

Bases de Données Avancées- Le Projet

Objectifs

L'objectif principal du projet BD est de mettre en œuvre les outils enseignés dans le cours et maîtriser les phases de mise en place d'une BDR de la conception à l'interrogation et la mise à jour. Ce projet doit vous permettre de :

- ✓ Se familiariser avec les étapes de mise en place d'une base de données,
- ✓ Démontrer vos capacités de choix des solutions de modélisation,
- ✓ S'exercer à la mise en œuvre des outils vus en cours,
- ✓ Comprendre l'utilité de SQL et développer vos compétences dans l'écriture des requêtes simples et avancées de type LDD, LMD, LID et LCD.

Description et démarche du projet

L'idée de ce projet est de migrer des données complexes disponibles comme des données csv dans des sites de données publiques (Open Data) vers une base de données structurée générée par un SGBDR (MySQL, MariaDB, etc). La base est ensuite enrichie avec un ensemble de requêtes SQL permettant d'exploiter, d'interroger et d'analyser ces données.

La réalisation de ce projet nécessite le passage par plusieurs phases :

I. **Choix du sujet (voir Annexe 1) :**

- a. Consulter les sites de données ouvertes et identifier une thématique qui vous intéresse ;
- b. Dans la thématique choisie, consulter les fichiers de données ouvertes afin de sélectionner un jeu de données qui peut être converti en base de données. Il est préférable que le jeu de données soit riche dans les deux dimensions (lignes, colonnes) mais avec une taille raisonnable pour pouvoir l'importer. Il vous est possible de supprimer des lignes si le fichier est volumineux.

II. **Conception de la base (voir Annexe 2) :**

L'objectif de cette phase est de retrouver le modèle logique qui peut correspondre à la base de données relationnelle à créer avec le jeu de données sélectionné. Pour ce faire, il vous est conseillé de :

- a. Lire les informations relatives à ce jeu de données dans le site ;
- b. Etudier les champs dans le fichier (significations et types des données) ;
- c. Imaginer une application (ou problématique à résoudre) associé à ce jeu de données ;
- d. Regrouper les champs par entité (les champs caractérisant un objet unique concret ou abstrait) et ajouter des identifiants pour ces entités s'ils n'y sont pas définis par défaut ;

- e. Retrouver les liens logiques entre les différentes entités en identifiant les entités pères et les entités fils.
- f. Modéliser les liens logiques entre les tables en ajoutant des clés externes dans les tables fils ou des tables de jonction.

III. **Création de la base (voir Annexe 3) :**

- a. Selon l'étude réalisée à l'étape 2, proposer un schéma relationnel complet qui définit :
 - Les tables, leurs champs et un domaine (type) pour chaque champ,
 - Le champ clé primaire dans chaque table,
 - Les contraintes de référence : clés externes et leurs références
- b. Créer les tables sous le SGBDR choisi avec des requêtes SQL LDD.
- c. Peupler les tables en important les données csv ou Excel vers les tables créées.

IV. **Interrogation simple et mise à jour de la base :**

- a. Concevoir un ensemble de requêtes d'interrogation basique de la base portant sur une ou plusieurs tables et les implémenter en SQL (au moins 12 requêtes réparties entre requêtes simples, requêtes avec jointure requêtes avec sous-requêtes simples) ;
- b. Ecrire des requêtes SQL permettant d'ajouter quelques lignes dans certaines tables de la base (au moins 2 requêtes) ;
- c. Concevoir un ensemble de requêtes de mise à jour des données portant sur une ou plusieurs tables et les implémenter en SQL (au moins 2 requêtes).

V. **Analyse et synthèse de la base :**

- a. Concevoir un ensemble de requêtes de synthèse (avec des fonctions d'agrégation) et les implémenter en SQL (au moins 10 requêtes).
- b. Concevoir des requêtes composées avec des opérateurs ensemblistes et/ou des sous-requêtes.

VI. **Interrogation avancée de la base**

Concevoir des requêtes complexes nécessitant l'utilisation des vues, des CTE, des jointures externes, fonctions de domaines, etc.

VII. **Optimisation des requêtes SQL**

Pour optimiser vos requêtes, ajouter des index sur les champs candidats, optimiser l'écriture de certaines requêtes.

Organisation et Evaluation du projet

Le projet est à faire en **binôme** (attention, la notation peut individualisée). Le dépôt du projet se fera dans l'espace dédié sur « cours en ligne » et doit comporter au moins les éléments suivants :

- Un fichier texte contenant une description brève du projet et le modèle relationnel établi dans la phase 2. Le document texte doit contenir aussi la liste des requêtes SQL (afin de comprendre les résultats affichés à la suite de l'exécution des scripts associés).

- Le fichier script qui a permis de créer la base avec toutes les composantes demandées (tables peuplées).
- Le fichier script contenant les requêtes SQL écrites dans les phases 4, 5 et 6.

Critères d'évaluation :

- Respect des consignes du projet ;
- Degré de réemploi des connaissances acquises en cours ;
- Intérêt des données et opérationnalité (facilité d'utilisation, absence de bugs, ...) ;
- Originalité (utilisation de fonctionnalités non vues en cours, imagination).

ANNEXES

Annexe 1 : Choisir des données Open Data

Les données ouvertes ou Open Data sont des données auxquelles l'accès est totalement public et libre de droit. Veuillez consulter ce site pour une définition plus détaillée : <https://www.lebigdata.fr/open-data-definition>.

Certaines données sont extraites d'une Bases de Données d'une application dans un organisme publique ou privée et exportée dans un fichier csv avant d'être publiées. Il est donc possible de re-converter ces données en BDR. En consultant les sites ci-dessous, il vous est recommandé de choisir des données ou certaines entités (tables) paraissent évidentes. Il faudrait choisir un fichier riche en données en colonnes (champs) afin que vous puissiez les départager en tables. Il vous est possible choisir deux fichiers (ou plus) traitant le même sujet.

<https://www.data.gouv.fr/>

<https://opendata.paris.fr/>

<https://www.europeandataportal.eu/fr>

<https://ressources.data.sncf.com/>

<https://data.worldbank.org/>

<https://data.europa.eu/fr>

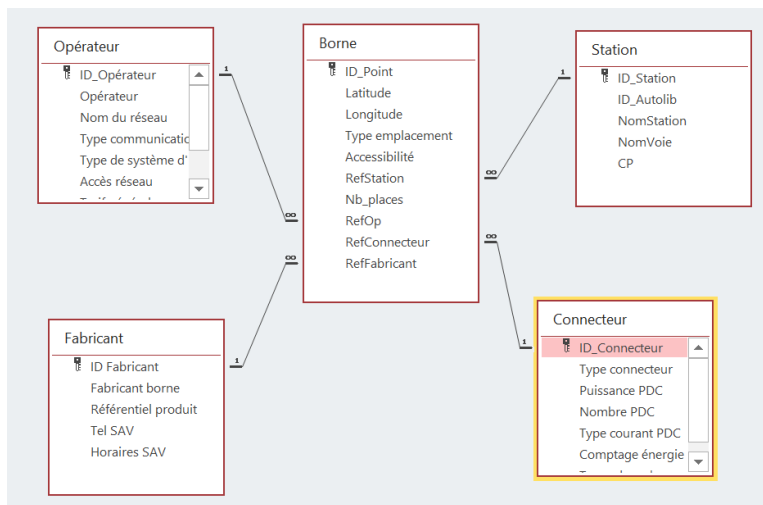
Annexe 2 : Conception de la base

La conception du modèle relationnel se basera essentiellement sur les informations fournies sur le jeu de données et la liste des champs dans le(s) fichier(s) des données. Dans certains cas, il est facile de regrouper les attributs par entité à travers la liste des champs. Si les données sont aussi fournies au format json, les tables peuvent correspondre au différents objets dans le document. Ci-dessous un exemple pour jeu de données intitulé « bornes-de-recharge-pour-vehicules-electriques » disponible dans le site *Opendata.Paris.fr*.

ID Station	ID Autolib	Nom de la Station	Nom voie	Code postal	Ville	Pays	Latitude	Longitude	Type localisation station	Type stationnement	emplacement borne	Accessibilité lieu	Nombre de places
------------	------------	-------------------	----------	-------------	-------	------	----------	-----------	---------------------------	--------------------	-------------------	--------------------	------------------

Opérateur	Nom du réseau	communication supervisée	système d'identification	Accès réseau	Tarif général	Tarif autopartage	Mode de paiement	Puissance abonnement ENEDIS	Types de recharge	Puissance PDC	Nombre PDC	Type connecteur	Type courant PDC	Comptage énergie	Fabricant borne	Référentiel produit	geo_shape	geo_point_2d
-----------	---------------	--------------------------	--------------------------	--------------	---------------	-------------------	------------------	-----------------------------	-------------------	---------------	------------	-----------------	------------------	------------------	-----------------	---------------------	-----------	--------------

Après simplification des données (suppression de quelques champs) le modèle logique pour ces données peut être comme suit :



Annexe 3 : Transformation des données csv en tables

Pour transformer votre fichier csv en tables, il faut d'abord étudier les données et identifier les différentes entités (tables) possibles. La séparation des données correspond à chaque table peut se faire :

1. Soit manuellement : copier les données relatives à chaque table dans une feuille Excel séparée tout en supprimant les doublons. N'oubliez pas de garder le lien entre les données (ou de le créer), ce lien permettra de définir les clés étrangères dans votre base ;
2. Soit en important tout le fichier csv dans une table SQL temporaire, et écrire par la suite des requêtes SQL permettant d'extraire les données correspondant à chaque table après l'avoir créé. Cette deuxième technique peut être plus rapide du fait qu'elle permet une insertion automatique d'identifiant si une entité n'en dispose pas.

Pour importer les données csv dans une table, plusieurs solutions sont possibles dont on cite :

- Transformer une feuille csv en un script SQL contenant une série de requêtes d'insertion. Un outil gratuit propose ce service dans le site : <https://www.site24x7.com/fr/tools/csv-en-sql.html>. Exécuter ensuite le script généré sur Workbench.
- Importer votre fichier csv dans une table en utilisant l'outil d'importation de l'éditeur Workbench de MySQL (sélectionner la table à peupler dans l'explorateur d'objets et activer « Table Data Import Wizard »).

