

0. Préface

Ce livre traite de la modélisation des données.

Il est destiné aux novices et aux personnes qui ont déjà une certaine expérience du domaine, aux informaticiens aussi bien qu'à des collaborateurs non-informaticiens impliqués dans des projets informatiques.

Pour ceux qui ne connaissent encore rien du sujet, modélisation des données signifie: créer, dans le contexte d'une organisation ou d'une application informatique la spécification des données qui seront enregistrées, stockées, exploitées. Plus précisément: définir la *structure* de stockage des données dans ce qu'on appelle une *base de données*. Tous ces termes sont définis à l'intention des novices dans le chapitre prochain. Le reste de cette préface ne leur sera d'ailleurs pas d'une grande utilité et ils peuvent sans autres passer au chapitre 1.

Pourquoi ce livre?

Question: encore un livre sur la modélisation des données?

Réponse: enfin un livre couvrant vraiment la modélisation des données.

La réalité, c'est qu'il n'existe que peu d'ouvrages modernes sur la modélisation des données; il n'en existe même pas du tout. On trouve bien des dizaines de livres sur les bases de données, sur les systèmes de gestion de bases de données, sur SQL, sur la normalisation des données. Sur la modélisation, point. Certes, chacun des ouvrages mentionnés aborde le sujet, mais il ne s'agit souvent que de maigres chapitres ne contenant que des exemples primitifs et scolaires. La majorité d'entre eux sont écrits par des théoriciens ou enseignants sans expérience dans la pratique, alors que nous nous intéressons ici à la modélisation des données dans un contexte réel.

Deux ouvrages font exception, mais ils datent des années 1980, de l'époque glorieuse de l'avènement des SGBD (hiérarchiques et en réseau d'abord, relationnels ensuite) et, dans le contexte francophone, de la méthode Merise. Le livre *Conception d'un système d'information – Construction de la base de données* d'Hubert Tardieu, Dominique Nancy et Daniel Pascot (1979), composante du patrimoine Merise, aborde le sujet et fournit des exemples intéressants, mais ne va pas assez loin dans le sens que nous préconisons et souffre en outre de l'orientation bases de données en réseau, répandues à l'époque. Le livre *Modélisation des données* du Suisse Max Vetter, paru en 1992 en français chez Dunod (dans une traduction de l'allemand malheureusement médiocre), résolument relationnel cette fois, souffre du même syndrome (trop de théorie, pas assez de pratique). Nous pourrions citer bien d'autres exemples. Tout cela n'est pas étonnant, car ces ouvrages étaient en premier lieu destinés à des étudiants dans les écoles. Le but recherché était d'enseigner les principes de base et de permettre aux lecteurs d'effectuer quelques exercices pour assimiler ces règles, sans jamais aborder les véritables problèmes que pose la modélisation des données dans la réalité.

Notre objectif ici est d'enseigner l'art de la modélisation des données, de trouver des solutions qui satisfont aux besoins métier des organisations qui entreprennent des projets d'informatique. Notre expérience de formateur dans ce domaine, de concepteur de bases de données dans la pratique et d'accompagnateur d'équipes de modélisation nous a montré que la majorité des informaticiens (qui ont rarement lu des ouvrages sur le sujet et sont souvent rentrés bredouilles de cours ne leur apportant que peu de soutien pour leur travail quotidien) se trouvent confrontés à d'énormes difficultés lorsqu'il s'agit de concevoir dans la réalité un système de données adéquat.

Un art, la modélisation des données?

La conception d'une base de données ne s'appuie-t-elle pas sur la théorie de la normalisation, à son tour basée sur l'algèbre relationnelle, et ne constitue-t-elle donc pas une science? Loin de là! Tout d'abord la théorie de la normalisation n'apporte que très peu d'aide à la conception d'une base de données, disons le immédiatement! Comparons-là à la grammaire dans le contexte linguistique. Un texte

peut être absolument impeccable sur le plan grammatical et n'exprimer que des âneries. Il en est de même des bases de données qui peuvent être normalisées jusqu'à la nième forme normale et malgré tout ne pas correspondre aux besoins de ceux qui l'exploitent. En outre, la normalisation, essentielle à l'époque du passage des systèmes de fichiers aux bases de données, a entretemps beaucoup perdu de son importance (et de sa gloire). Et qu'en est-il du formalisme entités – associations^{*)}, présenté en long et en large dans tous les ouvrages? Ici également, il ne s'agit pas d'une science exacte, mais d'une technique qui peut être bien ou mal appliquée.

La modélisation des données est véritablement un art. Elle exige du talent, du feeling, de l'expérience, une grande connaissance du domaine concerné et, bien sûr, beaucoup de travail. Comme pour d'autres arts et disciplines, certaines personnes possèdent un don pour cette activité et d'autres malheureusement moins ou pas du tout. Il est en tous cas illusoire d'espérer qu'il suffit d'appliquer quelques règles et raisonnements simplistes pour obtenir un modèle des données adéquat. Dans ce livre, nous présentons une approche empirique accompagnée de quantité de recettes. Très peu de règles absolues y sont énoncées. Concevoir une base de données implique dans un premier temps d'écouter et de comprendre. D'apprendre à connaître le domaine concerné et les besoins des futurs utilisateurs. De réfléchir longtemps à la meilleure solution et de tester exhaustivement celle qu'on va proposer. Sachant qu'une base de données restera généralement en place très longtemps, mieux vaut en outre qu'elle soit apte à évoluer en fonction des besoins et capable d'assumer des montées en charge importantes. Capacités qui dépendent aussi bien des qualités techniques du SGBD utilisé que d'une conception judicieuse.

L'activité de conception d'une base de données reste-elle d'actualité?

Dans les années 80, les ouvrages traitant de la modélisation des données parlaient de modèles de données globaux d'une entreprise, couvrant donc la totalité des besoins en matière de stockage d'informations de toute l'organisation. Cette approche sous-entendait une refonte complète du système d'information en remplacement d'applications isolées développées par le passé. Ce nouveau système d'information serait axé autour d'une base de données unique et intégrée utilisée par l'ensemble des programmes, sans exception. C'était également l'époque où la majorité des entreprises développaient encore leurs propres applications ou les faisaient développer par des sociétés de services. Les progiciels de gestion, tels que nous les connaissons aujourd'hui dans tous les domaines, n'existaient simplement pas encore ou n'étaient pas suffisamment riches en fonctionnalités pour satisfaire les besoins précis d'un grand nombre d'entreprises.

Ces temps ont évidemment bien changé! Des solutions commerciales existent dans la majorité des domaines, aussi bien pour les grandes entreprises que pour les PME, et très rares sont les organisations qui développent encore leur système de gestion elles-mêmes. D'autre part, le coût du développement logiciel est tel et les compétences requises si pointues que peu de sociétés ont encore les moyens de se lancer dans une telle aventure. Le mot d'ordre est aujourd'hui "Concentrons-nous sur notre métier et laissons le développement informatique aux sociétés dont c'est la profession!".

Exemple d'un exercice dans un manuel sur les bases de données: "Une bibliothèque veut informatiser ses opérations et vous demande de concevoir la base de données". Réponse correcte: " Plutôt que de démarrer un projet qui va vous mener à une catastrophe, achetez un progiciel. Il comprendra déjà la base de données nécessaire à la gestion d'une librairie".

Justification des projets de modélisation

Toutefois les progiciels standards couvrent rarement l'ensemble des besoins métiers qui se présentent dans chaque cas particulier. Il faudra donc soit les modifier ou les étendre (aventure extrêmement hasardeuse et coûteuse!), soit réaliser des systèmes annexes qui doivent évidemment communiquer avec ce progiciel. Pas une sinécure non plus, même si les éditeurs fournissent aujourd'hui les outils servant à faciliter une telle interconnexion. Les collaborateurs de sociétés de services attelés à de tels développements seront donc confrontés à l'activité de modélisation. Et, même s'ils sont plus rares aujourd'hui,

^{*)} Nous évitons ici le terme *relation*, utilisé dans un autre sens dans la théorie relationnelle.

les "mégaprojets" informatiques existent encore. Nous parlerons donc également de la modélisation des données dans une telle situation.

Souvent se présente en outre le besoin de développer des systèmes de dimension plus modeste, dévolus à des tâches particulières, qu'il faudra bien concevoir dans leur entièreté. Et s'il est un domaine où la conception de la base de données reste indispensable, ce sont les data warehouses ("entrepôts de données"), bases dans lesquelles sont enregistrées les données servant à la création des tableaux de bord, statistiques et rapports. En général ces bases reçoivent des informations provenant de l'ensemble des progiciels utilisés dans l'entreprise. Il s'agit d'entreposer ces données de manière à en permettre l'exploitation la plus judicieuse et la plus efficace.

Bref, la modélisation des données reste d'actualité. Et lorsqu'elle doit être effectuée, elle constitue certainement l'aspect le plus important du développement d'un système: si la base de données est mal conçue, aucune programmation, même géniale, ne sauvera le projet. Le présent ouvrage met, principalement pour des raisons pédagogiques, l'accent sur la modélisation des données pour une application "insulaire", une application qui n'est donc pas fortement intégrée au système de gestion global. Ce qui ne signifie pas qu'une telle base de données existe en dehors de tout contexte. Très probablement un échange de données avec le reste du système de gestion sera nécessaire, nous parlerons de ce que cela implique.

Finalement, une modélisation des données peut s'avérer nécessaire avant la sélection d'un outil de gestion, ceci dans le but de définir les besoins. Lequel parmi les candidats correspond-il le mieux aux besoins? On n'a souvent considéré que les fonctionnalités des solutions pour répondre à cette question. Un examen approfondi de la base de données sous-jacente à ces produits nous paraît tout aussi important.

Résumé et déclaration d'intention

Cet ouvrage n'est donc pas un livre sur les bases de données ou les SGBD en général, ni sur la normalisation, ni sur la théorie relationnelle, même si ces aspects y sont abordés et y reçoivent la place qu'ils méritent.

On n'y parle pas non plus (ou très peu) des problèmes liés à l'implémentation d'une base de données avec un SGBD particulier, ni des performances, ni de la sécurité: il s'agit d'aspects concernant l'administrateur de la base et les ouvrages traitant de ces sujets abondent. Le problème des performances ne sera abordé que dans le contexte du choix entre une solution par rapport à une autre.

Il n'est question ici que d'un seul sujet: la modélisation des données d'une application ou d'une entreprise (d'une collectivité) dans le but de concevoir une base de données dont le contenu correspond aux exigences de cette entreprise et de ses utilisateurs. Que cette base de données contienne les informations voulues et que ces informations y soient structurées de manière aussi judicieuse que possible. L'accent est essentiellement mis sur les besoins du métier. Tout le reste n'est que technologie.

Et, finalement, l'approche est conçue en vue de l'implémentation avec un **SGD relationnel**. Les systèmes en réseau ou hiérarchiques sont relégués au musée, les SGBD objets n'ont pas réussi leur percée et les bases de données XML constituent encore de la musique d'avenir.

Nous faisons par contre un usage étendu de certaines notions liées à l'orientation objet, ceci dans la mesure où elles apportent vraiment un avantage et peuvent être mises en œuvre sans problème.