

# Bases de données (CS443)

## #1 Introduction, problématiques

Arthur Baudet

Grenoble INP - Esisar

2023–2024



Cours de L. Gonnord, basé sur

- équipe BD de Lyon<sup>1</sup>
- équipe BD de Lille<sup>1</sup>

Avec leur autorisation, of course.

# Contexte

- Nous sommes dans l'ère du BIG DATA
  - Les données sont générées, récoltées, exportées, échangées, vendues
  - Puis mises en entrée d'algorithmes d'analyse ou de modélisation
    - Requêtes, Statistiques, Data Mining, Apprentissage. . .
  - Enjeux commerciaux, scientifiques, sécuritaires considérables - et moyens qui vont avec.
  - Les données sont un bien précieux de chaque organisation
    - Pour le fonctionnement de quasiment toutes les "applications"
    - Pour améliorer l'efficacité et la qualité de l'entreprise
    - Pour guider les évolutions
    - Pour satisfaire à des exigences légales

# Fortes exigences sur les données

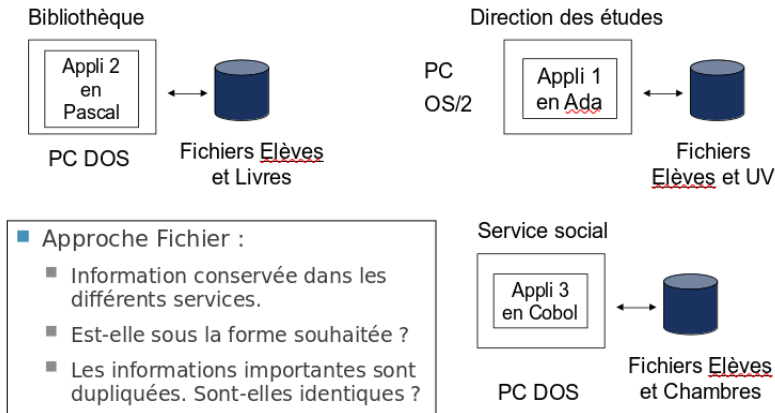
- Stocker les données de façon fiable
- Gérer des accès partagés
- Pouvoir effectuer des recherches complexes et rapides
- être évolutif, réactif, sécurisé
- et bien d'autres choses !

Il faut des outils professionnels : les SGBD

- Exemples : PostgreSQL, Oracle, SQL Server, ...

- ① Au commencement : les fichiers
- ② Maintenant, les Bases de Données
- ③ Conception ?

# L'exemple d'une application



# Problèmes de l'approche (multi)-fichier

- Fichiers de natures différentes, manipulés par des programmes différents, dans des langages différents avec des formalismes différents.
- Difficulté à saisir les liens entre les données : pas de partage de données entre les utilisateurs ; centralisation physique, mais pas logique.
- Dépendance entre les données et les traitements
- Absence de gestion de la sécurité des données
- Pour une nouvelle application, où sont les informations utiles ?

# Comparaison BD/Tableur - à raffiner en TP

Différences sur...	Tableur	Base de données
Utilisation principale	Calculs	Gestion et traitement des données
Structuration des données	Aucune	Structuration et cohérence forte
Contrôles d'intégrité des données	Aucuns	Vérification stricte des valeurs possibles de chaque donnée
Accès aux données	Mono-utilisateur	Multi-utilisateurs
Confidentialité des données	Aucun contrôle	Vérification des droits d'accès de chaque utilisateur
Taille des données	- Une table - Quelques dizaines de lignes	- Plusieurs tables - Plusieurs milliers de lignes par table
Traitement sur les données	Quantitatifs	Qualitatifs et quantitatifs
Interrogations des données	Réalisée par des procédures spécifiques	Langage "universel" : SQL

## ► Analyse en TP.



# Plan

- ① Au commencement : les fichiers
- ② Maintenant, les Bases de Données
- ③ Conception ?

# Abréviations courantes

- SGBD : Système de Gestion de Bases de Données
- DBA : DataBase Administrator

# Définitions

## Définition

Une Base de Données (BD) est une collection de données **électroniques** stockées de façon **pérenne** et reliées entre elles par un **contexte applicatif**.

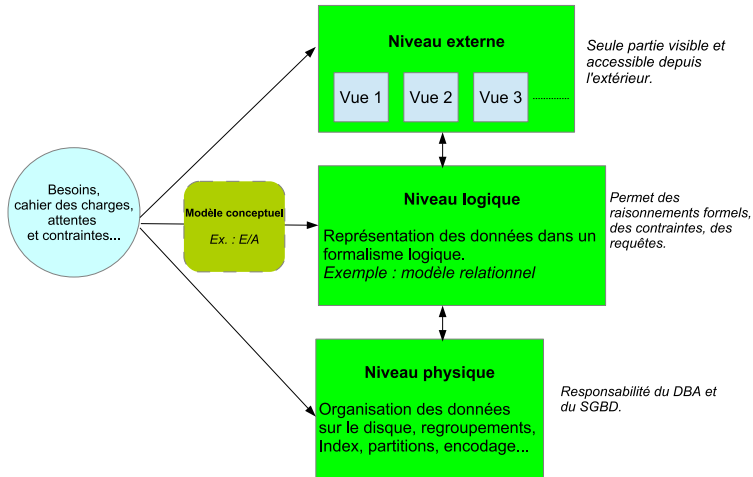
- Quelques exemples :
  - BD de la scolarité de l'université
  - BD des restaurants lyonnais pour la gestion d'un guide
  - BD des ventes d'une entreprise

# Contenu d'une BD

Une BD est toujours justifiée par des besoins !

- Ce sont les besoins applicatifs qui guident le contenu, l'organisation, la portée d'une BD.
- La conception des BD est donc un sous-problème de la conception logicielle. . .
- . . . mais plusieurs logiciels pourront accéder à une même base de données !

# Niveaux (souhaitables) d'abstraction et d'interactions



# Indépendance entre les niveaux

## Indépendance logique

- On peut modifier le modèle logique en conservant les mêmes vues externes.
  - Séparation entre BD et applications
  - Permet des évolutions "indépendantes" des deux niveaux

## Indépendance physique

- On peut modifier des choix physiques en conservant le modèle logique
  - Sémantique conservée, requêtes inchangées
  - Tout en opérant des réglages physiques

Indépendance (très) peu aboutie dans les modèles "NoSQL".

# Résumé : avantage des BD

- Description unique et globale des données
- Langage commun de manipulation.
- Centralisation (pas de duplication, donc pas d'incohérence possible).
- Indépendance entre données et traitements
- Séparation des descriptions logique et physique des données
- Contrôle sémantique des données
- Sécurité (Qui a le droit de lire ou d'écrire, contrôle d'intégrité).
- Partage des données

# Plan

- ① Au commencement : les fichiers
- ② Maintenant, les Bases de Données
- ③ Conception ?



## Une affaire de spécialiste ?

C'est une étape cruciale à considérer avec soin !

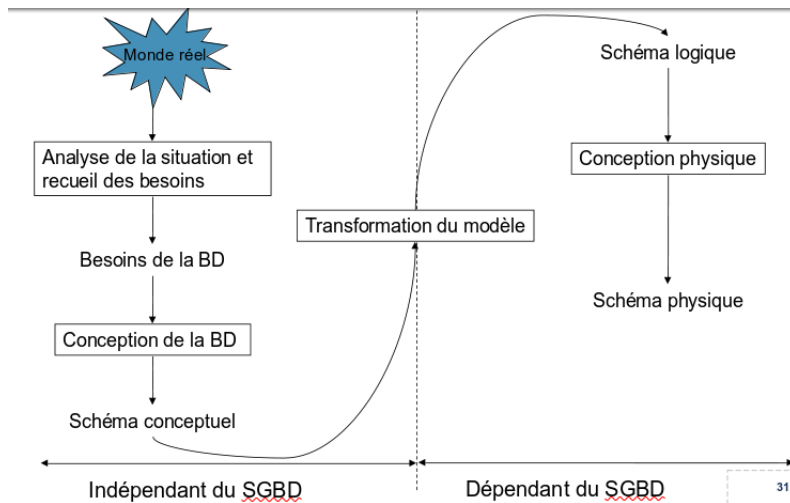
- Conception = modéliser au mieux les données réelles (choix d'un modèle)
- Adaptée aux besoins (réponses aux requêtes)
- Interactions nécessaires avec les experts des données et les concepteurs d'applications
  - Non spécialistes des bases de données
  - Le modèle conceptuel facilitera ces échanges

### Important

Une bonne conception est certes coûteuse, mais nécessaire !

- Pour les applications : correction, performance.
- Pour la gestion des données : faciliter la maintenance, les évolutions.
- Pour la science des données : sémantique des données, les valeurs manquantes : Une BD bien conçue est bien plus facile à analyser

# Étapes de conception



- Conception
- Aspects langages et requêtes
- Nouveau 2023-24 : NoSQL et SGDB non relationnels ?

## Disclaimer

Cours fait par un non-spécialiste. N'hésitez pas à critiquer et reporter d'éventuels bugs/imprécisions/erreurs.

# Organisation du cours

- CM : Arthur Baudet
- TD, TDM : Arthur Baudet
- Les annonces de cours : mailing via Chamilo
- Les mails : [CS443] dans l'objet.

## Supports

- Les transparents : sur Chamilo. **ne pas imprimer**
- Les supports de TD : distribués et sur Chamilo
- Les supports de TP : git dédié.