

Bases de Données

Travaux dirigés
Arbres B+

Exercice 1 :

- 1 Construire un arbre B+ contenant les clés suivantes: (2,3,5,7,11,17,19,23,29,31).
On suppose que l'arbre est initialement vide, que les clés sont insérées dans l'ordre croissant leur valeur, et que chaque nœud peut contenir au plus 4, 3 ou 2 clés.
- 2 Décrire les différentes étapes composant les opérations suivantes:
 - a) Recherche des enregistrements de clé 11.
 - b) Recherche des enregistrements dont la clé est comprise entre 7 et 17 inclus.
 - c) Insertion de la clé 9, de la clé 10, de la clé 8, puis de la clé 40.
 - d) Suppression de la clé 29, puis de la clé 5.

Exercice 2 :

On se propose d'étudier le comportement d'un B+-arbre dans lequel on va procéder à des mises-à-jour.

L'état initial du B+-arbre est le suivant : les blocs B1, B2, B3 et B4 contiennent chacun trois pointeurs et au plus deux valeurs, ce sont des noeuds du B-arbre; les blocs B5 à B11 sont les feuilles du B-arbre et contiennent au plus trois valeurs. La notation *Bi signifie pointeur sur Bi , * signifie que le pointeur est vide, et la notation - signifie qu'il n'y a pas de valeur.

- 1 Représentez le B-arbre suivant : B1= (*B2, 25, *B3, 144, *B4), B2= (*B5, 9, *B6, -, *), B3= (*B7, 64, *B8, 100, *B9), B4= (*B10, 196, *B11, -, *), B5= (1,4,-), B6=(9,16,-), B7=(25,36,49), B8= (64,81,-), B9=(100,121,-), B10= (144,169,-), B11= (196,225,256).
- 2 Insérez 21, puis 11, puis 22. Insérez 32. Supprimez 4, puis 1, puis 16, puis 25. Revenez à la situation après l'insertion de 32 et supprimez 64.
- 3 Représenter par un diagramme les algorithmes de mise à jour des B+-arbres.
? Peux faire l'objet d'un mini-mémoire : programmation des algorithmes de mise-à-jour d'arbres B+, en paramétrant la taille des blocs.

Exercice 3 :

La relation $R(A,B,C,D,E)$ est organisée en B+-arbre selon l'attribut A. Des index denses (toutes les valeurs de clés sont représentées dans l'index) organisés en B+-arbres sont associés respectivement aux attributs B, D et E. Le nombre moyen de n-uplets auxquels on accède à partir d'une valeur de l'attribut d'indexation est de 10 pour B, 20 pour D et 100 pour E. La relation comporte 10000 n-uplets stockés sur 1000 pages, parmi lesquelles on suppose la répartition des n-uplets uniforme.

- 1 On suppose que l'attribut A est stocké sur 6 octets, l'attribut B sur 4, l'attribut C sur 10, l'attribut D sur 14, et l'attribut E sur 16, et qu'une adresse de page est stockée sur 4 octets. On suppose que les pages de la base ont 500 octets, disponibles pour stocker des données, et qu'elles sont en moyenne pleines aux 3/5.

Décrivez complètement toutes les structures de données et évaluez le nombre de pages de l'index (non dense) sur A. Mêmes questions pour les autres index (denses).

- 2 Proposer différentes méthodes pour évaluer la requête :

$?A (?A ??0 ?? ?? ?D?? ?R)$

- 3 Evaluer le nombre d'entrées/sorties de pages nécessaires dans chacun des cas et déduisez la meilleure manière de procéder.

Exercice 4 :

Un organisme gère des références bibliographiques. La base de données qu'il utilise dans ce but contient en particulier deux relations :

Editeur(Nom, Id_Editeur, Rue, Ville, Pays...) *Livre*(Id_Livre, Titre, Id_Editeur...)

Id_Livre et *Id_Editeur* désignent respectivement des identificateurs de livre et d'éditeur. La relation *Livre* contient 300.000 nuplets, à raison de 5 nuplets par page en moyenne. Elle est organisée en B+-Arbre sur *Id_Livre*, cet attribut étant codé sur 16 octets. La relation *Editeur* contient 100.000 nuplets, à raison de 10 nuplets par page en moyenne.

- 1 Les pages de la base de données sont de 2K octets, dont 2000 utilisables pour les données de la base. Les pages d'un B+-Arbre sont en moyenne aux trois quarts pleines.
 - a) Sachant qu'un identificateur de page est codé sur 4 octets, décrivez la structure des pages du B+-Arbre sur *Id_Livre*.
 - b) Quel est l'ordre de ce B-Arbre ?
 - c) Combien a t'il de niveaux ?
 - d) Combien a t'il en moyenne de pages par niveau ?
- 2 On veut établir la liste des Titres des livres édités dans un pays *P* qui regroupe un cinquième des Editeurs de la relation Editeur.
 - a) Quelle requête *R* en algèbre relationnelle va t'on poser à la base ?
 - b) Proposez un ordonnancement des opérations qui minimise le nombre d'entrées/sorties de pages de la base que le SGBD aura à effectuer pour établir le résultat de *R*.
 - c) On suppose que pour chaque opération relationnelle le SGBD dispose d'un choix d'algorithmes afin de pouvoir optimiser l'évaluation des requêtes. On sait de plus que *Id_Editeur* est codé sur 10 octets et *Titre* sur 100. Quelle serait selon vous la manière optimale d'évaluer la requête *R*, sachant que le SGBD dispose en mémoire de 100 emplacements où peuvent être copiées des pages de la base de données ? Décrivez soigneusement chaque étape du traitement, en particulier les relations intermédiaires construites par le SGBD et les algorithmes utilisés pour évaluer les différentes opérations relationnelles.
 - d) Combien de temps faudra t'il au SGBD pour établir le résultat, sachant qu'on peut effectuer en moyenne 50 entrées/sorties de page par seconde ?

Exercice 5 :

Soit la relation *R*(*Id_Personne*, *Nom*, *Prénom*, *Rue*, *#rue*, *Ville*, *CodePostal*, *Téléphone*)

Id_Personne est l'identificateur de la personne et la relation à un index primaire, structuré en B+-Arbre sur *Id_Personne*, qui est codé sur 4 octet. Il y a de plus un index secondaire en B+-Arbre sur *Nom* (26 caractères) et un sur *CodePostal* (6 caractères) .

On suppose :

- ?? que la relation a 50 000 nuplets.
- ?? que la taille d'un nuplet est de 150 octets.
- ?? qu'en moyenne les pages d'un B+-arbre sont aux trois quarts pleines.
- ?? que les pages ont 512 octets.

1. Etablir le nombre de pages de la relation ainsi que le nombre de niveaux de chaque index.
2. On veut faire une sélection $?_{Nom = Dupont} ?_{CodePostal = 75***}$ (*R*)
Proposer différents algorithmes pour faire cette sélection.
3. On suppose que le taux de sélectivité de chaque sélection élémentaire est de 1%, 10%, 50%. On suppose aussi qu'on a indépendance, au sens probabiliste du terme, entre les deux attributs de la sélection. Comparer le nombre d'E/S de page dans les cas suivants :
 - ?? sélection la plus sélective.
 - ?? sélection la moins sélective.
 - ?? cas intermédiaire.En cas de besoin on considérera que la zone dévolue aux pages en mémoire centrale contient 50 pages.
4. On reprend les questions 2 et 3 dans le cas où la condition de sélection comporte un opérateur d'union et non d'intersection.