Enterprise Cloud Computing

Technology, Architecture, Applications

Gautam Shroff

MBRIDGE

ENTERPRISE CLOUD COMPUTING

TECHNOLOGY, ARCHITECTURE, APPLICATIONS

GAUTAM SHROFF

常州大学图书馆
藏书章

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521760959

© G. Shroff 2010

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

ISBN 978-0-521-76095-9 Hardback
ISBN 978-0-521-13735-5 Paperback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to in
this publication, and does not guarantee that any content on such websites is,
or will remain, accurate or appropriate.

ENTERPRISE CLOUD COMPUTING

Technology, Architecture, Applications

Cloud computing promises to revolutionize IT and business by making computing available as a utility over the internet. This book is intended primarily for practicing software architects who need to assess the impact of such a transformation. It explains the evolution of the internet into a cloud computing platform, describes emerging development paradigms and technologies, and discusses how these will change the way enterprise applications should be architected for cloud deployment.

Gautam Shroff provides a technical description of cloud computing technologies, covering cloud infrastructure and platform services, programming paradigms such as MapReduce, as well as 'do-it-yourself' hosted development tools. He also describes emerging technologies critical to cloud computing. The book also covers the fundamentals of enterprise computing, including a technical introduction to enterprise architecture, so it will interest programmers aspiring to become software architects and serve as a reference for a graduate-level course in software architecture or software engineering.

Gautam Shroff heads TCS' Innovation Lab in Delhi, a corporate R&D lab that conducts applied research in software architecture, natural language processing, data mining, multimedia, graphics and computer vision. Additionally he is responsible for TCS' Global Co-Innovation Network (COIN), which works with venture-backed emerging technology companies to create and take to market solutions that have disruptive innovation potential. Further, as a member of TCS' Corporate Technology Board, he is part of the process of recommending directions to existing R&D efforts, spawning new R&D efforts, sponsoring external research and proliferating the resulting technology and intellectual property across TCS' businesses.

Preface

In today's world virtually all available information on any technical topic is just a few clicks away on the web. This is especially true of an emerging area such as cloud computing. So why write a book, and, who should read this book and why?

Every few years a new 'buzzword' becomes the rage of the technology world. The PC in the 80s, the internet in the 90s, service-oriented architecture in the early 2000s, and more recently 'cloud computing': By enabling computing itself to be delivered as a utility available over the internet, cloud computing could transform enterprise IT. Such a transformation could be as significant as the emergence of power utilities in the early twentieth century, as eloquently elucidated in Nicholas Carr's recent book *The Big Switch*.

Over the years large enterprises have come to rely on information technology to run their increasingly complex business operations. Each successive technology 'revolution' promises tremendous gains. It falls upon the shoulders of the *technical architects* in the IT industry to evaluate these promises and measure them against the often significant pain that is involved in adapting complex IT systems to new computing paradigms: The transition to cloud computing is no exception.

So, this book is first and foremost for technical architects, be they from IT departments or consulting organizations. The aim is to cover cloud computing **technology**, **architectures** and **applications** in detail, so as to be able to properly assess its true impact on enterprise IT.

Since cloud computing promises to fundamentally revolutionize the way enterprise IT is run, we also revisit many principles of enterprise architecture and applications. Consequently, this is also a book on the fundamentals of **enterprise computing**, and can therefore serve as a reference for a

此为试读,需要完整PDF请访问: www.ertongbook.com

graduate-level course in software architecture or software engineering. Alternatively, software professionals interested in acquiring the 'architect' tag may also find it a useful read.

From a personal perspective this book is also an attempt to capture my experience of a decade in the IT industry after an initial career in academic computer science: The IT industry seemed ever busier dealing with constant changes in technology. At the same time, every generation of professionals, in particular the technical architects, were constantly reinventing the wheel: Even though automation techniques, such as large-scale code generation using 'model driven architecture' often actually worked in practice, these were far from the panacea that they theoretically appeared to be.

Nevertheless, the academic in me continued to ask, what after all does an enterprise application *do*, and why should it be so complex? In 2004 I wrote an *interpreter* for what appeared to me to be a perfectly reasonable 3-tier architecture on which, I thought, *any* enterprise application should run. This was the seed of what became TCS' InstantApps platform. At the same time Salesforce.com was also experimenting with an interpretive architecture that later became Force.com. While software as a service was the rage of the industry, I began using the term Dev 2.0 to describe such interpretive hosted development platforms.

In the meantime Amazon launched its elastic computing cloud, EC2. Suddenly, the entire IT infrastructure for an enterprise could be set up 'in the cloud.' 'Dev 2.0 in the Cloud' seemed the next logical step, as I speculated in a keynote at the 2008 ACM SIGSOFT FSE conference. After my talk, Heather Bergman from Cambridge University Press asked me whether I would be interested in writing a book. The idea of a book had been in my mind for more than a year then; I had envisaged a book on software architecture. But maybe a technical book on cloud computing was more the need of the hour. And thus this book was born.

In my attempt to present cloud computing in the context of enterprise computing, I have ended up covering a rather vast landscape. Part I traces the evolution of computing technology and how enterprise architecture strives to manage change with continuity. Part II introduces cloud computing platforms and the economics of cloud computing, followed by an overview of technologies essential for cloud applications in Part III. Part IV delves into the details of cloud computing and how it impacts application development. The essentials of enterprise software architecture are covered in Part V, from an overview of enterprise data models to how applications are built. We also show how the *essence* of what an enterprise application does can be abstracted

using *models*. Part V concludes with an integrated picture of enterprise analytics and search, and how these tasks can be efficiently implemented on computing clouds. These are important topics that are unfamiliar to many architects; so hopefully, their unified treatment here using matrix algebra is illuminating. Finally, Part VI presents an overview of the industry ecosystem around enterprise cloud computing and concludes by speculating on the possible future of cloud computing for enterprises.

A number of people have helped bring this book to fruition: First of all, Heather Bergman who suggested that I write, helped me finalize the topic and table of contents, and led me through the book proposal process in record time. Once the first draft was written, Jeff Ullman reviewed critical parts of the book in great detail, for which I remain eternally grateful. Rob Schreiber, my PhD advisor from another lifetime, also took similar pains, even 20 years after doing the same with my PhD thesis; thanks Rob! Many of my colleagues in TCS also reviewed parts of the manuscript; in particular Ananth Krishnan, C. Anantaram, Puneet Agarwal, Geetika Sharma, Lipika Dey, Venkatachari Raghavan, Surjeet Mishra, Srinivasan Varadanarayanan and Harrick Vin. I would also like to thank David Tranah for taking over as my editor when Heather Bergman left Cambridge University Press soon after I began writing, and for shepherding the book through the publication process.

Finally, I am grateful for the continuous encouragement and support I have received over the years from TCS management, in particular F.C. Kohli, S. Ramadorai and Phiroz Vandrevala, as well as, more recently, N. Chandrasekaran. I would also like to thank E. C. Subbarao and Kesav Nori, who have been my mentors in TCS R&D, for serving as role models, influencing my ideas and motivating me to document my experience.

I have learned that while writing is enjoyable, it is also difficult: Whenever my intrinsic laziness threatened this project, my motivation was fueled by the enthusiasm of my family. With my wife, sister-in-law and mother-in-law all having studied at Cambridge University, I suspect this was also in no small measure due to the publisher I was writing for! Last but not least, I thank my wife Brinda, and kids Selena and Ahan, for tolerating my preoccupation with writing on weekends and holidays for the better part of a year.

I sincerely hope that you enjoy reading this book as much as I have enjoyed writing it.

Abbreviations

Term	Description
AJAX	Asynchronous JavaScript and XML
AMI	Amazon Machine Image
API	Application Programming Interface
BPMN	Business Process Modeling Notation
CGI	Common Gateway Interface
CICS	Customer Information Control System
CORBA	Common Object Request Broker Architecture
CPU	Central Processing Unit
CRM	Customer Relationship Management
CRT	Cathode Ray Tube
EAI	Enterprise Application Integration
EBS	[Amazon] Elastic Block Storage
EC2	Elastic Compute Cloud
ECA	Event Condition Action
EJB	Enterprise Java Beans
ERP	Enterprise Resource Planning
GAE	Google App Engine
GFS	Google File System
GL	General Ledger
GML	Generalized Markup Language
HDFS	Hadoop Distributed File System
HTML	Hypertext Transport Protocol and Secure Socket Layer
HTTP	Hypertext Transport Protocol
HTTPD	Hypertext Transfer Protocol Daemon

Term	Description
IA	[TCS] InstantApps
IaaS	Infrastructure as a Service
IBM	International Business Machines
IDL	Interface Definition Language
IDMS	Integrated Database Management System
IDS	Integrated Data Store [Database System]
IIS	Internet Information Server
IMS	[IBM] Information Management System
IT	Information Technology
ITIL	Information Technology Infrastructure Library
J2EE	Java 2 Enterprise Edition
JAAS	Java Authentication and Authorization Service
JCL	Job Control Language
JSON	JavaScript Object Notation
LDAP	Lightweight Directory Access Protocol
MDA	Model Driven Architecture
MDI	Model Driven Interpreter
MDX	Multidimensional Expressions [Query Language]
MVC	Model View Controller
MVS	Multiple Virtual Storage [Operating System]
OLAP	Online analytical processing
OMG	Object Management Group
PaaS	Platform as a Service
PKI	Public Key Infrastructure
REST	Representational State Transfer
RMI	Remote Method Invocation
RPC	Remote Procedure Call
SaaS	Software as a Service
SCM	Supply Chain Management
SGML	Standardized Generalized Markup Language
SNA	Systems Network Architecture
SOA	Service Oriented Architecture
SOAP	Simple Object Access Protocol
SQL	Structured Query Language
SQS	[Amazon] Simple Queue Service
SVD	Singular Value Decomposition

Term	Description
TCP/IP	Transmission Control Protocol/Internet Protocol
TCS	Tata Consultancy Services
T&M	Time and Materials
TP Monitor	Transaction Processing Monitor
UML	Unified Modeling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locater
VM	Virtual Machine
VMM	Virtual Machine Monitor
VPC	Virtual Private Cloud
VPN	Virtual Private Network
VSAM	Virtual Storage Access Method
VTAM	Virtual Telecommunications Access Method
W3C	World Wide Web Consortium
WSDL	Web Services Description Language
WYSIWYG	What You See is What You Get
XHTML	Extensible Hypertext Markup Language
XML	Extensible Markup Language

Contents

Preface *page* **xi**

List of abbreviations **xiv**

Part I Computing platforms **1**

Chapter 1
Enterprise computing: a retrospective **3**

1.1 Introduction 3
1.2 Mainframe architecture 5
1.3 Client-server architecture 7
1.4 3-tier architectures with TP monitors 10

Chapter 2
The internet as a platform **16**

2.1 Internet technology and web-enabled applications 16
2.2 Web application servers 19
2.3 Internet of services 22

Chapter 3
Software as a service and cloud computing **27**

3.1 Emergence of software as a service 27
3.2 Successful SaaS architectures 29

此为试读,需要完整PDF请访问: www.ertongbook.com

3.3 Dev 2.0 platforms 31
3.4 Cloud computing 32
3.5 Dev 2.0 in the cloud for enterprises 36

Chapter 4
Enterprise architecture: role and evolution 39

4.1 Enterprise data and processes 40
4.2 Enterprise components 40
4.3 Application integration and SOA 42
4.4 Enterprise technical architecture 44
4.5 Data center infrastructure: coping with complexity 47

Part II Cloud platforms 49

Chapter 5
Cloud computing platforms 51

5.1 Infrastructure as a service: Amazon EC2 51
5.2 Platform as a service: Google App Engine 56
5.3 Microsoft Azure 60

Chapter 6
Cloud computing economics 64

6.1 Is cloud infrastructure cheaper? 64
6.2 Economics of private clouds 67
6.3 Software productivity in the cloud 71
6.4 Economies of scale: public vs. private clouds 73

Part III Cloud technologies 75

Chapter 7
Web services, AJAX and mashups 77

7.1 Web services: SOAP and REST 77
7.2 SOAP versus REST 83
7.3 AJAX: asynchronous 'rich' interfaces 85
7.4 Mashups: user interface services 87

Chapter 8
Virtualization technology **89**

8.1 Virtual machine technology 89
8.2 Virtualization applications in enterprises 95
8.3 Pitfalls of virtualization 103

Chapter 9
Multi-tenant software **104**

9.1 Multi-entity support 105
9.2 Multi-schema approach 107
9.3 Multi-tenancy using cloud data stores 109
9.4 Data access control for enterprise applications 111

Part IV Cloud development **115**

Chapter 10
Data in the cloud **117**

10.1 Relational databases 118
10.2 Cloud file systems: GFS and HDFS 121
10.3 BigTable, HBase and Dynamo 123
10.4 Cloud data stores: Datastore and SimpleDB 128

Chapter 11
MapReduce and extensions **131**

11.1 Parallel computing 131
11.2 The MapReduce model 134
11.3 Parallel efficiency of MapReduce 137
11.4 Relational operations using MapReduce 139
11.5 Enterprise batch processing using MapReduce 142

Chapter 12
Dev 2.0 platforms **144**

12.1 Salesforce.com's Force.com platform 145
12.2 TCS InstantApps on Amazon cloud 148

12.3 More Dev 2.0 platforms and related efforts 153
12.4 Advantages, applicability and limits of Dev 2.0 154

Part V Software architecture 159

Chapter 13
Enterprise software: ERP, SCM, CRM 161

13.1 Anatomy of a large enterprise 161
13.2 Partners: people and organizations 164
13.3 Products 167
13.4 Orders: sales and purchases 168
13.5 Execution: tracking work 170
13.6 Billing 172
13.7 Accounting 174
13.8 Enterprise processes, build vs. buy and SaaS 176

Chapter 14
Custom enterprise applications and Dev 2.0 178

14.1 Software architecture for enterprise components 178
14.2 User interface patterns and basic transactions 180
14.3 Business logic and rule-based computing 188
14.4 Inside Dev 2.0: model driven interpreters 194
14.5 Security, error handling, transactions and workflow 198

Chapter 15
Workflow and business processes 203

15.1 Implementing workflow in an application 203
15.2 Workflow meta-model using ECA rules 205
15.3 ECA workflow engine 207
15.4 Using an external workflow engine 210
15.5 Process modeling and BPMN 211
15.6 Workflow in the cloud 216

Chapter 16
Enterprise analytics and search **217**

16.1 Enterprise knowledge: goals and approaches 218
16.2 Business intelligence 219
16.3 Text and data mining 225
16.4 Text and database search 235

Part VI Enterprise cloud computing **241**

Chapter 17
Enterprise cloud computing ecosystem **243**

17.1 Public cloud providers 244
17.2 Cloud management platforms and tools 246
17.3 Tools for building private clouds 247

Chapter 18
Roadmap for enterprise cloud computing **253**

18.1 Quick wins using public clouds 254
18.2 Future of enterprise cloud computing 257

References **264**

Index **269**

PART I

Computing platforms

Barely 50 years after the birth of enterprise computing, cloud computing promises to transform computing into a utility delivered over the internet. A historical perspective is instructive in order to properly evaluate the impact of cloud computing, as well as learn the right lessons from the past. We first trace the history of enterprise computing from the early mainframes, to client-server computing and 3-tier architectures. Next we examine how the internet evolved into a computing platform for enterprise applications, naturally leading to Software as a Service and culminating (so far) in what we are now calling cloud computing. Finally we describe how the 'enterprise architecture' function within IT departments has evolved over time, playing a critical role in managing transitions to new technologies, such as cloud computing.

此为试读,需要完整PDF请访问: www.ertongbook.com