

Enhancing Data Warehouse Design with the NFR Framework

Fábio Rilston Silva Paim, Jaelson F. B. Castro

Universidade Federal de Pernambuco, Centro de Informática
Av. Professor Luis Freire s/n, Cidade Universitária,
CEP 50740-540, Recife, PE, Brazil
{frsp,jbc@serpro.gov.br}

Abstract. In recent years, *Data Warehouse* has emerged as a powerful technology for integrating heterogeneous data into a multidimensional repository on behalf of decision-support analysis. The complex extraction, transformation and loading process involved, as well as the aggregational-intensive queries are governed by a multitude of quality factors such as integrity, accessibility, performance, and other domain-specific non-functional requirements (NFRs). This clearly advocates the use of an NFR approach in support of building a high-quality data warehouse specification. In this work we extend the NFR Framework [3] to define catalogues of major data warehouse NFR types and related operational methods, for latter reuse during the specification stage. We illustrate the contributions of our approach in a case study on a large data warehouse project.

1. Introduction

In recent years Data Warehouse (DW) has emerged as a powerful technology for integrating sparsely distributed operational data into a comprehensive analytical fashion on behalf of an old enterprise's dream: to predict and thus make decisions upon their (near) future. The design of such systems is rather different from the design of the conventional operational systems that supply data to the warehouse. The former not only involves information requirements of decision makers, but also the structure and allocated requirements of the latter. Software engineers are required to deal with the complex process of extracting, transforming, and aggregating data while managing to deploy a solution that precisely, timely integrates with a number of heterogeneous source-provider systems; presents analytical results in an accurate, reliable form; offers flexibility at the front-end where ad-hoc queries are to be launched; and do all this with the support of a complete, non-redundant dimensional model. Thus, both operational and strategical visions have to be wrapped up in a multidimensional package to meet corporative analytical requirements that pervade pure decision-support functionality as well as strong

quality constraints like *integrity*, *accessibility*, *performance* and domain-specific non-functional requirements such as *multidimensionality* [2]. This clearly advocates the use of Requirements Engineering techniques to build a precise data warehouse specification.

To pursue this goal, we innovate by proposing a methodological approach for requirements analysis of data warehouse systems in [1]. Our approach provides an interactive, phase-oriented method to guide requirements engineers throughout the data warehouse specification process. The approach, however, is rather general with regard to exploring non-functional requirements and the alternative paths developers would have to probe into in order to understand both positive and negative influences of a certain quality requirement to the data warehouse design process.

The NFR Framework (Chung et al. [3]) fills in this gap by enabling developers to produce tailored solutions that embrace the quality characteristics of a particular domain, including priorities, related operational methods and reasoning about the influence of a non-functional choice to the system design. To deal with the large number of possible development alternatives, developers can consult the Framework's design catalogues, which organize past experience, standard techniques, knowledge about particular non-functional requirements as well as their tradeoffs and interdependences.

In this work, we adopt the NFR Framework to complement the requirements specification phase of our methodology. We start by defining a hierarchical tree of the major data warehouse non-functional requirements. Most significant non-functional requirements are further decomposed into catalogues of operational methods that, in conjunction with the Framework structure, enable engineers to select from the combination of qualitative and implementation factors that best meet users' decision-support needs.

This work is organized as follows. In section 2 we briefly present our approach for requirements analysis of data warehouse systems. In section 3 we describe the NFR Framework. The main set of data warehouse non-functional requirements, and their related Operationalization Catalogues, are discussed in Sections 4 and 5 respectively. Section 6 illustrates the contributions of the extended NFR Framework to the design of a large-scale governmental data warehouse. Section 7 summarizes our conclusions.

2. Analysing Data Warehouse Requirements

Data Warehouse systems offer efficient access to integrated and historical data from heterogeneous information sources to help managers in planning and decision-making. The data within the warehouse is extracted from the sources, consolidated, aggregated and accumulated in multidimensional data structures to support strategic analysis, powered by a technology named *OnLine Analytical Processing* (OLAP) [4]. To achieve good OLAP performance, the multidimensional model classifies data into *facts*, numeric data that quantifies a specific business activity that one wishes to analyze (e.g. *quantity of products sold*); and *dimensions*, a hierarchical classification chain of qualitative values through

