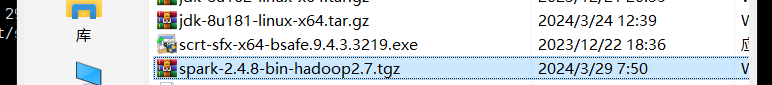
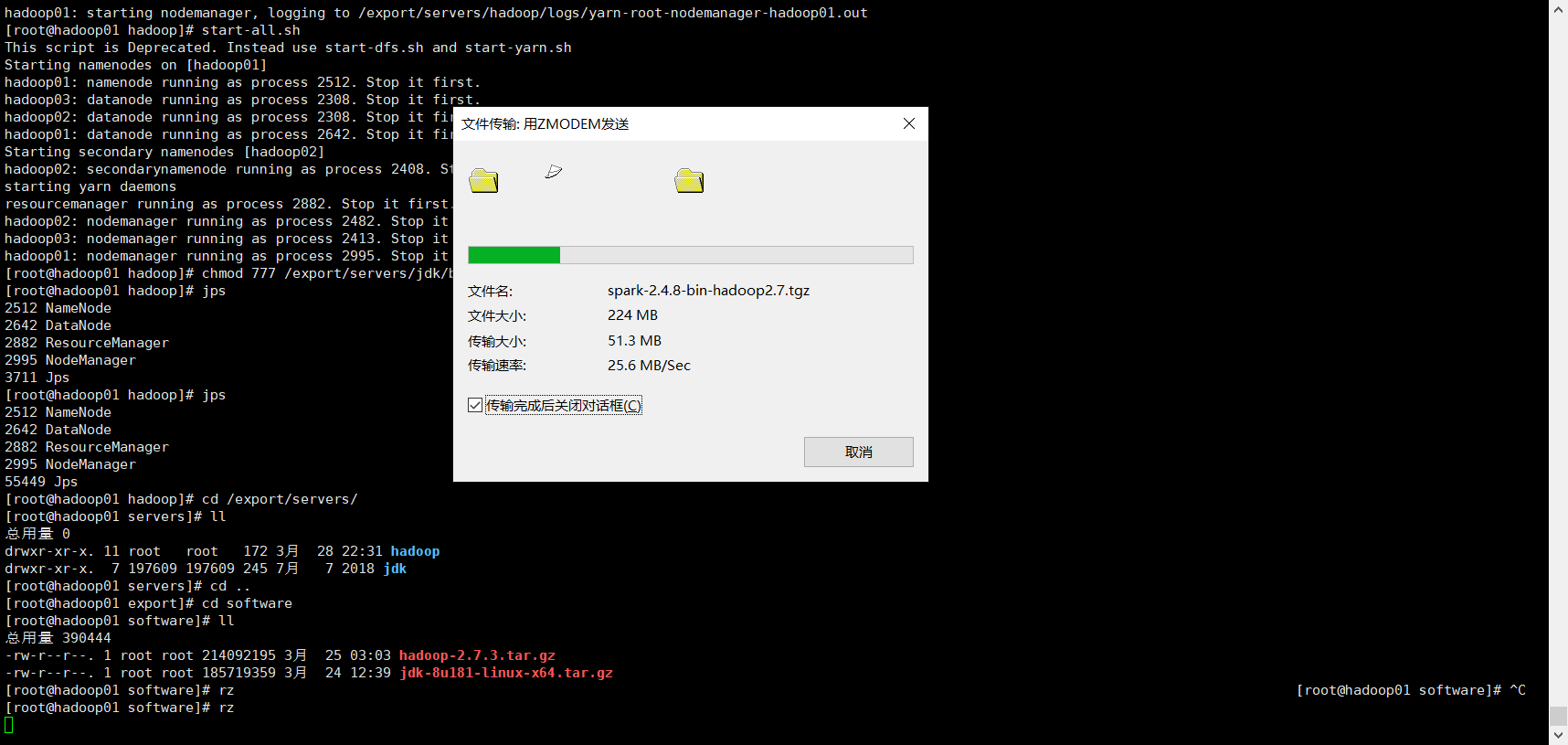
S

第一步下载Spark安装包，本次我用的是

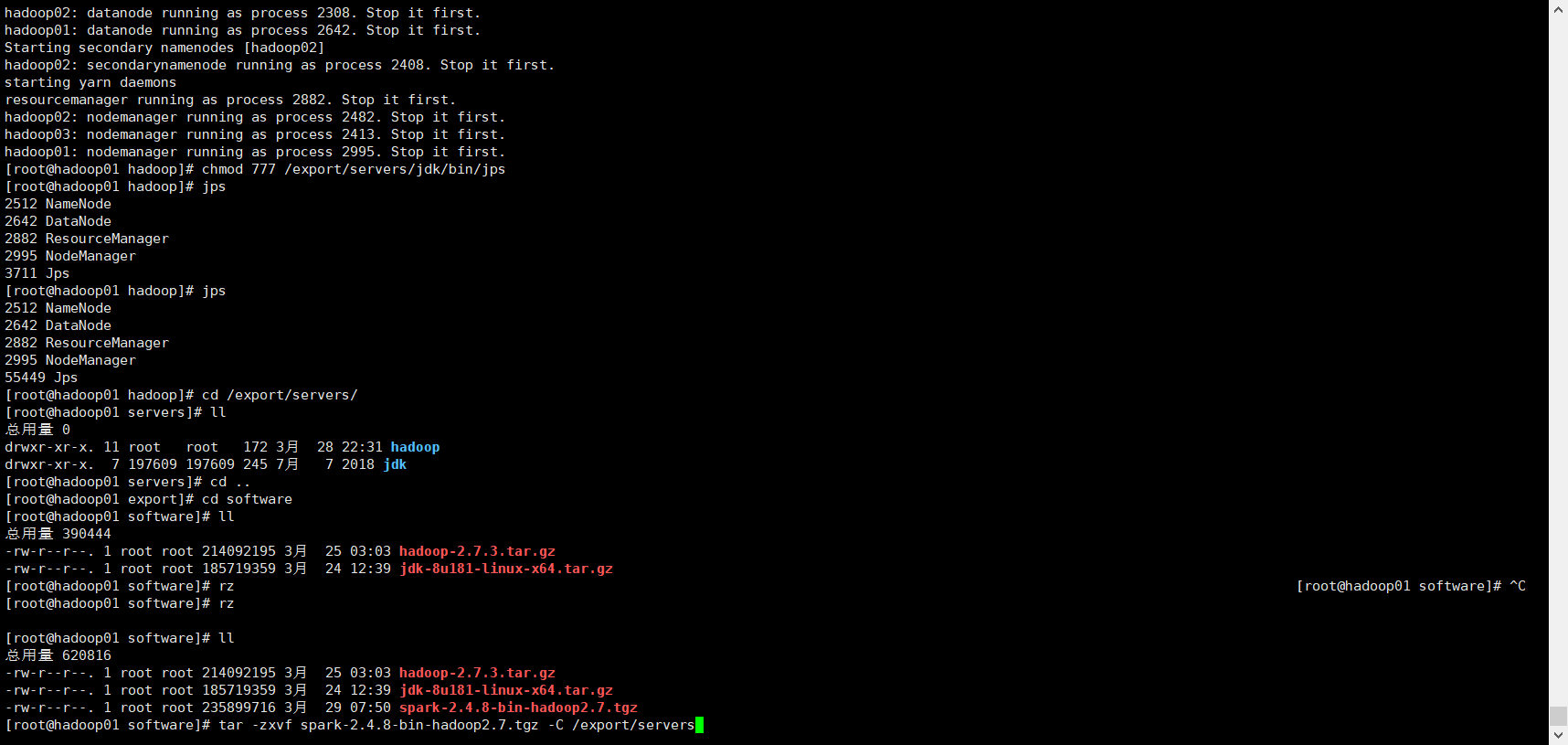
spark-2.4.8-bin-hadoop2.7.tgz，下载时间比较长，

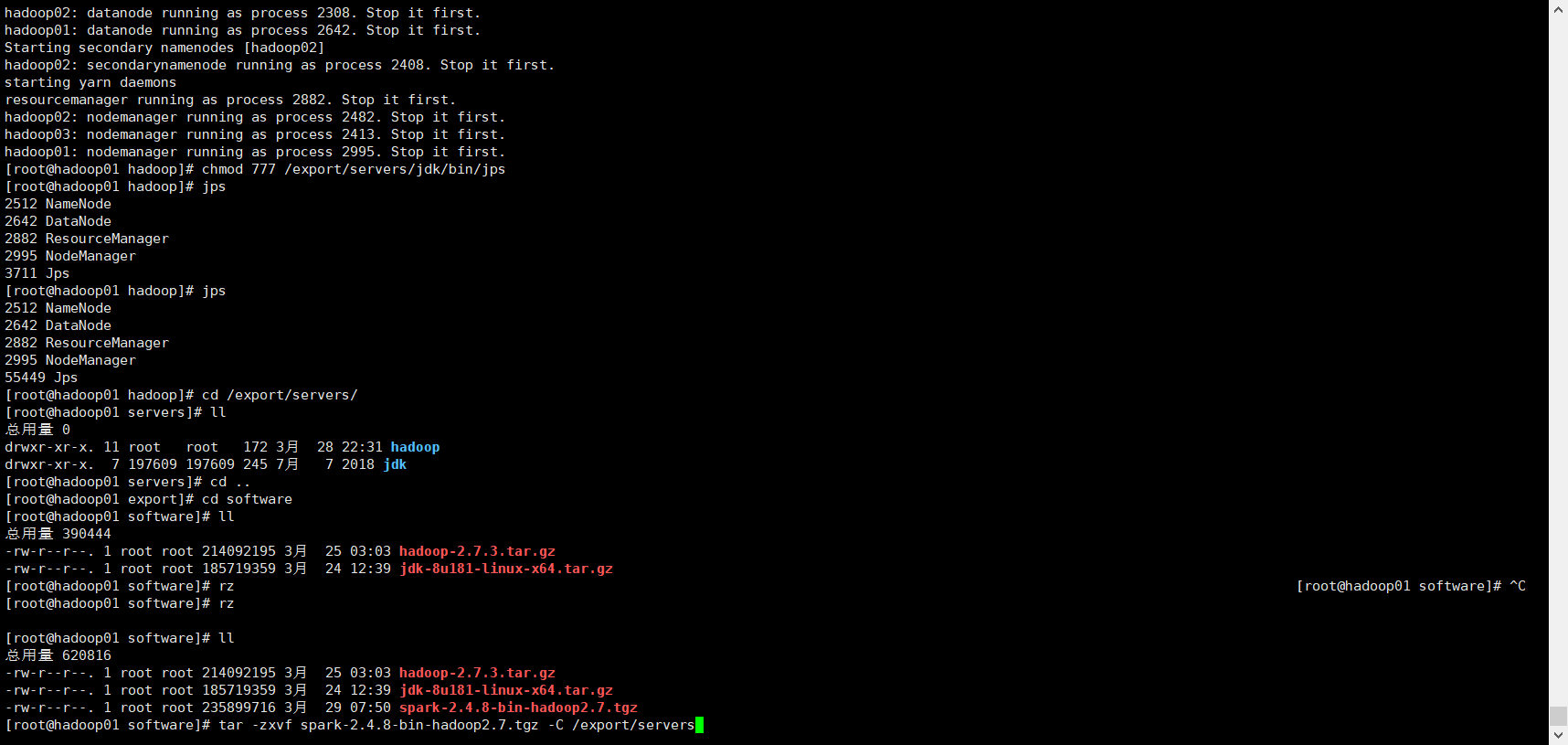
下载好安装包以后，输入命令cd /export/software 切换目录，用 rz 命令上传到节点上





上传完毕，输入命令：tar -zxvf spark-2.4.8-bin-hadoop2.7.tgz -C /export/servers 进行解压安装。



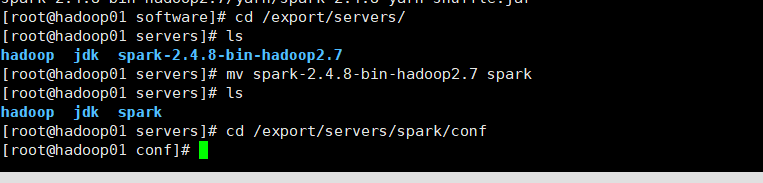


解压安装完成，命令：cd /export/servers/ 切换到/export/servers/目录，输入命令：

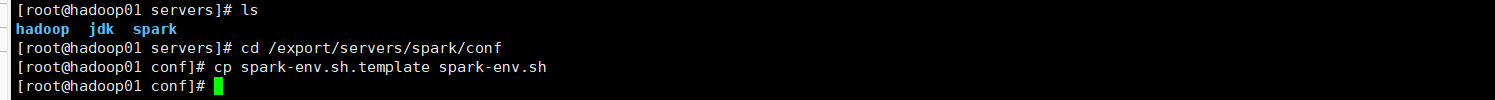
mv spark-2.4.8-bin-hadoop2.7 spark 进行重命名

接下来进行主要步骤配置spark 集群文件

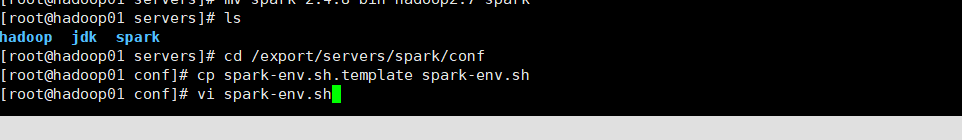
输入命令：cd /export/servers/spark/conf 进入需要配置的文件目录中，



用命令：cp spark-env.sh.template spark-env.sh 复制一份配置文件，方便在此基础之上进行修改



用命令：vi spark-env.sh 打开文件进行编辑



输入i 切换到内容插入模式，将以下内容复制进配置文件末尾，

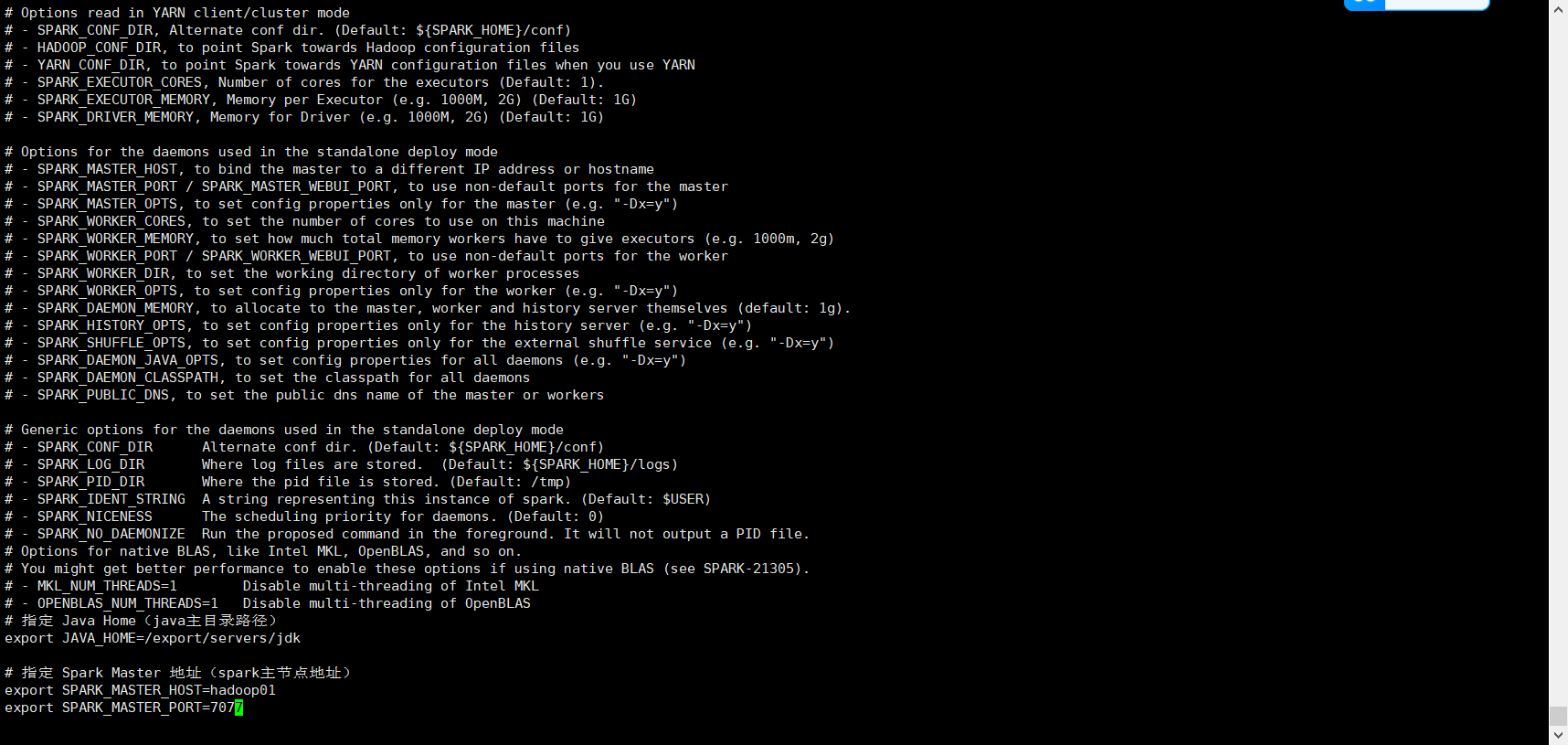
# 指定 Java Home（java主目录路径）

export JAVA\_HOME=/export/servers/jdk

# 指定 Spark Master 地址（spark主节点地址）

export SPARK\_MASTER\_HOST=hadoop01

export SPARK\_MASTER\_PORT=7077



然后按Esc键退出插入模式，再用 :wq 保存退出.

#标蓝为方便复制用

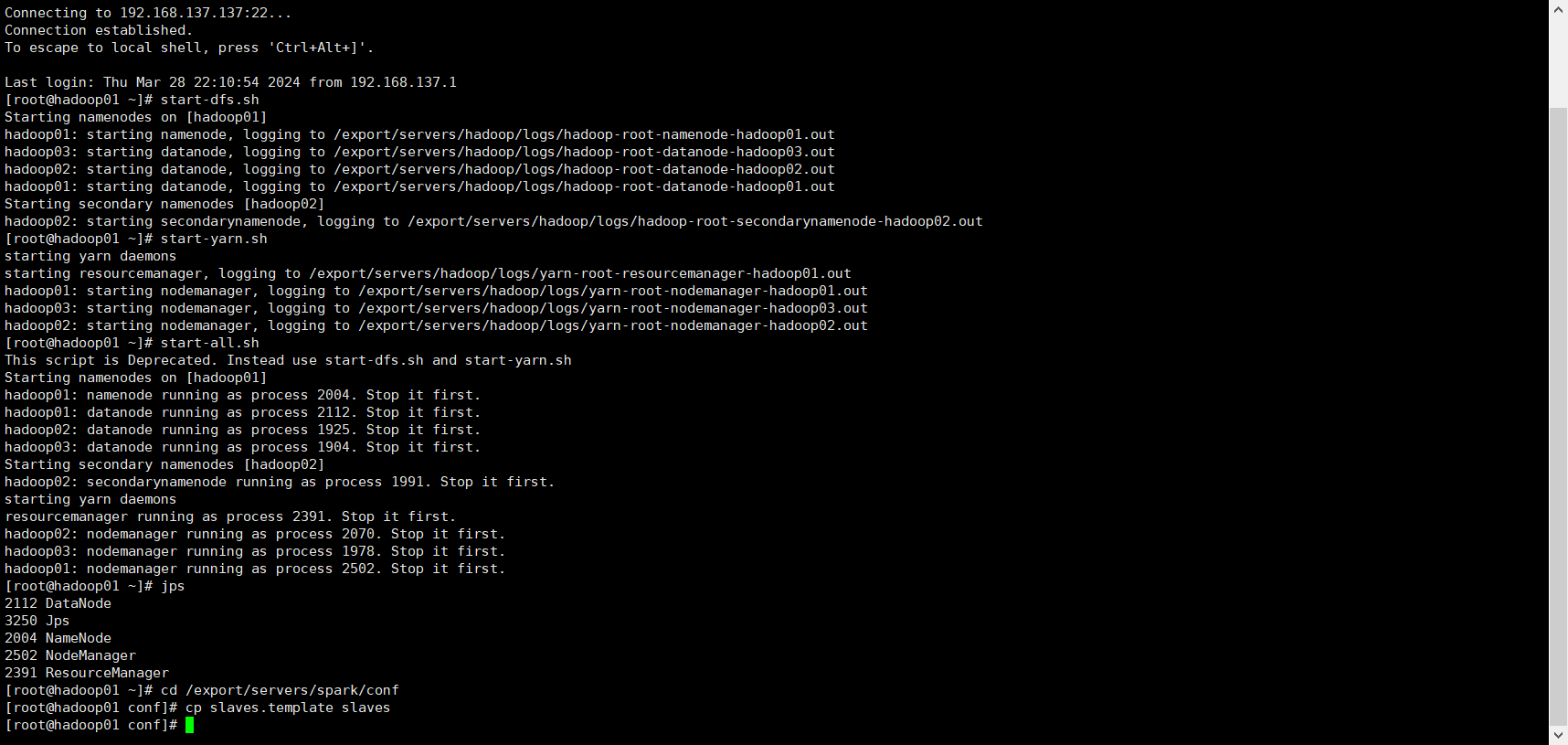
命令：cd /export/servers/spark/conf

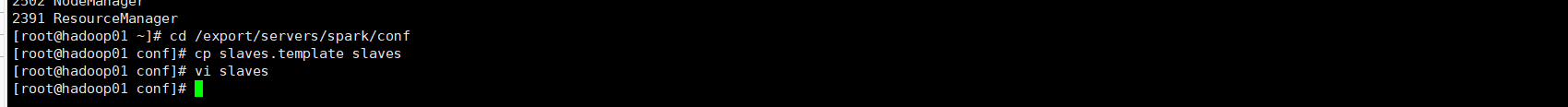
命令：cp slaves.template slaves

命令：vi slaves

修改配置文件 slaves, 以指定从节点为止, 从在使用 sbin/start-all.sh 启动集群的时候, 可以一键启动整个集群所有的 Worker(工作节点)

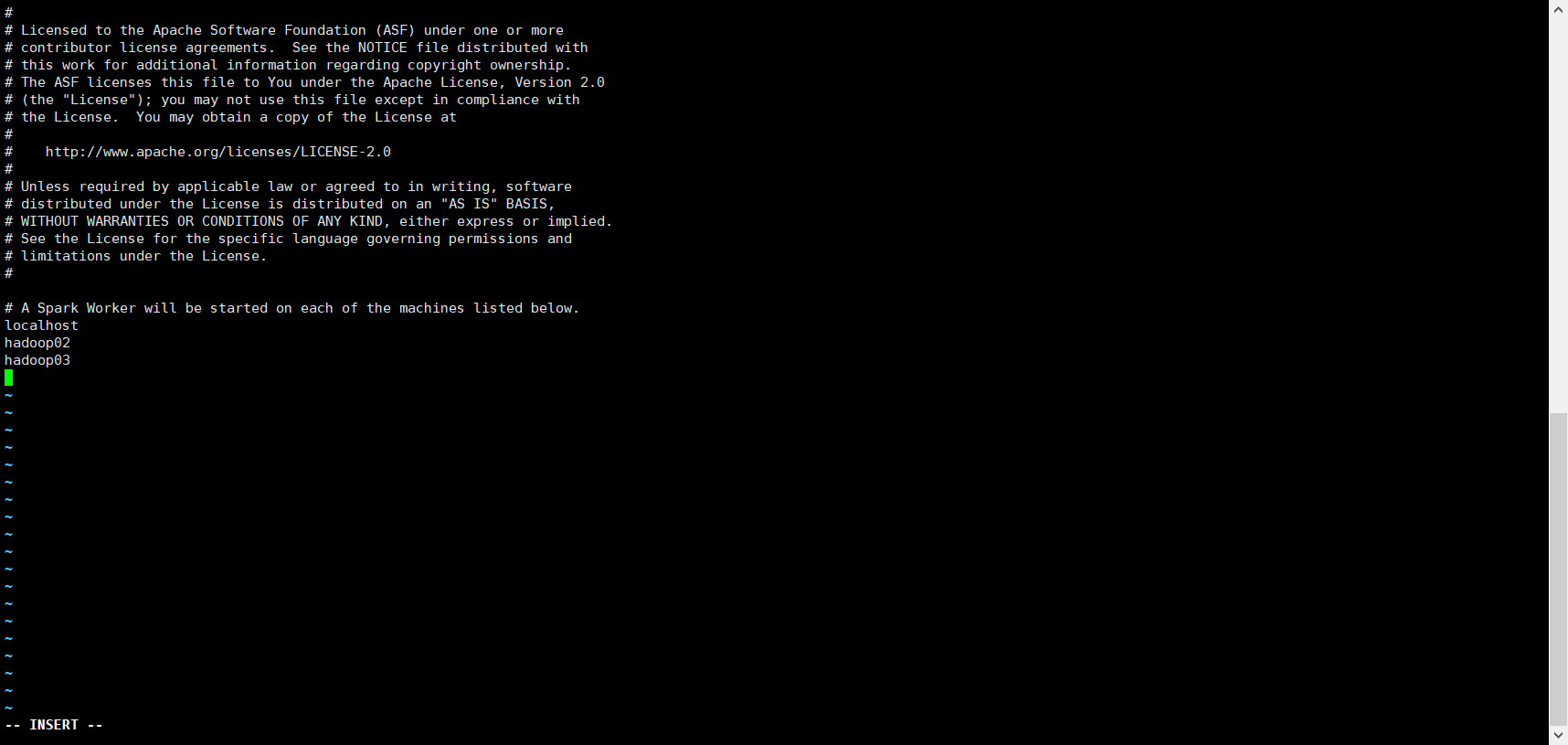
进入配置目录, 并复制一份新的配置文件, 以供在此基础之上进行修改





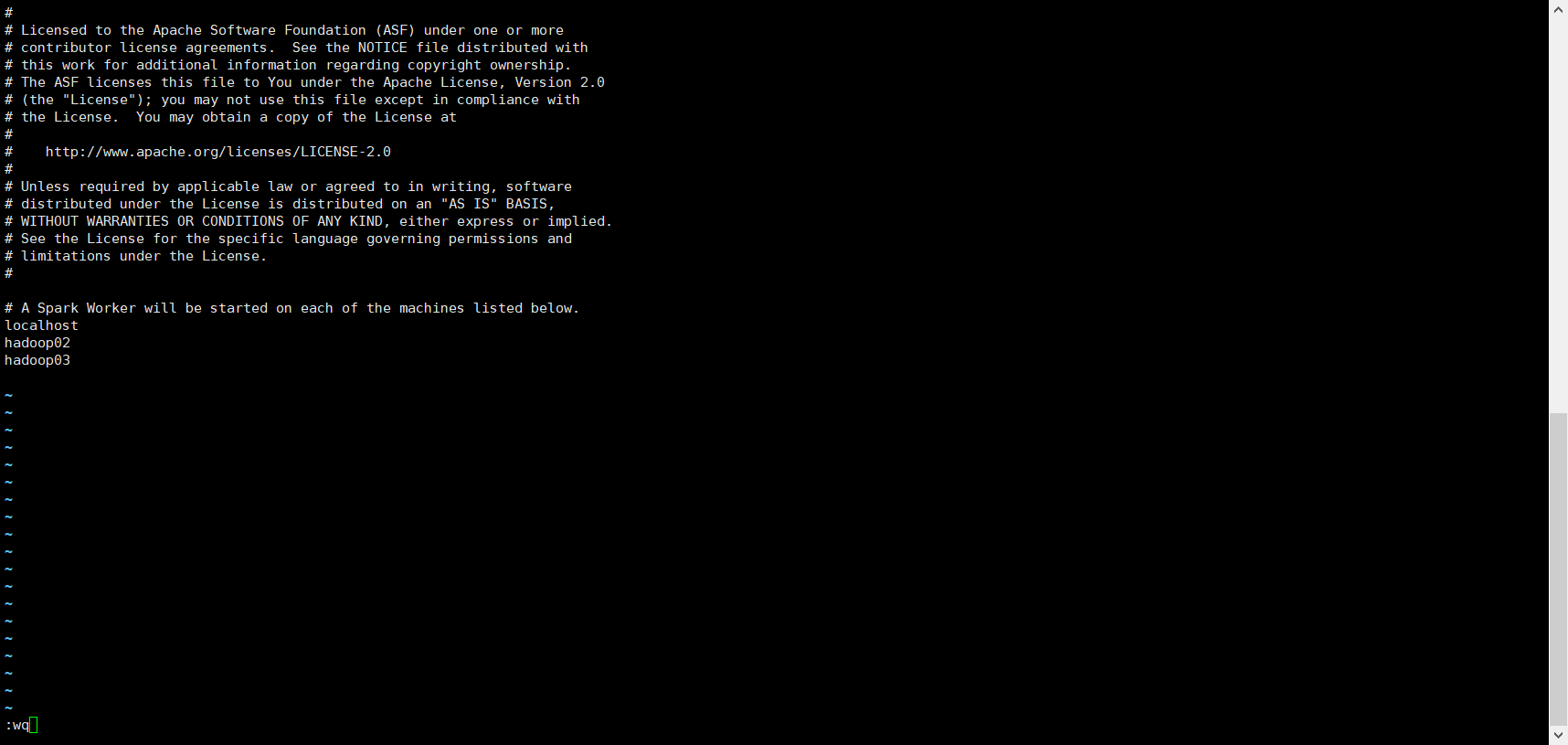
使用命令

cd /export/servers/spark/conf 切换目录，再用命令：cp slaves.template slaves 复制slaves配置文件，接着命令 vi slaves 开始编辑文件内容配置



按i切换到插入模式，添加内容为从节点名称：hadoop02、hadoop03;之后按Esc键，再用

：wq 命令保存退出。



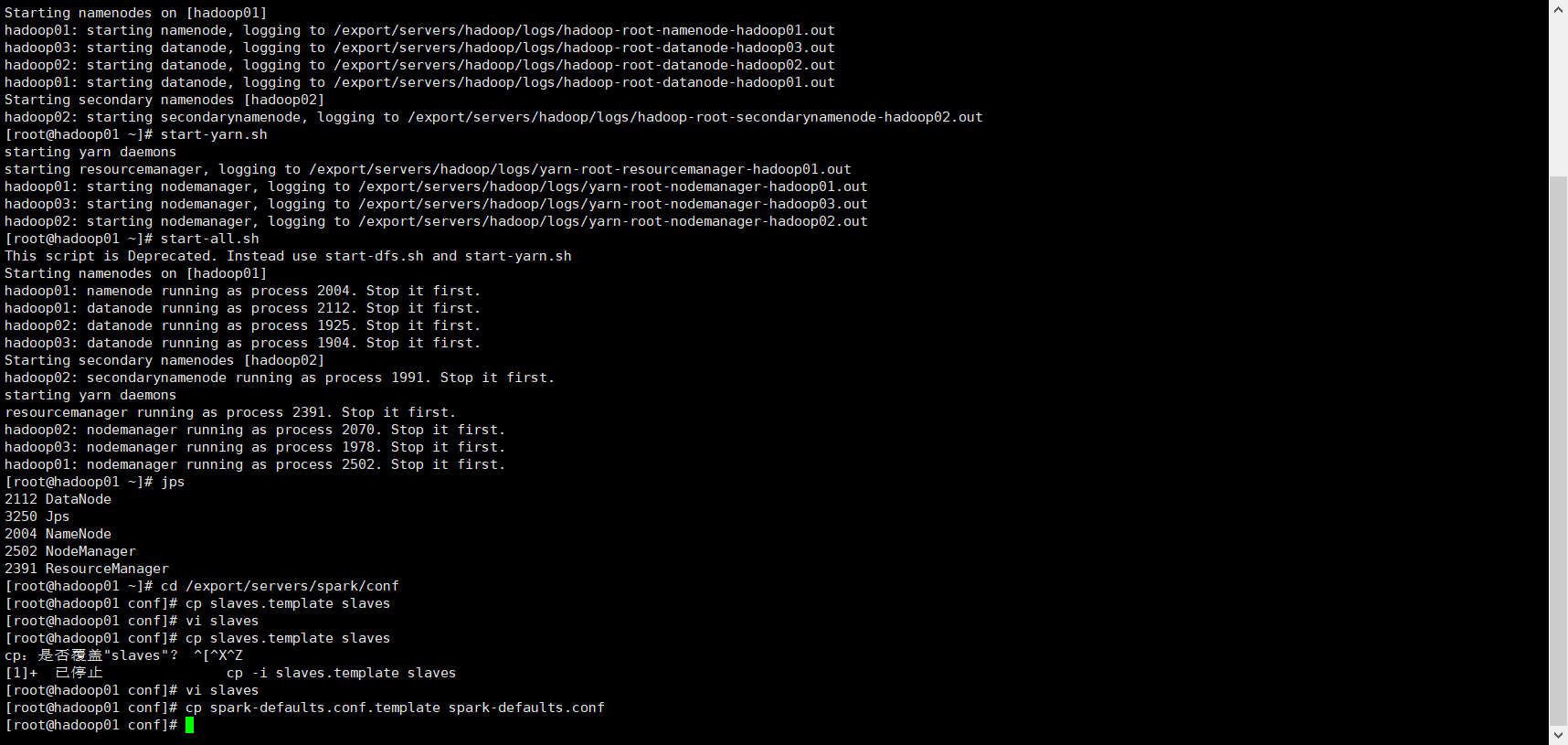
配置 HistoryServer （日志服务）

a.默认情况下, Spark 程序运行完毕后, 就无法再查看运行记录的 Web UI 了, 通过 HistoryServer 可以提供一个服务, 通过读取日志文件, 使得我们可以在程序运行结束后, 依然能够查看运行过程

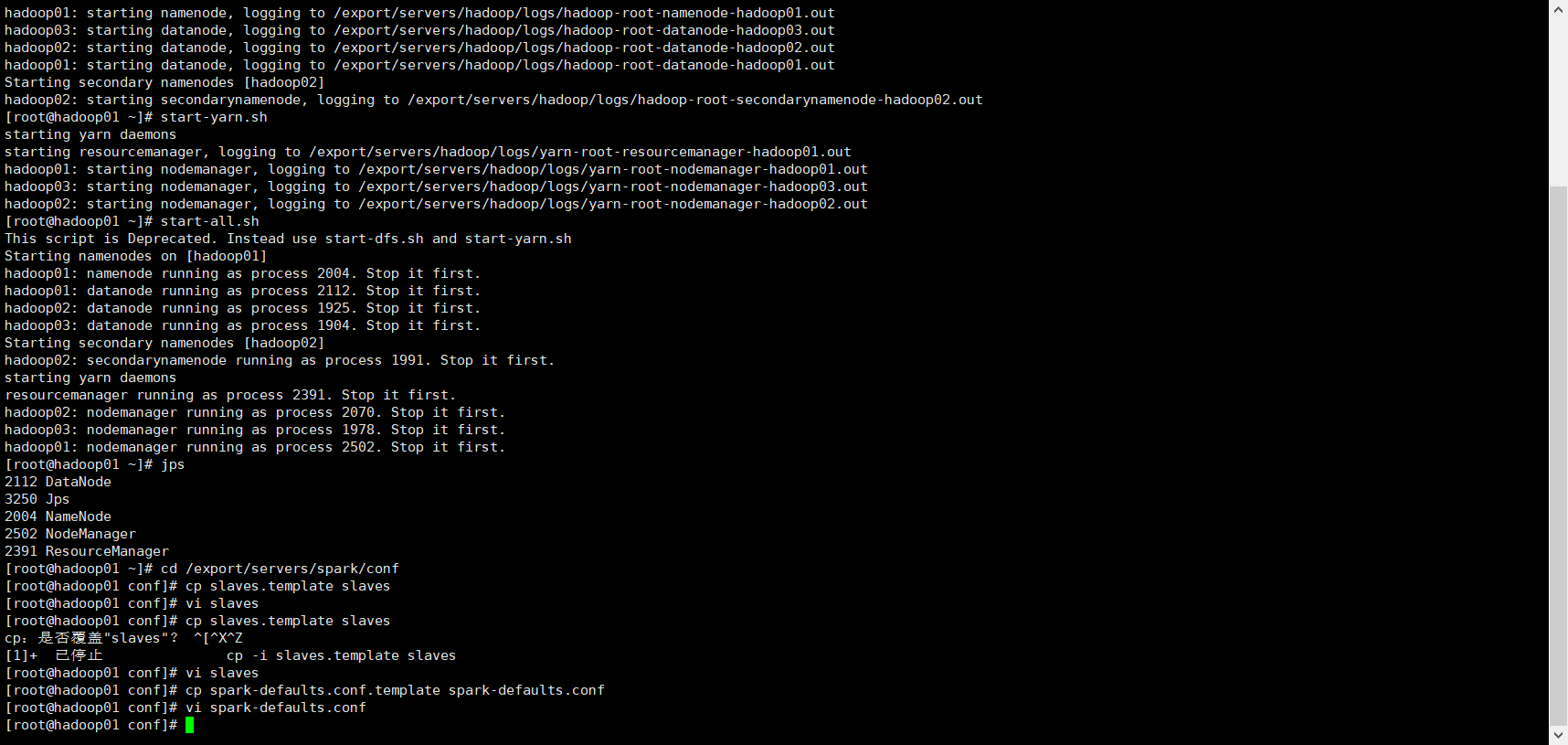
b.复制 spark-defaults.conf, 以供修改

使用命令 cd /export/servers/spark/conf 切换目录，在这里之前已经切换好了就不用再切换了

输入命令： cp spark-defaults.conf.template spark-defaults.conf 复制文件 spark-defaults.conf.template



接着输入命令：vi spark-defaults.conf 进行配置，

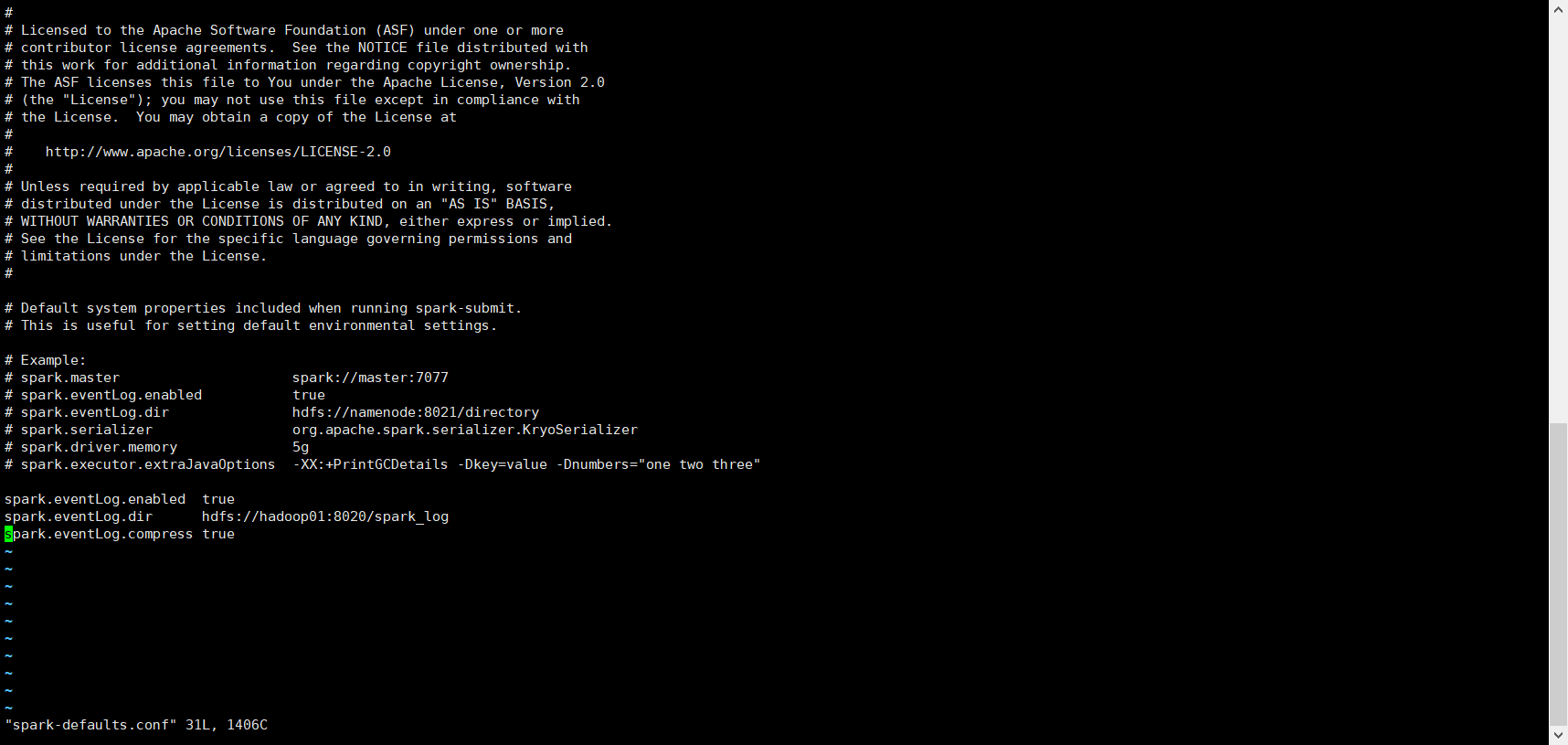


将以下内容复制到spark-defaults.conf末尾处, 通过这段配置, 可以指定 Spark 将日志输入到 HDFS 中

spark.eventLog.enabled true

spark.eventLog.dir hdfs://hadoop01:8020/spark\_log

spark.eventLog.compress true



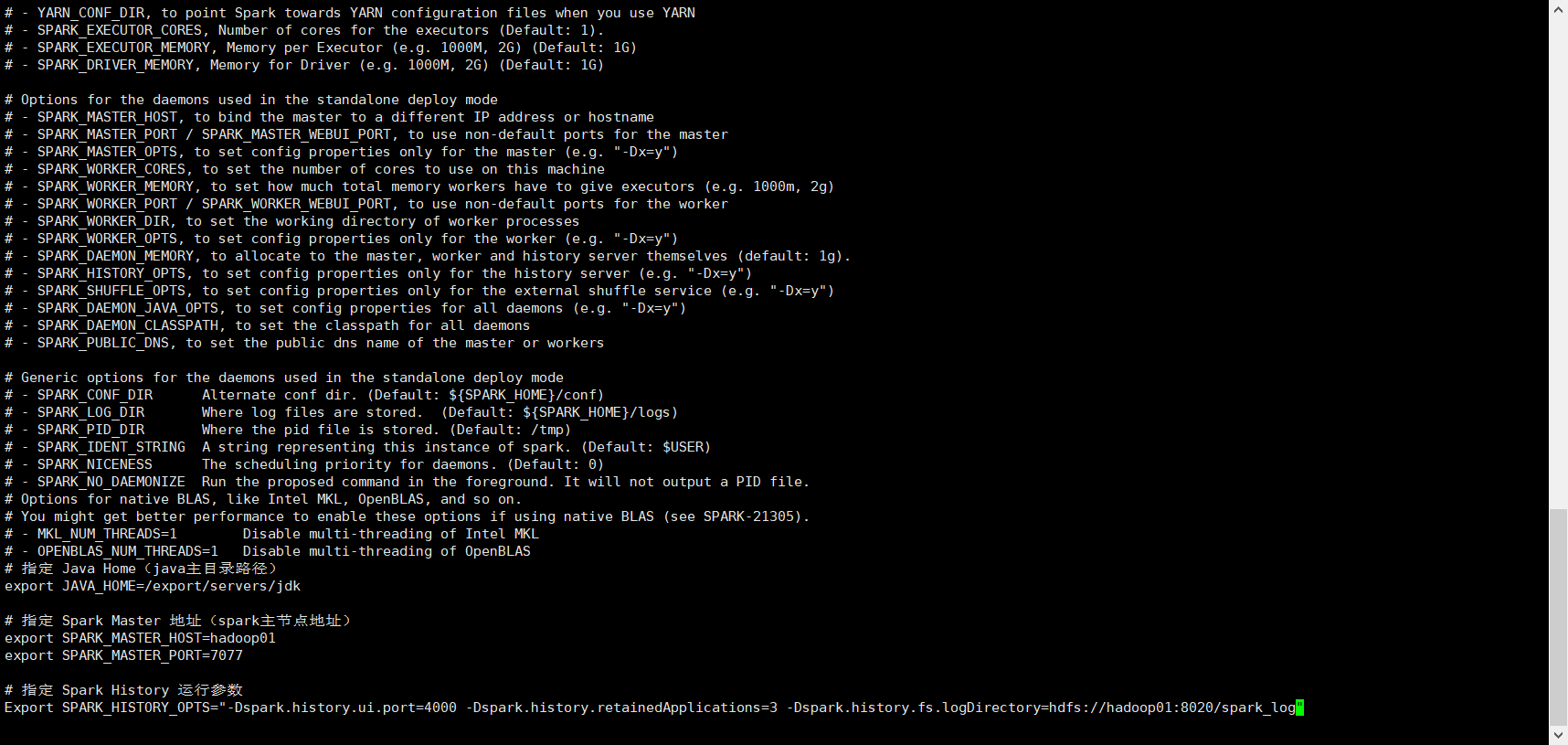
将以下内容复制到spark-env.sh的末尾, 配置 HistoryServer 启动参数, 使得 HistoryServer 在启动的时候读取 HDFS 中写入的 Spark 日志

添加内容：

# 指定 Spark History 运行参数

export SPARK\_HISTORY\_OPTS="-Dspark.history.ui.port=4000 -Dspark.history.retainedApplications=3 -Dspark.history.fs.logDirectory=hdfs://hadoop01:8020/spark\_log"

先用命令：vi spark-env.sh 打开文件spark-env.sh，接着按i进入插入模式添加内容

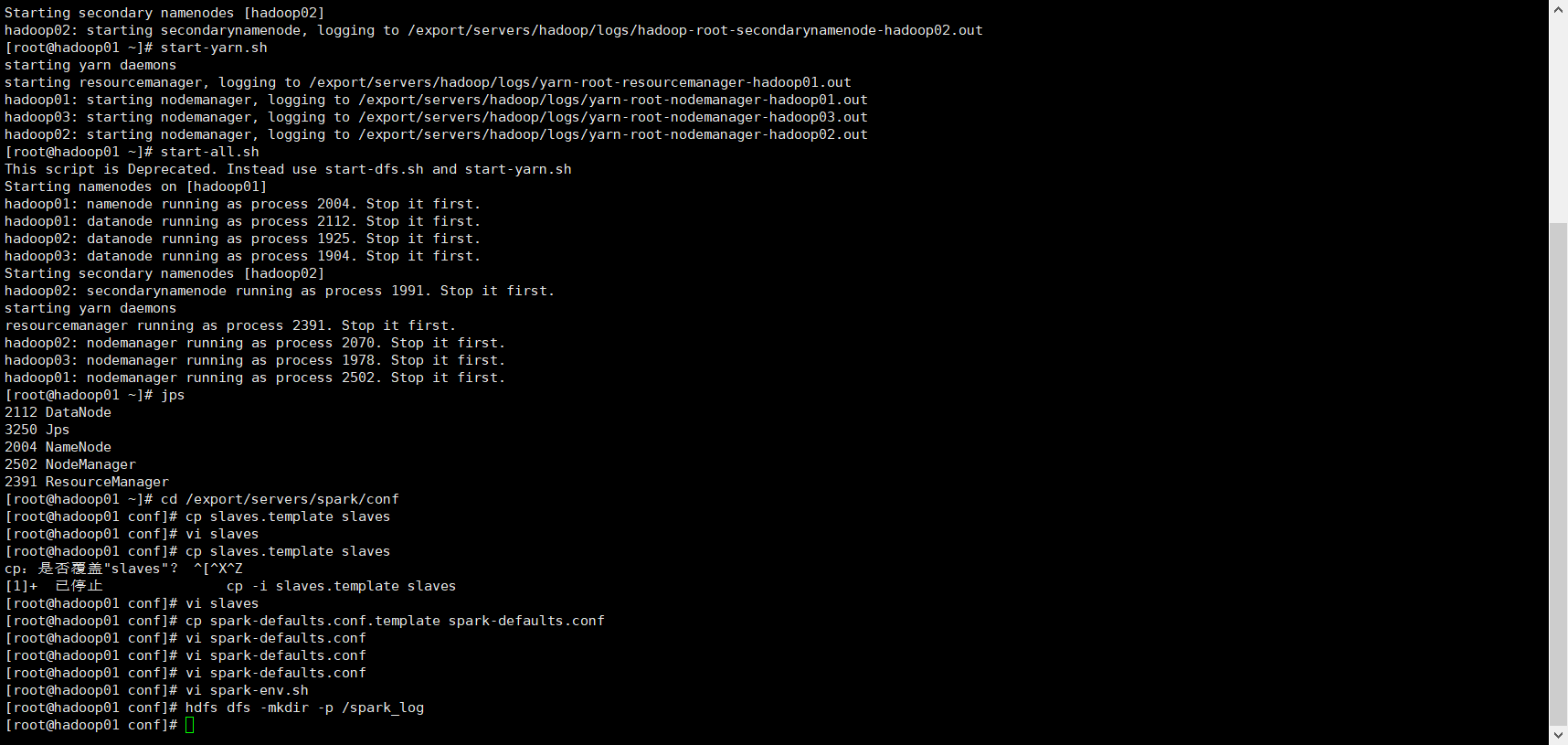


然后Esc键退出插入模式，命令；wq保存退出。

为 Spark 创建 HDFS 中的日志目录

命令：

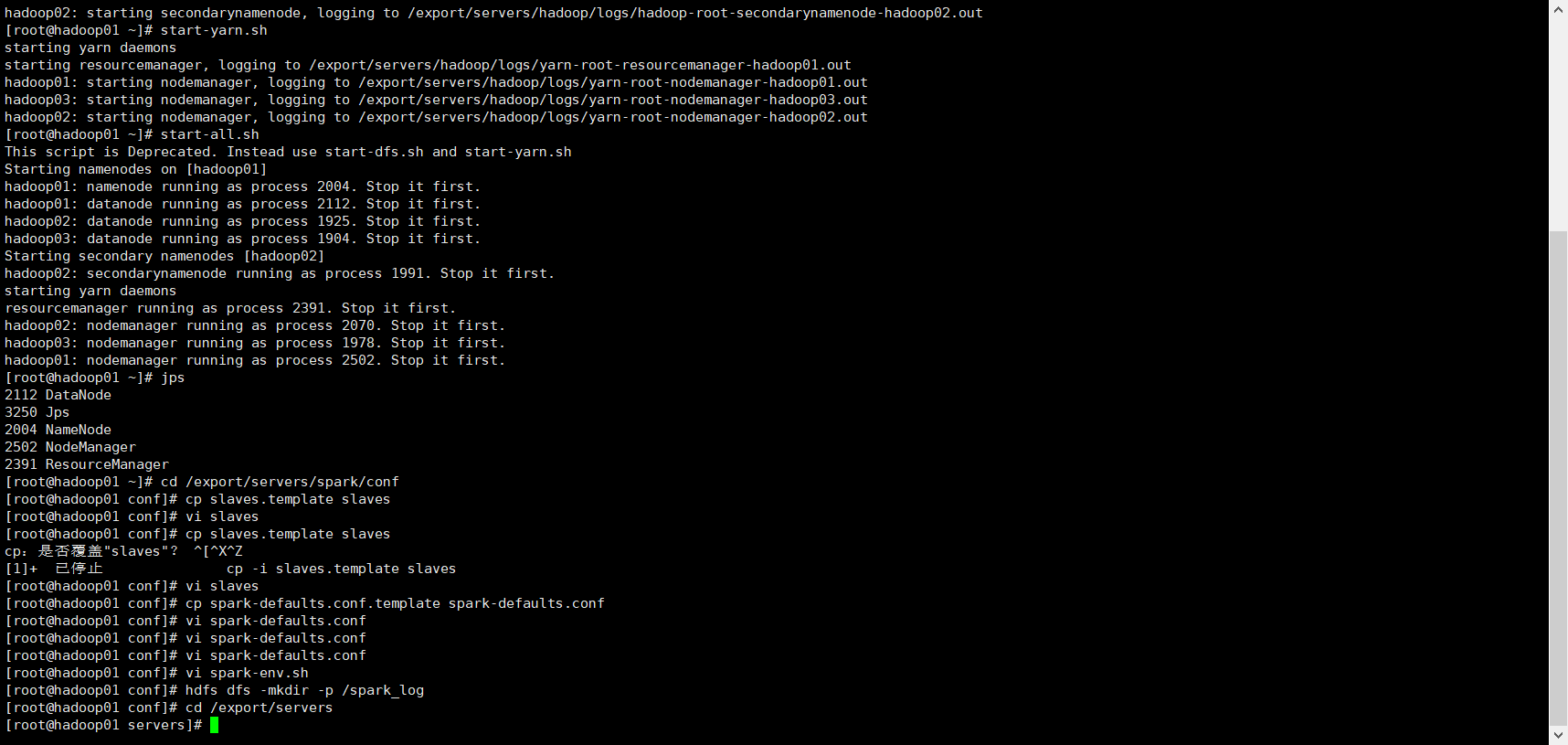
hdfs dfs -mkdir -p /spark\_log



到这里配置完成后，我们开始进入到分发和运行的步骤

1. 将 Spark 安装包分发给集群中其它机器

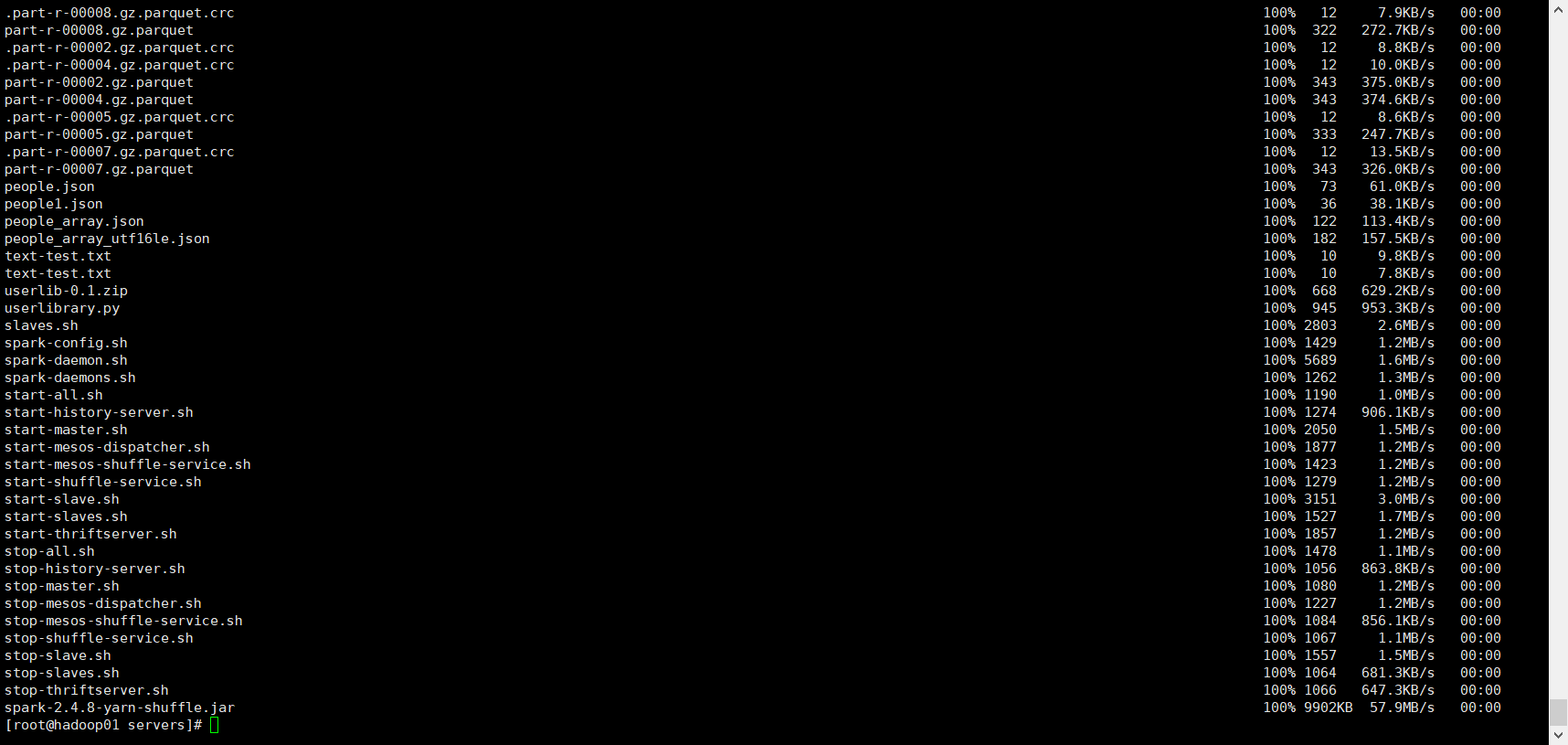
先用命令：cd /export/servers 切换目录，分别再用下面两条命令进行分发。

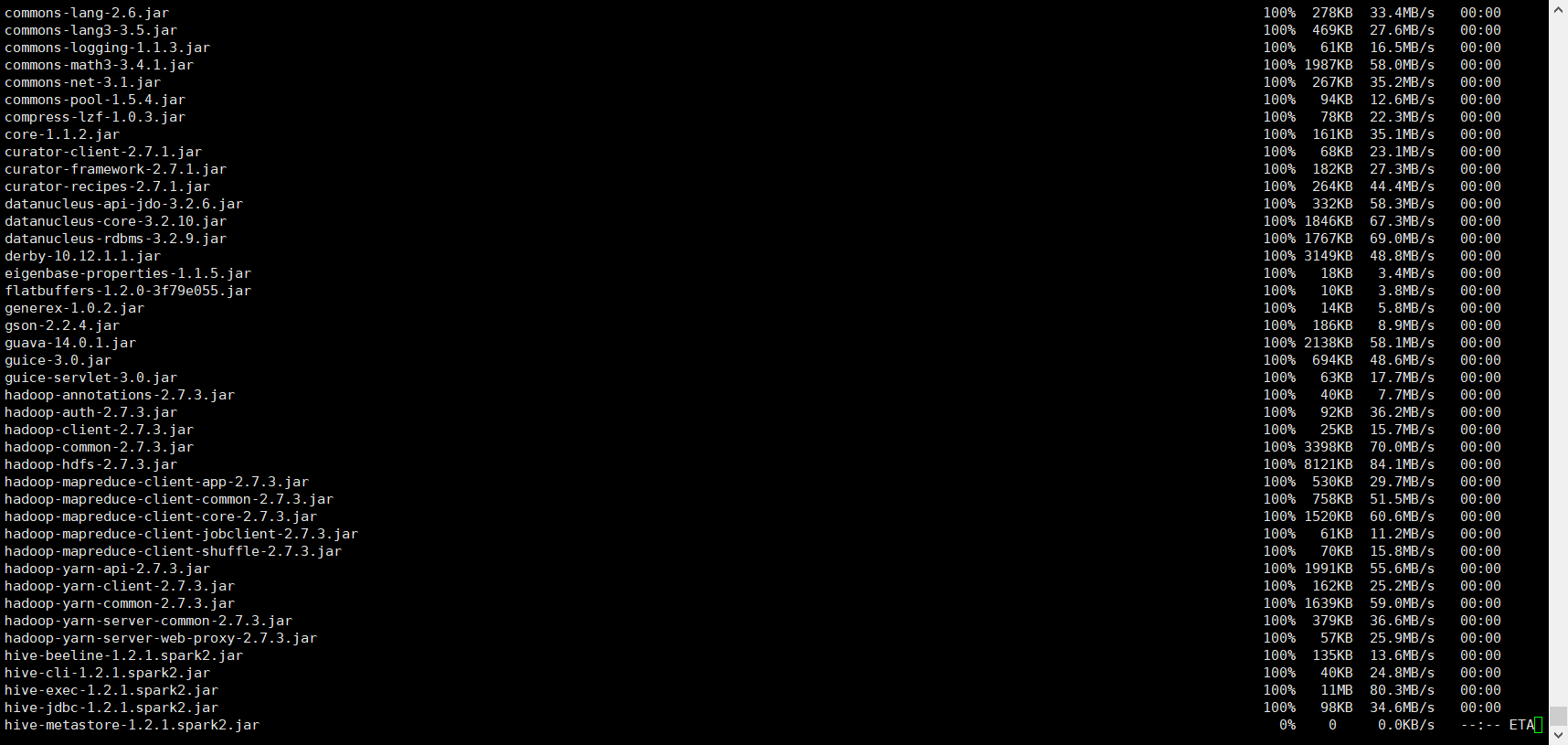


分别这两条命令给其他两个节点分发spark安装包的

scp -r spark root@hadoop02:$PWD

scp -r spark root@hadoop03:$PWD





启动 Spark Master 和 Slaves, 以及 HistoryServer

先使用命令：cd /export/servers/spark 切换目录

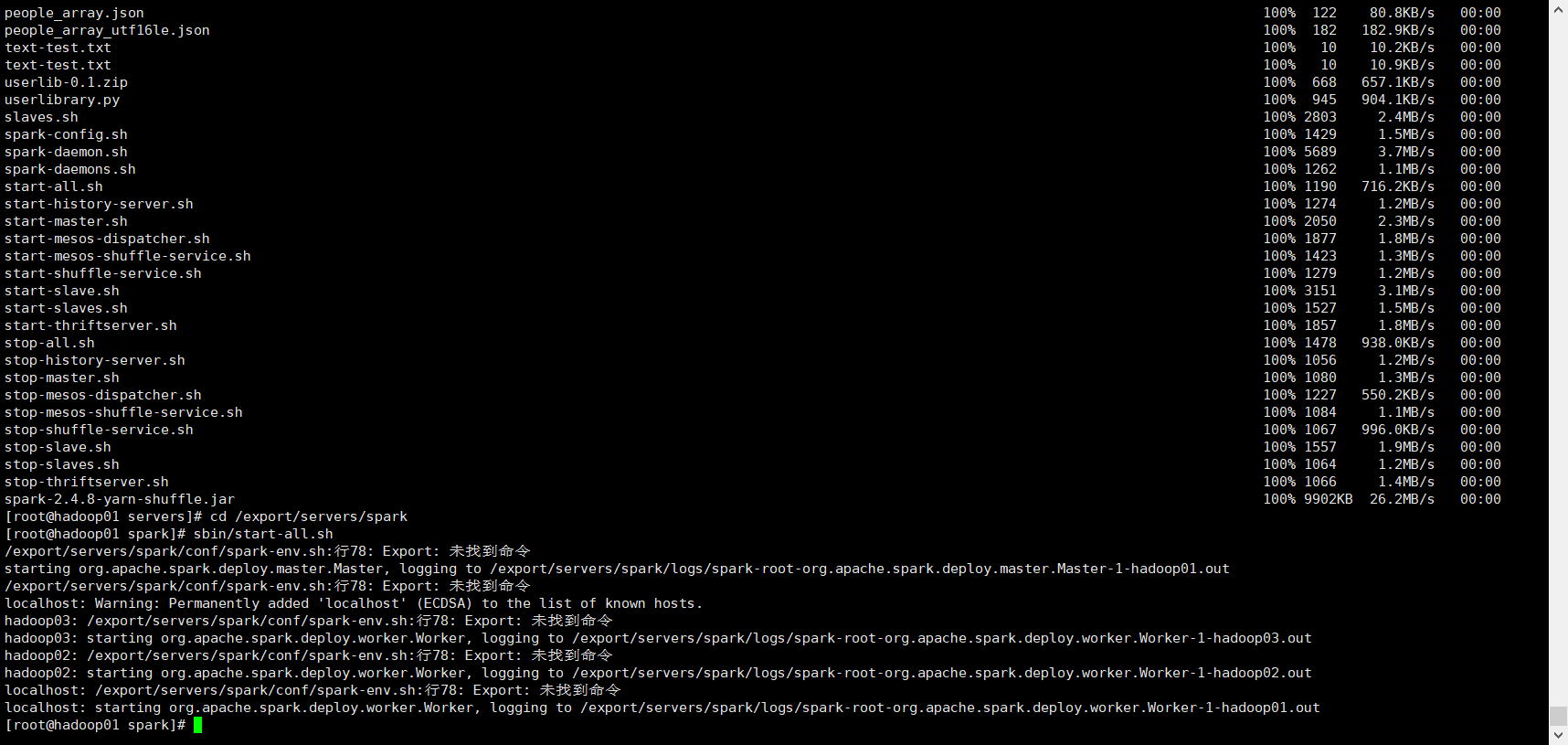
再用以下两条命令：sbin/start-all.sh

sbin/start-history-server.sh

启动spark集群。

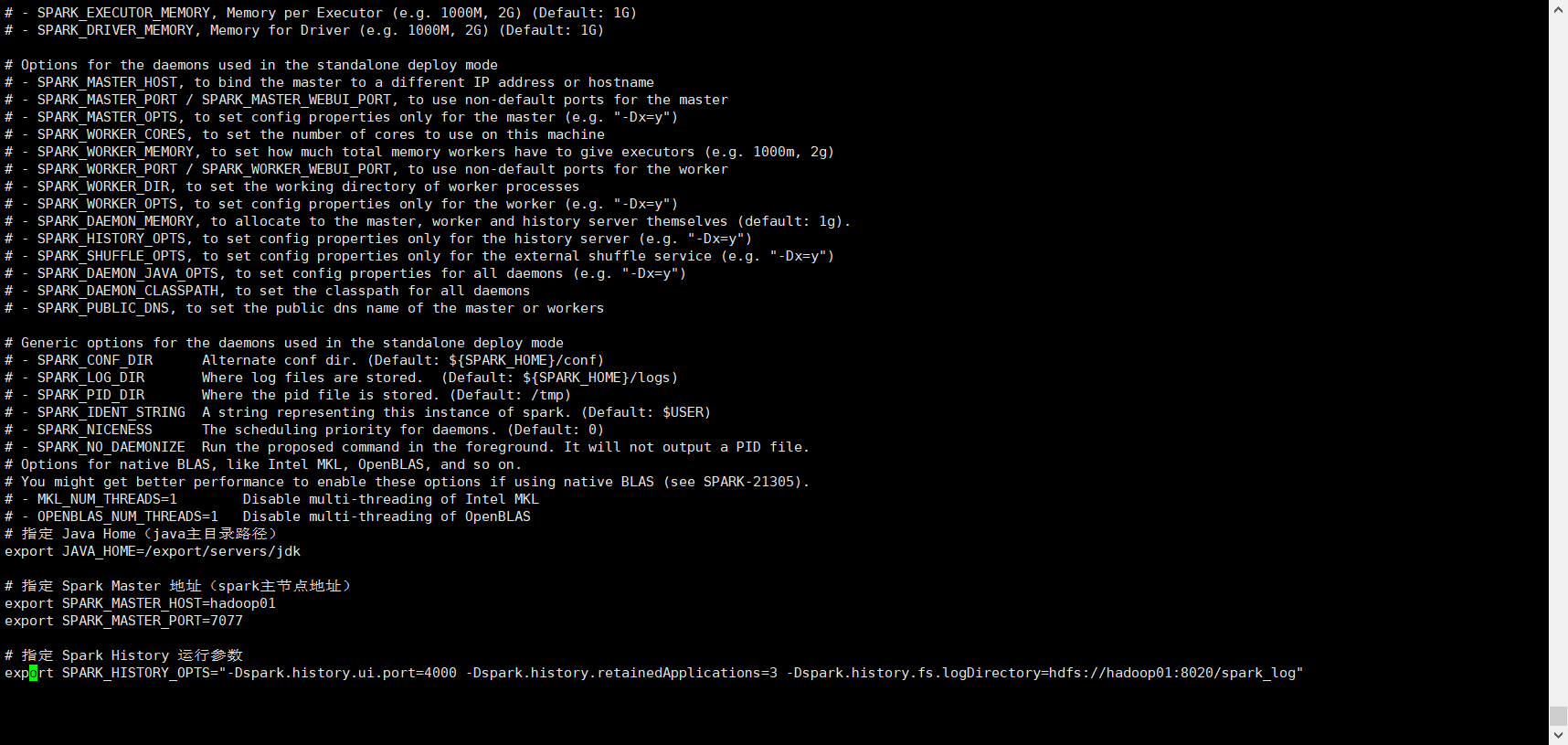
再使用：

sbin/start-all.sh启动时，如图出现这样的错误



此时需要排错，先切换目录cd /export/servers/spark/conf；

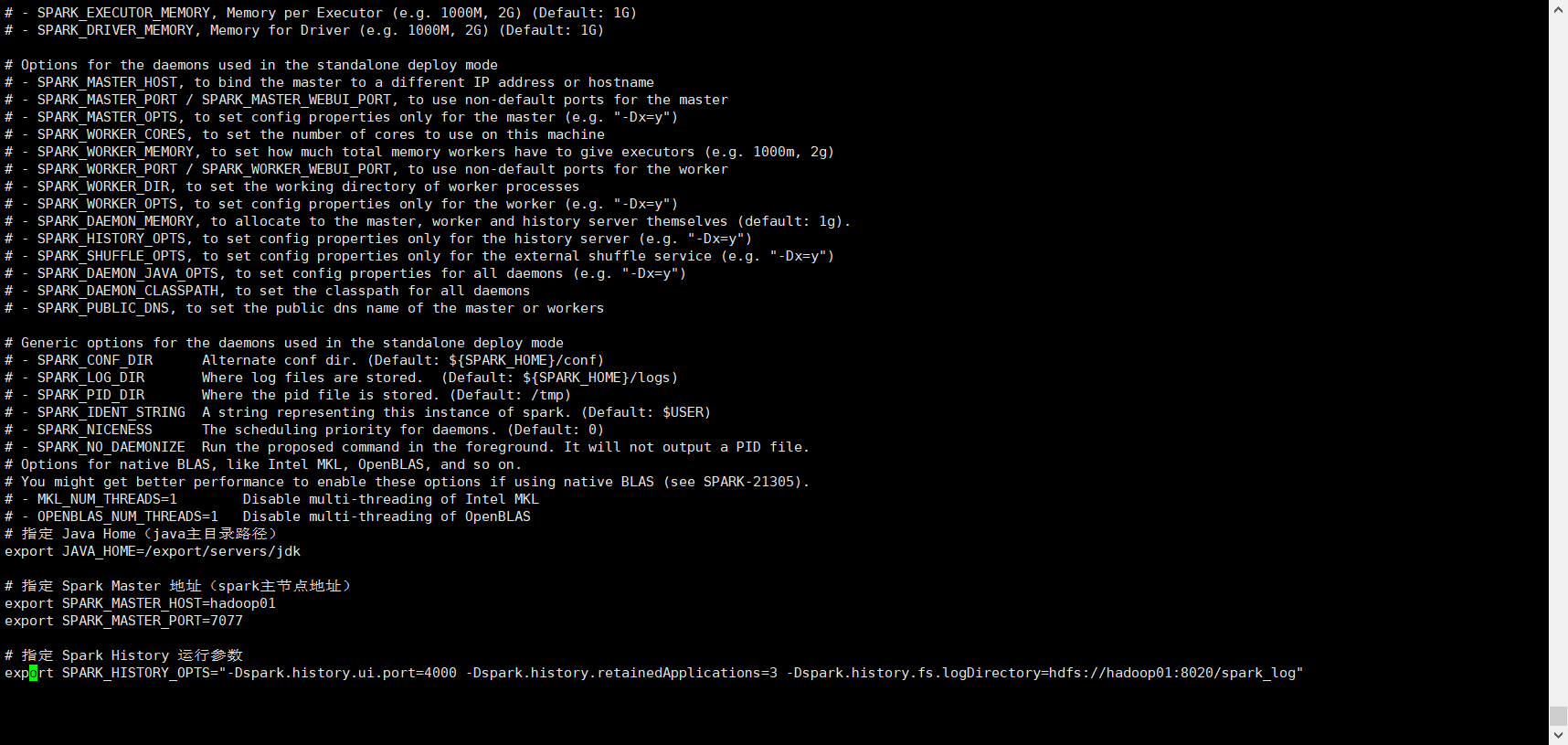
用命令vi spark-env.sh 打开文件，找到78行，检查错误。



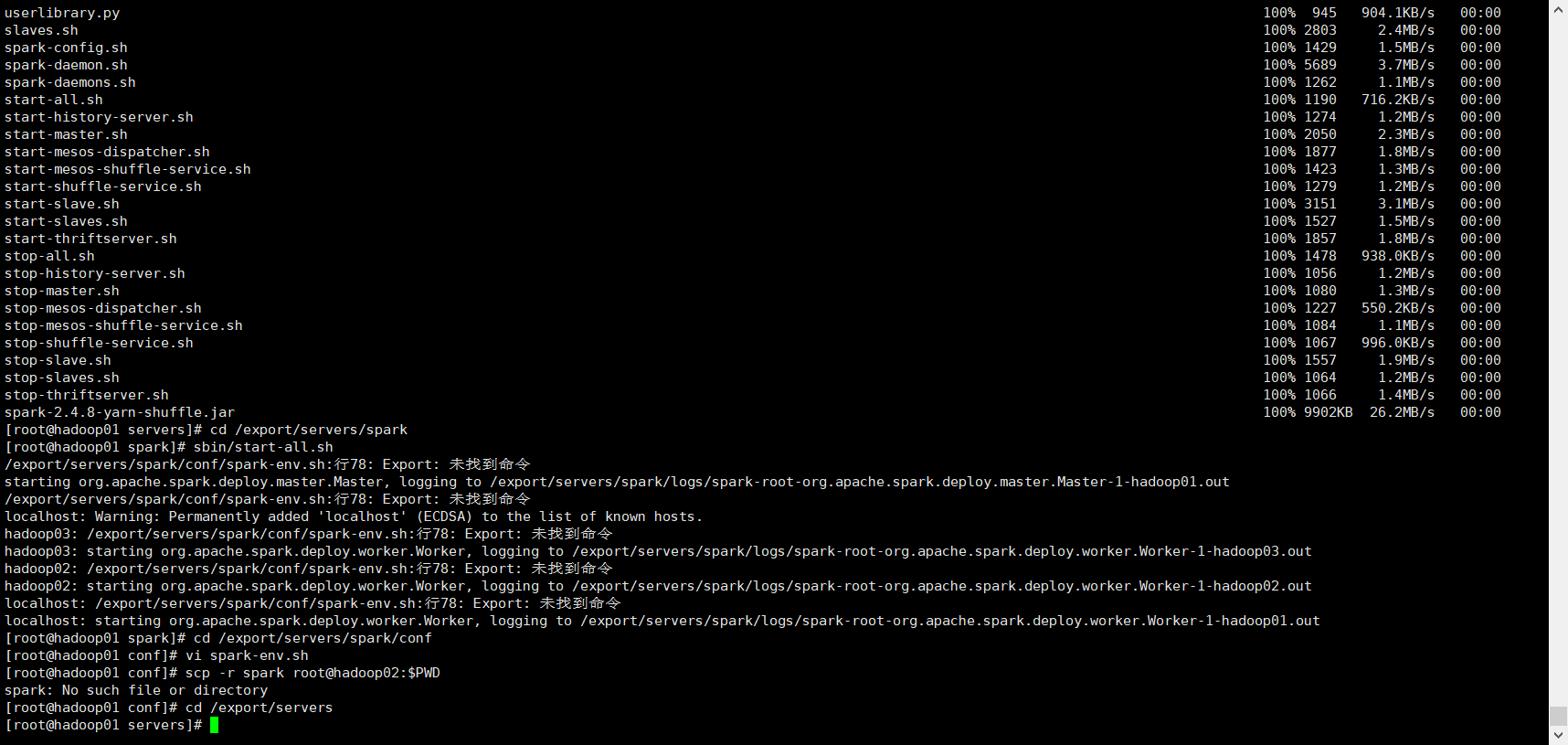
这里我将Eport改为小写字母开头export，保存退出后，重新分发一下。在这里备注；

插入命令：i

保存退出用 ：wq



还是先用命令：cd /export/servers 切换目录以后，进行重新分发。



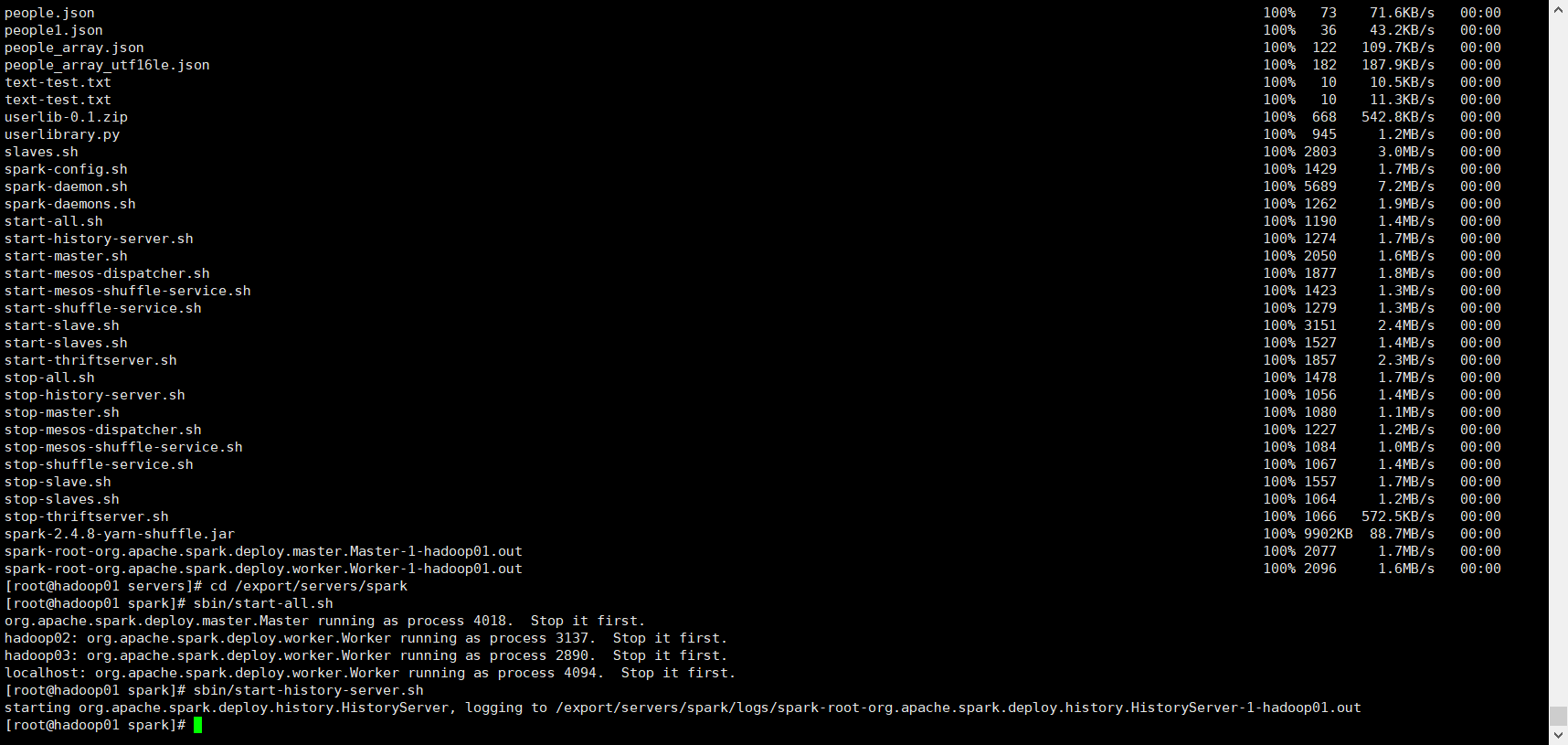
排错修改后，用命令切换目录cd /export/servers/spark

再分别用以下两条命令启动

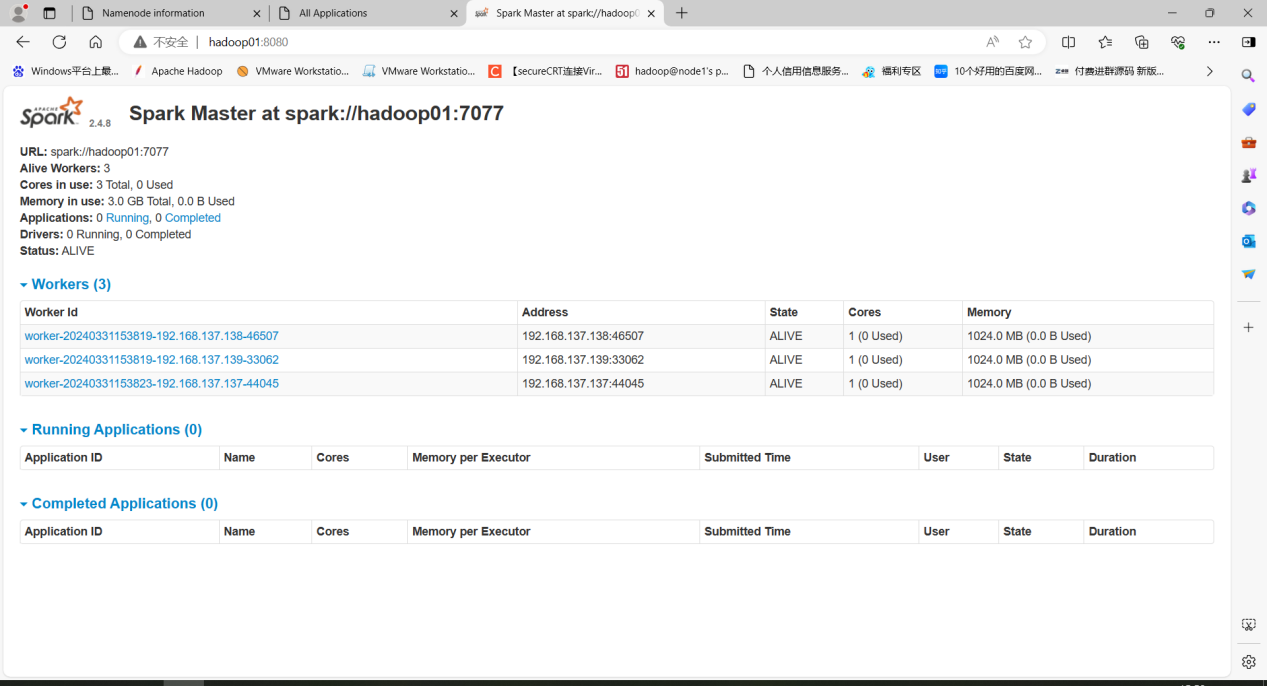
sbin/start-all.sh

sbin/start-history-server.sh

这时可以看到spark集群正常启动了，并无报错。



接着打开浏览器地址栏里面输入 hadoop01：8080 打开spark集群管理页。



到这里spark集群搭建完成了。