



Chapter X

Metrics for Data Warehouse Quality

Manuel Serrano
University of Castilla-La Mancha, Spain

Coral Calero
University of Castilla-La Mancha, Spain

Mario Piattini
University of Castilla-La Mancha, Spain

ABSTRACT

This chapter proposes a set of metrics to assess data warehouse quality. A set of data warehouse metrics is presented, and the formal and empirical validations that have been done with them. As we consider that information is the main organizational asset, one of our primary duties should be assuring its quality. Although some interesting guidelines have been proposed for designing “good” data models for data warehouses, more objective indicators are needed. Metrics are a useful objective mechanism for improving the quality of software products and also for determining the best ways to help professionals and researchers. In this way, our goal is to elaborate a set of metrics for measuring data warehouse quality which can help designers in choosing the best option among more than one alternative design.

INTRODUCTION

It is known that organizations are very rich in data but poor in information. Today technology has made it possible for organizations to store vast amounts of data obtained at a relatively low cost; however, these data fail to provide information (Gardner, 1998).

Data warehouses have appeared as a solution to this problem, supporting decision-making processes and new kinds of applications as marketing.

A data warehouse is defined as a “collection of subject-oriented, integrated, non-volatile data that supports the management decision process” (Inmon, 1997). Data warehouses have become the key trend in corporate computing in the last few years, since they provide managers with the most accurate and relevant information to improve strategic decisions. Also the future for data warehousing is promising. Jarke et al. (2000) forecast a market of 12 millions American dollars for the data warehouse market for the next few years. However, the development of a data warehouse is a difficult and very risky task. It is essential that we can assure the information quality of the data warehouse as it became the main tool for strategic decisions (English, 1999).

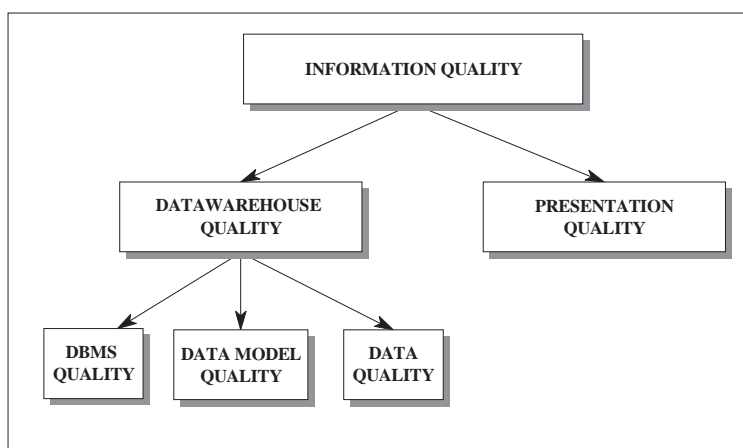
Information quality of a data warehouse comprises data warehouse system quality and presentation quality (see Figure 1). In fact, it is very important that data in a data warehouse reflect correctly the real world, but it is also very important that data can be easily understood. In data warehouse system quality, as in an operational database (Piattini et al., 2000), three different aspects could be considered: DBMSs quality, data model quality and data quality.

In order to assess DBMS quality, we can use an international standard like ISO 9126 (ISO, 1999), or some of the existing product comparative studies. This type of quality should be addressed in the product selection stage of the data warehouse life cycle.

Data quality must address mostly the extraction, filtering, cleaning and cleansing, synchronization, aggregation, loading, etc., activities of the life cycle. In the last few years, very interesting techniques have been proposed to assure data quality (Bouzeghoub & Kedad, 2002).

Last, but not least, data warehouse model quality has a great influence in the overall information quality. The designer has to choose the tables, processes, indexes and data partitions, representing the logical data warehouse and facilitating its functionality (Jarke et al., 2000).

Figure 1. Information and Data Warehouse Quality



16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/metrics-data-warehouse-quality/9210

Related Content

The Expert's Opinion

Mohammad Dadashzadeh (1991). *Journal of Database Administration* (pp. 30-34). www.irma-international.org/article/expert-opinion/51085

Semantic Resolution in Multi-Database Environments

Elizabeth R. Towell and William D. Haseman (1995). *Journal of Database Management* (pp. 23-30). www.irma-international.org/article/semantic-resolution-multi-database-environments/51152

Detecting Expressional Anomie in Social Media via Fine-grained Content Mining

Qingqing Zhou and Ming Jing (2020). *Journal of Database Management* (pp. 1-19). www.irma-international.org/article/detecting-expressional-anomie-in-social-media-via-fine-grained-content-mining/245297

Towards an Ontology for Information Systems Development—A Contextual Approach

Mauri Leppänen (2007). *Contemporary Issues in Database Design and Information Systems Development* (pp. 1-36). www.irma-international.org/chapter/towards-ontology-information-systems-development/7019

Ensuring Correctness, Completeness, and Freshness for Outsourced Tree-Indexed Data

Tran Khanh Dang (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 2204-2222). www.irma-international.org/chapter/ensuring-correctness-completeness-freshness-outsourced/8031