Active Categorization of Sensibility Degrees of Datawarehouse Established on User Visibilities

Ayasha Siddiqua, Shaista Sabeer, Nivedita Soni

1Lecturer Jazan University, Saudi Arabia
2Lecturer Jazan University, Saudi Arabia
3Technical Analyst HCL

Abstract: A data warehouse arsenals confidential information about the privateness of someone and significant commercial activity. This builds admittance to this origin a risk of revelation of information. Therefore the importances of enforcing security which ensure the data discretion by demonstrating an accession control policy. In this centering, several suggestions were made, but none are conceived as a standard for access direction to data warehouses. In this paper, we will demonstrate our advance that appropriates first to exploit the permits determined in the data sources in order to help the executive to define access licenses to the data warehouse, and then our system will automatically give the sensitivity level of each data warehouse component according to the permits granted to an object in the data warehouse.

Keywords: Data Warehouse, Privacy, Profession profile, Crucial data, Tractable

I. INTRODUCTION

A data warehouse is a vital component for high focus looking for build the right strategically decision. It is a database of terabytes data stacked away historically, from usable systems to have a authorize view and a rich origin for decision makers. It demonstrates a source of critical business concern data and the data of customer’s secrecy such as medical and financial information protected by laws. Accordingly, they should not be approachable without access assure.

Because of the vital data stored in the Data Warehouse, it is crucial to check its secrecy. According to [6], concealment in the context of Data Warehouse is conceived an significant necessity, which must be checked by an authority management mechanics which is the stipulation and performance of access rights in the information in general and more specifically in Data Warehouses.

In general, schematic security demands are summed up by the signifier CIA (confidentiality, integrity and availability). All other security measure necessities such as authentication, authorization, access control, etc., can be attributed to these three basic attributes. Secrecy is determined as the absence of revealing of wildcat information. Unity is defined as the absence of the wildcat modification of data, and availability that ensures the continuity of service [8].

Our destination is to assure the discretion of the Data Warehouse by accession control, since there is no criterion that allots this significant expression, and mechanisms of circumpection assigned for OLTP arrangements cannot be employed for the Data Warehouse because in functional organizations, access assure is determined on the tables, rows, columns, etc. While in an Data Warehouse, we have a large number of exploiters with dissimilar analysis needs, attempting access to the multidimensional Data Warehouse [10,11].

In this paper, we center on the abstract patterning process of the Data Warehouse by extending an approach path furnishing accession control, based on the job visibility of a user who depicts its access rights. We employ the RBAC access assures insurance (Role-based Access Control) that concentrates on grouping users allowing to their professing or their roles. With this approach path, we are able to assort the Data Warehouse mechanically by giving their sensitivity levels, in order to describe user accomplishes on sore data.

II. STATE OF ART AND DEDUCTIVE REASONING

Lately, a number of Data Warehouse protection examples have been advised. In this division we will coordinate these explore works according to two approaches, the integration of security in the modeling process of the Data Warehouse, and the Data Warehouse access assure models already in place.

2.1. Protection at the Abstract Level of the Data Warehouse

2.1.1. The Datawarehouses Protection Based on Permits Determined at the Origins

Rosemary and Scion (2000) advise a theoretical access that exploits the accession licenses determined in the data source, rather than producing new access mechanisms. They use the revision interrogations to assert that they comply with the limitations defined in the data beginnings, as well as creating comparative views in order to denigrate the risk to infer sore information.
Saltor and al. (2002) since there are resemblances amongst the computer architecture of federate information and the computer architecture of a Data Warehouse, the authors advised the employ of the multi-level accession licenses convention determined for the federate information without being changed to build a ensure Data Warehouse, this authority strategy depicts the multi level accession rules.

2.1.2. Concealment of the Information in a Data Warehouse throughout the Modelization

Amongst the works that have been acquired on incorporating protection into the modelization of warehouses, we determine:

Fernandez-Al Madinah and al. (2006) have acquired an accession control and auditing model particular for Data Warehouse, grounded on two access direction insurances: MAC and RBAC. They define the protection rules throughout the moderate procedure of a conceptual model, by containing the concept of “user profile”, which comprises of an isolated table comprising all user selective information (identity, classification level: top secret, secret, confidential or unknown). This example remains a purely theoretic model since no solution for its execution has been aimed.

Villarroel and al. (2006) have determined an OCL extension Object Constraint Language employing UML2.0 extension mechanisms to conclude issues of privateness; this university extension defines the security necessities of the components during the abstract modeling of Data warehouses.

Soler and al. (2008) have employed elongation mechanicses furnished by the CWM (Common Warehouse Meta-model) to cover the comparative package and build a star schema, which constitutes protection and verification rules caught throughout the conceptual phase of the Data Warehouse.

Trujillo and al. (2009) have acquired a methodology comprising of four phases: analysis, modeling, execution and establishment, 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 methodological analysis offers all the security requirements throughout the life cycle of the Data Warehouse.

Rodriguez and al. (2011) delivered an UML 2.0 denotation of the activity diagram. This suggestion, called BPSEC (Business Security Process), appropriates to determine a set of security necessities (access control, detection of attack risks, non-repudiation, integrity, confidentiality and security audit), which amends the quality of the business processes examples, and alters to secure a DataWarhouse during its growth, taking into account this necessity.

Blanco and al. (2015) have acquired an automatic MDA computer architecture to assure Data Warehouse; this computer architecture is compiled of a logic model and its translations from the conceptual example using the UML denotation and the CWM computer software. They determined these restraints in the metadata layer that associates the Data Warehouse with the OLAP tools. This proposition comprises of models and translations.

2.1.3. Illations Direction

Triki and al. (2013) advised an example to assure multidimensional information versus the illations in the abstract phase; this access assumes that the Data Warehouse strategy is already designed. It reserves finding both cases of illations:

- Exact Illation: where the assesses of the deduced data are exact.
- Partial Illation: where the data values are partly disclosed, entailing that the user can have an idea of the value of information. This approach comprises of three steps:

Step 1. A field practiced describes the sensible components to protect by questioning the Data Warehouse designer.
Step 2. Construct the graphical record illations from the class plot, assigning the components that are specific or overtone illations.
Step 3. Confront the Data Warehouse with UML notations foregrounding both types of illations.

Blanco and al. (2010) advised an overture based on the state plot to find illations on the aim level. This suggestion centers sore petitions and its developments, but they do not allow inferring the data from the approachable data. The access is delivered as an OLAP protection model of 3 states:

- Static example: demonstrates the Fernandez Medina 2007 UML visibility particular to Data Warehouses, contributing a new kind of dominate appointed Joint Rules, which demonstrates the necessity perquisites to certain combining’s allowing to a specific grammar.
- Dynamic example: or the transactions-state example that has the aim of increasing the static model, in order to assure that discreetness is not endangered by treating the evolutions of the compounding defined with JR done the coating of OLAP functioning.
- Session Assure: This step gets of concern to the user sessions to examine them by assuring each event in order to discover any possible illation.

Sweeney (2002) depicts a real data illation case, using a presentation of the recognition of sore information based on information covering of an policy group, those data guessed anonymous, as well as a voter enrollment list, which appropriated the sensing of the name of the former regulator William Weld and his anamneses, by linking shared assigns. He determined a protection example called k-anonymity admitting support policies that allow debarring illations, which are:
• Classifying rows of the tables to be printed in a random manner in order to not expose sensitive data.
• Avoid having a row with a singular value in the table to be printed.
• Take into account the old interpretation of the information during the structure of the new table.

2.2. The Protection in the Controlling Level of the Datawarehouse

Online Analytical Processing (OLAP) has become growingly a significant factor in conclusion networks. The OLAP host is alleged to furnish access based on authorities determined for each user. He may deny the admittance to information of assess, a proportion, and/or outside a level in a pecking order. Access rights can be explicable assigned on tables / columns of the tables of the Data Warehouses. Nevertheless, the OLAP host alone cannot defend access to forbade data. Explore works has been done to fortify the access rights and authorities of users, and to foreclose any poisonous user to infer prohibited information from data which he has access to. [9] Kirkgoze and al. (1997) determined a assure model for Data Warehouses which comprises in the enlargement of a custom cube with its own proportions and hierarchies. This example is founded on the AMAC direction policy. It is a denotation of the MAC example that specifies the projects that the user can execute allowing to its role within the establishment. The advantage of this example is the tractability in allotting roles to different practical cubes.

Priebe and Pernul (2000) researched the protection troubles in the Goal Design, a design that directs to analyze the desegregation of a distributed information system. During this explore, they have acquired a proper method acting to the OLAP world. This method complies the conventional methodological analysis in the amplification of databases, while containing the multi-dimensional expression during the abstract phase.

Priebe and Pernul (2001) extend their explore on the conception of access control mechanics to assure information concealment, and they have produced an access control chemical mechanism in the form of a language carrying the security-related restraints, during the conceptual phase. This is a language based on MDX MulDimensionnale Xpression.

Eavis and Althamimi (2012) confronted an certification model reposes on algebra particularly planned for OLAP. It is object-pointed and employs query rescript eclipses to assure access to coherent data all around all levels of the abstract model. The process is fundamentally crystalline to the user; a presentment is furnished in the case where a subset of the original request is returned. The outcome is a nonrational and hefty access for the database authentication that is uniquely accommodated to the OLAP field.

Triki and al. (2013) advised an approach path that does not demand extra marching after each feeding phase of the Data Warehouse. It is established on Bayesian electronic network in order to defend a Data Warehouse versus illations; they employ a control module, which attempts to forbid a user to infer husbanded information from the approachable data by using Min and Max collections functions.

2.3. Equivalence and Deduction of Existent Work

We confronted in the former section, explore works that aim figuring out problems related the circumspection of access to the Data Warehouse. Some accesses appear to be applicable and allow for an satisfactory level of privateness but still insufficient. In this division, we confront a relative table of these brings based on criteria, and adopted by a synthesis.

2.3.1. Equivalence

<table>
<thead>
<tr>
<th>Source</th>
<th>Used Approach</th>
<th>Used Technique</th>
<th>Validation</th>
<th>Inferences Management</th>
<th>Data Sensitivity</th>
<th>Traceability of access</th>
</tr>
</thead>
<tbody>
<tr>
<td>[19]</td>
<td>Access permission from the source</td>
<td>SQL grant/revoke</td>
<td>no</td>
<td>Yes</td>
<td>no</td>
<td>No</td>
</tr>
<tr>
<td>[20]</td>
<td>Secure DW from Federated Information Systems</td>
<td>OCL</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>[10]</td>
<td>ACA</td>
<td>UML 2.0</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[22]</td>
<td>UML 2.0</td>
<td>OCL</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[31]</td>
<td>MDA</td>
<td>MDA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[14]</td>
<td>MDA</td>
<td>UML 2.0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[8]</td>
<td>protect a DW against inferences</td>
<td>MDA</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[17]</td>
<td>Authentication Framework</td>
<td>MDX</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[16]</td>
<td>access control mechanism</td>
<td>MDX</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 1 Explore Works Equivalence
The valuations of some significant accesses that cover the circumspection of Data Warehouse for the growth of a assured 4-dimensional example, is confronted in Table 1 allowing to several criteria are:

**Employed access:** the name of the aimed access.

**Used proficiency:** the technologies employed in the aimed approach.

**Transmutation among examples:** growth of a logic example of ED from an abstract model taking account protection in every example.

**Illations direction:** detection of feasible decisions from the approachable data.

**Validation:** carrying out of the aimed solution.

**Data dynamical sensibility:** automatic conclusion of data sensibility level.

**Traceability of accession:** to trace the user deals on the Data Warehouse for analytic thinking and decision making.

2.3.2 Synthesis

Data Warehouse auspices versus illegal immigrant accession was felt outside sensible question since various years [10] [16] [17]. According to the authors [25], model the accession control of the Data Warehouse is the march of construction an abstract model that demands to be stored in the Data Warehouse. This example is a delegacy of the realism or a part of the realism. Following the analyze of existent works, we found the following:

- The circumspection of the Data Warehouse has traditionally been conceived in the authoritative carrying out of a Data Warehouse[16] [17], nevertheless the latest works [14] [18] believe its inclusion body in the growth levels which can develop more full-bodied lineament results, and the system can adapt these security demands in a more innate way.

- In the explore work [5], the authors aimed an reflex MDA architecture established on the UML denotation to assure a Data Warehouse, but the doubtfulness that develops in this case is: What is the use of precisying the protection level in the user visibility as well as for the elements of the Data Warehouse since the licenses are defined.

- The most of explore works, particularly those demanded in the abstract modeling phase relied on the CWM meta-model, in order to acquire a secure Data Warehouse. Knowing that the CWM example is based on three criteria, namely UML, MOF and XMI to decent constitute all protection and audit dominates defined in the abstract modeling of Data Warehouses.

- Most of the acts are modeling the access control founded on MAC and RBAC policies, while the user profile is believed a detached table that comprises the requirement data for a static access of a user without appreciating the antecedences of the user documented. An MDA computer architecture for a reflex secure conception of a Data Warehouse is applied in [14, 13], but the two accesses were unable to realize the safety rules that are complex.

- Although the permits present the main axis to en-sure secret access to the warehouse, however, the absence of a standard that supports the accuracy of these permissions can cause incompatibilities and illations as consequences. In this sense, we find the work of [2], which proposed the use of the permits schema de-fined for federated databases without any modification to build a secure Data Warehouse, and [3] that propose revising requests in order to verify that they comply with the limitations determined in the sources.

- Some writers [19] [20] aimed to make the access assure model in the Data Warehouse, from data beginnings, while others [16] [26] have conceived this suggestion as difficult since the source data arise from unlike systems (with different policies). And functional systems use the relational model while the OLAP organizations use the multidimensional example.

- Also note that the conception of illation was cited in several works as a necessity element to ensure circumspection, and whose mastery is essential. In this sense, there is the work of [1] which advised an access that can detect partial and precise inferences, [4] aim an approach based on the state-diagram to detect illations in the abstract phase without considering the possibility of illations from accessible data. However, despite the high risk of inferences, it is not sufficiently taken into account in the abstract phase.

We determine that majority of the explore works impress the task of relegating data according to their level of sensibility (very sensitive, sensitive, and confidential) to the data owner. Experiencing that allowing to the role of the user, the information proprietor allots a level of information sensibility in order to access data having the same level of sensitiveness or lower. The owner of the Data Warehouse may then allot a lower level of sensibility to vital information. This consequence, however, a trouble of information discreetness loss. In addition, the licenses defined in the origins are not sufficiently overworked to help the owner well decide the permits of a Data Warehouse user. In the next part, we propose an advance that defeats these restrictions.

III. ACTIVE DIRECTION OF PROTECTION CHARGES

3.1. Need

Circumspection of a Data Warehouse is established on the accession assure example that defines the licenses allowed to each exploiter. Allowing analyzing mentioned in the state department, accession license to an object of the Data Warehouse is allowed to a user allowing to his function. The sensibility level attains data approachable to the user, with a predisposition level defined on the object of Data Warehouse allowing to its function. Such position is hard to deal
by the data proprietor, who can assign a not proper level of sensitivity to an object of Data Warehouse. What may abuse its discreetness? For this we aim a different way of defining user licenses by using the licenses determined in the data sources such as suggestions that can help the data proprietor. Then our accession control model can generate the sensibility level of each object in the Data Warehouse established on these licenses. This classification will help us to trace user actions on sore data. In the difference of this section we demonstrate the computer architecture of our contribution, and then we detail our contribution which comprises of two levels.

3.2. Architecture

Figure 1 General Architecture Aimed for Controlling Access to Data warehouses Based on User Visibilities

In this discussion section we demonstrate the architecture of our proposal which its composite plant of three parts (Figure 1):

1. **Active categorization of Data Warehouse predisposition degrees founded on user visibilities** Compares of determining the access licenses of each user allowing to his role, and giving the sensibility level of information.

2. **Discovering illations by combine several visibilities** Appropriates discovering the sensible compounding of licenses that can acquire illations, for a user that aggregates multiple functions.

3. **Dealing visibilities from uses** Appropriates to deal the approach to the Data Warehouse entertained in the Cloud Computing. The documentary is to downplay the dealings, and gain the functioning of the discussions.

In this research paper, we demonstrate our access for the first part of our computer architecture which comprises in demonstrating a DAC model (Dynamic Access Control) founded on the PPU (Profession Profile of a User) established of several mechanisms (Figure 1):

**Appointment Direction of licenses:**

This is to determine the permissions of a role on a information with a given assigned perquisite, taking into account the licenses determined on the data sources level. And advise them to the executive, information owner of the ED, when as-signing licenses.

- **Information categorization mechanics:** Automatic generation of data sensitivity levels based on defined permissions.
- **Traceable of sore information:** Is a mechanics to trace user activities on the information with a high level of sensitivity.
- **Sending alerts Mechanics:** Depending upon user access trackable mechanism, our alert organization can send presentments to the executive during a sought violation of license on sensitive information.

This component of our computer architecture, that appropriates dealing dynamically of sensibility degrees of Data Warehouse founded on user profiles, is presented in a meta-model form. It’s an extension of meta-model CWM (Common Warehouse Meta-model) and presents the user’s profession profile. It comprises five classes conceived the core of our part.

3.3. Active Direction of Data Warehouse Sensibility Degrees Founded on User Visibilities

The following is the classification of our work.

The license is an aboriginal axis in an access assures mechanism, the direction of these licenses is an unmanageable task for the executive. In this level, we aim to use the permits determined in the data sources as hypnotisms that will help the information proprietor to well define the permission of a role on an object in the Data Warehouse allowing to a given privilege $P(R_{i}, P_{rj}, O_{k})$ where
Each function belongs to at a hierarchical degree:

- **P**: License (0, 1).
- **Ri**: Function of the user.
- **Prj**: Perquisite license (Read, Write, Modification).
- **Ok**: Objective of the Data Warehouse (Table fact, dimension, column, column value).

According to our meta-example, each aim of a Data Warehouse is confronted by the class "ObjectDW" which can be a conception, a proportion, a base, a column or a value of a column. In order to help the owner to decide the permits of an "ObjectDW", our system suggests the source permissions of an "ObjestDW" to help the owner in the determination licenses phase of the Data Warehouse. The "Object Source" class demonstrates the comparable object of an "Object DW" in the "Source" information for each role source "Role Source" by considering the licenses conceded in the source "Source Permission".

### 3.3.1 Active Propagation of Data Sensibility Degrees

The sensibility degree of an object o is ascertained by the number of functions that have license to read it, relative to the total number of functions in each hierarchal level. In order to mechanically give a percentage that demonstrates the sensitivity level of object to, our system is based on the following rules:

- Each function belongs a hierarchical level Well-defined N = {1, 2,...,n} Where P is the total count of hierarchical degrees.
- Each hierarchical degree has a constant C[N]=N
- Each function belongs a hierarchical level H={h1,h2,...,hp}
A function may acquire access rights of another role.
The objectives in the Data Warehouse are grouped in the set \( O = \{o_1, o_2, \ldots, o_k \} \)
\( \{R_i, Pr, Ok\} = \{0, 1\} \), With Prj a audience.
The object conferred by several roles is less sore than the object that is got at by a small number of roles. The sensitivity level is calculated as a percentage.

The relationship that appropriates computing the sensibility level is as follows:
\[
\text{The sensitivity level} (\%) = \left[ 1 - \frac{\sum R_i (O_k)}{\sum R_f} \times C_i \right] \times 100
\]

- \( \Sigma (P(Ok)) \): The amount of the functions in a hierarchal level \( i \) which have the right to consult an object.
- \( CT = \Sigma C_i \): The amount of the hierarchical levels constants.
- \( \Sigma R_i \): The amount of functions in a hierarchical level.
- \( Ci \): The Constant of a hierarchy Level \( i \).

The following algorithm demonstrates the advised method for determining the sensitivity degree of each object in the Data Warehouse:

**Data warehouse Algorithm Classification**

**Input:** \( P \): Total number of Hierarchical Levels
\( OT \): Total number of warehouse objects

**Output:** \( SE[] \): Table of sensitivity levels

**Begin**

```java
double[] SE = 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=newdouble[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++;
}
while(h<=P){
    for(int k=1; k<OT+1; k++){
        SE[k]=1-(PP[h][k]/(double)CT))*100;
        System.out.println("Sensitivity level of the object \["+k+"\] is \"+ SE[k] + "%\";)
    }
    h++;
}
return SE;
**End.**
```

The loop one of our algorithmic programs appropriates to compute the constant of each hierarchal level which is the heightening order count of the levels. The loop two employs the function "Permissions (h, k)" for determinant the number of functions that have license to read an object \( O \), manifolds by the hierarchical degree constant. And the loop three finds a percentage that confronts the sensitiveness level of the object \( k \) as depending upon the number of functions that have permission to read the object, the hierarchal degrees of the roles and their constants.

### 3.3.1.1 Option of Thresholds

Each aim of the DataWarehouse has a threshold. This is a varying argument (accommodation cursor of the predisposition level), which depends upon the context of the accompany such as:
• Action type.
• Flow (crisis, war).
• Consequences (internal or external).
• Business strategy

The sore information of an aim of the DataWarehouse is the data whose sensibility degree is greater than or equal to the threshold of this object.

3.3.1.2 Trackable and Alarm

In order to describe user accomplishes on sore information, and send alarms to an information owner to inform them of efforts assault licenses, our organization is founded on sensibility degrees fathered. Thus, tractable appropriates deciphering the accomplishes of the exploiters on the sore information, whose sensibility level outmatches the doorway defined by the data owner. Then our organization sends an alert to the proprietor if the action is an attempt to break the licenses. The figure 6 clarifies how our establishment treats the user’s request:

![Diagram](image_url)

Figure 6 Trackable and Alarm

If the requested information is appropriated, 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 an appraisal comprising the details of the attempted violation of access licenses, in order to react. And of course access will be refused.

IV. CONCLUSION

Because of the vital data stored in the DataWarehouse, it is significant to assure its secrecy. The circumspection in the circumstance of DataWarehouse is conceived an significant necessity, which must be assured by an authority direction mechanics which is the stipulation and executing of access corrects in the informations in general and more specifically in Data Warehouses. In this analyze appropriated us to see the major troubles of the circumspection of the information in the warehouse. To reduce these dangers, we have aimed a answer based on the user profile, which comprises in the resolution of access licenses allowing to the user role using the access rights determined in the sources, generate the level of sensibility of each object in the DataWarehouse allowing to these licenses, trace the access and discover violation efforts of access rights on a sore information (a data with a high level of sensitivity). The objective of this answer is to bring down the exposure of the data in a DataWarehouse, and assist the owner of the DataWarehouse to well deal the access assure of the users.

REFERENCES


