Contents

Preface xxxi

How to Use This Book xxxiii

Acknowledgments xxxv

About the Author xxxvii

Chapter 1 What Is a DBA? 1

Why Learn Database Administration? 3

A Unique Vantage Point 4

DBA Salaries 4

Database Technology 6

The Management Discipline of Database Administration 9

A Day in the Life of a DBA 12

Evaluating a DBA Job Offer 14
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database, Data, and System Administration</td>
<td>15</td>
</tr>
<tr>
<td>Data Administration</td>
<td>15</td>
</tr>
<tr>
<td>Database Administration</td>
<td>19</td>
</tr>
<tr>
<td>System Administration</td>
<td>20</td>
</tr>
<tr>
<td>DBA Tasks</td>
<td>20</td>
</tr>
<tr>
<td>Database Design</td>
<td>21</td>
</tr>
<tr>
<td>Performance Monitoring and Tuning</td>
<td>22</td>
</tr>
<tr>
<td>Ensuring Availability</td>
<td>24</td>
</tr>
<tr>
<td>Database Security and Authorization</td>
<td>24</td>
</tr>
<tr>
<td>Governance and Regulatory Compliance</td>
<td>26</td>
</tr>
<tr>
<td>Backup and Recovery</td>
<td>26</td>
</tr>
<tr>
<td>Ensuring Data Integrity</td>
<td>27</td>
</tr>
<tr>
<td>DBMS Release Migration</td>
<td>29</td>
</tr>
<tr>
<td>Jack-of-All-Trades</td>
<td>29</td>
</tr>
<tr>
<td>The Types of DBAs</td>
<td>31</td>
</tr>
<tr>
<td>System DBA</td>
<td>31</td>
</tr>
<tr>
<td>Database Architect</td>
<td>32</td>
</tr>
<tr>
<td>Database Analyst</td>
<td>33</td>
</tr>
<tr>
<td>Data Modeler</td>
<td>33</td>
</tr>
<tr>
<td>Application DBA</td>
<td>34</td>
</tr>
<tr>
<td>Task-Oriented DBA</td>
<td>36</td>
</tr>
<tr>
<td>Performance Analyst</td>
<td>36</td>
</tr>
<tr>
<td>Data Warehouse Administrator</td>
<td>36</td>
</tr>
<tr>
<td>Staffing Considerations</td>
<td>37</td>
</tr>
<tr>
<td>How Many DBAs?</td>
<td>37</td>
</tr>
<tr>
<td>DBA Reporting Structures</td>
<td>40</td>
</tr>
<tr>
<td>Multiplatform DBA Issues</td>
<td>42</td>
</tr>
<tr>
<td>Production versus Test</td>
<td>44</td>
</tr>
<tr>
<td>The Impact of Newer Technology on DBA</td>
<td>46</td>
</tr>
<tr>
<td>Procedural DBAs: Managing Database Logic</td>
<td>46</td>
</tr>
</tbody>
</table>
Contents

The Internet: From DBA to eDBA 50
The Personal DBA and the Cloud 53
NoSQL, Big Data, and the DBA 55
New Technology Impacts on DBA 56

DBA Certification 56
The Rest of the Book 58
Review 58
Bonus Question 59

Chapter 2 Creating the Database Environment 61
Defining the Organization's DBMS Strategy 61
Choosing a DBMS 63
DBMS Architectures 68
DBMS Clustering 71
DBMS Proliferation 73
Hardware Issues 73
Cloud Database Systems 74

Installing the DBMS 75
DBMS Installation Basics 75
Hardware Requirements 76
Storage Requirements 76
Memory Requirements 78
Configuring the DBMS 80
Connecting the DBMS to Supporting Infrastructure Software 81
Installation Verification 81
DBMS Environments 82

Upgrading DBMS Versions and Releases 82
Features and Complexity 87
Complexity of the DBMS Environment 87
Reputation of the DBMS Vendor 89
Support Policies of the DBMS 89
<table>
<thead>
<tr>
<th>Chapter 3</th>
<th>Data Modeling and Normalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Modeling Concepts</td>
<td>108</td>
</tr>
<tr>
<td>Entity-Relationship Diagramming</td>
<td>110</td>
</tr>
<tr>
<td>The Components of a Data Model</td>
<td>113</td>
</tr>
<tr>
<td>Entities</td>
<td>113</td>
</tr>
<tr>
<td>Attributes</td>
<td>115</td>
</tr>
<tr>
<td>Keys</td>
<td>120</td>
</tr>
<tr>
<td>Relationships</td>
<td>122</td>
</tr>
<tr>
<td>Discovering Entities, Attributes, and Relationships</td>
<td>124</td>
</tr>
<tr>
<td>Conceptual, Logical, and Physical Data Models</td>
<td>125</td>
</tr>
<tr>
<td>What Is Normalization?</td>
<td>128</td>
</tr>
<tr>
<td>The Normal Forms</td>
<td>128</td>
</tr>
<tr>
<td>First Normal Form</td>
<td>129</td>
</tr>
<tr>
<td>Second Normal Form</td>
<td>129</td>
</tr>
<tr>
<td>Third Normal Form</td>
<td>132</td>
</tr>
</tbody>
</table>
A Normalized Data Model 133
Further Normal Forms 134
Normalization in Practice 135
Additional Data Modeling Issues 135
Summary 136
Review 137
Bonus Question 138
Suggested Reading 138

Chapter 4 Database Design 141
From Logical Model to Physical Database 141
Transform Entities to Tables 142
Transform Attributes to Columns 142
Build Referential Constraints for All Relationships 146
Build Physical Data Structures 147
Database Performance Design 150
Designing Indexes 150
Hashing 158
Clustering 159
Interleaving Data 160
Denormalization 160
When to Denormalize 161
Prejoined Tables 164
Report Tables 164
Mirror Tables 165
Split Tables 165
Combined Tables 168
Redundant Data 168
Repeating Groups 169
Derivable Data 170
Hierarchies 171
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special Physical Implementation Needs</td>
<td>173</td>
</tr>
<tr>
<td>Denormalization Summary</td>
<td>173</td>
</tr>
<tr>
<td>Views</td>
<td>175</td>
</tr>
<tr>
<td>Data Definition Language</td>
<td>177</td>
</tr>
<tr>
<td>Temporal Data Support</td>
<td>177</td>
</tr>
<tr>
<td>A Temporal Example</td>
<td>178</td>
</tr>
<tr>
<td>Business Time and System Time</td>
<td>179</td>
</tr>
<tr>
<td>Summary</td>
<td>180</td>
</tr>
<tr>
<td>Review</td>
<td>181</td>
</tr>
<tr>
<td>Bonus Question</td>
<td>181</td>
</tr>
<tr>
<td>Suggested Reading</td>
<td>182</td>
</tr>
</tbody>
</table>

**Chapter 5** Application Design 185

Database Application Development and SQL 186

SQL 186

Set-at-a-Time Processing and Relational Closure 189

Embedding SQL in a Program 191

SQL Middleware and APIs 192

Application Infrastructure 193

Object Orientation and SQL 199

Types of SQL 200

SQL Coding for Performance 202

Querying XML Data 203

Defining Transactions 205

Transaction Guidelines 207

Unit of Work 207

Transaction Processing Systems 207

Application Servers 209

Locking 210

Types of Locks 212

Lock Time-outs 213
Deadlocks 214
Lock Duration 215
Lock Escalation 219
Programming Techniques to Minimize Locking Problems 220
Locking Summary 220
Batch Processing 221
Summary 222
Review 222
Bonus Question 223
Suggested Reading 223

Chapter 6 Design Reviews 227
What Is a Design Review? 227
Rules of Engagement 228
Design Review Participants 229
Knowledge and Skills Required 232
Types of Design Reviews 232
Conceptual Design Review 233
Logical Design Review 235
Physical Design Review 236
Organizational Design Review 237
SQL and Application Code Design Review 238
Pre-Implementation Design Review 239
Post-Implementation Design Review 239
Design Review Output 239
Additional Considerations 240
Dealing with Remote Staff 240
Mentorship and Knowledge Transfer 240
Summary 241
Review 241
Suggested Reading 242
Chapter 7  Database Change Management  243
  Change Management Requirements  244
    The Change Management Perspective of the DBA  246
  Types of Changes  247
    DBMS Software  248
    Hardware Configuration  248
    Logical and Physical Design  248
    Applications  249
    Physical Database Structures  250
  Impact of Change on Database Structures  250
    The Limitations of ALTER  252
  Database Change Scenarios  254
  Comparing Database Structures  257
  Requesting Database Changes  258
  Standardized Change Requests  259
  Communication  260
  Coordinating Database and Application Changes  260
  Compliance  261
  DBA Scripts and Change Management  262
Summary  262
  Review  263
  Suggested Reading  263

Chapter 8  Data Availability  265
  Defining Availability  267
    Increased Availability Requirements  268
  Cost of Downtime  271
    How Much Availability Is Enough?  273
  Availability Problems  274
    Loss of the Data Center  274
    Network Problems  275
Contents xvii

Loss of the Server Hardware 276
Disk-Related Outages 278
Operating System Failure 279
DBMS Software Failure 279
Application Problems 279
Security and Authorization Problems 280
Corruption of Data 280
Loss of Database Objects 281
Loss of Data 282
Data Replication and Propagation Failures 283
Severe Performance Problems 283
Recovery Issues 284
DBA Mistakes 284
Outages: Planned and Unplanned 286
Ensuring Availability 287
Perform Routine Maintenance While Systems Remain Operational 288
Automate DBA Functions 290
Exploit High-Availability Features 291
Exploit Clustering Technology 292
Database Architecture and NoSQL 296
Summary 296
Review 297
Suggested Reading 298

Chapter 9 Performance Management 299
Defining Performance 299
A Basic Database Performance Road Map 302
Monitoring versus Management 304
Reactive versus Proactive 306
Chapter 11  Database Performance  349

Techniques for Optimizing Databases  349
  Partitioning  350
  Raw Partition versus File System  351
  Indexing  352
  Denormalization  355
  Clustering  356
  Interleaving Data  360
  Free Space  360
  Compression  361
  File Placement and Allocation  362
  Page Size (Block Size)  364
Database Reorganization  365
  Determining When to Reorganize  369
  Automation  371

Summary  371
Review  371
Suggested Reading  372

Chapter 12  Application Performance  373

Designing Applications for Relational Access  373
Relational Optimization  374
  CPU and I/O Costs  376
  Database Statistics  376
Query Analysis  378
Joins  379
Access Path Choices  381
Additional Optimization Considerations 391
  View Access 391
  Query Rewrite 392
  Rule-Based Optimization 393
Reviewing Access Paths 394
  Forcing Access Paths 398
SQL Coding and Tuning for Efficiency 399
  A Dozen SQL Rules of Thumb 400
  Additional SQL Tuning Tips 406
  Identifying Poorly Performing SQL 406
Summary 407
Review 407
Suggested Reading 408

Chapter 13  Data Integrity 409
Types of Integrity 409
  Database Structure Integrity 410
    Types of Structural Problems 410
    Managing Structural Problems 411
Semantic Data Integrity 414
  Entity Integrity 416
  Unique Constraints 417
  Data Types 417
  Default Values 419
  Check Constraints 419
  Triggers 426
  Referential Integrity 433
Temporal Database Systems 444
Summary 446
Review 447
Suggested Reading 448
# Chapter 14 Database Security

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Breaches</td>
<td>449</td>
</tr>
<tr>
<td>Database Security Basics</td>
<td>451</td>
</tr>
<tr>
<td>Database Users</td>
<td>455</td>
</tr>
<tr>
<td>Granting and Revoking Authority</td>
<td>456</td>
</tr>
<tr>
<td>Types of Privileges</td>
<td>457</td>
</tr>
<tr>
<td>Granting to PUBLIC</td>
<td>460</td>
</tr>
<tr>
<td>Revoking Privileges</td>
<td>461</td>
</tr>
<tr>
<td>Label-Based Access Control</td>
<td>463</td>
</tr>
<tr>
<td>Security Reporting</td>
<td>465</td>
</tr>
<tr>
<td>Authorization Roles and Groups</td>
<td>466</td>
</tr>
<tr>
<td>Roles</td>
<td>466</td>
</tr>
<tr>
<td>Groups</td>
<td>467</td>
</tr>
<tr>
<td>Other Database Security Mechanisms</td>
<td>468</td>
</tr>
<tr>
<td>Using Views for Security</td>
<td>468</td>
</tr>
<tr>
<td>Using Stored Procedures for Security</td>
<td>470</td>
</tr>
<tr>
<td>Encryption</td>
<td>470</td>
</tr>
<tr>
<td>Data at Rest Encryption</td>
<td>472</td>
</tr>
<tr>
<td>Data in Transit Encryption</td>
<td>472</td>
</tr>
<tr>
<td>Encryption Techniques</td>
<td>472</td>
</tr>
<tr>
<td>SQL Injection</td>
<td>473</td>
</tr>
<tr>
<td>SQL Injection Prevention</td>
<td>475</td>
</tr>
<tr>
<td>Auditing</td>
<td>477</td>
</tr>
<tr>
<td>External Security</td>
<td>478</td>
</tr>
<tr>
<td>Job Scheduling and Security</td>
<td>479</td>
</tr>
<tr>
<td>Non-DBMS DBA Security</td>
<td>480</td>
</tr>
<tr>
<td>DBMS Fixpacks and Maintenance</td>
<td>480</td>
</tr>
<tr>
<td>Summary</td>
<td>481</td>
</tr>
<tr>
<td>Review</td>
<td>481</td>
</tr>
<tr>
<td>Suggested Reading</td>
<td>482</td>
</tr>
</tbody>
</table>
Chapter 15  Regulatory Compliance and Database Administration  483

A Collaborative Approach to Compliance  486

Why Should DBAs Care about Compliance?  487

Metadata Management, Data Quality, and Data Governance  488

Metadata  488

Data Quality  489

Data Governance  489

Database Auditing and Data Access Tracking  490

Database Auditing Techniques  493

Privileged User Auditing  495

Data Masking and Obfuscation  496

Data Masking Techniques  497

Database Archiving for Long-Term Data Retention  498

The Life Cycle of Data  499

Database Archiving  500

Components of a Database Archiving Solution  505

The Impact of e-Discovery on DBA  506

Closer Tracking of Traditional DBA Tasks  507

Database Change Management  508

Database Backup and Recovery  508

Summary  511

Review  511

Suggested Reading  512

Chapter 16  Database Backup and Recovery  515

The Importance of Backup and Recovery  515

Preparing for Problems  516

Backup  517

Full versus Incremental Backups  521

Database Objects and Backups  523
General Disaster Recovery Guidelines 563
  The Remote Site 564
  The Written Plan 564
  Personnel 569
Back Up the Database for Disaster Recovery 569
  Tape Backups 570
  Storage Management Backups 572
  Other Approaches 573
  Some Guidelines 573
Disaster Prevention 575

Disaster and Contingency Planning Web Sites 576

Summary 576
Review 576
Suggested Reading 577

Chapter 18  Data and Storage Management 579

Storage Management Basics 579
Files and Data Sets 583
  File Placement on Disk 584
  Raw Partitions versus File Systems 586
  Temporary Database Files 587
Space Management 587
  Data Page Layouts 588
  Index Page Layouts 592
  Transaction Logs 594
Fragmentation and Storage 595
Storage Options 596
  RAID 597
  JBOD 604
  Storage Area Networks 604
Chapter 19  Data Movement and Distribution  613

Loading and Unloading Data  614

The LOAD Utility  614

The UNLOAD Utility  618

Maintaining Application Test Beds  621

EXPORT and IMPORT  622

Bulk Data Movement  623

ETL Software  623

Replication and Propagation  623

Messaging Software  624

Other Methods  625

Distributed Databases  626

Setting Up a Distributed Environment  627

Data Distribution Standards  629

Accessing Distributed Data  630

Two-Phase COMMIT  631

Distributed Performance Problems  632

Summary  633

Review  634

Bonus Question  634

Suggested Reading  635
Chapter 20  Data Warehouse Administration  637
What Is a Data Warehouse?  637
  Analytical versus Transaction Processing  638
Administering the Data Warehouse  640
  Too Much Focus on Technology?  641
Data Warehouse Design  641
Data Movement  644
Data Cleansing  645
Data Warehouse Scalability  649
Data Warehouse Performance  650
Data Freshness  654
Data Content  654
Data Usage  655
  Financial Chargeback  655
  Backup and Recovery  656
  Don't Operate in a Vacuum!  657
Summary  658
Review  658
Suggested Reading  659

Chapter 21  Database Connectivity  661
Multitier, Distributed Computing  661
  A Historical Look  661
Business Issues  663
  What Is Client/Server Computing?  663
  Types of Client/Server Applications  667
Network Traffic  670
  Database Gateways  671
  Database Drivers  672
  Connection Pooling  674
Databases, the Internet, and the Web 675

Internet-Connected Databases 676

Web Development and Web Services 680

Summary 681

Review 682

Suggested Reading 682

Chapter 22 Metadata Management 685

What Is Metadata? 685

From Data to Knowledge and Beyond 686

Metadata Strategy 687

Data Warehousing and Metadata 688

Types of Metadata 689

Repositories and Data Dictionaries 691

Repository Benefits 693

Repository Challenges 693

Data Dictionaries 695

Summary 696

Review 696

Suggested Reading 697

Chapter 23 DBA Tools 699

Types and Benefits of DBA Tools 699

Data Modeling and Design 700

Database Change Management 701

Table Editors 707

Performance Management 708

Backup and Recovery 714

Database Utilities 715

Data Protection, Governance, Risk, and Compliance Tools 716

Data Warehousing, Analytics, and Business Intelligence 721
Programming and Development Tools 724

Miscellaneous Tools 726

Examine Native DBA Tools 728
Evaluating DBA Tool Vendors 729

Homegrown DBA Tools 732

Summary 733

Review 733

Chapter 24 DBA Rules of Thumb 735

Write Down Everything 735
Keep Everything 736
Automate! 737
Share Your Knowledge 739
Analyze, Simplify, and Focus 741
Don't Panic! 742
Measure Twice, Cut Once 743
Understand the Business, Not Just the Technology 743
Don't Become a Hermit 745
Use All of the Resources at Your Disposal 745
Keep Up-to-Date 746
Invest in Yourself 747
Summary 748
Final Exam 748

Appendix A Database Fundamentals 753

What Is a Database? 753
Why Use a DBMS? 754

Advantages of Using a DBMS 755

Summary 759
Bibliography  793
Database Management and Database Systems  793
Data Administration, Data Modeling, and Database Design  799
Database Security, Protection, and Compliance  802
Data Warehousing  804
SQL  805
Object Orientation and Database Management  807
Operating Systems  807
Related Topics  808
DB2  812
IMS  813
MySQL  813
Oracle  814
SQL Server  815
Sybase  816
Other Database Systems  817

Glossary  819

Index  853