

E. F. CODD AND RELATIONAL THEORY

A Detailed Review and Analysis of Codd's Major Relational Writings

by

C. J. Date

Lulu Press, Inc. (2019)
ISBN: 978-1-68470-527-6

304 pages: \$24.99¹ +++ ???

Background

As everyone knows, the foundation of relational database theory is the relational model. The relational model was invented back in 1969-1970 by E. F. (“Ted”) Codd, who received the 1981 ACM Turing Award—the most prestigious award in computer science—for that work. Of course, Codd published a large number of writings dealing with the relational model and relational theory, and relational database technology in general, over the course of his career. Of those various writings, however, the ones that were (and still are) the most significant by far are as follows:

- “Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks” (1969)
- “A Relational Model of Data for Large Shared Data Banks” (1970)
- “A Data Base Sublanguage Founded on the Relational Calculus” (1971)
- “Further Normalization of the Data Base Relational Model” (1971)
- “Relational Completeness of Data Base Sublanguages” (1972)
- “Interactive Support for Nonprogrammers: The Relational and Network Approaches” (1974)

¹ An epub version (ISBN 978-1-68470-528-3) will be available very soon and will be priced more cheaply. However, I had little control over the detailed format and layout of that version, and I’m sorry to have to say that I find the result, though readable, much less satisfactory than the regular printed version.

- A two-part paper, “Is Your DBMS Really Relational?” and “Does Your DBMS Run by the Rules?” (1985)
- A book, *The Relational Model for Database Management Version 2* (1990)

These writings of Codd’s were staggering in their originality. Among other things, they changed, permanently, the way database management was perceived in the IT world; more specifically, they transformed what had previously been nothing but a ragbag of tricks and ad hoc techniques into a solid scientific endeavor. They also, not incidentally, laid the foundation for an entire multibillion dollar industry. Together, they provided the basis for a technology that has had, and continues to have, a major impact on the very fabric of our society. Thus, it’s no exaggeration to say that Codd is the intellectual father of the modern database field. Here for the record is a summary of his major contributions:

- His biggest overall achievement was to make database management into a science; in other words, he put the field on to a solid scientific footing, by providing a theoretical framework within which a variety of important problems could be addressed in a scientific manner.
- As a consequence of the previous point, he introduced a welcome and sorely needed note of clarity and rigor into the database field.
- He introduced not only the relational model in particular, but the whole idea of a data model in general.
- He stressed the importance of the distinction, regrettably still widely misunderstood, between model and implementation.
- He saw the potential of using the ideas of predicate logic as a foundation for database management.
- He defined both a relational algebra and a relational calculus as a basis for dealing with data in relational form.
- He defined (albeit only informally) what was probably the first relational language, “Data Sublanguage ALPHA.”
- He introduced the concept of functional dependence and defined the first three normal forms (1NF, 2NF, 3NF).
- He defined the key notion of essentiality.

To repeat, Codd's writings are the foundation texts on which the entire relational database field is based. Yet very few people have actually read those texts, or even have much awareness of what they contain. So one thing the present book does is explain in each case just what the contribution was, how and why it was important, and also to some extent what it wasn't. Note, however, that the book isn't meant as an adulatory or purely uncritical review; rather, it provides a detailed analysis of the texts in question, with the aim, in part, of seeing how well they hold up to such careful analysis after all this time. Thus, portions of the book are quite critical (even severely so, on occasion). However—this shouldn't need saying, but I don't want to be misunderstood here—my overall aim is to be constructive. Basically, what I do in the book is this: First, I highlight and describe the many fundamental things that Codd got right; at the same time, I also identify and explain aspects on which Codd wasn't as clear as he might have been or was, in some cases, almost certainly wrong.

Let me add too that, although it might be immodest to say so, there's no one more qualified than myself to write such a book. I knew Codd well. We worked as colleagues for many years, first in IBM (starting in 1970) and subsequently in a series of companies—Codd & Date International and various subsidiaries of that company—that, along with Sharon Weinberg, Codd and I formed in the early 1980s. In fact, I was one of Codd's earliest supporters—especially during the first half of the 1970s, when he was almost literally a voice crying in the wilderness—and I was instrumental through my writing and lecturing in bringing his ideas to a wider audience.

Target Audience

The book is aimed at anyone professionally interested in data or database management, including college professors and students, database designers, “data modelers,” application developers, language and DBMS designers and implementers, and—perhaps most of all—anyone interested in the history of the database field. In fact I think it belongs in all database libraries; it might also serve as a textbook for various college courses. Readers are assumed to have a basic understanding of the relational model—possibly of SQL also, although that's not required. Familiarity with the content of my book *Relational Theory for Computer Professionals: What Relational Databases Are Really All About* (O'Reilly, 2013) would be more than sufficient.

Book Structure

The book is arranged into three principal parts, as follows:

- I. Setting the Scene
- II. Codd's Early Writings
- III. Codd's Later Writings

There are also three appendixes. A detailed table of contents appears after the “Competition” discussion immediately following.

Competition

I’m aware of only one book that might be seen even remotely as a competitor to the present book, and that’s one by myself, *The Database Relational Model: A Retrospective Review and Analysis* (Addison-Wesley, 2001). In fact I believe that earlier book still provides a useful overview of Codd’s achievement overall—but it’s much more superficial than the present book (it’s really little more than a booklet), and it doesn’t even begin to get into the same kind of detail and extensive discussion as the present book does.

I’d like to add that the book has already received detailed reviews by four technically knowledgeable reviewers—all acquaintances of mine, admittedly, but they were instructed not to spare my feelings in their comments, and they didn’t, and the text has benefited from their comments accordingly.

Detailed Table of Contents

Preface

PART I SETTING THE SCENE

Chapter 1 What’s a Relational DBMS?

“Data looks relational”

“Relational operators are available”

A little history

Why databases must be relational

The relational model defined

Types

The RELATION type generator

Relation variables

Relational assignment

Deriving relations from relations

Integrity constraints

Concluding remarks

Appendix: SQL departures from the relational model

PART II CODD'S EARLY WRITINGS

Chapter 2 The First Two Papers

What's the relational model?
What's a relation?
What's a domain?
What's a time-varying relation?
What's a key?
What about the operators?
What about data independence?
What about integrity?
And what about nulls?
Appendix: Relations revisited

Chapter 3 The Completeness Paper

Structure of the paper
Preliminary definitions
Relational algebra
Relational calculus
Codd's reduction algorithm
Calculus vs. algebra
Miscellaneous comments

Chapter 4 The ALPHA Paper

Overview
Data definition
Retrieval operations
Update operations
More on retrieval
Miscellaneous comments
Concluding remarks

Chapter 5 The Further Normalization Paper

Informal overview
Toward a more formal treatment
Integrity constraints
Keys

First normal form
Second normal form
Third normal form
Functional dependency precisely defined
“Admissible states”
Logical data independence
Miscellaneous comments

Chapter 6 The Essentiality Paper

Historical background
Essentiality
Structure of the paper
Objectives and strategies
Questions concerning owner-coupled sets
Essentiality revisited
The comparative example

PART III CODD’S LATER WRITINGS

Chapter 7 The Relational Model Version 1

Overview
Structural features
Integrity features
Manipulative features
Concluding remarks
Acknowledgments
References and bibliography
Appendix: Codd’s twelve rules

Chapter 8 The Relational Model Version 2

Overview
General remarks
A survey of RM/V2
Major areas of concern
Miscellaneous comments
Concluding remarks
References and bibliography

Postscript
Appendix: The fundamental laws of database management

APPENDIXES

Appendix A The Turing Award Website Piece

Name
Citation
Major content essay
Capsule biographical information items
Portrait
Annotated bibliography
Attachment
Sidebar

Appendix B Formal Definitions

Appendix C Consolidated List of References

Index

End of Overview