

Université Pierre et Marie Curie - Master d'informatique

Administration des Bases de Données Réparties (ABDR)

EXAMEN - 11 janvier 2005

Documents autorisés – Durée : 2h.

Répondre aux questions sur la feuille du sujet dans les cadres appropriés. La taille des cadres suggère celle de la réponse attendue. Utiliser le dos de la feuille précédente si la réponse déborde du cadre. Le barème est donné à titre indicatif. Ne pas dégrafer le sujet. Ecrire à l'encre bleue ou noire.

La **qualité de la rédaction** sera fortement prise en compte. *Illustrer, argumenter et justifier* chaque réponse. Préciser toutes les hypothèses supplémentaires sur lesquelles repose votre raisonnement.

Exercice 1 : Contrôle de concurrence

5 pts

Une entreprise utilise le SGBD Sybase pour son application de gestion de stock. Le contrôle de concurrence est garanti grâce au protocole de verrouillage à 2 phases.

Le schéma des données est :

Stock (np, quantité, prix, qmv)

np est le n° d'un produit

quantité est la quantité restante en stock.

qmv est la quantité moyenne vendue pendant les 30 derniers jours

Vente (nv, montantTotal, date)

nv est le n° d'une vente (clé)

montantTotal est la somme des montants des articles d'une vente

VenteArticle (nv, np, quantité, montant)

L'application de gestion est composée des transactions suivantes :

T1(n) : vente de n articles : chaque article p est vendu en quantité q, avec mise à jour des relations Stock, Vente et VenteArticle.

T2(m) : m réapprovisionnements. Chaque réapprovisionnement correspond à l'ajout en stock de 10 unités d'un produit p'.

Les valeurs n, m, q, p et p' sont aléatoires.

5 transactions de type T1 et une transaction de type T2 s'exécutent simultanément. Le niveau d'isolation des transactions est sérialisable.

Question 1

L'exécution d'une transaction T1 est parfois refusée car une mise à jour ne satisfait pas la contrainte d'unicité de *nv*. Expliquer pourquoi. Proposer une solution pour éviter ce problème ?

Question 2

a) Dans quels cas des interblocages peuvent-ils se produire ?

b) La valeur de m a-t-elle une influence sur la fréquence des interblocages ? Le nombre de transactions T1 simultanées a-t-il une influence sur la fréquence des interblocages ?

c) Proposer une solution pour éviter tout interblocage ? Votre solution a-t-elle un impact sur la cohérence des données et sur les performances de l'application ?

Question 3

L'application de gestion est complétée avec la transaction T3 qui actualise la quantité moyenne vendue de chaque produit (attribut *qmv*).

La durée d'exécution moyenne de T3 est 100 fois celle de T1. Chaque fois que T3 s'exécute, de nombreuses transactions T1 et T2 sont bloquées. Expliquer pourquoi. Proposer une solution pour que T3 puisse s'exécuter sans bloquer T1 ni T2.

Exercice 2 : Stockage**5 pts**

Une application contient les données

Employé (ne, ns, nom, prénom, ville, age, commentaire)

ne est le n° d'employé, *ns* est son n° de service

Service (ns, nom, description, budget)

On donne les informations suivantes : la taille d'une page est de 1000 octets

	Emp	Service
Cardinalité	10^6	10 000
taille d'un tuple (octets)	100	50
domaine de la clé	$[1, 10^6]$	$[1, 10^4]$

On suppose uniforme la distribution des valeurs d'attributs.

On propose 2 possibilités de stockage au choix. Pour chaque stockage les tuples sont triés sur disque selon l'ordre croissant d'un certain attribut :

P1 : tous les employés triés par **ns**, puis tous les services triés par ns

P2 : pour chaque ns croissant : le tuple de Service est juxtaposé avec les Employés de ce service.

Question 1

Soient les requêtes :

R1 : Select * from Service where $ns \leq v$

R2 : Select * from Emp e, Service s where $e.ns = s.ns$ and $ns = 1000$

a) Quels index ajouter pour réduire la taille des données lues par les requêtes.

b) On exécute successivement R1 puis R2. En fonction des valeurs de ν , indiquer quel est le stockage P_i le plus avantageux ? C'est celui pour lequel le nombre total de pages lues, pour traiter les requêtes R1 et R2, est minimal. On suppose que les index sont en mémoire.

Question 2

On effectue les mises à jour suivantes

U1 : insérer 1000 nouveaux employés travaillant dans 10 services existants.

U2 : insérer 1000 nouveaux employés travaillant dans 10 nouveaux services.

U3 : modifier l'attribut *commentaire* de 1000 employés de telle sorte que cela double la taille des tuples modifiés.

U4 : modifier l'attribut *description* de 1000 services de telle sorte que cela double la taille des tuples modifiés

Expliquer l'effet de ces mises à jour sur P1 et P2. Expliquer comment maintenir les propriétés initiales des stockages P1 et P2.

Exercice 3 : Dictionnaire du SGBD**5 pts**

Le dictionnaire du SGBD contient les vues suivantes pour gérer les rôles et les privilèges.

UP (u, p, t) : le privilège p sur la table t est accordé directement à l'utilisateur u .

UR (u, r) : le rôle r est accordé directement à l'utilisateur u .

RP (r, p, t) : le privilège p sur la table t est accordé directement au rôle r .

RR ($r1, r2$) : le rôle $r2$ est accordé directement au rôle $r1$.

La relation «accordé à» est transitive : si un privilège (p, t) est accordé à un rôle r , et si r est accordé à l'utilisateur u , alors le privilège (p, t) est considéré comme étant accordé **indirectement** à l'utilisateur u .

Expliquer comment obtenir les informations suivantes. Préciser l'algorithme et les requêtes SQL nécessaires.

a) Quels sont les utilisateurs possédant, directement **ou indirectement**, un privilège sur la table x ?

b) Quels sont les utilisateurs qui ne possèdent aucun des privilèges du rôle y ?

c) Pour simplifier l'obtention des informations demandées aux questions a) et b), on souhaite modifier l'algorithme de la commande **grant** qui accorde un privilège (p, t) ou un rôle r à un utilisateur ou à un autre rôle. Proposer une solution pour maintenir avec UP la liste de tous les privilèges directs et indirects des utilisateurs.

Exercice 4 : Réservation de billet d'avion**5 pts**

Soit le schéma :

Vol (nv, siège, état) *nv* est le n° d'un vol, l'état d'un siège est 'libre' ou 'occupé'.

Vol a 10⁶ tuples (approximativement 5000 vols avec 200 sièges par vol).

Soit l'application suivante :

La transaction T1, exécutée une fois par heure :

```
Select avg(dispo) from ( Select nv, count(siège) as dispo
                        From Vol
                        Where état = 'libre'
                        Group by nv);

Commit;
```

La transaction T2 exécutée 100 fois par secondes :

```
select siège from Vol where état='libre' and nv='AF618';
-- du code permettant au voyageur de choisir son siège
-- et de stocker le siège choisi dans la variable &choix
update Vol set état = 'occupé' where nv = 'AF618' and siège = &choix;
commit;
```

Question 1

Décrire ce que fait T1 en une phrase

Question 2

Quel est le problème avec T2 ? Est-ce que ce problème est le même pour les 2 niveaux d'isolation supportés par Oracle (read committed et serializable) ?

- Analyser le problème avec précision et concision. Décrire les conséquences.
- Expliquer comment résoudre le problème. Décrire quels nouveaux problèmes peuvent être engendrés par votre solution au problème initial. Expliquer comment les résoudre.

Question 3

Quels sont les index les plus appropriés pour cette application ?

Question 4

On remplace maintenant T2 par T2' :

```
Select min(siège) into &choix
From Vol
Where état = 'libre' and nv = 'AF618'

If(&choix != null) {
    Update Vol set état = 'occupé' where nv = 'AF618' and siège = &choix ;
}
else affiche('plus aucun siège disponible')
commit;
```

Remarque : T2' ne contient plus aucune interaction avec le voyageur.

Décrire les avantages et inconvénients des 2 niveaux d'isolation lorsque la charge de l'application est constituée uniquement de la transaction T2' (100 transactions soumises par secondes). Quel niveau d'isolation conseillez vous?

Question 5

Quel que soit le niveau d'isolation choisi à la question précédente, on constate que les performances ne sont pas satisfaisantes. Le système ne parvient pas à traiter 100 transactions T2' par secondes. Donner 2 hypothèses à ce problème de performance. Expliquer comment vérifier ces hypothèses expérimentalement (quels indicateurs de performance consulter ? Quelles valeurs sont attendues pour ces indicateurs ?). Proposer des solutions pour remédier au problème.