



Bases de données avancées

Projet de modélisation

Modalités

Le projet est à réaliser en binômes. Vous devez rendre via Moodle une archive zip contenant quatre fichiers dont les noms sont spécifiés comme suit (XXX et YYY étant les noms des deux membres du binôme).

1. XXX_YYY_modelisation.pdf pour la modélisation
2. XXX_YYY_tables.sql pour la création des tables
3. XXX_YYY_anomalies.sql pour la détection des anomalies dans les données (s'il y a lieu)
4. XXX_YYY_donnees.sql pour la consultation des données avec des requêtes paramétrées simples

I) Présentation générale

L'objectif du projet est la modélisation, le peuplement, et la mise en place d'une base de données dont les données sont issues du fichier Open Data de la Ville de Paris « Que faire à Paris ? » [https://opendata.paris.fr/explore/dataset/que-faire-a-paris-/](https://opendata.paris.fr/explore/dataset/que-faire-a-paris/)

a) Modélisation des données

Vous devez rendre un court rapport de trois à quatre pages qui détaille votre modélisation des données.

1. Identifier toutes les dépendances fonctionnelles présentes dans les données, et celles qui selon vous devraient être présentes. Donner un schéma des dépendances fonctionnelles identifiées.
2. Si les données ne sont pas en première forme normale¹, proposer des attributs supplémentaires et expliquer comment les données seront décomposées. (Ne pas rajouter d'attributs clefs primaires numériques.)
3. Si certains attributs sont calculés à partir d'autres données, identifiez-les et retirez-les de la modélisation. (Vous pourriez rajouter une fonction pl/pqsql qui fait le calcul dans la partie implémentation mais c'est optionnel.)
4. Proposer une décomposition et identifier les clefs candidates de toutes les relations. Votre décomposition est-elle en FNBC ?
5. Fournir un schéma entité/relation (voir le cours de BD de L2/L3 d'Amélie GHEERBRANT pour la syntaxe graphique) et un schéma relationnel (qui correspondent bien l'un à l'autre).

1. Pour être en première forme normale, tous les attributs doivent être atomiques, c'est-à-dire qu'aucun attribut ne peut être décomposé en relation.

6. identifier un maximum de contraintes (cardinalité, règles de gestion) et préciser comment chacune pourrait être gérée (par exemple `CHECK`, `NOT NULL`, `UNIQUE`, trigger, fonctions d'insertion/mise à jour sûre, ...); un récapitulatif doit être fourni sous forme de tableau.
7. Identifier les index qui devraient être rajoutés et justifier brièvement.

b) Création et indexation des tables

Après cette phase de modélisation, vous pouvez implémenter la création de vos tables SQL, en prenant soin d'intégrer un maximum de contraintes. Il n'est pas obligatoire d'implémenter des triggers dans ce projet.

c) Peuplement des tables

Vous devez prévoir un script entièrement automatique qui peuplera vos tables à partir du fichier source au format CSV. Ceci peut se faire en plusieurs étapes, en créant avec la commande `\copy` une table temporaire dans laquelle vous importerez les données publiques à partir du fichier, puis en faisant des requêtes sur cette table pour peupler les tables de votre schéma. Le script sera lancé avec la commande `\i XXX_YYY_tables.sql` où `XXX` et `YYY` sont les noms du binôme.

Faites le tout en SQL, et évitez sauf absolue nécessité d'utiliser `pl/pgsql`. Consulter la documentation sur les fonctions sur les chaînes de caractères comme `split_part` et `substring` qui admet des expressions rationnelles. Si vous trouvez des anomalies dans les données (par exemple des dépendances fonctionnelles qui ne sont pas respectées) faites un choix arbitraire parmi les valeurs présentes (par exemple prendre le minimum parmi les valeurs non-nulles et non-vides).

Écrire des requêtes pour extraire des données que vous auriez identifiées comme anormales ou incorrectes. Les requêtes seront dans un fichier `XXX_YYY_anomalies.sql`. Utiliser quelques affichages et la commande `\prompt` pour expliquer ce que fait chaque requête et pour ne pas que les résultats défilent à la suite les uns des autres.

d) Consultation des données

Préparer un script `XXX_YYY_data.sql` qui permet de consulter les données dans vos tables avec des requêtes paramétrées utilisant `PREPARE` et `\prompt`.

e) Annexe : liste des attributs présents dans le fichier

id	54380
url	https://www.paris.fr/evenements/yowl-en-concert-54380
titre	YOWL en concert au Supersonic Records !
chapeau	Les Londoniens YOWL viendront au Supersonic Records vous conter leurs récits indie menaçants, oscillant entre proto-punk intense et ballades souffrantes !
description	<p>Supersonic Records & Take Me Out présentent...</p><p>YOWL en concert !</p><p>Tickets : https://link.dice.fm/te7746778696 </p><p>YOWL</p><p>(Indie rock / Post-punk - Londres, UK)</p>
date_de_debut	2024-05-15T22:00:00+02:00
date_de_fin	2024-05-16T01:30:00+02:00
occurrences	2024-05-15T20:00:00+02:00_2024-05-15T23:30:00+02:00
description_de_la_date	Le mercredi 15 mai 2024 de 19h00 à 22h30
url_de_l_image	https://cdn.paris.fr/qfapv4/2024/02/15/huge-a0.jpeg
texte_alternatif_de_l_image	YOWL en concert au Supersonic Records !
credit_de_l_image	
mots_cles	Concert,Musique
nom_du_lieu	Supersonic Records
adresse_du_lieu	9 Rue Biscornet
code_postal	75012
ville	Paris
coordonnees_geographiques	48.8502469839762, 2.36970702296574
acces_pmr	
acces_mal_voyant	
acces_mal_entendant	
transport	Métro -> 1 : Bastille (Paris) (304m)+ Bus -> 29618791 : Lyon / Daumesnil - Ledru Rollin (Paris) (246m)+
url_de_contact	
telephone_de_contact	
email_de_contact	
url_facebook_associee	https://fb.me/e/1ITtFUYl2
url_twitter_associee	
type_de_prix	payant
detaill_du_prix	Billet : 13.50 EUR
type_d_acces	obligatoire
url_de_reservation	https://link.dice.fm/te7746778696
url_de_reservation_texte	DICE
date_de_mise_à_jour	2024-02-15T11:59:14+01:00
image_de_couverture	
programmes	
en_ligne_address_url	
en_ligne_address_url_text	
en_ligne_address_text	
title_event	YOWL en concert au Supersonic Records !
audience	Tout public.
childrens	
groupe	Aucun