

# Gestion des données et accès à l'information

## SQL en Python

Olivier Schwander <olivier.schwander@sorbonne-universite.fr>

2023-2024

L'objectif de ce TD est d'accéder à une base SQL à partir d'un programme Python, de façon à extraire des données et à effectuer divers traitements dessus.

Pour des raisons de simplicité, on utilise désormais SQLite, un système de base de données relationnelles qui a la particularité de ne pas s'utiliser avec un mécanisme client-serveur, mais directement comme une bibliothèque au sein d'un programme.

Les données utilisées ici proviennent du portail de données publiques de Météo-France, en particulier du jeu de données *Données SYNOP essentielles OMM* ([https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=90&id\\_rubrique=32](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32)) qui regroupent des observations des principales stations météorologiques françaises. Une partie de ces observations (les stations, les dates et les températures) a été insérée dans une base de données SQLite.

La documentation du module sqlite3 est disponible à l'adresse <https://docs.python.org/3/library/sqlite3.html> pour Python3.

**Remarque:** il n'y a pas de type `date` en SQLite, on utilise donc un simple type `TEXT` pour stocker des dates. Il est possible d'interpréter une colonne contenant des dates comme de vraies dates avec les fonction `date` et `datetime`: c'est utile en particulier pour les comparaisons ou les tris.

## Exercice 1 - *Prise en main*

Téléchargez le fichier `temperature.db` disponible sur la page du cours.

La structure de la base est la suivante:

```
CREATE TABLE stations (  
    id INT,  
    name TEXT,  
    PRIMARY KEY (id)  
);  
CREATE TABLE temperatures (  
    id INT,  
    date TEXT,  
    station INT,  
    temp REAL,  
    PRIMARY KEY (id)  
);
```

### Question 1

Décrivez la structure de la base, sous forme de texte puis de schéma entité-association.

## Exercice 2 - *En Python*

On accède à la base de donnée de la façon suivante

```
import sqlite3
db = sqlite3.connect("temperature.db")
```

### Question 2

Affichez tout le contenu de la base à l'aide du fragment de code suivant:

```
for db in con.iterdump():
    print(line)
```

Commentez.

## Exercice 3 - *Analyse de données*

### Question 3

Affichez toutes les températures relevées par la station ST-PIERRE.

### Question 4

Affichez toutes les températures relevées par la station ST-PIERRE. en janvier 2016.

### Question 5

Affichez toutes les températures relevées en France métropolitaine.

### Question 6

Tracez toutes les températures relevées par la station ST-PIERRE.

```
query = """SELECT temperatures.date, temperatures.temp FROM stations,temperatures
           WHERE stations.id = temperatures.station AND stations.name = 'ST-PIERRE';"""
data = db.execute(query)
```

Attention, data n'est pas une vraie liste:

```
print(type(data))
```

On peut quand même le parcourir avec une boucle for:

```
for line in data:
    print(line)
```

Ou le convertir en une vraie liste:

```
data = list(data)
print(type(data))
```

Pour l'affichage, on peut se servir de ce qui précède pour appeler directement les fonctions de `matplotlib`. Pour une utilisation statistique, il peut être utile de connaître la fonction `pandas.read_sql_query` qui construit un dataframe à partir d'une requête sql:

```
import pandas as pd
df = pd.read_sql_query(query, db)
print(df.describe())
```

### **Question 7**

Tracez toutes les températures sur un mois il y a un an à la station **ST-PIERRE**.

**Remarque:** les dates doivent être calculées avec Python, et non pas écrites directement.

### **Question 8**

Tracez toutes les températures sur un mois il y a un an, en France métropolitaine.

### **Question 9**

Calculez la moyenne annuelle de chaque station, en utilisant Python pour le calcul de la moyenne.

### **Question 10**

Calculez la moyenne annuelle de chaque station, en utilisant SQL pour le calcul de la moyenne.

### **Question 11**

Comparez les temps d'exécutions pour les deux questions précédentes.

### **Question 12**

Tracez toutes les températures sur un mois il y a un an, en France métropolitaine, en déterminant les stations à partir des informations sur les stations données à l'adresse [https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=90&id\\_rubrique=32](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32).

### **Question 13**

Modifiez la structure de la base pour insérer les coordonnées des stations, puis insérez les coordonnées.