

## Examen de session 2 (durée 2h)

27 juin 2019

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

### Évaluation de requêtes

**Question 1** La table  $R$  a 200 000 lignes, dont 100 rentrent dans un bloc. La table  $S$  a 40 000 lignes, dont 40 rentrent dans un bloc. Il y a 11 blocs disponibles dans le buffer en mémoire centrale. Supposer qu'il y ait au plus 50 lignes de  $S$  qui peuvent être en jointure avec une même ligne de  $R$ . Calculer le nombre d'accès au disque nécessaires pour effectuer un *merge join* entre  $R$  et  $S$ . Inclure le coût du tri des deux tables ainsi que le coût de la matérialisation des tables triées; ne pas inclure le coût de l'écriture du résultat de la jointure.

**Question 2** Supposer qu'une jointure  $R \bowtie S$  est calculée par *block nested-loop join*. Supposer que  $R$  occupe 100 blocs et  $S$  occupe 30 blocs. Il y a 12 blocs en mémoire centrale qui peuvent être utilisés pour calculer cette jointure, dont 1 sera utilisé pour écrire le résultat un bloc à la fois. Le coût de cette jointure dépend de la table choisie pour la boucle externe et du nombre de blocs utilisés en mémoire centrale. Calculer le coût optimal (i.e. le plus petit nombre d'accès au disque) pour cette jointure, et dire quels choix permettent de l'obtenir.

### Optimisation

**Question 3**  $R(A, B)$ ,  $S(B, C)$  et  $T(C, D)$  sont trois tables dont on veut calculer la jointure naturelle. Supposer les statistiques suivantes sur la base de données :

- tailles des tables (nombre de lignes) :
  - $R$  : 3000
  - $S$  : 6000
  - $T$  : 10000
- chaque bloc du disque peut accueillir jusqu'à 10 enregistrements.
- nombre de valeurs distinctes de l'attribut  $B$  dans  $R$  : 10
- nombre de valeurs distinctes de l'attribut  $B$  dans  $S$  : 300
- nombre de valeurs distinctes de l'attribut  $C$  dans  $S$  : 30
- nombre de valeurs distinctes de l'attribut  $C$  dans  $T$  : 1000

1. Quel est le meilleur ordre pour la jointure  $R \bowtie S \bowtie T$ ? et le pire? Utiliser comme critère de coût la taille des résultats intermédiaires (on peut ignorer la taille du résultat final).
2. Avec référence au plan logique choisi au point 1. sur quels attributs faut-il définir un index pour que chaque jointure puisse être implémentée par *index join*? S'autoriser au plus deux index. S'il y a plusieurs façons de choisir ces deux index, proposer celle qui minimise le coût du plan physique. Pour calculer le coût du plan physique on supposera :
  - pipeline pour tous les opérateurs
  - index primaires
  - coût 1 pour la recherche d'une clef dans un index.

### Formes normales

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

$\mathcal{F} = \{$   
 $G \rightarrow B$   
 $F \rightarrow DB$   
 $DG \rightarrow F$   
 $FE \rightarrow A$   
 $B \rightarrow F$   
 $AE \rightarrow C$   
 $DE \rightarrow C$   
 $\}$



- 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 le dépendances fonctionnelles redondantes.)

## Transactions

**Question 5** Les deux transactions suivantes modifient deux éléments de type entier (A et B) dans la base de données

T1:

READ(A,t);  $t \leftarrow t+2$   
 $t := t+2$ ;  
 WRITE(A,t);  $t \leftarrow t+3$   
 READ(B,t);  $t \leftarrow t+3$   
 $t := t+3$ ;  
 WRITE(B,t);

T2:

READ(B,s);  
 $s := s*2$ ;  
 WRITE(B,s);  
 READ(A,s);  
 $s := s+3$ ;  
 WRITE(A,s);

- Y a-t-il deux ordonnancements sériels des deux transactions qui n'ont pas le même effet sur la base de données?
  - Si oui, donner un état initial de la base de données pour lequel les deux ordonnancements produisent un état final différent.
  - Si non, dire quel est l'effet sur A et B de tout ordonnancement sériel des deux transactions (en fonction des valeur initiaux de A et B).
- Donner un ordonnancement sérialisable non-sériel de T1 et T2.
- Donner un ordonnancement non-sérialisable de T1 et T2.
- Combien y a-t-il d'ordonnements serialisables de T1 et T2 (y compris les ordonnancements sériels)?