Broadview®
www.broadview.com.cn


# 分布式系统
## 常用技术及案例分析

柳伟卫 编著

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

# 分布式系统
## 常用技术及案例分析

柳伟卫 编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

## 内 容 简 介

本书全面介绍在设计分布式系统时所要考虑的技术方案，内容丰富、案例新颖，相关理论与技术实践较为前瞻。本书不仅仅介绍了分布式系统的原理、基础理论，同时还引入了大量市面上常用的最新分布式系统技术，不仅告诉读者怎么用，同时也分析了为什么这么用，并阐述了这些技术的优缺点。希望本书可以成为读者案头的工具书，供读者随手翻阅。

本书分为三大部分，即分布式系统基础理论、分布式系统常用技术以及经典的分布式系统案例分析。第一部分主要介绍分布式系统基础理论知识，总结一些在设计分布式系统时需要考虑的范式、知识点以及可能会面临的问题，其中包括线程、通信、一致性、容错性、CAP 理论、安全性和并发等相关内容；同时讲述分布式系统的常见架构体系，其中也包括最近比较火的 RESTful 风格架构、微服务、容器技术等。第二部分主要列举了在分布式系统应用中经常用到的一些主流技术，并介绍这些技术的作用和用法；这些技术涵盖了分布式消息服务、分布式计算、分布式存储、分布式监控系统、分布式版本控制、RESTful、微服务、容器等领域的内容。第三部分选取了以淘宝网和 Twitter 为代表的国内外知名互联网企业的大型分布式系统案例，分析其架构设计以及演变过程；这部分相当于是对第二部分零散的技术点做一个“串烧”，让读者可以结合技术的理论，看到实战的效果。

本书主要面向的读者是对分布式系统感兴趣的计算机专业的学生、软件工程师、系统架构师等。



**图书在版编目（CIP）数据**

分布式系统常用技术及案例分析 / 柳伟卫编著. —北京：电子工业出版社，2017.2
ISBN 978-7-121-30771-3

Ⅰ. ①分… Ⅱ. ①柳… Ⅲ. ①分布式操作系统－研究 Ⅳ. ①TP316.4

中国版本图书馆 CIP 数据核字（2016）第 323368 号

责任编辑：陈晓猛
印　　刷：北京京科印刷有限公司
装　　订：三河市良远印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：787×980　1/16　　印张：43.75　　字数：980 千字
版　　次：2017 年 2 月第 1 版
印　　次：2017 年 5 月第 2 次印刷
定　　价：99.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819　faq@phei.com.cn。

# 前　言

## 写作背景

我一直想写一本关于分布式系统方面的书。一方面是想把个人多年工作中涉及的分布式技术做一下总结，另一方面也想把个人的经验分享给广大的读者朋友。由于我的开发工作大都以Java为主，所以一开始的主题设想是“分布式Java”，书也以开源方式发布在互联网上（网址为https://github.com/waylau/distributed-java）。

后来，陈晓猛编辑看到了这本开源书，以及我关于分布式系统方面的博文，问我是否有兴趣出版分布式相关题材的图书。当然书的内容不仅仅是“分布式Java”。

对于出书一事，我犹豫良久。首先，本身工作挺忙，实在无暇顾及其他；其次，虽然我之前写过超过一打的书籍（可见https://waylau.com/books/），但多是开源电子书，时间、内容方面自然也就不会有太多约束，几乎是“想写就写，没有时间就不写”，这个跟正式出版还是存在比较大的差异的；最后，这本书涉及面相对较广，需要查阅大量资料，实在是太耗费精力。

但陈晓猛编辑还是鼓励我能够去尝试做这个事情。思索再三，于是我便答应。当然，最后这本书还是在规定时间内完成了。它几乎耗尽了我写作期间所有的业余和休息时间。

“不积跬步，无以至千里；不积小流，无以成江海。”虽然整本书从构思到编写完成的时间不足一年，但书中的大部分知识点，却是我在多年的学习、工作中积累下来的。之所以能够实现快速写作，一方面是做了比较严格的时间管理，另一方面也得益于我多年坚持写博客和开源书的习惯。

## 内容介绍

本书分为三大部分，即分布式系统基础理论、分布式系统常用技术以及经典的分布式系统案例分析。第一部分为第1章和第2章，主要介绍分布式系统基础理论知识，总结一些在设计分布式系统时需要考虑的范式、知识点以及可能会面临的问题。 第二部分为第3章到第8章，主要列举了在分布式系统应用中经常用到的一些主流技术，并介绍这些技术的作用和用法。第

三部分为第 9 章和第 10 章，选取了以淘宝网和 Twitter 为代表的国内外知名互联网企业的大型分布式系统案例，分析其架构设计以及演变过程。

**第 1 章**介绍分布式系统基础理论知识，总结一些在设计分布式系统时需要考虑的范式、知识点以及可能会面临的问题，其中包括线程、通信、一致性、容错性、CAP 理论、安全性和并发等相关内容。

**第 2 章**详细介绍分布式系统的架构体系，包括传统的基于对象的体系结构、SOA，也包括最近比较火的 RESTful 风格架构、微服务、容器技术、Serverless 架构等。

**第 3 章**介绍常用的分布式消息服务框架，包括 Apache ActiveMQ、RabbitMQ、RocketMQ、Apache Kafka 等。

**第 4 章**介绍分布式计算理论和应用框架方面的内容，包括 MapReduce、Apache Hadoop、Apache Spark、Apache Mesos 等。

**第 5 章**介绍分布式存储理论和应用框架方面的内容，包括 Bigtable、Apache HBase、Apache Cassandra、Memcached、Redis、MongoDB 等。

**第 6 章**介绍分布式监控方面常用的技术，包括 Nagios、Zabbix、Consul、ZooKeeper 等。

**第 7 章**介绍常用的分布式版本控制工具，包括 Bazaar、Mercurial、Git 等。

**第 8 章**介绍 RESTful API、微服务及容器相关的技术，着重介绍 Jersey、Spring Boot、Docker 等技术的应用。

**第 9 章和第 10 章**分别介绍以淘宝网和 Twitter 为代表的国内外知名互联网企业的大型分布式系统案例，分析其架构设计以及演变过程。

## 源代码

本书提供源代码下载，下载地址为 https://github.com/waylau/distributed-systems-technologies-and-cases-analysis。

## 勘误和交流

本书如有勘误，会在 https://github.com/waylau/distributed-systems-technologies-and-cases-analysis 上进行发布。由于笔者能力有限，时间仓促，难免错漏，欢迎读者批评指正。读者也可以到博文视点官网的本书页面进行交流（www.broadview.com.cn/30771）。

您也可以直接联系我：

博客：https://waylau.com

邮箱：waylau521@gmail.com

微博：http://weibo.com/waylau521

开源：https://github.com/waylau

# 致谢

首先，感谢电子工业出版社博文视点公司的陈晓猛编辑，是您鼓励我将本书付诸成册，并在我写作过程中审阅了大量稿件，给予了我很多指导和帮助。感谢工作在幕后的电子工业出版社评审团队对于本书在校对、排版、审核、封面设计、错误改进方面所给予的帮助，使本书得以顺利出版发行。

其次，感谢在我十几年求学生涯中教育过我的所有老师。是你们将知识和学习方法传递给了我。感谢我曾经工作过的公司和单位，感谢和我一起共事过的同事和战友，你们的优秀一直是我追逐的目标，你们所给予的压力正是我不断改进自己的动力。

感谢我的父母、妻子 Funny 和两个女儿。由于撰写本书，牺牲了很多陪伴家人的时间。感谢你们对于我工作的理解和支持。

最后，特别要感谢这个时代，互联网让所有人可以公平地享受这个时代的成果。感谢那些为计算机、互联网所做出贡献的先驱，是你们让我可以站在更高的“肩膀”上！感谢那些为本书提供灵感的佳作，包括《分布式系统原理与范型》《Unix Network Programming》《Enterprise SOA》《MapReduce Design Patterns》《Hadoop: The Definitive Guide》《Learning Hbase》《Advanced Analytics with Spark》《Pro Git》《Docker in Action》《淘宝技术这十年》《Hatching Twitter》，等等，详细的书单可以参阅本书最后的“参考文献”部分。

柳伟卫

2016 年 11 月 13 日于杭州

# 目　录


# 第 1 章
# 分布式系统基础知识

# 1.1 概述

自从有了计算机，人类的生活就发生了巨大的变化，原来人力需要数年才能解决的计算问题，计算机“分分钟”就搞定了。回顾历史，在 19 世纪 50 年代，计算机在诞生初期，主要用于军事方面，而天才的数学家图灵建造了计算机来破译德军的“英格玛（Enigma）”密码系统。那时候的计算机如一个庞大笨重的巨人，足足占据了好几间房子的空间。在 19 世纪 80 年代，计算机的体积越来越小，开始被广泛应用于科研计算，但那时候计算机的价格仍然是极其昂贵的，只有少数几个科研单位或者大学才能买得起。今天，个人计算机或者以手机为代表的智能设备已经走进寻常百姓家了。每个人几乎都拥有手机，手机不仅仅是通信工具，还能发语音、看视频、玩游戏，让人与人之间的联系变得更加紧密。智能手环随时监控你的身体状况，并根据你每天的运动量、身体指标来给你提供合理的饮食运动建议。出门逛街甚至不需要带钱包了，吃饭购物搭车时使用手机就可以支付费用，多么方便快捷。智能家居系统更是你生活上的“管家”，什么时候该睡觉了，智能家居系统就自动拉上窗帘，关灯；早上起床了，智能家居系统会自动拉开窗帘，并播放动人的音乐，让你可以愉快地享受新的一天的来临；你再也不用担心家里的安全情况，智能家居系统会帮你监控一切，有异常情况时会及时发送通知到你的手机，让你第一时间掌握家里的状况。

毫无疑问，计算机改变了人类的工作和生活方式，而计算机系统也正在进行一场变革。没错，任何一个手机应用，或者智能 App，都离不开背后那个神秘的巨人——分布式系统。正是那些看不见的分布式系统，每天处理着数以亿计的计算，提供可靠而稳定的服务。

本章就揭开分布式系统的神秘面纱。

## 1.1.1 什么是分布式系统

《分布式系统原理与范型》一书中是这样定义分布式系统的：

“分布式系统是若干独立计算机的集合，这些计算机对于用户来说就像单个相关系统”。

这里面包含了 2 个含义：

- 硬件独立；
- 软件统一。

什么是硬件独立？所谓硬件独立，是指计算机机器本身是独立的。一个大型的分布式系统，会由若干台计算机来组成系统的基础设施。而软件统一，是指对于用户来说，用户就像是跟单个系统打交道。就好比我们每天上网看视频，视频网站对我们来说就是一个系统软件，它们背后是如何运作的，部署了几台服务器，每台服务器是干什么的，这些对用户来说是透明的，不

可见的。用户不关心背后的这些服务器，用户所关心的是，今天访问的这个网站能提供什么样的节目，视频运行是否流畅，卡不卡帧、清晰度如何等。

而软件统一的另外一个方面是指，分布式系统的扩展和升级都比较容易。分布式系统中的某些节点发生故障，不会影响整体系统的可用性。用户和应用程序交互时，不会察觉哪些部分正在替换或者维修，也不会感知到新加入的部分。

## 1.1.2 集中式系统 VS.分布式系统

集中式系统主要部署在 HP、IBM、SUN 等小型机以上档次的服务器中，把所有的功能都集成到主服务器上，这样对服务器的要求很高，性能也极其苛刻。它们的主要特色在于年宕机时间只有几小时，所以又统称为 z 系列（zero，零）。AS/400 主要应用在银行和制造业，还用于 Domino，主要的技术在于 TIMI（技术独立机器界面）、单级存储，有了 TIMI 技术可以做到硬件与软件相互独立。RS/6000 比较常见，用于科学计算和事务处理等。这类主机的单机性能一般都很强，带多个终端。终端没有数据处理能力，运算全部在主机上进行。现在的银行系统大部分都是这种集中式的系统，此外，在大型企业、科研单位、军队、政府等也有应用。

集中式系统主要流行于 20 世纪。现在还在使用集中式系统的，很大一部分原因是为了沿用原来的软件，而这些软件往往很昂贵。优点是便于维护，操作简单。但这样的系统也有缺陷，就是不出问题还好，一出问题，就会造成单点故障，所有功能就都不能正常工作了。由于集中式的系统相关的技术只被少数厂商所掌握，个人要对这些系统进行扩展和升级往往也比较麻烦，一般的企业级应用很少会用到集中式系统。图 1-1 是一个典型的集中式系统的示意图。

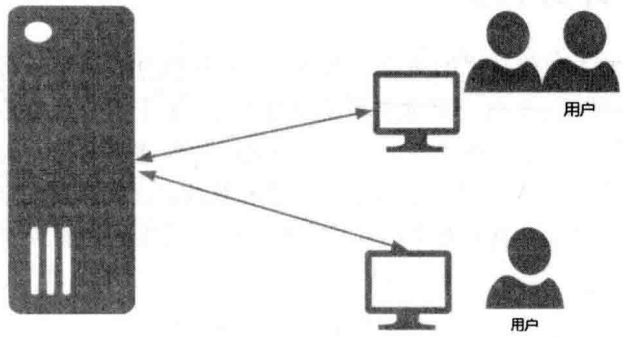


图 1-1 集中式系统示意图

而分布式系统恰恰相反。分布式系统是通过中间软件来对现有计算机的硬件能力和相应的软件功能进行重新配置和整合。它是一种多处理器的计算机系统，各处理器通过互连网络构成统一的系统。系统采用分布式计算结构，即把原来系统内中央处理器处理的任务分散给相应的处理器，实现不同功能的各个处理器相互协调，共享系统的外设与软件。这样就加快了系统的