

Exercicis Optimització

lunes, 13 de febrero de 2023 11:47

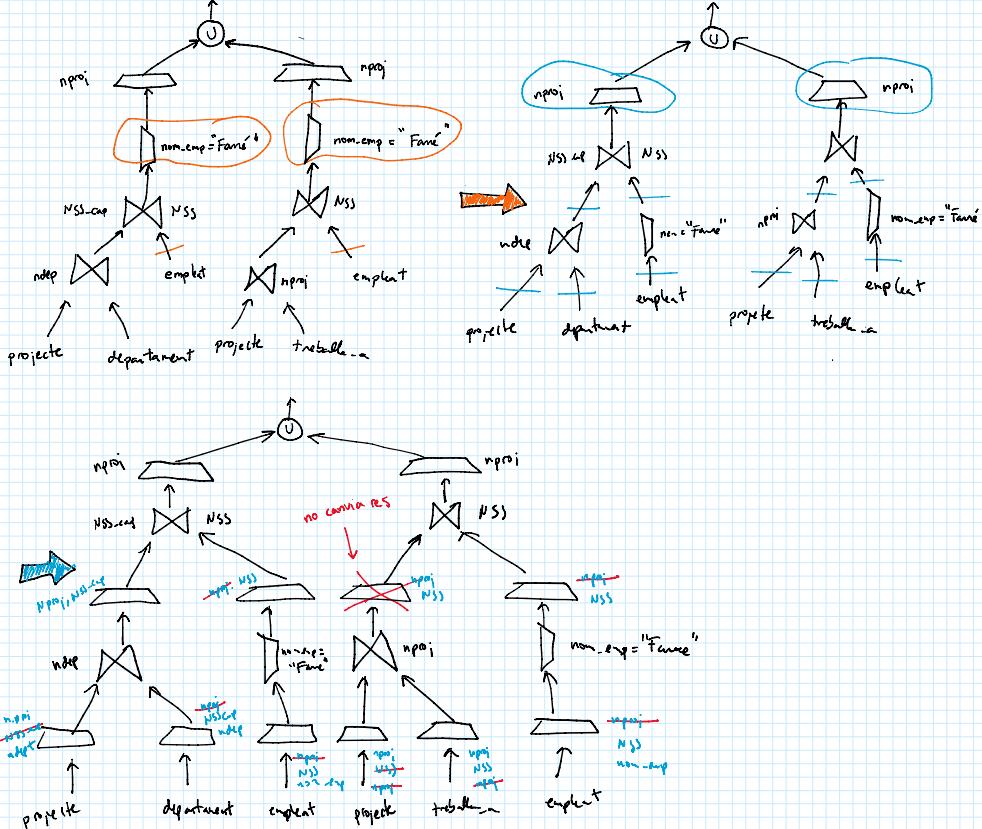
1. Disposant de la següent BD:

```
empleat(NSS, nom_emp, adreça, salari, ndept, NSS_responsable)
departament(ndept, nom_dept, NSS_cap)
projecte(nproj, nom_proj, ubicació, ndept)
treball_a(NSS, nproj, hores)
```

i de la consulta en SQL:

```
select p.nproj
from projecte p, departament d, empleat e
where p.ndept = d.ndept and d.NSS_cap = e.NSS and e.nom_emp = 'Farré'
union
select p.nproj
from projecte p, treball_a t, empleat e
where p.nproj = t.nproj and t.NSS = e.NSS and e.nom_emp = 'Farré';
```

Dibuixeu un arbre sintàctic que correspongui a la consulta i realitzeu-hi l'optimització sintàctica.



2. Suposeu una base de dades per a una cadena de llibreries que disposa del següent esquema expressat en el model relacional:

```
llibre(ISBN, titol, autor, tema, preu, estoc_total)
llibrena(CIF, nom, adreça, ciutat, codi_postal, valor_inventari)
estoc(CIF, ISBN, quantitat)
```

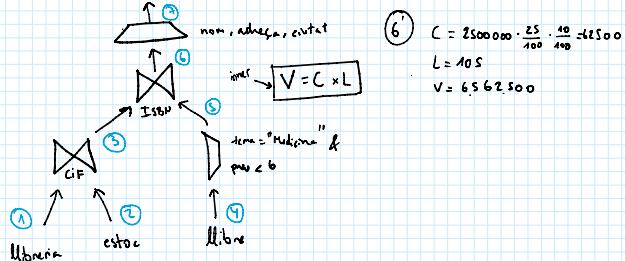
Suposeu, també, les següents dades emmagatzemades al diccionari de dades:

- Cardinat(libreria) = 1.100.000
- 3 bytes, 10 bytes, 8 bytes, 10 bytes, 5 bytes, estoc: total 6 bytes
- Cardinat(libreria) = 50
- CIF: 3 bytes, nom: 14 bytes, adreça: 20 bytes, ciutat: 6 bytes, codi_postal: 3 bytes, valor_inventari: 8 bytes
- Cardinat(estoc) = 2.500.000
- quantitat: 4 bytes
- 25% dels llibres tenen un preu inferior a 6 euros, dels quals 10% són de medicina.

i la consulta:

```
select l.nom, l.adreça, l.ciutat
from llibreria l, estoc e, llibre i
where i.CIF = e.CIF and i.ISBN = e.ISBN and i.tema = 'Medicina' and i.preu < 6;
```

Proposar un arbre sintàtic equivalent a la consulta, tot indicant el volum de dades que es traeix al llarg de tot l'àrbol.



- | | |
|------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| $\textcircled{1} \quad C = 50$
$L = 3 + 14 + 10 + 6 + 3 + 8 = 54$
$V = 50 \times 54 = 2700$ | $\textcircled{4} \quad C = 1100 \ 000$
$L = 3 + 14 + 10 + 6 + 3 + 8 = 54$
$V = 1100 \ 000 \times 54 = 59400 \ 000$ |
| $\textcircled{2} \quad C = 2.500.000$
$L = 3 + 3 + 4 = 10$
$V = 2500 \ 000 \times 10 = 25000 \ 000$ | $\textcircled{5} \quad C = 11000 \ 000 \times \frac{25}{100} = 27500$
$L = 47$
$V = 27500 \times 47 = 1292500$ |
| $\textcircled{3} \quad C = 2500 \ 000$
$L = 54 + 10 - 3 = 61$
$V = 2500 \ 000 \times 61 = 152500000$ | $\textcircled{6} \quad C = 2500 \ 000$
$L = 61 + 47 - 3 = 105$
$V = 25000 \ 000 \times 105 = 2625000000$ |

$$\textcircled{7} \quad C = 1500 \ 000$$

$$L = 14 + 20 + 6 = 40$$

$$V = 1500 \ 000 \times 40 = 60000000$$

4. Disposant de la següent BD:

Persona(idPersona, nom_p, edat, gènere, email, telèfon, adreça, ciutat)
 Estudi(idEstudi, nom_e, descripció, nivell)
 Titol(idPersona, idEstudi, any_inici, any_final)

- a) Proposar un arbre sintàctic equivalent a la següent consulta, i realitzar-hi l'optimització sintàctica

```
SELECT nom_p, email
FROM Persona p, Titol t
WHERE p.idPersona = t.idPersona
AND gènere = 'Dona' AND any_final < any_inici + 5;
```

A més a més, el diccionari de dades conté la següent informació:

- index cluster per idPersona
- index no cluster per email
- index no cluster per gènere
- NDIST(idPersona) = 25.000
- MAX(idPersona) = 25.000
- MIN(idPersona) = 1
- NDIST(gènere) = 2
- MAX(gènere) = 'Home'
- MIN(gènere) = 'Dona'
- NDIST(nom_p) = 24.38
- MAX(nom_p) = 'Xènia Vilalta'
- MIN(nom_p) = 'Abril Arnau'
- NDIST(email) = 30.000
- MAX(email) = 'xenia.vilalta@alumnes.edu'
- MIN(email) = 'abril.arnau@alumnes.edu'
- NDIST(ciutat) = 4
- MAX(ciutat) = 'Tarragona'
- MIN(ciutat) = 'Barcelona'
- NDIST(edat) = 31
- MAX(edat) = 57
- MIN(edat) = 18
- card(Persona) = 25.000

- b) Indicar quin és el millor algoritme d'execució que podria escollir l'optimitzador físic per a la següent consulta:

```
SELECT nom_p, email
FROM Persona
WHERE idPersona > 12.356 AND gènere = 'Dona' AND edat = 19;
```

5. Disposant de la següent BD:

Persona(idPersona, nom_p, edat, gènere, email, ciutat, idEstudi)
 Estudi(idEstudi, nom_e, descripció, nivell)

A més a més, el diccionari de dades conté la següent informació:

- En quan a la relació Persona:
- index cluster per idPersona, b_i = 250
 - index no cluster per edat, b_i = 300
 - NDIST(idPersona) = 30.000
 - MAX(idPersona) = 30.000
 - MIN(idPersona) = 1
 - NDIST(gènere) = 2
 - MAX(gènere) = 'Home'
 - MIN(gènere) = 'Dona'
 - NDIST(nom_p) = 24.568
 - MAX(nom_p) = 'Xènia Vilalta'
 - MIN(nom_p) = 'Abril Arnau'
 - NDIST(email) = 30.000
 - MAX(email) = 'xenia.vilalta@alumnes.edu'
 - MIN(email) = 'abril.arnau@alumnes.edu'
 - NDIST(ciutat) = 4
 - MAX(ciutat) = 'Saragossa'
 - MIN(ciutat) = 'Alacant'
 - NDIST(edat) = 31
 - MAX(edat) = 57
 - MIN(edat) = 18
 - B=75 (relació sola en el fixer)
 - card(Persona) = 30.000

En quan a la relació Estudi:

- index cluster per idEstudi, b_i = 50
- index no cluster per nivel, b_i = 75
- NDIST(idEstudi) = 125
- MAX(idEstudi) = 125
- MIN(idEstudi) = 1
- NDIST(nom_e) = 98
- MAX(nom_e) = 'Veterinària'
- MIN(nom_e) = 'Administració'
- NDIST(descripció) = 98
- MAX(descripció) = 'Zzzzzzzzzzzz'
- MIN(descripció) = ''
- NDIST(nivell) = 4
- MAX(nivell) = 4
- MIN(nivell) = 1
- B=5 (relació sola en el fixer)
- card(Estudi) = 125

- a) Proposar un arbre sintàctic equivalent a la següent consulta, i realitzar-hi l'optimització sintàctica.

```
SELECT e.nom_e, e.descripció
FROM Persona p, Estudi e
WHERE p.idEstudi = e.idEstudi AND p.ciutat = 'Lleida' AND p.edat > 35 AND e.nivell = 2;
```

- b) Indicar quin és el millor algoritme d'execució, o quina combinació d'ells, que podria escollir l'optimitzador físic per a la consulta anterior.

 Join (i les seleccions) → 380

- Bucle imbricats:
 - Bpersona = 75 (bucle intern)
 - Estudi = 5 (bucle extern)
 - Cost = 5 * 75 = 380
- Ordenació - fusió
 - Persona ordenada per id (cal ordenar)
 - Estudi ordenat per idEstudi (no cal ordenar)
 - Ordinar Persona:

$$75 \times [\log_2 75] = 75 \times 7 = 525$$
 - Cost fusió:

$$75 + 5 = 80$$

- Bucle amb index
 - Index primari a Estudi per idEstudi
 - Lectura seqüencial de Persona
 - Bpersona = 75
 - Accés a l'index d'estudi NOMBRES per les Persones que compleixen la selecció → 4231
 - Per cada una d'elles com l'index és primaria, trobarem una tupla d'estudi que fa join → l'hauria de llegir 4231
 - Cost = 75 + 4231 + 4231 = 8531

 Selecció persona → 75 + 11 + 60 = 146

- Seleccionar Persona
 - Cost = 75
- Guardar selecció Persona
 - Cost = 11 (ordenat per idPersona)
- Join del resultat anterior (selecció) amb Estudi
 - Bucle imbricats:
 - BseleccióPersona = 11 (bucle intern)
 - Estudi = 5 (bucle extern)
 - Cost = 5 * 11 = 60

$$M \times [\log_2 M] = 44$$

$$\text{fusió: } M + S = 16$$

- Bucle amb index
 - Seqüència de la selecció de persona → Bsel.Persona = 11
 - Per cada tupla de la selecció, llegir l'index → card(selPersona) = 4.231
 - Per cada idEstudi, llegir una tupla de l'estudi → 4231

* Usen index primari (cluster) per idPersona

$$\text{cost} = [sl(idPersona > 12356) + 8 \text{ persona}]$$

$$Bpersona = \left[\frac{\text{card}(persona)}{b} \right] = \frac{25000}{50} = 500, \text{ s} = 500$$

$$sl(idPersona > 12356) = \frac{25000 - 12356}{25000} = 0,505$$

$$\text{cost} = [0,505 \times 500] = [150,00] = 150$$

↑
ordenat per idpersona

* index secundari per gènere

$$\text{cost} = [sl(gènere = 'Dona') \times \text{card}(persona)]$$

$$= \frac{1}{2} \times 25000 = 12500$$

↑
ordenat per gènere

Selecció Estudi

- Index cluster per nivell? NO
- Index no cluster per nivell? SI
 - Cost = sl(e.nivell = 2) * card(Estudi) = $\frac{1}{4} \times 425 = 31,25 = 32$

- Intersecció indexes?

- No aplicable doncs la condició és simple

- Seqüencial

- Cost Bestudi = 5

- Guardar la selecció d'estudi

$$\text{and } (selció) = 32$$

$$Bestudi = \frac{125}{50} = 5 \Rightarrow bestudi = 25$$

$$\text{card } (selció) = \frac{32}{25} = 2 \quad \text{Resultat ordenat com estudi}$$

per id Estudi

Join entre els dos resultats

- Bluckles implicats
 - Bsel.Persona = 11 (bucle intern)
 - Bsel.Estud = 2 (bucle extern)

$$\text{Cost} = 2 + 2 * 11 = 24$$

- Ordenació - fusió (merge join)
 - Sel Persona ordenada per idPersona (cal ordenar)
 - Sel Estudi ordenada per idEstudi (no cal ordenar)

- Ordenar sel Persona

$$M \times [\log_2 M] = M \times 4 = 44$$

- Fusió

$$M + 2 = 46$$

- Bucle amb index?

No aplicable perquè sobre les seleccions no hi ha indexos

(A) Selecció Persona → 75 + 11 + 5 + 2 + 21 = 112

- Index cluster (primari)
 - Index cluster per ciutat? NO
 - Index cluster per edat? NO
- Índex no cluster
 - Índex no cluster per ciutat? NO
 - Índex no cluster per edat? SI

Cost = sl(p.edat > 35) * card(Persona) = $\frac{21-11}{21-11} \times \frac{30000}{39} = \frac{2}{39} \times 30000 = 16924$

B persona = $\frac{\text{card}(persona)}{b}$

$$B persona = \frac{30000}{50} = 600 \Rightarrow b persona = \frac{30000}{600} = 50$$

B selecció = $\frac{4231}{600} = \frac{4231}{400} = 10,6$

↑ ordenat per idpersona (següent)

(B) $S + 2 + 11 + 5 = 15$

- Selecció d'estudi: cost = 5
- Guardar selecció d'estudi cost = 2 (ordenat per idEstudi)
- Join del resultat anterior amb persona:
 - Bucle imbricats:
 - Bpersona = 75 (bucle intern)
 - Bseleccióestud = 2 (bucle extern)
 - Cost = 2 + 2 * 75 = 152

o Ordenació - fusió

- Persona no ordenada per idEstudi
- Selecció estudi ordenada per idEstudi
- Ordenar persona

$$75 + [\log_2 75] = 75 + 7 = 825$$

o Fusió

$$75 + 7 = 82$$

o Bucle amb index no es pot fer perquè sobre la selecció d'estudi no hi ha cap index

(C)

Selecció persona → 75 + 11 + 60 = 146

- Seleccionar Persona
 - Cost = 75
- Guardar selecció Persona
 - Cost = 11 (ordenat per idPersona)
- Join del resultat anterior (selecció) amb Estudi
 - Bucle imbricats:
 - BseleccióPersona = 11 (bucle intern)
 - Estudi = 5 (bucle extern)
 - Cost = 5 * 11 = 60

$$M \times [\log_2 M] = 44$$

$$\text{fusió: } M + S = 16$$

- (C)

8473

7. Suposeu el següent esquema relacional:
- ```

vehicle(matricula, any_matriulacio, tipus_vehicle)
moto(matricula, numero_motor)
camio(matricula, capaçat, mida_maxima_paquet)
conductor(NSS, anys_experiencia, matricula)

```
- i les següents dades enmagatzemades al diccionari de dades:
- Cardinalitat(vehicle) = 100
    - b = 10
  - Cardinalitat(moto) = 15
    - b = 15
  - Cardinalitat(camio) = 85
    - b = 85
    - DISIT(capaçat) = 50
    - MAX(capaçat) = 44.000
    - MIN(capaçat) = 1.000
    - b = 11
  - Cardinalitat(conductor) = 80
    - b = 16
  - Index primari a vehicle per matrícula (b=100)
  - Index primari a moto per matrícula (b=100)
  - Index primari a camió per matrícula (b=100)
  - Index secundari a camió per capacitat (b=100)
  - Index primari a conductor per NSS (b=100)

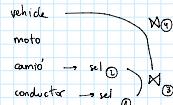
axi com la consulta

```

select o.NSS, v.any_matriulacio, c.capaçat
from vehicle v, camio c, conductor o
where o.anys_experiencia > 15
and c.capaçat > 3000
and o.matricula = c.matricula
and c.matricula = v.matricula;

```

- Els demana:
- Proposar un altre sintàctic equivalent a la consulta, optimitzat sintàcticament.
  - Indicar quin és el millor algoritme d'execució de la consulta que podria escollir un optimitzador físic tenint en compte les diferents operacions que s'han d'executar, tenint en compte que:
    - 40% dels conductors tenen més de 15 anys d'experiència.



### 1. Selecció conductor anys-experiencia

- Index cluster anys-exp? NO
- Index no cluster anys-exp? NO
- Intersecció indexos NO APLICABLE
  - No té indexos
  - Condició simple
- Sequencial

$$\text{cost} = \frac{B}{b} = \frac{80}{16} = 5$$

- Guardar selecció del conductor

$$\text{cond (selecció)} = \frac{\text{sel (anys-exp > 15)} * 80}{b_{\text{conductor}}} = \frac{40\% * 80}{16} = \frac{32}{16} = 2$$

### 2. Selecció de camió per capacitat

- Index cluster per capacitat? NO
- Index no cluster per capacitat? SI
  - Cost =  $\lceil \frac{3000}{50} \rceil = 60$
- Intersecció d'indexos NO (condició simple)
- Sequencial:
  - Cost =  $B_{\text{camio}} = 85/11 = 7,727 \approx 8$
- Guardar selecció de camió
  - Card(resultat) = 82
  - Cost =  $82/11 = 7,4545 \approx 8$  (ordenat per matrícula)

### 3. Join de les seleccions (per matrícula)

- Buclecs imbricats:
  - BselectConductor = 2 (extern)
  - BselectCamio = 8 (intern)
  - $2 * 8 = 16$
- Ordenació - fusió
  - Select conductor = ordenat per NSS -> ordenar
  - Selecció camió ordenat per matrícula -> ja ordenat
- Cost ordenar selecció conductor:
 
$$2 * \lceil \log_2 2 \rceil = 2$$
- Cost fusió:
 
$$2 + 8 = 10$$
- Cost de guardar el join
 

Cond (join) → -matrícula és foran a conductor  
b join - matrícula és prouva a camió

$$\text{cond (join)} = \text{cond (sel conductor)} = 32$$

$$\frac{1}{16} = \frac{1}{16} + \frac{1}{8} = b_i = 6,15 \approx 6$$

cost =  $\frac{32}{6} = 6$

ordenant per matrícula