Preface

Origin

The material in this book is the result of courses given at Stanford University as "File and Database Structures" since 1971. Initially little coherent published material was available, even though a large number of references could be cited. In particular, no clear definition of the concept of a *schema* was available. Now many practical and scholarly sources for material exist but problems of emphasis and analysis remain.

This book brings together knowledge in the area of database management in a structured form suitable for teaching and reference. The first edition has found broad acceptance in course sequences where quantitative approaches are stressed. Uninended, but gratifying, is the place this book has found as a programmer's reference and analysis guide.

Analyses to predict logical correctness and adequate performance of systems prior to investment in an implementation effort should be expected from professional system designers. An analysis of system development methods assigns a cost ratio of 100 to 1 when errors are found in testing rather than caught in the design stage. Many more cross-references have been added and the extensive index has been further expanded to help the professional reader. The background sections have also been completely redone to reflect the best recent references. Other major changes made are discussed in the objective section.

Much has happened in the database field since the first edition was published. The area has become an accepted academic discipline. Conferences and journals covering the area abound. Selection of material is now more difficult than finding material. At the same time commercial interest in database management has exploded creating a serious shortage of qualified people. Already now the number of professionals working in the database area exceeds the number employed in traditional computing fields as compilers and operating systems. Not everyone working today with databases has the backbground to deal well with the complexities that can arise, so that complaints of system inadequacies are common.

Much of the literature remains descriptive or presents one-sided experiences, inadequate to provide a basis for the transfer of concepts to diverse applications. An effort has been made here to develop concepts from the wealth of material and to present the subject in such a way that the concepts which evolve can be applied in practice. An engineering attitude to the problems of database organization has been used in order to combine formality with applicability.

I hope that this greatly revised text will continue to fill the needs and that it will help extend and improve the teaching of the data-processing aspects of computer science.

Objective

This book is intended to present the methods, the criteria for choices between alternatives, and the principles and concepts, that are relevant to the practice of database design. No actual systems are described completely, nor are systems surveyed and compared, although specific and realistic examples are used throughout to illustrate points being made. The material provides the basis to allow the reader to understand, recognize the implications, and evaluate database approaches. Databases in this sense are a broader concept than database management systems. The design of a database involves understanding the meaning of the data, the systems chosen, be they database management systems or traditional file systems, just help in the implementation.

This book includes two major parts:

1. The description and analysis of file systems (Chaps. 2 to 6)
2. The description and analysis of database systems (Chaps. 7 to 10)

The first part is intended to provide a solid foundation for the latter part, since the issues arising in database design are difficult to discuss if file-design concepts are not available to draw upon. A number of subjects which pertain to both files and databases, anmely reliability, protection of privacy, integrity, and coding are presented in the third part, consisting of Chaps. 11 to 14. If the material is taught in two distinct courses, the third part should not be ignored in either course.

The audience for this book ranges from students of computing who have finished a reasonably complete course in programming to applications and systems programmers who wish to synthesize their experiences into a cmore formal structure. The material covered should be known by systems designers or systems analysts faced with implementation choices. It probably presents too much detail to be of interest to management outside the database management area itself.

The revision of this book has been major. Nearly every paragraph has been rewritten, and some sections have been completely replaces. The number of tables and examples is also increased. The general outline, however, and the underlying principles could remain the same. Distribution of databases over multiple sites is now considered throughout the book. The equipment table in Chap. 2 deals with the much wider range of storage devices now available. The prevalence of B-trees for indexes has permitted a rewrite of Sec. 3-4 which is more modern and simpler. A new method for dealing with growth of direct-access areas is part of Sec. 3-5. New results dealing with device interference are presented in Chap. 5.

Chapter 7 includes the modern concepts of semantic modeling. a new section, 7-2, deals with the formal semantics now available for database design. Section 7-3 defines the conceptual tools for establishing the structural relationships between files. The entire design process is described step by step in Sec. 7-5. The introduction of commercial relational database implementations permits now a consistent description of these systems in Sec. 9-2. the performance of databases using relational operations can be predicted using the information introduced in Sec. 9-3.

The concept of using transactions to access the database is used throughout but has a major impact on the handling of reliability issues in Chap. 11 and integrity maintenance in Chap. 13. New sections have been added there.

Design Methodology

This book presents a comprehensive collection of database design tools. In order to employ them, a strategy of problem decomposition, followed by a structured design process is advised. A top-down design requires the underlying primitives to be understood. Since this book starts with the basics, the design process is initiated with concepts from Chap. 10.

The categorization of database approaches given in Chap. 10 helps to set the initial objective for a database. Chapter 7 provides the means to construct a model which integrates the requirements of multiple applications which share the database. The schemas in Chap. 8 provide methods to describe the model in machine-readable form. Existing database systems, described in Chap. 9 and referenced in Appendix B, suggest available implementation alternatives.

If the database is to be directly supported by a file system, the basic file choices in Chap. 3 and their combinations shown in Chap. 4 provide the alternatives. The data representation can be chosen using the material of Chap. 14.

The performance of the chosen approach can be predicted following the outline shown in Chap. 5. Factors relevant to specific database systems or file systems appear where they are discussed, but the terminology and variable definitions are consistent throughout, so that cross-referencing is simple. The structural model defined in Chap. 7 provides the framework for the translation of application loads to the load to be handled by the database. Transaction performance in database systems is estimated using the performance analyses for the prevalent approaches from Chap. 9. An Optional file design may be selected after application of the load parameters to the performance formulas from Chaps. 3 and 4. The formulas also require the hardware description parameters introduced in Chap. 2.

Problems of reliability, protection, and integrity (Chaps. 11, 12, and 13) require a close scrutiny of the available operating system. The long-term maintenance of a database is guided by considerations presented in Chap. 15.

Curricula

Modern curricula give increased emplasis to files and databases. This book matches the suggested courses well. The material of Chaps. 1 to 6 covers course CS-5 (Introduction to File Processing) and Chaps. 7 to 14 cover course CS-11 (Database management System Design) as specified in the report by the ACM Curriculum on Computer Science [Austing et al[78]]. The quantitative approach taken in this book causes that algorithmic and analytic material assigned to courses CS-6 and CS-7 is included as w3ell, albeit limited by relevance to databases. I do not agree with the recommendation that students write a database management system during CS-11; this is apt to be an exercise in trivializing real issues. Design and performance prediction of a nontrivial database application is part of the course taught at Stanford and enables a broad exposure to important concepts beyond programming. I agree with Ralston and Shaw[80] that mathematical competence is necessary for a computer science student.

The text provides all the material for the file and database subjects of Courses C1, C2, C3, C4, and D2 specified in the Curriculum Recommendations for Graduate professional programs in Information Systems [Ashenhurst[72]]. The author feels, however, that these courses are easier to teach using sa depth-first approach to the subjects versus the breadth-first approach advocated in the proposal. I have been impressed by constructive comments received from readers who used the book for self-study.

In many schools files and databases are still neglected, possibly because of a shortage of teachers. Students who enter industry or commerce with a bachelor's-level education in computing or data-processing feel this void sharply. It is now reasonable to expect that students majoring in computing and computer applications should be familiar with this subject area [Teichrow[71], Sprowls[75]]. Projections regarding the future use of computers give a considerable weight to the importance of the database area [Dolotta et al[76]], so that we can expect an increasing demand for educational services in this area.

Terminology

We are grateful that the terminology in the area of database and file management is becoming more consistent. Within this book a major effort has been made to define all terms and to use only one term per concept. some terms were changed since the first edition because usage developed differently than I had foreseen. All terms are listed in the index. In order to aide both experienced readers and users of the references, Appendix A cites common alternate terminology and compares it with the terminology used in this text. The introductory chapter is mainly devoted to definitions. It is assumed that subjects such as programming and basic functions of operating systems are familiar to the reader.

Most of the program examples throughout the text use a simple subset of PL/1. The variable names are chosen so that they will aid in the ccomprehension of the prtogram, they are printed in lowercase. Keywprds which are to be recognizes by translating programs appear in uppercase, The programs are designed to be obvious to readers familiar with any procedure-oriented programming language. A number of introductory PL/1 texts [Hume[75], Richardson[75], Mott[72]] can be used to explain features that are not recognized. Som PL/1 texts, unfortunately, omit the statements required for the manipulation of data files.

Many of the examples illustrate features of actual systems and applications, but are of necessity incomplete. An effort has been made to note simplifying assumptions. The same should be done in students' design assignments, so that awareness of real-world complexities is fostered without overwhelming the design with trivia.

Exercises

The exercises listed in each chapter have been kept relatively simple. It is strtongly suggested that an analysis of some of the systems described in the referenced literature be made part of some of the assignments. The analysis or comparison of actual systems may seem to be an excessively complex task, but has been shown to be manageable by students when the material of this book has been assimilated. Appendix B provides references to a number of database systems.

The primary exercise when this course is being taught at Stanford is a design project. Early in the course students prepare an objective statement for a database application of interest to them. Some individual research may be needed to obtain estimates of expected data quantities and transaction load frequencies. The students prepare a structural mpdel of desribing their database while Chap. 7 is being covered. The project is fleshed out with a schema description and a perfomance prediction of selected important transactions for the appplication. Exercises related to this project appear throughout the text and are labeled with a superscript [p].

References

Source material for this book came from many places and experiences. References are not cited throughout the text since the intent is to produce primarily a text and reference book which integrates the mny concepts and ideas available for database design. An extensive background section at the end of every chapter cites the major sources used and indicates further study material. The references provide a generous foothold for students intending to pursue a specific topic in depth. The references can also direct research effort toward the many yet unsolved problems in the area.

The bibliography has been selected to include some important material for each of the subject areas introduced. The volume of database publications is now such that a comprehensive bibliography is beyond the scope of a textbook. Only sources that are easy to obtain, such as books and journals, have been chosen. Papers which appear in conference proceedings containing much relevant material are cited only by author and proceedings, and do not appear individually in the bibliography. Typically only one or two early publications and some instances of recent work in an area are cited. To provide a complete view, references in the recent material have

to be traced. Trade publications, research reports, theses, and computer manuals are referenced only when used directly, although much relevant information can be found there. Up-to-date information on computer and software systems is best obtained from manufacturers.

I apologize to the authors of work I failed to reference, either due to application of these rules, or because of lack of awareness on my part. A large, annotated, bibliography is being maintained by me and is available. I prefer to distribute the bibliography in computer-readable form since it is too large to be effectively scanned without computer assistance.

This book would not have been written without the inspiration, support, and just plain hard work by my wife, Voy. The appreciation she has received from her students, and users of computer manuals she has written, has encouraged me to attempt to presennt this material in as straightforward a fashion as she has been able to present PL/1 programming.

Gio Wiederhold