



PROCESS ANALYSIS AS DATA WAREHOUSE BASIS

Stanislava Šimonová¹ - Michal Vávra²

Key words: process analysis, data warehouse, data modeling

Abstract:

Organizace běžně pracují se svými operativními daty transakčních systémů, přičemž se nabízí příležitost využít tato data pro analytické výstupy ve smyslu vytvoření datového skladu (příp. datových tržišť). Výchozí situace pro fiktivní firmu X je - při budování datového skladu vycházet ze tří aspektů, jednak je to existující procesní analýza, jednak je to existující datový model transakčního databázového systému a jednak požadavky uživatelů. Provázání těchto tří aspektů je zásadní podmínkou pro vytvoření modelu datového skladu.

Initial status and first decisions

It is not possible to find out organization, where employees working without touch of modern information and communication technology. Good prepared and optimized technology can help to make business more effectively and with profit. The question is, how to transfer business process map (result of business analysis) to data model – as the basis for data warehouse structure. The way to useful data warehouse has more points. Before starting the way to useful data model, it is necessary to take up with start position. Company X, which made a decision start working on own data warehouse, is leader on market of automotive business. Company has more that 3000 employees in 5 countries of Europe and plan expand to another countries. Company use more in-house prepared application for business support. All this systems are based on Microsoft technology and transaction systems. That means that all reports are real time calculated. This wastes lot of time to get it. On the beginning is “overheating” of IT resources for more and more analysis and ad hoc queries to current systems, because management needs more and more data for their decision. And in this situation there is a question – does company need a data warehouse? Is it ready for one? But the decision is not so easy, before the answer from top management there is necessary run over lot of meetings about following question: will the data warehouse help the executives and managers to do better planning and make better decisions? Is it going to improve the bottom line? Is it going to increase market share? If so, by how much? What are the expectations? What does the management want to accomplish through the data warehouse? If the project fails, how much money will go down the drain? Management has to decide on the type of data warehouse to be built and where to keep it, where the data are going to come from and even whether you have all the needed data, who will be using the data warehouse, how they will use it, and at what times. There are only few of question, which are needed to

¹ Stanislava ŠIMONOVÁ, Univerzita Pardubice, Fakulta ekonomicko-správní, Studentská 85, Pardubice, tel. +420 466 036 009, e-mail Stanislava.Simonova@upce.cz

² Michal VÁVRA, Hewlett-Packard s r.o., Vyskočilova 1/1410, Praha 4, tel.+420 261 307 111, e-mail michal.vavra@centrum.cz

solve. And other important question – to build or buy? No one builds a data warehouse totally from scratch by in-house programming. There is no need to reinvent the wheel every time. The possible way can be compromise between two points, between build and buy. There are lot of transformations of data, that we can made in house, programs for loading the data warehouse storage can be also wrote by our programmers. And the buy option could lead to quicker implementation.

The business requirements

Most important think in whole project is one axiom: let business requirements drive your data warehouse, not technology [3]. Data warehousing is not only about technology, it is about solving users' need for strategic information. Do not plan to build the data warehouse before understanding the requirements. Start by focusing on what information is needed and not on how to provide the information. The basic structure and the architecture to support the user requirements are more important [1, 3]. In the development process is user that means person, which will use this system, the most important component of all data warehouse. Without this integration it is possible to take system, but users can not use this. Because data are perfect – but not for one business, these data can not help users get right decision. For example data are group by wrong area, from reports we can not see, what our point of view is. And in this case there is an expensive data storage place, it is written in annual report, that there is the data warehouse – but this is all. And this is wrong way. A data warehouse is not a single software or hardware product you purchase to provide strategic information. It is, rather, a computing environment where users can find strategic information, an environment where users are put directly in touch with the data they need to make better decisions. It is a user-centric environment. Users' requirements are building elements of our data warehouse.

User-oriented developing is important. But well prepared infrastructure is of course also very important. Without good interaction between user requirements and corresponding hardware is building of data warehouse mission impossible. The overall functional components of a data warehouse are data acquisition, data storage, and information delivery. These three functional components form the general architecture of the data warehouse. There must be the proper technical infrastructure to support these three functional components. The phases must also include tasks to define the architecture as composed of the three components and to establish the underlying infrastructure to support the architecture. The design and construction phase for these three components may run somewhat in parallel. But there is a “gap” in the development process – the gap between process analysis and data model design. The question is: How to transfer description of processes in company to facts and dimensions in data warehouse?

Process analysis – the start point

Process analysis can be defined as "a logical series of related transactions that converts input to results or output". The process we are considering is a "business process," which can be defined as "a chain of logical connected, repetitive activities that utilizes the organization's resources to refine an object for the purpose of achieving specified and measurable results or products for internal or external customers." [4]. An operation is composed of processes designed to add value by transforming inputs into useful outputs. Inputs may be materials, labor, energy, and capital equipment. Outputs may be a physical product (possibly used as an input to another process) or a service. Processes can have a significant impact on the performance of a business, and process improvement can improve a firm's competitiveness [2]. The first step to improving a process is to analyze it in order to understand the activities,

their relationships, and the values of relevant metrics. Process analysis generally involves define the process boundaries that mark the entry points of the process inputs and the exit points of the process outputs. Next step is to construct a process flow diagram that illustrates the various process activities and their interrelationships. Very important is also identification the bottleneck, to use the analysis to make operating decisions and to improve the process.

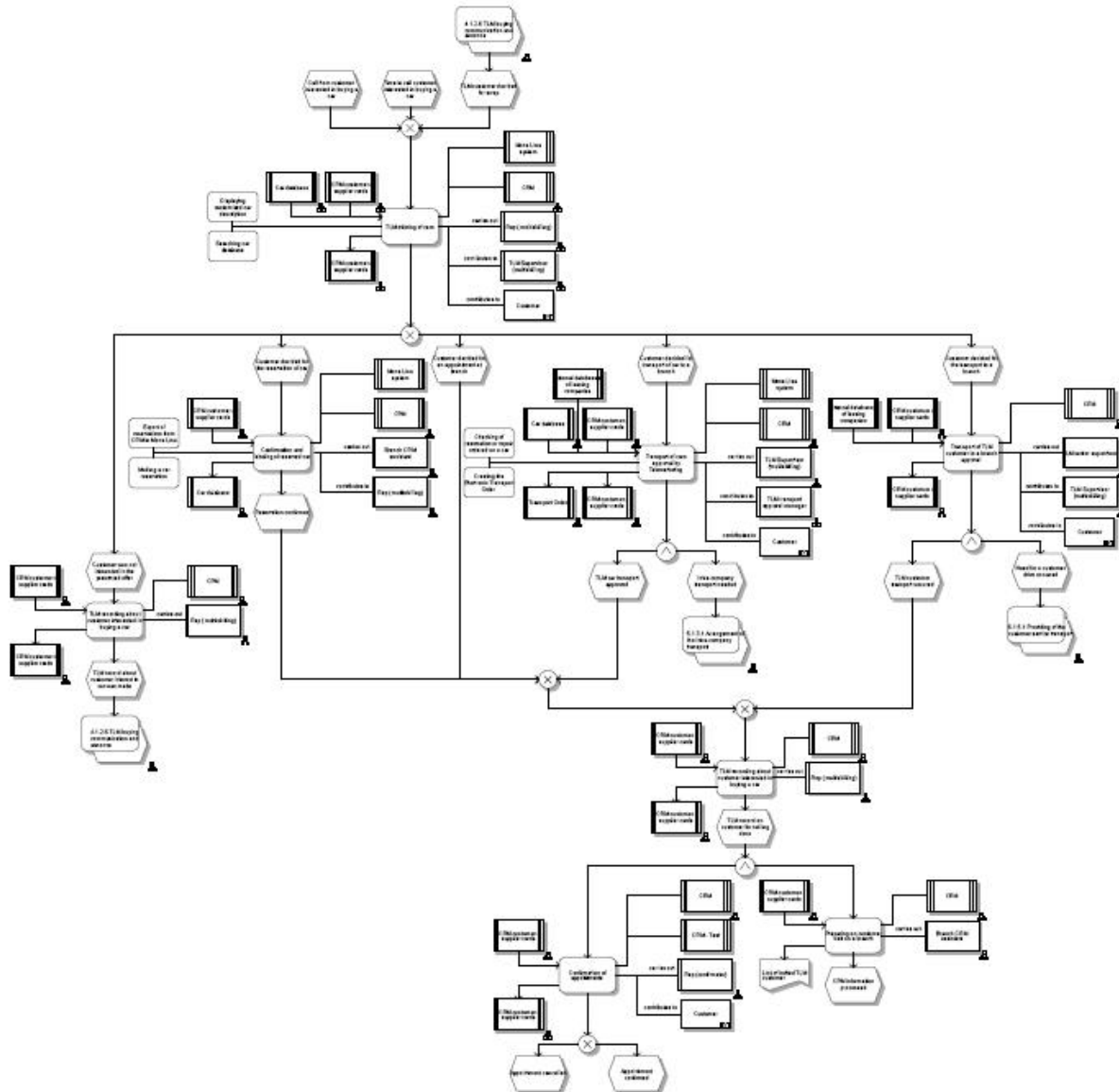


Figure 1 Process map of Company X (example, source [5])

There are more ways, how to get process analysis. The first one is complete outsourcing of whole analysis; it means that we choose another company with experience in business process mapping. This way is quick, but expensive. And there is here also one risk – know how theft. Management of company must be sure, that won't take into business secret another person. On the second pole it is possible to found in house made analysis. This type of analyses is cheaper, but only few companies has well educated business analyst to do this mission. Between these two ways is the third one – cooperation of external company and our specialist. In this case we have persons with required knowledge and our specialist can help them. Decision, which way we will use, is on our turn. In any case (of course, when process analysis will end successfully) are outputs maps of processes in the company. These maps are

sets of documents with process flow diagram, description of process, IS function in the model, reports and files in the model data cluster in the model etc. refer to figure 1.

Information packages

The background is – there is process analysis, there is a description of database objects (existent or new created tables view). That means, it is known, which tables store related data. To help to produce good data model for data warehouse it is possible to use some schema of database object. According to business requirements about data warehouse outputs it is possible to start design structure of this. Other idea for determining and recording information requirements for a data warehouse are information packages [3]. This concept helps to give a concrete form to the various insights and opinions expressed during the process of collecting requirements. The users are unable to describe fully what they expect to see in the data warehouse. We can not determine how each class of users will use the new system. So, when requirements cannot be fully determined, then is needed another concept to gather and record the requirements. The users tend to think in terms of business dimensions and that is why is necessary to analyze this business dimensions.

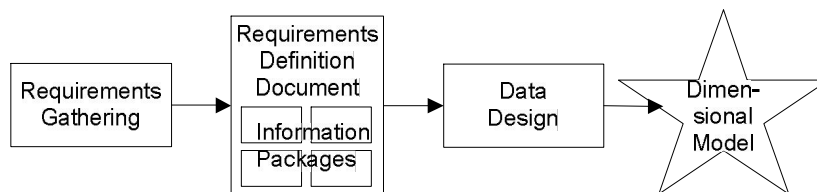


Figure 2 From requirements to data design (source [3]).

This methodology for determining requirements for a data warehouse system is based on business dimensions. It incorporates the basic measurements and the business dimensions along which the users analyze these basic measurements. Here we can come up with the measurements and the relevant dimensions that must be captured and kept in the data warehouse. What is necessary to do first in definition phase? Compile information packages for all the subjects for the data warehouse. Once you have firmed up the information packages, you'll be able to proceed to the other phases. With information packages it is able to define the common subject areas, design key business metrics, presentation of data, grade aggregation of users, decide data quantity for user analysis or query, how will be data accessed based on data granularity. Also it is possible to estimate data warehouse size and determine the frequency for data refreshing. Based on this prepared information packages it is able to start with data design. Data design consists of putting together the data structures. The results of the requirements gathering phase is documented in detail in the requirements definition document. An essential component of this document is the set of information package diagrams. Now is time for design decisions - to select the subjects from the information packages for the first set of logical structures to be designed and also determining granularity of data. After this comes choosing of business dimensions to be included in the first set of structures, choosing the facts (metrics or units of measurements – sale units etc.) and also how far back to store historical data.

Dimensional model of Company X

Based on rule describe in previous text it is possible to start build data model for the data warehouse of company X (for example figure 3). After this the next steps of successfully implementation of data warehouse will come. But it means not the end of the way to build something for company business. The questions are: How we will know, that our work was

effective and useful? How do we know that data warehouse is a success? Do we need 3 or 5 years to see if we get the ROI (return on investment) proposed in our plan? There are some such indications of success that can be observed within a short time after implementation.

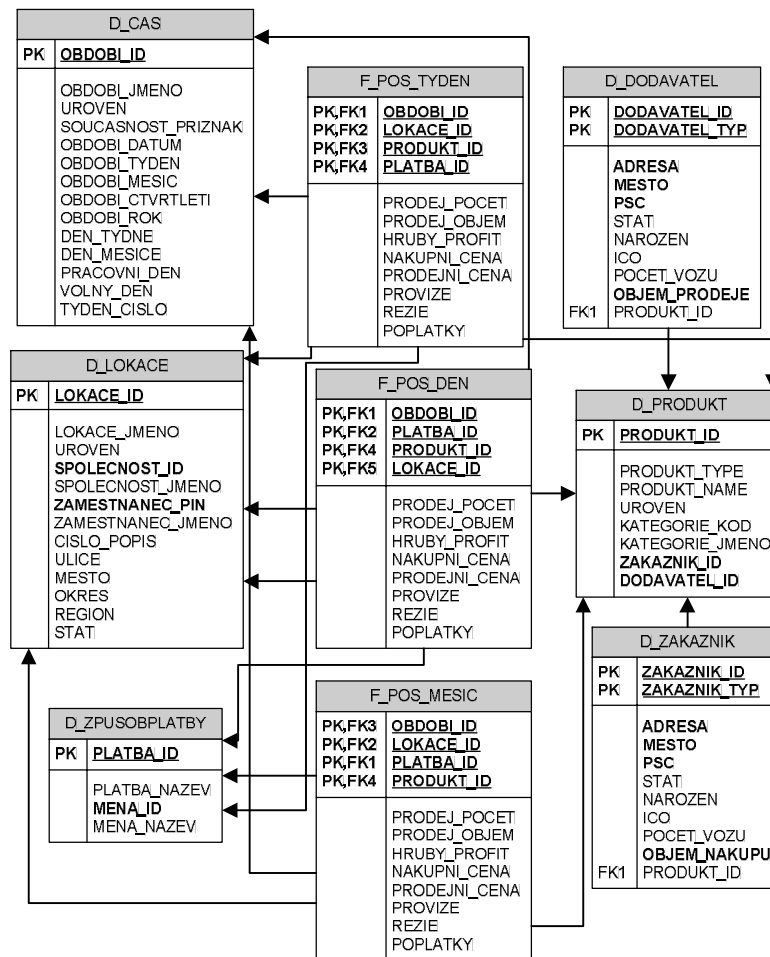


Figure 3 Data model for data warehouse of company X

Success indicate generally the following happenings - rapid increase in the number of queries and reports requested, queries becoming more sophisticated, number of active users steady increase and users spending more time in the data warehouse looking for solutions. Effective / quality data of data warehouse are basis for strategic information of the company.

References:

- [1] INMON, W. H. *Building the Data Warehouse*. New York: John Wiley & Sons, Inc. 2001. 410 s. ISBN 0-471-08130-2.
- [2] Net MBA – Business Knowledge center. *Process analysis* [online]. [cit. 2008-06-18]. Available from WWW: <<http://www.netmba.com/operations/process/analysis/>>.
- [3] PONNIAH, P. *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. New York: John Wiley & Sons, Inc. 2001. 518 s. ISBN 0-471-22162-7.
- [4] University of Central Florida. *What is Process Analysis* [online]. [cit. 2008-06-02]. Available from WWW: <http://oeas.ucf.edu/process_analysis/what_is_pa.htm>.
- [5] VÁVRA, M. *Datový sklad – od procesní analýzy k implementaci*. Pardubice: Univerzita Pardubice, 2008. 89 s. Bakalářská práce na FES Univerzity Pardubice.