

Preface

Database research has about 35 years of rich history of productivity which has lead to the most relevant and important developments into the software engineering discipline. Naturally, database technologies, architectures and conceptual frameworks have been consolidated in the last decades. Moreover, over the last decade, database management has evolved in such a way that databases have become a key component at the heart of current computing environments and modern information systems. This has provoked a deep impact and significant changes in the way organizations and institutions operate and make their business decisions. It is worthwhile mentioning some facts that have promoted such a growth: the ubiquity of powerful and easy-to-use personal computer database management products, new modeling techniques and tools, most importantly, those based on object oriented thinking, the emergence of client-server processing, the decreasing price of hardware and software, and the imperious need to properly and efficiently manage huge amounts of information.

The history of the database evolution fits into the following paradigm: a problem with data management arises, database technology then copes with the problem in an efficient way and leads users to figure out new problems which in turn feed the evolution of the technology. Perhaps this can be seen as a never-ending cycle of management information needs, new technological advances and successful database software products, which in turn open the gate to new management information needs, and so on. This chain often occurs with any efficient person or program. Their success leads the world to give them more and more tasks!

The design of database applications is a crucial factor in the success of information systems in any organization. Data is one of the most valuable assets to an organization whose validity, consistency and accuracy are vital. A database management systems (DBMS) greatly contributes with these purposes by providing data persistence, efficient access to data, and data integrity guaranteed by integrity constraints. Current database systems offer the SQL query language standard, which has become the query and manipulation language par excellence. By isolating the conceptual schema from the implementation schema, database systems guarantee data independence from storage techniques. Also, by means of the management of users and their privileges, the DBMS can provide secure control access to data. While the control of concurrent access to data is managed through different protocols of transaction scheduling and varied locking techniques, backups and database recovery strategies allow the database recovering after hardware and software failures. All of these areas have opened research fields, exciting challenges, and major technological and conceptual changes in many features through their evolution. This has happened in such a way that database knowledge has become an essential part of the education and research related to computer science.

From another perspective of this fantastic growth of database frameworks, a vast diversity of users, with their areas of interest, particular application requirements and own technological needs, have become interested in databases. In recent years, these facts have promoted a general tendency towards new and original database research areas and practice fields, framed into new approaches. Other new and powerful issues of database research are programming languages, active databases, temporal databases, spatial databases, multimedia databases, and databases and the Web.

The growth of database effectiveness was accompanied by a huge increase of users and user profiles. The need of users for information sources has naturally and extensively grown. We believe this encyclopedia will help to close the breach between both aspects.

The *Encyclopedia of Database Technologies and Applications* is a collaborative effort that addresses the evolution of database management, technologies and applications along with the progress and endeavors of new research areas.

Leading specialists in each area, researchers with profuse publications on the topics covered by this volume, practitioners with a vast experience in the development of database systems in organizational environments, and teachers with accumulated experience teaching graduate and undergraduate courses have contributed articles on their field of expertise to this encyclopedia.

The potential audience of this encyclopedia is widely diverse. This encyclopedia is intended for computing students who have some knowledge on databases, for teachers giving not only introductory but advanced courses on databases, for researchers with concerns about specific concepts in their area of interest, and for practitioners facing database implementation choices. Inexperienced readers and students will also find in this encyclopedia an invaluable basis to increase their knowledge about databases, in addition to the main issues and applications related to databases. Experienced teachers will discover a comprehensive compendium of teaching resources. On the other hand, this database encyclopedia is also a valuable reference book for professionals, designers, application programmers, and database practitioners. This important new publication brings together a discussion of a wide spectrum of database issues, giving researchers, scholars, students, and professionals the access to the latest knowledge related to database science, technology and the trends in this promising and continuously evolving field. We can say that the main endeavor of this encyclopedia has been to grant access to essential core material for education, research, and practice on database systems.

This encyclopedia is a compilation of entries with a standardized structure and style sorted by title. It is broader than most database books since it provides a comprehensive and balanced treatment of the most important issues, concepts, languages, and definitions of the core terms in each subject. It recommends concrete references to current technology and includes profuse links to relevant literature on each topic. It offers a sound grounding in the foundations of database technology and the state of the art and also provides solid ideas about current trends and evidence on how this field is likely to develop in the near future. Many articles offer an analytical approach, so that the concepts presented can be applied not only to the wide variety of actual databases but also can be used as a fundamental stone to build future systems.

The areas covered in this encyclopedia are under an exceptionally fast development and spreading. For this reason, each article of this volume includes a list with the Key Terms and concepts of each topic along with their definitions. The choice of such terms has been a great success of our contributors.

Topics covered by this encyclopedia include the analysis of information requirements, the development of conceptual data models, the improvement of relational and object oriented schemas, SQL and XML concerns, database integrity, database performance and optimization, transactions, backup and recovery strategies, file structures and indexing, and the development of challenging database applications. Many articles have numerous examples on how to apply the material.

ORGANIZATION OF THIS ENCYCLOPEDIA

This encyclopedia has been organized alphabetically by title, in order the reader to find quickly the article of his interest.

This preface introduces the main research areas exposed. It is not intended to detail all the issues and results, but to give an initial orientation of the work introduced in the research articles. A short description of each area covered follows:

Knowledge management is generally seen as a problem of treating the capture, organization, and retrieval of information. Related to this area, the design of knowledge management systems including in that design the human and social factors, constitute a novel and promising research discipline that has been covered in few articles. The reader will find in them particular applications of this approach.

Data integration is the subject of various articles. As the storing and retrieving of data are not enough in newest organizational environments, new approaches to data integration have been emerged. One of them is the *data warehousing* approach. Conventional and current database systems are (and typically have been) designed and optimized for On-Line Transaction Processing (OLTP). In this sense, the databases can manage huge quantities of transactions executing concurrently, modifying or reading generally small and changeable amounts of data. However, recently applications in the context of the management of a complex enterprise require a far-reaching, overall view of all aspects of its activities. The core characteristic of the *data warehousing* approach is that the data of interest come from heterogeneous sources. Since those data are often of a different format and complexity; for instance they can come from different databases, installed from different platforms, or simply from online multimedia data stores, Web pages, spreadsheets, or even conventional files. This poses not only difficulties inherent to its access, but new challenges on the way to provide a uniform, integrated and transparent access to such heterogeneous and distributed information sources. *Federated database systems* are one of the most recent advances that allow distributed databases and other data sources to be virtually integrated.

Related to this new trend, special attention has been paid to the analysis of operational data in order to discover tendencies, singularities, patterns of behavior and perhaps some anomalies. Different analytical tools, called *decision support systems* (DSS) and *data mining*, enable the analysts to make better decisions. All of these related issues have promoted numerous research efforts whose main results can be read in several articles of this encyclopedia.

Several entries have been devoted to *object-oriented database management system* (OODBMS) issues. These systems have generated developments and challenges in the database community, mainly promoted by the genuine requirement of support to the management of data, which is not very well carried out by the relational technology. The advances accomplished through this approach are well-known and are giving rise to the so called “advanced application areas”. Although the market for OODBMSs is still small, this technology has rapidly gained maturity and has notoriously influenced other paradigms. This scenario has given birth to a variety of applications, which are in constant growth and have foreseen the convenience of using this technology.

The World Wide Web can be seen as a database, although is inherent chaos, without enforcing any condition to the sites members of this net. For this reason, the challenge is to create a structure inside the chaos by means of special queries. Many articles provide an excellent snapshot of current research, development activities, and trends in the area of *Web or Internet databases*. The Web as a huge database, and its implications regarding the progressive adaptation of core database techniques to Web usage, are quickly evolving and have created a massive demand for Web-based information systems, posing interesting challenges for researchers and practitioners.

A commencement of organization has been accomplished through the consolidation of the eXtensible Markup Language (XML) as a new standard adopted by the World Wide Web Consortium (W3C). It has facilitated the management of electronic data, mainly due to emerging database system’s current application needs interacting with each other over the Web. Nowadays the researchers talk about the Second-Generation Web. The conjunction of XML and related technologies have resulted in the so-called *semantic Web*, in such a way that the computers will be able to deal with the meaning (“semantics”) of Web data. Several articles are devoted to these topics, covering the state-of-the-art in databases for the Internet era, in addition to the basic material of the main issues related to databases accessed through and interacting on the World Wide Web. These articles conform a strong starting point for practitioners interested in the development of database applications on the Web.

Nowadays, the profound knowledge of the relational model, the evolution to the object-relational model, object-oriented, and deductive models are a core need to effectively face the development of most industrial applications. In this regard, the articles which focus on the related areas of *modeling* and *CASE tools* reflect a wide range of current research and practice areas, given that the quality of the results obtained in the early stages of design are a main factor of the success of an information system. The discussion of design techniques covers topics, which range from the introduction of the concepts of classic models to a novel view on modern metadata management problems. Research on conceptual modeling, its state of the art, modern database paradigms, requirements engineering, CASE tools implementation, and comparison are extensively explored.

Data models’ evolution over time, due to the changes of the user requirements, is a current concern and the subject of articles mainly dedicated to the case of databases. Versioning is one of the useful approaches to manage the evolution of conventional and non-conventional databases. This encyclopedia offers articles providing insight on *database evolution* and *versioning* areas.

Query languages constitute a key piece of data models. Some topics covered extensively are the use of database front-ends to support a useful way to translate queries from natural language to SQL, the use of specific algorithms to provide efficient *query processing* in the case of static time series of data, the description and overcoming of the weaknesses of *relational query languages* in expressing set comparison queries, *query optimization* as an active area of research recently renewed due to the appearance of data warehousing, and the analysis of different approaches for the efficient *evaluation of queries*.

Database integrity is another main subject addressed in this encyclopedia. Articles provide not only a background on basic knowledge about this concern but innovative perspectives and implementations.

Several articles cover *Active database systems*, concepts on active rules and the way these rules are supported in representative relational systems, discussions about how active rules can be generated for integrity maintenance and tested for termination, as well as current applications and trends related to this challenging paradigm.

Temporal issues are also addressed in this volume. Temporal data are data, which represent some aspect of time, incorporating this concept to conventional databases. Some implementation issues and approaches to implement a temporal DBMS as a layer on top of an existing non-temporal DBMS, aspects such as the adaptation of the framework for concurrent processing of operations ensuring the ACID properties of transactions, and the extension of the SQL language to incorporate temporal issues are some of the topics explored in contributed articles.

Multimedia technologies are gaining popularity day after day and recently they constitute an important component of the study of computer graphics, image processing, real-time systems, information retrieval, computer vision, and many other related fields. Modern database systems have enhanced the capabilities of traditional database systems to cope with any kind of data, including text, image, audio, and video, and to provide access to these data via SQL-like query languages. To make such immense amount of data available, a vast research effort is being developed focusing on design models of *multimedia databases* and novel methodologies to provide efficient image indexing and retrieval. This encyclopedia includes articles focusing this research field.

Digital libraries are systems that contain organized collections of objects, in such a way that they resemble traditional libraries containing paper documents. Generally, most of the information contained in a digital library consists in documents. Throughout the last decades the production, storing, retrieving, and management of digital documents has increased explosively, provoking an increased need for solutions for those activities.

From a different perspective, the huge amount of text data available in current applications poses the challenge of effectively searching them for queries of interest. A text database is a system that maintains a text collection and provides fast and accurate access to it. As traditional database technologies are not well suited to handle *text databases*, this novel approach faces the challenge of solving the need of specific technology to manage such kind of data. Digital libraries and text databases are the main subject of several articles.

Fundamental issues of database technology such as transactions, concurrency control, reliability, and recovery are explored into a set of entries, with emphasis on *distribution*, *replication* and *parallelism*. The purpose of these writings is to offer the reader not only the state of the art subjects but to disseminate the knowledge and challenges associated with the scale-up of distributed systems.

Another core concept in the database area is *security*. Conventional approaches address discretionary security by controlling access privileges and modes of data users. On the other hand, mandatory security provides a higher level of assurance by restricting the access to data items to cleared database users. This last approach is adequate for such commercial and military applications requiring a multilevel secure database management system. Relevant authors explore these topics in a set of articles.

In previous years, *Free Software* and *Open Source databases* (OSDBMS) have evolved in a remarkable way, becoming a valuable alternative to the commercial databases. The articles devoted to this topic not only focus on the tendencies outlined by OSDBMS, but also on the most pertinent licenses and in the projects of the database and applications developed by the Free/Open Source Software Development community and on the central role played by this kind of software.

GIS (Geographic Information Systems) allows users to model, store, and analyze information describing physical properties of the real world. Spatial data coming from maps, digital images, routes, field elevations, and data coming from a diversity of censuses, economic studies, market indicators, etc. are the kind of material the systems deal with. Many articles address issues such as data quality, query languages, and integrity within this framework.

The articles which focus on other advanced database applications deal with information retrieval by means of the creation and updating of *database indexes* and their application to newest kind of non-conventional data. These articles also focus on specialized databases such as *biometric databases*, which research basic problems related with pattern recognition, biometric authentication and intelligent surveillance, *bioinformatics* as a recent discipline into the information science, which concerns the creation and maintenance of databases of biological information, to store, extract, analyze, and the use of biological information.

The emergence of powerful portable computers and the development of fast reliable networks have promoted the birth and growing of a new dimension in data processing and communication: *mobile computing*. Mobility and portability have created a new repertory of applications and have produced a deep impact in the data management field. In this sense, *mobile database systems* have emerged as a paradigm bringing support to transactions provided by mobile clients and creating a broad and promising research field in the past decade. Entries exploring mobile databases focus on data dissemination over limited bandwidth channels, location-dependent querying of data, and concurrency control and recovery mechanisms.

In summary, this encyclopedia offers a unique and exciting opportunity to find out, not only about the state of the art of fundamental database concepts, but about current trends, novel technology, and challenging applications to meet the needs of current data management systems.