

Université d'Ottawa
CSI 3717 – Examen Final
Professeur(s): Iluju Kiringa

19 décembre, 2003
14:h00-17h00
Durée: 3 hrs

A livre fermé; aucune aide permise, sauf une page A4 écrite au recto et verso

Nom de famille: _____

Prénom: _____

Numéro d'étudiant: _____

Il y a 11 questions et un total de 100 points.

Cet examen doit contenir 18 pages,
incluant cette page couverture.

1 – Choix multiple	/ 32
2 – Définitions des concepts	/ 8
3 – Diagramme ER	/ 10
4 – Algèbre relationnelle	/ 6
5 – SQL	/ 4
6 – Index Matching	/ 5
7 – Hash-Join	/ 5
8 – Sérialisabilité	/ 10
9 – Formes normales	/ 10
10 – Propriétés ACID	/ 5
11 – Le protocole “Two Phase Locking”	/ 5
<hr/>	
Total	/ 100

LIRE IES INSTRUCTIONS SUR LA PAGE SUIVANTE EN PREMIER !

Répondez aux questions suivantes: 1 – 9 et 10 ou 11.

L'examen consiste en deux parties: 40% choix multiple; 60% essai.

La répartition des points est donnée pour chaque question.

Tout votre travail doit être contenu dans ce cahier d'examen. Utilisez le verso des pages si nécessaire.

Note: Pour les questions à choix multiple, choisissez la réponse la plus inclusive. Par exemple, si l'on vous demandait "Qui est le premier ministre du Canada?" et que les réponses possibles sont:

- A. Paul Martin
- B. Le leader du parti libéral
- C. Le leader du parti majoritaire à la Chambre des Communes
- D. Le leader du parti requis par la Gouverneure-Générale pour conduire le présent parlement

vous choisirez (D) parce que cette réponse est la plus inclusive et la plus générale possible.

"Le règlement de l'université affirme que tricher est une offense académique. Commettre une offense académique est suivi de sanctions qui peuvent aller jusqu'à l'exclusion du programme. Si vous êtes surpris entraîné de tricher, il n'y aura pas un second avertissement".

1 Choix multiple — 32 points

Partie A — 1 point Un des items suivants n'est **PAS** une caractéristique des système de fichiers:

- A. acces lent
- B. duplication des données
- C. isolation des données
- D. requêtes statiques

Partie B — 1 point Dans le modèle ANSI-SPARC de référence (i.e. l'architecture standard de référence), si le lieu de stockage physique des données change, alors

- A. la base des données doit être recompilee
- B. les niveaux supérieures ont besoin d'être au courant de ce changement
- C. rien ne change aux niveaux au-dessus du niveau physique
- D. seul le niveau immédiatement au-dessus du niveau physique a besoin d'être au courant du changement

Partie C — 1 point Une des fonctions suivantes n'est **PAS** une fonction d'aggrégation:

- A. COUNT
- B. MEDIAN
- C. SUM
- D. AVG

Partie D — 1 point Conceptuellement, quel est l'ordre temporel d'évaluation d'une requête GROUP BY?

- A. WHERE, HAVING, GROUP BY
- B. WHERE, GROUP BY, HAVING
- C. HAVING, WHERE, GROUP BY
- D. L'ordre est irrelevant

Partie E — 1 point Lequel des items suivants n'est **PAS** un avantage de SQL?

- A. portabilité
- B. standard universel
- C. plusieurs manières différentes d'exprimer la même requête
- D. offre l'accès à la fois interactif et imbriqué

Partie F — 1 point Que fait une clause HAVING?

- A. filtre les groupes, sujet à une condition
- B. filtre des groupes de lignes
- C. filtre des groupes
- D. filtre des groupes de lignes ayant la même valeur de colonne

Partie G — 3 points Supposez que nous avons une relation déclarée par l'instruction SQL suivante:

```
CREATE TABLE FOO(
    name VARCHAR(50) PRIMARY KEY,
    salary INT
    CHECK(salary <= (SELECT AVG(salary) FROM FOO)) );
```

Initialement, le contenu de FOO est:

name	salary
'Sally'	10000
'Joe'	20000
'Sue'	30000

Nous essayons d'exécuter la suite suivante de modifications, l'une après l'autre:

```
INSERT INTO FOO VALUES('Fred', 12000);
UPDATE FOO SET salary = 20000 WHERE name = 'Sue';
INSERT INTO FOO VALUES('Sally', 13000);
DELETE FROM FOO WHERE name = 'Joe';
```

À la fin de l'exécution de ces instructions, la somme des salaires de tous les tuples dans FOO est:

- A. 42,000
- B. 62,000
- C. 65,000
- D. 72,000

Partie H — 3 points Supposez que nous voulons contraindre les données de manière à n'avoir aucun département avec un total des salaires des employés qui dépassent \$1,000,000. L'assertion suivante est censée réaliser cette contrainte:

```
CREATE ASSERTION cheap CHECK( NOT EXISTS(Q));
```

Quelle requête Q réalise le mieux cette contrainte?

- A. `SELECT * FROM Emps WHERE SUM(salary) > 1000000`
- B. `SELECT dept, SUM(salary) FROM Emps GROUP BY dept`
- C.

```
SELECT SUM(salary)
FROM Emps, Managers
WHERE id = mgr
GROUP BY Emps.dept
HAVING SUM(salary) > 1000000
```

D.

```
SELECT dept
FROM Emps
GROUP BY dept
HAVING SUM(salary) > 1000000
```

Partie I — 3 points Supposez la relation `Enroll(SID, CID, term, grade)` qui stocke les données académiques d'étudiants dans une base de données universitaire. Par exemple, un tuple `(123, 'SE470b', 'Winter 2003', 9.0)` représente le fait que l'étudiant 123 a pris le cours SE470b en hiver 2003 et y a obtenu une note de 9.0.

Quelqu'un dans le bureau du registraire s'est plaint d'avoir obtenu des résultats différents en répétant la même requête dans une transaction:

```
T1:
SELECT AVG(grade) FROM Enroll;      Q1
SELECT AVG(grade) FROM Enroll;      Q2
COMMIT;
```

Laquelle des mesures suivantes est suffisante et nécessaire pour assurer que Q1 et Q2 retournent le même résultat?

- A. Mettre le niveau de transaction pour toutes transactions, y compris T1, a SERIALIZABLE
- B. Mettre le niveau de transaction pour toutes transactions, y compris T1, a REPEATABLE READ
- C. Mettre le niveau de la transaction T1 a SERIALIZABLE
- D. Mettre le niveau de la transaction T1 a REPEATABLE READ

Partie J — 1 point Lequel des aspects suivants n'est **PAS** un aspect des transactions?

- A. RAID
- B. graphes wait-for
- C. two-phase commit
- D. verrous

Partie K — 1 point Lequel des problèmes suivants peut apparaître dans un environnement multi-utilisateurs?

- A. changement perdu (“Lost update”)
- B. données non validées (“Uncommitted data”)
- C. données inconsistantes
- D. Tous les problèmes ci-dessus

Partie L — 1 point 2PL ne fait **PAS** la chose suivante:

- A. permet des deadlocks
- B. permet seulement des plans sérialisables
- C. relâche les verrous seulement au moment de la validation
- D. relâche les verrous progressivement

Partie M — 1 point Comment BCNF diffère-t-elle de la troisième forme normale (3NF)?

- A. Elles sont les mêmes
- B. BCNF a la propriété que pour toute dépendance fonctionnelle (DF) non-triviale $X \rightarrow Y$, X contient une clé

- C. 3NF ne contient aucune dépendance transitive d'une DF de la clé primaire
- D. BCNF contient des dépendance partielles de la clé primaire

Partie N — 1 point La différence entre les protocoles d'acc es simultan é pessimiste et optimiste est:

- A. les protocoles pessimistes ne permettent pas l'apparition des conflits
- B. les protocoles pessimistes ne permettent pas l'apparition des deadlocks
- C. les protocoles optimistes supposent que les conflits sont rares
- D. toutes les réponses ci-haut

Partie O — 2 points Laquelle des assertions suivantes n'est **PAS** un avantage de l'utilisation d'un SGBD?

- A. Les programmes d'application ne sont pas exposés aux détails de la représentation des données et du stockage.
- B. Le SGBD planifie l'acc es simultan é aux données.
- C. L'administration des données n'est pas centralisée.
- D. Le SGBD protège les données des effets des faillites du système.
- E. Le SGBD assure le respect des contraintes d'int égrité sur les données stockées.

Partie P — 2 points Laquelle des assertions suivantes n'est **PAS vraie**?

- A. Un modèle de données est une collection d'éléments langagiers pour décrire les données.
- B. L'élément central pour décrire les données en modèle relationnel est la notion de schéma.
- C. Le schéma physique spécifie les détails du stockage du modèle des données.
- D. Le schéma conceptuel définit les données en utilisant le modèle des données.
- E. Un schéma externe contient une collection d'une ou plusieurs vues sur les relations dans la base des données.

Partie Q — 2 points Laquelle des assertions suivantes est **vraie**?

- A. Le but du design conceptuel des bases de données est de produire un schéma logique de la base de données.

- B. Le but de la décomposition du schéma logique est de produire un schéma physique de la base de données.
- C. Le fichier des enregistrements est une abstraction importante en SGBD et il est implémenté par les méthodes d'accès du SGBD.
- D. Un "heap file" est un fichier d'enregistrements qui sont stockés sur un "heap".
- E. Les requêtes à plages ("range queries") sont bien supportées par les indexes à hachage.

Partie R — 2 points Laquelle des assertions suivantes n'est **PAS vraie**?

- A. Dans une page de mémoire pour des enregistrements à longueur fixe organisée en mode "packed", l'espace libre est toujours située à la fin de la page.
- B. Dans une page de mémoire pour des enregistrements à longueur fixe organisée en mode "unpacked", l'espace libre est dispersé sur toute la page.
- C. Dans un RAID, la redondance et la répartition des données sont combinées pour respectivement accroître la fiabilité et la performance.
- D. Les niveaux de RAID qui utilisent le miroitement maintiennent des copies identiques des données sur deux disques différents.
- E. Un "bitmap" est une chaîne de bits pour maintenir l'information sur les lots libres dans une page de mémoire pour des enregistrements à longueur fixe organisée en mode "packed".

Partie S — 2 points Laquelle des assertions suivantes est **vraie**?

- A. Dans un arbre ISAM, seules des pages internes sont affectées lors des modifications.
- B. Dans arbre ISAM, on efface toute feuille vide.
- C. Dans un arbre ISAM, on ne maintient pas des pages de débordement.
- D. Pendant l'effacement dans un arbre ISAM, on efface toutes les pages de débordement.
- E. Tant les arbres ISAM que les arbres B+ supportent des requêtes à plages ("range queries").

Partie T — 2 points Laquelle des assertions suivantes est **vraie**?

- A. Dans une table de hachage statique, les pages primaires peuvent être effacées.
- B. Dans une table de hachage extensible, on peut avoir des pages de débordement.
- C. Le hachage statique est similaire aux arbres B+.
- D. Durant l'effacement dans une table de hachage statique, on efface toutes les pages de débordement.
- E. Aucune des assertions ci haut n'est vraie.

2 Définitions des concepts — 8 points

Donnez une définition claire de **deux lignes** pour chacun des concepts suivants:

1. plan sériel

2. transaction

3. index groupé

4. dépendance fonctionnelle

5. RAID (Donnez l'idée.)

6. "deadlock"

7. atomicité

8. serialisabilité par rapport aux conflits

3 Diagramme ER — 10 points

SITE possède une base de données qui contient de l'information sur des professeurs. Cette information est stockée en utilisant les spécifications suivantes:

- Chaque professeur a un NAS; il a aussi un nom, ainsi que l'année dans laquelle il a été engagé.
- Il y a des professeurs permanents et des nonpermanents. Les professeurs permanents ont une indication de l'année de leur obtention de permanence. Les nonpermanents ont une indication de l'année à laquelle ils peuvent faire une demande de permanence.
- Parmi les professeurs permanents, il y en a qui sont des professeurs associés et d'autres des professeurs complets. Ces derniers ont une indication de l'année de leur promotion au rang de professeur complet.
- Les professeurs nonpermanents doivent être experts en un domaine scientifique donné.
- Tous les domaines d'expertise ont une identification ainsi qu'une description.
- Un professeur complet peut superviser un professeur nonpermanent sur son domaine d'expertise. Aussi, chaque expertise d'un nonpermanent sur un domaine doit avoir un professeur complet qui la supervise.

Dessinez un diagramme ER pour ce design. Indiquez toutes les contraintes appropriées.

4 Algèbre relationnelle — 6 points

Considérez le schéma relationnel suivant:

Suppliers(sid : int, sname : string, adress : string, rating : real)

Parts(pid : int, pname : string, color : string, fid : int)

Catalog(sid : int, pid : int, cost : real)

La relation Catalog liste les prix pour les pièces ("parts") fournies par les fournisseurs. Formulez les requêtes suivantes en **algèbre relationnelle**:

- Trouvez les sids des fournisseurs qui fournissent chaque pièce rouge ou qui fournissent chaque pièce verte.
- Trouvez les pids des pièces fournies par chaque fournisseur à un prix en dessous de \$200.
- Trouvez les noms des fournisseurs qui fournissent une pièce verte qui coûte moins de \$50.

5 SQL — 4 points

Considérez le schéma de la question précédente.

1. Formulez la requête suivante en SQL:

”Trouvez les noms des fournisseurs qui fournissent une pièce verte qui coûte moins de \$50.”

2. Formulez une instruction SQL pour créer la relation Catalog, en incluant la contrainte d'intégrité suivante (Utilisez un CHECK):

“Chaque pièce a un minimum de 2 fournisseurs et un maximum de 10 fournisseurs qui la fournissent.”

6 Index Matching — 5 points

Considérez le schéma relationnel suivant pour la relation Suppliers:

$Suppliers(\underline{sid} : \text{int}, sname : \text{string}, adress : \text{string}, rating : \text{real})$

Pour chacun des indexes suivants, dites si cet index correspond a la condition de sélection donnée. Donnez les conjoints primaires de chaque correspondance. I.e. repondez par **oui** ou **non** et si oui donnez les conjoints primaires.

(1) Un arbre B+ avec clé de recherche $\langle Suppliers.sid \rangle$

(a) $\sigma_{Suppliers.sid > 20,000}(Suppliers)$

(b) $\sigma_{Suppliers.sid < 20,000}(Suppliers)$

(c) $\sigma_{Suppliers.sid = 20,000}(Suppliers)$

(2) Un arbre B+ avec clé de recherche $\langle Suppliers.sid, Suppliers.rating \rangle$

(a) $\sigma_{Suppliers.sid > 20,000 \wedge Suppliers.rating < 4}(Suppliers)$

(b) $\sigma_{Suppliers.rating < 4}(Suppliers)$

(c) $\sigma_{Suppliers.sid = 20,000}(Suppliers)$

(2) Un index a hachage avec clé de recherche $\langle Suppliers.sid, Suppliers.rating \rangle$

(a) $\sigma_{Suppliers.rating < 4}(Suppliers)$

(b) $\sigma_{Suppliers.sid > 20,000 \wedge Suppliers.rating < 4}(Suppliers)$

(c) $\sigma_{Suppliers.sid = 20,000}(Suppliers)$

7 Hash-Join — 5 points

Expliquez comment l'algorithme de hash-join marche.

8 Sérialisabilité — 10 points

Partie A — 4 points Considérez les plans suivants:

$$S_1 = R_2(X), R_1(Y), W_2(X), R_3(X), W_1(Y), W_3(X), R_2(Y), W_2(Y), C_1, C_2, C_3$$

$$S_2 = R_2(X), R_1(Y), W_2(X), R_2(Y), R_3(X), W_1(Y), W_3(X), W_2(Y), C_1, C_2, C_3$$

Pour chacun de ces plans, dites s'il est sérialisable par rapport aux conflits; s'il ne l'est pas dites pourquoi.

Partie B — 6 points Pour chacun des plans suivants:

$$S_1 = R_1(X), R_2(X), R_3(Y), W_1(X), R_2(Z), R_2(Y), W_2(Y), W_1(Z), C_1, C_2, C_3$$

$$S_2 = R_1(X), W_1(Y), R_2(Y), W_2(Z), R_3(Z), W_3(X), C_1, C_2, C_3$$

$$S_3 = W_3(X), R_1(X), W_1(Y), R_2(B), W_2(Z), R_3(Z), C_1, C_2, C_3$$

repondez aux questions suivantes:

- (1) Quel est le graphe de précédance du plan?
- (2) Donnez un pla sériel équivalent pour chaque plan qui est sérialisable par rapport aux conflits.

9 Formes normales — 10 points

La table donnée ci bas montre les détails des rôles joués par des acteurs et actrices dans des films.

fNo	fTitle	dirNo	dName	actorNo	actorName	role	timeOnScreen
F1100	Happy Days	D101	Jim Alan	A1020	Sheila Toner	Jean	15.45
F1100	Happy Days	D101	Jim Alan	A1222	Peter Watt	Tom	25.38
F1100	Happy Days	D101	Jim Alan	A1020	Sheila Toner	Silvia	15.45
F1109	Snake Bite	D076	Sue Ramsay	A1567	Steven McDonald	Tim	19.56
F1109	Snake Bite	D076	Sue Ramsay	A1222	Peter Watt	Archie	10.44

- Decrivez pourquoi la table ci haut n'est pas en premiere forme normale (1NF).
- La table ci haut est susceptible d'anomalies de modification. Donnez des exemples de la maniere dont des anomalies liées a l'insertion, l'efacement et la modification peuvent survenir dans cette table.
- Identifiez les dépendances fonctionnelles présentes dans la table ci haut. Indiquez toute suppositions que vous faites en répondant a cette question si nécessaire.
- Utilisant les dépendances fonctionnelles identifiées dans la partie (c), decrivez et illustrez le processus de normalisation en convertissant la table ci haut en troisieme forme normale (3NF). Identifiez les clés primaires et secondaires de vos nouvelles relations.

10 Propriétés ACID — 5 points

Considérez les propriétés ACID. En une page, décrivez pourquoi elles sont importantes pour les transactions.

11 Le protocole “Two Phase Locking” — 5 points

Expliquez comment le protocole “two-phase locking” fonctionne.