

国家社科基金重大项目(编号: 12&ZD220)资助  
国家自然科学基金面上项目(编号: 4112030)资助



普通高等教育“十一五”国家级规划教材 计算机系列教材

IBM大学合作项目书籍出版资助  
微软研究院全球精品课程资助计划



# 分布式系统及云计算概论

## (第2版)

陆嘉恒 文继荣 编著



清华大学出版社

国家社科基金重大项目(编号: 12&ZD220)资助  
国家自然科学基金面上项目(编号: 4112030)资助



普通高等教育“十一五”国家级规划教材 计算机系列教材

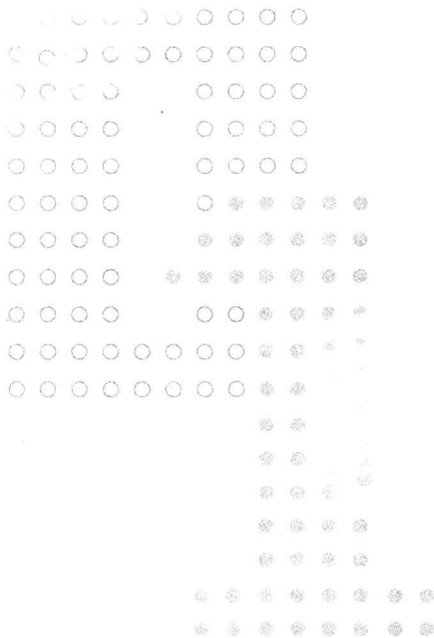


IBM大学合作项目书籍出版资助  
微软研究院全球精品课程资助计划

陆嘉恒 文继荣 编著

# 分布式系统及云计算概论

## (第2版)



清华大学出版社  
北京

## 内 容 简 介

云计算是一个新兴的术语,很多技术还处在起步阶段。云计算涉及的范围非常广,包括分布式计算、并行计算、效用计算等。本书从分布式系统的角度出发,系统地对云计算进行全面介绍,既有分布式系统和云计算系统的理论分析和内核技术阐述,又有对各大IT公司的云计算软件产品的使用方法的介绍和比较分析。本书作者队伍强大,有海内外一流高校的教授和研究学者,也有IT公司的云计算技术的开发和管理人员。本书可以作为高年级本科生、研究生的教材,也可供云计算的应用开发人员、行业专业人士以及相关学科的研究者作参考。

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

分布式系统及云计算概论/陆嘉恒,文继荣编著. --2版.--北京:清华大学出版社,2013

计算机系列教材

ISBN 978-7-302-34519-0

I. ①分… II. ①陆… ②文… III. ①分布式操作系统—教材 ②计算机网络—教材 IV. ①TP316.4  
②TP393

中国版本图书馆CIP数据核字(2013)第277181号

责任编辑:汪汉友

封面设计:傅瑞学

责任校对:时翠兰

责任印制:杨艳

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](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载: <http://www.tup.com.cn>, 010-62795954

印刷者:北京富博印刷有限公司

装订者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:21.25 字 数:486千字

版 次:2011年5月第1版 2013年12月第2版 印 次:2013年12月第1次印刷

印 数:1~1700

定 价:39.00元

---

产品编号:053956-01

近年来,随着分布式计算技术的发展,一种新兴的计算模式——云计算受到了研究者的极大关注。云计算已经成为当前计算机数据管理,计算机网络,计算机系统结构研究的前沿问题之一。它提供了一种新型的共享基础架构的方法,各服务提供商将巨大的系统池连接在一起提供各种 IT 服务。云计算是大规模分布式计算技术不断演进的结果;并行计算、网格计算、效用计算、服务计算等都为云计算的发展提供了技术基础。随着云计算的发展,这项兴起中的技术,正在深刻地影响乃至改变科学研究的方式以及人们学习、工作和生活方式。

虽然研究者针对云计算技术进行了大量的研究,但是到目前为止学术界和行业内还未形成对云计算的相关技术和产品的统一标准和规范。特别是在国内,关于云计算的研究较少,还处在探讨的初步阶段。迄今为止,国内尚缺乏全面系统介绍分布式系统与云计算技术的教材,而广大的大专院校学生和软件开发人员又迫切需要了解这一新兴的技术,本书的出版填补了这一空白。作者对基本概念给出了深入细致的介绍,对应用中的技术进行了详细具体的描述。对概念方面的内容,辅以详细的图表和相关的对比进行说明,对技术方面的内容和实际应用给出了细致的指导和相关代码,有效地提高了读者对技术问题的认识。相信本书可以为广大科研工作者和同学们提供重要的入门材料和学习参考。

中国人民大学教授 王珊

在计算机网络技术高速发展的今天,云计算已成为当今信息技术领域最受瞩目的新兴概念之一。微软公司把未来的计算定义在云+端、软件+服务。可以预见今后几年内会有越来越多的技术人员会参与到云计算系统的开发和研究中来。然而,现在介绍云计算及其相关技术的书籍还相当少,很多读者都无法找到相应的参考书籍。我们可以看到这本书是在云计算系统实际开发工作的基础上完成的。本书广泛地介绍了云计算的相关系统与技术,所介绍的内容可以为广大读者提供参考,开阔思路。

微软云计算平台上有三大非常重要的核心技术,一个是 Windows Azure;另一个是 SQL Azure;还有一个是 Windows Azure Platform AppFabric。本书对 Azure 云计算平台相关技术进行了详细的介绍,并通过一定的实例说明了如何使用开发 Azure 平台下的应用程序。希望读者能通过本书的学习,深入了解 Azure 云计算平台,并在此基础上做出更出色的云计算系统或进行更好的云计算研究。

微软亚洲研究院院长 洪小文

计算技术已经深刻地改变了我们的工作、学习和生活,电子技术和通信技术的飞速发展和进一步融合以及体系结构的进步更是将其提到前所未有的高度。分布式云计算是一项正在兴起中的计算技术,其发展脚步正在逐渐加快,云计算产品也在不断上市,其潜力巨大,正在扭转整个行业的面貌,也在悄然改变人们使用计算机的习惯。它的出现,有可能完全改变用户现有的以桌面为核心的使用习惯,而转移到以 Web 为核心,使用 Web 上的存储与服务。相信人类会因此迎来一个新的信息化时代!

云计算技术将对信息产业产生深远的影响,然而目前国内外还缺乏系统、深入地介绍相关技术的书籍。我非常欣喜地看到此书的出版,这本书系统、清晰、全面地介绍了云计算的概念、架构、关键技术以及各企业最新研究动向及研究成果。希望能帮助大家更好地了解云计算概念及其发展现状。更为重要的是我相信不管是云计算研究者、初学人员还是 IT 人或者云计算的最终用户,都能从本书中找到自己想要的东西。

云计算是一个美好的愿景,相信通过众多学者研究人员的努力,它必将引领我们的生活进入一个新的信息时代!

IBM 中国研究院院长 李实恭

可以预见,在未来几年内会有越来越多的人参与到云计算系统的研发中。然而,我们发现在介绍云计算及其相关技术的书籍方面还是一片空白。即使在澳大利亚,很多读者都无法找到相应的参考书籍。所幸,能够在中国看到这样一本系统介绍云计算领域的书籍。本书首先对分布式计算、云计算、云服务等做了简略介绍,特别是比较了效用计算、网格计算、集群计算与云计算。然后分类介绍 Google、IBM、Amazon、Greenplum、Hadoop、HBase、Microsoft Azure 等的概念与关键技术。这些都是当下最流行的与云计算直接相关联的技术,本书对其中每一类问题均提供具有代表性的实例。并对每种技术应用给出注重实效的规则和参考资料与习题。这本书不仅内容丰富,而且极佳的表达风格使之具有很好的可读性。

这本书还系统、清晰、全面地介绍了按需分配的虚拟化和云计算的概念、架构、关键技术以及最新研究动向,相信这本书能帮助大家更好地了解虚拟化和云计算技术。

澳大利亚墨尔本大学云计算与分布式系统试验室主任

Rajkumar Buyya 教授

随着计算机网络的进一步发展和对海量数据计算能力的要求,各种大型计算能力的计算机硬件不断出现。此外,全球信息系统万维网也非常流行。这些软硬件技术或设备的出现,为提出一种新型的计算模型提供了可能。本书考察分布式系统与云计算的概念和技术。云计算是分布式系统中一个有巨大商业前景的学科前沿。它是虚拟化、效用计算、基础设施即服务、平台即服务和软件即服务等概念混合演进的结果。

云计算作为一个新兴的术语,目前还没有确切的定义,很多技术还处在起步阶段。它涉及的范围非常广。本书主要从分布式系统的角度对云计算进行解释,并通过对现有各大IT公司云计算技术的分析,使读者对云计算建立一种直观概念。本书是对云计算技术的全面介绍,同时关注系统实现和设计问题。对于本科生、研究生、应用开发人员、行业专业人士以及相关学科的研究者而言,本书应该也是比较全面的云计算书籍。

云计算是分布式计算、并行计算和网格计算的发展,或者说是这些计算机科学概念的商业实现。云计算目前还处在起步阶段,但由于IT界各大公司的争相角逐,云计算在不久会迎来它的黄金发展机遇,给IT带来全面的革新。本书提供分布式系统和云计算的全面知识,介绍云计算相关的技术和开源项目,并讨论和设计相关云计算平台的应用程序开发设计。写本书的一个重要动机就是,在当前云计算到底为何物,如何具体实现云计算这些基本问题尚不明了的情况下,帮助读者入门与提高,这是一项富有挑战性的任务。我们希望通过本书的出版,与不同领域、不同层次读者探讨、交流,为进一步促进这一领域的发展做出贡献。

### 1. 本书的组织

本书除第1章绪论外其他内容分为3篇。

第一篇 分布式系统,包含第2章和第3章。这部分对分布式计算进行介绍,为读者进行后续章节的学习打下良好的知识基础。

第2章是分布式计算的入门,引入分布式计算的定义及其相关概念,简单地介绍背景知识,方便读者理解本书中所提到的材料,并对人们非常关心的分布式计算的安全性与容错性问题进行重点论述。

第3章讨论客户—服务器端架构的问题。首先讲解客户—服务器端的相关协议,然后分别介绍面对有连接和无连接情况时客户—服务器端的构架,最后讨论有状态和无状态的服务器问题。

第二篇 云计算技术,包含第4章~第9章。这部分是对云计算的综述,从各个方面与角度,向读者全面地介绍云计算。我们选取了具有代表性的云平台和技术,这些代表性的技术更是现阶段研究学习的重点与未来继续开发新平台与新技术的重要参考。

第4章是云计算的入门章节,在这一章中经过综合比较分析,给出本书认为的云计算的定义,简述云计算的发展历史,详细分析云计算的优点与缺点,如果以前你对是否使用云计算或如何使用云计算还有疑虑,相信这一章将帮助你理清思路,本章对云计算所涉及的服务进行了讨论,最后是与云相关技术的比较,本章分别从网格计算、效用计算、集群计算、并行和分布计算等方面与云计算进行了具体比较。

第5章就 Google 公司云平台的 Google 文件系统(GFS)技术、Bigtable 技术、Map-Reduce 三大技术分别进行详细的介绍与论述,并针对每个具体技术给出应用的例子,方便学习及理解。

第6章讲述的是 Yahoo!公司的云平台技术。Yahoo!公司在云计算方面投入巨大,涉及云计算的许多方面并支持许多与此相关的开源项目,而本章介绍的正是其中的重要部分包括一种灵活通用的表存储平台 PNUTS 云平台,分析大型数据集的框架 Pig,提供团体服务的集中化服务平台 ZooKeeper。

第7章就 Greenplum 云平台技术进行介绍。Greenplum 是一个商业的分析型云数据库,同样也用来支持下一代数据仓库和大规模的分析过程。它支持 SQL 和 Map-Reduce 并行处理,可以提供低成本的管理太字节(TB)到皮字节(PB)数量级数据的能力。相信在未来云计算发展过程中,云数据库将占有重要的一席之地。

第8章讲的是 Amazon 公司的 Dynamo 云平台技术。Amazon 公司提供的 EC2 和 S3 想必已为大家所熟知,但其中的关键技术所利用的平台 Amazon Dynamo,可能读者并不了解。本章分别从 Amazon Dynamo 的研发背景和 Dynamo 系统体系结构对其进行讲解。

第9章介绍 IBM 公司与云计算相关的项目,包括云风暴、智能商业服务、智慧地球、Z 系统以及虚拟化的动态基础架构。IBM 公司在云计算领域的工作的影响力十分巨大,并且在云计算的普及方面以其针对企业的云服务推动了领域的发展。

第三篇 分布式云计算的程序开发,包含第10章~第16章。这一部分是分布式云计算的程序开发部分,在了解了众多的与云计算相关的知识与技术之后,本书讲解各种主流的进行分布式云计算的程序开发系统与平台,并给出对应的开发实例,方便读者进行实践。

第10章对 Apache 开源项目 Hadoop 系统进行介绍。Hadoop 主要由 Hadoop 核心组件、MapReduce 以及 HDFS 组成。本章对 Hadoop 生态系统所包含的项目及其自身体系结构进行介绍,并对 Hadoop 集群相关安全策略进行说明。

第11章详细介绍 MapReduce,它是 Google 提出的 MapReduce 编程模型的开源实现,被广泛应用于海量数据的处理。本章对 MapReduce 的计算模型以及工作机制进行深入分析,并结合 MapReduce API 和实例介绍如何采用 MapReduce 框架编写并行程序。

第12章详细介绍 Hadoop 的另一核心子项目 HDFS。作为 Hadoop 所采用的分布式文件系统,本章着重对 HDFS 的体系结构以及数据在 HDFS 中的执行流程进行详细介绍。



第 13 章主要讨论 HBase 分布式数据库系统。HBase 的设计思想主要来源于 Google Bigtable 技术,为面向列存储的分布式数据库。本章将分别对 HBase 的体系结构、数据模型、物理视图和概念视图进行介绍,并说明如何采用 HBase Java API 对 HBase 数据库进行操作。

第 14 章详细介绍 Hive 数据仓库。Hive 的存储建立在 Hadoop 文件系统之上,它向上为用户封装了 Hive QL 编程语言,使用方法类似于 SQL 操作,这极大方便了用户的使用。

第 15 章讲解基于 Google App Engine 系统的开发。首先对 Google App Engine 做简介,然后讲解如何使用 Google App Engine,同时给出基于 Google App Engine 的应用程序开发实例。

第 16 章讲解基于 Windows Azure 系统的开发。首先对微软公司推出的 Windows Azure 系统做简介,然后讲解 Windows Azure 服务的使用方法,并给出应用开发实例。

## 2. 致教师

本书旨在提供分布式系统最近发展云计算领域的一个全面而深入的概述。本书作为国内第一本云计算领域的教材,可以作为高年级本科生、研究生的云计算导论或高级课程的教材。此外,本书也可以作为讲授云计算的高级课程的参考书籍。

如果您打算使用本书作为讲授云计算导论的教材,您会发现第一篇是介绍分布式系统的基本概念;第二篇主要对云计算的基本概念和相关技术进行介绍。第三篇主要介绍相关云计算平台编程技术,第 13 章和第 14 章要求掌握基本原理,第 11 章、第 12 章、第 15 章和第 16 章则可以安排学生进行相应的编程作业。此外,每章后面还有相关的参考文献,供教师安排学生进行更深入的学习指导。

本书的很多知识,可以作为独立的自学材料。每章后面我们都设计了相关习题,这些习题可以作为课后巩固知识的作业。同时,本书的网站([datasearch.ruc.edu.cn/cloud-book/](http://datasearch.ruc.edu.cn/cloud-book/)或清华大学出版社的网站)还提供了相应的教学辅助资料,如课程纲要和课程 PPT。

## 3. 致学生

作为国内第一本全面介绍云计算及相关技术的书籍,希望通过本书的学习能够激发起你对云计算研究的兴趣,了解计算机科学的最新发展方向。我们力求以清晰全面易懂的方式提供该方面的知识讲解。我们还是力求卓越,通过实际的编程实例使学生们在学习的过程中,真正走进云计算环境中。如果你在以后的研究或学习中能够深入了解云计算的相关知识和编程技术,那么它将会给你的未来职业提供一个非常美好的前景。

为学习本书,应该具有如下知识和技能。

- 具有一些分布式系统的基础知识。当然,本书尽力在介绍云计算之前对相关分布式的概念等知识进行介绍,帮助读者回忆或学习这方面的知识。
- 较强的计算机编程能力或阅读代码能力。特别是能够熟练使用 Java 或 C 语言。同时,还要有一定的 UNIX 和 Linux 知识背景。

#### 4. 致专业人员

本书旨在涵盖目前云计算领域的广泛课题。这样,本书是国内关于该主题的第一本内容全面的参考书。由于每章内容的相对独立性,可以只关心最感兴趣的课题。希望从事分布式系统和云计算的应用程序员和相关领域的管理者都可以从本书受益。

书中关于云计算的介绍很多都是目前研究领域的最新成果,可以帮助专业人员及时掌握云计算的最新进展,并把握住云计算的发展机遇。通过对目前云计算领域中比较流行的平台和技术的介绍,以及一定的编程实例说明,相信可以很快地帮助专业人员了解、掌握并做进一步深入的研究工作。

#### 5. 本书资源网站

本书的网站地址是 <http://datasearch.ruc.edu.cn/cloudbook/>,该网站为本书读者和对云计算感兴趣的人提供了一些补充资料,主要包括:

- 每章的幻灯片,提供 Microsoft PowerPoint 格式的教案;
- 相关章节的源代码;
- 课程大纲和教学计划;
- 补充读物目录;
- 本书勘误更正表。

欢迎到本网站指正本书的错误信息,一旦勘误被最终确认,会及时更新到勘误表中,并在此表示感谢!同时,也欢迎读者到本网站发表对本书的评论或建议信息,以帮助我们以后的工作。

#### 6. 致谢

在本书的形成过程中,感谢众多研究生和本科生的工作,特别是马中瑞、廖承炫、冯博亮、王仲远、胡一、黄超、黄飞、刘胜文、吕瑛、孙荣丽、林春彬、田宗起、程明、王海涌、陈东伟、周凯、陈新星、张林、徐文韬等在资料收集和文献整理等方面的工作。

本书涉及面甚广,内容丰富,涉及的相关技术众多。但需要指出的是,在全书的撰写和相关技术的研究中,尽管作者投入了大量的精力、付出艰辛的努力,然而受知识水平所限,书中不当之处在所难免,望读者批评指正并不吝赐教。如果有任何问题或建议,可发送电子邮件至 [jiahenglu@gmail.com](mailto:jiahenglu@gmail.com) 或 [jiahenglu@ruc.edu.cn](mailto:jiahenglu@ruc.edu.cn)。

作 者

2013 年 8 月于北京

第1章 绪论	/1
1.1 分布式计算与分布式系统	/1
1.1.1 分布式计算简介	/1
1.1.2 分布式系统的实例	/2
1.1.3 分布式系统的目标	/4
1.2 云计算	/5
1.2.1 简介	/5
1.2.2 云计算的优点和缺点	/7
1.3 本书概要	/10
1.4 小结	/11
习题	/12

## 第一篇 分布式系统

第2章 分布式系统入门	/15
2.1 分布式系统的定义	/15
2.1.1 分布式与集中式	/15
2.1.2 分布式与计算机网络	/16
2.1.3 分布式系统层次结构	/17
2.1.4 分布式系统分类	/18
2.2 分布式系统中的软硬件	/19
2.2.1 硬件	/19
2.2.2 软件	/20
2.3 分布式系统中的主要特征	/21
2.3.1 容错性	/21
2.3.2 安全性	/22
2.4 小结	/23
习题	/23
第3章 客户—服务器端架构	/25
3.1 客户—服务器模式的基本概念和优点	/25

- 3.1.1 客户—服务器模式的基本概念 /25
- 3.1.2 客户—服务器模式优点 /26
- 3.2 客户—服务器端架构和体系结构 /26
  - 3.2.1 面向连接服务与无连接服务 /26
  - 3.2.2 应用程序的层次结构 /27
  - 3.2.3 客户—服务器模型体系结构 /29
- 3.3 客户—服务器模型的进程通信 /32
  - 3.3.1 进程通信中客户—服务器模型的实现方法 /32
  - 3.3.2 客户—服务器模型的进程通信协议 /34
- 3.4 客户—服务器端模型的变种 /37
  - 3.4.1 移动代码 /37
  - 3.4.2 移动代理 /38
  - 3.4.3 网络计算机 /39
  - 3.4.4 瘦客户 /40
  - 3.4.5 移动设备和自组网络 /40
- 3.5 小结 /41
- 习题 /42

## 第二篇 云计算技术

- 第4章 分布式云计算概述 /45
  - 4.1 云计算入门 /45
    - 4.1.1 云计算的定义 /45
    - 4.1.2 云计算的发展历史 /46
    - 4.1.3 云计算的特点 /48
  - 4.2 云服务 /49
    - 4.2.1 使用云平台的理由 /49
    - 4.2.2 云平台的服务类型 /50
    - 4.2.3 云平台服务的安全性 /51

- 4.2.4 云平台服务的供应商 /51
- 4.2.5 云平台服务的优势和面临的挑战 /54
- 4.3 云计算比较 /55
  - 4.3.1 集群计算和云计算 /55
  - 4.3.2 网格计算和云计算 /57
  - 4.3.3 效用计算和云计算 /59
  - 4.3.4 并行计算、分布计算和云计算 /60
- 4.4 小结 /61
- 习题 /62

**第5章 Google公司的三大技术 /63**

- 5.1 Google 文件系统 /63
  - 5.1.1 前言 /63
  - 5.1.2 设计概要 /64
  - 5.1.3 系统交互 /68
  - 5.1.4 主控服务器操作 /69
  - 5.1.5 容错和检测 /70
- 5.2 Bigtable 技术 /71
  - 5.2.1 Bigtable 简介 /71
  - 5.2.2 Bigtable 数据模型 /72
  - 5.2.3 API /73
  - 5.2.4 Bigtable 所依赖的框架 /74
  - 5.2.5 Bigtable 实现的关键 /75
  - 5.2.6 Bigtable 性能优化方案 /78
  - 5.2.7 Bigtable 应用实例 /81
  - 5.2.8 经验总结 /83
- 5.3 MapReduce 技术 /84
  - 5.3.1 前言 /84
  - 5.3.2 编程模型 /85
  - 5.3.3 实例 /85
  - 5.3.4 输入输出类型 /85
  - 5.3.5 更多实例 /86

- 5.3.6 MapReduce 执行 /86
- 5.4 小结 /89
- 习题 /89

## 第6章 Yahoo!公司的云平台技术 /91

- 6.1 什么是 PNUTS——灵活通用的表存储平台 /91
  - 6.1.1 前言 /91
  - 6.1.2 PNUTS 概述 /92
  - 6.1.3 PNUTS 的设计和性能 /93
  - 6.1.4 PNUTS 的系统结构 /93
  - 6.1.5 PNUTS 的数据存储和检索 /94
  - 6.1.6 副本和一致性 /95
  - 6.1.7 其他数据库系统功能 /97
  - 6.1.8 数据库服务 /98
- 6.2 Pig 系统简述 /98
  - 6.2.1 Pig 的定义 /98
  - 6.2.2 Pig 简介 /98
- 6.3 ZooKeeper 系统简述 /99
  - 6.3.1 什么是 ZooKeeper /99
  - 6.3.2 ZooKeeper 项目介绍 /99
- 6.4 小结 /100
- 习题 /100

## 第7章 Greenplum 数据库技术 /101

- 7.1 什么是 Greenplum /101
- 7.2 Greenplum 分析数据库 /101
- 7.3 Greenplum 数据库的体系结构 /102
  - 7.3.1 无共享大规模并行处理体系结构 /102
  - 7.3.2 Greenplum 的分段单元服务 /103
  - 7.3.3 数据分布和并行扫描 /103
  - 7.3.4 容错能力和先进的复制技术 /104

7.3.5	全局并行查询优化技术	/105
7.3.6	gNet 软件互联	/105
7.3.7	并行数据流引擎	/106
7.3.8	统一的分析处理	/107
7.3.9	基于标准,建立在开源 PostgreSQL 数据库系统之上	/107
7.4	Greenplum 的关键特性和优点	/107
7.5	小结	/108
	习题	/108
<b>第 8 章</b>	<b>Amazon 公司的 Dynamo 技术</b>	<b>/109</b>
8.1	Dynamo 初步介绍	/109
8.2	Dynamo 的背景资料	/110
8.2.1	系统的假设和需求	/110
8.2.2	服务层协议	/111
8.2.3	设计考虑因素	/112
8.3	Dynamo 系统体系结构	/113
8.3.1	系统接口	/114
8.3.2	分割算法	/114
8.3.3	复制	/115
8.3.4	数据版本	/116
8.3.5	Dynamo 中的 get()和 put()的操作	/118
8.3.6	临时性故障处理	/118
8.3.7	处理永久的错误:同步复制	/119
8.3.8	成员关系和故障检测	/119
8.3.9	增加/删除存储结点	/120
8.4	小结	/121
	习题	/121
<b>第 9 章</b>	<b>IBM 公司的云计算技术</b>	<b>/122</b>
9.1	IBM 公司的云计算概述	/122

- 9.2 云风暴 /123
- 9.3 智能商业服务 /124
- 9.4 智慧地球计划 /124
- 9.5 Z系统 /125
- 9.6 虚拟化的动态基础架构技术 /126
  - 9.6.1 虚拟化 /127
  - 9.6.2 虚拟化的云计算技术 /128
  - 9.6.3 实现虚拟化策略的关键解决方案 /129
- 9.7 小结 /132
- 习题 /132

### 第三篇 分布式云计算的程序开发

#### 第10章 Hadoop 系统 /135

- 10.1 Hadoop 简介 /135
  - 10.1.1 Hadoop 系统的由来 /135
  - 10.1.2 Hadoop 的作用 /135
  - 10.1.3 Hadoop 的优势 /136
  - 10.1.4 Hadoop 应用现状及发展趋势 /136
- 10.2 Hadoop 项目及其结构 /137
- 10.3 Hadoop 体系结构 /139
- 10.4 Hadoop 集群安全策略 /141
- 10.5 小结 /143

#### 第11章 MapReduce 详解 /145

- 11.1 MapReduce 简介 /145
- 11.2 MapReduce 计算模型 /146
  - 11.2.1 MapReduce Job /146
  - 11.2.2 Hadoop 中 hello world 程序 /147



- 11.2.3 运行 MapReduce 应用程序 /150
- 11.2.4 新的 API /152
- 11.2.5 MapReduce 的数据流和控制流 /154
- 11.3 MapReduce 工作机制 /155
  - 11.3.1 MapReduce 作业的执行流程 /155
  - 11.3.2 提交作业 /156
  - 11.3.3 初始化作业 /158
  - 11.3.4 分配任务 /160
  - 11.3.5 执行任务 /162
  - 11.3.6 更新任务执行进度和状态 /163
  - 11.3.7 完成作业 /164
- 11.4 开发 MapReduce 应用程序 /164
  - 11.4.1 系统参数的配置 /165
  - 11.4.2 配置开发环境 /167
  - 11.4.3 编写 MapReduce 程序 /168
  - 11.4.4 本地测试 /170
  - 11.4.5 在集群上运行 /170
- 11.5 小结 /173

## 第 12 章 HDFS 详解 /174

- 12.1 HDFS 简介 /175
- 12.2 HDFS 的相关概念 /176
- 12.3 HDFS 的体系结构 /177
- 12.4 HDFS 的基本操作 /179
  - 12.4.1 HDFS 的命令行操作 /179
  - 12.4.2 HDFS 的 Web 界面 /180
- 12.5 HDFS 中的读写数据流 /180
  - 12.5.1 文件的读取 /180
  - 12.5.2 文件的写入 /182
  - 12.5.3 一致性模型 /183
- 12.6 小结 /184