

BASES DE DONNEES

Éléments de conception et d'exploitation d'une Base de Données

Equipe Bases de Données, F. De Marchi

LYON 1 - UFR Informatique - Laboratoire LIRIS

31 janvier 2021

1 Présentation

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, ...

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.

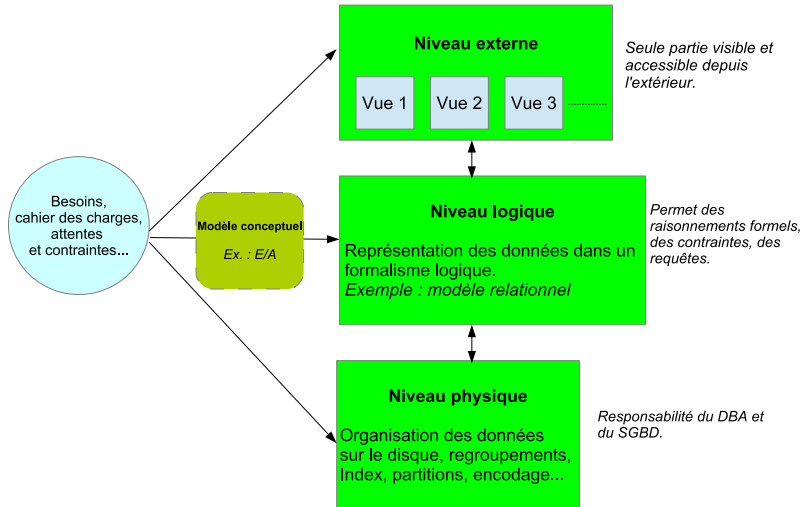
- Quelques exemples :
 - BD de la scolarité de l'université
 - BD des bâtiments de l'université
 - BD des personnels de l'université

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 d'abstraction



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".

Conception d'une base de données

Etape cruciale à considérer avec soins !

- Conception = construire le modèle logique
- De façon à modéliser au mieux les données réelles
- Et pouvoir répondre aux besoins (requêtes)
- Nécessite de nombreuses interactions

Important pour la science des données

- L'analyste doit comprendre la sémantique des données
- L'analyste doit pouvoir détecter des incohérences, des fragilités dans les données
- On a parfois besoin de construire une Base de Données AVANT l'analyse (intégration, nettoyage, . . .)