

Thème 4 : Les données structurées

TP 2 : Le stockage des données

2.1 Les formats de données

Les données sont stockées sur des fichiers qui ont différents formats.

→ Allez sur le site http://nrj.renouvelables.free.fr/sntdata/co/grain_format.html puis complétez le tableau suivant :

Formats de type texte	Formats de type binaire
CSV	ODS
XML	XLS
JSON	XLSX
vCard	DBF

Exercice : Reconnaître les différents formats

→ Voici une table codée sous trois formats de type texte différents. Observez attentivement :

CSV	XML	JSON
Nom, Prénom, Profession, Décès Hugo, Victor, Écrivain, 1885 Baudelaire, Charles, Écrivain, 1867	<Auteurs> <Auteur> <Nom>Hugo</Nom> <Prénom>Victor</Prénom> <Décès>1885</Décès> </Auteur> <Auteur> <Nom>Baudelaire</Nom> <Prénom>Charles</Prénom> <Décès>1867</Décès> </Auteur> </Auteurs>	[{ "Nom" : "Hugo", "Prénom" : "Victor", "Décès" : 1885 }, { "Nom" : "Baudelaire", "Prénom" : "Charles", "Décès" : 1867 }]

→ Décrivez les principaux aspects du format CSV :

La première ligne contient les descripteurs, ensuite, chaque ligne correspond à un objet. Les valeurs sont séparées par des virgules.

→ Décrivez les principaux aspects du format XML :

Les descripteurs et les valeurs sont délimités par des balises, comme dans le langage HTML.

→ Décrivez les principaux aspects du format JSON :

Les descripteurs et les valeurs sont délimités par des crochets, des accolades, des double-points et des guillemets, comme dans le langage CSS.

2.2 Les supports de stockage

• Les fichiers de données sont stockés sur des **supports de stockage** :

→ Citez deux supports de stockage internes :

Disque dur, SSD (Solid State Drive, c'est le stockage utilisé dans vos ordinateurs fournis par la Région)

→ Citez deux supports de stockage externes :

Disque dur, SSD, clé USB, carte SD, CD-ROM, DVD-ROM, disquette...

Il faut réaliser des sauvegardes car ces supports peuvent subir des dommages entraînant des altérations ou des destructions de données.

• Les supports de stockage peuvent être locaux ou distants : **le cloud**.

→ Regardez la vidéo située à l'adresse suivante jusqu'à 2'00 :

https://www.youtube.com/watch?v=5YawCCUxa_E

→ Donnez une définition du cloud :

Le Cloud est un ensemble d'ordinateurs rassemblés dans des fermes à serveurs, pas nécessairement en France. On peut les utiliser aujourd'hui car le prix du stockage a beaucoup baissé et les connexions sont rapides avec la fibre.

• Attention, le cloud c'est très pratique... mais n'oubliez pas qu'il s'agit en réalité de l'ordinateur d'un autre ! Cela veut dire que vos données sont **copiées** ailleurs dans le monde et que **vous n'avez plus de contrôle total sur ces données**. Il vaut mieux privilégier les sauvegardes locales pour les données les plus sensibles.

• Les grandes bases de données sont stockées dans des fermes à serveurs (= *data centers*), des centres de données possédant des serveurs dédiés (des machines très puissantes avec une importante capacité de stockage sur disques). Les quantités de données stockées sont de 10^{21} (zettaoctet) voire 10^{24} (yottaoctet). Ces *data centers* fonctionnent en permanence et sont connectés à internet.

3. Quel accès aux données ?

Certaines données sont dites ouvertes (*OpenData*) : leurs producteurs considèrent qu'il s'agit d'un bien commun, qui doit être accessible à tous.

Allez sur le site <https://www.data.gouv.fr/fr/>

→ Choisissez un exemple de données ouvertes :

Nombre de personnes vaccinées dans un département, horaires de bus, statistiques sur le nombre d'entrées à la piscine municipale...

→ Pourquoi ces données sont-elles accessibles à tous ?

Parce qu'elles présentent un intérêt pour le grand public et ne sont pas confidentielles.

→ Quels usages peut-on en faire ?

Elles peuvent servir à réaliser des recherches, des études, un aménagement urbain, etc.

- Mais en parallèle se développe un **marché des données**, dans lequel des entreprises collectent et revendent des données sans transparence pour les usagers.