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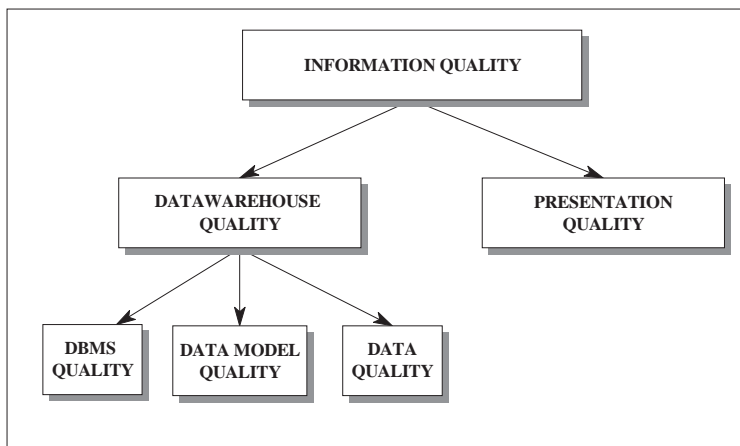
16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/metrics-data-warehouse-quality/9210

Related Content

Semantic Resolution in Multi-Database Environments
Elizabeth R. Towelland William D. Haseman (1995). *Journal of Database Management (pp. 23-30).*
www.irma-international.org/article/semantic-resolution-multi-database-environments/51152

SinGAN-Based Asteroid Surface Image Generation
Yundong Guo, Jeng-Shyang Pan, Chengbo Qiu, Fang Xie, Hao Luo, Huiqiang Shang, Zhenyu Liuand Jianrong Tan (2021). *Journal of Database Management (pp. 28-47).*
www.irma-international.org/article/singan-based-asteroid-surface-image-generation/289792

A Conceptual Design Methodology for Fuzzy Relational Databases
Z.M. Ma (2005). *Journal of Database Management (pp. 66-83).*
www.irma-international.org/article/conceptual-design-methodology-fuzzy-relational/3332

Interesting Knowledge Patterns in Databases
Rajesh Natarajanand B. Shekar (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1652-1662).*
www.irma-international.org/chapter/interesting-knowledge-patterns-databases/7997

A Conceptual Design Methodology for Fuzzy Relational Databases
Z.M. Ma (2005). *Journal of Database Management (pp. 66-83).*
www.irma-international.org/article/conceptual-design-methodology-fuzzy-relational/3332