



TD 02 - Optimisation des requêtes

Exercice 01 :

Pour chacune des requêtes suivantes concernant les relations : $R(A,B,C,D)$ et $T(C,D,E)$ dites si elle vous semble optimale et pourquoi. Si la requête ne vous semble pas optimale, en donner une formulation meilleure.

- $\sigma_{A=2}\pi_A(R)$
- $\pi_C\sigma_{A=2,E=3}(R \alpha T)$

Exercice 02 :

Soit la base de données suivante :

Livre(num,prix,reduc) ; **DVD**(num,prix,reduc) ; **Info**(num,titre,année) ;

Le pourcentage possible des réductions (reduc) à comme valeurs possible { 30,40,50,60,70 %}, les numéros des livres sont tous différent de numéros des DVD. La relation Info donne le titre et l'année d'édition des livres et des DVD. Les produits ont tous été édités dans les 10 dernières années. On donne les cardinalités suivantes :

$$\text{Card(Livre)} = 10000 \quad - \quad \text{Card(DVD)} = 5000 \quad - \quad \text{Card(Info)} = 15000$$

Soit la requête R1 suivante :

$$\pi_{\text{num,titre,prix}} \left[\sigma_{\text{reduc} \geq 60\%} \left(\sigma_{\text{annee}=2007} \left((\text{Livre} \cup \text{DVD}) \bowtie_{\text{num}} \text{Info} \right) \right) \right]$$

- Calculer les Facteurs de sélectivité :

$$\sigma_{\text{reduc} \geq 60\%}(\text{Livre}) - \sigma_{\text{reduc}=30\%}(\text{DVD}) - \sigma_{\text{année}=2002}(\text{Info})$$

- Calculer les cardinalités suivantes :

$$(\text{Livre} \cup \text{DVD}) \bowtie_{\text{num}} \text{Info} - R1 - (\sigma_{\text{reduc} \geq 60\%}(\text{Livre})) \bowtie_{\text{num}} (\sigma_{\text{annee}=2007}(\text{Info}))$$

- Dessiner l'arbre algébrique et donner un plan d'exécution pour évaluer R1 de telle sorte que : l'union soit traitée après toutes jointures et les sélections et les projections soient traitées le plus tôt possible.

Exercice 03 :

Soit le schéma suivant :

Emp (ne, salaire, age, ns) -- un employé est identifié par son numéro ne

Service (ns, np, budget, statut) -- ns fait référence à un numéro de service.

Projet (np, code, libellé) -- un projet est identifié par son numéro np

La taille d'un n-uplet est de 20 octets pour Emp, 40 octets pour Service, 2000 octets pour Projet. Les attributs ne, ns et np ont chacun 4 octets. La cardinalité des relations est de 20 000 pour Employés, 5000 pour Service et 1000 pour Projet. La relation Service représente l'association N-M d'un service avec un projet. La clé de Service est composée des attributs ns et np (ainsi, l'attribut ns n'est pas unique dans Service). Chaque Service, identifié par ns à en moyenne 10 Projets. Les données sont stockées sur disque dans des pages de 4000 octets. Les attributs sont indépendants et leur distribution est uniforme.

Soient les fonctions auxiliaires :

ntp(R) le nombre de tuples par pages pour la relation R,

D(R, c) le nombre de valeurs distinctes du domaine de l'attribut R.c,

largeur(R) la taille d'un tuple de R,

et **arr(x)** l'arrondi de x par excès à une valeur entière,

Un index sur l'attribut c est dit **plaçant** si les données sont triées sur le disque dans l'ordre des valeurs de c. Un index sur l'attribut c est dit **non plaçant** si les données ne sont pas triées sur le disque dans l'ordre de c.

Le modèle de coût estime le nombre de pages à lire et écrire, sans prendre en compte l'écriture du résultat final. Le coût d'une lecture séquentielle de la relation R est égal au nombre de pages de R (noté page(R)).

Le coût d'une sélection avec un prédicat **pred** de la forme **a op v** où **a** est un attribut numérique et **op** est l'opérateur = (égal), < (inférieur à) ou > (supérieur à), est :

$coût(\sigma_{pred}(R)) = card(\sigma_{pred}(R))$ si l'attribut a est indexé par un index non plaçant,

$coût(\sigma_{pred}(R)) = arr(page(R) * card(\sigma_{pred}(R)) / card(R))$ si l'attribut a est indexé par un index plaçant,

$coût(\sigma_{pred}(R)) = page(R)$ si l'attribut a n'est pas indexé.

Question 1

- 1) Donner le nombre de services distincts et le nombre moyen de services par projet
- 2) Donner le nombre de tuples par page des 3 relations.
- 3) Donner la taille des relations en nombre de pages.

Question 2

Rappel des formules évaluant la cardinalité de la sélection et de la jointure :

$$card(\sigma_F(R)) = SF(\sigma_{(F)}) * card(R)$$

où $SF(\sigma_{A=valeur}) = 1 / D(R, A)$

$$SF(\sigma_{A > valeur}) = (\max(A) - valeur) / (\max(A) - \min(A))$$

$$SF(\sigma_{A < valeur}) = (valeur - \min(A)) / (\max(A) - \min(A))$$

$\text{card} (R \bowtie_{A=B} S) = \text{card}(S)$ si A est clé de R, et B est clé étrangère de S.

sinon $\text{card} (R \bowtie_{A=B} S) = \text{SFj} * \text{card}(S) * \text{card}(R)$ où $\text{SFj} = 1 / \max(D(R, A), D(S, B))$

On donne $\text{SFj} = 1/500$ pour la jointure entre Emp et Serv.

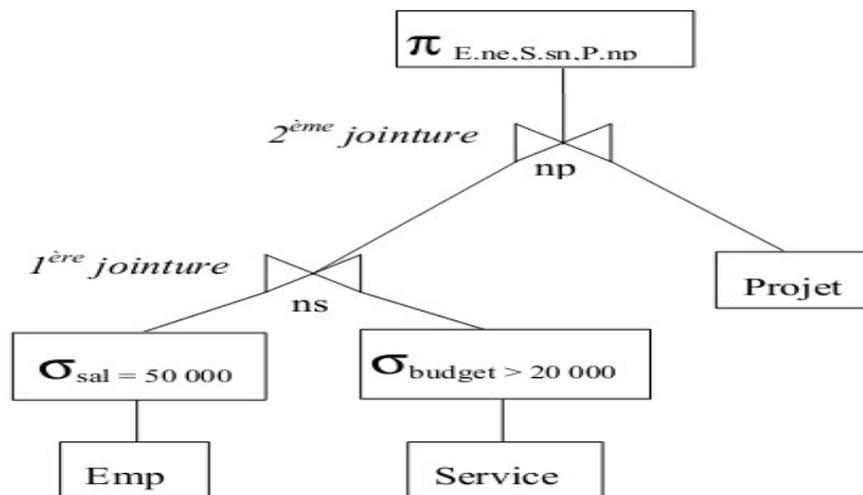
Soit la requête :

```

Select E.ne, S.ns, P.np
From Emp E, Service S, Projet P
Where E.Salaire = 50000
and S.budget > 20000
and E.ns=S.ns and S.np=P.np
    
```

On suppose une répartition uniforme des salaires (par tranche de 5000 dans l'intervalle [10000, 105000]) et des budgets (dans l'intervalle [10000, 30000]). Il y a un index plaçant sur l'attribut salaire de Emp, sur l'attribut ns de Service et sur l'attribut np de Projet.

On considère le plan d'exécution P1 suivant :



1) Estimez le nombre de n-uplets résultant de chacune des opérations.

$\text{card} (\sigma_{\text{sal} = 50\,000} (\text{Emp})) = \dots$

$\text{card} (\sigma_{\text{budget} > 20\,000} (\text{Service})) = \dots$

$\text{card} ((\sigma_{\text{sal} = 50\,000} (\text{Emp})) \bowtie_{\text{ns}} (\sigma_{\text{budget} > 20\,000} (\text{Service}))) = \dots$

$\text{card} (\text{résultat final}) =$

2) Calculez le coût du plan d'exécution P1, en précisant les algorithmes de jointure utilisés.

coût ($\sigma_{sal = 50\,000}$ (Emp)) = ...

coût ($\sigma_{budget > 20\,000}$ (Service)) = ...

coût (1^{ère} jointure) = ...

coût (2^{ème} jointure) = ...

coût total = ...