

THE PEARSON DIGITAL ENTERPRISE SERIES FROM THOMAS ERL 

Foreword by **David Linthicum**

SECOND EDITION

Cloud Computing

Concepts, Technology, Security & Architecture

by Top-Selling Author **Thomas Erl**
with Eric Barceló Monroy

with contributions from Professor Zaigham Mahmood and Dr. Ricardo Puttini



FREE SAMPLE CHAPTER |





human



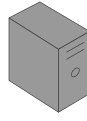
administrator



manager



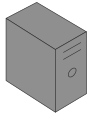
attacker



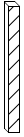
physical server



virtual server



server (attacker)



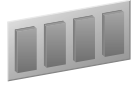
physical firewall



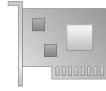
virtual firewall



CPU



memory



network adapter



physical network



virtual network



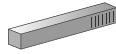
VI manager



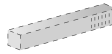
hypervisor



virtualization platform



physical network device



virtual network device



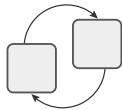
connection ports or virtual switch



container



internal container logic



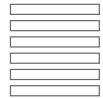
container cluster



container engine



container image



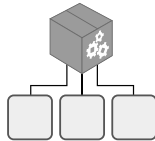
container image layers



package



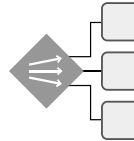
image registry



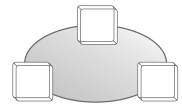
container package manager



package repository



deployment optimizer



container network



router



core switch



top-of-rack switch



container build file



schema or data model



policy



general machine processable document



human readable document



ready-made environment



management system



remote administration system



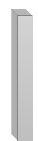
actively processing



software program or application



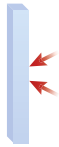
product, system or application



agent or intermediary



traffic monitor



network intrusion monitor



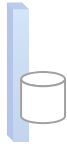
activity log monitor



authentication log monitor



VPN monitor



data loss protection monitor



machine learning system



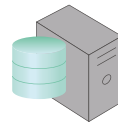
AI system



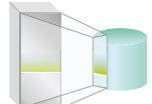
biometric scanner



multi-factor authentication (MFA) system



identity & access management (IAM) system



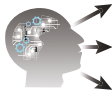
digital virus scanning & decryption system



malicious code analysis system



data loss prevention (DLP) system



intrusion detection system (IDS)



penetration testing tool



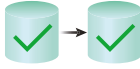
user behavior analytics (UBA) system



third-party software update utility



trusted platform module (TPM)



data backup & recovery system



cybersecurity solution



malware



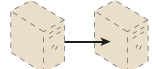
malicious packet



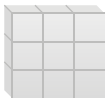
virus



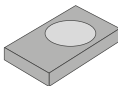
virtual desktops



live VM migration



multitenant application



hard disk



hard disk with enclosure



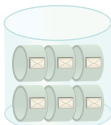
storage device (internal)



storage controller



databases



message queue



repository or storage device



shared storage



state data in memory



service with state data (stateful service)



repository with state data



grid service



service or proxy



service composition



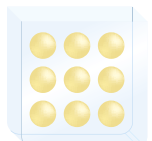
service layer



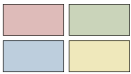
service contract (chorded circle notation)



decoupled service contract



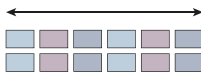
service inventory



LUNs



LUN migration



storage replication



live storage migration



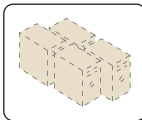
security element or locked resource



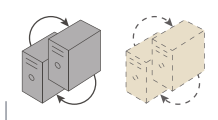
message



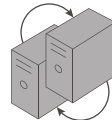
encrypted message



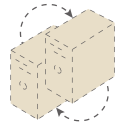
resource pool



resource clusters



physical server cluster



virtual server cluster



malicious component or program



trusted attacker



attacker



malicious service agent



private key



public key



heartbeat message



conflict symbol



web browser



web user interface



folder



workstation



mobile computer



mobile devices



business process/ workflow logic



file system



runtime



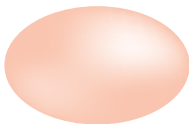
transition arrow



symbols used in conceptual relationship diagrams



organization



zone or region



internet



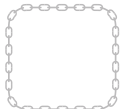
virtual private network



cloud



host boundary



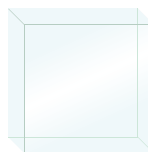
container chain boundary



logical network perimeter or logical boundary



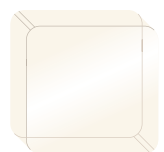
pod boundary



general physical boundary



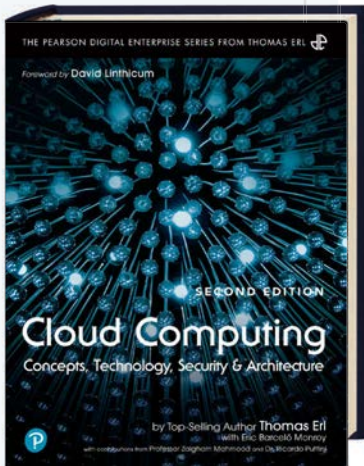
container boundary



system or program boundary

ARCITURA EDUCATION

PROFESSIONAL ACCREDITATION PROGRAMS



This book is an official supplement for the following Arcitura certification programs:

- Cloud Technology Professional
- Cloud Architect
- Cloud Security Specialist
- Cloud Computing Consultant
- Containerization Architect
- Cybersecurity Specialist
- Digital Transformation Technology Professional
- Digital Transformation Technology Architect

For more information, visit: www.arcitura.com



Learn more about Arcitura courses and certifications at:
youtube.com/@arcitura

THE PEARSON DIGITAL ENTERPRISE SERIES FROM THOMAS ERL

Learn from Top-Selling Authors and Leading Industry Experts

ABOUT THE SERIES

The Pearson Digital Enterprise Series from Thomas Erl aims to provide the IT industry with comprehensive, unbiased coverage of the contemporary practices and technology innovations that are driving the international adoption and evolution of digital transformation and the realization of digital enterprises. Each title in this book series is authored in relation to other titles so as to establish a library of complementary knowledge. Although the series covers a broad spectrum of topics, each title is authored in compliance with common language, vocabulary, and illustration conventions so as to enable readers to continually explore cross-topic research and education.



Book Series Website: DigitalEnterpriseBookSeries.com



Book Series LinkedIn Group: [LinkedIn.com/groups/2954416](https://www.linkedin.com/groups/2954416)



ABOUT THE SERIES EDITOR

Thomas Erl is a best-selling IT author and the series editor of the Pearson Digital Enterprise Series from Thomas Erl (www.thomaserl.com/books). You can find Thomas on the Thomas Erl YouTube channel (youtube.com/@terl). He is also the host of the *Real Digital Transformation* podcast series (available via Spotify, Apple, Google Podcasts, and most other platforms) and also publishes the LinkedIn newsletter *The Digital Enterprise*. Over 100 articles and interviews by Thomas have been published in numerous publications, including *CEO World*, *The Wall Street Journal*, *Forbes*, and *CIO Magazine*. Thomas has also toured over 20 countries as a keynote speaker for various conferences and events.

As CEO of Arcitura Education (www.arcitura.com), Thomas has led the development of curricula for internationally recognized, vendor-neutral training and accreditation programs. Arcitura's portfolio currently consists of over 100 course modules, over 100 Pearson VUE exams, and over 40 certification tracks, covering topics such as Digital Transformation, Robotic Process Automation (RPA), DevOps, Blockchain, IoT, Containerization, Machine Learning, Artificial Intelligence (AI), Cybersecurity, Service-Oriented Architecture (SOA), Cloud Computing, and Big Data Analytics. Thomas is also the founder and senior advisor at Transformative Digital Solutions (www.transformative.digital), as well as a freelance LinkedIn Learning instructor and courseware author.

ThomasErl.com



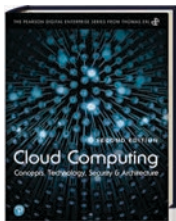
[LinkedIn.com/in/thomaserl](https://www.linkedin.com/in/thomaserl)



[YouTube.com/@terl](https://www.youtube.com/@terl)



[YouTube.com/@arcitura](https://www.youtube.com/@arcitura)



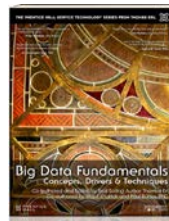
Cloud Computing: Concepts, Technology, Security & Architecture (Second Edition)
by T. Erl, E. Barceló
ISBN: 9780138052256
Paperback



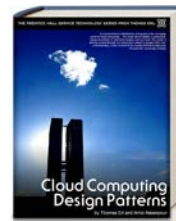
A Field Guide to Digital Transformation
by T. Erl, R. Stoffers
ISBN: 9780137571840
Paperback, 278 pages



Service-Oriented Architecture: Analysis & Design for Services and Microservices (Second Edition)
by T. Erl, P. Merson, R. Stoffers
ISBN: 9780133858587
Paperback, 393 pages



Big Data Fundamentals: Concepts, Drivers & Techniques
by P. Buhler, T. Erl, W. Khattak
ISBN: 9780134291079
Paperback, 218 pages



Cloud Computing Design Patterns
by T. Erl, R. Cope, A. Naserpour
ISBN: 9780133858563
Paperback, 564 pages



Next Generation SOA: A Concise Introduction to Service Technology & Service-Oriented
by T. Erl, C. Gee, J. Kress, B. Maier, H. Normann, P. Raj, L. Shuster, B. Trops, C. Utschig-Utschig, P. Wik, T. Winterberg
ISBN: 9780133859041
Paperback, 208 pages



SOA with Java: Realizing Service-Oriented with Java Technologies
by T. Erl, S. Roy, P. Thomas, A. Tost
ISBN: 9780133859034
Paperback, 590 pages



Cloud Computing: Concepts, Technology & Architecture
by T. Erl, Z. Mahmood, R. Puffini
ISBN: 9780133387520
Paperback, 528 pages



SOA with REST: Principles, Patterns & Constraints for Building Enterprise Solutions with REST
by R. Balasubramanian, B. Carlyle, T. Erl, C. Pautasso
ISBN: 9780137012510
Paperback, 577 pages



SOA Governance: Governing Shared Services On-Premise & in the Cloud
by S. Bennett, T. Erl, C. Gee, R. Laird, A. Manes, R. Schneider, L. Shuster, A. Tost, C. Venable
ISBN: 9780138156756
Paperback, 675 pages



SOA with .NET & Windows Azure: Realizing Service-Oriented with the Microsoft Platform
by D. Chou, J. de Vadoss, T. Erl, N. Gandhi, H. Kommalapati, B. Loesgen, C. Schittko, H. Wilhelmsen, M. Williams
ISBN: 9780131582316
Paperback, 893 pages



SOA Design Patterns
by T. Erl
ISBN: 9780136135166
Paperback, 865 pages



Web Service Contract Design and Versioning for SOA
by T. Erl, H. Haas, A. Karmarkar, C. Liu, D. Orchard, J. Pasley, A. Tost, P. Walmsley, U. Yalcinlap
ISBN: 9780136135173
Paperback, 826 pages



SOA Principles of Service Design
by T. Erl
ISBN: 9780132344821
Paperback, 573 pages



Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services
by T. Erl
ISBN: 9780131428980
Paperback, 534 pages

Praise for the Second Edition

“This book is a solid and comprehensive overview of the cloud concepts and the real mechanics of how the cloud works. It does a nice job of not only explaining cloud function and architecture but the business impact. It’s a great foundation for creating a cloud roadmap.”

—*Jo Peterson, VP Cloud and Security, Clarify360*

“The book provides a comprehensive and well-researched guide to cloud computing. It covers a wide range of topics, from the basics of cloud computing to advanced architectural considerations. The book is written in a clear and concise style and is packed with valuable information and Case Studies. I highly recommend this book to anyone who wants to learn more about cloud computing.”

—*Jorge Blanco, Managing Director, Corporate Reinvention and Education Director, Glumin*

“With its comprehensive insights and vendor-neutral approach, this book truly shines as a beacon of knowledge in the complex world of cloud computing. The author’s emphasis on informed decision-making and alignment with business objectives sets the stage for success in adopting this transformative technology. By navigating the dangers and challenges, the book empowers readers to make strategic choices, safeguarding their organizations from potential pitfalls. From fundamental concepts to practical considerations, each chapter provides a roadmap for unlocking the full potential of cloud computing. With its invaluable case studies, architectural insights, and focus on service quality metrics and SLAs, this book is an indispensable resource for anyone seeking to harness the power of the cloud. It serves as a cornerstone in the Cloud Certified Professional program, providing a solid foundation for further exploration in this dynamic field. A must-have for every forward-thinking professional.”

—*Valther Galván, Chief Information Security Officer*

“This is the most complete book on cloud computing concepts available. It starts with a solid foundation and builds from there. The case studies complement the content and illustrate the possibilities and difficulties.”

—*Emmett Dulaney, University Professor and Author*

Praise for the First Edition

(Affiliations were current when the first edition was released, but may have changed.)

“Cloud computing, more than most disciplines in IT, suffers from too much talk and not enough practice. Thomas Erl has written a timely book that condenses the theory and buttresses it with real-world examples that demystify this important technology. An important guidebook for your journey into the cloud.”

—*Scott Morrison, Chief Technology Officer, Layer 7 Technologies*

“An excellent, extremely well-written, lucid book that provides a comprehensive picture of cloud computing, covering multiple dimensions of the subject. The case studies presented in the book provide a real-world, practical perspective on leveraging cloud computing in an organization. The book covers a wide range of topics, from technology aspects to the business value provided by cloud computing. This is the best, most comprehensive book on the subject—a must-read for any cloud computing practitioner or anyone who wants to get an in-depth picture of cloud computing concepts and practical implementation.”

—*Suzanne D’Souza, SOA/BPM Practice Lead, KBACE Technologies*

“This is a great book on the topic of cloud computing. It is impressive how the content spans from taxonomy, technology, and architectural concepts to important business considerations for cloud adoption. It really does provide a holistic view to this technology paradigm.”

—*Kapil Bakshi, Architecture and Strategy, Cisco Systems Inc.*

“I have read every book written by Thomas Erl and *Cloud Computing* is another excellent publication and demonstration of Thomas Erl’s rare ability to take the most complex topics and provide critical core concepts and technical information in a logical and understandable way.”

—*Melanie A. Allison, Principal, Healthcare Technology Practice,
Integrated Consulting Services*

“Companies looking to migrate applications or infrastructure to the cloud are often misled by buzzwords and industry hype. This work cuts through the hype and provides a detailed look, from investigation to contract to implementation to termination, at what it takes for an organization to engage with cloud service providers. This book really lays out the benefits and struggles with getting a company to an IaaS, PaaS, or SaaS solution.”

—Kevin Davis, Ph.D., *Solutions Architect*

“Thomas, in his own distinct and erudite style, provides a comprehensive and a definitive book on cloud computing. Just like his previous masterpiece, *Service-Oriented Architecture: Concepts, Technology, and Design*, this book is sure to engage CxOs, cloud architects, and the developer community involved in delivering software assets on the cloud. Thomas and his authoring team have taken great pains in providing great clarity and detail in documenting cloud architectures, cloud delivery models, cloud governance, and economics of cloud, without forgetting to explain the core of cloud computing that revolves around Internet architecture and virtualization. As a reviewer for this outstanding book, I must admit I have learned quite a lot while reviewing the material. A ‘must have’ book that should adorn everybody’s desk!”

—Vijay Srinivasan, *Chief Architect - Technology, Cognizant Technology Solutions*

“This book provides comprehensive and descriptive vendor-neutral coverage of cloud computing technology, from both technical and business aspects. It provides a deep-down analysis of cloud architectures and mechanisms that capture the real-world moving parts of cloud platforms. Business aspects are elaborated on to give readers a broader perspective on choosing and defining basic cloud computing business models. Thomas Erl’s *Cloud Computing: Concepts, Technology & Architecture* is an excellent source of knowledge of fundamental and in-depth coverage of cloud computing.”

—Masykur Marhendra Sukmanegara, *Communication Media & Technology, Consulting Workforce Accenture*

“The richness and depth of the topics discussed are incredibly impressive. The depth and breadth of the subject matter are such that a reader could become an expert in a short amount of time.”

—Jamie Ryan, *Solutions Architect, Layer 7 Technologies*

“Demystification, rationalization, and structuring of implementation approaches have always been strong parts in each and every one of Thomas Erl’s books. This book is no exception. It provides the definitive, essential coverage of cloud computing and, most importantly, presents this content in a very comprehensive manner. Best of all, this book follows the conventions of the previous service technology series titles, making it read like a natural extension of the library. I strongly believe that this will be another bestseller from one of the top-selling IT authors of the past decade.”

—*Sergey Popov, Senior Enterprise Architect SOA/Security, Liberty Global International*

“A must-read for anyone involved in cloud design and decision making! This insightful book provides in-depth, objective, vendor-neutral coverage of cloud computing concepts, architecture models, and technologies. It will prove very valuable to anyone who needs to gain a solid understanding of how cloud environments work and how to design and migrate solutions to clouds.”

—*Gijs in 't Veld, Chief Architect, Motion10*

“A reference book covering a wide range of aspects related to cloud providers and cloud consumers. If you would like to provide or consume a cloud service and need to know how, this is your book. The book has a clear structure to facilitate a good understanding of the various concepts of cloud.”

—*Roger Stoffers, Solution Architect*

“Cloud computing has been around for a few years, yet there is still a lot of confusion around the term and what it can bring to developers and deployers alike. This book is a great way of finding out what’s behind the cloud, and not in an abstract or high-level manner: It dives into all of the details that you’d need to know in order to plan for developing applications on cloud and what to look for when using applications or services hosted on a cloud. There are very few books that manage to capture this level of detail about the evolving cloud paradigm as this one does. It’s a must for architects and developers alike.”

—*Dr. Mark Little, Vice President, Red Hat*

“This book provides a comprehensive exploration of the concepts and mechanics behind clouds. It’s written for anyone interested in delving into the details of how cloud environments function, how they are architected, and how they can impact business. This is the book for any organization seriously considering adopting cloud computing. It will pave the way to establishing your cloud computing roadmap.”

—*Damian Maschek, SOA Architect, Deutsche Bahn*

“One of the best books on cloud computing I have ever read. It is complete yet vendor technology neutral and successfully explains the major concepts in a well-structured and disciplined way. It goes through all the definitions and provides many hints for organizations or professionals who are approaching and/or assessing cloud solutions. This book gives a complete list of topics playing fundamental roles in the cloud computing discipline. It goes through a full list of definitions very clearly stated. Diagrams are simple to understand and self-contained. Readers with different skill sets, expertise, and backgrounds will be able to understand the concepts seamlessly.”

—*Antonio Bruno, Infrastructure and Estate Manager, UBS AG*

“*Cloud Computing: Concepts, Technology & Architecture* is a comprehensive book that focuses on what cloud computing is really all about.... This book will become the foundation on which many organizations will build successful cloud adoption projects. It is a must-read reference for both IT infrastructure and application architects interested in cloud computing or involved in cloud adoption projects. It contains extremely useful and comprehensive information for those who need to build cloud-based architectures or need to explain it to customers thinking about adopting cloud computing technology in their organization.”

—*Johan Kumps, SOA Architect, RealDolmen*

“This book defines the basic terminology and patterns for the topic—a useful reference for the cloud practitioner. Concepts from multitenancy to hypervisor are presented in a succinct and clear manner. The underlying case studies provide wonderful real-worldness.”

—*Dr. Thomas Rischbeck, Principal Architect, ipt*

“The book provides a good foundation to cloud services and issues in cloud service design. Chapters highlight key issues that need to be considered in learning how to think in cloud technology terms; this is highly important in today’s business and technology environments where cloud computing plays a central role in connecting user services with virtualized resources and applications.”

—Mark Skilton, *Director, Office of Strategy and Technology,
Global Infrastructure Services, Capgemini*

“The book is well organized and covers basic concepts, technologies, and business models about cloud computing. It defines and explains a comprehensive list of terminologies and glossaries about cloud computing so cloud computing experts can speak and communicate with the same set of standardized language. The book is easy to understand and consistent with early published books from Thomas Erl... It is a must-read for both beginners and experienced professionals.”

—Jian “Jeff” Zhong, *Chief Technology Officer (Acting) and
Chief Architect for SOA and Cloud Computing, Futrend Technology Inc.*

“Students of the related specialties can fulfill their educational process with very easily understood materials that are broadly illustrated and clearly described. Professors of different disciplines, from business analysis to IT implementation—even legal and financial monitoring—can use the book as an on-table lecturing manual. IT specialists of all ranks and fields of application will find the book as a practical and useful support for sketching solutions unbound to any particular vendor or brand.”

—Alexander Gromoff, *Director of Science & Education,
Center of Information Control Technologies, Chairman of BPM Chair in Business
Informatics Department, National Research University “Higher School of Economics”*

“*Cloud Computing: Concepts, Technology & Architecture* is a comprehensive compendium of all the relevant information about the transformative cloud technology. Erl’s latest title concisely and clearly illustrates the origins and positioning of the cloud paradigm as the next-generation computing model. All the chapters are carefully written and arranged in an easy-to-understand manner. This book will be immeasurably beneficial for business and IT professionals. It is set to shake up and help organize the world of cloud computing.”

—Pethuru Raj, *Ph.D., Enterprise Architecture Consultant, Wipro*

This page intentionally left blank

Cloud Computing

Concepts, Technology, Security & Architecture

SECOND EDITION

Thomas Erl
Eric Barceló Monroy

with contributions from
Professor Zaigham Mahmood and Dr. Ricardo Puttini



HOBOKEN, NJ • BOSTON • INDIANAPOLIS • SAN FRANCISCO
NEW YORK • TORONTO • MONTREAL • LONDON • MUNICH • PARIS • MADRID
CAPE TOWN • SYDNEY • TOKYO • SINGAPORE • MEXICO CITY

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

The authors and publisher have taken care in the preparation of this book, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact intlcs@pearson.com.

Visit us on the Web: informit.com

Library of Congress Control Number: 2023939909

Copyright © 2024 Arcitura Education Inc.

Cover image: Thomas Erl

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permissions, request forms and the appropriate contacts within the Pearson Education Global Rights & Permissions Department, please visit www.pearson.com/permissions.

ISBN-13: 978-0-13-805225-6

ISBN-10: 0-13-805225-5

\$PrintCode

Vice President

Mark L. Taub

Director, ITP Product Management

Brett Bartow

Executive Editor

Nancy Davis

Production Editors

Mary Roth

Tonya Simpson

Publishing Coordinator

María Pareni Barceló

Nieves

Technical Reviewers

Jo Peterson

Emmett Dulaney

Development Editor

Eleanor Bru

Indexer

Cheryl Lenser

Proofreader

Paula Lowell

Composer

Bumpy Design

Photos

Thomas Erl

Graphics

Kamilla Bieska

To my family and friends

—Thomas Erl

To Eva, Pareni, Víctor, and Diego with all my love

—Eric Barceló Monroy

This page intentionally left blank

Pearson's Commitment to Diversity, Equity, and Inclusion

Pearson is dedicated to creating bias-free content that reflects the diversity of all learners. We embrace the many dimensions of diversity, including but not limited to race, ethnicity, gender, socioeconomic status, ability, age, sexual orientation, and religious or political beliefs.

Education is a powerful force for equity and change in our world. It has the potential to deliver opportunities that improve lives and enable economic mobility. As we work with authors to create content for every product and service, we acknowledge our responsibility to demonstrate inclusivity and incorporate diverse scholarship so that everyone can achieve their potential through learning. As the world's leading learning company, we have a duty to help drive change and live up to our purpose to help more people create a better life for themselves and to create a better world.

Our ambition is to purposefully contribute to a world where:

- Everyone has an equitable and lifelong opportunity to succeed through learning.
- Our educational products and services are inclusive and represent the rich diversity of learners.
- Our educational content accurately reflects the histories and experiences of the learners we serve.
- Our educational content prompts deeper discussions with learners and motivates them to expand their own learning (and worldview).

While we work hard to present unbiased content, we want to hear from you about any concerns or needs with this Pearson product so that we can investigate and address them.

Please contact us with concerns about any potential bias at <https://www.pearson.com/report-bias.html>.

This page intentionally left blank

Contents at a Glance

Foreword.....	xxxix
About the Authors	xliii
Acknowledgments	xlv
CHAPTER 1: Introduction.....	1
CHAPTER 2: Case Study Background	11
PART I: FUNDAMENTAL CLOUD COMPUTING	21
CHAPTER 3: Understanding Cloud Computing	23
CHAPTER 4: Fundamental Concepts and Models.....	51
CHAPTER 5: Cloud-Enabling Technology.....	79
CHAPTER 6: Understanding Containerization	115
CHAPTER 7: Understanding Cloud Security and Cybersecurity	159
PART II: CLOUD COMPUTING MECHANISMS.....	193
CHAPTER 8: Cloud Infrastructure Mechanisms.....	195
CHAPTER 9: Specialized Cloud Mechanisms	227
CHAPTER 10: Cloud Security and Cybersecurity Access-Oriented Mechanisms	269
CHAPTER 11: Cloud Security and Cybersecurity Data-Oriented Mechanisms.....	311
CHAPTER 12: Cloud Management Mechanisms	325
PART III: CLOUD COMPUTING ARCHITECTURE.....	341
CHAPTER 13: Fundamental Cloud Architectures.....	343
CHAPTER 14: Advanced Cloud Architectures	371
CHAPTER 15: Specialized Cloud Architectures	415
PART IV: WORKING WITH CLOUDS	457
CHAPTER 16: Cloud Delivery Model Considerations.....	459
CHAPTER 17: Cost Metrics and Pricing Models	479
CHAPTER 18: Service Quality Metrics and SLAs.....	503
PART V: APPENDICES	519
APPENDIX A: Case Study Conclusions.....	521
APPENDIX B: Common Containerization Technologies.....	527
Index	543

This page intentionally left blank

Contents

Forewordxxxix
About the Authors	xliii
Acknowledgments	xlv
Chapter 1: Introduction	1
1.1 Objectives of This Book	3
1.2 What This Book Does Not Cover	3
1.3 Who This Book Is For	3
1.4 How This Book Is Organized	4
Part I: Fundamental Computing	4
<i>Chapter 3: Understanding Cloud Computing</i>	4
<i>Chapter 4: Fundamental Concepts and Models</i>	4
<i>Chapter 5: Cloud-Enabling Technology</i>	4
<i>Chapter 6: Understanding Containerization</i>	4
<i>Chapter 7: Understanding Cloud Security and Cybersecurity</i>	5
Part II: Cloud Computing Mechanisms	5
<i>Chapter 8: Cloud Infrastructure Mechanisms</i>	5
<i>Chapter 9: Specialized Cloud Mechanisms</i>	5
<i>Chapter 10: Cloud Security and Cybersecurity Access-Oriented Mechanisms</i>	5
<i>Chapter 11: Cloud Security and Cybersecurity Data-Oriented Mechanisms</i>	6
<i>Chapter 12: Cloud Management Mechanisms</i>	6
Part III: Cloud Computing Architecture	6
<i>Chapter 13: Fundamental Cloud Architectures</i>	6
<i>Chapter 14: Advanced Cloud Architectures</i>	6
<i>Chapter 15: Specialized Cloud Architectures</i>	7
Part IV: Working with Clouds	7
<i>Chapter 16: Cloud Delivery Model Considerations</i>	7
<i>Chapter 17: Cost Metrics and Pricing Models</i>	7
<i>Chapter 18: Service Quality Metrics and SLAs</i>	8
Part V: Appendices	8
<i>Appendix A: Case Study Conclusions</i>	8
<i>Appendix B: Common Containerization Technologies</i>	8

- 1.5 Resources 8
 - Pearson Digital Enterprise Book Series 8
 - Thomas Erl on YouTube. 8
 - The Digital Enterprise* Newsletter on LinkedIn 9
 - Cloud Certified Professional (CCP) Program 9

Chapter 2: Case Study Background 11

- 2.1 Case Study #1: ATN 12
 - Technical Infrastructure and Environment 12
 - Business Goals and New Strategy 13
 - Roadmap and Implementation Strategy 13
- 2.2 Case Study #2: DTGOV 15
 - Technical Infrastructure and Environment 15
 - Business Goals and New Strategy 16
 - Roadmap and Implementation Strategy 17
- 2.3 Case Study #3: Innovartus Technologies Inc. 18
 - Technical Infrastructure and Environment 18
 - Business Goals and Strategy 19
 - Roadmap and Implementation Strategy 19

PART I: FUNDAMENTAL CLOUD COMPUTING 21

Chapter 3: Understanding Cloud Computing 23

- 3.1 Origins and Influences 24
 - A Brief History 24
 - Definitions 25
 - Business Drivers 26
 - Cost Reduction* 26
 - Business Agility* 27
 - Technology Innovations. 28
 - Clustering*. 28
 - Grid Computing*. 28
 - Capacity Planning* 29
 - Virtualization*. 30
 - Containerization*. 31
 - Serverless Environments* 31

3.2 Basic Concepts and Terminology	32
Cloud	32
Container	33
IT Resource	33
On Premises	35
Cloud Consumers and Cloud Providers	35
Scaling	36
<i>Horizontal Scaling</i>	36
<i>Vertical Scaling</i>	36
Cloud Service	37
Cloud Service Consumer	39
3.3 Goals and Benefits	39
Increased Responsiveness	40
Reduced Investments and Proportional Costs	40
Increased Scalability	42
Increased Availability and Reliability	43
3.4 Risks and Challenges	44
Increased Vulnerability Due to Overlapping Trust Boundaries	44
Increased Vulnerability Due to Shared Security Responsibility	44
Increased Exposure to Cyber Threats	46
Reduced Operational Governance Control	46
Limited Portability Between Cloud Providers	48
Multiregional Compliance and Legal Issues	49
Cost Overruns	49

Chapter 4: Fundamental Concepts and Models51

4.1 Roles and Boundaries	52
Cloud Provider	52
Cloud Consumer	52
Cloud Broker	53
Cloud Service Owner	54
Cloud Resource Administrator	55
Additional Roles	57
Organizational Boundary	57
Trust Boundary	58

- 4.2 Cloud Characteristics 59
 - On-Demand Usage 59
 - Ubiquitous Access 60
 - Multitenancy (and Resource Pooling). 60
 - Elasticity 60
 - Measured Usage. 62
 - Resiliency 62
- 4.3 Cloud Delivery Models 62
 - Infrastructure as a Service (IaaS). 64
 - Platform as a Service (PaaS) 64
 - Software as a Service (SaaS) 66
 - Comparing Cloud Delivery Models 67
 - Combining Cloud Delivery Models 68
 - IaaS + PaaS*. 68
 - IaaS + PaaS + SaaS*. 71
 - Cloud Delivery Submodels 72
- 4.4 Cloud Deployment Models. 74
 - Public Clouds 74
 - Private Clouds. 74
 - Multiclouds 77
 - Hybrid Clouds. 77

Chapter 5: Cloud-Enabling Technology 79

- 5.1 Networks and Internet Architecture 80
 - Internet Service Providers (ISPs) 80
 - Connectionless Packet Switching (Datagram Networks) 82
 - Router-Based Interconnectivity 83
 - Physical Network* 84
 - Transport Layer Protocol* 84
 - Application Layer Protocol* 84
 - Technical and Business Considerations 84
 - Connectivity Issues* 84
 - Network Bandwidth and Latency Issues* 87
 - Wireless and Cellular* 88
 - Cloud Carrier and Cloud Provider Selection* 89

5.2 Cloud Data Center Technology	89
Virtualization	89
Standardization and Modularity	90
Autonomic Computing	91
Remote Operation and Management	91
High Availability	91
Security-Aware Design, Operation, and Management	92
Facilities	92
Computing Hardware	92
Storage Hardware	93
Network Hardware	94
<i>Carrier and External Networks Interconnection</i>	<i>94</i>
<i>Web-Tier Load Balancing and Acceleration</i>	<i>94</i>
<i>LAN Fabric</i>	<i>95</i>
<i>SAN Fabric</i>	<i>95</i>
<i>NAS Gateways</i>	<i>95</i>
Serverless Environments	95
NoSQL Clustering	96
Other Considerations	98
5.3 Modern Virtualization	99
Hardware Independence	99
Server Consolidation	99
Resource Replication	100
Operating System–Based Virtualization	100
Hardware–Based Virtualization	102
Containers and Application–Based Virtualization	103
Virtualization Management	104
Other Considerations	104
5.4 Multitenant Technology	105
5.5 Service Technology and Service APIs	107
REST Services	107
Web Services	108
Service Agents	110
Service Middleware	110
Web-Based RPC	111
5.6 Case Study Example	111

Chapter 6: Understanding Containerization 115

6.1	Origins and Influences	116
	A Brief History	116
	Containerization and Cloud Computing	117
6.2	Fundamental Virtualization and Containerization	117
	Operating System Basics	117
	Virtualization Basics	118
	Physical Servers	118
	Virtual Servers	118
	Hypervisors	119
	Virtualization Types	119
	Containerization Basics	121
	Containers	121
	Container Images	121
	Container Engines	121
	Pods	122
	Hosts	122
	Host Clusters	124
	Host Networks and Overlay Networks	125
	Virtualization and Containerization	125
	Containerization on Physical Servers	125
	Containerization on Virtual Servers	126
	Containerization Benefits	127
	Containerization Risks and Challenges	128
6.3	Understanding Containers	129
	Container Hosting	129
	Containers and Pods	130
	Container Instances and Clusters	133
	Container Package Management	133
	Container Orchestration	136
	Container Package Manager vs. Container Orchestrator	139
	Container Networks	139
	Container Network Scope	140
	Container Network Addresses	142
	Rich Containers	144
	Other Common Container Characteristics	145

6.4 Understanding Container Images	145
Container Image Types and Roles	145
Container Image Immutability	147
Container Image Abstraction	147
<i>Operating System Kernel Abstraction</i>	147
<i>Operating System Abstraction Beyond the Kernel</i>	148
Container Build Files	149
<i>Container Image Layers</i>	149
How Customized Container Images Are Created	151
6.5 Multi-Container Types	152
Sidecar Container	152
Adapter Container	154
Ambassador Container	155
Using Multi-Containers Together	157
6.6 Case Study Example	158

Chapter 7: Understanding Cloud Security and Cybersecurity159

7.1 Basic Security Terminology	160
Confidentiality	160
Integrity	161
Availability	161
Authenticity	162
Security Controls	162
Security Mechanisms	163
Security Policies	163
7.2 Basic Threat Terminology	163
Risk	163
Vulnerability	163
Exploit	163
Zero-Day Vulnerability	164
Security Breach	164
Data Breach	164
Data Leak	164
Threat (or Cyber Threat)	164

Attack (or Cyber Attack)	164
Attacker and Intruder	164
Attack Vector and Surface	165
7.3 Threat Agents	165
Anonymous Attacker	166
Malicious Service Agent	167
Trusted Attacker	167
Malicious Insider	167
7.4 Common Threats	168
Traffic Eavesdropping	168
Malicious Intermediary	168
Denial of Service	169
Insufficient Authorization	171
Virtualization Attack	172
Overlapping Trust Boundaries	173
Containerization Attack	174
Malware	175
Insider Threat	177
Social Engineering and Phishing	178
Botnet	178
Privilege Escalation	181
Brute Force	182
Remote Code Execution	182
SQL Injection	183
Tunneling	184
Advanced Persistent Threat (APT)	185
7.5 Case Study Example	187
7.6 Additional Considerations	188
Flawed Implementations	188
Security Policy Disparity	188
Contracts	189
Risk Management	190
7.7 Case Study Example	191

PART II: CLOUD COMPUTING MECHANISMS 193**Chapter 8: Cloud Infrastructure Mechanisms195**

8.1 Logical Network Perimeter	196
Case Study Example.	198
8.2 Virtual Server	200
Case Study Example.	201
8.3 Hypervisor	205
Case Study Example.	206
8.4 Cloud Storage Device	207
Cloud Storage Levels	208
Network Storage Interfaces.	208
Object Storage Interfaces	209
Database Storage Interfaces.	210
<i>Relational Data Storage</i>	210
<i>Non-Relational Data Storage</i>	210
Case Study Example.	211
8.5 Cloud Usage Monitor	214
Monitoring Agent.	214
Resource Agent	215
Polling Agent.	215
Case Study Example.	216
8.6 Resource Replication	220
Case Study Example.	221
8.7 Ready-Made Environment	224
Case Study Example.	225
8.8 Container	226

Chapter 9: Specialized Cloud Mechanisms227

9.1 Automated Scaling Listener	228
Case Study Example.	230
9.2 Load Balancer	234
Case Study Example.	235

- 9.3 SLA Monitor 236
 - Case Study Example. 238
 - SLA Monitor Polling Agent*. 238
 - SLA Monitoring Agent* 238
- 9.4 Pay-Per-Use Monitor 242
 - Case Study Example. 245
- 9.5 Audit Monitor 247
 - Case Study Example. 247
- 9.6 Failover System 249
 - Active–Active. 249
 - Active–Passive 252
 - Case Study Example. 254
- 9.7 Resource Cluster 259
 - Case Study Example. 262
- 9.8 Multi-Device Broker 263
 - Case Study Example. 265
- 9.9 State Management Database. 265
 - Case Study Example. 266

Chapter 10: Cloud Security and Cybersecurity
Access-Oriented Mechanisms 269

- 10.1 Encryption 271
 - Symmetric Encryption. 272
 - Asymmetric Encryption. 272
 - Case Study Example. 273
- 10.2 Hashing 274
 - Case Study Example. 275
- 10.3 Digital Signature 276
 - Case Study Example. 278
- 10.4 Cloud-Based Security Groups 280
 - Case Study Example. 282
- 10.5 Public Key Infrastructure (PKI) System 284
 - Case Study Example. 286

10.6 Single Sign-On (SSO) System	287
Case Study Example	289
10.7 Hardened Virtual Server Image	290
Case Study Example	291
10.8 Firewall	292
Case Study Example	293
10.9 Virtual Private Network (VPN)	293
Case Study Example	294
10.10 Biometric Scanner	295
Case Study Example	296
10.11 Multi-Factor Authentication (MFA) System	297
Case Study Example	298
10.12 Identity and Access Management (IAM) System	298
Case Study Example	301
10.13 Intrusion Detection System (IDS)	301
Case Study Example	302
10.14 Penetration Testing Tool	302
Case Study Example	304
10.15 User Behavior Analytics (UBA) System	304
Case Study Example	305
10.16 Third-Party Software Update Utility	306
Case Study Example	308
10.17 Network Intrusion Monitor	308
Case Study Example	308
10.18 Authentication Log Monitor	309
Case Study Example	309
10.19 VPN Monitor	309
Case Study Example	310
10.20 Additional Cloud Security Access-Oriented Practices and Technologies	310

**Chapter 11: Cloud Security and Cybersecurity
Data-Oriented Mechanisms311**

11.1 Digital Virus Scanning and Decryption System	312
Generic Decryption	313
Digital Immune System	313
Case Study Example.	315
11.2 Malicious Code Analysis System	315
Case Study Example.	316
11.3 Data Loss Prevention (DLP) System	317
Case Study Example.	318
11.4 Trusted Platform Module (TPM).	319
Case Study Example.	320
11.5 Data Backup and Recovery System	320
Case Study Example.	322
11.6 Activity Log Monitor	322
Case Study Example.	322
11.7 Traffic Monitor.	323
Case Study Example.	323
11.8 Data Loss Protection Monitor	323
Case Study Example.	324

Chapter 12: Cloud Management Mechanisms325

12.1 Remote Administration System	326
Case Study Example.	331
12.2 Resource Management System	331
Case Study Example.	333
12.3 SLA Management System	334
Case Study Example.	336
12.4 Billing Management System	337
Case Study Example.	339

PART III: CLOUD COMPUTING ARCHITECTURE 341**Chapter 13: Fundamental Cloud Architectures343**

13.1 Workload Distribution Architecture	344
13.2 Resource Pooling Architecture	346
13.3 Dynamic Scalability Architecture.	350
13.4 Elastic Resource Capacity Architecture	353
13.5 Service Load Balancing Architecture	355
13.6 Cloud Bursting Architecture	358
13.7 Elastic Disk Provisioning Architecture	359
13.8 Redundant Storage Architecture	363
13.9 Multicloud Architecture.	365
13.10 Case Study Example	368

Chapter 14: Advanced Cloud Architectures.371

14.1 Hypervisor Clustering Architecture	373
14.2 Virtual Server Clustering Architecture	379
14.3 Load-Balanced Virtual Server Instances Architecture	380
14.4 Nondisruptive Service Relocation Architecture	383
14.5 Zero Downtime Architecture	388
14.6 Cloud Balancing Architecture	389
14.7 Resilient Disaster Recovery Architecture.	391
14.8 Distributed Data Sovereignty Architecture.	393
14.9 Resource Reservation Architecture.	395
14.10 Dynamic Failure Detection and Recovery Architecture	399
14.11 Rapid Provisioning Architecture.	402
14.12 Storage Workload Management Architecture	406
14.13 Virtual Private Cloud Architecture	411
14.14 Case Study Example.	413

Chapter 15: Specialized Cloud Architectures415

- 15.1 Direct I/O Access Architecture 417
- 15.2 Direct LUN Access Architecture 419
- 15.3 Dynamic Data Normalization Architecture. 421
- 15.4 Elastic Network Capacity Architecture 423
- 15.5 Cross-Storage Device Vertical Tiering Architecture. 424
- 15.6 Intra-Storage Device Vertical Data Tiering Architecture. . . 429
- 15.7 Load-Balanced Virtual Switches Architecture 432
- 15.8 Multipath Resource Access Architecture 434
- 15.9 Persistent Virtual Network Configuration Architecture. . . 436
- 15.10 Redundant Physical Connection for Virtual Servers Architecture 439
- 15.11 Storage Maintenance Window Architecture. 441
- 15.12 Edge Computing Architecture. 449
- 15.13 Fog Computing Architecture 450
- 15.14 Virtual Data Abstraction Architecture. 452
- 15.15 Metacloud Architecture. 453
- 15.16 Federated Cloud Application Architecture. 454

PART IV: WORKING WITH CLOUDS 457

Chapter 16: Cloud Delivery Model Considerations.459

- 16.1 Cloud Delivery Models: The Cloud Provider Perspective. 460
 - Building IaaS Environments. 460
 - Data Centers* 461
 - Scalability and Reliability* 463
 - Monitoring* 463
 - Security* 464

Equipping PaaS Environments	464
<i>Scalability and Reliability</i>	465
<i>Monitoring</i>	467
<i>Security</i>	467
Optimizing SaaS Environments	467
<i>Security</i>	470
16.2 Cloud Delivery Models: The Cloud Consumer Perspective	471
Working with IaaS Environments	471
<i>IT Resource Provisioning Considerations</i>	472
Working with PaaS Environments	473
<i>IT Resource Provisioning Considerations</i>	474
Working with SaaS Services	475
16.3 Case Study Example	476

Chapter 17: Cost Metrics and Pricing Models479

17.1 Business Cost Metrics	480
Up-Front and Ongoing Costs	480
Additional Costs	481
Case Study Example	482
Product Catalog Browser	482
<i>On-Premises Up-Front Costs</i>	482
<i>On-Premises Ongoing Costs</i>	483
<i>Cloud-Based Up-Front Costs</i>	483
<i>Cloud-Based Ongoing Costs</i>	483
17.2 Cloud Usage Cost Metrics	485
Network Usage	485
<i>Inbound Network Usage Metric</i>	485
<i>Outbound Network Usage Metric</i>	486
<i>Intra-Cloud WAN Usage Metric</i>	486
Server Usage	487
<i>On-Demand Virtual Machine Instance Allocation Metric</i>	487
<i>Reserved Virtual Machine Instance Allocation Metric</i>	487
Cloud Storage Device Usage	488
<i>On-Demand Storage Space Allocation Metric</i>	488
<i>I/O Data Transferred Metric</i>	488

Cloud Service Usage	488
<i>Application Subscription Duration Metric</i>	488
<i>Number of Nominated Users Metric</i>	489
<i>Number of Transactions Users Metric</i>	489
17.3 Cost Management Considerations	489
Pricing Models	491
Multicloud Cost Management	493
Additional Considerations	495
Case Study Example	496
Virtual Server On-Demand Instance Allocation	497
Virtual Server Reserved Instance Allocation	499
Cloud Storage Device	501
WAN Traffic	501

Chapter 18: Service Quality Metrics and SLAs503

18.1 Service Quality Metrics	504
Service Availability Metrics	505
<i>Availability Rate Metric</i>	505
<i>Outage Duration Metric</i>	506
Service Reliability Metrics	507
<i>Mean Time Between Failures (MTBF) Metric</i>	507
<i>Reliability Rate Metric</i>	507
Service Performance Metrics	507
<i>Network Capacity Metric</i>	508
<i>Storage Device Capacity Metric</i>	508
<i>Server Capacity Metric</i>	508
<i>Web Application Capacity Metric</i>	508
<i>Instance Starting Time Metric</i>	509
<i>Response Time Metric</i>	509
<i>Completion Time Metric</i>	509
Service Scalability Metrics	509
<i>Storage Scalability (Horizontal) Metric</i>	510
<i>Server Scalability (Horizontal) Metric</i>	510
<i>Server Scalability (Vertical) Metric</i>	510
Service Resiliency Metrics	511
<i>Mean Time to Switchover (MTSO) Metric</i>	511
<i>Mean Time to System Recovery (MTSR) Metric</i>	512
18.2 Case Study Example	512

18.3 SLA Guidelines. 513

18.4 Case Study Example 516

Scope and Applicability 516

Service Quality Guarantees 516

Definitions 517

Usage of Financial Credits 517

SLA Exclusions 518

PART V: APPENDICES 519

Appendix A: Case Study Conclusions.521

A.1 ATN 522

A.2 DTGOV 522

A.3 Innovartus 524

Appendix B: Common Containerization Technologies527

B.1 Docker 528

 Docker Server 528

 Docker Client 529

 Docker Registry 530

 Docker Objects 532

 Docker Swarm (Container Orchestrator) 533

B.2 Kubernetes 534

 Kubernetes Node (Host) 534

 Kubernetes Pod 535

 Kubelet 536

 Kube-Proxy 536

 Container Runtime (Container Engine). 537

 Cluster 538

 Kubernetes Control Plane 539

Index543

This page intentionally left blank

Foreword

by David S. Linthicum

Finally, an owner's manual for cloud computing.

Most enterprises got cloud computing wrong. Not “Going out of business” wrong, but the majority ended up with under-optimized cloud-based systems that failed to return the value stakeholders expected.

What happened? Most people blame over-hyped technology, “cloud washing,” and faster-than-needed movement to cloud-based platforms. The honest answer is that there were and still are not enough qualified cloud computing solutions designers and builders to go around. Even the cloud salespeople started with too little cloud expertise to adequately advise their clients.

It's hard to gain experience and qualifications with a complex new technology that requires a mostly custom solution for every implementation, especially when the cloud “pioneers” are in such high demand that they rarely have time to teach others their skills.

For far too long we've worked off the assumption that if something works, it's also optimized. The unoptimized result in a cloud deployment is a solution that removes value from the business over time. Keep replicating these mistakes and you will soon enjoy a negative value from the use of cloud computing.

Back in 2008 and 2009, when cloud computing hype first arose in the fast-moving technology market, promises of 10-fold cloud ROIs were common. Instead of getting \$10 back on every dollar invested, most enterprises only return about \$0.50 back on each dollar invested.

Think of the problem this way: Flying from LA to New York on a budget carrier in coach class will cost about 1% of the fare on a private jet. Both planes will get you from Point A to Point B, but too many enterprise clouds are chartered jets. As with flight costs, many happy medium choices are available in cloud computing that will result in a satisfactory compromise between efficiency and costs. This compromise requires understanding the data, security, governance, and required application behaviors that need to be addressed with a carefully configured cloud computing architecture and enabling technology that creates a fully optimized solution pattern.

The Missing Manual

What we have is an education problem rather than a technology problem. Most enterprises faked their way through their initial cloud implementations using bits and pieces of what they understood from more traditional technology platforms. There are too many vastly reaching assumptions about the capabilities of the emerging cloud computing technology.

Of course, no single source can provide all the knowledge of what “the cloud” is and does. This book stands out as a source of practical knowledge that offers a comprehensive understanding of cloud technology and how it can be effectively utilized to solve most business problems using standard and advanced cloud architecture concepts. Better put, this book provides you with the knowledge needed to find the value of cloud computing that was initially promised.

Like most good owner’s manuals, this book includes the basics that serve as a “quick start” guide as well as advice to successfully leverage cloud mechanisms. Erl then delves into advanced concepts that can only be learned through experience. The basics will get you through a cloud job interview. Erl’s discussions on advanced concepts surpass what most of us in the cloud architecture field have currently considered.

What I find most engaging is that Erl does not focus on specific technology brands, understanding that those technologies will quickly evolve. Good solutions begin as a concept. Unfortunately, we often misunderstand what those solutions should do or be by inserting specific branded technology too early in the process. This is especially true when designing and building cloud computing solutions. Erl leaves brands out of the discussion, making the concepts in this book much more useful and applicable across different technologies and through the evolution of technologies over time.

With the heart of a teacher, Erl puts what others understand into a useful aggregation of that knowledge. Read this book to educate yourself on cloud computing concepts, designs, architecture, and other advanced concepts in a structure that builds upon other concepts in logical ways. The information imparted will make sense to those who are in the early days of their cloud journey, as well as to those who are more advanced. This is a manual that’s useful to all levels and for all needs. It’s a reference you will return to many times in your own cloud computing journey to ensure you’re doing things correctly.

Finally, Find the Value of Cloud Computing

I suspect that most of you are here because you've seen cloud computing overwhelm your business and you are wondering how to fix it. This is the only well-structured and complete manual you'll need to figure out how you can get cloud computing right. Turn the concepts presented in this book into optimized solutions that maximize the value returned to the business.

This book is about making the right choices, understanding why those choices are made, and determining the best choices for the business. If there is a user manual for cloud computing, both advanced and basic concepts, this is it.

It will help you better understand the correct application of any technology and its usefulness in solving your problems. Indeed, avoiding going down many of the "rabbit holes" that can either waste time, or more likely lead you to the wrong decisions.

Happy computing.

David S. Linthicum

Author, Speaker, Educator, and Consultant

This page intentionally left blank

About the Authors



Thomas Erl

Thomas Erl is a best-selling IT author and series editor of the Pearson Digital Enterprise Series from Thomas Erl. Thomas has authored and co-authored 15 books published by Pearson Education and Prentice Hall dedicated to contemporary business technology and practices. You can find Thomas on the Thomas Erl YouTube channel (youtube.com/@terl). He is also the host of the *Real Digital Transformation* podcast series (available via Spotify, Apple, Google Podcasts, and most other platforms) and also publishes the weekly LinkedIn newsletter *The Digital Enterprise*. Over 100 articles and interviews by Thomas have been published in numerous publications, including *CEO World*, *The Wall Street Journal*, *Forbes*, and *CIO Magazine*. Thomas has also toured over 20 countries as a keynote speaker for various conferences and events.

At Arcitura Education (www.arcitura.com), Thomas leads the development of curricula for internationally recognized, vendor-neutral training and accreditation programs. Arcitura's portfolio currently consists of over 100 courses, over 100 Pearson VUE exams, and over 40 certification tracks, covering topics such as Cloud Computing Architecture, Security, and Governance, as well as Digital Transformation, Robotic Process Automation (RPA), DevOps, Blockchain, IoT, Containerization, Machine Learning, Artificial Intelligence (AI), Cybersecurity, Service-Oriented Architecture (SOA), and Big Data Analytics. Thomas is also the founder and senior advisor at Transformative Digital Solutions (www.transformative.digital) and a freelance LinkedIn instructor and courseware author.

www.thomaserl.com

**Eric Barceló Monroy**

Eric Barceló Monroy is an IT professional with extensive experience in IT strategic planning, operational and administrative process re-engineering, system implementation project management, and IT operations. He has a proven track record of implementing systems that exceed user expectations while reducing costs and improving response times. He has held various high-level positions in both the private and public sectors, including Director of Information Technology at Farmacéuticos MAYPO, a pharmaceutical distributor; Vice-president of Telecommunications and Technology Operations at iExplore, an internet-based adventure travel agency; and Director of Information Technology and Telecommunications at the Ministry of Education in Tabasco, Mexico, where he oversaw the implementation of telecommunication networks among schools and develops and delivers computer literacy training programs for faculty.

Additionally, he is a partner and Technical Consulting Director at EGN, a cloud technology consulting and training firm, where he provides IT consultancy on state-of-the-art topics like Big Data, Cloud Computing, Virtualization, Advanced Networking, and Strategic IT Management. Eric is a Certified Cloud Computing Technology Professional, Certified Cloud Virtualization Specialist, and Certified Cloud Architect, among others. He is also a VMware Certified Professional, Red Hat Certified System Administrator, Red Hat Certified Engineer, and Certified Amazon Web Services Solutions Architect.

Acknowledgments

We would like to acknowledge the co-authors of the first edition of this book:

- Prof Zaigham Mahmood, Derby, UK
- Ricardo Puttini, PhD, Core Consulting

Acknowledgments for the second edition in alphabetical order by last name:

- Gustavo Azzolin
- Jorge Blanco, Managing Director, Corporate Reinvention and Education Director, Glumin
- Emmett Dulaney, University Professor and Author
- Valther Galván, Chief Information Security Officer
- David Linthicum, Deloitte Consulting
- Vinícius Pacheco, University of Brasília, Brazil
- Jo Peterson, VP Cloud and Security, Clarify360
- Pamela J. Wise-Martinez, Global Chief Architect, Whirlpool Corporation
- Matthias Ziegler

Acknowledgments for the first edition in alphabetical order by last name (affiliations were current when the first edition was released, but may have changed):

- Ahmed Aamer, AlFaisaliah Group
- Randy Adkins, Modus21
- Melanie Allison, Integrated Consulting Services
- Gabriela Inacio Alves, University of Brasilia
- Marcelo Ancelmo, IBM Rational Software Services
- Kapil Bakshi, Cisco Systems
- Toufic Boubez, Metafor Software
- Antonio Bruno, UBS AG
- Dr. Paul Buhler, Modus21
- Pethuru Raj Cheliah, Wipro
- Kevin Davis, Ph.D.
- Suzanne D’Souza, KBACE Technologies
- Yili Gong, Wuhan University
- Alexander Gromoff, Center of Information Control Technologies
- Chris Haddad, WSO2
- Richard Hill, University of Derby
- Dr. Michaela Iorga, Ph.D.
- Johan Kumps, RealDolmen
- Gijs in ’t Veld, Motion10
- Masykur Marhendra, Consulting Workforce Accenture
- Damian Maschek, Deutsche Bahn
- Claynor Mazzarolo, IBTI
- Charlie Mead, W3C
- Steve Millidge, C2B2

- Jorge Minguez, Thales Deutschland
- Scott Morrison, Layer 7
- Amin Naserpour, HP
- Vicente Navarro, European Space Agency
- Laura Olson, IBM WebSphere
- Tony Pallas, Intel
- Cesare Pautasso, University of Lugano
- Sergey Popov, Liberty Global International
- Olivier Poupeney, Dreamface Interactive
- Alex Rankov, EMC
- Dan Rosanova, West Monroe Partners
- Jaime Ryan, Layer 7
- Filippas Santas, Credit Suisse
- Christoph Schittko, Microsoft
- Guido Schmutz, Trivadis
- Mark Skilton, Capgemini
- Gary Smith, CloudComputingArchitect.com
- Kevin Spiess
- Vijay Srinivasan, Cognizant
- Daniel Starcevich, Raytheon
- Roger Stoffers, HP
- Andre Toffanello, IBTI
- Andre Tost, IBM Software Group
- Bernd Trops, talend
- Clemens Utschig, Boehringer Ingelheim Pharma
- Ignaz Wanders, Archimiddle

- Philip Wik, Redflex
- Jorge Williams, Rackspace
- Dr. Johannes Maria Zaha
- Jeff Zhong, Futrend Technologies

Special thanks to the research and development teams at Arcitura Education (www.arcitura.com) that produced the Cloud Computing, Cloud Architecture, Containerization Architecture, Cloud Security and Cybersecurity courses upon which this book is based.

Chapter 12



Cloud Management Mechanisms

12.1 Remote Administration System

12.2 Resource Management System

12.3 SLA Management System

12.4 Billing Management System

Cloud-based IT resources need to be set up, configured, maintained, and monitored. The systems covered in this chapter are mechanisms that encompass and enable these types of management tasks. They form key parts of cloud technology architectures by facilitating the control and evolution of the IT resources that form cloud platforms and solutions.

The following management-related mechanisms are described in this chapter:

- Remote Administration System
- Resource Management System
- SLA Management System
- Billing Management System

These systems typically provide integrated APIs and can be offered as individual products, custom applications, or combined into various product suites or multifunction applications.

12.1 Remote Administration System

The *remote administration system* mechanism (Figure 12.1) provides tools and user interfaces for external cloud resource administrators to configure and administer cloud-based IT resources.

A remote administration system can establish a portal for access to administration and management features of various underlying systems, including the resource management, SLA management, and billing management systems described in this chapter (Figure 12.2).

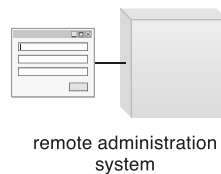
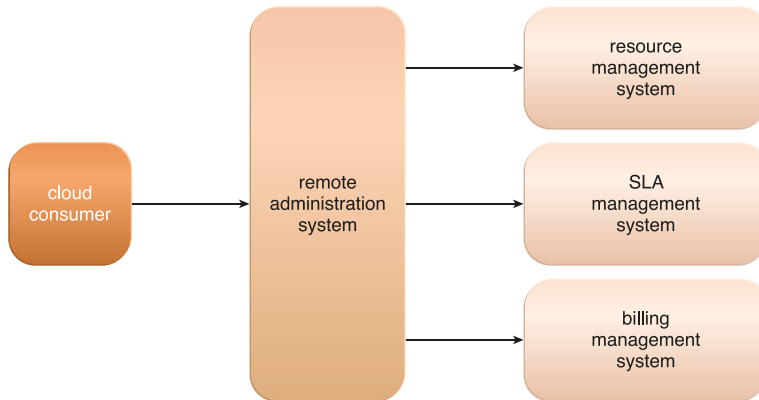


Figure 12.1

The symbol used in this book for the remote administration system. The displayed user interface will typically be labeled to indicate a specific type of portal.

**Figure 12.2**

The remote administration system abstracts underlying management systems to expose and centralize administration controls to external cloud resource administrators. The system provides a customizable user console, while programmatically interfacing with underlying management systems via their APIs.

The tools and APIs provided by a remote administration system are generally used by the cloud provider to develop and customize online portals that provide cloud consumers with a variety of administrative controls.

The following are the two primary types of portals that are created with the remote administration system:

- *Usage and Administration Portal* – A general-purpose portal that centralizes management controls to different cloud-based IT resources and can further provide IT resource usage reports. This portal is part of numerous cloud technology architectures covered in Chapters 13 to 15.
- *Self-Service Portal* – This is essentially a shopping portal that allows cloud consumers to search an up-to-date list of cloud services and IT resources that are available from a cloud provider (usually for lease). The cloud consumer submits its chosen items to the cloud provider for provisioning. This portal is primarily associated with the rapid provisioning architecture described in Chapter 14.



usage and administration portal



self-service portal

Figure 12.3 illustrates a scenario involving a remote administration system and both the usage and administration and self-service portals.

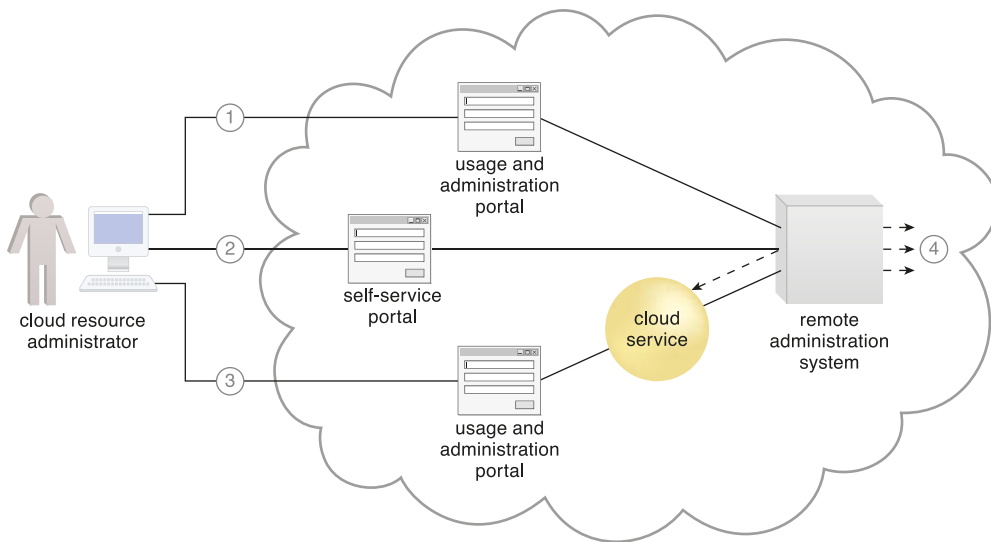


Figure 12.3

A cloud resource administrator uses the usage and administration portal to configure an already leased virtual server (not shown) to prepare it for hosting (1). The cloud resource administrator then uses the self-service portal to select and request the provisioning of a new cloud service (2). The cloud resource administrator then accesses the usage and administration portal again to configure the newly provisioned cloud service that is hosted on the virtual server (3). Throughout these steps, the remote administration system interacts with the necessary management systems to perform the requested actions (4).

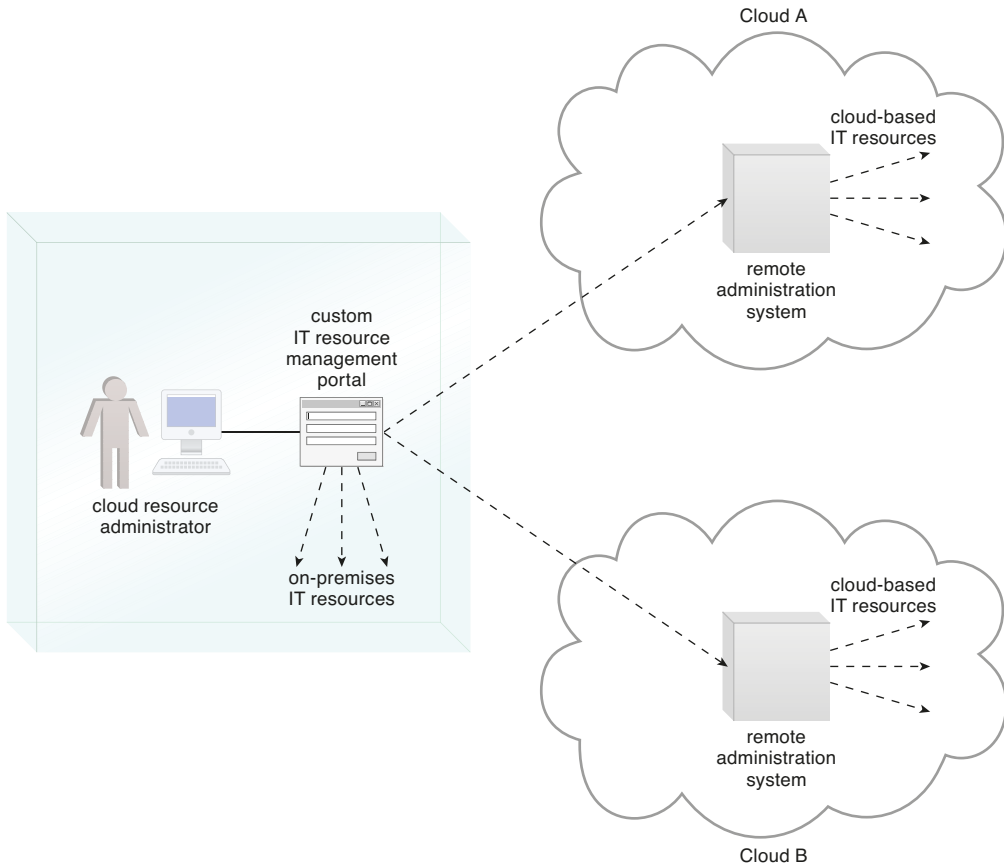
Depending on:

- the type of cloud product or cloud delivery model the cloud consumer is leasing or using from the cloud provider,
- the level of access control granted by the cloud provider to the cloud consumer, and
- which underlying management systems the remote administration system interfaces with,

...the following tasks can commonly be performed by cloud consumers via a remote administration console:

- configuring and setting up cloud services
- provisioning and releasing IT resource for on-demand cloud services
- monitoring cloud service status, usage, and performance
- monitoring QoS and SLA fulfillment
- managing leasing costs and usage fees
- managing user accounts, security credentials, authorization, and access control
- tracking internal and external access to leased services
- planning and assessing IT resource provisioning
- capacity planning

While the user interface provided by the remote administration system will tend to be proprietary to the cloud provider, there is a preference among cloud consumers to work with remote administration systems that offer standardized APIs. This allows a cloud consumer to invest in the creation of its own front-end with the foreknowledge that it can reuse this console if it decides to move to another cloud provider that supports the same standardized API. Additionally, the cloud consumer would be able to further leverage standardized APIs if it is interested in leasing and centrally administering IT resources from multiple cloud providers and/or IT resources residing in cloud and on-premises environments (Figure 12.4).

**Figure 12.4**

Standardized APIs published by remote administration systems from different clouds enable a cloud consumer to develop a custom portal that centralizes a single IT resource management portal for both cloud-based and on-premises IT resources.

CASE STUDY EXAMPLE

DTGOV has been offering its cloud consumers a user-friendly remote administration system for some time and recently determined that upgrades are required to accommodate the growing number of cloud consumers and the increasing diversity of requests. DTGOV is planning a development project to extend the remote administration system to fulfill the following requirements:

- Cloud consumers need to be able to self-provision virtual servers and virtual storage devices. The system specifically needs to interoperate with the cloud-enabled VIM platform's proprietary API to enable self-provisioning capabilities.
- A single sign-on mechanism (described in Chapter 10) needs to be incorporated to centrally authorize and control cloud consumer access.
- An API that supports the provisioning, starting, stopping, releasing, up-down scaling, and replicating of commands for virtual servers and cloud storage devices needs to be exposed.

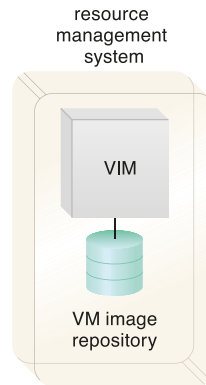
In support of these features, a self-service portal is developed and the feature set of DTGOV's existing usage and administration portal is extended.

12.2 Resource Management System

The *resource management system* mechanism helps coordinate IT resources in response to management actions performed by both cloud consumers and cloud providers (Figure 12.5). Core to this system is the virtual infrastructure manager (VIM) that coordinates the server hardware so that virtual server instances can be created from the most expedient underlying physical server. A VIM is a commercial product that can be used to manage a range of virtual IT resources across multiple physical servers. For example, a VIM can create and manage multiple instances of a hypervisor across different physical servers or allocate a virtual server on one physical server to another (or to a resource pool).

Figure 12.5

A resource management system encompassing a VIM platform and a virtual machine image repository. The VIM may have additional repositories, including one dedicated to storing operational data.



Tasks that are typically automated and implemented through the resource management system include:

- managing virtual IT resource templates that are used to create prebuilt instances, such as virtual server images
- allocating and releasing virtual IT resources into the available physical infrastructure in response to the starting, pausing, resuming, and termination of virtual IT resource instances
- coordinating IT resources in relation to the involvement of other mechanisms, such as resource replication, load balancer, and failover system
- enforcing usage and security policies throughout the lifecycle of cloud service instances
- monitoring operational conditions of IT resources

Resource management system functions can be accessed by cloud resource administrators employed by the cloud provider or cloud consumer. Those working on behalf of a cloud provider will often be able to directly access the resource management system's native console.

Resource management systems typically expose APIs that allow cloud providers to build remote administration system portals that can be customized to selectively offer resource management controls to external cloud resource administrators acting on behalf of cloud consumer organizations via usage and administration portals.

Both forms of access are depicted in Figure 12.6.

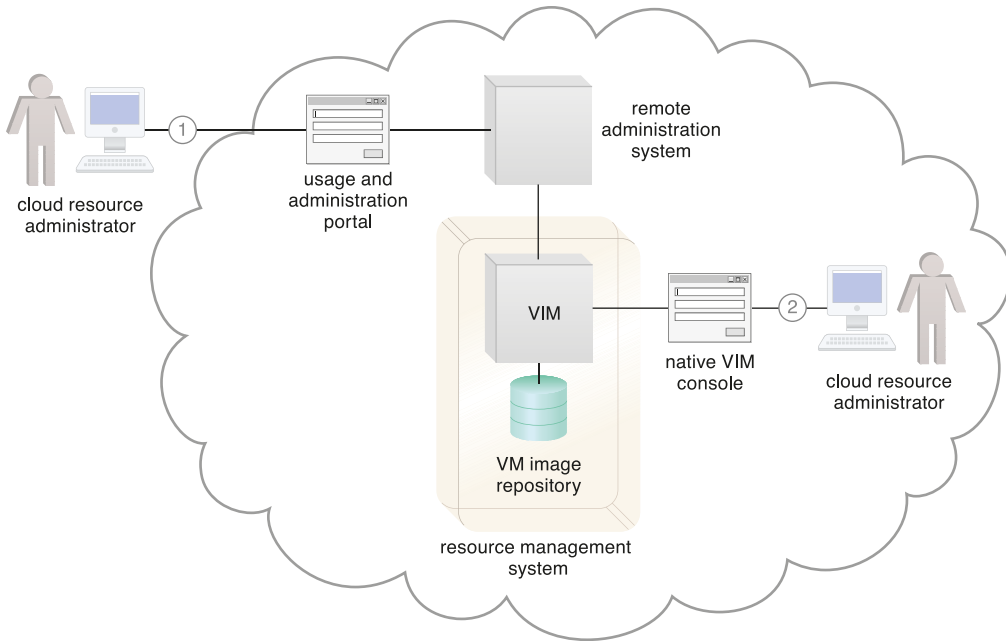


Figure 12.6

The cloud consumer's cloud resource administrator accesses a usage and administration portal externally to administer a leased IT resource (1). The cloud provider's cloud resource administrator uses the native user interface provided by the VIM to perform internal resource management tasks (2).

CASE STUDY EXAMPLE

The DTGOV resource management system is an extension of a new VIM product it purchased and provides the following primary features:

- management of virtual IT resources with a flexible allocation of pooled IT resources across different data centers
- management of cloud consumer databases
- isolation of virtual IT resources at logical perimeter networks
- management of a template virtual server image inventory available for immediate instantiation

- automated replication (“snapshotting”) of virtual server images for virtual server creation
- automated up–down scaling of virtual servers according to usage thresholds to enable live VM migration among physical servers
- an API for the creation and management of virtual servers and virtual storage devices
- an API for the creation of network access control rules
- an API for the up–down scaling of virtual IT resources
- an API for the migration and replication of virtual IT resources across multiple data centers
- interoperation with a single sign-on mechanism through an LDAP interface

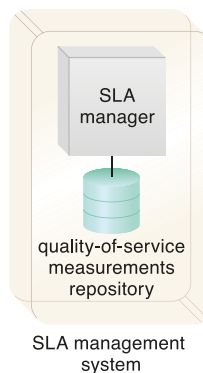
Custom-designed SNMP command scripts are further implemented to interoperate with the network management tools to establish isolated virtual networks across multiple data centers.

12.3 SLA Management System

The *SLA management system* mechanism represents a range of commercially available cloud management products that provide features pertaining to the administration, collection, storage, reporting, and runtime notification of SLA data (Figure 12.7).

Figure 12.7

An SLA management system encompassing an SLA manager and QoS measurements repository.



An SLA management system deployment will generally include a repository used to store and retrieve collected SLA data based on predefined metrics and reporting parameters. It will further rely on one or more SLA monitor mechanisms to collect the SLA data that can then be made available in near-realtime to usage and administration portals to provide ongoing feedback regarding active cloud services (Figure 12.8). The metrics monitored for individual cloud services are aligned with the SLA guarantees in the corresponding cloud provisioning contracts.

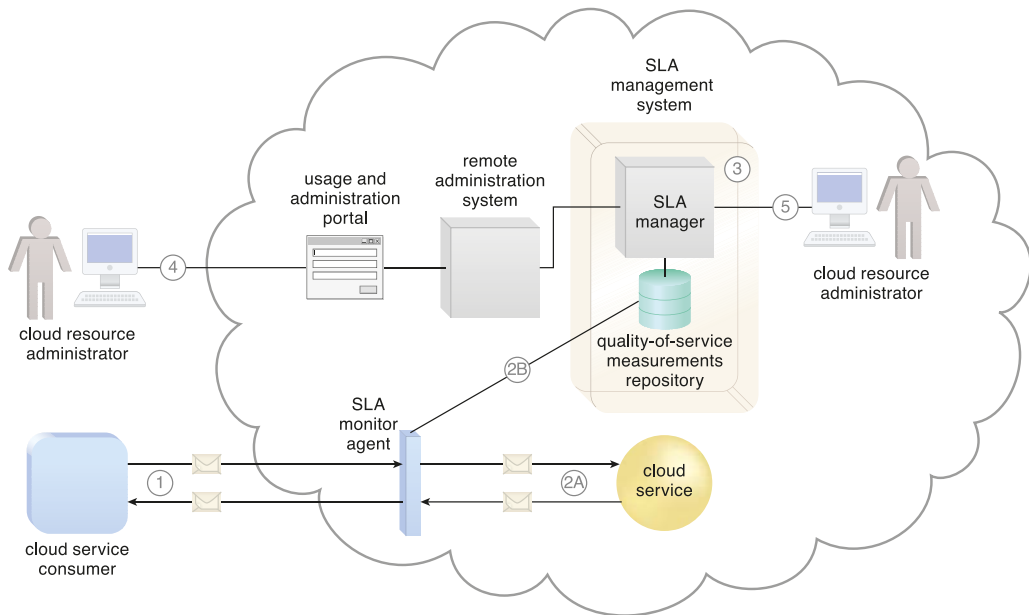


Figure 12.8

A cloud service consumer interacts with a cloud service (1). An SLA monitor intercepts the exchanged messages, evaluates the interaction, and collects relevant runtime data in relation to quality-of-service guarantees defined in the cloud service's SLA (2A). The data collected is stored in a repository (2B) that is part of the SLA management system (3). Queries can be issued and reports can be generated for an external cloud resource administrator via a usage and administration portal (4) or for an internal cloud resource administrator via the SLA management system's native user interface (5).

CASE STUDY EXAMPLE

DTGOV implements an SLA management system that interoperates with its existing VIM. This integration allows DTGOV cloud resource administrators to monitor the availability of a range of hosted IT resources via SLA monitors.

DTGOV works with the SLA management system's report design features to create the following predefined reports that are made available via custom dashboards:

- *Per-Data Center Availability Dashboard* – Publicly accessible through DTGOV's corporate cloud portal, this dashboard shows the overall operational conditions of each group of IT resources at each data center, in realtime.
- *Per-Cloud Consumer Availability Dashboard* – This dashboard displays realtime operational conditions of individual IT resources. Information about each IT resource can only be accessed by the cloud provider and the cloud consumer leasing or owning the IT resource.
- *Per-Cloud Consumer SLA Report* – This report consolidates and summarizes SLA statistics for cloud consumer IT resources, including downtimes and other time-stamped SLA events.

The SLA events generated by the SLA monitors represent the status and performance of physical and virtual IT resources that are controlled by the virtualization platform. The SLA management system interoperates with the network management tools through a custom-designed SNMP software agent that receives the SLA event notifications.

The SLA management system also interacts with the VIM through its proprietary API to associate each network SLA event with the affected virtual IT resource. The system includes a proprietary database used to store SLA events (such as virtual server and network downtimes).

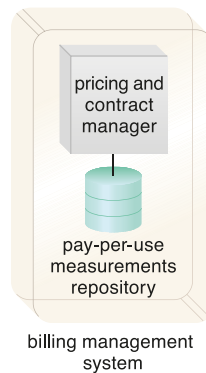
The SLA management system exposes a REST API that DTGOV uses to interface with its central remote administration system. The proprietary API has a component service implementation that can be used for batch processing with the billing management system. DTGOV utilizes this to periodically provide downtime data that translates into credit applied to cloud consumer usage fees.

12.4 Billing Management System

The *billing management system* mechanism is dedicated to the collection and processing of usage data as it pertains to cloud provider accounting and cloud consumer billing. Specifically, the billing management system relies on pay-per-use monitors to gather runtime usage data that is stored in a repository that the system components then draw from for billing, reporting, and invoicing purposes (Figures 12.9 and 12.10).

Figure 12.9

A billing management system comprised of a pricing and contract manager and a pay-per-use measurements repository.



The billing management system allows for the definition of different pricing policies, as well as custom pricing models on a per-cloud-consumer and/or per-IT-resource basis. Pricing models can vary from the traditional pay-per-use models, to flat-rate or pay-per-allocation models, or combinations thereof.

Billing arrangements can be based on pre-usage and post-usage payments. The latter type can include predefined limits or it can be set up (with the mutual agreement of the cloud consumer) to allow for unlimited usage (and, consequently, no limit on subsequent billing). When limits are established, they are usually in the form of usage quotas. When quotas are exceeded, the billing management system can block further usage requests by cloud consumers.

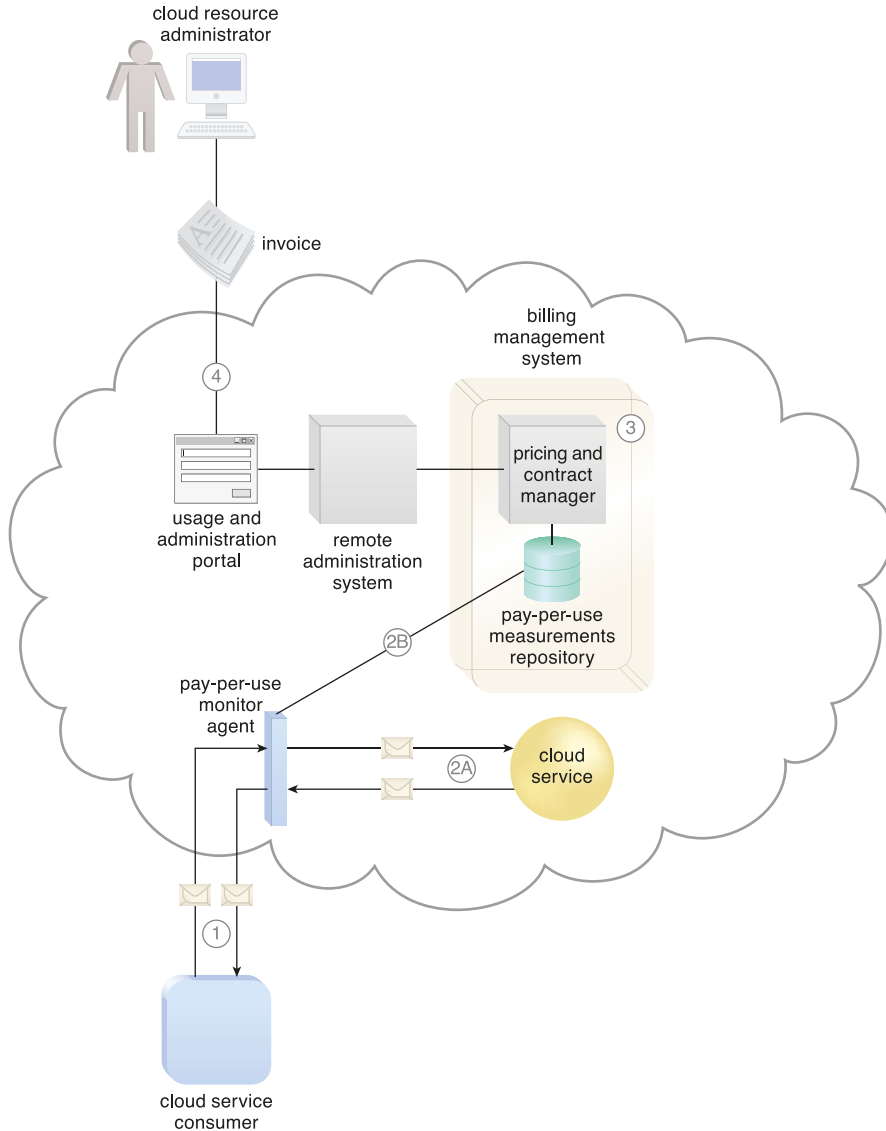


Figure 12.10

A cloud service consumer exchanges messages with a cloud service (1). A pay-per-use monitor keeps track of the usage and collects data relevant to billing (2A), which is forwarded to a repository that is part of the billing management system (2B). The system periodically calculates the consolidated cloud service usage fees and generates an invoice for the cloud consumer (3). The invoice may be provided to the cloud consumer through the usage and administration portal (4).

CASE STUDY EXAMPLE

DTGOV decides to establish a billing management system that enables them to create invoices for custom-defined billable events, such as subscriptions and IT resource volume usage. The billing management system is customized with the necessary events and pricing scheme metadata.

It includes the following two corresponding proprietary databases:

- billable event repository
- pricing scheme repository

Usage events are collected from pay-per-use monitors that are implemented as extensions to the VIM platform. Thin-granularity usage events, such as virtual server starting, stopping, up-down scaling, and decommissioning, are stored in a repository managed by the VIM platform.

The pay-per-use monitors further regularly supply the billing management system with the appropriate billable events. A standard pricing model is applied to most cloud consumer contracts, although it can be customized when special terms are negotiated.

This page intentionally left blank

Index

A

abstraction of container images, 147–148

access-oriented security mechanisms.

See mechanisms

accidental insider, 177

active-active failover system (specialized mechanism), 249–251

active-passive failover system (specialized mechanism), 252–254

activity log monitor mechanism (security), 322

adapter

component, 154

container, 154–155

advanced persistent threat (APT), 185–187

Advanced Research Projects Agency Network (ARPANET), 24

Advanced Telecom Networks (ATN) case study. *See case study examples*

adware, 176

agent

monitoring, 214

polling, 215–216

resource, 215

service, 110

malicious, 167

threat, 165–167

ambassador

component, 156

container, 155–156

anonymous attacker, 166

application

configuration baseline, 403

layer protocol, 84

multitenant, 105–107

package, 403

packager, 403

subscription duration metric, 488–489

usage, 470

virtualization, 103

APT (advanced persistent threat), 185–187

APT groups, 187

architectures

cloud balancing, 389–391, 413–414

cloud bursting, 358–359, 368–369

cross-storage device vertical tiering, 424–429

direct I/O access, 417–418

direct LUN access, 419–421

distributed data sovereignty, 393–394

dynamic data normalization, 421–422

dynamic failure detection and recovery, 399–402

dynamic scalability, 350–353

edge computing, 449–450

elastic disk provisioning, 359–362

elastic network capacity, 423–424

elastic resource capacity, 353–355

federated cloud application, 454–455

fog computing, 450–451

hypervisor clustering, 373–378

intra-storage device vertical data tiering, 429–431

load balanced virtual server instances, 380–383

load balanced virtual switches, 432–433

metacloud, 453–454

multicloud, 365–367

multipath resource access, 434–436

nondisruptive service relocation, 383–387

persistent virtual network configuration, 436–438

rapid provisioning, 402–405

redundant physical connection for virtual servers, 439–441

redundant storage, 363–365

resilient disaster recovery, 391–393

resource pooling, 346–350

resource reservation, 395–399

service load balancing, 355–358

storage maintenance window, 441–448

storage workload management, 406–411

virtual data abstraction, 452–453

- virtual private cloud, 411–413
 - virtual server clustering, 379–380
 - workload distribution, 344–346
 - zero downtime, 388–389
 - ARPANET (Advanced Research Projects Agency Network), 24**
 - asymmetric distribution, 234**
 - asymmetric encryption (security mechanism), 272–273**
 - ATN (Advanced Telecom Networks) case study. *See* case study examples**
 - attack, 164. *See also* threat**
 - attack surface, 165**
 - attack vector, 165**
 - attackers, 164–165. *See also* threat agent**
 - audit monitor mechanism (specialized), 247–248**
 - in cross-storage device vertical tiering architecture, 428
 - in distributed data sovereignty architecture, 394
 - in dynamic failure detection architecture, 402
 - in resource pooling architecture, 349
 - in resource reservation architecture, 399
 - in storage workload management architecture, 410
 - in zero downtime architecture, 389
 - authentication**
 - authentication log monitor mechanism, 309
 - IAM (identity and access management), 298–301
 - location-based, 298
 - MFA (multi-factor authentication) system, 297–298
 - risk-based, 298
 - weak, 171–172
 - authenticity (characteristic), 162**
 - authorization**
 - IAM (identity and access management), 299
 - insufficient, 171–172
 - automated scaling listener mechanism (specialized), 228–233**
 - in load balanced virtual server instances architecture, 383
 - in storage workload management architecture, 410
 - autonomic computing, 91**
 - availability (characteristic), 161–162**
 - data center, 91
 - IT resource, 43
 - NoSQL storage devices, 97
 - availability rate metric, 505–506**
- B**
- backup and recovery system mechanism (security), 320–322**
 - bandwidth, 87–88**
 - behavioral identifiers, 295**
 - billing management system mechanism (management), 337–339**
 - biometric scanner mechanism (security), 295–296**
 - bot, 176**
 - botnet, 178–180**
 - boundary**
 - logical network perimeter, 59
 - organizational, 57–58
 - trust, 44, 58
 - broadband networks, 80–89**
 - brute force, 182**
 - build files, 149–150**
 - business agility, 27–28, 40**
 - business case, mapping to SLA, 513**
 - business cost metrics, 480–485**
 - business drivers, cloud computing, 26–28**
- C**
- CA (certificate authority), 284**
 - capacity planning, 29**
 - capacity watchdog system, 380–382**
 - capital costs, 481**
 - carrier and external networks, interconnection, 94**

CASB (Cloud Access Security Brokers), 310**case study examples**

- ATN (Advanced Telecom Networks), 12
 - background*, 12–14
 - business cost metrics*, 482–485
 - cloud bursting architecture*, 368–369
 - cloud security*, 191–192
 - conclusion*, 522
 - hashing*, 275–276
 - IAM (identity and access management), 301
 - load balancer*, 235–236
 - network intrusion monitor*, 308
 - ready-made environment*, 225–226
 - SSO (single sign-on), 289
 - state management database*, 266–267
 - traffic monitor*, 323

DTGOV, 12

- authentication log monitor*, 309
- automated scaling listener*, 230–233
- background*, 15–18
- billing management system*, 339
- cloud-based security groups*, 282–283
- cloud delivery model*, 476–477
- cloud storage device*, 211–213
- cloud usage monitor*, 216–219
- conclusion*, 522–524
- data backup and recovery system*, 322
- data loss prevention (DLP) system*, 318
- data loss protection monitor*, 324
- digital signature*, 278–279
- digital virus scanning and decryption system*, 315
- failover system*, 254–258
- firewall*, 293
- hardened virtual server images*, 291
- hypervisor*, 206–207
- IDS (intrusion detection system), 302
- logical network perimeter*, 198–199
- pay-per-use monitor*, 245–246
- penetration testing tool*, 304
- PKI (public key infrastructure), 286
- pricing models*, 496–501

- remote administration system*, 331
- resource cluster*, 262–263
- resource management system*, 333–334
- resource replication*, 221–223
- service technologies*, 111–114
- SLA management system*, 336
- SLA monitor*, 238–242
- SLA template*, 516–518
- third-party software update utility*, 308
- threat mitigation*, 187–188
- UBA (user behavior analytics) system, 305
- virtual private network (VPN)*, 294
- virtual server*, 201–204
- VPN monitor*, 309–310

Innovartus Technologies Inc., 12

- activity log monitor*, 322
- audit monitor*, 247–248
- background*, 18–19
- biometric scanner*, 296
- cloud balancing architecture*, 413–414
- conclusion*, 524–525
- containers and containerization*, 158
- encryption*, 273–274
- malicious code analysis system*, 316
- multi-device broker*, 265
- multi-factor authentication (MFA) system*, 298
- service quality metrics*, 512–513
- TPM (trusted platform module)*, 320

CCP (Cloud Certified Professional) program, 9**cellular networks, 88–89****certificate authority (CA), 284****characteristics. See cloud characteristics****CIEM (Cloud Infrastructure Entitlement Management), 310****cipher, 271****ciphertext, 271****client (Docker), 529–530****Cloud Access Security Brokers (CASB), 310****cloud architectures. See architectures****cloud auditor (role), 57**

- cloud balancing architecture, 389–391**
 - Innovartus case study, 413–414
 - SaaS environments, 470
- cloud-based IT resource, 34**
 - versus on-premise IT resource, 84–89
 - versus on-premise IT resource in private clouds, 76
 - usage cost metrics, 485–489
- cloud-based security group mechanism (security), 280–283**
- cloud broker (role), 53–54**
- cloud bursting architecture, 358–359**
 - ATN case study, 368–369
- cloud carrier (role), 57**
 - selection, 89
- Cloud Certified Professional (CCP) program, 9**
- cloud characteristics, 59–62**
 - elasticity, 60
 - measured usage, 62
 - multitenancy, 60–61
 - on-demand usage, 59
 - resiliency, 62–63
 - resource pooling, 60–61
 - ubiquitous access, 60
- cloud computing, 2, 25–26**
 - business drivers, 26–28
 - containerization and, 117
 - goals and benefits, 39–43
 - history, 24–25
 - risks and challenges, 44–49
 - technology innovations, 28–32
 - terminology, 32–39
- cloud consumer (role), 35, 39, 52–53**
 - compliance and legal issues, 49
 - governance control, 46–47
 - perspective in cloud delivery models, 471–475
 - shared security responsibility model, 44–45
- cloud-controller-manager, 540**
- cloud delivery models, 62–73**
 - cloud consumer perspective, 471–475
 - cloud provider perspective, 460–470
 - combining, 68–71
 - comparing, 67–68
 - IaaS (Infrastructure-as-a-Service), 64
 - PaaS (Platform-as-a-Service), 64–66
 - SaaS (Software-as-a-Service), 66–67
 - submodels, 72–73
- cloud deployment models, 74–78**
 - hybrid, 77–78
 - multicloud, 77
 - private, 74–76
 - public, 74
- Cloud Infrastructure Entitlement Management (CIEM), 310**
- cloud mechanisms. *See* mechanisms**
- cloud-native delivery submodel, 73**
- cloud provider (role), 35, 52**
 - compliance and legal issues, 49
 - governance control, 46–47
 - perspective in cloud delivery models, 460–470
 - portability, 48
 - selection, 89
 - shared security responsibility model, 44–45
- cloud resource administrator (role), 55–57**
- Cloud Security Posture Management (CSPM), 310**
- cloud service, 37–39**
 - lifecycle phases, 489–490
- cloud service consumer (role), 39**
- cloud service owner (role), 54–55**
- cloud service usage cost metrics, 488–489**
- cloud storage device mechanism (infrastructure), 207–213**
 - in distributed data sovereignty architecture, 394
 - in multipath resource access architecture, 435
 - in resilient disaster recovery architecture, 393
 - in storage maintenance window architecture, 441–448
 - usage cost metrics, 488
 - in virtual private cloud architecture, 413
- cloud storage gateway, 264**

- cloud usage monitor mechanism (infrastructure), 214–219**
 - in cross-storage device vertical tiering architecture, 429
 - in direct I/O access architecture, 418
 - in direct LUN access architecture, 421
 - in dynamic scaling architecture, 353
 - in elastic disk provisioning architecture, 362
 - in elastic network capacity architecture, 423
 - in elastic resource capacity architecture, 354
 - in load balanced virtual switches architecture, 433
 - in nondisruptive service relocation architecture, 387
 - in resource pooling architecture, 350
 - in resource reservation architecture, 399
 - in service load balancing architecture, 358
 - in storage workload management architecture, 411
 - in workload distribution architecture, 345
 - in zero downtime architecture, 389
- Cloud Workflow Protection Platforms (CWPP), 310**
- cluster**
 - container, 133
 - database, 259
 - HA (high availability), 261
 - host, 124
 - Kubernetes, 538–539
 - large dataset, 259
 - load balanced, 261
 - resource, 259–263
 - server, 259
- clustering, 28**
 - NoSQL, 96–98
- Communication as a Service, 73**
- completion time metric, 509**
- compliance and legal issues, 49**
- computational grid, 28**
- computing hardware, 92–93**
- confidentiality (characteristic), 160, 272**
- configuration management, 137**
- connectionless packet switching (datagram networks), 82–83**
- container, 121**
 - build file, 149–150
 - clusters, 133
 - deployment, 137
 - deployment file, 134
 - engine, 121–122
 - orchestrator, 136–139
 - package manager, 134, 139
 - runtime (Kubernetes), 537–538
- container image, 121, 145–152**
 - abstraction, 147–148
 - basic, 145
 - build files, 149–150
 - customized, 145, 151–152
 - immutability, 147
 - types and roles, 145–146
- container network, 125, 139–143**
 - addresses, 142–143
 - scope, 140–142
- container runtime interface (CRI), 537**
- containers and containerization, 31, 33, 103, 226**
 - attack, 174–175
 - benefits, 127–128
 - characteristics, 145
 - Docker, 528–534
 - history, 116–117
 - hosting, 129–130
 - instances, 133
 - Kubernetes, 534–541
 - multi-container types, 152–157
 - orchestration, 136–139
 - package management, 133–136, 139
 - on physical servers, 125
 - pod, 130–132
 - rich containers, 144
 - risks and challenges, 128–129
 - terminology, 117–125
 - on virtual servers, 126–127
- content-aware distribution, 234**
- contracts, 189–190**
- control groups (Docker), 532**
- control plane, 122, 539–541**

cost(s)

- archiving, 495
- of capital, 481
- integration, 481
- locked-in, 482
- management of, 489–495
- ongoing, 481
- overruns, 49
- proportional, 40–42, 60
- reduction, 26–27
- sunk, 481
- up-front, 480

CPU pool, 347**credential management, 299****CRI (container runtime interface), 537****cross-storage device vertical tiering**

- architecture, 424–429

crypto jacking, 176**cryptography, 271–274****CSPM (Cloud Security Posture Management), 310****customized container image, 145, 151–152****CWPP (Cloud Workflow Protection**

- Platforms), 310

cyber activists, 165**cyber attack, 164****cyber criminals, 165****cyber threat, 164****cybersecurity threats, 46****D****daemon (Docker), 529****data backup and recovery system mechanism (security), 320–322****data block, 208****data breach, 164****data center, 89–99**

- autonomic computing, 91
- component redundancy, availability, 91
- facilities, 92
- hardware, 92–95
 - computing, 92–93
 - network, 94–95
 - storage, 93–94

IaaS-based IT resources, 461–462**NoSQL clustering, 96–98****persistence, 467****remote operation and management, 91****security awareness, 92****serverless environments, 95–96****standardization and modularity, 90****technical and business considerations, 98–99****virtualization, 89–90****data leak, 164****data loss prevention (DLP) system mechanism (security), 317–318****data loss protection monitor mechanism (security), 323–324****data normalization, 210–211****data-oriented security mechanisms.***See mechanisms***data storage, 210–211, 463**

- non-relational (NoSQL), 210–211

- relational, 210

database

- cluster, 259

- state management, 265–267

- storage interface, 210–211

Database as a Service, 72**datagram networks (connectionless packet switching), 82–83****decryption, digital virus scanning and decryption system, 313****delivery models, 62–73****denial of service (DoS), 169–170****deployment data store, 403****deployment models, 74–78****deployment optimizer, 134–135****design constraints, REST, 108****Desktop as a Service, 73*****The Digital Enterprise* (newsletter), 9****digital signature mechanism (security), 276–279**

- in PKI (public key infrastructure), 284–286

digital virus scanning and decryption system mechanism (security), 312–315**direct I/O access architecture, 417–418**

direct LUN access architecture, 419–421
 disaster recovery, 391–393
 DLP (data loss prevention) system mechanism (security), 317–318
Docker, 528–534
 client, 529–530
 daemon, 529
 objects, 532–533
 orchestration, 533
 registry, 530–531
 server, 528–529
Docker Pull command, 531
Docker Push command, 531
Docker Run command, 531
Docker Swarm, 533–534
 DoS (denial of service), 169–170
 DTGOV case study. *See* case study examples
 dynamic data normalization architecture, 421–422
 dynamic failure detection and recovery architecture, 399–402, 469
 dynamic horizontal scaling, 351
 dynamic IDS (intrusion detection system), 302
 dynamic malicious code analysis, 316
 dynamic relocation, 351
 dynamic scalability architecture, 350–353
 dynamic vertical scaling, 351

E

eavesdropping, traffic, 168
 edge computing architecture, 449–450
 Elastic Compute Cloud (EC2) services, 25
 elastic disk provisioning architecture, 359–362
 elastic network capacity architecture, 423–424
 elastic resource capacity architecture, 353–355, 470
 elasticity (cloud characteristic), 60
 encryption mechanism (security), 271–274
 asymmetric, 272–273
 symmetric, 272
 enterprise service bus (ESB) platform, 110
 etcd service, 539
 event triggers, 464, 467
 exploit (IT security), 163

F

failover system mechanism (specialized), 249–258
 active-active, 249–251
 active-passive, 252–254
 in dynamic failure detection architecture, 402
 in redundant physical connection for virtual servers architecture, 441
 in zero downtime architecture, 388–389
failure conditions, 464, 467
fast data replication mechanisms, 93
federated cloud application architecture, 454–455
firewall mechanism (security), 292–293
flawed implementations (IT security), 188
fog computing architecture, 450–451

G

gateway
 cloud storage, 264
 mobile device, 264
 XML, 264
governance control, 46–47
grid computing, 28–29
guest operating system, 30, 118

H

HA (high availability), 506
 cluster, 124, 261
hard disk arrays, 93
hardened virtual server image mechanism (security), 290–291
hardware
 computing, 92–93
 independence, 99
 network, 94–95
 obsolescence, 98
 storage, 93–94
 virtualization compatibility, 104
hardware-based virtualization, 102–103
hashing mechanism (security), 274–276
health monitoring, 137
heartbeats, 373

high availability. *See* HA (high availability)

history

- cloud computing, 24–25
- containers and containerization, 116–117

horizontal scaling, 36

host (physical server), 30, 35, 122–124, 129–130

host cluster, 124

host network, 125, 140

host operating system, 100

hot-swappable hard disks, 93

HTTPS, 273

hybrid cloud, 77–78

hypervisor clustering architecture, 373–378

hypervisor mechanism, 31, 102–103, 119, 205–207

- in dynamic scaling architecture, 353
- in elastic network capacity architecture, 424
- in hypervisor clustering architecture, 373
- in load balanced virtual switches architecture, 433
- in multipath resource access architecture, 435
- in persistent virtual network configuration architecture, 438
- in redundant physical connection for virtual servers architecture, 441
- in resilient disaster recovery architecture, 392
- in resource pooling architecture, 350
- in resource reservation architecture, 399
- in virtual private cloud architecture, 413
- in workload distribution architecture, 346
- in zero downtime architecture, 389

I

IaaS (Infrastructure-as-a-Service), 64

- cloud consumer perspective of, 471–473
- cloud provider perspective of, 460–464
- in combination with PaaS, 68–70
- in combination with PaaS and SaaS, 71
- in comparison with SaaS and PaaS, 67–68
- pricing models, 492
- submodels, 72–73

identifiers

- behavioral, 295
- physiological, 295

identity and access management (IAM)

mechanism (security), 298–301

IDS (intrusion detection system) mechanism (security), 301–302

images, 118

- container, 121, 145–152
- Docker, 532

immutability of container images, 147

inbound network usage cost metric, 485–486

Innovartus Technologies Inc. case study. *See* case study examples

insider threat, 177

instance starting time metric, 509

instances of containers, 133

insufficient authorization, 171–172

Integration as a Service, 73

integration costs, 481

integrity (IT security), 161

intelligent automation engine, 353

Internet

- architecture, 80–89
- versus cloud, 32–33
- service provider (ISP), 80–82

internetworks (Internet), 80

intra-cloud WAN usage metric, 486

intra-storage device vertical data tiering architecture, 429–431

intruders, 165

intrusion detection system (IDS) mechanism (security), 301–302

I/O

- caching, 93
- data transferred metric, 488

ISP (Internet service provider), 80–82

IT resource, 33–35

- cloud-based versus on-premise, 84–89
- costs, 480–485
- provisioning considerations
 - of IaaS environments, 472–473
 - of PaaS environments, 474–475
- virtualization, 99–105

K

kernel, 147–148

kube-apiserver, 539

kube-controller-manager, 540

kube-proxy, 536

kube-scheduler, 539

kubelet, 536

Kubernetes, 534–541

cluster, 538–539

container runtime, 537–538

control plane, 539–541

kube-proxy, 536

kubelet, 536

node, 534–535

pod, 535

Kubernetes API, 539

L

lag strategy (capacity planning), 29

LAN fabric, 95

large dataset cluster, 259

latency, 87–88

layers (container images), 149–150

lead strategy (capacity planning), 29

legal issues, 49

live storage migration, 441

live VM migration, 374

load balanced cluster, 124, 261

load balanced virtual server instances
architecture, 380–383

load balanced virtual switches architecture,
432–433

load balancer mechanism (specialized),
234–236

in load balanced virtual server instances
architecture, 383

in load balanced virtual switches
architecture, 433

in service load balancing architecture,
355–357

in storage workload management
architecture, 411

in workload distribution architecture, 345

load balancing, 137

location-based authentication, 298

locked-in costs, 482

logical network perimeter mechanism
(infrastructure), 59, 196–197

in direct I/O access architecture, 418

in elastic network capacity architecture, 424

in hypervisor clustering architecture, 374

in load balanced virtual server instances
architecture, 383

in load balanced virtual switches
architecture, 433

in multipath resource access
architecture, 435

in persistent virtual network configuration
architecture, 438

in redundant physical connection for virtual
servers architecture, 441

in resource pooling architecture, 350

in resource reservation architecture, 399

in storage workload management
architecture, 411

in virtual server clustering architecture, 380

in workload distribution architecture, 346

in zero downtime architecture, 389

logical pod container, 122

LUN (logical unit number), 363

in direct LUN access architecture, 419–421
migration, 406

M

malicious code analysis system mechanism
(security), 315–316

malicious insider, 167, 177

malicious intermediary threat, 168–169

malicious service agent, 167

malicious software, 175–177

malicious tenant, 167

malicious users, 165

malware, 175–177

management plane, 122

match strategy (capacity planning), 29

mean time between failures (MTBF)
metric, 507

- mean time system recovery (MTSR) metric, 512**
 - mean time to switchover (MTSO) metric, 511**
 - measured usage (cloud characteristic), 62**
 - mechanisms**
 - access-oriented security, 270–310
 - authentication log monitor, 309*
 - biometric scanner, 295–296*
 - cloud-based security groups, 280–283*
 - digital signature, 276–279*
 - encryption, 271–274*
 - firewall, 292–293*
 - hardened virtual server images, 290–291*
 - hashing, 274–276*
 - identity and access management (IAM), 298–301*
 - intrusion detection system (IDS), 301–302*
 - multi-factor authentication (MFA) system, 297–298*
 - network intrusion monitor, 308*
 - penetration testing tool, 302–304*
 - public key infrastructure (PKI), 284–286*
 - single sign-on (SSO), 287–289*
 - third-party software update utility, 306–308*
 - user behavior analytics (UBA) system, 304–305*
 - virtual private network (VPN), 293–294*
 - VPN monitor, 309–310*
 - data-oriented security, 312–324
 - activity log monitor, 322*
 - data backup and recovery system, 320–322*
 - data loss prevention (DLP) system, 317–318*
 - data loss protection monitor, 323–324*
 - digital virus scanning and decryption system, 312–315*
 - malicious code analysis system, 315–316*
 - traffic monitor, 323*
 - trusted platform module (TPM), 319–320*
 - infrastructure, 196–226
 - cloud storage device, 207–213*
 - cloud usage monitor, 214–219*
 - hypervisor, 205–207*
 - logical network perimeter, 196–197*
 - ready-made environment, 224–226*
 - resource replication, 220–223*
 - virtual server, 200–204*
 - management, 326–339
 - billing management system, 337–339*
 - remote administration system, 326–331*
 - resource management system, 331–334*
 - SLA management system, 334–336*
 - specialized, 228–267
 - audit monitor, 247–248*
 - automated scaling listener, 228–233*
 - failover system, 249–258*
 - load balancer, 234–236*
 - multi-device broker, 263–265*
 - pay-per-use monitor, 242–246*
 - resource cluster, 259–263*
 - SLA monitor, 236–242*
 - state management database, 265–267*
- message digest, 274**
- metacloud architecture, 453–454**
- metrics**
 - application subscription duration, 488–489
 - availability rate, 505–506
 - business cost, 480–485
 - completion time, 509
 - on-demand storage space allocation, 488
 - on-demand virtual machine instance allocation, 487
 - inbound network usage cost, 485–486
 - instance starting time, 509
 - intra-cloud WAN usage, 486
 - I/O data transferred, 488
 - mean time between failures (MTBF), 507
 - mean time system recovery (MTSR), 512
 - mean time to switchover (MTSO), 511
 - network capacity, 508
 - network usage cost, 485–486
 - number of nominated users, 489
 - number of transactions users, 489
 - outage duration, 506
 - outbound network usage, 486

- reserved virtual machine instance
 - allocation, 487
 - response time, 509
 - server capacity, 508
 - service performance, 507–509
 - service quality, 504–513
 - service reliability, 507
 - service resiliency, 511–512
 - service scalability, 509–510
 - storage device capacity, 508
 - usage cost, 485–489
 - Web application capacity, 508–509
 - MFA (multi-factor authentication) system mechanism (security), 297–298**
 - middleware, service, 110**
 - middleware platforms, 110**
 - enterprise service bus (ESB), 110
 - orchestration, 110
 - migration**
 - live storage migration, 441
 - live VM, 374
 - LUN, 406
 - virtual server, 383–387
 - mobile device gateway, 264**
 - model**
 - “as-a-service” usage, 42
 - delivery, 62–73, 476–477
 - deployment, 74–78, 471–475
 - pricing, 491–493, 496–501
 - shared security responsibility, 44–45
 - monitor**
 - audit, 247–248
 - cloud usage, 214–219
 - IaaS-based IT resources, 463–464
 - PaaS environments, 467
 - pay-per-use, 242–246
 - SLA, 236–242
 - monitoring agent, 214**
 - MTBF (mean time between failures) metric, 507**
 - MTSO (mean time to switchover) metric, 511**
 - MTSR (mean time system recovery) metric, 512**
 - multicloud, 77**
 - architectures, 365–367
 - cost management, 493–495
 - multi-container types, 152–157**
 - multi-device broker mechanism (specialized), 263–265**
 - multi-factor authentication (MFA) system mechanism (security), 297–298**
 - multimodal biometric scanners, 295**
 - multipath resource access architecture, 434–436**
 - multitenancy, 60–61**
 - and resource pooling, 60–61
 - versus virtualization, 107
 - multitenant application, 105–107**
- N**
- namespaces (Docker), 532**
 - NAS (network-attached storage), 94**
 - NAS gateway, 95**
 - negligent insider, 177**
 - nested virtualization, 124**
 - network**
 - addresses, 142–143
 - broadband, 80–89
 - container, 125, 139–143
 - hardware, 94–95
 - host, 125, 140
 - orchestration, 137
 - overlay, 125, 140
 - pool, 347
 - storage interface, 208–209
 - usage cost, PaaS environments, 467
 - virtualization, 30
 - network-attached storage (NAS), 94**
 - network capacity**
 - in elastic network capacity architecture, 423–424
 - metric, 508
 - network intrusion monitor mechanism (security), 308**
 - NIST Cloud Reference Architecture, 25–26**

node, 122, 124

 Kubernetes, 534–535

nondisruptive service relocation architecture, 383–387

normalization, data, 210–211

NoSQL clustering, 96–98

NoSQL (non-relational) data storage, 210–211

number of nominated users metric, 489

number of transactions users metric, 489

O

object storage interface, 209

objects (Docker), 532–533

on-demand storage space allocation metric, 488

on-demand usage (cloud characteristic), 59

on-demand virtual machine instance allocation metric, 487

ongoing cost, 481

on-premise IT resource, 35

 versus cloud-based IT resource, 480–485
 in private cloud, 76

operating system

 abstraction, 147–148

 baseline, 403

 terminology, 117

 virtualization, 100–102

orchestration

 container, 136–139

 Docker, 533

 platform, 110

organizational boundary, 57–58

outage duration metric, 506

outbound network usage metric, 486

overlapping trust boundaries, 173–174

overlay network, 125, 140

P

PaaS (Platform-as-a-Service), 64–66

 cloud consumer perspective, 473–475

 cloud provider perspective, 464–467

 combination with IaaS, 68–70

 combination with IaaS and SaaS, 71

 comparison with IaaS and SaaS, 67–68

 pricing models, 492

 submodels, 72–73

package, 133–134

 management, 133–136, 139

 repository, 134

passive IDS (intrusion detection system), 302

pay-per-use monitor mechanism (specialized), 242–246

 in cross-storage device vertical tiering architecture, 429

 in direct I/O access architecture, 418

 in direct LUN access architecture, 421

 in dynamic scaling architecture, 353

 in elastic network capacity architecture, 424

 in elastic resource capacity architecture, 354

 in nondisruptive service relocation architecture, 387

 in resource pooling architecture, 350

penetration testing tool mechanism (security), 302–304

performance overhead (virtualization), 104

persistent virtual network configuration architecture, 436–438

phishing, 178

physical host, 30, 35

physical network, 84

physical RAM pool, 347

physical server, 118, 125

physical server pool, 346

physiological identifiers, 295

PKI (public key infrastructure) mechanism (security), 284–286

plaintext, 271

pod, 122, 130–132

 Kubernetes, 535

polling agent, 215–216

pool (resource), 346–349

 CPU, 347

 network, 347

 physical RAM, 347

 physical server, 346

 storage, 346

 virtual server, 346

- portability**
 - cloud provider, 48
 - virtualization solution, 104–105
- portal**
 - self-service, 327
 - usage and administration, 327
- power, virtualization, 30**
- pricing models, 491–493**
 - DTGOV case study, 496–501
- private cloud, 74–76**
- privilege escalation, 181**
- Process as a Service, 73**
- proportional costs, 40–42, 60**
- public cloud, 74**
- public key cryptography, 272**
- public key identification, 284**
- public key infrastructure (PKI) mechanism (security), 284–286**
- Q**
- quality of service (QoS), 504–513. *See also* SLA (service-level agreement)**
- R**
- ransomware, 176**
- rapid provisioning architecture, 402–405**
- ready-made environment mechanism (infrastructure), 224–226, 467**
- reduction, cost, 26–27**
- redundant physical connection for virtual servers architecture, 439–441**
- redundant storage architecture, 363–365**
- registry (Docker), 530–531**
- relational data storage, 210**
- reliability (characteristic)**
 - IaaS-based IT resources, 463
 - IT resource, 43
 - PaaS environments, 465–466
- reliability rate metric, 507**
- remote administration system mechanism (management), 326–331, 350**
- remote code execution, 182–183**
- replicas, 133**
- resiliency (cloud characteristic), 62–63**
- resilient disaster recovery architecture, 391–393**
- resilient watchdog system, 399–402**
- resource agent, 215**
- resource cluster mechanism (specialized), 259–263**
 - in service load balancing architecture, 358
 - in workload distribution architecture, 346
 - in zero downtime architecture, 389
- resource constraints, 395**
- resource management system mechanism (management), 331–334, 350**
- resource pool, 346–349**
- resource pooling architecture, 346–350**
- resource pooling (multitenancy), 60–61**
- resource replication mechanism (infrastructure), 220–223**
 - in direct I/O access architecture, 418
 - in direct LUN access architecture, 421
 - in elastic disk provisioning architecture, 362
 - in elastic network capacity architecture, 424
 - in elastic resource capacity architecture, 354
 - in hypervisor clustering architecture, 374
 - in load balanced virtual server instances architecture, 383
 - in load balanced virtual switches architecture, 433
 - in multipath resource access architecture, 435
 - in nondisruptive service relocation architecture, 387
 - in persistent virtual network configuration architecture, 438
 - in redundant physical connection for virtual servers architecture, 441
 - in resource pooling architecture, 350
 - in resource reservation architecture, 399
 - in service load balancing architecture, 358
 - in storage maintenance window architecture, 448
 - in virtual server clustering architecture, 380
 - in workload distribution architecture, 346
 - in zero downtime architecture, 389

resource replication, virtualization, 100
 resource reservation architecture, 395–399
 resources, website, 8
 response time metric, 509
 responsiveness (characteristic), IT
 resource, 40
 REST design constraints, 108
 REST service, 107–108
 rich containers, 144
 risk (IT security), 163
 risk assessment, 190
 risk-based authentication, 298
 risk control, 190
 risk management, 190–191
 risk treatment, 190
 rogue antivirus, 176
 roles, 52–57
 cloud auditor, 57
 cloud broker, 53–54
 cloud carrier, 57
 cloud consumer, 52–53
 cloud provider, 52
 cloud resource administrator, 55–57
 cloud service owner, 54–55
 router-based interconnectivity, 83–84
 RPC, Web-based, 111
 runtime, 117

S

SaaS (Software-as-a-Service), 66–67
 cloud consumer perspective, 475
 cloud provider perspective, 467–470
 combination with IaaS and PaaS, 71
 comparison with PaaS and IaaS, 67–68
 pricing models, 492
 submodels, 72–73
 SAN (storage area network), 94
 SAN fabric, 95
 SASE (Secure Access Service Edge), 310
 scalability
 cloud-based IT resource, 42–43
 IaaS-based IT resources, 463
 PaaS environments, 465–466
 supported by multitenant applications, 106
 scaling, 36–37
 cluster, 124
 container, 137
 horizontal, 36
 vertical, 36–37
 scheduling, 135
 scope, container network, 140–142
 secret key cryptography, 272
 Secure Access Service Edge (SASE), 310
 secure sockets layer (SSL), 273
 secure VPN (virtual private network), 294
 security
 ATN case study, 191–192
 breach, 164
 controls, 162
 cybersecurity threats, 46
 IaaS-based IT resources, 464
 mechanisms, 163. *See also* mechanisms
 PaaS environments, 467
 SaaS environments, 470
 shared responsibility model, 44–45
 terminology, 160–163
 Security as a Service, 72
 security policy, 163
 disparity, 188–189
 self-service portal, 327
 sequence logger, 403
 sequence manager, 403
 server
 capacity metric, 508
 cluster, 259
 consolidation, 99
 Docker, 528–529
 host, 122
 images, 403
 physical, 118, 125
 scalability (horizontal) metric, 510
 scalability (vertical) metric, 510
 usage, 487
 virtual, 118–119, 126–127
 virtual (physical host), 35
 virtualization, 200–204
 serverless environments, 31–32, 95–96

- service, 107–111**
 - agent, 110
 - discovery, 137
 - Docker, 532
 - middleware, 110
 - REST, 107–108
 - Web, 108–109
 - Web-based, 107
- service agent, 110**
 - malicious, 167
- service availability metrics, 505–506**
- service-level agreement. *See* SLA (service-level agreement)**
- service load balancing architecture, 355–358, 469**
- service performance metrics, 507–509**
- service quality metrics, 504–513**
- service reliability metrics, 507**
- service resiliency metrics, 511–512**
- service scalability metrics, 509–510**
- shared security responsibility model, 44–45**
- sidecar container, 152–153**
- Simple Object Access Protocol (SOAP), 108–109**
- single sign-on (SSO) mechanism (security), 287–289**
- SLA (service-level agreement), 38–39, 504**
 - DTGOV case study, 516–518
 - guidelines, 513–515
- SLA management system mechanism (management), 334–336**
 - in dynamic failure detection architecture, 402
 - in nondisruptive service relocation architecture, 387
- SLA monitor mechanism (specialized), 236–242**
 - in dynamic failure detection architecture, 402
 - in nondisruptive service relocation architecture, 387
- snapshotting, 93, 461**
- SOAP, 108–109**
- SOAP-based Web service, 108–109**
- social engineering, 178**
- software, virtualization (hypervisor), 102–103, 205–207**
- Software-as-a-Service. *See* SaaS (Software-as-a-Service)**
- spyware, 176**
- SQL (Structured Query Language), 184**
- SQL injection, 183–184**
- SSL (secure sockets layer), 273**
- SSO (single sign-on) mechanism (security), 287–289**
- state management database mechanism (specialized), 265–267**
- state-sponsored attackers, 165**
- static malicious code analysis, 316**
- storage**
 - hardware, 93–94
 - live migration, 441
 - pool, 346
 - replication, 364
 - virtualization, 30, 93
- storage area network (SAN), 94**
- Storage as a Service, 72**
- storage device, 207–213**
 - capacity metric, 508
 - levels, 208
 - usage, 488
- storage interface**
 - database, 210–211
 - network, 208–209
 - object, 209
- storage maintenance window architecture, 470**
- storage orchestration, 137**
- storage replication mechanism, in distributed data sovereignty architecture, 394**
- storage service gateway, 363**
- storage workload management architecture, 406–411**
- Structured Query Language (SQL), 184**
- sunk costs, 481**
- symmetric encryption (security mechanism), 272**

T

tenant application functional module, 470
tenant subscription period, 470
Testing as a Service, 73
third-party software update utility mechanism (security), 306–308
threat, 164

- advanced persistent threat (APT), 185–187
- botnet, 178–180
- brute force, 182
- DoS (denial of service), 169–170
- insider, 177
- insufficient authorization, 171–172
- landscape, 164
- malicious intermediary, 168–169
- malware, 175–177
- overlapping trust boundaries, 173–174
- phishing, 178
- privilege escalation, 181
- remote code execution, 182–183
- social engineering, 178
- SQL injection, 183–184
- terminology, 163–165
- traffic eavesdropping, 168
- tunneling, 184–185
- virtualization attack, 172–173

threat agent, 165–167

- anonymous attacker, 166
- malicious insider, 167
- malicious service, 167
- trusted attacker, 167

TLS (transport layer security), 273
TPM (trusted platform module) mechanism (security), 319–320
traffic eavesdropping, 168
traffic monitor mechanism (security), 323
transport layer protocol, 84
transport layer security (TLS), 273
Trojan, 176
trust boundary, 58

- overlapping, 44, 173–174

trusted attacker, 167

trusted platform module (TPM) mechanism (security), 319–320
trusted VPN (virtual private network), 294
tunneling, 184–185

U

UBA (user behavior analytics) system mechanism (security), 304–305
ubiquitous access (cloud characteristic), 60
UDDI (Universal Description, Discovery, and Integration), 109
union file system, 149, 532–533
up-front costs, 480
usage and administration portal, 327
usage cost metrics, 485–489

- cloud service, 488–489
- cloud storage device, 488
- inbound network, 485–486
- network, 485–486
- server, 487

user behavior analytics (UBA) system mechanism (security), 304–305
user management, 299
utility computing, 24

V

vertical scaling, 36–37
VIM (virtual infrastructure manager), 104, 331
virtual data abstraction architecture, 452–453
virtual firewall, 197
virtual infrastructure manager (VIM), 104, 331
virtual machine manager (VMM), 31
virtual machine monitor (VMM), 31
virtual network, 197
virtual private cloud architecture, 411–413
virtual private network (VPN) mechanism (security), 293–294
virtual private network (VPN) monitor mechanism (security), 309–310
virtual server, 118–119

- containerization on, 126–127

virtual server clustering architecture, 379–380

virtual server mechanism (infrastructure), 200–204

in elastic network capacity architecture, 424

images, hardened, 290–291

in load balanced virtual server instances architecture, 380–383

in load balanced virtual switches architecture, 433

in multipath resource access architecture, 435

in nondisruptive service relocation architecture, 383–387

in persistent virtual network configuration architecture, 436–438

in redundant physical connection for virtual servers architecture, 439–441

in resilient disaster recovery architecture, 393

in virtual private cloud architecture, 413

in zero downtime architecture, 298–299 lifecycles, 463

virtual server pool, 346

virtual switch

in elastic network capacity architecture, 423

in load balanced virtual switches architecture, 432–433

in persistent virtual network configuration architecture, 436–438

in redundant physical connection for virtual servers architecture, 439–441

in virtual private cloud architecture, 413

virtualization, 30–31, 89–90, 99–105

application-based, 103

attack, 172–173

hardware-based, 102–103

management, 104

versus multitenancy, 107

nested, 124

operating system-based, 100–102

software (hypervisor), 102–103, 205–207

storage, 93

terminology, 118–120

types of, 119–120

viruses, 176, 312–315

VMM (virtual machine manager), 31

volume, 132

volume cloning, 93

VPN (virtual private network) mechanism (security), 293–294

VPN monitor mechanism (security), 309–310

vulnerability (IT security), 163. *See also* threat

W

weak authentication, 171–172

Web application capacity metric, 508–509

Web-based

resource, 472

RPC, 111

service, 107

Web service, 108–109

SOAP-based, 108–109

Web Service Description Language (WSDL), 108

Web-tier load balancing, 94

wireless networks, 88–89

workload distribution architecture, 344–346

workload prioritization, 234

worm, 176

WSDL (Web Service Description Language), 108

X

XML, 108

XML gateway, 264

XML Schema Definition Language, 108

Y

YouTube, Thomas Erl on, 8

Z

zero-day vulnerability, 164

zero downtime architecture, 298–299, 388–389