Contents

About the Authors  xvii
Foreword  xix
Preface  xxi
  The Purpose of This Book  xxii
    Complexity  xxii
    Step-by-Step Guide  xxii
  How This Book Is Organized  xxiii
    Part I: Stages and Steps  xxiv
    Part II: At a Glance  xxv
  How to Use This Book  xxvi
  Who Should Read This Book  xxvi
    Business Representatives  xxvi
    Business Sponsors  xxvi
    Project Managers  xxviii
    Technicians  xxvix
  Comments  xxx

Acknowledgments  xxxi

PART I  STAGES AND STEPS  1

Chapter Zero  Guide to the Development Steps  3
  Business Intelligence Definition  4
  BI Decision-Support Initiatives  5
  Development Approaches  5
    The Traditional Development Approach  6
    The Cross-Organizational Development Approach  8
  Engineering Stages and the Development Steps  11
  Parallel Development Tracks  17
  BI Project Team Structure  20
    The Core Team  20
    The Extended Team  23
    The BI Arbitration Board  26
Chapter One   Step 1: Business Case Assessment   29

Business Justification   31
Business Drivers   33
Business Analysis Issues   35
   Information Needs   35
   Types of Data Sources   35
   Source Data Quality   37
Cost-Benefit Analysis   37
Risk Assessment   40
Business Case Assessment Activities   45
Deliverable Resulting from These Activities   48
Roles Involved in These Activities   49
Risks of Not Performing Step 1   49
Bibliography and Additional Reading   50

Chapter Two   Step 2: Enterprise Infrastructure Evaluation   51

Step 2, Section A: Technical Infrastructure Evaluation   53
The Hardware Platform   54
   Controlled Chaos   54
   Hardware Platform Requirements   55
The Middleware Platform   57
   DBMS Gateways   58
The DBMS Platform   58
   Criteria for Selecting a DBMS   59
Technical Infrastructure Evaluation Activities   61
Deliverables Resulting from These Activities   62
Roles Involved in These Activities   63
Risks of Not Performing Step 2, Section A   63
Step 2, Section B: Nontechnical Infrastructure Evaluation   64
The Effects of Stovepipe Development   65
The Need for Nontechnical Infrastructure   66
Enterprise Architecture   68
Enterprise Standards   71
Nontechnical Infrastructure Evaluation Activities   75
Deliverable Resulting from These Activities   76
Roles Involved in These Activities   77
Contents

Risks of Not Performing Step 2, Section B 78
Bibliography and Additional Reading 78
  Technical Infrastructure Evaluation 78
  Nontechnical Infrastructure Evaluation 79

Chapter Three  Step 3: Project Planning  81
Managing the BI Project 83
Defining the BI Project 84
  Project Goals and Objectives 85
  Project Scope 85
  Project Risks 85
  Project Constraints 86
  Assumptions 87
  Change-Control Procedures 88
  Issues Management Procedures 90
Planning the BI Project 90
  Activities and Tasks 90
  Estimating Techniques 92
  Resource Assignment 93
  Task Dependencies 94
  Resource Dependencies 95
  Critical Path Method 95
  Project Schedules 96
Project Planning Activities 98
Deliverables Resulting from These Activities 100
Roles Involved in These Activities 101
Risks of Not Performing Step 3 103
Bibliography and Additional Reading 103

Chapter Four  Step 4: Project Requirements Definition 105
General Business Requirements 108
  Interviewees for General Business Requirements 109
  Data Quality Requirements 110
  Business Requirements Report 111
Project-Specific Requirements 112
  Interviewees for Project-Specific Requirements 113
  Application Requirements Document 114
The Interviewing Process 116
  Interviewing Considerations 116
  Interviewing Tips 117
Risks of Not Performing Step 2, Section B  78
Bibliography and Additional Reading  78
   Technical Infrastructure Evaluation  78
   Nontechnical Infrastructure Evaluation  79

Chapter Three  Step 3: Project Planning  81
   Managing the BI Project  83
   Defining the BI Project  84
      Project Goals and Objectives  85
      Project Scope  85
      Project Risks  85
      Project Constraints  86
      Assumptions  87
      Change-Control Procedures  88
      Issues Management Procedures  90
   Planning the BI Project  90
      Activities and Tasks  90
      Estimating Techniques  92
      Resource Assignment  93
      Task Dependencies  94
      Resource Dependencies  95
      Critical Path Method  95
      Project Schedules  96
   Project Planning Activities  98
   Deliverables Resulting from These Activities  100
   Roles Involved in These Activities  101
   Risks of Not Performing Step 3  103
   Bibliography and Additional Reading  103

Chapter Four  Step 4: Project Requirements Definition  105
   General Business Requirements  108
      Intervieweess for General Business Requirements  109
      Data Quality Requirements  110
      Business Requirements Report  111
   Project-Specific Requirements  112
      Intervieweess for Project-Specific Requirements  113
      Application Requirements Document  114
   The Interviewing Process  116
      Interviewing Considerations  116
      Interviewing Tips  117
Chapter Five  Step 5: Data Analysis  125

Business-Focused Data Analysis  127
Top-Down Logical Data Modeling  128
  Project-Specific Logical Data Model  128
  Enterprise Logical Data Model  129
Logical Data Modeling Participants  131
  Standardized Business Meta Data  131
Bottom-Up Source Data Analysis  133
  Technical Data Conversion Rules  134
  Business Data Domain Rules  134
  Business Data Integrity Rules  135
Data Cleansing  136
  Data Quality Responsibility  137
  Source Data Selection Process  137
  Key Points of Data Selection  139
  To Cleanse or Not to Cleanse  140
  Cleansing Operational Systems  141
Data Analysis Activities  141
Deliverables Resulting from These Activities  143
Roles Involved in These Activities  144
Risks of Not Performing Step 5  145
Bibliography and Additional Reading  146

Chapter Six  Step 6: Application Prototyping  149

Purposes of Prototyping  151
  Time-Boxing  152
Best Practices for Prototyping  153
  Considerations for Prototyping  154
Types of Prototypes  156
  Show-and-Tell Prototype  156
  Mock-Up Prototype  156
  Proof-of-Concept Prototype  157
  Visual-Design Prototype  157
Chapter Seven  Step 7: Meta Data Repository Analysis  169

The Importance of Meta Data  172
   Meta Data Categories  173
Meta Data Repository as Navigation Tool  174
   Data Standardization  175
Meta Data Classifications  176
   Groupings of Meta Data Components  176
   Prioritization of Meta Data Components  179
Meta Data Repository Challenges  182
   Technical Challenges  182
   Staffing Challenges  183
   Budget Challenges  183
   Usability Challenges  183
   Political Challenges  184
The Logical Meta Model  184
   The Entity-Relationship Meta Model  185
Meta-Meta Data  186
Meta Data Repository Analysis Activities  186
Deliverables Resulting from These Activities  188
Roles Involved in These Activities  188
Risks of Not Performing Step 7  189
Bibliography and Additional Reading  190

Chapter Eight  Step 8: Database Design  191

Differences in Database Design Philosophies  193
   Operational Databases  193
   BI Target Databases  196
Logical Database Design 197
  The Star Schema 197
  The Snowflake Schema 200
Physical Database Design 201
  Implementation Options 201
  Physical Dataset Placement 201
  Partitioning 202
  Clustering 202
  Indexing 202
  Reorganizations 203
  Backup and Recovery 203
  Parallel Query Execution 204
Database Design Activities 204
Deliverables Resulting from These Activities 207
Roles Involved in These Activities 208
Risks of Not Performing Step 8 209
Bibliography and Additional Reading 209

Chapter Nine  Step 9: Extract/Transform/Load Design 211
  Implementation Strategies 213
Preparing for the ETL Process 215
  The Initial Load 216
  The Historical Load 217
  The Incremental Load 217
Designing the Extract Programs 219
Designing the Transformation Programs 221
  Source Data Problems 221
  Data Transformations 222
Designing the Load Programs 223
  Referential Integrity 224
  Indexing 224
Designing the ETL Process Flow 225
  The Source-to-Target Mapping Document 225
  The ETL Process Flow Diagram 225
  The Staging Area 228
Evaluating ETL Tools 229
ETL Design Activities 231
Deliverables Resulting from These Activities 233
Roles Involved in These Activities 233
Risks of Not Performing Step 9  234
Bibliography and Additional Reading  234

Chapter Ten  Step 10: Meta Data Repository Design  237
Meta Data Silos  239
   Sources of Meta Data  240
Meta Data Repository Solutions  242
   Centralized Meta Data Repository  242
   Decentralized Meta Data Repository  244
   Distributed XML-Enabled Meta Data Solution  245
Designing a Meta Data Repository  247
   Entity-Relationship Design  247
   Object-Oriented Design  248
Licensing (Buying) a Meta Data Repository  250
   Product Evaluation  251
   Vendor Evaluation  252
Meta Data Repository Design Activities  254
Deliverables Resulting from These Activities  255
Roles Involved in These Activities  256
Risks of Not Performing Step 10  257
Bibliography and Additional Reading  257

Chapter Eleven  Step 11: Extract/Transform/Load Development  259
Source Data Transformation  261
   Data Transformation Activities  261
   Underestimating Data Transformation Efforts  262
Reconciliation  263
   Calculating Reconciliation Totals  264
   Storing Reconciliation Statistics  266
Peer Reviews  267
ETL Testing  268
   Unit Testing  269
   Integration Testing  270
   Regression Testing  271
   Performance Testing  271
   Quality Assurance Testing  272
   Acceptance Testing  272
Formal Test Plan  273
ETL Development Activities  276
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverables Resulting from These Activities</td>
<td>277</td>
</tr>
<tr>
<td>Roles Involved in These Activities</td>
<td>278</td>
</tr>
<tr>
<td>Risks of Not Performing Step 11</td>
<td>279</td>
</tr>
<tr>
<td>Bibliography and Additional Reading</td>
<td>279</td>
</tr>
<tr>
<td><strong>Chapter Twelve  Step 12: Application Development</strong></td>
<td>281</td>
</tr>
<tr>
<td>Online Analytical Processing Tools</td>
<td>283</td>
</tr>
<tr>
<td>Advantages of OLAP Tools</td>
<td>284</td>
</tr>
<tr>
<td>OLAP Tool Features</td>
<td>285</td>
</tr>
<tr>
<td>Multidimensional Analysis Factors</td>
<td>287</td>
</tr>
<tr>
<td>Multivariate Analysis</td>
<td>289</td>
</tr>
<tr>
<td>Online Analytical Processing Architecture</td>
<td>289</td>
</tr>
<tr>
<td>Presentation Services</td>
<td>290</td>
</tr>
<tr>
<td>OLAP Services</td>
<td>291</td>
</tr>
<tr>
<td>Database Services</td>
<td>292</td>
</tr>
<tr>
<td>Development Environments</td>
<td>292</td>
</tr>
<tr>
<td>Application Development Activities</td>
<td>295</td>
</tr>
<tr>
<td>Deliverables Resulting from These Activities</td>
<td>297</td>
</tr>
<tr>
<td>Roles Involved in These Activities</td>
<td>298</td>
</tr>
<tr>
<td>Risks of Not Performing Step 12</td>
<td>299</td>
</tr>
<tr>
<td>Bibliography and Additional Reading</td>
<td>299</td>
</tr>
<tr>
<td><strong>Chapter Thirteen  Step 13: Data Mining</strong></td>
<td>301</td>
</tr>
<tr>
<td>Defining Data Mining</td>
<td>303</td>
</tr>
<tr>
<td>The Importance of Data Mining</td>
<td>305</td>
</tr>
<tr>
<td>Data Sources for Data Mining</td>
<td>306</td>
</tr>
<tr>
<td>Data Mining Techniques</td>
<td>307</td>
</tr>
<tr>
<td>Associations Discovery</td>
<td>307</td>
</tr>
<tr>
<td>Sequential Pattern Discovery</td>
<td>308</td>
</tr>
<tr>
<td>Classification</td>
<td>309</td>
</tr>
<tr>
<td>Clustering</td>
<td>309</td>
</tr>
<tr>
<td>Forecasting</td>
<td>309</td>
</tr>
<tr>
<td>Data Mining Operations</td>
<td>310</td>
</tr>
<tr>
<td>Predictive and Classification Modeling</td>
<td>310</td>
</tr>
<tr>
<td>Link Analysis</td>
<td>311</td>
</tr>
<tr>
<td>Database Segmentation</td>
<td>311</td>
</tr>
<tr>
<td>Deviation Detection</td>
<td>311</td>
</tr>
<tr>
<td>Applications of Data Mining</td>
<td>311</td>
</tr>
<tr>
<td>Data Mining Activities</td>
<td>313</td>
</tr>
<tr>
<td>Deliverables Resulting from These Activities</td>
<td>315</td>
</tr>
</tbody>
</table>
Contents

Roles Involved in These Activities 316
Risks of Not Performing Step 13 316
Bibliography and Additional Reading 317

Chapter Fourteen  Step 14: Meta Data Repository Development 319
Populating the Meta Data Repository 321
Meta Data Repository Interface Processes 324
    The Tool Interface Process 324
    The Access Interface Process 325
Meta Data Repository Testing 326
Preparing for the Meta Data Repository Rollout 327
    Meta Data Repository Directory 330
Meta Data Repository Development Activities 331
Deliverables Resulting from These Activities 332
Roles Involved in These Activities 333
Risks of Not Performing Step 14 334
Bibliography and Additional Reading 335

Chapter Fifteen  Step 15: Implementation 337
Incremental Rollout 339
Security Management 340
    Security Measures for BI Applications 340
    Security in a Multi-Tier Environment 341
    Security for Internet Access 344
Data Backup and Recovery 345
Monitoring the Utilization of Resources 347
    Computer Utilization 347
    Network Utilization 347
    Personnel Utilization 348
Growth Management 349
    Growth in Data 349
    Growth in Usage 350
    Growth in Hardware 351
Implementation Activities 352
Deliverables Resulting from These Activities 354
Roles Involved in These Activities 355
Risks of Not Performing Step 15 356
Bibliography and Additional Reading 356
Chapter Sixteen  Step 16: Release Evaluation  359
   The Application Release Concept  361
      Guidelines for Using the Release Concept  362
   Post-Implementation Reviews  364
      Organizing a Post-Implementation Review  366
      Post-Implementation Review Session Flow  368
   Release Evaluation Activities  369
   Deliverables Resulting from These Activities  371
   Roles Involved in These Activities  371
   Risks of Not Performing Step 16  374
   Bibliography and Additional Reading  375

PART II  AT A GLANCE  377

Chapter Seventeen  Human Resource Allocation Matrix  379

Chapter Eighteen  Entry & Exit Criteria and Deliverables Matrix  387

Chapter Nineteen  Activity Dependency Matrix  405

Chapter Twenty  Task/Subtask Matrix  411

Chapter Twenty-one  Practical Guidelines Matrix  455

Appendix  Work Breakdown Structure  491

Index  525