

Fondements des bases de données

SECTION :LGSSI 1

Informations générales

- Enseignant
 - Rahma Bouaziz
 - rahma.bouazizkammoun@gmail.com
- Support du cours
 - Google Classroom

75umrpr

Chapitre 1

GESTION DES FICHIERS

I. Introduction

- Les raisons principales pour le développement des unités de stockage externe des données:
 - La nécessité de stocker les données ou les informations d'une manière permanente.
 - La volatilité de la mémoire centrale d'un ordinateur.
- Ce stockage doit suivre un certain nombre de techniques et organisations permettant ainsi de retrouver facilement les données qui ont été déjà stockées pour les utiliser ultérieurement.

II. Définitions

- **Enregistrement** Un enregistrement d'un fichier contient des informations relatives à un même individu ou un même objet appartenant à ce dernier. Il est formé de plusieurs champs ou rubriques. Chaque champ peut être de type numérique, alphabétique, numérique, date, ...
- **Fichier** C'est un ensemble d'enregistrements stocké en mémoire secondaire (disque, bande magnétique, ...).
- **Indicatif ou Clé d'indicatif ou la clé d'un enregistrement** est un champ significatif pour l'enregistrement permettant ainsi d'identifier un et un seul individu dans un fichier.

Exemple : n° inscription, n° cin, n° bon, ...

III. Opérations sur les fichiers

• 1. Opérations concernant tous les articles

- **Création** nécessite comme paramètres : le nom du fichier, le type (texte, structuré, etc.).
- **Destruction** déclare la fin du cycle de vie d'un fichier (suppression totale d'un fichier).
- **Tri** Deux paramètres nécessaires pour l'opération de tri, qui sont : un critère et un ordre.

Exemple :Etant donné l'enregistrement Etudiant définit comme suit :
Etudiant CIN NINSC Nom Prénom Adresse Critère : NINSC Ordre :
ordre croissant

- **Réunion** de deux fichiers F1, F2 donne comme résultat F1/F2 + F2/F1 + F1 union F2
- **Fusion** obtenir un fichier qui contient tous les enregistrements de deux fichiers en entrées F1 et F2.
- **Eclatement** d'un fichier avec les critères d'éclatement. Le résultat de cette opération est deux fichiers différents.

III. Opérations sur les fichiers

2. Opérations concernant un seul enregistrement

- **Ajout ou création** d'un enregistrement .le programme doit vérifier s'il est déjà existant ou non. Cette vérification se fait sur le(s) champ(s) identificateur(s).
- **Consultation** C'est une opération de lecture. Le résultat peut être visualisé sur l'écran ou imprimé sur papier.
- **Mise à jour** un enregistrement signifie modifier le contenu d'un ou plusieurs champs de cet enregistrement.
- **Suppression** peut être logique ou physique.
 - Une suppression logique signifie le changement d'état d'un indicateur (0 : supprimé, 1 : existant).
 - Une suppression physique signifie la destruction totale de l'enregistrement.

IV. Les organisations

- L'**organisation d'un fichier** est la manière d'implantation physique du fichier sur le support (magnétique, optique, ...). On distingue principalement quatre types d'organisations :
 - **Organisation séquentielle** Les enregistrements sont rangés les uns à la suite des autres, dans l'ordre où ils sont présentés par l'utilisateur au système.
 - **Organisation directe** Les enregistrements sont rangés dans des emplacements où les adresses sont calculées à l'aide d'une fonction de hachage appliquée à la clé. Ce type d'organisation permet de retrouver rapidement l'enregistrement sans avoir passer par les précédents.
 - **Organisation séquentielle indexée** : associer à la clé d'un article son adresse relative à l'aide d'une table de matière appelée index.
 - **Organisation en fichier inversé** (Multi – clés) un fichier inversé est la liste des enregistrements (ou références aux enregistrements) organisés selon les valeurs de certains champs ou attributs ou rubriques. Un fichier inversé sur une propriété est un fichier indiquant les éléments pour lesquelles cette propriété est vraie dans le fichier de données.
- L'accès à un enregistrement d'un fichier est la manière avec laquelle on recherche cet enregistrement dans un support, on distingue :
 - Accès séquentiel
 - Accès direct

Chapitre 2

SYSTEME DE GESTION DE BASE DE DONNEES

I. Introduction

- De nos jours, la prise de décision demande une information complète, précise et rapide, généralement suivie de multiples interrogations. Un système de gestion de fichier (SGF) n'offre pas des services tels que l'interrogation (affichage) à l'écran, des aspects de prévention pour le conflit d'accès, des procédures de sauvegarde (reprise sur incident),...
- Le système de gestion de base de données (SGBD) n'a pas été inventé pour remplacer le système de gestion de fichier (SGF) associé au système d'exploitation, mais il apparaît comme un complémentaire nécessaire au (SGF).
- Les SGBD se distinguent clairement des systèmes de fichier par le fait qu'ils permettent la description des données (définition des noms, format et caractéristique) de manière séparée de leur utilisation (mise à jour et recherche).

II. Définitions

1. Base de données (BD) :

- Nous appelons BD l'ensemble des données, qui peuvent être manipulées par plusieurs utilisateurs ayant des vues différentes de ces données.
- Une BD doit satisfaire les critères suivants :
 - **L'exhaustivité** : La base de données contient 100% des données sur le sujet.
 - **La non-redondance** : Une donnée n'est stockée plus qu'une fois.
 - **La structuration** : Les données sont stockées tel que leurs traitement sera efficace.

II. Définitions

1. Base de données (BD) :

- La base de données doit satisfaire un certain nombre de contraintes possibles, pour contenir la réponse à toutes les questions possibles, elle doit être exhaustive, c'est à dire contient toute information sur le sujet à traiter. Pour que les informations soient valables, il faut qu'elles soient cohérentes, pour permettre leurs recherche, et structurées, c'est à dire organisées en groupes et sous-groupes avec des relations clairement définies.

II. Définitions

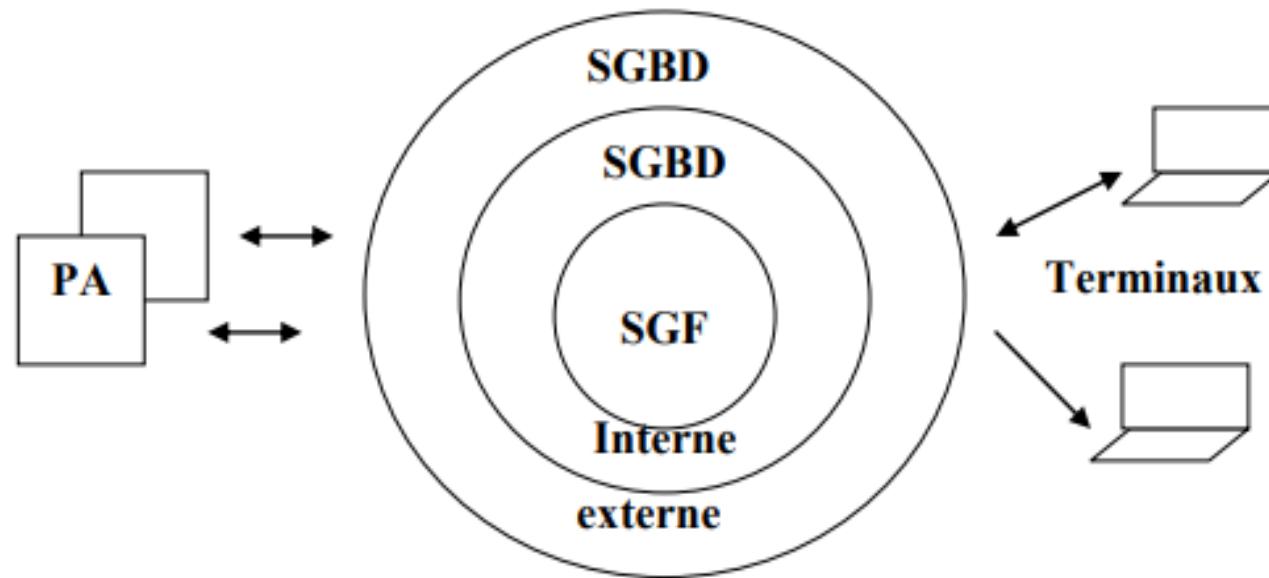
2. Système de gestion de base de données (SGBD)

- Un SGBD est un outil permettant **d'insérer, de modifier et de rechercher** efficacement des données spécifiques dans une grande masse d'informations (quelques milliards d'octets) partagées par tous les usagers.
- un SGBD peut apparaître comme un outil de rangement, de recherche, d'assemblage et de conversion des données.
- Un SGBD se compose de trois couches successives des fonctions, empilées depuis les mémoires secondaires vers les usagers :
 - La gestion des récipients de données sur mémoire secondaire compose la première couche, c'est le système de gestion de fichiers (SGF), il doit permettre la création, l'ouverture, la fermeture, la sauvegarde et la destruction d'un fichier.
 - La gestion des données stockées dans les fichiers, le placement et l'assemblage de ces données, la gestion des liens entre données et des structures permettant de les retrouver rapidement constituent la deuxième couche, c'est le système d'accès aux données (SGBD interne).
 - La présentation des données aux programmes d'application et aux usagers terminaux ayant exprimé leurs besoins en données à l'aide de langages plus ou moins élaborés constitue la fonction essentielle de la troisième couche (SGBD externe). Cette couche assure d'une part l'analyse et l'interprétation des requêtes des usagers, d'autre part, la mise en forme des données échangées avec le monde extérieur.

II. Définitions

- 2. Système de gestion de base de données (SGBD)

La figure suivante illustre ces trois couches de fonctions constituant un SGBD.



PA : Programme d'application

Figure 1 : Présentation d'un SGBD (3 couches)

III. Historique des SGBD

- **Les années 60** ont connu un premier développement des systèmes de fichiers :
 - apparaître les mémoires secondaires qui composent aujourd’hui le cœur du SGBD.
- **Le milieu des années 60** a vu la naissance de la première génération des SGBD :
 - la séparation de la description des données, des programmes d'applications et l'avènement des langages d'accès navigationnels.
 - visant essentiellement à optimiser les méthodes de placement des données sur les mémoires secondaires afin de réduire les temps d'accès.
- **Depuis 1970** La deuxième génération de SGBD a grandi dans les laboratoires, à partir du modèle relationnel.
 - simplifier l'accès aux données pour les utilisateurs externes
 - C'est le système qui doit déterminer le meilleur plan d'accès possible.
- **Le début des années 80** Les premiers systèmes de la deuxième génération sont commercialisés
 - une troisième génération de SGBD basée sur les langages d'accès plus puissants et plus naturels, supportent des types de données plus variés, incluant des possibilités de déduction et de répartition, est à l'étude dans les laboratoires de recherche.

IV. Avantages des bases de données

- Une base de données présente les avantages suivants :
 - **Facilité et rapidité des développements d'applications**
 - SGBD pour stocker et retrouver des informations de manière transparente.
 - La protection contre les conflits d'accès
 - Le retour sur train de transaction : après un service de mise à jour, on a la possibilité de le confirmer ou non.
 - La confidentialité des accès
 - **Facilite les évolutions d'applications**
 - L'ajout d'une nouvelle zone à un fichier déjà existant
 - La modification d'une zone

V. Objectifs d'un SGBD

- **1. Intégration et corrélation**

Dans les systèmes d'informations classiques, chaque application gère ses propres données ou fichiers. Dans toute entreprise, modélisée avec cette approche (classique), il y aura autant de couples (applications – fichiers) qu'il y a d'utilisations différentes de données. Il existe un risque d'avoir une redondance et non-compatibilité de données.

Il en résulte plusieurs inconvénients :

- Redondance et danger d'incohérence de données : la même donnée peut appartenir à plusieurs applications.
- La difficulté de mise en œuvre d'un traitement non prévu. Dans l'approche BD, un réservoir commun de données est constitué, il représente une modélisation aussi fidèle que possible de tout le réel organisationnel de l'entreprise. Cet ensemble de données uniques devra être très fidèle à l'univers réel & global modélisé, puisque toutes les recherches de l'information vont en dépendre : c'est l'objectif de corrélation.

V. Objectifs d'un SGBD

- **2 Flexibilité**

Dans les systèmes classiques, un changement de supports, des méthodes d'accès physiques, etc, entraîne une re-programmation (ré-écriture) des applications correspondantes. Dans l'approche BD, on peut isoler le niveau utilisation (applications) de toute modification intervenant au niveau de :

- L'organisation physique de donnée : changement de support ou de la méthode d'accès.
- L'organisation logique de donnée : la vue d'un utilisateur est indépendante d'autres vues externes).
- Du chemin d'accès des données : création de nouveaux indexes, clé inversée, etc

V. Objectifs d'un SGBD

- 3 Disponibilité

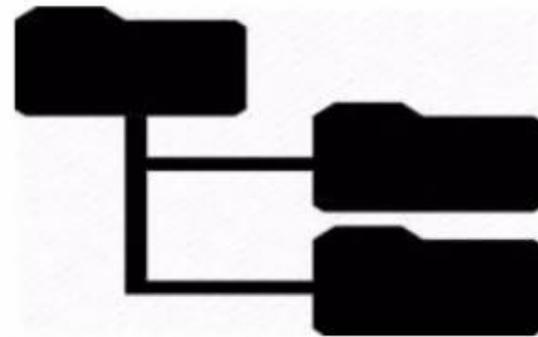
Tout utilisateur doit ignorer l'existence d'utilisateurs concurrents de cette base .

- 4 Sécurité

On va parler de deux notions de sécurité :

- Intégrité de données ou protection contre l'accès invalide aux données (erreurs ou pannes) et contre l'incohérence des données, ce sont les contraintes d'intégrité.
- Confidentialité : contre une modification ou consultation illégale des données, ce sont les droits d'accès.

Système de base de données vs système de fichiers



Consistance des données	Redondance des données
Accès facile aux données	Difficulté d'accès aux données
Contrainte d'intégrité	Problème d'intégrité
Mécanisme d'autorisation	Problème de sécurité
Gestion des données plus faciles grâce aux requêtes fournies par le SGBD	Difficulté de gestion des données

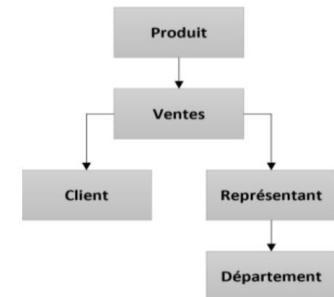
VI. Types des SGBDs

- 1 Modèle hiérarchique

Historiquement, les SGBDs de type hiérarchique sont apparus les premiers sur le marché.

Les données sont représentées sous forme **d'une structure arborescente** d'enregistrements.

C'est un arbre ordonné : chaque nœud sauf la racine a un seul arc incident de type (1 : N) et un ou plusieurs arcs émergents (1 : N). (1: N) : un père pour plusieurs fils.



VI. Types des SGBDs

- **2 Modèle réseau**

La structure de données est présentée sous la forme **d'un graphe**, d'un réseau connectant les entités entre elles.

Ce modèle permet la représentation N : M interentités.

Pour les modèles hiérarchique et réseau, les programmes ne sont pas indépendants de la structure logique de la base et du chemin d'accès aux données : ils doivent décrire comment retrouver les données et si, par exemple, on enlève un index, tous les programmes qui l'utilisaient doivent être réécrits. De plus les langages qui décrivent les bases de données réseaux ou hiérarchiques sont complexes.



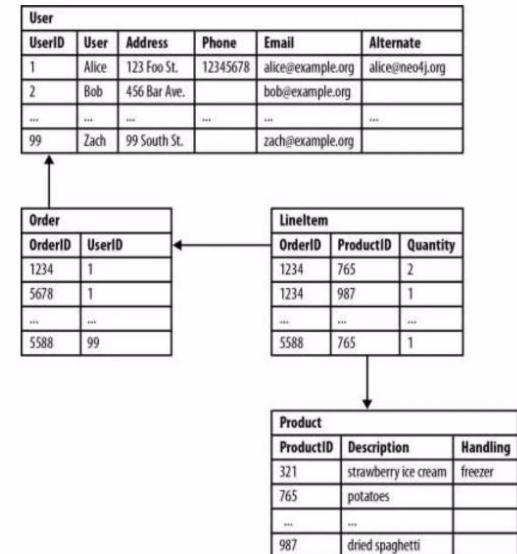
VI. Types des SGBDs

- **3 Modèle relationnel**

Il est fondé sur la théorie mathématique des relations et conduit à une représentation très simple des données sous forme **de tables** constituées **de lignes et de colonnes**.

Au modèle relationnel est associée une théorie qui ne peut être séparée du modèle : la théorie de la normalisation des relations qui a pour but :

- éliminer les comportements anormaux des relations lors de la mise à jour.
- éliminer les données redondantes
- mieux comprendre les relations sémantiques entre données.



VI. Types des SGBDs

- 3 Modèle relationnel - Les principaux SGBD

Nom	Commentaires
Oracle	Il s'agit d'un environnement de développement complet comportant notamment un noyau de SGBD relationnel puissant.
DB2	C'est un SGBD relationnel développé par IBM.
MySQL	C'est un SGBD relationnel appartenant à la famille des logiciels libres.
Postgres	C'est un SGBD relationnel appartenant à la famille des logiciels libres, offrant plus de fonctionnalités que MySQL.
Access	Commercialement présenté comme SGBD relationnel. Il appartient à la suite MS Office.
SQL Server	C'est un SGBD relationnel développé par Microsoft pour succéder à Access pour de grosses applications.
...	

VI. Types des SGBDs

Exemples :

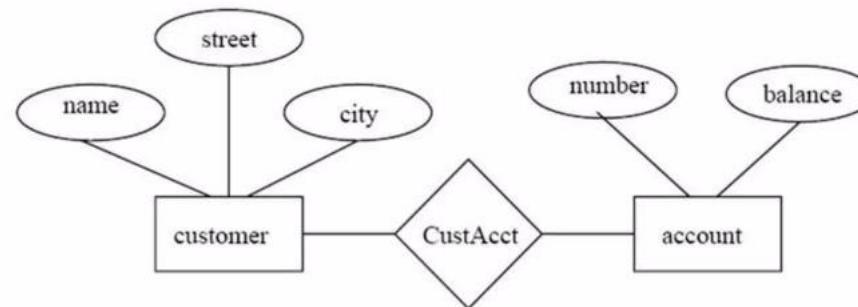


VI. Types des SGBDs

- **4 Modèle objet**

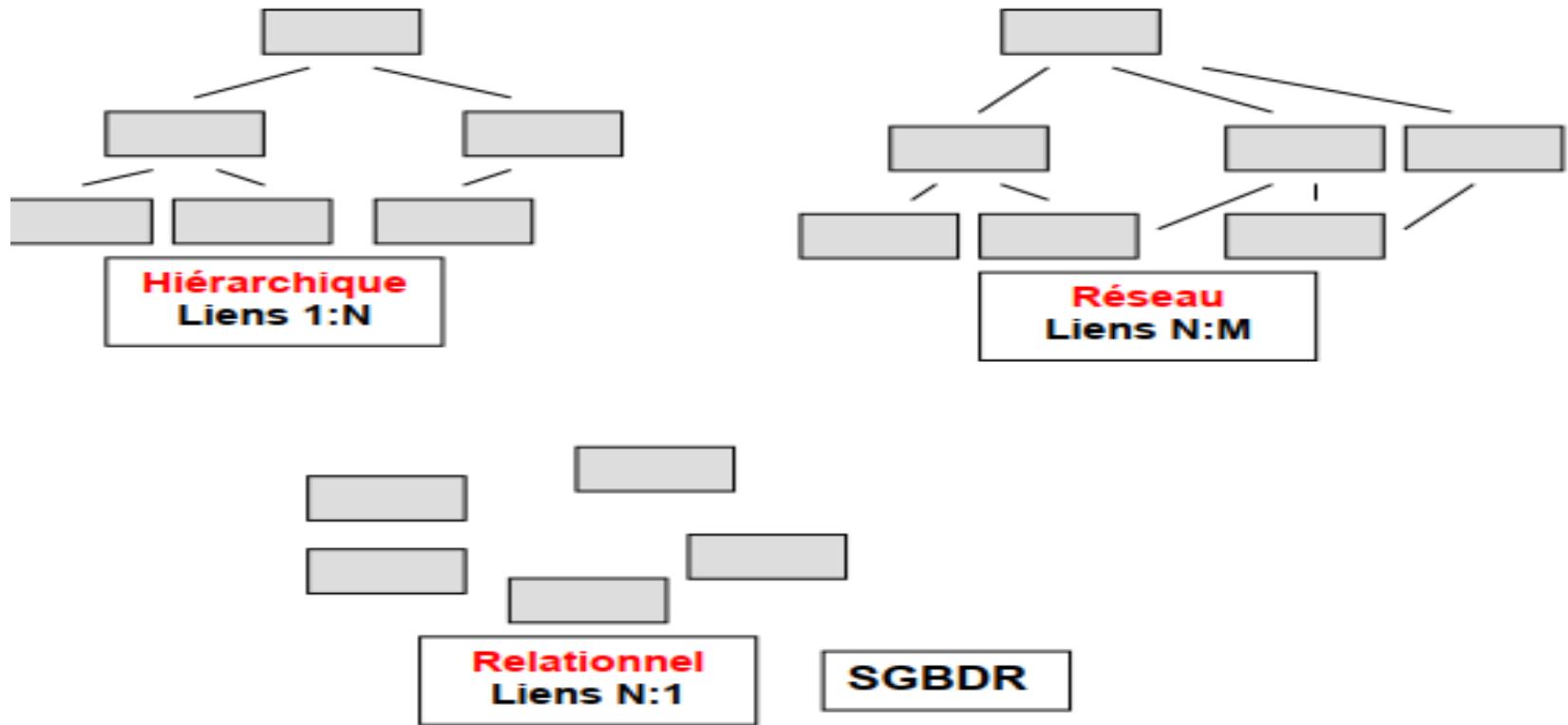
C'est un nouveau modèle des SGBDs.

Les données sont représentées sous forme **d'objets** au sens donné par les langages orientés objet : les données sont enregistrées avec les fonctions qui permettent de les manipuler.



Les différents modèles de données

- L'organisation des données au sein d'une BD a une importance essentielle pour faciliter l'accès et la mise à jour des données



VII. Conclusion

- Pour remédier aux inconvénients des approches classiques de gestion de données (fichier), les bases de données sont venues remplacer cette ancienne approche, c'était en 1960. dès cette date et jusqu'au nos jours, l'évolution des bases de données a vu l'apparition de trois générations :
 - BD hiérarchiques, réseau
 - BD relationnel
 - BD objet
- Un système de gestion d'une base de données doit permettre:
 - la description , la manipulation d'une BD
 - l'intégration corrélation, la flexibilité, la disponibilité et la sécurité de données.