

## 华东师范大学数据科学与工程学院实验报告

课程名称：分布式计算系统

年级：2022

上机实践成绩：

指导教师：徐辰

姓名：郭夏辉

学号：10211900416

上机实践名称：Flink 部署

上机实践日期：

2024.6.4

### 问题 1：配置 flink-1.12.1/conf/flink-conf.yaml 之后 start-cluster 出现问题

经过一段时间的检查，我发现自己忽略了实验指导文档中的一句话：

**注意：**配置 key/value 时候在 “:” 后面需要有一个空格，否则配置不会生效。  
此外，不要将注释与配置项写在同一行。

按照格式要求重新修改 flink-conf.yaml 之后，我再去 start-cluster 就好了。

### 问题 2：下载 Flink 时候报错 Error 404:Not Found

本来我还以为是网络问题，但检查一番后我发现自己的路径打错了，正确的应该是 `sudo wget https://archive.apache.org/dist/flink/flink-1.12.1/flink-1.12.1-bin-scala_2.11.tgz` 但是我输入的确是：`sudo wget https://archive.apache.org/dist/flink/flink-1.12.1/flink-1.12.1-bin.scala_2.11.tgz` 就一个小小的符号打错了就路径错误了。

### 问题 3：attached 方式最初结果不对

结果不对：

```
dase-local@10-23-192-237:~$ nc -lk 8888
hello ecnu
hello dase
hello ecnu

dase-local@10-23-192-237:~$ flink-1.12.1/bin/flink run flink-1.12.1/examples/streaming/SocketWindowWordCount
.jar --port 8888
Job has been submitted with JobID b383af79a2d423d832c7f56c61cde9f8

dase-local@10-23-192-237:~$ tail -f flink-1.12.1/log/flink-dase-local-taskexecutor-0-10-23-192-237.out
(a,2)
(b,1)
(b,2)
(c,1)
hello : 1
ecnu : 1
hello : 1
dase : 1
hello : 1
ecnu : 1
```

正确的应该是这样：

我都重启了一下，包括第一个终端的监听输入，第二个终端中的 job 就好了：

```
dase-local@10-23-192-237:~$ nc -lk 8888
hello ecnu
hello dase
```

```
dase-local@10-23-192-237:~$ flink-1.12.1/bin/flink run flink-1.12.1/examples/streaming/SocketWindowWordCount
.jar --port 8888
Job has been submitted with JobID 5ad85d5a08324ec20532fc1a0d9b1ddf
```

```
dase-local@10-23-192-237:~$ tail -f flink-1.12.1/log/flink-dase-local-taskexecutor-0-10-23-192-237.out
(a,2)
(b,1)
(b,2)
(c,1)
hello : 1
ecnu : 1
hello : 1
dase : 1
hello : 1
ecnu : 1
hello : 2
dase : 1
ecnu : 1
```

#### 问题 4: 提交 jar 包之后报错并终止运行

Socket Window WordCount	2024-05-24 19:35:08	58ms	2024-05-24 19:35:08	2 1 1	FAILED
Socket Window WordCount	2024-05-24 19:34:23	104ms	2024-05-24 19:34:23	2 1 1	FAILED

无论是 attached 还是 detached 模式，都要先启动 netcat 之后才能提交 jar 包，否则就是报错

#### 问题 5: 输出结果和输入的速度紧密相关，可能会导致和实验文档中的预期结果存在差异

输出结果和输入的速度紧密相关，如果不紧不慢地慢慢输 hi dase hi ecnu flink flink，输出就是全 1。但如果已经写好了要输入的内容，复制粘贴到 netcat 中，就输出 2 1 1

为了避免这一细节带来的影响，我后来输入数据都是讲内容写好，然后直接复制粘贴到 netcat 中。

## 华东师范大学数据科学与工程学院实验报告

课程名称：分布式计算系统

年级：2022

上机实践成绩：

指导教师：徐辰

姓名：郭夏辉

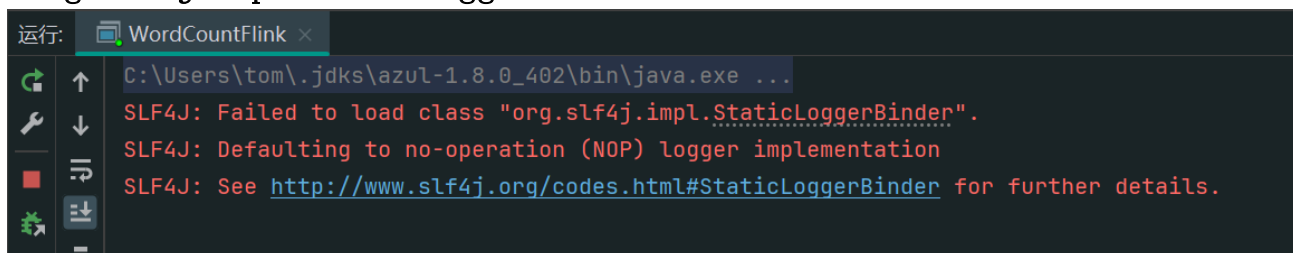
学号：10211900416

上机实践名称：Flink 编程

上机实践日期：

2024.6.11

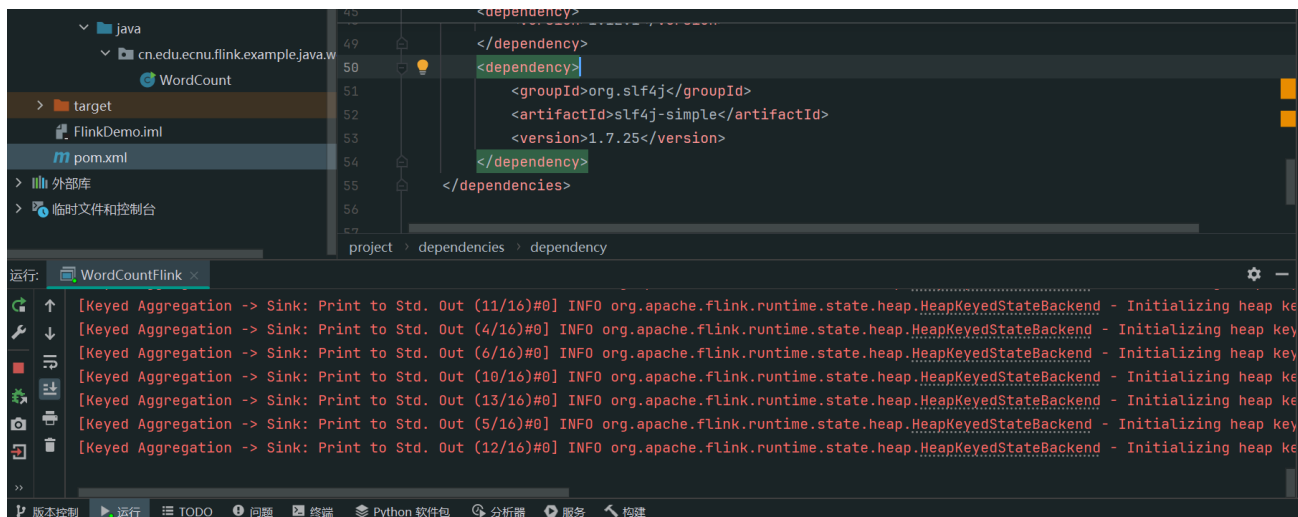
问题 1：IDEA 里面去运行程序时报错 SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".



参考这里的思路

(<https://blog.csdn.net/wang465745776/article/details/80384210>) pom.xml 加上这句话就好了：

```
<dependency>
    <groupId>org.slf4j</groupId>
    <artifactId>slf4j-simple</artifactId>
    <version>1.7.25</version>
</dependency>
```



问题 2：本地运行 scala 版本时，报错 java.lang.ClassNotFoundException

经过一段时间的排查，我发现程序运行时找不到 org.apache.flink.client.program，然后进一步检查发现 IDEA 中的报错：No Scala SDK in module

在把 Scala SDK 引入后，问题并没有解决。然后我又检查了一下，发现自己的 Flink 版本是错误的，在 pom.xml 中修改成如下的样子就好了：

```
<dependency>  
  <groupId>org.apache</groupId>  
  <artifactId>flink-clients_2.11</artifactId>  
  <version>1.12.1</version>  
</dependency>
```

## 华东师范大学数据科学与工程学院实验报告

课程名称：分布式计算系统

年级：2022

上机实践成绩：

指导教师：徐辰

姓名：郭夏辉

学号：10211900416

上机实践名称：Flink+Yarn

上机实践日期：

2024.6.18

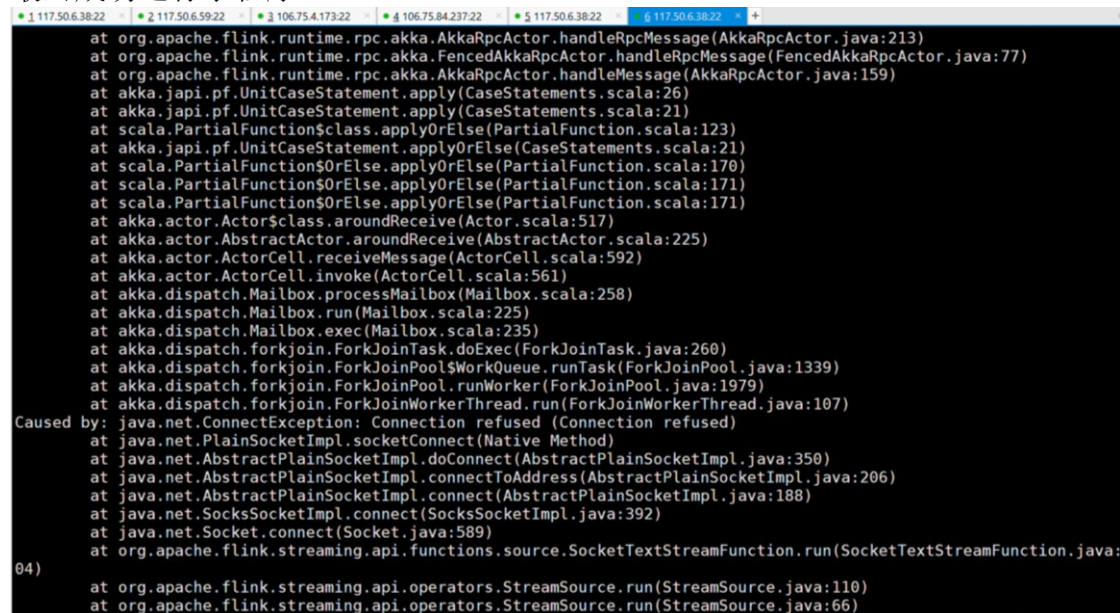
## 问题 1: hadoop: command not found

在进行单机伪分布式部署实验时，我修改了 `bashrc` 文件以使其生效，但不巧的是我创建云主机时镜像选错了，这台主机之前并没有安装 Hadoop。因此，我安装了 Hadoop，并确保 `bashrc` 文件生效。

但是随后我发现启动 HDFS 后缺少了 Namenode 进程。经过一系列的检查，我发现自己配置文件中有多处错误，并且忘记初始化 Namenode。最终，我修正了 Hadoop 的依赖文件，并成功初始化了 Namenode，顺利运行了 `wordcount` 程序。

## 问题 2 提交 JAR 包的机器错误

在进行在分布式部署实验时，我提交了 JAR 包，发现跑不起来。最后，我发现我是在主节点 `ecnu01` 上监听了 8888 端口并提交了 JAR 包，而不是客户端 `ecnu04`。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，最终成功运行了程序。



```
at org.apache.flink.runtime.rpc.akka.AkkaRpcActor.handleRpcMessage(AkkaRpcActor.java:213)
at org.apache.flink.runtime.rpc.akka.FencedAkkaRpcActor.handleRpcMessage(FencedAkkaRpcActor.java:77)
at org.apache.flink.runtime.rpc.akka.AkkaRpcActor.handleMessage(AkkaRpcActor.java:159)
at akka.japi.pf.UnitCaseStatement.apply(CaseStatements.scala:26)
at akka.japi.pf.UnitCaseStatement.apply(CaseStatements.scala:21)
at scala.PartialFunction$class.applyOrElse(PartialFunction.scala:123)
at akka.japi.pf.UnitCaseStatement.applyOrElse(CaseStatements.scala:21)
at scala.PartialFunction$OrElse.applyOrElse(PartialFunction.scala:170)
at scala.PartialFunction$OrElse.applyOrElse(PartialFunction.scala:171)
at akka.actor.Actor$class.aroundReceive(Actor.scala:517)
at akka.actor.AbstractActor.aroundReceive(AbstractActor.scala:225)
at akka.actor.ActorCell.receiveMessage(ActorCell.scala:592)
at akka.actor.ActorCell.invoke(ActorCell.scala:561)
at akka.dispatch.Mailbox.processMailbox(Mailbox.scala:258)
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:04)
at org.apache.flink.streaming.api.operators.StreamSource.run(StreamSource.java:110)
at org.apache.flink.streaming.api.operators.StreamSource.run(StreamSource.java:66)
```

## 问题 3 分布式实验时报错

`org.apache.flink.client.program.ProgramInvocationException: The main method caused an error: No Executor found. Please make sure to export the HADOOP_CLASSPATH`

之所以出现这样的报错是因为我缺少执行器（Executor），换言之就是环境变量没有设置 HADOOP\_CLASSPATH 或没有将 Hadoop 添加到类路径。我最终的程序是在 ecnu04 上运行的，所以需要在这上面像最开始 ecnu01 那样去配置 .bashrc 并使之生效。当然我也去在 ecnu02 和 ecnu03 上面配置 .bashrc 并 source ~/.bashrc, 最终就没有报错了。