# 第一次作业

目录

[1、基本环境](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1001)

[2、实验须知](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1002)

[实验命令介绍](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1003)

[实验说明](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1004)

[3、HDFS高级操作命令](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1005)

[管理集群](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1006)

[设置限制文件数量命令](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1007)

[数据检查](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1008)

[数据均衡](http://192.168.145.11:8080/DataStudio3.0/student/training/detail/141/training/outline#_Toc1009)

**1、基本环境**

环境详情：

|  |  |
| --- | --- |
| 环境说明 | 安装目录 |
| 集群环境已安装hadoop-2.7.1完全分布式（环境开启后，自行启动） | /usr/hadoop-2.7.1 |

**2、实验须知**

**实验命令介绍**

与基本常用命令相比，本节实验主要是练习文件所属用户或组，文件权限，文件副本数等文件的高级命令和一些常用于管理员的命令，如管理命令，数据检查，数据均衡命令等。普通用户一般用于HDFS上文件/文件夹的权限管理，但集群管理命令只能以HDFS超级用户身份运行。

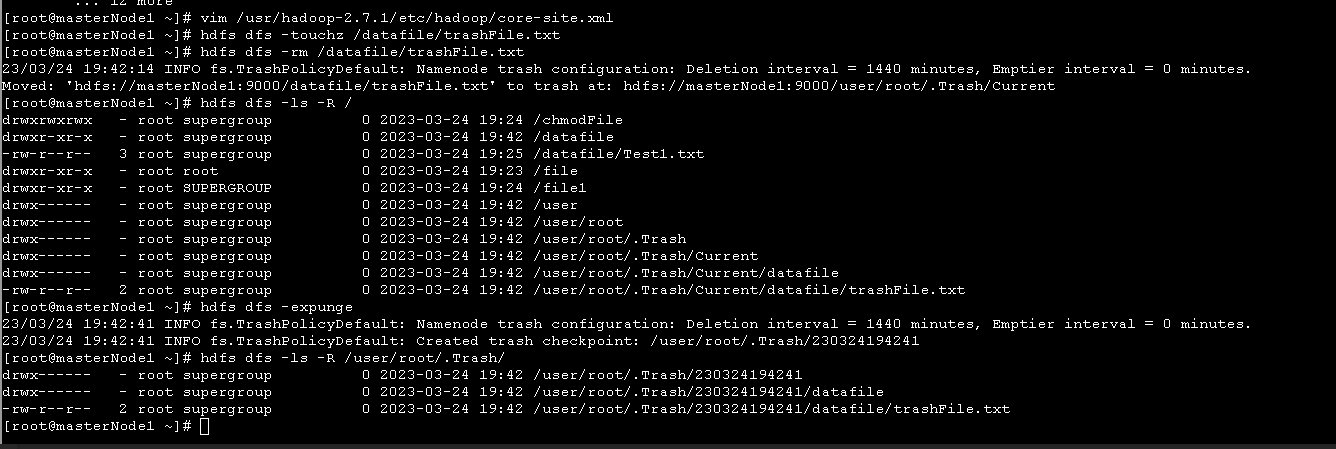
**实验说明**

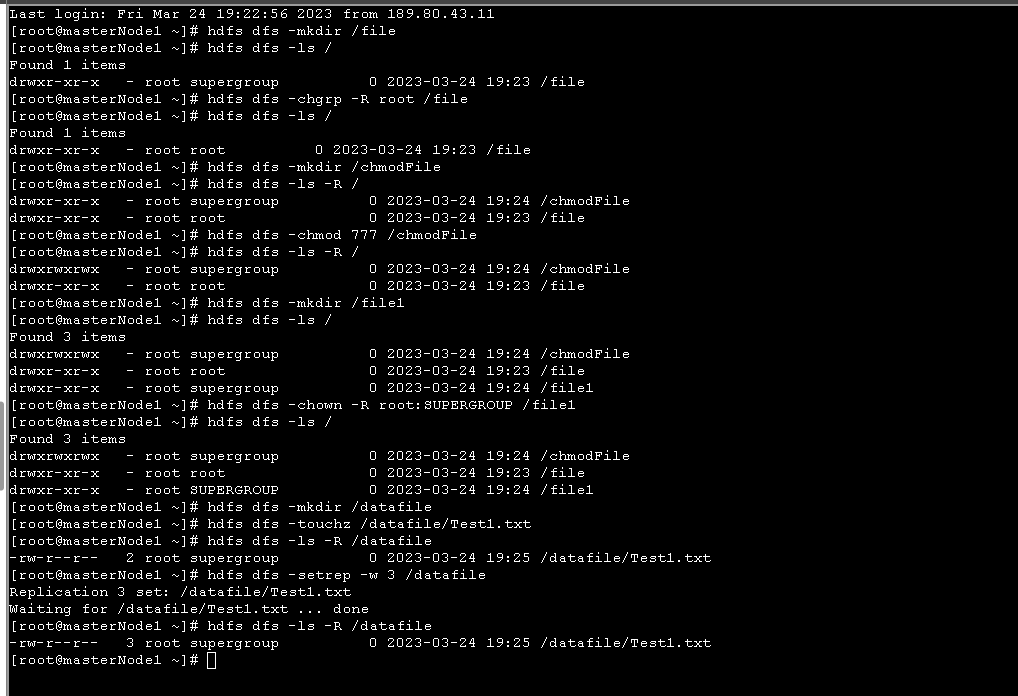
本次实验我们将在集群环境上进行，我们将对开启NameNode进程的节点（masterNode1）进行文件权限，文件所属组或用户的操作和集群管理命令、数据管理命令的验证。

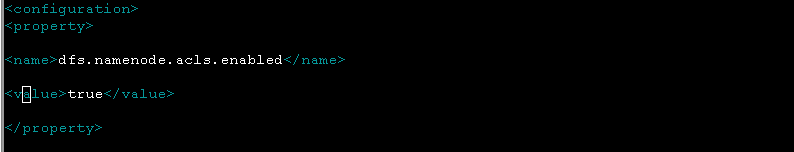
**3、HDFS高级操作命令**

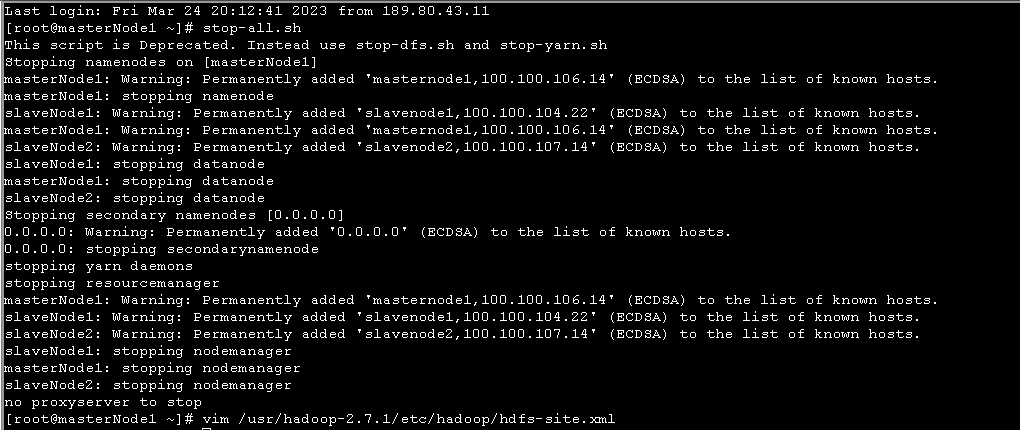
**管理集群**

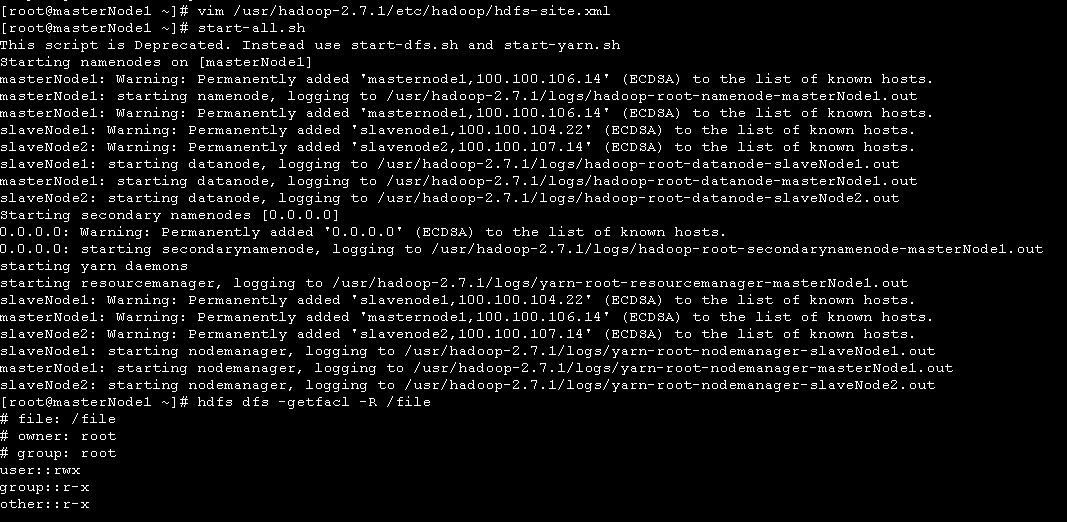
**管理帮助命令**

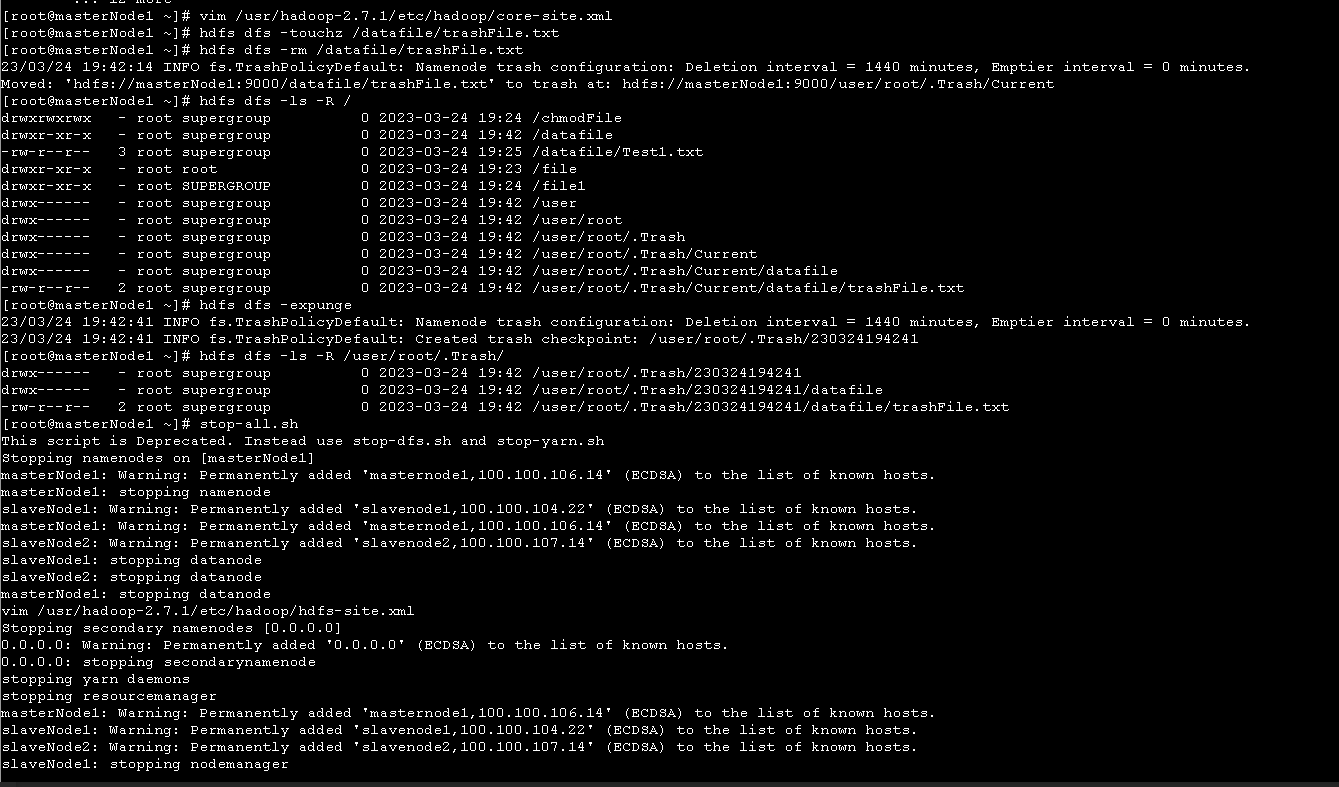


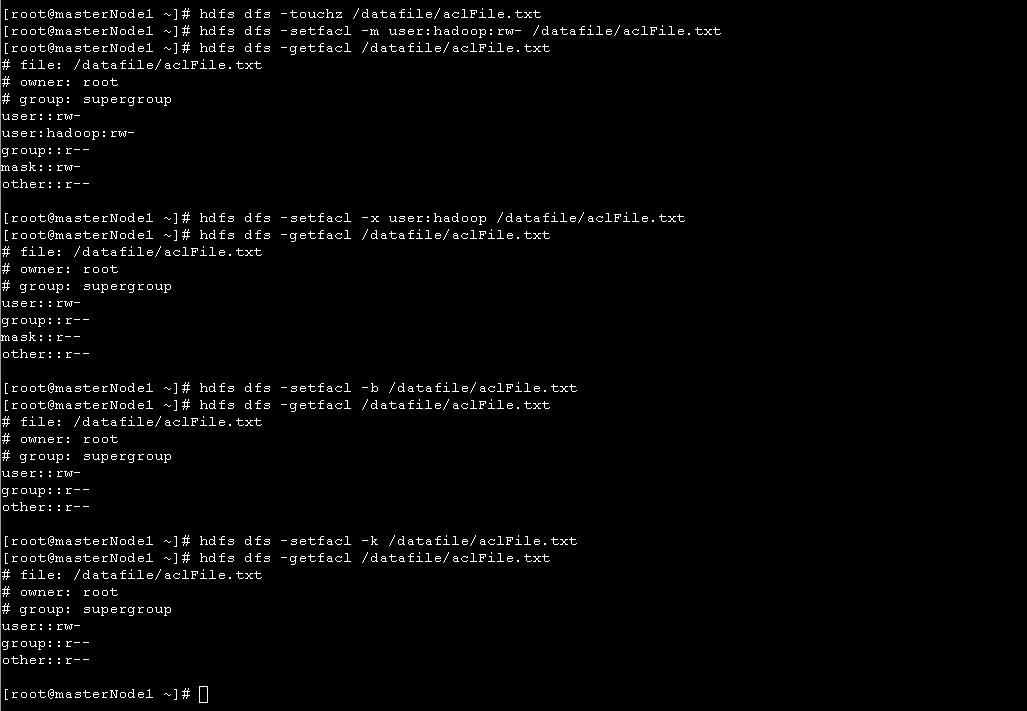




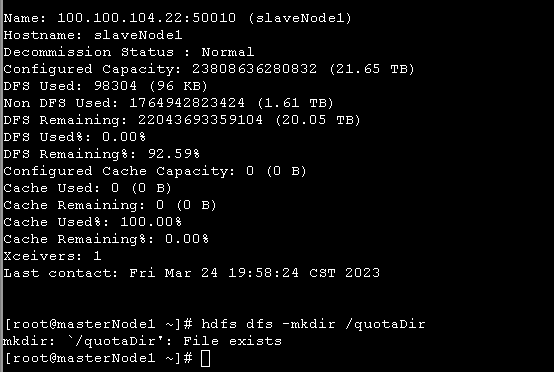


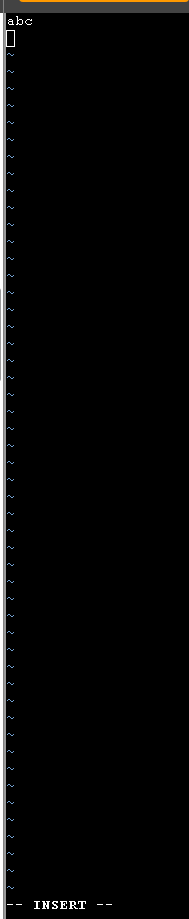


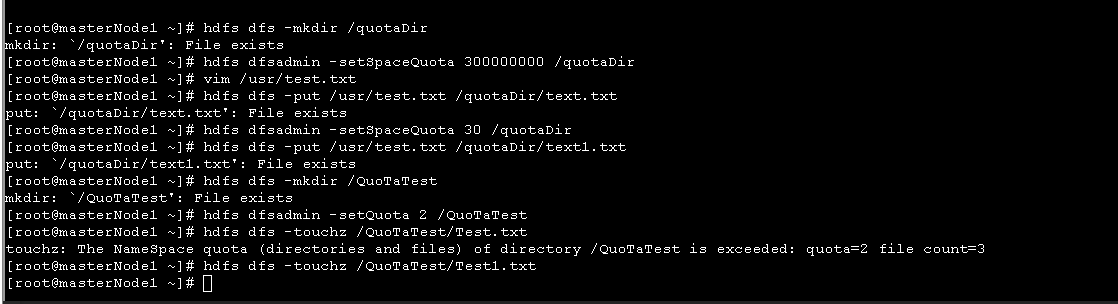


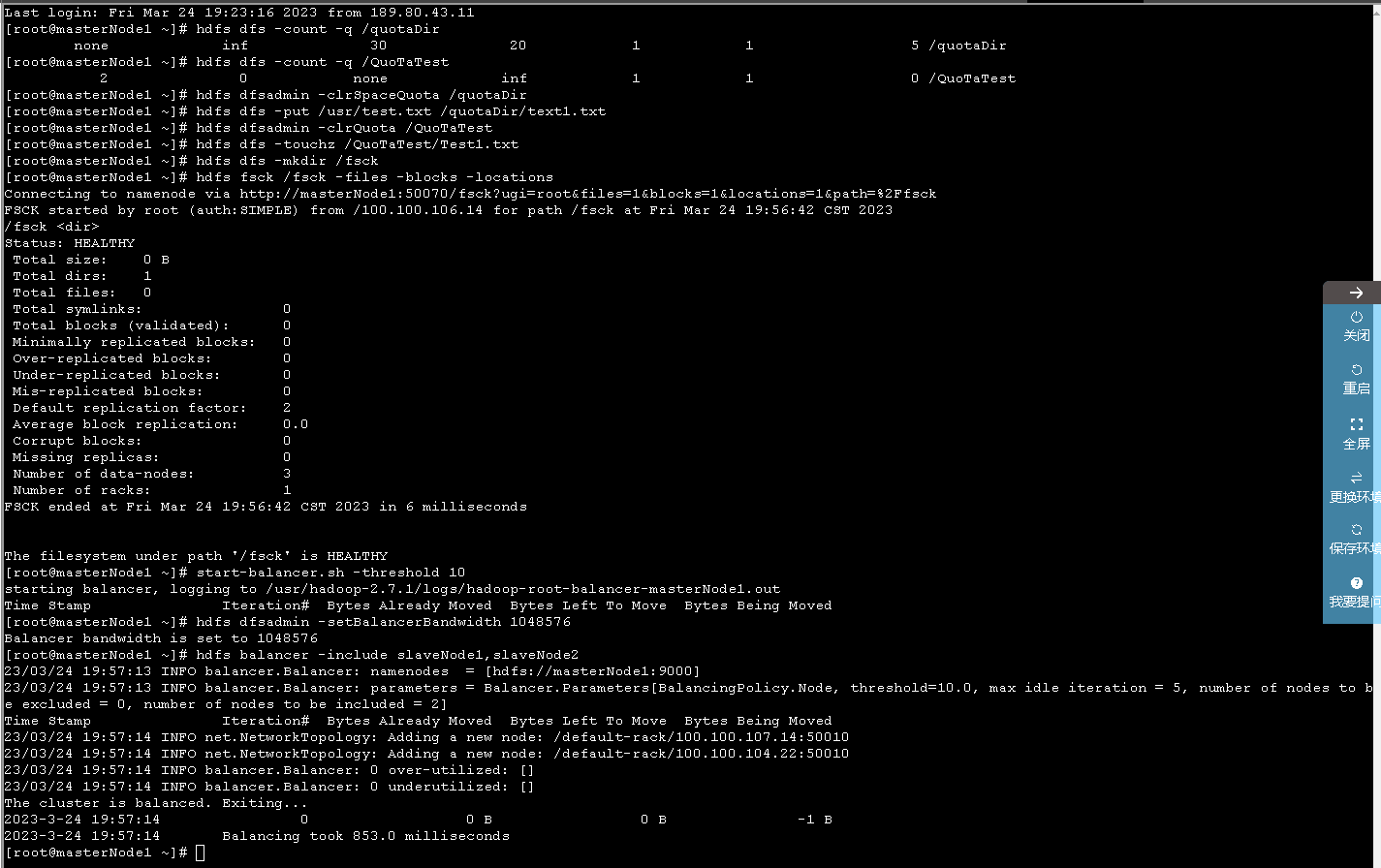






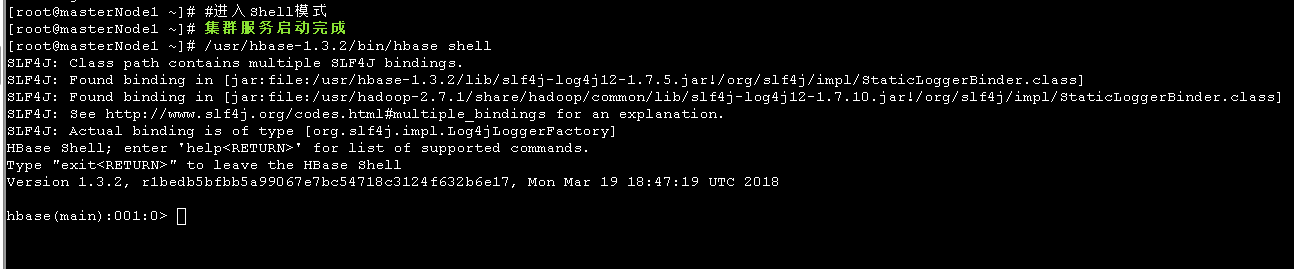




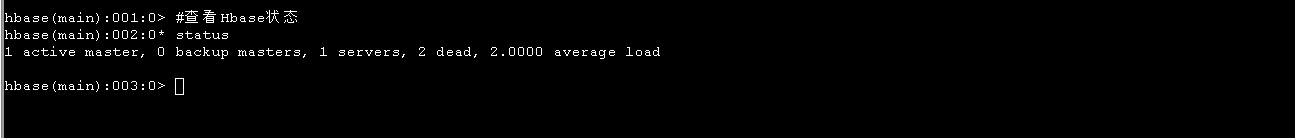


# 第二次作业

进入shell模式



查看HBase状态

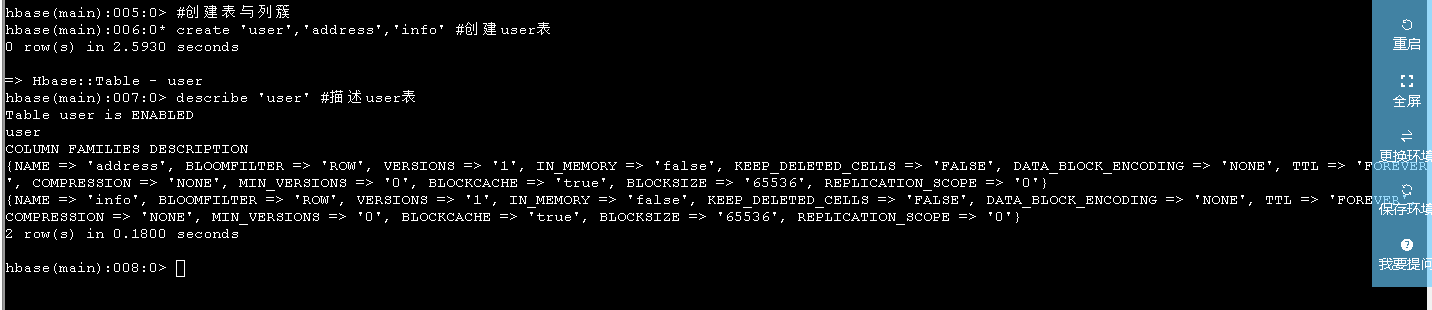
  
查看HBase 版本信息

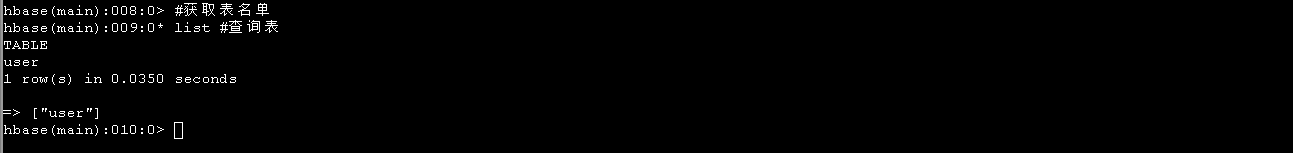
  
插入数据

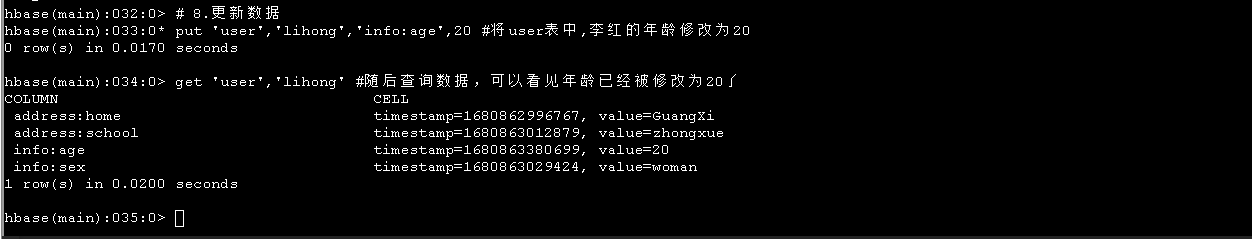
  
查询数据



创建表和列族

  
获取表名单

  
更新数据

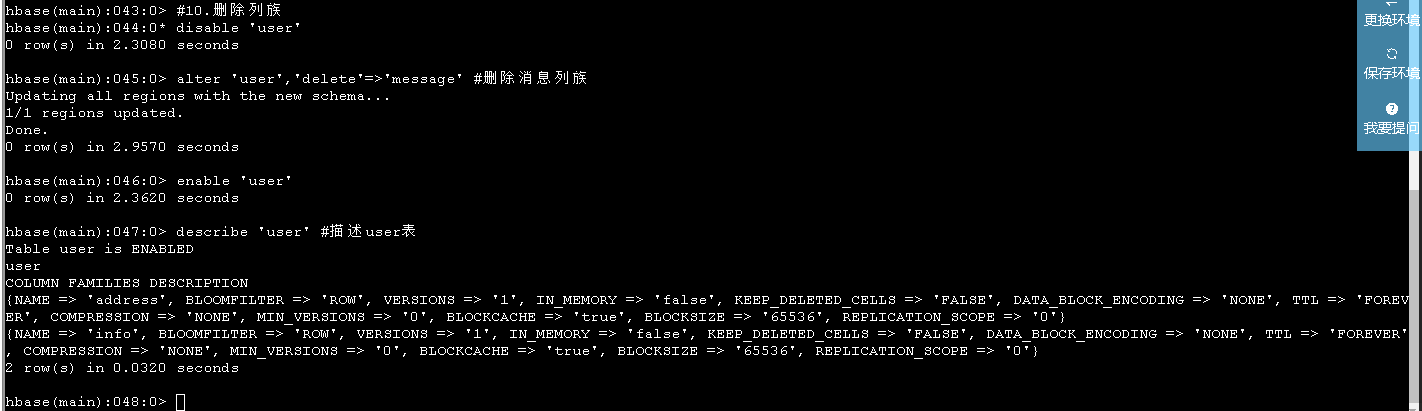


查询任何一列中包含“ x “的数据

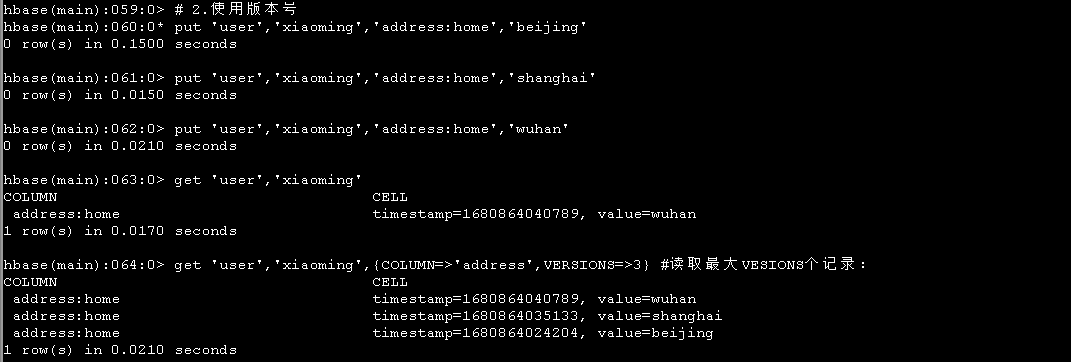
  
添加列族



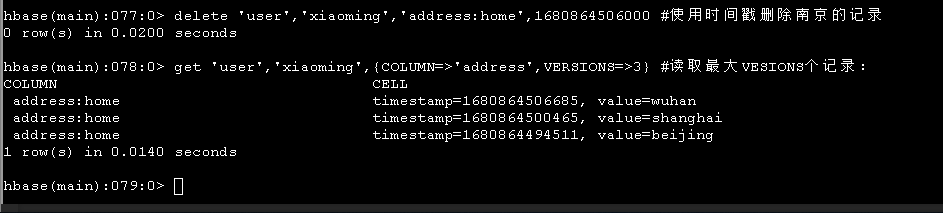
删除列族



查询使用版本号

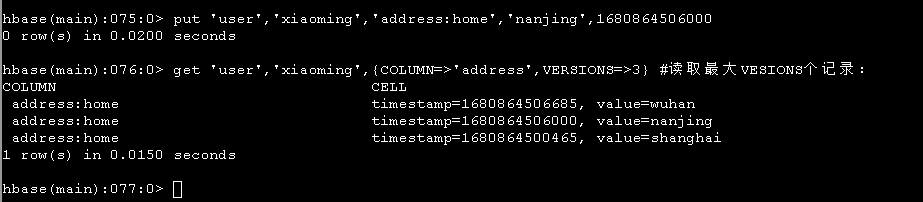


使用时间戳删除南京的记录

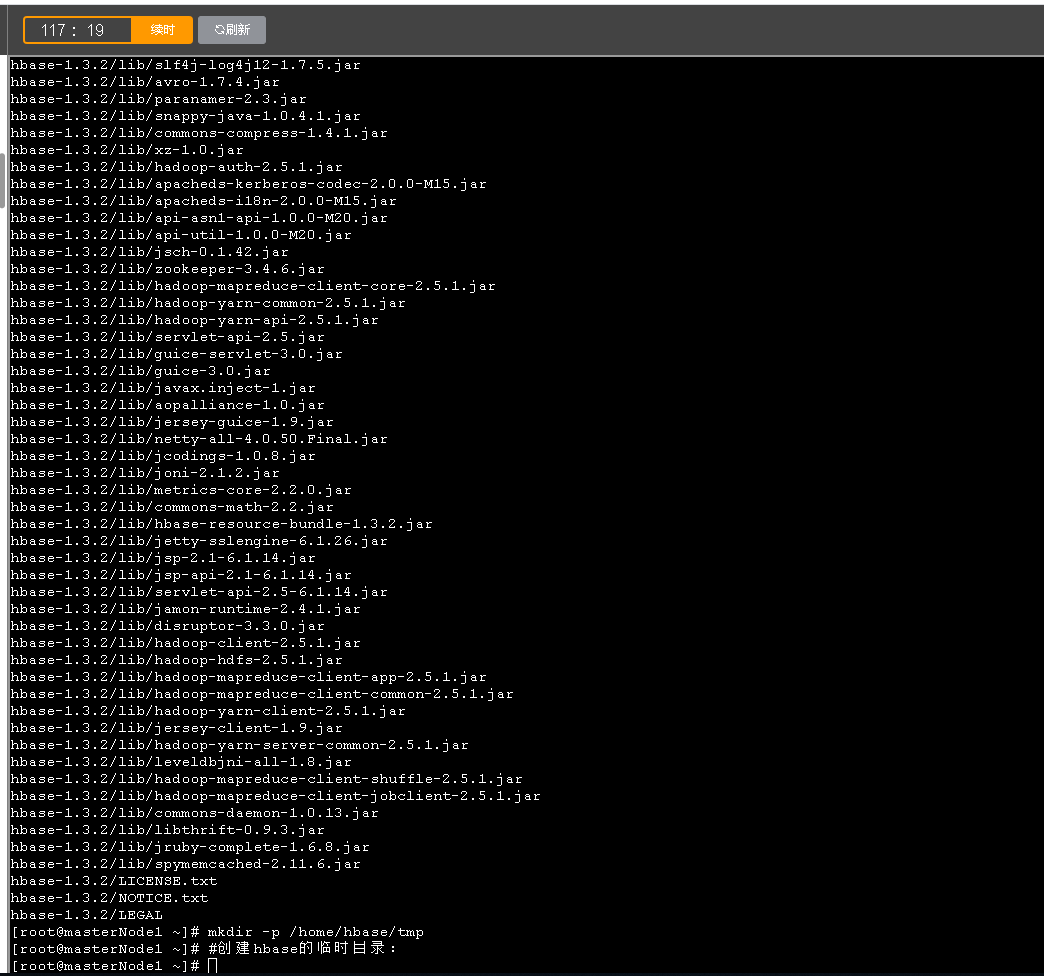
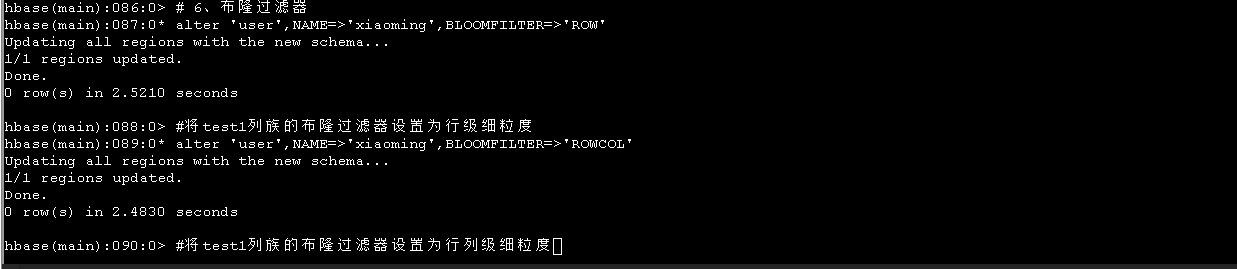
修改版本保存数



读取最大VESTONS 个记录



布隆过滤器



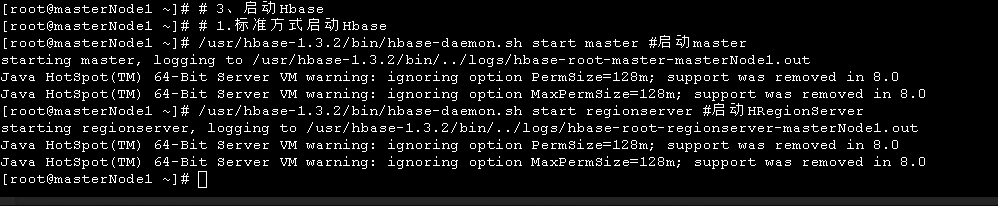
修改JKS 路径



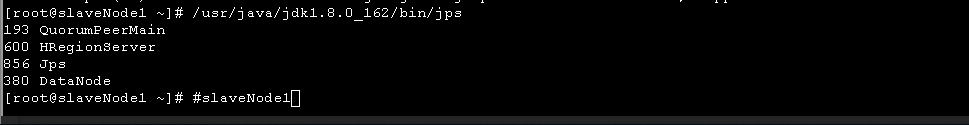
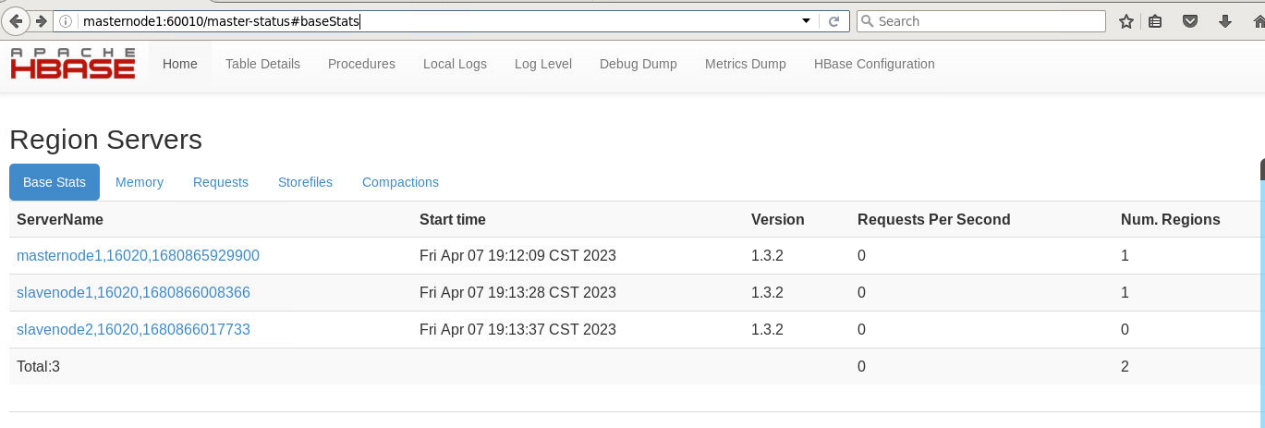
修改JAVE\_HOME 路径



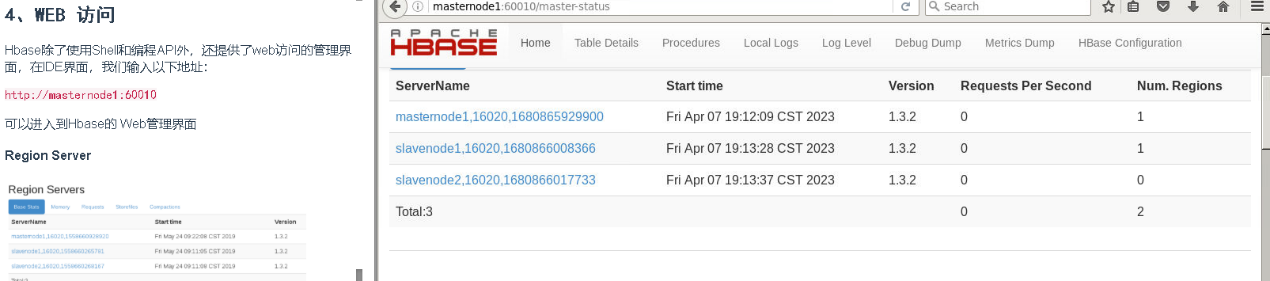
启动HBase 标准方式启动



配置HBase 集群 regineserver 信息



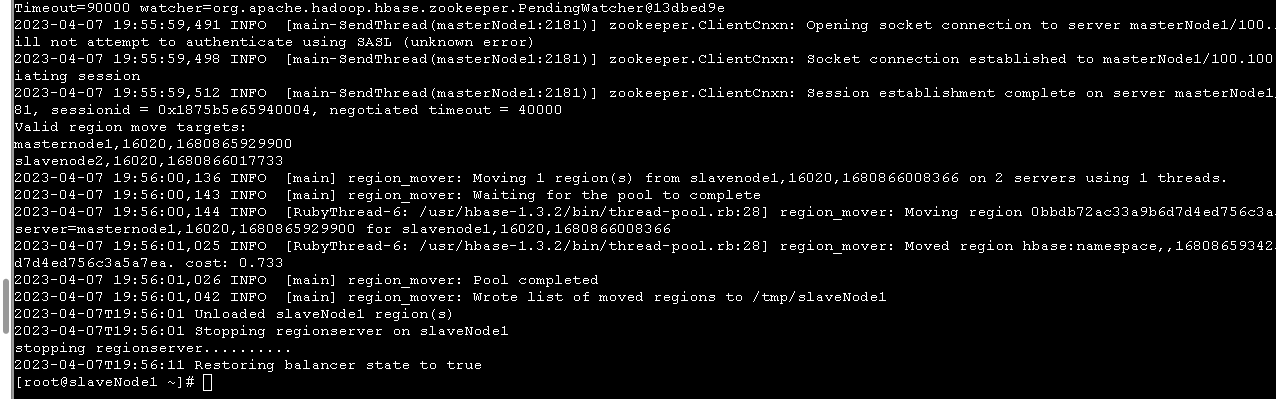
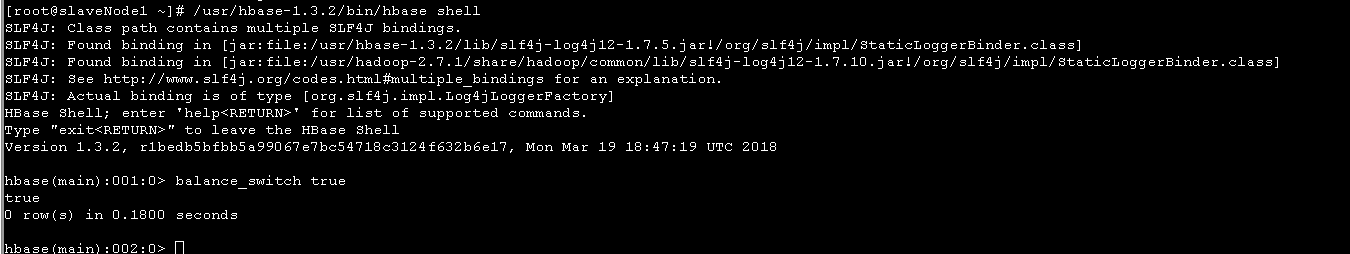
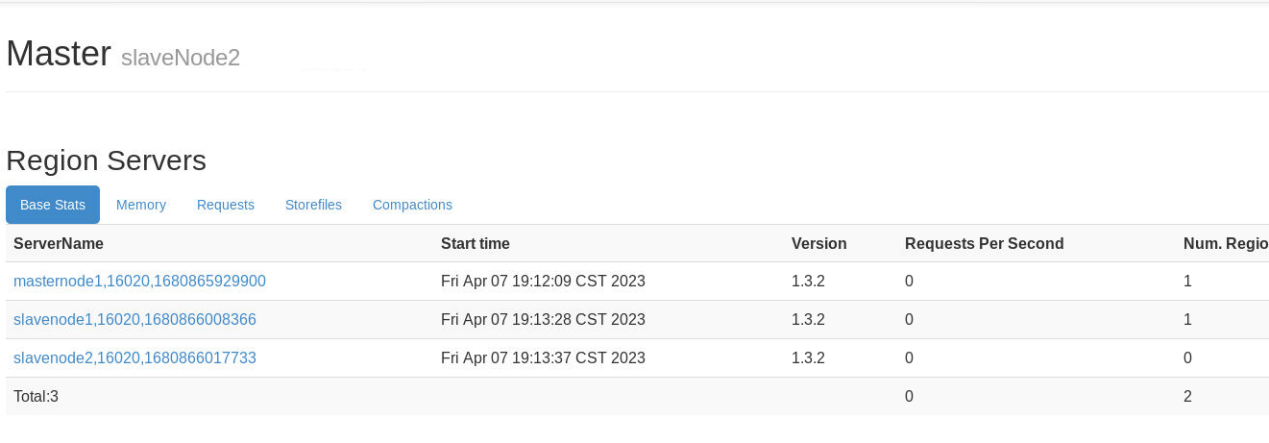
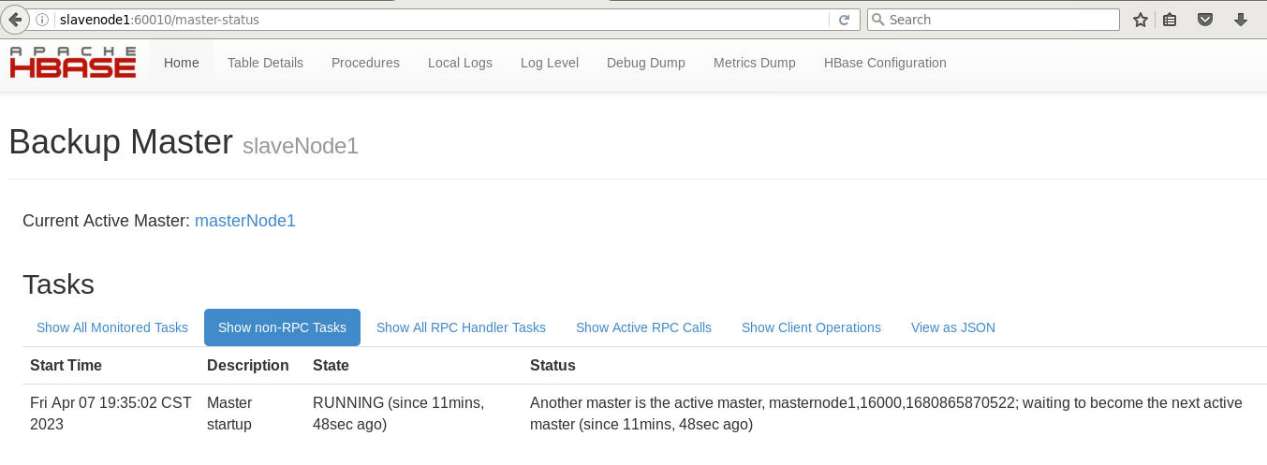
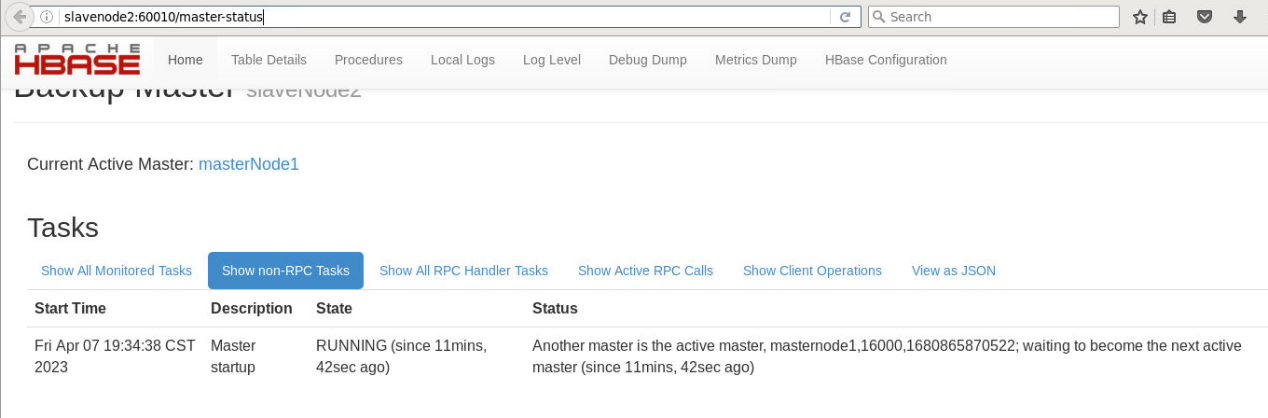
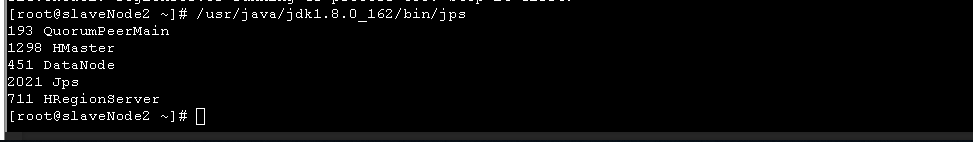
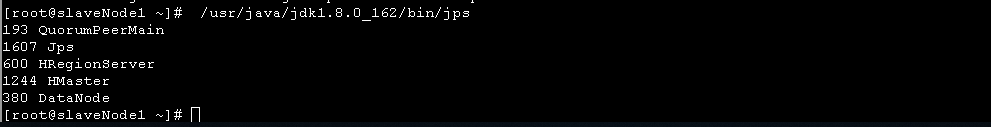
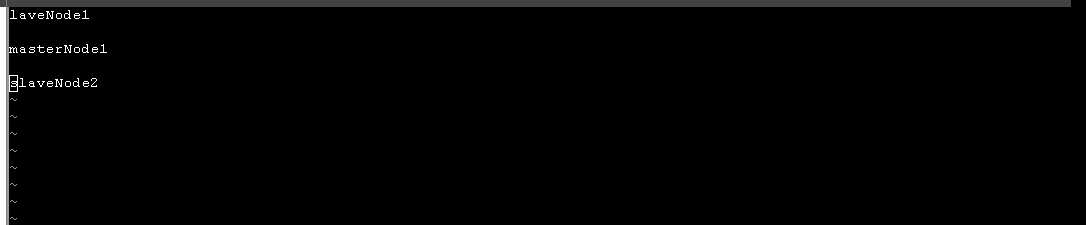
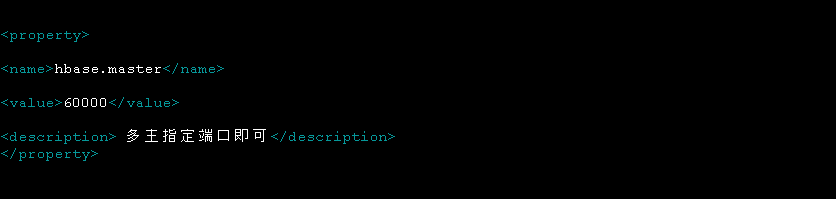
通过JPS 查看服务状态



修改核心配置文件



创建 backup – master 文件

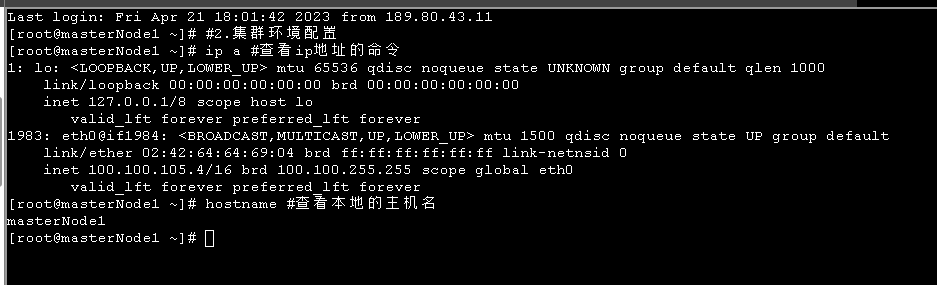


# 第三次作业

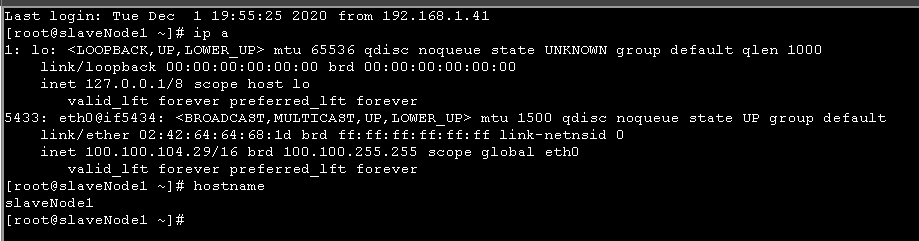
集群环境配置

查看ip 地址命令

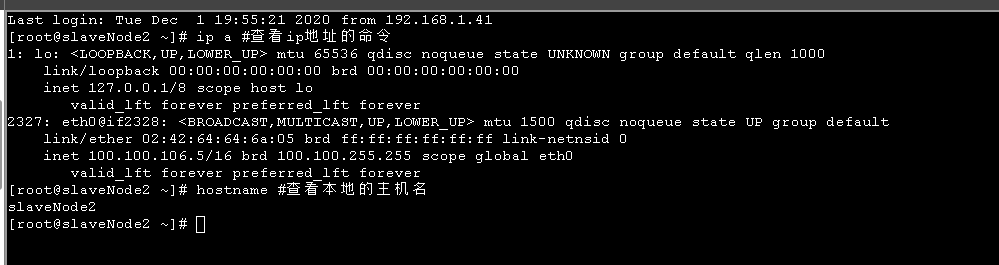
查看本地主机名



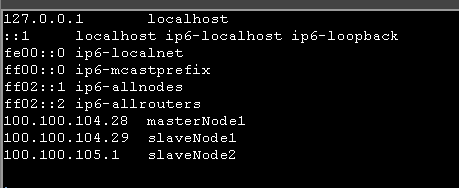
slaveNode1



slaveNode2

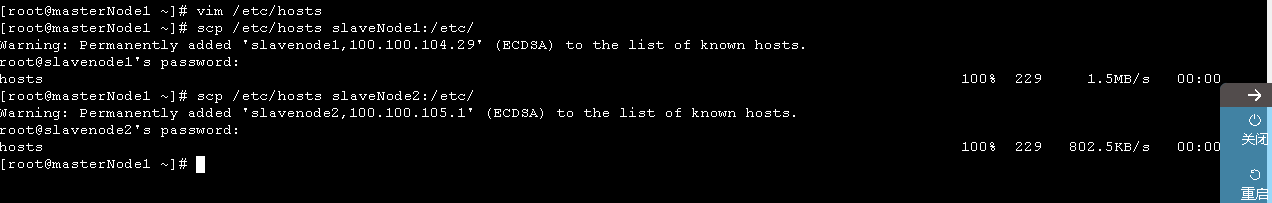


向/etc/hosts文件中添加IP和hostname，格式如下图所示，其中间使用制表符隔开。

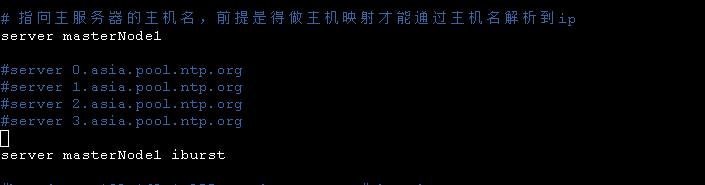
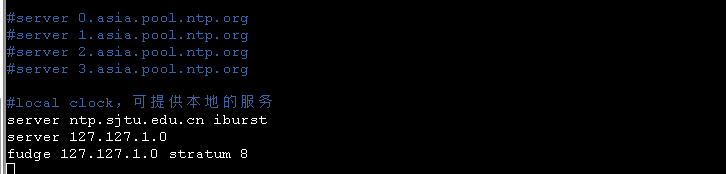


分发masterNode1节点的/etc/hosts文件到其余节点，此时我们可以使用其余节点的hostname进行分发，分发命令如下：

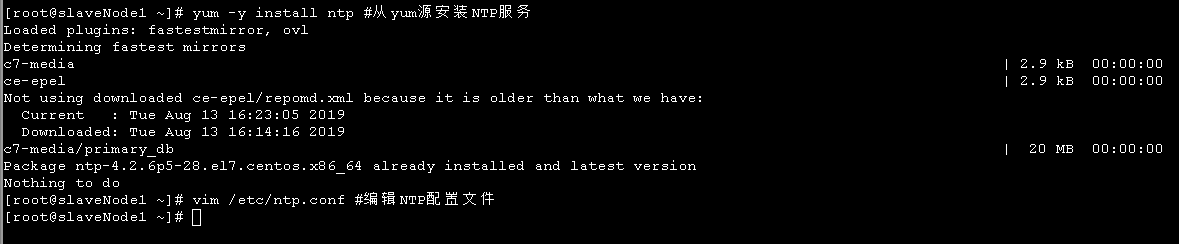
|  |
| --- |
| #在masterNode1执行以下命令  scp /etc/hosts slaveNode1:/etc/  #需要输入SSH连接密码（datastudio）  scp /etc/hosts slaveNode2:/etc/ |



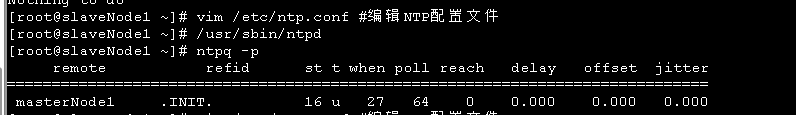
让masterNode1节点作为NTP服务器的服务端。注释掉原文件所有以‘server’开头的配置项



从yum 源安装 NTP服务

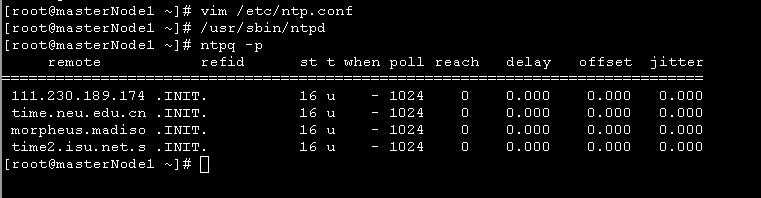


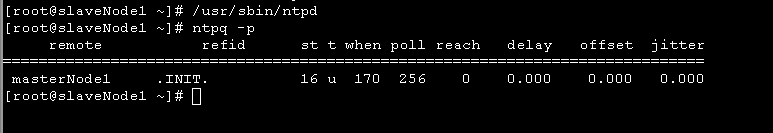
编辑NTP配置文件

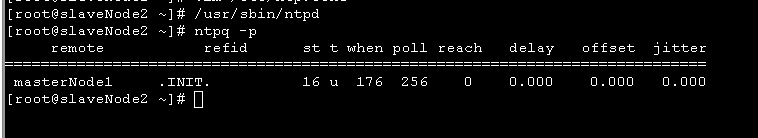


查看时间同步是否成功：

|  |
| --- |
| #在各节点执行  ntpq -p |







SSH免密登录



#在masterNode1,slaveNode1,slaveNode2执行

/usr/sbin/sshd

#生成ecdsa密钥

ssh-keygen -t ecdsa -f /etc/ssh/ssh\_host\_ecdsa\_key -N ""

#生成ed25519密钥

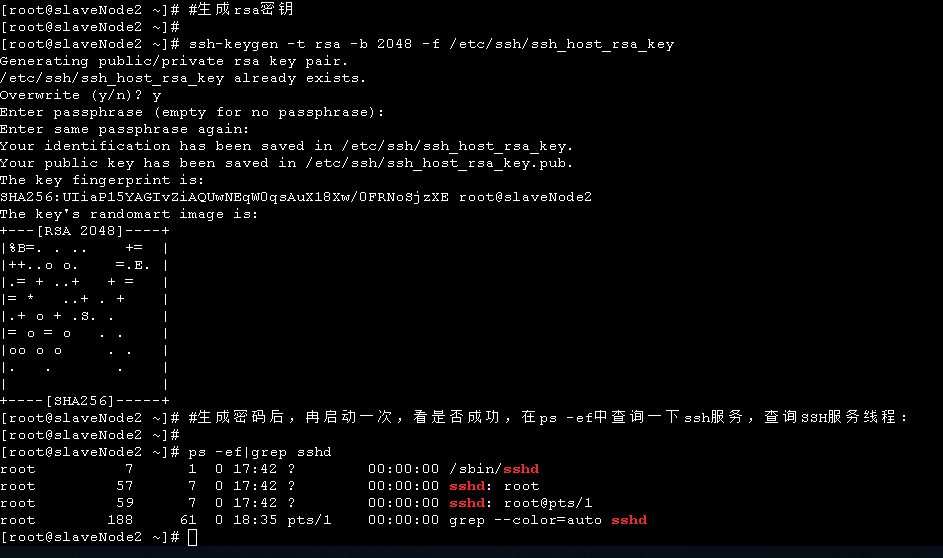
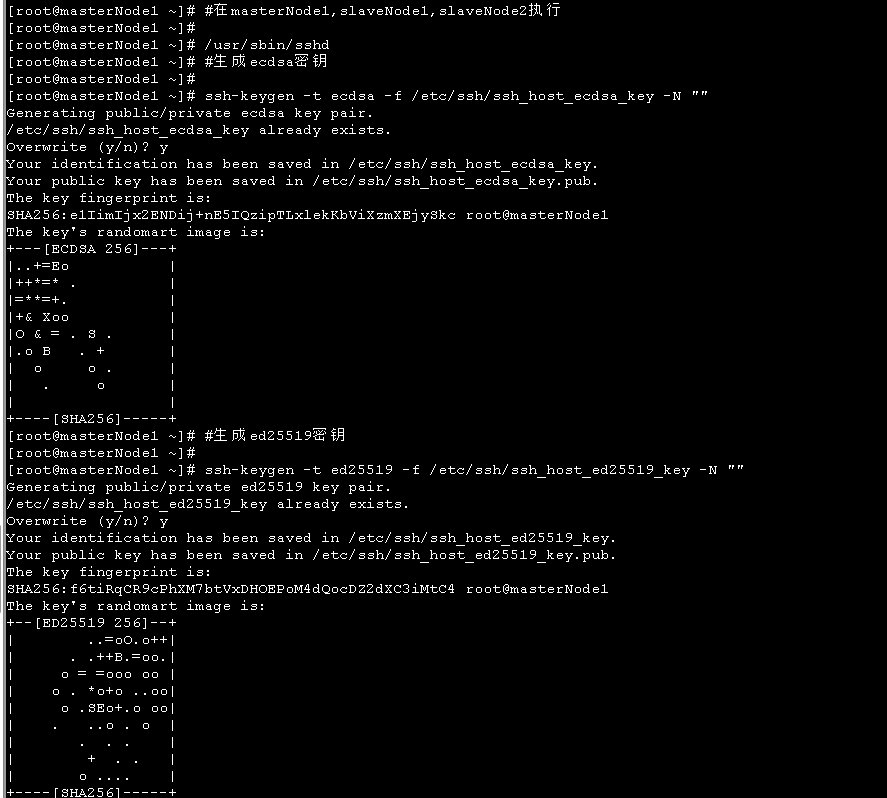
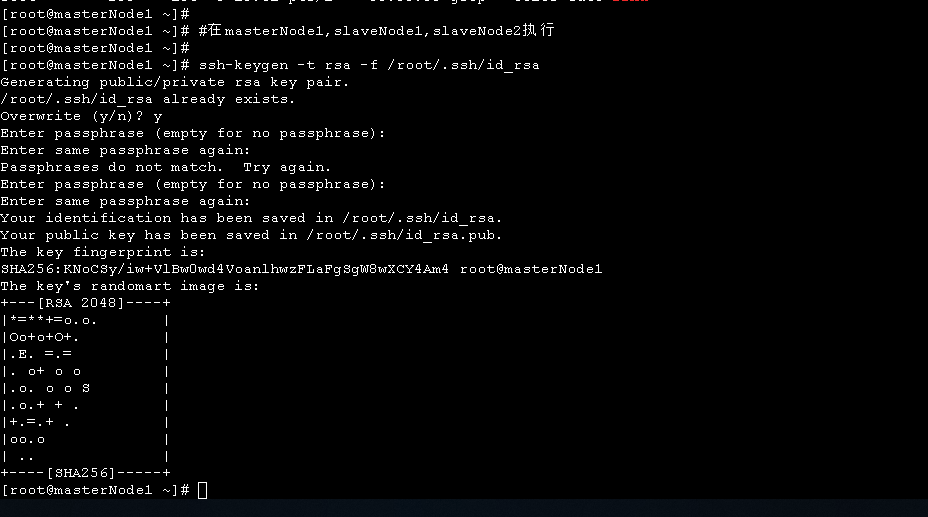
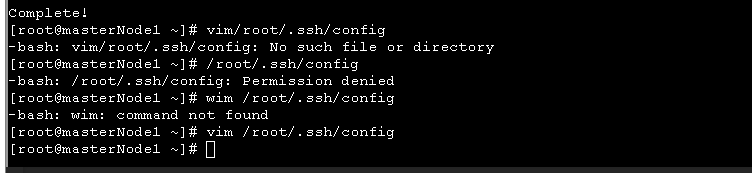
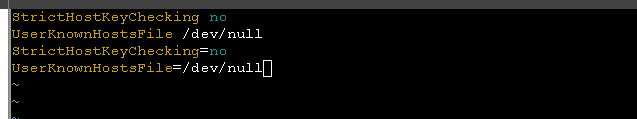
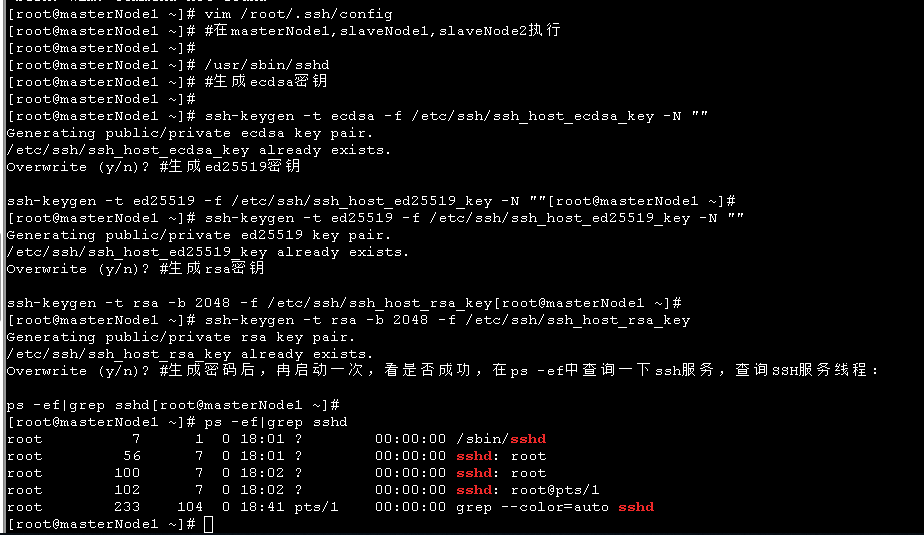
ssh-keygen -t ed25519 -f /etc/ssh/ssh\_host\_ed25519\_key -N ""

#生成rsa密钥

ssh-keygen -t rsa -b 2048 -f /etc/ssh/ssh\_host\_rsa\_key

#生成密码后，再启动一次，看是否成功，在ps -ef中查询一下ssh服务，查询SSH服务线程：

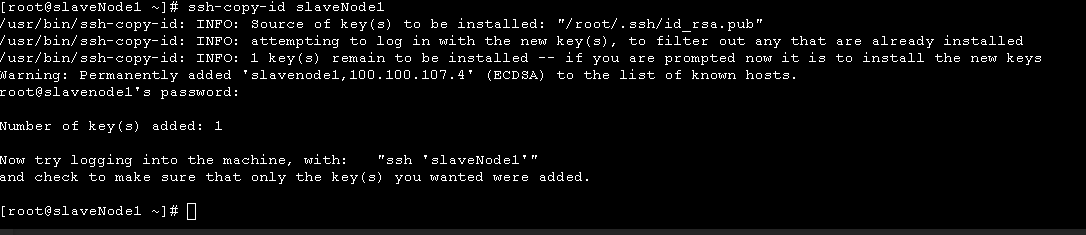
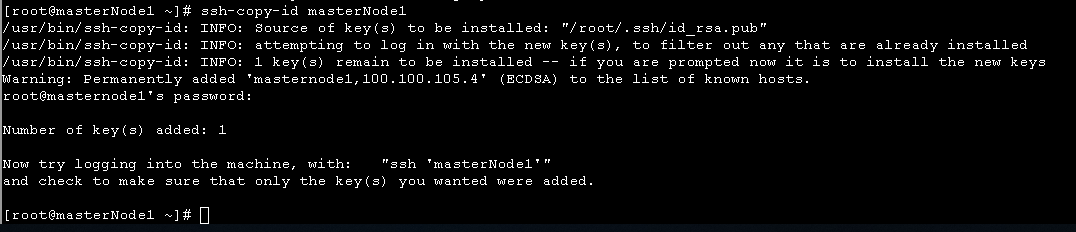
ps -ef|grep sshd

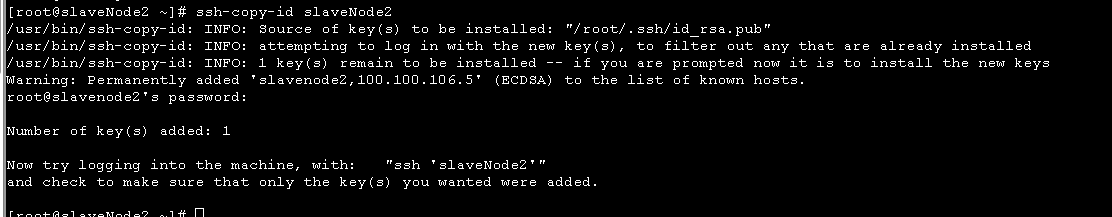


将证书发送到本机及其他机器上，需要输入登录密码datastudio

|  |
| --- |
| #在masterNode1,slaveNode1,slaveNode2执行  ssh-copy-id masterNode1  ssh-copy-id slaveNode1  ssh-copy-id slaveNode2 |

执行成功后，如图所示：

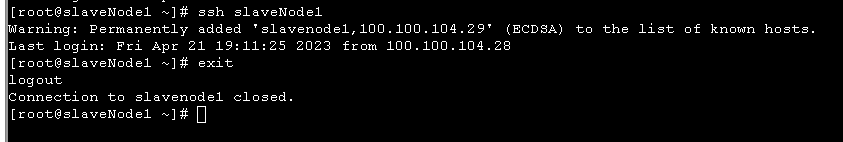


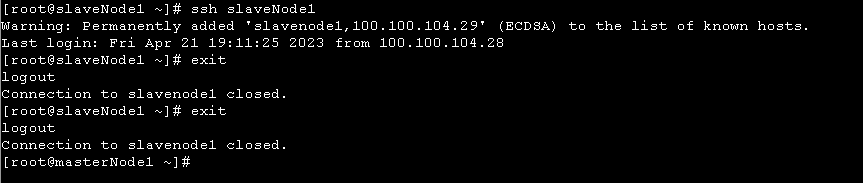


测试验证，使用在各节点中使用SSH命令分别进行登录测试，查看能否无密码登入，均可登录即设置成功，可使用exit命令退出。

|  |
| --- |
| #在masterNode1,slaveNode1,slaveNode2执行验证  ssh masterNode1  ssh slaveNode1  ssh slaveNode2 |

截图如下：

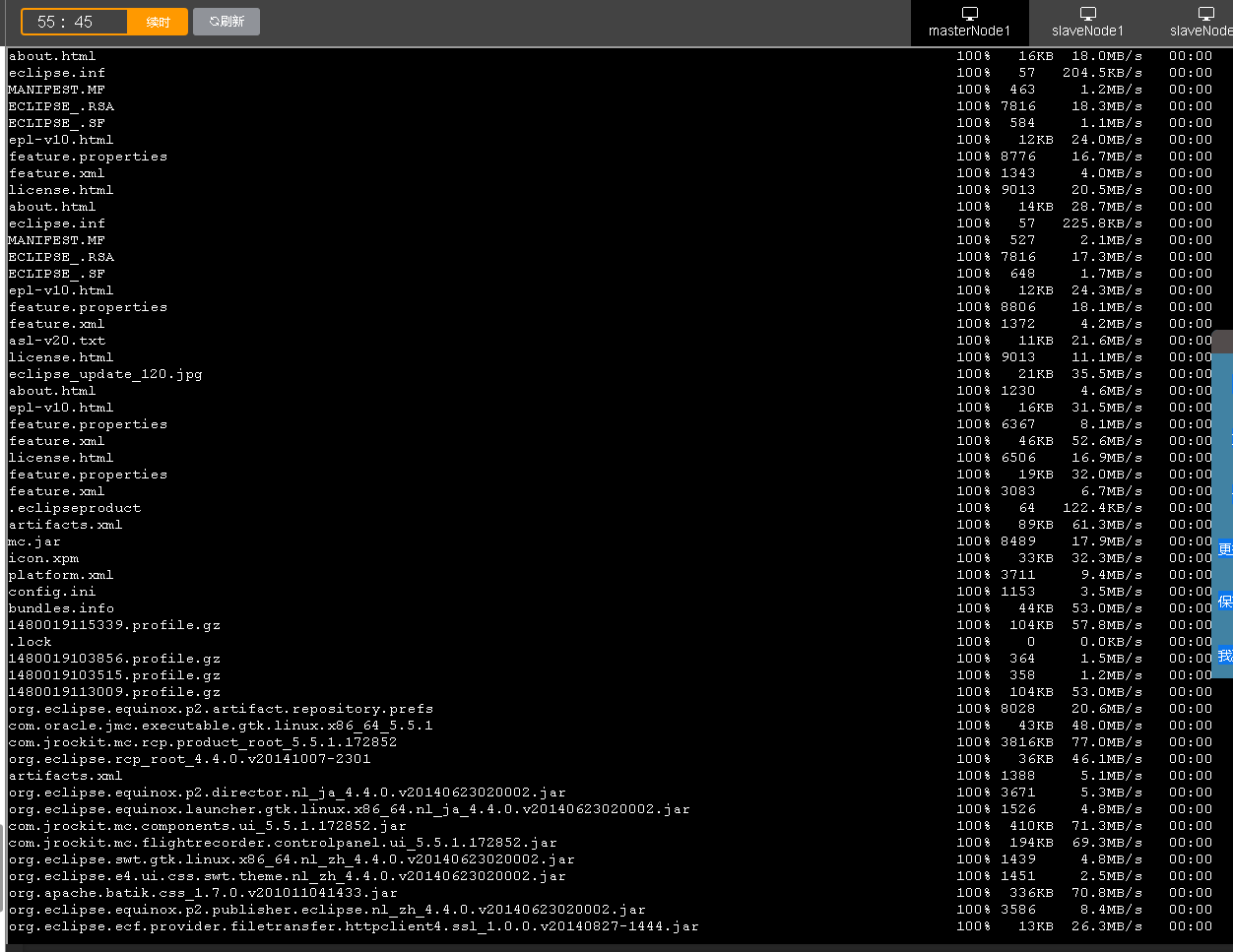


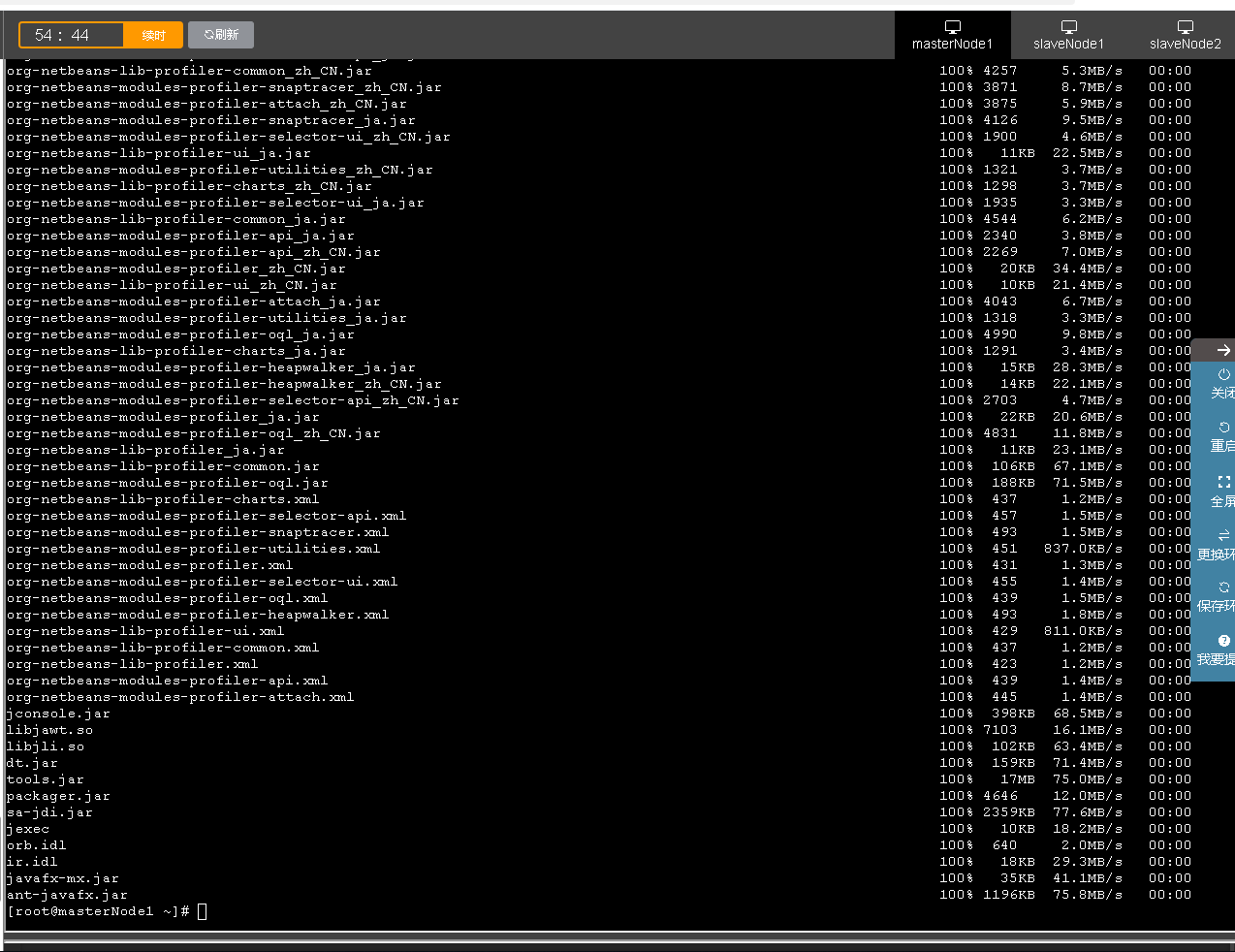


**安装JDK**

Hadoop程序是由Java编写，我们需要在运行Hadoop集群的所以节点上安装Java环境，才能使Hadoop集群可以正常运行。JDK安装包文件地址见资源环境表。集群中JDK安装流程如下：

|  |
| --- |
| #在masterNode1执行以下操作  #创建/usr/java目录，用于存放JDK文件  mkdir /usr/java  #将JDK解压至/usr/java目录中  tar -zxvf /mnt/aosa/java/jdk-8u162-linux-x64.tar.gz -C /usr/java/  #将/usr/java目录分发至其它节点（slaveNode1，slaveNode2）中  scp -r /usr/java/ slaveNode1:/usr/  scp -r /usr/java/ slaveNode2:/usr/ |

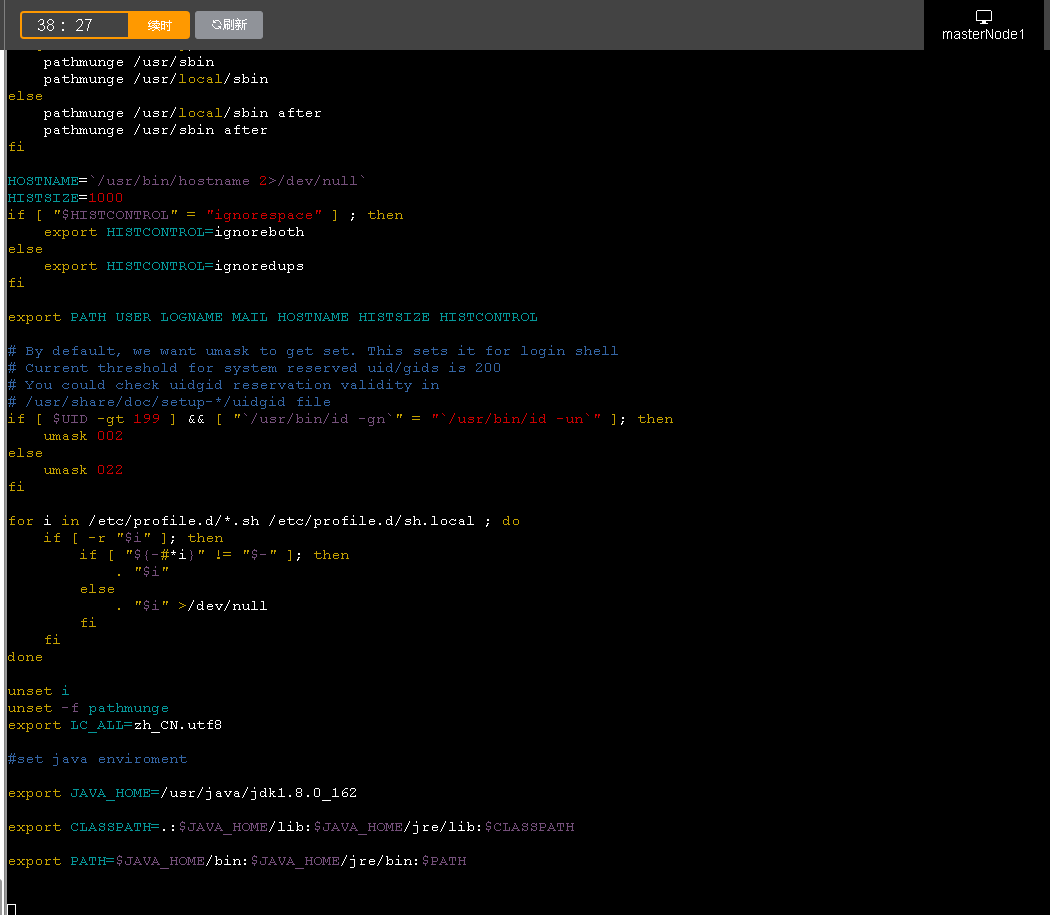




为了使用方便，我们可以为JDK添加环境变量，在/usr/profile文件末尾追加以下内容，注意JDK的环境路径。

|  |
| --- |
| #set java enviroment  export JAVA\_HOME=/usr/java/jdk1.8.0\_162  export CLASSPATH=.:$JAVA\_HOME/lib:$JAVA\_HOME/jre/lib:$CLASSPATH  export PATH=$JAVA\_HOME/bin:$JAVA\_HOME/jre/bin:$PATH |

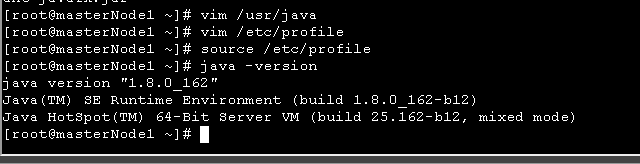
修改结果如下图：



编译环境变量后，查询Java版本信息，验证配置是否正确：

|  |
| --- |
| #在各节点执行  source /etc/profile  java -version |

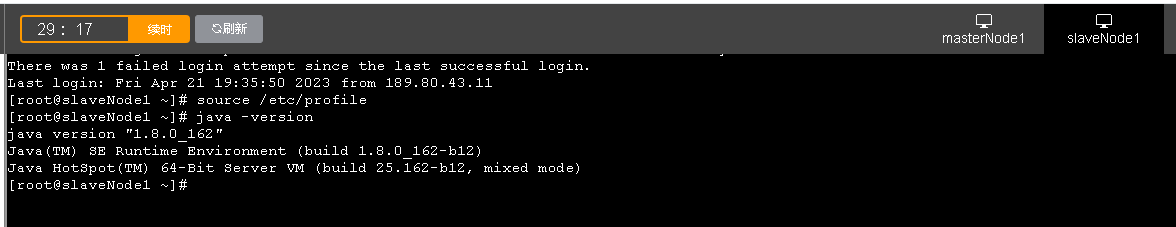
Masternode1



Slavenode1:

刚开始没把为JDK添加环境变量这一步在s1,s2运行导致了错误





Slavenode2:

