# 实验一 Linux集群基础

# 任务1 使用ssh远程登录服务器

文本

描述已自动生成

# 任务2 配置免密登录

## 生成公钥和私钥

文本

描述已自动生成

## 将公钥内容复制到服务器

文本

描述已自动生成

# 任务3 pwd、mkdir、cd命令

文本

描述已自动生成

# 任务4 cp、vim、ls、mv、rm命令

文本

描述已自动生成

# 任务5 cat、head、scp、awk、grep命令

## head、tail命令

文本

描述已自动生成

## 重定向符'>'的使用

文本

描述已自动生成

## cat命令

文本

描述已自动生成

文本

描述已自动生成

## scp命令

配置thumm01和thumm02的免密登录

文本

描述已自动生成

scp命令执行





## Awk命令

背景图案

中度可信度描述已自动生成



## Grep命令

文本

描述已自动生成

文本

描述已自动生成

# 任务6 阻塞和非阻塞时间对比

文本

描述已自动生成

# 任务7 多节点任务处理

## 集群主机之间免密登录配置

文本

描述已自动生成

## 集群批管理

图片包含 表格

描述已自动生成

## 在多主机上并行执行任务

背景图案

中度可信度描述已自动生成

# 任务8 尝试提出一种可以加快多节点处理速度的方法并验证

我们考虑将长文本分割为6份和12份来加快多节点处理速度。具体地，对WMT16数据集进行7000次的重复扩充，得到2G+的大数据集。



接着以类似任务7的方法将其平均分割为6份小数据集

绿色的标志

描述已自动生成

将6份小数据集分别发送至thumm01~thumm05，运行pssh "grep '^t' ~/multi-nodes/part > ~/multi-nodes/result"命令，记录时间

图形用户界面

描述已自动生成

相似地，将大数据集平均分为12份，每两份发送至一个节点，运行相同命令，记录时间。

图形用户界面

中度可信度描述已自动生成

发现real time较之前有了小部分提升。