

“十三五”普通高等教育规划教材

# Linux 系统 与大数据应用

主编 夏辉 杨伟吉 金鑫



提供电子课件、课后习题答案、源代码

<http://www.cmpedu.com>



机械工业出版社  
CHINA MACHINE PRESS

“十三五”普通高等教育规划教材

# Linux 系统与大数据应用

主 编 夏 辉 杨伟吉 金 鑫

副主编 李淑霞 刘 澍 李 强



机械工业出版社

本书主要内容包括 Linux 基本命令, 大数据主要框架 Hadoop 的搭建和应用, Spark 框架的应用, 大数据科学计算、Python 语言、网络爬虫分析等主要大数据分析应用的工具等。本书分别从系统使用者、网络管理者、shell 程序开发者、应用程序开发者和内核开发者的角度, 全方位地介绍 Linux 操作环境、操作命令, 以及基本的开发方法; 同时, 本书面向大数据应用的开发者, 介绍基于 Linux 系统的大数据计算平台、存储平台, 以及几个广泛使用的应用开发和分析工具; 最后是综合案例和应用, 使读者初步认识 Linux, 熟练使用 shell 命令, 掌握系统管理原理, 熟悉基于 Linux 的大数据开发平台, 并学会几种典型的大数据存储和开发方法。

本书既可作为高等学校计算机软件技术课程的教材, 也可作为管理信息系统开发人员的技术参考书。

本书配套授课电子课件, 需要的教师可登录 [www.cmpedu.com](http://www.cmpedu.com) 免费注册, 审核通过后下载, 或联系编辑索取 (QQ: 2850823885; 电话: 010-88379739)。

## 图书在版编目 (CIP) 数据

Linux 系统与大数据应用 / 夏辉, 杨伟吉, 金鑫主编. —北京: 机械工业出版社, 2019.6

“十三五”普通高等教育规划教材

ISBN 978-7-111-63192-7

I. ①L… II. ①夏… ②杨… ③金… III. ①Linux 操作系统—高等学校—教材 IV. ①TP316.85

中国版本图书馆 CIP 数据核字 (2019) 第 140698 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

责任编辑: 郝建伟 责任校对: 张艳霞

责任印制: 张 博

三河市宏达印刷有限公司印刷

2019 年 9 月第 1 版·第 1 次印刷

184mm×260mm·18.25 印张·449 千字

0001—2500 册

标准书号: ISBN 978-7-111-63192-7

定价: 59.00 元

电话服务

客服电话: 010-88361066

010-88379833

010-68326294

封底无防伪标均为盗版

网络服务

机工官网: [www.cmpbook.com](http://www.cmpbook.com)

机工官博: [weibo.com/cmp1952](http://weibo.com/cmp1952)

金书网: [www.golden-book.com](http://www.golden-book.com)

机工教育服务网: [www.cmpedu.com](http://www.cmpedu.com)

# 前 言

随着云时代的来临，大数据（Big Data）也受到了越来越多的关注。《着云台》的分析师团队认为，大数据通常用来形容一个公司创造的大量非结构化数据和半结构化数据，这些数据在下载关系型数据库用于分析时会花费过多时间和金钱。大数据分析常和云计算联系在一起，因为实时的大型数据集分析需要像 MapReduce 一样的框架来向数十、数百或甚至数千的计算机分配工作。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。“大数据”在物理学、生物学、环境生态学等领域以及军事、金融、通信等行业存在已有时日，却因为近年来互联网和信息行业的发展而引起人们的关注。

大数据作为时下最火热的 IT 行业的词汇，随之而来的数据仓库、数据安全、数据分析、数据挖掘等围绕大数据商业价值的利用逐渐成为行业人士争相追捧的利润焦点。随着大数据时代的来临，大数据分析也应运而生，大数据分析是指对规模巨大的数据进行分析。大数据可以概括为 5 个 V，即数据量（Volume）大、速度（Velocity）快、类型（Variety）多、价值（Value）、真实性（Veracity）。

本书围绕大数据应用开发基础，在内容的编排上力争体现新的教学思想和方法。本书内容编写遵循“从简单到复杂”“从抽象到具体”的原则。书中通过各个章节穿插了很多案例，提供了大数据应用开发从入门到实际应用所必备的知识。Linux 编程基础、Linux 系统用户与组管理、Linux 系统编辑器和软件安装、Linux 系统网络及其服务配置、大数据挖掘的 shell 基础、Linux 系统下的 Python 基础、大数据开发平台、大数据应用开发工具和大数据应用与案例，学生除了要在课堂上学习程序设计的理论方法，掌握编程语言的语法知识和编程技巧外，还要进行大量的课外练习和实践操作。为此本书每章都配备有课后习题，并且每章都有一个综合案例，除此之外，每章节还安排了实验的题目，可供教师实验教学使用。

本书共分 10 章。第 1 章是 Linux 系统概述，第 2 章介绍 Hadoop 平台常用的 Linux 命令，第 3 章介绍 Linux 系统用户与组管理，第 4 章介绍 Linux 系统编辑器和软件安装，第 5 章介绍 Linux 系统网络及其服务配置，第 6 章介绍大数据挖掘的 shell 基础，第 7 章介绍 Linux 系统下的 Python 基础，第 8 章介绍大数据开发平台，第 9 章介绍大数据应用开发工具，第 10 章介绍大数据应用与案例。

本书内容全面，案例新颖，针对性强。本书中所介绍的实例都已在 Windows 10 操作系统下调试运行通过。从应用程序的设计到应用程序的发布，读者都可以按照书中所讲述内容实施。作为教材，每章均附有习题。

本书由夏辉负责书的整体策划，并负责编写第 8 和 10 章，杨伟吉老师负责编写第 4 和 6 章，金鑫老师负责编写第 2、5 和 7 章，李淑霞老师负责编写第 1、3 和 9 章，李强负责所有实验的编写和审核，刘澍老师负责编写课后习题、制作电子课件，并最终完成全文书稿的修订、完善、统稿和定稿工作，参与本书编写的还有王学颖、吴鹏。本书由王学颖教授承担

内容的主审工作，吴鹏博士为本书编写提供了技术支持和帮助，并且对本书初稿在教学过程中存在的问题提出了宝贵的意见。本书也借鉴了中外参考文献中的原理知识和资料，在此一举感谢。

本书配有电子课件、课后习题答案、每章节案例代码、实验代码，以方便教学和自学参考使用，如有需要请到 <http://www.cmpedu.com> 下载。

由于时间仓促，书中难免存在不妥之处，敬请读者原谅，并提出宝贵意见。

编 者

# 目 录

前言	
<b>第 1 章 Linux 系统概述</b> .....1	
1.1 认识 Linux 系统.....1	
1.1.1 Linux 系统版本.....1	
1.1.2 文件与目录.....5	
1.2 Linux 系统的发展.....8	
1.2.1 早期的 Linux 系统.....8	
1.2.2 Linux 系统发展.....9	
1.2.3 Linux 系统举例.....10	
1.2.4 Linux 基本思想和特征.....10	
1.3 Ubuntu 介绍.....12	
1.3.1 Ubuntu 概述.....12	
1.3.2 Ubuntu 的版本与应用.....12	
1.3.3 Ubuntu 下载安装.....13	
1.3.4 Ubuntu 评价.....14	
1.4 Linux 系统下大数据平台.....15	
1.4.1 Linux 系统下大数据平台介绍.....16	
1.4.2 Linux 系统下大数据平台架构.....16	
1.4.3 大数据平台发展前景.....21	
1.5 本章小结.....24	
实践与练习.....24	
<b>第 2 章 Hadoop 平台常用的 Linux 命令</b> .....25	
2.1 文件和目录.....25	
2.1.1 Ubuntu 系统目录结构.....25	
2.1.2 创建/删除目录.....26	
2.1.3 查看文件.....28	
2.1.4 查看路径.....29	
2.1.5 tree 命令.....29	
2.2 文件操作.....30	
2.2.1 创建文件.....30	
2.2.2 查看文件内容.....31	
2.2.3 清屏.....32	
2.3 帮助和历史.....33	
2.3.1 help 命令.....33	
2.3.2 man 命令.....34	
2.3.3 自动补全 Tab.....34	
2.3.4 查看历史 history.....35	
2.4 文件的其他操作.....36	
2.4.1 复制/剪切命令.....36	
2.4.2 重命名.....37	
2.4.3 重定向.....38	
2.4.4 管道命令.....40	
2.4.5 链接快捷方式.....41	
2.4.6 文件搜索.....42	
2.4.7 压缩文件和解压缩.....44	
2.5 系统常用操作.....45	
2.5.1 日历 cal.....45	
2.5.2 时间 date.....46	
2.5.3 进程操作.....47	
2.5.4 系统状态查看和操作.....50	
2.6 本章小结.....51	
实践与练习.....51	
<b>第 3 章 Linux 系统用户与组管理</b> .....53	
3.1 用户与组账号.....53	
3.1.1 添加账号.....53	
3.1.2 切换账号.....56	
3.1.3 远程登录.....57	
3.2 超级用户 root.....59	
3.2.1 切换超级用户.....59	
3.2.2 加入用户组.....60	
3.2.3 退出.....61	
3.3 用户操作权限.....61	
3.3.1 文件权限管理概述.....61	
3.3.2 修改权限.....62	
3.4 系统的安全性.....64	
3.4.1 账号的安全性.....64	
3.4.2 常见漏洞的安全性.....66	
3.4.3 SELinux.....71	
3.5 应用案例：Hadoop 集群创建的用户.....75	

3.6 本章小结	76	第6章 大数据挖掘的 shell 基础	141
实践与练习	76	6.1 大数据开发的 shell 基础	141
<b>第4章 Linux 系统编辑器和软件安装</b>	<b>79</b>	6.2 shell 的模式和类型	141
4.1 Linux 主要编辑器介绍	79	6.2.1 shell 的运行模式	142
4.1.1 gedit 编辑器	79	6.2.2 shell 的类型	143
4.1.2 vim 编辑器	82	6.3 shell 编程	144
4.2 HTTP Server 的介绍和安装	87	6.3.1 bash 简介	144
4.2.1 Apache HTTP Server	87	6.3.2 shell 命令行	146
4.2.2 安装与配置实例	87	6.3.3 shell 脚本语句和命令	149
4.3 Linux 大数据包导入库	91	6.4 应用案例: shell 编程的 数据挖掘	167
4.3.1 pip 安装	91	6.5 本章小结	169
4.3.2 自带工具	94	实践与练习	170
4.3.3 导入大数据计算库	95	<b>第7章 Linux 系统下的 Python 基础</b>	<b>171</b>
4.4 应用案例: Hadoop 集群修改 配置文件	97	7.1 Linux 中的 Python	171
4.5 本章小结	101	7.2 Python 基础	173
实践与练习	101	7.2.1 基本数据类型	173
<b>第5章 Linux 系统网络及其服务配置</b>	<b>104</b>	7.2.2 流程控制语句	180
5.1 网络配置	104	7.3 Python 函数	185
5.1.1 IP 地址查看和配置	104	7.4 Python 类和对象	191
5.1.2 SSH 和 Telnet	108	7.4.1 面向对象	191
5.1.3 SSH 工具	112	7.4.2 类和对象	193
5.2 Xshell 工具	113	7.4.3 构造函数	194
5.2.1 服务器配置与连接	113	7.4.4 继承	199
5.2.2 客户端与服务器通信	116	7.5 文件的操作	202
5.3 FTP 服务器	118	7.6 Python 的类库	207
5.3.1 FTP 服务器简介	119	7.7 应用案例: 数据挖掘相关 Python 类库应用	211
5.3.2 FTP 服务器安装配置	120	7.8 其他相关知识	216
5.4 Samba 服务器	123	7.9 本章小结	218
5.4.1 Samba 简介	123	实践与练习	218
5.4.2 安装与配置实例	124	<b>第8章 大数据开发平台</b>	<b>220</b>
5.5 Apache Web 服务器	130	8.1 大数据开发平台简介	220
5.5.1 Apache Web 服务器简介	130	8.1.1 大数据的应用与前景	220
5.5.2 Apache Web 服务器安装与配置	130	8.1.2 Hadoop 简介	220
5.5.3 虚拟主机配置	133	8.2 Hadoop 框架介绍	221
5.6 应用案例: Hadoop 平台的 网络和服务器配置	138	8.2.1 Hadoop 框架及组件介绍	221
5.7 本章小结	139	8.2.2 Hadoop 的核心组件	226
实践与练习	139		

8.2.3 Hadoop 的安全性	228	9.3 Hive 基础与应用	259
8.3 Hadoop 安装与配置	230	9.3.1 Hive 安装	259
8.3.1 下载安装 Hadoop	230	9.3.2 Hive 数据模型和查询语言	262
8.3.2 Hadoop 配置	231	9.3.3 Hive 数据分析	269
8.3.3 词频统计示例	233	9.4 本章小结	269
8.4 Spark 简介	233	实践与练习	270
8.4.1 Spark 特点	234	<b>第 10 章 大数据应用与案例</b>	271
8.4.2 Spark 生态系统	234	10.1 数据挖掘工具和主要算法	271
8.5 Spark 安装与配置	244	10.1.1 Mahout 安装和配置	271
8.5.1 下载与安装 Spark	244	10.1.2 K-Means 算法和 Canopy 算法	273
8.5.2 Spark 集群配置	244	10.1.3 贝叶斯算法	274
8.6 本章小结	245	10.2 Hadoop 应用案例: World count	
实践与练习	246	词频统计案例	276
<b>第 9 章 大数据应用开发工具</b>	247	10.2.1 实训目的和要求	276
9.1 数据抽取 ETL	247	10.2.2 运用	276
9.1.1 ETL 概述	247	10.3 Spark 应用案例: Spark 进行	
9.1.2 ETL 工具	249	电商数据检索	279
9.1.3 网络爬虫技术及应用	250	10.3.1 实训目的和要求	279
9.2 Hbase 原理和模型	250	10.3.2 运用	279
9.2.1 Hbase 安装和部署	250	10.4 本章小结	281
9.2.2 Hbase 应用	253	实践与练习	281
9.2.3 Hbase 优化和存储	255	<b>参考文献</b>	283

# 第 1 章 Linux 系统概述

Linux 内核最初是由芬兰人李纳斯·托瓦兹 (Linus Torvalds) 在赫尔辛基大学上学时出于个人爱好而编写的。

Linux 是一套免费使用和自由传播的类 UNIX 操作系统，是一个基于 POSIX 和 UNIX 的多用户、多任务、支持多线程和多 CPU 的操作系统。

Linux 能运行主要的 UNIX 工具软件、应用程序和网络协议，支持 32 位和 64 位硬件。Linux 继承了 UNIX 以网络为核心的设计思想，是一个性能稳定的多用户网络操作系统。

本章主要介绍 Linux 系统的版本以及文件与目录结构，使读者对 Linux 系统有一个初步了解；同时本章也将对 Linux 系统的发展进行介绍，在读者对于 Linux 系统有一定了解后再对 Ubuntu 版本及 shell 命令进行初步学习，从而为之后的学习建立基础。

## 1.1 认识 Linux 系统

Linux 系统具有多个版本，Linux 的发行版就是将 Linux 内核与应用软件做一个打包。

### 1.1.1 Linux 系统版本

目前 Linux 系统具有多个版本，市面上使用较多的发行版本有：Ubuntu、RedHat、CentOS、Debian、Fedora、SUSE、TurboLinux、BluePoint、RedFlag、Xterm、SlackWare 等，我们将在本节对各个版本进行简单介绍。

#### 1. Ubuntu

Ubuntu (乌班图) 是一个以桌面应用为主的 Linux 操作系统，其名称来自非洲南部祖鲁语或豪萨语的“ubuntu”一词，意思是“人性”“我的存在是因为大家的存在”，是非洲传统的一种价值观，类似华人社会的“仁爱”思想。

- 中文名称：友帮拓、优般图、乌班图。
- 开发商：Canonical 公司、Ubuntu 基金会。
- 产品类型：自由开放源代码。
- 初始版本：2004 年 10 月 20 日。

Ubuntu 基于 Debian 发行版和 Gnome 桌面环境，而从 11.04 版起，Ubuntu 发行版放弃了 Gnome 桌面环境，改为 Unity，与 Debian 的不同在于它每 6 个月会发布一个新版本。Ubuntu 的目标在于为一般用户提供一个最新的、稳定的、主要由自由软件构建而成的操作系统。Ubuntu 具有庞大的社区力量，用户可以方便地从社区获得帮助。2013 年 1 月 3 日，Ubuntu 正式发布面向智能手机的移动操作系统。Ubuntu 基于 Linux 的免费开源桌面 PC 操作系统，十分契合英特尔的超极本定位，支持 x86、64 位和 PPC 架构。2014 年 2 月 20

日, Canonical 公司于北京中关村召开了 Ubuntu 智能手机发布会, 正式宣布 Ubuntu 与国产手机厂商魅族合作推出 Ubuntu 版 MX3。

## 2. Debian

广义的 Debian 是指一个致力于创建自由操作系统的合作组织及其作品, 由于 Debian 项目众多内核分支中以 Linux 宏内核为主, 而且 Debian 开发者所创建的操作系统中绝大部分基础工具来自于 GNU 工程, 因此 Debian 常指 Debian GNU/Linux。

- 英文名称: Debian。
- 开发商: Debian Project。
- 产品类型: 自由软件。
- 内核类型: 宏内核 (Linux)、微内核 (Hurd)。
- 软件管理: dpkg。
- 发行时间: 1993 年 8 月 16 日。

非官方内核分支还有只支持 x86 的 Debian GNU/Hurd (Hurd 微内核), 只支持 Amd64 的 Dyson (OpenSolaris 混合内核) 等。这些非官方分支都存在一些严重的问题, 没有实用性, 如 Hurd 微内核在技术上不成熟, 而 Dyson 则基础功能仍不完善。

## 3. Red Hat

Red Hat 是全球最大的应用开源技术的厂家, 其产品 Red Hat Linux 也是全世界应用最广泛的 Linux 操作系统。Red Hat 公司总部位于美国北卡罗来纳州, 在全球拥有 22 个分部。

Red Hat 在 2014 年 6 月发布了旗舰版企业操作系统——Red Hat Enterprise Linux (RHEL) 7。基于 Red Hat Enterprise Linux 7 操作系统, 企业可整合裸机服务器、虚拟机、基础设施即服务 (IaaS) 和平台即服务 (PaaS), 以构建一个强大稳健的数据中心环境, 满足不断变化的业务。

## 4. CentOS

CentOS (Community Enterprise Operating System, 社区企业操作系统) 是 Linux 发行版之一, 它是来自于 Red Hat Enterprise Linux 依照开放源代码规定释出的源代码所编译而成。由于出自同样的源代码, 因此有些要求高度稳定性的服务器以 CentOS 替代商业版的 Red Hat Enterprise Linux 使用。两者的不同在于 CentOS 并不包含封闭源代码软件。

- 中文名称: 社区企业操作系统。
- 类型: 计算机软件。
- 初始版本: 2014 年 7 月 7 日。

CentOS 是一个基于 Red Hat Linux 提供的可自由使用源代码的企业级 Linux 发行版本。每个版本的 CentOS 都会获得十年的支持 (通过安全更新方式)。新版本的 CentOS 大约每两年发行一次, 而每个版本的 CentOS 会定期 (大概每 6 个月) 更新一次, 以便支持新的硬件。这样, 建立一个安全、低维护、稳定、高预测性、高重复性的 Linux 环境。

CentOS 是免费的, 可以像使用 RHEL 一样去构筑企业级的 Linux 系统环境, 但不需要向 Red Hat 支付任何的费用。CentOS 的技术支持主要通过社区的官方邮件列表、论坛和聊天室获得。

## 5. Fedora

Fedora 是一个知名的 Linux 发行版, 是一款由全球社区爱好者构建的面向日常应用的

快速、稳定、强大的操作系统。它允许任何人无论现在还是将来自由地使用、修改和重发布。它由一个强大的社群开发，这个社群的成员以自己的不懈努力，提供并维护自由、开放源码的软件和开放的标准。Fedora 项目由 Fedora 基金会管理和控制，得到了 Red Hat 公司的支持。Fedora 是一个独立的操作系统，可运行的体系结构包括 x86（即 i386~i686），x86\_64 和 PowerPC。

- 英文名称：Fedora。
- 软件许可：主要为 GNU GPL。
- 源码模式：自由及开放源代码软件。
- 内核类型：宏内核（Linux）。
- 发行时间：2003 年 11 月 16 日。

最早 Fedora Linux 社区的目标是为 Red Hat Linux 制作并发布第三方的软件包，然而当 Red Hat Linux 停止发行后，Fedora 社区便集成到 Red Hat 赞助的 Fedora Project，目标是开发出由社区支持的操作系统（事实上，Fedora Project 除了由志愿者组织外，也有许多 Red Hat 的员工参与开发）。Red Hat Enterprise Linux 则取代 Red Hat Linux 成为官方支持的系统版本。

## 6. SUSE

SUSE Linux 原来是德国的 SUSE Linux AG 公司发行维护的 Linux 发行版，是属于此公司的注册商标。第一个版本出现在 1994 年年初。2004 年这家公司被 Novell 公司收购。

- 英文名称：SUSE。
- 起源：德国。
- 原属公司：SUSE Linux AG 公司。
- 初发行：1994 年初。

SUSE 支持在安装的时候调校 NTFS 硬盘的大小，以便顺利把 Linux 安装到一台已经安装了 Windows 2000 Server 或 Windows XP 的计算机上。此外，SUSE 亦会自动侦测很多常见的 Windows 调制解调器并为它们安装驱动程序。

SUSE 也收录了 Linux 下的多个桌面环境，如 KDE 和 GNOME 及一些视窗管理员（如 Window Maker、Blackbox 等）。YaST2 安装程序也会让使用者选择使用 GNOME、KDE 或者不安装图形界面。SUSE 已经为使用者提供了一系列多媒体程序，如 K3B（CD/DVD 烧录）、Amarok（音乐播放器）和 Kaffeine（影片播放器）。它也收录了图片处理器，以及其他文字阅读/处理软件，如 PDF 格式文件阅读软件等。

相比以往，现在所有的开发人员及使用者能够测试 SUSE 的产品并一起开发新版本的 SUSE。在以往，SUSE 的开发工作都是于内部进行的。SUSE 10.0 是第一个给予公众测试的版本。为了配合这个转变，用户除了能够购买盒装版本的 SUSE 外，也可以从网络上免费下载。一系列的改变让 2005 年 10 月 6 日推出的 SUSE Linux 有 3 个版本——“OSS 版”（完全地开放原始码）、“试用版”（同时包含开放原始码的程序及专属程序如 Adobe Reader、Real Player 等，还有“盒装零售版”，可以免费下载，还可以安装在硬盘上，并且没有使用限制或限期。

## 7. TurboLinux

TurboLinux 是拓林思公司最近发行的 Linux 版本，已在日本和中国取得了巨大的成功，

在美国也有一定的业绩。当前版本为 4.0，基于 Linux 2.2.10 内核。目前 TurboLinux 的总公司位于日本，成立于 1992 年，由 Cliff 和 Iris Miller 在美国创建。主要客户是亚洲客户，是面向亚洲语发行的 UNIX-like 版本系统。

TurboLinux 优点在于：具有简单易用的图形安装程序，友好的图形桌面界面 KDE、GNOME 等；并且具有丰富的软件包，包括系统管理工具、网络分析程序、服务程序包（如 Apache）等；提供了完整的源代码程序；提供了预配置安装功能。

## 8. BluePoint

BluePoint Linux 为第一个在 FrameBuffer 上进行汉化的中文 Linux 版本。

和其他中文版 Linux 不同的是，BluePoint 首创了一个版本同时支持 GB/BIG5/ASCII 繁简汉字，多内码同屏显示和指定窗口内码的功能。在中文桌面环境上，BluePoint Linux 全面支持中文 TTF/GBK 字库、多级点阵字库及中文打印。针对 Linux 操作系统应用软件较少的弱点，BluePoint 集成了不少中文软件包、服务器用软件包和图形化应用开发工具。

而且 BluePoint Linux 采用了自主开发的中文 Linux 内核与中文 X Window 技术，使得从开机启动到具体应用，整个中文环境与 Linux 系统无缝结合。虽然 BluePoint Linux 是在国内开发的，用 BIG5 的用户也不用担心兼容性的问题，在开始安装的时候就可以选择简体或者繁体安装界面，而在进入 BluePoint Linux 系统之后，GB/BIG5 的切换也非常简单。

## 9. Red Flag

Red Flag（红旗）Linux 是由北京中科红旗软件技术有限公司开发的一系列 Linux 发行版，包括桌面版、工作站版、数据中心服务器版、HA 集群版和红旗嵌入式 Linux 等产品。目前在中国各软件专卖店可以购买到光盘版，同时官方网站也提供光盘镜像免费下载。Red Flag Linux 是中国较大、较成熟的 Linux 发行版之一。

嵌入式领域是红旗软件的重要发展方向之一。红旗嵌入式 Linux 是红旗软件面向嵌入式设备而开发的通用型嵌入式平台。

- 中文名称：红旗 Linux。
- 源码模式：开源。
- 内核类型：Linux。
- 发行时间：1999 年 8 月。

Red Flag Linux 特色在于：具有完善的中文支持，与 Windows 相似的用户界面，通过 LSB 4.1 测试认证，具备了 Linux 标准基础的一切品质农历的支持和查询；X86 平台对 Intel EFI 的支持；Linux 下网页嵌入式多媒体插件的支持，实现了 Windows Media Player 和 RealPlayer 的标准 JavaScript 接口；前台窗口优化调度功能，支持 MMS/RTSP/HTTP/FTP 协议的多线程下载工具；界面友好的内核级实时检测防火墙，KDE 登录窗口、注销窗口、主皮肤的主题支持，可缩放的系统托盘；源代码已经进入 KDE 项目，GTK2 Qt 打开关闭文件对话框的统一等。

红旗软件不仅专注于嵌入式平台的研究，而且与第三方合作伙伴开展广泛的协作，共同向客户提供成熟的嵌入式 Linux 软硬件整体解决方案，缩短客户产品的上市时间，这也正是红旗软件在嵌入式领域的价值所在。

## 10. Slackware

Slackware Linux 是由 Patrick Volkerding 开发的 GNU/Linux 发行版。与很多其他的发行

版不同，它坚持 KISS(Keep It Simple Stupid)的原则。一开始，配置系统会有一些困难，但是更有经验的用户会喜欢 Slackware 的透明性和灵活性。Slackware 很多特性体现出了 KISS 原则，例如，不依赖图形界面的文本化系统配置、传统的服务管理方式和不解决依赖的包管理方式。它的最大特点就是安装灵活，目录结构严谨，版本力求稳定而非追新。Slackware 的软件包都是通常的 `tgz(tar/gzip)`或者 `txz(xz)`格式文件再加上安装脚本。`tgz/txz` 对于有经验的用户来说，比 RPM 更为灵活，并避免了 APT 等管理器可能带来的依赖地狱。

Slackware 的软件套件管理系统很独特。它的软件套件管理系统和别的发行版本一样可以很容易地安装、升级、移除包，但是它不会试着去追踪或者管理涉及哪些依赖关系（也就是保证系统拥有所有的安装包内的程序需要的系统库）。如果所需要的先决条件不能满足，在程序执行之前不会有提醒和指示。

Slackware 的包都经过 `gzip` 压缩和 `tarball` 打包，但文件扩展名是 `.tgz`，而不是 `.tar.gz`。它们的结构是这样的：当在根目录下解压缩和释放，其中的文件会放置于它们的安装位置。因此可以不使用 Slackware 的包工具来安装包，而仅仅使用 `tar` 和 `gzip` 命令，如果包中有 `doinst.sh` 脚本，一定要运行它。

相对的，Red Hat 的 RPM 是 CPIO 档案，Debian 的 `deb` 文件是 `ar` 档案。它们都包括一些依赖关系的信息，包管理器工具可以使用这些信息来寻找和安装先决条件。它们在先决条件满足前是不会安装新包的（虽然可以强制进行）。

关于追踪或者无视依赖关系孰优孰劣的争论并不很热闹，这多少让人想起了持续甚久的“VI 对 Emacs”的“宗教战争”。Slackware 解决问题的方法被技巧熟练的用户群很好地接受了。

## 11. XTerm

XTerm 最先是 Jim Gettys 的学生 Mark Vandevoorde 在 1984 年夏天为 VS100 写的独立虚拟终端，当时 X 的开发刚刚开始。很快人们就发现它作为 X 的一部分比作为独立的程序更为有用，于是它开始针对 X 而开发。

Gettys 曾讲述过有关的故事，“XTerm 内部如此恐怖的部分原因是它最初被计划开发成一个能驱动多个 VS100 显示器的单独进程。”

作为 X 参考实现的一个部分之后多年，1996 年左右，XTerm 开发的主干转移至了 XFree86（从 X11R6.3 版本派生出来），暂时由 Thomas E. Dickey 维护，有许多 XTerm 变体可用，大多数的 X 虚拟终端都是从 XTerm 的变体起步的。

### 1.1.2 文件与目录

由于 Linux 操作系统的操作界面并不同于 Windows 系统的图形操作界面，其优点是建立了不受任何商品化软件的版权制约、全世界都能自由使用的 UNIX 兼容产品，所以在初步接触时，需要对 Linux 系统的文件目录进行系统的学习。

#### 1. 文件

普通文件 (regular file)：就是一般存取的文件，由 `ls -al` 显示出来的属性中，第一个属性为 `[-]`，例如 `[-rwxrwxrwx]`。另外，依照文件的内容，又大致可以分为以下几种。

1) 纯文本文件 (ASCII)：这是 UNIX 系统中最多的一种文件类型，之所以称为纯文本文件，是因为内容是可以直接读到的数据，例如数字、字母等。设置文件几乎都属于这

种文件类型。例如，使用命令“`cat ~/.bashrc`”就可以看到该文件的内容（`cat` 是将文件内容读出来）。

2) 二进制文件 (binary): 系统其实仅认识且可以执行二进制文件 (binary file)。Linux 中的可执行文件 (脚本, 文本方式的批处理文件不算) 就是这种格式。例如, 命令 `cat` 就是一个二进制文件。

3) 数据格式文件 (data): 有些程序在运行过程中, 会读取某些特定格式的文件, 那些特定格式的文件可以称为数据文件 (data file)。例如, Linux 在用户登入时, 都会将登录数据记录在 `/var/log/wtmp` 文件内, 该文件是一个数据文件, 它可以通过 `last` 命令读出来。但使用 `cat` 时, 会读出 Linux 乱码。因为它是属于一种特殊格式的文件。

- 目录文件 (directory): 就是目录, 第一个属性为 [d], 例如 [drwxrwxrwx]。
- 连接文件 (link): 类似 Windows 下面的快捷方式。第一个属性为 [l], 例如 [lrwxrwxrwx]。
- 设备与设备文件 (device): 与系统外设及存储等相关的一些文件, 通常都集中在 `/dev` 目录, 分为以下两种。

块设备文件: 就是存储数据以供系统存取的接口设备, 简单而言就是硬盘。例如一号硬盘的代码是 `/dev/hda1` 等文件。

字符设备文件: 即串行端口的接口设备, 例如键盘、鼠标等。第一个属性为 [c]。

- 套接字 (sockets): 这类文件通常用在网络数据连接。可以启动一个程序来监听客户端的要求, 客户端就可以通过套接字来进行数据通信。第一个属性为 [s], 最常在 `/var/run` 目录中看到这种文件类型。
- 管道 (FIFO, pipe): FIFO 也是一种特殊的文件类型, 它主要的目的是, 解决多个程序同时存取一个文件所造成的错误。

## 2. 目录结构

`/`: 根目录所有的目录、文件、设备都在`/`之下, `/`就是 Linux 文件系统的组织者, 也是最上级的领导者。

首先手动输入命令 `ls/`(打开目录)

```
ls /
```

结果如图 1-1 所示, 即 Linux 操作系统目录。

```
root@localhost:~# ls /
```

图 1-1 打开目录

以下是对这些目录及该目录下文件的简单介绍。

- `/bin`: `bin` 是 `binary` 的缩写, 这个目录存放着最经常使用的命令。
- `/boot`: 这里存放的是启动 Linux 时使用的一些核心文件, 包括一些连接文件以及镜像。
- `/dev`: `dev` 是 `device` (设备) 的缩写, 该目录下存放的是 Linux 的外部设备, 在

Linux 中访问设备的方式和访问文件的方式是相同的。

- /etc: 这个目录用来存放所有的系统管理所需要的配置文件和子目录。
- /home: 用户的主目录, 在 Linux 中, 每个用户都有一个自己的目录, 一般该目录名以用户的账号命名。
- /lib: 这个目录里存放着系统最基本的动态连接共享库, 其作用类似于 Windows 里的 DLL 文件。几乎所有的应用程序都需要用到这些共享库。
- /lost+found: 这个目录一般情况下是空的, 当系统非法关机后, 这里就存放了一些文件。
- /media: linux 系统会自动识别一些设备, 例如 U 盘、光驱等, 当识别后, Linux 会把识别的设备挂载到这个目录下。
- /mnt: 系统提供该目录是为了让用户临时挂载别的文件系统, 可以将光驱挂载在 /mnt/上, 然后进入该目录就可以查看光驱里的内容了。
- /opt: 这是给主机额外安装软件所摆放的目录。例如, 安装一个 Oracle 数据库就可以放到这个目录下。默认是空的。
- /proc: 这个目录是一个虚拟的目录, 它是系统内存的映射, 可以通过直接访问这个目录来获取系统信息。

这个目录的内容不在硬盘上而是在内存里, 也可以直接修改里面的某些文件, 例如, 可以通过下面的命令来屏蔽主机的 ping 命令, 使别人无法 ping 你的机器。

```
echo 1 > /proc/sys/net/ipv4/icmp_echo_ignore_all
```

- /root: 该目录为系统管理员, 也称作超级权限者的用户主目录。
- /sbin: s 就是 Super User 的意思, 这里存放的是系统管理员使用的系统管理程序。
- /selinux: 这个目录是 Red Hat/CentOS 所特有的目录, selinux 是一个安全机制, 类似于 Windows 的防火墙, 但是这套机制比较复杂, 这个目录就是存放 selinux 相关的文件的。
- /srv: 该目录存放一些服务启动之后需要提取的数据。
- /sys: 这是 Linux 2.6 内核的一个很大的变化。该目录下安装了 2.6 内核中新出现的一个文件系统 sysfs。sysfs 文件系统集成了下面 3 种文件系统的信息: 针对进程信息的 proc 文件系统、针对设备的 devfs 文件系统以及针对伪终端的 devpts 文件系统。该文件系统是内核设备树的一个直观反映。

当一个内核对象被创建时, 对应的文件和目录也在内核对象子系统中被创建。

- /tmp: 这个目录用来存放一些临时文件。
- /usr: 这是一个非常重要的目录, 用户的很多应用程序和文件都放在这个目录下, 类似于 Windows 下的 Program Files 目录。
- /usr/bin: 系统用户使用的应用程序。
- /usr/sbin: 超级用户使用的比较高级的管理程序和系统守护程序。
- /usr/src: 内核源代码默认的放置目录。
- /var: 这个目录中存放着在不断扩充着的东西, 通常将那些经常被修改的目录放在这个目录下, 包括各种日志文件。

在 Linux 系统中，有几个目录是比较重要的，平时需要注意不要误删除或者随意更改内部文件。

最后对于 Linux 系统还应注意以下几点。

1) /etc: 这是系统中的配置文件，如果更改了该目录下的某个文件可能会导致系统不能启动。

2) /bin, /sbin, /usr/bin, /usr/sbin: 这些是系统预设的执行文件的放置目录，比如 ls 就是在 /bin/ls 目录下的。

3) 值得提出的是，/bin, /usr/bin 是给系统用户使用的指令（除 root 外的通用用户），而 /sbin, /usr/sbin 则是给 root 使用的指令。

4) /var: 这是一个非常重要的目录，系统中的程序都会有相应的日志产生，而这些日志就被记录在这个目录下，具体在 /var/log 目录下，另外 mail 的预设放置也是在这里。

## 1.2 Linux 系统的发展

Linux 操作系统是基于 UNIX 操作系统发展而来的一种克隆系统，它诞生于 1991 年的 10 月 5 日（这是第一次正式向外公布的时间）。以后借助于 Internet 网络，并通过全世界各地计算机爱好者的共同努力，已成为今天世界上使用最多的一种 UNIX 类操作系统，并且使用人数还在迅猛增长。

### 1.2.1 早期的 Linux 系统

1993 年，大约有 100 余名程序员参与了 Linux 内核代码编写/修改工作，其核心开发组由 5 人组成，此时 Linux 0.99 的代码大约有 10 万行，用户大约有 10 万左右。这为后面 Linux 系统的正式出现奠定了基础。

1994 年 3 月，Linux 1.0 发布，代码量 17 万行，当时是按照完全自由免费的协议发布，随后正式采用 GPL 协议。1995 年 1 月，Bob Young 创办了 Red Hat（红帽），以 GNU/Linux 为核心，集成了 400 多个源代码开放的程序模块，开发出了一种冠以品牌的 Linux，即 Red Hat Linux，称为 Linux “发行版”，在市场上出售。这在经营模式上是一种创举。1996 年 6 月，Linux 2.0 内核发布，此内核有大约 40 万行代码，并可以支持多个处理器。此时的 Linux 已经进入了实用阶段，全球大约有 350 万人使用。

1998 年 2 月，以 Eric Raymond 为首的一批年轻一代认识到 GNU/Linux 体系的产业化道路的本质，并非是什么自由哲学，而是市场竞争的驱动，创办了“Open Source Initiative”（开放源代码促进会），在互联网世界里展开了一场历史性的 Linux 产业化运动。2001 年 1 月，Linux 2.4 发布，它进一步地提升了 SMP 系统的扩展性，同时也集成了很多用于支持桌面系统的特性：USB，PC 卡（PCMCIA）的支持，内置的即插即用等功能。2003 年 12 月，Linux 2.6 版内核发布，相对于 2.4 版内核，2.6 在对系统的支持方面都有很大的变化。2004 年的第 1 月，SUSE 被 Novell 收购，Asianux，MandrakeSoft 也在 5 年中首次宣布季度赢利。2004 年 3 月，SGI 宣布成功实现了 Linux 操作系统支持 256 个 Itanium 2 处理器。

## 1.2.2 Linux 系统发展

尽管 Linux 是最流行的开源操作系统，但是相对于其他操作系统的漫长历史来说，Linux 的历史非常短暂。

20 世纪 80 年代，Andrew Tanenbaum 创建了一个微内核版本的 UNIX，名为 MINIX（代表 minimal UNIX），它可以在小型的个人计算机上运行。这个开源操作系统在 20 世纪 90 年代激发了林纳斯·托瓦兹开发 Linux 的灵感。

1991 年 10 月 5 日，林纳斯·托瓦兹为了给 Minix 用户设计一个比较有效的 UNIX PC 版本，自己动手写了一个“类 Minix”的操作系统。整个故事从两个在终端上打印 AAAA...和 BBBB...的进程开始，当时最初的内核版本是 0.02。林纳斯·托瓦兹将它发到了 Minix 新闻组，很快就得到了反应。林纳斯·托瓦兹在这种简单的任务切换机制上进行扩展，并在很多热心支持者的帮助下开发和推出了 Linux 的第一个稳定的工作版本。1991 年 11 月，Linux 0.10 版本推出，0.11 版本随后在 1991 年 12 月推出，当时将它发布在 Internet 上，免费供人们使用。当 Linux 非常接近于一种可靠的/稳定的系统时，Linus 决定将 0.13 版本称为 0.95 版本。1994 年 3 月，正式的 Linux 1.0 出现了，这差不多是一种正式的独立宣言。截至那时为止，它的用户基数已经发展得很大，而且 Linux 的核心开发队伍也建立起来了。

核心的开发和规范一直是由 Linux 社区控制着，版本也是唯一的。实际上，操作系统的内核版本指的是在林纳斯本人领导下的开发小组开发出的系统内核的版本号。自 1994 年 3 月 14 日发布了第一个正式版本 Linux 1.0 以来，每隔一段时间就有新的版本或其修订版公布。

一般地，可以从 Linux 内核版本号来区分系统是 Linux 稳定版还是测试版。以版本 2.4.0 为例，2 代表主版本号，4 代表次版本号，0 代表改动较小的末版本号。在版本号中，序号的第二位为偶数的版本表明这是一个可以使用的稳定版本，如 2.2.5，而序号的第二位为奇数的版本一般有一些新的东西加入，是不一定很稳定的测试版本，如 2.3.1。这样稳定版本来源于上一个测试版升级版本号，而一个稳定版本发展到完全成熟后就不再发展。

Linux 内核的发展过程中，还不得不提一下各种 Linux 发行版的作用，因为正是它们推动了 Linux 的应用，从而也让更多的人开始关注 Linux。一些组织或厂家，将 Linux 系统的内核与外围实用程序（Utilities）软件和文档包装起来，并提供一些系统安装界面和系统配置、设定与管理工具，就构成了一种发行版本（distribution），Linux 的发行版本其实就是 Linux 核心再加上外围的实用程序组成的一个大软件包而已。

相对于 Linux 操作系统内核版本，发行版本的版本号随发布者的不同而不同，与 Linux 系统内核的版本号是相对独立的。因此把 SUSE、Red Hat、Ubuntu、Slackware 等直接说成是 Linux 是不确切的，它们是 Linux 的发行版本，更确切地说，应该叫作“以 Linux 为核心的操作系统软件包”。根据 GPL 准则，这些发行版本虽然都源自一个内核，并且都有自己各自的贡献，但都没有自己的版权。Linux 的各个发行版本，都是使用林纳斯主导开发并发布的同一个 Linux 内核，因此内核层不存在兼容性问题。每个版本都不一样，只是在发行版本的最外层才有所体现，而绝不是 Linux 本身特别是内核不统一或是不兼容。

Linux 快速从一个人项目进化成为一个全球数千人参与的开发项目。对于 Linux 来