

Méthodologie

Projet 3 - SQL

Méthodologie

Découvrir les différents types de données

Type de donnée

SQL

Type de variable	Caractéristique
INT pour <i>integer</i>	Nombre entier
FLOAT	Nombre décimal
VARCHAR	Donnée textuelle / chaîne de caractère

Taille / contrainte de variable	Caractéristique
Exemple : 5	La cellule peut contenir 5 caractères max

Clé	Caractéristique
Primaire	Identifiant unique pour chaque donnée
Etrangère (secondaire)	Fait référence à la clé primaire d'une autre table. Permet de faire la jointure avec l'autre table

	Nom des colonnes	Type de données	Taille	Clé	Description
CONTRAT.CSV	Contrat_ID	INT		Clé primaire	Id unique pour les contrats
	No_voie	INT			Numéro dans la voie pour l'adresse du logement assuré
	B_T_Q	VARCHAR	1		Indicateur éventuel de répétition pour l'adresse du logement assuré sur un caractère
	Type_de_voie	VARCHAR	5		Type de voie pour l'adresse du logement assuré: rue, av (Avenue), rte (Route), ...
	Voie	VARCHAR	100		Libellé de la voie pour l'adresse du logement assuré
	Code_dep_code_commune	VARCHAR	10	Clé étrangère	Concaténation du code département et code commune pour avoir une clé unique
	Code_postal	INT	5		Code postal pour l'adresse du logement assuré
	Surface	INT	4		Surface du logement
	Type_local	VARCHAR	15		Type de logement : appartement ou maison
	Occupation	VARCHAR	15		Type d'occupation : locataire ou propriétaire
	Type de contrat	VARCHAR	20		Type de contrat : mise en location, résidence principale ou secondaire
	Formule	VARCHAR	10		Type de formule : classique ou intégral
	Valeur_declaree_biens	VARCHAR	12		Fourchette de valeur du bien
REGION.CSV	Prix_cotisation_mensuel	INT	4		Prix des cotisations mensuel du bien correspondant
	Code_dep_code_commune	VARCHAR	10	Clé primaire	Concaténation du code département et code commune pour avoir une clé unique
	reg_code	INT	2		Numéro de la région
	reg_nom	VARCHAR	25		Nom de la région
	aca_nom	VARCHAR	25		Nom de la ville
	dep_nom	VARCHAR	40		Nom du département
	com_nom_maj_court	VARCHAR	40		Nom de la commune
	dep_code	VARCHAR	3		Numéro du département
	dep_nom_num	VARCHAR	40		Nom du département

Méthodologie

Découvrez la conception de schéma relationnel

Schéma relationnel

SQL



SQL Power architect

CONTRAT	
Contrat_id: VARCHAR	NOT NULL [PK]
Code_dep_code_commune: VARCHAR	NOT NULL [FK]
Surface: INTEGER	NOT NULL
Code_postal: INTEGER	NOT NULL
Voie: VARCHAR	NOT NULL
Type_de_voie: VARCHAR	NOT NULL
B_T_Q: VARCHAR	NOT NULL
Type_local: VARCHAR	NOT NULL
Occupation: VARCHAR	NOT NULL
Type_contrat: VARCHAR	NOT NULL
Formule: VARCHAR	NOT NULL
Valeur_declaree_biens: VARCHAR	NOT NULL
Prix_cotisation_mensuel: INTEGER	NOT NULL

REGION	
Code_dep_code_commune: VARCHAR	NOT NULL [PK]
reg_code: INTEGER	NOT NULL
Nom de la région: VARCHAR	NOT NULL
aca_nom: VARCHAR	NOT NULL
dep_nom: VARCHAR	NOT NULL
com_nom_maj_court: VARCHAR	NOT NULL
dep_code: INTEGER	NOT NULL
dep_nom_num: VARCHAR	NOT NULL

Méthodologie

Découvrez la création et le chargement d'une base de données

Chargement d'une base de données

SQL

Requête SQL	Définition
CREATE DATABASE	Créer une nouvelle databse qui contiendra les 2 tables
SHOW DATABASE	Voir les différentes database présente dans mon environnement
USE DATABASE	Utilisé la database correspondante
CREATE TABLE	Créer une table à l'intérieur de la database sélectionné précédemment

Utilisation de python (VSCode) pour insérer les données du fichier csv dans la table précédemment créé dans SQL.

```
# Fonction pour remplacer valeur vide par Null
def replace_empty_with_null(row):
    return [None if field == '' else field for field in row]

# Ouvrir csv
with open("C:\\Users\\simon\\OneDrive\\Documents\\Openclassrooms - Business Intelligence Analyst\\Projet 3 - sql\\Donnée\\Contrat+(4).csv", mode='r') as file:
    csv_data = csv.reader(file, delimiter=',')

    # Eviter les en-têtes si il y en a
    next(csv_data)

    # Requête SQL pour importer les données (nom de colonne)
    query = f"INSERT IGNORE INTO {table_name} (Contrat_ID, No_voie, B_I_Q, Type_de_voie, Voie, Code_dep_code_commune, Code_postal, Surface, Type_local, Occupation, Type"
    #INSERT IGNORE = MySQL ignorera les erreurs de doublons pour les clés primaires
    # Parcourir chaque ligne pour les insérer
    for row in csv_data:
        # Remplacer chaque valeur vide par des Null pour chaque ligne
        row = replace_empty_with_null(row)

        # Si chaque ligne contient 14 colonnes ALORS importer la ligne dans la BDD
        if len(row) == 14:
            cursor.execute(query, row)
        else:
            print(f"Ligne non importer car nombre de colonne différent : {row}")

    # Importer les données dans la table
    connection.commit()

print("Les données ont été importées.")
```

Méthodologie

Découvrez la rédaction de requêtes SQL

Rédaction de requête

SQL

Requête SQL	Définition
SELECT	Sélectionner toutes les colonnes souhaitées ...
FROM	... depuis la table correspondante.
LEFT JOIN	Joindre une 2 ^e table pour filtrer les données et utiliser les colonnes de cette 2 ^e table. <i>LEFT JOIN = garder uniquement les données identiques à la table de gauche.</i>
WHERE	Filtrer les données selon une condition. <i>Exemple : filtrer les données pour garder uniquement celle appartenant à la région Pays de la Loire.</i>
ORDER BY	Trier selon une colonne <i>ASC = croissant</i> <i>DESC = décroissant</i>
GROUP BY	Grouper les données selon une condition <i>Exemple : grouper les contrats par région.</i>

```
SELECT c.Contrat_ID, c.Surface
FROM table_contrat c
LEFT JOIN table_region r ON r.Code_dep_code_commune = c.Code_dep_code_commune
WHERE r.Nom_commune = 'Caen'
```

Q	Nom_commune varchar	Contrat_ID int	Surface int
	Filter	Filter	Filter
>	CAEN	103791	35
>	CAEN	103792	99
>	CAEN	103793	40
>	CAEN	103794	20

Rédaction de requête

SQL

Quel est le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens ?

```
SELECT Valeur_declaree_biens, COUNT(Contrat_ID) as nb_contrat
FROM table_contrat
GROUP BY Valeur_declaree_biens
ORDER BY nb_contrat DESC
```

* Valeur_declaree_biens varchar(100)	nb_contrat bigint
Filter	Filter
0-25000	22720
25000-50000	6815
50000-100000	696
100000+	104