ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

COMPARATIVE STUDY ON OLAP SECURITY AND INTEGRITY IN DATA WAREHOUSING

C.N.Ravi*, Dr. C. Nalini Chidambaram**
***Research Scholar, Dept of CSE, Bharath Institute of Higher Education and Research, Bharath University, Chennai
**Professor, Dept of CSE, Bharath Institute of Higher Education and Research, Bharath University, Chennai.
Email: mail:nalinicha2002@gmail.com

Received on: 15.10.2016 *Accepted on: 22.11.2016*

Abstract

According to W H Inman, A data warehouse is a subject-oriented, integrated, non volatile and Time-variant collection of data in support of management's decisions. The tasks of data warehouse are data cleaning, data integration, data consolidation and summarization. Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. To facilitate complex analyses and visualization, the data in a warehouse is typically modeled multi dimensionally called cube. In this paper, we surveyed and evaluated the literature related to the various research papers, The important points, Concepts, Algorithm, Limitations and Future enhancements explained in each paper are given as study report. The literature survey essence is summarized and it has submitted in comparative statement format.

Keywords: Security; OLAP; Integrity; Multidimensional.

Introduction

Data warehouse uses the Update driven approach rather than query driven approach, in which data from multiple, heterogeneous sources is integrated in advance and stored in a warehouse/repository for direct querying and analysis . Building a data warehouse is a very challenging issue because compared to software engineering it is quite a young discipline and does not yet offer well-established strategies and techniques for the development process. Different Data warehouse tools are used different Levels such as data modeling tools, ETL tools, Multi dimensional data base tools, reporting and analysis tools. We address issues related to the protection of private information in Online Analytical Processing (OLAP) systems, where a major privacy concern is the adversarial inference of private information from OLAP query answers. Many commercial products and services are now available, and all of the principal database management system vendors now have offerings in these areas.

Decision support places some rather different requirements on database technology compared to traditional on-line transaction processing applications. This paper provides an overview of data warehousing and OLAP technologies, important points, Concepts, Algorithm, Limitations and Future enhancements of various implementations.

1.1. General Security Considerations

Obviously a lot of communication takes place in a data warehouse system, creating the need for proper communication security measures. The data load process (transferring the source data from operational databases to the data warehouse) defines new requirements for a network infrastructure. Independent (possibly distributed) source databases have to be consolidated over a network. As the data may be highly sensitive it is essential to protect it from eavesdropping and similar secrecy threats. For the communication between the front-end applications and the OLAP server (or the data warehouse in 2-tier environments) usually a client/server connection will be utilized, possibly to remote sites. Even though the information on this channel is most likely aggregated and less complete, it may be highly security critical.

The use of the Internet or other possibly insecure networks for the above mentioned connections makes suitable security measures necessary. As only some tools support encrypted communication on application level, virtual private network (VPN) technology might be appropriate. Authentication and audit are other security measures that have to be installed in a data warehouse environment. A clear proof of a user's identity is needed in order to apply appropriate security restrictions and to avoid access by unauthorized users. Corresponding user identification and authentication mechanisms check the authenticity of the pretended identity, either inside the frontend tools, or by making use of authentication mechanisms provided by modern operating systems or tools allowing a single sign-on strategy. The essential decision for auditing in data warehouses is to put it in the right place in the architecture. Using only the auditing capabilities of an underlying (relational) DBMS holding the data warehouse will not satisfy the needs, as logging the access to tables of the star/snowflake schema (or similar objects)will not expose the actual multidimensional queries performed (especially in MOLAP based systems). The conclusion is that auditing should also be performed on the multidimensional level of an OLAP engine (i.e. at the same level where authorization semantics are defined).

1.2 Privacy and Access Control Policy

Access control on the back-end side involves controlling the access to the data warehouse and the source databases by the extract/ transform/load processes and the access to these procedures (invoking as well as administration). In a

role-based authorization model for the Administrative processes in a data warehouse is presented identifying two categories of roles. We have identified Communication security, user identification and authentication, auditing, and access control as important security issues.

2. OLAP Security System Design

Deriving the access control policies from the operational data sources is very difficult although some research efforts are made in this area Data from different systems (with different policies) will be consolidated. Also, the users of the operational systems are not the same as the users of the data warehouse. However, the main problem is, that the relational model is predominate in operational systems while OLAP systems make use of the nontraditional multidimensional model. Access control schemes do not map easily. Protection is not defined in terms of tables, but dimensions, Hierarchical paths, granularity levels. The need for proper OLAP security design arises.

2.1. System Settings

Let there be one data warehouse server and a number of users C1; . . . ; CU in the system. For each user Ck (k 2 ½1; U_), let Gk be the set of cells in the data cube that Ck has the right to access. We refer to Gk as the set of pre known cells of Ck. Since the data warehouse server needs to properly enforce access control on the sensitive cells, we assume that it knows Gk for k 2 ½1; U_ as pre knowledge.

2.2. Models of Privacy Intrusion Attacks and Defensive Countermeasures

A user Ck can be an adversary intending to compromise the private information about sensitive cells x (x 62 Gk) that it has no right to access. In order to do so, Ck may launch an inference attack by inferring private information about x from the pre known cells in Gk as well as the historic query answers in Qk. The issue of effectively and efficiently computing and managing privacy preserving OLAP data cubes [has attracted the interest from a large community of Database and Data Warehousing researchers. This problem indeed plays a critical role for both centralized and distributed environments.

Applications where computing privacy preserving data cubes is relevant embrace a large range of cases, spanning from Business Intelligence (BI) systems to Data Mining and Analysis tools, and from sensor network data analysis tools to social network data components. Despite the equal dignity‖ of both the target environments (i.e., centralized and distributed), the distributed case is indeed more relevant, as today very large organization very often make use of a distributed infrastructure for producing, managing and delivering knowledge.

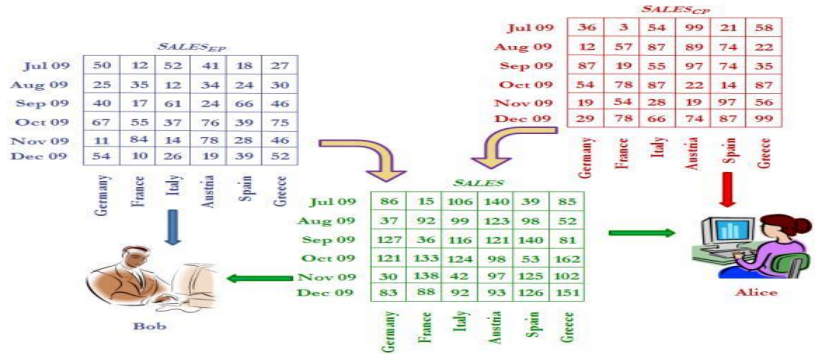

Figure 1. A privacy preserving distributed OLAP scenario.

Fig. 1 depicts an applicative example of such a scenario. Here, Alice and Bob are two analysts working at different companies federated to the same main organization, respectively. Alice and Bob perform OLAP on sale data extracted from the own respective legacy databases, with a focus on computer and electronic parts for Alice and Bob, respectively. To this end, Alice build and query the two-dimensional data cube *SALESCP*, and Bob build and query the two-dimensional data cube *SALESEP*, respectively.

SALESEP have the same *multidimensional schema* constituted by the dimensions *Time* (with granularity *Year*) and *Zone* (with granularity *Country*), respectively, and the measure *Sale*. As regards data, both data cubes *SALESCP* and *SALESEP* store aggregations on sales occurred in EU countries during the second half of the 2009.

Base paper Topic	**Authors**	**Journal Details Concepts explained**	**Technology / Algorithm used**	**Conclusions**
Data security and privacy in data mining: research issues & preparation	1.Dileep Kumar Singh, #IT Resource Centre Madan Mohan Malaviya Engineering College, Gorakhpur, India 2. Vishnu Swaroop *Dept. of Computer Science& Engineering, Madan Mohan Malaviya Engineering College Gorakhpur, India	International Journal of Computer Trends and Technology-Volume-4-Issue 2- 2013 role of self-regulation, data protection laws	Not Applicable	Federal legislation is necessary
A Pragmatic Approach to Conceptual Modelling of OLAP Security	Torsten Priebe, Günther Pernul Department of Information Systems, University of Essen, University. 9, D-45141 Essen, Germany {priebe,pernul}@wi-inf.uni-essen.de	To appear in Proc. 20th International Conference on Conceptual A conceptual (or Modeling (ER semantic) model, 2001), Access control November 27-30, policies2001, Yokohama, Japan	Unified Modeling Language (UML), ADAPTed UML, Multidimension al Security Constraint Language (MDSCL)	This can be applied real life situations
A security concepts of	Remi kirkgoze,	Published in:		

OLAP.	Nevena Katic, Instuite of Software Tech(IFS) Vienna university of Technology, Austria.	Discretionary Proceeding Access Control DEXA '97 Proceedings of the 8th International (DAC) Workshop on Database and Security Expert Systems Applications subjects(S), Page 0619 Attributes (A) Objects' (O). IEEE Computer Society Washington, DC, USA ©1997	Mandatory Security Model,(MAM) Adaptive Mandatory Security model used(AMSM)	Its is flexible in assigning roles
Modelling conflict of interest in the design of secure data warehouses A security algorithm for on-line analytical processing data cube	Salah Triki, Hanene Ben-Abdallah, Jamel Feki Laboratoire Mir@cl, Faculté des Sciences Economiques et de Gestion de Sfax, Route de l'Aéroport Km 4 – 3018 Sfax, BP. 1088 Salah.Triki@fsegs.rnu.tn , Hanene.BenAbdallah@ fsegs.rnu.tn, Jamel.Feki@fsegs.rn **1.**Narander Kumar Department of Computer Science B. B. Ambedkar University Lucknow (U.P.), 226025, India	published in "KEOD 2010 – International Conference on Knowledge Mandatory Engineering and Access Control Ontology (MAC), Development, Spain (2010)" Role-Based Access Control (RBAC) International Journal of Computer Applications (0975 Encryption and – 8887) Volume 79 Decryption – No 14, October 2013	Using MAC develop a new Concept Secure Data Warehouse(SEC DW) Calculate the cipher $Y=2^{X.}$ Calculate the Plain text $Z=\log_2 Y$	Developing a Case tool set of SECDW Finding a new algorithm SEAOLAP
Privacy-preserving OLAP: an	Nan Zhang, Member, IEEE, Wei Zhao,	Authorized licensed common use limited to: The aggregate		protects private information

Information-theoretic approach	Fellow, IEEE	George Washington University. Downloaded on February 23,2010 at 08:21:26 EST from IEEE Xplore. Restrictions apply.	functions (e.g., COUNT, SUM, MIN, MAX, MEDIAN)	1.Inference Control 2.Input/Output perturbation	against both exact and partial disclosure
Towards a theory for privacy preserving distributed OLAP	Alfredo Cuzzocrea ICAR-CNR and University of Calabria Cosenza, Italy cuzzocrea@si.deis.unical.it Elisa Bertino CERIAS and Purdue University West Lafayette, IN, USA bertino@cs.purdue.edu	PAIS'12, March 30, 2012, Berlin, Germany. Copyright 2012 ACM 978-1-4503-1143-4/12/03…$10.00.	CUR matrix decomposition method,	Not Applicable	Not Applicable
Securing OLAP data cubes against privacy breaches	Lingyu Wang, Sushil Jajodia, and Duminda Wijesekera Center for Secure Information Systems George Mason University Fairfax, VA 22030-4444, USA {lwang3,jajodia,dwijesek}@gmu.edu	Proceedings of the2004 IEEE Symposium on Security and Privacy (S&P'04) 1081-6011/04 $ 20.00 © 2004 IEEE	specifying Authorization objects in data cubes. An authorization is usually a triple: (*object,subject,*(*signed*) *action*),	Data cube separation	protecting sensitive data in OLAP data cubes from unauthorized accesses and malicious inferences.

A prototype model for data warehouse security based on metadata	N. Katic 1 G. Quirchmayr 2 J. Schiefer 1.Institute of Software Technology (E188) Vienna University of Technology Resselgasse 3/188, A-1040 Vienna Austria {katic, stolba, js, tjoa} @ifs.tuwien.ac.at 2Institute of Applied Computer Science and Information Systems University of Vienna Liebiggasse 4, A-1010 Vienna Austria	Ninth International Workshop on Meeting Date 26 -28 Aug 1998 Print ISBN: 0-8186-8353-8 INSPEC Accession Number: 6040729 Conference Location : Vienna Digital Object Identifier : 10.1109/DEXA.199 8.707417	metadata driven approach implemented	Audit Requirements and Network requirements	Security aspects should be considered in the design phase
Towards the secure modelling of OLAP users' behaviour	Carlos Blanco 1, Eduardo Fernández-Medina 2, Juan Trujillo 1. Dep. Of Mathematics, Statistical and Computation. Facultad de Ciencias University of Cantabria. Av. De los Castros s/n. 39071. Santander. Spain	7th VLDB Workshop, SDM 2010, Singapore, September 17, 2010. Proceeding **DOI** 10.1007/978-3-642-15546-8_8	Sensitive information assignment rules (SIAR) authorization rules (AUR)	Joint Rules (JR)	security rules (joint rules) has been added to our static modelling proposal in order to define sensitive combinations of information and the security privileges needed to query this info.

3. Conclusion

In this paper, we make a comparative study of different approaches used for data warehouse Security design. In the literature survey, different authors [1, 2, 3, 4,..,10,] have proposed different techniques at different levels i.e. conceptual level ,logical level and physical level. Our comparative study is based on following criteria: Proposal, Framework/Architecture, Approach or technique proposed, Schema used, whether the design can be extended to logical and physical design also, Case study and Tool used.

References

1. Castano, S., Fugini, M., Martella, G., Samarati, P.: Database Security. ACM Press, 1994.
2. Chaudhuri, S., Dayal, U.: An Overview of Data Warehousing and OLAP Technology. 1996.
3. Cognos Incorporated: Schrittweise Anleitungen für Transformer. Cognos Power-Play Version 6.0. 1998.
4. Denning, D.E., Schlörer, J.: Inference Controls in Statistical Databases. In IEEE Computer Vol. 16(7); July 1983.
5. Doshi, V., Jajoda, S., Rosenthal, A.: A Pragmatic Approach to Access Control in Data Warehouses. Via private communication,1999.
6. Essmayer, W., Wagner, R., Kapsammer, E., Tjoa, A.M.:Meta-Data for Enterprise-Wide Security Administration. In Proceedings of the Third IEEE Computer Society Metadata Conference; NIH Campus, Bethesda, Maryland, April 6-7, 1999.
7. Edgar W., Oscar M., Wolfgang E., Franz L.,Werner W,.2001. An Authorization Model for Data Warehouses and OLAP. In : Workshop on Security in Distributed Data Warehousing
8. Eric S. K. Y. 1997. Towards modelling and reasoningsupport for early-phase requirements engineering. In,the Third IEEE International Symposium onRequirements Engineering. Annapolis, MD, U.S.A Michael S. J., (1992). The Z Notation: A reference manual (2nd edition ed.). Prentice Hall International Series in computer science.
9. Lujan-Mora, S., Trujillo, J. and Song, I. Y., 2002. Extending the UML for multidimensional modeling. In: 5th International Conference on the Unified Modeling Language. Dresden, Germany, Springer-Verlag. LNCS 2460.
10. OMG 2005. MOF QVT final adopted specification. Ravi S. Sandhu and Edward J. Coyne and Hal L. Feinstein and Charles E. Youman, 1996. Role-Based Access Control Models, pp 38--47. IEEE Computer Vol. 29 Num. 2.
11. Rodolfo V., Eduardo F., Mario P., Juan T. 2006. A UML 2.0/OCL Extension for Designing Secure Data Warehouses. Journal of Research and Practice in Information Technology, Vol. 38, Num. 1 pp. 31—43 Sandhu,

R. S., Coyne E.J., Feinstein H.L. and Youman C.E., 1996. Role-Based Acces Control Models, IEEE Computer 29(2): 38-47, IEEE Press.

12. Inmon, H., *Building the Data Warehouse*. Third Edition ed. 2002, USA: John Wiley & Sons.
13. Kimball, R., *The Data Warehouse Toolkit*. 2002: John Wiley & Sons.
14. Lodderstedt, T., D. Basin, and J. Doser. *SecureUML: A UML-based modeling language for model-driven security*. in *UML 2002. The Unified Modeling Language. Model Engineering, Languages Concepts, and Tools. th International Conference*. 2002. Dresden, Germany: Springer.
15. Mouratidis, H. and P. Giorgini, *Integrating Security and Software Engineering: Advances and Future Vision*. 2006: Idea Group Publish.