International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:1, 2008

Towards Development of Solution for Business Process-Oriented Data Analysis

M. Klimavicius

***Abstract*—**This paper proposes a modeling methodology for the development of data analysis solution. The Author introduce the approach to address data warehousing issues at the at enterprise level. The methodology covers the process of the requirements eliciting and analysis stage as well as initial design of data warehouse. The paper reviews extended business process model, which satisfy the needs of data warehouse development. The Author considers that the use of business process models is necessary, as it reflects both enterprise information systems and business functions, which are important for data analysis. The Described approach divides development into three steps with different detailed elaboration of models. The Described approach gives possibility to gather requirements and display them to business users in easy manner.

***Keywords*—**Data warehouse, data analysis, business process management.

I. Introduction

DURING the last ten years, the interest to analyze data has increased significantly, because the competitive advantages that information can provide for decision-making process. Data warehouse systems represent a single source of information for analyzing the development and results of an enterprise organization in a changing environment. The data in the data warehouse describes events and a status of business processes, products and services, goals and organizational units. Nowadays, a key to survival in the business world is ability to analyze, plan and react to changing business conditions as fast as possible. However the ability to change is bound to many constraints, such as staff knowledge, business supporting systems, etc. Business operations depend on enterprise information systems, it mean that changes in business processes requires changes in supporting information systems. A change in operational information systems requires changes in data warehouse, which is a central repository of atomic and summarized data from different operational systems.

Data warehouses integrate data from multiple heterogeneous information sources and transform them into a multidimensional representation for decision support applications. Apart from a complex architecture, involving data sources, the data staging area, operational data stores, the global data warehouse, the client data marts, etc., a data warehouse is also characterized by a complex lifecycle [2]. Business processes are often the starting point for software development and define requirements for software systems to be designed. Research and industry have addressed the alignment of business processes and information technology only marginally. Most software developers are not aware of business processes or are not able to read the models, as different modeling languages with different diagrams and notations are used in both domains. Author proposes to fulfill needs of developers and business users by applying business-oriented approach for development of data analysis solution. Development stages of business process-oriented approach are as follows [3]:

- Development starts with business process modeling;
- Requirements are elicited from business process diagrams;
- Business processes are linked to logical data warehouse model;
- New logical data structures of dependent data mart are generated.
- Physical data warehouse structures are created.

II. Related Work

There are many conceptual models available for business processes, databases or data warehouses, but there are no models available that focus on the relationship between the data warehouse development and the business processes.

Widely used are approaches to model some particular stage of data warehouse development. Data warehouse models are available for multidimensional modeling [16] and ETL process modeling [15]. These researches solves, particular problem in data warehouse development lifecycle, but do not address the link to business processes.

Boehnlein and Ende [7] propose to use structured entity relationship model for derivation of data warehouse structures. They show how data modeling technique is used for derivitation of initial data warehouse structures from the conceptual schemes of operational sources. Approach is not practical for automatic creation of multidimensional tables because of varying requirements.

An integrated view on data warehousing and business processes was introduced by Stefanov, List and Korherr [17]

Manuscript received November 30, 2007. This work has been partly supported by the European Social Fund within the National Programme "Support for the carrying out doctoral study programm's and post-doctoral researches" project "Support for the development of doctoral studies at Riga Technical University".

M. Klimavicius is with the Faculty of Computer Science and Information Technology, Riga Technical University, Meza street 1/3-537, LV-1048, Riga, Latvia (phone: 371-29446356; e-mail: maris.klimavicius@gmail.com).

in terms of a model that allows showing where and how a data warehouse is used by business processes, and which parts of the business processes depend on which parts of the data warehouse. The approach is not targeted for development of data warehouse, as one proposed by author of this paper.

Mazon, Trujillo, Serrano and Piattini [18] applied the MDA framework to data warehouse repository development, and aligned multidimensional conceptual data models with code. Speaking in terms of MDA, they aligned a Platform Independent Model (PIM) with a Platform Specific Model (PSM) and defined a transformation between them. Approach proposed in this paper can be seen on top of this work targeting the Computational Independent Level (CIM) level, as enterprise context is aligned with the data warehouse data model.

In this paper we present a business process-oriented approach to the development of data warehouse. We suggested a comprehensive modeling technique, which does not only focus on a certain step of the data warehouse development. In order to derive initial data warehouse structures we use business process models instead of operational data models

III. Data Warehouse Development

Most techniques that are used by organizations to build a data warehousing system use either a top-down and bottom-up development approach. In the top-down approach [1], an enterprise data warehouse is built in an iterative manner, business area by business area, and underlying dependent data marts are created as required from the enterprise data warehouse content. In the bottom-up approach [13], independent data marts are created with the view to integrating them into an enterprise data warehouse at some time in the future. There is still a lot of discussions about the similarities and differences among these architectures, but despite these differences there are two main steps in data warehouse development, which are very closely connected – requirements gathering and information modeling. In an ideal world, information modeling would be easy – just add every kind of information you need to the information providers (data store objects, info cubes, etc), make those available to end users, and you have done. However, there is a fundamental dilemma in developing and information model. Every information model will always be a compromise of some sort between business requirements, performance requirements and fundamental data warehouse design aspects [8] Fig. 1 illustrates the information-modeling dilemma.

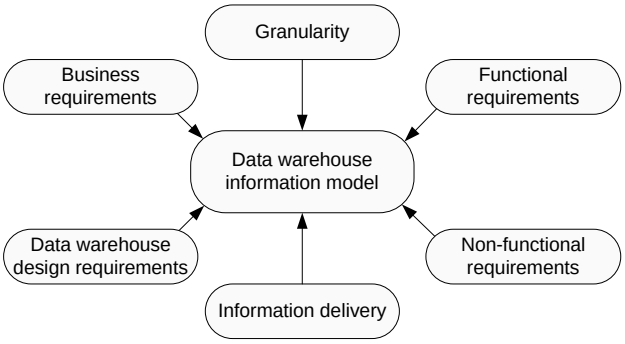

Fig. 1 Information modeling dilemma

Common answer to the question “What information do you need? “is: “I need everything!”. Business requirements usually conflict with non-functional requirements like performance and usability. There is always a trade-off between performance and amount of information available, and the more complex is the information model, the more complex is to use it and more difficult it is for occasional end users to create and execute meaningful queries with consistent and complete result sets. The art of information modeling is to define an information model that best suits to all these requirements without sacrificing too much of the business side of requirements; Business requirements are most important requirements. Business users must understand the process of information modeling.

Usually business users do not understand complicated data models; therefore it is hard for them to follow this process. The authors propose approach, which raises the modeling warehouse to a higher level that is more clear and easy to follow by business users.

IV. Methodology

It is well known fact, that modeling techniques can facilitate communication between representatives of different systems and increase the development speed. Most of business users have a comprehensive understanding of their business processes, which they want to explore and analyze. Business users know how the enterprise works and developers are familiar with information systems running that business. Major problem in building a data warehouse is to identify and consider information needs of potential users. The methodology consists of 6 steps:

1. Development of logical data warehouse data model;
2. Development of business process models;
3. Searching for decision points;
4. Eliciting of requirements and data organization according to data warehouse data structures;
5. Development of new logical data structures;
6. Creation of physical data warehouse structures.

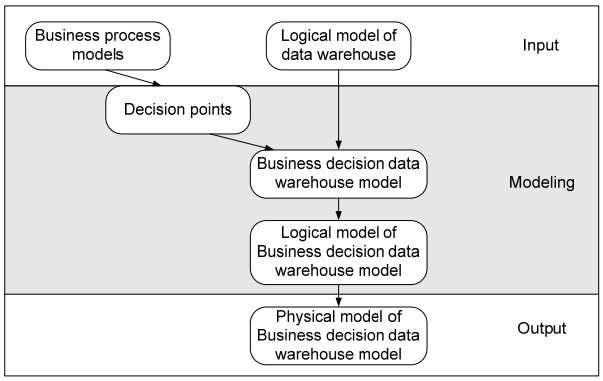

Fig. 2 Business process-oriented data analysis solution development

Business users are accustomed to their own vocabularies and concepts, and data interpretation is greatly improved by knowledge about context.

Surprisingly, information about the relationship between the data warehouse data and the organization's processes is not made available to the data warehouse users. In this paper, I present an approach that uses business process models and modeling techniques to record the knowledge about this relationship. I propose an approach that allows us to show:

• What kind of data are behind every business process and how they influence decision making;

• Which parts of the data warehouse data is created by which business process or part of it;

• How business processes impacts values of the data warehouse data

• Where the products and deliverables of the processes are mirrored in the data warehouse data Structure

V. Architecture

For many modeling purposes, one large general model is not advisable [19]. Creating several smaller, specific domain models allows for a better separation of concerns. Separate issues can be expressed in separate domain models, which are then related to each other by the use of central meta-data repository and conformed data marts. It is the reason why author selected enterprise data warehouse architecture for this solution. I propose to use separate data mart (Fig. 3) with data organized according to business processes.

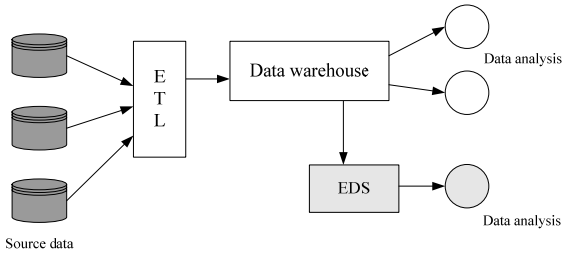

Fig. 3 Solution architecture

Such architecture provides with logical data warehouse model, which is essential input data in order to do modeling.

VI. Business Process Modeling

It is well known fact, that modeling techniques can facilitate communication between representatives of different systems and increase the development speed. Most of business users have a comprehensive understanding of their business processes, which they want to explore and analyze. Business users know how the enterprise works and developers are familiar with information systems running that business. Major problem in building a data warehouse is to identify and consider information needs of potential users. By using business process models it is possible to elicit information requirements. Example of business process of Budget allocation in The Treasury of the Republic of Latvia is shown on Fig. 4.

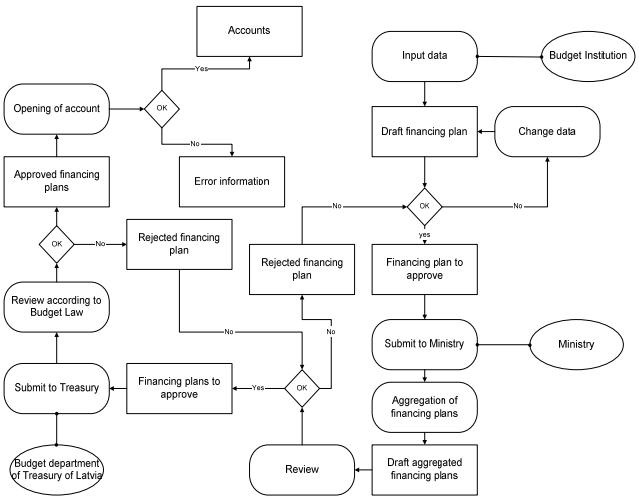

Fig. 4 Business process diagram of Budget allocation

The process contains operational activities connected in chronological sequence as well as other resources such as the human resources/organizations, information resources and the data that contain the information itself. There are three views focusing on functions, data, and the organization.

• Data view. The data view contains events and statuses. Events such as "customer order received," "completion notice received," or "invoice written" are information objects that are represented by data. Statuses such as "customer status" and "article status" are also represented by data.

• Function view. The function view contains the description of the activities to be performed, the enumeration of the individual sub functions that belong to the over-all relationship, and relationships that exist between the functions.

• Organization view. The structure and the relationships between users and organization units constitute the organization view. Users are assigned to organization units, which are formed according to criteria such as the same

function or the same work object.

Business process modeling integrates typical business practices, processes and information flows, data stores and system functions.

Development starts with comprehensive enterprise business process model. This model shows the main functions of the organization. It has nothing specific that would be exactly connected to such warehouse development. The aim of this model is to define the starting point from which further requirements specification and information modeling has to be done.

During the further steps requirements are elicited from business process models. It is done by searching decision points and choosing them together with data view. For example (Fig. 4), after Ministry has reviewed financing plans, sent by Budget institutions, it sends them to Budget department of Treasury. After review decision is made there are two possible statuses available – financing plans to approve and rejected financing plans. These data will associated with appropriate data warehouse data in next stage.

VII. Logical Data Model

The main logical data model in data warehousing is the multidimensional model, also called star schema. It is said to provide intuitive and high performance data analysis. The multidimensional paradigm allows data access in a way that comes more natural to human analysts. The data is imagined as located in n-dimensional space, with the dimensions representing the different ways the data can be sorted (e.g., sorted by time, by customer type, etc.). I am using the same approach in definition of data warehouse data structures in order to make this process more human readable.

Example of business decision data warehouse model is shown on Fig. 5. This diagram shows the process of budget allocation (Fig. 4) in more detailed view. The author adds extra perspective that describes the data warehouse dimension and the facts.

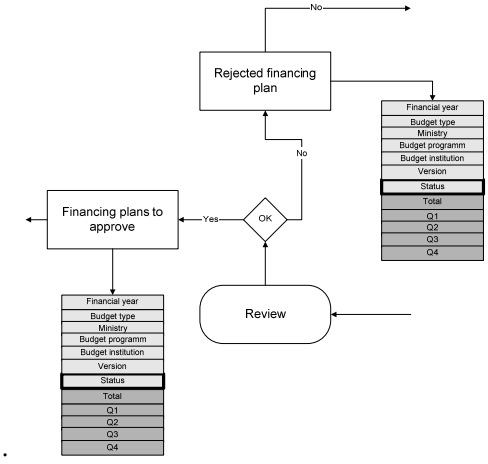

Fig. 5 Review stage of Budget allocation process

Budget allocation is one of the starting points in Budget execution process of Treasury of the Republic of Latvia. The aim of this business process is approval and accounting of financial plans of institutions and ensuring of opening and closing of budgetary book accounts in accordance with structure of accounts. Now it would be possible to do reporting on business process objects.

The final stage is to convert logical data warehouse model to physical model. This stage is performed manually depending on the particular data warehouse vendor.

VIII. Conclusion

In this work I described business process models oriented development of data mart. I have demonstrated theoretical aspects of methodology, but as a practical tool to model data analysis solution I propose ARIS framework and Event process chain (EPC) as a modeling notation. With the data warehouse perspective, it becomes possible to create models that show where and how business processes use decision support data. Organizing development in such way involves business users more in development, and I consider that such involvement reduces risks of data warehouse failures and help to develop the data warehouse that better fits business requirements. Visualization of data warehouse information model provides with necessary information to decide which dimensions and facts are important in every business process. Advantage of business process models driven approach is popularity of business process modeling. Enterprises already have developed business process models, it is necessary only to adopt them for data warehouse needs.

References

[1] W. H Inmon, R. D.Hackethorn, Using the Data Warehouse. New York: John Wiley & Sons (1994).

[2] P. Vassiliadis, C. Quix, Y. Vassiliou, M. Jarke. Data Warehouse Process Management, Information Systems, Vol. 26, No. 3, 2001.

[3] M.Klimavicius, U.Sukovskis, Business process driven data warehouse development, Scien-tific Proceedings of Riga Technical University, Computer Science Series, Applied computer Systems, 6th issue, Vol. 22, Riga, Latvia, RTU(2005).

[4] M.Klimavicius, U.Sukovskis, Risk management methodology for data warehousing projects, ISD'2005 Proceedings of the Fourteenth International Conference on Information Systems Development: Pre-Conference, Karlstad, Sweden, Karlstad University Studies (2005).

[5] V. Stefanov, B. List and J. Schiefer: Bridging the Gap between Data Warehouses and Busi-ness Processes - A Business Intelligence Perspective for Event-Driven Process Chains, Pro-ceedings of the 9th International Enterprise Distributed Object Computing Conference (EDOC 2005), Enschede, The Netherlands, IEEE Press (2005).

[6] V.Stefanov, B. List and B. Korherr: Extending UML 2 Activity Diagrams with Business Intelligence Objects, Proceedings of the 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2005), Copenhagen, Denmark, Lecture Notes in Com-puter Science, Springer Verlag (2005).

[7] Michael Boehnlein, Achim Ulbrich-vom Ende, Deriving Initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems, DOLAP 1999, ACM Second International Workshop on Data Warehousing and OLAP, Kansas City, Missouri, USA, ACM(1999).

[8] Kevin McDonald, Andreas Wilmsmeier, David C. Dixon, W.H. Inmon, Mastering the SAP Business Information Warehouse, New York, Wiley Publishing (2002).

[9] IDS Scheer AG, ARIS Methods manual, IDS Scheer (2004).
[10] Vaclav Repa, Business Processes Based Information Systems Development, BIS99 Confer-ence, Proceedings of the 3rd International Conference on Business Informaiton Systems, Poznan, Poland, Springer-Verlag (1999).
[11] Sergio Lujan-Mora, Panos Vassiliadis, Juan Trujillo, Data Mapping Diagrams for Data Warehouse Design with UML, Conceptual Modeling - ER 2004, 23rd International Confer-ence on Conceptual Modeling, Shanghai, China, Lecture Notes in Computer Science, Springer (2004)
[12] W. H. Inmon, Building the Data Warehouse, Fourth Edition, Indianapolis, Indiana, USA, Wiley Publishing (2005).
[13] R.Kimball, L.Reeves, M.Ross, W.Thornthwaite - The Data Warehouse Lifecycle Toolkit, John Wiley & Sons (1998).
[14] P.Ponniah, Data Warehousing Fundamentals: A Comprehensive Guide for IT Profes-sionals, John Wiley & Sons(2001).
[15] Vassiliadis, P., Simitsis, A., Skiadopoulos, S.: Conceptual Modeling for ETL Processes. In: Proc. of 5th Intl. Workshop on Data Warehousing and OLAP (DOLAP 2002), McLean, USA (2002).
[16] Luján-Mora, S., Trujillo, J., Song, I.: Extending UML for Multidimensional Modeling. In: Proc. of the 5th Intl. Conf. on the Unified Modeling Language (UML'02). Volume 2460 of LNCS., Dresden, Germany (2002).
[17] Stefanov, V., List, B. and Korherr, B., Extending UML 2 Activity Diagrams with Business Intelligence Objects, Proceedings DaWaK, Springer Verlag, (2005).
[18] Mazon, J.-N., Trujillo, J., Serrano, M. and Piattini, M. Applying MDA to the development of data warehouses. In Proceedings DOLAP'05, New York, NY, USA. ACM Press, (2005).
[19] Breton E. and Bézivin J. Weaving Definition and Execution Aspects of Process Meta-models. In 35th Hawaii Int. Conf. on System Sciences, (2002).