

BDW - Systèmes de Gestion de Bases de Données

Fabien Duchateau

fabien.duchateau [at] univ-lyon1.fr

Université Claude Bernard Lyon 1

2024 - 2025



<https://perso.liris.cnrs.fr/fabien.duchateau/BDW/>

Gestion de données

Prise de conscience de l'importance des informations :

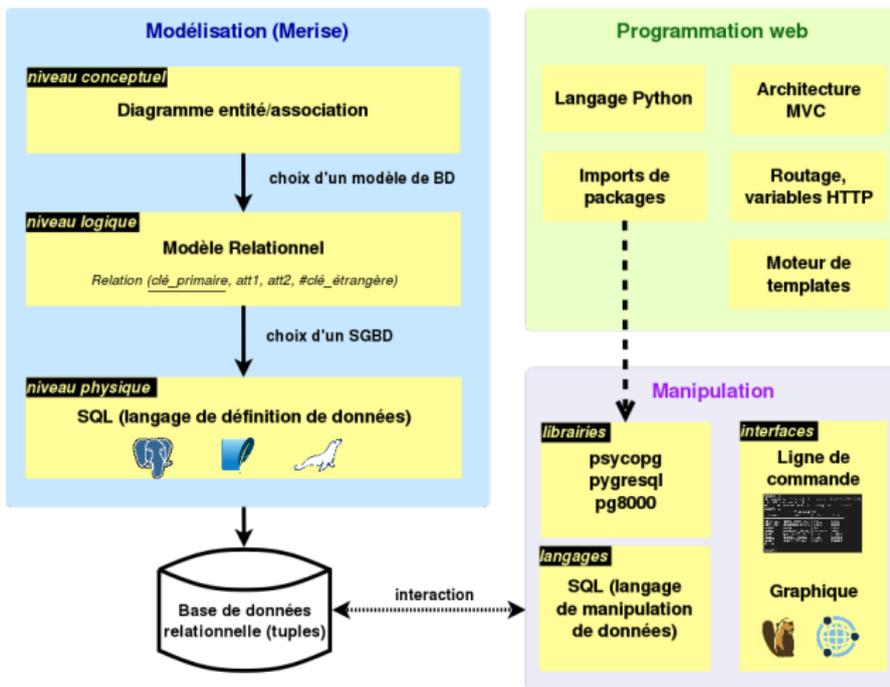
- ▶ Organisations (entreprises, collectivités) et individuelles (web)
- ▶ Nombreux domaines (biologie, astronomie, service public, etc.)
- ▶ Société de l'information, de la connaissance

Organiser, exploiter, partager les données \Rightarrow gestion de données

Rubin M.R., Taylor M., *The Knowledge Industry in United States : 1960-1980*, Princeton University Press (1984)

Curien N., Muet P.A., *La société de l'information*. Documentation Française (2004)

Positionnement dans BDW



Ces diapositives utilisent **le genre féminin** (e.g., chercheuse, développeuses) plutôt que **l'écriture inclusive** (moins accessible, moins concise, et pas totalement inclusive)

Plan

Introduction à la gestion de données

Systèmes de gestion de bases de données (SGBD)

Données

Un ensemble de données, c'est :

- ▶ Des objets :
 - ▶ Un nom, par exemple *Fabien*
 - ▶ Un cours, par exemple *bases de données et programmation web*
 - ▶ Une date, par exemple *septembre 2024*
 - ▶ ...

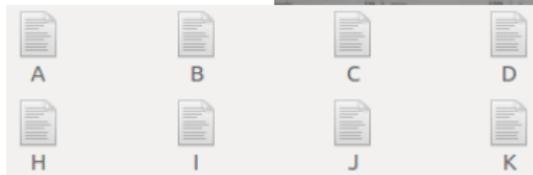
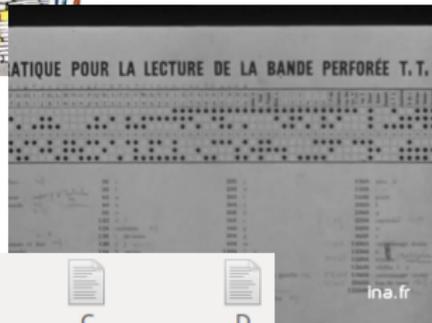
- ▶ Mais aussi des des liens ou des relations entre ces objets :
 - ▶ *Fabien* enseigne le cours *bases de données et programmation web* en *septembre 2024*

Exemples de collections de données

- ▶ Une banque gère un ensemble de données sur ses clientes, sur les comptes bancaires, sur ses employées, les produits et services proposés, etc.
- ▶ La société qui gère les Velovs a un ensemble de données sur les abonnées, la flotte de vélos, les locations en cours, etc.
- ▶ Une équipe de biologistes modélise les protéines, les complexes protéiques, les expérimentations réalisées, les interactions découvertes, etc.

Évolution du stockage des données

1. Fiches papier
2. Cartes perforées
3. Fichiers informatiques
4. Bases de données



<https://www.ina.fr/video/CAF97059686>

Gestion de données par fichiers informatiques

Quand on utilise des fichiers pour gérer un ensemble de données :

- ▶ Chaque application définit et gère ses fichiers
 - ▶ la manipulation des fichiers est directement intégrée dans le programme
- ▶ Un fichier est une suite d'enregistrements contenant des données logiquement liées
 - ▶ bibliothèques dans les différents langages pour simplifier la lecture et l'écriture dans ces fichiers

```
id_facture, nom_client, adresse_client, tél_client, montant
00001, Gaston, "10 rue de la gare 59000 LILLE", 0908070605, 175
00002, Kyla, "3 rue de la poste 56000 VANNES", 0102030405, 34.89
00003, Tintin, "274 rue du bois 75000 PARIS", 0304050607, 67
00003, Mariette, "17 avenue de la liberté 33000 BORDEAUX", 0706050403, 102.54
```

Gestion de données par fichiers informatiques (2)

Inconvénients de la gestion par fichiers :

- ▶ **Redondance** (coût de stockage et d'accès)
 - ▶ *les coordonnées d'une cliente seront potentiellement stockées dans plusieurs fichiers*
- ▶ **Incohérence** (souvent liée à la redondance)
 - ▶ *lors d'un changement d'adresse d'une cliente, il faut vérifier que cette adresse est mise à jour dans tous les fichiers où elle apparaît (encore faut-il savoir où ?)*
- ▶ **Difficulté d'accès** (les utilisatrices ne peuvent pas faire des requêtes non prévues dans les programmes)

Gestion de données par fichiers informatiques (3)

Inconvénients de la gestion par fichiers (suite) :

- ▶ **Anomalies dues à la concurrence**

- ▶ *deux opérations lisent en même temps le solde d'un compte puis ajoutent chacun une somme à ce compte, seul le dernier ajout sera pris en compte*

- ▶ **Manque de sécurité**, tout le monde ne doit pas pouvoir voir ou modifier les mêmes données

- ▶ *une employée ne peut voir que son salaire, pas celui des autres*

- ▶ **Gestion de l'intégrité** (les données obéissent à des contraintes que chaque programme doit vérifier avant de modifier)

- ▶ *un compte ne peut avoir un solde inférieur à -500€*

En résumé...

Inefficacité de la gestion de données par fichiers
⇒ introduction des **bases de données**

Une base de données est une collection de données :

- ▶ Enregistrée (sur un support adressable)
- ▶ Dont la structure ne dépend pas de l'application mais des données à stocker
- ▶ Cohérente
- ▶ De redondance minimale
- ▶ Accessible de manière concurrente par plusieurs utilisatrices

Plan

Introduction à la gestion de données

Systèmes de gestion de bases de données (SGBD)

Définition

Un Système de Gestion de Bases de Données (SGBD, DBMS en anglais) est une application qui permet de définir et manipuler un ensemble de données (les bases de données)

access application computer conceptual data
database DBMS design development
external information internal language level management
model navigational object operational performance products
programming query record relational retrieved security
software SQL storage store structures support system
table transaction types users views XML

Grands principes des SGBD

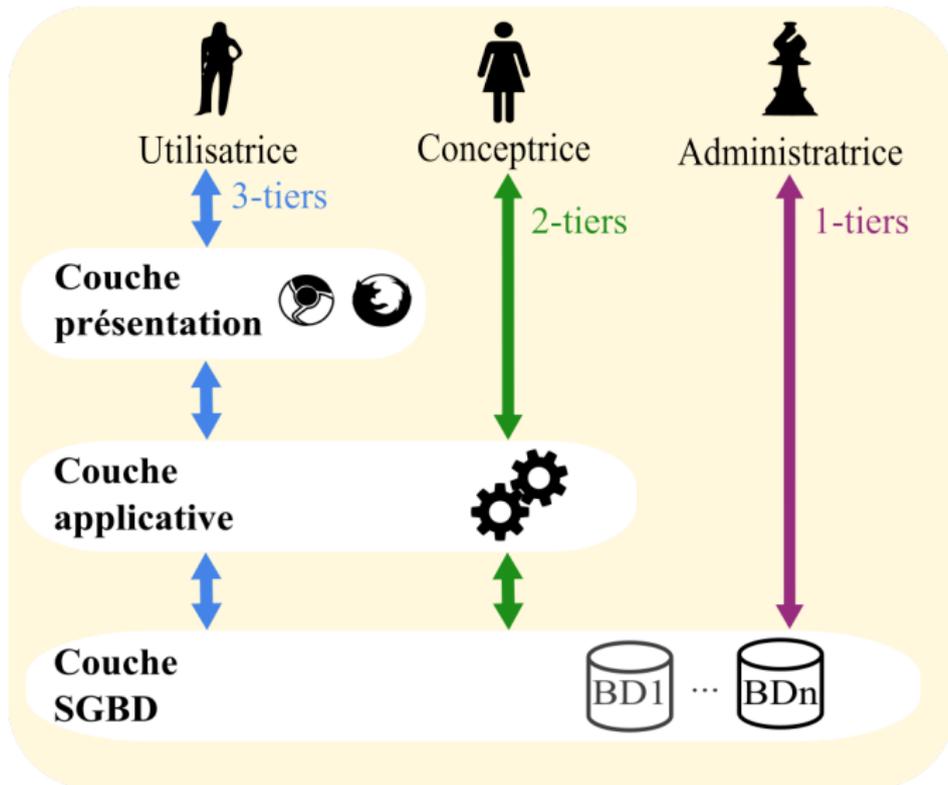
- ▶ **Abstraction** = le SGBD est un médiateur entre l'utilisatrice et l'ordinateur et doit donc présenter les données de manière intuitive et permettre de les manipuler à un niveau abstrait
- ▶ **Universalité** = le SGBD doit permettre de représenter l'ensemble des données d'une entreprise ou organisation quelconque tout en offrant des fonctionnalités variées
- ▶ **Indépendance** = le SGBD distingue ce que voient les utilisatrices (clarté) et ce qui est effectivement géré (efficacité) à travers trois niveaux indépendants

Qui utilise les SGBD ?

Trois niveaux (idéalement) indépendants pour trois profils :

- ▶ Niveau vue/externe (profil utilisatrice)
 - ▶ quelles données sont manipulables ?
- ▶ Niveau conceptuel/logique (profil conceptrice)
 - ▶ quelles est la structure des données stockées ?
- ▶ Niveau physique/interne (profil administratrice)
 - ▶ comment sont organisées des données sur le support physique ?
 - ▶ comment elles sont stockées ?
 - ▶ comment accéder rapidement aux données (index) ?

Utilisation d'un SGBD selon le profil



Fonctionnalités d'un SGBD

Définition d'une base de données :

- ▶ Construire le schéma d'une base de données
 - ▶ l'organisation / la structuration des données
 - ▶ les types de données
 - ▶ les contraintes d'intégrité (cohérence) sur les données

Un **Langage de Description de Données** (LDD) unique, commun aux différentes applications

- ▶ Ce n'est pas l'application qui guide la structuration mais les données à représenter

Fonctionnalités d'un SGBD (2)

Manipulation des données :

- ▶ Outils et mécanismes permettant de faire communiquer la base de données et les applications qui en font usage
- ▶ Création, interrogation, mise à jour et suppression de données (CRUD)

Un **Langage de Manipulation de Données (LMD)** :

- ▶ Spécifie ce que l'on veut faire ou obtenir plutôt que comment le faire ou l'obtenir
- ▶ Garantit une indépendance entre données et applications

<https://fr.wikipedia.org/wiki/CRUD>

Fonctionnalités d'un SGBD (3)

Garantie de l'intégrité :

- ▶ Contraintes d'intégrité spécifiées dans le schéma de la base et préservées par le SGBD
- ▶ Possibilité de programmer des contraintes complexes

Reprise après panne :

- ▶ Restauration de la BD dans un état cohérent et récent
- ▶ Journalisation des actions (logs)

Gestion de la concurrence :

- ▶ Empêcher les incohérences lors d'accès multiples

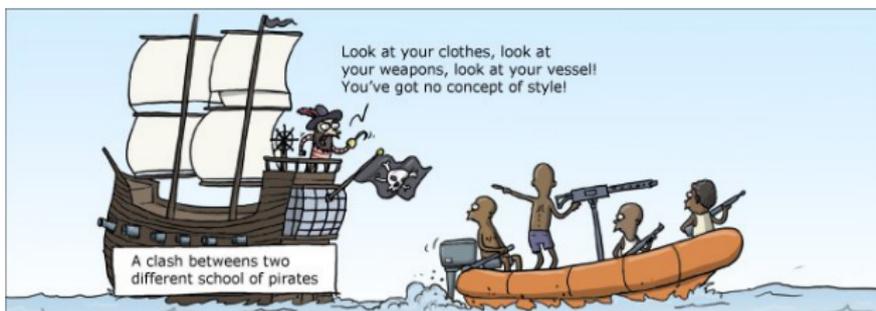
Mécanisme de transactions :

- ▶ Validation ou annulation d'une suite d'opérations liées

Fonctionnalités d'un SGBD (4)

Sécurité et confidentialité :

- ▶ Mise en commun des données
- ▶ Logins et mots de passe
- ▶ Privilèges et droit d'accès
- ▶ Mécanisme de vues



<http://wumo.com/>

Interactions avec le SGBD

Utilisation du SGBD :

- ▶ Interpréteur de commandes
- ▶ Interface graphique
- ▶ Langage de programmation
 - ▶ Java, Python, PHP, C#...
 - ▶ via des bibliothèques pour envoyer des requêtes au SGBD.

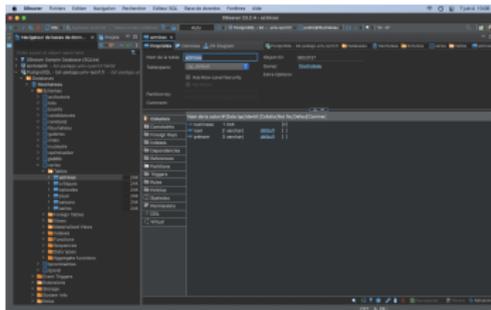
```

fduchateau> \d saisons
Table "series.saisons"
-----
Column          | Type          | Collation | Nullable | Default
-----
id_saison       | integer       |           | not null |
date_lancement | date          |           |          |
nom_serie       | character varying(42) |           |          |

Indexes:
  "saisons_pkey" PRIMARY KEY, btree (id_saison)
Foreign-key constraints:
  "saisons_nomserie_fkey" FOREIGN KEY ("nomserie") REFERENCES series("nomserie")
Referenced by:
  TABLE "episodes" CONSTRAINT "episodes_id_saison_fkey" FOREIGN KEY (id_saison) REFERENCES series.saisons(id_saison)

fduchateau> select * from episodes join saisons using(id_saison);
 id_saison | numero | titre                                     | date_lancement | nomserie
-----
          1 |      1 | The Skank Reflex Analysis                 | 2011-09-22     | The Big Bang Theory
          2 |      1 | The Date Night Variable                  | 2012-09-27     | The Big Bang Theory
          3 |      1 | Winter is coming                         | 2011-04-17     | Game of Thrones
          3 |      2 | The Kingsroad                            | 2011-04-17     | Game of Thrones
          4 |      1 | Pilot                                    | 2014-03-19     | The 100
          5 |      3 | Le Table de Breccan                      | 2000-03-01     | KeanuLott
(6 rows)

```



<https://fr.wikipedia.org/wiki/L4G>

Modèle de données

Un modèle de données définit un mode de représentation de l'information :

- ▶ Un mode de représentation des données (LDD)
- ▶ Un mode de représentation des contraintes sur ces données (LDD)
- ▶ Un ensemble d'opérations pour manipuler les données (LMD)

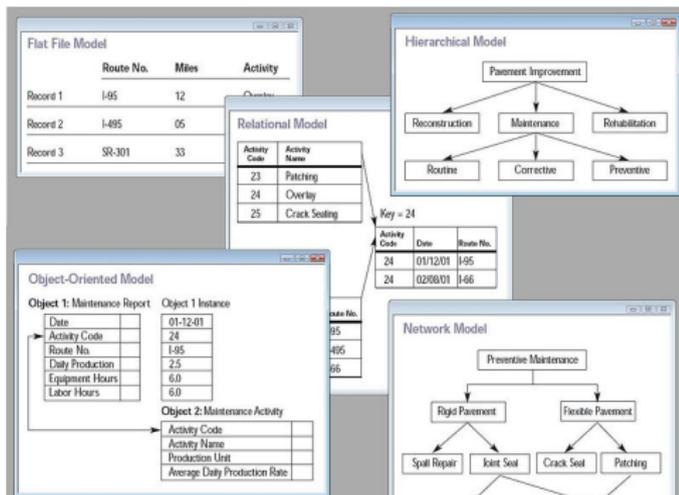
Il est indépendant de la représentation physique des données, ce qui simplifie :

- ▶ L'administration
- ▶ L'optimisation
- ▶ L'utilisation

Modèle de données (2)

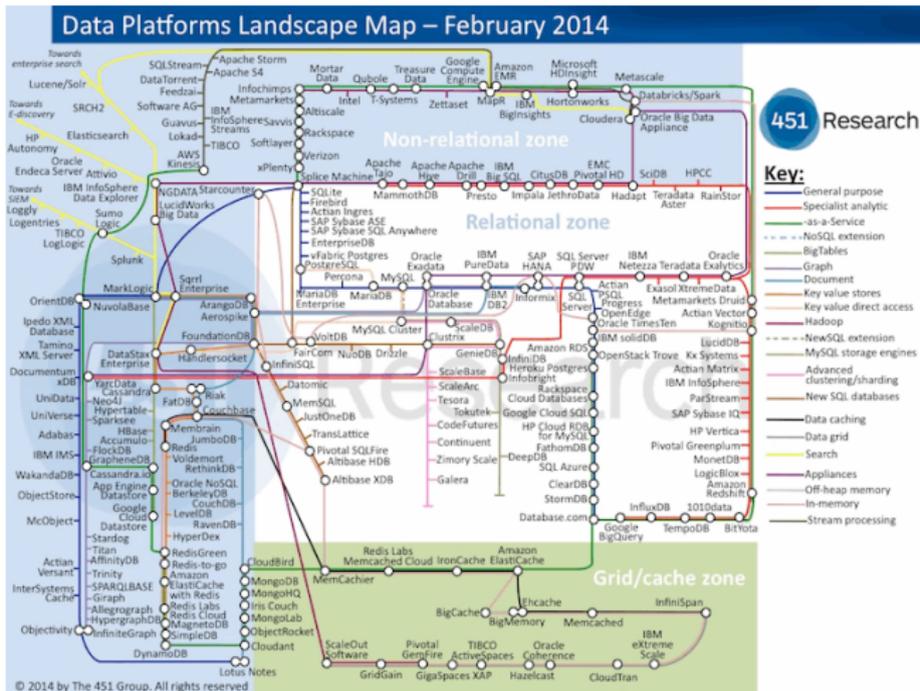
Plusieurs modèles de données :

- ▶ Hiérarchique
- ▶ Réseau
- ▶ Entité-attribut-valeur
- ▶ Relationnel
- ▶ Objet
- ▶ Document
- ▶ En étoile
- ▶ ...



https://en.wikipedia.org/wiki/Database_model

Aperçu des SGBD existants



<https://451research.com/451-research-data-platform-map>

En résumé...

Un Système de Gestion de Base de Données (SGBD) est un logiciel permettant à plusieurs utilisatrices d'accéder à un grand nombre de données

Un SGBD supporte un modèle de données, qui détermine la manière de représenter les informations