# DW Access Control Model

**Attia El Sayed**[†], **Ali El Bastawissy**[††], **Ibrahim El Imam**[†††]
[†]Collage of Computing& I.T., Arab Academy for Science, Technology & Maritime Transport, Cairo, Egypt
[††] Prof. Dr., Faculty of Computers & Information, Cairo University, Cairo, Egypt
[†††] Prof. Dr., Collage of Computing& I.T., Arab Academy for Science, Technology & Maritime Transport, Cairo, Egypt

**Summary**
Data warehouse collects and integrates critical data of organization from multiple sources and stores them for a long time in multidimensional model. This nature motivates researchers to propose set of models to secure these data. In this paper, we present an enhanced authorization model in order to close open security holes in Data Warehouse (DW) and On-Line Analytical Process (OLAP). We focus on adding much flexibility to formats of security rules besides controlling access to combination of fact and dimension to grant user the required permissions to perform his work only. As an integral part of this model, we present new methodology to avoid bad security impacts which are result from using Slowly Changing Dimension (SCD) techniques which track time-variant dimension modifications.

***Keywords:***
*Security, Data Warehouse, OLAP, SCD Techniques*

## 1. Introduction

Great efforts are directed to data warehouse field during last period, Data Warehouses (DW) contain historical, consolidated, and summarized data to support business decision systems at many levels. It is considered the central part in the decision support systems. It provides rapid responses to iterative complex analytical queries. The complementary part of data warehouse within any decision support system is (OLAP) which enables data warehouse to be used effectively for online analysis. It provides data aggregation techniques which organize and summarize large amounts of data which is stored in data warehouse. Data warehouse contains sensitive data which motivates researchers to do efforts in order to empower access control in the data warehouse environment.
In this paper, several data warehouse security models are referenced with indication to benefits and drawbacks of them. These drawbacks vary among low flexibility in security rules formatting, missing ability to present security rules in different DW building phases or inappropriate consideration for objects in the security rules. Therefore, new security model is proposed to solve these issues and satisfy more security requirements.
An attempt to propose new mechanism to save the user accessibility on entity (e.g. dimension or fact) or couple of entities (e.g. fact and dimension) in the data warehouse in case of using (SCD) techniques for the first time to the best of our knowledge.

Manuscript received October 5, 2010
Manuscript revised October 20, 2010

## 2. Related Work

Many security models are proposed for data warehouse and OLAP. For example, the authors propose model [1] which focuses on access and security management in OLAP and N-dimensional cube. This model intends to generate security profile which contains security restrictions for each role in date warehouse. Each user in data warehouse has a role to perform his work. According to his role, user can retrieve data from data warehouse. Set of advantages are accompanied with this model such as: applying close-world assumption which means access permission is not allowed until it is granted explicitly; ability to grant read to specific values and level of details in dimension. Although this model is tailored specifically to control security in data warehouse and offers the preceding advantages, it has many drawbacks which prevent satisfying all information security needs in data warehouse. It cannot control user access to coupled dimension and fact among multiple dimensions. This drawback results in miss the ability to permit measure within fact according to linked dimension and vice versa. Predicate clause should be flexible enough to contain dynamic and complex structures.
In 2006, new model for DW security is presented [5]. The authors of it aimed to propose model which allows specifying security rules in conceptual multidimensional modeling phase. Moreover, audit rules are added to analyze user behaviors. Additionally, Unified Modeling Language (UML) extension is presented to design secure multidimensional model. Set of remarkable advantages is presented in this model such as: providing mechanism to represent security requirements at conceptual level; authorization rules consider couple of fact and dimensions; applying row level security; flexible representation of subjects. These great advantages do not prevent from existence of some defects. No methodology to apply these rules in the next phases of data warehouse building process. There is no way to permit or prevent specific classification hierarchy level. Presentation of these rules is joined with using UML in building conceptual model for data warehouse.

In 2008, new model is presented to provide a middleware-enabled policy-based framework that applies access control policy to both base tables and materialized views [7]. Combining the fast response of materialized view and controlling user access is a great advantage of this model. Some missing issues are highlighted such as: missing ability to control access to hierarchy levels; nodes which are in the same level cannot share data; auditing mechanism is not considered.

## 3. Motivating Example

To illustrate the security needs for data warehouse, suppose there is data mart model appeared in the figure 1. The requirement is enabling user "Ahmed" to perform the following tasks:

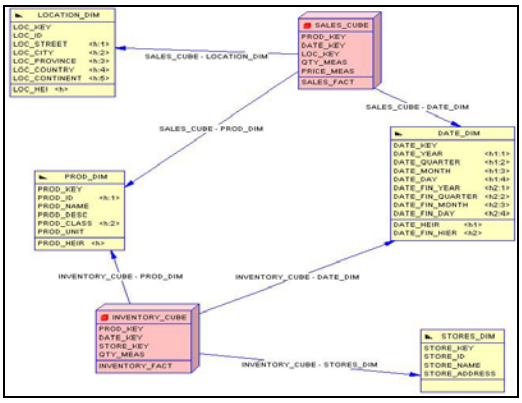


Figure 1: Data Mart Model

- Retrieving all details about products.
- Retrieving total "quantity" of sold products.
- Retrieving total "quantity" of stored products.
- Retrieving total "quantity" of each product that was sold from "Egypt" in financial year.
- Retrieving total "quantity" of each product class that was saved in store called "North".

## 4. DW Access Control Model

DW Access Control Model (DWACM) is an extension to OLAP security model. It applies close-world assumption to prevent access of unauthorized user until he is granted explicitly. It presents new set of much flexible predicates to control user access in DW and OLAP. These predicates are stated in terms of subjects (active entities that will apply operations on objects), objects (passive entities that will receive subjects operations) and actions (operations that will be applied on objects). They consider READ operation as the permitted action as it involves all other OLAP operation. They intend to add security restrictions based on the association between FACT and DIMENSION besides implementing them on single entity. They permit access on multiple hierarchy levels and attributes based on the connected fact and access on measures according to linked dimension. The formats of these predicates are:

- SimpleFactPredicate SFP(S,A,F,Meas[],C) where S is subject, A is action, F is fact, Meas[] is set of accessible measures within F and C conditions used to permit set of data stored in F.

- SimpleDimensionPredicates SDP(S,A,D,Att[],C) where D is dimension, Att[] is set of accessible attributes within D.

- JoinAttributeMeasurePredicate JamP(S,A,F,D,Att[],Mea[],C) which permits access on Att[] within D, Mea[] within F in case of enquiry linked F and D together.

- JoinPredicate JP(S,A,F,D) which is special case of the JamP predicate to permit access on coupled F and D together involving all measures and attributes.

The following notes should be considered:

- C ➔ can be inserted as simple or complex conditions to filter data or it can be inserted as NULL to permit all data stored in the object.
- Meas[] ➔ can be replaced by ALL to grant access permissions to all measures within mentioned fact in the predicate.
- Att[] ➔ can be replaced by ALL to grant access permissions to all attributes within mentioned dimension in the predicate.

These new predicates strengthen multiple drawbacks appeared in OLAP security model and present much effective model to satisfy required security requirements. These predicates provide the following added security flexibilities:

- User can be authorized to use some dimensions with and only with certain facts
- User can be allowed to access the hierarchy level of aggregation in dimension according to the connected fact
- User can be permitted to access the measure within fact based on the coupled dimension
- Security policy can be defined during any phase of DW building process (Conceptual, Logical or Physical)

These predicates are classified into two categories each of them has two types of predicates:

1. Simple Predicates
    a. SimpleFactPredicate SFP(S,A,F,Meas[],C)
    b. SimpleDimensionPredicate SDP(S,A,D,Att[],C)

2. Joined Predicate
    a. JoinAttributeMeasurePredicate JamP(S,A,F,D,Att[],Mea[],C)
    b. JoinPredicate JP(S,A,F,D)

This categorization is very helpful and fruitful in defining conflict resolution policy. This policy intends to determine user authorizations according to entered query type and the set of granted predicates to user. One of the following two scenarios is followed:

1. If the user enquiries single entity fact or dimension, security checks are performed according to simple predicates of this user only.

2. If the user enquiries connected fact and dimension, security checks are performed according to joined predicates of this user only.

For example, suppose the following predicates are granted to user Hassan:

SFP(Hassan,READ,SALES_CUBE)
JP(Hassan,READ,INVENTORY_CUBE,PROD_DIM )

Then there is query entered by Hassan against SALES_CUBE only without any connected dimensions, all stored data are retrieved to him. But, if Hassan executes query against couple of SALES_CUBE and any dimension, he is not authorized to retrieve any data. This restriction appears as a result for first predicate. The second predicate results in if the query is performed against a couple of INVENTORY_CUBE and PROD_DIM, all data are retrieved to Hassan. But, if Hassan executes query against INVENTORY_CUBE or PROD_DIM separately, he is not authorized.
Proceeding, two notes are highlighted regarding conflict and resolution policy. 1) All predicates related to the same person are complementary only if they belong to the same predicate category. 2) Last format of predicate overwrites all previous predicates from the same category for the same object or join.
Referencing to the requirements mentioned in section 3, important notes are appeared regarding two conformed dimensions “PROD_DIM” and “DATE_DIM”. These notes are:

- Access permissions to aggregate data on level of PROD_ID from dimension “PROD_DIM” in case of its association with “SALES_CUBE”.
- Access permissions to aggregate data on level of PROD_CLASS from dimension “PROD_DIM” in case of its association with “INVENTORY_CUBE”
- Access permissions to measure “QTY_MEAS” within “SALES_CUBE” in case of its association with “PROD_DIM”
- Access permissions to retrieve data from dimension “DATE_DIM” in case of its association with “SALES_CUBE”.

By using DWACM, these security requirements with consideration for preceding notes are satisfied by generating the following predicates:

- SP(“Ahmed”,READ,PROD_DIM)
- SaP(“Ahmed”,READ,SALES_CUBE,QTY_MEAS )
- SaP(“Ahmed”,READ, INVENTORY _CUBE,QTY_MEAS)
- JamP(Ahmed,READ,SALES_CUBE,PROD_DIM,[PROD_ID,PROD_NAME,PROD_DESC], [QTY_MEAS],NULL)

  JamP(Ahmed,READ,SALES_CUBE,LOC_DIM,[LOC_COUNTRY],[QTY_MEAS],LOC_COUNTRY=“EGYPT)

  JamP(Ahmed,READ,SALES_CUBE,DATE_DIM,[DATE_FIN_YEAR],[QTY_MEAS],NULL)

- JamP(Ahmed,READ,INVENTORY_CUBE, PROD_DIM,[PROD_CLASS],[QTY_MEAS],NULL)

  JamP(Ahmed,READ,INVENTORY_CUBE, STORES_DIM,[STORE_NAME],[QTY_MEAS],STORE_NAME=“NORTH”)

However dimension table attributes are more stable and static, many dimensions are still subject to change more slowly and unpredictably. This issue forces to track time-variant attributes to satisfy business requirements. There are three fundamental techniques for handling these modifications. First technique intends to replace the old value of the attribute with the new value. Second technique adds new row with new value and generating new surrogate key to this new record. Third technique intends to add new attribute to save old value and writes the new value in the prior current attribute [6] [7].
The implementation of these techniques may limit the accessibility of authorized users. To explain the problem, suppose we have the following values in PRODUCT dimension:

| Key | Name | ID | Class |
|---|---|---|---|
| 1 | tea | 10 | Drink |
| 2 | coffee | 11 | Drink |
| … | … | … | … |

PROD_DIM values

There is access permission is granted to user "Ali" using the following predicate:
SDP(Ali,READ,PROD_DIM,[NAME,ID],"Name='tea'")

The value of attribute Name of first entity is changed into "ice tea". This modification changes "Ali" to unauthorized user to retrieve this data in case of using any SCD techniques as appeared below.

Type1: Overwrite Dimension Attribute

| Key | Name | ID | Class |
|---|---|---|---|
| 1 | ice tea | 10 | Drink |
| 2 | Coffee | 11 | Drink |
| … | … | … | … |

Type2: Add New Dimension Row

| Key | Name | ID | Class | From | To |
|---|---|---|---|---|---|
| 1 | tea | 10 | drink | 10/5/ 2001 | 11/7/ 2002 |
| 2 | coffee | 11 | drink | 15/5/ 2001 | |
| 3 | ice tea | 10 | drink | 12/7/ 2002 | |
| … | … | … | … | … | … |

Type3: Add New Dimension Attribute

| Key | Name | Name1 | ID | Class |
|---|---|---|---|---|
| 1 | ice tea | Tea | 10 | Drink |
| 2 | Coffee | coffee | 11 | Drink |
| … | … | … | … | |

In order to avoid these security impacts, new methodology is proposed. It intends to generate new list of predicates which is adequate to the new value of the attribute. It replaces the old one automatically after receiving confirmation of data warehouse security administrator. Based on the used type of SCD technique, the format of the new list is determined as described in the following subsections.

*SCD Type1 & SCD Type2*

The effects of using type1 and type2 SCD techniques are avoided by formulating new condition C1 that includes the new value. This new condition overwrites the old condition in the following predicates:

SDP(S,A,D,Att[],C1)

JamP(S,A,F,D,Att[],Mea[],C1)

So, the predicate of our example will be:
SDP(Ali,READ,PROD_DIM,[NAME,ID],"Name='tea' or Name='ice tea'")

*SCD Type3*

The effects of using type3 SCD techniques are avoided by formulating new condition C1 that includes the new value and granting access to new attribute if it has OldValue. These modifications applied on the following two predicates:

SDP(S,A,D,Att[],C1)
JamP(S,A,F,D,Att[],Mea[],C1)

So, the predicate of our example will be:
SDP(Ali,READ,PROD_DIM,[NAME1,NAME,ID],"Name='ice tea' or Name1='tea'")

Proceeding to generating the new list of predicates, the following algorithm is used to apply this new list. This algorithm states:

**IF** *modified Attr. is used in the condition* **THEN**
  **IF** *new value is not permitted* **THEN**
    **IF** *old value is permitted* **THEN**
      - *Generate new list of modified predicates according to the type of SCD technique*
      - *Send this list to security administrator for approval*
      **IF** *security admin approved* **THEN**
        - *Replace the old predicates with new predicates automatically*
      **END**
    **END**
  **END**
**END**

Algorithm 1

## 5. Conclusion

In this work we enhance basic OLAP security model to control access rights to combination of linked fact and dimension by adding much flexible predicates. These predicates provides administrator with required flexibility that is used to specify the permitted attributes and hierarchy levels according to a coupled fact. On the contrary, permitted measures can be determined according to the connected dimension. They can be defined in any DW building phases. They are categorized simple and joined types; each of them

includes two predicates. This classification leads to introduce new concept for conflict resolution in order to determine user permitted actions based on the category of predicates and the type of entered query. Finally, new methodology is proposed to avoid missing user's accessibility on the permitted objects in case of using any SCD technique. This approach intends to modify the implemented predicates to be adequate with new values of the modified attributes according to the implemented SCD technique.

**Acknowledgment**

I want to present cordial thanks to Allah, my family, my friends, my supervisors and everybody encourages me to accomplish this research.

## References

[1] Remzi kirgoze, Nevena Katic, Mladen Stolba, A Min Tjoa. A Security Concept for OLAP. IEEE Computer Society, 1997.
[2] Edgar Weippl, Oscar Mangisengi, Wolfgang Essmayr, Franz Lichtenberger, Werner Winiwarter. An Authorization Model for Data Warehouses and OLAP. Workshop on Security in Distributed Data Warehousing, 2001.
[3] Arnon Rosenthal, Edward Sciore. View Security as The Basis For Data Warehouse Security. 2nd International Workshop on Design and Management of Data Warehouse, 2000.
[4] E. Fernández-Medina, J. Trujillo, R. Villarroel, M. Piattini, Access control and audit model for the multidimensional modeling of data warehouses. Decision Support Systems, 2006.
[5] Ralph Kimball. The Data Warehouse Lifecycle Toolkit, Second Edition. John Wiley & Sons, 2008
[6] W. H. Inmon. Building the Data Warehouse, Fourth Edition. John Wiley & Sons, 2005
[7] Rafae Bhatti, Dengfeng Gao, Wen-Syan Li. Enabling policy-based access control in BI applications. Data&Knowledge Engineering, 2008.

**Attia Elsayed B.Sc.** of Computer Science, faculty of computers and Informatics, Zagazig University 2005. This paper was presented to obtain Master degree in computer science.

**Ali Bastawissy** IS Prof. Doctor and Director of the Centre of Studies Development of computers and information Systems, Cairo University.

**Ibrahim F. Imam** Prof. Doctor, Department of Computer Science, Collage of Computing& I.T., Arab Academy for Science, Technology & Maritime Transport, Cairo, Egypt.