

COURS – Données Structurées – SNT

Sommaire

A) Introduction

- 1) Définitions
- 2) Collecte de données
- 3) Base de données

B) Bref historique

C) Fichiers de données

- 1) Formats de fichiers
- 2) Exemples

D) Base de données relationnelles

- 1) Définition
- 2) Schéma relationnel
- 3) Langage SQL

E) Stocker des données

- 1) Définitions
- 2) Supports de stockage
- 3) Lieux de stockage

F) Impact environnemental

- 1) A l'échelle mondiale
- 2) En France

G) Protection des données personnelles

- 1) Collecte des données
- 2) La CNIL
- 3) Le RGPD

A) Introduction

1) Définitions

Donnée : Une donnée est une *valeur*, une *suite d'octets* traitée par ordinateur. A chaque donnée correspond un **type** : Ce peut être un texte, un nombre, une date, une image... Les données se sauvegardent :

- Dans de simples fichiers si elles sont peu nombreuses
- Sinon, dans des bases de données vers des Serveurs de fichiers

Les données présentent plusieurs *caractéristiques*. Elles peuvent être :

- Privées ou placées dans le domaine public
- Vraies ou fausses
- Obtenues honnêtement ou dérobées par des moyens illicites
- Protégées par chiffrement

Information : Une donnée devient une *information* dès qu'elle est interprétée par une personne. Le nom, l'âge ou la date de naissance constituent des informations pour une personne.



Exemple :

Prenons l'exemple d'une maison de 9 m de longueur par 5 m de largeur, située près de la forêt, dans laquelle vit Julie Durant, une femme très sympathique.

Terme	Donnée	Information	Donnée non structurée	Donnée structurée	Donnée personnelle	Métadonnée
Définition	Valeur traitée par un ordinateur, de type numérique, texte ou date, décrivant un objet, une personne, un événement	Valeur interprété par un humain, lui permettant de se représenter le monde réel	Donnée difficile à traiter pour un ordinateur	Donnée facile à traiter pour un ordinateur	Toute donnée se rapportant à une personne physique identifiée ou identifiable (CNIL)	Donnée incluse dans un fichier, apportant des précisions sur ce fichier
Exemple	La valeur « 5 »	5 : largeur de la maison en mètres	La largeur est de 5 m	"Largeur";5;"m" (fichier .csv)	"Durant"	Date de prise de la photo incluse dans le fichier

Remarques :

- En droit, une personne (juridique donc) peut être ;
 - Morale (entreprise, association, établissement public)
 - Ou physique.
- Une donnée personnelle n'est pas une propriété de la personne.

2) Collection de données

Une donnée est généralement associée à un **descripteur**. Une collection est un ensemble de données partageant les mêmes descripteurs. Un moyen de les structurer est d'utiliser une table.

Exemple : les **descripteurs** sont « Prénom », « Taille », « Age » les **objets** sont « Thierry », « Julie », « Pascal », les **valeurs** sont les nombres à l'intersection

Prénom	Taille (m)	Age (ans)
Thierry	1,82	17
Julie	1,71	25
Pascal	1,93	39

Remarques :

- Les termes génériques collection, objet ou descripteur proviennent du programme de SNT.
- Les termes table, enregistrement ou champ s'utilisent pour les bases de données relationnelles.

3) Base de données

Classiquement, une base de données est un **logiciel** qui gère plusieurs collections de données, en relation les unes avec les autres. Par exemple, la base de données d'une bibliothèque conserve les données sur les livres, abonnés, emprunts, ...

B) Bref historique

Chronologie des évolutions dans le traitement des données :

- 1800 : Utilisation des cartes et bandes perforés.
- 1928 : Invention de la bande magnétique, par Fritz Pfleumer.
- 1956 : Invention du disque dur, par Reynold Johnson.
- 1967 : Invention de la disquette rigide par IBM
- 1970 : Invention du modèle relationnel des SGBDR, par Edgar Frank Codd.
- 1979 : Création du premier tableur VisiCalc, par Dan Bricklin.
- 1982 : Commercialisation du CD (Compact Disc), inventé par Philips en 1979.
- 1984 : Apparition sur le marché de la mémoire flash, inventé par Fujio Masuoka en 1980.
- 2013 : Charte du G8 pour l'ouverture des données publiques.
- 2018 : GRPD applicable dans l'UE.

C) Fichiers de données

1) Formats de fichiers

Il est utilisé plusieurs formats de fichier texte lorsqu'une application doit sauvegarder ou communiquer des données à une autre application. Exemples de formats de fichier très classiques :

- CSV (Comma-separated values)
- XML (Extensible Markup Language)
- JSON (JavaScript Object Notation)

2) Exemples

Considérons un groupe de trois élèves :

- Laura, âgée de 12 ans, aime la natation.
- Marc, âgé de 15 ans, aime la musique.
- Julie, âgée de 17 ans, aime l'informatique.

```
Prénom ; Age ; Loisir
Laura ; 12 ; natation
Marc ; 15, musique
Julie, 17, informatique
```

Format CSV : Sa structure ressemble à celle d'un tableau (séparateur « ; » ou « , »). Le fichier obtenu est de taille réduite. Convertissable en fichier tableur (.xls ou .ods)

Format XML : Sa structure ressemble à celle d'un document au format HTML. Les données, encadrées par des balises, sont présentées de manière hiérarchique. Le fichier obtenu est assez volumineux

```
<?xml version="1.0"
encoding="UTF-8"?>
<groupe>
  <eleve>
    <prenom>Laura</prenom>
    <age>12</age>
    <loisir>natation</loisir>
  </eleve>
  <eleve>
    <prenom>Marc</prenom>
    <age>15</age>
    <loisir>musique</loisir>
  </eleve>
  <eleve>
    <prenom>Julie</prenom>
    <age>17</age>
    <loisir>informatique</loisir>
  </eleve>
</groupe>
```

Format JSON : Sa structure ressemble à celle d'un objet en JavaScript.

```
[
  { "Prénom": "Laura",
    "Age": 12,
    "Loisir": "natation"
  },
  { "Prénom": "Marc",
    "Age": 15,
    "Loisir": "musique"
  },
  { "Prénom": "Julie",
    "Age": 17,
    "Loisir": "informatique"
  }
]
```

Les données sont présentées à l'aide d'objets (entre accolades) ou de tableaux (entre crochets)

D) Base de données relationnelle

1) Définition

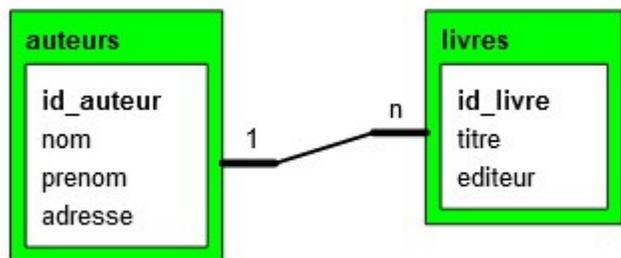
De nombreux **logiciels** utilisent de simples fichiers pour les sauvegardes. Une base de données s'impose lorsque le volume d'informations à **sauvegarder** et à gérer devient très important

Une base de données est un ensemble **hiérarchisé** et structuré permettant de **stocker** un grand volume d'informations. Elle est représentée physiquement par un ensemble de fichiers stockés en mémoire, exploités et gérés (création, recherche, mise à jour) par un **système de gestion** de base de données (SGBD)

Il existe de nombreux **types** de bases de données. Le modèle de base de données le plus répandu est le modèle de base de données **relationnelle**. Dans ce modèle, les informations sont organisées dans des **tables**. Une base de données renferme plusieurs tables reliées entre elles

2) Schéma relationnel

L'exemple ci-dessous comporte deux tables nommées auteurs et livres. Il montre qu'un auteur peut écrire plusieurs livres, mais que chaque livre n'a qu'un seul auteur. Pour les deux tables, il a été choisi, pour le premier champ, une clé primaire. Elle identifie de manière unique les enregistrements



Remarque : Il existe plusieurs manières de représenter une base de données relationnelle. Le MCD (Modèle Conceptuel de Données), le MPD (Modèle Physique des Données), le MLD (Modèle Logique des Données), sont employés dans les modélisations MERISE. Le modèle entité-association est utilisé dans les modélisations ERM (Entity-Relationship model)

3) Langage SQL

SQL (Structured Query Language) est un langage permettant d'accéder aux systèmes de gestion de bases de données. Il permet de récupérer, stocker, insérer, supprimer des champs et des enregistrements dans des tables.

Exemple de visualisation de données

```
SELECT livres.titre FROM livres
WHERE auteurs.nom = "Beaudeau" OR auteurs.nom = "Vernes";
```

Exemple d'enregistrement de données

```
INSERT INTO livres (titre, editeur) VALUES ("Les oiseaux",
"Flamarion");
```

E) Stocker des données

1) Définitions

Les données sont placées dans des fichiers qu'il convient :

- De stocker : Les données sont vivantes, la rapidité d'accès aux données est importante
- De sauvegarder :
 - Elle repose sur la copie périodique de données afin de les restaurer sur les systèmes en cas de besoin
 - Le délai entre deux sauvegardes dépend de la PDMA (perte de données maximale admissible) ou RPO (recovery point objective)
 - Elle peut être effectuée à chaud ou à froid
 - Elle peut être complète ou différentielle (incrémentale)
- D'archiver : Il s'agit de conserver des données de manière fiable sur le long terme. La rapidité d'accès aux données est sans importance

Remarques :

- Les bandes magnétiques restent un support privilégié d'archivage des données en raison de leur très grande capacité, de leur bon rapport qualité/prix et de leur caractère amovible qui permettent de les délocaliser aisément
- Les bandes perforées, utilisées jusqu'à la fin des années 1980, présentaient les avantages d'une grande robustesse et d'une simplicité sur le plan technique

2) Support de stockage

Les principaux supports de stockage sont :

- RAM (Random Access Memory) ou mémoire vive
- ROM (Read Only Memory) en disque dur HDD (Hard Disk Drive)
- ROM (Read Only Memory) en disque dur SSD (Solid State Drive)
- Mémoire flash SSD (Solid State Drive)
- Les cartes de mémoires SD (Secure Digital)

3) Lieux de stockage

Les données peuvent être stockées :

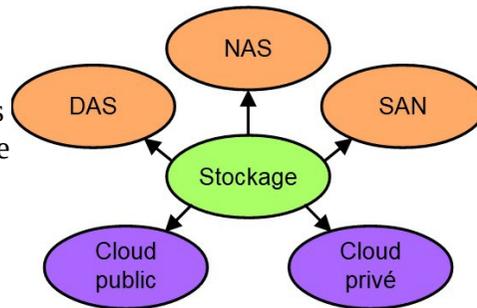
- Sur le réseau local : ENT
- Sur le réseau internet global
- Sur le Cloud

Données stockées localement (Serveur)

- DAS (Direct Attached Storage) : La baie de stockage (armoire contenant des disques durs) est directement reliée à un ordinateur. Si cet ordinateur est un serveur, le stockage devient accessible aux ordinateurs du réseau local.
- NAS (Network Attached Storage) : La baie de stockage est directement reliée au réseau. Elle est ainsi accessible à l'ensemble des ordinateurs du réseau local.
- SAN (Storage Area Network) : La baie de stockage est reliée à plusieurs serveurs par un réseau dédié. Ce système permet d'optimiser le stockage.

Données stockées sur internet (Cloud)

- Solution fiable, gérée par des professionnels de l'informatique.
- Les données peuvent être consultées par l'hébergeur (le secret d'entreprise ne se met jamais dans le Cloud).
- L'impact environnemental n'est pas négligeable.
- On ne sait pas en quel lieu sont stockées les données.



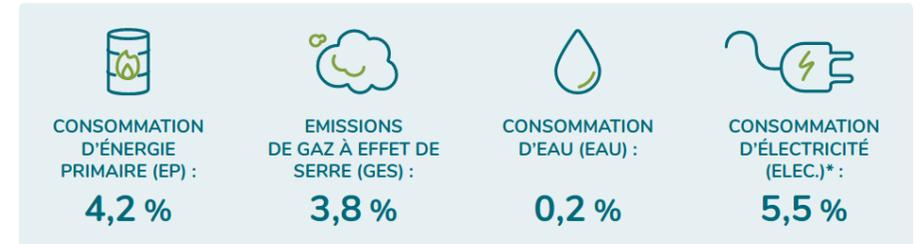
F) Impact environnemental

1) A l'échelle mondiale

Diverses organisations étudient l'impact environnemental du numérique et les moyens de le réduire : L'Adème (Agence de l'Environnement et de la Maîtrise de l'Energie), l'INR (Institut du Numérique Responsable), GreenPeace, Negaoctet...

Rapport de GreenIT de 2019 : <https://www.greenit.fr/empreinte-environnementale-du-numerique-mondial/>

Contribution du numérique à la facture environnementale

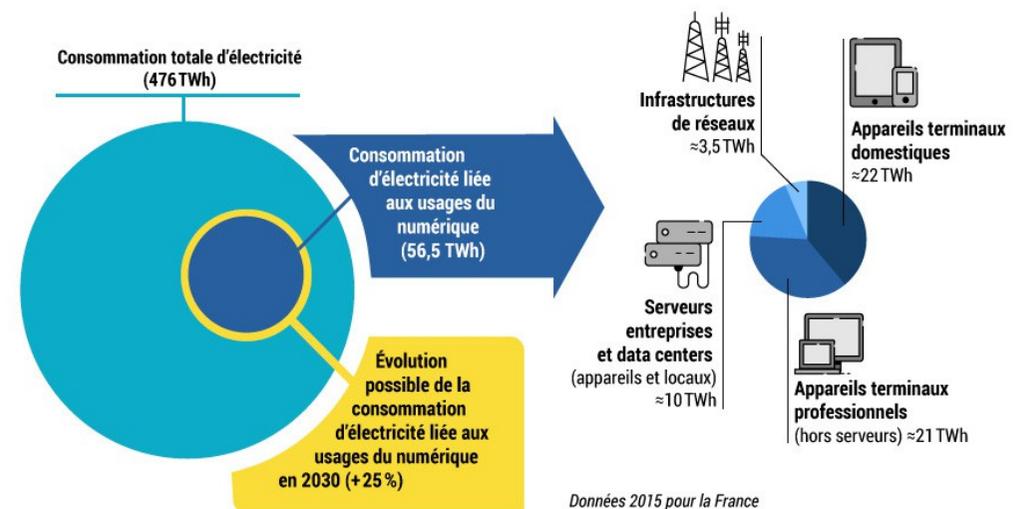


Bilan énergie primaire

Bilan EP	Fabrication	Utilisation	Total
Utilisateurs	30%	30%	60%
Réseau	3%	20%	23%
Centres informatiques	2%	15%	17%

2) En France

Rapport de négaWatt de 2020 : <http://decrypterlenergie.org/la-revolution-numerique-fera-t-elle-exploser-nos-consommations-denergie/>



G) Protection des données personnelles

1) Collecte de données

GAFAM, BATX

Des sociétés privées telles que Google ou FaceBook développent des techniques de collecte des données personnelles, de manière opaque pour l'utilisateur. Ces données sont ensuite traitées et commercialisées, généralement à des fins publicitaires.

- GAFAM : Google, Apple, Facebook, Amazon et Microsoft (géants du numérique américains)
- BATX : Baidu, Alibaba, Tencent et Xiaomi (géants du numérique chinois)

Affaire Snowden

En 2013, Edward Joseph Snowden a révélé l'existence de plusieurs programmes de surveillance de masse américains et britanniques, basés sur la captation des métadonnées, des appels téléphoniques, des systèmes d'écoute sur internet (courriel, réseaux sociaux).

Ces programmes portent les noms de PRISM, XKeyscore, Boundless Informant et Bullrun pour le gouvernement américain, Tempora, Muscular et Optic Nerve pour le gouvernement britannique.

Système de crédit social

C'est un système chinois de réputation des citoyens. Chacun d'entre eux se voit attribuer une note, échelonnée de 350 à 950 points. Ce système de surveillance utilise l'intelligence artificielle et la reconnaissance faciale.

Avec une note élevée, on a plus de chances de trouver un emploi. Ceux qui ont un faible score peuvent être interdits d'avion ou de train.

2) La CNIL

Créée par la loi Informatique et Libertés du 6 janvier 1978, cette AAI (autorité administrative indépendante) est un organisme public qui agit au nom de l'Etat, composée de 18 membres élus ou nommés.

Elle est chargée de veiller à la protection des données personnelles contenues dans les fichiers et traitements informatiques ou papiers, aussi bien publics que privés. Ainsi, elle est chargée de veiller à ce que l'informatique soit au service du citoyen et qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

Elle a un rôle d'alerte, de conseil et d'information vers tous les publics mais dispose également d'un pouvoir de contrôle et de sanction.



3) Le RGPD

Le RGPD, ou GDPR (General Data Protection Regulation) en anglais, est un règlement de l'UE (Union européenne) qui constitue le texte de référence en matière de protection des données à caractère personnel.

Sa date d'entrée en vigueur est le 25 mai 2018.

Remarque : Le RGPD s'applique... dans l'UE seulement.