Preface

This fourth edition of *Readings in Database Systems* is being issued at an interesting time in the history of the field. The database industry has undergone significant consolidation in recent years. It is now dominated by three companies, two of which are not database vendors per se. IBM and Microsoft manage large portfolios of products and services; database systems are one—but only one—of their crown jewels. The third major player, Oracle, is nominally a database company, but is as active in enterprise applications as it is in core database systems. The era of the "database wars" is over, and it has been a long time since a database startup company has made major noise. The argument is sometimes made that database management systems—like many other components in the computing industry—are a victim of their own success: they have reached a level of maturity at which significant industrial innovation has become impossible.

Even if this were an accurate assessment of the entrenched database industry, it is a very narrow view of database research. The research field itself is healthier than it has ever been. There has been a surge of database faculty hiring in recent years, including at some of the leading research institutions that traditionally ignored the field. New conferences have emerged both in database systems design, and in more algorithmic fields like data mining. Lessons from the database literature are being applied in a host of forward-looking research areas at universities, from bioinformatics to sensor networks to next-generation Internet architectures.

This external interest in database technologies is not confined to academia. Industrial software systems are increasingly turning to database system innovations to solve other problems. Rumor has it, for example, that Microsoft's next operating system will have a single unified store for all files and data, based on their database engine. Web-based e-commerce services depend to a large extent on transactional messaging technologies developed in the database community. Text-based web services like search engines also owe a debt to database innovations in parallel query processing. The list goes on.

It would seem, then, that while the core industrial database products have gelled, the ideas that they encapsulate have become increasingly influential. A much more optimistic and realistic view of database research is that the field is in a position to make a bigger impact on computing in the large than it ever has before, in part because the community has solved many of its own challenges and is courting other areas for collaborations. This cross-fertilization could result in major changes in the traditional database industry, and in other aspects of computing.

This book is intended to provide software technologists—both professionals and students—with a grounding in database research past and present, and a technical context for understanding new innovations. It is also designed to be a reference for anyone already active in database systems. This set of readings represents what we perceive to be the most important issues in the database area: the core material for any DBMS professional to study.

The book opens with two introductory articles we wrote to set the stage for the research papers collected here. The first article presents a historical perspective on the design of data models and query languages; the second provides an architectural overview of the anatomy of a database system. These articles are intended to provide an organized, modern introduction to basic knowledge of the field, which in previous

editions was represented by a sampling of seminal research papers from the late Ted Codd and the pioneering INGRES and System R projects. A true database aficionado should still read those original papers [Cod70,ABC+76, SWK76, Sto80, CPS+81], since they give a snapshot of the excitement and challenges of the time. However we felt that after three decades it was hard for readers to get a substantive basis for the field in its current richness by reading the early papers. Hence with some notable regret we chose not to include them in this edition.

For the remaining papers we have selected, we provide chapter introductions to discuss the context, motivation, and, when relevant, the controversy in the area. These introductions summarize the comments we make during lectures in our graduate courses, and place the papers in the broader perspective of database research. The comments are often explicitly intended to be *opinions*, not necessarily statements of fact—they are intended as conversation-starters. We hope this style encourages students and independent readers to critically evaluate both the papers and our editorial remarks.

This edition of the book contains a host of new papers, including a number of chapters in new areas. Four of the papers were written expressly for the book: the two introductory articles, Brewer's paper on search engine architecture, and Jacobs' paper on application servers. The remaining papers we chose from both the classical literature and from recent hot topics. We selected papers based on our assessment both of the quality of research and its potential for lasting importance. We have tried to assemble a collection of papers that are both seminal in nature and accessible to a reader who has a basic familiarity with database systems. We often had two or more papers to choose from. In such cases we selected what we felt was the best one or the one discussing the broadest variety of issues. In some areas such as transaction management, all of the research is very detail-oriented. In these cases we tried to favor papers that are accessible. In areas like data mining with a strong mathematical component, we tried to select papers that are both accessible to software systems experts, and that deal non-trivially with systems challenges.

This book has been greatly improved by the input of many colleagues, including: Paul Aoki, Eric Brewer, David DeWitt, Mike Franklin, Johannes Gehrke, Jim Gray, James Hamilton, Wei Hong, Guy Lohman, Sam Madden, Chris Olston, Tamer Ozsu, Raghu Ramakrishnan, Andreas Reuter, and Stuart Russell. We particularly thank Eric Brewer and Dean Jacobs for their contributions of new material. Thanks are also due to the students of CS286 and CS262 at Berkeley, and 689.3 at MIT; their comments have been a major influence on our choice of papers and our presentation of the material.

References

[ABC+76] Morton M. Astrahan, Mike W. Blasgen, Donald D. Chamberlin, Kapali P. Eswaran, Jim Gray, Patricia P. Griffiths, W. Frank King III, Raymond A. Lorie, Paul R. McJones, James W. Mehl, Gianfranco R. Putzolu, Irving L. Traiger, Bradford W. Wade, and Vera Watson. System R: Relational Approach to Database Management. *ACM Transactions on Database Systems (TODS),* 1(2):97-137, 1976.

[CPS+81] Donald D. Chamberlin, Franco Putzolu, Patricia Griffiths Selinger, Mario Schkolnick, Donald R. Slutz, Irving L. Traiger, Bradford W. Wade, Robert A. Yost, Morton M. Astrahan, Michael W. Blasgen, James N. Gray, W. Frank King, Bruce G. Lindsay, Raymond Lorie, James W. Mehl and Thomas G. Price. A History and Evaluation of System R, *Communications of the ACM* 24:10 (1981), 632-646.

[Cod70] E. F. Codd. A Relational Model of Data for Large Shared Data Banks, *Comm. ACM* 13(6), June 1970, p377-387.

[SWK76] M.R. Stonebraker, E. Wong, and P. Kreps. The Design and Implementation of INGRES. *ACM Transactions on Database Systems (TODS)*, 1(3):189-222, September 1976.

[Sto80] M. Stonebraker. Retrospection on a Database System. *ACM Transactions on Database Systems (TODS)*, 5(2):225-240, 1980.