

Support de
cours

BASE DE DONNEES ET SYSTEMES DE GESTION DE BASE DE DONNEES (SGBD)

CHAPITRE 0 A – GENERALITES

Présenté par Anani ATADEGNON

Sommaire

Historique	3
1. Introduction	3
2. Qu'est-ce qu'une base de données ?	4
2.1. Notion de base de données	4
2.1.a. Description Générale.....	4
2.1.b. Base de Données informatisée	4
2.1.c. Enjeux des Bases de Données	4
2.2. Qu'attendre d'une base de données ?	5
3. Quelques définitions.....	5
4. Différents niveaux d'abstractions	6

CHAPITRE 0A - GENERALITES

Historique

Depuis l'antiquité, toute société organisée a éprouvé, au cours de son évolution, la nécessité d'accumuler et d'organiser des informations.

L'efficacité de l'organisation d'un grand nombre de données s'est révélée indispensable dans les domaines les plus variés : dans les entreprises à la disposition des employés, dans les bibliothèques, la disposition de tout un chacun, dans les compagnies aériennes pour les réservations de places, etc.

A partir de 1945 avec l'invention des ordinateurs, la notion d'organisation des informations prendra beaucoup plus d'ampleur. Surtout à partir de 1972, la découverte des circuits intégrés a permis aux ordinateurs d'avoir beaucoup plus de mémoire, de vitesse d'exécution et de capacité de stockage ; ceci a permis à un bon nombre d'individus de prendre conscience du parti qu'ils pouvaient tirer de ces machines pour résoudre les problèmes d'archivage et de consultation.

Des outils informatiques adaptés à ces problèmes ont alors été créés et facilitent de manière très efficace la tâche des non professionnels qui souhaitent gérer des fichiers.

Les bases de données et les systèmes de gestion de bases de données constituent les deux instruments de cette organisation. Tous deux reposent sur l'usage de "modes d'emploi" appelés langages qui permettent de définir et de manipuler les différents types de données présentés.

1. Introduction

Ce cours a pour but d'expliquer les bases de données et les systèmes de gestion de bases de données de façon générale et en particulier l'utilisation de SQL (Structured Query Language) qui s'est affirmé convenir aussi bien aux non-spécialistes qu'aux programmeurs professionnels.

Au terme de cette étude, l'étudiant pourra, au fur à mesure qu'il assimile les constituantes de ce langage, pénétrer le fonctionnement d'un logiciel utilisant SQL, et ainsi atteindre progressivement ses possibilités d'exploitation d'une base de données de son choix.

Les bases de données ont pris aujourd'hui une place essentielle dans l'informatique, plus particulièrement en gestion. Au cours des trente dernières années, des concepts, méthodes et algorithmes ont été développés pour gérer les données sur mémoires secondaires ; ils constituent aujourd'hui l'essentiel de la discipline « Bases de Données ». 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. De plus, une théorie fondamentale sur les techniques de modélisation des données et les algorithmes de traitement a vu le jour. Les bases de données constituent donc une discipline s'appuyant sur une théorie solide et offrant de nombreux débouchés pratiques.

Dans la suite de ce cours, nous intéresserons essentiellement aux bases de données informatisées.

2. Qu'est-ce qu'une base de données ?

2.1. Notion de base de données

2.1.a. Description Générale

Il est difficile de donner une définition exacte de la notion de base de données. Une définition très générale pourrait être :

Définition 1 | Base 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 de données). Bien entendu, dans le cadre de ce cours, nous nous intéressons aux bases de données informatisées.

2.1.b. Base de Données informatisée

Définition 2 | 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.

Le résultat de la conception d'une base de données informatisée est une description des données. Par description on entend définir les propriétés d'ensembles d'objets modélisés dans la base de données et non pas d'objets particuliers. Les objets particuliers sont créés par des programmes d'applications ou des langages de manipulation lors des insertions et des mises à jour des données. Cette description des données est réalisée en utilisant un modèle de données. Ce dernier est un outil formel utilisé pour comprendre l'organisation logique des données.

2.1.c. Enjeux des Bases de Données

Comme cela a été précisé dans l'introduction de ce chapitre, les bases de données ont pris une place importante en informatique, et particulièrement dans le domaine de la gestion. L'étude des bases de données a conduit au développement de concepts, méthodes et algorithmes spécifiques, notamment pour gérer les données en mémoire secondaire (*i.e.* disques durs). En effet, dès l'origine de la discipline, les informaticiens ont observé que la taille de la RAM ne permettait pas de charger l'ensemble d'une base de données en mémoire. Cette hypothèse est toujours vérifiée, car le volume des données ne cesse de s'accroître sous la poussée des nouvelles technologies du WEB. Ainsi, les bases de données de demain devront être capables de gérer plusieurs dizaines de téraoctets de données, géographiquement distribuées à l'échelle d'Internet, par plusieurs dizaines de milliers d'utilisateurs dans un contexte d'exploitation changeant (on ne sait pas très bien maîtriser ou prédire les débits de communication entre sites) voire sur des nœuds volatiles. En physique des hautes énergies, on prédit qu'une seule expérience produira de l'ordre du pétaoctet (1 000 000 000 000 000 octets) de données par an. Comme il est peu probable de disposer d'une technologie de disque permettant de stocker sur un unique disque cette quantité d'informations, les bases de données se sont orientées vers des architectures distribuées ce qui permet, par exemple, d'exécuter potentiellement plusieurs instructions d'entrée/sortie en même temps sur des disques différents et donc de diviser le temps total d'exécution par un ordre de grandeur.

2.2. Qu'attendre d'une base de données ?

Contrôle de la redondance des données

Si plusieurs utilisateurs ou applications accèdent aux mêmes données, celles-ci doivent être enregistrées dans la machine à un seul endroit.

Cohérence des données

La transaction est enregistrée sur disque au moment où elle est terminée.

Partage des données

Les données stockées peuvent être partagées.

Sécurité

On doit pouvoir définir des droits d'accès ou des restrictions sur la base.

Intégrité des données

L'intégrité sera assurée s'il n'y a ni redondance, ni inconsistance. De plus, dans l'application, un ensemble de contraintes doit être défini pour que les données restent intègres : une date de naissance libellée comme suit 36/15/1998 est impensable.

Indépendance des données pour l'application

Différents services offerts par le système de base de données ne doivent avoir aucun impact sur l'application :

1. Définition d'index,
2. Définition de données distribuées ou partagées,
3. Représentations internes des données,
4. Système de backup,
5. Notion de vue,
6. Accès à la base.

3. Quelques définitions

Entité

Un ensemble d'informations liées entre elles.

Attribut

Les entités ont des propriétés appelées attributs.

Domaine

Chaque attribut a un ensemble de valeurs qui est appelé le domaine.

Clé

Une clé est définie à partir de colonnes. Une valeur de chaque colonne permettant d'identifier univoquement un enregistrement de l'entité.

Relation

Une relation permet de lier des entités entre elles.

Administrateur de base de données (DBA : Data Base Administrator)

Personne ou groupe de personnes responsable du contrôle de la base de données.

Système de gestion de base de données (DBMS : Data Base Management System)

Ensemble de softwares (applications) nécessaire aux contrôles de la base de données.

Dictionnaire de données (DD)

Un ensemble d'entités représentant les caractéristiques de la base de données appelé aussi méta-database.

Programme de l'application

Un programme qui utilise une base de données est appelé un programme de l'application.

Host langage

Un programme de l'application est écrit dans un langage de programmation. Ce langage est appelé Host langage (Java, C/C++, Cobol...)

Langage de définitions de données (DDL : Data Definition Language)

Langage permettant d'interfacer le dictionnaire pour créer la base de données (création de tables, d'index...).

Langage de manipulation de données (DML : Data Management Language)

Langage utilisé pour manipuler les données.

4. Différents niveaux d'abstractions

Dans cette partie de notre cours, nous expliquerons comment les objectifs de contrôle de redondance, de données partagées et spécialement de l'indépendance des données peuvent être réalisées dans un système de base de données.

Nous débutons la discussion en présentant les trois niveaux d'abstraction. Nous discuterons également des deux différents langages utilisés dans le système de base de données, celui pour la définition des données (DDL) et celui pour la manipulation des données (DML).

Le modèle externe

Le modèle externe est en fait une vue pour l'utilisateur (end user). Il peut y en avoir plusieurs : jusqu'à un par utilisateur.

Par utilisateur, il peut être défini :

1. le host langage,
2. une vue (sous ensemble de la base de données),
3. accès à la base de données (login).

Le modèle conceptuel

Le modèle conceptuel est fréquemment appelé le «data model» de la base de données. Il définit toutes les entités, associations, index, ... à travers le langage de définition de données (DDL). La description de la base de données est enregistrée dans le dictionnaire. Contient l'ensemble des processus et librairies.

Le modèle interne

La représentation physique des données est spécifiée dans le modèle interne. L'utilisateur n'a pas accès à cette partie du modèle. Le modèle interne permet de cacher la complexité de l'implémentation de la base de données.

Les langages de la base de données

Il existe trois types de langages :

1. langage de définition de la base (création de table,...)
2. langage de manipulation de données (insertion de données, ...)
3. host langage : langage de programmation où les instructions des deux langages précédents peuvent être insérées.

Niveau utilisateur : vue limitée

