

--	--	--

## Systèmes de Gestion de Bases de Données – 3I009

EXAMEN - 1<sup>ère</sup> session du 5 janvier 2016

Durée : 2 heures

Documents autorisés

*Les téléphones mobiles doivent être éteints et rangés dans les sacs. Le barème sur 21<sup>1/2</sup> points (20 questions) n'a qu'une valeur indicative.*

### 1 Optimisation de schéma (5 pts)

Soient une table **R(A,B,C,D,E,F)** et deux ensembles de dépendances fonctionnelles  $\mathcal{F}$  et  $\mathcal{G}$  :

$$\mathcal{F} = \{AB \rightarrow C; A \rightarrow B; C \rightarrow BD; D \rightarrow A; E \rightarrow BF\} \quad (1)$$

$$\mathcal{G} = \{A \rightarrow BC; C \rightarrow D; D \rightarrow A; E \rightarrow BF\} \quad (2)$$

$\mathcal{G}$  est un ensemble minimal.

**Question 1** (1 point)

Montrez que  $\mathcal{G}$  est équivalent à  $\mathcal{F}$ .

Réponse :

**Question 2** (1 point)

Donnez les clés de la table **R**. Justifiez votre réponse.

Réponse :

**Question 3** ( $\frac{1}{2}$  point)

Est-ce que la table **R** est en 3e forme normale (3FN) ? Justifiez formellement votre réponse.

Réponse :

**Question 4** ( $1\frac{1}{2}$  points)

Est-ce que la décomposition suivante de **R** en (**R1**, **R2**, **R3**) est sans perte d'information (SPI) par rapport à  $\mathcal{G}$  ? Justifiez votre réponse en utilisant le méthode du tableau (montrez quelles DF sont utilisées)

- **R1(A,B,C,E)**
- **R2(C,D)**
- **R3(D,E,F)**

Réponse :

**Question 5** (1 point)

Donnez une décomposition SPI et SPD de **R** par rapport à  $\mathcal{G}$  (ensemble minimal), en tables qui sont en 3FN. Donnez pour chaque table les dépendances fonctionnelles associées et les clés.

**Réponse :**

## 2 Indexation : arbres B+ et tables de hachage (3 $\frac{1}{2}$ pts)

Dans cet exercice, on considère des arbres B+ d'ordre 2 (les noeuds et les feuilles ont entre 2 et 4 valeurs). Soit un arbre A1, contenant les valeurs suivantes : 3, 5, 10, 12, 16, 18, 22, 25, 29, 40, 45, 46, 47, 51, 60.

### Question 6 (1 point)

Dessinez l'arbre A1, sachant que la racine contient les valeurs 16, 20, 35, 50.

**Réponse :**

**Question 7** (1½ points)

On insère successivement dans A1 les valeurs 32, puis 35. Dessinez l'arbre A2 après insertion de ces valeurs. En cas d'éclatement d'une feuille en deux, la feuille de gauche contiendra les 3 plus petites clés, la feuille de droite les deux plus grandes clés.

**Réponse :**

**Question 8** (1 point)

**Dans l'arbre A1**, on supprime successivement les valeurs 18, 5, puis 10. Dessinez l'arbre A3 après suppression de ces valeurs. On considère, dès que c'est possible, la redistribution des clés avec le frère de gauche.

**Réponse :**

### 3 Transactions et concurrence (4 pts)

Soient  $T_1, T_2, T_3, T_4$  et  $T_5$  cinq transactions,  $x, y, z$  et  $t$  quatre granules d'une base de données. On note :

- $L_i(g)$  la lecture de la transaction  $T_i$  du granule  $g$
- $E_i(g)$  l'écriture de la transaction  $T_i$  du granule  $g$
- $V_i$  l'opération de validation de la transaction  $T_i$

#### Question 9 (1 point)

Soit  $S_1$  la séquence d'opérations donnée comme suit

$$S_1 = L_3(t), L_5(t), L_4(z), E_4(t), E_4(z), L_2(z), E_2(z), L_1(z), L_1(y), E_1(x), E_5(x), L_3(x), V_1, V_2, V_3, V_4, V_5$$

On suppose que les opérations sont exécutées dans l'ordre indiqué.

Préciser pour chaque granule la séquence d'opérations qui le concerne ainsi que les arcs de précedence  $T_i \rightarrow T_j$ .

**Réponse :**

granule	LA séquence	LES arcs
$x$	$E_1 \dots$	$T_{\dots} \rightarrow T_{\dots}$
$y$		
$z$		
$t$		

Cette séquence est-elle sérialisable ? **Justifier.**

**Réponse :** Barrer la mauvaise réponse

$S_1$  est

SERIALISABLE

NON SERIALISABLE

**Justification**

**Question 10** (1 point)

Soit  $S_2$  la séquence d'opérations donnée comme suit

$$S_2 = L_3(t), L_5(y), L_4(z), E_4(t), E_4(z), L_2(z), E_2(z), L_1(x), L_1(y), E_1(x), E_5(y), L_3(x), V_1, V_2, V_3, V_4, V_5$$

On suppose que les opérations sont exécutées dans l'ordre indiqué.

Préciser pour chaque granule la séquence d'opérations qui le concerne ainsi que les arcs de précedence  $T_i \rightarrow T_j$ .

**Réponse :**

granule	LA séquence	LES arcs
$x$	$L_1 \dots$	$T_{\dots} \rightarrow T_{\dots}$
$y$		
$z$		
$t$		

Cette séquence est-elle sérialisable ? **Justifier.**

**Réponse :** Barrer la mauvaise réponse

$S_2$  est

SERIALISABLE

NON SERIALISABLE

**Justification**

**Pour les questions suivantes, on ne considère que trois transactions  $T_1, T_2$  et  $T_3$  travaillant sur trois granules  $x, y, z$**

**Question 11** (1 point)

Soit  $S_3$  la séquence d'opérations donnée comme suit

$$S_3 = L_2(x), L_3(x), L_1(y), L_1(z), E_2(z), V_2, E_3(x), E_1(z), V_1, E_3(y), V_3$$

On voudrait appliquer le protocole de verrouillage en deux phases strict (2PL strict). Indiquer la séquence d'actions obtenue en sortie lorsqu'il n'y a pas d'interblocage. S'il y a un interblocage, dessiner le graphe des attentes à la place.

**Réponse :** Barrer la mauvaise réponse

 INTERBLOCAGE PAS D'INTERBLOCAGE

**Ordre en sortie ou Graphe des attentes**

**Question 12** (1 point)

Soit  $S_4$  la séquence d'opérations donnée comme suit

$$S_4 = L_1(x), L_2(y), L_2(z), L_3(y), E_3(y), V_3, E_1(z), V_1, E_2(x), V_2$$

On voudrait appliquer le protocole de verrouillage en deux phases strict (2PL strict). Indiquer la séquence d'actions obtenue en sortie lorsqu'il n'y a pas d'interblocage. S'il y a un interblocage, dessiner le graphe des attentes à la place.

**Réponse :** Barrer la mauvaise réponse

 INTERBLOCAGE PAS D'INTERBLOCAGE

**Ordre en sortie ou Graphe des attentes**

## 4 Algèbre relationnelle (4 pts)

On considère le schéma relationnel suivant qui permet de stocker des données géographiques :

**PAYS** (idP, NomP, Superficie, Population, *idC*\*)

**CONTINENT** (idC, NomC, Superficie)

**FRONTIERES** (*idP*\*, *idPF*\*, longueur)

**LANGUE** (CodeL, NomL)

**LANGUESPAYS** (*idP*\*, *CodeL*\*, pourcentage)

**MONTAGNE** (idM, NomM, Altitude)

**MONTAGNESPAYS** (*idP*\*, *idM*\*)

Les attributs soulignés représentent les clés primaires, les attributs avec astérisque représentent les clés étrangères. Un pays a un identifiant, un nom, on connaît sa population et sa superficie (en  $km^2$ ) et le continent auquel il appartient. Un continent a un identifiant, on connaît son nom et sa superficie. La table **FRONTIERES** stocke les couples de pays frontaliers, avec la longueur de chaque frontière. La relation **FRONTIERES** n'est pas symétrique, chaque couple de pays est stocké une seule fois, *i.e.* si (*idPays1*, *idPays2*, long) existe dans **FRONTIERES** alors (*idPays2*, *idPays1*, long) n'y est pas. Pour chaque langue on connaît son code (identifiant) et son nom. Les langues parlées dans chaque pays sont enregistrées dans la table **LANGUESPAYS**, pour chaque langue on connaît le pourcentage de la population qui la parle. Pour chaque montagne on connaît son identifiant, son nom et son altitude. Une montagne peut se trouver dans plusieurs pays (table **MONTAGNESPAYS**).

### Question 13 (1 point)

Les noms des pays où plus de 30% de la population parle français.

Réponse :

### Question 14 (1 point)

Les noms et les altitudes des montagnes qui se trouvent à la fois en France et en Italie.

Réponse :

### Question 15 (1 point)

Les langues qui sont parlées sur tous les continents où l'on parle l'anglais (*i.e.* il existe un pays de ce continent où on parle l'anglais).

Réponse :

**Question 16** (1 point)

Le nom de la montagne la plus haute.

Réponse :

**5 Optimisation de requêtes (5 pts)**

Un concert se déroule dans une ville à une date donnée. Les personnes sont identifiées par leur numéro. Elles achètent des billets de concert à un prix donné.

**Concert** (numC, ville, date)

**Persone** (numP, nom, prénom, âge)

**Billet** (numC\*, numP\*, prix)

Il y a 100 villes différentes :  $v_1, v_2, \dots, v_{100}$ . L'âge est dans  $[10, 110[$ . Le prix est dans  $[5, 55[$

**Question 17** (1 point)

Soit la requête R1 :

```
select p.prénom, c.ville
from Concert c, Persone p, Billet b,
where p.numP = b.numP and b.numC = c.numC
and p.age = 18 and b.prix < 10
```

P1 est le plan d'exécution de R1 pour lequel on traite les sélections puis les projections le plus tôt possible. Dessiner l'arbre de P1. Inscrive les feuilles de l'arbre sur les traits pointillés en bas du dessin, la racine est en haut.

Réponse :

.....

.....

.....

**Question 18** (1 point)

Pour chaque prédicat de sélection s1 à s3, quel est son facteur de sélectivité ?

s1 :  $age < 30$

s2 :  $ville = 'Paris' \text{ or } ville = 'Lyon'$

s3 :  $prix > 10 \text{ and } prix < 50$

**Réponse :**

s1                      facteur de sélectivité=

s2                      facteur de sélectivité=

s3                      facteur de sélectivité=

**Question 19** (1 point)

Coût d'une sélection. Les données sont stockées sans ordre particulier. Le coût pour lire une page d'une relation vaut 1. L'accès à un nuplet indexé se fait en une lecture de page.

Il y a 1000 billets stockés dans 100 pages. L'attribut prix est indexé.

Il y a 500 personnes stockées dans 50 pages. L'attribut âge est indexé.

On considère les requêtes R2 :  $\sigma_{prix=54} Billet$  et R3 :  $\sigma_{age < 60} Personne$

Pour chaque requête, quel est son coût avec et sans utiliser d'index ?

**Réponse :**

coût R2 avec index =

coût R2 sans index =

coût R3 avec index =

coût R3 sans index =

**Question 20** (2 points)

Les attributs  $age$ ,  $prix$ ,  $Personne.numP$  et  $Billet.numP$  sont indexés. On considère les plans

P1 :  $\sigma_{age < 60}(\sigma_{prix=54} Billet \bowtie Personne)$  et P2 :  $\sigma_{prix=54}(\sigma_{age < 60} Personne \bowtie Billet)$

Les jointures sont traitées par boucle imbriquées. Penser à utiliser des index lorsque cela permet de diminuer le coût d'un plan. Donner le coût des plans P1 et P2 en précisant les index utilisés.

**Réponse :**

P1 :    index utilisés :                       $cout_{P1} =$

P2 :    index utilisés :                       $cout_{P2} =$