



Bases de données

Filière : LPE/TMW(SS)

Prof. M. RAHMOUNI
mcl.rahmouni@yahoo.fr

Plan de cours

- ✎ Chapitre 1: Introduction aux bases de données
- ✎ Chapitre 2: Modèle Entité/Association
- ✎ Chapitre 3: Le Modèle Relationnel
- ✎ Chapitre 4: Dépendances Fonctionnelles et Normalisation
- ✎ Chapitre 5: Algèbre relationnelle
- ✎ Chapitre 6: Calcul relationnel
- ✎ Chapitre 7: Le Langage SQL (Structured Query Language)

2

Bases de données

Chapitre 1 : Introduction aux bases de données

Problématique

Dès les débuts de l'informatique, l'un des soucis majeurs de l'utilisateur fut de pouvoir stocker massivement des données et de pouvoir en disposer régulièrement afin d'en extraire de nouvelles informations, de les consulter et de les modifier.

La démarche classique consiste à réaliser pour un type d'information, un programme d'application.

Une entreprise doit conserver un volume élevé d'information:

- noms, adresses, salaire, adresse des fournisseurs, quantités, prix des items, bilan financier, etc.

Ces informations se retrouvent dans différents systèmes de traitement de fichiers.

- Système de gestion des stocks, système de facturation, système de préparation de paie, programme de gestion de personnel, etc.

Exemple d'un système décentralisé:

3

Problématique

Exemple d'un système décentralisé (suite)

Pour obtenir une information, l'employé doit:

1. *déterminer le système à consulter*
2. *trouver la bonne personne concernée.*
 - Perte de temps

De plus, certaines informations sont souvent conservées en plusieurs endroits.

- Duplication de données
- Gaspillage au niveau du volume de fichiers.

8

Introduction aux bases de données

Problématique

Résumons le système décentralisé... Principaux problèmes

de ce système:

- *Redondance de certaines informations*
- *Ne peut répondre rapidement aux demandes d'information provenant de fichiers multiples.*
- *Coûts élevés pour les modifications (plusieurs systèmes)*

Avec le temps, il y aura...

- *Accroissement inutile de:*
 - l'ensemble des fichiers;
 - la taille des fichiers;
 - temps d'accès.
- *Code développé par différents programmeurs et écrits dans différents langages.*
- *Formats de fichiers différents*
- *Inconsistance des données*
- *Lourd à supporter*

8

Introduction aux bases de données

Problématique

Solution : une banque de données commune, entièrement centralisée

➢ Base de données

Idéalement, il devrait y avoir...

- *Un seul exemplaire de chaque élément de données*
- *Tous les utilisateurs ont accès aux données en ne communiquant qu'avec la base (sans intermédiaire).*
- *Mesures de protection pour l'information confidentielle*
- *La complexité du stockage ne doit pas être apparente à l'utilisateur.*

8

Introduction aux bases de données

Un point dans le temps

- 1950-1960
 - Des fichiers séquentiels, du 'batch'
- 1960 – 1970
 - Le début des bases de données hiérarchiques
- 1970 – 1980
 - La naissance du modèle relationnel
- Début des années 90
 - Sql, l'aide à la décision
- Fin des années 90
 - Croissance du volume des données, Internet, modèle multi tiers

8

Les fichiers

- **Définition :** C'est une unité d'informations homogènes portant un nom et physiquement stockée sur une mémoire de masse permanente.
- **Intérêts**
 - Stockage permanent des informations et manipulation de gros volumes de données.
- **Exemples**
 - Une source de programme C, un annuaire de personnes "annuaire.dat" où chaque élément est un triplet <nom, bureau, téléphone>, un fichier image (.jpg), etc.

9

Propriétés d'un fichier :

- Un nom, par exemple « Etudiant.txt », un type, un propriétaire, une date de création, une date de dernière modification, une taille exprimée en octets, un chemin d'accès (localisation physique sur un disque), une description des droits d'accès par les autres utilisateurs.
- Chaque fichier a un contenu, soit une suite ordonnée d'octets, qui peut représenter n'importe quelle donnée binaire déterministe : un programme informatique, un document, un texte, etc.
- Chaque ligne d'un fichier s'appelle article ou enregistrement.

10

Actions sur les fichiers ... -1-

- **Création :** Ecriture et organisation des articles sur leur support physique.
- **Consultation :** Les données sont lues et éventuellement imprimées.
- **Fusion :** Réunir plusieurs fichiers en un seul fichier
- **Eclatement :** l'inverse de la fusion
- **Mise à jour :**

11

Actions sur les fichiers ... -2-

• Mise à jour :

- a) **Ajout :** insertion d'un ou plusieurs article / de nouvelles zones
- b) **Modification :**
 - Par substitution : correction d'une ou plusieurs rubriques d'un article
 - Par addition : addition de nouvelles valeurs dans une zone d'accumulation de l'enregistrement
 - Par soustraction : soustraction de certaines quantités d'une zone d'accumulation d'un egistrement
- c) **Suppression :**
 - D'articles : suppression d'un ou plusieurs enregistrement d'un fichier
 - De zones : élimination de certaines zones dans les articles d'un fichiers qui sont inutiles.

12

Le Système de Gestion des Fichiers

- Le **Système de Gestion des Fichiers (SGF)** est la partie du système d'exploitation permettant de stocker les informations, de les récupérer et de les organiser dans des fichiers.
- **Rôle** : Il gère :
 - Le catalogue des fichiers (emplacement, organisation, droits d'accès),
 - Les opérations d'enregistrement et de lecture (entrées/sorties) dans les fichiers,
 - La sécurité et contrôle des fichiers
 - Les accès concurrents,
 - La récupération de pannes.
- **Principe** :
 - Chaque fichier créé est décrit par un certain nombre de propriétés
 - La description d'un fichier est stockée sur disque dur, dans une zone accessible en lecture et en écriture au SGF

13

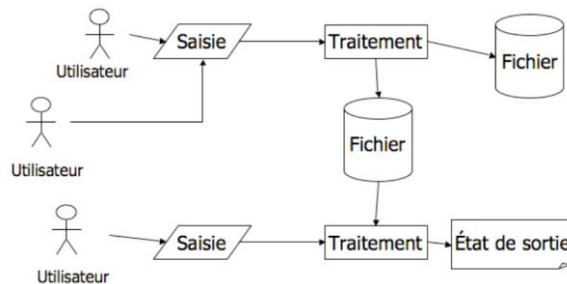
Organisation des fichiers

- » L'organisation des fichiers est la façon dont les données sont organisées sur le support externe. Elle constitue la structure interne du fichier, dont dépendent les algorithmes de lecture/écriture du SGF
- » L'organisation des enregistrements dans un fichier peut être :
Séquentielle, Relative, Aléatoire, Indexée, etc.
- » Dans la plupart des langages de programmation, l'organisation d'un fichier se décide au moment de sa création.
- » **But**: minimiser les E/S (transferts mémoire secondaire (MS)).

14

Les fichiers

» Organisation en fichiers



15

Accès aux enregistrements des fichiers

- Une méthode d'accès est une façon d'accéder à un enregistrement dans un fichier (séquentielle, indexée). C'est le mode d'exploitation d'un fichier dans un programme d'application.
- Elle dépend de l'organisation adoptée
- Trois façons d'accéder aux fichiers
 - Séquentiel : on ne peut accéder à un élément que si on lit tous les éléments qui le précèdent.
 - lire: fichier -> enregistrement (notion d'enregistrement courant)
 - direct (relatif)
 - lire: fichier -> no_enregistrement -> enregistrement
 - indexé (par la clé)
 - lire: fichier -> valeur_clé -> enregistrement

16

Les limites de l'utilisation des fichiers -1-

✎ Travailler directement sur des fichiers présente plusieurs inconvénients :

- Manipulation de données lourdes et compliquées → il faut être expert en programmation.
- Le programmeur doit connaître l'organisation des fichiers (séquentielle, indexé), la localisation physique des fichiers, la structure physique des enregistrements, le mode d'accès aux fichiers.
- Toute modification de la structure des enregistrements (ajout d'un champ par exemple) entraîne la réécriture de tous les programmes qui manipulent ces fichiers.

17

Les limites de l'utilisation des fichiers -2-

Les applications sont :

- rigides,
- contraignantes,
- longues et coûteuses à mettre en œuvre.

Les données associées sont :

- mal définies et mal désignées,
- redondantes,
- peu accessibles de manière ponctuelle,
- peu fiables.

Besoin d'une base de données !!!

18

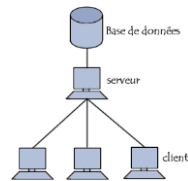
C'est quoi une Base de Données ?

- Une Base de Données (BD) est un ensemble structuré de données (1) enregistrées sur des supports permanents accessibles par l'ordinateur (2) pour satisfaire simultanément plusieurs utilisateurs (3) de manière sélective (4) en un temps opportun (5).

- (1) Organisation et description des données
- (2) Stockage sur disque
- (3) Partage de données
- (4) Confidentialité
- (5) Performance

Exemples :

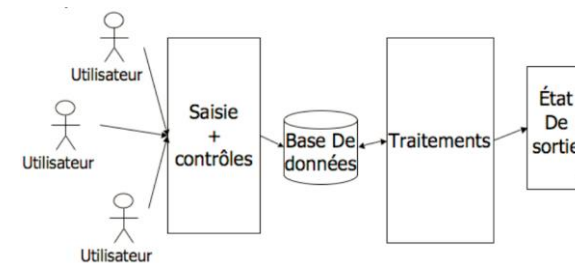
- Gestion des personnels, étudiants, cours, inscriptions, ... d'une université.
- Système de réservation de billets.
- Gestion des comptes clients d'une banque.



19

C'est quoi une Base de Données ?

• Organisation en base de données



20

Pourquoi une BD?

- Pour contourner les limites des fichiers
- **Disponibilité** des données pour une consultation, une saisie ou une mise à jour, tout en s'assurant des droits accordés à ces derniers.
- Limiter la **redondance** des données
- **Accès simultané** à la base par plusieurs utilisateurs
- **Partage et sécurité** de données
- **Fiabilité** de données : Transactions, Reprises sur pannes, Tolérance de pannes
- Langages **assertionnels** de requêtes : SQL, QBE
- Interfaces **conviviales** : 4-GL (Langage de quatrième Génération , ex: générateurs : de formes, de rapports, des applications) & interface Web
- **Locale et répartie**.

21

La gestion des bases de données

- Pour contrôler les données ainsi que les utilisateurs, le besoin d'un système de gestion s'est vite fait ressentir. → La gestion de la BD se fait grâce à un système appelé **Système de Gestion de Bases de Données (SGBD)**.
- Le SGBD est une collection de logiciels permettant de créer, gérer et interroger de manière efficace une BD indépendamment du domaine d'application.
 - Mono ou multiordinateur.
 - En général, peut gérer plusieurs BDs.
 - Peut aussi accéder aux BDs d'autres SGBDs.

22

Historique des SGBD

- **Années 1970** : premiers SGBD commerciaux (hiérarchique, réseau)
 - couplage encore assez fort entre données et programmes
 - difficulté d'accès aux données
- **Années 1970-1980** : premiers SGBD relationnels
 - modèle fondé sur une théorie mathématique
 - langage déclaratif d'accès aux données, "simple" à utiliser
- **Années 1990** : premiers SGBD orientés-objets (OO)
 - intégration de types de données plus divers

23

Fonctions d'un SGBD ... -1-

- 1) Persistance : Données stockées sur disque
- 2) Gestion du disque :
 - Techniques spécifiques pour de bonnes performances
 - Index, hash-coding
 - Regroupement des données sur disque
 - Optimisation des requêtes
 - Cache-mémoire
- 3) Description des données qui seront stockées grâce au Langage de Description de Données (LDD).
- 4) Manipulation des données (ajouter, modifier, supprimer des informations) grâce au Langage de Manipulation de Données (LMD).

24

Fonctions d'un SGBD ... -2-

- 5) Consultation des données et traitement des informations obtenues (sélectionner, trier, calculer, agréger,...) grâce au Langage d'Interrogation de Données (LID).

Ces langages (LDD, LMD, LID, LCD) fournissent des requêtes qui doivent être :

- Simples
- Déclaratives
- Optimisées avant leur exécution

État de l'art :

- SQL (Structured Query Language) : mélange d'algèbre relationnelle et de calcul relationnel
- QBE (Query By Example) : s'appuie sur le calcul relationnel, mais permet de formuler des requêtes et d'effectuer des manipulations de données à l'aide de représentations graphiques
- OQL (Object Query Language) : conçu pour s'intégrer avec des programmes écrits en langage objet (C++, Smalltalk, Java)
- XQUERY (XML Query language)

25

Fonctions d'un SGBD ... -3-

- 6) Partage des données
- Autoriser un accès aux informations à de multiples d'utilisateurs de façon simple.
 - Notion de transaction (begin, abort, commit).
 - Cohérence des mises à jour effectuées par un utilisateur.
 - Cohérence collective : sérialisabilité.
- 7) Sécurité des données
- Les données sont protégées contre les accès non autorisés.
 - Tous les utilisateurs ne peuvent pas tout faire sur toutes les données.
 - Notion de groupes d'utilisateurs.
 - Notion d'autorisation (lecture, écriture, exécution).
 - Granularité des autorisations : base de données, table, colonne, n-uplet, ...
 - Possibilité d'accorder ou de supprimer des droits.

26

Fonctions d'un SGBD ... -4-

- 8) Fiabilité et Contrôle des données :

- Vérification de contraintes d'intégrité sur les données (contraintes de domaines, d'existence,...)
- Résolution des problèmes d'accès multiples aux données (blocages, inter-blocages)
- Atomicité des transactions : transaction complètement effectuée ou pas du tout Résistance aux pannes.
- Prévention des procédures de reprise en cas d'incident (sauvegardes, journaux,...)
 - Si panne mémoire : restauration automatique de la base intégrant les dernières transactions validées avant la panne.
 - Si panne disque : restauration d'une sauvegarde et déroulement du journal archivé.
- Mise en place d'un mécanisme de réplication synchrone de la base dans une base miroir : mirroring

27

Fonctions d'un SGBD ... -5-

- 9) Indépendance logique / physique

- 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.
- L'organisation physique de la BD est transparente pour le développeur d'application et indépendante des structures de stockage utilisées.

28

Concurrence d'accès

- Plusieurs utilisateurs doivent pouvoir accéder en même temps aux mêmes données. Le SGBD doit savoir :
 - Gérer les conflits si les deux utilisateurs font des mises-à-jour.
 - Offrir un mécanisme de retour en arrière si on décide d'annuler des modifications en cours.
 - Donner une image cohérente des données si l'un fait des requêtes et l'autre des mises-à-jour.
- Le but** : éviter les blocages, tout en empêchant des modifications anarchiques.

29

Utilisateurs d'un SGBD / d'une BD



- Utilisateurs finaux / Interactifs (ad-hoc)
 - Cherchent les infos, sans connaître la BD par l'intermédiaire d'interfaces visuelles: 4-GL, Web...
 - A la rigueur peuvent utiliser le langage QBE (Query by Example) ou peuvent interroger directement les tables ou les vues dont ils ont le droit
- Programmeurs d'application
 - Construisent les interfaces pour les usagers interactifs
 - Créent les tables et les structures associées (vues, index,...) utilisées par ses applications.
- Spécialistes de SQL
- L'administrateur de la base DBA
 - définit et maintient la BD
 - a la priorité sur tous les autres usagers
 - Contrôle la BD (permettre l'accès aux données aux applications ou aux individus qui y ont droit)
 - conserve de bonnes performances d'accès à ces données
 - maintient des sauvegardes et des procédures de reprise après les pannes.

30

Niveaux de représentation des données /Architecture d'un SGBD

- L'architecture ANSI/SPARC définit 3 niveaux d'abstraction pour un SGBD:
 - Niveau interne (ou physique)**: Il définit la façon selon laquelle les données sont stockées et les méthodes pour y accéder
 - Niveau conceptuel (ou logique)**: Il définit l'arrangement des informations au sein de la BD sans soucis de l'implantation physique
 - Niveau externe (ou vue)**: Il définit les vues des utilisateurs (c-à-d une description pour chaque utilisateur de sa perception des données)
- Cette architecture permet d'avoir une indépendance entre les données et les traitements :
 - L'existence du niveau conceptuel et interne réels et uniques
 - Le niveau externe permet d'avoir des vues multiples de la base.



31

Avantages de l'utilisation d'un SGBD

- Centralisation des données \Rightarrow intégrité des données
- Contrôle centralisé de l'accès aux données \Rightarrow sécurité accrue
- Instructions de traitement très puissantes \Rightarrow grande rapidité de développement
- Indépendance vis-à-vis de la structure physique et logique des données \Rightarrow facilitée de la maintenance.

Les types de SGBDs

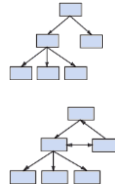
- Les SGBDs peuvent être classés selon les **modèles** des BD (hiérarchique, réseau, relationnel; relationnel objet, objet, ...)
- Les SGBDs les plus connus sont : Oracle, SQLServer, MySQL, PostgreSQL, Microsoft Access, Informix, DB2, Sybase,

32

Les Modèles de BD / types de SGBDs

Un modèle peut se définir comme une représentation abstraite de l'information et éventuellement des opérateurs de manipulation. Plusieurs modèles ont été définis :

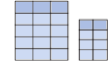
- » **Le modèle hiérarchique** : il a été introduit par IMS d'IBM (1964). Les données sont classées hiérarchiquement, selon une arborescence descendante (structure en arbre). Ce modèle utilise des pointeurs entre les différents enregistrements.
- » **Le modèle réseau** : il a été introduit par IDS2 (1964). Les données sont représentées sous la forme d'un graphe. Ce modèle utilise aussi des pointeurs entre des enregistrements.



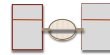
33

Les Modèles de BD / types de SGBDs ... -2-

- » **Le modèle relationnel** : il a été introduit par Codd (1970). Les données sont stockées dans des tables. Le succès de ce modèle auprès des chercheurs, concepteurs et utilisateurs est dû à la puissance et à la simplicité de ses concepts. Dans ce modèle, la manipulation de données se base sur la logique mathématique et la théorie des ensembles. Les SGBD relationnels dominent le marché et le langage SQL est un standard.



- » **Le modèle Entité Association** : il a été introduit par Chen en 1976. L'objectif de ce modèle est de décrire, à travers d'un diagramme, les données qu'on souhaite intégrer dans la base ainsi que les relations qui les lient.

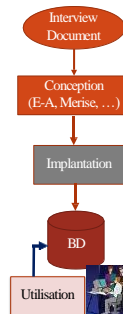


- » **Le modèle Objet, Relationnel-Objet, ...**

34

Cycle de vie d'une BD

- On appelle **conception d'une BD** la phase d'analyse qui aboutit à déterminer le futur contenu de la base, c-à-d décider les données qu'on va stocker dans la BD.
- Une fois que cette phase est terminée, il faudra transmettre son contenu au SGBD choisi (à travers le LDD)
- On appelle **implantation de la BD** cette phase qui consiste à décrire la BD dans le langage du SGBD et introduire les données initiales.
- Une fois l'implantation terminée, on peut commencer l'**utilisation** de la BD. Celle-ci consiste à exprimer des requêtes de m-à-j (ajout, suppression, modification, ...), des requêtes d'interrogation.
- On appelle cycle de vie d'une BD la suite des phases: **conception, implantation, utilisation.**



35

CONCEPTION DES BASES DE DONNEES

Les BD constituent le cœur du système d'information. La conception de ces bases est la tâche la plus ardue du processus de développement du système d'information.

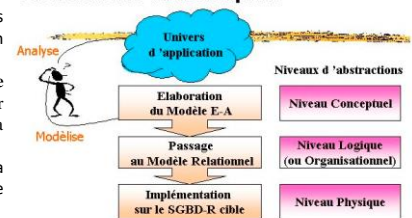
La modélisation se réalise en trois étapes principales qui correspondent à trois niveaux d'abstraction différents :

I. Niveau conceptuel : On appelle **conception d'une BD** la phase d'analyse qui aboutit à déterminer le futur contenu de la base, c-à-d décider les données qu'on va stocker dans la BD.

II. Niveau logique relationnelle : résulte de la traduction du schéma conceptuel en un schéma propre à un type de BD.

III. Niveau physique : est utilisé pour décrire les méthodes d'organisation et d'accès aux données de la base.

La démarche de Conception



36