

Bases de données

Projet

Modalités

Le projet est à réaliser en binôme.

Une modélisation préliminaire (avec une brève explication) doit être rendue sur Moodle le 10 avril 2022. Avant de procéder à la réalisation/implémentation du projet la modélisation doit être validée par votre chargé de TP.

Le projet sera à rendre sans doute dans la deuxième quinzaine de mai.

I) Présentation générale

L'objectif du projet est la modélisation, le peuplement, et la mise en place d'une base de données regroupant des données autour du commerce maritime au XVIII^e siècle, quelque soit sa localisation dans le monde.

a) Navires

Un navire est défini par un identifiant, un type (Flûte, Galion, Gabare...), une catégorie (qui est entier entre 1 et 5 correspondant à la taille), une nationalité, un volume de marchandise pouvant être transporté et un nombre de passager maximum (en plus de l'équipage).

La nationalité d'un navire peut varier en fonction du temps. En effet, certaines nations en guerre peuvent capturer les navires de leurs ennemis pour s'approprier le bateau ainsi que la cargaison.

b) Voyages

Un navire effectue des voyages, définis par une provenance et une destination (qui sont tous les deux des Ports). Chaque voyage possède une durée qui dépend de la distance entre la provenance et la destination. Au départ d'un voyage, le navire doit avoir sa cale pleine de cargaison et/ou transporté le nombre maximum de passagers (dans un souci de rentabilité). A l'arrivée, le navire se vide complètement (passagers et cargaisons), mouille 2 semaines (le temps que l'équipage se repose) et repart.

Un voyage boucle (provenance = destination) est possible.

Chaque voyage est aussi déterminé par une date de début et une date de fin. Par souci de simplicité, on considérera que les voyages respectent toujours les dates pré-déterminés, sauf en cas de prise du navire par une autre nation. Dans ce cas, le bateau n'arrive jamais. De plus, il existe 3 types de voyages en fonction de la distance : court (<1000km), moyen (entre 1000km et 2000km) et long (>2000km). Les voyages courts peuvent servir à transporter n'importe quel type de cargaison. En revanche, pour les voyages moyens et longs, il n'est pas possible de transporter de la cargaison considérée comme périssable (seulement de la cargaison sèche).

Enfin, il vous faudra créer des classes de voyages, différenciés par la localisation continentale (Europe, Amérique, Asie, Intercontinental...). Un voyage s'effectue généralement au sein d'un même continent. Il existe des voyages intercontinentaux, mais il ne peuvent s'effectuer qu'avec des navires de catégorie 5, et donc arriver dans les ports en capacité de les accueillir. On considérera cependant qu'une distance inférieure à moins de 1000km ne peut pas être intercontinental (utile surtout pour les échanges en mer Méditerranée, où 3 continents sont proches). On choisira alors le continent de la destination comme classe.

c) Étapes transitaires

Un voyage peut-être construit en plusieurs étapes. Les voyages les plus longs ($>2000\text{km}$) doivent obligatoirement avoir une étape (environ au milieu, $\pm 500\text{km}$), afin de se ravitailler. Lors d'un voyage, un navire peut décider de vendre une partie de sa cargaison, ou de remplir une partie de sa cale avec les produits disponibles à l'étape ou les deux à la fois (vendre pour faire de la place et la remplir aussitôt). De même, il peut prendre des passagers en plus, ou en déposer certains, ou les deux.

Pour les étapes transitaires, il n'y a pas de dates planifiées.

d) Ports

Un port est déterminé par un nom, une localisation, une nationalité, et une catégorie (entier entre 1 et 5). Un bateau peut mouiller dans un port si et seulement si sa catégorie est inférieure ou égale à celle du port. Lors d'un voyage avec des étapes, on considérera qu'un navire peut se ravitailler dans n'importe quel port, mais ne peut en revanche commercer ou débarquer dans les ports dont la catégorie est inférieure à la sienne.

e) Produits

Un produit est déterminé par un identifiant, un nom et un attribut qui indique s'il est considéré comme sec ou périssable. Exemple de produits : Oranges de Porto, Beurre (périssable); Boisseaux de Sel, Chandelles, Savons, Sucre, Rhum, Sucre, Café, Tabac (sec). Les produits pourront être regroupés par catégories possédant des caractéristiques spécifiques (e.g. durée de conservation, valeur au kilo etc).

f) Cargaison

Une cargaison appartient à un navire, mais correspond également à un voyage de ce navire. Elle est définie par un ensemble de produits, chaque produit ayant une quantité associée. Le volume total des produits multiplié par leur quantité ne doit pas dépasser la taille de la cale d'un navire (le volume pouvant être transporté).

g) Nations et Relations Diplomatiques

Une nation est déterminée par un nom. Chaque nation possède une liste de ports. On considérera que les ports ne peuvent changer de nationalité.

Il existe des relations diplomatiques entre chaque nation, qui peuvent être : alliés commerciaux, alliés, neutres, en guerre. Si deux nations sont en guerre, le voyage d'un navire

appartenant à une des deux nations ne peut se faire depuis ou vers un port de l'autre nation. Exemple : la nation R. et la nation U. sont en guerre. Un voyage d'un navire ayant pour nationalité U. ne peut avoir comme provenance ou destination un port de nationalité R.

Par ailleurs, si deux nations sont alliés commerciaux, un navire d'une des deux nations privilégiera toujours d'avoir ces deux nations comme couple (provenance/destination). Exemple : la nation R. et la nation U. sont en guerre. Un voyage d'un navire ayant pour nationalité U. ne peut avoir comme provenance ou destination un port de nationalité R.

II) Peuplement des tables

Les tables seront alimentées à partir de fichiers csv. Vous mettrez tous les fichier csv utilisés dans le répertoire CSV pour ne pas les mélanger avec les scripts sql.

Bien sûr le travail le plus ingrat consiste à produire les fichiers csv avec les données. Il est irréaliste de trouver sur internet les fichiers prêts à être utilisés directement.

Vous pourrez alimenter vos tables avec la technique qui consiste à

- importer les fichier csv avec COPY dans des tables temporaires,
- « oublier » les attributs inutiles en supprimant les colonnes avec (`ALTER TABLE ... DROP COLUMN ...`),
- alimenter vos tables à partir de ces tables temporaires, par exemple faire les couples nom, prénom à partir de deux tables, une avec des prénoms et une autres avec des noms.

III) Les requêtes

Imaginez 20 questions sur la base de données que vous avez modélisée, et écrivez les requêtes SQL pour y répondre. L'originalité des questions et la difficulté des requêtes (si tant est que celle-ci soit nécessaire) seront prises en compte dans la notation. Parmi vos requêtes, il faut au minimum :

- une requête qui porte sur au moins trois tables ;
- une 'auto jointure' (jointure de deux copies d'une même table)
- une sous-requête corrélée ;
- une sous-requête dans le FROM ;
- une sous-requête dans le WHERE ;
- deux agrégats nécessitant GROUP BY et HAVING ;
- une requête impliquant le calcul de deux agrégats (par exemple, les moyennes d'un ensemble de maximums)
- une jointure externe (LEFT JOIN, RIGHT JOIN ou FULL JOIN) ;
- deux requêtes équivalentes exprimant une condition de totalité, l'une avec des sous requêtes corrélées et l'autre avec de l'agrégation
- deux requêtes qui renverraient le même résultat si vos tables de contenaient pas de nulls, mais qui renvoient des résultats différents ici (vos données devront donc contenir quelques nulls), vous proposerez également de petites modifications de vos requêtes (dans l'esprit de ce qui a été présenté en cours) afin qu'elles retournent le même résultat

- une requête récursive (par exemple pour reconstituer le trajet effectué par un certain bateau sur un laps de temps recouvrant des voyages différents).

Prenez l’habitude de bien programmer vos requêtes SQL. Structurez vos requêtes et surtout indentez-les correctement. N’utilisez pas les sous-requêtes là où une jointure suffirait. N’utilisez pas les vues si c’est uniquement dans le but de simplifier l’écriture d’une requête complexe.

Aussi, il peut être intéressant d’avoir quelques requêtes donnant des statistiques de commerces sur une année (les passagers transportés sur une année, les quantités par produit par année entre deux ports etc)

IV) Le rendu

Le rendu final du projet sera constitué des éléments suivants :

- la modélisation (la version initiale et la version approuvée par le/la chargé/e de TP),
- un script sql qui crée toutes les tables (et des vues si nécessaire), n’oubliez pas d’implémenter le plus finement possible les contraintes dégagées lors de la phase de modélisation,
- des fichiers csv avec les données qui permettent d’alimenter les tables ; vous devez les mettre dans le sous-répertoire CSV pour ne pas les mélanger avec les scripts sql,
- un ou plusieurs scripts sql avec les commandes COPY qui alimentent les tables à partir de fichiers csv, effectuent éventuellement des transformations de tables, pour obtenir à la fin les tables de votre projet,
- le répertoire REQ qui contient les fichiers .sql avec les requêtes. Chaque fichier de requêtes doit contenir une seule requête avec, en commentaire, la description en français de ce qu’elle est censée faire. Le but est de pouvoir facilement exécuter vos requêtes pendant les soutenances. Les fichiers doivent être numérotés (par exemple `req1.sql`, `req2.sql` etc.)
- un fichier sql qui contient toutes les requêtes du répertoire REQ concaténées l’une après l’autre (avec les numéros de requêtes et les descriptions). Ce fichier sera utilisé par le jury de soutenance pour voir facilement la totalité de vos requêtes.
- le fichier README qui contient le nom et prénom et le numéro d’étudiant de chaque membre du binôme, et explique brièvement et clairement ce que contient chaque fichier et dans quel ordre il faut exécuter les différents scripts pour créer et alimenter la bases de données.

En cas d’utilisation de données publiques (fichiers csv disponibles publiquement), donnez dans README les références de sources.