

IBM Research Report

Risk-Management for Data Warehouse Systems

Robert Bruckner, Beate List
Institute of Software Technology
Favoritenstr. 9-11/188
Vienna University of Technology
A-1040 Vienna, Austria

Josef Schiefer
IBM Research Division
Thomas J. Watson Research Center
P. O. Box 704
Yorktown Heights, NY 10598



Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

Risk-Management for Data Warehouse Systems

Robert M. Bruckner¹, Beate List¹, Josef Schiefer²

¹Institute of Software Technology, Favoritenstr. 9-11 / 188
Vienna University of Technology
A-1040 Vienna, Austria
{bruckner, list}@ifs.tuwien.ac.at

²IBM Watson Research Center, 30 Saw Mill River Rd.
Hawthorne, NY 10532
josef.schiefer@us.ibm.com

Abstract. Data warehouse projects are notoriously difficult to manage and many of them end in failure. One explanation for the high failure rate is that managers are not taking prudent measures to assess and manage the risks involved in the data warehouse project. In this paper we introduce the goal-driven risk management approach of the easyREMOTE^{DWH} framework. We classify the risks of a data warehouse system into categories of risk sources and discuss their consequences to the data warehouse project. Further, we show the relevance of risks to the go/no-go decision of the data warehouse project. Finally, we present an approach for documenting data warehouse risks.

1. Introduction

Data warehouse projects are characterized by high decision stakes and high levels of system uncertainty. As such data warehouse projects are difficult to plan and develop with traditional project management methods. Because of the peculiarities and characteristics of data warehouse systems, the project management for such systems is a challenging task. Data warehouse systems are complex, large-scale and constantly evolving systems, which integrate the data of many software systems of an organization. They are used by a wide-range of users with very heterogeneous requirements.

Project managers of data warehouse projects have to cope with new major challenges, which also have a strong impact on the risk management activities. Following characteristics of data warehouse systems make a risk management different to traditional software projects:

User Heterogeneity. Because data warehouse systems are designed and constructed to address concrete business problems or opportunities, the users drive the data warehouse requirements. Data warehouse systems have to cope with a wide heterogenous range of users. The data warehouse requirements have to express all the needs of these users, but also arising conflicts between the user requirements have to be resolved.

Growth and Scalability. A user-driven data warehouse system typically grows fast - in numbers of users and volume of data. For instance ERP sources fuel this rapid expansion because they contain huge amounts of detailed data, which users may want to analyze inside a data warehouse. The constant growth of historical business data can cause bottleneck problems for the data warehouse system.

System Evolution. Data warehouses are high-maintenance systems. Reorganizations, product introductions, new pricing schemes, new customers, changes in production systems, and so on are going to affect the data warehouse. If the data warehouse is going to stay current (and being current is absolutely indispensable for user acceptance), changes to the data warehouse have to be made without delay. Therefore, the data warehouse system has to evolve with the business trends. Risk management has to consider the evolutionary development.

Open System Architecture. Ideally, all data warehouse applications should be supported by a common enterprise architecture that standardizes processes, components, and tools, eliminates redundant activities, and rationalizes business rules and data definitions. Proprietary data warehouse solutions and the lack of supporting existing standards for data warehouse systems can be a major risk factor for a data warehouse project.

System Integration. Data warehouse systems are fed from a wide range of software systems of the organization. Therefore, most of the effort for building a data warehouse system is integrating the data of those software systems.

Many data warehouse risks are reflected by one or more of these characteristics of a data warehouse. This paper presents the easyREMOTED^{DWH} approach, which addresses these risk sources and provides a solid framework for managing these risks including tool support.

The remainder of this paper is organized as follows. In section 2, we discuss the contribution of this paper and related work. In section 3, we classify data warehouse risks and discuss risk consequences to the data warehouse project. In section 4, we discuss the impact of data warehouse risks to the go/no-go decision of a data warehouse project. In section 5 and 6, we introduce the easyREMOTED^{DWH} risk management process and discuss its activities in detail. Section 7 shows templates of the easyREMOTED^{DWH} prototype, which can be used to document risks. Finally, in section 8 we present our conclusion.

2. Contribution and Related Work

Since the 1970s, both academics and practitioners have written about managing risks of software projects [2, 3, 5]. Unfortunately, much of what has been published on risk management is based on studies limited to a narrow portion of the development process of software projects. With a few exceptions [1, 6], there has been little attempt to understand the relative importance of various risks or to classify them in

any meaningful way. The characteristics of data warehouse systems are different to normal software projects [8] and therefore, there is a need for a more systematic investigation to identify the major risks that can impact a data warehouse project to classify these risks, and to develop appropriate risk management strategies.

Most risk management approaches do not explicitly support different stakeholder perspectives [3, 4], and those that do, often limit the number of stakeholders and assume that consensus can be reached [9]. Kontio presents in [7] the risk management method Riskit that is based on sound theoretical principles and thus avoids many of the limitations and problems that are common to many other risk management approaches in software engineering. Riskit also includes goals for representing different perspectives of stakeholders. While the Riskit method can be applied in many other domains, such as business planning, technology selection etc., it has been originally developed for software development projects. Therefore, its main features correspond to the risk management concepts and practices required in traditional software projects.

Boehm introduces in [2] a risk management approach for software projects. The projects and environments that were the basis of Boehm's work might not, however, be representative for data warehouse projects. Furthermore, both the organizational and technological landscape has changed considerable since Boehm's work appeared: new organizational forms and systems development environments have emerged, new mechanisms have evolved for acquiring systems (e.g. outsourcing and strategic alliances), and centralized, web-focused systems architectures open new business opportunities for organizations. For all of these reasons, we propose a reexamination of the risk management issue for data warehouse systems.

3. Categories and Consequences of Data Warehouse Risks

Risk management sets forth a discipline and environment of proactive decisions and actions to continuously assess what can go wrong in the data warehouse project, what risks are important to consider, and what strategies to implement to deal with those risks. Data warehouse projects face many kinds of risks besides those related to the project scope and the requirements (see Figure 1).

Dependency on external entities, such as subcontractors, external systems or other related projects, is also a common source of risks for data warehouse systems. The project management is fraught with risks caused by poor estimation, rejection of accurate estimates by managers, insufficient visibility into project status, and staff turnover. A disconnection between the data warehouse and business objectives can be the reason for unfulfilled business returns. The absence of an executive commitment can endanger the sponsorship for the data warehouse system. The failure of constant selling and promoting the data warehouse system internally to the users can result in a rejection of the new system. Escalated project scope by additional data subject areas or including other user groups makes an accurate identification and measurement of deliverables difficult. The bundling of a data warehouse project with other projects can be easier to justify or fund. Because data warehouse systems are complex and involve various kinds of technologies, technology risks can arise.

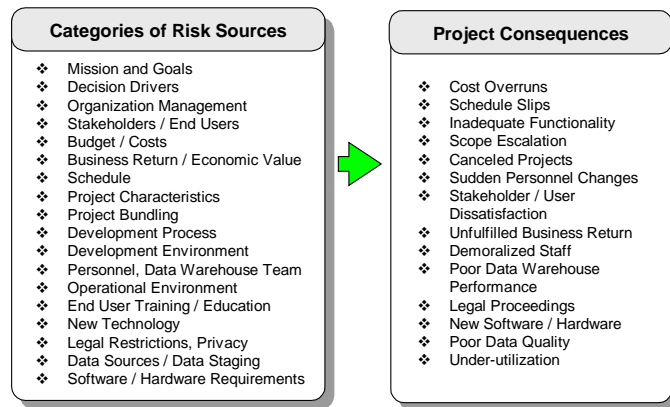


Figure 1: Risk Sources and Consequences for Data Warehouse Projects

Lack of knowledge is another source of risk, as with practitioners who have insufficient experience with technologies being used or with the application domain. Successful data warehouse project must respond to changing user community and their information needs. Lack of an ongoing post-implementation support can cause under-utilization of the data warehouse system.

4. Impact of DWH Risks to Go/No-Go Decision

The go/no-go decision of a data warehouse project is strongly influenced by the project risks. Besides the analysis of the risks, also measuring the required effort and determining the values for the stakeholders, makes a composite assessment of the overall worth of the data warehouse system possible.

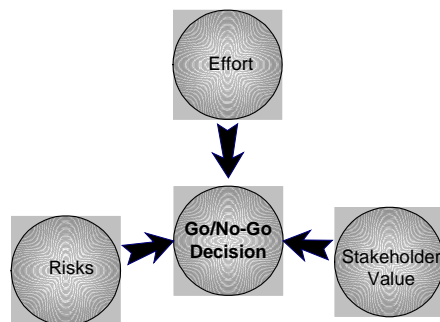


Figure 2: Go/No-Go Assessment for a Data Warehouse Systems

Figure 2 shows the three relevant dimensions for this assessment. Each of the surrounding bubbles in Figure 2 represents one of the factors that determine if the data warehouse system is worthwhile. The effort bubble can be assessed using function points, use case points or some other size measurement, and represents the cost of construction. The stakeholder value bubble is assessed using metrics such as satisfaction and dissatisfaction ratings. The risks bubble is a measure of the severity of risks determined by the risk analysis activity. Based on the composite assessment of all factors, the decision for continuing or stopping the data warehouse project must be made.

5. Risk Management Activities

Risk management is the application of tools and procedures to contain project risks within acceptable limits [10]. The management of risks for data warehouse systems should provide a standard approach to identify and document risk factors, evaluate their potential severity, and propose strategies for mitigating those risks [11]. Risk management includes the following activities (see Figure 3):

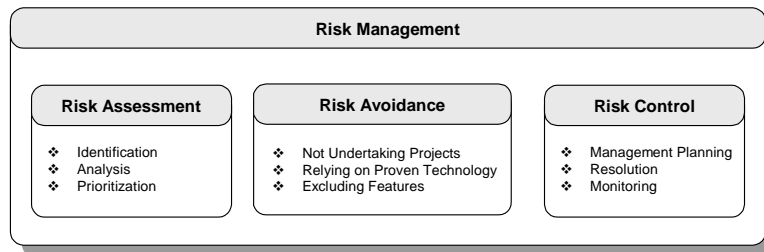


Figure 3: Risk Management Activities

Risk assessment is the process of examining the data warehouse project to identify areas of potential risks. The risk identification can be facilitated with lists of common risk factors of data warehouse projects. During risk analysis, the potential consequences of specific risks to the data warehouse project are examined. Risk prioritization helps to focus on the most severe risks by assessing the potential risk exposure from each.

Risk avoidance is one way to deal with a risk by simple avoiding risky things. Risks can be prevented by not undertaking certain projects, by relying on proven rather than cutting-edge technologies when possible, or by excluding features that will be especially difficult to implement correctly.

Risk control activities are performed to manage the identified top-priority risk factors. Risk management planning produces a plan for dealing with each significant risk, including mitigation approaches, contingency plans, owners, and timelines. Risk resolution involves executing the plans for mitigating each risk. Finally, a tracking process for resolving each risk item through risk monitoring is essential. It is

important to revise the contents and priorities of the risk items periodically to monitor how well the risk mitigation actions are working.

6. Risk Management Process

When a data warehouse team applies risk management, it proactively assesses risks of the data warehouse project continuously, and uses them for decision-making in all phases of the project. They carry the risks forward and deal with them until they are resolved or until they turn into problems and are handled as such. Figure 4 shows the steps of the proactive risk management process of the easyREMOTEDWH framework.

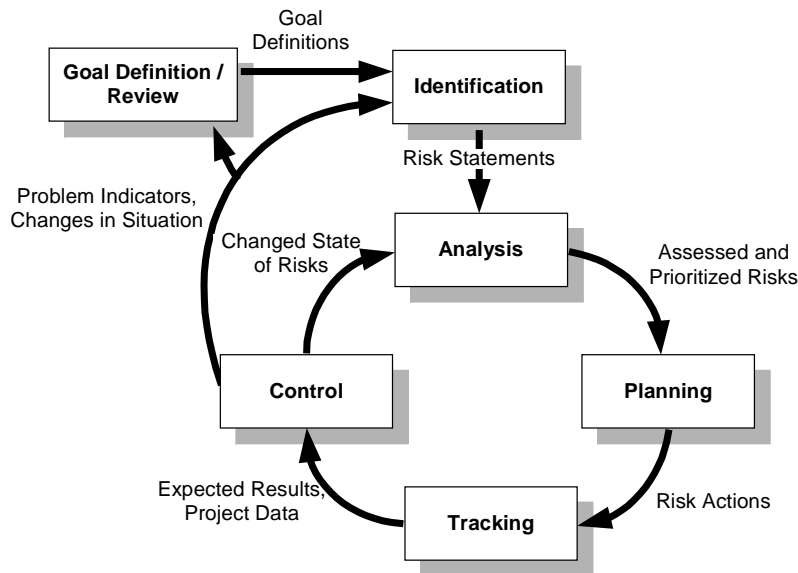


Figure 4: easyREMOTEDWH Risk Management Process

Goal Definition / Review

This step identifies and defines the project goals and stakeholders, including means to keep goal definitions up-to-date. Goals represent the perspectives of the envisioned data warehouse system. When risk scenarios are defined, their impact to the data warehouse project is specified by the stated project goals. This allows full traceability between risks, goals and stakeholders: each risk can be described by its potential impact on the agreed project goals, and each stakeholder can use this information to rank risks from their perspective.

Risk Identification. Risk identification offers the data warehouse team the opportunities, cues, and information that allow them to reveal major risks before they adversely affect the data warehouse project. Depending on experience, different stakeholders and members of the data warehouse team will see the data warehouse project and the related risks differently. Before the risks of a data warehouse system can be managed, they must be expressed clearly. When stating a risk, the data warehouse team must consider not only a symptom, but also a result. Hence, the statement of risk should include what is causing the situation to arise (risk condition) and the expected result (risk consequence). Therefore, risk statements are made up of a condition and at least one consequence of the condition. The condition is something perceived to be true today (it could be a problem, something neutral, or even a good thing) from which undesirable outcomes for the data warehouse system are expected. Section 7 shows how risk statements are implemented in the easyREMOTE^{DWH} prototype.

Risk Analysis

Risk analysis ensures that the data warehouse team is working on the right risks. Risk analysis is the conversion of risk data into risk decision-making information. The risk is composed of two factors: (1) risk probability and (2) risk impact. Both factors are used for determining the exposure of a risk.

Risk Probability. Risk probability is the likelihood that an event will actually occur. Risk probability must be greater than zero, or the risk does not pose a threat to the data warehouse project. Likewise, the probability must be less than 100 percent or the risk is a certainty.

Risk Impact. Risk impact measures the severity of adverse affects, or the magnitude of a loss, if the risk comes to pass. If the risk has a financial impact, a value in currency can be the preferred way to quantify the magnitude of loss. The impact to the schedule of the data warehouse project (in man days) can also measure the loss by a risk. Risks can also be quantified by a level of impact where a subjective scale (i.e. from 1 to 5) rates the viability of project success. High values indicate serious loss to the project. Medium values show loss to portions of the data warehouse project or loss of effectiveness.

Risk Exposure. To evaluate a list of risks, the overall threat of each risk needs to be calculated. Sometimes a high-probability risk has low impact and can be safely neglected; sometimes a high-impact risk has low probability and can safely neglected as well. The risks that have high exposure (= high probability and high impact) are the ones worth managing. Therefore, risk exposure is a function of both the probability of incurring a loss due to the risk and the potential magnitude of that loss.

Risk analysis weighs the threat of each risk to help decide which risks merit taking action. Managing risk takes time and effort away from other parts of the data warehouse project, so it is important for the data warehouse team to do only what is absolutely necessary to manage them. The data warehouse team should not spend much time in quantifying risks too precisely. The goal of risk analysis is to differentiate the most threatening risks from those, which do not need to be tackled immediately.

Risk Planning

Risk planning activities turn risk information into decisions and actions. Planning involves developing actions to address individual risks, prioritizing risk actions, and creating an integrated risk management plan. The goals of risk planning include (1) reduction of the probability that a risk will occur, (2) reduction of the magnitude of loss, or (3) change of the consequences of a risk.

In general, groups new to risk management have trouble dealing with more than 10 risks [11]. Therefore, a data warehouse team should consider only a limited number of major risks that must be managed. After the data warehouse team has ranked the risk exposure, it can focus on a risk management strategy and how to incorporate the risk action plans into the overall project plan.

Risk Contingency Strategy. The idea behind a contingency strategy is to have a fallback plan in place that can be activated in case all efforts to manage a risk fail. For example, a new release of a particular ETL (Extraction, Transformation, Loading) tool is needed so that the data of a particular legacy system can be loaded into the data warehouse system, but the arrival of the tool is at risk. The data warehouse team might devise a plan to use an alternate tool or develop an own solution for the loading of the data. Simultaneous development may be the only contingency plan that ensures that the data warehouse system is implemented in time. Deciding when to start the second parallel effort is a matter of watching the trigger value for the contingency plan. Often the data warehouse team can establish trigger values for the contingency plan based on the type of risk or the type of consequence for the data warehouse project that will be encountered.

Risk Tracking

Risk tracking activities include the monitoring of the risk statuses by the data warehouse team, and ensure that actions are taken to mitigate the risks. Risk tracking is essential to an effective action plan implementation. Therefore, risk metrics and triggering events help ensure that the planned risk actions are working. Tracking is the watchdog function of the risk action plan.

An effective way of tracking risks are risk status reports, which show risk exposure vs. time [11]. Risk status reporting can identify four possible risk management situations:

- A risk is resolved, completing the risk action plan.
- Risk actions are tracking the risk management plan, in which case the risk actions continue as planned.
- Some risk actions are not tracking the risk management plan, in which case corrective measures should be determined and implemented.
- The situation has changed significantly with respect to one or more risks and will usually involve reassessing the risks.

Risk Control

Risk control is the last step of the risk management process. After the data warehouse team has chosen risk metrics and triggering events, risks can be similarly controlled as data warehouse requirements. Therefore, risk management melds into the project management processes to control risk action plans, correct for variations from the plans, respond to triggering events, etc. Controlling a risk includes:

- Altering the mitigation strategy when it becomes ineffective.
- Taking action on a risk that becomes important enough to require mitigation.
- Taking a preplanned contingency action.
- Dropping to a watch-only mode at a specific threshold.
- Closing the risk when it no longer exists.

Williams states in [11] that risk information is only useful if it is accessible and easy to understand. Therefore, it is advantageous to hold the risk information in a database. Having the risk data online and accessible to stakeholders and the data warehouse team lets them adjust risk information or enter new risks as soon as they are identified. Furthermore, this information will be a significant benefit to the organization on future data warehouse projects.

7. Documenting Data Warehouse Risks

The risks of a data warehouse system need to be documented and managed in a way that allows the data warehouse team to communicate risk issues to stakeholders throughout the project's duration. In this section we show how to document data warehouse risks by risk statements and risk action statements facilitated by the easyREMOTEDWH prototype, which is implemented in Lotus Notes.

Risk Statements

During the risk identification and risk analysis activities, risk statements gather essential information about data warehouse risks. As mentioned in the previous section, risk statements have a condition-consequence structure. Table 1 shows the sections of a risk statement used by the easyREMOTEDWH prototype.

Section	Description
<i>Risk ID</i>	A number, which uniquely identifies the risk statement.
<i>Risk Category</i>	This section indicates the focus area (i.e. implementation, deployment, enterprise architecture management), or the risk factor category (i.e. decision drivers, schedule, budget/costs) that is used to identify the risk.
<i>Risk Rational</i>	The risk rationale is the reason behind the occurrence of the risk. It contains additional background information that helps to clarify the risk situation.
<i>Risk Condition</i>	A natural language statement describing an existing condition that could possibly cause a loss for the data warehouse project.
<i>Risk Consequence</i>	A natural language statement describing the loss that would occur to the data warehouse project if the risk became certain.
<i>Resolution Needed By</i>	A date, which indicates when the risk has to be at the latest resolved.

<i>Affected Requirement(s) / Use Cases</i>	This section lists requirements and use cases, which are affected by the risk.
<i>Related Risks</i>	This section lists dependent risks.
<i>Risk Status</i>	The current status of the risk. Possible statuses are: Unresolved, Resolved, Being Investigated.
<i>Risk Probability</i>	A percentage number representing the likelihood that the risk condition will actually occur, resulting in a loss.
<i>Risk Loss</i>	The risk loss is the magnitude of impact, if the risk will actually occur. This number could be the loss of money or time, or simply a number between 2 and 10 that indicates relative magnitude (see Table 2).
<i>Risk Exposure</i>	The risk exposure measures the overall threat of the risk to the data warehouse project by balancing the likelihood of the actual loss with the magnitude of the potential loss. The data warehouse team uses the risk exposure to rate and rank the risks.

Table 1: easyREMOTE^{DWH} Prototype: Risk Statement

Risk Action Statement

Risk action statements determine the necessary steps for mitigating a risk or defining a contingency plan for the risk. Table 2 lists the sections of a risk action statement form. The risk action statement form distinguishes between actions for the risk mitigation (“Planned Actions” section) and actions for a contingency strategy (“Risk Contingency Strategy” section).

Section	Description
<i>Risk Action ID</i>	A number, which uniquely identifies the risk action statement.
<i>Risk ID</i>	The number of the risk statement, which belongs to the risk action statement.
<i>Planned Actions</i>	This section lists planned actions the data warehouse team will take to manage the risk.
<i>Fit Criteria for Risk Action</i>	A fit criterion for risk actions specifies how the data warehouse team will know that the planned actions have been successfully done.
<i>Assigned To</i>	This section lists people assigned to perform the planned action items.
<i>Date To Finish</i>	The date, when each planned action item has to be completed.
<i>Risk Contingency Strategy</i>	This section describes the strategy in the case that the actions planned to manage the risk will not work.
<i>Fit Criteria for Risk Contingency Strategy</i>	The fit criteria for the risk contingency strategy specify how the data warehouse team will know that the risk contingency strategy has been successfully done.
<i>Risk Contingency Strategy Triggers</i>	This section lists the metrics and triggers that the data warehouse system will use to determine when the risk contingency strategy should be put into effect.

Table 2: easyREMOTE^{DWH} Prototype: Risk Action Statement

8. Conclusion

In this paper we stressed the importance of risk management for data warehouse projects. We have shown that data warehouse projects have characteristics that makes risk management different to traditional software projects. Using a systematic approach for risk management facilitates the control of these risks by the data warehouse team. We have shown different categories of risk sources and risk consequences for a data warehouse project. We introduced the easyREMOTE^{DWH} risk management process and discussed its activities in detail. The risk management approach of easyREMOTE^{DWH} is goal-driven and allows stakeholders to manage the risks from different perspectives. We also introduced templates for capturing risk information, which are implemented by the easyREMOTE^{DWH} prototype.

References

1. Boehm, B. W.: Theory-W software project Management: principles and examples, IEEE Trans. Softw. Eng. 15, 1989
2. Boehm, B.W.: Software Risk Management: Principles and Practices, IEEE, 1991
3. Charette, R.N.: Software Engineering Risk Analysis and Management, Intertext, New York, 1989
4. Garvey, P. R.; Phair, D. J., Wilson, J. A.: An Information Architecture for Risk Assessment and Management. IEEE, 1997
5. Jones, C.: Assessment and Control of Software Risks, Prentice-Hall, 1994
6. Keil, M.; Cule, P. E.; Lyytinen, K.; Schmidt, R. C.: A Framework for Identifying Software Project Risks, ACM, Nov. 1998
7. Kontio, J.: The Riskit Method for Software Risk Management, Computer Science Technical Reports, University of Maryland, 1997
8. List, B.; Schiefer, J.; Tjoa, A M.: Use Case Driven Requirements Analysis for Data Warehouse Systems, Data Warehousing 2000, Friedrichshafen
9. Pandelios, G.: Software Risk Evaluation and Team Risk Management, SEPG Conference, Pittsburgh, PA
10. Wiegers, Karl, E.: Software Requirements, Microsoft Press, 1999
11. Williams, R. C.; Walker, J. A.; Dorofee, A. J.: Putting Risk Management into Practice, Software Engineering Institute