

HZ BOOKS
华章教育

计 算 机 科 学 丛 书

云计算系统 与人工智能应用

[美] 黄铠 (Kai Hwang) 著 袁志勇 杜瑞颖 张立强 伍春香 蔡朝晖 译
香港中文大学 (深圳) 武汉大学

Cloud Computing for Machine Learning
and Cognitive Applications

KAI HWANG

CLOUD COMPUTING FOR
MACHINE LEARNING AND
COGNITIVE APPLICATIONS

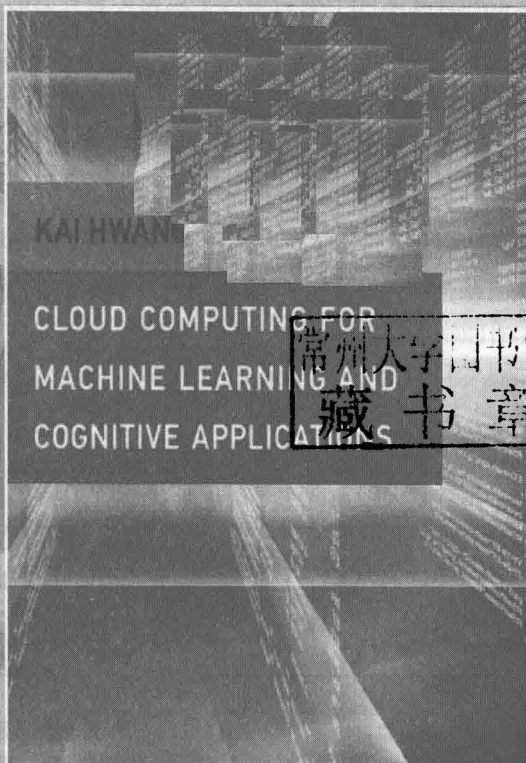
机械工业出版社
China Machine Press

计 算 机 科 学 丛 书

云计算系统 与人工智能应用

[美] 黄铠 (Kai Hwang) 著 袁志勇 杜瑞颖 张立强 伍春香 蔡朝晖 译
香港中文大学 (深圳) 武汉大学

Cloud Computing for Machine Learning
and Cognitive Applications



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

云计算系统与人工智能应用 / (美) 黄铠 (Kai Hwang) 著; 袁志勇等译. —北京: 机械工业出版社, 2018.5

(计算机科学丛书)

书名原文: Cloud Computing for Machine Learning and Cognitive Applications

ISBN 978-7-111-59883-1

I. 云… II. ①黄… ②袁… III. ①云计算 - 研究 ②人工智能 - 研究 IV. ① TP393.027
② TP18

中国版本图书馆 CIP 数据核字 (2018) 第 081967 号

本书版权登记号: 图字 01-2017-8949

Authorized translation from the English language edition entitled Cloud Computing for Machine Learning and Cognitive Applications (ISBN: 9780262036412) by Kai Hwang, Copyright by Kai Hwang.

Chinese simplified language edition published by China Machine Press.

Copyright by China Machine Press.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanic, including photocopying, recording, or by any information storage retrieval system, without the prior permission of the publisher.

本书中文简体字版由 Kai Hwang 授权机械工业出版社独家出版。未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分。

本书关注云计算、大数据、物联网、认知计算、机器学习的基本原理、智能应用和编程实战, 包括 AWS、微软 Azure 云、谷歌 DeepMind 和寒武纪神经芯片等大量案例。全书共四个部分, 第一部分介绍云计算、数据科学和自适应计算的基本原理, 第二部分涵盖云架构、虚拟机、Docker 容器和多云混搭服务等, 第三部分讲解机器学习、深度学习、类脑计算机和 AR/VR 等的原理, 第四部分讨论 MapReduce、Hadoop、Spark、TensorFlow 和 GraphX 云编程。

本书适合作为高等院校计算机相关专业高年级本科生和研究生的教材, 也适合互联网、物联网等领域的专业技术人员参考。

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: 殷 虹

印 刷: 北京瑞德印刷有限公司

版 次: 2018 年 5 月第 1 版第 1 次印刷

开 本: 185mm × 260mm 1/16

印 张: 27.25

书 号: ISBN 978-7-111-59883-1

定 价: 99.00 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: hzsj@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，不仅肇划了研究的范畴，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, Afred 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

译者序

Cloud Computing for Machine Learning and Cognitive Applications

本书根据黄铠教授在美国南加州大学完成的英文原著《Cloud Computing for Machine Learning and Cognitive Applications》翻译而成，英文原版由美国麻省理工学院出版社（The MIT Press）于2017年6月出版发行。本书适合作为高等院校计算机、电子信息等相关专业的教材，也适合作为专业技术人员的参考书籍。

本书由武汉大学计算机学院五位教学科研一线的教师翻译。其中，蔡朝晖副教授负责翻译第1章，伍春香副教授负责翻译第2章、第5章，张立强副教授负责翻译第3章、第4章、第8章，袁志勇教授负责翻译第6章、第7章、第9章和前言，杜瑞颖教授负责翻译第10章。杜瑞颖教授和蔡朝晖副教授共同负责翻译索引。最后由袁志勇教授和杜瑞颖教授对全书进行统稿。

本书中文版能够在国内出版，机械工业出版社华章公司做了大量工作，译者在此表示衷心感谢！

由于时间及水平所限，书中译文不当之处，恳请学术界同仁及广大读者批评指正。

译者

2018年3月于武汉

以“云大物移智”为标志的新技术革命开启了科技的新时代，我们看到了云计算、大数据、物联网、移动互联网与人工智能等新兴领域的蓬勃发展。这对计算机科学与信息工程专业人才的知识结构与创新能力提出了新的要求。在这样的智能时代，计算机与人工智能学科的教育及人才培养需要结合数理基础知识、算法与软件知识以及云平台核心知识与能力，特别是在领域应用与知识交叉方面需要新的教材，这也是出版本书的主要目的。

本书英文版在美国南加州大学完成，中文版的翻译由武汉大学计算机学院的五位老师承担：袁志勇，杜瑞颖，张立强，伍春香，蔡朝晖。全书的翻译与校对工作在袁志勇教授和杜瑞颖教授的协调下完成。我在此对他们专业的学术工作与敬业精神表示由衷的感谢。

全书共 10 章，强调云计算、大数据、物联网、认知计算、机器学习的基本原理与智能应用。本书可作为高等院校与研究院的基础教材，专业领域跨越计算机科学、人工智能、机器学习与大数据。本书之前的版本《云计算与分布式系统》曾作为美国南加州大学、清华大学与武汉大学的教材。我们为授课教师提供配套的课件与习题解答，请访问华章网站（www.hzbook.com）下载教辅资料。

计算机、互联网与人工智能的突飞猛进

自新旧千年交替以来，计算机和信息技术在规模、设备以及平台方面都发生了重大变化。全球数以千计的数据中心正在转化为云端，使数以亿计的个人、企业和政府用户受益。30 亿部智能手机正在用于与社交网络云进行互动。所有这些都大大改变了人类的活动和交往。物联网（IoT）和机器智能正在重塑我们的生活方式。这些信息技术的进展正在把我们的社会转变为由许多人工智能（AI）和自动化认知解决方案所支撑的规模经济。事实上，我们正在进入一个拥有云数据分析、智能机器人、机器学习和认知服务的时代。

为实现新的计算和通信模式，我们必须用新功能提升云计算生态系统，这些新功能涵盖机器学习、物联网感知、数据分析以及能模拟或增强人类智能的认知能力。最终目标是建立大数据产业，提供认知服务，在更高的效率下处理劳动密集型任务，从而弥补人类在这方面的缺陷。我们必须设计成功的云系统、网络服务和数据中心，用于存储、处理、学习和分析大数据，以发现新知识或做出重要决策。这些目标可通过硬件虚拟化、机器学习、经训练的深度学习、神经形态计算机架构以及认知服务来实现。例如，新的云服务可以包括学习即服务（LaaS）、分析即服务（AaaS）或安全即服务（SaaS）等。

如今，IT 公司、大型企业、大学和政府正逐渐将其数据中心转移到云设施中，以支持移动和网络应用。拥有类似于云的集群架构的超级计算机也正在转型，以处理大数据集或数据流。智能云对于支持社交、媒体、移动端、商业和政府运营的需求量极大。谷歌、亚马逊、微软、脸书、苹果、百度、阿里巴巴、腾讯和 IBM 等公司正在争相开发物联网设备、智能机器人、自动驾驶汽车、飞行汽车和认知系统。高科技产业正在进入一个挑战与机遇并存的新世界。

本书速览与导读

本书旨在帮助高年级本科生或研究生掌握现代云系统架构、机器学习算法、并行和分布式编程以及用于大数据挖掘、预测分析和认知服务应用的软件工具，并推动大数据和机器智能时代的职业发展和业务转型。它也可以作为专业人士、科学家或工程师的参考书。本书是我过去二十年的研究、教学和授课经验的结晶，读者可利用本书学习云和机器学习方面的技能。本书共四个部分，包含 10 章，现简要介绍如下。

- 第一部分有两章，介绍大数据应用中的云计算、数据科学和自适应计算的基本原理。这两章为后续八章提供了必要的理论基础和技术基础。
- 第二部分有三章，涵盖云架构、虚拟机、Docker 容器、移动云、物联网和多云混搭服务，案例研究包括 AWS、谷歌云、微软 Azure 云、IBM 智能云、Salesforce 云、SGI Cyclone、苹果 iCloud、NASA Nebula 云及 CERN 云。
- 第三部分有两章，主要介绍机器学习、深度学习、人工智能机器、智能机器人、神经形态处理器、类脑计算机、增强现实（AR）和虚拟现实（VR）的原理，涵盖谷歌脑计划、DeepMind、X-Lab 计划、IBM SyNapse 神经形态芯片、Bluemix 云和认知计划以及中国的寒武纪神经芯片。
- 第四部分有三章，介绍关于 MapReduce、Hadoop、Spark、TensorFlow 和 GraphX 的云编程范例、软件工具以及应用开发，最后一章专门讨论云性能、隐私和安全问题。所有云系统、编程范例、机器学习方法和软件工具在书中都有具体的应用示例。

将人工智能赋予云端和物联网平台

本书将大数据理论与智能云中的新兴技术相结合，并利用新的应用探索分布式数据中心。如今，信息物理系统（CPS）出现在智慧城市、自动驾驶、情感检测机器人、送货无人机、虚拟现实、增强现实以及认知服务中。为促进智能云或数据中心的有效大数据计算，本书采用技术融合方式将大数据理论与云设计原理及超级计算标准相结合。物联网感知技术可实现大规模的数据采集和筛选。在云端或物联网平台上，机器学习和数据分析有助于智能决策，自动且无人干预。

拥有 AI 特性的增强云和超级计算机是我们的根本目标。这些人工智能和机器学习任务在实际中的应用由 Hadoop、Spark 和 TensorFlow 程序库支持。数据分析师、认知科学家和计算机专业人士须共同努力解决实际问题。这种协作学习必须包含云端、移动设备、数据中心和物联网资源，最终目标是发现新知识或者做出重要决策。多年以来，我们一直在探索如何建造类脑计算机，它能在感知、记忆、识别和理解中模仿或增强人类功能。

当今，谷歌、IBM、微软、脸书、中国科学院以及百度都在探索云计算、机器学习和物联网应用中的人工智能。本书涵盖一些新的神经形态芯片和领先的研究中心构建的软件平台，它使认知计算成为可能。本书考察了在硬件、软件和生态系统方面的进步，不仅注重机器学习技术，包括模式识别、语音/图像理解以及低成本、低功耗需求的语言翻译和理解，而且强化了用手机、机器人、物联网平台、数据中心以及云计算构建未来网络空间的新方法。

读者对象和教师指南

本书是为满足日益增长的计算机科学和电气工程教育课程的需求而编写的。教师可选择

与课程相适应的不同章节进行讲授，从而满足各层次学生的需求。本书适合在高年级本科生和研究生的教学中使用，同时，也能帮助那些希望增进技能以迎接新的 IT 技术挑战的计算机专业人士。

讲授云计算课程应至少覆盖 8 章。如果教学课时有限，可跳过第 2 章和第 7 章。对于机器学习课程，可跳过第 3 章或第 10 章。对于高年级本科生课程，讲授 7 章（1, 2, 3, 4, 5, 6, 8）足够。任何课程都要覆盖云计算和机器学习两个主题，此外，本书也适合作为大数据科学、物联网应用和分布式计算课程的参考书。

计算机、互联网等高科技行业的工程师、科学家及管理者也能从本书中找到有益的参考。例如：英特尔工程师可能会发现本书有助于推动物联网的概念；亚马逊、谷歌、微软、Apache、Databricks 和 Cloudera 的软件工程师及大数据分析师可能会发现它在日常工作中有用；谷歌脑计划、X-Lab 计划和 DeepMind 的开发人员可在深度学习服务中使用书中知识进行语音、文字、图像理解以及医疗保健和自动驾驶开发；在新兴认知行业中，IBM 已确定将物联网和认知服务用于扩大其在社会和政府部门的业务；脸书、苹果和百度可能希望利用 AR 和 VR 技术探索新的 AI 特性、社交服务与个性化娱乐；最后，亚马逊和阿里巴巴云的买卖双方可能希望在电子商务和 P2P 社交服务之外拓展网络购物及交易业务。

黄铨

2018 年 5 月于深圳

作者简介

Cloud Computing for Machine Learning and Cognitive Applications



黄铠教授是计算机系统和互联网技术领域的国际知名资深学者。目前，他是香港中文大学（深圳）校长讲座教授，兼任中国科学院云计算中心首席科学家。他拥有加州大学伯克利分校的博士学位，主要研究领域为计算机体系结构、并行与分布式处理、云计算、大数据、物联网、移动互联网、网络安全与人工智能应用等方面，目前主要关注大数据在医疗保健、智慧城市与移动社交网络方面的应用。

加盟香港中文大学与中国科学院之前，他是美国南加州大学（USC）电子工程与计算机科学系的终身教授，现在已从该校退休。他曾在普渡大学任教多年，并先后在清华大学、香港大学、台湾大学和浙江大学担任特聘讲座教授。他在专业领域发表了 260 篇科学论文，截至 2018 年 4 月在谷歌学术搜索中被引次数超过 17400 次，h 指数为 55。他是 IEEE 计算机协会的终身会士。他于 2012 年获得国际云计算大会（IEEE CloudCom）终身成就奖，2004 年获得中国计算机学会（CCF）首届海外杰出贡献奖。

黄教授创作或合著了 10 余本英文学术专著，被翻译为五国语言。其中，有 5 本被翻译为中文，包括清华大学出版社出版的《高等计算机系统结构》（1995），以及机械工业出版社出版的《可扩展并行计算》（2000）、《云计算与分布式系统》（2013）与《认知计算与深度学习：基于物联网云平台的智能应用》（2018）等。这些书与本书是配套的关系。此外，他曾担任《并行与分布式计算》（JPDC）杂志主编 28 年，还曾担任 IEEE《云计算会刊》（TCC）、《并行和分布式系统》（TPDS）、《服务计算》（TSC）以及《大数据智能》杂志的编委。

多年来，黄教授在南加州大学和普渡大学共培养博士生 21 人，其中 4 人晋升为 IEEE 会士，1 人为 IBM 会士。他在 IEEE 与 ACM 国际会议和全球领先的大学进行过 60 多次主题演讲和杰出讲座。他曾在 IBM 研究院、Intel 公司、富士通研究院、麻省理工学院林肯实验室、加州理工学院喷气推进实验室（JPL）、台湾工业技术研究院（ITRI）、法国国家计算科学研究中心（ENRIA）和中国科学院计算所担任高级顾问或首席科学家。

出版者的话

译者序

前言

作者简介

第一部分 云平台、大数据与认知计算

第 1 章 云计算系统原理 2

1.1 可扩展计算的弹性云系统 2

1.1.1 云计算的驱动技术 2

1.1.2 可扩展的分布式 / 并行计算的
演化 3

1.1.3 云系统中的虚拟资源 5

1.1.4 云计算与本地计算 7

1.2 云平台架构与分布式系统的比较 8

1.2.1 基本云平台的架构 8

1.2.2 公共云、私有云、社区云和
混合云 10

1.2.3 物理集群与虚拟集群 12

1.2.4 云与传统并行 / 分布式系统的
比较 141.3 云服务模型、生态系统与可扩展性
分析 161.3.1 云服务模型: IaaS、PaaS 和
SaaS 16

1.3.2 云性能分析与可扩展性定理 18

1.3.3 云生态系统与用户环境 20

1.3.4 云计算的技术成熟度曲线 23

1.3.5 云计算与其他技术的关系 24

1.4 集群的可用性、移动性和优化 26

1.4.1 云服务器集群的可用性分析 26

1.4.2 虚拟集群操作中的容错 28

1.4.3 云中多服务器集群的排队
模型 29

1.4.4 云计算的多服务器集群优化 30

1.5 结论 32

习题 33

参考文献 35

第 2 章 数据分析、物联网与
认知计算 37

2.1 大数据科学与应用面临的挑战 37

2.1.1 数据科学与大数据特征 37

2.1.2 物联网的技术成熟度曲线 38

2.1.3 走向大数据工业 40

2.1.4 大数据应用概述 41

2.2 物联网与云平台的互动模式 44

2.2.1 IoT 感知与平台架构 45

2.2.2 IoT 价值链与发展路线 47

2.2.3 独立使用和与云平台结合的
IoT 应用 49

2.2.4 智慧城市与智慧社区的发展 52

2.3 在云平台上的数据收集、挖掘与
分析 54

2.3.1 数据质量控制与表达 54

2.3.2 数据挖掘与分析 57

2.3.3 在云平台上提升数据分析
能力 59

2.3.4 支撑大数据分析的云资源 61

2.4 神经形态硬件与认知计算 63

2.4.1 认知计算与神经形态处理器 63

2.4.2 IBM SyNAPSE 与相关的神经类
脑计算机项目 662.4.3 中国科学院的神经处理器
Cambricon 682.4.4 Google TPU 与相关的人工智能
应用 68

2.5 结论 70

习题	71	习题	105
参考文献	72	参考文献	107
第二部分 云系统架构与服务平台设计		第 4 章 云架构与服务平台设计	
第 3 章 虚拟机、Docker 容器和服务器集群		4.1 云架构与基础设施设计	
3.1 云计算系统中的虚拟化	74	4.1.1 公共云平台及其服务项目	109
3.1.1 虚拟化的基本概念	74	4.1.2 云服务的商业模型	111
3.1.2 虚拟化的实现层级	75	4.1.3 数据中心到云平台的转换	113
3.1.3 集群或云系统中的资源虚拟化	78	4.1.4 资源弹性配置方法	117
3.2 用于创建虚拟机的虚拟机监控器	79	4.2 虚拟集群的动态配置	119
3.2.1 虚拟机架构类型	79	4.2.1 虚拟集群配置项目	119
3.2.2 完全虚拟化和托管虚拟化	81	4.2.2 虚拟集群配置自适应	121
3.2.3 修改客户操作系统的半虚拟化	83	4.2.3 数据中心集群的虚拟化支持	122
3.2.4 平台虚拟化软件产品与工具包比较	85	4.2.4 VMware vSphere 6: 商用云操作系统	123
3.3 Docker 引擎和应用程序容器	86	4.3 AWS 云及其服务项目	124
3.4 Docker 容器和部署要求	89	4.3.1 三大云架构与服务融合	124
3.4.1 使用 Linux 内核函数创建的 Docker 容器	89	4.3.2 AWS EC2 计算引擎和 S3 存储云服务	127
3.4.2 虚拟机与 Docker 容器的比较	91	4.3.3 其他 AWS 云服务产品	129
3.4.3 从虚拟机到容器和 unikernel 的架构演变	92	4.4 Google AppEngine 与 Microsoft Azure	133
3.5 虚拟机管理和容器编排	94	4.4.1 Google AppEngine 及其计算引擎	133
3.5.1 虚拟机管理解决方案	94	4.4.2 Google 硬件/软件支持下的机器学习服务	137
3.5.2 用于灾难恢复的虚拟机迁移	95	4.4.3 Microsoft Azure 云及其服务项目	138
3.5.3 Docker 容器调度和编排	98	4.5 Salesforce、IBM SmartCloud 及其他云平台	141
3.6 Eucalyptus、OpenStack 和 VMware 云构建	100	4.5.1 用于 SaaS 服务的 Salesforce 云平台	141
3.6.1 私有云中的 Eucalyptus 虚拟集群	100	4.5.2 IBM SmartCloud 云平台、物联网及认知项目	143
3.6.2 用于构建私有云或公共云的 OpenStack 软件	102	4.5.3 SGI、NASA 和 CERN 建立的云平台	145
3.6.3 支持构建混合云的 VMware 虚拟化	103	4.6 结论	148
3.7 结论	105	习题	149
		参考文献	150

第5章 移动云、物联网、社交媒体与混搭云服务

5.1 无线互联网与移动云计算	153
5.1.1 移动设备与边际互联子网	153
5.1.2 WiFi、蓝牙和无线传感器网络	155
5.1.3 移动云计算的微云网	156
5.1.4 移动云与托管云	158
5.2 物联网感知以及与云的交互	160
5.2.1 本地与全球定位系统	161
5.2.2 构建移动云的无线接入网	162
5.2.3 物联网和云的互动框架与设备	164
5.3 社交媒体应用中的云计算	167
5.3.1 社交媒体大数据工业应用	167
5.3.2 社交网络与应用编程接口	170
5.3.3 社交图的特性与表示	172
5.3.4 智慧云的社交图分析	175
5.4 多云混搭架构与服务	176
5.4.1 混搭云的敏捷性与可扩展性	177
5.4.2 混搭云服务架构	179
5.4.3 混搭云服务的 Skyline 发现	182
5.4.4 混搭云服务的动态组成	184
5.5 结论	185
习题	185
参考文献	187

第三部分 机器学习原理与人工智能机器

第6章 机器学习算法与预测模型拟合

6.1 机器学习方法的分类	190
6.1.1 机器学习算法的种类	190
6.1.2 监督式机器学习算法	192
6.1.3 无监督机器学习算法	193
6.2 监督式回归与分类方法	193
6.2.1 预测的线性回归方法	194
6.2.2 机器学习的决策树	199

6.2.3 贝叶斯分类与训练样本	202
6.2.4 支持向量机	205
6.3 聚类方法与降维方法	207
6.3.1 聚类分析与 k 均值聚类	207
6.3.2 降维方法与强化学习	210
6.3.3 主成分分析方法	212
6.3.4 半监督学习方法	214
6.4 机器学习的建模方法	216
6.4.1 模型拟合的性能指标	216
6.4.2 降低模型过拟合的方法	217
6.4.3 避免欠拟合的方法	219
6.4.4 选择机器学习算法	219
6.5 结论	221
习题	221
参考文献	224

第7章 智能机器与深度学习网络

7.1 人工智能机器的发展	225
7.1.1 智能机器的技术成熟度曲线	225
7.1.2 谷歌人工智能产品及服务进展	226
7.1.3 IBM 与其他公司的认知服务	230
7.1.4 Intel、Nvidia 和 CAS/ICT 的深度学习芯片	231
7.2 增强现实、虚拟现实与区块链技术	233
7.2.1 增强、介导与虚拟现实	234
7.2.2 虚拟现实与相关产品	235
7.2.3 区块链在商业交易安全中的应用	236
7.3 深度学习人工神经网络	237
7.3.1 深度学习模仿人类认知功能	237
7.3.2 ANN 的演进和应用	239
7.3.3 人工神经元的数学描述	240
7.3.4 多层 ANN	241
7.3.5 ANN 的正向传播与反向传播	244

7.4 深度学习网络的分类	247
7.4.1 深度学习网络的类型	247
7.4.2 卷积神经网络	248
7.4.3 深度神经网络的连接性能	252
7.4.4 循环神经网络	252
7.5 深度学习应用与其他认知功能	255
7.5.1 受限玻尔兹曼机	255
7.5.2 深度信念网络	256
7.5.3 深度学习与脑功能开发	257
7.6 结论	258
习题	258
参考文献	261

第四部分 云编程、性能提升与 数据安全

第8章 基于Hadoop和Spark的 云计算

8.1 大型集群的可扩展并行计算	264
8.1.1 可扩展计算的特点	264
8.1.2 从MapReduce到Hadoop和Spark	265
8.1.3 大数据处理的应用软件库	266
8.2 Hadoop及其HDFS和YARN	267
8.2.1 MapReduce云计算引擎	267
8.2.2 用于并行矩阵乘法的MapReduce	271
8.2.3 Hadoop架构及其最新扩展	273
8.2.4 Hadoop分布式文件系统	276
8.2.5 Hadoop YARN资源管理系统	279
8.3 Spark核心组件和弹性分布式数据集	280
8.3.1 Spark通用核心组件	280
8.3.2 弹性分布式数据集	282
8.3.3 用于DAG的RDD	284
8.4 Spark SQL和流编程	287
8.4.1 具有结构化数据的Spark SQL	287

8.4.2 使用实时数据流的Spark Streaming	288
8.4.3 Spark Streaming应用示例	290
8.5 用于机器学习的Spark MLlib和用于图像处理的GraphX	291
8.5.1 用于机器学习的Spark MLlib库	292
8.5.2 MLlib应用示例	293
8.5.3 用于图像处理的Spark GraphX	294
8.5.4 GraphX编程示例	295
8.6 结论	299
习题	300
参考文献	304

第9章 TensorFlow、Keras、DeepMind和图分析

9.1 神经网络计算平台TensorFlow	306
9.1.1 TensorFlow的关键概念	306
9.1.2 张量、变量、输入和提取操作	309
9.1.3 分布式TensorFlow执行环境	310
9.1.4 TensorFlow程序的执行会话	313
9.2 用于深度学习的TensorFlow系统	315
9.2.1 分层TensorFlow系统架构	315
9.2.2 TensorFlow在不同主机上的安装	317
9.2.3 分布式资源共享的TensorFlow生态系统	319
9.2.4 TensorFlow用于手写数字识别	320
9.2.5 TensorFlow用于认知服务	323
9.3 Google DeepMind及其他AI计划	326
9.3.1 强化深度学习算法	326
9.3.2 政策网络与价值网络的互动	328

9.3.3 AlphaGo 计划中的强化学习	330	策略	347
9.3.4 英国 DeepMind Health 项目	332	10.2.2 云性能指标	350
9.4 预测软件库、Keras 库、DIGITS 库和图库	333	10.2.3 雷达图表示的云性能模型	353
9.4.1 用于认知应用的预测软件库	333	10.3 云基准结果的性能分析	357
9.4.2 用于深度学习的 Keras 库和 DIGITS 5	335	10.3.1 可扩展云性能的弹性分析	357
9.4.3 云中的图并行计算	338	10.3.2 横向扩展、放大和混合扩展性能	358
9.4.4 社交网络中的社区检测	339	10.3.3 扩展策略的优缺点	360
9.5 结论	342	10.4 云安全和数据隐私保护	362
习题	342	10.4.1 云安全和隐私问题	362
参考文献	343	10.4.2 云安全基础设施	364
第 10 章 云性能、安全和数据隐私	344	10.4.3 移动云和安全威胁	369
10.1 引言	344	10.5 云和数据中心的信任管理	370
10.1.1 什么是云性能和 QoS	344	10.5.1 分布式入侵和异常检测	370
10.1.2 如何保护云和共享数据	345	10.5.2 云中基于信誉的信任管理	371
10.2 云性能指标和基准	346	10.5.3 多个数据中心的 P2P 信任覆盖网络	374
10.2.1 自动扩展、横向扩展和放大		10.6 结论	377
		习题	377
		参考文献	379
		索引	381

云平台、大数据与认知计算

本部分包括两章，介绍云计算、数据科学和认知计算在大数据应用领域中的基本原理。第1章介绍云计算原理，第2章介绍支持大数据科学技术的必要背景。我们尝试将三个相互支持的领域——云计算、认知服务和大数据科学——进行整合。

这两章为阅读本书后续内容奠定了必要的理论基础和技术基础。第二部分将介绍虚拟化、云架构、移动云、社交媒体以及混合服务。第三部分内容涉及机器学习以及数据分析原理与算法。第四部分探讨大数据编程支持和云性能提升。

第1章 云计算系统原理

摘要：这一章致力于建立智能云和高效云的必要基础知识，这些云在编程能力和生产力方面具备用户友好特性。我们将介绍基础云模型和通用构架，特别是解决采用弹性云时的可扩展性、可用性、移动性和性能问题。这里指定了基本的云服务模型，第2章将给出用于支撑云端大数据计算的专用架构。后续章节将详细介绍这些硬件、软件和网络技术。

第2章 数据分析、物联网与认知计算

摘要：这一章将介绍数据科学和云分析，同时涵盖物联网和认知计算的相关问题。我们首先回顾大数据的特征，然后学习数据挖掘和机器学习技术的基础知识。接下来介绍关于大数据分析的基础构架。最后，我们给出认知计算和神经信息学的关键概念，并将IBM、谷歌以及中科院的神经形态处理器、类脑计算机和认知研究项目作为示例进行综述。

云计算系统原理

1.1 可扩展计算的弹性云系统

多年以来，传统的计算机系统都强调高性能计算（High-Performance Computing, HPC）系统中批处理的原生速度性能。现阶段每天有数十亿人使用互联网，因此，针对网络计算的新需求需要高吞吐量计算（High-Throughput Computing, HTC）系统，这些系统是用并行和分布式计算技术构建的。这种需求已经引发了许多数据中心升级到可以同时为数百万用户提供服务的互联网云。本章重点介绍如何利用低成本服务器、分布式存储系统和高带宽网络构建 HTC 云，以便在 Web、云和物联网（Internet of Things, IoT）服务中推进大数据计算。主要内容包括云系统设计原理、硬件和软件基础设施以及虚拟化资源管理。

1.1.1 云计算的驱动技术

云计算背后的关键驱动力是宽带和无线网络的普及、存储成本的下降以及互联网计算软件的逐步改进。云端用户能够在高峰时段请求更多容量、降低成本、尝试新服务，并消除不必要的容量占有，而服务供应商可以通过复用、虚拟化和动态资源配置来提高系统利用率。

云计算的概念从集群、网格和效用计算发展而来，集群和网格计算需要并行利用大量计算机。效用计算和软件即服务（Software as a Service, SaaS）提供计算资源。云计算利用动态资源为终端用户提供大量服务，它将用户解放出来，通过将作业外包给云供应商，使用户专注于应用程序的开发。

关键技术

硬件、软件和网络技术的不断改进使云得以实现（见表 1.1）。这些技术在使云计算成为现实方面发挥了重要作用。今天，这些技术大多数已经足够成熟，可以满足日益增长的需求。在硬件领域，多核 CPU、内存芯片和磁盘阵列的快速发展，使得建立拥有巨大的存储空间且更快的数据中心成为可能。资源虚拟化使得用 HTC 快速部署云和灾难恢复成为可能。

表 1.1 硬件、软件和网络中的关键技术

技术	要求和好处
快速平台配置	快速、有效和灵活的云资源配置，以为用户提供动态计算环境
按需的虚拟集群	满足用户需求的预分配的虚拟化虚拟机集群，以及根据负载变化重新配置的虚拟集群
多租户技术	用于分布式软件的 SaaS，可以满足大量用户的同时使用和所需的资源共享
海量数据处理	互联网搜索和 Web 服务通常都需要进行海量数据处理，特别是要支持个性化服务
Web 规模通信	支持电子商务、远程教育、远程医疗、社交网络、电子政务和数字娱乐等
分布式存储	个人记录和公共档案信息的大规模存储，要求云上的分布式存储
授权和计费服务	许可证管理和计费服务有益于效用计算中的各类云计算

SaaS 的供应、Web 2.0 标准和互联网性能的进步都促进了云服务的涌现。今天，云应该

能在巨大数据量之上满足众多租用者的需求。大规模分布式存储系统的可用性是当今数据中心的基础。当然，在许可证管理和自动计费技术方面取得的进步也使云计算受益匪浅。

企业内部的私有云更容易得到保护，因此比公共云更可靠。一旦私有云技术变得更加成熟和安全，就可以开放或转换成公共云。因此，未来公共云和私有云之间的界限可能会变得模糊，未来大多数云可能是混合的。

技术的融合

云计算依托以下四个方面技术的融合（如图 1.1 所示）：（1）硬件虚拟化和多核芯片使得在云中的动态配置成为可能；（2）效用和网格计算技术为云计算奠定了必要的基础；（3）最近，面向服务的架构（Service Oriented Architecture, SoA）、Web 2.0 和混搭平台的进步推动了云的进一步发展；（4）自主计算和自动化的数据中心运作使云计算成为可能。

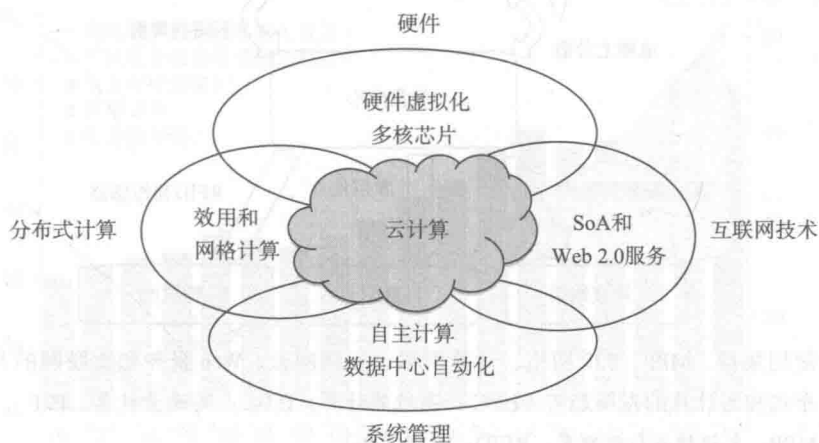


图 1.1 技术融合促成互联网的云计算

云计算探讨了多核和并行计算技术。要实现数据密集型系统的构想，需要集成硬件、互联网和数据中心。今天的互联网技术强调 SoA 和 Web 2.0 服务。效用和网格计算是云计算所需的分布式计算基础。最后，我们不能忽视数据中心的广泛应用，以及使用虚拟化技术将资源配置过程自动化。

效用计算基于用户从云或物联网服务供应商处获取计算资源的商业模式。这在计算机科学与工程的所有方面都面临着许多技术挑战。例如，用户需求新的高效网络处理器、可扩展的内存和存储方案、分布式操作系统（OS）、机器虚拟化中间件、新的编程模型、有效的资源管理和应用程序开发。这些硬件和软件的进步对于促进不同物联网应用领域的移动云计算是必要的。

1.1.2 可扩展的分布式 / 并行计算的演化

一般的计算趋势是通过互联网增加共享网络资源的利用。图 1.2 阐述了 HPC 系统和 HTC 系统的演化。在 HPC 方面，超级计算机（大规模并行处理器（Massively Parallel Processors, MPP））逐渐被协同计算机集群所替代，这是出于共享计算资源的愿望。集群通常是一个同构计算节点的集合，这些节点在物理上处于近距离范围且彼此连接。在 HTC 方面，对等（Peer-to-Peer, P2P）网络起源于分布式文件共享和内容分发应用。

从 HPC 范式到 HTC 范式的重点已经有了战略性的变化。P2P、云计算和 Web 服务平台更