

Panorama général de la problématique des Bases de Données

L'ensemble des informations dont nous disposons explose littéralement. La croissance du *World Wide Web* a encore accru ce développement, en fournissant l'accès à des bases de données très diverses avec une interface commune.



- **Pourquoi les Bases de Données?**

- Les bases de données ont pris aujourd'hui une place essentielle dans l'informatique, plus particulièrement en gestion.
- Au cours des 30 dernières années, des concepts, méthodes et algorithmes ont été développés pour gérer des données sur mémoires secondaires ; ils constituent aujourd'hui l'essentiel de la discipline « Bases de Données » (BD). Cette discipline est utilisée dans de nombreuses applications.
- Il existe un grand nombre de Systèmes de Gestion de Bases de Données (SGBD) qui permettent de gérer efficacement de grandes bases de données.

Les bases de données constituent donc une discipline s'appuyant sur une théorie solide et offrant de nombreux débouchés pratiques.

1- Définition des Bases de Données

- « Un ensemble organisé d'informations avec un objectif commun »

Peu importe le support utilisé pour rassembler et stocker les données (papier, fichiers, etc.), dès lors que des données sont rassemblées et stockées d'une manière organisée dans un but spécifique, on parle de base de données.



Plus précisément, on appelle base de données un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation (ajout, mise à jour, recherche).

- « Collection homogène et structurée d'informations ou de données qui existent sur une longue période de temps et qui décrivent les activités d'une ou plusieurs organisations »

■ Organisation : une bibliothèque

Données : les livres, les emprunts, les emprunteurs, ...

■ Organisation : une Université

Données : étudiants, enseignants, cours, etc, ...

- Dans une entreprise, il convient de faire appel à l'approche base de données lorsque les données à gérer sont
 - de natures diverses (exemple : étudiants, cours, enseignants, salles, ...)
 - et possèdent de nombreux liens entre elles (exemple : un étudiant suit un cours, un cours est assuré par un enseignant, ...).

- Base de Données Informatisée :

Une base de données informatisée est un ensemble structuré de données enregistrées sur des supports accessibles par l'ordinateur, représentant des informations du monde réel et pouvant être interrogées et mises à jour par une communauté d'utilisateurs.



2- Histoire des Bases de Données

Le terme database est apparu en 1964 par Charles Bachman en General Electric pour désigner une collection d'informations partagées par différents utilisateurs d'un système d'informations militaire. Les premières bases de données hiérarchiques sont apparues au début des années 1960.

En 1969, il créa le modèle de données réseau pour des applications informatiques pour lesquelles le modèle hiérarchique ne convient pas. Charles Bachman a reçu le prix Turing en 1973 pour ses « contributions exceptionnelles à la technologie des bases de données ».

En 1968 Dick Pick crée Pick, un système d'exploitation contenant un système de gestion de base de données « multivaluée » (SGBDR MV).

En 1970, Edgar F. Codd note dans sa thèse mathématiques sur l'algèbre relationnelle qu'un ensemble d'entités est comparable à une famille définissant une relation en mathématiques et que les jointures sont des produits cartésiens. Cette thèse est à l'origine des bases de données relationnelles. Edgar F. Codd a reçu le prix Turing en 1981.

Le modèle Entité-Association a été inventé par Peter Chen en 1975 ; il est destiné à clarifier l'organisation des données dans les bases de données relationnelles.

Apparues dans les années 1990, les bases de données objet-relationnel utilisent un modèle de données relationnel tout en permettant le stockage des objets. Dans ces bases de données les associations d'héritage des objets s'ajoutent aux associations entre les entités du modèle relationnel

Le langage de requête SQL pour les bases de données relationnelles, développé dans le cadre du projet System R d'IBM, est maintenant le langage de requête standard.

SQL a été normalisé à la fin des années 1980, et la norme actuelle, SQL: 1999, a été adoptée par l'American National Standards Institute (ANSI) et l'Organisation internationale de normalisation (ISO).

3- Modèles des Bases de Données

- Les modèles de données correspondent à la manière de structurer l'information dans une base de données.
- Ils reposent sur les principes et les théories issus du domaine de la recherche en informatique et permettent de traduire la réalité de l'information vers une représentation utilisable en informatique.

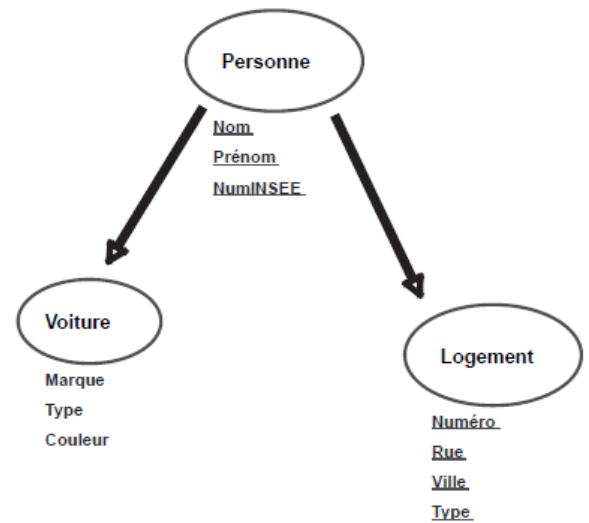
a- Modèle hiérarchique et modèle réseau :

- On manipule des graphes ou des arbres.
- Les nœuds de ces structures constituent les informations et les liens entre ces données les arêtes.
- À ce moment, on n'est pas encore capable de séparer complètement le niveau logique du niveau physique d'un système de bases de données

- **Modèle Hiérarchique**

Le modèle « **hiérarchique** » propose une classification arborescente des données à la manière d'une classification scientifique. Dans ce type de modèle, **chaque enregistrement n'a qu'un seul possesseur** ;

Par exemple, **une commande n'a qu'un seul client**. Cependant, notamment à cause de ce type de limitations, ce modèle ne peut pas traduire toutes les réalités de l'information dans les organisations.



- **Modèle Réseau**

- Le modèle « réseau » est une extension du modèle précédent : il permet des liaisons transversales, utilise une structure de graphe et lève de nombreuses limitations du modèle hiérarchique.
- Dans les deux cas, les enregistrements sont reliés par des pointeurs : on stocke l'adresse de l'enregistrement auquel il est lié.

- **Des SGBDs de type hiérarchique ou réseau sont encore employés pour des raisons d'efficacité lorsque la structure des données s'y prête.**

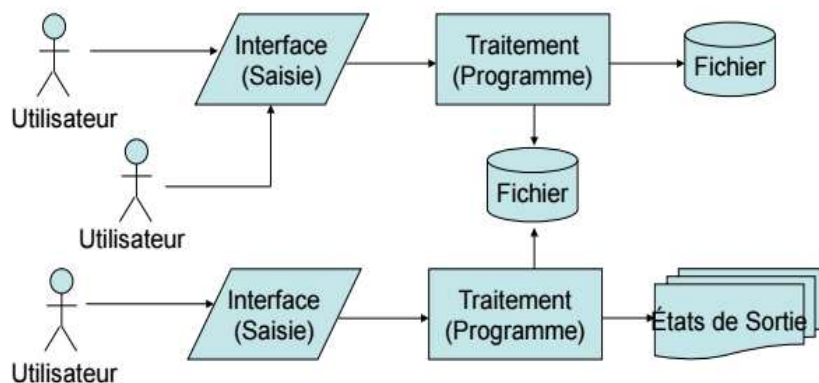
- **On utilise à cet effet des SGBD de conception ancienne, comme**

IMS (Bull) pour le modèle réseau

ou IDMS (Computer Associate).

- **Modèle Relationnel** (sera détaillé par la suite)

4- Organisation des Données en Fichiers



- L'utilisation de fichiers impose d'une part, à l'utilisateur de connaître l'organisation (séquentielle, indexée, ..) des fichiers qu'il utilise afin de pouvoir accéder aux informations dont il a besoin et, d'autre part, d'écrire des programmes pour pouvoir effectivement manipuler ces informations.
- Pour des applications nouvelles, l'utilisateur devra obligatoirement écrire de nouveaux programmes et il pourra être amené à créer de nouveaux fichiers qui contiendront peut-être des informations déjà présentes dans d'autres fichiers.

Limites:

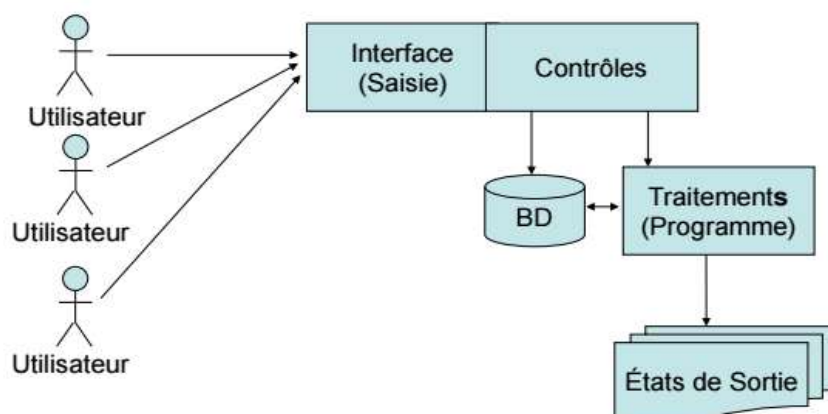
- Non indépendance des données: Saisie et traitement en fonction des fichiers (un ou plusieurs programmes par fichier)
- Redondance de données.
- Risque de problème d'intégrité.
- Pas de contrôle de concurrence entre utilisateurs.
- Manque de sécurité.
- Lourdeur d'accès aux données.

D'où le recours à un logiciel chargé de gérer les fichiers constituant une base de données, de prendre en charge les fonctionnalités de protection et de sécurité et de fournir les différents types d'interface nécessaires à l'accès aux données.

Ce logiciel (le SGBD) est très complexe. En particulier, une des tâches principales du SGBD est de masquer à l'utilisateur les détails complexes et fastidieux liés à la gestion de fichiers. D'où la définition :

« Un SGBD est un logiciel de haut niveau qui permet de manipuler les informations stockées dans une base de données »

5- Organisation en Base de Données (BD)



- Une BD est gérée par un **Système de Gestion de Bases de Données (SGBD)**: Assure la **structuration**, le **stockage**, la **consultation** et la **mise à jour** des données.

Exemple:

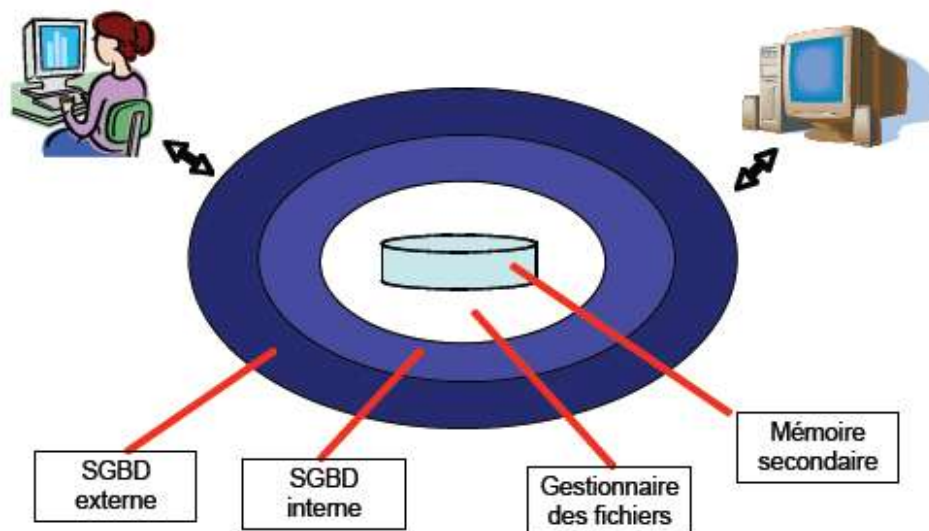
Gestion des étudiants, internat, inscription, cours, notes, emploi du temps, ... d'une école (université).

- **Avantage des Bases de Données**

- Minimum de redondance de données.
 - Partage entre plusieurs applications
 - Contrôle d'intégrité.
 - Indépendances des données et des traitements
 - Sécurité d'accès.
 - Ensemble compact de **tables** contenant les données de la BD. Ces tables sont physiquement des **fichiers**.
- Tout cela grâce au SGBD

6- Système de Gestion de Bases de Données (SGBD)

Système de Gestion de Bases de Données (SGBD)



La gestion et l'accès à une base de données sont assurés par un ensemble de programmes qui constituent le **Système de Gestion de Base de Données (SGBD)**

Un SGBD reçoit des demandes de manipulation du contenu de BD et effectue les opérations nécessaires sur les fichiers qui permettent l'ajout, la modification et la recherche de données.

Un SGBD héberge généralement plusieurs bases de données, qui sont destinées à des thématiques différentes.

Actuellement, la plupart des SGBDs fonctionnent selon un mode client/serveur. Le serveur (la machine qui stocke les données) reçoit des requêtes de plusieurs clients et ceci de manière concurrente.

Le serveur analyse la requête, la traite et retourne le résultat au client.

- **Objectifs des SGBDs**

a- **Indépendance physique** : La façon dont les données sont définies doit être indépendante des structures de stockage utilisées.

b- **Indépendance logique** : Un même ensemble de données peut être vu différemment par des utilisateurs différents. Toutes ces visions personnelles des données doivent être intégrées dans une vision globale.

Vue logique: Représente la vision de l'utilisateur qui ne se soucie pas des détails de stockage et d'accès aux données (*Vue physique*).

Exemple: Vue logique d'une BD de bibliothèque

Il y a trois **tables**: livres, emprunteurs et emprunts contenant les données utilisées (cote, titre, ...)

c- **Accès aux données** : Définition de données se fait par le **Langage de Description de Données LDD** (structure, contraintes, ...).

L'accès aux données se fait par l'intermédiaire d'un **Langage de Manipulation de Données (LMD)**. Ce langage permette d'obtenir des réponses aux requêtes en un temps « raisonnable ». Le LMD doit donc être optimisé, minimise le nombre d'accès disques, et tout cela de façon totalement transparente pour l'utilisateur.

d- **Administration centralisée des données (intégration)** : Toutes les données doivent être centralisées dans un réservoir unique commun à toutes les applications.

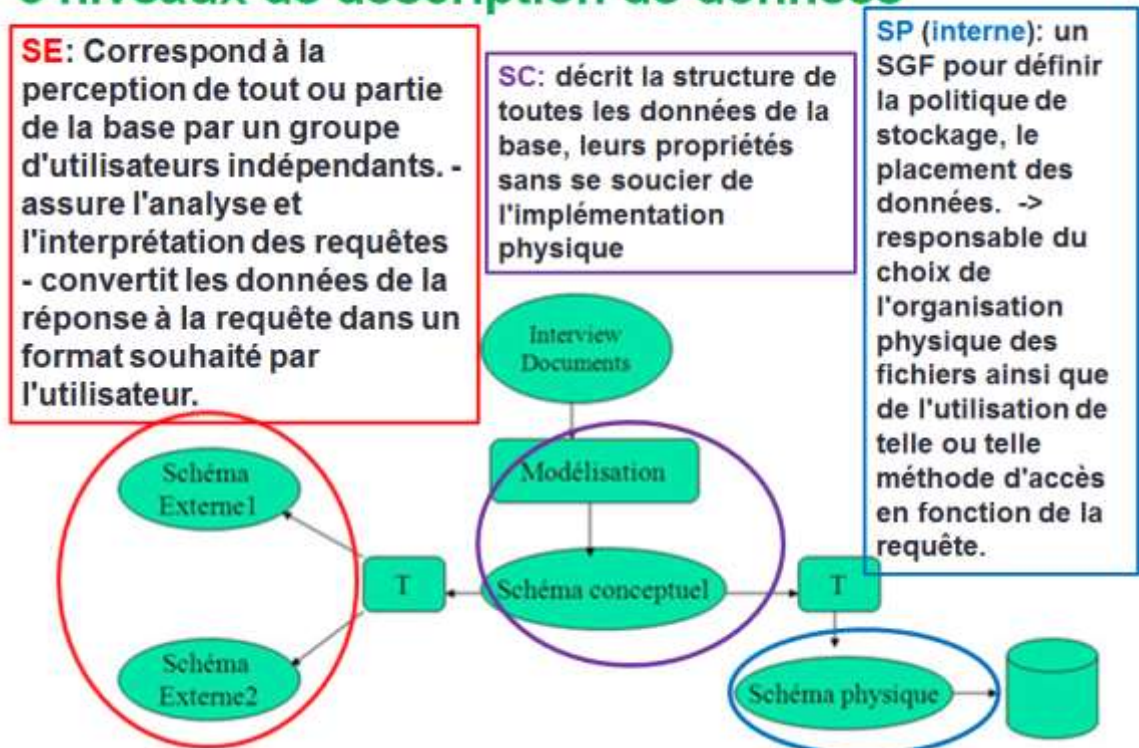
e- **Non-redondance des données** : Afin d'éviter les problèmes lors des mises à jour, chaque donnée ne doit être présente qu'une seule fois dans la base.

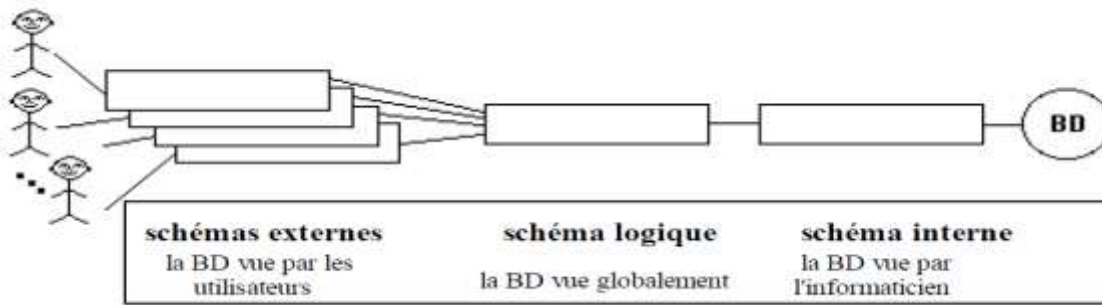
f- **Cohérence des données** : Les données sont soumises à un certain nombre de contraintes d'intégrité qui définissent un état cohérent de la base. Elles doivent pouvoir être exprimées simplement et vérifiées automatiquement à chaque insertion, modification ou suppression des données. Les contraintes d'intégrité sont décrites dans le **Langage de Description de Données (LDD)**.

- g- **Partage des données** : Il s'agit de permettre à plusieurs utilisateurs d'accéder aux mêmes données au même moment de manière transparente.
Si ce problème est simple à résoudre quand il s'agit uniquement d'interrogations, cela ne l'est plus quand il s'agit de modifications dans un contexte multiutilisateur, car il faut : permettre à deux (ou plus) utilisateurs de modifier la même donnée « en même temps » et assurer un résultat d'interrogation cohérent pour un utilisateur consultant une table pendant qu'un autre la modifie.
- h- **Sécurité des données** : Les données doivent pouvoir être protégées contre les accès non autorisés. Pour cela, il faut pouvoir associer à chaque utilisateur des droits d'accès aux données.
- i- **Résistance aux pannes** : *Que se passe-t-il si une panne survient au milieu d'une modification, si certains fichiers contenant les données deviennent illisibles ?*
Il faut pouvoir récupérer une base dans un état « sain ». Ainsi, après une panne intervenant au milieu d'une modification 2 solutions sont possibles :
---- soit récupérer les données dans l'état dans lequel elles étaient avant la modification,
---- soit terminer l'opération interrompue.

- **Description de Données**

3 niveaux de description de données





Schémas externes (SE)

- Schéma externe du professeur de base de données :

Etudiant _BD : nom, prénom, note1, note2, note_finale
 tel que Etudiant _BD résulte de la combinaison de Etudiant et Inscription du SL,
 tels qu'il existe une Inscription de cet étudiant pour le cours BD
 ($n^{\circ}\text{étudiant dans Etudiant} = n^{\circ}\text{étudiant dans Inscription et nom_cours dans Inscription} = \text{BD}$), et
 tel que $\text{note_finale} = (\text{note1} + \text{note2})/2$

Schéma logique (SL)

Etudiant : nom, prénom, date de naissance, $n^{\circ}\text{étudiant}$
 Enseignant : nom, prénom, statut, $n^{\circ}\text{compte_bancaire}$
 Cours : nomC, cycle, nom_enseignant
 Inscription : $n^{\circ}\text{étudiant}$, nom_cours, note1, note2

Schéma interne (SI)

Etudiant : fichier FEtud,
 contenu : nom, prénom, date de naissance, $n^{\circ}\text{étudiant}$
 indexé sur $n^{\circ}\text{étudiant}$,
 index secondaire sur nom+prénom

• Quelques SGBD connus et utilisés

Il existe de nombreux systèmes de gestion de bases de données

* PostgreSQL: <http://www.postgresql.org/>

* MySQL: <http://www.mysql.org/>

* Oracle: <http://www.oracle.com/>

* IBM DB2: <http://www-306.ibm.com/software/data/db2/>

* Microsoft SQL: <http://www.microsoft.com/sql/>

* Sybase: <http://www.sybase.com/linux>

* Informix: <http://www-306.ibm.com/software/data/informix/>

* Access

7- Les Acteurs des Bases de Données

L'administrateur de la BD:

- Créer la BD en précisant la structure ainsi que les détails de stockage sur disque.
 - Contrôler l'accès à la BD afin de le permettre aux applications et aux personnes qui y ont droit.
 - Conserver une bonne performance d'accès aux données (bonne organisation de stockage, index,)
 - Sauvegarder la BD et assurer les reprises après panne.
- Toutes ces tâches sont réalisées à travers des outils du SGBD.

Le programmeur:

Développe des applications qui traitent les données de la base.

L'utilisateur final:

N'a accès qu'aux données qui lui sont utiles:

- Via une application
- Via une interface du SGBD afin d'interroger directement les tables auxquelles il a droit.

8- Cycle de vie d'une Base de Données

1- Conception de la base

Déterminer les **informations** qu'il conviendra de mettre dans la base de données ainsi que sa **structure**.

2- Implantation des données

Transmettre la **structure** de la BD et les contraintes d'intégrité au SGBD choisi (au moyen du LDD)

3- Utilisation

Interroger et mettre à jour les données (au moyen du LMD)

4- Maintenance

Correction, évolution

Etapes de conception de BD

