

Partie II

N.B. : Cette deuxième partie concerne exclusivement les étudiants inscrits au cours INFOH303. Les consignes ci-dessous s'ajoutent et complètent les consignes données précédemment dans le cadre de la partie 1, qui restent d'application.

Introduction

Grâce à votre modèle entité-association et votre modèle relationnel, l'équipe *implémentation* de l'*Air Travel Association for Statistics, Computing and Optimization* (ATASCO) a pu créer une base de données PostgreSQL contenant les données transmises par les compagnies aériennes. Le schéma de cette base de données correspond au modèle relationnel finalement arrêté, qui se trouve en pièce jointe (entities.pdf, relational.pdf). Un backup de cette base de données est également disponible en pièce jointe (infoh303_202122.sql).

Il vous est maintenant demandé de charger ces données et de réaliser des requêtes.

Installation

Vous pouvez créer la base de données en utilisant le moyen de votre choix (pgadmin.org, ligne de commande, etc.). L'installation de PostgreSQL et le chargement des données font partie de l'exercice.

Requêtes

Les requêtes peuvent également être effectuées en utilisant le moyen de votre choix (pgadmin.org, ligne de commande, etc.) mais doivent impérativement être écrites en SQL. Vous ne pouvez donc pas utiliser un ORM (Object–relational mapping) qui automatiserait ces requêtes pour vous.

Proposez le SQL de chaque requête ainsi que si possible, l'algèbre relationnelle et le calcul relationnel tuple. Dans le cas de l'algèbre relationnelle et du calcul tuple, il se peut que cela nécessite une syntaxe non vue au cours ou au TP. Dans ce cas, vous n'êtes pas tenus de les réaliser.

Requête 1

Comptez le nombre de vols effectués par des avions de fret.

Requête 2

Sélectionnez les pilotes qui ont voyagé comme passagers.

Requête 3

Donnez les caractéristiques du vol (identifiant, ville de départ, ville d'arrivée, heure de départ, heure d'arrivée) ayant compté le plus grand nombre de passagers à ce jour.

Requête 4

Sélectionnez le nom des pilotes n'ayant conduit que des avions de ligne dans la base de données.

Requête 5

Les compagnies aériennes doivent avoir accès à des statistiques de base. Pour chacune des compagnies suivantes, 'ABX Air Inc', 'ADVANCED AIR, LLC', il vous est demandé de sélectionner, pour chaque jour, la distance moyenne parcourue par ses avions.

Requête 6

Afin d'améliorer leurs offres pour le voyageurs business, l'association souhaite que vous sélectionniez les destinations où l'aller-retour est possible en un jour.

Un aller-retour est considéré possible en un jour si le vol aller décolle après 7 heures du matin, que le vol de retour décolle le même jour que le vol aller et qu'il s'écoule au minimum 7 heures entre l'atterrissage et le décollage du retour (sans quoi le voyageur n'aurait pas le temps pour ses réunions). Attention : seuls les vols de ligne rentrent en ligne de compte pour le calcul.

Requête 7

Afin d'éviter le gaspillage des ressources, il vous est demandé de sélectionner les vols comptant moins de 20 passagers et de calculer, par compagnie, la moyenne des sièges occupée sur ces vols. Ordonnez les résultats par moyenne croissante.

Requête 8

Afin de s'assurer que les pilotes ne travaillent pas trop de jours d'affilé, comptez le maximum de jours consécutifs qu'un pilote a volé. Attention: la date peut changer entre l'heure de décollage et l'heure d'atterrissage et il faut tenir compte de cette possibilité. Ordonnez les résultats en commençant par les pilotes ayant volé le plus de jours consécutifs.

Requête 9

Le fichier csv_file.csv contient l'id de certains pilotes que l'ATASCO va nommer comme experts internes en 2022. On vous demande de charger ce fichier dans une base de données en utilisant des requêtes SQL. La nouvelle table à créer doit contenir trois colonnes : une clé primaire piloteld référençant l'id du pilote, une date et un champ booléen « nouveau » qui indique si le pilote a déjà servi comme expert ou non. Le csv contient deux colonnes : une colonne de texte contenant la chaîne de caractère « existing-expert-- » + id ou « new-expert » + id, et la date du début de sa nomination.

Requête 10

Pour finir, l'association souhaite créer des groupes de discussion permettant aux aéroports d'échanger sur différents sujets: mobilité, économie, écologie, etc. L'idée est que l'association crée ces différents groupes et permette aux différents aéroports de s'inscrire pour les sujets auxquels ils souhaitent participer. Écrivez le DDL nécessaire à la création des tables requises, et créez-les dans votre base de données. Insérez au minimum un aéroport pour chacun des 3 sujets énoncés ci-dessus.

Ces requêtes sont en très grande partie réalisables sur la base de ce qui a été vu au cours. Quelques méthodes non vues peuvent toutefois vous aider :

- le casting (<https://www.postgresql.org/docs/14/sql-expressions.html#SQL-SYNTAX-TYPE-CASTS>) permet de convertir le type d'une donnée. Un *timestamp* peut par exemple être converti en *date* en utilisant *NomDeLaColonneContenantLeTimestamp::date* dans votre requête ;
- la opérations sur les *Date/Time* sont très importantes et peuvent être trouvées ici : <https://www.postgresql.org/docs/14/functions-datetime.html> ;
- la requête 8 peut être réalisée au moyen d'une window function (<https://www.postgresql.org/docs/14/tutorial-window.html>) et des fonctions associées (<https://www.postgresql.org/docs/14/functions-window.html>);
- pour la requête 9, nous vous conseillons d'utiliser une table temporaire (<https://www.postgresql.org/docs/14/sql-createtable.html>) pour charger les données. Vous pouvez éclater une chaîne de caractère en utilisant la fonction *split_part* (<https://www.postgresql.org/docs/14/functions-string.html>). Vous pouvez vérifier si une chaîne de caractères en contient une autre en utilisant des expressions régulières ou du pattern matching (<https://www.postgresql.org/docs/14/functions-matching.html#FUNCTIONS-POSIX-REGEXP>). Enfin, Postgres permet des expression conditionnelles (<https://www.postgresql.org/docs/14/functions-conditional.html>).

Rapport

Pour la deuxième partie, on vous demande de déposer sur l'UV un fichier unique sous format pdf contenant:

1. un corrigé de votre modèle entité-association et votre modèle relationnel : en vous référant au corrigé, relevez les améliorations possibles de votre modèle et expliquez-les. Si vos choix vous semblent justes mais différents de ceux du corrigé, défendez-les.
2. vos requêtes (1 par page) écrites en SQL et, si possible avec la syntaxe vue au cours ou au TP, en algèbre relationnelle et en calcul tuple.

Informations pratiques

La deuxième partie, en version PDF, devra être rendue sur l'UV pour le 15 mai à 23h59.

La remise du projet dans les délais est obligatoire (une copie par groupe suffit) et aucun retard ne sera accepté pour la remise des documents du projet. Il n'est pas possible de changer les groupe entre la partie 1 et 2.

Après le dépôt sur l'UV des documents demandés ci-dessus, la défense du projet aura lieu dans les jours qui suivent selon un horaire à déterminer. Au sein du groupe, chacun d'entre vous devra être capable de présenter et de justifier vos choix de modélisation ainsi que vos requêtes. Sur base de notre évaluation des acquis au sein de la défense, il se peut que nous n'attribuions pas la même note à chaque membre du groupe.

Vous pouvez développer sur votre propre machine et présenter vos projets sur un ordinateur portable que vous apporterez lors de la défense.

Objectifs évalués lors de la défense orale du projet

Pour les étudiants inscrits en INFOH303, la note finale de votre projet sera calculée sur un total de 10 points répartis entre les objectifs suivants. En particulier, vous devez être capables de :

- Discuter et défendre votre modèle entité-association et sa traduction relationnelle et les comparer par rapport à la solution donnée en justifiant vos choix de modélisation. (3 points)
- Expliquer l'écriture de vos requêtes (algèbre, calcul relationnel tuple et SQL) au moment de la défense. Attention, il s'agit d'une évaluation orale qui ne porte que sur les requêtes que les évaluateurs vous demanderont de présenter. Faute de temps, l'ensemble des requêtes que vous aurez préparées ne sera pas systématiquement évalué. Lors de la défense vous devez aussi apporter votre base de données et pouvoir y exécuter des requêtes. Par exemple, en plus des requêtes demandées dans le présent énoncé, les évaluateurs pourront vous demander d'exécuter une ou plusieurs requêtes dont ils vous fourniront la syntaxe SQL au moment même de la défense. (5 points)
- Réaliser une présentation de bonne qualité faisant montre d'une bonne maîtrise transversale de la matière. (2 points)

Nota Bene :

1. Toutes les questions au sujet du projet seront abordées UNIQUEMENT pendant les séances de travaux pratiques (pas de mails ou messages sur Teams).
2. L'utilisation d'SQLite n'est pas autorisée dans le cadre du projet

3. Les données sont issues de https://www.transtats.bts.gov/DL_SelectFields.asp?gnoyr_VQ=FIL&QO_fu146_anzr=Nv4%20Pn44vr45 et ont été enrichies de données fictives complémentaires à des fins d'exercice.