Preface

In the ten years since the previous edition of *Readings in Database Systems*, the field of data management has exploded. Database and data-intensive systems today operate over unprecedented volumes of data, fueled in large part by the rise of "Big Data" and massive decreases in the cost of storage and computation. Cloud computing and microarchitectural trends have made distribution and parallelism nearly ubiquitous concerns. Data is collected from an increasing variety of heterogeneous formats and sources in increasing volume, and utilized for an ever increasing range of tasks. As a result, commodity database systems have evolved considerably along several dimensions, from the use of new storage media and processor designs, up through query processing architectures, programming interfaces, and emerging application requirements in both transaction processing and analytics. It is an exciting time, with considerable churn in the marketplace and many new ideas from research.

In this time of rapid change, our update to the traditional "Red Book" is intended to provide both a grounding in the core concepts of the field as well as a commentary on selected trends. Some new technologies bear striking resemblance to predecessors of decades past, and we think it's useful for our readers to be familiar with the primary sources. At the same time, technology trends are necessitating a re-evaluation of almost all dimensions of database systems, and many classic designs are in need of revision. Our goal in this collection is to surface important long-term lessons and foundational designs, and highlight the new ideas we believe are most novel and relevant.

Accordingly, we have chosen a mix of classic, traditional papers from the early database literature as well as papers that have been most influential in recent developments, including transaction processing, query processing, advanced analytics, Web data, and language design. Along with each chapter, we have included a short commentary introducing the papers and describing why we selected each. Each commentary is authored by one of the editors, but all editors provided input; we hope the commentaries do not lack for opinion.

When selecting readings, we sought topics and papers that met a core set of criteria. First, each selection represents a major trend in data management, as evidenced by both research interest and market demand. Second, each selection is canonical or near-canonical; we sought the most representative paper for each topic. Third, each selection is a primary source. There are good surveys on many of the topics in this collection, which we reference in commentaries. However, reading primary sources provides historical context, gives the reader exposure to the thinking that shaped influential solutions, and helps ensure that our readers are well-grounded in the field. Finally, this collection represents our current tastes about what is "most important"; we expect our readers to view this collection with a critical eye.

One major departure from previous editions of the Red Book is the way we have treated the final two sections on Analytics and Data Integration. It's clear in both research and the marketplace that these are two of the biggest problems in data management today. They are also quickly-evolving topics in both research and in practice. Given this state of flux, we found that we had a hard time agreeing on "canonical" readings for these topics. Under the circumstances, we decided to omit official readings but instead offer commentary. This obviously results in a highly biased view of what's happening in the field. So we do not recommend these sections as the kind of "required reading" that the Red Book has traditionally tried to offer. Instead, we are treating these as optional end-matter: "Biased Views on Moving Targets". Readers are cautioned to take these two sections with a grain of salt (even larger that the one used for the rest of the book.)

We are releasing this edition of the Red Book free of charge, with a permissive license on our text that allows unlimited non-commercial re-distribution, in multiple formats. Rather than secure rights to the recommended papers, we have simply provided links to Google Scholar searches that should help the reader locate the relevant papers. We expect this electronic format to allow more frequent editions of the "book." We plan to evolve the collection as appropriate.

A final note: this collection has been alive since 1988, and we expect it to have a long future life. Accordingly, we have added a modicum of "young blood" to the gray beard editors. As appropriate, the editors of this collection may further evolve over time.

Peter Bailis
Joseph M. Hellerstein
Michael Stonebraker