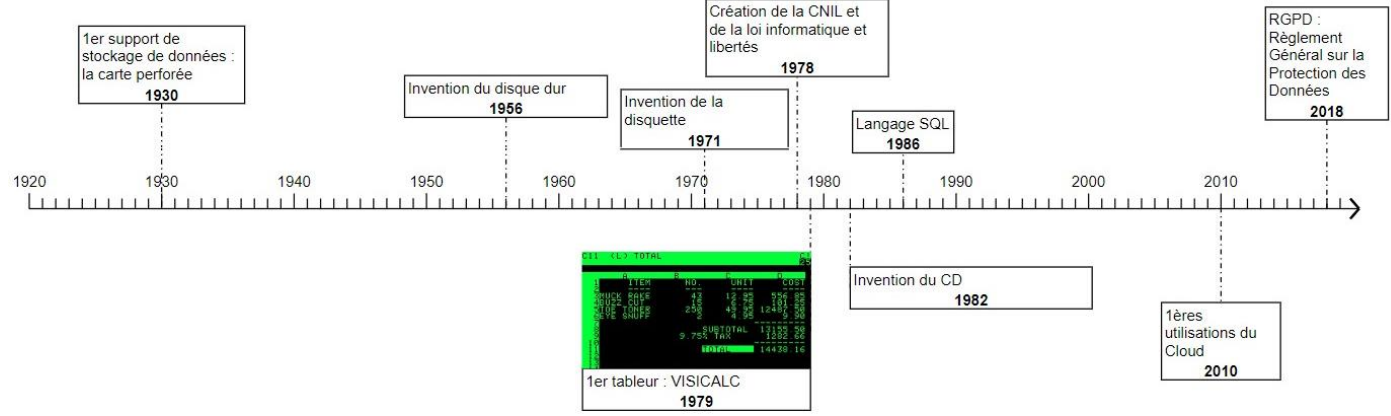




# Synthèse thème 3 : les données structurées et leur traitement

## Repères historiques



## Vocabulaire

### Les données :

**Donnée** : élément se rapportant à un objet, une personne ou un événement

**Donnée personnelle** : donnée identifiant directement (nom, prénom, photo...) ou indirectement (empreinte digitale, ADN, géolocalisation...) une personne physique

**Métadonnée** : donnée particulière qui apporte des informations sur la donnée principale

### Les données structurées :

**Tables de données** : données organisées sous forme de tableaux

**Descripteur** : 1ère ligne d'un tableau permettant de caractériser les données

**Objet** : une ligne contenant une valeur des descripteurs d'une table de données

**Collection** : ensemble des objets partageants les mêmes descripteurs

	A	B	C	D	E
1	Nom	Prénom	Sexe	Date de naissance	Ville de naissance
2	Martin	Gabriel	M	05/03/2005	Bressuire
3	Bernard	Camille	F	17/12/2001	Nantes
4	Thomas	Camille	M	06/02/2003	Poitiers
5	Petit	Alice	F	31/01/2006	Bordeaux
6	Robert	Louise	F	05/03/2005	Paris

Annotations: Blue arrow points to the header row (Descripteurs), green arrow points to a data row (Objet), red arrow points to the entire table (Collection).

## Formats des données structurées

### Format .CSV

```
1 Nom,Prénom,Sexe,Date de naissance,Ville de naissance
2 Martin,Gabriel,M,05/03/2005,Bressuire
3 Bernard,Camille,F,17/12/2001,Nantes
4 Thomas,Camille,M,06/02/2003,Poitiers
5 Petit,Alice,F,31/01/2006,Bordeaux
6 Robert,Louise,F,05/03/2005,Paris
```

Le format **.CSV** (Comma Separated Values) est un format de texte où chaque valeur est généralement séparée par une virgule ou un point virgule, ce qui permet d'enregistrer les données dans un tableau

### Format .XML

```
1 <eleve>
2   <nom>Martin</nom>
3   <prenom>Gabriel</prenom>
4   <sexe>M</sexe>
5   <Date_de_naissance>05/03/2005</Date_de_naissance>
6   <Ville_de_naissance>Bressuire</Ville_de_naissance>
7 </eleve>
8 <eleve>
9   <nom>Bernard</nom>
10  <prenom>Camille</prenom>
11  <sexe>F</sexe>
12  <Date_de_naissance>17/12/2001</Date_de_naissance>
13  <Ville_de_naissance>Nantes</Ville_de_naissance>
14 </eleve>
```

Le format **.XML** (Extensible Markup Language) est un format utilisé pour l'échange automatisé entre systèmes d'informations

### Format .JSON

```
1 {
2   "Nom": "Martin",
3   "Prénom": "Gabriel",
4   "Sexe": "M",
5   "Date de naissance": "05/03/2005",
6   "Ville de naissance": "Bressuire"
7 }
8 {
9   "Nom": "Bernard",
10  "Prénom": "Camille",
11  "Sexe": "F",
12  "Date de naissance": "17/12/2001",
13  "Ville de naissance": "Nantes"
14 }
```

Le format **.JSON** (JavaScript Objet Notation) est un format de données textuelles dérivé du langage JavaScript

# Tri, filtrage et calcul de données sous tableur

Il existe 3 principales opérations sur un tableau :

- Le tri dans l'ordre **croissant** ou **décroissant** suivant une colonne ;
- La **projection** sur une colonne ;
- Le **filtrage** suivant une condition fixée.

Nom	Prénom	Age
Minot	Alban	42
Panama	Rodrigue	27
Pascal	Blaise	78

## Exemples :

Trier le tableau dans l'ordre décroissant de la colonne « Age » :

Nom	Prénom	Age
Pascal	Blaise	78
Minot	Alban	42
Panama	Rodrigue	27

Projeter sur la colonne « Prénom » :

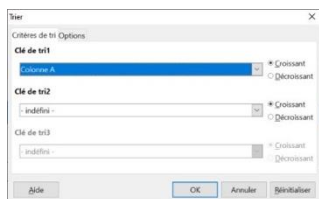
Prénom
Blaise
Alban
Rodrigue

Filtrer avec la condition « Nom commençant par P » :

Nom	Prénom	Age
Pascal	Blaise	78
Panama	Rodrigue	27

## Le tri :

### Menu Données / Trier



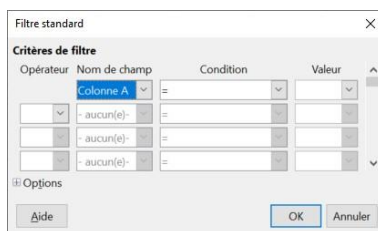
## Les filtres :

### Menu Données / AutoFiltre

	A	B	C	D	E
1	Nom	Prénom	Sexe	Date de naissance	Ville de naissance
2	Robert	Louise	F	05/03/05	Paris
3	Martin	Gabriel	M	05/03/05	Bressuire
4	Bernard	Camille	F	17/12/01	Nantes
5	Thomas	Camille	M	06/02/03	Poitiers
6	Petit	Alice	F	31/01/06	Bordeaux

Une liste déroulante par descripteur pour plusieurs possibilités de tri

### Menu Données / Filtre / Filtre standard



## Les calculs :

F1	A	B	C	D	E
1	Nom	Prénom	Sexe	Date de naissance	Ville de naissance
2	Bernard	Camille	F	17/12/01	Nantes
3	Martin	Gabriel	M	05/03/05	Bressuire
4	Petit	Alice	F	31/01/06	Bordeaux
5	Robert	Louise	F	05/03/05	Paris
6	Thomas	Camille	M	06/02/03	Poitiers



## Quelques exemples de fonctions :

Fonctions	Actions
=NBVAL(A:A)	Compter le nombre de données dans une plage de valeurs (ici, la colonne A)
=NB.SI(D2:D6;"<01/01/05")	Compter le nombre de données répondant à un critère (ici, nombre de dates de la plage D2 à D6 antérieures au 1 <sup>er</sup> janvier 2005)
=NB.SI.ENS(C2:C6;"F";B2:B6;"Camille")	Compter le nombre de données répondant à plusieurs critères (ici, nombre de fille de la plage C2 à C6 dont le prénom, plage B2 à B6, est Camille)

# Données sous Python

## Type de variable « tableau » :

Un tableau est représenté en Python par une variable de type **list**, avec des **crochets**.

Ex: tab = [2, 3, 7, 10, 22]

```
>>> tab[0]          >>> tab[3]
```

```
2                    10
```

Indices	0	1	2	3	4
Valeurs	2	3	7	10	22

Un tableau à double entrée est représenté en Python par un **tableau de tableaux**.

Ex: tab2 = [[7, 15, 2], [7, 10, 12], [16, 7, 0], [19, 14, 27]]

```
>>> tab2[0][2]      >>> tab[2][0]
```

```
2                    16
```

	0	1	2
0	7	15	2
1	7	10	12
2	16	7	0
3	19	14	27

On obtient le **nombre d'éléments** d'une liste Python avec **len()**.

Ex: >>> len(tab)

```
5
```

>>> len(tab2)

```
4 # tab2 contient 4 listes !
```

# Le Cloud

Le **cloud** (nuage en français) permet l'accès en ligne à des applications, un espace de stockage et d'autres ressources. Les données, au lieu d'être stockées sur le disque dur des ordinateurs ou sur des clés USB, sont alors accessibles par Internet (**centres de données ou data center**). Il est également possible d'utiliser des logiciels collaboratifs à distance.

Le cloud est actuellement utilisé massivement par les entreprises et les individus.