IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB7495

Chapter 11

Methodology of Schema Integration for New Database Applications: A Practitioner's Approach

Joseph Fong
City University of Hong Kong

Kamalakar Karlapalem
Hong Kong University of Science & Technology

Qing Li and Irene Kwan
Hong Kong Polytechnic University

A practitioner's approach to integrate databases and evolve them so as to support new database applications is presented. The approach consists of a joint bottom-up and top-down methodology; the bottom-up approach is taken to integrate existing database using standard schema integration techniques (B-Schema), the top-down approach is used to develop a database schema for the new applications (T-Schema). The T-Schema uses a joint functional-data analysis. The B-schema is evolved by comparing it with the generated T-schema. This facilitates an evolutionary approach to integrate existing databases to support new applications as and when needed. The mutual completeness check of the T-Schema against B-Schema derive the schema modification steps to be performed on B-Schema to meet the requirements of

Previously Published in the *Journal of Database Management, vol.10, no.1*, Copyright © 1999, Idea Group Publishing.
This chapter appears in the book, Human Computer Interaction Development and Management by Tonya Barrier.

the new database applications. A case study is presented to illustrate the methodology.

There has been a proliferation of databases in most organizations. These databases are created and managed by the various units of the organization for their own localized applications. Thus the global view of all the data that is being stored and managed by the organization is missing. Schema integration is a technique to present such a global view of an organization's databases. There has been a lot of work done on schema integration. Batini et al. (1986) and Özsu amd Valduriez (1991) present surveys of work in this area. But all these techniques concentrate on integrating database schemas without taking into consideration of new database applications. This paper presents a practical approach to schema integration to support new database applications by comparing the existing databases against data requirements of the new applications. If the existing databases are inadequate to support new applications, then they are evolved to support them.

In any schema integration methodology all the database schemas have to be specified using the same data model. The proposed approach uses an extended entity relationship(EER) data model. Therefore, the first step in the schema integration methodology is to translate a non-EER database schema to an EER database schema. A joint bottom-up and top-down approach for schema integration to support new database applications is proposed. The bottom-up approach is taken to integrate existing databases using standard schema integration techniques whereas the top-down approach is used to come up with the database schema for the new applications. The top-down approach uses the joint functional-data analysis. The B-schema generated by bottom-up approach is evolved by comparing it with the T-schema generated by the top-down approach. This facilitates a stream lined approach to evolve integrated databases to support new applications.

Conventional approaches that have been widely used in database community for database design can be classified as top-down, and are therefore suitable for designing databases from scratch to support new applications. On the other hand, research in the area of heterogeneous distributed databases over the last decade has emphasized on bottom-up approaches towards global schema derivation by integrating existing databases. These two kinds of approaches are complementary in many aspects, and thus can be combined into a unified framework for schema integration.

Fong et al. (1994) developed a hierarchical comparison scheme using three major criteria for comparing relationships in two schemas. The paper classified the relationship integration by taking into account the degree of

Related Content