

Bases de Données I (J. Wijsen)

27 janvier 2012

NOM + PRENOM :

Orientation + Année :

Cet examen contient 14 questions. Durée : 2 heures et 30 minutes.

Question 1 Soit R un nom de relation avec schéma AB . Soit S un nom de relation avec schéma A . Considérez la requête q_1 :

$$q_1 = \{x \mid \exists y (R(x, y) \wedge \neg (S(x) \wedge S(y)))\}$$

Donnez une requête en algèbre relationnelle qui est équivalente à q_1 . Détaillez votre réponse.

Chaque tuple $\{A:a, B:b\}$ dans R contribue "a" à la réponse sauf si S contient à la fois $\{A:a\}$ et $\{A:b\}$.

Les tuples de R qui ne contribuent pas à la réponse sont donc tous dans :

$$S \bowtie_{A \rightarrow B} S.$$

La requête suivante est donc équivalente à q_1 :

$$\pi_A (R - (S \bowtie_{A \rightarrow B} S)).$$

Question 2 Considérez la requête :

$$\{x, y \mid R(x, y) \vee \exists y(R(x, y) \wedge \neg S(x))\}$$

Cochez chaque case (éventuellement plusieurs) qui précède une expression correcte :

- Cette requête est *domain independent*.
 Cette requête n'est pas *domain independent*.

Expliquez en détail.

.../5

Prenons $R \begin{array}{c|c} 1 & 2 \\ \hline a & c \end{array}$ $S \begin{array}{c|c} 1 & 2 \\ \hline & \end{array} \rightarrow S$ est vide

Pour cette base de données, $\langle a, b \rangle$ est dans la réponse pour un b quelconque, car $\exists y (R(a, y) \wedge \neg S(a))$ est vrai.

↪ Notez que ceci n'est pas "le même y " que celui au début de la requête!

$x \rightarrow a$
 $y \rightarrow b$

Question 3 Considérez la requête :

$$\{y \mid \exists x \neg \exists z (\neg R(x, y) \vee R(z, y))\}$$

Cochez chaque case (éventuellement plusieurs) qui précède une expression correcte :

- Cette requête est *domain independent*.
 Cette requête n'est pas *domain independent*.

Expliquez en détail.

.../5

Pour toute base de données, la réponse sera toujours vide. Supposons que la réponse n'est pas vide pour une bd. Alors, il existe b et a tel que la bd satisfait :

$$\neg \exists z (\neg R(a, b) \vee R(z, b)),$$

ce qui est équivalent à :

$$R(a, b) \wedge \forall z (\neg R(z, b)).$$

Donc, bd doit contenir $R(a, b)$ et ne pas contenir $R(a, b)$, une contradiction.

$x \rightarrow a$
 $y \rightarrow b$

Question 4 Quelle est la motivation principale pour passer de 2PL à Strict 2PL ? Expliquez de façon précise à l'aide d'un exemple.

Voir Syllabus.

.../5

Question 5 Prouvez qu'il existe une requête q en algèbre SPJRUD qui n'est équivalente à aucune requête en algèbre SPJRD (i.e., l'algèbre sans l'union). N'oubliez pas de préciser les attributs des relations qui figurent dans q .

.../5

Cette question est résolue
dans le Syllabus.

Question 6 Considérez le schéma avec attributs $ABCDEFGH$ et les DF suivantes :

$$\Sigma = \begin{array}{lll} AB \rightarrow CD & EF \rightarrow GH & BCD \rightarrow A \\ G \rightarrow E & B \rightarrow F & F \rightarrow A \end{array}$$

Cochez chaque case (éventuellement plusieurs) qui précède une expression correcte :

- Ce schéma est en BCNF.
- Ce schéma n'est pas en BCNF.
- Ce schéma est en 3NF.
- Ce schéma n'est pas en 3NF.

Détaillez les calculs qui mènent à cette conclusion.

Chaque clé doit contenir B, parce que .../5
 B n'apparaît pas à droite d'une flèche.

$\Sigma \models B \rightarrow BFACD$

Donc, aucune clé ne contient un attribut parmi A, C, D, F.

$\Sigma \models BE \rightarrow \underline{BFACD} \textcircled{E} GH$ BE est une clé

$\Sigma \models BG \rightarrow \underline{BFACD} \textcircled{G} EH$ BG est une clé

$\Sigma \not\models BH \rightarrow G$, donc BH n'est pas une clé

Il est correct de conclure que BE et BG sont les seules clés.

Ce schéma n'est pas en 3NF (et donc pas en BCNF) parce que

Σ contient $B \rightarrow F$,
 B n'est pas superclé, et
 F ne fait partie d'aucune clé.

Question 7 Considérez le schéma avec attributs $ABCDE$ et les DF suivantes :

$$\Sigma = \left\{ \begin{array}{l} AB \rightarrow D \quad ABC \rightarrow E \quad CDE \rightarrow AB \quad ADE \rightarrow C \end{array} \right\}$$

Cochez chaque case (éventuellement plusieurs) qui précède une expression correcte :

- Ce schéma est en BCNF.
- Ce schéma n'est pas en BCNF.
- Ce schéma est en 3NF.
- Ce schéma n'est pas en 3NF.

Détaillez les calculs qui mènent à cette conclusion.

.../5

$\Sigma \models ABC \rightarrow DE$, donc ABC est superclé.

On peut vérifier que AB , AC , BC ne sont pas superclés. Donc ABC est une clé.

$\Sigma \models CDE \rightarrow AB$, donc CDE est superclé.

On peut vérifier que CD , CE , DE ne sont pas superclés. Donc CDE est une clé.

Le schéma est en 3NF, parce que chaque attribut fait partie d'une clé.

Pas en BCNF, car Σ contient $AB \rightarrow D$, mais AB n'est pas superclé.

Question 8 Les 20 km de Bastogne est une course à pieds organisée annuellement depuis 1980. La 33ème édition aura lieu le 27 mai 2012. La table RESULTATS stocke les résultats de toutes les éditions passées :

- Chacun des athlètes porte un dossard avec un numéro unique. Le dossard d'un athlète peut varier d'une année à l'autre.
- CP est le code postal de la localité belge de l'athlète; pour les athlètes non domiciliés en Belgique, cette colonne donne le pays de l'athlète. Supposons que pour une même édition, les attributs Nom et CP constituent une identification unique pour chaque athlète. Donc, si deux participants à une même édition ont le même nom, ils auront une valeur différente pour CP. Par contre, il n'est pas sûr que l'athlète Anne Dua avec code postal 3700 en 1980 est la même personne physique qu'Anne Dua avec code postal 3700 en 2011.
- Nat donne la nationalité de l'athlète (B=Belgique, F=France,...).
- Pour chaque édition, on fait deux classements : un pour les athlètes masculins et un pour les athlètes féminins. Pour une édition donnée, le Rang d'un athlète masculin est n s'il y a exactement $n - 1$ athlètes masculins ayant réalisé un meilleur temps. De façon similaire, le Rang d'une femme est n s'il y a exactement $n - 1$ femmes ayant couru plus vite. Notez qu'en 2011, Hugo Claus et Ivo Michiels ont réalisé le même temps et sont classés 2èmes tous les deux; par conséquent, aucun athlète n'avait le Rang 3 en 2011.
- Moyenne donne la vitesse moyenne (en km/h) de chaque athlète. Le parcours a toujours été exactement 20 km; la vitesse moyenne pourrait donc être calculée à partir du Temps.

RESULTATS									
Edition	Année	Dossard	Nom	CP	Nat	Sexe	Rang	Temps	Moyenne
1	1980	17	Anne Dua	3700	B	F	1	1:15:00	16.00
1	1980	29	Eric Point	8100	F	M	1	1:00:00	20.00
32	2011	37	Tom Michiels	9473	B	M	1	1:00:00	20.00
32	2011	17	Hugo Claus	France	B	M	2	1:01:15	19.59
32	2011	105	Ivo Michiels	9000	F	M	2	1:01:15	19.59
32	2011	39	Jef Geeraerts	9000	B	M	4	1:01:19	19.57
32	2011	99	Jules Verne	France	F	M	934	1:30:00	13.33
32	2011	123	Caroline Smets	2100	B	F	200	1:30:00	13.33
32	2011	1957	Anne Dua	3700	B	F	722	1:45:00	12.63 → 11.43

Quelles sont les dépendances fonctionnelles pour cette table ?

.../10

Edition → Année

Année → Edition

Edition, Dossard → Nom, CP, Nat, Sexe, Temps
 (les dossards servent à identifier les athlètes dans la course)

Edition, Nom, CP → Dossard

Edition, Sexe, Rang → Temps

Edition, Sexe, Temps → Rang

Temps → Moyenne

(Je préfère ne pas mettre Moyenne → Temps, parce que 1:45:00 et 1:45:01 donnent la même valeur pour moyenne).

Question 9 Considérez l'exécution suivante :

$R_1(A)W_2(A)R_3(A)R_1(B)W_2(B)R_1(E)$

Non

Est-ce que cette exécution est possible en 2PL? Complétez l'exécution avec des demandes de verrous ou argumentez pourquoi cette exécution n'est pas possible en 2PL.

.../5

<p>On n'a pas beaucoup de liberté pour mettre les verrous :</p>	<div style="display: flex; flex-direction: column; align-items: center;"> <div style="margin-bottom: 10px;">$S_1(A)$</div> <div style="margin-bottom: 5px;">$R_1(A)$</div> <div style="margin-bottom: 5px;">$S_1(B)$</div> <div style="margin-bottom: 5px;">$S_1(E)$</div> <div style="margin-bottom: 5px;">$U_1(A)$</div> <div style="margin-bottom: 10px;">$X_2(A)$</div> <div style="margin-bottom: 5px;">$W_2(A)$</div> <div style="margin-bottom: 10px;">$R_1(B)$</div> <div style="margin-bottom: 5px;">$U_1(B)$</div> <div style="margin-bottom: 5px;">$X_2(B)$</div> <div style="margin-bottom: 5px;">$W_2(B)$</div> <div style="margin-bottom: 10px;">$R_1(E)$</div> </div>	<p>$U_2(A)$ doit se faire après $X_2(B)$.</p> <p>Alors T_3 effectue $R_3(A)$ au moment où T_2 possède $X_2(A)$, ce qui est impossible en 2PL.</p>
---	--	---

Question 10 Cochez la case qui précède une expression correcte :

- L'exécution de la question 9 est sérialisable.
- L'exécution de la question 9 n'est pas sérialisable.

Argumentez votre réponse.

.../5

Le graphe de précédence ne contient pas de cycle :

```

graph LR
    T1((T1)) --> T2((T2))
    T2 --> T3((T3))
            
```

Question 11 La société *Rainparcs* exploite plusieurs villages de vacances. Ces villages s'appellent Borinage, Ardennes, Côte, etc. Chaque village de vacances contient plusieurs bungalows de vacances. Chaque bungalow est d'un certain type (hyacinthe, rose, tulipe, etc.). La première ligne de la table **TYPE** indique que les bungalows de type "hyacinthe" peuvent accueillir 6 personnes, contiennent une télévision mais pas de lave-vaisselle. La première ligne de la table **VILLAGES** indique que le village "Borinage" contient 12 bungalows de type "hyacinthe". Les bungalows peuvent être loués pour des périodes d'une semaine commençant le lundi. Les premières deux lignes de la table **LOCATIONS** indiquent que le client 111 a loué 3 bungalows de type "tulipe" pour les semaines commençant le 1 août 2011 et le 8 août 2011. Un même client ne peut pas louer deux bungalows dans deux villages différents pour la même période (mais il peut louer plusieurs bungalows dans un même village). Chaque client est identifié par un numéro unique. La table **CLIENT** donne les noms et adresses des clients.

TYPE	Type	Personnes	Télévision	LaveVaisselle
	hyacinthe	6	1	0
	rose	6	1	1
	dahlia	6	0	1
	tulipe	2	1	1
	tournesol	4	0	0

VILLAGES	Village	Type	Nombre
	Borinage	hyacinthe	12
	Borinage	tulipe	15
	Borinage	tournesol	35
	Ardennes	rose	27
	Ardennes	tulipe	5
	Ardennes	dahlia	5
	Côte	tournesol	17
	Côte	rose	23

LOCATIONS	Client	Village	Nombre	Type	Semaine
	111	Borinage	3	tulipe	1 août 2011
	111	Borinage	3	tulipe	8 août 2011
	111	Borinage	2	tournesol	1 août 2011
	111	Ardennes	1	tulipe	15 août 2011
	222	Ardennes	2	rose	15 août 2011
	333	Borinage	1	hyacinthe	1 août 2011
	444	Borinage	1	hyacinthe	22 août 2011

CLIENTS	Client	Nom	Rue	CodePostal	Ville	Pays
	111	M. Ed Green	5 Av. de Binche	7000	Mons	Belgique
	222	Mme Anne Dupont	10 Champ de Mars	7000	Mons	Belgique
	333	Mme Alice Dufour	20 Av. d'Anvers	1000	Bruxelles	Belgique
	444	M. Jean Pinto	20 Av. de Charleroi	7000	Mons	Belgique

Pour ces quatre tables, donnez toutes les contraintes de type PRIMARY KEY, FOREIGN KEY et UNIQUE.

.../5

TYPE PK (Type)

VILLAGES PK (Village, Type)
FK (Type) REFS TYPE

LOCATIONS PK (Client, Type, Semaine)
FK (Client) REFS CLIENTS
FK (Village, Type) REFS VILLAGES

CLIENTS PK (Client)

Notez que LOCATIONS n'est pas en BCNF.

Question 12 Écrivez une requête en calcul relationnel pour répondre à la question suivante :

Donnez les villages où chaque type de bungalow est présent.

Pour l'exemple, la réponse est vide : Borinage n'a pas de bungalow de type *rose*, Ardennes n'a pas de bungalow de type *hyacinthe*, Côte n'a pas de bungalow de type *hyacinthe*.

.../5

$$\{v \mid (\exists t \exists m \text{ VILLAGES}(v, t, m))$$

$$\wedge (\forall y \forall u_1, \forall u_2, \forall u_3$$

$$\text{TYPE}(y, u_1, u_2, u_3) \rightarrow \exists t' \exists m' \text{ VILLAGES}(v, y, m'))\}$$

Question 13 Écrivez une requête en algèbre relationnelle pour répondre à la question suivante :

Donnez le nom de chaque personne qui n'a jamais loué un bungalow de type *hyacinthe*.

Pour l'exemple, la réponse contient *M. Ed Green* et *Mme Anne Dupont*.

.../5

Soit $R := \pi_{\text{client}} \left(\sigma_{\text{Type}="hyacinthe"} \text{LOCATIONS} \right)$

R donne les clients
ayant loué "une hyacinthe"

Soit $S := \left(\pi_{\text{client}} \text{CLIENTS} \right) - R,$
les autres clients.

La requête demandée :

$$\pi_{\text{Nom}} \left(\text{CLIENTS} \bowtie S \right)$$

Question 14 Écrivez une seule requête en SQL pour répondre à la question suivante :

Donnez le village avec le plus grand nombre de bungalows de type *tulipe*.

Pour l'exemple, c'est le village *Borinage* avec 15 bungalows de type *tulipe*.

.../5

```
SELECT v. Village
FROM VILLAGES AS v
WHERE v. Type = "tulipe"
AND NOT EXISTS
    ( SELECT *
      FROM VILLAGES AS w
      WHERE w. Type = "tulipe"
        AND w. Nombre > v. Nombre )
```

