Wiley CIO Series

CLOUD COMPUTING AND ELECTRONIC DISCOVERY

JAMES P. MARTIN
HARRY CENDROWSKI

WILEY

CLOUD COMPUTING AND ELECTRONIC DISCOVERY

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

Architecting the Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS, and IaaS) by Michael Kavis

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

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

Confessions of a Successful CIO: How the Best CIOs Tackle Their Toughest Business Challenges by Dan Roberts and Brian Watson

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

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

Decoding the IT Value Problem: An Executive Guide for Achieving Optimal ROI on Critical IT Investments by Gregory J. Fell

Enterprise Performance Management Done Right: An Operating System for Your Organization by Ron Dimon

Information Governance: Concepts, Strategies and Best Practices by Robert F. Smallwood

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

Leading the Epic Revolution: How CIOs Drive Innovation and Create Value Across the Enterprise by Hunter Muller

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 World (Second Edition) by Gregory S. Smith

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

Trust and Partnership: Strategic IT Management for Turbulent Times by Robert Benson, Piet Ribbers, and Ronald Billstein

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 (Second Edition) by Dan Roberts

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

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.

CLOUD COMPUTING AND ELECTRONIC DISCOVERY

James P. Martin

Harry Cendrowski

WILEY

Cover image: © istock/polygraphus
Cover design: Wiley

Copyright © 2014 by John Wiley & Sons, Inc. 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 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:
Martin, James P., author.
Cloud computing and electronic discovery/James P. Martin, Harry Cendrowski.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-76430-5 (cloth); ISBN 978-1-118-94745-6 (ebk); ISBN 978-1-118-94744-9 (ebk) 1. Cloud computing—Law and legislation—United States. 2. Electronic discovery (Law)—United States. 3. Privacy, Right of—United States. 4. United States. Electronic Communications Privacy Act of 1986. I. Cendrowski, Harry, author. II. Title.
KF390.5.C6M365 2014
347.73'72—dc23
2014013668

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE

In general, *cloud computing* describes technologies that allow applications and data to be hosted on a computer external to a business's own computing resources and firewall (i.e. a "remote computer"). From a personal perspective, it means that an individual can have access to convenient solutions for little or no cost, for example, to host family photos or videos. One of the promises of cloud computing is the end user doesn't really need to know how it works, or where the data resides; it just works, and will be available when you need it, wherever you need it. This promise has also resulted in profound misunderstandings of what cloud computing actually involves, and what it means from a litigation perspective.

Cloud computing services provide computing resources on an as-needed basis; this is why cloud computing was sometimes referred to as *utility computing*, a term that certainly did not have the marketing cachet of *cloud computing*. Cloud computing solutions generally require a reasonable periodic service fee and little additional hardware cost to access a computing solution. The reduction of costly IT assets, avoidance of software license costs, and removal of software maintenance tasks provide an attractive economic model; the end user is given a turnkey solution supported and maintained by the service provider, and hosted at a remote location. Cloud computing is enabled by rapid, reliable Internet and mobile data communications, which means that applications and data are available "everywhere," simultaneously, and transparently. The convenience and financial benefits of cloud solutions are changing business models fundamentally and have resulted in mass migration of data to the cloud. Much of this data, of course, would be of interest to parties during criminal proceedings or civil litigation.

A key issue from a legal perspective is that an investigator or litigant cannot access data held in the cloud through traditional discovery techniques. Discovery of data within a cloud computing solution likely falls under the restrictions of the Electronic Communications Privacy Act of 1986 (ECPA),[1] and specifically, Title II of the ECPA, which is called the Stored Communications Act (SCA).[2] Under the SCA, third parties that provide communication services or remote computing services to the public are generally prohibited from releasing the data; the SCA defines a series of procedures for the government to access

[1] 18 U.S.C. §§2510–2522.

[2] 18 U.S.C. §§2701–2712, although the term "Stored Communications Act" does not appear anywhere within the body of the legislation.

the data. This law, now almost 30 years old, is the primary law that regulates disclosure of such data. It was written at a time when telephones actually rang, when e-mail was considered a novel new technology for computer geeks, and when conversations on portable phones could be intercepted with a standard FM radio. Today, judges use this law to rule on cases involving data created and stored by devices that would have been considered magic (or certainly at least in the realm of science fiction) in 1986.

This book is our attempt to briefly explain the way that data held by a third-party provider (i.e., in a cloud computing solution) potentially affects legal proceedings and discovery of electronic information. This work is divided into three topical sections:

Section One explains the basics of cloud computing technologies, how data is stored, and (at a high level) the technical aspects of hosted solutions that can affect production of data. This is intended to be a technical guide for non-technicians, offering a brief glimpse behind the technological curtain.

Section Two describes the SCA as well as the prior laws that protected technological communications of the day. This will hopefully provide the reader with insights into legal concepts that still shape cases today, and the common themes of privacy issues. We also describe some of the limitations of the current laws in interpreting modern systems and devices.

Section Three surveys many of the precedent-setting cases involving interpretation of hosted data and access of such data by litigants or the government. Many of these cases are still active and may be modified on appeal. Rapid technological advancements mean that issues may arise that have not been previously considered by the courts in the current context, and interpretation in those situations can widely vary.

The issues presented here often walk hand-in-hand with privacy issues. However, we limit this discussion primarily to litigation settings. Recent revelations of widespread government surveillance programs are well beyond the scope of this work. We sincerely hope this book provides practical insight into the current world of hosted data and its potential impact on legal proceedings, and wish you the best as you encounter these issues in the future.

James P. Martin
Harry Cendrowski
May 2014

ACKNOWLEDGMENTS

We are sincerely grateful to many individuals for their unique contributions to this book as well as their steadfast support and encouragement. First and foremost, we would like to thank the Wiley team, including John DeRemigis, Sheck Cho, Stacey Rivera, and the staff at John Wiley & Sons for their assistance and support during the development and writing process.

We would also like to thank all the contributing authors and advisors to the process, without whom the production of this book would not have been possible:

Matthew P. Breuer
Deirdre Fox
Virginia Kim
Sarah Marmor
Christopher Thieda

Their professional insights and advice were instrumental in the production of this work, and their dedication and commitment are sincerely appreciated. Thank you also to the countless individuals who provided perspectives on the use of emerging technologies, expectations of privacy, and the proliferation of smart devices.

SECTION ONE

Cloud Computing: Basics of Technologies and Applications

CHAPTER 1

Cloud Computing Definitions and Technical Considerations

Christopher Thieda

The introduction of cloud computing has taken technology users by the hand and brought them into a new realm of possibilities. Whether the purpose is for personal, corporate use, or anything in between, today's everyday tech users have been exposed to a multitude of cloud practicalities. Cloud computing applications allow computer users to conveniently rent access to fully featured applications, to software development and deployment environments, and to computing infrastructure assets such as network-accessible data storage and processing. Those that have exposure to common applications such as Google Apps or Microsoft Office 365 likely already have experience with cloud computing, even though they may not have realized it.

The term *cloud computing* has a variety of definitions, mostly because it has become a powerful marketing term. The National Institute of Standards and Technology, the federal technology agency that works with industry to develop and apply technology, offers this definition:

> Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[1]

Today, technical questions remain that occasional users might not dare to ask regarding how virtualized models actually operate, where data

actually resides, or who actually controls access to the data and applications, but for some users that are financially dependent on or have sensitive data involved with their cloud solution, those questions should be addressed. Parties to litigation will also naturally be concerned with the answers to those questions as well. Of course, there are numerous advantages to cloud computing from the perspective of the customer. Scalability, cost efficiency, ease of implementation, and optimal resource allocation are some of the main benefits that stem from virtualization. Conversely, concerns have risen concerning cloud practices regarding security, storage location, and intrusion protection. For parties and their counsel involved in litigation, cloud computing has increased the complexity of electronic discovery. In this chapter, we will address the different cloud computing models, the issues of cloud computing applications, and the legal regulations involving virtual data capture. Cloud computing is a developing area, and the strengths, weaknesses, delivery models, and legal implications of its use are constantly in flux.

Virtualization is the key technology involved in cloud computing. In a virtual computing model, an organization can obtain the exact hardware and/or software solutions required, at the exact time it is required, without the need for a large capital commitment. Virtualization allows hardware and software owners to partition their resources and provide the exact quantity of resources needed to satisfy their customers. This model has existed for a while, but has been advancing in recent years due to the common availability of low-cost, high-speed data communications infrastructure.

There are three main service models seen in today's cloud computing environments. We will focus on: cloud Infrastructure as a Service (IaaS), which allows organizations to outsource hardware, cloud Platform as a Service (PaaS), which allows organizations to outsource operating systems and web infrastructure, and cloud Software as a Service (SaaS), which allows companies to outsource applications. These layers create the core of cloud computing. Since they share the commonality as components of the cloud, each of the three layers accomplish specific tasks and have the capabilities to complement one another in an entirely virtual environment. IaaS is the substitution of virtual solutions for hardware that is commonly used within a company's network. PaaS is created for users to be able to build and implement their own virtual, web-based solutions. SaaS is centered around supporting users entirely through web-based resources, and it is the most commonly seen model in today's cloud market. Every cloud layer provides a differentiation factor versus standard enterprise networking while providing a broad range of possibilities for users looking to delve into the world of virtualization. Most consumers will typically contract with an SaaS vendor to provide a web solution, and may not be aware that the

infrastructure and platform levels have also, in turn, been outsourced to other cloud vendors.

IaaS

Cloud computing has provided organizations with the advantage of configuring their network based on using resources in the most efficient manner. IaaS is the foundation of the three cloud layers. It is a virtualized availability of hardware that can substitute for pertinent networking items such as servers, firewalls, and load balancers. Instead of purchasing a physical server and firewall with a set amount of data capacity, virtual network solutions are available where storage and computing power is scalable depending on the organization's requirements. Virtual machines have also created a way for users to obtain similar functionality to preexisting hardware while eliminating data center space and recurring physical support costs including maintenance, power consumption, and expertise to operate the hardware. The elimination of overhead costs and flexibility are the main reasons why companies choose to source their infrastructure through the cloud.

Although there are many benefits of virtualizing an environment, network administrators must have a thorough knowledge of networking and how infrastructures should be constructed in order to properly configure their cloud requirements. Administrators must be well-versed in dealing with different virtualized operating systems and interfaces. An example of an important resource to be familiar with when dealing with a cloud-based infrastructure is a hypervisor. A hypervisor is software that enables users to monitor and control servers that are built on hosted environments. Hypervisors are an extremely useful technology piece to remotely allocate shared resources that can have a large impact concerning how efficiently data is transferred.

There are two types of hypervisors, depending on how they are implemented. The first is a type-1 hypervisor, which is built directly on the server platform and communicates with resources designated by the service provider. The second is a type-2 hypervisor, which is built on a preexisting host operating system and can interact with associated virtual systems thereafter. A type-1 hypervisor is more commonly used in business practices, as it minimizes any latency potential and maximizes networking efficiency from its direct source of interaction with the server. Type-2 hypervisors are still a useful way to virtually manage servers and can be effective when the operating system is communicating with input-output style computing processes, similar to how personal web surfing is conducted. VMware, a popular cloud service provider, offers both type-1 and type-2 hypervisors with their operating

systems. VMware's ESXi is an example of a type-1 hypervisor, whereas their VMware Server software is a type-2.

It is important to understand hypervisors and how they work, as it could aid in reducing potential security threats. If an organization is looking to minimize risk against their virtualized infrastructure, they could implement an efficient hypervisor strategy to stop any malicious attacks from taking down their entire network. Hypervisors can be set up by separating virtual servers with the intention of preventing compromised network channels from negatively impacting other servers. Instead of an attack on a host causing a severe security breach for all virtual machines associated with it, hypervisors can be set up by segregating how information is transferred. Hypervisors also provide the advantage of transferring data using encrypted communication methods such as Internet Protocol Security, commonly referred to as IPsec.

The most significant element concerning cloud computing infrastructure is that there are differences regarding how cloud networks can be implemented. This breakdown is categorized into three cloud computing groups: public, private, or hybrid. Each is distinctively separated in terms of how the software, firmware, or infrastructure is hosted.

Public Cloud

Public cloud computing provides users with the availability of hosted online resources through service providers. This is the most common cloud application seen in today's market due to the integration ease for new users and its convenient bundles that can be purchased according to requirements and usage. Instead of having hardware on site and needing to constantly create data center space, public cloud computing is an alternative hosting solution.

Because it is externally hosted, the added benefit of network flexibility also comes along with vagueness regarding how data is stored and where it resides. As previously mentioned, public cloud is a service provided and sourced entirely through a service provider's infrastructure; the service provider is providing services to hundreds if not thousands of organizations. Thus, public cloud solutions are hosted through an infrastructure with a mixture of data from other entities. In a traditional computing model, organizations operate with enterprise hardware they actually own and control, and have the benefit of physically knowing where data is being stored at all times. Cloud users operate with the hindrance of not having direct physical control over the hosted network (i.e., the actual hardware) within which their data lies. During discovery, this limits the ability of investigators to directly access the data of interest. This transparency limitation can also cause the potential for unfavorable variances with data security.