

LES BASES DE DONNES

1. Introduction :

L'informatique évolue vers le traitement de masses d'informations de plus en plus grandes dans des environnements répartis géographiquement où doivent cohabiter des matériels hétérogènes. Dans ce contexte, les bases de données sont utilisées de façon intensive pour de nombreux domaines d'application tels que le domaine médical, les administrations ou les associations. Les applications concernées par l'utilisation d'un SGBD possèdent des caractéristiques différentes tant au niveau du volume de données concernées qu'au niveau de la complexité de ces données et des traitements informatiques à réaliser. Néanmoins, le regroupement des données dans une base de données gérée par un système de gestion de base de données apporte de nombreux avantages dans la plupart des cas d'utilisation. De manière intuitive, il est possible de définir une base de données de la façon suivante :

2. Une base de données (BD) est un ensemble structuré de données persistantes, représentant une partie du monde réel, servant de support à des applications informatiques et qui peut être interrogeable par le contenu, ces données sont mémorisées sur des supports accessibles par un ordinateur pour satisfaire simultanément plusieurs utilisateurs de façon sélective et en temps très court. Elles constituent le cœur du système d'information.

2.3 La démarche de conception d'une base de données

2.3.1 Principe général

Il est communément reconnu que la conception d'une base de données doit se faire en utilisant une méthode de conception qui définit la démarche à suivre. Plusieurs méthodes de conception existent à cet effet et nous citerons comme exemple la méthode MERISE. Pour certaines méthodes, on dispose même d'un outil logiciel d'aide à la conception appelé aussi un Atelier de Génie Logiciel (AGL) constitué d'un ensemble de logiciels permettant l'automatisation d'un certain nombre de tâches lors des différentes phases du processus de conception (génération automatique de la structure de la B.D., de programmes d'accès et de manipulation, etc.). Néanmoins, quelle que soit la méthode utilisée, la conception d'une base de données passe par un processus de modélisation permettant de modéliser une certaine partie du monde réel afin de caractériser les entités qu'on manipule (étudiants, Comptes Bancaires, Ouvrages, etc.). De plus on essaye de caractériser les attributs de ces entités en fonctions des problèmes que doit résoudre l'existence de la B.D. : Gestion de la scolarité, Gestion de Prêt d'ouvrages dans une bibliothèque, etc. Le cas le plus général est celui où la B.D. est partagée par plusieurs utilisateurs. Ces utilisateurs n'ont pas tous la même vue des données de la base, et n'ont pas tous à voir la base dans sa totalité car chaque utilisateur n'est concerné que par une partie de celle-ci.

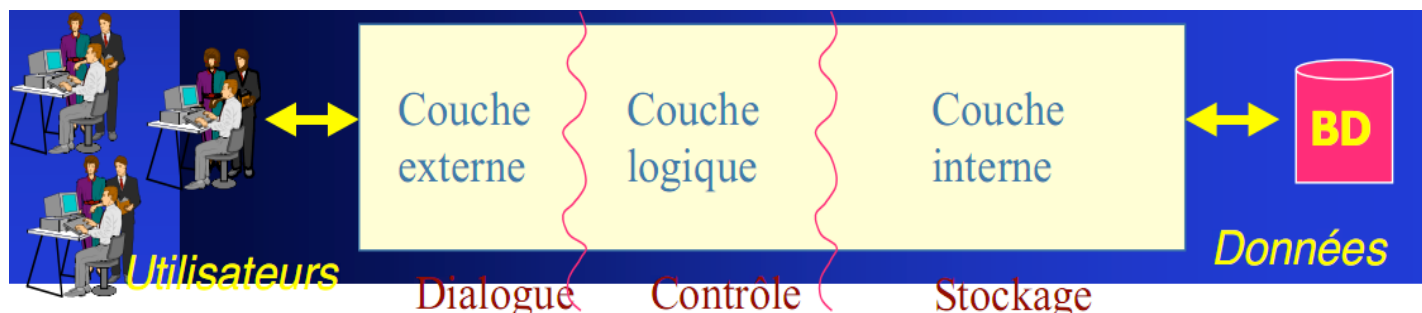
2.3.2 Niveaux de description d'une Base de Données

La description d'une base de données peut se faire à différents niveaux, suivant que l'on regarde plus du côté de l'utilisateur que du côté du stockage des données sur les supports physiques. On distingue communément trois (3) niveaux de description d'une B.D. :

A- niveau conceptuel

B- niveau externe

C- niveau interne



A. Niveau conceptuel

Le schéma conceptuel est la partie fondamentale dans l'architecture d'une base de données. Il a pour but de décrire en termes abstraits mais fidèles une certaine réalité d'une organisation et de ses processus de gestion qui ont nécessité la mise en place d'une B.D.

Le passage du monde réel au schéma conceptuel correspond à un processus de modélisation où les objets du monde réel ayant les mêmes caractéristiques sont classés en catégories et désignés par des noms (Etudiants, Véhicules, etc.). Le SGBD fournit un langage de description qui permet de spécifier le schéma conceptuel.

Dans le processus de modélisation, le concepteur de la B.D. spécifie le schéma conceptuel en utilisant les possibilités offertes par un modèle de données.

A.1 Définition d'un modèle de données

Un modèle de données est un outil formel destiné à décrire la réalité de manière indépendante de tout traitement informatique.

Un modèle de données doit permettre de regrouper les objets du monde réel auxquels on s'intéresse en classes d'objets de nature identique.

A.1.2 Classification des modèles

Un schéma conceptuel est donc le résultat d'un processus de modélisation fait en respectant les possibilités d'un modèle de données. Le modèle de données est une caractéristique de tout SGBD. Il existe trois grandes classes de modèles de données qui se distinguent par la nature des associations qu'ils permettent de modéliser. Ce sont : Les modèles hiérarchiques, Les modèles réseaux, Les modèles relationnels.

A.1.3 Historique sur les modèles BD

Fichier indexé : fichier permettant un accès direct à un enregistrement en fonction d'une clef, qui est un sous-ensemble des attributs de l'enregistrement. Accès très rapide à un enregistrement pour la mise à jour et l'interrogation. Peu souple pour l'interrogation à partir d'un attribut autre que la clef, ou pour joindre plusieurs fichiers ensemble. Utilisé aujourd'hui dans l'implémentation des bases de données relationnelles et objets.

BD hiérarchique : stockage efficace des données pour favoriser certains accès, en utilisant les liens de hiérarchie entre les données. Moins souple pour l'interrogation des données et la navigation à travers les données. Encore en opération aujourd'hui dans plusieurs grandes organisations (banques, assurances, gouvernement), à cause de leur efficacité, mais aussi à cause des investissements majeurs requis pour les remplacer.

BD réseau : les liens entre les données sont prédéterminés et spécifiés dans le schéma de base de données. Navigation plus souple que dans les BD hiérarchiques, mais moins flexible que celles des BD relationnelles.

BD relationnelle : les données sont regroupées sous formes de relations (tables). La navigation est très souple; les liens entre les tables ne sont pas prédéfinis; ils sont établis de manière dynamique, directement dans la requête d'interrogation. En contrepartie, si les requêtes ne sont pas optimisées, elles peuvent être très longues à exécuter.

BD orientée objets : les données sont représentées par des objets. Plus souple que les BD relationnelles au niveau des structures de données admises. On peut y stocker les objets d'un programme orienté objets. Permet le partage des données entre plusieurs programmes à la fois. Simplifie la programmation orientée objets pour assurer la persistance des données, par rapport à une BD relationnelle. Navigation parfois plus compliquée qu'une BD relationnelle. Moins standardisée que les BD relationnelles. Dépendante des langages de programmation.

BD relationnel objets : BD relationnel offrant en plus des mécanismes pour stocker des objets, ou pour stocker des structures plus complexes. Par exemple, une table peut "contenir" une autre table, ou stocker un objet d'un programme Java ou C++. Toutefois, les mécanismes d'accès aux objets sont moins souples que ceux des bases de données orientées objets.

XML : Extensible Markup Language : fichier texte ordinaire contenant des données structurées sous forme de balises que l'on peut imbriquer les unes dans les autres. On peut visualiser son contenu, à l'aide d'outils, sous forme d'un arbre. On peut l'utiliser pour échanger des données entre des applications, ou pour stocker de manière simple des données d'une application.

B. Niveau externe

Ce niveau correspond à la vision de tout ou partie du schéma conceptuel par un groupe d'utilisateurs concerné par une application. Il s'agit de décrire à l'aide d'un schéma externe ou Vue la façon dont seront perçues les données par un programme d'application.

Exemple :

Dans une base de données de gestion universitaire, un groupe d'utilisateurs concerné par les INSCRIPTIONS des étudiants n'a pas besoin d'avoir une vision globale de la base et peut se limiter à la partie qui englobe les informations relatives aux étudiants et aux Modules.

Un schéma externe peut être considéré comme un sous schéma du schéma conceptuel. Grâce à cette notion de schéma externe, chaque groupe d'utilisateurs perçoit les données à sa façon. Par exemple, une donnée vue comme donnée numérique par un groupe peut être vue comme une chaîne de caractères par un autre (cas d'une date par exemple). Un groupe peut ne pas voir certaines caractéristiques (attributs) d'une entité (ex : note obtenue dans un module) qui seront par contre visibles par un autre groupe (ou application).

C. Niveau interne

Ce niveau a pour but de spécifier comment les données sont stockées sur les supports physiques. Cette spécification est faite par le biais d'un schéma physique ou schéma de stockage. Ce schéma permettra par exemple de :

- Décrire la structure des fichiers qui constituent la base de données (nom d'un fichier, organisation, adresse sur le support, etc.)
- Définir les méthodes d'implantation (fichier plat, inversé, etc.)
- Préciser les chemins d'accès aux enregistrements (index, chaînage, calcul d'adresse, etc.)

2.4 Objectifs d'une base de données

Les objectifs que l'on assigne généralement aux bases de données et aux systèmes qui les supportent sont les suivants :

- la centralisation,
- l'indépendance entre les données et les traitements,
- la structuration de données complexes,
- le partage des données,
- l'intégrité et la cohérence,
- la confidentialité,
- la sécurité.

La centralisation des données a pour objet de limiter la redondance, c'est-à-dire d'éviter la présence de duplicata de données. L'information n'étant pas dupliquée, il s'ensuit une unicité de la saisie ainsi qu'une centralisation des contrôles qui peuvent être réalisés par le SGBD.

2.4.1 Partage de l'information

Une base de données permet le partage d'un ensemble unique d'informations par plusieurs utilisateurs. Cependant, il faut que cette mise en commun soit faite tout en préservant la vue particulière que chaque utilisateur peut avoir des informations, et en s'assurant que la simultanéité des traitements qui peuvent être effectués ne risque pas de dégrader l'intégrité de la base de données.

2.4.2 Organisation des données indépendamment des programmes

Afin de construire un ensemble d'informations structurées non redondant et qui soit partageable par plusieurs utilisateurs, il est nécessaire de faire abstraction des traitements particuliers de tel ou tel utilisateur (ou programme) pour tenter d'organiser les informations en fonction de leur nature et des liens réels qui existent entre elles. C'est de cette manière qu'on arrivera à garantir le maximum d'indépendance entre données et programmes.

2.5 Rôle d'une Base de données

Contrairement aux approches classiques, la création d'une base de données qui soit partagée par plusieurs utilisateurs est le reflet d'une évolution dans la gestion de l'entreprise. Son rôle est de rendre possible :

- La centralisation de l'information : l'information n'est plus éparpillée dans différents fichiers à différents endroits.
- L'intégration (tout ce qui se fait dans un service est visible par d'autres services)
- La diffusion de l'information archivée (si l'information est disponible à un seul endroit, elle est facile à diffuser)

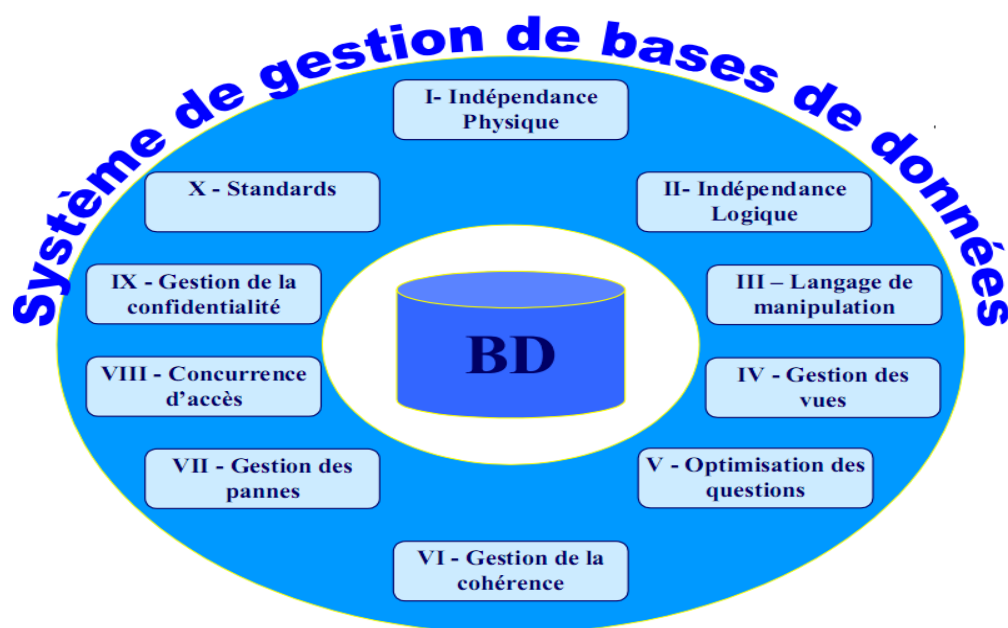
Ceci a pour avantages :

- d'améliorer la cohérence de l'information (une seule valeur pour une même information)
- de réduire les redondances (une même information n'est stockée si possible qu'une seule fois)
- de réduire les efforts de saisie et de mise à jour des informations (i.e. une information qui doit être stockée une seule fois ne sera saisie qu'une seule fois. De même que sa mise à jour ne se fera qu'une seule fois)

3. La gestion des bases de données

Afin de pouvoir contrôler les données ainsi que les utilisateurs, le besoin d'un système de gestion s'est vite fait ressentir. La gestion de la base de données se fait grâce à un système appelé **SGBD** (système de gestion de bases de données) ou en anglais DBMS (Database management system). Le SGBD est un ensemble de services (applications logicielles) permettant de gérer les bases de données, c'est-à-dire :

- permettre l'accès aux données de façon simple
- autoriser un accès aux informations à de multiples utilisateurs
- manipuler les données présentes dans la base de données (insertion, suppression, modification)



3.1 Définition

Un SGBD peut être vu comme un système informatique (un logiciel) spécialisé dans le traitement de gros volumes d'informations et permettant à différents utilisateurs d'interagir avec la base de données.

Un SGBD doit permettre de définir la structure de la base et d'y introduire les données correspondantes. De plus, une fois la base créée, il faudra d'une part la mettre à jour et d'autre part l'exploiter ou l'interroger.

3.2 Fonctions principales d'un SGBD

Un SGBD possède 3 fonctions principales :

- 1- Fonction Description
- 2- Fonction Manipulation
- 3- Fonction Utilisation

3.2.1 Description des données

La modélisation conduit à la définition des entités qui vont constituer les données de la base, de préciser leurs caractéristiques ainsi que les liaisons qui existent entre elles. Ceci se fait grâce à un langage de description de données ou LDD (Data Definition Language ou DDL en anglais) offert par le SGBD. Le but essentiel de ces langages est de fournir une indépendance totale des données vis à vis des supports où elles seront stockées (on peut comparer la DATA DIVISION d'un programme COBOL à la description de la base)

3.2.2 Manipulation des données

Lorsque la structure de la base est décrite, il faut pouvoir stocker les informations correspondantes.

Ceci implique nécessairement un certain nombre de mécanismes pour construire des enregistrements correctement structurés et qu'il faut ensuite écrire sur un support physique (mémoire secondaire). De plus, il faudra pouvoir accéder à de tels enregistrements pour tous les problèmes de mise à jour et d'interrogation.

Pour ce faire, le SGBD fournit un langage de manipulation de données ou LMD (Data Manipulation Language en anglais). Ce langage est en général construit autour de quelques primitives d'accès et de manipulation (Get, Retrieve, Update, etc.).

3.2.3 Utilisation des données

A ce niveau, on désire par exemple interroger la base de données, c'est à dire rechercher parmi l'ensemble des entités stockées celles qui répondent à des critères de choix très divers. C'est une fonction qui définit directement le lien entre l'utilisateur et les données au sein d'une application.

4. Les différents types d'utilisateurs d'un SGBD

Il s'agit donc de voir les différents rôles que doivent jouer un individu ou un groupe d'individus pour concevoir, créer, mettre en œuvre et exploiter une base de données.

- **Administrateur de BDs (DBA) : s'occupe de**
 - la sécurité des données;
 - la coordination et du contrôle de l'usage de la BD;
 - La gestion du matériel et du logiciel du SGBD.
- **Concepteur de BDs : s'occupe de définir la structure des données.**
- **Analystes et programmeurs : développent les systèmes de BD.**
- **Utilisateur : accède à la DB pour l'interroger et la modifier.**