

Data Warehouse Design Using Process-Oriented Requirement Analysis: A survey

Pallavi Kalokhe

*Department of Information Technology
Gharda Institute Of Technology, Lavel, Khed, Maharashtra, India*

Kranti Gajmal

*Department of Information Technology
Gharda Institute Of Technology, Lavel, Khed, Maharashtra, India*

Abstract- Intelligent and comprehensive analysis systems are a powerful instrument for companies to analyze their business. The implementation of such systems for an enterprise-wide management and decision support can be very different from traditional software implementations. Because analysis systems are strongly data-driven, the development process is highly dependent on its underlying data, which is generally stored in a data warehouse. Data warehouse systems generally concern many organizational units. Therefore, the collection of unambiguous, complete, verifiable, consistent and usable requirements can be a very difficult task. Use cases are considered as standard notation for object-oriented requirement modeling. This paper explains how use cases can be used to elicit requirements for data warehouse systems, and how to involve the organizational context in the modeling process.

Keywords: Data Warehouse, Business Process, Process Orientation.

I. INTRODUCTION

A data warehouse is multidimensional modeled and is used for the storage of historicized, cleansed, validated, synthesized, operative, internal and external data. Stakeholders of a data warehouse system are interested in analyzing their business processes in a comprehensive and flexible way. What they actually need is a view of their business processes and its data. For this purpose data warehouses are modeled multidimensional, which corresponds to a typical view of its users. This analysis view of the business processes can be very different to the general view even though the underlying process is the same. Hence it is necessary to elicit requirements from the stakeholder of a data warehouse, which belong to their analysis views. The design of data warehouse system is highly dependent on these requirements. Often the users, the system analysts and developers don't speak the same language. Such communication problems can make it difficult to turn description of an analysis system into a technical specification of a data warehouse system that all parties can understand. Therefore, it will not deliver the expected effect to the company. In these cases often departments will develop data marts for their own purposes, which can be considered as stovepipes and makes an enterprise-wide analysis system impossible.

The challenge is to model a data warehouse system in a way that is both precise and user-friendly. Each symbol describing the analysis process should be intuitive for the user and have defined semantics, so that the developers can use the description as a general, but precise specification of the data warehouse system. Use cases have two advantages, which make them suitable for representing the requirements for a data warehouse system. Important characteristics that such a model should possess are:

- *A high level of abstraction.* The model should be at the level of the users' views of the desired system. It should not indiscriminately mix this high level information with information that is relevant to lower levels of the development process.
- *Human-readability.* The language in which the model is to be expressed will be used for validating the specification: that is, for presenting the specification to users for their views on its contents. Human understandability is thus the prime concern.
- *Precision.* A high-level specification language, for project scope agreement, delineating the system boundaries and naming major objects, rules and processes, is required.
- *Specification completeness.* It is important that the model captures all aspects of a specification.
- *Map ability to later phases.* A requirements definition phase will typically be followed by a detailed analysis and design phases. Therefore, the model should possess a structure suitable for mapping on the later phases.

This paper shows an object-oriented approach for a process-oriented requirement analysis, which grasps the stated model characteristics.

II. RELATED WORK AND BUSINESS PROCESS ORIENTATION

Building a data warehouse is different than developing transaction systems, whereby the requirement analysis process for the latter is supported by numerous methods. Up to now the data warehouse design process has not been supported by a formal requirement analysis method although there are some approaches for requirement gathering. Requirements can and must be gathered before the data warehouse design process otherwise only those parts are captured, which are in the basic corporate model. Data warehouse is designed to support the business process rather than specific query requirements. The process approach for data warehouse design enables organizations to focus on the performance of business processes and drift away from traditional task or department performance measurement. The change to process centering is not primarily a structural one, but process centering is a shift in perspective, in which tasks and processes exchange places. In practice this means that the traditional hierarchical organization remains and the process view is integrated into the industrial way of organizing. Therefore business processes cross organizational boundaries and often tends to be inefficient because of changing responsibilities, long delay times and so forth. The support of different views of the models is an opportunity to capture multidimensional and aggregated views of data warehouses. Basically use case models and object models can be applied to software processes and business processes, which use the same notation but represent different functionality. As argued above, we aim at analyzing business processes and focus therefore on the business process approach provided by these models and described in.

The use case model provides an overview of the area of interest, while the object model is an internal model and describes accurately each business process with its tasks and resources in order to make the model clearer. This paper presents the adaptation of the use case and object model in order to support the data warehouse requirement analysis. The aim is to provide a formal requirement engineering method that is easy to use, quickly to understand, covering all major data warehouse model characteristics and is therefore a means of communication between all involved parties in the requirement process.

III. DATA WAREHOUSE REQUIREMENTS

Requirements to the data warehouse system determine what data must be available, how it is organized, and how often it is updated. Furthermore the requirements enable the stakeholders to communicate the purpose. It is advantageous to start the requirement collection process with a macro business discovery (see Figure 1), which is necessary to identify the key business processes, key business metrics, measures and requirements. A high-level model has to be developed depicting how business is conducted.

Data Warehouse Requirements

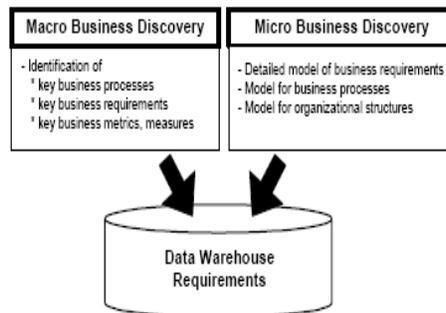


Figure 1: Requirement Discovery Process for Data Warehouse Requirements

The micro business discovery is an in-depth analysis of the requirements of the organization as defined by the macro business discovery. It describes the business requirements in more detail by considering these requirements in context of the models for the business processes and the organizational structures. Figure 2 shows the impacts of data warehouse requirements [4]. The figure demonstrates very well, that data warehouse requirements directly affect technical aspects of the data warehouse system.

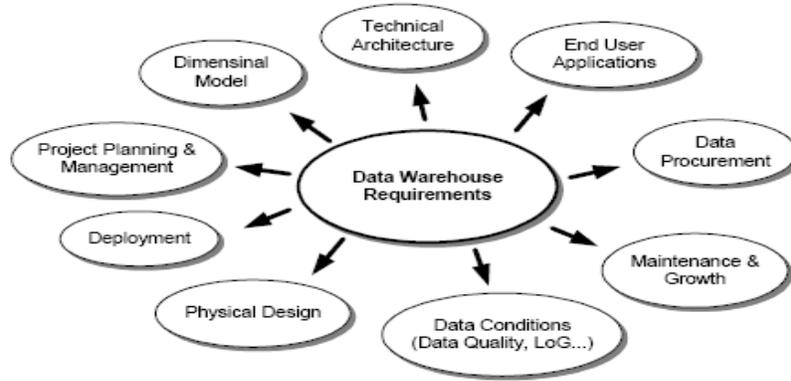


Figure 2: Impact of Warehouses Requirements

In the requirement collection process data warehouse requirements have to be driven by the business requirements. They arise from the macro and micro business discovery and result in a comprehensive technical specification. On the other hand data warehouse requirements are derived from the business requirements and their aim is to provide a comprehensive, precise specification for the data warehouse team. Current use case oriented approaches for the requirement analysis are used to describe technical aspects and ignore the business view. By integrating this central aspect in the requirement modeling process here present an approach of how to gather the business requirements from stakeholders corresponding to their business processes. In the requirement analysis process it is important to consider the organizational context of the stakeholders.

IV. ORGANIZATIONAL CONTEXT

A data warehouse is intended to provide an enterprise-wide decision support for an organization. It should address the users of all hierarchy levels of the organization. This broad spectrum of different types of end users requires information at different granularity levels to meet their specific needs.

In the often-found structure of traditional organizations as figure 3(a) different requirements for data processing and data analysis can be identified, which correspond to the different layers of the organization. As we move upward through the layers of the hierarchy, for example, the level of granularity required decreases. Executives use highly summarized information, while line managers work at a detailed level. Administrative users need to create and maintain individual data items such as orders or customer records. Professional users and line managers require statistical analysis tools. Executives require systems that highlight anomalies or geographically show key indicators, and allow drill-down in problem areas. During the micro business discovery requirements of data warehouse users of all different levels in the hierarchy are identified.

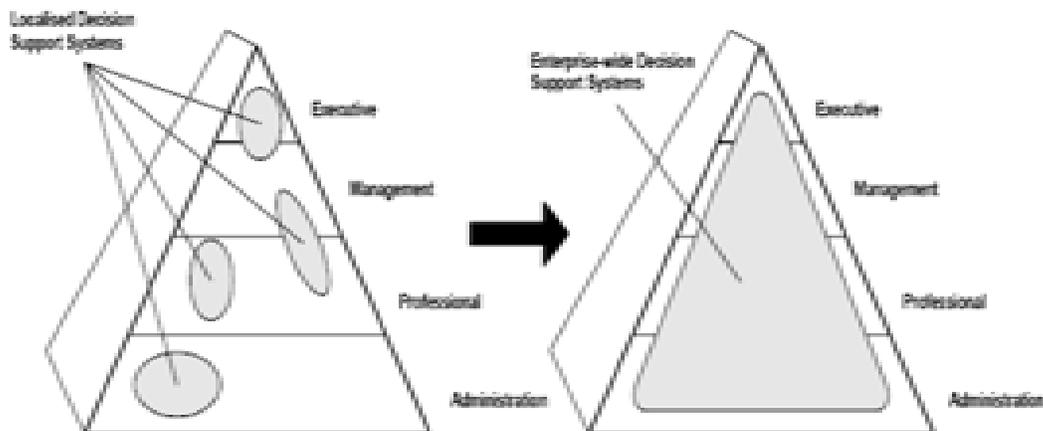


Figure 3 (a)structure of traditional organizations

(b)an enterprise-wide decision support system

Figure 3 (b) shows an enterprise-wide decision support system with its organizational hierarchy, which can achieve these business goals. For gathering the meshed requirements of such a decision support system, it is necessary to use models which represent all business requirements on each hierarchy level and which allow requirement consolidation of all levels.

V. USE CASE MODEL

The Use Case model captures business processes in the company that satisfy the customer’s interests and the interests of others outside the company (partners, suppliers, etc). Our use case model describes business processes in the company that are analyzed and provide information to the data warehouse user and the company’s staff.

Analysis System and Subsystems-The analysis system or subsystem is the modeling concept to symbolize the business or area of responsibility.

Analyst-The analyst represents a role that someone or something in the environment can play in relation to the analysis system. In our model analyst represent the environment that analyzes business processes belonging to the specified business system.

Use Case-The use case is constructing for a business process that is analyzed.

Uses Association-The uses association is a must association and describes use cases that are aggregated into other use cases. This concept must be applied when parts of the use case are analyzed by other analysts.

VI. OBJECT MODEL

The object model provides a clear picture how the use case is structured internally in order to realize the analysis capabilities of the process (backward engineering) and to describe the analysis requirements of various actors (forward engineering). The internal structure consists of fact and dimension objects.

Dimension Objects

Davenport states in [3] that a process is a structured, measured set of activities designed to produce a specified output for a particular customer or market. He notes that a business process is a specific ordering of work activities across time and place, with a beginning, an end, and clearly identified inputs and outputs: a structure of action. Hammer argues in [7] that a business process is a group of tasks that together create a result of value to the customer. Hammer’s definition consists of activity, result, customer and input objects, and Jacobson focuses on the customer, product, service, and deliverable and measure objects. Davenport provides the most comprehensive definition, which basically includes the objects of both other definitions. The objects in the process definitions highlight key business process characteristics, but represent also classical data warehouse dimensions (e.g. organization, customer, product, service, time, etc.) and data warehouse facts (measure). As these key business process objects can be found in any classical business process, we propose a standard set of dimension objects representing data warehouse dimensions (Figure 4):

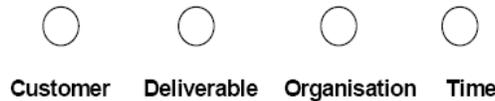


Figure 4: Data Warehouse Main Dimension Objects in Classic Business Processes

Customer Dimension Object. The single most important word in the definition of process is ‘customer’ and a process perspective on a business is the customer’s perspective [8].

Deliverable Dimension Object. Davenport and Hammer do not specify the output [3] or result of value [7] of the process in their definition. The customer does not see or care about the company’s organizational structure or its management philosophies; the customer sees only the company’s products and services, all of which are produced by its processes. As the output or result of the value of a process might be the opposite of the defined process deliverables, and products or services are too specific for a requirement analysis model.

Organization Dimension Object-

The analysis of the organizational structure detects occurrences that are caused by certain units and therefore organization dimension object is required.

Time Dimension Object-

A business process is a specific ordering of work activities across time with a beginning and an end [3]. The end of the process represents a pointing time where the results of the process are delivered and cannot be changed anymore.

Fact Object-

Performance is measured against cost, longevity, service and quality. These measures (e.g. turnover, profit, ratio rates etc.) represent facts in data warehouse notation and fact object in this paper.

VII. CONCLUSION

This paper presents the adaptation of use case and object models for modeling business requirements for data warehouse systems to support the data warehouse design process. We showed how data warehouse requirements are derived from business requirements and their organization context. Use case model is an excellent means of both expressing requirements with regard to the data warehouse and providing a comprehensive picture of what the data warehouse is intended to perform. It illustrates the function of the business, the process analysts and the business process to be analyzed with its aggregations. The object model is an internal model that captures different analysis perspectives of the business process.

REFERENCES

- [1] Amatory, S. Murray, D., Data Warehousing in the real World – A practical Guide for Building Decision Support Systems, Addison-Wesley Publishing 1997.
- [2] Coad, P., Yourdon, E., Object-Oriented Analysis. Englewood Cliffs, N.J.: Prentice Hall 1991.
- [3] Davenport, Th., Process Innovation: Reengineering Work through Information Technology, Harvard Business School Press 1993.
- [4] Davis, A., Hisa, P., Status Report: Requirements Engineering, IEEE Software November 1993.
- [5] Debevoise, N. T., The Data Warehouse Method, Prentice-Hall, 1999.
- [6] Finkelstein, A., Requirements Engineering Research: Coordination and Infrastructure, Requirements Engineering 1, 1996.
- [7] Hammer, M., Champy, J., Reengineering the Corporation – A Manifesto for Business Revolution, Harper Collins Publishers 1993.
- [8] Hammer, M., Beyond Reengineering, Harper Collins Publishers 1996.
- [9] IEEE, IEEE Guide to Software Requirements Specifications, IEEE Inc., 1984
- [10] Inmon, W. H., Building the Data Warehouse, Wiley & Sons 1996.