<table>
<thead>
<tr>
<th>Chapter 1 Evolution of Decision Support Systems</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The Evolution</td>
<td>2</td>
</tr>
<tr>
<td>The Advent of DASD</td>
<td>4</td>
</tr>
<tr>
<td>PC/4GL Technology</td>
<td>4</td>
</tr>
<tr>
<td>Enter the Extract Program</td>
<td>5</td>
</tr>
<tr>
<td>The Spider Web</td>
<td>6</td>
</tr>
<tr>
<td>Problems with the Naturally Evolving Architecture</td>
<td>7</td>
</tr>
<tr>
<td>Lack of Data Credibility</td>
<td>7</td>
</tr>
<tr>
<td>Problems with Productivity</td>
<td>9</td>
</tr>
<tr>
<td>From Data to Information</td>
<td>12</td>
</tr>
<tr>
<td>A Change in Approach</td>
<td>14</td>
</tr>
<tr>
<td>The Architected Environment</td>
<td>16</td>
</tr>
<tr>
<td>Data Integration in the Architected Environment</td>
<td>18</td>
</tr>
<tr>
<td>Who is the User?</td>
<td>20</td>
</tr>
<tr>
<td>The Development Life Cycle</td>
<td>20</td>
</tr>
<tr>
<td>Patterns of Hardware Utilization</td>
<td>22</td>
</tr>
<tr>
<td>Setting the Stage for Re-engineering</td>
<td>23</td>
</tr>
<tr>
<td>Monitoring the Data Warehouse Environment</td>
<td>25</td>
</tr>
<tr>
<td>Summary</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2 The Data Warehouse Environment</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Structure of the Data Warehouse</td>
<td>33</td>
</tr>
<tr>
<td>Subject Orientation</td>
<td>34</td>
</tr>
<tr>
<td>Day 1 to Day n Phenomenon</td>
<td>39</td>
</tr>
</tbody>
</table>
Granularity 41
   The Benefits of Granularity 42
   An Example of Granularity 43
   Dual Levels of Granularity 46
Exploration and Data Mining 50
Living Sample Database 50
Partitioning as a Design Approach 53
   Partitioning of Data 53
Structuring Data in the Data Warehouse 56
Auditing and the Data Warehouse 61
Data Homogeneity and Heterogeneity 61
Purging Warehouse Data 64
Reporting and the Architected Environment 64
The Operational Window of Opportunity 65
Incorrect Data in the Data Warehouse 67
Summary 69

Chapter 3  The Data Warehouse and Design 71
Beginning with Operational Data 71
Process and Data Models and the Architected Environment 78
The Data Warehouse and Data Models 79
   The Data Warehouse Data Model 81
   The Midlevel Data Model 84
   The Physical Data Model 88
The Data Model and Iterative Development 91
Normalization and Denormalization 94
   Snapshots in the Data Warehouse 100
Metadata 102
   Managing Reference Tables in a Data Warehouse 103
Cyclicity of Data — The Wrinkle of Time 105
Complexity of Transformation and Integration 108
Triggering the Data Warehouse Record 112
   Events 112
   Components of the Snapshot 113
   Some Examples 113
Profile Records 114
Managing Volume 115
Creating Multiple Profile Records 117
Going from the Data Warehouse to the Operational Environment 117
Direct Operational Access of Data Warehouse Data 118
Indirect Access of Data Warehouse Data 119
   An Airline Commission Calculation System 119
   A Retail Personalization System 121
   Credit Scoring 123
Indirect Use of Data Warehouse Data 125
<table>
<thead>
<tr>
<th>Chapter 4</th>
<th><strong>Granularity in the Data Warehouse</strong></th>
<th>139</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw Estimates</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>Input to the Planning Process</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>Data in Overflow</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>Overflow Storage</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>What the Levels of Granularity Will Be</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>Some Feedback Loop Techniques</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Levels of Granularity — Banking Environment</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>Feeding the Data Marts</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5</th>
<th><strong>The Data Warehouse and Technology</strong></th>
<th>159</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Managing Large Amounts of Data</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>Managing Multiple Media</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>Indexing and Monitoring Data</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Interfaces to Many Technologies</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Programmer or Designer Control of Data Placement</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Parallel Storage and Management of Data</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>Metadata Management</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>Language Interface</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>Efficient Loading of Data</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td>Efficient Index Utilization</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>Compaction of Data</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>Compound Keys</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>Variable-Length Data</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>Lock Management</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Index-Only Processing</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Fast Restore</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Other Technological Features</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>DBMS Types and the Data Warehouse</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>Changing DBMS Technology</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td>Multidimensional DBMS and the Data Warehouse</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>Data Warehousing across Multiple Storage Media</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td>The Role of Metadata in the Data Warehouse Environment</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td>Context and Content</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>Three Types of Contextual Information</td>
<td>186</td>
</tr>
<tr>
<td></td>
<td>Capturing and Managing Contextual Information</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Looking at the Past</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Refreshing the Data Warehouse</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>191</td>
</tr>
</tbody>
</table>
## Chapter 6  The Distributed Data Warehouse  193

- Types of Distributed Data Warehouses  193
  - Local and Global Data Warehouses  194
    - The Local Data Warehouse  197
    - The Global Data Warehouse  198
    - Intersection of Global and Local Data  201
  - Redundancy  206
  - Access of Local and Global Data  207
  - The Technologically Distributed Data Warehouse  211
  - The Independently Evolving Distributed Data Warehouse  213
- The Nature of the Development Efforts  213
  - Completely Unrelated Warehouses  215
  - Distributed Data Warehouse Development  217
    - Coordinating Development across Distributed Locations  218
    - The Corporate Data Model — Distributed  219
    - Metadata in the Distributed Warehouse  223
- Building the Warehouse on Multiple Levels  223
- Multiple Groups Building the Current Level of Detail  226
  - Different Requirements at Different Levels  228
  - Other Types of Detailed Data  232
  - Metadata  234
- Multiple Platforms for Common Detail Data  235
- Summary  236

## Chapter 7  Executive Information Systems and the Data Warehouse  239

- EIS — The Promise  240
- A Simple Example  240
- Drill-Down Analysis  243
- Supporting the Drill-Down Process  245
- The Data Warehouse as a Basis for EIS  247
- Where to Turn  248
- Event Mapping  251
- Detailed Data and EIS  253
- Keeping Only Summary Data in the EIS  254
- Summary  255

## Chapter 8  External Data and the Data Warehouse  257

- External Data in the Data Warehouse  260
- Metadata and External Data  261
- Storing External Data  263
- Different Components of External Data  264
- Modeling and External Data  265
- Secondary Reports  266
- Archiving External Data  267
- Comparing Internal Data to External Data  267
- Summary  268
Chapter 9  Migration to the Architected Environment 269
   A Migration Plan 270
   The Feedback Loop 278
   Strategic Considerations 280
   Methodology and Migration 283
   A Data-Driven Development Methodology 283
      Data-Driven Methodology 286
      System Development Life Cycles 286
      A Philosophical Observation 286
   Summary 287

Chapter 10  The Data Warehouse and the Web 289
   Supporting the eBusiness Environment 299
   Moving Data from the Web to the Data Warehouse 300
   Moving Data from the Data Warehouse to the Web 301
   Web Support 302
   Summary 302

Chapter 11  Unstructured Data and the Data Warehouse 305
   Integrating the Two Worlds 307
      Text — The Common Link 308
      A Fundamental Mismatch 310
      Matching Text across the Environments 310
      A Probabilistic Match 311
      Matching All the Information 312
      A Themed Match 313
         Industrially Recognized Themes 313
         Naturally Occurring Themes 316
         Linkage through Themes and Themed Words 317
         Linkage through Abstraction and Metadata 318
   A Two-Tiered Data Warehouse 320
      Dividing the Unstructured Data Warehouse 321
      Documents in the Unstructured Data Warehouse 322
      Visualizing Unstructured Data 323
   A Self-Organizing Map (SOM) 324
      The Unstructured Data Warehouse 325
      Volumes of Data and the Unstructured Data Warehouse 326
   Fitting the Two Environments Together 327
   Summary 330

Chapter 12  The Really Large Data Warehouse 331
   Why the Rapid Growth? 332
   The Impact of Large Volumes of Data 333
      Basic Data-Management Activities 334
      The Cost of Storage 335
      The Real Costs of Storage 336
   The Usage Pattern of Data in the Face of Large Volumes 336
Chapter 13 The Relational and the Multidimensional Models as a Basis for Database Design

The Relational Model

The Multidimensional Model

Snowflake Structures

Differences between the Models

The Roots of the Differences

Reshaping Relational Data

Indirect Access and Direct Access of Data

Servicing Future Unknown Needs

Servicing the Need to Change Gracefully

Independent Data Marts

Building Independent Data Marts

Summary

Chapter 14 Data Warehouse Advanced Topics

End-User Requirements and the Data Warehouse

The Data Warehouse and the Data Model

The Relational Foundation

The Data Warehouse and Statistical Processing

Resource Contention in the Data Warehouse

The Exploration Warehouse

The Data Mining Warehouse

Freezing the Exploration Warehouse

External Data and the Exploration Warehouse

Data Marts and Data Warehouses in the Same Processor

The Life Cycle of Data

Mapping the Life Cycle to the Data Warehouse Environment

Testing and the Data Warehouse