

Contents

Foreword	xvii
How to Use This Book	xix
Acknowledgments	xxiii

PART I CONCEPTS AND CONTEXT

CHAPTER 1 The Business Demand for Data, Information, and Analytics	3
Just One Word: Data	3
Welcome to the Data Deluge	4
Data Volume, Variety, and Velocity	4
Taming the Analytics Deluge	6
The Importance of Analytics	7
Analytics Challenges	7
Analytics Strategy	8
Too Much Data, Too Little Information	8
The Difference between Data and Information	8
The Role of BI in Creating Actionable Information	10
The Information Backbone	10
Data Capture versus Information Analysis	10
The Roles of BI and Operational Systems	11
Operational BI Blurs the Lines	11
Where Data Warehousing Fits in	11
The Five Cs of Data	12
Common Terminology from Our Perspective	14
References	19

PART II BUSINESS AND TECHNICAL NEEDS

CHAPTER 2 Justifying BI: Building the Business and Technical Case	23
Why Justification is Needed	23
Building the Business Case	24
Review Organization's Business Initiatives and Processes	25
Solicit BI Sponsorship	25
Enlist BI Stakeholders	26
Identify Business Processes Affected by BI	26
Document Business Benefits	27
Determine Business Value (Tangible Benefits)	27

Building the Technical Case	28
Technology and Product Short Lists	28
Convincing Business People	30
Convincing the Technologists	31
Assessing Readiness	32
Data and Data Quality	32
Expertise and Experience	33
Organizational and Cultural Change	34
Financial and Resource Commitment	34
Creating a BI Road Map	35
Developing Scope, Preliminary Plan, and Budget	35
Project Scope	36
Project Plan	36
Project Budget	37
Calculating Benefits and ROI	39
Obtaining Approval	40
Common Justification Pitfalls	40
Overzealous Business Sponsor	41
CIO is Sole Sponsor	41
Intangible or Too High-Level Benefits	41
Confusion Between BI Technology and Business Value	41
CHAPTER 3 Defining Requirements—Business, Data and Quality	43
The Purpose of Defining Requirements	43
Goals	44
Deliverables	45
Roles	47
BI Team Participants	47
Business Participants	48
Other IT Participants	49
Defining Requirements Workflow	49
Business Requirements	49
Data (and Data Quality) Requirements	51
Functional Requirements	52
Regulatory/Compliance Requirements	53
Technical Requirements	54
Reverse Engineering (When Necessary)	54
Putting It All Together	55
Prioritizing Requirements	55
Interviewing	56
Preparation for Interviews	57

Conducting the Interviews	58
Reviewing Interview Content	58
Interview Follow-ups	59
Documenting Requirements	60

PART III ARCHITECTURAL FRAMEWORK

CHAPTER 4 Architecture Framework	65
The Need for Architectural Blueprints	65
Architectural Framework	66
Information Architecture	67
Data Architecture	68
The Rise of the Enterprise Data Warehouse	68
Data Warehousing Replaces the Data Warehouse	69
Technical Architecture	72
Business Intelligence	74
Data Warehouse and BI Data Stores	75
Data Integration	76
Source Systems	76
BI Technology Keeps Evolving	77
Product Architecture	78
Metadata	78
What is It?	79
What to do About It	79
Security and Privacy	80
Getting Started	80
Implementing the Plan	81
Avoiding Accidents with Architectural Planning	81
The Signs of Accidental Architecture Addiction	82
Recovering from an Accidental Architecture	83
Do Not Obsess over the Architecture	83
CHAPTER 5 Information Architecture	85
The Purpose of an Information Architecture	85
Data Integration Framework	86
DIF Information Architecture	87
Data Preparation	88
Data Franchising	91
BI and Analytics	95
Data Management	98
Metadata Management	98

Operational BI versus Analytical BI	100
Shift All Reporting to the Application-Specific Environment	101
Shift All Reporting to the DW-Based BI Environment	102
Blend Application-Specific and DW BI Environments	102
Master Data Management	103
Identify the Data	104
Find the Problem Areas	104
Assess a Solution	105
CHAPTER 6 Data Architecture.....	107
The Purpose of a Data Architecture	107
History	108
Prehistory	108
In the Beginning	108
Data Warehousing Goes Public	109
The Data Mart	110
Multiple Data Marts	112
Operational Data Store (ODS)	113
Federated DWs	114
BI Accidental Architecture	115
Hub-and-Spoke	116
Data Architectural Choices	118
Data Categories	119
Selecting a Data Architecture	119
The Same But Different	122
Analytical Data Architecture (ADA)	123
Data Integration Workflow	128
Data Integration Workflow—Hub-and-Spoke	129
Data Workflow of the System of Integration (SOI)	130
Data Workflow of the System of Analytics (SOA)	133
Data Workflow—Rise of EDW Again	136
Operational Data Store	137
The Rationale for an ODS	138
ODS Reexamined	140
ODS is Dead, Long Live ODS	142
References	142
CHAPTER 7 Technology & Product Architectures	143
Where are the Product and Vendor Names?	143
Evolution Not Revolution	144
Technology Platforms	145
Enterprise Applications	145

Data Management	146
Technology Architecture	147
Business Intelligence and Analytics	149
Information Access and Data Integration	151
Databases	158
Product and Technology Evaluations	165
BI Product Vendors	165
Dazed and Confused	166
Technology and Product Evaluations	167
Product Migration	168

PART IV DATA DESIGN

CHAPTER 8 Foundational Data Modeling	173
The Purpose of Data Modeling	173
Definitions—The Difference Between a Data Model and Data Modeling	173
Three Levels of Data Models	174
Conceptual Data Model	175
Logical Data Model	175
Physical Data Model	176
Data Modeling Workflow	177
Where Data Models Are Used	178
Entity-Relationship (ER) Modeling Overview	179
ER Building Blocks	179
Types of Entities and Attributes	180
Relationship Cardinality	181
Types of Relationship Models	181
Keys	187
Referential Integrity	188
Normalization	189
Normalization Levels	190
Normalizing an Entity	190
First Normal Form—1NF	190
Second Normal Form—2NF	191
Third Normal Form—3NF	192
ER Model Recap	193
Limits and Purpose of Normalization	194
CHAPTER 9 Dimensional Modeling	197
Introduction to Dimensional Modeling	197
High-Level View of a Dimensional Model	198
Facts	198

Fact Table Keys	199
Fact Table Measures	201
Types of Facts	202
Dimensions	203
Dimension Hierarchy	204
Dimension Keys and Surrogate Keys	205
Schemas	208
Star Schema	208
Snowflake Schema	210
Multidimensional Schema	211
Multifact Star Models	212
Entity Relationship versus Dimensional Modeling	213
Comparing Approaches	213
Why the Structures Matter	215
Purpose of Dimensional Modeling	216
Mapping to a Business Report	217
Mapping to an OLAP Analysis or Spreadsheet	217
Fact Tables	218
Achieving Consistency	220
Conforming Dimensions	220
Conforming Facts	220
Advanced Dimensions and Facts	221
Date (or Calendar) Dimension	221
Time Dimension	225
Date and Time Dimensions across Time Zones	228
Role Playing Dimensions	229
Degenerative Dimensions	230
Event Tables	232
Consolidated Fact Tables	233
Dimensional Modeling Recap	234
CHAPTER 10 Business Intelligence Dimensional Modeling.....	237
Introduction	237
Hierarchies	237
Balanced Hierarchies	238
Ragged Hierarchies	240
Unbalanced Hierarchies	241
Working with Variable-Depth Hierarchies	241
Outrigger Tables	244
Slowly Changing Dimensions	245
SCD Type 0: Keep Original	246

SCD Type 1: Overwrite Existing Data	246
SCD Type 2: Create New Row	249
SCD Type 3: Track Changes Using Separate Columns but Overwrite the Rest	251
SCD Type 4: Add Mini-Dimension	254
SCD Type 5: Combine Types 4 + 1	257
SCD Type 6: Combine Types 1 + 2 + 3 (=6)	259
SCD Type 7: Use both Types 1 + 2	260
Causal Dimension	262
Multivalued Dimensions	263
Pick One Value, Ignore Others	264
Extend the Attributes	264
Bridge Table	264
Junk Dimensions	265
Misguided Attempts	266
Recommended Solution	266
Value Band Reporting	268
Heterogeneous Products	269
Alternate Dimensions	270
Hot Swappable Dimension	270
Custom Dimension Groups	271
Too Few or Too Many Dimensions	272

PART V DATA INTEGRATION DESIGN

CHAPTER 11 Data Integration Design and Development	275
Getting Started with Data Integration	275
Holistic Approach	275
Incremental and Iterative Approaches	276
Productive, Efficient Methods	277
Data Integration Architecture	277
Data Integration Requirements	280
Prerequisites	280
Business Requirements	282
Data Sourcing System Requirements	284
Data Integration Design	285
Create Stage-Related Conceptual Data Integration Model	285
Create Stage-Related Logical Data Integration Model	286
Create Stage-Related Physical Data Integration Model	287
Design Stage-Related Source to Target Mappings	288
Design Overall Data Integration Workflow	289

Data Integration Standards	290
For Standards, Adopt a Software Development Mindset	290
Development Project Standards	291
Data Integration Reusable Components	292
Data Integration Development Standards	293
Loading Historical Data	295
Same as the Old	296
Historical Data is Different or Dirty	297
Data Integration Prototyping	298
Data Integration Testing	298
CHAPTER 12 Data Integration Processes	301
Introduction: Manual Coding versus Tool-Based Data Integration	301
Why We Hand Code	302
Why We Use Tools	304
Selecting a Tool	305
Two Tools May Be Better Than One	309
Data Integration Services	309
Access (Extract) and Delivery (Load) Services	312
Data Ingestion Services	312
Data Profiling Services	317
Data Transformation Services	320
Data Quality Services	324
Process Management Services	328
Operations Management Services	331
Data Transport Services	332
PART VI BUSINESS INTELLIGENCE DESIGN	
CHAPTER 13 Business Intelligence Applications	337
BI Content Specifications	337
Revise BI Applications List	339
BI Personas	340
Casual Consumers	341
Analysts	341
Power Users	342
Data Scientists	342
BI Design Layout—Best Practices	343
Focus on the Purpose	343
Design Layout	343
Data Design for Self-Service BI	348
The Last Data Preparation Step	348

When Inconsistency is Reintroduced	349
OLAP Cubes and In-Memory Columnar Databases	350
Matching Types of Analysis to Visualizations	351
Comparative Analysis	351
Time-Series or Trending Analysis	351
Contribution Analysis	352
Correlation Analysis	354
Geographic Data	354
Distribution Analysis	357
CHAPTER 14 BI Design and Development	359
BI Design	359
BI User Interface (UI) Standards	359
Create Privacy, Security and Access Standards	361
Designing Each BI Application	362
BI Development	367
Prototyping Lifecycle	369
BI Application Development Tasks	370
BI Application Testing	372
CHAPTER 15 Advanced Analytics	375
Advanced Analytics Overview and Background	375
The Window to the Future	376
Don't Ignore the Past	376
Advanced Analytics in Action	376
Predictive Analytics and Data Mining	377
Setting Up a Predictive Analytics or Data Mining Program	377
Tasks for Developing and Using Models	378
Selecting Tools	380
Architecture for Predictive Analytics and Data Mining	380
Techniques for Predictive Analytics and Data Mining	381
Resources and Skills	382
Roadblocks to Success	382
Analytical Sandboxes and Hubs	383
Analytical Sandboxes	383
Analytical Hubs	384
Hub and Sandbox Design Principles	387
Hub and Sandbox Architecture Options	388
Advice for Hubs and Sandboxes	394
Big Data Analytics	395
Scope	396
The Program	397

Hybrid Architecture	397
The Big Data Team	398
Big Data Analytics Worst Practices	400
Data Visualization	401
Why Data Visualization is Needed	401
What Data Visualization is Not	402
Reference	402
CHAPTER 16 Data Shadow Systems	403
The Data Shadow Problem	403
Are There Data Shadow Systems in Your Organization?	405
What Kind of Data Shadow Systems do You Have?	406
Data Shadow System Triage	407
Mild and Moderate	407
Serious	407
Very Serious	408
The Evolution of Data Shadow Systems in an Organization	408
Filling in a Gap	408
As They Grow, It is Harder to Turn Back	410
IT Group Response to Data Shadow Systems	411
Damages Caused by Data Shadow Systems	412
The Benefits of Data Shadow Systems	413
Moving beyond Data Shadow Systems	414
Stopping the Blame Game	415
Business and IT—In It Together	415
Changing the Approach	416
Changing the Culture	417
Misguided Attempts to Replace Data Shadow Systems	417
Migrating to BI Tool	418
Create a Standard Set of Reports	418
Renovating Data Shadow Systems	418
Data Integration Considerations	419
Analytical Process Considerations	421
Balanced Priorities	421
PART VII ORGANIZATION	
CHAPTER 17 People, Process and Politics	425
The Technology Trap	425
Meeting Expectations	425
Getting People, Process, and Politics under Control	426
The Business and IT Relationship	427

Users or Customers?	427
Who's in Charge?	427
Communication Shortcomings	428
Roles and Responsibilities	429
The Business Group's Role in the Front Office	429
IT's Role in the Back Office	430
Roles that Span Both Business and IT	430
Building the BI Team	431
Project Sponsorship and Governance	431
Project Management Team	434
Project Development Team (Core Team)	435
Extended Project Team	440
Training	441
Types of Training	441
Training for the IT Group	442
Training for the Business Group	442
Training Methods	443
Data Governance	444
Who's Responsible, Business or IT?	445
Getting Started with Data Governance	445
Creating a Data Governance Organization	446
Working with COEs	446
Keeping Data Governance on Track	447
CHAPTER 18 Project Management.....	449
The Role of Project Management	449
Establishing a BI Program	450
Business Strategy Drives the BI Program	451
Business Sponsors for the BI Program	452
Determining the Portfolio of BI Projects	452
Sponsorship, Governance, and Business Participation	453
Build the BI Program Like a Software Firm	455
Balancing a BI program	457
Internal Marketing of the BI Program	459
BI Assessment	460
Discovery Phase—Assess Current Environment	461
Analysis Phase—Identify the Gaps	463
Recommendations Phase—Scope, Priorities, and Budget	465
Work Breakdown Structure	465
WBS Defined	467
WBS Design Principles	467

Multiple WBSs for BI	468
BI Architectural Plan	470
BI Projects Are Different	472
Project Methodologies	473
Battling Project Methodologies	474
The False Choice between Project Methodologies	476
Hybrid BI Project Methodology	477
BI Project Phases	479
Scope and Plan Phase	480
Analysis and Definition Phase	481
Architect and Design Phase	482
Build, Test, and Refine Phase	482
Implementation Phase	483
Deploy and Roll-Out Phase	484
BI Project Schedule	484
BI Swim Lane	487
Data and Database Swim Lane	488
Data Integration Swim Lane	489
Infrastructure Swim Lane	490
Program and Project Management Swim Lane	491
CHAPTER 19 Centers of Excellence	493
The Purpose of Centers of Excellence	493
BI COE	494
Why Businesses Need BI COEs	494
Building a Business Case for a BI COE	496
BI COE Deliverables	498
Skills Needed	499
Organization and Funding	501
Data Integration Center of Excellence	501
Why Businesses Need DI COEs	502
Building a Business Case for a DI COE	504
Organizational Models for Building the DI COE	507
DI COE Deliverables	509
Skills Needed	510
Enabling a Data-Driven Enterprise	511
Reference	512
Index	513