



生物数据处理技术

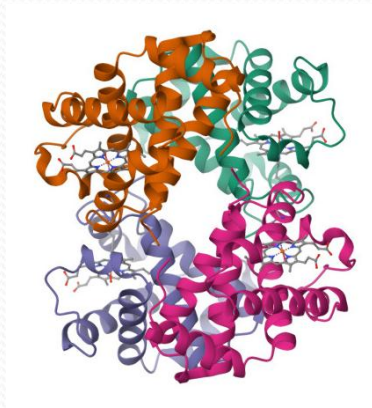
解增言

生物信息学院





斑头雁 (*Anser indicus*)



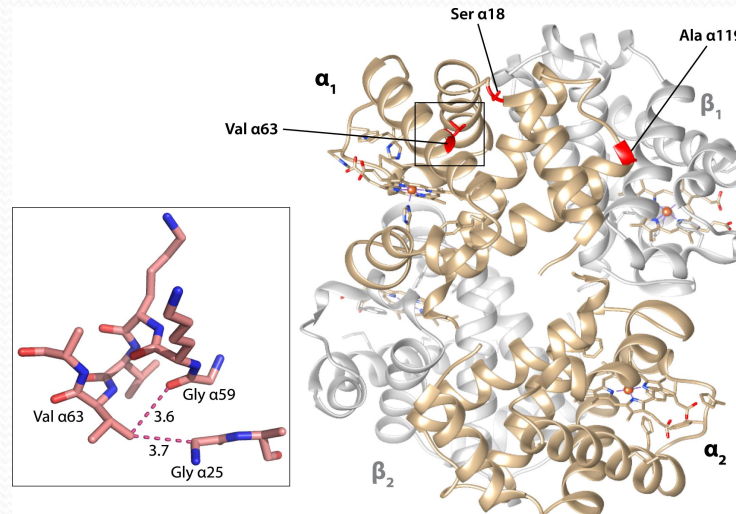
灰雁 (*Anser anser*)

>ACT80359.1 hemoglobin alpha A subunit [Anser indicus]

MVLSAADKTNVKGVF^{SKIS}GHAEYGAETLERMFTAYPQTKTYFP^{PH}FDLQH^{GS}SAQIKAHGKKV^VAALVEA
VNHIDDIAGALSKLS^{DL}HAQKL^RVD^PVNF^KFLGHCFLV^VVAIH^{HP}SALT^AEVHASLDKFLCAVGT^VLTAK
YR

>ACT80855.1 hemoglobin alpha A subunit [Anser anser]

MVLSAADKTNVKGVF^{SKI}G^HGHAEYGAETLERMFTAYPQTKTYFP^{PH}FDLQH^{GS}SAQIKAHGKKV^AAALVEA
VNHIDDIAGALSKLS^{DL}HAQKL^RVD^PVNF^KFLGHCFLV^VVAIH^{HP}SALT^PEVHASLDKFLCAVGT^VLTAK
YR



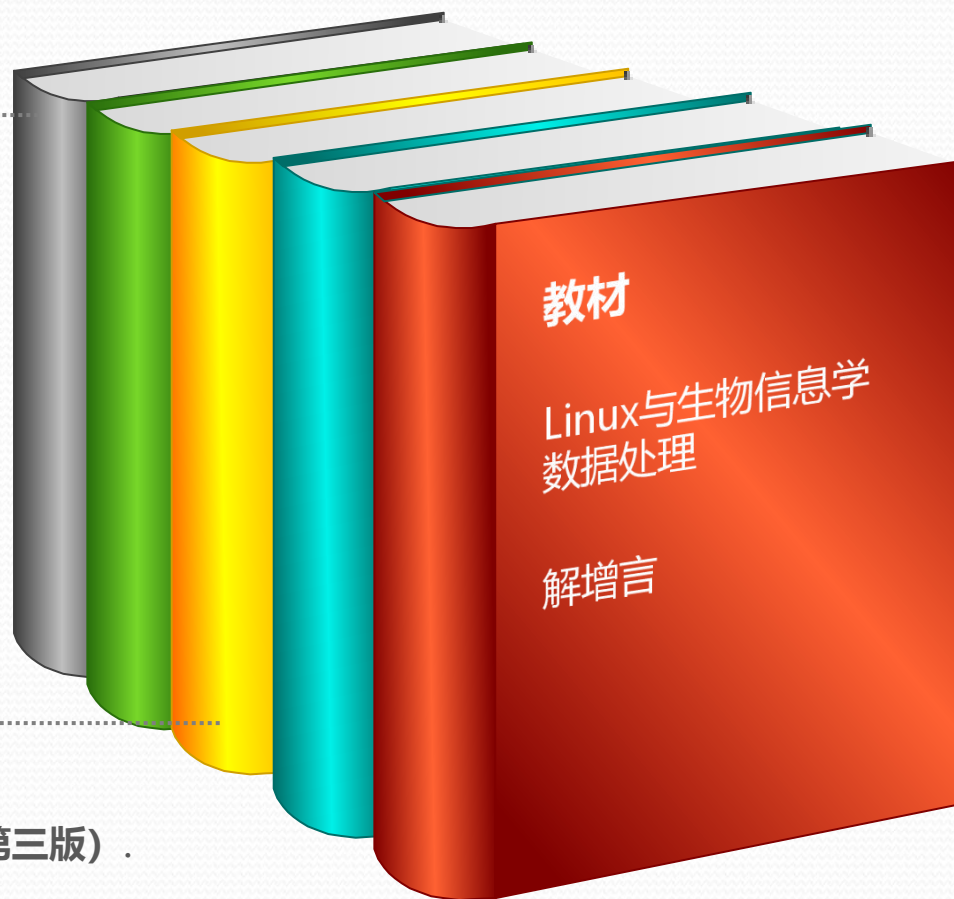
教材与参考书

参考书1

鸟哥的Linux私房菜
(基础篇第四版)

参考书2

Daniel J. Barrett,
Linux命令速查手册 (第三版) .
中国电力出版社, 2018



课程网站与上机账号

∞课程网站:

∞ <https://www.xueyinonline.com/detail/236158134>

∞ <http://www.linuxstudio.cn>（免费注册）

∞作业及练习服务器:

∞ www.linuxstudio.cn

∞用户名：姓名全拼（如wangbin）

∞密码：姓名全拼2023（如wangbin2023）。请及时登录并用passwd命令更改密码



考核方式

平时成绩 (100分 × 60%) :

- 作业: 20分 (每次2分, 共10次)
- 随堂测试: 20分 (每次5分, 共4次)
- 雨课堂: 20分
- 课程网站学习情况: 20分
- 学银慕课学习情况: 20分
- 缺课: 每次-5分
- 附加任务: 教材找错 (每处错误确认后+1分, 第一个找到的同学有效, 最高至平时成绩满分)

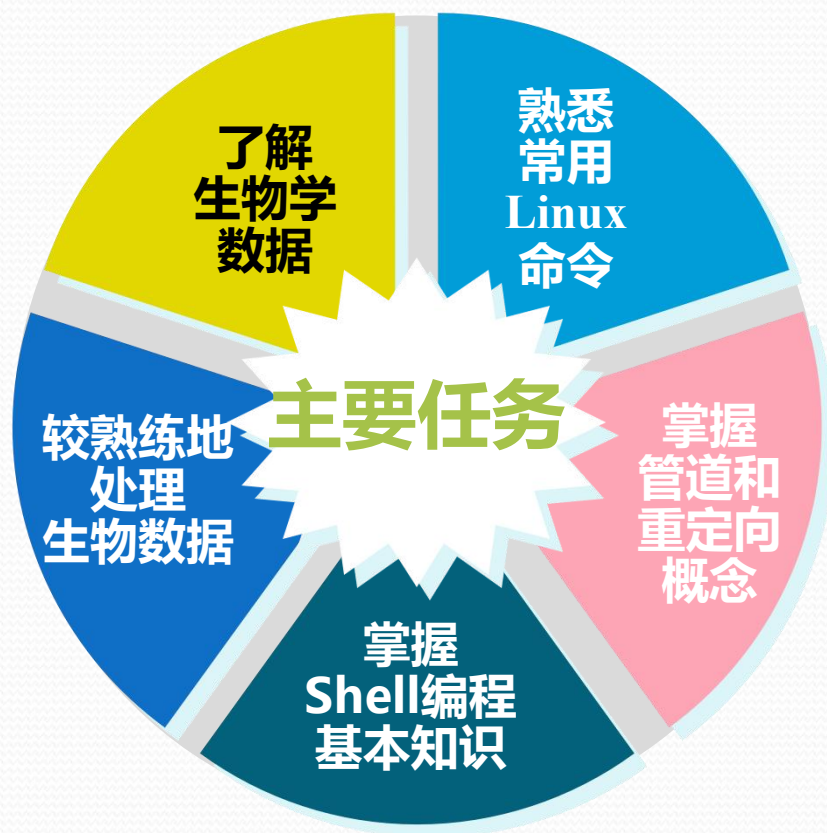
期末考试 (100分 × 40%) :

- 单选、多选、填空、简答 (除简答外, 题目来自 www.linuxstudio.cn 题库)

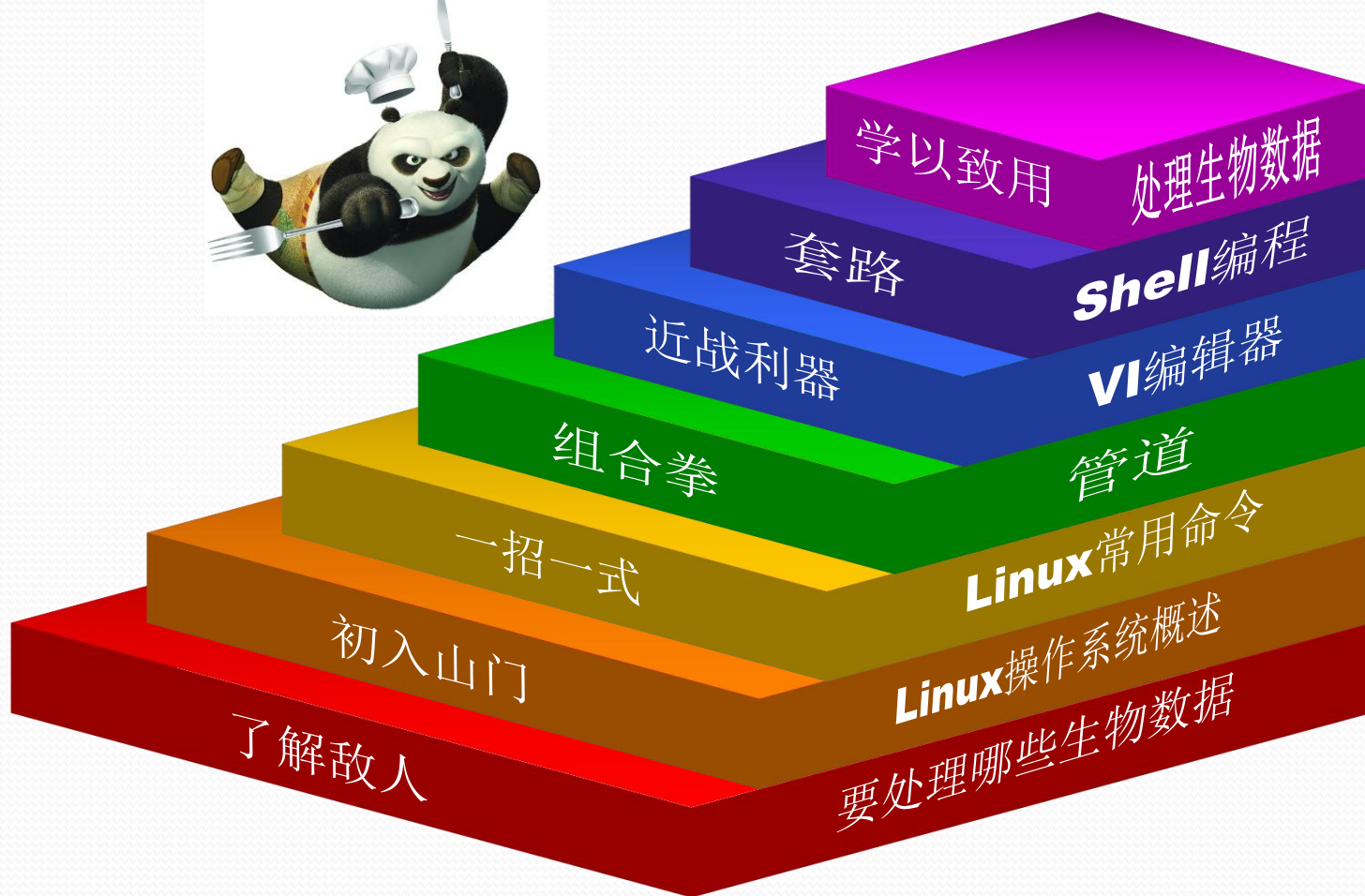


课程简介

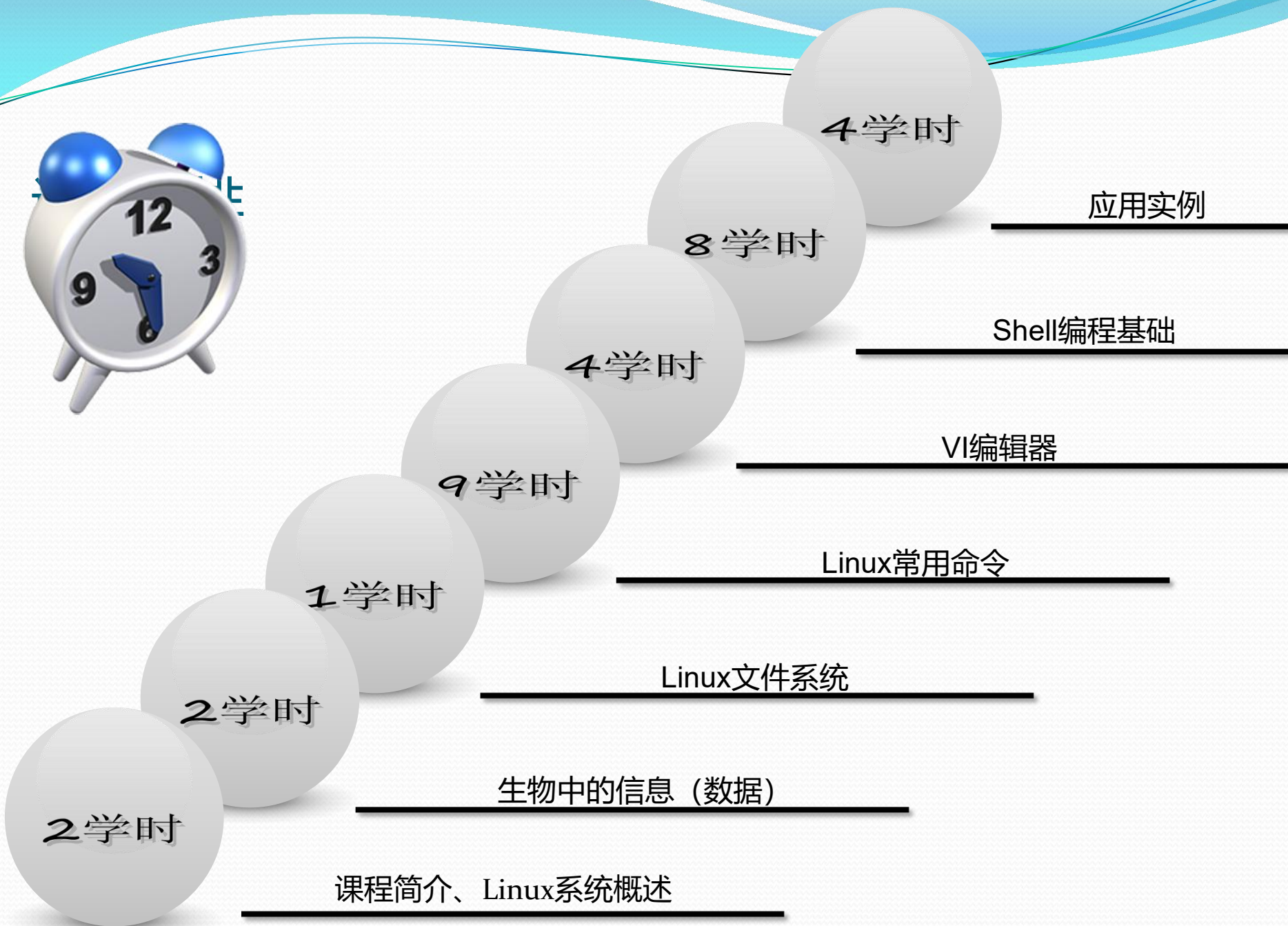
生物数据处理技术课程主要内容是学习利用Linux命令、管道及Shell编程，处理生物学数据（主要是分子生物学数据）。



课程内容



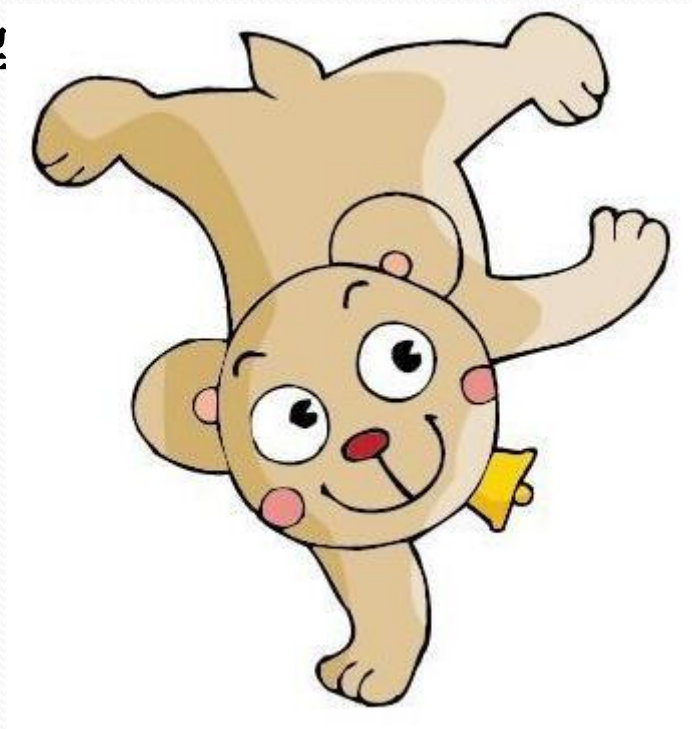
课程内容



上课形式

∞ 翻转课堂

- ✓ 课前：观看教学视频，提前练习命令
- ✓ 课内：练习、作业、测试、解
- ✓ 课后：复习



怎样获得一个好成绩

- ☞ 按时预习：上课前完成指定的内容，完成雨课堂推送的课件预习；
- ☞ 完成课堂任务：上课时专心听讲，及时完成雨课堂的随堂练习；
- ☞ 按时完成作业：根据要求，按时完成服务器上的作业练习；
- ☞ 学银慕课：按时完成慕课教学视频学习及测试，积极参加慕课平台的讨论；
- ☞ 课程网站：按时完成课程网站的文字材料学习及练习，多做自主测试，参加课程阶段测试；
- ☞ 按时上课：按时签到，不迟到，不旷课。
- ☞ 最重要的一点：多练习、理解Linux命令，并用来解决实际问题。

Linux系统概述

重庆邮电大学 生物信息学院
解增言





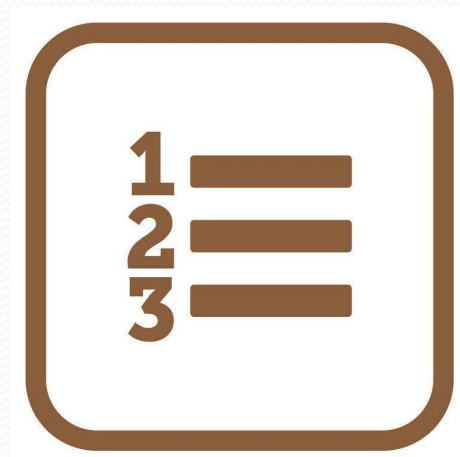
Mac OS

UNIX[®]
00011110 00011110 00011110 00011110 00011110 00011110 00011110

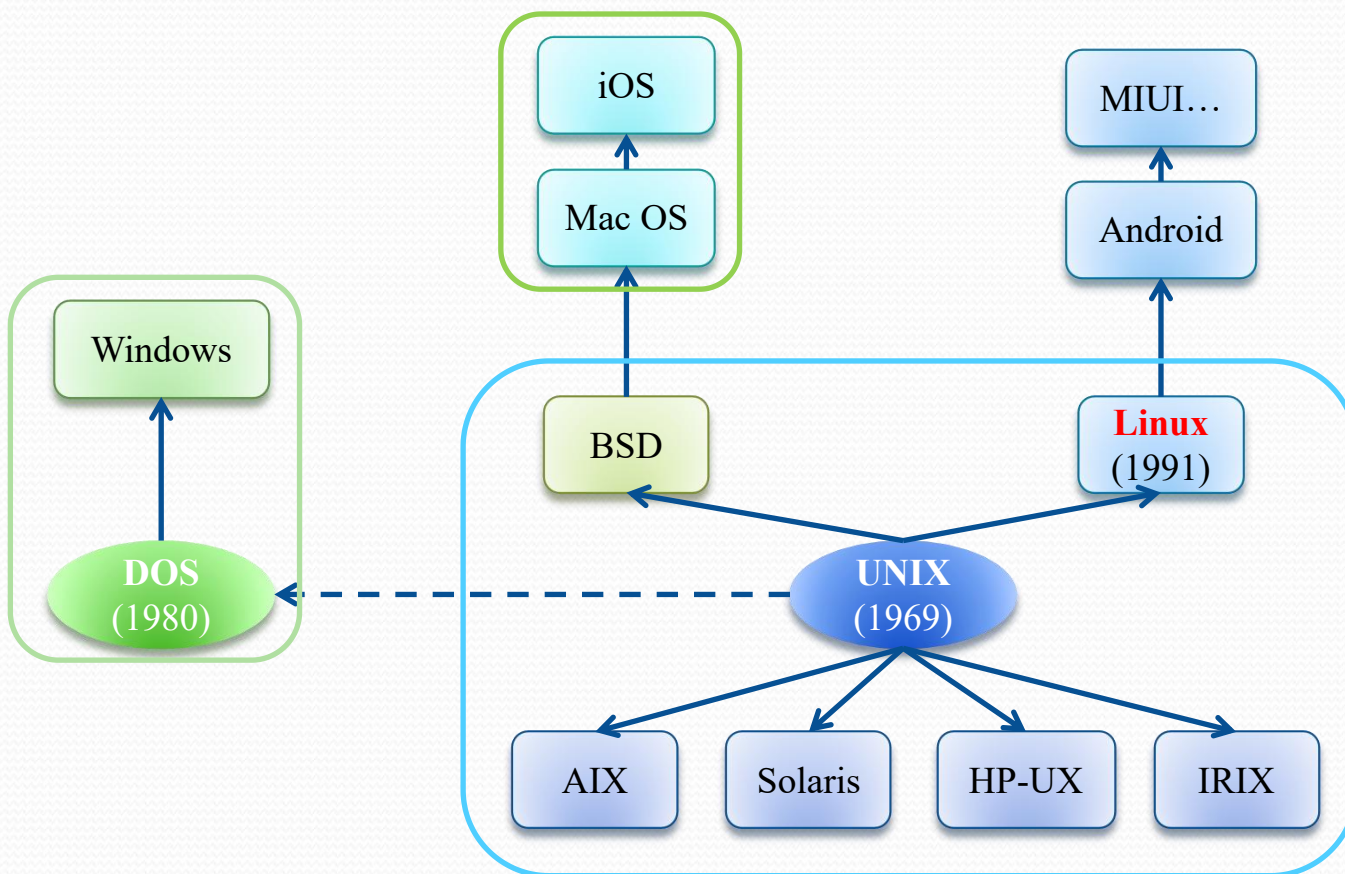


内容

- 一、Linux的发展历史
- 二、Linux系统的特点
- 三、Linux系统应用
- 四、Linux系统安装
- 五、Shell编程简介



一、Linux的发展历史



操作系统演变

Linux的发展历史

- ∞ 1969, Bell Laboratory, **UNIX**
- ∞ 1984, Richard Stallman, **GNU** (GNU is Not Unix) and GPL (General Public License)
- ∞ 1987, Andrew S. Tanenbaum, **MINIX**
- ∞ 1991, Linus Torvalds, **Linux**内核, based on MINIX
- ∞ 1994.3, **Linux** 1.0 released
- ∞ 1995, Bob Young, **RedHat**
- ∞ 2022.8.1, Linux Kernel 5.19



Linux不同的发行版



2021年Linux发行版排行榜

- ☞ 1. CentOS
- ☞ 2. Ubuntu
- ☞ 3. Kali Linux
- ☞ 4. Arch Linux
- ☞ 5. Debian
- ☞ 6. Deepin （深度）
- ☞ 7. Linux Kylin （优麒麟）
- ☞ 8. Fedora
- ☞ 9. Manjaro
- ☞ 10. OpenSUSE



根据阿里云镜像下载统计数据

二、Linux系统的特点

∞ 命令行操作方式

∞ 自由软件

∞ 安全性高

∞ 多任务多用户

∞ 性能强大

∞ 网络支持好

∞ 扩展性好

三、Linux系统的应用

∞ 软件开发

∞ 网络服务器

∞ 超级计算机

∞ 特殊领域，如生物信息学

超级计算机Top500中

Linux操作系统比例：

1998 年：首次入榜

2012 年：94%

2013 年：95%

2014 年：97%

2015 年：97.2%

2016 年：99.6%

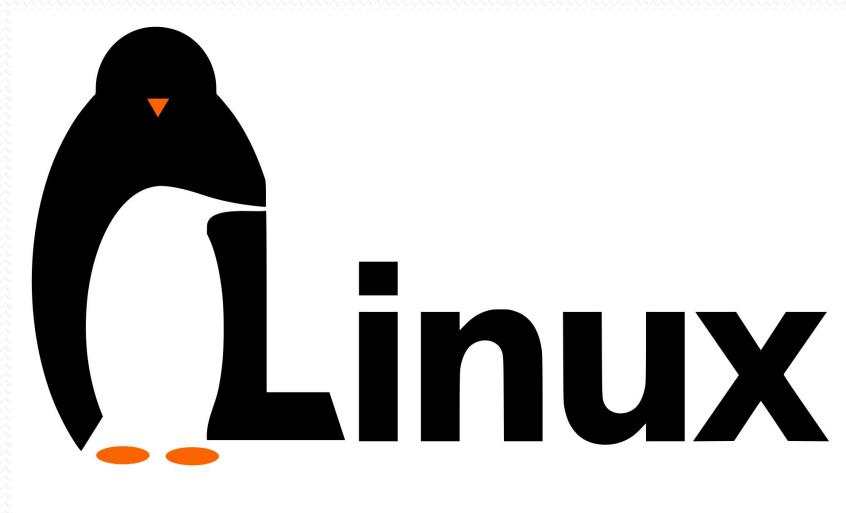
2017 年：100%

四、Linux系统安装

☞ 完整安装

☞ 虚拟机安装

☞ Windows 10/11 内置Linux子系统（WSL）（**推荐**）



完整安装

∞ 单系统：只安装Linux系统

∞ 双系统：Windows+Linux

∞ 分区：

建议>100G

SWAP 建议内存*2

/boot 200M

/home 剩余空间

性能好；但需要重新分区，安装复杂，使用不便。

虚拟机安装

先安装虚拟机软件：

VMware

Virtual Box

Virtual PC



不影响原系统；但性能较差，使用不太方便。

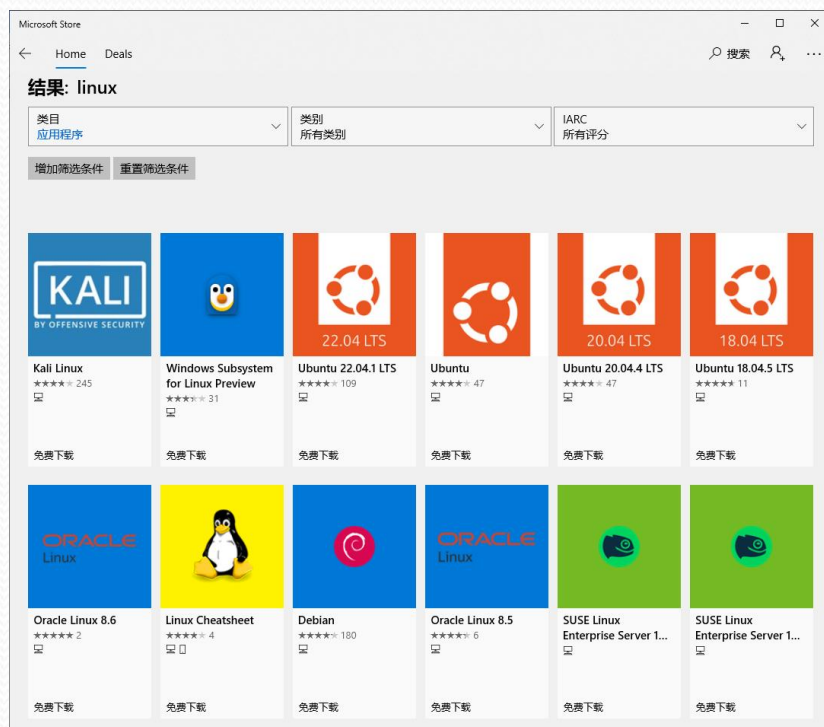


VirtualBox

Windows内置的Linux子系统

✧ 安装方便，不需要重新分区

✧ 使用方便，可共享宿主系统的文件和网络



五、Shell编程简介

编程语言按照运行方式可分为：

编译语言：C, C++, Fortran, ...

运行速度快，但开发耗时，适用于需要经常运行的软件或程序开发

脚本语言：Perl, Python, PHP, Awk, Ruby...

容易编写，不需要编译，直接运行，但运行速度慢，适用于日常数据处理

Shell是一种脚本语言

Shell是一种脚本语言，拥有脚本语言的所有特点。另外，编写shell程序需注意以下几个原则：

- ∞ 一个程序做一件事
- ∞ 善于利用正则表达式
- ∞ 输出时尽量利用标准输出，而不是写到文件
- ∞ 尽量避免不必要的提示信息
- ∞ 利用已有的工具完成任务最难的部分。写程序之前找一找这样的工具，避免重复前人的工作

Shell的 “Hello world” 程序

```
#!/bin/bash
```

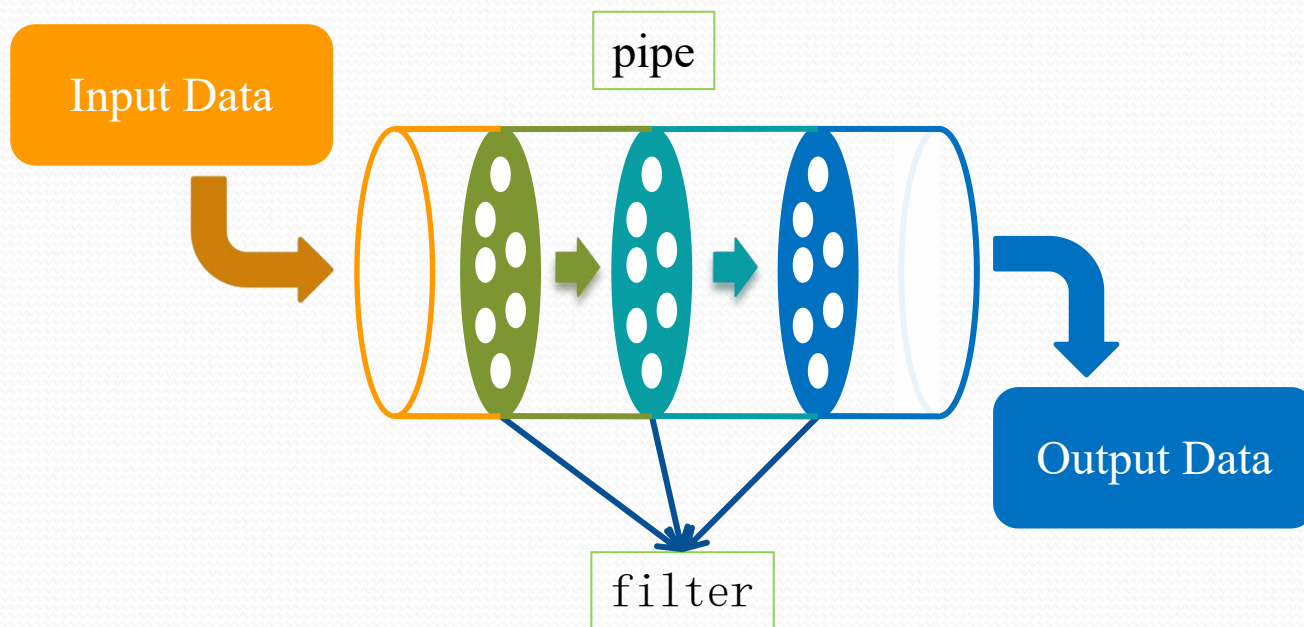
```
# This is a simple example of shell script
```

```
echo 'Hello world!'
```

总结

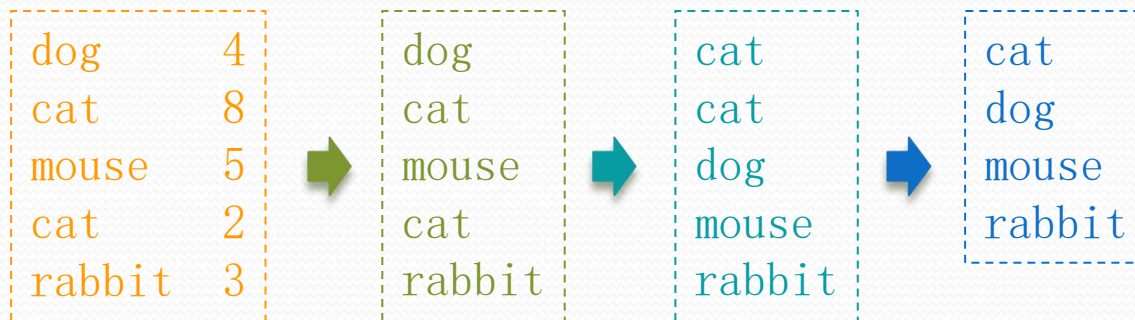
- ❧ Linux是基于UNIX的开源操作系统；
- ❧ Linux最早由Linus Torvalds开发，由开源社区共同开发维护；
- ❧ Linux安全性好，性能强大；
- ❧ Linux主要用于软件开发、服务器、超级计算机及一些专业领域；
- ❧ Shell是用户与操作系统内核之间沟通的桥梁，也是Linux的脚本语言的解释器。

Linux下的文本处理原理：管道



命令管道: `cat input.txt | cut -f1 | sort | uniq > output.txt`

数据变化:





The End