Lecture Notes in Computer Science

The LNCS series reports state-of-the-art results in computer ... research, development, and education, at a high level and in both ... and electronic form. Enjoying tight cooperation with the R&D ... with numerous individuals, as well as with prestigious organizatio... societies, LNCS has grown into the most comprehensive compute... research forum available.

The scope of LNCS, including its subseries LNAI and LNBI, spans the whole range of computer science and information technology includ... interdisciplinary topics in a variety of application fields. The type of material published traditionally includes

- proceedings (published in time for the respective conference)
- post-proceedings (consisting of thoroughly revised final full papers)
- research monographs (which may be based on outstanding PhD work, research projects, technical reports, etc.)

... several color-cover sublines have been added featuring, ... collection of papers, various added-value components; the ... include

- ... (textbook-like monographs or collections of lectures given ... courses)
- ... surveys (offering complete and mediated coverage ...)
- ... (introducing emergent topics to the broader community)

... the printed book, each new volume is published electronic...

... on LNCS can be found at

... publication should be sent to
... Tiergartenstr. 17, 69121 Heidelberg, Germany
... .com

ISSN 0302-9743

ISBN 978-3-642-03729-0

Lecture Notes in Computer Science
LNCS LNAI LNBI

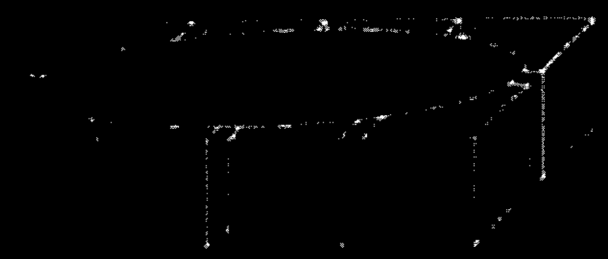
Pedersen • Mohania
Tjoa (Eds.)

Knowledge Di...

LNCS 5691

D... an
K... overy

11th International Conference, DaWaK 2009
Linz, Austria, August/September 2009
Proceedings



Springer

Lecture Notes in Computer Science 5691

Torben Bach Pedersen   Mukesh K. Mohania
A Min Tjoa (Eds.)

Data Warehousing and Knowledge Discovery

11th International Conference, DaWaK 2009
Linz, Austria, August 31–September 2, 2009
Proceedings

Springer

Volume Editors

Torben Bach Pedersen
Aalborg University
Department of Computer Science
Selma Lagerlöfsvej 300, 9220 Aalborg Ø, Denmark
E-mail: tbp@cs.aau.dk

Mukesh K. Mohania
IBM India Research Lab
Plot No. 4, Block C, Institutional Area
Vasant Kunj, New Delhi 110 070, India
E-mail: mkmukesh@in.ibm.com

A Min Tjoa
Vienna University of Technology
Institute of Software Technology and Interactive Systems
Favoritenstr. 9-11/188, 1040 Wien, Austria
E-mail: amin@ifs.tuwien.ac.at

Library of Congress Control Number: 2009932136

CR Subject Classification (1998): H.2, H.4, H.3, J.1, H.2.8, H.3.3, I.5.3

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-642-03729-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-03729-0 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12737444 06/3180 5 4 3 2 1 0

Preface

Data warehousing and knowledge discovery are increasingly becoming mission-critical technologies for most organizations, both commercial and public, as it becomes increasingly important to derive important knowledge from both internal and external data sources. With the ever growing amount and complexity of the data and information available for decision making, the process of data integration, analysis, and knowledge discovery continues to meet new challenges, leading to a wealth of new and exciting research challenges within the area.

Over the last decade, the International Conference on Data Warehousing and Knowledge Discovery (DaWaK) has established itself as one of the most important international scientific events within data warehousing and knowledge discovery. DaWaK brings together a wide range of researchers and practitioners working on these topics. The DaWaK conference series thus serves as a leading forum for discussing novel research results and experiences within data warehousing and knowledge discovery. This year's conference, the 11th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2009), continued the tradition by disseminating and discussing innovative models, methods, algorithms, and solutions to the challenges faced by data warehousing and knowledge discovery technologies.

The papers presented at DaWaK 2009 covered a wide range of aspects within data warehousing and knowledge discovery. Within data warehousing and analytical processing, the topics covered data warehouse modeling including advanced issues such as spatio-temporal warehouses and DW security, OLAP on data streams, physical design of data warehouses, storage and query processing for data cubes, advanced analytics functionality, and OLAP recommendation. Within knowledge discovery and data mining, the topics included stream mining, pattern mining for advanced types of patterns, advanced rule mining issues, advanced clustering techniques, spatio-temporal data mining, data mining applications, as well as a number of advanced data mining techniques. It was encouraging to see that many papers covered emerging important issues such as spatio-temporal data, streaming data, non-standard pattern types, advanced types of data cubes, complex analytical functionality including recommendations, multimedia data, mssing and noisy data, as well as real-world applications within genetics and within the clothing and telecom industries. The wide range of topics bears witness to the fact that the data warehousing and knowledge discovery field is dynamically responding to the new challenges posed by novel types of data and applications.

From 124 submitted abstracts, we received 100 papers from 17 countries in Europe, North America and Asia. The Program Committee finally selected 36 papers, yielding an acceptance rate of 36%.

We would like to express our most sincere gratitude to the members of the Program Committee and the external reviewers, who made a huge effort to review the papers in a timely and thorough manner. Due to the tight timing constraints and the high number of submissions, the reviewing and discussion process was a very challenging task, but the commitment of the reviewers ensured that a very satisfactory result was achieved.

We would also like to thank all authors who submitted papers to DaWaK 2009, for their contribution to making the technical program so excellent.

Finally, we extend our warmest thanks to Gabriela Wagner for delivering an outstanding level of support within all aspects of the practical organization of DaWaK 2009. We also thank Amin Anjomshoaa for his support with the conference management software.

August 2009

Torben Bach Pedersen
Mukesh Mohania
A Min Tjoa

Organization

Program Chairs

Torben Bach Pedersen	Aalborg University, Denmark
Mukesh Mohania	IBM India Research Lab, India
A Min Tjoa	Vienna University of Technology, Austria

Publicity Chair

Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy

Program Committee

Alberto Abello Gamazo	Universitat Politecnica de Catalunya, Spain
Elena Baralis	Politecnico di Torino, Italy
Ladjel Bellatreche	Poitiers University, France
Petr Berka	University of Economics, Prague, Czech Republic
Jorge Bernardino	Instituto Superior de Engenharia de Coimbra, Portugal
Elisa Bertino	Purdue University, USA
Mokrane Bouzeghoub	CNRS - Université de Versailles SQY, France
Stephane Bressan	National University of Singapore, Singapore
Peter Brezany	University of Vienna, Austria
Robert Bruckner	Microsoft, USA
Erik Buchmann	Universität Karlsruhe, Germany
Jesús Cerquides	Universitat de Barcelona, Spain
Zhiyuan Chen	University of Maryland Baltimore County, USA
Sunil Choenni	The Netherlands Ministry of Justice, The Netherlands
Frans Coenen	University of Liverpool, UK
Bruno Cremilleux	Université de Caen, France
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy
Agnieszka Dardzińska	University of North Carolina at Chapel Hill, Poland
Karen C. Davis	University of Cincinnati, USA
Kevin Desouza	University of Washington, USA
Curtis Dyreson	Utah State University, USA
Todd Eavis	Concordia University, USA
Johann Eder	University of Klagenfurt, Austria
Tapio Elomaa	Tampere University of Technology, Finland
Roberto Esposito	Università di Torino, Italy

Vladimir Estivill-Castro	Griffith University, Australia
Christie Ezeife	University of Windsor, Canada
Jianping Fan	UNC-Charlotte, USA
Ling Feng	Tsinghua University, China
Eduardo Fernandez-Medina	Universidad de Castilla-La Mancha, Spain
Ada Fu	Chinese University of Hong Kong, Hong Kong
Dragan Gamberger	Ruder Boškovic Institute, Croatia
Chris Giannella	Information Systems Security Operation of Sparta, Inc., USA
Matteo Golfarelli	University of Bologna, Italy
Eui-Hong (Sam) Han	iXmatch Inc., USA
Wook-Shin Han	Kyungpook National University, Korea
Jaakko Hollmén	Helsinki University of Technology, Finland
Xiaohua (Tony) Hu	Drexel University, USA
Jimmy Huang	York University, Canada
Farookh Khadeer Hussain	Curtin University of Technology, Australia
Ryutaro Ichise	Japan National Institute of Informatics, Japan
Mizuho Iwaihara	Kyoto University, Japan
Alípio Mário Jorge	University of Porto, Portugal
Murat Kantarcioglu	University of Texas at Dallas, USA
Jinho Kim	Kangwon National University, Korea
Sang-Wook Kim	Hanyang University , Korea
Jörg Kindermann	Fraunhofer Institute, Germany
Jens Lechtenboerger	Westfälische Wilhelms-Universität Münster, Germany
Wolfgang Lehner	Dresden University of Technology, Germany
Sanjay Madria	University of Missouri-Rolla, USA
Jose Norberto Mazón López	University of Alicante, Spain
Anirban Mondal	University of Tokyo, Japan
Ullas Nambiar	IBM Research, India
Jian Pei	Simon Fraser University, Canada
Evaggelia Pitoura	University of Ioannina, Greece
Stefano Rizzi	University of Bologna, Italy
Monica Scannapieco	University of Rome"La Sapienza", Italy
Alkis Simitsis	HP Labs, USA
Il-Yeol Song	Drexel University, USA
Koichi Takeda	Tokyo Research Laboratory, IBM Research, Japan
Dimitri Theodoratos	New Jersey Institute of Technology, USA
Christian Thomsen	Aalborg University, Denmark
Igor Timko	Free University of Bozen-Bolzano, Italy
Juan-Carlos Trujillo Mondéjar	University of Alicante, Spain
Panos Vassiliadis	University of Ioannina, Greece
Millist Vincent	University of South Australia, Australia
Wolfram Wöß	Johannes Kepler Universität Linz, Austria
Robert Wrembel	Poznan University of Technology, Poland
Xiaofang Zhou	University of Queensland, Australia
Esteban Zimanyi	Université Libre de Bruxelles, Belgium

External Reviewers

Timo Aho
Jussi Kujala
Ryan Bissell-Siders
Marc Plantevit
Francois Rioult
Ke Wang
Jinsoo Lee
Julius Köpke
Marcos Aurelio Domingues
Nuno Escudeiro
Tania Cerquitelli
Paolo Garza
Ibrahim Elsayed
Fakhri Alam Khan
Yuzhang Han
Xiaoying Wu

Table of Contents

Pattern Mining

Data Cubes

Data Mining Applications

Analytics

Data Mining

Clustering

Spatio-Temporal Mining

Rule Mining

Olap Recommendation

New Challenges in Information Integration

Laura M. Haas[1] and Aya Soffer[2]

[1] IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA
[2] IBM Haifa Research Lab, Haifa University Campus, Mount Carmel, Haifa, 31905 Israel
laura@almaden.ibm.com, ayas@il.ibm.com

Abstract. Information integration is the cornerstone of modern business informatics. It is a pervasive problem; rarely is a new application built without an initial phase of gathering and integrating information. Information integration comes in a wide variety of forms. Historically, two major approaches were recognized: data federation and data warehousing. Today, we need new approaches, as information integration becomes more dynamic, while coping with growing volumes of increasingly dirty and diverse data. At the same time, information integration must be coupled more tightly with the applications and the analytics that will leverage the integrated results, to make the integration process more tractable and the results more consumable.

Keywords: Information integration, analytics, data federation, data warehousing, business intelligence solutions.

1 Introduction

Information integration is the cornerstone of modern business informatics. Every business, organization, and today, every individual, routinely deals with a broad range of data sources. Almost any professional or business task we undertake causes us to integrate information from some subset of those sources. A company needing a new customer management application may start by building a warehouse with an integrated and clean record of all information about its customers from legacy data stores and newer databases supporting web applications. A healthcare organization needs to integrate data on its patients from many siloed laboratory systems and potentially other hospitals or doctors' offices. Individuals planning their trip to Austria may integrate information from several different web sites and databases.

There are many information integration problems [1]. Different environments, data sources, and goals have led to a proliferation of information integration technologies and tools [2], each addressing a different piece of the information integration process, for a particular context. There are tools to help explore data on the web, tools to track metadata in an enterprise, and tools to help identify common objects in different data sources. Other technologies focus on information transformation, specifying what data should be transformed and how to transform it, or actually doing the transformation to create the needed data set.

Two major technologies for information integration are data warehousing and data federation. Data warehousing materializes the integrated information, typically leveraging Extract/Transform/Load (ETL) tools to do scalable processing of complex

T.B. Pedersen, M.K. Mohania, and A M. Tjoa (Eds.): DaWaK 2009, LNCS 5691, pp. 1–8, 2009.

Towards a Modernization Process for Secure Data Warehouses

Carlos Blanco[1], Ricardo Pérez-Castillo[1], Arnulfo Hernández[1],
Eduardo Fernández-Medina[1], and Juan Trujillo[2]

[1] Dep. of Information Technologies and Systems, Escuela Superior de Informática
ALARCOS Research Group - Institute of Information Technologies and Systems
University of Castilla-La Mancha, Paseo de la Universidad, 4, 13071, Ciudad Real, Spain
{Carlos.Blanco,Ricardo.PdelCastillo,Arnulfonapoleon.Hernandez,
Eduardo.Fdezmedina}@uclm.es
[2] Dep. of Information Languages and Systems, Facultad de Informática,
LUCENTIA Research Group, University of Alicante, San Vicente s/n. 03690,
Alicante, Spain
jtrujillo@dlsi.ua.es

Abstract. Data Warehouses (DW) manage crucial enterprise information used for the decision making process which has to be protected from unauthorized accesses. However, security constraints are not properly integrated in the complete DWs' development process, being traditionally considered in the last stages. Furthermore, legacy systems need a reverse engineering process in order to accomplish re-documentation for detecting new security requirements as well as system's design recovery to enable migration and reuse. Thus, we have proposed a model driven architecture (MDA) for secure DWs which takes into account security issues from the early stages of development and provides automatic transformations between models. This paper fulfills this architecture providing an architecture-driven modernization (ADM) process focused on obtaining conceptual security models from legacy OLAP systems.

1 Introduction

Data Warehouses (DWs) manage business' historical information used to take strategic decisions and usually follow a multidimensional approach in which the information is organized in facts classified per subjects called dimensions. In a typical DW architecture, ETL (extraction/transformation/load) processes extract data from heterogeneous Data Sources and then transform and load this information into the DW repository. Finally, this information is analyzed by Data Base Management Systems (DBMS) and On-Line Analytical Processing (OLAP) tools.

Since data in DWs are crucial for enterprises, it is very important to avoid unauthorized accesses to information by considering security constraints in all layers and operations of the DW, from the early stages of development as a strong requirement to the final implementation in DBMS or OLAP tools (Thuraisingham, Kantarcioglu et al. 2007).

In this way, DWs' development can be aligned with the Model Driven Architecture (MDA 2003) approach which proposes a software development focused on models at

T.B. Pedersen, M.K. Mohania, and A M. Tjoa (Eds.): DaWaK 2009, LNCS 5691, pp. 24–35, 2009.

different abstraction levels which separate the specification of the system functionality and its implementation. Firstly, system requirements are included in business models (CIM). Then, conceptual models (PIM) represent the system without including information about specific platforms and technologies which are finally specified in logical models (PSM). Moreover, automatic transformations between models can be defined by using several languages such as Query / Views / Transformations (QVT) (OMG 2005).

Furthermore, MDA architectures support reverse engineering capabilities which consists of analysis of legacy systems to (1) identify the system's elements and their interrelationships and (2) carry out representations of the system at a higher level of abstraction (Chikofsky and Cross 1990). Reverse engineering can be used in the development of DWs to accomplish re-documentation for detecting new security requirements as well as system's design recovery to enable migration and reuse. Nevertheless, reverse engineering takes part in a whole reengineering process (Müller, Jahnke et al. 2000). MDA provides the needed formalization to reengineering process to converge in so-called Architecture-Driven Modernization (ADM), another OMG initiative (OMG 2006). ADM advocates reengineering processes where each artifact involved in these processes is depicted and managed as a model (Khusidman and Ulrich 2007).

We have proposed an MDA architecture to develop secure DWs taking into account security issues in the whole development process (Fernández-Medina, Trujillo et al. 2007). To achieve this goal we have defined an access control and audit model specifically designed for DWs and a set of models which allow the security design of the DW at different abstraction levels (CIM, PIM and PSM). This architecture provides two different paths (a relational path towards DBMS and a multidimensional path towards OLAP tools) and includes rules for the automatic transformation between models and code generation.

This paper improves the architecture by defining an architecture-driven modernization (ADM) process which permits re-documentation and platform migration. Since most of DWs are managed by OLAP tools by using a multidimensional approach, this ADM process is focused on the multidimensional path, obtaining conceptual security models (PIM) from logical multidimensional models (PSM) and legacy OLAP systems.

This paper is organized as follows: Section 2 will present the related work on secure DWs; Section 3 will briefly show our complete MDA architecture for developing secure DWs and will underline the difference between our previous works and the contribution of this paper; Section 4 will present the defined ADM process; Section 5 will use an application example to validate our proposal; Section 6 will finally present our conclusions and future work.

2 Related Work

There are relevant contributions focused on secure information systems development, such as UMLSec (Jürjens 2004) which uses UML to define and evaluate security specifications using formal semantics, or Model Driven Security (MDS) (Basin, Doser et al. 2006) which uses the MDA approach to include security properties in high-level system models and to automatically generate secure system architectures. Within the context of MDS, SecureUML (Lodderstedt, Basin et al. 2002) is proposed as an extension of UML for modeling a generalized role based access control.

However, these proposals do not consider the special characteristics of DWs. In this area, solely Priebe and Pernul propose a complete methodology for develop secure DWs (Priebe and Pernul 2001). This methodology deals with the analysis of security requirements, the conceptual modeling by using ADAPTed UML, and the implementation into commercial tools, but does not establish the connection between levels in order to allow automatic transformations. They use SQL Server Analysis Services (SSAS) creating a Multidimensional Security Constraint Language (MDSCL) by extending multidimensional expressions (MDX) with hide statements for cubes, measures, slices and levels.

Although MDA philosophy has been applied to develop secure DWs (Fernández-Medina, Trujillo et al. 2007) and data reverse engineering field has been widely studied in literature (Aiken 1998; Blaha 2001; Cohen and Feldman 2003; Hainaut, Englebert et al. 2004), there is little research on reengineering of data warehouses following an MDA approach and security concerns are not considered. These reengineering works are performed for: re-documentation, model migration, restructuring, maintenance or improvement, tentative requirements, integration, conversion of legacy data.

3 MDA Architecture for Secure DWs

Our architecture to develop secure DWs proposes several models improved with security capabilities which allow the DW's design considering confidentiality issues in the whole development process, from an early development stage to the final implementation. This proposal has been aligned with an MDA architecture (Fernández-Medina, Trujillo et al. 2007) providing security models at different abstraction levels (CIM, PIM, PSM) and automatic transformations between models (Figure 1).

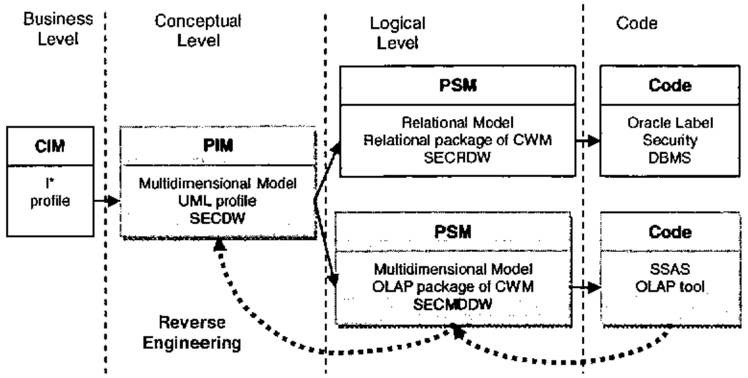


Fig. 1. MDA architecture for Secure DWs

Firstly, security requirements are modeled at business level (CIM) by using a UML profile (Trujillo, Soler et al. 2008) based on the i* framework (Yu 1997), which is an agent oriented approach centered on the agents' intentional characteristics. Then, transformation from secure CIM models to conceptual model (PIM) is achieved applying a methodology described by using the OMG Software Process Engineering Metamodel Specification standard (SPEM) (Trujillo, Soler et al. 2008).

Conceptual models (PIM) are defined according to a UML profile, called SECDW (Fernández-Medina, Trujillo et al. 2007) which has been specifically created for DWs and complemented by an Access Control and Audit (ACA) model focused on DW confidentiality (Fernández-Medina, Trujillo et al. 2006). In this way, SECDW allows the representation of structural aspects of DWs (such as facts, dimensions, base classes, measures or hierarchies) and security constraints which permit the classification of authorization subjects and objects in three ways (into roles (SecurityRole), levels (SecurityLevel) and compartments (SecurityCompartment)) and the definition of several kinds of security rules (Sensitive information assignment rules (SIAR), authorization rules (AUR) and audit rules (AR)).

Multidimensional modeling at the logical level depends of the tool finally used and can be principally classified into online analytical processing by using relational (ROLAP), multidimensional (MOLAP) and hybrid (HOLAP) approaches. Thus, our architecture considers two different paths: a relational path towards DBMS and a multidimensional path towards OLAP tools.

The relational path uses a logical relational metamodel (PSM) called SECRDW (Soler, Trujillo et al. 2008) which is an extension of the relational package of the Common Warehouse Metamodel (CWM 2003) and allows the definition of secure relational elements such as secure tables or columns. Moreover, this relational path is fulfilled with the automatic transformation from conceptual models (Soler, Trujillo et al. 2007) and the eventual implementation into a DBMS, Oracle Label Security.

Furthermore, this MDA architecture was recently improved with a new multidimensional path towards OLAP tools in which a secure multidimensional logical metamodel (PSM), called SECMDDW (Blanco, García-Rodríguez de Guzmán et al. 2008) considers the common structure of OLAP tools and allows to represent a DW model closer to OLAP platforms than conceptual models. SECMDDW is based on a security improvement of the OLAP package from CWM and is composed of: a security configuration metamodel which represents the system's security configuration by using a role-based access control policy (RBAC); a cube metamodel which defines both structural cube aspects such as cubes, measures, related dimensions and hierarchies, and security permissions for cubes and cells; and a dimension metamodel with structural issues of dimensions, bases, attributes and hierarchies, and security permissions which are related to dimensions and attributes.

This path also deals with the automatic transformation from conceptual models by using QVT transformations (Blanco, García-Rodríguez de Guzmán et al. 2008) and the final secure implementation into a specific OLAP platform, SQL Server Analysis Services (SSAS), by using a set of Model-to-Text (M2T) rules.

4 Modernizing Secure DWs

Modernizing DWs provides us several benefits such us to generate diagrams on a high abstraction level in order to identify security lacks in an easy way and to include new security constraints which solve these identified problems. Transformation rules are then applied obtaining an improved logical model and the final implementation. By using the MDA philosophy the system can be also migrate to different technologies (MOLAP, ROLAP, HOLAP, etc.) and different final tools. Since most DWs are managed by OLAP tools using a multidimensional approach (MOLAP), in this section we present a modernization process focused on the multidimensional path obtaining conceptual models from multidimensional logical models (Figure 1).

In a first stage, the multidimensional logical model according to SECMDDW is obtained from the source code of the OLAP tool. To achieve this goal is applied a static analysis (Canfora and Penta 2007) which is a reengineering method based on the generation of lexical and syntactical analyzers for the specific tool. In this way, code files are analyzed and a set of code-to-model transformations create the corresponding elements into the target logical model.

Once logical multidimensional model is obtained several set of QVT rules carry out a model-to-model transformation towards the corresponding conceptual model. Since the source metamodel (SECMDDW) presents three kinds of models (roles configuration, cubes and dimensions) three sets of transformations have been developed (Figure 2). Each transformation is composed of several QVT relations which are focused on transforming structural and security issues.

Role2SECDW transformation creates the security configuration of the system based on a set of security roles. This is an example of a semantic gap between abstractions levels, because conceptual level is richer than logical level and includes support to the definition of security levels, roles and compartment. This transformation presents the relations "RoleFiles2Package" and "Role2SRole" which transform the "RoleFiles" into a "Package" and create security roles "SRole" for each role detected at the logical level. Figure 3 shows the implementation of this transformation and Figure 4 the graphical representation for the "Role2SRole" relation.

Cube2SECDW transformation analyzes cube models and generates at the conceptual level structural aspects and security constraints defined over the multidimensional elements. Table 1 (left column) shows the signatures for the relations included in this transformation.

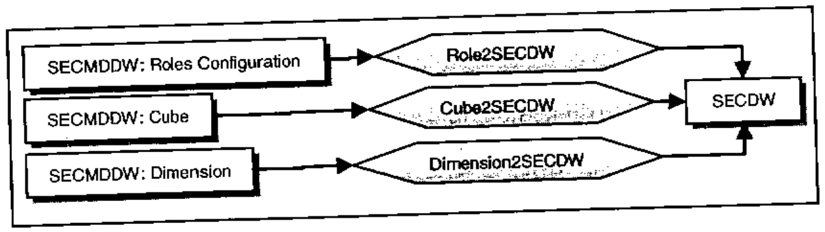


Fig. 2. PSM to PIM transformation overview


```
transformation Role2SECDW (psm:SECMDDW, pim:SECDW) {
  key SECDW::SRole {rootPackage, name};
  top relation RoleFiles2Package {
      xName : String;
      checkonly domain psm rf:SECMDDW::SecurityConfiguration::RoleFiles {
            name = xName };
      enforce domain pim pk:SECDW::Package { name = xName };
      where { rf.ownedRoles->forAll (r:SECMDDW::SecurityConfiguration::Role |
            Role2SRole(r, pk)); }  }
  relation Role2SRole {
      xName : String;
      checkonly domain psm r:SECMDDW::SecurityConfiguration::Role { ID = xName };
      enforce domain pim pk: SECDW::Package{
            ownedMember = sr : SECDW::SRole { name = xName } };  }}
```

Fig. 3. Role2SECDW transformation

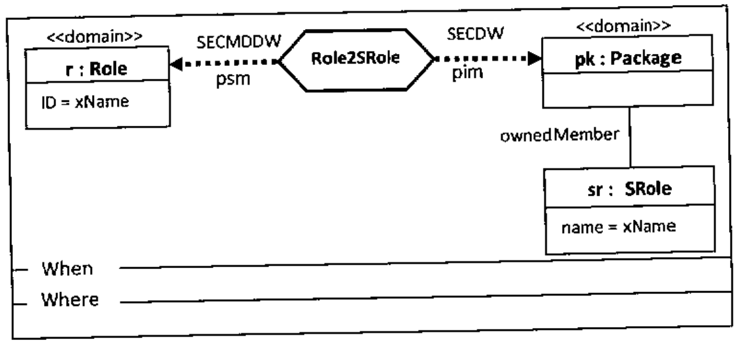


Fig. 4. Graphical representation of Role2SRole relation

Table 1. Relations for Cube2SECDW and Dimension2SECDW transformations

transformation Cube2SECDW	**transformation Dimenssion2SECDW**
top relation CubeFiles2Package {...} relation Cube2SFact {...} relation Measures2SFA {...} relation Measure2SProperty {...} relationDimension2SDimension{...}	top relation DimensionFiles2Package {...} relation Dimension2SDimension {...} relation attribute2SProperty {...} relation hierarchy2SBase {...} realtion attribute2SBaseProperty {...}
relation CubePermission2SClass {...} relation CellPermission2SProperty{...}	relation DimensionPermission2SClass {...} relation AttributePermission2SProperty{...}

There are a set of structural rules which transform cubes into secure fact classes ("Cube2SFact" relation) and their related measures and dimensions into secure properties ("Measures2SFA" and "Measure2Property" relations) and secure dimension classes ("Dimension2SDimension" relation). Security permissions related with cubes or cells are transformed into security constraints at the conceptual level ("CubePermission2SClass" and "CellPermission2SProperty" relations).

```
transformation Cube2SECDW (psm:SECMDDW, pim:SECDW) {
  key SECDW::SFact {rootPackage, name};
  top relation CubeFiles2Package {
        xName : String;
        checkonly domain psm cf:SECMDDW::Cubes::CubeFiles { name = xName };
        enforce domain pim pk:SECDW::Package { name = xName };
        where { cf.ownedCubes->forAll (c:SECMDDW::Cubes::Cube | Cube2SFact(c, pk)); } }
  relation Cube2SFact {
        xName : String;
        checkonly domain psm c:SECMDDW::Cubes::Cube { ID = xName };
        enforce domain pim pk: SECDW::Package {
              ownedMember = f : SECDW::SFact { name = xName } };
        where { c.ownedMeasureGroups->forAll (mg:SECMDDW::Cubes::MeasureGroup |
        (mg.ownedMeasures->forAll (m:SECMDDW::Cubes::Measure | Measures2SFA(m, f)));}}
  relation Measures2SFA {
        xName : String;
        checkonly domain psm m:SECMDDW::Cubes::Measure { ID = xName };
        enforce domain pim f:SECDW::SFact {
              attributes = sfa:SECDW::SFA { name = xName } };  }}
```

Fig. 5. Cube2SECDW transformation

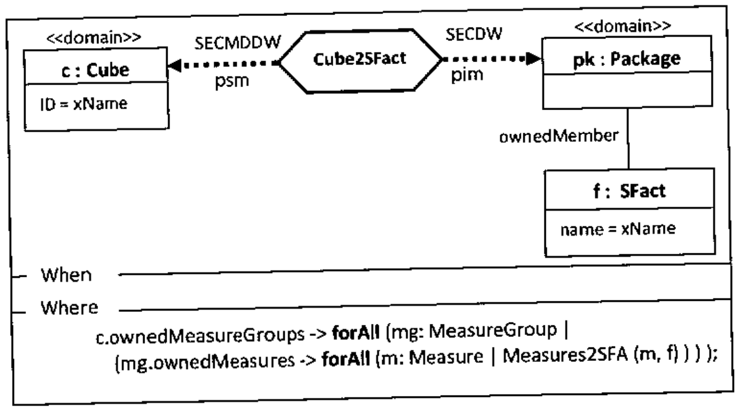


Fig. 6. Graphical representation of Cube2SFact relation

The implementation of some relations is shown in Figure 5 and Figure 6 presents the "Cube2SFact" relation in a graphical way.

Dimension2SECDW transformation focuses on dimension models and creates at the conceptual level structural aspects such as dimension and base classes, properties and hierarchies ("Dimension2SDimension", "attribute2SProperty", "hierarchy2SBase" and "attribute2SBaseProperty" relations) and security constraints related with dimensions, bases and properties ("DimensionPermission2SClass" and "AttributePermission2SProperty" relations). This transformation is composed of several relations which signatures are shown in Table 1 (right column).

The implementation of some relations is shown in Figure 7 and Figure 8 presents the "DimensionPermission2SClass" relation in a graphical way.


```
transformation Dimension2SECDW (psm:SECMDDW, pim:SECDW) {
  key SECDW::SDimension {rootPackage, name};
  key SECDW::SRole {rootPackage, name};
  top relation DimensionFiles2Package {
        xName : String;
        checkonly domain psm df:SECMDDW::Dimensions::DimensionFiles { name = xName };
        enforce domain pim pk:SECDW::Package { name = xName };
        where { df.ownedDimensions->forAll (d:SECMDDW::Dimensions::Dimension |
              Dimension2SDimension(d, pk)); } }
  relation Dimension2SDimension {
        xName : String;
        checkonly domain psm d:SECMDDW::Dimensions::Dimension {ID = xName };
        enforce domain pim pk: SECDW::Package {
              ownedMember = sd : SECDW::SDimension {
              ownedSecInf = si : SECDW::SecureInformation { }, name = xName } };
        where { d.ownedDimensionPermissions->forAll
        (dp:SECMDDW::Dimensions::DimensionPermission |
        (dp.deniedSet.oclIsUndefined()) implies (DimensionPermission2SClass (dp, si, pk)) ); } }
  relation DimensionPermission2SClass {
        xRoleID : String;
        checkonly domain psm dp:SECMDDW::Dimensions::DimensionPermission {
              roleID = xRoleID };
        enforce domain pim sd :SECDW::SecureInformation {
              securityRoles = sr : SECDW::SRole { name = xRoleID } };
        enforce domain pim pk:SECDW::Package { ownedMember = sr : SECDW::SRole {} };
        when{ dp.deniedSet = ''; } }}
```

Fig. 7. Dimension2SECDW transformation

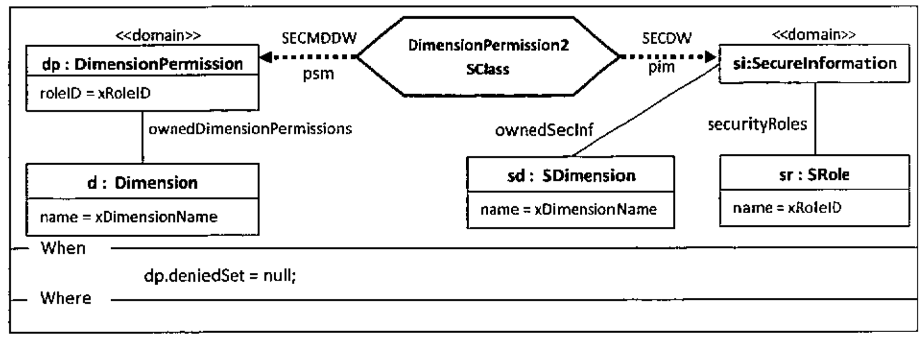


Fig. 8. Graphical representation of DimensionPermission2SClass relation

5 Example

This section shows the defined ADM process by using an example in which the transformation rules are applied into a PSM multidimensional model to obtain the corresponding PIM model. This example uses a DW which manages airport's information about trips involving passengers, baggage, flights, dates and places. This information is analyzed for the airport staff, companies or passengers, and can be used for many purposes, for instance companies can decide to reinforce certain routes with a great number of passengers or can offer to passengers a special price for their top

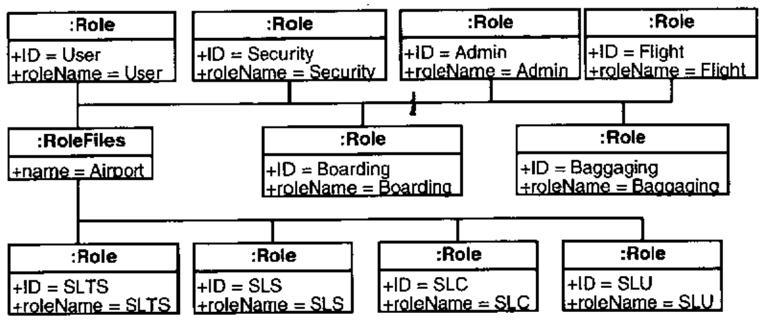


Fig. 9. PSM multidimensional model for security configuration

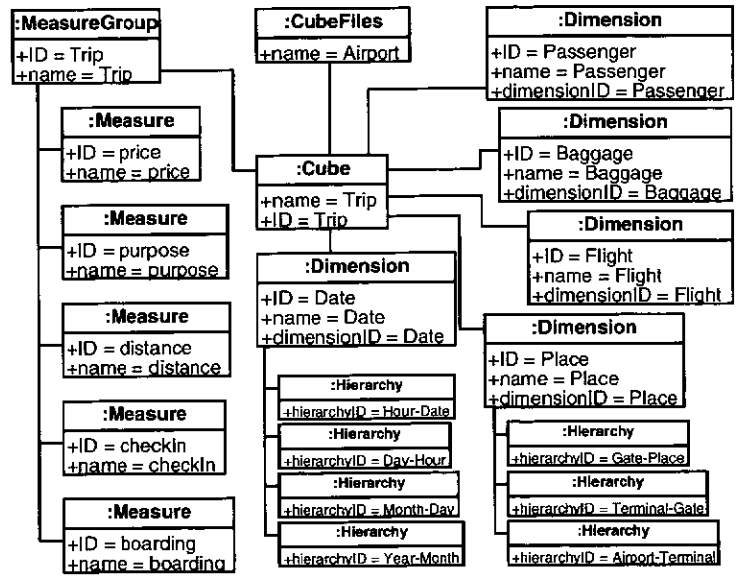


Fig. 10. PSM multidimensional model for cubes

destinations. The source multidimensional PSM model is composed of three parts: security configuration (Figure 9), cubes (Figure 10) and dimensions (Figure 11). Figure 12 finally shows the PIM model obtained after applying the ADM process.

Firstly, **Role2SECDW** transformation analyzes the security configuration model (Figure 9) and creates roles in the PIM model. The conceptual level (PIM) is richer and supports the specification of security levels, compartments and roles, but logical models (PSM) only include information of roles. Thus, transformation rules can only transform each role in the logical model into a role in the conceptual model.

Then, logical cube models (Figure 10) are processed by the **Cube2SECDW** transformation. It creates in the PIM model (Figure 12) the following structural aspects: the secure fact class "Trip", its measures and its related dimensions and hierarchies. Since security permissions related with cubes were not defined, security constraints are not established in the PIM model.

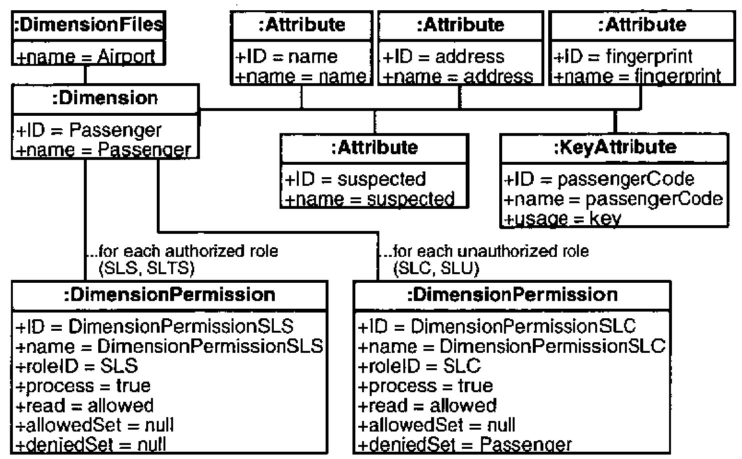


Fig. 11. PSM multidimensional model for dimensions

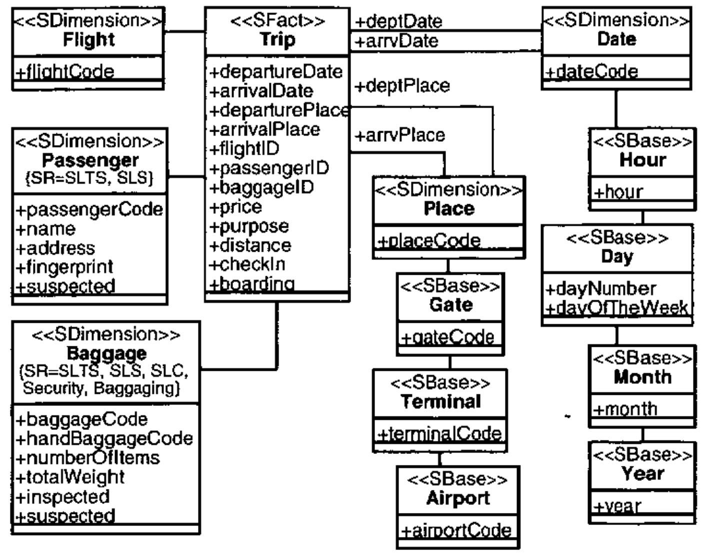


Fig. 12. PIM model

Finally, **Dimension2SECDW** process logical dimension models. Figure 11 shows the PSM model for "Passenger" dimension in which have been defined some attributes and dimension permissions to authorize and deny accesses to certain roles. This structural information is transformed into a secure dimension class "Passenger" with secure properties in the PIM model (Figure 12). Positive security permissions are also transformed by including the authorized roles ("SLTS" and "SLS") as stereotypes of the "Passenger" dimension.

6 Conclusions

We have proposed an MDA architecture for developing secure DWs taking into account security issues from early stages of the development process. We provide security models at different abstraction levels and automatic transformations between models and towards the final implementation.

This work has fulfilled the architecture providing an architecture-driven modernization (ADM) process which allows us to automatically obtain higher abstraction models (PIM). Firstly, code analyzers obtain the logical model from the implementation, and then, QVT rules transform this logical model into a conceptual model. In this way, existing systems can be re-documented and this design at higher abstraction level (PIM) can be easier analyzed in order to include new security constraints. Furthermore, once PIM model is obtained the DW can be migrated to other platforms or final tools.

Our further works will improve this architecture in several aspects: dealing with the inference problem by including dynamic security models which complement the existing models; including new PSM models (such as XOLAP); and giving support to other final platforms (such as Pentaho).

Acknowledgments. This research is part of the ESFINGE (TIN2006-15175-C05-05) Project financed by the Spanish Ministry of Education and Science, the QUASIMODO (PAC08-0157-0668) Project financed by the FEDER and the Regional Science and Technology Ministry of Castilla-La Mancha (Spain), the SISTEMAS (PII2109-0150-3135) Project financed by the Regional Science and Technology Ministry of Castilla-La Mancha (Spain) and the MITOS (TC20091098) Project financed by the University of Castilla-La Mancha (Spain).

References

Aiken, P.H.: Reverse engineering of data. IBM Syst. J. 37(2), 246–269 (1998)

Basin, D., Doser, J., et al.: Model Driven Security: from UML Models to Access Control Infrastructures. ACM Transactions on Software Engineering and Methodology 15(1), 39–91 (2006)

Blaha, M.: A Retrospective on Industrial Database Reverse Engineering Projects-Part 1. In: Proceedings of the 8th Working Conference on Reverse Engineering (WCRE 2001), Suttgart, Germany. IEEE Computer Society Press, Los Alamitos (2001)

Blanco, C., García-Rodríguez de Guzmán, I., et al.: Applying QVT in order to implement Secure Data Warehouses in SQL Server Analysis Services. Journal of Research and Practice in Information Technolog (in press) (2008)

Canfora, G., Penta, M.D.: New Frontiers of Reverse Engineering. IEEE Computer Society, Los Alamitos (2007)

Cohen, Y., Feldman, Y.A.: Automatic high-quality reengineering of database programs by abstraction, transformation and reimplementation. ACM Trans. Softw. Eng. Methodol. 12(3), 285–316 (2003)

CWM, OMG: Common Warehouse Metamodel (CWM) (2003)

Chikofsky, E.J., Cross, J.H.: Reverse Engineering and Design Recovery: A Taxonomy. IEEE Softw. 7(1), 13–17 (1990)

Fernández-Medina, E., Trujillo, J., et al.: Model Driven Multidimensional Modeling of Secure Data Warehouses. European Journal of Information Systems 16, 374–389 (2007)

Fernández-Medina, E., Trujillo, J., et al.: Access Control and Audit Model for the Multidimensional Modeling of Data Warehouses. Decision Support Systems 42, 1270–1289 (2006)

Fernández-Medina, E., Trujillo, J., et al.: Developing secure data warehouses with a UML extension. Information Systems 32(6), 826–856 (2007)

Hainaut, J.-L., Englebert, V., et al.: Database reverse engineering: From requirements to CARE tools. Applied Categorical Structures. SpringerLink. 3 (2004)

Jürjens, J.: Secure Systems Development with UML. Springer, Heidelberg (2004)

Khusidman, V., Ulrich, W.: Architecture-Driven Modernization: Transforming the Enterprise. DRAFT V.5, OMG: 7 (2007), `http://www.omg.org/docs/admtf/07-12-01.pdf`

Lodderstedt, T., Basin, D., Doser, J.: SecureUML: A UML-based modeling language for model-driven security. In: Jézéquel, J.-M., Hussmann, H., Cook, S. (eds.) UML 2002. LNCS, vol. 2460, p. 426. Springer, Heidelberg (2002)

MDA, OMG: Model Driven Architecture Guide (2003)

Müller, H.A., Jahnke, J.H., et al.: Reverse engineering: a roadmap. In: Proceedings of the Conference on The Future of Software Engineering, Limerick, Ireland. ACM Press, New York (2000)

OMG. MOF QVT final adopted specification

OMG, ADM Glossary of Definitions and Terms, OMG: 34 (2006), `http://adm.omg.org/ADM_Glossary_Spreadsheet_pdf.pdf`

Priebe, T., Pernul, G.: A pragmatic approach to conceptual modeling of OLAP security. In: Kunii, H.S., Jajodia, S., Sølvberg, A. (eds.) ER 2001. LNCS, vol. 2224, p. 311. Springer, Heidelberg (2001)

Soler, E., Trujillo, J., et al.: A Set of QVT relations to Transform PIM to PSM in the Design of Secure Data Warehouses. In: IEEE International Symposium on Frontiers on Availability, Reliability and Security (FARES 2007), Viena, Austria (2007)

Soler, E., Trujillo, J., et al.: Building a secure star schema in data warehouses by an extension of the relational package from CWM. Computer Standard and Interfaces 30(6), 341–350 (2008)

Thuraisingham, B., Kantarcioglu, M., et al.: Extended RBAC-based design and implementation for a secure data warehouse. International Journal of Business Intelligence and Data Mining (IJBIDM) 2(4), 367–382 (2007)

Trujillo, J., Soler, E., et al.: An Engineering Process for Developing Secure Data Warehouses. Information and Software Technology (in Press) (2008)

Trujillo, J., Soler, E., et al.: A UML 2.0 Profile to define Security Requirements for DataWarehouses. Computer Standard and Interfaces (in Press) (2008)

Yu, E.: Towards modelling and reasoning support for early-phase requirements engineering. In: 3rd IEEE International Symposium on Requirements Engineering (RE 1997), Washington, DC (1997)