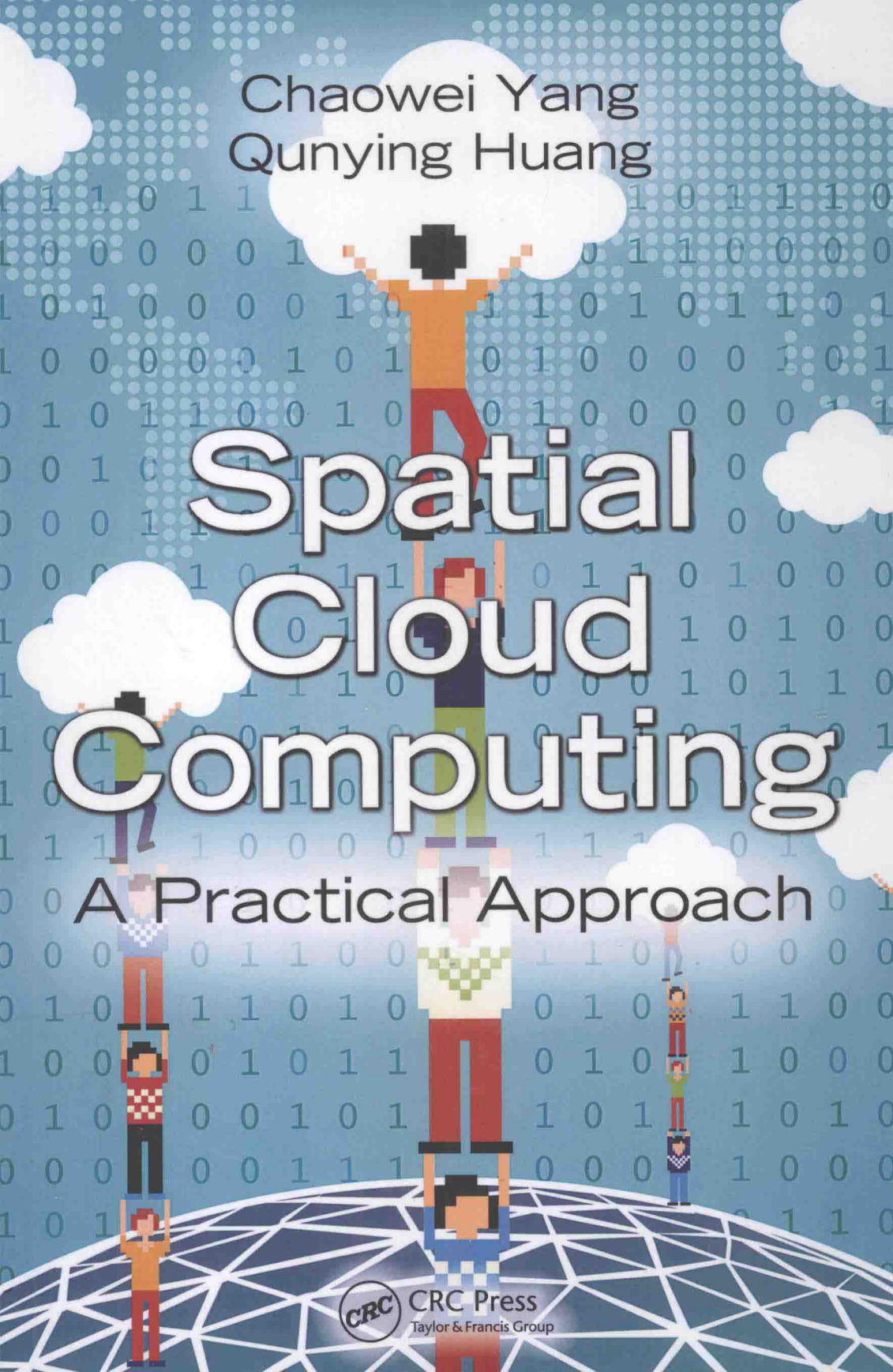
Chaowei Yang
Qunying Huang

Spatial Cloud Computing

A Practical Approach



CRC Press
Taylor & Francis Group

Spatial Cloud Computing

A Practical Approach

Chaowei Yang
Qunying Huang

With the collaboration of

Zhenlong Li
Chen Xu
Kai Liu

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20131227

International Standard Book Number-13: 978-1-4665-9316-9 (Hardback)

Library of Congress Cataloging-in-Publication Data

Yang, Chaowei.
Spatial cloud computing : a practical approach / authors, Chaowei Yang, Qunying Huang.
pages cm
Summary: "A computing cloud is a set of network enabled services, providing scalable, inexpensive computing platforms on demand, which can be accessed in a simple way. This book helps its readers understand the process of how to deploy and customize geospatial applications onto clouds, as well as how to optimize clouds to make them better support geospatial applications. It also discusses and presents the strategies for customizing different types of applications to better utilize the cloud capabilities, such as on-demand services"--Provided by publisher.
Includes bibliographical references and index.
ISBN 978-1-4665-9316-9 (hardback)
1. Geospatial data. 2. Cloud computing. I. Huang, Qunying. II. Title.

G70.217.G46Y36 2014
004.67'82--dc23 2013037852

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

WHY DID WE WRITE THIS BOOK?

There are several motivations that led to the writing of this book. We started utilizing cloud computing for geoscience applications around 2008, when cloud computing was just starting to take shape. During the past several years, many cloud computing related books have been published in the computer science domain. But there is no such book detailing the various aspects of how the geoscience community can leverage cloud computing. Our first motivation with this book was to fill this gap to benefit the geoscience community to cover the various aspects of why and how to adopt cloud computing for geosciences (Parts I and II).

Our second motivation came from our well-cited 2011 spatial cloud computing publication of the *International Journal of Digital Earth*. The paper introduced the general concepts and the benefits that can be brought about by cloud computing to geoscience research and applications. We also received inquiries on how to achieve those benefits and how to use cloud computing in a pedagogical fashion. This book in one aspect responds to the requests with Parts II and III on how to cloud-enable geoscience applications step by step.

We have conducted a series of research and development initiatives for using cloud computing for geoscience applications. The projects, detailed in Parts II, III, and IV, range from migrating a Web portal onto a cloud service to investigating the readiness of cloud computing for geosciences using both commercial cloud services and open-source solutions. We also believed that firsthand experience would be very useful if documented in a systematic fashion for geoscientists and geoscience application developers to evaluate, select, plan, and implement cloud operations for their geoscience applications. This was our third motivation that enlightened us to write this book.

We combined our experience gained during the past six years to write this systematically progressive book for demonstrating how geoscience communities can adopt cloud computing from concepts (Part I),

migrating applications to cloud services (Part II), cloud-enabling geoscience applications (Part III), cloud readiness tests and federal cloud-adoption approaches (Part IV), and the future research direction of cloud computing for geosciences (Part V). We expect this book to provide systematic knowledge for readers who wish to get a sense of spatial cloud computing, adopt cloud computing for their applications, or conduct further research in spatial cloud computing.

HOW DID WE WRITE THE BOOK?

In 2012, CRC Press/Taylor & Francis (Irma Britton) saw the need for a cloud computing book for the geoscience communities and agreed with the author team to materialize such an effort. During the past year, we followed 13 steps to ensure a well-written and structured book for our audience: (1) Drs. Chaowei Yang, Qunying Huang, Chen Xu, and Mr. Zhenlong Li and Mr. Kai Liu worked to define the structure and content of the book with each of the chapter authors, who are the developers and researchers for relevant projects. (2) Each chapter was initially written by the authors with the participation of one or several editors. (3) To ensure that the content of each chapter corresponded to the overall book design, Yang was responsible for the review of each chapter in Parts I, II, and V; Xu was responsible for the review of Part IV; and Li was responsible for the review of Chapters 4, 5, and Part III. (4) Structural and content comments were provided to the authors of each chapter to ensure that the overall organization of the book was integrated. (5) Authors of each chapter revised and reviewed the entire chapter by themselves. (6) An internal review of a chapter by authors of other relevant chapters was conducted to ensure the smooth flow of chapters. (7) Authors of each chapter revised and restructured the book chapter as needed. (8) Each chapter was sent out for review by two to four external reviewers. (9) Authors of each chapter and section (part) editors collaborated to address the external review comments. (10) Yang, Xu, and Li did a final review and proof of the chapters. (11) The chapters and the entire book were finalized with Taylor & Francis editors after being formatted by Nanyin Zhou. (12) Huang, Yang, Li, Xu, and Liu worked together to develop the online content including lecture slides for each chapter and online code, scripts, virtual machine images, videos, and documents for readers to easily repeat the cloud deployment and migration processes described in the book. (13) The online content is published on the Taylor & Francis Web site for the book. The book is written by authors who have firsthand experience to ensure the content is well covered. The content was also checked to ensure its organization as a single volume with the project's team leaders (all editors) and principal investigator (Yang) reviewing and approving all content.

WHAT IS THIS BOOK ABOUT?

This book comprehensively introduces knowledge of spatial cloud computing through practical examples in 17 chapters from 5 aspects including: (a) What are the essential cloud computing concepts and why do geosciences need cloud computing? (b) How can simple geoscience applications be migrated to cloud computing? (c) How can complex geoscience applications be cloud-enabled? (d) How can a cloud service be tested to see if it is ready to support geoscience applications? (e) What are the research issues in need of further investigation?

Part I introduces the geoscience requirements for cloud computing in Chapter 1, summarizes the architecture, characteristics, and concepts of cloud computing in Chapter 2, and discusses the enabling technologies of cloud computing in Chapter 3.

Part II introduces the general procedures and considerations when migrating geoscience applications onto cloud services. Chapter 4 demonstrates how to use cloud services through deploying a simple Web application onto two popular cloud services: Amazon EC2 and Windows Azure. Chapter 5 introduces the common procedures for deploying general geoscience applications onto cloud platforms with needs for server-side scripting, database configuration, and high performance computing. Chapter 6 discusses how to choose cloud services based on general cloud computing measurement criteria and cloud computing cost models.

Part III demonstrates how to deploy different geoscience applications onto cloud services. Chapter 7 explains how users can interact with cloud services using ArcGIS in the Cloud as an example. The other three chapters demonstrate how consumers can cloud-enable three different complex geoscience applications: (1) cloud-enabling databases, spatial index, and spatial Web portal technologies to support GEOSS Clearinghouse, (2) cloud-enabling stand-alone model simulations and model output visualization for Climate@Home, and (3) leveraging elastic cloud resources to support disruptive events (e.g., dust storm) forecasting.

Part IV examines the readiness of cloud computing to support geoscience applications using open-source cloud software solutions and commercial cloud services. Chapter 11 introduces and compares three commercial cloud services: Amazon EC2, Windows Azure, and Nebula. In Chapter 12, the readiness of these three cloud services are tested with the three applications described in Part III. Chapter 13 introduces four major cloud computing open-source solutions including CloudStack, Eucalyptus, Nimbus, and OpenNebula; their performance and readiness are tested and compared in Chapter 14. Chapter 15 presents the background, architecture design, approach, and coordination of GeoCloud, which is a cross-agency initiative to define common operating system and software suites for geoscience applications.

Finally, Part V reviews the future research and developments for cloud computing in Chapters 16 and 17. Chapter 16 introduces data, computation, concurrency, and spatiotemporal intensities of geosciences and how cloud services can be leveraged to solve the challenges. Chapter 17 introduces the research directions from the aspects of technology, vision, and social dimensions.

ONLINE CONTENT OF THE BOOK

To help readers better use this book for different purposes, the following online content is provided at: http://www.crcpress.com/product/isbn/9781466593169.

- *Lecture slides for each chapter*—To serve educational purposes, this book provides slides for instructors to assist them in teaching the content. The slides are closely mapped to the book chapter content.
- *Key questions*—Five to ten questions that lead a reading of the book are available for each book chapter. The answers for those questions can be found through the context of the chapters as a review of the core content.
- *Virtual machine images of the application examples used in this book*—Chapters 4, 5, 7, 8, 9, and 10 include different levels of examples, from a simple Web application to complex geoscience applications, such as GEOSS Clearinghouse (Chapter 8), Climate@Home (Chapter 9), and dust storm forecasting (Chapter 10). The images contain the source code and data for those examples available for Amazon EC2. Therefore, audiences can directly launch cloud virtual machines from those images and test those examples.
- *Manuals for deploying the application examples*—Details of deploying workflow applications onto cloud services are included (Chapters 4, 5, 7, 8, 9, and 10). In addition, Chapters 12 and 14 also include the detailed workflow for testing the cloud services.
- *Scripts for installing and configuring application examples and cloud services.*
- *Videos to show step-by-step deployment of the application examples.*

WHO IS THE AUDIENCE?

To thoroughly understand spatial cloud computing, especially in supporting the computing needs of geoscience applications, we wrote this book based on our last decade's investigation into many projects in collaboration with a variety of agencies and companies to solve the computing problems of geoscience applications. The reading of the book should progress

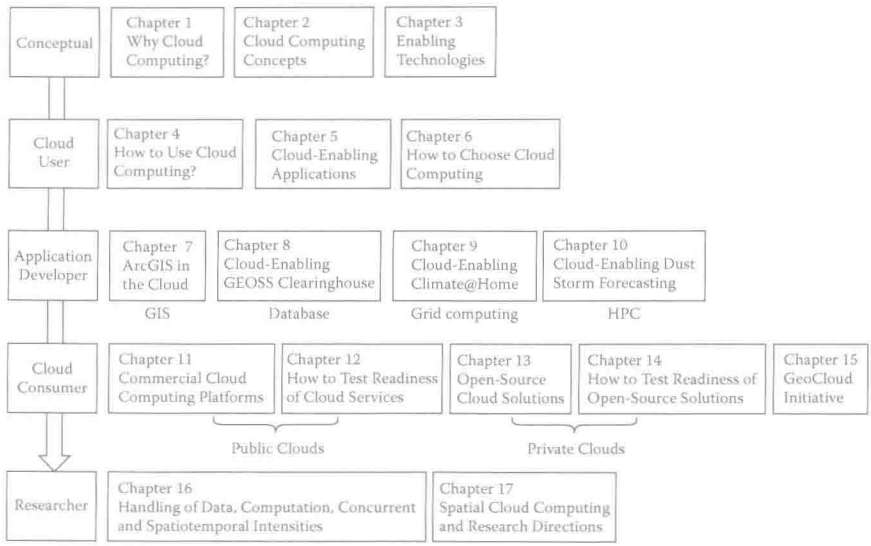


Figure P.1 Reading guide of the book.

in the sequence of the parts and book chapters. But some of them can be omitted based on interest. Figure P.1 depicts the workflow of the chapters for a reader in the knowledge progression sequence.

This book can be used as follows:

(1) As a textbook by professors and students who plan to learn different aspects of cloud computing with the combination of the online slides and examples for class lectures. Each chapter includes lecture slides and is appropriate to serve as independent lecture content. The chapters of Parts II to IV include detailed examples, source code, and data, which could be used for class practice to provide students with hands-on experiences of cloud usage and deployment. These examples can also be used as homework to reinforce what students learned from the lecture. In addition, the examples are carefully selected and considered ranging from simple to complex so that students with different levels of background can follow along. Five to ten questions are provided for each chapter to help students summarize the core content of the respective chapter.

(2) A manual for cloud-enabled application developers with the guidelines is progressively provided in Parts II, III, and IV. This book first provides a general guideline of how to deploy applications onto cloud services (Chapter 4). And then based on the guideline, a common workflow for deploying geoscience applications onto cloud services is introduced (Chapter 5). Based on this common workflow, three practical examples are used to demonstrate (a) how to cloud-enable three different types of

geoscience applications (database, grid computing, and high performance computing [HPC]), and (b) how to handle special requirements of different applications (Chapters 8, 9, and 10). In addition to demonstrating how to use cloud services, this book also provides guidelines on how to choose suitable cloud services (Chapter 6) and how to test cloud services (Chapters 12 and 14).

(3) A reference for geoscientists. The book provides different aspects of cloud computing, from driving requirements (Chapter 1), concepts (Chapter 2), and technologies (Chapter 3) to applications (Chapters 8, 9, and 10), from cloud provider selection (Chapter 6) to testing (Chapters 12 and 14), from commercial cloud services (Chapters 4, 5, 11, and 12) to open-source cloud solutions (Chapters 13 and 14), and from using cloud computing to solve contemporary research and application issues (Chapter 16) to future research topics (Chapter 17). Geoscientists with different research and science domain backgrounds can easily find the cloud computing knowledge that will fit their requirements.

(4) A reference for general IT professionals and decision makers. This book provides references to the concepts, the technical details, and the operational guidelines of cloud computing. The first 15 chapters provide incremental descriptions about different aspects of cloud computing. Chapters 4, 5, and 7 through 15 are closely related to daily IT operations. Decision makers can use Chapters 1 to 3 to build a foundational understanding of cloud computing; then skip to Chapter 6 for considerations related to cloud service selection; and find useful information in Chapters 11, 13, and 15, which cover both commercial and private cloud introductions and evaluations that are most relevant to their decision making.

Acknowledgments

The authors have gained their experience from participating in projects totaling over $10 million in funding from a variety of agencies and companies including the National Aeronautics and Space Administration (NASA) OCIO, NASA MAP Program, NASA Applied Sciences Program, NASA High End Computing Program, National Science Foundation (NSF) EarthCube Program, NSF I/UCRC Program, NSF CNF Program, and NSF Polar Programs, U.S. Geological Survey (USGS) SilvaCarbon Program, Federation of Earth Science Information Partners (FGDC), Microsoft Research, intergovernmental Group on Earth Observation (GEO), ESIP, Association of American Geographers (AAG), Cyberinfrastructure Specialty Group (CISG), ISPRS WG II/IV, the international association of Chinese Professionals in Geographic Information Sciences (CPGIS), and Amazon Inc. The nonauthor project collaborators include Mirko Albani, John Antoun, Jeanette Archetto, Jeff de La Beaujardiere, Don Badrak, Celeste Banaag, David Burkhalter, Robert Cahalan, Songqing Chen, Lloyd Clark, Guido Colangeli, Corey Dakin, Hazem Eldakdoky, David Engelbrecht, John D. Evans, Daniel Fay, Bill Fink, Paul Fukuhara, Andrian Gardner, Pat Gary, Tom Giffen, Steve Glendening, Yonanthan Goitom, Lon Gowen, Sue Harzlet, Mohammed Hassan, Thomas Huang, Haisam Ido, Mahendran Kadannapalli, Ken Keiser, Paul Lang, Wenwen Li, Matthew Linton, Lynda Liptrap, Michael L. Little, Stephen Lowe, Richard Martin, Jeff Martz, Roy Mendelssohn, Dave Merrill, Lizhi Miao, Mark Miller, Nick Mistry, Matthew Morris, Doug Munchoney, Aruna Muppalla, Steve Naus, Slobodan Nickovic, Erik U. Noble, Robert Patt-Corner, Goran Pejanovic, Pete Pollerger, Chris Rusanowski, Todd Sanders, Gavin A. Schmidt, Alan Settell, Khawaja S. Shams, Bryan J. Smith, William Sprigg, Mike Stefanelli, Joe Stevens, Nicola Trocino, Tiffany Vance, Archie Warnock, Mike Whiting, Paul Wiese, Lisa Wolfisch, Huayi Wu, Yan Xu, and Abraham T. Zeyohannis.

The chapters were reviewed by external experts including Michael Peterson at the University of Nebraska at Omaha, Ian Truslove of the National Sea Ice Data Center, Chuanrong Zhang at the University of Connecticut, Stefan

Falke at Northrop Grumman, Xuan Shi at the University of Arkansas, Long Pham at NASA Goddard, Jian Chen at Louisiana University, Jin Xing at McGill University, Marshall Ma at Rensselaer Polytechnic Institute, Thomas Huang at NASA JPL, Chris Badure at Appalachia State University, Doug Nebert at FGDC, Rick Kim at the National University of Singapore, Rui Li at Wuhan University, and Alicia Jeffers at the State University of New York at Geneseo. Nanyin Zhou at George Mason University spent a significant amount of time formatting the book. Peter Lostritto at George Mason University proofed several chapters.

Thanks also go to CRC Press/Taylor & Francis acquiring editor Irma Britton, her assistant and production coordinator Arlene Kopeloff, Joselyn Banks-Kyle who helped to ensure that the manuscript was formatted according to standards adopted by CRC Press/Taylor & Francis. They provided insightful comments and were patient when working with us. Students who worked on the projects and participated in writing the chapters are greatly appreciated.

Finally, we would like to thank our family members for their tolerance and for bearing with us as we stole family time to finish the book.

Chaowei Yang would like to thank his wife Yan Xiang, his children Andrew Yang, Christopher X. Yang, and Hannah Yang.

Qunying Huang would like to thank her husband Yunfeng Jiang.

Zhenlong Li would like to thank his wife Weili Xiu and his son Mason J. Li.

Chen Xu's deepest gratitude goes to his beloved wife, Jianping Han, for her understanding, support, and patience.

Kai Liu would like to thank his wife, Huifen Wang.

Contents