

Archivage d'entrepôts de données multidimensionnelles

Faten Atigui, Franck Ravat, Olivier Teste, Gilles Zurfluh

IRIT (UMR 5505) Institut de Recherche en Informatique de Toulouse
118 route de Narbonne F-31062 Toulouse, France
{atigui, ravat, teste, zurfluh}@irit.fr

Résumé. Les données d'un entrepôt sont rafraîchies périodiquement et conservées de manière permanente. Cependant, les décideurs portent généralement un intérêt moindre pour les données anciennes. Dans cet article, nous proposons un mécanisme permettant de synthétiser les données les plus anciennes. Nous définissons un modèle conceptuel d'archivage de données multidimensionnelles. Nous présentons, ensuite, le modèle logique correspondant et les principes permettant d'interroger des schémas multidimensionnels archivés.

1 Introduction

Un entrepôt de données (ED) est une collection de données thématiques, intégrées, non volatiles et historisées pour des fins décisionnelles. Les données pertinentes pour la prise de décision sont collectées à partir des sources par le biais des processus d'Extraction-Transformation-Chargement. Dans un ED, les données extraites sont souvent structurées selon un format multidimensionnel (MD) qui organise l'information en termes de sujets d'analyse (faits) et d'axes d'analyse (dimensions) au sein d'un schéma en étoile Kimball (1996).

Dans un ED, les données sont conservées de manière permanente et sont rafraîchies de manière récurrente. De ce fait, l'ED possède de gros volume de données dans lequel le décideur risque de « se perdre » lors de ses analyses. De plus, les données historisées perdent de leur intérêt avec le temps : alors que la granularité des informations doit généralement être importante pour des données récentes ; elle peut être plus faible pour des données anciennes. Par exemple, un décideur peut analyser ses ventes au niveau de la granularité produit sur les cinq dernières années tandis que pour les périodes antérieures, ces analyses au niveau du produit seraient absurdes (les produits n'existent plus à l'heure actuelle) et donc le décideur fera des analyses au niveau de la gamme du produit (qui n'a pas évolué dans le temps). Afin de faciliter la tâche du décideur et de mieux répondre à ses besoins, il est préférable de garder uniquement l'information nécessaire à ses analyses. L'idée est donc d'offrir un environnement d'analyse MD adapté aux besoins des décideurs en leur permettant de supprimer dans le temps les niveaux de granularité inutiles pour leurs analyses.

Notre objectif est donc de proposer un modèle de données MD permettant de représenter l'archivage de données afin de ne conserver que les données nécessaires aux analyses décisionnelles. Ce mécanisme permet de garder une vue synthétique sur les données les moins récentes et dont les détails sont superflus pour la prise de décision. De plus, cette solution permettrait d'anticiper le problème de temps de réponse aux requêtes MD dès les premières phases

de modélisation MD. Dans cet article, après avoir effectué une analyse des travaux existants (section 2) nous proposons :

- Les modèles de représentation conceptuelle (sections 3 et 4) et logique (section 5) de données MD archivées ;
- Les principes d'interrogation d'un schéma MD archivé (section 5).

2 Etat de l'art

Dans cette section, nous exposons les travaux en relation avec notre contribution. D'une part, certains travaux ont présenté des techniques de synthèse de données. Dans ce contexte, nous pouvons souligner le travail de Ravat et al. (1999), dans lequel une extension du concept de classe d'objets a été proposée afin de spécifier les concepts de classe temporelle et de classe d'archives. Boly et al. (2007) proposent un ensemble de fonctions d'oubli permettant de garder une synthèse sur les données historisées dans les ED relationnelles. Skyt et al. (2008) proposent une technique de réduction de données dans les bases de données multidimensionnelles. D'autre part, dans le domaine des ED MD, les propositions relatives à la gestion des données temporelles et à l'évolution de schémas Favre et al. (2007) nous ont été utiles. L'étude de l'évolution dans les ED a fait l'objet de plusieurs travaux comme présentés dans Golfarelli et Rizzi (2009). Les approches existantes peuvent être classées en quatre types Wrembel (2009), à savoir, l'évolution de schémas et de données (Hurtado et al. (1999), Fan et Poulouvasilis (2004)), les extensions de versions (Ravat et al. (2006), Rizzi et Golfarelli (2007)), la simulation (Balmin et al. (2000), Bellahsene (1998)) et les extensions temporelles (Abelló et Martín (2003), Malinowski et Zimányi (2008)).

La gestion de l'évolution telle qu'elle est proposée dans la littérature ne répond pas de manière pertinente à nos besoins. En effet, les travaux précédents permettent de conserver les évolutions de données ou de schémas mais ils ne proposent pas de faire l'agrégation des données les plus anciennes dans le but de conserver uniquement les données utiles pour les prises de décision.

3 Modèle multidimensionnel avec archives : principes de base

L'archivage de données est un mécanisme qui vise à conserver uniquement l'information utile pour la prise de décision en utilisant les niveaux d'agrégation adéquats. Ce mécanisme permet donc de synthétiser les données les moins récentes. L'objet de cette section est de décrire les concepts de notre modèle MD avec archives.

3.1 Concepts

A l'instar des autres modèles MD Romero et Abelló (2009), notre modèle repose sur les concepts de faits et de dimensions. Les définitions suivantes explicitent ces concepts Ravat et al. (2008).

Définition 1 *Un schéma MD S est défini par $(F^S, D^S, Star^S)$ où :*

- $F^S \neq \emptyset$: est un ensemble non vide de faits ;

- $D^S \neq \emptyset$: est un ensemble non vide de dimensions ;
- $Star^S : F^S \mapsto 2^{D^S}$ associe chaque fait à un ensemble de dimensions.

Définition 2 Un fait noté $F_i \in F^S$ est défini par (N^{F_i}, M^{F_i}) où :

- N^{F_i} : est le nom du fait ;
- $M^{F_i} \neq \emptyset : \{(m_1^{F_i}, f_1), \dots, (m_w^{F_i}, f_w)\}$ est un ensemble non vide de couples de mesures $m_j^{F_i}$ associées à des fonctions d'agrégation f_k .

Définition 3 Une dimension notée $D_i \in D^S$ est définie par $(N^{D_i}, A^{D_i}, H^{D_i})$ où :

- N^{D_i} : est le nom de la dimension ;
- $A^{D_i} \neq \emptyset : \{a_1^{D_i}, \dots, a_u^{D_i}\}$ est un ensemble non vide d'attributs de dimension ;
- $H^{D_i} \neq \emptyset : \{H_1^{D_i}, \dots, H_w^{D_i}\}$ est un ensemble non vide de hiérarchies.

Définition 4 Une hiérarchie notée $H_j^{D_i} \in H^{D_i}$ est définie par $(N^{H_j^{D_i}}, P^{H_j^{D_i}}, AF^{H_j^{D_i}}, \{(p_u, af_v), \dots\})$ où :

- $N^{H_j^{D_i}}$: est le nom de la hiérarchie ;
- $P^{H_j^{D_i}} : \langle p_1, \dots, p_n \rangle \neq \emptyset$: est une liste non vide de paramètres (attribut identifiant un niveau de granularité d'analyse) ;
- $AF^{H_j^{D_i}}$: un ensemble d'attributs faibles permettant de compléter la sémantique des paramètres ;
- $\{(p_u, af_v), \dots\}$: permet l'association d'attributs faibles aux paramètres.

Ces concepts de base vont permettre de définir les éléments d'un schéma MD archivé. Ce dernier est composé d'un schéma courant et d'un ensemble d'archives. Le schéma courant représente le schéma indiquant le plus de niveaux de granularité d'analyse et stocke les données les plus récentes de l'entrepôt. Une archive contient les données synthétisées pour une période donnée. Celle-ci est toujours construite à partir d'un schéma précédent : la première archive est construite à partir du schéma courant, la seconde est construite à partir du schéma de la première archive, etc.

Les définitions suivantes permettent de spécifier les concepts de constellation et d'archive.

Définition 5 Une constellation archivée C est définie par (N^C, S_0^C, A^C) où :

- N^C : est le nom de la constellation ;
- S_0^C : est le schéma MD courant de la constellation ;
- $A^C : \langle A_1^C, \dots, A_n^C \rangle$: est une liste d'archives.

Définition 6 Une archive $A_i \in A^C$ est construite à partir d'un schéma précédent et définie par $(N^{A_i}, SO^{A_i}, S^{A_i}, T^{A_i})$ où :

- N^{A_i} : est le nom de l'archive ;
- SO^{A_i} que l'on notera SO est le schéma MD origine de l'archive défini par $(F^{SO}, D^{SO}, Star^{SO})$. Ce schéma est soit le schéma courant s'il s'agit de la première archive soit le schéma de l'archive précédente ;
- S^{A_i} : est le schéma MD de l'archive défini par $(F^{S^{A_i}}, D^{S^{A_i}}, Star^{S^{A_i}})$ où :
 - $F^{S^{A_i}}$: est l'ensemble de faits de l'archive issus du schéma précédent tel que : $F^{S^{A_i}} \subseteq F^{SO}$;
 - $D^{S^{A_i}}$: est l'ensemble de dimensions de l'archive issues du schéma précédent tel que : $D^{S^{A_i}} \subseteq D^{SO}$;

- $Star^{S^{A_i}} : F^{S^{A_i}} \mapsto 2^{D^{S^{A_i}}}$ associe chaque fait de l'archive à un ensemble de dimensions tel que : $\forall j, \exists k, F_j \in F^{S^{A_i}}, F_k \in F^{SO}, Star^{S^{A_i}}(F_j) \subseteq Star^{SO}(F_k)$;
- $T^{A_i} = [T_{Debut}, T_{Fin}]$: intervalle temporel durant lequel l'archive est valide. Le début d'une archive A_i représente la fin de l'archive A_{i+1} .

3.2 Application

Une entreprise souhaite faire l'analyse des montants et des quantités de ses ventes. Plus précisément, cette analyse s'effectue en fonction des clients (codeC, nom, prénom, ville, pays et continent), des produits (codeP, gamme et secteur) et du temps (date, mois et année). Le schéma courant qui répond à ce besoin est représenté dans la figure 1a. Au bout de 5 ans, ces niveaux de détails n'ont plus d'intérêt pour les analyses des décideurs. Dans un premier temps, le décideur souhaite conserver une synthèse sur les quantités et les montants des ventes en fonction des villes de clients et des gammes de produits. Une première archive est construite (figure 1b) pour la période $T_1 = [2005, 2007[$. Dans un second temps, le décideur souhaite garder les données les moins récentes de manière encore plus synthétisée. L'archive A_2 construite à partir de A_1 permet de stocker les quantités et les montants des ventes annuelles selon les gammes et les secteurs de produits pour la période de temps précédant 2005 (figure 1c).

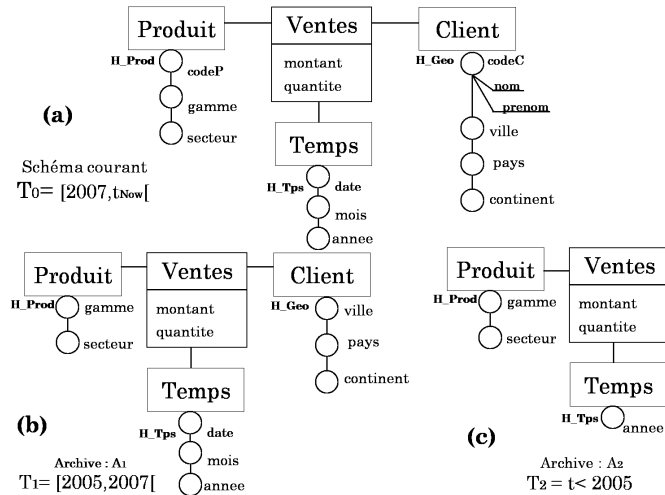


FIG. 1: Exemple d'archivage d'entrepôt de données multidimensionnelles

4 Schéma en constellation et hiérarchies multiples

4.1 Principes

Une constellation est un schéma MD qui présente plusieurs faits analysés en fonction de plusieurs dimensions dont certaines sont partagées. L'archivage d'une constellation à hiérarchies multiples partageant un ou plusieurs niveaux de granularité présente quelques particula-

rités. En effet, dans un schéma en constellation, les niveaux d'archivage peuvent être comme suit :

- Tous les faits liés à la dimension sont archivés par rapport à des niveaux de granularité identiques : la dimension est conservée avec les niveaux choisis,
- Les faits liés à la dimension sont archivés par rapport à des niveaux de granularité différents (éventuellement de hiérarchies différentes) : la dimension est divisée en un ensemble de dimensions simples où chacune est liée à son propre fait archivé par rapport à ses niveaux de granularité.

Afin de décrire les archives relatives à un schéma en constellation, nous proposons d'étendre le concept de dimension. Une dimension est soit une dimension simple, soit une dimension composée. Cette dernière regroupe plusieurs dimensions simples ayant les caractéristiques suivantes :

- Ces dimensions simples sont issues d'une même dimension dans un schéma d'origine ;
- Un même fait ne peut pas être associé à deux dimensions simples provenant de la même dimension composée.

Définition 7 : *Dimension*

Pour une archive $A_i \in A^C$ définie par $(N^{A_i}, SO^{A_i}, S^{A_i}, T^{A_i})$ où $S^{A_i} = (F^{S^{A_i}}, D^{S^{A_i}}, Star^{S^{A_i}})$
 $\forall D_j^{S^{A_i}} \in D^{S^{A_i}} = \{D_1, \dots, D_x\}$ un ensemble de dimensions simples, $D^{S^{A_i}}$ issue de $D_k^{SO} \forall j \in [1, x]$, D_j est définie comme suit :

- Les noms des dimensions simples correspondent à la concaténation du nom de la dimension composée et d'un compteur allant de 1 à x .
- $A^{D_j} \in A^{D_k^{SO}}$: l'ensemble d'attributs de la dimension est issu de l'ensemble des attributs de la dimension correspondante dans le schéma précédent ;
- $H^{D_j} \in H^{D_k^{SO}}$: l'ensemble de hiérarchie appartenant à la dimension est issu des hiérarchies de la dimension correspondante dans le schéma précédent.

Cette définition induit deux contraintes :

Soit D_x et D_y deux dimensions simples appartenant à la même dimension composée D :

- C1. Chaque dimension simple est associée à un fait qui ne peut pas être lié à une autre dimension simple issue de la même dimension composée :
 $Star^{-1}(D_x) \not\subseteq Star^{-1}(D_y)$;
- C2. La liste des attributs d'une dimension simple est différente de celle des attributs d'une autre dimension simple appartenant à la même dimension composée :
 $\exists i, A_i \in A^{D_x} \Rightarrow A_i \notin A^{D_y}$.

Comme pour le concept de dimension, le formalisme graphique a été étendu, le schéma MD visualise les données à l'instant t et les données qui vont servir à l'archive suivante. La figure 3 (a) montre le schéma de l'archive A_1 . Les ronds colorés représentent les niveaux de granularité du schéma de l'archive suivante (A_2). Les ronds colorés en noir servent à archiver tous les faits liés à la dimension. Quand les ronds sont incomplets, ceci montre que la dimension est partagée et que les faits auxquels elle est reliée sont archivés par rapport à des niveaux différents. Les ronds colorés en blanc vont disparaître dans le schéma suivant (cf. figure 3 (a) et (b)). Quand une dimension est composée, l'ensemble de ses dimensions simples est regroupé dans un rectangle discontinu (cf. figure 3 (b)).

4.2 Application

La figure 2 présente un schéma en constellation permettant l'analyse des montants et des quantités de ventes ainsi que les prix d'achat. L'analyse des ventes se fait par rapport aux dimensions « Produits », « Clients » et « Temps ». L'analyse des achats se fait uniquement par rapport aux dimensions Produits et Temps. Ce schéma présente le schéma courant avec les niveaux d'analyse les plus détaillés. La figure 3 (a) montre l'archive A_1 construite à partir du schéma courant et valide durant $[2005, 2007[$. Cette archive permet l'analyse des ventes et des achats journaliers et annuels selon les gammes et les secteurs de produits. Le décideur propose de garder les ventes selon les pays et les zones des clients, non plus leurs villes. Etant donné que ces deux niveaux appartiennent à deux hiérarchies différentes, le niveau d'archivage proposé et celui de granularité plus faible et commun aux deux hiérarchies ; il s'agit du niveau ville. Enfin, la figure 3 (b) présente l'archive A_2 où les ventes et les achats sont analysés en fonction des gammes et des secteurs de produit. Cependant, l'analyse des ventes se fait de manière annuelle alors que l'analyse des achats se fait de manière journalière. La dimension Temps est divisée en « Temps_1 » et « Temps_2 ».

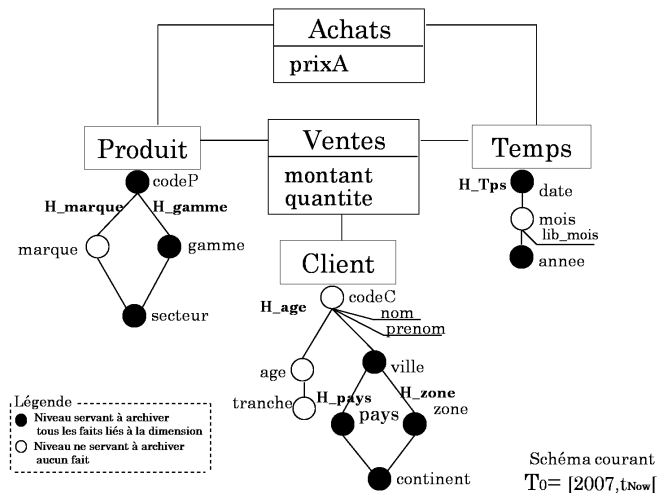


FIG. 2: Archivage de schéma en constellation à hiérarchies multiples : schéma courant

5 Principes d'interrogation

Dans cette section, nous définissons les principes d'interrogation de schémas MD archivés. Pour ce faire, nous présentons tout d'abord le modèle logique qui correspond aux différentes archives au niveau relationnel. Nous proposons d'implanter le schéma conceptuel au niveau physique en utilisant les vues matérialisées Oracle¹. L'utilisation des vues matérialisées est avantageuse puisque le calcul, le stockage, la mise à jour et le rafraîchissement sont effectués automatiquement par le SGBD. Ceci permet de garantir la cohérence des données du schéma

1. <http://www.oracle.com/index.html>

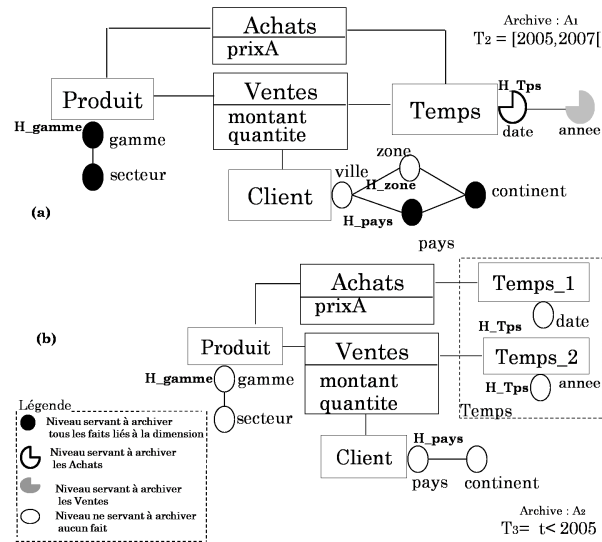


FIG. 3: Archivage de schéma en constellation à hiérarchies multiples : $A_1(a)$, $A_2(b)$

courant et des différentes archives. Ensuite, nous présentons les principes d'interrogation d'un schéma archivé.

5.1 Modèle logique

Dans le cadre de cet article, nous avons décidé de traduire les schémas courants et les archives en R-OLAP² dénormalisé. Plus de détails sur la démarche de transformation ont été présentés dans Atigui et al. (2010) et Atigui et al. (2012).

La définition d'un schéma R-OLAP dénormalisé s'effectue comme suit : chaque dimension est convertie en une relation et sa clé primaire correspond au paramètre racine de la dimension. De la même manière, chaque fait est converti en une relation où la clé primaire correspond à la concaténation des différentes clés étrangères qui référencent les dimensions liées à ce fait. En ce qui concerne les archives :

- Seules les dimensions qui ont changé de schéma (un ou plusieurs niveaux de granularité ont été supprimés) sont transformées en relation. Une dimension composée est transformée en un ensemble de relations (une pour chaque dimension simple) ;
- Tout fait est transformé en une relation de la même manière que dans un schéma courant.

En plus de ces schémas, un modèle MD archivé nécessite le stockage d'un ensemble de méta-données. Ces dernières permettent de stocker les informations relatives aux structures, aux contenus et aux temps de validité des différentes archives. Ces informations fournissent un support pour la gestion des archives ainsi que pour l'interrogation.

Exemple 1 L'exemple ci-dessous montre différentes relations R-OLAP associées aux faits et dimensions du schéma MD archivé présentés dans les figures 2, 3 (a), et 3 (b). Les dimensions et les faits du schéma courant correspondent aux relations suivantes :

$SC.Produit(\text{code}_P, \text{gamme}, \text{marque}, \text{secteur})$;

2. ROLAP : Relational On Line Analytical Processing

SC.Client(codeC, nom, prenom, age, tranche, ville, pays, zone, continent);

SC.Temps(date, mois, lib_mois, annee);

SC.Ventes(codeP#, codeC#, date#, montant, quantite);

SC.Achats(codeP#, date#, prixA)

L'archive A_1 est dérivée à partir du schéma courant, nécessite la création de nouvelles relations *Produit* et *Client* (changement de schéma) et des relations de faits :

A_1 .*Produit*(gamme, secteur); A_1 .*Client*(ville, pays, zone, continent)

A_1 .*Ventes*(gamme#, ville#, SC.Temps.date#, montant, quantite);

A_1 .*Achats*(gamme#, SC.Temps.date#, prixA)

L'archive A_2 contient les relations de la dimension composée *Temps*, de la dimension *Client* et les relations de fait :

A_2 .*Client*(pays, continent); A_2 .*Temps_1*(date); A_2 .*Temps_2*(annee)

A_2 .*Ventes*(A_1 .Produit.gamme#, pays#, annee#, montant, quantite)

A_2 .*Achats*(A_1 .Produit.gamme#, date#, prixA)

5.2 Principes d'interrogation

Dans un entrepôt de données archivées, les données qui intéressent l'utilisateur peuvent être réparties entre le schéma courant et les différentes archives. Les requêtes d'interrogation peuvent porter sur des schémas différents. L'indication de la durée de validité des données à interroger permet de préciser l'ensemble des archives à utiliser. L'exécution d'une requête R qui porte sur plusieurs schémas se fait suivant plusieurs étapes :

1. Déterminer les différents schémas à interroger $S = \{S_1, S_2, \dots, S_n\}$ en utilisant les durées de validité spécifiées par l'utilisateur, si aucune durée n'est spécifiée, tous les schémas sont interrogés;
2. Décomposer la requête en un ensemble de sous-requêtes adéquates pour interroger un schéma S_i de S ;
3. Exécuter chaque requête R^{S_i} ;
4. Restituer les résultats retournés par les différentes requêtes R^{S_i} de manière séparée, ainsi qu'un résultat global permettant de répondre à la requête de départ R par fusion et synthèse des résultats des sous-requêtes R^{S_i} .

La figure 4 montre le processus d'interrogation. Pour l'implantation de ce processus, nous proposons d'utiliser une extension des principes d'interrogation des bases de données réparties Özsu et Valduriez (1999).

6 Conclusion

Dans cet article, nous avons présenté une approche pour la synthétisation de données MD dans un entrepôt. Le modèle conceptuel proposé comprend un schéma courant et un ensemble de schémas archivés. Pour les schémas en constellation, le concept de dimension a été étendu afin d'archiver les dimensions partagées par plusieurs faits. Ensuite, nous avons présenté le modèle logique ROLAP dénormalisé ainsi que les principes d'interrogation d'un schéma archivé.

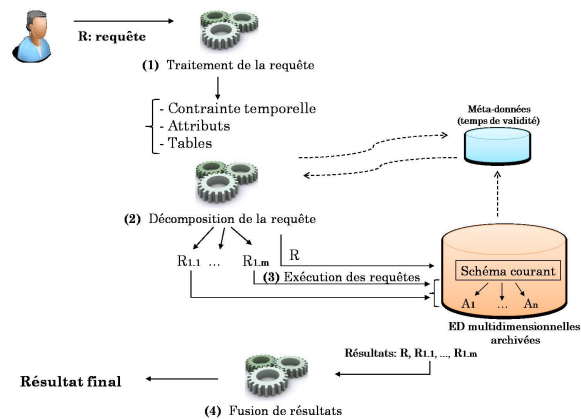


FIG. 4: Principes d'interrogation d'ED multidimensionnelles archivées

Dans de futurs travaux, nous envisageons dans un premier temps d'automatiser le processus d'archivage en intégrant une démarche dirigée par les modèles. Ceci nécessite le développement d'un ensemble de méta-modèles et des règles de transformation entre modèles qui permettent de générer le code final de manière automatique Atigui et al. (2012). Dans un second temps, nous souhaitons proposer différents types de modélisation logique et analyser les principes d'interrogation d'entrepôts de données MD archivées qui leur sont associés. Enfin, nous envisageons d'appliquer l'approche sur des cas d'études du monde réel et de la faire évaluer par différents utilisateurs.

Références

- Abelló, A. et C. Martín (2003). A bitemporal storage structure for a corporate data warehouse. In *ICEIS (1)*, pp. 177–183.
- Atigui, F., F. Ravat, O. Teste, et G. Zurfluh (2010). Démarche dirigée par les modèles pour la conception d'entrepôts de données multidimensionnelles. In *Journées Bases de Données Avancées (BDA), Toulouse, 18/10/2010-22/10/2010*.
- Atigui, F., F. Ravat, O. Teste, et G. Zurfluh (2012). Modélisation conjointe des données et des processus pour l'implantation de schémas d'entrepôts. *Journal of Decision Systems*. (In press).
- Balmin, A., T. Papadimitriou, et Y. Papakonstantinou (2000). Hypothetical queries in an olap environment. In *VLDB*, pp. 220–231.
- Bellahsene, Z. (1998). View adaptation in data warehousing systems. In G. Quirchmayr, E. Schweighofer, et T. J. M. Bench-Capon (Eds.), *DEXA*, Volume 1460 of *Lecture Notes in Computer Science*, pp. 300–309. Springer.
- Boly, A., S. Goutier, et G. Hébrail (2007). Des fonctions d'oubli intelligentes dans les entrepôts de données. In *EGC*, pp. 223–234.
- Fan, H. et A. Poulouvasilis (2004). Schema evolution in data warehousing environments - a schema transformation-based approach. In *ER*, pp. 639–653.

- Favre, C., F. Bentayeb, et O. Boussaïd (2007). Evolution de modèle dans les entrepôts de données : existant et perspectives. In *3èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 07)*, Poitiers, Volume B-3 of *Revue des Nouvelles Technologies de l'Information*, Toulouse, pp. 21–36.
- Golfarelli, M. et S. Rizzi (2009). A survey on temporal data warehousing. *IJDWM* 5(1), 1–17.
- Hurtado, C. A., A. O. Mendelzon, et A. A. Vaisman (1999). Maintaining data cubes under dimension updates. In *ICDE*, pp. 346–355.
- Kimball, R. (1996). *The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses*. John Wiley.
- Malinowski, E. et E. Zimányi (2008). A conceptual model for temporal data warehouses and its transformation to the er and the object-relational models. *Data Knowl. Eng.* 64(1), 101–133.
- Özsu, M. T. et P. Valduriez (1999). *Principles of distributed database systems (2nd ed.)*. Upper Saddle River, NJ, USA : Prentice-Hall, Inc.
- Ravat, F., O. Teste, R. Tournier, et G. Zurfluh (2008). Algebraic and graphic languages for olap manipulations. *IJDWM* 4(1), 17–46.
- Ravat, F., O. Teste, et G. Zurfluh (1999). Towards Data Warehouse Design . In *Proceedings of the 8th International Conference on Information and Knowledge Management - ACM CIKM'99 , Kansas City, 02/11/99-06/11/99*, pp. 359–366. ACM-Press Susan Gauch.
- Ravat, F., O. Teste, et G. Zurfluh (2006). A multiversion-based multidimensional model. In *DaWaK*, pp. 65–74.
- Rizzi, S. et M. Golfarelli (2007). X-time : Schema versioning and cross-version querying in data warehouses. In *ICDE*, pp. 1471–1472.
- Romero, O. et A. Abelló (2009). A survey of multidimensional modeling methodologies. *IJDWM* 5(2), 1–23.
- Skyt, J., C. S. Jensen, et T. B. Pedersen (2008). Specification-based data reduction in dimensional data warehouses. *Inf. Syst.* 33(1), 36–63.
- Wrembel, R. (2009). A survey of managing the evolution of data warehouses. *IJDWM* 5(2), 24–56.

Summary

Data integrated in a datawarehouse is periodically updated. Nevertheless, decision-makers are usually less interested in old information. In this paper, we present a new mechanism for summarizing data across the time. In particular, we propose a conceptual model that represents multidimensional data archiving. Moreover, we propose a logical model and principles for querying multidimensional archived schema.