

Côté Labo – Exploitation d'une base de données MySQL

Exemple d'utilisation de données libres de droit

Propriétés	Description
Intitulé long	Utilisation et exploitation d'une base de données MySQL issue de données libres de droit
Formation concernée	BTS Services Informatiques aux Organisations
Matière	SI3 : Exploitation des données
Présentation	<p>Ce côté labo propose un exercice de programmation en SQL. Les données sont importées du site de l'ANSES (Agence Nationale de Sécurité Sanitaire) à partir de fichiers csv. Elles sont libres de droit et peuvent être utilisées dans des entreprises commerciales.</p> <p>Cette base de données peut être réutilisée pour le <i>back-office</i> de nombreuses applications.</p>
Notions	<p>D1.1 – Analyse de la demande A1.1.1 Analyse du cahier des charges d'un service à produire D4.1 – Conception et réalisation d'une solution applicative A4.1.1 Proposition d'une solution applicative A4.1.3 Conception ou adaptation d'une base de données</p> <p>Savoir-faire Extraire et modifier les données d'une base de données Implémenter une base de données à partir d'un schéma existant</p> <p>Savoirs associés Langages et outils d'interrogation et de manipulation d'une base de données</p>
Pré-requis	<p>Modèle relationnel SQL LID, LMD et LDD Installation de MySQL + phpMyAdmin en local ou sur un serveur distant réalisé par le professeur.</p>
Outils	MySQL ou MariaDB, Apache2, phpMyAdmin, PHP
Mots-clés	SQL, MySQL, MariaDB
Durée	6 heures
Auteur(es)	Fabrice Missonnier, relecture par Luc Frebourg, Apollonie Raffalli, Gaëlle Castel et Yann Barrot
Version	v 1.0
Date de publication	octobre 2016

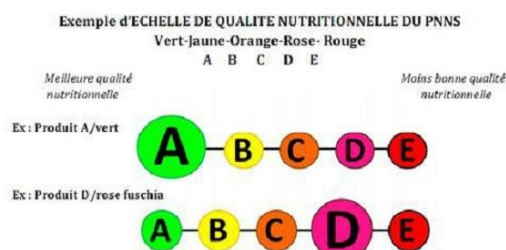
I Contexte

La direction marketing a décidé de travailler sur un projet transversal pour diversifier les activités de l'entreprise GSB. L'objectif du laboratoire n'est plus d'axer sa politique stratégique uniquement vers les entreprises (B2B) mais aussi vers les particuliers (B2C).

Depuis peu, GSB commercialise des compléments alimentaires et se doit de démarcher, en plus des médecins, des nutritionnistes et des pharmaciens. Elle souhaite se démarquer de la concurrence en proposant des applications mobiles gratuites à destination des diététiciens de leurs clients potentiels.

Après de nombreuses réunions entre le DSI, le chef de projet « Applications mobile » et le directeur marketing, il a été décidé de développer une nouvelle application qui devra non seulement répondre à un besoin commercial mais aussi correspondre à des objectifs légaux applicables aux entreprises agro-alimentaires. Elle sera fournie sur le Play Store de Google et permettra au client, avec son smartphone, de lire le code barre d'un produit et de lui en fournir la composition nutritionnelle. Par exemple, le client d'une grande surface pourra savoir de quoi est composé précisément le produit qu'il est en train d'acheter.

A terme, il sera possible de générer des alertes consommateur en fonction de la composition d'un produit (par exemple des conseils sous forme de code multicolore sur les aliments) :



Elle permettra aussi aux nutritionnistes de composer des repas spécifiques en fonction des besoins de leurs patients, tout en traçant très exactement toutes les compositions nutritionnelles des aliments proposés (important en cas d'allergies, de régimes spéciaux, etc.).

Avant de débiter le projet, le DSI de GSB veut s'assurer que les données de l'application répondront réellement aux besoins des nutritionnistes et de leurs clients. Cependant, il paraît très difficile pour le laboratoire de constituer une base d'échantillons alimentaires de tous les produits que l'on trouve habituellement dans le commerce. Le directeur de GSB confirme qu'il est impossible de mobiliser des équipes de scientifiques du laboratoire pour faire ces analyses : ce serait trop coûteux et prendrait beaucoup trop de temps.

Un des collègues nutritionniste, lors d'une conférence, a entendu parler de données libres de droit, utilisables dans un domaine commercial, qui référence des aliments ainsi que leurs compositions nutritionnelles. Elles sont fournies par l'agence nationale de sécurité sanitaire, régie par l'État.

Votre mission est de produire des tests sur ces données libres de droit téléchargées sur le site de l'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail). La base référence un millier d'aliments courants ainsi que leurs compositions nutritionnelles. Elle est récupérée sur Internet. Vous serez en charge de vérifier si le contenu est exploitable, cohérent et s'il convient pour la future application.

Pour mener à bien votre mission, l'équipe de développement a déjà travaillé en amont et vous fournira les requêtes SQL d'insertion de données.

II Installation de l'environnement de développement

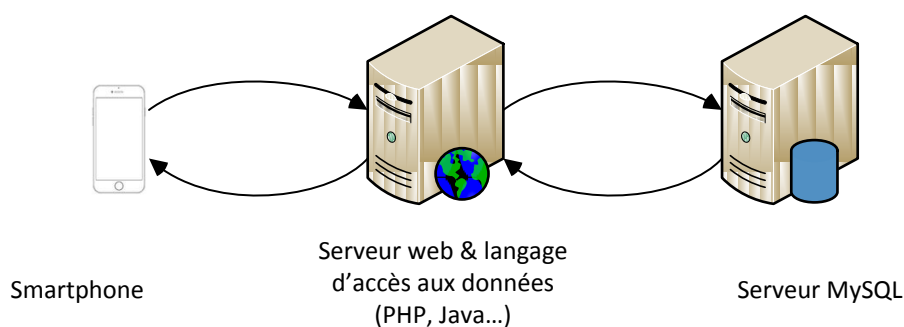
Les données initiales sont fournies sur le site de l'ANSES

<https://pro.anses.fr/tableciqua/>

Elles sont sous licence ouverte promouvant « la réutilisation la plus large en autorisant la reproduction, la redistribution, l'adaptation et l'exploitation commerciale des données ». Le laboratoire GSB a donc le droit de les utiliser et de les exploiter de manière commerciale.

Un travail a déjà été effectué par l'équipe projet pour migrer les données brutes (format csv) en SQL compatible avec MySQL.

Lorsqu'elle sera en production, la base de données sera installée sur un serveur distant MySQL. Ce serveur pourra être utilisé directement par les applications web.



Pour cette mission, les tests se font dans un environnement local où chaque développeur installe sa propre base sur sa machine ou sur son serveur. Dans tous les cas, LAMP, WAMP ou EasyPHP est déjà installé et comporte les logiciels suivants :

- le système de gestion de bases de données relationnelles MySQL,
- le serveur web Apache,
- l'interpréteur de pages dynamiques PHP.

L'outil d'interface web phpMyAdmin permet d'accéder en local à la base de données. Il est, lui-aussi, déjà installé.

III Mission

Le travail à fournir dans cette mission comporte trois étapes :

- l'importation de la base de données sur la machine de développement,
- la description du schéma relationnel de la base,
- la programmation des requêtes SQL permettant de tester les besoins des nutritionnistes.

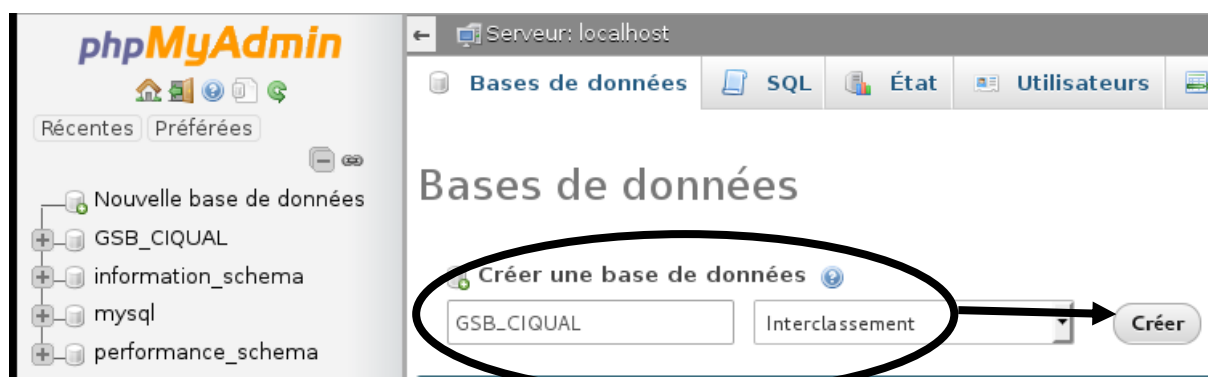
1.1 Créer la base de données CIQUAL

Travail à faire 1

La première étape est de créer la base de données sur la machine locale en se connectant avec l'invite de commande (à adapter en fonction de la localisation du serveur de base de données).

```
[adgsb@localhost ~]$ mysql -h localhost -u root -p
MariaDB [(none)]> CREATE DATABASE GSB_CIQUAL;
MariaDB [(none)]> exit
```

Il est aussi possible de créer la base avec phpMyAdmin :



1.2 Description du schéma relationnel de la base CIQUAL

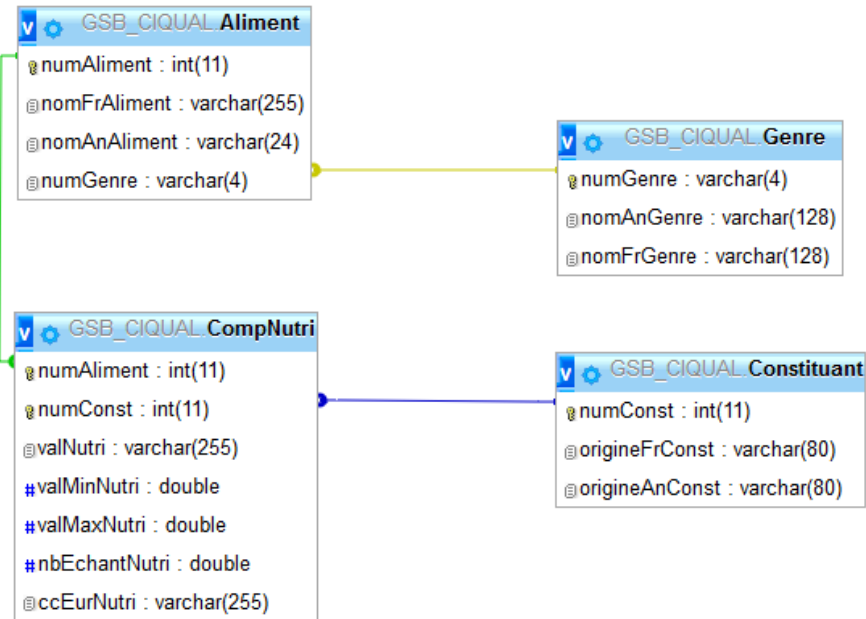
La deuxième étape est de créer le schéma relationnel de la base de données CIQUAL. Le fichier GSB_CIQUAL-Donnees.sql ne comporte que les requêtes d'insertion des tuples (enregistrements). Par rapport au fichier csv, une table « sources scientifiques » a été supprimée car elle ne correspond pas aux besoins du projet.

On trouve à ce lien un fichier au format pdf expliquant en détail la signification de l'ensemble des champs du fichier au format csv :

https://pro.anses.fr/tableciqual/Documents/Ciqual_2013_LISEZMOI_csv.pdf

Travail à faire 2

A partir du fichier et du schéma de relation représenté graphiquement dans la capture ci-après, écrire les requêtes SQL de création des tables Aliment, Genre, ComNutri et Constituant.



Lorsque les tables sont créées, on peut importer les données dans la base CIQUAL. Le fichier étant volumineux, il n'est pas possible d'utiliser l'outil d'importation de phpMyAdmin (taille maximum de 2 Mo). On peut utiliser directement la commande `mysql` dans un terminal :

```
[adgsb@localhost ~]$ mysql GSB_CIQUAL -h localhost -u root -p < GSB_CIQUAL-Donnees.sql
```

1.3 Proposer des requêtes de test SQL

Une réunion d'avant-projet a eu lieu entre le chef de projet, le directeur marketing (maître d'ouvrage), un nutritionniste du laboratoire et trois informaticiens. Le premier développeur est un expert en développement Android. Le second a des compétences en *back-office* (c'est vous-même !) et le troisième est votre collègue qui a effectué la migration de la base. L'objet de cette réunion a été de réfléchir sur la faisabilité du projet. Est-ce que la base de données CIQUAL est réutilisable en l'état dans le projet ? Est-elle cohérente ?

Travail à faire 3

Vous avez listé pendant la réunion un certain nombre de questions techniques. Vous vous donnez une demi-journée pour y répondre en SQL.

Le travail peut être effectué soit sous phpmyadmin, soit en mode console.

1. Quelle est l'utilité de la table Genre ?
2. Le logiciel étant proposé à une clientèle francophone, est-ce que tous les noms des genres ont été écrits en français ?
3. Votre collègue développeur Android ne peut afficher que 10 genres à la fois sur un smartphone 5'. Donnez les 10 premiers genres d'aliments de la base.

4. Toujours à la demande de votre collègue graphiste, il souhaiterait placer un bouton « Suivant » qui liste les 10 genres d'aliments suivants. Programmez la requête correspondante.
5. Le chef de projet pense nécessaire de pouvoir effectuer une recherche simple par genre et par nom d'aliment. Listez, par exemple, les genres qui ont "biscuit" dans leur intitulé français.
6. Au fait, combien y a-t-il de genres dans la base ?!

Les genres et les sous-genres sont identifiés dans la base par un point. Par exemple le genre 13 et ses sous-genres (13.1 et 13.2) correspondent aux lignes suivantes :

numGenre	nomAnGenre	nomFrGenre
13	Fish and batrachians	Poissons et batraciens
13.1	Fish and batrachians	Poissons et batraciens non transformés
13.2	Fish and batrachians products	Produits à base de poissons

7. Donnez tous les sous-genres.

A priori, il est possible que les intitulés des sous-genres soient mal renseignés. Dans l'exemple ci-dessus, en anglais, le sous-genre a le même nom que son genre. Il faut trouver une solution pour identifier les sous-genres qui ont le même nom que leur genre.

8. En utilisant la fonction `LENGTH` de MySQL, donnez le nombre de caractères maximum qui identifie un genre et un sous-genre (<http://sql.sh/fonctions/length>). Le genre est-il toujours codé sur 2 caractères (01, 02 par exemple) ?

Le sous-genre reprend forcément le numéro du genre puis un point et le numéro du sous-genre.

9. En utilisant la fonction `SUBSTRING`, affichez les deux premiers caractères de tous les genres. Attention : comme toutes les fonctions d'agrégat, sous MySQL, il ne faut pas placer d'espace entre le nom de la fonction et la parenthèse.
10. En utilisant la requête précédente, indiquez les intitulés des sous-genres qui ont le même nom français et/ou anglais que leur genre. Votre collègue vous note, pendant la réunion, qu'il faudra certainement utiliser une requête avec une « auto-jointure » (`SELF JOIN`). Cette requête retourne deux résultats :

NumeroSousGenre	NumeroGenre	NomSousGenre	NomGenre
13.1	13	Fish and batrachians	Fish and batrachians
14.1	14	Shellfish and mollusks	Shellfish and mollusks

11. Pour que la base soit cohérente, il faut modifier le nom des sous genres pour qu'ils soient différents de leurs genres
 - le sous-genre 13.1 "Fish and batrachians" deviendra "Unprocessed fish and batrachians",
 - le sous-genre 14.1 "Shellfish and mollusks" deviendra "Unprocessed shellfish and mollusks"

On souhaite maintenant travailler sur les aliments. D'après votre collègue, la base comporterait environ 1 400 aliments répertoriés dans une centaine de genres.

12. Vérifiez si le nombre d'aliments correspond bien à ce qui a été dit en réunion.
13. On souhaite effectuer une recherche par nom d'aliment. Donnez, par exemple, tous les aliments qui ont le mot clé « Pain » dans leur nom français. Fournissez aussi la requête qui permet de lister tous les aliments dont le numéro est compris entre deux entiers fournis par l'utilisateur dans l'interface graphique (9500 et 9700 par exemple).
14. L'application doit aussi permettre de trier les aliments. Proposez la requête qui trie par nom français de A à Z, puis celle qui trie de Z à A.
15. Il est potentiellement gênant d'avoir des aliments dont le genre est « 00 » ('non défini'). Vérifiez s'il existe de tels aliments dans la base.
16. Affichez le nombre d'aliments dont le genre est « 01.1 » ('Farines et amidons').
17. Donnez le nom du genre (numéro de genre et nom français du genre) de l'aliment « Riz rouge, cuit ».
18. Toujours dans un souci de trier les données dans l'interface graphique, votre collègue souhaiterait savoir combien il existe d'aliments par numéro de genre. Il faut obtenir le numéro de genre et le nombre d'aliments. Ce serait d'ailleurs encore mieux, dans le résultat de la requête précédente, d'avoir le nom français du genre affiché !
19. Afin de connaître les genres les plus utilisés, donnez les noms français des genres qui identifient plus de 10 aliments. Le résultat doit comporter les mêmes colonnes que la requête précédente.
20. Il est intéressant de connaître les aliments correspondant au genre d'un autre. Par exemple, on voudrait savoir quels aliments ont le même genre que la pêche melba (tri par famille d'aliments). En une seule requête, donnez le numéro et le nom des aliments qui ont le même genre que la « Pêche melba » (aliment numéro 39401).
21. Toujours pour tester de la cohérence de la base, quels sont les genres qui n'apparaissent jamais dans les aliments ?
22. Vérifiez qu'il existe bien 57 constituants dans la base.
23. Donnez tous les constituants classés par numéro croissant.

La composition de chaque aliment est donnée par une quantité (grammes, ml, etc.) de constituants. Par exemple, la composition nutritionnelle des « Pâtes alimentaires cuites » est donnée par la requête suivante :

```
SELECT Aliment.nomFrAliment, CompNutri.*, Constituant.origineFrConst
FROM Aliment, CompNutri, Constituant
WHERE Aliment.numAliment = CompNutri.numAliment
      AND Constituant.numConst = CompNutri.numConst
      AND Aliment.nomFrAliment = "Pâtes alimentaires cuites"
```

nomFrAliment	numAliment	numConst	valNutri	valMinNutri	valMaxNutri	nbEchantNutri	ccEurNutri	origineFrConst
Pâtes alimentaires cuites	9811	327	641	NULL	NULL	NULL	A	Energie, Règlement UE 1169/2011 (kJ/100g)
Pâtes alimentaires cuites	9811	328	151	NULL	NULL	NULL	A	Energie, Règlement UE 1169/2011 (kcal/100g)
Pâtes alimentaires cuites	9811	332	633	NULL	NULL	NULL	A	Energie, N x facteur Jones, avec fibres (kJ/100g)
Pâtes alimentaires cuites	9811	333	150	NULL	NULL	NULL	A	Energie, N x facteur Jones, avec fibres (kcal/100g...
Pâtes alimentaires cuites	9811	400	62,1	54.3	82.2	4	D	Eau (g/100g)
Pâtes alimentaires cuites	9811	10110	115	0	344	2	D	Sodium (mg/100g)
Pâtes alimentaires cuites	9811	10120	24,5	14	28.4	1	D	Magnésium (mg/100g)
Pâtes alimentaires cuites	9811	10150	70	37	87	1	D	Phosphore (mg/100g)
Pâtes alimentaires cuites	9811	10190	51,7	24	83	1	D	Potassium (mg/100g)
Pâtes alimentaires cuites	9811	10200	16,5	6	17.7	1	D	Calcium (mg/100g)
Pâtes alimentaires cuites	9811	10251	< 0,335	0.2	0.574	1	D	Manganèse (mg/100g)
Pâtes alimentaires cuites	9811	10260	< 0,65	0.5	2.5	1	D	Fer (mg/100g)

24. Est-ce que tous les constituants sont remplis pour chaque aliment ?
25. Fournissez les constituants qui sont dosés en mg/100g (indiqué dans le champ origineFrConst).
26. Comptez le nombre de constituants dosés en mg/100g (dans l'origine française puis dans l'origine anglaise).

L'application doit permettre, à terme, de scanner un code barre, un *flashcode* ou un *QR code* et de fournir la composition nutritionnelle de l'aliment qui s'en rapproche le plus dans la base.

27. On a donc besoin de récupérer la composition nutritionnelle d'un aliment fourni dans l'interface graphique. Donner, par exemple, la composition nutritionnelle du "Cidre doux".
28. Vous avez trouvé dans le code SQL de votre collègue la requête suivante. Que fait-elle ?

```

1 SELECT numAliment
2 FROM CompNutri
3 GROUP BY numAliment
4 HAVING COUNT(*) < (SELECT COUNT(*)
5                    FROM Constituant)

```

29. Votre graphiste souhaite produire une interface de création d'aliment. Que doit-on lui indiquer d'ajouter dans la fenêtre de création ?
30. En utilisant la requête précédente, créez une vue nommée `LesAlimentsRenseignes` où seront stockés l'ensemble des numéros d'aliments qui ont les 57 constituants renseignés.
31. A partir de la vue, vérifier que tous les aliments de la base ont bien l'ensemble des 57 constituants nutritionnels renseignés.

La dernière étape est de vérifier les compositions nutritionnelles des aliments les plus riches en certains composants. Par exemple, le diététicien souhaite savoir quels sont les aliments qui contiennent le plus de Fer, le plus de Sucre, ou le moins de sel. Il souhaite aussi connaître les éléments les plus énergétiques.

32. Donnez les aliments ayant du "Sucre" dans leur composition. La requête retournera tous les attributs d'aliment ainsi que la valeur nutritionnelle (`valMaxNutri`) du sucre.
33. Donnez l'aliment contenant le plus de Fer.
34. Quel est l'aliment le plus énergétique (valeur maximale de `valNutri`, colonne du Règlement UE 1169/2011 (kJ/100g)) ?

35. Donnez la valeur maximale de protéines (en g/100g) des aliments qui contiennent le mot 'viande'.

36. Conclure sur l'usage de la base de données : est-ce que l'ensemble des genres, des aliments et de leur composition nutritionnelle est renseignée ? La base est-elle cohérente pour le projet que l'on souhaite réaliser ?