

---

# **Systems Design in a Database Environment**

Dr. Kenneth S. Brathwaite  
AKI Group, Inc.



---

# **Systems Design in a Database Environment**

**Dr. Kenneth S. Brathwaite**  
AKI Group, Inc.

**Intertext Publications**  
**McGraw-Hill Book Company**

New York St. Louis San Francisco Auckland Bogotá  
Hamburg London Madrid Mexico Milan Montreal  
New Delhi Panama Paris São Paulo  
Singapore Sidney Tokyo Toronto

**Library of Congress Catalog Card Number 89-83767**

**Copyright © 1989 by Multiscience Press, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this book may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system without the prior written permission of the publisher.**

**10 9 8 7 6 5 4 3 2 1**

**ISBN 0-07-007250-7**

**Intertext Publications/Multiscience Press, Inc.  
One Lincoln Plaza  
New York, NY 10023**

**McGraw-Hill Book Company  
1221 Avenue of the Americas  
New York, NY 10020**

**Composed in Ventura Publisher by Context, Inc.**

---

# Preface

The work reported in this book is the result of research conducted in systems design in the database environment since 1979 and active participation in designing systems for several large corporations since 1984.

The main objective of the book is to provide readers with definitive information on topics that are crucial to the understanding of what is required to develop a system from its conceptual stage through analysis, design, testing, implementation, and performance monitoring. In this volume, I provide readers with information on traditional topics such as data analysis, user requirement analysis and user survey development, design methodologies, logical database design, and development of security controls.

It addresses topics of interest to data administrators, database administrators, systems programmers, systems analysts, application programmers, information processing managers, data processing managers, and database users. The book can be used effectively by practitioners in business as well as in government, and also for an introductory course in a business or technical school. The readers need no preparation beyond the application programmer's level, but would benefit more if some experience were obtained in database environments, VSAM and IMS in particular.

As the title suggests, the volume provides information on topics that are interrelated and interdependent. Thus the chapter on data analysis techniques includes information on data analysis, logical database design, physical database design, and development of controls. This chapter is meant to be a synopsis of the whole book with each subsequent chapter giving increasingly more detailed coverage of the various phases of systems design.

Chapter 1 introduces various data analysis techniques and shows how these techniques are used in the development of database systems. Chapter 2 discusses the place of user requirements in system design and outlines the survey method of collecting and analyzing

the user requirements. Chapter 3 outlines a database design methodology that was developed, tested, and implemented by a major public utility.

Chapter 4, on data models and entity-relationship diagrams, discusses the role these play in database design and shows how they are integral parts of data analysis and are the major input to the physical design phase. Chapter 5, on the normalization process, discusses normalization in a mathematically rigorous way and may be omitted by most readers. Chapter 6 discusses logical and physical database design. It is during these phases that the user requirements are first developed into physical models and stored in database files. The chapter on design of specific systems and databases shows how one can apply the techniques of database design developed in previous chapters to database management systems (DBMS) such as IMS and DB2.

Chapters 8, 9, 10, and 11 of the volume discuss topics that, although not part of the design process, very often determine the success or failure of database design. Chapter 8 discusses the role of data dictionaries in system design. Chapter 9 discusses the use of CASE tools as an aid to systems design. Chapter 10 is on data security in a database environment. Chapter 11 indicates how security controls can be built into database systems.

The case studies discussed in Part Two are illustrated for the express purpose of strengthening the discussion on user requirement analysis and the selection of CASE tools.

Finally, as with any large undertaking, errors may remain in this volume. I have worked diligently to maintain the accuracy of this work. However, if errors remain, I solicit your assistance in bringing them to my attention. I, in turn, will do my utmost to ensure that they are corrected.

## ACKNOWLEDGMENTS

I am grateful for the comments and suggestions I received from Vic Howard, Stanley Locbe, Francis Chin, and Jay Louise Weldon. The initial draft of this manuscript was ably typed by Jane Cuffy and Andrea Drayton. Their efforts are appreciated.

*Ken S. Brathwaite  
Plainfield, New Jersey  
January, 1989*

---

# Contents

<b>Preface</b>	<b>xv</b>
<b>Chapter A Database Design</b>	
<b>Chapter 1 Data Analysis Techniques</b>	<b>1</b>
1.1 Introductory Remarks	3
1.2 Definitions of Some Data Analysis Terms	3
1.3 Data as a Resource	5
1.4 Management Control of the Data Resource	5
1.5 Data Ownership Philosophies	5
1.6 Different Views of Data	6
1.7 Data Analysis	6
1.7.1 Analysis Using Different Models	10
1.8 Functional Analysis	12
1.8.1 Functional Analysis Example	13
1.9 Data Analysis Documentation	14
1.9.1 Data Analysis Documentation Examples	14
1.10 The Entity Model	15
1.10.1 Translation of User View to an Entity Model	19
1.10.2 Selection and Identification of Entities	20
1.11 Database Design — Schema Development	21
1.11.1 Formulating the DBMS Specific Logical Database Schema	22
1.12 Documenting the Database Design	23
1.13 The Role of Data Dictionary/Directory Systems	24
1.13.1 Features and Functions of DD/D Systems	24
1.14 The Normalization Process	25
1.15 Normalization	25
Questions	26
<b>Chapter 2 User Requirement Analysis</b>	<b>27</b>
2.1 Introductory Remarks	27

2.2	Data Collection Methods	27
2.3	The Survey Method	28
2.4	An Example of User Survey	29
2.5	Analysis of User Responses	37
2.6	Deliverables from Analysis	42
2.7	Discussion of Some Deliverables	42
2.8	Remarks in Conclusion	42
	Questions	44
<b>Chapter 3</b>	<b>Database Design Methodology</b>	<b>45</b>
3.1	Review of Existing Methodologies	46
3.2	Detailed Discussion of Database Design	46
3.2.1	Inputs to Design Process	47
3.2.2	The Entity-Relationship (E-R) Approach	49
3.2.3	Definitions and Terminologies	49
3.3	The Data Analysis Phase	50
3.4	Conducting the Data Analysis Phase	51
3.5	The Entity Model	52
3.6	Approaches to Entity Model	53
3.7	Stages of Integration of Entity Models	53
3.8	Entity Modeling Case Study	54
3.9	Application of Revisions to Entity Model	55
3.10	Deriving Entity Models from Traditional or Flat Files	56
3.11	Superimposition of Entity Models	57
3.12	Clustering of Entity Classes	58
3.13	Application View and Logical Schema Design	59
3.14	Logical Schema — Case Study	60
3.15	Translation of the Logical Schema into Physical Databases	60
3.16	Summary	62
	Questions	62
<b>Chapter 4</b>	<b>Data Models and Entity Relationship Diagrams</b>	<b>63</b>
4.1	Introductory Remarks	65
4.2	Some Defining Terms	65
4.3	Entity And Entity Classes	66
4.4	Super Entities And Entity Subtypes	66
4.5	Types Of Relationships	67
4.5.1	One-to-One Relationship	67
4.5.2	One-to-Many Relationship	67
4.5.3	Many-to-Many Relationship	68
4.5.4	Mutually Exclusive Relationship	68

4.5.5	Mutually Inclusive Relationship	68
4.5.6	Mandatory Relationship	69
4.5.7	Optional Relationship	69
4.6	Translation of E-R Diagrams to Logical Models	70
4.6.1	Identification of Business Entities	70
4.6.2	Determination of E-R Diagram for Problems	71
4.6.3	Conversion of the E-R Diagram to a Logical Data Model	71
4.6.4	Conversion of Many-to-Many Relationship	72
4.6.5	Handling of Repeating Groups	73
4.6.6	Translataion of Data Models to Logical Schemas	74
4.7	Overview of DBMS Structures	74
4.8	The Relational Data Model	75
4.8.1	Advantages of a Relational Data Model	76
4.8.2	Disadvantages of a Relational Data Model	77
4.9	The Hierarchical Data Model	77
4.9.1	Representation of Relationships	78
4.9.2	Storage Operations with a Hierarchical Data Model	78
4.9.3	Advantages of a Hierarchical Data Model	79
4.9.4	Disadvantages of a Hierarchical Data Model	79
4.10	The Network Data Model	79
4.10.1	Advantages of the Network Data Model	80
4.10.2	Disadvantages of the Network Data Model	80
4.11	Concluding Remarks	81
	Questions	81

## Chapter 5      The Normalization Process      83

5.1	Introductory Remarks	83
5.2	Normalization	83
5.3	First Normal Form	83
5.4	Second Normal Form	84
5.5	Update Problems With Unnormalized Relations	85
5.6	Third Normal Form	86
5.7	Normalization In The Relational Data Model	86
5.7.1	Basic Concepts of the Relational Model	87
5.7.2	First Normal Form (1NF)	89
5.7.3	Functional Dependency (FD)	91
5.8	Partial Dependency	93
5.8.1	Second Normal Form (2NF)	93
5.8.2	Another Harmful FD Type	94



## **X Systems Design in a Database Environment**

5.8.3	Transitive Dependencies	94
5.9	Relational Data Model (3NF)	94
5.10	Boyce-Codd Normal Form (BCNF)	95
5.10.1	Redundancy Reduction Due to BCNF	95
5.10.2	Decomposing a Relation into a Collection of BCNF Relations	95
5.10.3	Natural Joins	96
5.10.4	A Better Way to Represent the Information	96
5.11	Fourth Normal Form (4NF)	97
5.12	Network — Relational Comparison	97
5.12.1	Using Relational Design Methods for Network Databases	97
5.12.2	Multiple Data Model Implementations	98
5.13	Summary	98
	Questions	98
<b>Chapter 6</b>	<b>Logical and Physical Database Design</b>	<b>99</b>
6.1	Introduction	99
6.2	The Systems Development Life Cycle	99
6.2.1	The User Requirement Phase	100
6.2.2	The Logical Design Phase	101
6.2.3	The Physical Design Phase	101
6.2.4	The Evaluation Phase	102
6.2.5	Data-Related Activities During the SDLC	103
6.3	Logical Design of Databases	103
6.3.1	Mapping to a Hierarchical Data Model	103
6.3.2	Mapping to a Relational Data Model	106
6.3.3	Mapping to a Network Data Model	106
6.4	Physical Design of Databases	107
6.5	Selection of Access Methods	108
6.5.1	Hierarchical Sequential Databases	108
6.5.2	HISAM Databases	109
6.5.3	Hierarchical Direct Databases	111
6.6	Summary	113
	Questions	114
<b>Chapter 7</b>	<b>Design of Specific Systems and Databases</b>	<b>115</b>
7.1	Introductory Remarks	115
7.2	Prototyping Tools	116
7.3	Case Tools	117
7.3.1	Some Definitions of CASE	117
7.4	Database Specific Utilities	118
7.4.1	IMS Space Management Utilities	118

7.4.2	IMS Database Tuning Utilities	119
7.4.3	DB2 Database Utilities	119
7.5	The Data Dictionary	120
7.6	Space Calculations	120
7.6.1	Input to the Space Calculations	120
7.6.2	Use of Subtrees in Space Calculations	121
7.6.3	Calculating Subtree Length	122
7.6.4	Subtree Sizes Without Data Set Groups	122
7.6.5	Subtree Sizes With Data Set Groups	123
7.6.6	Basic Space Calculations Formulas	124
7.7	Time Estimates	125
7.8	Performance Tuning of IMS Databases	126
7.9	Use of Secondary Indexes	126
7.9.1	Characteristics of Secondary Indexes	127
7.10	Design Considerations of IMS	128
7.10.1	Determining the Size of Control Intervals	129
7.11	Performance Tuning of DB2 Databases	129
7.11.1	Performance Tuning Hints	129
7.12	Some Concluding Remarks	130
	Questions	131

<b>Chapter 8</b>	<b>The Data Dictionary in Systems Design</b>	<b>133</b>
8.1	Introduction	133
8.2	What Is a Data Dictionary?	134
8.3	The Concept of Metadata	134
8.4	Active Versus Passive Data Dictionaries	134
8.5	The Role of the Data Dictionary in the SDLC	135
8.5.1	Requirements Specifications	136
8.5.2	Relationship to Existing Systems	136
8.5.3	Detail Design	136
8.5.4	Generated Code	137
8.5.5	Operational Information	137
8.5.6	Final Documentation	137
8.6	Interfaces to the Data Dictionary	137
8.6.1	The End User	138
8.6.2	The Systems Analysis Phase	138
8.6.3	The Systems Design Phase	138
8.6.4	The Program Coding Phase	139
8.6.5	The Testing Phase	139
8.6.6	The Implementation Phase	139
8.6.7	The Maintenance Phase	139
8.7	The Data Dictionary as a Documentation Tool	139
8.7.1	Documenting the System Design Phase	140
8.7.2	Documenting the Detailed Design Phase	141
8.7.3	Documenting the Physical Design Phase	141

8.7.4	Documenting the Implementation Phase	142
8.7.5	Documenting the Structured Maintenance Phase	142
8.8	The Data Dictionary and Data Security	143
8.8.1	Access Management	144
8.8.2	Privacy Transformations	144
8.8.3	Cryptographic Controls	145
8.8.4	Security Kernels	145
8.9	Data Dictionary Standards	145
8.9.1	Standard Formats for Data Dictionary Entries	146
8.9.2	Standards for Programs Interfacing with a Data Dictionary	146
8.9.3	Security Standards	146
8.10	Examples of Data Dictionary Entries	147
8.10.1	The Xerox Data Dictionary	147
8.10.2	MSP Datamanager	150
8.10.3	The IBM DB/DC Data Dictionary	152
8.11	Summary	155
	Questions	156
<b>Chapter 9</b>	<b>Systems Design Using Case Tools</b>	<b>157</b>
9.1	Several Definitions Of Case	157
9.2	Categories Of Case Tools	158
9.2.1	Well-Equipped Toolkits	159
9.2.2	Structured Diagramming Tools	159
9.2.3	Prototyping Tools	159
9.2.4	The CASE Repository	160
9.2.5	Data Design Toolkits	160
9.2.6	Programming Toolkits	160
9.2.7	Maintenance Toolkits	161
9.2.8	Project Management Toolkits	161
9.3	Demonstrated Use of Case Tools in the SDLC	161
9.3.1	CASE in the Planning Environment	161
9.3.2	CASE in the Systems Design Environment	162
9.3.3	CASE in the Systems Development Environment	163
9.3.4	Samples of Deliverables from CASE Tools	163
9.4	Selection Criteria for Case Tools	166
9.4.1	Analysis of Selection Criteria	170
9.5	Vendors of CASE Tools	171
9.5.1	Vendor List	171
9.6	Getting CASE in Place	179
9.6.1	Determine the Methodology	179

9.6.2	Select a CASE Product	180
9.6.3	Implement the CASE Product	180
9.7	Remarks in Conclusion	181
	Questions	181
<b>Chapter 10</b>	<b>Data Security in a Database Environment</b>	<b>183</b>
10.1	Conducting a Threat Analysis	185
10.1.1	Threat Analysis Case Study	186
10.1.2	Analysis of Results	187
10.1.3	Conducting a Risk Assessment	189
10.1.4	Achieving Database Privacy	194
10.1.5	Access Management	194
10.1.6	Privacy Transformations	194
10.1.7	Cryptographic Controls and Data Transformation	195
10.1.8	Database Integrity	195
	Questions	196
<b>Chapter 11</b>	<b>Development of Security Controls</b>	<b>197</b>
11.1	Top-Level Management Involvement in Database Security, Privacy, and Integrity	198
11.2	Administrative Controls	200
11.3	Auditing and Monitoring the Database	201
11.4	Types of Protection Mechanism	203
11.4.1	The Access Matrix as a Protection Mechanism	203
11.4.2	Rules Governing Accessing Decisions	206
11.4.3	Protection Levels of Access Matrix	207
11.4.4	The Access Matrix Monitor	208
11.4.5	Administrative Controls in Access Matrix Environment	209
11.4.6	The Software Version of the Access Matrix Monitor	209
11.4.7	Design of the Monitor	210
11.4.8	Functions of Monitor	210
11.4.9	Design of Subject and Object Directories	210
11.5	Subjects and Objects of the Access Matrix	211
11.5.1	Processing Details of Access Matrix Model	211
11.5.2	Processing Alternatives of Access Matrix Model	212
11.6	The Risk Assessment as an Aid to Selecting a Protection Mechanism	213

11.7	Resource Access Control Facility (RACF) as a Protection Mechanism	214
11.7.1	RACF Concepts	214
11.8	Summary	216
	Questions	216
<b>Part B</b>	<b>Case Histories and CASE Tools</b>	<b>217</b>
<b>Appendix A</b>	<b>Systems Development Case Study — Marketing</b>	<b>219</b>
1.0	Introduction	219
2.0	Program Narrative	220
3.0	System Flow Chart	222
4.0	Description of Input	223
5.0	Description of Output	223
6.0	Detailed Processing Requirements	223
7.0	Edit/Valuation Rules	226
8.0	Description of Subprograms and Utilities	227
9.0	Record Layouts	227
10.0	Spool Database	227
<b>Appendix B</b>	<b>A Guide to Selecting Case Tools</b>	<b>259</b>
B.1	AMS' Response to Questionnaire: Selection Criteria for CASE Tools	259
B.2	Index Technology's Response to Questionnaire: Selection Criteria for CASE Tools	264
B.4	Designaid's Response to Questionnaire: Selection Criteria for CASE Tools	277
B.5	Optima, Inc — Response to Questionnaire: Selection Criteria for CASE Tools for Health Insurance Plan	279
B.6	APS Development Center — Response to Questionnaire: Selection Criteria for CASE Tools	285
B.7	Maestro's Response to Questionnaire: Selection Criteria for CASE Tools	288
B.8	Response from the Information Engineering Facility (IEF) to Questionnaire: Selection Criteria for CASE Tools	290
	<b>Bibliography</b>	<b>297</b>
	<b>Index</b>	<b>301</b>

---

**Part**

**A**

# **Database Design**



# **Data Analysis Techniques**

## **1.1 INTRODUCTORY REMARKS**

This chapter discusses a structured method of conducting the data analysis phase of database design: data analysis development life cycle. It is during this phase that formulations of the user requirements are developed.

The user requirements form the basis of the system design. Therefore, the success or failure in obtaining adequate information about the user requirements determines the success or failure of the entire system.

The structured methodology discussed in this chapter begins by defining data as a corporate resource which must be planned for like any other corporate resource. It details how the users' needs for processing that data are collected, analyzed, and formulated into data models, showing the relationship among the various entities or collections of data elements.

The Chapter concludes by discussing the normalization of these data models and the translation of these models into physical databases.

## **1.2 DEFINITIONS OF SOME DATA ANALYSIS TERMS**

Some of the terms used in data analysis are transplanted from the old approaches of system analysis. Other terms, like attribute, entity, and relationship, are new and resulted from efforts to adopt



#### 4 Systems Design in a Database Environment

structured methods of designing systems with the data as the driving force rather than the processing of that data. The following table gives some definitions of these new terms.

**Table 1.1 Definitions and Terminologies**

ITEM	DEFINITION
DATA ITEM	The smallest unit of named data. A data item is often referred to as a field or data element.
DATA	The values taken by various data items are called data. For example, the value of the data element, Customer Name, is data.
INFORMATION	Data that is processed, accessed, and assimilated or used in the decision-making process. Information results from the analysis and synthesis of data.
ENTITY (OR ENTITY CLASS OR ENTITY TYPE)	A fundamental thing of interest to an organization. An entity may be a person, place, thing, concept, or event, that is real or abstract; it should have a unique identifier.
ATTRIBUTE	A descriptive value of property associated with an individual entity.
RELATIONSHIP	An association between two or more entities.
ENTITY MODEL (SNAPSHOT)	A diagrammatical representation of the relationships between the entity classes. The representation allows us to include only those entities that are required to solve the user's data processing problem. It depends on the relationships you are concerned with.
LOGICAL SCHEMA (EXTERNAL STRUCTURE/SCHEMA)	The mapping of the entity model into the constructs or constraints of the database management system (DBMS). In general, the logical schema indicates how the model will be stored and accessed.
DATA ANALYSIS	The determination of the fundamental data resources of an organization. It deals with the collection of the basic entities and the relationships between those entities.