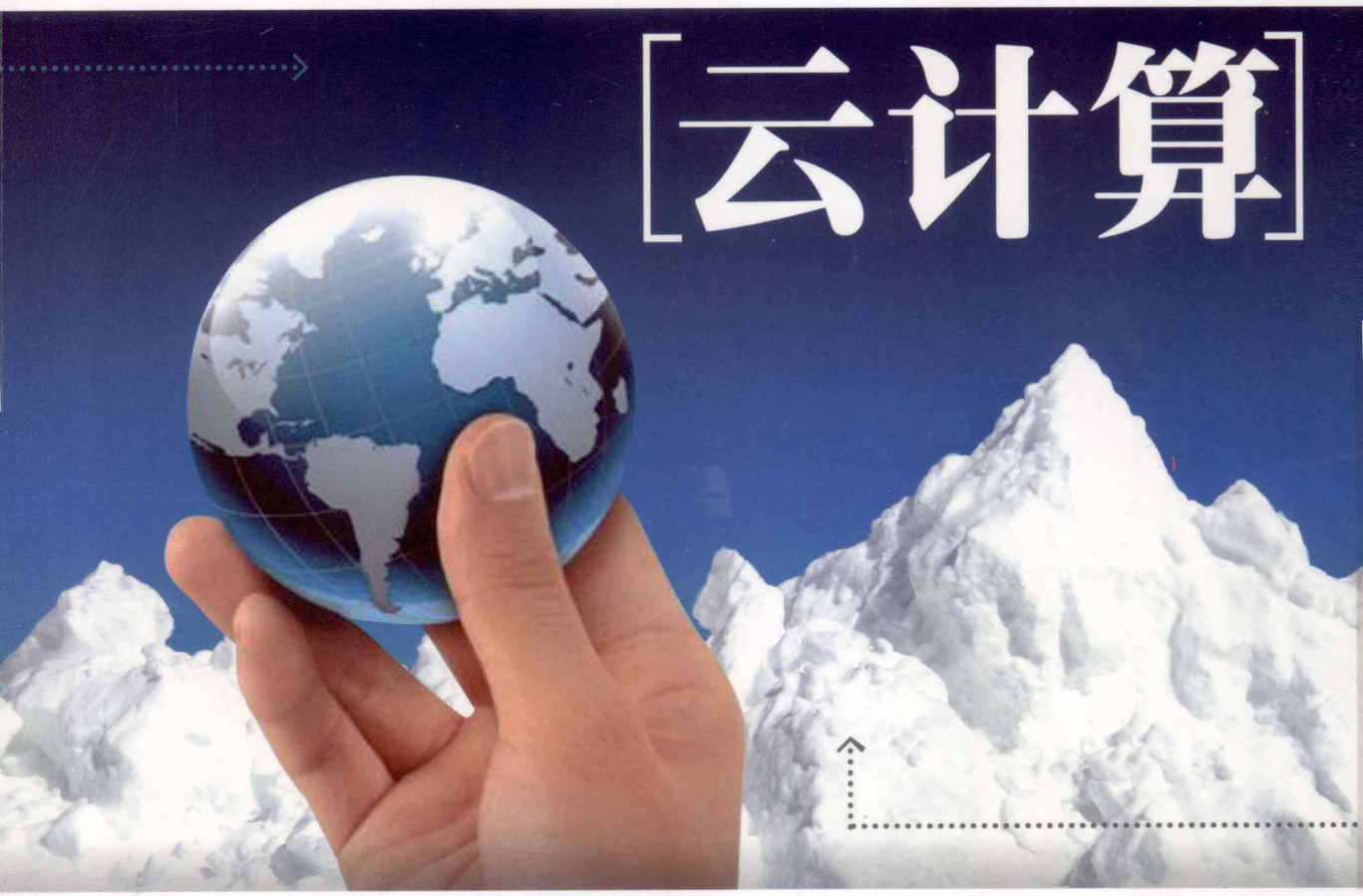
Hadoop
核心技术剖析与云端实现

Cloud Computing

深入浅出[云计算]



循序渐进讲述云计算的基本概念

基于Hadoop开源云计算平台，讲解如何构建一个基于云计算的应用系统

以Hadoop源代码为对象·深度剖析云计算核心技术

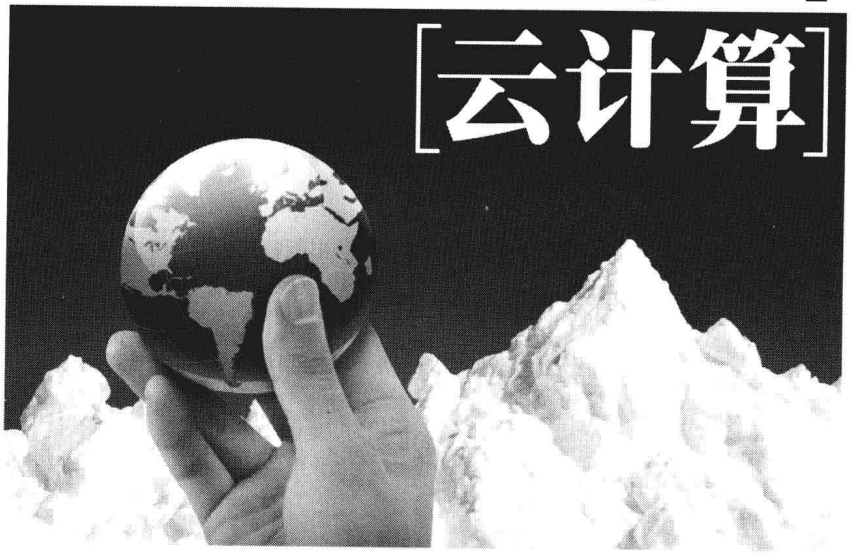
云技术与云端实现，讲述基于Hadoop云计算平台的4个高级应用框架，以及解决实际问题的思路与方法

鲍亮 陈荣 编著

清华大学出版社

Cloud Computing

深入浅出[云计算]



鲍亮 陈荣 编著

清华大学出版社
北京

内 容 简 介

本书作者以多年实际研发项目为背景，通过项目实战与代码分析，深入浅出地讲述云计算的基本概念，云计算的核心技术细节以及使用云计算平台解决实际问题的思路与方法。

全书共分 4 篇。第 1 篇循序渐进地介绍云计算的基本概念，学习云计算需要掌握的基本知识和云计算环境搭建方法；第 2 篇基于 Hadoop 开源云计算平台，讲解如何构建一个基于云计算的应用系统，了解云计算应用系统的设计方法；第 3 篇以开源的 Hadoop 云计算平台为分析对象，在源代码层次上对分布式文件系统、MapReduce 计算模型、NoSQL 数据库和集群管理算法与技术等云计算核心技术进行深度剖析；第 4 篇为云计算应用篇，介绍了基于 Hadoop 云计算平台的 4 个高级应用框架，读者可以结合自己的应用需求与场景，使用这些框架解决实际问题。

本书理论联系实践，既有理论深度又有实用价值，可作为高校教材使用，也可作为云计算研发人员以及爱好者的学习和参考手册。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989　13701121933

图书在版编目（CIP）数据

深入浅出云计算/鲍亮，陈荣编著. —北京：清华大学出版社，2012.10
ISBN 978-7-302-30238-4

Ⅰ.①深… Ⅱ.①鲍… ②陈… Ⅲ.①计算机网络—基本知识 Ⅳ.①TP393

中国版本图书馆 CIP 数据核字（2012）第 228217 号

责任编辑：夏非彼
封面设计：王　翔
责任校对：闫秀华
责任印制：何　芊

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：北京世知印务有限公司
装 订 者：三河市新茂装订有限公司
经　　销：全国新华书店
开　　本：190mm×260mm　　**印　　张**：28.25　　**字　　数**：723 千字
版　　次：2012 年 10 月第 1 版　　**印　　次**：2012 年 10 月第1 次印刷
印　　数：1～4000
定　　价：59.00 元

产品编号：045543-01

前　言

“云计算”这一概念最早由 IBM 和 Google 于 2007 年底提出，其本质是一种大规模的、由经济性所驱动的新一代分布式计算技术。采用云计算技术，可以将一组抽象的、虚拟的、动态可伸缩的以及受控的计算能力、存储、平台和服务通过互联网按需发布给海量用户。与传统的分布式计算技术（如集群、网格、P2P 等）相比，云计算的核心特征有其灵活性与经济性：第一，将基础设施架构在大规模的廉价服务器集群之上，通过多个服务器之间的冗余以及应用程序与底层服务协作开发等机制，使得软件具有高可用性，并最大限度地利用资源；第二，采用面向市场的业务模型，为用户提供诸如计算能力、存储和网络传输等基础服务，从中收取使用费用。

基于新型的系统架构，云计算实现了应用系统的可扩展性和高可用性；基于新的业务模型，云计算能够满足企业降低成本、提高工作效率以及简化 IT 管理的需求。目前，云计算技术已经广泛应用于电子商务、电信服务、工作流管理与客户关系管理等多个行业中。由于其突出的经济性、性能与可靠性优势，云计算自出现至今短短 5 年的时间，已得到学术界和产业界的广泛关注。

目前市场上与云计算技术相关的书籍较多，但经过作者调研，目前尚无以开源的云计算平台为对象，在源代码层次上对分布式文件系统、MapReduce 计算模型、NoSQL 数据库和集群管理算法与技术等云计算核心技术进行深度剖析的书籍。本书创作团队核心成员自 2005 年起就一直从事分布式计算、SOA 等相关方向的研发工作，具有丰富的项目实践经验。2009 年初，作者以实际研发项目为背景，组织团队分析、理解著名的开源云计算平台 Hadoop 的源代码，经过两年多的努力，其部分成果形成本书最为核心的第 9~12 章。通过项目实战与代码分析，我们积累了大量云计算平台的使用经验，对云计算的核心技术有了较为深刻的理解，认为有必要将自己的经验和认识整理出来，以满足广大读者希望一方面掌握云计算基本概念，能够使用云计算平台解决实际问题；另外一方面希望深入理解云计算核心技术细节的迫切心情，这也正是书名《深入浅出云计算》的由来。

本书适合不同层次的读者阅读。建议读者根据自己的兴趣和目的有选择性地阅读：希望了解云计算的基本概念、本质和发展趋势的读者，可以重点阅读第 1、2 章和附录；对于云计算的初学者，可以重点阅读 1~8 章；已经掌握云计算基本概念，具有一定实践基础，想进一步深入学习、研究云计算核心技术的读者，可以重点阅读 9~12 章；想利用云计算解决各类实际计算问题的读者，可以重点阅读 13~16 章。

感谢书籍创作团队核心成员王玉操、贾世达、王磊、葛军、李朝印、张翔、李江、杨阳和彭恒的辛勤努力。感谢我的导师陈平教授在云计算研究方面对我们的悉心指导。

由于云计算起源于企业界，是企业界推动学术界发展的一项新兴应用技术，相关资料数量不多，加之作者水平有限，时间紧迫，因此书中难免存在错误与不当，敬请读者批评指正。建议和意见请发至作者邮箱 baoliang@mail.xidian.edu.cn。

编　者

试读结束，需要全本PDF请购买 www.ertongbook.com

目　录

第 1 篇　初始云计算

第3篇 深入云计算

试读结束，需要全本PDF请购买 www.ertongbook.com

第1篇

初始云计算

本篇是入门篇，主要介绍云计算的基本概念、学习云计算需要掌握的基本知识和云计算环境搭建方法 3 个部分的内容。通过本篇的学习，读者可以了解到云计算的相关概念，云计算的历史以及云计算在产业界、学术界和政府机构等领域的发展现状，掌握 Java 网络编程、序列化与反序列化、数据库基础理论和分布式算法等基础知识，并能够搭建一个小型的云计算环境。

第 1 章　云计算介绍

自从 2007 年 10 月份云计算诞生至今，这一技术在短短的 4 年时间里对整个 IT 行业产生了巨大的影响。学术界、产业界和政府都对云计算产生了浓厚的兴趣：全球范围内讨论云计算技术的学术活动如火如荼；谷歌、亚马逊、IBM、微软等 IT 巨头大力推动云计算技术的宣传和产品的普及；各国政府纷纷斥巨资打造大规模的数据中心与计算中心。云计算技术目前已经得到了业界的高度认同，逐渐走向成熟。那么，云计算该如何定义？云计算的发展历史如何？云计算目前的现状如何？本章将重点解答这些问题，以帮助读者初步理解云计算。

1.1　云计算相关概念

云计算（Cloud Computing）这一概念于 2007 年 10 月 8 日正式出现，其标志性事件是谷歌和 IBM 宣布联合加入云计算的研究工作，并给出云计算的定义。同年 11 月 15 日，IBM 上海和阿莫科（Armok，NY）同时发布了 Blue Cloud，Blue Cloud 是一系列的云计算产品，使得共同的数据中心像互联网一样运作。

2007 年 10 月 8 日，纽约时报报道了谷歌和 IBM 联合加入云计算研究的新闻，如图 1.1 所示。

图 1.1　谷歌和 IBM 联合加入云计算研究的新闻

1.1.1 云计算的定义

自云计算这一概念诞生至今，尚未形成业界广泛认可的统一定义。本书将列举 4 种有代表性的云计算定义，并对每种定义方式进行解读。

1. IBM 的云计算定义

2007 年 10 月，IBM 的 Greg Boss 等人以技术白皮书[①]的形式给出了云计算的定义："云计算是同时描述一个系统平台或者一类应用程序的术语。云计算平台按需进行动态部署（Provision）、配置（Configuration）、重新配置（Reconfigure）以及取消服务（Deprovision）等。在云计算平台中的服务器可以是物理或虚拟的服务器。高级的云计算通常包含一些其他的计算资源，如存储区域网络（SANs）、网络设备、防火墙以及其他安全设备等。

在应用方面，云计算描述了一类可以通过互联网进行访问的可扩展应用程序。这类云应用基于大规模数据中心及高性能服务器来运行网络应用程序与 Web 服务。用户可以通过合适的互联网接入设备及标准的浏览器访问云计算应用程序。"

IBM 技术白皮书中的云计算定义原文：

Cloud computing is a term used to describe both a platform and type of application. A cloud computing platform dynamically provisions, configures, reconfigures, and deprovisions servers as needed. Servers in the cloud can be physical machines or virtual machines. Advanced clouds typically include other computing resources such as storage area networks (SANs), network equipment, firewall and other security devices.

Cloud computing also describes applications that are extended to be accessible through the Internet. These cloud applications use large data centers and powerful servers that host Web applications and Web services. Anyone with a suitable Internet connection and a standard browser can access a cloud application.

IBM 的定义明确指出云计算概念的内涵包含两个方面：平台和应用。平台即基础设施，其地位相当于 PC 机上的操作系统，云计算应用程序需要构建在平台之上；云计算应用所需的计算与存储通常在"云端"完成，客户端需要通过互联网访问计算与存储能力。

2. 加州大学伯克利分校的云计算定义

2009 年 2 月 10 日，加州大学伯克利分校电子工程和计算机学院的 Michael Armbrust 等人发布技术报告《Above the Clouds: A Berkeley View of Cloud Computing》[②]，介绍了对云计算

[①] IBM 云计算技术白皮书.http://download.boulder.ibm.com/ibmdl/pub/software/dw/wes/hipods/Cloud_computing_wp_final_8Oct.pdf

[②] 加州大学伯克利分校的云计算定义. http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf

试读结束，需要全本PDF请购买 www.ertongbook.com