Bases de Données I (J. Wijsen)

6 janvier 2020

Durée : exactement 150 minutes

Question 1 Supposons que l'on ajoute l'opération $\delta_{A\to B}$ à l'algèbre SPJRUD, qui est définie comme suit.

Syntaxe : Si E est une expression algébrique, $A, B \in sorte(E)$ avec $A \neq B$, alors $\delta_{A\to B}(E)$ est une expression algébrique telle que $sorte(\delta_{A\to B}(E)) = sorte(E)$.

Sémantique : Soit $\mathcal{I}$ une instance de base de données. Alors $[\![\delta_{A\to B}(E)]\!]^{\mathcal{I}}$ contient chaque tuple t de $[\![E]\!]^{\mathcal{I}}$ pour lequel il n'existe pas de tuple s dans $[\![E]\!]^{\mathcal{I}}$ tel que $s(A) = t(A)$ et $s(B) \neq t(B)$.

Par exemple,

R	A	B	C
	0	0	0
	0	0	1
	1	0	0
	1	1	0

$\Longrightarrow$

$\delta_{A\to B}(R)$	A	B	C
	0	0	0
	0	0	1

Donnez une expression SPJRUD qui est équivalente à $\delta_{A\to B}(E)$ ou argumentez pourquoi une telle expression n'existe pas.

$\boxed{\ldots/10}$

L'expression F suivante renvoie tout tuple $t \in [\![E]\!]^{\mathcal{I}}$ pour lequel il existe $s \in [\![E]\!]^{\mathcal{I}}$ tel que $s(A) = t(A)$ et $s(B) \neq t(B)$. On utilise l'abréviation $\sigma_{A\neq B}(G) := G - \sigma_{A=B}(G)$. Soit B' un attribut tel que $B' \notin sorte(E)$ et

$$F := E \bowtie \pi_{AB}(\sigma_{B\neq B'}(\pi_{AB}(E) \bowtie \rho_{B\to B'}(\pi_{AB}(E)))). \tag{1}$$

La projection π_{AB} extérieure dans (1) peut être remplacée par π_A : il suffit de connaître toute valeur pour A qui n'est pas associée à plusieurs valeurs de B.

Notez que dans la solution (1), il n'y a pas de besoin de connaître les attributs autres que A et B.

L'expression demandée est donc :

$$E - F.$$

Une autre solution correcte est comme suit. Soit $sorte(E) \setminus \{A, B\} = \{C_1, \ldots, C_n\}$ avec $n \geq 0$, et $D_1, \ldots, D_n \notin sorte(E)$. L'expression demandée peut alors s'écrire comme suit :

$$E - \pi_{sorte(E)}(\sigma_{B\neq B'}(E \bowtie \rho_{B\to B'}(\rho_{C_1\to D_1}(\rho_{C_2\to D_2}(\ldots\rho_{C_n\to D_n}(E)))))). \tag{2}$$

Dans la solution (2), il est important de renommer tout attribut autre que A. Ce renommage peut être évité en ajoutant une projection :

$$E - \pi_{sorte(E)}(\sigma_{B\neq B'}(E \bowtie \rho_{B\to B'}(\pi_{AB}(E)))). \tag{3}$$

Observez que les solutions (2) et (3) dépendent de $sorte(E)$, en utilisant $C_1, \ldots, C_n$ dans le renommage et $sorte(E)$ comme argument de la projection. Ceci n'est pas un souci pourvu que les expressions soient bien définies pour **tout ensemble** $sorte(E)$ contenant A et B. Une solution ne peut pas se limiter au seul cas $sorte(E) = \{A, B, C\}$ qui est utilisé dans l'exemple.

Question 2 Cochez la case qui précède une expression correcte.

- ☒ L'opération $\delta_{A \to B}$ est non-monotone.
- ☐ L'opération $\delta_{A \to B}$ est monotone.

Argumentez pourquoi l'expression choisie est correcte.

$\boxed{\ldots/5}$

Soient s, t deux tuples sur $sorte(E)$ tels que $s(A) = t(A)$ et $s(B) \neq t(B)$. Soient $\mathcal{I}$ et $\mathcal{J}$ deux bases de données telles que $[\![E]\!]^{\mathcal{I}} = \{s\}$ et $[\![E]\!]^{\mathcal{J}} = \{s, t\}$.

Alors, $[\![\delta_{A \to B}(E)]\!]^{\mathcal{I}} = \{s\}$ et $[\![\delta_{A \to B}(E)]\!]^{\mathcal{J}} = \{\}$.

Donc, $[\![E]\!]^{\mathcal{I}} \subseteq [\![E]\!]^{\mathcal{J}}$, mais $[\![\delta_{A \to B}(E)]\!]^{\mathcal{I}} \not\subseteq [\![\delta_{A \to B}(E)]\!]^{\mathcal{J}}$.

Plusieurs étudiants ont observé que pour toute base de données $\mathcal{I}$, on a que $[\![\delta_{A \to B}(E)]\!]^{\mathcal{I}} \subseteq [\![E]\!]^{\mathcal{I}}$. Cette observation est correcte, mais ne prouve pas la non-monotonicité. Notez, par exemple, que $[\![\sigma_{A=B}(E)]\!]^{\mathcal{I}} \subseteq [\![E]\!]^{\mathcal{I}}$, mais pourtant $\sigma_{A=B}$ est monotone.

Notez aussi que pour chaque expression E en SPJRUD, si E est non-monotone, alors E utilise la différence. L'inverse n'est cependant pas vraie. En effet, il existe des expressions monotones en SPJRUD que l'on ne sait pas exprimer en SPJRU. L'usage de la différence n'implique donc pas la non-monotonicité.

Question 3 Pour chacune des trois requêtes suivantes, cochez la case qui précède une expression correcte. Si vous cochez la deuxième case (et seulement dans ce cas), donnez, dans le cadre qui suit, une base de données $\mathcal{I}$ et un domaine $\mathbf{dom}$ tel que $\mathbf{dom} \supseteq adom(\mathcal{I})$ et l'interprétation de la requête par rapport à $adom(\mathcal{I})$ est différente de l'interprétation par rapport à $\mathbf{dom}$. Par exemple,

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

☐ Cette requête est *domain independent*.

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

R: A	R: B
a	b

S: B
b

— Résultat par rapport à $adom(\mathcal{I})$: $\{(b, b)\}$
— Résultat par rapport à $\{a, b, c\}$: $\{(b, b), (c, b)\}$

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

☒ Cette requête est *domain independent*.

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

...../6

1. $\{x \mid T(x) \wedge \forall y \forall z(R(y, z) \rightarrow R(z, y))\}$

☒ Cette requête est *domain independent*.

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

Notez que cette requête renvoie tous les tuples de T si la relation R est symétrique, et renvoie l'ensemble vide si R n'est pas symétrique.

2. $\{x \mid T(x) \rightarrow \neg\forall y \forall z(R(y, z) \rightarrow R(z, y))\}$

☐ Cette requête est *domain independent*.

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

Une formule $T(x) \rightarrow \psi$ est vraie pour tout x qui n'appartient pas à T. Donc,

R: A	R: B
a	a

T: A

— Résultat par rapport à $adom(\mathcal{I})$: $\{a\}$
— Résultat par rapport à $\{a, b\}$: $\{a, b\}$

3. $\{x \mid T(x) \wedge \exists y(\neg R(x, y))\}$

☐ Cette requête est *domain independent*.

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

R: A	R: B
a	a

T: A
a

— Résultat par rapport à $adom(\mathcal{I})$: $\{\}$
— Résultat par rapport à $\{a, b\}$: $\{a\}$

Question 4 On utilise une table pour encoder les résultats du *Concours Eurovision de la chanson*.

Ce concours est organisé annuellement dans une ville (*Ville*) du pays d'accueil (*Hôte*). Le pays d'accueil est le pays gagnant de l'année précédente.

Selon une procédure de vote, les pays sont classés (*Position*) comme 1, 2, 3,. . ., sans ex-æquos (c'est-à-dire, deux pays différents n'occuperont jamais la même position).

Date	*Année*	*Ville*	*Hôte*	*Pays*	*Position*
3 May	1986	Bergen	Norway	Belgium	1
13 May	2017	Kiev	Ukraine	Portugal	1
13 May	2017	Kiev	Ukraine	Belgium	4
8 May	2018	Lisbon	Portugal	Israel	1
8 May	2018	Lisbon	Portugal	Italy	5
8 May	2018	Lisbon	Portugal	Portugal	26
18 May	2019	Tel Aviv	Israel	The Netherlands	1
18 May	2019	Tel Aviv	Israel	Italy	2
18 May	2019	Tel Aviv	Israel	Israel	23

Quelles sont les dépendances fonctionnelles que l'on peut raisonnablement imposer sur ces données ?

. . ./5

Année, *Position* → *Pays*
Année, *Pays* → *Position*
Année → *Date*
Année → *Ville*
Ville → *Hôte*

Concernant la dernière DF, deux pays peuvent avoir une ville avec le même nom, par exemple, la ville de Brest existe en France et en Bélarus. Cependant, dans le contexte du *Concours Eurovision de la chanson*, il est raisonnable d'imposer *Ville* → *Hôte*. Notez, par exemple, qu'il est très invraisemblable que la France et le Bélarus choisissent tous les deux Brest comme ville d'accueil.

Question 5 Voici une base de données pour le *Concours Eurovision de la chanson*. La dernière ligne de la table *GAGNEURS* exprime que l'Israel a gagné le concours dans la ville de Lisbon en 2018. La table *CLASSEMENT* stocke les classements de toutes les années.

La première ligne de la table *CHANSONS* exprime qu'en 1986, Sandra Kim a participé au concours avec la chanson *J'aime la vie*, représentant la Belgique ; cette chanson était en langue française. Pour information, en 2018, la Belgique a été éliminée au stade des demi-finales et, par conséquent, n'a pas eu de classement.

GAGNEURS

Année	*Ville*	*Gagneur*
1986	Bergen	Belgium
1987	Brussels	Ireland
2017	Kiev	Portugal
2018	Lisbon	Israel

CLASSEMENT

Année	*Pays*	*Position*
1986	Belgium	1
1987	Ireland	1
2017	Portugal	1
2017	Belgium	4
2017	Israel	23
2018	Israel	1
2018	Portugal	26

CHANSONS

Pays	*Année*	*Chanson*	*Artiste*	*Langue*
Belgium	1986	J'aime la vie	Sandra Kim	French
Belgium	2017	City Lights	Blanche	English
Belgium	2018	A Matter of Time	Sennek	English
Ireland	1987	Hold me now	Johnny Logan	English
Portugal	2017	Amar pelos dois	Salvador Sobral	Portugese
Portugal	2018	O Jardim	Cláudia Pascoal	Portugese
Israel	2017	I Feel Alive	IMRI	English
Israel	2018	Toy	Netta	English
Iceland	2018	Our Choice	Ari Ólafsson	English

Donnez toutes les clés primaires, les clés étrangères et les contraintes UNIQUE en utilisant la syntaxe du cours.

.../5

```
GAGNEURS
   PRIMARY KEY(Année)
   FOREIGN KEY(Année, Gagneur) REFERENCES CLASSEMENT
CLASSEMENT
   PRIMARY KEY(Année, Pays)
   UNIQUE(Année, Position )
   FOREIGN KEY(Pays, Année) REFERENCES CHANSONS
CHANSONS
   PRIMARY KEY(Pays, Année)
   FOREIGN KEY(Année) REFERENCES GAGNEURS
```

Question 6 Pour la base de données de la question 5, écrivez une requête en algèbre relationnelle pour répondre à la question suivante.

> Les pays peuvent chanter dans une langue différente d'une année à l'autre. Quels pays ont toujours chanté dans la même langue ?

Pour la base de données de la question 5, le résultat est :

Pays
Ireland
Israel
Portugal
Iceland

La Belgique n'est pas dans la réponse car ce pays a déjà chanté en français et en anglais.

…/5

$\sigma_{A \neq B}(E)$ est une abréviation pour $E - \sigma_{A=B}(E)$. Soit

$$R \quad := \quad \pi_{\mathit{Pays,Langue}}(\mathit{CHANSONS}).$$

La requête demandée est :

$$\pi_{\mathit{Pays}}(\mathit{CHANSONS}) - \pi_{\mathit{Pays}}(\sigma_{\mathit{Langue} \neq \mathit{Taal}}(R \bowtie \rho_{\mathit{Langue} \to \mathit{Taal}}(R))).$$

Question 7 Pour la base de données de la question 5, écrivez une requête en calcul relationnel pour répondre à la question suivante.

> Pour une année donnée, un pays peut ne pas apparaître dans la table *CLASSEMENT* pour causes de non-participation ou élimination au stade des demi-finales. Quels pays ont obtenu un classement à chacune de leurs participations ?

Pour la base de données de la question 5, le résultat est :

Pays
Ireland
Israel
Portugal

Noter, par exemple, que l'Islande n'est pas dans la réponse : ce pays a participé en 2018 mais n'a pas obtenu de classement.

…/5

$$\left\{ p \;\middle|\; \begin{array}{c} \exists y \exists u \exists v \exists w (\mathit{CHANSONS}(p, y, u, v, w)) \\ \wedge \\ \forall y \forall u \forall v \forall w (\mathit{CHANSONS}(p, y, u, v, w) \to \exists i (\mathit{CLASSEMENT}(y, p, i))) \end{array} \right\}$$

Question 8 Soient

$$\begin{aligned} \mathcal{A} &= ABCDEF \\ \Sigma &= \{AB \to F, AF \to B, AD \to C, CE \to A, ACD \to BEF\} \end{aligned}$$

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

☐ Le schéma $(\mathcal{A}, \Sigma)$ est en 3NF.

☒ Le schéma $(\mathcal{A}, \Sigma)$ n'est pas en 3NF.

Détaillez les arguments qui mènent à cette conclusion.

…/5

Σ est équivalent à $\Sigma' = \{AB \to F, AF \to B, AD \to C, CE \to A, ACD \to B, ACD \to E, ACD \to F\}$, un ensemble de DFs singulières.

Puisque D n'apparaît pas à droite d'une flèche, D fera partie de chaque clé. Deux clés sont AD et CDE.

Il est facile de vérifier l'affirmation suivante :

pour tout ensemble $U \subseteq ABCDEF$, si $A \notin U$ et $CE \nsubseteq U$, alors $U^{*,\Sigma} = U \neq ABCDEF$.

Par contraposition :

pour tout ensemble $U \subseteq ABCDEF$, si $U^{*,\Sigma} = ABCDEF$, alors $A \in U$ ou $CE \subseteq U$.

En conséquence, chaque clé doit inclure $\{A\}$ ou $\{C, E\}$. On peut donc conclure qu'il n'y pas une troisième clé.

Le schéma n'est pas en 3NF, parce que
- Σ' contient $AB \to F$, et
  - AB n'inclut aucune clé, et
  - F ne fait partie d'aucune clé.

Question 9 Considérez l'exécution suivante :

$$R_1(A)R_2(B)R_3(B)W_2(B)W_1(C)R_2(C)W_3(A)$$

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

$R_1(A)$		
	$R_2(B)$	
		$R_3(B)$
	$W_2(B)$	
$W_1(C)$		
	$R_2(C)$	
		$W_3(A)$