

## 5. Bases de données

Bonjour à tous. Bienvenue dans cette cinquième et dernière séquence du codage binaire. Nous avons vu comment représenter les informations sous forme de bits, les images, les sons, comment étaient organisées les données en mémoire. Maintenant, nous allons parler des bases de données.

Alors les bases de données, à quoi ça sert ? En pratique dans la vie courante, on a souvent à gérer des stocks, un ensemble de clients, des ventes, un inventaire de bibliothèque, un site web, plein de choses qui contiennent énormément d'informations qu'on va vouloir manipuler, mais sans forcément manipuler les informations telles qu'elles sont stockées dans l'ordinateur parce ce qu'on est des humains, on veut manipuler les informations de manière simple. On ne veut pas avoir à s'embêter à savoir que les données ont été stockées dans l'ordinateur de telle façon, codées avec un certain nombre de bits, etc. On veut avoir une façon simple de manipuler toutes ces données. Les bases de données vont servir à ça. On va avoir des logiciels qui s'appellent des systèmes de gestion de base de données qui sont des médiateurs entre l'humain et les données telles qu'elles sont stockées dans l'ordinateur. Par exemple, MySQL est un tel système.

Les bases de données sont souvent manipulées par le modèle qu'on appelle le modèle relationnel. Dans le modèle relationnel, on a un principe important qui s'appelle les relations qui sont des tableaux contenant les données qu'on va vouloir manipuler. Ici dans mon exemple, j'ai 2 tableaux. Le premier qui s'appelle "Films" qui va contenir 3 colonnes : une avec le titre des films, une avec le réalisateur et une avec l'acteur principal. Et ensuite, un deuxième tableau qui contient les séances avec le titre à nouveau des films, le cinéma où le film est joué et l'heure de cette séance. En pratique, on aurait pu utiliser un seul tableau avec toutes ces informations en même temps, simplement on aurait eu beaucoup de duplications d'informations puisqu'on aurait répété par exemple le réalisateur et l'acteur pour chacune des séances du même film. On aurait eu beaucoup d'informations stockées plusieurs fois, ce qui n'est pas très utile en pratique et il n'y a pas forcément de raison indispensable de stocker ces informations de manière mélangée. Donc on va les stocker dans 2 tableaux différents avec chacun leur rôle.

Avec le modèle relationnel ensuite, on va pouvoir faire des requêtes. Une requête, c'est par exemple "Où est-ce que je peux voir un film avec tel ou tel acteur ?" Et pour ça, on va avoir un langage qui a été développé pour que les humains puissent relativement facilement faire des requêtes dans la base de données pour trouver ce genre d'information. Ici, on va écrire une requête dans le langage SQL qui va être "Trouve-moi les cinémas - c'est la ligne select Cinéma - en regardant dans les tables "Films" et "Séances" - que j'ai définis précédemment - et parmi ces tables, tu vas trouver les lignes pour lesquelles le titre du film est le même titre que la séance et l'acteur est l'acteur que je cherche". Ça, c'est une requête qui s'écrit dans un langage relativement compréhensible pour un humain et qui permet de chercher dans la base de données très efficacement dans mes 2 tableaux la réponse à ma question. Ce qui est bien, c'est que là on se fiche de savoir comment les données sont stockées concrètement dans l'ordinateur, on va simplement écrire dans un langage assez simple et avoir la réponse de manière très efficace. En pratique, ce langage va être traduit en langage machine et ensuite être exécuté rapidement par l'ordinateur qui lui sait où se trouvent les données concrètement.

L'intérêt des bases de données, ça va être de répondre à des questions comme ça de manière très facile sans avoir à être programmeur.

Les bases de données vont aussi avoir d'autres intérêts, par exemple protéger les données parce que les données de nos jours ça a une très grande valeur. Par exemple, les comptes d'une entreprise, nos photos de vacances ou plein d'autres choses, c'est des choses qui vont être stockées dans des systèmes informatiques et qu'il ne faut absolument pas perdre sinon ça pose énormément de problèmes. Et évidemment, les systèmes informatiques ne sont pas infailibles. On a des pannes matérielles, des bugs logiciels, des attaques de pirates éventuellement qui font qu'on va parfois perdre ces données. Les bases de données vont mettre en œuvre des systèmes qui vont permettre d'éviter ce genre de problème par exemple en assurant de la réplication entre les données, c'est-à-dire qu'on va stocker les données à plusieurs endroits et si jamais il y a un des serveurs qui tombe en panne, on pourra toujours accéder de la même façon aux données stockées ailleurs. Pour l'utilisateur, ça ne va rien changer sauf que ses données ne seront jamais perdues.

Les bases de données vont aussi servir à partager les données. On a souvent des ensembles de données manipulées par plusieurs utilisateurs. Par exemple, on peut imaginer d'avoir un système contenant plein de morceaux de musique avec 2 utilisateurs, Alice et Bob, qui y accèdent et vont pouvoir éventuellement mettre des likes, comme sur votre réseau social préféré, pour certains morceaux qu'ils apprécient. On peut imaginer Alice qui ouvre un morceau, qui le lit. Au même moment, on va avoir Bob qui va lire le même morceau. Et comme Alice se rend compte qu'elle aime bien le morceau, elle va ajouter un like sur ce morceau. Et un peu plus tard, Bob va faire la même chose, il va ajouter un like. Et là, on va se rendre compte que le like d'Alice va avoir disparu et que celui de Bob est resté. Ce qui s'est passé, en fait c'est que quand Bob a ouvert son morceau, il n'y avait pas encore le like d'Alice dessus, quand il a ajouté son like, il n'y avait que le sien et quand il l'a refermé et qu'il l'a enregistré il y avait uniquement son like qui y était est celui d'Alice a disparu parce qu'il a été écrasé.

Les bases de données, pour éviter ce genre de problème, vont définir ce qu'on appelle des transactions qui sont en fait des opérations qui vont garantir que ce genre de problème ne va pas arriver. Ça va permettre de faire des accès concurrents entre différents utilisateurs, des modifications de la base de données sans qu'il y ait de conflit si jamais ils accèdent à la même chose en même temps. En gros, il y en a un qui va faire sa modification et ensuite l'autre va faire la sienne de façon à ce que les likes ne soient pas écrasés.

Ce qu'il faut retenir de cette séquence, c'est que tout d'abord l'ordinateur stocke les données de manière pas forcément compréhensible pour un humain, avec plein de bits utilisés pour chacune des données, éventuellement de la compression, un ordre particulier, une organisation de mémoire particulière. Tout ça, un humain ne veut pas avoir à s'en servir quand il va accéder à ces données. Lui veut directement manipuler ces données de manière simple. Les systèmes de gestion de bases de données vont servir à ça, elles vont fournir un moyen pratique d'y accéder, vraiment compréhensible pour les humains et aussi un moyen fiable puisqu'ils vont éviter d'avoir des pertes de données ou éventuellement d'avoir des problèmes d'accès concurrents si jamais plusieurs utilisateurs partagent les mêmes données en même temps et les modifient en même temps.