

INFO 009-1/2 : Bases de données

Examen écrit - Juin 2017

*Livres fermés. Durée: 3 heures 1/2.
Répondez à chaque question sur une feuille séparée
sur laquelle figure votre nom, prénom et section.*

Les étudiants en géographie ne doivent pas répondre aux questions 2b, 4b et 5.

1. *Modélisation.*

Une base de données utilisée pour la gestion d'une conférence est organisée comme décrit ci-après. On distingue deux types de participants à la conférence: les chercheurs et les bénévoles (ces deux catégories étant mutuellement exclusives). Pour ces deux types de participants, leur numéro d'inscription sert d'identifiant unique. Ils sont décrits par leur nom, leur prénom, leur université, et leur pays d'origine. Pour les participants chercheurs on conserve également un booléen permettant de savoir s'ils ont payé leur droit d'inscription. Pour les participants bénévoles, leur numéro de gsm est repris dans la base de données.

Les présentations données à la conférence sont regroupées au sein de "sessions": une session est définie par un identifiant, et est caractérisée par un sujet. Chaque session sera supervisée par un participant chercheur.

L'horaire de la conférences est discrétisé en créneaux horaires, que l'on représente donc par un entier. Une présentation est caractérisée par un titre et un résumé, son créneau horaire et la session à laquelle elle appartient. Une présentation est toujours donnée par un seul participant chercheur, et se déroule dans une salle définie par sa localisation, composée d'un numéro de bâtiment, d'un numéro d'étage, et d'un numéro de local. Pour chaque salle, on conserve également son nombre de sièges, ainsi que l'ensemble du matériel fourni au sein de celle-ci (pointeurs, projecteurs, tableaux, ...) chaque élément étant décrit par une ligne de texte.

Il ne peut y avoir qu'une seule présentation par session pour un créneau horaire donné. Par contre, plusieurs sessions peuvent se dérouler en parallèle.

Afin de pouvoir faciliter l'organisation d'un horaire pour les bénévoles, on encode dans la base de données l'ensemble des présentations auxquelles ils voudraient assister.

- (a) Dessinez un diagramme entités-relations conforme à la description ci-dessus. Précisez les clés des ensemble d'entités et des relations, ainsi que les contraintes d'intégrité éventuellement non représentées dans la diagramme.
- (b) Effectuez la conversion de ce modèle vers le modèle relationnel.

2. *Théorie des dépendances.*

- (a) Soit $R(A, B, C, D, E)$ un schéma de relation et $F = \{A \rightarrow BC, BE \rightarrow D\}$ un ensemble de dépendances fonctionnelles associées à R .
 - (i) Donnez la (les) clé(s) de R
 - (ii) Ce schéma est-il en BCNF? Justifiez.
 - (iii) Donnez l'algorithme de décomposition en BCNF, et appliquez le si nécessaire à R . Cette décomposition est-elle sans pertes? Conserve-t-elle les dépendances? Justifiez.
- (b) Soit $R(A, B, C, D, E)$ un schéma de relation. Définissez son ensemble de dépendances F par rapport auquel R est en BCNF, mais pas en 4FN.

3. Langages d'interrogation.

Une organisation de courses à pied dispose d'une base de données dont le schéma est le suivant :

- circuit(identifiant, lieu_de_départ, nom, distance, dénivelé_positif);
- course(identifiant_circuit, année, jour, mois);
- coureur(n_dossard, nom, prénom);
- performances(identifiant_participant, identifiant_circuit, année, temps);
- ravitaillement(identifiant_circuit, année, position_km, type).

Note : la distance est donnée en kilomètres, le dénivelé en mètres, le temps en secondes. L'optimisation des requêtes n'est pas prise en compte.

Exprimez les requêtes suivantes en algèbre relationnelle étendue et en SQL.

- Donnez les noms et prénoms des coureurs ayant déjà effectué une course de plus de 20km avec un dénivelé positif supérieur à 500m.
- Donnez le nom et le prénom des coureurs qui ont participé à toutes les courses de 2017 où au moins un ravitaillement en nourriture était disponible.
- Effectuez un classement (numéro de dossard, nom et prénom) pour les coureurs ayant couru en 2017 : un participant sera classé s'il a effectué au moins une course au cours de l'année. Sa position sera déterminée par la moyenne de ses temps moyens au kilomètre sur l'ensemble de ses courses de 2017. Si des coureurs sont à égalité, le classement les affichera par ordre de dossard.

4. Implémentation du modèle relationnel

- Quelle est la différence entre un produit cartésien et un joint? Décrivez deux algorithmes qui permettent de calculer un joint avec une complexité inférieure à celle du calcul d'un produit cartésien. Sous quelle hypothèse cette complexité est-elle effectivement inférieure?
- Citez les noms des 4 niveaux d'isolation possibles pour une transaction, et expliquez brièvement ce que chacun d'eux garantit.

5. Bases de données déductives

Soient les prédicats suivants : bus(B) qui est vrai uniquement lorsque B est un bus, arret_bus(X) qui est vrai uniquement lorsque X est un arrêt de bus, et prochain_arret_bus(B,X,Y), qui est vrai uniquement lorsque Y est l'arrêt suivant l'arrêt X sur la ligne du bus B.

- Écrivez le prédicat arret_accessible_avec_bus(B,X,Y), qui est vrai uniquement lorsque Y est un arrêt pouvant être atteint avec le bus B depuis l'arrêt de bus X.
- Écrivez le prédicat inaccessible_2bus(X,Y), qui est vrai uniquement lorsque l'arrêt de bus Y ne peut être atteint depuis l'arrêt de bus X en utilisant deux bus ou moins.
- Décrivez brièvement la méthode qui permet de calculer l'extension d'un prédicat intentionnel.

6. Entrepôts de données XML

- Donnez un exemple de contexte dans lequel il serait préférable de stocker des données dans le format XML, plutôt que dans une base de données SQL. Justifiez.
- Donnez un DTD définissant un élément "reponse_examen" comprenant un nom d'étudiant, un prénom, un matricule, une section, ainsi que la réponse à trois questions. La seconde question est divisée en deux sous-questions a) et b).

7. Travaux pratiques

- Décrivez brièvement le sujet du travail pratique réalisé.
- Expliquez brièvement le lien entre PHP et HTML.