

MANAGEMENT STRATEGIES FOR THE CLOUD REVOLUTION

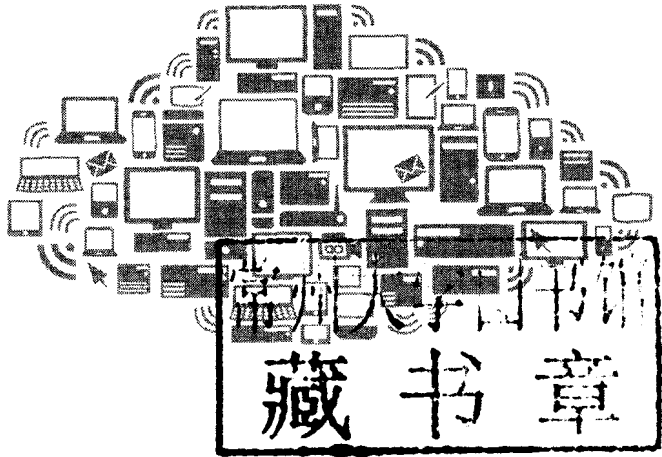


**How Cloud Computing Is
Transforming Business and Why
You Can't Afford to Be Left Behind**

CHARLES BABCOCK

(((MANAGEMENT STRATEGIES FOR)))

THE CLOUD REVOLUTION



**Mc
Graw
Hill**

New York Chicago San Francisco
Lisbon London Madrid Mexico City Milan
New Delhi San Juan Seoul Singapore
Sydney Toronto

Copyright © 2010 by Charles Babcock. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 1 5 4 3 2 1 0

ISBN: 978-0-07-174075-3

MHID: 0-07-174075-9

Design by Lee Fukui and Mauna Eichner

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional person should be sought.

—From a declaration of principles jointly adopted by a committee of the
American Bar Association and a committee of publishers

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please visit the Contact Us pages at www.mhprofessional.com.

ACKNOWLEDGMENTS

To all the outstanding people in the Silicon Valley and elsewhere who have tried to educate me through e-mail, phone interviews, numerous face-to-face sit-down sessions, and the occasional beer at the Thirsty Bear or 21st Amendment, I owe an everlasting debt of thanks.

They include Amit Pandey, CEO of Terracotta, who illustrated what in-memory software could do versus disk-store systems; early Linux advocate Mark Towfiq, who co-led the modernization of e-commerce at Walmart.com alongside Ari Zilka, currently Terracotta's CTO; Bogomil Balkansky at VMware for explaining key elements of virtualization; Winston Bumpus, president of the Distributed Management Task Force (and chief of standards at VMware); Simon Crosby, CTO at the XenSource unit of Citrix Systems, also for highlighting key elements of virtualization; Margaret Lewis, at AMD, for a cross-vendor point of view on industry trends; the open source

leader Rod Johnson, former CEO of SpringSource, now part of VMware, for illustrating how open source works, along with Brian Behlendorf, an original member of the Apache Web server project; Sanjiva Weerawarana, CEO at WSO2, for his understanding of lightweight Web services; and Executive Vice President Paul Cormier at Red Hat for his perspective on early Web search operations. And thanks also to Marty Goetz, president of the former Applied Data Research, who ushered a newcomer into the complexities of large system software and who, many years ago, was instrumental in starting software's march as a force independent of any particular brand of hardware.

Much of the thinking in this book found its first expression in truncated form in the United Business Media publication *InformationWeek*, my current employer. Two editors there assisted me in direct and indirect ways to make this book possible. One is Art Wittmann, now editor of *InformationWeek Analytics* and a steadying voice in the debate over the direction of IT, the network, and the cloud; the second is Chris Murphy, an editor with rare skill at working with the strengths (and weaknesses) of writers. Thanks also to Government CIO and Plug into the Cloud Editor John Foley at InformationWeek.com, who helped guide the process that led to this book's creation, and Editor in Chief Rob Preston, who in a period of constrained resources still gave the go-ahead and offered encouragement and advice.

As much as I am indebted to them, it is still correct to say that any mistakes in this work are purely my own and no one else is accountable.

INTRODUCTION

There's a kind of awe associated with what's called "cloud computing," an impression that something momentous is afoot, as if the man behind the curtain was about to unveil something really big—and this time, for real.

I think it's those big data centers that we keep hearing about, the ones that Google, Yahoo!, Microsoft, Amazon, Sales force.com, and Facebook have been building. In Chicago, Microsoft threw open the doors to its newest data center in September, and a truck backed in over the concrete floor, depositing a container filled with racks of servers. Instead of being unloaded, the container was plugged in, and more than 2,000 servers instantly came to life. There were 11 similar containers already in operation and room for 44 more, while on the

second story of the building, hardware in a more traditional, raised-floor data center was already humming away.

Microsoft's facility is designed for 300,000 servers, and according to Microsoft's president of servers and tools, Bob Muglia, as best he knows, it's the largest data center on earth.

A short while before Microsoft opened its doors, Google had opened a window on what had previously been the secret design of its own data centers. A Google camera crew showed an unpretentious-looking technician, possibly a recent high school graduate, mounting a razor scooter and scooting along the warehouse floor to a server unit. He extracted a failed server from the rack and inserted a new server, a unit that appeared to be about 3.5 inches thick, with a sheet metal baffle to keep the heat-generating parts separate from the cooler parts of the machine. That's not how they do it in the enterprise data center. This is not your father's data center.

Google and Amazon.com pioneered these concepts, and Microsoft and others have picked them up and produced their own implementations. When data centers such as this are built out of what are basically PC parts, with one server cluster consisting of thousands of servers, when very-large-scale parallel processing software is applied to the cluster, and when the governing software routes jobs around hardware failures, you have something new, a "cloud" data center. It is a string of 12 or more such data centers around the world that powers the marvelous Google search engine. And more are being built next to 2 cents per kilowatt hour sources of hydroelectric energy rather than the 11 cents per kilowatt hour energy that powers the computers on which this book was edited. Energy

makes up a quarter of the expense of running a data center; cloud data centers take advantage of low-cost energy sites. The enterprise data center, with its need to be close to headquarters or manufacturing, can't do that.

Granted, some claim that "the cloud" is just another cycle in our seemingly endless series of technology enthusiasms, only to be followed by disappointment. Gartner says that "the cloud" is at the peak of its "hype" cycle, where the highest hopes are invested in it, and at the same time, it's at the top of the list of innovations likely to be adopted in the coming year. That in itself is a rare convergence.

The last hype cycle brought us the dot-com boom, followed by an even more dramatic bust. That boom reflected a fever for Web traffic and led to investment in sites meant to attract a million visitors a week, with imaginary profits to follow. The cloud is more real than the dot-com boom.

The cloud is a set of major productivity gains in computing, each of which is a multiplier of standard computer power in its own right. These multipliers are converging in this new style of data center, combined with a new empowerment of the end user. We are close to moving beyond the world of known computing patterns into a field of dreams, where such data centers are built partly in the belief that end users will not be able to resist their raw compute power or the powerful services that will be created there. I believe that at some point, these data centers will be linked together, backing each other up over the Internet until the old Sun Microsystems dictum, "the network is the computer," finally comes true. This self-reinforcing grid of computer power will reach out to end

users in all sorts of unforeseen ways, finally becoming an all-enveloping embrace.

Have you been bumped into recently by someone walking down the street who is so absorbed in his iPhone or some other electronic device that he can't be bothered to notice the traffic around him? Well, it's going to get a lot worse. The range and depth of digital services that will flow out of the cloud will be more engaging than those currently available. Within a year, even the most detached observer will say that a fundamental shift is underway, with the human culture that's captured on the display of a small digital device being primary and other influences, such as education, literature, fine arts, and film, being secondary. Even skeptics will concede, most of them disapprovingly, that a revolution of historic proportions is taking place.

It will be hard to know how to position your company in the face of this inexorable, omnipresent shift to a more intense digital culture. But with an understanding of what the cloud is all about and how it's likely to evolve, it will be possible to form a strategy for survival and advancement in the coming era.

As we shall explore in this book, at its heart, the cloud is a shift in how end users will do the bulk of their computing. It's assumed at this early stage that "services that previously resided in the client, including e-mail, photo and video storage and office applications" will move off the PC device and into the cloud, according to a paper by Google's leading data center engineers. One needs only to look to MySpace, flickr, YouTube, and Facebook to see that such a shift is already underway.

But a more serious part of their computing, the way they conduct business, which was formerly done on their BlackBerry, netbook, laptop, or PC, will also move into the cloud. New layers of computing will be added to old patterns. Even as *the data centers on the Internet* get larger, the devices on which end users do their direct computing are likely to shrink, two contrary trends that must be reconciled if you're going to end up in the right position to be part of the cloud revolution.

But to the business strategist, the cloud means a good deal more than that. There will be a shift toward being able to rely on large clusters of servers on the Internet for either steady-state operations or the occasional needed surges of compute power—at *prices below the cost of running the corporate data center*. Businesses large and small will have the power to do things that they couldn't do before, do them faster, and reach customers more effectively when they make the right moves.

A new platform has emerged with which to engage customers and provide universal access to the business. Many new possibilities for interacting with prospects and engaging with customers are taking shape. The people you will be capable of *reaching tomorrow wouldn't have dreamed of walking through your doors today*.

If anything, business is going to find it harder to sell to well-informed consumers, who roam about the Internet at will as if they owned the world. The cloud has many potentially unpleasant connotations for traditional businesses—*instant, acidic reviews by the most superficially disgruntled consumers, sharing their upset with millions*. At the same time, it's going to offer new opportunities to *relate to customers*

and understand why they came to you in the first place—and what they may want next.

In its most popular form, such as Amazon's Elastic Cloud Compute (EC2), cloud computing is a reorganization of external resources into a hitherto hard to conceive of set of computing services. Computing cycles of nearly any magnitude can be tapped at will. The amount of resources devoted to the job expands if, say, a surge in customer traffic makes it advisable to do so. And this expansive data center accessed through the Internet can be utilized at low hourly rates with the swipe of a credit card.

Perhaps the single most compelling feature of the cloud is that it is programmatically accessible by outsiders, the end users who have work for all those concentrated processing units. Automated processes have been built in to make cloud services readily available to anyone, regardless of location, as long as that person can pay the hourly rate. It's something like iTunes. You upload a small set of information related to yourself and get back a favorite song, without having to appear at a store and sort through bins. However, in the cloud, it's an enterprise application that goes out over the wire and the results of processing all that data come back.

No single technology is responsible for the advent of the cloud. Broadband communications, Web standards, multicore servers, and the ability to manage large groups of computers as if they were a single machine—these are the components of cloud computing. Mix them together, along with a tendency to organize business applications as services, and things

will never be the same again. This new computing power will change the way companies will do business.

Today, cloud computing is most frequently thought of as an external resource, the public cloud. Tomorrow, you will find your organization reorganizing its data center around cloud principles. If this is done adroitly, your internal cloud will be smaller and less expensive than the former data center. That's because for years corporate data centers have been overbuilt. Now they will be right-sized and will align easily with an external cloud that can absorb the spikes that you send it. You will be provisioning your own facilities for near steady-state operation, rather than workload peaks. When unusual demands occur, say, in accounting at the end of the quarter or in the holiday rush at the end of the year, you will be able to move them off to the external cloud. You'll have to pay for the time you use, but immense savings will be gained by avoiding that former compulsive overprovisioning.

This hybrid cloud, a mix of external public resources and reorganized internal resources, and how it will affect what your company can do are what this book is about. No such hybrid clouds have been designed from the ground up yet—it's too early—but they're evolving out of today's infrastructure. In effect, your data center of the future is a hybrid cloud.

Cloud computing will solve the problem of overprovisioning and the tendency of data center budgets to invest heavily in keeping the lights on and the computers running, when what they really should be doing is solving new problems. The cloud will also bring its own complexities and management

challenges, and some of them will prove worthy successors to the challenges of the past.

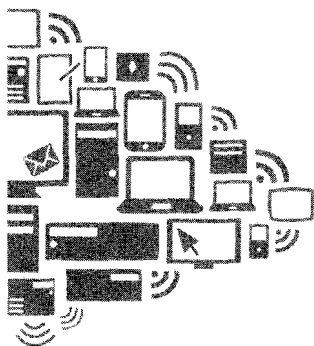
But most of all, the cloud will bring a new way of doing things and a whole new set of opportunities. *Management Strategies for the Cloud Revolution* is about this break from the shackles of the past and the competitive landscape that is likely to emerge as a result.

CONTENTS

	ACKNOWLEDGMENTS	ix
	INTRODUCTION	xi
1	THE CLOUD REVOLUTION	1
2	THE AMORPHOUS CLOUD	25
3	VIRTUALIZATION CHANGES EVERYTHING	51
4	JUST OVER THE HORIZON, PRIVATE CLOUDS	69
5	THE HYBRID CLOUD	87
6	OVERCOMING RESISTANCE TO THE CLOUD	103
7	IT REORGANIZES	127

CONTENTS

8	DANGERS ABOUND: SECURITY IN THE CLOUD	145
9	YOUR CLOUD STRATEGY: WHAT KIND OF COMPANY DO YOU WANT?	163
10	CALCULATING THE FUTURE	185
11	NEBULA: NASA'S STRATEGIC CLOUD	207
APPENDIX A	NIST DEFINITION OF CLOUD COMPUTING	221
APPENDIX B	<i>INFORMATIONWEEK</i> ANALYTICS, JUNE 2009	227
APPENDIX C	CLOUD COMPUTING'S PORTABILITY GOTCHA: TRANSFER FEES CAN LEAD TO LOCK-IN AS DATA STORES GROW	231
	GLOSSARY	235
	INDEX	245



THE CLOUD REVOLUTION

In works of art, from the photos of Ansel Adams to the paintings of ancient Chinese artists, clouds have often been given tangible form and purpose. In Adams's arid West, they served as a backdrop to granite peaks, holding out the promise of rain. To the Chinese, an all-encompassing mist allows special features to emerge out of the mountain landscape, or sometimes there will be a series of ridges as far as the eye can see, their bases cloaked in clouds—an illusion of infinity.

For many years the cloud has played a more prosaic role among the squares, rectangles, and circles of the architecture diagrams of technology projects, but its meaning has been ambiguous. "The cloud" was a euphemism for everything that was beyond the data center or out on the network. The action

that affected the project at hand was in the data center; the cloud was a mishmash of remotely connected parts and network protocols that didn't have much to do with the immediate problem. No matter how nonartistic the systems architect, he could always represent the cloud—an offhand, squiggly circle in the background of his scheme.

As business use of the Internet has grown, the cloud has moved from a throwaway symbol in the architect's diagram to something more substantial and specific: it has become the auxiliary computing, supplied through Web site applications and Web services, such as credit checks and customer address lookups, that backed up the operation of standard business applications in the enterprise data center. Businesses built around Web services, such as Google, Amazon.com, and eBay, produced a new type of data center that was more standardized, more automated, and built from mass-produced personal computer parts. Access to these data centers was kept under wraps for several years as their builders sought to maintain a competitive advantage. As the notion caught on that it was possible to provide more and more powerful services over the Internet, cloud computing came to mean an interaction between an end user, whether a consumer or a business computing specialist, and one of these services “in the cloud.”

When Microsoft appeared on the scene determined to stake a larger claim to this new form of computing, it started talking about its facilities in Chicago and Ireland as a new type of data center. Google, which played a key role in establishing the type, began illustrating key features of its data centers, and