



教育部高等学校计算机类专业教学指导委员会-华为ICT产学合作项目
数据科学与大数据技术专业系列规划教材

华为信息与网络
技术学院指定教材

云计算 原理与实践

王伟 ● 主编

郭栋 张礼庆 邱娟 张静轩 张东启 谭一鸣 ● 编著



系统、完整的数据科学与大数据技术专业解决方案

名校名师打造大数据领域精品力作

系统学习完整云计算知识框架

提供在线实验平台，从原理到实践掌握云计算

 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS



教育部高等学校计算机类专业教学指导委员会-华为ICT产学合作项目
数据科学与大数据技术专业系列规划教材

华为信息与网络
技术学院指定教材

云计算 原理与实践

王伟 ● 主编

郭栋 张礼庆 邱娟 张静轩 张东启 谭一鸣 ● 编著



人民邮电出版社

北京

图书在版编目 (CIP) 数据

云计算原理与实践 / 王伟主编. — 北京: 人民邮电出版社, 2018. 8
数据科学与大数据技术专业系列规划教材
ISBN 978-7-115-48303-4

I. ①云… II. ①王… III. ①云计算—教材 IV.
①TP393.027

中国版本图书馆CIP数据核字(2018)第157780号

内 容 提 要

本书系统地讲解了云计算的知识框架, 包括云计算的三大认识角度(商业模式、计算范式、实现方式)、四个关键技术(计算、存储、网络、安全)、四种开发运维维度(云原生应用、云操作系统、云端软件、云计算运维), 以及三大应用场景(桌面云、软件开发云、大数据与人工智能)。

本书在大部分的章中安排了一个或多个基于开源软件的实训内容, 以帮助读者更有效地动手实践, 包括 GitHub、Hadoop、OpenStack、KVM、Docker、Ceph、Mininet、Mesos、Kubernetes 等。

本书适合作为高等院校本科生、研究生的云计算及其相关课程的教材, 也可作为相关研究人员和工程技术人员的参考资料。

-
- ◆ 主 编 王 伟
编 著 郭 栋 张礼庆 邱 娟 张静轩
张东启 谭一鸣
责任编辑 张 斌
责任印制 沈 蓉 彭志环
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
大厂聚鑫印刷有限责任公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 26 2018 年 8 月第 1 版
字数: 679 千字 2018 年 8 月河北第 1 次印刷
-

定价: 65.00 元

读者服务热线: (010)81055256 印装质量热线: (010)81055316
反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字 20170147 号

教育部高等学校计算机类专业教学指导委员会-华为 ICT 产学合作项目
数据科学与大数据技术专业系列规划教材

编 委 会

- 主 任 陈 钟 北京大学
副主任 杜小勇 中国人民大学
周傲英 华东师范大学
马殿富 北京航空航天大学
李战怀 西北工业大学
冯宝帅 华为技术有限公司
张立科 人民邮电出版社
秘书长 王 翔 华为技术有限公司
戴思俊 人民邮电出版社

委 员 (按姓名拼音排序)

- | | | | |
|-----|----------|-----|---------|
| 崔立真 | 山东大学 | 段立新 | 电子科技大学 |
| 高小鹏 | 北京航空航天大学 | 桂劲松 | 中南大学 |
| 侯 宾 | 北京邮电大学 | 黄 岚 | 吉林大学 |
| 林子雨 | 厦门大学 | 刘 博 | 人民邮电出版社 |
| 刘耀林 | 华为技术有限公司 | 乔亚男 | 西安交通大学 |
| 沈 刚 | 华中科技大学 | 石胜飞 | 哈尔滨工业大学 |
| 嵩 天 | 北京理工大学 | 唐 卓 | 湖南大学 |
| 汪 卫 | 复旦大学 | 王 伟 | 同济大学 |
| 王宏志 | 哈尔滨工业大学 | 王建民 | 清华大学 |
| 王兴伟 | 东北大学 | 薛志东 | 华中科技大学 |
| 印 鉴 | 中山大学 | 袁晓如 | 北京大学 |
| 张志峰 | 华为技术有限公司 | 赵卫东 | 复旦大学 |
| 邹北骥 | 中南大学 | 邹文波 | 人民邮电出版社 |

毫无疑问，我们正处在一个新时代。新一轮科技革命和产业变革正在加速推进，技术创新日益成为重塑经济发展模式和促进经济增长的重要驱动力量，而“大数据”无疑是第一核心推动力。

当前，发展大数据已经成为国家战略，大数据在引领经济社会发展中的新引擎作用更加突显。大数据重塑了传统产业的结构和形态，催生了众多的新产业、新业态、新模式，推动了共享经济的蓬勃发展，也给我们的衣食住行带来根本改变。同时，大数据是带动国家竞争力整体跃升和跨越式发展的巨大推动力，已成为全球科技和产业竞争的重要制高点。可以大胆预测，未来，大数据将会进一步激起全球科技和产业发展浪潮，进一步渗透到我们国计民生的各个领域，其发展扩张势不可挡。可以说，我们处在一个“大数据”时代。

大数据不仅仅是单一的技术发展领域和战略新兴产业，它还涉及科技、社会、伦理等诸多方面。发展大数据是一个复杂的系统工程，需要科技界、教育界和产业界等社会各界的广泛参与和通力合作，需要我们以更加开放的心态，以进步发展的理念，积极主动适应大数据时代所带来的深刻变革。总体而言，从全面协调可持续健康发展的角度，推动大数据发展需要注意以下五个方面的辩证统一和统筹兼顾。

一是要注重“长与短结合”。所谓“长”就是要目标长远，要注重制定大数据发展的顶层设计和中长期发展规划，明确发展方向和总体目标；所谓“短”就是要着眼当前，注重短期收益，从实处着手，快速起效，并形成效益反哺的良性循环。

二是要注重“快与慢结合”。所谓“快”就是要注重发挥新一代信息技术产业爆炸性增长的特点，发展大数据要时不我待，以实际应用需求为牵引加快推进，力争快速占领大数据技术和产业制高点；所谓“慢”就是防止急功近利，欲速而不达，要注重夯实大数据发展的基础，着重积累发展大数据基础理论与核心共性关键技术，培养行业领域发展中的大数据思维，潜心培育大数据专业人才。

三是要注重“高与低结合”。所谓“高”就是要打造大数据创新发展高地，要结合国家重大战略需求和国民经济主战场核心需求，部署高端大数据公共服务平台，组织开展国家级大数据重大示范工程，提升国民经济重点领域和标志性行业的大数据技术水平和应用能力；所谓“低”就是要坚持“润物细无声”，推进大数据在各行各业和民生领域的广泛应用，推进大数据发展的广度和深度。

四是要注重“内与外结合”。所谓“内”就是要向内深度挖掘和深入研究大数据作为一门学科领域的深刻技术内涵，构建和完善大数据发展的完整理论体系和技术支撑体系；所谓“外”就是要加强开放创新，由于大数据涉及众多学科领域和产业行业门类，也涉及国家、社会、个人等诸多问题，因此，需要推动国际国内科技界、产业界的深入合作和各级政府广泛参与，共同研究制定标准规范，推动大数据与人工智能、云计算、物联网、网络安全等信息技术领域的协同发展，促进数据科学与计算机科学、基础科学和各种应用科学的深度融合。

五是要注重“开与闭结合”。所谓“开”就是要坚持开放共享，要鼓励打破现有体制机制障碍，推动政府建立完善开放共享的大数据平台，加强科研机构、企业间技术交流合作，推动大数据资源高效利用，打破数据壁垒，普惠数据服务，缩小数据鸿沟，破除数据孤岛；所谓“闭”就是要形成价值链生态闭环，充分发挥大数据发展中技术驱动与需求牵引的双引擎作用，积极运用市场机制，形成技术创新链、产业发展链和资金服务链协同发展的态势，构建大数据产业良性发展的闭环生态圈。

总之，推动大数据的创新发展，已经成为了新时代的新诉求。刚刚闭幕的党的十九大更是明确提出要推动大数据、人工智能等信息技术产业与实体经济深度融合，培育新增长点，为建设网络强国、数字中国、智慧社会形成新动能。这一指导思想为我们未来发展大数据技术和产业指明了前进方向，提供了根本遵循。

习近平总书记多次强调“人才是创新的根基”“创新驱动实质上是人才驱动”。绘制大数据发展的宏伟蓝图迫切需要创新人才培养体制机制的支撑。因此，需要把高端人才队伍建设作为大数据技术和产业发展的重中之重，需要进一步完善大数据教育体系，加强人才储备和梯队建设，将以大数据为代表的新兴产业发展对人才的创新性、实践性需求渗透融入人才培养各个环节，加快形成我国大数据人才高地。

国家有关部门“与时俱进，因时施策”。近期，国务院办公厅正式印发《关于深化产教融合的若干意见》，推进人才和人力资源供给侧结构性改革，以适应创新驱动发展战略的新形势、新任务、新要求。教育部高等学校计算机类专业教学指导委员会、华为公司和人民邮电出版社组织编写的《教育部高等学校计算机类专业教学指导委员会-华为 ICT 产学合作项目——数据科学与大数据技术专业系列规划教材》的出版发行，就是落实国务院文件精神，深化教育供给

侧结构性改革的积极探索和实践。它是国内第一套成专业课程体系规划的数据科学与大数据技术专业系列教材，作者均来自国内一流高校，且具有丰富的数据教学、科研、实践经验。它的出版发行，对完善大数据人才培养体系，加强人才储备和梯队建设，推进贯通大数据理论、方法、技术、产品与应用等的复合型人才培养，完善大数据领域学科布局，推动大数据领域学科建设具有重要意义。同时，本次产教融合的成功经验，对其他学科领域的人才培养也具有重要的参考价值。

我们有理由相信，在国家战略指引下，在社会各界的广泛参与和推动下，我国的大数据技术和产业发展一定会有光明的未来。

是为序。



中国科学院院士 郑志明

2018年4月16日

在 500 年前的大航海时代，哥伦布发现了新大陆，麦哲伦实现了环球航行，全球各大洲从此连接了起来，人类文明的进程得以推进。今天，在云计算、大数据、物联网、人工智能等新技术推动下，人类开启了智能时代。

面对这个以“万物感知、万物互联、万物智能”为特征的智能时代，“数字化转型”已是企业寻求突破和创新的必由之路，数字化带来的海量数据成为企业乃至整个社会最重要的核心资产。大数据已上升为国家战略，成为推动经济社会发展的新引擎，如何获取、存储、分析、应用这些大数据将是这个时代最热门的话题。

国家大数据战略和企业数字化转型成功的关键是培养多层次的大数据人才，然而，根据计世资讯的研究，2018 年中国大数据领域的人才缺口将超过 150 万人，人才短缺已成为制约产业发展的突出问题。

2018 年初，华为公司提出新的愿景与使命，即“把数字世界带入每个人、每个家庭、每个组织，构建万物互联的智能世界”，它承载了华为公司的历史使命和社会责任。华为企业 BG 将长期坚持“平台+生态”战略，协同生态伙伴，共同为行业客户打造云计算、大数据、物联网和传统 ICT 技术高度融合的数字化转型平台。

人才生态建设是支撑“平台+生态”战略的核心基石，是保持产业链活力和持续增长的根本，华为以 ICT 产业长期积累的技术、知识、经验和成功实践为基础，持续投入，构建 ICT 人才生态良性发展的使能平台，打造全球有影响力的 ICT 人才认证标准。面对未来人才的挑战，华为坚持与全球广大院校、伙伴加强合作，打造引领未来的 ICT 人才生态，助力行业数字化转型。

一套好的教材是人才培养的基础，也是教学质量的重要保障。本套教材的出版，是华为在大数据人才培养领域的重要举措，是华为集合产业与教育界的高端智力，全力奉献的结晶和成果。在此，让我对本套教材的各位作者表示由衷的感谢！此外，我们还要特别感谢教育部高等学校计算机类专业教学指导委员会副主任、北京大学陈钟教授以及秘书长、北京航空航天大学马殿富教授，没有你们的努力和推动，本套教材无法成型！

同学们、朋友们，翻过这篇序言，开启学习旅程，祝愿在大数据的海洋里，尽情展示你们的才华，实现你们的梦想！



华为公司董事、企业 BG 总裁 阎力大

2018 年 5 月

“云计算”在 2006 年时还是个未知概念，到今天“云计算”已经家喻户晓、落地生根，并快速地发展壮大，“像用电一样使用信息服务”的云计算理想虽然还未完全实现，但距离这个目标已经非常接近了。

在 IT 领域，基本上每 2~3 年就会进行一次产品技术的更新换代，云计算经过多年发展，无论在技术深度还是在技术广度上均会有显著的延展。可以明显感受到云计算技术这些年来的发展和进步，以及云计算技术在整个信息基础设施重构中所起到的作用。

云计算可以看作是分布式计算、并行计算和网格计算等计算范式的集大成者。云计算的发展借鉴了这些不同计算模式的优点。以并行计算为例，并行计算求解问题的大致过程为：对于一个给定的应用问题，首先，计算科学家将这个应用转化为一个数值或非数值的计算问题；然后计算机科学家对此计算问题设计并行算法，并通过某种并行编程语言实现它；最后应用领域的专家在某台具体的并行计算机上运行应用软件求解此问题。而云计算在解决大数据的问题上同样如此：首先，科学家将实际应用问题转化成一个大数据的计算问题；然后对该问题进行并行算法设计（例如 MapReduce），并通过某个具体语言或软件进行实现（例如 Hadoop）；最后在具体的计算系统上（如数据中心）运行该算法求解该问题。

计算机技术的发展使得“计算科学”已经与传统的“理论科学”和“实验科学”并列成为推动科技发展和社会文明进步的三大科学，并朝着第四范式的“数据科学”迈进。云计算作为一种研究的工具，已经逐渐越来越深度地融入传统的“理论科学”“实验科学”“计算科学”，以及新兴的“数据科学”中。如化学家哈姆弗雷·戴维爵士（Sir Humphrey Davy）曾经明智地指出：“没有什么比应用新工具更有助于知识的发展。在不同的时期，人们的业绩不同，与其说是天赋智能所致，倒不如说是他们拥有的工具特性和软资源（非自然资源）不同所致。”

美国总统信息技术咨询委员会（PITAC）在致总统的《计算科学：确保美国竞争力》的报告中，有这样一段描述：“虽然计算本身也是一门学科，但其具有促进其他学科发展的作用。21 世纪科学上最重要的和经济上最有前途的研究前沿，有可能通过熟练掌握先进的计算技术和运用计算科学得到解决”。云计算就是这一论断的最好注解，是本世纪目前能够对其他学科产生重要影响的前沿技术，通过云计算，很多以往无法想象的问题都将迎刃而解。

我本人长期从事并行计算研究，并行计算“结构—算法—编程—应用”一体化的研究方法是我在不断实践的过程中逐渐总结出来的。虽然并行计算技术不断得到发展，但一体化的研究方法依然具有其指导价值。很高兴看到王伟老师所编写的这本云计算教材正是这一方法论在云计算教学中的应用。该教材从云计算的基本概念入手，通过云计算的计算、存储、网络和安全四个关键技术来阐述云计算的结构，然后通过云原生应用、云操作系统和云端软件来叙述云计算时代的算法和编程，最后通过桌面云、开发云以及大数据与人工智能来阐述云计算的应用场景。

我相信这本书会使许多人受益，希望本书的出版能够对云计算乃至计算机学科的发展起到好的促进作用！

中国科学院院士 陈国良

2018年5月

云计算自 2006 年出现至今已经 12 年，按照中国传统文化正好是一个生肖的轮转。在过去的 12 年，云计算的应用模式、核心技术与实践方式都发生了飞速的发展，正是需要全面梳理过去 12 年云计算发展的时候。在“教育部高等学校计算机类专业教学指导委员会-华为 ICT 产学合作项目”的资助下，王伟老师等编写的《云计算原理与实践》一书恰到好处地满足了这方面的需要。作为计算机系统领域的教育工作者、研究者与实践者，我从教育、研究与实践三个方面谈一下对本书的感受。

我是一名从事过“操作系统”“计算机系统工程”“分布式系统”与“计算机系统设计与实践”等计算机系统本科生与研究生核心课程教学的高校教师。从计算机系统教学的视角来看，该书源于王伟等老师在同济大学 5 年多的课堂教学课件整理而来，从概念与基础、原理与技术、开发与运维以及应用与案例等方面系统性地阐述了云计算的相关知识。本书的介绍深入浅出，便于理解，对于希望系统性了解云计算的原理与实践的师生具有很好的帮助。

正如本书总结的，云计算作为一种新的商业模式、计算范式与实现方式，涵盖了计算机系统的各个方面的支撑技术，包括操作系统、系统虚拟化、分布式系统、软件定义网络、系统安全与编程模型等。从一名从事计算机系统领域研究者的角度来看，本书的内容不仅概括了计算机系统领域的经典知识与最新研究成果，并且以云计算为纲领将各个领域的知识与成果进行了有机组织，对于有志于开展计算机系统领域研究的科研工作者也能提供很好的参考作用。

云计算的成功在于实践。因此，本书着重突出实践的重要性。通过 Docker 容器、分布式存储系统 Ceph、软件定义网络 Openflow、全同态加密算法 HElib、Node.js、Mesos、大视频运维、OpenStack 与 DevCloud 等主流软件与系统的实践介绍，本书对于有志于在云计算领域开展工作的实践者也有很强的参考价值。

云计算仍在快速的发展中。王伟老师等作者通过在云计算领域多年教育、研究与实践的经验，通过《云计算原理与实践》的分享，必将对云计算的推广与深入发展起到重要的推动作用。希望本书对各位读者有所帮助，在学习与工作中更上一层楼！

上海交通大学教授、并行与分布式系统研究所所长
华为操作系统首席科学家、操作系统内核实验室主任
陈海波

2018 年 5 月

“云计算”一词出现在 2006 年左右。当时谷歌 (Google) 公司的 CEO 施密特 (Eric Schmidt) 在一次会议上提出了“云计算”这个概念。如果说是谷歌为云计算命名,那么亚马逊 (Amazon) 公司则为云计算明确了商业模式。亚马逊在谷歌提出云计算的概念后不久,就正式推出了 EC2 云计算服务模式。从此之后各种有关云计算的概念层出不穷,“云计算”开始流行。

所有科学的认识都是以一种层次递进、螺旋上升的方式发展的,每一种技术都有其适用的场景和范围。云计算糅合了各种技术,并不代表云计算比其他技术更优秀,而是说明云计算技术确实是针对“大用户”“大数据”“大系统”,甚至“大智能”发展出来的一种新的实现机制。

云计算的一个特别之处,在于从技术上创造性地给出了一种灵活可靠的组织机制,通过将各种资源进行快速调度和组合,来满足不同业务应用的需求。这种不但适用于业务模块,而且适用于底层硬件资源的“积木式重组”思想,重新定义了反映在上层的计算资源的使用方式、服务的提供方式,以及社会化大生产的协作过程,为我们解决互联网带来的“大”问题和创新服务模式提供了一种全新的思路。

云计算技术使大量的硬件资源通过虚拟化技术结合成一个有机整体,再通过数据传输、负载均衡等技术来相互依赖、相互作用,完成预设功能,形成一个标准概念上的“系统”。这个系统的特征,是在物理上分散,在逻辑上集中,即所谓的分布式集中。

数年前,亚马逊 EC2 云基础设施平台就分布在全球 8000 多个机架上,而当时的谷歌在全球有将近 100 万台服务器。依靠大规模分布式软件架构的支持,这种大量资源的逻辑集中,一是意味着通过技术手段充分利用这些资源,我们可以满足“大用户”的需求,解决“大数据”的问题;二是通过对不同资源(硬件、应用)进行调度,我们可以基于一个云平台提供多种服务,即各种 IaaS、PaaS 和 SaaS,以及 XaaS(一切皆为服务),满足各类用户的需求。

从服务的角度说,云计算的落地是已然确定的,大家使用的 Salesforce.com 的在线 CRM 应用、苹果的 iCloud、谷歌的 App Engine、亚马逊的 S3 和 EC2、阿里云、腾讯云、华为云、UCloud 等都是云计算服务。作为互联网服务的一种延伸,云计算服务或许可被称为 Internet 2.0。云计算服务很自然带有一些互联网服务的特征,但它也有自己的特点,形成了新的商业模式,如“按需使用”“多租户支持”等。

从技术应用的角度看,云计算平台是一种新的计算资源的使用和管理思路。在云计算之前,企业应对大用户、大数据问题时的唯一选择是购买更多更高性能的服务器。而云计算的出现,创新性地以“分布式集中”的方式,将分散的廉价计算资源组合在一起,发挥“群体”的功能,来应对大用户和大数据的压力,从而形成了新的“机器管理机器”的思路,在技术实现上有完整的体系架构。

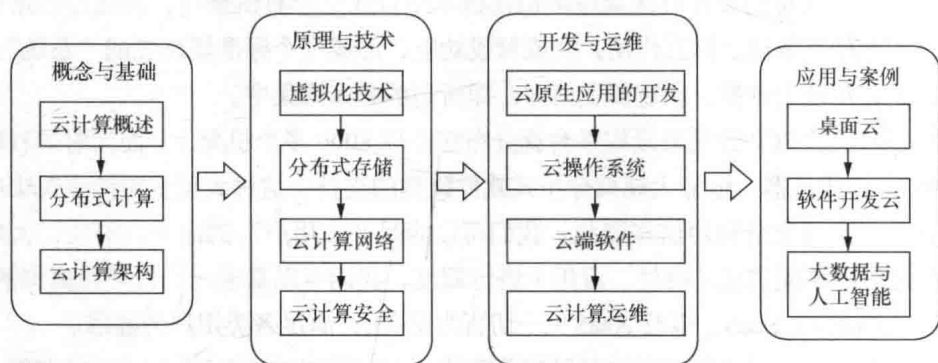
分布式系统通过网络将物理上分散的计算资源连接起来解决问题,用网络技术就可以解决分布式系统中的“距离”问题。然后,在网络协同的基础之上建立资源的调度、存储体系,为上层开发打基础。这样,我们就可以针对“计算”“通信”“存储”和“管理”提供一个实现云计算的“通用”方案。

因此，云计算既是一种商业模式，也是一种计算范式，还是一种实现方式。本书就是这样一种尝试，希望通过一种多维的角度将云计算相关的原理与实践方法传递给读者。

我们希望这本书首先能作为一本供大专院校和相关研究机构使用的云计算入门教材，其次也尽量使其成为能引起普通读者共鸣的一本读物。

本节提出了“三四四三”的云计算知识框架，包括云计算的三大认识角度（商业模式、计算范式、实现方式），四个关键技术（计算、存储、网络、安全），四种开发运维维度（云原生应用、云操作系统、云端软件、云运维），三大应用场景（桌面云、开发云、大数据与人工智能）。

本书共分四大部分：第一部分概念与基础主要包括云计算概述、分布式计算和云计算架构；第二部分原理与技术主要包括虚拟化技术、分布式存储、云计算网络以及云计算安全；第三部分开发与运维主要包括云原生应用的开发、云操作系统、云端软件以及云计算运维；第四部分应用与案例包括桌面云、软件开发云、大数据与人工智能三个云计算重要的应用领域。全书总体结构如下图所示：



云计算的知识框架全景图

本节是作者所开设的“云计算原理与实践”这门课程的实践成果，特别感谢 2012 年至 2017 年期间选修该门课程的学员，每次开课教室都坐得满满的，你们对新知识的热情与渴望是推动本课程建设的原动力，同时也为本课程提出了很多好的建议。

本节的另一大特点就是采用了众多的开源软件作为实践基础。所谓“云起之时，开源有道”。回顾历史，在 2000 年左右，国内软件开发领域最热门的操作系统、语言、开发工具、数据库等基本都是大型商业公司的产品。尽管当时 Linux 已经存在，但是还不算主流。我们当时所工作的 IT 环境，大部分服务器使用的是 Windows Server 或者 Sun Solaris 这样的操作系统。市场上需求最火爆的开发平台是 Visual C++、Visual Basic 和已经基本消失不见的 Delphi。然而近 20 年后的今天，当再次审视当前所处的环境时，我们会惊讶地发现，开源社区的产品已出现在各个领域：从操作系统、开发工具、编程语言，到中间件、数据库，再到虚拟化、基础架构云、应用平台云等。

可以说当前的时代是名副其实的开源的时代，企业可以通过开源社区的创新构建一个完全开源的企业架构堆栈，个人也可以通过开源软件获得云之力。

为此，在整本书系统讲述云计算原理的同时，我们还在大部分章中安排一个或多个相关的开源软件，以帮助读者进行有效的实践，这些开源项目包括了：GitHub、Hadoop、OpenStack、KVM、Xen、Docker、Ceph、Mininet、Mesos、Kubernetes、CloudwareHub 等。可以说当下的云计算生态中，开源软件已经成为一个不可或缺的部分。

同时，我们在每章内容后配套了丰富的辅导材料帮助读者进行复习、思考和实践等活动，主要包括下面四个类型的材料：

- 课内复习：帮助读者复习本章的一些基本核心概念，从章节内容中基本上就可以找到对应答案；
- 课外思考：围绕本章核心内容的相关问题，一方面引发读者的思考，另一方面需要读者查找一些课外资料才能找到答案，甚至包括一些开放问题；
- 动手实践：围绕本章内容，结合对应的开源软件和工具，开展动手实践的活动，这也是本书的特色之一，动手实践已经成为新工科背景之下课程内容的必要组成部分；
- 论文研习：阅读所推荐的学术论文，深度调研与本章内容相关的话题，培养读者学术论文阅读与理解的能力，从中也可以找到很多云计算领域的最新前沿内容。

本书的写作、出版是在“教育部高等学校计算机类专业教学指导委员会-华为 ICT 产学合作项目”专家组的大力支持下完成的，感谢教育部高等学校计算机类专业教学指导委员会专家陈钟教授、周傲英教授、杜小勇教授、马殿富教授在本书成书过程中的指导。本书在编写过程中，参考和引用了大量国内外的著作、论文和研究报告。由于篇幅有限，本书仅仅列举了主要的参考文献。作者向所有被参考和引用相关文献的作者表示由衷的感谢，他们的辛勤劳动成果为本书提供了丰富的资料。如果有的资料没有查到出处或因疏忽而未列出，请原作者见谅，并请告知我们，以便再版时补上。

衷心感谢人民邮电出版社的所有工作人员，从本书的策划开始，正是在他们无数次的帮助下，多次满足我在书稿上的各种苛刻要求，才使本书顺利出版。还要感谢华为公司的技术人员，他们的帮助和指导使得该书能够更加接地气。

苏轼在《稼说》一文中提出学习的主张“博观而约取，厚积而薄发”，这是我们多年教育工作的共鸣，其精髓就是勤于积累和精于应用。一个好的教育，是一个灵魂对另一个灵魂的呼唤；一门好的课程，是一个生命对另一个生命的碰撞。

最后，欢迎读者关注我们的公众号（嘉数汇，微信号 Datahui），获取配套的课件、扩展阅读材料和实践信息等。

编者

2018 年 3 月于上海

数字课程资源与实训平台的使用说明

与本书配套的数字课程资源与实训平台发布在我们自建的课程网站上，请登录网站后开始使用配套的课程资源 <https://github.com/willtongji/ppcloudcomputing>。

1. 注册/登录

访问课程平台网站，单击“注册”按钮，在注册页面输入用户名、密码及常用的邮箱进行注册。已经注册的用户直接输入用户名和密码登录即可进入“我的课程”页面。

2. 课程绑定

单击“我的课程”页面右上方“绑定课程”按钮，按照网站提示输入教材封底防伪标签上的密码，单击“确定”按钮完成课程绑定。

3. 访问课程

完成课程绑定后即可访问与本书配套的课程资源和实训，根据网页提示进行使用。账号自登录之日起一年内有效，过期作废。

4. 实训平台

课题组自主开发了一套基于 Web 的实训平台，为本书相关章节配套了实训，避免读者安装复杂的环境。读者只需浏览器，就可以一键生成对应的实训环境，在网页上对照教程即可一步步体验目前主流的不同开源云计算软件的安装、配置、使用、应用等实训任务。让读者轻松地就可以和云计算进行亲密接触，大大增加了读者的学习实践体验。目前主要的实训任务如下。

- Git 实训任务（分布式版本控制系统）
- Hadoop 实训任务（分布式实例）
- OpenStack 实训任务（云计算架构）
- KVM 实训任务（虚拟化技术）
- Docker 实训任务（轻量级虚拟化技术）
- Ceph 实训任务（分布式存储技术）
- Mininet 实训任务（云计算网络）
- Kubernetes 实训任务（云原生）
- Mesos 实训任务（云操作系统）
- Spark 实训任务（大数据应用）
- TensorFlow 实训任务（人工智能应用）

如有账号问题，请发邮件至 ppcc2018@sina.com。

目 录 CONTENTS

第 1 章 云计算概述	1	2.2 分布式计算的理论基础	42
1.1 初识云计算	2	2.2.1 ACID 原则	42
1.1.1 云计算的定义	3	2.2.2 CAP 理论	43
1.1.2 计算模式的演进过程	6	2.2.3 BASE 理论	46
1.1.3 云计算简史	11	2.2.4 最终一致性	46
1.1.4 云计算的推动力	12	2.2.5 一致性散列	47
1.2 云计算的公共特征与分类	14	2.3 分布式系统概述	50
1.2.1 云计算的公共特征	14	2.3.1 分布式系统的基础知识	50
1.2.2 云计算的分类	15	2.3.2 分布式系统的特性	51
1.2.3 与云计算有关的技术	19	2.3.3 分布式存储系统实例： Apache Hadoop	52
1.3 云计算的三元认识论	21	2.4 分布式系统的进阶	54
1.3.1 云计算作为一种商业模式	21	2.4.1 分布式存储系统	54
1.3.2 云计算作为一种计算范式	22	2.4.2 分布式计算系统	56
1.3.3 云计算作为一种实现方式	23	2.4.3 分布式资源管理系统	58
1.4 云计算的开源方法论	25	2.5 典型的分布式系统	60
1.4.1 开源定义和相关概念	25	2.5.1 网格系统	60
1.4.2 开源的价值和意义	26	2.5.2 P2P 系统	63
1.4.3 开源发展历程	27	2.5.3 透明计算	66
1.4.4 开源是方法论	28	2.5.4 区块链系统	69
1.4.5 开源给云计算人才培养带来的 挑战	30	2.6 本章小结	73
1.5 实践：GitHub	32	2.7 复习材料	73
1.5.1 GitHub 简介	32	第 3 章 云计算架构	75
1.5.2 使用 GitHub	33	3.1 云计算的本质	76
1.5.3 GitHub 界面总览	37	3.1.1 革命性概念：IT 作为服务	76
1.6 本章小结	38	3.1.2 云计算系统工程	76
1.7 复习材料	38	3.1.3 云数据中心	77
第 2 章 分布式计算	40	3.1.4 云的工作负载模式	78
2.1 分布式计算概述	41	3.1.5 云计算的规模效应	79
2.1.1 基本概念	41	3.2 云计算的架构	80
2.1.2 分布式计算的原理	42	3.2.1 计算架构的进化	80
		3.2.2 一般云计算架构的二维视角	83

3.3 云栈和云体	86	4.7.2 容器与虚拟机的对比	123
3.3.1 逻辑云栈	86	4.7.3 容器背后的内核知识	124
3.3.2 逻辑云体	89	4.8 实践: Docker 容器	126
3.3.3 一切皆为服务	90	4.8.1 安装 Docker	126
3.4 软件定义的数据中心	90	4.8.2 运行第一个 Docker 容器	127
3.4.1 数据中心的历史	91	4.9 本章小结	128
3.4.2 继续发展的推动力	93	4.10 复习材料	129
3.4.3 软件定义的必要性	94	第 5 章 分布式存储	130
3.4.4 软件定义数据中心的架构分析	95	5.1 分布式存储的基础	131
3.4.5 软件定义数据中心的发展	97	5.1.1 基本概念	131
3.5 实践: OpenStack	98	5.1.2 分布式存储分类	132
3.5.1 初识 OpenStack	98	5.1.3 分布式存储的发展历史	133
3.5.2 OpenStack 组件介绍	101	5.2 文件存储	136
3.5.3 体验使用 OpenStack	101	5.2.1 单机文件系统	136
3.6 本章小结	103	5.2.2 网络文件系统	137
3.7 复习材料	103	5.2.3 并行文件系统	137
第 4 章 虚拟化技术	104	5.2.4 分布式文件系统	138
4.1 虚拟化的定义	105	5.2.5 高通量文件系统	138
4.2 服务器虚拟化	106	5.3 从单机存储系统到分布式存储系统	139
4.2.1 x86 架构对虚拟化的限制	107	5.3.1 单机存储系统	139
4.2.2 全虚拟化	108	5.3.2 分布式存储系统	140
4.2.3 半虚拟化	108	5.4 实践: 分布式存储系统 Ceph	143
4.2.4 硬件辅助虚拟化	109	5.4.1 概述	143
4.3 商用虚拟机技术	109	5.4.2 设计思想	144
4.4 新型硬件虚拟化	111	5.4.3 整体架构	145
4.4.1 硬件虚拟化背景	111	5.4.4 集群部署	146
4.4.2 硬件虚拟化的代表	113	5.5 本章小结	149
4.4.3 硬件虚拟化的未来	114	5.6 复习材料	149
4.5 实践: Xen 虚拟化技术	115	第 6 章 云计算网络	151
4.5.1 Xen 的历史	115	6.1 基本概念	152
4.5.2 Xen 功能概览	115	6.1.1 计算机网络	152
4.5.3 Xen 实际操作	116	6.1.2 覆盖网络	153
4.6 实践: KVM 虚拟化技术	119	6.1.3 大二层网络	154
4.6.1 KVM 简介	119	6.1.4 租户网络	155
4.6.2 KVM 的基本安装操作	119	6.2 数据中心网络: 云计算的骨架	155
4.7 轻量级虚拟化	122		
4.7.1 容器技术简介	122		