Bases de Données I, Mons, 17 janvier 2022
Partie « avant la pause »

Cette partie contient les questions 1 à 6.

NOM + PRÉNOM :	
Orientation + Année :	

Question 1 On utilise une table pour encoder les résultats des courses de cyclisme. Une course est soit une course d'une seule journée (par exemple, Milan-San Remo), soit un tour avec plusieurs étapes (par exemple, le Tour de France).

La deuxième ligne de la table ci-dessous exprime qu'Oliver Naesen est arrivé à la dixième place dans l'onzième étape du Tour de France 2016, avec un délai de 6 secondes. Cette étape avait lieu le 13 juillet. Le gagneur a toujours un classement égal à 1 et un délai égal à 0.

NA indique que le numéro d'étape est "non applicable" pour les courses d'une seule journée. Supposez que chaque coureur est identifié par un nom unique. Un coureur peut changer d'équipe, mais pas au cours d'une même course. Deux courses peuvent avoir lieu à une même date. Cependant, pour un tour avec plusieurs étapes, il n'y a jamais deux étapes sur une même journée. Un coureur ne peut pas participer à deux courses sur une même journée. Dans chaque course, les coureurs sont classés 1, 2, 3, ... sans ex æquo (i.e., deux coureurs ne peuvent pas avoir la même valeur pour l'attribut *Classement*). Cependant, deux coureurs peuvent avoir le même délai, par exemple, dans un sprint.

Coureur	*Équipe*	*Course*	*Étape*	*Jour*	*Mois*	*Année*	*Classement*	*Délai*
Arnaud Démare	FDJ	Milan-San Remo	NA	19	mars	2016	1	0
Oliver Naesen	IAM Cycling	Tour de France	11	13	juillet	2016	10	6
Julian Alaphilippe	Quick Step	Tour de France	3	8	juillet	2019	1	0
Michael Matthews	Sunweb	Tour de France	3	8	juillet	2019	2	26
Jasper Stuyven	Trek	Tour de France	3	8	juillet	2019	3	26
Julian Alaphilippe	Quick Step	Milan-San Remo	NA	23	mars	2019	1	0
Oliver Naesen	AG2R	Milan-San Remo	NA	23	mars	2019	2	0

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

.../10

Question 2 Dans cette question, on s'intéresse seulement au Tour de France. La table *CLASSEMENTS* stocke les classements des étapes. L'attribut *Temps* est le temps parcouru dans le format $hh : mm : ss$, c'est-à-dire, en heures, minutes, secondes. La table *COUREURS* liste l'affiliation des coureurs. La table *ÉTAPES* liste, pour chaque étape, le numéro de l'étape, les villes de départ et d'arrivée, ainsi que la distance de l'étape. Par exemple, l'onzième étape du Tour de France 2016 parcourait 162 kilomètres entre Carcassone et Montpellier.

CLASSEMENTS

Édition	*Étape*	*Position*	*Coureur*	*Temps*
1970	10	1	Eddy Merckx	05 : 45 : 21
2016	4	1	Julian Alaphilippe	06 : 35 : 04
2016	4	53	Oliver Naesen	06 : 39 : 14
2016	11	10	Oliver Naesen	03 : 59 : 32
2019	3	1	Julian Alaphilippe	05 : 01 : 45
2019	3	2	Michael Matthews	05 : 01 : 45
2019	5	1	Oliver Naesen	02 : 10 : 07
⋮	⋮	⋮	⋮	⋮

COUREURS

Nom	*Année*	*Équipe*
Eddy Merckx	1970	Molteni
Oliver Naesen	2016	IAM Cycling
Julian Alaphilippe	2016	Quick Step
Julian Alaphilippe	2019	Quick Step
Oliver Naesen	2019	AG2R
Michael Matthews	2019	Sunweb
⋮	⋮	⋮

ÉTAPES

Année	*Numéro*	*Départ*	*Arrivée*	*Distance*
1970	10	Limoges	Saint-Flour	230
2016	4	Saumur	Limoges	237
2016	11	Carcassone	Monpellier	162
2019	3	Binche	Épernay	215
2019	4	Reims	Nancy	213
2019	5	Nancy	Metz	105
⋮	⋮	⋮	⋮	⋮

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

…/10

Question 3 Pour la base de données de la question 2, écrivez une requête en algèbre relationnelle pour répondre à la question :

> Quelle est l'étape la plus courte ? Supposez qu'il n'y ait pas deux étapes avec la même distance.

Pour la base de données de la question 2, le résultat est :

Année	*Numéro*
2019	5

En effet, la cinquième étape du Tour de France 2019 avait une distance de 105 kilomètres ; toutes les autres étapes sont plus longues.

Pour cette question, on peut utiliser la sélection $\sigma_{A<B}(E)$ avec $A, B \in sorte(E)$. La sémantique est :

$$[\![\sigma_{A<B}(E)]\!]^{\mathcal{I}} = \{t \in [\![E]\!]^{\mathcal{I}} \mid t(A) < t(B)\}.$$

$\boxed{\ldots/10}$

Question 4 Pour la base de données de la question 2, écrivez une requête en calcul relationnel qui rend le nom de chaque cycliste qui a gagné au moins une étape dans chaque Tour de France auquel il a participé. Pour la base de données de la question 2, le résultat est comme suit (si l'on ignore les rangées manquantes indiquées par " $\vdots$ ") :

Eddy Merckx
Julian Alaphilippe

Oliver Naesen n'est pas dans le résultat, car il a participé au tour de 2016, mais n'apparaît pas en 1ère position d'une étape en 2016. Noter qu'Alaphilippe est dans la réponse, même s'il n'a pas gagné d'étape dans le tour de 1970 auquel il n'a pas participé.

$\boxed{\ldots/10}$

Question 5 Supposons que l'on ajoute un nouvel opérateur δ_X^Y à l'algèbre SPJRUD, où X et Y sont des ensembles non-vides d'attributs. Cet opérateur prend en entrée une relation sur $X \cup Y$ et renvoie, comme résultat, une relation sur X. Si la relation d'entrée satisfait la dépendance de jointure $\bowtie[X, Y]$, alors le résultat est la projection, sur X, de la relation d'entrée. Par contre, si la relation d'entrée ne satisfait pas la dépendance de jointure $\bowtie[X, Y]$, alors le résultat est la relation vide. Par exemple, soit R la relation suivante :

R	B	C	D	E
	0	0	0	0
	1	1	1	1

On peut vérifier que R satisfait $\bowtie[BC, CDE]$. En conséquence, on obtient :

$\delta_{BC}^{CDE}(R)$	B	C
	0	0
	1	1

, c'est-à-dire, la projection de R sur BC.

Par contre, puisque R ne satisfait pas $\bowtie[BC, DE]$, on obtient :

$\delta_{BC}^{DE}(R)$	B	C

, c'est-à-dire, la relation vide.

De façon formelle, on peut définir δ_X^Y comme suit :

Syntaxe : Si E est une expression algébrique et X, Y sont des sous-ensembles non-vides de $sorte(E)$ tels que $X \cup Y = sorte(E)$, alors $\delta_X^Y(E)$ est une expression algébrique telle que $sorte(\delta_X^Y(E)) = X$.

Sémantique : Soit $\mathcal{I}$ une instance de base de données. Si $[\![E]\!]^{\mathcal{I}}$ satisfait la dépendance de jointure $\bowtie[X, Y]$, alors $[\![\delta_X^Y(E)]\!]^{\mathcal{I}} := [\![\pi_X(E)]\!]^{\mathcal{I}}$. Si $[\![E]\!]^{\mathcal{I}}$ ne satisfait pas la dépendance de jointure $\bowtie[X, Y]$, alors $[\![\delta_X^Y(E)]\!]^{\mathcal{I}}$ est la relation vide.

Démontrez que l'opérateur δ_X^Y est non-monotone.

…/5

Note : L'opérateur est non-monotone si (et seulement si) $X \nsubseteq Y$ et $Y \nsubseteq X$. L'opérateur est monotone dans le cas particulier où $X \subseteq Y$ ou $Y \subseteq X$. Effectivement, une dépendance de jointure $\bowtie[X, Y]$ avec $X \subseteq Y$ ou $Y \subseteq X$ est satisfaite par toute relation sur $X \cup Y$.

Question 6 On dénote par SPJRUD$^\delta$ l'algèbre que l'on obtient si l'on ajoute le nouvel opérateur de la question 5 à l'algèbre SPJRUD. Cochez la case qui précède une expression correcte :

☒ Pour chaque expression E en SPJRUD$^\delta$, il existe une expression E' en SPJRUD telle que $E' \equiv E$. Le nouvel opérateur est donc redondant.

☐ L'algèbre SPJRUD$^\delta$ est plus expressive que SPJRUD.

Argumentez votre choix. Si vous avez cochez la première case, expliquez la construction de E' à partir de E.

…/5

Réponse succincte : Soit A un attribut de X, et B un attribut tel que $B \notin XY$. Définissons E' comme suit :

$$E' := E - \pi_{XY}(E \bowtie \rho_{A \to B}(\pi_A((\pi_X(E) \bowtie \pi_Y(E)) - E))).$$

On peut vérifier que pour toute instance de base de données, les deux expressions suivantes sont équivalentes :

(1) la sous-expression $\rho_{A \to B}(\pi_A((\pi_X(E) \bowtie \pi_Y(E)) - E))$ renvoie une relation non-vide sur le schéma $\{B\}$;

(2) E renvoie une relation sur XY qui ne respecte pas $\bowtie[X, Y]$.

Puis, on peut vérifier :

— si (1) est vrai, alors E' renvoie une relation vide ; et

— si (1) est faux, alors E' renvoie la même relation que E.

Donc, E' est équivalent à $\delta_X^Y(E)$.

Par ailleurs, si on utilise une projection sur l'ensemble vide d'attributs (ce qui n'est pas interdit par les définitions vues au cours), on peut définir E' de façon plus simple :

$$E' := E - (E \bowtie \pi_{\{\}}((\pi_X(E) \bowtie \pi_Y(E)) - E)).$$

Noter que pour une relation R,

$$\pi_{\{\}}(R) = \begin{cases} \text{la relation vide } \{\} & \text{si } R \text{ est vide} \\ \text{la relation } \{\{\}\}\text{, qui contient le tuple vide,} & \text{si } R \text{ est non-vide} \end{cases}$$

Erratum du 23 janvier 2024 : L'expression demandée n'est pas E', mais plutôt $\pi_X(E')$. Effectivement, dans l'énoncé, il est demandé de renvoyer la projection sur X.

Bases de Données I, Mons, 17 janvier 2022
Partie « après la pause »

Cette partie contient les questions 7 à 10.

NOM + PRÉNOM :	
Orientation + Année :	

Question 7 Considérez le schéma-DF suivant (dénotons ce schéma-DF par δ) :

$$(ABCDEF, \{A \to B, A \to C, B \to C, E \to F\})$$

Considérez l'ensemble suivant avec deux schémas-DF (dénotons cet ensemble par Δ) :

$$\Big\{(ABCD, \{A \to B, B \to C\}),\ (DEF, \{E \to F\})\Big\}$$

Cochez la case qui précède une expression correcte :

☐ Δ n'est pas une décomposition de δ parce que Δ ne préserve pas le contenu.

☐ Δ est une décomposition de δ qui préserve les dépendances fonctionnelles.

☐ Δ est une décomposition de δ qui ne préserve pas les dépendances fonctionnelles.

Argumentez votre choix.

…/6

Question 8 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 $\mathbf{dom}$. Pour rappel, $adom(\mathcal{I})$ est l'ensemble de toutes les constantes qui apparaissent dans $\mathcal{I}$. Par exemple,

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

☐ Cette requête est *domain independent*.

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

R:

A	B
a	b

S:

B
b

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

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

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

☒ Cette requête est *domain independent*.

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

.../9

1. $\Big\{y \mid T(y) \lor \forall x\Big(R(x, y) \to \exists z\big(R(y, z)\big)\Big)\Big\}$

☐ Cette requête est *domain independent*.

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

2. $\Big\{x, y \mid \exists z\Big(S(x, y, z) \lor \exists u\big(S(u, x, z)\big)\Big)\Big\}$

☐ Cette requête est *domain independent*.

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

3. $\big\{x \mid \exists y\big(R(y, x) \lor \neg R(x, y)\big)\big\}$

☐ Cette requête est *domain independent*.

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

Question 9 Soient

$$\begin{aligned}\mathcal{A} &= ABCDEFG \\ \Sigma &= \{DE \to FG, DG \to BE, BF \to D, ABC \to G, DE \to AB, BEF \to C\}\end{aligned}$$

Cochez la case qui précède une expression correcte :

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

Détaillez les arguments qui mènent à cette conclusion.

$\boxed{\ldots/15}$

Question 10 Considérez l'exécution suivante :

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

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.

…/10

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