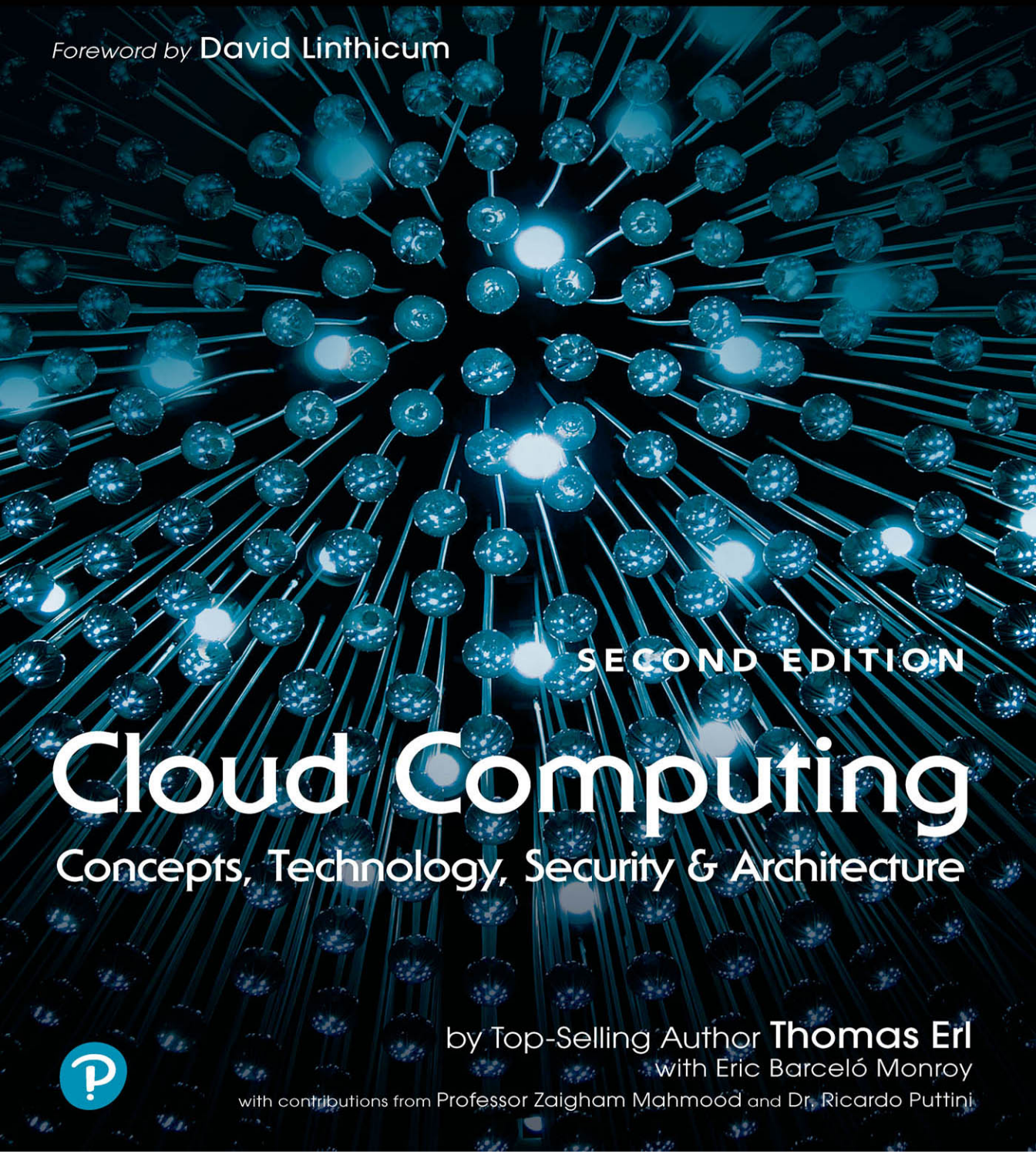
THE PEARSON DIGITAL ENTERPRISE SERIES FROM THOMAS ERL



Foreword by David Linthicum

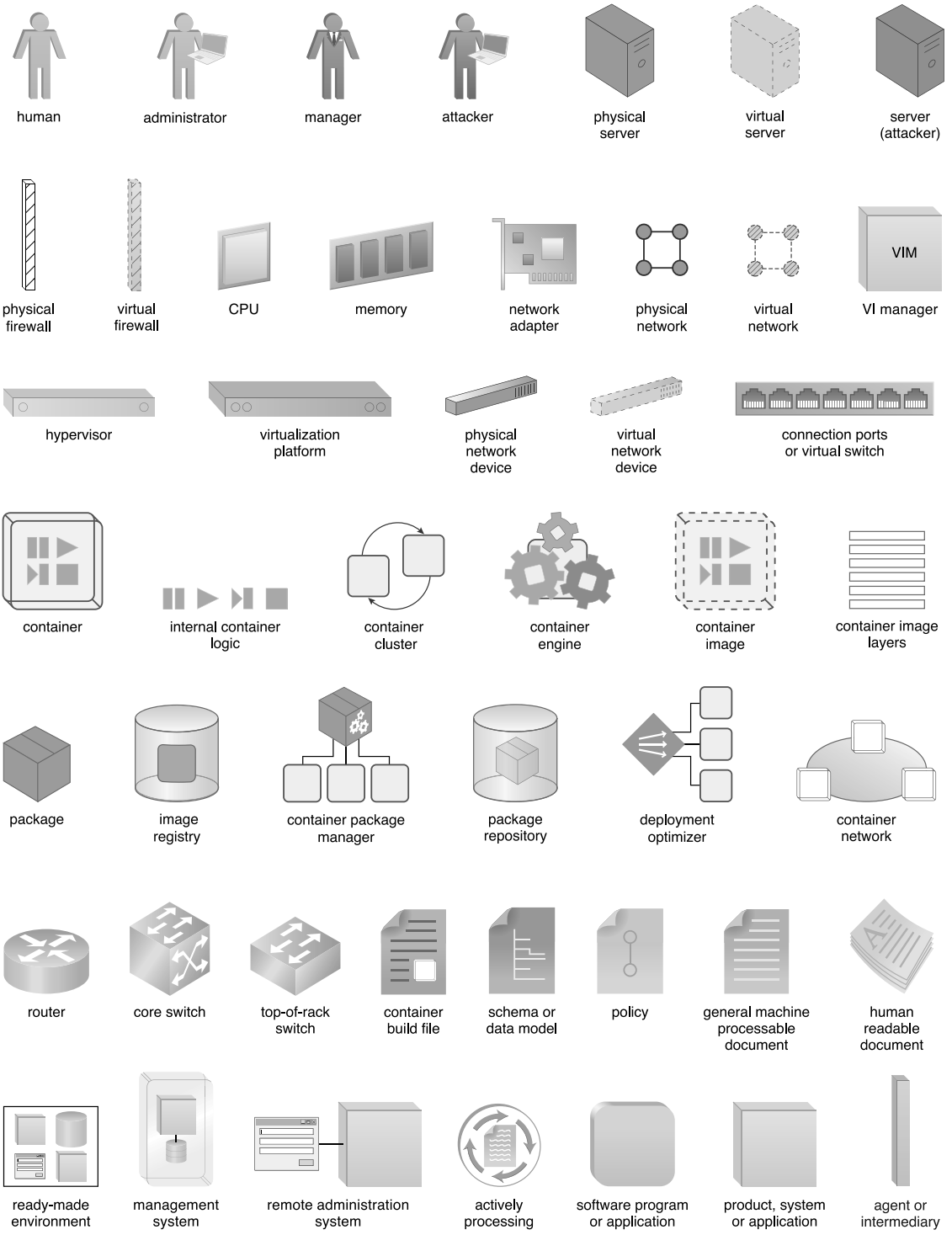
SECOND EDITION

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

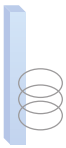

human
administrator
manager
attacker
physical server
virtual server
server (attacker)
physical firewall
virtual firewall
CPU
memory
network adapter
physical network
virtual network
VIM
VI manager
hypervisor
virtualization platform
physical network device
virtual network device
connection ports or virtual 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

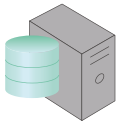

data loss protection monitor

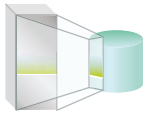

machine learning system


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

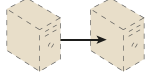

cybersecurity solution

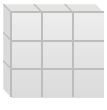
malware

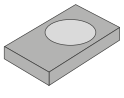

malicious packet


virus


virtual desktops

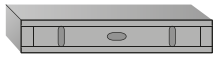

live VM migration


multitenant application

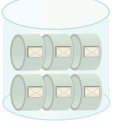

hard disk


hard disk with enclosure


storage device (internal)

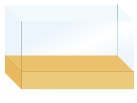

storage controller


databases


message queue

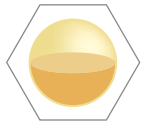

repository or storage device


shared storage


state data in memory


service with state data (stateful service)


repository with state data


grid service

service or proxy

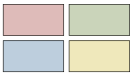
service composition

service layer

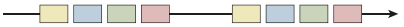
service contract (chorded circle notation)

decoupled service contract

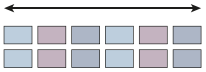
service inventory



LUNs



LUN migration



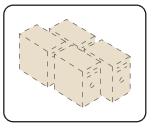
storage replication

live storage migration

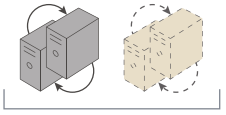
security element or locked resource

message

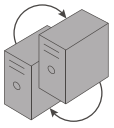
encrypted message



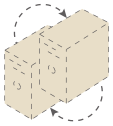
resource pool



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

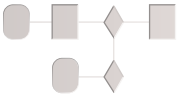
web user interface

folder

workstation

mobile computer

mobile devices

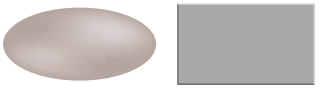


business process/ workflow logic

file system

runtime

transition arrow



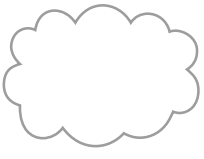
symbols used in conceptual relationship diagrams

organization

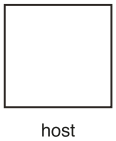
zone or region

internet

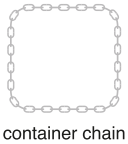
virtual private network



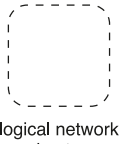
cloud



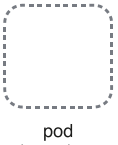
host boundary



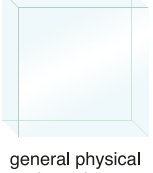
container chain boundary



logical network perimeter or logical boundary



pod boundary



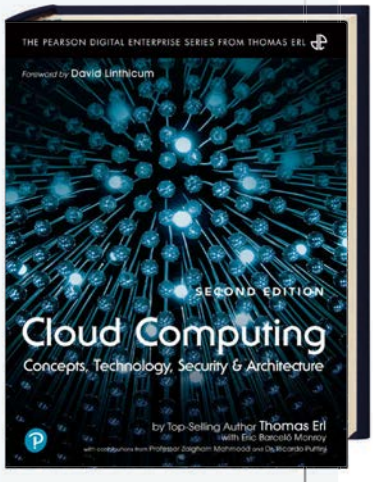
general physical boundary

container boundary

system or program boundary

Arcitura®
the IT education company

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

itcc PCC PROFESSIONAL CERTIFICATION COALITION Pearson | VUE OnVUE Acclaim Credly

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

ThomasErl.com

LinkedIn.com/in/thomaserl

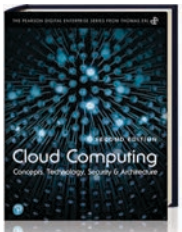
YouTube.com/@terl

YouTube.com/@arcitura

ABOUT THE SERIES EDITOR

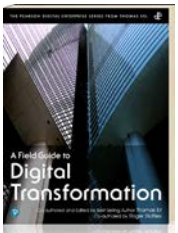
Thomas Erl is a best-selling IT author and the series editor of the Pearson Digital Enterprise Series from Thomas Erl (www.thomaserl.com/books). You can find Thomas on the Thomas Erl YouTube channel (youtube.com/@terl). He is also the host of the *Real Digital Transformation* podcast series (available 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.



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

ISBN: 9780138052256
Paperback

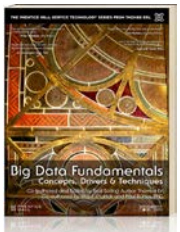


A Field Guide to Digital Transformation
by T. Erl, R. Stoffers

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

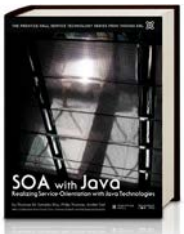
ISBN: 9780134291079
Paperback, 218 pages

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-Orientation
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-Orientation 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. Puttini

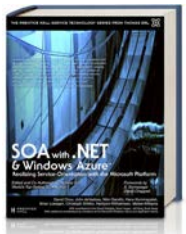
ISBN: 9780133387520
Paperback, 528 pages

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

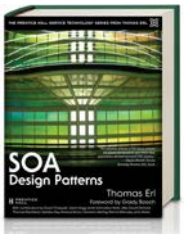
SOA Governance: Governing Shared Services On-Premise & in the Cloud
by S. Bennett, T. Erl, C. Gee, R. Laird, A. Manes, R. Schneider, L. Shuster, A. Tost, C. Venable

ISBN: 9780138156756
Paperback, 675 pages



SOA with .NET & Windows Azure: Realizing Service-Orientation 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

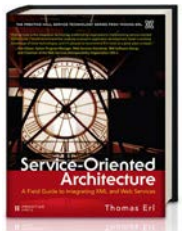
ISBN: 9780136135166
Paperback, 865 pages

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. Yalcinalp

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

Visit us on the Web: informit.com

Library of Congress Control Number: 2023939909

Cover image: Thomas Erl

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

Compositor
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

This page intentionally left blank

Contents

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 underwhelm 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, Deutshe 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
- Filippos 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

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.

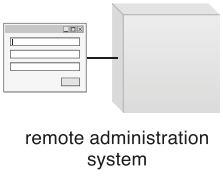


Figure 12.1
The symbol used in this book for the remote administration system. The displayed user interface will typically be labeled to indicate a specific type of portal.

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).

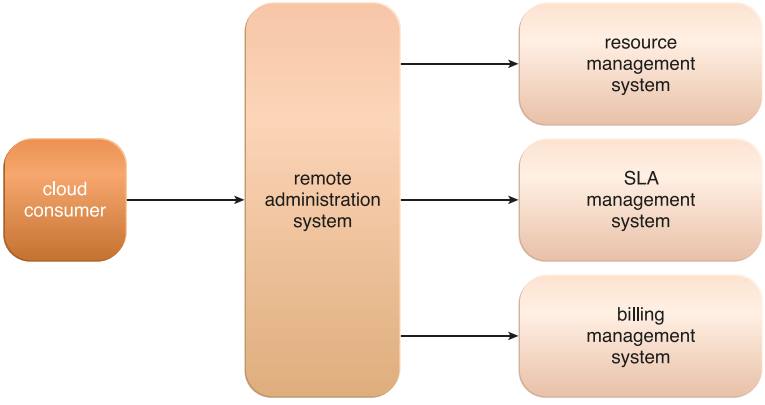


Figure 12.2

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

- *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

- *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.

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

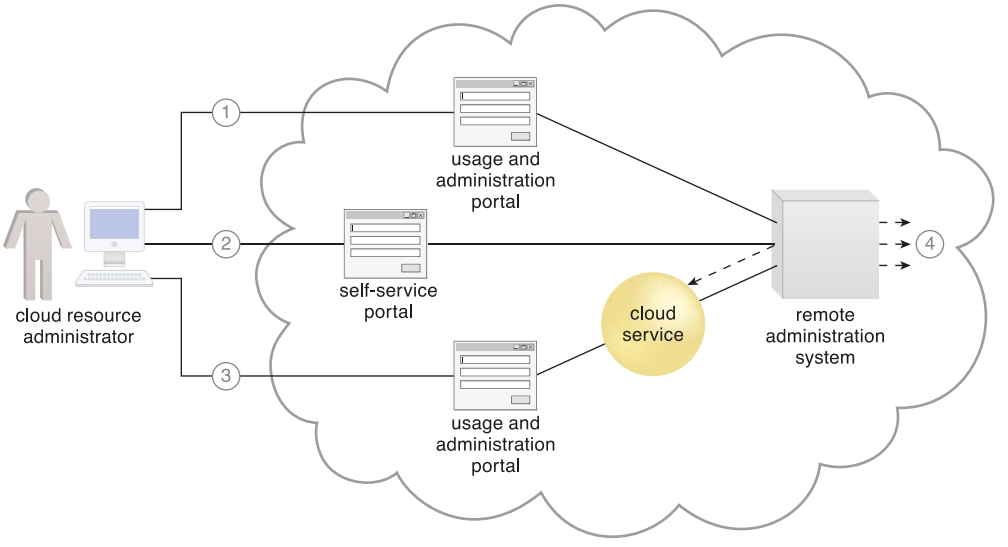


Figure 12.3

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).

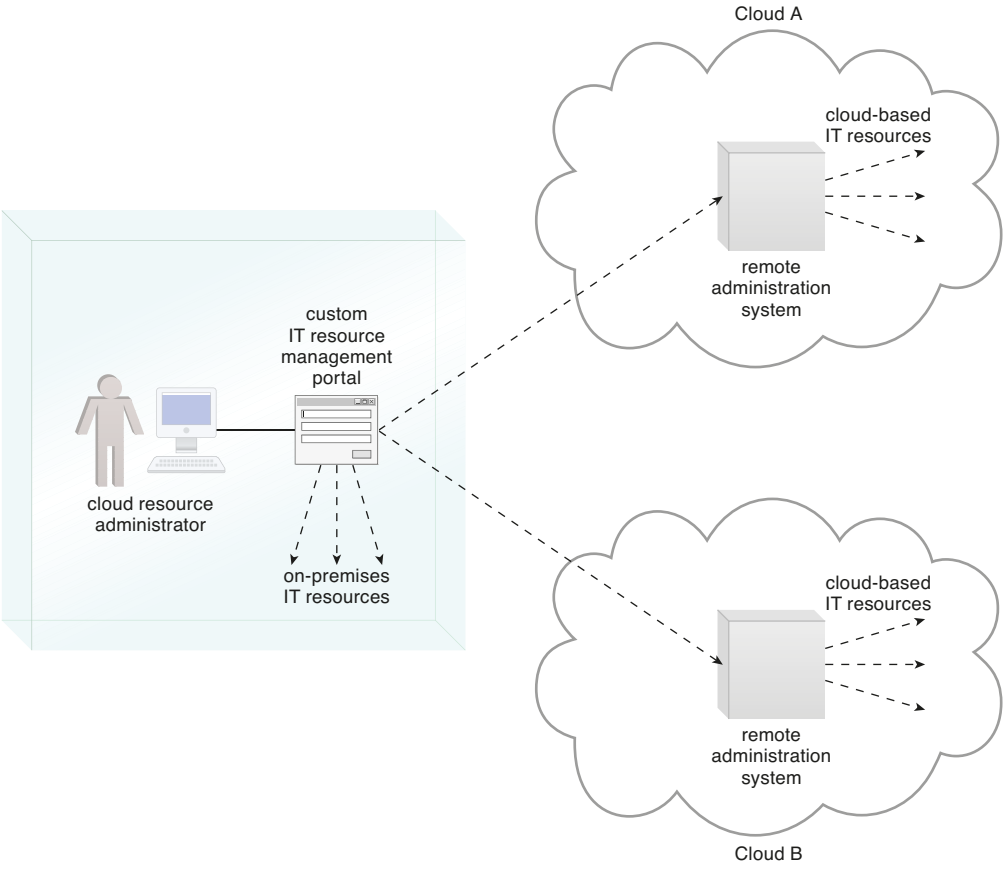


Figure 12.4

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).

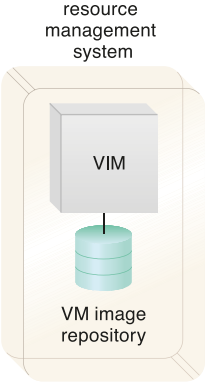


Figure 12.5
A resource management system encompassing a VIM platform and a virtual machine image repository. The VIM may have additional repositories, including one dedicated to storing operational data.

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.

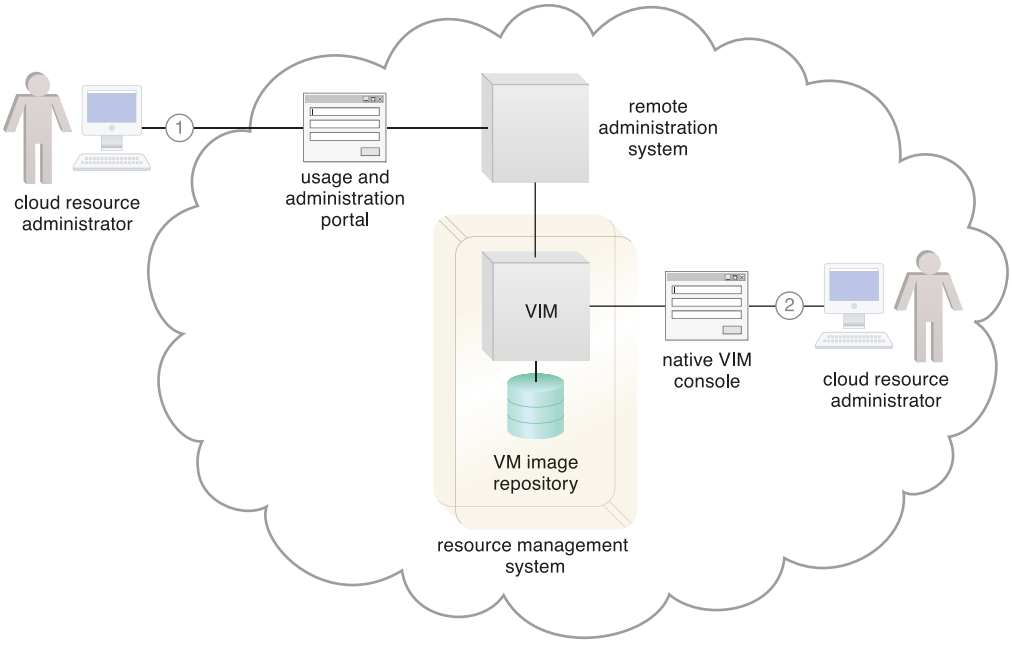


Figure 12.6

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

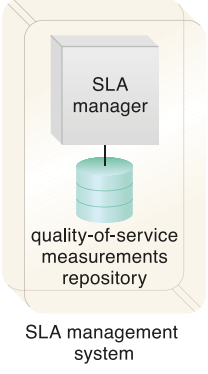
- automated replication ("snapshotting") of virtual server images for virtual server creation
- automated up–down scaling of virtual servers according to usage thresholds to enable live VM migration among physical servers
- an API for the creation and management of virtual servers and virtual storage 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.

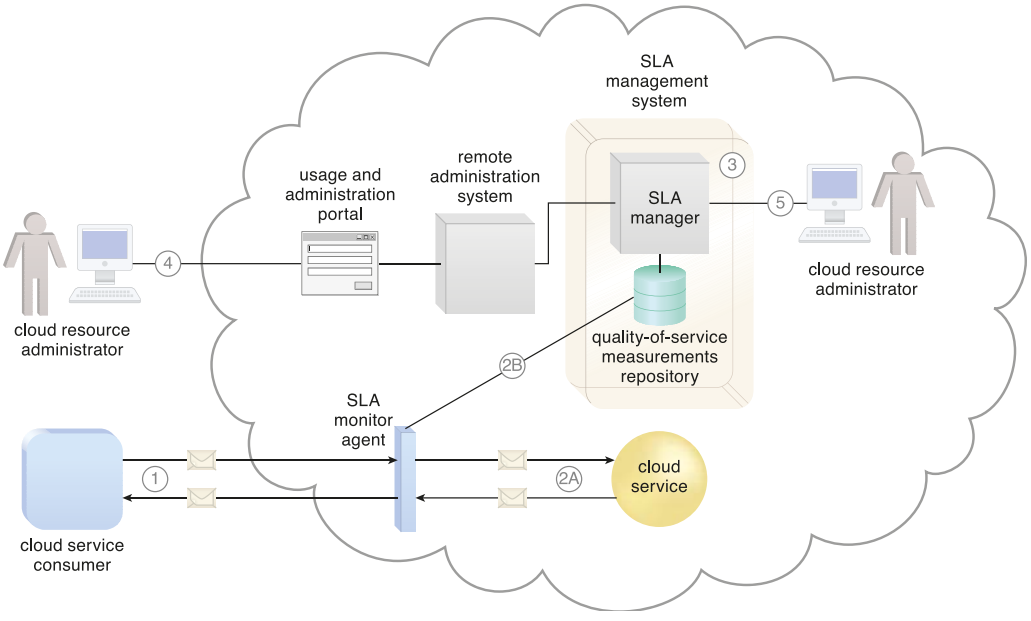


Figure 12.8

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.

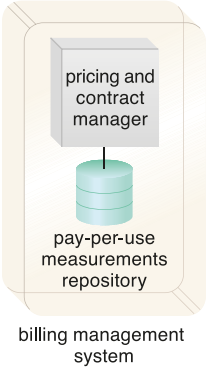
The SLA management system also interacts with the VIM through its proprietary API to associate each network SLA event with the affected virtual IT resource. The system includes a proprietary database used to store SLA events (such as virtual server and network downtimes).

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.

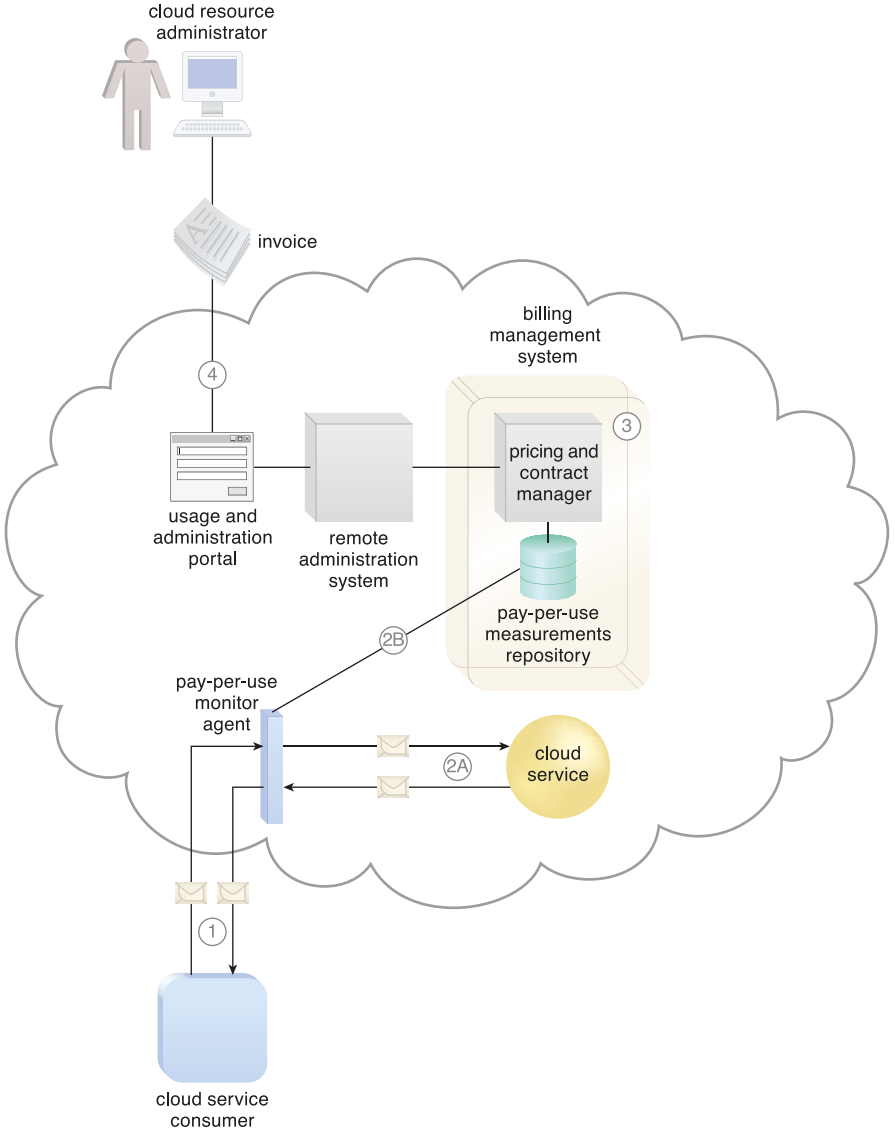


Figure 12.10

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

B

C

D

K

L

M

N

S

T

U

V