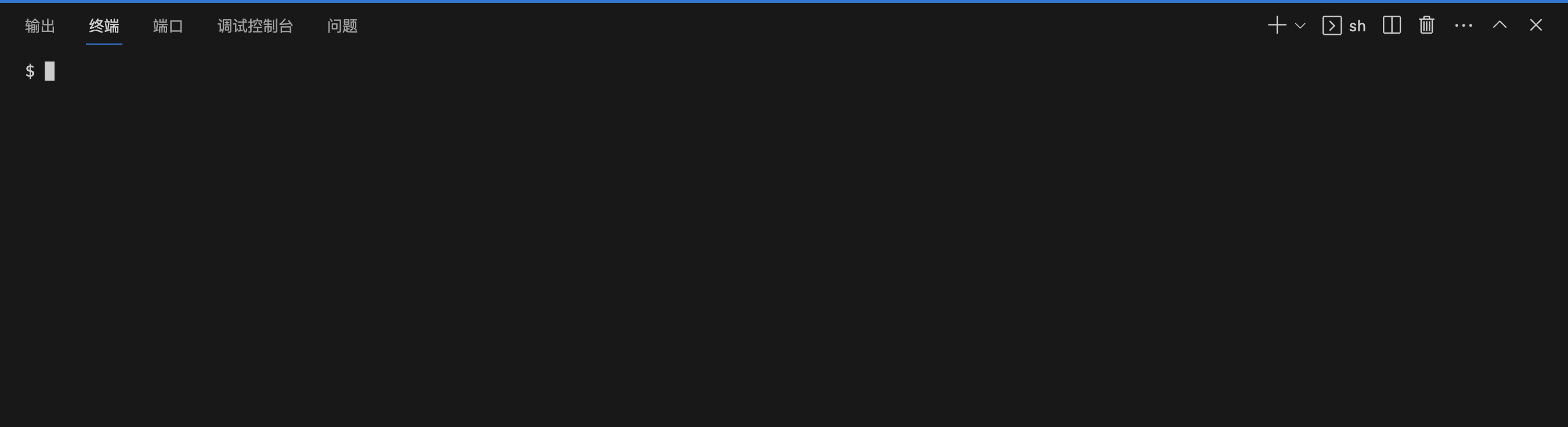
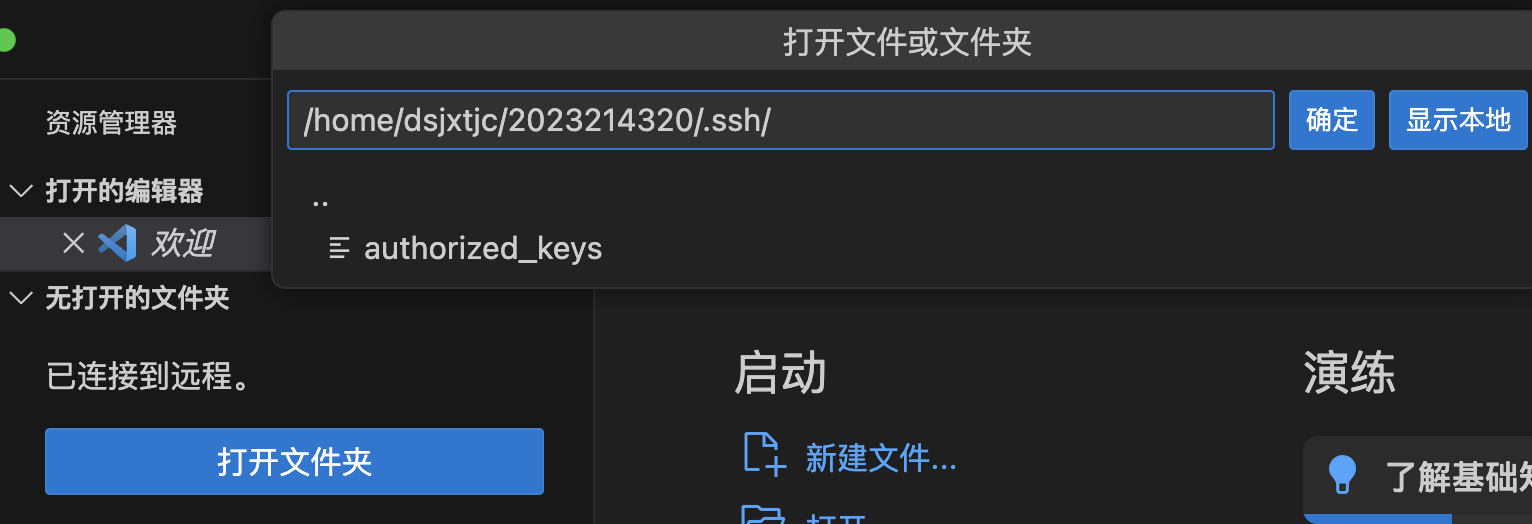
# 三、Linux常用命令

## 任务1. 使用ssh远程登录服务器 （0.5 分）



## 任务2. 配置免密登录 （0.5 分）



## 任务3. pwd、mkdir、cd命令

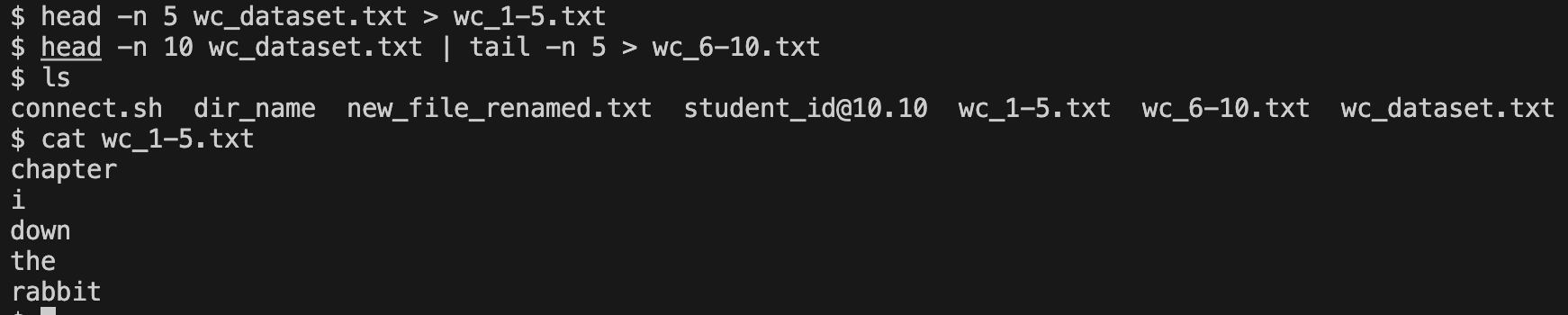
## 任务4. cp、vim、nano、ls、mv、rm命令

## 任务5. cat、head、scp、awk、grep等文本处理命令（0.5 分）

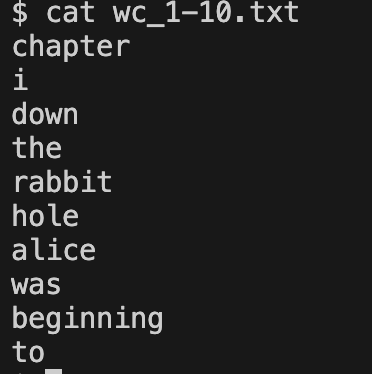
head和tail命令结合



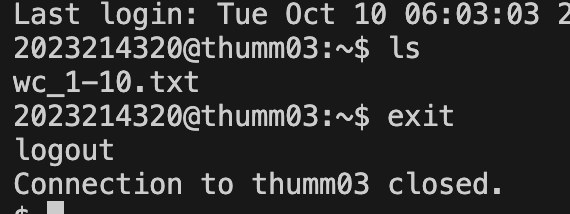
重定向符'>'的使用



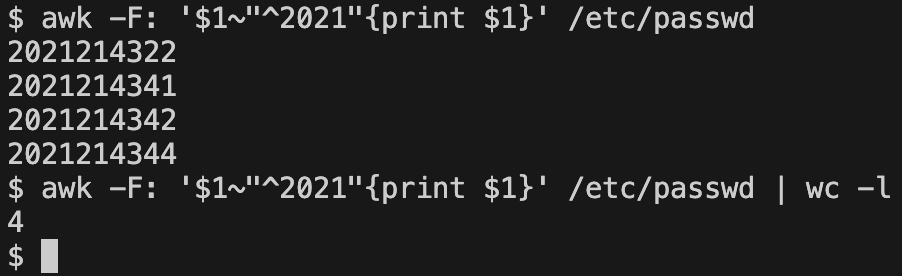
Cat命令的使用



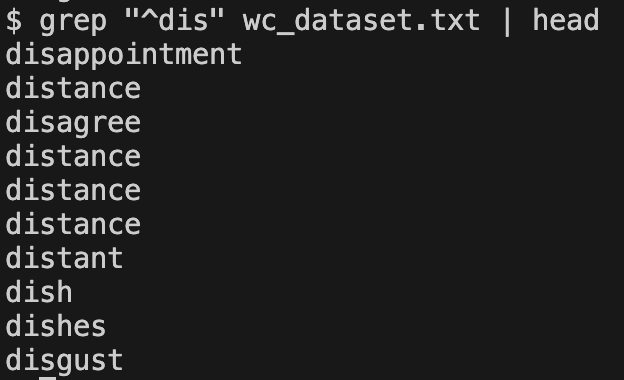
Scp命令的使用



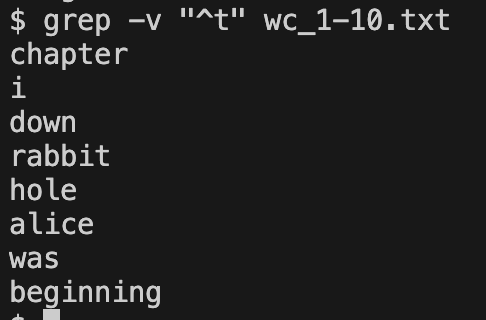
Awk命令的使用



Grep命令的使用



反向过滤添加参数-v



## 任务6.阻塞与非阻塞时间对比 （0.5 分）

# 四、多主机任务处理

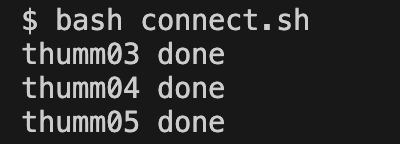
## 任务7. 多节点任务处理

**集群主机之间免密登录配置 （1.5 分）**

代码：



实验结果截图：



## 多结点任务处理（1.5 分）

制作数据集，使用脚本，

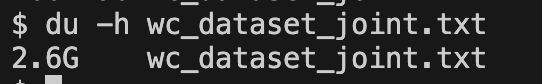
For I in {1…200}

do

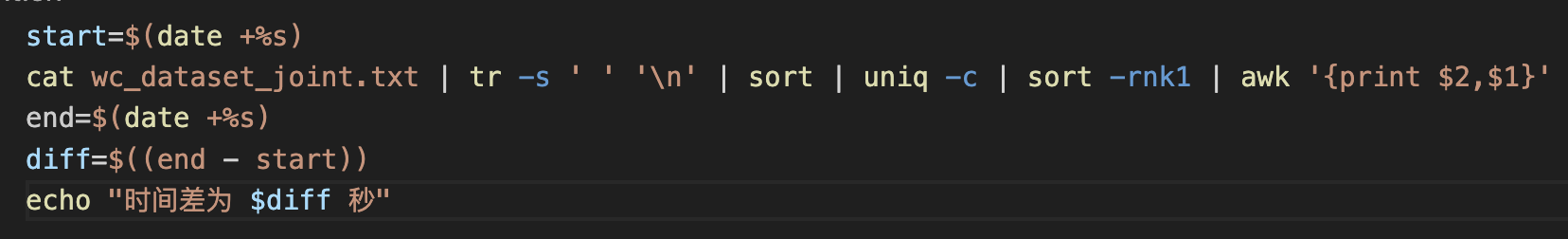
cat wc\_dataset.txt >> wc\_dataset\_joint.txt

done

结果如图显示，得到一个2G的数据集。使用du -h 命令查看文件大小。



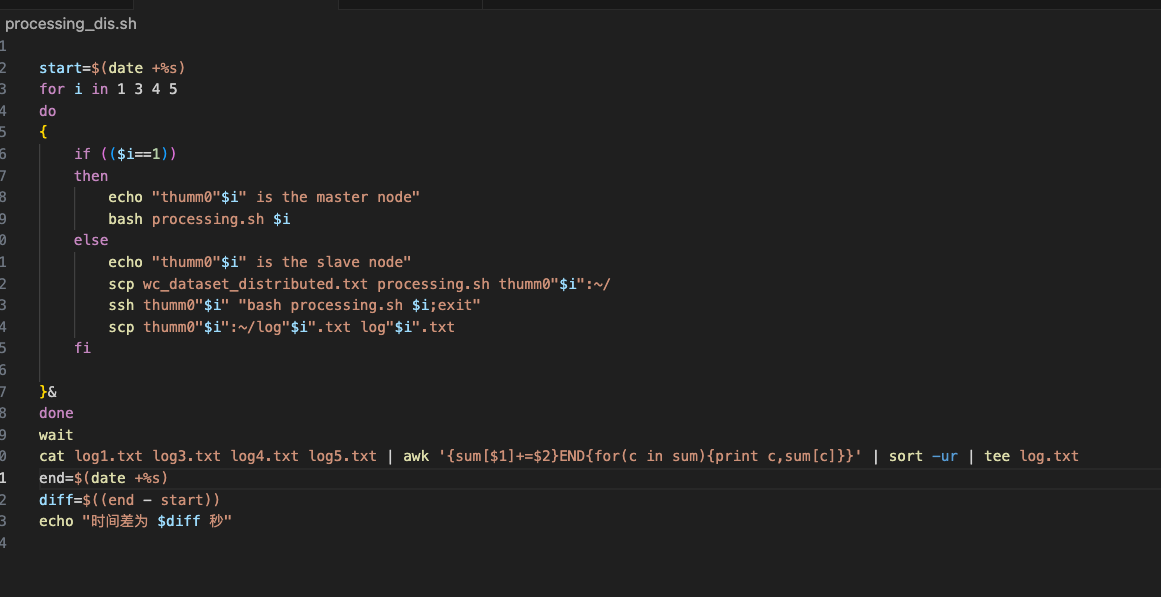
在一台机器上使用检索命令进行统计词频，得到最终结果，shell脚本代码和结果如下，单机上处理的时间是659秒

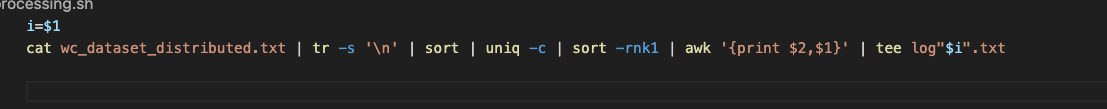




进行并行处理后，得到的时间是161秒。

代码解释：第一个脚本程序主机直接执行processing.sh脚本程序进行处理文本词频统计，从机先将事先处理好的数据集进行分发，并且将脚本程序发送到各个从机上。执行处理脚本，将得到的词频统计文件发到主机上。



Processing.sh处理程序：接受外面传来的对应的服务器参数，对已经分好的数据集进行词频统计。

最后统计程序执行的ddddddddddd总时间：161秒