

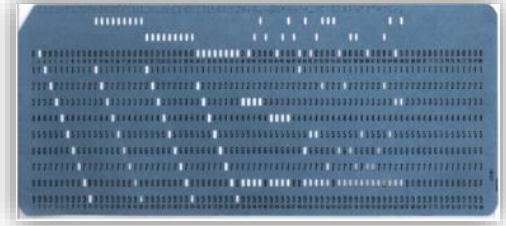
Synthèse : les données structurées



- **1725 :** le Français Basile Bouchon invente un système de rubans perforés pour automatiser un métier à tisser. Les trous, ou leur absence, déterminent les mouvements des pièces mécaniques du métier : c'est en quelque sorte un système binaire (« trou » ou « pas trou »).

- **1728 :** Jean-Baptiste Falcon, l'assistant de Bouchon, remplace les rubans par des cartes. Ces cartes perforées, premier support de stockage de données, seront en usage jusque dans

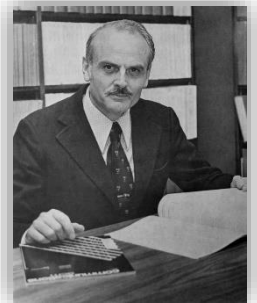
les années 1960.



- **1956 :** Une équipe d'IBM menée par l'Américain Reynold Johnson invente le premier disque dur, l'IBM 350. Constitué de 50 disques de 610 mm de diamètre, il peut contenir 3,75 mégaoctets de données (à peine la taille d'une chanson en mp3) et pèse environ une tonne ! Aujourd'hui, un disque dur de 10 téraoctets (1 Téra = 1 million de mégas) tient dans une main.

- **1970 :** Le développement des supports de stockage entraîne mécaniquement une hausse de la quantité de données disponible.

Mais comment retrouver et traiter facilement des données dans ces conditions ? C'est à cette fin que l'Anglais Edgar Codd met au point en 1970 le modèle relationnel : une représentation logique et rationnelle des données permettant de les rendre exploitables.



- **1979 :** Les tableurs, apparus dans les années 60, permettent de visualiser des données présentées sous forme de tableaux. En 1979, Visicalc est le premier à fonctionner sur un ordinateur individuel, l'Apple II. Il sert à la comptabilité et contribue à faire des ordinateurs un outil de travail dans les entreprises. 700 000 copies sont vendues en 6 ans.



- **1984 :** Grâce au Japonais Fujio Masuoka, la mémoire flash est mise sur le marché par Toshiba en 1984. Ce type de mémoire rapide et sans pièce mécanique est utilisée dans les clefs USB, les appareils photo, etc. et remplace progressivement les disques durs.



- **En 1997 :** C'est en 1997 que le terme Big Data apparaît pour caractériser de grands ensembles de données de nature multiple (textes, images, sons, etc.) et en constante évolution. Par exemple, les données des utilisateurs d'un réseau social constituent du Big Data. Elles peuvent être exploitées pour des raisons commerciales (publicités ciblées, etc.) ou encore scientifiques (études sociologiques, etc.).

- **2009 :** L'Open Data prône le libre accès des données. Ainsi en 2009, le président Obama lance l'Open Government Initiative dans le but de rendre l'État américain plus transparent. En 2013, le G8 (les 8 plus grandes puissances économiques du monde) signe une charte dans ce sens. Les résultats sont à considérer d'un œil critique. En effet, toutes les données ne sont pas librement accessibles, mais seulement certaines, et leur flux peut diminuer si ceux qui les génèrent estiment certains sujets trop sensibles.



REPERES HISTORIQUES



1. LES DONNEES PERSONNELLES

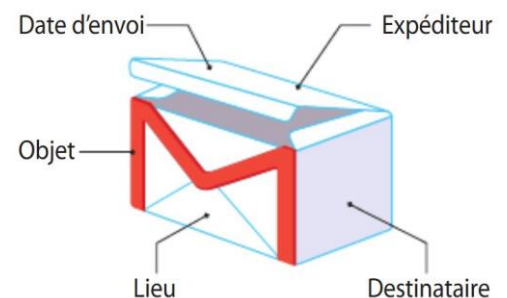
Une **donnée** est un élément se rapportant à un **objet**, une personne ou un événement. Une **donnée personnelle** est une donnée **identifiant** directement ou indirectement une personne physique.

Exemple : Nom, numéro de téléphone, photographie, date de naissance, empreinte digitale, etc...

2. LES METADONNEES

Une **métadonnée** est une donnée particulière qui apporte des informations sur la donnée principale.

Exemple : Pour un fichier de musique, la donnée principale est la chanson et le nom de l'artiste ou le nom de l'album sont des métadonnées.



Les données structurées

1. LES TABLES DE DONNEES

On organise les données sous la forme d'un tableau appelé « **table de données** ». Une liste de **descripteurs** permet de caractériser les données. Un objet est un élément d'une table. Une collection regroupe des objets partageant les mêmes descripteurs. Les valeurs de tous les descripteurs d'un objet sont précisées. Les données sont alors dites « structurées ».

Collection

Descripteurs	Prénom	Nom	Âge
Une valeur du descripteur « Prénom »	Antoine	Ledoux	15
	Pauline	Darcis	16
	Karim	Chanhoun	15
Un objet	Sarah	Goldberg	14

2. LES FORMATS ET REPRESENTATION DES DONNEES

Les données sont organisées de manière à s'adapter à des traitements spécifiques. La manière dont elles sont organisées est le **format**. Les formats CSV, JSON et XML comptent parmi les plus utilisés. Pour stocker des données au format CSV, on écrit les descripteurs sur la première ligne. Puis, sur les lignes suivantes, sont placées les valeurs des descripteurs pour chaque objet. Les mots sont séparés généralement par des points-virgules.

Données au format CSV

```
1 prenom;nom;classe
2 Antoine;Ledoux;seconde
3 Pauline:Darcis:premiere
```

Le traitement des données peut être réalisé de diverses manières : recherche, tri, estimation, calcul, etc. Pour réaliser ces opérations, des phrases logiques, appelées « requêtes » sont exprimées dans un langage informatique. Le plus utilisé aujourd'hui est le SQL. Une requête est constituée de mots-clés dans un ordre précis



Exemple :

La requête ci-dessus sélectionne le nom des élèves de la table **Seconde7** qui ont plus de 15 ans et moins de 19 ans.

Sélectionner nom dans Seconde7 avec age>15 et age<19

Pour obtenir une information, il est courant de faire des requêtes sur plusieurs tables en même temps.

Exemple :

Cette requête fait la même chose que la précédente sur les tables **Seconde7** et **TS3**.

Sélectionner nom dans Seconde7, TS3 avec age>15 et age<19

Les données dans le nuage (cloud)

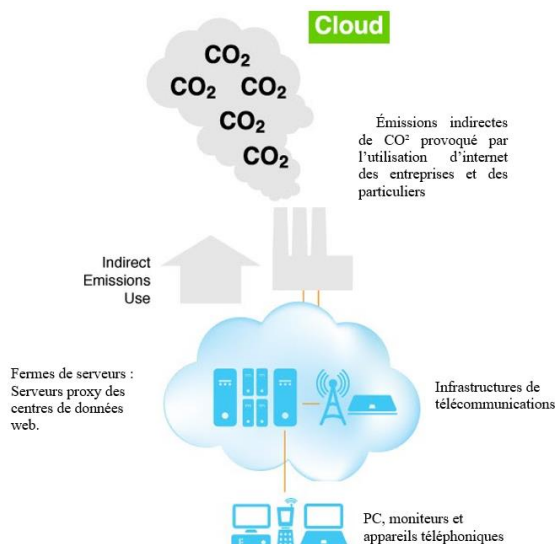
1. LE CLOUD

Le cloud ou **cloud computing** désigne l'accès à des ressources informatiques (stockage, logiciels, puissance de calcul, données) situées dans des serveurs informatiques distants par l'intermédiaire d'un réseau.

Il est possible d'automatiser le stockage de ses données sur le cloud en paramétrant la **synchronisation** des fichiers sur son ordinateur ou son téléphone. Le partage des données y est également facilité



Les outils collaboratifs de Google cloud



2. L'IMPACT DU CLOUD SUR LA CONSOMMATION ENERGETIQUE

Avec l'augmentation de la quantité de données stockées et traitées dans les centres de données, ou data centers, le cloud est devenu un des premiers consommateurs d'électricité dans le monde. Les entreprises doivent adapter leurs technologies pour réduire leur impact écologique.

Exemple :

Les centres de données consomment 10 % de l'électricité mondiale. Un centre de données consomme autant d'électricité que 30 000 habitants européens.

Base de données

Ensemble d'informations structurées accessibles au moyen d'un logiciel.

Collection

Elle regroupe des objets partageant les mêmes descripteurs (par exemple, la collection des contacts d'un carnet d'adresses).

Cloud computing :

Utilisation de ressources situées dans des serveurs informatiques distants.

Descripteur :

Élément servant à décrire une donnée.

Données :

Information diverse pouvant être stockées (dates, températures, position géographiques, ...)

Données personnelles :

Informations identifiant une personne.

Format :

Type d'un fichier numérique.

Identifiant :

Code permettant d'identifier une personne.

Métadonnées :

Informations relatives à un fichier image, son, vidéo, etc.

Objet :

Élément d'une table de données.

Requête :

Texte composé de mots-clés permettant une action sur des données.

Synchronisation :

Copie des données stockées en local sur un serveur.

Table de données :

Données organisées en tableau.

Traitement de données :

Ensemble d'actions permettant d'extraire de l'information.

Valeur d'un descripteur :

Valeur d'une caractéristique d'un objet