Chapter 18

Data Warehousing Requirements Collection and Definition:
Analysis of a Failure

Nenad Jukic
Loyola University Chicago, USA

Miguel Velasco
University of Minnesota, USA

ABSTRACT

Defining data warehouse requirements is widely recognized as one of the most important steps in the larger data warehouse system development process. This paper examines the potential risks and pitfalls within the data warehouse requirement collection and definition process. A real scenario of a large-scale data warehouse implementation is given, and details of this project, which ultimately failed due to inadequate requirement collection and definition process, are described. The presented case underscores and illustrates the impact of the requirement collection and definition process on the data warehouse implementation, while the case is analyzed within the context of the existing approaches, methodologies, and best practices for prevention and avoidance of typical data warehouse requirement errors and oversights.

1. INTRODUCTION

When it comes to the failure rates of IT projects, estimates differ widely, depending on the understanding of what constitutes a failure. There are number of measurements that can be used to assess the success or failure of a large IT project. The success measurements can be divided into two categories (Nelson, 2005). One category encompasses process-based measures: project on schedule, project on budget, and project meeting requirements. The other category includes outcome-based measures: is the project result

actually used, does the project result provide value for the organization, and does the project result enable learning that helps prepare the organization for the future. In addition to the fact that some IT projects are complete failures and some are definite successes, other concepts, such as "failed success" and "successful failure" have been recognized (Nelson, 2005). A failed success is defined as successful from the process perspective, but a failure from the outcome perspective. On the other hand, a successful failure fails on one or more process-based measures, but ultimately delivers solutions that succeed from the outcomes perspective.

Data warehousing has become a standard practice for many companies worldwide (Jukic, 2006). Within the past decade data warehousing projects have been receiving a growing amount of attention and resources in the majority of large and mid-size organizations. A recent study reports typical cost for creating a one terabyte data warehouse of several million USD with a typical implementation time of several years (Gray, 2006). There are no definitive numbers on the failure rate of data warehousing projects, but estimates vary from as little as 20% to as high as 90% (Watson et al., 1999; Frolick & Lindsey, 2003; Watson, 2005; Hwang & Xu, 2007). As is the case with estimating the failure rates of all IT projects, one of the reasons for the wide discrepancy in estimated failure rates of data warehousing projects is the absence of an agreement of what constitutes a failure. For example, there is no unambiguous answer to the question: does abandoning the initial design, scope, strategy, infrastructure or technology of a data warehouse design constitutes a failure? In some cases the answer is a definitive yes. On the other hand, in cases where such abandonment is accompanied by learning lessons that allow for the adoption of successful alternatives which eventually result in a properly designed and used data warehouse, it is appropriate to exempt such cases from a label of

outright failure, given that the outcome is in line with the concept of a successful failure.

Although it is hard to establish a precise overall failure rate of data warehousing projects, the fact remains that some data warehousing projects (just like any other IT projects) fail. The literature indicates that there are many, often intertwined, factors that can cause data warehouse project failure, such as budget overruns, unacceptable performance, poor quality data, weak sponsorship, and lack of long-term planning, etc (Stackowiak, 1997; Adelman & Moss, 2000; Goldman, 2001; Frolick & Lindsey, 2003; Hayen, Rutashobya, & Vetter, 2007).

As do most information system development processes, data warehousing projects follow some form of a System Development Life Cycle (SDLC). SDLC is the overall process of developing information systems through a multi-step process including steps such as planning, analysis, design and implementation (Dennis, Wixom, & Roth, 2006). One popular data warehouse-focused variation of the SDLC is the Data Warehousing Lifecycle (Kimball, Ross, Thornthwaite, Mundy, & Becker, 2007) illustrated in Figure 1. Certain steps (such as product selection, project initiation, etc.) are omitted for brevity. The depicted steps are common to any data warehousing project:

- Data warehouse requirement collection and definition
- Data warehouse modeling and design
- Extraction, transformation and load (ETL - a process that moves the relevant data from underlying operational databases and other applicable data sources into the data warehouse) design and development
- Business intelligence (BI) design
- BI development
- Deployment
- Use/maintenance/growth.

Figure 1 illustrates that the design and development of a data warehouse, as well as the

Related Content

Quality Management in Web Warehouses
Adriana Marotta, Laura González, Lorena Etcheverry, Bruno Rienzi, Raúl Ruggia, Flavia Serraand Elena Martirena (2012). *Business Intelligence Applications and the Web: Models, Systems and Technologies (pp. 1-25).*

www.irma-international.org/chapter/quality-management-web-warehouses/58409

Identification and Categorization of Disruptive Innovations According to the Strategic Scope of the Firm
Vincent Sabourin (2018). *Disruptive Technologies for Business Development and Strategic Advantage (pp. 31-55).*

www.irma-international.org/chapter/identification-and-categorization-of-disruptive-innovations-according-to-the-strategic-scope-of-the-firm/206830

What is Business Intelligence?
Éric Foleyand Manon G. Guillemette (2010). *International Journal of Business Intelligence Research (pp. 1-28).*

www.irma-international.org/article/business-intelligence/47193

Improving Competitive Intelligence Through System Synamics
Özge Pala, Dirk Vriensand Jac A.M. Vennix (2004). *Information and Communications Technology for Competitive Intelligence (pp. 129-157).*

www.irma-international.org/chapter/improving-competitive-intelligence-through-system/22564

Document Retrieval using Efficient Indexing Techniques: A Review
Shweta Gupta, Sunita Yadavand Rajesh Prasad (2016). *International Journal of Business Analytics (pp. 64-82).*

www.irma-international.org/article/document-retrieval-using-efficient-indexing-techniques/165011