

Bases de Données Réparties- MI034 – 1^{er} examen réparti du 17 mars 2015

Exercice 1 : Arbres B+

6 pts

Dans les 3 premières questions, tous les arbres sont d'ordre **1** (i.e., il y a 1 ou 2 valeurs par nœud). On utilise la syntaxe suivante pour représenter un nœud de l'arbre : $N(v_1, v_2, \dots)$ où N est le nom du nœud et les v_i sont les valeurs. Quand la feuille F déborde, on garde les 2 plus **petites** valeurs dans F , la plus grande valeur sera dans la nouvelle feuille. S'il faut choisir une valeur pour un nœud intermédiaire, la choisir, autant que possible, identique à une valeur existant dans une feuille.

1) Soit l'arbre A_0 composé d'une racine $N_1(20)$, et de deux feuilles $F_1(16, 18)$ et $F_2(20)$. A_0 a 2 niveaux.

On insère successivement dans A_0 les **nombre pairs consécutifs décroissants** $\{14, 12, \dots\}$ jusqu'à ce que l'arbre gagne un niveau. On obtient A_1 après l'insertion ayant provoqué l'ajout du 3^{ème} niveau. Quels sont les nœuds de A_1 et leur contenu ?

2) Soit l'arbre B_0 composé d'un seul nœud (servant de feuille et de racine) $N_1(30)$. B_0 a 1 niveau.

On insère dans B_0 successivement les nombres pairs consécutifs **croissants** $\{10, 12, \dots\}$ jusqu'à ce que l'arbre gagne deux niveaux. On obtient B_1 après l'insertion ayant provoqué l'ajout du 3^{ème} niveau. Quels sont les nœuds de B_1 et leur contenu ? Rmq : les nœuds créés sont nommés N_2, N_3, \dots

3) Lors d'une suppression, on considère si possible la redistribution à gauche puis à droite, seulement entre des nœuds ayant le même père. L'arbre C_0 est composé :

d'une racine R

de 4 valeurs $\{3, 5, 9, 11\}$ dans les nœuds intermédiaire nommés N_i

et de 10 valeurs $\{1, 2, 3, 4, 5, 6, 7, 9, 11, 12\}$ dans les feuilles nommées F_j

a) Que contient R ?

b) On supprime successivement dans C_0 les valeurs 1 et 2. On obtient C_1 Représenter seulement les nœuds modifiés.

c) On supprime successivement dans C_0 les valeurs 1, 2, 3, 4, 5. Représenter l'arbre C_2 obtenu.

d) On supprime successivement dans C_0 les valeurs 4, 5, 6, 7, 11, 12. Représenter l'arbre C_3 obtenu.

4) On considère un arbre d'ordre **50** (50 à 100 valeurs dans une feuille ou un nœud intermédiaire).

a) Combien de niveau faut-il au minimum pour indexer un million de valeurs dans les feuilles ? Préciser le nombre de nœuds dans chaque niveau.

b) Si on insère successivement les valeurs dans l'ordre croissant $(2, 3, \dots, 10^6)$, dans un arbre d'ordre 50 contenant seulement la racine $R(1)$. Quel sera le nombre de niveaux de l'arbre obtenu ?

Exercice 2 : Table de hachage extensible

6 pts

Dans toutes les tables de hachage considérées, un paquet ne peut pas contenir plus de **4** valeurs. Lors d'une suppression, si un paquet devient vide, on tente de le fusionner avec un autre paquet, si ce n'est pas possible, il reste vide.

Utiliser la syntaxe suivante :

$L(v_1, \dots, v_n)$ PL= n avec L le nom (A, B, ...) du paquet, n la profondeur locale et les valeurs v_i .

Rmq : utiliser la lettre suivante de l'alphabet pour nommer un nouveau paquet.

$R[L, L', \dots]$ pour le répertoire avec L, L', \dots les noms des paquets.

On peut aussi préciser le contenu d'une case particulière avec $R[i]=L$ (avec $R[0]$ étant la 1^{ère} case).

1) La table de hachage T_0 contient *seulement* les valeurs 1, 4, 5, 6, 9, 14, 25. De combien de paquets a-t-on besoin au **minimum**? Préciser leur contenu, leur PL, le contenu du répertoire.

2) On insère 7 dans T_0 . On obtient T_1 . Représenter T_1 . Préciser ce que contient chaque case du répertoire R

3) On insère 13 dans T_1 obtenu à la question précédente. On obtient T_2 . Représenter T_2 .

4) On considère la table T_3 avec les paquets A(10), B(16), C(17,31), D(30), E(36). Représenter T_3 . Rmq : vous devez préciser le contenu de chaque case du répertoire ($R[0]$ est différent de A, la taille du répertoire est différente de 5).

5) On supprime 10 dans T_3 . On obtient T_4 . Représenter T_4

6) On supprime 30 dans T_4 obtenu à la question précédente. On obtient T_5 . Représenter T_5 .

Exercice 3 : Optimisation de requêtes

5 pts

Une base contient

Hotel(hid, nom, catégorie, ville, tel) // *hid* est un numéro d'hôtel

Résa (client, num, date, nbjours) // le client a réservé une chambre pour *nbjours* à l'hôtel *num*.

Les catégories de prix vont de 1 à 5, la durée maximale de réservation est de 30 jours. Il y a 100 villes. On suppose la distribution uniforme des valeurs d'un attribut. Les attributs sont indépendants.

On a :

card(Résa)=10 000 et card(Hotel) = 1000

Soit la requête R1 suivante :

```
select *
from Resa r, Hotel h
where h.hid = r.num
and r.nbjours between 1 and 15
and h.catégorie = 3 and h.ville='Lyon';
```

Pour simplifier la notation les prédicats sont appelés p_i comme suit :

$p1$: $h.hid = r.num$ $p2$: $r.nbjours$ between 1 and 15
 $p3$: $h.catégorie = 3$ $p4$: $r.ville='Lyon'$

- 1) Exprimer la requête R1 en algèbre relationnelle
- 2) Quels sont les facteurs de sélectivités des sélections exprimées dans R1

2) Détailler le calcul de card(R1)

3) On considère les expressions suivantes pour évaluer R1.

E1 : $\sigma_{p4}(\sigma_{p3}(\sigma_{p2}(\text{Résa} \bowtie_{p1} \text{Hotel})))$

E2 : $[\sigma_{p4}(\sigma_{p3}(\text{Hotel}))] \bowtie_{p1} [\sigma_{p2}(\text{Résa})]$

E3 : $\sigma_{p2}([\sigma_{p4}(\sigma_{p3}(\text{Hotel}))] \bowtie_{p1} \text{Résa})$

La métrique utilisée pour estimer le coût des plans d'exécution est le nombre de tuple. On rappelle les formules de coût vues en TD (ex. 2 pages 9) :

coût($\sigma_p(R)$) = card($\sigma_p(R)$) si l'attribut du prédicat p est indexé
 = card(R) sinon

coût($R \bowtie_a S$) = card(R) si $S.a$ est indexé
 = card(R) \times card(S) sinon

On *néglige* le coût des opérateurs traités en pipeline et l'écriture éventuelle de résultats intermédiaires.

- a) Il n'y a aucun index. Calculer le coût des plans E1 à E3 et dire quel est le meilleur ? Justifier en détaillant le mode de calcul pour chacun des plans
- b) Même question que a) mais en considérant qu'il y a un index sur Hotel.hid et un autre sur Resa.num.

Exercice 4 : Questions diverses

3 pts

1. Soit l'objet A obtenu par une copie en surface de l'objet B. A est-il égal en profondeur à B ? Justifier.
2. Rappeler les relations logiques qui existent entre les propositions « A égal en surface à B », « A identique à B », « A égal en profondeur B »
3. Pourquoi l'organisation d'une relation sous forme d'arbre-B+ est-elle plus efficace pour les requêtes par intervalle que l'organisation sous forme de fichier haché (statiquement ou dynamiquement) ?
4. Rappeler ce qu'est la cohérence à terme des répliques (préciser, notamment, pour quel type de réplification cette notion a-t-elle un sens)
5. La cohérence à terme est-elle une propriété applicable à une transaction ou bien à un ensemble de transactions ? Justifier
6. Dans un contrôle de concurrence par estampillage, si une transaction est abandonnée, est-elle relancée avec la même estampille ou avec une nouvelle estampille ? Justifier