

Entreposage virtuel de demandes marketing : de l'acquisition des objets complexes à la capitalisation des connaissances

Cécile Favre¹, Fadila Bentayeb¹, Omar Boussaid² et Nicolas Nicoloyannis² *

Laboratoire ERIC, Université Lumière Lyon2

5, avenue Pierre Mendès-France, 69676, Bron Cedex, France

¹ {cfavre, bentayeb}@eric.univ-lyon2.fr, ² {boussaid, nicoloyannis}@univ-lyon2.fr

Résumé

Les données bancaires se présentent sous différents formats et proviennent de multiples sources hétérogènes. Pour les interroger et les analyser, il est nécessaire de procéder à une intégration des données dans un format unifié. Nous proposons donc une combinaison d'une intégration par médiation et par entreposage. Celle-ci a pour but de permettre la construction de contextes d'analyse (cube de données). A partir d'une requête utilisateur, le système doit pouvoir extraire les données nécessaires et pertinentes à l'analyse et alimenter un système d'aide à la décision.

1 Introduction

Le travail de recherche que nous exposons ici est réalisé dans le cadre d'une Convention Industrielle de Formation par la Recherche (CIFRE), en partenariat avec le Crédit Lyonnais. Le fonctionnement d'un établissement bancaire comme le Crédit Lyonnais est générateur d'un flot très important de données diverses. Celles-ci sont accessibles via un ensemble d'applications, qui composent le système d'informations du Crédit Lyonnais. Ces applications ont été construites de façon indépendante, et sont autonomes et très hétérogènes. Par ailleurs, pour compléter la politique marketing nationale et réaliser ainsi des ventes additionnelles, les responsables commerciaux en région Rhône-Alpes sont amenés à faire des demandes marketing au niveau local. Une demande marketing 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. La concrétisation de cette demande se traduit par l'extraction d'une liste de clients répondant à certains critères (profils de clients). Jusqu'à présent, la gestion des demandes marketing se fait de façon manuelle, ce qui pose un certain nombre de problèmes. Une demande marketing est construite à partir de plusieurs informations en provenance de multiples sources (bases de données ou programmes) et constitue alors un objet complexe. Pour interroger et analyser ces objets complexes, nous proposons une démarche basée sur un entreposage virtuel des demandes de marketing offrant une aide efficace à la prise de décision.

La combinaison de l'intégration par médiation et par entreposage nous semble être une solution pertinente pour la gestion et l'analyse des demandes marketing. Dans notre cas, l'intégration

*Nous remercions Jean-Marc CROS et Michel Rougié, représentants le Crédit Lyonnais, partenaire du laboratoire ERIC dans le cadre d'une Convention Industrielle de Formation par la Recherche (CIFRE).

des données est un contexte particulier. Du fait que les sources de données soient des bases de données ou des applications. Ces dernières génèrent des informations multiples de façon fréquente. La médiation nous dispense ainsi des tâches de rafraîchissement des données qui risquent d'être très coûteuses. Sa vocation est seulement de fournir à l'utilisateur des données. Cependant, celles-ci sont sous un format non approprié à l'analyse. L'approche d'entrepôt contribue alors à remédier à ce problème. Elle permet à l'utilisateur de construire des contextes d'analyses ciblant ses objectifs. C'est ce double besoin de gérer et d'analyser efficacement les demandes marketing qui nous amène à proposer une architecture d'entrepôt virtuel de ces données. Le traitement d'une requête utilisateur par un médiateur et le rapatriement des données utiles à celle-ci par des agents "wrappers" nécessite un travail coopératif pouvant être effectué par un Système Multi Agents (SMA). Par ailleurs, les performances commerciales sont obtenues par la réalisation des opérations de ventes basées sur des ciblage affinés des clients. L'analyse de ces performances permet de générer des connaissances qu'il faut gérer et exploiter à leur tour. C'est un autre objectif que nous nous proposons d'étudier. Cette connaissance, qu'il faut réintroduire comme source d'informations dans notre dispositif d'entrepôt virtuel, permet alors de formaliser le savoir-faire des experts en ciblage et de le prendre en compte comme un élément important dans le processus d'aide à la décision.

La suite de ce papier est organisée de la façon suivante. Dans la section 2, nous présentons la base des demandes marketing. Nous exposons dans la section 3 notre approche d'entrepôt virtuel de données bancaires qui sera suivie par une discussion (section 4).

2 La base des demandes marketing

Le Crédit Lyonnais dispose de trois applicatifs qui permettent d'accéder à des données répondant à notre objet d'étude. Le premier applicatif est dédié au marketing avec les informations les plus utilisées sur les clients. Le second est un requêteur qui permet d'obtenir des informations plus détaillées sur ces derniers. Enfin, le troisième est un entrepôt de données orienté production (ventes réalisées). Il permet à l'aide de *Business Object* de visualiser les ventes réalisées selon les structures commerciales, les produits, les périodes d'analyse, etc.

Nous avons commencé par concevoir une base de données (cf modèle de la figure 1) pour stocker les différentes demandes marketing qui étaient traitées initialement de façon manuelle sous forme de fiches de liaison papier. Pour ce faire, deux aspects sont considérés : le versionnement et la généricité. La prise en compte des versions est nécessaire pour gérer l'évolution de la structure commerciale. Quant au caractère générique, il doit être introduit pour produire plusieurs sortes de demandes pouvant avoir des descripteurs spécifiques. Cette base de données sera alors considérée comme une source de données supplémentaire.

3 Notre approche

3.1 Intégration des données par médiation-entrepôt

Dans l'approche dite matérialisée, les données sont extraites de différentes sources et combinées pour être stockées de manière centralisée selon un schéma global dans une base cible. Ceci correspond à la mise en place d'une solution d'entrepôt de données [1, 2]. Cette centralisation physique des données permet à l'utilisateur d'avoir une vue globale des différentes sources. L'entrepôt constitue un support pour les applications décisionnelles. Une des difficultés de cette approche concerne le processus d'alimentation de l'entrepôt (d'ETL : Extracting, Transforming

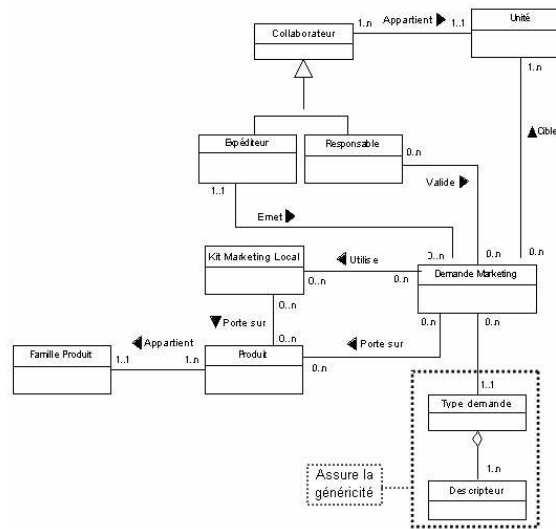


FIG. 1 – Modèle Conceptuel des Données pour la base des demandes marketing

and Loading) qui est déjà complexe et qui risque d'être pénalisé par les rafraîchissements fréquents des données.

Dans l'approche dite non matérialisée, une solution fréquente est la médiation [5]. Un système de médiation comprend un médiateur avec un schéma global et des adaptateurs (wrappers). C'est dans les termes de ce schéma global qu'est exprimée la requête utilisateur. Celle-ci est alors décomposée par le médiateur en sous-requêtes compréhensibles par les adaptateurs. Les différents adaptateurs sont chargés d'aller collecter les données dans les sources locales. Les réponses rendues par les différents adaptateurs sont ensuite combinées par le médiateur pour retourner les données à l'utilisateur. Ainsi, le problème de rafraîchissement des données ne se pose pas. Lorsqu'un utilisateur soumet une requête, la réponse est construite à partir des données sources qui constituent les données les plus récentes. L'approche médiateur a été utilisée, pour la recherche d'informations dans différents projets, tel que PICSEL [3] et autres. Mais elle n'a pas pour objectif l'analyse de celles-ci.

L'approche que nous préconisons permet de capitaliser les avantages de ces deux courants. Nous proposons donc une architecture d'entrepôt virtuel de données basée sur la médiation et l'entreposage. Cette architecture permet de construire des cubes de données à la volée à partir de données fournies par un système de médiation. Ce système d'aide à la décision présenté dans le schéma de la figure 2 permet la sélection et l'extraction des données ; l'intégration des données ; la construction de cube(s) de données représentant chacun un contexte d'analyse ciblé ; et enfin, l'analyse des performances d'une demande de marketing et l'extraction de connaissances qui sera réinjectée dans le dispositif pour les demandes à venir.

3.2 Mise en oeuvre

Pour mettre en oeuvre notre architecture, il est nécessaire d'avoir recours à un système coopératif afin que les différentes tâches s'effectuent de manière cohérente. La mise en place d'un Système Multi-Agents (SMA), désignant un ensemble plus ou moins étendu d'acteurs et communiquant entre eux [4], nous semble alors très pertinent. Le SMA doit permettre de placer l'utilisateur au cœur du processus. En effet, à partir d'une requête utilisateur, le système doit pouvoir l'analyser, sélectionner et interroger les sources de données originelles pertinentes à la construction

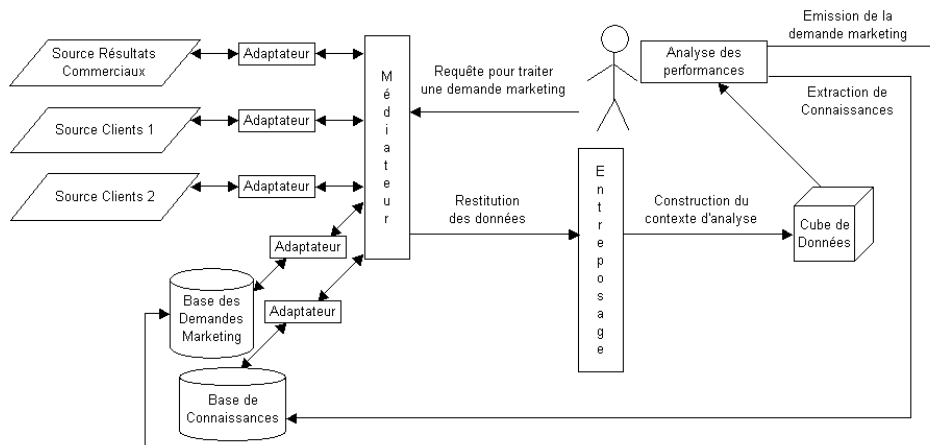


FIG. 2 – Architecture d’entrepotage virtuel

des réponses. Ces différentes tâches peuvent être assimilées à des services offerts par des acteurs spécifiques, agissant dans un système dont l’objectif final est la mise en oeuvre d’un processus décisionnel complet. La technologie "agent" permet une approche modulaire du problème, facilitant ainsi la maintenance du système (ajout de services ou d’agents par exemple). Dans notre cas, de nouvelles sources pourraient être considérées.

4 Discussion

Le principe de notre architecture est de créer un contexte d’analyse à partir des données retournées par le médiateur en réponse à une requête soumise par l’utilisateur. Ainsi il est nécessaire de se pencher sur le problème de l’approximation de requête afin d’obtenir toujours les données nécessaires et utiles à l’analyse. La construction d’un cube permet de définir un contexte d’analyse. Il s’agit donc de pouvoir aider l’utilisateur à déterminer en fonction des données renvoyées par le médiateur, les mesures et les dimensions du cube à construire. En outre, cette architecture se veut virtuelle dans la mesure où des cubes de données sont construits dynamiquement. Cependant, nous devons, dans un souci de performances, poser le problème de leur matérialisation. Le cube est construit à partir d’une requête. De ce fait, si celle-ci est récurrente, il peut s’avérer intéressant de matérialiser le contexte d’analyse. Par ailleurs, l’analyse des données doit permettre une extraction de connaissances. Il faut donc gérer ces connaissances, les représenter, les stocker afin de pouvoir les réinjecter dans le processus d’aide à la décision.

Références

- [1] W. H. INMON, *Building the data warehouse (2nd ed.)*, John Wiley & Sons, Inc., (1996).
- [2] R. KIMBALL, *The Data Warehouse toolkit*, John Wiley & Sons, (1996).
- [3] M.-C. ROUSSET, A. BIDAULT, C. FROIDEVAUX, H. GAGLIARDI, F. GOASDOUÉ, C. REYNAUD, AND B. SAFAR, (2002), *Construction de médiateurs pour intégrer des sources d’information multiples et hétérogènes : Picisel*, Journal I3 : Information - Interaction - Intelligence, 2, pp. 9–59.
- [4] K. SYCARA, K. DECKER, A. PANNU, M. WILLIAMSON, AND D. ZENG, December 1996, *Distributed intelligent agents*, IEEE Expert, 11, pp. 36–46.
- [5] G. WIEDERHOLD, *Mediators in the architecture of future information systems*, in Readings in Agents, M. N. Huhns and M. P. Singh, eds., Morgan Kaufmann, San Francisco, CA, USA, (1992), pp. 185–196.