

Rio ne répond plus

Projet base de données L2

Le projet s'effectue par groupe de 2 personnes. Vous pouvez contacter vos encadrants aux adresses :

Jules Chouquet	:	<code>chouquet@irif.fr</code>
Gianluca Curzi	:	<code>gianluca.curzi@unito.it</code>
Sidi Mohamed Beillahi	:	<code>beillahi@irif.fr</code>
Valia Mitsou	:	<code>vmitsou@liris.cnrs.fr</code>
Cristina Sirangelo	:	<code>cristina@irif.fr</code>

Les règles anti-plagiat de l'Université s'appliquent à votre travail et toute utilisation d'éléments extérieurs devra être explicitement évoquée et sourcée. Vous pouvez discuter des différents aspects du projet avec les autres groupes, mais toute copie partielle ou totale de code entre plusieurs groupes est interdite.

Date rendu	:	Dimanche 16 décembre à 23h59 (sous réserve de modification : la date pouvant être avancée en fonction de celle des soutenances)
Date soutenance	:	annoncé bientôt en amphi
Format	:	Rapport entre 4 et 8 pages au format PDF à déposer sur Moodle (annexes possibles, mais considérées à la discrétion de l'examineur). Vous pouvez utiliser https://www.freepdfconvert.com/fr pour convertir votre rapport en PDF.
Consignes	:	Vous incluez votre modèle entités-associations et un schéma relationnel de vos tables. Vous expliquerez les choix qui ont motivé votre modélisation, et les limitations de votre modèle. Vous discuterez de la qualité de votre schéma en termes de résistance aux anomalies et adéquation aux formes normales vues en cours. Vous décrierez également les requêtes que vous avez choisies d'implémenter, et les choix que vous avez faits pour le calcul de l'indice prédictif de remplissage (voir Section 4). Vous débutez votre soutenance par une présentation de votre travail (10 minutes maximum : reprise des points marquants du rapport et petite démo). Nous vous poserons ensuite des questions sur vos choix de modélisation, les contraintes d'intégrité utilisées, les points forts et les points faibles de votre approche, le fonctionnement de vos requêtes, etc.

1 Scenario

Vous êtes journaliste à *Lepique.fr*, un site sérieux d'informations sportives. Suite à une avarie généralisée des serveurs de la boîte, un grand nombre de données ont été perdues à jamais, notamment des données concernant des événements terminés depuis plusieurs années. On vous charge *Manu Militari* de mettre au point le plus rapidement possible une base de données concernant les jeux olympiques de l'été 2016, à Rio, pour lesquels on enregistre encore beaucoup de requêtes sur le site.

Un certain nombre de difficultés s'offrent à vous, elles se constatent à travers notamment la variété des requêtes des utilisateurs. Il faut être capable de distinguer les sports d'équipes et les sports individuels, par exemple : si tous les sportifs d'une équipe remportent une médaille lors d'une victoire, cela ne fait qu'une médaille supplémentaire pour le pays représenté. Et on veut bien sûr être capable de donner des informations sur les médailles remportées par des sportifs, ou remportées sous le drapeau d'une certaine nationalité.

Certains sports présentent des modèles de qualification différents, une finale de course à pied comporte plus de participants opposés qu'une finale de handball. Il y a donc plusieurs types de sports, revêtant

diverses caractéristiques. On peut vouloir connaître les temps réalisés en course ou en natation, ou bien les scores en sport d'équipe...

Un même athlète peut également participer à plusieurs compétitions, c'est souvent le cas en natation par exemple : Michael Phelps aura pu remporter plusieurs médailles, dans la même catégorie de sport, mais pour des épreuves différentes, il faut pouvoir en tenir compte. La base doit être structurée de façon assez fine pour que l'on puisse récupérer des informations sur les athlètes, les résultats des épreuves, les statistiques par pays et par sport, par catégorie de sport (équipe, course chronométrée), en tenant compte de toutes les nuances possibles.

Il n'est bien sûr pas demandé d'être exhaustif dans la création et le remplissage de la base, tous les sports, tous les athlètes, ou tous les pays ne devront pas être représentés. Mais elle devra contenir une variété et quantité suffisante de données pour que l'on soit sûr de savoir comment s'y prendre aux JO d'été de Paris en 2024 par exemple.

Les requêtes proposées devront être implémentables dans votre base, mais naturellement, vous devrez montrer que votre façon de la construire permet bien, à terme, de considérer toutes les requêtes de ce type, pour peu que l'on enrichisse la table en suivant votre modèle.

En plus de vos cours et TPs, vous êtes encouragés à consulter la documentation en français de PostgreSQL disponible ici :

<https://docs.postgresql.fr/10/>

Les sections I. Tutoriel (<https://docs.postgresql.fr/10/tutorial.html>) et II. Langage SQL (<https://docs.postgresql.fr/10/sql.html>) sont les plus intéressantes. Il y est expliqué comment compter, calculer un maximum, ajouter des contraintes lors de la création d'une table (empêcher un attribut d'être négatif par exemple), faire des opérations mathématiques, etc.

Vous pouvez également consulter les fichiers **Installation et commandes basiques pour PostgreSQL** et **Quelques fonctionnalités PostgreSQL** disponibles sur Moodle :

<https://moodlesupd.script.univ-paris-diderot.fr/course/view.php?id=3673>

Vous y trouverez notamment de l'aide sur la création de requêtes interactives (ce qui peut être utile pour la présentation orale) et la manipulation des dates.

2 Peuplez vos tables

Maintenant que vos tables sont créées, vous pouvez les remplir avec les informations adaptées (athlètes, pays, sports, épreuves,...). Insérez au moins une centaine de tuples au total dans votre base de données. Vous pouvez par exemple écrire dans votre langage de programmation préféré (Java, Python, ...) une boucle for qui affiche des tuples générés automatiquement (avec la syntaxe attendue par PostgreSQL), et les recopier ensuite dans votre code PostgreSQL.

Vous pouvez aussi utiliser directement les structures de contrôle intégrées à PostgreSQL :

<https://docs.postgresql.fr/8.3/plpgsql-control-structures.html>

Vous pouvez bien sûr vous renseigner chez la concurrence en récupérant des données, par exemple aux adresses suivantes :

lequipe.fr/Jo/JoPodium40_1_H.html

lequipe.fr/jeux-olympiques/tableau-classement-medailles-jeux-olympiques-rio-2016-E.html

lequipe.fr/jeux-olympiques/jo-rio-2016-programme-calendrier-par-sport-date.html

3 Effectuez des requêtes

La base étant prête, il va maintenant s'agir de vérifier que les demandes des utilisateurs concernant les JO d'été 2016 ne resteront plus lettre morte. Pour cela, commencez par implémenter *obligatoirement*

les 12 requêtes suivantes (vous pouvez utiliser au besoin des opérateurs trouvés dans la documentation en ligne de PostgreSQL) :

Difficulté ★

1. La liste des athlètes italiens ayant obtenu une médaille
2. Le nom et la nationalité des médaillés du 100m, 200m, et 400m avec à chaque fois le type de médaille (or, argent, bronze)
3. Les membres de l'équipe féminine de handball de moins de 25 ans
4. Les médailles gagnées par Michael Phelps, avec l'épreuve et le temps correspondants
5. La liste des sports pratiqués en équipe
6. Le meilleur temps réalisé au marathon

Difficulté ★★

1. La moyenne des temps réalisés au 200 mètres nage libre par nationalité
2. Le nombre de médailles par pays représentés (rappel : une seule médaille est comptée pour une équipe)
3. Pour chaque épreuve, le nom et la nationalité de l'athlète ayant obtenu la médaille d'or, ainsi que le nom et la nationalité de celui ayant obtenu la médaille d'argent (tableau résultat avec 5 attributs)
4. Les athlètes qui n'ont obtenu aucune médaille d'or
5. Les sports individuels dans lesquels la France n'a pas obtenu de médaille
6. Les coureurs qui n'ont jamais mis plus de dix secondes au 100m

Implémentez ensuite au minimum 4 requêtes au choix parmi les 6 suivantes :

Difficulté ★★★

1. L'athlète qui a concouru sept jours consécutifs et a gagné à chaque fois
2. Les pays qui ont eu une médaille dans chaque catégorie sportive (pas forcément à toutes les épreuves de cette catégorie)
3. Les cinq catégories sportives pour lesquelles il y a le moins d'épreuves
4. Le pourcentage de médailles remportées par des femmes (y compris en équipe)
5. Le nombre total de points marqués par l'équipe féminine de handball qui a marqué plus de points que chaque équipe masculine de handball tout au long des jeux
6. Les pays qui ont obtenu plus de médailles que la France dans chaque sport

Enfin, créez et implémentez 3 requêtes supplémentaires de *votre choix* qui vous semblent pertinentes et sont suffisamment complexes. Utilisez des jointures entre plusieurs tables, des requêtes imbriquées, des conditions **EXISTS** ou **IN**, des opérateurs **GROUP BY**, **HAVING**, etc.

Expliquez dans le rapport (Rendu 2bis) quelles sont les difficultés que vous avez pu rencontrer lors de l'implémentation de toutes vos requêtes et comment vous les avez surmontées. Détaillez également les 3 requêtes que vous avez choisies précédemment. Enfin, trouvez *une requête* qui vous semble difficile (ou impossible) à effectuer dans votre modèle, et suggérez une modification de votre modélisation initiale qui faciliterait l'implémentation de cette requête.

4 Organisation et prévisions

L'équipe logistique de Paris 2024 doit planifier les horaires de ses volontaires pendant les jeux. Cependant, ils veulent envoyer des volontaires dans des endroits où ils sont nécessaires car généralement, les volontaires ne sont pas en nombre suffisant pour gérer des milliers d'athlètes avec du personnel. À Rio 2016, les événements de la même catégorie ont lieu au même endroit. Par conséquent, les organisateurs de

Paris 2024 souhaitent connaître pour chaque jour et catégorie, combien d'événements de cette catégorie ont eu lieu ce jour-là et sur le nombre de participants à chaque événement.

Modifiez votre base de données pour proposer à votre tour un indice prédictif du nombre de volontaires nécessaire pour chaque endroit. Vous calculerez cet indice selon la méthode et les critères de votre choix en vous basant sur combien d'événements de même catégorie ont eu lieu au même jour et le nombre de participants à chaque événement. Utilisez-le ensuite pour concevoir des requêtes qui aideront à réguler la répartition des volontaires. Par exemple, proposez une requête pour récupérer les 10 endroits avec l'indice le plus petit et les 10 avec l'indice le plus grand. Expliquez dans le rapport (Rendu 2bis) le fonctionnement de votre indice et les raisons qui vous ont guidées dans votre choix. Détaillez ses points forts, ses limites et comment l'améliorer.