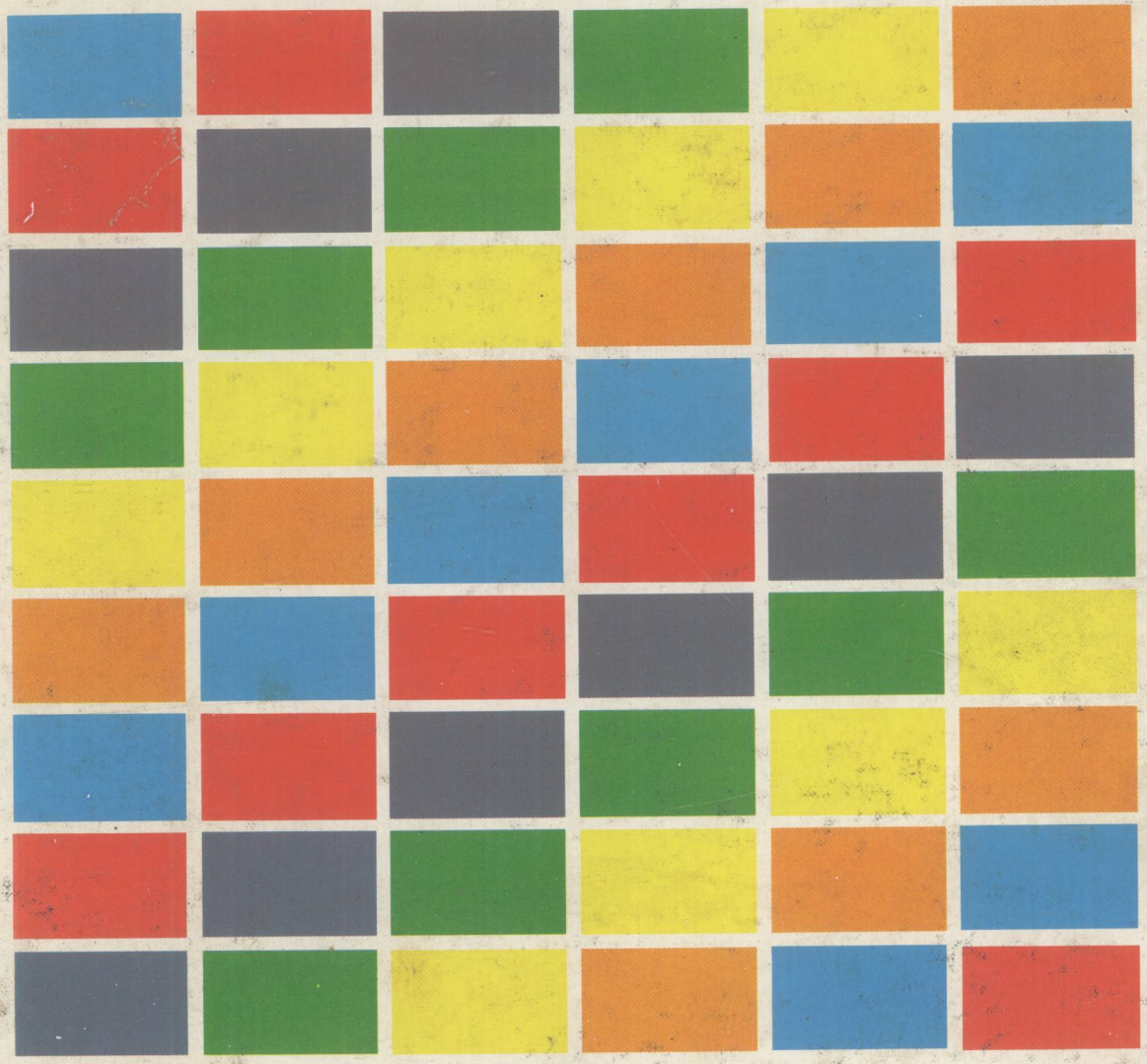
Richard C. Perkinson

Data Analysis
The Key to Data Base Design



QED® Information Sciences, Inc.

Data Analysis: The Key to Data Base Design

Richard C. Perkinson

QED® Information Sciences, Inc.
Wellesley, Massachusetts

I would like to thank several publishers for permitting me to use material from their publications. Figure 1.2 is reprinted from the December 8, 1980, issue of COMPUTERWORLD, copyright 1980 by CW Communications/Inc. Figure 1.8 was adapted with the permission of Nolan & Norton Company, Lexington, Mass., and based on Richard L. Nolan, "Managing the Crisis in Data Processing," *Harvard Business Review*, March-April 1979. Specified brief excerpts in Chapter 2 of this book are from *Management: Tasks, Responsibilities, Practices*, by Peter F. Drucker (copyright 1973, 1974, Peter F. Drucker) and are reprinted by permission of Harper & Row, Publishers, Inc. Figure 8.8 was adapted with the permission of Cullinet Software, Inc., Westwood, Mass.

©1984 by QED® Information Sciences, Inc.
QED Plaza • Box 181
Wellesley, MA 02181

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or transmitting, or by any information storage or retrieval system, without written permission of the copyright owner.

Library of Congress Catalog Card Number 83-63212
ISBN 0-89435-105-2

Printed in the United States of America.

Data Analysis: The Key to Data Base Design

ALSO FROM QED®

MANAGEMENT

Strategic Planning for Information Systems
Applying Business Planning Techniques to Information Systems
Managing for Productivity in Data Processing
Effective Methods of EDP Quality Assurance
Managing Systems Maintenance
A Guide to EDP Performance Management
The Handbook for Data Center Management
The Data Center Disaster Consultant
Handbook for Data Processing Educators

TECHNICAL

Computer Control and Audit
Humanizing Office Automation: The Impact of Ergonomics on Productivity
Handbook of Screen Format Design
Architecture and Implementation of Large Scale IBM Computer Systems
1001 Questions and Answers to Help You Prepare for the CDP® Exam
Handbook of COBOL Techniques
High Level COBOL Programming
Humanized Input: Techniques for Reliable Keyed Input
CICS/VS Command Level Reference Guide for COBOL Programmers
EDP System Development Guidelines
Time Is of the Essence: The DP Professional's Guide to Getting the Right Things Done
Managing Data Base: Four Critical Factors
Data Base Analysis: The Key to Data Base Design

DATA BASE

Data Base Management Systems in the Eighties
Creating and Planning the Corporate Data Base System Project
Data Base Techniques
IMS Design and Implementation Techniques
Design Guide for CODASYL Data Base Management Systems
The Data Base Monograph Series
Data Base Systems: A Practical Reference
Information Resource/Data Dictionary Systems

TELECOMM/DATA COMM

Transnational Data Regulation: The Realities
Distributed Processing: Current Practice and Future Developments

THE QED® PERSONAL COMPUTING SERIES

The IBM Personal Computer: What You Should Know
Learning to Use the IBM Personal Computer
How to Buy Software for Personal Computers
Peachtree Software for Personal Computers: Introduction & Description

QED® LEADING EDGE LEARNING SOFTWARE

PC PLUS: Learn The IBM PC . . . Plus Computing Fundamentals
DOS PLUS: Learn to Use IBM PC DOS

FOR INFORMATION, CONTACT:

QED® Information Sciences, Inc.
QED Plaza • P.O. Box 181
Wellesley, Massachusetts 02181

试读结束：需要全本请在线购买： www.ertongbook.com

Acknowledgments

I wish to thank several people at QED Information Sciences, Inc., without whom this book would never have come to such a successful fruition.

First, let me thank Peters P. Jones, my colleague at QED, who reviewed every word of this text. The "how to" list found in Appendix E is almost entirely his creation. It synthesizes beautifully all the steps in data analysis. I know the readers of this text will appreciate that contribution when they do their first data analysis. I would like to acknowledge another colleague, Parker Shannon, who did the original planning on which the banking case study in Appendix F is based.

Second, I wish to acknowledge the many contributions and hard work of QED's art department, in particular, Donna St. Martin and Sybil Norwood, whose artistic and timely contributions are greatly appreciated by this author.

Last, to the many students in my classes who have made countless suggestions and improvements in this manuscript, I say thank you.

PREFACE

The inability of most data base designers to articulate to their user community the steps necessary to design a data base successfully, or even, in a basic sense, to group data together, is the reason for this book. This inability to articulate is for the most part unintentional on the part of the designers. The designers themselves have never really been asked to describe how they do it. They offer excuses, saying the task is "intuitively obvious," or that their techniques are "based on my eleven years experience," or some other such remark and never really analyze for anyone how they do what they do.

Implementation of the data base designs have gotten bogged down in the syntax and technology of the data base management system (DBMS) chosen. Lost in the shuffle are the user and the application—the raison d'être for the whole thing. Unnecessary constraints are applied to the design by the data base administrator (DBA) or designer based solely on the methodology of the DBMS, not the restrictions of the data.

This book is an effort to relieve some of this user discomfort with data base. Through the technique of *data analysis*, users will be able to understand data and communicate to the designer their needs more accurately. For the designer, data analysis provides a simple means to a desired end. It allows reliable prognostication of the performance and reliability of a design while it is still in the paper and pencil stage. The products of the handbook can easily be mapped to any commercially available

DBMS. The entire technique is a five-step process from initial data analysis to final physical data base design.

In addition to espousing data analysis and data base design techniques, this text will cover what the role of data is in an organization; when and how to use data dictionaries; and why and how to use data administration and the data base administration function.

In short, this book is an attempt to bridge the gap between structured analysis and design and management disciplines, such as business systems planning (BSP) or long-range systems planning (LRSP), by providing a simple method of going from one to the other in either direction by following a few simple rules.

CONTENTS

Preface

Acknowledgments

PART ONE Data Analysis

1 Introduction to Data Analysis 3

What Part I Is All about
Data Identification Processes
Relevance of Data Analysis and Design
Characteristics of an Organization
Traditional Data Processing Approaches
Procedure-Oriented versus Data-Oriented Design
Corporate Data Processing Growth

2 Functional Analysis 23

Determining What To Do
Stages of Functional Analysis

Sample Functional Analysis Problem
Introduction to the Data Flow Diagram

3 Normalization: Creating the Relations **37**

Grouping Data
Normalization: First, Second, and Third Normals Forms
Relevance of Normalization
Business Functions and Relationships
Relational Considerations: Projection, Joining, and Partitioning

4 Relational Mapping **55**

Relational Maps: Types and Nature
Transaction Mapping: Frequency Ratios along Defined Paths

5 Transaction Analysis **67**

What Is a Transaction Map?
Building Transaction Maps for E-Z Chair
Composite Overlay Mapping

6 Constructing the Data Model **79**

Quantitative Analysis of the Functional Model
The Transaction Path Analysis Form
Using Mathematical Probability to Predict Entity Frequency
The Loaded Matrix
Implications and Factors in Physical Design

PART TWO Data Base Design

7 Data Base Management Systems: Their Architecture and Evolution **101**

Primary DBMS Architectures
Evolution of Data Base
TOTAL
Information Management System
CODASYL
Inverted Lists
Relational DBMS

8 Data Base Design Methodology **121**

Record Grouping and DBMS Alternatives
Review of Schema Design
Logical Synthesis
Summary of Subschema Design
Overview of Integrity Security and Audit Design
Integrity Design Summary
Factors Affecting System Operation and Performance
Summary of Operations and Performance Design
Data Dictionary: Essential Design Tool
The Role of the Data Base Administrator Creating the DBA Function in an Organization
Job Functions of the DBA
Tools of the DBA
Interfaces of the DBA
The DBA's Staff

9 Designing for Network Architectures **155**

IDMS Schema (CODASYL Architecture)
TOTAL Schema (Network Architecture)
Implication of DBMS Selection
IDMS (CODASYL)
TOTAL Subschema
Security and Audit Design
CODASYL Data Integrity
TOTAL Data Integrity
IDMS and CODASYL Operations and Performance Design Considerations

10 Designing for Hierarchical Architectures **169**

IMS Schema (Hierarchical Architecture)
IMS Subschema
Summary of Subschema Design
IMS Data Integrity
IMS Operations and Performance Design

11 Designing for Relational Architectures **181**

System/38 Schema
QBE: Query-By-Example
System/38—Subschema
QBE Subschema
System/38 Data Integrity
QBE Data Integrity
System/38 and QBE—Performance

试读结束：需要全本请在线购买： www.ertongbook.com

12 Designing for the Inverted List Structure 193

ADABAS Schema (Inverted List Structure)
System 2000 Schema—Fully Inverted List Structures
ADABAS Subschema
System 2000 Subschema
Inverted List Data Integrity
ADABAS Operations and Performance Design
System 2000's Operation and Performance
Review

Appendixes

Appendix A **Usage Maps for the Sample Problem** 205
Appendix B **Glossary of Terms** 211
Appendix C **Solution to Normalization Sample Problem** 221
Appendix D **Solution to Functional Analysis Sample Problem** 223
Appendix E **The Nine Deliverables** 227
Appendix F **Two Sample Case Studies** 235

Bibliography 283

FIGURES

FIGURE 1.1	Stages of Data Analysis	5
FIGURE 1.2	Price Performance	8
FIGURE 1.3	People Costs versus Hardware Costs	10
FIGURE 1.4	Files Answer Questions	13
FIGURE 1.5	Typical Data Flow at E-Z Chair Company	14
FIGURE 1.6	Departmental View of the Data Base	16
FIGURE 1.7	Traditional Departmentalized Files and Application Intercommunication	17
FIGURE 1.8	Nolan's Six Stages of EDP Growth	21
FIGURE 2.1	Annual Report of the E-Z Chair Corporation	26
FIGURE 2.2	Sample Relational Map from E-Z Chair Functional Analysis	33
FIGURE 2.3	Sample Data Flow Diagram for E-Z Chair Corporation	34

FIGURE 3.1 Sample E-Z Chair Corporation Order Form 39
FIGURE 3.2 Sample Problem 46
FIGURE 3.3 Stage 1—Normalization of Data Flow 53
FIGURE 4.1 Example of One-to-Many Set 56
FIGURE 4.2 Example of Tree or Hierarchical Data Base Architecture 57
FIGURE 4.3 Sample Hierarchy 58
FIGURE 4.4 Many-to-Many Relationship 58
FIGURE 4.5 Sample Relationship Types 59
FIGURE 4.6 Many-to-One Relationship of ORDERS to CUSTOMERS 60
FIGURE 4.7 Additional One-to-Many Relationships between ORDER and LINEITEM 60
FIGURE 4.8 The Many-to-Many PRODUCT to ORDER Relationship 60
FIGURE 4.9 The One-to-Many PRODUCT to LINEITEM Relationship 61
FIGURE 4.10 The LINEITEM Junction Allows Removal of the ORDER to PRODUCT Many-to-Many Relationship 61
FIGURE 4.11 Relational Map After Introducing the INVOICE Entity 62
FIGURE 4.12 Relational Map Adding the INVLINE Junction 63
FIGURE 4.13 Final Relational Map After Adding the BACKORDER Entity 64
FIGURE 4.14 Ratio of Entities in a Relation 64
FIGURE 4.15 Relational Map with Frequency Ratios 65
FIGURE 4.16 Stage 2—Map Relationships Data Flow Diagram 66
FIGURE 5.1 Usage Map for Transaction A 69
FIGURE 5.2 Usage Map for Transaction B 70
FIGURE 5.3 Usage Map for Transaction C 70
FIGURE 5.4 Usage Map for Transaction D 71

FIGURE 5.5 Usage Map for Transaction E 71
FIGURE 5.6 Usage Map for Transaction F 72
FIGURE 5.7 Usage Map for Transaction G 72
FIGURE 5.8 Usage Map for Transaction H 72
FIGURE 5.9 Usage Map for Transaction I 73
FIGURE 5.10 Sample Combined Transaction Map 73
FIGURE 5.11 Composite Overlay Map 75
FIGURE 5.12 Complete Logical Structure 76
FIGURE 5.13 Stage 3—Transaction Analysis Data Flow Diagram 77
FIGURE 6.1 Transaction Load Calculation Form 81
FIGURE 6.2 Transaction: Create an Order 82
FIGURE 6.3 Transaction: Create an Invoice 84
FIGURE 6.4 Transaction: Backorder Update 85
FIGURE 6.5 Transaction: Customer Backorder Inquiry 86
FIGURE 6.6 Transaction: Customer Order Inquiry 86
FIGURE 6.7 Transaction: Product Invoice Inquiry 87
FIGURE 6.8 Sample Load Matrix Form 90
FIGURE 6.9 E-Z Chair Load Matrix 92
FIGURE 6.10 E-Z Chair Loaded Data Model 93
FIGURE 6.11 Finding the Stress Period 94
FIGURE 6.12 Alternative Path Creation 94
FIGURE 6.13 Stage 4—Calculate Total Load—Data Flow Diagram 96
FIGURE 7.1 Simple Data Structures 103
FIGURE 7.2 Indexed Data Structures 104
FIGURE 7.3 Chain Structures 106
FIGURE 7.4 Complex Structures 108
FIGURE 7.5 Example of TOTAL Structure and Linkage Path 110
FIGURE 7.6 Typical Information Management System Hierarchy 112
FIGURE 7.7 Sample CODASYL Network and Subschema 114
FIGURE 7.8 Typical Inverted List Architecture 116

FIGURE 7.9 Relational DBMS Structure 118
FIGURE 8.1 Stages of Data Base Design 123
FIGURE 8.2 Sample E-Z Chair Loaded Composite Map with Preliminary Record Grouping 126
FIGURE 8.3 User-Views of E-Z Chair Problem 130, 131
FIGURE 8.4 Consolidated User-Views 132
FIGURE 8.5 Control Evaluation Table 135
FIGURE 8.6 Managing Systems Maintenance: Data Accessibility versus Storage 138
FIGURE 8.7 Components of Response Time 141
FIGURE 8.8 IDMS Integrated Data Dictionary 144
FIGURE 8.9 DB/DC Data Dictionary System of IBM 145
FIGURE 8.10 LEXICON Data Dictionary Product of Arthur Andersen & Company 146
FIGURE 8.11 Data Dictionary System of the Future 148
FIGURE 9.1 Sample E-Z Chair Corporation IDMS Schema 157
FIGURE 9.2 Sample E-Z Chair TOTAL Schema 158
FIGURE 9.3 TOTAL Subschemas 160
FIGURE 9.4 IDMS Recovery 162
FIGURE 9.5 Bachman Diagram 163
FIGURE 9.6 Example of Bachman Diagram 165
FIGURE 9.7 TOTAL Order File Example 167
FIGURE 10.1 Sample E-Z Chair IMS Schema 171
FIGURE 10.2 Sample IMS DBD Source Code 172
FIGURE 10.3 IMS Program Data Independence 173
FIGURE 10.4 Sample IMS Program Specification Block Coding 175
FIGURE 10.5 IMS Integrity Utilities 177
FIGURE 11.1 System/38 Table Arrays 184
FIGURE 11.2 System/38 Physical and Logical Files 185
FIGURE 11.3 QBE Table Skeleton 187

FIGURE 11.4 QBE Query Example 187
FIGURE 11.5 System/38 Pseudo-hierarchy 188
FIGURE 11.6 Creating a QBE Snapshot 190
FIGURE 12.1 Sample ADABAS Schema 194
FIGURE 12.2 Overview of System 2000 Physical File Organization 196
FIGURE 12.3 ADAMINT DML 197
FIGURE 12.4 ADAMINT Logical Record 198
FIGURE 12.5 ADABAS Indexing Search for Descriptor "Smith" 201

试读结束：需要全本请在线购买： www.ertongbook.com