

TD8

Evaluation de requêtes - Optimisation

Exercice 1

Soit la requête :

```
select titre
from livres
where annee = 2009 or (annee = 2009 and id_auteur = 'Gard');
```

posée sur la relation de schéma suivant :

```
livres (cote, isbn, titre, id_auteur, annee)
```

On suppose qu'il existe un index B+ sur l'attribut `annee` et que l'optimiseur connaît deux implantations de l'opérateur de sélection : un algorithme de parcours séquentiel et un algorithme utilisant un index B+, et une seule implantation de l'opérateur de projection, sous sa forme la plus simple (algorithme sans élimination de doublons ni tri).

Q1. Expliquer ce que produirait la phase « analyse/traduction » de l'évaluation de cette requête.

Q2. Donner tous les plans possibles pour cette requête.

Q3. Quel plan semble le « meilleur » intuitivement ? Évaluez sa complexité en fonction du nombre n de pages disque utilisées pour stocker la relation `livres`.

Exercice 2

On considère les relations :

```
livres (cote, isbn, titre, id_auteur, annee)
auteurs (id_auteur, nom, prenom, labo, email)
```

et la requête :

```
R1 : select * from livres natural join auteurs
where annee = 2009 and labo = 'LINA';
```

On suppose que :

- les relations `livres` et `auteurs` occupent respectivement n et m pages sur disque.
- l'optimiseur connaît trois implantations de la jointure : sous la forme de 2 boucles imbriquées ; sous la forme de 2 boucles imbriquées utilisant un index sur la table intérieure ; sous la forme d'un tri-fusion.
- qu'il n'y a qu'une seule implantation de la sélection.

Q1. Donner tous les plans possibles pour cette requête, en rappelant l'ordre de grandeur du nombre d'accès disque pour chacun des opérateurs élémentaires.

Q2. Quelles informations manque-t-il pour pouvoir calculer de manière plus précise le nombre total de pages accédées ?

Exercice 3

(d'après TD, Université Pierre et Marie Curie) On considère de nouveau les relations :

livres (cote, isbn, titre, id_auteur, annee)

auteurs (id_auteur, nom, prenom, labo, email)

où id_auteur est une clé étrangère dans la relation livres. Les hypothèses sont :

- le nombre de tuples de la relation livres est $T_l = 200000$; le nombre de tuples de la relation auteurs est $T_a = 10000$.
- la relation livres mémorise des tuples dont la valeur de l'attribut annee est comprise entre 2000 et 2010.
- la taille en octets des attributs de la relation livres est de 2, 3, 4, 2 et 1 octets respectivement pour les attributs cote, isbn, titre, id_auteur, annee.
- la taille en octets des attributs de la relation auteurs est de 2, 4, 4, 3 et 4 octets respectivement pour les attributs id_auteur, nom, prenom, labo, email.
- les valeurs des attributs sont indépendantes et la distribution de ces valeurs dans leur domaine est uniforme.
- le nombre de valeurs distinctes des attributs de la relation livres est noté nd1, nd2, nd3, nd4, nd5 respectivement. Il est noté nd'1, nd'2, nd'3, nd'4, nd'5 pour les attributs de auteurs.

Q1. Proposer des valeurs pour les paramètres nd1, nd2, nd3, nd4, nd5, nd'1, nd'2, nd'3, nd'4 et nd'5.

Q2. Pour les requêtes suivantes, donner la cardinalité (nombre de tuples) et la taille en octets du résultat :

1. `select id_auteur, nom, prenom from auteurs ;`
2. `select * from auteurs where labo = 'LINA';`
3. `select * from auteurs a, livres l where a.id_auteur = l.id_auteur;`
4. `select * from livres where annee >= 2007 ;`

Q3. Quelles opérations produisent un résultat de taille supérieure à la taille de leurs opérandes ?

Q4. On considère les deux formes suivantes de la même recherche :

- `select a.nom from auteurs a, livres l where a.id_auteur = l.id_auteur and annee = 2009;`
- `select a.nom from auteurs a, livres l where annee = 2009 and a.id_auteur = l.id_auteur;`

a. Donner leur arbre syntaxique en supposant que les opérateurs disponibles sont la sélection, la projection, la jointure, l'union et la différence et que les arbres sont construits en respectant l'ordre des prédicats de la clause `where`.

b. Calculer le volume en octet de l'information manipulée pendant leur exécution si celle-ci suit l'arbre syntaxique. Conclure.

Q5. Une des heuristiques utilisée pour la recherche d'un plan d'exécution consiste à effectuer le plus tôt possible les sélections puis les projections, avant les jointures. Pour cela, donner les règles de transformation que l'on peut utiliser.