M1 - Bases de Données Réparties
Devoir sur table du 27 mars 2007

Eléments de corrigé

Sans document – Durée : 1 heure

Exercice 1 : Index : Arbre B+ 7 pts

La relation **M**(m, x, y, z, t) contient 10^9 nuplets.

Question 1. Le nombre de valeurs distinctes de *M.t* est 1 200 000. L'attribut *M.t* est indexé par un arbre B+ d'ordre 500.

a) Combien de niveaux (en incluant le niveau de la racine) possède cet index ?

3 niveaux

b) L'index est construit en minimisant le nombre de nœuds. Quel est le nombre de clés dans la racine?

1 seule clé

c) L'index est maintenant construit en **maximisant** le nombre de nœuds. Quel est le nombre de clés dans la racine?

: 3 clés

Question 2. L'attribut *M.x* est indexé par un arbre B+ d'ordre 1. Le nombre de valeurs distinctes de *M.x* est 30 parmi les entiers [1,30]. L'index est construit en minimisant le nombre de nœuds.

a) Combien de niveaux (en incluant le niveau de la racine) possède cet index ?

4 niveaux

b) Quels nœuds faut-il traverser depuis la racine pour atteindre la clé 14 ? Pour chaque nœud traversé, donner les clés qu'il contient.

Il faut traverser (19)→(7,13)→(15,17)→(13,14)

Question 3. L'attribut *M.y* est indexé par un arbre B+ d'ordre 3. L'attribut *M.y* a 1801 valeurs entières distinctes dans [0, 1800]. L'index est construit en minimisant le nombre de nœuds.

a) Combien de feuilles possède l'arbre ?

301 feuilles

b) Les feuilles sont remplies en maximisant le nombre de feuilles pleines. Combien de feuilles ne sont pas pleines ?

2 feuilles non pleines

c) Le niveau juste au-dessus des feuilles contient K nœuds nommés N_0 , N_1, ..., N_i, ..., N_{K-1} avec $0 \leq i < K$

Quelles est la première valeur contenue dans N_i (*i.e.* la plus petite valeur dans le nœud N_i)? Donner une réponse en fonction de i, sous la forme : a*i + b

42i+6

Question 4. L'attribut *M.z* est indexé par un arbre B+. L'attribut *M.z* a 10^6 valeurs entières distinctes dans [1, 10^6]. Combien de pages, contenant des nuplets de M, faut-il lire pour répondre aux requêtes suivantes ?

a) R1 : Select * from M where z = 3.10^5

1000 pages

b) R2 : Select count(*) from M where z <= 10^4 .

: 0 page

Il suffit de parcourir les feuilles de l'index puis de compter le nombre de références vers des nuplets, sans nécessité de lire les nuplets par la suite.

c) R3 : Select count(distinct z) from M where z <= 10^4 .

: 0 page

Exercice 2 : Hachage extensible 7 pts

On considère une table de hachage extensible. La profondeur globale de son répertoire est PG=5.
On nomme certains paquets A, B, C, D. On donne la profondeur locale (PL) des paquets :

Paquet A : PL = 2
Paquet B : PL = 1
Paquet C et D : PL = 4
PL = 5 pour tous les autres paquets.

On décrit partiellement le contenu du répertoire :

la 1ère case du répertoire référence le paquet A,
la 2ème case référence B,
la 3ème case référence C,
la 11ème case référence D.

Question 1

a) Quelle est la taille du répertoire ?

: 32

b) Combien de cases du répertoire sont associées avec le paquet B ?

Réponse : 16 cases

c) La 7ème case référence-t-elle un des paquets A à D ? Répondre par A, B, C, D ou Autre.

7ème case C6 pointe vers un autre paquet

d) Au total, combien de paquets possède la table de hachage ?

: 8 paquets

Question 2 : Aucun paquet n'est plein.

a) On insère la valeur 122. Dans quel paquet est-elle insérée ? Répondre par A, B, C, D ou Autre.

Insertion dans le paquet D

b) On insère la valeur 46. Dans quel paquet est-elle insérée ? Répondre par A, B, C, D ou Autre.

Insertion dans le paquet F (Autre)

Question 3 : On suppose que chaque paquet contient exactement 2 valeurs (les plus petites possible).

a) Parmi tous les paquets, quelle est la plus grande valeur présente ?

Réponse : 62

b) Quelles sont les valeurs à supprimer pour provoquer la division par 2 de la taille du répertoire, *i.e.* telle que PG devienne égale à 4 ? Répondre en indiquant les valeurs dans l'ordre croissant.

Réponse : 22, 30, 54, 62

c) Suite à cette division, quelle sera maintenant la profondeur locale de C?

PL = 4

Question 4 : Un paquet peut contenir jusqu'à 4 valeurs. On veut insérer des valeurs dans le paquet A de l'arbre obtenu à la question précédente (question 3) pour que le répertoire retrouve sa taille initiale (*i.e.* telle que PG = 5).

a) Combien d'éclatements successifs du paquet A sont nécessaires ?

2 éclatements

b) Au total, combien de paquets contient la table ainsi obtenue ?

:9 paquets

c) Combien de valeurs, au minimum, faudra-t-il insérer ? Donner la liste des valeurs à insérer (si possible, choisir des petites valeurs).

Insérer 4 valeurs :Réponse : 16, 32, 48,64

Exercice 3 : Optimisation des requêtes : jointures 6 pts

Question 1 : Soient les relations A(a, x) et B(y, a) stockées dans la base sans être triées. Il existe seulement des index non plaçant sur A.a et B.a. Les valeurs des attributs *a* sont des entiers positifs. On note :

- Card(R) : la cardinalité de R
- D(R.a) : le nb de valeurs distinctes de l'attribut R.a
- Page(R) : le nb de pages de R

On suppose que D(A.a) > 100 et D(B.a) > 40.

On calcule le coût des requêtes en nombre de pages de données lues (on néglige le coût de lecture des nœuds d'un index).

a) La requête R1 est : select * from B order by a. Quel est le coût pour traiter R1 en utilisant l'index B.a?

Card(B)

b) La requête R2 est : $\sigma_{a=10}$ (A). Quel est le coût pour traiter R2 en utilisant l'index A.a?

Cout = card(A) / D(A.a)

c) La requête R3 : A $\bowtie_a \sigma_{a=4}$ (B). Quel est le coût pour traiter R3 en utilisant les deux index A.a et B.a ?

Réponse : card(A) + card(B) / D(B.a)

Question 2 : Soient 4 relations : **A**(<u>a</u>, x) **B**(y, a) **C**(z, a) **D**(t, a).

Les clés sont soulignées ; deux attributs de même nom sont liés par une contrainte d'intégrité référentielle (par exemple, B.a fait référence à A.a). On traite les requêtes en effectuant seulement des jointures et sans aucun produit cartésien.

a) Combien d'ordre de jointure sont possibles pour traiter la requête suivante ? : A $\bowtie_a$ B $\bowtie_a$ C $\bowtie_a$ D

: 12

b) On modifie le schéma des relations B, C, D pour avoir :

A(<u>a</u>, x) **B**(<u>b</u>, a) **C**(<u>c</u>, b) **D**(<u>d</u>, c).

Combien d'ordre de jointure sont possibles pour traiter la requête suivante ? : A $\bowtie_a$ B $\bowtie_b$ C $\bowtie_c$ D

: 6