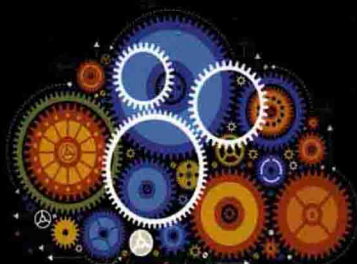


深入理解云计算

基本原理和应用程序编程技术

[澳] 拉库马·布亚 (Rajkumar Buyya)
[印] 克里斯坦·维奇拉 (Christian Vecchiola) 著
[印] S. 泰马莱·赛尔维 (S. Thamarai Selvi)
刘丽 米振强 熊曾刚 译

Mastering Cloud Computing
Foundations and Applications Programming



MASTERING
CLOUD COMPUTING
FOUNDATIONS AND APPLICATIONS PROGRAMMING

MK
MORGAN KAUFMANN

Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi



机械工业出版社
China Machine Press

计 算 机 科 学 丛 书

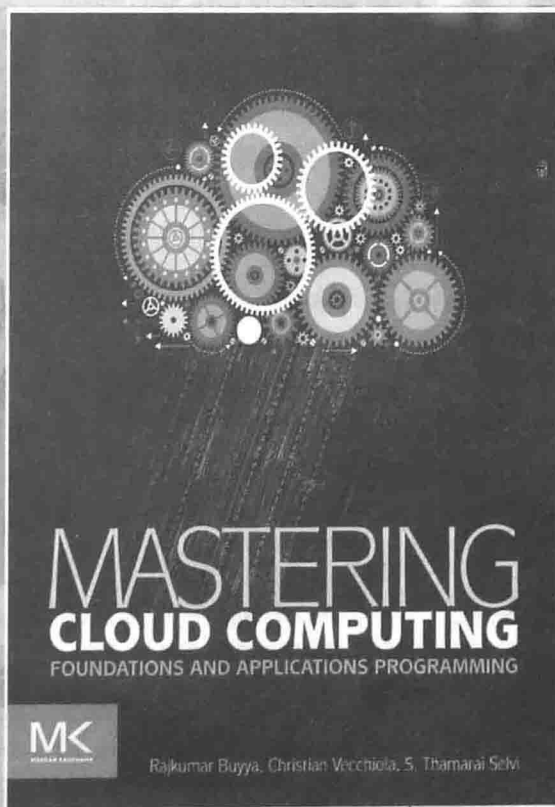
深入理解云计算

基本原理和应用程序编程技术

[澳] 拉库马·布亚 (Rajkumar Buyya)
[印] 克里斯坦·维奇拉 (Christian Vecchiola) 著
[印] S. 泰马莱·赛尔维 (S. Thamarai Selvi)

刘丽 米振强 熊曾刚 译

Mastering Cloud Computing
Foundations and Applications Programming



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

深入理解云计算：基本原理和应用程序编程技术 / (澳) 布亚 (Buyya, R) 等著, 刘丽, 米振强, 熊曾刚译. —北京: 机械工业出版社, 2015.4

(计算机科学丛书)

书名原文: Mastering Cloud Computing: Foundations and Applications Programming

ISBN 978-7-111-49658-8

I. 深… II. ①布… ②刘… ③米… ④熊… III. 计算机网络—研究 IV. TP393

中国版本图书馆 CIP 数据核字 (2015) 第 052134 号

本书版权登记号: 图字: 01-2013-7841

Mastering Cloud Computing: Foundations and Applications Programming

Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi

ISBN:978-0-12-411454-8

Copyright © 2013 by Elsevier Inc. All rights reserved.

Authorized Simplified Chinese translation edition published by the Proprietor.

Copyright © 2015 by Elsevier (Singapore) Pte Ltd. All rights reserved.

Printed in China by China Machine Press under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macau SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书简体中文版由 Elsevier (Singapore) Pte Ltd. 授权机械工业出版社在中国大陆境内独家出版和发行。本版仅限在中国境内 (不包括香港特别行政区、澳门特别行政区及台湾地区) 出版及标价销售。未经许可之出口, 视为违反著作权法, 将受法律之制裁。

本书封底贴有 Elsevier 防伪标签, 无标签者不得销售。

本书从云基础知识、云应用编程和云平台三个方面, 介绍云计算的起源、发展、核心技术、编程技巧和实际应用, 基于 Aneka 平台详细讲解并行计算、高吞吐量计算和数据密集型计算的范式。此外, 还涉及亚马逊 Web 服务、谷歌 AppEngine 和微软 Azure 三大云平台, 以及云计算在科学、工程、游戏、社交等领域的最新应用。

本书内容严谨、结构清晰、实例丰富, 既可作为高等院校计算机相关专业的教材, 也适合云技术研发人员阅读参考。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 曲 熠

责任校对: 董纪丽

印 刷: 北京瑞德印刷有限公司

版 次: 2015 年 4 月第 1 版第 1 次印刷

开 本: 185mm×260mm 1/16

印 张: 22

书 号: ISBN 978-7-111-49658-8

定 价: 69.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域中取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近百两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：www.hzbook.com

电子邮件：hzjsj@hzbook.com

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



华章科技图书出版中心

云计算是在传统计算机科学基础上发展起来的新兴技术。随着移动设备的大量普及、网络与计算成本的大幅降低以及用户计算需求的不断提升,传统的PC计算模式必将向以云数据中心为核心的云计算模式转移。同时,云计算技术所提供的强大的计算能力、低廉的成本以及按需服务的模式将会从根本上推动计算密集型应用的进一步发展,以及全球计算能力的全方位释放。

由于涉及计算机科学领域的专业知识和技术,所以对普通研发人员而言,云计算技术颇有些阳春白雪的意味。究其根源,云计算技术是学术研究与工程开发的集合体。一方面,研究人员在有好的想法或者概念的情况下,需要一个良好的云计算模型与原型系统来验证所提出的方法;另一方面,工程人员则由于专业背景知识的缺乏,而很难进入真正的云计算技术开发领域。

作为云计算技术领域的先驱者和领导者,本书作者——澳大利亚墨尔本大学计算机科学系著名教授 Rajkumar Buyya,长期从事网格计算及云计算技术研究,研究成果得到学术界与工业界的广泛认可。本书集成了 Buyya 教授在云计算领域多年的研究成果,从云计算基础、应用平台和编程技术三个方面,利用 11 个章节详细介绍云计算技术的起源、发展、核心技术及其范式、典型云计算平台以及编程技巧等重要内容。本书内容丰富、深入浅出,并包含配套习题,适合不同层次的云计算技术研发人员使用。同时,本书清晰的思路、严谨的技术框架和详尽的实例讲解也使其成为适用于计算机相关学科本科生及研究生教学的不可多得的优秀教材。

参与本书翻译工作的人员均多年从事云项目相关工作,对云计算前沿问题的研究及教学都有较深刻的理解。本书前言、第 1~4 章及索引由刘丽(北京科技大学自动化学院副教授)翻译,第 5~8 章由米振强(北京科技大学计算机与通信工程学院讲师)翻译,第 9~11 章由熊曾刚(湖北工程学院计算机与信息科学学院教授)翻译,全书由刘丽统稿。北京科技大学的研究生张森、翟颖奇、夏毓娴、徐安琪、谢翔、潘梦圆、李萌、何苗、陆源等同学协助进行资料收集,并参与了部分章节的初稿翻译工作。

感谢机械工业出版社引进此书并为本书的出版付出大量努力,使 IT 从业人员和计算机相关专业学生从中受益。特别感谢本书作者 Rajkumar Buyya 教授对翻译过程中遇到的问题进行解答。

需要说明的是,本书翻译工作和云计算研究工作得到以下项目的资助和支持:国家自然科学基金项目“互联云环境中基于效用模型的跨云协同服务优化研究(No. 61370132)”、“大数据环境下基于视觉主题模型的视觉数据分类方法研究(No. 61370092)”、“互联云环境下面向数据中心的服务资源分配与调度机制研究(No. 61472033)”,国家高技术研究发展计划(863 计划)项目“城市多模式数据系统互联技术与支撑环境(No. 2013AA01A601)”,以及湖北省自然科学基金项目“云计算环境下内容语义信任度量与评估方法研究

(No. 2013CFC005)”和湖北省高等学校优秀中青年科技创新团队计划项目“云计算环境下智能信息处理技术研究(No. T201410)”。

由于译者对云计算相关变革性技术的理解有待加深,而且许多新出现的专业术语还没有公认的译法,所以在翻译过程中难免出现一些不够清楚的表述,若有不妥之处,恳请广大读者批评指正,电子邮箱 liuli@ustb.edu.cn。

刘丽

2015年1月

随着互联网与 Web 技术的快速发展和普及,以及手持计算机、移动设备、传感器设备功能的不断强大,人与人之间的交互方式、商业行为以及获取和提供服务的模式都在发生变化。低成本的计算与通信驱动了从单一计算方式向以数据中心为核心的计算方式的转变。尽管并行与分布式计算在 IT 行业已经存在多年,但其新的形式——多核和云计算为 IT 行业带来了彻底的变革。这种发展趋势将促使 IT 行业从 PC 应用开发模式转变为支持数百万用户同时使用软件的云数据中心模式。

计算向商业服务模式变革,这种计算服务类似于传统的公共基础设施服务,如水、电、煤气和电话。因此 IT 服务被当作与水、电、煤气和电话一样的“计算公共基础设施”,通过共有传输网络来交付使用并计费。在这种计算环境下,用户按需获取服务,而不管该服务由哪里提供。一些计算模式已经提出交付这种效用计算服务的构想,云计算是最新出现的实现这一目标的计算模式。

云计算如今已成为 IT 行业的又一个流行术语。众多 IT 厂商承诺提供存储、计算及应用托管服务,其服务范围涵盖几大洲,并能提供基于服务等级协议(Service Level Agreement, SLA)的服务性能保障和运行时间承诺。云服务模式允许用户基于订阅方式访问基础设施、平台及应用,也就是通常所说的基础设施即服务(IaaS)、平台即服务(PaaS)、软件即服务(SaaS)。这种服务模式大大降低了计算和应用的成本,但是要实现应用和服务开发及交付的一致性、可扩展性、可靠性是极其复杂的。

已有一些云计算技术和云平台产品,如谷歌 AppEngine、微软 Azure 和 Manjrasoft Aneka。谷歌 AppEngine 利用大量 IT 基础设施为托管的 Web 应用提供可扩展的运行环境。微软 Azure 为在云计算环境中开发和部署应用提供了大量 Windows 服务实例。Manjrasoft Aneka 能够灵活地创建云应用并将其部署在各种基础设施上,如亚马逊公共云 EC2。

随着应用开发从 PC 向云数据中心迁移,需要大量掌握云计算技能的人员。面对这一挑战,大学教育在培养下一代 IT 专业人员方面发挥了重要作用,帮助学生学习和掌握新的相关技术与工具。这样,大学需要以较小的投入建立云计算教学环境,而 Manjrasoft Aneka 比较适合建立这种云应用平台,它允许用户利用已有计算机网络组建私有云/企业云,提供软件开发工具包(SDK),支持多种编程模型(如线程、任务、MapReduce)的应用编程接口(API),支持在多核服务器、私有云、公共云等不同基础设施上无缝地部署和执行应用。

如今,专业开发人员需要创建云应用和服务。云计算研究人员、从业者以及供应商努力让用户了解云计算的好处并充分利用其潜在能力。然而,由于云计算是一种新兴的计算范式,所以对于云计算的精确定义,不同的云计算专家会给出不同的答案。因此,尽管现在比以往更好地实现了真正的效用计算,但是,与云计算服务方交互的复杂性使得对于云计算的认可和应用还只限于领域专家。本书旨在用简单的方式向读者讲解云计算基础知识、技术及编程技能,让更多普通程序员和软件工程师能轻松地开发云应用程序。

本书结构

本书介绍云计算的基本原理及相关范式,阐述云计算架构模型中虚拟化技术的概念,并

展示包括 Aneka 云计算应用平台在内的著名云计算技术产品，详细讲解并行计算、高吞吐量计算和数据密集型计算的范式，以及如何将这些范式应用于云应用程序开发。本书还研究了来自科学界、工程界、游戏和社交网络领域的多个应用案例，阐述了各应用案例的架构以及云计算技术的应用方式。这些案例研究有助于读者对云计算原理的理解。最后，本书详细阐述了许多源于云计算快速应用的开放性研究问题和机遇，我们希望这有助于激发读者在未来的研发过程中解决这些问题。另外，了解本书相关内容可浏览 Web 站点 (<http://www.buyya.com/MasteringClouds>)，其中包含更多在线资源。

全书分为三部分，共 11 章：

第一部分 基础

第 1 章 导论

第 2 章 并行计算与分布式计算原理

第 3 章 虚拟化

第 4 章 云计算架构

第二部分 云应用编程与 Aneka 平台

第 5 章 Aneka：云应用平台

第 6 章 并行计算：线程编程

第 7 章 高吞吐量计算：任务编程

第 8 章 数据密集型计算：MapReduce 编程

第三部分 工业云平台与新进展

第 9 章 工业云平台

第 10 章 云应用

第 11 章 云计算高级主题

本书将引导读者进入云计算领域，从理论基础讲解开始，让学生和专业人员通过在 Aneka 平台上实际开发云应用程序来理解和掌握相关概念。第三部分介绍业界其他云技术和解决方案（亚马逊 Web 服务、谷歌 AppEngine 和微软 Azure）及其实际应用，阐述云计算的发展趋势和发展前景。

读者对象

由于云计算迅速崛起为一种主流计算模式，所以必须深入理解其核心概念和特性，并掌握如何设计和实现云计算的应用程序与系统。这是如今的软件架构师、工程师和开发人员应具有的基本技能，因为多数应用都将被迁移到云环境。随着技术的成熟，具备此技能尤其重要。本书涵盖云计算的起源、理论和实际开发技术，读者对象更广泛，可作为研究生、IT 从业者、开发人员、工程师等设计和实现云计算解决方案的参考书。此外，书中最后关于相关研究的展望更加吸引云计算领域的研究人员探究其将带来的新挑战。

云计算正在不断获取相当可观的商业利益且发展势头强劲，本书为云计算领域做出了非常适时的贡献。本书主要针对研究生和 IT 专家，例如系统架构师、软件工程师、应用程序员等。在未来的 20 年里，云计算将是对科学研究和社会生活产生重大影响的五大新兴技术之一，因此认真理解和掌握本书的知识将帮助读者置身 IT 领域的前沿。

用书指南：理论、实验室、项目

鉴于云计算范式的重要性及其在业界的快速崛起，教育机构应该更新其课程体系，增加

云计算或相关领域的一门或多门专业课程，例如“并行计算”和“分布式系统”。我们建议大学设置面向本科生或研究生的云计算专业，及计算科学学士和相关硕士学位，相信此书将是该专业的优秀教材。如果学生已经了解并行和分布式计算的概念，可以跳过第2章。

对于想用云计算丰富其课程体系的学校，建议分别在两个学期开设两门课程：“云计算基础课程”和“云计算高级课程”。本书第1~6章适用于云计算基础课程，第7~11章适用于云计算高级课程。

除了理论学习以外，我们强烈推荐实验室使用本书，书中给出了很多实验指导。实验练习和作业题包括数学函数的并行执行、大量数据排序的并行处理、图像处理和数据挖掘。在Aneka云软件系统上，学校很容易利用已有的Windows系统的计算机局域网络搭建私有云（企业云）计算环境。学生可利用此环境学习各种云应用编程模型和接口的实例并在Aneka平台上实现，如第6章的线程编程、第7章的任务编程、第8章的MapReduce编程。章后编程习题可作为实验作业让学生自己编写程序实现。

学生也可以在其本科毕业设计中开发处理实际问题的云应用程序。例如，学生可以协助其他科学领域或工程领域（比如生命医学科学、机械科学）的研究人员，利用云计算强大的计算能力开发符合实际需求的应用。请阅读并学习第10章的各种应用案例。

教辅资源^①

教师与学生需要的教辅资源可以从Elsevier网站下载（booksite.elsevier.com/9780124114548），也可以从作者为本书所做的主页（<http://www.buyya.com/MasteringClouds>）下载。

^① 关于本书教辅资源，使用教材的教师需通过爱思唯尔的教材网站（www.textbooks.elsevier.com）注册并通过审批后才能获取。具体方法如下：在www.textbooks.elsevier.com教材网站查找到该书后，点击“instructor manual”便可申请查看该教师手册。有任何问题，请致电010-85208853。——编辑注

首先感谢所有学者和云计算开发者为本书探讨的各种概念与技术所做出的贡献。特别感谢 Manjrasoft 公司、墨尔本大学云计算与分布式系统实验室 (CLOUDS)、墨尔本风险投资公司的所有员工和顾问, 他们为 Aneka 云应用平台的开发、应用示范和文档的准备以及 Aneka 技术的商业化做出了贡献。他们是: Chu Xingchen、Srikumar Venugopal、Krishna Nadiminti、Christian Vecchiola、Dileban Karunamoorthy、Chao Jin、Rodrigo Calheiros、Michael Mattess、Jessie Wei、Enayat Masoumi、Ivan Mellado、Richard Day、Wolfgang Gentzsch、Laurence Liew、David Sinclair、Suraj Pandey、Abhi Shekar、Dexter Duncan、Murali Sathya、Karthik Sukumar、Ravi Kumar Challa 和 Sita Venkatraman。

感谢澳大利亚研究委员会 (ARC) 与创新、工业、科学和研究部 (DIISR) 对我们的研究工作及其商业化尝试的支持。

感谢墨尔本大学的所有同事, 尤其是 Rao Kotagiri 教授、Iven Mareels 教授 和 Glyn Davis 教授, 他们对我们的研究和研究成果的传播给予指导与支持。

感谢 Aneka 技术的研发同事和用户, 他们对本书的应用案例做出了直接或间接的贡献。特别感谢 ADRIN/ISRO 的 Raghavendra Kune, 他利用 Aneka 建立了卫星图像处理应用并发表了文章。感谢 MSRIT 的 Srinivasa Iyengar, 他利用 Aneka 开发了数据挖掘应用程序, 并在云计算早期将 Aneka 引荐到学术界。

感谢 CLOUDS 实验室人员对本书部分章节的校对, 他们是: Rodrigo Calheiros、Nikolay Grozev、Amir Vahid、Anton Beloglazov、Adel Toosi、Deepak Poola、Mohammed AlRokayan、Atefeh Khosravi、Sareh Piraghaj 和 Yaser Mansouri。

感谢家人 Smrithi Buyya、Soumya Buyya 和 Radha Buyya 在本书编写过程中对我的爱和理解。

我们诚挚地感谢 Elsevier 出版社的外审专家对本书提出的意见和建议, 使得很多章节具有更好的组织方式和呈现效果, 这为我们提高全书质量带来了极大帮助。

最后, 我们要感谢 Elsevier 出版社的员工在书稿准备过程中给予的热情帮助和指导。特别感谢 Todd Green 鼓励我们承接此项目并促成此书出版。Elsevier 出版社的员工太棒了!

Rajkumar Buyya 教授

澳大利亚墨尔本大学和 Manjrasoft 公司

Christian Vecchiola 博士

澳大利亚墨尔本大学和 IBM 研究院

S. Thamarai Selvi 教授

印度安那大学金奈分校 Madras 技术学院

出版者的话

译者序

前 言

致 谢

第一部分 基础

第 1 章 导论	2
1.1 云计算简介	2
1.1.1 云计算构想	3
1.1.2 云计算定义	4
1.1.3 进一步了解云计算	6
1.1.4 云计算参考模型	7
1.1.5 特性和优势	9
1.1.6 面临的挑战	10
1.2 云计算起源	10
1.2.1 分布式系统	11
1.2.2 虚拟化	12
1.2.3 Web 2.0	13
1.2.4 面向服务的计算	14
1.2.5 效用计算	15
1.3 构建云计算环境	16
1.3.1 应用程序开发	16
1.3.2 基础设施和系统开发	16
1.3.3 云计算平台和技术	17
本章小结	18
习题	19
第 2 章 并行计算与分布式计算原理	20
2.1 计算时代	20
2.2 并行计算与分布式计算	21
2.3 并行计算基本要素	21
2.3.1 什么是并行处理	21
2.3.2 并行处理硬件架构	22
2.3.3 并行编程方法	25

2.3.4 并行性的级别	25
2.3.5 注意事项	26
2.4 分布式计算基本要素	26
2.4.1 通用概念和定义	27
2.4.2 分布式系统组件	27
2.4.3 分布式计算架构模式	28
2.4.4 进程间通信模型	35
2.5 分布式计算技术	37
2.5.1 远程过程调用	37
2.5.2 分布式对象框架	38
2.5.3 面向服务的计算	42
本章小结	48
习题	48
第 3 章 虚拟化	50
3.1 简介	50
3.2 虚拟化环境特点	51
3.2.1 更强的安全性	52
3.2.2 执行管理	53
3.2.3 可移植性	54
3.3 虚拟化技术分类	54
3.3.1 执行虚拟化	54
3.3.2 其他类型的虚拟化	61
3.4 虚拟化和云计算	62
3.5 虚拟化的利与弊	64
3.5.1 虚拟化技术的优点	64
3.5.2 虚拟化技术的缺点	64
3.6 技术实例	66
3.6.1 Xen: 半虚拟化	66
3.6.2 VMware: 完全虚拟化	67
3.6.3 微软 Hyper-V	72
本章小结	75
习题	75
第 4 章 云计算架构	76
4.1 简介	76

4.2 云计算参考模型	76
4.2.1 架构	76
4.2.2 基础设施即服务和硬件 即服务	78
4.2.3 平台即服务	80
4.2.4 软件即服务	83
4.3 云的种类	85
4.3.1 公共云	85
4.3.2 私有云	86
4.3.3 混合云	88
4.3.4 社区云	90
4.4 云计算经济特性	91
4.5 云计算面临的挑战	93
4.5.1 云计算定义	93
4.5.2 云计算互操作性和标准	93
4.5.3 可扩展性和容错性	94
4.5.4 安全、可信和隐私	94
4.5.5 组织方面	95
本章小结	95
习题	95

第二部分 云应用编程与 Aneka 平台

第 5 章 Aneka: 云应用平台	98
5.1 框架概述	98
5.2 Aneka 容器结构	100
5.2.1 Aneka 平台基础: 平台抽象层	101
5.2.2 构造服务	101
5.2.3 基础服务	103
5.2.4 应用服务	105
5.3 构建 Aneka 云平台	106
5.3.1 基础设施组织	107
5.3.2 逻辑组织	107
5.3.3 私有云部署模式	109
5.3.4 公共云部署模式	110
5.3.5 混合云部署模式	111
5.4 云编程和云管理	112
5.4.1 Aneka SDK	112

5.4.2 管理工具	115
本章小结	116
习题	116
第 6 章 并行计算: 线程编程	117
6.1 单机并行计算简介	117
6.2 线程编程应用	118
6.2.1 什么是线程	119
6.2.2 线程 API	120
6.2.3 线程并行计算技术	121
6.3 Aneka 多线程方式	130
6.3.1 线程编程模型简介	131
6.3.2 Aneka 线程和普通线程	132
6.4 Aneka 线程编程应用	135
6.4.1 Aneka 线程应用模型	135
6.4.2 域分解: 矩阵乘法	136
6.4.3 功能分解: Sine、Cosine、 Tangent	142
本章小结	147
习题	148
第 7 章 高吞吐量计算: 任务编程	149
7.1 任务计算	149
7.1.1 任务特性	150
7.1.2 计算类别	150
7.1.3 任务计算框架	151
7.2 基于任务的应用模型	152
7.2.1 高度并行应用	153
7.2.2 参数化应用	153
7.2.3 消息传递接口应用	155
7.2.4 具有任务依赖性的 工作流应用	156
7.3 基于任务的 Aneka 编程	159
7.3.1 任务编程模型	159
7.3.2 用任务模型开发应用	160
7.3.3 开发参数化应用	174
7.3.4 管理工作流	176
本章小结	178
习题	179
第 8 章 数据密集型计算:	
MapReduce 编程	181
8.1 什么是数据密集型计算	181

8.1.1 数据密集型计算特性	182
8.1.2 未来的挑战	182
8.1.3 历史背景	183
8.2 数据密集型计算技术	186
8.2.1 存储系统	186
8.2.2 编程平台	193
8.3 Aneka MapReduce 编程	199
8.3.1 MapReduce 编程模型简介	199
8.3.2 应用实例	216
本章小结	226
习题	226

第三部分 工业云平台与新进展

第 9 章 工业云平台	228
9.1 亚马逊 Web 服务	228
9.1.1 计算服务	229
9.1.2 存储服务	232
9.1.3 通信服务	239
9.1.4 其他服务	240
9.1.5 总结	241
9.2 谷歌 AppEngine	241
9.2.1 架构和核心概念	241
9.2.2 应用程序生命周期	245
9.2.3 成本模型	247
9.2.4 结论	247
9.3 微软 Azure	248
9.3.1 Azure 核心概念	248
9.3.2 SQL Azure	252
9.3.3 Windows Azure 平台设备	253
9.3.4 结论	253
本章小结	254
习题	254
第 10 章 云应用	256
10.1 科学应用	256
10.1.1 医疗保健:	
云心电图分析	256

10.1.2 生物学:	
蛋白质结构预测	257
10.1.3 生物学: 基因表达数据用	
于癌症诊断分析	258
10.1.4 地球科学: 卫星	
图像处理	259
10.2 商业和消费应用	260
10.2.1 CRM 和 ERP	260
10.2.2 效率型应用	262
10.2.3 社交网络	265
10.2.4 媒体应用	265
10.2.5 多人在线游戏	268
本章小结	269
习题	269
第 11 章 云计算高级主题	270
11.1 云能效	270
11.2 基于市场的云管理	273
11.2.1 面向市场的云计算	273
11.2.2 MOCC 参考模型	274
11.2.3 支持 MOCC 的技术	
和实现	278
11.2.4 结论	282
11.3 云联盟和互联云	282
11.3.1 特性和定义	282
11.3.2 云联盟栈	283
11.3.3 关注点	288
11.3.4 云联盟技术	301
11.3.5 结论	304
11.4 第三方云服务	304
11.4.1 MetaCDN	304
11.4.2 SpotCloud	306
本章小结	307
习题	307
参考文献	309
索引	317

第一部分
Mastering Cloud Computing: Foundations and Applications Programming

基础

导 论

计算正在发生变革，它将转化为一种商业化服务模式，像提供水、电、煤气和电话等基础设施服务一样来交付计算服务。这种模式下，用户根据需求获得计算服务，而不需要知道该服务由哪里提供。一些计算模式（如网格计算）提出了交付这种效用计算服务的构想，近年来新兴的云计算模式则将实现了这一蓝图。

云计算是一项技术进步，它关注如何设计计算系统和开发应用程序，以及如何利用现有的服务构建软件。云计算基于动态交付概念，不仅是服务的动态交付，还包括计算能力、存储、网络、信息技术（IT）基础设施等。资源可以通过互联网并基于使用付费的方式由云计算供应商提供。今天，任何人都可以用信用卡订阅云服务，并在一定时间内为应用程序部署和配置服务，还可根据需求增加或减少配置资源，且只需要按资源的使用时间和使用量付费。

本章简要概述了云计算模式，介绍其发展前景，讨论其核心特征，并追溯其发展历程。本章还介绍了云计算的关键技术，以及关于云计算发展方向的深入见解。

1.1 云计算简介

1969 年，原高级研究计划署网络（ARPANET）的首席科学家之一，互联网奠基人 Leonard Kleinrock 说：

目前，计算机网络仍处于起步阶段，但是随着其成长和成熟，我们可能会看到计算服务的普及，像电力和电话等公共事业服务一样，跨国界服务于每个家庭和办公室。

这种计算模式的设计蓝图是基于服务的交付模式，在 21 世纪，整个计算机行业将产生巨大变革，计算服务将按需求即时交付，类似于水、电、电话、燃气等公共服务。同样，只有当用户（消费者）访问计算服务时，他们才需要向服务供应商支付费用。此外，消费者不再需要将大量投资用于构建和维护复杂的 IT 基础设施。

在这样的交付和使用模式中，用户可以按需求获得服务，而不用考虑服务提供方位于何处。这种模式曾经被称为效用计算，近期（2007 年起）改称为云计算。IT 基础设施称为“云”，企业和用户可以从任何地方按需访问各种应用服务。因此，可以认为云计算是动态交付计算服务的新模式，它拥有采用虚拟化技术构建的先进数据中心，可实现资源的整合和有效利用。

云计算允许用户租用基础设施、运行环境和服务，按需使用与付费。云计算原理得到广泛应用，不同的人对云计算有不同的认识。大型企业首席信息官和技术主管看到了云计算的机遇，按企业商业需要部署和扩展基础架构。终端用户应用云计算服务，可以从任何地方通过互联网随时访问自己的文件和数据。还有许多关于云计算的观点^①，其中，最直白的观点可

① 2007 年 5 月 Web 2.0 发布会上，Joyent 公司副总裁 Rob Boothby 组织了一系列云计算研讨，有趣的是，不同的人对云结算有不同的认识。首席执行官、技术主管、IT 公司创始人、IT 分析师都给出了各自的观点，正是从那时开始，云计算概念逐渐被人们熟知。研讨会视频发布在 YouTube 上，网址为 www.youtube.com/watch?v=6PNuQHUiV3Q。

以概括如下:

我不在乎服务器在哪儿、谁管理它们、文档存储在哪儿,或者应用程序托管在哪儿。我只是希望它们始终可用,并且可以通过连接到互联网的任何设备访问它们。我愿意为我需要的服务付费。

这一描述与我们使用其他服务的方式非常相似,例如水、电等。换句话说,云计算已经由 IT 服务转变为公共事业服务。通过多种技术的有效融合,云服务模式已经可以实现,而且已经相对成熟。Web 2.0 技术在应用云计算构建计算系统中发挥了核心作用。云计算可以把互联网转化成一个丰富的、足以满足各种复杂需求的应用和服务的交付平台。面向服务架构将计算资源抽象为服务,使得云计算能提供计算服务能力;而虚拟化赋予了云计算用于构建生产和企业系统时必要的可定制化特性、可控性和灵活性。

除了能灵活地构建新的系统和应用之外,云计算环境还能对现有系统进行扩展。动态配置 IT 资源比购买额外的硬件和软件更有吸引力,这些硬件和软件资源的需求量很难估计并且受限于时间因素。这是云计算最重要的优势之一,也使其成为了技术热点。随着云计算系统的广泛部署,支持云计算的基础技术与系统逐步完善和标准化,这是实现云计算长远目标的基础。云计算提供了一个开放的环境,计算、存储和其他服务都将作为计算服务进行交易。

4

1.1.1 云计算构想

云计算为任何一个拥有信用卡的人提供虚拟硬件、运行环境和服务,用户需要时就可以使用,不需要事先委托。计算系统总体架构被转换成各种计算服务的集合,这些服务被配置和组合在一起以便快速部署系统,以小时为单位而不是以天为单位,并且没有维护成本。这种模式起初遭到怀疑,而现在已经在多个领域和行业中得到实际应用(见图 1-1)。需求紧紧跟随着技术的发展而发展,丰富了云计算所提供的服务集合,使云服务变得更加成熟,成本也更低。

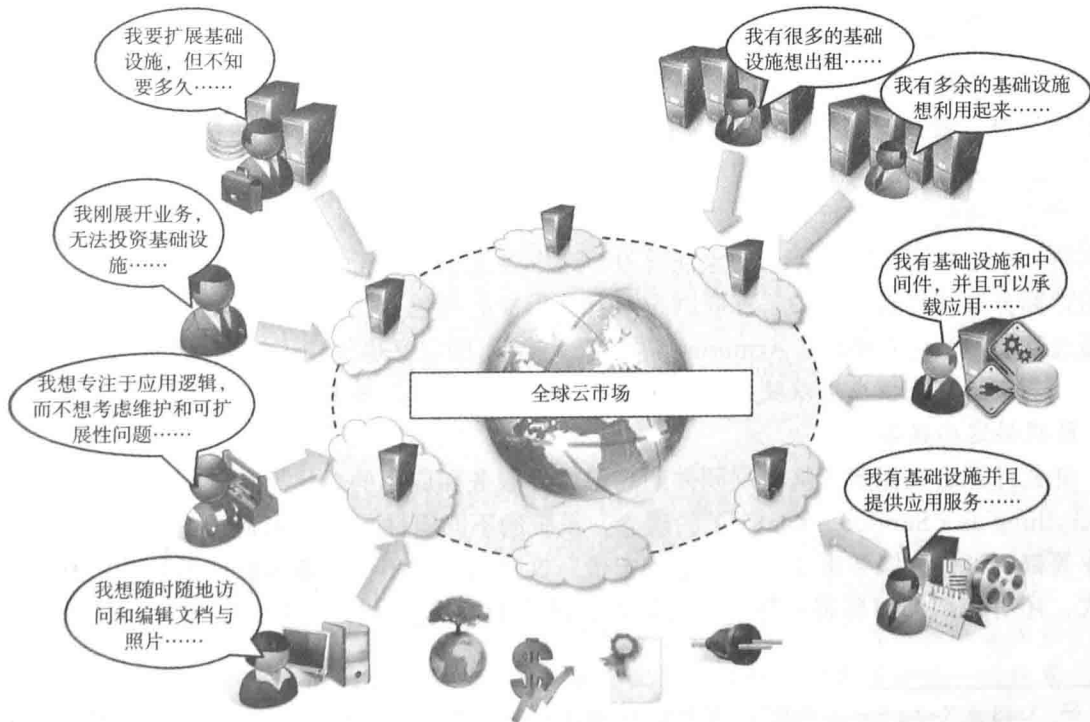


图 1-1 云计算构想

尽管技术在不断演变,然而云计算的使用往往仅限于在某一时刻提供单一服务,更常见的是,一组相关服务由同一供应商提供。缺乏有效的标准化措施,使得服务从一个供应商迁移到另一个供应商变得非常困难。云计算的长期构想是将 IT 服务作为公用基础设施服务,在没有技术和法律限制的开放市场中进行交易。在这样的云市场环境中,云服务供应商和消费者把云服务作为公共基础设施服务进行交易,发挥其核心作用。

一些实现云计算模式的技术已经存在,可以从不同的角度利用云的各种服务功能。对于无处不在的存储和按需计算能力的需求是应用云计算的最常见场景。对于应用和系统开发者而言,在不具有计算资源或者没有能力扩展现有资源的情况下,一个可扩展的应用程序运行系统是非常有吸引力的选择。终端用户更愿意基于 Web 方式访问文件,并采用成熟的应用软件处理文件。

在这种情况下,服务的发现主要通过人为干预完成:人们浏览互联网,以确定能够满足其需求的服务。可以想象,未来只需在提供云计算服务的全球化数字市场中输入我们所需要的服务,就可以自动匹配服务。这种云服务市场能自动发现服务,并集成现有的软件系统,从而使用户能够在应用系统中透明地利用云资源。云服务交易的全球化平台使服务供应商更容易被消费者发现,潜在地增加了收益,也消除了消费者和供应商之间的界限,即不再只能拥有这两种身份中的一种。例如,为了履行对客户的承诺,一个云服务供应商可能会成为其竞争供应商的消费者。

可以通过制定描述云服务的统一标准,以及不同云技术之间的交互技术标准,来建立全球云计算交易平台。已有大量应用向云计算转移,并迅速普及。此外,将云计算的核心功能应用于大型数据中心,可以减少或消除消费者对于基础设施的需求,优化数据中心资源,充分利用其为多用户提供共享服务的能力。这种模式将减少能源浪费和碳排放,从而一方面促进了绿色 IT 的实现,另一方面提高了收益。

1.1.2 云计算定义

云计算已经成为流行术语,广泛用于指代各种技术、服务和概念。它往往与硬件虚拟化或按需计算、效用计算、IT 外包、平台和软件即服务,以及其他 IT 热门技术有关。图 1-2 描述了已有云计算定义中的不同概念。

术语云曾用于电信行业,作为系统结构图的网络抽象表示。之后,它成为最流行的计算机网络——互联网(Internet)的象征符号。这个含义也适用于云计算,指一种以互联网为中心的计算方式。在云计算中,互联网扮演着基础性的角色,代表介质或平台,众多云计算服务通过互联网交付和访问。Armbrust 等人[28]给出的定义也体现了这方面的含义:

云计算是指应用以服务形式通过互联网交付使用,并且数据中心的硬件和软件能提供这些服务。

此云计算定义涵盖了从底层硬件到高层软件服务和应用的总体架构。产生了一切皆服务(Everything as a Service, XaaS^①)的理念,系统的不同组件——IT 基础设施、开发平台、数据库等都可以作为服务来交付、定量和定价。这种新模式不仅显著地影响着我们构建软件的方式,还对我们如何部署软件、使用软件、设计 IT 系统结构产生影响,甚至决定着公司投

① XaaS 是 X-as-a-Service 的缩写,其中 X 可以被以下任一字符代替: S (软件), P (平台), I (基础设施), H (硬件), D (数据库), 等等。