Preface

Databases are essential ingredients of modern computing systems. Although database concepts, technology and architectures have been developed and consolidated in the last decades, many aspects are subject to technological evolution and revolution. Thus, writing a textbook on this classical and yet continuously evolving field is a great challenge.

Key features

This book provides a new and comprehensive treatment of databases, dealing with the complete syllabuses for both an introductory course and an advanced course on databases. It offers a balanced view of concepts, languages and architectures, with concrete reference to current technology and to commercial database management systems (DBMSs). It originates from the authors' long experience in teaching, both in academia and in industrial and application settings.

The book is composed of four main parts and a fifth part containing three appendices and a bibliography. Parts I and II are designed to expose students to the principles of data management and for teaching them how to master two main skills: how to query a database (and write software that involves database access) and how to design its schema structure. These are the fundamental aspects of designing and manipulating a database that are required in order to make effective use of database technology.

Parts III and IV are dedicated to advanced concepts, required for mastering database technology. Part III describes database management system architectures, using a modern approach based upon the identification of the important concepts, algorithms, components and standards. Part IV is devoted to the current trends, focusing on object-oriented databases, active databases, data warehouses and the interaction between databases and the World Wide Web.

Appendices cover three popular database systems: Microsoft Access, IBM DB2 and Oracle.

A number of features make this book unique.

- We make a clear distinction between the basic material (Parts I and II) and the advanced material (Parts III and IV), which often marks the progression from an introductory course to an advanced course. A coherent treatment makes the transition from basic to advanced material quite smooth and progressive and makes this book perfectly suitable to serving the needs of a single, comprehensive course.

- We provide the foundations of the relational model and theory, but we never introduce theory for its own sake. Each concept is immediately applied to SQL, with plenty of examples and exercises.

- The discussion of design techniques starts with the introduction of the elements of the E-R (Entity-Relationship) model and proceeds through a well-defined, staged process through conceptual design to the logical design, which produces a relational schema. Several detailed examples are developed, to help the reader to follow the process.

- We deal with the fundamental issues of database technology (transactions, concurrency control, reliability, recovery) first on a single-server architecture and then on a multi-server architecture with emphasis on distribution, replication and parallelism.

- We emphasize standards (including SQL-2, SQL-3, ODL, OQL, CORBA, ODBC, JDBC, X-OPEN) in order to give to the student an understanding of the way in which the concepts developed in this book are deployed in the commercial world.

- We include appendices on Access, DB2 and Oracle, particularly helpful for hands-on courses. We focus on the ease of use of Access, on the interactive tools of DB2 and the richness of Oracle's PL-SQL as a database programming language.

- We provide a Web site (http://www.mcgraw-hill.co.uk/atzeni/) with slides and exercises with solutions that can be used by lecturers and students.

Detailed Content Overview

Chapter 1 covers the use of database technology in modern information systems. We cover basic aspects, such as the difference between data and information, the concepts of data model, schema and instance, a multi-level organization of the database architecture with the fundamental notion of data independence and the classification of database languages and users.

Part I – Relational Databases

This part introduces the concepts of relational databases and then focuses on SQL programming, one of the main objectives of this textbook.

Chapter 2 describes the relational model, by introducing the basic notions of domain, attribute, relation schema and database schema, with the various integrity constraints: primary key and referential constraints; null values are also briefly discussed.

Chapter 3 illustrates the foundations of the languages for the relational model. First we describe relational algebra, a simple and important procedural language; then we introduce declarative languages like relational calculus (on domains and on tuples with range restrictions) and Datalog.

Chapter 4 provides a thorough description of SQL, by focusing on both the Data Definition Language, used to create the schema of a database and the Data Manipulation Language, which allows for querying and updating the content of the database. The chapter also includes advanced features of SQL, such as programming language interfaces and dynamic SQL.

Part II – Database Design

This part covers the conceptual and logical design of relational databases. The process starts with the analysis of user requirements and ends with the production of a relational database schema that satisfies several correctness criteria. We believe that a student must initially learn about database use before he or she can concentrate on database design with sufficient confidence and therefore we postpone design until after the mastering of a query language.

Chapter 5 introduces the design methodology and describes the E-R conceptual model, with the classical notions of entity, relationship, attribute, identifier and generalization. Business rules are also introduced, as a formalism for the description of additional user requirements.

Chapter 6 illustrates conceptual design, which produces an E-R conceptual description of the database, starting from the representation of user requirements. Simple design principles are illustrated, including methods for the systematic analysis of informal requirements, the identification of the main concepts (entities and relationships), top-down refinement strategies, suggestions for achieving certain qualities of the schemas and schema documentation.

Chapter 7 focuses on logical design, which produces a relational database schema starting from the conceptual schema. We discuss the various design options and provide guidelines that the designer should follow in this phase.

Chapter 8 discusses schema normalization and the correctness criteria that must be satisfied by relational schemas in order to avoid anomalies and

redundancies. Normalization is used for verification: although it is an important design technique, we do not believe that a designer can really use normalization as the main method for modelling reality. He or she must, however, be aware of normalization issues. Also, the development is precise but not overly formal: there are no abstract algorithms, but we cover instead specific cases that arise in practice.

Part III – Database Technology

This part describes the modern architectures of database management systems.

Chapter 9 is focused on the technology required for operating a single DBMS server; it discusses transactions, concurrency control, buffer management, reliability, access structures, query optimization and physical database design. This chapter provides a database administrator with the fundamental knowledge required to monitor a DBMS.

Chapter 10 addresses the nature of architectures that use a variable number of database servers dispersed in a distributed or parallel environment. Again, transactions, concurrency control and reliability requirements due to data distribution are discussed; these notions are applied to several architectures for data management, including client-server, distributed, parallel and replicated environments.

Part IV – Database Evolution

This part discusses several important extensions to database technology.

Chapter 11 describes object database systems, which constitute a new generation of database systems. We consider both the 'object-oriented' and the 'object-relational' approaches, which are the two alternative paths towards object orientation in the evolution of database systems. We also consider multimedia databases and geographic information systems. The chapter also describes several standards, such as ODM, OQL and CORBA.

Chapter 12 describes active database systems; it shows active rules as they are supported in representative relational systems (Oracle and DB2) and discusses how active rules can be generated for integrity maintenance and tested for termination.

Chapter 13 focuses on data analysis, an important new dimension in data management. We describe the architecture of the data warehouse, the star and snowflake schemas used for data representation within data warehouses and the new operators for data analysis (including drill-down, roll-up and data cube). We also briefly discuss the most relevant problems of data mining, a novel approach for extracting hidden information from a data warehouse.

Chapter 14 focuses on the relationship between databases and the World Wide Web, which has already had a deep influence on the way information systems and databases are designed and accessed. It discusses the notion of Web information systems, the methods for designing them and the tools and techniques for coupling databases and Web sites.

Appendices

Three appendices conclude the volume, with descriptions of three popular DBMSs:

Appendix A deals with Microsoft Access, which is currently the most widespread database management system on PC-based platforms. Access has a simple yet very powerful interface, not only for programming in SQL and QBE, but also for adding forms, reports and macros in order to develop simple applications.

Appendix B describes the DB2 Universal Database, the latest member of one of the major families of DBMSs produced by IBM. The emphasis of the presentation is on its interactive tools and its advanced functionality.

Appendix C covers Oracle, a major product in the market of DBMSs and specifically focuses on the procedural extension of SQL (called PL/SQL), which is supported by Oracle for writing applications.

Use as a textbook

The book can be covered in a total of approximately 50–70 lecture hours (plus 30–40 hours dedicated to exercises and practical experiences).

Our experience is that Parts I and II can be covered as a complete course in about 30 taught hours. Such a course requires a significant amount of additional practical activity, normally consisting of several exercises from each chapter and a project involving the design, population and use of a small database. The appendixes provide useful support for the practical activities.

Parts III and IV can be covered in a second course, or else they can be integrated in part within an extended first course; in advanced, project-centred courses, the study of current technology can be accompanied by a project dedicated to the development of technological components. Part IV, on current trends, provides material for significant project work, for example, related to object technology, or to data analysis, or to Web technology. The advanced course can be associated with further readings or with a research-oriented seminar series.

An international textbook

Making the book reflect the international nature of the subject has been a challenge and an opportunity. This book has Italian authors, who have also

given regular courses in the United States, Canada and Switzerland, was edited in the United Kingdom and is directed to the worldwide market. We have purposely used a neutral version of the English language, thus avoiding country-specific jargon whenever possible. In the examples, our attitude has been to choose attribute names and values that would be immediately understandable to everybody. In a few cases, however, we have purposely followed the rules of different international contexts, without selecting one in particular. The use of car registration numbers from France, or of tax codes from Italy, will make the reader aware of the fact that data can have different syntax and semantics in different contexts and so some comprehension and adaptation effort may be needed when dealing with data in a truly worldwide approach. It should also be noted that when dealing with money values, we have omitted the reference to a specific currency: for example, we say that a salary is '40 thousand', without saying whether it is dollars (and which dollars: US, Canadian, Australian, Hong Kong, …), or Euros, or Pounds Sterling.

Additional support

Supplementary material, including overhead slides, solutions to exercises and additional exercises can be found at the book Web site:

http://www.mcgraw-hill.co.uk/atzeni/

The authors can be contacted through the site.

The Authors

Paolo Atzeni and Riccardo Torlone are professors at Università di Roma Tre. Stefano Ceri and Stefano Paraboschi are professors at Politecnico di Milano. They all teach courses on information systems and database technology and are active members of the research community. Paolo Atzeni and Stefano Ceri have many years of experience in teaching database courses, both in European and in North American universities. They have also presented many courses for professional database programmers and designers. All the authors are active researchers, operating on a wide range of areas, including distributed databases, deductive databases, active databases, databases and the Web, data warehouses, database design and so on. They are actively participating in the organization of the main International Conferences and professional Societies dealing with database technology; in particular, Paolo Atzeni is the chairman of the EDBT Foundation and Stefano Ceri is a member of the EDBT Foundation, VLDB Endowment and ACM Sigmod Advisory Committee. Their appointments include being co-chairs of VLDB 2001 in Rome.

Acknowledgements

The organization and the contents of this book have benefited from our experiences in teaching the subject in various contexts. All the students attending those courses, dispersed over many schools and countries (University of Toronto, Stanford University, Università dell'Aquila, Università di Roma 'La Sapienza', Università di Roma Tre, Politecnico di Milano, Università di Modena, Università della Svizzera Italiana) deserve our deepest thanks. Many of these students have field-tested rough drafts and incomplete notes, and have contributed to their development, improvement and correction. Similarly, we would like to thank people from companies and government agencies who attended our courses for professionals and helped us in learning the practical aspects that we have tried to convey in our textbook.

We would like to thank all the colleagues who have contributed, directly or indirectly, to the development of this book, through discussions on course organization or the actual revision of drafts and notes. They include Carlo Batini, Maristella Agosti, Giorgio Ausiello, Elena Baralis, Giovanni Barone, Giampio Bracchi, Luca Cabibbo, Ed Chan, Giuseppe Di Battista, Angelo Foglietta, Piero Fraternali, Maurizio Lenzerini, Gianni Mecca, Alberto Mendelzon, Paolo Merialdo, Barbara Pernici, Silvio Salza, Fabio Schreiber, Giuseppe Sindoni, Elena Tabet, Letizia Tanca, Ernest Teniente, Carson Woo and probably some others whom we might have omitted. We thank the reviewers of the English edition for a number of very useful suggestions concerning the organization of the book and the specific content of chapters.

We thank the very many people who have contributed to the birth of this book inside McGraw-Hill. We are grateful to Gigi Mariani and Alberto Kratter Thaler who have worked with us to the Italian edition of this work. We are deeply indebted to David Hatter, who endorsed our project enthusiastically and was able to put together an effective team, together with Ellen Johnson and Mike Cotterell. These three people have dedicated an enormous amount of effort to the production process. In particular, we thank Ellen for her support in the translation, David for his careful copy-editing and for the many terminological suggestions and Mike for his professional and patient processing of our manuscript through its numerous revisions.

We would also like to thank our families, for the continuous support they have given to us and for their patience during the evenings, nights and holidays spent on this book. Specifically, Paolo Atzeni would like to thank Gianna, his wife and his children Francesco and Laura; Stefano Ceri wishes to thank Maria Teresa, his wife and his children Paolo and Gabriele; Stefano Paraboschi wishes to thank Paola, his wife; Riccardo Torlone wishes to thank Maria Paola.