
DATA MODELING FUNDAMENTALS

A Practical Guide for
IT Professionals

Paulraj Ponniah



WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

CONTENTS

PREFACE	xvii
ACKNOWLEDGMENTS	xxi
I INTRODUCTION TO DATA MODELING	1
1 Data Modeling: An Overview	3
Chapter Objectives / 3	
Data Model Defined / 4	
What Is a Data Model? / 5	
Why Data Modeling? / 6	
Who Performs Data Modeling? / 9	
Information Levels / 10	
Classification of Information Levels / 11	
Data Models at Information Levels / 13	
Conceptual Data Modeling / 17	
Data Model Components / 18	
Data Modeling Steps / 20	
Data Model Quality / 26	
Significance of Data Model Quality / 27	
Data Model Characteristics / 27	
Ensuring Data Model Quality / 28	
Data System Development / 29	
Data System Development Life Cycle / 29	
Roles and Responsibilities / 33	
Modeling the Information Requirements / 33	
Applying Agile Modeling Principles / 34	
Data Modeling Approaches and Trends / 35	
Data Modeling Approaches / 36	
Modeling for Data Warehouse / 38	

Other Modeling Trends / 39
Chapter Summary / 41
Review Questions / 41

2 Methods, Techniques, and Symbols 43

Chapter Objectives / 43
Data Modeling Approaches / 44
 Semantic Modeling / 44
 Relational Modeling / 45
 Entity-Relationship Modeling / 46
 Binary Modeling / 46
Methods and Techniques / 47
 Peter Chen (E-R) Modeling / 48
 Information Engineering / 50
 Integration Definition for Information Modeling / 51
 Richard Barker's Model / 53
 Object-Role Modeling / 55
 eXtensible Markup Language / 57
 Summary and Comments / 60
Unified Modeling Language / 61
 Data Modeling Using UML / 61
 UML in the Development Process / 64
Chapter Summary / 68
Review Questions / 68

II DATA MODELING FUNDAMENTALS 71

3 Anatomy of a Data Model 73

Chapter Objectives / 73
Data Model Composition / 74
 Models at Different Levels / 74
 Conceptual Model: Review Procedure / 76
 Conceptual Model: Identifying Components / 77
Case Study / 81
 Description / 81
 E-R Model / 84
 UML Model / 87
Creation of Models / 89
 User Views / 90
 View Integration / 92
 Entity Types / 96
 Specialization/Generalization / 98
 Relationships / 98
 Attributes / 100

Identifiers /	101
Review of the Model Diagram /	103
Logical Model: Overview /	104
Model Components /	104
Transformation Steps /	107
Relational Model /	109
Physical Model: Overview /	111
Model Components /	111
Transformation Steps /	112
Chapter Summary /	113
Review Questions /	113

4 Objects or Entities in Detail

115

Chapter Objectives /	115
Entity Types or Object Sets /	116
Comprehensive Definition /	116
Identifying Entity Types /	120
Homonyms and Synonyms /	125
Category of Entity Types /	127
Exploring Dependencies /	130
Dependent or Weak Entity Types /	131
Classifying Dependencies /	132
Representation in the Model /	133
Generalization and Specialization /	134
Why Generalize or Specialize? /	136
Supertypes and Subtypes /	137
Generalization Hierarchy /	138
Inheritance of Attributes /	140
Inheritance of Relationships /	140
Constraints /	141
Rules Summarized /	144
Special Cases and Exceptions /	144
Recursive Structures /	145
Conceptual and Physical /	145
Assembly Structures /	147
Entity Type Versus Attribute /	148
Entity Type Versus Relationship /	148
Modeling Time Dimension /	149
Categorization /	150
Entity Validation Checklist /	153
Completeness /	153
Correctness /	154
Chapter Summary /	155
Review Questions /	155

5 Attributes and Identifiers in Detail **157**

- Chapter Objectives / 157
- Attributes / 158
 - Properties or Characteristics / 158
 - Attributes as Data / 161
 - Attribute Values / 162
 - Names and Descriptions / 163
- Attribute Domains / 164
 - Definition of a Domain / 164
 - Domain Information / 165
 - Attribute Values and Domains / 166
 - Split Domains / 167
 - Misrepresented Domains / 167
 - Resolution of Mixed Domains / 168
- Constraints for Attributes / 169
 - Value Set / 169
 - Range / 170
 - Type / 170
 - Null Values / 170
- Types of Attributes / 171
 - Single-Valued and Multivalued Attributes / 171
 - Simple and Composite Attributes / 171
 - Attributes with Stored and Derived Values / 172
 - Optional Attributes / 173
- Identifiers or Keys / 175
 - Need for Identifiers / 175
 - Definitions of Keys / 175
 - Guidelines for Identifiers / 176
 - Key in Generalization Hierarchy / 177
- Attribute Validation Checklist / 178
 - Completeness / 178
 - Correctness / 179
- Chapter Summary / 180
- Review Questions / 180

6 Relationships in Detail **183**

- Chapter Objectives / 183
- Relationships / 184
 - Associations / 184
 - Relationship: Two-Sided / 186
 - Relationship Sets / 187
 - Double Relationships / 187
 - Relationship Attributes / 189
- Degree of Relationships / 190
 - Unary Relationship / 191
 - Binary Relationship / 191

Ternary Relationship /	193
Quaternary Relationship /	193
Structural Constraints /	194
Cardinality Constraint /	195
Participation Constraint /	198
Dependencies /	200
Entity Existence /	200
Relationship Types /	201
Identifying Relationship /	202
Nonidentifying Relationship /	204
Maximum and Minimum Cardinalities /	204
Mandatory Conditions: Both Ends /	206
Optional Condition: One End /	206
Optional Condition: Other End /	207
Optional Conditions: Both Ends /	208
Special Cases /	209
Gerund /	209
Aggregation /	210
Access Pathways /	211
Design Issues /	215
Relationship or Entity Type? /	215
Ternary Relationship or Aggregation? /	216
Binary or <i>N</i> -ary Relationship? /	216
One-to-One Relationships /	217
One-to-Many Relationships /	219
Circular Structures /	219
Redundant Relationships /	221
Multiple Relationships /	221
Relationship Validation Checklist /	222
Completeness /	223
Correctness /	224
Chapter Summary /	225
Review Questions /	225

III DATA MODEL IMPLEMENTATION 227

7 Data Modeling to Database Design 229

Chapter Objectives /	229
Relational Model: Fundamentals /	231
Basic Concepts /	231
Structure and Components /	233
Data Integrity Constraints /	238
Transition to Database Design /	242
Design Approaches /	243

- Conceptual to Relational Model / 243
- Traditional Method / 244
- Evaluation of Design Methods / 245
- Model Transformation Method / 246
 - The Approach / 246
 - Mapping of Components / 249
 - Entity Types to Relations / 250
 - Attributes to Columns / 250
 - Identifiers to Keys / 252
 - Transformation of Relationships / 252
 - Transformation Summary / 267
- Chapter Summary / 269
- Review Questions / 269

8 Data Normalization **271**

- Chapter Objectives / 271
- Informal Design / 272
 - Forming Relations from Requirements / 272
 - Potential Problems / 273
 - Update Anomaly / 275
 - Deletion Anomaly / 275
 - Addition Anomaly / 276
- Normalization Methodology / 276
 - Strengths of the Method / 277
 - Application of the Method / 277
 - Normalization Steps / 277
- Fundamental Normal Forms / 278
 - First Normal Form / 278
 - Second Normal Form / 279
 - Third Normal Form / 281
 - Boyce-Codd Normal Form / 284
- Higher Normal Forms / 285
 - Fourth Normal Form / 286
 - Fifth Normal Form / 287
 - Domain-Key Normal Form / 288
- Normalization Summary / 290
 - Review of the Steps / 290
 - Normalization as Verification / 291
- Chapter Summary / 292
- Review Questions / 292

9 Modeling for Decision-Support Systems **295**

- Chapter Objectives / 295
- Decision-Support Systems / 296
 - Need for Strategic Information / 296

History of Decision-Support Systems /	297
Operational Versus Informational Systems /	299
System Types and Modeling Methods /	299
Data Warehouse /	301
Data Warehouse Defined /	301
Major Components /	302
Data Warehousing Applications /	305
Modeling: Special Requirements /	305
Dimensional Modeling /	308
Dimensional Modeling Basics /	309
STAR Schema /	312
Snowflake Schema /	318
Families of STARS /	321
Transition to Logical Model /	322
OLAP Systems /	325
Features and Functions of OLAP /	325
Dimensional Analysis /	326
Hypercubes /	328
OLAP Implementation Approaches /	330
Data Modeling for OLAP /	332
Data Mining Systems /	334
Basic Concepts /	334
Data Mining Techniques /	338
Data Preparation and Modeling /	339
Data Preprocessing /	339
Data Modeling /	341
Chapter Summary /	342
Review Questions /	343

IV PRACTICAL APPROACH TO DATA MODELING 345

10 Ensuring Quality in the Data Model 347

Chapter Objectives /	347
Significance of Quality /	348
Why Emphasize Quality? /	348
Good and Bad Models /	349
Approach to Good Modeling /	351
Quality of Definitions /	351
Importance of Definitions /	352
Aspects of Quality Definitions /	353
Correctness /	353
Completeness /	354
Clearness /	357
Format /	358

Checklists / 358	
High-Quality Data Model / 360	
Meaning of Data Model Quality / 360	
Quality Dimensions / 361	
What Is a High-Quality Model? / 363	
Benefits of High-Quality Models / 364	
Quality Assurance Process / 365	
Aspects of Quality Assurance / 365	
Stages of Quality Assurance Process / 366	
Data Model Review / 369	
Data Model Assessment / 370	
Chapter Summary / 373	
Review Questions / 373	
11 Agile Data Modeling in Practice	375
Chapter Objectives / 375	
The Agile Movement / 376	
How It Got Started / 377	
Principles of Agile Development / 378	
Philosophies / 378	
Generalizing Specialists / 379	
Agile Modeling / 379	
What Is Agile Modeling? / 380	
Basic Principles / 380	
Auxiliary Principles / 381	
Practicing Agile Modeling / 381	
Primary Practices / 381	
Additional Practices / 382	
Role of Agile DBA / 383	
Agile Documentation / 383	
Recognizing an Agile Model / 384	
Feasibility / 384	
Evolutionary Data Modeling / 385	
Traditional Approach / 385	
Need for Flexibility / 386	
Nature of Evolutionary Modeling / 386	
Benefits / 387	
Chapter Summary / 388	
Review Questions / 388	
12 Data Modeling: Practical Tips	391
Chapter Objectives / 391	
Tips and Suggestions / 392	
Nature of Tips / 392	
How Specified / 392	
How to Use Them / 392	

Requirements Definition / 393	
Interviews / 393	
Group Sessions / 394	
Geographically Dispersed Groups / 394	
Documentation / 395	
Change Management / 395	
Notes for Modeling / 396	
Stakeholder Participation / 396	
Organizing Participation / 397	
User Liaison / 397	
Continuous Interaction / 398	
Multiple Sites / 399	
Iterative Modeling / 399	
Establishing Cycles / 399	
Determining Increments / 400	
Requirements: Model Interface / 400	
Integration of Partial Models / 401	
Special Cases / 401	
Legal Entities / 402	
Locations and Places / 403	
Time Periods / 405	
Persons / 407	
Bill-of-Materials / 409	
Conceptual Model Layout / 409	
Readability and Usability / 409	
Component Arrangement / 410	
Adding Texts / 416	
Visual Highlights / 417	
Logical Data Model / 417	
Enhancement Motivation / 418	
Easier Database Implementation / 418	
Performance Improvement / 418	
Storage Management / 419	
Enhanced Representation / 419	
Chapter Summary / 421	
Review Questions / 421	
Bibliography	423
Glossary	425
Index	433