HZ BOOKS
华章科技

Linux/Unix
技术丛书

构建高可用Linux服务器

（第3版）

余洪春 著

Build High Availability Linux Servers
Third Edition

- Linux运维领域公认经典畅销书，曾被《程序员》杂志和51CTO等权威IT媒体评为“10大最具技术影响力的图书”和“最受读者喜爱的原创图书”，运维工程师必备工具书
- 基于实际生产环境，从Linux服务器构建与优化、高可用Linux集群构建、MySQL高可用架构设计、Puppet自动化运维等多角度讲解了构建高可用Linux服务器的方法和技巧

机械工业出版社
China Machine Press

构建高可用Linux服务器

（第3版）

Build High Availability Linux Servers
Third Edition

余洪春 著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

构建高可用 Linux 服务器 / 余洪春著 . —3 版 . —北京：机械工业出版社，2014.9
（Linux/Unix 技术丛书）

ISBN 978-7-111-47787-7

I. 构… II. 余… III. UNIX 操作系统 – 网络服务器 IV. TP316.81

中国版本图书馆 CIP 数据核字（2014）第 204103 号

构建高可用 Linux 服务器

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：孙海亮　　责任校对：殷　虹
印　　刷：北京市荣盛彩色印刷有限公司　　版　　次：2014 年 10 月第 3 版第 1 次印刷
开　　本：186mm × 240mm　1/16　　印　　张：27.25
书　　号：ISBN 978-7-111-47787-7　　定　　价：79.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjg@hzbook.com

Praise 本书赞誉

如何构建高可用的 Linux 服务器，这是很多 Linux 系统管理员和运维人员都感兴趣的话题，这也是他们努力学习的方向之一。本书作者在该领域纵横多年，在大量的一线生产环境中积累了丰富的经验和最佳实践。难能可贵的是，他将这些宝贵的知识系统梳理后总结在了这本书中，旨在与所有同行分享。如果你是一位 Linux 系统管理员，或运维工程师，或项目实施工程师，只要你细心研读本书的内容并跟随书中的案例去实践，相信一定会受益匪浅。

——刘天斯　腾讯系统架构师

在 IT 领域，好书很多，烂书也不少，毫无疑问，本书是一本好书。建议大家用最强悍的执行力来学习本书中的案例，用最细腻的心思去揣摩每个案例背后的原理。如果你能吃透这本书，那么你就能在短短几个月的时间内将作者几年来积累的知识和经验化为己有，从而在短时间内使自己的技能产生质的飞跃。

——曹亚孟　合力金软运维经理

本书的内容全部来自于企业的实际生产环境，非常注重实践性和实用性，书中所有案例都可以供大家在解决实际问题时参考和借鉴。本书从 Linux 服务器的构建、生产环境下服务器的故障诊断与排除、生产环境下的 Shell 脚本、高可用 Linux 集群建设、MySQL 性能调优与高可用、自动化运维、VPN 在企业中的部署应用、Linux 防火墙等多个方面阐述了构建高可用 Linux 服务器的方法与最佳实践。强烈推荐！

——崔晓辉　大众网高级系统管理员

本书作者从事 Linux 运维相关的工作已近 10 年，不仅主导过多个 Linux 集群相关项目，还从事过 Linux 教学方面的工作，积累了相当丰富的经验。利用工作之余，他把自己多年来积累的实践经验整理到了这本书中，可谓是精华中的精华！本书结合一线生产环境的真实案例讲解了 Linux 集群、MySQL 性能优化、自动化运维和系统安全相关的内容，同时还包括一些常见问题和故障的排除方法。本书尤其适合那些已经有 2～3 年 Linux 服务器管理与运维经验的读者，相信本书能在工作中助你们一臂之力。

——侯心刚　巨人网络运维中心运维部经理

前言 Preface

我的系统架构师之路

2004 年我初识 UNIX 开源系统。那会儿我正在一家大型国有企业做系统管理员，负责值守公司的 Windows Server 2000 服务器。当时“震荡波”和“冲击波”这两种病毒很猖狂，虽然我们在防毒方面投入了大量的精力和金钱（当时购买的都是正版 Windows 2000 系统和正版瑞星杀毒软件），但新上线的机器，偶尔也有遗漏的时候，没有打补丁的机器无一幸免，所以对这个问题比较头疼。有一次去朋友的公司（某省太平洋寿险下面的一个分支机构）参观，我发现他们的服务器和终端系统都很奇怪，只有纯字符界面，没有任何图形界面。一问才知道是 UNIX 系统，因为运行机制不一样，所以 Windows Server 下的病毒丝毫影响不了它们。而且这些服务器很稳定，基本上不宕机。当时很是羡慕，心想要是哪一天我们的服务器也换成 UNIX 系统，那“冲击波”和“震荡波”就奈何不了我们的服务器了，而且也不会有蓝屏现象，那该多好。

后来有幸到北京一家大型广告公司上班，公司的核心业务是 CDN 系统，所用的服务器基本上都是 CentOS 系统，内部用的文件服务器是 Samba，Web 服务器是 Apache 和 Nginx，NAT 路由器是 iptables，装的几乎全部是 CentOS 5.1 x86_64，仅有一台服务器安装的是 Windows Server 2003（供程序员开发 .NET 程序之用）。公司的这套 CDN 分布式系统要负责处理所有的流量，即使在高峰期 PV 和并发量特别大的时候，网站也非常稳定。

我当时对 CentOS 系统产生了浓厚的兴趣，尝试改掉自己多年使用 Windows 的习惯，换成了纯字符操作，用 Vim 编辑 CentOS 下面的配置文件，并且尝试用 Shell 完成自动化工作。渐渐地，我发现自己越来越喜欢 CentOS 系统了。公司有一台 vsftpd 服务器，3 年没有重启了，这很令人吃惊。后来我又得知另外一位从事 Linux 运维工作的朋友所在的公司有一台很老的 RH9 服务器，因为负责的是公司的核心业务，已经 9 年没有重启了，当时更加感到震撼，这更加坚定了我要学好 Linux/UNIX 的决心。

后来在担任项目实施工程师期间，接触到一些客户，他们的核心网站的并发量并不是太大，但比较重要，所以他们都要求部署 Linux 集群，有时指定要部署 LVS 或 HAProxy 负载均衡器。在项目

实施的过程中，我发现 LVS/HAProxy 的负载均衡能力确实非常强大，其能力可以与硬件级的 F5 负载均衡器相媲美。很快我就被 Linux 集群这门艺术迷住了，我自己也研究了 Nginx + Keepalived 这种负载均衡高可用架构，并且在许多项目中成功实施，客户反映效果也不错，所以我开始在很多开源社区推广这些技术。

现阶段我的职务是公司的系统架构师，主要工作是设计、实施及维护本公司的电子商务网站，相对于 CDN 分布式系统而言，它没有节点冗余，所以对 Linux 集群技术的要求更高。我前期对所有的网站应用都做了双机高 HA、LVS/HAProxy + Keepalived 和 Nginx + Keepalived，以及 DRBD + Heartbeat + NFS 文件高可用，MySQL 数据库用的是 DRBD 双主多从架构。后期随着流量和规模日益增大，新机器上线也日益频繁，我采用了 Puppet 自动化运维来管理线上机器，避免重复劳动。另外，由于电子商务网站涉及支付的问题，所以对安全性的要求非常高，我们平时都会从网络安全（硬件防火墙）、系统安全、代码安全和数据库安全这些方面着手，尽力避免一切影响网站安全的行为。虽然工作辛苦，但看着自己架构的网站顺利稳定运行，心里还是很有成就感的，这也是我目前工作的主要动力。

撰写此书的目的

从事系统集成/管理/架构方面的工作已经有 9 年了，在工作期间，我曾经有幸担任了一段时间的红帽 RHCE 讲师，到东北大学等高校推广红帽 Linux 系统。在教学过程中我发现，很多学生在进入企业后都无法胜任自己的工作，更谈不上正确地规划自己的职业道路了。一方面因为企业的生产环境具有一定的复杂性和危险性，另一方面市场上入门书居多，缺乏能真正指导读者解决实际问题的书籍。例如，很多书都只是给出了比较基础的操作及理论，而相对于线上环境，根本没有涉及如何安全操作、如何避免误操、PV、UV、并发数、磁盘 I/O 压力及数据库压力等相关话题。

之所以写这本书，一方面是想对自己这些年的工作经验和心得进行一次系统的梳理和总结；另一方面是想将自己的经验分享给大家，希望能帮助大家少走弯路。通过本书中的项目实践（包括 Linux 集群、MySQL 高可用方案及 Puppet 自动化运维工具的使用）和线上环境的 Shell 脚本，大家能迅速进入工作状态。书中所提供的 Shell 脚本和 iptables 脚本均来自于线上的生产服务器，大家均可直接拿来用。关于 Linux 集群的项目实践和 MySQL 的高可用方案，大家也可以根据实际项目的需求直接采用，以此来设计自己公司的网站架构。

希望大家能通过本书掌握 Linux 的精髓，轻松而愉快地工作，从而提高自己的技术水平，这是我非常希望看到的，这也是我写本书的初衷。

第 3 版与第 2 版的区别

本书是第 3 版，相对于第 1 版和第 2 版而言改动比较大，第 3 版涉及的 Linux 服务器系统以现在主流的 CentOS 5.8 x86_64 为主（第 2 版以 CentOS 5.5 为主）。在写作过程中吸收了读者对上一版

本的许多意见和建议，继续修正第2版的排版错误、人为错误及其他问题。如果大家能够完全掌握第3版新增的章节和内容，相信无论是在平常的自动化运维工作方面还是系统架构设计方面都会有自己的认识和见解了。

具体改动如下：

- 考虑到XEN虚拟化目前应用范围不是特别大，所以删除了第2版第2章Linux服务器虚拟化章节，部分内容并进了附录。
- 考虑到目前企业中多采用商业版的邮件系统，所以删除了第2版第8章如何构建开源免费的企业级邮件系统。
- 限于篇幅，再加上编辑希望第3版是一本纯粹的技术书籍的原因，删除了第2版第9章系统管理员在企业中的职业定位及发展方向，部分内容会在我的个人博客发出。
- 由于现在MySQL在互联网项目中的比重日益增大，所以我特意将第2版第5章构建高可用Linux集群中的MySQL部分重新整理，增加了MySQL性能调优级高可用案例分享章节；考虑到第2版中MySQL双主多从高可用配置方案的受众群范围较小（游戏行业），这里用线上的DRBD + Heartbeat双机高可用方案来代替；还增加了利用sysbench对磁盘I/O作性能测试等新内容，对MySQL有兴趣的读者朋友也可以重点关注这一章节。
- 构建高可用Linux集群章节增加了千万级PV网站系统架构拓朴图，限于篇幅，集群章节的内容暂时没有考虑前端有CDN的系统架构方案。
- 另外，现在自动化运维是系统运维的流行趋势，所以增加了分布式自动化部署管理工具Puppet章节，对负载均衡技术有兴趣的朋友可以关注Nginx在Puppet部署中的应用。

读者对象

本书的读者对象包括：

- 系统管理员或系统工程师
- 网络管理员或企业网管
- 项目实施工程师
- 开发人员

如何阅读本书

本书的内容是对实际工作经验的总结，涉及大量的知识点和专业术语，建议经验还不是很丰富的读者先了解第1章的内容，这一章比较基础，如果大家在学习过程中根据这章的讲解进行操作，定会达到事半功倍的效果。

系统管理员和系统工程师们则可以重点关注第3章、第4章、第5章及第6章的内容，这些都与日常工作息息相关的，建议大家多花些精力和时间，抱着一切从线上环境考虑的态度去学习。

对于网络管理员来说，如果基础不扎实，建议先学习第1～3章的内容，然后将重点放在第7

章和第 8 章上。

对于项目实施工程师而言，由于大多数都是从事系统集成相关工作的，因此建议顺序学习全书的内容，重心可以放在第 4 章和第 5 章上。

对于开发人员来说，由于其只需对系统有一个大概的了解，重点可以放在第 1 章和第 3 章上。

大家可以根据自己的职业发展和工作需要选择不同的阅读顺序和侧重点。

关于勘误

尽管我花了大量时间和精力去核对文件和语法，但书中难免还会存在一些错误和纰漏，如果大家发现问题，希望及时反馈给我，相关信息可发到我的邮箱 yuhongchun027@ gmail. com。尽管我无法保证每一个问题都会有正确的答案，但我肯定会努力回答并且指出一个正确的方向。

如果大家对本书有任何疑问或者想与我进行 Linux 方面的技术交流，可以访问我的个人博客：http://yuhongchun. blog. 51cto. com。另外，我在 51CTO 和 CU 社区的用户名均为“抚琴煮酒”，大家也可以直接通过此用户名与我在社区进行交流。

致谢

感谢我的家人，他们在生活上对我的照顾无微不至，让我有更多精力和动力去工作和创作。

感谢东北大学信息技术学院的付冲教授，感谢他在我生活困难的时候伸出援手。

感谢老男孩在网站架构设计方面给出的专业性指导意见，他的经验和专业知识让我受益匪浅。

感谢朋友刘鑫，他和我一起花了大量时间研究和调试 HAProxy + Keepalived。

感谢朋友胡安伟，他为本书提供了许多精美的插图，并就 Linux 集群相关内容提出了许多宝贵的意见。

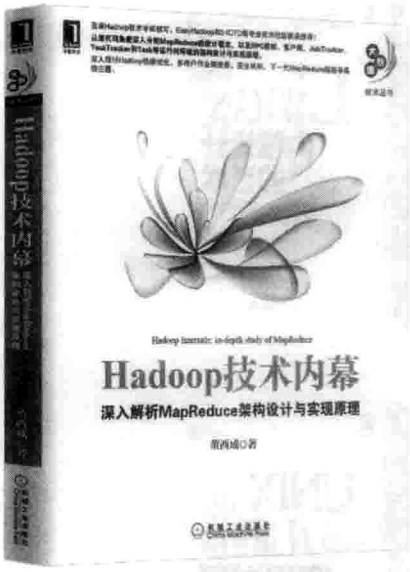
感谢朋友崔晓辉、曹孟亚及同事 ritto. zhao，崔晓辉和 ritto 为本书提供了大量的线上 Shell 脚本。

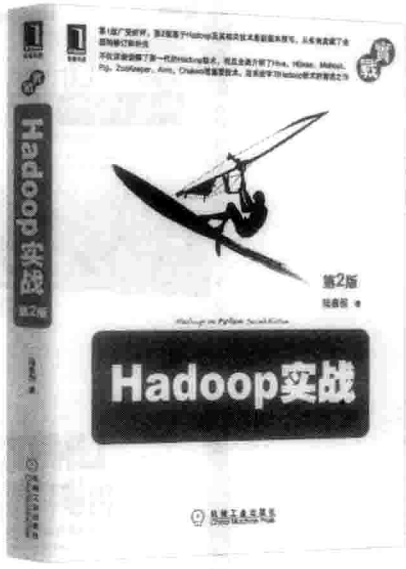
感谢 51CTO 的编辑们，尤其是赵克衡、杨赛、张浩，感谢你们的信任和帮助，没有你们，就不会有本书的面世，更不会延续到第 3 版。

感谢朋友三宝这么多年来对我的信任和支持，是他在我苦闷的时候陪我聊天，从始至终都支持我和信任我。

感谢在工作和生活中给予我帮助的所有人，感谢你们，正是因为有了你们，本书才能问世。

推荐阅读





Hadoop技术内幕：深入解析MapReduce架构设计与实现原理

资深Hadoop技术专家撰写，EasyHadoop和51CTO等专业技术社区联袂推荐！

从源代码角度深入分析MapReduce的设计理念，以及RPC框架、客户端、JobTracker、TaskTracker和Task等运行时环境的架构设计与实现原理。

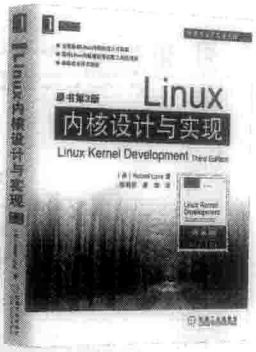
深入探讨Hadoop性能优化、多用户作业调度器、安全机制、下一代MapReduce框架等高级主题。

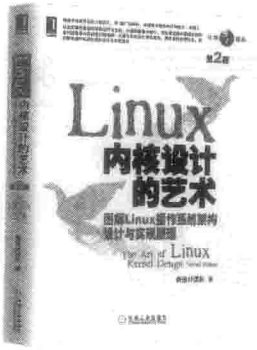
Hadoop实战（第2版）

畅销书，第1版广受好评，第2版基于Hadoop及其相关技术最新版本撰写，从多角度做了全面的修订和补充。

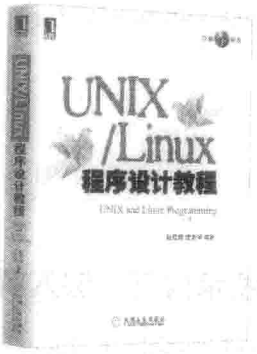
不仅详细讲解了新一代的Hadoop技术，而且还全面介绍了Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa等重要技术，是系统学习Hadoop技术的首选之作。

推荐阅读

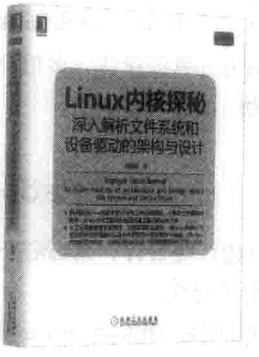

Linux
内核设计与实现
Linux Kernel Development


Linux
内核设计
的艺术


LINUX
内核精髓
O'REILLY


UNIX
/Linux
程序设计教程


Linux
高性能服务器编程


Linux内核探秘
深入解析文件系统和
设备驱动的架构与设计

目　录 *Contents*

第 1 章 Chapter 1

Linux 服务器的构建与优化

在从事目前的系统架构师工作之前，很长一段时间我从事的是系统管理员/高级系统管理员工作。在企业日常运营中，我的工作涉及的内容主要有电子商务网站的运维、内网开发环境的部署、公司外包项目的实施等。在这些工作中，我用到的系统绝大多数是免费开源的 CentOS 5.8 x86_64 系统，它的稳定和高效令我印象深刻。本章将以 CentOS 5.8 x86_64 的生产服务器为平台，逐步介绍它的 Kickstart 无人值守安装、网络配置、日志分析、性能状态监控，以及它的最小化优化等内容，这些都是构建高性能及高可用的 Linux 系统的基础，希望对大家有所帮助。

1.1 使用 PXE + DHCP + Apache + Kickstart 无人值守安装 CentOS 5.8 x86_64

CentOS 5.8 x86_64 的安装方法挺多的，最常见的有光盘安装、Kickstart 无人值守安装、优盘安装及 ISO 硬盘安装等。现阶段的工作由于需要大规模将 CentOS 5.8 x86_64 系统应用于集群环境，所以这种无人值守安装的方法主要用于在公司内网批量安装新服务器系统，这种方法极大地简化了用光盘重复安装 CentOS 5.8 x86_64 的过程，再加上通过应用分布式自动化运维工具 Puppet 进行批量部署，达到了自动化运维的目的，避免了重复性劳动，极大地提高了工作效率。

首先，我们来介绍一下与之相关的原理和概念。

1. 什么是 PXE

严格来说，PXE 并不是一种安装方式，而是一种引导方式。进行 PXE 安装的必要条件是在要安装系统的计算机中包含一个 PXE 支持的网卡（NIC），即网卡中必须有 PXE Client。PXE（Pre-boot Execution Environment）协议可以使计算机通过网络启动。此协议采用的是 C/S 结构，即大家熟