with WEBSITE

CLOUDONOMICS

The Business Value of Cloud Computing

Joe Weinman

Additional praise for

Cloudonomics: The Business Value of Cloud Computing

It is a business imperative to do more with less—and do everything faster. *Cloudonomics* offers a much-appreciated framework for sorting through cloud computing options and the marketing hype.

—Lorraine Cichowski, SVP and CIO, Associated Press

Joe Weinman is one of the foremost thinkers in cloud computing. He has captured a topic of fierce complexity and expressed it with elegant prose and simple, powerful, and compelling mathematical rigor. Weinman delivers his arguments with a clarity and logic that is unassailable. *Cloudonomics* should be required reading for every enterprise CIO seeking a way through the clutter and hype of vendors' cloud solutions and looking for a set of crystal clear, mathematically grounded, and meticulously presented arguments that show the way through the morass of cloud computing.

—Simon Crosby, PhD, CTO, Bromium; founder and former CTO, XenSource; former CTO, Data Center and Virtualization, Citrix

Cloudonomics is a seminal work on cloud based on an axiomatic mathematical theory, and not on popular opinions and baseless assumptions. Based on rigorous quantitative analysis, the book is amazingly simple to read with real-world examples in a lucid language that will resonate with both the technical and business professional. This book is a must-read for every professional interested in cloud and is an invaluable reference for any advanced course on cloud computing.

—Ravi Rajagopal, VP, Cloud Strategy and Solutions, CA Technologies; Adjunct Professor, New York University

Joe Weinman's masterful book looks past the hype to offer new insights into the impact of cloud computing. *Cloudonomics* is must-reading for anyone interested in a more analytically based understanding of the cloud's transformative potential.

—Christopher S. Yoo, John H. Chestnut Professor of Law, Communication, and Computer & Information Science and founding director of the Center of Technology, Innovation and Competition, University of Pennsylvania

The cloud is redefining how technology is being used for businesses. *Cloudonomics* is a book that will help you understand the economics and importance of this change and what it means for your industry.

—Om Malik, founder, GigaOM & Structure: the Cloud Computing conference

Cloudonomics

The Business Value of Cloud Computing

JOE WEINMAN

John Wiley & Sons, Inc.

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

Library of Congress Cataloging-in-Publication Data:

Weinman, Joe, -
Cloudonomics: the business value of cloud computing / Joe Weinman.
p. cm.
Includes index.
ISBN 978-1-118-22996-5 (pbk); 978-1-118-28696-8 (ebk); 978-1-118-28402-5 (ebk); 978-1-118-28288-5 (ebk);
1. Cloud computing—Economic aspects. 2. Information technology—Management. I. Title.
QA76.585.W45 2012
004.6782—dc23

2012012399

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Paige, Ali, and Sierra

Contents

Preface

In the course of human history, there have been a number of bona fide revolutions in the interdependent arenas of technology, society, religion, economics, and politics: flint tools, money, writing, agriculture, democracy, printing, steam power, capitalism, mass production, telephony, and electricity, to name a few. We are 65 or 70 years into one such revolution—the *information age*—which has permeated every corner of the earth and beyond—from video games to war games to baseball games, and from first-world stock markets to third-world fish markets, to out-of-this-world interstellar probes.[1] Oh yes, and musical greeting cards, talking dolls, and intelligent thermostats too.[2]

Does the advanced age of this advanced age signal impending retirement? Some argue that "the opportunities for gaining IT-based advantages are already dwindling,"[3]; however, this sounds suspiciously similar to alleged pronouncements, such as "everything that can be invented has been invented" or "there is a world market for maybe five computers."[4]

The revolution is accelerating, not slowing.

Technologies such as quantum computing, digital electro-holographic displays, brain-computer interfaces, natural-language interaction via speech-to-text and semantic processing, homomorphic encryption, and new electronic components such as HP's nanoscale memristor, Intel's

[1] Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (Norton, 2003). Robert Jensen, "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *Quarterly Journal of Economics* 122, No. (2007): 879–924.

[2] www.nest.com/living-with-nest/.

[3] Nicholas Carr, "IT Doesn't Matter," *Harvard Business Review* (May 2003). 81(5) 41–49.

[4] These quotes are of questionable authenticity. See Kevin Maney, "Tech Titans Wish We Wouldn't Quote Them on This Baloney," *USA Today*, July 5, 2005. www.usatoday.com/tech/columnist/kevinmaney/2005-07-05-famous-quotes_x.htm.

three-dimensional chips, and on-chip optical interconnects are still in their infancy. Innovative cognitive computers are now being designed by IBM to use "brainlike" neurosynaptic chips. Other disturbing anomalies, from quantum entanglement to apparently faster-than-light neutrinos, may form the foundation for future disruption.[5]

In this ocean of innovation, *cloud computing* is the latest of successive waves that have eroded the shoreline of prior paradigms, such as the mainframe, the minicomputer, and the personal computer.

Cloud computing is a tsunami of transformation exemplifying Schumpeterian creative destruction: amassing immense wealth for companies that didn't exist a few years ago—such as Google, Facebook, Amazon, Salesforce.com, and Zynga; disrupting long-standing business models and ecosystems including publishing, advertising, television, the recording industry, telecommunications, and retailing; and reordering relationships within the computing industry: among hardware vendors, licensed software vendors, distributors, value-added resellers, and systems integrators, to name a few.

The cloud is both an existential threat and an irresistible opportunity. Virtually any summary of key trends or chief information officer (CIO) focus areas ranks cloud computing at or near the top of the list. A recent Gartner survey of 2,000 CIOs places cloud computing as the number-one technology priority.[6] Most if not all of the rest of the top priorities—virtualization, mobility, collaboration, business intelligence—enable, are enabled by, or otherwise relate to the cloud.

Wharton fellow and author Jeremy Rifkin would consider this to be a natural consequence of "the Age of Access."[7] He has argued that the market economy—in which people own and trade goods—is being replaced by the network economy—where people pay to access them. Why bother owning something if you can access it anytime, anywhere? People don't want drills but the holes that they make; people don't want CDs or applications but the music or functionality that they provide.

Although Rifkin has positioned this trend as something new, in many respects it represents a return to a prior age. After all, the idea of content ownership is relatively new: Before *owning* records, CDs, or MP3s, people *accessed* audio content via the concert hall or radio; before VHS, Beta, or

---

[5] Adrian Cho, "From Geneva to Italy Faster than a Speeding Photon?" *Science*, September 30, 2011, Vol. 333, No. 6051, p. 1809.

[6] Gartner, "Gartner Executive Programs Worldwide Survey of More than 2,000 CIOs Identifies Cloud Computing as Top Technology Priority for CIOs in 2011," *Gartner Newsroom*, January 21, 2011. http://www.gartner.com/it/page.jsp?id=1526414.

[7] Jeremy Rifkin, *The Age of Access: The New Culture of Hypercapitalism, Where All of Life Is a Paid-For Experience* (Tarcher, 2001).

DVDs, there was broadcast and cable television, movie theaters, and, even earlier, plays and operas. In the Age of Access 2.0, however, the logic, characteristics, and payment models are certainly changing. Rather than traveling to theaters and opera halls, the content comes to you; unlike broadcast, it is personalized, contextualized, and on demand.

So the cloud is the *new*, new thing, but what it actually *is*, is a matter of disagreement. Cloud computing, so named by Ram Chellappa of Emory University, is, at a high level, computing that is done *somewhere* out there in an undisclosed location away from your own laptop, smartphone, tablet, or data center.[8] The cloud model applies to the discovery and acquisition of applications, services, and content, such as eBooks from ebookstores; tablet and smartphone apps from app stores; songs from Lady Gaga on Vevo or blockbuster movies on demand; and customer relationship management software executed far from your device. The cloud is at the heart of social networks such as Facebook and LinkedIn, social games such as Zynga's FarmVille, microblogs such as Twitter, and texting, messaging, and mail such as AOL Instant Messenger, RIM's BlackBerry Messenger, Microsoft Hotmail, and Google Gmail. But it also applies to the core infrastructure (computer servers and data storage), the utility software (middleware and databases), and the currency of the digital economy, "big data" (petabytes of information), that together enable those applications to run at the scale of millions or hundreds of millions of users.

Like the parable of the blind men describing an elephant—one feels the legs and says the elephant is like a pillar, the other the tail and says it is like a rope—or perhaps like real cloud gazers—one sees a rabbit, one his Aunt Martha—everyone has a different perspective on the cloud. Some see a new technology, say, virtualization; others, a new silo-busting integrated development and operations model; others, a means of delivering software functionality; still others, an ecosystem that enables applications spanning mobile devices, networks, and cloud-based resources and services. They are all correct in their own way.

For the purposes of this book, however, we consider cloud computing primarily from the business, financial, and economic perspective: *Cloudonomics*, to use the term I coined in the summer of 2008 for Cloudonomics.com and a blog post for the popular technology site GigaOM.com, syndicated to *BusinessWeek*.[9] As such, we consider core characteristics of

[8] www.bus.emory.edu/ram/.

[9] Joe Weinman, "The 10 Laws of Cloudonomics," *GigaOM.com*, September 7, 2008, http://gigaom.com/2008/09/07/the-10-laws-of-cloudonomics/. Joe Weinman, "The 10 Laws of Cloudonomics," *BusinessWeek*, September 6, 2008, www.businessweek.com/technology/content/sep2008/tc2008095_942690.htm.

the cloud—on-demand resources, usage-based charging, resource sharing, geographic dispersion, and the like—and how they map to and drive business—and even societal—value.

I would claim that such a perspective is one of the most important ways to evaluate and exploit the cloud, since *unless a technology drives compelling value, it will end up in the dustbin of history*. Remember the CueCat barcode reader?[10]

The value of the core characteristics of the cloud has been proven time and again in domain after domain: hotel chains, airlines, electric utilities, taxi services, manufacturing service providers, and others. Taxis offer transportation capacity on an on-demand, pay-per-use basis. Banks rent resources as well—principal—on a pay-per-use basis—interest. Companies can buy workers' services from what could be called the "labor cloud" either on a flat-rate basis—a salary—or a pay-per-use basis—hourly wages.

Ubiquitous access and location independence are key. When you order a *physical* book from Amazon.com or DVD from Netflix, you don't really need to know what distribution center it was sent from as long as it arrives on time, and when you order an eBook from Amazon.com or stream a video from Netflix, you don't really need to know what data center it came from. Either way, you trust that the provider has figured out the appropriate locations to get you what you want within the time frame that you would like and to which you agree via the terms of service.

In the pay-per-use model of the cloud, we see the same charging model used by hotels and barber shops. In on-demand provisioning, we see the same resource allocation strategy used in accessing energy resources by turning on an electric switch, or financial resources by tapping into a home equity line of credit. Geographically dispersed data centers and content-delivery networks echo the approach used by coffee shops and fast food chains to distribute their wares globally. Resource sharing of computer servers in a cloud data center is not that different from sharing servers—waiters and waitresses—in a restaurant.

These are more than casual analogies; the point is that the same immutable principles—say, resource utilization improvements from demand aggregation or diminishing returns from investments in geographic dispersion for latency reduction—apply *regardless* of domain. I call these the *Laws of Cloudonomics*. The Laws of Cloudonomics are not restricted to cloud computing any more than the Law of Gravity is restricted to apples.

From these parallels—and an analysis of underlying, abstract models—we can determine that there are quite a few characteristics and behaviors

[10] Dan Tynan, "The 25 Worst Tech Products of All Time," *PCWorld*, May 26, 2006. www.pcworld.com/article/125772-8/the_25_worst_tech_products_of_all_time.html.

that contravene simplistic thinking, a sort of freakonomics of the cloud. For example, branch expansion is doomed to fail eventually. Rational customers often should be *delighted* to pay *more* for cloud services—really—even when there are *no differences* in characteristics such as performance or security. It can cost nothing to accelerate performance. Even after both heavy and light users switch rate plans from flat rate to pay-per-use or vice versa to save money, a firm or industry can maintain revenues. As the cost of IT plummets, IT spend will stay the same—or increase.

This book doesn't focus on industry market projections or vendor offerings but rather on strategy, business models, customer value, and their relationships. The intent of this book is to be multidisciplinary, seminal, evergreen, rigorous, forward-looking, and irreverent and to appeal to a broad range of customers, prospects, strategists, venture capitalists, investors, technologists, executives, service providers, and academics, both within the field of cloud computing and beyond.

This book is first and foremost *multidisciplinary*, drawing on illustrative industry examples but also a broad range of fields: strategy, economics, psychology, system dynamics, calculus, statistics, computing technology, and theoretical computer science, with forays into botany, biology, and physics. Most readers should find most of the book quite readable, and experts in a variety of fields should find the breadth of *Cloudonomics* of interest.

Second, I hope that much of this book is *seminal*. I believe I've been first to explore a number of areas as they relate to the cloud, such as the architectural implications of cost optimization, analysis of latency for interactive applications using packing of spherical caps, and computational intractability of networked resource allocation.

Third, rather than conducting a *Consumer Reports*–style analysis of ratings of the different vendors and service providers that would be out-of-date before the book is even published, the intent is for the book to be *evergreen* (i.e., usable for a long time to come) in assessing architectural and business alternatives, developing new business models, and incorporating cloud into your own business strategy.

The suffix "-nomics" has been used for important insights into business trends and strategy—Eric Qualman's *Socialnomics* on social media, Ken Doctor's *Newsono*mics regarding the digitalization of news, and Don Tapscott and Anthony Williams' *Wikinomics* covering IT-based collaboration—to the more quantitative *Freakonomics*, presidential economic policies such as Reaganomics, Clintonomics, and Obamanomics, and humorous insurance advertising—Geckonomics. However, *Cloudonomics* is quantitatively rigorous. I've skimmed the surface of much of the math for this volume. However, it is worth noting that most claims are supported by detailed proofs. My research has taken me far afield of "the cloud" and uncovered unusual connections. For example, cell site or content delivery

network node placement is related to cannonball stacking, the Tammes problem—solved by evolution—of designing pollen spores, and the Thomson problem of minimum energy electron configurations, which in turn relates to baryon density isosurfaces of Skyrmions, whatever they are.[11]

References are provided to more detailed papers and to my simulation Web site: ComplexModels.com, which provides easy-to-use Monte Carlo simulations illustrating the Laws of Cloudonomics.

Of course, being overly quantitative can lead to precise yet *incorrect* conclusions due to quirks, biases, and anomalies in human behavior, which I will touch on. These experiments have led to more than one Nobel Prize in economics[12] and provide fascinating insights into what Duke University professor Dan Ariely calls "predictable irrationality."[13]

Cloudonomics posits forward-looking scenarios for cloud computing and industry ecosystem evolution and revolution. For example, braking at high speed through a turn on a wet road used to be the domain of professional stunt drivers; now any driver can perform at the same level using standard antilock braking systems (ABS), electronic stability control, and traction control. Similarly, proprietary virtual server provisioning tools, for example, are not likely to offer sustainable competitive advantage to a cloud service provider. This technology is diffusing just like ABS.

Finally, the book is *irreverent*, challenging conventional wisdom. The truth in cloud is often counterintuitive and nuanced; better mental models mean better business and technology strategies and investments.

I hope to provide a lens to view the world of cloud computing economics, so that you can consider my questions and evaluate my arguments but draw your own conclusions. I do not necessarily provide all the answers but present models with which you can perceive, comprehend, and thereby exploit the cloud for your own uses, whether as a customer, service provider, equipment manufacturer, software vendor, venture capitalist, or investor. This is not the end of the dialogue, but the beginning.

---

[11] Gerald Edward Brown and Mannque Rho, *The Multifaceted Skyrmion* (World Scientific Publishing, 2010).

[12] Technically, the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, nobelprize.org.

[13] Dan Ariely, *Predictably Irrational: The Hidden Forces that Shape Our Decisions* (HarperCollins, 2008).

Acknowledgments

A book like this owes so much to so many that it is impossible to fully trace the directed causal graph. First and foremost, however, I'd like to thank the wonderful team at John Wiley & Sons, Inc. Sheck Cho, executive editor at Wiley, immediately saw the potential of this book and has been extremely professional, flexible, insightful, transparent, collaborative, and patient. I'd also like to thank the rest of the terrific Wiley team, especially Natasha Andrews-Noel, Stacey Rivera, and Helen Cho, who helped make this book a reality. Thanks to Bennett Ruiz of AT&T, Barrie Sosinsky, and Hunter Muller of HMG Strategy for helping make the connection with Wiley. And thanks to Zick Rubin and Brenda Ulrich at the Law Office of Zick Rubin, who were both knowledgeable and responsive.

I believe that a book such as this is immeasurably enriched by data. Beyond the extensive references, thanks are particularly due Greg Orelind of Alexa, who kindly permitted the use of the Alexa pageview data illustrating demand variability; Marty Kagan and Greg Unrein of Cedexis for HTTP response time data; Ali Kafel and Dave Connolly of Sonus Networks; Stephan Beckert and Olivia Vandenbussche of TeleGeography; and James Miller of the FCC.

The usual disclaimers apply; I take full responsibility for any errors, which, sadly, have a nonzero probability of existing in a book of this scope.

Any delineation of the main causal path of events leading to my involvement in the cloud would have to include Eric Shepcaro, Allan Leinwand, Om Malik, and Alistair Croll. I worked for Eric beginning at the turn of the millennium when he was AT&T's senior vice president of application networking. We were pioneers in introducing new hosting services—called utility computing at the time—and, thanks to Eric, I was an active participant in Don Tapscott's IT & Competitive Advantage program—the syndicated research effort that led to *Wikinomics* and included in-depth collaboration with a host of thought leaders: Don himself, David Ticoll, Joe Pine, Anthony Williams, Rena Granofsky, Paul Strassmann, Erik Brynjolfsson, Charlie Fine, Mike Dover, and others.

Eric also introduced me to Chris Albinson and Allan Leinwand of Panorama Capital, who invited me to join their Technology Advisory Board. Allan also introduced me to Om Malik. My first official cloud event was Om's Structure, in June 2008, where I was on a panel moderated by Alistair, who asked a number of thought-provoking questions, which in turn led to my first blog posts for GigaOM.com, including "The 10 Laws of Cloudonomics." This book is a 100,000-word-plus expansion of the "Laws" post and a number of others I've done for GigaOM.com.

Om is the epicenter of the cloud, between his own social and professional network, the focus GigaOM.com and GigaOM Pro put on it, and the Structure event, which I've had the pleasure of participating in as MC, moderator, and panelist since its inception. At Giga Omni Media, I've had the good fortune to work with Paul Walborsky, Stacey Higginbotham, Surj Patel, Derrick Harris, Carolyn Pritchard, Celeste LeCompte, Mike Sly, a host of cloud innovators and executives, and the magnificent Magnify Communications team: Stacey Tomlinson, Erin McMahon Lyman, and Jill Short Milne.

Extra-special thanks to Carl Brooks (now at The 451 Group) and Jo Maitland (now at GigaOM) of TechTarget, who were overly kind in naming me to their prestigious list of Cloud Computing Leaders.

I'd also like to thank UBM TechWeb, which owns and operates Interop, Cloud Connect, Light Reading, Heavy Reading, *InformationWeek*, and the former *Business Communications Review* (*BCR*, now moved online as No Jitter). Particular thanks go to Alistair Croll of Bitcurrent, Eric Krapf of *No Jitter*, and Blair Klein of AT&T. Alistair is always thinking several steps ahead of everyone else, and has been organizing both the Interop cloud summits as well as TechWeb's Cloud Connect, where I've served as track chair of the cloud economics track and had the opportunity to learn from many cloud thought leaders and to work with the TechWeb events leadership and crew: Lenny Heymann, Steve Wylie, Manuela Farrell, Andy Saldana, Paige Finkelman, Amy Jones, and Emily Johnson. Blair is AT&T's social media expert and also PR lead; it was she who originally asked me "Have you ever considered writing anything for external publication?" and suggested *BCR*; Eric Krapf was then editor in chief and published my first cloud article—"The Evolution of Networked Computing Utilities"—before cloud was cool. At Light Reading, special thanks to Ray Le Maistre and Carol Wilson, at *InformationWeek*, John Foley and Charlie Babcock.

At AT&T Labs, I was also fortunate to work with Ralph Wyndrum, Norm Shaer, Rick Kent, Rudy Alexander, Bernie McElroy, Stan Quintana, Steve Fisher, Sam Glazer, Clayton Lockhart, Tom Siracusa, Dave Belanger, Chuck Kalmanek, Ed Amoroso, Rick Schlichting, and dozens of other technical experts and executives and to interact with literally thousands of AT&T customer executives and their teams—in groups ranging from 1 to 1,000—thanks to executive/sales senior leaders Ron Spears, Donna

Henderson, John Finnegan, Bennett Ruiz, Andrea Messineo, Norihiko Minato, Gopi Gopinath, Bernard Yee, and Andrew Edison; countless excellent country managers; regional sales vice presidents; sales teams; and PR, marketing, events, and legal leads Greg Brutus, Peter J. Butler, June Chan, Fenny Fang, Karen Ko, Mary Beth Asher, Dagmar Hettler-Gentil, Niall Hickey, Donna Cobb, Eileen Whelan, Andrea Montesano, Christine O'Leary, Linda Plesnick, Don Parente, Joanne Murphy, Wendy Weinstein, Janet Wyles, Sara Vincent, and Karen Wentworth. I enjoyed working with AT&T strategy, hosting, and cloud product management executives Bill Archer, Steve Sobolevitch, Steve Mucchetti, Steve Caniano, Chris Costello, Tim Connors, Amy Anderson, Toby Ford, and others. I was fortunate to work with the team at Fleishman Hillard, including Morri Berman, Patrick Yu, Gioconda Beekman, Winnie Leung, and Brad Mays.

At HP, I had the pleasure of working with great colleagues: Erwan Menard, Sandeep Johri, Dave Shirk, Rick Halton, Russ Daniels, Dave Collins, Sybille Schieg-Heimann, Reem El-Tonsy, Julia Ochinero, Andrea Nicole Garcia, Paul Battaglia, Blaithin Underhill, Rebecca Lawson, and Emil Sayegh, as well as my global team and business and technical colleagues. Special thanks to Jujhar Singh for his support in permitting me to work on the book in my spare time and to Jan Tarantino for her critical role. As at AT&T, at HP I also had the pleasure of working with a large group of excellent regional sales VPs, country managers, account teams, and cloud experts too numerous to name. And I also had the superb support of IVORY Europe: Andrea Lee, Kristina Dalborg, Robbie Crittall, and Harry Whitbread.

I've been involved in dozens of cloud initiatives, events, planning committees, and conferences that have helped introduce me to customers with unique problems and insights and cloud thought leaders and innovators. In addition to GigaOM Structure and Techweb/UBM events, I've been fortunate to participate in numerous events and initiatives. Some highlights: Debbie Landa and her team at Dealmaker Media/Under the Radar; George Gilder and Telecosm; Sharon Nakama, Gary Kim, Anamarcia Lacayo, and the rest of the team at the Pacific Telecommunications Council; Frank Gens at IDC and John Gallant at IDG Enterprise; Karen Tucker and the Churchill Club; Stuart Sharrock and Katz Kiely of ICIN and ITU; Tier 1/The 451 Group; CDM Media; TM Forum; IIR/Informa; Cloud Expo; Milken Global Institute; Mobile World Congress; BSS/OSS World; SIIA/All About the Cloud; GDS International; Argyle Executive Forum; Simon Torrance and STL Partners/Telco 2.0; Capacity Media; SWIFT/Sibos, and IEEE Technology Time Machine conference planners including Rico Malvar, Gerhard Fettweis, Maurizio Decina, and Roberto DeMarco.

I've also learned immensely from the academic/technical community: Special thanks to Ravi Rajagopal, a VP at CA Technologies but also an adjunct professor at NYU Polytechnic and the first to use my Cloudonomics

papers in academia; Christopher Yoo, polymath professor of law at the University of Pennsylvania; Michael Lightner, past president of the IEEE and a professor at University of Colorado, Boulder, who gave selflessly of his time in ensuring the rigor and breadth of axiomatic cloud theory; Rico Malvar of Microsoft Research and the IEEE Technology Time Machine team including faculty of the Hong Kong University of Science and Technology; and Lucy Hood of the University of Southern California Marshall School.

The home of cloud computing, via both the invitation-only @Clouderati group and as an extended community is primarily on Twitter, where I am @joeweinman. Most of these individuals are also bloggers, executives, research analysts, CTOs, CIOs, and consultants. A special glue holds this group together, and I consider these leading cloud experts to be not only colleagues but friends and thought-leaders. Literally from A to Z, they are: Vanessa Alvarez, Dave Asprey, David Berlind, Linda Bernardi, Randy Bias, Benjamin Black, Simone Brunozzi, Rachel Chalmers, Sam Charrington, Jean-Luc Chatelain, Adrian Cockcroft, Peter Coffee, Reuven Cohen, Tim Crawford, Simon Crosby, Ellen Daley, William Fellows, Rodrigo Flores, Will Forrest, Mike Fratto, Jay Fry, Vijay Gill, Barton George, Bernard Golden, Charles Golvin, Christofer Hoff, Jason Hoffman, Sam Johnston, Ton Kalker, Jeff Kaplan, John Keagy, Ben Kepes, Michael Krigsman, Maribel Lopez, Bob Lozano, Tom Lounibos, William Louth, Barry Lynn, Lori Mac Vittie, Dave McCrory, Dave McCandless, Brenda Michelson, Marten Mickos, Rich Miller (Data Center Knowledge), Rich Miller (Telematica), Lew Moorman, Ofir Nachmani, Greg Ness, Vinay Nichani, Deb Osswald, Geva Perry, Antonio Piraino, Justin Pirie, Eric Pulier, Elizabeth Rainge, Surendra Reddy, Christian Reilly, Dave Roberts, Guy Rosen, Scott Sanchez, Steve Shah, James Staten, Krishnan Subramanian, Shlomo Swidler, Mark Thiele, Lew Tucker, James Urquhart, Jinesh Varia, Werner Vogels, Simon Wardley, James Watters, John Willis, Dan Young, Jian Zhen, Joe Ziskin.

Thanks to the cloud itself for helping research and connect, not just Twitter, but Google, Google Books, Google Scholar, Alexa, JSTOR, Wikipedia, and hundreds of other online resources such as newspaper and magazine articles.

Final thanks go to my family. To my wife, Carolyn, for helpful editorial comments and for her patience as I spent every night, weekend, and vacation day for six months typing away from behind a wall of books, in the presence of a monotonically increasing list of delayed household projects. To my mother, Elisa, and late father, Joe, for being two wonderful, loving, dedicated parents. Finally, to my daughters, Paige, Ali, and Sierra, for their love and support.

About the cover: The cover illustrates the three perspectives of the book: the real world, shrouded in actual meteorological clouds, symbolizing objectives such as optimized customer experience and issues such as cognitive biases; the world of IT, symbolized by a circuit board; and the abstract mathematics underlying both of them, symbolized by the grid.

CHAPTER 1

A Cloudy Forecast

The cloud—shorthand for "cloud computing"[1]—is transforming all spheres of our world: commerce, entertainment, culture, society, education, politics, and religion. Cloud start-ups are forming on a daily basis, and billions of dollars in wealth are being created as companies craft innovative strategies to exploit this opportunity. Conversely, long-standing corporate icons that have failed to do so are becoming history instead of making it.

The concept of a public cloud—shared, on-demand, pay-per-use resources, accessible over a wide-area network, available to a broad range of customers—might appear to be a recent breakthrough, but there is nothing new under the sun, not even the cloud. The ancient Romans implemented the information superhighway of their time, constructing an unprecedented wide-area network with thousands of route miles of roads, called *viae*, using state-of-the-art engineering, following documented standards.[2] The public network, made of public roads, or *viae publicae*, was complemented by and interoperable with metro networks, the *viae vicinales*, and private networks, the *viae privatae*, creating an Internet of sorts. The roads of the Romans carried people, goods, and soldiers, but, perhaps most important, they also served as a communications network, enabling information, coordination, and control of the far-flung republic and then empire.

These *viae* were multiprotocol networks—carrying pedestrians, animals, carts, military chariots, horses, and their riders—with class of service—military and chariots in the center lane, pedestrians and slower vehicles to the side.[3] Net neutrality was assumed: Any citizen could traverse the *viae publicae* and even had certain rights of passage on the *viae privatae*.[4] By order of Caesar, the core of the network had congestion management: Transport carts were banned from the network core—the narrow, winding streets at the heart of Rome—from dawn until dusk.[5] A complementary architecture was used for streaming content delivery: the aqueducts.

A variety of service providers—inns, taverns, posthouses, and the like—became embedded in the fabric of this network, offering value-added services. Each inn—called a *caupona*—offered lodging to travelers on an on-demand, pay-per-use basis: The traveler merely showed up, stayed, and paid. The inns serviced different classes of customers, from peasants to citizens and free men, and there were laws concerning security and limitations of liability. According to an edict issued by the *praetor*, a senior regulatory official of the time, the proprietor, or *caupo*—the cloud service provider of the age—was responsible for ensuring that the traveler's belongings were neither stolen nor damaged while resident at the service provider's facility.[6] Acts of the gods, such as fires, were excluded. Authentication, via the presentation of credentials or tokens, *tesserae hospitalitatis*, was required before service could be rendered.[7] Advertising and branding were important even then. In ancient Pompeii, the Elephant Inn had a logo: a painting of a pygmy defending an elephant entrapped by a snake. The signage also offered capacity status updates: *hospitium hic locatur* (i.e., "inn to let").[8]

Even before Rome, the Greeks had inns, the Persians had public roads, the Assyrians had aqueducts,[9] the Babylonians extended credit, and over 4,000 years ago, during the dawn of Western civilization in Sumeria, the advanced production, facilities, and power technologies of the time—farm implements, water rights, and oxen—were offered for access under a pay-per-use model: leasing. Thousands of years later, in the Middle Ages, knights' armor—the intrusion-prevention hardware of the time—was also leased.[10] It can even be argued that key elements of today's cloud computing environments have been anticipated by early biological systems: ant colonies will determine the shortest path to "content," such as sources of food, and exhibit behavioral plasticity, that is, will dynamically allocate resources—worker ants—to foraging, patrolling, nest maintenance, and midden work, that is, refuse pile maintenance.[11]

Clouds Everywhere

This proven architectural and business model, since applied to modern hotels, electricity, coffee shops, taxi fleets, rental car services, and others, has now come to computing, and in computing—as in meteorology—the cloud these days is covering a lot of ground. Now, as never before, information technology (IT) and cloud computing are having a broad impact.

The cloud is pervading the prosaic patterns of everyday existence. Teens, tweens, and even toddlers are tapping on touch screens or thumb-typing text messages. Even untethered applications, or "apps," need to be purchased and downloaded via a cloud-based app store, but, more

important, many applications require additional cloud-based services to function. Natural-language interfaces are enabled by cloud-based speech processing and semantic analysis; search requires the near-infinite processing and storage power of the cloud; social gaming is mediated via the cloud; high scores are uploaded to the cloud; apps and content are updated from the cloud; and status updates, files, photos, videos, reviews, and check-ins are shared via the cloud.

The cloud complements the consumerization of IT, and broadens and deepens its democratization. Businesses used to dictate the desktop, laptop, and software used by employees. But if applications in the cloud process data in the cloud, "bring your own device" is a viable strategy—if not without security and interoperability concerns—potentially reducing corporate expenditures while enabling consumer-employees to make fashion and status statements as well as live a blended work-family lifestyle. Democratization of IT means that not only device access but the creation and modification of applications can expand beyond the IT shop, unleashing a torrent of innovation and motivation through empowerment.

Gaming is moving to the cloud as well. Traditionally, you bought a console and cartridges or discs at a physical store. Then you could order over the Web. Then you could take delivery over the Web, via game downloads from an app store. Then you could use your console over the Internet, with up to four-fifths of gamers using connected consoles to play online, download games, or chat.[12] Now, with "cloud gaming," even high-performance games—formerly requiring advanced consoles built to exploit state-of-the-art computing engines—are being played in real time *on* the net *over* even 3G networks, with polygons and video generated remotely but displayed on relatively low-performance endpoints, such as smartphones.[13] It would be a mistake to consider gaming merely to be the province of, say, 14- to 24-year-old males. Gaming is not only popular across many demographics, but represents the state of the art in everything from interfaces to performance that will trickle down into more mundane business applications. Moreover, "games" can represent a new era in collaboration: A long-standing problem in HIV research—the protein structure of the Mason-Pfizer monkey virus retroviral protease—was recently solved by global players of the online game Foldit, illustrating "the power of online games to channel human intuition and three-dimensional pattern-matching skills to solve challenging scientific problems."[14]

The conduct of commerce is undergoing a revolution, with new players in online retailing, group coupons, video distribution, and blogging—to name a few—dramatically disrupting market ecosystems and driving long-established players out of business, while creating fortunes for some in the process. Behind the scenes, cloud-based collaboration, innovation markets, and contests are enabling companies to tap into the smartest and most

creative minds in any field, regardless of geographic location. Procter & Gamble explained the straightforward math[15]: Fewer than 10,000 researchers *within* the firm, 1.5 million *outside*. Or let's go beyond blue-chips: Sites like Mom Invented, which let moms go to market with mechanisms to prevent kids from unrolling toilet paper, are, well, on a roll.[16]

In short, the cloud is disrupting every dimension of business, whether it is the research, engineering, or design of new products and services; their manufacturing, operations, and delivery; or any point along the customer interface and its myriad moments of truth[17]—branding, awareness, catalog, trial, customization, order processing, delivery, installation, support, maintenance, or returns.

Consider the customer engagement life cycle and how IT and the cloud can play a role. Positioning, branding, and advertising today often require social media, and product positioning and customer awareness require trials, demos, and virtual tours. In developing solutions to meet customer requirements, cloud-mediated collaboration such as telepresence, 3-D models, electronic whiteboards, and contests can be vital. For product delivery, mobile tracking and installation support are needed. For service delivery, content delivery networks and continuous online connections can be essential. Billing and payment have gone online and mobile. Support and repair can be handled by user-driven knowledge bases, frequently asked questions (FAQs), and online chat. And the returns and recycling processes are being augmented by collaborative consumption, which is creating a cloud-enabled means of recycling, well, everything.[18] This includes eBay but also a plethora of niche sites.

The cloud can be used to cut costs and to create value. *The New York Times* digitized archival copies of the paper from 1851 to 1980 for customer Web retrieval and was able to convert 11 million articles in less than a day for less than $1,000.[19] *The Washington Post* managed to process 17,481 pages of Hilary Clinton's daily schedule as First Lady in nine hours, for a cost of $144.62.[20] The cost reduction is useful, but these cases are more interestingly viewed in terms of unlocking hidden value and creating a time advantage for reporters to search for scoops. Moreover, the fact that IT was bypassed is both a threat to legacy organizations and an example of the empowerment created by the democratization of IT.

The cloud is also radically reshaping the relationship among governments, the governed, and nongovernmental organizations, impacting regional balances of power and global stability. Arguably the most powerful man on earth—the President of the United States—has had to enlist Twitter to achieve his political objectives. Meanwhile, other world leaders are in prison, or worse, due to movements initiated and coordinated through cloud-based social networks. Throughout history, there has been an asymmetry between governments—organized megaliths with hierarchical,