

Thème : données structurées

Repères historiques

- 1930 : utilisation de _____, premier support de stockage de données
- 1956 : invention du _____, permettant de stocker de plus grandes quantités de données, avec un accès de plus en plus rapide
- 1970 : invention du modèle relationnel (E. L. Codd) pour la structuration et l'indexation des _____
- 1979 : création du premier tableur, _____
- 2009 : Open Government Initiative du président _____
- 2013 : charte du G8 pour l'_____ publiques.

I. Les données

Une donnée est la _____ d'une information, décrivant un objet, une personne, un événement... Elle peut être de _____ variée (nombre, texte, image, son, vidéo...)

Les données sont stockées en mémoire dans des _____. Selon leur nature, elles ne sont pas représentées de la même façon : le _____ décrit une représentation et est identifié par _____ du fichier. Les programmes travaillent avec des fichiers de certains formats adaptés. Par exemple, Paint ouvre les images au format PNG, BMP...

Une donnée _____ est une donnée identifiant directement ou indirectement une personne physique. Exemples : nom, numéro de téléphone, photographie, date de naissance, empreinte digitale...

Une _____ est une donnée particulière apportant des informations sur la donnée principale. Lors de l'enregistrement, des métadonnées peuvent être associées au fichier. Leur nature varie selon le _____ de fichier. Exemple : pour tout fichier, la date de la dernière modification est une métadonnée. Pour un fichier de musique, le nom de l'album peut être renseigné et constituer une métadonnée.

Les données ouvertes (ou _____) sont partagées et peuvent être réutilisées librement. Elles sont produites par des états, des entreprises, des associations...

II. Les données structurées

(a) Les tables de données

On parle de données structurées lorsqu'on les organise sous la forme d'un tableau, appelé _____. Une liste de _____ permet de caractériser les données. Un _____ est un élément d'une table. Une _____ regroupe des objets partageant les mêmes descripteurs. Les _____ de tous les descripteurs d'un objet sont précisées. Les données sont alors dites _____.

Collection		
Descripteurs		
Prénom	Nom	Âge
Antoine	Ledoux	15
Pauline	Darcis	16
Karim	Chanhoun	15
Sarah	Goldberg	14

Une valeur

Un objet

(b) Format et représentation des données

Les données sont organisées de manière à s'adapter à des traitements spécifiques. La manière dont elles sont organisées est le _____. Les formats CSV (_____), JSON (_____) et XML (_____) sont parmi les plus utilisés car ils sont simples, ouverts et interopérables.

Données	CSV	JSON	XML						
<table><tr><td>Nom</td><td>Prenom</td></tr><tr><td>DUPONT</td><td>Jean</td></tr><tr><td>DURAND</td><td>Paul</td></tr></table>	Nom	Prenom	DUPONT	Jean	DURAND	Paul	Nom,Prenom DUPONT,Jean DURAND,Paul	[{ "Nom": "DUPONT", "Prenom": "Jean" }, { "Nom": "DURAND", "Prenom": "Paul" }]	<eleve> <nom>DUPONT</nom> <prenom>Jean</prenom> </eleve> <eleve> <nom>DURAND</nom> <prenom>Paul</prenom> </eleve>
Nom	Prenom								
DUPONT	Jean								
DURAND	Paul								

III. Le traitement des données structurées

(a) Opérations sur les données structurées

On peut effectuer différents traitements sur une table de données, comme celle ci-contre :

Nom	Prenom	Age	Classe	Ville
DUPONT	Jean	16	2de	Toulon
DURAND	Paul	17	1re	Hyères

- Calculs sur les _____ d'un descripteur
Exemple : calcul de l'âge moyen des élèves de cette table
- _____ des objets dont les valeurs répondent à des contraintes spécifiques, filtrage de ces derniers
Exemple : filtrage des élèves ayant 16 ans ou plus
- _____ des objets selon des critères précis portant sur leurs valeurs
Exemple : tri des élèves dans l'ordre alphabétique de leur nom

(b) Traitement de données structurées

Le traitement des données peut être réalisé de diverses manières :

- A l'aide d'un tableur, permettant de manipuler visuellement les données et de les représenter graphiquement. Les calculs se font à l'aide de _____ constituées à partir de _____ prédéfinies. Le filtrage et le tri se font à l'aide d'assistants
Exemple : pour calculer l'âge moyen d'un élève de 2de, on _____ pour ne conserver que les élèves de 2de puis on applique la fonction _____ à la colonne Age (_____) via la fonction _____
- A l'aide de bibliothèques d'un langage de programmation, par exemple **pandas** pour le langage Python. Les données doivent se trouver dans un fichier CSV, JSON ou XML dont il faut lire le contenu avant traitement.
- A l'aide d'un SGBD (Système de Gestion de Bases de Données), logiciel spécialisé dans lequel on exprime des requêtes (souvent dans le langage _____, proche de l'anglais) portant sur les descripteurs.

Par exemple, on peut sélectionner le nom et le prénom des élèves de 16 ans qui habitent à Toulon :
En français : SELECTIONNER nom, prenom DANS eleves AVEC age = 16 ET ville = 'Toulon'
En langage SQL : SELECT nom, prenom FROM eleves WHERE age = 16 AND ville = 'Toulon'

IV. Les données dans le nuage

(a) Informatique en nuage

Le _____ ou _____ désigne l'accès à des _____ informatiques (stockage, logiciels, puissance de calcul, données...) situées dans des _____ informatiques _____ par l'intermédiaire d'un réseau.

Par exemple, on peut automatiser le stockage de ses données en paramétrant la _____ des fichiers, utiliser des programmes sans les installer, _____ des fichiers et les éditer collaborativement.

(b) L'impact du cloud sur la consommation énergétique

Avec l'augmentation de la quantité de données stockées et traitées dans des _____ ou _____, le cloud est devenu l'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 _____ de l'électricité mondiale. Un centre de données consomme autant d'électricité que _____ habitants européens.