IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4470

Deploying the Business Case to Reduce the Risk of Large Scale Data Warehousing

Reinhard Jung and Stefan Schwarz
University of St. Gallen, Mueller-Friedberg-Strasse 8, CH-9000 St. Gallen,{Reinhard.Jung | Stefan.Schwarz}@unisg.ch

ABSTRACT

Managers of large data warehousing projects often put project failure down to organizational resistance. Technical requirements are usually not considered to be crucial for project success. However, the majority of the scientific work on data warehousing concentrates on technical aspects. As a consequence, a comprehensive framework or method for the introduction of a data warehouse is still missing. This paper takes this contradiction into account and deals with the management of organizational risks in large scale data warehouse projects. We base our research on information gathered from large organizations which develop and/or run data warehouses.

This paper is structured as follows: After a short introduction the planning process of data warehousing is outlined. The first step is the strategic decision for data warehousing which is followed by the definition and evaluation of the initial project which is also called first increment. In the third section we present a business case strategy for both the initial project and subsequent data mart projects. Furthermore, the interdependencies between the initial phase and subsequent projects are discussed.

1 INTRODUCTION

1.1 Competence Center Data Warehousing Strategy (CC DWS)

In January 1999 the Competence Center Data Warehousing Strategy (CC DWS) was founded at the University of St.Gallen (Switzerland). The CC DWS is a joint research project of the Institute of Information Management and ten partner corporations from the insurance, logistics, telecommunications and consulting industry, and the Swiss department of defense. Some of our partners already have one ore more data warehouses, others are in the middle of the development process. CC DWS research focuses on methodological aspects of data warehousing. In this paper we present our findings as regards the business case for data warehousing (DWH) projects. However, it will become obvious that organizational aspects are strongly interrelated with the business case issue.

1.2 Related work and research method

In a review of current research on data warehousing a comprehensive and accepted framework to organize organizational risk of large scale data warehousing projects could not be found. However, literature focussing on various relevant aspects is available.

Project prioritization which is of significant importance for data warehousing especially in large organizations is a rather broad field of research. However, most of the available publications are based on discrete algorithms for benefit assessment which we consider inappropriate in the case of data warehousing. A general approach for the prioritization of IT projects is presented by Jung (Jung 1995, pp. 544). Fröschle and Niemeier have worked on benefit assessments for IT infrastructure investments (Fröschle, Niemeier 1988, pp. 190-197). Case study research about the assessment of decision support systems is introduced by Belcher and Watson (Belcher and Watson 1992) and Gallivan (Gallivan1994, pp. 65-77). The problems that arise from replacing legacy systems or their integration into modern ones are discussed by Robertson (Robertson 1997). Especially his illustrations of the complexity caused by legacy system integration is of special interest.

Our work in the areas of data warehousing is based on action research. By interviewing employees of CC DWS partner corporations and associated corporations we gathered information on the main areas of interest and analyzed best practices in order to find possible solutions. As the interviews delivered mostly soft facts, their generalization has to be done in an interactive and revolving process.

None of the partner corporations interpreted the business case issue in such a broad sense as we do in this paper. Therefore, each corporations' practice only covers part of the comprehensive business case strategy. However, some of them adapted the generalized findings to their individual settings and thereby contributed to the validation of our results.

In every qualitative research project a certain level of bias can be observed. In order to avoid biases interim results were controversially discussed on a regular basis both within the research team and with employees of the partner corporations.

1.3 Basic assumptions

Our work is based upon a general three-tier DWH architecture (cf. Figure 1) which proved to be suitable for describing a variety of different real-life architectures including those of our partner corporations.

Other authors, e.g. Bontempo, Zagelow 1998, Gardner 1998, Kimball et al. 1998, present quite similar architectures. The basis of every data warehouse architecture is the operational IT environment and especially its data sources. The next layer deals with the so-called ETL processes, i.e. extraction, transformation, and loading of detailed data into the core data warehouse (core DWH) or the operational data store (ODS). In contrast to transactional databases, the core data warehouse comprises both actual and historical data in order to support all kinds of analyses. The ODS is designed to serve real-time applications such as call centers. Since the data within these central components of the architecture is detailed and not aggregated, it is necessary to define specific views on this core data for the business units. If the views are materialized, i.e. if controlled redundancy is introduced, we call the set of views for one business unit a data mart. These views can provide aggregate data and denormalized or even multidimensional data structures. The top layer of the architecture can comprise end-user tools for ad hoc queries, online analytical processing (OLAP), data

mining, and real-time applications.

2 PLANNING FOR DATA WAREHOUSING

2.1 The strategic decision for data warehousing

The trend towards data warehousing is a result of current market forces and conditions. The initial trigger is to be seen in the transition from producer oriented markets to consumer oriented markets. The IS which have been implemented over the past decades are orientated towards divisions or products, i.e. there is usually at least one IS with its own database for every product. An example for this state of affairs is the insurance industry: there is usually a dedicated IS for life insurance, liability insurance etc. However, today's customers are demanding individualized and integrated services, e.g. consolidated invoices, which cannot be granted by an uncoordinated number of product oriented IS.

In large organizations an enterprise-wide data warehouse, either physically or logically centralized, cannot be realized in a single step ("big-bang approach"). Instead, it takes several steps or projects to provide a comprehensive and management-oriented database (Sigal 1998). Therefore, the decision for data warehousing is usually a strategic one, which is or should be made by the top management (also cf. Clemons 1991). However, it is very difficult or even impossible to make strategic decisions based on monetary criteria. Thus, the decision for or against data warehousing is made on qualitative criteria because the derivation of the important decision parameters based on a relatively long timespan.

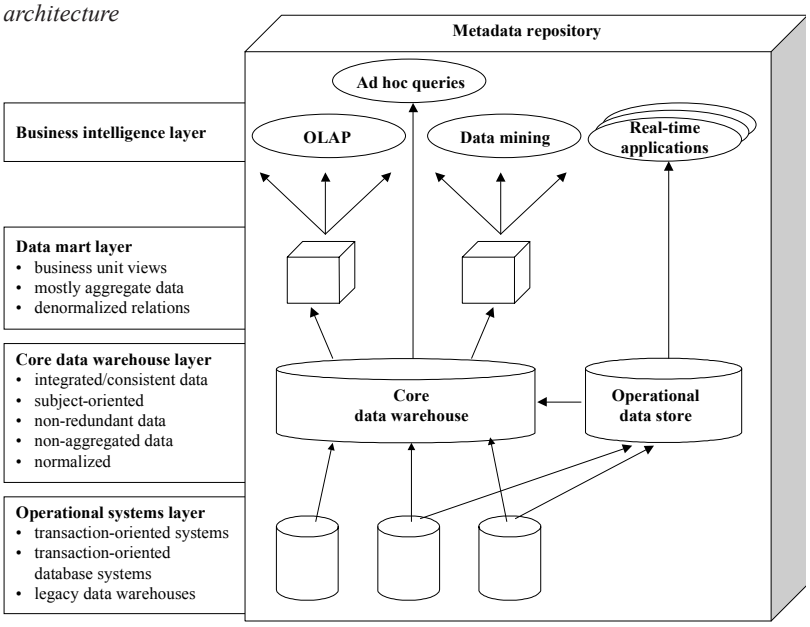
2.2 Defining and evaluating the initial project

The scope of the initial project is crucial for the success of the entire warehouse and especially for the funding of subsequent projects (Devlin 1997, p. 311). In Figure 2 the different projects types as regards the initial project are depicted:

- Project type A: The informational need of one business unit is to be covered by a single data mart. Further projects include the implementation of additional data marts and the integration of additional data sources if needed.
- Project type B: A core data warehouse which is limited as regards its coverage is provided as a corporate infrastructure. Subsequently, business units will be asked to formulate their requirements in order to be able to specify data mart projects. A discussion of influencing factors as regards the prioritization of these projects can be found below.
- Project type C: The informational need of all business units including the top management is to be covered by the core data warehouse. Therefore, on the one hand all relevant data sources have to be integrated and on the other hand several data marts have to be implemented.

While project type A is adequate for corporations independently of their size, type C is only suitable for corporations with both a small number of business units and a small number of OLTP systems. As a consequence, type C is viable only for medium and small size corporations. Project type B bears a high risk potential and should therefore be avoided. Without at least one business intelligence application and therefore without a reference project it will be almost impossible to provide visible benefits which could ensure an ongoing funding and development of the DWH.

Figure 1: Data flow (examples) in a general three-tier data warehouse architecture

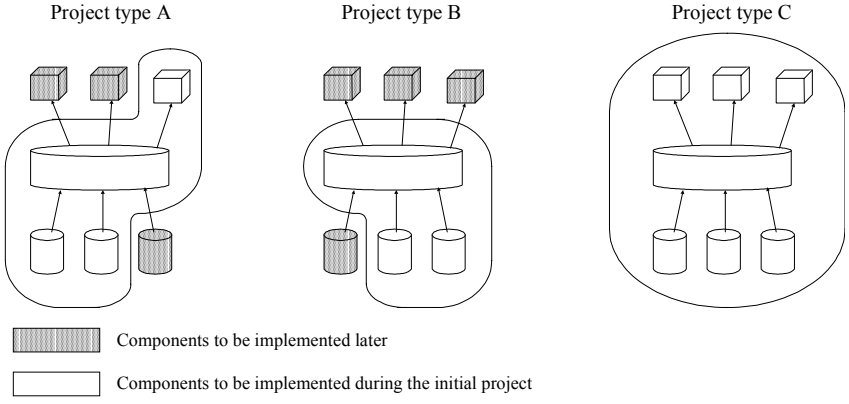


3 DATA WAREHOUSE BUSINESS CASE STRATEGY

The preceding sections concentrate on mainly static aspects of a DWH business case. In the following we are going to integrate these aspects into a strategy. We present strategy patterns and structure them in order to establish a DWH strategy. This strategy provides recommendations depending on context factors we identified in cooperation with our competence center partners. Based on the general architecture described above (cf. Figure 1) the DWH strategy problem can be divided into two sub tasks:

- Design and implementation of the initial project as described in section 2.2.
- Design and implementation of several subsequent data mart projects. Within such a project a new data mart is built and additional base data might be added if needed. Sutter also favors a subdivision of the entire development process into small

Figure 2: Different project types for the initial project



projects. He uses a case study to explain the benefits of such action which he mainly sees in the lower risk of DWH project as a whole (Sutter 1998, pp. 49-51).

Because of significant differences, these two phases have to be dealt with separately. The most important difference between the initial project and a subsequent data mart lies in the range of addressees. Whereas the initial project is mainly to be judged as a common infrastructure project and therefore has a variety of addressees, the implementation of a data mart only provides benefits for the business unit it is built for.

Another aspect that differs between design and implementation of the initial project and that of data marts is the complexity. An ideal core DWH must be flexible enough to be adjusted to all future business driven information needs. These information needs are usually unknown in the beginning. Furthermore, information needs are generally expressed from different business units and may therefore have contradicting intentions. The different intentions lead to different requirements. An example for this, which can be found in almost every corporation, is the definition of management terms. The definition of the term "turnover" for instance can be quite different if seen from the controlling point of view or from a sales perspective. Both business units have good reasons for their definition so that sometimes considerable integration and standardization efforts have to be made. This short example clarifies the additional requirements the design of a core DWH has to cover in comparison to the design of a data mart.

Similar to most infrastructure projects, the design and implementation of the initial project suffer from an asynchronous occurrence of costs (now) and benefits (later). Furthermore these costs are, compared to a data mart project, significantly higher.

Common IT practice suggests that the project sponsor of a data mart project should be the head of the business unit the mart is designed for (a detailed discussion can be found in Inmon et al. 1997). The initial project serves the whole corporation. Therefore, the determination of the project sponsor for the initial project is not easy. As a consequence, corporations have different perspectives on that issue. In many cases the CIO signs responsible for this project, in other cases the top management takes this role. It is not possible to formulate an ideal strategy as regards this issue but two general guidelines can be given:

- Since the initial project has an impact on different business units and, therefore, bears a lot of potential for conflicts, the project sponsor should be the manager the units are reporting to. According to his position he would be capable of moderating the search for compromises.
- The project sponsor of the initial project will have a special role throughout the development process of the DWH. Therefore, he is required to understand the implications and opportunities of data warehousing. One of his foremost tasks will be to communicate and promote the idea of an active data management within the corporation.

3.1 Prioritization of data mart projects

As described above, the initial project is an infrastructure project to be carried out in the first place. The data mart projects have to be scheduled subsequently. They can be prioritized according to different strategies. In the following we describe three basic prioritization strategies. These strategies suggest the existence of distinct approaches. However, usually mixtures of these strategies can be found. The pros and cons of the three strategies are described below:

- The "internal customer" strategy

  Most of today's modern corporations are service oriented. Each business unit has to find either an internal or an external customer for its "products". As a consequence, internal customers have to pay for IT services. According to this idea, the prioritization of data mart projects can be done based on monetary considerations. Each business unit has to prognosticate the benefits a data mart is likely to provide. Afterwards, it has to determine the internal price it is willing to pay for the data mart. Due to the fact that the benefits resulting from better decisions are almost intangible, the latter task is very difficult.

  When the implementations costs and the benefits for all the possible data mart projects are determined, the prioritization can be done based on available approaches (Dué 1989, Goodhue et al. 1992, Sassone 1987).

  Especially in large corporations it could be one aim of a DWH project to replace decentralized legacy DWH systems. For the owners of these legacy systems, i.e. the business units, a replacement often appears to be unnecessary, because the old system delivers all the information needed. As a consequence, the business unit might not be willing to pay for the implementation of a new DWH at all. This attitude can, from the point of view of the business unit, be considered as rational. However, it most likely leads to inefficiencies for the overall system. The only solutions for this problem is to offer a funding by overheads.
- The "technical" strategy

  In most cases, the IT department dominates the DWH projects. In a first step the IT department tries to determine the information needs of each business unit. This is often done through interviews or analyses of the information profile. Afterwards, the IT department derives requirements and tries to meet them at minimal costs. However, cost minimization guided by a technical perspective can lead to an inferior sequence of projects as far as business benefits are concerned. The costs of a specific data mart project are highly dependent on its position in the sequence of all data mart projects. This is mainly a result from the effort for improving data quality and for the loading of the data. A suitable scheduling of the projects can help to enable an efficient implementation and therefore might avoid a lot of redundant work.

  From a technical point of view it can also be appropriate to start with a project which requires only little effort in order to gain experience easily and deliver results very early ("quick wins"). On the one hand, the "quick wins" are often considered to have an accelerating effect on the overall project. On the other hand, a low-effort project is normally not a technical challenge. Typically, the number of source systems is very limited. As a consequence, a progressive learning curve might not be achieved. Furthermore, data integration which considers only few systems cannot provide significant improvements as regards information supply. This can lead to a negative attitude of the management towards the entire DWH project.
- The "political" strategy

  The "political" strategy can be found in corporations which are focused on specific business functions. Since customer orientation became a common paradigm, many corporations pay more and more attention to their marketing and/or sales activities. Every action is judged according to its grade of support as regards these activities. As a consequence, the prioritization of the data mart projects is also dominated by this paradigm. Thus, the full benefits from data warehousing might not be achieved. We will prove this statement by some general considerations concerning the service industry. A customer assigns a value to a service depending on the processes or activities he recognizes. The sequence of those processes or activities is called *service chain.* Mostly, the value a customer assigns to the overall service chain is higher than

the sum of values he assigns to the individual activities. The choice of activities and their order provide the added value. This leads to the conclusion that the service chain should be completely supported, rather than only partially.

Some authors even claim, that a poor performance of one activity is hardly to be compensated by other, better performing activities. If we adapt this argumentation to the issue of information supply, the benefit of an excellent information supply of a specific functions can easily be overcompensated by an inadequate information supply of other functions.

3.2 Business case aspects of the warehousing process

DWH project failures can be put down to various reasons, technical ones only seldom prevail. In most cases the following aspects prove to be more crucial:

- After designing and implementing the core data warehouse, it is not possible to convince enough business units of the value of a business unit specific data mart. Without a powerful project sponsor a no-go decision as regards the overall DWH project becomes likely.
- Data marts which have already been implemented are very useful for demonstrating benefit potentials. As a consequence, the demand for new data marts is likely to be strong. The first data mart projects in a corporation are often selected due to the low effort required and "quick wins". Subsequent projects are typically much more complex. A huge number of project applications with a high degree of complexity can lead to an overstrain of the DWH team. As a consequence, the demand for data marts stimulated in the beginning cannot be satisfied. If this point is reached, the success of the overall DWH project becomes questionable.
- No internal marketing procedure for the DWH project exists. Additionally, only few project applications have been formulated with, as the worst case, non-overlapping semantic focuses. Due to the low number of applications the DWH team has to accept almost every application in order not to run out of projects. Furthermore, a low number of project applications renders an optimal implementation sequence very difficult. As a consequence a lot of design and implementation work has to be done redundantly.

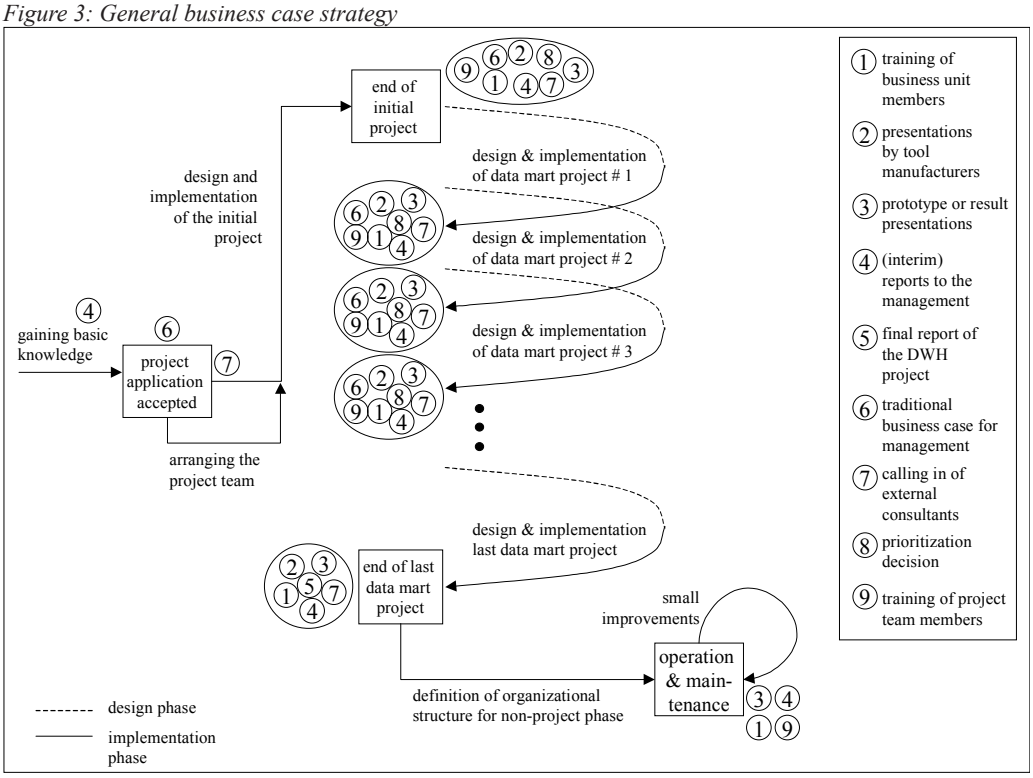
In order to avoid situations as described above, a suitable strategy has to be formulated. In the following we identify activities that constitute viable strategies. We demonstrate how these activities can be ordered to build different strategies with regard to context factor combinations that often occur in practice (cf. Figure 3).

A core team (often only one to two persons) has to gain basic DWH knowledge before a DW H project can be started. The team should be able to give a rough estimation of a DWH project's costs and benefits. Since the first decision is a strategic one, the core team has to provide a complete list of pros and cons.

The next deliverable is the formal application for the initial project. The corresponding activities are crucial for the project because they will have a strong impact on top management support for further activities (also cf. Watson, Haley 1998, pp. 34). As described above, at this early stage only the costs of the initial project can be estimated. Most of the benefits are intangible. Nevertheless, in many corporations top management requires a traditional business case.

If the project application is accepted, the next activity to be carried out is the staffing of the project. Several reasonable compositions for the project team exist. The main points to be considered are a good balance of IT and business knowledge in the team, full-time project participants, and a project sponsor as described in section 3. It is often useful to call in external consultants at this stage of the project, because the internal DWH knowledge is often limited. The project management, however, should remain internally. On the one hand a project led by a consultant often gains quick wins. On the other hand our practical experience shows that in such projects DWH knowledge is of little persistence and the motivation of its members is low. A good possibility to gain highly persistent knowledge is to train the members of the DWH team which often improves the motivation of the team members.

Figure 3: General business case strategy



Once the initial project is finished, the (positive) results should be presented in order to ensure the further development of the DWH project. A presentation is appropriate if the results of the initial project are easy to understand and its business value is easy to be recognized. Additionally, tool presentations may help business units to assess the benefits a data mart can provide.

Internal marketing activities like these have to be carried out carefully because too many appli-

cants can easily increase the pressure on the overall project (cf. section 3.1). As the presentation of the results helps to attract internal customers, the report to the management helps to extend their support. At this stage of the project the first benefits should have been realized and can be partly measured more precisely.

The next step of the process is the design and implementation of the first data mart. As a result of the internal marketing activities, a pool of data mart project applications should exist. One of the projects has to be selected considering the guidelines described in section 3.1. If the internal customer strategy is chosen, a business case has to be elaborated to support the decision. Since the initial project and the data mart projects differ significantly, the calling in of consultants can be reasonable. They should be able to understand the information needs of the specific business units and transform them into a corresponding specification. Due to the variety of tools that are likely to be deployed, the project team has to be trained. As shown in Figure 3 the activities described above have to be repeated for each data mart project.

After all data marts have been added to the DWH system, it should have reached a status which should allow to change the organizational embedding. At this point it is very likely that implementation efforts will decrease significantly. Therefore, the project phase of the DWH lifecycle is completed. An organizational unit which is responsible for operation and maintenance of the DWH has to be built. A mistake often made at this stage is to staff the new organizational unit exclusively with technically skilled employees. This renders the adaptation of the DWH to new business requirements more difficult because the unit lacks business knowledge.

Even at this stage internal marketing activities are of great importance. Otherwise the DWH might lose its management attention. A continuous improvement reporting to the management and to business units involved can be a basic means of internal marketing. Furthermore, ongoing training of the business unit members and suitable support services help to ensure the acceptance of the DWH.

4 SUMMARY

In this paper we presented some building blocks for a comprehensive data warehousing process model. In cooperation with our partners we were able to identify restrictions and to develop guidelines for the preparation of business cases. Finally, we tackled some organizational aspects of data warehousing. In further research, the business case issue has to be refined. A detailed catalogue of potentials which can be enabled by data warehousing would facilitate the assessment of business benefits.

5 REFERENCES

Belcher, L.W., Watson, H.J.: Assessing the Value of Conoco's EIS, MIS Quarterly, Vol. 17 (1993), pp. 239-253.

Bontempo, C., Zagelow, G.: The IBM data warehouse architecture, CACM, Vol. 41 (1998), No. 9, pp. 38-48.

Clemons, E.K.: Evaluation of Strategic Investments in Information Technology; CACM, Vol. 34 (1991), No. 1, pp. 22-36.

Devlin, B.: Data Warehouse – from Architecture to Implementation; Addison-Wesley: Reading et al. 1997.

Dué, R.T.: Determining Economic Feasibility: Four Cost/Benefit