

Bases de données avancées

Exercices dirigés : organisation physique

1. Types d'organisation physique

La table **Commande**(*numéro*, client, produit, date, destination) contient 120 000 lignes. Les tailles des champs sont (en octets) : *numéro*(10), *client*(40), *produit*(40), *date*(10) et *destination*(100).

Une page disque a 4 Ko, dont l'entête occupe 96 octets. Une adresse de page occupe 10 octets. Une lecture de page prend en moyenne 10 ms en lecture aléatoire et 1 ms en lecture séquentielle.

A. Pour chacun des types d'organisation physique suivants, quel sera le temps d'accès moyen à une commande étant donné son numéro ?

Même question si l'on cherche les commandes pour un produit commandé donné. On suppose qu'il y a en tout 1000 produits commandés et qu'ils apparaissent d'une manière uniforme dans les commandes.

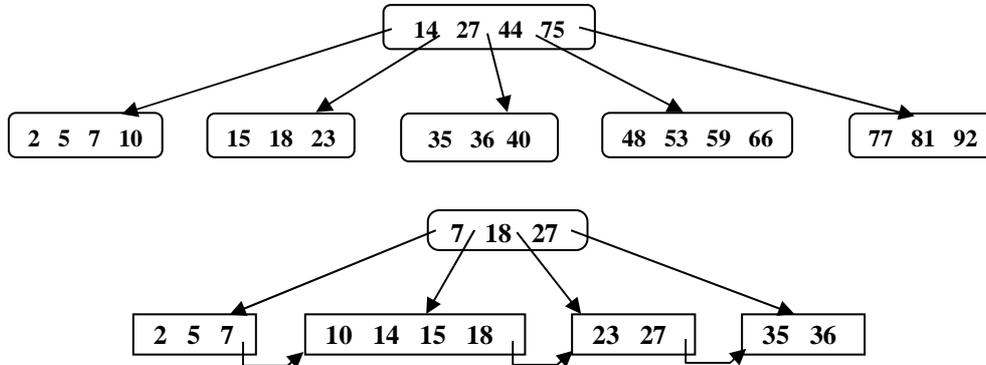
1. Organisation séquentielle.
2. Organisation séquentielle triée.
3. Organisation séquentielle indexée simple et généralisée.
4. Hachage, en considérant qu'il a été réalisé lorsque la table contenait 40 000 commandes et qu'on avait prévu 20% de plus d'entrées.

B. On construit un arbre B d'ordre maximal sur le numéro de commande. Quelle est sa hauteur minimum? Même question si l'on construit un arbre B+.

C. On construit un index arbre B+ sur le produit commandé. Sans calculer sa hauteur comme au point précédent, on considère qu'il possède 3 niveaux (deux pour les nœuds internes et un pour les feuilles). Quel est le temps nécessaire pour retrouver à l'aide de cet index toutes les commandes pour un produit donné ? Quelle est la différence dans le cas d'un index clustérisé ?

2. Opérations sur les arbres B/B+

- A. Soit les arbres B et B+ d'ordre 2 ci-dessous, où seules les clés d'indexation ont été représentées. Illustrez l'évolution de ces arbres par rapport à la séquence suivante d'opérations : insertion 3, 17, 1, 50, suppression 27, 36, 7. Considérer ensuite le cas avec rotation pour les insertions.



- B. Soit un index de type arbre B+ à n niveaux. Quel est le coût minimum de l'insertion d'une nouvelle entrée dans l'index ? Même question si la feuille d'insertion éclate, sans que l'éclatement se propage plus haut. Pareil si l'éclatement de la feuille est évité par rotation. Quel est le coût maximum d'une insertion ?

On considère que les nœuds sur le chemin parcouru jusqu'à la feuille d'insertion sont gardés en mémoire jusqu'à la fin de l'opération, donc il n'est pas nécessaire de les relire du disque si on les accède plusieurs fois.