

Base de données

Aujourd'hui il paraît trivial à l'internaute, tapant un mot-clé sur un moteur de recherche, d'obtenir instantanément une kyrielle de réponses d'ailleurs plus ou moins pertinentes avec la question posée.

Par exemple, l'[interrogation](#) "base de données" sur Google donne 20 millions de références en moins d'1/3 de seconde.

Evidemment des recherches aussi peu précises ne renseignent pas vraiment sur le sujet et il faut affiner la question pour obtenir des résultats plus significatifs.

Ainsi si on cherche plus précisément un titre comme "*introduction pratique aux bases de données relationnelles*" on obtient simplement 40 000 réponses.

Il a fallu de nombreuses années de recherche en informatique et en mathématique théorique et des milliards de dollars d'investissement dans de gigantesques [fermes de serveurs](#) pour que les informations innombrables contenues sur la toile puissent être accessibles avec une telle efficacité.

Les puissants algorithmes qui permettent cet exploit sont d'ailleurs précautionneusement et jalousement protégés par leurs propriétaires.

Mais serez-vous surpris, ayant fait ce type de recherche, d'être ensuite envahi de publicités portant sur le sujet que vous avez exploré ?

C'est l'effet [big brother](#) qui fait que ces grands dévoreurs d'informations que sont Google, Facebook, Apple et quelques autres entretiennent d'immenses bases de données sur nous, nos goûts, nos comportements, nos déplacements et bien d'autres choses encore à venir avec l'émergence de l'[internet des objets](#).

Les gigantesques silos à données, accumulant des millions voire des milliards de téraoctets ne servent pas seulement à indexer et conserver les références de milliards de documents textes, voix, images, données formatées que ces dévoreuses d'information récoltent en permanence mais ils gardent, dans de gigantesques bases de données, les informations personnelles que nous leur avons confiées volontairement ou involontairement.

On est vraiment très loin des premiers ordinateurs qui ne faisaient que généraliser les capacités des tabulatrices c'est-à-dire des totalisations, des multiplications, des divisions sur les champs numériques de fichiers cartes, par définition, séquentiels.

Les enregistrements étaient à l'origine de longueur fixe, de la taille des cartes de 80 colonnes soit 80 caractères alphanumériques ce qui obligeait à se limiter au strict nécessaire dans les fichiers.

On se souvient d'ailleurs de la grande peur du bug de l'an 2000 lié au fait que, pour économiser de la place dans les fichiers, les dates avaient été codées sur seulement 2 caractères et que forcément 2000 codé 00 se trouvait ainsi devenir une date antérieure à 1999 codé 99.

L'apparition du [disque](#) en 1956 avec l'annonce par IBM du RAMAC 305 permit de se libérer de la contrainte de lecture sé-

quentielle qui imposait de faire défiler 99 cartes avant de lire la 100^{ème} lorsqu'on voulait accéder seulement à celle-ci.

Elle libéra plus tard, avec l'apparition de fichier [séquentiel indexé](#), de l'accès par la position ramenant la recherche à une lecture séquentielle d'un index pour connaître la position d'un enregistrement recherché sur le critère d'une clé unique comme par exemple un matricule.

Vinrent ensuite les premières bases de données [hiérarchiques](#) et [réseaux](#) qui visaient à éliminer la redondance de données et aussi à assurer une meilleure cohérence en évitant les duplications de champs, source d'erreurs à la saisie et encore plus à la mise à jour.

Elles permettaient aussi l'utilisation simultanée par plusieurs programmes puis avec l'avènement du télétraitement par plusieurs utilisateurs.

Mais déjà à la fin des années 60, [Edgard F. Codd](#), dont la formation de mathématicien incitait à définir un cadre rigoureux aux concepts des bases de données, s'appliquait à trouver un formalisme unique pour définir la manière dont les données pouvaient être stockées, mises à jour et extraites des bases de données. Il fallait aussi obtenir des réponses cohérentes à des requêtes tout en permettant la mise à jour des informations.

Ainsi naquit le concept de base de données relationnelles qui domine encore aujourd'hui le marché toujours aussi dynamique des fournisseurs de gestionnaires de Bases de Données dont les plus représentatifs sont, par ordre alphabétique, IBM avec [DB2](#), Microsoft avec [SQL Server](#), [Oracle](#) et [Sybase](#) pour ne citer que les principaux.

Le langage SQL, désormais universellement connu, fut formalisé pour la première fois par Donald Chamberlain et Raymond Boyce dans un [article](#) paru en 1974 dans ACM : "*SEQUEL: a structured english query language*".

Le nom de SEQUEL ne put être retenu parce que c'était déjà une marque déposée par [Hawker Siddeley Aircraft Company](#) et le standard s'imposa sous le nom de SQL.

Raymond Boyce, prématurément disparu en 1974, laissa son nom à la postérité relationnelle pour son travail avec Ted Codd sur la normalisation des tables relationnelles avec la [BCNF](#) ou [Boyce and Codd Normal Form](#).

Sans ces précurseurs, les incroyables progrès de la technologie et les avancées de la science informatique en matière de recherche et de stockage, les gigantesques bases de données qui font la richesse des leaders de l'internet n'auraient pas vu le jour.

Nous sommes quelques uns à avoir partagé l'enthousiasme de ces chercheurs alors que la technologie balbutiante ne nous donnait pas vraiment encore l'occasion d'exploiter les immenses possibilités que leurs travaux laissaient entrevoir.

Nous étions loin des pétaflops, des pétaoctets et des 1/3 de seconde de temps de réponse d'un google d'aujourd'hui même si nous rêvions déjà d'un mythique *sub second response time*, ce [fameux](#) délai qui évite à la pensée de décrocher et qui joue un si grand rôle dans le succès des appareils modernes comme le smartphone.

