



# **Data Modeling Essentials**

*This page intentionally left blank*

# Data Modeling Essentials

Third Edition

Graeme C. Simsion and Graham C. Witt



<i>Publishing Director</i>	Diane Cerra
<i>Senior Editor</i>	Lothlórén Homet
<i>Publishing Services Manager</i>	Simon Crump
<i>Project Manager</i>	Kyle Scroelen
<i>Editorial Coordinator</i>	Cyrina Demma
<i>Cover Design</i>	Dick Hannus, Hannus Design Associates
<i>Cover Image</i>	Creatas
<i>Composition</i>	Cepha Imaging Pvt. Ltd.
<i>Copyeditor</i>	Broccoli Information Management
<i>Proofreader</i>	Jacqui Brownstein
<i>Indexer</i>	Broccoli Information Management
<i>Interior printer</i>	Maple-Vail Book Manufacturing Group
<i>Cover printer</i>	Phoenix Color Corp.

Morgan Kaufmann Publishers is an imprint of Elsevier.  
500, Sansome Street, Suite 400, San Francisco, CA 94111

This book is printed on acid-free paper.

© 2005 by Elsevier Inc. All rights reserved.

Designations used by companies to distinguish their products are often claimed as trademarks or registered trademarks. In all instances in which Morgan Kaufmann Publishers is aware of a claim, the product names appear in initial capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, scanning, or otherwise—without prior written permission of the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: [permissions@elsevier.com](mailto:permissions@elsevier.com). You may also complete your request online via the Elsevier homepage (<http://elsevier.com>) by selecting "Customer Support" and then "Obtaining Permissions."

#### **Library of Congress Cataloging-in-Publication Data**

Application submitted.

ISBN: 0-12-644571-6

For information on all Morgan Kaufmann publications, visit our Web site at [www.mkp.com](http://www.mkp.com) or [www.books.elsevier.com](http://www.books.elsevier.com)

*This new edition of Data Modeling Essentials is dedicated to the memory of our friend and colleague, Robin Wade, who put the first words on paper for the original edition, and whose cartoons have illustrated many of our presentations.*



*This page intentionally left blank*

# Contents

*Preface* xxiii

- ■ ■ Part I
  - The Basics 1**
  - ■ ■ Chapter 1
    - What Is Data Modeling? 3**
    - 1.1 **Introduction 3**
    - 1.2 **A Data-Centered Perspective 3**
    - 1.3 **A Simple Example 4**
    - 1.4 **Design, Choice, and Creativity 6**
    - 1.5 **Why Is the Data Model Important? 8**
      - 1.5.1 Leverage 8
      - 1.5.2 Conciseness 9
      - 1.5.3 Data Quality 10
      - 1.5.4 Summary 10
    - 1.6 **What Makes a Good Data Model? 10**
      - 1.6.1 Completeness 10
      - 1.6.2 Nonredundancy 11
      - 1.6.3 Enforcement of Business Rules 11
      - 1.6.4 Data Reusability 11
      - 1.6.5 Stability and Flexibility 12
      - 1.6.6 Elegance 13
      - 1.6.7 Communication 14
      - 1.6.8 Integration 14
      - 1.6.9 Conflicting Objectives 15
    - 1.7 **Performance 15**



## viii ■ Contents

1.9	<b>Where Do Data Models Fit In?</b>	<b>20</b>
1.9.1	Process-Driven Approaches	20
1.9.2	Data-Driven Approaches	20
1.9.3	Parallel (Blended) Approaches	22
1.9.4	Object Oriented Approaches	22
1.9.5	Prototyping Approaches	23
1.9.6	Agile Methods	23
1.10	<b>Who Should Be Involved in Data Modeling?</b>	<b>23</b>
1.11	<b>Is Data Modeling Still Relevant?</b>	<b>24</b>
1.11.1	Costs and Benefits of Data Modeling	25
1.11.2	Data Modeling and Packaged Software	26
1.11.3	Data Integration	27
1.11.4	Data Warehouses	27
1.11.5	Personal Computing and User-Developed Systems	28
1.11.6	Data Modeling and XML	28
1.11.7	Summary	28
1.12	<b>Alternative Approaches to Data Modeling</b>	<b>29</b>
1.13	<b>Terminology</b>	<b>30</b>
1.14	<b>Where to from Here?—An Overview of Part I</b>	<b>31</b>
1.15	<b>Summary</b>	<b>32</b>
■ ■ ■	<b>Chapter 2</b>	
	<b>Basics of Sound Structure</b>	<b>33</b>
2.1	<b>Introduction</b>	<b>33</b>
2.2	<b>An Informal Example of Normalization</b>	<b>34</b>
2.3	<b>Relational Notation</b>	<b>36</b>
2.4	<b>A More Complex Example</b>	<b>37</b>
2.5	<b>Determining Columns</b>	<b>40</b>
2.5.1	One Fact per Column	40
2.5.2	Hidden Data	41
2.5.3	Derivable Data	41
2.5.4	Determining the Primary Key	41
2.6	<b>Repeating Groups and First Normal Form</b>	<b>43</b>

2.6.5	Determining the Primary Key of the New Table	46
2.6.6	First Normal Form	47
<b>2.7</b>	<b>Second and Third Normal Forms</b>	<b>47</b>
2.7.1	Problems with Tables in First Normal Form	47
2.7.2	Eliminating Redundancy	48
2.7.3	Determinants	48
2.7.4	Third Normal Form	51
<b>2.8</b>	<b>Definitions and a Few Refinements</b>	<b>53</b>
2.8.1	Determinants and Functional Dependency	53
2.8.2	Primary Keys	54
2.8.3	Candidate Keys	54
2.8.4	A More Formal Definition of Third Normal Form	55
2.8.5	Foreign Keys	55
2.8.6	Referential Integrity	56
2.8.7	Update Anomalies	57
2.8.8	Denormalization and Unnormalization	58
2.8.9	Column and Table Names	59
<b>2.9</b>	<b>Choice, Creativity, and Normalization</b>	<b>60</b>
<b>2.10</b>	<b>Terminology</b>	<b>62</b>
<b>2.11</b>	<b>Summary</b>	<b>63</b>
■■■	Chapter 3	
	<b>The Entity-Relationship Approach</b>	<b>65</b>
<b>3.1</b>	<b>Introduction</b>	<b>65</b>
<b>3.2</b>	<b>A Diagrammatic Representation</b>	<b>65</b>
3.2.1	The Basic Symbols: Boxes and Arrows	66
3.2.2	Diagrammatic Representation of Foreign Keys	67
3.2.3	Interpreting the Diagram	68
3.2.4	Optionality	69
3.2.5	Verifying the Model	70
3.2.6	Redundant Arrows	71
<b>3.3</b>	<b>The Top-Down Approach: Entity-Relationship Modeling</b>	<b>72</b>
3.3.1	Developing the Diagram Top Down	74
3.3.2	Terminology	75

<b>3.4</b>	<b>Entity Classes</b>	<b>76</b>
3.4.1	Entity Diagramming Convention	77
3.4.2	Entity Class Naming	78
3.4.3	Entity Class Definitions	80