

--	--	--

Systèmes de Gestion de Bases de Données – 3I009

EXAMEN - 1^{ère} session du 11 janvier 2018

Durée : 2 heures – CORRIGÉ

Documents autorisés

Les téléphones mobiles doivent être éteints et rangés dans les sacs. Le barème sur 21^{1/2} points (23 questions) n'a qu'une valeur indicative.

1 Questions de cours et de TME (3 pts)

Question 1 (1 point)

Exprimez $T = R(XY) \div S(X)$ la division algébrique de la table $R(XY)$ par la table $S(X)$ en fonction des opérateurs de base de l'algèbre relationnelle. Vous pouvez utiliser des variables pour simplifier l'expression

Question 2 (1 point)

Quel est le seul type de jointure accepté par RAEval ? Que doit on faire obligatoirement pour effectuer une jointure d'un autre type ?

Question 3 (1 point)

Comment le SGBD fait-il pour éliminer les doublons lors de l'évaluation d'une requête `SELECT DISTINCT ...` ? Donner deux solutions possibles

2 Dépendances fonctionnelles (3 pts)

Soit une table **Voyages(N,P,D,R,A,S,T,B)** qui stocke des informations sur des voyages avec le Nom (N) et le Prénom (P) du voyageur, la date de Départ (D), la date de retour (R), l'Adresse de destination (A), le prix du Séjour (S), le moyen de Transport (T) et le prix du Billet (B). On observe l'ensemble de dépendances fonctionnelles suivant sur **Voyages** :

$$\mathcal{F} = \{ \text{N P D} \rightarrow \text{R A S T B}; \quad \text{D R A} \rightarrow \text{S}; \quad \text{D A T} \rightarrow \text{B}; \quad \text{D A S} \rightarrow \text{R}; \quad \text{R A S} \rightarrow \text{D}; \}$$

Question 4 (1 point)

Donnez les deux clés (minimales) de la table Voyages par rapport à \mathcal{F} . Expliquer brièvement comment vous les obtenez.

**Question 5** ($\frac{1}{2}$ point)

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

**Question 6** ($\frac{1}{2}$ point)

Montrez que \mathcal{F} contient au moins une dépendance fonctionnelle redondante et donnez les autres dépendances qui permettent de la retrouver.

**Question 7** (1 point)

Est-ce que la décomposition suivante de **Voyage** est sans perte d'information (SPI) par rapport à \mathcal{F} ? Justifiez votre réponse en utilisant la méthode du tableau (montrez quelles DF sont utilisées)

- Reservation(N,P,D,R,A)
- PrixSéjour(D,R,A,S)
- PrixBillet(D,A,T,B)

**3 Contrôle de Concurrency (3 $\frac{1}{2}$ pts)**

Soient les transactions T1 à T6 accédant aux granules A à D . On considère l'exécution S_1 suivante, où L_i et E_i représentent une lecture et une écriture faites par la transaction T_i (on suppose que les

opérations sont exécutées dans l'ordre d'arrivée) :

$$L_1(B)L_6(C)L_2(B)L_1(A)E_4(C)L_3(A)E_6(B)E_2(A)L_1(C)L_3(B)E_5(A)L_5(B)$$

Question 8 (1 point)

Préciser pour chaque granule la séquence d'opérations le concernant et les arcs de précédence, notés $T_i \rightarrow T_j$.

Question 9 ($\frac{1}{2}$ point)

Précisez quels sont les circuits existants dans le graphe de précédence correspondant.

Question 10 ($\frac{1}{2}$ point)

Choisir une seule transaction à enlever parmi les six afin que l'exécution obtenue S_2 soit sérialisable. Donner l'exécution en série équivalente.

Question 11 (1 point)

Soient les transactions T1 à T4 accédant aux granules A à C. On considère le début d'exécution S_3 suivant où **OP** est une opération à préciser. Le gestionnaire de concurrence utilise un verrouillage en 2 phases strict pour contrôler l'exécution des opérations. Au moment de **OP**, toutes les transactions ont encore des opérations à effectuer (aucune des transactions n'a terminé).

$$L_1(A)E_4(B)E_2(C)L_2(B)L_4(B)E_2(B)L_3(A)E_1(A)L_1(B)L_3(C)OP \dots$$

Donner un exemple d'opération **OP** qui produit un inter-blocage dans lequel toutes les 4 transactions sont impliquées. Dessiner le graphe d'attente

Question 12 ($\frac{1}{2}$ point)

Donner l'exécution effective des opérations jusqu'à l'inter-blocage.

4 Algèbre relationnelle (4 pts)

Vol (nvol, villeD, villeA, heureD, heureA, distance) **Appareil** (idAvion, distMin, distMax)

Membre (pid, nom, debutContrat)

Habilité (pid*, idAvion*, poste)

On considère un schéma relationnel sur les vols, les appareils utilisés et le personnel navigant. Pour chaque vol, identifié par son numéro (nvol), on connaît sa ville de départ (villeD), sa ville d'arrivée (villeA), son heure de départ (heureD), son heure d'arrivée (heureA) et la distance parcourue (distance). Pour chaque avion, identifié par un numéro (idAvion), on connaît la distance minimale qu'il doit parcourir pour être rentable (distMin) et la distance maximale qu'il peut parcourir (distMax) sans avoir besoin de ravitaillement. Pour chaque membre du personnel, identifié par son pid, on connaît son nom et le début de son contrat ; on connaît également, avec la table **Habilité**, les appareils dans lesquels il est habilité à naviguer et dans quel poste.

Exprimer en **algèbre relationnelle** les requêtes qui retournent les informations suivantes.

Remarque. Seuls les opérateurs de l'algèbre relationnelle sont autorisés : π , σ , \bowtie , ρ , \times , \div , \cup , \cap , $-$.

Question 13 (1 point)

Pour chaque membre du personnel, les vols sur lesquels il peut opérer. Un membre peut opérer sur un vol si il est habilité à voler dans un appareil rentable pour ce vol. Un appareil est rentable pour un vol si la distance du vol est comprise entre la distance minimale et la distance maximale de l'appareil. Le résultat doit être un ensemble de paires (*pid*, *nvol*).

Question 14 (1 point)

Les correspondances possibles entre deux vols (v_1, v_2) tels que la ville de départ de v_2 est la ville d'arrivée de v_1 et l'heure de départ de v_2 est supérieure à l'heure d'arrivée de v_1 . On retournera des paires de numéros de vols.

Question 15 (1 point)

Les villes où il n'existe aucun vol ni en partance vers 'Paris' ni en arrivée depuis 'Paris'.

Question 16 (1 point)

Les identifiants des membres qui sont habilités à voler sur tous les appareils rentables pour un vol de 1000km.

5 Optimisation de requêtes (5 pts)

On considère le schéma d'une agence de voyage.

Personne (numP, prenom, age, ville, statut, civilité, profil) *ville fait référence à VillePays.*

VillePays (ville, pays) *ville est unique.*

Destination (numD, villeD, catégorie) *villeD fait référence à VillePays.*

Séjour (numP, numD, année, commentaire)

On connaît les cardinalités et les tailles suivantes (ne pas confondre cardinalité et taille) :

Relation	Cardinalité	Rmq	Taille en pages
Personne	2 000	seules 50 villes sont référencées dans Destination seules 100 villes sont référencées dans Personnes	1000 pages
VillePays	4 000		
Destination	100		
Séjour	20 000	200 séjours par numD et 10 séjours par numP	1000 pages

On connaît les domaines des attributs suivants :

Relation	Attribut	Nb valeurs	Rmq
Personne	age	50	dans [20, 70[
Personne	ville	100	'Paris', ...
Personne	statut	5	'A', 'B', 'C', 'D' ou 'E'
Personne	civilité	2	'Mme' ou 'Mr'
VillePays	pays	200	'Fr', 'Es', ...
Destination	villeD	50	'Rio', ...
Destination	catégorie	4	'mer', 'montagne', ...
Séjour	numD	100	1, ..., 100
Séjour	année	10	dans [2008, 2017[

Question 17 (1 point)

On considère un prédicat s_i et le facteur de sélectivité $SF_i = SF(\sigma_{s_i} Personne)$. Calculer SF_i .

s_1 : age BETWEEN 25 AND 34

s_2 : age < 25 OR age >= 60

s_3 : ville = 'Paris' AND age = 25

s_4 : civilité = 'Mr' OR statut <> 'A'

NB. l'opérateur <> signifie **différent de**

Question 18 (1 point)

Quelle est la cardinalité $C_i = \text{card}(R_i)$ des requêtes suivantes ? Justifier votre réponse.

$R_1 = \text{Personne} \bowtie_{\text{ville}} \text{VillePays}$

$R_2 = \sigma_{\text{civilité}='Mr'} (\text{Personne} \bowtie_{\text{numP}} \text{Séjour})$

$R_3 = \text{Séjour} \bowtie_{\text{numD}} \text{Destination} \bowtie_{\text{numP}} \text{Personne}$

$R_4 = \sigma_{\text{année}=2016} \text{Séjour} \bowtie_{\text{numP}} \text{Personne} \bowtie_{\text{numD}} (\sigma_{\text{villeD}='Rio'} \text{Destination})$

Question 19 (1 point)

Par les questions suivantes, tous les coûts sont exprimés en nombre de pages lues. Le coût d'un accès par **index** est $\text{coût}(\sigma_{a \text{ op } x}(R)) = \text{card}(\sigma_{a \text{ op } x}(R))$ où op est un opérateur ($=$, $<$, $>$) et x est une valeur. Cf. cours : lire une page pour chaque nuplet accédé à partir des ROWID obtenus en parcourant l'index.

Pour T_3 et T_4 , la jointure est faite par boucles imbriquées et son coût est : $\text{coût}(A \bowtie B) = \text{coût}(A) + P(A) * P(B)$ avec $P(R)$ étant la taille de R exprimée en nombre de pages.

Pour chaque plan T_i calculer son coût noté $\text{coût}(T_i)$.

T_1 : $\sigma_{\text{age}=25}\text{Personne}$ utilise l'index $\text{Personne}(\text{âge})$.

T_2 : $\sigma_{\text{numD}=3}\text{Séjour}$ utilise l'index sur $\text{Séjour}(\text{numD})$.

T_3 : $(\sigma_{\text{age}=25}\text{Personne}) \bowtie_{\text{numP}} \text{Séjour}$ utilise l'index $\text{Personne}(\text{âge})$.

T_4 : $\sigma_{\text{age}=25}((\sigma_{\text{numD}=3}\text{Séjour}) \bowtie_{\text{numP}} \text{Personne})$ utilise l'index $\text{Séjour}(\text{numD})$.

Question 20 (1 point)

On rappelle qu'une jointure par **hachage** accède à tous les nuplets de A puis à tout ceux de B, A et B pouvant être des opérations de sélection, et son coût est : $\text{coût}(A \bowtie B) = \text{coût}(A) + \text{coût}(B)$.

Soit la requête H_1 composée d'une jointure et deux sélections :

select * **from** Séjour s, Personne p

where s.numP = p.numP **and** s.numD = 3 **and** p.âge = 25

Proposer un plan pour H_1 qui utilise les **index** sur $\text{Séjour}(\text{numD})$ et $\text{Personne}(\text{âge})$ et évalue la jointure par **hachage**. Détailler les principales étapes de l'exécution de H_1 et donner son coût.

Question 21 (1 point)

Soit la requête J_1 :

select * from Séjour s, Personne p
where s.numP = p.numP **and** s.année = 2016 **and** p.âge = 25

J_1 est évaluée par **boucles imbriquées avec index sur l'attribut de jointure**, et utilise les **index** Séjour(numP) et Personne(âge). Expliquer les principales étapes de l'exécution. Donner le coût du plan.



Soit la requête J_2 : **select * from** Séjour s, Personne p
where s.numP = p.numP **and** s.année = 2016 **and** p.âge > 60

Le plan optimal pour évaluer J_2 a un coût égal à 1400. Préciser les index et l'algorithme de jointure que ce plan utilise (parmi les index et les algorithmes indiqués ci-dessous). Justifier.



6 Triggers (3 pts)

On considère le schéma relationnel suivant :

CLIENT (numClient, nomClient, adresse)

PRODUITS(numP, nomP, prix, qteStock)

COMMANDE (numC, numClient, dateCom, numP, qte, cout)

Un client est identifié par un numéro (*numClient*). Il a un nom (*nomClient*) et une adresse (*adresse*).

Un produit a un numéro (*numP*), un nom (*nomP*) un prix unitaire (*prix*). L'attribut *qteStock* donne le nombre de produits en stock.

Une commande a un numéro de commande (*numC*). Elle est passée à une date (*dateCom*), pour le client (*numClient*). Elle concerne un produit (*numP*). Les attributs *qte* et *cout* indiquent la quantité du produit commandé et le cout total de la commande.

Question 22 (1½ points)

Lorsqu'on passe une commande, la quantité en stock doit être mise à jour. Ecrire un trigger AFTER gérant la quantité en stock des produits en fonction des commandes.

Question 23 (1½ points)

Ecrire un trigger BEFORE qui fait une réduction de 10 % aux clients qui achètent plus de 3 produits le même jour (il y a une commande par produit). La réduction s'applique à partir de la troisième commande le même jour.