IADIS International Journal on Computer Science and Information Systems
Vol. 9, No. 1, pp. 43-54
ISSN: 1646-3692

TOWARDS A FRAMEWORK INCORPORATING FUNCTIONAL AND NON FUNCTIONAL REQUIREMENTS FOR DATAWAREHOUSE CONCEPTUAL DESIGN

Mohammed El Mohajir. *Computer sciences department, Faculty of sciences, University of Sidi Mohamed Ben Abdelah, Fez, Morocco. m.elmohajir@ieee.ma*

Ismail Jellouli. *Computer sciences department, Faculty of sciences, University Abdelmael Essaadi, Tetuan, Morocco. ismail.jellouli@yahoo.fr*

ABSTRACT

Data warehouse (DW) architectures, which are at the core of many new generation information systems, are designed to provide decision makers and professional analysts with intuitive and performing means to explore and get intelligent data from large heterogeneous databases. Most methods dealing with DW projects focus on functional requirements related to data quality; nonfunctional requirements like security, performance or operability are considered in a complete separate way. The lack of a global approach that gather concurrently both aspects from the requirements analysis to the conceptual design causes that many DW projects fail to meet their goals in satisfying users' needs and expectations. This work aims at developing a framework that encompasses both functional and non-functional requirements during the whole process of data warehouse design. This is achieved by integrating three additional aspects in the requirements analysis workflow, namely organizational, decisional and business process views. We select the telecom industry for our case study to demonstrate the viability of our approach.

KEYWORDS

Datawarehousing, functional requirements, non functional requirements, multidimensional structures.

1. INTRODUCTION

Building a data warehouse system is rather different from the development of conventional operational systems; not only the structures of the underlying source systems have to be considered but also the goals and strategies of the company. Therefore, requirement analysis and conceptual design are two of the key steps within the data warehouse building process. They are to a great extent responsible for the success of a data warehouse project since, during these two phases, the expressivity of the multidimensional schema is completely defined.

The purpose of requirements analysis is to identify which information is relevant for decision-making by either examining the users' needs in a requirement-driven approach, or the availability of data operational sources in a data-driven approach. While the Conceptual design aims at deriving an implementation-independent and expressive conceptual schema for the DW on the basis of the gathered requirements.

Several gaps have been identified in those phases such as the lack of standards for requirement analysis, the lack of mechanisms to ensure that the users' requirements are well mapped into the conceptual schema and then validate the obtained schema in the conceptual phase. To overcome these issues, several researchers have proposed some approaches such as (Giorgini et al. 2008) which uses goals diagram to capture users' requirements and (Tria et al. 2011) which exploits graphs to represent and validate the conceptual schema.

On the other hand, in the majority of data warehouse development projects, the non-functional requirements are almost neglected or at best presented as a second class type of requirement, frequently hidden inside notes with no mapping to the conceptual model, the experiences in the broader area of software engineering show that capturing non-functional requirements without mapping them into the conceptual model may determine an information loss.

Our goal is therefore to propose a framework for conducting a datawarehouse project where both the functional and the non-functional requirements are concurrently taken into consideration. Goals, rationale and organizational diagrams as well as roles and business process models are used for the analysis phase. While the entity graph representation, the secure multidimensional model, and the OCL constraints concepts are used at the conceptual phase. A refinement process will help us to accurately map the resulted models to the final datawarehouse multidimensional schema.

The organization of this paper is as follows: we describe in section 2 the main gaps identified in DW design while the third section discusses the proposed approach covering the requirements analysis and the conceptual design. Section 4 summarizes the most relevant literature related to requirement analysis and conceptual design in DW design followed by our conclusions in section 5.

2. GAPS IN DATA WAREHOUSE DESIGN

The data warehouse development process is a set of steps that a DW goes through when developed. Typically, this process is composed of the following phases:

- Requirement analysis where the purpose to which DW is intended is identified. This is accomplished on the basis of the decision makers' goals and needs. These are then documented in a form that is amenable to analysis, communication, and subsequent implementation.
- Conceptual phase aims at deriving an implementation independent and expressive conceptual schema for the DW.
- Logical phase takes the conceptual schema and creates a corresponding logical schema.
- Physical phase addresses all the issues related to the specific suite of tools chosen for implementation.

Existing approaches for DW development contain many gaps that hamper the process of building a data warehouse with an efficient structure for managing information and analyzing data. As examples of these gaps, we can cite:

- Requirement analysis in DW development is typically overlooked in many ways: Architects using the data driven approach for the development of the DW ignore the requirements analysis phase and they only focus on analyzing the operational data sources to elaborate the target DW conceptual schema.
- The lack of standards for requirements analysis in DW projects: Only few research proposals exist for conducting DW requirements analysis.
- DW requirements analysis results are not necessarily taken into consideration in the conceptual phase.
- Some types of requirements are not considered in the requirements analysis phase: Architects focus only on what information the DW is expected to provide, and no much interest is given to modeling non-functional requirements to provide details on how this information should be provided for compliance and accurate use.
- The lack of a mechanism that allows end users to validate their queries on the DW schema at the conceptual level before the implementation of the physical target DW.
- Datawarehousing projects are for a long-term period, and it is therefore very difficult to anticipate fully the evolving requirements.

Our work aims to overcome some of these issues. Our proposal is to combine several approaches such as (Soler & Trujillo 2008)(Tria et al. 2011)(Villarroel & Trujillo 2006) to come up with an hybrid approach as we will detail it in the next section.

3. OUR HYBRID APPROACH

Due to the lack of a general approach that allows the designer of a DW to build up a conceptual schema that is aligned with both functional and non-functional requirements, we propose a hybrid approach that fills this gap. Our method combines and takes advantage of several existing approaches. Its main objectives and contributions are summarized as follows:

- Our approach enables concurrent and consistent modeling of both the functional and the non-functional requirements from the initial requirements analysis phase up to the conceptual design stage.
- Our approach is hybrid in the sense that it is both data-driven and requirement-driven. The requirement driven approach allows complying fully with the business and the user requirements while the data driven approach permits the depiction of the data contained in the operational databases.
- Our approach encompasses an enhanced functional requirements analysis method (FRAM) that provides the decisions makers and the end users with the organizational, the decisional and the business process views. These additional diagrams should primarily help capturing commonly overlooked and unformulated business requirements.
- Our approach also incorporates the graph method (Tria et al. 2011) to allow the validation of the end users' queries against the conceptual schema and therefore ensures at an earlier stage the full adequacy of the final delivered multidimensional architecture with the full initial requirements.

3.1 Approach Overview

Functional requirements define what a DW is supposed to do whereas non-functional requirements define how a system is supposed to be. The combined approach allows the elaboration of the conceptual schema based on a detailed analysis of these two kinds of requirements. The following figure illustrates the different phases of our approach. Details of each phase are given in the subsequent sections.

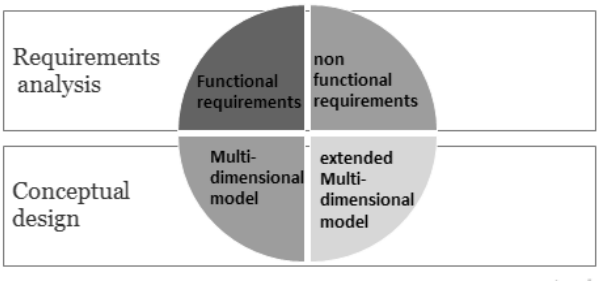


Figure 1. An overview of our hybrid approach.

3.2 Functional requirements modeling

The objective of this stage is to identify the functional requirements and map them into an appropriate multidimensional schema that also takes into account the existing data models in the operational systems. This phase is divided into two steps:

- In the functional requirements analysis step, rather than interviewing the decision makers as proposed by GrHyMM (Tria et al. 2011), we adopt the FRAM method that consists in supporting the end users in expressing their requirements by providing them detailed models that cover the organizational, decisional, and business process views.
- In the second step, we use GrHyMM (Tria et al. 2011) to generate the multidimensional schema in accordance with previously expressed functional requirements.

3.2.1 Functional Requirements analysis phase (FRAM)

We will follow for this phase our functional requirements analysis methodology (FRAM) to explicit the users' functional requirements. The different steps of FRAM are illustrated in figure 2 and the details of each step are given hereafter.

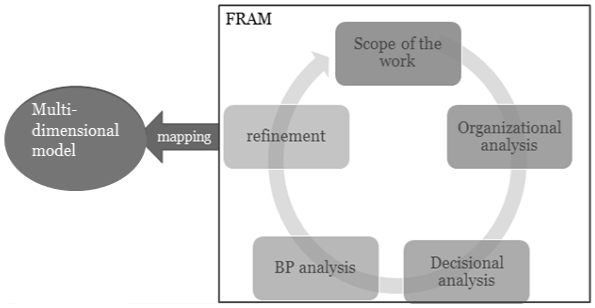


Figure 2. Functional requirements analysis method (FRAM).

a) Scope of the work

This step aims to elucidate the main objectives as well as the general business needs that motivate the development of the future DW, and also to identify the main actors who should be involved in the project lifecycle and who will interact actively with the final product.

b) Organizational analysis

This phase aims to study the organizational chart of the enterprise and the functional structures of the various departments and the interaction between them. The diagram will also display the various hierarchies, roles and services that exist within the enterprise.

The interest of this step is to help the designer to better understand the profiles of the decision makers who are concerned with the DW project, and suggest a specific role they can play in expressing the initial requirements or in supporting the development process.

This organizational analysis will also help the designer in getting better insight on the communication aspect within the domain of interest, from the semantic point of view. This will help in accurately understanding the description of the business models at different levels, and the related vocabulary. The designer will therefore be in a better position to understand the context and efficiently assimilate the business meaning of the requirements expressed by the different actors.

c) Decisional analysis

This step aims to investigate the business goals expressed by the decision-makers. We use the strategic rationale model proposed by GrHyMM according to three levels:

- Strategic goals or high-level objectives to be reached by the organization.
- Decision goals to answer how strategic goals can be satisfied.
- Information goals to define which information is needed for the decision making.

On the basis of the goals diagrams described above, we will identify the relevant functional requirements for the elaboration of the appropriate multidimensional schema.

d) Business Process analysis

The objective behind studying the organization business processes is to elucidate the unformulated requirements. Indeed, decision makers often, consciously or not, are unable to provide full details of the information they want to extract from the datawarehouse. Therefore, within the scope of the project, we will use the business process graph as a complementary tool for expressing additional low level stakeholders' goals. We will associate for each new goal specific metrics that permit the measurement of its achievement. These metrics will be possibly used as input for the datawarehouse model.

e) Refinement

The aim of this step is to refine the strategic rationale model initially designed with the goals expressed explicitly by the decision-makers by integrating the goals derived from the subsequent business processes analysis phase.
The obtained final diagram should exhaustively cover all the functional requirements needed to build the new enterprise datawarehouse.

3.2.2 Conceptual multidimensional modeling phase

The aim of this phase is to come up with a conceptual schema that meets not only the functional requirements obtained from the requirements analysis phase but also the data model mapped from the existing operational data stores. To achieve this goal, we use the GrHyMM framework that provides an automated process for reengineering the data sources to produce the conceptual schema using a set of requirement-derived constraints. The overall architecture of this method is illustrated in the following figure and a detailed description of the various steps can be found in (Tria et al. 2011).

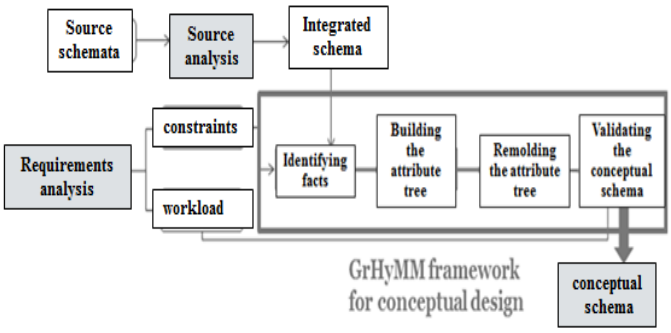


Figure 3. The overall architecture of GrHyMM framework.

3.3 Non functional requirements modeling

The objective of this phase is to identify the non-functional requirements and then build a representation of them to enrich the multidimensional schema obtained from the previous phase. This stage is divided into two steps. One is the non-functional requirements analysis phase in which we will identify the non-functional requirements by using the goal diagram. Then, we come to the elaboration of an extended multidimensional model where both functional and non-functional requirements are represented.

3.3.1 Non functional requirements analysis phase

We intend during this phase to capture and depict the non-functional requirements (NFR) that are relevant to the business context and inline with the APEX goals. These additional requirements should enhance the quality of the services offered by the future decision-support system. A detailed list of these NF requirements can be found in (Paim & Castro 2002).

Each type of a non-functional requirement needs a specific approach and model for its illustration. In this paper, for our case study, we only consider the security type of requirements. Following the approach proposed by (Soler & Trujillo 2008), we will describe how access to the data can be managed, which information can be accessed and by whom, and what are the constraints under which the information can or cannot be accessed. We use the security requirement model to specify these requirements

3.3.2 Non functional multidimensional modeling phase

In this phase, we enrich the conceptual schema derived from the initial functional requirements with the security requirements. The enhanced representation is based on UML 2.0/OCL extension proposed by (Villarroel & Trujillo 2006).

The UML 2.0/OCL profile provides the necessary stereotypes, tagged values and constraints and allows therefore representing and expressing the main security aspects in the datawarehouse conceptual model. Elements such as security levels on data, compartments and user roles on the main elements of a MD modeling such as facts, dimensions and classification hierarchies can be specified. Constraints attached to any of these elements can also be specified.

3.4 Case study: Telecom company DW(sales department)

In this paper, we use our approach to elaborate a DW to study the performance of a sales department of a telecommunication company. To achieve this goal we proceed as stated in the following sections.

3.4.1 Functional requirements modeling

a) functional Requirements analysis phase (FRAM)

We aim in this phase to identify the functional requirements of the future DW. First, we need to identify the objectives, the scope and the various parts involved in the project.

- The objectives: In our case study, we suppose that the objective for the decision-makers is to monitor and analyze the performance of the sales department.
- The scope: The sales department.
- The direct concerned parties: decision-makers, sales employees.

Organizational analysis: The organizational analysis shows the roles hierarchy of the actors. Diagram 4 represents the functional hierarchy of the telecom agency whereas diagram 5 describes the different roles. This analysis allows identifying the actors' profiles like access rights and security for the non-functional part.

Decisional analysis: At this stage, we draw the goals of the decision-makers (figure 6) according to three levels: strategy, decision and information. These goals will lead us to the definition of the appropriate functional requirements.

In the figure, the strategic goal is to increase the profit. This is achieved by increasing the sales and decreasing the expenses (decision goals). These decision goals depend on information on customers and product sales (information goals).

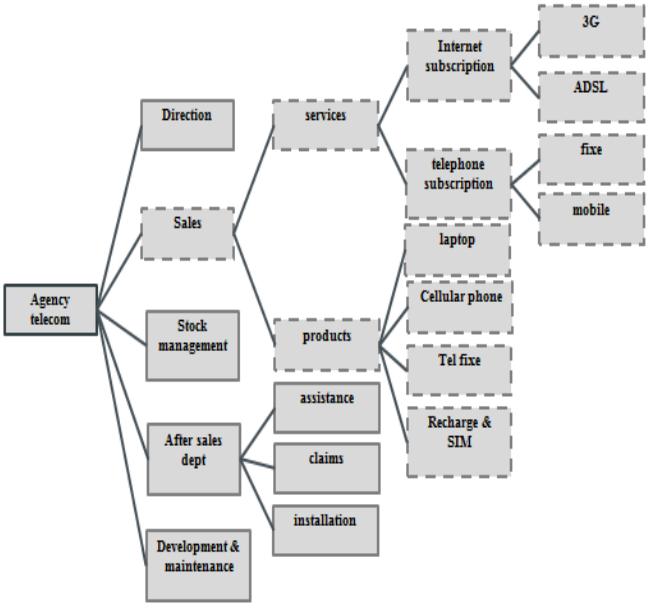


Figure 4. Organizational diagram of the telecom agency (by functions).

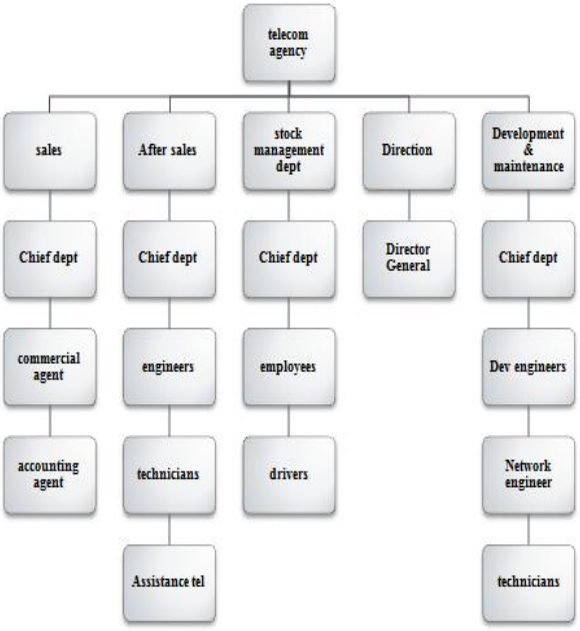


Figure 5. Organizational diagram of the telecom agency (by roles).

Business process analysis: This phase will help us identifying additional functional requirements that couldn't be extracted in the previous stage. The analysis of the business process can undergo the following steps:

- Step 1: We identify the relevant business processes: like "BP product sales", "BP subscription", "BP promotion". As an example, we will focus here only on the « subscription BP ».
- Step 2: We identify the goals of the Business Process. We have the following goals for the "subscription BP":
  - ✓ Goal 1: Provide a subscription service of a high quality.
  - ✓ Goal 2: Retain the existing customers and get new customers each month.
  - ✓ Goal 3: Availability of the service in various areas.
- Step 3: We associate specific metrics for each goal as shown in figure 7.

In the refinement phase, we update the strategic rationale model shown in figure 6 according to the new requirements and goals obtained from the analysis of the business process. Indeed, the BP goal is a decision goal for the decision maker and the BP metric becomes a requirement information goal. The new refined model is shown in figure 8.

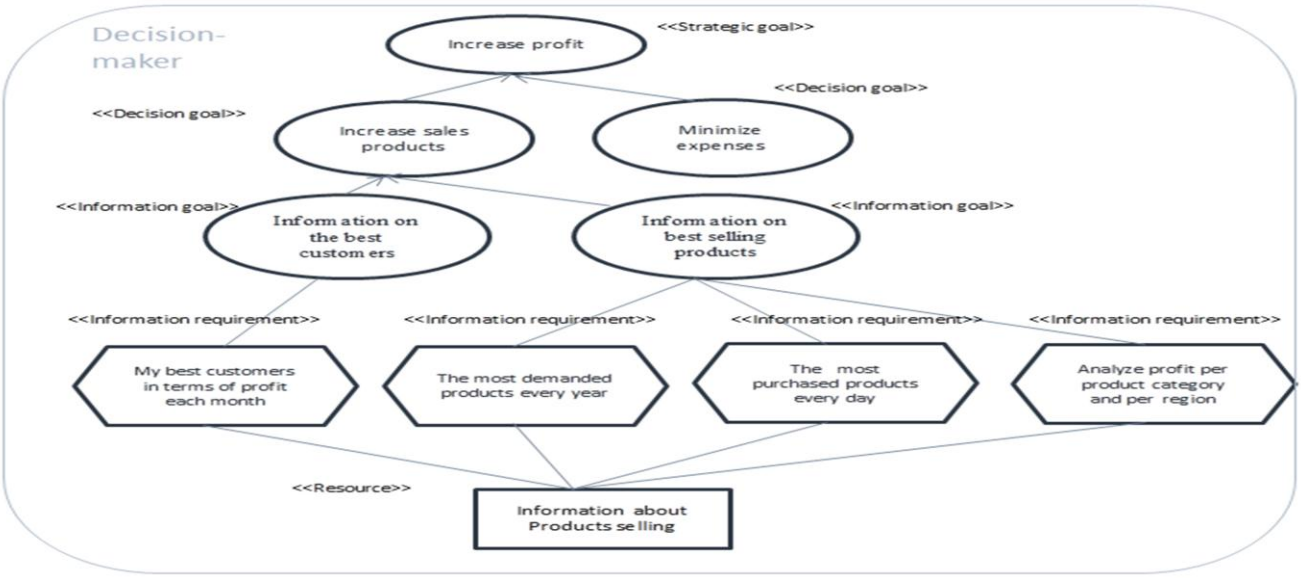


Figure 6. Strategic rationale model for the sales department.

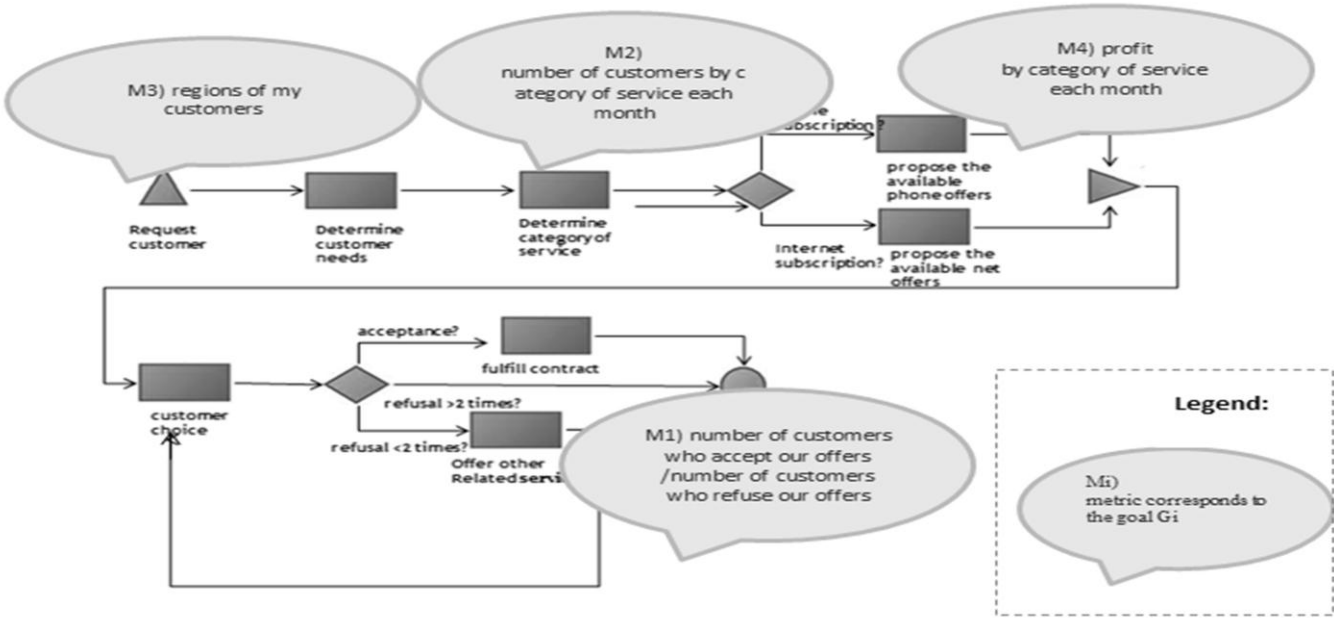
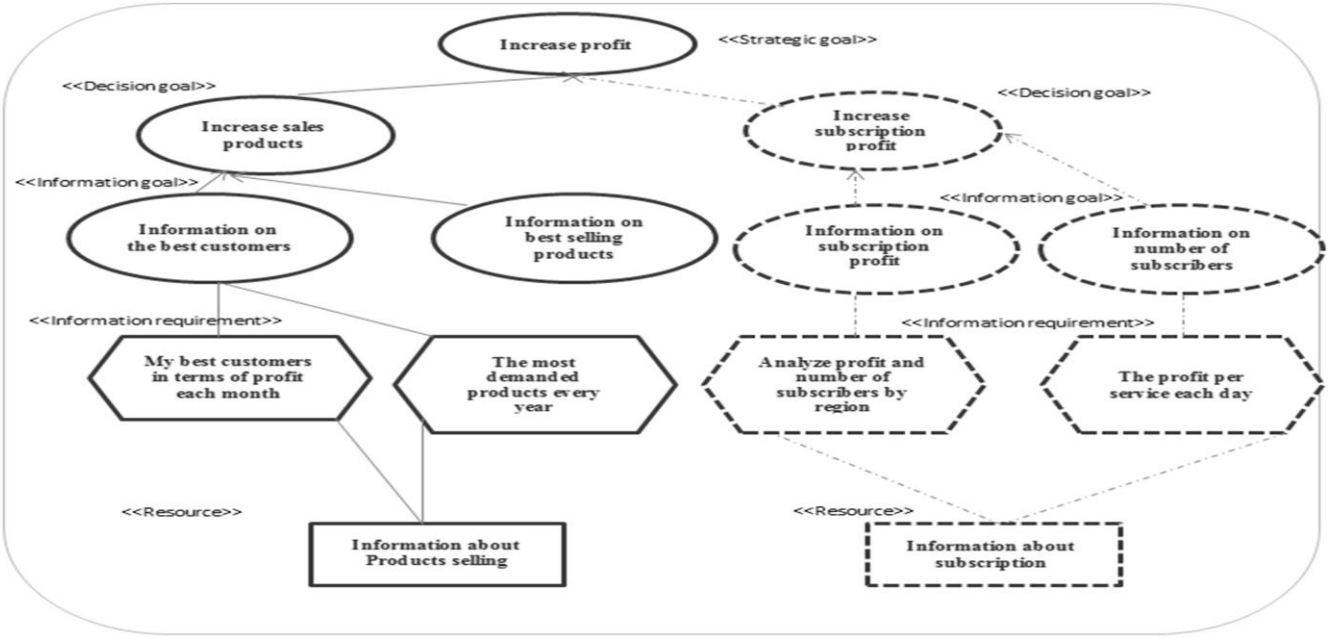


Fig 7: Metrics analysis in subscription BP diagram.



Legend:
-------: Obtained from BP analysis
——— : Obtained from decision maker interviewing

Figure 8. The refined strategic model for the sales department.

b) functional Requirements multidimensional modeling

In this phase, the functional requirements are mapped into an appropriate multidimensional schema using the approach proposed in (Tria et al. 2011). The following figure illustrates the multidimensional model obtained in our case.

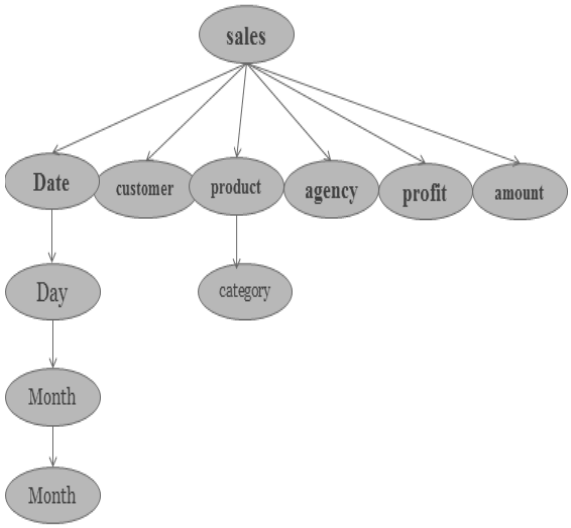


Figure 9. The multidimensional model for sales DW.

3.4.2 Non Functional requirements modeling

a) Non functional Requirements analysis phase

The security NF requirements are integrated into the appropriate multidimensional model using the approach proposed in (Soler & Trujillo).

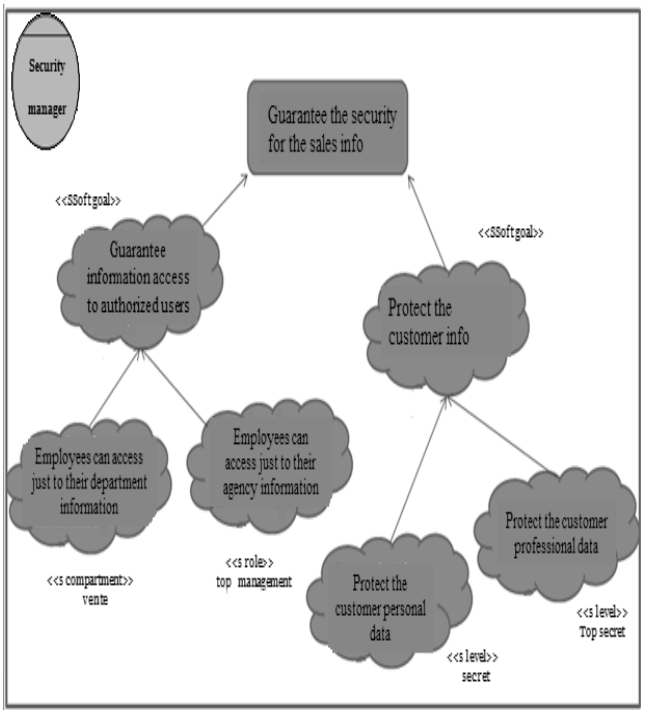


Figure 10. The security requirement model.

b) NF requirements in the conceptual phase

The aim of this phase is to incorporate the security requirements in the conceptual model using the approach proposed in (Villarroel & Trujillo 2006). The goals may become constraints on the basis of the OCL, An example of these constraints is described in the following list:

- For confidentiality reasons, access to sales information is denied to users whose working area is considered as not relevant to a specific domain of interest.
- The tagged value logType has been defined for sales class, specifying the value frustratedAttempts. This stereotype specifies that the system has to record, for future audit, the situation in which a user tries to access information in this fact class, and so where the system denies it because of the lack of the access rights.
- Users can be denied access to any sales data recorded before the date of their hiring contract.

The outcome of this phase, as it is shown in figure 11, is an extended multidimensional model in which we represented both the functional and the non-functional security requirements. This is according to the example we have taken for this study.

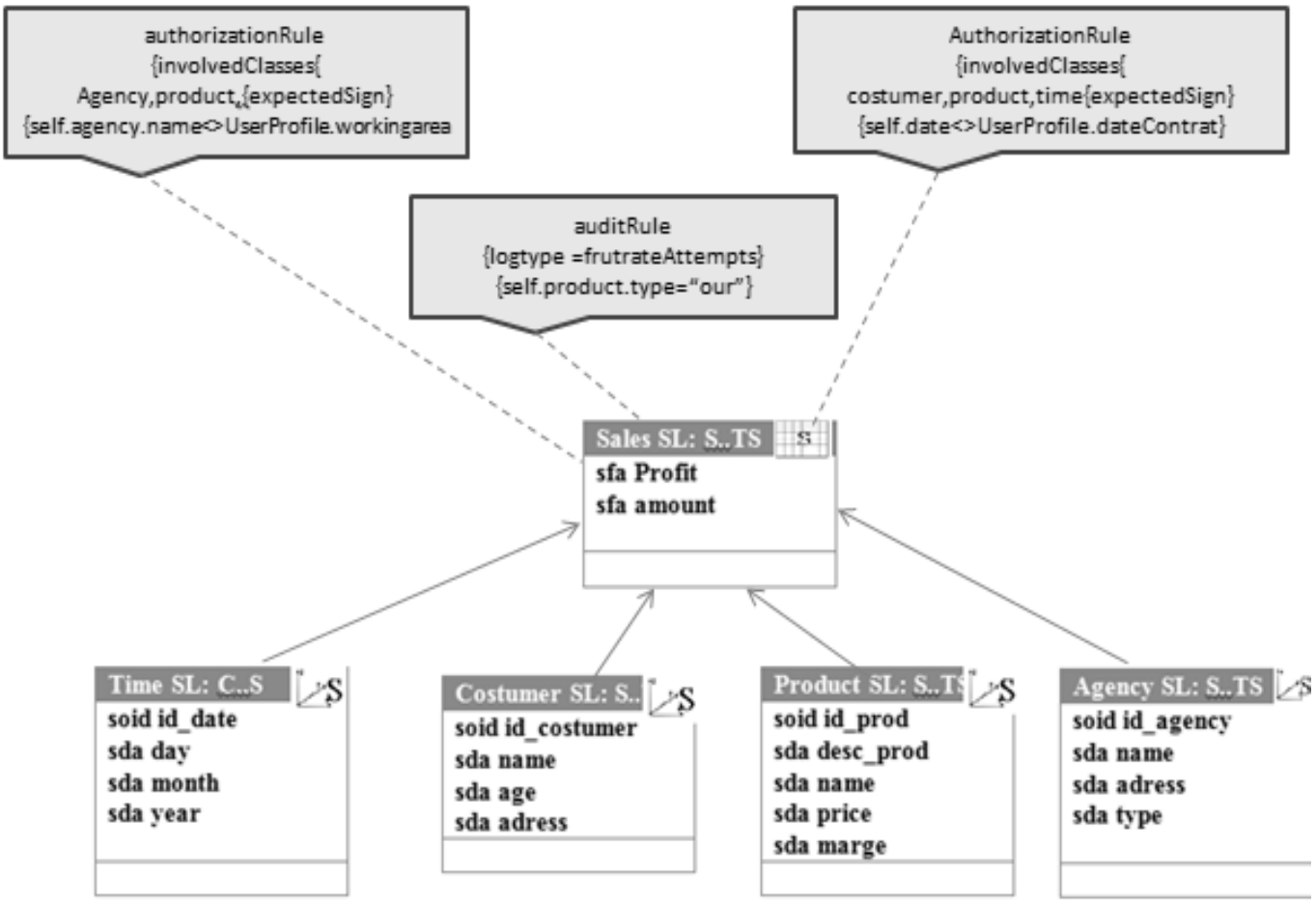


Figure 11.The extended conceptual schema for sales department.

4. RELATED WORK

The main purpose of a Data Warehouse is to support the decision making process by providing key indicators derived from collected heterogeneous and distributed data. To achieve this goal, DWs developers follow different approaches. In (Frendi & Salinesi 2003) the authors have proposed a framework for classifying those approaches from four different perspectives that are system, subject, usage and development. This classification allowed identifying some gaps such as the lack of guidance for the phase of requirement gathering which is a vital input for the DW design and development processes. To overcome this issue, it has been suggested to combine both the end-user and the operational system based approaches. However, we underlined that the main limitation of the proposed solution is the lack of a mapping process that permit a good level of traceability between the requirements analysis phase and the conceptual modeling phase. We have therefore enriched the methodology by conveying additional details and diagrams to help ensuring more accuracy and adequacy between the requirements and the design steps within the project lifecycle.

We stress out that in this paper we focused on requirement analysis and conceptual design as two main phases in the DW development process. The state of the art related to the different approaches used in those phases is illustrated in (Golfarelli 2009). Regarding the requirement analysis, three different approaches can be identified. The first one is the supply-driven approach that starts with an analysis of the operational data sources in order to identify all the available and relevant data, while the second one is a goal-driven approach that focus on elucidating the business strategy by interviewing the decision-makers (see also (Winter & Strauch) and (Boehnlein & Vom Ende 2000). However, some mixed techniques have been proposed in (Bonifati et al. 2007) and (Guo et al. 2006) to overcome the drawbacks of each single approach.

A concrete example for those approaches is illustrated in (Giorgini et al. 2008) and can be used within both data and mixed strategies. Indeed, the authors proposed a goal-oriented approach for datawarehousing requirement analysis by integrating two

different perspectives. The organizational modeling view, centered on stakeholders, and the decisional modeling view, focused on decision makers. Also, they proposed a set of rules that will govern the transition from the requirement analysis to the conceptual design. However, the main limitation of this approach is the lack of a formal mapping process between the requirements analysis phase and the conceptual design. The lack of representation of the non-functional requirements is also inconvenient. The second example illustrated in (Tria et al. 2011) explains the benefits of the mixed strategies in deriving the conceptual model on the basis of an automatic process of reengineering of data sources and by using a set of requirement-derived constraints. A formal validation process is adopted to map and to ensure correctness of the interpretation of the users' requirements in the conceptual model. The limitation however remains with the lack of representation of the non-functional aspect.

Most of the requirement analysis approaches adopted for DWs; especially (Giorgini et al. 2008) and (Tria et al. 2011), have focused purely on the business information needs expressed by the decision makers, without taking care of any nonfunctional requirements such as security or performance. Incorporating nonfunctional issues in the early modeling stages of the development is a cornerstone for building a DW that satisfies all user expectations and company policies. To overcome this issue, the authors in (Paim & Castro 2002) defined a catalogue of major data warehouse non functional requirements types but without suggesting any methodology for mapping those requirements in the global conceptual model. Thus, several works have focused only on the implementation of a specific type of the NFR from the requirement level to the operational physical level. For example, regarding the security aspect, paper (Saurabh & Nagpal 2011) suggests an MDA approach to come up with a secure data warehouse model. However a comparative analysis of the data warehouse design approaches from the security perspectives (Saroop & Munar 2011) shows that most of the researchers start considering security issues only from the design phase without any representation at the requirement level.

The analysis of the various approaches proposed for building a datawarehouse highlight the importance of taking into account the functional and the nonfunctional aspects in meeting and satisfying users' needs and expectations. However, existing approaches put emphasis only on one aspect rather than the other. Our work is a tentative towards building an hybrid approach that allows the elaboration of the conceptual schema on the basis of both the functional and the non-functional requirements. The security aspect is taken as an example.

A datawarehouse is not just a collection of data that takes into account the functional and non functional specifications in order to satisfy decision-makers needs and expectations; it is also a crucial component in modern organizations that influence the business strategy and its processes. However, there are no models available that are capable of linking the conceptual representation to the business processes. In (Stefanov & Scheifer) the authors extend the Event-Driven Process Chain, a business process modeling language, with an additional perspective to make this relationship explicit in a conceptual model. (Boehnlein & Vom Ende 2000) proposed a new process oriented approach to the development of the data warehouse structures, it includes a comprehensive modeling technique which use business process models instead of operational data models to derive the initial data warehouse schematic structures.

5. CONCLUSIONS

The state of the art conducted in this study revealed that even if software and datawarehouse systems belong to different application domains, there are some commonalities within their development processes. The gaps identified in existing DWs approaches, such as the lack of standards for requirement analysis and the lack of formal traceability between the requirement analysis phase and the conceptual phase were alleviated in software engineering. Therefore, many researchers are trying to adapt the solutions adopted in the software domain to DWs. A concrete example is illustrated in the GRanD approach [1] which has adapted the Agent oriented software engineering methodology to the DW domain (i.e. tropos methodology). On the other hand, the topic of conceptual modeling in datawarehousing has been widely explored, but the issues related to user requirement analysis and its relationships with the subsequent design steps have been only partially studied. Moreover, capturing and modeling non-functional requirements has not been integrated in most of the proposed approaches. Thus, we believe that the efforts in this area should be twofold. On the one hand, we need to elaborate a comprehensive and detailed approach that takes into account all types of requirements in the initial analysis and specifications phases. On the other hand, we should come up with a unified framework where these requirements models can be formally transformed and propagated to the conceptual, logical and physicals models. Our study can be considered as an introductory attempt towards the construction of this framework.

We have highlighted the importance of considering the organizational, decisional and business process views when tackling the functional requirement analysis. Our case study related to the telecom industry provided us a concrete basis to demonstrate the viability of our proposals.

As a future work we will carry a more detailed study on the comprehensive representation of the additional models we can define for the various types of requirements during the analysis phase. We will then need to elaborate and validate a methodology that will help achieving a formal transformation of these requirements models to support the generation process of the conceptual and physical models. These will ensure both completeness and full implementation of the requirements specifications.

REFERENCES

Boehnlein M., Ulbrich vom Ende A., 2000. Business process oriented development of data warehouse structures. *Proceedings of Data Warehousing.*

Bonifati A., Cattaneo F., Ceri S., Fuggetta A., Paraboschi S., 2007. Designing data marts for data warehouses. *ACM Transactions on Software Engineering and Methodology (TOSEM)*.

Frendi M., Salinesi C., 2003, Requirements Engineering for Data Warehousing.

Giorgini P., Rizzi S., Garzetti M., 2008. GRAnD: A Goal-Oriented Approach to Requirement Analysis in Data Warehouses. *Decision Support System*, pp 4 -21.

Golfarelli M., 2009. From User Requirements to Conceptual Design in Data Warehouse Design.

Guo Y., Tang S., Tong Y., Yang D., 2006. Tiple-driven data modeling methodology in data warehousing: a case study. *DOLAP '06 Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*.

Paim F. R. S., Castro J. F. B., 2002. Enhancing Data Warehouse Design with the NFR Framework. *workshop on requirements engineering (WER)*.

Saroop S., Kumar M., 2011. Comparative Analysis of Data warehouse Design Approaches from Security Perspectives. *International journal of computer trends and technology*.

Saurabh A. K., Nagpal B., 2011. A survey on current security strategies in data warehouses. *International Journal of Engineering Science and Technology (IJEST)*.

Soler E., Trujillo J., 2008. Towards Comprehensive Requirement Analysis for Data Warehouses: Considering Security Requirements. *Proccedings of The Third International Conference on Availability, Reliability and Security (ARES)*.

Stefanov V., Schiefer J.,2008. Bridging the Gap between Data Warehouses and Business Processes: A Business Intelligence Perspective for Event-Driven Process Chains.

Tria F. D., Lefons E., Tangorra F., 2011. GrHyMM: A Graph Hybrid Multidimensional Model. *ER'11 Proceedings of the 30th international conference on Advances in conceptual modeling: recent developments and new directions*.

Villarroel R., Trujillo J., Fernández-Medina E., Piattini M., 2006. A UML 2.0/OCL Extension for Designing Secure Data Warehouses. *Journal of Research and Practice in Information Technology*.

Winter R., Strauch B., 2003. A method for demand-driven information requirements analysis in data warehousing. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03)*.