

Wiley CIO Series

Foreword by
JIM STOGDILL
*General Manager,
Radar,
O'Reilly Media*



BIG DATA **BIG ANALYTICS**

**EMERGING BUSINESS INTELLIGENCE AND
ANALYTIC TRENDS FOR TODAY'S
BUSINESSES**

Michael Minelli • Michele Chambers • Ambiga Dhiraj

BIG DATA, BIG ANALYTICS



WILEY CIO SERIES

Founded in 1807, John Wiley & Sons is the oldest independent publishing company in the United States. With offices in North America, Europe, Asia, and Australia, Wiley is globally committed to developing and marketing print and electronic products and services for our customers' professional and personal knowledge and understanding.

The Wiley CIO series provides information, tools, and insights to IT executives and managers. The products in this series cover a wide range of topics that supply strategic and implementation guidance on the latest technology trends, leadership, and emerging best practices.

Titles in the Wiley CIO series include:

The Agile Architecture Revolution: How Cloud Computing, REST-Based SOA, and Mobile Computing Are Changing Enterprise IT by Jason Bloomberg

Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses by Michele Chambers, Ambiga Dhiraj, and Michael Minelli

The Chief Information Officer's Body of Knowledge: People, Process, and Technology by Dean Lane

CIO Best Practices: Enabling Strategic Value with Information Technology by Joe Stenzel, Randy Betancourt, Gary Cokins, Alyssa Farrell, Bill Flemming, Michael H. Hugos, Jonathan Hujsak, and Karl D. Schubert

The CIO Playbook: Strategies and Best Practices for IT Leaders to Deliver Value by Nicholas R. Colisto

Enterprise IT Strategy, + Website: An Executive Guide for Generating Optimal ROI from Critical IT Investments by Gregory J. Fell

Executive's Guide to Virtual Worlds: How Avatars Are Transforming Your Business and Your Brand by Lonnie Benson

Innovating for Growth and Value: How CIOs Lead Continuous Transformation in the Modern Enterprise by Hunter Muller

IT Leadership Manual: Roadmap to Becoming a Trusted Business Partner by Alan R. Guibord

Managing Electronic Records: Methods, Best Practices, and Technologies by Robert F. Smallwood

On Top of the Cloud: How CIOs Leverage New Technologies to Drive Change and Build Value Across the Enterprise by Hunter Muller

Straight to the Top: CIO Leadership in a Mobile, Social, and Cloud-based (Second Edition) by Gregory S. Smith

Strategic IT: Best Practices for IT Managers and Executives by Arthur M. Langer

Strategic IT Management: Transforming Business in Turbulent Times by Robert J. Benson

Transforming IT Culture: How to Use Social Intelligence, Human Factors and Collaboration to Create an IT Department That Outperforms by Frank Wander

Unleashing the Power of IT: Bringing People, Business, and Technology Together by Dan Roberts

The U.S. Technology Skills Gap: What Every Technology Executive Must Know to Save America's Future by Gary Beach

BIG DATA, BIG ANALYTICS

EMERGING BUSINESS INTELLIGENCE AND
ANALYTIC TRENDS FOR TODAY'S BUSINESSES

Michael Minelli
Michele Chambers
Ambiga Dhiraj



WILEY

John Wiley & Sons, Inc.

Cover image: © nobeastsofierce/Alamy
Cover design: John Wiley & Sons, Inc.

Copyright © 2013 by Michael Minelli, Michele Chambers, and Ambiga Dhiraj. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the Web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Minelli, Michael, 1974-

Big data, big analytics : emerging business intelligence and analytic trends for today's businesses / Michael Minelli, Michele Chambers, Ambiga Dhiraj.

pages cm

Includes bibliographical references and index.

ISBN 978-1-118-14760-3 (cloth); ISBN 978-1-118-22583-7 (ebk);

ISBN 978-1-118-23915-5 (ebk); ISBN 978-1-118-26381-5 (ebk)

1. Business intelligence. 2. Information technology. 3. Data processing.

4. Data mining. 5. Strategic planning. I. Chambers, Michele. II. Dhiraj, Ambiga, 1975-III. Title.

HD38.7.M565 2013

658.4'72—dc23

2012044882

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*To my wife Jenny and our three incredible children,
Jack, Madeline, and Max. Also to my parents, who have
always been there for me.*

—Mike

*To my son Cole, who is the light of my life and the person
who taught me empathy. Also to my adopted family and
support system, Lisa Patrick, Pei Yee Cheng, and Patrick
Thean. Finally, to my colleagues Bill Zannine, Brian
Hess, Jon Niess, Matt Rollender, Kevin Kostuik, Krishnan
Parasuraman, Mario Inchiosa, Thomas Baeck, Thomas
Dinsmore, and Usama Fayyad, for their generous support.*

—Michele

*To Mu Sigmans all around the world for their passion
toward building the decision sciences industry.*

—Ambiga

CONTENTS

FOREWORD xiii

PREFACE xix

ACKNOWLEDGMENTS xxi

CHAPTER 1 What Is Big Data and Why Is It Important? 1

A Flood of Mythic “Start-Up” Proportions 4

Big Data Is More Than Merely Big 5

Why Now? 6

A Convergence of Key Trends 7

Relatively Speaking . . . 9

A Wider Variety of Data 10

The Expanding Universe of Unstructured Data 11

Setting the Tone at the Top 15

Notes 18

CHAPTER 2 Industry Examples of Big Data 19

Digital Marketing and the Non-line World 19

Don’t Abdicate Relationships 22

Is IT Losing Control of Web Analytics? 23

Database Marketers, Pioneers of Big Data 24

Big Data and the New School of Marketing 27

Consumers Have Changed. So Must Marketers. 28

The Right Approach: Cross-Channel Lifecycle Marketing 28

Social and Affiliate Marketing 30

Empowering Marketing with Social Intelligence 31

Fraud and Big Data 34

Risk and Big Data 37

Credit Risk Management 38

Big Data and Algorithmic Trading 40

Crunching Through Complex Interrelated Data 41

Intraday Risk Analytics, a Constant Flow of Big Data 42

Calculating Risk in Marketing	43
Other Industries Benefit from Financial Services' Risk Experience	43
Big Data and Advances in Health Care	44
"Disruptive Analytics"	46
A Holistic Value Proposition	47
BI Is Not Data Science	49
Pioneering New Frontiers in Medicine	50
Advertising and Big Data: From Papyrus to Seeing Somebody	51
Big Data Feeds the Modern-Day Donald Draper	52
Reach, Resonance, and Reaction	53
The Need to Act Quickly (Real-Time When Possible)	54
Measurement Can Be Tricky	55
Content Delivery Matters Too	56
Optimization and Marketing Mixed Modeling	56
Beard's Take on the Three Big Data Vs in Advertising	57
Using Consumer Products as a Doorway	58
Notes	59

CHAPTER 3 Big Data Technology 61

The Elephant in the Room: Hadoop's Parallel World	61
Old vs. New Approaches	64
Data Discovery: Work the Way People's Minds Work	65
Open-Source Technology for Big Data Analytics	67
The Cloud and Big Data	69
Predictive Analytics Moves into the Limelight	70
Software as a Service BI	72
Mobile Business Intelligence is Going Mainstream	73
Ease of Mobile Application Deployment	75
Crowdsourcing Analytics	76
Inter- and Trans-Firewall Analytics	77
R&D Approach Helps Adopt New Technology	80
Adding Big Data Technology into the Mix	81
Big Data Technology Terms	83
Data Size 101	86
Notes	88

CHAPTER 4 Information Management 89

- The Big Data Foundation 89
- Big Data Computing Platforms (or Computing Platforms That Handle the Big Data Analytics Tsunami) 92
- Big Data Computation 93
- More on Big Data Storage 96
- Big Data Computational Limitations 96
- Big Data Emerging Technologies 97

CHAPTER 5 Business Analytics 99

- The Last Mile in Data Analysis 101
- Geospatial Intelligence Will Make Your Life Better 103
- Listening: Is It Signal or Noise? 106
- Consumption of Analytics 108
- From Creation to Consumption 110
- Visualizing: How to Make It Consumable? 110
- Organizations Are Using Data Visualization as a Way to Take Immediate Action 116
- Moving from Sampling to Using All the Data 121
- Thinking Outside the Box 122
- 360° Modeling 122
- Need for Speed 122
- Let's Get Scrappy 123
- What Technology Is Available? 124
- Moving from Beyond the Tools to Analytic Applications 125
- Notes 125

CHAPTER 6 The People Part of the Equation 127

- Rise of the Data Scientist 128
- Learning over Knowing 130
- Agility 131
- Scale and Convergence 131
- Multidisciplinary Talent 131
- Innovation 132
- Cost Effectiveness 132

Using Deep Math, Science, and Computer Science	133
The 90/10 Rule and Critical Thinking	136
Analytic Talent and Executive Buy-in	137
Developing Decision Sciences Talent	139
Holistic View of Analytics	140
Creating Talent for Decision Sciences	142
Creating a Culture That Nurtures Decision Sciences Talent	144
Setting Up the Right Organizational Structure for Institutionalizing Analytics	146

CHAPTER 7 Data Privacy and Ethics 151

The Privacy Landscape	152
The Great Data Grab Isn't New	152
Preferences, Personalization, and Relationships	153
Rights and Responsibility	154
Playing in a Global Sandbox	159
Conscientious and Conscious Responsibility	161
Privacy May Be the Wrong Focus	162
Can Data Be Anonymized?	164
Balancing for Counterintelligence	165
Now What?	165
Notes	167

CONCLUSION 169

RECOMMENDED RESOURCES 175

ABOUT THE AUTHORS 177

INDEX 179

FOREWORD: BIG DATA AND CORPORATE EVOLUTION

When my friend Mike Minelli asked me to write this foreword I wasn't sure at first what I should put on paper. Forewords are often one part book summary and one part overview of the field. But when I read the draft Mike sent me I realized that this is a really good book, and it doesn't need either of those. Without any additional help from me it will give you plenty of insight into what is happening and why it's happening now, and it will help you see the possibilities for your industry in this transition to a data-centric age. Also, the book is just full of practical suggestions for what you can do about them. But perhaps there's an opportunity to establish a wider context. To explore what Big Data means across a broad arc of technological advancement. So rather than bore you with a summary of a book you're going to read anyway, I'll try to daub a bit of paint onto the big picture of what it all might mean.

This foreword is based on the thesis that Big Data isn't merely another technology. It isn't just another gift box en route to the world's systems integrators via the conveyor belt of Gartner hype cycles. I believe Big Data will follow digital computing and internetworking to take its place as the third epoch of the information age, and in doing so it will fundamentally alter the trajectory of corporate evolution. The corporation is about to undergo a change analogous to the rise of consciousness in humans.

So let's start at the beginning. The Industrial Age was an era of vast changes in society. We harnessed first steam and then electricity as prime movers to unleash astonishing increases in productivity. The result was the first sustained growth of wealth in human history.

Those early industrial concerns required vast pools of labor that gradually grew more specialized. To coordinate the efforts of all of those people, management developed systems of rules and hierarchy of authority. At massive scale the corporation was no longer the direct exercise of an owner's will, it was a kind of organism.

It was an organism whose systems of control were born out of the Napoleonic bureaucracy of the French State and its emphasis on specialized function, fixed rules, and rigid hierarchy. The "bureau" in bureaucracy literally means desk, and paper was both the storage mechanism in them and the signaling mechanism between them.