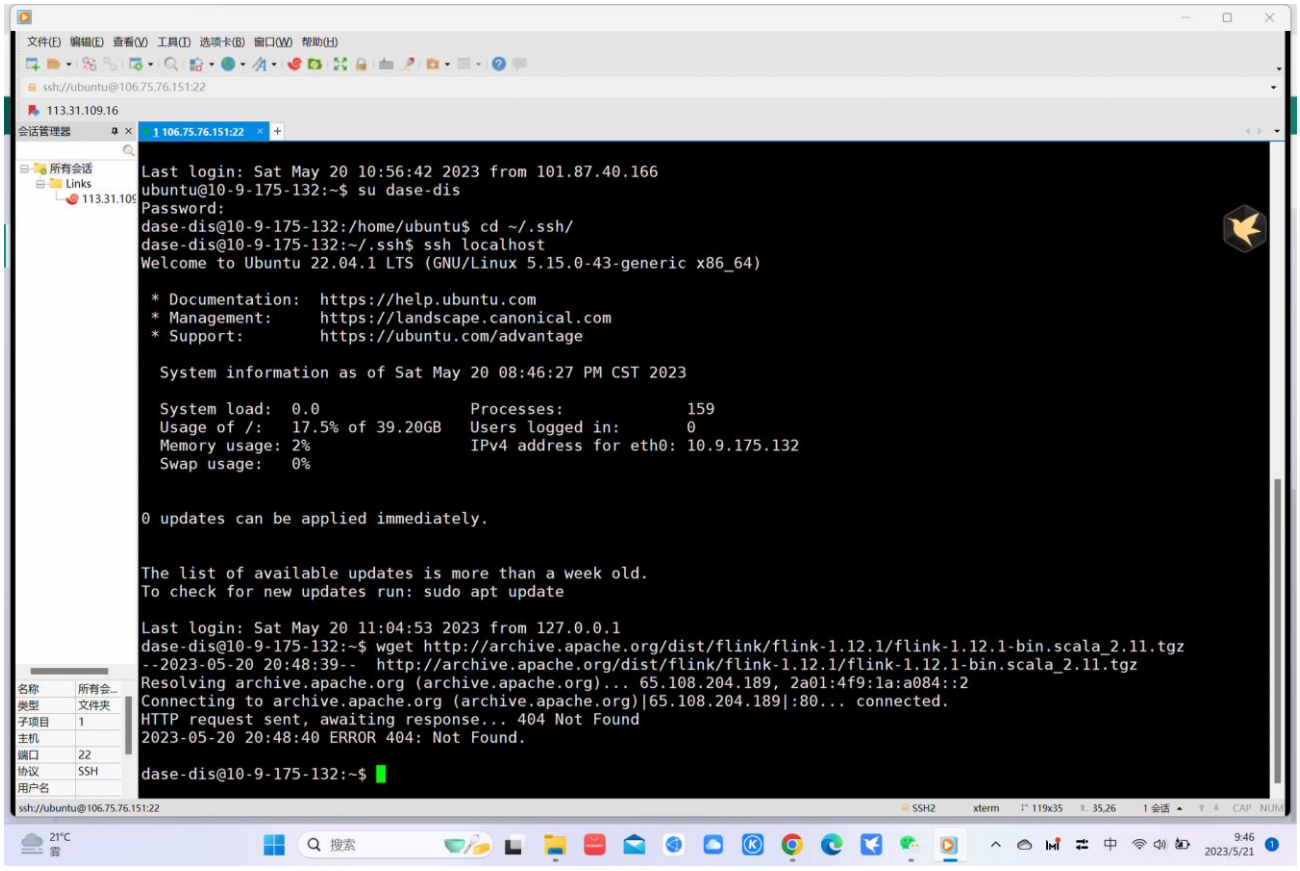
华东师范大学数据科学与工程学院实验报告

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：Flink 部署　　**上机实践日期：**　　2022.05.18

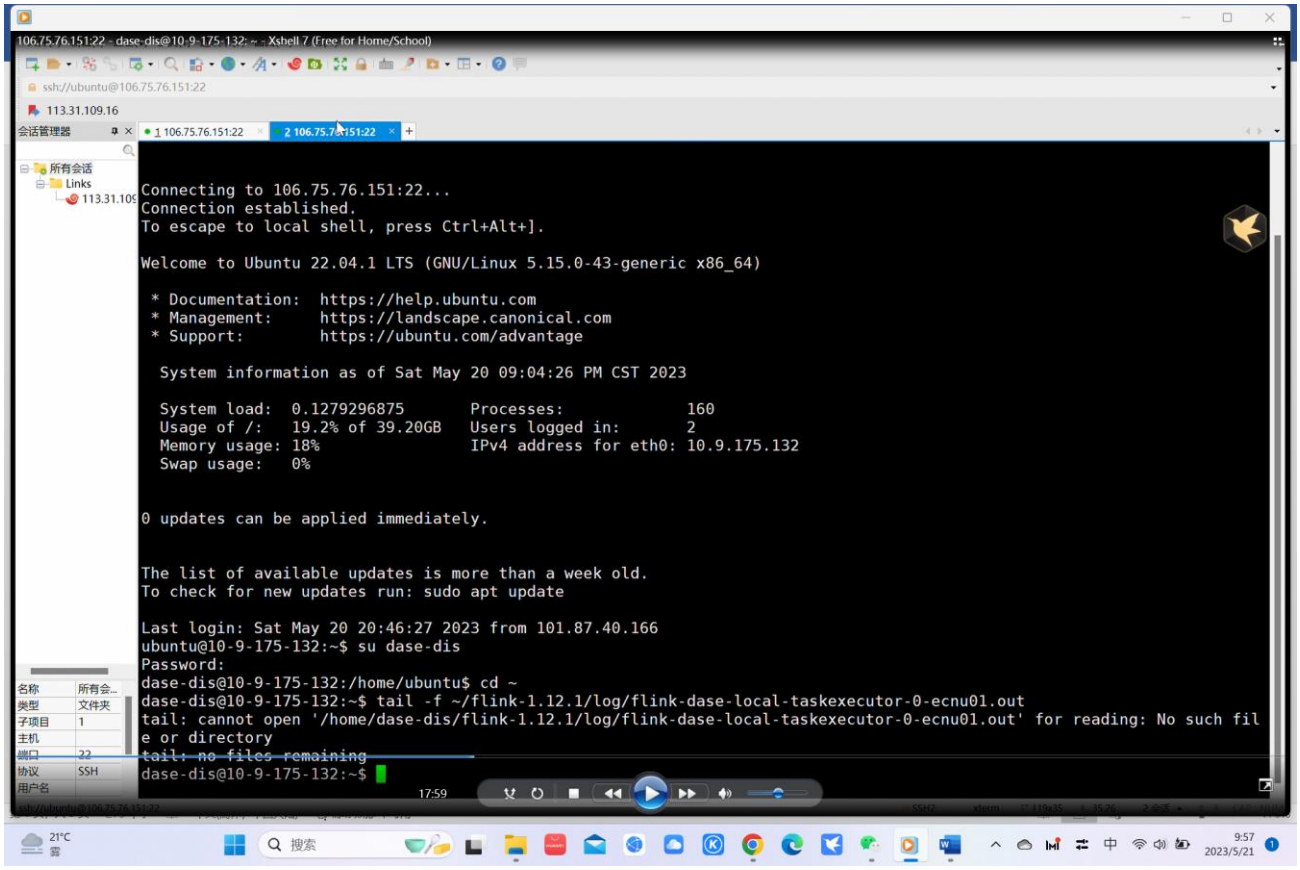
问题一



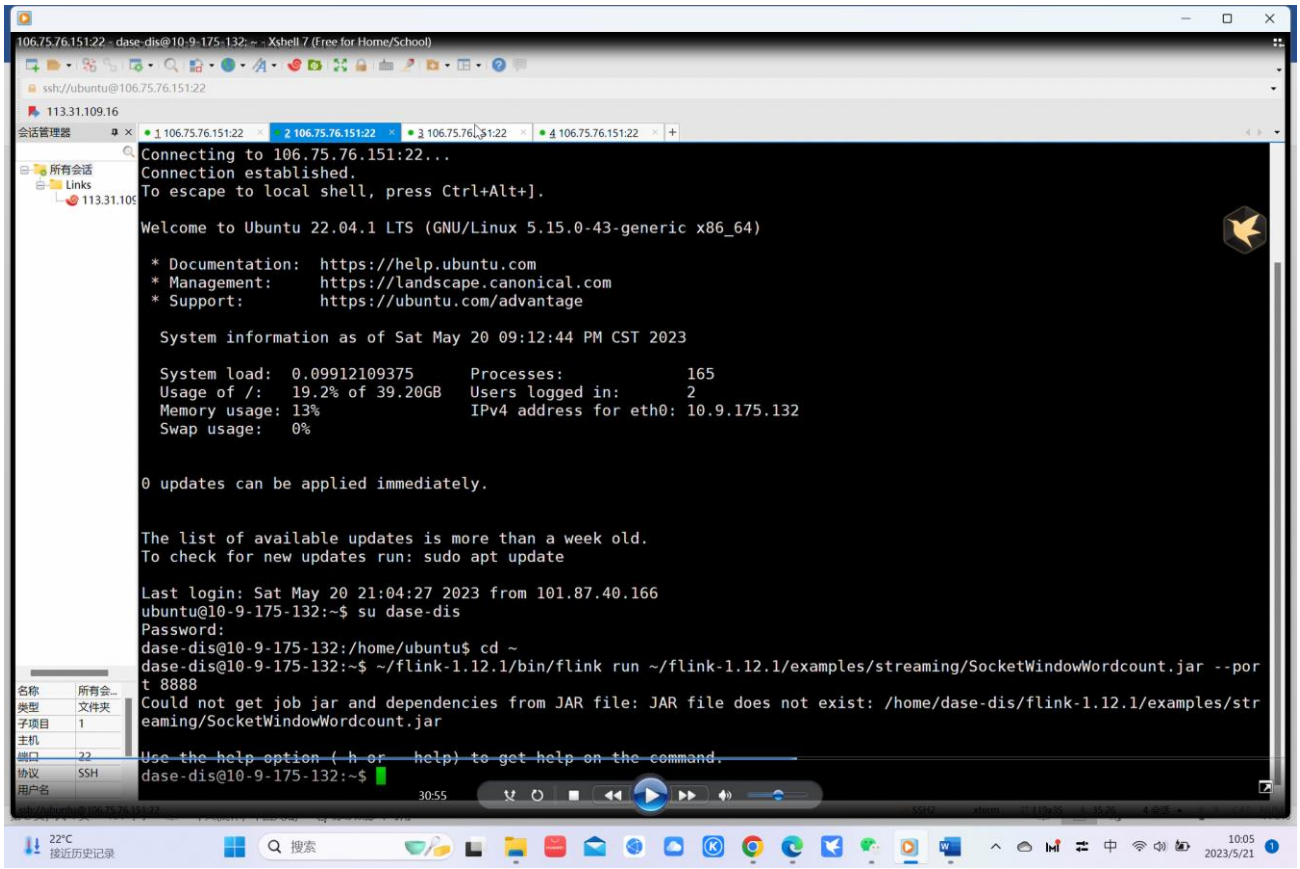
在实验的一开始就没能顺利下载 Flink 文件。原因：路径输错。

问题二

在进行 Flink 单机伪分布式部署实验时，我在 flink shell 中输入了命令，并在另一个终端中输入 tail 命令准备获取 flink 执行程序的输出，但发现输出文件不存在。于是，我将工作路径移动到输出文件所在的路径下并输入 ls，才发现不同主机上输出文件的名称是不同的，而我却直接把实验手册上的文件名称放在了 tail 语句的后面。于是，我重新进行实验并把 tail 命令中的输出文件名称换成自己的云主机上的输出文件名称，最终看到你了输出。



问题三



单机提交 JAR 包运行时，系统报错说没有这个 JAR 包。原因：路径中 JAR 包名称输错。

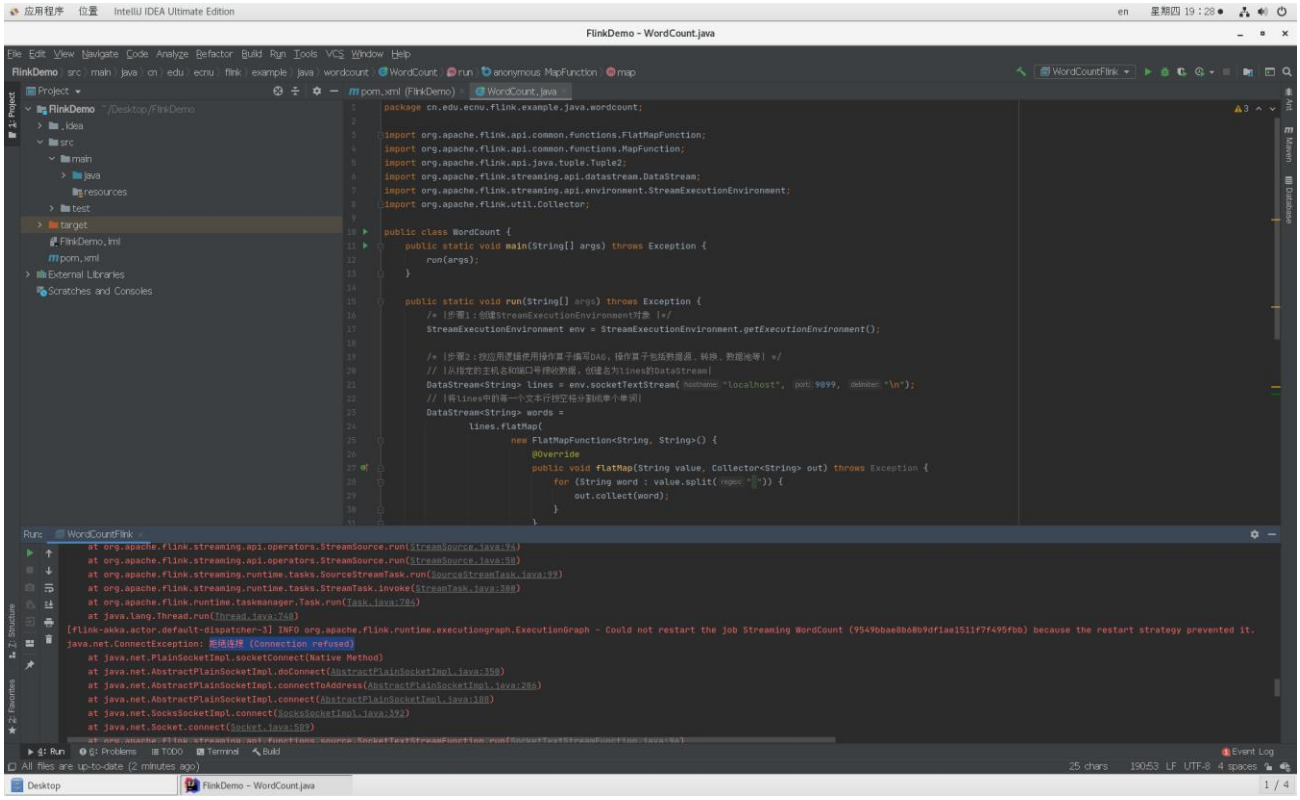
华东师范大学数据科学与工程学院实验报告

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：Flink 编程　　**上机实践日期：**2022.06.01

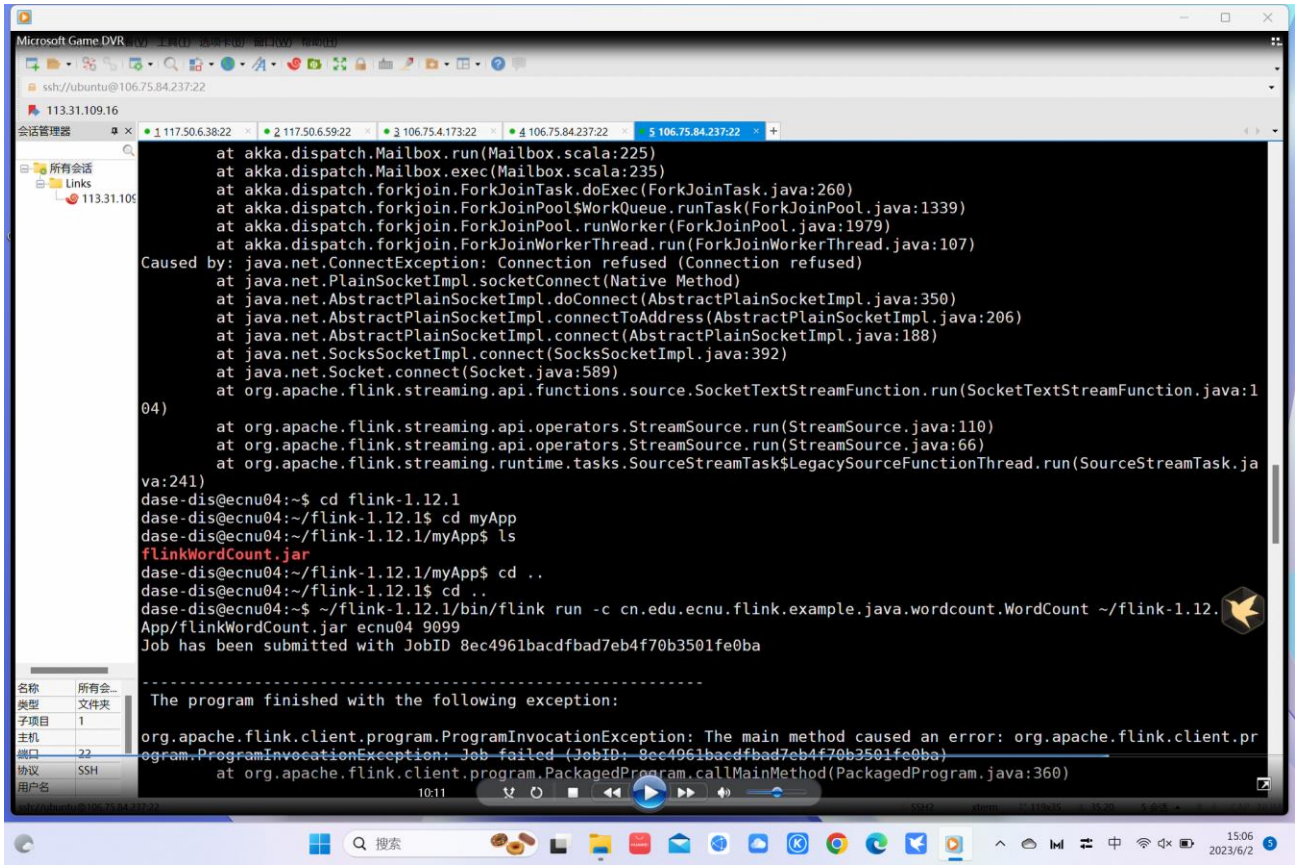
问题一



在本地调试 flink 的 wordcount 代码时，我在本地 shell 输入 nc -lk 8888 命令准备监听 8888 端口，准备在 8888 端口输入数据流并查看程序的输出结果，但出现了“拒接连接”的报错。助教在查看了我的代码后，发现这段代码在读取 DataStream 时，限定从 9099 端口读取输入。于是，我把我把程序配置中的 Program Arguments 改成了 localhost 9099，在本地 shell 中监听 9099 端口，运行 WordCount 并在本地 shell 输入数据流，最终看到了词频统计的结果。

问题二

接着，我把原先的 Maven 项目打包，传入云主机并决定在分布式环境下运行这个 maven 项目，但每次开始运行后就会立即出现大段的报错。助教告诉我，代码中限制从 9099 端口读取输入的数据流，但分布式情况下，每台主机的 9099 端口并不是一样的。于是，我打开虚拟机，把读取数据流的那一行代码从 DataStream<String> lines = env.socketTextStream("localhost", 9099, "\n");改成 DataStream<String> lines = env.socketTextStream(args[0], Integer.parseInt(args[1]), "\n");，重新将项目打包并传入云主机进行分布式环境下的运行，最终看到了正确的结果。


Microsoft Game DVR
ssh://ubuntu@106.75.84.237:22
113.31.109.16
会话管理器
1 117.50.6.38:22
2 117.50.6.59:22
3 106.75.4.173:22
4 106.75.84.237:22
5 106.75.84.237:22
所有会话
Links
at akka.dispatch.Mailbox.run(Mailbox.scala:225)
at akka.dispatch.Mailbox.exec(Mailbox.scala:235)
at akka.dispatch.forkjoin.ForkJoinTask.doExec(ForkJoinTask.java:260)
at akka.dispatch.forkjoin.ForkJoinPool$WorkQueue.runTask(ForkJoinPool.java:1339)
at akka.dispatch.forkjoin.ForkJoinPool.runWorker(ForkJoinPool.java:1979)
at akka.dispatch.forkjoin.ForkJoinWorkerThread.run(ForkJoinWorkerThread.java:107)
Caused by: java.net.ConnectException: Connection refused (Connection refused)
at java.net.PlainSocketImpl.socketConnect(Native Method)
at java.net.AbstractPlainSocketImpl.doConnect(AbstractPlainSocketImpl.java:350)
at java.net.AbstractPlainSocketImpl.connectToAddress(AbstractPlainSocketImpl.java:206)
at java.net.AbstractPlainSocketImpl.connect(AbstractPlainSocketImpl.java:188)
at java.net.SocksSocketImpl.connect(SocksSocketImpl.java:392)
at java.net.Socket.connect(Socket.java:589)
at org.apache.flink.streaming.api.functions.source.SocketTextStreamFunction.run(SocketTextStreamFunction.java:104)
at org.apache.flink.streaming.api.operators.StreamSource.run(StreamSource.java:110)
at org.apache.flink.streaming.api.operators.StreamSource.run(StreamSource.java:66)
at org.apache.flink.streaming.runtime.tasks.SourceStreamTask$LegacySourceFunctionThread.run(SourceStreamTask.java:241)
dase-dis@ecnu04:~$ cd flink-1.12.1
dase-dis@ecnu04:~/flink-1.12.1$ cd myApp
dase-dis@ecnu04:~/flink-1.12.1/myApp$ ls
flinkWordCount.jar
dase-dis@ecnu04:~/flink-1.12.1/myApp$ cd ..
dase-dis@ecnu04:~/flink-1.12.1$ cd ..
dase-dis@ecnu04:~$ ~/flink-1.12.1/bin/flink run -c cn.edu.ecnu.flink.example.java.wordcount.WordCount ~/flink-1.12.1/myApp/flinkWordCount.jar ecnu04 9099
Job has been submitted with JobID 8ec4961bacdfbad7eb4f70b3501fe0ba
The program finished with the following exception:
org.apache.flink.client.program.ProgramInvocationException: The main method caused an error: org.apache.flink.client.program.ProgramInvocationException: Job failed (JobID: 8ec4961bacdfbad7eb4f70b3501fe0ba)
at org.apache.flink.client.program.PackagedProgram.callMainMethod(PackagedProgram.java:360)
名称 所有会...
类型 文件夹
子项目 1
主机
端口 22
协议 SSH
用户名
10:11
搜索
15:06
2023/6/2

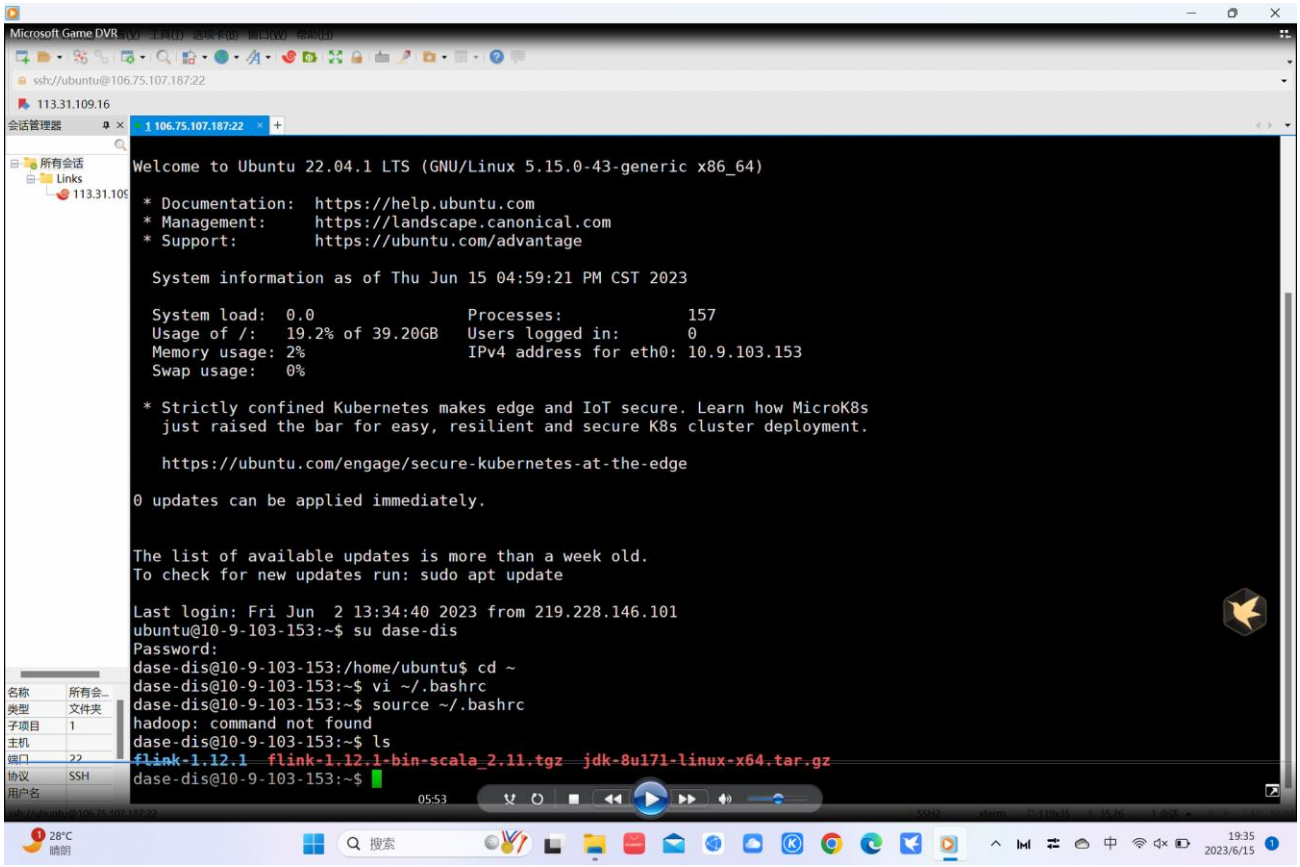
华东师范大学数据科学与工程学院实验报告

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：基于 Yarn 部署 Flink　　**上机实践日期：**2022.06.15

问题一

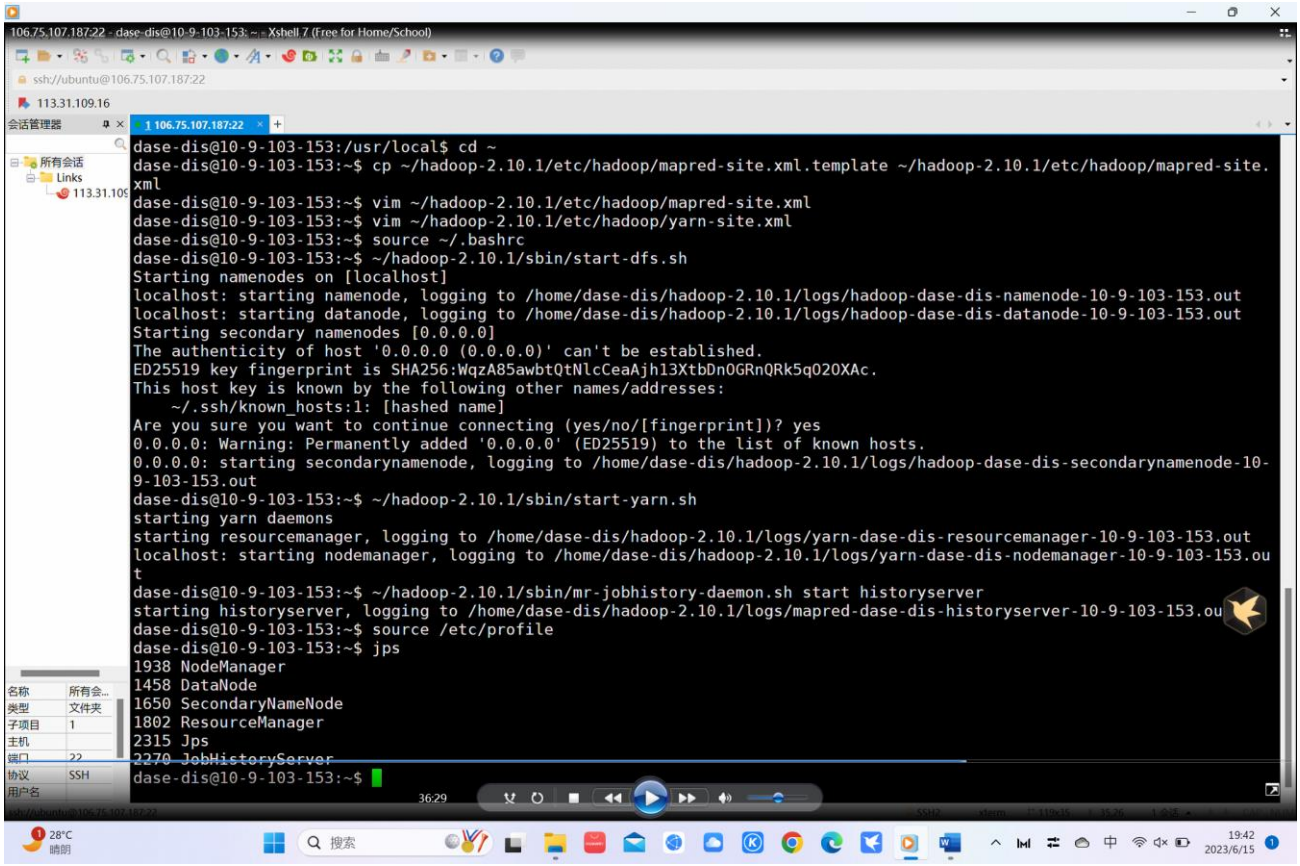


在进行单机伪分布式部署实验时，我修改 bashrc 文件后想要让它生效，结果发现原来做 flink 伪分布式实验的这台主机没有装 Hadoop。于是，我在这台主机里安装了 Hadoop 并使 bashrc 文件生效。

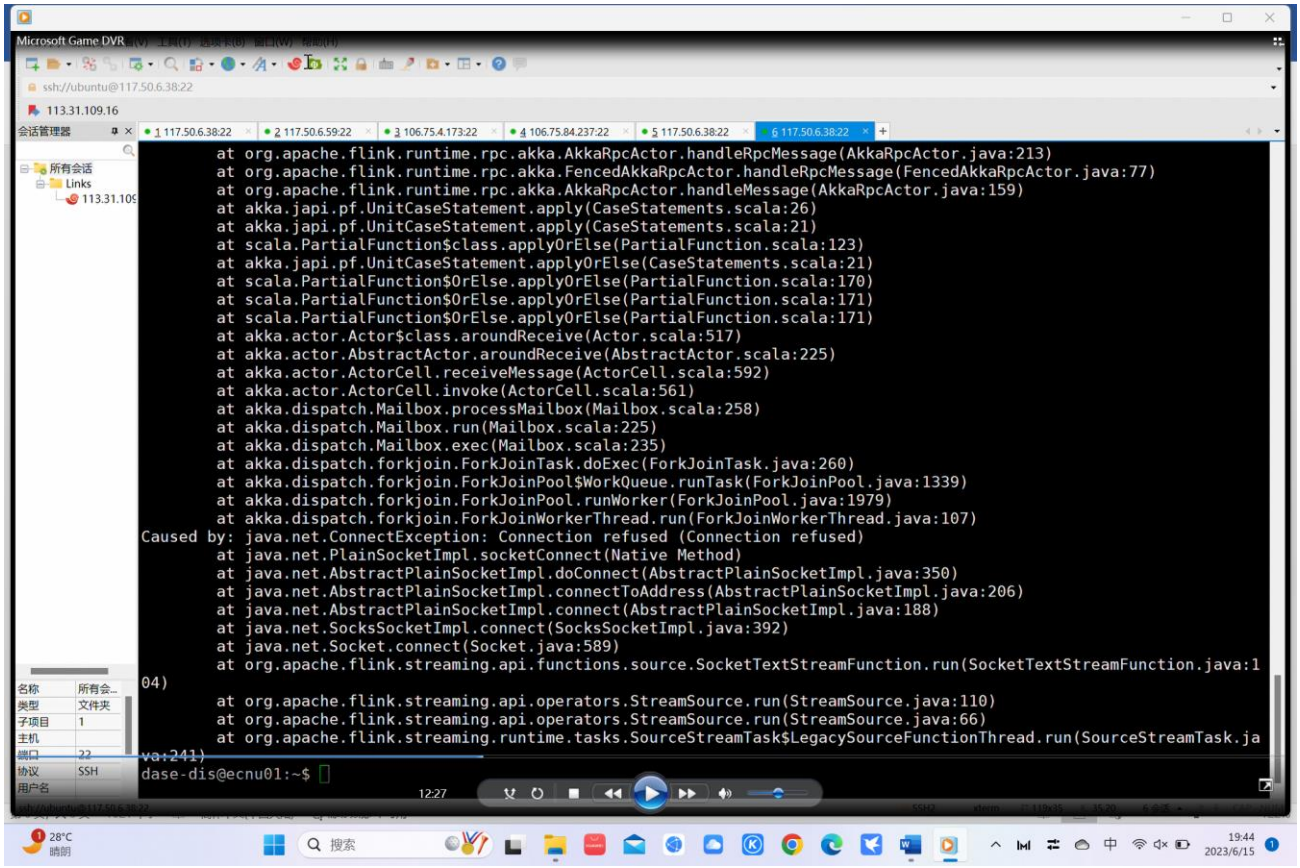
但紧接着，我发现启动 HDFS 后没有 Namenode 进程。于是，我检查了自己装配 Hadoop 的过程，发现自己不仅在配置文件里打错了好几个地方，还忘记初始化 Namenode。最后，我重新修改了 Hadoop 依赖文件并初始化了 Namenode，最终成功运行了 wordcount 程序。

问题二

在进行分布式部署实验时，我提交了 JAR 包，发现跑不起来。最后，我偶然发现我是在主节点 ecnu01 上监听了 8888 端口并提交了 JAR 包，而不是客户端 ecnu04。最后，我重新打开了两个连接到 ecnu04 上的 shell，在上面重新监听了 8888 端口并提交了 JAR 包，最终成功地运行了程序。



伪分布式实验图



分布式实验图