

# LES DONNÉES STRUCTURÉES

Très fortement inspiré de l'excellent site InfoForAll

## I Stocker une donnée en informatique

### Connaissance n°1 Le bit

L'unité de base du stockage de l'information en informatique est le **bit**. Il s'agit d'une case mémoire qui ne possède que deux états :

Vrai ou Faux, Haut ou Bas ou encore plus simple : 0 et 1 .

### Remarque n°1.

Si on prend deux bits, on aura alors 4 possibilités différentes :  $2^2$

00 01 10 11

Si on prend trois bits, on aura alors 8 possibilités différentes :  $2^3$

000 001 010 011 100 101 110 111

etc...

### Connaissance n°2 L'octet

En réalité, on ne transmet jamais un seul bit à la fois. On transmet toujours les bits par groupe de 8 bits consécutifs : c'est ce qu'on nomme un **octet**. Puisqu'on a 8 bits, on a alors  $2^8$  possibilités. Soit 256 nombres différents allant de 0 à 255.

0 0 0 0 0 0 0 0 (première possibilité : 0)

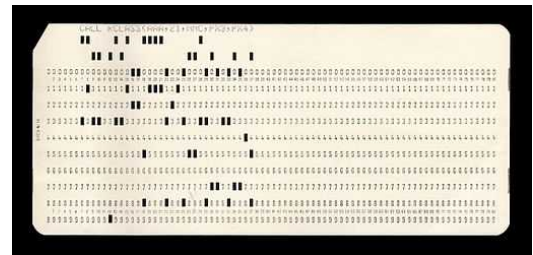
0 0 0 0 0 0 0 1 (deuxième possibilité : 1)

⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮

1 1 1 1 1 1 1 1 (dernière possibilité : 255)

### Connaissance n°3 Les débuts

Au début de l'utilisation des systèmes automatisés, on utilisait des systèmes mécaniques et les données étaient codées sous forme de cartes perforées : on voyait bien les 0 et les 1 !



Carte perforée (Mutatis mutandis CC BY-SA sur Wikimedia)

Puis vint un autre système de stockage, plus rapide mais toujours fortement mécanique : la bande magnétique.

L'avantage par rapport aux fiches ? Plus de vitesse et la possibilité de modifier l'enregistrement sans sortir du scotch !



Bande magnétique (CC BY-SA sur Wikimedia)

En 1956, la société IBM crée un premier disque dur de 5 Mo. L'ensemble mesurait alors

1,52 x 1,73 x 0,74 m !

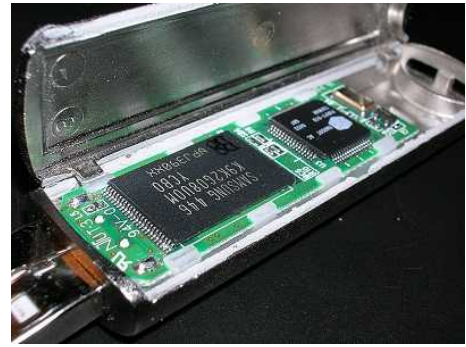
Le coût ? Environ 10 000 \$ le Mo !



L'IBM 3330 (source <https://www.cs.auckland.ac.nz>)

Dans les années 1984 sont apparus des systèmes de stockage de masse purement électroniques, sans pièce en mouvement : la carte SD et la clé USB.

Ce type de technologie remplace maintenant de plus en plus les disques durs classiques. Il s'agit des fameux disques durs SSD.



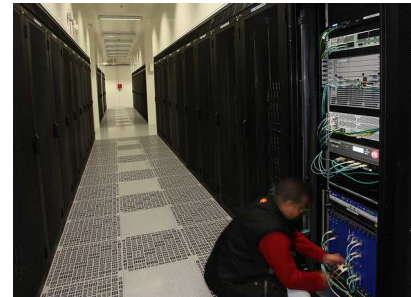
Intérieur d'une clé USB (image libre de droit, Wikipedia)

En 2007, le stockage a même l'air d'être devenu immatériel : on passe au stockage dans le "Cloud" avec Dropbox par exemple.



(image licence Pixabay, Image par Pete)

Mais en réalité, sous cette belle image de nuage tout mignon, se cache des Data Centers, c'est à dire un entassement d'ordinateurs. Et ils chauffent, quoi qu'on en dise...



(image récupérée sur le site du lycée Brassens)

### Remarque n°2.

les prix ont baissé de façon drastique :

- en 1958 : 10000 \$ le Mo
- en 2000 : 0.02 \$ le Mo...
- en 2010 : 0.08 \$ le Go, 8 cents pour mille fois plus de mémoire !
- et ça continue à baisser...

### Connaissance n°5

#### Les sous-unités de l'octet

Sous-unité	Valeur en octet	Equivalent au stockage
Un kilo (k) représente $10^3$	1000	Un petit texte
Un Méga (M) représente $10^6$	1000000	un livre ou une image HD
Un Giga (G) représente $10^9$	1000000000	un film
Un Téra (T) représente $10^{12}$	1 puis douze zéros	
Un Péta (P) représente $10^{15}$	1 puis quinze zéros	capacité mémoire d'un cerceau humain selon certaines études
Un Exa (E) représente $10^{18}$	1 puis dix-huit zéros	
Un Zetta (E) représente $10^{21}$	1 puis vingt-et-un zéros	Production annuelle d'informations par Internet

### Connaissance n°6

En réalité, on ne stocke pas juste des données.

Donner un rapport de 30 000 pages écrits à un système informatique ne sert pas à grand chose. Pour être efficace, il faut fournir au système les informations dans un format bien défini.

C'est ce qu'on nomme des **données structurées**.

### Connaissance n°7 Les données structurées

Une **donnée structurée** est une donnée présentée selon un format connu et fixe.

Il peut s'agir par exemple d'un tableau à deux entrées.

#### Exemple n°1.

Identifiant	Type	Utilisation	Couleur
0	Voiture	Sport	Rouge
1	Hélicoptère	Tourisme	Marron
2	Voiture	Sport	Rouge

#### Remarque n°3.

Cette fois, on voit qu'on pourra demander facilement (sans IA) à l'ordinateur de renvoyer les couleurs des véhicules : il suffit de trouver la bonne colonne. C'est beaucoup plus facile.

### Connaissance n°8

- Les colonnes sont nommées les **descripteurs**.

Toutes les données présentes dans ce tableau partagent les mêmes descripteurs et les fournissent dans le même ordre :

ID - Type - Catégorie - Couleur.

- Chaque ligne se nomme un **enregistrement**

La seule ligne qui ne soit pas un enregistrement est la toute première, celle qui est grisée. Il s'agit de l'en-tête (header en anglais). Elle permet de connaître le nom et l'ordre des descripteurs.

- La **collection** est le nom donné à l'ensemble des enregistrements, donc au tableau.

#### Exemple n°2.

Collection	Attributs			
	Identifiant	Type	Utilisation	Couleur
Enregistrements	0	Voiture	Sport	Rouge
	1	Hélicoptère	Tourisme	Marron
	2	Voiture	Sport	Rouge