GLOBAL EDITION

Business Intelligence, Analytics, and Data Science

A Managerial Perspective

FOURTH EDITION

Ramesh Sharda • Dursun Delen • Efraim Turban

Pearson

FOURTH EDITION
GLOBAL EDITION

BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE:

A Managerial Perspective

Ramesh Sharda
Oklahoma State University

Dursun Delen
Oklahoma State University

Efraim Turban
University of Hawaii

With contributions to previous editions by

J. E. Aronson
The University of Georgia

Ting-Peng Liang
National Sun Yat-sen University

David King
JDA Software Group, Inc.

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

VP Editorial Director: Andrew Gilfillan
Senior Portfolio Manager: Samantha Lewis
Content Development Team Lead: Laura Burgess
Content Developer: Stephany Harrington
Program Monitor: Ann Pulido/SPi Global
Editorial Assistant: Madeline Houpt
Project Manager, Global Edition: Sudipto Roy
Acquisitions Editor, Global Edition: Tahnee Wager
Senior Project Editor, Global Edition: Daniel Luiz
Managing Editor, Global Edition: Steven Jackson
Senior Manufacturing Controller, Production, Global Edition: Trudy Kimber
Product Marketing Manager: Kaylee Carlson
Project Manager: Revathi Viswanathan/Cenveo Publisher Services
Text Designer: Cenveo® Publisher Services
Cover Designer: Lumina Datamatics, Inc.
Cover Art: kentoh/Shutterstock
Full-Service Project Management: Cenveo Publisher Services
Composition: Cenveo Publisher Services

Pearson Education Limited
KAO Two
KAO Park
Harlow
CM17 9NA
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com

Authorized adaptation from the United States edition, entitled Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th edition, ISBN 978-0-13-463328-2, by Ramesh Sharda, Dursun Delen, and Efraim Turban, published by Pearson Education © 2018.

ISBN 10: 1-292-22054-6
ISBN 13: 978-1-292-22054-3

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

10 9 8 7 6 5 4 3 2 1
14 13 12 11 10

Typeset in ITC Garamond Std-Lt by Cenveo Publisher Services.

Printed and bound by Vivar, Malaysia.

Brief Contents

Preface 19

About the Authors 25

Chapter 1 **An Overview of Business Intelligence, Analytics, and Data Science** 29

Chapter 2 **Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization** 79

Chapter 3 **Descriptive Analytics II: Business Intelligence and Data Warehousing** 153

Chapter 4 **Predictive Analytics I: Data Mining Process, Methods, and Algorithms** 215

Chapter 5 **Predictive Analytics II: Text, Web, and Social Media Analytics** 273

Chapter 6 **Prescriptive Analytics: Optimization and Simulation** 345

Chapter 7 **Big Data Concepts and Tools** 395

Chapter 8 **Future Trends, Privacy and Managerial Considerations in Analytics** 443

Glossary 493

Index 501

This page intentionally left blank

Contents

Preface 19
About the Authors 25

Chapter 1 An Overview of Business Intelligence, Analytics, and Data Science 29

1.1 OPENING VIGNETTE: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics 30
1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics 37
1.3 Evolution of Computerized Decision Support to Analytics/Data Science 39
1.4 A Framework for Business Intelligence 41
Definitions of BI 42
A Brief History of BI 42
The Architecture of BI 42
The Origins and Drivers of BI 42
▸ **APPLICATION CASE 1.1 Sabre Helps Its Clients Through Dashboards and Analytics** 44
A Multimedia Exercise in Business Intelligence 45
Transaction Processing versus Analytic Processing 45
Appropriate Planning and Alignment with the Business Strategy 46
Real-Time, On-Demand BI Is Attainable 47
Developing or Acquiring BI Systems 47
Justification and Cost–Benefit Analysis 48
Security and Protection of Privacy 48
Integration of Systems and Applications 48
1.5 Analytics Overview 48
Descriptive Analytics 50
▸ **APPLICATION CASE 1.2 Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities** 50
▸ **APPLICATION CASE 1.3 Siemens Reduces Cost with the Use of Data Visualization** 51
Predictive Analytics 51
▸ **APPLICATION CASE 1.4 Analyzing Athletic Injuries** 52
Prescriptive Analytics 52
Analytics Applied to Different Domains 53
▸ **APPLICATION CASE 1.5 A Specialty Steel Bar Company Uses Analytics to Determine Available-to-Promise Dates** 53
Analytics or Data Science? 54

1.6 Analytics Examples in Selected Domains 55
Analytics Applications in Healthcare—Humana Examples 55
Analytics in the Retail Value Chain 59
1.7 A Brief Introduction to Big Data Analytics 61
What Is Big Data? 61
▸ **APPLICATION CASE 1.6 CenterPoint Energy Uses Real-Time Big Data Analytics to Improve Customer Service** 63
1.8 An Overview of the Analytics Ecosystem 63
Data Generation Infrastructure Providers 65
Data Management Infrastructure Providers 65
Data Warehouse Providers 66
Middleware Providers 66
Data Service Providers 66
Analytics-Focused Software Developers 67
Application Developers: Industry Specific or General 68
Analytics Industry Analysts and Influencers 69
Academic Institutions and Certification Agencies 70
Regulators and Policy Makers 71
Analytics User Organizations 71
1.9 Plan of the Book 72
1.10 Resources, Links, and the Teradata University Network Connection 73
Resources and Links 73
Vendors, Products, and Demos 74
Periodicals 74
The Teradata University Network Connection 74
The Book's Web Site 74
Chapter Highlights 75
Key Terms 75
Questions for Discussion 75
Exercises 76
References 77

Chapter 2 Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization 79

2.1 OPENING VIGNETTE: SiriusXM Attracts and Engages a New Generation of Radio Consumers with Data-Driven Marketing 80
2.2 The Nature of Data 83
2.3 A Simple Taxonomy of Data 87
▸ **APPLICATION CASE 2.1 Medical Device Company Ensures Product Quality While Saving Money** 89

2.4 The Art and Science of Data Preprocessing 91

▸ **APPLICATION CASE 2.2 Improving Student Retention with Data-Driven Analytics** 94

2.5 Statistical Modeling for Business Analytics 100

Descriptive Statistics for Descriptive Analytics 101

Measures of Centrality Tendency (May Also Be Called Measures of Location or Centrality) 102

Arithmetic Mean 102

Median 103

Mode 103

Measures of Dispersion (May Also Be Called Measures of Spread Decentrality) 103

Range 104

Variance 104

Standard Deviation 104

Mean Absolute Deviation 104

Quartiles and Interquartile Range 104

Box-and-Whiskers Plot 105

The Shape of a Distribution 106

▸ **APPLICATION CASE 2.3 Town of Cary Uses Analytics to Analyze Data from Sensors, Assess Demand, and Detect Problems** 110

2.6 Regression Modeling for Inferential Statistics 112

How Do We Develop the Linear Regression Model? 113

How Do We Know If the Model Is Good Enough? 114

What Are the Most Important Assumptions in Linear Regression? 115

Logistic Regression 116

▸ **APPLICATION CASE 2.4 Predicting NCAA Bowl Game Outcomes** 117

Time Series Forecasting 122

2.7 Business Reporting 124

▸ **APPLICATION CASE 2.5 Flood of Paper Ends at FEMA** 126

2.8 Data Visualization 127

A Brief History of Data Visualization 127

▸ **APPLICATION CASE 2.6 Macfarlan Smith Improves Operational Performance Insight with Tableau Online** 129

2.9 Different Types of Charts and Graphs 132

Basic Charts and Graphs 132

Specialized Charts and Graphs 133

Which Chart or Graph Should You Use? 134

2.10 The Emergence of Visual Analytics 136

Visual Analytics 138

High-Powered Visual Analytics Environments 138

2.11 Information Dashboards 143

▸ **APPLICATION CASE 2.7 Dallas Cowboys Score Big with Tableau and Teknion** 144

Dashboard Design 145

▸ **APPLICATION CASE 2.8 Visual Analytics Helps Energy Supplier Make Better Connections** 145

What to Look for in a Dashboard 147

Best Practices in Dashboard Design 147

Benchmark Key Performance Indicators with Industry Standards 147

Wrap the Dashboard Metrics with Contextual Metadata 147

Validate the Dashboard Design by a Usability Specialist 148

Prioritize and Rank Alerts/Exceptions Streamed to the Dashboard 148

Enrich the Dashboard with Business-User Comments 148

Present Information in Three Different Levels 148

Pick the Right Visual Construct Using Dashboard Design Principles 148

Provide for Guided Analytics 148

Chapter Highlights 149

Key Terms 149

Questions for Discussion 150

Exercises 150

References 152

Chapter 3 Descriptive Analytics II: Business Intelligence and Data Warehousing 153

3.1 OPENING VIGNETTE: Targeting Tax Fraud with Business Intelligence and Data Warehousing 154

3.2 Business Intelligence and Data Warehousing 156

What Is a Data Warehouse? 157

A Historical Perspective to Data Warehousing 158

Characteristics of Data Warehousing 159

Data Marts 160

Operational Data Stores 161

Enterprise Data Warehouses (EDW) 161

Metadata 161

▸ **APPLICATION CASE 3.1 A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry** 161

3.3 Data Warehousing Process 163

3.4 Data Warehousing Architectures 165

Alternative Data Warehousing Architectures 168

Which Architecture Is the Best? 170

3.5 Data Integration and the Extraction, Transformation, and Load (ETL) Processes 171

Data Integration 172

▸ **APPLICATION CASE 3.2 BP Lubricants Achieves BIGS Success** 172

Extraction, Transformation, and Load 174

3.6 Data Warehouse Development 176

▸ **APPLICATION CASE 3.3 Use of Teradata Analytics for SAP Solutions Accelerates Big Data Delivery** 177

Data Warehouse Development Approaches 179

Additional Data Warehouse Development Considerations 182

Representation of Data in Data Warehouse 182

Analysis of Data in Data Warehouse 184

OLAP versus OLTP 184

OLAP Operations 185

3.7 Data Warehousing Implementation Issues 186

Massive Data Warehouses and Scalability 188

▸ **APPLICATION CASE 3.4 EDW Helps Connect State Agencies in Michigan** 189

3.8 Data Warehouse Administration, Security Issues, and Future Trends 190

The Future of Data Warehousing 191

3.9 Business Performance Management 196

Closed-Loop BPM Cycle 197

▸ **APPLICATION CASE 3.5 AARP Transforms Its BI Infrastructure and Achieves a 347% ROI in Three Years** 199

3.10 Performance Measurement 201

Key Performance Indicator (KPI) 201

Performance Measurement System 202

3.11 Balanced Scorecards 203

The Four Perspectives 203

The Meaning of Balance in BSC 205

3.12 Six Sigma as a Performance Measurement System 205

The DMAIC Performance Model 206

Balanced Scorecard versus Six Sigma 206

Effective Performance Measurement 207

▸ **APPLICATION CASE 3.6 Expedia.com's Customer Satisfaction Scorecard** 208

Chapter Highlights 209

Key Terms 210

Questions for Discussion 210

Exercises 211

References 213

Chapter 4 **Predictive Analytics I: Data Mining Process, Methods, and Algorithms 215**

4.1 OPENING VIGNETTE: Miami-Dade Police Department Is Using Predictive Analytics to Foresee and Fight Crime 216

4.2 Data Mining Concepts and Applications 219

▸ **APPLICATION CASE 4.1 Visa Is Enhancing the Customer Experience While Reducing Fraud with Predictive Analytics and Data Mining** 220

Definitions, Characteristics, and Benefits 222

How Data Mining Works 223

▸ **APPLICATION CASE 4.2 Dell Is Staying Agile and Effective with Analytics in the 21st Century** 224

Data Mining versus Statistics 229

4.3 Data Mining Applications 229

▸ **APPLICATION CASE 4.3 Bank Speeds Time to Market with Advanced Analytics** 231

4.4 Data Mining Process 232

Step 1: Business Understanding 233

Step 2: Data Understanding 234

Step 3: Data Preparation 234

Step 4: Model Building 235

▸ **APPLICATION CASE 4.4 Data Mining Helps in Cancer Research** 235

Step 5: Testing and Evaluation 238

Step 6: Deployment 238

Other Data Mining Standardized Processes and Methodologies 238

4.5 Data Mining Methods 241

Classification 241

Estimating the True Accuracy of Classification Models 242

▸ **APPLICATION CASE 4.5 Influence Health Uses Advanced Predictive Analytics to Focus on the Factors That Really Influence People's Healthcare Decisions** 249

Cluster Analysis for Data Mining 251

Association Rule Mining 253

4.6 Data Mining Software Tools 257

▸ **APPLICATION CASE 4.6 Data Mining Goes to Hollywood: Predicting Financial Success of Movies** 259

4.7 Data Mining Privacy Issues, Myths, and Blunders 263

▸ **APPLICATION CASE 4.7 Predicting Customer Buying Patterns—The Target Story** 264

Data Mining Myths and Blunders 264

Chapter Highlights 267

Key Terms 268

Questions for Discussion 268

Exercises 269

References 271

Chapter 5 Predictive Analytics II: Text, Web, and Social Media Analytics 273

5.1 OPENING VIGNETTE: Machine versus Men on Jeopardy!: The Story of Watson 274

5.2 Text Analytics and Text Mining Overview 277

▶ **APPLICATION CASE 5.1 Insurance Group Strengthens Risk Management with Text Mining Solution** 280

5.3 Natural Language Processing (NLP) 281

▶ **APPLICATION CASE 5.2 AMC Networks Is Using Analytics to Capture New Viewers, Predict Ratings, and Add Value for Advertisers in a Multichannel World** 283

5.4 Text Mining Applications 287

Marketing Applications 287

Security Applications 287

▶ **APPLICATION CASE 5.3 Mining for Lies** 288

Biomedical Applications 290

Academic Applications 292

▶ **APPLICATION CASE 5.4 Bringing the Customer into the Quality Equation: Lenovo Uses Analytics to Rethink Its Redesign** 292

5.5 Text Mining Process 294

Task 1: Establish the Corpus 295

Task 2: Create the Term–Document Matrix 295

Task 3: Extract the Knowledge 297

▶ **APPLICATION CASE 5.5 Research Literature Survey with Text Mining** 299

5.6 Sentiment Analysis 302

▶ **APPLICATION CASE 5.6 Creating a Unique Digital Experience to Capture the Moments That Matter at Wimbledon** 303

Sentiment Analysis Applications 306

Sentiment Analysis Process 308

Methods for Polarity Identification 310

Using a Lexicon 310

Using a Collection of Training Documents 311

Identifying Semantic Orientation of Sentences and Phrases 312

Identifying Semantic Orientation of Documents 312

5.7 Web Mining Overview 313

Web Content and Web Structure Mining 315

5.8 Search Engines 317

Anatomy of a Search Engine 318

1. Development Cycle 318

2. Response Cycle 320
Search Engine Optimization 320
Methods for Search Engine Optimization 321
▸ **APPLICATION CASE 5.7 Understanding Why Customers Abandon Shopping Carts Results in a $10 Million Sales Increase** 323
5.9 Web Usage Mining (Web Analytics) 324
Web Analytics Technologies 325
Web Analytics Metrics 326
Web Site Usability 326
Traffic Sources 327
Visitor Profiles 328
Conversion Statistics 328
5.10 Social Analytics 330
Social Network Analysis 330
Social Network Analysis Metrics 331
▸ **APPLICATION CASE 5.8 Tito's Vodka Establishes Brand Loyalty with an Authentic Social Strategy** 331
Connections 334
Distributions 334
Segmentation 335
Social Media Analytics 335
How Do People Use Social Media? 336
Measuring the Social Media Impact 337
Best Practices in Social Media Analytics 337
Chapter Highlights *339*
Key Terms *340*
Questions for Discussion *341*
Exercises *341*
References *342*

Chapter 6 **Prescriptive Analytics: Optimization and Simulation** 345

6.1 OPENING VIGNETTE: School District of Philadelphia Uses Prescriptive Analytics to Find Optimal Solution for Awarding Bus Route Contracts 346
6.2 Model-Based Decision Making 348
Prescriptive Analytics Model Examples 348
▸ **APPLICATION CASE 6.1 Optimal Transport for ExxonMobil Downstream through a DSS** 349

Identification of the Problem and Environmental Analysis 350

Model Categories 350

▸ **APPLICATION CASE 6.2 Ingram Micro Uses Business Intelligence Applications to Make Pricing Decisions** 351

6.3 Structure of Mathematical Models for Decision Support 354

The Components of Decision Support Mathematical Models 354

The Structure of Mathematical Models 355

6.4 Certainty, Uncertainty, and Risk 356

Decision Making under Certainty 356

Decision Making under Uncertainty 357

Decision Making under Risk (Risk Analysis) 357

6.5 Decision Modeling with Spreadsheets 357

▸ **APPLICATION CASE 6.3 Primary Schools in Slovenia Use Interactive and Automated Scheduling Systems to Produce Quality Timetables** 358

▸ **APPLICATION CASE 6.4 Spreadsheet Helps Optimize Production Planning in Chilean Swine Companies** 359

▸ **APPLICATION CASE 6.5 Metro Meals on Wheels Treasure Valley Uses Excel to Find Optimal Delivery Routes** 360

6.6 Mathematical Programming Optimization 362

▸ **APPLICATION CASE 6.6 Mixed-Integer Programming Model Helps the University of Tennessee Medical Center with Scheduling Physicians** 363

Linear Programming Model 364

Modeling in LP: An Example 365

Implementation 370

6.7 Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking 372

Multiple Goals 372

Sensitivity Analysis 373

What-If Analysis 374

Goal Seeking 374

6.8 Decision Analysis with Decision Tables and Decision Trees 375

Decision Tables 376

Decision Trees 377

6.9 Introduction to Simulation 378

Major Characteristics of Simulation 378

▸ **APPLICATION CASE 6.7 Syngenta Uses Monte Carlo Simulation Models to Increase Soybean Crop Production** 379

Advantages of Simulation 380

Disadvantages of Simulation 381

The Methodology of Simulation 381

Simulation Types 382
Monte Carlo Simulation 383
Discrete Event Simulation 384
▸ **APPLICATION CASE 6.8 Cosan Improves Its Renewable Energy Supply Chain Using Simulation** 384
6.10 Visual Interactive Simulation 385
Conventional Simulation Inadequacies 385
Visual Interactive Simulation 385
Visual Interactive Models and DSS 386
Simulation Software 386
▸ **APPLICATION CASE 6.9 Improving Job-Shop Scheduling Decisions through RFID: A Simulation-Based Assessment** 387
Chapter Highlights 390
Key Terms 390
Questions for Discussion 391
Exercises 391
References 393

Chapter 7 Big Data Concepts and Tools 395

7.1 OPENING VIGNETTE: Analyzing Customer Churn in a Telecom Company Using Big Data Methods 396
7.2 Definition of Big Data 399
The "V"s That Define Big Data 400
▸ **APPLICATION CASE 7.1 Alternative Data for Market Analysis or Forecasts** 403
7.3 Fundamentals of Big Data Analytics 404
Business Problems Addressed by Big Data Analytics 407
▸ **APPLICATION CASE 7.2 Top Five Investment Bank Achieves Single Source of the Truth** 408
7.4 Big Data Technologies 409
MapReduce 409
Why Use MapReduce? 411
Hadoop 411
How Does Hadoop Work? 411
Hadoop Technical Components 412
Hadoop: The Pros and Cons 413
NoSQL 415
▸ **APPLICATION CASE 7.3 eBay's Big Data Solution** 416
▸ **APPLICATION CASE 7.4 Understanding Quality and Reliability of Healthcare Support Information on Twitter** 418

7.5 Big Data and Data Warehousing 419

Use Cases for Hadoop 419

Use Cases for Data Warehousing 420

The Gray Areas (Any One of the Two Would Do the Job) 421

Coexistence of Hadoop and Data Warehouse 422

7.6 Big Data Vendors and Platforms 423

IBM InfoSphere BigInsights 424

▸ **APPLICATION CASE 7.5 Using Social Media for Nowcasting the Flu Activity** 426

Teradata Aster 427

▸ **APPLICATION CASE 7.6 Analyzing Disease Patterns from an Electronic Medical Records Data Warehouse** 428

7.7 Big Data and Stream Analytics 432

Stream Analytics versus Perpetual Analytics 434

Critical Event Processing 434

Data Stream Mining 434

7.8 Applications of Stream Analytics 435

e-Commerce 435

Telecommunications 435

▸ **APPLICATION CASE 7.7 Salesforce Is Using Streaming Data to Enhance Customer Value** 436

Law Enforcement and Cybersecurity 437

Power Industry 437

Financial Services 437

Health Sciences 437

Government 438

Chapter Highlights 438

Key Terms 439

Questions for Discussion 439

Exercises 439

References 440

Chapter 8 Future Trends, Privacy and Managerial Considerations in Analytics 443

8.1 OPENING VIGNETTE: Analysis of Sensor Data Helps Siemens Avoid Train Failures 444

8.2 Internet of Things 445

▸ **APPLICATION CASE 8.1 SilverHook Powerboats Uses Real-Time Data Analysis to Inform Racers and Fans** 446

▸ **APPLICATION CASE 8.2 Rockwell Automation Monitors Expensive Oil and Gas Exploration Assets** 447

IoT Technology Infrastructure 448

RFID Sensors 448
Fog Computing 451
IoT Platforms 452
▸ **APPLICATION CASE 8.3 Pitney Bowes Collaborates with General Electric IoT Platform to Optimize Production** 452
IoT Start-Up Ecosystem 453
Managerial Considerations in the Internet of Things 454
8.3 Cloud Computing and Business Analytics 455
Data as a Service (DaaS) 457
Software as a Service (SaaS) 458
Platform as a Service (PaaS) 458
Infrastructure as a Service (IaaS) 458
Essential Technologies for Cloud Computing 459
Cloud Deployment Models 459
Major Cloud Platform Providers in Analytics 460
Analytics as a Service (AaaS) 461
Representative Analytics as a Service Offerings 461
Illustrative Analytics Applications Employing the Cloud Infrastructure 462
▸ **MD Anderson Cancer Center Utilizes Cognitive Computing Capabilities of IBM Watson to Give Better Treatment to Cancer Patients** 462
▸ **Public School Education in Tacoma, Washington, Uses Microsoft Azure Machine Learning to Predict School Dropouts** 463
▸ **Dartmouth-Hitchcock Medical Center Provides Personalized Proactive Healthcare Using Microsoft Cortana Analytics Suite** 464
▸ **Mankind Pharma Uses IBM Cloud Infrastructure to Reduce Application Implementation Time by 98%** 464
▸ **Gulf Air Uses Big Data to Get Deeper Customer Insight** 465
▸ **Chime Enhances Customer Experience Using Snowflake** 466
8.4 Location-Based Analytics for Organizations 467
Geospatial Analytics 467
▸ **APPLICATION CASE 8.4 Indian Police Departments Use Geospatial Analytics to Fight Crime** 469
▸ **APPLICATION CASE 8.5 Starbucks Exploits GIS and Analytics to Grow Worldwide** 470
Real-Time Location Intelligence 471
▸ **APPLICATION CASE 8.6 Quiznos Targets Customers for Its Sandwiches** 472
Analytics Applications for Consumers 472

8.5 Issues of Legality, Privacy, and Ethics 474
Legal Issues 474
Privacy 475
Collecting Information about Individuals 475
Mobile User Privacy 476
Homeland Security and Individual Privacy 476
Recent Technology Issues in Privacy and Analytics 477
Who Owns Our Private Data? 478
Ethics in Decision Making and Support 478
8.6 Impacts of Analytics in Organizations: An Overview 479
New Organizational Units 480
Redesign of an Organization through the Use of Analytics 481
Analytics Impact on Managers' Activities, Performance, and Job Satisfaction 481
Industrial Restructuring 482
Automation's Impact on Jobs 483
Unintended Effects of Analytics 484
8.7 Data Scientist as a Profession 485
Where Do Data Scientists Come From? 485
Chapter Highlights 488
Key Terms 489
Questions for Discussion 489
Exercises 489
References 490
Glossary 493
Index 501

This page intentionally left blank

Preface

Analytics has become the technology driver of this decade. Companies such as IBM, SAP, IBM, SAS, Teradata, SAP, Oracle, Microsoft, Dell and others are creating new organizational units focused on analytics that help businesses become more effective and efficient in their operations. Decision makers are using more computerized tools to support their work. Even consumers are using analytics tools, either directly or indirectly, to make decisions on routine activities such as shopping, health/healthcare, travel, and entertainment. The field of business intelligence and business analytics (BI & BA) has evolved rapidly to become more focused on innovative applications for extracting knowledge and insight from data streams that were not even captured some time back, much less analyzed in any significant way. New applications turn up daily in healthcare, sports, travel, entertainment, supply-chain management, utilities, and virtually every industry imaginable. The term *analytics* has become mainstream. Indeed, it has already evolved into other terms such as data science, and the latest incarnation is deep learning and Internet of Things.

This edition of the text provides a managerial perspective to business analytics continuum beginning with descriptive analytics (e.g., the nature of data, statistical modeling, data visualization, and business intelligence), moving on to predictive analytics (e.g., data mining, text/web mining, social media mining), and then to prescriptive analytics (e.g., optimization and simulation), and finally finishing with Big Data, and future trends, privacy, and managerial considerations. The book is supported by a Web site (pearsonglobaleditions.com/sharda) and also by an independent site at dssbibook.com. We will also provide links to software tutorials through a special section of the Web sites.

The purpose of this book is to introduce the reader to these technologies that are generally called *business analytics* or *data science* but have been known by other names. This book presents the fundamentals of the techniques and the manner in which these systems are constructed and used. We follow an EEE approach to introducing these topics: **Exposure**, **Experience**, and **Exploration**. The book primarily provides **exposure** to various analytics techniques and their applications. The idea is that a student will be inspired to learn from how other organizations have employed analytics to make decisions or to gain a competitive edge. We believe that such **exposure** to what is being done with analytics and how it can be achieved is the key component of learning about analytics. In describing the techniques, we also introduce specific software tools that can be used for developing such applications. The book is not limited to any one software tool, so the students can **experience** these techniques using any number of available software tools. Specific suggestions are given in each chapter, but the student and the professor are able to use this book with many different software tools. Our book's companion Web site will include specific software guides, but students can gain **experience** with these techniques in many different ways. Finally, we hope that this **exposure** and **experience** enable and motivate readers to **explore** the potential of these techniques in their own domain. To facilitate such **exploration**, we include exercises that direct them to Teradata University Network and other sites as well that include team-oriented exercises where appropriate. We will also highlight new and innovative applications that we learn about on the book's Web site.

Most of the specific improvements made in this fourth edition concentrate on four areas: reorganization, new chapters, content update, and a sharper focus. Despite the many changes, we have preserved the comprehensiveness and user friendliness that have made the text a market leader. Finally, we present accurate and updated material that is not available in any other text. We next describe the changes in the fourth edition.

What's New in the Fourth Edition?

With the goal of improving the text, this edition marks a major reorganization of the text to reflect the focus on business analytics. This edition is now organized around three major types of business analytics (i.e., descriptive, predictive, and prescriptive). The new edition has many timely additions, and the dated content has been deleted. The following major specific changes have been made.

- ***New organization.*** The book recognizes three types of analytics: descriptive, predictive, and prescriptive, a classification promoted by INFORMS. Chapter 1 introduces BI and analytics with an application focus in many industries. This chapter also includes an overview of the analytics ecosystem to help the user explore all the different ways one can participate and grow in the analytics environment. It is followed by an overview of statistics, importance of data, and descriptive analytics/visualization in Chapter 2. Chapter 3 covers data warehousing and data foundations including updated content, specifically data lakes. Chapter 4 covers predictive analytics. Chapter 5 extends the application of analytics to text, Web, and social media. Chapter 6 covers prescriptive analytics, specifically linear programming and simulation. It is totally new content for this book. Chapter 7 introduces Big Data tools and platforms. The book concludes with Chapter 8, emerging trends and topics in business analytics including location analytics, Internet of Things, cloud-based analytics, and privacy/ethical considerations in analytics. The discussion of an analytics ecosystem recognizes prescriptive analytics as well.
- ***New chapters.*** The following chapters have been added:

  Chapter 2. *Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization* This chapter aims to set the stage with a thorough understanding of the nature of data, which is the main ingredient for any analytics study. Next, statistical modeling is introduced as part of the descriptive analytics. Data visualization has become a popular part of any business reporting and/or descriptive analytics project; therefore, it is explained in detail in this chapter. The chapter is enhanced with several real-world cases and examples (75% new material).

  Chapter 6. *Prescriptive Analytics: Optimization and Simulation*

  This chapter introduces prescriptive analytics material to this book. The chapter focuses on optimization modeling in Excel using linear programming techniques. It also introduces the concept of simulation. The chapter is an updated version of material from two chapters in our DSS book, 10th edition. For this book it is an entirely new chapter (99% new material).

  Chapter 8. *Future Trends, Privacy and Managerial Considerations in Analytics* This chapter examines several new phenomena that are already changing or are likely to change analytics. It includes coverage of geospatial analytics, Internet of Things, and a significant update of the material on cloud-based analytics. It also updates some coverage from the last edition on ethical and privacy considerations (70% new material).
- ***Revised Chapters.*** All the other chapters have been revised and updated as well. Here is a summary of the changes in these other chapters:

  Chapter 1. *An Overview of Business Intelligence, Analytics, and Data Science* This chapter has been rewritten and significantly expanded. It opens with a new vignette covering multiple applications of analytics in sports. It introduces the three types of analytics as proposed by INFORMS: descriptive, predictive, and prescriptive analytics. A noted earlier, this classification is used in guiding the complete reorganization of the book itself (earlier content but with a new figure). Then it includes several new examples of analytics in healthcare and in the retail industry. Finally, it concludes with significantly expanded and updated coverage of the analytics ecosystem to give the students a sense of the vastness of the analytics and data science industry (about 60% new material).

Chapter 3. ***Descriptive Analytics II: Business Intelligence and Data Warehousing*** This is an old chapter with some new subsections (e.g., data lakes) and new cases (about 30% new material).

Chapter 4. ***Predictive Analytics I: Data Mining Process, Methods, and Algorithms*** This is an old chapter with some new content organization/flow and some new cases (about 20% new material).

Chapter 5. ***Predictive Analytics II: Text, Web, and Social Media Analytics*** This is an old chapter with some new content organization/flow and some new cases (about 25% new material).

Chapter 7. ***Big Data Concepts and Analysis*** This was Chapter 6 in the last edition. It has been updated with a new opening vignette and cases, coverage of Teradata Aster, and new material on alternative data (about 25% new material).

- ***Revamped author team.*** Building on the excellent content that has been prepared by the authors of the previous editions (Turban, Sharda, Delen, and King), this edition was revised primarily by Ramesh Sharda and Dursun Delen. Both Ramesh and Dursun have worked extensively in analytics and have industry as well as research experience.
- ***Color print!*** We are truly excited to have this book appear in color. Even the figures from previous editions have been redrawn to take advantage of color. Use of color enhances many visualization examples and also the other material.
- ***A live, updated Web site.*** Adopters of the textbook will have access to a Web site that will include links to news stories, software, tutorials, and even YouTube videos related to topics covered in the book. This site will be accessible at dssbibook.com.
- ***Revised and updated content.*** Almost all the chapters have new opening vignettes that are based on recent stories and events. In addition, application cases throughout the book have been updated to include recent examples of applications of a specific technique/model. New Web site links have been added throughout the book. We also deleted many older product links and references. Finally, most chapters have new exercises, Internet assignments, and discussion questions throughout.
- ***Links to Teradata University Network (TUN).*** Most chapters include new links to TUN (teradatauniversitynetwork.com).
- ***Book title.*** As is already evident, the book's title and focus have changed substantially.
- ***Software support.*** The TUN Web site provides software support at no charge. It also provides links to free data mining and other software. In addition, the site provides exercises in the use of such software.

The Supplement Package: www.pearsonglobaleditions .com/sharda

A comprehensive and flexible technology-support package is available to enhance the teaching and learning experience. The following instructor and student supplements are available on the book's Web site, pearsonglobaleditions.com/sharda:

- ***Instructor's Manual.*** The Instructor's Manual includes learning objectives for the entire course and for each chapter, answers to the questions and exercises at the end of each chapter, and teaching suggestions (including instructions for projects). The Instructor's Manual is available on the secure faculty section of pearsonglobaleditions .com/sharda.

- ***Test Item File and TestGen Software.*** The Test Item File is a comprehensive collection of true/false, multiple-choice, fill-in-the-blank, and essay questions. The questions are rated by difficulty level, and the answers are referenced by book page number. The Test Item File is available in Microsoft Word and in TestGen. Pearson Education's test-generating software is available from www.pearsonglobaleditions .com/sharda. The software is PC/MAC compatible and preloaded with all the Test Item File questions. You can manually or randomly view test questions and drag-and-drop to create a test. You can add or modify test-bank questions as needed.
- ***PowerPoint slides.*** PowerPoint slides are available that illuminate and build on key concepts in the text. Faculty can download the PowerPoint slides from pearsonglobaleditions.com/sharda.

Acknowledgments

Many individuals have provided suggestions and criticisms since the publication of the first edition of this book. Dozens of students participated in class testing of various chapters, software, and problems and assisted in collecting material. It is not possible to name everyone who participated in this project, but our thanks go to all of them. Certain individuals made significant contributions, and they deserve special recognition.

First, we appreciate the efforts of those individuals who provided formal reviews of the first through third editions (school affiliations as of the date of review):

Ann Aksut, Central Piedmont Community College
Bay Arinze, Drexel University
Andy Borchers, Lipscomb University
Ranjit Bose, University of New Mexico
Marty Crossland, MidAmerica Nazarene University
Kurt Engemann, Iona College
Badie Farah, Eastern Michigan University
Gary Farrar, Columbia College
Jerry Fjermestad, New Jersey Institute of Technology
Christie M. Fuller, Louisiana Tech University
Martin Grossman, Bridgewater State College
Jahangir Karimi, University of Colorado, Denver
Huei Lee, Eastern Michigan University
Natalie Nazarenko, SUNY Fredonia
Joo Eng Lee-Partridge, Central Connecticut State University
Gregory Rose, Washington State University, Vancouver
Khawaja Saeed, Wichita State University
Kala Chand Seal, Loyola Marymount University
Joshua S. White, PhD, State University of New York Polytechnic Institute
Roger Wilson, Fairmont State University
Vincent Yu, Missouri University of Science and Technology
Fan Zhao, Florida Gulf Coast University

We also appreciate the efforts of those individuals who provided formal reviews of this text and our other DSS book—*Business Intelligence and Analytics: Systems for Decision Support*, 10th Edition, Pearson Education, 2013.

Second, several individuals contributed material to the text or the supporting material. Susan Baskin of Teradata and Dr. David Schrader provided special help in identifying new TUN and Teradata content for the book and arranging permissions for the same. Dr. Dave Schrader contributed the opening vignette for the book. This vignette also included material developed by Dr. Ashish Gupta of Auburn University and Gary Wilkerson of the University of Tennessee–Chattanooga. It will provide a great introduction to analytics. We also thank INFORMS for their permission to highlight content from *Interfaces*. We also recognize the following individuals for their assistance in developing this edition of the book: Pankush Kalgotra, Prasoon Mathur, Rupesh Agarwal, Shubham Singh, Nan Liang, Jacob Pearson, Kinsey Clemmer, and Evan Murlette (all of Oklahoma State University). Their help for this edition is gratefully acknowledged. Teradata Aster team, especially Mark Ott, provided the material for the opening vignette for Chapter 7. Aster material in Chapter 7 is adapted from other training guides developed by John Thuma and Greg Bethardy. Dr. Brian LeClaire, CIO of Humana Corporation led with contributions of several real-life healthcare case studies developed by his team at Humana. Abhishek Rathi of vCreaTek contributed his vision of analytics in the retail industry. Dr. Rick Wilson's excellent exercises for teaching and practicing linear programming skills in Excel are also gratefully acknowledged. Matt Turck agreed to let us adapt his IoT ecosystem material. Ramesh also recognizes the copyediting assistance provided by his daughter, Ruchy Sharda Sen. In addition, the following former PhD students and research colleagues of ours have provided content or advice and support for the book in many direct and indirect ways:

Asil Oztekin, Universality of Massachusetts-Lowell
Enes Eryarsoy, Sehir University
Hamed Majidi Zolbanin, Ball State University
Amir Hassan Zadeh, Wright State University
Supavich (Fone) Pengnate, North Dakota State University
Christie Fuller, Boise State University
Daniel Asamoah, Wright State University
Selim Zaim, Istanbul Technical University
Nihat Kasap, Sabanci University

Third, for the previous edition, we acknowledge the contributions of Dave King (JDA Software Group, Inc.). Other major contributors to the previous edition include J. Aronson (University of Georgia), who was our coauthor, contributing to the data warehousing chapter; Mike Goul (Arizona State University), whose contributions were included in Chapter 1; and T. P. Liang (National Sun Yet-Sen University, Taiwan), who contributed material on neural networks in the previous editions. Judy Lang collaborated with all of us, provided editing, and guided us during the entire project in the first edition.

Fourth, several vendors cooperated by providing case studies and/or demonstration software for the previous editions: Acxiom (Little Rock, Arkansas), California Scientific Software (Nevada City, California), Cary Harwin of Catalyst Development (Yucca Valley, California), IBM (San Carlos, California), DS Group, Inc. (Greenwich, Connecticut), Gregory Piatetsky-Shapiro of KDnuggets.com, Gary Lynn of NeuroDimension Inc. (Gainesville, Florida), Palisade Software (Newfield, New York), Promised Land Technologies (New Haven, Connecticut), Salford Systems (La Jolla, California), Sense Networks (New York, New York), Gary Miner of StatSoft, Inc. (Tulsa, Oklahoma), Ward Systems Group, Inc. (Frederick, Maryland), Idea Fisher Systems, Inc. (Irving, California), and Wordtech Systems (Orinda, California).

Fifth, special thanks to the Teradata University Network and especially to Susan Baskin, Program Director; Hugh Watson, who started TUN; and Michael Goul, Barb Wixom, and Mary Gros for their encouragement to tie this book with TUN and for providing useful material for the book.

Finally, the Pearson team is to be commended: Samantha Lewis, who has worked with us on this revision and orchestrated the color rendition of the book; and the production team, Ann Pulido, and Revathi Viswanathan and staff at Cenveo, who transformed the manuscript into a book.

We would like to thank all these individuals and corporations. Without their help, the creation of this book would not have been possible.

R.S.

D.D.

E.T.

Global Edition Acknowledgments

For his contributions to the content of the Global Edition, Pearson would like to thank Bálint Molnár (Eötvös Loránd University, Budapest), and for their feedback, Daqing Chen (London South Bank University), Ng Hu (Multimedia University, Malaysia), and Vanina Torlo (University of Greenwich).

Note that Web site URLs are dynamic. As this book went to press, we verified that all the cited Web sites were active and valid. Web sites to which we refer in the text sometimes change or are discontinued because companies change names, are bought or sold, merge, or fail. Sometimes Web sites are down for maintenance, repair, or redesign. Most organizations have dropped the initial "www" designation for their sites, but some still use it. If you have a problem connecting to a Web site that we mention, please be patient and simply run a Web search to try to identify the new site. Most times, the new site can be found quickly. We apologize in advance for this inconvenience.

About the Authors

Ramesh Sharda (MBA, PhD, University of Wisconsin–Madison) is the Vice Dean for Research and Graduate Programs, Watson/ConocoPhillips Chair, and a Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). He cofounded and directed OSU's PhD in Business for the Executives Program. About 200 papers describing his research have been published in major journals, including *Operations Research*, *Management Science*, *Information Systems Research*, *Decision Support Systems*, and the *Journal of MIS*. He cofounded the AIS SIG on Decision Support Systems and Knowledge Management (SIGDSA). Dr. Sharda serves on several editorial boards, including those of *Decision Sciences Journal*, *Decision Support Systems*, and *ACM Data Base*. He has authored and edited several textbooks and research books and serves as the coeditor of several Springer book series (Integrated Series in Information Systems, Operations Research/Computer Science Interfaces, and Annals of Information Systems) with Springer. He is also currently serving as the Executive Director of the Teradata University Network. His current research interests are in decision support systems, business analytics, and technologies for managing information overload.

Dursun Delen (PhD, Oklahoma State University) is the Spears Endowed Chair in Business Administration, Patterson Foundation Endowed Chair in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University (OSU). Prior to his academic career, he worked for a privately owned research and consultancy company, Knowledge Based Systems Inc., in College Station, Texas, as a research scientist for 5 years, during which he led a number of decision support and other information systems–related research projects funded by several federal agencies including the Department of Defense (DoD), National Aeronautics and Space Administration (NASA), National Institute for Standards and Technology (NIST), Ballistic Missile Defense Organization (BMDO), and Department of Energy (DOE). Dr. Delen has published more than 100 peer-reviewed articles, some of which have appeared in major journals like *Decision Sciences*, *Decision Support Systems*, *Communications of the ACM*, *Computers and Operations Research*, *Computers in Industry*, *Journal of Production Operations Management*, *Artificial Intelligence in Medicine*, *International Journal of Medical Informatics*, *Expert Systems with Applications*, and *IEEE Wireless Communications*. He recently authored/coauthored seven textbooks in the broad areas of business analytics, data mining, text mining, business intelligence, and decision support systems. He is often invited to national and international conferences for keynote addresses on topics related to data/text mining, business analytics, decision support systems, business intelligence, and knowledge management. He served as the General Cochair for the Fourth International Conference on Network Computing and Advanced Information Management (September 2–4, 2008, in Seoul, South Korea) and regularly chairs, tracks, and minitracks at various information systems and analytics conferences. He is currently serving as Editor-in-Chief, Senior Editor, Associate Editor, or Editorial Board Member for more than a dozen academic journals. His research and teaching interests are in data and text mining, business analytics, decision support systems, knowledge management, business intelligence, and enterprise modeling.

Efraim Turban (MBA, PhD., University of California, Berkeley) is a Visiting Scholar at the Pacific Institute for Information System Management, University of Hawaii. Prior to this, he was on the staff of several universities, including City University of Hong Kong; Lehigh University; Florida International University; California State University, Long Beach; Eastern Illinois University; and the University of Southern California. Dr. Turban is the author of more than 100 refereed papers published in leading journals, such as *Management Science*, *MIS Quarterly*, and *Decision Support Systems*. He is also the author of 20 books, including *Electronic Commerce: A Managerial Perspective* and *Information Technology for Management*. He is also a consultant to major corporations worldwide. Dr. Turban's current areas of interest are Web-based decision support systems, social commerce, and collaborative decision making.

Business Intelligence, Analytics, and Data Science

A Managerial Perspective

This book deals with a collection of computer technologies that support managerial work—essentially, decision making. These technologies have had a profound impact on corporate strategy, performance, and competitiveness. Collectively, these technologies are called *business intelligence*, *business analytics,* and *data science*. Although the evolution of the terms is discussed, these names are also used interchangeably. This book tells stories of how smart people are employing these techniques to improve performance, service, and relationships in business, government, and non-profit worlds.

This page intentionally left blank

CHAPTER

1

An Overview of Business Intelligence, Analytics, and Data Science

LEARNING OBJECTIVES

- Understand the need for computerized support of managerial decision making
- Recognize the evolution of such computerized support to the current state—analytics/data science
- Describe the business intelligence (BI) methodology and concepts
- Understand the different types of analytics and see selected applications
- Understand the analytics ecosystem to identify various key players and career opportunities

The business environment (climate) is constantly changing, and it is becoming more and more complex. Organizations, both private and public, are under pressures that force them to respond quickly to changing conditions and to be innovative in the way they operate. Such activities require organizations to be agile and to make frequent and quick strategic, tactical, and operational decisions, some of which are very complex. Making such decisions may require considerable amounts of relevant data, information, and knowledge. Processing these, in the framework of the needed decisions, must be done quickly, frequently in real time, and usually requires some computerized support.

This book is about using business analytics as computerized support for managerial decision making. It concentrates on the theoretical and conceptual foundations of decision support, as well as on the commercial tools and techniques that are available. This book presents the fundamentals of the techniques and the manner in which these systems are constructed and used. We follow an EEE approach to introducing these topics: **Exposure**, **Experience**, and **Exploration**. The book primarily provides exposure to various analytics techniques and their applications. The idea is that a student will be inspired to learn from how other organizations have employed analytics to make decisions or to gain a competitive edge. We believe that such **exposure** to what is being done with analytics and how it can be achieved is the key component of learning about analytics. In describing the techniques, we also give examples of specific software tools that can be used for developing such applications. The book is not limited to any one software tool, so students can **experience** these techniques using any number of available software tools. We hope that this exposure and experience enable and motivate readers to explore the potential of these techniques in their own domain. To facilitate such **exploration**, we include exercises that direct the reader to Teradata University Network (TUN) and other sites that include team-oriented exercises where appropriate.

This introductory chapter provides an introduction to analytics as well as an overview of the book. The chapter has the following sections:

1.1 Opening Vignette: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics 30
1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics 37
1.3 Evolution of Computerized Decision Support to Analytics/Data Science 39
1.4 A Framework for Business Intelligence 41
1.5 Analytics Overview 48
1.6 Analytics Examples in Selected Domains 55
1.7 A Brief Introduction to Big Data Analytics 61
1.8 An Overview of the Analytics Ecosystem 63
1.9 Plan of the Book 72
1.10 Resources, Links, and the Teradata University Network Connection 73

1.1 OPENING VIGNETTE: Sports Analytics—An Exciting Frontier for Learning and Understanding Applications of Analytics

The application of analytics to business problems is a key skill, one that you will learn in this book. Many of these techniques are now being applied to improve decision making in all aspects of sports, a very hot area called sports analytics. Sports analytics is the art and science of gathering data about athletes and teams to create insights that improve sports decisions, such as deciding which players to recruit, how much to pay them, who to play, how to train them, how to keep them healthy, and when they should be traded or retired. For teams, it involves business decisions such as ticket pricing, as well as roster decisions, analysis of each competitor's strengths and weaknesses, and many game-day decisions.

Indeed, sports analytics is becoming a specialty within analytics. It is an important area because sports is a big business, generating about $145B in revenues each year, plus an additional $100B in legal and $300B in illegal gambling, according to Price Waterhouse.[1] In 2014, only $125M was spent on analytics (less than 0.1% of revenues). This is expected to grow at a healthy rate to $4.7B by 2021.[2]

[1]"Changing the Game: Outlook for the Global Sports Market to 2015," Price Waterhouse Coopers Report, appears at https://www.pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sports-market-to-2015.pdf. Betting data from https://www.capcredit.com/how-much-americansspend-on-sports-each-year/.

[2]"Sports Analytics Market Worth $4.7B by 2021," Wintergreen Research Press Release, covered by PR Newswire at http://www.prnewswire.com/news-releases/sports-analytics-market-worth-47-billion-by-2021-509869871.html, June 25, 2015.

The use of analytics for sports was popularized by the *Moneyball* book by Michael Lewis in 2003 and the movie starring Brad Pitt in 2011. It showcased Oakland A's general manager Billy Beane and his use of data and analytics to turn a losing team into a winner. In particular, he hired an analyst who used analytics to draft players able to get on base as opposed to players who excelled at traditional measures like runs batted in or stolen bases. These insights allowed them to draft prospects overlooked by other teams at reasonable starting salaries. It worked—they made it to the playoffs in 2002 and 2003.

Now analytics are being used in all parts of sports. The analytics can be divided between the front office and back office. A good description with 30 examples appears in Tom Davenport's survey article.[3] Front-office business analytics include analyzing fan behavior ranging from predictive models for season ticket renewals and regular ticket sales, to scoring tweets by fans regarding the team, athletes, coaches, and owners. This is very similar to traditional customer relationship management (CRM). Financial analysis is also a key area, where salary caps (for pros) or scholarship limits (colleges) are part of the equation.

Back-office uses include analysis of both individual athletes as well as team play. For individual players, there is a focus on recruitment models and scouting analytics, analytics for strength and fitness as well as development, and PMs for avoiding overtraining and injuries. Concussion research is a hot field. Team analytics include strategies and tactics, competitive assessments, and optimal roster choices under various on-field or on-court situations.

The following representative examples illustrate how three sports organizations use data and analytics to improve sports operations, in the same way analytics have improved traditional industry decision making.

Example 1: The Business Office

Dave Ward works as a business analyst for a major pro baseball team, focusing on revenue. He analyzes ticket sales, both from season ticket holders as well as single-ticket buyers. Sample questions in his area of responsibility include why season ticket holders renew (or do not renew) their tickets, as well as what factors drive last-minute individual seat ticket purchases. Another question is how to price the tickets.

Some of the analytical techniques Dave uses include simple statistics on fan behavior like overall attendance and answers to survey questions about likelihood to purchase again. However, what fans say versus what they do can be different. Dave runs a survey of fans by ticket seat location ("tier") and asks about their likelihood of renewing their season tickets. But when he compares what they say versus what they do, he discovers big differences. (See Figure 1.1.) He found that 69% of fans in Tier 1 seats who said on the

Tier	Highly Likely	Likely	Maybe	Probably Not	Certainly Not
1	92	88	75	69	45
2	88	81	70	65	38
3	80	76	68	55	36
4	77	72	65	45	25
5	75	70	60	35	25

FIGURE 1.1 Season Ticket Renewals—Survey Scores.

[3]Thomas Davenport, "Analytics in Sports: The New Science of Winning," International Institute for Analytics White paper, sponsored by SAS, February 2014. On the SAS Web site at: http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analytics-in-sports-106993.pdf. (Accessed July 2016)

survey that they would "probably not renew" actually did. This is useful insight that leads to action—customers in the green cells are the most likely to renew tickets, so require fewer marketing touches and dollars to convert, for example, compared to customers in the blue cells.

However, many factors influence fan ticket purchase behavior, especially price, which drives more sophisticated statistics and data analysis. For both areas, but especially single-game tickets, Dave is driving the use of dynamic pricing—moving the business from simple static pricing by seat location tier to day-by-day up-and-down pricing of individual seats. This is a rich research area for many sports teams and has huge upside potential for revenue enhancement. For example, his pricing takes into account the team's record, who they are playing, game dates and times, which star athletes play for each team, each fan's history of renewing season tickets or buying single tickets, as well as factors like seat location, number of seats, and real-time information like traffic congestion historically at game time and even the weather. See Figure 1.2.

Which of these factors are important? How much? Given his extensive statistics background, Dave builds regression models to pick out key factors driving these historic behaviors and create PMs to identify how to spend marketing resources to drive revenues. He builds churn models for season ticket holders to create segments of customers who will renew, won't renew, or are fence-sitters, which then drives more refined marketing campaigns.

In addition, he does sentiment scoring on fan comments like tweets that help him segment fans into different loyalty segments. Other studies about single-game attendance drivers help the marketing department understand the impact of giveaways like bobbleheads or T-shirts, or suggestions on where to make spot TV ad buys.

Beyond revenues, there are many other analytical areas that Dave's team works on, including merchandising, TV and radio broadcast revenues, inputs to the general manager on salary negotiations, draft analytics especially given salary caps, promotion effectiveness including advertising channels, and brand awareness, as well as partner analytics. He's a very busy guy!

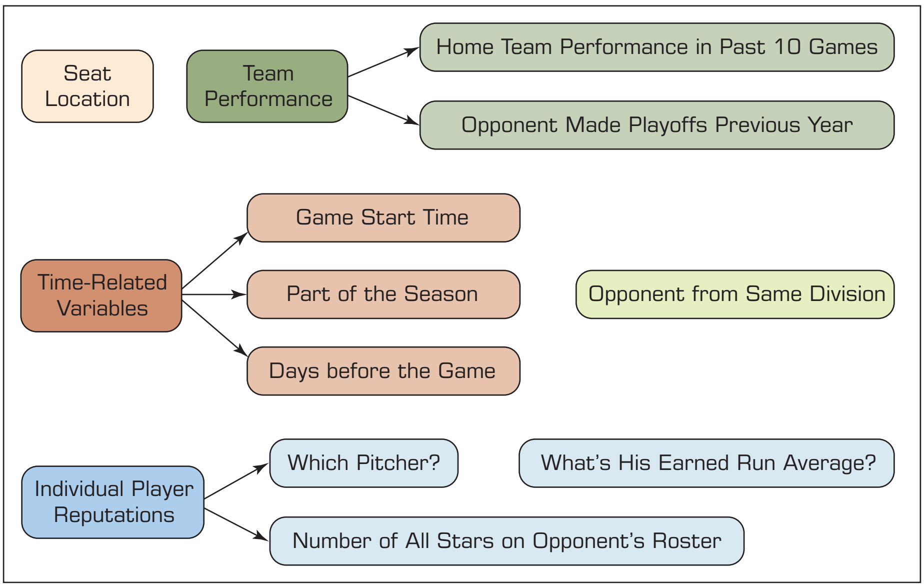

FIGURE 1.2 Dynamic Pricing Previous Work—Major League Baseball. *Source:* Adapted from C. Kemper and C. Breuer, "How Efficient is Dynamic Pricing for Sports Events? Designing a Dynamic Pricing Model for Bayern Munich", *Intl. Journal of Sports Finance*, 11, pp. 4-25, 2016.

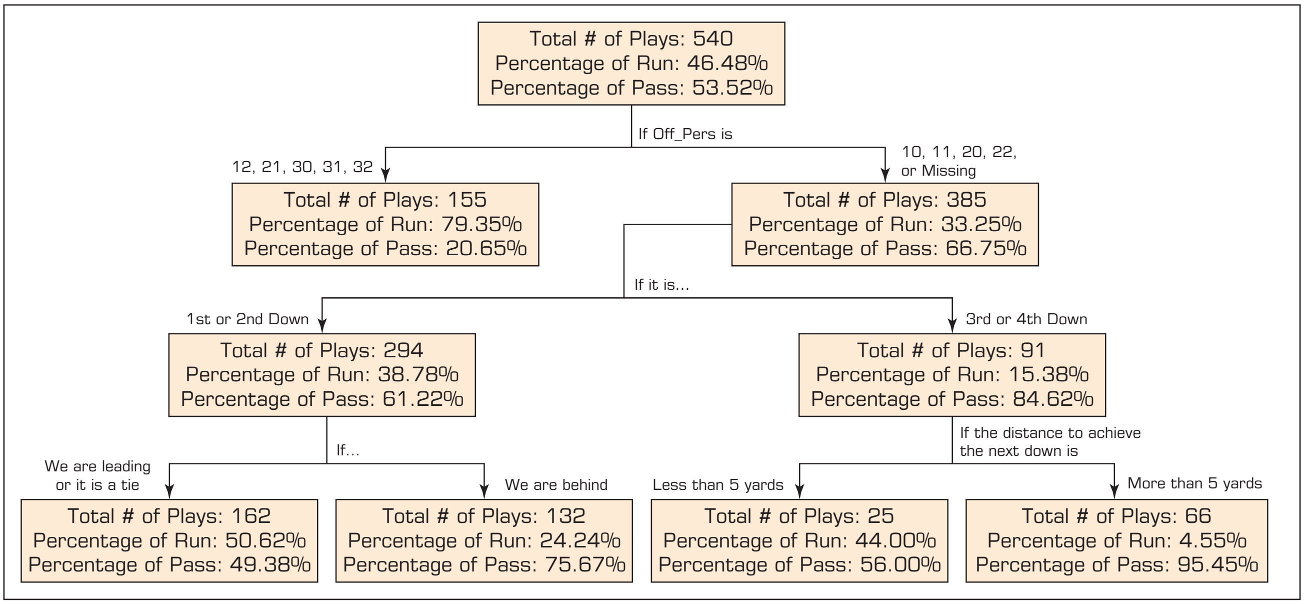

FIGURE 1.3 Cascaded Decision Tree for Run or Pass Plays.

Example 2: The Coach

Bob Breedlove is the football coach for a major college team. For him, it's all about winning games. His areas of focus include recruiting the best high school players, developing them to fit his offense and defense systems, and getting maximum effort from them on game days. Sample questions in his area of responsibility include: Who do we recruit? What drills help develop their skills? How hard do I push our athletes? Where are opponents strong or weak, and how do we figure out their play tendencies?

Fortunately, his team has hired a new team operations expert, Dar Beranek, who specializes in helping the coaches make tactical decisions. She is working with a team of student interns who are creating opponent analytics. They used the coach's annotated game film to build a cascaded decision tree model (Figure 1.3) to predict whether the next play will be a running play or passing play. For the defensive coordinator, they have built heat maps (Figure 1.4) of each opponent's passing offense, illustrating their tendencies to throw left or right and into which defensive coverage zones. Finally, they built some time series analytics (Figure 1.5) on explosive plays (defined as a gain of more than 16 yards for a passing play or more than 12 yards for a run play). For each play, they compare the outcome with their own defensive formations and the other team's offensive formations, which helps Coach Breedlove react more quickly to formation shifts during a game. We will explain the analytical techniques that generated these figures in much more depth in Chapters 2–5 and Chapter 7.

New work that Dar is fostering involves building better high school athlete recruiting models. For example, each year the team gives scholarships to three students who are wide receiver recruits. For Dar, picking out the best players goes beyond simple measures like how fast athletes run, how high they jump, or how long their arms are to newer criteria like how quickly they can rotate their heads to catch a pass, what kinds of reaction times they exhibit to multiple stimuli, and how accurately they run pass routes. Some of her ideas illustrating these concepts appear on the TUN Web site; look for the BSI Case of Precision Football.[4]

---

[4]Business Scenario Investigation BSI: The Case of Precision Football (video). (Fall 2015). Appears on http://www.teradatauniversitynetwork.com/About-Us/Whats-New/BSI–Sports-Analytics—Precision-Football//,Fall 2015. (Accessed September 2016)

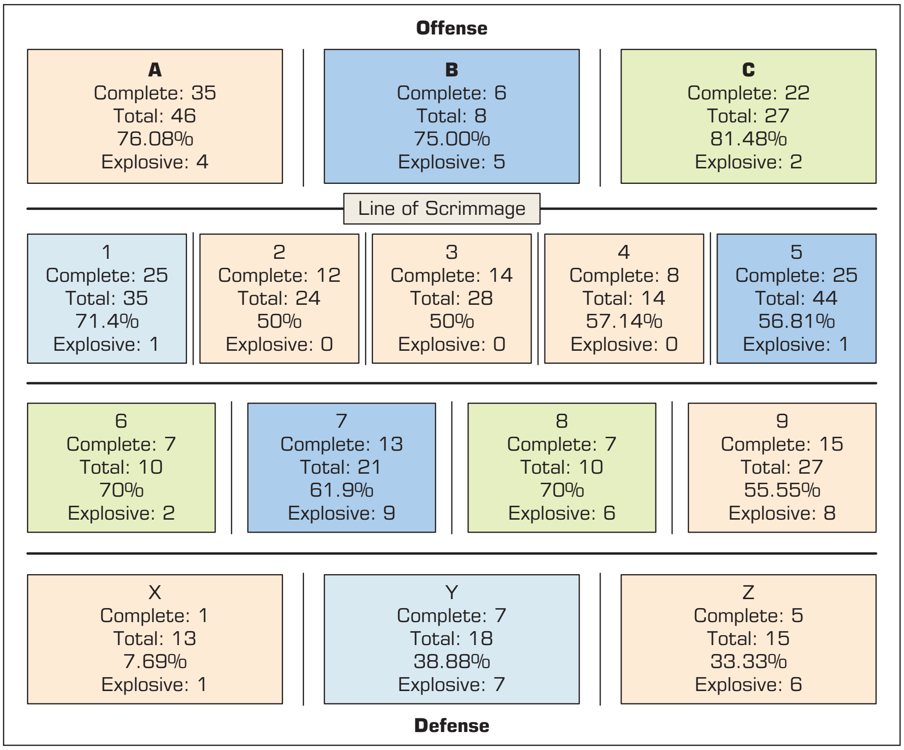

FIGURE 1.4 Heat Map Zone Analysis for Passing Plays.

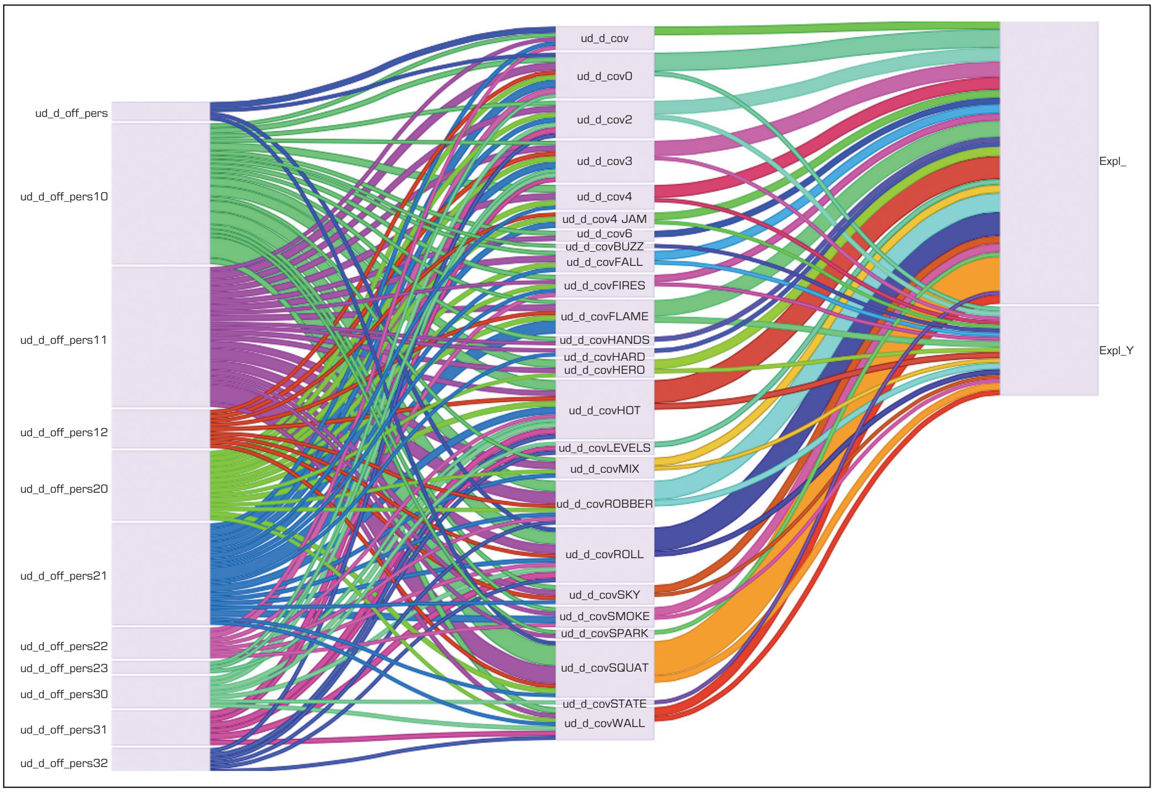

FIGURE 1.5 Time Series Analysis of Explosive Plays.

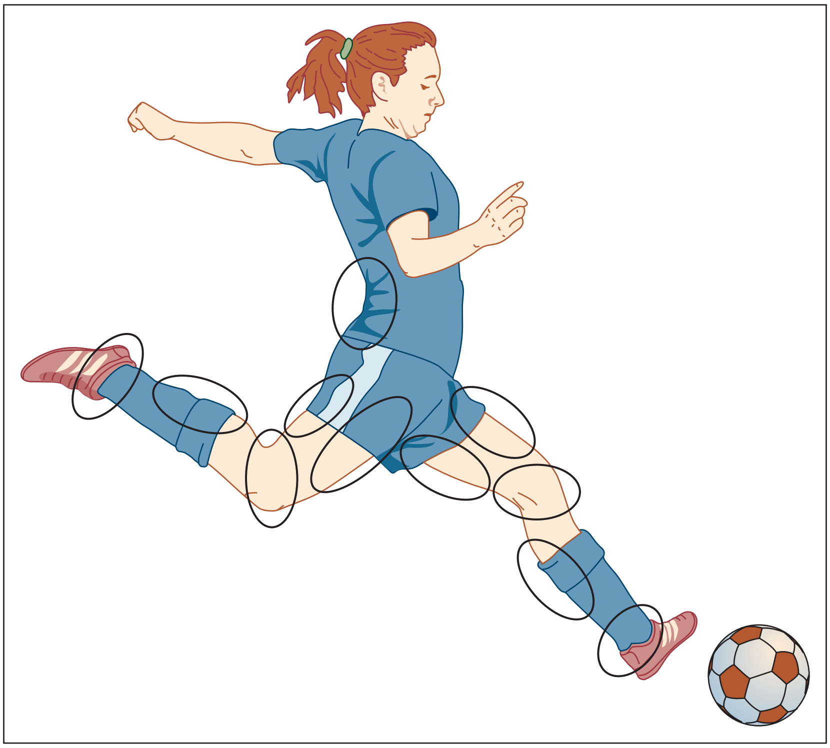

FIGURE 1.6 Soccer Injury Models.[5]

Example 3: The Trainer

Dr. Dan Johnson is the trainer for a women's college soccer team. His job is to help the players stay healthy and to advise the coaches on how much load to put on players during practices. He also has an interest in player well-being, including how much they sleep and how much rest they get between heavy and light practice sessions. The goal is to ensure that the players are ready to play on game days at maximum efficiency.

Fortunately, because of wearables, there is much more data for Dr. Dan to analyze. His players train using vests that contain sensors that can measure internal loads like heartbeats, body temperature, and respiration rates. The vests also include accelerometers that measure external loads like running distances and speeds as well as accelerations and decelerations. He knows which players are giving maximal effort during practices and those who aren't.

His focus at the moment is research that predicts or prevents player injuries (Figure 1.6). Some simple tasks like a Single Leg Squat Hold Test—standing on one foot, then the other—with score differentials of more than 10% can provide useful insights on body core strengths and weaknesses (Figure 1.7). If an athlete is hit hard during a match, a trainer can conduct a sideline test, reacting to a stimulus on a mobile device, which adds to traditional concussion protocols. Sleep sensors show who is getting adequate rest (or who partied all night). He has the MRI lab on campus do periodic brain scans to show which athletes are at risk for brain injury.

[5] "Women's Soccer Injuries," National Center for Catastrophic Sports Injury Research Report, NCAA. NCAA Sport Injury fact sheets are produced by the Datalys Center for Sports Injury Research and Prevention in collaboration with the National Collegiate Athletic Association, and STOP Sports Injuries. Appears at https://www.ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf. (Accessed November 2016).

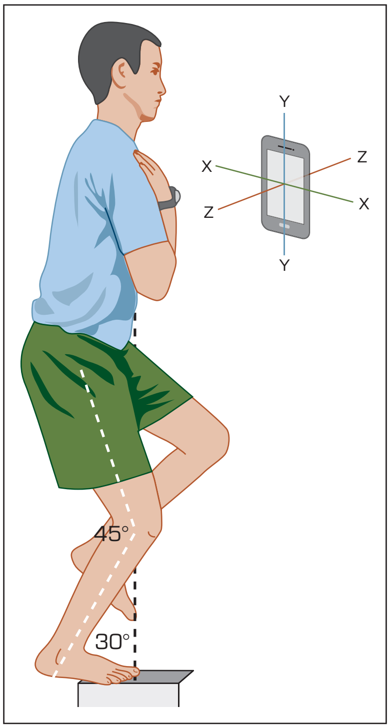

FIGURE 1.7 Single Leg Squat Hold Test–Core Body Strength Test
(*Source:* Figure adapted from Gary Wilkerson and Ashish Gupta).

QUESTIONS ABOUT THESE EXAMPLES

1. What are three factors that might be part of a PM for season ticket renewals?
2. What are two techniques that football teams can use to do opponent analysis?
3. How can wearables improve player health and safety? What kinds of new analytics can trainers use?
4. What other analytics uses can you envision in sports?

What Can We Learn from These Vignettes?

Beyond the front-office business analysts, the coaches, trainers, and performance experts, there are many other people in sports who use data, ranging from golf groundskeepers who measure soil and turf conditions for PGA tournaments, to baseball and basketball referees who are rated on the correct and incorrect calls they make. In fact, it's hard to find an area of sports that is *not* being impacted by the availability of more data, especially from sensors.

Skills you will learn in this book for business analytics will apply to sports. If you want to dig deeper into this area, we encourage you to look at the Sports Analytics section of the Teradata University Network (TUN) a free resource for students and faculty. On this Web site, you will find descriptions of what to read to find out more about sports analytics, compilations of places where you can find publically available data sets for analysis, as well as examples of student projects in sports analytics and interviews of sports professionals who use data and analytics to do their jobs. Good luck learning analytics!

Source and Credits: Contributed by Dr. Dave Schrader, who retired after 24 years in advanced development and marketing at Teradata. He has remained on the Board of Advisors of the Teradata University Network, where he spends his retirement helping students and faculty learn more about sports analytics. The football visuals (Figures 1.3–1.5) were constructed by Peter Liang and Jacob Pearson, graduate students at Oklahoma State University, as part of a student project in the spring of 2016. The training visuals (Figures 1.6 and 1.7) are adapted from the images provided by Prof. Gary Wilkerson of the University of Tennessee at Chattanooga and Prof. Ashish Gupta of Auburn University.

1.2 Changing Business Environments and Evolving Needs for Decision Support and Analytics

The opening vignette illustrates how an entire industry can employ analytics to develop reports on what is happening, predict what is likely to happen, and then also make decisions to make the best use of the situation at hand. These steps require an organization to collect and analyze vast stores of data. From traditional uses in payroll and bookkeeping functions, computerized systems have now penetrated complex managerial areas ranging from the design and management of automated factories to the application of analytical methods for the evaluation of proposed mergers and acquisitions. Nearly all executives know that information technology is vital to their business and extensively use information technologies.

Computer applications have moved from transaction processing and monitoring activities to problem analysis and solution applications, and much of the activity is done with cloud-based technologies, in many cases accessed through mobile devices. Analytics and BI tools such as data warehousing, data mining, online analytical processing (OLAP), dashboards, and the use of the cloud-based systems for decision support are the cornerstones of today's modern management. Managers must have high-speed, networked information systems (wireline or wireless) to assist them with their most important task: making decisions. In many cases, such decisions are routinely being automated, eliminating the need for any managerial intervention.

Besides the obvious growth in hardware, software, and network capacities, some developments have clearly contributed to facilitating growth of decision support and analytics in a number of ways, including the following:

- **Group communication and collaboration.** Many decisions are made today by groups whose members may be in different locations. Groups can collaborate and communicate readily by using collaboration tools as well as the ubiquitous smartphones. Collaboration is especially important along the supply chain, where partners—all the way from vendors to customers—must share information. Assembling a group of decision makers, especially experts, in one place can be costly. Information systems can improve the collaboration process of a group and enable its members to be at different locations (saving travel costs). More critically, such supply chain collaboration permits manufacturers to know about the changing patterns of demand in near real time and thus react to marketplace changes faster.
- **Improved data management.** Many decisions involve complex computations. Data for these can be stored in different databases anywhere in the organization and even possibly outside the organization. The data may include text, sound, graphics, and video, and these can be in different languages. Many times it is necessary to transmit data quickly from distant locations. Systems today can search, store, and transmit needed data quickly, economically, securely, and transparently.
- **Managing giant data warehouses and Big Data.** Large data warehouses (DWs), like the ones operated by Walmart, contain humongous amounts of data. Special

methods, including parallel computing, Hadoop/Spark, and so on, are available to organize, search, and mine the data. The costs related to data storage and mining are declining rapidly. Technologies that fall under the broad category of Big Data have enabled massive data coming from a variety of sources and in many different forms, which allows a very different view into organizational performance that was not possible in the past.

- **Analytical support.** With more data and analysis technologies, more alternatives can be evaluated, forecasts can be improved, risk analysis can be performed quickly, and the views of experts (some of whom may be in remote locations) can be collected quickly and at a reduced cost. Expertise can even be derived directly from analytical systems. With such tools, decision makers can perform complex simulations, check many possible scenarios, and assess diverse impacts quickly and economically. This, of course, is the focus of several chapters in the book.
- **Overcoming cognitive limits in processing and storing information.** According to Simon (1977), the human mind has only a limited ability to process and store information. People sometimes find it difficult to recall and use information in an error-free fashion due to their cognitive limits. The term *cognitive limits* indicates that an individual's problem-solving capability is limited when a wide range of diverse information and knowledge is required. Computerized systems enable people to overcome their cognitive limits by quickly accessing and processing vast amounts of stored information.
- **Knowledge management.** Organizations have gathered vast stores of information about their own operations, customers, internal procedures, employee interactions, and so forth, through the unstructured and structured communications taking place among the various stakeholders. Knowledge management systems have become sources of formal and informal support for decision making to managers, although sometimes they may not even be called *KMS*. Technologies such as text analytics and IBM Watson are making it possible to generate value from such knowledge stores.
- **Anywhere, anytime support.** Using wireless technology, managers can access information anytime and from anyplace, analyze and interpret it, and communicate with those involved. This perhaps is the biggest change that has occurred in the last few years. The speed at which information needs to be processed and converted into decisions has truly changed expectations for both consumers and businesses.

These and other capabilities have been driving the use of computerized decision support since the late 1960s, but especially since the mid-1990s. The growth of mobile technologies, social media platforms, and analytical tools has enabled a different level of information systems (IS) support for managers. This growth in providing data-driven support for any decision extends to not just the managers but also to consumers. We will first study an overview of technologies that have been broadly referred to as BI. From there we will broaden our horizons to introduce various types of analytics.

SECTION 1.2 REVIEW QUESTIONS

1. What are some of the key system-oriented trends that have fostered IS-supported decision making to a new level?
2. List some capabilities of information systems that can facilitate managerial decision making.
3. How can a computer help overcome the cognitive limits of humans?

1.3 Evolution of Computerized Decision Support to Analytics/Data Science

The timeline in Figure 1.8 shows the terminology used to describe analytics since the 1970s. During the 1970s, the primary focus of information systems support for decision making focused on providing structured, periodic reports that a manager could use for decision making (or ignore them). Businesses began to create routine reports to inform decision makers (managers) about what had happened in the previous period (e.g., day, week, month, quarter). Although it was useful to know what had happened in the past, managers needed more than this: They needed a variety of reports at different levels of granularity to better understand and address changing needs and challenges of the business. These were usually called management information systems (MIS). In the early 1970s, Scott-Morton first articulated the major concepts of DSS. He defined DSSs as "interactive computer-based systems, which help decision makers utilize *data* and *models* to solve unstructured problems" (Gorry and Scott-Morton, 1971). The following is another classic DSS definition, provided by Keen and Scott-Morton (1978):

> Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems.

Note that the term *decision support system*, like *management information system* and several other terms in the field of IT, is a content-free expression (i.e., it means different things to different people). Therefore, there is no universally accepted definition of DSS.

During the early days of analytics, data was often obtained from the domain experts using manual processes (i.e., interviews and surveys) to build mathematical or knowledge-based models to solve constrained optimization problems. The idea was to do the best with limited resources. Such decision support models were typically called operations research (OR). The problems that were too complex to solve optimally (using linear or nonlinear mathematical programming techniques) were tackled using heuristic methods such as simulation models. (We will introduce these as prescriptive analytics later in this chapter and in a bit more detail in Chapter 6.)

In the late 1970s and early 1980s, in addition to the mature OR models that were being used in many industries and government systems, a new and exciting line of models had emerged: rule-based expert systems. These systems promised to capture experts' knowledge in a format that computers could process (via a collection of if–then–else rules or heuristics) so that these could be used for consultation much the same way that one

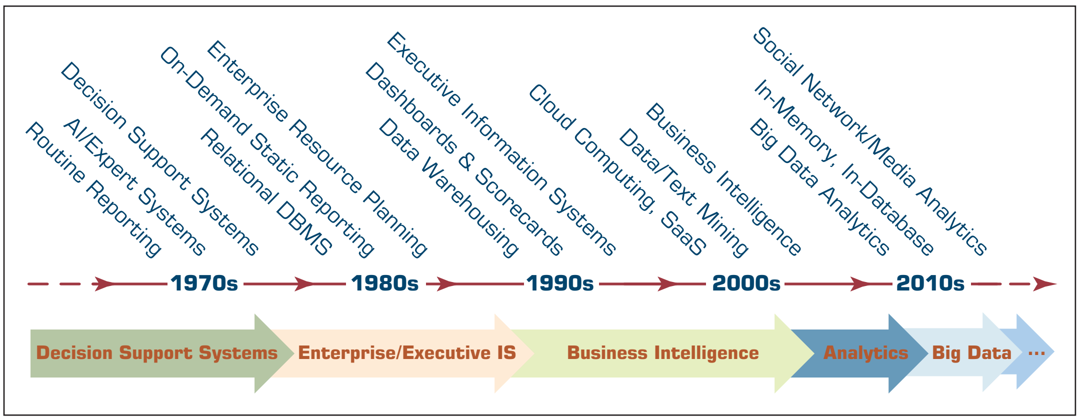

FIGURE 1.8 Evolution of Decision Support, Business Intelligence, and Analytics.

would use domain experts to identify a structured problem and to prescribe the most probable solution. ESs allowed scarce expertise to be made available where and when needed, using an "intelligent" DSS.

The 1980s saw a significant change in the way organizations captured business-related data. The old practice had been to have multiple disjointed information systems tailored to capture transactional data of different organizational units or functions (e.g., accounting, marketing and sales, finance, manufacturing). In the 1980s, these systems were integrated as enterprise-level information systems that we now commonly call enterprise resource planning (ERP) systems. The old mostly sequential and nonstandardized data representation schemas were replaced by relational database management (RDBM) systems. These systems made it possible to improve the capture and storage of data, as well as the relationships between organizational data fields while significantly reducing the replication of information. The need for RDBM and ERP systems emerged when data integrity and consistency became an issue, significantly hindering the effectiveness of business practices. With ERP, all the data from every corner of the enterprise is collected and integrated into a consistent schema so that every part of the organization has access to the single version of the truth when and where needed. In addition to the emergence of ERP systems, or perhaps because of these systems, business reporting became an on-demand, as-needed business practice. Decision makers could decide when they needed to or wanted to create specialized reports to investigate organizational problems and opportunities.

In the 1990s, the need for more versatile reporting led to the development of executive information systems (EISs; DSSs designed and developed specifically for executives and their decision-making needs). These systems were designed as graphical dashboards and scorecards so that they could serve as visually appealing displays while focusing on the most important factors for decision makers to keep track of the key performance indicators. To make this highly versatile reporting possible while keeping the transactional integrity of the business information systems intact, it was necessary to create a middle data tier known as a DW as a repository to specifically support business reporting and decision making. In a very short time, most large to medium-sized businesses adopted data warehousing as their platform for enterprise-wide decision making. The dashboards and scorecards got their data from a DW, and by doing so, they were not hindering the efficiency of the business transaction systems mostly referred to as (ERP) systems.

In the 2000s, the DW-driven DSSs began to be called BI systems. As the amount of longitudinal data accumulated in the DWs increased, so did the capabilities of hardware and software to keep up with the rapidly changing and evolving needs of the decision makers. Because of the globalized competitive marketplace, decision makers needed current information in a very digestible format to address business problems and to take advantage of market opportunities in a timely manner. Because the data in a DW is updated periodically, it does not reflect the latest information. To elevate this information latency problem, DW vendors developed a system to update the data more frequently, which led to the terms *real-time data warehousing* and, more realistically, *right-time data warehousing*, which differs from the former by adopting a data-refreshing policy based on the needed freshness of the data items (i.e., not all data items need to be refreshed in real time). DWs are very large and feature rich, and it became necessary to "mine" the corporate data to "discover" new and useful knowledge nuggets to improve business processes and practices, hence the terms *data mining* and *text mining*. With the increasing volumes and varieties of data, the needs for more storage and more processing power emerged. Although large corporations had the means to tackle this problem, small to medium-sized companies needed more financially manageable business models. This need led to service-oriented architecture and software and infrastructure-as-a-service analytics business models. Smaller companies, therefore, gained access to analytics capabilities on an as-needed basis and paid only for what they used, as opposed to investing in financially prohibitive hardware and software resources.

In the 2010s, we are seeing yet another paradigm shift in the way that data is captured and used. Largely because of the widespread use of the Internet, new data generation mediums have emerged. Of all the new data sources (e.g., radio-frequency identification [RFID] tags, digital energy meters, clickstream Web logs, smart home devices, wearable health monitoring equipment), perhaps the most interesting and challenging is social networking/social media. This unstructured data is rich in information content, but analysis of such data sources poses significant challenges to computational systems, from both software and hardware perspectives. Recently, the term *Big Data* has been coined to highlight the challenges that these new data streams have brought on us. Many advancements in both hardware (e.g., massively parallel processing with very large computational memory and highly parallel multiprocessor computing systems) and software/algorithms (e.g., Hadoop with MapReduce and NoSQL) have been developed to address the challenges of Big Data.

It's hard to predict what the next decade will bring and what the new analytics-related terms will be. The time between new paradigm shifts in information systems and particularly in analytics has been shrinking, and this trend will continue for the foreseeable future. Even though analytics is not new, the explosion in its popularity is very new. Thanks to the recent explosion in Big Data, ways to collect and store this data, and intuitive software tools, data-driven insights are more accessible to business professionals than ever before. Therefore, in the midst of global competition, there is a huge opportunity to make better managerial decisions by using data and analytics to increase revenue while decreasing costs by building better products, improving customer experience, and catching fraud before it happens, improving customer engagement through targeting and customization all with the power of analytics and data. More and more companies are now preparing their employees with the know-how of business analytics to drive effectiveness and efficiency in their day-to-day decision-making processes.

The next section focuses on a framework for BI. Although most people would agree that BI has evolved into analytics and data science, many vendors and researchers still use that term. So Section 1.4 pays homage to that history by specifically focusing on what has been called BI. Following the next section, we introduce analytics and will use that as the label for classifying all related concepts.

SECTION 1.3 REVIEW QUESTIONS

1. List three of the terms that have been predecessors of analytics.
2. What was the primary difference between the systems called MIS, DSS, and Executive Support Systems?
3. Did DSS evolve into BI or vice versa?

1.4 A Framework for Business Intelligence

The decision support concepts presented in Sections 1.2 and 1.3 have been implemented incrementally, under different names, by many vendors that have created tools and methodologies for decision support. As noted in Section 1.3, as the enterprise-wide systems grew, managers were able to access user-friendly reports that enabled them to make decisions quickly. These systems, which were generally called EISs, then began to offer additional visualization, alerts, and performance measurement capabilities. By 2006, the major *commercial* products and services appeared under the term *business intelligence* (BI).

Definitions of BI

Business intelligence (BI) is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies. It is, like DSS, a content-free expression, so it means different things to different people. Part of the confusion about BI lies in the flurry of acronyms and buzzwords that are associated with it (e.g., business performance management [BPM]). BI's major objective is to enable interactive access (sometimes in real time) to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analyses. By analyzing historical and current data, situations, and performances, decision makers get valuable insights that enable them to make more informed and better decisions. The process of BI is based on the *transformation* of data to information, then to decisions, and finally to actions.

A Brief History of BI

The term *BI* was coined by the Gartner Group in the mid-1990s. However, as the history in the previous section points out, the concept is much older; it has its roots in the MIS reporting systems of the 1970s. During that period, reporting systems were static, were two dimensional, and had no analytical capabilities. In the early 1980s, the concept of EISs emerged. This concept expanded the computerized support to top-level managers and executives. Some of the capabilities introduced were dynamic multidimensional (ad hoc or on-demand) reporting, forecasting and prediction, trend analysis, drill-down to details, status access, and critical success factors. These features appeared in dozens of commercial products until the mid-1990s. Then the same capabilities and some new ones appeared under the name BI. Today, a good BI-based enterprise information system contains all the information executives need. So, the original concept of EIS was transformed into BI. By 2005, BI systems started to include *artificial intelligence* capabilities as well as powerful analytical capabilities. Figure 1.9 illustrates the various tools and techniques that may be included in a BI system. It illustrates the evolution of BI as well. The tools shown in Figure 1.9 provide the capabilities of BI. The most sophisticated BI products include most of these capabilities; others specialize in only some of them.

The Architecture of BI

A BI system has four major components: a *DW*, with its source data; *business analytics*, a collection of tools for manipulating, mining, and analyzing the data in the DW; *BPM* for monitoring and analyzing performance; and a *user interface* (e.g., a **dashboard**). The relationship among these components is illustrated in Figure 1.10.

The Origins and Drivers of BI

Where did modern approaches to data warehousing and BI come from? What are their roots, and how do those roots affect the way organizations are managing these initiatives today? Today's investments in information technology are under increased scrutiny in terms of their bottom-line impact and potential. The same is true of DW and the BI applications that make these initiatives possible.

Organizations are being compelled to capture, understand, and harness their data to support decision making to improve business operations. Legislation and regulation (e.g., the Sarbanes-Oxley Act of 2002) now require business leaders to document their business processes and to sign off on the legitimacy of the information they rely on and report to stakeholders. Moreover, business cycle times are now extremely compressed; faster, more informed, and better decision making is, therefore, a competitive imperative. Managers need the *right information* at the *right time* and in the *right place*. This is the mantra for modern approaches to BI.

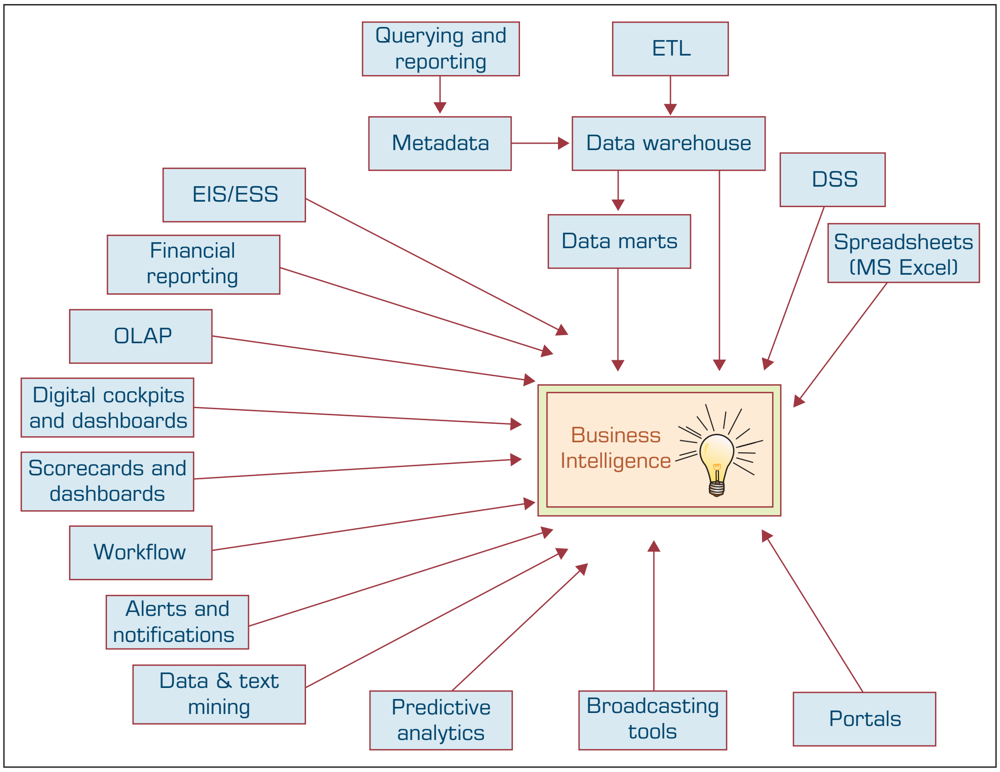

FIGURE 1.9 Evolution of Business Intelligence (BI).

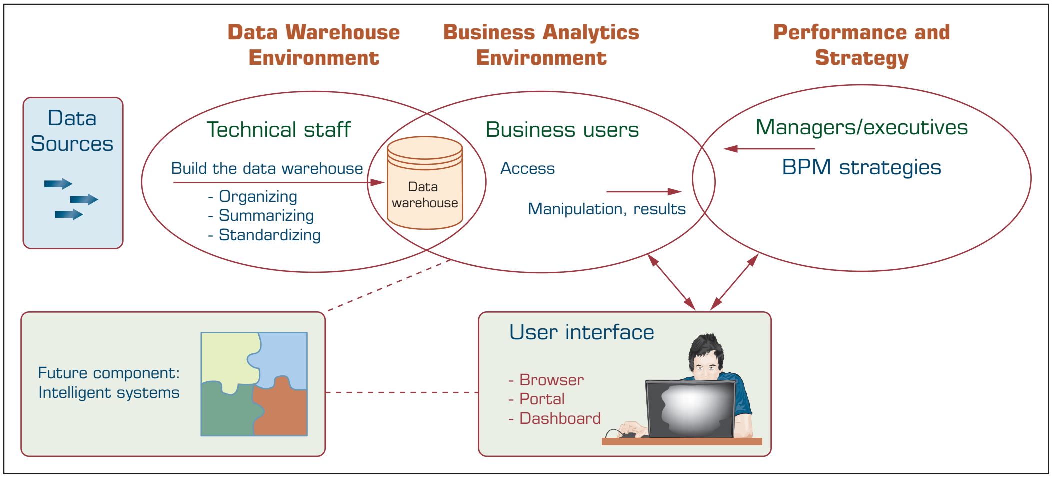

FIGURE 1.10 A High-Level Architecture of BI. (*Source:* Based on W. Eckerson, *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions.* The Data Warehousing Institute, Seattle, WA, 2003, p. 32, Illustration 5.)

Organizations have to work smart. Paying careful attention to the management of BI initiatives is a necessary aspect of doing business. It is no surprise, then, that organizations are increasingly championing BI and under its new incarnation as analytics. Application Case 1.1 illustrates one such application of BI that has helped many airlines as well as, of course, the companies offering such services to the airlines.

Application Case 1.1

Sabre Helps Its Clients Through Dashboards and Analytics

Sabre is one of the world leaders in the travel industry, providing both business-to-consumer services as well as business-to-business services. It serves travelers, travel agents, corporations, and travel suppliers through its four main companies: Travelocity, Sabre Travel Network, Sabre Airline Solutions, and Sabre Hospitality Solutions. The current volatile global economic environment poses significant competitive challenges to the airline industry. To stay ahead of the competition, Sabre Airline Solutions recognized that airline executives needed enhanced tools for managing their business decisions by eliminating the traditional, manual, time-consuming process of aggregating financial and other information needed for actionable initiatives. This enables real-time decision support at airlines throughout the world to maximize their (and in turn Sabre's) return on information by driving insights, actionable intelligence, and value for customers from the growing data.

Sabre developed an Enterprise Travel Data Warehouse (ETDW) using Teradata to hold its massive reservations data. ETDW is updated in near-real time with batches that run every 15 minutes, gathering data from all of Sabre's businesses. Sabre uses its ETDW to create Sabre Executive Dashboards that provide near real-time executive insights using a Cognos BI platform with Oracle Data Integrator and Oracle Goldengate technology infrastructures. The Executive Dashboards offer their client airlines' top-level managers and decision makers a timely, automated, user-friendly solution, aggregating critical performance metrics in a succinct way and providing at a glance a 360-degree view of the overall health of the airline. At one airline, Sabre's Executive Dashboards provide senior management with a daily and intraday snapshot of key performance indicators in a single application replacing the once-a-week, 8-hour process of generating the same report from various data sources. The use of dashboards is not limited to the external customers; Sabre also uses them for their assessment of internal operational performance.

The dashboards help Sabre's customers to have a clear understanding of the data through the visual displays that incorporate interactive drill-down capabilities. It replaces flat presentations and allows for a more focused review of the data with less effort and time. This facilitates team dialog by making the data/metrics pertaining to sales performance available to many stakeholders, including ticketing, seats sold and flown, operational performance including the data on flight movement and tracking, customer reservations, inventory, and revenue across an airline's multiple distribution channels. The dashboard systems provide scalable infrastructure, graphical user interface support, data integration, and aggregation that empower airline executives to be more proactive in taking actions that lead to positive impacts on the overall health of their airline.

With its ETDW, Sabre could also develop other Web-based analytical and reporting solutions that leverage data to gain customer insights through analysis of customer profiles and their sales interactions to calculate customer value. This enables better customer segmentation and insights for value-added services.

Questions for Discussion

1. What is traditional reporting? How is it used in the organization?
2. How can analytics be used to transform the traditional reporting?
3. How can interactive reporting assist organizations in decision making?

What We Can Learn from This Application Case

This case shows that organizations that earlier used reporting only for tracking their internal business activities and meeting the compliance requirements set out by the government are now moving toward generating actionable intelligence from their transactional business data. Reporting has become broader as organizations are now trying to analyze the archived transactional data to understand the underlying hidden trends and patterns that will enable them to make better decisions by gaining insights into problematic areas and resolving them to pursue current and future market opportunities. Reporting has advanced to interactive online reports, which enable the users to pull and build quick custom reports and even present the reports aided by visualization tools that have the ability to connect to the database, providing the capabilities of digging deep into summarized data.

Source: Teradata.com, "Sabre Airline Solutions," Terry, D. (2011), "Sabre Streamlines Decision Making," http://www.teradatamagazine.com/v11n04/Features/Sabre-Streamlines-Decision-Making/ (Accessed July 2016).

A Multimedia Exercise in Business Intelligence

TUN includes videos (similar to the television show *CSI*) to illustrate concepts of analytics in different industries. These are called "BSI Videos (Business Scenario Investigations)." Not only are these entertaining, but they also provide the class with some questions for discussion. For starters, please go to http://www.teradatauniversitynetwork.com/Library/Items/BSI–The-Case-of-the-Misconnecting-Passengers/ or www.youtube.com/watch?v=NXEL5F4_aKA. Watch the video that appears on YouTube. Essentially, you have to assume the role of a customer service center professional. An incoming flight is running late, and several passengers are likely to miss their connecting flights. There are seats on one outgoing flight that can accommodate two of the four passengers. Which two passengers should be given priority? You are given information about customers' profiles and relationships with the airline. Your decisions might change as you learn more about those customers' profiles.

Watch the video, pause it as appropriate, and answer the questions on which passengers should be given priority. Then resume the video to get more information. After the video is complete, you can see the slides related to this video and how the analysis was prepared on a slide set at www.slideshare.net/teradata/bsi-how-we-did-it-the-case-of-the-misconnecting-passengers.

This multimedia excursion provides an example of how additional available information through an enterprise DW can assist in decision making.

Although some people equate DSS with BI, these systems are not, at present, the same. It is interesting to note that some people believe that DSS is a part of BI—one of its analytical tools. Others think that BI is a special case of DSS that deals mostly with reporting, communication, and collaboration (a form of data-oriented DSS). Another explanation (Watson, 2005) is that BI is a result of a continuous revolution, and as such, DSS is one of BI's original elements. Further, as noted in the next section onward, in many circles BI has been subsumed by the new terms *analytics* or *data science*.

Transaction Processing versus Analytic Processing

To illustrate the major characteristics of BI, first we will show what BI is not—namely, transaction processing. We're all familiar with the information systems that support our transactions, like ATM withdrawals, bank deposits, cash register scans at the grocery store, and so on. These *transaction processing* systems are constantly involved in handling updates to what we might call *operational databases*. For example, in an ATM withdrawal transaction, we need to reduce our bank balance accordingly; a bank deposit adds to an account; and a grocery store purchase is likely reflected in the store's calculation of total sales for the day, and it should reflect an appropriate reduction in the store's inventory for the items we bought, and so on. These **online transaction processing (OLTP)** systems handle a company's routine ongoing business. In contrast, a DW is typically a distinct system that provides storage for data that will be used for *analysis*. The intent of that analysis is to give management the ability to scour data for information about the business, and it can be used to provide tactical or operational decision support, whereby, for example, line personnel can make quicker and/or more informed decisions. We will provide a more technical definition of DW in Chapter 2, but suffice it to say that DWs are intended to work with informational data used for **online analytical processing (OLAP)** systems.

Most operational data in enterprise resources planning (ERP) systems—and in its complementary siblings like *supply chain management* (SCM) or *CRM*—are stored in an OLTP system, which is a type of computer processing where the computer responds immediately to user requests. Each request is considered to be a *transaction*, which is a computerized record of a discrete event, such as the receipt of inventory or a customer order. In other words, a transaction requires a set of two or more database updates that must be completed in an all-or-nothing fashion.

The very design that makes an OLTP system efficient for transaction processing makes it inefficient for end-user ad hoc reports, queries, and analysis. In the 1980s, many business users referred to their mainframes as "black holes " because all the information went into them, but none ever came back. All requests for reports had to be programmed by the IT staff, whereas only "precanned" reports could be generated on a scheduled basis, and ad hoc real-time querying was virtually impossible. Although the client/server-based ERP systems of the 1990s were somewhat more report-friendly, it has still been a far cry from a desired usability by regular, nontechnical, end users for things such as operational reporting, interactive analysis, and so on. To resolve these issues, the notions of DW and BI were created.

DWs contain a wide variety of data that present a coherent picture of business conditions at a single point in time. The idea was to create a database infrastructure that was always online and contained all the information from the OLTP systems, including historical data, but reorganized and structured in such a way that it was fast and efficient for querying, analysis, and decision support. Separating the OLTP from analysis and decision support enables the benefits of BI that were described earlier.

Appropriate Planning and Alignment with the Business Strategy

First and foremost, the fundamental reasons for investing in BI must be aligned with the company's business strategy. BI cannot simply be a technical exercise for the information systems department. It has to serve as a way to change the manner in which the company conducts business by improving its business processes and transforming decision-making processes to be more data driven. Many BI consultants and practitioners involved in successful BI initiatives advise that a framework for planning is a necessary precondition. One framework, developed by Gartner, Inc. (2004), decomposes planning and execution into *business, organization, functionality*, and *infrastructure* components. At the business and organizational levels, strategic and operational objectives must be defined while considering the available organizational skills to achieve those objectives. Issues of organizational culture surrounding BI initiatives and building enthusiasm for those initiatives and procedures for the intra-organizational sharing of BI best practices must be considered by upper management—with plans in place to prepare the organization for change. One of the first steps in that process is to assess the IS organization, the skill sets of the potential classes of users, and whether the culture is amenable to change. From this assessment, and assuming there is justification and the need to move ahead, a company can prepare a detailed action plan. Another critical issue for BI implementation success is the integration of several BI projects (most enterprises use several BI projects) among themselves and with the other IT systems in the organization and its business partners.

If the company's strategy is properly aligned with the reasons for DW and BI initiatives, and if the company's IS organization is or can be made capable of playing its role in such a project, and if the requisite user community is in place and has the proper motivation, it is wise to start BI and establish a BI Competency Center within the company. The center could serve some or all of the following functions (Gartner, 2004):

- The center can demonstrate how BI is clearly linked to strategy and execution of strategy.
- A center can serve to encourage interaction between the potential business user communities and the IS organization.
- The center can serve as a repository and disseminator of best BI practices between and among the different lines of business.
- Standards of excellence in BI practices can be advocated and encouraged throughout the company.
- The IS organization can learn a great deal through interaction with the user communities, such as knowledge about the variety of types of analytical tools that are needed.
- The business user community and IS organization can better understand why the DW platform must be flexible enough to provide for changing business requirements.
- It can help important stakeholders like high-level executives see how BI can play an important role.

Another important success factor of BI is its ability to facilitate a real-time, on-demand agile environment, introduced next.

Real-Time, On-Demand BI Is Attainable

The demand for instant, on-demand access to dispersed information has grown as the need to close the gap between the operational data and strategic objectives has become more pressing. As a result, a category of products called *real-time BI applications* has emerged. The introduction of new data-generating technologies, such as RFID and other sensors is only accelerating this growth and the subsequent need for real-time BI. Traditional BI systems use a large volume of *static* data that has been extracted, cleansed, and loaded into a *DW* to produce reports and analyses. However, the need is not just reporting because users need business monitoring, performance analysis, and an understanding of why things are happening. These can assist users, who need to know (virtually in real time) about changes in data or the availability of relevant reports, alerts, and notifications regarding events and emerging trends in social media applications. In addition, business applications can be programmed to act on what these real-time BI systems discover. For example, an SCM application might automatically place an order for more "widgets" when real-time inventory falls below a certain threshold or when a CRM application automatically triggers a customer service representative and credit control clerk to check a customer who has placed an online order larger than $10,000.

One approach to real-time BI uses the DW model of traditional BI systems. In this case, products from innovative BI platform providers provide a service-oriented, near-real-time solution that populates the DW much faster than the typical nightly *extract/transfer/load* batch update does (see Chapter 3). A second approach, commonly called *business activity management* (BAM), is adopted by pure-play BAM and/or hybrid BAM-middleware providers (such as Savvion, Iteration Software, Vitria, webMethods, Quantive, Tibco, or Vineyard Software). It bypasses the DW entirely and uses **Web services** or other monitoring means to discover key business events. These software monitors (or **intelligent agents**) can be placed on a separate server in the network or on the transactional application databases themselves, and they can use event- and process-based approaches to proactively and intelligently measure and monitor operational processes.

Developing or Acquiring BI Systems

Today, many vendors offer diversified tools, some of which are completely preprogrammed (called *shells*); all you have to do is insert your numbers. These tools can be purchased or leased. For a list of products, demos, white papers, and more current product information, see product directories at tdwi.org. Free user registration is required. Almost all BI applications are constructed with shells provided by vendors who may themselves create a custom solution for a client or work with another outsourcing provider. The issue that companies face is which alternative to select: purchase, lease, or build. Each of these alternatives has several options. One of the major criteria for making the decision is justification and cost–benefit analysis.

Justification and Cost–Benefit Analysis

As the number of potential BI applications increases, the need to justify and prioritize them arises. This is not an easy task due to the large number of intangible benefits. Both direct and intangible benefits need to be identified. Of course, this is where the knowledge of similar applications in other organizations and case studies is extremely useful. For example, The Data Warehousing Institute (tdwi.org) provides a wealth of information about products and innovative applications and implementations. Such information can be useful in estimating direct and indirect benefits.

Security and Protection of Privacy

This is an extremely important issue in the development of any computerized system, especially BI that contains data that may possess strategic value. Also, the privacy of employees and customers needs to be protected.

Integration of Systems and Applications

With the exception of some small applications, all BI applications must be integrated with other systems such as databases, legacy systems, enterprise systems (particularly ERP and CRM), e-commerce (sell side, buy side), and many more. In addition, BI applications are usually connected to the Internet and many times to information systems of business partners.

Furthermore, BI tools sometimes need to be integrated among themselves, creating synergy. The need for integration pushed software vendors to continuously add capabilities to their products. Customers who buy an all-in-one software package deal with only one vendor and do not have to deal with system connectivity. But, they may lose the advantage of creating systems composed from the "best-of-breed" components.

SECTION 1.4 REVIEW QUESTIONS

1. Define *BI*.
2. List and describe the major components of BI.
3. Define *OLTP*.
4. Define *OLAP*.
5. List some of the implementation topics addressed by Gartner's report.
6. List some other success factors of BI.

1.5 Analytics Overview

The word *analytics* has largely replaced the previous individual components of computerized decision support technologies that have been available under various labels in the past. Indeed, many practitioners and academics now use the word *analytics* in place of BI. Although many authors and consultants have defined it slightly differently, one can view **analytics** as the process of developing actionable decisions or recommendations for actions based on insights generated from historical data. According to the Institute for Operations Research and Management Science (INFORMS), analytics represents the combination of computer technology, management science techniques, and statistics to solve real problems. Of course, many other organizations have proposed their own interpretations and motivations for analytics. For example, SAS Institute Inc. proposed eight levels of analytics that begin with standardized reports from a computer system. These reports essentially provide a sense of what is happening with an organization. Additional technologies have enabled us to create more customized reports that can be generated on an ad hoc basis. The next extension of reporting takes us to OLAP-type queries that allow a user to dig deeper and determine specific sources of concern or opportunities. Technologies available today can also automatically issue alerts for a decision maker when performance warrants such alerts. At a consumer level we see such alerts for weather or other issues. But similar alerts can also be generated in specific settings when sales fall above or below a certain level within a certain time period or when the inventory for a specific product is running low. All of these applications are made possible through analysis and queries on data being collected by an organization. The next level of analysis might entail statistical analysis to better understand patterns. These can then be taken a step further to develop forecasts or models for predicting how customers might respond to a specific marketing campaign or ongoing service/product offerings. When an organization has a good view of what is happening and what is likely to happen, it can also employ other techniques to make the best decisions under the circumstances. These eight levels of analytics are described in more detail in a white paper by SAS (sas.com/news/sascom/analytics_levels.pdf).

This idea of looking at all the data to understand what is happening, what will happen, and how to make the best of it has also been encapsulated by INFORMS in proposing three levels of analytics. These three levels are identified (informs.org/Community/Analytics) as descriptive, predictive, and prescriptive. Figure 1.11 presents a graphical view of these three levels of analytics. It suggests that these three are somewhat independent steps and one type of analytics applications leads to another. It also suggests that there is actually some overlap across these three types of analytics. In either case, the interconnected nature of different types of analytics applications is evident. We next introduce these three levels of analytics.

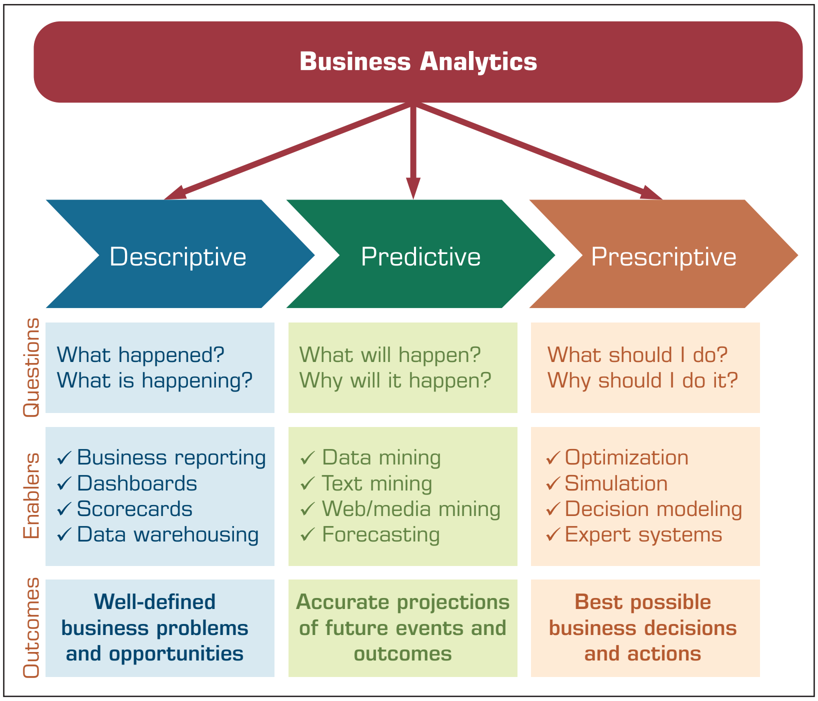

FIGURE 1.11 Three Types of Analytics.

Descriptive Analytics

Descriptive (or reporting) analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences. First, this involves the consolidation of data sources and availability of all relevant data in a form that enables appropriate reporting and analysis. Usually, the development of this data infrastructure is part of DWs. From this data infrastructure we can develop appropriate reports, queries, alerts, and trends using various reporting tools and techniques.

A significant technology that has become a key player in this area is visualization. Using the latest visualization tools in the marketplace, we can now develop powerful insights in the operations of our organization. Application Cases 1.2 and 1.3 highlight some such applications. Color renderings of visualizations discussed in these applications are available online or the book's companion Web site (dssbibook.com).

Application Case 1.2

Silvaris Increases Business with Visual Analysis and Real-Time Reporting Capabilities

Silvaris Corporation was founded in 2000 by a team of forest industry professionals to provide technological advancement in the lumber and building material sector. Silvaris is the first e-commerce platform in the United States specifically for forest products and is headquartered in Seattle, Washington. It is a leading wholesale provider of industrial wood products and surplus building materials.

Silvaris sells its products and provides international logistics services to more than 3,500 customers. To manage various processes that are involved in a transaction, they created a proprietary online trading platform to track information flow related to transactions between traders, accounting, credit, and logistics. This allowed Silvaris to share its real-time information with its customers and partners. But due to the rapidly changing prices of materials, it became necessary for Silvaris to get a real-time view of data without moving data into a separate reporting format.

Silvaris started using Tableau because of its ability to connect with and visualize live data. Due to dashboards created by Tableau that are easy to understand and explain, Silvaris started using Tableau for reporting purposes. This helped Silvaris in pulling out information quickly from the data and identifying issues that impact their business. Silvaris succeeded in managing online versus offline orders with the help of reports generated by Tableau. Now, Silvaris keeps track of online orders placed by customers and knows when to send renew pushes to which customers to keep them purchasing online. Also, analysts of Silvaris can save time by generating dashboards instead of writing hundreds of pages of reports by using Tableau.

Questions for Discussion

1. What was the challenge faced by Silvaris?
2. How did Silvaris solve its problem using data visualization with Tableau?

What We Can Learn from This Application Case

Many industries need to analyze data in real time. Real-time analysis enables the analysts to identify issues that impact their business. Visualization is sometimes the best way to begin analyzing the live data streams. Tableau is one such data visualization tool that has the capability to analyze live data without bringing live data into a separate reporting format.

Sources: Tableau.com,"Silvaris Augments Proprietary Technology Platform with Tableau's Real-Time Reporting Capabilities," http://www.tableau.com/sites/default/files/case-studies/silvaris-business-dashboards_0.pdf (accessed July 2016); Silvaris.com, "Overview," http://www.silvaris.com/About/ (accessed July 2016).

Application Case 1.3

Siemens Reduces Cost with the Use of Data Visualization

Siemens is a German company headquartered in Berlin, Germany. It is one of the world's largest companies focusing on the areas of electrification, automation, and digitalization. It has an annual revenue of 76 billion euros.

The visual analytics group of Siemens is tasked with end-to-end reporting solutions and consulting for all of Siemens internal BI needs. This group was facing the challenge of providing reporting solutions to the entire Siemens organization across different departments while maintaining a balance between governance and self-service capabilities. Siemens needed a platform that could analyze their multiple cases of customer satisfaction surveys, logistic processes, and financial reporting. This platform should be easy to use for their employees so that they can use this data for analysis and decision making. In addition, the platform should be easily integrated with existing Siemens systems and give employees a seamless user experience.

They started using Dundas BI, a leading global provider of BI and data visualization solutions. It allowed Siemens to create highly interactive dashboards that enabled Siemens to detect issues early and thus save a significant amount of money. The dashboards developed by Dundas BI helped Siemens global logistics organization answer questions like how different supply rates at different locations affect the operation, thus helping them to reduce cycle time by 12% and scrap cost by 25%.

Questions for Discussion

1. What challenges were faced by Siemens visual analytics group?
2. How did the data visualization tool Dundas BI help Siemens in reducing cost?

What We Can Learn from This Application Case

Many organizations want tools that can be used to analyze data from multiple divisions. These tools can help them improve performance and make data discovery transparent to their users so that they can identify issues within the business easily.

Sources: Dundas.com, "How Siemens Drastically Reduced Cost with Managed BI Applications," http://www.dundas.com/resource/getcasestudy?caseStudyName=09-03-2016-Siemens%2FDundas-BI-Siemens-Case-Study.pdf (accessed July 2016); Wikipedia.org, "SIEMENS," https://en.wikipedia.org/wiki/Siemens (accessed July 2016); Siemens.com, "About Siemens," http://www.siemens.com/about/en/ (accessed July 2016).

Predictive Analytics

Predictive analytics aims to determine what is likely to happen in the future. This analysis is based on statistical techniques as well as other more recently developed techniques that fall under the general category of **data mining**. The goal of these techniques is to be able to predict if the customer is likely to switch to a competitor ("churn"), what the customer would likely buy next and how much, what promotions a customer would respond to, whether this customer is a creditworthy risk, and so forth. A number of techniques are used in developing predictive analytical applications, including various classification algorithms. For example, as described in Chapters 4 and 5, we can use classification techniques such as logistic regression, decision tree models, and neural networks to predict how well a motion picture will do at the box office. We can also use clustering algorithms for segmenting customers into different clusters to be able to target specific promotions to them. Finally, we can use association mining techniques to estimate relationships between different purchasing behaviors. That is, if a customer buys one product, what else is the customer likely to purchase? Such analysis can assist a retailer in recommending or promoting related products. For example, any product search on Amazon.com results in the retailer also suggesting other similar products that a customer may be interested in. We will study these techniques and their applications in Chapters 3 through 6. Application Case 1.4 illustrates one such application in sports.