Additional praise for *Taming the Big Data Tidal Wave*

This book is targeted for the business managers who wish to leverage the opportunities that big data can bring to their business. It is written in an easy flowing manner that motivates and mentors the non-technical person about the complex issues surrounding big data. Bill Franks continually focuses on the key success factor . . . How can companies improve their business through analytics that probe this big data? If the tidal wave of big data is about to crash upon your business, then I would recommend this book.

—Richard Hackathorn, President, Bolder Technology, Inc.

Most big data initiatives have grown both organically and rapidly. Under such conditions, it is easy to miss the big picture. This book takes a step back to show how all the pieces fit together, addressing varying facets from technology to analysis to organization. Bill approaches big data with a wonderful sense of practicality—“just get started” and “deliver value as you go” are phrases that characterize the ethos of successful big data organizations.

—Eric Colson, Vice President of Data Science and Engineering, Netflix

Bill Franks is a straight-talking industry insider who has written an invaluable guide for those who would first understand and then master the opportunities of big data.

—Thornton May, Futurist and Executive Director, The IT Leadership Academy
Taming the Big Data Tidal Wave
Wiley & SAS Business Series

The Wiley & SAS Business Series presents books that help senior-level managers with their critical management decisions.

Titles in the Wiley & SAS Business Series include:

- *Activity-Based Management for Financial Institutions: Driving Bottom-Line Results* by Brent Bahnub
- *Branded! How Retailers Engage Consumers with Social Media and Mobility* by Bernie Brennan and Lori Schafer
- *Business Analytics for Customer Intelligence* by Gert Laursen
- *Business Analytics for Managers: Taking Business Intelligence beyond Reporting* by Gert Laursen and Jesper Thorlund
- *Business Intelligence Competency Centers: A Team Approach to Maximizing Competitive Advantage* by Gloria J. Miller, Dagmar Brautigam, and Stefanie Gerlach
- *Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy* by Olivia Parr Rud
- *Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors* by Clark Abrahams and Mingyuan Zhang
- *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring* by Naeem Siddiqi
- *Customer Data Integration: Reaching a Single Version of the Truth*, by Jill Dyche and Evan Levy
- *Demand-Driven Forecasting: A Structured Approach to Forecasting* by Charles Chase
- *Executive’s Guide to Solvency II* by David Buckham, Jason Wahl, and Stuart Rose
- *Fair Lending Compliance: Intelligence and Implications for Credit Risk Management* by Clark R. Abrahams and Mingyuan Zhang
- *Manufacturing Best Practices: Optimizing Productivity and Product Quality* by Bobby Hull
- *Marketing Automation: Practical Steps to More Effective Direct Marketing* by Jeff LeSueur
- *Mastering Organizational Knowledge Flow: How to Make Knowledge Sharing Work* by Frank Leistner
- *Performance Management: Finding the Missing Pieces (to Close the Intelligence Gap)* by Gary Cokins
- *Performance Management: Integrating Strategy Execution, Methodologies, Risk, and Analytics* by Gary Cokins
- *Retail Analytics: The Secret Weapon* by Emmett Cox
- *Social Network Analysis in Telecommunications* by Carlos Andre Reis Pinheiro
- *The Data Asset: How Smart Companies Govern Their Data for Business Success* by Tony Fisher
- *The Executive’s Guide to Enterprise Social Media Strategy: How Social Networks Are Radically Transforming Your Business* by David Thomas and Mike Barlow
- *The New Know: Innovation Powered by Analytics* by Thornton May
- *The Value of Business Analytics: Identifying the Path to Profitability* by Evan Stubbs
- *Visual Six Sigma: Making Data Analysis Lean* by Ian Cox, Marie A Gaudard, Philip J. Ramsey, Mia L. Stephens, and Leo Wright

For more information on any of the above titles, please visit [www.wiley.com](http://www.wiley.com).
Taming the Big Data Tidal Wave

Finding Opportunities in Huge Data Streams with Advanced Analytics

Bill Franks
This book is dedicated to Stacie, Jesse, and Danielle, who put up with all the nights and weekends it took to get this book completed.
Contents

Foreword xiii
Preface xvii
Acknowledgments xxv

PART ONE THE RISE OF BIG DATA ..................................................... 1

Chapter 1 What Is Big Data and Why Does It Matter? .......... 3

What Is Big Data? 4
Is the “Big” Part or the “Data” Part More Important? 5
How Is Big Data Different? 7
How Is Big Data More of the Same? 9
Risks of Big Data 10
Why You Need to Tame Big Data 12
The Structure of Big Data 14
Exploring Big Data 16
Most Big Data Doesn’t Matter 17
Filtering Big Data Effectively 20
Mixing Big Data with Traditional Data 21
The Need for Standards 22
Today’s Big Data Is Not Tomorrow’s Big Data 24
Wrap-Up 26
Notes 27
Chapter 2  Web Data: The Original Big Data .......................... 29
  Web Data Overview 30
  What Web Data Reveals 36
  Web Data in Action 42
  Wrap-Up 50
  Note 51

Chapter 3  A Cross-Section of Big Data Sources and the Value They Hold .................................................. 53
  Auto Insurance: The Value of Telematics Data 54
  Multiple Industries: The Value of Text Data 57
  Multiple Industries: The Value of Time and Location Data 60
  Retail and Manufacturing: The Value of Radio Frequency Identification Data 64
  Utilities: The Value of Smart-Grid Data 68
  Gaming: The Value of Casino Chip Tracking Data 71
  Industrial Engines and Equipment: The Value of Sensor Data 73
  Video Games: The Value of Telemetry Data 76
  Telecommunications and Other Industries: The Value of Social Network Data 78
  Wrap-Up 82

PART TWO  TAMING BIG DATA: THE TECHNOLOGIES, PROCESSES, AND METHODS ........................................ 85

Chapter 4  The Evolution of Analytic Scalability ................. 87
  A History of Scalability 88
  The Convergence of the Analytic and Data Environments 90
  Massively Parallel Processing Systems 93
  Cloud Computing 102
  Grid Computing 109
CONTENTS

MapReduce  110
It Isn’t an Either/Or Choice!  117
Wrap-Up  118
Notes  119

Chapter 5  The Evolution of Analytic Processes................. 121

The Analytic Sandbox  122
What Is an Analytic Data Set?  133
Enterprise Analytic Data Sets  137
Embedded Scoring  145
Wrap-Up  151

Chapter 6  The Evolution of Analytic Tools and Methods ............................................................ 153

The Evolution of Analytic Methods  154
The Evolution of Analytic Tools  163
Wrap-Up  175
Notes  176

PART THREE  TAMING BIG DATA: THE PEOPLE AND APPROACHES.............................................. 177

Chapter 7  What Makes a Great Analysis?......................... 179

Analysis versus Reporting  179
Analysis: Make It G.R.E.A.T.!  184
Core Analytics versus Advanced Analytics  186
Listen to Your Analysis  188
Framing the Problem Correctly  189
Statistical Significance versus Business Importance  191
Samples versus Populations  195
Making Inferences versus Computing Statistics  198
Wrap-Up  200
Chapter 8  What Makes a Great Analytic Professional? ..... 201

Who Is the Analytic Professional?  202
The Common Misconceptions about Analytic Professionals  203
Every Great Analytic Professional Is an Exception  204
The Often Underrated Traits of a Great Analytic Professional  208
Is Analytics Certification Needed, or Is It Noise?  222
Wrap-Up  224

Chapter 9  What Makes a Great Analytics Team? ............... 227

All Industries Are Not Created Equal  228
Just Get Started!  230
There’s a Talent Crunch out There  231
Team Structures  232
Keeping a Great Team’s Skills Up  237
Who Should Be Doing Advanced Analytics?  241
Why Can’t IT and Analytic Professionals Get Along?  245
Wrap-Up  247
Notes  248

PART FOUR  BRINGING IT TOGETHER: THE ANALYTICS CULTURE ................................................................. 249

Chapter 10  Enabling Analytic Innovation ...................... 251

Businesses Need More Innovation  252
Traditional Approaches Hamper Innovation  253
Defining Analytic Innovation  255
Iterative Approaches to Analytic Innovation  256
Consider a Change in Perspective  257
Are You Ready for an Analytic Innovation Center?  259
Wrap-Up  269
Note  270
Chapter 11 Creating a Culture of Innovation and Discovery ............................................................ 271

Setting the Stage 272
Overview of the Key Principles 274
Wrap-Up 290
Notes 291

Conclusion: Think Bigger! 293

About the Author 295

Index 297