

Université Lumière Lyon 2
École Doctorale Informatique et Information pour la Société
THÈSE pour obtenir le grade de Docteur en Informatique présentée et soutenue publiquement par
Cécile Favre
le 12 Décembre 2007

Évolution de schémas dans les entrepôts de données : mise à jour de hiérarchies de dimension pour la personnalisation des analyses

Préparée au sein du laboratoire ERIC
sous la direction de M. Omar Boussaid et Mme Fadila Bentayeb
Thèse au format PDF

COMPOSITION DU JURY Mme Rokia MISSAOUI Rapportrice (Professeur, Université du Québec en Outaouais) M. Gilles ZURFLUH Rapporteur (Professeur, Université Toulouse 1) Mme Corine CAUVET Examinatrice (Professeur, Université Aix-Marseille 3) M. Djamal BENSLIMANE Examineur (Professeur, Université Lyon 1) M. Abdelkader Djamel ZIGHED Examineur (Professeur, Université Lyon 2) Mme Fadila BENTAYEB Co-directrice de thèse (Maître de Conférences, Université Lyon 2) M. Omar BOUSSAID Directeur de thèse (Maître de Conférences-HDR, Université Lyon 2) M. Michel ROUGIE Invite (LCL, Direction Rhône-Alpes Auvergne)

Table des matières

| | |
|--|-----------|
| Contrat de diffusion . | 1 |
| Résumé . | 3 |
| Introduction générale . . | 5 |
| 1.1 Contexte . . | 5 |
| 1.2 Problématique . . | 9 |
| 1.3 Contributions . . | 11 |
| 1.4 Organisation du mémoire . . | 12 |
| Entrepôt de données : conception et exploitation . | 15 |
| État de l'art . | 17 |
| Modèle d'entrepôt de données à base de règles . . | 19 |
| Mise à jour des hiérarchies de dimension . . | 21 |
| Vers une évaluation des modèles d'entrepôt de données évolutifs . | 23 |
| La plateforme WEDriK . | 25 |
| Développement industriel : la plateforme MARKLOC . . | 27 |
| Conclusion générale . | 29 |
| Publications . | 31 |
| Bibliographie & Annexes . . | 33 |

Contrat de diffusion

Ce document est diffusé sous le contrat Creative Commons « Paternité – pas d'utilisation commerciale - pas de modification » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

Résumé

Cette thèse a été réalisée dans le cadre d'une Convention Industrielle de Formation par la REcherche (CIFRE) en collaboration avec l'établissement bancaire LCL- Le Crédit Lyonnais. Elle s'inscrit dans le domaine des entrepôts de données. Ces derniers constituent un élément incontournable de l'architecture décisionnelle, sur lesquels reposent alors des outils permettant l'analyse en ligne des données (OLAP : On Line Analytical Processing) pour l'aide à la décision. Le schéma de l'entrepôt, qui détermine les possibilités d'analyse, est conçu en fonction des sources de données disponibles d'une part et des besoins d'analyse d'autre part. Or, il est difficile d'être exhaustif dans le recensement des besoins d'analyse des utilisateurs au moment de la conception du schéma de l'entrepôt. En outre, de nouveaux besoins individuels peuvent émerger. L'émergence de nouveaux besoins d'analyse individuels fait alors apparaître la nécessité d'une personnalisation des analyses, qui placerait l'utilisateur au cœur du processus décisionnel. Dans cette thèse, nous proposons une solution à la personnalisation des analyses dans les entrepôts de données. Cette solution se base sur une évolution du schéma de l'entrepôt guidée par les utilisateurs. Il s'agit en effet de recueillir les connaissances de l'utilisateur et de les intégrer dans l'entrepôt de données afin de créer de nouveaux axes d'analyse. Afin de développer cette solution, nous avons proposé quatre contributions majeures : 1) Notre première contribution consiste en la définition d'un modèle formel d'entrepôt de données évolutif, basé sur des règles «si-alors», que nous appelons règles d'agrégation. Ce modèle est composé d'une partie «fixe» et d'une partie «évolutive». La partie fixe est constituée de la table des faits et des tables de dimension qui lui sont directement reliées. La partie évolutive est composée d'un ensemble de hiérarchies de dimension qui sont mises à jour. Pour assurer la généralité de notre approche, nous proposons également un méta-modèle qui permet de décrire tout entrepôt de données évolutif. 2) Notre modèle d'entrepôt évolutif est soutenu par une architecture qui permet de modéliser le processus de personnalisation. Cette architecture comprend quatre modules : – un module d'acquisition des connaissances utilisateurs sous forme de règles d'agrégation ; – un module d'intégration des règles d'agrégation dans l'entrepôt de données ; – un module d'évolution du schéma ; – un module d'analyse permettant à l'utilisateur de réaliser des analyses sur le nouveau schéma. 3) Pour mettre en œuvre cette architecture globale, nous proposons un modèle d'exécution avec l'approche relationnelle, qui vise à gérer l'ensemble des processus liés à l'architecture globale. Il est fondé sur la transformation des règles d'agrégation en une table relationnelle de mapping qui permet le stockage, la vérification des règles, la création des niveaux de hiérarchie. 4) Nous nous sommes par ailleurs intéressés à l'évaluation de la performance de notre modèle d'entrepôt de données évolutif. Or, l'évaluation de la performance des modèles est généralement basée sur une charge (ensemble de requêtes utilisateurs). Lorsqu'un changement au niveau du schéma de l'entrepôt de données se produit, la charge doit être mise à jour. Dans ce contexte, nous proposons ici une méthode de mise à jour incrémentale de la charge. Pour valider nos différentes contributions, nous avons développé la plateforme WEDriK (data Warehouse Evolution Driven by Knowledge), qui permet la personnalisation des analyses. Elle se base sur un entrepôt de données évolutif stocké dans le SGBD relationnel Oracle d'une part et sur une interface Web programmée en PHP d'autre part. Les problèmes posés dans ce mémoire sont directement issus de la réalité de l'entreprise LCL avec laquelle nous avons collaboré. LCL a constitué un véritable terrain d'application pour mettre en œuvre nos solutions de personnalisation. Nous nous sommes également intéressés à la personnalisation dans sa définition plus classique, dans le cadre de la gestion des interfaces et de la recherche

Évolution de schémas dans les entrepôts de données : mise à jour de hiérarchies de dimension pour la personnalisation des analyses

d'information, au travers du travail d'ingénierie que nous avons réalisé pour cette entreprise durant le développement d'une plateforme pour la gestion des demandes de marketing local : la plateforme MARKLOC.

Mots-clés : entrepôt de données, évolution de schéma, hiérarchie de dimension, mise à jour, personnalisation, utilisateur, règles d'agrégation, analyse en ligne, performance, évolution de charge.

Introduction générale

1.1 Contexte

Nos travaux de thèse ont été réalisés dans le cadre d'une Convention Industrielle de Formation par la REcherche (CIFRE). L'objectif d'une convention CIFRE est d'associer, autour d'un projet de recherche, trois partenaires : une entreprise, un laboratoire et un(e) doctorant(e). L'Association Nationale de la Recherche Technique (ANRT) a accepté les partenaires suivants pour la convention portant l'identifiant 735/2003 : l'établissement bancaire LCL-Le Crédit Lyonnais (note LCL par la suite), le laboratoire ERIC (Equipe de Recherche en Ingénierie des Connaissances) et moi-même dans le rôle de doctorante. Le projet de recherche a été défini en collaboration avec les différents partenaires et avait pour but de pouvoir apporter des solutions, issues d'un travail de recherche, à des besoins réels. Ces besoins, pour LCL, se situent dans le contexte du marketing local. LCL, qui constitue la première banque nationale de proximité en France, dispose d'une double politique marketing. Sur un plan national, des actions multi-canaux (courrier, téléphone, courriel) sur des produits identifiés sont menées auprès des clients. En complément de ces actions nationales, des actions au niveau local peuvent être menées pour répondre à des besoins spécifiques émergeant au niveau local (rattrapage de retard sur des objectifs commerciaux, création d'une nouvelle agence, etc.). Ces actions sont proposées par les

responsables commerciaux eux-mêmes, à différents niveaux hiérarchiques. En effet, l'organisation commerciale de LCL correspond à une structure hiérarchique pyramidale. Au début du projet, cette hiérarchie se présentait sous la forme de la pyramide de la Figure 1.1 qui inclut également à sa base un niveau client. Ce niveau ne fait pas partie à proprement parler de la structure organisationnelle hiérarchique, mais il permet de situer les clients, qui ont bien évidemment un rôle prépondérant pour l'établissement.

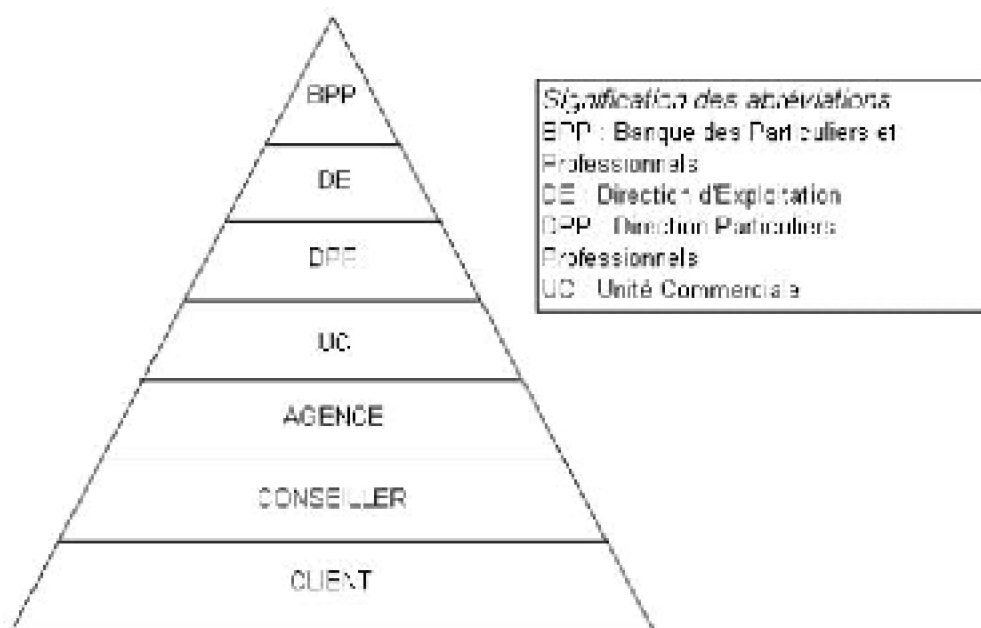


FIG. 1.1 – Structure hiérarchique de LCL au début du projet

Au sommet de cette pyramide, on trouvait la banque des particuliers et des professionnels (BPP), elle était composée d'un ensemble de huit directions d'exploitation (DE) qui correspondaient à un découpage géographique (Figure 1.2) : le Nord Ouest (DENO), Paris (DEP), le Bassin Parisien Sud (DEBPS), l'Ouest (DEO), l'Est (DEE), le Sud-Ouest (DESO), la Méditerranée (DEM) et Rhône-Alpes Auvergne (DERAA). Dans le cadre de cette thèse, nous avons travaillé au niveau d'une DE, en l'occurrence, celle de la région Rhône-Alpes Auvergne, au service à la fois de cette direction, mais également de toutes les unités des niveaux hiérarchiques qui dépendent de son périmètre géographique. En effet, chaque DE est ensuite organisée selon des niveaux

hiérarchiques successifs, découpés selon des périmètres géographiques dirigés par des responsables commerciaux. Ainsi, chaque DE était composée d'un ensemble de directions particuliers-professionnels (DPP), chaque DPP était composée d'un ensemble d'unités commerciales (UC), chaque UC comprenait un ensemble d'agences. Chaque agence dispose d'un ensemble de conseillers qui sont en contact direct avec les clients. La DERA, avec laquelle nous avons collaboré, comprenait une dizaine de DPP (citons par exemple la DPP DRÔME ARDÈCHE HAUTES-ALPES), une soixantaine d'UC (une des UC de la DPP citée en exemple est l'UC ANNONAY) et environ 300 agences bancaires telles que nous les connaissons.

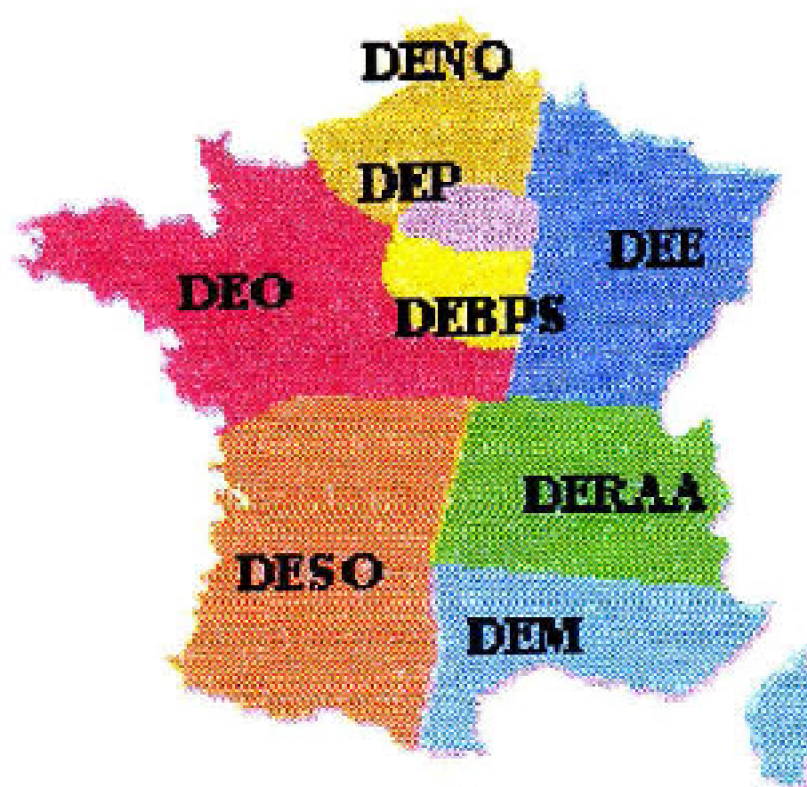


FIG. 1.2 – Périmètre géographique des directions d'exploitation

En observant cette pyramide, nous pouvons constater le rôle primordial des conseillers commerciaux qui sont au contact des clients et, par conséquent, réalisent les ventes. Il s'agit alors, pour l'ensemble de la hiérarchie, de tout mettre en œuvre pour aider et optimiser leur travail. Ainsi, pour compléter la politique marketing nationale et réaliser ainsi des ventes additionnelles, les responsables commerciaux en région Rhône-Alpes sont amènes à faire des demandes de marketing local afin de mener des actions. Une

demande de marketing local est la formulation d'une demande de ciblage de clients pour une action marketing ponctuelle, qui n'est autre qu'une opération spécifique à un produit ou à un événement particulier. Elle se traduit par l'extraction d'une liste de clients répondant à certains critères, autrement dit elle cible certains profils de clients. Cette liste permet aux conseillers commerciaux de contacter leurs clients pour un motif précis, en l'occurrence celui lié à l'action marketing en question. Par exemple, un responsable peut faire une demande de marketing afin de mener une action pour vendre des plans d'épargne logement dans l'agence Annonay suite à un retard sur les objectifs à atteindre pour ce produit. Auparavant, la gestion des demandes de marketing local se faisait de façon manuelle, via des fiches de liaison papier, transmises par fax pour la validation de niveau hiérarchique en niveau hiérarchique. Une fois les accords successifs des différents niveaux hiérarchiques d'un point de vue commercial, un accord de faisabilité devait être donné par le Pole Outils et Méthodes de la direction d'exploitation, qui réalisait le ciblage si c'était possible. Ce fonctionnement posait un certain nombre de problèmes, au-delà du fait qu'il n'était pas pratique. Tout d'abord, il y avait parfois un problème au niveau des informations données pour la demande. En effet, le modèle de fiche de liaison « officielle » avait évolué, mais il subsistait différentes versions utilisées dans le réseau commercial, versions non conformes aux attentes pour la réalisation du ciblage. De plus, comme ces demandes étaient formulées sur un support papier, la recherche d'informations était trop coûteuse en temps et en ressources humaines donc inefficace. En outre, il n'y avait pas de suivi des résultats commerciaux liés à cette demande. Ainsi, si la demande n'était pas relayée au niveau des conseillers, on ne pouvait que constater après la fin de l'action que celle-ci n'avait pas donné de bons résultats commerciaux. Enfin, ne disposant pas d'un système informatisé, il était très difficile, voire impossible, d'analyser les demandes elles-mêmes et donc de capitaliser les connaissances acquises lors des précédentes demandes. Cette collaboration industrielle et les problèmes concrets que nous avons été amenés à résoudre ont fait émerger deux types d'objectifs : des objectifs liés à un travail d'ingénierie et d'autres relevant d'un travail de recherche. En effet, il était nécessaire d'apporter une solution pour la gestion des demandes marketing en disposant d'un système informatisé qui permette de gérer l'ensemble du processus lié à celles-ci : de l'acquisition de la demande, à l'analyse des résultats, en passant par la validation hiérarchique. Préalablement, nous avons réalisé une étude conséquente de l'existant et des besoins utilisateurs en exploitant différents moyens (observation, interview et enquête), afin de déterminer le cahier des charges à satisfaire. En faisant émerger les besoins réels et en fournissant une réponse à ces besoins, nous nous assurons de l'adhésion future des utilisateurs au système qui serait développé. L'objectif en termes d'ingénierie a été atteint dans la mesure où LCL dispose aujourd'hui d'une plateforme dédiée, que nous avons développée, nommée MARKLOC. Cette plateforme est supportée par une base des demandes marketing que nous avons conçue et créée, ainsi qu'un processus de flux de travail que nous avons implémenté pour assurer le suivi de la chaîne de validation et de réalisation. Cette plateforme fait partie intégrante de l'intranet de la DE Rhône-Alpes Auvergne et est utilisée quotidiennement par les responsables commerciaux des différents niveaux hiérarchiques pour différentes raisons : émission de demandes, validation de demandes, suivi de résultats commerciaux liés à une demande, etc. Nous reviendrons plus en détails sur les différents aspects techniques de cette

plateforme dans le chapitre 8. Par la suite, notre objectif était de proposer un système décisionnel qui aide LCL à analyser efficacement ces demandes de marketing local, pour permettre une capitalisation des connaissances. Pour atteindre cet objectif, différentes sources de données hétérogènes et autonomes potentiellement intéressantes étaient à notre disposition, dont la base des demandes marketing que nous avons conçue. Il était donc nécessaire d'avoir recours à un processus d'intégration de données dédiée à l'analyse. Nous avons donc conçu un entrepôt de données en fonction des sources dont nous disposions et des besoins d'analyse que nous avons identifiés. Une des caractéristiques phares de la modélisation des entrepôts est l'historisation des données. Les schémas multidimensionnels classiques permettent de garder effectivement trace de l'historique des faits, de leur évolution, grâce à une dimension Temps. Paradoxalement, ils s'approprient mal à prendre en compte les changements pouvant être opérés sur les dimensions, aussi bien du point de vue de leurs données, que de leur structure. Ceci est dû à l'hypothèse d'orthogonalité des dimensions qui est généralement faite [EK01]. Ainsi faire l'hypothèse que chacune des dimensions est orthogonale à la dimension Temps signifie faire l'hypothèse que chacune des dimensions est temporellement invariante. Néanmoins, cette hypothèse n'est pas vérifiée dans la pratique puisque l'évolution de la structure et des données des dimensions est bien réelle. En effet, des produits apparaissent ou disparaissent, les structures organisationnelles se modifient, de nouveaux clients arrivent, certains partent et pour d'autres clients, leurs caractéristiques peuvent être modifiées (statut familial, pouvoir d'achat, etc.). Cette évolution des dimensions est due à l'évolution des sources de données d'une part et à celle des besoins d'analyse d'autre part, comme c'est le cas chez LCL. En premier lieu, les sources de LCL évoluent. En effet, LCL est un établissement au sein duquel se produisent de nombreux changements, parmi lesquels des changements organisationnels. Par exemple, en juin 2003 s'opérait le rachat par le groupe Crédit Agricole SA. S'en suivait quelques années plus tard une réorganisation complète de la structure commerciale que l'on retrouvait dans les sources de données de LCL. En second lieu, les besoins d'analyse peuvent eux-mêmes évoluer. De nouveaux besoins peuvent émerger, en réaction à l'évolution des données par exemple, ou tout simplement parce que les utilisateurs ont des besoins qui n'avaient pas été recensés lors de la conception de l'entrepôt de données et parce qu'il est difficile de prédire les besoins à venir. LCL est un établissement regroupant des employés exerçant divers métiers et ayant donc des besoins d'analyses variés. Ainsi, de nouveaux besoins individuels peuvent émerger.

Les sources de données, tout comme les besoins d'analyse ne sont pas figés dans le temps. Étant donné que le schéma de l'entrepôt est conçu en fonction des sources de données disponibles et des besoins d'analyse, qu'advient-il de ce schéma lors de leur évolution ? C'est dans ce contexte que nous nous sommes intéressés à la problématique de l'évolution de schémas dans les entrepôts de données.

1.2 Problématique

Les entrepôts de donnée sont nés au sein des entreprises pour répondre à un besoin crucial d'analyse pour l'aide à la décision. Un entrepôt de données peut être vu comme une grosse base de données modélisée pour accueillir, après nettoyage et homogénéisation, les informations en provenance des différents systèmes de production de l'entreprise. L'un des points-clés de la réussite de l'entrepôt réside alors dans la conception du schéma de l'entrepôt qui doit permettre de répondre aux besoins d'analyse pour l'aide à la décision. Il est vrai que la création de l'entrepôt n'est pas un objectif final. L'important, c'est l'exploitation qui s'en suivra. La vocation de l'entrepôt de données est de supporter l'analyse des données. Ainsi, l'entrepôt de données constitue un support pour les outils d'analyse en ligne issus de la technologie OLAP («On-Line Analytical Processing»). Ce sont ces outils qui permettent d'accompagner les décideurs d'une entreprise en leur fournissant une possibilité de naviguer facilement dans les données, et de construire ainsi des informations opératoires. Ainsi l'entrepôt de données permet d'offrir des contextes d'analyses aux utilisateurs pour les aider dans leur prise de décision. Il s'avère alors que les utilisateurs peuvent avoir besoin de contextes d'analyses spécifiques, répondant à des besoins particuliers voire individuels. L'émergence de nouveaux besoins d'analyse individuels fait alors apparaître la nécessité d'une personnalisation des analyses, qui placerait l'utilisateur au cœur du processus décisionnel. En effet, le processus d'entrepôt de données est souvent associé à une idée de technologie centrée utilisateur. Cependant, cette place centrale se limite au fait que l'utilisateur mène son analyse en naviguant dans les données et ce, selon le schéma de l'entrepôt prédéfini. Il est donc intéressant de redonner tout son sens à l'expression «centrée utilisateur». C'est l'un des objectifs de cette thèse. La personnalisation n'est pas un principe nouveau, elle fait l'objet de nombreux travaux dans des domaines tels que la recherche d'informations, les bases de données. Cependant, elle constitue un axe de recherche récent dans le domaine des entrepôts de données, alors même que les caractéristiques de ces derniers lui sont favorables. En effet, la volumétrie des données connue pour être importante dans les entrepôts de données et le rôle central que joue l'utilisateur dans le processus décisionnel, au niveau de l'analyse en ligne, sont deux éléments qui justifient pleinement le recours à la personnalisation. Il est vrai que le principe général de la personnalisation, dans un contexte de recherche d'informations par exemple, est de fournir, parmi une multitude de résultats possibles, la réponse (qui peut être un ensemble de résultats) qui sera la plus pertinente pour l'utilisateur. Cette personnalisation doit se baser sur des éléments concernant l'utilisateur lui-même : ses préférences, ses habitudes, etc. La conception de magasins de données vise à répondre aux objectifs métiers mais elle ne constitue pas une réponse réelle au besoin de personnalisation. Or, les décideurs qui exploitent les entrepôts de données ont d'importantes connaissances métiers et d'autres connaissances qu'il nous semble intéressant d'exploiter. Ainsi, la question que nous nous posons alors est comment intégrer la connaissance utilisateur dans l'entrepôt de données pour la personnalisation des analyses.

1.3 Contributions

L'objectif général de cette thèse est de proposer une approche qui met à la disposition des analystes des moyens pour intégrer leurs connaissances dans l'entrepôt de données, à des fins d'analyses spécifiques futures. Compte tenu de l'émergence de nouveaux besoins d'analyse, la conception et la construction d'un entrepôt de données dont le schéma est fixe ne satisfait plus les attentes des décideurs. En effet, la construction des axes d'analyse doit évoluer en même temps que les besoins. Précisons que par axes d'analyse, nous entendons, dans l'ensemble de ce mémoire, les possibilités d'analyses. Ainsi, notre première contribution dans le cadre de la personnalisation des analyses consiste en la définition d'un modèle formel d'entrepôt de données évolutif basé sur des règles de type «si- alors», que nous appelons règles d'agrégation. Nous l'avons nommé modèle *R-DW* pour «Rule-based Data Warehouse». Il est composé d'une partie «fixe» et d'une partie «évolutive». La partie fixe est constituée de la table des faits et des tables de dimension qui lui sont directement reliées. La partie évolutive est composée de l'ensemble des hiérarchies, définies lors de la conception du schéma de l'entrepôt, qui peuvent subir des mises à jour. En effet, des niveaux de hiérarchie peuvent être supprimés, d'autres peuvent être créés par l'application des règles d'agrégation. Ainsi, le modèle d'entrepôt de données que nous proposons a un schéma flexible qui permet de prendre en compte l'évolution des besoins d'analyse. Pour assurer la généralité de notre proposition, nous définissons un méta-modèle qui permet de décrire tout entrepôt de données évolutif. Par ailleurs, afin d'impliquer l'utilisateur dans le processus d'évolution de schéma de l'entrepôt de données, nous proposons une démarche basée sur une architecture globale qui nous permet de modéliser le processus de personnalisation. Notons que le terme «utilisateur» désigne à la fois l'utilisateur final mais aussi l'administrateur. Par exemple, lors de la modification de la structure interne de l'entreprise, l'administrateur peut faire évoluer le schéma de l'entrepôt en exprimant des connaissances pour l'ensemble des utilisateurs finaux. Ainsi, notre deuxième contribution est la proposition d'une architecture globale qui vise à soutenir notre modèle d'entrepôt évolutif. Cette architecture globale comprend quatre modules : – un module d'acquisition des connaissances utilisateurs sous forme de règles d'agrégation ; – un module d'intégration des règles d'agrégation dans l'entrepôt de données ; – un module d'évolution de schéma de l'entrepôt qui permet la mise à jour des hiérarchies de dimension – un module d'analyse permettant à l'utilisateur d'avoir de nouvelles analyses OLAP basées sur le nouveau schéma. Notre troisième contribution consiste à mettre en œuvre notre démarche. En effet, nous proposons un modèle d'exécution avec l'approche relationnelle (ROLAP : Relational OLAP), qui gère l'ensemble des processus liés à l'architecture globale. Ce modèle d'exécution est constitué des étapes suivantes. Premièrement, les règles d'agrégation sont représentées dans une table relationnelle, que nous appelons table de mapping. Deuxièmement, le contenu de cette table est contrôlé pour vérifier la cohérence des règles d'agrégation. Troisièmement, si les règles d'agrégation sont cohérentes, le nouveau niveau de granularité est créé ; il prend la forme d'une table

relationnelle liée à une ou plusieurs autres tables selon que le niveau a été ajouté en fin de hiérarchie ou inséré entre deux niveaux existants de la hiérarchie. Par ailleurs, nous nous sommes intéressés à l'évaluation de la performance de notre modèle d'entrepôt de données évolutif. Or, l'évaluation de la performance dans les entrepôts de données est généralement basée sur une charge (ensemble de requêtes utilisateurs). Lorsqu'un changement se produit au niveau du schéma de l'entrepôt de données (suppression d'un niveau de hiérarchie par exemple), la charge doit être mise à jour. Ainsi notre quatrième contribution réside dans la proposition d'une méthode de mise à jour incrémentale de la charge. Pour valider nos différentes contributions, nous avons développé une plateforme appelée WEDriK1 (data Warehouse Evolution Driven by Knowledge). Cette plateforme se base d'une part sur un entrepôt de données évolutif stocké dans le SGBD relationnel Oracle et d'autre part sur une interface Web qui permet l'interaction avec les utilisateurs. Les problèmes posés dans ce mémoire sont directement issus de la réalité de l'entreprise LCL avec laquelle nous avons collaboré. LCL a constitué un véritable terrain d'application pour mettre en œuvre nos solutions de personnalisation. De plus, nous nous sommes également intéressés à la personnalisation dans sa définition plus classique, dans le cadre de la gestion des interfaces et de la recherche d'information, au travers du travail d'ingénierie que nous avons réalisé pour cette entreprise durant le développement de la plateforme dédiée à la gestion des demandes de marketing local MARKLOC, que nous présentons également dans ce mémoire.

1.4 Organisation du mémoire

Ce mémoire est organisé comme suit. Dans la première partie, nous évoquons le contexte de nos travaux sur deux chapitres. Le chapitre 2 présente différents aspects relatifs à la modélisation, à la conception des entrepôts de données et à leur mise en œuvre. Le chapitre 3 est consacré à un état de l'art sur l'évolution de schéma d'une part et sur la personnalisation d'autre part. Nous y détaillons les principaux travaux proposés qui ont traité de la problématique de l'évolution de schéma et plus généralement, de l'évolution de modèle dans les entrepôts de données. Nous proposons également une classification de ces travaux selon divers critères que nous jugeons pertinents. Nous évoquons également le thème de la personnalisation à travers différents domaines qui sont la recherche d'informations, l'interaction homme-machine, les bases de données. Nous présentons également les travaux émergents sur la personnalisation dans les entrepôts de données. Dans la deuxième partie, nous présentons nos contributions scientifiques qui s'articulent selon trois chapitres. Le chapitre 4 présente notre approche d'évolution de schéma guidée par les utilisateurs. Cette approche vise une personnalisation des analyses en prenant en compte les connaissances des utilisateurs eux-mêmes pour étendre les hiérarchies de dimension et par conséquent, enrichir les possibilités d'analyse de l'entrepôt. Nous y détaillons la formalisation de notre modèle *R-DW* d'entrepôt de données évolutif, le méta-modèle qui permet d'assurer la généralité de notre approche,

ainsi que l'architecture globale qui permet d'impliquer l'utilisateur dans le processus d'évolution de schéma. Vient ensuite le chapitre 5 qui vise à présenter la mise à jour des hiérarchies de dimension pour la personnalisation des analyses telle que nous la concevons. Nous y détaillons le processus de mise à jour et de sa propagation dans le schéma. Par ailleurs, nous proposons un modèle d'exécution dans le contexte relationnel qui permet de mettre en œuvre notre démarche de personnalisation jusqu'à la mise à jour des hiérarchies de dimension. Puis, dans le chapitre 6, nous posons le problème de l'évaluation de notre modèle évolutif et nous proposons une méthode de mise à jour incrémentale de la charge en réponse à l'évolution subie par le schéma de l'entrepôt. Dans la troisième partie, nous détaillons les développements que nous avons réalisés à la fois dans le contexte industriel et dans le contexte scientifique pour valider nos contributions. D'un point de vue chronologique, c'est le développement et l'exploitation de la plateforme MARKLOC pour le compte de LCL qui ont finalement suscité la problématique de recherche sur la personnalisation des analyses que nous traitons dans ce mémoire. Pour des raisons de cohérence, nous abordons d'abord le développement réalisé pour la validation de nos propositions, avant d'évoquer le développement industriel. Ainsi, le chapitre 7 présente la validation de nos propositions à travers la réalisation d'un prototype et l'utilisation des données réelles de LCL pour appliquer notre démarche. Préalablement, nous y évoquons donc la construction d'un entrepôt de données test (LCL-DW) avec les sources de données de LCL. Le chapitre 8 aborde nos contributions en termes d'ingénierie au niveau du projet fait en collaboration avec l'entreprise LCL sur les demandes de marketing local. Ainsi, nous y détaillons les éléments constitutifs de la plateforme MARKLOC et nous nous attachons à décrire comment nous avons mis en œuvre le concept de personnalisation dans son développement. Enfin, le chapitre 9 conclut ce mémoire. D'une part nous dressons le bilan de cette thèse et de nos contributions. D'autre part, nous présentons nos perspectives de recherche.

Entrepôt de données : conception et exploitation

Chapitre 2

État de l'art

Chapitre 3

Modèle d'entrepôt de données à base de règles

Chapitre 4

Mise à jour des hiérarchies de dimension

Chapitre 5

Vers une évaluation des modèles d'entrepôt de données évolutifs

Chapitre 6

La plateforme WEDriK

Chapitre 7

Développement industriel : la plateforme MARKLOC

Chapitre 8

Conclusion générale

Chapitre 9

Publications

Publications

Bibliographie & Annexes

Bibliographie & Annexes