



# Bases de données

## Projet

### Modalités

Le projet est à réaliser en binômes. Deux notes seront attribuées dans le cadre du projet : la première pour une modélisation préliminaire, la deuxième pour l'ensemble de la réalisation de projet.

Une modélisation préliminaire (avec une brève explication) doit être rendue sur Moodle le 11 avril 2021. Cette modélisation donnera lieu à une note. 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 ; les modalités définitives vous seront communiquées ultérieurement.

### I) Présentation générale

L'objectif du projet est la modélisation, le peuplement, et la mise en place d'une base de données d'un site de e-commerce.

Le choix de produits ou des classes de produits à vendre est laissé à chaque binôme.

#### a) Produits

Les produits seront regroupés en classes, par forcément disjointes, et les classes peuvent former une hiérarchie.

Le choix exact des informations à stocker pour chaque produit dépend largement du type de produit. Par exemple la marque est pertinente pour les produits industriels, alors que pour un commerce de fruits et légumes frais savoir si le produit est local et/ou bio est cruciale pour les clients. Bien sûr il faut connaître le prix de produit. Mais le prix varie dans le temps et il est important de garder l'information sur tous les changements de prix (par exemple pour suivre l'influence du prix sur le volume de vente).

Le produit peut-être en stock (peut-être avec le nombre exacte d'items disponibles immédiatement), disponible sur commande (avec un délai) ou temporairement indisponible.

#### b) Clients

Pour les clients il faut garder les informations habituelles : nom, prénom, adresse, mail, peut-être le numéro de téléphone.

Dans certains cas il peut être pertinent de garder les informations comme la date anniversaire pour envoyer au client des offres « promotionnelles » ou tout simplement des vœux.

On peut garder aussi la date d'inscription du client sur le site.

### c) Commandes

Finalement ce qui intéresse le plus le commerçant ce sont les commandes. Pour chaque commande il faut enregistrer la date de la commande (ou plutôt le timestamp c'est-à-dire la date et l'heure). Chaque commande peut concerner plusieurs produits. Chacun de ces produits peut être livré séparément donc pour chaque produit de la commande il faut maintenir son état : livré (avec la date d'expédition et la date de livraison), en livraison (avec la date d'expédition), en préparation, en attente (pour le produit qui n'est pas en stock et a été commandé chez le fabriquant, sans doute la date prévue de livraison dans notre dépôt). La commande peut contenir l'adresse de livraison si elle est différente de l'adresse de facturation (on considère que l'adresse de facturation est l'adresse du client). Tant que la commande n'a pas été envoyée elle peut être totalement ou partiellement annulée par le client, ce qui implique le remboursement.

Pour les commandes qui peuvent être payées par chèque il faut noter si la commande a été effectivement payée.

Pour certains types de produits il faut sans doute prévoir la possibilité de refus du client à la livraison (produit arrivé abîmé).

Pour certain produits (tous ?) il faut prévoir la possibilité de retour, limité en temps, contre un remboursement. Il faudra enregistrer les demandes de retour et les retours effectifs. Pour les retours il peut être intéressant de noter la raison (par exemple produit défectueux).

Et n'oubliez pas que pour passer une commande il faut d'abord constituer un panier. D'habitude on peut faire un panier sans être identifié comme un client (l'identification est nécessaire pour passer la commande).

On ne demande pas de modéliser les interactions avec les banques en cas de paiement en ligne. Il suffit de noter que la commande a été payée.

### d) Les avis et les notes

Il n'y a plus de site e-commerce sans avis et sans notation de produits par les clients. Bien sûr il faut éviter la situation où quelqu'un qui n'a pas acheté un produit donne son avis sur ce produit.

## II) Peuplement des tables

Les tables seront alimentées à partir de fichiers csv. Vous mettrez tous les fichiers 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.

Cependant il y a beaucoup de fichiers disponibles qui permettent facilement d'obtenir une partie de données : par exemple <https://www.data.gouv.fr/fr/datasets/la-liste-annuelle-des-prenoms-des-nouveaux-nes-ville-de-nancy/>, <https://www.data.gouv.fr/fr/datasets/liste-de-prenoms-et-patronymes/>, <https://www.insee.fr/fr/statistiques/3536630> permettent d'obtenir des noms et des prénoms,

<https://www.data.gouv.fr/fr/datasets/base-d-adresses-nationale-ouverte-bano/> fournit une très longue liste d'adresses, <http://www.generatedata.com/#generator> est un générateur de données en format csv.

Il y a beaucoup d'autres données de ce type. Ces fichiers ne correspondent pas exactement à vos tables mais 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.

Pour les commandes sans doutes un petit programme en java ou python permettra de générer des données aléatoires qui correspondent aux données dans vos tables.

La tâche la plus compliquée concerne sans doute les produits. Mais peut-être trouverez-vous des données appropriées sur internet. Sinon il reste toujours la méthode pénible de fabriquer un fichier csv de produits à la main.

### III) Les requêtes

Imaginez 20 questions sur le site de e-commerce que vous avez modélisé, 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)
- deux requêtes équivalentes exprimant une condition de totalité, l'une avec des sous requêtes corrélées et l'autre avec de l'agrégation
- une jointure externe (LEFT JOIN, RIGHT JOIN ou FULL JOIN) ;
- 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

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 pour éviter d'écrire des requêtes complexes. Utilisez les USING uniquement là où c'est justifié.

## 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.