# An Authorization Model to Access the Summarized Data of Data Warehouse

Shaukat Ali [1], Azhar Rauf [2], Shah Khusro[2], Muhammad Zubair[1], Haleem Farman[1], Sajid Ullah[1]

[1.] Department of Computer Science, Islamia College Peshawar, Pakistan
[2.] Department of Computer Science, University of Peshawar, Pakistan
shaukat@icp.edu.pk

**Abstract:** Data warehouse is large repository of data. Many useful reports can be generated from data warehouse. These reports having useful information and the information will be more accurate if the report generated from large amount of data. The amount of data is directly proportional to time span, more is the time duration data will be in large volume and vice versa. For example, there will be more volume of data in an year as compare to a month. In this paper a Role Based Access technique is introduced to restrict users to summarized data. Different users may have different access authorizations to summarized data. This restriction on the basis of summarized data will improve the security because more summarized data having more information as compare to less summarized data. [Shaukat Ali, Azhar Rauf Shah Khusro, Muhammad Zubair, Haleem Farman, Sajid Ullah. **An Authorization Model to Access the Summarized Data of Data Warehouse**. *Life Sci J* 2014;11(6s):608-610] (ISSN:1097-8135). http://www.lifesciencesite.com. 125

Keywords: Data Warehouse Security, Access Control

## 1. Introduction

Data warehouse is a large repository of data, which can be used for reporting purposes and useful information can be retrieved in these reports. A report is actually summarized data retrieved in a query. This summarize data contains some useful and crucial information which needs security from unauthorized user. Large size of data warehouses and the strategic purpose of the data warehouse systems, the query outputs are nearly always summarized data (Martyn, 2004), and the aggregate query outputs could result important data resulting information leakness to low profile user in organization.

A data warehouse is an integrated collection of data source designed to help managerial decision making and other relevant functions. It contains both highly grained level and summarized level historical data relating to different categories and subjects areas. All data unite is time oriented. The data warehouse enterprise wide decision support system and normally does not have much data updating operation. It empowers the managerial users to perform analysis on data stored in data warehouse (Anahory, 1997); (Katic, 1998).

It is obvious that summarized data has useful information. The information from more summarized data may be more accurate and sensitive as compared to a report which is obtained from less summarized data. Normally, the executives or high level users in an organization need more summarized reports for long or short term decision making, while the low profile users need less summarized report for day-to-day operations. This paper proposes a role based access control mechanism which enables users access to summarized data on the basis of his/her role in the organization. If the user has executive role in the organization then he/she may be able to access data at more summarized level and get more information, but the user with low level will get less summarized report to get less information because he/she may not be authorized to it. This security mechanism is very useful to apply for insiders.

The table (Table 1) shows the proposed time intervals and the user access roles to that time span's summarized data.

Table 1: Table of different users' roles and their access time span for summarized data

| Level of Summarization | User category |
|---|---|
| Day | Ordinary User |
| Month | Classified User |
| Year | Confidential User |
| All Time | Executive User |

### A. Importance of Security for Data Warehouse

Security is the protection of data from un-authorized users (Jan, 2010). Security is an important aspect for any type of data (Ali, 2013). Security is much important for data warehouses as compare to other data source, because data warehouse consolidated from very different data source in a single large data repository. The leakness of information from data warehouses may be very crucial in sense that all data in collected to a single point and vulnerable to attacks. The data is important when some information can be retrieved from it. That information is more valuable which is extracted from large amount of data. The information from data warehouse and databases is normally extracted in the form of reports. The data warehouse reports is more

informative and important with security point of view as it is retrieved from large amount of aggregated and consolidate data. The summarized report may be different, for example, current detail data, lightly summarized data, and highly summarized data reports. Low summarized reports are retrieved from less span of time and highly summarized from large span of time. Difference in summarization level reports may need different level of security layers.

### B.  Motivating Example

Data warehouse is dimensional in nature and subject oriented. Consider a data mart of subject Sale. The most common dimensions of Sale data mart are Time, Location, and Product. A user can summarized data of this Sale data mart with different levels, which may carry different summarized information for some decision making. If the data is summarized at low level then, it may not carry sufficient information for an intruder (insider).
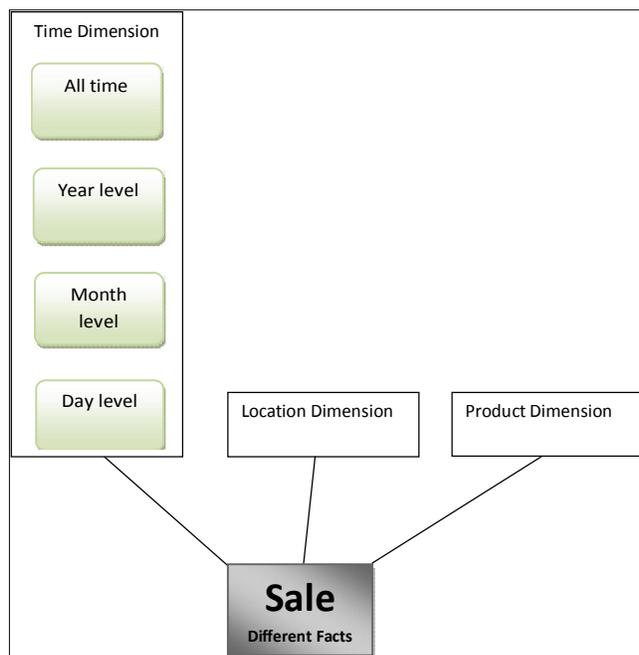


**Figure 1: Star schema of a Sale data mart**

Consider the figure 1, here different level of time dimension is shown. These different dimension's levels show the different summarization levels, because more is the time span high will be the summarization level and vice versa. If more the summarization level more information can be gained and vice versa. If a user has high level of authorization he/she should be allowed to access more summarization level. There should be an authorization mechanism which restrict low level authorization user to high level of summarized data

because high level of summarized data has more information which should be hidden from low profile users.

### 2. Related Work

Data warehouse integrates data from different resources in a single repository, therefore, Security can be considered as an important factor in data warehouses (Santos, 2011); (Vela, 2006). Some researchers propose the role based access control mechanism which is used for securing data warehouse (Fugkeaw, 2008). The Stolba et al. (Stolba, 2006) highlights the security issues in federated data warehouse. They proposed security measurements for assurance of patient privacy of medicine data.

A Data Warehouses contains consolidated, historical, and summarized data to support decision makers at different level (El Sayed, 2010). An access control model is proposed in this work for data warehouse. The users are restricted to data warehouse's data on bases of roles. The authors do not consider the access roles to the summarized data.

A hybrid approach is used for the securing data warehouse from unauthorized access (Ahmad, 2010). The author combined the different classical approaches to secure data before transferring to data warehouse.

Thuraisingham and Iyer (Thuraisingham, 2007) proposed an Extended Role Based Access Control for securing enterprise data warehouse. The authors claimed that their technique gives better privacy and confidentiality for organizations' data.

### 3. Proposed Work

Whenever a user wants to summarize the data with respect to time dimension, his/her authorization role is checked for different summarize data. If a user qualifies for a level of summarization then he/she is given privileges to summarized data for a particular period of time otherwise he/she is restricted. The level of summarization is directly proportional to time duration of data. More is the time duration high will be the summarization level. It can be written mathematically as follow:

$$level\ of\ summarization \propto time\ span$$

Different authorization levels with respect to time span are shown mathematically as follow:

$$Authorization\ level\ of\ \text{Ordinary User} = \sum_{n=1}^{1} D_n \qquad Eq.1$$

$$Authorization\ level\ of\ \mathbf{Classified\ User} = \sum_{n=1}^{30} D_n \qquad Eq.II$$

$$Authorization\ level\ of\ \mathbf{Confidential\ User} = \sum_{n=1}^{365} D_n \qquad Eq.III$$

$$Authorization\ level\ of\ \mathbf{Executive\ User} = \sum_{n=1}^{\infty} D_n \qquad Eq.IV$$

Where "**D**" is the number of days (time span) used in the summarized report while querying data.
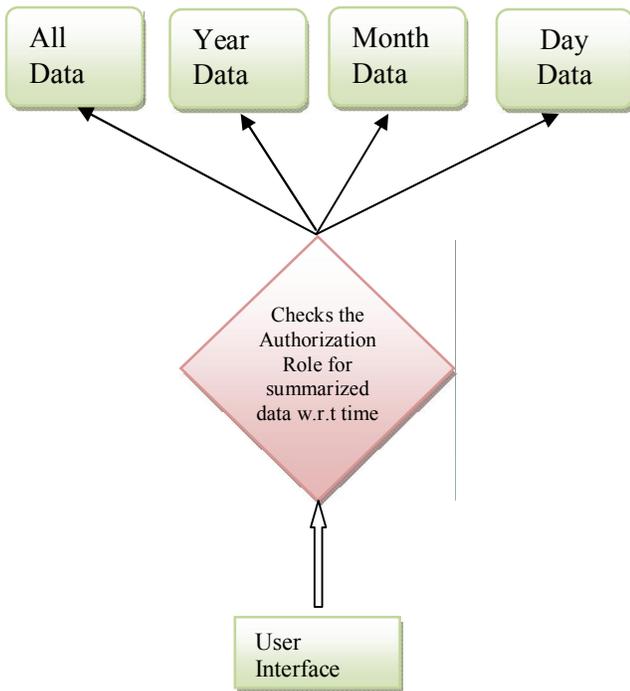


**Figure 2: Architecture of the proposed access control mechanism**

Figure 2 shows the architecture of the proposed role based authorization model. When a user poses query, his/her authorization level is checked. If a user has authorization to summarized data according to equation 1, so he/she can summarize data upto day level. If a role satisfies equation 2, he/she can summarize data upto month duration. Roles according to equations 3 and 4 can summarize data upto an year or for whole duration respectively. If a user exceeds from his/her authorization level his/her query will restrict.

6/14/2014

## 2. Conclusion

This paper proposes a novel role based access control mechanism on summarized data. This summarized data is based on time span of data, more is the time span of data selected in a query more information it will have and vice versa. Therefore, in this paper the roles are restricted to different time span of data to preserve data security.

## References
1. Ahmad, S. and R. Ahmad (2010). "An improved security framework for data warehouse: A hybrid approach" International Symposium in Information Technology (ITSim), Kuala Lumpur, Malaysia, IEEE.
2. Anahory, S. and D. Murray (1997). Data Warehouse, Addison-Wesley.
3. El Sayed, A., A. El Bastawissy, et al. (2010). "DW Access Control Model" IJCSNS **10**(10): 193 - 197.
4. Fugkeaw, S., P. Manpanpanich, et al. (2008). "Securing Web OLAP Based on RBAC Model", International Conference on Advanced Computer Theory and Engineering, Phuket IEEE.
5. Jan, Z., M. Shah, et al. (2010). "Access Control Mechanism for Web Databases by Using Parameterized Cursor". 5th International Conference on Future Information Technology (FutureTech), Busan, IEEE.
6. Katic, N., G. Quirchmay, et al. (1998). "A Prototype model for data warehouse security based on metadata". 9th International Workshop on Database and Expert Systems Applications, Vienna, Austria, IEEE.
7. Martyn, T. (2004). "Reconsidering multi-dimensional schemas." ACM SIGMOD Record **33**(1): 83-88.
8. Santos, R. J., J. Bernardino, et al. (2011). "A survey on data security in data warehousing: Issues, challenges and opportunities" IEEE International Conference on Computer as a Tool (EUROCON), Lisbon, IEEE.
9. Stolba, N., M. Banek, et al. (2006). "The security issue of federated data warehouses in the area of evidence-based medicine" The First International Conference on Availability, Reliability and Security, IEEE.
10. Thuraisingham, B., M. Kantarcioglu, et al. (2007). "Extended RBAC-based design and implementation for a secure data warehouse." International Journal of Business Intelligence and Data Mining **2**(4): 367-382.
11. Ali, S., Rauf, A., and Javed, H., "A Technique for Handling Range and Fuzzy Match Queries on Encrypted Data" International Arab Journal of Information Technology, Vol. 10, No. 3, May 2013.
12. Vela, B., E. Fernández-Medina, et al. (2006). "Model driven development of secure XML databases." ACM SIGMOD Record **35**(3): 22-27.