Information Systems Reengineering and Integration

Second Edition

Joseph Fong

Information Systems Reengineering and Integration

Second Edition

Springer

Joseph F.P. Fong
Department of Computer Science
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong
China

British Library Cataloguing in Publication Data
A catalouge record for this book is available from the British Library

Library of Congress Control Number: 2006923904

ISBN-10: 1-84628-382-5 2nd edition Printed on acid-free paper
ISBN-13: 978-1-84628-382-6 2nd edition
ISBN 981-3083-15-8 1st edition

First published 1997
Second edition 2006

Printed in the United States of America (SBA)

9 8 7 6 5 4 3 2 1

Springer Science+Business Media, LLC
springer.com

To the memory of my parents Fong Chung Lung and Sun Oil Yuk

Preface

Over the part three decades, there has been a tremendous investment made in information systems. Such systems have evolved from file systems, through database systems, and we are now seeing the emergence of management information systems (MIS) and executive information systems (EIS). With the advent of each new form of technology, there has been a need to redesign and re-implement existing information systems.

In recent years, a great deal of resources have been put into the area of reengineering. Reengineering involves the redesign of existing information systems, while using as much of the existing systems as possible. That is, the approach taken is to transform the existing information systems into the format needed for the new technology, rather than to throw away the old systems. Such an approach has obvious benefits, particularly if it can be automated and/or supported by methods and tools.

Very often, a large company has multiple heterogeneous databases for MIS operations. The company needs to integrate them into a corporate database for its decision support systems. Subsequently, schema integration must be performed to resolve the conflicts between two databases with respect to data name, data type, and data semantics. Schema integration must be done before data integration, which is mainly concerned with the automation of loading data from source databases into an integrated database. Furthermore, in reality, user demands are changing daily. It is essential for companies to enhance

and evolve the existing database schemas to meet the new data requirements.

This text will focus upon practical approaches to information systems reengineering and integration, including:

- The conversion and integration of hierarchical or network database systems into relational database technology, or from a relational to an object-oriented database and XML database.

- The integration of multiple databases, and also between a database system and an expert system to produce MIS (management information systems) and EIS (executive information systems).

The text will summarize the concepts, the approach to be taken and the benefits to be gained in these two crucial technological areas. It will focus upon proven methods and tools for:

- Converting hierarchical and network databases to relational technology, or from relational to object-oriented databases, or from relational to XML databases.

- Reengineering existing systems to produce MIS and EIS.

The book will describe in detail:

- Database conversion techniques

- Reverse engineering and forward engineering for data modeling

- A reengineering methodology for information systems

- Techniques of schema and data intergration

From a professional point of view, this book proposes a general solution for the problem of system migration to new database technology. It offers a systematic software engineering approach for reusing existing database systems built with "old" technology by converting them into the new database technology. As a result, investment in the existing information systems can be protected by upgrading database systems and expert systems, rather than phasing them out.

This book focuses on methodologies for information systems reengineering and integration. It applies many examples, illustrations, and case studies of procedures for reusing existing database systems and information systems. The objective is to make the methodologies very practical for readers to follow. Even though there are many technical terminologies used in the book, the techniques proposed are simple enough for students or computer professionals to follow. The content of the book is divided into nine chapters.

Chapter 1 gives an overview of information systems, and deals with its past history, its evolution to management information systems, its problems encountered in file systems, its solution found in database systems and expert systems, and the need for the reengineering of existing database systems and information systems. It also describes database conversion, the merge of multiple databases, and the integration of the expert systems and the database systems into an expert database system. It show how to apply data transformation for electronic data interchange on the Internet.

Chapter 2 describes basic theories and data structures of various data models, including hierarchical, network, relational, object-oriented, and XML. Their pros and cons are discussed. Expert systems technology is explained. The advanced expert database systems are introduced. The basic concepts discussed include data definition language, data manipulation language, forward chaining, backward chaining, procedural language and non-procedural language, data type definition, and XML schema definition.

Chapter 3 covers various techniques in schema translation from nonrelational (e.g., hierarchical or network) to relational, and from relational to object-oriented and XML databases. Reverse engineering is adopted to recover original schema's semantics into the conceptual model of the Extended Entity Relationship (EER) model. Forward engineering is used to map the EER model into relational or Unified Model Language (UML), a conceptual model for an object-oriented database.

Chapter 4 shows a methodology of converting data from nonrelatonal database to relational database, and from relational database to object-oriented database, and also from relational database into XML database. Unload and upload processing in a logical level approach is adopted to do the task.

Chapter 5 explains a methodology of emulating SQL by using a hierarchical or network database data manipulation language. The methodology can be used in program translation from relational

database programs to nonrelational database programs. The objective is to provide a relational interface to the nonrelational database so that the users can use SQL to access a hierarchical or network database. It also presents a methodology of translating SQL query into OSQL (Object SQL or Object Query Language) and XQL (XML Query Language).

Chapter 6 summarizes the database conversion methodology for converting hierarchical or network databases to relational databases. The methodology is in three phases: schema translation, transaction translation, and data conversion. The first and second phases provide a relational interface to a nonrelational database as a temporary solution in the database conversion (migration) process. The third phase provides a permanent solution to convert data from nonrelational database to relational database after nonrelational database programs are phased out or rewritten. A case study of constructing an XML view of a relational database involving schema and data transformation from relational into XML is presented.

Chapter 7 offers a state-of-the-art methodology for integrating two relational database schemas by resolving their name, data type, and data semantics conflicts with user supervision. The relational or object-relational data integration can only be done after relational or object-relational schemas integration for the loading of data into the integrated databases is performed. A Frame model metadata is introduced to store data operation for encapsulation in the object-oriented database.

Chapter 8 lays out the rules in integrating expert systems and database systems for the purpose of reengineering. The technique is to transform both expert systems rules and database systems relations into a common Frame model metadata. This Frame model metadata offers object-oriented-like database functions by treating each frame as an object and a collection of objects as a class. Coupling classes, active classes, static classes, and integrated classes are introduced to implement an expert database system (EDS). The users can then apply EDS to develop new applications.

Chapter 9 summaries the methodologies proposed by the book. The main theme is that knowledge engineering is a requirement for information systems reengineering and integration. We need users' knowledge to assist system developers in reusing existing database systems and expert systems in order to develop new applications. The final result is database systems upgrade, multiple database intergration and expert systems enhancement to knowledge-based systems. As knowledge engineering becomes important in data processing, the

resultant knowledge-based system, i.e., the expert database system, will become a very important asset to companies.

Acknowledgments

This book is a tribute to the University of Sunderland in United Kingdom since the author developed most of the methodologies there.

The author thanks Professor Shi-Ming Huang as the original inventor of Frame model metadata used in the book, and also to Dr. Chuk Yau for his joint articles with the author, which contribute the MIS overview in the book. The author appreciates the assistance of the word processing work from Shing-han Li in formatting and drawing the diagrams for the book. Special thanks go to Dr Reggie Kwan and Frances Fong for their review and proofreading of the book. The book is the result of the combined research of the author and his project students over more than a decade.

Joseph Fong
City University of Hong Kong
Hong Kong

Contents