}$$