

Chapitre 1 : Données structurées

1. Données

A. Données et structure

Regardez la vidéo suivante et répondez aux questions



Q1. Définir les mots suivants et donner des exemples :

- Donnée :

- Descripteur :

- Collection de données :

- Base de données :

Q2. Lorsque plusieurs collections de données sont organisées dans un tableau, on dit que les données sont structurées. Sur l'exemple ci-contre, entourer une donnée, un descripteur, une collection de données, une base de données.

Q3. Toute information permettant de caractériser les données qui sont représentées sont importantes. Ces informations sur la base de données sont appelées **métadonnées** (elles décrivent les données). Quelles métadonnées sont visibles ici? En quoi les métadonnées sont une source d'informations importante?

Sexe et âge	Aucun diplôme, brevet des collèges	CAP, BEP ou équivalent	Baccalauréat ou équivalent	Bac + 2	Supérieur à Bac + 2	Non déterminé	Ensemble
25-34 ans							
Femmes	10,9	15,6	20,5	13,7	39,0	0,3	100,0
Hommes	13,0	18,3	22,6	12,9	32,8	0,4	100,0
Ensemble	11,9	16,9	21,5	13,3	36,0	0,3	100,0
35-44 ans							
Femmes	13,7	15,4	20,0	20,1	30,6	0,2	100,0
Hommes	14,9	22,2	20,4	15,9	26,4	0,2	100,0
Ensemble	14,3	18,7	20,2	18,1	28,6	0,2	100,0
45-54 ans							
Femmes	17,9	23,9	17,4	17,9	22,7	0,2	100,0
Hommes	20,1	30,3	15,4	13,9	20,1	0,2	100,0
Ensemble	19,0	27,0	16,4	16,0	21,4	0,2	100,0
55-64 ans							
Femmes	30,0	27,0	15,7	13,6	13,4	0,2	100,0
Hommes	26,4	36,7	11,3	9,9	15,6	0,1	100,0
Ensemble	28,3	31,7	13,6	11,9	14,5	0,2	100,0
Ensemble							
Femmes	18,3	20,7	18,3	16,4	26,0	0,2	100,0
Hommes	18,7	27,1	17,3	13,2	23,5	0,2	100,0
Ensemble	18,5	23,8	17,8	14,8	24,8	0,2	100,0

Lecture : en 2020, 32,8 % des hommes âgés de 25 à 34 ans ont un diplôme supérieur à bac + 2.
 Champ : France hors Mayotte, personnes vivant en logement ordinaire, de 25 à 64 ans.
 Source : Insee, *enquête Emploi*.

Table 1 : Diplôme le plus élevé selon l'âge et le sexe en 2020

Q4. Définir alors ce que sont des données personnelles.

Q5. À quoi sert de structurer des données? Citer des exemples où cette structuration est utilisée. Définir ce qu'est un algorithme.

B. Stocker des données

Pour stocker des données de manière persistante, il faut les stocker dans un _____. Le _____ à choisir dépend du type de données à stocker et de la communauté qui l'utilise (exemples de format : .jpg ou .png pour des images, .csv pour les tables, .txt). Les informations relatives aux données stockées dans ces fichiers sont appelées _____.

Or, toute base de données a une taille, et stocker des données prend donc de la place. La taille d'un fichier peut s'évaluer en octet (o) ou en byte (B). À titre quantitatif, un caractère dans un fichier (une lettre, un chiffre, un symbole, ...) a une taille de 8 octet. Cela correspond à 1B.

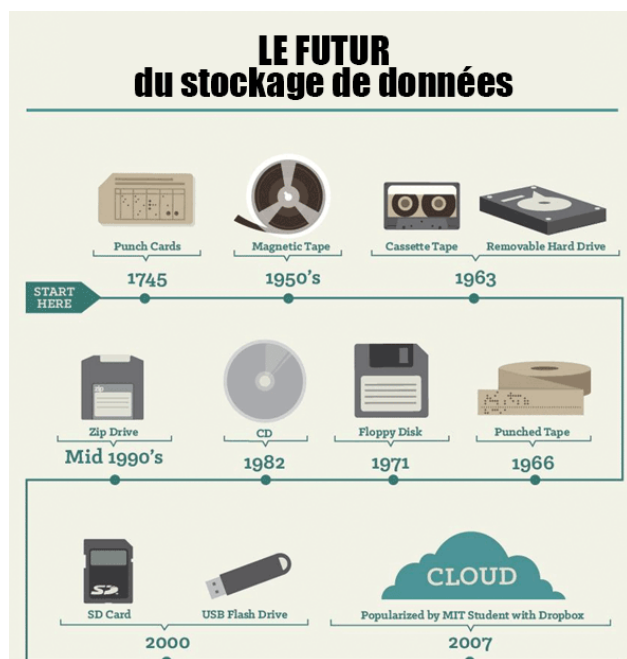


Figure 1 : évolution des moyens de stockage de données.

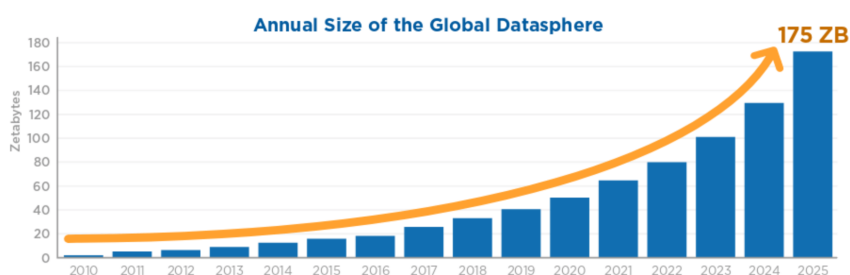


Figure 2 : Evolution et prédiction de la taille de la datasphère au cours du temps (source : IDC, 2018). Note : 1ZB=10²²B.

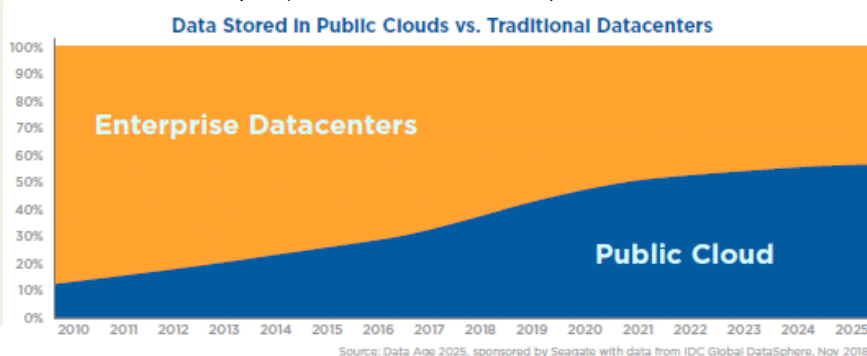


Figure 3 : pourcentage de données stockées dans le cloud public par rapport aux données stockées dans les data centers.

Q1. Commenter l'évolution des différents types de support de stockage : comment les données ont-elles été stockées?

Q2. Les données mondiales sont estimées en 2020 à environ 60 ZB (« zetabytes »). Commenter l'évolution de stockage des données depuis 2010. Comment cela peut-il s'expliquer?

Q3. Quel type de stockage semble prendre de l'ampleur? À quoi correspond-il? Pourquoi prend-il autant d'importance?

Exposés à venir :

- (1) Où sont stockées les données ? (Livre p.86)
- (2) Protéger les données personnelles (Livre p.90)

La présentation doit durer 15 min au maximum.

Elle doit comporter :

- une partie introductive dans laquelle il faut :
 - présenter la problématique, contextualiser le sujet, et présenter les documents étudiés
 - Présenter la question que l'on se pose, et à laquelle nous allons proposer une réponse. Il faut définir les termes importants s'ils sont nouveau.
 - Présenter le plan qui va être suivi pour répondre à cette question (dans un premier temps, nous allons discuter..., puis nous allons étudier,... Et enfin nous allons mettre en évidence que....)
- Développer les parties annoncées. Vous pouvez, pour cela, vous aider des questions posées à la fin des corpus de documents. La présentation peut se faire en s'appuyant sur des figures, des graphiques, des images, des supports visuels (powerpoint, ...). On rappelle que toute donnée chiffrée, tout graphique, tout visuel doit s'accompagner de sa source.
- Faire une conclusion : qui synthétise la discussion qui a été faite et propose une ouverture.

Vous présenterez par groupe de 2 ou 3, et toutes les personnes doivent participer à l'oral de manière équitable.

Les autres membres de la classe qui ne présentent pas les corpus sont priés de lire les deux études de documents, de définir les mots importants, et de répondre aux questions par écrit également.