



Nursery

A la découverte des métiers du digital



Le parcours

Découvrir différents métiers et leurs composantes au travers des présentations et de la pratique personnelle

Les ateliers

Atelier 4 : Découverte des métiers de la Data



Atelier 4

Découverte de l'univers DATA

Découverte des métiers de la Data

Sommaire :

L'évolution du stockage de données

L'évolution de la collecte de données

Le big data

Les différents métiers de la Data

Evolution du stockage de données

Découverte de l'univers DATA

❖ 1956 :

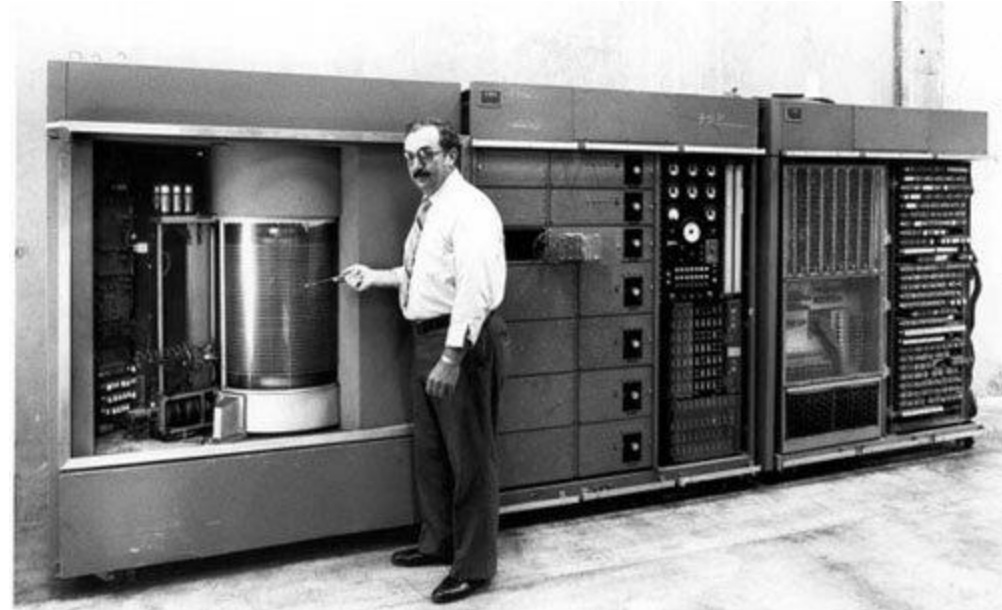
1^{er} DISQUE DUR:

IBM RAMAC 305

Capacité : 5 Mo

Poids : environ 1 Tonne

Tarif : 50 000\$ soit 10 000 \$/Mo



Evolution du stockage de données

Découverte de l'univers DATA

❖ Début des années 80 :

Support fixe :

DISQUE DUR

Capacité : 5 Mo

Format : 5,25 pouces

Tarif : 1500\$ soit 300 \$/Mo



Support Amovible :

DISQUETTE 3,5 pouces

Capacité : 1,44 Mo

Format adapté à une poche de chemise

Inventé par Sony, démocratisé par Apple sur ses Macintosh.

Pour la plupart des ordinateurs c'est le seul outil de stockage.



Evolution du stockage de données

Découverte de l'univers DATA

❖ Fin des années 80 :

Support fixe :

DISQUE DUR

Capacité max : 1 Go

À cette époque, la capacité moyenne des disques durs est de 10 Mo.

Support Amovible :

CD



Capacité : 500 à 700 Mo

Inventé par Sony et Philips.

D'abord utilisé pour la musique, il servira plus tard à stocker des données via la démocratisation des lecteurs-graveurs.

Evolution du stockage de données

Découverte de l'univers DATA

❖ Fin des Années 90 :

Support fixe :

DISQUE DUR



Capacité max : 25 Go en 1998

Le standard pour les PC de bureau est de 2 Go

Support Amovible :

DVD



Capacité : 4,7 Go

Inventé par Sony et Philips

Evolution du stockage de données

Découverte de l'univers DATA

❖ Années 2000 :

Support fixe :

DISQUE DUR

Capacité max : 500 Go (2005), 3 To (2010)

Vers 2002, les disques durs de 40 Go sont courants pour des PC de bureau.

Support Amovible :

CompactFlash

Capacité : 1 Mo au début
puis jusqu'à 512 Go

Inventé par SanDisk et Toshiba

*Utilisé pour les téléphones, appareils photo,
Caméscope, consoles de jeux et lecteurs MP3*

En 1991, 1 Go de mémoire Flash coûtait 45 000 \$



Evolution du stockage de données

Découverte de l'univers DATA

❖ Années 2010 :

Support fixe :

DISQUE DUR

Capacité max : 3 To (2010), 10 To (2015)

En 2010, le standard pour les PC de bureau est de 1 To (à partir de 0,1 €/Go) et de 500 Go pour les PC portables.

Support Amovible :

Mémoire Flash Nand, Carte SD & Clé USB

Capacité : 32Go à 1 To aujourd'hui

Evolution du stockage de données

Découverte de l'univers DATA

❖ Depuis 2015 :

Support fixe :

DISQUE DUR SSD

Jusqu'à 30 To (Samsung PM1643)

250 Go à 2 To pour les modèles les plus courants en 2019

Le Cloud :

1.000 milliards d'octets (1 To)

pour quelques euros par mois

Démocratisé par Amazon qui souhaitait rentabiliser en dehors des périodes de forte affluence, ses immenses serveurs informatiques qui étaient sous-utilisés.

Evolution du stockage de données

Découverte de l'univers DATA

❖ Depuis 2015 :

Support fixe :

DISQUE DUR SSD

Jusqu'à 30 To (Samsung PM1643)

250 Go à 2 To pour les modèles les plus courants en 2019

Le Cloud :

1.000 milliards d'octets (1 To)

pour quelques euros par mois

Démocratisé par Amazon qui souhaitait rentabiliser en dehors des périodes de forte affluence, ses immenses serveurs informatiques qui étaient sous-utilisés.

Les bases de données

Découverte de l'univers DATA

❖ 4 ÉVOLUTIONS MAJEURES

1. **Les BDD hiérarchiques** Chaque enregistrement dépendait d'un seul enregistrement parent.
2. **Les BDD réseau** Les relations n-n (plusieurs parents / plusieurs enfants) sont désormais permises et offrent de nouvelles possibilités. Mais cela implique une trop grande dépendance entre les données et les programmes.
3. **Les BDD relationnelles** C'est le type de bases que l'on connaît et que l'on pratique aujourd'hui. Dans ce type de bases de données, les données sont organisées en tables. Ces BDD sont simples à maintenir, à faire évoluer et indépendantes de leur support.
4. **Les BDD objet** A moyen terme, les principes du modèle objet (une entité complexe et évolutive), une fois mûrs et bien standardisés, rejoindront les bases de données relationnelles afin de simplifier les requêtes complexes.

Les bases de données

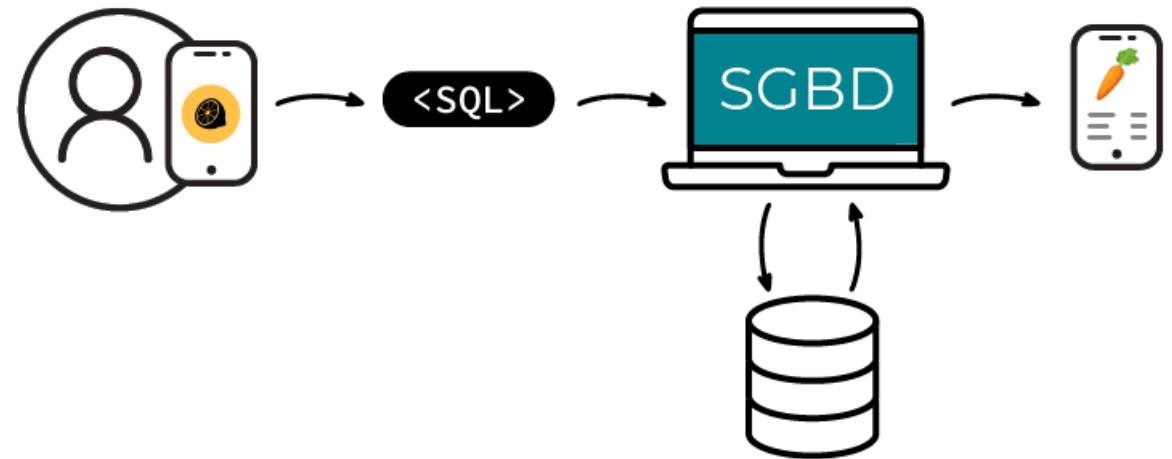
Découverte de l'univers DATA

Système de Gestion de Bases de Données Relationnelles (SGBDR)

Le SGBD ou SGBDR est le logiciel qui va vous permettre de manipuler les données d'une base.

C'est ce logiciel qui commande les interactions avec votre base pour y récupérer, ajouter, modifier ou supprimer des données.
(CREATE READ UPDATE DELETE -> CRUD)

Il va utiliser le langage SQL (Structured Query Language)



Les bases de données

Découverte de l'univers DATA

Système de Gestion de Bases de Données Relationnelles (SGBDR)

Table also called **Relation**

© guru99.com

CustomerID	CustomerName	Status
1	Google	Active
2	Amazon	Active
3	Apple	Inactive

Primary Key

Domain
Ex: NOT NULL

Tuple OR Row
Total # of rows is **Cardinality**

Column OR Attributes
Total # of column is **Degree**

Les bases de données

Découverte de l'univers DATA

Système de Gestion de Bases de Données Relationnelles (SGBDR)

- **MySQL**

C'est le plus connu et utilisé, car il était (auparavant) open-source avant d'être racheté par Oracle.

The Facebook logo, consisting of the word "facebook" in white lowercase letters on a blue rectangular background.

- **Oracle Database**

C'est très cher, mais utile pour traiter un très gros volume de données. Pour les très grandes entreprises donc.



- **PostgreSQL**

C'est "l'autre" grand SGBD open-source disponible sur le marché.



- **MariaDB**

C'est une « copie » améliorée de MySQL créée par le fondateur de MySQL lui-même. C'est un SGBD open-source qui devient de plus en plus populaire.



Le déluge de données

Découverte de l'univers DATA

❖ Aujourd'hui :

**Tous les appareils et leurs utilisateurs
génèrent des données en masse...**

Le déluge de données

Découverte de l'univers DATA

❖ Depuis 2015 :

3,8 Zetta Octets de données en 2015

**L' équivalent d'une pile de Blu-Ray qui
ferait 7 fois le tour de la terre !**

1 Zo

= mille milliards de
milliards d'octets

1 000 000 000 000 000 000 000 octets

Le déluge de données

Découverte de l'univers DATA

❖ Depuis 2015 :

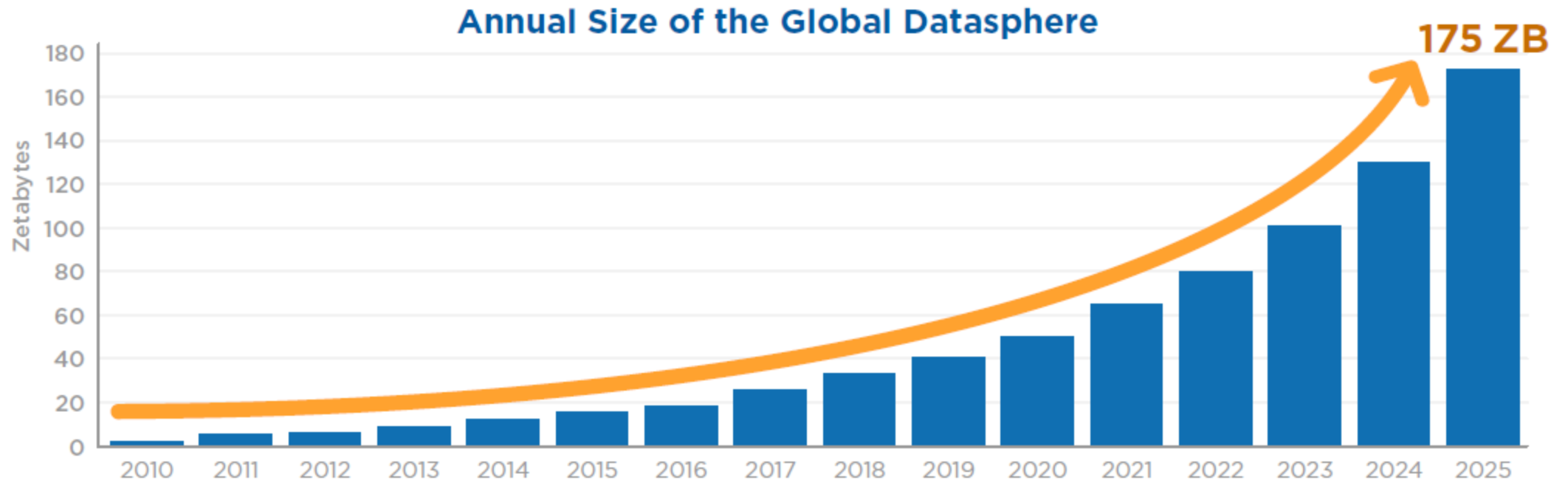
- 60% de croissance par an des volumes d'informations
- 1 Boeing produit 20 To / heure de données
- 250 milliards d'e-mail envoyés chaque jour (80% de spam)
- 72h de vidéo déposées chaque minute sur Youtube

Le déluge de données

Découverte de l'univers DATA

❖ Les prévisions d'ici 2025 :

Figure 1 - Annual Size of the Global Datasphere



Le déluge de données

Découverte de l'univers DATA

❖ Le problème n'est plus le stockage

80% de l'information

est **non-structurée**

95% de l'information

est **non-exploitée**

L'ère du BIG DATA

Découverte de l'univers DATA

L'objectif principal du Big Data est de réussir à faire apparaître des enseignements (insights) et des connexions entre de gros volumes de données de nature hétérogène qui seraient impossible à obtenir avec les méthodes classiques d'analyse des données.

Métiers du BIG DATA

Découverte de l'univers DATA

- Data Analyst et Data Miner
- Data Scientist
- Architect Big Data
- Consultant Business Intelligence

L'ère du BIG DATA

Découverte de l'univers DATA

Data Analyst

-> recueille, analyse et exploite des données existantes. Il doit vérifier que les données ne sont pas obsolètes et qu'elles peuvent être transformées en énormes données statistiques. L'exploitation des données lui permet de faire par exemple des rapports financiers, d'améliorer les tableaux de bord, de répondre aux problématiques de la structure...

- Le Data miner, cherche les données utiles à l'entreprise, il les formate et nettoie ses différentes données, afin qu'elles soient utilisables par tous les départements de l'entreprise.

L'ère du BIG DATA

Découverte de l'univers DATA

Data Scientist

-> mi informaticien, mi mathématicien, doit être en mesure de rendre intelligible des données brutes pour son entreprise, en utilisant des techniques analytiques avancées (Deep Learning, Machine Learning, analyse de textes...).

Il cherche à construire de nouveaux modèles de big data qui dans la plupart du temps diffèrent des systèmes déjà existants afin d'élargir les possibilités d'exploitation des données.

L'ère du BIG DATA

Découverte de l'univers DATA

Architect Big Data

-> c'est la personne qui se charge de récolter toutes les données brutes qui peuvent provenir de différentes sources, internes ou externes.

Son rôle est de créer une infrastructure pour intégrer, centraliser, protéger et maintenir les données utiles à l'entreprise.

L'ère du BIG DATA

Découverte de l'univers DATA

Consultant Business Intelligence

-> le Consultant Business Intelligence a pour principale mission de proposer aux décideurs d'une entreprise un outil d'aide à la décision permettant d'avoir une vue d'ensemble de l'activité de l'entreprise.

Extraire/Transformer/Charger les données dans le but de produire des rapports ou tableaux de bords lisibles et compréhensibles par les décideurs.