

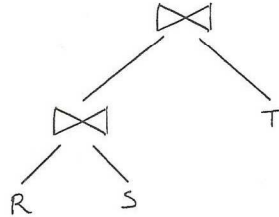
Examen de session 1 (durée 2h)

15 mai 2018

Documents autorisés : dix feuilles A4 recto-verso manuscrites ou imprimées

Optimisation

Question 1 $R(A, B)$, $S(B, C)$ et $T(C, D)$ sont trois tables dont on veut calculer la jointure naturelle. Supposer que l'optimiseur a choisi le plan logique suivant :



Supposer également les statistiques suivantes sur la base de données :

- nombre de lignes qui peuvent être stockées dans un bloc du disque :

R	S	T	$R \bowtie S$	$R \bowtie S \bowtie T$
10	20	10	8	5

- tailles des tables

R	S	T
r	s	t

- nombre de valeurs distincts de l'attribut B dans S : 50
- nombre de valeurs distincts de l'attribut B dans R : 20
- nombre de valeurs distincts de l'attribut C dans T : 100

Donner une estimation de chacune des valeurs suivantes (chaque estimation sera une expression utilisant les paramètres r , s et t) :

- le nombre de lignes de $R \bowtie S$;
- le nombre de blocs occupés par $R \bowtie S$;
- le nombre de lignes de $(R \bowtie S) \bowtie T$;
- le nombre de blocs occupés par $(R \bowtie S) \bowtie T$

Supposer maintenant que l'optimiseur choisit le plan physique suivant :

- jointure à une passe pour $(R \bowtie S)$, avec R chargée complètement en mémoire centrale ;
- matérialisation du résultat de $R \bowtie S$;
- block nested loop* pour la jointure entre $(R \bowtie S)$ et T , avec boucle externe sur $(R \bowtie S)$.

En utilisant les nombres estimés ci-dessus, calculer le coût de ce plan physique, c'est-à-dire le nombre d'accès aux blocs du disque (sans distinguer entre accès en lecture et en écriture). Remarquer que cela comporte non seulement le coût des deux jointures, mais également le coût de matérialiser le résultat intermédiaire et écrire le résultat final.

Donner le coût du plan physique comme une expression de r , s et t ; ensuite l'évaluer pour $r = 100$, $s = 2000$ et $t = 1000$, et donner le résultat.

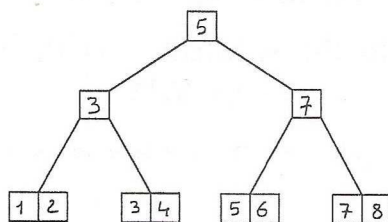
Question 2 Soit $R1(\underline{C}, D, E)$ et $R2(\underline{E}, F)$ $R3(\underline{A}, B, C)$ trois tables avec clefs primaires C , E , et A , respectivement. Supposer que $R1$ a 7500 lignes, $R2$ a 2500 lignes, et $R3$ a 800 lignes. Supposer que chaque bloc de chaque table peut contenir jusqu'à 100 lignes. Estimer la taille de $R1 \bowtie R2 \bowtie R3$ (en donner une borne supérieure).

Supposer maintenant que les index construits sur la clef primaire de chaque table soient des arbres B^+ de hauteur au plus 3.

- Donner un plan physique efficace pour calculer la jointure. Indiquer aussi si les résultats intermédiaires sont matérialisés ou pas dans ce plan.
- Estimer le nombre d'accès à disque nécessaires pour calculer la jointure selon ce plan (sans prendre en compte le coût de l'écriture du résultat).

Indexation

Question 3 Considérer l'arbre B^+ suivant



Supposer que chaque bloc du disque peut contenir au plus 3 clefs et 4 pointeurs.

Supposer que les opérations suivantes sont effectuées dans l'ordre :

- suppression de la clef 5 ;
- suppression de la clef 4 ;
- insertion de la clef 9 ;
- insertion de la clef 11.

Montrer l'arbre après chacune de ces opérations.

Formes normales

Question 4 Soit $R(A, B, C, D, E, F, G)$ un schéma de relation avec dépendances fonctionnelles :

$\mathcal{F} = \{$
 $DA \rightarrow C$
 $F \rightarrow DE$
 $B \rightarrow E$
 $FA \rightarrow G$
 $E \rightarrow F$
 $GA \rightarrow C$
 $DB \rightarrow F$
 $\}$

- Donner toutes les clefs de R .
- Trouver une décomposition de R en *forme normale de Boyce-Codd* sans perte d'information. Est-ce que cette décomposition préserve toutes les dépendances fonctionnelles ?
- Trouver une décomposition de R en *troisième forme normale* sans perte d'information et sans perte de dépendances fonctionnelles. (Se rappeler que il est nécessaire d'abord de minimiser $\mathcal{F}$, c'est-à-dire d'abord minimiser toutes les parties gauches, et ensuite éliminer toutes les dépendances fonctionnelles redondantes.)

Triggers

Question 5 On considère le schéma de base de données suivant pour la gestion d'un réseaux de machines :

Ordinateur (IP, nom)
Utilisateur (nom-util, password)
Session (nom-util, IP, timestamp-debut, timestamp-fin)

Les clefs primaires sont soulignées. Les clefs étrangères sont les suivantes :

$\text{Session}(\text{nom-util}) \subseteq \text{Utilisateur}(\text{nom-util}) ;$
 $\text{Session}(\text{IP}) \subseteq \text{Ordinateur}(\text{IP}) ;$

Écrire les triggers nécessaires pour garantir le comportement suivant de la base de données :

- Chaque nouvelle session insérée est automatiquement complétée avec timestamp-debut égal au timestamp courant, et timestamp-fin égal null, avant d'être insérée.
- À l'insertion d'une nouvelle session, toute session en cours du même utilisateur sur la même machine sera automatiquement terminée, c'est-à-dire son timestamp de fin prendra la valeur du timestamp de début de la nouvelle session.
- Toute suppression d'une session sera empêchée. Après une tentative de suppression, la session sera terminée (au timestamp courant).