

**操作系统课程设计**

**学生姓名 张瑞晨**

**学 号 202283290159**

**院 系 计算机与网络空间安全学院**

**专 业 计算机科学与技术**

**指导老师 韩进**

**二Ｏ二四 年 1 月 9 日**

Linux操作系统入门

张瑞晨

南京信息工程大学计算机与网络安全学院，江苏 南京 210044

摘要：本次课程设计的核心内容主要围绕Linux操作系统的环境配置、系统软件的安装以及服务部署等多个方面展开，同时涵盖了大数据平台Hadoop的伪分布式环境搭建的详细步骤和实践操作。在课程中，我们详细探讨了如何通过配置SSH、Apache、Samba等关键系统服务，来全面掌握Linux系统的网络管理、文件共享和Web服务部署等核心操作技能。特别是在Hadoop的安装和配置环节，我们深入剖析了HDFS、YARN等核心组件的运行机制，并针对环境变量配置错误、权限管理不当、SSH无密码登录失败等常见问题，提供了切实可行的解决方案。通过这一系列的学习和实践，不仅显著提升了对Linux系统的实际操作能力，还培养了在面对复杂问题时进行有效分析和解决的能力，为未来在大数据技术和系统管理领域的深入学习奠定了坚实的基础。

关键词：Linux 操作系统、SSH、Hadoop、伪分布式部署

目录

[一、 LINUX操作系统安装的步骤 3](#_Toc944210385)

[1.1 Linux镜像文件选择与下载 3](#_Toc4182471)

[1.2 镜像安装工具下载及启动盘制作 5](#_Toc593644904)

[1.3 进入BIOS设置U盘启动 6](#_Toc763822202)

[二、LINUX操作系统命令的应用与相应功能 6](#_Toc374971353)

[2.1常见的进程相关命令应用与相应功能 6](#_Toc1795592555)

[2.2其他常用命令 8](#_Toc537747290)

[三.LINUX操作系统系统软件的安装与配置 10](#_Toc296016633)

[3.1 ssh服务安装 10](#_Toc2093194398)

[4. LINUX操作系统下HADOOP的伪分布安装与配置 12](#_Toc997956666)

[1.java环境安装检测 12](#_Toc695670068)

[2.创建用户 12](#_Toc24484064)

[3.更改host内容 12](#_Toc1629363554)

[4.安装 Hadoop 13](#_Toc136384361)

[5. 测试是否成功安装hadoop 13](#_Toc877805381)

[6. hadoop单机/伪分布式 14](#_Toc1699109581)

[5.总结与设计中遇到问题的解决 14](#_Toc266100157)

[问题1.JAVA\_HOME 环境变量配置错误 14](#_Toc1093056947)

[问题 2：Hadoop 启动时提示 SSH 认证失败 15](#_Toc591853609)

[问题3：错误修改环境变量导致无法进入系统 15](#_Toc552401185)

[课程设计总结 16](#_Toc1499516853)

# LINUX操作系统安装的步骤

Linux操作系统的安装过程包括下载、制作启动盘、安装系统及后期配置等环节。以下是详细的Linux系统安装全过程，适用于安装Ubuntu、CentOS、Debian等主流Linux发行版。

## Linux镜像文件选择与下载

这里我们选择ubuntu进行下载。

方法1：

直接去官网下载最新版。

下载链接为： https://ubuntu.com/#download 选择桌面版的系统下载。

此种方法速度实在太慢，因此选取其他方法。

方法2：

使用清华源镜像下载。此方法速度较快。

访问以下网址<https://mirrors.tuna.tsinghua.edu.cn/ubuntu-releases/>

选择适合自己的版本镜像文件下载，此处选择了ubuntu24下载。



图 1 镜像源选择与下载

## 镜像安装工具下载及启动盘制作

使用 Win32 Disk Imager 制作启动盘，Win32 Disk Imager 是一个轻量级的工具，主要用于将 ISO 文件 或 镜像文件 写入 U 盘，从而制作启动盘。可以用它来将 Windows 安装镜像或 Linux 系统镜像写入 U 盘，使其成为可引导盘。安装完成后，打开 Win32 Disk Imager。

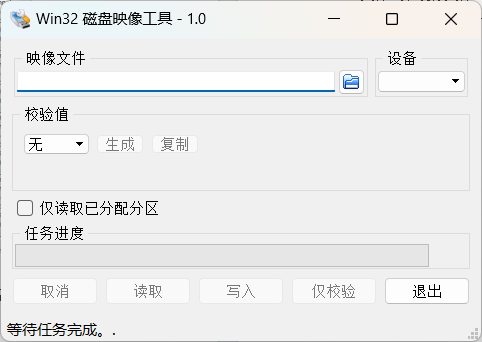


图 2 win32DiskImager 界面

选择 ISO 文件,点击 左上角的文件夹图标，选择下载的操作系统 ISO 文件。在工具界面的下拉框中，选择目标 U 盘设备。选择正确的盘符，避免误操作擦除其他硬盘数据。在 Write 按钮的右侧，确保选择 Write（写入）模式。如果之前已经有数据，程序会提醒清空 U 盘中的内容。点击 Write 按钮，工具将开始将 ISO 文件内容写入到 U 盘中。

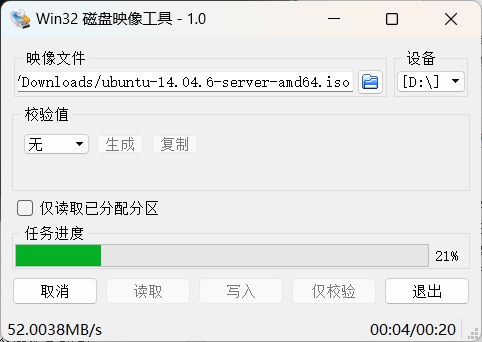


图 3 启动盘制作

写入完成后，会弹出提示框，确认写入已完成。此时，U 盘已经制作成了启动盘。

## 进入BIOS设置U盘启动

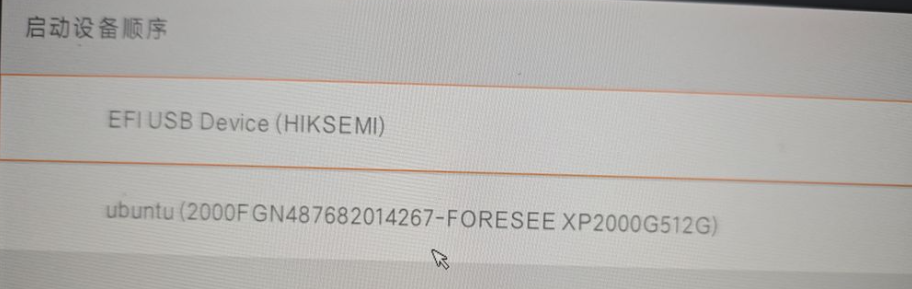


图 4 设置U盘启动

如上图所示，将制作好的linux启动盘插入电脑。开机时按f12（不同电脑不同）进入BIOS，寻找到启动设备顺序一栏目，设置U盘为第一启动项。接着重新启动电脑，便能够进入安装程序了。这里我安装的带桌面gui版本系统，后续直接按照系统教程流程安装即可。

# 二、LINUX操作系统命令的应用与相应功能

## 2.1常见的进程相关命令应用与相应功能

1. htop

htop 是一个交互式的进程查看器和系统监视工具，广泛用于Linux和类Unix操作系统中。与传统的 top 命令相比，htop 提供了更为友好和可视化的界面，允许用户更直观地查看系统性能和进程信息。

htop 的主要作用是实时显示系统中各个进程的状态、资源使用情况以及其他系统信息，帮助用户更好地监控和管理系统性能。相比于传统的 top 命令，htop 提供了更强的交互性、支持颜色显示，并可以方便地进行排序、筛选和管理进程。

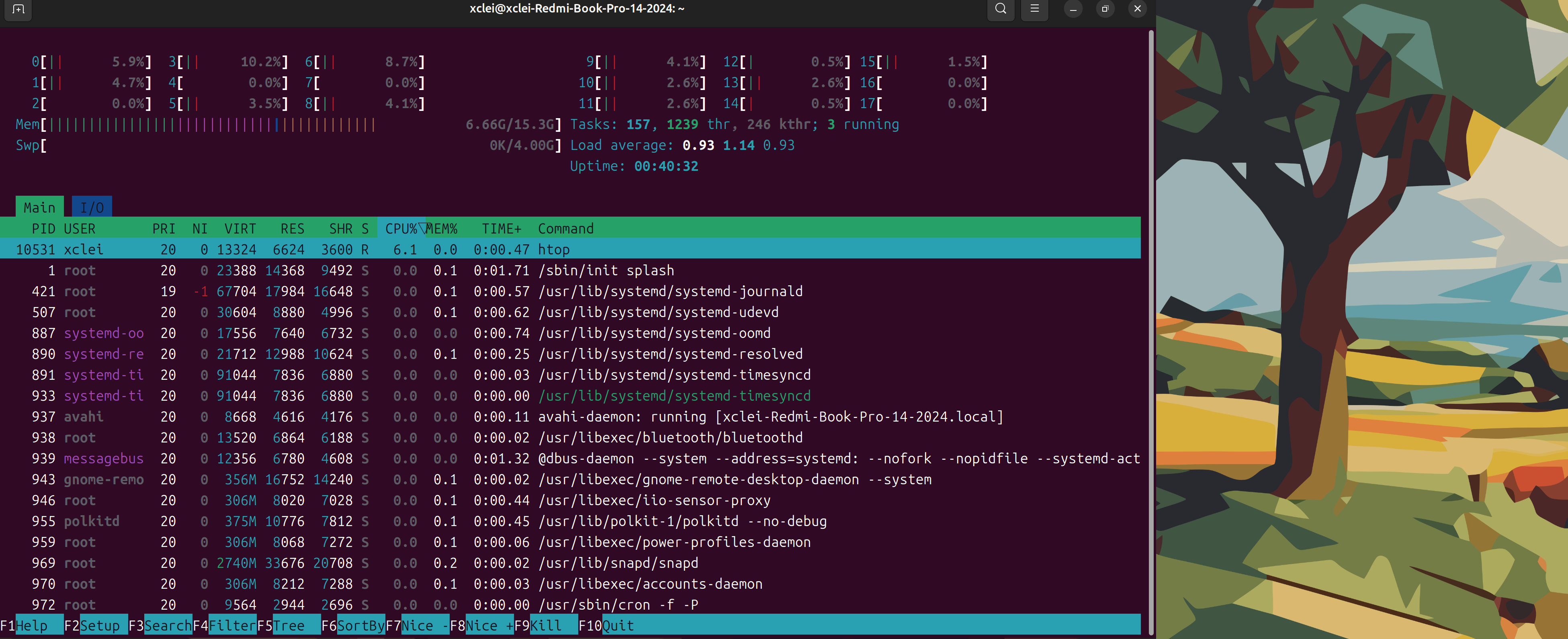


图 5 htop命令的应用（总体资源使用情况）

如上图3所示，屏幕顶端显示了以下内容：

(1) CPU 使用情况：以不同的颜色显示不同 CPU 核心的使用情况，图形化显示CPU的负载（如，绿色为用户空间占用、蓝色为系统空间占用、红色为等待I/O的进程等）。

(2)内存使用情况：显示当前物理内存的使用情况，包含已用空闲缓存和交换空间的大小。

(3)交换区（Swap）使用情况：显示交换分区的使用情况。

(4）负载平均值：显示系统的负载情况（1分钟、5分钟、15分钟的平均负载）。

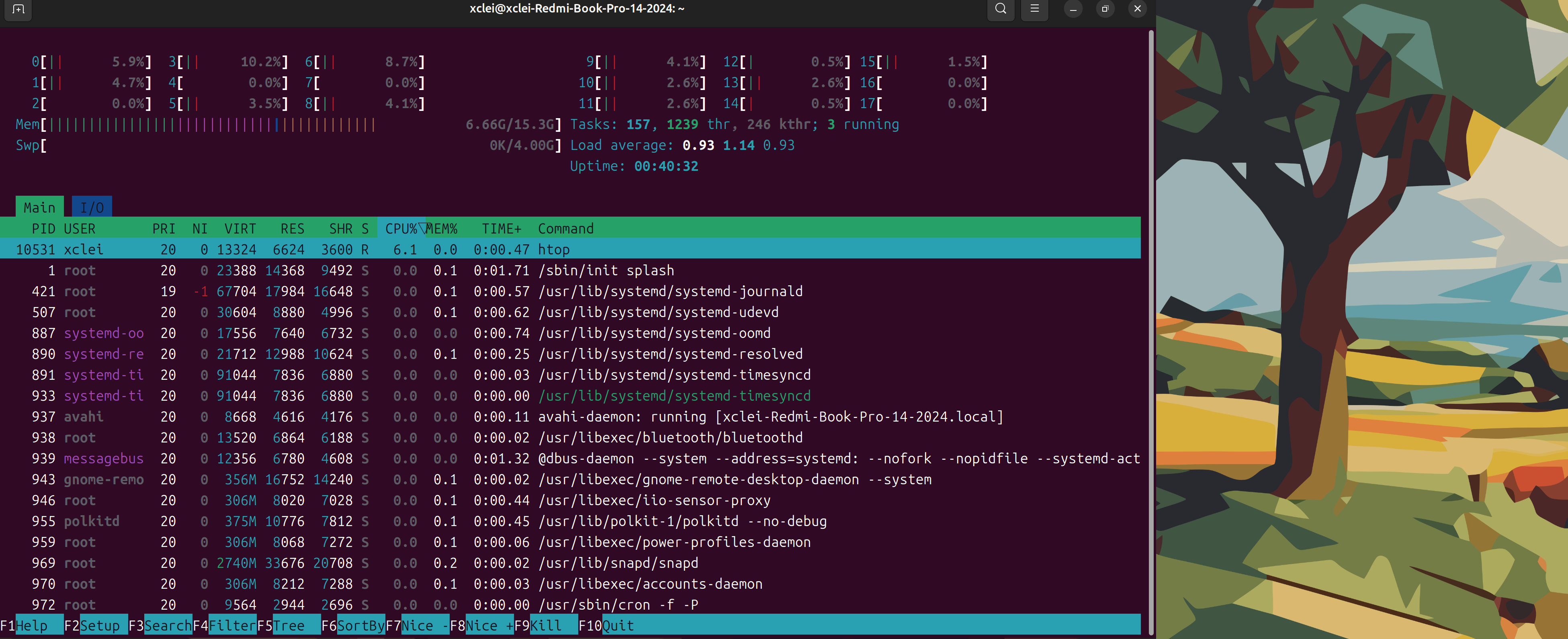


图 6 htop命令的应用（进程相关）

如图4所示，图中不同字段分别有如下表示：

PID：进程的唯一标识符（进程ID）。

USER：进程的所有者。

PR：进程的优先级。

NI：进程的"nice"值，影响进程的调度优先级。

VIRT：进程占用的虚拟内存大小。

RES：进程占用的物理内存大小。

SHR：进程使用的共享内存的大小。

S：进程的状态，常见的状态有：

S：睡眠（Sleeping）

R：运行中（Running）

Z：僵尸（Zombie）

T：停止（Stopped）

%CPU：进程占用的CPU百分比。

%MEM：进程占用的内存百分比。

TIME+：进程的累计CPU时间。

COMMAND：执行的命令或进程名称。

主要操作：

F6 键选择排序方式（如按照 CPU、内存、PID 排序等）。

F3 键搜索特定进程，快速筛选进程列表。

F4 键进行自定义筛选，过滤进程。

F9：杀死进程，你可以选择一个进程并杀死它。

F10：退出 htop。

通过上述htop命令即其相关操作，用户能够轻松掌控系统资源使用情况并灵活地进行进程相关信息的查看以及进程的控制。

1. kill

kill命令是常见的用于终止进程的命令，kill 命令用于向进程发送信号，通常用于终止进程。

(1)kill <PID>：发送默认的终止信号（SIGTERM）给指定进程ID（PID）。

(2)kill -9 <PID>：强制杀死指定的进程（发送 SIGKILL 信号）.

(3)killall <process\_name>：终止指定名称的所有进程。



图 7 kill命令的使用

如图5所示，将刚刚htop中获取到的进程pid记录下来，并用kill命令进行终止。命令执行结束后对应的火狐浏览器进程被关闭。

1. pgrep

根据名称查找进程ID

pgrep 用于查找匹配某个模式的进程。

pgrep <process\_name>：根据进程名称查找并返回进程ID。

pgrep -u <username> <process\_name>：查找指定用户的某个进程

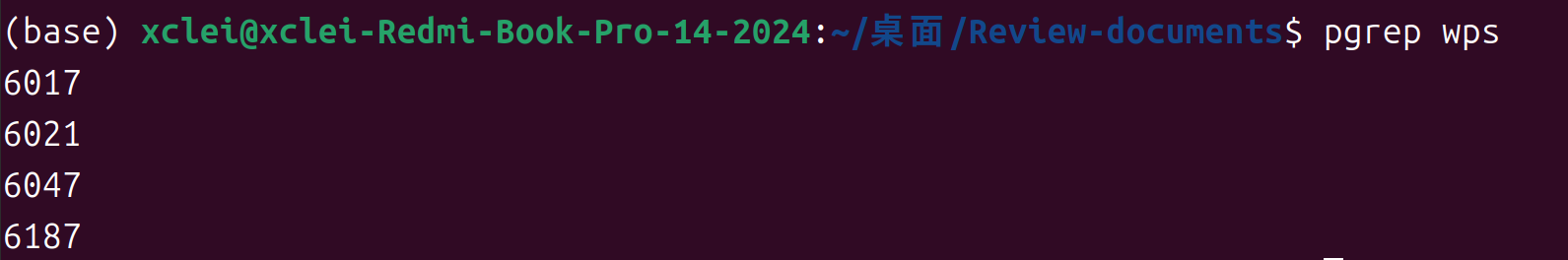


图 8 pgrep查找进程对应pid

## 2.2其他常用命令

1、dmesg

dmesg 命令的主要功能是输出内核日志消息，包括：

系统启动信息（如硬件检测、驱动加载）

硬件设备的识别和状态（如磁盘、网卡、USB 设备等）

内核模块的加载和卸载

系统错误和警告信息（如内存错误、硬件故障等）

I/O 设备的连接或断开事件

这些消息可以帮助用户诊断内核和硬件问题，比如硬盘、网络接口、声卡等设备的故障，以及系统崩溃时的错误信息。

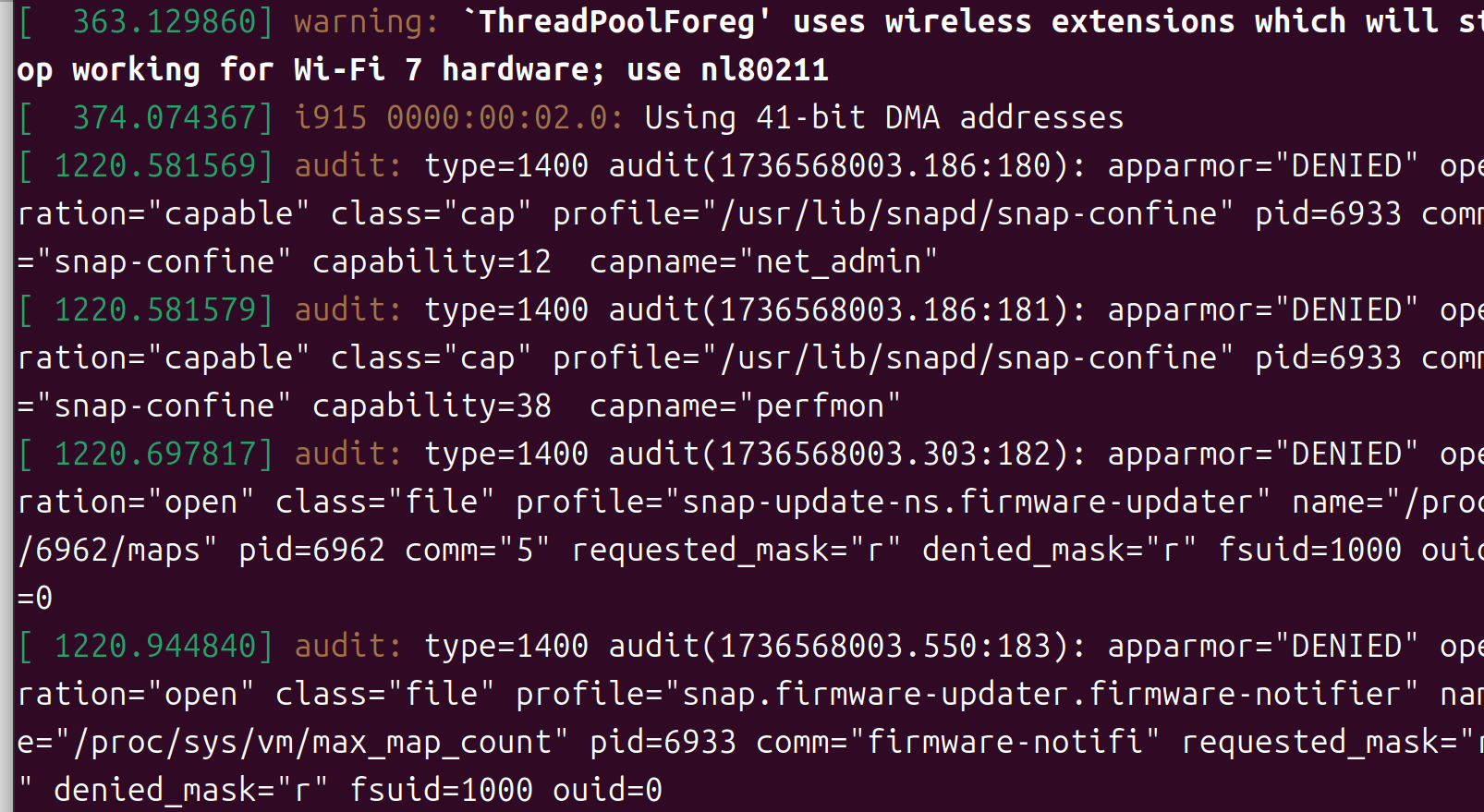


图 9 dmesg内容

| **字段** | **说明** |
| --- | --- |

|  |  |
| --- | --- |
| [时间戳] | 消息发生的时间，通常是从系统启动起的秒数 |

|  |  |
| --- | --- |
| 内核模块 | 生成消息的内核模块或驱动程序 |

1. netstat

netstat（network statistics）是 Linux 系统中常用的网络状态查看工具，用于显示与网络相关的各种信息，包括网络连接状态、端口使用情况、路由表、网络接口统计等。

尽管在现代系统中，netstat 已被 ss 命令逐渐取代，但它依然是网络诊断和故障排查的重要工具，许多场景下仍然适用。



图 10 netstat命令

netstat 的主要功能包括：

显示当前系统的网络连接（TCP、UDP、UNIX 套接字等）

显示正在监听的端口和服务（识别哪些端口正在开放）

显示网络接口统计信息（如数据包发送和接收情况）

显示路由表（查看系统的路由配置）

显示多播组信息

这些功能非常有助于网络性能监控和故障诊断，尤其是检测是否存在恶意连接或非预期的端口开放。

# 三.LINUX操作系统系统软件的安装与配置

## 3.1 ssh服务安装

SSH（Secure Shell）是一种用于远程登录和管理服务器的加密网络协议，提供对远程系统的安全访问，取代了传统的 Telnet、FTP 等不安全的协议。SSH 通过加密通信保护用户的密码和数据，常用于服务器管理、文件传输和远程执行命令。

安装命令(使用apt命令安装)：

sudo apt update

sudo apt install openssh-server



图 11 ssh下载

2.配置 SSH 服务

SSH 服务的配置文件通常位于 /etc/ssh/sshd\_config。可以通过编辑该文件来自定义 SSH 的行为。

sudo vim /etc/ssh/sshd\_config

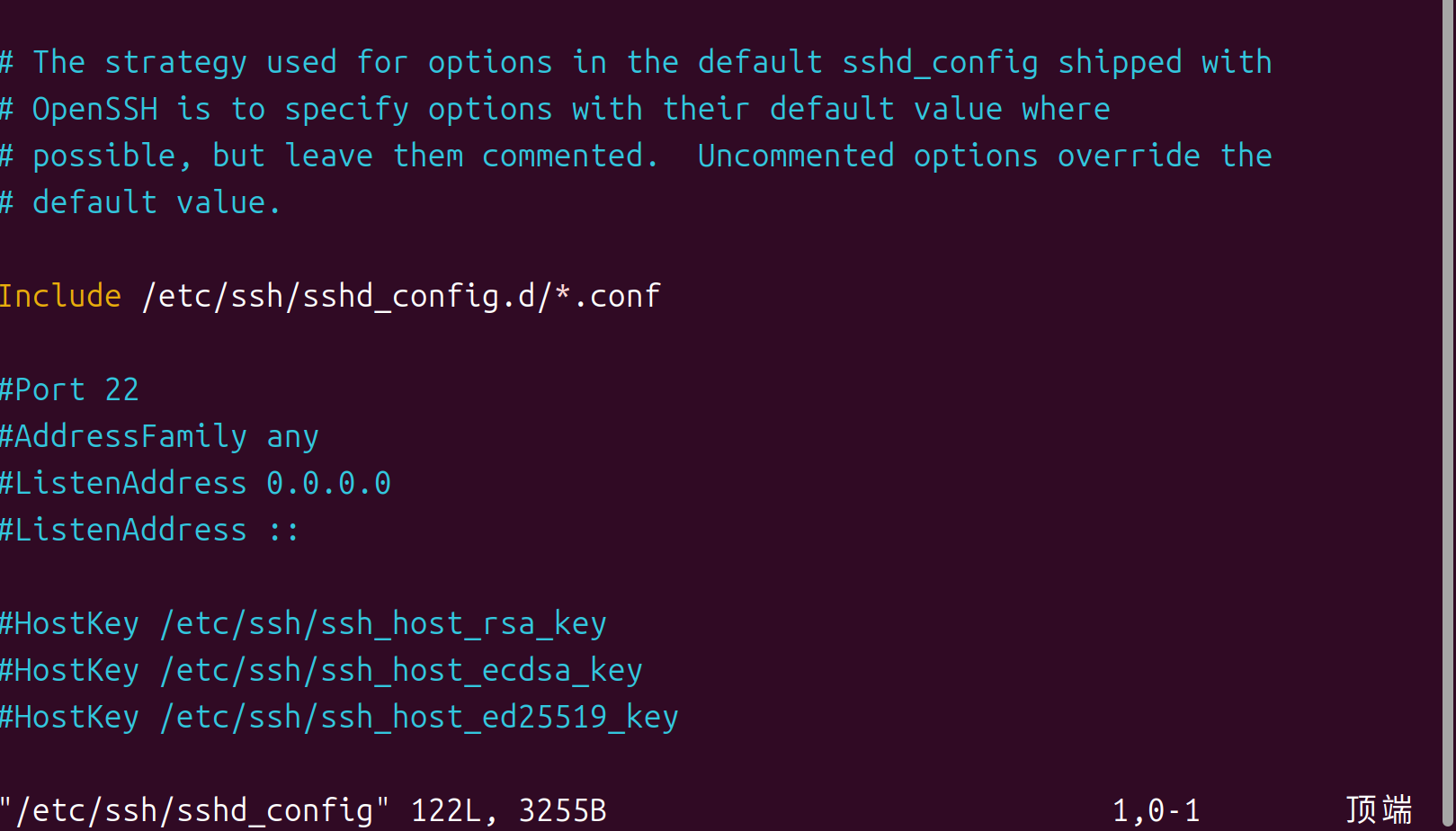


图 12 ssh配置文件

|  |  |
| --- | --- |
| Port 22 | SSH 服务的默认端口号（可更改为其他端口） |

|  |  |
| --- | --- |
| PermitRootLogin no | 是否允许 root 用户通过 SSH 登录（建议禁用） |

|  |  |
| --- | --- |
| PasswordAuthentication yes | 是否允许密码验证登录 |

|  |  |
| --- | --- |
| MaxAuthTries 3 | 最大认证尝试次数 |

|  |  |
| --- | --- |
| AllowUsers | 指定允许登录的用户 |

3.启动ssh服务

sudo systemctl start ssh

设置 SSH 服务开机自启

sudo systemctl enable ssh

4.在另一台电脑上的登陆

在另一台电脑上，通过 SSH 客户端连接到服务器。

Windows 客户端：使用 PuTTY

打开 PuTTY 客户端。

在 Host Name 输入服务器的 IP 地址。

在 Port 输入 22 或自定义端口号。

点击 Open，输入用户名和密码即可登录。

# LINUX操作系统下HADOOP的伪分布安装与配置

Hadoop 是一个用于大数据处理的开源框架，它可以分布式存储和处理大规模数据。在单机环境下进行 Hadoop 安装时，我们通常采用 伪分布式模式，即所有 Hadoop 的守护进程都在单个节点上运行，但每个进程都模拟一个独立的分布式环境。伪分布式模式适用于开发、测试和学习目的，适合资源有限的环境。

以下是 Hadoop 在 Linux 系统下的伪分布式模式安装与配置步骤。

## 1.java环境安装检测

Hadoop需要安装java，先检查一下

java -version

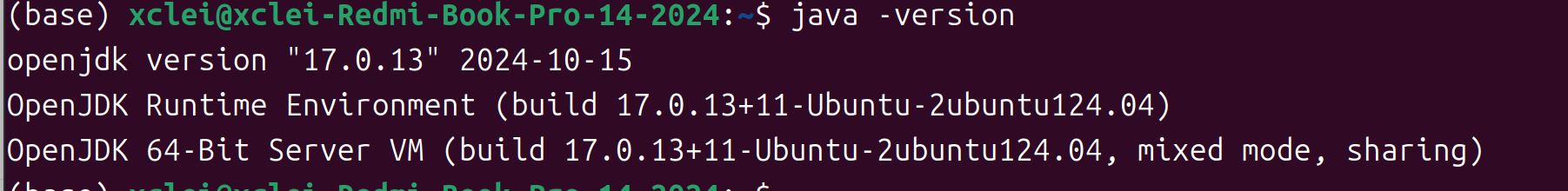


图 13 java安装检测

之前已经安装过，则无需再次装JDK。

## 2.创建用户

为 Hadoop 创建一个专门的用户，这样可以保证 Hadoop 的运行与系统其他部分隔离。

sudo useradd hadoop

sudo passwd hadoop

## 3.更改host内容

设置主机名和 /etc/hosts 配置

伪分布式模式要求集群节点能通过主机名进行通信。配置 /etc/hosts 文件，使其支持主机名解析。

sudo vim /etc/hostname

修改 /etc/hosts

编辑 /etc/hosts，添加以下内容：

127.0.0.1 localhost 127.0.1.1 hadoop-node

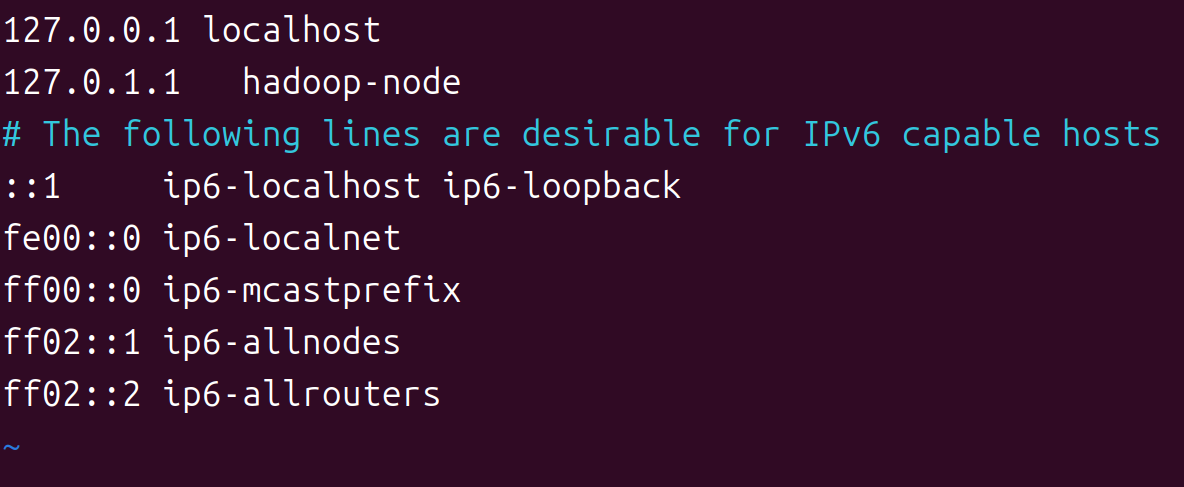


图 4 hadoop 配置文件更新

修改后更新配置

sudo hostname hadoop-node

## 4.安装 Hadoop

用wget从官网下载安装包安装：

wget

https://downloads.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz

解压并安装

sudo tar -zxf hadoop-3.4.0.tar.gz -C /usr/local

更改文件名并增加权限

sudo mv ./hadoop-3.4.0/ ./hadoop

sudo chown -R hadoop ./hadoop/

## 测试是否成功安装hadoop

./bin/hadoop version

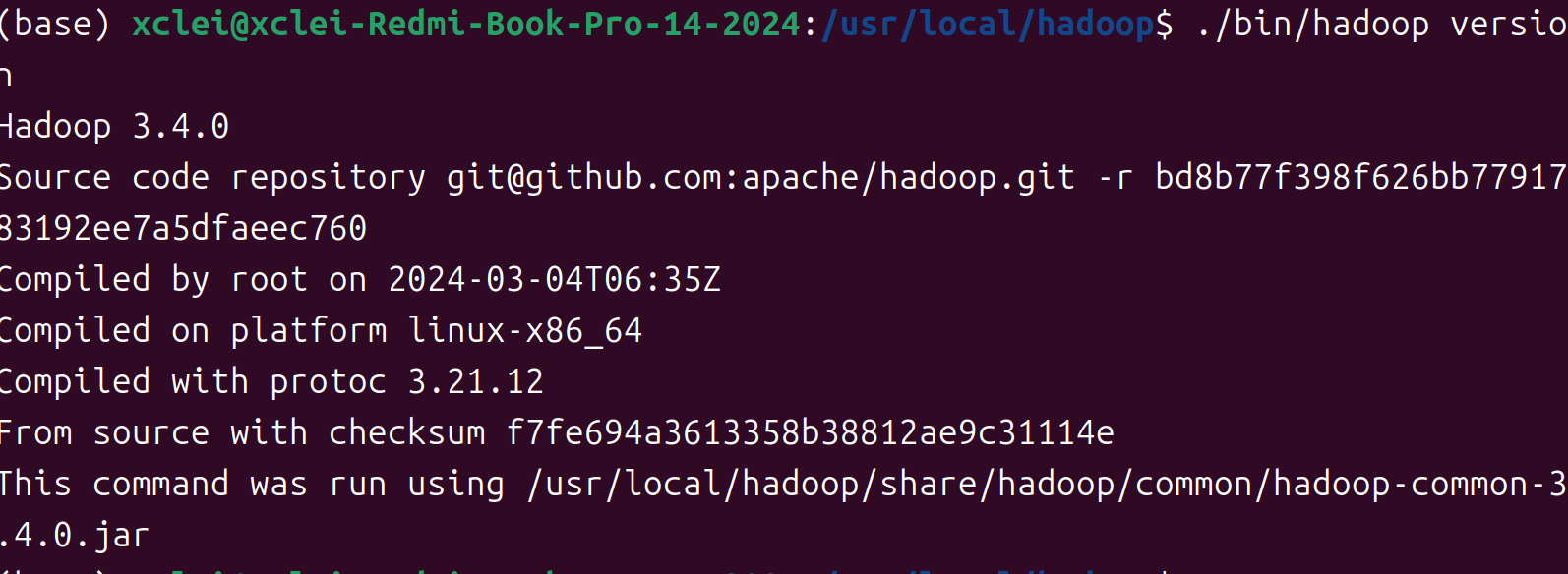


图 15 hadoop安装测试

能够看到HADOOP已经成功安装。

## hadoop单机/伪分布式

1.进行NameNode格式化

sudo ./bin/hdfs namenode -format

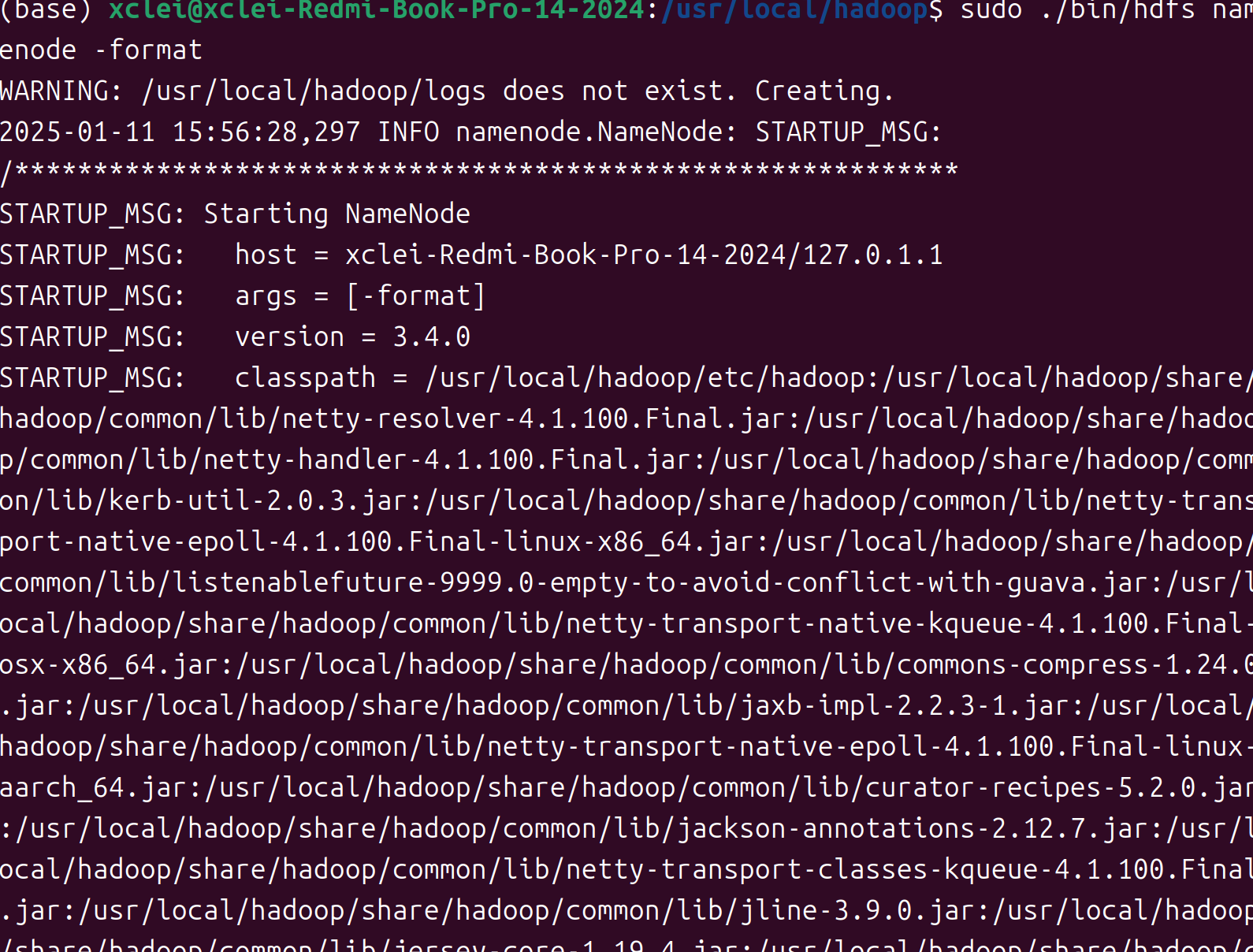


图 16 NameNode格式化

2.启动NameNode 和 DataNode 守护进程

./sbin/start-dfs.sh

并输入jps检测

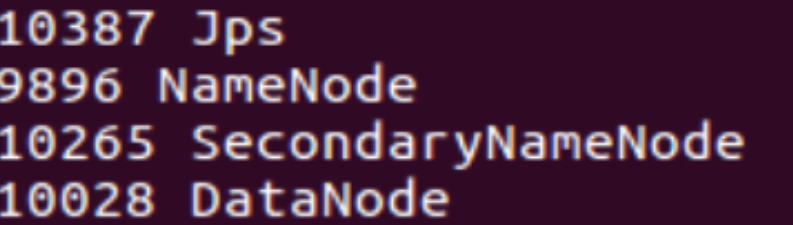


图 17 jps查看

能够看到hadoop伪分布式成功部署。

# 5.总结与设计中遇到问题的解决

## 问题1.JAVA\_HOME 环境变量配置错误

ERROR: JAVA\_HOME /usr/lib/jvm/java-17-openjdk-amd64/bin/java does not exist.

环境变量 JAVA\_HOME 的路径配置错误，指向了 bin/java 文件，而不是 JDK 的根目录。Hadoop 启动脚本需要 JAVA\_HOME 指向 JDK 的安装目录。

解决方法：

修改 Hadoop 的 hadoop-env.sh 文件：

vim /usr/local/hadoop/etc/hadoop/hadoop-env.sh

添加JDk根目录

export JAVA\_HOME=/usr/lib/jvm/java-17-openjdk-amd64

## 问题 2：Hadoop 启动时提示 SSH 认证失败

Starting namenodes on [xclei-Redmi-Book-Pro-14-2024]

xclei-Redmi-Book-Pro-14-2024: xclei@xclei-redmi-book-pro-14-2024: Permission denied (publickey,password).

Starting datanodes

localhost: xclei@localhost: Permission denied (publickey,password).

Hadoop 依赖 SSH 进行节点之间的通信，提示的错误说明当前用户未配置 SSH 的无密码登录。

解决方法：

生成 SSH 密钥对：  
ssh-keygen -t rsa

添加授权文件

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

设置权限：

chmod 700 ~/.ssh chmod 600 ~/.ssh/authorized\_keys

## 问题3：错误修改环境变量导致无法进入系统

在修改环境变量的时候，不小心错误地设置了 PATH 环境变量，导致无法进入系统终端。

覆盖了原有的 PATH，导致系统无法找到常用命令。

导致在进入系统输入用户密码的时候导致无限循环，进步去系统。

解决方法：

我用手机搜了搜有很多人都犯过这个错误，解决方法很简单：在开机的时候一直按shift键选择高级启动选项，选择recovery（恢复）模式。选择root进入命令行，删除错误的环境变量即可。

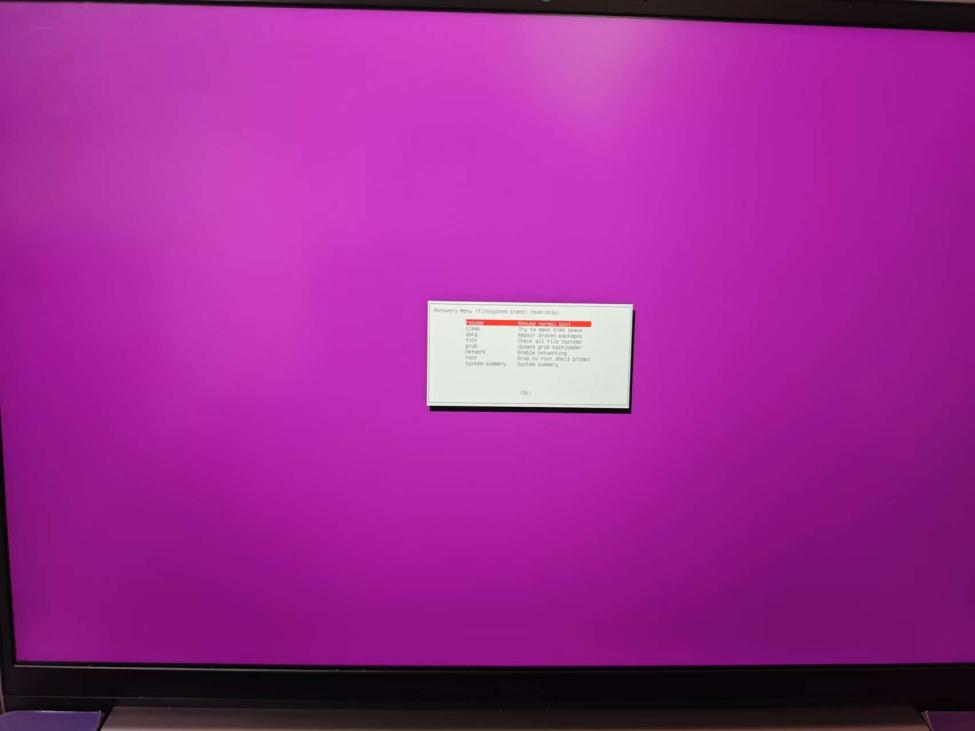


图 18进入系统恢复模式

## 课程设计总结

通过本次 Linux 操作系统课程设计，我对 Linux 系统的基本操作、系统软件的安装与配置以及大数据平台 Hadoop 的部署有了更加深入的理解和实践经验。在课程设计过程中，我完成了系统软件的配置与服务的部署，包括 SSH 服务的配置、Apache Web 服务器的搭建以及 Samba 文件共享服务的实现。同时，我成功部署了 Hadoop 的伪分布式环境，并解决了多个实际操作中的问题，如环境变量配置错误、权限管理不当、SSH 无密码登录失败等。在这个过程中，我不仅掌握了各项技术的具体操作步骤，还学会了如何分析错误日志，快速定位并解决系统问题。

在系统软件的安装与配置中，我学会了通过命令行进行服务的管理，熟悉了各类系统文件的修改方法。SSH 服务的配置让我理解了公钥认证机制和用户权限管理的重要性，Apache 的安装让我掌握了 Web 服务的基本配置方法，而 Samba 的搭建则增强了我对跨平台文件共享的理解。这些实践操作让我更加熟悉 Linux 系统环境，为今后的学习和工作奠定了坚实的基础。

在 Hadoop 的伪分布式环境部署过程中，我深入学习了 HDFS 和 YARN 的基本架构，掌握了核心配置文件的修改方法以及 Hadoop 的启动和管理流程。然而，在实际部署时遇到了诸多挑战，例如 Hadoop 启动时的权限错误、环境变量配置错误导致服务无法启动、SSH 无密码登录配置不成功等。这些问题让我意识到，Linux 系统的使用不仅仅是对命令的简单执行，更需要有逻辑思维和解决问题的能力。我通过查看系统日志和错误信息，不断分析问题的根本原因，最终成功解决了这些问题。这让我深刻体会到，在实际操作中，面对问题时需要保持冷静，善于利用系统提供的提示信息来找到解决方案。

本次课程设计最大的收获在于，我学会了如何从错误中学习和成长。在遇到问题时，我不再盲目地重复操作，而是通过分析日志、查阅文档、搜索解决方案等方式，逐步定位并解决问题。这种问题分析与解决的能力，是我在本次课程设计中获得的最宝贵的经验。此外，我还意识到，在系统配置中一定要谨慎操作，特别是在修改系统环境变量时，务必要确保配置正确，避免因错误的配置影响系统的正常运行。

总的来说，本次课程设计让我对 Linux 系统的使用有了更为全面和深入的理解，特别是在系统管理、服务配置和大数据平台部署方面，积累了丰富的实践经验。通过本次课程设计，我不仅提升了动手能力和技术水平，还培养了独立思考和解决问题的能力。这些收获将对我未来的学习和工作产生深远的影响。Linux 操作系统的应用场景十分广泛，本次课程设计为我未来进一步探索服务器管理、大数据技术以及云计算平台等领域提供了重要的基础。我将继续深入学习 Linux 系统，不断提升自己的技术能力，为未来的发展做好准备。

1.内容

内容主要包括三个方面：

一是熟悉和掌握LINUX操作系统的使用，通过命令查看LINUX作系统的进程.实现LINUX操作系统的进程管理方面；

二是熟悉和掌握LINUX操作系统下系统软件的安装与配置，包括SAMBA,APACHE,SSH；

二是掌握LINUX操作系统下HADOOP的伪分布安装与配置；

2. 操作系统课程设计报告要求

主要包括：

1. LINUX操作系统安装的步骤
2. LINUX操作系统命令的应用与相应功能（至少写出三个）
3. LINUX操作系统系统软件的安装与配置（三个至少写一个）
4. LINUX操作系统下HADOOP的伪分布安装与配置
5. 总结与设计中遇到问题的解决（这一部分的评分占比较高，大家注意）

3. 评分标准

1. 报告内容完整，30%

2. 报告工具操作，演示，步骤清楚， 20%

3. 报告完成了目标，20%；

4. 总结的问题的具体，体现实际操作的经验总结，30%