



Votre numéro d'anonymat :

--	--	--

Bases de données – LI341

Examen - première session du 13 décembre 2011

Durée : 2 heures – CORRIGÉ

Documents autorisés

Les téléphones mobiles doivent être éteints et rangés dans les sacs. Le barème sur 20 points (17 questions) n'a qu'une valeur indicative.

1 Requêtes (8 pts)

On considère le schéma relationnel suivant, qui permet de gérer l'installation de décorations de Noël dans différentes mairies de l'Ile de France :

DÉCORATION(code, type, taille, couleur, prix)KIT(kit, type, description)COMPOSITION(kit, code, quantité)MAIRIE(code-postal, nom, rue, téléphone)AFFECTATION(kit, code-postal, date)

Chaque *décoration* de Noël est identifiée par un code et caractérisée par un type de décoration (“sapin”, “guirlande”, “maisonnette”, etc.), sa taille, sa couleur et son prix unitaire. Un *kit de décoration* est identifié par un code (attribut *kit*) et décrit par un type (“sapin décoré”, “crèche”, “village éclairé”, etc.) et une description détaillée. Chaque kit regroupe plusieurs décorations, chacune dans une certaine quantité (table Composition). Une mairie est identifiée par son code postal, et elle possède un nom (“Mairie de Bondy”), la rue dans laquelle elle se situe et un numéro de téléphone. Une mairie reçoit des kits de décoration de Noël, qui sont installés à une certaine date (table Affectation).

Question 1 (1 point)

Caractériser en français le résultat de la requête suivante :

$$\{k.kit \mid Kit(k) \wedge \forall c, d (Composition(c) \wedge Décoration(d) \wedge c.kit = k.kit \wedge c.code = d.code \rightarrow d.couleur = 'rouge')\}$$

Solution:

Les kits composés uniquement de décorations en rouge.

DÉCORATION(code, type, taille, couleur, prix) KIT(kit, type, description) COMPOSITION(kit, code, quantité)
MAIRIE(code-postal, nom, rue, téléphone) AFFECTATION(kit, code-postal, date)

Exprimez dans le(s) langage(s) indiqué(s) les requêtes suivantes :

Question 2 (1 point)

Le noms des mairies ayant reçu des kits devant être installés à des dates différentes (algèbre).

Solution:

$$\pi_{nom}(Mairie \bowtie \pi_{code-postal} (\rho_{kit \rightarrow k, code-postal \rightarrow m, date \rightarrow d}(Affectation) \bowtie_{m=code-postal \wedge d \neq date} (Affectation)))$$

$$\pi_{nom}(Mairie \bowtie \sigma_{d \neq date}(\rho_{date \rightarrow d}(\pi_{date, code-postal}(Affectation)) \bowtie Affectation))$$

Question 3 (2 points)

Les types de kits composés uniquement de décorations en rouge (algèbre et SQL).

Solution:

Les types de kits où *tous les kits* sont composés uniquement de décorations en rouge.

$$\pi_{type} Kit - \pi_{type}(Kit \bowtie \pi_{kit}(\sigma_{couleur <> 'rouge'} Decoration \bowtie Composition))$$

```
select type
from Kit k
where type not in (select c.type
                   from Kit k, Decoration d, Composition c
                   where d.code=c.code
                   and k.kit = d.kit
                   and d.couleur <> 'rouge');
```

Les types de kits où *au moins un kit* est composé uniquement de décorations en rouge.

$$\pi_{type}(Kit \bowtie (\pi_{kit} Kit - \pi_{kit}(\sigma_{couleur <> 'rouge'} Decoration \bowtie Composition)))$$

```
select type
from Kit k
where kit not in (select c.kit
                  from Decoration d, Composition c
                  where d.code=c.code
                  and d.couleur <> 'rouge');
```

DÉCORATION(code, type, taille, couleur, prix) KIT(kit, type, description) COMPOSITION(kit, code, quantité)
MAIRIE(code-postal, nom, rue, téléphone) AFFECTATION(kit, code-postal, date)

Question 4 (1 point)

Les types de décorations utilisés dans *tous* les kits de type “sapin décoré” (algèbre).

Solution: algèbre

Les types des décorations utilisées dans *tous* les kits de type “sapin décoré” :

$$\pi_{type}(Décoration \bowtie (\pi_{kit, code} Composition \div \pi_{kit}(\sigma_{type='sapin décoré'} Kit)))$$

Les types de décorations utilisés dans *tous* les kits de type “sapin décoré” :

$$\pi_{type, kit}(Décoration \bowtie Composition) \div \pi_{kit}(\sigma_{type='sapin décoré'} Kit)$$

Question 5 (1 point)

Pour chaque kit et pour chaque couleur, le nombre et le prix total des décorations de cette couleur contenues dans ce kit (SQL).

Solution:

```
select c.kit, d.couleur, count(d.code), sum(c.quantité * d.prix)
from Composition c, Décoration d,
where k.code = d.code
group by c.kit, d.couleur;
```

Question 6 (2 points)

Les codes des kits qui sont composés d'un nombre égal de décorations de couleur "rouge" et de couleur "bleue" (SQL).

Solution:

```
select c.kit
from Composition c, Décoration d
where k.code = d.code and d.couleur='rouge'
group by c.kit
having count(d.code) = (select count (d.code)
                        from Composition e, Décoration f
                        where e.code = f.code and f.couleur='bleu'
                        and e.kit = c.kit);
```

2 Dépendances fonctionnelles (5 pts)

Soit la relation

CRISE (Agence, Banque, Croissance, Dette, Emplois, Note, Pays, Taux, Zonemonetaire)

et deux ensembles de dépendances fonctionnelles \mathcal{F} et \mathcal{G} :

$$\mathcal{F} = \{ AP \rightarrow N, \quad BNP \rightarrow C, \quad CD \rightarrow EN, \quad E \rightarrow D, \quad N \rightarrow T, \quad P \rightarrow CDZ, \quad T \rightarrow C \}$$

$$\mathcal{G} = \{ CD \rightarrow EN, \quad E \rightarrow D, \quad N \rightarrow T, \quad P \rightarrow CDZ, \quad T \rightarrow C \}$$

Chaque attribut est indiqué par sa première lettre et AP désigne l'ensemble d'attributs $\{A, P\}$. On peut montrer que \mathcal{G} est une couverture minimale.

Question 7 (1 point)

Montrez formellement que \mathcal{F} et \mathcal{G} sont équivalents.

Solution: \mathcal{G} est un sous-ensemble de \mathcal{F} . Il suffit de montrer qu'on peut retrouver les DF dans \mathcal{F} à partir de \mathcal{G} .

$$P \rightarrow CDZ, CD \rightarrow EN \models P \rightarrow EN \models AP \rightarrow N$$

$$N \rightarrow T, T \rightarrow C \models N \rightarrow C \models BNP \rightarrow C$$
Question 8 ($\frac{1}{2}$ point)

Donner la ou les clés (minimales) de la table CRISE ? Prouver votre réponse (une réponse sans justification ne sera pas acceptée).

Solution: ABP font partie de toutes les clés.

ABP+=ABPCDZENT

Question 9 (1½ points)

Vérifiez avec la méthode tableau si la décomposition de la table CRISE en

- R1(A,B,C,D,P),
- R2(C,D,E,N),
- R3(P,T,Z)

est SPI. Indiquez le tableau initial, les dépendances fonctionnelles utilisées et le tableau final.

Solution:

	A	B	C	D	E	N	P	T	Z
R1	a	b	c	d	x15	x16	p	x18	x19
R2	x21	x22	c	d	e	n	x27	x28	x29
R3	x31	x32	x33	x34	x35	x36	p	t	z

$P \rightarrow CDZ$: $x33=c$, $x34=d$, $x19=z$

$CD \rightarrow EN$: $x15=e$, $x16=n$, $x35=e$, $x36=n$

$N \rightarrow T$: $x18=t$

	A	B	C	D	E	N	P	T	Z
R1	a	b	c	d	e	n	p	n	z
R2	x21	x22	c	d	e	n	x27	x28	x29
R3	x31	x32	c	d	e	n	p	t	z

Question 10 (1 point)

Donnez une décomposition de CRISE qui est SPI, SPD et où chaque relation est en 3e FN. Donnez pour chaque relation créée ses dépendances fonctionnelles et ses clés.

Solution: On utilise \mathcal{G} (couverture minimale) :

étape 1 :

R1(CDEN)

R2(ED)

R3(NT)

R4(PCDZ)

R5(TC)

étape 2 : ajouter table pour clé

R6(ABP)

étape 3 : regrouper R1 et R2 :

R12(CDEN), $CD \rightarrow EN$, $E \rightarrow D$, clés : CD et CE

R3(NT), $N \rightarrow T$, clés : N

R4(PCDZ), $P \rightarrow CDZ$, clés : P

R5(TC), $T \rightarrow C$ clés : T

R6(ABP) \emptyset clés : ABP

Question 11 (1 point)

Si la note d'un pays a est meilleure (plus grande) que la note d'un pays b , le taux pour a est inférieur au taux pour b . Donnez la formule logique qui exprime cette contrainte.

Solution:

$$\forall x, y (CRISE(x) \wedge CRISE(y) \wedge x.note \geq y.note \rightarrow x.taux \leq y.taux)$$

3 Concurrency (2 pts)

On considère les transactions t_1, t_2, t_3 et t_4 qui accèdent à trois granules x, y, z d'une base de données et les deux exécutions suivantes :

$$\begin{aligned} H_1 &: e_1(x) \ l_2(y) \ l_4(y) \ e_1(y) \ e_2(z) \ e_3(z) \ l_3(x) \ l_3(y) \ l_4(z) \\ H_2 &: e_2(z) \ l_2(y) \ l_4(y) \ e_3(z) \ e_1(x) \ e_1(y) \ l_4(z) \ l_3(y) \ l_3(x) \end{aligned}$$

où e_i et l_i représentent respectivement, des opérations d'écriture et de lecture de la transaction t_i .

Question 12 (1 point)

H_1 et H_2 sont-elles sérialisables ? Montrez les graphes de précédence correspondants, en indiquant sur chaque arc le(s) granule(s) pour le(s)quel(s) la précédence correspondante existe.

Solution: Pour H_1 :

- sur x : $T1 \rightarrow T3$,
- sur y : $T2 \rightarrow T1$, $T1 \rightarrow T3$, $T4 \rightarrow T1$
- sur z : $T2 \rightarrow T3$, $T2 \rightarrow T4$, $T3 \rightarrow T4$

Pour H_2 :

- sur x : $T1 \rightarrow T3$,
- sur y : $T2 \rightarrow T1$, $T1 \rightarrow T3$, $T4 \rightarrow T1$
- sur z : $T2 \rightarrow T3$, $T2 \rightarrow T4$, $T3 \rightarrow T4$

On a les mêmes conflits pour H_1 et H_2 .

Question 13 (1 point)

Les exécutions H_1 et H_2 sont-elles équivalentes ? **Détaillez votre raisonnement.**

Solution: Non, les opérations de la transaction $T2$ ne sont pas exécutées dans le même ordre : dans H_2 elle commence par la lecture sur y , tandis que dans H_1 elle commence par une écriture sur z .

4 Optimisation (5 pts)

Soit $R(A,B,C,D)$ et $S(E,F,G)$ deux relations telles que :

- A est clé primaire de R ,
- a est la taille de l'attribut A ,
- T_R est la taille d'un n-uplet de R ,
- T_S est la taille d'un n-uplet de S ,
- N_R est la cardinalité de R ,
- N_S est la cardinalité de S .

4.1 Exercice 1

Dans un premier temps, on effectue la jointure

$$R \bowtie_{A=E} S$$

(equi-jointure) sur une seule machine. On suppose que le coût de traitement est approximé par la quantité totale d'information traitée, comme vu en TD.

Question 14 (1 point)

Soit N_Q le nombre de n-uplets du résultat de la jointure. Donner la taille T_Q des n-uplets de ce résultat, et en déduire le coût total C_Q de la jointure (en utilisant T_R , T_S , N_R , N_Q et N_S).

Solution:

$$TQ = TR + TS$$

$$CQ = NS * TS + NR * TR + NQ * TQ$$

4.2 Exercice 2

On suppose maintenant qu'on dispose de deux machines, M1 et M2. R est stockée sur la machine M1, S est stockée sur la machine M2. La jointure est demandée par un utilisateur sur la machine M1. On suppose que le coût de transfert d'un octet vaut deux fois son coût d'accès local. On propose deux stratégies pour effectuer cette jointure, dont on veut comparer le coût.

4.2.1 Stratégie S_1

La stratégie S_1 consiste à transférer entièrement S de M2 vers M1, puis effectuer la jointure sur M1 où le résultat peut être fourni à l'utilisateur, soient les coûts suivants, correspondants aux différentes étapes :

1. Lecture de S : C_{S11}
2. Transfert de S de M2 vers M1 : C_{S12}
3. Jointure $R \bowtie_{A=E} S$ sur M1 : C_{S13}

Question 15 ($1\frac{1}{2}$ points)

Donner les coûts C_{S11} , C_{S12} , C_{S13} et $C_{S1} = C_{S11} + C_{S12} + C_{S13}$ de la stratégie S_1 .

Solution:

$$CS11 = NS * TS$$

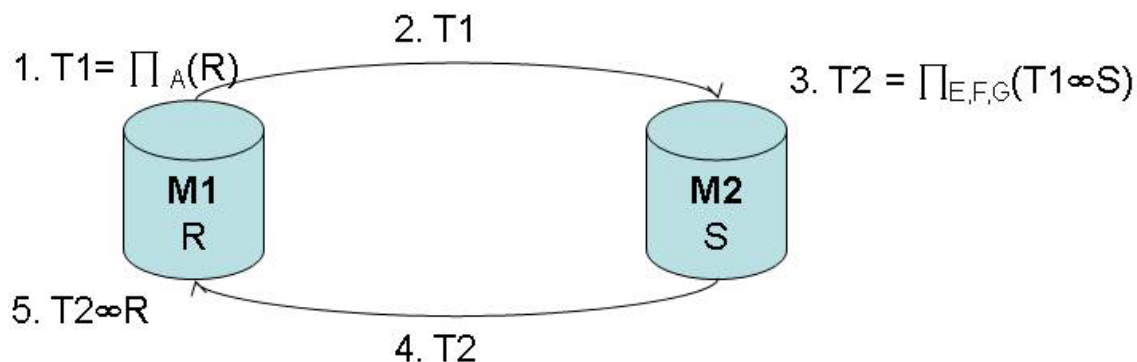
$$CS12 = 2 * NS * TS$$

$$CS13 = CQ$$

$$CS1 = CQ + 3 * NS * TS$$

Stratégie S_2

La stratégie S_2 consiste à utiliser une semi-jointure (\ltimes), comme indiqué sur la figure ci-dessous.



Les coûts sont les suivants, correspondants aux différentes étapes :

1. On effectue $\Pi_A(R)$ sur M1, coût C_{S21}
2. On transfère le résultat T1 sur M2, coût C_{S22}
3. Sur M2, on effectue $\Pi_{E,F,G}(T1 \bowtie S)$, coût C_{S23}
4. On transfère le résultat T2 sur M1, coût C_{S24}
5. Enfin, sur M1, on calcule $T2 \bowtie R = R \bowtie S$, coût C_{S25}

On suppose que la relation R reste en mémoire (donc pas de deuxième coût d'accès) entre la première et la dernière étape.

Question 16 ($\frac{1}{2}$ point)

Caractériser en français les n-uplets de T2 en fonction des n-uplets des tables R et S.

Solution: Les n-uplets de S qui joignent avec au moins un n-uplet de R

Question 17 (2 points)

Donner les coûts C_{S21} , C_{S22} , C_{S23} , C_{S24} , C_{S25} et $C_{S2} = C_{S21} + C_{S22} + C_{S23} + C_{S24} + C_{S25}$ de la stratégie S_2 .

Solution:

$$CS21 = NR * TR + NR * a$$

$$CS22 = 2 * (NR * a)$$

$$CS23 = NS * TS + (NQ * TS)$$

$$CS24 = 2 * (NQ * TS)$$

$$CS25 = NQ * TQ$$

$$CS2 = NR * TR + 3 * (NR * a) + NS * TS + 3 * (NQ * TS) + NQ * TQ$$

Question bonus (1 points)

On suppose que $N_S = N_R = 1000$, $T_S = 10$, $T_R = 20$, $a = 5$. Exprimer les coûts C_{S1} et C_{S2} en fonction de N_Q .

Solution:

$$CS1 = 60000 + 30 NQ$$

$$CS2 = 45000 + 60 NQ$$