RELIABILITY AND AVAILABILITY OF CLOUD COMPUTING

ERIC BAUER

RANDEE ADAMS

WILEY

IEEE PRESS

Table of Contents

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

Technical Reviewers

Xuemei Zhang
Principal Member of Technical Staff
Network Design and Performance Analysis
AT&T Labs

Rocky Heckman, CISSP
Architect Advisor
Microsoft

RELIABILITY AND AVAILABILITY OF CLOUD COMPUTING

Eric Bauer
Randee Adams

IEEE PRESS

A JOHN WILEY & SONS, INC., PUBLICATION

cover image: © iStockphoto

cover design: Michael Rutkowski

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

Published simultaneously in Canada.

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 also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Bauer, Eric.

Reliability and availability of cloud computing / Eric Bauer, Randee Adams.

p. cm.

ISBN 978-1-118-17701-3 (hardback)

1. Cloud computing. 2. Computer software–Reliabillity. 3. Computer software–Quality control. 4. Computer security. I. Adams, Randee. II. Title.

QA76.585.B394 2012

004.6782–dc23

2011052839

To our families and friends for their continued encouragement and support.

FIGURES

TABLES

EQUATIONS

INTRODUCTION

Cloud computing is a new paradigm for delivering information services to end users, offering distinct advantages over traditional IS/IT deployment models, including being more economical and offering a shorter time to market. Cloud computing is defined by a handful of essential characteristics: on-demand self service, broad network access, resource pooling, rapid elasticity, and measured service. Cloud providers offer a variety of service models, including infrastructure as a service, platform as a service, and software as a service; and cloud deployment options include private cloud, community cloud, public cloud and hybrid clouds. End users naturally expect services offered via cloud computing to deliver at least the same service reliability and service availability as traditional service implementation models. This book analyzes the risks to cloud-based application deployments achieving the same service reliability and availability as traditional deployments, as well as opportunities to improve service reliability and availability via cloud deployment. We consider the service reliability and service availability risks from the fundamental definition of cloud computing—the essential characteristics —rather than focusing on any particular virtualization hypervisor software or cloud service offering. Thus, the insights of this higher level analysis and the recommendations should apply to all cloud service offerings and application deployments. This book also offers recommendations on architecture, testing, and engineering diligence to assure that cloud deployed applications meet users' expectations for service reliability and service availability.

Virtualization technology enables enterprises to move their existing applications from traditional deployment scenarios in which applications are installed directly on native hardware to more evolved scenarios that include hardware independence and server consolidation. Use of virtualization technology is a common characteristic of cloud computing that enables cloud service providers to better manage usage of their resource pools by multiple cloud consumers. This book also considers the reliability and availability risks along this evolutionary path to guide enterprises planning the evolution of their application to virtualization and on to full cloud computing enablement over several releases.

AUDIENCE

The book is intended for IS/IT system and solution architects, developers, and engineers, as well as technical sales, product management, and quality management professionals.

ORGANIZATION

The book is organized into three parts: *Part I, "Basics," Part II, "Analysis,"* and *Part III—,"Recommendations."* Part I, "Basics," defines key terms and concepts of cloud computing, virtualization, service reliability, and service availability. Part I contains three chapters:

- *Chapter 1, "Cloud Computing."* This book uses the cloud terminology and taxonomy defined by the U.S. National Institute of Standards and Technology. This chapter defines cloud computing and reviews the essential and common characteristics of cloud computing. Standard service and deployment models of cloud computing are

reviewed, as well as roles of key cloud-related actors. Key benefits and risks of cloud computing are summarized.

- *Chapter 2, "Virtualization."* Virtualization is a common characteristic of cloud computing. This chapter reviews virtualization technology, offers architectural models for virtualization that will be analyzed, and compares and contrasts "virtualized" applications to "native" applications.
- *Chapter 3, "Service Reliability and Service Availability."* This chapter defines service reliability and availability concepts, reviews how those metrics are measured in traditional deployments, and how they apply to virtualized and cloud based deployments. As the telecommunications industry has very precise standards for quantification of service availability and service reliability measurements, concepts and terminology from the telecom industry will be presented in this chapter and used in Part II, "Analysis," and Part III, "Recommendations."

Part II, "Analysis," methodically analyzes the service reliability and availability risks inherent in application deployments on cloud computing and virtualization technology based on the essential and common characteristics given in Part I.

- *Chapter 4, "Analyzing Cloud Reliability and Availability."* Considers the service reliability and service availability risks that are inherent to the essential and common characteristics, service model, and deployment model of cloud computing. This includes implications of service transition activities, elasticity, and service orchestration. Identified risks are analyzed in detail in subsequent chapters in Part II.
- *Chapter 5, "Reliability Analysis of Virtualization."* Analyzes full virtualization, OS

virtualization, paravirtualization, and server virtualization and coresidency using standard reliability analysis methodologies. This chapter also analyzes the software reliability risks of virtualization and cloud computing.

- *Chapter 6, "Hardware Reliability, Virtualization, and Service Availability."* This chapter considers how hardware reliability risks and responsibilities shift as applications migrate to virtualized and cloud-based hardware platforms, and how hardware attributed service downtime is determined.
- *Chapter 7, "Capacity and Elasticity."* The essential cloud characteristic of rapid elasticity enables cloud consumers to dispense with the business risk of locking-in resources weeks or months ahead of demand. Rapid elasticity does, however, introduce new risks to service quality, reliability, and availability that must be carefully managed.
- *Chapter 8, "Service Orchestration Analysis."* Service orchestration automates various aspects of IT service management, especially activities associated with capacity management. This chapter reviews policy-based management in the context of cloud computing and considers the associated risks to service reliability and service availability.
- *Chapter 9, "Geographic Distribution, Georedundancy, and Disaster Recovery."* Geographic distribution of application instances is a common characteristic of cloud computing and a best practice for disaster recovery. This chapter considers the service availability implications of georedundancy on applications deployed in clouds.

Part III, "Recommendations," considers techniques to maximize service reliability and service availability of applications deployed on clouds, as well as the design for

reliability diligence to assure that virtualized applications and cloud based solutions meet or exceed the service reliability and availability of traditional deployments.

- *Chapter 10, "Applications, Solutions and Accountability."* This chapter considers how virtualized applications fit into service solutions, and explains how application service downtime budgets change as applications move to the cloud. This chapter also proposes four measurement points for service availability, and discusses how accountability for impairments in each of those measurement points is attributed.
- *Chapter 11, "Recommendations for Architecting a Reliable System."* This chapter covers architectures and techniques to maximize service availability and service reliability via virtualization and cloud deployment. A simple case study is given to illustrate key architectural points.
- *Chapter 12, "Design for Reliability of Virtualized Applications."* This chapter reviews how design for reliability diligence for virtualized applications differs from reliability diligence for traditional applications.
- *Chapter 13, "Design for Reliability of Cloud Solutions."* This chapter reviews how design for reliability diligence for cloud deployments differs from reliability diligence for traditional solutions.
- *Chapter 14, "Summary."* This gives an executive summary of the analysis, insights, and recommendations on assuring that reliability and availability of cloud-based solutions meet or exceed the performance of traditional deployment.

ACKNOWLEDGMENTS

The authors were greatly assisted by many deeply knowledgeable and insightful engineers at Alcatel-Lucent, especially: Mark Clougherty, Herbert Ristock, Shawa Tam, Rich Sohn, Bernard Bretherton, John Haller, Dan Johnson, Srujal Shah, Alan McBride, Lyle Kipp, and Ted East. Joe Tieu, Bill Baker, and Thomas Voith carefully reviewed the early manuscript and provided keen review feedback. Abhaya Asthana, Kasper Reinink, Roger Maitland, and Mark Cameron provided valuable input. Gary McElvany raised the initial architectural questions that ultimately led to this work. This work would not have been possible without the strong management support of Tina Hinch, Werner Heissenhuber, Annie Lequesne, Vickie Owens-Rinn, and Dor Skuler.

Cloud computing is an exciting, evolving technology with many avenues to explore. Readers with comments or corrections on topics covered in this book, or topics for a future edition of this book, are invited to send email to the authors (Eric.Bauer@Alcatel-Lucent.com, Randee.Adams@Alcatel-Lucent.com, or pressbooks@ieee.org).

Eric Bauer
Randee Adams

I

BASICS

1

CLOUD COMPUTING

The U.S. National Institute of Standards and Technology (NIST) defines cloud computing as follows:

> Cloud computing is a model for enabling ubiquitous, 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
>
> [NIST-800-145].

This definition frames cloud computing as a “utility” (or a “pay as you go”) consumption model for computing services, similar to the utility model deployed for electricity, water, and telecommunication service. Once a user is connected to the computing (or telecommunications, electricity, or water utility) cloud, they can consume as much service as they would like whenever they would like (within reasonable limits), and are billed for the resources consumed. Because the resources delivering the service can be shared (and hence amortized) across a broad pool of users, resource utilization and operational efficiency can be higher than they would be for dedicated resources for each individual user, and thus the price of the service to the consumer may well be lower from a cloud/utility provider compared with the alternative of deploying and operating private resources to provide the same service. Overall, these characteristics facilitate outsourcing production and delivery of these crucial “utility” services. For example, how

many individuals or enterprises prefer to generate all of their own electricity rather than purchasing it from a commercial electric power supplier?

This chapter reviews the essential characteristics of cloud computing, as well as several common characteristics of cloud computing, considers how cloud data centers differ from traditional data centers, and discusses the cloud service and cloud deployment models. The terminologies for the various roles in cloud computing that will be used throughout the book are defined. The chapter concludes by reviewing the benefits of cloud computing.

1.1 ESSENTIAL CLOUD CHARACTERISTICS

Per [NIST-800-145], there are five essential functional characteristics of cloud computing:

1. on-demand self service;
2. broad network access;
3. resource pooling;
4. rapid elasticity; and
5. measured service.

Each of these is considered individually.

1.1.1 On-Demand Self-Service

Per [NIST-800-145], the essential cloud characteristic of “on-demand self-service” means “a consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service’s provider.” Modern telecommunications networks offer on-demand self service: one has direct dialing access to any other telephone