Chapter 7.11
Information Quality: How Good are Off-the-Shelf DBMS?

Felix Naumann
Humboldt-Universität zu Berlin, Germany

Mary Roth
IBM Silicon Valley Lab, USA

ABSTRACT

Commercial database management systems (DBMS) have come a long way with respect to efficiency and more recently, with respect to quality and user friendliness. Not only do they provide an efficient means to store large amounts of data and intuitive query languages to access the data, popular DBMS also provide a whole suite of tools to assess, store, manage, clean, and retrieve data in a user-friendly way. Some of these feature address database experts, others are targeted at end-users with little or even no database knowledge. The recent developments in the field of autonomic computing drive the ease-of-use even further. In this chapter we study how well a typical DBMS meets the goal of providing a high-quality data storage and retrieval facility. To this end, we draw on an established set of information quality criteria and assess how well an exemplary DBMS fares. While quality criteria are usually defined for a set of data, we extend, wherever possible, the definitions to the systems that manage this data.

THE QUALITY-IN-QUALITY-OUT PRINCIPLE

Arguably the most widespread architecture to store, manage, and retrieve structured data is the relational database management system (DBMS) architecture. Starting with System R (Astrahan, 1979) of IBM, which evolved to the IBM DB2 database system, today there are many commercial systems storing petabytes of data. Other prominent examples are Oracle database,[1] Microsoft's SQL Server,[2] and MySQL.[3] Other data models, such as the object-oriented model or the hierarchical model are also widespread but not discussed here. The information quality provided by a database is not due to the data model itself, but to the system carefully managing the data. Thus, database systems with other data models enjoy the same information quality properties. Research and development for DBMS follows two main directions: scalability and usability. With the growing demand to store more and more data, databases systems have scaled in the hardware they use and in the software managing the data. Additionally, administrators of databases and end-users of the data demand more and more functionality that either adds value to the DBMS or makes its use easier. In this chapter, we analyze how well modern DBMS are able to meet user demands, or at least help database administrators (DBAs) meet user demands regarding their everyday work with the DBMS or applications built on top. Here, user demands are expressed as a set of information quality criteria taken from the empirical study of Wang and Strong (1996).

Information quality is a measure to assess the value of data to perform the task at hand (Wang, 1996). Other definitions mention fitness for use (Tayi, 1998) or user satisfaction (Delone, 1992). As DBMS are one of the most common means to generate, manage, and provide this data, it is worthwhile to examine how they influence the quality of the information they handle. This influence is both explicit within the core functionality of a DBMS and implicit through tools that help data providers, developers, managers, and consumers derive the most value from the data.

To examine DBMS with respect to the quality of information they are able to supply, we apply a large set of IQ criteria to DBMS as an entire system. Usually, IQ criteria are used to assess the quality of information, and not the quality of a system. Addressing this mismatch, we analyze not the DBMS itself, but its ability to provide high quality data. DBMS are not the sole source of high information quality, but they are designed to at least not diminish quality. While the well-known *garbage-in-garbage-out* principle holds for any system dealing with data, we postulate the *quality-in-quality-out* principle for modern, well-designed DBMS. For instance, if data is generated and inserted into a DBMS in a timely manner, a good DBMS will not unduly delay the accessibility of the data to users. Another example is the completeness of information: DBMS are developed to always return complete (and correct) answers to queries. Only if the stored base data is incomplete or incorrect will a DBMS answer with an inferior result. In this spirit we analyze several quality dimensions and provide details on if and how a typical DBMS meets IQ demands. In this chapter we ignore the issue of software quality and assume a DBMS that correctly implements the SQL standard and its added functionality.

Structure of this Chapter

In the following section we introduce IBM's DB2 database system as a typical representative of commercial DBMS. Additionally, we present three different roles that DBMS users acquire, each with different needs towards information quality and each with different demands on a DBMS. In third section we enumerate a comprehensive set of information quality criteria and discuss for each, if and how a DBMS used in different roles affects them. We conclude with a discussion of future work in the fourth section and a summary of our findings in the last section.

Database Management Systems and Their Users

A software product, such as a database management system (DBMS), is used by many different persons with different educational backgrounds, different IT needs, and, most importantly, with different roles inside an organization. These roles are the basis of our assessment of DBMS quality.

Related Content