

M1 Informatique
Bases de données avancées
Examen de session 1 (durée 2h)
19 mai 2017

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

Modélisation E/R et formes normales

Question 1 1. Produire un diagramme E/R qui modélise les informations suivantes concernant des séances de présentation de livres. Chaque livre a un titre, une (et une seule) maison d'édition, une année de publication et possiblement plusieurs auteurs. Les auteurs ont un nom, une date de naissance et une période d'activité. (Différents sites peuvent inviter un auteur à présenter un des ses livres.) Un site est caractérisé par un nom, une adresse et une ou plusieurs catégories (association, librairie, école, etc.). Une présentation d'un livre est effectuée par un seul auteur (qui doit être un auteur du livre même). Chaque présentation concerne un seul livre, est prévue à une date, dans un site donné, et a une durée. Un même livre peut être présenté plusieurs fois (à plusieurs dates), possiblement par des auteurs différents, et avec durées différentes. Mais un livre ne sera jamais présenté plusieurs fois dans le même site.

Traduire le diagramme E/R en un schéma relationnel. Exprimer les dépendances fonctionnelles sur le schéma obtenu. Vérifier que il est bien en forme normale de Boyce-Codd, par rapport à ces dépendances.

Question 2 Soit $R(A, B, C, D, E, F, G)$ un schéma de relation avec dépendances fonctionnelles :

$$\mathcal{F} = \{ \\ E \rightarrow B \\ D \rightarrow CG \\ CFB \rightarrow E \\ CF \rightarrow D \\ F \rightarrow G \\ DB \rightarrow A \\ AB \rightarrow E \}$$

- 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.)

Indexation

Question 3 On considère un *index hash extensible* construit sur l'attribut A d'une table R . Supposer que l'index utilise des valeurs de hachage à 4 bits. Supposer qu'une valeur de A occupe 20 bytes, qu'un pointeur occupe 8 bytes et que qu'un bloc de la table de hachage a une capacité de 64 bytes (hypothèse irréaliste pour faciliter les calculs). Simuler l'insertion de six enregistrements avec valeurs de hachage de A : 1111, 1110, 1101, 0001, 0000, dans cet ordre.

Supposer l'index hash initialement vide. Montrer l'état de l'index hash après chaque insertion. En particulier montrer, après chaque insertion :

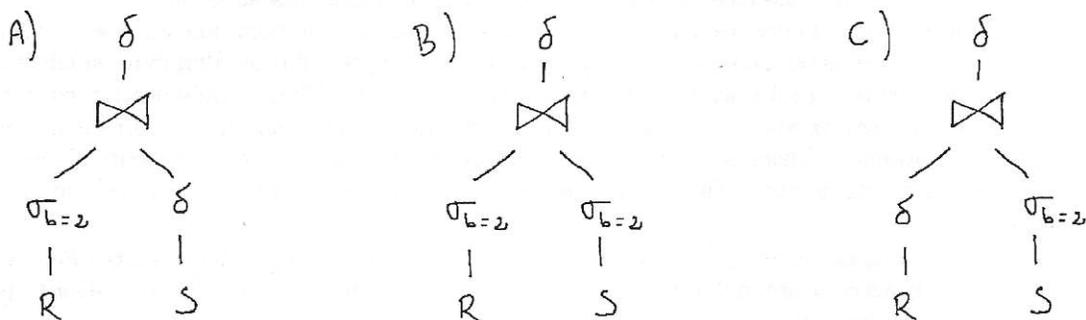
- le contenu des blocs de l'index (indiquer les valeurs de hachage au lieu des valeurs de A),
- la structure des pointeurs aux blocs,
- les valeurs des compteurs de chaque bloc et du compteur global i .

Évaluation de requêtes

Question 4 Les tables $R(a, b)$ et $S(b, c)$ ont r et s lignes, respectivement. Dans les deux tables l'attribut b prends valeurs $1, 2, \dots, 10$, avec la même probabilité (c'est-à-dire, dans chaque table, le nombre de lignes ayant valeur i de l'attribut b est le même, pour tout i). On suppose aussi que les deux tables ont des doublons et en particulier que chaque ligne apparaît 3 fois. On veut calculer la requête suivante :

SELECT DISTINCT * FROM R NATURAL JOIN S WHERE R.b = 2 (1)

On considère les trois plans d'exécution logiques dans la figure ci-dessous. Supposer que les opérateurs algébriques σ et \bowtie maintiennent les doublons, et que δ soit l'opérateur d'élimination des doublons.



1. Quels plans parmi A) B) et C) calculent la requête (1)?
2. Pour chacun des plans retenus au point 1, calculer la somme des tailles des résultats intermédiaires, comme une expression $t(r, s)$ dépendante de r et s . Remarquer que par "résultat intermédiaire" on entend le résultat de chaque noeud du plan, exception faite des feuilles et de la racine.
3. En utilisant $t(r, s)$ comme mesure de qualité du plan d'exécution, déterminer le meilleur plan parmi ceux retenus au point 1., dans chacun de ces deux cas :
 - i) $r = 6000, s = 9000$;
 - ii) $r = 900, s = 300$.

Question 5 On considère deux tables de schéma $R(b, d)$ et $S(b, c)$. On suppose que :

- Chaque bloc du disque a une capacité de 4KiB (4096 bytes), dans lesquels 96 bytes sont réservés pour l'en-tête du bloc.
- R contient 30 000 lignes et chaque ligne occupe 80 bytes;
- S contient 160 000 lignes et chaque ligne occupe 50 bytes;

Répondre aux questions suivantes, en montrant les détails des calculs :

1. Combien de blocs occupent R et S, respectivement?
2. Si l'on dispose d'un buffer de 3 blocs en mémoire centrale, calculer le coût de la jointure naturelle de R et S en utilisant *block nested loop join*, avec matérialisation du résultat. Est-ce qu'il est plus efficace de calculer $R \bowtie S$ ou $S \bowtie R$?
3. On suppose maintenant un buffer de 11 blocs en mémoire centrale, on souhaite utiliser un algorithme de *hash join*.

- Combien de partitions doivent être générés au minimum pour que chaque partition de R n'occupe pas plus de 10 blocs ? (supposer hachage uniforme).
- Quel est le coût du hachage de R et de S pour obtenir un tel nombre de partitions ?
- Calculer le coût du hash join qui en dérive (sans compter l'écriture du résultat).

Transactions

Question 6 On considère l'ordonnancement suivant de trois transactions :

R2(X), R1(X), W3(X), R2(Y), R1(Z), R3(Y), W3(Z)

En montrant le graphe de precedence, donner tous les ordonnancements sériels équivalents par conflit, s'il y en a.

Question 7 On considère trois transactions :

T1: R1(A), W1(B)

T2: R2(B), W2(D)

T3: R3(C), W3(A)

Lesquels des ordonnancements suivants sont possibles sous *verrouillage à deux phases (2PL)* avec verrou unique L(*lock*) / U (*unlock*) ?

(1) R1(A), R2(B), R3(C), W2(D), W3(A), W1(B)

(2) R3(C), R1(A), W3(A), R2(B), W2(D), W1(B)

Motiver les réponses. Quand c'est possible ajouter les opérations L/U à l'ordonnancement, de façon à respecter 2PL.