PROCEEDINGS OF

REBNITA 2005

1ST INTERNATIONAL WORKSHOP ON REQUIREMENTS ENGINEERING FOR BUSINESS NEED AND IT ALIGNMENT

29-30 AUGUST 2005
THE SORBONNE, PARIS at RE'05

REBNITA is supported by National ICT Australia Ltd
www.nicta.com.au

Editors: Karl Cox, Eric Dubois, Yves Pigneur, Steven J. Bleistein, June Verner, Alan M. Davis, Roel Wieringa

ISBN: 0 7334 2276 4
University of New South Wales Press

Contents

Designing Data Warehouses: From Business Requirement Analysis to Multidimensional Modeling

Jose-Norberto Mazón[1], Juan Trujillo[1], Manuel Serrano[2], and Mario Piattini[2]
[1]*Dept. of Software and Computing Systems*
University of Alicante
Apto. Correos 99. E-03080
{jnmazon,jtrujillo}@dlsi.ua.es
[2]*Alarcos Research Group*
University of Castilla – La Mancha
Paseo de la Universidad, 4; 13071 Ciudad Real
{Manuel.Serrano,Mario.Piattini}@uclm.es

Abstract

Most of the data warehouse projects still fail because the final data warehouse does not properly meet business goals. Designers start a data warehouse project with the conceptual or logical modeling of the multidimensional schema, and unfortunately, not much attention has been paid on the requirement analysis phase. However, this phase is very important, because it can include the understanding of the business context in which the data warehouse is pretended to work. This is a crucial issue, since the aim of data warehouses is to provide enough information in a suitable way to improve decision making and accomplish with business goals. In this paper, we propose an approach to take into account business context and their business goals in data warehouse requirement analysis phase. First of all, we adapt i notation to model business environment and goals for data warehouses requirement analysis. Then, from i* models, a multidimensional model which satisfies business goals is obtained. To avoid an arbitrary use of our approach, we provide a set of guidelines to correctly specify i* diagrams and transform them into a multidimensional model. Finally, we apply our approach to a case study to show its benefit.*

1. Introduction

Even though a decade later, data warehouses still pay a central role in current decision support systems, since they are oriented to provide adequate information to improve the decision making process [4]. Nowadays, it is widely accepted that the basis for designing the data warehouse repository is the multidimensional modeling [4,5,6,8]. Conceptual models have been provided to be able to represent main properties of the multidimensional modeling that satisfy final user requirements [5,8]. Nevertheless, even though we use conceptual models, many data warehouse projects still fail because traditionally not much attention has been paid on the requirement analysis phase. Therefore, the final data warehouse may not reflect organization needs and may not deliver the expected support of the decision making process [3,12]. This process is crucial in organizations, since making better decisions allows them to improve business processes by achieving business goals. Moreover, several studies have shown that more than 80% of data warehouse projects fail to meet business goals [12]. Often, business goals are ignored as a result of poor communication between IT and business professionals during requirement analysis. Therefore, it is obvious that an effort is needed to develop data warehouses within a business context by incorporating explicit understanding of the business into data warehouse requirement analysis using some organizational modeling technique [19]. Then designers will be able to develop data warehouses that provide organizations with the necessary information to fulfill their business goals. A summary of main benefits that organizations pretend to achieve with the use of a data warehouse can be viewed in figure 1. These benefits can be achieved if the data warehouse is understood within its business environment.

Thus, we present an approach for including business issues (i.e. business goals) in data warehouse requirement analysis and then, transform requirements in a multidimensional model that helps to fulfill business goals. Since the i* technique provides understanding of the organizational environment and goals in requirement analysis phase [19], we adapt i* diagrams to requirement analysis in data warehouses. We also structure business goals that data warehouse helps to achieve into strategic, decision and information goals. This allows developers to have a better understanding of business and, then users can communicate better their ideas. Finally, from these i* diagrams, we obtain a multidimensional conceptual model which provides organization with the adequate information to fulfill business goals. This model is designed using our UML (Unified Modeling Language) profile [7,8,14]. Furthermore, we provide, based on our experience in designing real world data warehouses, a set of design guidelines to correctly specify these i* diagrams and transform them into the corresponding multidimensional model.

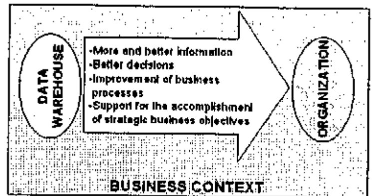


Figure 1. Some benefits derived from developing the data warehouse within a business context (adapted from [1]).

The rest of this paper is structured as follows. Section 2 presents the most relevant related work for requirement analysis in data warehouses. Our approach for requirement analysis and its guidelines for properly specify the i* and multidimensional models are presented in section 3. We show the benefit of our approach in section 4 by means of a little case study. Finally, in Section 5 we present our conclusions and sketch some future works.

2. Related Work

In this section, we will make a brief description of the most relevant approaches for requirement analysis in data warehouses. We want to point out that, most of approaches have a main drawback: they are not part of a global methodology in which we can directly obtain, from the requirements, the corresponding conceptual multidimensional schema that provides adequate information to fulfill business goals.

Böhnlein et al. [2] derive a data warehouse from business process models. They point out that it is relevant to focus on goals and strategies of the company for an efficient decision making, since this information cannot be only extracted by analyzing operational data sources. Therefore, it is crucial to situate the data warehouse within a business context and analyze this context. However, as Winter and Strauch [17,18] point out, only a detailed business process analysis is not feasible because decision processes consist of tasks which are often unique and unstructured, and decision makers often refuse to disclose their process in detail. Thus, they present a methodology based on determining information requirements which data warehouse users need in decision processes, and matching information requirements with actual information supply (operational sources), because these seem to be more stable, more concrete, and better accessible. Nevertheless, they do not use any notation to represent business goals and understand the business context in which the data warehouse works.

Schiefer et al. [12] present a method, easyREMOTEDWH (easy Requirements Modeling Technique for Data Warehouses), which considers data warehouse requirements from different stakeholders perspectives, according to several levels of abstraction. They include an interesting business point of view to represent business objectives and needs. Unfortunately, they do not present a notation or guidelines to properly specify requirements.

Prakash et al. [11] also propose a requirement elicitation process for data warehouses grouping requirements in several levels of abstraction. Their process consists of identifying information that support decision making via information scenarios. The process starts with the determination of the goals of an organization. Secondly, the decision making needs are specified, and finally, the information needed to cover these decisions is identified. In this process, they use a Goal-Decision-Information (GDI) diagram. Although they show how to obtain the GDI diagram and the information scenarios, the relationships between information scenarios and requirements are not properly specified. Moreover, they only represent interaction between decision makers and the data warehouse, and how to obtain organization needs are not considered. Finally, they lack in using guidelines to specify data warehouse requirements.

(information needed by user is provided by the data warehouse).

Although SD models describe business environments and dependencies between the users and the data warehouse, only external relationships among actors are showed. However, intentional constructs within each actor stay hidden [20]. On the other hand, the SR model provides a more detailed level of modeling internal intentional relationships of each actor. Intentional elements (goals, tasks, resources and softgoals) and their relationships (means-end and task-decomposition) are represented. Regarding data warehouses, we are interested in represent goals, tasks and resources as intentional elements. Following, we explain how to build the SD and the SR models for data warehouses.

3.3.2. Building the Strategic Dependency Model

Several guidelines to build the SD model for data warehouses are given. These guidelines are based on representing actors and dependencies between them.

Guideline 1. Discover business actors. These actors are decision makers (e.g. managers, top executives...). The data warehouse under construction is also considered as an actor. We have to represent these actors in a SD model.

Guideline 2. Determine strategic goals of organization from decision makers. These goals must be represented by means of goal dependencies between every actor and the data warehouse.

Guideline 3. Information required by decision makers is represented as a resource dependency between each actor and the data warehouse, since this information is provided by the data warehouse.

3.3.3. Building the Strategic Rationale Model

Guidelines to build SR model for data warehouses are given. These guidelines are based on representing internal intentional elements and relationships. Here, we also specify dependencies between actors with a more level of detail.

Guideline 4. For each actor who is a decision maker, intentional elements are obtained (in this case, goals and tasks). Several guidelines are given for obtaining and representing them.

Guideline 4.1. Refine main strategic goals (obtained in guideline 2), following a top-down strategy in order to obtain possible strategic subgoals. We have to keep on refining until obtaining decision goals. Strategic and decision goals are represented as goals. Relationships between them are represented as means-end links, since these links are used to describe how goals are achieved.

Guideline 4.2. Refine decision goals (obtaining subgoals) until obtaining information goals (top-down strategy). Each of these goals is represented as a goal. Relationships between them are represented as means-end links.

Guideline 4.3. Each information goal previously obtained is related to the analysis of some measure used to achieve that goal. This analysis describes an information requirement and it is represented as a task. Decision makers carry out this task in order to obtain information from the data warehouse to achieve required information goals.

Guideline 5. For the data warehouse actor, every task and resource (and their relationships) needed in order to provide adequate information (according to the previous guideline) is represented.

Guideline 5.1. For each resource dependency according to guideline 3, providing the adequate information is the objective for data warehouse actor. Then a goal is required to provide such information.

Guideline 5.2. Measures according to guideline 4.3, must be represented. These measures are represented as resources. However, if they are derived measures, then they are presented as tasks in order to calculate them. Both, resources and tasks, are linked to main goal with a means-end relationship. Measures needed to calculate derived attributes are represented as resources (linked by means of a decomposition link to its corresponding task).

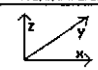
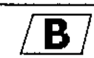
Guideline 5.3. Analysis of each measure must be provided within a context. This is represented as tasks. Every task is related to main goal with a means-ends relationship. Within context of analysis, there are several levels of aggregation to analyze measures. These levels of aggregation are represented as resources. These resources are linked to every task that represents the context of analysis by means of a decomposition link.

3.4. From i* Model to Multidimensional Model

We must be sure that each of the tasks and resources reflected in the SD model for data warehouse actor must be addressed by a multidimensional model. This model must be useful to fulfill business goals.

In this paper we follow our UML profile for the conceptual design of data warehouses following the multidimensional paradigm [7,8,15]. The most important feature of this paradigm is dividing data into facts (composed of measures) and dimensions; to provide data on a suitable level of granularity, hierarchies are defined on the dimensions. This profile is defined by a set of stereotypes and tagged values to elegantly represent these main multidimensional properties at the conceptual level using a UML class diagram (see table 1). Due to lack of space we refer reader to [7,8,15] for a further explanation.

Table 1. Main stereotypes of the UML profile.

Stereotype	Description	Icon
Fact class	Represent facts consisting of measures	
Dimension class	Represent dimensions consisting of dimension attributes and hierarchy levels	
Base class	Represent dimension hierarchy levels	

Following, we describe several guidelines to specify a multidimensional class diagram by using our UML profile [7,8,15] for the multidimensional modeling at the conceptual level. This conceptual multidimensional schema is defined from i* models by identifying fact and dimension classes with their corresponding base classes (i.e. classification hierarchies) from the SD model. Identifying attributes within fact and dimension classes should be completed from operational sources.

The following guidelines are used to define a multidimensional class diagram from a SD model:

Guideline 6. Create a fact class for each main goal in the data warehouse actor. For each resource representing a measure we create an attribute. For each task representing a derived measure we create a derived attribute.

Guideline 7. Resources that represent the context of analysis become dimension classes.

Guideline 8. Levels of aggregation (i.e. base classes) are also specified from resources which represent the context of analysis. We want to point out that these base classes have not any attribute. due to the fact that these attributes stay in the operational sources. In this paper, modeling operational sources is not still considered, so enriching levels of hierarchies with attributes from operational sources will be considered in a future research. However, this model can be used as a prototype in order to know if business goals can be achieved by information requirements.

4. Case Study

The aim of this section is to exemplify the usage of our requirement analysis approach. We have selected a case study presented in Chapters 2 and 3 of Kimball's book [6] to show how we can obtain requirements within a business context.

Kimball's retail case study presents a brief description of the retail business which embraces both retail sales and inventory. This retail business is composed of several grocery stores spread over several regions. In each store several products are sold. At the grocery store, management is concerned with the logistics of ordering, stocking, and selling products while maximizing profit. The profit ultimately comes. among other things. attracting as many customers as possible in a highly competitive pricing environment. Thus, some of the most significant management decisions have to do with pricing and promotions used to increase the number of customer, since they include temporary price reductions in a grocery store. One of the most important tasks of managers is to determine whether a promotion is effective or not. Therefore, retail sales business process deals with analyzing what quantity of products are selling in which stores on what days under what promotional conditions.

In this case study, Kimball deals with several kinds of inventory models of a store. We are interested in the inventory snapshot example, where the inventory levels are measured every day and are placed in separate records in the database. Main management objective is making decisions to optimize inventory levels in order to decrease inventory costs. These decisions are related to make sure the right product is in the right store at the right time to minimize out-of-stocks (where the products is not available on the shelf to be sold) and reduce overall inventory carrying costs. So, the inventory management needs the ability to analyze daily quantity-on-hand inventory levels by product and store. Inventory manager is also concerned with measure the velocity of inventory movement

Guideline 7. Dimension classes are elicited from resources which represent the context of analysis. In this case, we have to create the following dimension classes linked to "sales" fact class: "store", "date", "product", and "promotion". We have also have to link "store", "date", and "product" to "inventory" fact class.

Guideline 8. Specify levels of aggregation by means of resources which represent the context of analysis. Levels of aggregation are defined with base classes. They are all represented in figure 7.

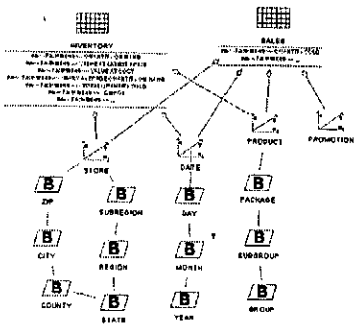


Figure 7. Multidimensional class diagram created from the i* models.

Once we have the multidimensional schema, if we go back to the first main information requirement above-describe: "analyzing what quantity of products are selling in which stores on what days under what promotional conditions", we can easily see that this requirement can be answered by navigating the obtained multidimensional schema of figure 7. In concrete, the measure specified in the fact class comes from the resource "quantity sold", and the dimension classes and base classes come from resources "product", "date", "promotion", and "store" (see figure 6). For an overview of what multidimensional elements are created to fulfill each goal, see table 2.

5. Conclusion and Future Work

In this paper, we have presented a requirement analysis approach for understanding the data warehouse within its business context. Business goals of organizations must be understood by designers to develop a data warehouse which properly support expected decision making and allow organizations to derive business value. In order to understand business environment, we represent actors and business goals which are pretended to be fulfilled with information provided by a data warehouse using i* technique. First of all, these actors, business goals, information and their relationships are modeled in a SD model. Secondly, we have structured data warehouse goals into strategic, decision and information goals. We have built a SR model to show these goals for each actor. In this SR model, we also represent tasks and resources that actors need to fulfill goals, so we properly represent the information requirements needed to achieve all goals. Finally, we have shown how to build a multidimensional schema from these i* models by following our own approach for the conceptual design of data warehouses with UML. This multidimensional schema provides required information for fulfilling business goals. Moreover, we have also provided, based on our experience in designing real world data warehouses, a set of guidelines to correctly specify i* models, thereby avoiding an arbitrary use of them.

Immediate planned future work involves formalizing and organizing proposed guidelines into a process model. It is also planned to add quality measures to these i* models to provide more objective indicators of quality. Then, these measures must be both formally and empirically validated. On the other hand, a future interesting experiment is focused on analyzing the understandability of these diagrams with stakeholders in real world data warehouse projects (like in [13]). This experiment will allow us to validate our approach. Furthermore, we plan to add softgoals in order to gather security and quality constraints [9]. Furthermore, we also consider adding the specification of operational data sources to this approach (by adding them as actors in SR and SD models). This allows us to enrich levels of hierarchies (i.e. base classes) with attributes. Further future works refer to provide an overall methodology for data warehouse design starting from the requirement analysis phase.

6. Acknowledgements

This work has been partially supported by the METASIGN project (TIN2004-00779) from the Spanish Ministry of Education and Science, by the DADASMECA project (GV05/220) from the Valencia Government, and by the MESSENGER (PCC-03-003-1) and DIMENSIONS (PBC-05-012-2) projects from the Regional Science and Technology Ministry of Castilla-La Mancha (Spain).

Table 2. Goals and their corresponding multidimensional elements which allow their achievement.

Goals	Multidimensional elements			
	Fact class	*Fact attributes (measures)*	*Dimension classes*	*Base classes (levels of aggregation)*
Increase number of customers	Sales	Quantity sold	Store, date, product, promotion	Store: ZIP, city, county, subregion, region, state. Date: day, month, year. Product: package, subgroup, group.
Decrease inventory costs	Inventory	Quantity on hand, quantity sold, value at latest price, value at cost, daily average quantity on hand, GMROI	Store, date, product	Store: ZIP, city, county, subregion, region, state. Date: day, month, year. Product: package, subgroup, group.

References

[1] Agosta, L. The Essential guide to Data Warehousing. Prentice Hall, New Jersey, 2000.

[2] Böhnlein M., Ulbrich-vom Ende, A. (2000). Business Process Oriented Development of Data Warehouse Structures. In: Proceedings of Data Warehousing 2000, Physica Verlag.

[3] Bruckner R., List B., Schiefer J. (2001). Developing requirements for data warehouse systems with use cases. In Proc. 7th Americas Conf. on Information Systems (AMCIS'01), Boston, USA (2001). 329-335.

[4] Inmon. W. Building the Data Warehouse (2nd Edition). New York. Wiley & Sons.

[5] Jarke, M., Lenzerini, M., Vassiliou, Y. and Vassiliadis, P. Fundamentals of Data Ware-houses, Ed. Springer. (2000).

[6] Kimball R., Ross M. (2002) The Data Warehouse Toolkit, second edition, John Wiley & Sons. (2002).

[7] Luján-Mora S., Trujillo J., Song I-Y. (2002). Extending UML for Multidimensional Modeling. 5th International Conference on the Unified Modeling Language (UML 2002), LNCS 2460, 290-304. (2002).

[8] Luján-Mora S., Trujillo J., Song I-Y. (2002). Multidimensional modeling with UML package diagrams. 21st Intl. Conference on Conceptual Modelling (ER2002). Volume 2503 of LNCS. pp. 199-213. Springer-Verlag (2002).

[9] Paim F.R.S., Castro J.B. (2002b). Enhancing data warehouse design with the NFR Frame-work. 5th Workshop on Requirements Engineering (WER2002), Valencia, Spain, pp 40-57.

[10] Paim F.R.S., Castro J.B. (2003). DWARF: An Approach for Requirements Definition and Management of Data Warehouse Systems. 11th IEEE International Requirements Engineering Conference (RE'03). Monterey Bay, California, USA, 2003.

[11] Prakash N., Singh Y., Gosain A. (2004). Information scenarios for data warehouse re-quirements elicitation. 23th Intl. Conference on Conceptual Modelling (ER2004). Volume 3288 of LNCS. pp. 205-216. Springer-Verlag (2004).

[12] Schiefer J., List B., Bruckner R. (2002). A holistic approach for managing requirements of data warehouse systems. Proc. Americas Conf. on Information Systems, 2002.

[13] Siau K., Lee L. Are use case and class diagrams complementary in requirements analysis? An experimental study on use case and class diagrams in UML. Requirements Eng (2004) 9:229-237.

[14] Trujillo J., Luján-Mora S., Song I-Y. (2003). Applying UML For Designing Multidimensional Databases And OLAP Applications. Advanced Topics in Database Research. Vol. 2 2003: 13-36.

[15] Trujillo J., Palomar M., Gómez J., Song I-Y. Designing data warehouses with OO conceptual models. IEEE Computer, special issue on Data Warehouses 34 (2001) 66-75

[16] Van Lamsweerde A. (2000). Requirements engineering in the year 00: a research perspective. Invited paper for ICSE'2000 – 22nd International Conference on Software Engineering, Limerick, Ireland. ACM Press, New York, June 2000.

[17] Winter R., Strauch B. A method for demand-driven Information Requirements Analysis in Data Warehousing Projects. Proceedings of the 36th Hawaii International Conference on Sys-tem Sciences, 2003

[18] Winter R., Strauch B. Information Requirements Engineering for Data Warehouse Sys-tems. ACM Sysmposium on Applied Computing (SAC'04) Nicosia, Cyprus.

[19] Yu, E. Modelling Strategic Relationships for Process Reengineering, Ph.D. thesis. University of Toronto, 1995.

[20] Yu, E. "Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering", Proc. 3rd Int. Symp. on Requirements Engineering. Annapolis, 1997, pp 226-235.