Applying the i* framework to the development of data warehouses*

Jose-Norberto Mazón[1], Jesús Pardillo[1], Emilio Soler[2], Octavio Glorio[1], and Juan Trujillo[1]

[1] Universidad de Alicante, Spain
{jnmazon,jesuspv,oglorio,jtrujillo}@dlsi.ua.es
[2] Universidad de Matanzas, Cuba
emilio.soler@umcc.cu

Abstract. Data warehouse design has been traditionally guided by an in-depth analysis of the underlying operational data sources, thus overlooking an explicit stage in which information requirements of decision makers are addressed. This scenario has prompted that the deployed data warehouse often fails in delivering the expected support of the decision making process. To overcome this problem, we propose to use the i* framework for modeling goals and requirements within our *model driven architecture* (MDA) approach for the development of data warehouses. Our current and short term research also includes the reconciliation between information requirements and data sources, and the modeling of quality-of-service requirements (e.g., security). Finally, an Eclipse-based tool is being implemented as a proof of concept of our research.

1 Introduction

Data Warehouse (DW) systems are used by decision makers to analyze the status and the development of an organization. These systems are based on large amounts of data integrated from heterogeneous sources into multidimensional (MD) models, which are special data models allowing data access in a way that comes more natural to human analysts. Generally speaking, designers depict data into facts and dimensions in a conceptual MD model. Facts are usually measures of business processes (e.g., how many products are sold, how many patients treated, how long something takes, etc.), and dimensions represent the context for analyzing these measures (e.g., time, customer, product, etc.).

Since the DW integrates several operational data sources, the development of conceptual MD models has been traditionally guided by their detailed analysis. However, several studies have pointed out that most of these conceptual MD models fail to address the required information as a result of a poor communication between DW developers and decision makers. Actually, information

* This work has been supported by the ESPIA (TIN2007-67078) project and by the grants AP2005-1360 and AP2006-00332 from the Spanish Ministry of Science, and by the DADS (PBC-05-012-2) project from the Castilla-La Mancha Ministry of Science.

needs cannot be understood by only analyzing the operational data sources, and a requirement analysis stage is needed in order to model the information requirements of decision makers. Moreover, from this requirement model, a suitable conceptual MD model can be derived and reconciled with the available data sources. Importantly, this stage should be based on a *goal-oriented requirement engineering* (GORE) framework since (i) the DW aims at providing adequate information to support the decision making process, thus helping to fulfil goals of an organization, (ii) requirements for DWs are difficult to specify from scratch, since decision makers often only express general expectations about which goals the DW should support, and (iii) DW systems have a lot of kind of stakeholders with different interrelated goals that must be modeled to easily obtain a conceptual MD model that satisfy them.

The remainder of this contribution is structured as follows: objectives of our research are described in the next section. Our scientific contributions are presented in section 3. Our conclusions are reported in section 4. Finally, in section 5, ongoing and future work are sketched.

2 Objectives of the research

Our research is focused on defining a GORE approach, based on *i**, for modeling goals that the DW supports, thus easier obtaining information requirements. Furthermore, this approach is combined with our *model driven architecture* (MDA) framework for the development of DWs that has been described in [4]. This framework is based on defining a *computation independent model* (CIM) which addresses goals and requirements, a *platform independent model* (PIM) to specify MD properties at the conceptual level, and a *platform specific model* (PSM) tailored to a specific database technology. Therefore, *i** is used for defining a CIM, while a PIM for MD modeling is derived by establishing a formal transformation between these models via the *query/view/transformation* (QVT) language. The main advantage is that the conceptual MD model, represented in a PIM, meets every goal and requirement defined in the CIM. Furthermore, this PIM obtained from requirements is reconciled with data sources to obtain a hybrid PIM that provides the adequate information to fulfil business goals without disagreeing with data sources [3]. It is worth noting that combining *i** and MDA in DW development, via the use of the QVT language, assures the traceability between goals, requirements and the necessary MD elements related to them. This is an advantage of our proposal, since other works only propose informal guidelines to obtain a conceptual MD model from information requirements which also prevents the automatization of the process.

Finally, the DW is not just *data* but a whole system, where users require that the information has some characteristics when it is provided (security, performance tuning, etc.). These characteristics are constraints that the DW must fulfil to satisfy user expectations. We have named them quality-of-service (QoS) requirements, because they are additional issues that must be fulfilled by the DW to add quality in the way that the information is supplied and used. Infor-

mally speaking, information requirements answer *what* information the DW is expected to provide, and QoS requirements answer *how* this information should be provided for a right use. Therefore, these QoS requirements must be considered in the CIM by extending the i* notation. We have first focused on security requirements, since the extreme importance of the information managed by DW systems makes essential to specify security issues from the early stages of the MD modeling process, and enforce them [5].

3 Scientific contributions

To fulfil our research objectives, the *i** modeling framework and MDA have been integrated via the profiling mechanism of the *unified modeling language* (UML). In this way, *i** has been adapted to requirement analysis in DWs, allowing us to model both information [1, 2] and security requirements [6] at the CIM level (see Fig. 1). Moreover, in [3], we have developed an approach for reconciling data sources and requirements based on a set of *multidimensional normal forms* which assure several desirable properties in the conceptual MD model.

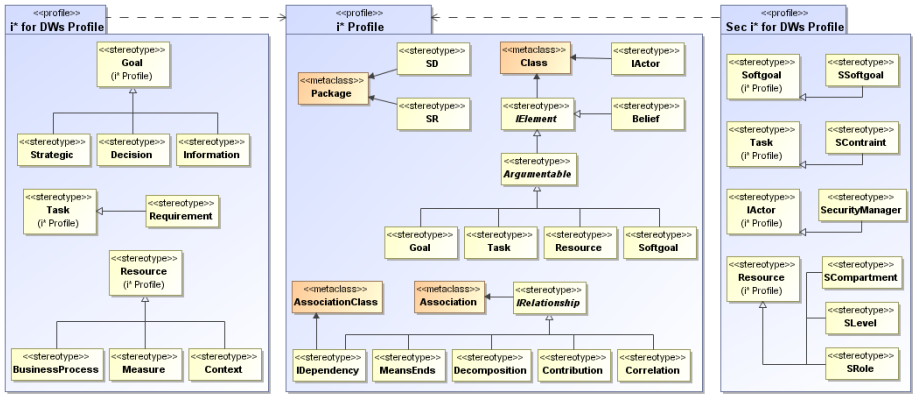

Fig. 1. UML profiles for i* modeling in the DW domain.

This UML profile for i* has been implemented in an Eclipse-based tool that provides support for our MDA-based approach for the development of DWs (see Fig. 2). By using this tool we can define a CIM and apply a set of QVT transformations to obtain the corresponding PIM.

4 Conclusions

DW projects overlook an explicit requirement analysis phase when MD models are defined. Therefore, DW fails to give the adequate support to decision making.

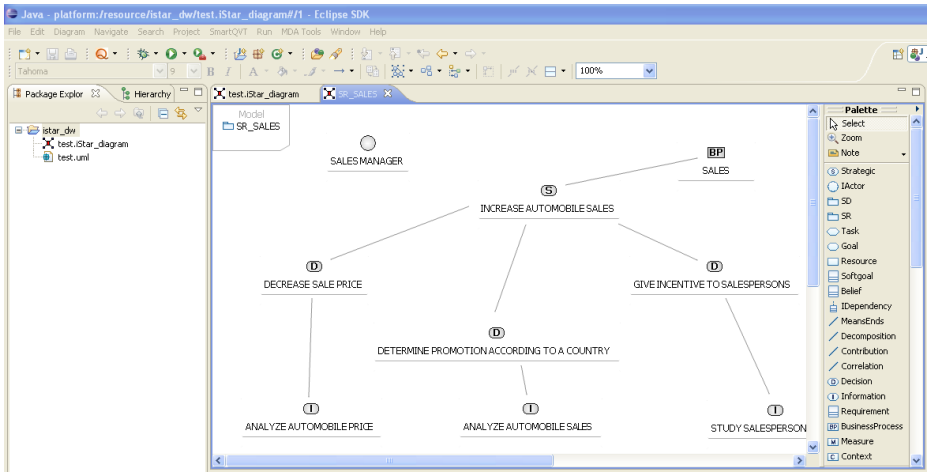

Fig. 2. Snapshot from our tool for i* modeling in the DW domain.

Our research aims to use the *i** framework within our MDA approach for the development of DW in order to avoid this important drawback.

5 Ongoing and future work

Our immediate future work comprises the improvement of our proposal by considering other further issues of GORE (e.g., more complex mechanisms for reasoning about goals or prioritization of goals). Furthermore, other QoS issues (apart from security) should be considered.

References

1. Mazón J.N., Trujillo J., Serrano M., Piattini M.: Designing data warehouses: from business requirement analysis to multidimensional modeling. REBNITA 2005
2. Mazón J.N., Pardillo J., Trujillo J.: A Model-Driven Goal-Oriented Engineering Approach for Data Warehouses. RIGiM 2007. LNCS 4802, pp. 255–264
3. Mazón J.N., Trujillo J., Lechtenbörger J.: Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms. Data Knowl. Eng. **63**(3) (2007) 725–751
4. Mazón, J.N., Trujillo, J.: An MDA approach for the development of data warehouses. Decision Support Systems (in press, **doi:10.1016/j.dss.2006.12.003**)
5. Soler E., Trujillo J., Fernández-Medina E., Piattini M.: A Framework for the Development of Secure DWs based on MDA and QVT. ARES 2007, pp. 294-300
6. Soler E., Stefanov V., Mazón J.N., Trujillo J., Fernández-Medina E., Piattini M.: Towards Comprehensive Requirement Analysis for Data Warehouses: Considering Security Requirements. ARES 2008