An Authorization Model for Data Warehouses and OLAP

Edgar Weippl, Oscar Mangisengi, Wolfgang Essmayr, Franz Lichtenberger, Werner Winiwarter
Software Competence Center Hagenberg
Hauptstr. 99, A-4232 Hagenberg, Austria
E-Mail: edgar.weippl@scch.at

Keywords

Security, Authorization, Data Warehouse, OLAP

Abstract

In this paper we discuss a basic authorization model for data warehouses and OLAP offering greater expressiveness and highly increased usability with respect to security. Based on a simple formal notation we will show that access privileges can be expressed more intuitively than using SQL's grant statements.

1. Introduction

Data warehouses (DWHs) contain historical, consolidated, and summarized data for supporting business decision systems at many levels. Providing rapid responses to iterative complex analytical queries, On-Line Analytical Processing (OLAP) enables DWHs to be used effectively for online analysis. OLAP's multidimensional data model and data aggregation techniques organize and summarize large amounts of data to facilitate quick evaluation using online analysis and graphical tools.

A DWH containing vast amounts of information can be a considerable asset if tools can extract the required knowledge. Companies may want to sell aggregated data to other companies or share it with its subsidiaries. Therefore security, currently too frequently neglected, should be brought to the attention of both DWH administrators and managers who decide on who should have access to what data.

Despite the fact that many DWHs are implemented on relational databases (DBs) that already offer a wide variety of security mechanisms such as authentication, access control and auditing, it is not trivial to correctly grant permissions to the underlying tables. The reason why it is such a demanding task to set the access privileges accordingly is that relational DBs supporting SQL have not been primarily designed for DWHs and thus SQL's grant statements do not allow defining access privileges intuitively. Moreover, it is well-known that security features in information systems are particularly prone to user errors during configuration.

In this paper we, therefore, elaborate on access privileges in the specific context of DWH and introduce a formal notation that addresses SQL's shortcomings. Building on our proposal of a more expressive language it is easy feasibly to implement access privileges for a given security policy. Readers may doubt whether yet another authorization model is really necessary. However, we think that the greater expressiveness and the highly increased usability justify the effort, especially because we envision a tool to translate the DWH authorization language into SQL statements – similar to, though less formal than Rosenthal's approach described in [BR01].

2. Related Work

Recently, a number of security models have been proposed for data warehouse and OLAP. For example, [KKST97] propose a security model based on mandatory access control for OLAP-cubes. They define security constraints for each role in the data warehousing environment, so that each role defines a sub-cube of the data warehouse's N-dimensional cube. An advantage of this approach to handling security is its flexibility of assigning roles to different virtual sub-cubes.

A model for data warehouse security based on metadata is presented in [KQST98]. The authors elaborate on requirements of and impacts on the

selection of an adequate security model for a data warehouse environment.

Building upon on a similar mindset [Bhar00] focuses on ideas that can contribute to warehouse security. The concepts take common operations in to account such as replication control, aggregation and generalization, ex-aggregation and misleading, anonymity, and user profile based security.

[RS00] suggest to base data warehouse security on view security. This paper aims to provide a theory that permits automated inference of many permissions for the warehouse, in a way that minimize both the learning curve for administrators and the amount of new software that vendors would need to implement.

Moving from theory to practical challenges [PP00] focus mainly on the technical issues such as authorization and access control. They explore these security issues in the context of the GOAL project; they compare the implementation in commercial ROLAP (relational OLAP as opposed to MOLAP = multidimensional OLAP) based products such as Microsoft SQL Server 2000, MicroStrategy, Cognos PowerPlay, and Oracle Express.

However, to the best of our knowledge there is no access control model that focus specifically on expressiveness and usability in the context of DWHs.

3. Motivating Example

To illustrate the security concepts for data warehousing we use the data warehouse star schema model as follows.

We classify users based on the required roles:

- All users want to project unit sales of a product for a store and for a day. In this case all users can only access the detail data or drill-through of OLAP operation.
- Sales managers want to find out the total of unit sales and total of revenue generated by each product, subcategory, and category in the dimension product for a month, quarter or a year in the dimension time for a store. In addition, they can access measures, such as unit sales and total revenue, of the fact sales. They have privilege for OLAP operations, such as drilling-through, rolling-up, drilling-down.
- Distribution managers compare and find out unit sales in different stores at the city, regional level for each product, subcategory, and category and for month, quarter, and year. They can apply OLAP operation such as rolling-up, drilling-down, slice and dice, drilling-through, and measure (e.g., sales).
- A director can compare and find out unit sales and revenue for all levels of dimension hierarchies until the ALL hierarchy level. A director can access all OLAP operations.

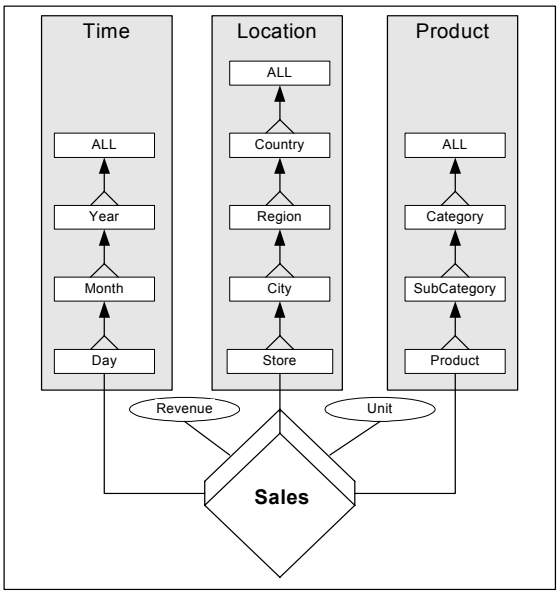

Figure 1: A typical 3-D data cube

4. Authorization Model for OLAP

Authorization is the granting of a right or privilege, which enables a subject to have legitimate access to a system's objects or a system itself. Subjects represent active entities usually acting on behalf of principals such as users whereas objects are considered to be of passive nature such as, for instance, a database table, view, or any other object that can be created within the system.

An authorization model is typically stated in terms of subjects, access types, objects, and predicates. Due to space limitations imposed on this paper we will ignore predicates, which allow the definition of content- and context-based access restrictions. We will elaborate on these topics in reverse order, starting with objects.

Unlike in operating systems (OSs) or traditional databases the *objects* that are to be protected are not files or tables but the *dimensions* of the multidimensional cube, the *hierarchies* within one dimension and, obviously, the *facts* (frequently also referred to as measures) themselves.

In contrast to the read/write/execute privileges known from OS security, DWH basically only requires read access for users. There are, however, different operations specific to OLAP on which we feel access control should be based. Of these, six *actions* such as *read*, *drill-down*, *roll-up*, *slice*, *dice* and *drill-through* are commonly agreed to be fundamental for DWHs. Read privileges are obviously needed to read a specific fact; drill-down rights allow viewing data in greater detail, i.e. to split data within one dimension according to the hierarchy within the dimension in question, whereas the roll-up operation is the opposite. Slicing the multidimensional cube along one dimension or viewing sub-cubes (dice) may also reveal additional insight into the data. Finally, drill-through refers to the operation of accessing the original OLTP data upon which the DWH has been built.

For the sake of simplicity we will assume that *subjects* are simply users and not roles. However, the extension to a model for role-based access control seems to be straightforward. Nonetheless we want to emphasize the duality between the hierarchy of dimensions in the DWH and a role hierarchy of DWH users. A simple example clearly highlights this fact. The probably best-known introductory example (see section 3) in data warehousing is that of a retail chain; the three dimensions used are (1) location, (2) time, and (3) product. Considering the roles of managers one would probably distinguish between *location* and *product* managers. Location managers could be regional managers for individual states or on a higher level country managers. Similarly, product managers could either be managers for individual products or product groups.

As our authorization model supports fundamentally different objects not all actions can be performed on each object. For instance, facts can only be read, whereas on objects like dimensions all types of access can be supported.

As mentioned before, we propose to specify access privileges for users of a DWH and do not consider which access rights these users have in the source DBs. For one operation, however, the DWH access control mechanism will have to refer to the source DB. As the drill-through operation does not really involve data processed in the DWH but is instead propagated to the underlying database, the local authorization scheme will obviously be decisive when considering whether a specific read request is permissible. A more detailed exposition of this special case lies, however, not within the scope of this paper. Ideas of solving this problem might be found in [RSD99].

We formally specify our simplified authorization model for DWHs (only containing the access types read, drill down, roll up, and slice) based on the following (finite) sets:

S … set of subjects

F … set of facts

D … set of dimensions

T … set of allowed operations; in our case T = {read, drill-down, roll-up, slice}

For every dimension $d \in D$ we have a set Hd of hierarchy levels of dimensions (e.g. year, month, etc.) with $ALL \in Hd$ for every $d \in D$.

For every operation $t \in T$ and dimension $d \in D$ we have a set $P_{t,d}$ of predicates applicable to t in dimensions d. We do, however, not elaborate on the structure of these predicates but we want to state that the mere noun of a hierarchy level of dimensions like country is regarded a valid predicate.

Thus, an access record (s, F', DA) is a triple with $s \in S$, $F' \subseteq F$, $F \neq \{\}$, and DA a total map that assigns to every dimension $d \in D$ a nonempty set of pairs (t, p) with $t \in T$, $p \in P_{t,d}$ such that this set

$$DA : D \rightarrow (T \overset{\text{partial}}{\rightarrow} P)$$

$$\text{with } P := \bigcup_{\substack{t \in T \\ d \in D}} P_{t,d}$$

$$DA_d \neq \{\} \text{ for all } d \in D$$

$$DA_d(t) \in \text{ for all } d \in D, t \in T$$

of pairs forms a (partial) map, i.e.

In addition to the previous specification we apply the rule of a *closed world* assumption so that operations are not permissible unless explicitly granted. As there are no negative authorizations in our model, contradictions are not possible. Additionally, the evaluation whether an access operation should be permitted is rather efficient. Moreover, overlapping sets of authorization have

no adverse effect on our model because negative permissions are not allowed.

We will use the previous example (see Section 3) to illustrate authorizations for DWHs. For instance:

```
(Alice, {Unit},
        {(Time, {(Read, ALL), (Drill-Down,
        Month)}),
        (Location, {(Drill-Down, Country),
        (Roll-Up, Region)}),
        (Product, {(Drill-Down, Category),
        (Roll-Up, Product)})
        }).
```

This allows Alice to access all unit *sales* but not the *revenues* of each product, subcategory or category. However, she would lack privileges to retrieve aggregated data on per year or per day basis and would neither be permitted to access data for individual cities or stores.

Adding the permission set

```
(Alice, {Unit, Revenue},
        {(Time, {(Drill-Down, Year)}),
        (Location, {(Drill-Down, Country),
        (Roll-Up, Store), (Slice, City =
        "Linz")}),
        (Product, {(Drill-Down, Category),
        (Roll-Up, Product)})
        }).
```

to Alice's permissions would give her full access to data of stores located in the city of Linz. Given these two sets of permissions she can fulfill her job supervising stores located in Linz and compare their performance to other stores in aggregated form.

5. Conclusion and Future Work

In this work we develop a basic authorization model for data warehouses and OLAP offering greater expressiveness and highly increased usability with respect to security. We logically separate the authorization model into a subject and an object model, the latter being the center of interest of that paper. The object model contains the elements of a data warehouse schema, i.e. dimensions, hierarchies, and facts as well as their OLAP operations including read, drill-down, roll-up, slice, and dice. With respect to the subject model we only refer to users but motivate the correlation between role- and dimension-hierarchies. We furthermore sketch an informal authorization language illustrating how to intuitively formulate access restrictions for data warehouse data and OLAP operations especially when compared to standard SQL's grant statements and formally specify the corresponding simplified authorization model. [BR01] illustrate an approach how to automatically translate arbitrary security policies into SQL, which we plan to apply to our authorization model in order to guarantee enforcement of the security policy within the backend systems. Future work will also have to deal with special security requirements with respect to aggregation functions different to SUM, like variance or average, for instance. Likewise the issue of derived authorizations has to be addressed in the context of data warehouses especially when serving a drill-through operation that directly forwards access to the OLTP data sources.

6. References

[AOSS00] A. Abello, M. Oliva, J. Samos, and F. Saltor. Information System Architecture for Secure Data Warehouses. Technical Report LSI-00-26-R, Dep. of Information Systems, University of Catalunya, Spain, April 2000.

[Bhar00] B. Bhargava. Security in Data Warehousing (Invited Talk). Proceedings of the 3rd Data Warehousing and Knowledge Discovery (DAWAK'00), 2000.

[BR01] S. Barker and A. Rosenthal. Flexible Security Policies in SQL. Proc. 15th Annual IFIP WG 11.3 Working Conf. on Database and Application Security, Niagara on the Lake, Ontario, Canada, July 15-18, 2001.

[Kimb96]R. Kimball. Data Warehouse Tool Kits. John Wiley & Sons, 1996.

[KKST97] R. Kirkgöze, N. Katic, M. Stolba, and A M. Tjoa. A Security Concept for OLAP. Proceedings of the 8th. International Workshop on Database and Expert Systems Applications (DEXA'97), IEEE Computer Society, 1997.

[KQST98] N. Katic, G. Quirchmayr, J. Schiefer, M. Stolba, and A M. Tjoa. A Prototype Model for Data Warehouse Security based on Metadata. Proc. 9th Int. Conf. on Database and Expert Systems Applications (DEXA'98), Aug. 26-28, 1998, Vienna, Austria. IEEE Computer Society, Vol. 8, pp. 300-308, 1998

[ML99] J.D. Moffett, and E.C. Lupu. The Use of Role Hierarchies in Access Control. 4th ACM Workshop on Role Based Access Control (RBAC), George Mason University, Fairfax, VA, 28-29 October 1999.

[OSM00] S. Osborn, R. Sandhu, and Q. Munawer. Configuring Role-Based Access Control to Enforce Mandatory and Discretionary Access Control Policies. ACM Transaction on Information and System Security, Vol. 3, No. 2, Pages 85-206, May 2000.

[PP00] T. Priebe, and G. Pernul. Towards OLAP Security Design – Survey and Research Issues. Proc. of the 3rd ACM International Workshop on Data Warehousing and OLAP (DOLAP 2000), Washington, DC, November 10, 2000.

[RS00] A. Rosenthal, and E. Sciore. View Security as the Basis for Data Warehouse Security. Proceedings of the International Workshop on Design and Management of Data Warehouse (DMDW'2000), Manfred A. Jeusfeld, Hua Shu, Karlstad Martin Staudt, Gottfried Vossen (Editor), Sweden, June 2000.

[RSD99] A. Rosenthal, E. Sciore, and V. Doshi. Security Administration for Federations, Warehouses, and other Derived Data. Proc. IFIP WG 11.3 Working Conf. on Database Security, 1999, Seattle, WA, USA.

[SCFY95] R.S. Sandhu, E.J. Coyne, H.L. Feinstein, and C.E. Youman. Role-Based Access Control Models. IEEE Computer, Volume 29, Number 2, pages 38-47, February 1996.

[TSSL++97] B. Thuraisingham, L. Schlipper, P. Samarati, T.Y. Lin, S. Jajodia, and C. Clifton. Security issues in Data warehousing and data mining: panel discussion. Proc. 11th IFIP WG 11.3 Working Conf. on Database Security, Lake Tahoe, California, USA, Aug. 1997. Database Security XI: Status and Prospects, T.Y. Lin and S. Quian (Eds.), Chapman & Hall, 1997.