Contrôle d'accès aux entrepôts de données fondé sur le profil utilisateur

Amina El, Nouria Harbi, Hassan Badir

Contrôle d'accès aux entrepôts de données fondé sur le profil utilisateur

Amina El ouazzani*
Nouria Harbi **, Hassan Badir*

* Université Abdelmalek Essaadi ENSA LabTIC 1818 Tanger MAROC
{ a.elouazzani2000,h.badir}@gmail.com,
http://www.ensat.ac.ma/
**Université Lumière Lyon 2 Laboratoire ERIC 69635 Lyon, Cedex FRANCE
nouria.harbi@univ-lyon2.fr

Résumé. Un entrepôt de données ou Data Warehouse (DW) peut être utilisé comme un mécanisme très puissant pour découvrir les informations cruciales de l'entreprise. Il est donc important d'appliquer des mesures de sécurité qui garantissent la confidentialité des données qu'il contient. Dans ce sens, plusieurs propositions ont été présentées, néanmoins, aucune n'est considérée comme un standard dans la gestion des accès aux entrepôts de données. Dans cet article, nous allons d'abord analyser le problème de sécurité au sein d'un entrepôt et ses enjeux, ainsi, nous présenterons une architecture d'un entrepôt sécurisé qui consiste à gérer les accès en fonction des profils des utilisateurs. Cette solution a l'avantage de fonctionner d'une manière indépendante de la plate-forme cible.

1 Introduction

Les entrepôts de données occupent une place importante dans les organisations, ils sont considérés comme étant le système de soutien et l'élément clé dans les processus de prise de décisions stratégiques. Cependant, les infractions en matière de sécurité et de confidentialité des entrepôts continuent à poser une menace en raison de la grande sensibilité de l'information qui peut y être découverte, tel que des informations sur la vie privée des individus.
Dans ce cadre, plusieurs gouvernements ont promulgué des lois pour la protection des informations sur la vie privée de leurs citoyens. Parmi ces lois, HIPAA (Health Insurance Portability and Accountability Act HHS (1996)) vise à protéger les données médicales des patients américains en obligeant les établissements du secteur des soins de la santé de suivre des règles strictes de sécurité, de même GLBA (Gramm Leach Bliley Act GPO 1999) oblige les organisations financières américaines à protéger les données de leurs clients. Par conséquent, étant donné que la sécurité des entrepôts de données présente une préoccupation urgente dans la plupart des entreprises compte tenu de l'importance de l'information dont ces entrepôts regorgent, il est essentiel de définir des mesures de confidentialité.
Dans ce papier, nous nous concentrons sur la gestion des accès à l'entrepôts de données en proposant une approche de sécurité basée sur le profil de l'utilisateur qui décrit les droits d'accès à ce dernier, en utilisant la politique de contrôle d'accès RBAC (Role based Access Control)

qui se focalise sur le regroupement des utilisateurs selon leurs métiers. Avec cette approche, nous sommes en mesure de restreindre davantage l'accès des utilisateurs aux données qui leur sont interdites.
Après la présentation de la problématique dans la section 1, le reste de cet article est structuré comme suit. La section 2 présente une vue d'ensemble des travaux connexes. La Section 3 décrit notre proposition à savoir une architecture d'entrepôt sécurisé dont l'accès est basé sur les profils des utilisateurs. Enfin, la section 4 présente nos conclusions et perspectives.

2 Etat de l'art des travaux existants

Récemment, un certain nombre de solutions de sécurité des entrepôts de données ont été proposés, nous avons organisés les travaux selon deux approches, la première est celle de l'intégration de la sécurité dans le processus de la modélisation des entrepôts (niveau conception), et la deuxième présente les modèles de contrôle d'accès pour un entrepôt déjà mise en place (niveau exploitation).

2.1 La sécurité dans la modélisation des entrepôts (niveau conceptuel)

Parmi les travaux qui ont été développés sur l'intégration de la sécurité dans la modélisation des entrepôts, on trouve :

Fernandez-Medina et al. (2006) ont développé un modèle de contrôle d'accès et d'audit (ACA) spécifique aux entrepôts de données, qui repose sur deux politiques de gestion des accès : MAC et RBAC. en intégrant la notion de « profil utilisateur ». Ce modèle reste un modèle purement théorique car aucune solution concernant son implémentation n'a encore été proposée.

Soler et al. (2008) ont utilisé des mécanismes d'extension fournis par le CWM (Common Warehouse Metamodel) pour étendre le package relationnel et construire un schéma en étoile, qui représente les règles de sécurité et de vérification.

Soler et al. (2009) ont développé une méthodologie comprenant quatre phases : analyse, modélisation, implémentation et validation, qui couvrent les cinq niveaux d'abstraction qui sont : analyse des besoins, niveau conceptuel, niveau logique, niveau physique et l'examen post-développement, ce dernier étant une nouvelle discipline introduite par Lujan et Trujillo (2004).

Rodriguez et al. (2011) présentent une extension d'UML 2.0 du diagramme d'activité. Cette proposition, libellée comme BPSec (Business Security Process), permet de définir un ensemble d'exigences de sécurité (contrôle d'accès, détection des risques d'attaques, non-répudiation, intégrité, confidentialité et vérification de la sécurité).

2.2 L'intégration de la sécurité dans les entrepôts existants (niveau exploitation)

Le serveur OLAP (Online Analytical Processing) est censé d'assurer des accès à l'entrepôt de données en fonction des habilitations de chaque utilisateur. Cependant, le serveur OLAP tout seul ne peut pas protéger l'accès aux données interdites. Des travaux ont été réalisés pour renforcer les droits d'accès/habilitations, pour interdire tout utilisateur malicieux d'inférer des données qui lui sont interdites.

Kirkgoze et al. (1997) ont défini un modèle qui consiste à élaborer un cube personnalisé possédant ses propres dimensions et hiérarchies. Ce modèle repose sur la politique de gestion AMAC. Il s'agit d'une extension du modèle MAC qui permet de spécifier les tâches que l'utilisateur peut exécuter selon son rôle au sein de l'organisation.

Priebe et Pernul (2001) ils ont créé un mécanisme de contrôle d'accès sous forme d'un langage exprimant, il s'agit d'un langage basé sur MDX « MulDimensionnale Xpression », celuici étant un langage de requête spécialisé dans l'interrogation et la manipulation des données multidimensionnelles.

Triki et al. (2011) ont proposé une approche basée sur les réseaux Bayésiens. Pour protéger un entrepôt de données contre les inférences, elle vise à interdire à un utilisateur d'inférer des données protégées à partir des données qui lui sont accessibles en utilisant les fonctions d'agrégations Min et Max.

Eavis et Althamimi (2012) ont présenté une approche intuitive et puissante pour l'authentification de base de données qui est uniquement adaptée au domaine OLAP. Il est orienté objet et utilise des règles de réécriture de requêtes afin d'assurer l'accès aux données cohérentes.

2.3 Synthèse des travaux existants

En générale, les mesures de sécurité pour les entrepôts sont définies dans la mise en œuvre finale au-dessus des systèmes commerciaux, car il n'y a pas une norme pour l'échange et l'interopérabilité des métadonnées. Bien que la proposition du méta modèle CWM (Common Warehouse Metamodel) basé sur trois standards, à savoir UML, MOF et XML est largement acceptée comme la norme pour l'échange et l'interopérabilité des métadonnées.
Les travaux présentés sont cohérents et complémentaires. D'abord E Soler et al (2006) ont étendu le méta modèle CWM pour représenter correctement toutes les règles de sécurité et d'audit définies dans la modélisation conceptuelle des entrepôts de données Dans la suite de leurs travaux, Soler et al. (2008) et Soler et al. (2009) se sont basés sur le modèle MDA pour définir des règles de passage formelles entre le modèle conceptuel de l'entrepôt de données et le modèle logique, en exploitant le query/view/transformations (QVT) proposé par le modèle MDA. Soler et al. (2009) ont fait usage des mécanismes d'extension fournis par le CWM pour étendre le paquet relationnel afin de construire un schéma en étoile, qui représente les règles de sécurité et de vérification capturées pendant la phase de modélisation conceptuelle de l'entrepôt.

Concernant les politiques d'accès, peu d'auteurs fondent leurs travaux sur la politique DAC, bien que celle-ci soit intéressante, l'application d'une politique DAC implique un faible contrôle du flot d'information.

3 Proposition d'une approche de sécurité des entrepôts

3.1 Profil utilisateur

Le profil de l'utilisateur est une représentation des préférences et les droits d'accès d'un utilisateur individuel. Il contient les informations nécessaires pour l'authentification, et le niveau de sécurité pour accéder aux données de l'entrepôt, alors que l'entrepôt peut utiliser les détails de catégorie des données pour déterminer le contrôle d'accès.
Le modèle de contrôle d'accès à base de rôles adopté est le RBAC étendu. Il s'agit d'une politique d'accès répondant mieux aux besoins des grandes entreprises, gérant beaucoup de permissions pour un plus grand nombre d'utilisateurs. En effet, il est possible, pour certaines personnes, de cumuler plusieurs fonctions et par conséquent, plusieurs rôles. L'objectif est alors d'associer des profils aux profils, voire des rôles au rôle concerné. Il se crée une hiérarchie entre les rôles et donc une relation d'héritage. En effet, le rôle descendant hérite des permissions et restrictions du rôle ascendant.

3.2 Architecture d'entrepôt sécurisé basée sur la gestion des accès à base de profil

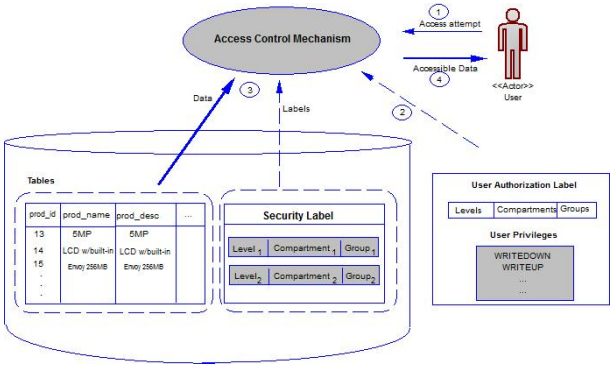

FIG. 1 – *scénario de gestion des accès à base de profil utilisateur.*

Dans l'architecture proposée, quelque soit la façon dont les utilisateurs se connectent à la table protégée (via une application, une interface Web ou SQL * Plus), le résultat est le même. Il n'y a pas de problème de sécurité des applications, puisque la politique d'accès est fixée à la table, et ne peut pas être contournée.

Pour atteindre les objectifs explicités dans l'architecture, on doit fixer les droits d'accès des utilisateurs de l'entrepôt. Comme dans les systèmes d'information, cette tâche de sécurité peut être traitée au niveau conceptuel ou au niveau logique, et les droits d'accès fixés sont appliqués par le serveur OLAP.

3.3 Le modèle d'authentification proposé

Le modèle conceptuel d'authentification proposé est chargé de vérifier les informations de l'identification de l'utilisateur. Il est constitué d'un ensemble des tables représentants les métadonnées nécessaires pour authentifier et autoriser l'utilisateur.
Par exemple, la table des utilisateurs stocke les informations de l'identification de base de l'utilisateur (login, mot de passe), tandis que la table des autorisations enregistre le fait qu'un utilisateur donné peut ou non accéder à certaines informations, tout en gardant la trace de chaque transaction effectuée par un utilisateur authentifié, durant une période autorisée indiquée dans la table des permissions d'accès.
La figure 2 illustre une version légèrement simplifiée du schéma d'autorisation de la base de données autorisation :

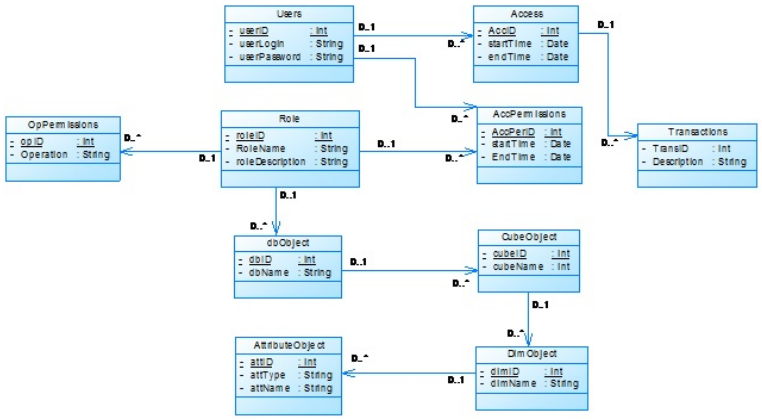

FIG. 2 – *Le modèle conceptuel d'authentification proposé.*

4 Conclusion et perspective

Dans cet article, nous avons essayés de présenter l'important rôle que jouent les entrepôts de données dans une organisation pour les prises de décisions stratégiques. Vue la sensibilité de leurs contenus, leur sécurité est une nécessité pour les entreprises. C'est pour cela que nous avons tenté de cerner les problèmes liés à la sécurité des entrepôts au niveau conceptuel et au niveau exploitation (OLAP) à travers une étude des travaux effectués dans la sécurisation des accès aux entrepôts de données.
En se basant sur ces travaux de recherche, nous avons proposé une architecture simplifiée et

fonctionnelle pour le contrôle d'accès aux entrepôts de données en adaptant la politique profil d'utilisateurs afin de garantir la confidentialité des données.
Notre travail est orienté vers l'avenir, car nous avons l'intention d'améliorer la solution proposée dans ce papier afin d'avoir une solution hybride combinant la sécurité liés aux Bases de données et surtout aux entrepôts de données et celles liées à l'environnement cloud computing.

Références

Eavis, T. et A. Althamimi (2012). Olap authentication and authorization via queryre-writing. *The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications*, 130–139.

Fernandez-Medina, E., J. Trujillo, R. Villarroel, et M. Piattini (2006). Access control and audit model for the multidimensional modeling of dws. *Decision Support Systems*, 1270–1289.

Kirkgoze, R., N. Katic, M. Stolba, et A. Tjoa (1997). A security concept for olap. *Proceedings of the 8th International Workshop on Database and Expert System Applications (DEXA'97)*, 619–626.

Lujan, S. et J. Trujillo (2004). A data warehouse engineering process. *Proceedings of the 3rd International Conference on Advances in Information Systems (ADVIS'04)*, 20–22.

Priebe, T. et G. Pernul (2001). A pragmatic approach to conceptual modeling of olap security. *Proceedings of the 20th International Conference on Conceptual Modeling (ER'01) 2224*, 311–324.

Rodriguez, A., E. Fernandez-Medina, J. Trujillo, et M. Piattini (2011). Secure business process model specification through a uml 2.0 activity diagram profile.

Soler, E., J. Trujillo, C. Blanco, et E. Fernandez-Medina (2009). Designing secure data warehouse by using mda and qvt. *Journal of Universal Computer Science 8 15*, 1607–1641.

Soler, E., J. Trujillo, E. Fernandez-Medina, et M. Piattini (2008). Building a secure star schema in data warehouses by an extension of the relational package from cwm. *Computer Standards and Interfaces 30*, 341–350.

Triki, S., H. Ben-Abdallah, N. Harbi, et O. Boussaid (2011). Securing data warehouses: A semi-automatic approach for inference prevention at the design level. *1st International Conference on Model and Data Engineering, Lecture Notes in Computer Science (LNCS) by Springer-Verlag*, 311–324.

Summary

A data warehouse (DW) is a powerful mechanism that can be used to discover the critical business information, so it's important to specify security measures to ensure the confidentiality of data. In this sense, many proposals has been presented, however, any one is considered a standard in the management of access to data warehouse, in this article, we will firstly discuss confidentiality problems in data warehouse, and we present our secure architecture to manage access based on the user profile. This solution has the advantage of operating independently of the target platform , on any secure management system database.