

Recherche, industrie, informatique : Grenoble a-t-elle eu en main les clés du BIG DATA ? (7/12 - 2021)

Publié par [ACONIT \(Association pour un Conservatoire de l'Informatique et de la Télématique\)](#), le 28 juin 2021 750



une synthèse par Xavier Hiron

à partir d'un article détaillé de Christian Jullien (disponible sur le site de l'ACONIT)

Depuis les années 2000 et l'avènement des technologies du numérique, la plupart d'entre nous sommes familiers d'échanges au quotidien d'importantes quantités de données. Selon nos usages qui évoluent avec le temps, ordinateurs, tablettes, smartphones échangent ces données via les réseaux de communication, le plus souvent en lien avec des serveurs distants. Les systèmes ajustent leurs performances grâce aux progrès d'une l'électronique aux débits de plus en plus conséquents.

Dans le même temps, nous prenons conscience de la « criticité » de l'information, surtout lorsque celle-ci affecte notre vie personnelle ou nos organisations. Le terme **Big Data** est alors employé pour exprimer ces **nouveaux ordres de grandeur concernant la capture, la recherche, le partage, le stockage, l'analyse et la présentation de données**. L'exploitation de cette masse considérable représente aujourd'hui un enjeu majeur pour les géants du WEB. En effet, nous sommes passés en quelques dizaines d'années de la simple gestion de lots de fiches à celle de fantastiques quantités de textes, images et sons. Dans ce domaine, la recherche logicielle grenobloise a joué un rôle crucial.

Fin des années 1960 : naissance du concept de bases de données

Dès les premiers calculateurs programmés en langage machine, dans les années 1950, des langages évolués se développent : **FORTAN**, **COBOL**, [ALGOL](#), **PL1**. Les données étant alors organisées en fichiers, leur dépendance avec les programmes est forte. Chaque modification de la structure

d'un fichier nécessite de recompiler les programmes. Puis de nouveaux enjeux apparaissent et, pour y répondre, un marché se crée pour **gérer des systèmes évolués de bases de données (SGBD)**. Ceux-ci permettent la description et l'organisation de structures de données, la manipulation de ces données de manière normalisée et l'amélioration du partage, de la concurrence d'accès, de la protection et de la sauvegarde.

Les premiers SGBD proposent un **modèle hiérarchique** de l'organisation des données, mais manquent de souplesse et sont sources d'erreurs, l'accès aux données nécessitant de coûteux parcours d'arbres. D'où l'évolution vers un **modèle réseau** organisé en collections d'objets (SETs), permettant de gérer des dépendances plus complexes. L'exécution des programmes devient alors soumise à tous les liens définis par la structure des données.

Ce modèle fournit la première approche commerciale des SGBD. Dès 1962, **Charles Bachman** et General Electric proposent le logiciel *Integrated Data Store* (IDS), suivi par IBM, avec *Information Management System* (IMS). En 1967, le consortium **CODASYL** publie un standard normalisant les interfaces des SGBD en deux volets : le DML (*Data Manipulation Language*) pour la manipulation des données, et le DDL (*Data Definition Language*) pour leur définition.

SOCRATE, un projet de recherche grenoblois

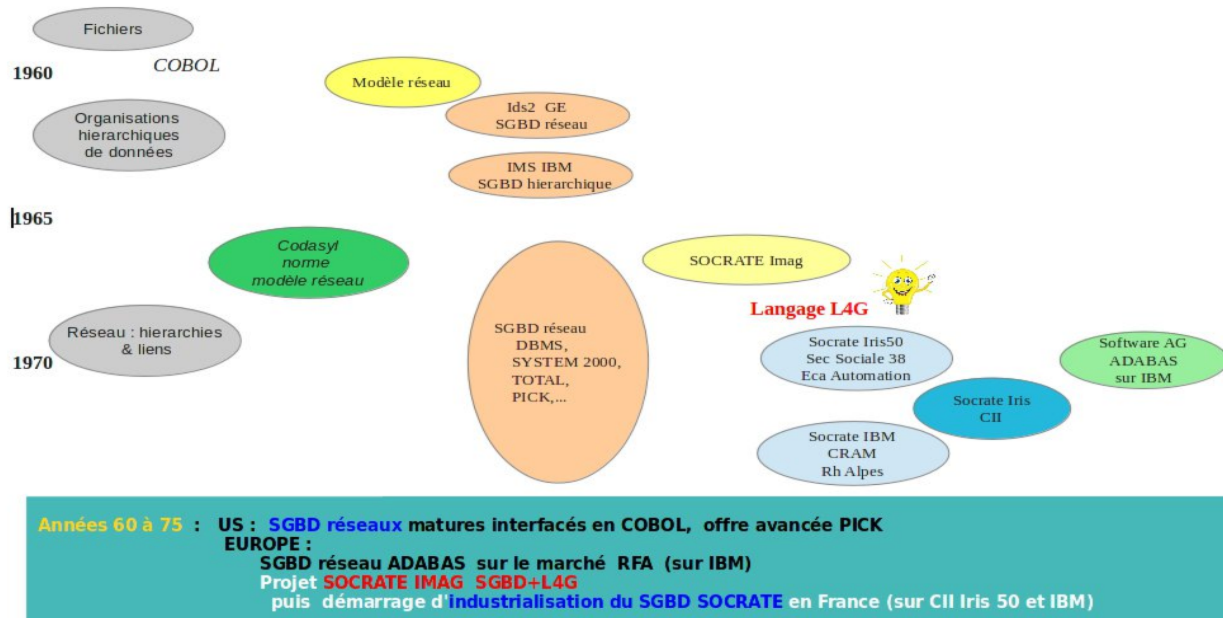
Au milieu des années 1960, quelques chercheurs grenoblois perçoivent l'importance de répondre aux nouveaux besoins de l'informatique de gestion (paie, comptabilité). Des contacts existent avec la recherche américaine, mais il fallut beaucoup de persuasion aux pionniers tels que **Jean Kuntzmann, Louis Bolliet ou François Peccoud** pour faire admettre cette informatique dans le milieu scientifique.

Grâce au plan calcul et à l'implantation d'un centre scientifique IBM à Grenoble, l'orientation s'affirme. Parmi les premiers travaux de l'IMAG figure dès 1969 un projet financé par le nouvel INRIA. L'équipe du professeur **Jean-Raymond Abrial** travaille à la réalisation d'un prototype de SGBD avancé, du nom de **SOCRATE**. Pour programmer des accès et modifier des données sans se soucier de leur implantation, ce prototype intègre un puissant **langage de manipulation de quatrième génération (L4G)** capable de se substituer à COBOL. Des administrations locales (CHU-Grenoble, rectorat, sécurité sociale), en phase de démarrage de leur informatique, soutiennent cette innovation.

Dans SOCRATE, les langages de définition, manipulation et d'écriture sont enregistrés grâce à une fonction **bibliothécaire** et un **outil macrogénérateur**. Un algorithme innovant gère l'implantation physique des objets en organisant des espaces virtuels quasi infinis sur les espaces disques de l'époque. En réduisant au maximum les accès en mémoire cachée, SOCRATE devient un outil simple et performant. Entre 1970 et 1973, prototype et développements sont réalisés sur un ordinateur **IBM 360/67**, en langage PL1. Ce prototype sera ensuite utilisé au CHU de Grenoble, puis au rectorat. L'année suivante, sa production sur **IRIS 50** puis **IRIS 80** de la **CII** est assurée pour le compte de la sécurité sociale. Cette version industrielle est réalisée en trois ans par une équipe d'ingénieurs de la société ECA Automation. Puis ce sera au tour de la Caisse régionale d'assurance maladie (CRAM) d'en commander une version pour **IBM 370**.

C'est le début d'une longue aventure industrielle. Sur le plan technique, SOCRATE étant utilisable en

langages de programmation usuels, les bases de données sont transférables d'un matériel à l'autre. Les modes de fonctionnement sont multiples : conversationnel, transactionnel, temps-réel... Au début des années 1970, les challengers européens SOCRATE et ADABAS (Allemagne) font face aux SGBD CODASYL, qui acquièrent leur pleine maturité et abondent une forte production américaine.



Évolution technologique : modèle relationnel, systèmes distribués, modélisation...

Cependant, dès la fin des années 1960, les recherches américaines convergent pour améliorer le modèle CODASYL. Chez IBM, les travaux algébriques d'**Edgar Frank Codd** débouchent sur une solution cohérente et rigoureuse d'arrangement de données. Ce modèle relationnel, toujours utilisé, fit l'objet d'une importante publication scientifique (Communications of the ACM, juin 1970 : *A Relational Model of Data for Large Shared Data Banks*). A sa suite, plusieurs développements de SGBD-R vont être mis en œuvre :

- SYSTEM-R au centre de recherches IBM de San José (**langage SEQUEL**), en 1970 ;
- RDBMS, INGRES à l'université de Berkeley (**langage QUEL**), en 1973 ;
- MRDS sur Honeywell, le premier SGBD-R à être commercialisé, en 1976.

Dans le cadres d'échanges entre le centre de recherche IBM de Grenoble et celui de San José, **Michel Adiba** contribue à résoudre des problèmes de **partage d'accès aux données** sur systèmes distribués (technique dite de « clichés » ou « snapshots »). Résoudre les problèmes de distribution des données s'avère être aussi un aspect primordial pour les systèmes d'informations (**Claude Delobel** en définit les **dépendances fonctionnelles**). Toujours en lien avec l'INRIA, le réseau **CYCLADES**, le nouveau centre de recherche CII et l'IMAG réalisent à partir de 1972 des avancées internationalement reconnues dans le domaine de la modélisation des SGBD (projets **SIRIUS**, **POLYPHEME** et **URANUS**).

ECA Automation et l'industrialisation de SOCRATE

ECA Automation ouvre alors une agence à Grenoble, en partenariat avec la firme **TELEMECANIQUE**, à l'initiative **Robert Morin**, ancien thésard sur SOCRATE. Ensemble, ils en produisent une version pour le mini-ordinateur **SOLAR**. Ce matériel, utilisant des mots de 16bits, cible principalement le marché des applications temps-réel ou embarquées de l'industrie, lesquelles manipulent des données de plus en plus complexes et volumineuses. Ce développement SOLAR s'accompagne de la création d'une carte électronique de traitement de l'adressage, dénommée VSS 65. Par ailleurs, SOCRATE est porté sur matériel **SIEMENS**, compatible IBM. Avec le succès des mini-ordinateurs, ce système devient un produit apprécié des professionnels et s'exporte outre-Atlantique.



Les années 1980 : des SGBD-R américains aux évolutions de CLIO

Dans la foulée de SYSTEM-R, une importante percée commerciale de produits est observée, du système **ORACLE** à **MySQL**, la version libre des développeurs WEB, en 1995. La **normalisation du langage SQL** est éditée par l'ANSI en 1986, puis adoptée par l'ISO en 1987 (ISO/CEI 9075:1986). La programmation d'applications s'en trouve pérennisée.

En France, dans les années 1980-90, SOCRATE doit changer de nom, la marque étant déjà déposée. Le nouveau produit, plus industriel, s'appellera **CLIO**, avec un accompagnement commercial au sein de **SYSECA (ex ECA-Automation) du groupe THOMSON-CSF**. Grenoble fournit le centre de développement, les services parisiens animant l'administratif et le marketing, pour un effectif global d'environ 80 personnes. Un club utilisateurs contribue aux évolutions du produit. Suivant diverses configurations, plus de 1000 sites en France seront équipés de CLIO, intégrant des références prestigieuses comme EDF, SNCF, Défense Nationale, Aérospatiale, COGEMA, banques et assurances. CLIO aura même un client à Los Angeles. Mais bien que leader dans son pays, la concurrence des **SGBD-R** se fait sentir. Même avec une normalisation de CLIO, comment faire face aux géants US tel qu'ORACLE, d'autant que THOMSON n'a pas la culture de l'édition logicielle ?

Les dirigeants de CLIO vont alors miser sur un de ses points forts : le L4G. Ce langage permet de développer des applications à des coûts très compétitifs. Le projet de production d'un **langage de 5ème génération, MUST**, est lancé. La branche des services informatiques (BSI) du groupe Thomson renforce le projet en rachetant l'activité américaine **NOMAD** au groupe Dun&Bradstreet. Une base commerciale est acquise aux USA. Un effectif d'environ 40 personnes travaille sur ce projet à Grenoble entre 1986 et 1990... Pendant ce temps, autre stratégie : en 1984, des amis fondent le futur **HARDIS Group** et éditent un produit de type L4G pour la gamme AS400 d'IBM. Depuis, HARDIS compte parmi les fleurons des entreprises grenobloises !

Années 1990 : à l'IMAG, un SGBD-R prototype nommé MICROBE et autres projets

Dans la cadre d'un projet national, une équipe de chercheurs met au point un SGBD écrit en langage **PASCAL**, qui sera utilisé au **CNET** dans le cadre d'un projet européen de CAO en microélectronique. L'équipe du CNET en effectue la mise au point. MICROBE intègre une technologie de traitement de l'arbre relationnel très performante, en mode « pipe-line ». Les **besoins spécifiques de la CAO** mettent en évidence les lacunes des SGBD-R pour gérer des structures complexes et les réflexions convergent soit vers le concept naissant des **SGBD Orientés Objets (SGBD-OO)**, soit vers une extension du modèle relationnel (concepts de surrogates). Les travaux du CNET donneront lieu à une publication internationale à la DAC 86 à Las Vegas.

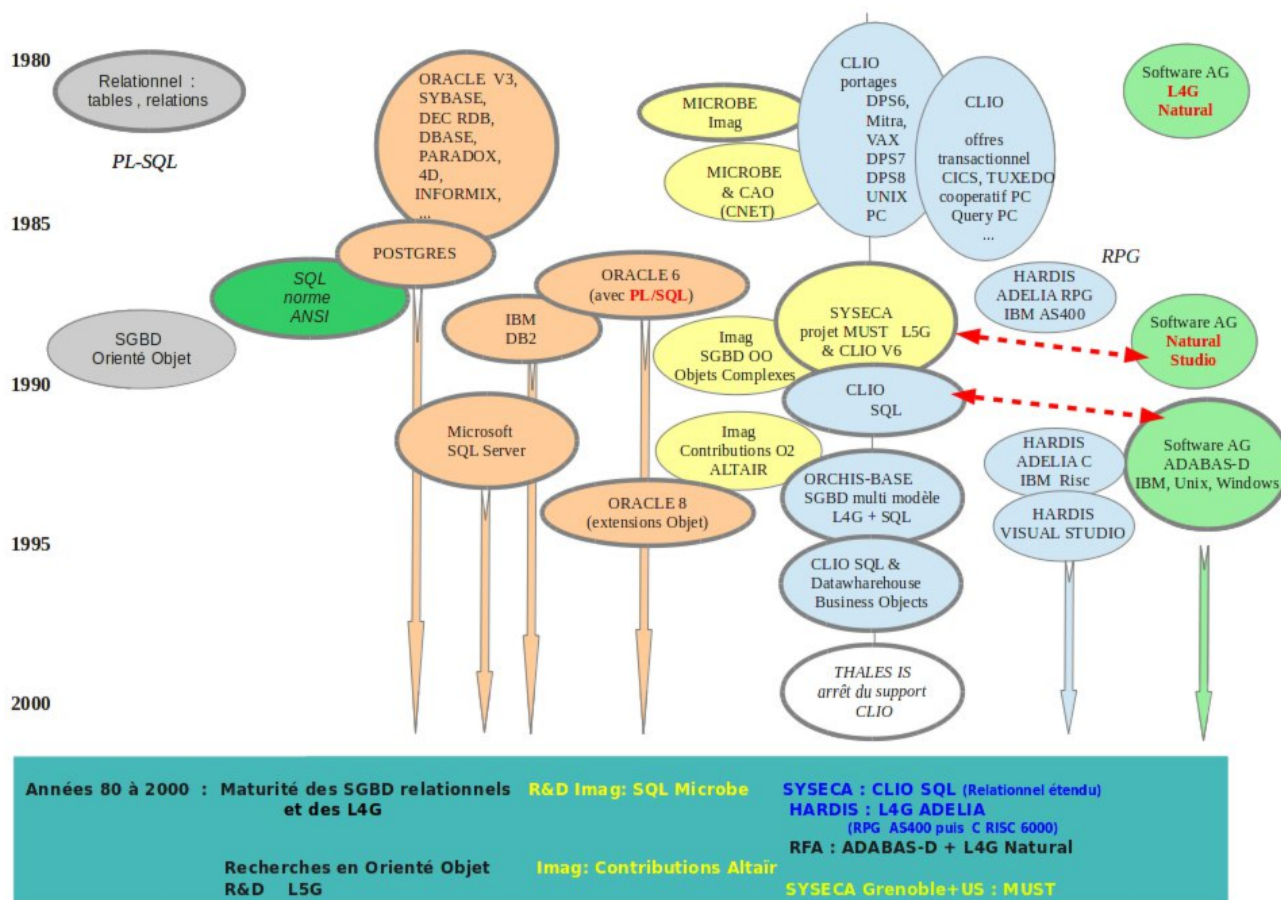
Dans les années 1990, les équipes de l'IMAG se concentrent sur le concept de SGBD étendu aux aspects multimédia et CAO (projets **TIGRE, FAKIR, STORM**), en liaison avec des industriels (SIG MATRA, base de sons de l'Institut de la Communication Parlée, applications médicales). La compétence est forte et la capacité à résoudre des cas concrets est appréciée des industriels. Cependant, les SGBD orientés objets n'émergent pas et le SQL relationnel étendu se concrétise par des extensions dans les produits d'ORACLE, qui reste leader sur le marché. La norme SQL 99 intégrera ces concepts.

THOMSON/SYSECA se désengage de l'édition logicielle

En 1990, SYSECA a fortement investi dans le projet MUST, qui apporte ses premiers résultats, mais sans produire de ventes. Grâce à l'expérience de **Jean-Claude Favre**, qui a travaillé à San José, une version industrielle de **CLIO SQL** intégrant des idées issues de CLIO, MICROBE et de la CAO du CNET voit le jour, grâce à l'enthousiasme d'une équipe de 10 personnes. A la demande du Ministère de l'industrie, le centre de Grenoble est chargé d'industrialiser le SGBD prototype **SABRINA**, issu du projet **SABRE**. Mais malgré un travail très important de restructuration pour amener ce système à un état industriel, sa vente fut un échec.

Des décisions stratégiques conduisent alors SYSECA à arrêter le projet MUST et ses efforts commerciaux sur CLIO. A Grenoble, l'effectif passe en quelques mois de 140 à 28 personnes, se terminant en plan social. CLIO-SQL sera préservé et packagé avec une documentation moderne par l'équipe restreinte, pour satisfaire aux besoins des clients. Nommé **ORCHIS-base**, il restera utilisé au travers des **interfaces ODBC de Microsoft**, développées à Grenoble. Un partenariat avec

Business Objects (société rachetée en 2008 par SAP) permettra aux clients de mettre en œuvre **des entrepôts de données** dès 1995, **prémices du Big Data**. Jusqu'au départ en retraite du dernier ingénieur support CLIO vers 2015, la maintenance des produits était assurée par l'agence THALES-Services de Grenoble.



Dernières avancées côté IMAG et conclusions

L'essor du Big Data et l'émergence du « data science » ont remis à l'ordre du jour l'exploitation de grosses masses d'informations provenant d'observations du monde « réel ». **Christine Collet** et son équipe ont revisité les outils scientifiques et technologiques existants pour gérer et exploiter ces masses de données, répondant ainsi à de nouvelles applications ciblées (**polystores de systèmes**). Il faut noter que les réponses aux défis à relever ne peuvent être obtenues sans une forte synergie entre les scientifiques de tous les instituts, établissements et entreprises collectant et exploitant des masses de données dans des domaines comme la santé publique, la médecine, le sport, la biologie, la physique, l'astrophysique, l'énergie, la défense, l'environnement, les sciences de l'univers, la cosmologie, les sciences humaines et sociales, etc.

Ainsi se sont écoulées 50 ans de « l'école grenobloise des bases de données », selon les termes de **Jacques Voiron**, ancien professeur. Durant cette période, Grenoble a-t-elle eu en main les clés du Big Data ? Elle en a eu les compétences, avec la maîtrise industrielle en édition de SGBD. Grâce à des alliances industrielles en Europe, les grenoblois auraient pu continuer à être un acteur majeur du domaine. L'abandon du projet **UNIDATA** par les français a certainement été un moment clé dans la

perte de cette maîtrise européenne. Le seul SGBD industrialisé en France à une échelle importante aura donc été SOCRATE/CLIO, à Grenoble. Malheureusement, la culture de son actionnaire n'intégrait pas l'édition de logiciels.

- - -

NOTE :

Plus d'une trentaine de références de noms de personnes, marques ou produits est détaillée sous forme de liens dans le travail posté sur le site de l'ACONIT, cité en entrée de cet article.