

华东师范大学数据科学与工程学院实验报告

课程名称：分布式编程模型与系统

年级：2020

上机实践成绩：

指导教师：徐辰

姓名：温兆和

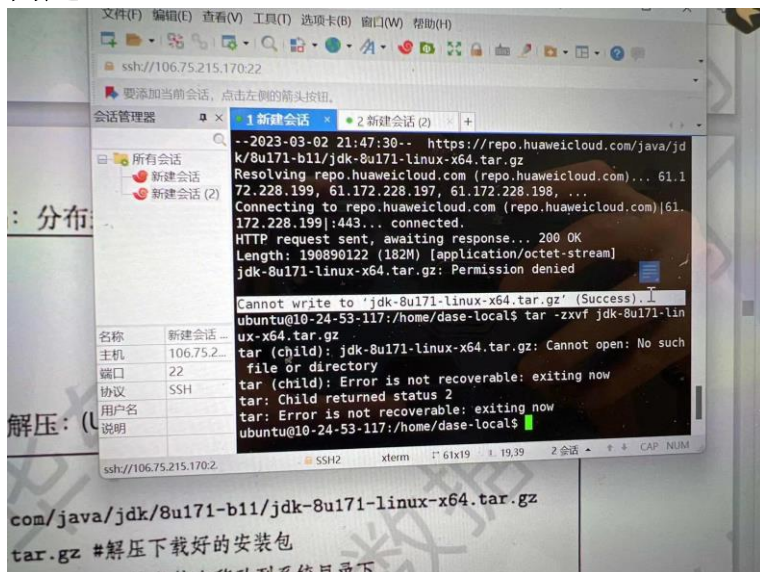
学号：10205501432

上机实践名称：准备工作

上机实践日期：

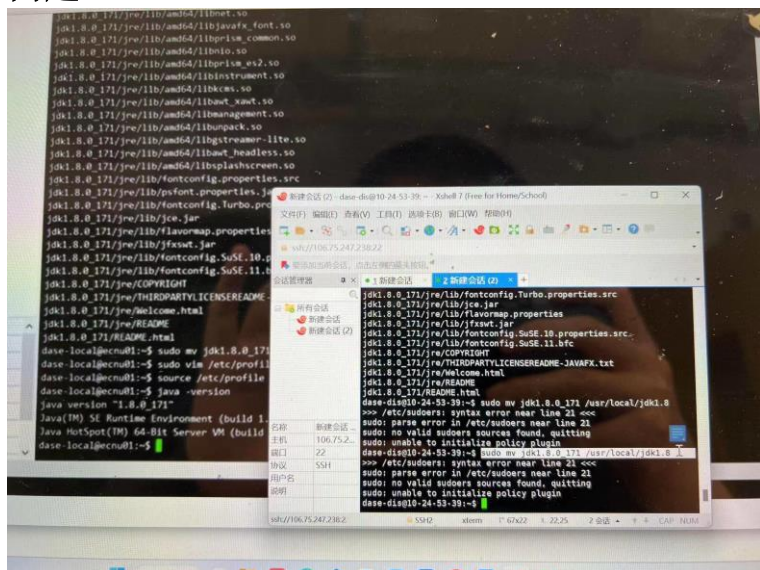
2022.03.02

问题一

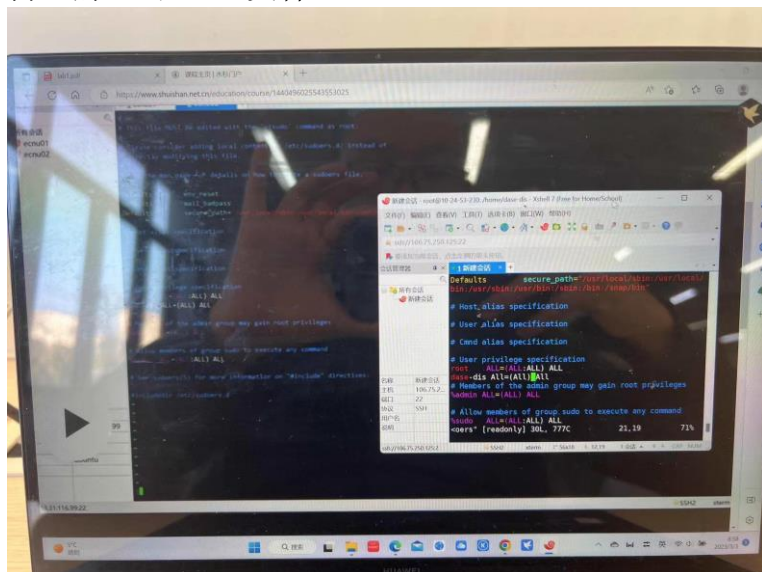


在云主机上配置 JAVA 环境时，我打算进入/home/dase-local 路径解压其中已经配置好的 JAVA 包，但是发现云主机里并没有这个路径。随后助教告诉我这个带有 JAVA 包的路径需要自己通过 sudo useradd 命令创建，并且在我的电脑上，这个路径是/home/dase-dis，而不是/home/dase-local。随后解压 JAVA 包的时候，云主机再一次告诉我当前路径下没有那个 JAVA 包。后来助教告诉我，所有的包都在/software_prepare/路径下，所以我的指令应该是 tar zxvf /software_prepare/jdk1.8，而不是 tar zxvf jdk1.8。

问题二

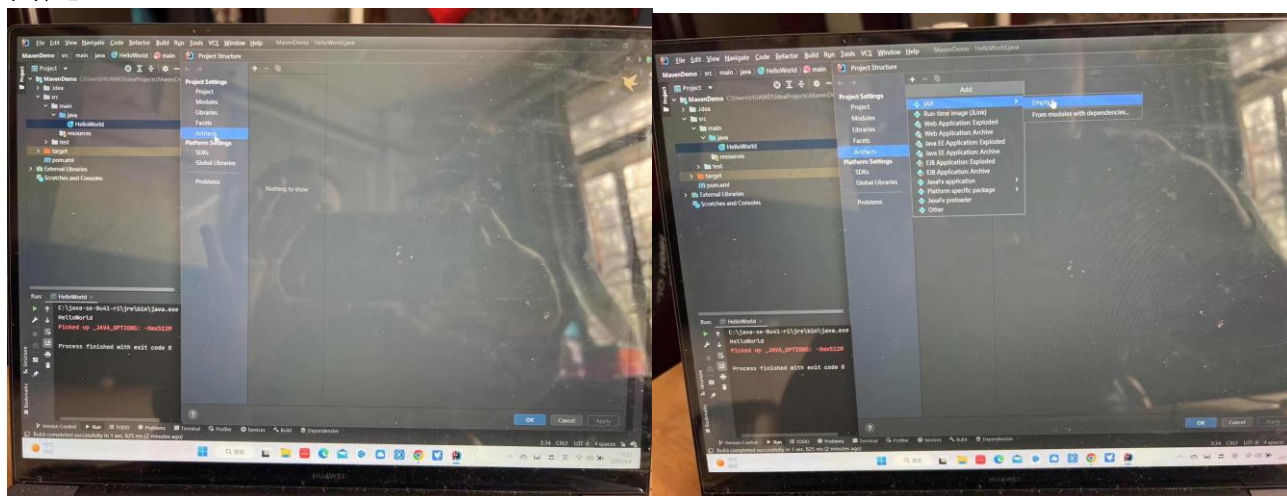


当我想把 jdk 文件夹移动到系统文件夹下的时候，发现系统报错”no valid sudoers source found”。后来助教告诉我，可能是 etc/sudo 这个文件的语法写错了。于是我打开自己的 etc/sudo 文件。



助教看了之后，发现我在修改文件时，把所有”ALL”都打成了”A11”，所以造成了错误。修正这个错误以后，我成功地把 jdk 包移动到了系统目录之下。

问题三



当我在 IDEA 中建立了一个名叫 HelloWorld 的 maven 项目并决定将其打包成 jar 包时，发现 IDEA 的 Project Structure-Artifacts 下什么都没有。助教让我点击左上角的 “+” 号并点击 JAR 按钮，我却发现里面是空的。最后我在助教的指引下点击了”From modules with dependencies”，我才成功地把 maven 项目做成了 jar 包。

但是，新的云平台的网址变成了 <https://edu.ucloud.cn/>

华东师范大学数据科学与工程学院实验报告

课程名称：分布式编程模型与系统

年级：2020

上机实践成绩：

指导教师：徐辰

姓名：温兆和

学号：10205501432

上机实践名称：Hadoop 1.x 部署

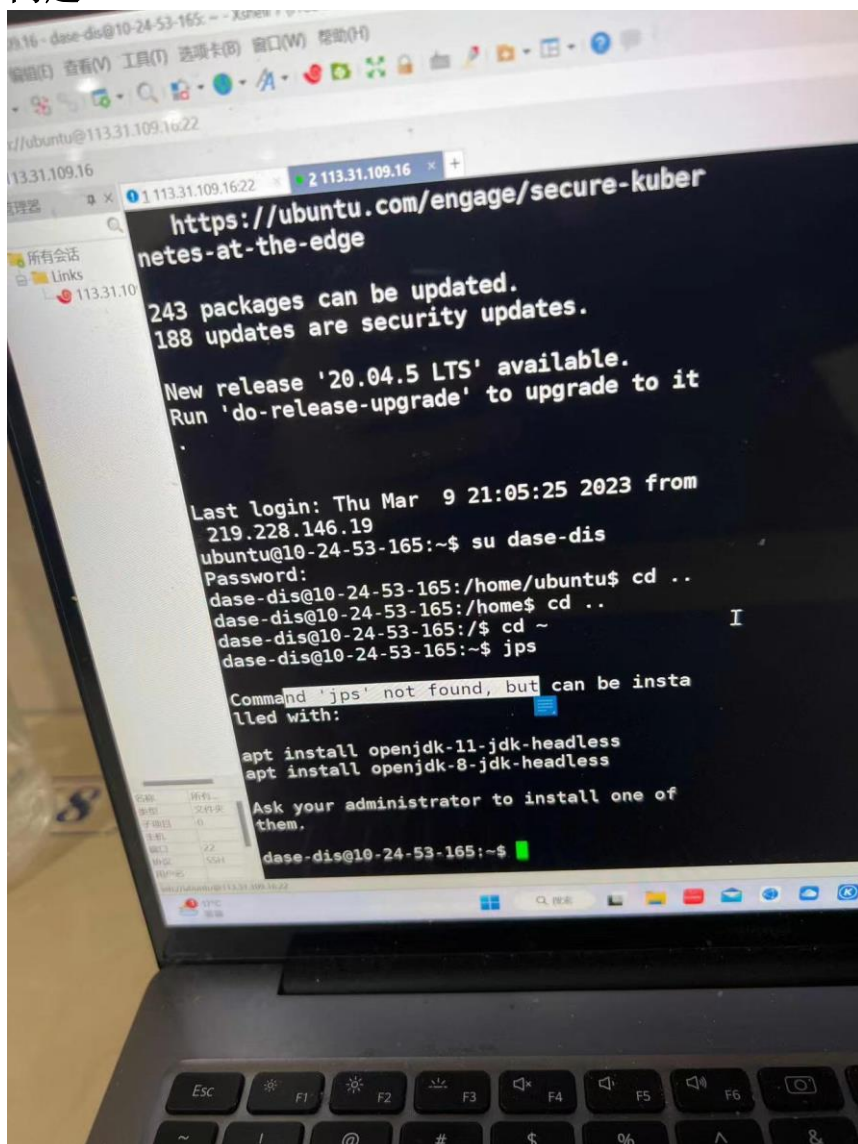
上机实践日期：

2022.03.09

问题一

在进行单机集中式部署实验时，我在运行 wordcount 程序时，未能及时打开另一个终端、使其与云主机建立联系并在其中用 jps 命令查看云主机中的进程运行情况。而当我做好另一个终端中的准备工作并打算再次运行 wordcount 程序时，发现该程序的运行结果已经被保存在本地，导致这个程序无法再被重新运行比较长的一段时间。在询问助教后，助教提示我删除 /output/grep 路径。我在删除这个路径后，wordcount 程序又能重新运行比较长的一段时间，我也就有足够的时间在另一个终端里查看云主机中的进程信息。

问题二



在解决问题一之后，我一度发现第二个终端里没有 jps 命令。经过助教的提示，我首先输入 source /etc/profile 命令后再输入 jps 命令，就成功地看到了云主机中的进程信息。

经过自己的努力和助教的帮助，我顺利完成了实验二中的后面两个实验。截图为证。

```

113.31.109.16 - dase-dis@10-24-53-165: ~/hadoop-1.2.1 - Xshell 7 (Free for Home/School)
ssh://ubuntu@113.31.109.16:22
113.31.109.16
23/03/11 11:35:19 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
23/03/11 11:35:19 INFO mapred.JobClient: Data-local map tasks=33
23/03/11 11:35:19 INFO mapred.JobClient: File Output Format Counters
23/03/11 11:35:19 INFO mapred.JobClient: Bytes Written=2996713
dase-dis@10-24-53-165:~/hadoop-1.2.1$ cd userlogs
bash: cd: userlogs: No such file or directory
dase-dis@10-24-53-165:~/hadoop-1.2.1$ cd logs
dase-dis@10-24-53-165:~/hadoop-1.2.1/logs$ cd userlogs
dase-dis@10-24-53-165:~/hadoop-1.2.1/logs/userlogs$ ls
job_202303111114_0001  job_202303111114_0002  job_202303111114_0003
dase-dis@10-24-53-165:~/hadoop-1.2.1/logs/userlogs$ cd job_202303111114_0003
dase-dis@10-24-53-165:~/hadoop-1.2.1/logs/userlogs/job_202303111114_0003$ ls
attempt_202303111114_0003_m_000000_0 attempt_202303111114_0003_m_000013_0 attempt_202303111114_0003_m_000026_0
attempt_202303111114_0003_m_000001_0 attempt_202303111114_0003_m_000014_0 attempt_202303111114_0003_m_000027_0
attempt_202303111114_0003_m_000002_0 attempt_202303111114_0003_m_000015_0 attempt_202303111114_0003_m_000028_0
attempt_202303111114_0003_m_000003_0 attempt_202303111114_0003_m_000016_0 attempt_202303111114_0003_m_000029_0
attempt_202303111114_0003_m_000004_0 attempt_202303111114_0003_m_000017_0 attempt_202303111114_0003_m_000030_0
attempt_202303111114_0003_m_000005_0 attempt_202303111114_0003_m_000018_0 attempt_202303111114_0003_m_000031_0
attempt_202303111114_0003_m_000006_0 attempt_202303111114_0003_m_000019_0 attempt_202303111114_0003_m_000032_0
attempt_202303111114_0003_m_000007_0 attempt_202303111114_0003_m_000020_0 attempt_202303111114_0003_m_000033_0
attempt_202303111114_0003_m_000008_0 attempt_202303111114_0003_m_000021_0 attempt_202303111114_0003_m_000034_0
attempt_202303111114_0003_m_000009_0 attempt_202303111114_0003_m_000022_0 attempt_202303111114_0003_r_000000_0
attempt_202303111114_0003_m_000010_0 attempt_202303111114_0003_m_000023_0 job-acls.xml
attempt_202303111114_0003_m_000011_0 attempt_202303111114_0003_m_000024_0
attempt_202303111114_0003_m_000012_0 attempt_202303111114_0003_m_000025_0
dase-dis@10-24-53-165:~/hadoop-1.2.1/logs/userlogs/job_202303111114_0003$ cd ~/hadoop-1.2.1
dase-dis@10-24-53-165:~/hadoop-1.2.1$ ./bin/stop-all.sh
stopping jobtracker
dase-dis@localhost's password:
localhost: stopping tasktracker
stopping namenode
dase-dis@localhost's password:
localhost: stopping datanode
dase-dis@localhost's password:
localhost: stopping secondarynamenode
dase-dis@10-24-53-165:~/hadoop-1.2.1$ jps
12189 Jps
dase-dis@10-24-53-165:~/hadoop-1.2.1$

```

华东师范大学数据科学与工程学院实验报告

课程名称：分布式编程模型与系统

年级：2020

上机实践成绩：

指导教师：徐辰

姓名：温兆和

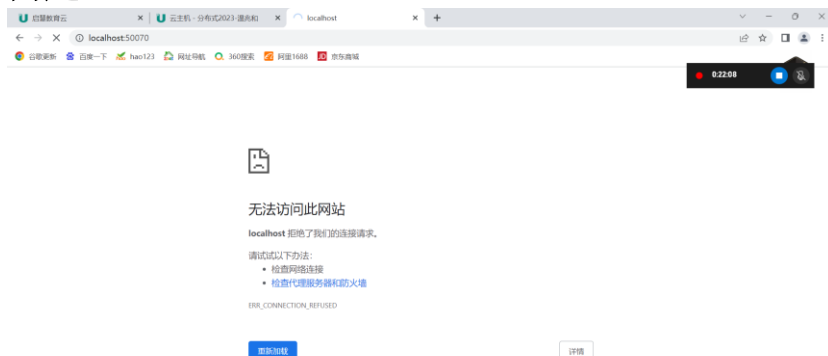
学号：10205501432

上机实践名称：Hadoop 2.x 部署

上机实践日期：

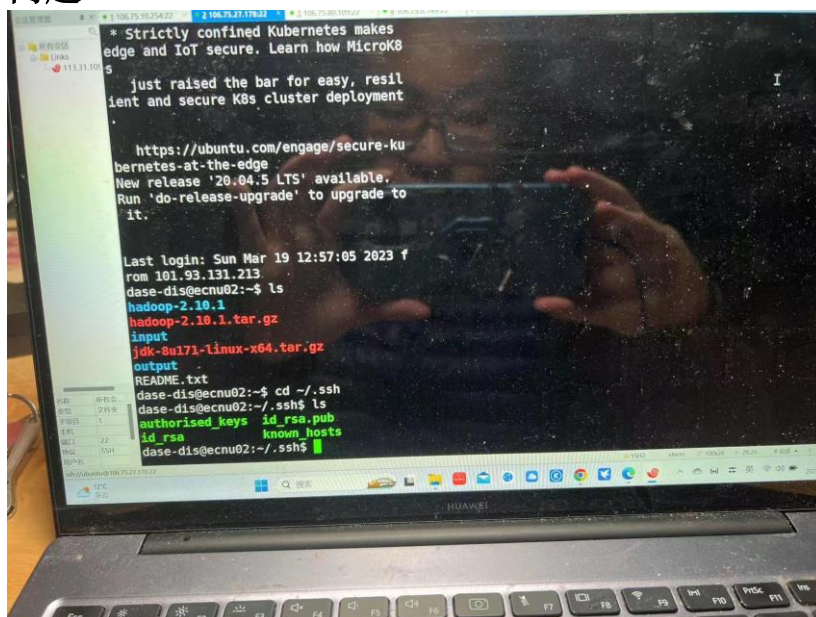
2022.03.16

问题一



我在进行 HDFS 单机伪分布式部署实验时，想要在本机查看节点内部情况。我在浏览器输入 localhost:50070 后，发现无法访问网站。在观看实验指导视频后，我发现自己忘了把端口 50070 映射到本地。我在本地操作系统的 shell 中输入 localhost:50070:localhost:50070 [dase-dis@106.75.27.178](#) 命令后再次访问节点，就看到了节点内部的情况。

问题二

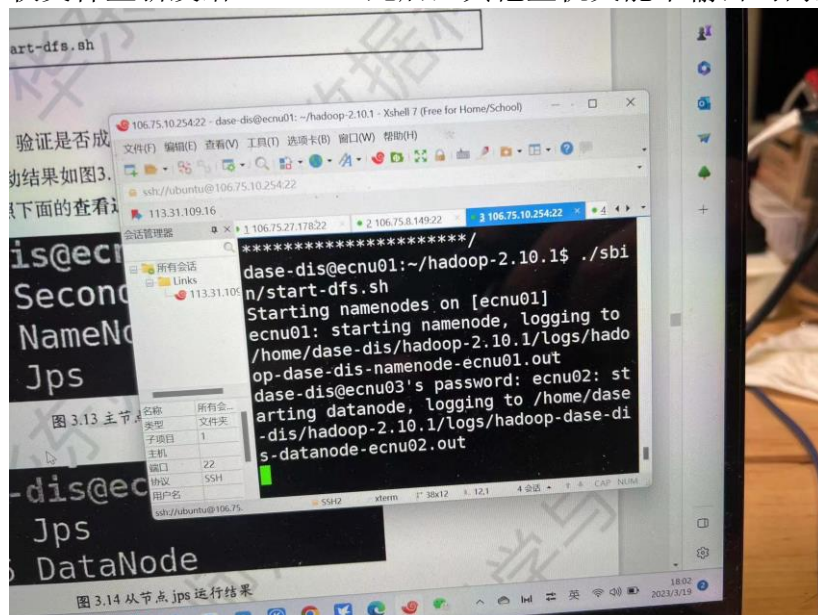


在进行 HDFS 分布式部署实验时，我需要实现四台主机单机和它们互相之间的免密钥登录。在执行完所有操作之后，我发现所有主机在登录时还是需要输入密码。在将自己的操作过程

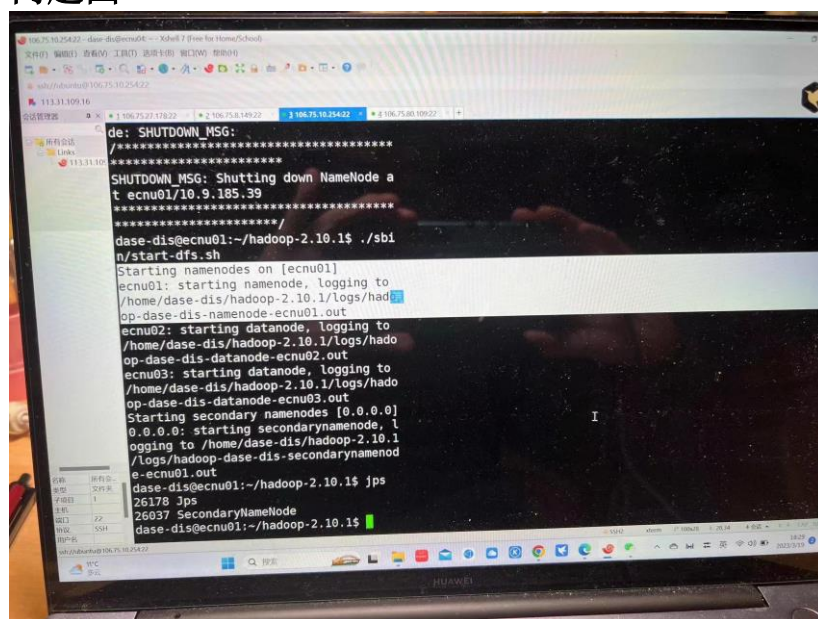
录屏并发给助教后，助教告诉我错误的原因是 `authorized` 被错打成 `authorised` 了。

问题三

在进行 HDFS 分布式部署实验时，我将主机 `ecnu03` 的本地连接关闭后重新连上，但发现其他主机无法免密钥登录 `ecnu03`，但另外三个主机互相之间还是可以免密钥登录。我六神无主，然后就询问了助教。后来助教让我把 `ecnu03` 上的公钥重新发给 `ecnu01`，`ecnu01` 上的授权文件重新发给 `ecnu03`。此后，其他主机又能不输密码而直接连上 `ecnu03` 了。

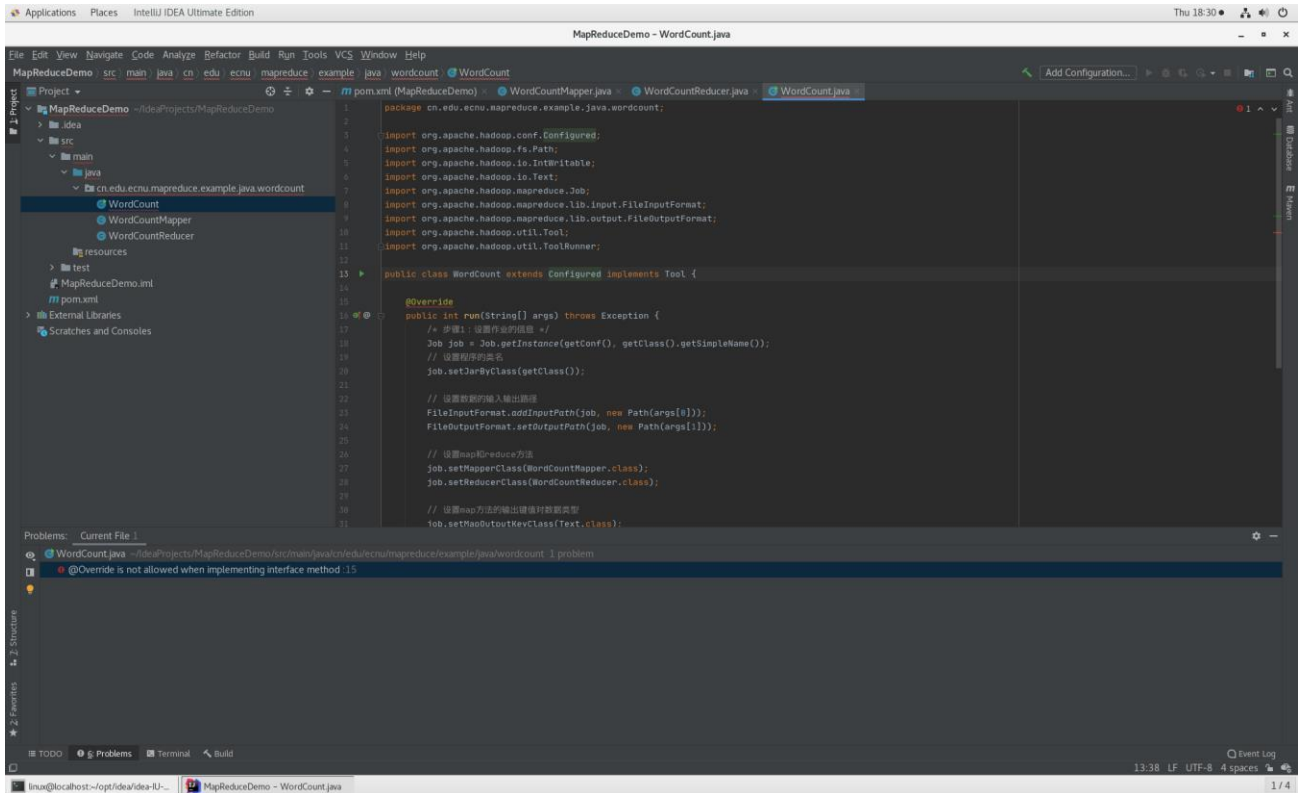


问题四

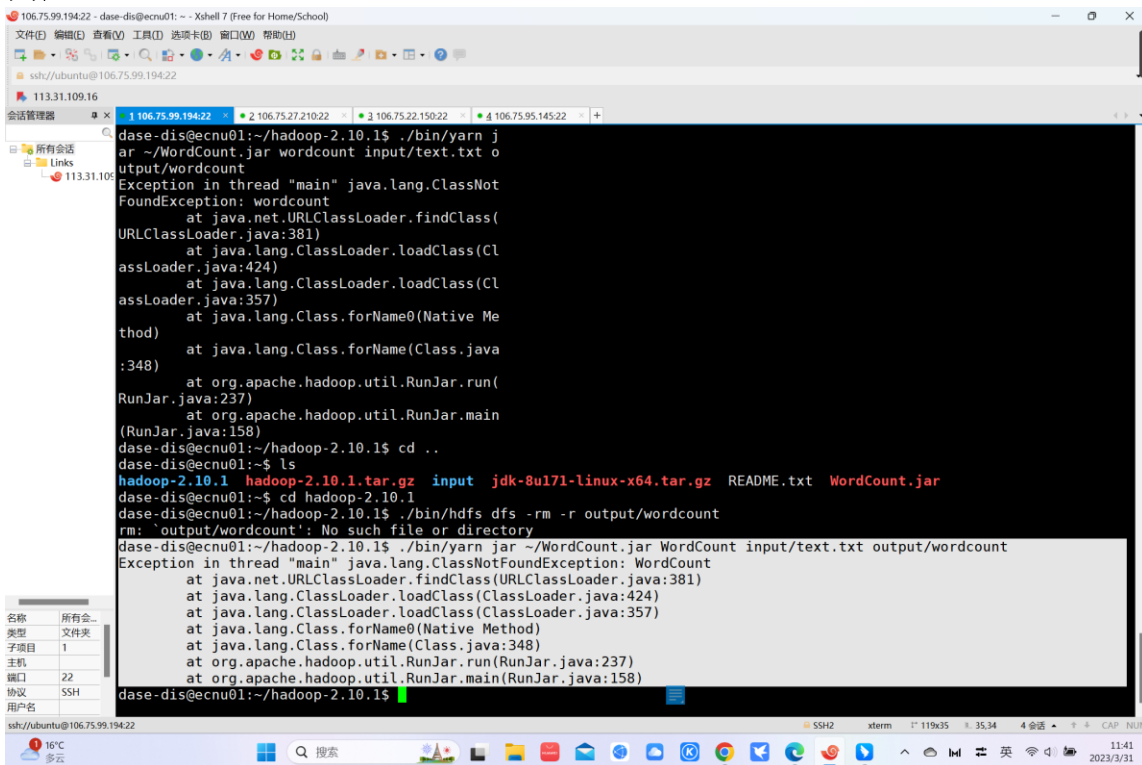


在进行 HDFS 分布式部署实验时，我不小心多次格式化了 `namenode`，导致我在启动 `hdfs` 服务后，`ecnu01` 中没有 `namenode` 进程。我当时已经冥冥之中意识到了问题的原因，但在询问助教和搜索 `csdn` 博客后，我的问题解决了。我本想破罐子破摔，毕竟还有 `secondarynamenode`，但在进行 MapReduce 分布式部署实验时，我发现如果没有 `namenode`，这个实验就做不下去了。于是我又询问了另一个助教。他在查看了 `namenode.log` 日志后发现 `tmp` 文件夹已经损坏，需要杀死所有进程，把四台主机上的 `tmp` 文件全部删除并重新格式化 `namenode`。我在这样做并重启了 `hdfs` 服务，发现 `ecnu01` 的进程列表里可以看见 `namenode` 了。

2022.03.30



问题三



最后在云主机上我还遇到了 jar 包里没有 WordCount 类的报错。孙助教告诉我，要把类的名称写全，即把 package 的名称放在类的前面。于是，我再次输入命令行，把类的名称从 WordCount 改成了 cn.edu.ecnu.mapreduce.example.java.wordcount.WordCount，本地调试好的程序就能顺利地云主机上运行了。