HAL
open science

User Profile Management to protect sensitive data in Warehouses

Amina Elouazzani, Nouria Harbi, Hassan Badir

To cite this version:

Amina Elouazzani, Nouria Harbi, Hassan Badir. User Profile Management to protect sensitive data in Warehouses. International Journal of Next-Generation Computing, 2018. hal-02054436

HAL Id: hal-02054436

https://hal.science/hal-02054436v1

Submitted on 12 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

User Profile Management to protect sensitive data in Warehouses

Amina ELOUAZZANI*, Nouria HARBI** and Hassan BADIR*
* LabTIC Laboratory, National School of Applied Sciences Tangier,Morocco
**Eric Laboratory, Lyon II University, France

A data warehouse presents a rich source of information on the activities of the company and the privacy of individuals. So this source can be used as a very powerful mechanism for discovering the crucial information of company. Hence the importance of implementing security measures which guarantee the data confidentiality by establishing an access control policy. In this direction, several propositions were made, but none are considered as a standard for access management to data warehouses. In this article, we will present our approach that allows first to exploit the permissions defined in the data sources in order to help the administrator to define access permissions to the data warehouse, and then our system will automatically generate the sensitivity level of each data warehouse element according to the permissions granted to an object in the data warehouse. This makes it possible to specify sensitive data in order to protect them against illegal access and to detect inferences.

Keywords: Data Warehouse, Confidentiality, Business profile, Crucial information, Traceability.

1. INTRODUCTION

A data warehouse DW is a critical factor for high direction looking to make the right strategic decision. It is a database of terabytes data stored historically, from operational systems to have a clear view and a rich source for decision makers. It presents a source of critical business data and the data of customers privacy such as medical and financial data protected by laws. Consequently, they should not be accessible without access control.

Because of the critical information stored in the DW, it is important to check its privacy. According to DEVBANDU and STUBBLEBINE [2000] , confidentiality in the context of DW is considered an important requirement, which must be ensured by an authorization management mechanism which is the specification and execution of access rights in the databases in general and more specifically in DWs.

In general, conventional security requirements are summarized by the acronym CIA (confidentiality, integrity and avail-ability). All other security requirements such as authentication, authorization, access control, etc., can be assigned to these three basic properties. Confidentiality is defined as the absence of disclosure of unauthorized information. Integrity is defined as the absence of the unauthorized modification of information, and availability that ensures the continuity of service LANDWEHR [2001].

Our goal is to ensure the confidentiality of the DW by access control, since there is no standard that manages this important aspect, and mechanisms of confidentiality specified for OLTP systems cannot be used for the DW because in operational systems, access control is defined on the tables, rows, columns, etc. While in an DW, we have a large number of users with different analysis needs, seeking access to the multidimensional DW E.FERNANDEZ-MEDINA et al. [2006], J.TRUJILLO et al. [2009].

In this paper, we focus on the conceptual modeling process of the DW by offering an approach providing access control, based on the job profile of a user who de-scribes its access rights. We use the RBAC access control policy (Role-based Access Control) that focuses on grouping users according to their professions or their roles. With this approach, we are able to classify the DW automatically by generating their sensitivity levels, in order to trace user actions on sensitive

data.

2. STATE OF THE ART AND SYNTHESIS

Recently, a number of DW security models have been proposed. In this section we will organize these research works according to two approaches, the integration of security in the modeling process of the DW, and the DW access control models already in place.

2.1 Security at the conceptual level of the DW

2.1.1 *The DWs security based on permissions defined at the sources*

—**ROSENTHAL and SCIORE [2000]**propose a theoretical approach that exploits the access permissions defined in the data source, rather than creating new access mechanisms. They use the rewrite queries to verify that they comply with the restrictions defined in the data sources, as well as creating relational views in order to minimize the risk to infer sensitive data.

—**SALTOR et al. [2002]** since there are similarities be-tween the architecture of a federated database and the architecture of a DW, the authors proposed the use of the multi-level access permissions pattern defined for the federated database without being modified to construct a secure DW, this authorization scheme describes the multi-level access rules.

2.1.2 *Confidentiality of the data in a DW during the modelization.* Among the works that have been developed on integrating security into the modelization of warehouses, we find:

—**E.FERNANDEZ-MEDINA et al. [2006]** have developed an access control and auditing model (ACA) specific for DW, based on two access management policies: MAC and RBAC. They specify the security rules during the modelization process of a conceptual model, by incorporating the concept ofuser profile, which consists of an isolated table containing all user information (identity, classification level: top secret, secret, confidential or unknown). This model remains a purely theoretical model since no solution for its implementation has been proposed.

—**R.VILLARROEL et al. [2006]** have defined an OCL ex-tension Object Constraint Language using UML2.0 extension mechanisms to resolve issues of confidentiality; this extension specifies the security requirements of the elements during the conceptual modeling of DWs.

—**SOLER et al. [2008]** have used extension mechanisms provided by the CWM (Common Warehouse Meta-model) to extend the relational package and build a star schema, which represents security and verification rules captured during the conceptual phase of the DW.

—**J.TRUJILLO et al. [2009]** have developed a methodology consisting of four phases: analysis, modeling, implementation and validation, which covers the five levels of abstraction: requirements analysis, conceptual level, logic level, the physical level and the post-development review, the latter being a new discipline introduced by Lujan and Trujillo (2004). This methodology offers all the security requirements throughout the life cycle of the DW.

—**RODRIGUEZ et al. [2011a]** presented an UML 2.0 extension of the activity diagram. This proposition, called BPSec (Business Security Process), allows to define a set of security requirements (access control, detection of attack risks, non-repudiation, integrity, confidentiality and security audit), which improves the expressiveness of the business processes models, and enables to secure a DW during its development, taking into account this requirement.

—**BLANCO et al. [2015]** have developed an automatic MDA architecture to secure DW; this architecture is composed of a logic model and its transformations from the conceptual model using the UML extension and the CWM package. They defined these constraints in the metadata layer that connects the DW with the OLAP tools. This proposition consists of models and transformations.

2.1.3 *Inferences Management*

—**S.TRIKI et al. [2013]**proposed a model to secure multidimensional data against the inferences in the conceptual phase, this approach assumes that the DW scheme is already designed. It allows to detect both types of inferences:

✓ Precise Inference: where the values of the derived data are accurate.

✓ Partial Inference: where the data values are partially disclosed, meaning that the user can have an idea of the value of data. This approach consists of three steps:

—**Step 1.** A domain expert identifies the sensitive elements to protect by interrogating the DW designer.

—**Step 2.** Build the graph inferences from the class diagram, specifying the elements that are specific or partial inferences.

—**Step 3.**Present the DW with UML annotations highlighting both types of inferences.

—**C.BLANCO et al. [2010]** proposed an approach based on the state diagram to detect inferences on the design level. This proposition focuses on sensitive requests and its evolutions, but they do not take into account to infer the data from the accessible data. The approach is presented as an OLAP security model of 3 states:

✓ Static model: presents the Fernandez Medina 2007 UML profile specific to DWs, adding a new kind of rule named Joint Rules, which presents the necessary privileges to certain combinations according to a specific grammar.

✓ Dynamic model: or the transactions-state model that has the objective of enhancing the static model, in order to ensure that confidentiality is not compromised by treating the evolutions of the combinations defined with JR through the application of OLAP operations.

✓ Session Control: This step makes it of interest to the user sessions to analyze them by checking each event in order to detect any possible inference.

—**L.SWEENEY [2002]** Describes a real data inference case, using a demonstration of the identification of sensitive data based on data crossing of an insurance group, those data supposed anonymous, as well as a voter registration list, which allowed the detection of the name of the former governor William Weld and his medical records, by linking shared attributes. He defined a protection model called k-anonymity including support policies, that allows avoiding inferences, which are:

✓ Sorting rows of the tables to be published in a random manner in order to not disclose sensitive information.

✓ Avoid having a row with a unique value in the table to be published.

✓ Take into account the old version of the data during the construction of the new table.

2.2 The security in the operating level of the DW

Online Analytical Processing (OLAP) has become increasingly an important component in decision support systems. The OLAP server is supposed to provide access based on authorizations defined for each user. He may refuse the access to data of measure, a dimension, and/or beyond a level in a hierarchy. Access rights can be explicitly specified on tables / columns of the tables of the DW. However, the OLAP server alone cannot protect access to prohibited data. Research works has been done to strengthen the access rights and authorizations of users, and to prevent any malicious user to infer prohibited data from data which he has access to.

—**R.KIRKGOZE et al. [1997]** defined a secure model for DWs which consists in the elaboration of a custom cube with its own dimensions and hierarchies. This model is based on the AMAC management policy. It is an extension of the MAC model that specifies the tasks that the user can perform according to its role within the organization. The advantage of this model is the flexibility in assigning roles to different virtual cubes.

—**PRIEBE and PERNUL [2000]** explored the security problems in the Goal Project, a project that aims to study the integration of a distributed information system. During this research, they have developed a proper method to the OLAP world. This method follows the traditional methodology in the elaboration of databases (analysis of needs, conceptual model, logical and physical), while incorporating the multi-dimensional aspect during the conceptual phase.

—**PRIEBE and PERNUL [2001]** continue their research on the creation of access control mechanisms to ensure data confidentiality, and they have created an access control mechanism in the form of a language expressing the security-related constraints, during the conceptual phase. This is a language based on MDX MulDimensionnale Xpression.

—**EAVIS and ALTHAMIMI [2012]** presented an authentication framework build on algebra especially designed for OLAP. It is object-oriented and uses query rewrite rules to ensure access to consistent data across all levels of the conceptual model. The process is essentially transparent to the user, a notification is provided in the case where a subset of the original request is returned. The final result is an intuitive and powerful approach for the database authentication that is uniquely adapted to the OLAP field.

—**S.TRIKI et al. [2013]** proposed an approach that does not require additional processing after each alimentation phase of the DW. It is based on Bayesian networks in order to protect a DW against inferences; they use a control module, which seeks to prohibit a user to infer protected data from the accessible data by using Min and Max aggregations functions.

2.3 Comparison and Synthesis of existing work

We presented in the previous section, research works that propose solving problems related to the confidentiality of access to the DW. Some approaches seem to be relevant and provide an acceptable level of privacy but still insufficient. In this section, we present a comparative table of these works based on criteria, and followed by a synthesis.

2.3.1 *Comparison.* The evaluation of some important approaches that address the confidentiality of DW for the development of a secured multidimensional model, is presented in Table 1 according to several criteria are:

— **Used approach:** the name of the proposed approach.

— **Used technique:** the technologies used in the proposed approach.

— **Transformation between models:** development of a logic model of ED from a conceptual model taking account security in every model.

— **Inferences management:** detection of feasible conclusions from the accessible data.

— **Validation:** : implementation of the proposed solution.

— **Data dynamic sensitivity:** automatic determination of data sensitivity level.

— **Traceability of access:** to trace the user transactions on the DW for analysis and decision making.

	Citation	Used Approach	Used Technique	Validation	Inferences Management	Data sensitivity	Traceability of access
Source	A.ROSENTHAL and SCIORE [2000]	Access permissions from the source	SQL grant/revoke	No	Yes	No	No
	SALTOR et al. [2002],	Secure DW from Federated Information Systems	No	No	No	No	No
Modelization	E.FERNANDEZ-MEDINA et al. [2006]	ACA	UML 2.0, OCL	No	No	No	No
	R.VILLARROEL et al. [2006]	UML 2.0	UML 2.0	No	No	No	No
	SOLER et al. [2008]	MDA	MDA	No	No	No	No
	C.BLANCO et al. [2010]	UML 2.0	the state diagram	Yes	Yes	No	No
	BLANCO et al. [2015]	MDA	UML 2.0, MDA	Yes	No	No	No
Operation	S.TRIKI et al. [2013]	protect a DW against inferences	Bayesian networks	Yes	Yes	No	No
	EAVIS and ALTHAMIMI [2012]	Authentication Framework	-	No	No	No	No
	PRIEBE and PERNUL [2001]	access control mechanism	MDX	No	No	No	No

Table I: Research works comparison

2.3.2 *Synthesis*

DW protection against illegal access was felt beyond reasonable doubt since several years E.FERNANDEZ-MEDINA et al. [2006], PRIEBE and PERNUL [2001], EAVIS and ALTHAMIMI [2012]. According to the authors KHAJARIA and KUMAR [2011],modeling the access control of the DW is the process of building an abstract model that needs to be stored in the DW. This model is a representation of the reality or a part of the reality. Following the study of existing works, we found the following:

—The confidentiality of the DW has traditionally been considered in the definitive implementation of a DW PRIEBE and PERNUL [2001], EAVIS and ALTHAMIMI [2012], however the most recent works BLANCO et al. [2015], RODRIGUEZ et al. [2011b] consider its inclusion in the development stages which can produce more robust quality solutions, and the system can accommodate these security requirements in a more natural way.

—In the research work R.MOUSSA and BADIR [2013], the authors proposed an automatic MDA architecture based on the UML extension to secure a DW, but the question that arises in this case is: What is the use of precising the security level in the user profile as well as for the elements of the DW since the permissions are specified.

—The majority of research works, especially those involved in the conceptual modeling phase relied on the CWM meta-model, in order to develop a secure DW. Knowing that the CWM model is based on three standards, namely UML, MOF and XMI to properly represent all security and audit rules defined in the conceptual modeling of DWs.

—Most of the works are modeling the access control based on MAC and RBAC policies, while the user profile is considered an isolated table that contains the necessary data for a static access of a user without taking account the priorities of the user authenticated.

—An MDA architecture for an automatic secure conception of a DW is applied in BLANCO et al. [2015], INMON [1991] but the two approaches were unable to understand the safety rules

that are complex.

—Although the permissions present the main axis to en-sure confidential access to the warehouse, however, the absence of a standard that supports the accuracy of these permissions can cause inconsistencies and inferences as consequences. In this sense, we find the work of SALTOR et al. [2002], which proposed the use of the permissions schema de-fined for federated databases without any modification to build a secure DW, and A.ROSENTHAL and SCIORE [2000]that propose rewriting requests in order to verify that they comply with the restrictions defined in the sources.

—Some authors A.ROSENTHAL and SCIORE [2000] , SALTOR et al. [2002] proposed to make the access control model in the DW, from data sources, while others PRIEBE and PERNUL [2001] , FERNNDEZ-MEDINA et al. [2007] have considered this proposition as difficult since the source data arise from different systems (with different policies). And operational systems use the relational model while the OLAP systems use the multidimensional model.

—Also note that the concept of inference was cited in several works as an essential element to ensure confidentiality, and whose mastery is crucial. In this sense, there is the work of S.TRIKI et al. [2013] which proposed an approach that can detect partial and accurate inferences, C.BLANCO et al. [2010] propose an approach based on the state-diagram to detect inferences in the conceptual phase without considering the possibility of inferences from accessible data. Nevertheless, despite the high risk of inferences, it is not sufficiently taken into account in the conceptual phase.

We find that most of the research works affects the task of classifying data according to their level of sensitivity (very sensitive, sensitive, and confidential) to the data owner. Knowing that according to the role of the user, the data owner assigns a level of data sensitivity in order to access data having the same level of sensitivity or lower. The owner of the DW may then assign a lower level of sensitivity to a critical data. This results, however, a problem of information confidentiality loss. In addition, the permissions defined in the sources are not sufficiently exploited to help the owner well determine the permissions of a DW user. In the next section, we propose an approach that overcomes these limitations.

3. DYNAMIC MANAGEMENT OF SECURITY LEVELS

3.1 Motivation

Confidentiality of a DW is based on the access control model that specifies the permissions granted to each user. According to studies cited in the state of the art section, access permission to an object of the DW is granted to a user according to his role. The sensitivity level makes data accessible to the user, with a sensitivity level defined on the object of DW according to its role. Such situation is difficult to manage by the data owner, who can assign a not proper level of sensitivity to an object of DW. What may abuse its confidentiality. For this we propose a different way of defining user permissions by using the permissions defined in the data sources such as suggestions that can help the data owner. Then our access control model can generate the sensitivity level of each object in the DW based on these permissions. This classification will help us to trace user actions on sensitive data. In the remainder of this section we present the architecture of our contribution, and then we detail our contribution which consists of two stages.

3.2 Architecture

In this section we present the architecture of our proposition which its composites of three parts (Figure 1): **1.Dynamic classification of DW sensitivity levels based on user profiles** Consists of defining the access permissions of each user according to his role, and generating the sensitivity level of data. **2. Detecting inferences by combining several profiles** allows detecting the sensitive combinations of permissions that can produce inferences, for a user that combines multiple roles. **3. Managing profiles from uses** Allows to manage the access to the

DW hosted in the Cloud Computing (CC). The objective is to minimize the traffic, and increase the performance of the treatments.
In this article, we present our approach for the first part of our architecture which consists in establishing a DAC model (Dynamic Access Control) based on the PPU (Profession Profile of a User) constituted of several mechanisms (Figure 1): **1.1 Assignment Management of permissions:** This is to define the permissions of a role on a data with a given specified privilege, taking into account the permissions defined on the data sources level. And suggest them to the administrator, data owner of the ED, when as-signing permissions. **a. Data classification mechanism:** Automatic generation of data sensitivity levels based on defined permissions. **b. Traceability of sensitive data:** Is a mechanism to trace user actions on the data with a high level of sensitivity. **c. Sending alerts Mechanism:** Depending on user access traceability mechanism, our alert system can send notifications to the administrator during an attempted violation of permission on a sensitive data.

This part of our architecture, that allows managing dynamically of sensitivity levels of DW

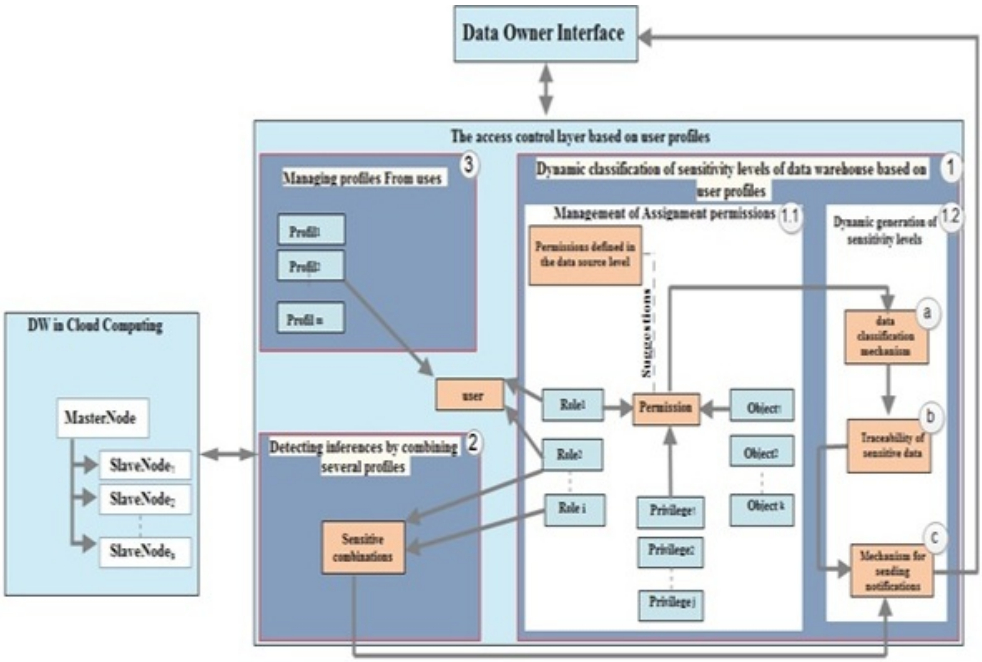


Figure 1 Global architecture Proposed for controlling access to DWs based on user profiles

based on user profiles, is presented in a meta-model form. Its an extension of meta-model CWM (Common Warehouse Meta-model) and presents the users profession profile. It contains five classes considered the core of our contribution (Figure 2):

—**PermissionDW:** present the permission of access to a role with a specific privilege on an object of the DW.

—**SourcePermission:** for each object in the DW, we precise the object and the source permission.

—**AutomaticSecurityLevels:** generation of the level of sensitivity of an object according to the defined permissions.

—**Threshold:** Specified by the data owner for each object, to define the sensitive data.

—**Traceability:** according to the sensitivity level of the object, we trace the actions made on sensitive data.Alert: sending alerts allows claiming an attempt of unauthorized access to a sensitive data by a user.

3.3 Dynamic management of DW sensitivity levels based on user profiles

Our solution is structured following two parts:

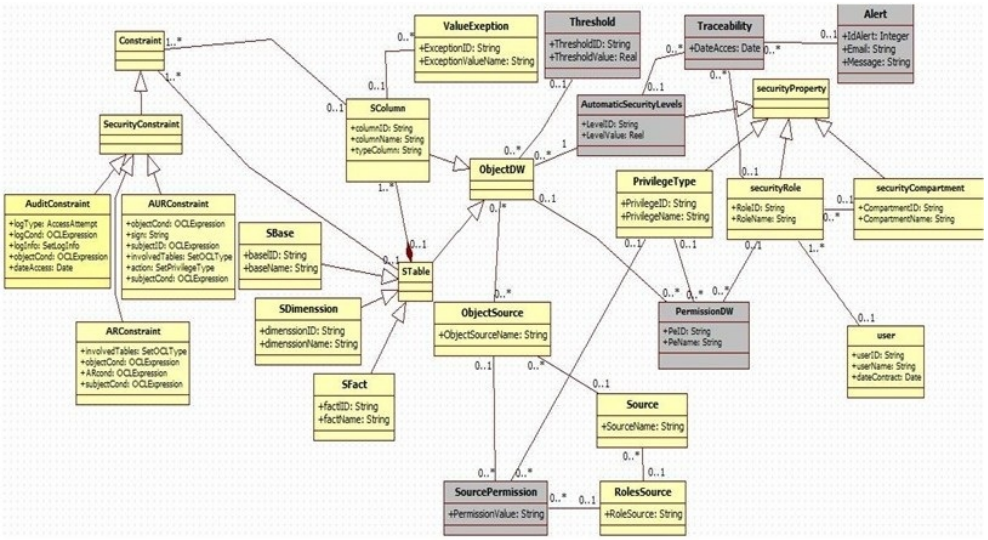


Figure 2 Meta-Model proposed for managing dynamically the sensitivity levels of DW based on user profiles

3.3.1 *Management of assignment permissions .*

The permission is a primordial axis in an access control mechanism, the management of these permissions is a difficult task for the administrator. In this phase, we propose to use the permissions defined in the data sources as suggestions that will help the data owner to well define the permission of a role on an object in the DW according to a given privilege $\mathrm{P}(\mathrm{R}_i, Pr_j, O_k) where$:

P : Permission $(0, 1)$.

R_i**:** Role of the user.

Pr_j**:** Privilege permission (Read, Write, Modification).

O_k**:** Object of the DW (Table fact, dimension, column, column value).

	R_1	R_2	R_3	R_i
N_1	✓			
N_2		✓		
N_3				✓
N_i		✓		

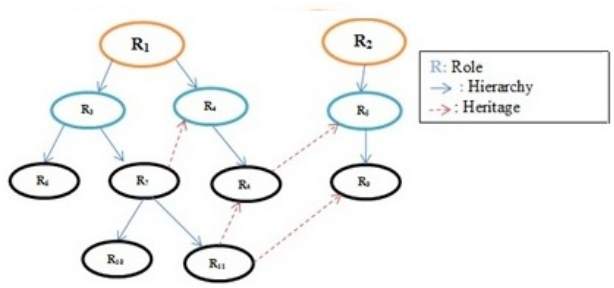
—A user can have one or more roles. Whith N is a hierarchical level,U is a user and O is an Object

	R_1	R_2	R_3	R_i
U_1	✓			
U_2		✓	✓	
U_3			✓	✓
U_m		✓		

—The role can consult one or more objects.

	R_1	R_2	R_3	R_4	R_i
O_1	1	1	0	0	1
O_2	1	1	1	1	0
O_3	0	0	0	0	1
O_k	1	0	0	0	1

—A role can inherit permissions from one or more roles.



According to our meta-model, each object of a DW is presented by the class "ObjectDW" which can be a fact, a dimension, a base, a column or a value of a column. In order to help the owner to determine the permissions of an "ObjectDW", our system suggests the source permissions of an "ObjestDW" to help the owner in the determination permissions phase of the DW. The "ObjectSource" class presents the corresponding object of an "ObjectDW" in the "Source" database for each role source "RoleSource" by viewing the permissions granted in the source "SourcePermission".

3.3.2 *Dynamic generation of data sensitivity levels.* The sensitivity level of an object o is determined by the number of roles that have permission to read it, relative to the total number of roles in each hierarchical level. In order to automatically generate a percentage that presents the sensitivity level of object o, our system is based on the following rules:

—The roles are grouped in the set R $=r_1, r_2, ...r_i$

—Each role belongs to a hierarchical level Well-defined N= 1, 2,...n Where P is the total number of hierarchical levels.

—Each hierarchical level has a coefficient CN=N.

—Each role belongs to a hierarchical level H=$h_1, h_2, ..., h_p$

—A role may inherit access rights of another role.

—The objects in the DW are grouped in the set O=$o_1, o_2, ..., o_k$

—R_i, Pr_j, O_k= 0, 1, With Pr_j is a consultation.

—The object consulted by several roles is less sensitive than the object that is accessed by a small number of roles. The sensitivity level is calculated as a percentage.

relationship that allows calculating the sensitivity level is as follows:

$$[1 - (\sum(\sum(P(O_k))/(\sum(R_j))) * C_i)/CT] * 100 \quad (1)$$

The sensitivity level (%) = Such as:

—$\sum(P(O_k))$: The sum of the roles in a hierarchical level i which have the right to consult an object.

—$CT = \sum(C_1, C_2, ..., C_n)$: The sum of the hierarchical levels coefficients.

—$\sum(R_j)$: The sum of roles in a hierarchical level.

—C_i : The Coefficient of a hierarchy Level i.

The following algorithm presents the proposed method for defining the sensitivity level of each object in the DW:

ALGORITHM DW Classification

```
Input: P: Total number of Hierarchical Levels
       OT: Total number of warehouse objects
Output: SL []: Table of sensitivity levels
Begin
       double [] SL = new double [OT+1];
       int n = 1, h = 0, CT=0;
       double[] C = new double[P+1];
       double[][] PP = new double[P][OT];
       double[] PK=new double[OT];
While(n ≤ P){
       C[n] =n;
       CT = (int) (CT + C[n]);
       n++;
}
While(h  < P){
       For(int k=0; k  < OT; k++){
       PP[h][k] = Permissions(h,k)*C[h];
       PK[k] =PK[k] + PP[h][k];
}
h++;
}
h=1;
while(h ≤ P){
       for(int k=1; k  < OT+1; k++){
              SL[k]=(1-(PP[h][k]/(double)CT))*100;
              System.out.println("Sensitivitylevelofthe object["+k+"] is " + SL[k] + "%");
}
h++;
}
f return SL;
}
End.
```

The loop one of our algorithm allows to calculate the coefficient of each hierarchical level which is the increasing order number of the levels. The loop two uses the function "Permissions (h, k)" for determiner the number of roles that have permission to read an object O, multiplied by the hierarchical level coefficient. And the loop three obtains a percentage that presents the sensitivity level of the object k as depending on the number of roles that have permission to read the object, the hierarchical levels of the roles and their coefficients.

3.3.3 *Choice of thresholds.* Each object of the DW has a threshold. This is a variable parameter (adjustment cursor of the sensitivity level), which depends on the context of the company such as:

—Activity type.

—Period (crisis, war).

—Events (internal or external).

—Business strategy ...

The sensitive data of an object of the DW are the data whose sensitivity level is greater than or equal to the threshold of this object.

3.3.4 *Traceability and alert.* In order to trace user actions on sensitive data, and send alerts to a data owner to inform them of attempts violation permissions, our system is based on sensitivity levels generated. Thus, traceability allows tracing the actions of the users on the sensitive data, whose sensitivity level exceeds the threshold specified by the data owner. Then our system sends an alert to the owner if the action is an attempt to violate the permissions. The figure 3 clarifies how our system treats the users request:

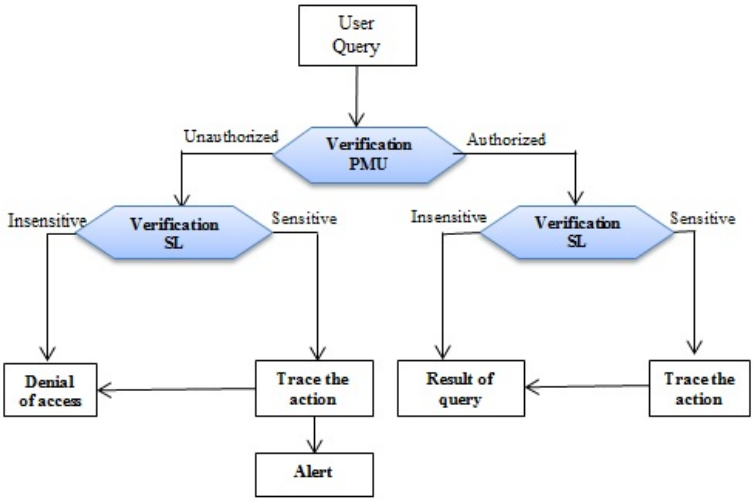


Figure 3 Tractability and alert

—If the requested data are allowed, according to the user's profile, the classification module checks their sensitivity level. If a data item is sensitive, the action will be plotted in the traceability class of our meta-model, and the user will receive the result of the requested query. If the requested data are not sensitive the user receives the result without plotting the action.

—If the requested data are not allowed according to the user's profile, the classification module checks their level of sensitivity. If a data is sensitive, the owner of the data receives a notification containing the details of the attempted violation of access permissions, in order to react. And of course access will be denied.

4. CASE STUDY AND IMPLEMENTATION

4.1 Example

Take as an example the ED presented in Figure 5, which consists of two tables of fact: "Diagnosis" and "Hospitalization" and a set of dimensions: "Illness", "Diagnosis_detail", "Diagnostic_group"...

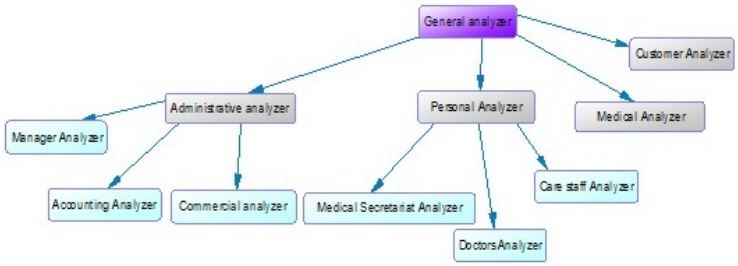


Figure 4 Example of a hierarchy of roles

The figure 4 shows the hierarchy of roles operating the data warehouse for analysis and decisions making:

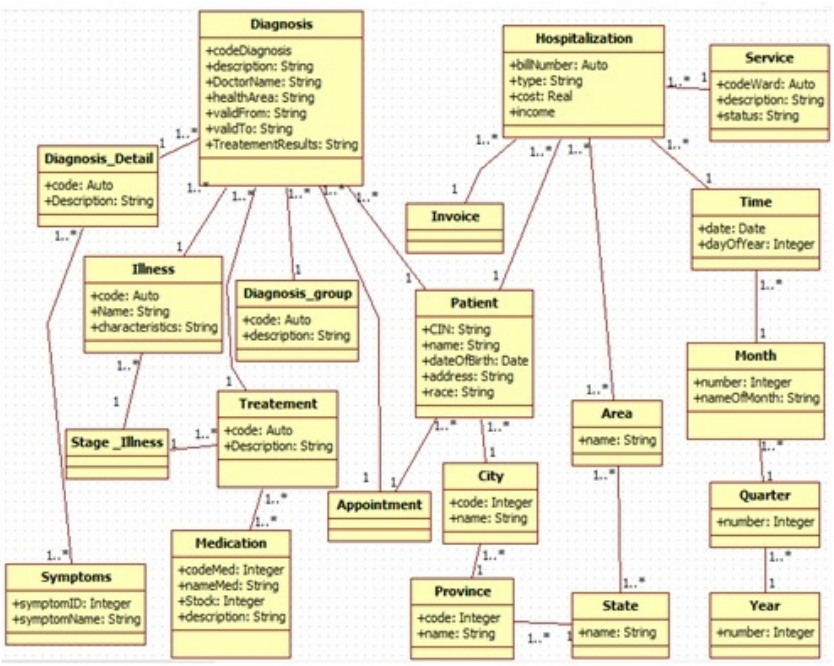


Figure 5 Data warehouse TEST

We have three hierarchical levels (Tableau 2), each level contains a set of roles and a coefficient:

— **Hierarchy level 1 (H1)** contains the role General Analyzer with the coefficient C1 = 1.

— **Hierarchy level 2 (H2)** contains the roles Administrative Analyzer, Personal Analyzer, Medical Analyzer, Customers Analyzer with the coefficient C2 = 2.

— **Hierarchy level 3 (H3)** contains the roles Manager Analyzer, Accounting Analyzer, Business Analyzer, Medical Secretariat Analyzer, Doctors Analyzer, Personal Care Analyzer with the coefficient C3 = 3.

	Coefficient of $\mathbf{H_1(C_1 = 1)}$	**Coefficient of $\mathbf{H_2(C_2 = 2)}$**	**Coefficient of $\mathbf{H_3(C_3 = 3)}$**
Hierarchical level 1 (H_1)	General analyzer		
Hierarchical level 2 (H_2)		—Administrative analyzer, —Personal analyzer, —Medical analyzer, —Customers analyzer	
Hierarchical level 3 (H_3)			—Manager analyzer, —Accounting analyzer, —Commercial analyzer, —Medical Secretariat analyzer, —Doctors analyzer, —Care staff analyzer

Table II: the roles of each hierarchical level

The distribution of permissions performed by the data owner, on the objects "Service", "Patient", "Treatment" is shown in the table below:

—**Let's calculate the sensitivity level for the object "Service"**
$[1 - (\frac{(\frac{1}{1}*1)+(\frac{4}{4}*2)+(\frac{6}{6}*3))}{6}] * 100$ The object "Service" is accessible by all roles, so normally it must have 0% as a sensitivity level.
Sensitivity level (%) = 0% Such as:

—The sum of the permissions P of the object "Service" in the hierarchical level H1 is equal to 1.

—The sum of the coefficients of the hierarchical levels $CT = \sum_1^n C_i = 1 + 2 + 3$.

—The sum of the roles R in the hierarchical level H1 is equal to 1.

—The coefficient C of the hierarchical level H1 is equal to 1.

	General Analyzer	Administrative Analyze	Personal Analyzer	Customers Analyzer	Medical Analyzer	Manager Analyzer	Accounting Analyzer	Business Analyzer	Medical Secretariat Analyzer	Doctors Analyzer	Personal Care Analyzer
H_1	Service										
H_2		Service	Service	Service	Service						
				Patient	Patient						
H_3						Service	Service	Service	Service	Service	Service
									Patient		
										Treatment	Treatment

Table III: Distribution of permissions

—**Let's calculate the sensitivity level for the object "Patient":** The object "Patient" is accessible by the roles: "Customers Analyzer", "Analyzer Medical Secretariat" and "Medical Analyzer", its sensitivity level is calculated as follows: $[1 - (\frac{(\frac{0}{1}*1)+(\frac{2}{4}*2)+(\frac{1}{6}*3))}{6}] * 100$ Sensitivity level (%) = 75% Such as:

—The sum of the permissions P of the object "Patient" in the hierarchical level H1 is equal to 0.
—The sum of the coefficients of the hierarchical levels $CT = \sum_1^n C_i = 1 + 2 + 3$.
—The sum of the roles R in the hierarchical level H1 is equal to 1.
—The coefficient C of the hierarchical level H1 is equal to 1.

The tableau 4 shows the threshold of each object in the DW (Dimension, Fact, Column) specified by the data owner, in order to plot user access to sensitive data.

Object	Threshold
Service	50%
Patient	50%
StageIllness	30%
Treatment	50%

The object "Patient" is a sensitive object because its sensitivity level is higher than the threshold, and consequently, our system will trace the actions of the users on this data and send alerts to the owner in the case of a user who is not allowed try to access it.
The object "Service" is an insensitive object because its sensitivity level is below the threshold, and consequently, our system will not trace the actions of the users on this object.

4.2 Implementation

4.2.1 *Tools and development environment* . For the implementation of our contribution and the realization of the experiments, we used a machine DELL PRECISION T1700 under Windows 7 professional 64-bit. This machine has a 3.40 Ghz Intel Core i7-4770 processor and 8 GB of RAM. We used the following software tools:

—The Java language with the Eclipse development environment. This choice was motivated by the advantages offered by this language in terms of portability, robustness and the availability of many libraries;
—The Oracle version 12c DBMS to design the test data warehouse (Figure 5). This choice was mainly based on the availability of the OLAP option, which includes an OLAP analytics engine, workspaces, and an Analytic Workspace Manager (AWM).;
—The MySQL relational DBMS to design our Meta Model (Figure 2). This choice was mainly argued by its simplicity of use and its interfaces to perform various operations.

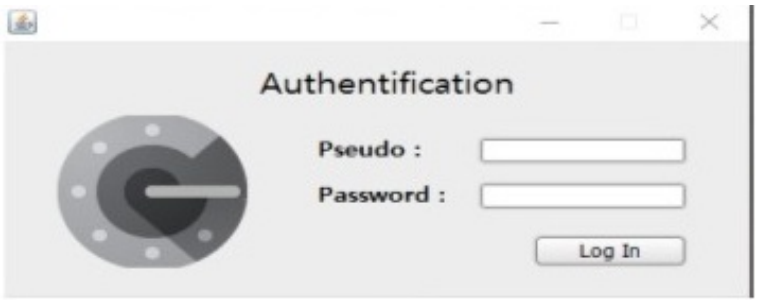
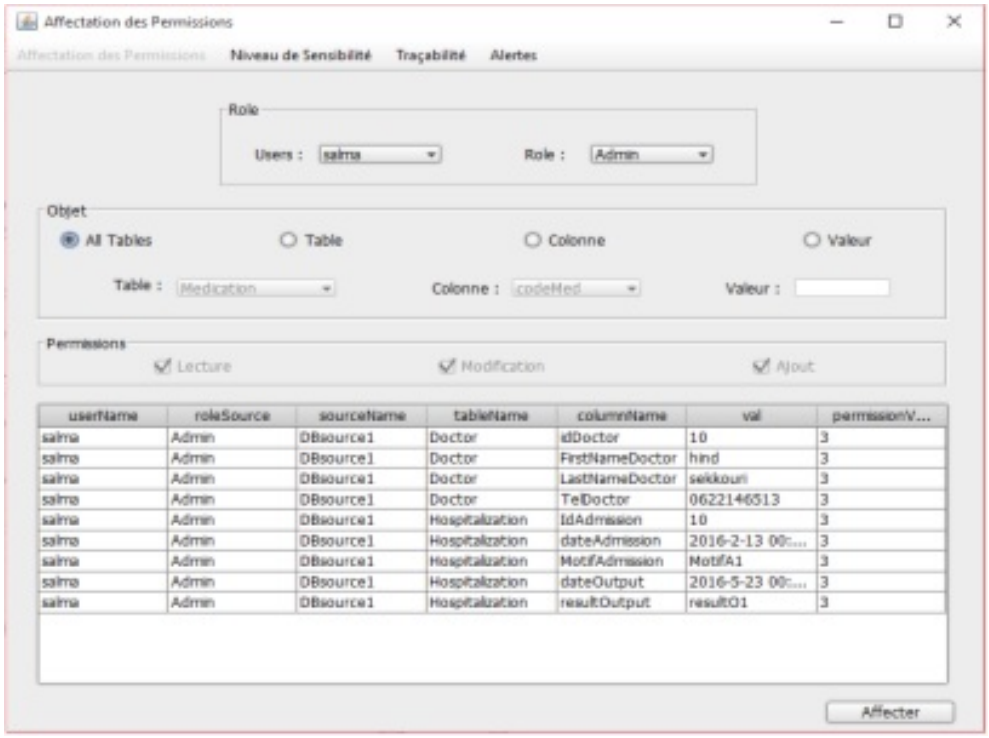


Figure 6 Authentication interface

4.2.2 *Authentication interface.* The administrator must authenticate by specifying the login and password in order to access the interfaces. The program checks this information in our meta-model (Figure 2).

4.2.3 *Permissions assignment interface*



userName	roleSource	sourceName	tableName	columnName	val	permissionV...
salma	Admin	DBsource1	Doctor	idDoctor	10	3
salma	Admin	DBsource1	Doctor	FirstNameDoctor	hind	3
salma	Admin	DBsource1	Doctor	LastNameDoctor	sekkouri	3
salma	Admin	DBsource1	Doctor	TelDoctor	0622146513	3
salma	Admin	DBsource1	Hospitalization	IdAdmission	10	3
salma	Admin	DBsource1	Hospitalization	dateAdmission	2016-2-13 00:...	3
salma	Admin	DBsource1	Hospitalization	MotifAdmission	MotifA1	3
salma	Admin	DBsource1	Hospitalization	dateOutput	2016-5-23 00:...	3
salma	Admin	DBsource1	Hospitalization	resultOutput	resultO1	3

Figure 7 Permissions assignment interface

The assignment of permissions is a task that only the administrator does it. In this interface we choose the role we will assign to a user.
The program shows us information about this user and its authorized objects at the source database level, what has already been discussed inOuazzani et al. [2016],in order to help the administrator to define the permissions. This interface makes it possible to specify the type of object (table / column / value), and the privileges to be authorized.

4.2.4 *Sensitivity Level Interface.* According to the choice of objects (tables / columns / values) made by the administrator, this interface makes it possible to display their sensitivity level calculated by the algorithm proposed.

4.2.5 *Traceability interface .* Traceability makes it possible to find the history actions made by each user. The interface is used to display the details of the actions that are allowed.
Only the administrator has the right to access it, the interest is to keep a trace for use in module 2 which can detect inferences between the profiles of a user.
Actions can be numerous. In order to facilitate the research, the administrator has the option of sorting the data according to the date, the object or the order of the users.

4.2.6 *Alert interface .* The alert is a signal that prevents a danger. In our program, it presents the unauthorized actions made by the users, that is, if a user who does not have the right to access a particular table and requests a request to access it anyway, the program reports an alert. These are controlled by the administrator. Alerts can be sorted by date, subject, or user order.

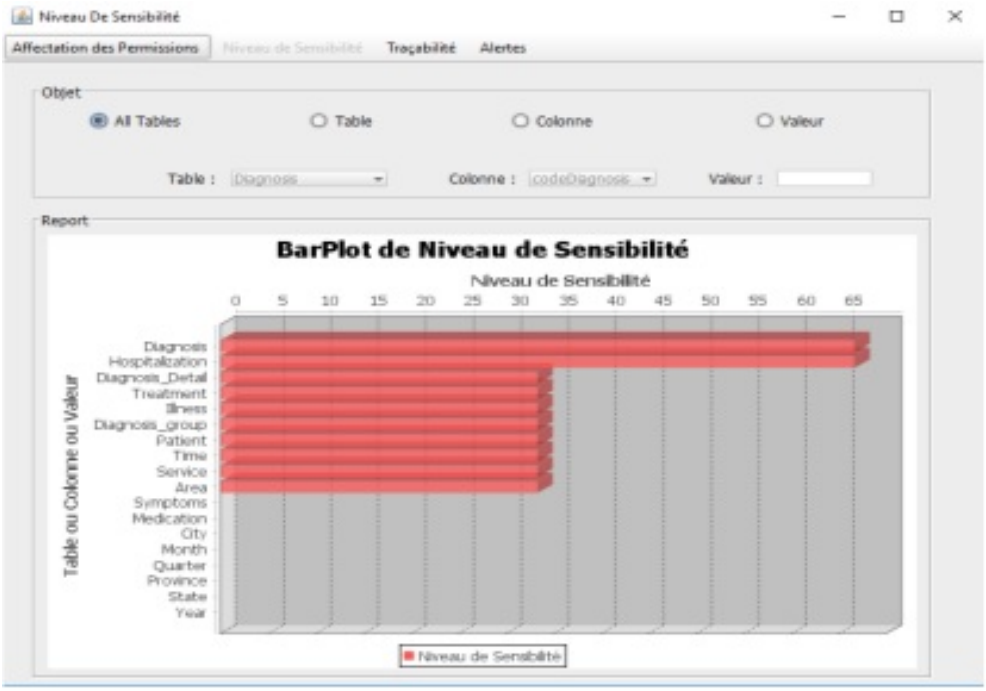
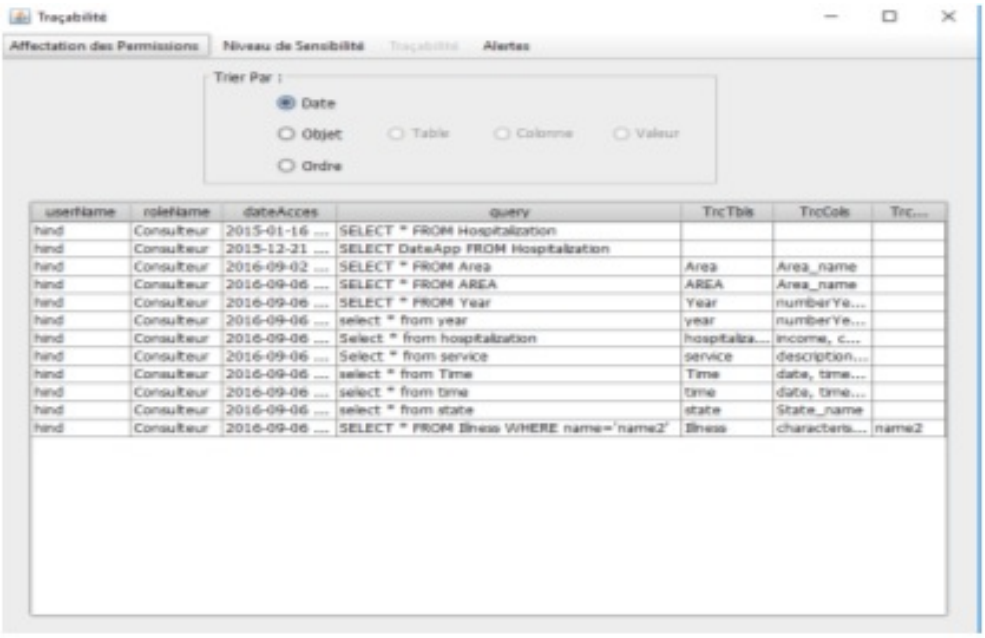


Figure 8 Sensitivity level interface



userName	roleName	dateAcces	query	TrcTbls	TrcCols	Trc...
hind	Consulteur	2015-01-16 ...	SELECT * FROM Hospitalization			
hind	Consulteur	2015-12-21 ...	SELECT DateApp FROM Hospitalization			
hind	Consulteur	2016-09-02 ...	SELECT * FROM Area	Area	Area_name	
hind	Consulteur	2016-09-06 ...	SELECT * FROM AREA	AREA	Area_name	
hind	Consulteur	2016-09-06 ...	SELECT * FROM Year	Year	numberYe...	
hind	Consulteur	2016-09-06 ...	select * from year	year	numberYe...	
hind	Consulteur	2016-09-06 ...	Select * from hospitalization	hospitaliza...	income, c...	
hind	Consulteur	2016-09-06 ...	Select * from service	service	description...	
hind	Consulteur	2016-09-06 ...	select * from Time	Time	date, time...	
hind	Consulteur	2016-09-06 ...	select * from time	time	date, time...	
hind	Consulteur	2016-09-06 ...	select * from state	state	State_name	
hind	Consulteur	2016-09-06 ...	SELECT * FROM Illness WHERE name='name2'	Illness	characteris...	name2

Figure 9 Traceability interface

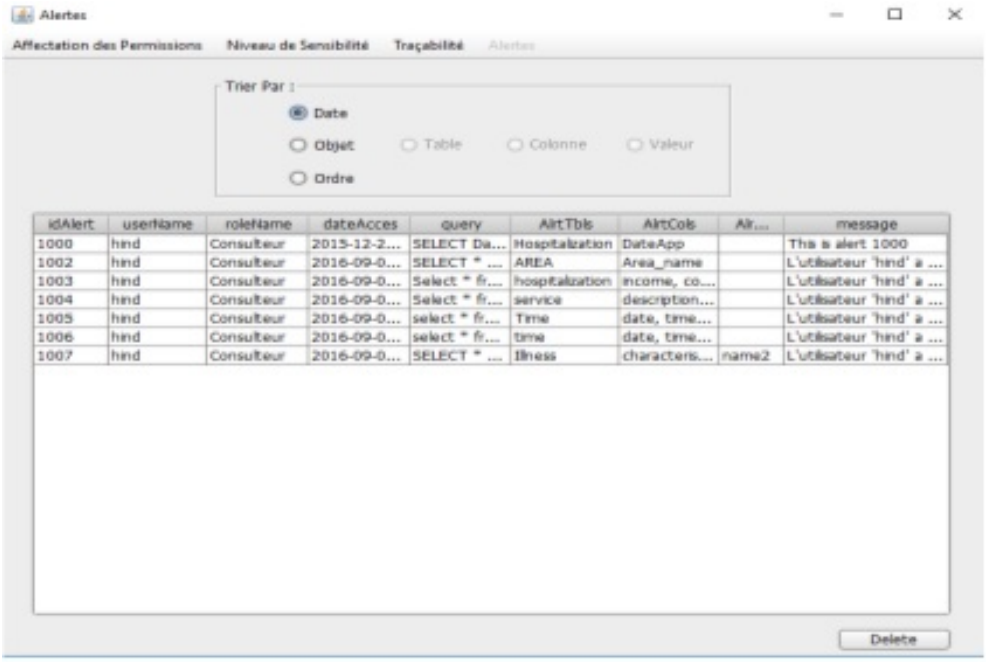


Figure 10 Alert interface

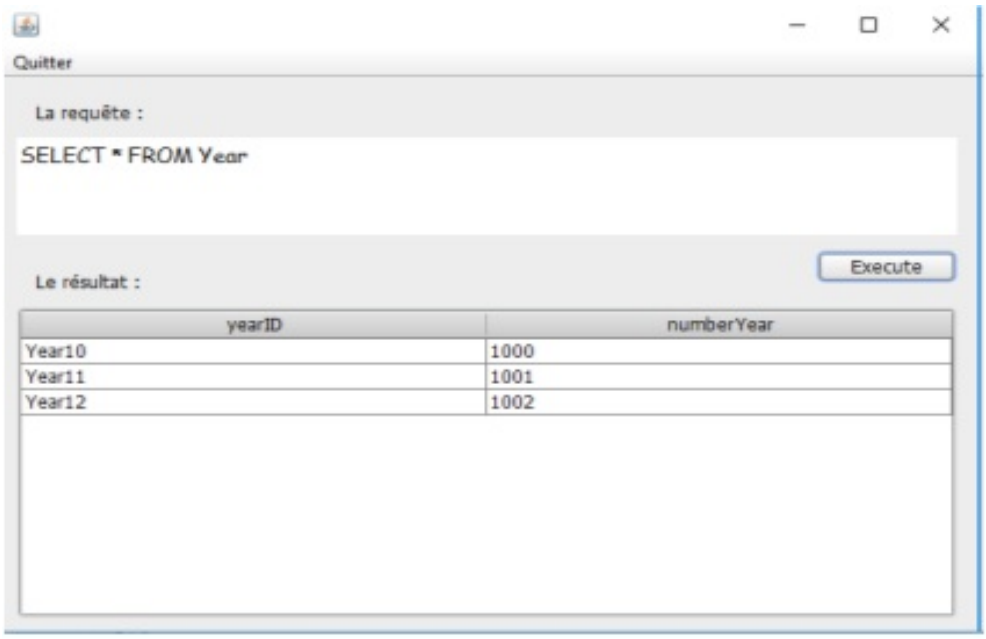


Figure 11 User interface

4.2.7 *User interface.* After user authentication, he has to activate a profile among his profiles, knowing that each profile concerns only one role.
The user interface allows the user to execute one or more queries according to the activated profile.
The user formulate his request with SQL (Structured Query Language) a computer language used to manipulate data from a database or other information systems.
If the user has sufficient permissions, the system displays the results of the desired query, and records the information automatically in the traceability table.

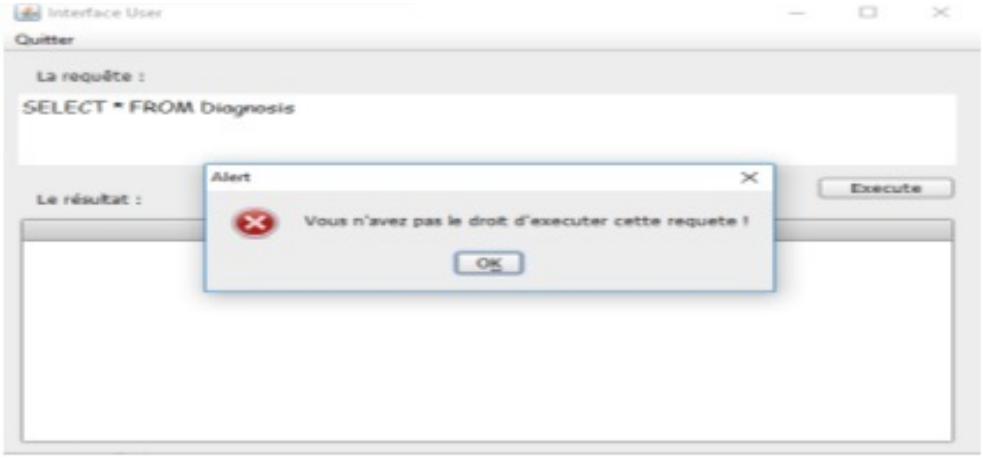


Figure 12 Exemple d'excution d'une requte

In the case where the user does not have the right to access data expressed in the request, the program does not display any results and automatically saves the information in the alert table.

5. CONCLUSION

Because of the critical information stored in the DW, it is important to check its privacy. The confidentiality in the context of DW is considered an important requirement, which must be ensured by an authorization management mechanism which is the specification and execution of access rights in the databases in general and more specifically in DWs. In this study allowed us to see the major problems of the confidentiality of the data in the warehouse. To reduce these risks, we have proposed a solution based on the user profile, which consists in the definition of access permissions according to the user role using the access rights defined in the sources, generate the level of sensitivity of each object in the DW according to these permissions, trace the access and detect violation attempts of access rights on a sensitive data (a data with a high level of sensitivity). The objective of this solution is to reduce the vulnerability of the data in a DW, and help the owner of the DW to well manage the access control of the users.

6. PERSPECTIVES

The results show that the research prospects in this direction are numerous, however we have fixed three perspectives that we see interesting in the context of this work, which are:

—Implement our approach in a CC environment: Migration data warehouses to CC should improve satisfaction of users and increase productivity, which requires a high performance.

From the moment that a data will be entrusted to an external service provider, the establishment of an access control mechanism should not increase the burden of processing, because the aim is to have an evolutionary and productive system which data are well protected against prohibited access.

—Detect inferences by combining accessible data: The detection of inferences by the combination of data, has not taken much interest in researchers, despite its importance for data warehouses that use multiple data sources. Queries that use the combination of several multidimensional elements tend to be more sensitive than data separately. These combinations must be detected in order to avoid the inference of unauthorized data.

References

A.ROSENTHAL AND SCIORE, S. 2000. View security as the basis for data warehouse security.

BLANCO, C., GUZMN, I. G. R. D., FERNNDEZ-MEDINA, E., AND TRUJILLO, J. 2015. An architecture for automatically developing secure olap applications from models. *Information and Software Technology 59*, 1–16.

C.BLANCO, FERNANDEZ-MEDINA, E., TRUJILLO, J., AND JURJENS, J. 2010. Towards the secure modelling of olap users behaviour.

DEVBANDU, P. AND STUBBLEBINE, S. 2000. Software engineering for security: a roadmap. *Finkelstein, A. (ed.) The Future of Software Engineering, ACM Press, New York*, pp.227–239.

EAVIS, T. AND ALTHAMIMI, A. 2012. Olap authentication and autho-rization via queryre-writing. *The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications*, 130–139.

E.FERNANDEZ-MEDINA, TRUJILLO, J., VILLARROEL, R., AND PIATTINI, M. 2006. Access control and audit model for the multidimensional modeling of dws. *Decision Support Systems 12701289*.

FERNNDEZ-MEDINA, E., TRUJILLO, J., VILLARROEL, R., AND PIATTINI, M. 2007. Developing secure data warehouses with a uml extension. *Information Systems 32(6)*, 826–856.

INMON. 1991. Building the data warehouse.

J.TRUJILLO, E.SOLER, E.FERNNDEZ-MEDINA, AND M.PIATTINI. 2009. A uml 2.0 profile to define security requirements for data warehouses. *Computer Standards and Interfaces 31(5), 969-983*.

KHAJARIA, K. AND KUMAR, M. 2011. Evaluation of approaches for modeling of security in data warehouses. *Advances in Computing and Communications Springer Berlin Heidelberg 26*, pp. 9–18.

LANDWEHR, C. 2001. Computer security. *Inter Journal of Information Security 13*.

L.SWEENEY. 2002. k-anonymity: A model for protecting privacy. *Advances in Computing and Communications Springer Berlin Heidelberg 25*.

OUAZZANI, A., RHAZLANE, S., N.HARBI, AND H.BADIR. 2016. Dynamic management of data warehouse security levels based on user profiles. *Information Science and Technology (CiSt),4th IEEE International Colloquium*, 59–64.

PRIEBE, T. AND PERNUL, G. 2000. Towards olap security design survey and research issues. *Proceedings of the 3rd ACM International Workshop on Data Warehousing and OLAP 3340*.

PRIEBE, T. AND PERNUL, G. 2001. A pragmatic approach to concep-tual modeling of olap security. *Proceedings of the 20th Interna-tional Conference on Conceptual Modeling (ER01) 2224*, 311–324.

R.KIRKGOZE, KATIC, N., STOLBA, M., AND TJOA, A. 1997. A security concept for olap. *A security concept for olap. Proceedings of the 8th International Workshop on Database and Expert System Applications*, 619–626.

R.MOUSSA AND BADIR, H. 2013. Data warehouse systems in the cloud: rise to the benchmarking challenge. *Journal International of Computers and Their Applications 245*.

RODRIGUEZ, A., FERNANDEZ-MEDINA, E., TRUJILLO, J., AND PIATTINI, M. 2011a. Secure business process model specification through a uml 2.0 activity diagram profile. *24*.

RODRIGUEZ, A., FERNANDEZ-MEDINA, E., TRUJILLO, J., AND PIATTINI, M. 2011b. Secure business process model specification through a uml 2.0 activity diagram profile.

ROSENTHAL, A. AND SCIORE, S. 2000. View security as the basis for data warehouse securitye. *DMDW*.

R.VILLARROEL, FERNANDEZ-MEDINA, E., AND PIATTINI, M. 2006. A uml 2.0/ocl extension for designing secure data warehouses. *Journal of Research and Practice in Information Technology 23*, 31–43.

SALTOR, F., OLIVA, M., ABELLO, A., AND SAMOS, J. 2002. Building se-cure data warehouse schemas from federated information systems.

SOLER, E., STEFANOV, V., AND MAZON, N. 2008. Towards comprehensive requirement analysis for data warehouses. *Considering Security Requirements IEEE, Los Alamitos*, 104–111.

S.TRIKI, BEN-ABDALLAH, H., BOUSSAID, O., AND HARBI, N. 2013. Scurisation des entrepts de donnes: de la conception lexploitation. *Rapport de thse*.

Mrs. Amina El ouazzani is PhD from National School of Applied Sciences, Tangier Morocco. His research interests are in cloud computing and the confidentiality confidentiality of data warehouses.

Dr. Nouria Harbi is an associate professor at the University Lumiere 2 Lyon, France. she received her PhD in computer science from the INSA Lyon, France in 1990. His research interests are in Multidimensional modeling of complex data, Security of decision-making information systems: Data warehouse access control based on profile management, Data integrity in decision-making information systems, Data mining and security (intrusion detection), Confidentiality and availability of data in Cloud Computing.

Dr. Hassan Badir is an associate professor at the University Abdelmalek Essadi , Tetouan Morocco , he received Ph.D. degree in Database complex Design from Claude Bernard University Lyon1 in 2004. He studies in the fields of Computer Science, specifically Business Intelligence, Database, optimization, Cloud Computing and Big data. - Head of Computer Science and Complex Systems Master - Co-Investigator of Human Heredity and Health African Bioinformatics Network - Head and founding member of the Moroccan Innovation and New trends of Informatio System Society.