

Cours

Les bases de données et le langage SQL

Rédigé par

David ROUMANET
Professeur BTS SIO



Changement

Date	Révision

Sommaire

A Introduction.....	1
A.1 Présentation.....	1
A.2 Fonctionnement bases relationnelles.....	2
A.2.1 Exemple : logiciel de messagerie instantanée.....	2
A.2.1.a Fichier texte.....	2
A.2.1.b Fichier XML.....	3
A.2.1.c Gestionnaire de base de données.....	3
A.3 Recherche et gestion d'index.....	4
A.3.1 Attentes des utilisateurs.....	4
A.3.2 Les index.....	5
A.4 Utilité de l'aspect relationnel.....	6
A.4.1 Les associations.....	6
A.4.2 Les contraintes.....	6
B SQL.....	7
B.1 Les commandes d'interrogation.....	8
C Titre N°2.....	9
D Annexes.....	10
D.1 Sources.....	10
D.2 Autres.....	10
E Ce qu'il faut retenir.....	12

Nomenclature :

- **Assimiler** : cours pur. Explication théorique et détaillée (globalement supérieur à 4 pages).
- **Décoder** : fiche de cours, généralement inférieure à 5 pages.
- **Découvrir** : Travaux dirigés. Faisable sans matériel.
- **Explorer** : Travaux pratiques. Nécessite du matériel ou des logiciels.
- **Mission** : Projet encadré ou partie d'un projet.
- **Voyager** : Projet en autonomie totale. Environnement ouvert : Vous êtes le capitaine !

A Introduction

Ce qui fait la force d'internet, c'est l'arrivée du web 2.0 autour de 2003-2004.

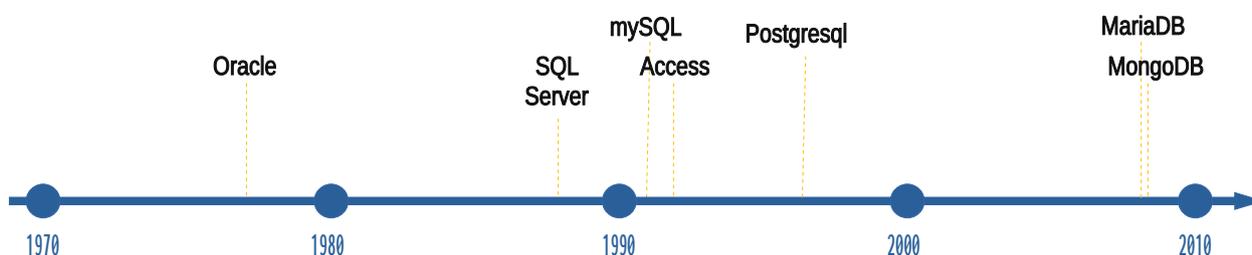
Ce nouvel Internet permet enfin aux utilisateurs d'avoir plus d'interactivité et ainsi permettre aux plateformes commerciales, de pouvoir vendre en ligne. Pour cela, une des briques essentielles, correspond aux moteurs de base de données.

A.1 Présentation

Les bases de données ne sont que des fichiers contenant des informations. Le format des informations y est normalisé, mais surtout, il y a un logiciel qui gère les écritures et les lectures multiples dans ces fichiers.

Ce logiciel est appelé moteur de base de données et il fait l'objet de féroces concurrence. On peut citer quelques moteurs (et bases) célèbres :

- [Oracle](#), par la société Oracle. Un système de base de données relationnelle, né en 1979.
- [SQL Server](#) voit le jour en 1989
- **Microsoft Access** est un système de base de données sur poste de travail, conçu autour de 1992.
- [MySQL](#) est inventé en 2001 par Michael WIDENIUS, et tire son nom de l'ajout du prénom de sa fille ('My') et du terme SQL.
- [MariaDB](#) est un fork (une fourchette) de MySQL, lorsque Oracle rachète MySQL en 2009, et c'est toujours Michael WIDENIUS qui donne le prénom de sa deuxième fille à son logiciel... Maria !
- [PostgreSQL](#) est le système open-source de base de données probablement le plus robuste et fiable, inventé en 1996. Il s'agit du remaniement du système **Ingres** (1985).
- [MongoDB](#) propose un système de base de données qui n'utilise pas le langage SQL et n'est pas relationnel, mais réparti, dans les années 2009
- Autour de 2010 on trouve également **Redis**, **Elasticsearch** et d'autres bases dites "no SQL".



L'utilisation des bases de données était utile, bien avant le web 2.0, mais la facilité de mise en œuvre de cette technologie en fait désormais, le composant principal des applications.

A.2 Fonctionnement bases relationnelles

Sans entrer dans le détail, tout ce qui est stocké sur un ordinateur est un fichier : une base de données est donc un fichier dont le format est déterminé pour permettre une recherche rapide.

Nous pourrions créer nous-même un tel fichier.

A.2.1 Exemple : logiciel de messagerie instantanée

A.2.1.a Fichier texte

La première idée est d'enregistrer les messages dans un fichier texte, au fur et à mesure que nous les recevons.

```
Manon ; Salut à tous ; 16:24:22
Herbert ; Salut à toi ; 16:24:39
Evy ; Coucou à toi, Manon ; 16:24:40
Emma ; salut, t'es nouvelle ? ; 16:25:04
Manon ; non, mais je ne faisais que lire ; 16:25:45
Bob ; Salut à tous ; 16:25:59
Bob ; moi, je suis nouveau ; 16:26:14
```

Pour effectuer une recherche sur un auteur, il faut alors une fonction qui lit chaque ligne, et prend le texte jusqu'au premier ; qu'il trouve.

Pour trouver les dates, il faut rechercher le troisième ; et prendre le texte jusqu'à la fin.

Si on fait évoluer le logiciel pour prendre en compte les messages privés, il faut ajouter un ; à la fin et ajouter le nom du destinataire. Il faut également changer la recherche de date pour qu'elle s'arrête au 4^e ; !

Ce format est celui des fichiers CSV (Comma Separated Value).

Cependant, le stockage d'images ou d'éléments multimédia serait compliqué, **il faudrait interdire la valeur ASCII correspondant au ; pour ne pas trouver ce symbole dans les octets d'une image** (problème d'encodage).

A.2.1.b Fichier XML

Une amélioration pourrait venir de l'utilisation de balises plutôt que de symboles, pour séparer les informations.

```
<auteur>Manon</auteur><msg>Salut à tous</msg><heure>16:24:22</heure>
<auteur>Herbert</auteur><msg>Salut à toi</msg><heure>16:24:39</heure>
<auteur>Evy</auteur><msg>Coucou à toi, Manon</msg><heure>16:24:40</heure>
<auteur>Emma</auteur><msg>salut, t'es nouvelle ?</msg><heure>16:25:04</heure>
<auteur>Manon</auteur><msg>non, mais je ne faisais que lire</msg><heure>16:25:45</heure>
<auteur>Bob</auteur><msg>Salut à tous</msg><heure>16:25:59</heure>
<auteur>Bob</auteur><msg>moi, je suis nouveau</msg><heure>16:26:14</heure>
```

Ce format présente l'avantage qu'il y a toujours une étiquette à trouver. On peut insérer de nouvelles étiquettes sans risque de tout décaler.

Cependant, lors du changement de version (et donc l'ajout de destinataire privé), une partie du fichier n'aurait pas de champ <destinataire></destinataire> : une recherche pourrait planter ou bien le développeur devrait prévoir l'absence d'étiquette (absence de champ).

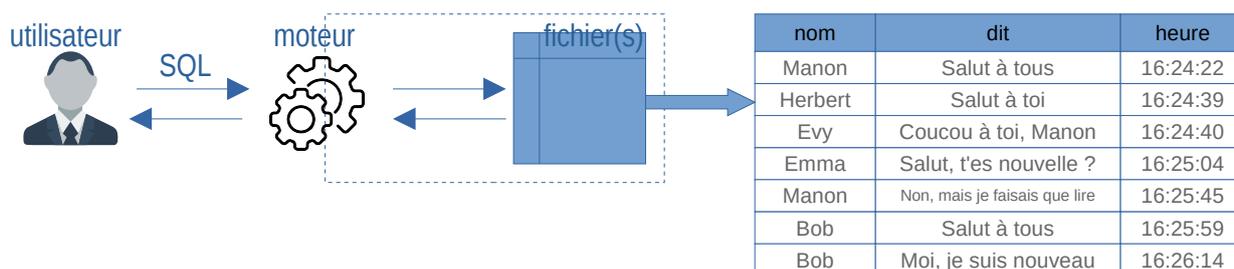
Un autre problème surviendrait après quelques années d'utilisation : **le fichier serait de plus en plus volumineux, une recherche prendrait de plus en plus de temps.**

A.2.1.c Gestionnaire de base de données

Cette fois, il s'agit de laisser gérer le fichier à un moteur à qui nous pourrions faire des requêtes, et qui se débrouillerait pour nous fournir une réponse rapidement.

La problématique devient alors : comment communiquer notre besoin à ce fameux moteur de base de données ?

Et la réponse est devenue : SQL !

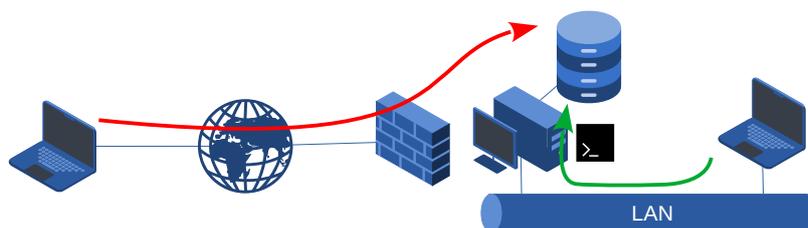


Ainsi, la question "afficher tous les noms d'expéditeurs ayant dit 'Salut à tous'" devient en SQL :

```
SELECT nom
FROM tableau
WHERE dit = 'Salut à tous';
```

Le moteur de base de données peut fonctionner de deux manières :

- Dans une console : l'utilisateur lance alors une commande du système pour ouvrir l'éditeur SQL et peut ensuite interroger la base. On peut accéder à la console du système (par SSH par exemple) et ainsi afficher les commandes SQL sur son propre ordinateur. Ce n'est cependant pas le fonctionnement le plus répandu.
- Comme un service réseau : il est alors possible de se connecter sur le port réseau de la machine sur laquelle est installée la base de données et son moteur et de transmettre des requêtes SQL. Ce mode est très utilisé avec les langages de programmation, mais présente une vulnérabilité aux hackers. Ainsi, c'est la norme de ne pas accepter de connexions depuis des machines inconnues.



Voici quelques numéros de ports TCP par défaut de quelques bases de données connues :

Editeur	Port
MySQL	3306
MariaDB	3306 ou 3307
PostgreSQL	5432
Microsoft SQL server	1433
Oracle	1521

Généralement, MariaDB est utilisée en remplacement de MySQL, elle utilise donc le même port.

A.3 Recherche et gestion d'index

A.3.1 Attentes des utilisateurs

Sans rentrer dans les détails, il est utile de comprendre qu'une base de données est un ensemble de fichiers et d'algorithmes. Toutes les bases de données visent des objectifs similaires :

- Fiabilité des données et des requêtes : en laissant le moteur de base de données gérer les données, l'utilisateur attend que ses données ne contiennent pas d'erreur lors de manipulations.
- Disponibilité des informations : le langage d'interrogation doit permettre de restituer les informations demandées, sous toutes ses formes (partielles, complètes, triées, filtrées, etc.)
- Rapidité d'exécution : les performances de la base de données doivent rester acceptables, quelle que soit sa taille. Pour cela, le moteur de base de données doit utiliser des techniques permettant d'optimiser ses recherches.

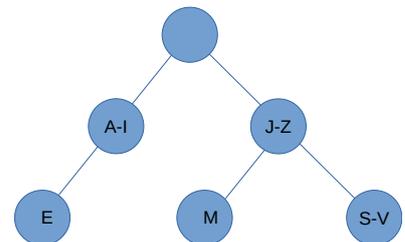
A.3.2 Les index

Selon les concepts découverts par CODD, le format des tables de données doit répondre à des critères particuliers et particulièrement le fait que les informations ne doivent pas être redondantes, mais unique.

Pour cela, le concepteur crée des clés (un champ dont le contenu est unique) : la force de ce système est qu'il devient alors possible d'indexer les données.

L'indexation des éléments d'une base de données est similaire à l'index d'un livre : chaque notion est associée à un numéro de page, permettant au lecteur d'atteindre rapidement l'information recherchée, sans devoir lire chaque page du livre. Il en va de même avec la base de données.

utilisateur			
prenom	nom	heure	ID
Manon	DESOURCES	16/12/01	1
Herbert	FRANK	05/10/03	2
Evy	TEUMOI	18/08/93	3
Emma	LADH	20/11/16	4
Bob	MORANE	16/10/53	6



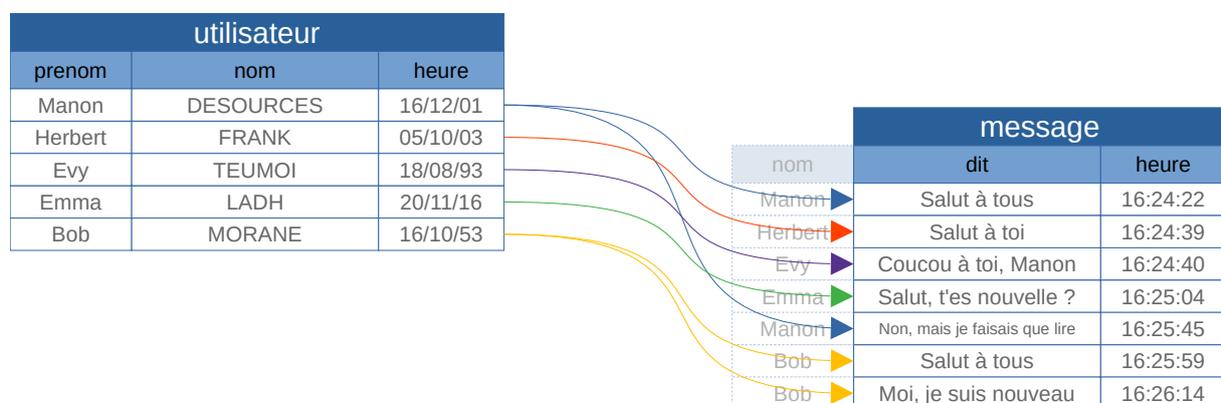
Le moteur se crée un arbre de recherche en fonction de la quantité de données ou du type de données. Ce procédé sort du cadre de ce cours, mais il est abordé dans les notions de mathématiques en BTS SIO.

A.4 Utilité de l'aspect relationnel

Telle que nous l'avons présenté précédemment, une base de données ne contient que des tables de données. Cette vision serait réductrice : le moteur de base de données gère de nombreuses autres fonctionnalités, dont les contraintes entre les tables.

A.4.1 Les associations

Par exemple, nous pourrions avoir deux tables : l'une contenant les informations sur le profil de l'utilisateur (prénom, nom, date de naissance) et une table contenant les messages (message, destinataire et date). Ce qui permet à notre messagerie instantanée de fonctionner, c'est l'association entre un utilisateur et ses messages.



Une base de données relationnelle contient donc les données nécessaires dans les tables, mais également des informations sur les relations.

A.4.2 Les contraintes

Le moteur de la base s'assure en permanence (à chaque insertion ou suppression de données) que les contraintes sont respectées.

Toujours dans notre exemple, un message ne peut pas être émis par une personne qui n'existe pas dans la table utilisateur. De même en cas de suppression d'un utilisateur, il faudra d'abord supprimer ses messages, pour qu'ils ne soient pas orphelins (sans auteur).

Il y a d'autres fonctionnalités intéressantes (gestion des accès, déroulement des requêtes, automatisation d'actions) qui feront l'objet d'un autre cours.



Ce qu'il est important de se rappeler et de comprendre, c'est qu'une base de données relationnelle est capable de stocker d'énormes quantités de données et d'avoir de très bonnes performances, grâce aux index.

B SQL

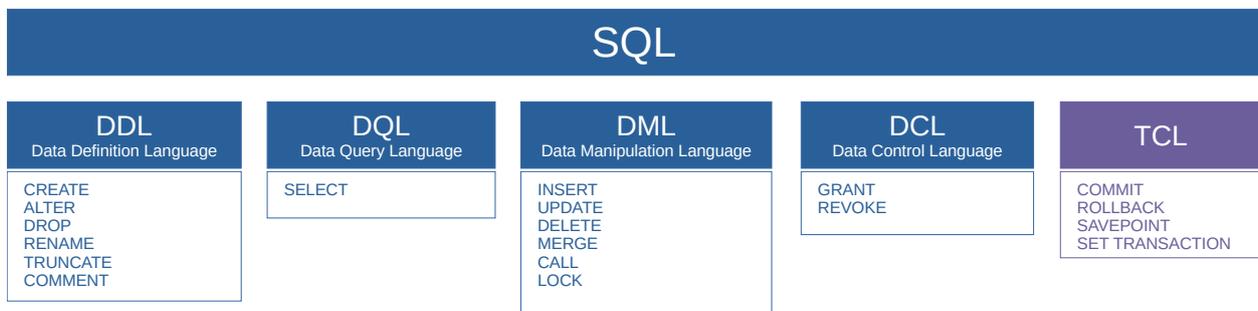
SQL est l'acronyme de [Structured Query Language](#) (langage de requêtes structurées).

La première version de SQL date de 1974, mais a été normalisée en 1986 et depuis cette date, de nombreux éditeurs l'ont adoptée. Ce langage est lié à un ingénieur – Edgar Frank Codd – qui a travaillé sur le modèle relationnel. Son idée est de traiter les relations entre les données de façon logique et mathématique.

SQL comporte quatre domaines d'actions :

- **DDL** (Data Definition Language). La définition des bases de données, les opérations permettant de définir les structures, les tables, de les modifier et de les supprimer.
- **DML** (Data Manipulation Language). La définition des données, l'ajout, la suppression, la lecture et la suppression de celles-ci.
- **DQL** (Data Query Language). La création de requêtes sur les structures, pour récupérer des données dans un format pratique.
- **DCL** (Data Control Language). La gestion des contrôles sur les accès et droits aux données et aux structures.

Il existe un cinquième domaine d'action qui concerne le déroulement des transactions (TCL pour Transaction Control Language) qui serait un peu comparable à TCP pour le protocole IP.



L'objet de ce cours est de permettre l'usage de l'ensemble de ces commandes.

B.1 Les commandes d'interrogation

C Titre N°2

D Annexes

D.1 Sources

Les indexes dans les SGBDR : <https://www.guru99.com/fr/indexing-in-database.html>

D.2 Autres

À EFFACER LORSQUE LE DOCUMENT EST TERMINÉ

Un encart de base

Attention, il est important de vérifier blabla que blabla. Cela peut arriver lorsque blabla ou bien que truc active le machin et dépasse la limite du bidule. Il est donc important de ne pas laisser les choses dériver sans un contrôle des valeurs sinon blabla.

Il reste cependant la possibilité d'écrire en dessous du nombre de lignes.

Un encart d'attention + puce attention



Attention, il est important de vérifier blabla que blabla. Cela peut arriver lorsque blabla ou bien que truc active le machin et dépasse la limite du bidule. Il est donc important de ne pas laisser les choses dériver sans un contrôle des valeurs sinon blabla. Il reste cependant la possibilité d'écrire en dessous du nombre de lignes.

Un encart stop + puce stop



Attention, il est important de vérifier blabla que blabla. Cela peut arriver lorsque blabla ou bien que truc active le machin et dépasse la limite du bidule. Il est donc important de ne pas laisser les choses dériver sans un contrôle des valeurs sinon blabla. Il reste cependant la possibilité d'écrire en dessous du nombre de lignes.

Un encart de réflexion



Attention, il est important de vérifier blabla que blabla. Cela peut arriver lorsque blabla ou bien que truc active le machin et dépasse la limite du bidule. Il est donc important de ne pas laisser les choses dériver sans un contrôle des valeurs sinon blabla. Il reste cependant la possibilité d'écrire en dessous du nombre de lignes.

Un encart de note



Attention, il est important de vérifier blabla que blabla. Cela peut arriver lorsque blabla ou bien que truc active le machin et dépasse la limite du bidule. Il est donc important de ne pas laisser les choses dériver sans un contrôle des valeurs sinon blabla.

Il reste cependant la possibilité d'écrire en dessous du nombre de lignes.

E Ce qu'il faut retenir...

Insertion > Autres sauts > Saut Manuel...

Choisir le style de page « Résumé »

Attention : résumé s'enchaîne avec Style de page droite ► Mettre une page « Résumé » à la place d'une page Gauche.

Les [liens](#) sont sous et surlignés