

# **A Manager's Guide to Data Warehousing**

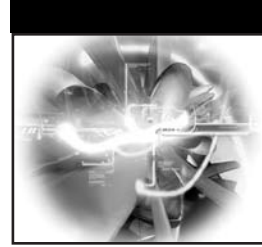
Laura L. Reeves





# **A Manager's Guide to Data Warehousing**





# **A Manager's Guide to Data Warehousing**

---

Laura L. Reeves



Wiley Publishing, Inc.

## A Manager's Guide to Data Warehousing

Published by  
**Wiley Publishing, Inc.**  
10475 Crosspoint Boulevard  
Indianapolis, IN 46256  
[www.wiley.com](http://www.wiley.com)

Copyright © 2009 by Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-0-470-17638-2

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

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, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at [www.wiley.com/go/permissions](http://www.wiley.com/go/permissions).

**Limit of Liability/Disclaimer of Warranty:** The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Web site may provide or recommendations it may make. Further, readers should be aware that Internet Web sites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Library of Congress Cataloging-in-Publication Data

Reeves, Laura L.

A manager's guide to data warehousing / Laura L. Reeves.  
p. cm.

Includes index.

ISBN 978-0-470-17638-2 (paper/website)

1. Data warehousing—Management. I. Title.

QA76.9.D37R44 2009

005.74068—dc22

2009007401

**Trademarks:** Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc. is not associated with any product or vendor mentioned in this book.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.



## About the Author

Laura L. Reeves started designing and implementing data warehouse solutions in 1986. Since then she has been involved in hundreds of projects. She has extensive experience in end-to-end data warehouse development, including developing comprehensive project plans, collecting business requirements, developing business dimensional models, designing database schemas (both star and snowflake designs), and developing enterprise data warehouse architecture and strategies. These have been implemented for many business functions for private and public industry.

Laura co-founded StarSoft Solutions, Inc., in 1995 and has been a faculty member with The Data Warehousing Institute since 1997. She is a contributing author of *Building a Data Warehouse for Decision Support* (Prentice Hall, 1996) and a co-author of the first edition of *The Data Warehouse Lifecycle Toolkit* (Wiley, 1998). Laura graduated magna cum laude from Alma College with a bachelor of science degree in mathematics and computer science, with departmental honors.







## Credits

**Executive Editor**

Robert Elliott

**Development Editor**

Sara Shlaer

**Technical Editor**

Jonathan Geiger

**Production Editor**

Melissa Lopez

**Copy Editor**

Luann Rouff

**Editorial Manager**

Mary Beth Wakefield

**Production Manager**

Tim Tate

**Vice President and Executive  
Group Publisher**

Richard Swadley

**Vice President and Executive  
Publisher**

Barry Pruett

**Associate Publisher**

Jim Minatel

**Proofreader**

Josh Chase, Jen Larsen, and  
Kyle Schlesinger, WordOne

**Indexer**

Robert Swanson

**Cover Image**

© Digital Vision





# Acknowledgments

I have been very blessed with great family, friends, and colleagues. I would like to thank the many clients and colleagues who have challenged me, pushed me, and collaborated with me on so many initiatives over the years. I appreciate the opportunity to work with such high-quality people. I want to acknowledge the contributions that have been made to the data warehousing industry and to me personally by the amazing people who worked at Metaphor. I want to express my gratitude to my dear friend and colleague Paul Kautza for his belief in me and for all his hard work all these years.

Thanks are also due to the dedicated staff at Wiley who believed in me and had great patience to help see this project through. Thanks to Bob Elliott for being the impetus to get this project started and to Sara Shlaer and Rosanne Koneval for their detailed efforts to produce a quality product. I want to express appreciation to Cindi Howson for her insight on business intelligence tools.

I want to extend a sincere and special thank you to Jonathon Geiger for his meticulous comments and suggestions.

I also want to thank two very special people who have provided unflagging support and encouragement every step of the way: my friends Ingrid Korb and Paula Johnson. I am not sure I could have done this without you!

Of course, none of this would be possible without the dedication, sacrifice, love, and support given to me by my family: Mark, Ryan, Michael, and Leah.





# Contents

<b>Introduction</b>	<b>xxiii</b>
<b>Part One    The Essentials of Data Warehousing</b>	<b>1</b>
<b>Chapter 1    Gaining Data Warehouse Success</b>	<b>3</b>
The Essentials of Data Warehousing	3
What Is a Data Warehouse?	4
Differences Between Operational and DW Systems	4
The Data Warehousing Environment	4
What Is a Data Model?	6
Understanding Industry Perspectives	7
Design and Development Sequence	8
Why Build a Data Warehouse?	11
The Value of Data Warehousing	12
The Promises of Data Warehousing	15
Keys to Success	16
Developing and Maintaining Strong Business and Technology Partnerships	17
Identifying True Business Requirements	17
Shifting to a Global Perspective	18
Overcoming Unrealistic Expectations	19
Providing Clear Communication	20
Treating Data As a Corporate Asset	21
Effectively Leveraging Technology	21
Roadblocks to Success	22
Believing the Myth: “If You Build It, They Will Come”	22
Falling into the Project Deadline Trap	23

	Failing to Uphold Organizational Discipline	23
	Lacking Business Process Change	24
	Narrowing the Focus Too Much	25
	Resting on Your Laurels	27
	Relying on the Technology Fix	27
	Getting the Right People Involved	28
	Finding Lost Institutional Knowledge	29
	Summary	30
<b>Chapter 2</b>	<b>The Executive's FAQ for Data Warehousing</b>	<b>31</b>
	Question: What is the business benefit of a data warehouse?	32
	Answer	32
	Question: How much will it cost?	33
	Answer	33
	Question: How long will it take?	34
	Answer	35
	Question: How can I ensure success?	36
	Answer	36
	Question: Do other companies really build these in 90 days?	37
	Answer	37
	Question: How will we know we are doing this right?	38
	Answer	38
	Question: Why didn't this work last time? What is different this time?	39
	Answer	39
	Question: Do we have the right technology in place?	39
	Answer	40
	Question: Are we the only company with data warehouse problems?	40
	Answer	41
	Question: Will I get one version of the truth?	41
	Answer	42
	Question: Why can't we just use our current systems?	43
	Answer	44
	Question: Will the data warehouse replace our old systems?	45
	Question: Who needs to be involved?	45
	Question: Do we know where we are going? How will we know when we get there?	46
	Answer	46
	Question: How do we get started and stay focused?	47
	Answer	47
	Summary	48

<b>Part Two</b>	<b>The Business Side of Data Warehousing</b>	<b>49</b>
<b>Chapter 3</b>	<b>Understanding Where You Are and Finding Your Way</b>	<b>51</b>
	Assessing Your Current State	51
	What Is Your Company's Strategic Direction?	52
	What Are the Company's Top Initiatives?	54
	How Healthy Is Your Data?	55
	Does the Business Place Value on Analysis?	56
	Reflecting on Your Data Warehouse History	57
	Understanding Your Existing Reporting Environment	58
	Finding the Reporting Systems	59
	Compiling an Inventory	60
	Identifying the Business Purpose	61
	Discovering the Data You Already Have	63
	Understanding the People	65
	Tracking Technology and Tools	65
	Understanding Enterprise Resources	66
	Netting It All Out	68
	Introducing the Case Studies	70
	The Call Center Data Warehouse Project	70
	In Real Life	70
	Giant Company	71
	Agile, Inc.	72
	Summary	72
<b>Chapter 4</b>	<b>Successful IT–Business Partnerships</b>	<b>75</b>
	What a Partnership Really Means	75
	What the Business Partners Should Expect to Do	76
	Business Executives and Senior Management	78
	The Executive Business Sponsor	78
	Business Managers	81
	The Business Champion	82
	Business Analysts	83
	Helping the Business Analyst Deal with Change	85
	Business User Audience	86
	Project Manager	86
	What You Should Expect from IT	88
	CIO/IT Executive Sponsor	89
	Data Warehouse Manager	89
	Business Systems Analyst	90
	Source System Analyst	91
	Data Modeler/Data Architect	92

ETL Developer(s)	93
Business Intelligence Application Developer	94
Other Supporting Roles	95
Tips for Building and Sustaining a Partnership	95
Leveraging External Consulting	97
Building Strong Project Teams	98
Effective Communication	99
Netting Out Key Messages	99
Presenting in Business Terms	100
Meeting Preparation	101
Presentation Tips	102
When to Communicate	103
Partnerships Beyond a Project	104
The Decision-Making Process	104
Executive Steering Committee	104
DW Business Support Team	106
Enterprise Considerations	107
In Real Life	107
A Glimpse into Giant, Co.	107
Insight from Agile, Inc.	108
Summary	109
<b>Chapter 5    Setting Up a Successful Project</b>	<b>111</b>
Defining the Project	111
Setting Up the Project Charter	112
Documenting Project Scope	117
Developing a Statement of Work	117
How Much Will It Cost?	120
Project Approval	122
Starting the Project	122
Launching the Project	123
Managing a Successful Project	124
Issue Tracking	124
Using Project Change Control	125
Discussing Change in Business Terms	126
Managing Expectations	128
In Real Life	129
Structured Projects with Giant	129
Freedom for Creativity at Agile, Inc.	130
Summary	131
<b>Chapter 6    Providing Business Requirements</b>	<b>133</b>
What Requirements Are Needed?	134
Peeling Back the Layers of Requirements Gathering	134



Who Provides Input?	137
Who Gathers the Requirements?	137
Providing Business Requirements	138
Strategic Requirements	138
Broad Business Requirements	140
Business Analyses	143
Business Data Requirements	145
Systems and Technical Requirements	147
Communicating What You Really Need	149
What Else Would Help the Project Team?	150
Data Integration Challenges	151
Assess Organizational Motivation	151
Complete Picture of the Data	152
What If No One Is Asking?	152
Practical Techniques for Gathering Requirements	153
Interview Session Characteristics	153
Individual Interviews	153
Group Interviews	153
Project Team Participation	154
Interview Tips	154
Who Needs to Be Included?	155
Setting a Good Example	156
Preparing for Interview Sessions	157
Conducting the Interview Sessions	157
Capturing Content: Notes vs. Tapes	157
Running the Interview	158
Concluding the Interview	158
Putting the Pieces Together	158
Individual Interview Documentation	159
Responsibilities	159
Business Themes	159
Business Data	160
Consolidated Requirements Documentation	161
Executive Summary	161
Consolidated Business Themes	162
Candidate Business Analyses	162
Consolidated Business Data Requirements	162
Identification of Non-Data Warehouse Requirements	163
Common Requirements Gathering Challenges	163
Sifting Through Reports	163
Listing Data Elements	164
Developing Functional Specifications	164
Moving Beyond Immediate	164
Lack of Requirements	165
The Cynic	165

Setting Attainable Goals	166
Exploring Alternatives	167
Setting Priorities	168
In Real Life	170
A Glimpse into Giant Company	170
Insight from Agile, Inc.	170
Summary	171

### **Part Three    Dealing with the Data    173**

#### **Chapter 7    Modeling the Data for your Business    175**

The Purpose of Dimensional Models	176
Ease of Use	176
Query Performance	177
Understanding Your Data	177
What Is a Dimensional Model?	178
Dimensions	178
Facts	180
Using Both Parts of the Model	180
Implementing a Dimensional Model	181
Diagramming Your Dimensional Model	182
The Business Dimensional Model	182
Business Dimensions	183
Fact Groups	184
A Call Center Case Study	186
Call Center Dimensions	187
Date Dimension	187
Time Dimension	187
Customer Dimension	189
Employee Dimension	191
Call Dimension	191
Call Outcome Dimension	194
Employee Task Dimension	195
Call Center Fact Groups	196
Calls Fact Group	196
Call Center Time Tracking Fact Group	196
Call Forecast Fact Group	198
Working with the Model	199
Business Dimensional Model Index	200
Enterprise Considerations	200
Conformed Dimensions	200
Conformed Facts	202
Practical Guidelines	202
Guidelines for a Single Dimension	202

Guidelines for a Single Fact Group	203
Characteristics of the Model across the Enterprise	204
Business Participation in the Modeling Process	205
Creating the First Draft	205
Preparing for Modeling Sessions	205
Brainstorming the Framework	206
Drafting the Initial Dimensions	206
Drafting the Initial Fact Groups	207
Documenting the Model	208
Logging Questions and Issues	208
Building the Business Measures Worksheet	209
Preliminary Source to Target Data Map	211
Completing or Fleshing Out the Model	211
Working Through the Issues	211
Completing the Documentation	212
Working Through All the Data Elements	212
Refining the Model	213
Business Reviews of the Model	213
Small Business Reviews	214
When Are You Done?	214
Gaining Final Commitment	215
Expanding Business Data Over Time	215
Enhancing Dimensions	215
Adding More Fact Groups	215
Reflecting on Business Realities: Advanced Concepts	216
Supporting Multiple Perspectives: Multiple Hierarchies	216
Tracking Changes in the Dimension: Slowly Changing Dimensions	216
Depicting the Existence of a Relationship: Factless Fact Tables	218
Linking Parts of a Transaction: Degenerate Dimensions	219
Pulling Together Components: Junk Dimensions	221
Multiple Instances of a Dimension: Role Playing	222
Other Notation	224
Dimension Connectors	224
Clusters of Future Attributes	225
Notation Summary	225
Taking the Model Forward	225
Translating the Business Dimensional Model	226
Dimension Table Design	226
Translating Fact Groups	227
Physical Database Design	228
In Real Life	228
A Glimpse into Giant Co.	229
Insight from Agile, Inc.	229
Summary	230

<b>Chapter 8</b>	<b>Managing Data As a Corporate Asset</b>	<b>231</b>
	What Is Information Management?	232
	Information Management Example—Customer Data	235
	IM Beyond the Data Warehouse	239
	Master Data Management	240
	Master Data Feeds the Data Warehouse	242
	Finding the Right Resources	242
	Data Governance	243
	Data Ownership	243
	Who Really Owns the Data?	244
	Your Responsibilities If You Are “the Owner”	246
	What are IT’s Responsibilities?	247
	Challenges with Data Ownership	247
	Data Quality	248
	Profiling the Data	249
	How Clean Does the Data Really Need to Be?	250
	Measuring Quality	250
	Quality of Historical Data	251
	Cleansing at the Source	253
	Cleaning Up for Reporting	254
	Managing the Integrity of Data Integration	254
	Quality Improves When It Matters	256
	Example: Data Quality and Grocery Checkout	
	Scanners	257
	Example: Data Quality and the Evaluation of Public	
	Education	257
	Realizing the Value of Data Quality	258
	Implementing a Data Dictionary	259
	The Data Dictionary Application	259
	Populating the Data Dictionary	261
	Accessing the Data Dictionary	263
	Maintaining the Data Dictionary	263
	Getting Started with Information Management	264
	Understanding Your Current Data Environment	264
	What Data Do You Have?	265
	What Already Exists?	266
	Where Do You Want to Be?	267
	Develop a Realistic Strategy	268
	Sharing the Information Management Strategy	269
	Setting Up a Sustainable Process	270
	Enterprise Commitment	270
	The Data Governance Committee	270
	Revising the Strategy	271

	In Real Life	271
	A Glimpse into Giant, Co.	272
	Insight from Agile, Inc.	272
	Summary	274
<b>Part Four</b>	<b>Building the Project</b>	<b>275</b>
<b>Chapter 9</b>	<b>Architecture, Infrastructure, and Tools</b>	<b>277</b>
	What Is Architecture?	278
	Why Do We Need Architecture?	278
	Making Architecture Work	281
	Data Architecture	282
	Revisiting DW Goals	283
	Components of DW Data Architecture	285
	A Closer Look at Common Data Warehouse Architectures	286
	Bottom-Up Data Architecture	286
	Top-Down Data Architecture	290
	Publish the Data: Data Marts	294
	Adopting an Architecture	295
	Technical Architecture	297
	Technical Architecture Basics	298
	Components of Technical Architecture	299
	Infrastructure	300
	Technical Architecture in Action	300
	What You Need to Know about Technical Architecture	301
	Navigating the Technology Jungle	302
	Weighing Technology Options	303
	Best of Breed	303
	End-to-End Solutions	303
	Deciding Not to Buy a Tool	304
	Finding the Right Products	304
	Requests for Information or Proposals	305
	Business Participation in the Selection Process	305
	Understanding Product Genealogy	306
	Understanding Value and Evaluating Your Options	306
	Cutting through the Marketing Hype	308
	The Value of References	309
	Making Architecture Work for You	310
	Just-In-Time Architecture	311
	In Real Life	311
	Architecture at Giant	311
	Agile Ignores the Need for Architecture	312
	Summary	313

<b>Chapter 10   Implementation: Building the Database</b>	<b>315</b>
Extract, Transform, and Load (ETL) Fundamentals	315
What Work Is Being Done?	315
ETL System Functionality	317
Extraction	318
Transformation	318
Load	322
The Business Role in ETL	323
Why Does the Business Need to Help?	323
Defining Business Rules	324
Defining Expected Results—The Test Plan	325
Development Support	326
Testing the ETL System—Is the data Right?	326
Why Does It Take So Long and Cost So Much?	327
Balancing Requirements and Data Reality	329
Discovering the Flaws in Your Current Systems	330
Applying New Business Rules	331
Working Toward Long-Term Solutions	332
Manually Including Business Data	333
Tracking Progress—Are We There Yet?	333
What Else Can You Do to Help?	334
Encouragement and Support	334
Ensuring Continued Business Participation	335
Proactive Communication	336
In Real Life	337
Building the Data Warehouse at Giant, Co.	337
Agile, Inc., Builds a Data Warehouse Quickly	338
Summary	339
 <b>Chapter 11   Data Delivery: What you Finally See</b>	 <b>341</b>
What Is Business Intelligence?	341
Business Intelligence without a DW	342
BI in Action	343
Tabular Reports	343
Parameter-Driven Reports	343
Interactive Reports—Drilling Down and Across	344
Exception Reports	344
Other BI Capabilities	345
Complex Analysis	345
BI Building Blocks	346
Data Content—Understanding What You Have	346
Navigation—Finding What You Need	347

Presentation—How Do You Want to See Results?	347
Delivery—How Do You Receive the Results?	351
Supporting Different Levels of Use	352
Construction of the BI Solution	354
Planning for Business Change	354
Design—What Needs to Be Delivered?	355
Development	357
Testing BI Applications and Validating Data	358
Additional Responsibilities	359
Security—Who Can Look at the Data?	359
System Controls—Who Can Change What?	360
Planning a Successful Launch	361
Marketing the Solution	361
Learning to Use the Data without a Technical Degree	362
Learning about the Data	362
Learning about the BI Tool/Application	362
Ensuring That the Right Help Is Available	363
In Real Life	364
BI at Giant Company	364
Agile, Inc. Dives into BI	365
Summary	366
<b>Part Five   Next Steps—Expanding On Success</b>	<b>367</b>
<b>Chapter 12   Managing the Production Data Warehouse</b>	<b>369</b>
Finishing the Project	369
Recapping the BI Application Launch	369
Post-Implementation Review	370
Looking Back—Did you Accomplish Your Objectives?	371
Adopting the Solution	371
Tracking Data Warehouse Use	372
Getting the Rest of the Business Community on Board	372
Business Process Change	374
Changing How Data Is Used	374
Streamlining Business Processes	374
Encouraging Change	375
The Production Data Warehouse	375
Staffing Production Activities	376
Maintaining the Environment	376
Keeping Up with Technology	376
Monitoring Performance and Capacity Planning	378
Maintaining the Data Warehouse	380
Maintaining the ETL System	380
Maintaining the BI Application	381

Tracking Questions and Problems	382
Fixing Bugs	384
When the Data Warehouse Falls Short	384
Common Causes for a Stalled Warehouse	385
Jump-Starting a Stalled Data Warehouse	388
Conducting an Assessment	388
Determining What Can Be Salvaged	389
Developing a Plan to Move On	390
Aligning DW Objectives with Business Goals	391
Getting It Right This Time	392
Launching the Improved Data Warehouse and BI Solution	393
In Real Life	394
Lack of Support for the Production DW at Giant Co.	394
Unleashing BI at Agile, Inc.	395
Summary	396
<b>Chapter 13 Achieving Long-Term Success</b>	<b>397</b>
Planning for Expansion and Growth	397
Exploring Expansion Opportunities	398
Prioritization of Feedback	399
Managing Enterprise DW Resources	400
Creating an Enterprise Data Warehouse Team	400
The Centralized Enterprise Data Warehouse Team	401
The Virtual Enterprise Data Warehouse Team	401
Enterprise DW Team Responsibilities	403
Funding the Enterprise DW Team	404
Pushing into the Future	405
Embedded Business Intelligence	405
Operational Business Intelligence	406
Real-Time Data Warehousing	407
Unstructured Data	408
Monitoring Industry Innovation	409
Moving Toward Business Value	410
Measuring Success One Step at a Time	410
Adjusting Expectations to Reality	412
Keeping the Momentum Going	413
Celebrating Progress	416
Success Can Be Attained	417
Conclusion	419
<b>Glossary</b>	<b>421</b>
<b>Index</b>	<b>429</b>





# Introduction

Many executives, managers, business analysts, and nontechnical personnel are highly motivated to learn more about data warehousing. They want to understand what data warehouses are and how they work. More important, many are truly interested in doing their part to ensure success when implementing a data warehouse in their company. They are not interested in learning how to write code or tune a database.

Unfortunately, most data warehouse publications available today are written for the people who design and build them. Some are from a project management perspective and others provide a great deal of technical depth. While these are very valuable to the technical team, they do not help the nontechnical audience. This book was written to provide a resource for those nontechnical people.

## Overview of the Book

---

The information in this book has been gathered over years of working on data warehouse projects. Hundreds of hours have been invested in learning what works well and what does not. One constant thread over the years is the need to develop and strengthen the partnership between business and systems personnel. There has always been a need to help nontechnical people in the organization understand the different parts of a data warehouse and what needs to be done to build and maintain one.

This book covers the topics and questions that come up repeatedly in executive briefings, classes, meetings, and casual conversation. It also includes coverage of topics related to how organizations frequently get into trouble.

The goal is to minimize the frustration of all participants and to ultimately help organizations to build and maintain valuable data warehouse environments.

The book provides a sound introduction to data warehousing concepts and then moves on to explain the process of developing and creating a data warehouse project. A description of each step is provided, including details about how the business should participate. This book does not provide the technical level of detail that IT team members need to know, such as specific coding techniques or how to best set parameters to improve performance with a specific technology. The book does provide what is needed to understand these areas so that you are better prepared to have clear communication across the organization. By knowing what is being designed or developed, you can be more effective in sharing your ideas, requirements, and concerns.

More technical readers will benefit by gaining a complete picture of data warehousing, with an emphasis on how representatives from the business community can help you be more successful. This will provide you with ideas about how to interact more clearly with the business community. Use this as a resource; highlight the most pertinent chapters or sections that would be beneficial for your business counterparts to read.

## **How This Book Is Organized**

---

This book is divided into five major parts. Each part focuses on a specific aspect of data warehousing.

- The first part shares the essentials of data warehousing. This can help executives and managers get a realistic understanding of the fundamentals and gain insight into both the most common factors for success and how to avoid roadblocks.
- The second part of the book takes a look at the business management of a data warehouse. This includes how to cultivate a strong partnership between the business and IT communities, what is entailed in setting up a data warehouse project, and how to effectively communicate your business requirements for the data warehouse. It is just as important for different business groups to work together in order to support a consistent view of the data.
- The third part of the book focuses on the data itself. How the data is organized is critical to ensure that you, the business, can understand and exploit what is available. Without the right data, presented in a useful manner, the data warehouse will not be used, which jeopardizes the entire investment. This section of the book gives you the tools you need to ensure that the data meets your needs. Beyond how data is processed or

stored, the issues surrounding how organizations view data ownership and management are also addressed.

- Part four delves into data warehouse architecture, what is involved to efficiently build the data warehouse, and finally what must be done to deliver the data. While most of the work during this part of a project is more technical in nature, there is still a need for active business participation. This section will help managers and business participants understand what work is being done and how they can help.
- Part five explores what is needed to launch the data warehouse and wrap up the project. Looking beyond a single project, the work needed to maintain and grow the data warehouse is discussed. The causes of, and recommendations to address, a stalled data warehouse are presented. A data warehouse can have initial success that decreases over time unless specific steps are taken. Suggestions are provided that can help you sustain your success.

## Who Should Read This Book

---

The target audience for this book is anyone who is responsible for, working on, or paying for a data warehouse. In particular, it is targeted toward nontechnical readers in order to help them understand the basics of data warehousing, and, more importantly, how to successfully build one. For the more technical readers, this book provides the tools you need in order to be able to communicate with your business partners.

The information is presented in a layered manner. The basic concepts are presented early in the book, while more in-depth coverage is provided in later chapters. This book is designed to serve a variety of levels of need. Some readers may benefit from reading only the first part, returning later to learn more about a specific area. Other readers may benefit by reading all of the content from start to finish.

While all readers could benefit from reading the entire text, I realize that not everyone has the same passion for data warehousing. Realistically, different types of readers will benefit from various parts of the book:

- Executives and senior managers should read Part 1. Then, based upon what the organization is facing, subsequent chapters may be worthwhile as different issues crop up. For example, if your company has problems getting the data loaded into the data warehouse, it would be helpful to read Chapter 10 to get a better understanding of what is really involved in building the database.

- Middle managers, both business and IT, will benefit from Part 1 but will also find Part 2 to be important. These parts help you get the project set up properly and, most important, learn about how to provide business requirements. Skimming the rest of the book can help all managers learn what is available to help them with the rest of the project. Then, as needed, the specific chapters can be studied in detail when the organization is working on that part of a data warehouse initiative.
- Business personnel involved with a data warehouse can skim the chapters on setting up a project, but will find it helpful to study in depth Chapter 6 to learn about providing requirements, Chapter 7 to understand how the data should be organized, and Chapter 11 to learn about how the data can be delivered. It is recommended that the rest of the book be reviewed to familiarize yourself with the content. Specific chapters can be studied in more detail when the organization is working on that area.
- Everyone on the data warehouse project team should read this entire book. It can provide a common ground for dialogue and more meaningful discussions between the business and technical personnel. The project team can provide this book to their business counterparts and suggest specific chapters for them to read to better support the project.
- Technical staff can also benefit from this book, which can help anyone with in-depth technical knowledge to communicate with their managers and business counterparts about data warehousing. This provides the backdrop for all interactions with the business community.

When designed and implemented properly, a data warehouse is a valuable tool for an organization to improve how it runs. Getting the right design requires active participation of knowledgeable business personnel. Keeping things on track requires the support of middle managers to ensure that everything is progressing smoothly. Executives and senior managers will be able to ask meaningful questions about data warehouse projects and be able to understand the answers. All of these things require that business and technical staff have a common understanding of the different parts of a data warehouse and what is involved in a successful project. It is hoped that this book provides the foundation you and your organization need to achieve success with all of your data warehouse initiatives.

# **A Manager's Guide to Data Warehousing**



**PART**

**I**

# The Essentials of Data Warehousing

---

## In This Part

---

**Chapter 1:** Gaining Data Warehouse Success

**Chapter 2:** The Executive's FAQ for Data Warehousing





# **Gaining Data Warehouse Success**

You've heard about them. You may have used one. You may be asked to pay for one. But what is a data warehouse and why should you invest any time, energy, and money on one? The short answer is that a data warehouse can help your organization to be more profitable, run more efficiently, and meet the challenges of today's marketplace. Yet it is not a quick, simple, or inexpensive undertaking to build a data warehouse.

There is often a disconnect between the technical side that builds and maintains a data warehouse and the business side that will use it. This book will help bridge that gap. Both business managers and IT managers will learn what is involved with building and deploying a successful data warehouse. Executives and senior managers will also find this book helpful, especially Part 1, in order to be able to provide effective oversight and support. This book will also be beneficial for all business and technical personnel involved with a data warehouse, providing a common foundation for better communication. Managers on both sides need the knowledge and information that will enable them to help their organization build and use a data warehouse most effectively, and this book is the path to that knowledge.

This chapter explains the value of a data warehouse and highlights what is needed for success. To help frame the discussion, the chapter begins with some definitions.

## **The Essentials of Data Warehousing**

---

Data warehousing is not new. Most large organizations have been investing in data warehousing for years. Currently, cost-effective technology is creating more possibilities for small and medium-size companies to build and deploy data warehouse solutions too. There are many stories about wild successes, and just as many about failed projects. With so much buzz about data warehousing, it is often assumed that everyone already knows the basics. However, many

people are being exposed to these concepts for the first time. To ensure a common understanding, it is worth taking the time to boil things down to the essence of data warehousing.

## **What Is a Data Warehouse?**

A *data warehouse (DW)* is the collection of processes and data whose overarching purpose is to support the business with its analysis and decision-making. In other words, it is not one thing per se, but a collection of many different parts. Before looking more closely at the specific parts of a data warehouse environment, it is helpful to compare the characteristics and purpose of a data warehouse with an operational application system.

### ***Differences Between Operational and DW Systems***

Applications that run the business are called *online transaction processing systems (OLTPs)*. OLTP systems are geared toward functions such as processing incoming orders, getting products shipped out, and transferring funds as requested. These applications must ensure that transactions are handled accurately and efficiently. No one wants to wait minutes to get cash from an automated teller machine, or to enter sales orders into a company's system.

In contrast, the purpose and characteristics of a data warehousing environment are to provide data in a format easily understood by the business community in order to support decision-making processes. The data warehouse supports looking at the business data over time to identify significant trends in buying behavior, customer retention, or changes in employee productivity. Table 1-1 lays out the primary differences between these two types of systems.

The inherent differences between the functions performed in OLTP and DW systems result in methodology, architecture, tool, and technology differences. Data warehousing emerged as an outgrowth of necessity, but has blossomed into a full-fledged industry that serves a valuable function in the business community.

Now that the differences between data warehouse and OLTP systems have been reviewed, it is time to look deeper into the makeup of the data warehouse itself.

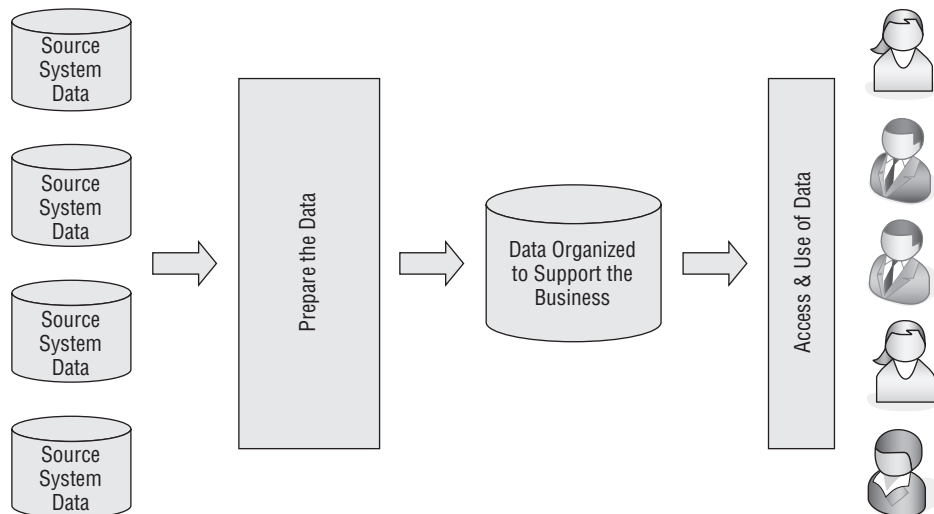
### ***The Data Warehousing Environment***

There are many different parts of a data warehouse environment, which encompasses everything from where the data lives today through where it is ultimately used on reports and for analysis. Each of the main parts of the data

warehousing environment, shown in Figure 1-1, are described in the following sections. This figure indicates how the data flows throughout the environment.

**Table 1-1** Comparison of Online Transaction Processing and Data Warehousing Systems

AREA OF COMPARISON	OLTP	DATA WAREHOUSING
System purpose	Support operational processes	Support strategic analysis, performance, and exception reporting
Data usage	Capture and maintain the data	Exploit the data
Data validation	Data verification occurs upon entry	Data verification occurs after the fact
Update frequency	Data is updated when business transactions occur (e.g., client uses debit card, web order is placed)	Data is updated by periodic, scheduled processes
Historical data requirement	Current data	Multiple years of history
Data integration and balancing	Data is balanced within the scope of this one system	Data must be integrated and balanced from multiple systems



**Figure 1-1** Basic data warehousing environment

*Source systems*, shown on the left side of Figure 1-1, are where data is created or collected by operational application systems that run the business. These are often large applications that have been in place for a long time. Examples of source systems include the following:

- Order processing
- Production scheduling
- Financial trading systems
- Policy administration
- Claims handling
- Accounts payable/receivable
- Employee payroll

The entire midsection of Figure 1-1 is devoted to the preparation and organization of data. First, the data must be extracted from the source systems. Next, the data needs to be transformed to prepare it for business use. It must be cleansed, validated, integrated together, and reorganized. Finally, the data is loaded into structures that are designed to deliver it to the business community. The entire process is referred to as the *extract, transform, and load (ETL)* process.

The database in which the data is organized to support the business is called a *data mart*. A data mart includes all of the data that is loaded into a single database and used together for analysis. Data marts are often developed to meet the needs of a business group such as marketing or finance. The key to a successful data mart is to create it in an integrated manner. It is also recommended that data be loaded into only one data mart and then shared across the organization to ensure data consistency.

Finally, an application or reporting layer is provided to facilitate access and analysis of the data. This is where business users access reports, dashboards, and analytical applications. Collections of these reports and analyses are called *business intelligence*.

There is one more critical concept that warrants some attention: the mechanism used to help organize data, which is called a *data model*.

### ***What Is a Data Model?***

A data model is an abstraction of how individual data elements relate to each other. It visually depicts how the data is to be organized and stored in a database. A data model provides the mechanism for documenting and understanding how data is organized.

There are many different types of data modeling, each with a specific goal and purpose. As organizations modified how data was structured to

support reporting and analysis, a new data modeling technique, now called *dimensional modeling*, emerged. Ralph Kimball, a pioneer in data warehousing, can be credited with crystallizing these techniques and publishing them for the benefit of the industry. (For more information about dimensional modeling, see Chapter 7.)

The data and processes to perform the work shown in Figure 1-1 are collectively called the data warehouse. These basic concepts have been fine-tuned and relabeled by many different players in the data warehousing field. The two most common philosophies are discussed in the next section.

## ***Understanding Industry Perspectives***

At the end of the day, everyone faces the same challenge: getting the data into the hands of the business user to turn it into information that can be used to make decisions. The definitions provided so far provide the backdrop for how terms are used in this book. There are many brilliant and talented people in the data warehousing industry, many of whom have different philosophies about how to design and build a data warehouse. It is worthwhile to look more closely at two of the most frequently used philosophies.

The first is from Ralph Kimball and colleagues, as described in *The Data Warehouse Lifecycle Toolkit, Second Edition* (Wiley, 2008). Ralph Kimball is a clear thought leader in the data warehousing industry and has written several books that provide detailed information essential for practitioners. That book describes the enterprise data warehouse as:

*The complete end-to-end data warehouse and business intelligence system (DW/BI System). Although some would argue that you can theoretically deliver business intelligence without a data warehouse and vice versa, that is ill-advised from our perspective. Linking the two together in the DW/BI acronym reinforces their dependency. Independently, we refer to the queryable data in your DW/BI system as the enterprise data warehouse, and value-add analytics as BI (business intelligence) applications.*

A second definition worth looking at is from Bill Inmon, a prolific author and another leader in the data warehousing industry, from *Building the Data Warehouse, Fourth Edition* (Wiley, 2005):

*The data warehouse is a collection of integrated subject-oriented databases designed to support the DSS (decision support system) function, where each unit of data is relevant to some moment in time. The data warehouse contains atomic data and lightly summarized data.*

Bill's definition is also described and expanded by Claudia Imhoff and colleagues in *Mastering Data Warehouse Design* (Wiley, 2003):

*... It [the DW] is the central point of data integration for business intelligence and is the source of data for the data marts, delivering a common view of enterprise data.*

This second viewpoint is incomplete without also including their definition of a data mart. Again, expanding on Bill Inmon's definition, Claudia states in *Mastering Data Warehouse Design*:

*A data mart is a departmentalized structure of data feeding from the data warehouse where data is denormalized [organized] based on the department's need for information. It utilizes a common enterprise view of strategic data and provides business units with more flexibility, control, and responsibility. The data mart may or may not be on the same server or location as the data warehouse.*

To bring this second viewpoint into the proper context, *Mastering Data Warehouse Design* further defines business intelligence:

*Business intelligence is the set of processes and data structures used to analyze data and information used in strategic decision support. The components of Business Intelligences are the data warehouse, data marts, the DSS (decision support system) interface and the processes to 'get data in' to the data warehouse and to 'get information out'.*

The single definition provided by Ralph Kimball is comprehensive. You must look at the full set of definitions set forth by Bill Inmon and Claudia Imhoff to fully understand their perspective. There is much more common ground between these differing philosophies than perceived at first glance. While there are distinct differences, the common theme is that data warehousing must provide the method to prepare and deliver data to the business community to support reporting and analysis. Chapter 9 provides a comprehensive discussion about these different approaches to data warehousing.

The key point here is that there are multiple ways these terms can be interpreted. Understanding which definition is being used is critical to being able to understand what is being discussed and worked on. An organization can avoid confusion by selecting one set of definitions to be used, which enables everyone to use a common language.

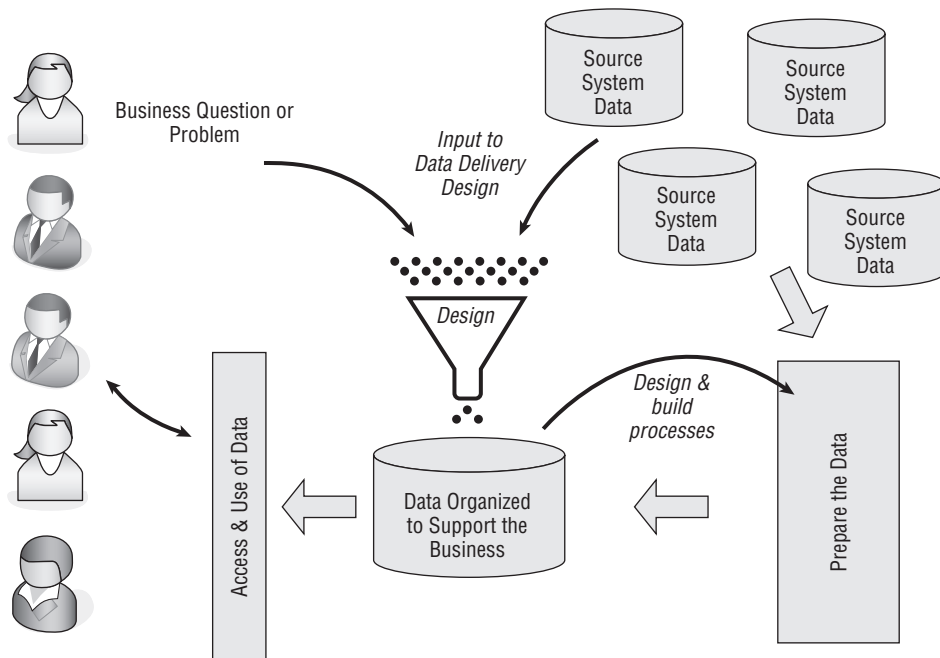
Regardless of labels and terminology, all data warehouse initiatives are trying to accomplish the same thing. Now that the basic parts of the data warehouse have been defined, it is time to look at the order in which they are created.

### ***Design and Development Sequence***

Earlier in this chapter, you looked at how data flows through the data warehouse environment. While this correctly illustrates how data flows in the

completed environment, this is not the recommended sequence for designing and developing a data warehouse. A better way to design the environment is to start from the business user perspective. Figure 1-2 shows the correct order to successfully design and implement a data warehousing environment. Both the technical and business team members play a role throughout. Chapter 4 describes the different roles and responsibilities. Each step in the design process is described as follows:

1. An understanding of what the business is trying to accomplish and how success is measured should be the foundation for all data warehousing initiatives. The starting point for designing the data warehouse is with the business community. Chapter 6 covers what you need to know to effectively provide business requirements.
2. Once the business requirements are understood, the data in the underlying source systems needs to be studied. Many business people have a vision for what they want to do, but it is not always tied to the reality of the organization's actual data. In preparation for modeling data, Chapter 7 introduces techniques to help you understand your data.



**Figure 1-2** Optimal data warehouse design and development sequence

3. The foundation for successful data warehousing, now and into the future, is properly structuring the data. Data must be organized to support the business perspective. This provides ease of use and improved query performance. This design is created based on a knowledge of the business requirements, as well as the reality of the existing data. Chapter 7 focuses on how the business and technical team members can work together to develop appropriate data models for this data delivery layer.
4. After defining how the data will be organized, the design for getting the data from the source systems to the database can be created. Decisions about the architecture and tools needed to prepare the data can be made in the proper context. Too often these decisions are made before you know what is to be delivered. Chapter 9 provides the background needed to understand data warehouse architecture and technology. Chapter 10 deals with the challenges of preparing the data for business reporting and analytical use.
5. While the data is being prepared, the data access and application layer can be designed. This includes the design of basic reports, business intelligence, and analytical applications, and performance dashboards or other end user tools. Chapter 11 focuses on final delivery of data to the business community.

Many different project methodologies are available for all systems' development efforts. There are even multiple methodologies specifically targeted toward data warehousing. These have evolved over several decades. Most organizations already have adopted some type of project methodology or project life cycle. It is important to understand how your organization runs projects to ensure that the data warehouse project is adhering to the strategic direction for all information systems. Several basic building blocks are found in any methodology. These primary components, discussed throughout the book, are as follows:

- **Project definition, planning and management** is outlined in Chapters 4 and 5.
- **Defining business requirements** is discussed in Chapter 6.
- **Designing the data delivery database** is covered in Chapter 7.
- **Defining the architecture** is discussed in Chapter 9 and includes development of the database
- **Processes for building the database** are reviewed in Chapter 10.
- **Developing reports/analyses and providing education/support** is presented in Chapter 11 and includes deployment of the final results.



These basic components need to be in place regardless of the chosen methodology, approach, philosophy, or technology. The sequence outlined earlier helps to ensure overall success by tying all other activities to business requirements. This sequence also helps build a foundation that will withstand the test of time.

## Why Build a Data Warehouse?

Since the first application systems were built, businesses have struggled with the proliferation of data. Systems have been implemented to automate business processes such as order processing, policy administration, shipping, and manufacturing scheduling. Each of these applications captures a lot of data. This data has the potential to be used in several ways:

- To understand what is really happening in the business
- To identify historical trends
- To predict future opportunities
- To measure performance

Originally, business requests for data were run directly against production application systems. This was not an efficient way to gather data. For example, a request for data might query which days over the last three years have the highest ATM transaction rates in the Midwest. Unfortunately, in order to gather this data, the data query slowed—or worse, interrupted—the ability of the application systems to perform their primary function. It is unacceptable to reduce the speed with which the ATMs complete transactions, even to support the corporate marketing department. Needless to say, the ability to run these reports against the production application systems was revoked.

However, understanding the results of such a request was valuable, so a copy of the production data was made to enable reporting and basic analysis to be done without affecting the actual operational systems. When working with a copy, modifications were made to how the data was stored in order to make reporting easier. In addition, if historical data was not retained in the operational systems, then the history was kept in the “copy”. New system design techniques were developed to improve how the copy of the data was created and organized. Specific hardware, software tools, and technologies also emerged. These have grown into what we know as the data warehousing industry today.

In spite of these technical advancements, organizations still struggle with some basic data issues:

- **Direct data access:** You can’t directly access the data; you have to request it from a data systems department.

- **Data usefulness:** You can't get the data you need; what is collected is not useful for your purposes.
- **Poor data quality:** The data, if and when you get it, is often not correct. You can't rely on it.
- **Facilitating exception reporting:** There is too much data; you have to dig through it to find what's important.
- **Timeliness of data:** By the time reports get to you, the window of opportunity no longer exists.
- **Flexibility:** You need total flexibility to look at data from a variety of perspectives.
- **Data integration:** You need to perform analysis across all lines of business but the data is stored all over, so you can't easily compare it.
- **Silo reporting environments:** Different departments use different reporting tools and you can't share reports.
- **Unclear definitions of data:** Everyone has their own report, but the figures do not match and you don't know why.

These basic data problems that drove the initial work in the data warehousing industry are the same core problems facing businesses today. If anything, the problems are compounded by ever-increasing mountains of data. While addressing basic data-related issues is certainly a big part of the rationale to build a data warehouse, there are also bigger, more critical business issues that need to be addressed. The data warehouse can provide a vehicle to deliver the information needed to help a business address these issues, which is what drives the true value.

## **The Value of Data Warehousing**

Data warehousing is much bigger than simply delivering reports in a timely manner. It is not the data, the technologies, or the reports that impact the business. Rather, it is the ability of your staff to harness the information to make better, fact-based, insightful decisions. The data warehouse is simply a tool that enables your staff to be more effective. The types of things that can be done using a data warehouse include the following:

- **Tracking and trending key performance indicators:** A data warehouse can provide reports that indicate which product lines are popular in various regions, which employees have generated the most sales, or whether certain ad campaigns are correlated with successful sales.
- **Measuring business performance:** Using reports from the data warehouse, actual and forecasted performance can be compared. For example,

claims managers can see how close they are to reaching the target of making a first payment on claims within the first ten days of opening a new claim.

- **Reporting and understanding financial results:** A data warehouse can help identify departments that have exceeded their monthly budgets, highlight suppliers who have consistently met profitability goals, and single out products that have contributed the most (or least) to the bottom line.
- **Understanding customers and their behavior:** Exception reports that highlight changes in consumer purchase patterns can help identify shifts in the marketplace or erosion of brand loyalty. For example, early identification of changes in payment patterns might indicate that a customer is under financial pressure and could benefit from a courtesy call to prevent more serious problems.
- **Identifying high-value customers:** Using the data warehouse to identify the lifetime value of customers helps with the development of loyalty programs and improves customer service. Some customers may generate many business transactions, but they may not actually be profitable. Other customers may contribute consistently to the organization's profits without a lot of hands-on interaction or support.
- **Attracting and retaining high-value customers:** Data warehouse reports can help you to develop a profile of high-value customers so that initiatives can be created to seek out new customers with a similar profile. This may mean offering low-cost incentives early on so that the organization has the opportunity to develop a strong long-term relationship.
- **Better selection or development of new products:** Having integrated data in a common place, the data warehouse can help streamline the product development process by enabling all groups involved to quickly access market research test results, product packing cost scenarios, and projected product sales.
- **Understanding which products should be scaled back or eliminated:** Using the data warehouse, reports can be generated to highlight products with lagging sales. Additional analyses can be run to determine the cost effectiveness of continuing to carry these items in stores. The data warehouse can also be used to help develop plans identifying when trendy items should be marked down to clear out any remaining inventory.
- **Understanding business competitors:** The data warehouse can provide reports to compare internal sales volumes with external competitor sales

figures. This can help identify fluctuations in the overall marketplace and how well the organization is maintaining its market share.

- **Identifying opportunities to improve business flow and processes:** The data warehouse can be used to track how business transactions flow within the organization to identify bottlenecks, the need for more training, or when systems capacity can no longer keep up with demand.
- **Understanding the impact of highly qualified professionals:** A data warehouse can also be used effectively in not-for-profit scenarios. For example, data warehouse reports can help identify teachers who meet specific criteria and to track how a teacher's students perform on educational assessments over time.

These uses apply across many different industries. For example, industries that have realized data warehousing success include the following:

- Consumer packaged goods
- Financial services
- Manufacturing
- Utilities and telecommunications
- Pharmaceuticals
- Insurance
- Healthcare
- Shipping and transportation
- Educational institutions
- Nonprofit organizations

Furthermore, data warehousing has been successfully deployed across a wide variety of business functions, such as sales, marketing, finance, purchasing, manufacturing quality, human resources, inventory management, customer relationship marketing, call centers, and more.

Providing timely, reliable data enables business professionals to make informed decisions. It is not enough to provide a static report; the information must be flexible enough to make possible the identification of anomalies, so that interesting results can be further explored. This easy exploration enables business users to get to the bottom of problems, allowing corrective action to be taken quickly. Sometimes the data demonstrates unexpected positive results; these also warrant further research to determine whether they indicate extenuating circumstances or whether the results can be repeated.

The return on any individual decision may be small. A lot of small, better decisions can add up to a significant impact on the bottom line. While there have certainly been cases, often highly publicized, where a single decision

resulted in multimillion-dollar results, it is more likely for a business to see benefits from many small decisions. The data warehouse provides the keys that unlock the data's value.

There are several common expectations regarding what a data warehouse should be able to do. Many organizations fall short of meeting these expectations, so it is worth taking some time to understand these expectations and what needs to be in place in order to achieve them.

## The Promises of Data Warehousing

---

Since its inception, data warehousing has offered the promise of helping to improve your business. A data warehouse is expected to provide both of the following:

- A single version of the truth
- The capability to access all data whenever it is needed

Unfortunately, many organizations have invested millions of dollars in data warehousing without realizing either of these goals. There are many reasons for these struggles and failures, many of which are addressed in this book. First, it is worth looking a little more closely at these expectations.

The idea of having a single version of the truth is appealing because so much time and energy is wasted in chasing down discrepancies. It is reasonable to want to trust report results so that decisions based upon those reports are sound. A data warehouse can indeed provide a single repository for all of your data, but that alone is not enough to ensure that all reports will be consistent. Clean, trusted data from the data warehouse is often pulled out, further manipulated, possibly loaded into yet another database, and finally presented on reports. The use of different formulae for calculations, and the criteria used to include or exclude data from the result set, can cause significant differences in what is shown on a report. The effort of loading data into a data warehouse is not enough to fulfill the promise of a single version of the truth. This requires organizational discipline and a commitment to data management, which is discussed in detail in Chapter 8.

The second big promise of data warehousing is that any data that is desired will be available at your fingertips whenever it is needed. Indeed, a data warehouse can make data more accessible to many different types of users across an organization, but it is too expensive to load all of the company's data into a data warehouse. The audience and business impact of some data is not significant enough to justify the expense of including it in the data warehouse. An organization must determine what data is needed to help the business decision-making processes. Then, the most useful data can be loaded

into the data warehouse over a period of time, perhaps even a number of years. Often this is viewed as the final goal: The data is available in the data warehouse.

However, data in a database does not automatically mean that it is accessible to the business community. The data must be made available through reports, dashboards, or analysis tools that are combined with appropriate education about how to leverage the data as part of the day-to-day decision-making processes.

While sound database design and the use of technology are important, many other factors need to be addressed in order to achieve success. In the next section, you'll learn about the most common roadblocks to this success, and how to remove them.

## Keys to Success

There are several key factors for success in building a data warehouse. These factors are relevant regardless of the industry or size of company, and have remained constant over the years. As mentioned earlier, some organizations have invested millions of dollars to obtain the best of everything, and yet struggle to deliver any value to the business. Conversely, some organizations have built a data warehouse with little more than bailing wire and an abacus and have achieved great success. While technology and tools certainly play a role in building a data warehouse, they are clearly not the primary factors for success or failure.

### PRIMARY FACTORS FOR SUCCESS

**Although several factors contribute to a successful data warehouse, two are primary:**

- **A strong partnership between the business and systems communities**
- **Ensuring that the data warehouse project is driven by true business requirements**

**These two major success factors have more to do with people than technology. Take a look at any project plan. How many tasks are focused on communication—that is, people—compared to technical tasks? These technical tasks are indeed important, but specific actions to ensure meaningful communication are critical to the project's success. All other efforts can amount to nothing if the business community and management are not committed.**

## ***Developing and Maintaining Strong Business and Technology Partnerships***

It is one thing to affirm the importance of a strong partnership between the business stakeholders and the technology units involved in a data warehouse project, but it is another thing to put that into practice.

Partnerships often start out strong in the beginning. There is joint participation in developing a project's scope and objectives. A business sponsor may be designated and may help launch the project. This is where the successful partnerships begin to emerge. Once the project is underway, however, this is *not* the time for the business community to sit back and wait for the next briefing from IT. Now is the time to dig in and help the project move forward. This includes gathering true business requirements and assisting with understanding the data and helping to make decisions along the way.

Staying involved in a project on a daily or weekly basis ensures that there is a good understanding of what is happening on a project. This regular communication also helps to identify potential problems and get the appropriate help in a timely manner. Detailed involvement helps to ensure that decisions are made within the proper business context. What may seem impossible to a technical project team may be easily accomplished by a business representative, and vice versa.

In addition, it is easier to garner support from the rest of the business stakeholders when one of their own comes to them with questions or the need to clarify something. Too often, a technical or systems representative may not be allowed the same level of access or may not be able to ask questions in a way that is meaningful to the business. By working together, the project team members can put together a strategy to get the business input that is necessary to drive the data and technical decisions. A strong partnership that includes a lot of mutual involvement also builds ownership of the data warehouse across the business group. This is important for adoption and use of the data warehouse when it is deployed.

**REFERENCE** Chapter 4 provides more details about how to develop and maintain strong partnerships.

## ***Identifying True Business Requirements***

True business requirements are not a list of data elements, data sources, or even reports. These requirements must be more fundamental to the business itself. For example, the need to better measure marketing campaign performance, understand and better manage loss ratios, and understand and track student



performance are all business reasons that explain the need for gathering data and generating a report, in contrast to a demand to “give me the numbers.”

These clearly defined business requirements also tend to stand the test of time. The specific data elements may change, but the need to monitor competitor performance will always be important. These requirements also remain fairly constant unless an organization completely changes business industry.

**REFERENCE** Chapter 6 provides more details about how to collect, understand, and communicate business requirements.

### ***Shifting to a Global Perspective***

Many organizations structure their employee performance expectations and goals on the results of that specific group. This creates an environment where the staff members look out for their own interests, sometimes at the expense of the greater good. This culture makes it more difficult to successfully implement a data warehouse. The ability to shift the emphasis to a more global perspective increases the success of the data warehouse. Rewarding and encouraging employees, business or technical, who are looking for opportunities to collaborate, leverage, and integrate across the organization is important. The larger the organization, the more challenging this can be.

High-level managers need to look for a balance between individual department goals and enterprise goals. This is done to set priorities and direction for the organization, and happens annually during the budgeting process. This same thought process can be applied to the data warehouse environment. Looking for the largest benefit for the most number of people across the organization can determine what data is loaded first and which reports are developed. This broad perspective is also needed to create standard data definitions and calculations.

#### **SUCCESS TIPS: ENCOURAGING AN ENTERPRISE VIEWPOINT**

It may be necessary to modify individual job descriptions and annual performance goals to encourage behavior that is beneficial across the enterprise. Reward employees who fight for the greater good. Many individual contributors on a data warehouse project can recognize myopic decisions, but there may not be a way to raise these issues. Strive to provide a mechanism for raising concerns and identifying problems without fears about retribution. This can open up a dialogue that improves the decision-making process and encourages taking broader requirements or benefits into consideration.



## Overcoming Unrealistic Expectations

It is unrealistic to think that many years' worth of data problems and challenges can be corrected in a matter of weeks. In many organizations, addressing fundamental data problems can actually take years. While setting forth a vision for the future, it is critical that realistic expectations are set for each individual project along the way. Informal expectations exist whether you explicitly set them or not. It is important to know what employees expect a project to provide, how much it will cost, and how long before the data warehouse will be available for use. Project expectations must be compared to what is actually planned. Organizations that regularly monitor these expectations can identify gaps and can take action sooner. This does not mean that other development problems go away or that project dates never shift; it simply means that there is a common understanding of what is happening and why.

Although there is always interest in getting things done faster, a business must be patient in order to get the right things done. Although it is a cliché, there are times when simply adding more resources will not result in faster completion of a task. You cannot assign nine people and expect a baby to be born in one month.

### SUCCESS TIPS: MANAGING EXPECTATIONS

**It is crucial that what a specific project will and will not deliver is documented and communicated. This *project scope* must be in business terms and should describe the business benefit. The scope needs to be communicated often and modified as appropriate. Moreover, this communication must be delivered across the entire project team, the target user audience, and associated business and systems management.**

**Another aspect of setting proper expectations is specifying a realistic budget and timeline for the project. Apply experience and knowledge of the organization to estimating data warehousing projects. If all other systems development projects are being completed in 10% less time than industry averages, then this might also be true for data warehousing. If the organization bureaucracy adds 15% overhead to industry average budgets or timelines, then adjust accordingly. Clearly, experiences with other data warehousing projects across the organization also need to be considered. This enables others in the organization to leverage what has already been learned.**

Initial project expectations and time schedules may shift throughout the life of the project. Successful organizations communicate changes in terms of impact to the business. This helps ensure that there is a common understanding of both what the project will be delivering and when.

### ***Providing Clear Communication***

The need for regular open communication is critical to the success of a data warehouse project. This involves ensuring that there is a regular forum to discuss progress and problems. Successful communication is not simply conducting project status meetings and sending out project status reports. Those are indeed important communication vehicles, but the intended audience for those is the project team and project management personnel. Regular communication provides the best vehicle for questions and concerns to be raised and then addressed. This helps expectations to be managed.

The communication must be open to be helpful. This means that teams are encouraged to share both good news and bad news. Openly admitting that a problem exists provides the opportunity for others in the organization to help provide solutions.

**REFERENCE** Chapter 4 provides tips for how to ensure effective and timely communication about data warehouse initiatives.

#### **SUCCESS TIPS: EFFECTIVE COMMUNICATION**

Be sure to employ a multi-layered approach to communication. Organizations don't have clear communication by accident; they develop specific communication plans and then follow through. These plans need to include an approach to different levels of the organization, including senior management, project sponsors, systems management, and the business groups who are the targeted users of the data warehouse. The frequency and format for communication to each of these audiences is different.

Senior management may only need quarterly updates in an executive briefing format. Middle management, from both the business and the systems community, may need a more detailed form of an executive briefing. It is helpful to have these on a monthly basis. This is often done by including a brief time slot in regularly scheduled staff meetings. Monthly updates for the target user group may also be enough. Again, the opportunity to share progress can be easily included in regular staff meetings.

**NOTE** It is critical to continue regular communication when there is no significant progress to report. The fact that work has slowed and milestones have not yet been reached is just as important as sharing significant progress. Reporting that hard work continues and that the next milestone is expected to be met by a specific date is useful.

### ***Treating Data As a Corporate Asset***

The realization that data is more than what is collected and stored on a computer is just the beginning. Successful organizations realize that while the technical staff provides the mechanics for collecting and storing data, it is the business that really drives the meaning and use of data. Data is owned by the business itself. This drives the business to take a hands-on role in what data means, where it is collected, and how it is used.

For example, the business must determine the criteria regarding who is considered a new customer. Is it when a new customer identifier is assigned in the order entry system? Is it when a customer is issued a new credit card or when that customer uses the card for the first time? Perhaps a customer is also considered new when he or she makes a purchase after a twelve-month lack of purchases. Even seemingly simple concepts require a bit more thought and understanding. Organizations that truly leverage their data do not leave these types of decisions to the technical staff, but work together to clarify and store useful data. Treating data as a true corporate asset is much broader than the data warehouse. This encompasses all systems across the organization.

**REFERENCE** Chapter 8 provides more details about what is involved in handling data as the asset it can be, with the emphasis on what this means for the data warehouse.

### ***Effectively Leveraging Technology***

The selection and use of technology is usually very much a focus for data warehousing. Purchasing and maintaining hardware and software are some of the most concrete costs of a data warehouse project. There are a myriad of options for every component. Even after a specific vendor or technology is selected, there are more options for different levels of capability and/or configuration. Most companies have sound practices in place for the evaluation, selection, and implementation of technology. Operational applications are very much defined by construction and then deployment using a set of technologies. Successful data warehouses are implemented by investing enough in technology to empower the technical staff to be efficient in both development and ongoing maintenance and enhancements. The same investments must also ensure effective business use of the data warehouse.

The key here is that buying the least (or most) expensive technology may not be effective for your organization. Having a vision of what needs to be accomplished for the business helps drive decisions about technology.

**REFERENCE** Chapter 9 contains suggestions for finding the right technology to meet your requirements.

**SUCCESSFUL TIPS: DEPLOYING TOOLS SUCCESSFULLY**

**When implementing technology, many organizations purchase the best-of-breed tool, but then cut corners when setting it up. Squeezing a sophisticated piece of technology onto an existing computer server is not likely to allow that tool to work effectively. It may even limit the functionality and will most certainly degrade the performance. Don't try to save money by configuring the minimum supported environment. Look for the recommended configuration that will deliver the results you desire. If the technology does not work well, the effort to build the data warehouse may have been in vain.**

## Roadblocks to Success

Many organizations have failed in their data warehousing efforts. Some have struggled for years, while others have failed in a big and highly visible way. Some data warehouses have not failed outright, but have never achieved their full potential. While each case has unique characteristics, several common themes regularly contribute to failure or the lack of total success.

### ***Believing the Myth: "If You Build It, They Will Come"***

There are still many companies who invest in building a data warehouse by pulling data from core operational systems into a common database, expecting that the business will then begin to use it. This is a good exercise in data movement and organization, but an approach that rarely leads to success. The data is organized in ways that are not tied to actual business needs and requirements. This is an approach that seems 'safe' from a systems perspective. All design and development work is done without venturing into the business community. The belief is that once the data is organized, the business can be consulted to determine how to deliver it to them.

This approach does not take into consideration what the business community is really doing and wants to do in the future. It is also limited to the way that data is captured and stored in the underlying operational systems. The reality of the business often differs from what is captured. Data is grouped and reported in ways that may not be reflected in the underlying systems.

Such an approach also requires a large investment in infrastructure and building this database before there is any link to business value. This was once a very popular approach to data warehousing, but it is no longer a leading one. Unfortunately, many organizations are already well on their way to building this type of data warehousing environment.

**RECOMMENDATION** It is never too late to gather business requirements, a task that can identify the data needed to deliver specific business analyses and reports. Adding new data to the data warehouse and possibly restructuring and/or enhancing the existing data can derive value.

### *Falling into the Project Deadline Trap*

Many organizations have invested in setting up procedures and processes that follow a project methodology. Clearly, an important part of the process is defining a project's scope. Detailed project plans are developed and used to measure progress. Group and individual performance are often based upon the ability to meet project deadlines. While project planning and management are critical to coordinating multiple resources and keeping projects moving forward, this focus on the project plan itself can cause data warehousing initiatives to go awry.

The time needed to extract, transform, and load data for analytical applications (see Chapter 10) is one of the least understood and most often underestimated tasks of any systems development project. This often leads to project delays. One common way to manage these potential delays is to reduce the scope of what is to be delivered. This can lead to finally delivering something on time, but the scope has been pared down so that no business value is delivered. In other cases, testing and quality checks are reduced or eliminated in order to maintain the project schedule. Unfortunately, many organizations reward this type of behavior. Delivering something on time is a prevailing theme.

**RECOMMENDATION** Reward project teams for producing quality results. This includes validated data and well-tested applications. This also means that applications indeed support the business in a valuable way. It must not be acceptable to reduce the scope so that little or no value is delivered. Part of this change in deadline focus also means that project change requests are viewed in a positive light. Project issue logs should be full of open items needing to be addressed. Teams should not be penalized for being honest about progress.

### *Failing to Uphold Organizational Discipline*

Despite an impression that data warehousing is glamorous, most of the work to build one is not. The type of work needed to develop and implement a successful data warehouse is detailed and painstaking. Project tasks are often

tackled with excitement, but as the realities of the project sink in, enthusiasm begins to wane. Data analysis may sound fine in a project plan, but having to crawl through hundreds of data fields is time-consuming, and can be difficult as well. Systems professionals have an aptitude and interest in this type of detailed data work. Business professionals rarely do—after all, if they did they would probably have become systems professionals!

The primary areas of effort for a data warehouse revolve around the data itself. When thinking about data, most people think about the technical aspects—where the data is stored, what computer programs may access or use it and who to call in systems if you have a problem. The technical care and feeding of data is only a part of the picture. It is also important to understand the business definition of each data element, why it matters to the organization, and how the data should or should not be used. Resources must be assigned to work on these business aspects of data. This is where true discipline comes into play. It is much easier to simply move data around than it is to figure out what data is needed and then determine where to find it. This discipline is needed not only for the initial project, but also going forward into the future. The need to develop and adhere to good data practices never ceases.

Organizations that have not been successful with other initiatives that require long-term discipline need to be up-front about this. Strategies must be developed to ensure long-term discipline. A lack of data discipline does not mean failure, it simply means that the business must be willing to accept and use data as it is. This may be sufficient to support today's business needs.

**RECOMMENDATION** Adequate resources need to be allocated to build and maintain a sustainable data warehouse environment. Many organizations allocate enough systems resources for project development but neglect to allocate business resources. Business representatives must be responsible for providing guidance, answering questions, and resolving data problems during a data warehouse project. Additionally, to ensure proper organizational discipline, business processes must be put in place to address long-term maintenance of data definitions and data preparation rules. In addition, resources must be allocated to support and address ongoing business data responsibilities.

### ***Lacking Business Process Change***

The impact that a data warehouse can have on an organization goes far beyond getting reports faster. Automating what is done today does not dramatically change business results. It is much more valuable to change how things are done. The data warehouse can be a catalyst to help make these changes. The data warehouse should be used to provide historical information and to

highlight the most important things that need attention today. This can be done at many different levels within the organization.

For example, the executive team might see a sharp increase in claims activity and evaluate the need for setting up additional financial reserves. The claims managers might need to review staffing plans in order to keep up with the increase in activity. The actuaries might need to perform analysis to identify underlying causes for the increased claims activity (unless it is a result of a natural disaster). The underwriters might need to reevaluate which policies are issued and renewed to ensure that the company continues to maintain financial stability. On an individual level, the data warehouse can also be used to have customer service representatives proactively contact claimants to communicate claim status, especially if it is taking longer than expected to resolve a claim. All of these activities are done periodically as a natural course of running the business, but changing these processes to utilize the data warehouse can influence the timing and sequence of these tasks.

Unfortunately, many organizations simply use the data warehouse as a glorified report generator. While it is true that the data warehouse can provide more accurate, integrated, and timely reports, significant long-term benefits are derived when report results and analyses drive changes to how business is conducted.

**RECOMMENDATION** Use of a data warehouse can be daunting, especially for people who have not been able to get to data themselves in the past. Staff members can feel intimidated and unprepared to meet the new expectations. Appropriate education must be provided to ensure that the group has the skills necessary to use the data warehouse. This education must include both the mechanics of using the warehouse and an understanding of how to use or analyze the data effectively.

Care must also be taken when making changes to business processes. When developing process changes, it is helpful to include employees, who often have great ideas and insight that cannot be found elsewhere. This inclusion also creates a more open attitude for embracing the changes. Business process changes may not be done in a single sweeping effort, but evolve over time, as use of the data warehouse becomes a regular part of daily work.

### ***Narrowing the Focus Too Much***

Because a broad goal of loading all of the data and then looking for opportunities to improve is rarely successful, many organizations have wedded themselves to the other end of the spectrum. With this approach, the requirements for a project are defined in a concrete, specific manner to deliver a set of key performance indicators, or perhaps the 25 most important data elements



have been identified. While this can narrow the scope of a project, it can also lead to significant challenges in the future.

With this type of narrow focus, the emphasis is indeed on the business, which is good, but often only the data specified is extracted and loaded. After looking at the requested data, it becomes obvious that some of it is as expected, some of it is not what was intended, and other data may not yield any insight even though it is was requested. This results in the need to change the data that is needed, which requires going back to the beginning.

Another common problem with too narrow a focus is apparent when follow-up questions are asked. Often, data is requested at the summarized level to represent an entire company or group. When the more detailed questions are asked to better understand why a certain number is high or low, the data is not available. This was a common downfall of the executive information systems that were popular in the 1980s. Although the development and maintenance of performance dashboards is much easier, the same potential pitfall exists.

**RECOMMENDATION** Because both too broad a focus and too narrow a focus on the business is bad, there must be a middle ground that is more successful. The best balance is achieved when the underlying data needed to create performance indicators is loaded in the data warehouse and made available to support detailed analysis. Executives and high-level managers can drill down into the data themselves or rely on their employees to follow up on the more detailed analysis. Either way, the data is available to support the needed analysis.

Another benefit of loading detailed data is that often very interesting and possibly valuable data may live alongside the specific elements requested. By pulling all useful business data, rather than just one field of immediate interest, the data warehouse supports multiple levels of analysis and future changes. The effort required to pull along these additional fields is often not significant. The most effort is needed to integrate even one element from each source, rather than multiple elements.

For example, it can be hard work aligning product codes from non-integrated systems, but the amount of effort needed to integrate the product codes is the same whether one uses one data element from each source or fifteen elements from each source. Given that the level of effort does not increase dramatically, why not bring forward other useful business data at the outset? To increase long-term benefits, it is critical that designs for data delivery and preparation *not* be limited to a specific set of business measures. The design of the data access and analysis components of the data warehouse can indeed focus on the specific requirements at hand.



### ***Resting on Your Laurels***

Some organizations have realized great business success for many years. The people in these companies have little incentive to change. The way that business has been done in the past still works well and continues to yield profits. As long as success continues, it is extremely difficult to effectively deploy a data warehousing environment. A similar culture may also exist in larger organizations, where individual employees may not feel directly connected to business results.

The process of developing a data warehouse requires that business rules regarding how data is processed must be reviewed and possibly changed. Rules and guidelines for integrating data from independent systems must be defined. Even the basic content and quality of existing databases is scrutinized. This is not easy, in any situation. When there is no real business reason to expend the effort, the chances for success dwindle.

Nonetheless, data warehouses can be effectively deployed in highly successful organizations when there is a culture and mindset to stay ahead of the competition.

**RECOMMENDATION** Take an honest assessment of the need for business change. If the business can likely continue for the next three to five years as is, then perhaps the time is not right for data warehousing. If change is necessary to continue market strength or to adapt with the marketplace, then data warehousing is likely to help the organization. All data warehouse initiatives must be positioned as part of an overall strategy to take the company into the future. Data warehousing must be accompanied by other initiatives to help the organization change business processes and the culture. Individuals must see how their work contributes to the overall success of the company.

### ***Relying on the Technology Fix***

Organizations often seek a simple solution to a set of very complex challenges. No matter how sophisticated technology becomes, it will not solve fundamental organizational and business issues. Purchasing and installing the latest technology will not overcome fundamental business process flaws. Technology does not solve data quality or integration problems either, but it can help a company address their challenges.

Businesses sometimes believe that swapping technologies will correct the problems of a data warehouse. If a data warehouse is floundering, the easiest thing to point at is a specific tool. Before changing technologies, however, it is worthwhile to evaluate the current environment. Newer versions of existing

software may resolve outstanding issues and provide additional functionality, but if the current version of a tool does not meet the organization's needs, then it may be worthwhile to survey the marketplace for other options.

Keep in mind that each tool has its own strengths and weaknesses. You may be trading in a known set of weaknesses for the promise of a new set of strengths without really understanding the new weaknesses. The cost to swap tools may be greater than the potential benefit of the new tool. Moreover, there are additional costs to retrain both technical and business users of a new tool; and still other costs associated with converting applications and uses of the existing tool.

**RECOMMENDATION** Evaluate all aspects of the current data warehouse environment, not just technology. Keep in mind that technology alone does not address business, process, or data problems.

The bottom line? Evaluate and understand all aspects of a struggling data warehouse before making drastic changes. Removing any single roadblock may not yield the desired results, as multiple roadblocks may need to be addressed with a comprehensive strategy.

### ***Getting the Right People Involved***

Once a data warehouse project is approved, the challenges of staffing the project team begin. Unfortunately, the best resources to help are usually already overbooked, so the available employees are assigned to the project. These people are often lacking the necessary experience or skills needed.

The most successful data warehouse projects are filled with the most valuable business and technical employees. These people have a lot of knowledge about the company, how it works, where data is stored, and the secret handshakes needed to gain access.

It is difficult to allocate these highly valued resources to a new project because they often play an integral role in the daily functioning of the business and other special initiatives, yet these are the people who help resolve crisis after crisis. These resources must be allowed to participate in the data warehouse project. By leveraging their knowledge and experience, the existing data problems can be addressed, enabling the data warehouse to serve as the foundation of the company's efforts to move forward.

Another common mistake is hiring new staff to fill the business and technical roles. While adding new staff members is beneficial in the long run, this can create a significant problem at first. New employees do not typically know enough about how the business is run or what data is currently used. Nonetheless, new employees do bring a wealth of ideas about what is possible,

and often have a realistic understanding of the difficulties in weaving their way around the data in its current form.

**RECOMMENDATION** Even if the most experienced staff members cannot participate full-time on a data warehouse project, these valuable resources need to provide regular guidance and direction. From a business perspective, get other staff members to pick up some of the daily responsibilities for key personnel. This may mean brining in additional help—you can't expect everyone else in the group to go from 110% to 120% effort. From a technical perspective, it is important to keep your staff involved and informed about what the data warehouse is, and how and why it is built. It is risky to turn the entire project over to a third party because they will eventually leave with all the business and technical knowledge.

At a minimum, hire contractors to backfill some of the daily operations work so that key technical personnel (DBAs, architects, and lead developers) are able to participate in the project. Another technical resource constraint is access to the key individuals who understand and intimately know the source system and data. The data warehouse team must have regular access to these individuals in order to understand the existing data. Major errors can be made in pulling data without the appropriate background knowledge. This is often not discovered until much closer to deployment.

Finally, if the appropriate resources are not being made available from a business and/or technical perspective, then the true priority of the data warehouse must be evaluated.

### ***Finding Lost Institutional Knowledge***

Over the years, application systems have been developed to automate business processes and apply business rules and logic. In addition to these systems that run the business, a series of programs, SAS routines, or complex spreadsheets have been developed to meet reporting and analytical needs. Many of these reports and analyses have now been in place for years, but over time the fundamental purpose of a report may be lost.

In addition, the business rules underlying the criteria for the report have been embedded in code that is simply run on a regular basis. Too often, a description of these rules and the business rationale behind them has been lost. Therefore, when the data warehouse team is applying pressure to understand the current business practices, the business community may honestly no longer have that knowledge. This is a fairly common situation, and one that does not warrant looking for someone to blame. It is important to acknowledge that this is the case and work from there.

**RECOMMENDATION** One's first impulse is to read the programs or dissect the spreadsheets. After many hours of tedious work, the business logic can be

**found, but the rationale cannot. The result is merely a better understanding of the status quo, which is probably not meeting all of the business' needs. As an alternative, step back and figure out what is really needed to understand and run the business. This is a chance to define, or redefine, how you want to look at the group, department, or company. Starting from a clean slate seems scary, but it can be a much faster, and more direct and more useful approach than trying to replicate an existing, potentially archaic set of reports.**

## **Summary**

---

Building a data warehouse is challenging, but it can also provide great rewards. This chapter offered a general overview of data warehousing and illustrated how it can serve as a potential foundation for realizing business value. The next chapter has answers to some of the most frequently asked questions about data warehousing, and subsequent chapters provide more in-depth coverage about how to build and sustain a successful data warehouse.

# The Executive's FAQ for Data Warehousing

When you are in a leadership role within an organization, you are charged with the responsibility of setting corporate goals and defining the strategic initiatives to meet them. You influence and inspire others in the organization to help fulfill the vision and achieve these goals. This is simple in theory, but much tougher in practice. You need to ask questions and challenge the status quo. As a business or IT leader, it is easier to ask questions and understand the answers within your part of the organization, but there can be a high degree of frustration when working across IT and business boundaries. When business leaders are asked whether they are getting what they need from their IT organization to run their business, or when IT leaders are asked whether they are getting what they need from the business to build a data warehouse that will meet business challenges, the answer from both is often the same: *no*.

Part of the problem is an assumption that both business and IT inherently understand each other's needs, when in reality the only way to truly understand the requirements is to work side by side and learn from each other. As leaders, we want to see a return on our investment (ROI). In this case, that investment is the partnership between business and IT. If you don't invest in the partnership, then you will never see the return you want in your data warehouse.

This chapter is intended to provide a big dose of reality. It consists of a high-level summary of questions that are frequently asked by executives and senior managers regarding DW development. These are complex issues and often require additional dialogue to clarify the question and provide an appropriate answer for each organization. Therefore, this chapter shares only a common interpretation of the question and the essence of the answer. These topics are covered in much more detail throughout the rest of the book, so if the short answer doesn't meet your needs, then follow the references to subsequent chapters to learn more.

## Question: What is the business benefit of a data warehouse?

---

Proper implementation and effective use of a data warehouse can provide significant value to an organization. The magnitude and frequency of value recognition can vary greatly based on your application and industry. For example, a telecommunications company built an integrated data warehouse that enabled them to identify potential customers that were ideal for them to cross-sell other services, and to develop a targeted cross-sales campaign. This single campaign resulted in approximately \$50 million dollars in additional revenue. More often, the value is not realized by one thing, but a collection of many different things. A retail organization developed a data warehouse that enabled them to better analyze sales by store and cost of goods sold. This in turn enabled them to better target promotions and more effectively negotiate with their suppliers. The results were a series of negotiations that reduced cost of goods, and promotions that increased revenue, each adding between \$20,000 and \$200,000 to the bottom line.

## Answer

While it is nice to know that others have seen success, what you really want to know is “what is the benefit to *me*?” Unfortunately, it depends on many different factors. Without research and a sound understanding of your organization and the challenges you face, you can’t really put a number on the potential benefits. However, given the preceding examples, it is hoped that you can begin to think of areas in your own business that, with the right information, you could make more profitable or cost effective.

Historically, successful data warehouses have not resulted in a decreased head count. Typically, the data warehouse enables the organization to make improvements that result in increased revenue and profit. The magnitude of the benefit is relative to the size of the organization. If the organization has \$10 million in sales, then it is unlikely that the data warehouse will immediately provide an additional \$5 million to the bottom line. Conversely, for an organization with revenues in the billions, \$5 million in return is more likely. The types of benefits that have been observed in other organizations include the following:

- Increased understanding of what really drives business
- Identification of market opportunities
- Proactive identification of potential problems and the launch of preventative measures

- Identification of opportunities to streamline and improve business processes
- Development of more effective marketing strategies through target marketing, and improved customer relationship management
- Improved quality of products or services
- Increased customer satisfaction
- Identification and monitoring of trends in the industry, such as shifts in customer behavior or changes in competitors' pricing strategy

The key to understanding the potential benefit to your organization is understanding the challenges facing the organization and the data that the organization already collects. Chapter 5 provides more details about how to set up a successful project, while Chapter 6 discusses how to gather and understand business requirements.

## Question: How much will it cost?

---

The cost of a data warehouse is just as wide-ranging as the business benefits. The cost depends on a wide variety of factors, including size of the organization, the industry, and data complexity. For example, the data needed to support actuaries within an insurance company is more complex and requires much more history than the data needed to support the sales organization of a retailer. Data warehouse budgets can be measured in tens of thousands to multiple millions per year. Whereas a small organization is unlikely to spend millions each year, a multibillion-dollar global enterprise may spend millions on their data warehouse initiatives.

## Answer

Most IT organizations are quite capable of compiling costs for technology, including hardware, software, maintenance, and support. These costs are defined by the volume of data to be stored, the number of different technologies to be used, and the number of people who will access the environment.

Data volume has a big impact on the size of the hardware needed to extract, transform, load, and store. How quickly these processes must run also affects the size and class of the machine needed. However, the business needs to be involved to establish the level of detail and the length of history needed for analysis. It is not acceptable for the business to simply want "all data for all time," which has a cost associated with it. This is just not practical.

Many different tools can be used to build a data warehouse environment, ranging from the programming language used for other application development to the purchase of specialized technologies to cleanse, transform, and load the database. Some organizations support multiple data access tools, which may include a basic report writer, multiple business intelligence tools, and perhaps a tool to perform statistical analysis.

Another category of costs that vary greatly are the people. There are costs for the IT team members to design and build the data warehouse. There are also costs associated with the business people who are active participants on the project team. The amount of time and effort required depends upon the scope of the project and the state of the data itself. If data is well defined, understood and accurate, then less time will be needed to investigate and correct data quality problems. The worse the data is today, the more effort is needed to get it into shape to be useful tomorrow. There will be effort from both IT and the business to understand data problems and figure out what to do to correct them.

Once a data warehouse has been released for use, there are ongoing costs to ensure that proper support and education are provided. Without this continued support, many data warehouses fail. Ensure that the organization plans for these ongoing maintenance and enhancement expenses. Usually, these are offset by the success of the data warehouse.

Chapter 4 explains the different roles for both business and IT project team members. Chapter 9 discusses the tools and technologies that may be used. Chapter 10 explores what is involved in preparing the data to be loaded into the data warehouse. Chapter 12 provides insight into what is needed to support the data warehouse once it is put into production.

---

## **Question: How long will it take?**

---

It depends. You must first define the business problem you are trying to address. This will help define what are you trying to build. There is a dramatic difference between the time required to build a data warehouse from scratch versus building a few new reports to deliver via a new dashboard using a tool that you have already been using. Variables that can have a big impact on how long it takes include the following:

- How many different business groups are to be supported?
- How many different data sources will be included?
- Is the data well defined and well understood by the business?
- Is the data well structured and documented from a systems perspective?
- How easy is it to integrate the data from these different sources?
- Is this the first data warehouse project?



- How much experience does your staff have in building a data warehouse?
- Will critical subject matter experts (from both the business and IT) be available when needed? Will they have time to actively participate in the project?
- How quickly can you make decisions needed to design and build the data warehouse? Sample topics requiring decisions include defining the formulas for business measures and how to handle anomalies in the source system data.

## Answer

In order to provide a glimpse into how long it may take, let's look at building a data warehouse from scratch. After having worked on hundreds of different data warehouse projects, I've found that there are clear breakpoints for estimating efforts. The first part of a project gathers true business requirements through conversations with real business users, and develops the dimensional model that will meet those business needs. This helps both to clarify what is actually needed and to gain a good understanding of the data. This effort should include exploring possible sources of data to populate the database. The dimensional model provides the vision for what the data needs to look like to support the business reporting and analyses that need to be done.

Based upon years of experience, it typically takes six to eight weeks to gather requirements, and another six to eight weeks to develop a comprehensive dimensional model, the first time. The number of different groups that will participate in providing requirements, and their availability, determines the time needed to gather and document the requirements. The number of data sources, the sheer number of data elements, and the complexity of the data define the time needed to complete the dimensional model. Generally, these can be completed in three to four months. Although you may use some external resources to accomplish this process, it is critical to have your internal resources intimately involved.

Many people are surprised at the length of time described here. In order to develop a sound, sustainable data warehouse, it is critical the requirements and design be done thoroughly. When the first project builds a solid foundation, then subsequent projects are likely to take less time and re-work is minimized.

Once the requirements have been collected and the data has been modeled, you should have sufficient information to be able to estimate the effort required to build the data warehouse and develop the business intelligence solution. This can range from a matter of weeks for small projects that start with solid reliable data to many months for larger, more complex data.

The effort to build a robust data warehouse and business intelligence solution is quite different from starting with a sketch of a dashboard that

needs to be built. The key performance indicators on that dashboard may come from ten different sources that do not integrate well. This often results in the development of *one-off reports* that populate only a portion of the dashboard. These one-off reports often do not have the underlying detailed data used to build them, which limits the user's ability to drill down into more detail, thereby decreasing the overall value of the dashboard.

In short, the answer to how long it will take depends on where you are today and what you are trying to build.

---

## **Question: How can I ensure success?**

---

As a leader in the organization, there are several things that you can do to help ensure the success of the data warehouse.

### **Answer**

The first thing you can do is *learn about data warehousing*. Invest your time to understand what a data warehouse is and what it takes to build and maintain one. Take the time to learn more about the components and what needs to be done for the areas where you will participate. This helps by improving the level of communication between business and IT staff members. For the IT readers, learn about the business. Learn about the industry that your organization is part of and how the company works. This also improves the level of communication with your business counterparts.

The next most important thing is to listen. Too much time is spent talking and jumping to conclusions. Step back and really listen. Understand what is being communicated. If you don't understand, then ask questions to gain a better understanding. Too often, important messages are discounted or ignored when brought up internally. In many cases, external consultants are brought in and conclude the same thing that your own people already did. This is not to say that there is never a need to get an impartial assessment or opinion, but make sure that you take the time to really listen to your own people.

Finally, follow through on your commitments. When others in the organization challenge the project team, stand up for what is needed. Help the project team gain access to the right business personnel, support their efforts to work toward a single definition for each data element, and ensure that sufficient resources are allocated. Too many managers support a data warehouse initially, but then don't allocate the time and resources necessary to ensure that it will be designed and implemented correctly.

Chapter 4 describes the importance of partnership, and the specific roles and responsibilities for both business and IT personnel involved in a data warehouse.

## **Question: Do other companies really build these in 90 days?**

---

The ability to deliver in 90 days is closely related to the previous question—how long will it take? Again, the most important factor to consider is what you are trying to build. If you are putting in place the foundation for the future of the data warehouse, bringing in all new technology and starting from scratch (or starting over) to meet business needs, then you are unlikely to accomplish this in 90 days. However, if you are building new reports and analyses that use data that is already loaded, this can often be done in 90 days or less.

## **Answer**

Making time estimates can be very deceptive if you only look at how good the technology is. The tools can be set up fairly quickly and can move data around faster than ever before, so why can't we build a data warehouse in 90 days? A good data warehouse is not about the technology. The value lies in the effort made to define the data elements and develop the business rules needed to clean and integrate the data. This requires a great deal of thought and discussion among and between business groups and IT. Then decisions must be made. Regardless of how fast technology can implement those decisions, it still takes time to research and make them.

It is also important to understand what is actually being delivered from a 90-day project. Often, it is not a complete robust environment but a quick prototype. If the prototype is not quite what is needed, then another 90-day effort is launched to change what is delivered. Many organizations iterate many times, hoping to get it right. It is important to understand and agree upon what must be in place for success. Number one on the list? The data warehouse must be driven by well-defined business requirements.

The amount of time required to put a robust data warehouse in place can vary greatly. Rather than approach this by demanding a completed project in 90 days, determine what results you want and then figure out what has to be done to develop a sustainable, long-term solution.

Once you have a solid foundation in place, the project schedules can be shortened. The work to enhance a dimension or to add another fact table is much less than the work to start at the beginning. It is also much faster to build new BI reports and perform analyses using data that is already loaded into the data warehouse. Many of these subsequent projects can be done in 90 days, including the testing and release activities. Ultimately, investing up front to get a solid foundation in place will save the organization a great deal of time and money.

## Question: How will we know we are doing this right?

---

There is no single “right” way to build a data warehouse, which makes it difficult to assess how well you are doing. Therefore, there is no cookie-cutter scorecard that you can use to benchmark your progress against the industry. While you may not know whether you’ll attain the projected benefits, reasonable adherence to industry best practices and sound project management will help you determine whether things are going “right.” In the end, you will be able to tell when the data warehouse is able to support the business to meet its objectives. When everyone relies on the data warehouse to get their data and use it for their analyses, then it is working. Another sign of success is an increasing amount of requests for additional data and more reports or analyses. This means that there is confidence in the data warehouse environment, and that users see new ways that it could help.

## Answer

As with any project, you can track and monitor progress compared to the project plan. Are the tasks being completed? Are project deliverables being finished? If not, you need to ask what is holding things up; and then you need to apply sound management and judgment to figure out what can be done to help. For example, if it is taking a long time to gather requirements, ask the project team why. If the project team is having difficulty getting meetings scheduled with key individuals, then work with the appropriate managers to see what is happening. Too often, a lot of time is wasted while the project team waits on other people. Eventually, the situation will be escalated, but your active involvement can keep things moving. Requests that come through business channels often get higher priority than requests from an unknown project manager or IT person.

It can be helpful to set up periodic reviews and checkpoints that go beyond basic status meetings. When possible, these should involve others across the organization with data warehousing experience. This helps leverage the experiences of other groups. Some reviews should be scheduled at regular intervals, regardless of what project deliverables have been completed. This helps to identify problems before they get out of hand. Checkpoints are typically done when major milestones have been met. These enable a review of the results, such as a review of the data architecture, data model, or dashboard design.

Although there are no guarantees for success, strong executive sponsorship, partnership between the business and IT, hard work, sound planning, and a lot of communication certainly help.

---

## **Question: Why didn't this work last time? What is different this time?**

---

Just as there are many factors for success, unfortunately there are just as many that contribute to failure. For every highly publicized successful data warehouse success story, many more have not achieved those amazing results. Most data warehouse projects deliver some benefit, some have limited results, and there are a few downright failures. When any project comes to an end, time should be spent on evaluation. It is helpful for the technical team to compile their ideas of what worked well, what could be improved and how, as well as things that did not work at all. The same should be done from a business perspective. Then, a complete review should pull both perspectives into a complete assessment of the project. More details about a post-implementation review can be found in Chapter 12.

This evaluation must be done in an honest, nonjudgmental manner in order to get to the heart of what happened. Rather than look for someone to blame, examine what did not work as planned and explore opportunities for improvement. Collect ideas from everyone, not just the high-level people and managers. Individual contributors often have great insight into what could be improved.

## **Answer**

Taking the time for a review is good closure for any project, but it is particularly important for those that had problems. This should be done for projects that are canceled before they are completed too.

In order to get the next project started on the right foot, you may need more help. This may require hiring someone with strong data warehouse experience or bringing in consultants to help train and mentor your staff while working on the project. In order to change the outcome, different activities and steps must be taken. These new participants often have great ideas and can recommend what needs to be done. In order to be more successful with this attempt, you need to listen to their advice and do what they recommend.

Chapter 12 shares details about data warehouse projects that have stalled, along with ideas about how to correct them.

---

## **Question: Do we have the right technology in place?**

---

You can purchase a number of different kinds of technology to help build and use a data warehouse. Chapter 9 provides an overview of the different categories of technology that may be used and some insight into how to determine

what products are best for you. However, once the data warehouse has been built, you may still wonder whether you have the right technology in place.

## **Answer**

In order to determine the viability of your technology choices, several key questions must be answered:

- Does the environment work?
- Is the environment stable?
- Does it work fast enough?
- Can you maintain it?
- Do you have enough functionality to meet both current and near-future needs?
- Can you afford to keep using it?

While these questions can be asked of the overall environment, if there are any major concerns, then you need to look more closely at each of the different parts. Perhaps the ETL system is working fine but you don't have acceptable query performance or the data access tool cannot perform all the calculations that you need.

There may also be a different level of satisfaction between the IT and business groups. The data access tool may be easy to install and keep up and running, but if it cannot perform all the calculations that the business groups need, then it really is not working.

Problems with the current environment can result from many contributing factors. Sometimes the tool itself no longer meets your needs. However, in many cases, a closer look may reveal that the tool does provide that function but is limited by how you have implemented it. Before tossing out any tools, make sure you look carefully at how they are installed and are being used. Perhaps you merely need to ensure you are running the current release of that product, or you may be able to make simple changes that enable you to better exploit what you already have. Look at Chapter 9 to learn more about technical architecture, and at Chapter 12 to understand what is needed to maintain the technical environment of a production data warehouse.

## **Question: Are we the only company with data warehouse problems?**

---

It can seem as though everyone else is having fantastic success with little or no trouble building and leveraging their data warehouse. The vendors describe

how wonderful their products are and how much success their other clients are having. You can find dozens of articles that make it seem like the entire world has figured out how to develop an effective warehouse. This can be intimidating and make others feel behind the times. The truth is that even the most successful organizations face problems with their data warehouses. Those that are able to address their problems and overcome challenges are the organizations that are seeing the biggest results. Based upon years of experience and the opportunity to interact with many different organizations, I can tell you that you are definitely not alone!

## **Answer**

It is useful to be able to put things in the right perspective. This can be done through conversations with peers in other organizations and networking through industry interest groups. Now that the data warehousing industry has really come of age, many resources are available to you. Seek out industry groups with a focus on data warehousing or related topics. There are many IT associations that have data warehouse or business intelligence subgroups. The Data Warehousing Institute offers major conferences, seminars, and local chapters. Any of these provide the opportunity to meet people with similar issues on their mind.

There is also a great deal of interest from business groups to learn more about and discuss data warehousing. Explore business industry groups and ask colleagues for their ideas. Check out social networking sites such as LinkedIn. If you don't find a group, take the initiative and start one. There are probably dozens of people who are just as interested in these topics—it only takes one person to get something started.

There are also opportunities to meet and exchange ideas with others at product vendor user conferences. These groups have general concerns and can be engaged in dialogue about topics that are specific to the technology that you are using.

You are not alone. The good news is that there are often proven techniques to address the specific challenges facing your data warehouse. You just need to reach out to find these proven methods.

## **Question: Will I get one version of the truth?**

---

How many meetings have you sat in on and looked at a profit number on four reports that were all different? Often, data is pulled from many different systems, by different people, using different criteria. This results in reports that provide different and possibly conflicting data. Too much time is spent



trying to figure out which set of numbers is correct, which leaves little time to truly interpret and understand the data, and less time to make decisions. One fundamental objective of data warehousing is to provide a single environment for all reporting. The objective is to indeed provide a single source of the truth.

## **Answer**

The short answer is yes, provided you have the right level of executive commitment—that is, willingness to invest the time, effort, and business expertise needed to fix the problem. While this is a clear objective, it is not easy to attain. There are some technical challenges, but working toward a single version of data typically requires a very significant level of commitment and effort from the business community. The business should drive the effort to achieve a single version of the data, and IT can help to maintain it.

There are two major areas that warrant discussion. First, it is critical to set up names and definitions for the data that consistently and accurately describe what it means. How data is pulled out of source systems and manipulated changes what the resulting data looks like. Consistent and meaningful business rules must be defined so that IT can build the appropriate mechanisms and controls. Who decides what the data means and defines these business rules? The business does, not IT, and this is typically the largest part of the overall effort.

Providing consistent data requires a lot of communication between different parts of the organization. Each group must clarify the data it needs and compare that with similar data used by other groups. Sometimes the data is supposed to be the same, but over time the meaning and use have drifted apart. Hard decisions must be made to bring these back together. In other cases, there are legitimate differences that need to be clarified. Unique labels must be used to differentiate between the data elements.

Second, the organization must be committed to working toward one version of the truth. It seems like a desirable and reasonable objective to everyone, but if working toward this goal impedes the progress of an individual project, the willingness to comply begins to fade. The development of this single version of the truth must be a long-term strategic objective. If the organization is not committed from the top down to dedicate the people, time, effort, and dollars required to define your business data, you will never reach a single version of the truth. There must be strong senior management support to make the investment in time and money to ensure that the right long-term steps are taken. If not, your organization will continue to experience a proliferation of disparate data and reports that do not match.

You can indeed realize a single version of the truth if your organization works diligently to make it happen. There must be ongoing commitment from the business to help define the data, and IT must ensure that individual projects



are chartered to adhere to these and then follow through—even if it means that your project takes a few weeks longer. If you insist on taking shortcuts, then you jeopardize the ability to maintain and deliver this consistent view of data. Chapter 8 provides much more detail about what is involved in working toward a consistent view of the data.

## **Question: Why can't we just use our current systems?**

---

In order to answer that question you need to understand the difference between running the business and managing the business. “Running” the business is the regular set of activities that keeps the business going, such as keeping factories operational, processing orders, cutting checks for claim payments, ensuring that the ATM network is working, and that the airplanes are taking off. These are the functions that define the business itself. These are what keep the company running. The activities that “manage” the business are those that track and monitor how well the business is running, and make decisions about the direction and strategy of the organization.

Operational reports are those that are produced from the systems that keep the business running. They report the current state and help make adjustments that can be put into action right now. Because the data is current, a report that shows the number of calls on hold will return different results each time it is run. If too many calls are on hold, additional staff may be pulled from the break room or supervisors may handle a few calls. The reports provide immediate feedback and the resulting decisions are used to keep that business function running smoothly.

In contrast, management reports are used to identify trends and patterns. Perhaps there is higher call volume on the third Tuesday morning of every month. Identifying this pattern would enable the managers to ensure that additional staff were scheduled to work that day to avoid long wait times. Monitoring the overall call volume is necessary to plan for the future. If business continues to grow at the current pace, then three more people will be needed by summer to handle the call volume. This is important to ensure appropriate lead time to hire and train new employees.

Depending upon your role in the company, you may use both operational and management reports. The operational reports provide guidance as to what to do now, whereas the management reports can help provide perspective so that you can plan for tomorrow, next week, and next year.

Most organizations already have both types of systems in place. Operational applications include those that handle the general ledger, accounts payable/receivable, order processing, policy administration, claims handling,

travel reservations, manufacturing, and production controls. Many organizations also have separate databases used to support reporting and analysis, or perhaps a full-fledged data warehouse. These reporting environments exist because historically the technologies were not able to handle both operational and management reporting at the same time. This is one of the primary reasons that data warehousing started.

## **Answer**

There have been significant advancements in technology to help address these technical challenges. Databases can handle a lot more data and provide good performance. Often, the data that is needed for reporting exists in more than one operational system. Products have been developed to enable business rules to be set up to support integration of this data on-the-fly. As you hear about these advancements, it seems even more puzzling to not just use the current systems.

While these technologies are interesting, the real challenge is the nature and state of the data itself. These technologies will only work if the organization's data can support them. The following key requirements need to be met:

- You need to have good performance of queries to support reporting and analysis.
- The performance of the operational application system itself must not be effected.
- The underlying system must retain sufficient history to support the business needs.
- The data from different systems must be aligned to integrate easily or the business rules to apply to the data must be straightforward to be done quickly on-the-fly (without affecting performance of either purpose).
- The data must be clean and accurate.
- There must be sufficient reference data available to support analysis from different perspectives. This includes groupings and categorizations that are used for reporting but are not required to keep the transactions flowing.

Most organizations cannot meet these key objectives by using their data directly in their operational systems, and this continues to be the case for a long time. If you can't effectively use the data in your current operational systems, you need to build a data warehouse. This book will help you understand what a data warehouse is and highlight the key things that need to happen to ensure its success.

---

## **Question: Will the data warehouse replace our old systems?**

---

The short answer is no, a data warehouse does not take the place of any operational application systems. The systems that run the business will continue to function. The data created and captured by those application systems will be used for business analysis and therefore become the input to the data warehouse environment. There is a possibility that existing reporting systems (used to manage the business) may be retired, as the data warehouse can support the reporting and analysis currently done with that system. Often, these reporting systems are not meeting the needs of the business, creating the impetus for building a data warehouse. These older reporting systems may not have all the data needed to support the business today, and they were likely built using outdated technology. When the appropriate data is available through the data warehouse, newer tools can be used to recreate critical reports and analyses needed to provide continuous support to the business.

Some reporting and analysis systems have a great deal of sophisticated logic already built and are meeting a specific business need. In order to maintain consistency across the organization, the data warehouse may be used to feed data into this specialized system.

---

## **Question: Who needs to be involved?**

---

Broad organizational involvement and active participation is needed for a successful data warehouse project. The exact level of involvement varies by group within an organization, but active participation and commitment to the success of the data warehouse is necessary from both the business and IT groups. Executive sponsorship and active ongoing support is critical. This support can be a significant factor in the success of the data warehouse.

There is also a need for assistance from representatives from the target audience of the data warehouse. Most of these business participants will provide input, review and critique the design, and assist the team in understanding and properly handling data-related issues. This does not usually require a large time commitment, but their involvement is important to ensure that what is designed and built will indeed be helpful. These participants should represent more than just the initial target audience. Other closely related functions or groups that will benefit from using the data should also provide input. This enables the project team to take these additional needs into account early in the process in order to avoid or minimize future changes needed to support these other groups.

It is extremely important to have one or more individuals from the business be full-time members of the project team. These representatives serve as the daily liaison between the project team and the rest of the business community. They also actively assist the project to understand and document requirements. This includes overall business requirements as well as the definition of individual data elements and the business rules for processing and handling the data. While this business analyst does not need to know all the answers, he or she does have to know who else to turn to.

A core set of IT personnel will form the project team. This is usually comprised of a project manager, a systems analyst, a data modeler or architect and developers. There are often two different types of developers: those who write the ETL system and those who build the business intelligence (BI) application. Other technical people are needed periodically to support network, security, and database needs.

Chapter 4 describes the specific roles and responsibilities for both business and IT personnel involved in a data warehouse.

## **Question: Do we know where we are going? How will we know when we get there?**

---

It may seem as though there is a lot of activity but the actual vision and final result are not clear. This can be a result of a number of factors. Sometimes there is no formal vision or strategy in place. Other times, there is a vision but it has not been clearly communicated. In both scenarios, a lot of effort can be spent without any real, sustainable progress. It can be easy to fall into this pattern. For example, when an immediate need crops up to address one part of the business, we scramble to get a quick set of reports created. Then another area has a different urgent need, so a dashboard is created to support that group. Without a clear and communicated vision, these activities can consume your organization. The challenge is to balance these short-term needs while working toward long-term goals.

## **Answer**

First, it is important to figure out where you are in the process. Techniques for how to do this can be found in Chapter 3. Based upon these findings, there may be a need to better understand business requirements (Chapter 6), stabilize the technical environment (Chapter 9), or perhaps develop a more strategic view of the data (Chapters 8 and 9).

Successful data warehouse environments are built over time with a clear direction and goal in mind. This includes the business objectives and both

technical and data architectures that show how this can be accomplished. The technical architecture defines what tools and technologies will be used to perform the many different functions required. This includes the hardware and database management system to effectively store the data, and the data access or business intelligence tool needed to access that data. Data architecture provides the guidelines outlining how data will be organized to achieve the business goals.

Take the time to set up a vision and long-term strategy for data warehousing. This provides the roadmap showing where you are going and provides the capability to determine how close you are getting to that vision. There will be benefits along the way, so you should not have to wait for the entire vision to be built before you see any results. The vision is what provides guidance so that each individual effort or project is helping to take the organization in that direction. By following through on the vision, you *will* know “when you get there.”

## **Question: How do we get started and stay focused?**

After reading about data warehousing or attending a conference, there is often a high level of excitement and enthusiasm. The potential is clear and everyone wants to get moving. Rather than rush forward, however, take a small step back and do an assessment of your current situation and conduct some solid planning.

## **Answer**

The first place to start is to ensure that you have a good grasp of what you already have in place. Chapter 3 can help you with this. After that, you can define one or more projects that address the action items that were identified when you evaluated the current situation.

Some organizations may only have limited reporting or data warehouse efforts under their belt. This is really an opportunity to start with a clean slate. When possible, adhere to proven methods—there is no need for you to blaze a new trail. The proven methods have established a core of best practices that have evolved from years of experimentation, with many scars to prove them. Take advantage of what others have learned in order to build the best data warehouse that you can.

Many organizations are years into a variety of data warehouse or similar efforts—usually with varied levels of success. This is the time to collaborate between the business and IT groups to set the vision for data warehousing in your organization. With a clear vision in place, individual projects can be

defined to work toward that goal. Step by step, you can move closer to that goal. When possible, craft these individual projects so that some business value can be realized with each step. The benefits will grow as you move toward your vision.

## **Summary**

---

The process to design, build, and use a data warehouse that will give you a solid return on your investment is complex and challenging. The good news is that there are proven methods, techniques, and tools to accomplish these steps successfully. Understanding what has already been achieved in the data warehousing industry will help your organization improve the quality and value of its data warehouse. This part of the book is just the tip of the iceberg. This basic overview may be sufficient for some people, but most people from both business and IT will benefit from more in-depth coverage of these topics throughout the rest of this book. As suggested in the last question, it is important to have a good understanding of what you currently have in place in order to address the organization's reporting and analytical needs. The next chapter shares an approach to doing just that.

**PART**

**II**

# **The Business Side of Data Warehousing**

---

## **In This Part**

---

**Chapter 3:** Understanding Where You Are and Finding Your Way

**Chapter 4:** Successful IT-Business Partnerships

**Chapter 5:** Setting Up a Successful Project

**Chapter 6:** Providing Business Requirements





# Understanding Where You Are and Finding Your Way

Most organizations already have some mechanism in place to meet their reporting needs. This mechanism may be a set of reports that run directly against the production application systems or reports generated from an existing data warehouse. Many organizations have multiple data warehouse attempts under their belt. Some data warehouses have been successful, while others have never delivered value. In many cases, reports and analysis are not supported in a formal system at all. Data is pulled from wherever it can be found, and manually merged into spreadsheets to create reports and presentations, and used for analysis.

Because you are not starting with a clean slate, it is important to get a realistic understanding of what reporting and/or data warehouse components and resources are available and to what extent they are being used.

Making an honest assessment of what has been accomplished to date is often one of the most difficult things to do. Several years of hard work may have been invested to build the system currently in place. It usually takes longer and costs more than what was originally planned, and the results may have fallen short of expectations. What should be done? Take a deep breath, step back, and take a little time for reflection. This chapter lays out the steps to help:

- Understand the current business situation.
- Identify how reporting is currently accomplished.
- Assess the effectiveness of existing reporting mechanisms.
- Determine what to do next.

## Assessing Your Current State

---

Rather than jumping in either to learn about technology or to dissect an existing data warehouse, it is better to step back a little further to gain a

wider perspective. Once you have a good understanding of the current overall business environment, then it is time to dig into assessing the data that is already in place. The success or failure of the current data warehouse might highlight deeper issues within the organization.

## What Is Your Company's Strategic Direction?

Reading the corporate vision statement or even the beginning text of the annual report can provide a great deal of insight into the goals and direction of the organization. This direction dictates what is most important to the organization and where time and energy should be spent. This is much larger than any data warehouse effort, but it can provide insight into how a data warehouse can help support the organization's efforts to achieve that vision. A few examples of company goals and their associated opportunities include the following:

- The goal to deliver high-quality products could be supported by using manufacturing, product returns and defects, and sales data.
- The objective to provide superior customer service could benefit from looking at a comprehensive view of each customer across the enterprise, and by tracking service calls and customer feedback.
- The objective to retain one's status within the industry as an innovator could be supported by providing access to reporting and analysis of research and product development data.
- The vision to ensure financial strength and stability could be supported through timely access to financial results from the general ledger and tracking organization reserves and cash flow.

In addition to the stated vision for the organization, there are also other characteristics of how the organization works that can affect the success of any data warehouse project. Each of the following scenarios describes a different company. A brief description of the company's situation is followed by the potential impact on any data warehouse initiatives:

- **Survival mode:** The company is cutting costs and simply trying to stay afloat. Long-term investments are not an option at the moment. Quick turnaround and limiting expenditures are the name of the game.

*Ramification:* This short-term focus is not good for strategic, long-term investments such as a data warehouse. If progress and value are not realized in a matter of months, then projects are cut. For a data warehouse project, this does not allow the organization to address fundamental data issues and enterprise integration. When data warehousing principles are

applied to individual component solutions, there may be immediate, short-term benefits, but you end up with multiple non-integrating data marts. These solutions are often discarded because they do not address a broad perspective but focus only on one thing at a time. A short-term fix does not equal a sustainable long-term solution.

- **Staying ahead of the game:** Business has always been good. New competition has entered the marketplace, increasing competitive pressures. Things are OK today, but the company must continue to build its strategic advantage.

*Ramification:* The company does not want to waste resources on throw-away solutions. There is a commitment to building a long-term strategic advantage. Investment in a data warehouse will support the business initiatives for strategic growth. It is worth taking the time to build the data warehouse correctly.

- **Changing with the times:** A well-established business needs a face-lift to compete with new, nimble companies. There is a strong, stable client base, and the organization is financially strong, but it's not as flexible as newer companies and is perceived as being a bit stodgy. The organization is looking for new business channels and product offerings to capture the interest and business of the younger generation.

*Ramification:* The organization already has policies and procedures in place for just about everything. Business practices are engrained and there isn't much interest in change because "this is how it has always been done" and it appears to be successful. Looking at the bigger picture, however, there is a need to shift the organization's culture to be more open and creative to address the challenges of the future. The data warehouse can help to provide new ways to get things done. Specific plans need to be made to help the organization through this cultural change. Concentrated efforts will be required to get the data warehouse built, but then it can serve as a catalyst for change.

- **Back to basics:** The company has branched out over the past several years. This has begun to dilute the long-standing core competencies of the company. Not all of the expansion efforts have been successful. There is a need to focus on the core business once again.

*Ramification:* Mature systems for the core business keep things running but do not provide an overall view of the entire enterprise. Rather than just abandon all of the new business areas, it is necessary to carefully analyze them, keeping the things that are most profitable and eliminating the unprofitable ventures that deviate too far from the core business. This requires having a thorough understanding of what is working and what is not. Because the organization has been taking risks with the expansion,

it will likely be comfortable delving into the data warehouse arena. New things do not scare them. In fact, the data warehouse can help channel all the creative energy into the core business.

- **Global innovator:** The company has always been considered a leader in the industry. Now is the time to take advantage of the changing marketplace. There are great opportunities to expand into global markets, but this aggressive growth must be done carefully to avoid overextending the organization's resources.

*Ramification:* Now is the time to ensure that the business has a sound foundation from which to launch new initiatives. This foundation includes strong business processes, application systems, and a data warehouse. The data warehouse can be used to help monitor the impact of expansion on the organization's ability to deliver high-quality services to customers.

These scenarios simply highlight how the vision and culture of the organization can affect data warehouse efforts. There is an impact whether you are starting from scratch or you already have a data warehouse in place.

## What Are the Company's Top Initiatives?

Every organization's attention and resources are limited. These resources tend to be allocated for the biggest crisis and for initiatives that will help the company meet its strategic objectives. Therefore, it is important to identify the focus of the company's resources today. If the data warehouse project is not directly supporting those things, then it will be put on the back burner. This will affect the time available for the business to spend on the project, as attention will be focused elsewhere. No matter how many IT people are assigned, the project will either experience delays as other higher-priority tasks take precedence or, worse, forge ahead without input from the business. In either case, the data warehouse will fall short.

Sometimes a data warehouse is not part of an immediate solution to help the company. If so, it may be best to set the data warehouse plans aside for the now. When the organization is ready, then dust off the plans and start again. Too often, IT organizations continue to forge ahead without a strong business objective or purpose. The rest of the company's energy is elsewhere. Usually these initiatives fail. Data may be loaded, but it may fail to reflect what is really needed, and is therefore never used to help the business. Don't be afraid to call a time-out. Stop spending money and wasting time. The project can be restarted when the resources will be used to move the company forward.

Another way to help determine where the data warehouse fits into the big picture is to look at management or oversight studies that your organization may have already done. These are usually conducted by high-level consulting

organizations, and their findings are well documented. Review the findings. Ask yourself how, if at all, the data warehouse can support the organizations' efforts to meet current challenges. Often it is clear how an existing or planned data warehouse can directly support the company's goals.

## How Healthy Is Your Data?

Data is needed to help an organization achieve its strategic goals, but if the data is not accurate, current, integrated, and accessible to business users, it will inhibit the company's ability to meet those goals. Therefore, an honest assessment of the state of the data must be done.

Business and IT staff generally have a good sense of the overall state of the data for your company, but certain aspects of the data can be overlooked. Is there clean, reliable data within each individual application system, but the data does not integrate easily, if at all? Perhaps your newest systems have good data quality, but the older systems do not. Don't conduct a detailed data quality study at this time, but rather think about the data that is currently available. Sometimes the data is sound, but it is only available to those who know the secret handshakes that unlock it from storage. Some symptoms of poor data health include the following:

- Regularly finding duplicate data, within and across application systems. For example, there are multiple places where customer data is stored, and it is well known that there are multiple instances of some customers.
- Pulling data together from more than one system is too difficult for all except the most data-savvy business analysts.
- There are often miscommunications between report requests and getting the actual data needed to make a decision.
- There are often errors in the data that require restatements and corrections.
- There are many concerns about the accuracy of the data.
- Data is not requested because it takes too long to get in time to make a decision.
- Costs to access and maintain the data are escalating.
- There is little documentation about what data means or how it is stored. This is a concern because the last person who really knew the data just retired.

The organizations with the biggest problems often expect a data warehouse to perform miracles. Significant data problems do not materialize over night. Generally, it has taken years for the application systems to grow and evolve into their current state. Many systems are well designed at first, with sound

database designs, but become disorganized over time, such as when the application is modified to meet changing business needs. Systems professionals don't design systems or make changes to intentionally produce bad data. Intense demands are placed on these systems to provide functionality that was never part of the original design, and the changes must be done in a timely manner. As this happens repeatedly, even the cleanest database design can be degraded.

If there are long-standing problems with the company's data, then it will take a combined effort of IT and business professionals to correct these deficiencies. This includes business decisions about how to handle the data, changes to the production application systems, changing business processes, and adding cleansing and transformation rules to the data warehouse ETL processes. The data warehouse team may be at the center of it all. The bigger the problems, the longer it typically takes to correct them. It *will* take time.

## **Does the Business Place Value on Analysis?**

An organization's readiness for a data warehouse reflects the value that is placed on fact-based decision-making and reliance on the results from business analysis. A data warehouse is a mechanism for the organization to better leverage data and support more sophisticated analyses. How can you tell if your organization is ready? Several characteristics indicate the value that an organization places on analysis, including the following:

- There are clearly defined roles and/or jobs for business analysis.
- There is a career path for business analysts, which ensures more senior positions without going into management.
- The people who perform business analysis are well respected throughout the company.

If an organization does not place a lot of value on analysis, this can limit the willingness of the business community to participate in the design and development of a data warehouse. Without this partnership, organizations often fail to realize benefit from their data warehouse. Certain industries have developed a strong appreciation for fact-based decision-making and the value of analysis. The financial solvency of the insurance industry is highly dependent upon the accuracy of analyses performed by the actuarial groups. This analysis helps the company to understand risks, determine the types of risks to handle, and set competitive rates. Consumer packaged goods and retail industries have been investing and developing strong analytics to improve inventory management, identify customer purchase patterns, and develop effective promotions.

Some organizations do not have a strong track record of using data and may not have much experience performing sophisticated business analysis. This does not mean these organizations should avoid building a data warehouse; it just means that additional effort must be spent on developing analytical skills. This also means that it will take longer for the data warehouse to evolve to support the business because the business itself must evolve in its use of data and the types of analyses that will be performed.

## **Reflecting on Your Data Warehouse History**

In order to get a good understanding of the big picture, it is worth taking some time to reflect on the experiences that your organization has had with a data warehouse. In many cases, more than one data warehouse or data mart are in place. Questions that you can ask about each data warehouse include the following:

- Is this an enterprise data warehouse or a collection of non-integrating data marts?
- Does IT consider the data warehouse to be a success? If not, why not?
- Does the business community consider the data warehouse to be a success? If not, why not?
- Were business goals set for the data warehouse? Have these been met?
- Who, from the business community, was involved in designing and building the data warehouse? Was this active participation or periodic status updates?
- Does the business feel that this is their data warehouse, or did IT build it with the hope that the business would find it useful?
- Is the data warehouse still being used today? If not, why not?
- From an overall perspective, does the organization have a broad data warehouse strategy in place today?

The objective is to get an overview of what has happened in the past. If previous data warehouse projects have been successful, then the organization will collectively be more positively inclined toward any new projects. If all previous data warehouse attempts are considered failures, then you are likely to run into resistance in getting another one started. It is critical to understand the root causes of any failures so that new projects can develop specific strategies to avoid past mistakes. The project team needs to effectively communicate what is being done differently this time so that this effort will be a success.

Digging into the details about individual project success or failure is explored in the next section. At this point in time, it is important to simply get an



appreciation for what the organization has experienced. This will give you insight into people's attitudes about data warehousing. Are people fed up and think that a data warehouse will never work for your company? Perhaps the company experienced some early success and everyone is clamoring for more.

If there are deep-seated problems with data, and business analysis is not highly valued, then you are likely to face serious challenges building a data warehouse. This does not mean that you should stop; it just means that you need to address these fundamental issues at the same time and not expect the data warehouse project itself to fix problems that have been building for decades. Specific activities must be planned to address these overarching issues. The data warehouse can be a catalyst for changing how the organization uses and values data.

## **Understanding Your Existing Reporting Environment**

---

Now that you have taken a step back to survey the big picture, it is time to get your arms around how reporting is being done today. Many reports are generated from some sort of reporting system, which may or may not be a data warehouse. The objective here is three-fold:

- Find the reporting environments.
- Learn about the data that is available via that environment.
- Identify the technology used by that environment.

Each of these steps is described in more detail in this section. One or two people from IT or the business community can do this work. While some of the information to be collected is technical, the people gathering the information do not need to have deep technical skills. This should be done by someone who is very organized and persistent, often a senior business analyst, a project manager, or a senior systems analyst. The effort to dig deeper into the current reporting environment can span the enterprise or may be focused in only one area, such as sales or finance.

If the work is to collect information across the enterprise, then the people doing this work must be authorized and empowered by enterprise management. This gives the team the authority to cross division or department boundaries. The purpose is to get a basic understanding of the big reporting systems or data warehouses across the enterprise. At the enterprise level, the work described in this chapter can be completed in several weeks. This type of effort will not be able to identify every set of personal spreadsheets that are used for reporting, but should be able to touch on those environments that produce the most critical reports for the organization.



When focused in a single area, the effort should still take only several weeks, but may include more research into all types of reporting, such as individual spreadsheets.

## **Finding the Reporting Systems**

The first step is to compile a list of all the different systems that are used for reporting that exist across the organization. Some of these will be highly visible and labeled as a data warehouse or data mart. Many other systems are usually in place to support reporting and analysis. Some of these will look like a data mart, even if not labeled as such. Other reporting systems may not look anything like a data warehouse, but provide reports for the business. The challenge now is to find these other reporting systems. If you have a well-established enterprise data warehouse strategy, these other systems may want to stay hidden so that they are not expected to participate and/or comply with the enterprise data warehouse. There needs to be a joint effort between the business and IT to locate these systems. These can be found by the following methods:

- Searching for systems with names that include reporting, analysis, benchmark, scorecard, dashboard, executive information system, or decision support.
- Reviewing the IT budgets to locate maintenance and support for reporting systems. The IT budgets may also include new projects that will soon fall into a data warehouse-like category.
- Reviewing the budgets of key business groups for which analysis is usually performed. Systems may be developed, maintained, and supported completely outside of IT. Business groups build these to get their work done. Groups that commonly have their own analytical systems include actuary, market research, business planning, forecasting, brand management, merchandising, and financial reporting.
- Asking the database administration group to provide a list of reporting and analytical databases.
- Looking for groups or projects that use sophisticated data access and business intelligence tools. The purchasing department may be a good resource to identify who is spending money on these tools. Some of the tools to look for include Business Objects, Cognos, Microsoft Analysis Services, Hyperion Essbase, and SAS. This is not a complete list, of course, but points to the type of tool you are looking for.

In many cases, reports are created outside the realm of a formal system developed and maintained by IT. These may be a collection of spreadsheets or

reports in an MS Access database that are handled by someone in the business group. Often, reports created this way are generated on a regular basis and the business depends upon them. This qualifies them to be included in this inventory. This is not an exercise to compile a complete inventory of all reports or spreadsheets, but rather to discover the core collection of reports. These are often handled like a real “system,” just not under the care of IT. Additional questions that can help locate these include the following:

- Where do your standard reports come from?
- Is there historical data stored within the business group, such as in spreadsheets, SAS data sets, or MS Access?
- Is there a research group that collects and maintains its own data?
- Is there a process (often manual) that is used to produce the most useful reports for this group?

In addition, is there someone who everyone depends upon to get data and to produce reports? If so, then explore where and how this work gets done. When a business analyst is doing this work, these informal reporting environments are often already known. However, if a project manager or IT analyst is doing this research, then this can be much harder. When the business groups cooperate, this yields more accurate and complete information. This partnership can be encouraged through business management.

If your goal is to understand reporting across the enterprise, then only the largest and most critical reporting environments need to be included. It could easily take months to try to track down every instance of this less formal type of reporting for an enterprise. It would be appropriate to dig into this much detail if you are studying a single part of the organization.

## **Compiling an Inventory**

Now that you have a starter list of the systems of interest, the team can continue their work by collecting more information about each of them. When asked to do this type of survey, people usually jump directly to researching the tools and technologies that are used. That level of detail will be helpful, but it is important to get a basic understanding of each of these systems first. Start with the biggest and most well-known systems, possibly the data warehouse or a data mart. Then progress to those systems with the broadest potential impact. Leave for last any reporting systems that seem to be narrow in focus, such as the reporting system used by facilities management to track the replacement of light bulbs.

In order to collect this information, you need to find the person(s) who knows the most about each of these systems. If possible, identify both a business and technical person for each system. The technical resource, is the

individual who is responsible for keeping the system running, who may report directly into the business unit and not IT.

**TIP** Don't send out a questionnaire and expect to get responses. Schedule a face-to-face meeting to gather the information. This provides you with an opportunity to begin building a relationship with representatives from each of these systems. Personal meetings also enable you to observe attitudes toward data warehousing.

Several layers of detail will be helpful for learning about each system. Each of the layers is described in the following sections. It may be necessary to meet with several different people about a system before getting a complete picture. You may need to speak with the project manager, business application owner, lead IT developer, or the primary business user. The layers start with the business purpose and then dig deeper into some of the mechanics of how the system works.

**TIP** As you begin finding and researching these reporting systems, make sure that everyone knows that this is not a witch hunt. You are not planning to fire the three people responsible for the last project failure. The goal is to gain an honest assessment of what assets the organization has in place. Some reflection about how you got to the current point may be useful to catalog lessons learned, and things not to repeat. The most value comes from understanding exactly where you are. Then your effort can be applied to figuring out where you want to be and developing a plan that outlines how to get there. Without the reassurance that this is not an exercise in laying blame, many individuals will be hesitant to speak openly about what issues the organization is facing.

### *Identifying the Business Purpose*

You first need to get a basic understanding of what the system does. For example, a reporting system may provide financial reports to the manager of each distribution center. This is the most important information you need to understand because it forms the basis of your insight into the overall value and impact that this system is having.

Aspects of reporting that can get the conversation going include the following (be sure to determine who is able to answer these questions—a systems group or a business group):

- Describe the overall business purpose or function of this system.
- What are the top five business questions that the system helps to answer?
- Why was this system built?

- What was the business justification for the project?
- Who is the business sponsor for the system?
- What business groups is this system intended to support, or who is the target audience?
- How many business users are using it? How frequently? To answer what type of questions?
- Who is the recipient of the results of these reports and/or analyses?
- Is this part of the data warehouse initiative? If not, why not?
- When was this system built?
- What are the biggest known challenges or problems?
- What are the biggest benefits from this system? Who or which group benefits?
- At what stage is this system in its overall life cycle: Design and development? Initial deployment? In production for multiple years?
- Are any return on investment (ROI) figures available? Do they indicate cost savings or improved business results, such as increased profit?
- At the end of a session, always ask what else should be known about this system or the group. For example, what are the expectations and/or concerns about a data warehouse?

The objective of asking these questions is to get an understanding of how these systems are helping the business. For some systems, you may find many of these questions to be difficult or impossible to answer. Not getting answers is just as important as getting a lot of detail. The lack of a business purpose or focus is a common problem for many data warehouse-like systems.

When you receive positive feedback for a reporting system that is not a data warehouse, this is good. Business needs are being met. There may not be a need to build data warehouse at this time. However, you still need to look beyond the immediate users of any reporting system to determine if there are other areas that need access to this system and/or data. Frequently, the existing reporting system is not meeting the needs of the business, which is why there is interest in data warehousing.

If you are looking at a data warehouse and your answers reveal sound business responses, congratulations! This indicates that the data warehouse is supporting the business and adding value to the organization. This does not mean that everyone should sit back and get comfortable, because there is always more that can be done. Effort can be spent on improving and gaining even more value, rather than on remediation work. Unfortunately, for many data warehouses, these questions cannot be answered at all, or the responses are phrased only in technical terms. This failure can indicate a

variety of problems. It could be that the appropriate people were not polled, which may indicate communication problems. A lack of business-oriented responses may also indicate that business results have not been captured and tracked. With further exploration, the true value of the data warehouse may be discovered. Don't be shocked if no concrete value can be identified from the data warehousing environment. Your organization is not alone. Many data warehouses are less than successful, and some are a complete failure.

### ***Discovering the Data You Already Have***

Of course, another critical part of these systems is the data itself. You don't need a list of specific data elements, which is at too fine-grained a level of detail. More useful is an overview of the type of data that is stored and used in this reporting environment. Questions to help you understand each reporting environment include the following:

- What type of data is included?
- What level of detail is included? Individual transactions or summaries? At what summary level?
- How much history is provided?
- How often is the data updated?
- From which source systems is data pulled?
- What is an estimated size of the data (terabytes or megabytes)?
- What is the confidence level in the data (high, medium, or low)?
- Are there any plans to make significant changes to the data?
- Is a specific data warehouse architecture or methodology being used for this system (e.g., Inmon or a Kimball approach)?
- How is the data organized? Is it in flat files? Is the data structured dimensionally?

Document what you learn about the data. Table 3-1 shows an example of a data summary. Complete what you can, but don't take months to track these details down. Simply capture what you can in one or two meetings. When this is collected across all the different data warehouse-like systems, it is easier to highlight areas of overlap and possible redundancy. This alone cannot be used as justification to eliminate one system versus another. There are often many valid differences that require additional research and analysis to determine what should or could be done. This is simply a mechanism to understand what you already have in place.

**Table 3-1** Data Summary

TYPE OF DATA	DATA SOURCE	DATA ORGANIZATION	LEVEL OF DETAIL	HISTORY	UPDATE FREQUENCY	EST. SIZE	CONFIDENCE*
Retail Sales	Point-of-Sale System	Dimensional	Transaction Line Item	3 years	Daily	650 GB	Medium
Store Inventory	Logistics Management System	Dimensional	Inventory Transactions (In and Out)	3 years	Daily	25 GB	High
Sales Forecast	Logistics Management System	Dimensional	Weekly Forecast by Product and Store	2 years	Weekly	10 GB	High
Promotion Calendar	Marketing Group Spread Sheets	Dimensional	Weekly Promotions	2 years	Monthly	1 GB	Low

\* This is not a quantitative measurement of data quality, but rather the perception of the data: Does the business community feel the data is reliable and trustworthy?

### ***Understanding the People***

One of the most useful things to learn about any reporting environment is who knows the most about it. This includes both IT and business professionals who are involved in building, maintaining, and using the reporting environment. You can determine which people should be contacted and interviewed by asking relevant questions such as the following:

- Who answers questions about this system? Who knows the most within the business community? Who supports this from a technology perspective?
- Who creates new reports or analyses?
- Is anyone responsible for standard reports and analyses?
- Who is responsible for data quality? (It may be helpful to ask instead: Who do you go to if there is a problem with the data?)
- Is there a team of people who are dedicated to supporting this environment or is this system only a small part of their job responsibilities? To whom do they report?
- How long has the support team been in place? Who is the most senior member of that team? How long have they been on the team? How long has this senior team member been with the company?
- Are there any more experienced staff members who help the team? To whom do they report and how often is their expertise used? Is it easy to get their help or does it require the coordination of several managers?
- Are there any enterprise data warehouse resources available to the team? Have they been used? If not, why not?
- Have any third-party organizations assisted the group with this system? Explore who was involved with the original design and development, as well as ongoing support.

These questions are geared to help understand the breadth and depth of the people who are currently involved in supporting and using these reporting systems. Such questions also explore how much each individual group knows about what resources are available to them and whether they are aware of any data warehousing activity.

### ***Tracking Technology and Tools***

While it is important to understand the business value of a data warehouse, you also need a basic understanding of the underlying components that together deliver that value. You need to know what you have in place, how well it is

working, and what plans are already established to address or change these components. This is much more technical, but it is still worthwhile to collect.

There are many different ways that reporting systems can be created. Even for a data warehouse there are many components involved. You need to understand what technology is being used to extract, prepare, publish, and use the data. Chapter 9 provides more details about each of these components when they are part of a data warehouse. For now, ask the following questions to help identify the technology and tools currently being used:

- What technologies or products are used?
- Are these are the most current release of the product?
- Is this component stable?
- Does this component perform acceptably?
- Is this component supportable?
- Is this component sustainable for the future?
- Are there any plans in place that will modify this technical environment?

Table 3-2 shows an example of how to track this technology inventory. By documenting a summary of technical details, you can keep track of all the different moving parts of the system. You may ask what technology each system uses for the technical components listed in Table 3-2. More information about what each of these different technologies and tools does is provided in Chapter 9.

In addition to the Technology Inventory Matrix, it is helpful to document the answers to the other questions. One group may be using the most current release of a tool, which is meeting their needs. Another group may use an earlier version of the same product but without the same level of satisfaction. The Technology Inventory Matrix provides a handy cross-reference for finding the appropriate detail for a single tool. You need to look at the details for each group that is using that tool.

### ***Understanding Enterprise Resources***

As you did for each system, be sure to take a moment to understand what is already in place at the enterprise level. This may include data, technology, people, policies, and procedures. If these have not already been noted with a specific system, then be sure to explore these resources now. There may be a well-established enterprise data warehouse team, a documented data warehouse architecture, a central library of all project documents to share best practices, or you may not find anything at an enterprise data warehouse level.



**Table 3-2** Sample Technology Inventory Matrix

TECHNICAL COMPONENT	SALES DATA MART	FINANCIAL REPORTING SYSTEM	CORPORATE MARKETING CAMPAIGN MANAGEMENT
Data Modeling Tool	ER-Win	None	None
Data Extract Tool	Informatica	Informatica	SAS Programs
Transformation Tool	Informatica	Informatica	SAS Programs
Data Movement	Informatica	Informatica	SAS
Data Cleansing	NA	NA	NA
DBMS	Oracle	Oracle	Essbase
Database Load	Oracle	Informatica	Essbase
End User Data Access	Business Objects	Cognos	Essbase
Data Dictionary	Internal Data Dictionary	Internal Data Dictionary	CRM Data Dictionary

**DOCUMENT WHAT YOU LEARN**

Going through the exercise of finding and learning about the many reporting environments across the enterprise is challenging and enlightening; but in order to leverage what is learned, the findings must be documented! Taking notes is not enough. Can others find your notes? Can anyone else read your writing? Taking the time to document what you learn ensures that the details are available even when you are not. Large, complex organizations have too many systems to keep track of unless you carefully document each one.

This should not be an effort that takes months—spending an hour or two to summarize the findings of each system is invaluable. Organize the findings and publish them for the benefit of the entire organization. You should plan to document the following:

- A brief summary of each reporting environment, which includes a short abstract of the system's purpose and contact information if someone wants to learn more. This can include the technical findings too.

*(continued)*

**DOCUMENT WHAT YOU LEARN (continued)**

- Data Summary Matrix
- Technology Inventory Matrix

**This can form the basis of an enterprise data warehouse web site that provides information about all data warehousing activity across the company. Clearly indicate that the information collected at this time reflects the current state. Later, additional information can be added to this site to share the direction and strategy of the organization's future data warehouse efforts.**

## Netting It All Out

Think about what has been learned. Were you surprised by anything? Are things better or worse than you expected? Too often we spend all our time performing tasks. While things must get done, it is also helpful to sit quietly and think about what has been observed. After taking some time for reflection, it is helpful to bring together everything that was collected through the inventory.

If you did not document what you learned about each reporting environment, then take the time to do so now. Once each separate system has been documented, look for common themes across these systems. Are there common areas of concern? Is the same or similar data flowing into multiple systems? Are multiple tools in place that perform the same function?

The answers to these questions will highlight what is in place and working, and what problem areas need to be looked at more closely. This also highlights any activity that is currently outside of any formal data warehousing initiatives. These outliers would not exist in a perfect world, but reality reveals that not every report or analysis will be supported via a data warehouse. The challenge is determining which things make business sense, and should therefore be supported by the data warehouse, and which things to leave alone.

It is impossible to list all possible findings here and make recommendations. Business and IT managers should review these findings together; and with the help of the team who compiled the findings, the group must determine what should happen next. This requires an understanding of what you already have compared to the organization's strategic direction. Combining this with a basic understanding of data warehousing, the group can assess what needs to be done next. To show how this may play out, several scenarios are provided here, with the recommendation for what to do next:

- The organization has a data warehouse in place, but it is stalled. It is not helping the business today. There is a great deal of interest to leverage the investment that has already been made.

*Recommendation:* A short project can be undertaken to study what needs to be done to improve the data warehouse's ability to support the business. Chapter 12 discusses several ideas to help jump-start a stalled data warehouse. This will likely spawn one or more additional projects to address the specific recommendations of the study.

- The organization has a well-defined data warehouse strategy with a core set of data already loaded into the data warehouse. It supports several key business groups but additional analyses are needed, which require additional data.

*Recommendation:* Define a project to expand the data warehouse by loading this additional data. This project should also design and deliver the reports and analyses needed to continue to support the business. Keep up the good work!

- There are multiple non-integrating data marts. Each was built a number of years ago and has provided value, but there are new requirements for which the data must be used together.

*Recommendation:* Assess the effort that would be involved to merge the appropriate data marts. Depending upon how each is designed, this may be an opportunity to create a new solution or it may be better to enhance one of the marts to accommodate the other data. This will likely result in a project for the new design or expansion.

- One of the most relied upon reporting systems is not able to keep up with the demand for new reports and analyses. There have been rumblings for years about how much it costs to maintain this system. It is no longer cost effective to continue to run this system as is.

*Recommendation:* If there is business need to continue using the data that currently lives in the old reporting system, then this is a good candidate to be loaded into a data warehouse. A new project can be defined to design and build the solution. This must focus on what the business wants to do with that data, rather than simply replace each existing report. Where possible, additional functionality should be delivered to the business that will increase the value of using this data once it is in the data warehouse.

- The organization has dozens of reporting systems and several separate data warehouses already. There doesn't appear to be any cohesive strategy to tackle reporting and analysis.

*Recommendation:* Initiate a project to develop a data warehouse strategy and define what is needed to realize that strategy. Chapter 9 provides a description of data warehouse architectures and what needs to be put in place.

- While reviewing the overall direction of the organization, it becomes clear that its biggest need is the ability to get a complete view of customers and what products and services they buy. This data is currently in many different reporting systems. There is no data warehouse in place today.

*Recommendation:* Because this is a big priority for the organization, a project should be started to design and build a data warehouse to provide this critical customer data.

There is no single answer for how to move forward. Experience shows that once you know both what you have and where you want to be, there are usually clear steps for moving in that direction.

## **Introducing the Case Studies**

---

There are two different types of case studies included throughout the book. The first type provides examples of the concepts presented in the chapters. The second type of case study allows us to follow the progress of two different companies on their data warehouse project. Each serves a different purpose, but they are intended to clarify and reinforce the main concepts covered in each chapter.

### **The Call Center Data Warehouse Project**

Sample findings and project deliverables are included for a data warehouse project for a call center. Many organizations run their own call center operations or outsource them. This is a critical operation because it is typically how external parties interact with the company. Both the public's impression of a company and customer satisfaction can be greatly affected by a good or bad experience when calling that company. Because this transcends a single industry, the call center example highlights the universal principles of the book.

### **In Real Life**

To highlight how the concepts presented in each chapter play out in real life, the experience of two organizations will be shared. These are not actual companies, but rather a compilation of many different organizations. The two organizations that will be tracked represent opposite ends of the spectrum in terms of size and culture.

### ***Giant Company***

The first organization, which we will call Giant Company, is an extremely large conglomerate. There are thousands of employees in IT alone; and there are multiple divisions, each run as a separate company. Because of the large number of people employed, there are many layers of management and well-structured positions and career paths. As the company grew, many policies and procedures were put in place to assist in the overall management of such a large concern. As employees are brought on board, there is a pre-defined training plan based upon the position. This helps everyone get up to speed as fast as possible.

Within each of the divisions, the business is organized into departments with specialized functions. Each department knows what its inputs are, how to get its business processes done, and to whom results should be sent. Much of the business process and decision-making for the daily work has been built into the production application systems. This has enabled the business to expand and ensures consistency. However, knowledge about the rationale and details of the processes has been lost over time. Everyone knows how to keep the systems working, but how they work and why is a mystery.

The IT function is centralized, but it is structured to support each of the divisions. To keep it manageable, a clear project methodology is used. This provides consistency from project to project and helps individuals to understand how to get things done. With so much structure, it is relatively easy to get everyone to participate in enterprise initiatives. Everyone is used to having to comply with corporate standards for everything—from naming a data element to requesting a new laptop. Following corporate policies is rewarded, whereas questioning the status quo or trying to circumvent well-established procedures is highly frowned upon.

Within each division there have been attempts at building a data warehouse, with a range of results. Some have mature and valuable data marts in place. Others have been struggling to deliver anything. There have also been several attempts to build an enterprise data warehouse. The goal each time was to consolidate all data from across the enterprise and then serve it up to individual groups as needed. Most of the efforts to date have been more focused on infrastructure and moving data into a large database. There has not yet been any true delivery of this data into the business community.

With such a variety of data warehouse experiences, it is especially critical to compile an inventory of what has been done. There is a need to learn more about industry best practices and how these apply within Giant Company. Understanding what hasn't worked for this organization will help other

initiatives to avoid making the same mistakes. With so many layers of management and narrowly defined jobs, it will be challenging to get the right people together to help ensure a successful data warehouse. All employees, businesses, and systems must learn to be more flexible and collaborate. Each person involved with the data warehouse must feel that he or she can challenge the status quo in order to make the solution more useful and valuable.

### ***Agile, Inc.***

The second business, called Agile, Inc., is a small entrepreneurial company. It has been in business for several decades, so it has well-established application systems to support it. The company is not highly structured and each group is encouraged to be creative and get things done. This empowerment is welcome and can be a powerful tool for the data warehouse efforts, but there are also risks associated with this environment.

For example, people are not used to coordinating across groups. It is more of an “every man for himself” environment. Commonly, if someone else is taking too long, an employee will not wait but simply forge ahead and develop his or her own solution. This has resulted in a proliferation of many point solutions that address single business issues. Data is repeated all over the place. The result of compiling an inventory of all these different systems indicates just how bad the problem is. There is a strong need to get a single integrated view of the entire business. Due to market pressures, the organization must leverage technology to get more productivity from the existing staff.

It is necessary to harness all this individual energy to coordinate a common solution that benefits everyone. This requires collaboration and compromise, and it will take some time to get a foundation in place. Patience is difficult. The senior executives believe that building an enterprise data warehouse environment will help the business in the long run. They also understand that this commitment to a strategic solution requires dedication and persistence.

Each chapter will conclude with observations about how these two organizations are dealing with the challenges presented.

## **Summary**

---

Before jumping into any new data warehouse projects, or even if you are in the middle of one, it is helpful to step back and look at the big picture of the organization. Taking some time to consider the strategic direction of the business can help define where data warehouse efforts can be best leveraged. It is also important to have a realistic understanding of the different reporting systems that are currently being used.

This chapter provided suggestions for how to compile this background information and explained how this foundation can guide what should be done next as you get ready to start a new data warehouse design and development project.

Sometimes the next steps should be focused on leveraging what has already been built. The book follows the path of starting a new data warehouse project, and the next chapter begins by exploring the partnership between business and IT groups, including a definition of the roles and responsibilities of data warehouse project participants.





# Successful IT–Business Partnerships

Cumulative experience over the past twenty plus years of implementing data warehouses, as well as talking to hundreds of corporations and consulting organizations about key factors that drove successful data warehouse implementations, have consistently shown that one of the most important factors for a successful data warehouse is a strong partnership between the business and systems communities. Other texts and articles typically focus on what IT must do to improve that partnership. This chapter focuses on what the business community can do to build and maintain a strong relationship with IT. Chapter highlights include the following:

- What a partnership really means
- Roles and responsibilities for both IT and the business
- Working with external consultant organizations
- Ideas to improve communication
- Tips for developing strong project teams
- Review of ongoing responsibilities

## What a Partnership Really Means

---

Wikipedia offers the following definition of partnership:

*A partnership is a type of business entity in which partners (owners) share with each other the profits or losses of the business undertaking in which all have invested.*

This definition relates well to the type of partnership that is necessary to build and sustain a successful data warehouse. Both the business and IT community must take ownership of the data warehouse and be able to share

equally in the success or failure of the process. The strength of any partnership develops over time, as you work together toward a common goal.

Partnerships are strong when each individual in the relationship invests equally. Everyone must pull his or her own weight. The strongest partnerships are developed when teams or individuals are dependent on each other in order to be successful. This is clear when looking at a sports team. While success may result from the performance of one superstar, a team is more likely to achieve success when it contains multiple strong contributors. In the 1980s, Michael Jordan alone was not enough to ensure success for the Chicago Bulls. However, in the 1990s when he was surrounded by other strong partners, the Bulls were unstoppable.

How a partnership works for a data warehouse may not be as clear as the relationship between players on a sports team. For any partnership to succeed, all partners must have a clear understanding of what is expected of them, take ownership of those expectations, and follow through by fulfilling their own responsibilities.

Organizations that are committed to building a strong partnership between the business and IT communities typically fall somewhere on a continuum. At the low end of the continuum, organizations designate people for both groups and encourage them to learn, understand, and partner with each other. At the midrange of the continuum, organizations create performance goals and incentives based on the ability of the groups to build strong partnerships. Finally, on the high end of the continuum, organizations build competency centers or centers of excellence where both the business and IT partners are reassigned into this new organization with integrated goals and a single management team. No matter where your organization falls on the continuum, all parties must be committed to building a true partnership. If this commitment does not exist, then your chances of a successful project decrease dramatically.

Many organizations have well-developed, strong relationships between business and systems. In order to capitalize on that experience, let's take a look at how you will contribute to the partnership through your role.

## **What the Business Partners Should Expect to Do**

---

Before delving into specific roles and responsibilities, it is worthwhile to take a step back to reflect on how the various parts of the business are often perceived. Business professionals can be intimidating to systems people. Salespeople are outgoing and gregarious. Marketing people are creative and trendy. Folks from finance are detail oriented and demand that everything balance to the penny. Engineers are methodical. Actuaries are incredibly intelligent and detail-aware.

The higher up the management chain, the more intimidating. Ultimately, senior executives can be very intimidating. Executives absorb information quickly and form opinions quickly too. They have the ability to immediately hone in on any areas where there is uncertainty and doubt. Good executives also have the ability to listen with an open mind, and are willing to modify their strategy based upon new information. Therefore, if you are a member of the business community, be mindful of how you might be perceived and act as a partner. If you are a member of the IT community, remember these are your partners and act accordingly.

Why take the time to reflect on this? Too often we fall into behavior patterns of dealing with our peers or others within our own area of the business. As we interact with those outside of our realm, we each need to take our audience into consideration. This includes limiting the use of acronyms and jargon in our discussions. Systems people are often criticized for excessive use of jargon. Business people are just as guilty. This can be as simple as using industry terminology such as market share, margin calls, or loss ratios. It may be using abbreviations for projects and initiatives that are underway within your organization. If these initiatives do not involve major systems development, the technical staff may not know anything about them. Both business and IT have to understand the success factors of the project and develop a common level of understanding to achieve them.

Yes, that does mean the systems staff should learn the business basics, including terminology and fundamental functions or processes. If the technical project team members are struggling with understanding the business, you, their business partner, can help! Recommend internal education classes for them to attend. Often, introductory seminars are offered to new business employees. Sometimes these are offered as online self-paced studies. Arrange for the technical team members to participate. Compile a list of good books that lay out the principles of your industry. Consider the textbooks that you used in college if the information is still current. Encourage the systems staff to purchase these books or loan them out if your department already has them. Recommend periodicals that are targeted to each functional area, and don't overlook online sources of information such as articles and blogs.

It is hoped that the systems people you are working with will be open to your recommendations. The goal is not to turn the technical staff into full-fledged brand managers, brokers, or accountants, but to give them a good background about what you do. This process will go a long way in establishing a common understanding between you and the IT partners as you seek to solve the business problem.

There are several different partnership roles that need to be filled by representatives from the business community. While each has a different level of responsibility and time commitment, each is important to the overall success

of the project. The following sections describe the partnership roles that each level within the business community is responsible for filling.

## **Business Executives and Senior Management**

Every data warehouse project needs to have executive-level buy-in and support. At least one person should serve as the executive business sponsor. This does not mean that all other executives and senior managers do not have to play a role. The entire senior staff should have a basic understanding of data warehousing principles, as presented in Chapter 1. These executives need to grasp the organization's strategic vision for data warehousing, and some of them will play key roles in shaping that strategic vision and presenting it to the senior staff in a clear, concise manner. Development of this vision is primarily a systems function and is discussed in Chapter 9.

Implementing a successful data warehouse strategy does not happen with one monolithic project but rather through a series of individual projects. There must be executive support for both each of these projects and the overall vision. Once the first data warehouse project is complete, it is important for senior management to have continued involvement in order to ensure long-term success. This includes encouraging their staff to use the data warehouse for business analysis, and using reports or a dashboard themselves. As part of supporting the overall data warehouse strategy, senior staff members need to set long-term priorities and strategic direction for the data warehouse environment.

## **The Executive Business Sponsor**

This is one of the most critical roles for ensuring a successful data warehouse project. Without executive support, many projects chew up a lot of time and money but never deliver anything of value to the company. The executive business sponsor must be someone who sees the value in managing data as a corporate asset and wants this opportunity to guide the process. An effective partnership between business and IT must start with the executive business sponsor. You cannot simply assign a title to someone and consider the role to be filled. The most effective executive business sponsors take time to learn about data warehousing and understand the value it can have to the organization.

For example, it might mean taking the time to understand how the data warehouse can be used to streamline vendor contract negotiations using historical performance and projected demand. Taking this further, it might entail understanding how the data warehouse can help track the timeliness and quality of vendor deliveries. An effective executive business sponsor truly believes in the data warehouse and has a vision for how it can be used to help the organization.

Unfortunately, the executive business sponsor role is often viewed only as the person who can authorize project funding. The sponsor's responsibilities stretch far beyond funding a project; they continue throughout the entire life of the project. One of the sponsor's major contributions to the project is political clout. If the project is important to senior staff, then others will take notice.

This support needs to be highly visible. The sponsor should set the stage and share his or her expectations and commitment to the project at the project kick-off meeting. It is important to have critical project communications sent through the office of this person, and have the help of his or her administrative assistant to schedule meetings and set up design reviews. It never ceases to amaze me how a project team can struggle to get on someone's calendar for weeks, but when the request comes from the executive sponsor's office, a meeting can be arranged this week!

The project team members all work hard to keep the project moving forward in a timely manner, and often this happens without direct intervention from the sponsor. This level of interaction between the individual team members and the sponsor is often minimal. Most of the interaction is through the *business champion*, whose role is described later in this chapter. The partnership is developed through efficient communication and regular briefings. When things are running smoothly, regular briefings are all that is required to keep the sponsor informed. When problems arise, the team must be able to rely on the sponsor's help. The most common tasks requiring help from the executive business sponsor include the following:

- Gaining the cooperation of others within the business community
- Ensuring that the data warehouse project priority is set high enough that business team members are not assigned to other projects and have no time left for this project
- Preventing the team from losing project focus
- Running interference in the corporate culture to keep the team from getting bogged down in organizational politics
- Garnering appropriate business resources and ensuring that they devote the needed attention to the project

If the executive business sponsor stays informed, these issues will not be a surprise. As members of senior staff, executive business sponsors are well informed of the organization's current priorities, which enables them to help fight for the data warehouse. If something else has taken a higher priority, then the sponsor can explain this to the project team and assist in adjusting expectations across the organization. The sponsor can also assist in the development of a new schedule for the project if necessary.

Other staff members should be cultivated to share the executive business sponsor responsibilities. Identify individuals in the organization who are

interested or have the skills to take over sponsorship in the future. If you are the business sponsor, begin to groom your own replacement. That way, you can move onward and upward when the opportunity arises without jeopardizing all you have worked for.

#### **EXECUTIVE BUSINESS SPONSOR JOB DESCRIPTION**

The executive business sponsor will:

- Serve as the primary contact for other senior staff members to understand data warehousing and its strategic value to the organization
- Understand how the data warehouse will support business decision-making
- Review summarized project deliverables
- Assist with the resolution of major issues
- Ensure that the appropriate business staff members are allocated to the project
- Remove major roadblocks on behalf of the project team
- Have a broad perspective of functions and groups across the enterprise
- Be authorized to make decisions that resolve issues and set direction

The executive business sponsor is not be expected to attend weekly project status meetings or review detailed project plans.

Desired characteristics of the executive business sponsor:

- Be well respected in the organization
- Be a creative problem solver
- Have a track record of success
- Be well-connected politically
- Work in the business area that the data warehouse will support
- Be a visionary

**NOTE** It is common for senior managers to ask external consultants for their opinion about how a project is progressing or how the organization is doing with data warehousing. Business executives routinely ask external consultants, “Is there anything else you need? Are my people doing what they are supposed to be doing?” While it is interesting to get an independent perspective, why not do the same with internal staff? Walk through the area where the project team sits to see what is going on. Just have casual conversations about what is on their mind—ask them what else they need to keep the project moving.

## Business Managers

Managers responsible for the business groups who will be using the data warehouse also have a role to play in the project. Business managers need to have a clear understanding of the data-driven business requirement that their group is facing. These managers need to get involved at the beginning and stay involved throughout. This involvement is typically not very time consuming, but it does require learning about the project and staying informed. Managers need to know what a data warehouse is, be familiar with the overall vision for the data warehouse, and understand what the goals are for individual data warehouse projects.

Some of the staff members are likely to have larger roles in the project. At a minimum, this group will be the recipient of the results of the project. The business manager's staff will provide broad requirements, and help define detailed formulas and rules for handling business data. This means that the business manager needs to know what data will be available and what analyses will be supported. One would expect to have this level of understanding of any other departmental project on which people are working.

While senior staff sets the strategic direction and helps to get a data warehouse project launched, the middle managers play a critical role in keeping things moving. These are the people who allocate resources for all of the work that needs to get done. The staff members who are representing the business community on the project cannot be overloaded with other responsibilities. Check with the staff to see if they feel that they have sufficient time to devote to the data warehouse project. Review all of the things that they have been asked to do, as many small requests can add up to more work than any single person can complete.

In addition, take the time to ask the project manager or IT managers if the partners assigned are doing what is expected for the data warehouse. What else could the group do to help the project? Others may be hesitant to “get someone in trouble” by expressing concerns about their availability or participation on a project. By proactively asking, you are likely to get honest feedback. Rarely is there intentional dereliction of duty, but simply too much work, too little time, and no clear understanding of how things should be prioritized. Consider strengthening the partnership focus by making the goals of the project part of the individual's performance goals. This level of interest and focus reinforces the messages from senior management about the importance of the data warehouse.

**NOTE** Lack of support from middle managers can reduce the effectiveness of IT-business partnerships. Senior management may support the data warehouse, and individual analysts often see the benefit, but time is not allocated by their middle managers to work on the project. This gap in middle-management support can cause all other partnership efforts to break down.



## The Business Champion

Similar to the role of the executive business sponsor, there should be one representative from the business community to serve as the business champion for each individual project. This person will spearhead and maintain business involvement on the project and work as an active partner with the IT organization. The business champion is typically a senior member of the group and may also have management responsibilities. The data warehouse project must be one of the primary responsibilities of the business champion. While there must be one person designated as the primary business champion, other people in the business community can help with this work too. However, if multiple people are assigned the responsibility, it is too easy to just assume that someone else is taking care of things.

The business champion should have direct and frequent access to the executive business sponsor, either as part of his or her current job responsibilities or having this avenue opened up for the champion role. The champion usually participates in, or could be one of the primary drivers for, getting the project started. This includes setting up the project charter and project scope (as described in Chapter 5). The champion has the responsibility to make decisions about project details on behalf of the business community, but is not expected to be responsible for performing detailed project tasks or producing project deliverables. The champion works closely with the project manager and other business and systems resources to keep the project on track. He or she must be one of the data warehouse's primary advocates, and is responsible for giving regular briefings to a variety of audiences across the business community.

### **BUSINESS CHAMPION JOB DESCRIPTION**

**The business champion will:**

- **Advocate for business process changes**
- **Understand the overall project life cycle**
- **Monitor project progress**
- **Help resolve issues**
- **Communicate and market the purpose and benefits of the project**
- **Cultivate relationships across business groups**
- **Understand the project's purpose and the team roles**
- **Work to develop the business justification and impact of the data warehouse**

*(continued)*



**BUSINESS CHAMPION JOB DESCRIPTION (continued)****Desired characteristics of the business champion:**

- Well respected in the business community
- Strong project management exposure
- Motivated and resourceful
- Excellent communication and presentation skills
- Knowledge of the business functions and processes
- Ability to learn complex technical concepts quickly and apply them to business situations

## Business Analysts

Business analysts are typically those who perform most reporting and analysis for the group. These people understand the intricacies and challenges of the current data. Many times, the business analyst function has been boiled down to data gathering and report generating. The length of time required to collect, validate, and format the data leaves little time to perform other analysis. Too often, highly skilled professionals are not able to leverage their true talents to perform more extensive analysis that would enable them to better understand what is causing changes to the business and make business decisions to take corrective action.

Without a well-designed data warehouse, it can take a long time to pull together basic reports to track business performance measures showing longitudinal trends in sales or profit margins. Developing these reports at a regional level may consume all of the time available to meet the deadline. This would not allow time to pull data to look at sales by individual markets or stores. To take this further, it would be interesting to be able to see other geographic breakdowns or perhaps by-product characteristics. More in-depth analysis would be required to study the impact that changes to the company's pricing or promotion strategies have had on sales. The data warehouse project needs the type of person who is interested in looking at these deeper levels of analysis and is not satisfied with producing basic reports.

The data warehouse project team needs at least one business analyst to work as a full participant on the project. This person must be knowledgeable about the business and the data used to support the business. He or she will be an advocate for the group. While the project needs one person to be assigned, several may be allocated, especially if each represents a different group.

Over time, some business analyst roles have evolved to be full-time data gatherers and report creators. When this happens, these business analysts are often one step removed from the business itself and are not involved in any decision making. This can be confirmed by looking closely at their skill sets and daily tasks. This type of business analyst often has strong technical skills and may actually perform work that is much more like an IT application developer than a business analyst.

The project risk is that too often the business participation roles for a data warehouse project are assigned to these people. Their insight is invaluable for finding data, assessing data quality, and defining calculations, but their views are limited to what they are currently asked to do. Over time, the business community learns what types of requests can be fulfilled by tomorrow morning and what cannot be done in time to support a key decision. As this experience builds up, more complex questions are no longer asked, leaving only what can actually be done as the entire set of what the business analysts are asked to do. In addition, these types of business analysts are not privy to what is done with their results—they deliver them to the right person and their responsibilities are complete.

**NOTE** Do not abdicate true business participation to data gatherers.

Limiting IT's access to the business community to these business analysts limits the breadth of requirements that are gathered. Duplicating what is done today will limit the value that is realized from the data warehouse.

#### **BUSINESS ANALYST JOB DESCRIPTION**

**The business analyst will:**

- **Use the data warehouse for daily work**
- **Translate business strategies and objectives into analyses to support those objectives**
- **Research the business meaning and use of data elements**
- **Represent business preferences and priorities and balance that with technical realities**
- **Understand basic principles of the industry and how business performance is measured and monitored**
- **Identify business process changes that will harness the data warehouse's capabilities**
- **Be a business subject matter expert**

*(continued)*

**BUSINESS ANALYST JOB DESCRIPTION (*continued*)**

- Understand who else in the business community can answer questions (i.e., he or she doesn't need to be able to answer everything, but knows who can)
- Understand consequences of technical decisions on the business, or be willing to ask for clarification to gain an understanding
- Assist with documenting and gaining approval for the calculations used for business measures
- Identify business-related issues that arise on the project and drive these to resolution
- Assist with gathering and validation of requirements
- Participate in development of and reviews of the data model
- Assist with designing report layouts and user interfaces for accessing reports/analyses

**Desired characteristics:**

- Strong analytical and reasoning skills
- Good understanding of technology, but does not need to be a programmer
- Good communication skills
- Able to educate others
  - IT: About the business and uses of the data
  - Business: About data warehousing, how the data looks, and what applications/reports are needed
- Can get to the heart of the matter, and is not rigidly tied to the status quo
- Knows the ins and outs of the data—what it really means
- Detailed and patient
- Enthusiastic
- Interested in figuring out better ways to get things done

***Helping the Business Analyst Deal with Change***

It's necessary to help the business analysts adapt to the new data warehouse environment. Initially, there may be a fear that their jobs may be eliminated when everyone begins to pull their own data. However, that rarely happens. More often, the scope and responsibilities of these more technical business analysts change. Rather than being data gatherers, the business analyst can truly perform analyses. When high-quality data is easily available, there is time to pursue the next layers of questions. Rather than stop with basic information about the business, there is now time to dig in and understand what is causing

fluctuations in the business. Many business analysts are thrilled to have their time freed up to pursue these more sophisticated analyses, although this is not always the case.

Over the years, the people hired to fill the business analyst role may not actually have the business and analytical skills needed to move into this new role. They may feel that the data warehouse will expose this lack of skill. Sometimes, it has simply been too long since those skills were used, and they have become rusty or forgotten. The fears and concerns of these people need to be addressed. Perhaps some people will switch to the business intelligence developer role described later in this chapter. Others may need a refresher course in business analysis.

If business analysts feel threatened by the data warehouse, they may become a significant roadblock. This risk can be mitigated by keeping your eyes open to any resistance and by reassuring these staff members that their contributions now and into the future are valued. While their input is critical to understanding how reporting and analysis is done today, these people cannot take the place of true business decision makers on the project.

## **Business User Audience**

The target audience for a data warehouse project should be defined up front. One or more specific business groups or functions should be targeted as the data warehouse audience for each project. This group is also called the *users*. This does not mean that every targeted user will be expected to create new analyses or create reports from scratch, but that the results from an analysis will be provided to support their business function. For example, a market research analyst may use the data warehouse as the source for data to feed a complex statistical analysis. That same data warehouse can be used to create a basic performance report that is automatically generated and distributed to every sales associate at each retail store. These automated reports can provide sales results with a comparison to sales targets. If a sales associate in a retail store sees that she needs to only sell one more business blazer to reach the platinum bonus level, you can easily predict where her efforts will be spent today. Both of these groups are users of the data warehouse, yet in vastly different ways.

A cross section of business personnel will be interviewed during the requirements gathering process. These people will need to participate in the interview, review subsequent requirements documentation, and may also participate in reviews of the data warehouse data model.

## **Project Manager**

Historically, project management for a data warehouse was a systems function. Project managers reported up through the systems department and were often

recruited from the ranks of designers and developers. This may still be the case in your organization. However, project management has emerged as a separate profession in its own right and now there is specialized training and professional certification available for project managers. These techniques are being applied for all work across many organizations, not just systems development projects. These professionals are typically part of a department or office that focuses on project management. The project management role is included here because the project manager may be from a project office and not from systems.

The project manager must be the central hub for the project. This is who makes sure that everything is where it should be, and helps others find what they are looking for. One of the most important responsibilities is ensuring effective communication. This does not mean that the project manager is the only team member who can speak, but it does mean that the project manager makes sure that the right audiences get their information at the appropriate level of detail in a timely manner. In fact, it is critical that the project manager *not* become a bottleneck for communication. The project manager is responsible for making sure that everyone is on the same page.

#### PROJECT MANAGER JOB DESCRIPTION

The project manager will:

- Develop and maintain full-scale project plans
- Direct and manage project development from beginning to end
- Develop and execute an ongoing communication plan
- Identify and manage project dependencies and critical path items
- Administer change control
- Track progress toward project milestones
- Track issues and progress toward resolution
- Monitor and report project status
- Coordinate resources from cross-functional groups
- Ensure that all project deliverables are completed

Desired characteristics:

- Excellent interpersonal skills
- Strong analytical and reasoning skills
- Strong organizational skills
- Very detailed and rigorous about follow-through

(continued)

**PROJECT MANAGER JOB DESCRIPTION (continued)**

- Excellent communication skills
- Task oriented
- Strong problem-solving skills
- Enthusiastic

**NOTE** The project team members from the business community must be empowered to make decisions. There are many small things that come up throughout a DW project. The business people assigned to the project should have the authority to talk to anyone they need to across the organization. If a request is made for a significant block of time, this clearly needs to be done with the cooperation of their management. E-mail is indispensable, but it can take more of everyone's time to e-mail back and forth discussing what you need than if you simply walked over and asked the question. Common sense must prevail. If someone does not have the time to spend, they will tell you. Usually, even the busiest people carve out a little time to help out.

Now that the roles and responsibilities of the business community have been presented, it is time to turn our attention to the project team members from IT.

## What You Should Expect from IT

---

The input of the business community is invaluable to a project team. Likewise, a data warehouse cannot be built without the talent, technical expertise, and rigor that the systems staff offers. The data warehousing industry has developed technologies and methods to design and implement robust environments that support businesses today and into the future. You must trust the technical team members to construct components to fit together to provide you with an environment that will support your business requirements. Let the systems staff do what they do best: make technical design decisions and build the capabilities. You really should not care which hash algorithm is used to load large tables into a database as long as you can get the data you need in a timely manner. There are many different roles on a data warehouse project that are filled by systems professionals. Many roles are consistent regardless of what kind of system is being developed. The IT roles for a data warehouse project are described in the following sections. There are many other venues and

publications that describe these roles in great detail, and are not duplicated in this book.

## **CIO/IT Executive Sponsor**

The data warehouse should have one executive sponsor from the IT organization that will partner closely with the executive business sponsor. This is typically the CIO, but in large organizations this could be a senior manager in the IT organization. This person should be ultimately responsible for the technology side of the data warehouse and should have management responsibility over the IT staff on the data warehouse. This person usually shares the funding responsibility with the executive business sponsor and will own control for the technical aspects of the data warehouse environment. It is very important that the executive sponsors from the two organizations are seen as equals and that they are on the same page in terms of goals and objectives for the data warehouse.

## **Data Warehouse Manager**

The data warehouse manager's role represents the driving force behind the data warehouse environment from the IT organization. This individual is often given direct responsibility for the funds for the data warehouse project. This is the primary visionary for data warehousing in the organization, and the lead systems professional who is responsible for building and sustaining the data warehouse. The data warehouse manager continues to have responsibility for the data warehouse after a project is complete and oversees support, maintenance, and growth.

The data warehouse manager must set the overall direction of the data warehouse. This is done in conjunction with key business management. This includes making sure that a sound technical and data architecture is developed. The data warehouse manager is the focal point for blending industry best practices with internal best practices to develop an overall methodology and techniques that will ensure the long-term success of data warehousing within the organization. This includes adapting system development life cycle (SDLC) practices and the definition of standard deliverables to meet the unique data warehousing needs.

If there are many data warehousing projects underway, the data warehouse manager would then be involved in coordinating and leveraging resources across the projects. These resources could include people as well as technology.

**DESIRED CHARACTERISTICS OF THE DATA WAREHOUSE MANAGER**

- Sound analytical and reasoning skills
- Strong presentation and communication skills
- Able to educate others
  - IT: About the business and uses of the data
  - Business: About data warehousing, how the data looks, and what applications/reports are needed
- Understands the business
- Possesses project management experience
- Has exposure to the technologies involved
- Understands the organization's culture and how things get done
- Exhibits leadership
- Understands and can manage risk
- Able to manage and coordinate multiple projects and priorities

## **Business Systems Analyst**

The business systems analyst role is similar to the business analyst, but this person needs to have sound knowledge of data warehousing design principles. The business systems analyst will participate in, or even drive, the business requirements gathering process and will be a key player in building the partnership with the business community. The business systems analyst will also be expected to do the following:

- Participate in the development of the dimensional model
- Help research data-related issues
- Be a liaison between systems and the business communities for daily project tasks
- Develop and oversee testing plans
- Develop project documentation and deliverables

The business systems analyst role is often blended with the business analyst role. This is effective only if a single person can understand the business clearly enough and delve deeply enough into the details of data warehouse design and development.



**DESIRED CHARACTERISTICS OF THE BUSINESS SYSTEMS ANALYST**

- Excellent problem-solving skills
- Understanding of business processes and functions
- Strong technical background
- Able to facilitate and communicate technical issues to a variety of audiences
- Understands the system development life cycle
- Good writing skills
- Quick learner
- Persistence to work through complex issues and design challenges
- Enthusiastic, motivated, and curious about learning the business

**NOTE** The most important characteristic to look for in candidates for the position of business systems analyst for a data warehouse project is curiosity. If a person is sincerely interested in what the business does and how things work, then this will feed their ability to gather requirements and then translate them into a solution.

## Source System Analyst

The source system analyst is not a full-time member of the data warehouse project team but plays a critical role. The data warehouse team must have one source system analyst designated for each source system that will feed the data warehouse. This person needs to have experience and knowledge of the underlying systems that will be used to populate the data warehouse. The source system analyst will also be expected to do the following:

- Support the development of the dimensional model
- Provide file layouts, data models, and data element definitions that exist for the source systems
- Research data-related issues
- Serve as the subject matter expert for the source system

When there is limited documentation for source systems, the source system analyst is often the only person who knows what each data element actually means. The names and how the source systems use each field may be a mystery without the active participation of this highly valuable resource.

**DESIRED CHARACTERISTICS OF THE SOURCE SYSTEM ANALYST**

- **Strong problem-solving skills**
- **Persistence to work through complex issues and design challenges**
- **Patience to support detailed, field by field analysis**
- **Willing to put forth extra effort to help the data warehouse initiative**
- **Understands the critical nature of his or her knowledge**
- **Able to juggle multiple priorities, including communicating to management concerns about the scheduling of their own time**
- **Shows interest in learning about data warehousing**

## **Data Modeler/Data Architect**

The data architect's responsibilities are to develop a big picture strategy for how data will be handled, stored, and processed to support the requirements of a specific project and to accommodate the requirements across the enterprise. This is the heart of the data warehouse. The data architect sets this vision. This includes driving or participating in the following:

- Defining the data structures and philosophy for the staging area
- Defining the data structures and philosophy for the data warehouse
- Participating in the requirements gathering process
- Understanding specific project requirements
- Gathering or researching enterprise data requirements and standards
- Ensuring that the data warehouse architecture fits in the overall enterprise data vision

Data modeling takes things to a lower level of detail. This entails understanding individual data elements and how they relate to each other; and then applying what was learned during the requirements gathering process in order to develop data models to support the business. There may be a need for detailed data modeling to support the staging and preparation processes. There is also a need for a dimensional model to reflect the specific data needed to support the business objectives of the project. The model must also be flexible enough to be extended as new data and requirements are discovered over time. The basics of a well-designed dimensional model should be able to sustain a data warehouse for years. The data modeler will be responsible for the following:

- Designing data models and databases to meet current and future needs
- Understanding how the data model will support the business reporting and analyses
- Researching data content and availability to ensure that the model is grounded in reality
- Researching and recommending approaches to handle data integration
- Performing detailed data analysis to understand the data as it currently exists
- Developing logical database designs

**NOTE** Dimensional modeling for a traditional data modeler can prove to be very challenging. If a data modeler or architect is fighting standard data warehousing practices, take action. Education may not be enough to change their ways. If this continues, consider reassigning that individual to a traditional project and get someone who will effectively support the data warehouse. That way, you can still leverage their skills and experience without causing delays and poor design choices for the data warehouse.

#### DESIRED CHARACTERISTICS OF THE DATA MODELER/DATA ARCHITECT

- Understands the business requirements and the business point of view
- Understands the impact of data design on the rest of the project
- Understands data warehouse design best practices and approaches
- Understands how BI tools work and present information to users
- Understands the benefits and purpose of dimensional modeling and is an expert modeler
- Curious and motivated
- Able to communicate effectively with both business and technical groups

## ETL Developer(s)

One of the most complex, challenging, and time-consuming parts of a data warehouse project is preparing the data for access and analysis. The main steps are to extract, transform, and load (ETL) the data. There are usually multiple systems professionals who perform the design and development of the ETL processes to populate the data warehouse. The lead ETL developer may sit in on requirements gathering sessions. It is also beneficial for the lead ETL developer to participate in the creation of the dimensional data model. This helps to ensure a full understanding of the complete model.

ETL developers work with the rest of the project team to define rules about how to clean, validate, and integrate the data. This often involves working directly with the business community. In some organizations, this must be done by working through the business system analyst. The ETL developers spend a lot of time researching and resolving issues discovered in the data as it is prepared for the data warehousing environment.

#### **DESIRED CHARACTERISTICS OF THE ETL DEVELOPER**

- Ability to adapt when changes occur
- Understand the full data warehouse life cycle
- Appreciation and desire to work through data quality issues
- Experienced system developer
- Ability and interest to learn new techniques and approaches
- Creative problem solver

## **Business Intelligence Application Developer**

The business intelligence application developer is responsible for creating and maintaining queries, reports, and applications. This usually involves using business intelligence (BI) tools. These tools need to be configured or set up properly to access the data mart. Each BI tool has a semantic layer, or metadata, that it uses to traverse the data and manage defined metrics. Once the BI tool is configured, a wide range of activities need to be performed, including the following:

- Designing and developing a suite of report templates
- Designing and developing a deployment strategy for reports
- Designing and developing performance dashboards or scorecards
- Designing and developing BI report and application documentation and training materials
- Supporting business users to help them increase their knowledge and use of the BI tool

**NOTE** There may be two different types of BI developers. The first has deep technical skills and is responsible for installing and configuring the BI technology. They also design and deploy complex reports and analyses. The second type of BI developer is less technical and may be filled from the business community. This second role is responsible for creating reports and delivering analyses to support

the business community. This second type often directly supports requests from senior management.

#### **DESIRED CHARACTERISTICS OF THE BI APPLICATION DEVELOPER**

- Strong understanding of the business requirements
- Interested in helping the business leverage the data warehouse
- Flexible and able to adapt to ongoing changing requests for information
- Works well with the business community
- Good communication skills
- Able to translate business requests into BI report specifications
- Strong analytic abilities

## **Other Supporting Roles**

A number of other technical roles are required for successful completion of a data warehousing project. These include a variety of technical functions, including database administration, security, network administration, systems architecture, training, testing, and operations support. These people are involved in making sure that the foundation technologies are set up and working properly and that the project is deployed successfully.

The roles for the business and IT communities have been described here individually, but the real strength of a data warehouse emerges when these people all work together. Ideas for helping to build and maintain this relationship are covered next.

## **Tips for Building and Sustaining a Partnership**

Many factors shape a strong partnership between the business and IT. This section describes a few that may seem to be common sense but are often overlooked. If you're from the business community, you should strive to ensure all of the following:

- Make time available to answer questions and clarify concepts. Too often it is assumed that others understand what you tell them, but you need to be patient and explain in detail.
- Listen to systems personnel to understand what they are trying to communicate. You need to be reasonable. IT knows you want everything at your fingertips for all time in any format that you think of now or in the

future. You need to be realistic; what do you need to get your job done today?

- Be willing to make trade-offs—you can't have everything right away. Help IT decide where to start and the order to proceed from there. This includes looking at the cost versus the value of your request. It is critical to long-term success.
- Understand the long-term impact of shortcuts taken today.
- Appreciate the courage it takes to deliver “bad” news. Someone has to stand up and say that the emperor has no clothes!
- Listen to your own people; external consultants and experts are not the only people who know what needs to be done. Help the team by offering creative business solutions to issues or areas where progress has slowed. Make sure that business resources are made available when needed and cut through the red tape whenever possible.
- Proactively follow up with the team if you have not heard anything. Teams can be reluctant to schedule briefings unless a significant milestone has been met. It is still worthwhile to touch base, even if it is only to report that everyone is hard at work. Difficulties and challenges that have been discovered can be raised and discussed and any potential impact on the budget or the schedule can be discussed early.
- Take the time to thank members of the project team for their efforts.

If you are from the IT community, you should do the following:

- Talk to the business. Communication is the key to a successful partnership. You must provide open and honest communication. This includes admitting when the team does not have the skills to get the work done or when you tried something and it did not pan out. We all tend to find it easier to share positive communications, but we owe it to each other to share it all. You need to be willing to ask for help when you need it.
- Invest time in learning about the business. Having a basic understanding about the core business functions of the data warehouse audience is critical. Without the basics, it is difficult to conduct meaningful conversations between the business and IT team members.
- Be able to communicate concerns and needs in a manner that resonates with the business community. This can mean that technical requests need to be put into business terms. For example, it is not acceptable to request to purchase more disk space “because we need it.” This is not likely to be approved without further dialogue. A more useful way to phrase the

same request would be “because additional disk space is needed to load the consumer-level data required to perform market basket analysis.”

- Develop realistic project plans and schedules. If the project will realistically take eight months, then do not let the business push you into promising you can complete it in two.
- Make sure the project team stands firm when necessary. If the business community is asking for things that are not feasible, it is acceptable to say no, as long as a sound rationale can be offered.

In addition to internal resources, it is common to also engage external help for the data warehouse project. The next section describes how to best utilize these external resources.

## Leveraging External Consulting

Each organization must determine what resources are available and where external help is needed. Some organizations go so far as to completely out-source the data warehouse design, development, and support. When getting any external help for the data warehouse, there are several important considerations. A data warehouse will contain critical data and support strategic planning for the company. It is important that the organization have a clear understanding of what the data warehouse contains and how it was built. This is needed to ensure that the organization can properly maintain and sustain its use over time, especially when all the consultants leave.

The most effective way to get this level of knowledge is to have consultants work side by side with your internal resources. This enables skills to be learned by your people. A significant side benefit of working on a data warehouse project is gaining insight into the meaning and use of the organization’s data. When external consultants are the only people working on the project, they are the only ones who gain that knowledge and take it with them when they move to their next engagement. Conversely, if your staff is also involved, then that knowledge remains with the company.

**NOTE** Key IT staff members who are strong candidates for the data warehouse team are usually already tied up with many other responsibilities. When possible, hire external help to backfill some of their other responsibilities and allocate at least some of their time to the data warehouse. This provides relief in areas that the organization already knows, and provides growth opportunities for key staff members.

When consultants are finishing up an engagement, schedule an exit interview. This is a common practice when employees leave, but it is rarely done

with consultants. If a large team of consultants is involved, conduct the interview with only one or two of the more senior members. Ask what went well, and where the organization could improve. This also ensures that it is clear what has been completed, and confirms the next steps to be taken. Make sure you receive complete and accurate documentation on all aspects of their engagement. Too often, when winding up an engagement, consultants work quickly to complete their final deliverables, then dash out the door to catch a plane.

## **Building Strong Project Teams**

Data warehouse projects are not the only projects that require interaction between systems and business people. The data warehouse project team members must continue to interact with the business throughout, not just at the beginning of the project. To improve these interactions, basic team building techniques can be used. These are typically used to build relationships within a group such as bringing together the whole sales force or human resources groups. These groups have a common bond simply because they have similar jobs. The data warehouse team consists of people from many varied backgrounds and positions. Many are on the team because they care about the company and want to contribute to its success. These people all work hard and enjoy a good challenge (they won't be disappointed!).

This cross-functional team needs to work well, so get them together for an outing, a late afternoon reception, or a long lunch. Run several exercises so that everyone gets to know each other as individuals. Team members quickly find commonalities: a shared college, children of similar ages, an avid sports allegiance, and so on. When a personalized relationship exists, it is much easier to get together to ask questions and to work through issues.

After getting started, several things can be done to help the team run smoothly. Have the core team members sit in the same area. This fosters open spontaneous dialogue. Encourage everyone to talk face-to-face, not just via e-mail. The more everyone "interfaces," the more they will operate as a single team, not as individuals completing just one part of a puzzle.

### **YOU KNOW YOU HAVE A HEALTHY TEAM WHEN:**

- **They all argue with each other, but also defend attacks on any single team member.**
- **They have heated discussions that end with much better ideas generated and more optimal solutions being developed and everyone happily goes to lunch together.**

*(continued)*



**YOU KNOW YOU HAVE A HEALTHY TEAM WHEN: *(continued)***

- There are inside jokes.
- Team members or data elements are given nicknames.
- They form a sports team.

Another area that is essential to building strong partnerships is effective communication. When we fail to communicate, we leave the door open for doubt, uncertainty, and blame. When we communicate well, we open the door to trust and partnership.

## Effective Communication

---

There is often a communication gap between the business and systems communities. The systems groups give briefings and share presentations. Unfortunately, these are often not effective. Even when a great deal of effort has been put into these briefings, they can still fall short. This is often due to a lack of communication skills. Many technical people can design and develop complex and amazing systems, but are unable to discuss what they have done in plain English. It is hoped that someone in IT management has developed these skills and can help others grow these skills too. The same problems exist on the business side as well. The business people understand their world but are not always effective at communicating it to others. Without effective communication it is almost impossible to develop a strong partnership. The most critical areas to improve the effectiveness of this communication follow.

## Netting Out Key Messages

When preparing to present to any audience, get back to basics. Ask yourself the following questions about the audience:

- What is their level of exposure to data warehousing or the business? Assess their level of understanding of this project.
- How long do you have for the meeting?
- Who called the meeting? If you did not request the meeting, make sure that you know who asked for it and what they want.
- Why were you asked to participate? If you requested the meeting, then make sure you have a clear purpose.

- What do you want to accomplish? For example, a meeting's purpose may be to educate senior managers about data warehousing basics, to share project status, or to ask for more funding.

One way to help hone in on what messages to communicate in a meeting is to ask yourself what you hope to gain as a result of the meeting. Are you expecting the group to make a decision? Is this intended to be an informational session? One technique to help boil things down to the essentials is to identify three to five things that you want the audience to walk away knowing. Much more than that and you have too much content for a single session.

Each session should have a clear and fairly specific objective. You will not be able to provide an introduction to data warehouse concepts, details about dimensional modeling, and a review of the project scope in a single half-hour meeting. It would be more beneficial to set up several meetings so that a single topic can be covered thoroughly. A series of short meetings can be much more effective than trying to cram everything you ever wanted know about data warehousing and your project into one meeting. Such broad, all-encompassing sessions provide too much information to absorb and the weary audience often leaves confused, having missed the key point of the meeting. This also leaves the attendees without a desire to come to another meeting.

Getting down to the key messages is one of the most difficult things to accomplish. We are often surrounded by a sea of details. From the systems side, there are so many exciting things to share, such as using new interfaces, protocols, and development techniques. However, these are unlikely to be of much interest to the business community. The business community only wants to know about things that will impact them directly. For example, details regarding a version of the database interface that does not work properly with the current version of the data access tool and is impacting the performance of everything else hosted on the same server is really not important to the business. What matters to them is that due to technical compatibility issues, the project team needs to extend the development schedule by three weeks. Note also whether the resources needed to resolve the current problem(s) are working on it. If not, then you need to share what you need to get it resolved.

## **Presenting in Business Terms**

Another important, yet difficult aspect of communication is sharing everything in business terms. This can be particularly challenging for IT team members, especially if there is little interaction between groups. Too often, the technical team members can only see and understand the development side of things, and they don't have any idea how what they do impacts the business community. Although it is not necessary for every member on the team to be able to translate technical topics into business impact, at least one person must have that ability.

Many times there are detailed discussions about data sources and individual data elements. Due to how the current source systems work, it is possible (and unfortunately fairly common) for data that is needed to allow integration with other systems to not be captured. The content may be known at the time that the transaction was being processed, but it was not stored in the system. This is usually because that piece of data is not needed to complete that specific function, but it may be critical to support reporting and analysis later. Below is a sample business description of a data integration problem.

*The order entry system tracks every product and service in detail, to ensure that the correct products are delivered. When invoicing the clients, the products and services are grouped to support tiered pricing, and customer volume discounts are applied. Given that the order entry system is focused on delivery, it does not contain any pricing. The invoicing system does not retain the link between the individual products and services and the line items on the invoice. The products are delivered appropriately and invoices are accurately generated. Unfortunately, this missing link limits the detailed analyses that can be done. High-level analysis of the entire invoice by customer can be done, but detailed analysis of the profitability at the product level of detail cannot be easily done.*

There must be enough substance to help everyone understand the problem itself—that there is missing data. Then, there must be a description about what that means. In this case, high-level analysis can be achieved with order and invoice data, but detailed analysis cannot be supported. This could be significant if the primary objective of the project was to support product level pricing and profitability analysis. Without that link to the business capability, the business community usually blindly trusts the project team to do the right thing. Then, during deployment the impact of the missing data link is understood. This is where many problems between the business community and the systems groups arise.

Now, the business is not satisfied because their true requirement is not being met, and the systems team members are frustrated because they feel like the business agreed to move forward without the link. Looking at the nitty-gritty, both are right. Ensuring that each issue or topic is presented in business terms can help prevent this type of communication problem from happening.

## Meeting Preparation

Preparation is the one factor that you can control in order for your meetings and briefings to be successful. Just as you go through a checklist of things you need before you go camping, you should review a checklist before going into a management briefing:

- Check the technical setup in the room to ensure that the presentation can be projected.

- Confirm that the seating capacity of the room is appropriate for the number of people who will attend. Make sure that you count the formal invitees as well as any team members who will also be attending.
- Bring handouts printed in a readable font (if you expect anyone over 40 years old, use a 12-point font or pass out reading glasses).
- Bring supporting material so that many questions can be addressed during the meeting. This builds momentum.
- While getting ready for any briefing, work with other team members to brainstorm likely questions from the audience. Practice what answers or responses will be given.
- Schedule the room 30 minutes in advance of the actual meeting to allow time for setup. It is too stressful to be pacing in the hallway while a group is finishing up their meeting. This also means that you are setting up while the attendees are waiting for the meeting to begin.

Over time, these guidelines will become a natural part of planning a meeting. A specific checklist may not be needed, but the list should still be reviewed in your head prior to any meeting.

### ***Presentation Tips***

Presentations are more effective if you observe the following suggestions:

- Limit the total number of words per slide. If the font is less than 22, you have too many words. Consider simplifying the bullets or divide the content into several slides. Another alternative is to produce a supporting document that contains all the details, while the presentation captures the highlights. This enables the audience to refer to the details in the document in the future.
- Estimate an average of three minutes per slide of content. Don't count title pages, but then add five minutes to the beginning and the end for introductions and wrap-up.
- Keep the background of the presentation light. While dark slides with white lettering seem like a good idea, they tend to be too dark and can make the audience sleepy. Light backgrounds are crisp and easy to read.
- Pay attention to nonverbal audience feedback. If the audience is reading ahead, then you need to pick up the pace. Condense what you need to communicate and get to the bottom line. This may also indicate that the content is too detailed.
- Keep a running list of things that need additional action. Summarize these at the end of the meeting, including who in the room is responsible for taking that action (at least until the task is assigned to someone else).

## When to Communicate

Communication is at the heart of a partnership. It needs to be timely, accurate, and relevant to be effective. Too often, unless there is a really big topic, there is limited formal communication between the data warehouse project team and management. Project status is regularly reported throughout the project management channels, but that is not sufficient. The project status simply notes if the project is on schedule or not. It may even highlight one or two big issues, but this is usually not in enough detail to ensure that all appropriate parties understand the impact. Moreover, the data warehouse is often only one of dozens of projects reported through this channel. Unless the project is tagged “red” to indicate that the project is behind schedule, managers may only briefly glance at the status.

Each different level of the organization has a different recommended frequency of communication. Clearly, the project team (including business analysts) should have regular, probably weekly, meetings to review progress and share concerns or roadblocks. As part of business group or department meetings, the rest of the business analysts can be kept up to date. The business champion often schedules weekly meetings to be kept informed. If no big issues need resolution and the team is simply working, the business champion may only want biweekly meetings.

Initially, there may be interest from management to meet regularly to learn more about data warehousing, information management (as discussed in Chapter 8), or other related topics. Once the data warehouse is underway, the executive sponsor should be formally briefed every quarter. A special meeting should be called if any major challenges facing the data warehouse arise between these briefings.

**NOTE** Regular project briefings are always recommended. At a minimum, weekly or biweekly meetings with the business champion, and quarterly for the executive business sponsor, are suggested. This is important even when there is nothing to report—no deliverables have been completed. In addition, project teams usually tolerate a fairly high level of static and loss of momentum without wanting to bother management and ask for help. Business management may be able to clear these hurdles quickly and get the team moving forward again. If you wait until there is news, it may be a long time. Letting everyone know that the project is still alive and working is important.

Effective communication helps to build a successful data warehouse. In order to ensure longevity of the data warehouse, there is additional work that must be done. After the initial project is complete, subsequent projects can be set up to run simultaneously. The next section shares ideas to help manage these ongoing challenges.

## **Partnerships Beyond a Project**

---

Additional responsibilities are needed to sustain a data warehouse environment that will outlive a single data warehouse project. These vary in size and scope according to the size of your enterprise and the organizational politics in your company.

Good working relationships are formed when teams are fully engaged on a project. On a well-structured project, representatives from the business and technical groups meet weekly. This regular interaction builds a strong relationship.

Once that project ends, the relationship should continue. The technical team that is supporting the data warehouse must continue to work with the business groups who are using it. This encompasses typical maintenance activities such as tracking and resolving problems and collecting enhancement requirements for new reports, analyses, and additional data.

The data warehouse manager can be the driving force to consciously maintain these relationships. Periodic briefings and planning sessions should continue over time, enhancements will be packaged together and released on a regular basis, and new projects will be initiated. All of these should occur as a result of both business and systems requirements. The dialogue must continue to ensure that the data warehouse environment delivers value into the future.

## **The Decision-Making Process**

Many of the things that make a data warehouse project interesting also increase the challenges. Serving the needs of multiple groups of people is always challenging. The product development group has a different perspective than the logistics planning group. Delivering data that is useful to multiple functions of the business implies that each of these groups will want input into what the solution looks like. Each group will be concerned about when their specific requirements will be addressed. This creates a suite of issues that cannot be addressed by an individual project team or by systems management. It requires that mechanisms be put into place to ensure that decisions are made by the business community in a manner that looks at what is best for the overall organization. These challenges only increase when multiple data warehouse projects are running at the same time.

### ***Executive Steering Committee***

Often, multiple departments or divisions of a company have requirements for the data warehouse. A committee of executives is needed to help set the direction and priorities for the data warehouse. This committee could be addressing

the diverse needs of a single project or balancing the needs of multiple projects. This committee should be comprised of the executive business sponsors from each data warehouse project, augmented by representatives from other areas that have a vested interest in the data warehouse.

The committee should have a representative from each of the business groups who are currently involved with, using, or planning to use, the data warehouse. Each person who is named to be on the committee must be able to see the big picture, not just the immediate needs of his or her own group. Each member of the committee must be empowered to make decisions. This includes making compromises that may involve deferring their own solutions for others that will have a larger impact on the business as a whole.

The committee may meet more frequently when initially building the data warehouse or planning a major enhancement project. On an ongoing basis, executive steering committees typically meet less often, perhaps on a quarterly basis. In preparation, each area should submit their requests to the data warehouse manager. Each individual data warehouse project team will work with their business champion to develop any requests that need to go to the committee. These requests can be project change requests and/or proposals for new projects, or may be issues that need strategic and cross-functional input to resolve.

The specific executive business sponsor should be briefed about the requests from their own group or project. Each request should include the business benefit, the estimated effort, the risk of not moving forward, and other time constraints or windows that need to be met. The data warehouse manager will compile these to be presented at the committee meeting. There may be instances when the committee's insight is needed for enterprise direction between their regularly scheduled meetings.

The merits of each request must be considered. Every request that makes it this far is legitimate and will help the organization in some way. However, resources are limited, and everything can't be done at the same time. Based upon overall business priorities, the optimal sequence of addressing these requests must be determined. In fact, the whole list does not need to be prioritized. Just select those requests that can be met next. This does not mean the others will not be done at all, but simply acknowledges that they are not as critical as other requests at this time.

Between committee meetings, decisions can be made with the data warehouse manager and appropriate executive business sponsors. Sometimes, an enhancement can be scoped, designed, built, and tested before the committee meets again. If resources are available and all of the groups involved agree, then there is no need to wait. It is helpful for project teams to set up guidelines outlining when requests must go through the committee.

The executive steering committee is still needed to set priorities and make decisions that cross business group boundaries, even when multiple data



warehousing projects are not involved. Some enhancements may support only one group's needs. Again, this provides a mechanism to determine what will be worked on next.

#### **GUIDELINES FOR STEERING COMMITTEE REQUESTS**

- **A collection of enhancements and/or bug fixes that will exceed a threshold number of hours of effort, as set by this organization**
- **Any project or enhancement that will cost more than a threshold determined by this organization**
- **Projects that have enterprise implications**

Other topics that the executive steering committee may also be requested to assist with include the resolution of cross-functional conflicts or issues that have not been able to be resolved via other avenues. The specific details of the problem may not actually warrant the attention of this high-level person, but if the data warehouse team is not able to get the cooperation of the parties who should be able to resolve the issue, it needs to be escalated.

The committee must also approve and agree with the strategic data warehouse architecture (details in Chapter 9). The first part is to understand and approve the strategy. Next, the steering committee must help to enforce it. This means that projects that do not follow the data warehouse architecture are halted until they determine how they will comply.

### ***DW Business Support Team***

In addition to the executive steering committee, it is beneficial to have a core group of people who can provide daily input and guidance for the data warehouse. During a data warehouse project, there are regular status meetings and periodic briefings. When there is a lot of project activity, weekly meetings are appropriate for many organizations. The team may meet less often when there is not as much activity. This ensures that all of the different groups who have a vested interest in the results of the project stay informed.

When the project is complete, issues continue to come up that need to be addressed, including the following:

- Enhancements to the underlying operational system to better support the data warehouse
- Changes in the business, such as the introduction of new product lines
- Migration to new versions of software—data warehouse tools or other system tools



- Business processes are streamlined
- New, special cases appear in the data, combinations of things that simply had not arisen up to this point
- New types of analyses may be needed
- Query performance may slow down and need some attention

Even if you are not in the middle of a data warehouse project, you will still find it beneficial for a cross-functional group of business representatives to meet regularly with the data warehouse manager. This team may also include other technical data warehouse staff members. This is a forum to raise any concerns or issues that you may have.

This team is most effective if they have regularly scheduled meetings. Topics are introduced and responsibility is assigned to research and follow up, and the results from previous topics or issues are discussed. This meeting should be strictly limited to one hour, as this will encourage continued participation. This is a long-established technique to ensure ongoing growth and support for the data warehouse.

## Enterprise Considerations

The complexities and challenges of developing strong data warehouse teams are even greater for large enterprises. One approach that can help leverage skills and experience across the organization is to form an *enterprise data warehouse team*. This team can be a group that reports directly to a single manager or a group distributed throughout the organization with dotted-line responsibility to an enterprise team. Chapter 13 contains details about how such a team can be structured and funded.

## In Real Life

---

Many organizations struggle with bridging the gap between the business and systems communities. Let's take a look at how this is working for the data warehouse teams in the big and small companies we are tracking.

### A Glimpse into Giant, Co.

In order to manage the sheer number of employees within the organization, well-defined roles and responsibilities have been developed. These are quite rigid, which can create big problems if someone reaches beyond his or her area of responsibility. Employees have a very strong territorial attitude toward their work, especially in IT. There are set channels for who talks to whom and

when. The responsibility to interact on a daily basis with the data warehouse project team has been assigned to a business analyst who produces the group's standard reports today. Any of the project team's attempts to talk directly to other business people are thwarted.

Specifically within IT there is a great deal of concern about who is working on what. Because the actual lines of delineation can blur on a data warehouse project, this is causing anxiety and concern for the IT project team members. Successful people follow the rules and there is concern that stepping outside these historical boundaries will impact their individual success.

To overcome these organizational challenges both business and IT management need to stay tuned in to the data warehouse team members. Variations that will get the work done better and in a timelier manner should be encouraged.

Rather than set data warehouse roles in stone, they have been drafted in clay. As the project moves forward and the data warehouse environment evolves, they will be reviewed and revised as needed. In order to get the project team members to interact with each other and the business people involved with the project, an afternoon outing was planned, which included bowling and pizza. This provided an opportunity for everyone to mingle. The chance to get to know one another and the shared interest and responsibility to build a successful data warehouse has changed the dynamics of how the group interacts. All levels of management continue to encourage open communication and team problem solving, which had helped to develop a stronger and valued partnership between the business and IT.

## **Insight from Agile, Inc.**

There is easy and direct access into the business community for the data warehouse project team and everyone is caught up in the excitement of building a data warehouse. When getting things started, there are many volunteers to help out. However, because there is so much going on, there really is no time to spare. There are daily crises that need attention. This is constantly taking up all of the resources, and the data warehouse project never moves forward.

Instead of allowing everyone to sign up to be on the project, select individuals must be chosen to participate. Then, and this is the hard part, time must be allocated for those chosen individuals. Their time cannot be filled with the next crisis. The organization will never get out of a crisis management mode until strategic work, including the data warehouse, is completed. Then, the data warehouse can help more quickly to address these situations.

## **Summary**

---

A strong partnership between the business and IT communities is one of the best predictors of the long-term success of any data warehouse project. To accomplish this, both the business and IT communities, at all levels, must play an active role to ensure data warehouse success. This involvement begins when a project is started and must continue throughout. The key to a successful partnership is open communication, as well as joint and equal ownership of success or failure. The next chapter focuses on what needs to happen to get a project started.



# Setting Up a Successful Project

Opportunities for data warehousing can be identified using the techniques described in Chapter 3. Now that the organization has determined that a data warehouse would be beneficial, it is time to set up a DW project. This involves more than just getting permission and some funding. This chapter takes you step by step through the process of starting a successful DW project:

- Setting up a project charter
- Defining the scope of work
- Formally launching the project
- Suggestions for managing change on a DW project

Setting the project up correctly is a critical step toward ensuring success. Some organizations have very rigorous project management procedures already in place. If so, please continue to leverage those methods and techniques. If not, then the following techniques can be adopted. The goal is to clearly articulate what you are trying to accomplish.

## Defining the Project

---

The business community should be involved up front, setting overall goals and budgeting for the entire data warehouse and for specific data warehouse projects. In some organizations, this is done without assistance from IT. The business must determine what they want, and then IT must help to define what is entailed in delivering that solution. Business involvement must not stop after the project is defined. As discussed in Chapter 4, the business and systems communities need to work together.

Since the organization has progressed to this point, there is usually a good idea of where there are needs that a data warehouse could help address. A specific project may have been identified while looking at the strategic direction

of the organization, or one group may have specific requirements such as the need for integrated sales and marketing data to improve promotion effectiveness. To start a DW project, certain standard tasks need to be accomplished. The first step is usually to create a *project charter*, a high-level proposal to get resources allocated to further define the project.

## Setting Up the Project Charter

A project charter is the written proposal that is used to communicate the business case to fund the project. This is typically developed by business representatives with input from their IT counterparts. Then, the project charter is used by the organization's management to decide whether there is enough potential value to pursue the project. The objective is to get permission to move to the next step of fully fleshing out the project, which is described in the next section.

It can be dangerous for a business group to set up a charter for a DW project without any input from IT. For example, it is easy for a business group to define a project that is not feasible due to limitations in the organization's data. It is also just as dangerous for IT to set up a project charter without appropriate input from the business. This can lead to a project that moves data but does not support real business objectives. This effort to develop the project charter should be led by an individual, either business or IT, who has strong project management skills, knowledge of the business area to be supported, and, if possible, a basic understanding of data warehousing.

The project charter is a broad description of what the project sets out to accomplish. It must be written so that anyone can read and understand the content without having intimate knowledge of the data warehouse or this specific project. It is a general-purpose document that should be understandable to others in the organization. It should not provide details that are generally known by anyone working for the organization, such as a description of what the claims handling group does. A general description of the project is needed, along with the project deliverables, or what the project will produce.

Estimated effort and schedule should also be included in the project charter. Given general project experience and some data warehouse knowledge, a general timeline can be set. For example, a project to develop a comprehensive data warehouse strategy could take a couple of weeks to get the project off the ground, four weeks to gather requirements, and another month to create, document, present, and refine the strategy. This would total a ten-week effort. This general timeframe can be used to develop a rough cost estimate. Using a fully loaded personnel cost of \$100,000 per year, this ten-week project with an average of three people would be about \$60,000. This estimate is from a very high level and may change as specific project details are defined.

It is helpful to identify project costs that are an investment in setting up a good data warehouse foundation, such as hardware and software

purchases. These should be split from the specific project costs. Typically, the project-specific costs and only a portion of the greater investment are used to determine the return on investment for this specific project.

The other side of the funding issue is estimating the return on this investment, which is challenging to put your finger on. Data warehouses do not usually provide fixed cost savings or reduced headcount, but rather enable the business to do more than what is currently possible. This empowers them to make better business decisions that will increase profits or, in some cases, help keep the company solvent.

The key to understanding value is to identify what the data warehouse will do. This is what is included in the project charter objectives. For example, translate a goal of helping the organization to retain customers into a financial value. This can be done by identifying the current retention rate and the target rate. What is the average “value” or revenue per customer? Use this number to apply to every customer who is retained. For example, if the retention rate is currently 73% and the data warehouse helps move this to 76%, this is a 3% increase. If each customer’s revenue is \$150 per year and you have 13 million customers, then this would be increased revenue of \$58.5 million per year. Therefore, if you only give 10% of the credit to the data warehouse, this is still almost \$6 million of increased revenue in the first year.

Using this business case approach, organizations can usually identify five to ten goals that the data warehouse can help achieve. Identify the potential financial impact for each. Often, the potential return on the data warehouse investment is absurdly large. Justify the data warehouse on a fraction of the potential return. This type of investigation cannot be done by systems alone, but must be done by the business community.

**TIP** Build a business-IT partnership from the beginning. The overall purpose of the project must come from the business, and project logistics must come from systems.

The easiest way to understand project deliverables is to see a sample. Throughout this book, examples are included for a call center data warehouse. The project charter is the first. It can be extremely challenging to define and accurately estimate the cost and effort to design and build a complete end-to-end data warehouse. It is much more effective to split this into two separate projects. The first project can gather business requirements and create the data model for the database organized to support the business. A second project can be defined to build what was designed in the first project. Figure 5-1 shows a project charter for a first project to gather business requirements and develop the data model to support a call center. The sections of the project charter are as follows:

- Basic project information such as the project name, and the names of the executive business sponsor, business champion, IT sponsor, and data warehouse manager

## Call Center DW Requirements Project Charter

Executive Sponsor:	James Finch
Business Champion:	Debbie Jordan
DW Manager:	Laura Swanson

### Background

The organization has been experiencing an increase in customer sales, mostly through the internet sales channel. This has also increased the demand on our call center to provide follow on customer support. We need to increase the efficiency of the call center. The organization has also struggled with providing timely, reliable, consistent, information to the users for reporting and analysis purposes. The business support team has been overburdened with producing reports. Many extracts and manual processes have been created to meet these needs. There is a significant backlog of requests.

### Project Description

The Call Center Data Warehouse Requirements Project will focus on understanding the business requirements for a data warehouse solution. The project will also produce a dimension data model that will serve as the cornerstone for the data warehouse. This data model will reflect the business perspective of the data needed to produce call center performance measurements.

This should be considered a design project and will not actually build any parts of the data warehouse. However, as a result of clarifying requirements and design, the project will be able to more accurately define a subsequent project to actually build the data warehouse, business intelligence applications to support the management of the Call Centers. Without the detailed design, there are too many unknown factors to develop an accurate plan.

### Project Objectives

The project will create the design for the data warehouse. The objectives are to:

- Clarify the specific business objectives for the data warehouse.

- Develop a data model that is easily understood by the target data warehouse business users.

- Understand the data sources that may be used to populate the data warehouse.

- Identify data integration and quality issues that will need to be addressed when constructing the data warehouse.

- Develop a realistic estimate of the cost, schedule, and resources that will be needed to build and deploy the data warehouse.

### Project Assumptions

- A total of 12 requirement gathering sessions will be conducted and documented.

- A maximum of 2 data sources will be analyzed and used as the basis for developing the dimensional model.

- A single integrated dimensional model will be developed.

- Data element and business measure definitions will be developed and documented by business personnel.

**Figure 5-1** Sample project charter (continued)



IT Sponsor

- Background providing a brief explanation of how you got to the point of starting this project.
- Project description to provide a summary of what the project will do.
- Project objectives, which are the high-level goals that will be accomplished when the project is completed. These should provide insight into why this project is being undertaken.
- Project assumptions are the major factors that were used to estimate the effort and resources being requested.
- Project major deliverables details the specific outputs that will be created by this project.
- Estimated project size provides several parameters that are needed to get project funding. This includes an estimated cost, which is often a size using thresholds that are useful to the organization. This section also includes estimated benefits and project schedule.
- Success factors are the things that must be in place for this project to be completed successfully.
- Project approval is the signatures of the executive business sponsor and IT sponsor. This may also include the business champion and DW manager named in the document.

Looking at this sample project charter, you may be surprised at the length of time estimated to gather requirements and develop a data model. There are many different approaches to building a DW, some of which would affect this estimate. Before judging this to be too long or too short, it is important to understand the work that will be done during these twenty weeks. Chapter 6 provides details about what is involved for requirements gathering, while Chapter 7 describes the data modeling effort. This work is intended to define a foundation that, when built, will hold up for years to come. The details regarding what work is actually being planned should be defined in additional project documents, which are described in the next section.

Once the project charter is completed, it is presented to the person or group who approves project funding. Each organization has a different approval process. For many, approval means that this project is fully funded and is ready to get started. If this is true for your organization, then it may be prudent to complete more detailed planning prior to submitting the project charter. For many other organizations, approval of a project charter is simply the permission to move to the next step of planning, described in the next section. This approval usually means that the funds are budgeted for this project. If the project charter is not approved, then it is important to understand why. The data warehouse may not be a high enough priority at this time or

perhaps the proposal was not clear enough to convey the potential benefits. Clarify what else would be needed to resubmit the proposed project charter and try again. For those projects ready to move forward, more detailed planning is needed, and it is time to engage a project manager.

## Documenting Project Scope

Project scope articulates both what a project will include and what it will *not* include. This is a vehicle that must be used to manage the expectations of all the different people involved. The information should be written at a high level in plain, easily understood terms, and should not be too long or else no one will bother to read it. Documenting project scope is a joint effort between the business champion, the project manager, and the data warehouse manager.

Combined, these people have insight into business needs and exposure and/or experience with data warehouse design, and the project manager provides the structure. Further investigation must be done to flesh out the big picture that was laid out in the project charter. This research includes reviewing candidate data sources, data warehouse strategies, and business objectives.

Figure 5-2 shows a sample project scope. This shows the minimum information that is needed to supplement the project charter. The scope must include what is in scope, what is out of scope, a description of the scope management process to be used, and identify project risks. It may not be appropriate to publish the financial information, such as employee costs, that is included in the project charter to a wide audience. If this is the case, then the project scope can be expanded to include nonsensitive project information from the project scope. That is, different documents may have a different audience, with the charter for appropriate managers and the scope for general use.

The overall project vision is set in the project charter and the project scope provides additional parameters. Now, it is time to look in more detail at the work to be done. This is needed to develop more realistic cost and schedule estimates.

## Developing a Statement of Work

A statement of work provides more detail than the project charter and scope. It describes the high-level project tasks and identifies project team members by role. While a statement of work is required from external consulting organizations, it can also be effective to quantify the work to be done by internal resources too. This is the contract that defines the work that will be done.

## Call Center DW Requirements Project Scope

### In Scope

This project will

Collect business requirements across primary business units including sales, marketing, call center, product planning, and finance

Develop a single integrated data model to support the business intelligence needs.

Develop a comprehensive project plan to identify tasks that need to be completed to fully develop and implement a business intelligence solution.

Explore and prioritize scope of possible subsequent implementation efforts. These will be developed and evaluated jointly by business and systems management.

### Out of Scope

This project will not

Develop or implement any software or applications. This is strictly a design effort. The actual development will be a subsequent project that will utilize the results of this design effort.

Address requirements of human resources, accounting, or facilities management areas of the organization.

### Scope Management Process

The project team will track project issues to determine the potential impact on the overall project scope, price, and/or timing of the project. For those issues that require changes to the project charter, change management procedures will be used.

A Project Change Request (PCR) will be the vehicle for communicating change. The PCR must describe the change, the rationale for the change, and the effect the change will have on the project.

### Project Risks

The major factors that could impact the ability for this project to be successful include:

Lack of clear communication from the business community of their business challenges, reporting, and analytical requirements.

Inability to access the appropriate business people for interviews, reviewing project documents, and to answer questions.

Inability to get timely answers from the business community for data-related questions.

Inability to have access to the appropriate systems people who have detailed knowledge of the data sources to be modeled. This includes getting timely answers from the systems community about data-related questions.

**Figure 5-2** Sample project scope

A complete statement of work can be quite lengthy, so only excerpts are shown in Figure 5-3 to aid in understanding what can be included. This is often developed by a project manager in conjunction with the data warehouse manager. Input from the people responsible for performing the tasks is also important. The people who will do the work usually have the best idea of how long the work will take, both effort and elapsed time. This should be reviewed and understood by the business champion too.

## Call Center DW Requirements Statement of Work (Excerpts)

### Project Assumptions

<Copy from Project Charter and include any additional major assumptions that were used to define the tasks to be performed, the timeline, or the cost.>

### Core Project Team

<Each project role and the person assigned to fill that role should be included here.>

### High-Level Project Tasks

(This section should list the major blocks of work that need to be done.)

#### Project Planning and Management Tasks

The objective of this task is to provide a project plan for the Call Center Data Warehouse Design project. The following subtasks will be performed:

- Prepare a comprehensive project plan for the Call Center Data Warehouse Design project.
- Monitor progress against the project plan for the duration of the project.
- Maintain project communications.
- Provide direction to the project team members.
- Measure and evaluate progress against the Project Plan.
- Resolve deviations from the Project Plan.
- Prepare and publish status reports.
- Review and administer the Change Control Procedure.

#### User Requirements Definition Tasks

The User Requirements Definition task consists of a number of subtasks directed at obtaining a clear understanding of the business requirements and developing data requirements to support those business requirements. The following subtasks are included:

- Conduct a maximum of eight business user group interviews comprised of between three to five business users at a time. Each session is estimated to last two hours each, to:
  - Understand the business environment
  - Understand analytical needs
  - Understand information needs
- Conduct a maximum of four individual interviews estimated to last one hour each to:
  - Understand strategic direction and vision for the organization
  - Understand the business environment
  - Understand analytical needs
  - Understand the information needs
- Analyze and Summarize Findings.

**Figure 5-3** Excerpts from a Statement of Work (continued)

<p>Write up interviews and review with interviewees</p> <p>Consolidate interview results</p> <p>Review data requirements</p> <p>Develop User Requirements Definition Document</p> <p><b>Develop Dimensional Data Model Tasks</b></p> <p>The Develop Dimensional Data Model task consists of a number of subtasks directed at defining the end user's view of the data. This Dimensional Data Model is used as the foundation for both designing the logical database and configuring Business Intelligence Tool Metadata. The following subtasks are included:</p> <ul style="list-style-type: none"> <li>Analyze business data elements from two data sources.</li> <li>Identify the business dimensions and attributes.</li> <li>Determine the relationships between attributes within each dimension.</li> <li>Identify business facts needed.</li> <li>Determine dimensionality of fact columns.</li> <li>Document Dimensional Data Model.</li> </ul> <p><b>Project Completion Criteria</b></p> <p>This project will be complete when all of the following deliverables have been completed and approved by the Business Executive Sponsor and the IT sponsor.</p> <p style="padding-left: 40px;">&lt;Insert the list of deliverables listed in the project charter. Update the list if additional deliverables have been identified.&gt;</p> <p><b>Statement of Work Approval</b></p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; text-align: center;"> <hr style="border: none; border-top: 1px solid black; margin-bottom: 5px;"/>         Executive Business Sponsor       </td> <td style="width: 50%; text-align: center;"> <hr style="border: none; border-top: 1px solid black; margin-bottom: 5px;"/>         IT Sponsor       </td> </tr> </table>	<hr style="border: none; border-top: 1px solid black; margin-bottom: 5px;"/> Executive Business Sponsor	<hr style="border: none; border-top: 1px solid black; margin-bottom: 5px;"/> IT Sponsor
<hr style="border: none; border-top: 1px solid black; margin-bottom: 5px;"/> Executive Business Sponsor	<hr style="border: none; border-top: 1px solid black; margin-bottom: 5px;"/> IT Sponsor	

**Figure 5-3** (continued)

The first step in developing a statement of work is determining what tasks need to be included. The second step is estimating the effort and cost to complete this work. This must be done prior to gaining full approval for the project.

### ***How Much Will It Cost?***

Data warehousing is not cheap. It requires the investment of many resources, especially key individuals across the organization. Without leveraging the appropriate business staff, the organization may never see a return on the investment spent on hardware and software. Don't just buy things; you must invest time and energy participating in the project. The work involved is described throughout the rest of the book.

If this is your first data warehouse project, the technical foundation must be put in place. The investment in hardware and software must be made to get started. The platforms and technologies selected need to support growth and

expand as the data warehouse grows. It is worth building a strong foundation so that other data warehouse projects can leverage what has been built. If you are adding to an existing data warehouse or building a new data mart, the overall cost may be significantly less, if you can leverage the technology that is already in place. In this case, most of the investment will be centered on the data and delivery into the business community. Remember to highlight the costs that are an investment for the entire enterprise so that these can be isolated from the project-specific costs. Project-specific costs should be used to determine the return on investment for this project.

Building a data warehouse is not like ordering a laptop computer. Unfortunately, there is no central website where you can make a series of choices and see the price build as you go. While basic hardware and software costs can be estimated, however, some of the biggest costs result from the people required to do the work. The most costly part of the project is during the construction phase. Many resources are needed for extracting, transforming (or preparing), and loading the data for analysis.

Data warehouse budgets range from tens of thousands into the millions of dollars. These costs are influenced by the following:

- Size of the data
- Complexity of the business itself
- Overall size of the organization

Some small companies have extremely large volumes of data and high levels of complexity. Some large companies have a lot of data, but the analyses are straightforward.

For new data warehouse efforts, it is advisable to break the project up into smaller pieces. It can be extremely difficult to estimate the effort required to design and build a data warehouse before full requirements are gathered. After all, you don't really know what you are going to build! That is why you must set up an initial project to gather requirements (Chapter 6), define the overall data architecture (Chapter 9), and then design the solution (Chapter 7). Once the data model is designed, you know what will be built. Then it will be reasonable to estimate what it will cost to build and deploy the data warehouse. For subsequent work, it is still advisable to break things up into more manageable projects with shorter timeframes to deliver results.

The most accurate estimates of effort are created with the involvement of the people who will actually be doing the work. The data modeler should be consulted prior to estimating the time needed to develop the data model. In addition to the estimates to complete specific tasks, the overall effort must include project management and administrative overhead. The task estimates and overhead can be combined to provide a realistic estimate of the effort needed. This can be translated into costs by applying the organization's

loaded costs for personnel to the effort. The total estimated project cost is the sum of the cost for hardware, software, and personnel. This estimated cost should be compared to what was estimated in the project charter. Any major differences must be explained. The project, including the scope, statement of work, and this new, typically more accurate estimated cost, must be approved before starting the work.

**WARNING** Do not fall into the trap whereby projects don't have the time or money to do the right thing the first time. Otherwise, a second project is needed to make corrections and clean up the data warehouse. This is a very costly way to do business, one that is appallingly more common than you would think! Don't give in to short-term goals at the expense of long-term objectives.

## Project Approval

This is the second time that this project must go through an approval process. This may be done as a single step depending on how project funding and approval works within the organization. Usually, funds for the project are budgeted when the project charter is approved. Now that more details of the project are defined, the approval process may be simpler. Often, as long as the project estimated cost is within the budgeted amount, the approval process may be as simple as providing the completed statement of work to the executive and IT sponsors for their signatures. Sometimes approval of the business champion and the data warehouse manager are all that is needed.

If the project scope varies from the original project charter (business objectives and content, price, or schedule) additional steps should be taken. The project charter itself should be revised, and presented again to gain the appropriate approvals. This helps to ensure that there is a common understanding of what the project is going to do, and that the revised cost and time schedules are clear.

Although this may seem like a lot of work just to get started, it does not need to take a lot of time to put these documents together, perhaps a couple of weeks, and they will go a long way toward setting up a realistic and attainable project.

## Starting the Project

---

Now that the project is fully approved, it is time to get things really moving. This is not intended to be a text on project management or a step-by-step guide for the project team. However, it is worth mentioning what happens once the project is approved. To get things started, the project manager must work with business and IT management to do the following:



- Allocate resources, including enterprise resources as needed.
- Develop the detailed project plan that outlines the specific tasks, the effort to complete them, the timing, and the resources assigned.

The project manager typically drives this effort, but he or she needs the cooperation of managers to allocate the people, and then those people need to provide input regarding the development of the project plan. Once these initial steps are completed, it is time to formally get the project off the ground.

## Launching the Project

There are two different aspects to a project launch. First, the core project team needs to get together and ensure that they understand the project charter, the scope, and the statement of work. This project kick-off is the official beginning of the project. The individual team members need to understand the objectives in order to flesh out the project plan. Each of these core team members will provide input to the project manager about his or her tasks. The business champion and business analyst(s) who are assigned to the project are sufficient representation from the business community. The core project team may meet several times to ensure that the project plan is set.

The second project launch is more public. This is a meeting to introduce the project to the members of the business community and their managers. All of the people who will be part of the requirements-gathering process should attend this meeting. They need to learn about the purpose of the project, why it is important to their group, and what they will be expected to do. (More details about business involvement in the requirements-gathering process is included in Chapter 6.)

Let everyone know that the executive business sponsor is on the agenda. Attendance increases when others know who will be speaking. The sponsor should welcome everyone and introduce the project. This conveys the importance of the project to the senior staff. The executive business sponsor should participate even if it is only for the first five minutes. Figure 5-4 shows a sample agenda for a project launch. This single launch with the business community can save time in each requirements-gathering session.

If a formal project launch is not feasible, then an introductory letter can be sent from the executive business sponsor to provide a general description of the project, general expectations of the people invited to participate on the project, and a statement about why this is important to the business.

Now that the detailed project plan has been developed and the project has formally been launched, the next step of the project is gathering business requirements. Once started, change is likely to follow. It is worth taking a moment now to look at how change can be managed successfully.

### Call Center DW Requirements Project Launch Agenda

Welcome and Introductions	Debbie Jordan Business Champion
Project Objectives and Benefits	James Finch Executive Business Sponsor
Introduction to Data Warehousing	Laura Swanson Data Warehouse Manager
High-Level Task Overview	Jenny Larson Project Manager
Business Responsibilities	Jenny Larson
Expectations for Your Time	Jenny Larson
Getting Started	Jenny Larson
Questions & Answers	

**Figure 5-4** Sample project launch agenda

**WARNING** If business resources are frequently diverted from the data warehouse project, it is difficult to believe that the data warehouse is a priority for the business.

## Managing a Successful Project

Getting a project started on the right foot is just the beginning. Obviously, completing the project tasks well is also important. There are other project-related characteristics of successful projects that are worth reviewing.

Regardless of the project's size, there will certainly be surprises. Just about the only guarantee is that there will be change. Successful project teams handle change openly and directly. It is easier to set up mechanisms to deal with issues and change before anything comes up. Many organizations have formal procedures in place, but they are not used. In addition to regular project status meetings and published status reports, several other activities must be performed.

### Issue Tracking

Dozens of details need attention. Keeping track of these details helps ensure that nothing falls through the cracks. To begin with, each major task of the project will have details that need to be addressed before that task is

complete. For example, during requirements gathering, it can be helpful for sample reports and analyses to be provided to the project team. After each requirements-gathering session, (discussed in Chapter 6), the project team should compile a list of sample reports to collect. If these are not provided, then a reminder can be sent. This is not a showstopper for the project, but it is worth noting. A regular review of task-level issues should be a regular part of the project work.

In addition to keeping track of details that need attention, there is a set of bigger issues that also need attention. These are items that impact the project itself. These often start out as task-level details that grow in significance as they are better understood or they are difficult to resolve. These project-level issues must also be recorded as they are identified. These are the issues that can have an impact on the project resources or timeline. If progress is not being made on these project-level issues, then the decision-making processes, discussed earlier, should be used to escalate the problem.

For example, suppose the project team needs assistance coordinating with a third-party software vendor. The data needed to support the sales effectiveness calculation is only embedded in a proprietary system that was purchased directly by the sales organization, not through IT. Arrangements must be made to get the appropriate vendor resources to work with the project team to get this data. This type of coordination often requires the intervention of business managers.

Some organizations have a culture in which it is believed that if there are any problems, then the team is not doing a good job. In reality, if there are no issues on the table, then they have not been identified or they are being kept under the table. A key characteristic of successful projects is that issues are identified, tracked, and discussed. Some of these issues will result in changes to the project's scope. The change control process that is described next can be used to do this.

**WARNING** Don't create an environment in which project teams are afraid to make changes to their projects. Keep an open mind for monitoring project effort and schedule. As more is learned, revisions should be made to reflect reality. Too often, teams are penalized for making changes, and are hesitant to do so. This results in reduced quality and value of the final results.

## Using Project Change Control

Invoking change control procedures is often viewed as something to do only as a last resort, and is avoided at all costs by many project teams. This is often used as a negative measure of the performance of a project team or project manager. In reality, formal project changes are positive. Change control is a

mechanism to acknowledge what will actually be developed and delivered, which may not be what was originally planned.

Too often, change control procedures are used only when there are increased costs for a project. This is indeed the mechanism by which additional funding is approved, but this is not the only way change control can help. Formal change control procedures should also be used when there are modifications to the scope of a project, or perhaps what data is included, that do not change the cost or schedule. This is a mechanism to ensure that everyone has a common understanding of what the project will deliver. These formal project changes are a natural result of the business and IT team members learning and adjusting as the project progresses. Successful projects use change control to manage expectations and to communicate how the project will meet business objectives.

## **Discussing Change in Business Terms**

Many choices need to be made throughout the life of any data warehouse project. The biggest decisions are often made when the project charter and scope are crafted. The dozens of smaller decisions, however, can add up in surprising ways, and it is unrealistic to expect IT staff members to make most of these decisions. Clearly, technical decisions should be made by IT—after all, this is their specialty. However, decisions about what data is important, the rules to integrate data or determine business calculations, are not their responsibility. These decisions must be made by the business. In order for good decisions to be made, everyone involved must understand the question at hand and the potential ramifications.

The project team must be able to articulate the question or problem and, when possible, offer several alternative solutions. The problems need to be communicated in terms that make sense to the business community. It is critical that the ramifications of each alternative also be communicated in business terms. Several examples are provided in Table 5-1 to demonstrate how to communicate in business terms and to share the impact of a problem or question. When possible, this translation should be done prior to meeting with the business decision-maker.

From a business or management perspective, you may need to take the lead to ensure that questions are posed in business terms. Keep asking questions until the problem is clear enough that you feel comfortable making a decision. As you work with more technical team members, they will learn what you expect and, over time, be more prepared to clarify their concerns.

**Table 5-1** Posing Questions in Business Terms

ISSUE FROM IT PERSPECTIVE	BUSINESS TRANSLATION	RAMIFICATIONS
What data elements do we need for inventory?	Who knows about the international inventory spreadsheet?	The project team must have a clear understanding of the international inventory data to include it in the data warehouse. Without it, all inventory analysis will be limited to the U.S. business.
Do we need everything in the pricing table?	Where can I learn more about how tiered pricing works?	Pricing is complex and cannot be accurately included in the data warehouse without a detailed understanding of tiered pricing because it represents 70% of the business.
The sales forecast is not supported by an application supported by IT. Does this mean we don't need to worry about it?	Does anyone have any documentation or presentation about how the sales forecasting is done? If not, who can help me understand the forecasting process and what the data warehouse can do to support it?	Sales performance requires comparing actual sales to the forecast. This is one of the primary goals of the data warehouse. Without the sales forecast, this goal cannot be met.
We don't need to include data about the call center because there are no requirements.	We are supposed to ask the call center director before talking to anyone in that organization. However, she never returns our calls or e-mails. How else can we get through to the call center managers?	While the call center is not the primary audience for this project, they do use some of this data. If their requirements are not reviewed, then significant changes may be needed in the future if they want to use the data warehouse. Also, any delay in gathering those requirements will jeopardize our target to implement prior to the next business planning cycle.

*(continued)*

**Table 5-1** *(continued)*

<b>ISSUE FROM IT PERSPECTIVE</b>	<b>BUSINESS TRANSLATION</b>	<b>RAMIFICATIONS</b>
There are five different places that look like they contain loss ratio. Which one should I use?	Is there a common definition and formula for loss ratio used across the organization? If so, what is it? If not, why not? Who can provide these to the team?	A limited understanding of such a key business measure can reduce the overall value of the data warehouse. The groups supported may be reduced and the actual loss ratio calculation may not be implemented in a way that is beneficial to any group. The business must retain control over the definition of this critical calculation.
The purchasing data is too hard to understand. It will take too long to figure it out; can we postpone it to the second phase?	The IT team needs help understanding the complexities of the purchasing data. This requires time from the purchasing application system expert and the manager of the purchasing group. How can we get their cooperation in a timely manner? More time is needed to understand what, if any, impact this may have on the project budget or schedule.	Without the purchasing data, the data warehouse will only be able to support two-thirds of the objectives for the project.

## **Managing Expectations**

It can be challenging to keep expectations of the business community and IT staff aligned with the actual project goals. Documenting the project charter, project scope, and statement of work are major steps toward keeping those expectations in line. As discussed earlier, these documents should clearly lay out the project objectives, the major project tasks, and the deliverables. Once the project is running, status meetings and published status reports provide feedback about progress. The project team tracks details that need attention and issues that must be resolved. Project change control may be invoked to reflect decisions made. These activities are often done within the context of the project team. However, there are often many others in IT or the business who

are watching and waiting for project results. While a status report can provide a summary of what is happening, it may not be enough to keep expectations aligned with actual project goals.

It can be useful to publish the project scope regularly—even if nothing has changed. As individuals learn more about the data warehouse and interact with the project, they form their own ideas about what they think the project is about. The scope must be the place where these different ideas converge so that everyone has a common understanding.

Another key strategy for managing expectations is for everyone on the project team to listen to feedback. If there is a single person or group that does not seem to understand what is happening, this should be brought to the project manager's attention. This can be addressed with a formal briefing or perhaps through information discussions.

The project has been defined and launched. Techniques for ongoing management have been presented. Let's take a look at how the two companies we are tracking are doing.

## **In Real Life**

---

The steps described in this chapter are similar to steps defined in other project management practices. Project planning and management are necessary, but should not become the primary focus. These are tools to help ensure that the project is able to run smoothly and meet the stated objectives. The companies we have been following highlight the extremes of too much or too little project structure.

## **Structured Projects with Giant**

Giant Company adopted a project methodology years ago and has a strong project office staffed with certified project managers. These people work on all types of initiatives, not just systems-related projects. Every project must complete each step in the methodology without exception. The project managers do not specialize in any type of project, and the one assigned to the data warehouse project is new to this field.

Giant Company faced challenges when using the corporate project methodology for the data warehouse project. All projects across the enterprise are defined in a stepwise fashion and the business group must set up the project charter and scope. Unfortunately, for the data warehouse project, this was completed without input from any of the enterprise data warehouse team members or the people who will be assigned to the project. While the business goals are captured, there has not been a sanity check for feasibility. The project

charter and scope were handed to the project manager to define the project. A statement of work was not developed, and there was great pressure to just start working. Development of the project plan immediately identified a major problem with the project timeframe and budget; and it turns out that it will be impossible to complete this project as defined in the charter and scope.

Fortunately, the seasoned project manager was able to see how the charter and scope defined an impossible project, and called a time-out. Additional research was done to develop a statement of work to identify the actual work to be completed on the project. The project charter and scope were also revised. This new, attainable project was approved and the project was able to move forward. The project manager also pushed to get changes made to the project methodology.

## **Freedom for Creativity at Agile, Inc.**

In contrast, the smaller company, Agile, Inc., wants to avoid bureaucracy and empower everyone to make decisions. Teams can form projects quickly, execute, and then move on to the next thing. While creativity flows, each individual project stays focused on one narrow objective. This results in many individual, point solutions that do not work together. More structure is needed to create and grow a foundation that will support the organization as a whole.

The ability to move quickly also creates challenges for those individual projects. Ideas come up and work starts. The actual objectives are clear to everyone, but no one takes the time to formally document them. The objectives change frequently, depending on what is happening in any given week. It is difficult to develop a solution to address a moving target. In addition, what is needed grows each time the business and systems team members meet. As more thought goes into what is needed, more data and/or reports are identified. With this happening repeatedly, it is almost impossible to know when the project is done.

A balance must be reached so that there is some structure to ensure the long-term viability of the project deliverables. This does not mean that project methods take over and consume everyone's time and stifle creativity. The experienced project manager for Agile, Inc., has partnered with the business champion to restart and run the data warehouse project using a more structured approach. The executive and IT sponsors have also agreed to try this approach. The data warehouse project is put on hold for a couple of weeks to allow time to formally define the project: charter, scope, and statement of work. These key documents will be used throughout the life of the project to help manage change and maintain an appropriate focus. In order to maintain flexibility, the change control process is expected to be used, but will help the organization to understand the impact that each change will have on the project. This new



structured approach was viewed with skepticism, but is being tolerated due to management's support.

## **Summary**

---

Each organization has its own process for defining, approving, and launching projects. Regardless of how the steps are labeled, they share a common primary purpose. Defining the project entails the following:

- Proposing the project using the project charter. This includes describing the project, defining objectives, and estimating the project size (cost and schedule).
- Further detailing the project using the project scope to specify what is included, what is excluded, and known risk factors.
- Identifying the specific work that will be done in the statement of work. This includes developing a detailed cost estimate based upon the specified work.

Once the project is completely approved, then a detailed project plan can be developed and resources allocated. The project is kicked-off for the team members and introduced to the business participants as well. Finally, successful management includes tracking issues, utilizing change control procedures, and managing expectations. Now that the project is set up, it is time to dive in and get to work. The next chapter delves into the challenges of defining business requirements.



## Providing Business Requirements

Now that the project is set up, it is time to get started on the work. The next step of the project is to collect requirements. The intent is the same as with any other system design project: to ensure that the ultimate system serves the intended purpose. However, the requirements needed to design a data warehouse are distinctly different from those for operational systems. Operational system requirements are very specific to ensure that the business function or transaction is completed accurately. For example, a concrete set of criteria must be met to complete an online product purchase. Data warehouse requirements are often general and much more fluid. In a company, there may be a need to produce a report of the top twenty-five corporate customers and their year-to-date purchases, but tomorrow the hot topic needing attention may be identifying bottlenecks in the distribution channel. These differences create challenges for projects that use the same approach for a data warehouse project as for an operational system project.

This can lead to a communication gap between business and technical staff members when trying to gather requirements for a data warehouse. The systems community feels like the business does not know what they want because when asked, “What report do you want first?” they get a different answer each time. Likewise, there is frustration on the part of the business because requests are made, but IT does not deliver what is really needed. The systems people really want to deliver something useful. The business people really do know what they are doing, but numerous fluctuating factors create new reporting and analytical needs. The challenge for a data warehouse project is to both support the business today and be able to adapt to meet future demands too.

This chapter explains what requirements are needed for the DW project and how to effectively gather them. It will help anyone who participates in the

project by providing requirements. It also includes details for the project team. The chapter is organized to provide the following:

- An overview of the different kinds of requirements that are needed for a successful data warehouse. This section contains information that is useful for everyone involved with the data warehouse.
- Details about what these business requirements look like, including questions to help find them. This section is geared toward those who are providing the requirements, but is still important for the rest of the project team to understand.
- Techniques for effectively gathering requirements. This section is geared more specifically toward the project team members who are responsible for driving this process.
- A description of how to document the requirements. The project team can use this to help them capture what is collected. This can also help those providing requirements to understand what documentation to expect.
- A review of how to ensure that the project will be able to deliver value to the organization. This includes a process to help reset the priorities and scope of the project based upon the requirements that have been collected. This section is helpful for the project manager, business champion, data warehouse manager, and other managers supporting this project.

## **What Requirements Are Needed?**

---

Before getting too far on a data warehouse project, it is important to understand the following:

- Basics of the business itself
- Challenges facing the organization
- Strategic plans to move the organization forward

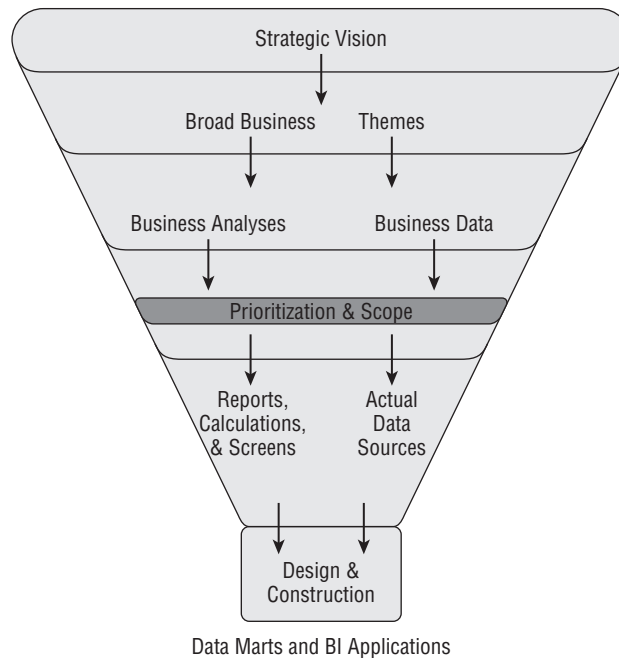
This provides the backdrop for designing an environment to support the business now and into the future. With this basic understanding of the organization, it is easy to dive directly into defining report layouts.

Having a solid background is useful to help the project team to understand what is shared during the requirements gathering process.

## **Peeling Back the Layers of Requirements Gathering**

Multiple levels of requirements are needed to build a successful data warehouse environment. These range from high-level strategic planning to detailed

data analysis. Each level represents a different type of information that is gathered. Figure 6-1 shows the different layers, each progressively more detailed.



**Figure 6-1** Layers of requirements gathering

A brief description of each of these layers follows:

- **Strategic requirements** provide insight into the vision and overall goals of the organization. Strategic requirements should focus on the big picture and look at the entire enterprise. These requirements need to be at a high level of detail. These are used to help define the charter and scope of the data warehouse project. More information can be found in the next section about strategic requirements.

The next levels of requirements focus on the areas defined by the project charter and scope. The process described in this chapter helps identify the following:

- **Broad business themes** include business goals and challenges. These are the issues and topics that the business group is working on. There may be too many of these broad business requirements to tackle at once. Select the most critical areas for further research and requirements gathering.

- **Business analyses** are often mentioned specifically during the interview process. These include analyses that are currently performed on a regular basis, samples of recent ad hoc analyses, and analyses that cannot be completed in the current environment. Business themes can be identified by looking at the underlying purpose of a business analysis. The goal of identifying business analyses is not to compile a report inventory, but rather to determine why a report is useful.
- **Business data** is mentioned in the interview to support reporting requirements and analyses. This is the type of data needed, not the specific file or column name.

The broad business requirements, business analyses, and business data are typically gathered at the same time through interviews. The interview content is analyzed to net out each of these types of requirements. The next major section of this chapter, “Providing Business Requirements,” explains these three levels in great detail.

Before jumping into more detailed requirements gathering or design work, it is important to take a moment to determine whether the results of the requirements gathering process so far are aligned with the project charter and scope. If not, this is the opportunity to set new priorities and then modify the project charter and scope as necessary. Then, with confirmation of the original project scope or a revised project scope, more detailed requirements can be gathered. These are typically done in later stages of a data warehouse project, but are included here to show a complete picture of all the requirements.

- **Specific reports, calculations, and screens** that will be part of the solution must be defined to support the business analyses specified earlier. Definition of calculations begins during the data modeling process, which is described in Chapter 7. Specific reports, analyses, and screens are defined as part of the data delivery process detailed in Chapter 11.
- **Actual data sources** are now selected to be the source for the business data that must be included, as defined by the business analyses. More detailed analysis of the data content is also needed, as discussed in Chapter 8. Business rules will need to be defined for the process that will extract, transform, and load the data into the database. Chapter 10 provides more details about developing the ETL process.

All of these requirements must be collected for the first data warehouse project. The project charter and scope will help determine the group of people needed to provide requirements input. The project scoping and prioritization step will further narrow the focus for this project. Once the initial data warehouse is built, then the process begins again. Subsequent projects do not need to start over, but can confirm whether the broad business themes, business analyses, and business data are still relevant. Any additional requirements can

be added at this time. The remaining business requirements can be prioritized again, to determine what is to be included in subsequent projects to expand the data warehouse.

This same approach can be applied to defining the requirements for an entire data warehouse program or for a single data warehouse project. The only difference would be the breadth of the audience providing the requirements. This approach helps organizations to look beyond the immediate need for a specific report or type of data.

Some organizations balk at taking the time to gather and document business requirements in this manner, preferring to jump directly to defining report layouts and screens. Without the proper business context and rationale behind the reports, however, the long-term value of the data warehouse can be limited. Focusing only on reports causes the project team to work on delivering only the data needed to create those reports. It is easy to overlook other data that is stored with the specific data needed for the reports. Often this other data could be included with minimal added effort or cost. Because the business environment changes constantly, the types of reports and analyses must also change. The requirements gathering process detailed here can help organizations to be prepared to adapt to these changes.

### ***Who Provides Input?***

The project charter and scope help to determine who should be interviewed. Representatives from the groups that are involved with or support the business goals should be included. More details about how to determine who will be interviewed are included later in this chapter. It is important to remember that while the charter and scope help determine the people to interview, they should not limit the topics of discussion. The process of gathering requirements may show that the most important area needing support by the data warehouse is not in the current project scope. The last part of this chapter provides ideas for what to do in those situations.

### ***Who Gathers the Requirements?***

The work to gather these business requirements is typically done by a business analyst, business systems analyst, and/or the data modeler. The people filling these roles form the core group that is responsible for conducting the interviews, analysis, and interpretation of the input and documentation of these findings. Because the requirements gathering process is focused on the business and does not address technology, it may be tempting to perform this task strictly as a business activity. However, because the requirements are used to drive the design, it is important to include more technical data warehouse team representatives.

The real art of data warehousing is melding together the business vision with the reality of the data you actually have. This requires the systems representatives to be part of the requirements gathering process too. This gives them the background and understanding of the business they need in order to make sound decisions during the design stage.

It does not matter who owns the budget for this work—a team of people who represent both the business and systems perspectives must complete the work. Again, partnership is key here.

Now with a basic overview of what kinds of requirements are needed, it is time get into more detail. The next section describes the first three levels of requirements and is intended to help those who are expected to provide business requirements to the data warehouse team.

## **Providing Business Requirements**

---

This section provides more details about the business requirements that need to be collected. It will help anyone who is providing input to understand what the project team is looking for. While these are divided here to aid in understanding, often the requirements are provided all together, interwoven throughout an interview session.

### **Strategic Requirements**

To be able to put together strategies and plans, interview sessions need to stay at a high conceptual level. These requirements need to highlight the biggest areas of need for data. This includes identification of data problems and concerns, such as the lack of timely reporting of data or questionable data accuracy. Any data-related concerns need to be tied to business impact. For example, without accurate sales figures, your account reviews are not effective. The tie into the business is critical. This will help highlight the areas where data warehousing can best help. After getting a good overview of the general requirements of the business, specific trouble areas can be explored. Executives and senior management typically provide strategic requirements.

Keep in mind that the purpose is to gather enough details to be able to develop a strategic plan to address the information needs of the organization. Perhaps the purpose is to gather enough information to be able to propose a data warehouse project itself. Some questions that can help identify strategic requirements include the following:

- Where do you see the organization in five years?
- What are the goals for the next year?



- How do you measure progress toward these goals?
- What will keep you from reaching those goals?
- How is information used today?
- What are your biggest concerns about data today?
- What do you need to be able to do in the future?
- What else should we be aware of?
- What does this project need to do for you to consider it a success?

#### **SAMPLE STRATEGIC REQUIREMENT**

### **Consistent Management Reporting**

**Currently, each department of the organization determines its own performance goals and tracks its performance to these goals. This is what is reported to enterprise management. Because these are all developed separately, there is no common view of the entire organization. There is a need to define ways to measure performance across the company and to ensure that these are reported to upper management consistently.**

### **Business Integration and Efficiency**

**In order to achieve peak efficiency in running the business, there needs to be great coordination and cooperation across the organization. This requires increased communication between different groups across the entire enterprise. Several business integration efforts are currently underway that should help the organization to better focus on the mission. Data is at the heart of enabling effective management of the business from the top down.**

If strategic requirements were not collected prior to a specific data warehouse project, it is worth taking the time to do so now. This strategic viewpoint can help put the rest of the requirements into the proper context.

The following types of requirements are gleaned from interview sessions with a wide range of representatives from the business community. Each session typically yields insight into broad business requirements, business analyses, and needed business data. The next major section of this chapter provides more detail about selecting who to interview and how to run the sessions. Note that the requirements are not usually presented in a linear fashion as described below. Rather, the discussion will flow back and forth.

Some specific questions are provided to ensure that each is covered sufficiently during the entire interview session.

## **Broad Business Requirements**

This is the basic bread-and-butter content that is needed as the foundation for design and development of the data warehouse. It is still too soon to begin crawling through individual reports, or tables and columns of data. A basic understanding of the business function itself is needed first. Then, explore with the interview group the challenges facing their part of the organization, how success is measured, and what problems the group is facing. This discussion should also cover what reporting and analytical needs exist. Some questions that can help identify these broad requirements include the following:

- What are your primary business objectives?
- How is your performance measured and how often?
- What are the biggest business issues facing your group today?
- How are you planning to address these issues?
- How could better information assist you?
- What is the potential impact of addressing this problem?
- What is the risk of not addressing this problem?
- What is the competitive environment like for your organization?
- How quickly is the marketplace changing?
- How is the company responding to these changes?
- What else would you like to be able to do?
- Are there any other things that the project team should know about that could have an impact on the project?

### **SAMPLE BUSINESS THEMES**

#### **Customer Service**

---

**As a provider of high-value products, superior customer service is critical to overall success of the company. The ability to identify trends in customer complaints, both the frequency and the severity, is important. Early identification of a problem would enable the organization to take action to correct the problem and avoid any additional issues.**

*(continued)*

**SAMPLE BUSINESS THEMES (continued)**

## **Business Performance Monitoring**

---

Monitoring business performance measures on a consistent/timely basis at a detailed product level is very difficult and time consuming today. This is largely due to the complexity of the industry and the breadth of products that are offered. There is a need to analyze the business at a high level and seamlessly drill to lower levels of detail to identify issues and anomalies as they are uncovered. This requires data access to be flexible enough to support many different perspectives.

Monitoring actual performance against annual goals set forth in the business plan on a monthly basis is essential. While the official performance reviews occur monthly, it is important to have access to the data more frequently. A month-to-date capability would enable the organization to identify what is needed to meet that month's goals, while there is still time to make adjustments.

## **Supplier Performance Analysis**

---

It is important to be able to monitor and measure the performance of each of the suppliers. Specific performance expectations are part of each supplier contract, so it is important to be able to see which suppliers are meeting those goals. There are quarterly supplier reviews that require a complete picture of the company's business with that supplier. This includes their on-time delivery rate, the quality of the products, responsiveness to problems, and how quickly the company pays their invoices. Because this information comes from multiple locations, this is currently a very time-consuming process. The goal would be for this information to be available more often, and perhaps even published via a web interface for each supplier.

## **Profitability Analysis**

---

The ability to understand the company's profitability is critical. Historically, this has been determined at a high level. However, there has been limited ability to drill down to see which areas of the business are contributing to the profitability. It is important to be able to understand both the areas of biggest growth and the problem areas. If the supporting detail were available, then each product line would be scrutinized and appropriate strategies developed to improve profitability.

(When working with a nonprofit or government agency this is often expressed as *fiscal responsibility*. There is still a profitability goal—it just happens to be zero. The same types of analyses and tracking need to be done.)

In addition to general business themes, data-related issues often arise. These are often the impetus behind building a data warehouse. These are important and should be noted, but are not sufficient alone. The other business challenges facing the organization are what provide the project team with what is needed to build a successful, sustainable data warehouse.

#### **SAMPLE DATA-RELATED BUSINESS THEMES**

### **Data Accuracy**

---

There are concerns about the accuracy of data that is used for decision-making. There are some issues related to how data is captured. More work is needed to specifically identify these problems and to develop recommendations to tighten controls at the point of data entry. In other cases, errors can be introduced during the processes of manually extracting and integrating data from multiple sources. Because this is often done for each analysis, there is also the possibility for variations to occur in handling the data during this manual process. As a result of the data quality concerns, the organization must spend a significant amount of time auditing and verifying the accuracy of information to be used in reports and analysis.

### **Single Source of Data**

---

A complete integrated source of data is needed to support analysis across the organization. Currently, data that needs to be integrated is often dispersed throughout the company. The different data must be located and then integrated. This is a difficult, time-consuming, and error-prone process. Development of a single integrated database could serve as the consolidation point for all divisional systems. A key to successful implementation of a single database is developing and maintaining consistent definitions of each data element. This would result in a one-time effort required to integrate and clean the data. The resulting database could then be made available to everyone. The single database would help ensure consistency of answers.

### **Flexible Data Access**

---

Flexible data access is critical to meeting the current and future demands on the company. The current storage methods and data access tools are difficult to use and require a great deal of technical expertise and knowledge about where the data is physically located. Only a limited number of people are able to access the data directly. Significant cost and lengthy turnaround times often result from this method for accessing data.

*(continued)*

**SAMPLE DATA-RELATED BUSINESS THEMES (continued)**

The company needs to have a mechanism in place to allow flexible data access to those requiring it, without the need for programming. The data access mechanism should provide an easy to use interface to run pre-defined reports, have the ability to drill down and change parameters, and easily create new reports. A user should be able to refine the parameters at will, without programming or significant system changes. This would enable more people to access the data and improve the efficiency of servicing requests. Ideally, requesters should be able to answer simple questions themselves.

## External Data Access

There is a need to provide data to people outside of the company. This includes customers and vendors. As data is made available to a wider audience, it is important to ensure that it is secure. Changes cannot be made to the data in the data warehouse. All modifications must flow through proper channels. In addition, access must be provided in a structured manner that minimizes the possibility of users pulling data in ways that are not statistically valid.

In addition to discussions about the business, general types of analyses and specific analyses are usually mentioned.

## Business Analyses

Depending on the audience, there may also be discussion about analyses that are currently performed. The key is to be able to understand why these are done and what happens with the results. The focus should always be on the final result, not the intermediate steps. Frustration and pain are often expressed about how difficult or how long it takes to get data or a report. This is a major reason why the data warehouse project is started. It is also useful to be able to tie a specific business analysis to a broad business theme discussed above. This helps to clarify the underlying purpose for performing the analysis.

It may be easier to begin a discussion about business analyses and then progress to the rationale behind it. In other cases, the business challenge itself may be mentioned first, and then analyses that are performed to address this situation can be discussed. Questions that can help uncover business analyses include the following:

- What reports do you rely on to make decisions?
- Do you do anything more with the reports to make them more useful? If so, what else is done? Is other data added?
- What analyses do you perform on a regular basis?

- Why is this important to the business?
- How does the analysis support the decision-making process?
- Do other groups receive the results of these analyses?
- What other analyses would you like to perform?
- Describe the types of requests that you have for ad hoc reports or analyses.
- How do you access data for reporting and analysis today? What recommendations do you have to improve data access?
- How long does it take to perform these analyses?
- What percentage of the time is spent gathering data? Reconciling data? Performing analysis?
- Do you have enough time to research additional questions?
- If all of the data issues were resolved and everything were at your fingertips, what would you do?

#### **SAMPLE BUSINESS ANALYSES**

The following analyses help support the business theme to improve marketing effectiveness.

### **Promotion Planning**

---

Study the past performance of other similar promotions. Develop a plan, including the budget, target audience, promotion format, and the desired response rate, increase in sales, or other result.

### **Promotion Performance Tracking**

---

Once the promotion has been launched, it is important to track how things are progressing. For a multi-week promotion, detailed, daily progress is needed to ensure that the promotion is running smoothly across the country. Actual sales need to be tracked and compared to the prior year and the pre-promotion baseline. This enables adjustments to be made as necessary during the promotion period.

*(continued)*

**SAMPLE BUSINESS ANALYSES (continued)**

## Promotion Financial Results

---

While most of the analyses are focused on the actual execution of the promotion itself, there is also a need to track the financial impact. This means that the promotion budget must be compared to actual expenditures. This is important to ensure that a promotion is effective. Some promotions may get a stellar response from the marketplace, but the cost to execute the promotion may outweigh the increased profit. If the objective of the promotion was to improve brand loyalty, it may still be successful. If the object of the promotion was to increase revenue and profit by .05%, then this may not have been a successful promotion.

## Post Promotion Analysis

---

After the promotion is completed, the full results need to be understood. This includes a high-level perspective and details by market, store, and product. This is similar to tracking the performance during a promotion, but it also needs to include a summary of the results.

**TIP** When providing requirements, consider the final result that you need. This is often several steps beyond receiving an existing report. Reports are often manipulated manually, merged together, and additional calculations performed. This ultimate result is really what is needed. If you could start here, what else would you want to do?

Also during the interview session, the discussion will touch upon a variety of different types of data—both data that is used currently and data that is needed but may not be available.

## Business Data Requirements

Again, in subsequent stages of the project, there will be plenty of time spent looking at individual data elements, specific reports, and the business rules for processing the data. At this point of the project, it is most important to get a complete picture of all of the data that would be useful to the business community. This is a time for the business representatives to share the realities of the data used regularly and to be able to share a vision about other data that

would valuable, if available. Some questions that can help identify the types of data that are needed include the following:

- What are the primary types of data that you use regularly?
- How do you get this data today?
- Are you able to access the data yourself? If not, why not? If so, how?
- What level of detail do you use? Is this sufficient?
- How much history is needed to support your analyses?
- How often does the data need to be refreshed?
- What other data would you like to be able to use?
- Is this data captured anywhere? If so, who would know more about this?
- Does this data have any special security or confidentiality requirements that must be considered?

#### **SAMPLE BUSINESS DATA REQUIREMENTS**

### **General Ledger**

---

**Core financial data is needed from the general ledger, including any manual journal entries to serve as the foundation for financial analysis.**

### **Accounts Receivable**

---

**Details about the financial accounts receivables are needed to identify and track trends in how customers make payments. This is also needed to identify characteristics of accounts that can predict late payments.**

### **Accounts Payable**

---

**Data about what actual payments are made for each claim is needed. This includes both claim payments and expense payments.**

### **Revenue Forecast**

---

**Information about the forecasted revenue for the coming year is needed. The forecast must be integrated with actual results to determine performance.**

*(continued)*



**SAMPLE BUSINESS DATA REQUIREMENTS (continued)**

Then, the underlying details can be studied to determine what is driving any variations.

## Vendor profiles

Information about vendors beyond basic contact information is needed. This includes which products they can provide, historical information about what has been purchased from them before, and details regarding reliability and quality.

**TIP** If a person can provide specific file and field names of the data while trying to identify business data that is needed, you are not talking to the right person. Try to identify who receives that data. The recipient of the data is likely to be the true business decision maker.

The bulk of the discussion should be focused on the business themes, analyses, and business data. There is also a need to understand the current technical environment in more depth.

## Systems and Technical Requirements

While most of the emphasis is on business requirements, it is also useful to touch base with the people who are currently responsible for the care and feeding of the primary source systems that are likely to be loaded initially. This is where the systems community has the opportunity to share their requirements for the data warehouse. This also provides the project team with the chance to get a realistic perspective of the data that is currently used. This insight into the source systems or existing reporting environments can help the data warehouse project team to better understand the effort needed to include this data in the data warehouse.

The information that is discovered during a systems interview should be included with all of the other requirements. Most of what is discussed will be used to flesh out the business data requirements sections. Questions that are useful to discover systems details include the following:

- What systems collect and maintain this type of data?
- What level of detail is collected?
- How much historical data is available?
- How is the data structured?

- What type of documentation exists for this data? Do the data elements have up-to-date definitions?
- How accurate is this data?
- Are there any known peculiarities with this source of data? If so, please describe them.
- How do the candidate source systems interact with each other?
- Does the data integrate easily?
- Describe the current user reporting and analysis environment.

There may also be requirements that should be included in the business themes documentation. These are themes that represent what the IT organization needs to effectively build and manage the data warehouse. These are requirements that the data warehouse project may need to address. Often, IT management identifies these needs. The same questions used for other business groups can be used to determine IT business requirements.

#### **SAMPLE IT BUSINESS REQUIREMENTS**

### **System Usage and Tracking**

---

**In order to better support the business community, it is important to be able to track the percentage of time that applications are available for internal use and for external use. It is also helpful to track who is using which systems, how often, and for how long. This helps the systems staff to ensure appropriate capacity, address performance issues, and identify applications that are not providing value. If an application is not being used, that typically indicates that there is a problem. Perhaps the application is too difficult to use, or the performance is not acceptable, or the functionality no longer meets the needs of that group. This becomes even more critical as applications are released to external users, such as customers and vendors. This information is needed to enable the systems community to develop and support valuable business applications.**

### **Education**

---

**The entire organization has already begun to learn about information management and data warehousing concepts. This education must continue and be at an appropriate level of detail for each audience. These concepts need to be understood to help integrate them into day-to-day work, set priorities,**

*(continued)*

**SAMPLE IT BUSINESS REQUIREMENTS (continued)**

**and manage expectations. Value will not be realized from the investment in data warehousing if education is not provided regarding its use.**

## **Support**

**The current problem tracking and support mechanisms need to be expanded to include the data warehouse. New help desk procedures must be developed to send data warehouse problems to the appropriate people for resolution. People with knowledge about business analysis, the data, and the technical workings of the data warehouse must be part of the team to effectively support the organization.**

So far in this section, the different types of requirements have been laid out. As mentioned earlier, requirements are rarely shared in this concise manner. Discussion can be prompted using the questions previously listed. Not a lot of work is necessary to prepare to be interviewed, but some thought ahead of time is useful. The next section shares a few ideas about how to get ready.

## **Communicating What You Really Need**

Many people are concerned about what they need to do to get ready for an interview. The primary preparation is doing your job, but there are a few other things that you could do in advance that would make the interview session more productive. Take a moment to step back to think about what you do on a regular basis:

- What reports or analyses do you create or use on a daily, weekly or quarterly basis?
- What things do you need to do for annual processes such as budgeting?
- How smoothly do these processes run?
- Do you run out of time before you can dig just a little deeper to explain or understand an unusual result?
- If you had the time, what else would you do?

As you go through your work, try to think about why you do what you do. Too often, we get so bogged down with daily challenges that we don't take the time to ask ourselves how we could do something better. It may be helpful

to jot down any ideas that you have now, so that you remember to share them during the session. Then, take this one step further and allow yourself to dream. What would a perfect world look like to help you make decisions?

Too often, the first thing that comes to mind is to have all of the data available as soon as possible. Of course, because there is a cost associated with preparing, loading, and maintaining data, this is not feasible. Therefore, consider what data would help you the most. Again, try not to focus on what is wrong with the data today, but think about what you are trying to accomplish. Then let the team track down and address problems with the data.

## **What Else Would Help the Project Team?**

As you participate in the requirements gathering process, it would be helpful for you to understand what the project team is hoping to glean from these sessions. As discussed earlier, there is a need to understand the business itself, what functions are performed, and how information is used today. The team is also looking for specific clues that will help drive the next steps of the project. The discussions will highlight the areas that are most important to the business and identify problem areas that need more research. The project team is not expecting to find all of the answers up front, simply to identify who and where to go to next.

There are also other specific things that the project team will be looking for. The team will sort through all that is shared to distill the essence of what you really need. Remember that the team represents both the business and the technical perspectives. Many clues are provided during an interview session that shed light on how the data should be organized to support business needs. These are not communicated in data modeling terms, but the details are mentioned. For example, an interviewee may state, “I need to track sales performance by employee, by product line, by individual product, by month and week, by sales division, and by customer state.”

Two different things can be gleaned from such a statement. First, the business needs to track sales performance. Further questions may determine that sales performance is measured by unit sales, dollar sales, and unit sales as a percentage of quota. The project team should recognize these as part of a dimensional model, called facts. Every word mentioned after “by” is a characteristic that is used for reporting. In this case, there is a need for employee, product, customer, and date data. Second, for products, the specific attributes of product line and product itself were mentioned. These are used to group the sales performance measures and/or to select a subset of data to look at. These are also clues to help design the dimensional model. The project team should recognize these characteristics as belonging in a dimension. Chapter 7 provides details about dimensional modeling.

## **RUNNING THE BUSINESS VS. MANAGING THE BUSINESS**

A basic concept that you must keep in mind is the difference between what is needed to *run* the business versus what is needed to *manage* the business. Decisions made to run the business are very tactical in nature and can be put into place immediately. For example, if an organization is currently placing too many calls on hold, then perhaps all customer service representatives who are on break should get back to the phones to help out.

Decisions that require historical trending and result in decisions that will take time to implement are typically used to manage the business. For example, a property and casualty insurance company must build and present a business case to a specific state's insurance board in order to change rates. When gathering requirements for a data warehouse, it is important to focus on those things that can help manage the business, rather than requirements that need to be addressed by the operational systems.

If there is currently a lack of reporting and analytical capabilities, it is harder to imagine what would be helpful to manage the business. The requirements that are communicated tend to revolve around data-related problems with running the business. These are often requirements for fixes or enhancements to the current operational application systems. These requirements should be noted and shared with the IT team that supports that operational system. Later, the discussion can return to explore what is needed to help manage the business.

### ***Data Integration Challenges***

One of the most frustrating things for the business community is trying to use data from different sources for a single report or analysis. Often these systems were not designed to work together, and do not contain a specific data element that can be directly linked between these sources. Most likely, someone has an understanding of these sources and probably is performing this integration manually. This is surely the case if this data shows up on a current report. The project team wants to learn about where integration challenges exist and also who has the most experience or knowledge about how this is accomplished today. This is discovered by asking, "Who would you go to if you needed this to be done?" This may lead to another person, and perhaps yet another. Each person in the chain generally understands his or her own piece. The project team will work with these people to understand the current business rules used to integrate the data. This is used to help define the business rules for the ETL system to prepare the data for analytic use.

### ***Assess Organizational Motivation***

The project team can also get a good idea about biases and willingness to help the data warehouse project during the interview process. Some groups

may be excited and highly motivated to help get the data warehouse built. Other groups may be more reserved and have serious doubts. In some cases, legitimate issues are raised. The sooner the team is aware of potential problems the better. In other cases, concerns may be due to a lack of understanding of the project, about data warehousing in general, or uncertainty about the project's impact on the work environment. The project team may need to take specific actions to address concerns in either case.

If a group has been starved for data, it may be necessary to spend more time educating and showing demonstrations. This may be more useful closer to delivery of the completed data warehouse. If this group is needed to help the design and development of the data warehouse, then the education must happen soon so that they can be more effective throughout the life of the project.

### ***Complete Picture of the Data***

Get a comprehensive list of the business data that is needed. This needs to include a description of the data, and if this is already being captured, then note the actual name of the system. Any primary concerns and issues with this as a data source should also be documented here. This should include data from major operational systems, information data sources, and data that is not currently being captured. The purpose is to get a broad view of all the data that would be useful. This is not a commitment that this project will address all of these sources, but it does acknowledge the need.

### **What If No One Is Asking?**

If there is a data warehouse project in process, and no one is talking to anyone in the business community, then this is a major problem. Someone from the business community must be proactive and get involved. This could be the business champion, a strong business manager, or perhaps the executive sponsor. While this may be disruptive to the current project, business involvement now will save time and money later. This may also mean the difference between success and failure.

Start by looking at the project scope and objectives. Review where the project is in the life cycle. Is it just starting, in the middle of developing the staging processes, or just about to be deployed? Other business groups may be participating, but for some reason your group was excluded. This may be due to project scope or it may simply be an oversight.

It is never too late for the business and IT groups to begin working closely together. Make an honest assessment of what you have and where you are as an organization. Even when you already have a data mart in production, it is still useful to take stock of what you have. Then, specific plans can be made to move the data warehouse forward.

## Practical Techniques for Gathering Requirements

---

This section discusses the techniques for preparing for, conducting, and documenting the findings from interview sessions. While there are other approaches to gather requirements, including facilitated sessions, prototyping, and report analysis, only the interviewing technique is covered in detail here. These other techniques can help refine requirements and provide more detail after the interviews have been conducted. However, as long as the business requirements are collected, the process to get there should not be an issue.

The information here is targeted to help those project team members who are responsible for gathering the requirements. This may be interesting to others who will provide input, but it could be skipped otherwise. If so, start reading again with the section “Putting the Pieces Together” to understand what happens next.

### Interview Session Characteristics

The requirements gathering process is focused on the business in general, so the sessions take the form of a dialogue, not technical specifications. This is not a process of reviewing screens, or a list of data elements. Because this is focused on gaining an understanding of the business itself, the easiest and most effective format is to conduct interview sessions. Most of the sessions are with small groups, but some are with individuals. The interviews are run by the lead business analyst or the dimensional modeler. A business systems analyst may also be able to run the sessions.

#### *Individual Interviews*

Based upon the seniority and availability of the person to be interviewed, it may be most appropriate to conduct an individual session. This is usually the most effective format for executives and senior management. The optimal amount of time is an hour, but most high-level people can clearly and concisely share their thoughts in a half hour.

#### *Group Interviews*

To accommodate a larger number of people, without an excessive number of sessions, other members of the business community can be interviewed in small groups. The group should be no larger than five individuals and last about two hours. A larger group makes it more difficult for each person's perspective to be heard. The group should be comprised of people with a common set of interests. If each person in the group represents a completely different business function, then each may lose interest while the others are taking their turn. This leaves the door open for viewing cell phones and other personal devices, which in turn can pull that person from the room.

Another thing to be aware of is not mixing levels of management within a single group. Make sure that everyone in a session is at the same peer level. It can be difficult for people to share openly if their boss and their boss's boss are present during the session. The second-level managers do not have a problem because their most critical report shows up every Monday morning in their e-mail. However, it may take the business analyst all weekend to pull the data and prepare that report for delivery.

**NOTE** There are often immediate benefits to having a group of key individuals together for this type of session. For example, one person may share a frustration about the inability to get to a specific set of data, and another person will speak up that such information is immediately available. Usually, there are still challenges that need to be addressed, but sometimes an immediate need can be met, even if it is in a manual fashion.

### ***Project Team Participation***

There may be multiple representatives from the project team in attendance. The more who can participate in the process, the better. However, be sensitive to the total number of people in the session; try not to overwhelm the interviewees with too many team members. If the project team is large, consider taking turns. This enables each person on the project team to have the opportunity to learn how the sessions are conducted and hear firsthand what is being shared. Observing other skilled interviewers is one of the best ways to develop interviewing skills. It is also a good idea to let the interviewees know in advance the number of team members who will be in the session, especially for the individual interviews.

**TIP** One person should be designated as the lead interviewer. One person should be designated to take notes. All others are silent observers. This enables the session leader to follow a complete train of thought without interruption.

Another benefit to having other team members sit in on the sessions is the opportunity to meet these critical business participants. Someone may have worked for years in IT without having met these important business representatives.

### ***Interview Tips***

Other logistical details can improve the effectiveness of the interview process, including the following:

- Send out invitations, perhaps even from the executive business sponsor.
- Draft a list of potential questions to initiate discussion.



- Make sure that you know where the interview will take place, and be early.
- Schedule the sessions at the interviewee's location.
- Don't schedule too many interviews in a single day. Limit the interviews to two or three a day. Too many sessions begin to blur together—what was asked and who shared what.
- Take the time to review your notes immediately following the interview. Number and label each page. Fill in blanks and finish writing out abbreviated notes and ideas that you rushed to get down during the meeting. Do this immediately; otherwise, you won't remember after the next session.
- The project team members should regroup after each session. Share the key points that were made. Did these reinforce what was expected or were there any surprises? Discuss what could be improved for running subsequent sessions.
- Be sure to leave time between sessions to allow for coffee, lunch, and traveling to the next session.
- Make a list of any follow-up that is required from this meeting, such as questions that need to be answered or ideas to better reach these individuals. This may be to forward the project scope or to schedule a briefing to share the enterprise data warehouse strategy.
- Send a thank-you note after the session to each interviewee.

With the overall characteristics of the interviews as background, it is time to look at who should be interviewed.

## Who Needs to Be Included?

Consider carefully who should be included in the requirements gathering process. It needs to be a cross section of the business community. Strong representation from the primary target audience is critical. One also needs to spend time with the other groups that interact and interface with the core audience, in addition to spending time with the representatives from the next areas that will be included in the data warehouse environment.

The project team can usually come up with an initial list of candidates to be interviewed. This must be reviewed with the data warehouse manager and the business champion. Once the list is refined, the executive business sponsor should give final approval. Sometimes critical people who should be included are considered unapproachable by the project team. Typically, these are higher-level executives to whom the project team does not generally have

access. This is where the sponsor must ensure that all of the appropriate people are included.

**TIP** Make this “invitation” to be included in the process sound like a privilege. When people see that it is not a waste of time, others may wish to be included too.

Look for those who can truly provide useful content. Often this is from people who have been around for a long time, but it is also worthwhile to meet with people who are relatively new as well. New people understand the pain of lacking the secret handshakes needed to get reports and data. These people also have the benefit of other experiences and may have ideas about what could be done, while people who have been around for a long time may not have been exposed to these other ideas.

It is also important to look for people who need to be included for political reasons. If you are interviewing three of four vice presidents, then you should probably touch base with the fourth vice president just to make sure all the perspectives are covered and no feelings are bruised.

Be sure to adjust as you go. During the interview process, additional people may be identified who should be interviewed also. Take the time to include them in the process. When requirements session are done well, people who have been resistant in the past suddenly clamor to be included. They want to make sure that their perspective is accurately represented.

**TIP** In general, 8–12 interview sessions tend to be sufficient. This is a mix of individual and group sessions. More sessions results in repeated content and extends the time required for documentation. Fewer sessions may not provide a broad enough base to ensure flexibility for future growth.

### ***Setting a Good Example***

As a business manager or sponsor, your attitude toward requirements gathering has a big impact on other participants and the attitude of the entire organization. If you are willing to be interviewed (taking your “immunization shot” first to prove that it really does not hurt!), then others will see that it is important to you and, it is hoped, how it will be beneficial to the organization.

**NOTE** This is the first time that business in general has the opportunity to get involved. This is important to ensure that what is built meets the needs of the organization. This is the first step to building a relationship between team members and key individuals from the business community. This will be helpful when further information is needed in the future. The project team will know who the players are (and vice versa) and will have a face to go with a name or position.

## Preparing for Interview Sessions

Background information should be provided to the IT team whether it is an internal team or external consultants. It is easy to justify spending time providing background information when you bring in an external consultant. This is required to bring the consultant up to speed. You should not assume that long-time employees have a good understanding of the business.

Because many organizations have already made some attempt to build a data warehouse, employees have probably already been asked for their requirements, multiple times. The people who need to provide input play critical roles in the organization and their time is at a premium. The project team can get ready by doing the following:

- Read the annual report to understand strategic objectives.
- Read everything you can about any prior data warehouse attempts and other related efforts.
- Arrange for the interview team to get a solid background on the business. The business analyst on the team can often provide this.
- Make sure you know who you are interviewing, what their role is, and how long they have been in that position.
- Identify interviews where the audience has a negative attitude toward data warehousing. Be ready to listen to their concerns and share how this project will overcome prior challenges. Be ready to solicit their ideas too.
- Conduct a project launch or kick-off session with everyone who will be interviewed. This was discussed in Chapter 5.

A well-prepared interview team can ensure that the sessions run smoothly.

## Conducting the Interview Sessions

The time has come to talk to the business representatives. The team is prepared. Everyone who was invited to the session has arrived. Begin with a round of introductions. All members of the meeting should share their name, title, and what role they assume on the project or in the business. Be sure to share what roles (lead interviewer, note taker, and/or observer) each of the team members is filling for this session. Then, have each interviewee share an overview of his or her job responsibilities. Once the basics are covered, you can get started with the heart of the session. Start with broad, open-ended questions.

### ***Capturing Content: Notes vs. Tapes***

These sessions are valuable to the entire team, not necessarily just those who were able to attend the meeting. For each interview, one team member

should be designated to take notes. It is helpful to make an audio recording of the session. These are to be used as input into the documentation process. Make sure that this is acceptable to everyone in the interview. The tapes will be helpful for even the most compulsive note takers. The interview discussions can be very fast, and it will interrupt the flow if you have to ask someone to “repeat the last six things.” Knowing that the ideas are captured on the tape keeps the conversation moving quickly and enables the great thoughts to be captured.

### ***Running the Interview***

After the responsibilities of each interviewee have been shared, begin asking broad questions about goals, objectives, and performance. Sample questions that can be used were provided earlier in this chapter. Explore business themes, analyses, and data, as mentioned. Be sure to give each interviewee the opportunity to share. Take the time to specifically address each one, asking for his or her specific input.

### ***Concluding the Interview***

Leave ten to fifteen minutes at the end of the session for wrap-up discussion. Ask if there is anything else that the project team should be aware of that could impact the data warehouse project. Also ask whether the interviewees have any questions about the project. As part of the wrap-up, review the next steps. The participants will be asked to review the interview documentation for accuracy. Once all of the interview documents are completed, they will be published. This enables others to read what was discussed in this session and for this group to read what was shared in the other interviews. Be sure to set proper expectations—just because something was discussed does not mean that it will be included in the data warehouse with this project. Offer to be available for questions and thank everyone for their time.

## **Putting the Pieces Together**

---

The interview sessions are helpful, but in order for the value to last, what was learned must be documented. This enables the interviewees to validate what was learned, and it helps other project team members who were not in that specific session. In addition to ensuring that the information is documented for future use, the process of writing documentation also helps to crystallize what was shared. What may make sense during a meeting may not make sense later. These topics must be understood in order to be able to write even a short paragraph about them.

## Individual Interview Documentation

Document each interview session. The interview documentation is not simply a transcription of the meeting. This is an analysis and interpretation of what was said. The essence of the discussion must be captured, rather than the word for word dialogue. Nor is a bulleted list of what was learned during an interview session sufficient. Each person who reads items on a list interprets it in his or her own way. This could result in ten or fifteen different meanings, all of which may fail to reflect what the person who wrote the list meant.

A brief paragraph should be enough to ensure that everyone knows what each item means. For example, “marketing performance monitoring” could mean tracking the response rate of a promotion, tracking promotion expenses to the budget, or even tracking the performance of each of the brand managers who work in the marketing department. All three could be useful, but which one was expressed as a requirement and which one is included in the project scope? Thorough requirements documentation would clarify this.

### ***Responsibilities***

This is the easiest section of the interview summary to write. This is where the basic job responsibilities of each interview participant are documented. This should not be lengthy or too detailed, but simply provide an overview of what this person’s job is at the time of the interview. This helps in the future, when people may have changed jobs, to be able to see what perspective they were representing at that time.

### ***Business Themes***

This is the hardest section to document. The business theme is a topic that was touched upon during the interview. Most likely, the topics are mentioned multiple times. A complete description of the topic is needed. Think about the heart of the matter. Each time something is mentioned, it may be another aspect or detail that, when collected, provides a more complete picture. These should be grouped into logical categories of related themes. For example, multiple themes may have been mentioned in the area of product development, finance, claims management, sales and marketing, or school accountability. Under each of these categories, the individual themes will be described. Most of the themes are related to areas of the business itself, but several themes are much more specific to data and systems. These are important, but they should not be the only requirements that are collected.

## ***Business Data***

This section should document the data that was mentioned by the business. This is not limited to specific databases that are managed by IT. This must represent all of the different types of data that are used for analysis. This will include the data that is captured and stored by the primary application systems within the organization. This section should also include a description of spreadsheets or personal databases that are maintained within the business community. If the specific system where this data can be found is identified, then it should be noted. The lack of a specific system name should not exclude business data from being listed in the interview documentation.

### **INFORMAL DATA SOURCES**

**The systems community can be myopic about what constitutes a data source. Typically, data sources are only considered to be “real” if the data is stored in a database that is backed up and maintained by computer operations. However, from a data warehousing perspective, data that resides outside of this traditional viewpoint must be considered. The primary characteristics of data are as follows:**

- **There is a process in place to collect the data.**
- **There is a process in place to validate the data.**
- **The data is stored.**
- **The data is distributed to others in the business community.**
- **Business decisions rely on having this data available.**

**This is real data. The data may be collected by an administrative assistant over the phone, stored in a spreadsheet, and then e-mailed to all of the managers. Through the collection process, anomalies are identified because the same person typically provides and enters the data. If there are huge differences between the current data collection and the last data collection, then this would be questioned when it is reported.**

**The value of including data like this in the data warehouse is that there would be a consistent mechanism for integrating this data with the rest of the data in the data warehouse. This data will be used for decision making, whether it is easily linked with data from the data warehouse or not. Identify this more informal data during the requirements gathering process. Later, techniques for including this information in the environment are explored.**

Each of the individual interview documents should be distributed to each interview participant for feedback. Once each participant has provided feedback, the updated documents can be published for everyone else to read. If the document contains a highly sensitive topic or content, consider excluding this confidential information from the public documents. The rest are typically published as part of an overall requirements document.

## Consolidated Requirements Documentation

Now that each individual interview has been validated, it is time to consolidate the findings across the organization.

### *Executive Summary*

Although a great deal of time and effort goes into gathering and documenting the business requirements, many people only need to have an overview of this information. This is the first section of the document, but it should be written last. The executive summary provides a high-level snapshot of the business issues and themes facing the organization. It should include the following sections:

- **Purpose:** Include a brief description of why and how these requirements were collected.
- **Business Theme Summary:** This should be a three-to-five page summary of the major business themes that were identified. This is a high-level summary of the business themes that resonate across the entire organization. More details about each theme are included in the next section of the requirements document.
- **Critical Findings and Recommendations:** This is where important conclusions and recommendations of the project team should be documented. If there has been a management study or perhaps a report from an oversight committee, it can be beneficial to tie the data warehouse project to these recommendations. The specific recommendation should be listed with the data warehouse benefit.
- **Interview Process:** A brief description of how the requirements were gathered is useful. The description may look like the following paragraphs.

*The objective of the interview process is to provide the organization with a business requirements document that outlines the critical business themes and issues and identifies key data required for analysis.*

*The methodology employed for gathering business requirements included interviewing a number of staff members who utilize information for decision making. These individuals represent a cross section of the company's functions. It was during this interview process that the critical business issues and high-level data requirements were identified.*

- **List of Participants:** Include a complete list of all of the individuals who were interviewed.



### ***Consolidated Business Themes***

This is a consolidation of what was documented for each interview session. When a business theme was discussed in more than one session, make sure that this provides a complete, cross-functional description reflecting all of the needs. Even if a theme was only mentioned in one session, it should still be included here. In the future, this will be the primary source for requirements. The individual session documentation will serve as a reference for finding out more about a topic. Examples of how the business themes are documented were provided previously in this chapter as the samples.

**NOTE** Why not just jump to a consolidated document? It is much more difficult to analyze what was said in multiple sessions at once than to understand and document each step along the way. It is also helpful to keep the interviewees involved and feeling like their participation is specifically benefiting the project. When the requirements are melded together, it is more difficult to determine where they actually originated.

### ***Candidate Business Analyses***

Each business theme will have at least one analysis that helps address those challenges. If specific analyses are not mentioned, then the project must reflect on the theme to determine what might be helpful to support it. The process to identify analyses is usually driven by the business representatives on the project team. If there are too many business themes to fully develop the analyses, then this may be done for those themes that will be the focus of this project. The last section of this chapter reviews how to ensure that the project charter and scope align with what was learned during this requirements gathering process.

In other cases, specific business analyses may have been mentioned during the interview process and documented for that session. If so, this would be a consolidation of all the different analyses mentioned. Sometimes you will have a collection of analyses that have been identified, which can help you to identify an overall business theme. While looking at the types of analyses that are currently performed can be helpful, this should not be an exercise to create a complete report inventory.

### ***Consolidated Business Data Requirements***

This is a consolidation of all of the different data that was identified in each interview session. Again, if the same data was mentioned multiple times, then a complete description of that data should be put together. If the source of the business data is well known, then the more detailed name of the database



or application system should be noted here. If it is known that this data is not captured currently, that should also be noted. Finally, include any details that were shared about where this type of data may be found. The intent is to provide a complete picture about the data that the business community would like to use for analysis and reporting.

### ***Identification of Non-Data Warehouse Requirements***

While conducting interview sessions, the project team often learns about other concerns that are not specific to a data warehouse. If the biggest and most often reported data-related pain point is the need for consistent and accurate customer and vendor contact information, then perhaps there is a need to develop a centralized directory database that will be used by all operational applications. This is a highly visible and critical business requirement, but not something that a data warehouse delivers. Note these types of non-data-warehouse requirements in the documentation, but clearly label them. During the interview session, identify such items as *not* a data warehouse function, and indicate that you can't comment on if, when, or how this can or will be addressed, but that the project team is willing to share this information with the appropriate systems personnel.

## **Common Requirements Gathering Challenges**

Many other techniques are used to gather requirements for other systems projects, although a lot of the techniques are not successful in the long run for a data warehouse. This section reviews the most common pitfalls and challenges that you are likely to face while gathering requirements.

### ***Sifting Through Reports***

While looking at the existing reports can be done, it will not actually provide the project with requirements that result in a design to support the organization into the future. Rather, this provides details about what has been developed in the past. One thing that is used to improve this approach is to capture the frequency with which the report is run. However, this does not tell you whether anyone looks at the report results or what the report is used for. Typically, an existing report is simply one step in a complex set of activities that results in the report that is actually provided to management to be used for decision making.

It can be helpful to dig into more detail for those reports identified by the business as being useful. This may be done during an interview or as a follow-up. Either way, it is important to understand why a given report is useful and what it is used for. Studying reports should not take the place of understanding the business themes.

### ***Listing Data Elements***

Because a data warehouse is for storing data, business groups may put together a list of data elements that they think they need. They may, for example, compile a list of the key performance indicators that are needed to manage the business. These are often developed independently of any actual data source research.

The problem with using this list as the basis for developing a data warehouse is that it may involve pulling data from many sources and limits the work to just the elements on the list. Once these are delivered through the data warehouse, change begins. The first five requested elements might not turn out to be what was expected. The second five data elements may be exactly what was requested but turn out to not be helpful to the business. This leaves only the last five data elements, which are part of what was wanted. Now, work must begin to go in search of the next suite of data elements that are of interest to the business. This requires starting back at the underlying data sources.

### ***Developing Functional Specifications***

The purpose of developing functional specifications is to define what a computer application will do, how it looks, and what user interactions are needed. This is an important step in the design of an operational system. However, it does not have the same value for data warehouse design. The use of a data warehouse is quite fluid and can change from person to person and from week to week. Data warehouse design is much more focused on the optimal organization of the data needed for analysis. However, once the database has been designed, it can be beneficial to develop functional specifications for the extract, transform and load processes to populate the data warehouse.

### ***Moving Beyond Immediate***

Many people are unable to get beyond the immediate data or report frustrations. These frustrations may also be more operational in nature and not anything that can be addressed with a data warehouse anyway. Listen carefully; you may identify an opportunity to address a problem in an existing system. This helps the overall organization, but it does not provide data warehouse requirements. If this happens during an interview, encourage the group to let off some steam, and then direct the discussion toward the future. Tell the group to imagine that all of these issues have been resolved. All the data is clean, integrated, and at their fingertips. Let this thought sink in a moment, and then ask the group what they would do now. This usually leads the interview session into exploring the types of business requirements that are needed to take the organization forward.

### ***Lack of Requirements***

There may be a general sense that if the organization had a data warehouse it would help the business to be more successful. However, there may not be any more specific or detailed requirement that can be found. The processes that are currently in place are used again and again. Everyone knows how to execute their part of the process. Often, these individuals do not have a real need for a data warehouse. In fact, a data warehouse may require more effort. Initially these individuals would need to learn a new tool, have to figure where to find data they need (which is not where it used to be), and then produce their part of the puzzle. From their perspective, this is a lot more work to get to the same answers they get today.

If this is the case, more research must be done to ferret out true requirements. Perhaps a different group is really driving the need. Sometimes this comes out of requirements at the highest management levels. Real business requirements must be identified before moving forward at all!

### ***The Cynic***

The project team must be prepared to explain how this project is different and more likely to succeed when other attempts have failed. It is also worthwhile to ask this type of individual what needs to be in place to make this initiative successful. Explore all areas—from data to technology to organization and staffing. Why did previous attempts not succeed?

This may be a single individual, a small group, or in some cases most of the business community itself. There must be a joint effort between the business and systems team members to address these concerns.

If left untreated, the business people may not invest their time and energy to a level that is needed, which will then contribute to this project failing too. There must be enough confidence that everyone is willing to try their best for success. Unfortunately, in the worst case, the executive business sponsor may need to partner with other executives to send a message that this is important and everyone must be encouraged to cooperate.

#### **DANGERS OF GATEKEEPERS**

**It is easy to push the responsibility to participate in the project to the people who support reporting and analysis today. These people are often business analysts within the business community. Indeed, these people are knowledgeable about the inner workings of the source systems. They have the technical expertise to pull data from every corner of the organization. The problem with relying only on these business analysts is that their knowledge is**

*(continued)*

**DANGERS OF GATEKEEPERS (continued)**

based upon what is being asked today. Superior analysts have an understanding of why reports are run, but many others do not. They simply run reports that are already put in place and deliver the results to the appropriate person.

Over time, the business community gets a reasonable understanding of how long it takes to get different types of reports, so business people filter what is requested. If an answer is needed this afternoon but it always takes three days to get the data, it will not be requested. In addition, some reports require so much effort that the business community does not ask, as it would divert resources from other required work.

These business analysts can actually become a roadblock for a data warehouse project. The ability for individuals to access data directly can be perceived as a threat to their jobs. It is important to ensure that these people understand that the goal is to free them from manual data gathering so that they can spend time actually performing analysis.

To get true business requirements, talk directly to the people who receive reports and results of analyses. These gatekeepers provide good insight that is needed for the data warehouse, but don't limit the input for the project to their perspective.

Now that the business requirements have been collected and documented, it is time to revisit the objectives of the project.

---

## Setting Attainable Goals

---

Before moving to the next step, take a moment to look again at the project charter and scope. Those project documents were created without the in-depth insight that has been gleaned during the requirements gathering process. In many instances the original project intent is reinforced by the business requirements. If so, then there is no need to go through the steps described next. However, there are also many cases where the business requirements do not align with the project charter. Rather than forge ahead with the original plan, now is the time to reevaluate the charter and scope.

This mismatch between requirements and the project charter is often identified by members of the project team. This discrepancy must be communicated to the business champion, DW manager, executive business sponsor, and IT sponsor. In order to address the situation, the requirements findings can be used to propose other possible project(s), exploring the potential business value of each. In a joint effort between key IT and business managers, the different alternatives can be prioritized, and finally the project charter and scope must be modified to reflect the new direction. This would certainly be a

case to invoke the project change control procedure, and the entire statement of work—including estimated effort, timeline and costs—would need to be revised. The first step is to figure out the different project alternatives.

**NOTE** In some cases, you might discover that the biggest business requirements are not in reporting and analysis, but in other areas of the systems and business processes. With limited resources, perhaps this is not the right time to spend money on data warehousing. It may be best to postpone building the data warehouse until these other issues are addressed.

## Exploring Alternatives

The first step to prepare for a joint prioritization session is to figure out what different possible projects could look like. Key members of the project team, often those who were involved in gathering the requirements, typically do the initial legwork. The objective is to figure out what complete business analyses could be supported by different business data. There might be too many business analyses and different types of data to be able to perform detailed analysis on all of this. The clear outliers can be eliminated. For example, do not include data that is not available, and don't spend time digging into the details of analyses mentioned by a one person from an auxiliary group.

Using a matrix, show the relationship between each business analysis and the data needed to support it. A sample matrix is shown in Table 6-1. Each row represents an analysis detailed in the requirements document. Each column represents a different type of business data. At this point, a specific system should be identified as the source of this business data. It is not worth going through this process for data that is not being captured. The body of the matrix is filled in when a specific analysis requires that data.

**Table 6-1** Matrix Showing the Relationship between Analyses and Business Data

BUSINESS ANALYSIS	SHIPMENTS	INVENTORY	SALES FORECAST	FINANCIAL BUDGET
Sales Performance Monitoring	X			
Sales Tracking	X			
Sales Trending	X			
Sales Rep. Performance	X		X	
Business Group Performance	X	X	X	X

This matrix can be used to understand what data is needed to support which analyses. For example, to support Business Group Performance, as shown in Table 6-1, all four data sources would need to be loaded. The matrix also shows that if Shipments data were loaded to support Sales Performance Monitoring, then three other analyses would also be performed.

In addition to simply creating the matrix, it is useful to think about the possible benefit that these analysis groups could provide. Conversely, the complexity of loading each data source must be evaluated.

Another area of consideration is the ease of integrating the sources. If all the sources under consideration use a common set of identifiers that can be used to link the data, it will be quite straightforward to process. However, if each source has its own identifiers and no one has developed a map between the sources, this will increase the complexity of any integration efforts. This is the type of background information that is needed during the joint prioritization session. The team that gathered the requirements typically does this with the help of the project manager, the business champion, and the DW manager. This group should have the background necessary to evaluate both potential business value and technical feasibility.

#### **UNDERSTANDING THE ANALYSIS AND DATA RELATIONSHIP**

**One of the most unfortunate things to observe is when a data warehouse is delivered to the business community only to have the business express frustration that it doesn't do what is needed. This often results from a series of small decisions about what to include and exclude from the project. Some data is more complex and requires a great deal more effort than was realized going into the project. Often, the only option from a project team perspective is to scale back. Agreeing to such a decision is often done without realizing the business impact of leaving some data out. All such decisions, even at the time of initial prioritization, need to be done with an understanding of what analyses can be completed with or without the data in question. It is also important to note if an analysis is only partially supported with a specific set of data.**

A completely different way to study various implementation alternatives is to isolate the analyses needed to support a specific group of people. Again, using the analysis–data source matrix, you can determine how many sources would be needed to fully support that group. Perhaps most of the analyses can be supported with the implementation of three data sources. This may be a recommendation to be made during the prioritization session.

### **Setting Priorities**

Now that combinations of business data and analyses have been explored, it is time to make decisions about what should be designed and subsequently built.

The executive steering committee needs to participate in the process of setting priorities. Each individual group will have its own highest-priority item, and each group would like its own needs met first. Because this is not possible, a decision must be made determining what would bring the most value to the entire organization.

Working to define these priorities can be done by getting key business and technical representatives in a room together, for half a day. This may seem like a lot of time to ask for, but it will be time well spent. If priorities can be set with less time, then the meeting will simply adjourn early. The participants must be able to articulate the value and impact of decisions on their own group. They must also be able to compromise for the greater good.

Conducting a joint prioritization session is especially valuable for the project team. Later, if there is any negative feedback or frustration about what is included in this project, the discussion needs to be held with business management, not the project team. The session also provides a clearer understanding of the business value and impact of each proposed alternative and the recommended direction. The goal is to be able to pick the area that will have the biggest impact on the business yet still have manageable scope to ensure successful completion.

**NOTE** Remember that this is not a process to determine what will ever be loaded into the data warehouse, but what will be loaded first.

Go into the session armed with the business themes, business analysis/data matrix, possible project alternatives, and some idea about the feasibility of building these solutions. At this point, it may be too hard to quantify a specific dollar amount to place on the potential value. Another way to get your arms around this is to compare the different themes with each other. Which has a bigger impact? Which is harder or easier to build? The goal is to identify which areas of the business will have the biggest potential benefit and are the most feasible to build. Then, more work can be done to articulate the estimated financial impact of being able to address these high-value, feasible challenges. Pick one or two of these themes that use similar data to begin building the data warehouse.

Now that you have a short list of business themes to work on, you can spend time getting into more detail about each one. Take a good look at the only the top few. It is not worth taking the time to determine whether one theme should be ranked as number eighteen or nineteen. By the time that the first data warehouse project is done, it may have become the next most important thing to work on.

The business decisions made here may impact what this project will deliver. Project change control must be used to revise the project plan, statement of



work, scope and perhaps the charter too. Changes now will help ensure that the data warehouse will support the highest priorities of the business.

Once the business priority is set, there must be patience and understanding that data will not be ready for use overnight. Data may not even be available in a matter of weeks. It may be multiple *months* before anything can be delivered. This is sometimes difficult to understand because technological advances have made it easier than ever to quickly move data around. However, often the data as it exists today is of questionable accuracy, is likely not integrated with other data across the organization, and may not even be captured anywhere. Decisions need to be made about how to properly process the data, rather than simply move it around. These rules are needed to clean, integrate, and structure the data in a more meaningful way. Let's see how the two companies we are tracking are progressing.

## **In Real Life**

---

Many organizations struggle with gathering business requirements. Techniques used for other types of systems development are not working. Now we will take a look at how our case study companies are doing.

### **A Glimpse into Giant Company**

To manage all of the employees in such a large organization, specific roles and responsibilities are clearly and somewhat rigidly defined. There are people in the business community who serve as the liaison between their group and systems. The data warehouse project team is not allowed direct access to business decision makers. This greatly inhibits the team's ability to gather business requirements.

To address this challenge, all of the managers involved, both from the business groups and IT, met to discuss the situation. With encouragement from the executive and IT sponsors, they agreed to loosen the definition of roles and empowered their staff to begin to interact directly and be flexible. This did not require that changes be made through the HR department; just the agreement between the managers to work smarter. This enabled the project team to work directly with both the business liaisons and the business decision makers.

### **Insight from Agile, Inc.**

This organization is small and nimble. Everyone is empowered to get things done. The organization does not spend much time planning—everyone just dives right in. As the company has grown, there are challenges in keeping



everyone moving in the same direction. There are always new crises that need to be addressed. Both business and IT staff spin from one crisis to the next, which leaves little time to invest in long-term efforts. This is usually viewed as an advantage, rather than a challenge.

However, developing a successful data warehouse requires spending time on things that may not have immediate benefits but will indeed make a huge difference in the future. The organization will never stop spinning unless there is a conscious effort to make it happen. Senior management must make a commitment, and stick to it, to allocate at least some people to this effort. This is easy in the beginning, but as time goes on there is a big temptation to divert the data warehouse team members to help on some immediate situation. Resist! Unless this investment is made, the organization will never get out of the mode of grabbing data to help address an immediate need. Taking the time to gather complete business requirements and looking at priorities helps the organization to see the potential value that can be realized, even from loading the first data source. While getting started with more structure was difficult, the business impact of building the data warehouse helps everyone to focus on getting the data warehouse built. Not everyone's needs are addressed up front, but those that will be met will help the entire organization to be more successful.

---

## Summary

---

Having a good understanding of the business in general and then a deeper understanding of business challenges and themes provides the necessary background to design a data warehouse that will meet the business's needs now and into the future. Well-documented requirements are not only useful for the initial project, but also tend to remain accurate for years. These capture the essence of the business and thus the basics remain constant.

Having this good foundation provides the project team with the correct perspective as they begin to develop a data model that will help the business to address whatever challenges have been identified. The next chapter discusses how to create the data model—in a way that makes sense to the business community and captures the details needed to construct the database.



# PART

## III

# Dealing with the Data

---

## In This Part

---

**Chapter 7:** Modeling the Data for your Business

**Chapter 8:** Managing Data As a Corporate Asset



# Modeling the Data for your Business

Regardless of what technology will be used, the heart and soul of the data warehouse is the data itself. How the data is organized can have a significant impact on how well the environment will work. A great deal of thought and care must go into the design of the data. This chapter presents the concepts covering how the data should be organized to support reporting and analysis. The business team will learn the following:

- Purpose of dimensional models
- Basic components of a dimensional model
- How to contribute to the modeling process

A case study of the dimensional model for call center operations is used to reinforce the concepts presented. This chapter also includes content that is geared toward the more technical members of the project team, including the following:

- A technique to document the dimensional model, specifically geared toward the business community
- A process for developing the dimensional model with participation from the business community
- An introduction to several advanced modeling concepts
- How to take the model forward

This chapter is not intended to provide an in-depth guide to dimensional modeling. It does share the basic concepts that can be used to communicate more effectively between the business and technical team members. Before diving into the modeling concepts, the first section reviews the rationale for dimensional modeling and why it is of interest to business people.

## The Purpose of Dimensional Models

---

A specific modeling technique has evolved in order to support the types of queries and analyses that businesses require. This technique is called *dimensional modeling*. This approach has been applied to data warehousing for nearly thirty years and is supported by a wide variety of database platforms and data access or business intelligence tools. Dimensional models support the business perspective of the data, and today's technology ensures that they can be effectively implemented.

Dimensional modeling is a formal data modeling technique that is used to organize and present data for analytical and reporting use. The focus is on the business perspective and representation of data. The goal is to free the data that has been captured and stored by the operational systems and make it available to the business community. Regardless of how data is structured, business people will ask questions based upon their frame of reference. This perspective is driven by the basic characteristics of the industry and how the company is organized, so why not organize the data to reflect this business perspective? The two primary goals for dimensional modeling are *ease of use* and *query performance*. These are the principles that guide the entire dimensional modeling process.

There are other data modeling techniques that play an important part in overall systems development. They help ensure that the data itself and the relationships between different data elements are clearly defined. For operational systems, it is important that the data be organized to facilitate transaction processing. This includes ensuring transaction integrity and speed. The type of modeling used for operational system design is called *entity-relationship (E-R) modeling*. This may also be referred to as *normalized modeling*. One specific form of E-R modeling represents the data in *third normal form (3NF)*. There is a complete discipline surrounding this approach to data modeling. This is mentioned to acknowledge the value and purpose of E-R modeling for operational system design. The following two sections examine the main objectives of dimensional modeling.

### Ease of Use

In order to ensure that people will use a data warehouse, the data must be presented in a manner that makes sense to them. If it is too confusing or does not mirror the way the business runs, then people are not likely to use it. Therefore, the dimensional model must cleanly represent the basic components of the business. In addition, the model must be presented in terms that are used by the business.

A well-designed dimensional model should be obvious to the business community and be met with a confirmation that it is indeed correct. If it causes

a great deal of confusion within the business community, then the model has not yet captured the business perspective.

**TIP** Don't do what is easier to implement, do what is easier for the business users.

## Query Performance

The second, and equally important, goal of a dimensional model is to ensure good query performance. If requests do not run in a timely manner, the data warehouse will not be used and will not be helpful to the business. Dimensional modeling takes the need for this query performance into account as part of the inherent design approach. The data is organized in order to provide consistent performance both for queries that are requested up front and for those that crop up later. All possible queries cannot be defined in advance, so the technique takes this into account to provide support for unpredictable access patterns.

## Understanding Your Data

---

The requirements gathering process detailed in Chapter 6 identified the different kinds of data that are needed for the DW. The prioritization process, also discussed in Chapter 6, helps ensure that the project charter and scope align with the business requirements, including analyses and the data sources needed. Now, detailed data analysis is needed to really understand what is stored in the operational databases that have been identified. A robust operational database is likely to have hundreds of tables and thousands of individual data elements.

A lot of legwork should be done by the core project team to determine which data elements may be useful. The systems staff members responsible for maintaining that application can be invaluable in this research. The objective is to narrow down the sheer number of data elements that need to be studied. This is also called *source data analysis*. This can range from running simple queries to much more sophisticated analysis. At a minimum, simple queries should be run against the application system's data structures. The types of queries include the following:

- Has this data element ever been populated?
- When was the last time this table was updated?
- What percentage of the rows in this table contain data?
- What are the possible values for this data element?
- What is the frequency with which these values are used?

The results of these queries can be used to determine which elements would be useful for reporting and analytical purposes. If the data element has never been populated, then this does not need further study. If a data element has not been changed in years, this may indicate core reference data that is still valid or it may indicate that this data is no longer being captured and/or maintained by the operational system. Obsolete data elements do not need to be addressed in the modeling process.

It is also helpful to have business analysts participate in this analysis. They can often help the team to understand the data content. Over time, the original name and meaning of a column can be lost or changed. It is helpful to know how the data is used today. The objective is not to isolate data elements that have been requested for reporting, but to identify all data elements that contain business data. This is often done through an elimination process. For example, data elements that are obsolete or those that are used to drive security for the operational system do not contain business data that would need to be included in the data warehouse.

*Data profiling* is a more comprehensive set of activities intended to look at data across all systems, operational and analytical. This is described in more detail in Chapter 8.

**TIP** Look at all the data elements, not just those specifically identified for reporting.

After there is a clear understanding of what the business is trying to accomplish and a familiarity with the data that is available, it is time to look at how that data should be organized.

## What Is a Dimensional Model?

---

A *dimensional model* is a data model organized for the purpose of user understandability and high performance. There are two basic parts of a dimensional model: the *dimensions* and the *facts*. These are the building blocks that comprise all dimensional models, simple or complex.

### Dimensions

Dimensions are groupings of data elements in major business categories. Common dimensions include the following:

- Customers
- Products
- Dates

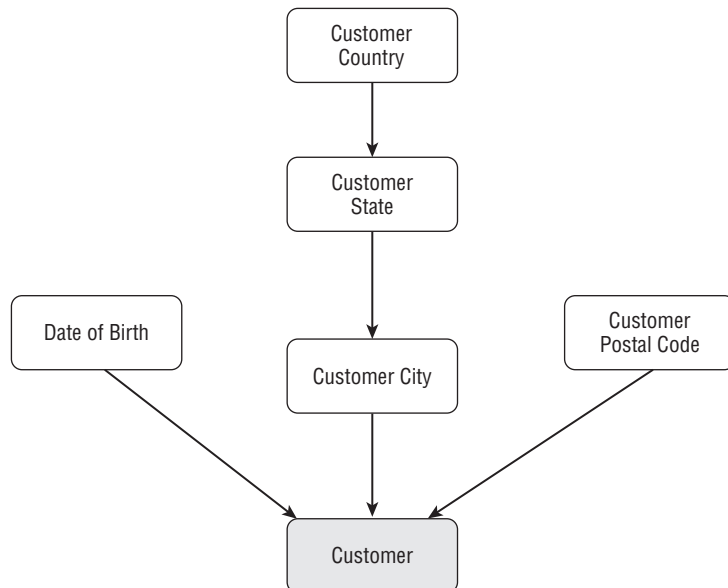


- Suppliers
- Vendors
- Accounts

The individual data elements are called *dimensional attributes*, or *reference data*. The dimensional attributes are used as row and column headings for reports. They are used to create lists of options to determine what to include or exclude on a report. The relationship between these dimensional attributes creates drill paths or the ability to navigate up and down a hierarchy.

The need for dimensional data is often recognized while gathering business requirements. It may not be directly communicated, but realized when someone needs a report by region, by week, and by product category. Each of the terms following the word “by” is a dimensional attribute. These should be included in dimensions to support that type of reporting.

An example of a customer dimension is shown in Figure 7-1. This is a highly simplified example that only shows the customer’s address and date of birth attributes. Some of these attributes relate to each other in a hierarchy, while others are simply additional characteristics of the customer. Any of these attributes could be used to constrain a query or for use in a report.



**Figure 7-1** Sample customer dimension

**TIP** A dimension can include attributes that are descriptive and that relate to each other, creating hierarchies.

The arrows represent one-to-many or one-to-one relationships. This notation is used because it is more meaningful to business users to visualize the drill paths, rather than the more traditional representation of “crow’s feet.” In fact, if looking at the arrows causes technical staff members to furrow their brows, this is the same reaction that crow’s feet have to business staff members. The intent of the business dimensional model is to communicate with the business.

## **Facts**

*Facts* are the measurement of business events. These are captured as specific information about a business event or transaction. These are measured, monitored, and tracked over time. Facts are typically the amounts and counts that show up as the body of reports. Facts are used as the basis for all calculations. Examples of facts include units ordered, retail price, amount paid, claim payment amount, gross margin, budgeted dollars, revenue forecast, and loan balance, among others.

The facts are only interesting within the proper context, and the context comes from the dimensions. For example, the fact that a company had \$10,000 in sales is not useful unless you know that it was from red shoes, in the Chicago market, the week before Valentine’s Day.

## **Using Both Parts of the Model**

Dimensional models can represent very complex businesses. Although there can be many different aspects of a business, most people deal with only a few variables at a time. Most of us can easily draw a three-dimensional cube, but it is much more difficult to draw a four dimensional cube. Similarly, while the overall model may contain over ten dimensions, often only three are reflected on a single report at a time.

The way that people think about data is often defined by the layout of a spreadsheet: rows and columns with perhaps a separate worksheet to represent another variable. For example, Figure 7-2 is a common sales performance report showing the monthly profit results for the current year, reported by product category. Each region of the organization is represented as a separate page. This report is constrained to the current year of data, and uses product, date, and sales organization dimensions. The Product Category attribute is from the Product dimension. The constraint for the current year is on the year attribute of the Date dimension. The months are also from the Date dimension. The Sales Regions (per sheet) are from the Sales Organization dimension. The fact itself is the Sales Profit, listed in thousands of dollars.

**Sales Profit for Northeast Region**  
For Current Year to Date  
(Profit in thousands of Dollars)

Product Category	Months						
	Jan	Feb	Mar	Apr	May	Jun	Jul
Camping Accessories	19	22	37	52	65	84	83
Women's Clothing	63	68	87	65	62	74	69
Men's Clothing	72	81	80	94	87	103	78
Athletic Shoes	201	214	194	183	191	192	199
Fishing Accessories	6	3	8	11	21	17	18
Backpacks	3	4	4	12	16	19	15
Tents	88	83	91	137	139	189	120
Total	452	475	501	554	581	678	582

**Figure 7-2** Sample Sales Results Report

The dimensions and facts are used together to create basic reports such as this one, or more sophisticated analyses. More detail about using dimensional models can be found in the section “A Call Center Case Study.”

## Implementing a Dimensional Model

Many different database technologies are available today to store data. Many of these have been developed specifically to support data warehouses. It is useful to have a basic understanding of these in order to be able to put the dimensional model into the appropriate context. A dimensional model is not inherently tied to a specific technology, and it can be implemented in a variety of different ways. Different types of databases include the following:

- **Relational databases** are one of the most commonly used databases for both operational and data warehouse systems today. When a dimensional model is stored in a relational database it is called a *star schema*. This is due to the visual appearance of the dimension tables surrounding the fact table.
- **Multi-dimensional databases** are specifically designed to support a dimensional view of the data. The data is stored in proprietary array structures. When a dimensional model is stored in a multi-dimensional database, it is called a *cube* (even though it may actually contain more than three dimensions).
- **Proprietary databases** are also available on the market. Many of these are designed specifically to support reporting and analytical use. There is a wide range of different methods used to physically organize and access the data in these environments. Once the dimensional model is completed, there is often no additional design work required to determine how the data will be structured in these environments. This is all handled by the proprietary system.

When database software and hardware are fine-tuned and bundled together, they are called a *data warehouse appliance*. These concepts are included here because they are commonly used and most readers have likely heard several of these terms.

## Diagramming Your Dimensional Model

There are different ways to document and present dimensional models. One of the most common ways that dimensional models are depicted are as tables to be stored in a relational database. The dimensional model can be documented using the same modeling tool that is used to develop any other data models for the relational database. Each of the dimensional attributes is included and represented using logical names that should be meaningful to the business. This type of table diagram is easily understood by systems professionals, but it is not as clear to business professionals.

Another method to document your dimensional model is to present business diagrams. The intent is to visually present the model in terms that more closely reflect the interface that will ultimately be presented for access. This is called *business dimensional modeling*, and it can be documented using any visual diagramming tool.

Careful analysis must be performed and dimensional modeling principles must be followed in either case. The primary difference is how the model is presented to the business. The diagram notation of the business dimensional model is reviewed next. The process to develop a dimensional model is discussed later in this chapter.

## The Business Dimensional Model

---

With the increasing variety of options for building the data warehouse, it is important to split the business perspective from the technical perspective regarding the data. The *Business Dimensional Model (BDM)* is a data model that is specifically geared toward working with the business community. It serves as an abstraction layer that insulates the model from technical implementation details. The model also serves as a communication vehicle between the business and systems groups. The model shows diagrams of the dimensions and the facts so that the details can be reviewed and discussed in business terms. This also separates the business discussion from any technical discussions.

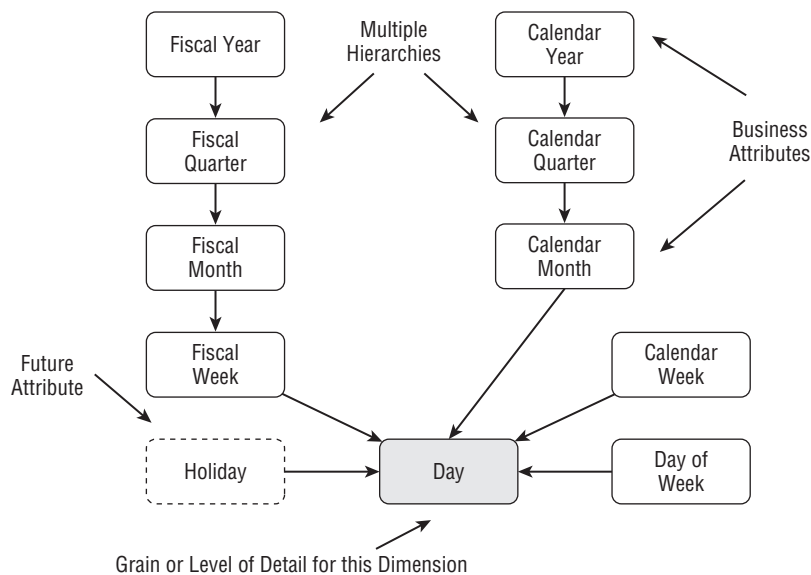
The business side of the problem can be addressed in enough detail so that the systems team can implement the model in any technology. Limited technical tips are shared at the end of this chapter to guide the project team when implementing the dimensional model in a relational database management

system. Other technical implementation details are not covered in detail in this book. The project team should work directly with the technology vendors for guidelines and recommendations specific to whatever tool is being used.

A dimensional model can reflect a wide range of data from multiple data sources. The focus is on understanding the dimensionality of the business itself and the facts that are needed to measure that part of the business. The business dimensional model will reflect all of the data to be included in the data mart. A summary of all of the notation used for the business dimensional model is provided at the end of the chapter.

## Business Dimensions

Each business dimension of the model will be designed and diagrammed separately. The business attributes are included to fully describe the dimension. Each business attribute of the dimension is depicted with a rectangle, as shown in the sample dimensions in Figure 7-3. For each dimension, you need to identify the lowest level of detail that exists. This is also called the *grain* of the dimension. This lowest-level attribute is shaded. The relationship between each of the business attributes is noted with an arrow. The direction of the arrow shows the direction to drill down to see more detail. These are also commonly referred to as *hierarchies*. Note that this organization's fiscal year does not align with the calendar year. The grain for this dimension, or the lowest level of detail, is the Day attribute. Each individual business attribute is included in the dimension diagram.



**Figure 7-3** Date dimension example

Sometimes a specific business attribute is requested but is not captured in any current source systems. To facilitate communication, this can be included in the dimension diagram, but the notation is different. The dotted rectangle indicates that the data element is either not captured or is not to be included in the initial implementation. In this example, the Holiday attribute is planned for the future. This ensures that the model reflects true business needs, but it also helps set and maintain expectations that the element is not going to be available at this time.

The dimension diagram itself is only part of the documentation that is needed for the dimension. The diagram shows each attribute with a useful business label. In Figure 7-3, note that each of the date hierarchies is uniquely named. This helps ensure that the correct attribute is easily selected for reporting. It is also important to have a clear definition of each attribute, and several sample values. The sample values are often what will spark recognition of what an attribute represents. This does not need to be a complete list of all the possible values for this attribute. Table 7-1 shows the additional documentation that is needed to support the dimension diagram.

#### DATA NAMES AND DEFINITIONS

**The business community must take a lead role in developing clear, meaningful names for each of these attributes. The business must also be responsible for creating the definition of each attribute. These are critical to ensure that the model is easily understandable and accurately documented. Meaningful data definitions are one part of overall data governance. This topic is important enough that Chapter 8 is devoted to discussing data ownership and governance issues.**

## Fact Groups

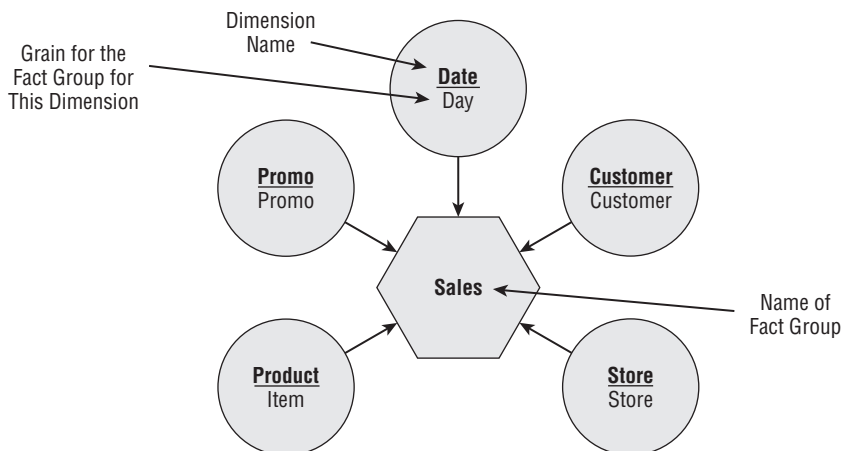
The second part of the model contains the facts, which is where the business measurements are stored. Modeling the facts is much more than simply creating a list of the business measures that are needed. Each of these facts must be reflected within the proper context. This can be understood by looking at how the data is captured and how the business uses each fact. The dimensions that are relevant to these facts are shown. The grain, or lowest level of detail that applies, is also identified for each applicable dimension. Often, several facts will have the same dimensionality and identical grain. These facts can be put together into a *fact group*.

**Table 7-1** Date Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Day	Calendar day information was reported	6/31/2003, 9/1/2003
Holiday	This is a future attribute to indicate the name of the federal holiday recognized by the U.S. government.	Memorial Day, President's Day
Fiscal Week	The week of the fiscal year that this date rolls up to; used for reporting purposes	FY 2003 Week 16
Fiscal Month	The month of the fiscal year that this date rolls up to; used for reporting purposes	FY 2003-10, FY 2003-11
Fiscal Quarter	The quarter of the fiscal year that this date rolls up to; used for reporting purposes	FY 2003-Q1, FY 2003-Q2
Fiscal Year	The year that this date rolls up to based upon the company's fiscal calendar	FY 2003
Calendar Month	The calendar month used for reporting purposes	March 2003, October 2003
Calendar Quarter	The calendar quarter used for reporting purposes	CY 2003-Q1, CY 2003-Q2
Calendar Year	The calendar year used for reporting purposes	CY 2003
Calendar Week	The calendar week ending Saturday and used for reporting purposes	Week Ending 11/21/1998
Day of Week	Attribute to indicate the calendar day of the week	Wednesday, Saturday

Figure 7-4 shows a fact group for retail sales. The name of the fact group is in the hexagon at the center of the diagram. Each dimension that applies to this fact group is diagrammed as a bubble around the hexagon. The specific grain is also included in each bubble. This is important because facts can have similar dimensionality, but be at different levels of detail. If you have two facts that have the same dimensions but one is daily and the other is monthly, these must be split into two separate fact groups. You must design only single-grain fact groups. Mixing the grain of a fact group can cause query results to be incorrect.

**TIP** Never mix the grain of facts in a fact group.

**Figure 7-4** Retail Sales fact group

The name and definition for each fact must be documented. Table 7-2 shows this information for a Retail Sales Fact Group. The table also includes a column that indicates what to do with the facts as you drill up and down through hierarchies and group the facts by other attributes from the dimensions. This is the default *aggregation rule* for each fact. The aggregation rule is also needed if a query excludes a dimension entirely. For example, if you want to know the number of units sold by Calendar Month, then the other dimensions are not included in this request. The Units Sold can be summed and then grouped by the Calendar Months to fulfill this request. The most common default aggregation rules are to sum, or count, the fact.

**Table 7-2** Retail Sales Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Units Sold	Number of consumer units sold	SUM
Sales Dollars	Dollar amount for the units sold	SUM

Using the diagrams for the dimensions and fact groups, you can see what types of queries can be supported. A more complete case study is shown next to better illustrate this capability.

## A Call Center Case Study

Most large organizations have some sort of call center. These centers are sometimes run internally or may be outsourced to third-party organizations.



The call center in this example is run internally to handle customer calls. These calls may be to place an order, check the status of an order, request information, file a complaint, or share a compliment. This case study shows the business dimensional model for this organization. While this shows a mostly complete picture of a dimensional model, keep in mind that it has been simplified to facilitate an introduction to the modeling approach, rather than to represent a complete model for a real call center. Clearly, there are many nuances and details that are specific to each organization that cannot be addressed here.

## Call Center Dimensions

The following sections describe each of the dimensions needed. Both the dimensional diagram and the supporting table provide a complete description of each of these dimensions.

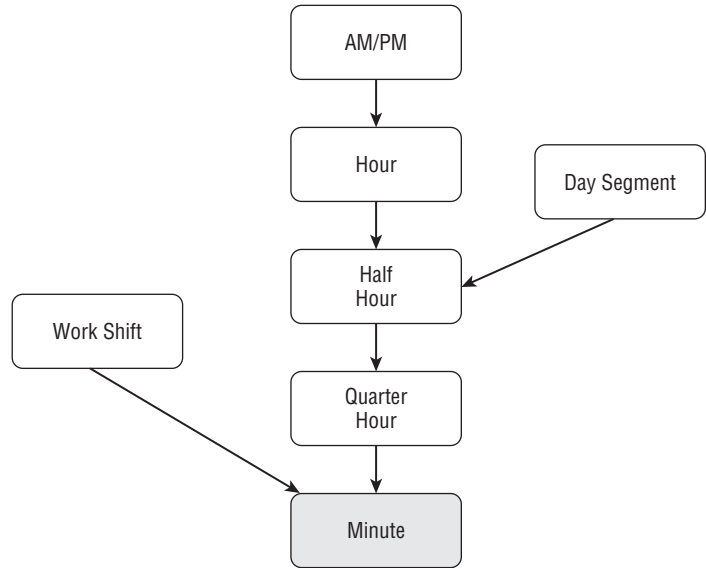
### ***Date Dimension***

The Date dimension for this call center case study is the same dimension shown previously in Figure 7-3. This supports analysis by fiscal and calendar years. Definitions of each of the attributes are provided in Table 7-1.

### ***Time Dimension***

Because call patterns can vary drastically throughout the day, it is important to be able to track the calls at a fine level of detail. Keeping the time of day attributes separate from the date attributes helps simplify the model, reducing the size of the dimension, which provides implementation benefits. The basic hierarchy of the Time dimension shows a straightforward rollup of time. The Time dimension can also be used to track the work shift of the organization. Although the work shifts currently change on the hour, the company is discussing changing the shifts around. In order to accommodate any possible future definition of the work shift, it is shown to the minute. The Time dimension diagram is shown in Figure 7-5.

Time dimensions are helpful to facilitate the representation of different parts of the day. If the only need for time or timestamp data is to calculate the elapsed time between two events, this would not be needed. If specific attributes are needed for grouping and reporting, these can easily be stored in a Time dimension. Also needed are special considerations to accommodate the shift between standard and daylight saving time. Separate rows can be included in the table to represent the second hour between 1:00 A.M. and 2:00 A.M. that occurs when the clocks are set back in the fall.



**Figure 7-5** Call Center Time dimension

The data definitions for the Call Center Time dimension are included in Table 7-3.

**Table 7-3** Call Center Time Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Minute	The specific minute of the day. These begin at midnight and finish at 11:59 P.M.	1:01 A.M.; 6:32 P.M.
Work Shift	Three shifts are worked on a daily basis. This attribute describes to which work shift this minute of the day belongs. These shifts are stored and noted based upon the local time zone. For example, the morning shift runs from 7:00 A.M. to 4:00 P.M. local time. In other words, people are working the morning shift in Maine before the headquarters is even open in Oregon.	Morning, Afternoon

(continued)

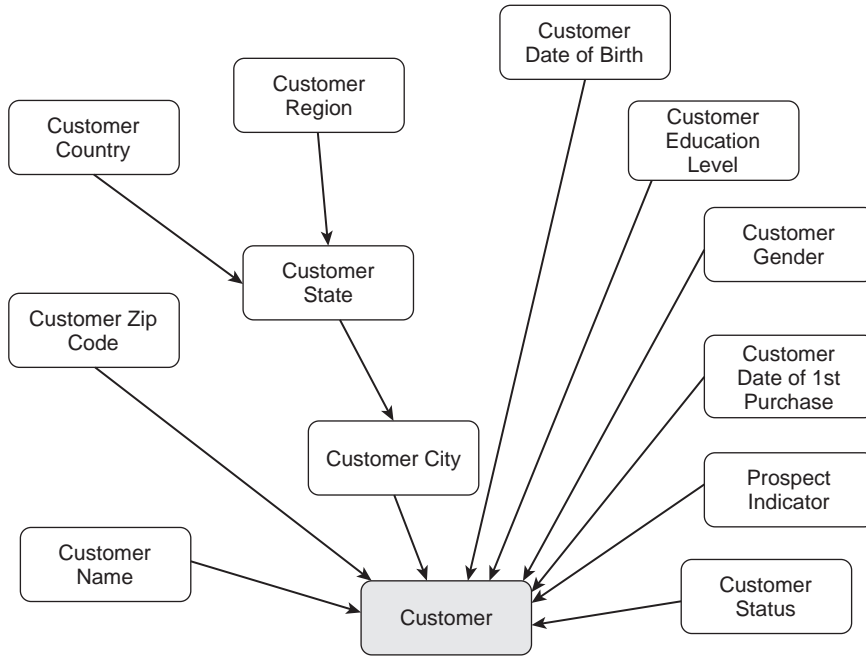
**Table 7-3** (*continued*)

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Quarter Hour	This attribute contains the minutes grouped into 15-minute segments. This is needed to track employee time at the quarter-hour level of detail. This is represented by the beginning minute of the 15-minute range.	4:30 P.M. quarter hour, 4:45 P.M. quarter hour
Half Hour	This attribute contains the minutes grouped into 30-minute segments. This could also be described as two quarter hours. This is represented by the beginning minute of the 30-minute range.	4:00 P.M. half hour, 4:30 P.M. half hour, 5:00 P.M. half hour
Hour	This attribute contains the specific hour of the day. This is represented by the beginning minute of the 60-minute range. Note that these could be stored in 24-hour, or military, format (e.g., 14:00) if this is meaningful and useful to the business community.	1:00 A.M. hour; 2:00 P.M. hour
AM/PM	This is a quick representation of the notation for the minutes between midnight and noon, and noon to midnight.	AM, PM
Day Segment	There are divisions of time that are not cleanly aligned with the clock notations of AM and PM or the employee shifts. These are defined by the patterns when customers call our hotline.	Before work (6:00 A.M. to 8:00 A.M.); Morning (8:00 A.M. to 11:30 A.M.); Lunchtime (11:30 A.M. to 1:30 P.M.); Overnight (11:00 P.M. to 6:00 A.M.),

### **Customer Dimension**

The Customer dimension provides information about the individuals who place calls to the company. While some attributes are known through doing business with this customer, the organization may also purchase external demographic data. Figure 7-6 shows the Call Center Customer dimension.

The data definitions for the Call Center Customer dimension are included in Table 7-4.



**Figure 7-6** Call Center Customer dimension

**Table 7-4** Call Center Customer Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Customer	A unique identifier for the individual consumer	100, 101
Customer Name	Name of the customer	Kate Twaddle, Sarah Mogensen
Customer Zip Code	Zip code for the primary residence for this customer	48801, 60663
Customer City	Name of the city of the primary residence for this customer	Harbor Beach, Peoria
Customer State	Name of the state of the primary residence for this customer	Illinois, Texas
Customer Country	Name of the country of the primary residence for this customer	US
Customer Region	Sales region to which the customer belongs	East, Central, Southeast

(continued)

**Table 7-4** (*continued*)

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Customer Date of Birth	Date the customer was born. This is used to determine the customer's age.	04/24/1947, 08/29/1972
Customer Education Level	Highest level of education that this customer completed	High School, Associates, Bachelors, Masters
Customer Gender	Gender of the customer	Male, Female
Customer Date of First Purchase	Date when the customer made his or her first product purchase	03/14/1998, 07/03/2004
Prospect Indicator	Denotes whether this person has even been a customer, or whether this person is known through a marketing campaign or has requested information	Customer, Prospect
Customer Status	Indicates whether this is customer is still an active customer. The customer status is set to not applicable if this is a prospect.	Active, Inactive, Not Applicable

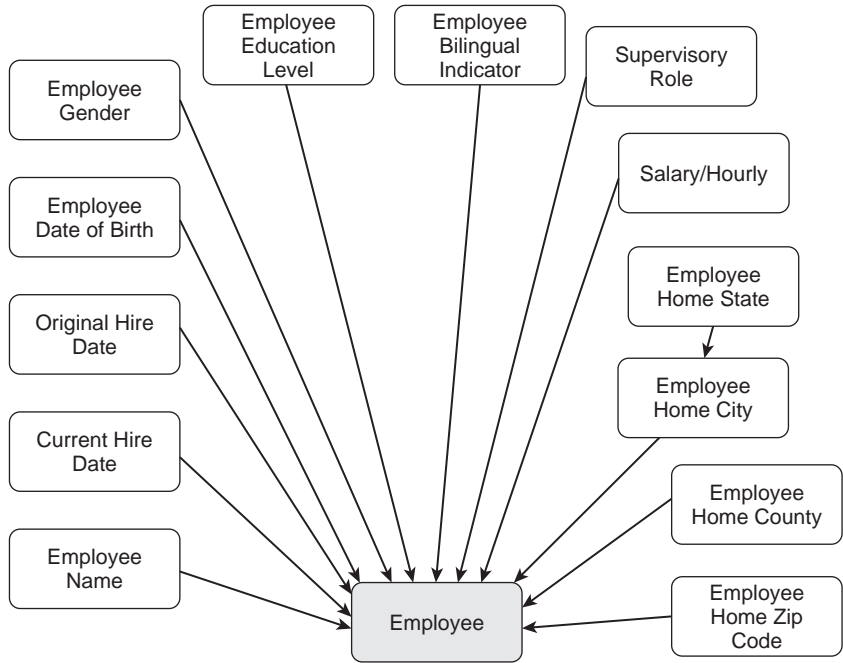
### ***Employee Dimension***

Managing a call center is highly dependent upon the customer service representatives, the people who answer the phones. There is a high turnover rate for call center employees. Productivity of these employees is critical. Figure 7-7 shows the Cell Center Employee dimension.

The data definitions for the Call Center Employee dimension are included in Table 7-5.

### ***Call Dimension***

Each specific call must be tracked by the call dimension. There is very little that is known about the call itself that is not already included in other dimensions. This is what is used to link the participation of multiple different employees for a variety of activities. This can happen if a call is transferred. The data warehousing term for this is a *degenerate dimension*. Figure 7-8 shows the Call dimension.



**Figure 7-7** Call Center Employee dimension

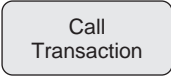
**Table 7-5** Call Center Employee Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Employee	Uniquely identifies each employee	2000, 3874
Employee Name	Name of the employee	Debbie Smith, Paul Jones
Current Hire Date	Most recent date that this person was hired.	09/01/2005
Original Hire Date	Date that this person was hired for the first time	09/01/2005
Employee Date of Birth	Employee's birth date	12/17/1975
Employee Gender	Gender of the employee	Male, Female
Employee Education Level	Indicates the highest level of education that this employee has completed	High School, Associates, Bachelors, Masters

(continued)

**Table 7-5** (continued)

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Bilingual Employee Indicator	Indicates whether the employee speaks more than one language fluently	Bilingual, Not Bilingual
Supervisory Role	Indicates whether this person has any managerial responsibilities. This includes having both budget responsibilities and personnel responsibilities.	Supervisor, Non-management
Salary/Hourly	Indicates whether the employee is paid based upon hourly or salaried basis	Salaried, Hourly
Employee Home State	Name of the state of the primary residence for this employee	Iowa, Florida
Employee Home City	Name of the city of the primary residence for this employee	San Francisco, Nashville
Employee Home County	Name of the county of the primary residence for this employee	Cook, Kane
Employee Home Zip Code	The five-digit postal code of the primary residence for this employee	40441, 48060



Call  
Transaction

**Figure 7-8** Call Center Call dimension

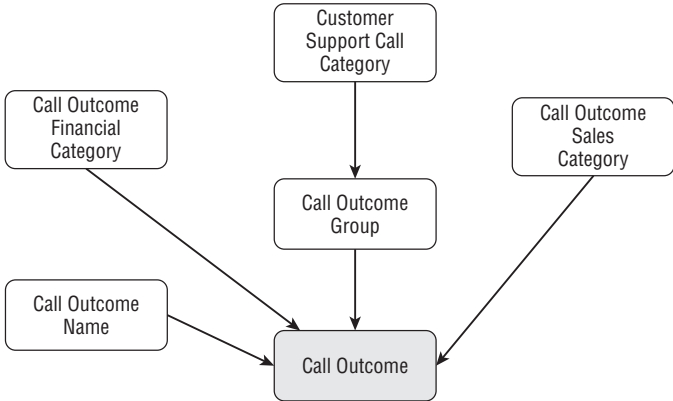
The data definitions for the Call Center Call dimension are included in Table 7-6.

**Table 7-6** Call Center Call Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Call Transaction	This is the unique identifier that is assigned to a call when it comes into the call center. This is tracked until the call is completed.	123, 446

### Call Outcome Dimension

When handling incoming calls, the purpose and type of call are tracked. The sales, finance, and call center groups each categorize these calls differently. Each of these groupings is shown in Figure 7-9.



**Figure 7-9** Call Center Call Outcome dimension

The data definitions for the Call Center Call Outcome dimension are included in Table 7-7.

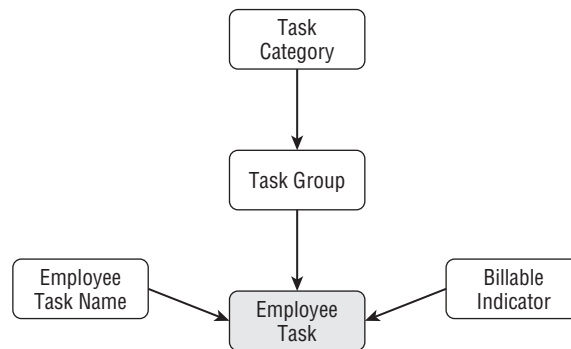
**Table 7-7** Call Center Call Outcome Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Call Outcome	A unique identifier for each possible result of a phone call	3900, 3901
Call Outcome Name	The label for each of the call outcomes, which describes the purpose of the call	Logged Complaint, Customer Question, Answered Prospect's Question, Placed Order
Call Outcome Financial Category	A grouping of call outcomes used by the financial group	Revenue Generation, Sales Support, Customer Service
Call Outcome Group	A collection of detailed call outcomes	Customer Complaints, Order Related, Prospect Activity
Customer Support Call Category	A grouping of call outcomes used by the call center management	Order Handling, Information Distribution, Problem Handling
Call Outcome Sales Category	A grouping of call outcomes used by the sales department	Customer, Prospects



### Employee Task Dimension

Employee time tracking is based upon whatever activity or task that employee was working on. This is needed to be able to help understand the dynamic between employee activity and call productivity. Figure 7-10 shows the Employee Task dimension.



**Figure 7-10** Call Center Employee Task dimension

The data definitions for the Call Center Employee Task dimension are included in Table 7-8.

**Table 7-8** Call Center Employee Task Dimension Attribute Definitions

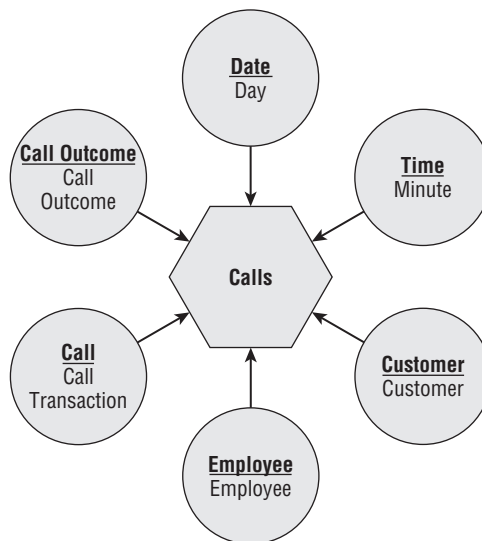
ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Employee Task	A unique identifier for each specific customer task	500, 600
Employee Task Name	The name of the unique task indicating how an employee's time is spent	Handling Phone Orders, Logging Customer Complaints, Vacation, Training, Call Follow-Up
Billable Indicator	Identifies employee tasks that can be charged to the customer	Billable, Non-Billable
Task Group	A grouping of employee tasks for management reporting	Internal Administration, Problem Handling, Customer Sales and Marketing
Task Category	A high-level grouping of employee tasks for management reporting	Customer, Internal

## Call Center Fact Groups

Now that each of the dimensions is defined, it is time to look at the different facts that are needed. These are documented next.

### *Calls Fact Group*

Some of the most critical facts that need to be tracked for a call center are those related to the handling of incoming calls. The Calls fact group is shown in Figure 7-11. Note that not every dimension defined earlier is included in this diagram, only those that are relevant to the calls facts. The minute that the call began for this employee is recorded for the Time dimension.



**Figure 7-11** Call Center Calls fact group

The specific facts that are included in this fact group are described in Table 7-9.

**NOTE** Having the Number of Calls set to one for each row allows a SUM function, rather than having to COUNT instances. The SUM function is much faster than the COUNT function.

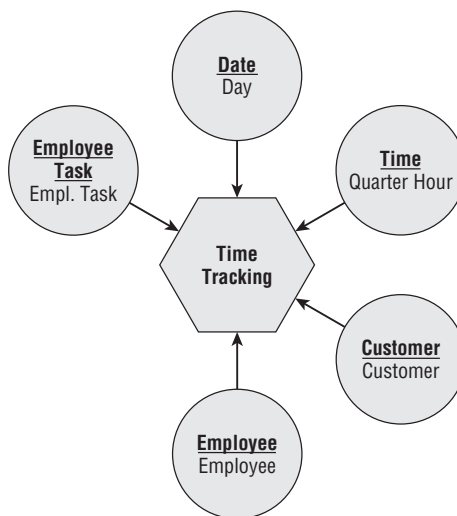
### *Call Center Time Tracking Fact Group*

The ability to understand how employees are spending their time is critical for planning resources. It is also important to be able to track patterns in performance compared to how time is spent. Perhaps more frequent, but shorter breaks make customer service representatives more productive. This is the type of information that needs to be learned by studying the data. To that

end, Figure 7-12 shows the Time Tracking fact group. These facts have some of the same dimensions as the Calls fact group, but the Task dimension is new. In addition, it is important to note that the employee time is not tracked at a minute level of detail but in quarter-hour increments. The facts themselves are defined in Table 7-10.

**Table 7-9** Call Center Calls Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Call Minutes	Total number of minutes spent on the phone with a customer	SUM
Wait Minutes	Number of minutes that the customer was put on hold before a customer service representative spoke to them.	SUM
Number of Call Transfers	Number of times that this call was transferred to another customer service representative. The first time a call is being handled, this is set to zero. Each subsequent transfer has this set to one. The Total Number of Transfers can be determined by summing the Number of Call Transfers grouped by the Call Transaction.	SUM
Number of Calls	Number of calls that were handled. This is set to one for each row at the lowest level.	SUM



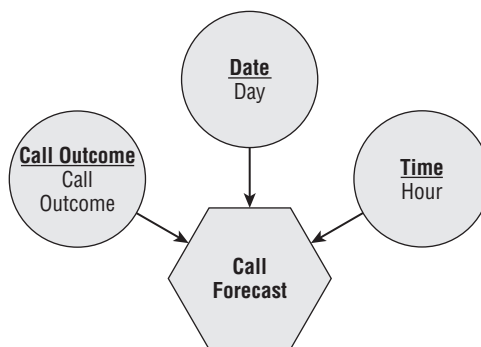
**Figure 7-12** Call Center Time Tracking fact group

**Table 7-10** Call Center Time Tracking Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Regular Hours Worked	Number of hours worked at this specific task at a regular or salaried pay rate. Time is tracked in quarter-hour increments.	SUM
Overtime Hours Worked	Number of hours worked at this task at an overtime pay rate	SUM

### Call Forecast Fact Group

In order to be able to schedule employees effectively, it is helpful to have an estimate of the expected call volume. The forecast is used to develop employee work schedules. Development of these work schedules is operational in nature. The call volume forecast is useful in the data warehouse to compare with actual call volumes. This enables fine-tuning of the forecasting process. Figure 7-13 shows the Call Forecast fact group. Note that the Time dimension is at the hour level of detail. Call volume is not forecast at the minute level of detail. Table 7-11 has the definitions for the Call Forecast facts.

**Figure 7-13** Call Center Call Forecast fact group**Table 7-11** Call Center Call FORECAST Fact Descriptions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Forecast Call Minutes	Expected number of minutes to be spent on the phone with customers	SUM
Forecast Number of Calls	Number of calls expected during this hour	SUM

## Working with the Model

Business analysts should work with the diagrams to explore how the model works. This helps ensure that the desired reports and analyses can be run. The following section walks through how to do this using the Call Center business dimensional model. Figure 7-14 shows a calendar monthly report of the number of calls by region. The calendar month perspective aligns with how the call center interfaces with the customers, and the number of calls fact is from the Calls fact group (refer to Figure 7-11).

REGION	CENTRAL REGION # of Calls											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
EAST	16,124	13,923	10,736	11,282	9,102	15,682	8,624	10,646	11,281	10,851	21,448	25,448
CENTRAL	12,459	10,759	8,296	8,718	7,034	12,118	6,664	8,226	8,717	8,385	16,573	19,664
SOUTHEAST	5,863	5,063	3,904	4,103	3,310	5,703	3,136	3,871	4,102	3,946	7,799	9,254
MOUNTAIN	28,583	24,682	19,032	20,000	16,136	27,800	15,288	18,872	19,998	19,236	38,021	45,112
WEST COAST	10,261	8,860	6,832	7,180	5,792	9,980	5,488	6,775	7,179	6,905	13,648	16,194
Total	73,291	63,288	48,799	51,283	41,374	71,283	39,201	48,390	51,278	49,322	97,489	115,672

**Figure 7-14** Sample Call Center Report

The report can be changed easily for an internal look by swapping out the Calendar Month for the Fiscal Month on the report. Both of these attributes are in the Call Center Date dimension (refer to Figure 7-3).

More detail can be pulled by drilling down on the customer regions to get a customer state. By looking at the Customer dimension in Figure 7-6, you can see that regions drill down into states. When you use the drill-down feature of your business intelligence tool, you will see this same report by Customer State.

These first two changes show how you can change a report by using other attributes within the dimensions that are already on the report. Reporting is not limited to these dimensions. Next, you can change the report to see the number of calls (the same fact) by employee education level, which is in the Employee dimension shown in Figure 7-7.

You can also change the fact on the report too. Using another fact from the same fact group, you can replace the number of calls with the call minutes fact from the Calls fact group (refer to Figure 7-11).

Finally, you can change the fact to one that is from a different fact group entirely. This is where you need to ensure that you have consistency between the dimensions that are on the report. In this case, you can change the fact to Regular Hours Worked from the Employee Time Tracking fact group shown in Figure 7-12.

This should help crystallize how dimensional models work. When the model is implemented, the business intelligence tools provide functionality to drill down and modify reports as shown in this example. This should also help you understand that the tool can only perform these functions if the data is organized in a manner to support it.

## Business Dimensional Model Index

A business dimensional model can include many different dimensions and fact groups. In order to help understand the contents of the model, a summary matrix can be created. Table 7-12 shows the index for the Call Center example. The columns represent each of the dimensions. Each row of the index represents a fact group. The body of the index shows where the fact group uses that dimension.

**Table 7-12** Call Center Business Dimensional Model Index

FACT GROUP NAME	DATE	TIME	CUSTOMER	EMPLOYEE	CALL	CALL OUTCOME	EMPLOYEE TASK
Calls	X	X	X	X	X	X	
Time Tracking	X			X			X
Call Forecast	X	X	X				

## Enterprise Considerations

There are often requirements to combine data from different parts of the organization. Dimensional models, when designed and implemented properly, can provide the capability to integrate data at query time. Often, the requirement to integrate data may not have been expressed during requirements gathering. This may be due to limitations of the current environment, so this is not even considered to be possible. This can also happen as the data warehouse grows over time. New data sources can be added that did not even exist when the original data warehouse was built. This section covers two fundamental concepts, conformed dimensions and conformed facts, which are essential to ensuring that all of the parts of the data warehouse work together.

### Conformed Dimensions

A *conformed dimension* is one that is shared between two or more fact tables. This enables integration between different fact tables at query time. This is a foundational principle that enables the longevity of a data warehousing environment. By using conformed dimensions, facts can be used together,

aligned along these common dimensions. The beauty of using conformed dimensions is that facts that were designed independently of each other, perhaps over a number of years, can be integrated using these conformed dimensions. An example of the conformed dimension concept can be seen through the Customer dimension. If a customer is assigned an identifier of 1000, then any reference to that customer is made using the identifier of 1000. This enables data about that customer from across the enterprise to be combined whenever it occurs, even if it was not a specific requirement for a data mart. This supports the ability to integrate data from any fact groups along their common dimensions at query time.

The biggest challenge in designing a conformed dimension is that it requires cooperation between different business groups. Too often there is a feeling that this will force everyone in the organization to a single rollup of the data and that only common characteristics can be included. This is not the case.

For example, consider that there are four different data elements to describe Customer Type. Analysis must be done to determine whether there is a legitimate business reason for each. Over time, implementation and enforcement of standards can become lax. Variations in the data can creep in. If three of the Customer Type data elements are truly supposed to represent the same thing, then the data should be cleaned up to reflect this; but if the fourth instance of Customer Type really is different, then this can be included. Each Customer Type must be labeled so that you can tell them apart. Table 7-13 shows sample values from these four data elements.

**Table 7-13** Customer Type Sample Values

CUSTOMER TYPE A	CUSTOMER TYPE B	CUSTOMER TYPE C	CUSTOMER TYPE D
Small Retail	SM RTL	Retail	Direct Bill
National Chain	LG CHN	L Chain	Payment on Delivery
Regional Chain	SM CHN	R Chain	Regular Invoice
Single Store	STORE	Single	Electronic Invoice

The first three may be melded into an attribute called Customer Type. The fourth should be renamed to Customer Payment Type.

If all of the attributes in a dimension are conformed, this by default ensures that all of the hierarchies are the same. However, if there are legitimate business reasons to represent different hierarchies, then each can be accommodated. All the attributes within each hierarchy should be uniquely named and clearly defined. This uniqueness must exist between all of the hierarchies, not just within each one. This can often be achieved by simply adding a qualifying word

such as “Sales” or “Finance” to each attribute within its respective hierarchy. This enables different groups to look at the business in their own way.

In addition to providing integration capabilities, another benefit of conformed dimensions is the sharing of attributes across the organization. Using the preceding example, the Customer Payment Type may have only been available to the billing group in the past. This may be an interesting way for the sales and marketing teams to look at Customer. This sharing can yield fresh insight.

Conformed dimensions should be created once and then shared as appropriate across the enterprise. This leverages development and maintenance efforts and helps ensure consistency.

## Conformed Facts

It is also necessary to ensure consistency of facts. A single fact should be defined once and remain consistent if it is used anywhere else in the enterprise. These are *conformed facts*. Again, if there are variances in meaning, then these are not the same fact and each should be uniquely named.

There are cases where the fact is intended to be the same but variations have crept in over time. This provides the opportunity to tighten up the data. Each fact should have a single definition.

**TIP** When possible, facts should be built and loaded once, and then shared with any other groups that need to access that fact. If, for performance reasons, the fact is loaded into more than one database, the definition and rules to create the fact must be strictly enforced to ensure consistency. The result should be the same regardless of which fact group, across the enterprise, is used.

## Practical Guidelines

There is no magic number for the right number of dimensions and fact groups you should have. The number of dimensions and fact groups will be determined by the industry you are in, the data that the organization has, and the specific characteristics of your organization. However, several general guidelines may be helpful.

### *Guidelines for a Single Dimension*

The more attributes that you have in a dimension, the richer the possible analyses. Whenever possible, seek out additional attributes. For example, the claims handling system may contain only a limited set of details about the policyholder. In order to create a complete and robust customer dimension, it would



be worthwhile to explore other sources such as a corporate customer master file or perhaps the policy administration system for additional customer data.

If you have many dimensions in your model, several are likely to have deep hierarchies and dozens of attributes, but you may also find that several dimensions have fewer than five, or perhaps just one attribute. Make sure you have looked for other sources, and explored additional possible attributes or hierarchies from the business community. If you can't find any, then this may in fact just be what you have. One other avenue to explore is whether this single attribute dimension really belongs within one of the other dimensions.

### ***Guidelines for a Single Fact Group***

In general, you are likely to find four to ten dimensions on any single fact group. Fewer than four dimensions may indicate that the dimensions are too big—with everything stuffed into a dimension. Too few dimensions may indicate that other dimensions still need to be identified.

You can also go overboard with too many dimensions. This can affect usability and perhaps impede the performance of the database. What is reasonable? It is common for transaction detail fact groups to have between ten and fifteen dimensions. Once you have more than fifteen dimensions, you should be concerned. Consider whether several of the dimensions should be combined. Over twenty dimensions and you could have some serious performance issues. If after further analysis you determine that there are in fact more than twenty dimensions, build extra time into your project schedule to explore the performance ramifications.

How can you tell which dimensions should and can be combined? It is important to look closely at the hierarchies. From a business perspective, you want to be able to follow a hierarchy from the top to the natural bottom. For example, the business may look at Regions, which are comprised of Districts and then Zones. Finally, the Zones are made up of Sales Territories. Each of these is part of the sales organization hierarchy. Suppose that Region, District, and Zone attributes are stored in one dimension, while the Sales Territory is in another. Walking through this example may be an indicator that perhaps Sales Territories belong in the first dimension with the other attributes. The best way to make this determination is to ask the business users what makes sense to them.

Classic dimensional analysis leverages drilling up or down hierarchies. When performing analysis against a design that has many dimensions, analysis will be done by leaving some dimensions out of the query. For example, for a travel reservation fact group, some people will want to understand which travel agents or web reservation systems are driving more business. Conversely, the business people who work with the hotel chains will want to understand more about the number and type of reservations by hotel brand and property location.

### ***Characteristics of the Model across the Enterprise***

As a data warehouse grows over time or a broad model is developed up front, there can be many dimensions and fact groups. It is not unheard of to have thirty or forty different dimensions as you look across an enterprise. You are likely to find is that a core set of dimensions is used repeatedly. These must be conformed! The other dimensions are typically used for fact groups that pertain to only one part of the business. Using an education example, the school district, student, and teacher dimensions will be used for many fact groups. Dimensions about student disciplinary actions, extracurricular activities, or teacher licensure pertain only to some of the fact groups.

**TIP** While enterprise-wide models may have over thirty dimensions and fact groups, keep in mind that any single fact group will have between five and fifteen dimensions.

Many organizations are already involved in data warehousing, perhaps even having multiple data warehouses. These have often been built without using conformed dimensions. By bringing the dimensions and facts together into a single model, this can be a good start toward understanding what will be needed to integrate these stovepipe databases.

Many enterprise-level dimensional models can be complex, because of the breadth that is included. Most users have an interest only in a subset of this model. The scope of what is presented to each group of users can be tailored to meet its own specific needs. Very few people in the organization may see or have a need to use the entire model.

#### **ENTERPRISE DESIGN GUIDELINES**

- Design broadly.
- Don't try to tackle all data for the enterprise at once.
- Design at the lowest level of detail that is captured.
- Work with multiple user groups who use the same data.
- Always consider an enterprise viewpoint—who else could use this data?
- Always design and plan for conformed dimensions.
- The dimensional model can be implemented in several increments.
- Each increment should deliver business value.
- Review and update the model with each increment.

The basic principles of dimensional modeling have been presented. Now it is time to take a practical look at how the model is created.

## **Business Participation in the Modeling Process**

---

This section outlines the steps involved in developing the business dimensional model. The description of the flow is geared more toward the project team, but it would still be valuable for any reader. The core project team, including the business analyst, must prepare for the modeling sessions and then develop a first draft. With subsequent refinements of the model, additional business representatives need to be included to get appropriate feedback. Finally, the model can be shared with anyone from the data warehouse target business audience.

### **Creating the First Draft**

One of the keys to getting the first cut of the model is having the right resources working together. The dimensional modeler is the team member who drives this entire process. The business analyst(s) who helped gather the business requirements is needed to ensure that the discussion is put into the proper context. The systems staff members who support and maintain the source systems are needed to help understand the incoming data elements. The final resource to include in the initial modeling sessions is the lead ETL developer. This person will be taking the model forward for implementation and is the individual most likely to take notes about any implementation-related issues that crop up during modeling. Key business people are not needed yet. Once an initial draft of the model is sketched out, then their assistance is needed. All of the participants should have at least an introductory knowledge of the parts of the dimensional model: facts and dimensions.

### **Preparing for Modeling Sessions**

In order to make the modeling sessions as productive as possible, the project team needs to do several things. If possible, block out a conference room for three straight days. As the model is drafted, the parts that have been diagrammed are posted on the walls. Changing conference rooms wastes a lot of time removing and putting the diagrams back on the walls. It also presents a big disruption in the continuity of thought, so try to stay in the same room if possible. The following checklist details the items you need for these sessions.

**MODELING CHECKLIST**

- **Business requirements document**
- **Data source documentation, data models and/or file layouts, and definitions of the data**
- **Flip charts**
- **Markers**
- **Tape, magnets, or pushpins to hang up flip chart pages (you could use self-sticking flip charts too)**
- **A spreadsheet with each source data element**
- **Electronic means for documenting the business dimensional model**
- **Connection to the network to look at live data content**

It is helpful to have three different people working electronically: one to document the diagrams, one to take notes for the ETL process, and one to look at the actual data and other source system references.

***Brainstorming the Framework***

Begin by simply starting a list of possible dimensions and fact groups. The intention is to get some rough ideas. Don't spend a lot of time debating whether or not something is a dimension or will simply be an attribute. Just capture the ideas. For the possible fact groups, think about the kinds of business processes you are working with. How do you measure them? What is tracked about those business events? It is helpful to think about the business itself, not what is tracked in the computer systems.

***Drafting the Initial Dimensions***

Now, each of the dimensions can be designed. It is often easiest to start with the Date dimension, which tends to be relatively straightforward and less controversial than other dimensions.

Once you have picked which dimension you want to work on, you can look for this type of data in the source systems. This is when the source system analysts are valuable. Identify the lowest level of detail that exists anywhere for that dimension. Next, add additional attributes and hierarchies. Look at the source system reference tables, and look at the source system transaction tables for more labels, descriptions, and groupings to include in the dimension. When a good first draft is done, post it on the wall in the conference room. Then, check the list for other possible dimensions to work on next.

It is common to continue to find attributes to add to dimensions that have already been drafted. Simply add the new attribute to the diagram and go back to what you were working on.

The objective is to get a first draft of all of the dimensions, not to get a perfect version before moving on to the next dimension. If the group has spent more than ten or fifteen minutes discussing or debating a set of data elements, then log the questions and move on. These issues will be reviewed and resolved later. If the team can't figure out what to do fairly quickly, then other people may be needed to provide more information or it may be too early in the modeling process. Later, it may seem immediately obvious how to handle a data element when the rest of the model is fleshed out.

Business dimensions should include attributes that contain complete values or descriptions. It is not acceptable to include only codes in the dimension. These codes are often well understood by people who currently access this data because more complete descriptions are not readily available. Over time, people have gotten used to these codes and know their meaning. However, it is difficult for new people to learn them. Many of the current users of the data have forgotten what it was like when they were trying to figure the data out. The objective is to empower people who have not had access to the data in the past. These people are not likely to know what these codes stand for. Useful business names and descriptions should be included in the dimensions. Codes can also be included in the dimensions as needed.

To help create a meaningful visual representation of the dimension, place the lowest-level attribute at the bottom center of the diagram. No other elements should appear lower than this element. The main or primary hierarchies should be placed up the center of the diagram. This shows their importance and the eye is drawn quickly to them. If other attributes describe similar or related things, then these should be diagrammed together. For example, if the Education Level Completion Date is needed, it should be placed near the Education Level attribute. Keep the attributes in the same general location with each revision of the model. This helps reduce confusion and keeps people from thinking that attributes have been removed. If there is a need to clean up a dimension, do this all at once, and if possible prior to presenting the model to a wide group of people.

### ***Drafting the Initial Fact Groups***

Once the first cut of all the dimensions is completed, it is time to take a look at the facts. Start with facts that are straightforward and well understood. Locate the transaction tables in the source systems where these facts are captured. By studying the context of these facts in the source system, you can see which dimensions will apply. The source system dependencies and keys are helpful to understand the data itself. After you understand the facts, you

can just work your way through each of the dimensions to decide which will apply. The dimension diagrams should still be hanging off the walls of the conference room to facilitate discussion. You must specify the level of detail, or grain, of each dimension for these specific facts.

**TIP** Remember: It is important to write down questions and not get bogged down during the sessions.

In order to avoid confusion, it is recommended that dimensions be placed in a consistent manner around the fact group name in the hexagon. Typically, start with the Date dimension and then place the others in a clockwise fashion. The order of the dimensions should be of relative importance to the business. The core dimensions, such as Customer and Product, should be after the Date dimension. Dimensions that are only used by a limited number of fact groups should be included around the ten o'clock position. Once the order has been established, keep it consistent for all fact group diagrams. This helps both the project team and the business users to quickly find what they are looking for.

It often takes three to five days to work through the dimensions and fact groups to get the first rough draft of the model. The time it takes depends upon the number of sources being used, the total number of data elements included in the model, the quality of documentation regarding the source data, and the experience of the modeling team.

### ***Documenting the Model***

Begin documenting the diagrams electronically as soon as possible. The longer you wait to begin this documentation, the harder it is and the less accurate it becomes. The optimal case is to have someone begin creating the diagrams during the modeling sessions. Depending upon how fast the model is developing, the documenter may be able to keep pace. If not, it is helpful to end the sessions with an hour left in the day to complete the documentation.

### ***Logging Questions and Issues***

It is also critical to log all of the questions electronically as they are noted. At the end of the day, these issues should be reviewed, and the descriptions and wording should be completed. Sometimes a quick note makes sense at the moment, but the meaning may be lost completely a week or even a few days later.

The issue log should include an issue number, a topic, a full description, the date, the person to whom the issue is assigned, the date the issue was closed, and the current status of the issue. At this point in the process, the issues do not need to be assigned. Over time, these will be assigned to

the appropriate people for follow-up. The discovery and resolution of each issue should be included in the issue log. This provides a record of the reasoning behind modeling decisions.

### ***Building the Business Measures Worksheet***

While developing the dimensional model, many different facts may be identified. Some are found as data elements that are stored in the source systems. Sometimes the facts are the result of a calculation. These calculated facts are called *derived facts*. Facts that can not be derived but need to be stored are called *base facts*. The complete set of facts, base and derived, is referred to as *business measures*. The base facts are typically what are included in the dimensional model and stored in the database. Derived facts are typically calculated on-the-fly, when a report is created.

It is helpful to begin to compile the list of all the business measures that will be needed while the dimensional model is being created. The list can be used to do the following:

- Review the derived facts to ensure that all of the base facts needed for these calculations have been included in the model.
- Determine whether a derived fact should be stored in the database, if the calculation is extremely complex or time consuming. It may be more efficient to perform the calculation during the ETL process.
- Set up the data access environment later, as described in Chapter 11.

Table 7-14 shows a sample of a business measures worksheet. The definition of base columns must match the definition of that fact in the fact group documentation. Note that the derived fact Average Minutes per Call must be recalculated each time. It is not mathematically valid to apply an average to a measure that is already an average.

These business measures can be reported by any attribute in any dimension that applies to the base facts used in the calculation. It is not necessary to document specific ways that each of these measures is to be reported. For example, Total Hours Worked can be reported by any attribute in the Date, Time, Customer, Employee, or Employee Task dimensions. The business dimensional model and the Business Measures Worksheet can be used to ensure that the required reports can be produced.

The business community is responsible for determining what business measures they need. Members of the project team may be able to do some background work to pull together a starter list of measures based upon reports that have been requested. The most time-consuming part of developing this worksheet is for the business community to agree on the official formulas that are to be used. It can take several months for different groups across the

business to reach a consensus on these calculations. There may be legitimate differences in how calculations are to be performed. If so, each different calculation can be included in the Business Measures Worksheet with a unique name. The description should also include details about when this version of a measure should be used rather than another.

**Table 7-14** Sample Business Measures Worksheet

<b>BUSINESS MEASURE NAME</b>	<b>BUSINESS MEASURE DESCRIPTION</b>	<b>MEASURE TYPE</b>	<b>AGGREGATION RULE</b>	<b>FORMULA</b>
Call Minutes	Total number of minutes spent on the phone with a customer	Base Fact	Sum	sum(Call Minutes)
Number of Calls	Number of calls that were handled	Base Fact	Sum	sum(Dollar Sales)
Average Minutes per Call	Average number of call minutes per call handled	Derived Fact	Recalculate	sum(Call Minutes)/ sum(Number of Calls)
Regular Hours Worked	Number of hours worked at this specific task at a regular or salaried pay rate. Time is tracked in quarter-hour increments.	Base Fact	Sum	sum(Regular Hours Worked)
Overtime Hours Worked	Number of hours worked at this task at a pay rate that is overtime	Base Fact	Sum	sum(Overtime Hours Worked)
Total Hours Worked	Total number of regular and overtime hours worked	Derived Fact	Sum	Regular Hours Worked + Overtime Hours Worked
Calls per Hours Worked	Ratio of the number of calls handled per every hour (regular and overtime) worked	Derived Fact	Recalculate	Calls Handled/(Total Hours Worked)



### ***Preliminary Source to Target Data Map***

When developing the business dimensional model, use a spreadsheet that contains all of the data elements from the source system that were identified during detailed data analysis. For each element that was included in the dimensional model, the name of the dimension or fact group should be noted. In addition, for those data elements that were determined to have no reporting or analytical value, this should also be noted. By keeping track of where these elements are placed in the dimensional model, this can be used as an early source-to-target data map to help with the ETL process, which is described in Chapter 10.

The final notes that need to be included are those about what additional processing or transformations will need to be done during implementation to get to the dimensional model representation. These are not the final business rules for processing, but notes about the discussions and perhaps ideas. The final business rules will be defined later in the project.

After all of the dimensions and fact groups that were identified during the initial brainstorming have been drafted, the first cut of the business dimensional model is done. There are still several weeks of work ahead, but the basics are in place and it should be clear which areas still need the most work.

### **Completing or Fleshing Out the Model**

Once the initial draft of the model is completed, some of the hardest work still lies ahead. The easiest and most straightforward data is already modeled. Now, the hardest and most complex data must be studied. The model moves toward completion by addressing issues, making sure that source data is represented properly, and having business group reviews.

### ***Working Through the Issues***

As you developed the initial draft of the model, each of the detailed questions and issues were logged. Now it is time to review these. Often, what seemed like a very difficult topic up front is actually easy to resolve. With the entire model laid out, the placement of these other data elements is often clear. The remaining issues tend to fall into three categories: those requiring technical research, those requiring business research, and those that need further consideration to see how they should be included in the model.

Often, nuanced knowledge of a source system database is only found with key individuals who have been working with that data for years. Many of the issues that need to be addressed require the insight of this experienced systems resource. One member of the core data warehouse project team must

be assigned the responsibility of working with the key systems people to get answers to these questions.

The second type of issue requires further business input. Again, someone from the project team must be assigned responsibility for following up on these issues. These can be the most challenging issues to resolve, and they often require the business community to examine differences between various parts of the organization (e.g., how business is accounted for literally in the general ledger) or other basic integration principles. This project may be the first time that different groups have needed to face these issues. Often, the problems have existed for years but have not been visible or have been ignored. Business policy decisions take much longer to reach resolution than the other types of issues you will work on. Resolving these issues can take a long time, and they require that the appropriate business participants fully understand the problem and then agree on a resolution.

The last type of issue has to do with modeling decisions. Usually the business needs and the data are understood. The challenge is reflecting this in the model. This may require research on modeling techniques or perhaps just taking the time to think about the problem.

As decisions are made or questions are clarified, this must all be documented in the issue log. This is where all of the work being done by each of the team members comes together. Topics that have been discussed at length but resolved can be included in the log as documentation for the future. This captures background information and what influenced important design decisions.

### ***Completing the Documentation***

A very important part of documenting the model is filling in the dimensional attribute definitions. This is time consuming, especially if the source system does not have good documentation. This is best done by project team members who understand the data and can express it in business terms. This is typically a representative from the business community.

### ***Working Through All the Data Elements***

As the issues are being worked on, the model will continue to be refined. As the model evolves, continue to track the placement of the source system data elements. This is needed to ensure that data is not inadvertently left out. Before the model can be deemed complete, each data element from the source systems must be placed in a dimension, a fact group, or designated as not included (i.e., these are being left out intentionally). Elements are left out when they not populated, obsolete, or are used for internal operations of the source system.

### ***Refining the Model***

As the project team works through the issues, the model can be updated regularly. This may be daily or twice a week. The model updates can be made and then shared with the rest of the team. The team must continue to alternate between researching the issues and updating the model. Before everything is finalized, but after the model has stabilized, it is important to begin getting business input.

### **Business Reviews of the Model**

It is important to get the true business perspective before the model is too far along. Although the core project team includes at least one business analyst, it is helpful to bring in one or two additional business people. These people should be carefully chosen and be patient, tolerant of change, and be willing to work with the project team. When including new people to review the dimensional model, it is important to consider their perspective. Take into consideration their exposure to the data warehouse project, their previous experience with systems design or data modeling, and any specific areas of interest. For example, business analysts for the claims department are not likely to have an in-depth interest in policy quotes or agent commissions. Focus the discussion for this group on the claims-related data. Ideas for presenting the dimensional model are highlighted in the accompanying sidebar text.

#### **IDEAS FOR PRESENTING THE MODEL**

- **Basic dimensional modeling concepts should be reviewed. The audience needs to have a basic understanding of dimensions and facts.**
- **It is helpful to present the model using an overhead projector and transparencies. While this is rather low tech, it does enable changes to the model to be seen by the entire group as they are made. Adding attributes to the diagrams can be tricky, and it takes a moment to find room, type the label, and draw the arrow. The point is not to test the team on their speed of diagramming in front of an audience, but to capture the ideas. This is fastest with transparencies and a pen. This low tech approach could be replaced with more high tech use of an electronic white board, if available. Be sure to test the capabilities prior to review session.**
- **A common business question can be used to show the parts of the model and how to use it. Walk through an example similar to that described in the call center case study, working with the model. As each part of the question is posed, show the appropriate dimension or fact group.**

*(continued)*

**IDEAS FOR PRESENTING THE MODEL (continued)**

- An example that pertains to your own business is the most effective way for the group to fully understand the concepts.
- Once the group has this introduction, it is time to get into more detail. Each of the dimensions needs to be presented and discussed. Then, the fact groups must be reviewed. This is painstaking, but necessary to ensure completeness and accuracy of the model.
- Share only the parts of the model of direct interest to the audience. This helps to ensure their interest and keep them from stepping out for a moment (often never to return).
- After looking through the model, get sample queries or report ideas from the group. You can work through each question by showing the pertinent dimension, identify which attribute would be used, and then look at the fact group. Sometimes you can start with a fact and then figure out what dimensions can be used. Engaging in this type of dialogue is what validates the model.
- You can maintain the electronic documentation, the diagrams, and the issue log during the session, but not on a machine with everyone watching.

***Small Business Reviews***

As the model is refined, it is helpful to bring in even more business people for small review sessions. These review sessions can focus on one area, perhaps only on one or two of the dimensions, and the focus can be on their part of the business. For example, for an insurance model, it can be helpful to review the claims parts of the model with a different group than the policy administration parts.

***When Are You Done?***

Continue working on the issue log, revising the model, and conducting business reviews until you have addressed all of the issues. As you work through the remaining open issues, the topics may be far outside of the scope of the initial implementation. It is perfectly fine to specify an issue to be addressed in the future, as long as the entire project team agrees, including the business representative(s).

When changes to the model are minor tweaks, rather than drastic revisions, the model is nearly complete. New business questions or challenges are easily met by walking through the model.

### ***Gaining Final Commitment***

The model should be presented to everyone who participated in the requirements gathering process. The meeting should be like the previous business review sessions, and minimal changes should result from this session. Each of the business people who participated in the development and refinement of the model need to play an active role in this session. It is much more effective to have business-to-business dialogue about the model than for systems people to feel they need to defend their work. If changes need to be made, this is still a good time to do so.

If the model does indeed represent the business perspective, the audience will be pleased, but perhaps not openly thrilled. This should be viewed as success! It means that the model makes sense to them, which is the goal.

### **Expanding Business Data Over Time**

Dimensional models do indeed change and expand over time. This happens as a natural course of business. There are two common ways for dimensional models to expand: enhancing dimensions and adding new facts.

#### ***Enhancing Dimensions***

As more, newer, or cleaner sources become available, new attributes can be added to existing dimensions. As long as these new attributes do not change the grain of that dimension, there is no impact on existing fact groups. The business community gains the ability to create new reports using any of the existing facts.

Dimensions can be fairly easily rebuilt, as long as the dimension does not have 50 million instances. If the dimension is extremely large, this can still be done, but it requires some additional planning and time. Businesses that deal with these extreme volumes are well versed in dealing with this much data, as it pervades all systems development, not just data warehousing.

#### ***Adding More Fact Groups***

Another common way that dimensional models expand is with the addition of new fact groups. Usually this happens when new facts, often from new data sources, are loaded. These new facts must be studied to determine their dimensionality and grain. The new facts may be added to existing fact groups as long as they have the same dimensionality and grain, although the technical team may decide to implement these separately. New fact groups can be added without impacting or changing any of the existing fact groups. Existing users of the data warehouse may not see any change. The new data can be released to only those who need access to it. Of course, all dimensions used by the existing and new fact groups must be conformed.

## **Reflecting on Business Realities: Advanced Concepts**

---

For many people, the basic dimensional modeling concepts presented so far will be sufficient for their needs. For those who want to learn about dimensional modeling in more detail or who are about to participate in the modeling process, several additional topics are included here. This is not intended to make you a dimensional modeler, but it describes techniques that you can use to address many common, complicated situations. The business dimensional model notation to identify these situations is presented. This notation provides the technical team with an indication that a condition exists. Standard dimensional modeling techniques can then be used to complete the actual database design.

### **Supporting Multiple Perspectives: Multiple Hierarchies**

It is a common misconception that all groups within the organization must agree on the structure and content of one hierarchy. Not only is this not true, but multiple hierarchies are encouraged because this enables multiple viewpoints to be represented in the model. In the Date dimension presented earlier (refer to Figure 7-3), note that there are two hierarchies: one to support the calendar year rollup and one to support the fiscal year rollup.

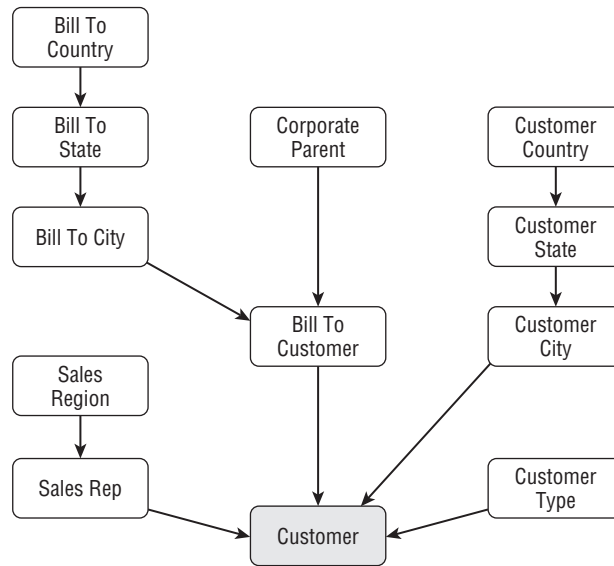
Another example of the benefits of multiple perspectives can be seen for customers. In this example, the “ship to” location is what is called Customer. Figure 7-15 shows multiple customer hierarchies: the customer’s location, the customer’s bill to location and the company’s sales organization hierarchies. Each hierarchy is diagrammed separately in the business dimensional model so that each is clearly visible. It is important to ensure that each is labeled uniquely.

### **Tracking Changes in the Dimension: Slowly Changing Dimensions**

Most of the data in the data warehouse that is tracked over time are facts. This is a natural and basic part of dimensional modeling. However, there will be changes to values in dimensions that need to be addressed. There are also cases where the dimensional attributes must be captured at the time that a business event occurred. There are technical techniques to address tracking these dimension changes over time.

From a business perspective, you need to evaluate which changes must be recorded and kept. If a person’s name changes, is that important for analytical results? Perhaps not. In many cases, the current value of an attribute

is sufficient. In other cases, and for some industries, these changes must be tracked over time. For example, the characteristics of an insurance policy must be kept as they were at the time of an accident to accurately track and understand claims data. It is not correct to show last year's claim with this year's policy characteristics. This is an example of capturing the dimensional data at the time that a business event occurs. The business dimensional model can note when the dimensional data must be tracked over time.

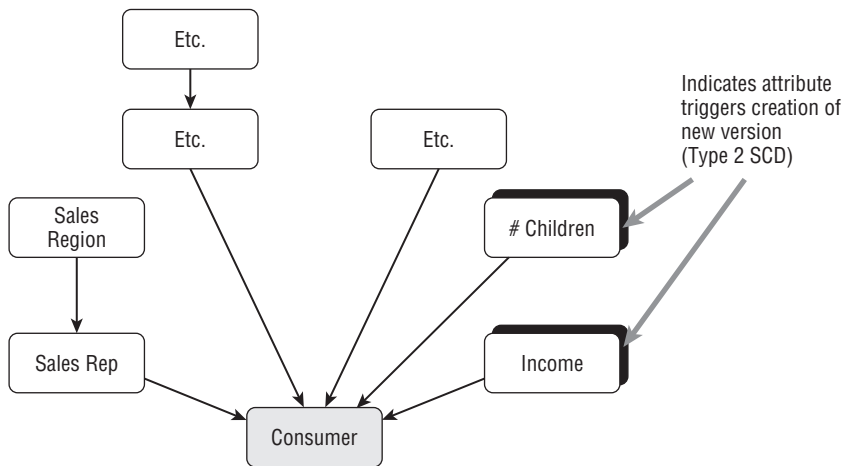


**Figure 7-15** Multiple hierarchies in a Customer dimension

*Slowly changing dimensions* is the terminology that is used in the data warehousing industry for dealing with these dimension attributes. There are several different ways to address these changes, depending upon the business requirements. The business dimensional model has notation to assist with this.

By default, all dimensional attributes are expected to contain only the current value. This is called a Type 1 slowly changing dimension. Tracking point-in-time history is called a Type 2 slowly changing dimension and is diagrammed differently. The attribute is shown with a shadow that appears as if there are versions. Figure 7-16 shows an example of a dimension that requires versioning. This technique creates a version of that dimension data at each point in time of interest.

To better define these requirements, the business must decide which attribute changes should trigger a new version. If there are 100 attributes of an individual policyholder, is a new version needed when any of these change? Perhaps you only need a new version when an attribute that is used as a policy rating factor changes, such as marital status or student grade point average.



**Figure 7-16** Slowly changing dimension notation

The business must consider that there is a cost associated with this versioning. There should be a balance between implementation complexity, cost to implement, and the business value of doing this. That's not to say this should be avoided, but it should be approached practically.

**TIP** For more information about the detailed implementation options for slowly changing dimensions, please refer to *The Data Warehouse Toolkit, Second Edition* (Wiley, 2002).

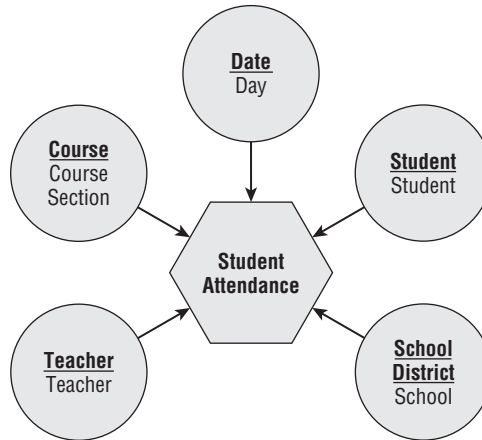
## Depicting the Existence of a Relationship: Factless Fact Tables

There are several cases where the relationship between the dimensions is important but there is no measurable fact. This is called a *factless fact group* or *table* in dimensional modeling. One example of a factless fact group is tracking daily student attendance, as shown in Figure 7-17. What is tracked is whether a student was in class (described as a course section) with a teacher at a school on a specific day. The existence of the event is what is important, but there is no way at present to measure a student's level of attention. We are not able to say that a student was only 63% attentive, so no fact is needed. The design team may decide to include an artificial fact. This would be the fact of attendance, where the value is always 1.

Other common factless fact groups include the following:

- **Product coverage:** What products are in a store on promotion? This is needed to determine which products were promoted but did not sell.
- **Approved vendors:** What products can be purchased from which vendors?





**Figure 7-17** Factless fact group

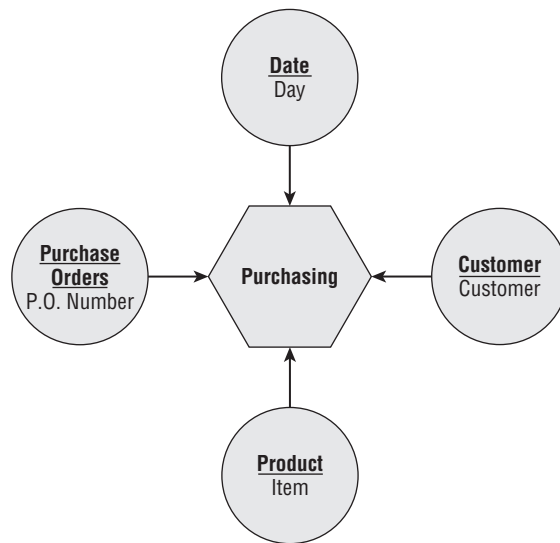
Of course, your business may have facts that would be included in these fact groups. Just because these are included as examples here does not mean that they must be factless.

**NOTE** A factless fact table is a complex associative entity from other data modeling techniques. For data warehousing, this relationship is tracked over time, so there is a need for a date too. While there is not a specific fact coming from the source system, it can be helpful to include an artificial fact, which is always set to 1. This can ensure that the business intelligence tools properly interact with this table.

## Linking Parts of a Transaction: Degenerate Dimensions

Sometimes all of the interesting attributes to describe a business transaction are placed into other dimensions but there is still one important piece of data left over: the transaction identifier. The *degenerate dimension* is not the same as a single attribute dimension. The degenerate dimension value is an identifier for each specific transaction. A single attribute dimension would contain a list of possible values that further describe the transaction, such as a transaction type—for example, containing the values of open claim, closed claim, or claim payment.

This transaction identifier must be retained in order to be able to pull together the different parts of a single business transaction. For example, to implement purchase order line-item detail, you still need to know that all of the items belong to the same purchase order. Because this is an important piece of data but is not a fact, this must be considered a dimension. However, this special case is called a degenerate dimension. Figure 7-18 shows the fact group that includes the Purchase Order dimension.



**Figure 7-18** Purchasing fact group with Purchase Order degenerate dimension

The degenerate dimension may be eliminated from the Fact Group Diagram and listed with the facts themselves in the Fact Definition Table, as shown in Table 7-15.

**Table 7-15** Purchase Order Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Cases Ordered	Number of cases of the item that were ordered	SUM
Sales Dollars	Dollar amount that the customer is expected to pay for the cases ordered	SUM
P.O. Number	Purchase order number for this customer order	Degenerate Dimension

Other situations that often result in a degenerate dimension include the following:

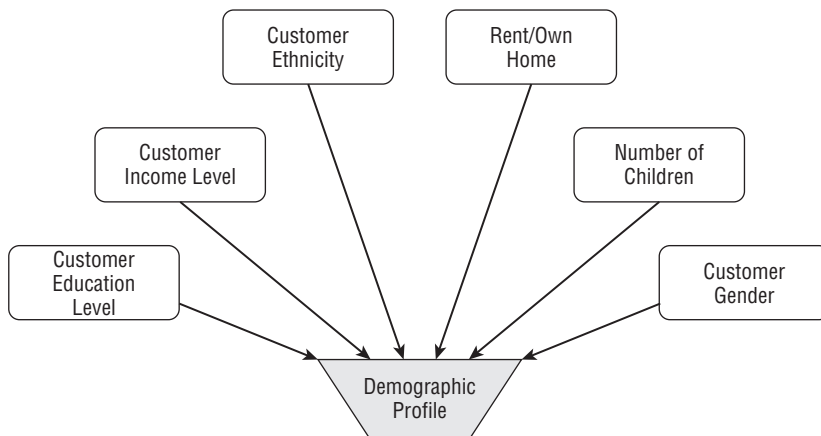
- **Retail sales checkout:** Again, you need to be able to track the specific items that were purchased at the same time. The ticket number is needed to support market basket analysis, to see what products are purchased together.
- **Call Center example:** Earlier in the chapter, the Call Transaction dimension is degenerate. This enables the tracking of calls even if multiple

call outcomes are possible, such as placing an order or asking about an existing order. This also helps track a single customer call if the call must be transferred to another customer service representative to address the customer's needs.

## Pulling Together Components: Junk Dimensions

In some instances, several single attributes don't really fit into another dimension. Sometimes this happens when these attributes relate in a general way but do not have a single lowest-level attribute to define the grain of a new dimension. Sometimes, you simply find yourself with a handful of attributes "left over" and need to find a home for them. These can be grouped together into a single dimension called a *junk dimension*. A new attribute must be introduced to provide a single attribute to place at the lowest level of this junk dimension. This new attribute is called a *derived attribute*, and it is created solely for the data warehouse. This derived attribute will not be found in any of the underlying source systems.

An example of a junk dimension is diagrammed in Figure 7-19. The derived attribute is shown as a trapezoid. A customer demographic profile can be designed as a junk dimension.

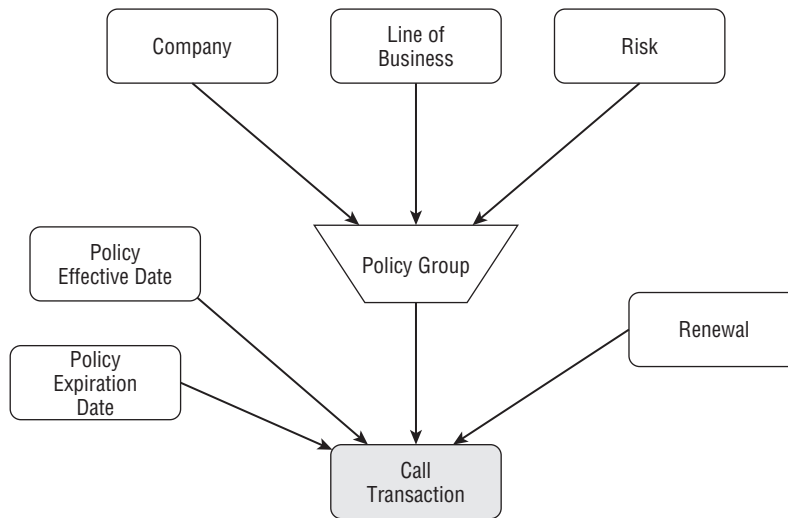


**Figure 7-19** Customer demographic junk dimension diagram

The relationship between the customer demographic profile and the customer is captured via a fact table. The relationship itself can be designed as a factless fact table, with the customer, customer demographics, and date that this relationship was observed. This could also be recorded as part of a sales transaction via a Sales fact group.

**NOTE** A derived attribute may be stored at the physical level but not displayed to the business users.

A derived attribute can also be created to tie together parts within a dimension in order to facilitate reporting. This is useful when, for example, you have three important identifiers that are used alone and together such as Line of Business, Company, and Risk for an insurance policy dimension. Figure 7-20 shows this derived attribute within the body of a dimension.

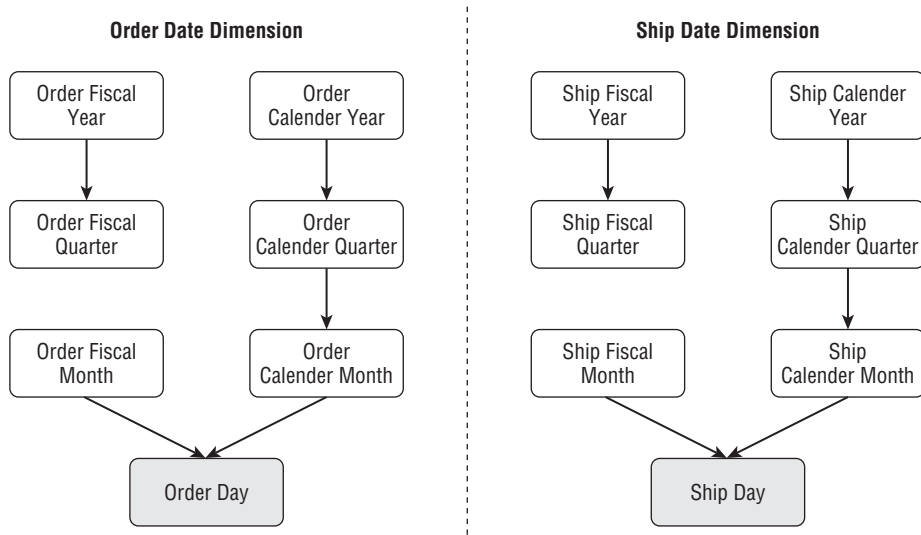


**Figure 7-20** Derived attribute within the Policy dimension

## Multiple Instances of a Dimension: Role Playing

There are many transactions whereby what originally seems like a single dimension occurs multiple times for a fact group. This must be fully explored to ensure that this is really the case. If so, each needs to be identified for query and analysis purposes as a separate role for that dimension. In order to minimize confusion, each attribute must be uniquely labeled. Typically, this means that each attribute must have a prefix to clarify which role is being represented.

One of the most common instances of role playing is for the Date dimension. Is it common for a single business transaction—say, for a product shipment—to have multiple dates. These would include the product order date and the product ship date. Figure 7-21 shows the two separate dimensions, whereas Figure 7-22 shows that each role of the dimension is included in the fact group diagram.



**Figure 7-21** Sample role-playing dimension diagrams

**TIP** Develop one instance of a role-playing dimension at first. Note each role as it is identified. After the model has been reviewed and is stable, then copy the dimension and rename each attribute with the appropriate prefix.

The role-playing dimensions must also be represented in the fact group diagrams. Figure 7-22 shows both the Order Date and Ship Date dimensions as they relate to the Shipments fact group.



**Figure 7-22** Fact group for role-playing example

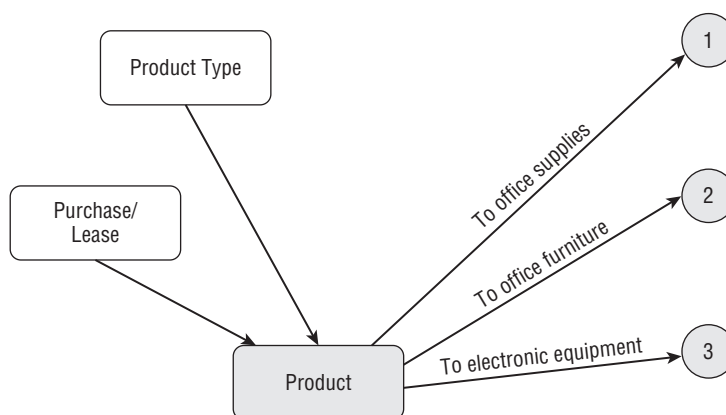
Another example of a role-playing dimension might be a school district that is responsible for educating a student. There is a district, usually where the student resides, that is legally responsible for educating that student. In some cases, a student is actually receiving instruction from a different school district, perhaps for special services or because a parent works for the other district.

## Other Notation

The business dimensional model should be easily printed on 8½'' by 11'' paper. By storing the diagrams and tables in a word processing document, many users can print it without requiring special software. A few additional notations can be used to make the model easier to read.

### Dimension Connectors

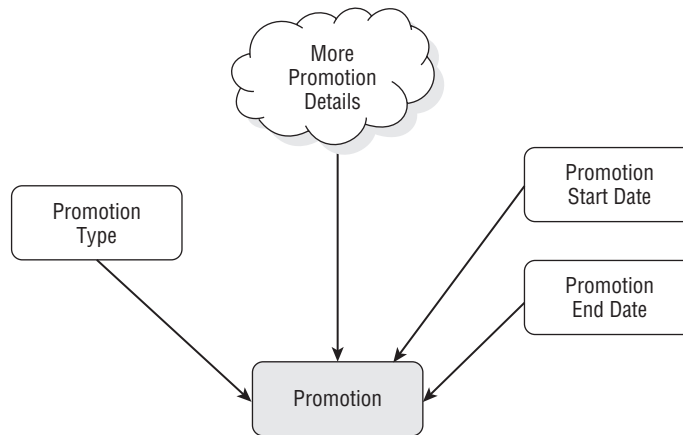
If you have a dimension that has too many attributes to fit on a single page, you can diagram the dimension across multiple pages. The best way to do this is to include the basic, most important attributes on the main page. Then use a connector, as shown in Figure 7-23, to link to the other pages. Try to group the attributes on each page based upon their meaning. For example, if you are trying to represent a dimension for all the different products that are purchased by a large enterprise, you might group the attributes by office supplies, office furniture, and electronic equipment. There would then be a separate page with the attributes specific to office supplies, office furniture, and electronic equipment.



**Figure 7-23** Dimension connector example

### Clusters of Future Attributes

The notation for an attribute that is needed in the future is a dotted rectangle. This is effective when specific attributes are defined. However, in some cases many possible attributes could be helpful but they are not yet defined. This can occur when the source of data does not exist yet or the source is completely out of scope. It may still be helpful to capture the team's ideas about where these would fit in the future. Figure 7-24 shows a cloud to represent this cluster of future attributes.

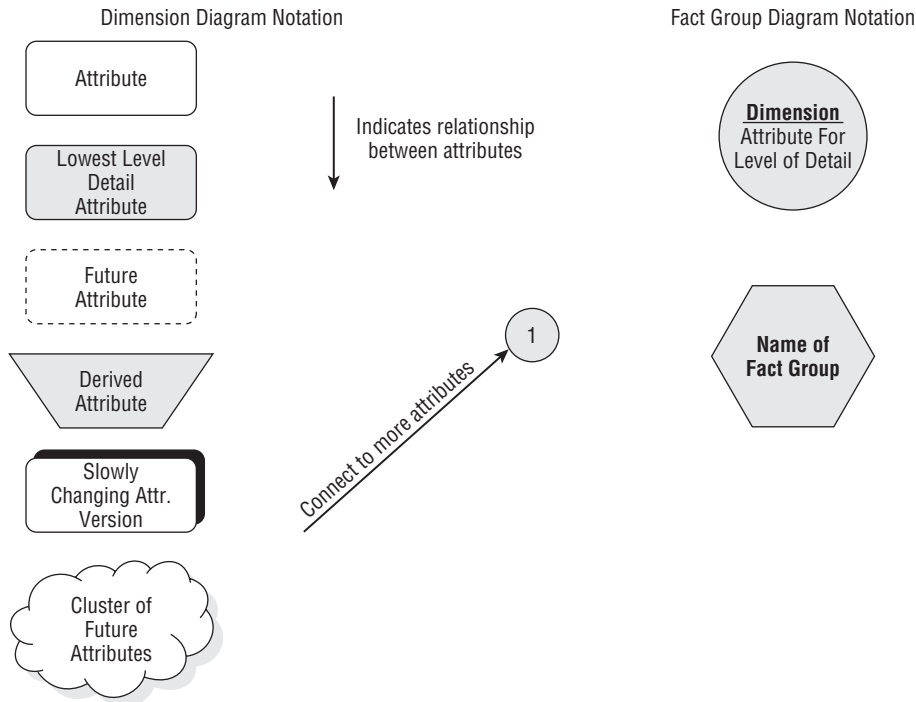


**Figure 7-24** Cluster of future attributes

If these future attributes are needed to support the core purpose of your project, you have a serious problem that needs to be addressed immediately. Otherwise, noting the placement of these attributes gives you a great head start when these are fully defined in the future. It also increases the confidence of users that you haven't forgotten their ideas.

### Notation Summary

Figure 7-25 shows a summary of the notation for the business dimensional model.



**Figure 7-25** Notation summary

## Taking the Model Forward

Now that the business dimensional model has been reviewed and approved by the business community, it is time to move forward into more traditional systems activities. While this is not intended to provide full technical detail of such activities, it will highlight the steps needed to take the model forward. This section is intended to assist the project team in understanding what happens next on the project.

### Translating the Business Dimensional Model

The business dimensional model can be implemented in a variety of technologies. If all of the data will be loaded into a multi-dimensional cube or other proprietary database, then no logical table design is needed, but it may be necessary to decide at which levels of detail you want to store the facts. If the business dimensional model is to be implemented in a typical relational database, then additional design decisions must be made. The business dimensional model must be translated into logical table structures. The next



section describes this process. This is straightforward, as the biggest challenge, understanding the business, has already been addressed.

### ***Dimension Table Design***

Each business dimension will become a single dimension table, and each dimension table should have a single surrogate key. This is a unique non-meaningful identifier that is assigned to each row in that dimension. Each business attribute is represented as a column on that dimension table. In some cases, a business attribute can be translated into several columns. An attribute of Vendor may be included on the table as Vendor Code, Vendor Short Name, and Vendor Name. Each of these may be represented in the business dimensional model, but these are really just different labels for the same true business attribute. If it helps reduce confusion, each should be included in the model diagrams. If there are too many attributes for that dimension (some can have over fifty), then it may clean things up to only show the business attribute once and introduce these additional details in the table design.

Using the call center case study from earlier in the chapter, Figure 7-26 shows what the Employee dimension would look like as a table stored in a relational database.

EMPLOYEE KEY
Employee ID
Employee Name
Current Hire Date
Original Hire Date
Employee Date of Birth
Employee Gender
Employee Education Level
Employee Bilingual Indicator
Supervisory Role
Salary/Hourly
Employee Home State
Employee Home City
Employee Home County
Employee Home Zip Code

**Figure 7-26** Call Center Employee Dimension Table

### Translating Fact Groups

A fact group often translates to a fact table. However, from an implementation perspective, you may decide to split a fact group into more than one physical fact table. This is typically done if the facts come from different sources or the timing of the data's availability differs. In these cases, splitting the fact group may offer efficiencies during the staging process. Regardless of how the facts are physically stored, the first thing to understand is the dimensionality and grain.

Each fact group will translate to a fact table. The primary key from each dimension that applies to the fact group will be stored on the fact table. After the keys, each fact becomes a column on the fact table. Figure 7-27 shows the Call Center Calls fact group translated into the logical table.

DATE KEY
TIME KEY
CUSTOMER KEY
EMPLOYEE KEY
CALL TRANSACTION KEY
CALL OUTCOME KEY
Call Minutes
Wait Minutes
Number of Call Transactions
Number of Calls

**Figure 7-27** Call Center Calls Fact Table

**NOTE** A primary key is a unique identifier of a row in a relational table. When the primary key is used as a link to another table, it is called a *foreign key* in the second table.

The logical key of the fact table is comprised of all the foreign keys from the dimension tables included in that fact group. For the Call Center fact table (refer to Figure 7-27), the logical key is comprised of the Date Key, Time Key, Customer Key, Employee Key, Call Transaction Key, and Call Outcome Key. In general, there is no need to create a separate surrogate key for the fact table. This would not be used for processing queries.

### Physical Database Design

Minimal design work is needed to move between a logical design and a physical database design. This is typically the step where performance considerations are taken into account. Because a dimensional model addresses

query performance as one primary purpose of the design approach, not much is needed other than some physical decisions such as physical partitioning and indexing strategies.

## **In Real Life**

---

There are many challenges in creating a dimensional model. Some of the common ones can be observed in the big and small company groups we have been following.

### **A Glimpse into Giant Co.**

In this organization, members of the enterprise modeling team do most of the modeling. The modelers are responsible for all databases, be they operational or data warehouses. Most of their experience is with operational system design. In addition, the modeling effort is being done without participation from the business community. The business was involved in providing requirements, but the interaction between them and IT ended there. Because initial involvement is all that is expected from them for other systems design, that is all that was expected of them for the DW project.

The challenge is that the modelers are not well versed in thinking dimensionally; and with no direct business interaction, many of the basic dimensional modeling principles may be overlooked or ignored. The final data model is to be presented to several select representatives from the business community. This is mostly a formality and few changes are expected to result from this session.

To overcome these challenges, the process to develop a data model was refined to accommodate dimensional modeling best practices. This includes steps to ensure participation of business representatives during the modeling process and for reviews. This required the executive business sponsor to support the data warehouse project team's effort to get these changes put in place. Now the enterprise modeling procedures include data warehouse-specific steps. This helps not only this project, but also future projects too.

### **Insight from Agile, Inc.**

At this company, everyone is used to jumping right in to get things done. In fact, that is what happened during the first attempt to build a data mart. The team gathered a few reports that were needed and turned them over to IT. There was a quick database design effort, but not really a full data modeling effort, and the design was not documented. Construction began, but as each report was studied in more detail, this required changes to the data structures

and modifications to the ETL processing. This happened repeatedly, which frustrated the project team members and the business group waiting for their reports.

After weeks of running in circles, the project team took a moment to review what was happening. It was determined that too many resources and time had been wasted already and a long-term solution was not in sight. The project team enlisted the help of the business champion to refine the project plan by going back to develop a robust data model. This required invoking the change control process. Key business analysts were assigned to help with the modeling process. The model was successfully completed and well received by the rest of the business community. Everyone is anxious for the database to be built so that they can begin to use it.

## Summary

---

Developing a sound data model up front sets the stage for success. Dimensional modeling helps to organize data for reporting and analysis, often directly by the business audience. In order to facilitate business use, it is important that the model reflect the business perspective. The business dimensional model diagrams help to capture this perspective. A proven process to develop the dimensional model was also presented. The dimensional model can be implemented in a variety of different technologies.

Useful information is uncovered during the modeling process and can be used to bridge to subsequent stages of a data warehouse project. The preliminary source-to-target data map captures details that will be needed to develop the ETL system. Similarly, the Business Measures Worksheet content is needed to set up the data delivery. Before discussing these subsequent steps, the next chapter spends a bit more time looking at data-related topics, including information management, data governance, and data quality.

# Managing Data As a Corporate Asset

“Data as a corporate asset”—what does that really mean? In simple terms, any asset a business owns must contribute to the success of that business or it is not really an asset. Many businesses have assets on their books that are not making a positive impact on the success of the business. If you add up the total investment in data, including operational systems to run your business, business intelligence systems to manage your business, staff to build, maintain, and use these systems, it will probably be one of the largest assets on your books. Is all this data you collect really working for you? Is it really a corporate asset? How can you turn data into a real asset?

This chapter is an introduction to *information management* concepts and principles, which provide the mechanism to turn data into a corporate asset. This is accomplished through strong partnerships between business and IT communities and corporate commitment. The topics covered here include the following:

- Understanding master data management
- Data governance
- Understanding data ownership
- What “high data quality” means and how to get it
- Ideas for setting up a data dictionary
- Practical steps to get started with information management

These concepts are briefly introduced in this chapter, but are not intended to provide a comprehensive guide. Many other resources are available for any readers who are interested in learning more about these concepts and how to develop and execute a comprehensive information management strategy.

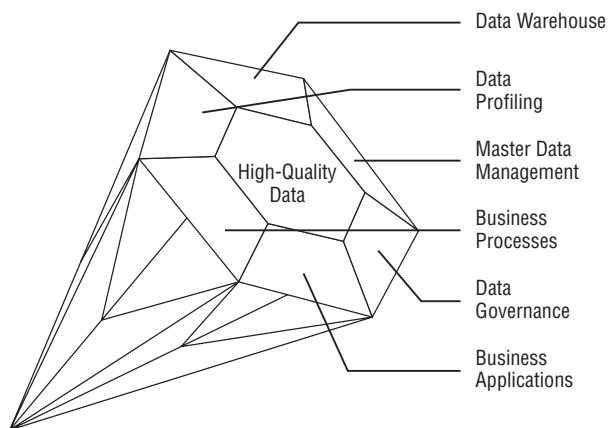
## What Is Information Management?

Data is what you collect and store. When that data is interpreted and used, it is turned into information. Companies that can effectively transform data into valuable information that is used to make business decisions are utilizing their data as an asset. A computer can be considered a corporate asset, yet it does not provide value until it is used; the same is true for your data. If you look at a group of competing companies, chances are good that the ones that have figured out how to turn data into valuable information and act on it are at the top of their industry.

Sometimes the terms and acronyms used across business and IT can be a bit overwhelming. Most terms can have multiple definitions depending on the context and scope of the discussion. Information management is no different. For this discussion, we'll use Wikipedia's definition:

*Information management (IM) is the collection and management of information from one or more sources and the distribution of that information to one or more audiences. This sometimes involves those who have a stake in, or a right to, that information. Management means the organization of and control over the structure, processing, and delivery of information.*

The first step in information management is focusing on the data itself. Data is the fundamental building block of the entire process, and the core asset. However, information management poses many challenges beyond simply sharing data between application systems. Information management also encompasses the people and business processes that are needed to ensure that the data is collected and handled properly. Figure 8-1 shows these different facets of information management, which include the organization's business applications, the data warehouse, business processes, data governance, master data management, data profiling, and high-quality data.



**Figure 8-1** Facets of information management

In order to facilitate understanding of the entire chapter, it is worth looking at basic definitions of these now. Again, due to the prevalence of existing terminology, the following definitions are shared via Wikipedia:

- **Master data management (MDM):** In computing, master data management comprises a set of processes and tools that consistently defines and manages the nontransactional data entities of an organization (also called *reference data*). MDM has the objective of providing processes for collecting, aggregating, matching, consolidating, quality-assuring, persisting, and distributing such data throughout an organization to ensure consistency and control in the ongoing maintenance and application use of this information. At a basic level, MDM seeks to ensure that an organization does not use multiple (potentially inconsistent) versions of the same master data in different parts of its operations, which can occur in large organizations.

- **Data governance:** This refers to a quality control discipline for assessing, managing, using, improving, monitoring, maintaining, and protecting organizational information<sup>1</sup>. It is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models that describe who can take what actions with what information, and when, under what circumstances, using what methods<sup>2</sup>.

Data governance encompasses the people, processes, and information technology required to create consistent and proper handling of an organization's data across the business enterprise, including the following goals:

- Increasing consistency and confidence in decision making
- Decreasing the risk of regulatory fines
- Improving data security
- Maximizing the income generation potential of data
- Designating accountability for information quality

These goals are realized by the implementation of data governance programs, or initiatives. Data governance initiatives improve data quality by assigning a team that will be responsible for the data's accuracy, accessibility, consistency, and completeness, among other metrics. This team usually consists of executive leadership, project management, line-of-business managers, and data stewards.

<sup>1</sup>IBM Data Governance. <http://www-306.ibm.com/software/tivoli/governance/servicemanagement/data-governance.html>. Retrieved on 2008-07-09.

<sup>2</sup>Data Governance Institute Data Governance Framework

- **Data quality:** Data is of high quality “if [it is] fit for [its] intended uses in operations, decision making, and planning” (J. M. Juran). Alternatively, data is deemed of high quality if it correctly represents the real-world construct to which it refers. These two views can often be in disagreement, even about the same set of data used for the same purpose.
- Data quality refers to the degree of excellence exhibited by the data in relation to the portrayal of the actual phenomena (GIS Glossary).
- The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use (Government of British Columbia).
- Data quality: The processes and technologies involved in ensuring the conformance of data values to business requirements and acceptance criteria.
- **Data profiling:** This is the process of examining the data available in an existing data source (e.g., a database or a file) and collecting statistics and information about that data. The purpose of these statistics may include the following:
  - Determine whether the existing data can easily be used for other purposes.
  - Provide metrics on data quality, including whether the data conforms to company standards.
  - Assess the risk involved in integrating data for new applications, including the challenges of joins.
  - Track data quality.
  - Assess whether metadata accurately describes the actual values in the source database.
  - Understand data challenges early in any data-intensive project in order to avoid late project surprises. Finding data problems late in the project can incur time delays and project cost overruns.
  - Have an enterprise view of all data, for uses such as master data management, for which key data is needed, or data governance, for improving data quality.

Some companies also look at data profiling as a way to involve business users in what traditionally has been an IT function. Line-of-business users can often provide context about the data, giving meaning to columns of data that are poorly defined by metadata and documentation.

Beyond simply working with the data, information management must also address the organization’s business processes. Business processes come



in a variety of flavors, from understanding how the business itself runs to processes that are added to specifically address the capture and maintenance of data needed for operational purposes and analytical purposes. Some examples of challenges that can affect business processes and impact information management include the following:

- Understanding and defining which group is responsible for ensuring that data about new customers is collected and entered properly
- Procedures regarding how changes to existing products are handled, including when a new product number is assigned, and who is responsible for updating the characteristics of the product packaging
- Procedures to ensure consistent and accurate data capture at the time of capture
- Handling and reporting of data errors. This can be used to understand gaps exist between what the system is expecting and what is being provided. This may require system changes or education for the people entering the data.
- Systematically checking the quality of the mailing addresses

These are just a few examples of things that can affect the business processes and impact information management. They are described in more detail in the sections on data quality and data governance.

Delivering high-quality data should be the primary goal of information management processes. The design of information management processes depends on a number of factors, including the following:

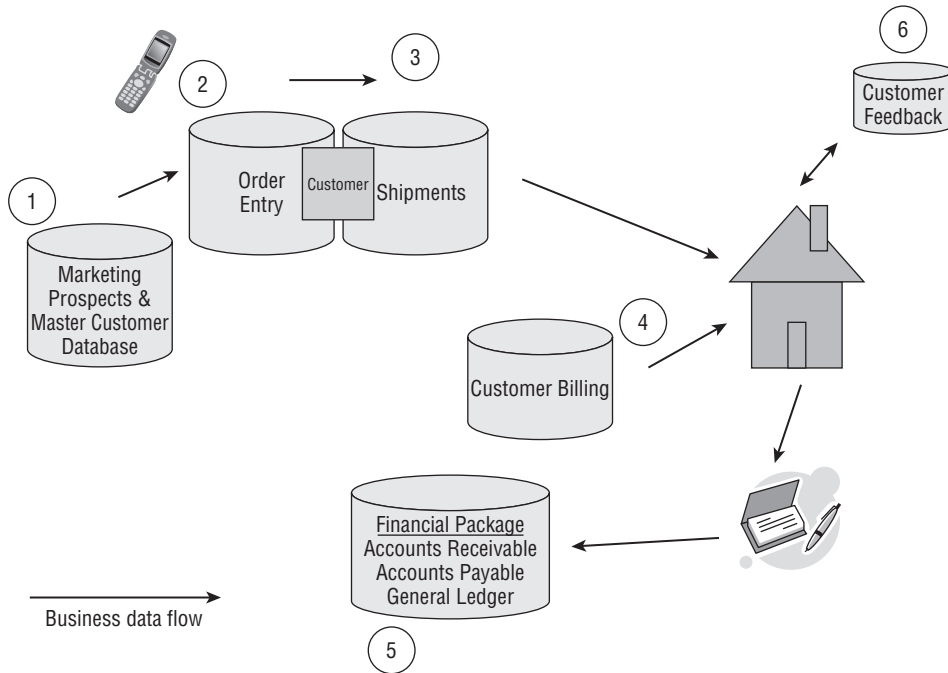
- Size of your organization
- Complexity of your business
- Quality of your source systems
- Current business processes
- Political nature of your organization

The next section shares an example of information management in action to facilitate an understanding from a data perspective.

## Information Management Example—Customer Data

In order to highlight how data should be managed, this section looks at an example of how a consumer direct retail organization can move from not having an integrated information management strategy to an environment with efficient information management processes. This example starts by describing how customer data is currently being handled, followed by ways

to improve. Figure 8-2 shows the original state of a logical business flow of customer data from consumer marketing through customer payments. This is a business pictorial of the data flow, and is not intended to represent a formal process flow diagram. It highlights the capture and storage of customer data.



**Figure 8-2** Non-integrated customer data flow

First, a promotional mailing is sent out to select customers and prospects using the Marketing Prospect and Master Customer Database. This generates orders, as shown in step two. Customer data is entered into the order entry system. The first step in taking an order is capturing the customer's name. If there are not too many customers waiting on hold, an initial search of the customers in the order entry system is done. The search screens are cumbersome and slow. If a name is not found or too many are found, the customer service representatives often enter this as a new customer. It is well known that there are duplicate entries of this data, but that is ignored so the order can be quickly captured, the product delivered, and more calls handled.

This same customer table is also used by the Shipments application in step 3. Once a shipment is made, the customer is billed for the purchased products, in step 4. Basic customer data is fed to the Customer Billing application, but a separate copy is maintained. There is no real business purpose for keeping a separate copy; this is simply how the applications have been developed over

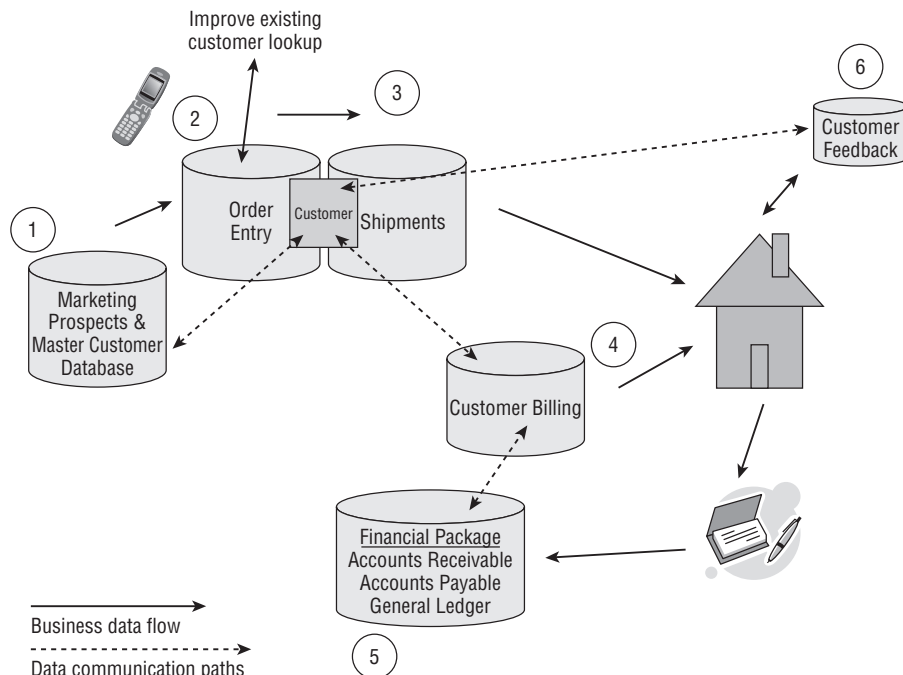
the years. If any changes are made to the data about a customer during the billing process, the data is updated only in the Customer Billing database.

In step 5, the customer data flows with the billing data into the third-party Financial Package. This package provides the accounts receivable, accounts payable, and general ledger functionality. All applications within this third-party software package use a common set of customer data.

Finally, in step 6, a new set of customer data is captured when feedback is received—via phone calls, e-mails, and letters. The data that is captured in step 6 is independent of the other processes and is not tied to any of the other processes. This is not a recommended information management strategy, but it reflects how customer data exists in the current systems.

The fact that there are multiple copies of customer data is quite common. Each application was designed and developed independently over a long period of time. Time and budget constraints often limit the ability of application development teams to ensure enterprise integration of customer data.

Information management takes a broader view by looking at the data across all application systems. Information management is concerned with integration and interoperability between the different systems. Improvements to how data flows between different parts of the organization can be made with business process changes and some small system enhancements. Figure 8-3 highlights where these improvements can be made.



**Figure 8-3** Improved customer data processes

The first change is to improve the customer search capability and speed up the order entry system. Adding a search alternative by customer number or home phone number can increase the number of existing customers who are identified. The customer service representatives are also trained regarding the importance of identifying existing customers. Additional discounts may be offered for these repeat customers once they are identified.

In many cases, it is not feasible to make significant changes to the order entry system. What can change is how the system is used. A little research into how application systems work often yields functionality that is not being used. In this case, the capability to automatically flag changes to customer data is enabled if the proper check box is marked on the entry screen. The customer service representatives need to be trained to check that box to flag any new or changed customers so that the changes can be automatically fed to other systems. While employees may resist having another task to do, the overall value can be emphasized.

More changes could be made by looking at each of the other application systems. Most important is setting up automated flows on a regular (hourly, daily, weekly) basis to enable the most current data to be shared across the organization.

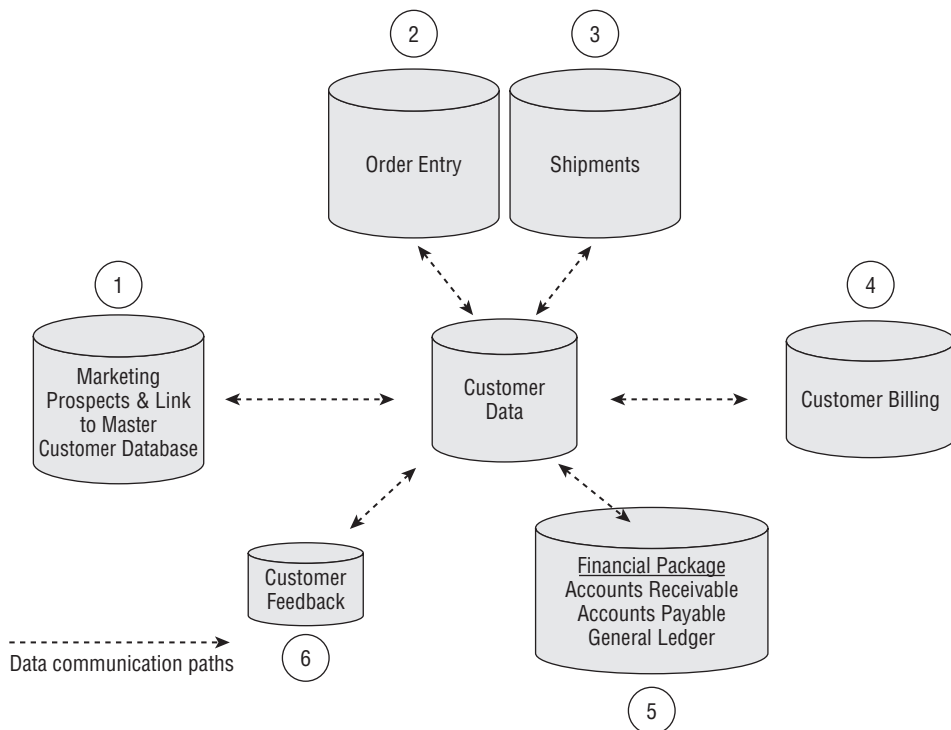
Taking a more aggressive approach, the customer data flow shown earlier could be revised so that a single customer database is used by all of the applications, as shown in Figure 8-4. This diagram demonstrates how all of these applications, from marketing through the financials, can tap into a single set of customer data. The marketing prospects are still maintained by marketing and flow to and from the centralized customer data. Note that the master customer database has been replaced completely with a link to the central customer data.

Additionally, this overall view addresses the need to collect data for all purposes, not just for a single application system. You don't need to know how the customer heard about a discount to capture his or her order, but in order to evaluate a promotion's overall effectiveness, you need to ask customers how they heard about the promotion and enter their response. This approach is also part of a master data management process, which is discussed later in this chapter.

While this last option has great appeal, most organizations get started with a blended approach of business process changes and system changes. Over time, progress can be made toward this vision for centralized, shared data.

The preceding example is a confined microcosm of a real corporate environment. Information management does not just happen. It comes about through a conscious, dedicated effort on the part of the entire organization. It cannot happen simply by having the IT people build systems to fix the problems. Successful information management environments demand a strong partnership between business and IT, as discussed in Chapter 4. This requires an in-depth

understanding of the business, how it works, and why it works that way. It also requires that you take a good look at how your business functions are currently set up. If there is a great deal of autonomy for each function, then you are likely to find a lot of segregation between the systems that support each function. Changes may be needed to your business processes to better support an overall information management strategy. This example is focused on the data itself. In order to attain the vision of integrated data, there is much more involved than changing application systems. It involves the many different facets mentioned earlier.



**Figure 8-4** Integrated customer data flow

## IM Beyond the Data Warehouse

Information management often comes to the surface when data warehouse initiatives are being discussed. Data warehousing is certainly part of an overall information management strategy, but it is not the whole thing. Information management encompasses all systems, operational and analytical.

In fact, the more comprehensive information management practices you have in place, the easier it is to build a data warehouse. This is because things

such as the definition and meaning of core data are already agreed upon and documented, other systems within the organization have already adopted these standards, and the primary sources for reference data are clear.

Without a comprehensive information management practice, much of the development work that is done on a data warehouse project is integrating and cleaning up the data as it comes into the data warehouse. This results in a consistent view for reporting and analysis, but the underlying data problems still exist. Each time data is loaded, the processes to clean up the data are run again.

Reflecting back on the customer data example in Figure 8-2, corrections are made only within the sphere of that single application, and are not fixed at the source. This can create confusion and may negatively affect customer relations. The company appears disorganized, and customers may be frustrated that a change must be requested multiple times, at different points in the process. A solid information management process helps eliminate the redundancy and incorrect information, improving overall customer relations.

## **Master Data Management**

---

There has been a surge of interest in *master data management (MDM)*. Master data management can seem like an intimidating and daunting task, but it doesn't have to be. Master data is core reference data that describes the key components of your business that must be used in a consistent manner across your entire company. For customers, this may include their name, mailing address, primary residence, income, education level, and other demographics. For employees, you would have similar data, as well as hire date, employment type (hourly or salaried), and professional qualifications. Product reference data typically includes details about the product such as size or flavor, dimensions of packages, and packaging material type. Many different processes and application systems in the company, including product development, manufacturing, marketing, sales, and distribution, need this basic reference data.

This basic reference data is called *master data*. Master data management is how the organization deals with the collection and maintenance of this enterprise reference data. Common master data is reflected as conformed dimensions in the data warehouse. Conformed dimensions are discussed in more detail in Chapter 7.

As in our customer example, many different applications need access to customer information, and typically each application has its own reference data. This means that there are multiple copies of customer data, including

address and contact information that may or may not be consistent. The same data element name may be used by different applications to capture and store different data. Perhaps each application uses a different business rule to determine whether the customer qualifies as a “premier customer.” There may also be duplicate customer data within and/or across applications. To make matters even more complicated, each application may have a different subset of customers.

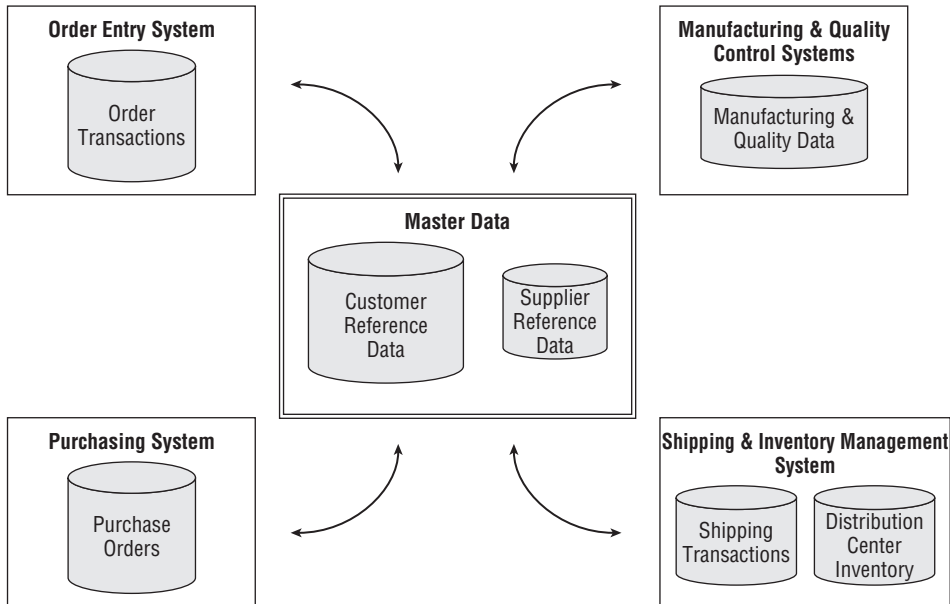
This can create severe problems for an organization. From an operations perspective, this can result in confusion and duplication of work. The sales department may contact a customer to offer a product promotion and be told not to call again until next summer. The next day, with no knowledge of the previous call, a service representative may call the same person to offer a discount on preventative maintenance services. Needless to say, that customer is likely to be annoyed by your company’s lack of coordination.

In addition to negatively affecting the company’s image, there is a significant cost associated with maintaining this data in so many places. There is the base cost for each application to capture the data and store it, and there is additional cost to get a consolidated view of your customers. Each time you need a consolidated view of your customers, you need to pull the customer lists from each system, combine them, eliminate duplicates, and clean them before any real analysis can be done. Prior to the implementation of a data warehouse, this is often done on a periodic—perhaps quarterly—basis, in order to create reports.

The overall objective of master data management is to provide common reference data to all applications across the enterprise. This means that each application uses the core customer data as the source of customer information for its processing and to, in turn, update the customer reference data with new information learned from business transactions in that system. This provides a single source for this core data. The business must decide which departments and/or application systems are allowed to change the master reference data.

You need a high level of coordination to keep applications from corrupting reference data needed for different purposes. For example, to support customer mailings, the marketing department wants to store a preferred first name. This may be a nickname. However, if that same field is used to validate credit card information, this will create problems. Often the personal preferences are different from the legal names used for financial transactions. The way to address these challenges is to identify the different purposes and store both versions: the legal first name and the preferred first name.

Figure 8-5 shows how the reference data is shared across applications, while the transaction data used remains within that application’s own database.



**Figure 8-5** Sample master reference data

## Master Data Feeds the Data Warehouse

Once master data management is in place, it provides an excellent source for the data warehouse. It makes it easier to build the dimensions of a DW using master reference data. This can also assist with the integration of fact data from multiple sources. All sources that use the master reference data can be integrated using these common keys, although this does not eliminate all of the challenges. For example, sometimes data from different systems is stored at different levels of detail, but when integrating data, it must be at the same level of detail. Therefore, the data with less detail must be able to be aggregated to the common higher level in order to combine it for reporting. Much of the work that is done in data cleansing and during the integration steps of the extract, transform, and load (ETL) process has already been accomplished in the source systems through their use of the common master data.

As with most things discussed in this book, it is very important to identify and engage the right resources needed to define and collect master reference data. This takes a continued strong partnership between the business and IT to be successful.

## Finding the Right Resources

It turns out that as companies are embarking on master data management efforts, they turn to the data warehouse teams to find the talent needed to get



the job done. The people in the organization who have already been looking at integration of data from multiple systems and taking an enterprise view of the business world are indeed the members of the data warehouse team. The necessary business knowledge often resides within this team. Depending upon the maturity of the data warehouse environment, your company may have the beginnings of the master reference data needed for analysis. This team can also be helpful in recruiting members from the business community and the operational systems teams.

Ultimately, a team that can represent the requirements for core reference data across the enterprise is needed to successfully implement a master data management program. A core part of this team will also be involved in data governance, which, again, requires a strong partnership between business and IT.

## **Data Governance**

---

Recall from earlier in the chapter that data governance initiatives improve data quality by assigning a team to be responsible for the data's accuracy, accessibility, consistency, and completeness, among other metrics. This team usually consists of executive leadership, project management, line-of-business managers, and data stewards.

Data governance is not a project but a cultural shift. As you can see from the roles listed above, in order for data governance to be successful, there must be a strong commitment and partnership between business and IT. At a high level, data governance is a set of business processes and procedures needed to ensure that the data is captured accurately and processed appropriately. Without a strong data governance initiative it is difficult for an organization to truly view its data as a corporate asset.

## **Data Ownership**

Several similar terms are used for the ideas presented in this discussion: data ownership, data stewardship, and data custodianship. They all define the responsibility of making decisions about how to define, process, and handle data. Your organization should pick the one that resonates best with your people.

Historically, the responsibility for making these decisions about how to process and handle data has often defaulted to the systems professionals who are developing or maintaining the application system. This is not usually a conscious decision, but just how things have fallen into place over the years. Too often the business community does not understand what data ownership is and how it impacts them, therefore, the business typically has not stepped

up to this role. Remember that true success in this environment is more easily achieved through a solid partnership between business and IT. It is time to take a step back and reassess how data ownership is handled.

### ***Who Really Owns the Data?***

“Owns” is a strong word, but someone must have the responsibility of managing and shepherding this process. There is usually some confusion about who really owns the data within an organization. Often, no one steps forward to claim data ownership, or, conversely, multiple groups claim ownership, including different business areas and systems.

Data is a by-product of how the business is run. Although systems professionals develop the applications that enable the business to keep running, the business community is responsible for driving what functions and rules are used for running the business. The business should drive the organization’s capability and function. For example, the accountants must set forth the criteria that constitute a legal accounting transaction, not the developers. Likewise, the business decides and drives what data must be captured, how it is to be used to keep the company running, and what is needed for reporting and analysis.

Often, people avoid being named an owner of anything. “Data ownership” sounds ominous and is perceived as just more work, or sometimes as being assigned an impossible task. Too often these responsibilities are assigned with no additional time, resources, or funding given to the data owner. Ownership often already exists, but usually informally. Think about who you would ask about the real meaning of the financial account structure. Who knows the most about details of insurance claim payments? The person who pops into your mind first is often the acting data owner.

Another way to try to identify business data ownership is to track the path for resolving data problems. The person who has to sift through a report presented to upper management with misleading or incorrect data is a candidate data owner. The data owner may already be identified as a *subject matter expert (SME)*. It may be best to ease into this by simply identifying the primary business contact for the different types of data. If “official” data ownership is too vague or unclear for business people to accept, then start by identifying primary contacts for the different kinds of data. Table 8-1 shows a sample of how data ownership can be tracked.

This process can be implemented at several levels. To get started, it may be helpful to identify ownership for an entire business application system, a module within a system, or a specific table. Begin at the higher level and work to lower levels when additional people have expertise in the different parts of an application. As the organization matures with data governance, ownership may shift toward business functions or business processes.

**Table 8-1** Tracking Data Ownership

DATA SOURCE	DATA AREA	BUSINESS DESCRIPTION	PRIMARY BUSINESS CONTACT	PRIMARY SYSTEMS CONTACT
Primus	Policy terms	Contains basic policy details	Ingrid	John
Primus	Policy ownership	Contains details about the people who have ownership of the insurance policy	Melanie	Steve
PLAN	Policy quotes	Contains basic information about the rating factors that were used in developing a quote for an auto policy	Ryan	Nancy
Primus	Claims handling	Reporting and payments for claims	Jonathon	Kate
SERVUS	Subrogation	Application used to track interactions with other insurers in order to get reimbursed for claims for which their policyholder was at fault	Leah	Debbie
REUP	Re-insurance	System that is used to manage the purchase of re-insurance to minimize overall risk of the company's portfolio	Michael	Nina

In addition to business data ownership, it is also helpful to identify who can assist with any technical questions about that data. This is usually called the primary IT contact, and it is not the data owner.

Sometimes data ownership is assigned to whomever seems to be willing to sign up, or someone is simply given ownership and burdened with the assignment. This is not a recommended selection method. As the organization matures with data governance, there are likely to be changes to who is recognized as the data owner. This shift can begin by selecting the business group that ultimately is responsible for that function in the business. Then, look into that group and select a strong, experienced person to be designated as the owner. Initially, it is easy to designate the person with the most knowledge, but the ultimate goal is to determine, from a business perspective, where this responsibility should truly be assigned.

The business must own the data, but IT must help to provide the mechanisms to implement and support the business requirements for collecting and handling the data.

### ***Your Responsibilities If You Are “the Owner”***

As stated earlier, often ownership carries a negative connotation. This happens when the responsibility is given to someone who does not have the means or authority to implement his or her decisions. It is helpful to understand what is involved, so that proper steps can be taken to ensure that the people assigned ownership responsibility can be successful. If you are the data owner, you need to be able to do the following:

- Define and help document what the data really means
- Be responsible for establishing quality expectations
- Be responsible for setting and enforcing policies concerning data quality
- Determine how the data should be used
- Identify and document caveats and special cases where the data can't or shouldn't be used in certain ways. Explain why.
- Assist in developing or refining processes for how data will be entered, including which group should be responsible for entry
- Make business decisions about how the data can be transformed or manipulated
- Help others to properly use and interpret the data itself
- Be empowered to make decisions about making changes to the data
- Be aware of interactions with and dependencies on the data
- Work across the organization to consolidate processes and share data
- Determine when historical data can be archived
- Work across groups to resolve differences of opinion about the data, and facilitate workable compromises
- Help others understand the impact of their use of the data or their requested changes
- Work with IT staff to develop strategies for developing solutions to address challenges with the data

While a single data owner can do many of the daily interactions, there will indeed be times when the question at hand spans multiple departments. This is when procedures for data governance are needed to reach the best resolution for the entire organization.

While much of the attention is on dealing with the data as it currently exists, the data owner can also play a valuable role for the future. Those with the most knowledge of today's data usually have the most knowledge about the problems with that same data. This understanding is needed in order

to develop strategies to address the problems and move the state of the data forward over time.

### ***What are IT's Responsibilities?***

Although data ownership is a business responsibility, there are several critical things that the information technology staff must provide. Following are a few examples of the IT staff's responsibilities:

- Provide the technical environment to store data.
- Ensure that the business rules are implemented correctly.
- Assist the data owner in researching data-related problems.
- Help design solutions to correct data problems and prevent them from happening again.

These are only a representative sample of responsibilities. The technical staff should be considered the group to provide the mechanism of support for all information management activities.

### ***Challenges with Data Ownership***

Over time, business rules and processing logic have become embedded into application systems. This ensures that the rules are applied in a consistent manner and that the people running the system are no longer required to know them at a detailed level. A negative side-effect of this is that institutional knowledge about why business is done that way, or sometimes even how business is done, is lost. The people who really understood it have retired, left the organization, or may have been promoted to very high levels and are not easily accessible. The people who are now responsible for this work know what steps need to happen to get the output that is needed on a periodic basis, but there is no understanding of the underlying business purpose. It can feel like being stonewalled in these cases, as you cannot get straight answers.

Acknowledge this situation if it exists, and give those people leeway. How could they possibly be expected to know the answers to detailed questions? This does not mean that you should just accept the status quo; it means that it will take longer to figure things out. Depending on how old the application system is and how much the business has changed, it may be better to just redefine what you want done. You can unravel the code and spreadsheets to see what is actually done, although it is unlikely that you can figure out why. Be prepared to discover that things are not what everyone thought. Some calculations or processes may not be done using the current business rules. Even worse, the rules may be applied in only some cases, while others

may continue using old logic. This can easily happen when business logic is implemented in multiple places.

As the owner of data, you need to look out for the care and feeding of that data. As a member of a specific business group, the needs of that group are only one of the perspectives that must be considered. The impact on the overall enterprise must also be considered. You cannot simply look out for the interests of your own group and ignore the needs of the rest of the organization. When that happens, other people start building their own databases—with duplicate, but not integrated or necessarily accurate, data. Data ownership is a very important part of the entire information management process and critical to being able to utilize data as a true asset in your organization.

## Data Quality

Data quality, as with most terms we use, can have many different meanings. Let's settle on Wikipedia's for this discussion: Data quality is "the state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use."

The need for reliable, accurate data is critical to support the decision-making process of each organization. The previous sections of this chapter pertain to all types of systems, but this section looks specifically at data quality in the data warehouse. Data quality problems have many different symptoms, including the following:

- Data can be missing completely (e.g., a sales transaction may be found that does not have any customer identification associated with it).
- Data can be incomplete (e.g., the customer's street address and city were entered, but the state and zip code were left blank).
- Data may not be available. For example, an auto policy was issued, but the system did not retain information about which channel this business came through. Policies can be issued to customers through agents, via the phone, and over the Internet. This is common when business changes are made in the source systems or new systems have been implemented. All policies issued prior to this system enhancement will never have this data associated with it, so while it looks like it is missing, it really never existed and it cannot be recreated now.
- Data can be wrong. For example, a sales transaction may identify a frequent shopper who does not match the customer identifier also captured on that transaction. Sometimes these errors have a big impact on the business and appropriate corrections are made to the source systems. However, the transactions that occurred with the erroneous data may not be corrected in the source system.

- Data may have multiple meanings. For example, the single field used to indicate the account representative responsible for this business could contain a link to another database for a list of promotional codes. The logic to determine how to use this field may be embedded in program code.
- Data may not be stored consistently. For example, sometimes systems are used in interesting and unusual ways. Clever codes may be used to indicate specific situations. Some account numbers require manual intervention prior to posting the numbers to the general ledger. This list of accounts may not be stored in a system but is known to the financial staff. For these account numbers, the actual amounts are to be pulled from the unit of measure field, rather than the amount field.

Often, the data warehouse team is left with the responsibility of sorting all of these situations out. Real data quality must be ensured by the source system, because only so much can be done to clean up data after the fact.

Poor data quality can impact the business. If the data in the underlying source system is not correct, then the input to the data warehouse will have problems. While some cleansing can occur when loading the data warehouse, this data may still have problems. As a result, reports created from this data warehouse will also inherit the bad data. Ultimately, this can result in poor business decisions being made. These poor business decisions are not caused by mistakes on the part of the staff, but bad conclusions can be drawn using erroneous facts. Many organizations are in full agreement up until the last statement. Then there is a strong denial that “we don’t make bad decisions!” Decisions are only as good as the data and assumptions that are used as guiding input.

### ***Profiling the Data***

One technique that is being used to proactively improve data quality is data profiling. Recall from earlier that data profiling is the process of examining the data currently available in an existing data source, such as a database or a file, and gathering statistics and information about that data. The purpose of these statistics may be any of the following:

- Find out whether existing data can easily be used for other purposes.
- Provide metrics on data quality, including whether the data conforms to company standards.
- Assess the risk involved in integrating data for new applications, including the challenges of joins.
- Track data quality.



- Assess whether metadata accurately describes the actual values in the source database. Understand data challenges early in any data-intensive project, in order to avoid surprises later in the project. Finding data problems late in the project can incur time delays and project cost overruns.
- Have an enterprise view of all data, for uses such as master data management, for which key data is needed, or data governance, for improving data quality.

Data profiling tools can be used to proactively sift through data elements and help determine the level of data quality. In addition, it is important to document the impact of data quality problems as they are discovered. This helps to justify the time and cost of cleaning the data. Data profiling is often done to identify and improve source systems themselves. Data profiling may also be done against these source systems to help design the ETL processes used to build the data warehouse. The goal is to improve the accuracy of how data is captured and stored in any system.

### ***How Clean Does the Data Really Need to Be?***

The level of data cleanliness is highly dependent upon the industry you are in, the type of data you are looking at, and what data has been available in the past. In addition, each department in the organization must decide what will be acceptable. Remember the last part of the definition we used earlier: “accuracy that makes data appropriate for a specific use.” If the data warehouse will be used to produce official financial results, then the data must balance exactly with the general ledger. However, if the data is being used to support management decision-making, then the quality of the data may not need to be quite so high. If management does not have any access to financial data until two months after the books are closed, having 80% accuracy within a week of the close would still be useful.

The level of acceptable data quality will also change over time. If this is the organization’s first serious look at the quality of the data, then the bar should be set at a reasonable level. If the organization has been working on improving data quality for several years already, then the target for data quality will be much higher.

### ***Measuring Quality***

There are entire tools and methodologies built around assessing and improving an organization’s data quality. It is certainly worthwhile having someone in the organization learn more about these and assess their applicability in your organization. This section covers some basic concepts, focused on the data warehouse, that everyone can easily understand, and which should be sufficient for getting started.



Typically, most data flows successfully through the *extract, transform, and load* (ETL) processes and all the preset checks and balances with no problems. Often, this will represent 75%–85% of the data. Basic validation rules should be part of the ETL processes. For example, these rules might ensure that the product number exists, that the zip code and state combination is valid, and that an employee number is for a current employee. If the data does not meet the criteria, then a warning or error is generated and processes are set in motion to inform the staff and generate error reports. Many measurements of data quality have a systems processing focus. Additionally, it is important to include business criteria to measure quality. Examples of common system measurements include the following:

- Total number of rows processed
- Number or percentage of rows with a warning condition
- Number or percentage of rows with an error condition

Examples of common business measurements include the following:

- Total number of units sold
- Total dollar amount of premium collected
- Change in market share compared to the previous period

The general system errors can be tracked over time to determine whether there is an increase or decrease in the number of errors encountered. This could indicate changes in the business processes or systems where the data warehouse team has not been notified. The business measurements should be compared to reasonable variations. If a swing in market share of greater than 0.5% is unheard of in your industry, this benchmark can be used to flag potential data problems.

### ***Quality of Historical Data***

The older the data, the more data quality problems you will find. The most accurate data tends to be the current data. Over time, enhancements are made to operational systems that tighten up controls for data entry and processing. Business processes improve to make it a standard part of each workflow to capture critical data. New systems are developed and implemented that include data collection to support reporting and analysis. All of these improve the quality of the data.

Historical data does not benefit from these improvements and therefore often has the worst quality. It is important to evaluate the business benefit of making bad historical data available. Keep in mind that the effort to work through data problems is often grossly underestimated. These estimates are usually based upon known problems. There are also many other problems

that have yet to be discovered and may prove to be difficult to diagnose and even more difficult to correct.

The business value of the oldest historical data must be scrutinized. For retail organizations, what was hot three years ago is unlikely to be an indicator of what will be hot this year, such as in the fashion industry. Historical data may indicate the volume of the fashionable items that are sold, but may yield little insight into which items are more likely to be trendy. Conversely, actuaries want all data possible from all hurricanes since data was collected. This long-term history is useful to feed statistical models.

A reasonable goal for data quality should be set—remember, accuracy appropriate for use. These goals can set to higher levels as the organization, both business and IT, learn more about the problems that exist with historical data and techniques to address them. Don't be afraid to back down from a goal that is too lofty to start with. Focus on getting better data into the hands of decision makers as soon as possible. This may even help to identify problems and determine the possible value of addressing those problem(s). You learned about data profiling in the section on data quality; historical data is an excellent place to use these tools and techniques.

#### **CASE STUDY: STARTING AT THE WRONG END OF THE SPECTRUM**

**This case study comes from a real-world scenario. Many other organizations may have had a similar experience. This is not a recommended approach, but is included to highlight the need to continuously review the business value of major project choices.**

**The organization needed to have five years of historical transactions to support its analysis. A new operational system was implemented two years ago. Two different sets of work needed to be done: one to load the history from the old system and one to develop a process to load historical and current data from the new system. A decision was made to start with the oldest data first. The strategy was to load the data in historical order, oldest to newest.**

**Once the ETL processes were built, loading the first three years was started. The goal was to be able to include all transactions. Many problems were discovered in the historical data, which did not meet the current data quality objectives. While most of the data was fine, 5% of the transactions had problems. Detailed research was required to discover how to handle these outliers. The ETL developers dutifully adhered to their instructions to clean all of the data. Many hours were spent researching the errors in a small number of transactions, using the time of the most skilled and experienced resources.**

**As usual, this took far longer than anyone expected. It took two years to load three years of the oldest data. Now the development to extract data from the new operational system could begin. At this point, the data warehouse still did not contain data that could be used to support any reporting or analysis.**

*(continued)*

**CASE STUDY: STARTING AT THE WRONG END OF THE SPECTRUM  
(continued)**

If the work to pull only current data from the new system had been developed first, the loads could have been running that entire time. After two years, the organization would have the most recent two years of history with no additional cost or effort.

Two lessons can be learned here:

- Periodically evaluate the effort and resources being put into chasing down data problems. Look at the work compared to the impact on the business. Understand that the cost to research and correct data problems may be more than what the potential business value, especially if the data represents a very small percentage of the business.
- Critically review the historical requirements to support business analysis and determine the most direct and cost effective way to have that data loaded.
- Once the minimal historical requirement is met, decisions can be made about the cost and value of loading more historical data.

***Cleansing at the Source***

When possible, the best solution to data quality problems is to address them at the source or where the data is originally captured. The first step is to ensure that research is done to determine the root cause of the problem. When the problem is in the data warehouse, it can be corrected there. More often, however, the root cause of data quality problems exists in the source systems and/or business processes.

Addressing the problem at the original source can mean changes need to be made to core application systems and/or business processes of the organization. This would be a reasonable alternative if nothing else were going on in the organization, but reality and experience show that this is often not a realistic option. Usually, the application support team does indeed agree with the recommended correction, but it may not be a priority when ranked against other changes that may be needed just to keep the business running, and it may not be implemented for 18 months! Changes to business processes can encounter similar delays. You may find you can change a business process more easily than changing the systems to capture the data, or you may find you can change the systems faster than you can change the business process. You need to understand all the options and take a pragmatic approach to the problem.

The data warehouse and operational application teams must work together to determine the impact of the change on the business, and how this may

impact other aspects of the operational environment currently running the business. In many cases, all of the changes are possible, but adding new data quality or cleansing functions can negatively affect the delivery schedule for other features. From the business perspective, what is more important to the organization's overall success?

When changes cannot be made at all or will be made too far into the future, the data warehouse team must assume the responsibility of addressing these data quality needs. However, no matter how good the data warehouse team is, they cannot manufacture data that is not captured anywhere. The data warehouse team needs to stay focused on the goals and objectives of the data warehouse. If they uncover issues that are outside the team's scope, they need to identify and document them and then turn them over to the appropriate group (data governance committee, MDM team, etc.). Data profiling can and should be used throughout this process.

### ***Cleaning Up for Reporting***

If efforts to implement data cleansing in the underlying source systems are not successful, the data warehouse team must include processes in the ETL system to clean the data before it is loaded into the data warehouse or convince the business to reevaluate its quality expectations. By building data cleansing logic into a systematic ETL process, the rules will be applied in a consistent manner for all reporting, rather than on a report-by-report basis. Labels and descriptions can be cleaned up. Duplicate data rows can be merged. The project team has a lot of flexibility with regard to cleaning up the data. This cleansing can be done without worrying about the impact on how the operational system functions.

### ***Managing the Integrity of Data Integration***

Data is often fine within the context of a single application. In fact, most application systems have quality checks on their own data to ensure that processing is performed correctly. Many data accuracy challenges arise when you try to integrate data from more than one system, such as when product and customer codes do not align.

Often the data from these systems has been used together for years. Sometimes this has been done in a previous reporting or data warehouse project for which the data has been integrated and coexists in the same database. If this is the case, then much of the hard work may already be done. If there are any major problems with the existing processes, the business community will know about it or it may be that no one is using this reporting environment. Take the time to review how the data is integrated and ensure that the business intent has been implemented.

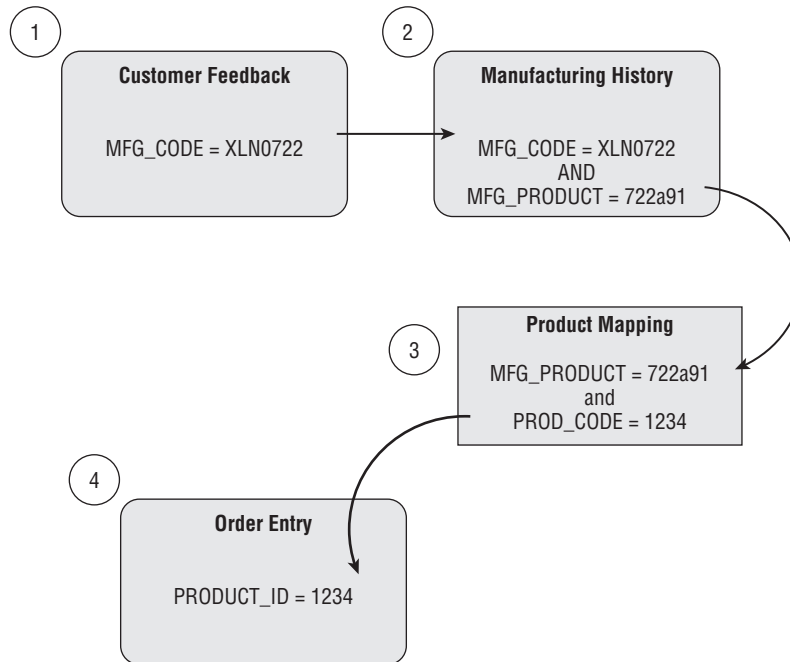
In other cases, data from these sources has not been integrated in a systematic way. If there is a business need to have this data together, it is likely that someone in the organization is already doing this by hand. Try to find a report where the data from these two different systems is being presented side by side today. Then, dig in to find out how this is being done. While there may be some manual manipulation of the data, there will still be some underlying queries that extract the data from the source. The business logic needed to integrate the data may be embedded in spreadsheets, SAS code, or someone's head.

Some organizations have employed strong information management practices for years, so that standard customer and product identifiers are used by operational systems throughout the organization. This improves the ease and accuracy of integrating data from multiple sources. Unfortunately, this is not always the case. A direct link may not exist, but often through a series of steps, the data can be linked to other identifiers that are included in the second source system.

Figure 8-6 shows a basic example of data integration. In order to integrate customer feedback on products with actual product sales, there is no direct path to follow. First, when a customer calls in, the manufacturing code that is stamped on the product is requested. This can be linked to a manufacturing history table from which the manufacturing product identifier is pulled. Using a product mapping table, the manufacturing product can be linked to the product code that is used in the order entry system. It is no wonder that it seems impossible to pull integrated data for reporting. This is straightforward if the mapping process is clear, but often this is not well understood by business users who simply want some reports. Sometimes this product mapping is developed and stored by a business analyst and is not generally available.

Sometimes there are no obvious links in the systems to support integration of data. Careful research needs to be done to ensure that this is really the case. Then, explore what would need to change in order to make this possible in the future. Sometimes it requires significant re-engineering of one or more underlying operational systems. If this is the case, the data integration requirement should be included in any system redesign efforts in the future.

Searching for which fields in the database can be linked together clearly requires strong technical skills. The people with these skills may exist only in the IT group responsible for supporting the application. However, do not leave the task of integration completely to the IT staff. Representatives from the business perspective must also assist in this work. Often, the business community is already familiar with the challenges of merging this data and can provide invaluable insight into how this can be done correctly.



**Figure 8-6** Data integration paths

### ***Quality Improves When It Matters***

Data that is required for a business transaction to run successfully is generally entered completely and accurately. In addition to the minimum data to complete a transaction, other data may be known at that moment that better describes the transaction. For example, it may be enough to capture the shipping address to place an order, but it may also be useful to know if this is the address of the customer's primary residence. Additional customer demographics may also be available. The level of detail applied can vary greatly depending upon the total number of fields that need to be entered and the amount of work that is waiting to be handled (e.g., customers in line, phone calls on hold, claims that need to be entered). Even required fields may not be captured carefully, and default values are often selected. If default values or other shortcuts are used to complete tasks that are valuable for analysis but not critical to complete the transaction, and there is no negative feedback or consequences for these shortcuts, why would any employee take the additional time to do something whose value they cannot perceive?

Data quality greatly improves when there is a clear impact. This impact can be feedback about the number of entry errors made by employees, public release of data that is incorrect, or the association of job performance with improved data quality. When individuals can see that their efforts make

a difference, data quality improves, even if no other systems changes are made. You need to make individuals accountable for their actions in the data collection process.

### ***Example: Data Quality and Grocery Checkout Scanners***

It is critical to ensure that the right people in the organization understand the importance and impact they have on data quality. This was evident when grocery stores began implementing scanners for checking out purchases. An initial benefit to the cashiers was that this sped up the checkout process. A larger goal was to collect store sales more accurately.

Experienced cashiers were often faster than the scanning technology at the time. In order to speed things up, cashiers would use the same approach they had always used—enter a product once and increase the quantity. For example, if a customer had many jars of baby food, one jar was scanned and then the quantity was change to reflect the total number of jars. While this was faster at the moment, it created a data quality nightmare because the store did not sell twenty jars of one item (e.g., pears), but a variety of different items (e.g., applesauce, peaches, green beans, and peas). Needless to say, the accuracy of item-level analysis was greatly reduced.

Organizational discipline and accountability was needed to ensure that each individual item was scanned, or that only identical items were entered for multiple quantities. Accuracy began to improve, but there was still not a clear understanding of the overall value of this accuracy until this same data was used for inventory replenishment. This took the need for data accuracy from a conceptual level to a specific, store-level requirement. This in turn raised the importance with store managers, who in turn emphasized this process change with their employees. Some organizations understood the value and pushed that need down to the individual cashier level early on. Those organizations were able to leverage that data sooner for a competitive advantage.

### ***Example: Data Quality and the Evaluation of Public Education***

This example is intended to illustrate the types of things that motivate behaviors that lead to higher quality data capture. It is not intended as a critique or judgment of the political aspects of any government programs.

This example demonstrates the importance of increased data quality in the public education system. Historically, the department of education within each state sets education goals and standards. The individual state also determined what financial and educational results were measured and tracked. For years, school districts submitted data to their respective states. Most of this data was used to determine how funds were allocated and distributed across the state. Because this was driven state by state, some states had more rigorous



requirements than others. Even the most progressive states had many data challenges.

The data entry function for education data was often relegated to a clerical function and was not the primary job responsibility for that person. Typically, a school receptionist or principal's secretary was responsible for entering all student data. This function was performed amid many others: answering the phone, dealing with parents in the office, and the many other tasks that keep a school running. From the secretary's perspective, the most important data elements to be carefully entered included student name, phone number, and address. These data elements are needed in the daily functions of the school. However, additional data elements were not being entered as carefully, such as ethnicity, parent's education level, or specific disability conditions. Many of these additional characteristics of a student may not be known at the time a student is enrolled. Additional characteristics about each student may be added later, as time permits.

When the No Child Left Behind Act was passed, additional data was required to be submitted to the U.S. Department of Education. This data would not just be submitted and sit in a database somewhere. It would be used to grade the performance of each school and district, and would be shared publicly. The school's performance could also have an impact on future funding. Needless to say, much more attention is now being paid to the quality of the data; however, our schools are much better equipped to interact with children, rather than data. In order to ensure proper reporting of school performance, each student's data must be accurate and submitted with the appropriate characteristics. Many school districts and schools now have staff members dedicated to data entry and validation. Others are providing better training to those who are involved with data entry.

Many of the challenges facing school districts and states are clearly related to data quality and data management issues. Now that the data matters, much more care and attention is paid to how and when data is captured and maintained. This improved data will also help support better analysis of the education process itself.

### ***Realizing the Value of Data Quality***

At the end of the day, working on data quality is painstaking and sometimes tedious. It requires attention to detail and tracking down many little details, and it often feels like a rather thankless job. Others begin to avoid you, taking alternative routes to get from place to place. Take heart—this process *will* provide value to the organization. For a data warehouse, the worst part is during the design and development, as all the issues need to be researched and followed up. Use all the resources at your disposal—data governance group, MDM group, and data profiling tools—to help you through the process. Once



the system goes into production, things should ease up. However, exceptions will inevitably arise and additional research will need to be done. Improved data quality helps the business to be better, and more accurately informed about what is actually happening. This enables better decisions to be made, which are ultimately reflected on the bottom line.

As all of the data quality issues are being addressed and you are moving through the project design and implementation process, you should be building an environment to catalog the data elements. The next section discusses options for accomplishing this task.

## Implementing a Data Dictionary

---

How much time is wasted when there is no documentation about your data? Each business analyst who is learning about a data source must ask the resident expert directly. Each business or IT person who has to write reports must ask the resident expert about the contents of the data files. Each application developer must ask the resident expert when fixing application problems or making enhancements. As data flows from one system to another, this also raises a new set of questions. Even worse, a person developing a report does not ask, but simply makes assumptions and delivers data that is not correct. The aftermath of correcting this type of situation is often more time consuming than reviewing the content in the first place.

Over the years, this can add up to a large number of wasted hours for the business and IT that could be greatly reduced if you only had some basic documentation about your data. A *data dictionary* can provide that place to capture and store the details about your data. The data dictionary is a single shared place to store definitions, descriptions, and uses of every data element. The notion of having a corporate data dictionary that is available to everyone is very appealing. However, a data dictionary is very difficult to develop and maintain. Most of the challenges are not related to the design and maintenance of such a system, but to getting everyone to agree on the definitions and keeping the contents complete and current.

**NOTE** A data dictionary is a mechanism to support the organization's data governance efforts. This is important to capture and communicate the business knowledge of the data.

## The Data Dictionary Application

Building or buying a data dictionary application is not always a straightforward process. Some of the challenges that can be encountered are more technical in nature. You can find limited off-the-shelf products that provide full data dictionary capabilities. Often these are tied to or integrated with other

application software, so the capabilities are good as long as you are using the other applications. Progress continues to be made in the marketplace in this arena, so it is worth taking some time to research current product availability.

If a data dictionary product does not meet your needs or is cost prohibitive, you can build a basic data dictionary internally. All you really need is a database to store the dictionary and an interface that enables others to read the contents—and, when appropriate, to be able to add and update the contents. One rapidly growing alternative is to develop your data dictionary by building a wiki for your organization. This is a very flexible, open, collaborative environment. If you do this, just make sure you have controls in place regarding who can post, edit, and contribute to the process.

This does not need to be an elaborate system, but it should contain some basic details. Table 8-2 shows several useful things to include in a data dictionary.

**Table 8-2** Possible Columns in a Data Dictionary

COLUMN NAME	DESCRIPTION
Technical Data Element Name	The actual field or column name in the system where this data is stored. There can be multiple technical data element names if the same business element is stored in multiple places.
Business Data Element Name	The name of the data element in business terms
Business Data Element Description	This describes the data in business terms.
Technical Data Element Description	A description of this data element that contains details useful to application developers about how this is manipulated or used in the system.
Technical Contact	The name of the application developer who would know the most about this data element.
Business Owner	The name of the person in the business who would be able to answer questions about this data element. (If ownership is not well defined or there is resistance, call this the “business contact.”)
Sample Values	A few examples of the content of this data element can be very helpful for quick understanding. Often, both IT and business people may be familiar with this data element, but a new business name may be assigned as part of the overall data management efforts, so this will help people realize which field this is.

These are just a few of the things that may be of interest. Additional details such as field type and length may be included for the benefit of a more technical audience. However, ensure that the business perspective is represented; otherwise, the data dictionary will only be useful for the technical staff.

The data dictionary application is more than just the database itself. It must enable designated users to enter and maintain the data dictionary. This needs to be done in bulk or on an individual basis. The bulk loads will likely be from spreadsheets that have been created during a development project. The individual maintenance will be done over time to keep the existing data dictionary current and accurate.

The other primary capability of the data dictionary application is the ability to view the contents of the data dictionary itself. This should be a straightforward, basic application for viewing the contents of the data dictionary database.

## **Populating the Data Dictionary**

Once the database or wiki is set up and an interface is provided for the data dictionary, it is like having a bound book with blank pages. It is only useful when the content is added. This is where the really hard work begins. With all of the company's databases to choose from and likely tens of thousands of data elements, how do you get started?

To begin, pick one system or database, and try to choose one that is being worked on for other reasons already. This may be for a set of major enhancements to a system, or the system has been identified to feed the data warehouse. If this data was loaded into the data warehouse, you already know that this data is of interest to the business and provides value.

The data warehouse team must work with the business and the technical teams to understand the contents of this database in order to effectively model and build the data warehouse components. The data warehouse team needs to develop clear business names and definitions for all the data to be published through the data warehouse. Because the data warehouse team will be performing some of this work anyway, it makes sense to combine the data dictionary work with the data warehouse. Make sure that the project scope and funding takes this work into account.

As always, when there are not enough resources, documentation tends to be ignored. Because populating the data dictionary is additional work, more resources should be allocated to the data warehouse team to ensure that this actually happens. This is not meant to imply that this is only a data warehouse issue, but sometimes it needs to be started there.

You can get a jump start by loading the technical details in from the database itself. This is a technical task, but the application developer or

database administrator can quickly provide this in a data file or spreadsheet. From a systems perspective, it can also help to determine what percentage of the rows do not have data, also referred to as *null*. If the application system does not enter data into this column, then there is no point in creating a business description.

Once the basics are populated, it is time to fill in the rest of the details. Representatives in the business community must provide the business names and definitions. Several different techniques can be used. It is important to pick an approach that fits with your corporate culture. This must be as painless as possible for those key business users, also called subject matter experts. It is much easier to get help from more people if you have an open, collaborative environment that easily enables people to edit and add to the dictionary.

## **APPROACHES TO CAPTURING INFORMATION FOR THE DATA DICTIONARY**

### **Direct Documentation**

---

**Some business people find it easier to simply fill in labels and descriptions themselves. They can do this in small increments between meetings, or often will stay late and quietly enter the descriptions. If a business person is inclined to document things, he or she may have already documented the key data that is frequently used.**

### **Interview**

---

**Some people never write anything down. They just know all of the content. The only way to gain access to this information is to sit down and ask them. Someone must be assigned to conduct sessions in which each element is looked at and the responses are entered into the data dictionary.**

### **Compilation**

---

**There may be bits and pieces of documentation that have been informally developed over the years. This may be in the form of handwritten notes, spreadsheets, or new employee presentations. All of these can be collected and entered into the data dictionary. Once this compilation of the different content is completed, the business group must validate the results.**

*(continued)*

**APPROACHES TO CAPTURING INFORMATION FOR THE DATA DICTIONARY**  
**(continued)****Ghostwriter**

**It may be necessary to create a draft of the business name and description. This may be done by a newer member in the business group or by someone from IT. Appropriate business people must validate this draft. It may be easier for the subject matter experts to critique something, rather than write it from scratch.**

Be careful to have realistic expectations about how quickly this will be done. This type of documentation is often just one more thing that these key individuals are being asked to do. Their participation is often provided on their own time, after their other immediate work is completed. Encourage and reward their participation. One big benefit to these key people is that once this is documented, fewer people will need to ask basic questions about the data. The data dictionary can be consulted first, and only more advanced questions will require the involvement of a subject matter expert.

**Accessing the Data Dictionary**

After you have good content in the data dictionary, the real value can be realized. This happens when the data dictionary is made available to the business and IT communities. The data dictionary can assist business people who are trying to understand a report, or application developers who need to know about the content of a table in the database.

The data dictionary should be published in a simple format. A basic search interface can be developed and the contents can be available through the company's intranet. The interface should be simple, with options to constrain what you are looking for, and be interactive with the data dictionary database. Take advantage of how other content is published across the organization. Use these existing applications as a starting point and go from there.

**Maintaining the Data Dictionary**

Often, a big initiative is set up to create and populate the data dictionary. This is well publicized and supported at multiple levels of management. However, implementing a data dictionary is a long-term commitment, not a one-shot effort. After the data elements from the first data source have been documented and loaded into the data dictionary, the work does not stop. Additional data sources and additional types of information can be added to the dictionary

and then captured. Each report, including its purpose and data elements, can also be added.

As business data elements are identified and documented, a cross-reference to each field or column containing that business data element can be captured.

A small amount of effort is needed to simply maintain the data dictionary over time. This entails communicating with the different project or application maintenance teams to ensure that as enhancements and corrections are made, the contents of the data dictionary are updated to reflect those changes. If you have a data governance team, it should assume the responsibility of maintaining the data dictionary; otherwise, the data warehouse team may need to take the lead.

Expansion of the data dictionary takes a lot more effort. This requires the involvement of not only the IT people, to explain what is put into a field or column, but also the active participation of the business community, to clarify what the data means and how it is used.

## **Getting Started with Information Management**

---

While the goals of information management are admirable, there is usually some sort of a catalyst that sparks interest and support. An information management push generally begins after the following has occurred:

- A major data issue, perhaps with external exposure of inaccurate data to a regulatory agency or the press.
- A key executive feels the pain and understands the solution.
- Challenges in the data have grown to a crisis point.
- Paralysis, inability to get data needed to conduct business.
- Changes in management personnel, whereby new staff members are used to a greater availability and quality of data for running the business.

While interest builds, this is the time to ensure that the breadth and severity of data challenges are documented and communicated. The real first step is to get your arms around what you have and where you are today.

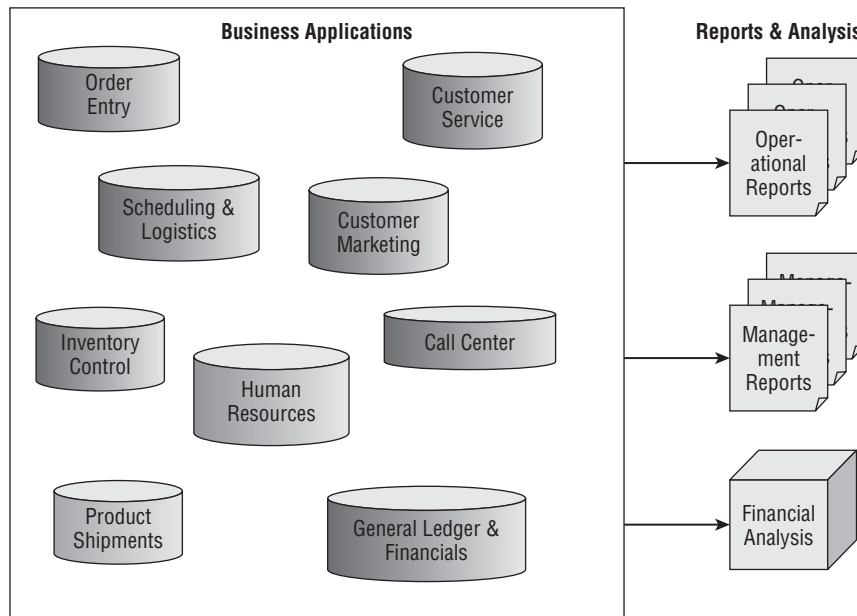
## **Understanding Your Current Data Environment**

It is important to gain a realistic understanding of the state of the data today. This understanding must come from both business and IT to really be effective. Two different paths need to be followed. One is to get an understanding of what data is stored where. The second is to gain some insight into the

organization's level of information management. In the end, it again is about the partnership—IT needs to manage where and how to store the data, and business needs to be able to understand what is available, and what it means to effectively use the data.

### ***What Data Do You Have?***

The organization may already have a sound understanding of the current data environment and it may already be diagrammed. If not, then a decision must be made about the level of detail that is currently needed. Perhaps a quick overview, such as what would be shared with a new employee, is all that is necessary. If there is a need to understand all of the data interfaces between operational systems, this will take longer. The goal is to get a fundamental understanding of the primary data the company already has. Taking the quicker route, a simple diagram can be documented. If this is not formally documented, it has likely been drawn repeatedly on white boards during meetings. A simple example is shown in Figure 8-7.



**Figure 8-7** Current data environment

Each of the major databases, usually associated with the main operational systems of the organization, is shown. Other types of data, including spreadsheets and departmental databases, should also be included. Do not limit the data to that which is managed and backed up by systems staff members.

Include the data that is loaded onto that PC in the corner of the office and the spreadsheets that are maintained and distributed on a weekly basis to the entire field sales force. If there are a small number of these sources, each can be named. If there are too many to list, indicate multiples and name a few that would resonate with most people in the organization. This is not intended to be a full data flow diagram or to provide specific database details. The point is simply to identify the major systems and to understand the levels of integration that currently exist.

The scope and level of detail needed to describe the current data depends on the size and complexity of the organization. A large, multi-division enterprise needs a very high-level view of the major systems, and acknowledgment that smaller departmental or personal systems exist. Then each individual division should have a more detailed diagram to describe the data that is used within the division. A single line of business organization may find one summary of the data to be sufficient.

### ***What Already Exists?***

The second set of information you want to understand is the existence and/or maturity of information management practices in the organization. To help determine what parts of the information management process already exist, ask the following types of questions:

- How well, if at all, do the application systems integrate today?
- Which kinds of data have multiple copies?
- Is there a single clean source for customer reference data?
- Are common product definitions in use across the organization?
- Which parts of the organization are interested in what data?
- What level of trust do you have in the accuracy of the data in the systems today?
- Are some systems better than others? If so, why?
- Do your application systems each have their own reference data?
- Are there duplicates in the reference data?
- How do your systems currently interact with one another?
- Are there common keys between systems—that is, some sort of system identifier to enable the data to be accurately linked together?

These questions will help you assess how far from a coordinated information management strategy you are. As you begin to understand and



communicate the level or lack of maturity of information management, make sure that this is handled carefully. Most system designers and developers are trying to do the best they can for the organization. Compromises and choices are made every day in order to keep the business moving forward. Deal honestly with the realities of your data, but don't point fingers or blame others. This must be a nonjudgmental statement of where things stand.

## Where Do You Want to Be?

Now that you have a better understanding of where you are and the state of integration or lack thereof, you can draft the vision outlining where you want to be. This is the beginning of a data architecture. The topic of data architecture is covered in more depth in Chapter 9. A full data architecture provides more specific details about how data will be structured and stored. The information management strategy needs to stay focused on what kind of data, customer, supplier, agent, or broker is needed by which systems. The information management strategy must also define the goal regarding how data is shared between systems.

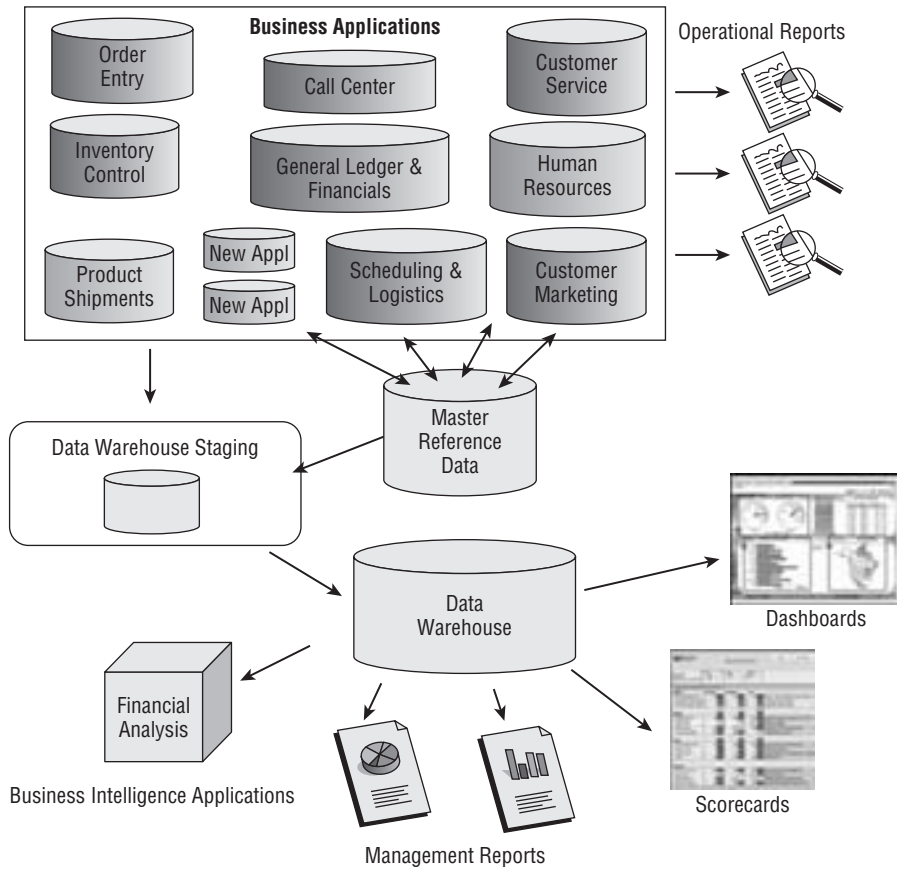
As stated earlier, because information management has many facets, the strategy must help address the business processes and data governance needed to achieve the goals.

While implementation of full information management plans takes many years, you need to have that vision first so that everyone knows what you are working toward.

Figure 8-8 shows an example vision describing how the data will be handled in the future. You can still see the different operational business applications. These all take advantage of the master reference data. This master reference data also feeds the data warehouse staging area, and ultimately the data warehouse itself. This diagram also shows how operational reporting continues to be supported directly by the business application systems, while the data warehouse supports management reporting and analysis.

The largest gaps between the current and future states are usually quite obvious. It is important to identify and understand these gaps. The business benefit of addressing the gaps can be studied, along with the impact of not addressing these gaps. Developing plans to fix these gaps can be made as part of developing the implementation plans for your information management strategy.

The information strategy should be explainable to people with different levels of technical skill. The information management strategy is much more about how the company will handle data than about technology.



**Figure 8-8** Information architecture vision

## Develop a Realistic Strategy

Developing a realistic information management strategy is probably the most critical aspect of the entire process. You need to have a dedicated data governance committee (described later in this chapter), built in partnership between business and IT, and that has both the responsibility and authority to execute the strategy. Setting your vision or long-term goals is beneficial, but only the beginning. Now a series of phases need to be defined to begin implementing your information management strategy. Define the steps that need to be taken over the next three to five years to reach the goal. The most specific details will be included for the first year. This often sets forth a series of projects. Several sample projects include:

- Set up basic data governance policies and procedures, including the following:
  - Determine roles and responsibilities.

- Identify resource requirements.
- Create a data governance committee.
- Select the pilot project to adopt the new data governance policies and procedures.
- Review the use and value of existing reporting and analytical environments (data warehouse, data mart, performance scorecard, executive dashboards, etc.). Determine the long-term viability of each.
- Recommend business process changes to improve the quality of data entered—for example, by a travel reservations staff.
- Gather requirements for and build a new financial data mart.
- Deploy two additional business intelligence applications using data already loaded in the data warehouse.
- Determine the training needs for both business and systems staff related to information management.
- Define projects to address operational system changes, such as making enhancements to the policy issuance system to capture additional data.
- Evaluate the organization's needs and select tools and technologies to support the information management strategy.

### ***Sharing the Information Management Strategy***

The organization's information management strategy needs to be written down to ensure clear communication with others. It should not be too lengthy; otherwise, few people will take the time to read it. Write the strategy in business terms that can be easily understood.

The information management strategy should include diagrams that show the current data environment (refer to Figure 8-7) and the vision (refer to Figure 8-8). Each diagram should include a one-page summary to support it. The formal documentation provides a reference for anyone in the organization to learn about the strategy or simply refresh their memory.

The development of an information management strategy usually spawns a series of projects to address the data and information challenges facing the organization. A summary of each of these projects should be communicated with the information management strategy. These projects are the actions that help the organization achieve the vision it has defined.

Senior management must be briefed about the concepts. Because information management affects the entire enterprise, securing management's support is critical to the success of the strategy. Once the strategy has been approved and accepted, others in the organization must also be briefed. The same presentation should be given to managers, supervisors, data warehouse project

team members, and any others who are interested. This clarifies the project's direction to everyone and signals that senior management supports it.

The information management strategy is a management tool, not a design specification. Additional work will be done within each project that falls under the information management umbrella to get to the technical specification level of detail.

## **Setting Up a Sustainable Process**

Many organizations get excited about information management concepts and draft their vision for the future. Then, when reality sinks in about how much hard work is involved and how much compromise is required, many of these efforts never get off the ground. This section provides some guidelines for ensuring that doesn't happen to your vision.

### ***Enterprise Commitment***

Visible support from senior management is critical to ensure success. This means that projects need to be funded. Groups must be encouraged and rewarded for working together. Small successes need to be celebrated. All systems projects need to include some responsibility for participating in the organization's information management strategy. Each business group must be charted with the responsibility to participate and support the information management work, including working on data quality, taking ownership of the data, and documenting what the data means.

### ***The Data Governance Committee***

A data governance committee should be formed to set and enforce strategy, and facilitate communication and resolution of issues. The committee provides a mechanism to guide the process of implementing the information management strategy.

The committee should be headed by a business executive and contain representatives from each of the business areas, including subject matter experts who know a lot about the data from their area. Each representative should also have an understanding of how others use data that is created and maintained within their own group.

The committee should have a standard meeting schedule. This may need to be weekly as you get started. Over time, monthly meetings are typically sufficient. The purpose of the meeting is to communicate and coordinate data-related issues. This is not intended to be a working session to research problems or recommend solutions. If topics require research, the appropriate

people are assigned the work and are expected to report back to the committee with the results.

The committee needs to serve as a decision-making body. Many data-related questions require cross-functional input, which the committee should provide. The committee should have the authority to handle most issues on its own, but for significant issues the committee can recommend a solution to be approved by senior management. This should only be done for big issues with broad implications or significant impact on funding.

The data governance committee also serves to provide a common ground for communication. Business decisions are often made that impact information management and specifically the data warehouse. Examples of business decisions that can affect information management include offering new lines of business, or acquiring a new subsidiary that will begin using the existing inventory control application system. Often, the application team directly responsible for the system is informed of these changes, but other downstream systems and groups are not always informed. The data governance committee provides a forum to discuss what is happening in different parts of the business. Appropriate follow-up can be done to ensure that data flows properly throughout the rest of the organization.

### ***Revising the Strategy***

Keep in mind that while the information management strategy must have a long-term view, it also needs to have the flexibility for refinement and the business changes and/or systems changes. As the organization learns what works and what doesn't, the information management strategy should reflect this corporate knowledge. Too often, an information management strategy sets forth unrealistic and unattainable goals. If this happens, revise the strategy to reflect more reasonable goals. Take some time to develop the first cut and then try to implement the strategy.

Sometimes the challenges that crop up could never have been anticipated, even if years were spent to develop the information management strategy. Treat information management as a general direction. Point the organization in that direction and start moving forward. There will be detours along the way, but most progress will still be in the desired general direction. While you are planning next year's projects, which are proposed under the information management umbrella, this is a perfect time to ensure that the direction still meets the organization's needs.

## **In Real Life**

---

After learning about information management, let's take a look at how this is being handled by the companies we have been tracking.

## **A Glimpse into Giant, Co.**

As the company has grown over the last fifty years, a lot of structure and procedures have been put into place. There is already a strong, well-developed set of enterprise reference data. It is not perfect, but it does a pretty good job of keeping the core customer, product, employee, and supplier data clean. All of the major application systems that have been written in house use these master data sources. This provides great paths for integrating data across sources, but several smaller applications were developed quickly, with no interface to the shared reference data.

The organization excels at following standards. This is beneficial, but unfortunately the organization has lost the ability to think creatively. Staff members simply follow the rules, whether they make sense or not. It is easier to stay in line than to step up and question the rules. These standards typically have been refined with an operational system perspective. Flexibility and the need for business-driven data for the data warehouse does not always fit as nicely. Many project resources are wasted trying to challenge the status quo. Sometimes this does work eventually, but in other cases, the existing rules prevail. This can affect the value that can be delivered to the business community. There needs to be some level of flexibility and a mechanism to evaluate the applicability of standards. This could help many projects, not just the data warehouse.

Using the principles outlined in this chapter, an information management strategy must be developed. The organization will be able to utilize much of what has been developed to date. The addition of a data governance committee and procedures will help the data warehouse team by providing a vehicle to raise issues and the needed support to come to resolution.

## **Insight from Agile, Inc.**

The organization prides itself on its entrepreneurial spirit. Do what you need to get the job done. This provides a high level of energy and creativity. However, this also has resulted in the organization having many different silos of business processes and application systems. In addition, many of the application systems are third-party packages that helped to get things running quickly.

This lack of coordinated effort and the resulting disparate systems has resulted in extreme challenges for data integration. This has also had a negative impact on data quality. Each project must fend for itself.

Starting down a more organized information management path is foreign to both the business and IT staff. This will only be successful if specific business problems can be addressed by the strategy. If no business problem is identified and no specific business value is delivered, then any information management efforts will fail. However, the organization has encountered many growing

pains and the disparate systems and data are now getting in the way of meeting customer needs. Now is the time to get started.

This will require education for everyone. Both the business and systems must change how the daily work is done. Structure and coordination are needed to ensure that everyone is working together. There will be many growing pains and false starts, but with enough commitment and fortitude, the organization will be able to make this cultural shift and leverage data across the enterprise. Communication and continuous refinement of the strategy will be critical to attaining the organization's information management goals.

### **IMPORTANT QUESTIONS TO ASK**

**This chapter reviewed many different aspects of information management. The following questions may help you to understand the information management strategy and procedures within your organization:**

- **Does your organization have a data governance or data management group?**
- **Where does the data management responsibility fall—business or IT?**
- **Do you have a specific strategy for information management?**
- **Are specific projects and tasks defined to get started with implementing your information management strategy?**
- **Do system development projects tie into this strategy?**
- **Who within the business is responsible for the information management strategy?**
- **List what you expect information management to do for you.**
- **List what business and IT issues and challenges you face regarding data.**
  - **Who is currently assigned to address these issues?**
  - **Who do you think should be working on these issues?**
- **What type of checks and balances are in place for data governance?**
- **How is data ownership assigned?**
- **What is in place to measure and improve data quality?**
- **Who is responsible for the design and evolution of the data dictionary?**
- **Have there been any previous attempts to develop an information management strategy? What has been tried for the following:**
  - **Data governance**
  - **Data quality**
  - **Implementation of a data dictionary**
- **Why were these attempts successful or not?**

## Summary

---

The ability to truly have your data work as a valuable corporate asset is highly dependent on your organization's ability to forge working partnerships between business and IT. Careful attention to information management strategies can provide a clear direction for the organization's data needs. All IT and business professionals can use this as a set of overarching guidelines. Information management is not a one-shot project, and it will never be complete. Getting started can be difficult and somewhat daunting, but it is worth the effort in the long run. This chapter has focused on the value of data; the next chapter discusses data warehouse architecture and tools.



# PART

# IV

## Building the Project

---

### In This Part

---

**Chapter 9:** Architecture, Infrastructure, and Tools

**Chapter 10:** Implementation: Building the Database

**Chapter 11:** Data Delivery: What You Finally See



# Architecture, Infrastructure, and Tools

This chapter is intended to introduce a nontechnical audience to architecture concepts. If you are a systems architect, this chapter is not intended for you. There is a lot of discipline and rigor associated with systems architecture, but these topics are beyond the scope of this book and are not covered in this chapter. This chapter will do the following:

- Provide an introduction to architecture
- Outline the benefits of architecture
- Discuss the basics of data architecture for a data warehouse
- Review the technical architecture for a data warehouse
- Share the fundamentals of finding tools that will work for your organization

For readers who are interested in gaining more in-depth knowledge of architecture, many fine resources are available. A search for books on enterprise architecture will yield many options. Several books specific to data warehousing that include architecture concepts include the following:

- *The Data Warehouse Lifecycle Toolkit, Second Edition*, by Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker (Wiley, 2008)
- *Building the Data Warehouse, Fourth Edition*, by William Inmon (Wiley, 2005)
- *Mastering Data Warehouse Design*, by Claudia Imhoff, Nicholas Galemmo, and Jonathon Geiger (Wiley, 2003)

## What Is Architecture?

---

There are many different definitions for architecture. Rather than introduce another one, a search on system architecture at [www.dictionary.com](http://www.dictionary.com) returns the following:

*A system architecture or systems architecture is the conceptual design that defines the structure and/or behavior of a system.*

There is no universally agreed upon definition of which aspects constitute a system architecture, and various organizations define it in different ways, including the following:

- The fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution (from ANSI/IEEE 1471-2000).
- A formal description of a system, or a detailed plan of the system at component level to guide its implementation (TOGAF).

Two different layers of architecture are addressed for a data warehouse in this chapter: *data architecture* and *technical architecture*. Data architecture is focused on where data is stored, how it is structured, and how it is intended to be used. Technical architecture provides the road map showing what technical capabilities are needed to perform the many different functions, from extraction through to end user access. The technical architecture is what defines hardware platforms, operating systems, databases, transformation tools and data access, analysis and reporting. This is how you can ensure that all of the capabilities you need are provided and that each of the parts fit together to provide the complete set of tools needed to develop, implement, and use your data warehouse.

Whether you are talking about data architecture or technical architecture, each must be put into the proper context. If you have a single project with a single data feed to your data warehouse, then the magnitude of the challenges and problems is greatly reduced. The larger the organization, and the more independent and autonomous the different parts of the business are, the harder it is to define and implement any architecture. The data warehouse architecture should encompass the needs across the entire organization.

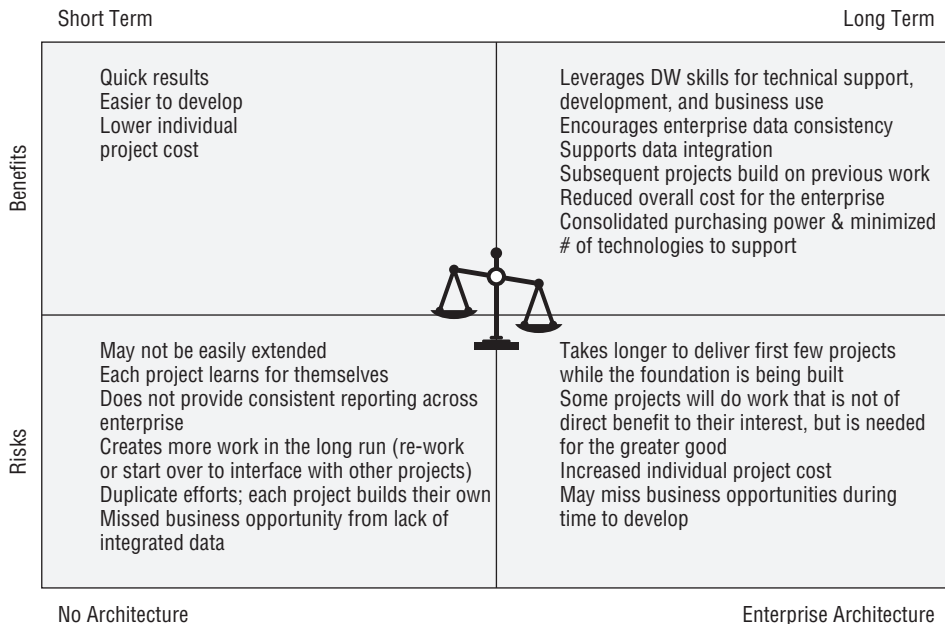
## Why Do We Need Architecture?

---

The reason to develop and use architecture in systems is to impose overall order on what can turn into chaos relatively quickly. It helps an organization to develop a common understanding of the data warehousing components

as implemented within the organization; leverage industry best practices; empower individual projects to work independently; provide the blueprint for an integrated view of the business; and minimize duplication of effort between projects.

Figure 9-1 describes the delicate balance of architecture. Both short-term and long-term risks and benefits must be considered. In today's tight, competitive marketplace, long-term investments are often abandoned for immediate results. While this may seem appropriate at first glance, it can set up all data warehouse initiatives for failure. Each different group is scurrying to meet its own business goals. This can spawn many different projects designed to support each individual group; and while a specific group's immediate needs may be met, the overall cost to the organization is larger due to duplicate efforts. The benefits of having an enterprise data warehouse architecture can be realized only when that architecture is followed.



**Figure 9-1** The balance of architecture

Without enterprise data warehouse architecture, each project does what is necessary to get its work done, resulting in multiple independent applications that do not integrate or work together. These are often referred to as *stovepipe applications*.

This also sets up a pattern of revisions and a great deal of rework that could have been avoided. Project teams often observe that there is not enough time to do it right the first time, but there is always time and money to do

it again. It is important as a management team to be in touch with what is really happening on projects. Individual project team members often wonder whether this tendency to rework projects is really what upper management wants, or whether this is simply a pattern that has emerged without conscious thought over time.

**NOTE** The biggest data warehousing disappointment occurs when an organization that lacks architectural direction results in the development of multiple non-integrating data warehouses or data marts. This lack of purpose fails to address one of the fundamental goals of data warehousing: supporting a single version of the truth.

#### **A CASE STUDY: THE NEED FOR ENTERPRISE RESPONSIBILITY**

The term *enterprise* is often interpreted as an extremely large organization with multiple, disparate lines of business that run independently. This type of large enterprise faces challenges with coordination and collaboration when working on enterprise systems development efforts, including a data warehouse. Similar coordination and collaboration challenges exist for small and medium-size companies too.

For example, a mid-size financial institution grew through acquisition, but the company continued to run each area autonomously. The business focus of each division was different enough that folding all operations together did not make sense. Each division was run as a separate company, each with its own budget and business goals. Within each division, there was no need to look at data from the other divisions. With a lack of overall enterprise requirements, architecture, or accountability, each division continued on its own path to develop solutions for its reporting needs. Each division was at a different stage of exploration or implementation of data warehousing capabilities. The focus of each was to support its own division.

As an enterprise, there was increasing need to look at more than just high-level financials across the organization. Within the organization structure, the people who needed this integrated view were the executive team. Although the audience is small in number, these are the users who have the greatest need for information to support the decision-making process as they set the direction of the organization.

Exploration of each division's business showed that while each division has a unique niche, they have some customers in common. Moreover, the different divisions have interests in similar industries too. By taking an enterprise view, the organization was able to leverage its customer relationships and cross sell products. The enterprise view also enabled the enterprise to identify industry-specific trends over time to head off what could have been big problems. The impact on each division was not large enough to cause concern, but when added together, the impact could have been severe.

---

Of course, an architecture is only helpful to the extent that it is adopted and used across the organization. The next section shares ideas for how to make this happen.

## **Making Architecture Work**

Architecture can provide great benefits to an organization, but it must be developed with common sense. It is easy to lose sight of the overall purpose, which is to provide structure to support the design and development of application systems. It is easy to get swept up in the realm of architecture and define strict, rigid guidelines that can impede the ability of a project team to build a system. Architecture should support and help teams, not prevent them from meeting their goals.

In addition, any architecture must continue to grow and change with the organization. It is critical that the architecture be flexible and adapt over time to ensure its value to the organization. A rigid architecture is often disregarded and never used. This is actually worse than no architecture, as time and resources were spent to develop it. Developing an architecture that both provides order and is able to help project teams deliver real solutions is a challenge.

Companies are under a lot of pressure to meet the many demands facing businesses today. It can be difficult to invest in long-term strategies, such as implementing architectures, when dealing with serious short-term problems. However, ignoring architecture is not the answer. Both business and systems management have some responsibility for architecture. From a systems perspective, it is important to ensure that your architecture is reasonable and that appropriate procedures are put in place so that project teams can utilize the architecture. All projects need to tie into the architectural vision. Many companies are good at these systems-related responsibilities. Unfortunately, many companies stop there.

Business management must also actively support the architecture. This means that everyone must be willing to make accommodations to ensure that the architecture is used. Each individual project must be not only allowed but actively encouraged to adopt the architecture. When getting started, this may add some time or additional expense to a project, but it should save time and money in the future.

As a joint business and IT management team, you need to work together to do the following:

- Understand the need for and purpose of the data warehouse architecture
- Support the enterprise viewpoint
- Not penalize projects or individuals who are taking the time to address enterprise-level needs

- Not pressure project teams such that immediate needs overshadow long-term benefits
- Look out for the long-term success of the organization
- Be willing to invest (time and money) for the greater good of the enterprise
- Understand the impact of short-term decisions on the long-term viability of the data warehouse
- Realize that many of the enterprise-level challenges have taken years to develop and will not be solved in the short term. These will never be solved if you don't get started.

Project teams can be easily coerced to deliver short-term solutions. Budget and schedule constraints are key motivating factors for most project managers and teams. This means that long-term enterprise concerns are pushed aside in the interest of meeting individual and group goals: In other words, get the project done yesterday. This is where the management team must step in to actively support efforts to develop long-term solutions.

Now that a general background about architecture has been provided, it is time to look more closely at data warehouse architecture. Recall that two different architectures are discussed in more detail: data and technology. While these two work together, each must be considered independently.

## **Data Architecture**

---

*Data architecture* addresses how the data is organized and structured to support the development, maintenance, and use of the data by application systems. The purpose of data architecture is to create a blueprint showing how the organization is going to process and store data.

While a comprehensive data architecture will encompass all operational and data warehouse data, this chapter focuses only on the data architecture for your data warehouse. This serves a different purpose and should be handled separately from your operational data architecture. A separate, distinct effort should define the data architecture for operational systems. These architectures interface with each other when data is extracted from the operational systems and may both interact in the area of master data management.

Advances in technology have made it easier and faster to move data than ever before. It is possible to keep more history and more details, and to develop prototypes quickly. The technical capabilities lure you into believing that you can do anything. Unfortunately, reality dictates otherwise. It is important to stop and ask whether all this really gets you where you want to be. More



important, where do you want to be? The data warehouse data architecture provides the vision for the organization's data.

How data is handled and structured in the data architecture is independent of your technology. This is why you need to pay such careful attention to your data architecture. Sound data architecture gives you flexibility and avenues for growth, and gracefully accommodates technical changes over time.

Although many organizations are getting started with data warehousing, the early adopters have been doing this for over two decades. In some cases, the original data design has been moved to a different database platform twice since the initial implementation. You don't go in thinking that you will change out your database, but if you have ten to fifteen years to reflect on, you probably have already done so. Who can predict what technology might be available in fifteen years? The bottom line is that you need to expect success and plan accordingly. This means ensuring that your data fits together properly across the enterprise.

## Revisiting DW Goals

There are many different opinions about, and approaches to, handling and structuring data in a data warehouse. Each different approach is trying to address a common set of objectives. Regardless of the data architecture, the goals are as follows:

- Extract, clean, integrate, and make the data available to the business in a timely manner.
- Be able to adapt to changes at any point in the process.
- Address enterprise integration requirements.

While these goals can be succinctly stated, they are actually complex and have broad implications—implications that are specific to your organization and the current state of the data. Examples of specific challenges that need to be addressed for the first goal include the following:

- The need to bring together data from disparate claims handling systems.
- The primary source system does not have a sound database design, which results in data being scattered throughout the database; duplicates; and inconsistencies.
- Extreme data volumes make it difficult to meet performance expectations for loading and/or querying data.
- Source systems do not retain history.
- There is no common definition of a product (this is common in financial services companies).

**NOTE** Many of the data challenges facing an organization have developed over a long period of time. Do not expect them to be resolved quickly and without effort.

For the secondary goals, the types of change that commonly occur include the following:

- Expansion into new businesses
- Changes in the marketplace
- Changes or enhancements to an underlying source system
- Replacement/rewrite of an underlying source system
- Mergers and acquisitions
- Need for data that was not originally requested or identified
- New groups of people who want to use the data
- Refinement of business rules, definitions, and groupings of the business
- Clarification about how you want to look at and measure the business itself
- Regulatory changes

The goal is to be able to gracefully adapt to these changes without having to completely rewrite the processes to build the data warehouse. A flexible design improves the time it takes to incorporate these changes.

The last goal is often the most difficult. Each organization must evaluate the need for, benefit of, and work required to achieve enterprise integration. The level of awareness of enterprise requirements is often related to where you fit in the organization. Individual project team members usually are focused on their immediate deliverables. From a systems perspective, pockets of people can be found who have enterprise concerns. These are often found in the architecture and database administration groups, which are responsible for looking after enterprise resources.

From a business perspective, business analysts typically concentrate on getting their reports and analyses completed on time. Often, you must look higher and higher in the organization to find the business people who have the need for enterprise integration. Decisions about enterprise integration must be made at these high levels in the business. The individual team members can help estimate the work involved, but the magnitude of the need and the benefits must be articulated by upper management. Without a clear management mandate and ongoing support, enterprise integration is difficult, if not impossible, to achieve.

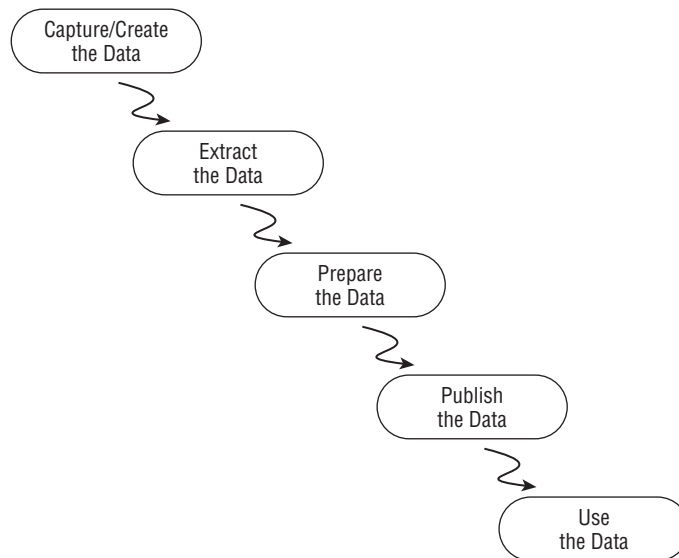
**NOTE** Don't ignore topics that are discussed in terms of the entire enterprise. Too often this is assumed to be applicable only to very large multi-divisional

**companies. Enterprise means across your entire company, regardless of the size. Similar problems have been observed for organizations of any size, from single line of business to global conglomerate.**

Data architecture provides the umbrella to help an organization meet these overarching data warehouse goals. It is time to look more closely at the different parts of the data warehouse data architecture. Keep in mind that the specific business and systems requirements of each organization must be taken into consideration when developing the data architecture.

## Components of DW Data Architecture

Any architecture can seem daunting to understand, but it does not need to be. As shown in Figure 9-2, the overall flow of data in a data warehouse consists of several basic layers, which need to be defined by the data architecture.



**Figure 9-2** Layers of data architecture data flow

The first layer of the data architecture is for capturing or creating the data. This is not within the realm of the data warehouse, but is done by the source systems. The specific architecture of the source systems is determined independently of the data warehouse, but this is where you begin. The second layer is the process to get the data from wherever it currently lives in the organization. This may be a simple extract from a database or a complex set of programs to pull the necessary data, or a utility to pull the data from a third-party application such as a financial system. The third layer is to prepare

the data for reporting and analytical use. This is the where most of the work for a data warehouse is done including validation, cleansing, and integration. The fourth step of data architecture is to publish the data that is ready for access. Here, data is made available for reporting and analysis. The final step is to begin using the data itself.

At each layer, several basic things need to be defined regarding the data:

1. What data will be stored here (reference and/or transaction data)?
2. What is the primary purpose of keeping the data here?
3. How will the data be structured?
4. What is the persistence of the data or how much history will be stored?  
If it is no longer needed, how is this handled?
5. Who will be able to use the data here?
6. What type of data access will they have?

There are many different alternatives for a data warehouse data architecture. A complete data architecture defines parameters and guidelines for the preceding questions for each of the layers illustrated previously in Figure 9-2. A review of the most widely adopted data architectures follows in the next section.

## **A Closer Look at Common Data Warehouse Architectures**

While there are many philosophical approaches to data warehouse architecture, two are by far the most widely adopted: the *bottom-up* approach and the *top-down* approach. These were introduced as Kimball/Inmon in Chapter 1, and each is explored in more detail here. Both are viable alternatives for implementing your data warehousing environment. It is important to understand the basic principles of each in order to select the approach that will work the best for your organization. Having a good grasp of the most prevalent data architectures can improve your understanding of any other approaches that your organization may consider or choose to adopt.

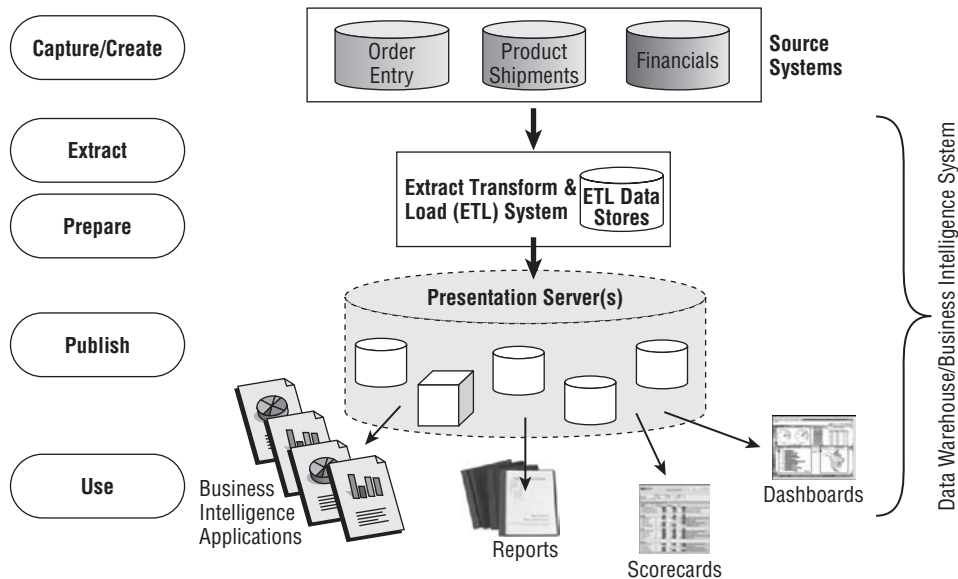
These data architectures are each defined within the context of an overall data warehouse methodology. The data architectures described next are pulled from each methodology.

### ***Bottom-Up Data Architecture***

The bottom-up approach has been well documented by Ralph Kimball and his colleagues in his *Data Warehouse Toolkit* series of texts. This approach has a strong focus on the end user and the delivery of value to the business. The goal is to ensure support across the enterprise for consistent reporting and analysis.

Much of the emphasis has been on how the data must be structured in order to support business reporting and analysis. This sets up the vision for how the data will work together today and into the future. The entire end-to-end architecture is called the *data warehouse/business intelligence system*.

The vision for data delivery must be supported by robust capabilities to build and maintain these data structures. Careful thought and planning is required to ensure a sustainable set of processes to build and maintain these integrated dimensional structures. Figure 9-3 shows the bottom-up architecture layers.



**Figure 9-3** Bottom-up data architecture approach

Each of these steps is described as the data flows from the source system; however, they are typically designed in the reverse. The first thing to be defined is the target dimensional data structure to support the business. Then, the work begins to determine what is needed to build a sustainable process to maintain these structures.

It is important to understand that integration of data across the enterprise is a foundation of this data architecture. This integration must be ensured from perspective of end user access. This is accomplished using conformed dimensions, as described in Chapter 7.

### Capture/Create the Data: Source Systems

Data is extracted from the underlying source systems. These may include major operational application systems, departmental databases, or data purchased

from third-party sources. The data architecture of this layer is determined by the source systems.

*The source system data architecture is the same regardless of the data warehouse data architecture approach.*

### **Extract the Data**

This is a step that is part of the extract, transform, and load system. This is simply the first step of gathering the data to feed to ETL processes. There is no specific data store defined for this layer. If the underlying source systems do not retain history, then the raw input files may be backed up for future use. Archiving the raw input files (by the source system or here) is useful to support reprocessing of the data if necessary over time.

*There are not substantial differences between the two data architecture approaches for data extraction.*

### **Prepare the Data: The ETL Data Stores**

This is where the hard work of building the data warehouse/business intelligence system begins. Many different functions are performed on the data in order to get it ready for use. The data architecture focuses on how the data will be stored, rather than the functions themselves. The functions are discussed in more detail in Chapter 10. Here, the data architecture of the ETL system is outlined:

1. What data will be stored here (reference and/or transaction data)? Data needed for running the ETL system itself will reside here. Often reference data is used for validation and preparing dimensions. The transactions data will flow through the system but will not necessarily be stored here.
2. What is the primary purpose of keeping the data here? Data is only kept here to keep the ETL system running—for example, data used for validation and audit tracking.
3. How will the data be structured? The data should be structured in a manner to expedite the periodic processing of the ETL system itself. Often this is achieved with simple flat-file or single-table structures. If the ETL challenges are great enough, more rigorous data structures may be used to facilitate ETL processing.
4. What is the persistence of the data or how much history will be stored? Reference data needed to ensure data quality will be kept. This could include a master list of products and customers. It is important to retain the keys that are assigned during ETL processing. Most transaction data will flow through the ETL system. Only the transactions necessary to successfully process the next periodic load will be retained. For example, 30 days of sales transactions may be retained in order to apply sales

adjustments appropriately. In this example, almost all adjustments flow through within two weeks, so 30 days accommodates the business need.

5. Who will be able to use the data here? Access to the data in the ETL data stores is limited to the programs and utilities that comprise the ETL system. The sole purpose of the ETL system is to prepare the data for reporting and analysis. No business users are allowed to access and query this environment. This allows pre-processing of data, tuning for batch processes, and eliminates the need for user security. The entire focus is on efficient preparation of the data.
6. What type of data access will they have? Access is through programs or utilities to run the ETL system itself.

*Both the bottom-up and top-down approaches utilize ETL data stores to efficiently and accurately process the data. However, the top-down approach also includes a physical data store here, called the data warehouse.*

### **Publish the Data: The Presentation Server**

The purpose of the presentation server is to deliver the data to the business user or analyst. Following are the primary considerations:

1. What data will be stored here (reference and/or transaction data)? All of the transaction detail data will be stored here. This is needed to support the business. Core reference data will be stored as dimensions here. All data that has potential business value should flow from the source system through the ETL system and be loaded here.
2. What is the primary purpose of keeping the data here? The data is to be used for business analysis and reporting. This includes basic querying, reporting, scorecards, dashboards, business intelligence, and analytical modeling.
3. How will the data be structured? The data is to be organized in dimensional structures. These may be star schemas in a relational database technology or inherent in a multidimensional cube technology. The dimensions are to be conformed to enable integration at query time across the enterprise.
4. What is the persistence of the data or how much history will be stored? All data needed to support the business requirements will be stored in these dimensional structures, including atomic transaction detail.
5. Who will be able to use the data here? Data in the presentation servers is organized to support many business users for their reporting and analytical needs. The data is intended to be accessed with interactive queries by the business community. Appropriate security measures must be put in place based upon business requirements to ensure that only authorized users are able to access the data.

6. What type of data access will they have? Most business users will have interactive query capabilities. Some may be recipients of more structured, static reports. Again, this is driven by the business requirements.

*The primary data storage for the bottom-up approach is here, in the presentation servers. These are architected using conformed dimensions as a foundation for this approach. The top-down approach may also physically store data in a similar manner, called data marts. Conformed dimensions are not required with the top-down approach, but they are highly recommended. If this part of the top-down approach is implemented with conformed dimensions, then this layer is quite similar to the bottom-up approach. With the top-down approach, this part of the architecture may also be implemented with a semantic layer on the data warehouse itself, rather than physically storing the data.*

### **Use the Data: Business Intelligence Applications**

Technologies used to access data today often include some mechanism to store data as part of how the tool works. The data architecture for using the data is dictated by how the organization's data access tool(s) function. The organization must understand each tool, and which is being used, so that appropriate measures can be taken to ensure that the data continues to be handled in a consistent, reliable manner.

*Both the bottom-up and top-down approaches have similar data architectures here, as it is driven by the data access technology.*

The data architecture role here is to ensure that the organization has a clear understanding of what data is stored as part of the business intelligence application. The goal is to have a structured, organized approach here too. For a technology that requires the creation of a multi-dimensional cube of data, a new cube should not be created for every new report that is needed. Rather, more robust cubes should be created to support a variety of reporting needs. These application-specific data stores must also have planned maintenance, including updates, corrections, and additions to the data. The data architecture must also recognize the need for backup and recovery of these data stores.

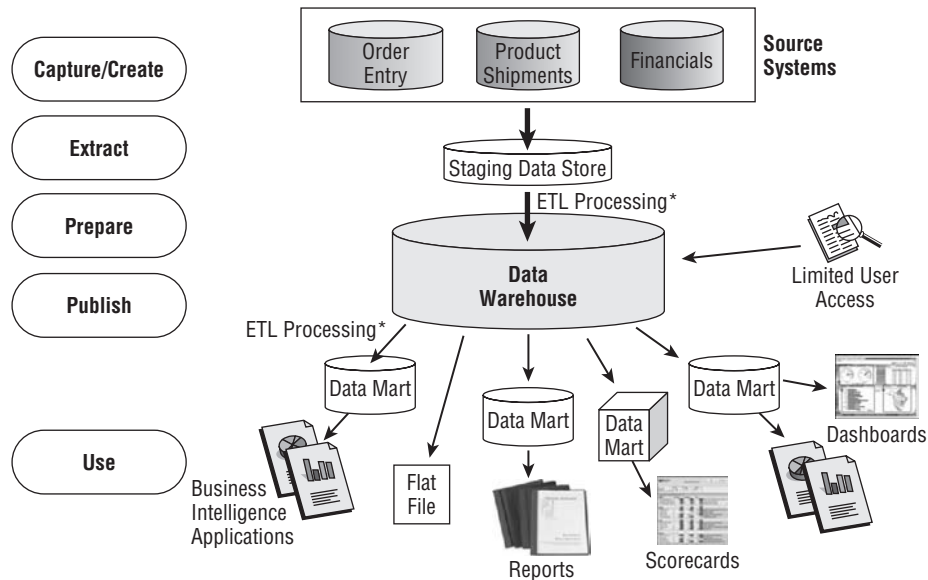
Now that the basic data architecture of the bottom-up approach has been discussed, it is time to take a similar look at the top-down approach.

### **Top-Down Data Architecture**

This approach was developed and popularized by Bill Inmon and Claudia Imhoff. Several other key individuals have also contributed to the refinement and success of this approach, which you can learn more about in *Corporate Information Factory, Second Edition*, by W. H. Inmon, Claudia Imhoff, and Ryan Sousa (Wiley, 2001) and *Mastering Data Warehouse Design* by Claudia Imhoff, Nicholas Gallemmo, and Jonathon Geiger (Wiley, 2003).



The most successful implementations of the top-down data architecture are grounded in business requirements. The overall approach is iterative. With each iteration, the scope of data is set based upon business objectives and requirements. Then work begins on the data. The data sources are identified and the data is extracted. Each layer of the data architecture is defined, starting with the source and moving forward, designing each step along the way—hence, the top-down label. Once the defined scope of data is addressed, a subsequent iteration is defined and the process repeats. Figure 9-4 shows the basic data architecture for the top-down data architecture approach.



\*Note: ETL Processes may include ETL Data Stores

**Figure 9-4** Top-down data architecture approach

### Capture/Create the Data

Data is extracted from the underlying source systems. These may include major operational application systems, departmental databases, or data purchased from third-party sources. The data architecture of this layer is determined by the source systems.

*The source system data architecture is the same regardless of the data warehouse data architecture approach.*

### Extract the Data: The Staging Data Store

Extracting data from the source systems is simply the first step in the overall ETL system. The goal here is to acquire the data from across the organization

as input to the rest of the ETL processing. This provides the raw input data to be processed, but not the other data needed for ETL processing. Where the extracted data is stored is referred to as the *staging data store*. The definition of this staging data store in the top-down architecture follows.

1. What data will be stored here (reference and/or transaction data)? This layer contains all of the raw input data, including both transaction and reference data.
2. What is the primary purpose of keeping the data here? The data is stored here as input for ETL processing. This may also be used as a source for the raw input data if the source system does not retain history. This may also serve the purpose of preserving data as provided by the source system. This supports reprocessing of the data if errors are found in the business rules or there are changes to how the data is to be processed. How will the data be structured? The data is structured as it is provided. This may be a copy of the source system structures or an extract file.
3. What is the persistence of the data or how much history will be stored? The amount of data that must be retained in the data staging area depends upon the characteristics of the underlying source systems. If the data can be easily provided again, then there is limited value in keeping another copy of the data here. However, if the data rolls off the source system database and is no longer available, there may be good reason to keep a copy of the data here. Even if the raw input data needs to be retained for possible future use, it may be more cost effective to archive the data, rather than keep it online. Data that has been archived can be retrieved for reprocessing if necessary.
4. Who will be able to use the data here? The data staging area is to be used by the data warehouse processing, either programs or utilities.
5. What type of data access will be available? The access will be in a batch mode, when the data warehouse processes are run. This may be monthly, daily, or perhaps more frequently if the business requires it.

*There are no substantial differences between the two data architecture approaches for data extraction. More definition is put around this area with the top-down approach.*

### **Prepare the Data: The Data Warehouse**

A lot of work is required to prepare the data for reporting and analysis. The top-down approach recommends creating a collection of the historical data as the basis for publishing and using all data. This collection of data is called the *data warehouse*. Data is cleansed, validated, and integrated prior to being loaded into this data warehouse. As noted in Figure 9-4, ETL data stores are needed to support the processing of the raw data in the staging data store for loading

into the data warehouse. The content and purpose of these ETL data stores is similar to those described for the bottom-up approach. The characteristics of the data warehouse are:

1. What data will be stored here (reference and/or transaction data)? Historical data, needed to satisfy business requirements, will be stored in the data warehouse. This includes all historical transactions and history of all reference data that can be captured within the scope of what has been addressed to date.
2. What is the primary purpose of keeping the data here? The purpose is to facilitate the integration of data across the enterprise and to provide a single platform to store the data. The data should support analytical and reporting needs of the organization. The data warehouse serves as the central component of the enterprise solution. All of the data is loaded here and then distributed as needed.
3. How will the data be structured? The data structures follow traditional, sound data design principles. The data is organized into subject areas. It is organized in a traditional third normal form. The modeling technique for this data organization is called entity-relationship modeling (see Chapter 7). This data represents the pure data relationships, and when done properly eliminates redundancy.
4. What is the persistence of the data or how much history will be stored? In theory, all of the history that may be needed by the business is to be retained in the data warehouse. All of the historical transactions are retained here. The data warehouse also includes all changes to any reference data for the entire period of time that the transaction data is needed. Data may be archived or deleted from the data warehouse when it is no longer needed by the business.
5. Who will be able to use the data here? Theoretically, business users may access the data in the data warehouse. However, the complexity of the data structures can make it difficult to write queries, and the data is not organized for optimal query performance. It is not recommended that large numbers of users access the data warehouse structures directly. Data that is needed by many people should be published, as described in the next layer. There are probably a handful of highly sophisticated, data-savvy users with strong technical skills who are capable of accessing the data warehouse directly. These are often market researchers, actuaries, or statisticians.
6. What type of data access will they have? The primary access to the data warehouse is through programs and utilities used to distribute data to the data marts. Limited user access to these tables is possible, but not recommended for ad hoc querying.

*Both the bottom-up and top-down approaches utilize ETL data stores to efficiently and accurately process the data. However, the bottom-up approach does not include the data warehouse data storage as defined here. The primary data storage for a bottom-up approach is in the next layer.*

### ***Publish the Data: Data Marts***

In order to more directly support each business area, a collection of the data needed for that specific area can be loaded into a data mart. Each data mart is intended to directly support the needs of a specific business group. It may be possible to define a semantic layer on top of the data warehouse to provide the data mart view. If so, the data is not physically moved from the data warehouse. The data architecture would reflect this here. It is also common for data to be physically moved from the data warehouse into data marts. This requires another layer of extraction, transformation, and loading (ETL) to move the data from the data warehouse to the data mart(s). The following describes the characteristics of this physical movement of data. As noted in Figure 9-4, additional ETL data stores may be needed to support ETL processes to load data marts. In some cases, the structures of the data warehouse itself may be modified to support the processes to load and maintain data marts. *Mastering Data Warehouse Design*, by Claudia Imhoff, Nicholas Gallemmo, and Jonathon Geiger (Wiley, 2003), is a good resource to learn more about these techniques.

1. What data will be stored here (reference and/or transaction data)? The data needed to support the specific business analysis is to be loaded. Theoretically, only summarized data is needed because the detailed transactions are in the data warehouse itself. In reality, all frequently used transaction data is loaded into the data mart. The reference data needed to support analysis is also loaded into the data mart.
2. What is the primary purpose of keeping the data here? The data marts are designed to support specific business groups or analyses.
3. How will the data be structured? The data is usually stored in dimensional structures in the data mart. The dimensional structures may be physically stored in a relational database or a multidimensional cube (this is a characteristic of the technical architecture). Sometimes data is stored in structures that are specifically tailored to the business application. For example, a flat file may be created to support a specific insurance catastrophe modeling software package. If a semantic layer is used, rather than physically moving the data, then that should be noted here.
4. What is the persistence of the data or how much history will be stored? The historical data requirement varies according to the specific business

need for this data mart. This could result in the same data being loaded into multiple data marts, with different amounts of history. For example, actuaries may use ten years of claims history, whereas claims management may only use five years.

5. Who will be able to use the data here? The data marts are designed specifically to support many business users for their reporting and analytical needs. This may be thousands of customer service representatives reviewing prior day performance, hundreds of suppliers tracking their progress toward annual goals, or an executive dashboard reviewed by a handful of executives to track corporate performance.
6. What type of data access will they have? The access is direct to the data mart via a data access or query environment or perhaps through interactive reports published through the company's intranet.

*Because integration is achieved in the data warehouse itself, the top-down approach does not require integration of these data marts. However, the use of conformed dimensions is recommended to enhance consistency and reduce redundant efforts. When the data marts are physically built using conformed dimensions, this is similar to the presentation server layer of the bottom-up approach. With the top-down approach, this part of the architecture may also be implemented with a semantic layer on the data warehouse itself, rather than physically storing the data.*

### **Use the Data: Business Intelligence Applications**

The characteristics of using the data are exactly the same for both the bottom-up and top-down approaches. The data architecture is driven by how the data access tool itself functions. The reason why it should be included in the overall data architecture is to ensure that appropriate measures are put in place so that the data continues to be treated in a consistent, reliable manner.

*Both the bottom-up and top-down approaches have similar data architectures here, as it is driven by the data access technology.*

### **Adopting an Architecture**

The IT professionals select or create the data architecture with the input of the business community. This chapter does not provide a step-by-step process to create the data architecture, but rather provides the background necessary to help the business and IT management understand the architecture. There needs to be a cross section of people involved in defining the data architecture. This group should have representation from the ETL and BI application developers because they often have the deepest understanding of many of the data-specific challenges facing the organization. It is also important to keep the overall business objectives of the data warehouse in mind.

Both the bottom-up and top-down approaches have developed strong advocates, and both have hundreds of implementations. For each, a few are wildly successful, many are moderately successful, some struggle, and a few fail outright. One of the biggest hurdles in defining an architecture is that many people become entrenched with one approach and are unable to look objectively at the overall situation. You want to involve intelligent, strong people, but they must still be able to consider alternatives to make the right decision for the organization.

Established industry approaches provide the benefit of the culmination of many years of experience. However, several common mistakes are made when adopting the approaches described here.

### **Common Mistakes Adopting the Top-Down Approach**

- It is too easy for IT to fall into their comfort zone and perform traditional entity-relationship modeling without frequent, ongoing communication with the business.
- The project can get caught up in foundation work with no line of sight to any real business purpose.
- Organizations can build the data warehouse foundation with no end user deliverable. The project can lose funding before finding a worthwhile business problem to solve.
- It is too easy to build non-integrated data marts. These marts are tailored to different business needs and the data is often transformed differently for each mart. This creates multiple, usually differing versions of what should be the same data.
- Building an entire data warehouse prior to delivering any business intelligence applications takes too long and costs too much.

### **Common Mistakes Adopting the Bottom-Up Approach**

- Building dimensional data marts that do not integrate.
- Too often, projects pull only data needed to support a specific report or target.
- Quickly populating the data marts without a sustainable set of processes to maintain them.
- Abandoning all system development rigor to move data faster.
- Moving data into dimensional structures and then asking whether this would be useful. This is often done with frequent iterations and without careful data analysis and an understanding of real business needs.

### Success Factors for Both Approaches

The mistakes with either approach are at odds with their true intent and purpose. Many of the risks outlined above can be mitigated by ensuring that all data warehouse efforts are grounded in business requirements. Studying the differences between the data architectures also yields insight into the common themes that contribute to success. These include the following:

- Pull all the business data from source systems.
- Keep focused on end users and how to deliver value.
- Develop the overall strategy and then implement in phases.
- Use common sense—make sure you understand what you are trying to accomplish and make decisions accordingly.
- Build the subject areas or parts that you need as you go. Add on later.
- There is no substitute for careful, detailed data analysis. You need to understand the actual data in the context of the source system. You also need to know how the business wants to look at the data.

This is not intended to be a complete list of success factors, but the things that should be kept in mind while making decisions about data architecture. This entire book is intended to help organizations learn techniques to be able to successfully design, implement, and use a data warehouse.

Now that data architecture has been discussed, it is time to move on and take a look at technical architecture.

## Technical Architecture

---

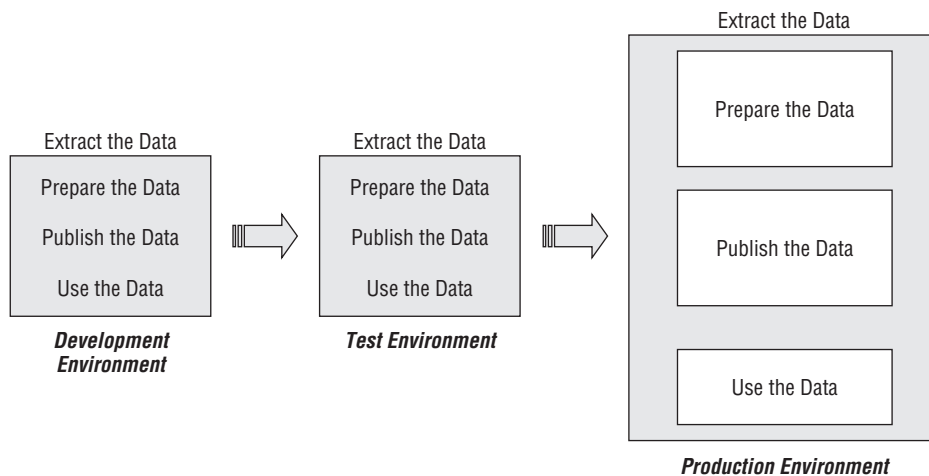
*Technical architecture* addresses the organization and structure of the collection of hardware and software technologies that are installed to support the development and delivery of the data warehouse.

This is the area where change is constant. Existing technologies are enhanced and new technologies emerge. Change is continuous for both software and hardware options. This text does not address specific technical products. Regardless of what happens with technology, the basic functionality needed to get the work done remains constant.

The responsibility for technical architecture belongs squarely with IT. The IT professionals have the education and skills needed to make solid technical recommendations. From a business perspective, you need to ensure that these recommendations and decisions are being made wisely. The following sections describe what you need to know about technical architecture.

## Technical Architecture Basics

While the specific technical specifications can be left to the IT professionals, it is important to ensure that the overall environment is set up for success. In fact, three separate environments are needed, as shown in Figure 9-5.



**Figure 9-5** Recommended technical environments

There should be a separate technical environment for development, test, and production work. This is no different from what should be in place for any other systems development environment. The production environment is needed to ensure the stability and reliability of the data warehouse. This is what the business depends on. Development work is often divided into smaller parts. Creating a separate development environment enables work to be done on each of these parts at the same time. Once all the individual parts are completed, the entire system can be moved to the test environment, where testing can be done to ensure that all the different parts are working properly. A full copy of the entire data warehouse does not typically exist in the development and test environments. Rather, a representative subset of data is available to support developing and testing activities.

These separate environments are also needed to support upgrades to the software tools too. It is too risky to install a new product release directly on a production environment. There are too many variables that could cause the product to not function when first loaded on your machines. The new software needs to be loaded and exercised, and any problems need to be corrected in order for the business to use the data warehouse.



Each of these environments must be able to support the complete end-to-end work associated with the data warehouse. This is shown as the work that flows with the data warehouse data architecture. Often, the development and test environments are each set up on a separate server. The production environment frequently involves multiple servers that are geared and tuned for that specific function. The specific number of servers needed is determined as part of your technical architecture.

## Components of Technical Architecture

As with data architecture, there are many layers of a technical architecture. On top of that, there may be a different product to deliver each major function. The layers are quite basic:

- Hardware
- Operating system
- Functional software such as database management system or query tool

The hardware and operating system layers for a data warehouse often align with the standards already in place within an organization. Examples of some of the types of functional software products that are needed for a data warehouse include the following:

- Data modeling
- Database management system (relational, columnar, multidimensional)
- Data extraction
- Transformation and manipulation, including cleansing
- Data loading
- Data access (simple query, reporting, business intelligence, scorecards, dashboards, specialized analytics)
- Data definitions, documentation, and other metadata

Sometimes these functions are combined into a single product. For example, many ETL software tools support a wide variety of cleansing and transformation functions. In other cases, these functions are sold as bundled solutions that include the hardware, operating system, and functional software. The bundled product offering saves you the effort of selecting each separately, and all of the technology in the bundle works well together. Taking this one step further, there are products designed specifically to support data warehousing. The technical design encompasses the hardware, operating system, and software

function. These are called *data warehouse appliances*. They are designed entirely to address the functional and performance requirements of a data warehouse.

Each of these functions needs to be represented in the technical architecture. The technical architecture should also define how these functions are split or combined across different hardware servers. The communication paths between the servers should also be identified.

## Infrastructure

Another term that is often used when discussing architecture is *infrastructure*. It is helpful to look at a general definition of infrastructure from Wikipedia:

*Infrastructure typically refers to the technical structures that support a society, such as roads, water supply, wastewater, power grids, flood management systems, telecommunications (Internet, telephone lines, broadcasting), and so forth. Most of these systems are typically owned and managed by governments or public utility companies. These various elements may collectively be termed civil infrastructure. Viewed functionally, infrastructure facilitates the production of goods and services; for example, roads enable the transport of raw materials to the production plant and distribution of finished products to markets.*

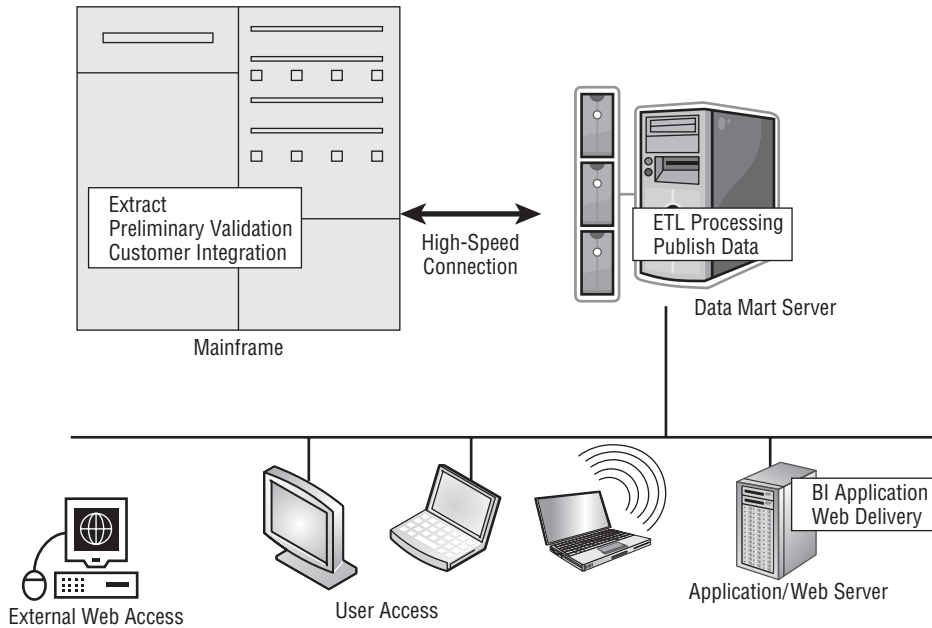
More specifically, infrastructure for the data warehouse is the foundation technology that supports the overall environment. This includes basic networking services such as providing shared network drives for storing the group's files, such as word processing, spreadsheets, and presentations. The existence of some sort of computer for each user can also be considered infrastructure. This is usually put in place when an employee is hired. The data warehouse project does not need to order laptops so that each business analyst has his or her own machine. Basic networking of computers, via a local area or wireless network, is another example of infrastructure.

Each project team does not need to design and install the infrastructure, but should ensure that there is enough capacity in the infrastructure to support the new data warehousing, reporting, and analytical functions.

## Technical Architecture in Action

The data warehouse is only one type of system that exists in an enterprise. The data warehouse technical architecture must fit within the parameters of the overall systems strategy. If the enterprise direction negatively impacts the data warehouse, then the enterprise strategy should be revised.

The technical architecture is where the basic principles are defined for each project to follow. This is the blueprint showing how the different functional requirements will be organized. Figure 9-6 shows a high-level technical architecture.



**Figure 9-6** Example high-level architecture

In this simple example, the source systems exist on a large mainframe computer. After the data is extracted from the source, preliminary data validation is done, taking advantage of corporate capabilities for standard enterprise codes. Because customer data will be pulled from multiple sources, initial integration of customers will also begin on the mainframe. Then, the data will be moved to a UNIX-based server to complete the cleansing and transformation work. The database to publish the data will also be stored on this server. If the performance of the ETL or database access degrades, these may be split to run on two different servers. A single server will be used to start, and another server running a Microsoft operating system will support the business intelligence/data access and web delivery functions. Finally, a variety of mechanisms enable end users to gain access, including direct network connection, wireless connection, and access via the Internet.

There are likely to be more detailed technical architecture diagrams to provide the necessary details to the different IT people who will be involved, but these are not usually anything you need to worry about.

### ***What You Need to Know about Technical Architecture***

Technical architecture is the responsibility of the IT organization. However, if business or IT managers have concerns about the technical architecture, this section provides some ideas about what can be asked to ensure that the right

issues are being addressed. When understanding the technical architecture, you need to determine the following:

- What parts of the technical architecture are new?
- What is the risk associated with that new technology?
- What steps are being taken to mitigate the risk?
- Do you have enough capacity to accommodate the new project with the existing technology?
- Does the technical architecture provide all the tools needed by the project team? Why or why not?
- How does the data warehouse architecture fit within the overall information technology strategies and architecture for the company?

For each product named as part of the architecture, you should also know the following:

- Will the tool provide the functionality that is needed?
- Can IT effectively support it?
- Will the technology deliver the performance needed?
- Is this a cost-effective choice? The cost estimate needs to include initial purchase or lease price, installation, maintenance, and support. It is also important to consider the effort required to develop solutions with the product. Some products may cost less but require three times the development effort.

It is usually not necessary for the business community and management to understand all the nuts and bolts of the technical architecture. It is necessary to ensure that the architecture will support the work that must be done.

## **Navigating the Technology Jungle**

---

The process of selecting hardware and software products is more difficult than ever before because of the many products from which to choose. Many provide similar functionality but work in distinctly different ways. How can you find the right fit? You must have a clear understanding of what you need in order to select the proper technology. These requirements can help you navigate this daunting sea of available solutions to find the appropriate suite of products for building and delivering the data warehouse.

## Weighing Technology Options

The data warehousing technology trends are very cyclical. The industry swings between an interest in having a single vendor provide end-to-end functionality and an interest in having multiple vendors each provide the “best of breed” functions. This trend is about more than just how customers purchase hardware and software; it is also reflected in product ownership. When the industry swings toward best of breed, there are many small companies that develop specialized products. The interest in end-to-end solutions is reflected by a consolidation in the market. Smaller companies merge or the larger, established companies buy up the companies with niche products. There is also a fairly consistent level of interest in building and deploying the data warehouse in house, without purchasing any specialized tools.

### ***Best of Breed***

Considerations for buying best of breed products include the following:

- The ability to get the richest set of capabilities for the purpose.
- Products are often finely tuned specifically for data warehouse use.
- The product may or may not work well (or at all) with other products.
- The complexity of problem resolution increases because the offending vendor must first acknowledge the problem. There can be a lot of finger-pointing before you get down to debugging and resolving the problem.

### ***End-to-End Solutions***

Considerations for buying end-to-end solutions include the following:

- Increased purchasing power with a single vendor.
- Product integration should be solid.
- You will not have the newest breakthrough technology in all areas. Usually, there will be one or two critical areas in which you will gain significant capabilities. These are likely the reason for the interest in using this vendor’s products in the first place.

**NOTE** Some end-to-end vendor solutions are created through acquisition and not by technical design. Tool integration may be rough initially, but usually improves over time.

### ***Deciding Not to Buy a Tool***

Many smaller organizations still look at the price tag for some of these tools and decide to build the data warehouse themselves, rather than buy a specific tool. For example, extract, transform, and load routines could be written programmatically, or an organization could develop its own report creation and delivery system. Sometimes a tool is purchased to perform one function, such as the end user data access, but other functions are developed in house. Considerations for building your own data warehouse functionality include the following:

- Cost estimates do not adequately reflect the long-term maintenance of a solution that is custom developed.
- Internal IT staff is usually inexperienced with software development for this type of functionality.
- The results are static and few organizations continue to invest in developing new capabilities for the tool.

In-house development is usually the most expensive alternative in the long run.

### **Finding the Right Products**

The technical architecture clarifies what types of functional software products are needed. Sometimes the organization already owns and uses a product, such as a database management system. If the organization has already been actively working on data warehousing, then enterprise standards may already exist for which products may be used. While the goal is often to purchase (and support) only one product to provide each of the functions, this is rarely the case. Organizations often support more than one relational database platform, and it is common to find multiple data access tools in use. Each organization must determine the costs and benefits of supporting multiple tools and then act accordingly.

For a specific project, product selection may default to simply utilizing products the company has already purchased. Product selection may mean choosing between two or more enterprise-approved products. In some cases, the project team may need to search for a new product.

**NOTE** Remember that the most important part of selecting any technology is having a clear understanding of what you are trying to accomplish.

With a clear set of functionality in mind, a short list of candidate products is usually easy to create. Rather than look in detail at fifty different products, the list can be reduced to a handful. IT organizations are usually good at

looking at a technology, understanding it, and making recommendations. The combination of business requirements and sound architecture help the IT team to make good decisions. It is important to do the general research, conduct several in-depth reviews, and then make a decision quickly (measured in weeks, rather than months). Given the rate of change in the data warehouse market, if too much time is spent on research, new products and major enhancements will be released before a decision is made. New features lure you into reevaluating a product, which can lead to a never-ending loop in which no product is ever purchased.

Pick a product that can help the organization today, and will likely work for the next several years. We often can't imagine what might be available in the next five years, so don't agonize over finding a single perfect product that will support the data warehouse for the next decade.

### ***Requests for Information or Proposals***

One of the most common mechanisms for collecting details about a product is to issue a formal request for information (RFI) or a request for proposal (RFP). Extensive effort is needed to develop the questions for these requests. Some companies hire consulting firms to write the RFI/RFP for them! After the RFI/RFP is issued, the vendors get to work. Once their responses are received, there is another big effort to read and understand them.

One main thing that is learned from an RFI or RFP response is which vendor has staff members who write well. Some vendors have departments that are dedicated to writing these responses. These groups are staffed with professional writers and often maintain a full database of questions and answers. RFI/RFP responses also can indicate which vendors are able to cleverly answer yes to any question or to provide details about their product strengths within the context of almost any (often nonrelated) question asked. Unless you are an organization that is mandated to follow this formal RFI/RFP process (such as a government agency), this is not a recommended method for selecting products.

### ***Business Participation in the Selection Process***

The products that the business community will be exposed to the most are those that interface to the data warehouse. These are the data access, reporting, and business intelligence products. The business community is often asked to participate in the selection of these tools. While this can be helpful, your impression of a product may depend to a large extent upon the skill and experience of the sales representative providing the demonstration. Looking at tools increases the understanding of what is possible, but do not allow the first impressions of a demonstration to be the single driving factor in selecting a BI tool.

### ***Understanding Product Genealogy***

With the frequency of mergers and acquisitions in the data warehousing industry, it is worthwhile to ensure that you have a good understanding of the history of any products you are considering working with. Find out how long the product has been in existence, other names it has been marketed under, and who may have owned it previously. After more than 20 years in the industry, I find it hard to keep up with some products that have been purchased and renamed multiple times. This information can help you with your research to learn about the product. Having a long history and multiple names does not inherently indicate a problem; it is simply one more thing to know about a product.

Another genealogy that is interesting to trace is that of the key product engineers and designers. Look at the experience of the key research and development personnel at that company. Some product engineers have masterminded multiple tools, with each one surpassing the former.

### **Understanding Value and Evaluating Your Options**

There is constant pressure on software vendors to deliver more functionality at a lower price point. This happens frequently with hardware. For example, in the late 1980s, a 10-gigabyte database server (hardware, operating system, and database bundled together) could be purchased for nearly \$1 million. Today, if you look for a promotion, you can buy a 2-gigabyte disk for a camera for around \$20! This reduction in the cost of disk space has greatly benefited all of us, making it affordable for us to store music, videos, and photos on our personal computers. There has been a similar reduction in the cost of the technology to produce, capture, and distribute music, photos, and videos.

If you look at the data warehousing industry, the technical platforms and disk alternatives have certainly expanded. New lower-cost options are available. These now enable mid- to small-tier companies to venture into data warehousing in ways that were not feasible in the past.

Although the price for some technology does indeed go down as it becomes more available, that is not always the case. Many of the high-end data warehousing tools are still quite costly. These software companies must still invest in highly skilled developers, and sales and support staff, to be able to market and support their products. As customers, we demand more functionality for less money, with superior support. This is a tall order for any company to achieve.

To remain viable, technology vendors are under constant pressure to win new business. Sales and marketing departments are highly skilled at taking customer demands and then refining their message to resonate with those requirements. Whether the products will actually meet your requirements is another issue. The message is intended to keep the company in the running to win your business.



Unfortunately, the pressure that you, as the customer, put on the data warehousing vendors contributes to the marketing hype. Intense competition also contributes to inflated marketing messages. For example, suppose that four out of five vendors under consideration claim to be able to have the initial data warehouse up and running in ninety days. What would your reaction be if the fifth vendor suggested that it will actually take six months to get the initial data warehouse running? In many cases, the fifth vendor would be cut from the selection process because they appear to be less efficient than the others. Unfortunately, the fifth vendor may be the only one with a realistic estimate. This pattern encourages all of the vendors to focus their message on the latest trends or hot topics.

Another thing that impedes your ability to gain a true understanding of a product is the implementation of “stopgap functionality.” Most vendors offer potential buyers a list of questions and answers about a product’s core functionality. Often, a specific function or feature emerges on these lists, one that has not been commonly offered. As soon as one competitor can provide this capability, the others quickly need to follow suit or they risk losing valuable business. Unfortunately, sometimes the capability is added in a convoluted ad hoc manner that makes it difficult to really utilize. However, the ability to deliver this feature can now be claimed during the sales pitch.

#### **A CASE STUDY: MISLEADING INTEGRATION FUNCTIONALITY**

**A commonly requested data access tool function is the capability to integrate data from spreadsheets with data from the data warehouse. This has been requested for years. The intent is to enable a business user to bring into an analysis data that is not loaded in the data warehouse. Many data access products have tackled this in different ways. Without revealing names, one product provided a very nice demonstration of this functionality by creating a report during the meeting. A query was saved and then a spreadsheet on the user’s desktop was selected to bring into the tool. The data was integrated using the product codes and a nice clean report showed the data warehouse results side by side with the spreadsheet input. This was simple, powerful, and compelling.**

**So, what is misleading? The technology works as advertised, as long as there is a column of data in both the query results and the spreadsheet that can be used to integrate the data. In reality, it is unlikely that the data on a spreadsheet will have codes or identifiers that will be cleanly and accurately integrated with the data from the data warehouse. The technology works, but it does not really provide the desired business functionality. This is not a defect in the data access product, but rather the reality of the data in any organization. The tool provides the functionality, but it can’t be implemented in a meaningful way.**

### ***Cutting through the Marketing Hype***

For anyone, business or IT, involved with selecting a product, you must be wary of sales and marketing messages. As with purchasing any other product, a good consumer is an informed consumer. Remember that the objective of those marketing messages is to entice you to purchase that particular product. Their goal is to get you to buy. As long as you keep that in mind, you can use the sales process to your benefit.

First, develop a partnering relationship with any vendors you may be working with. This will set the tone for how you interact over the long term. Good account representatives understand the value of honesty and hard work. This helps both you and their company to develop a solid, long-term relationship. Express your interest in understanding the product offering. Communicate that a single “no” answer will not immediately cause the vendor to be eliminated from the selection process. Pose your questions in a curious manner, not confrontationally.

Another way to help get clear answers is to work with the vendor’s consultants or systems engineers, rather than the sales representative. Most sales reps are honest, but even those with the highest integrity have been known to exaggerate from time to time. The consultants are the ones assigned to implement what was sold. If the product doesn’t do what you need, the consultants don’t want to be left holding the bag.

#### **INTERPRETING VENDOR RESPONSES**

**Question:** Can I see how you do *xyz*?

**Response:** I am sorry that I can’t show you that capability today because it is not loaded on this machine.

**Translation:** I don’t know how that works and I do not want to try it in front of you.

**Question:** How many customers are doing *abc*?

**Response:** Yes, our product supports the ability to *abc*. This is a relatively new feature. I have not implemented it for any of my clients yet. Let me do a little research to find a customer who is using that capability.

**Translation:** This feature has been added in order to enable us to move past certain hurdles in the sales process, but no one really uses it. As far as I can tell it should work, but I have not seen it myself.

**Question:** You mentioned that the product can do *lmn*. This is a critical feature for us. Can we see how that works?

*(continued)*

**INTERPRETING VENDOR RESPONSES (continued)**

**Response:** That capability is in a future release.

**Translation:** Someday you will be able to use that capability, but you really need to wait for development to finish and then wait for the bugs to be worked out before you should stake your entire project on this function.

When preparing for a sales meeting, a good vendor sales team will ask you to provide in advance a list of product(s) or services that are of interest, and any specific questions or concerns. This gives them the opportunity to properly address them during the meeting. This works to your benefit. Otherwise, the meeting may turn out to be a fact-finding mission for the sales team, which then leads to another meeting to really address your needs.

When any vendor is showing you product features, it is important to ensure that you walk away with a good understanding of how things are done. A simple-looking report may have taken several days and the help of advanced technical support staff to build. You need to know not only that the function is possible, but also whether or not it is feasible for your project. A colleague in London expressed this specific concern with a descriptive and colorful phrase. The question was asked, “Is that part of the base tool or was there some jiggery-pokery involved to produce that report?” What a wonderful expression! Using a little “jiggery-pokery” is not inherently a bad thing to get the results you want. It only becomes a problem when you are not aware that this is what is involved.

## The Value of References

One of the most useful things you can do when making product choices is to talk with other organizations that are using that product. The first path to find these other companies is to ask the vendor for references. Everyone’s first instinct is to request to talk to another company in the same industry with data of similar size and complexity. That puts any vendor in a tight spot. The only companies that meet these criteria are your top competitors. If those companies use this product, it is unlikely that they would be willing to speak with you. Clearly, the vendor should not be sharing too many details about these installations with you either. You would not want them telling others anything about your installation, would you?

Explore other possible references with that vendor. Request to speak with companies from other industries that support a similar business function (finance, human resources), size, or complexity. This should be sufficient to get a good idea of how the product is working in a similar setting.

Have discussions with people at different levels in the reference organization. It is helpful to touch base with other managers, but it can also be enlightening to have developers talk to each other too.

There are other avenues for seeking out references too. Tap into your network of friends and family. Compare notes with other organizations at industry (data warehouse and/or your industry) conferences and seminars. Some of the things that you may consider asking a reference include the following:

- How responsive is the vendor organization?
- Is there a difference in responsiveness for sales questions or problem resolution?
- How strong is the customer input process for product development?
- How are development priorities set?
- Explore the size of the installation: How many licenses? How many people use the tool? On how many servers is the software installed? How large is the data?
- What parts of the product do you use?
- Does the product work as advertised?
- What is the worst thing you have learned about the product, or your biggest disappointment?
- What is the best surprise or most positive thing you have learned about the product?
- How long have you used the product?
- How many major releases have you participated in?
- How smooth was the release upgrade process?
- How stable are new releases?

These are just a few of the questions that you can ask (or answer) that relate to the vendor and the product, but don't reveal details about how you are leveraging the technology.

**NOTE** Return the favor. After you have worked with a product, be willing to be a reference for your vendors. This may require permission from corporate communications.

---

## Making Architecture Work for You

---

The key to having architecture work for, rather than against, you is to maintain the balance between structure and flexibility. Architecture can get out of control. It can take on a life of its own when the purpose and business goals are lost. Architecture can also gain so much power that it gets in the way of delivering

value to the organization. Some organizations have separate enterprise groups that specialize in architecture. These groups typically define and maintain the enterprise architectures. They also assist individual projects and groups to understand and implement the architectures. The architecture group can act as gatekeepers who decide which projects meet the rigorous standards and which are not allowed to continue. It is much more effective when enterprise architecture groups focus on educating and helping projects, rather than simply enforcing compliance.

At the other extreme, some organizations work so fast and so hard that they never find the time to step back and develop an overall plan. This results in many disparate systems. Each project expends time and energy to pick products or determine a data strategy. These duplicate efforts are a waste of corporate resources.

The optimal architecture provides the structure and direction for each initiative. It is also flexible, and changes are made to the architecture when a reasonable business case is presented.

## **Just-In-Time Architecture**

While architecture provides the long-term vision, it may not be feasible for the first project to set up everything. With the vision in mind, parts of the architecture can be implemented at different times. Perhaps the data needed for the first project is small enough to be adequately supported using only a single server for the ETL, published data, and data access. When more data is added, additional servers can also be added. Each project should review which existing components could be leveraged. After that, the project can build out any missing parts that are needed to deliver the solution.

I must thank a very experienced colleague for describing this process as *just-in-time architecture*. This is similar to a manufacturing company whereby components are delivered just before they are needed. Each piece of the architecture can be put in place in time for the project that needs this level of support or functionality. Another way to state this is to “think globally and build locally.” The architecture provides the overall framework, and the implementation is performed as needed to support projects.

## **In Real Life**

---

To get a glimpse into common challenges with architecture, several issues facing our friends from Giant Company and Agile, Inc., are discussed next.

### **Architecture at Giant**

The company has a well-established centralized architecture group. All projects must go through an architecture review and approval process. The architecture has a heavy emphasis on operational application systems. The architects

themselves have years of experience, but limited exposure to any data warehousing concepts. Whenever a data warehouse project comes up for a review, the results are frustrating. The data warehouse project data and technical architectures do not fit neatly into the enterprise system architecture. The data warehouse projects are often sent back for changes and the review process begins again.

This has resulted in many reporting and analysis needs being met by information initiatives that circumvent the entire IT organization. The business groups fund the work directly. Many disparate data-warehouse-like databases have been built using a wide array of technologies. The immediate business needs are usually met, but the work is not extensible and is not leveraged across the enterprise.

The centralized architecture group must get data warehouse training. The architectures themselves must be expanded to accommodate data warehouse requirements. Then marketing and communication effort is needed to let the rest of the organization know of the changes in the architecture. New projects should use the new architecture. Strategies must be developed for existing projects so they can adopt the new architecture over time.

## **Agile Ignores the Need for Architecture**

Each individual division is empowered to do whatever it needs to get its work done, including the purchasing of technology. This has resulted in a proliferation of products, many of which provide similar functionality. In several cases, each group purchased its own copy of a technology and the organization now owns duplicate licenses.

In addition to the base cost of purchasing products, each group is responsible for the installation and support of its own platforms. Some applications are small and do not justify hiring a full-time systems administrator, so existing staff members are asked to take on these responsibilities.

From a data architecture perspective, each autonomous division looks out for its own interests. This has made it extremely difficult for the corporate executives to understand the dynamics of the company and to set the direction. These strategic decisions are made based on personal experience and limited facts.

The organization needs to invest some time and energy to develop an enterprise architecture for the data warehouse. The technical architecture can help to consolidate data warehouse technology. This will increase its purchasing power, and help coordinate the support and maintenance functions. Defining the data architecture is the just the first step toward delivering integrated data to the senior staff. To make the effort of developing the architecture worthwhile, the organization must follow through to ensure that it is implemented.

## Summary

---

This chapter has reviewed the fundamentals of data and technical architecture. What does this mean to you? Following are some basic questions that you should be able to answer about data warehousing in your organization:

- Do you have a data warehouse architecture? Where is it documented?
- Does the data warehouse architecture cover both data and technology?
- Do you have a basic understanding of the DW architecture?
- Who is responsible for revising and updating the architecture?
- Do you know how well your project is or is not adhering to the overall architecture? Why or why not?
- How do project teams learn about and leverage the DW architecture?
- Does architecture get in the way of getting things done on your project?
- Does the team need to learn more about the architecture in order to implement it effectively?
- Is the architecture too rigid, not adapting to changes in goals?
- Are enterprise ideas merely fleeting, with no actual follow-through?

These questions will help you gain a working understanding of where the organization stands and how much work is yet to be done. Use this as the catalyst to get something started, make adjustments mid-stream, and refine what is already in place.

While architecture is mostly a systems responsibility, it is important that both business and systems managers have a sound understanding of how architecture is being used within the organization. You don't need to be able to design one; and at the lowest level of details, you don't need to know the specific technical specifications, but you should have a solid understanding of the overall direction and how it will help the organization in the long run. The next chapter looks more closely at what is involved in building the data warehouse.





# Implementation: Building the Database

The IT work involved in building the database is one of the most difficult and complex parts of implementing a data warehouse environment. There are many resources to provide guidance for the project team and technical staff to help accomplish this goal. For the purposes of this text, the focus of database building will be at a high level. Rather than share the nitty-gritty design details, a summary is all that you need. However, it is important for all managers (business and IT) to have a basic understanding of what is happening during this part of the project.

This chapter provides a representative sample of the work that needs to be done, but is not intended to be an all-inclusive reference. It is also important to know how you can help and what is expected of you during this very technical part of the project. This chapter provides both a foundation and insight into how managers and other members of the business community should be involved.

## Extract, Transform, and Load (ETL) Fundamentals

---

The ETL (extract, transform, and load) work is the most frequently underestimated work that is done on a data warehouse project. It is also the least understood set of tasks by everyone other than the developers themselves. It would seem simple: Copy the data, move it around a little bit, and load it for use. In reality, the ETL process is rarely straightforward and simple. Having a good understanding of the work that must be done and the condition of the existing data will help the project team to better estimate the effort and length of time needed to build the data warehouse.

## What Work Is Being Done?

If the work is not that simple, what is actually happening? There are several major steps to developing an ETL system. While it is easy to move data

around, the goal here is to create a complete system that can be turned over to the operations group to run, and that is not as simple to accomplish. The same characteristics of any other production system must apply to the data warehouse too. The goal is to build a full system that runs without manual intervention.

#### CHARACTERISTICS OF A PRODUCTION SYSTEM

**A full production system must be able to do the following:**

- **Gracefully handle exceptions that are identified in the incoming data.**
- **Provide warning and error messages to flag conditions in the data that require further attention. These may be conditions that can be programmatically addressed in the system or they may require human intervention to resolve.**
- **Ensure restart and fallback capabilities in case the system is interrupted due to processing errors or environmental issues (e.g., a computer goes down).**
- **Provide an audit trail to trace how data flows through the system. This is needed to help track problems back through the ETL system and/or the underlying source systems.**
- **Include backup and recovery of the database itself.**

Several major steps are involved in the development of a production ETL system:

1. **ETL system requirements:** Up to this point, the requirements and design components of the project have focused on what the end result must look like. The data model reflects how the data is to be stored. Many detailed requirements have already been collected during other project activities, such as individual data element names and definitions. Data profiling activities should yield insight into what the current data looks like, and strong data governance may have already determined how each data element is to be handled. Additional requirements for the ETL system must also be defined. Examples of these requirements include processing rules, guidelines for compliance with legal requirements, a processing window, and what the audit trail must include.
2. **ETL system design:** The dimensional model is the target that the ETL system will build. The ETL system design provides the details about how to get from where the data is now to this target dimensional model. Some organizations require that every little detail be defined, including all of the specific rules for building the dimension, and that fact tables

be documented prior to starting to build the ETL system. Others sketch a quick high-level data flow on a napkin and then start building. There must be some balance here. It is impossible to track down every little detail that may be discovered in the data prior to starting work on the ETL system. However, it is important to create specifications defining how the ETL system will work. These are developed using the results from the requirements, including data element definitions, data mapping, and insight from source data analysis. This is even more critical if the construction work is to be done by third-party developers. Part of the overall design should also address functionality to keep the ETL system itself running, including an audit trail and backup/recovery capabilities.

3. **ETL system construction:** This is the actual development of the system itself, which may include writing programs or using technology to perform the work. This work can be divided up among different people or even different teams. Using the cohesive design, each team can work on its part. There may be some dependencies between the teams, but much of the work can be done at the same time.
4. **ETL system testing:** Because there are typically different people working on different parts of the system, it is important to conduct thorough and complete testing of the entire system. This also provides the opportunity to identify any bottlenecks and improve the system's performance. In order to prepare for testing, a series of test cases need to be developed to provide realistic conditions to determine whether the system is working properly. These test cases must represent actual business situations and need to be defined by representatives from the business community.

As discussed so far in this book, many of the facets of building a data warehouse are different from other traditional systems design and development. The creation of the ETL system itself is the most like any other systems development effort. The organization must apply its systems development methods to create a robust ETL system. This is not the time to abandon discipline and simply move data quickly.

## ETL System Functionality

What exactly does this ETL system need to do? The system will provide several common functions. While the concepts are consistent across the marketplace, the specifics are unique to each individual organization and its own source systems. The following sections describe the functionality and flow of an ETL system to directly populate dimensional data structures, called *presentation*

*servers* or *data marts*. Additional steps would be needed to populate a normalized data warehouse, but most of the functionality would be similar. Let's look at several of these common functions.

### ***Extraction***

Extracting the data can be a lot harder than it sounds. The business needs information about the latest transactions, or those that have been processed since the last time the ETL system was run. It is also necessary to get the reference data that describes those transactions. For example, if an existing customer moved but has since purchased more products, you need to know the details about the customer's new location and still get the purchase transaction(s).

The challenges in extracting data are directly related to the ability of a source system to provide new business transactions and supporting reference data. Some systems identify when changes have happened and can easily share the data. Other systems effectively do what is necessary to complete the transaction but are not designed to keep track of what has changed. This may mean that the data warehouse gets a copy of the entire customer file, and the ETL system needs to figure out what has changed. Clearly, there is a significant difference in the amount of work required to sift through the entire customer file versus getting details about only those customers who have updated information.

The ETL developers need to coordinate their work with that of IT teams who work with the source systems where data is to be extracted. These other application development and support teams may be overwhelmed by their own work. If so, the requests from the data warehouse team are just more work added to an already overloaded schedule.

**TIP** Include in the project plan tasks for the source system IT support teams. Get estimated effort and lead time requirements from their manager to ensure that the appropriate resources are available.

### ***Transformation***

The term *transformation* represents a wide variety of functions that are performed to take the data as it is and get it ready for what you need to support the business decision-making process. Some of these functions are direct and straightforward, while others can be complex and challenging. There are two distinctly different sets of work to be done by developers: one to build and maintain the dimensions, and another to build and maintain the fact tables.

The dimensions provide the data needed to enable selection and grouping of data in many different ways. Creating and maintaining the dimensions often

involves using data from multiple source systems. There may be some sources that are used purely to get descriptive data for the dimension. The functionality is easier to understand in a specific context, so let's look at common functions needed to build and maintain a Customer dimension:

- **Validate the customer:** Determine whether the incoming customer is already known.
- **Identify changes to known customers:** If the customer is known, have there been any changes to that customer information? This may require looking at data from several sources—perhaps the sales transactions, the corporate marketing customer database, and the accounting systems.
- **Handle changes to known customers:** For existing customers, make sure that changes are handled appropriately. Sometimes data is updated to reflect the current values. In other cases, historical versions of the customer are needed. This concept, first described in Chapter 7, is called *slowly changing dimensions*. Many other data warehouse books provide technical details about developing and maintaining slowly changing dimensions. Needless to say, this simply means that there is more work to be done and a little more complexity to be managed.
- **Identify new customers:** If this is a new customer, then look at the sources mentioned above to determine whether the customer is identified in those systems. If this is a new customer, then a new data warehouse identifier is assigned. This is called assigning a surrogate key.
- **Build a full description for new customers:** Once a new customer is identified, data must be located to fully populate all of the customer attributes. This may require looking at a variety of data sources. It also requires guidelines that prioritize the sources. Perhaps the customer master database from the marketing department should always be the first place to look for customer information. If the customer is not found there, then use the name and address data from the source where the customer was first identified and then request demographic data from a third-party data provider. The source must be selected for each data element.
- **Validate data relationships:** Follow the defined business rules to ensure that all of the data relationships are correct. For example, check the city, state, and zip code values to ensure that they represent a valid combination. This may include standard relationships such as the geography, as well as internal relationships such as customers living in Illinois belonging to the Midwest sales region.
- **Map customer data from multiple sources:** When data is being processed from more than one source system, there must be a method to associate the

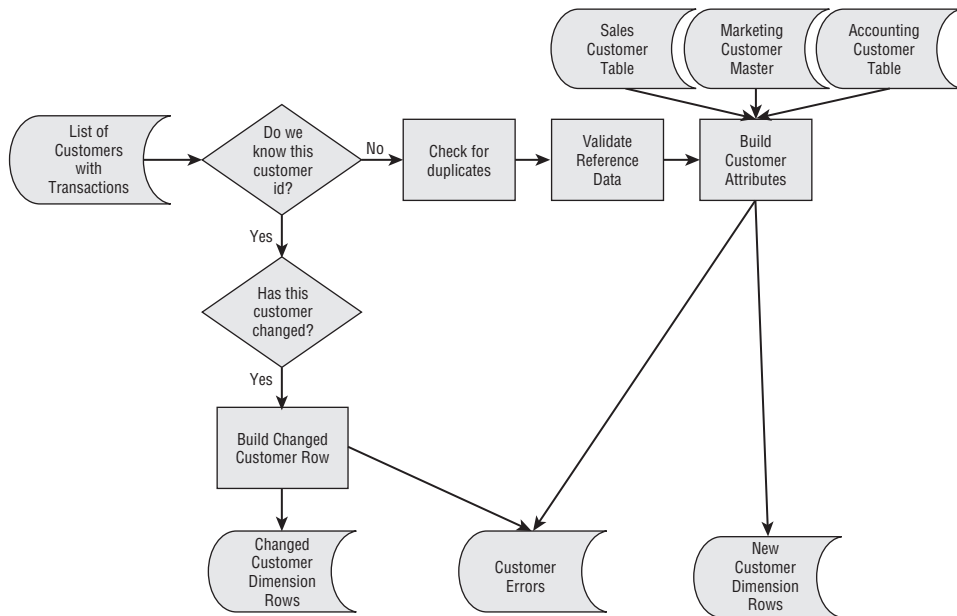
data from each source. Each source system may have different customer identifiers, sometimes called the production identifiers (IDs) or codes. For example, there must be a mapping indicating that the production ID for the customer “ABC” in the sales system is the same customer identified as “1003XJ” in the accounting system. Creating and maintaining this map is a critical part of the overall transformation process that enables integration of data from different sources/systems.

- **Identify and eliminate duplicate customers:** It is common to find the same customer in each of the different data sources. However, it may be the case that the same customer can be in data from the same data source. For example, there may be data for a customer named Laura Reeves and for a customer named L. L. Reeves. By looking at the last name, variations of the first name and initial, and other differentiating data such as date of birth and home address, you can determine whether this is indeed the same person. This processing is often called *de-duplicating* or *de-duping* the data.
- **Restructure the data:** The customer data must be reorganized to fit into the target database design.
- **Handle errors:** Throughout the processing of the customer reference data, rules are applied to accomplish the necessary function. If for some reason there is a problem processing the data, then a determination must be made about what to do next. Previous data profiling may have already captured what should be done. Some errors are minor and may simply require a default value being placed into the customer loyalty field. In other cases, the error may be severe enough that a customer cannot be identified anywhere. The business may require that the customer (and then any associated transactions) be placed in a “parking lot” for further review and *not* loaded into the data warehouse at all. The number of errors and how they are handled depends upon how the business needs to see the data. Figure 10-1 shows a high-level view of the processing required to build this simple example of a Customer dimension.

Similarly, there are common functions that are performed to process incoming transaction data to produce the fact tables. These are explained in the context of processing customer sales transactions. Common functions include the following:

- **Isolate transactions of interest:** Depending upon how the data is provided from the source system, some processing may be required to isolate the transactions that are needed to populate the sales fact table. The sales system may process other transactions such as product returns or inventory adjustments. If these are not needed to create the sales fact

table, then they can be ignored for this step (they may be needed for other fact tables).



**Figure 10-1** High-level Customer dimension data flow diagram

- **Check the existence of dimension reference data for each sales transaction:** This means that the critical descriptive data such as the product, customer, and sales date are valid. Transactions can flow through the sales system with missing or invalid customer identifiers. There must be a set of rules for how to handle each special case that is observed on the transaction. It is also critical to ensure that there is a row in the dimension table for each instance that is used for a transaction. This is known as *referential integrity*.
- **Assign appropriate data warehouse identifiers:** Once the critical reference data has been validated, the identifiers must be changed from those used by the underlying source systems to those used in the data warehouse environment, the surrogate keys.
- **Validate fact fields:** Although the actual facts will be specific to each transaction, there are often general guidelines that can be checked to determine if the fact is accurate. For example, a sales transaction of zero units may indicate that this is not an actual sale. A single sales transaction for an amount that is greater than the average weekly sales for the entire company is probably a data error.

- **Translate fact values for ease of use:** The business measures needed for reports may be different from how the data is stored in source systems. It is helpful to convert data to a common unit of measure for reporting. If a sales transaction records the sale of one case, this may be a case of 24 individual units or perhaps 36 individual units. Converting all of the sales to the individual units ensures consistent and meaningful reporting. Additionally, many facts work together. For example, the number of units sold, price, and dollar amount sold are all useful facts. Only two of these need to be stored physically; the third can be calculated on-the-fly. To make this as fast as possible, it is helpful to store the two facts that are used most often.
- **Unravel source system logic:** In many instances, the source system has core data stored in a manner to help that system run fast and enable each module to be relatively self-contained. This may mean that the pieces that are needed to get a complete picture of a sales transaction may be stored in several places or tables. To make things more challenging, there may be no direct links between these tables, although there may be a series of translation and lookup tables that are needed to get to the bottom of things. This may be why the data is so hard for you to use in the source system. The details are all there, but you need to understand the “secret handshakes” in order to get to the real numbers. The ETL system can apply this logic to provide the real sale numbers to be loaded into a fact table. This is done once and then everyone can access these facts without having to learn all of these details.
- **Perform complex calculations:** Many calculations can be done when creating a report or accessing data in an ad hoc manner. However, some calculations may be too complex and long-running to perform on-the-fly. The ETL system can perform these calculations and store the results as facts, which can then be directly accessed for reporting.

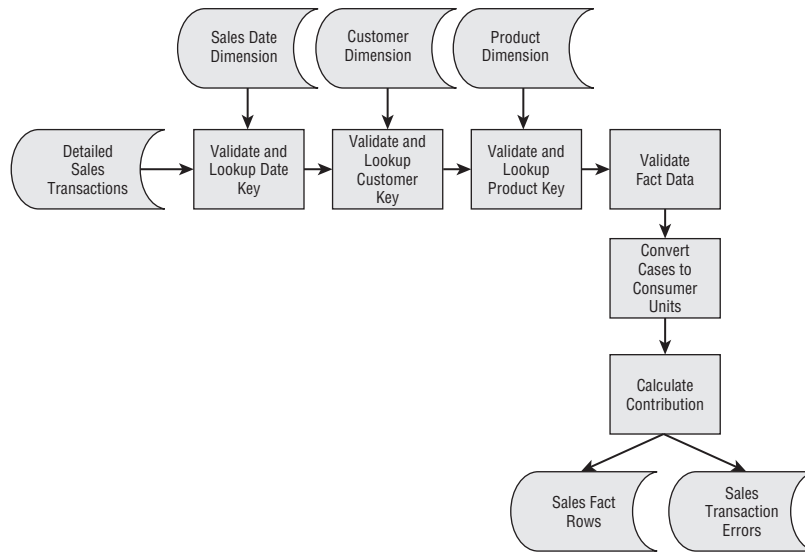
Figure 10-2 shows a general data flow diagram to build the sales fact table just described.

In addition to the functions described here, a great deal of work is also required to address anomalies and challenges in the data. Often these challenges surface when the ETL system is being developed. No matter how carefully the data is profiled in advance or how much detail is included in the design, new and unexpected things will show up in the data. The goal is to design the ETL system to be able to gracefully identify and address the unexpected.

## **Load**

After the extracted data has been processed, the final step is to load the data into a database. This may be loading directly into a star schema in a presentation server or loading data into a third normal form data warehouse.





**Figure 10-2** High-level sales fact table data flow diagram

## The Business Role in ETL

The participation of the business community is fairly obvious when gathering business requirements. Business participation in the construction of the data warehouse is less obvious but just as important. At this point in the project, the business people who are needed are those with in-depth knowledge about the current data and reporting methods. After looking at why these business people are needed, this section highlights the areas in which business involvement is necessary.

### Why Does the Business Need to Help?

There will be a variety of skill and experience levels on the development team. Without guidance from the business, you are expecting the project team to make these critical decisions about the data. If there is limited access to the business community, and pressure to deliver something, the project team is likely to forge ahead without business input. Consider what really happens: The technical team members decide what they think is better or worse, and then one of the ETL developer makes the decisions. *Do you really want the business performance results to be reported from a database that was loaded with rules defined by the technical team's best, and, it is hoped, educated, guess?*

Such an approach rarely delivers results that the business really wants. This often results in a subsequent project to redefine, rebuild, and correct the data to reflect what is really needed. Many organizations continue to apply pressure and do not commit business resources, yet fund second, third, and fourth efforts to get it right. Make the investment now so that there will not be

a need to redo the data warehouse repeatedly. In the following sections, you'll look at what needs to happen now, to get it right.

## Defining Business Rules

The overall design of the ETL system is clearly the responsibility of IT, so what role does the business play in the design process? The business must define and provide the rules about how to process the data. Fortunately, there are dozens of data elements that do not require much work. These data elements can be validated and then flow directly into the data warehouse. However, it is not always that straightforward. Some of the things where business input may be needed include the following:

- **Helping to pick the data source:** Suppose there are three different systems that have the customer's home address. Which one should be used? Perhaps you should always use the accounting system first, then look in the sales database, and, only if it is not found there, go to the customer feedback system. Candidate source systems will be researched by the IT team members, and their findings can help with the decision-making process.
- **Defining processing rules:** Representatives from the business community need to work hand-in-hand with the design team to define the rules for handling the data. What should happen when looking at a sales transaction and the customer does not exist in the reference data? Should the sales transaction be loaded with a default customer identifier? Should that sales transaction be set aside for further research before it is loaded? The business community must provide the input so the appropriate logic can be built into the ETL system.
- **Defining integration rules:** The business groups are often already using the data that will be loaded into the data warehouse. Suppose reports are being produced with data from multiple sources. How is this integration being done currently? Perhaps it is being processed in an existing reporting system. Perhaps it is pulled together manually to create the reports. The person who is doing this work has the most knowledge about the details regarding the matching rules and how to handle exceptions. The data warehouse team must work with the individuals who have this in-depth knowledge to determine whether this same logic should be applied in the ETL system. Sometimes there are no existing integration rules to even start with. In that case, the IT and business representatives need to work together to ensure that the data will be integrated appropriately.

**USING EXISTING LOGIC VS. STARTING OVER**

Over a number of years, very complex logic can be embedded into reporting systems. Sometimes the logic is applied to load data into a database, but in many cases the logic is applied when reports are created. This means that if the logic changes, it needs to be changed everywhere. Unfortunately, over time, there can be a divergence of this logic across different reports. In addition, this logic can be a series of rules that are applied in layers. For example, the logic may first get the subset of new customers. Next, logic is determined using a series of codes to identify the source of their business (agents, the web, or other channels). Finally, a complex formula is applied to determine their sales potential. This is a simple example to communicate the concept. In reality, business logic can be extremely complex, with many layers.

The challenge for the data warehouse team is to determine where to get these rules. Some organizations can have dozens of SAS programs with many lines of code to apply the business logic, and the original author of the logic is no longer with the company. The IT team members can reverse engineer the programs and/or reports to derive the existing business rules. Then, the business team members need to review those rules to determine whether they are still appropriate. However, another option is to start over and develop rules from scratch. This can be a better choice if the rules have been in place for a long time, the business itself has changed significantly, or the rules were not clearly defined and applied consistently in the current environment. Which option to pursue should be a joint decision between IT and the business.

## Defining Expected Results—The Test Plan

A comprehensive test plan is needed to ensure that the data is correct. The business must provide a set of test cases that can be used during development and testing. This gives the baseline for comparison with the data warehouse. For example, to validate earned premium, the business can provide a report with actual earned premium by state by calendar month for the last six months. This can be used to compare with the data pulled from the data warehouse.

If there are major changes in the business rules, an existing report will not suffice for comparison. A subset of raw data may need to be pulled and manipulated manually to provide the results for some test cases.

Some of the best people to help with the initial data validation are those individuals who, if required to use the new database, would immediately look for flaws. They would compare reports from the data warehouse with reports that they have been using themselves for years. This business participation also helps others to have more confidence in the data because people they respect have helped validate it.

Some organizations have separate testing and quality assurance groups. This requires that the test cases and instructions to run the system be provided. The people performing these tests may have very little exposure to data warehousing or the data you are loading. This requires that more care be given to developing and documenting these test cases.

## **Development Support**

Once the design work has been done, development can begin in earnest. This does not mean that the business representatives are off the hook. Although a great deal of thought goes into all the work leading to a complete ETL design, there may still be things that require business input.

As the team begins to work with the actual data, performing good detailed data analysis, there will be more surprises. These can be minimized with good data analysis, ETL requirements, and ETL design work up front. Data that was expected may not exist. Data that was not expected will show up. There will be many exceptions. Often, with a little legwork, the technical team can figure out what happened and adapt the development appropriately.

There will also be plenty of times when the business needs to be consulting about what to do next. How should it handle invoices for customers that do not exist in the customer master file? Although this should not happen, strange things are often found in the real data. Adjustments may need to be made to the underlying source system, but the ETL system must know what to do if this ever happens again. Should the invoice be loaded with an “unknown” customer? Should the invoice be held for further research? The development team can do either, but the business needs to define what they are expecting to see.

The business team members will have fewer daily tasks assigned to them, but they will be “on call.” When a problem is found or questions arise about how to understand a business rule that was defined during the design process, the team needs to be able to get answers quickly. They do not want to wait for permission to ask a question that may take five minutes to answer.

## **Testing the ETL System—Is the Data Right?**

Initial testing of the ETL system will be performed by the technical team members. This ensures that the programs and processes run without errors and that they correctly perform their designed function. This often includes basic checking to ensure that the data flows properly. The team will make sure that if 100,000 transactions came in, then 100,000 transactions were processed. Other checks may include confirming that the total units shipped on the incoming transactions matches the total units shipped from the resulting fact

table. The basic testing may also ensure that each customer is assigned a credit profile and that each product has a package type.

Testing that the programs and processes run correctly is only part of the overall testing and quality checks that must be done. Additional work is needed to ensure that the data itself is accurate. The ETL system may have correctly processed those 100,000 transactions, but if all of the sales were off by a decimal point, then the tests previously mentioned would not detect it. The business must be an integral part of the testing and quality assurance process to ensure that the database is correct.

A great deal of work is required to validate the database when it is first loaded. There will also be ongoing work to ensure that each time the ETL system runs, accurate data is loaded. Who decides if the data is right? The comprehensive test plan developed earlier by the business provides the basis to assess the data. Representatives from the business community may be involved in running the test cases, but their involvement is critical in evaluating the results. This needs to be someone who has the knowledge to recognize that an increase in market share of 14% is not possible, as changes in market share are usually measured in tenths of a percent.

**TIP** The ETL system should include many validation steps to check for data problems before the data is loaded. Leverage your data access or business intelligence tool to help with data validation. These tools are designed to identify exceptions, compare results with previous periods, and compare results to thresholds. Develop a set of queries or reports that can be run after the data has been loaded as a final data quality check.

## Why Does It Take So Long and Cost So Much?

If all of the systems in the organization contained clean, reliable data with consistent production identifiers, then developing a robust ETL system would not take so long. However, it is rare to find an organization that has such clean, integrated systems already in place. Therefore, with the reality that the source systems were designed without the requirement to integrate easily with one another, the data warehouse must overcome these problems.

A variety of factors contribute to the time it takes to build the ETL system. Several of the biggest contributors include the following:

- **Long decision cycle:** Many choices need to be made throughout the design and development of the ETL system. Often, decisions can be made based upon the requirements, or a single person or area needs to be consulted first. However, in many instances the decision impacts multiple areas. It can take a long time to get the appropriate people

together to understand the choice, come to an agreement, and commit to the direction.

- **New perspectives on the data:** As part of building the data warehouse, the data may be looked at in ways that have never been attempted before. As the business and the marketplace evolve, new requirements also emerge. These may result in identification of anomalies and challenges in the underlying source systems. As new perspectives are defined, the data may not yield meaningful or helpful results. For example, it may take several iterations to define the business rule to identify long-term customers. How many years are needed to consider someone a long-term customer? What if the customer leaves and then comes back again—do you count all of the years or just consecutive years?
- **Lack of business input:** There are also many demands on the business community. Business analysts who work with the data today are often swamped with demands for data, reports, and analyses. In addition to a full load of regular reporting, there are often ad hoc requests that need attention. However, these analysts have the most knowledge about how data is manipulated and which calculations are currently used. Sometimes, the team must go through a series of meetings before identifying the person who can really help.
- **A lot of hard work:** Often, developing the ETL system takes a lot of time and money simply because there is a lot of work that must be done. This can be due to the complexity and volume of data or because of a lack of well-structured and reliable data to begin with. Obviously, if the team is pulling data from two data sources, the integration work is much smaller than integrating data from ten different systems.
- **Insufficient data profiling:** Whether the organization is using a sophisticated data profiling tool or performing detailed data analysis by running queries against the database, it is important to understand what data is being stored. When the actual data contents have not been studied, problems will arise during ETL development. The types of data problems that emerge include finding that the data element is empty or that the contents of a data element are not what was expected. Thorough study of the data prior to designing and building the ETL system should obviate these issues.
- **Indirect communication:** Looking at the day-to-day work, questions often arise for which the team needs input from another IT resource or someone from the business community. If the project team must go through a liaison to gain access to other IT or business people, this can greatly lengthen the project schedule. While it may not seem like a big

deal to wait a couple of days to meet with a key IT or business person, over the life of the project this can add weeks to the schedule—just waiting to have the opportunity to ask a question. Moreover, if the question cannot be answered by that individual, then the process begins again to gain access to the next person in the chain. It is much better to create an environment where the data warehouse team is provided an open door directly to anyone (within reason!). This requires that both business and IT management communicate the importance of the data warehouse project to everyone else. It also means that the data warehouse team must act responsibly and respect everyone's time.

- **Experience level of the team:** There is a lot to learn when working on a data warehouse project, especially for the first time. It takes some time to learn data warehousing concepts and principles. It also takes some time to learn and *become efficient* with using any new technology.
- **Lack of access to the right IT people:** Usually, several key people have knowledge of how a system works. These IT people are in high demand, and the data warehouse team may have trouble scheduling time with them. These individuals often play a critical role in keeping the operational system running and may be involved in other systems development projects.

Each of these items can increase the amount of time needed to design and build the ETL system. This is also the part of the project where you are likely to have the largest number of resources all working on the project. A small core group can gather requirements and develop the data model, but now many more people can work at the same time to build the ETL system. Small organizations may still have a small core team (likely the same core team), whereas large enterprises may have a dozen or more people working on this. All this work translates into dollars, for internal business and IT staff and third-party consultants.

**TIP** Use what you know about the organization to anticipate common project impediments. Try to build time into the project schedule to deal with these. For example, if the team is inexperienced, then include time and resources for education and mentoring.

---

## Balancing Requirements and Data Reality

---

Organizations often have great ideas about what they want to do and how they would like to use data to help them. However, the data warehouse is limited by the data that the organization actually has. The business requirements can be



identified using the techniques discussed in Chapter 6. Many of the shortfalls of the current data are found during data profiling and/or detailed data analysis performed while data modeling. If the data is not captured anywhere, then it cannot be included in the database. Additional data challenges are discovered during the development of the ETL system.

It takes a lot of detailed, tedious work to track down and resolve all of these individual data issues. It is important to ensure that issues are well understood so that decisions can be made about how to deal with problems that arise. In some cases, getting to the bottom of the problem itself may take a lot more research. A decision must be made whether to work on the problem or to postpone it for the future. This must be a joint business and technical decision. Some problems can be put off with little or no immediate impact, but some data issues must be resolved in order to meet the overall objectives of the project.

For example, suppose the organization has been collecting customer demographic data for years. When customers call in, they are asked if they are willing to complete a short survey. This short survey collects additional demographics about each customer household. While it sounds interesting to use for analysis, most customers did not participate, so only 15% of the customers have any data. To make matters worse, the entry screens required the answers to be keyed in, rather than using a set list of options, so the data that has been collected has many different values and will require a lot of cleaning to make it useful. The question at hand is whether this is worth the effort.

Because the demographic analysis is not an immediate priority, and the work required is significant, this was postponed to a subsequent iteration. In the meantime, a better data solution is to modify the survey entry screen to capture pre-set options so that the data is consistent. In addition, the top five most important questions need to be included in the initial conversation with the customer, rather than as an optional survey. These decisions need to be based on a cost-benefit analysis—not a multi-week effort, but simply a checkpoint to ensure that resources are used wisely to deliver the most value in a timely manner.

Often, the data problems identified when working on a data warehouse project are data quality problems in the underlying source systems and/or business processes. It is important to dig down to find out the root cause of data quality problems. Then, decisions can be made to eliminate the problems from recurring.

## **Discovering the Flaws in Your Current Systems**

As a by-product of the detailed work that is done to extract and then transform the data, many anomalies and unusual data handling and storage techniques are uncovered. Sometimes fundamental flaws are identified regarding how



the source application system works. If these flaws impact how business is conducted, they must be addressed immediately. For example, if the formula used to calculate sales tax is not correct, this must be addressed. This would need to be corrected in the system, and follow-up research is also required to determine other ramifications. If the tax was too high, does that indicate that a refund must be paid to customers? If the tax amount was too low, what are the implications with the IRS? Addressing problems like this is difficult and highly charged with emotion. This may slow the project while the situation is being addressed. While this is usually unpleasant, it is imperative to discover and correct serious flaws.

More often, less serious problems are found while building the ETL system. The application system functions properly and business is successfully completed, but this type of problem is after the fact, in how data is reported and tracked. Many organizations have copies of production data that are used to support reporting. While these meet some basic needs, it is not enough because you are building a data warehouse environment.

Over the years, a lot of logic has been embedded into these reporting environments (databases, files, spreadsheets and/or report programs). Often the logic has not been well documented and the authors may no longer work in the group or for the organization. There is often sophisticated and complex logic applied in two primary ways. The first type of logic performs calculations. This is the most obvious and easy to understand. Calculations are used to determine commissioned sales, financial contribution, and gross margin. The second type of logic has just as much impact on the results as the specific formula that is used. It is this second type of logic that has been included and/or excluded from participating in the calculation. For example, all sales to non-profit organizations are not eligible for sales commission. This means that while the formula is the sum of sales dollars, the sales to customers designated as non-profit are excluded from the summation. The logic regarding what is included and excluded from reports and business measurements is much more difficult to track down, but just as important. Over the years, the business itself may have changed, but all of the underlying logic in the many different reports may not reflect the current view.

## Applying New Business Rules

As the data warehouse is built, the data will be cleaned and transformed using the currently defined rules. This often means that the data warehouse results will not match reports from the older systems. Clearly, the processes and results must be validated to ensure that the new rules have been implemented correctly. However, because the new and old systems are applying *different rules*, the results will not match.

This is often difficult for people to understand, and it creates concern because the reports that have been relied upon for years have been exposed as no longer accurate. In most cases, the reports were accurate and served the purpose as defined when they were developed. It is time to move on to how the business is being tracked and measured today. For most people, it is hard to give up the current way of doing things. It is comfortable; and while not optimal, it is well understood. The changes to business rules should be driven from business requirements and decisions made during design and development of the ETL system. It is helpful to have the business people who were involved in making these decisions communicate what has changed and the rationale behind the changes. Moving forward requires learning and changing—two things that many of us don't enjoy.

**TIP** Don't focus on why the old system was wrong, but emphasize how much better it is to use new measurements, formulas, and ways of looking at the business. This new perspective has evolved as a result of what was learned in the past. You are leveraging what was done in the past to make things better for the future.

## Working Toward Long-Term Solutions

Depending upon what you find, some issues may simply be too big to address at this time. If a problem originates in a major application system, the solution is likely to be much bigger than the data warehouse. Often the team that is responsible for the application system heartily agrees that there is a problem and they often agree that the problem must be addressed. However, their priorities may differ from those of the data warehouse team. The effort to address the underlying problem may be slated with the next maintenance release, which is scheduled to start 24 months out. The data warehouse may not be able to wait that long. Therefore, many techniques are employed and a lot of work is done by the ETL system to overcome these problems with underlying systems. Over the years, real change is so infrequent that too often data warehouse teams don't even think to ask for modifications and enhancements to the underlying source systems.

Don't live with the status quo. Regardless of how clever the data warehouse team is, it will be better for the long term to have clean data captured as close as possible to the business interaction or transaction itself. The following guidelines should help:

- What is the real source of this data problem and where should it be fixed?
- Encourage the data warehouse team to look for long-term solutions, rather than short-term patches.
- Follow up to ensure that modifications and enhancements are actually made to the source application systems.

In order for your organization to begin to shift in that direction, there must be a commitment and understanding with the management team, both business and IT, that this will be an investment, a change in how things are done, but it will result in long-term benefits. These benefits include saving the cost to fix data problems later, and improved data quality that results when data is captured closest to where the real details of that business interaction are known.

## **Manually Including Business Data**

There are instances where critical business data is not currently being captured in a formal operational system. This may be data that is needed to complete the required reports and analyses. Often this data is stored in spreadsheets on someone's hard drive or on a network shared drive. It is not acceptable to have anyone manually enter data into the data warehouse.

The first thing to do is explore the need for a small operational system to support the business function that captures this data. If the need is identified, a separate project should be initiated to address that need. Sometimes, the only need is to provide a place to collect this data in an organized, systematic manner. A small subsystem can be developed to provide a user interface that enables the data to be entered into a staging data table, rather than having the data entered on a spreadsheet. This staging data can be used as input to the ETL system. This adds more checks and balances to the data entry process and ensures that the data is integrated in a consistent and approved manner with the rest of the data in the data warehouse.

## **Tracking Progress—Are We There Yet?**

---

The design and development of the ETL system is very involved and can take quite a bit of time. It does not help to have only three tasks on a project plan: design, develop, and test. While these are the big tasks, it can be difficult to track real progress at such a high level. This may be all that is needed to share progress with upper management, but at a working level, more detail is needed to appreciate the work that has already been done and to see how much is left to do.

The tasks can be broken down to list each dimension and fact table, or even further if necessary to achieve small enough units of work to be done. As the detailed design, development, and testing is completed for each, it can be tracked, allowing a more accurate assessment of what percentage of the work has been completed. It is helpful to know that the ETL system is 60% complete; it is more interesting to know that eight of the ten dimension tables have been developed and one of the five fact tables is done.

Progress can even be measured in the number of known data problems. Then you can see progress as these are resolved. Without this, you don't even know how many known issues still need to be dealt with. Are there 25 or 125? Encourage the team to keep a log of all open questions. This can be used to track what still needs to be addressed. Assign a tracking number, a general category (perhaps the name of the dimension or data source to which the question relates), include a description of the problem or question, the date it was identified, the person responsible for finding the answer, and the status (opened, closed, deferred).

Over time, include notes about what has been learned and any decisions that have been made. This also serves as an audit trail showing how data issues have been resolved. This type of a log can easily have more than 100 items. The items can be prioritized to help the project team focus on the most important issues first.

At some point you are likely to get down to a dozen issues that remain open. You need to assess which of these must be addressed. From a systems perspective, can you move forward without having an answer? From a business perspective, is this issue critical? What happens if you can't find a resolution? If there is little or no impact (on systems or business), then this may be postponed to a future iteration.

## **What Else Can You Do to Help?**

---

While the ETL developers are working diligently, the rest of the team can feel helpless. However, there are several concrete steps during the development of the ETL system that you can take to contribute to the project's success.

### **Encouragement and Support**

Project teams are usually very aware of the time pressure to deliver results. They put in a lot of effort, often giving some of their own time to keep things moving forward. With so much work to be done, it can feel like a thankless job to the ETL developer. At times, more problems are identified than are being resolved. After a while, it can seem like the work will never end. It may still be weeks before the ETL system is complete.

Too often, the ETL developers only hear complaints and other negative feedback. Every time they need something, it feels like they are interrupting other more important work. The only time you hear from them is when they have found another problem. This is when managers and members of the business community need to step forward with a few words of encouragement. This can be a simple e-mail to thank them for their efforts. Even better, stop

by in person to thank them. If it has been a particularly bad week, bring in bagels, donuts, or even homemade cookies. You don't need to spend a lot of money, but a simple acknowledgment can give the team a sorely needed boost.

This type of support is often viewed as "management's job." However, it can also be given to the team by anyone in the business community. Remind the project team why this matters to you and how it will help you with your work every day.

**TIP** Business representatives can help the project team by keeping a positive attitude, even when other demands are looming. When asked, set aside time to answer questions and focus on what the data warehouse project needs.

## Ensuring Continued Business Participation

The construction of the database often takes longer than any other part of the project. For smaller projects this may still be weeks, but for many large initiatives, this can take months. Database construction is not as glamorous and exciting as brainstorming requirements and developing a data model. This is detailed, technical work. Many business team members, as well as the business community in general, can lose interest in the data warehouse.

Once the design work is done, the time demands on business partners decreases. Other daily work easily fills in and begins to take precedence over the remaining data warehouse work. While there is not a need for the business team members to develop programs and technical processes, their continued input is still highly valuable. Business input is needed to answer questions, clarify data requirements, make decisions about business rules for processing, and assist with testing and validating the data. It can be difficult to estimate when this participation will be needed for the next several months. The project team may have a better sense of what to expect in the next couple of weeks. While the participation is not constant, it is still important to ensure a successful data warehouse. Remember that any delays in helping the project team will add up and delay completion of the entire data warehouse.

### COPING WITH UNCERTAINTY

The one thing that is guaranteed on a data warehouse project is uncertainty. There is the uncertainty that results from the very nature of trying to keep up with the pace of business and the changing marketplace. There is the uncertainty that results from scrutinizing the organization's data in ways that have never been done before. The most important step toward dealing with

*(continued)*

**COPING WITH UNCERTAINTY (continued)**

this much uncertainty is openly acknowledging that it exists. There are also several things that everyone can do to help:

- **Be patient** with each other as the project evolves.
- **Remain flexible**—Realize that it may be best to get started with some work before you have every “t” crossed and every “i” dotted. Even if you achieved that level of detail, many things will change as soon as you move forward anyway. This does not give the project team carte blanche to run at full speed without any planning; it simply means that you must find a happy medium: Do enough planning and design to ensure that you have a strong direction, yet be willing to adapt as you go.
- **Communicate openly and frequently**—This is the key to ensuring that progress is being made toward the ultimate goal: a sustainable data warehouse that helps the business realize concrete value for the organization.
- **Apply sound project management**—Continue to utilize standard project management techniques to run the project. This includes maintaining the project plan, watching for scope creep, administering change control as needed, continuing to conduct periodic status meetings, and publishing status reports.

## Proactive Communication

Project communication should be driven from the project team out. The project manager is the spokesperson to share progress, concerns, and changes in timeline, deliverables, or costs. This is often through the standard project office or project management channels of the organization. Usually, weekly updates are submitted and included with all of the other initiatives across the organization. A one-page snapshot with a few bulleted points and a green/yellow/red indicator is *not* sufficient communication between the project team and the key stakeholders of the data warehouse.

During the early requirements gathering and data modeling phases of a project, there is natural and frequent interaction between the project team and the business community. Now that the team is focused on ETL system development, the nature of the communication must change. There will be many weeks when there are no major issues that need attention and development work continues. For very large projects, this may go on for several months. It seems that there is nothing specific to say, other than “we are still working.”

While regular status reports are published, there is often a drop-off in other communications. There are often no executive briefings, and no meetings with business sponsors and drivers. While these should continue, experience shows

that when there is no major deliverable or milestone accomplishment to report, project teams are quiet. If you don't hear from the project team, be proactive. Request a meeting or ask for an update.

Another reason to stay in close communication is that business and IT management can help remove many roadblocks for the team. Too often, project teams feel that they should be able to solve everything themselves. For example, one project team was running preliminary tests for one part of the ETL system. Every night, the server crashed and their processes never completed. With some investigation, the team discovered that the machine they were assigned for development and early testing was also being used by three other projects. The machine was overburdened, and did not have the resources to support all of these projects. From the ETL developer's perspective, they set up a schedule to take turns checking on the server throughout the night.

Over time, this decreased the productivity of the team (because they were all tired) and the testing took much longer than expected. With a broader perspective, other servers were available and could have been allocated to the data warehouse team. When you are too close to the problem, sometimes you don't see alternative solutions. Regular communication between the project team, business, and IT management can help identify these opportunities. Therefore, if you don't hear from the team, then seek them out to see how they are doing.

## **In Real Life**

---

Let's check in with the two companies we have been tracking now that they have moved into design and development of their ETL systems.

### **Building the Data Warehouse at Giant, Co.**

Job functions and roles are highly specialized at Giant Co. Everyone does their part and rarely looks beyond their immediate task. The organization has one group of people assigned to develop the technical and functional specifications for building the ETL system, but there was some trouble getting access to the right business people to develop the specifications. Therefore, the team just made the best decisions they could to complete the specs on time. After, with iterative development, there will be more work in the future. The developers are highly skilled and march forward to build what has been designed. However, as work progresses, many things crop up with the real data. The specifications do not include any instructions about how to handle these issues, and it takes too long for questions to be answered. The developers do what they can, but often data is excluded because there is no direction for handling it. This keeps the schedule on track, but as more and more data is left out, overall value begins to diminish. The entire project is still taking longer than expected.



A great deal of animosity also grew between the design and development teams. Changes continue to be made to the specifications to address discrepancies in the data. The developers feel that if the design team had done a better job this would not have happened. Likewise, the ETL designers feel that if the data modelers and analysts who gathered the requirements had done their jobs better, then there would not be so many changes.

With the situation deteriorating, the project manager called a time-out. The project team met to review the status and what would need to happen to get the project back on track. The project manager then met with the business and IT sponsors to share these findings. Although it was hard, it was decided to take a step back and develop a new plan of attack. These activities included the following:

- Compiling an inventory of what parts of the ETL system were working and what were not
- Comparing the project scope with the data currently being worked on, studying what needs to be added back in to meet the business objectives
- Documenting known data problems
- Revising the project plan, including cost and time estimates to complete the rest of the project
- Gaining approval for the revised project plan
- Developing specifications for remaining ETL work
- Assigning appropriate business people to assist with the specification and support the rest of the process
- Conducting team-building sessions to defuse the situation

While it was not easy to decide to revisit the requirements for the ETL system, it enabled the team to get back on track, with management support. Much of the work could be used. The experience and lessons learned were also documented while the issues were still fresh.

## **Agile, Inc., Builds a Data Warehouse Quickly**

There is a great deal of pressure to move quickly at Agile, Inc., to get the data warehouse up and running. The longer it takes before data starts moving, the more anxious the business gets. With the business requirements gathered and the database designed, the team gives in to the pressure to start building the database, with minimal time spent writing any specifications for the ETL system. The general consensus is that the team knows where the data is today and what it should look like, so it's time to get started. Everyone is empowered to make decisions and get their work done.



Sometimes business input is solicited to address data questions. This is not well received because everyone is so busy pulling data and producing reports for the upcoming quarterly management meetings. Therefore, the ETL developers just forge ahead and don't bother the business with questions. After a short development time, data is loaded and deemed ready for use. However, now that the business begins to look at the data, serious problems are identified. It turns out that many of the assumptions used to process the data were not valid. While the team met the short timeframe, the data is not usable.

After a great deal of finger pointing, a team is put together to conduct a review. The purpose is to determine whether it would be better to clean it up or start over. Because very little was documented and there was no cohesive system design for the ETL processes, it is determined to simply start over. This does not mean starting from scratch—the project team learned a lot about the data and about ETL development. Before diving in to build more processes, a joint business and IT data governance team is formed to study the data and develop standard definitions and guidelines regarding how data elements are to be handled. With this joint team to help answer data questions, the new project includes time to develop complete specifications. The project is going much better and the preliminary results are good. The data will be trustworthy and help meet the business requirements.

---

## Summary

---

The ETL system is at the heart and soul of the data warehouse and must be driven by detailed requirements for the data. Taking the time to understand and document these requirements helps to design and develop a robust production system. A lot of work is done by the ETL system to extract, transform, and load the database. This chapter introduced the basic functionality of these steps.

While much of this work is technical in nature, the business perspective must still be represented. Business representatives must be involved in defining business rules for handling the data and a test plan and detailed test cases. Once development begins, the business must continue to support the ETL developers by answering questions about the data promptly and enabling cross-functional group decisions when necessary.

The business needs to continue to provide input and guidance to the technical team members to ensure that the data is prepared properly. Continued partnership between business and IT can help each data warehouse project to successfully load high-quality data. With the data loaded, it is time to take a closer look at how it can be used. The next chapter focuses on delivering data into the hands of the business community.



# Data Delivery: What you Finally See

Data delivery is one of the most exciting parts of the entire DW project. You finally get to see the results of all the hard work. All of the input that went into gathering requirements and developing the data model fed into the construction of the database finally comes together. As soon as some sample data is loaded, the team can start building reports and screens where you can see the data. Excitement begins to build because the finished product is within sight. In many cases, this is what was “sold” as the data warehouse project result. Delivery of the data can take many different forms depending upon the job function and the need for data to support that work. In this chapter you will learn about the following:

- Business intelligence (BI)—what is it?
- The basic building blocks of a BI application
- Understanding different levels of use
- Building a BI application
- Planning for a successful launch

This chapter provides an introduction to these concepts but it is not intended to be a complete reference guide for BI. To get started, let’s take a look at what BI is and how it can work.

## What Is Business Intelligence?

---

Business intelligence is a broad all-encompassing term that is used to describe a wide variety of different types of data access. *Business intelligence* is the collection of one or more reports or analyses that provide insight into performance of a business organization. These reports and analyses are typically interactive to enable further understanding of specific areas of interest. These are used to support business professionals in their decision-making processes.

Over the years there have been many different labels used for this function, including analytical reporting, decision support, and now business intelligence. The term business intelligence has even been broadened to encompass processes, technologies, techniques, and tools that support actionable decision-making. Some also include the data warehouse itself within the scope of BI. For the purposes of this text, the term BI is used to refer to the actual use of the data warehouse to support decision making. The rest of this chapter explores this perspective in more depth. What does this mean to you? Simply that you need to clarify the use of the term BI within your organization to ensure effective communication and productive discussions.

#### **WHAT IS A QUERY?**

**The mechanism to get data out of a database is called a *query*. A query is comprised of several parts. First, what to include in the returned results is listed, such as customer name, product code, and number of units purchased. Second, a query must specify which database tables contain the information needed, such as the customer sales table. The query also includes a list of constraints that limit the results that are returned. For example, a query can be constrained to return only customers who live in Chicago, and only include pet food products, for sales in the last four weeks. A query can also include calculations (sales units \* price) and define how the results should be grouped or sorted. For example, the sales results could be grouped by customer gender and brand of the product.**

**The most common query language is SQL (structured query language), which is used to access many relational databases. Thankfully, there are many tools that provide user-friendly interfaces to help set up queries and generate the actual code that is sent to the database. These tools also provide a wide variety of presentation options, making the results much easier to understand and interpret. Queries are how the data is accessed in the data warehouse.**

### **Business Intelligence without a DW**

Some organizations have developed business intelligence solutions that are completely independent of the data warehouse environment. This can limit the long-term value of the business intelligence solution. These point solutions are usually focused on a single area of the business with a limited set of data geared toward supporting a specific set of reports. The data is often structured specifically to produce these reports. It can be difficult and costly to add more data and expand this type of application over time, as needs change.

The best business intelligence solutions are those that are built on top of a sound data warehouse foundation. This foundation enables the environment to flexibly grow and change over time. New data can be gracefully added to

the database and more analyses developed. The database is designed with the intention of supporting flexible reporting in ways not yet defined and with the expectation that more data will be added over time.

## BI in Action

There are many different ways to deliver data from the data warehouse to the business users. Often this data access is provided by a business intelligence tool. This section walks through a progression of data delivery that shows the types of things you can expect from a business intelligence tool.

### Tabular Reports

*Static tabular reports* are used to pull a subset of data from the data warehouse in a tabular format. These often provide a snapshot of business performance or a trend of business results over time. Figure 11-1 shows an example of a static tabular report. This report shows monthly sales for the third quarter, by vendor.

3rd Quarter Purchasing by Vendor							
Vendor Name	Year Ago September	July	August	September	Chg vs. Prev. Month	% Chg vs. Previous	% Chg vs. YA Month
Applicances	\$28,408	\$26,883	\$30,935	\$28,192	-\$2,743	-10%	-1%
Clothes For All	\$3,400	\$3,189	\$3,202	\$3,309	\$106	3%	-3%
Dream Homes	\$20,277	\$20,434	\$18,402	\$21,875	\$3,473	16%	7%
Enviro Electronics	\$21,482	\$20,880	\$21,201	\$22,185	\$984	4%	3%
Everything Entertainment	\$16,737	\$15,659	\$17,114	\$15,576	-\$1,539	-10%	-7%
Fill Your Pantry	\$6,177	\$5,932	\$5,698	\$5,335	-\$363	-7%	-16%
Making Good Stuff	\$15,794	\$14,674	\$15,649	\$16,572	\$923	6%	5%
Office and School Inc.	\$1,578	\$1,783	\$1,687	\$1,563	-\$124	-8%	-1%
Pack Your Bags Travel	\$5,158	\$5,499	\$5,352	\$5,968	\$616	10%	14%
Patriotic Gear	\$2,296	\$2,167	\$1,948	\$1,876	-\$72	-4%	-22%
Shoes Shoes Shoes	\$4,716	\$4,380	\$4,202	\$3,538	-\$664	-19%	-33%
Sports Galore	\$3,028	\$3,281	\$3,344	\$3,964	\$620	16%	24%
Warm Homes	\$13,204	\$14,102	\$17,443	\$20,046	\$2,603	13%	34%
Worldwide Stuff	\$3,220	\$3,052	\$3,672	\$3,556	\$385	11%	9%
Zippy Manufacturing	\$4,066	\$4,588	\$4,028	\$4,273	\$244	6%	5%
<b>Grand Total</b>	<b>\$149,541</b>	<b>\$146,503</b>	<b>\$153,378</b>	<b>\$157,827</b>	<b>\$4,449</b>	<b>3%</b>	<b>5%</b>

**Figure 11-1** Sample static tabular report

### Parameter-Driven Reports

Rather than build dozens of static tabular reports, several report templates can be created that are driven by user-defined parameters. The user is prompted to provide the input parameters when the report is requested. For example, a user who wants the report from Figure 11-1 could be prompted to select which quarter and vendor to include on the report.

While some tabular reports are static and unchangeable, more often the reports are parameter driven, so a business user could be prompted to select the appropriate time frame and sales region for the report.

## Interactive Reports—Drilling Down and Across

Reports are even more useful when they are interactive, enabling the user to drill down into more detail. Using drill-down capabilities, you can, for example, start with a regional report and then get more detail about each of the districts that make up the Midwest region. Interactive reports are the backbone of business intelligence solutions and provide straightforward presentation of selected data. Likewise, it is possible to roll data up to a higher level of detail.

Some tools also provide additional manipulation of the results by swapping the row and column headers. For example, you may be able to swap out the vendor for the product category. You may also be able to change which business measures are displayed on the report. Using these drilling and manipulation capabilities, it is possible to completely morph a report without having to go into a full development mode.

## Exception Reports

*Exception reports* are similar to tabular reports in their format but have the added capability to return results that meet specific criteria. For example, you can run an exception report if you want to list all sales representatives who are within 8% of meeting their sales quota for the month. The exception criterion in this example is the threshold of 8%. Users can be prompted for exception criteria. Exception reports can return only the results that meet the criteria, or exception conditions can be identified in the report itself. This is often done by using different colored text, shading the cell, or displaying a colored shape such as a circle or triangle. For example, perhaps all products whose sales are lower than last year's sales are shown in red. Figure 11-2 shows a sample of an exception report that includes only the vendors whose sales are less than zero percent of the previous month's sales. Note that cells whose sales are less than or equal to negative 10% of the previous month's sales are shaded for additional emphasis.

3rd Quarter Vendor Watch Report							
Exception Criteria: This report includes only Vendors where Current Month Sales are less than 0% of Previous Month Sales							
Vendor Name	Year Ago September	July	August	September	Chg vs. Prev. Month	% Chg vs. Previous	% Chg vs. YA Month
Applicances	\$28,408	\$26,883	\$30,935	\$28,192	-\$2,743	-10%	-1%
Everything Entertainment	\$16,737	\$15,659	\$17,114	\$15,576	-\$1,539	-10%	-7%
Fill Your Pantry	\$6,177	\$5,932	\$5,698	\$5,335	-\$363	-7%	-16%
Office and School Inc.	\$1,578	\$1,783	\$1,687	\$1,563	-\$124	-8%	-1%
Patriotic Gear	\$2,296	\$2,167	\$1,948	\$1,876	-\$72	-4%	-22%
Shoes Shoes Shoes	\$4,716	\$4,380	\$4,202	\$3,538	-\$664	-19%	-33%
Worldwide Stuff	\$3,220	\$3,052	\$3,672	\$3,556	-\$115	-3%	9%
<b>Grand Total</b>	\$63,132	\$59,856	\$65,256	\$59,635	-\$5,620	-9%	-6%

**Figure 11-2** Sample exception report

**AD HOC REPORTS**

The term *ad hoc report* has different meanings depending upon who you ask. Typically, those with technical backgrounds tend to think that ad hoc means you start with a blank slate and build a report from scratch and then run it to get your results. When most business people ask for ad hoc reporting, they really want parameter-driven reports. Be sure to clarify what is really needed when someone asks for ad hoc reporting.

***Other BI Capabilities***

Beyond creating reports, additional functions are included as part of BI. Some of the capabilities that fall under the BI umbrella include the following:

- Comparisons (actual sales vs. forecasted sales)
- Ranking (top/bottom  $n$ , top/bottom  $n\%$ )
- Totals and subtotals
- Time series comparisons (this week vs. same week one year ago)
- Ratios and penetration (market share-product sales as a percentage of the total category sales for this market)
- Exception processing to highlight conditions or limit the results included on the report. Exception processing includes the following:
  - Top  $n$ , bottom  $n$  (by % too)
  - Thresholds
  - % contribution of what is included in the report
  - % contribution to the whole data set, not limited to what is included on the report

These are used to help glean insight from the data.

***Complex Analysis***

*Complex analysis* can also be produced with a business intelligence application. This is where more sophisticated business logic, statistical analysis, or data mining is performed on the data to produce the final results. For example, the seasonally adjusted sales may be presented with baseline sales differentiated from promoted sales.

This is not an exhaustive list of BI capabilities but a quick overview. Many more variations are possible to produce results. This is intended to give you a basic idea of the range of things that are possible. In addition, note that not

everyone uses the terms discussed here in exactly the same way. Sometimes the BI labels are used because they align with the name of the software module used to create the reports. Sometimes the terms are rooted in how your organization has historically delivered data. It really does not matter whose terms are used, but it is important for everyone on the project team and, it is hoped, across the organization to use these terms in a consistent manner.

## **BI Building Blocks**

---

When you use a BI solution you tap into several different layers. These different layers are created to enable flexibility and the sharing of components across many reports and applications. When using a BI solution, you generally are not aware of these layers or how they work together. It is useful to discuss them here because it will help the entire team to have more meaningful discussions when refining requirements and developing the applications for the DW.

### **Data Content—Understanding What You Have**

The foundation of any BI solution is the data itself. All of the data must have useful business names and descriptions, rather than technical names used by the database. There are two parts of defining the data content.

First, the data itself must be set up so that it can be used to create and navigate through the BI reports and other display mechanisms. This is where the Business Dimensional Model described in Chapter 7 is used again. Each of the business attributes for each business dimension must be defined to the BI tool. This entails linking the business data element to the physical data element in the database. Additionally, the BI tool needs any relationships between each of the attributes to be defined. This is how the BI tool “knows” what to drill down to when asked. The labels that are set up here are what show up on reports and as labels for setting parameters. This is setting up the BI tool to present what is defined in the Business Dimensional Model as discussed in Chapter 7.

Second, the actual business measures that appear as the body of the reports must be defined. This work will use the Business Measures Worksheet described in Chapter 7. Think of this as setting up a library of all of the business measures that will be used. These may be basic facts, key performance indicators, or other business metrics. Each should have a descriptive name and a business description, and each should define the default handling of this measure for rolling the results up. This is also called the *default aggregation rule*. For sales units, for example, the default is usually to perform a sum. For an average price measure, the default aggregation rule should be to sum the dollar sales divided by the sum of the units sold, rather than the average of the average price itself.



The Business Dimensional Model and the business measures worksheet should utilize and align with any work that is done for master data management, as described in Chapter 8. This work is done by the BI developers and provides the basics needed to be able to create and run reports using a BI tool. This is also called *configuring the BI tool* or *defining the BI tool metadata*.

**NOTE** Developing the Business Measures Worksheet should begin during the dimensional modeling process. This helps to ensure that all of the underlying facts required to support these business measures are identified for inclusion in the model. This also allows time for the formulas to be approved, as it can take a long time to gain consensus across business groups for some of these business measures.

## Navigation—Finding What You Need

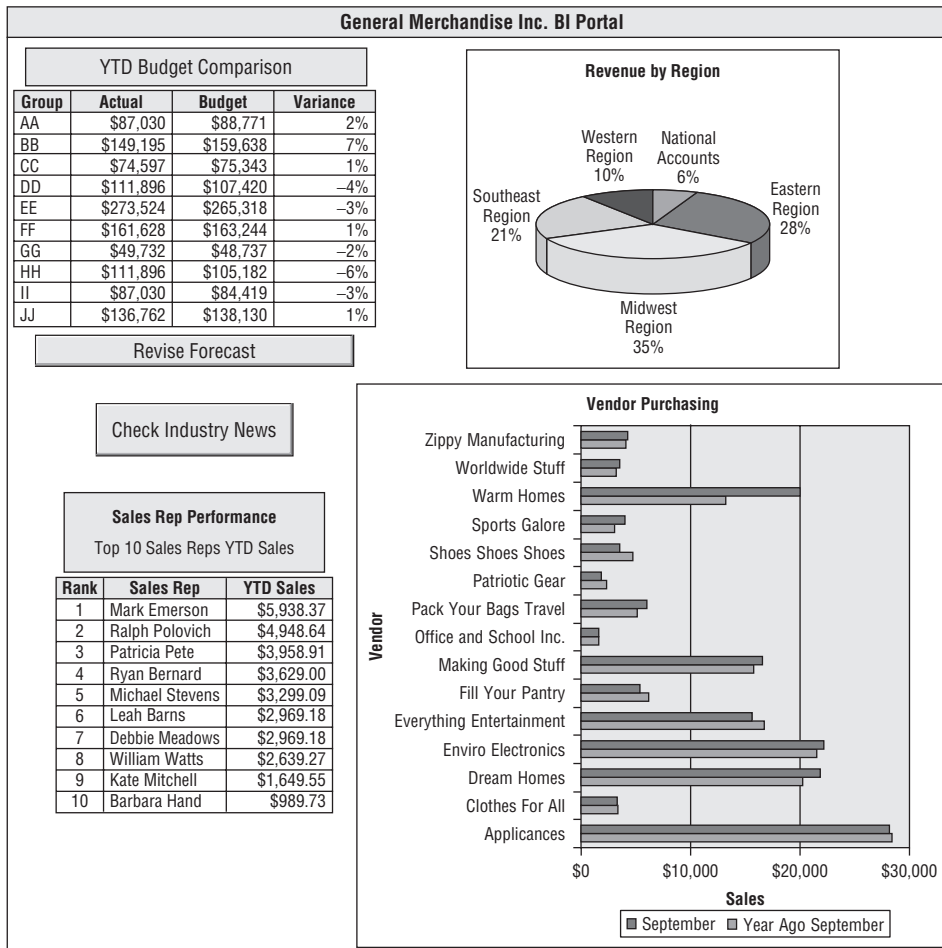
There is a wide range of possibilities for getting started using a data warehouse. The navigation layer provides the interface that enables the business users to locate and then use the data they are looking for. BI navigation may involve going to a shared network drive, finding the BI Reports folder, and then looking at the different reports that you can run. More likely, there will be a web interface that presents the different reports that can be accessed and run by the users. Figure 11-3 shows just one possibility of what a BI portal could look like. This is a conceptual drawing, rather than a screen shot from a specific product. A simple Internet search will yield many different examples of what the current technology can deliver. (Note that a BI dashboard could look like a BI portal; it really depends upon the tool and what your organization decides to call it.)

Perhaps the BI application is just one of many pieces that are included in the business group portal. In this case, the BI reports and data sit side by side with other applications and functions that are used by that business group.

## Presentation—How Do You Want to See Results?

Once you have navigated your way to find a report or screen of interest, there are many different ways that results can be presented. Some results are pre-defined and are displayed when requested. In other cases, parameters or variables may need to be set before the report is run. Generally, options and lists are provided to aid in the selection process. Once each of the parameters is set, the report is run. The results may be returned in a wide variety of formats, including the following:

- **Classic tabular reports:** Introduced earlier, these are a common method for presenting results. This is a familiar format that most people are comfortable with.



**Figure 11-3** Sample BI portal

- **Charts or graphs:** These are another common type of data presentation. These may be bar charts, pie charts, line graphs, or scatter charts. A graphical display of results helps to easily identify anomalies and outliers. This can help you to quickly hone in on areas for further research.
- **Maps:** These are another visual method to quickly convey data. They may be color-coded based upon selected data or bars may rise from the geography to indicate the value of the business measurement. For example, maps can be colored or coded to highlight concentrations of customers or to indicate areas where insured property is located.
- **Scorecards:** These are the display of key performance measures in a “report card” manner. These key measures are calculated and compared to target values as a means of determining how well the organization

is doing. Often the scorecard represents the results of many different analyses to determine each of the different measurements and areas of the business. These are combined into a single display to provide a comprehensive view of the entire business or a specific area of the business. The most useful scorecards are interactive, which means that they support the capability to dig deeper into the measures. Figure 11-4 shows what a scorecard might look like. Again, this is a line drawing to illustrate the concept. Actual scorecards also use color to help highlight areas to focus on.

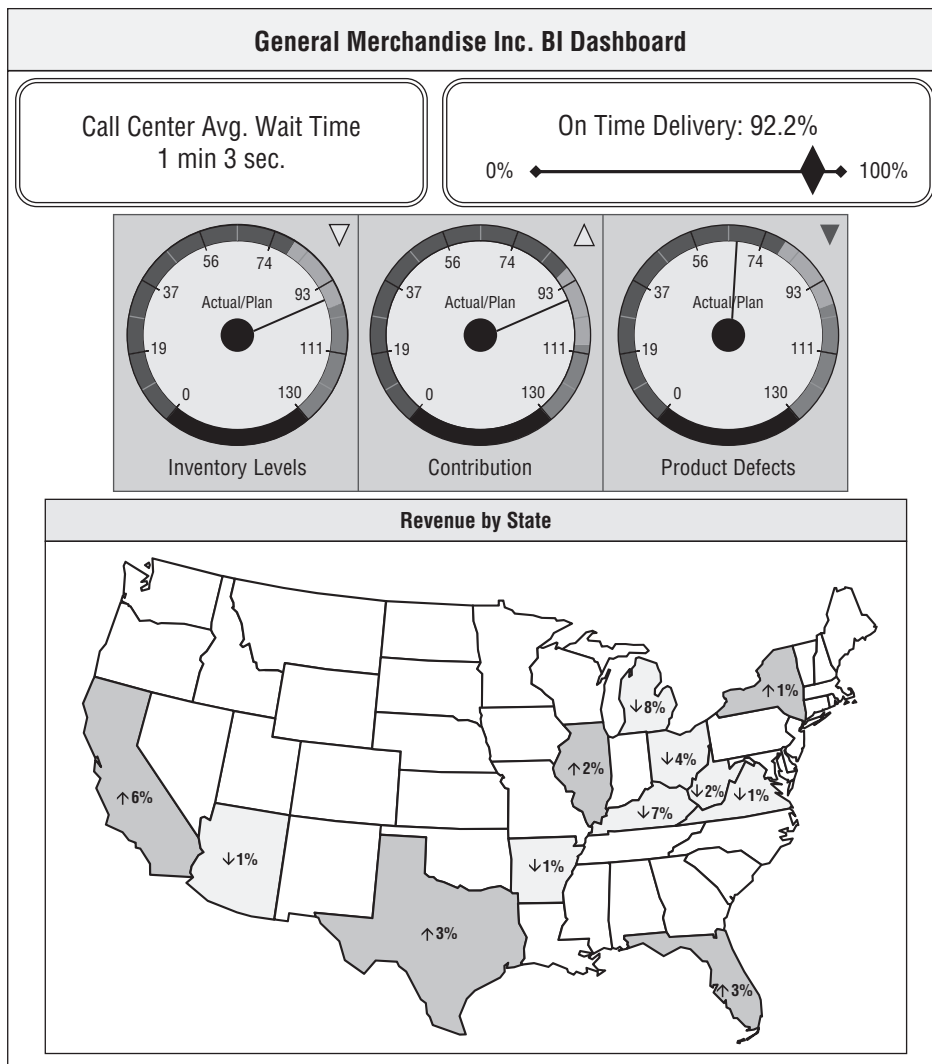
General Merchandise Inc. Performance Scorecard				
Financial	Status	Actual	Target	% off Target
Revenue	◆	\$57,032,110	\$56,000,000	1.8%
Expenses	◆	\$6,444,628	\$7,840,000	17.8%
Contribution	●	11.3%	14.0%	-19.3%
Customer				
New Customer Count	●	370	450	-17.8%
Satisfaction Ratings	◆	78	75	4.0%
Complaints	●	2300	2000	-15.0%
Manufacturing				
Production Volume	△	339,588	350,000	-3.0%
On Time Delivery	△	97.2%	95.0%	2.3%
Product Defects	●	1.1%	0.8%	-37.5%

**Figure 11-4** Sample scorecard

- **Dials and other visual representation:** These types of report can provide at-a-glance interpretations of results. Rather than see a tabular report with actual sales compared to forecasted sales, a dial can show the actual sales, with the sales forecast value as a line or bar across the dial at the appropriate target number. This is similar to a typical United Way fund drive poster. A giant thermometer shape is posted on the wall. The target is a line near the top of the thermometer shape. As employees donate, the thermometer is filled in. You can easily see how far the organization has moved toward the goal.
- **Dashboards:** This is a term that is used to describe the display of many different business measures at the same time. The business intelligence dashboard results combine a graphical and numeric display. The results may be shown as dials, and color is used to highlight various metrics. The intent is not to fill the screen with numbers, but rather to identify the most critical measures and present them in a way that enables a quick understanding. Figure 11-5 is a line drawing of a sample dashboard. Like scorecards, dashboards also make use of color to facilitate the interpretation of the data. Dashboards often contain a variety of financial and

operational results, compared to targets or goals. The intent is to provide a quick at-a-glance view of the state of the business, similar to how the dashboard of a car provides information about all critical functions.

As with most business intelligence applications, the ability to interactively drill into more detail increases its usefulness. For example, the dashboard may show a drop in customer satisfaction. Drilling down may show that the customers who were recently assigned to the newly created southeast sales region were unhappy with the change. There was not an overall customer satisfaction problem, but clearly some action is needed to address all customers in the new southeast region.



**Figure 11-5** Sample dashboard

While there are many different ways that results can be presented, there is still one more step to consider. How will someone receive these results?

## Delivery—How Do You Receive the Results?

There are many different ways that the results from a BI application can be delivered. The delivery mechanisms vary between BI tools and should be aligned with the needs and skills of the different target audiences. The following is simply a representative sample of what can be done.

- **BI tools** typically include the capability to directly display the results of reports and analyses. These are presented on the user's screen and may be enabled to support interaction, such as drilling down.
- **Standard office application output** includes word processing, spreadsheets and PDF, which are common formats for data delivery. These put the results from the BI tool into formats that are easily used by other commonly used office productivity tools for further manipulation and formatting. Spreadsheet output enables additional analysis to be performed. This type of output can be dropped onto a user's desktop.
- **E-mail attachments** are another way to receive output for standard office application output. A set of standard reports may be generated and automatically e-mailed to the appropriate people. In other cases, a person using the BI application can decide to send the results of his or her analysis to someone else by attaching the output to an e-mail message.
- **E-mail notification** is when the data from the data warehouse is included as part of the body of the e-mail message itself. This is often done for exception reporting. An e-mail is generated when certain exception conditions are met. For example, an e-mail may be sent only when the number of insurance quotes issued is more than 10% lower than the prior year's quotes for the same month to date. This enables agents to see that at the current rate, they will not make their goal for the month.
- **Non-PC delivery** on other devices is increasingly popular. This may be receiving e-mail messages on a personal digital assistant (PDA) or a sophisticated cell phone. The critical difference here is that the format of a message must be carefully designed to be read on such a small screen. The most critical content must be concisely displayed so that as much as possible can be viewed on the initial page. This may require quite a bit of testing to ensure that the results are presented in a useful and meaningful way.
- **Other file formats** may be created as a means to provide data to other specialized applications. These may provide additional analytical capabilities. Sometimes data feeds must continue to be created until all existing reporting systems can be shut off. This is usually the case when the

data warehouse does not have all of the required data yet, so the old system continues to run until the data warehouse can fully support the functionality.

With the many different possibilities for navigating, presenting, and delivering results from the BI application, it is important to understand who the audience is for your BI application. There are typically multiple types of users of the data warehouse. BI applications are often tailored to meet the specific needs of these different types of users.

## Supporting Different Levels of Use

---

The project team needs to have an understanding of the different audiences who will be using the data warehouse. Depending upon the audience, different data delivery methods can be used to target each group's specific needs. There will be differences between groups based upon frequency of use and/or job function. Each organization must determine which business roles and/or groups need access to data and what the expectations are for them to perform analysis. Using this input, the BI application can be tailored to meet these different audiences. The following list describes several different types of DW users in order to illustrate some of the possibilities:

- **Mainstream users:** These are users who access the data warehouse business intelligence application on a regular basis. The data is needed to help these people complete their day-to-day responsibilities. These regular users become very adept at using report templates to create their own reports. They often save these as new reports for themselves and/or their co-workers. Mainstream users do not perform advanced analysis, but use the results of analyses developed by others to support their work. They generally prefer to use structured navigation to get to the reports that they need. This may be by using a BI portal or by using a web interface of the BI application.
- **Business analyst:** This is someone who needs to track what is happening in the business and to understand the reason for changes in the business. This job requires them to go beyond looking at reports. They need to perform analyses on a regular basis, perhaps to support business reviews. Business analysts also work on requests to look into fluctuations in business results. These are the people who create new reports. However, a business analyst does not usually have the ability to change the building blocks of the BI application itself. Business analysts can have a wide range of preferences for diving into the BI application. BI portals and structured navigation work well, but the business analysts must be able to modify and extend the starting reports to address their requirements.

- **Casual Users:** These are users who occasionally come to the data warehouse to get answers. This is not part of their daily work. Casual users generally use the BI application in a structured manner. They access pre-defined reports, and are able to perform some limited manipulation of the results. They are not typically interested in performing more in-depth analysis, but rather are simply looking for a single report. The BI application must be set up to ensure ease of use for these casual users. Everything must be laid out intuitively and clearly labeled.

Casual users generally have structured and guided paths to get results from the BI application. Pulling up pre-defined reports is usually all they are allowed to do. They would have limited flexibility. When there is a need for a different report, casual users would ask someone else to create it for them.

- **Consumers of results:** There are many people in the organization who are consumers of the BI solution. These people may not even know that they are using a data warehouse at all. Information is provided to them as part of another business process or served up automatically for them. For example, when you make a purchases online, other recommended products are often offered by accessing the purchase patterns of other people who bought the product you just put into your electronic shopping cart. As another example, some people may sign into the network first thing every morning and see a company news page that includes sales results and quality measures. These are provided by the data warehouse.

As technology has expanded the reach of the BI delivery mechanisms, more and more people can benefit from the data warehouse, even though they may never actually log in and run a BI report themselves. Consumers of results have no idea that they are even accessing a data warehouse. The access is completely embedded in other applications, or results are simply sent electronically. They never go to a BI portal to run a report.

- **External users:** These users of the BI application are people who have business relationships with the organization and can benefit from seeing information about that relationship. External users may be customer, vendors, suppliers, agents, or other business partners. This use may be in the form of providing access to sales history to your customers, or providing business results to franchise managers. These users can have a profile like those already described, such as business analysts or casual users. If there are any external users, there are additional security responsibilities that need to be addressed to ensure the safety and privacy of all the data in the data warehouse. Most often, external users have structured access to the data warehouse. They may have limited flexibility



to drill down into more detail, but usually they are not allowed to make any changes to reports. External users have a similar profile as regular users, but because they are not employed by the organization, additional security requirements must be in place.

- **Power users:** These may be a business analyst or a BI application developer who reports to the business itself. These users have often been writing programs and complex spreadsheets for years to produce reports. They are often the most knowledgeable people in the organization about the strengths and pitfalls of the data. When power users are able to use the data warehouse, they are likely to build reports from scratch. Power users often have access to more capabilities and controls within the BI tool. They may set up the business measure library or refine reports developed by others so that they can be published with all of the other standard templates. The power user's role on a project is that of a BI developer. Once a project ends, their role in the business continues and they often develop new analyses and reports. Power users are typically the only people outside of IT who have the skills and interest to dig deeply into the BI tool's raw power.

## **Construction of the BI Solution**

---

Now that the basics of business intelligence have been discussed, let's take a closer look at what this looks like on a project. The major tasks sound similar to any other system construction effort, but the work itself is much different. The development work is typically done using robust business intelligence tools. As introduced in Chapter 4, there are two different types of BI developers. The first type is responsible for the more technical aspects, including configuration and setup of the BI tool itself. The second type of BI developer is responsible for the design and development of BI reports and applications. Because most of the development work is not programming, the resources to build the BI application could be from IT or the business. The tasks that are done are the same regardless of which group the BI developers report to. There are several steps involved in the construction of a BI application, but first take a moment to look at the bigger picture. What is the business really trying to accomplish?

## **Planning for Business Change**

The ability to realize the full benefit of the data warehouse often means that the business must change some processes. If business process change has not been studied yet, then take the time to do so before designing the BI application. Otherwise, it may simply automate existing processes or redirect where data is pulled. While the data warehouse project team may be able to provide some



ideas or input, evaluating and changing business processes should be done by the business group itself. Sometimes there are well-established groups within an organization that specialize in business process management. There are also plenty of external consulting organizations that can assist with business process change.

Initial changes may simply be to streamline current processes to take advantage of direct access to the data warehouse. Manual steps can be eliminated, and other complex spreadsheets may be able to be replaced with the BI application. Sometimes, this is the opportunity that you have been waiting for to make more significant changes. In either case, now is the time to consider how the BI application can help accomplish these changes.

## Design—What Needs to Be Delivered?

The design of the BI application itself is likely to be done shortly after dimensional modeling is complete. The project team has spent time learning and documenting the business requirements (see Chapter 6), and the dimensional model (see Chapter 7) provides details about what data elements will be available. Now is the time to think about what needs to be delivered.

The first step in this process is to identify each target audience and what type of use they need. Each different audience may have distinctly different requirements for the results they need to see. Who needs to look at what data? Market researchers have much different data needs than purchasing agents. The data used for analysis by the marketing department to plan promotions for the next year is different from what a sales rep running a set of reports to support an account review needs.

While BI developers often have a lot of knowledge about what the BI application needs to do, it is still worthwhile to get input from others in the business community. This helps to build ownership across groups who will be using the BI application. The design of the BI application can be done via design sessions, or individual and/or group meetings. The design focus should be on the high-priority business areas that were identified during initial requirements gathering, and they should be within the scope of the project. Based upon the experience of the project team, much of this work can be done within the team. However, it is important to continue to develop support, which can be done by sharing ideas with other representatives from the business community. The steps of the design process are as follows:

1. **List needed reports:** For existing reports that are still wanted, ensure that they are identified so the team can use this as a starting point. For new reports, draft a brief description of the report. For example, an Agency Performance Report would include premiums, claims, and new policy close rates compared to the previous year and to the current year's target.

2. **Prioritize candidate reports:** Even within the scope of a project, there may be dozens of reports that could be built. It is not feasible to deliver all of them at once. The project team can do some preliminary prioritization of these candidate reports. This task should be shared with the business champion, and possibly the executive and IT sponsors for their input and review. The rest of the BI application development efforts can continue with the approved priorities.
3. **Brainstorm reports:** Sketch out on a white board what each report might look like. Focus on the row and column headers and the business measures needed. If a series of reports is wanted, explore which report you want to see first, and then outline what should follow. Perhaps use a spreadsheet to capture these ideas.
4. **Prototype:** In some cases, it may be helpful to do some prototyping of reports. Once the dimensional model is complete, sample data can be loaded into a database so that the BI tool can be set up. Drafting candidate reports may be able to happen “live” with the tool and the sample data.
5. **Determine navigation:** It is also important to work with the business to draft ideas for navigation. Opinions on how reports should be organized can vary dramatically between a technical and business viewpoint. Technical people often like to group reports based upon type of data. Business people tend to group reports as they are used to walk through common business processes. This step should not be skipped even if the BI developers are from the business community. Discussions about navigation paths often occur after the reports are defined.
6. **Include security:** The project team must also pay attention to any data, report, or application security requirements. Who can access which reports? At what level of detail? These considerations must be part of the overall design.

The design process is not just an initial set of meetings, but should be a series of interactive sessions. In a conference room, it is very easy to dream up fantastic reports that are impossible to build. Often these ideas need further clarification. At the time the report is dreamed up, it makes sense, but the next morning the report may not be as clear. For example, suppose there is interest in building a report to show profit by customer category. Unfortunately, with further thought, it is realized that the profit measures are not going to be loaded initially with the customer dimension. This means that profit will not be available by customer category. Perhaps the standard profit fact is not what is really needed. This requirement could be met by providing estimated profit, which is the sales dollars less the estimated manufacturing and shipping costs. If the standard profit really is needed, then this reflects a gap in the initial

project delivery that must be addressed. Communication may be needed to reset expectations, or the team can estimate how much additional work is needed to get the required data.

On a more technical note, spend time designing a common look and feel to be used for all output from the data warehouse. Decide what headers and footers should be in place and create recommendations for the font and size of the report headings and body. All reports that are distributed through the common/standard library will adhere to these standards. This makes all of the output from the data warehouse look consistent and professional.

## Development

Although this part of the project is referred to as BI development, it is quite different from most systems development efforts. Most of this work is done using graphical interfaces, and does not typically involve much programming or coding. This work can be divided into two categories. The first is work that is generally done once to set things up. The second category is work that is done repeatedly on a project, on an ongoing basis. Each technology will have an impact on the specific tasks that need to be performed and what these tasks are called for that tool. Regardless of the technology used, several of the initial steps are common:

1. **Installing the technology:** This involves getting the software loaded and ready for use. The installation is done up front, by IT. Individual software vendors provide the details for installation.
2. **Configuring the BI tool:** This typically involves setting up the metadata. This is where the physical structures are mapped to the business labels for each data element. It also involves setting up relationships or drill paths between the dimensional attributes. The facts themselves are also defined. The Business Dimensional Model contains much of the business information needed to do this work, which is usually done by a BI developer, who may be from either IT or the business. Configuration tasks are only needed when there are changes to the data warehouse or new data is loaded.
3. **Setting up business measures:** This is done by setting up the calculations. The Business Measures Worksheet discussed in Chapter 7 should contain what is needed to set these up. Setting up business measures is also done by a BI developer up front.

**NOTE** If you are not using a BI tool, the development effort will entail more conventional coding, which requires technical skills and is done by IT.

Once the BI tool is installed and configured, development can begin in earnest. These steps are done to get the BI application off the ground, but the

same tasks continue to be done, even after the end of the project. Some of the steps that are repeated include the following:

- **Building report templates:** This involves setting up the report layouts, exception criteria, and parameters. It also includes developing how the results are presented, such as a tabular report, a chart, or a scorecard. Building the initial reports is done by a BI developer, either from IT or the business.
- **Creating additional reports:** Once the BI application is released for use, any other users who have been granted appropriate privileges can create other reports.
- **Setting up navigation:** Building the navigation paths is the work to organize and display how the starter set of reports can be located and accessed. This work is also done by a BI developer, but it can be a bit more technical. Often this is done by a BI developer from IT, but it could still be completed by a BI developer from the business community. This work is not typically done by other business users of the application.

The project work to build the BI application is almost done when the reports have been built and the navigation paths are set up. Now it is time to test everything, a very important step that must be done to ensure that the BI application works correctly.

## Testing BI Applications and Validating Data

The BI application must be tested in a similar manner to any other system that is tested. How the data has been defined in the BI tool must be validated. This includes ensuring that drill paths are set up properly and that business measures are being calculated correctly. Then, each report must be tested to ensure that it delivers what was intended. Some reports may be created properly but may not provide the expected insight. It is worth adapting these reports to be more useful, with the guidance of the business community.

The second type of testing that is done during BI application construction is validation of the data itself. Regardless of the extent of the testing and validation that was done during construction of the ETL system, more data problems will be discovered now. Due to the visual nature of BI tools, data problems are often quickly identified. A lot of this data validation occurs during the development of reports. As problems crop up, work with the entire project team to figure out the source of the problem and work toward a resolution. Data problems may result from errors in defining business measures in the BI tool, errors in the ETL processing rules, and sometimes even in the underlying source system. The investigation of data problems can be complex and time consuming.

**NOTE** One can quickly create reports using today's BI tools. Because many data problems may be identified, build enough contingency time into the project schedule to allow for this additional data validation and correction work.

## Additional Responsibilities

Building a BI application is more than just setting up a few reports and the job is done. As already discussed, the data must be defined, reports or results prepared for display, and the navigation path created. In addition, other administrative responsibilities need attention. You need to ensure that the right people have the appropriate level of access and you must prevent the wrong people from having access. This includes who can look at the data and who can make changes to the BI reports and applications.

### *Security—Who Can Look at the Data?*

Any security requirements should have been identified when gathering business requirements. Security requirements also surface when working on building the ETL system. However, the reality of securing the data may only sink in now that the data is going to be published. Security can be addressed in several different ways. Many security measures are part of how the BI application is constructed:

- Prohibit the loading of sensitive data into the data warehouse. If highly sensitive data elements are not included, then even if the security is broken, the sensitive data is not there. With medical records, this is often a method used to protect patient identity and comply with HIPPA privacy regulations. Name and social security number, street address, and phone number are not loaded into the data warehouse. A unique identifier, or surrogate key, should be assigned to each patient during the ETL processing. This surrogate key enables the patient data to be tracked and trended without revealing any personally identifiable information.
- Utilize the database management system security functionality. For most of the databases this is exactly the same security system that is used by the other operational applications. Therefore, if it is secure enough to store the ATM transactions, it is secure enough for the data warehouse.

**NOTE** Regardless of other layers of security, if the database management system itself is not secured, backdoor access may still be possible.

- Utilize the BI technology's security features. Most of the BI tools have some sort of security built into the tool. Take advantage of what is provided "out of the box."

- Leverage standard network security and firewalls. Many BI applications are released over the organization's intranet or the Internet. The BI application should adhere to all of the corporate standards for securing any application that is delivered using these methods.
- Build security into the BI application itself: There are several common situations in which data must be limited from view:
  - Often, organizational boundaries need to be recognized. For example, individual sales reps can look at their own results, but not others, whereas a district manager can see the results for the whole district and all of the reps who report to that district.
  - Security may be used to control access to pre-built reports. Access may be granted or restricted based by user group or individual user.
  - In addition to accessing a report, sometimes there is a need to secure data as you drill down. Consider a student reporting system. Identifiable student details such as name and social security number are not in the data warehouse, but if you drill down far enough you would be able to figure out the identity of some students (e.g., there is only one female student in third grade at Roosevelt Elementary School who is in a wheelchair). You can build into the BI application a control that limits the display of data if the set represents fewer than ten students.

It is important that all users of the data warehouse be informed about the organization's policies for use and distribution of data. This includes helping users understand what security is in place and why.

### ***System Controls—Who Can Change What?***

The security measures define who can access the BI application and what data will be visible. There will be many people who are granted access to the data but who should not be authorized to make changes to the BI environment and application.

BI technologies provide the capability to set up standard business measures and report templates. These can be defined and maintained by business analysts or power users. Only people who have been given the responsibility, and appropriate permissions, should be able to make changes to the standard components of the BI application. Some users may be given privileges to create their own business measures and reports. These are not part of the "library" that is available to the rest of the business community. These controls need to be in place to ensure the integrity of the BI environment itself.

Identify the individuals who will be given the responsibility of making changes to the BI environment. They must be provided with the appropriate training to know how to perform these functions in the BI tool. Additionally,

these individuals should have enough business experience to help ensure that any changes are valid and that new reports and measures are set up accurately.

## Planning a Successful Launch

---

Releasing a business intelligence application—whether it is a simple report or a full dashboard—is much more than just finishing the BI application and then putting a link on the web. Success requires appropriate education and support.

**NOTE** Regardless of how much planning and scheduling has been done to release the BI application, if it is not ready, then *do not move forward*. Postpone any education and launch activities until the data warehouse and BI application are ready for full use. Resolve any problems and then move forward.

## Marketing the Solution

As with the launch of any new product, releasing the BI application typically includes a marketing campaign too. You can do any or all of the following:

- Name the BI application, if you have not done so already. Use meaningful business names such as “Vendor Scorecard” or “Sales Dashboard.”
- Make posters to announce that the BI application is coming (using the business name).
- Include a note in the appropriate company newsletters.
- Arrange for a news flash to be posted on the main company intranet page.
- Send e-mail reminders to the target audience.
- Pass out buttons or candy.

These basic marketing activities need to communicate what the BI application will be able to do. This is best known and understood by the business representatives who have been involved with the project. The marketing messages should address the business value, such as likely benefits or how the BI application provides data needed to help the group attain its annual goals. Following are some sample messages:

- Early identification of at-risk customers enables us to take proactive steps to retain them.
- Easy access and use of accurate sales history can improve the accuracy of sales forecasting.



- By studying prescription claim payments we can identify opportunities to enroll customers in mail order prescription drug plans to reduce medical claim costs.

Simple things can generate positive emotion and momentum. When the marketing message comes from business peers and management, it carries much more weight than a message from a project manager.

## **Learning to Use the Data without a Technical Degree**

In order to ensure that the data warehouse and the BI application are put to good use, some sort of training must be provided. The type and extent of the training depends on the audience. Two different kinds of education are needed. The most obvious, and most commonly provided type, is teaching people how to use the BI tool and/or application. The other, and more important type, is teaching people about the data and how it should be used.

### ***Learning about the Data***

In order to understand and properly use the results from the BI application, time must be invested in learning about the data. This includes sharing core business definitions and how they can be found in the data. For example, it is important for people to understand that the new definition of a customer is more inclusive than what was used in the past.

Users need to learn what reports are available, what the reports contain, and how the data should be used effectively. This type of education is focused on business processes and how to apply the BI results to those processes.

Business analysts and power users need to learn more about what data is in the data warehouse, and become familiar with the Business Dimensional Model. This helps users understand what questions can be supported, and helps when forming new queries and reports.

**NOTE** It is critical to provide education about any business process changes linked with the BI application. Be sure to include the business rationale for the change.

### ***Learning about the BI Tool/Application***

The second type of education focuses on the mechanics of using the BI application itself. Casual users may need minimal training to use a well-designed web interface to access a set of structured reports. Mainstream users and business analysts need to learn the mechanics to set parameters and interact with the results. Some business analysts and power users will need additional technical training in order to leverage the capabilities of the BI tool. This may include attending training offered directly by the tool vendor to learn advanced features.



**TIP** Too often, training is viewed as a mechanism to learn which buttons to push. Learning how to rotate a report or change the format of a graph may be helpful, but it is much more critical for business users to learn what data is available and how it can be applied to their work.

The project team can put together the overall training plan to accommodate these different levels of data and tool education. Courses should be offered periodically to enable people to get more training as they are ready to dig deeper in using the environment. This can also serve as a refresher for anyone who has not used the data warehouse for a while and is ready to get started again.

## Ensuring That the Right Help Is Available

Beyond providing basic education, it is important to have resources available to assist with use of the BI application. While most organizations have a technical support line, that staff will not have the skills necessary to provide support for the data warehouse. Some typical technical support will be needed to ensure that the servers are up and running and that the database is online. More of the support requests stem from confusion about results on a report or how to get a certain type of data. The data warehouse help desk must be staffed with people who have experience using the BI tool and application. They must also have a good understanding of the reports and the underlying data. Often, they will be required to assist others in understanding what a report represents and how to interpret its results. This is far more than just ensuring that the website is available.

Many users will simply tap into the BI application to pull a report because that is all they need. A help desk format is sufficient for this type of interaction. However, it is important to provide enough support to those who will be using the data warehouse more extensively. For these users, it is beneficial to provide more proactive support. The work is the same, regardless of whether support is provided by IT or business personnel. If the support team is within the business unit, then they likely share work space with the other business users. If support is provided by IT, then they need to go into the business area to see what is going on. This may mean having “office hours” during which a representative is the floor (e.g., every Tuesday and Thursday afternoon) for questions and support. Alternatively, it may mean that the data warehouse/BI support team has a cubical assigned to them, and someone sits right in the business area every day.

The benefit of such close proximity is that the support team can jump in to help whenever the need arises. There is usually an intensity in the area when requests come in to provide reports or more detail. The intensity rises to a panic when the request comes from the senior executives or the board of directors. In order to fulfill these requests, people often gravitate to what they know best. When a data warehouse is new, the staff is more comfortable with the old system. It may not be pretty, but they know how to get the job done.

When a member of the data warehouse support team is embedded within the business group, this sense of intensity and excitement represents an opportunity to help. The support representative can step forward to learn more about what is needed. If the data is available through the data warehouse, then the support person can offer to pull it for the business team. When this occurs, the business analyst should observe to learn how this is done. Due to the pressure, it is OK for the support person to do the driving. The next time, the business analyst can drive, and the support person can provide step-by-step help.

Sometimes, business users are geographically dispersed across the country or the world. Other support mechanisms must be explored to help ensure success. This may mean that education and support are delivered via the web, which might include an online library of frequently asked questions (FAQs) and short video clips showing how to use certain functions of the BI application. This also has implications for scheduling the support staff to ensure that skilled resources are available to help anywhere (e.g., the offices located in Asia). These issues are similar to what any other system implementation faces with a global user base. Tap into any communication and support mechanisms that are already in place to address these global challenges.

The more times that the data warehouse can be used to support daily (as well as last-minute) requests, the more the business gains confidence. Over time, as users gain confidence in both the data and their own skills, they will rely more and more on the data warehouse/BI application to do their work. There must be a management commitment from both IT and the business to ensure that the right people are allocated to provide this level of support.

**NOTE** You have made a significant investment in time, personnel, and dollars. Make the investment worthwhile by putting in place the education and support needed to ensure success.

---

## In Real Life

The two companies that we have been following continue to make classic mistakes with their data warehouse projects. Let's take a look at how the data deployment has been going.

### BI at Giant Company

The project team's direction was set: Recreate all of the existing reports. There are hundreds of reports. The highest-priority reports have been identified, but there are still over 100 reports to recreate. The project team was able to estimate an average development effort by report. Finish the reports and the project will be done. This approach is endorsed by the representatives from

the business because the reports are currently being used. The only way to get rid of the old reporting system is to replace those reports.

As the team began to look at the reports in more detail, they realized that the original estimates were too low. It was going to take a lot more work to rebuild the existing reports. To make matters worse, the business analysts identified several additional reports per week. The processing and manipulation of data in many of the reports was extremely complex and the true business logic was not clear. After floundering for a few weeks and understanding the true effort to convert all of the reports, the team decided they needed to step back and regroup.

The organization is dependent on the current reports because this is the only way that the data can be accessed currently. Looking more closely at how these reports are used, it became clear that the reports themselves were just the tip of the iceberg. Rarely was a report used as is, to help the business. Instead, the data from the report is loaded into a spreadsheet, often combined with data from other reports. This new set of data was the beginning of further manipulation that was performed through a series of spreadsheets. Finally, another twist is that the current reports are static.

The project team developed a two-pronged approach. This was reviewed and approved by the business champion. The first plan of attack is to pare down the reports to be converted. To do this, the reports are grouped to identify patterns. It turns out that there are multiple versions of a single “report,” each creating a slightly different view.

For example, there is a sales top line report for the entire company, and then four reports (that look the same) for each of the sales regions. From there, dozens of other reports provide data for each district and unit. Using the flexible BI tool, a single report template can be defined with a parameter for which part of the sales organization should be included on the report. This single template can replace the dozens of sales top line reports in the old system.

To take things a step further, each of the report templates is designed to be interactive. This enables users to drill down into more detail as needed. The new focused effort to replace existing reports was able to get started while some of the team members focused on the second prong of the attack: better defining the BI application itself. Using the steps described in this chapter, the team was able to identify new reports and analyses to bring additional value to the results of the data warehouse. While the initial delivery of BI is via reports, this is really just the starting point. The business plans to use these reports to spark more in-depth analysis.

## **Agile, Inc. Dives into BI**

Limited business requirements were gathered on the data warehouse project at Agile, Inc. The primary requirement was to just “give me the data.” What

data? All of it. The approach is to load the data, set up a business intelligence tool, and let the users do what they want. No structured applications or navigation were built.

Once the data was loaded and released, there was little actual use. Needless to say, with the investment already spent, this is disappointing. Rather than just scrap the whole thing, the project team investigated the situation. The feedback was loud and clear: The data warehouse is too hard to use. To overcome the difficulties and resistance, more training was set up to teach people the technology, but it was still too difficult. Everyone began to think that the wrong BI tool was selected.

Luckily, the project team decided to look to the data warehousing best practices for ideas before forging ahead to find a new tool. The team learned that a suite of “starter” reports is recommended. Indeed, the business users did not want to build their own reports; they just wanted the results delivered and ready for them to study and analyze. The team also developed training about the data and how to use the starter reports. This was well received and the business group began using the data warehouse more frequently.

---

## Summary

---

There is more to using the data warehouse than simply installing a business intelligence tool. This chapter provided an introduction to what is possible with BI. A BI application has multiple layers, starting with the data content and providing a navigation path to get to the information you need. The layers also include the many different options for how data is presented, including tabular reports, charts, or a dashboard of indicators. You also learned that there are many different options for delivering these results.

BI development includes familiar systems development steps, including design, development, and testing, but this work is usually done with graphical interfaces, making it much easier than programming. In fact, the BI development work is often done by BI developers who come from the business community. To help the business community use and get value from the BI application, appropriate education and support must be provided. This helps the data warehouse to become an integral part of how you do business.

The project is nearing completion, but there are some additional areas that can help ensure success. The next chapter discusses what needs to be done to maintain the data warehouse. Factors that can contribute to impeding use of the data warehouse are shared along with recommendations for re-building momentum.

**PART**

**V**

# Next Steps—Expanding on Success

---

## In This Part

---

**Chapter 12:** Managing the Production Data Warehouse

**Chapter 13:** Achieving Long-Term Success



# Managing the Production Data Warehouse

The effort and energy needed to sustain a data warehouse does not end when the project tasks are all complete. This chapter shares what is needed to release the data warehouse for use and what steps are necessary to formally wrap up the project. It also looks forward, describing what can be done to encourage adoption of the data warehouse. Additional effort is also required to keep the data warehouse running. Finally, the chapter explores what can cause a data warehouse to fall short of expectations, and offers practical steps to get things back on track.

## Finishing the Project

---

The project has been underway for many weeks or possibly months. The requirements have been gathered and the overall solution designed. The ETL system was built and the data has been loaded and validated. A BI application has been built to facilitate use of the data warehouse. Only a few more things are needed to wrap up the project.

## Recapping the BI Application Launch

Ideas for launching the BI application were introduced in Chapter 11. These launch activities are some of the last tasks on the project. Highlights of the launch activities include the following:

- Marketing and communication to let potential users of the data warehouse know that it is ready for use. These efforts can also help pique interest by demonstrating how the data warehouse can support users with their work.

- Two kinds of education are needed:
  - Education about data content, meaning, and how to apply it to daily work is needed so that the data can be understood and used effectively.
  - Education about the mechanics of using the BI application is needed to help users get started. This includes training about how to navigate to find what you are looking for and how to access the data you need. Education should also be provided to meet more advanced user needs to help them change reports and leverage more advanced features of the BI tool for more sophisticated analyses.
- Support is necessary to ensure that any problems are identified and addressed in a timely manner. Support is also needed to help people understand what is available and how to use the BI application.

These are important activities that help the business users to learn what is possible with the data warehouse. These are often some of the last tasks on the project plan. Before the project is considered complete, take the time to review how things went.

## **Post-Implementation Review**

Once all of the other project tasks have been completed, the project team should meet again to reflect on the project. Don't wait months to get everyone together; meet soon so that the project experiences are still fresh. Conduct a systematic review of each of the different areas of the project; basically, walk through the high-level tasks of the project plan and other key topics, making sure to touch on the following:

- Requirements gathering
- Data modeling
- Data governance, ownership, and quality
- ETL system design and development
- BI application design and development
- End-user education
- End-user support
- Education for the project team members
- Political and/or organizational assistance and hurdles
- Process and accuracy of estimating effort and setting the project schedule



For each topic discussed, try to identify what happened, making sure to mention what worked well. For problem areas, discuss potential causes, whether it preventable, and what should be done differently the next time.

Due to the sensitivity of some areas, such as personnel, financials, or other politically charged topics, it may also be useful to conduct a separate project review with managers.

When the post-implementation review findings and recommendations are documented, the rest of the organization can also benefit from this team's experiences. Take the time to capture key findings, including recommendations for future projects.

### ***Looking Back—Did you Accomplish Your Objectives?***

The post-implementation project review is focused on the mechanics of running a data warehouse project. A project team may have completed every task on the plan, but that does not necessarily mean that the project is a success. Now is the time to go back to the initial project documents, as discussed in Chapter 5. The project charter should include the objectives. Be sure to also look at any project change control documents that may have modified the original objectives. Have these objectives been met? Were any additional project goals identified through the requirements gathering process? Has the data warehouse been able to support the attainment of those goals?

It may be too soon, right after the project ends, to determine whether the goals for the data warehouse have been met. It takes some time for the solution to settle in and for business users to harness the potential to realize value. While some organizations may begin to see immediate results, it is more typical to see the benefits after several months. While there are indeed some multi-million dollar success stories, success is more often attained by adding up the results of many smaller decisions.

## **Adopting the Solution**

---

The most successful data warehouses are those that are used as part of the daily business. Pulling up the BI application is not something that has to be done because the boss says you have to; it simply becomes how work gets done. It takes time for the BI application to become ingrained. People need to grow comfortable with using the BI application. They need to build their confidence that the data is really there and will still be there tomorrow and the next day. As the group gains experience using the data warehouse, real change can happen. In order to determine how usage patterns grow and change over time, use of the data warehouse must be tracked.

## Tracking Data Warehouse Use

The true measurement of a successful data warehouse is the business value. This may have been articulated when the overall project was reviewed, as discussed earlier in this chapter. Taking it further, Chapter 13 provides more suggestions for finding and tracking business value on an ongoing basis. At this point, it is also helpful to simply understand general data warehouse usage. This is needed to help you do the following:

- Monitor performance and plan for increases in capacity.
- Understand which business groups are using the DW.
- Identify which business groups are not using the DW.
- Identify groups or individuals who would benefit from additional education.

There are some simple ways to determine to what extent the data warehouse is being used:

- Track the number of different people who signed on to the BI application.
- Track the number of queries run against the data warehouse.
- Determine the average number of queries by type of user. Depending upon the how security is implemented, this may also be determined by individual users too.
- Count the number of new reports published for general use.
- Count the total number of reports/templates created and stored by users for their own use.
- Track the number of problems reported, and how many were resolved.
- Track how often the data is loaded and ready for use on time.

These statistics can be loaded into a dimensional structure that can be used to analyze fluctuations in use over time. While these measures can help you gain an overall impression of data usage, they do not provide any insight into what impact the data warehouse has had on the business. They can help the data warehouse team target groups and actions that increase the number of people using the data warehouse.

## Getting the Rest of the Business Community on Board

Now that the data warehouse project is complete, the business users' managers need to help their staff members begin to use the data warehouse. These business managers can get support from the data warehouse manager and

the data warehouse support team, but the primary emphasis must be from business management.

Several things can be done to help encourage use of the data warehouse, including the following:

- Communicate when and how the data warehouse was used to address specific business situations that have come up. Highlight situations where the use went beyond creating a new report to performing an analysis that yielded new insight.
- Openly share situations in which executives and senior management have tapped into and benefited from the results of the data warehouse. If an executive dashboard has been built, then make the effort to use it yourself. This sends a clear message that it is important.
- Reward those who are using the data warehouse environment for their work. Take it a step further by providing incentives to those who actively participate in ways that make it better. People can be rewarded for identifying enhancements, developing business processes that utilize the data warehouse, or improving data quality.

In addition to the business support, the data warehouse manager can also sponsor a variety of activities to encourage use of the data warehouse:

- Create interest groups that communicate regularly and meet periodically. These can be business users who want to create maps from the output of their analyses, people who want to learn more about the BI tool, or a group of IT developers who want to swap techniques for dealing with changes in the source systems. These can be started using a social network such as Facebook, and then the groups can meet in person.
- Offer “Lunch and Learn” sessions to offer informal education on specific topics. Interested staff can bring their lunch to a conference room to learn about new reports that are available, how to use specific parts of the data, how to interact with reports, and how to perform more advanced analyses. Offer a range of sessions and topics, from introductory through advanced, including some with a strong business focus and others that provide in-depth technical content. Once this format is established, the participants can determine what topics are needed.
- Create short articles and notices that can be placed on the organization’s intranet and company or departmental newsletters. This is an extension of the marketing that is done to launch the data warehouse. Additional marketing efforts are needed to let others in the organization know how the data warehouse is being used.

Encouraging or requiring use of the data warehouse/BI application must come from business management. No matter how desperately the IT people

want others to use and appreciate the data warehouse, the push for change must be from within the business community. This emphasis must go beyond the lure of easier and faster reporting, helping potential business users to see how the data warehouse can help them do their jobs better.

## **Business Process Change**

A great deal of attention is paid to the data side of data warehousing. Even gathering business requirements is often focused on understanding what data would help run the business better. However, there is another aspect of data warehousing that is glossed over: A data warehouse can help the business change how things are done. Building a data warehouse does not require business process changes, but successfully using information to run the business can drive the need to change. The data warehouse can help identify where there is success or failure. It can help drill down into the details to understand the underlying cause of problems. The biggest impact will be realized as you adapt how you conduct business based upon what is learned from the data.

Successful projects look for opportunities to incorporate business process changes. Improving how things are done is much more effective than automating exactly what is done today. Several different types of business process changes can result from a data warehouse project. The possibilities are numerous, so only a few are included here to generate ideas.

### ***Changing How Data Is Used***

One of the most obvious areas of change associated with a data warehouse is how data is reported, analyzed, and distributed throughout the organization. Simply arming the existing analytical staff with a data warehouse will certainly improve their ability to use data and respond to the business, but wider benefits can be realized by pushing access to the results of the data warehouse out into many parts of the organization. That way, everyone can see what is needed to help them set daily priorities, determine what is working and what is not, and compare their own performance to their peers. These changes are not simple, but they can result in large benefits to the entire organization.

### ***Streamlining Business Processes***

The data warehouse can also serve as a catalyst to change how other business processes work. The availability of more data to identify patterns and trends provides an opportunity to take a closer look at how things work. For example, consider an organization for which ten steps are currently required to implement a new set of terms on supplier contracts. Three of these steps

involve getting the required authorization to implement these terms. These three steps can't be changed. However, the remaining four steps involve gathering supporting detail for the appropriate people so that they feel comfortable signing off. Without a common source of data for reporting and analysis, each group will pull data its own way to support their own manager's needs. These reports are often not what is needed by the next manager in the chain. New reports are generated for each authorization step.

By having all of the data to support the entire authorization process in the data warehouse, this entire set of steps should be reviewed. A single shared set of reports that lay out the business case for the change in terms could be created, and then made available to each person giving authorization. If any managers have a question, they would be able to simply drill into more detail to understand the rationale behind the request. This type of process change is extremely difficult without a data warehouse in place to provide a common, clean source of data that can be used by each different group.

### ***Encouraging Change***

Management must make a commitment to encourage and seek true business change. The data warehouse can serve as a catalyst, but change must come from the business group itself.

The data warehouse is a tool to be used by the business community. This requires reviewing business processes and assessing their effectiveness. Visionaries and thought leaders in the business group are often the source for this true change. As they become more adept at using the data warehouse and the BI application, ideas will flow about how things can be done differently. The business community, especially managers, have the responsibility to push beyond creating reports, making the data warehouse an integral part of business process flows. This involves developing new business processes and any associated BI reports to support the new processes. It also involves getting people accustomed to the new ways of doing things.

Data warehouses that have had the largest impact over time are used by the business community to constantly reevaluate and fine-tune how business is done.

## **The Production Data Warehouse**

---

Once a data warehouse moves into production, two distinct types of work continue. First, there are activities and tasks that are geared toward keeping the current data warehouse up and running. This section focuses on the ongoing activities needed to maintain the production data warehouse. The second type of work is to enhance or expand the data warehouse. This is usually handled

as new projects. Chapter 13 includes a discussion of defining the work for the expansion and growth of the data warehouse.

## **Staffing Production Activities**

Once the data warehouse project is complete, that does not mean that all of the resources are now freed up for new projects. As with any other system, a team of people needs to be in place to provide support and maintain the data warehouse. The data warehouse manager, as described in Chapter 4, is the focal point for maintaining the production data warehouse. In addition to the DW manager, the following types of resources (which are outlined in Chapter 4) are needed (the activities for each of these are discussed in the following sections):

- Staff to provide technical support for the environment
- ETL development staff to maintain the ETL system
- BI developers and power users to maintain and develop new BI reports, analyses, and applications
- Helpdesk support
- Ongoing education

There are many different ways that these resources could be organized. For the purposes of this text, these are collectively called the *data warehouse support team*. They are also called *centers of excellence* for data warehousing or business intelligence. Regardless of what you call it, these functions are needed. Considerations for larger enterprises are discussed in Chapter 13.

## **Maintaining the Environment**

Several primary components must be maintained over time: the technical environment, the ETL system, and the BI application. The BI application typically gets attention through regular use, but often the technical platform does not get the appropriate level of care and the data loads are launched with no additional attention. These are each worthy of a closer look.

### ***Keeping Up with Technology***

It seems elementary to discuss basic maintenance, but too many organizations allow the technical environment to lag behind. While the work to keep the technology current is done by IT, the funding is often tied to business support of the data warehouse. Managers from both IT and the business need to ensure that the appropriate funding and personnel resources are allocated in order to do this work. While it is up to IT to do the work described next, it is helpful

for business managers to understand what is involved in maintaining the technical environment. This is intended to provide a high-level view of what needs to happen.

Once the data warehouse is built, it is important to keep all of its different components current, including any patches and new maintenance releases. Some organizations work actively with their vendors to get corrections to bugs and to push for enhancements. These organizations often partner with their vendors to beta test their new product releases or to install these new releases soon after they become available. Some organizations just like to push the envelope. Others find themselves forging ahead because they need the new product features to help them address their own business objectives.

Not all organizations are up to the challenge of getting involved with early releases of products. This is fine. You can monitor the results of other organizations. This can be accomplished by networking with other organizations that are using the same technologies. You can also participate in local or national user groups for that product. These groups typically work closely with the vendors (and may even be sponsored by them) and are well informed about the availability of new releases and any current issues. For example, if many installation problems are occurring, you can wait a little longer to ensure that the new product release is stable. Once the release is fairly stable, go ahead and install it.

### **The Importance of Testing**

Keep in mind that you can minimize risk by setting up a good test environment. In general, it is always important to have a good test environment (as recommended in Chapter 9), and even more so when adopting new technology early. This enables your organization to ensure that the new software will not break anything else in your environment. If it does, the problem can be studied and resolved without affecting your production data warehouse environment. This also enables your staff to learn to use the new features of this release. You should expect to find problems with the products, but you can play a critical role to ensure that they are fixed and that the new features work properly.

The test environment is also used by the ETL and BI developers for their work. If there is a lot of development activity, then consider setting up an isolated test environment dedicated to testing software releases. Depending upon the level of other development efforts and the frequency of software releases, this additional test environment may need to be permanent.

### **Falling Behind**

The worst-case scenario is when an organization has not kept up with releases at all. Over a number of years, a significant gap can develop between the current product and the version that you are running. Eventually, support for older



releases is dropped. That means you will not be able to get help if something stops working. In addition, as bugs and problems are identified, the solution is usually offered in a newer version of the product. If you are unwilling to move forward, you have to live with these bugs. Most organizations begin to panic when support will be dropped. The path to upgrade is then under pressure and is based upon the vendor support timeline, rather than a schedule that is convenient or affordable for your organization.

When support is dropped, it is usually after several releases, so getting your environment up to current standards often means jumping ahead two or three releases. This can be a big change; and there can be major differences in supported configurations, how to install the software, and how to use it. These changes are often less dramatic when taken one release at a time. It is usually less expensive in the long run to incur upgrade costs as each release becomes available. Costs associated with upgrades may include software, retraining of IT and business staff, and conversion of existing programs or reports. There may also be a delay of new development when resources are focused on upgrading the software. The costs and time required is typically less when moving up one release at a time compared to jumping several releases at once.

It is worth the investment to keep your data warehouse up to date with your technology. Once everything is current, make a commitment to keep it that way. This often means a steady stream of new product releases (some for maintenance and some for new features). Installation of these new product releases can often be bundled with other maintenance and enhancements that are made to your data warehouse environment.

### ***Monitoring Performance and Capacity Planning***

Once the data warehouse is launched, it takes time for its use to become a regular part of the day-to-day business. Data volumes will grow as you keep loading data. The number of users will increase as more people get comfortable with the environment. The number of queries and the complexity of queries will also grow over time. All of these factors can lead to performance or capacity issues.

The performance of the data warehouse must be monitored in two separate areas. First, the performance of the ETL system must be tracked and monitored. There is often a processing window of time that must be met to deliver the data in a timely manner. Often, this window is specified as part of a series of service objectives that are defined and agreed to by business and IT management. The performance of the ETL system is not usually observed by the users of the data warehouse unless the data is not ready on time. IT organizations are typically already equipped with the tools necessary to monitor performance of the ETL system.



The second type of performance that must be monitored is that of the BI application and any other data access. Data access performance is highly visible to the user community. When queries and reports do not run in a timely manner, IT is usually informed immediately. While IT may already have some tools in place to monitor query performance, explore what tools may be available from the BI tool vendor, possibly for an additional cost.

Another thing that must be monitored is utilization of resources. As the use of resources nears capacity, this can result in performance problems or the inability to meet the system and/or user needs. There are many different places where capacity must be monitored, including the ETL processing resources, the database itself (data storage and query processing), application servers, web servers, and sometimes simply the number of licenses for each different product. Someone must be responsible for keeping an eye on each of these different areas.

Business intelligence tools can be used to help you monitor your entire data warehouse environment. This requires that you capture statistics about the data warehouse itself. The ETL system can capture the number of rows processed, the size of the data processed, the number (and type) of warnings and errors encountered, and the time required to complete each step. Capture and store database query statistics, including the number of users who log in and the number of queries that are run. See what your BI tool can capture to help you understand which reports are being used and which are not.

Using the statistics just described, you can create trend reports, and compare actual results to targets. For example, you can compare total ETL processing time to the load window that you have. If the time used is steadily creeping toward the maximum time available, then you can investigate the situation to determine what is causing the increased time. Is more data flowing through the system? Are you waiting for other systems to complete before you can begin processing? You can perform an analysis to both figure out where to look and determine what needs to be done to ensure completion in the allotted timeframe.

In addition to comparing actual performance to targets or goals, you can also track each of these measurements over time. Is there an increase or decrease in the number of queries run against the data warehouse? If so, you can investigate to figure out why. Is there an increase in the number of rows being loaded each week? If so, you may need to adjust the schedule regarding the purchase of additional disk space.

Leverage the capabilities of BI to help manage the business of building and maintaining the data warehouse. As you have seen, there is more work to be done than merely keeping the technology up to date with sufficient capacity to maintain acceptable service levels. There is additional work to maintain the data warehouse itself.

## Maintaining the Data Warehouse

While it is important to keep your technology current, it is just as important to ensure that the data warehouse itself is maintained. To simply keep the environment running, the ETL system and the BI application require attention.

### *Maintaining the ETL System*

The data loads must continue to run as planned, and each load must have a quality assurance step to monitor the data's accuracy. Most successful data warehouses generate a need for more data, more reference data, and the inclusion of more sources. Adding new data sources is usually done by creating a new project and following the principles discussed throughout this book.

More time and energy are needed to maintain the data that you have already started with. Once the ETL system is developed, it should not be moved into production to simply run for the next five years. Changes and tweaks will always be made to the underlying source systems. This can affect what flows to the ETL system. It may be small changes that have minimal impact, but sometimes significant changes are required to the ETL processing. These changes are often the result of modifications to the underlying source systems. Examples of source system changes that impact the ETL system include the following:

- Changes to how data elements are used. For example, if there is a change to how pricing is performed, the data may be stored in all the same fields as before, but the contents may now contain pricing tiers, rather than an actual price.
- Modifications to key identifiers, such as changing all of the customer IDs
- The addition of new functionality that introduces new data. This is usually done via a major enhancement to a system to add screens for additional input or the development of new modules.
- Expansion of reference data to enable new and improved values—for example, adding more transaction types to be more specific about what business transaction occurred.

Changes to the underlying systems are made both to correct problems and to provide support for new features that are needed by the business. Because the data warehouse often deals with multiple source systems, a steady stream of changes is usually occurring to keep the existing ETL system working.

Communication with each of the source system development teams can help ensure that the data warehouse is aware of the changes that are coming. Even the most diligent and cooperative source system teams can forget about the data warehouse at times. You need to be proactive with your communication so that everybody has a good understanding of what they are working

on. Regular communication helps ensure that the source system application development team continues to consider the impact of their own enhancements and modifications on the data warehouse. While they do not need to limit what they do because of potential impact on the data warehouse, they do need to ensure that the data warehouse team is prepared to properly process any new data when it begins to flow from the source system. This can be accomplished by having a representative from the data warehouse attend change review meetings for the source systems or review the change/enhancement requests that have been approved. This provides insight into what the source system support team will be working on and enables the data warehouse team to begin exploring the potential impact on the data warehouse itself.

### ***Maintaining the BI Application***

The BI application provides a starting point for business users. As these initial reports and analyses are used, new ideas will be generated. Often these enhancement requests are formally submitted, but usually new reports are created directly by business users.

The beauty of a BI environment is that experienced users can indeed create their own reports and analyses. This can be done from scratch or through a series of modifications to existing reports, enabling them to morph into a completely new report. It does not require anyone else to step in and do the work for them.

However, such reports may not follow the standards developed for BI development and they may not run efficiently. They may meet the business need, however, and if a report is useful enough, it may be shared with colleagues for their use too. Nonetheless, rather than have a poorly performing report distributed throughout the business community, it is better for the author of the report to submit it to a central BI support team for a formal release. The BI support team may be a group of highly experienced business analysts within the business community. This support team can ensure that proper formatting and other standards are applied. Quality checks can be done to ensure the accuracy of calculations. Finally, the efficiency of the report can also be reviewed. Usually, in a short time frame the new report can be released into the BI environment as a new analysis in the library of reports. If the new analysis is valuable enough, it may even be included on a dashboard or other BI portal for general access. As more people gain experience, they can be taught how to prepare these reports or analyses for distribution.

The important thing is to provide a mechanism for new reports and analyses to be distributed to the rest of the business community. In addition, keep in mind that not every report or analysis must go through this process, just those that are to be included as part of the official BI application. Many reports will be created for personal use that do not need to be shared.

## Tracking Questions and Problems

When the data warehouse and BI solution is released for use, there must be a mechanism in place to collect and track questions, issues, and problems. This same mechanism may also be used to submit and track enhancement requests. More information about handling enhancement requests can be found in Chapter 13. The organization may already have a system in place to track problems. This is often used by the help desk. This application may be able to log and help track data warehouse and BI issues too. The data warehouse support staff needs access to and training for using the problem-tracking application.

While a full-fledged problem-tracking system can provide many features, a simple spreadsheet can also be used to log and track issues. The most important point is that these be documented in a single location so that all issues relating to the data warehouse environment are identified and monitored.

All feedback should be logged. This helps to understand how the support team is spending their time. Feedback could be a question regarding how to use the data warehouse or what the data means. Problems that are reported could include the data warehouse not working at all, reporting an error in the data, or a malfunction in getting the BI application to run. Sometimes feedback is a request for something, such as a new business measure, another report, or a new data element. The actions resulting from these requests will vary, but a common set of information should be tracked for all feedback, including the following:

- A unique tracking number
- Name and contact information for the person reporting the problem or requesting the enhancement
- The group for which this person works (sales, marketing, IT)
- Date the problem is identified
- The request type (bug, enhancement request)
- A brief description of the problem
- Severity of the problem (e.g., the system is down, important but can get by in the meantime, nice to have, critical enhancement)
- Part of the data warehouse/BI environment is involved (data, BI calculation, report, query performance, ETL system rules)—This may not be known at first, but when it is figured out, note it.
- The status of this issue (e.g., open, resolved, solution scheduled for release in the second quarter of next year)
- Date the problem is resolved

- Details about resolution of the problem—The question may have been answered or a technical glitch was fixed. Sometimes, resolution may be the generation of a bug fix request or enhancement request in order to start the process of addressing this issue.

Once the feedback has been logged, the support team should work toward a resolution. In many cases, questions can be immediately answered and problems can be resolved. Sometimes additional resources are needed to help identify the source of the problem and determine what to do about it. In some cases, IT and/or business management should be informed of an open problem. For example, if the entire data warehouse database is down, all IT and business managers involved should be informed.

A series of guidelines and procedures should be developed to lay out problem escalation paths. These are typically defined by the severity of the problem and how much time has passed since the problem was reported and what type of service levels have been agreed to. These guidelines may also include steps to take when service levels for the data warehouse have been compromised. This is not unique to a data warehouse. Most organizations already have guidelines and procedures in place for other systems that can serve as a starting point and can be adapted for the data warehouse.

Guidelines and procedures should include instruction for ongoing communication regarding open problems or issues. The problem-tracking system may have an inquiry feature whereby anyone can see the status of reported problems. However, if a problem has been reported but is still unresolved, then proactive communication is appropriate. An e-mail should be sent to the requestor so that he or she knows that the support team is aware that there is still an open problem. When possible, it is important to be specific about the status of open problems. If someone is actively working on the problem but hasn't figured it out yet, say so. If the problem won't even be looked at until the end of next week, then communicate that too. Once a decision has been made, the plan to address a problem or request should be communicated. Sometimes a correction is made and put into production overnight.

More often, however, the solution to a problem will be bundled with other features to be released at some point in the future. The timing and rationale is often the result of joint IT and business prioritization, as discussed in the next chapter. If the work to address the request is significant enough, it may be put into the pool of other requests that will go through a larger planning process, as discussed in the section "Planning for Expansion and Growth" in Chapter 13.

As decisions are made, it is important to ensure completion of the problem-solution loop. The original requestor must be informed. If he or she is not satisfied with the decision, then any outstanding concerns can be raised

with management and/or the data warehouse decision makers, such as the executive steering committee or the DW business support group.

**NOTE** If no one is asking questions or requesting enhancements, then it is unlikely that anyone is using the data warehouse! Usually a few questions will arise even if the data warehouse is simply a new place to get the same reports. Even more questions will be asked if the data warehouse is being used to think outside the box to better leverage data.

The problem-tracking system can also be used to help quantify how the data warehouse team is spending its time. It can be useful to see over time the number of items reported, the number of items resolved, and the average time to resolve issues. These can also be sorted by other attributes that are collected by the problem-tracking system, including the department or the severity of the problem.

## Fixing Bugs

There will always be some problems that need to be addressed. Identification of problems is accomplished using the tracking mechanism described in the previous section. Sometimes these problems can be addressed immediately. Sometimes the resolution of a problem requires changes to the ETL system or BI application. As with any other system, these modification need to follow an organized progression of steps, including development of the solution, testing, and then promoting the revised program or process into the production environment.

Many bug fixes can be handled by the data warehouse support team without any additional direction or prioritization. Usually, there is a threshold number of hours of effort that the DW support team is empowered to handle directly. Modifications that require more effort hours should be included in the planning process for expansion and growth, as long as the problem is not severe. If a bug fix is needed to keep the data warehouse running, then this must be addressed as soon as possible, even if the solution requires a significant effort.

## When the Data Warehouse Falls Short

---

As mentioned previously, many organizations have invested in data warehouse environment but have never realized any value or have seen a slowdown in the value. Many other data warehouses have stalled. This can happen when launching it initially, or after five years. Sometimes it is difficult to remember why it was built in the first place. Regardless of how you got here, you are

not alone. Don't be afraid to admit that everything is not going as planned. Once you understand why things have stalled, you can take specific steps to get started again.

## Common Causes for a Stalled Warehouse

Over the years, patterns have emerged which have resulted in industry-best practices. Sometimes the patterns that emerge help identify what can go wrong. There are several common causes for a data warehouse that has stalled, including the following:

- **Lack of business focus:** The data warehouse may have been built without direct input and partnering with the business. This is the root cause of many data warehouses that are not used or have limited adoption in the business community. In this case, the data warehouse usually represents a collection of well-organized data. It often doesn't have any connection to solving any business issues or supporting business units.

*Recommendation:* Work with representatives from the business audience to understand their requirements. Compare the data that you already have with these requirements. Perhaps minor enhancements to the data and the development of a few reports are needed. Adding training may breathe new life into the data warehouse. Unfortunately, sometimes the data as loaded cannot really help at all. If this is true, then study what is needed. Take the techniques and methods learned and apply them to starting over with a strong business focus this time.

- **Unstable technical platform:** If the products that have been selected are not installed and configured properly, then they will not work as advertised. Sometimes products don't work at all. In other cases, the products have sporadic problems. Either way, if the technology is not reliable, then no one will depend on the data warehouse to help them run their part of the business.

*Recommendation:* Stabilize the technology. Ensure that the development team and the business users have a sound, reliable environment to use. Take the time to pinpoint any problems. Upgrade the hardware, fix the configuration, and ensure that you are running the most current releases of the software. Approach the vendors with an open attitude. Explore the recommended configurations, rather than demand making your current setup work. Vendors are often hesitant to say that what you have will not work. They will try to develop many work-arounds to meet your demands. Make the environment the best that you can to exploit the products you have selected.



- **Lack of reports and/or BI applications:** The notion that “if you build it, they will come” has been proven wrong, again and again, by many different organizations. Just loading data into a database is not enough to launch a successful data warehouse—even if the data structures are perfect! The real value is obtained when people use the data. Most business people have no desire to crawl through the data and craft new reports from scratch. They want to look at snapshots of their business but have the capability to look into more detail if necessary. This type of use is best supported through the design and development of a series of report templates in combination with structured navigation.

*Recommendation:* Work with key business analysts to identify useful report templates. Invest the time to build a starter set of reports, and invest in the development of the interface to help the business users navigate through all that is available to find what is most useful for them. This may require the development of a BI portal or creation of a dashboard. Chapter 11 describes these in more detail.

- **Insufficient data:** Not having the data necessary to support the business is one of the biggest challenges to success. Insufficient data takes several different forms. There may be some important data that is missing. The data that is loaded may not be at a low enough level of detail to be useful. Finally, there may be serious data quality problems. Data-related issues do not go away without specific actions and they can stall a data warehouse for years.

*Recommendation:* It is critical to get to the bottom of any real or perceived data quality problems. Once they are identified and fixed, additional education must be provided to help the business users understand what the data represents and why it is accurate, even if it is different from what they are used to. When data is missing, either new data or more detail, additional projects must be scoped out to expand the ETL system to load the required data. Keep in mind that once the data is loaded, BI applications and reports must also be developed to push the new data out into the business community.

- **Personnel challenges:** Making sure the right people from IT and the business community are involved in the data warehouse can be difficult. Sometimes people are assigned to a data warehouse project because they had time available. Others are assigned because they bring a special skill or have expressed interest in data warehousing. In other cases, people simply do not have the skills needed to work effectively on the data warehouse. Even with the proper education and given time to adapt, some people just fight DW industry best practices and cause trouble



every step of the way. This can be true for project team members as well as business analysts being asked to use the data warehouse.

*Recommendation:* There are two different areas in which personnel issues can impede progress. First, the project team itself can face challenges. In defense of these opinionated employees, they are simply fighting for what they believe is right. However, if there is a major difference between a personal opinion and the organization's strategy that continues to be a problem, then that person should be reassigned to a different project. Often these are people with a lot of experience and skills that are valuable to the organization. There is simply a mismatch between their skills and what is needed for the data warehouse. In some cases, it may be better to leave a position empty than to fill it with someone who is not able to contribute effectively.

Second, the business community may face personnel challenges when implementing a data warehouse. Many people learn to do what is necessary to complete the set of tasks to produce reports, but they don't have time to perform further analysis. Over time, these roles do not have analytical and decision-making responsibilities. The people who fill these roles may not have the skills and background needed to perform more sophisticated analyses. Even with additional education, they may not be able to meet the demands of the data warehouse. These people may serve a vital function to develop BI applications, perform quality assurance tests, or assist with other support roles. Identify someone else in the organization to fill the role of true business analyst.

- **The wrong tools:** In some cases, the products and technologies that were selected years ago may have outlived their usefulness. Perhaps they never were a good fit. Take the time to assess how well your current products are meeting your needs. Acknowledge what is working well and identify any major problem areas. This will help you assess the viability of different technologies. You certainly do not want to disrupt what is already working well.

*Recommendation:* Take a quick survey of the marketplace. The landscape of possible tools changes significantly every few years. If you have a product, such as your ETL tool, that is serving its purpose well, don't spend a lot of time exploring the marketplace. It can be worth taking a look, however, if you have a need that is not being met.

- **Lack of business use:** There may have been sound requirements collected and great work done to prepare and load the data. Even after reports have been created, however, until someone starts to use them, you will not see any value. Sometimes the current way to do business is so ingrained that there is no impetus to begin using the data warehouse. When things are

working adequately, many individuals have little or no incentive to do their work differently.

*Recommendation:* Trying to change how things are done is very difficult and takes time. Communicating that the data warehouse is now available for use may not be enough to get people to try it. There must be a concerted effort on the part of business management to encourage use of the data warehouse. Sometimes providing the business reason to use the warehouse is all that is necessary. Making additional education available can also help. Teach people what the data means and how to use it for their daily work. Rewarding use of the data warehouse can provide the incentive needed to help people get started. Encourage creativity and empower employees to develop new ways to get their work done. While it does take some time for this to happen, it can be extremely beneficial for change to grow from within the organization. That way, people make changes to help them work smarter, not just harder.

Many additional things can go wrong that are not mentioned here. These are just the most common causes for a data warehouse effort to fall short. Typically, it is not a single factor that affects success or failure, but rather a combination of multiple things. Even when a project attempts to do all the right things, success is not guaranteed. The important thing is to acknowledge that there is a problem, try to identify the source of the problem, and then make a conscious effort to change.

## **Jump-Starting a Stalled Data Warehouse**

Just because the data warehouse has fallen short, it does not mean that everything must be scrapped. As discussed earlier, many different factors can contribute to a stalled data warehouse. The real question is what will be done about it. Before you can decide that, you need a realistic assessment of the situation.

### ***Conducting an Assessment***

The causes and recommendations described in the previous section can help you to diagnose some of the most common problems. Clearly, getting a good understanding of the root causes that are impeding your data warehouse process is a good first step. This can be done through an assessment process. Part of this is performing an audit to understand what is in place. In addition, you need to evaluate what is working and what is not. Then, recommendations should be developed to address any problems that have been identified.

An assessment is best done by someone who is not too close to the project. This enables the independent party to be more objective. This person may

be found internally, but external consultants are often needed, especially to develop recommendations to move forward.

Whether internal or external people do the assessment, it is important to listen to these individuals. Take the time to listen to the observations, concerns, and ideas from several different perspectives. Speak with representatives from the project team, the target business audience, IT managers, business managers, and senior management.

The purpose is to assess what they think is working and what is not. Be prepared for a wide range of reactions, including frustration, anger, confusion, hope, and sometimes complete lack of awareness. Understanding how each different group perceives the data warehouse can help you develop useful strategies for moving forward. For example, it serves no purpose to send an e-mail announcing new BI features to a group of users who never even look at the BI application.

Each of these groups can also provide great ideas about what needs to be done to improve the data warehouse environment. Many are feasible, but some may not be possible in the near term. Expectations need to be set regarding what may be possible. These discussions begin to build a communication channel that can be used to gather and clarify requirements and build interest in using improved data warehouse/BI capabilities. Use this channel to loop back and share strategies and plans as they are developed. Keeping everyone informed is a major step forward toward building a more successful data warehouse.

**NOTE** If you are having trouble figuring out what is and isn't working, hire an independent third-party to come in and perform an assessment. This is usually worth the cost. Third parties can make immediate suggestions for improvement and help you identify the major steps required to move forward.

Part of what is learned through an assessment is what is working today. This can help determine what should continue to be used.

### ***Determining What Can Be Salvaged***

Rather than simply toss everything out to start from scratch, you need to assess what components are viable. There are often things that have already been built that should continue to be used. For example, consider a situation in which a third normal form data warehouse was to be built without any dimensional data marts and is not being used. This is not the proper implementation for the top-down architecture. The third normal form data warehouse may be just fine, and building dimensional data marts or views would be a recommended next step.

At the opposite end of the spectrum, perhaps chaos has taken over. There are dozens of non-integrating dimensional data marts all over the organization. Here again, this is not recommended by the bottom-up architecture. This may require a coordinated and disciplined effort to begin to bring these together in a thoughtful and integrated manner. Many of the ETL processes developed for each individual data mart can still be used and enhanced to fit into this new integrated environment. This reuse can leverage IT resources and provide integrated data to the business.

However, there are still times when the effort to fix and maintain the existing components, such as the ETL processes or reports, is greater than starting over. This cost must take into account any effort to enhance existing capabilities and maintain them over time. If this cost is too high, then make sure that you take the lessons learned (about what worked well and what did not) forward. Knowing what can still be utilized, and what cannot, is vital input to creating an action plan.

### ***Developing a Plan to Move On***

To a large extent, getting a data warehouse moving again is much like setting up a new project. It includes work to scope out the effort, and define the goals and what needs to be built. Many organizations are good at this type of work and can assess the current state and craft a project to get things moving. Clearly, specific plans must be developed to correct and/or rebuild the technical components of the data warehouse. This must be done with an eye toward delivering business value, which requires a close partnership between the business and IT. This is the opportunity to do things the right way.

Part of scoping out the work to get the data warehouse moving again is getting funding. This can be challenging because the last investment did not pan out. This is why having a clear understanding of what contributed to the shortfall of the previous data warehouse initiatives is critical. The proposal for a new data warehouse project should spell out each of the primary challenges on previous projects, with an associated action plan to mitigate the risk. This book should provide insight into what works and help you to craft a new project that will be successful.

Because this is not the first project, planning and doing the work is not enough. You must include a series of tasks and activities that are focused on addressing cultural and organizational challenges. At a minimum, this must include sales and marketing efforts for the data warehouse and BI solution(s). Concerted efforts need to focus on how the data warehouse can become an active part of supporting the day-to-day business.

It is much harder the second or third time around to get support and build momentum for a data warehouse. People are tired of participating and not

seeing any benefit. You will likely encounter a great deal of doubt and cynicism about data warehousing, and there is often a pervasive attitude that it will never work and is a waste of time.

To turn things around, get back to basics and the fundamentals for successful data warehousing. Use what you learned from this book. Tap into the knowledge and experience of people who have been involved with the organization's data warehousing efforts over the years. After several years of valiant effort, sometime the best data warehouse resources get so frustrated that they move on to contribute in other ways. Seek them out to get their ideas. Some steps that can have a big impact include the following:

- Get a visible, long-term commitment to data warehousing from senior management.
- Senior management must work with middle managers to set the data warehouse as a priority. Middle managers must then allocate the necessary business resources to work on data warehouse projects—and then stick to this commitment. Do not steal their time back.
- The project team can get input from external data warehousing resources by reading and attending courses and conferences. Bring consultants on board to design and build the data warehouse and/or BI applications. Use external resources for reviews and mentoring.
- Empower all participants to speak up! Create an environment in which people are encouraged and rewarded for bringing issues and concerns to the attention of project managers and IT or business management.
- The project team must be encouraged to take the time to address concerns raised within the team or issues brought up elsewhere in the organization. If something is not working, then stop spinning, figure out what needs to be done, and then do it!

While these are important, the single most important thing that can be done is to link any data warehouse initiatives to real business requirements.

**NOTE** Even projects trying to adopt all the best practices can end up with a stalled data warehouse. An honest assessment of the situation usually yields several contributing factors. This entire book is dedicated to helping executives and business and IT managers understand what building a DW really entails. This understanding can help all future DW initiatives to identify potential roadblocks and take proactive steps to avoid them.

### ***Aligning DW Objectives with Business Goals***

The only way to align the data warehouse with business goals is to know what those business goals are and to have an understanding of what a data

warehouse can do. Representatives from the business must work with lead data warehouse and BI team members to figure this out. Each individual division of an organization will have its own objectives and goals. There will also be objectives that span all of the divisions. What is the strategic direction of the company? What are the biggest challenges facing the entire organization? As new business strategies and objectives are developed, take a look at how the data warehouse can continue to support these goals. This is similar to the type of work that is done to justify a data warehouse project in the first place. This may have happened years ago. It is worthwhile to go back to find the original rationale. Often, many of the same overarching issues are still facing the organization. New challenges may have cropped up due to market pressures, regulatory changes, or advancements in technology needed to run the business, such as manufacturing technology.

The data warehouse manager and steering committee need to understand the organization's strategic objectives so that they can prioritize all data warehouse work to support those objectives. This makes it easier to decide the order in which the work will be done. It also makes explaining the data warehouse priorities much easier. The more closely aligned the data warehouse is to the overall corporate strategies, the better the likelihood of retaining funding for maintenance and expansion.

### ***Getting It Right This Time***

As you've already seen, while many data warehouse projects try to use best practices, the project may still fall short. Subsequent projects should tap into the experiences of any previous data warehouse projects. Project teams often have insight into what could have been done differently. It is hoped that these ideas were captured during the post-implementation project review. If not, seek out the team members to get their perspective on the previous project. If there is a lack of confidence in the current staff, then external help is needed. This may entail hiring experienced data warehouse staff or bringing in consultants to design and build the data warehouse. In addition to correcting known problems, the new project should also do the following:

- Communicate regularly to a wide variety of audiences across IT and the business community.
  - Openly acknowledge problems of the past.
  - Assure everyone that steps are being taken to address these problems.
  - Ensure that communication works both ways. Share what is happening and then *listen*.
- Implement sound data governance to define the data. Then, develop and publish useful documentation, including what each data element means

in business terms, a description of each report, and what to do when problems occur.

- Create new business processes that leverage the new data warehouse environment.

Overall, it is important to articulate what was different on this project—that is, why this will work when other attempts have failed. Be sure to include management's commitment, the participation of business representatives throughout, and the adoption of industry best practices. Include specific details that highlight how the new environment is targeted at the business. These benefits are best identified and communicated by the business representatives on the project team. Leverage what has been learned from this book. Seek out additional resources for specific areas that need attention. Get a more detailed understanding of areas in which the organization is weak.

Everyone, including executives, managers, and developers, can contribute more effectively by learning more about their immediate data warehouse responsibilities and having an understanding of what else is involved in designing, building, and using a data warehouse. Use the lessons learned and the experience of the new project team to build a better data warehouse. After development of the improved data warehouse and BI application is complete, additional attention must be placed on releasing these for use.

## Launching the Improved Data Warehouse and BI Solution

When an organization has already attempted to launch a data warehouse, there is likely to be more resistance with subsequent launch efforts. People will be skeptical about what the data warehouse will be able to do for them. Several important things need to be considered when getting ready to release the new and improved data warehouse environment. Because there have been other launches, it is important to take the time to do this right. These are the same principles that apply to any launch, but there should be no shortcuts this time. To recap these important steps:

- Ensure that the environment is ready to be released. Test everything: data accuracy, load performance, query performance, usefulness of reports and analyses, and the interface to the BI application. Everything has to be working *well* before you actually launch the data warehouse.
- Consider a dress rehearsal. Select a small group of helpful business users and go through all the steps of your planned release. This “beta test” can identify items that need more attention and help the team to develop a smooth release plan.



- Develop and conduct training about the data and how to use it for analysis, and help the business people learn how the reports/analyses can help them do their daily work.
- Create a visible and public launch of the new and improved data warehouse. Make a big deal about it. Even if you are not in the marketing department, you can usually find several people who are great party planners. Tap into those skills to pull together a great celebration. Order a cake, decorate the conference room and the business areas with balloons, and have a ribbon-cutting ceremony. Pass out pencils with the data warehouse support number printed on them. Invite the target business users and everyone on the project team. This is truly a time to let everyone know that things have changed. Be proud of the work that has been done. Celebrate the progress made to date and get ready for success!
- The business needs to provide a lot of proactive, visible support, especially from upper management. Include key managers in launch meetings to share their expectations that the data warehouse environment *will be used* to help the company meet its goals. If this isn't possible in person, it can be shared via a memo or e-mail message.

While the launch should be carefully orchestrated, this is really just the beginning. Now follow-through is needed to ensure that the data warehouse is used and that business value is achieved. With the launch of the improved data warehouse, the project team should loop back to the steps outlined in the beginning of this chapter to finish the project and encourage use of the data warehouse.

## **In Real Life**

---

While the two companies we have been following are at opposite ends of the spectrum from a size perspective, both have many complexities to deal with. Let's see what challenges they are facing.

### **Lack of Support for the Production DW at Giant Co.**

Giant Company has been diligent about following their project management methodology. Resources are assigned based upon these project needs. Unfortunately, no one paid much attention to supporting the data warehouse once the project was complete. Sufficient funding was not secured and a single ETL developer was assigned the responsibility to keep things running. Problems were channeled through the standard help desk. Unfortunately, this left many areas open for problems. The help desk was not educated or prepared to deal



with the kinds of questions that began to come in about the data warehouse. Many of the questions regarding data meaning, accuracy of calculations on reports, and how to use the BI interface could not be answered by the remaining ETL developer.

Complaints began to flow in through the help desk and management channels. The data warehouse, including the BI application, was really geared toward helping the business. The business champion stepped up to work with IT and business managers to determine what to do. They proposed setting up a data warehouse center of excellence. This included the staff to effectively maintain the data warehouse and provide appropriate support for the business community. The executive steering committee approved the requested funding. Once these resources were put in place, they were able to set up the procedures and processes needed to help the business leverage its investment in the data warehouse.

## **Unleashing BI at Agile, Inc.**

The business users at Agile, Inc., were happy with the starter set of reports that were developed on their behalf. In order to encourage a lot of use, all of the users were given the authority to create their own calculations and publish reports for general use. Many of the business users grabbed this new opportunity to start creating their own reports and analyses. Dozens of reports and calculations were added. The users who were not interested in building their own analyses quickly learned to look through the published list of available reports to find things to try. Unfortunately, many of the old problems began to emerge. Too much time was spent in management meetings debating which numbers were correct. This was frustrating for everyone. Business managers decided to put a stop to this and figure out why the data warehouse had not fixed this fundamental problem.

A core team was brought together to study what was happening. The team included a data modeler, an ETL developer, a BI application developer, and a project manager. After a quick review, it was apparent that not all of the calculations that had been added were correct. Further review of the reports showed inconsistencies too. The permission allowing everyone to tinker with the BI environment was revoked. A high-level understanding of the data and how to use the BI tool was set as the minimum in order to be granted privileges to add to the BI environment.

The education offerings were aligned with these skills to provide a path for people to follow to improve their skills. A review process was also set up so that all new additions, reports, or calculations were reviewed on a weekly basis by core BI application developers, who served a quality control function. Some time was spent cleaning up the calculations and reports that had gotten out of control, and the new policies kept the situation from getting out of hand again.

## Summary

---

A great deal of time and energy are put into designing and building a data warehouse. Now that the construction has been completed, the fun really begins. This chapter provided practical suggestions for launching the data warehouse and getting the business community to embrace its use. It takes time to effect real change, and the data warehouse can be a useful tool to help an organization improve how business is conducted, which benefits the bottom line.

The work for a data warehouse is never really done. Once the project ends, ongoing support and maintenance activities begin. The technology, the ETL system, and the BI application need to be kept current and running efficiently.

Many data warehouses have already been built and are in production; some are successful and some are not. This chapter presented common causes for a stalled warehouse and provided practical recommendations for moving it forward. Now it is time to look toward the future. Chapter 13 covers planning for expansion and growth, how to deal with challenges across a large enterprise, what's next for data warehousing, and how to measure success.

# Achieving Long-Term Success

The data warehouse project has been completed and appropriate measures have been taken to maintain it. Now the business should be getting used to using the data warehouse and learning how to make it work for them. This is not the end, but really the beginning. This chapter looks to the future so that proper planning can be done to support expansion and growth of the data warehouse. Several ideas are offered to help manage data warehouse resources across a large enterprise. After a quick look at the future of data warehousing, the chapter ends by highlighting techniques to measure business value and sustain success.

## Planning for Expansion and Growth

---

Use of the data warehouse environment will generate new requirements, and problems will need to be fixed as part of maintaining it. The data warehouse support team may be able to accommodate some minor enhancements. Beyond this basic support, enhancement requests and new requirements will need to be addressed.

Periodically, these requests must be reviewed and evaluated. This can be done as part of a scheduled planning process. These requests may be bundled into releases. Each release should be handled as a separate project to ensure the coordination of sound design and development, complete testing, quality assurance, documentation, and associated education.

This planning is typically done annually in order to ensure appropriate funding for the upcoming year. In addition to maintenance releases, brand-new projects may be defined to tackle the work of adding new data sources or supporting completely different business groups. In preparation for planning, the requests for changes and/or additions to the data warehouse need to be identified.

## Exploring Expansion Opportunities

Usually, you can do many different things to enhance the data warehouse. A concerted effort must be put forth to collect the many requests and ideas. This work is often spearheaded by the data warehouse manager but carried out in collaboration with managers, business analysts, and/or members of the data warehouse support team. These requirements can be identified in a variety of ways:

- Review feedback captured by the problem-tracking system that was discussed in Chapter 12.
- Send an e-mail survey to current data warehouse users to ask what else they would like to see or be able to do.
- Schedule brief requirements review meetings with representatives from each of the major areas, using the data warehouse to gather and/or confirm their high-level requirements.
- Go back to the requirements gathering documentation to see what other business areas and opportunities were identified. Clarify with the business community which requirements are still relevant.
- Ask the DW business support team for their input (although they may be the ones performing this work).

This is not a full requirements gathering process, but rather a variety of ways to examine what the next data warehouse project(s) should tackle. If the total number of items needing attention is too great, then the business and IT managers must do some initial prioritization to reduce to number of requests to a manageable size.

For the requirements that are under consideration, additional research may be required to ensure that the real objective or requirement is understood, and to figure out what work is needed to design and develop a solution to meet this need. For each request, the findings need to be documented. This can be brief, but it should still be written down. Details that need to be documented include: a description of the request; the reason or business purpose; a description of the technical problem or challenge that needs to be addressed; an estimate of the magnitude of effort (large, medium, small); identification of the impact of this change on all areas of the data warehouse environment, including the ETL system, dimensional structures, and the BI data delivery layer. The business impact of the request should be captured when possible. This includes the benefit of making the change and the risk of not addressing the request.

The research for enhancements to the data warehouse may be done as the requests come in, rather than waiting for the next planning cycle. The collection of documented requests is used to help set the order in which they will be implemented.

## Prioritization of Feedback

After each individual request has been explored, the requests need to be examined as a whole. The goal is to determine how to best group the requests to provide the most benefit with the least amount of effort and/or rework. For example, instead of having four releases next year that all work on the product dimension, consider combining all of the product dimension requests into a single release. That way, the product dimension and the ETL processes to build and maintain it will be worked on only once.

Look at the combinations of requests in several different ways. Combine everything needed by department. Look at all of the requests that only require BI application changes. Group the requests by the business impact or return on investment. Group the requests based upon source system data that is involved. While representatives from the data warehouse support team can study the requests, it is also important to get input from different business groups too.

**NOTE** It is not recommended to put together technical-only releases. Resources will be consumed for some time to work on the release and when the results are unveiled, the solution will deliver exactly what the business had before. Look for ways to bundle technical foundation work with value-added enhancements.

To help the decision-making process, it is important to tie the releases to the potential business value. Each of these alternatives defines a new project, which needs a project charter as described in Chapter 5. Some organizations require a project charter to be developed for each alternative being considered. Other organizations prefer to examine the different alternatives and then require the project charter for the alternative(s) that have been selected.

Ultimately, a combined group of IT and business representatives must develop recommendations for what each release should contain. These recommendations and all new projects can then be presented to the executive steering committee for approval. Several alternatives may be developed and presented. The pros and cons for each can be discussed, and the steering committee can make an informed decision about which alternative to pursue.

Once the decisions are made, the direction can be communicated to the rest of the organization. If anyone is not satisfied with the direction, then their concerns can be raised through their management chain, either IT or business. This is when each representative on the steering committee can help by communicating the options and the business rationale for the decisions. This helps ensure that everyone understands (even if they don't agree) that the business is driving the content and priorities for the data warehouse, rather than IT setting these priorities.

**RUNNING ON EMPTY**

**There are only so many hours in a day. Depending upon what needs to be done, the total number of resources available to work on data warehousing can easily be exhausted—“exhausted” as in you can run out of time and each of the individuals can be fatigued. The more skilled the project teams becomes, the more is expected of them. Sometimes you can push everyone to work a little harder to get a specific deliverable finished, but you cannot expect people to give that extra effort with no end in sight. Ensure that sufficient resources, both IT and business personnel, are allocated to maintain the data warehouse. Keep an eye on the balance of new project resource needs and the available personnel—again, both IT and business. If you are falling short, you must add more resources or reduce the work.**

The result of the prioritization process is a new set of projects that need to follow the standard project planning procedures. Now that you have identified what projects will be started to expand and grow the data warehouse, it is worth taking some time to look at leveraging skilled data warehouse staff members across the organization.

---

## **Managing Enterprise DW Resources**

---

Many organizations find that putting together a project team and setting up ongoing support is sufficient. There is clear coordination from project to project because the same people are involved in all the work. If you work for a large enterprise, however, this is not the case. The staffing and management of people across the organization is significantly different. Often, people and their skill sets are listed in a human resources database. When forming a project team, this database is used to locate the people needed to fill the different roles. This is based on both who has the background or base skills and who is available at the right time.

This creates a much higher level of complexity to ensure consistency across the data warehouse environment. It is not unusual to have a project team of 10–15 people. If eight projects are under way across the enterprise, then you may have 80–120 people working in the data warehouse environment. Additional strategies are needed to leverage data warehouse knowledge and techniques. Some sort of centralization of data warehousing resources is needed.

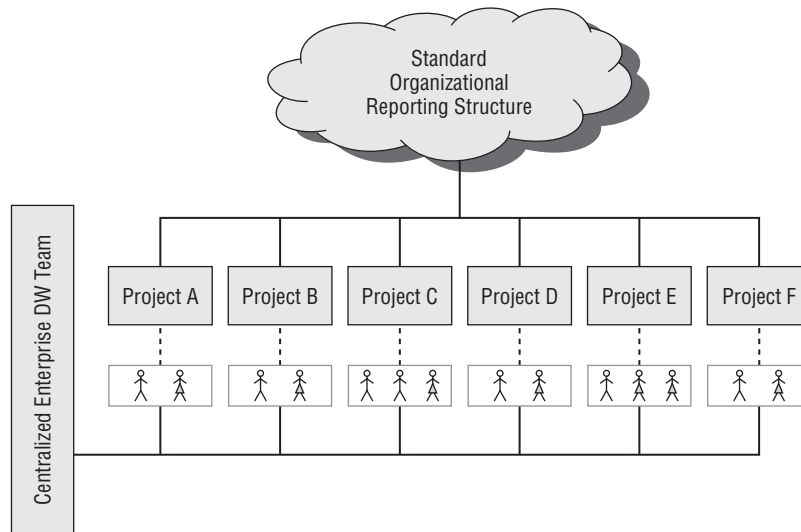
### **Creating an Enterprise Data Warehouse Team**

A single centralized team of data warehousing professionals can be formed. This is a collection of the most experienced people from across the organization. It may also be called a *center of excellence* for data warehousing. Their jobs

include the responsibility to advocate, implement, and help others work toward enterprise data warehousing goals. There are several different ways that this can be handled organizationally. This section shares two examples.

### ***The Centralized Enterprise Data Warehouse Team***

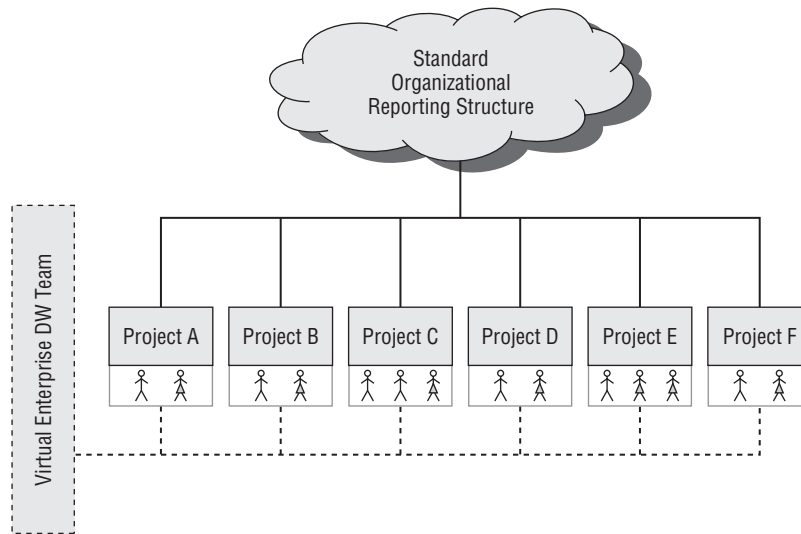
With this structure, the enterprise data warehouse personnel report directly to an enterprise group. While the staff members belong to the enterprise team, each person is assigned to and works on specific data warehouse projects. This ensures that they have a continued role in actively developing and delivering value to the organization. It also enables each to leverage his or her experience and mentor the project team while keeping an eye on what is best for the overall data warehouse initiatives across the organization. Figure 13-1 shows how the centralized data warehouse team is organized.



**Figure 13-1** Centralized enterprise DW team structure

### ***The Virtual Enterprise Data Warehouse Team***

The virtual enterprise data warehouse team is similar in function to a centralized team but it is organized differently. The highly skilled data warehouse staff members are full-time members of project teams and have the same management chain as all other project team members. These key data warehouse people have dotted-line responsibility to the enterprise data warehouse team. Highly skilled data warehouse resources are leveraged on each project. A portion of their time must be allocated to enterprise data warehouse work. This work must not be pushed aside. Figure 13-2 shows how the virtual enterprise data warehouse team is organized.



**Figure 13-2** Virtual enterprise DW team structure

Sometimes the concept of forming an enterprise team can take a wrong turn. This can result in an *ivory tower team*. This happens when the enterprise team does not have any actual project involvement. Their only work is at a conceptual level for data warehousing across the organization. This is not a recommended approach. This type of team tends to set up methods and policies to which all projects should adhere, but they lose touch with what is realistic. They become more of a burden and added expense, rather than helping data warehouse initiatives become more successful.

Regardless of how the team is organized, there are common skills and functions that this team can perform. The types of people who often are part of the enterprise data warehouse team include the following:

- Technical specialists who support the tools
- Project managers
- Data modelers or data architects
- Lead ETL developers
- Lead BI developers
- Business analysts

Not all members of every data warehouse project should belong to the enterprise data warehouse team, and each project should be staffed with a variety of people. The enterprise DW team should be the best and most experienced data warehouse personnel in the organization. They can lead, mentor, and teach the rest of the individual project team members. This also provides the opportunity for new people to begin to develop these



valuable skills. Over time, they may excel and become part of the enterprise project team.

## Enterprise DW Team Responsibilities

The enterprise data warehouse team serves as the central clearinghouse for ideas, techniques, and solutions. They can facilitate communication through e-mail, newsletters, and briefings. They can also coordinate interaction between different groups across the enterprise. Following are the key responsibilities of this team:

- **Dealing with technology:** Problems can arise when installing software for the first time or upgrading to new releases. Centralized resources can help resolve a problem once. If each project is independent, time and resources are wasted trying to figure out the same problem. A centralized group can also keep an eye on new technology as it emerges. This group can also test new products that may help the organization. Volume discounts may be possible by consolidating product purchases.
- **Adoption of industry best practices:** There is value in keeping current with what is happening in the data warehouse industry. Anyone involved with data warehousing projects should have access to the proper ongoing education needed to do their work. The central group can scan all of the different publications and regularly attend industry conferences. This group should also be responsible for helping to determine how these industry practices can be adapted and used by the organization.
- **Developing internal best practices:** Because each organization has its own existing procedures for running projects, it is important to ensure that these are fine-tuned to better support data warehousing. This may mean creating standards for what deliverables will look like from each data warehouse project.
- **Leveraging business experience:** As each business group learns to use the data warehouse, the insight they gain can support their work. An enterprise data warehouse team can be a focal point for sharing these ideas. Sometimes the same report or analysis can benefit other groups. The team can share suggestions about how to encourage use of the data warehouse or how to modify business processes to tap into it.
- **Providing a centralized location for a consolidated view of data warehousing across the enterprise:** The enterprise data warehouse team can compile information about each different data warehouse, as described in Chapter 3. The team can be responsible for summarizing the different initiatives that are in place, under way, or being planned.

Each organization must determine what responsibilities are best suited to their own enterprise data warehouse team. The team should provide a focal point for channeling communication, fostering cooperation, and helping to ensure that each data warehouse project is successful. However it is organized, the team must be responsive, and avoid becoming a bottleneck.

## **Funding the Enterprise DW Team**

Logically, it makes sense to gain economies of scale for technology and people, but who will pay for this? The budget for a centralized team may seem expensive at first, but remember that you are already paying for this as part of each and every project and existing standalone data warehouse. This seems to be more when you add it all together, rather than have these costs embedded in many different places.

Additional time is required for coordination and communication, but a cost savings is realized by having skilled resources available so that each project does not have to start from scratch (data architecture, selecting the technology, etc.). You can also avoid costs by preventing projects from making choices that will not work. Having a centralized data warehouse function will not prevent all mistakes, but it can certainly help each project avoid the most common errors and get things done sooner.

The real challenge is to determine a fair way to allocate the cost of these central data warehouse resources across all the different groups. A variety of funding models are used to pay for an enterprise data warehouse team. The first thing to understand is how other enterprise services are funded. If all database administration resources are centralized today, are the services charged back by hours worked or by size of the database? It can be straightforward to follow an existing pattern. The enterprise DW team costs can be shared across all of the groups who are using (directly or indirectly) the data warehouse. These costs can be allocated by the number of users, the size of their data, the number of queries that are run, or sometimes even by the size of the rest of their budget.

Another approach for funding the enterprise DW team is to charge for services. Chargeback, or consulting, rates must be determined, typically based upon the cost for an individual or perhaps as a blended cost of all of the team members. The biggest risk with this approach is that use of these internal consultants is often viewed as optional. These resources are either avoided (due to the high internal chargeback associated with such highly skilled and experienced employees) or their services are used only if the team is in big trouble. Either way, the organization still incurs the cost of individual project teams forging their own path.

Another common alternative is to automatically allocate the cost of the enterprise DW team across all projects or groups who are involved in data warehousing. This can be based upon the number of members on a project

team, the size of the data, or the number of users. This can help mitigate the risk of projects not wanting assistance from the enterprise DW team. If a group is already paying for the team, then there will be a higher level of interest in using its services.

## Pushing into the Future

---

Looking to the future is important to continue to push the data warehouse forward to support the organization. The work on a data warehouse is never done, but should become an integral part of how the company grows and changes. The data warehousing industry itself continues to grow and change too. There are advancements in technology and the emergence of new best practices. This section serves as an introduction to some of these new areas that are emerging and maturing at the time of this publication. It is not intended to provide in-depth definitions and/or recommendations for implementation.

## Embedded Business Intelligence

The concept of *embedded business intelligence* refers to use of the data warehouse when it is embedded into other business workflow processes. This often is done in a way that the business person or customer is not aware that the data warehouse is part of the process. For example, a data warehouse can be used to generate parts of a phone script to drive what questions are asked, and what options or offers are presented to the customer. This can also be embedded into other applications such as online ordering—for example, to suggest to a customer ordering hiking boots that other customers who purchased hiking boots also ordered sock liners and safari hats.

The reason that this differs from using the operational database itself for the application is that this requires historical data, often from multiple sources in the organization. The data warehouse is a logical source for this integrated historical data. Several things need to be in place to enable this type of operational business intelligence:

- The data warehouse must be able to support these queries. It must contain the data needed by the operational application and be able to provide an acceptable level of query performance.
- The operational system must be enhanced to tap into the data warehouse.
- Appropriate business processes must be in place so that the business community understands how to leverage the results that have been embedded in their operational applications. For example, this may simply be training about a new pop-up screen, such as how to use it and when it can be ignored.

Embedded business intelligence may also be implemented as a separate step in a workflow, but not directly embedded into an operational application system. For example, after the order scheduling process has finished, there may be exceptions flagged for further investigation. A dashboard or BI portal may be used to check on these exceptions to help determine what action needs to be taken. The key point here is that business intelligence is used as an integral part of operational business processes and taps into historical data from the data warehouse.

## **Operational Business Intelligence**

For the purposes of this book, the term “business intelligence” has been used to describe the delivery of data from the data warehouse to the appropriate business audience. Operational business intelligence is quite different. Operational business intelligence changes the focus to address more immediate business issues that deal with daily business operations. Operational BI is used by frontline managers and customer-facing employees to measure and monitor operational processes.

Typical business intelligence would help track and monitor claim payments, the number of claims being handled, and the amount of time spent on the phone, each compared to target values. It is often necessary to do this more frequently than a weekly or daily claims report. This operational tracking is needed to provide more immediate feedback of performance so that behavior can change in real time. A bigger push can be made to complete sales or to follow up on lingering claims to get to a final resolution.

This operational performance could be displayed on a big screen that shows the number of calls on hold and the number of calls handled so far on the current shift. This information can then be reported by a team or an individual claims handler. Some organizations prefer to publicly show results by team, and then show each person’s results on his or her own individual screen. Sometimes these measurements can be pulled directly from the operational system. However, it may be necessary to include historical trending and data from other systems that the operational system does not have. In this case, this is more in line with the requirements for real-time data warehousing, which is discussed in the next section.

When exploring operational BI, you need to consider a number of things, which revolve around the data in the operational systems. The following characteristics of data need to be evaluated to determine whether the operational system can support the business need:

- History
  - Does the historical data needed to support analysis already exist in the operational database?

- Does this history include both transactional data and reference data as it existed at the time of that transaction?
- Cleanliness
  - Is the data in the operational system clean enough for effective reporting?
  - Are there duplicate entries for the same customer?
- Performance
  - What are the performance characteristics of the current operational system?
  - Will the additional workload to support reporting and analysis affect the system's efficiency?
  - What kind of query performance will the report and analytical requests see?
- Integration
  - How will data be integrated between different operational systems?
  - How many exceptions are there for these integration rules?
  - What, or how many, changes are needed to the underlying data to support integration?
  - What is the impact of these changes on the functionality of the application system itself?

As with any other BI application, you need to determine the business requirements—that is, articulate what you are trying to accomplish. Then potential solutions can be explored to address the business need.

## Real-Time Data Warehousing

Many companies have an increasing interest in real-time data warehousing, but what is it? For many business people, real-time data warehousing is the capability to ask a question and have the request processed and the results returned immediately. However, this is not what most people in the data warehousing industry mean by real-time data warehousing. From an industry perspective, real-time data warehousing refers to providing access to real-time (when a business transaction completes) or near real-time (soon after a business transaction completes) data through a data warehouse.

For example, real-time data would enable you to see network usage for the last ten minutes, or view investment positions as of “now.” In addition, there are other closely related concepts that address direct access to data in the operational systems and closely link business intelligence capabilities with

the operational systems. Here, real-time data warehousing refers to the how quickly data is made available through the data warehousing environment.

The question at hand is what real time means to you. The objective is to get data to the business at the right time. The right time means getting the data to the business users in time to help them make decisions. This may mean real time in some cases, but in many others it means “by the start of business the next morning.” The goal is to get the data to the right people when it will be most useful for decision making.

Many organizations agonize over building a real-time environment, but their business does not operate at that speed. For insurance companies, for example, it may seem interesting to have real-time policy and claims data in order to watch loss ratios unfold. However, unless you are in the excess and surplus business, you can’t change your rates immediately anyway, so if you see a pattern emerge this afternoon, you would still need to build a case and present it to the state board of insurance before you could institute a rate change. In other words, if a business decision can’t be implemented immediately anyway, what is the value of building the infrastructure needed to support loading real-time data? In this case, the right time may be to continue with daily loads of data to monitor the business.

In some cases, the data may need to be refreshed every hour or perhaps every fifteen minutes. If data is truly needed up to the second, then a continuous trickle of data may be more appropriate. In any case, the infrastructure must be built to support these increased demands. The data may be loaded into an *operational data store (ODS)*, an integrated, up-to-the-moment operational database or into a partition of the data warehouse set aside for this high-performance real-time data loading. At the end of the day, if the solution that has been devised is working, then great! If your organization does not have a real need for time-critical data, then don’t worry about these issues. Continue to load and maintain the data with the timing that is necessary to support the speed of your decision-making processes. There are many considerations regarding the design, development, implementation, support, and cost of real-time or near real-time data warehousing, a subject beyond the scope of this book.

## **Unstructured Data**

In years past, it was ludicrous to think of searching through textual data in a timely manner to find useful tidbits. Think about searching the Internet, which offers an overwhelming volume of information, in many different formats. Therefore, we know it is possible. What does this mean for data warehousing?

Most data warehousing technology and methods have been focused on structured data—how to organize and use it. However, many other types of information are useful as well. Organizations often have maps, photos, contracts, and schematics and/or blueprints. For example, it would be helpful

for claim handling to be able to pull up photos of both vehicles involved in an accident or a photo of the wind damage from a severe thunderstorm.

A basic method to handle these unstructured elements or objects is to have descriptors assigned to them. These keywords can be stored and searched the same way as any other structured data. Once the object of interest is identified using traditional query methods, the object itself can be pulled up for viewing.

Advancements continue to be made for storing and gaining flexible access to these unstructured types of data. As breakthroughs are made, often for purposes other than a data warehouse, we need to think of ways we can harness these capabilities to support business decision-making processes.

## Monitoring Industry Innovation

As technology continues to leap forward, there are more and more opportunities to leverage the data warehouse. It is necessary to keep up with advancements in the tools and to keep the organization moving forward with efforts to utilize these improved technologies.

To put things into proper perspective, think back to the advent of the PC. Suddenly, each individual was expected to handle a whole new set of tasks. Not only did you have to draft the content, but everyone was expected to type reports and memos themselves. The entire approach to communicating with one another has changed as well. We have instant access to one another's calendars for scheduling meetings. Questions and answers can be handled via a series of e-mail messages, sometimes without direct interaction. Moreover, accessing e-mail is no longer limited to sitting at your desk, but can be used almost continuously with PDAs and cell phones.

As new generations enter the workforce, their expectations of how to use technology and what that interaction looks like is changing dramatically. Texting, instant messaging, blogging, and social networking are a regular part of the fabric of their lives, often starting in elementary school. They also have high expectations for technical interfaces because of their use of electronic gaming systems. These games offer incredibly realistic graphics and fluid 3-D movement. Consider the interface used for the Wii, whereby your actual physical motions drive the games. Even modest handheld systems provide rich, realistic interfaces. (You can blow bubbles in the game *Nintendogs* simply by blowing on the screen!)

With data warehousing, it is important to keep an eye on what new innovations are being introduced. Someone in the organization, perhaps the data warehouse manager, should regularly attend data warehouse industry conferences to see what new technologies are being introduced to the marketplace. This is also a good way to keep on top of new design, development, and deployment practices. Representatives should also attend any conferences or seminars offered by the vendors of the tools you are using in order to keep



current with new product features and enhancements. In addition to keeping an eye on technology, it is important to continue to monitor and measure the impact that the data warehouse is having on the organization.

**NOTE** Don't jump on the latest fad. It is fine to keep up with trends and changes in technology. Look around to see what is out there. However, always come back to the reality of what are you trying to accomplish. Implement products that support the business community to deliver business value. Just because a new feature is getting a lot of press does not mean that it will provide value to your organization.

## Moving Toward Business Value

---

Success can mean a wide variety of things to many different people. Only you can determine whether the financial investment and all the hard work was worth it. What return is the organization actually seeing? This is not usually easy to answer. Someone needs to follow up with the business groups to determine what types of progress and success they are seeing from use of the data warehouse.

### Measuring Success One Step at a Time

Although this may have been done as part of a post-implementation project review, it is still valuable to go back to the information you captured when the project was initiated and see what was learned during the requirements gathering process. The project objectives were captured in the project charter. Look at the pre-data warehouse baseline results. Collect feedback on what is happening now that the data warehouse is in place. It is also important to collect details about all the ways in which the data warehouse is being used. Sometimes, the biggest benefits were part of the original objectives.

What types of results do people see? One organization determined that increasing customer retention by 0.5% would result in \$4 million of additional revenue annually. Another organization was able to improve response rates from promotions by 1%, resulting in an average increase in sales of \$1.3 million per promotion. A third organization was able to streamline and increase the accuracy of their inventory replenishment process resulting in more consistent national distribution of products. This helped eliminate overstocks in some stores while others were out of stock. This reduced the post-holiday inventory that had to be discounted. Overall, this added an average of \$3 million dollars in revenue for each major holiday season. Many big successes like these are not widely publicized—organizations with these kinds of results are often unwilling to reveal anything to the general public.

It is much more common to see a decision that resulted in \$25,000 of new revenue, another decision that saves \$10,000, and yet another that results in



\$5,000 a month in add-on customer sales. For some companies, these are each rather small amounts, but when accumulated over time the financial results can be significant.

How can these results be identified and measured? It can be challenging for the data warehouse team to capture information about these benefits. This is much better done through the business management chain. It may be difficult for business people to attribute their successes to the data warehouse. A report or results of an analysis may contribute to making a decision, but someone's experience and insight from non-data-warehouse input also are used to make decisions. Link the collection of business value to public recognition.

Business managers need to work with their staff to determine when and how the data warehouse is used. While this can happen informally, the best results come from a concerted effort to capture and document business results. Ask everyone to submit the "decision of the week," the best new report, or results that made a difference each month. Provide incentives to the best submissions. These do not need to be costly rewards, but something that would be valued—perhaps being allowed to leave two hours early on Friday or getting to park in a favorable location. Purchase a "goofy" trophy that is awarded to the person who made the best use of BI in the current month. This can become a "badge of honor" as it gets passed around the group. Create an environment where people can "brag" about what they have been able to accomplish. If you do not have success stories, even small ones, then you need to jump in to figure out why. There may be challenges in getting the business to use the data warehouse.

Another tactic that can help quantify value is to compare current business results with the baseline results articulated at the beginning of a data warehouse project. If the organization's goal was to improve market penetration by 0.5%, what was the original penetration? What is it now? Every goal may not have been reached yet, but some progress is usually observed.

**NOTE** Be sure to communicate your success to the rest of the organization.

**Publish the simple usage statistics and business results as they are identified.**

**Success must also be shared with the executive business sponsor, the business champion, and other business and IT managers.**

Once mechanisms are in place to get people to share how they are using the data warehouse to drive business results, it is important to keep track. Someone needs to be responsible for documenting these results. While the first impulse is to get the data warehouse manager to do this, it is much more effective to get someone from the business community, perhaps the business champion, to take on this responsibility.

Part of this task is following up to assign a dollar value to what was done. Sometimes this can be provided as proof of results, but sometimes it

takes additional discussion to clarify. In many instances, decisions may have been made in a timelier manner, which may be difficult to quantify. The most exciting part is adding all of these individual results together. There is often a much bigger impact overall than expected. These composite results need to be shared with upper management and across the business and IT communities.

## **Adjusting Expectations to Reality**

The previous section focused on identifying and quantifying results. Even when concrete results can be measured, it is possible for the data warehouse to fall short of expectations. How do you know what people's expectations are? When a data warehouse project is under way, the official expectations should be documented in the project charter and/or scope. However, individuals may have their own expectations, which may or may not align with the official version. Individual expectations are often shared during the process of gathering requirements. Once the data warehouse goes into production, it is still important to keep in touch with expectations of the user community. Periodically, go out into the user community and listen to their concerns. This can easily be done when gathering requirements and enhancement requests for the enhancement planning process.

Sometimes individuals or groups seem to always have crazy ideas. Don't just write off their input. These are often the people who can push the organization forward into new and better ways of running the business. In other cases, individuals bring a wealth of previous experience to the organization. This can include ideas that no one has even dreamed of, let alone knew were possible.

Naturally, serious problems can arise if what is delivered and supported falls short of expectations, regardless of how unrealistic those expectations may be. After exploring expectations across the organization, it is important to follow up with a good dose of reality. Communicate the data warehouse priorities back to these groups. If their needs and expectations will not be met, let them know now. While this may not be an easy meeting, it is much better than waiting and allowing their hopes to grow. When possible, these discussions should be led by the business decision makers. Share the business rationale for the priorities that have been set. If an individual or group continues to have concerns, they should work directly with their management and the data warehouse steering committee.

**NOTE** As noted elsewhere in this book, communication is critical to data warehouse success. Too often, communication is viewed as telling others about what is happening. Keep in mind that good communication involves listening too.

## Keeping the Momentum Going

Executives and managers, within the business and IT, play a significant role in the ongoing success of the data warehouse. These groups must continue to strengthen their partnership and work jointly to achieve the data warehouse objectives. Management must continue to commit by allocating necessary resources and being willing to make decisions about priorities and direction of the data warehouse. This also requires a willingness to support data governance to ensure that the data is defined and used effectively. Finally, business management must communicate the value of information to the organization. This includes encouraging the use of data and underscoring the value of performing analysis and making fact-based decisions. Management sets the tone that drives the organization's attitude toward the data warehouse.

When a data warehouse is first released, there is often a great deal of enthusiasm and support. When there has been strong partnership between the business and IT throughout the life of a project, everyone feels a sense of ownership. The business is trained and use begins to take off.

Once the initial glamour wears off, frustration can build, but if problems and issues are addressed in a timely manner, this can be managed. As the data warehouse and BI application settle in to become the new status quo, several things can be put into place to help ensure long-term adoption of the data warehouse. While the following suggestions may be part of the data warehouse project, it is important that they continue as the data warehouse moves into production:

- **Include the use of the data warehouse and skills to perform analysis as part of the job description and responsibilities of appropriate personnel:** This sends a clear message that these skills are valued and identifies who is expected to perform these functions. This helps ensure that these skills continue to be cultivated for years to come. Otherwise, the level of expertise and skill can degrade over time.
- **Groom replacements for all critical data warehouse roles:** Both business and IT staff members will change positions over time, often within the organization and sometimes leaving. Although a key individual will serve as the data warehouse executive sponsor, another person should begin to learn and understand what that role entails. When the executive sponsor leaves that position, someone is ready to step up and shoulder those responsibilities.
- **Continue to offer introductory education:** Education that introduces the data, the reports, and analysis must continue to be offered regularly. This is the primary mechanism for new staff members to become proficient in using the BI solution.

- **Provide advanced education opportunities:** In addition to ensuring that new staff members learn the basics, it is important to offer a series of courses or tutorials to help current users increase their data warehouse skills. Education opportunities should include more in-depth training on the data, additional features, and capabilities of the BI technology and how to perform more advanced analyses.
- **Be practical:** Best practices are often the culmination of years of experience, yet rarely does one single project or organization do everything in every possible area completely and thoroughly. Ensure that the development and enforcement of standards does not overshadow reality. Make informed decisions about what is feasible for the organization today, and move toward your long-term vision.
- **Be flexible:** The entire management team must be willing to listen and then adjust as necessary to keep the data warehouse going.
- **Don't give up:** Data warehousing must be viewed as a new corporate lifestyle, rather than a single project with a set deadline for completion. The data warehouse is never "done." When problems crop up or a data warehouse stalls, many organizations simply cut the funding and dissolve the project teams. Then, after several months, a new effort is created to start over. Rather than stop completely before a project is perceived to have reached the point of delivering value, take a hard look at what progress has been made to date. With a little more patience and effort, some value can be seen.
- **Be sure to follow through:** Too often, great work is completed to set standards but then they are never adopted. The enterprise team can pull together recommendations for best practices or to enhance a project methodology to support data warehousing, but if these are never used by any projects, then this has been a waste of time. Once standards and recommendations are developed, additional effort must be expended to put these into practice.
- **Continue to address data governance issues:** These include defining the data, implementing master data management, and developing methods to address data quality issues.

These are a compilation of practical steps and areas that require ongoing attention to increase the long-term viability of a data warehouse. It is more important to keep an eye on how the data warehouse enables business users. When the data warehouse is helping support the business, this generates more use. Business managers can do several things to encourage continued use of the data warehouse:

- **Identify and publicly acknowledge successes, large and small:** This is the most important thing managers can do. As part of their regular progress reports, employees should be encouraged to include how they used the data warehouse. As successes are recognized by management, others within the group will also want recognition. This taps into the competitive nature of people to keep up with their peers.
- **Continue to identify business opportunities:** It is never too late to gather business requirements. This is not collecting desired reports, but understanding how data could support the business community. Learning what is needed to support different business groups is critical to ensuring successful reporting and analysis. Chapter 6 has specific techniques for gathering requirements, and explains how the business community can prepare and effectively participate.

In addition to partnering with IT on a project, the business community should also continue to seek new ways to leverage the data in the warehouse. This should happen without a formal requirements gathering process associated with a project. Some new ideas may be addressed immediately, while others may need to be included as part of maintenance or a new development project.

- **Keep an eye on how the data warehouse is being used:** What types of analyses and decision making is being done? Take the best ideas and share them across the group. This helps spread the best ideas and practices across the organization.
- **Allocate some time for visionaries to explore:** This enables them to take existing analyses one or two steps further or to otherwise be creative and follow up on new ideas they may have.
- **Work with your staff to determine what additional education would be beneficial:** There are many different types of education, not just how to use the BI application. Consider enrolling employees in courses or seminars to learn about business process reengineering, your industry analytics, or advanced statistical modeling.
- **Use the data warehouse yourself to measure the group's performance:** Make sure that the group knows that you are using it.
- **Encourage deeper use of data:** Don't stop by asking for a high-level summary of the business; dig deeper to identify patterns and trends. Over time, this will progress from looking at how things are to identifying situations that need attention. Next, move toward identification of potential problems so that proactive measures can be taken to avoid them. You can help drive this effort by the types of questions you ask.

- **Stay in touch with the organization's data warehouse staff:** Regular communication is natural when you are involved in a project. Keep up periodic communication after the project is complete too. Make sure you know what enhancements are planned, what is happening with the tools, and explore how other groups are using the data warehouse. Provide feedback to the data warehouse support team—not just what is broken, but what is working for you too.

This section has reviewed several areas that are often overlooked as a data warehouse moves into production. Perhaps not at first, but over a number of years, these concepts can easily be lost. Make a conscious effort to include them into the fabric of the organization. The goal is for the data warehouse to become a tool that is used regularly and relied upon, like using a word processor or checking e-mail. Getting to that point requires the encouragement and support of business management. No matter how desperately the IT people want others to use and appreciate the data warehouse, the impetus must come from within the business community.

## **Celebrating Progress**

A lot of attention is given to identifying and working through problems. It is even more important to ensure that success is also identified. Progress is often measured and included in executive briefings. Too often, this information is not shared with the rest of the organization. A lot of hard work goes into building a data warehouse, and executives and managers play a critical role in motivating and encouraging staff members who are involved. While each organization has its own ways to reward teams and celebrate progress, here are a few simple things that can be done to encourage the project team and the business users:

- **Publish a data warehouse/BI newsletter that highlights accomplishments:** These do not need to be limited to multimillion-dollar decisions. An accomplishment could be that a certain group got 100% of their staff members to complete the online tutorial or that three new reports were created to support the call center managers.
- **E-mail a thank-you to someone who put in extra effort to respond to a last-minute request from the executive team:** Copy the rest of the group on the e-mail so that others know about it too.
- **Surprise the ETL developers with balloons:** These are some of the hardest-working, least-known members of the project team.
- **Send thank-you notes to the support team after a difficult problem has been resolved:** This can be from business representatives to the IT team or from the IT team to a helpful person on the business side.

- **Have a data warehouse anniversary party.**
- **Invite members from the project team to see how the BI application works:** Have a business analyst provide the demonstration so that their excitement is seen. This is why everyone worked so hard.
- **Offer coupons that can be earned by sharing how an analysis helped you get your work done:** The coupons can be redeemed for whatever rewards your staff finds valuable, such as the opportunity to leave an hour early.

Although every organization has different things that are valued as rewards, food is often a great motivator (e.g., pastries, pizza, or bagels). What usually makes the biggest impact, and doesn't cost a thing, is when someone personally notices and appreciates your work. It is extremely powerful when business managers express their appreciation to IT staff members. Likewise, IT managers need to do the same for members of the business community who are really making a difference. Keep your eyes open for opportunities to thank one another.

**NOTE** The best reward is when the business group is able to meet or exceed its business goals. Many factors drive business success, but a data warehouse can indeed play a part. Take a moment to reflect on how the data warehouse helped when you cash your bonus check!

Sometimes it may seem as though you are just spinning your wheels and not making any tangible progress. The ability to see progress, however, is analogous to watching one's children grow and change. If you measure them every morning, you are unlikely to notice any change; but if you wait to take that measurement, then one morning you will suddenly notice growth of an inch! Similarly, you often don't notice how your children's faces are maturing until you see a major difference when looking at photographs from last summer or the previous school year.

Measuring the progress and maturity of a data warehouse is quite similar. You may not notice any change at all when you compare this week to last week, but if you stop every six months to review progress, you are likely to be pleasantly surprised. Therefore, add a reminder to your calendar to reflect on data warehouse and BI growth. After you take a moment to reflect, share your observations with others!

## Success Can Be Attained

There are plenty of data warehouse success stories. The types of benefits vary widely and depend upon many variables, including the organization's industry, the maturity of the business, its size, and the state of the marketplace.



The following examples are included to share real-world data warehouse initiatives that have had a positive impact. Specific financial results cannot be shared, but the general business scenarios are summarized. Overlay what the financials would look like if you could accomplish these objectives. Use these to spark ideas for what could be done within your organization:

- **Improved policy underwriting:** For an insurance company, policy renewals are a significant source of revenue. Using the data warehouse with policy and claims history, this company was able to identify the highest-risk customers and develop a profile for those customers. This profile was used to support getting approval for a rate change. The profile is now being used for underwriting to rate new policy quotes and for the renewal of existing policies.
- **Understanding total customer value:** There are many efforts to support customer relationship management. This organization needed to integrate data from multiple disparate divisions in order to get a complete picture of its customers. The data warehouse was built to include customer sales and demographics from each of the separate divisions. This enabled corporate marketing to identify cross-selling opportunities, which increased sales. They also were able to flag high-value customers so that the individual divisions could provide premier customer service, regardless of the customer status for that part of the business. These efforts improved overall customer satisfaction and customer retention. As a side note, the work done to develop an integrated customer dimension was used to help launch an enterprise master data management program.
- **Vendor performance management:** Purchasing data and vendor quality measurements were combined in a data warehouse. This enabled the vendor management team to identify the highest-volume vendors with the worst track record for on-time delivery and product defects. This in turn enabled the team to eliminate some of the worst offenders and to include performance standards in the new vendor contracts. With just-in-time delivery to the plants, this reduced manufacturing delays and the cost associated with defective parts.
- **Teacher supply and demand:** At both the state and the federal level, it is necessary to understand the future demand for teachers and to take appropriate steps to ensure that skilled professionals are prepared and ready to teach our children. From the state perspective, a data warehouse was built to include current student data, teacher certification details, population growth, and data about education majors still in college. These were used to identify what teachers would be needed, by grade and subject, and by geographic area.



Taking it further, this data was compared to teachers who were certified, both those currently working as teachers and those who were not but who had maintained their credentials. This was combined with the college student data to gain a realistic perspective of the true supply of teachers, which predicted a serious shortage of science teachers within the next five years, more severely in some of the rural geographic areas. In response, funds were provided for marketing to high school science students to recruit new teachers. Financial incentives were offered to science teachers willing to relocate to the areas with the greatest need. Unlike other examples that demonstrated a clear financial return, in this case the data warehouse enabled the organization to perform its core function: providing quality education to all of the children in the state.

- **Call center scheduling:** Effectively staffing a call center requires a sound understanding of calling patterns. The data warehouse was built to integrate call detail with employee payroll and scheduling data. This was used to identify call patterns and outcomes. Analysis was also done to evaluate the impact of employee experience on outcomes. Together, these were used to predict future call patterns, which were used to help call center managers ensure appropriate staffing. This resulted in a significant decrease in call wait times, an increase in customer satisfaction, and reduced staff costs for low-volume periods.

Although benefits like these are sometimes seen quickly, more often they occur over a number of years. These examples are intended to encourage you. If you were able to accomplish similar objectives, what would that mean financially? Put yourself into these situations to dream of the possibilities. A data warehouse can indeed be a useful tool for moving an organization forward and positively affecting the financial bottom line.

---

## Conclusion

The success of the data warehouse launch will result in more work. As the business gains experience with what has been delivered so far, more ideas will be generated. The expansion and growth of the data warehouse should be handled in a systematic and organized manner. Resources across the enterprise can be leveraged by creating an enterprise data warehouse team. This chapter shared several structural alternatives and possible responsibilities. It also introduced several concepts to provide a quick look at the types of things on the data warehousing industry horizon. Finally, the chapter looked at how you can measure business value and what you can do to ensure long-term success.

This book has covered a lot of ground, including planning, designing, constructing, and launching a data warehouse. The basic concepts have been shared in a way to provide you with a foundation for your own data warehouse efforts—enabling you to be more effective at communicating, understanding, and participating in all aspects of a data warehouse. The most important thread throughout this book is that there must be a strong partnership between business and IT in order to see sustainable data warehouse success. Whether you are working on the organization's first data warehouse project or the nth, take heart—it is within your reach to glean true business value from your data warehouse.



## Glossary

**Attribute** Individual data element that is represented and stored in a dimension. Each attribute contains data relating to that dimension.

**Base fact** A fact that is physically stored in the data warehouse, in a fact table.

**Business dimension** The part of a business dimensional model that diagrams the business abstraction of data elements or attributes used to describe the major categories of reference data, such as customer or product. The diagrams also show how these attributes relate to one another, including hierarchies, groupings, and drill paths. The business dimension reflects the business perspective regardless of the number of physical tables used to store the data. *See also* dimension.

**Business dimensional model (BDM)** A data model that provides a business abstraction of the data needed to support reporting and analysis. The model shows diagrams of dimensions and facts so that the details can be reviewed and discussed in business terms. It also separates the business perspective from technical implementation details.

**Business intelligence (BI)** The collection of one or more reports or analyses, using data from the data warehouse, that provide insight into the

performance of a business organization. These reports and analyses are typically interactive to enable further understanding of specific areas of interest. They are used to support business professionals in their decision-making processes.

**Business measures** The complete set of facts, base and derived, that are defined and made available for reporting and analysis.

**Conformed dimension** A dimension that is shared between two or more fact tables. It enables the integration of data from different fact tables at query time. This is a foundational principle that enables the longevity of a data warehousing environment. By using conformed dimensions, facts can be used together, aligned along these common dimensions. The beauty of using conformed dimensions is that facts that were designed independently of each other, perhaps over a number of years, can be integrated. The use of conformed dimensions is the central technique for building an enterprise data warehouse from a set of data marts.

**Conformed fact** A fact or measure whose definition is consistent across fact tables and data marts. A conformed fact, such as revenue, can be correctly added and compared across different fact tables.

**Dashboard (also called performance dashboard)** The presentation of key business measurements on a single interface designed for quick interpretation, often using graphics. The most effective dashboards are supported by a full data mart that enables drilling down into more detailed data to better understand the indicators.

**Data architecture** Describes how data is organized and structured to support the development, maintenance, and use of the data by application systems. This includes guidelines and recommendations for historical retention of the data, and how the data is to be used and accessed.

**Data cleansing** The process of verifying and correcting data using a series of business rules for validation, and specifying how to handle cases that fail the checks.

**Data dictionary** The place where information about data that exists in the organization is stored. This should include both technical and business details about each data element.

**Data element** The smallest unit of data that is named. The values are stored in a column or a field in a database.

**Data governance** The practice of organizing and implementing policies, procedures, and standards for the effective use of an organization's structured or unstructured information assets.

**Data mart** A collection of data that is organized for end user ease of use and performance. The term is typically used to describe all of the data that is loaded into a single database and used together for analysis. Data marts are often developed to meet the needs of a business group such as marketing or finance. The key to successful data marts is to create them in an integrated manner using conformed dimensions (explained in Chapter 7). It is also recommended that data be loaded into only one data mart and then shared across the organization to ensure data consistency.

**Data model** An abstraction of how individual data elements relate to each other. It visually depicts how the data is to be organized and stored in a database. A data model provides the mechanism to document and understand how data is organized.

**Data quality** Assessment of the cleanliness, accuracy, and reliability of data.

**Data warehouse** For the purposes of this book, a data warehouse is the collection of data and processes that gather data from across the organization, organize it for access, and deliver reports and analyses to the business community. Other well-known definitions of a data warehouse in the industry that are worth mentioning include the following:

(1) The complete end-to-end data warehouse and business intelligence system (DW/BI system). Although some would argue that you can theoretically deliver business intelligence without a data warehouse and vice versa, that is ill-advised from our perspective. Linking the two together in the DW/BI acronym reinforces their dependency. Independently, we refer to the queryable data in your DW/BI system as the enterprise data warehouse, and value-add analytics as BI (business intelligence) applications. (From *The Data Warehouse Lifecycle Toolkit, Second Edition*, by Ralph Kimball et al., John Wiley and Sons, 2008, pg. 602.)

(2) The data warehouse is a collection of integrated subject-oriented data bases designed to support the DSS (decision support system) function, where each unit of data is relevant to some moment in time. The data warehouse contains atomic data and lightly summarized data. (From *Building the Data Warehouse, Fourth Edition*, by Bill Inmon and Claudia Imhoff, John Wiley and Sons, 2005, pg. 495.)

... It [the DW] is the central point of data integration for business intelligence and is the source of data for the data marts, delivering a common view of enterprise data. (From *Mastering Data Warehouse Design*, by Claudia Imhoff, Nicholas Gallemmo, and Jonathon Geiger, John Wiley and Sons, 2003, pg. 400.)

**Degenerate dimension** A single attribute dimension whereby the only attribute is a reference identifier such as invoice number, P.O. number, or transaction ID. This is needed to support analysis of the individual parts of a business transaction (individual line items) and the entire business transaction (the whole purchase order).

**Derived attribute** An attribute that is created to facilitate the overall usefulness of the dimensional model. This enables different attributes to be pulled together using a single identifier that can help the technical implementation of the dimensional model. It is used when the underlying source systems do not have a data element to uniquely define this case, and is required to pull together nonrelated attributes of a junk dimension.

**Derived fact** A fact that is calculated on-the-fly and not stored in the database.

**Dimension** Major business categories of information or groupings to describe business data. Dimensions contain information used for constraining queries, report headings, and defining drill paths. Within a dimension, specific attributes are the data elements that are used as row and column headers on reports. Dimensional attributes are also considered to be reference data. When describing the need to report information by region, by week, and by month, the attributes following “by” are dimensions. Each of these would be included in a dimension.

**Dimensional model** A data model organized for the purpose of user understandability and high performance. In a relational database, a dimensional model is a star join schema characterized by a central fact table with a multi-part key. The components of this key are joined to a set of dimension tables, each defined by its own primary keys. In a multi-dimensional database, a dimensional model is the database cube.

**Dimensional modeling** A formal data modeling technique that is used to organize and represent data for analytical and reporting use. The focus is on the business perspective and the representation of data.

**Entity-relationship (E-R) model** A data model that is used to represent data in its purest form and to define relationships between different entities.

It is often the type of model used to design online transaction processing systems. *See also* normalized model.

**Extract, transform, and load (ETL)** The collection of processes that are used to prepare data for another purpose. This is typically applied to data warehousing, whereby the extract process collects data from the appropriate underlying source systems. The transformation processes perform cleansing, manipulation, and reorganization of the data in preparation for its intended use. Finally, the load processes put the data into the data structures where it is held for data delivery. While ETL processes are regularly discussed in the context of building the data warehouse, these techniques can also be used for moving and manipulating data for a variety of other purposes.

**Fact group** Part of a business dimensional model that diagrams a collection of individual facts that have identical dimensionality and level of detail (or grain).

**Factless fact table** A fact table that captures the existence of business events that do not have an associated quantitative measurement. The existence of the relationship is what is relevant.

**Facts** The fundamental measurements of the business. These are captured as specific information about a business event or transaction. They are measured, monitored, and tracked over time. Facts are typically the amounts and counts that show up as the body of reports. Facts are used for any and all calculations that are performed.

**Grain** The level of detail showing how fact data is stored and available for analysis.

**Hybrid OLAP (HOLAP)** OLAP technology that uses both multidimensional cubes and relational databases to support business functions.

**Information management** In its simplest form, this is the work associated with collecting, maintaining, applying, and leveraging data across an organization.

**Infrastructure** A basic foundation technology that all other initiatives in the organization can rely on and use. This includes basic networking services such as providing shared network drives for storing the group's files (e.g., word processing documents, spreadsheets, and presentations). The existence of some sort of computer for each user can also be considered infrastructure.

Basic networking of computers, via a local area or wireless network, is another example of infrastructure.

**Junk dimension** A dimension that brings together single attributes that may or may not have any true relationship to each other in order to simplify the model, improve query performance, and/or reduce data storage.

**Master data management (MDM)** The processes and tools to help an organization consistently define and manage core reference or descriptive data across the organization. This may involve providing a centralized view of the data to ensure that its use for all business processes is consistent and accurate.

**Multi-dimensional OLAP (MOLAP)** OLAP technology whereby the data is stored in proprietary array structures called *multi-dimensional cubes*. *See also* OLAP.

**Normalized model** A data model organized to clarify pure data relationships and targeted at gaining efficiencies in data storage and maintenance. This is used for the design of transaction processing systems. There are specific rules for normalization. Depending upon the number of rules followed (for different purposes) there are different “forms,” such as third normal form. *See also* entity-relationship model.

**Online analytical processing (OLAP)** A collection of common business analysis functions that are difficult to perform directly with SQL. Some of the specific functions that fall under the OLAP umbrella include time series comparisons, ranking, ratios, penetration, thresholds, and contribution to report or to the whole data population. Most business intelligence tools provide this type of functionality. The capabilities can be implemented in a variety of different data storage mechanisms. *See also* MOLAP, ROLAP, HOLAP.

**Online transaction processing (OLTP)** Online transaction processing (OLTP) systems are the fundamental systems used to run the business. These are also called operational systems or operational applications. They are often used as sources of data for the data warehouse.

**Operational data store (ODS)** A collection of data from operational systems, most often integrated together, that is used for some operational purpose. The most critical characteristic here is that this is used for some operational function. This operational dependency takes precedence and the ODS should not be considered a central component of the data warehousing



environment. An ODS can be a clean, integrated source of data to be pulled into the data warehousing environment.

**Query** The mechanism to get data out of a database. A query is comprised of constraints used to filter data out of the results, and defines the data elements to be included in the result set and possibly some mathematical computations, grouping, or sorting of the data.

**Relational OLAP (ROLAP)** OLAP technology that uses data is stored in relational database management systems. Data is usually organized dimensionally using a star or snowflake schema.

**Role-playing dimension** Instances of a dimension that legitimately has more than one value for a given business transaction, such as order date and shipped date. Each attribute with the dimension is uniquely identified to enable easy differentiation between the different roles, such as Order Date, Order Quarter and Shipped Date, and Shipped Quarter.

**Scorecard (or performance scorecard)** An application that helps organizations measure and align the strategic and tactical aspects of their businesses, comparing organizational and individual performance to goals and targets.

**Slowly changing dimension (SCD)** A dimension that accommodates changes to the reference data over time. Several dimensional modeling techniques are used to determine how to handle changes to the reference data stored in dimensions. This may be to retain only the current values (Type 1), to store different versions of the reference data (Type 2), or to retain one previous version of changes made to the entire dimension (Type 3).

**Snowflake schema** A variation of the star schema in which the business dimensions are implemented as a set of normalized tables. The resulting diagram resembles a snowflake.

**Source system** An operational system of records whose function it is to capture the transactions of the business. Source systems are often large online transaction processing systems, but could also be smaller departmental databases or spreadsheets that are maintained and used by members of the business community. These are the origin of the data used to build the data warehouse.

**Staging area** Place where data is stored while it is being prepared for use, typically where data used by ETL processes is stored. This may encompass

everything from where the data is extracted from its original source until it is loaded into presentation servers for end user access. It may also be where data is stored to prepare it for loading into a normalized data warehouse.

**Star schema** The implementation of a dimensional model in a relational database. The tables are organized around a single central fact table possessing a multi-part key, and each surrounding dimension table has its own primary key.

**Structured query language (SQL)** The programming language used to access data stored in a relational database.

**Technical architecture** Addresses the organization and structure of the collection of hardware and software technologies that are installed to support the development and delivery of the data warehouse.

**Third normal form (3NF)** The most common form of a normalized model. *See also* normalized model.



# Index

## A

- accessibility, data, 15–16, 142–143
- accounts receivable, 146
- ad hoc reports, 345
- aggregation rule, 186
- Agile, Inc., 72
  - architecture and, 312
  - BI and, 365–366
  - business requirements gathering
    - process and, 170–171
  - change control procedures and,
    - 130, 229
  - data cleansing and, 395
  - dimensional modeling and, 229
  - ETL production system and,
    - 338–339
  - freedom for creativity at, 130–131
  - IM and, 272–273
  - IT-business partnership of, 108
  - production data warehouse and,
    - 395
- appliance, data warehouse, 182
- architecture (system)
  - balance of, 279
  - benefits of, 278–282

- books on, 277
- choosing, 295–296
- data architect/data modeler,
  - 92–93
- data architecture, 282–297
  - bottom-up, 286–290, 296, 297
  - capture/create data layer, 285,
    - 287–288, 291
  - components of, 285–286
  - data flow, layers of, 285–286
  - data warehouse goals and, 280,
    - 283–285
  - defined, 278, 422
  - extract data layer, 285, 287, 288,
    - 291–292
  - IM and, 267–268
  - philosophical approaches to,
    - 286
  - prepare data layer, 285, 287,
    - 288–289, 292–293
  - publish data layer, 285, 287,
    - 289–290, 294–295
  - top-down, 290–297
  - use data layer, 285, 287, 290

architecture (system) (*continued*)  
  definition of, 278  
  development requirements,  
    281–282  
  Giant Company and, 311–312  
  just-in-time, 311  
  questions and, 313  
  structure/flexibility balance and,  
    310–311  
  technical architecture, 297–302  
    basics of, 298–299  
    components of, 299–300  
    defined, 278, 428  
    development environment, 298  
    functional software layer, 299  
    hardware layer, 299  
    high-level, 300–301  
    necessary knowledge about,  
      301–302  
    operating system layer, 299  
    production environment, 298  
    test environment, 298, 377  
  attribute definitions  
    customer dimension, 190–191  
    date dimension, 185  
    employee dimension, 192–193  
    time dimension, 188–189  
  attributes  
    cluster of future attributes, 225  
    defined, 179, 421  
    derived, 221, 222, 424

## **B**

base facts, 209, 421  
BDM. *See* business dimensional  
  model  
Becker, Bob, 277  
BI (business intelligence), 287–288  
  ad hoc reports, 345

  Agile, Inc. and, 365–366  
  application developers, 46, 94–95  
  applications, 290, 295  
  capabilities of, 345  
  complex analysis, 345–346  
  defined, 6, 8, 341–342, 421  
  embedded, 405–406  
  exception reports, 5, 13, 344  
    defined, 344  
    e-mail notification and, 351  
    issues with, 12  
  Giant Company and, 364–365  
  interactive reports, 344  
  operational, 406–407  
  other labels for, 342  
  parameter-driven reports, 343  
  portal, 347, 348  
  tabular reports, 343, 347  
  tools, 94  
    configuring, 347  
    metadata, defining, 347  
  without DW, 342–343  
BI solution/application  
  building blocks, 346–352  
  construction of, 354–361  
  data content, 346–347  
  delivery, 351–352  
  design process, 355–357  
  development process, 357–358  
  education and  
    learning about BI application,  
      362–363, 370  
    learning about data, 362, 370  
    resources/help, 363–364, 370  
  launching, 361–364, 369–370  
  layers, 346–352, 366  
  levels of use  
    business analysts, 352  
    casual users, 353  
    consumers of results, 353

- external users, 353–354
- mainstream users, 352
- power users, 354
- maintaining, 381
- marketing of, 361–362, 369
- navigation, 347
- presentation, 347–351
- security and, 359–360
- system controls and, 360–361
- testing, 358
- BI system, data warehouse, 287–288
- bottom-up data architecture, 286–290
- extract data layer, 285, 287, 288
- mistakes in, 296
- prepare data layer, 285, 287, 288–289
- publish data layer, 285, 287, 289–290
- success factors, 297
- bug fixes, production data warehouse, 384
- building DW projects. *See also* architecture; BI; ETL production system
  - data delivery, 341–366
  - time length for, 34–35, 37
- Building the Data Warehouse, Fourth Edition* (Inmon), 7
- business analyses
  - business data *v.*, 167–168
  - candidate, business themes and, 162
  - consolidated, 162
  - description of, 135
  - identifying, 143–144
  - samples, 144–145
    - post promotion analysis, 145
    - promotion financial results, 145
    - promotion performance tracking, 144
    - promotion planning, 144
- business analysts, 83–86
  - BI solution and, 352
  - gatekeeper danger and, 165–166
- business applications, IM and, 232
- business champion, 79, 82–83
- business community (business-IT partnership), 76–88
- business analysts, 83–86
- business champion, 79, 82–83
- business user audience, 86
- executive business sponsor, 78–80
- middle managers, 81
- project manager, 46, 86–88
- senior management, 77, 78
- business data
  - business analyses *v.*, 167–168
  - informal data sources and, 160
- business data requirements
  - consolidated, 162–163
  - description of, 136
  - documentation of, 160
  - identifying, 146
  - samples, 146–147
    - accounts payable, 146
    - accounts receivable, 146
    - general ledger, 146
    - revenue forecast, 146–147
    - vendor profiles, 147
- business dimensional model (BDM), 182–186, 347, 421. *See also* dimensional models
- business dimensions, 183–184
  - date dimension example, 183–184, 185, 187
  - defined, 421

- business integration/efficiency, 139
- business intelligence. *See* BI
- business measures, 209, 422
- Business Measures Worksheet, 209–210, 346, 347
- business performance monitoring (sample business theme), 141
- business processes
  - changes, DWs and, 24–25, 374–375
  - IM and, 232, 234–235
  - streamlining, 374–375
- business requirements (for DW projects), 9, 16, 17–18, 22–23, 133–171
- business analyses, 135, 143–145
  - consolidated, 162
  - description of, 135
  - identifying, 143–144
  - sample, 144–145
- business data requirements, 136, 145–147
  - consolidated, 162–163
  - description of, 136
  - documentation of, 160
  - identifying, 146
  - sample, 146–147
- business themes
  - broad, 135, 140–142
  - candidate business analyses and, 162
  - consolidated, 162
  - data-related, 136, 142–143
  - documentation of, 159
  - executive summary and, 161
- communication and, 149–152
- data models and, 171
- overview of, 134–138
- project charter and, 166–167
- project scope and, 166
- strategic, 135, 138–140
  - description, 135
  - identifying, 138–139
  - sample, 139
- business requirements gathering
  - process, 134–138, 153–158
- Agile, Inc., 170–171
- challenges
  - cynics and, 165
  - data integration, 151
  - developing functional specifications, 164
  - lack of requirements, 165
  - listing data elements, 164
  - moving beyond immediate data, 164
  - sifting through reports, 163
- documentation and, 158–166, 171
- Giant Company, 170
- interview sessions, 153–158
  - concluding of, 158
  - conducting, 157–158
  - documentation of, 158–166
  - executive summary and, 161
  - group, 153–154
  - individual, 153, 159–160
  - input for, 137
  - non-data warehouse requirements, 163
  - notes *v.* tapes in, 157–158
  - number of, 156
  - objective of, 161
  - preparing for, 157
  - project team participation, 154
  - setting good example in, 156
  - tips for, 154–155
  - who to include in, 155–156
- layers of, 135
- project team and, 150–152

business systems analyst, 90–91  
 business terms, IT topics in,  
   100–101, 126–128  
 business themes  
   broad, 135, 140–142  
   identifying, 140  
   sample, 140–141  
 candidate business analyses and,  
   162  
 consolidated, 162  
 data-related, 136, 142–143  
 documentation of, 159  
 executive summary and, 161  
 business user audience, 86  
 business value (DWs), 410–419  
   adjusting expectations to reality,  
     412  
   measuring, 410–412  
 businesses/companies. *See also*  
   partnerships  
   business analysis and, 56–57  
   cultures, 52–53  
   data warehouse history in, 57–58  
   dimensional modeling and,  
     205–215. *See also* dimensional  
     modeling  
   discipline, DWs and, 15, 23–24  
   initiatives of, 54–55  
   reporting environments of, 58–70  
   *running the business v. managing*  
     the business, 43, 151  
   visions for, 52  
 business-IT partnerships. *See*  
   partnerships

## C

calendar monthly report (Call  
 Center dimensional model), 199

Call Center data warehouse  
   project, 70  
   project charter, 114–116  
   project launch agenda, 124  
   project scope, 117, 118  
   statement of work, 119–120  
 Call Center dimensional model,  
   175, 186–200  
   calendar monthly report, 199  
   call dimension, 191, 193  
   date dimension, 183–184, 185,  
     187  
     attribute definitions, 185  
     diagram, 183  
   dimensions, 187–195  
   employee dimension, 191,  
     192–193  
     attribute definitions, 192–193  
     diagram, 192  
   employee task dimension, 195  
   fact groups, 196–198  
     call forecast, 198  
     calls, 196  
     time tracking, 196–198  
   index, 200  
   time dimension, 187–189  
     attribute definitions, 188–189  
     diagram, 188  
   working with, 199  
 call center scheduling (real-world  
   DW initiative), 419  
 call dimension, 191, 193  
 call forecast fact group, 198  
 call outcome dimension, 194  
 calls fact group, 196  
 candidate business analyses,  
   162  
 capacity planning, 378–379

- capture/create data layer (data architecture)
  - bottom-up approach, 285, 287–288
  - top-down approach, 291
- case studies. *See also* Agile, Inc.; Call Center data warehouse project; Giant Company Enterprise Responsibility, 280
  - spreadsheet integration
    - functionality, 307
  - Starting at Wrong End of Spectrum, 252–253
- casual users, BI solution and, 353
- celebrating progress, 416–417
- centers of excellence, 376, 400. *See also* data warehouse support team
- centralized enterprise DW team, 401
- champion. *See* business champion
- change control procedures, data warehouse projects and, 125–126, 167, 169, 336, 371
  - Agile, Inc. and, 130, 229
- charts/graphs, BI presentation and, 348
- CIO/IT executive sponsor, 89
- cluster of future attributes, 225
- communication. *See also* partnerships
  - business requirements and, 149–152
  - business-IT partnerships and, 31, 96, 99–103
  - data warehouse projects and, 20, 31, 412
  - IT topics in business terms, 100–101, 126–128
  - key messages and, 99–100
  - meeting preparation and, 101–102
  - multi-layered approach to, 20
  - presentation tips, 102
  - proactive, ETL production system and, 336–337
- companies. *See* businesses/companies
- compilation, data dictionary and, 262
- complex analysis, BI and, 345–346
- configuring BI tools, 347
- conformed dimensions, 200–202
  - customer type sample values, 201–202
- defined, 422
- MDM and, 240
- conformed facts, 202, 422
- consistent management reporting, 139
- consolidated business analyses, 162
- consolidated business data requirements, 162–163
- consolidated business themes, 162
- consolidated requirements
  - documentation, 161–163
- consumers of results, BI solution and, 353
- corporate asset, data as, 21, 231–274. *See also* information management
  - Corporate Information Factory, Second Edition* (Inmon, Imhoff, and Sousa), 290
- costs, for data warehouses, 33–34, 120–122
- cube, 181
- current data environment, IM and, 264–267



customer data example  
     (information management),  
     235–239  
 improved customer data  
     processes, 237–238  
 integrated customer data flow,  
     238–239  
 non-integrated customer data  
     flow, 236  
 customer demographic junk  
     dimension diagram, 221  
 customer dimension  
     Call Center, 189–191  
     attribute definitions, 190–191  
     diagram, 190  
     sample, 179–180  
 customer relationship  
     management (real-world DW  
     initiative), 418  
 customer service (sample business  
     theme), 140  
 customer type sample values,  
     201–202  
 cynics, business requirements  
     gathering process and, 165

## D

dashboards (performance  
     dashboards)  
     BI presentation and, 349–350  
     defined, 422  
     one-off reports and, 36  
 data. *See also* information  
     management  
     accessibility, DWs and, 15–16,  
     142–143  
     characteristics of, 160  
     as corporate asset, 21, 231–274  
     historical, quality of, 251–253

issues/problems, 11–12, 55–56  
 master, 240, 242. *See also* master  
     data management  
     timeliness, 12, 14  
     unstructured, 408–409  
     usefulness, 12  
 data accuracy (data-related  
     business theme), 142  
 data analysis, source, 177–178  
 data architect/data modeler,  
     92–93  
 data architecture, 282–297  
     bottom-up, 286–290  
     extract data layer, 285, 287, 288  
     mistakes in, 296  
     prepare data layer, 285, 287,  
     288–289  
     publish data layer, 285, 287,  
     289–290  
     success factors, 297  
 components of, 285–286  
 data flow, layers of, 285–286  
 data warehouse goals and, 280,  
     283–285  
 defined, 278, 422  
 IM and, 267–268  
 philosophical approaches to, 286  
 top-down, 290–297  
     capture/create data layer, 291  
     extract data layer, 291–292  
     mistakes in, 296  
     prepare data layer, 292–293  
     publish data layer, 294–295  
     success factors, 297  
     use data layer, 285, 287, 290  
 data cleansing, 242, 249, 250  
 Agile, Inc. and, 395  
 defined, 422  
 reporting and, 254  
 at the source, 253–254

- data content (BI solution/application), 346–347
- data definitions, 184
- data delivery, 341–366. *See also* BI
- data dictionary, 259–264
  - accessing, 263
  - application, 259–261
  - capturing information for
    - compilation, 262
    - direct documentation, 262
    - ghostwriter, 263
    - interview, 262
  - columns in, 260
  - defined, 422
  - implementation of, 259–264
  - maintaining, 263–264
  - populating, 261–263
- data element, 422
- data environment, IM and, 264–267
- data governance
  - defined, 233, 423
  - IM and, 232, 233, 243, 423
- data governance committee, 254, 268, 270–271, 272
- data integration
  - integrity of, 254–256
  - issues with, 12, 151
  - order entry system problem, 101
  - paths, 256
- data management, 15
- data marts, 294–295, 318
  - defined, 6, 8, 423
- data modeler/data architect, 92–93
- data models, 9, 10. *See also*
  - dimensional modeling
  - business requirements and, 171
  - defined, 6–7, 423
  - dimension modeling and, 7
  - 3NF and, 176, 428
- data ownership, 243–248
  - challenges with, 247–248
  - owners in
    - identification of, 244–245
    - responsibilities of, 246–247
  - tracking, 244–245
- data profiling, 178, 232, 234, 249–250
- data quality
  - definitions, 234, 248, 423
  - examples, 257–258
    - grocery checkout scanners, 257
    - public education evaluation, 257–258
  - high, 85, 232, 235, 339
  - historical data and, 251–253
  - IM and, 232, 234, 248–259, 423
  - improvement of, 256–257
  - measuring, 250–251
  - problems, 12
    - symptoms of, 248–249
  - value of, 258–259
- data sources
  - business analyses-data sources matrix, 167–168
  - informal, 160
- data stores (ETL), 288–289, 290, 291, 292, 293, 294
- data warehouses (DWs). *See also*
  - data warehouse projects;
  - production data warehouse
  - benefits of, 32–33
  - Building the Data Warehouse, Fourth Edition*, 7
  - business process changes and, 24–25, 374–375
  - business value and, 410–419

- business-IT partnerships and. *See* partnerships
- costs for, 33–34, 120–122
- data accessibility and, 15–16, 142–143
- data preparation and, 292–294
- The Data Warehouse Lifecycle Toolkit, Second Edition*, 7, 218, 277, 286, 423
- definitions, 4, 7–8, 423–424
- design and development
  - sequence, 8–11
- education about, 25, 36
- encouraging usage of, 372–374
- environment, 4–6
- expansion/growth and, 397–400
- failures, 22–30, 39
- FAQs, 31–48
- future of, 405–410
- global perspective and, 18
- goals of, 280, 283–285
- IM and, 232, 239–240
- infrastructure and, 300
- long-term success for, 397–420
- master data and, 242
- Mastering Data Warehouse Design*, 7, 8, 277, 290, 294, 424
- narrow focus and, 25–26
- OLTPs *v.*, 4, 5
- operational application systems and, 43–45
- organizational assessment and. *See* businesses/companies
- organizational discipline and, 15, 23–24
- problems with, 40–41
- production, 369–396
- promises of, 15–16
- rationale for, 11–12
- real-time, 407–408
- single version of truth and, 15, 41–43, 280
- stalled
  - causes for, 385–388
  - jump-starting of, 388–393
- success stories, 417
- successful
  - ensuring, 36
  - examples of, 417–419
  - factors for, 16–22
  - signs of, 38
- successful organizations and, 27
- technology and. *See* technology
- technology products/tools and, 302–310
- uses of, 12–14, 24–25
- value of, 12–15
- data warehouse appliance, 182
- data warehouse initiatives (real world), 418–419
- call center scheduling, 419
- customer relationship
  - management, 418
- policy underwriting, 418
- teacher supply and demand, 418–419
- vendor performance
  - management, 418
- data warehouse manager, 89–90
- data warehouse projects. *See also* Agile, Inc.; Call Center data warehouse project; Giant Company
- approval process, 116–117, 122
- architecture. *See* architecture
- building. *See also* architecture; BI; ETL production system
- data delivery, 341–366
- time length for, 34–35, 37

- data warehouse projects
    - (*continued*)
    - business community
      - roles/responsibilities, 76–88
    - business requirements, 9, 16, 17–18, 22–23, 133–171
    - business analyses, 135, 143–145
    - business data requirements, 136, 145–147, 160, 162–163
    - business themes, 135, 136, 140–142, 159, 161, 162
    - communication and, 149–152
    - data models and, 171
    - overview of, 134–138
    - project charter and, 166–167
    - project scope and, 166
    - strategic, 135, 138–140
  - change control procedures, 125–126, 167, 169, 336, 371
  - communication and, 20, 31, 412.
    - See also* partnerships
  - dangers of gatekeepers in, 165–166
  - defining, 111–112, 131
  - ETL production system. *See* ETL production system
  - expectations and, 19, 128–129
  - finishing, 369–371
  - issue tracking, 124–125
  - IT requirements for, 147–149
  - IT roles/responsibilities, 88–95
  - launching, 122–124, 369–396
  - lost institutional knowledge and, 29–30
  - managing, 124–129, 369–396
  - personnel in, 45–46
  - post-implementation review, 370–371
  - project charter for, 112–117
  - project deadline trap, 23
  - project scope, 19, 23, 117, 118
  - setting up, 47–48, 111–131
  - staff members for, 28–29
  - statement of work, 117–120
  - tracking DW use, 372
  - vision/strategy for, 46–47
- data warehouse support team, 106–107, 376
  - data warehouse team, enterprise, 400–405
    - centralized, 401
    - creating, 400–403
    - funding, 404–405
    - responsibilities, 403–404
    - virtual, 401–403
  - data warehouse/business intelligence system, 287–288
  - data warehousing industry, growth of, 11
  - Data Warehousing Institute, 41
  - databases
    - design, dimensional models and, 228
    - dimensional models and, 181–182
    - multi-dimensional, 181
    - proprietary, 182
    - relational, 181
  - data-related business themes, 136, 142–143
  - The Data Warehouse Lifecycle Toolkit, Second Edition* (Kimball et al.), 7, 218, 277, 286, 423
  - date dimension (Call Center), 183–184, 185, 187
    - attribute definitions, 185
    - diagram of, 183
  - default aggregation rule, 346
  - defining BI tool metadata, 347

- degenerate dimensions, 191, 219–221, 424
- defined, 424
- Purchase Order, 220
- derived attributes, 221, 424
  - defined, 424
  - within Policy dimension, 222
- derived facts, 209, 424
- development environment
  - (technical architecture), 298
- dials, BI presentation and, 349
- dictionary. *See* data dictionary
- dimension connectors, 224
- dimension table design, 226–227
- dimensional attributes. *See* attributes
- dimensional modeling, 175–230
  - advanced concepts, 216–225
  - Agile, Inc. and, 229
  - business participation in, 205–215
  - conformed dimensions and, 200–202
    - customer type sample values, 201–202
    - defined, 422
  - conformed facts and, 202, 422
  - data models and, 7
  - defined, 7, 424
  - degenerate dimensions and, 191, 219–221, 424
  - dimensions and. *See* dimensions
  - ease of use in, 176–177
  - factless fact tables and, 218–219, 425
  - Giant Company and, 229
  - goals of, 176–177
  - guidelines, 202–205
    - enterprise design, 204
    - for single dimension, 202–203
    - for single fact group, 203
  - multiple hierarchies and, 216, 217
  - notation for, 224–225
    - dimension connectors, 224
    - summary, 226
  - purpose of, 176–177
  - query performance and, 177
  - role-playing dimensions and, 222–224, 427
  - SCDs and, 216–218, 427
  - slowly changing dimensions in, 216–218
  - source data analysis and, 177–178
- dimensional models. *See also* Call Center dimensional model
  - adding fact groups to, 215
  - BDM, 182–186, 347, 421
  - building blocks of, 178
  - database design and, 228
  - databases and, 181–182
  - defined, 178, 424
  - development of, 205–215
    - brainstorming framework, 206
    - Business Measures Worksheet, 209–210, 346, 347
    - business reviews and, 213–214
    - completing/fleshing out, 211–213
    - data elements and, 212
    - first draft, 205
    - initial dimensions, 206–207
    - initial fact groups, 207–208
    - issues in, 211–212
    - logging questions/issues, 208–209
  - modeling checklist, 206
  - modeling sessions, 205–206

- dimensional models (*continued*)
  - refining, 213
  - source to target data map, 211
- diagramming, 182
- documentation of, 208, 212
- enhancing dimensions in, 215
- expansion of, 215
- facts and. *See* facts
- implementing, 181–182
- presenting, ideas for, 213–214
- Sales Results Report, 181
- translating, 226–228

dimensions, 178–180

- business, 421
- call, 191, 193
- call outcome, 194
- conformed, 200–202
  - customer type sample values, 201–202
- defined, 422
- MDM and, 240

customer, 179–180

date (Call Center), 183–184, 185, 187

- attribute definitions, 185
- diagram of, 183

defined, 424

degenerate, 191, 219–221, 424

dimensional modeling and, 180–181

employee (Call Center), 191, 192–193

- attribute definitions, 192–193
- diagram, 192

employee task, 195

enhancing, dimensional models and, 215

junk, 426

Policy, 222

role-playing, 222–224, 427

- slowly changing, 216–218, 427
- time (Call Center), 187–189
  - attribute definitions, 188–189
  - diagram, 188
- types of, 178, 179

documentation, 158–166, 171

- business data requirements and, 160
- business themes and, 159
- consolidated business
  - requirements and, 161–163
- data dictionary and, 262
- dimensional models and, 208, 212

drill-down capabilities, 344

DWs. *See* data warehouses

## E

education

- BI solution/application
  - learning about BI application, 362–363, 370
  - learning about data, 362, 370
  - resources/help, 363–364, 370
- DWs and, 25, 36
- public education evaluation, data
  - quality and, 257–258
- sample IT business requirement, 148–149

embedded business intelligence, 405–406

employee dimension (Call Center), 191, 192–193

- attribute definitions, 192–193
- diagram, 192

employee task dimension, 195

enhancement requests, 382, 383, 397, 412

enterprise DW resources, managing, 400–405

- enterprise DW team, 400–405
  - centralized, 401
  - creating, 400–403
  - funding, 404–405
  - responsibilities, 403–404
  - virtual, 401–403
- enterprises. *See also*
  - businesses/companies
  - data warehouse team, 107
  - defined, 280
  - design guidelines, dimensional modeling and, 204
  - Enterprise Responsibility case study, 280
- entity-relationship (E-R) model, 176, 424. *See also* normalized model
- E-R model. *See* entity-relationship model
- ETL data stores, 288–289, 290, 291, 292, 293, 294
- ETL developers, 93–94
- ETL (extract, transform, and load)
  - process, 93, 251, 315–339
  - defined, 6, 425
  - extraction in, 318
  - load in, 322
  - project deadline trap and, 23
  - transformation in, 319–322
- ETL production system
  - Agile, Inc. and, 338–339
  - business data in, manual inclusion of, 333
  - business role in, 323–329
    - design process, 324
    - development support, 326
    - test plan, 325–326
  - characteristics of, 316
  - construction, 317
  - continued business participation, 335
  - design, 316–317
  - development steps for, 316–317
  - encouragement/support and, 334–335
  - flaws in current systems, 330–331
  - functionality, 317–323
  - Giant Company and, 337–338
  - long-term solutions, 332–333
  - maintaining, 380–381
  - new business rules, 331–332
  - proactive communication and, 336–337
  - requirements, 316
    - data reality and, 329–330
  - testing, 317, 326–327
  - time to build, 327–329
  - tracking progress of, 333–334
  - uncertainty and, 335–336
- exception reports, 5, 13
  - defined, 344
  - e-mail notification and, 351
  - issues with, 12
- executive business sponsor, 78–80
- executive steering committees, 104–106
- executive summary (of business requirements), 161
  - business theme summary, 161
  - critical
    - findings/recommendations, 161
  - interview process, 161
  - purpose of requirements, 161
- expansion/growth, DWs and, 397–400
- expectations, DW projects and, 19, 128–129
- external consulting, 80, 97–98

- external data access (data-related business theme), 143
- external users, BI solution and, 353–354
- extract data layer (data architecture)
  - bottom-up approach, 285, 287, 288
  - top-down approach, 291–292
- extract, transform, and load process. *See* ETL process
- extraction (in ETL process), 318

## F

- fact groups, 184–186
  - adding, to dimensional models, 215
- Call Center dimensional model, 196–198
  - call forecast, 198
  - calls, 196
  - time tracking, 196–198
- defined, 425
- Retail Sales, 185, 186
- translating, 227–228
- factless fact tables, 218–219, 425
- facts
  - base, 209, 421
  - conformed, 202, 422
  - defined, 180, 425
  - derived, 209, 424
  - dimensional models and, 178, 180–181
- failures (DWs), 22–30, 39
- FAQs, for DWs, 31–48
- foreign key, 228
- functional software layer (technical architecture), 299

- functional specifications, developing, 164
- future, of data warehouses, 405–410

## G

- Galemmo, N., 277, 290, 294, 424
- gatekeepers
  - architecture group as, 311
  - dangers of, DW projects and, 165–166
- gathering process. *See* business requirements gathering process
- Geiger, J., 277, 290, 294, 424
- genealogy, technical product, 306
- general ledger, 146
- ghostwriter, data dictionary and, 263
- Giant Company, 71–72
  - architecture at, 311–312
  - BI at, 364–365
  - business requirements gathering process and, 170
  - dimensional modeling and, 229
  - ETL production system and, 337–338
  - IM and, 272
  - IT-business partnership of, 107–108
  - production data warehouse and, 394–395
  - structured projects with, 129–130
- global perspective, DWs and, 18
- glossary, 421–428
- grain, 183, 425
- graphs/charts, BI presentation and, 348
- grocery checkout scanners, data quality and, 257



group interviews, 153–154  
 growth/expansion, DWs and,  
 397–400

## H

hardware layer (technical  
 architecture), 299  
 hierarchies, 183  
   multiple, 216, 217  
 high-level technical architecture,  
 300–301  
 high-quality data, 85, 232, 235, 339.  
   *See also* data quality  
 high-value customers, 13, 418  
 historical data, quality of, 251–253  
 HOLAP (hybrid OLAP), 425  
 hybrid OLAP. *See* HOLAP

## I

IM. *See* information management  
 Imhoff, Claudia, 7, 8, 277, 290, 294,  
 423, 424  
 implementation (building  
 database). *See* ETL process  
 index, Call Center dimensional  
 model, 200  
 individual interviews, 153,  
 159–160  
 industry innovation, monitoring,  
 409–410  
 informal data sources, 160  
 information management (IM),  
 231–274, 425  
   Agile, Inc. and, 272–273  
   beyond data warehouses,  
   239–240  
   business applications and, 232  
   business processes and, 232,  
   234–235

catalyst for, 264  
 current data environment and,  
 264–267  
 customer data example, 235–239  
   improved customer data  
     processes, 237–238  
   integrated customer data flow,  
   238–239  
   non-integrated customer data  
     flow, 236  
 data architecture and, 267–268  
 data dictionary, 259–264  
   accessing, 263  
   application, 259–261  
   columns in, 260  
   compilation and, 262  
   defined, 422  
   direct documentation and, 262  
   ghostwriter and, 263  
   implementation of, 259–264  
   interview and, 262  
   maintaining, 263–264  
   populating, 261–263  
 data governance, 232, 233, 243,  
 423  
 data ownership, 243–248  
   challenges with, 247–248  
   identification of owners,  
   244–245  
   responsibilities of owners,  
   246–247  
   tracking, 244–245  
 data profiling, 178, 232, 234,  
 249–250  
 data quality, 232, 234, 248–259,  
 423  
   definitions, 234, 248, 423  
   grocery checkout scanners, 257  
   high, 85, 232, 235, 339  
   historical data and, 251–253

- information management (IM),
  - (*continued*)
  - improvement of, 256–257
  - measuring, 250–251
  - problems, 12, 248–249
  - public education evaluation, 257–258
  - value of, 258–259
- defined, 232
- DWs and, 232, 239–240
- existing practices of, 266–267
- facets of, 232
- getting started with, 264–273
- Giant Company and, 272
- MDM and, 232, 233, 240–243, 426
  - conformed dimensions and, 240
  - defined, 233, 426
- overview of, 232–235
- questions to ask, 273
- strategy
  - development, 268–270
  - revising, 271
- sustainable process and, 270–271
- infrastructure. *See also* architecture
  - data warehouse and, 300
  - defined, 300, 425
- Inmon, W. H., 7, 8, 63, 277, 286, 290, 423
- innovations, in technology, 409–410
- interactive reports, 344
- interview, data dictionary and, 262
- interview sessions (business requirements gathering process), 153–158
  - concluding of, 158
  - conducting, 157–158
  - documentation of, 158–166, 171
  - executive summary and, 161
  - group, 153–154
  - individual, 153, 159–160
  - input for, 137
  - non-data warehouse
    - requirements, 163
  - notes *v.* tapes in, 157–158
  - number of, 156
  - objective of, 161
  - organizational motivation and, 150–151
  - preparing for, 157
  - project team participation, 154
  - setting good example in, 156
  - tips for, 154–155
  - who to include in, 155–156
- issue tracking, 124–125
- IT community
  - IT business requirements (for DW projects), 147–149
  - education, 148–149
  - support, 149
  - system usage and tracking, 148
- jargon and, 77
- roles (business-IT partnerships)
  - BI application developers, 46, 94–95
  - business systems analyst, 90–91
  - CIO/IT executive sponsor, 89
  - data modeler/data architect, 92–93
  - data warehouse manager, 89–90
  - ETL developers, 93–94
  - source system analyst, 91–92
- technical architecture and. *See* technical architecture
- technical topics in business terms, 100–101, 126–128
- ivory tower team, 402

**J**

jargon, 77  
 joint prioritization process,  
   166–169, 177  
   business analyses-data source  
     matrix, 167–168  
   implementation alternatives in,  
     167–168  
   setting priorities, 168–170  
 junk dimensions  
   customer demographic, 221  
   defined, 426  
 just-in-time architecture, 311

**K**

keys  
   foreign, 228  
   primary, 228  
 Kimball, Ralph, 7, 277, 423

**L**

launching  
   BI solution/application, 361–364,  
     369–370  
   improved, 393–394  
   data warehouse projects,  
     122–124, 369–396  
   improved, 393–394  
 learning. *See* education  
 load (in ETL process), 322  
 long-term success for DWs,  
   397–420  
 lost institutional knowledge,  
   29–30

**M**

mainstream users, BI solution and,  
   352

management reporting, 43  
   consistent, 139  
 managing data warehouse  
   projects, 124–129, 369–396  
 managing enterprise DW  
   resources, 400–405  
*managing the business v. running*  
   the business, 43, 151  
 maps, BI presentation and, 348  
 marketing hype, technology  
   products and, 308–309  
 marketing, of BI  
   solution/application, 361–362,  
     369  
 master data, 240  
   DW and, 242  
 master data management (MDM),  
   240–243. *See also* information  
   management  
   conformed dimensions and, 240  
   defined, 233, 426  
*Mastering Data Warehouse Design*  
   (Imhoff, et al.), 7, 8, 277, 290,  
   294, 424  
 matrix, business analyses-data  
   source, 167–168  
 MDM. *See* master data  
   management  
 middle managers, 81  
 models/modeling. *See also*  
   dimensional modeling  
   BDM, 182–186, 347, 421  
   data, 9, 10  
   business requirements and, 171  
   defined, 6–7, 423  
   dimension modeling and, 7  
   3NF and, 176, 428  
   E-R, 176, 424  
   normalized, 176, 426

- MOLAP (multi-dimensional OLAP), 426
- monitoring industry innovation, 409–410
- monitoring performance, 378–379
- multi-dimensional databases, 181
- multi-dimensional OLAP. *See* MOLAP
- multiple hierarchies, 216, 217
- Mundy, Joy, 277

## N

- narrow focus, DWs and, 25–26
- navigation (BI solution/application), 347
- 90-day project, 37
- normalized model, 176, 426. *See also* entity-relationship model
- null, 262

## O

- ODS (operational data store), 408, 426
- OLAP (online analytical processing), 426. *See also* HOLAP; MOLAP; ROLAP
- OLTPs (online transaction processing systems), 426
  - DWs *v.*, 4. 5
- one version of truth, 15, 41–43, 280
- one-off reports, 36
- online analytical processing. *See* OLAP
- online transaction processing systems. *See* OLTPs
- operating system layer (technical architecture), 299
- operational application systems, DWs and, 43–45

- operational business intelligence, 406–407
- operational data store (ODS), 408, 426
- operational reports, 43
- order entry system data
  - integration problem, 101
- organizations. *See* businesses/companies
- ownership of data. *See* data ownership

## P

- parameter-driven reports, 343
- partnerships (business-IT), 9, 17, 75–109
  - Agile, Inc., 108
  - beyond projects, 104–107
  - business community
    - roles/responsibilities, 76–88
  - business analysts, 83–86
  - business champion, 79, 82–83
  - business user audience, 86
  - executive business sponsor, 78–80
  - middle managers, 81
  - project manager, 46, 86–88
  - senior management, 77, 78
- communication in, 31, 96, 99–103
- definition of, 75–76
- external consulting, 80, 97–98
- Giant Company, 107–108
- importance of, 36, 420
- IT roles, 88–95
  - BI application developers, 46, 94–95
  - business systems analyst, 90–91
  - CIO/IT executive sponsor, 89

- data modeler/data architect, 92–93
- data warehouse manager, 89–90
- ETL developers, 93–94
- source system analyst, 91–92
- lack of talking in, 152
- strength of, 76
- tips for, 95–99
- performance dashboards. *See* dashboards
- performance monitoring, 378–379
- performance scorecards. *See* scorecards
- Policy dimension, 222
- policy underwriting (real-world DW initiative), 418
- portal, BI, 347, 348
- post promotion analysis, 145
- post-implementation review, DW projects and, 370–371
- power users, BI solution and, 354
- prepare data layer (data architecture)
  - bottom-up approach, 285, 287, 288–289
  - top-down approach, 292–293
- presentation (BI solution/application), 347–351
- presentation servers, 289–290, 317, 318
- presentation tips, 102. *See also* communication
- primary key, 228
- prioritization process, 166–169, 177
  - business analyses-data source matrix, 167–168
  - implementation alternatives in, 167–168
  - setting priorities, 168–170
- production data warehouse, 369–396. *See also* data warehouse projects
  - Agile, Inc. and, 395
  - aligning DW objectives with business goals, 391–392
  - bug fixes, 384
  - Giant Company and, 394–395
  - maintaining, 380–381
  - maintaining environment, 376–379
  - performance
    - monitoring/capacity planning and, 378–379
  - staffing production activities, 376
  - stalled
    - causes for, 385–388
    - jump-starting of, 388–393
  - tracking questions/problems, 382–384
- production environment (technical architecture), 298
- production system. *See* ETL production system
- profiling data. *See* data profiling
- profitability analysis (sample business theme), 141
- progress, celebrating, 416–417
- project briefings, 103
- project charter, 112–117
  - business requirements and, 166–167
- Call Center DW, 114–116
- prioritization session and, 166–169, 177
- sections of, 113, 116
- project deadline trap, 23
- project manager, 46, 86–88

- project scope, 19, 23
  - business requirements and, 166
  - Call Center DW project and, 117, 118
  - prioritization session and, 166–169, 177
- project teams
  - business requirements gathering process and, 150–152
  - interview session (requirements gathering process) and, 154
  - members of, 46
  - tips for, 98–99
- promotion financial results, 145
- promotion performance tracking, 144
- promotion planning, 144
- proprietary databases, 181
- public education evaluation, data quality and, 257–258
- publish data layer (data architecture)
  - bottom-up approach, 285, 287, 289–290
  - top-down approach, 294–295
- Purchase Order degenerate dimension, 220

## Q

- queries. *See also* SQL
  - defined, 342, 427
  - dimensional modeling and, 177
  - source data analysis and, 177–178
  - SQL and, 342

## R

- real-time data warehousing, 407–408

- real-world data warehouse
  - initiatives, 418–419
- call center scheduling, 419
- customer relationship management, 418
- policy underwriting, 418
- teacher supply and demand, 418–419
- vendor performance management, 418
- reference data. *See* attributes
- relational databases, 181
- relational OLAP. *See* ROLAP
- reports. *See also* BI
  - ad hoc, 345
  - exception, 5, 13
  - BI and, 344
  - defined, 344
  - e-mail notification and, 351
  - issues with, 12
  - interactive, 344
  - parameter-driven, 343
  - tabular, 343
- report cards. *See* scorecards
- reporting environments, 58–70
- request for information (RFI), 305
- request for proposal (RFP), 305
- requirements for DW projects. *See* business requirements
- Retail Sales fact group, 185, 186
  - names/definitions for, 186
- return on investment (ROI), 31, 62
- revenue forecast (sample business data requirement), 146–147
- RFI (request for information), 305
- RFP (request for proposal), 305
- roadblocks. *See* failures
- ROI (return on investment), 31, 62
- ROLAP (relational OLAP), 427

role-playing dimensions, 222–224, 427

Ross, Margy, 277

*running the business v. managing the business*, 43, 151

## S

Sales Results Report (dimensional model), 181

SCDs (slowly changing dimensions), 216–218, 427

scope. *See* project scope

scorecards (performance scorecards)

BI presentation and, 348–349

defined, 427

SDLC (system development life cycle), 89

security, BI solution/application and, 359–360

senior management, 77, 78

silo reporting environments, 12, 272

single source of data (data-related business theme), 142

single version of truth, 15, 41–43, 280

slowly changing dimensions. *See* SCDs

SME (subject matter expert), 244

snowflake schema, 427

source data analysis, 177–178

source system analyst, 91–92

source systems, 6, 427

source to target data map, 211

Sousa, Ryan, 290

spreadsheet integration

functionality (case study), 307

SQL (structured query language)

defined, 428

OLAP and, 426

queries and, 342

staging area, 427–428

staging data store, 291, 292

stalled data warehouses

causes for, 385–388

jump-starting of, 388–393

star schema, 181, 428

Starting at Wrong End of

Spectrum (case study), 252–253

statement of work, 117–120

stopgap functionality, 307

stovepipe applications, 279

strategic business requirements, 135, 138–140

description, 135

identifying, 138–139

sample, 139

structured query language. *See* SQL

subject matter expert (SME), 244

supplier performance analysis (sample business theme), 141

support (sample IT business requirement), 149

support team, data warehouse, 106–107, 376

system architecture. *See* architecture

system controls, BI

solution/application and, 360–361

system development life cycle (SDLC), 89

system usage and tracking (sample IT business requirement), 148

systems community. *See* IT community

**T**

- tabular reports, 343, 347
- teacher supply and demand (real-world DW initiative), 418–419
- technical architecture, 297–302
  - basics of, 298–299
  - components of, 299–300
  - defined, 278, 428
  - development environment, 298
  - functional software layer, 299
  - hardware layer, 299
  - high-level, 300–301
  - necessary knowledge about, 301–302
  - operating system layer, 299
  - production environment, 298
  - test environment, 298, 377
- technology (for data warehouses)
  - choices, viability of, 39–40
  - fix, 27–28
  - industry innovation, 409–410
  - keeping up with, 376–378
  - leveraging, 21–22
- technology products/tools, 302–310
  - best of breed, 303
  - end-to-end solutions, 303
  - genealogy, 306
  - marketing hype and, 308–309
  - not buying, 304
  - references for, 309–310
  - selection process, 304–306
  - spreadsheet integration
    - functionality (case study), 307
  - trends, 303
  - value/pricing and, 306–307
- testing
  - BI solution/application, 358
  - DW management and, 377
  - ETL production system, 317, 325–327
  - test environment (technical architecture), 298, 377
- third normal form (3NF), 176, 428
- Thorntwaite, Warren, 277
- 3NF (third normal form), 176, 428
- time dimension (Call Center), 187–189
  - attribute definitions, 188–189
  - diagram, 188
- time tracking fact group, 196–198
- timeliness of data, 12, 14
- tools, technology. *See* technology products/tools
- top-down data architecture, 290–297
  - capture/create data layer, 291
  - extract data layer, 291–292
  - mistakes in, 296
  - prepare data layer, 292–293
  - publish data layer, 294–295
  - success factors, 297
- tracking
  - data ownership, 244–245
  - data warehouse use, 372
  - ETL production system, 333–334
  - issue tracking, 124–125
  - promotion performance tracking, 144
  - questions/problems, production DW and, 382–384
  - system usage and tracking (sample IT business requirement), 148
  - time tracking fact group, 196–198
- transformation (in ETL process), 319–322
- translating dimensional models, 226–228



translating fact groups, 227–228  
truth, single version of, 15, 41–43,  
280

## U

unstructured data, 408–409  
use data layer (data architecture),  
285, 287, 290

## V

value, business, 410–419  
vendor performance management  
(real-world DW initiative), 418  
vendor profiles (sample business  
data requirement), 147  
virtual enterprise DW team,  
401–403

## An ideal guide for the non-technical professional eager to learn more about data warehousing

**A successful data warehouse project can provide immense value for business enterprises or other organizations. Building and maintaining a data warehouse demands the combined efforts of both IT and non-technical personnel. While there are plenty of resources aimed at the technology professionals who design and build data warehouses, there has to date been no useful guide written for a non-technical audience. This book fills that void and serves as an ideal resource for business and IT managers and others from the non-IT side who want to do their part to ensure data warehousing success.**

**This helpful book provides a solid introduction to the fundamentals of data warehousing. The author details**

**each step of a data warehouse project, and provides a clear explanation of what's involved in efficiently building a data warehouse and what must be done to deliver the data. You'll examine the business management of a data warehouse and discover essential methods for cultivating a strong partnership between the business and IT elements of your organization. You can use this knowledge to be more effective when sharing your requirements and concerns during a project.**

***A Manager's Guide to Data Warehousing* explains what you need to create your data warehouse and establish long-term success. The book covers:**

- The most common factors for ensuring data warehousing success and the roadblocks that can prevent it
- How to ensure that business and technical staff have a common understanding of the data warehouse project

- How to effectively communicate your business requirements for the data warehouse
- The tools you need to make certain that data is organized and can be delivered as needed
- Ways to deploy the data warehouse and ensure sustainable success

**LAURA L. REEVES**, coauthor of *The Data Warehouse Lifecycle Toolkit*, has over 23 years of experience in end-to-end data warehouse development focused on developing comprehensive project plans, collecting business requirements, designing business dimensional models and database schemas, and creating enterprise data warehouse strategies and data architectures.



Wiley Computer Publishing

**Timely. Practical. Reliable.**

Visit our Web site at [www.wiley.com/compbooks/](http://www.wiley.com/compbooks/)