

SYNGRESS

MOVING TO THE CLOUD

Developing Apps in the New World of Cloud Computing

Dinkar Sitaram
Geetha Manjunath



SYNGRESS

MOVING TO THE CLOUD

Developing Apps in the New World of Cloud Computing



Dinkar Sitaram
Geetha Manjunath

Table of Contents

[Cover image](#)

[Front Matter](#)

[Copyright](#)

[Dedication](#)

[About the Authors](#)

[About the Technical Editor](#)

[Contributors](#)

[Foreword](#)

[Preface](#)

[Chapter 1. Introduction](#)

[Chapter 2. Infrastructure as a Service](#)

[Chapter 3. Platform as a Service](#)

[Chapter 4. Software as a Service](#)

[Chapter 5. Paradigms for Developing Cloud Applications](#)

[Chapter 6. Addressing the Cloud Challenges](#)

[Chapter 7. Designing Cloud Security](#)

[Chapter 8. Managing the Cloud](#)

[Chapter 9. Related Technologies](#)

[Chapter 10. Future Trends and Research Directions](#)

[Index](#)

Front Matter

Moving to the Cloud

Moving to the Cloud

Developing Apps in the New World of Cloud Computing

Dinkar Sitaram

Geetha Manjunath

Technical Editor

David R. Deily



ELSEVIER AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW
YORK • OXFORD • PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE •

SYDNEY • TOKYO

SYNGRESS

Syngress is an imprint of Elsevier

Copyright

Acquiring Editor: Chris Katsaropoulos

Development Editor: Heather Scherer

Project Manager: A. B. McGee

Designer: Alisa Andreola

Syngress is an imprint of Elsevier

225 Wyman Street, Waltham, MA 02451, USA

© 2012 Elsevier, Inc. All rights reserved.

Credits for the screenshot images throughout the book are as follows:

Screenshots from Amazon.com, Cloudwatch © Amazon.com, Inc.; Screenshots of Nimsoft © CA Technologies; Screenshots of Gomez © Compuware Corp.; Screenshots from Facebook.com © Facebook, Inc.; Screenshots of Google App Engine, Google Docs © Google, Inc.; Screenshots of HP CloudSystem, Cells-as-a-Service, OpenCirrus © Hewlett-Packard Company; Screenshots of Windows Azure © Microsoft Corporation; Screenshots of Gluster © Red Hat, Inc.; Screenshots from Force.com, Salesforce.com © Salesforce.com, Inc.; Screenshots of Netcharts © Visual Mining, Inc.; Screenshots of Yahoo! Pipes, YQL © Yahoo! Inc.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods or professional practices, may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information or methods described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas

contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Sitaram, Dinkar.

Moving to the cloud: developing apps in the new world of cloud computing / Dinkar Sitaram and Geetha Manjunath; David R. Deily, technical editor.

p. cm.

Includes bibliographical references.

ISBN 978-1-59749-725-1 (pbk.)

1. Cloud computing. 2. Internet programming. 3. Application programs--Development.

I. Manjunath, Geetha. II. Title.

QA76.585.S58 2011

004.6782--dc23

2011042034

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

For information on all Syngress publications visit our website at www.syngress.com

Typeset by: diacriTech, Chennai, India

Printed in the United States of America

11 12 13 14 15 10 9 8 7 6 5 4 3 2 1



Dedication

To Swarna, Tejas, and Tanvi for their encouragement and support.

— Dinkar

To my dear husband Manjunath, wonderful kids Abhiram and Anagha and my loving parents.

— Geetha

About the Authors



Dr. Dinkar Sitaram is a Chief Technologist at Hewlett Packard, Systems Technology and Software Division, in Bangalore, India. He is one of the key individuals responsible for driving file systems and storage strategy, including cloud storage. Dr. Sitaram is also responsible for University Relations, and Innovation activities at HP. His R&D efforts have resulted in over a dozen granted US patents. He is co-author of *Multimedia Servers: Applications, Environments and Design*. Morgan Kaufmann, 2000. Dr. Sitaram received his Ph. D from the University of Wisconsin-Madison and his B. Tech from IIT Kharagpur. He joined as a research staff member in IBM's Research Division at the IBM T. J. Watson Research Center. At IBM, Dr. Sitaram received an IBM Outstanding Innovation Award (an IBM Corporate Award) as well as IBM Research Division Award and several IBM Invention Achievement Awards for his patents and research. He also received outstanding paper awards for his work, and served on the editorial board of the Journal of High-Speed Networking. Subsequently, he returned to India as Director of the Technology Group at Novell Corp. Bangalore. The group developed many innovative products in addition to filing for many patents and standards proposals. Dr. Sitaram received Novell's Employee of the Year award. Before joining HP, Dr. Sitaram was CTO at Andiamo Systems India (a storage networking startup later acquired by Cisco), responsible for architecture and technical direction of an advanced storage management solution.



Geetha Manjunath is a Senior Research Scientist and Master Technologist at Hewlett Packard Research Labs in India. She has been with HP since 1997 working on research issues in multiple systems technologies. During these years, she has developed many innovative solutions and published many papers in the area of Embedded Systems, Java Virtual Machine, Mobility, Grid Computing, Storage Virtualization and Semantic Web. She is currently leading a research project on cloud

services for simplifying web access for emerging markets. As a part of this research, she conceptualized the notion of Tasklets and lead the development of a cloud-based solution called SiteOnMobile that enables consumers to access web tasks on low-end phones via SMS and Voice. The solution was awarded the NASCOM Innovation Award 2009 and has been given a status of “HP Legend”. It was also the winner of Technology Review India's 2010 Grand Challenges for Technologists (2010 TRGC) in the healthcare category.

Before joining HP, she was a senior technical member at Centre for Development of Advanced Computing (C-DAC), Bangalore for 7 years – where was a core member of PARAS system software team for a PARAM supercomputer and she lead a research team to develop parallel compilers for distributed memory machines.

She is a gold medalist from Indian Institute of Science where she did her Masters in Computer Science in 1991 and pursuing Ph. D at the time of this writing. She was awarded the TR Shammanna Best Student award from Bangalore University in the Bachelors degree for topping across all branches of Engineering. She holds four US patents with many more pending grant.

About the Technical Editor

David R. Deily (CISSP, MCSE, SIX SIGMA) has more than 13 years of experience in the management and IT consulting industry. He has designed and implemented innovative approaches to solving complex business problems with the alignment of both performance management and technology for increased IT effectiveness.

He currently provides IT consulting and management services to both midsize and Fortune 500 companies. His core competencies include delivering advanced infrastructure consulting services centered on application/network performance, security, infrastructure roadmap designs, virtualization / cloud, and support solutions that drive efficiency, competitiveness, and business continuity. David consults with clients in industries that include travel/leisure, banking/finance, retail, law and state and local governments.

Mr. Deily has held leadership roles within corporate IT and management consulting services organizations. He is currently a Senior Consultant at DATACORP in Miami, FL. He would like to thank his wife Evora and daughter Drissa for their continued support.

Contributors



Badrinath Ramamurthy is a senior technologist at Hewlett Packard, Bangalore, India. He has been with HP since 2003 and has worked in the areas of High Performance Computing, Semantic Web and Infrastructure Management. He currently works on HP's Cloud Services.

During 1994–2003 he served on the faculty of the CSE Department at the Indian Institute of Technology, Kharagpur. He spent the year 2002–2003 as a visiting researcher at IRISA, France.

Badrinath obtained a Ph.D. in computer science from Rensselaer Polytechnic Institute, NY, in 1994. He has over 30 refereed published research works in his areas of interest. He has served as the General Co-Chair for the International Conference on High-Performance Computing (HiPC) for the years 2006, 2007 and 2008.

In this book, Dr. Badrinath has contributed the section titled “Cells as a Service” in [Chapter 2](#).



Dejan Milojevic is a senior researcher and director of Open Cirrus Cloud Computing testbed at Hewlett Packard Labs. He has worked in the areas of operating systems and distributed systems for more than 25 years. Dr. Milojevic has published over 100 papers. He is an ACM distinguished engineer, IEEE Fellow and member of USENIX. He received B.Sc. and M.Sc. degrees from University of Belgrade and a Ph.D. from University of Kaiserslautern. Prior to HP Labs, he worked at Institute “Mihajlo Pupin”, and at OSF Research Institute.

In this book, Dr. Dejan has contributed the section titled “OpenCirrus” in [Chapter 10](#).



Devaraj Das is a co-founder of Hortonworks Inc, USA. Devaraj is an Apache Hadoop committer and member of the Apache Hadoop Project Management Committee. Prior to co-founding Hortonworks, Devaraj was critical in making Apache Hadoop a success at Yahoo! by designing, implementing, leading and managing large and complex core Apache Hadoop and Hadoop-related projects on Yahoo!'s production clusters. Devaraj also worked as an engineer at HP in Bangalore earlier in his career. He has a Master's degree from the Indian Institute of Science in Bangalore, India, and a B.E. degree from Birla Institute of Technology and Science in Pilani, India. In this book, Devaraj has shared his knowledge on advanced topics in Apache Hadoop, specially in section titled “Multi-tenancy and security” of [Chapter 6](#) and “Data Flow in MapReduce” in [Chapter 3](#).



Dibyendu Das is currently a Principal Member of Technical Staff in AMD India working on Open64 optimizing compilers. In previous avatars he has worked extensively on optimizing compilers for PA-RISC and IA-64 processors while at HP, performance/power analyses for Power-7 multi-cores at IBM and VLIW compilers for Motorola. Dibyendu is an acknowledged expert in the areas of optimizing compilers, parallel languages, parallel and distributed processing and computer architecture. Dibyendu has a Ph.D. in computer science from IIT Kharagpur and an M.E. and B.E. in computer science from IISc and Jadavpur University, respectively. He is an active quizzier and quiz enthusiast and is involved with the Karnataka Quiz Association. In this book, Dr. Dibyendu has contributed the section titled “IBM SmartCloud: pureXML” in [Chapter 3](#).



Gopal R Srinivasa is a Sr. Research SDE with Microsoft Research India. Before joining Microsoft, he worked for Hewlett-Packard, Nokia Siemens Networks, and CyberGuard Corporation. Along with cloud computing, his interests include software analytics and building large software systems. Gopal has a Masters' degree in computer science from North Carolina State University.

In this book, Gopal has shared his expert knowledge on Microsoft Azure in [Chapter 3](#) as well as the section titled “Managing PaaS” in [Chapter 8](#).



Nigel Cook is an HP distinguished technologist and technical director for the HP CloudSystem program. He has worked in areas of data center automation and distributed management systems for over 20 years, spanning environments as diverse as embedded systems for power utility control, telecom systems, and enterprise data center environments. At HP he created the BladeSystem Matrix Operating environment, and prior to that he served as chief architect on the Adaptive Enterprise and Utility Data Center programs. Prior to HP, he established and ran the US engineering operations of a software R+D development company specializing in telecom distributed systems. He received a BEng from University of Queensland, and is currently pursuing an MSc degree from University of Colorado, Boulder in the area of cloud computing based bioinformatics.

In this book, Nigel has contributed the section “HP CloudSystem Matrix” in [Chapter 2](#), as well as to the [Chapter 8](#) on “Managing the Cloud”.



Prakash S Raghavendra has been a faculty member at the IT Department of NITK, Surathkal from February 2009. He received his doctorate from the Computer Science and Automation Department (IISc, Bangalore) in 1998, after graduating from IIT Madras in 1994.

Earlier, Dr. Prakash worked in the Kernel, Java and Compilers Lab in Hewlett-Packard ISO in Bangalore from 1998 to 2007. Dr. Prakash has also worked for Adobe Systems, Bangalore from 2007 to 2009 in the area of flex profilers.

Dr. Prakash's current research interests include programming for heterogeneous computing, Web usage mining and rich Internet apps. Dr. Prakash has been honored with the 'Intel Parallelism Content Award' in 2011 and the 'IBM Faculty Award' for the year 2010.

In this book, Dr. Prakash has contributed about Adobe RIA in the section titled "Rich Internet Applications" in [Chapter 5](#).



Praphul Chandra is a Research Scientist at HP Labs India. He works on the simplifying web access and interaction project. His primary area of interest is complex networks in the context of social networks and information networks like the Web. At HP Labs, he also works on exploring new embedded systems architecture for emerging markets.

He is the author of two books – *Bulletproof Wireless Security* and *Wi-Fi Telephony: Challenges and Solutions for Voice over WLANs*. He joined HP Labs in April 2006. Prior to joining HP he was a senior design engineer at Texas Instruments (USA) where he worked on Voice over IP with specific focus on wireless local area networks. He holds an M.S. in electrical engineering from Columbia University, NY, a PG Diploma in public policy from University of London and a B.Tech. in electronics and communication engineering from Institute of Technology, BHU. His other interest areas are evolution and economics.

In this book, Praphul has shared his expert knowledge on Social networking in the section titled "Social Computing Services" in [Chapter 4](#).



Vanish Talwar is a principal research scientist at HP Labs, Palo Alto, researching management systems for next generation data centers. His research interests include distributed systems, operating systems, and computer networks, with a focus on management technologies. He received his Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign (UIUC). Dr. Talwar is a recipient of the David J Kuck Best Masters Thesis award from the Dept. of Computer Science, UIUC, and has numerous patents and papers, including a book on utility computing. In this book, Dr. Vanish has contributed to the [Chapter 8](#) titled “Managing the Cloud” and sections on “DMTF” and “OpenCirrus” in [Chapter 10](#).

Foreword

Prith Banerjee

Senior Vice President of Research and Director of HP Labs, Hewlett-Packard Company

Information is the most valuable resource in the 21st century. Whether for a consumer looking for a restaurant in San Francisco, a small business woman checking textile prices in Bangalore, or a financial services executive in London studying stock market trends, information *at the moment of decision* is key in providing the insights that afford the best outcome.

We now are sitting at a critical juncture of two of the most significant trends in the information technology industry – the convergence of cloud computing and mobile personal information devices into the Mobility/Cloud Ecosystem that delivers next-generation personalized experiences using a scalable and secure information infrastructure. This ecosystem will be able to store, process, and analyze massive amounts of information around structured, unstructured and semi-structured data. All this data will be accessed and analyzed at the speed of business.

In the past few years, the information technology industry began describing a future where everything is delivered as a service via the cloud, from computing resources to personal interactions. The future mobile internet will be 10 times the size of the desktop internet, connecting more than 10 billion “devices” from smartphones to wireless home appliances. Information access will then be as ubiquitous as electricity. Research advancements that the IT industry is making today will allow us to drive economies of scale into this next phase of computing to create a world where increasing numbers of people will be able to participate in and benefit from the information economy.

This book provides an excellent overview of all the transformations that are taking place in the IT industry around Cloud computing, and that, in turn, are transforming society. The book provides an overview of the key concepts of cloud computing, analyzes how cloud computing is different from traditional computing and how it enables new applications while providing highly scalable versions of traditional applications. It also describes the forces driving cloud computing, describes a well-known taxonomy of cloud architectures, and discusses at a high level the technological challenges inherent in cloud computing.

The book covers key areas of the different models of cloud computing: infrastructure as a service, platform as a service and software as a service. It then talks about paradigms for developing cloud applications. It finally talks about cloud-related technologies such as security, cloud management and virtualization.

HP Labs as the central research organization for Hewlett Packard has carried out research in many aspects of cloud computing in the past decade. The authors of the book are researchers in HP Labs India, and have contributed to many years of research on these topics. They have been able to provide their own personal research insight into the contents of the book and their vision of where this technology is headed.

I wish the readers of the book the best of luck in their journey to cloud computing!

Preface

First of all, thanks very much for choosing this book. We hope that you will like reading it and learn something new during the process. We believe the depth and breadth of the topics covered in the book will cater to a vast technical audience. Technologists who have a very strong technical background in distributed computing will probably like the real-life case studies of cloud platforms that enable them to get a quick overview of current platforms without actually registering for trials and experimenting with the examples. Developers who are very good in programming traditional systems will probably like the simple and complex examples of multiple cloud platforms that enable them to get started on programming to the cloud. It will also give them a good overview of the fundamental concepts needed to program a distributed system such as the cloud and learn new techniques to enable them to write efficient, scalable cloud services. We believe even research students will find the book useful to identify some open problems that are yet to be solved and help the evolution of cloud technologies to address all the current gaps.

Having worked on different aspects of systems technology particularly related to distributed computing for a number of years, we both were often discussing the benefits of cloud computing and what realignment in technology and mindset that the cloud required. In one such discussion, it dawned on us that a book based on real case studies of cloud platforms can be very valuable to technologists and developers, especially if we can cover the underlying technologies and concepts. We felt that many of the books available on cloud computing seemed to have a one-dimensional view of cloud computing. Some books equate cloud computing to just a specific cloud platform, say Amazon or Azure. Other books discuss cloud computing as if it is simply a new way of managing traditional data centers in a more cost-effective manner. There is also no dearth of books that hype the benefits of cloud computing in the ideal world.

In fact, the different perspectives about cloud computing that exist today remind us of the well-known story of the six blind men and the elephant. The blind man who caught hold of the elephant's tail insisted that the elephant is like a rope, while another who touched the elephant's tusks said that the elephant is like a spear, and so on. It definitely seemed to us that there is a need for a book that ties together the different aspects of cloud computing, both at the depth as well as breadth. However, we knew that covering all topics related to cloud in a single book, or even covering all popular cloud platforms as case studies, was not really feasible. We decided to cover at least three to four diverse case studies in each aspect of cloud computing and get into the technical depth in each of those case studies.

The second motivation for writing this book is to provide sufficiently deep knowledge to programmers and developers who will create the next generation of cloud applications. Many existing books focus entirely upon writing programs, without analyzing the key concepts or alternative implementations. It is our belief that in order to efficiently design programs it is necessary to have a good understanding of the technology involved, so that intelligent trade-offs can be made. It is also important to

design appropriate algorithms and choose the right cloud platform so that the solution to the given problem is scalable and efficient to execute on the cloud. For example, many cloud platforms today offer automatic scaling. However, in order to use this feature effectively, a high-level understanding of how the platform handles scaling is required. It is also important to select the right algorithm for special cloud platforms so that the solution to the given problem can be solved in the most efficient way for the use case and cloud platform (such as Hadoop MapReduce).

The challenge for us has been how to cover all the facets of cloud computing (provide a holistic view of the elephant) without writing a book that itself is as large as an elephant. To achieve this, we have adopted the following strategy. First, for each cloud platform, we provide a broad overview of the platform. This is followed by detailed discussion of some specific aspect of the platform. This high-level overview, together with a detailed study of a particular aspect of the platform, will give readers a deep insight into the basic concepts and features underlying the platform. For example, in the section on Salesforce.com, we start with a high-level overview of the features, followed by detailed discussion of using the call center features, programming under Salesforce.com, and important performance trade-offs for writing programs. Further sections cover the platform architecture that enables Salesforce.com, and some of the important underlying implementation details. The technology topics are also discussed in depth. For example, MapReduce is first introduced in [Chapter 3](#) with an overview of the concept and usage from a programming perspective. In later sections, a detailed look at the new programming paradigm that MapReduce enables along with fundamentals of functional programming, data parallelism and even theoretical formulation of the MapReduce problem are introduced. Many examples of how one can redesign an algorithm to suit the MapReduce platform are given. Finally, the internal architecture of the MapReduce platform, with details of how the performance, security and other challenges of cloud computing are handled in the platform, is described.

In summary, this book provides an in-depth introduction to the various cloud platforms and technologies today. In addition to describing the developer tools, platforms and APIs for cloud applications, it emphasizes and compares the concepts and technologies behind the platforms, and provides complex examples of their usage as invited content from experts in cloud platforms. This book prepares developers and IT professionals to become experts in cloud technologies, move their computing solutions to the cloud and also explore potential future research topics. It may be kindly noted that the APIs and functionality described in this book are as per the versions available at the time of the writing of this book. Readers are requested to refer to the latest product documentation for accurate information. Finally, since this area is evolving rapidly, we plan to continuously review the latest cloud computing technologies and platforms on our companion website <http://www.movingtocloudbook.com>.

Structure of the Book

[Chapter 1](#) of the book is the introduction and provides a high-level overview of cloud computing. We start with the evolution of cloud computing from Web 1.0 to Web 2.0, and discuss its evolution in the future. Next, we discuss various cloud computing models (IaaS, PaaS, and SaaS) and the cloud deployment models (public, private,

community and hybrid) together with the pros and cons of each model. Finally, the economics of cloud computing and possible cost savings are described.

[Chapter 2](#), [Chapter 3](#) and [Chapter 4](#) describe the three cloud service models (IaaS, PaaS, and SaaS) in detail – from a developer and technologist stand point. The platform models are explained using popular cloud platforms as case studies (for example, Amazon for IaaS and Windows Azure for PaaS) through sample programs, as well as an overview of the underlying technology. While describing program development, the book tries to follow a standard pattern. First, a simple *Hello World* program that allows users to get started is described. This is followed by a more complex example that illustrates commonly used features of the major APIs of the platform. The complex example also introduces the concepts underlying the platform (for example, MapReduce in Hadoop). These chapters will provide programmers interested in developing cloud applications a good understanding of the features and differences between the various existing cloud platforms. In addition, professionals who are interested in the technology behind cloud computing will understand key platform features that are needed to motivate a discussion of the technology and evaluate the suitability of a platform for their specific use case.

[Chapter 2](#) describes three important IaaS platforms – Amazon, HP CloudSystem Matrix, and a research prototype called Cells-as-a-Service. The first section of the chapter describes the Amazon storage services – S3, SimpleDB, and Relational Database Service with GUI and programming examples. The chapter also describes how to upload large files and multi-part uploads. The next section describes Amazon's EC2 cloud service. This contains descriptions of how to administer and use these services through the Web GUI, and also a code example of how to set up a document portal in EC2 using a running example called Pustak Portal (details of which are described towards the end of this Preface). Methods are presented for automatically scaling up and down the service using both Amazon Beanstalk as well as custom code (when Beanstalk is not suitable). The next sections of the chapter describe HP CloudSystem Matrix, and Cells-as-a-Service, a research prototype developed by HP Labs. Here again, after describing the basic features of the offering, the section describes how to set up the document portal in our running example (Pustak Portal). Methods for autoscaling up or autoscaling down the portal are described.

[Chapter 3](#) describes some important PaaS cloud platforms – Windows Azure, Google AppEngine, Apache Hadoop, IBM PureXML, and mashups. The Windows Azure section first describes a simple “Hello World” program that illustrates the basic concepts of Web and Worker roles, and shows how to test and deploy programs under Azure. Subsequently, the architecture of the Azure platform, together with its programming model, storage services such as SQL Azure, as well as other services such as security are described. These are illustrated with the running example of implementing Pustak Portal. In the Google App Engine section, the process of developing and deploying programs is described, together with use of the Google App Engine storage services and memory caching. Next IBM PureXML, which is a cloud service that exposes both a relational as well as XML database interface, is discussed. Examples of how to store data for a portal such as Pustak Portal are described. The next section describes Apache Hadoop, including examples of MapReduce programs, and how Hadoop Distributed File System can be used to provide scalable storage. The