

Chapitre :

Traitement de données en tables



Ce chapitre est à voir comme une première approche des bases de données, qui ne seront pas abordées ici mais seront vues en terminale.

On se rend compte que l'usage de données informatique est de plus en plus important, celles-ci devenant de plus en plus massives (on parle de big data). Elles doivent être stockées, gérées. Des recherches pour cela sont faites lors du projet Apollo dans les années 1960. Un système de gestion de base de données (SGBD), nommé Integrated Data Store, est créé en 1964 par Charles Bachman qui travaillait pour General Electric.

En 1970, Edgar Franck Codd, qui travaillait pour pour IBM, met au point une théorie sur les modèles relationnels permettant le partage de nombreuses données. Il développe alors un SGBD fonctionnant avec le langage SQL (Structured Query Language) qui sera vu en terminale.

Nous allons ici seulement voir le traitement de données en tables, une table étant une liste de p-uplets nommés partageant les mêmes descripteurs. Notre but sera, à l'aide de Python en particulier, de faire des recherches, de faire des tris et des fusions de telles tables.

Un format courant de tables de données est le format **CSV** (Comma-separated values). Dans ce format, les données sont représentées par des valeurs séparées par des virgules ou des points-virgules. Pour décrire une table, on note ses descripteurs (les titres des colonnes) sur une première ligne, séparés donc par exemple par des points-virgules. Chaque ligne suivante contient les différentes données séparées là encore par des points-virgules, en respectant l'ordre des colonnes.

Exemple Considérons ce tableau de données (provenant par exemple d'un logiciel de tableur) :

Prénom	Date de naissance	Couleur de yeux	couleur préférée
Andy	12/05/2001	bleu	orange
Bilal	24/03/2002	brun	vert
Céline	22/01/2002	marron	violet
...

Alors le fichier cvs correspondant (que le tableur peut fournir) contiendra les lignes suivantes :

Prénom ;Date de naissance ;Couleur de yeux ;couleur préférée

Andy ;12/05/2001 ;bleu ;orange

Bilal ;24/03/2002 ;brun ;vert

Céline ;22/01/2002 ;marron ;violet

...

À l'aide de Python, on peut ouvrir un fichier cvs et obtenir une liste de listes contenant ses données :

```
import csv
with open('Donnees.csv') as csvfile:
    data = list(csv.reader(csvfile))
```

La suite dans la fiche d'activité sur les données en tables :

⊗ **Activité** : Données en tables