

Déroulé du projet ecommerce

vidéo de présentation du projet complet : <https://youtu.be/HL15zD-kppc>

Semaine 1

vidéo de présentation du programme de la semaine 1 : https://youtu.be/MOWrK_M2xxs

Le but est de constituer un jeu de données en vue de son analyse ultérieure. On suppose ici qu'on part d'un ensemble de données existantes sous forme de fichiers csv que l'on va transformer et enrichir (ici par de la génération aléatoire de données). Le résultat de ces transformations sera une base de données composée de plusieurs relations chargées sous MySQL (appelée schéma cible).

Schéma cible :

Les clés sont soulignées et les clés étrangères sont en gras.

Clients(idClient, prénom, nom, email, telephone, dateNaissance, mdp, ville, pays)

Produits(idProduit, nomproduit, prix, poids, **idCategorie**) : idCategorie clé étrangère sur Categories

Categories(idCategorie, nomCategorie, **idCategorieMere**) : idCategorieMere clé étrangère sur Categories (on suppose qu'il existe une supercatégorie de nom top qui est sa propre catégorie mère). Categories va être construite à partir de products.csv

Commandes(idCommande, **idClient**, dateCommande, statut) : idClient clé étrangère sur Clients -> générer aléatoirement à partir de clients

LignesCommandes(**idCommande**, noligne, **idProduit**, quantité) : idCommande clé étrangère sur Commandes et idProduit clé étrangère sur Produits -> générer aléatoirement à partir de Commandes, Produits

Les données initiales sont fournies dans une archive constituée de plusieurs fichiers csv :

- customers.csv : des clients
- products.csv : des produits
- Villes.csv : des villes du monde

Travail à faire

- installation des outils : looping-mcd, wamp, notepad++
- récupération des jeux de données (fichiers csv)
- comprendre les jeux de données (les regarder avec un tableur)

- concevoir le schéma conceptuel (MCD) du schéma cible avec looping-mcd et générer le code SQL de création des relations
- créer une base de données appelée ecommerce avec phpmyadmin
- importer chacun des tableaux csv dans la base ecommerce (une relation créée par tableau csv)
- comparer le schéma de la BD cible avec les schémas des relations issues des tableaux csv

Supports de cours :

- utilisation de looping-mcd : <https://youtu.be/kqT9Y8q8ARc>
- utilisation de phpmyadmin: <https://youtu.be/hL2lWy2WuUE>
- cours sur modèle conceptuel de données : <https://www.youtube.com/watch?v=0tJvackhcjw>
- cours sur modèle relationnel de données : https://www.youtube.com/watch?v=SqJ_Z5G9CAE
- cours sur transformation MCD vers MLD : <https://www.youtube.com/watch?v=qCUY30MIHII>
- cours sur rétro-conception MLD vers MCD : <https://www.youtube.com/watch?v=WbjD4VWzzYw>
- cours sur SQL (langage définition de données) : <https://www.youtube.com/watch?v=AJqVC0fT7gg>

A rendre :

- MCD fait avec looping-mcd (cela peut être une copie d'écran)
- schéma relationnel produit par looping-mcd
- code SQL de création des relations
- copie d'écran de phpmyadmin montrant la liste des relations créées

Semaine 2

vidéo de présentation du programme de la semaine 2 : https://youtu.be/MOWrK_M2xxs

- installation de python et de la librairie python-mysql
- mettre en place des programmes python simples
 - connexion à la BD
 - affichage de 20 lignes d'une relation au choix
- générer aléatoirement des données en utilisant les fonctions de génération python (random et choice)

Support de cours :

- tutoriel python mysql : <https://python.antoinepernot.fr/cours.php?course=chap6>
- vidéo python mysql : <https://youtu.be/2R-BveCE-so>
- fonctions de génération aléatoire en python : <https://docs.python.org/3/library/random.html>

Rendu :

- programmes simples en python
- programme de génération aléatoire en python

Semaine 3

vidéo de présentation du programme de la semaine 3 : https://youtu.be/MOWrK_M2xxs

Il s'agit de peupler la BD cible avec les données importées à partir des fichiers csv plus des données générées aléatoirement (selon les règles décrites ci dessous) en utilisant python.

Règles de transformations :

- par création de relations, renommage des attributs (en SQL exécuté sous phpmyadmin) :
 - T1 : Relations clients par simple création puis renommage en SQL (-create table clients as select customer_id as idClient, ... from customers) en ajoutant deux attributs ville (varchar(37)) et pays (varchar(44)) ainsi que mdp (varchar(40))
 - T2, T3, T4, T5 : Relations Commandes, LignesCommandes, Produits, Catégories en créant simplement leur structure avec des create table (issus de looping-mcd)
- En écrivant du code python+SQL (en utilisant mysql-connector)
 - T6 : Ajouter des données dans vendeurs pour les attributs ville et région en prenant aléatoirement un couple ville – région dans Villes pour chaque client
 - T7 : ajouter un mot de passe généré aléatoirement pour chaque client
 - T8 : Ajouter des données dans commandes en générant aléatoirement entre 0 et 5 commandes par client (présent dans clients). L'idcommande sera généré automatiquement en tant qu'attribut auto-incrémenté, la dateCommande sera prise aléatoirement sur l'intervalle 1/1/2020-1/9/2022, le statut sera pris aléatoirement dans la liste {en cours, payée, livrée, annulée}
 - T9 : Ajouter des données dans LignesCommandes en générant aléatoirement entre 1 et 3 lignes par Commandes. Noline vaudra donc entre 1 et 3, idProduit sera pris aléatoirement dans Produits et quantité aléatoirement entre 1 et 4
 - T10 : Ajouter des données dans Categories en utilisant l'attribut category (varchar(157)) de products. Il s'agit ici de décoder la valeur de cet attribut en utilisant le séparateur | Par exemple une valeur comme « Arts, Crafts & Sewing | Beading & Jewelry Making | Beading Supplies | Fuse & Perler Beads » se découpe en 4 catégories avec « Fuse & Perler Beads » comme catégorie la plus précise qui a pour mère « Beading Supplies » qui a pour mère « Beading & Jewelry Making » qui a pour mère « Arts, Crafts & Sewing » (qui a pour mère top). Cette valeur va donc générer jusqu'à 4 lignes dans la

relation Categories (dépend si tout ou partie d'entre elles existait auparavant). IdCategorie sera généré automatiquement comme attribut auto-incrémenté.

- T11 : ajouter les données dans produits à partir de productso et en faisant une jointure avec Categories pour récupérer idCategorie (cf plus haut remplissage de Categories). Attention de transformer les valeurs de prix en réels en supprimant l'unité monétaire et en supprimant le cas échéant les produits ayant une valeur de prix non interprétable facilement (par exemple « from 1 seller »). Pareil pour le poids (normaliser les unités, 1 ounce=0.0625 pound, puis supprimer l'unité).

Travail à faire

- implémenter les différents règles de transformations soit en écrivant du code SQL (exécuté sous phpmyadmin), soit en écrivant du code python+SQL, soit un mixte des deux
- vérifier que les données sont bien chargées dans les différentes tables du schéma cible et que les règles de transformations ont bien été appliquées
- ajouter les contraintes d'intégrité (venant de looping-mcd ou ajoutées par vous même) de manière à ce que chaque relation ait une clé, une clé étrangère (si besoin), des contraintes domaines (si besoin)

Support de cours

- tutoriel python mysql : <https://python.antoinepernot.fr/cours.php?course=chap6>
- vidéo python mysql : <https://youtu.be/2R-BveCE-so>

Rendu

- un fichier texte donnant pour chaque règle de transformation (de T1 à T11) la séquence d'opérations à faire, une opération étant soit du code SQL, soit du code python+SQL (on en donnera le source)
- un fichier texte donnant la liste des contraintes d'intégrité ajoutées (en SQL)
- pour chaque table du schéma cible, le nombre de lignes créées (obtenu par un select count(*))

Semaine 4

vidéo de présentation du programme de la semaine 4 : https://youtu.be/MOWrK_M2xxs

- analyser les données via des requêtes de type SQL (au moins 15). Il doit y avoir avoir au moins 10 requêtes avec jointure et 5 requêtes avec group by (dont au moins 3 avec HAVING) :
 - compter le nombre de lignes par relations
 - nombre de produits par catégorie
 - produits les plus vendus
 - les meilleurs clients
 - ...
- définir 3 vues relationnelles (avec des CREATE VIEW) en choisissant 3 des requêtes précédentes (ne pas prendre les plus simples mais celles qui ont un réel intérêt applicatif donc en général celles qui ont un group by)

Support de cours :

-

Rendu :

- les 15 requêtes d'analyses en SQL (avec une explication en français de ce qu'elles font) et la définition des 3 vues (en SQL)

Une séance de présentation de l'état d'avancement du projet sera organisée cette semaine là. Chaque groupe va présenter ce qui a été fait et aura un retour sur les points de corrections éventuels à apporter. Cela permettra d'aborder au mieux ce qui reste à faire.

Semaine 5

Vidéo de présentation de la semaine 5 : <https://youtu.be/XHy3bc0l3vc> (jupyter et pandas), <https://youtu.be/1mVFfkKZ41M> (Exploration des données)

- installation de jupyter notebook : <https://jupyter.org/install>
- installation de dataprep : <https://docs.dataprep.ai/installation.html>

Préalable :

- se former à la librairie pandas (voir vidéo sur pandas dans la partie support de cours)


Exploration et qualité des données

- on va utiliser dataprep.eda pour explorer des données. C'est plus facile de le faire via un notebook jupyter en mode web. Pour commencer pratiquer dataprep avec le jeu de données du titanic (https://docs.dataprep.ai/user_guide/eda/titanic.html) puis celui sur le prix des maisons (https://docs.dataprep.ai/user_guide/eda/house_price.html)
- il est possible aussi d'utiliser pandas-profiling (maintenant ydata-profiling) à la place de dataprep.eda (<https://github.com/ydataai/ydata-profiling>)
- choisir une des relations de la BD ecommerce et appliquer la même démarche d'exploration (on pourra exporter la relation choisie en format csv)
- coder la transformation T3 (de la semaine 2) en utilisant pandas à la place du code SQL.

Rendu

- programme python ou notebook ayant servi à analyser un des fichiers csv de ecommerce. Il faudra commenter ce qui est fait pour expliquer l'intérêt pratique de chaque analyse effectuée
- programme pandas implémentant T3

Supports de cours

-  **DataPrep Python library for Easy Data Preparation and EDA** vidéo sur dataprep.eda
- <https://www.youtube.com/c/Coreyms/videos> : vidéos sur python couvrant EDA, pandas, matplotlib
- <https://www.youtube.com/watch?v=zZkNOdBWgFQ> : vidéo sur pandas
- <https://github.com/ue12/python-numerique/tree/master/notebooks> : un bon cours python avec pandas, matplotlib
- https://youtu.be/6rOB_DvfzOw vidéo sur ydata profiling

Semaine 6

Vidéo de présentation de la semaine 6 : https://youtu.be/MOWrK_M2xss

- installation de jupyter notebook : <https://jupyter.org/install>
- installation de matplotlib et seaborn : <https://matplotlib.org/stable/index.html> et <https://seaborn.pydata.org/installing.html>

Visualisation des données

- on va utiliser deux librairies de visualisation pour python : matplotlib et seaborn (plus haut niveau)
- se former à matplotlib : regarder les vidéos de la section support de cours
- se former à seaborn : regarder les vidéos de la section support de cours
- choisir un fichier csv parmi les fichiers ecommerce et en faire 2 ou 3 affichages simples
- se mettre à la place de l'administrateur du site de ecommerce et définir quelques statistiques intéressantes (au moins 2) : par exemple les 30 produits les plus vendus, la ventilation des ventes par catégorie de produits
- mettre en place la visualisation de ces statistiques avec soit matplotlib, soit seaborn
 - sélectionner les données à afficher via une requête SQL à exécuter dans phpmyadmin, puis exporter le résultat sous forme d'un fichier csv
 - afficher les données du fichier cvs exporté avec matplotlib ou seaborn

Rendu

- les programmes python permettant un affichage simple issu du jeu de données ecommerce (cela peut être un format notebook ou python)
- la définition des statistiques intéressantes sur le ecommerce (en français)
- pour chaque statistique intéressante, la requête SQL permettant de sélectionner les données utiles et le code python permettant l'affichage (notebook ou python)

Supports de cours

- <https://www.youtube.com/c/Coreyms/videos> : vidéos sur python couvrant EDA, pandas, matplotlib
- https://www.youtube.com/watch?v=O_OeWxpnUc0&t=26s , https://www.youtube.com/watch?v=O_OeWxpnUc0 et <https://www.youtube.com/watch?v=MILtbfrMGL4> et <https://www.youtube.com/watch?v=UO98IJQ3QGI> : vidéos sur matplotlib
- <https://www.youtube.com/watch?v=xYgflRzNPlo> : vidéo sur seaborn
- http://spacetelescope.github.io/pylunch/4-matplotlib/Intro%20to%20Matplotlib.slides.html# : tutoriel matplotlib

Semaine 7

Finalisation du projet (tout ce qui reste à faire doit être finalisé)
préparation de la soutenance de cette première partie

Rendu

- vidéo de démonstration de ce qui a été réalisé (de l'ordre de 5mn)
- archive de tout ce qui a été réalisé classé semaine par semaine (un répertoire pour chaque semaine)

Semaine 8

Soutenance de la première partie devant tous les étudiants

Rendu

- un ensemble de transparents (sous forme pdf)

Semaine 2/1/2024

vidéo de présentation de l'UE 2 : <https://youtu.be/ZZvuYSIEBno>

Travail à faire :

- appropriation du schéma de base de données sur l'agence de voyage (voir description du schéma, code SQL et données dans le répertoire Données/DonnéesAgenceVoyage)
- Identification des acteurs de l'application agence de voyage
- Identification des opérations pour chaque acteur
- Spécification des interactions entre les utilisateurs et l'application (au niveau de chaque opération et au niveau de l'application globalement) : graphe de navigation
- Choix de l'interface : structure d'une page, bandeau de navigation, aspects graphiques

Rendu de la semaine

- rétro-conception du schéma de la base de données agence de voyage avec looping-mcd
- création de la BD sous mysql et peuplement avec des données (vous avez des données qui vous sont fournies pour les relations vols, compagnies, aéroports)
- liste des acteurs de l'application agence de voyage et pour chacun d'entre eux les opérations qu'ils peuvent réaliser (dans un document .doc ou .pdf)
- graphe de navigation (sous forme numérique ou scan d'un document papier LISIBLE)
- modèle des pages (structure type d'une page) : sous forme numérique, sous forme numérique ou scan d'un document papier LISIBLE

Supports de cours

- <https://youtu.be/F1o-LxPsgYA> : vidéo sur la programmation BD-Web (dont présentation graphe de navigation)

Schéma de la BD agence de voyage

Semaine 9/1/2024

vidéo de présentation du module Couche présentation et interface :

<https://youtu.be/ZZvuYSIEBno>

Travail à faire :

- Codage de l'interface : html, CSS
- ajout de fonctionnalités de contrôle ou dynamiques en javascript
- évaluation des critères d'ergonomie de l'interface (vous pouvez demander à des amis de tester)

Rendu de la semaine

- archive avec l'ensemble des fichiers produits (html, CSS, javascript)
- copie d'écran de l'affichage sur plusieurs terminaux
- analyse de l'évaluation des critères d'ergonomie

Outils à installer

Supports de cours

- <https://www.w3.org/Style/Examples/011/firstcss.fr.html> : introduction à CSS
- <https://darchevillepatrick.info/index.htm> : site couvrant les différents aspects de la programmation web (html, css, javascript, jquery)
- <https://openclassrooms.com/fr/courses/1603881-apprenez-a-creer-votre-site-web-avec-html5-et-css3> : cours sur HTML5 et CSS

Semaine du 16/1/2024

vidéo de présentation du module développement application web :

<https://youtu.be/vFRtEyjRZKw>

vidéo sur le développement web : https://youtu.be/i5e_OtN2TXg

Travail à faire :

- réaliser des programmes simples en PHP
- apprendre l'API PDO pour gérer la connexion à la BD (faire quelques programmes simples)
- apprendre à gérer cookies et variables sessions
- définir comment générer les id des différentes relations
- choisir le mode de gestion des sessions
- coder la partie connexion/authentification
- raffiner le graphe de navigation construit pour la partie interface pour tenir compte des pages statiques/dynamiques et des sessions
- commencer le codage des opérations

Rendu de la semaine

- mode de génération des id et choix de la gestion de session
- code php de la partie connexion/authentification
- graphe de navigation complet

Supports de cours

- <https://sylvie-vauthier.developpez.com/tutoriels/php/grand-debutant/> : tutoriel PHP
- <https://youtu.be/F1o-LxPsgYA> : vidéo sur la programmation BD-Web (dont présentation graphe de navigation, cookies - sessions et API PDO)

Semaine 23/1/2024 et semaine 30/1/2024

vidéo de présentation du module développement application web :

<https://youtu.be/vFRtEyjRZKw>

semaine du 23/1, suivi de projets sur le travail des 3 semaines précédentes

Travail à faire :

- codage des opérations pour faire une application réaliste
- faire des tests (opération par opération puis de l'application)

Rendu de la semaine

- archive des fichiers php représentant l'application

Supports de cours

- <https://sylvie-vauthier.developpez.com/tutoriels/php/grand-debutant/> : tutoriel PHP
- <https://youtu.be/F1o-LxPsgYA> : vidéo sur la programmation BD-Web

semaine du 6/2/2024

Travail à faire :

- produire une documentation utilisateur et développeur
- préparer une vidéo de démonstration
- préparer la soutenance

Rendu de la semaine

- documentation utilisateur et développeur
- document explicatif de la méthode de test utilisée

semaine du 13/2/2024

soutenance finale des projets