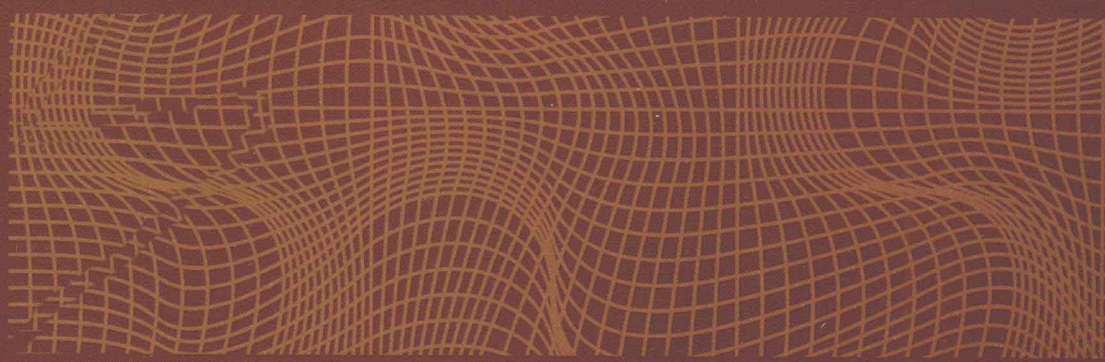
现代信息科学与技术基础

分布计算系统
（第三版）

Distributed Computing Systems

胡　亮　徐高潮　魏晓辉



高等教育出版社
HIGHER EDUCATION PRESS

现代信息科学与技术基础

分布计算系统

FENBU JISUAN XITONG

（第三版）

Distributed Computing Systems

胡　亮　徐高潮　魏晓辉

高等教育出版社·北京
HIGHER EDUCATION PRESS　BEIJING

内容提要

本书主要介绍分布计算系统的结构和实现技术。全书共分14章，主要讲述分布计算系统的基本概念、体系结构及重点设计问题，命名系统，通信，安全和保护，同步和并发控制，容错，多副本数据管理，资源管理与调度，分布式文件系统，分布式共享存储器，分布式程序设计语言，集群系统，网格系统，云计算系统等。

本书可作为高等院校本科高年级学生和研究生参考教材，也可供有关科技人员参考。

图书在版编目(CIP)数据

分布计算系统 / 胡亮，徐高潮，魏晓辉著. --3版.
-- 北京：高等教育出版社，2012.1
ISBN 978-7-04-034541-4

Ⅰ.①分… Ⅱ.①胡… ②徐… ③魏… Ⅲ.①分布式计算机系统 Ⅳ.①TP338.8

中国版本图书馆CIP数据核字(2011)第272078号

策划编辑 刘建元　　责任编辑 陈红英　　封面设计 李卫青　　责任印制 张泽业

出版发行 高等教育出版社
社　　址 北京市西城区德外大街4号
邮政编码 100120
印　　刷 北京市文林印务有限公司
开　　本 787×1092　1/16
印　　张 31.25
字　　数 670千字
购书热线 010-58581118
咨询电话 400-810-0598

网　　址 http://www.hep.edu.cn
　　http://www.hep.com.cn
网上订购 http://www.landraco.com
　　http://www.landraco.com.cn

版　　次 2003年12月第1版
　　2012年2月第3版
印　　次 2012年1月第1次印刷
定　　价 59.00元

物料号 34541-00

序

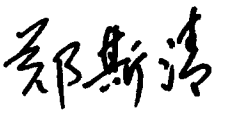
20世纪90年代以来，计算机硬件技术、软件技术以及高性能计算机网络技术的快速发展极大地拓展了分布计算的应用领域，推动了分布计算领域的科研活动，深刻地改变了人们使用计算机的方式。

分布式计算是一门计算机科学，它研究如何把一个需要非常巨大的计算能力才能解决的问题分成许多小的部分，然后把这些部分分配给许多计算机进行处理，最后把这些计算结果综合起来得到最终的结果。通过互联网使用世界各地的闲置计算能力可用来解决需要非常巨大的计算能力才能解决的问题，如分析外太空的电信号，寻找超过1000万位数字的梅森质数，寻找并发现对抗癌症更有效的药物等。

随着大规模计算能力和存储能力的需求不断增加，出现了许多新的分布式计算技术以及商业应用模式，网络计算与云计算是当前最受关注的分布计算模式。网络计算通过利用大量异构计算机的未用资源，将其作为嵌入在分布式电信基础设施中一个虚拟的计算机集群，为解决大规模的计算问题提供了一个模型。网格计算的焦点在于支持跨管理域计算的能力，这使它与传统的计算机集群或传统的分布式计算相区别。而云计算是网络计算、分布式计算、并行计算、效用计算、网络存储、虚拟化等传统计算机和网络技术发展融合的产物，是这些计算机科学技术的商业实现，是一种基于因特网的超级计算模式。通过使计算分布在大量的分布式计算机上，企业数据中心的运行将更与互联网相似。这使得企业能够将资源切换到需要的应用上，根据需求访问计算机和存储系统。

目前，国内系统阐述网格计算和云计算的文献日益增多，关于分布计算系统方面的专著却不多。《分布计算系统》是胡亮教授及其所领导的团队多年来潜心集群计算、分布计算、网格计算方面的研究成果，对分布计算系统的基本概念和体系结构、分布计算系统的进程通信、分布式程序设计语言、命名与保护、分布式同步和互斥机构、死锁问题及其处理技术、分布式数据管理、分布式文件系统、分布式调度等方面进行了深入浅出、全面系统的论述，还特别补充了网格计算与云计算相关领域的最新进展，对分布式计算、网格计算和云计算所涉及的共性技术内容进行了较为深入的讨论，是多年以来国内有关分布式计算的权威专著。本书的再次出版对有志于分布式计算及相关学科研究的科技工作者和研究生大有裨益。作为一门传统基础学科的著作，本书的工作与时俱新，具有开拓性。本书的

出版对国内分布式计算及其后续的研究必将起到很大的促进作用。



美国德克萨斯大学达拉斯分校
计算机科学、计算机工程、电信工程教授
2011 年 11 月 24 日于美国

第三版前言

由于计算机技术的快速发展、高速网络的快速普及以及分布计算系统应用领域的日益扩大，人们对分布计算系统的研究保持了长久的热情，研究兴趣日益浓厚。经过近30年的快速发展，分布计算系统的理论体系和相关技术已经日臻成熟，人们在此领域的研究工作不断深入和拓展，不断取得新的研究成果，进一步促进了该领域的快速发展，拓展了分布计算系统的应用范围。许多高等院校已经将分布计算系统作为计算机专业的核心课程。

本书的作者及其研究团队20多年来一直从事该领域的教学和科研工作，近几年在教学工作中又积累了许多新的素材，同时也积累了自己在分布计算系统方面的一些研究成果，为了反映这一领域最新的研究成果，本书在前两版的基础上增加了集群系统、网格系统和云计算系统三章内容，并对一些较为陈旧的内容做了删减，同时充实了一些新的内容，如分布计算系统的安全性。

为了保证本书内容的正确性、准确性和先进性，本书选材的原则是：对于分布计算系统方面的基础性、关键性技术，尽量选用权威性的原著作为参考资料，包括《Distributed System:Principles and Paradigms》(Andrew S. Tanenbaum 著，Prentice Hall，2007)、《分布式系统设计》(吴杰著，高传善译，机械工业出版社，2010)等；对于分布计算系统方面的最新成果和发展趋势，尽量选用最近出版的著作和研究成果，包括：《基于身份的密码学》(胡亮等，高等教育出版社，2011)、《云计算》(刘鹏，电子工业出版社，2010)、著名期刊和国际会议论文等。在此，我们对所引用著作的作者和出版者深表感谢。

本书主要介绍分布计算系统的结构和实现技术，侧重于讲授基本概念、基本原理和基本方法。全书共分为14章，主要讲述分布计算系统的基本概念、体系结构及重点设计问题，命名系统，进程通信，安全和保护，同步和并发控制，容错技术，多副本数据管理，资源管理与调度，分布式文件系统，分布式共享存储器，分布式程序设计语言，集群系统，网格系统，以及云计算系统。

第一章，绪论，介绍分布计算系统的基本概念、体系结构及重点设计问题，主要包括分布计算系统的透明性，分层体系结构，以及命名、差错控制、资源管理、保护、同步、调度等关键设计问题，引出分布计算系统方面的基础性、关键性技术。

第二章，命名系统，主要介绍命名系统的结构和功能、名字的结构、命名空间和名字解析等方面的问题，命名问题贯穿于分布计算系统的各个逻辑层面。

第三章，分布计算系统的进程通信，包括通信的层次结构，常见的通信类型，如消息传

递和远程过程调用，组通信等问题，通信问题是分布式计算得以实现的基础性问题。

第四章，安全和保护，主要介绍加密技术、身份认证、访问控制、网络安全与信任体系等方面的内容，安全与保护问题是分布计算系统资源得以正确共享的前提。

第五章，同步和并发控制，主要介绍逻辑时钟等同步机构、各种互斥算法（包括集中式互斥算法、基于逻辑时钟的互斥算法和基于令牌的互斥算法）、死锁问题（包括死锁预防和死锁检测）、并发控制问题（包括可串行化调度、基于锁的并发控制和基于时间戳的并发控制）。同步和并发控制能够保障多个用户在使用共享资源时不会产生相互干扰的问题。

第六章，容错，包括处理故障的方式、检查点算法、拜占庭故障的恢复、原子事务处理的恢复及可靠的组通信技术。随着资源的增多，系统失效的概率会大大增加，容错技术是提高系统可用性和可靠性的关键技术保障。

第七章，多副本数据管理，包括多副本数据的一致性模型、多副本更新和一致性管理问题、复制控制算法等。多副本技术是提高数据访问性能和提高容错和容灾的有效方法。

第八章，资源管理与调度，主要介绍资源管理方式、调度算法的分类、静态调度和动态调度等问题。资源管理与调度是提高资源利用率和分布式应用程序执行性能的重要前提。

第九章，分布式文件系统，主要介绍分布式文件系统的命名问题、分布式文件系统的共享访问问题和分布式文件系统的设计问题。分布式文件系统有利于用户以透明、高效的方式共享整个系统的信息资源。

第十章，分布式共享存储器，主要包括分布式共享存储器的实现算法、缓存一致性协议和应考虑的主要实现问题。分布式共享存储器是一种有利于用户开发分布式应用的开发环境，在此环境下，许多单机环境下的并发程序只需经过少量的修改就可以在分布式环境下运行。

第十一章，分布式程序设计语言，主要介绍分布式程序设计语言对并行性的支持问题、对通信和同步的支持问题、逻辑上分布地址空间的语言和逻辑上共享地址空间的语言。分布式程序设计语言是开发分布式应用的工具。

第十二到第十四章，通过集群系统、网格计算系统和云计算系统的介绍，体现最近几年的研究成果和该领域的发展趋势。

读者应该在学习过计算机网络、计算机组成原理和操作系统等课程之后学习本课程。如果读者已学习网络程序设计课程，可以略过本书的第三章。

车喜龙博士参与了第十三章的材料收集和组稿工作，赵阔副教授参与了第四章的材料收集和组稿工作，他们为本书的出版做出了重要贡献。

清华大学郑纬民教授再次为本书审稿，提出许多宝贵意见，美国德克萨斯大学达拉斯分校郑斯清教授为本书作序，作者在此表示诚挚的谢意。

衷心希望读者指出本书错误并提出改进意见，你们的意见和建议对我们改进和提高书稿质量水平至关重要。

作　者

于吉林大学计算机科学与技术学院

2011 年 10 月

第二版前言

由于计算机技术和网络技术的快速发展以及分布计算系统的应用领域日益扩大，人们对分布计算系统的研究兴趣日益浓厚，现在，分布计算系统已经是一个非常热门的研究领域。在过去的10年中，分布计算系统的理论体系和相关技术更加成熟，许多研究者在此领域进行了大量的研究工作，并取得了大量的研究成果。在此期间，作者在教学工作中也积累了许多新的素材，同时也积累了自己在分布计算系统方面的一些研究成果，本书力求反映这些最新的研究成果。分布式应用已涉及社会生活的各个方面，越来越多高等院校的计算机系开设了分布计算系统这门课程。

本书是在1994年高等教育出版社出版的《分布计算系统》(第一版)基础上重新修订、补充和完善而成。本书所选用的参考文献主要有：《Distributed System: Principles and Paradigms》(Andrew S. Tanenbaum 著，清华大学出版社，2002年影印版)、《分布式系统设计》(吴杰著，高传善译，机械工业出版社，2001年中译本)、《Distributed Operating System: Concepts and Practice》(Doreen L. Galli 著，人民邮电出版社，影印版)、《机群计算》(鞠九滨等著，吉林大学出版社，1999年)等。为了保证本书内容的正确性、准确性和先进性，作者选用参考资料时，尽量选用了最近出版的有权威性的原著和最新的研究成果。在此，我们对所引用著作的作者和出版者深表感谢。

本书介绍了用计算机网络组成的分布计算系统的结构和实现技术，侧重于基本概念、基本原理和基本方法。全书共分为12章：分布计算系统的基本概念和体系结构、分布计算系统的进程通信、分布式程序设计语言、命名与保护、分布式同步与互斥机构、死锁问题及其处理技术、容错技术、分布式数据管理、分布式文件系统的设计问题与实现方法、分布式调度、分布式共享存储器技术以及基于对象的分布式系统。

读者应该在学过计算机网络、计算机组成原理和操作系统之后学习本课程。

清华大学郑纬民教授审阅了全稿，对本书提出了不少宝贵意见，作者在此表示诚挚的谢意。

衷心希望读者指正错误和提出改进意见。

作者

于吉林大学计算机科学与技术学院

2003年9月

第一版前言

作者从1983年开始为研究生讲授“分布计算系统”课程，已讲授过八次，每次的内容均有较大变化。本书是在这一基础上重新组织材料写成的。分布计算系统的研究仍处于非常活跃的阶段。尽管其理论体系仍处于发展时期，但很多基础部分已趋于稳定和成熟。本书介绍的内容主要是一些基本概念、基本原理和基本方法，力求反映最新研究成果。使用的参考资料尽量选用有权威性的原著，以确保本书内容的正确性、准确性和先进性。

按照国际学术界大多数人的观点，分布计算系统可以分成紧密耦合式和松散耦合式两种。在现有的分布计算系统中，绝大多数是用计算机网络（主要是局部网络）支持的松散耦合式。所以本书介绍这种系统的结构和实现，主要包括进程通信、命名和保护、资源控制、分布式文件系统、工作站调度、分布式程序设计语言以及分布式共享存储器等问题。在说明基本原理和方法时，列举的例子都是当前国际上已实现的有代表性的著名系统。最后一章中介绍了五个典型的分布式系统，以便使读者对设计分布计算系统的几个方法和分布式系统的整体结构有个清楚的了解。由于篇幅所限，本书不讨论紧密耦合分布式系统、分布式数据库以及分布式系统的应用等问题。

国家教育委员会已把分布计算系统列为高等学校计算机专业基础选修课，本书是为这一课程编写的教材。学生应在学过计算机组成原理、操作系统之后学习本课程。当然，在上本课程以前如果对计算机网络有所了解则更好。通过本课程的学习，要求学生掌握分布计算系统的概念，组成分布计算系统的网络基础，分布式系统中的进程通信、命名与保护的特点，分布式同步机构及其在互斥、并发控制、失效恢复和多副本更新中的应用，分布式文件系统的设计与实现方法，分布式程序设计语言的特点，分布式共享存储器的概念，并对国际上几个著名的分布式系统有所了解。各章中每节平均讲授两学时。

中国科学技术大学陈国良教授审阅了全稿，对本书原稿提出不少宝贵意见，作者在此表示诚挚的谢意。

衷心希望读者指正错误和提出改进意见。

鞠九滨

于吉林大学计算机科学系

1993年10月

目　录