
Contents

Preface ix

Acknowledgments xiii

Chapter 1 What Is Big Data?1

The Arrival of Analytics 2

Where Is the Value? 3

More to Big Data Than Meets the Eye 5

Dealing with the Nuances of Big Data 6

An Open Source Brings Forth Tools 7

Caution: Obstacles Ahead 8

Chapter 2 Why Big Data Matters.....11

Big Data Reaches Deep 12

Obstacles Remain 13

Data Continue to Evolve 15

Data and Data Analysis Are Getting More Complex 17

The Future Is Now 18

Chapter 3 Big Data and the Business Case.....21

Realizing Value 22

The Case for Big Data 22

The Rise of Big Data Options 25

Beyond Hadoop 27

With Choice Come Decisions 28

Chapter 4 Building the Big Data Team29

- The Data Scientist 29
- The Team Challenge 30
- Different Teams, Different Goals 31
- Don't Forget the Data 32
- Challenges Remain 32
- Teams versus Culture 34
- Gauging Success 35

Chapter 5 Big Data Sources37

- Hunting for Data 38
- Setting the Goal 39
- Big Data Sources Growing 40
- Diving Deeper into Big Data Sources 42
- A Wealth of Public Information 43
- Getting Started with Big Data Acquisition 44
- Ongoing Growth, No End in Sight 46

Chapter 6 The Nuts and Bolts of Big Data47

- The Storage Dilemma 47
- Building a Platform 52
- Bringing Structure to Unstructured Data 57
- Processing Power 59
- Choosing among In-house, Outsourced, or Hybrid Approaches 61

**Chapter 7 Security, Compliance, Auditing,
and Protection63**

- Pragmatic Steps to Securing Big Data 64
- Classifying Data 65
- Protecting Big Data Analytics 66
- Big Data and Compliance 67
- The Intellectual Property Challenge 72

Chapter 8 The Evolution of Big Data77

- Big Data: The Modern Era 80
- Today, Tomorrow, and the Next Day 84
- Changing Algorithms 90

Chapter 9 Best Practices for Big Data Analytics93

- Start Small with Big Data 94
- Thinking Big 95
- Avoiding Worst Practices 96
- Baby Steps 98
- The Value of Anomalies 101
- Expediency versus Accuracy 103
- In-Memory Processing 104

Chapter 10 Bringing It All Together.....111

- The Path to Big Data 112
- The Realities of Thinking Big Data 113
- Hands-on Big Data 115
- The Big Data Pipeline in Depth 116
- Big Data Visualization 121
- Big Data Privacy 122

Appendix Supporting Data.....125

- “The MapR Distribution for Apache Hadoop” 126
- “High Availability: No Single Points of Failure” 142

About the Author 151**Index 153**