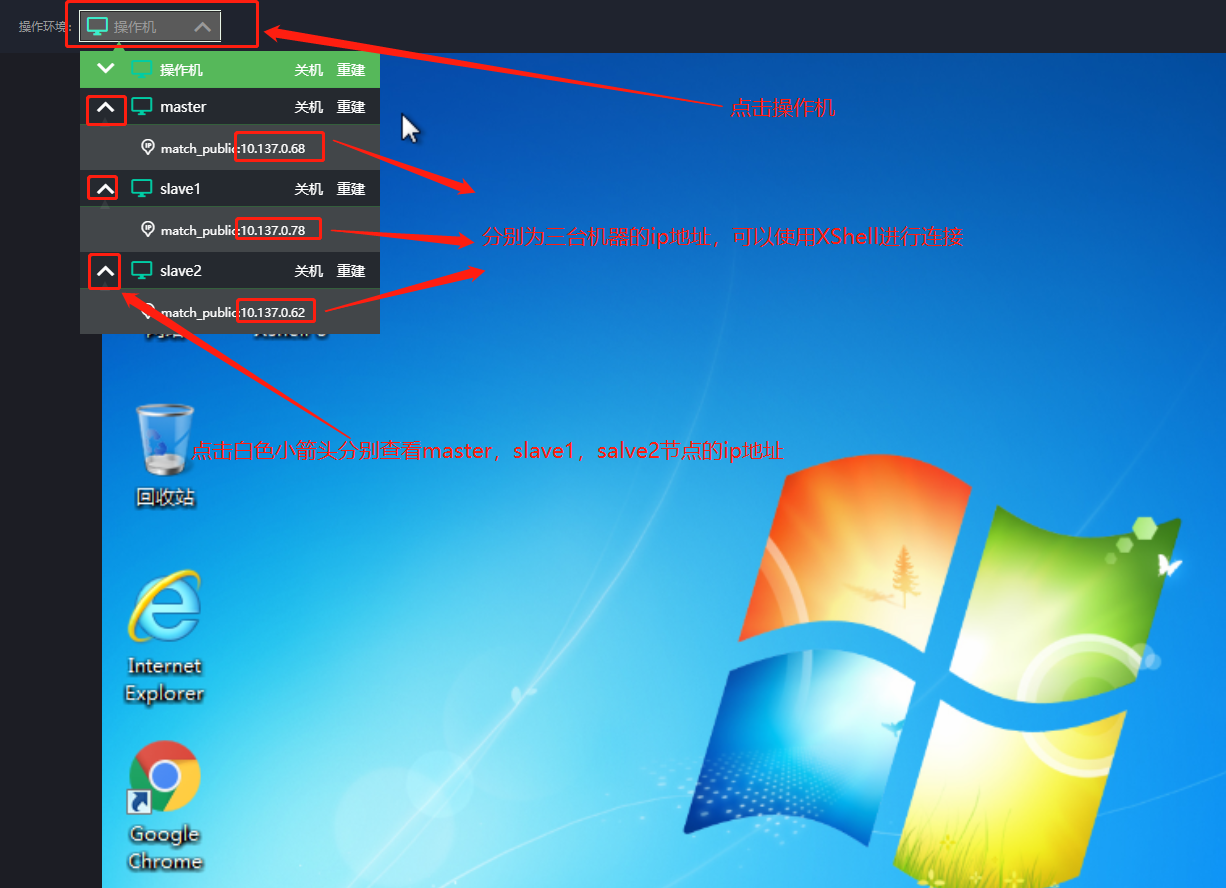
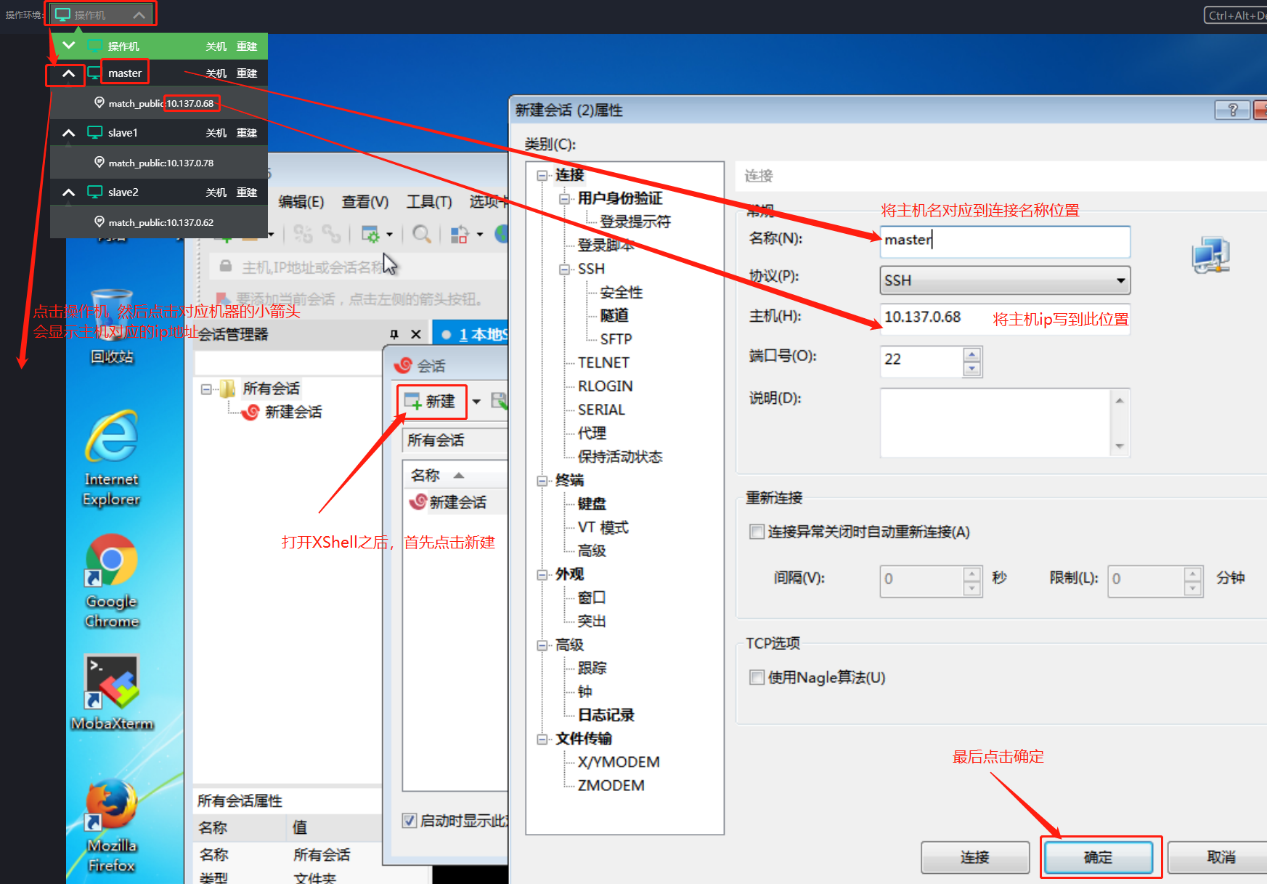
**安装hadoop**

在操作机上使用Xshell连接虚拟机：

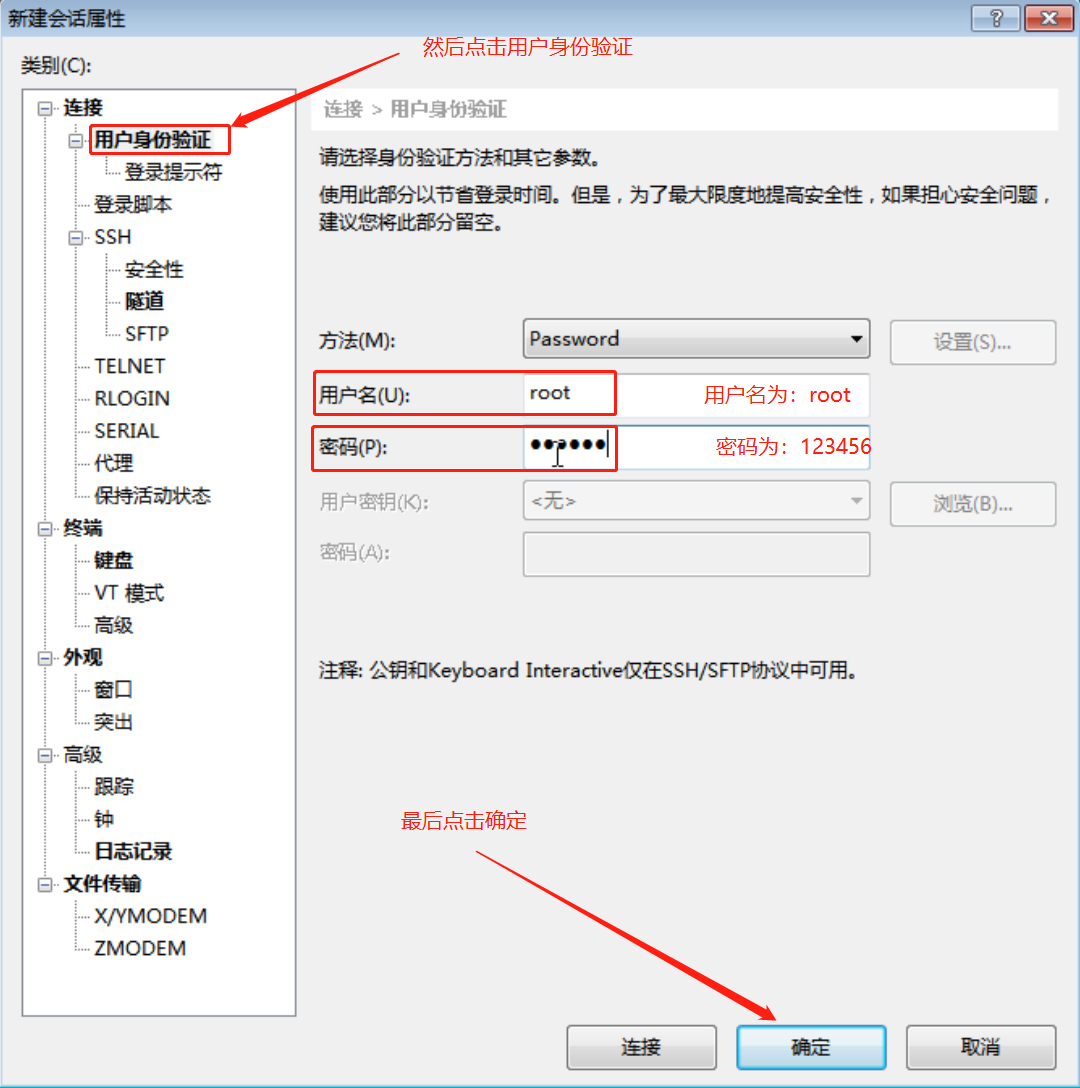
1. 可以通过平台查看所有虚拟机的ip地址，如下图所示：



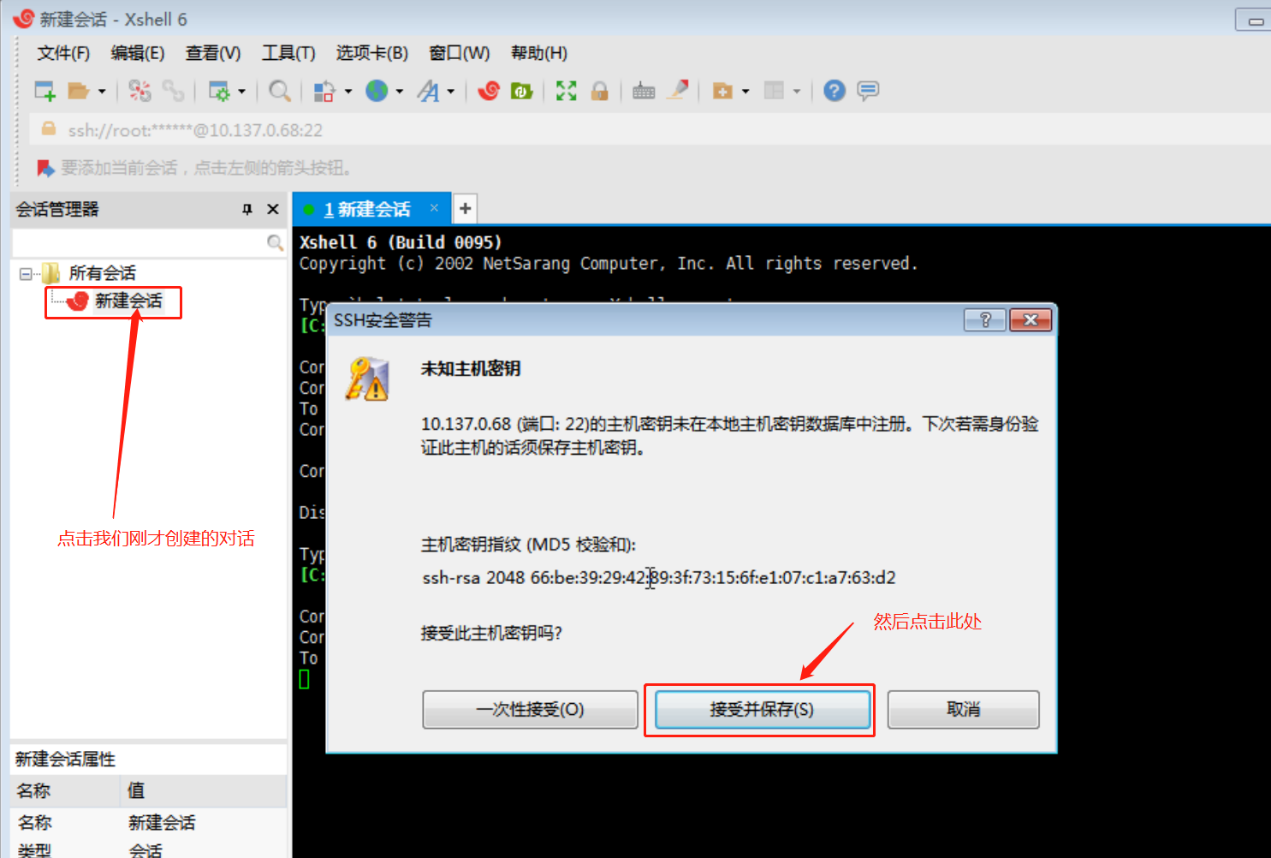
2. 打开Xshell将对应的主机名及ip地址输入到对应的输入框内，如下图所示：



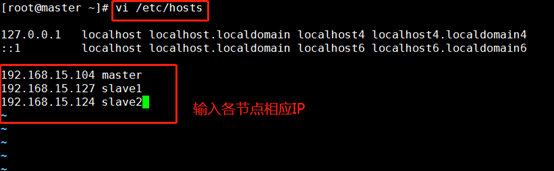
3. 点击用户身份验证将用户名和密码输入到对应的输入框内，如下图所示：



4. 双击我们创建的连接，最后点击接受并保存，如下图所示：



5. 修改 /etc/hosts 文件（三台机器都需要操作），操作过程如下图所示：



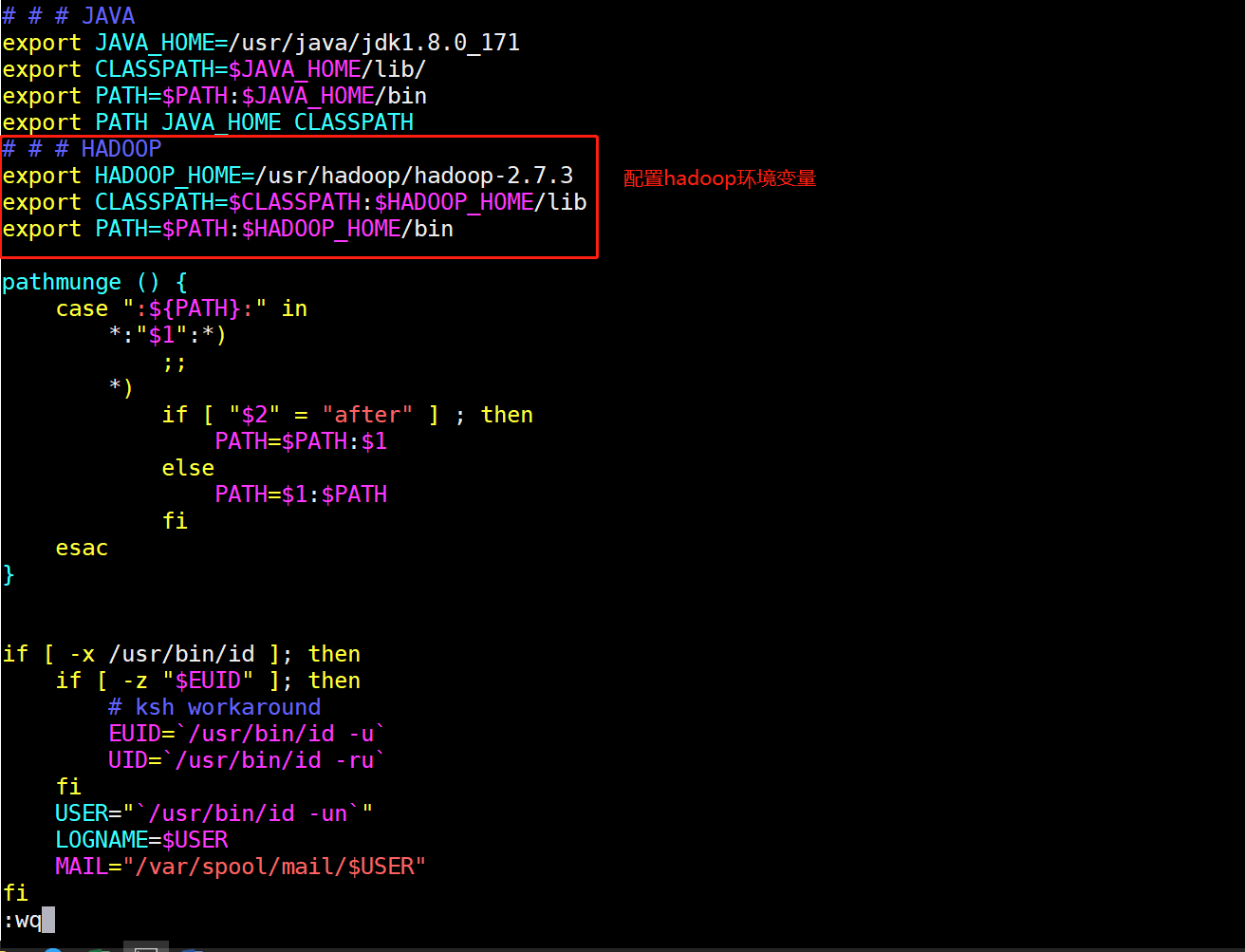
注意：图中的ip地址为操作是使用的机器的ip地址，需要将hosts文件中的ip地址更换为我们将第一步所得到的ip地址！！！  
  
6.当所有机器上的hosts文件配置完成之后，我们可以在master节点上输入 ssh slave1或者 ssh slave2 ，测试一下hosts文件是配置成功。  
  
7.创建对应工作目录/usr/hadoop（在master节点上操作）



8.解压hadoop到相应目录：  
我们已经把hadoop的安装包放在了主节点的 /opt/soft 目录下，可以通过tar命令将其复制到 我们刚才创建的hadoop工作目录中：

* mkdir -p /usr/hadoop
* tar -zxvf /opt/soft/hadoop-2.7.3.tar.gz -C /usr/hadoop/

解压后：



**配置环境变量**

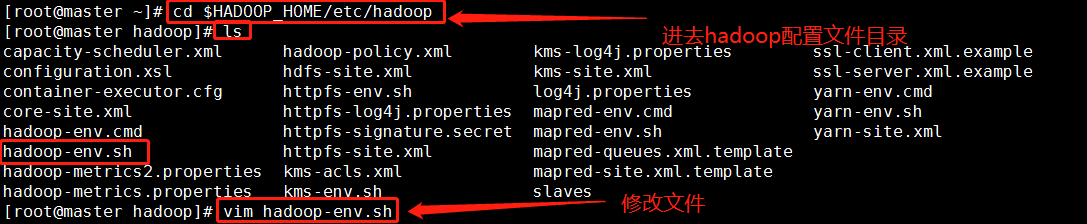
　　1.配置环境变量

* vim /etc/profile

export HADOOP\_HOME=/usr/hadoop/hadoop-2.7.3

export CLASSPATH=$CLASSPATH:$HADOOP\_HOME/lib

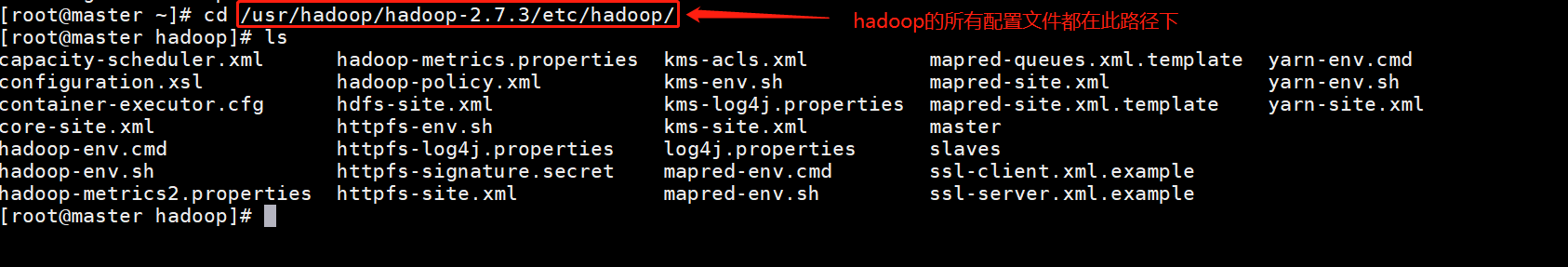
export PATH=$PATH:$HADOOP\_HOME/bin



　　使用以下命令使profile生效：

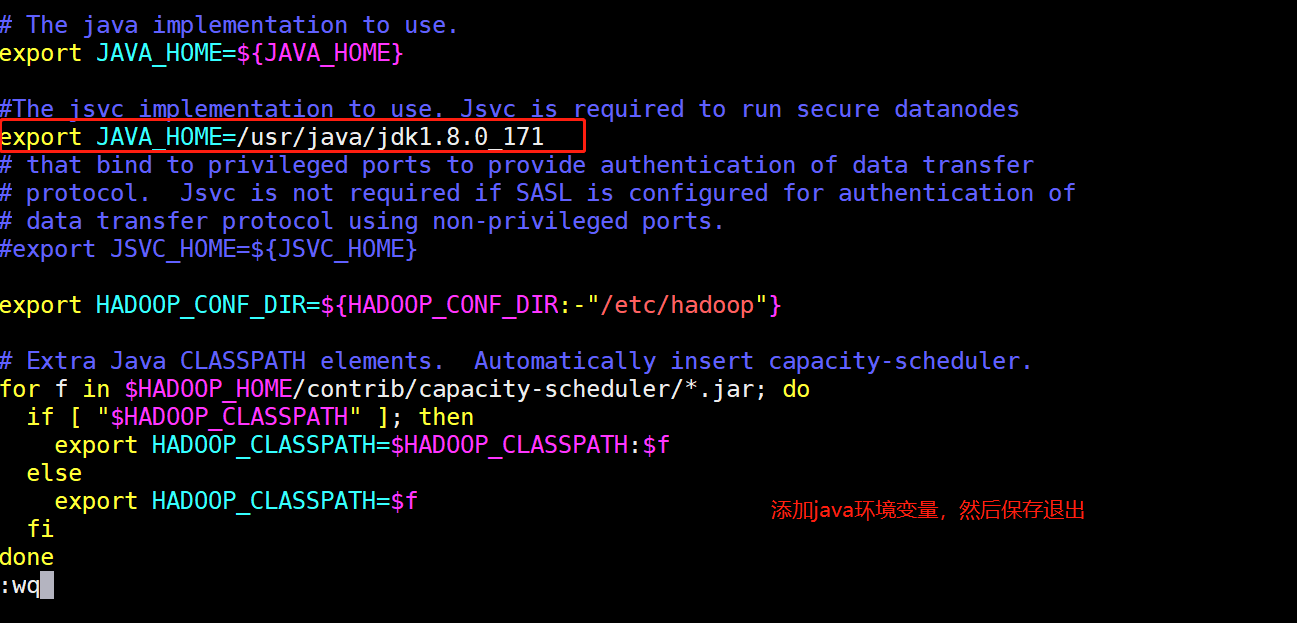
* source /etc/profile

　　2. 接下来我们需要对hadoop进行一些配置，所有的hadoop的配置文件都在 /usr/hadoop/hadoop-2.7.3/etc/hadoop中，所有我们需要使用 cd /usr/hadoop/hadoop-2.7.3/etc/hadoop命令进入到hadoop配置文件所在的目录中，如下图所示：



　　3. 使用vim命令编辑hadoop-env.sh文件：

* vim hadoop-env.sh



在配置文件中输入内容：export JAVA\_HOME=/usr/java/jdk1.8.0\_171  
输入完成后保存并退出

　　4.使用vim命令编辑core-site.xml文件，并在配置文件中加入以下内容：

* vim core-site.xml

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://master:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/usr/hadoop/hadoop-2.7.3/hdfs/tmp</value>

<description>A base for other temporary directories.</description>

</property>

<property>

<name>io.file.buffer.size</name>

<value>131072</value>

</property>

<property>

<name>fs.checkpoint.period</name>

<value>60</value>

</property>

<property>

<name>fs.checkpoint.size</name>

<value>67108864</value>

</property>

</configuration>



　　5.使用vim命令编辑yarn-site.xml文件，并在配置文件中加入以下内容：

<configuration>

<property>

<name>yarn.resourcemanager.address</name>

<value>master:18040</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>master:18030</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>master:18088</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>master:18025</value>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>master:18141</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

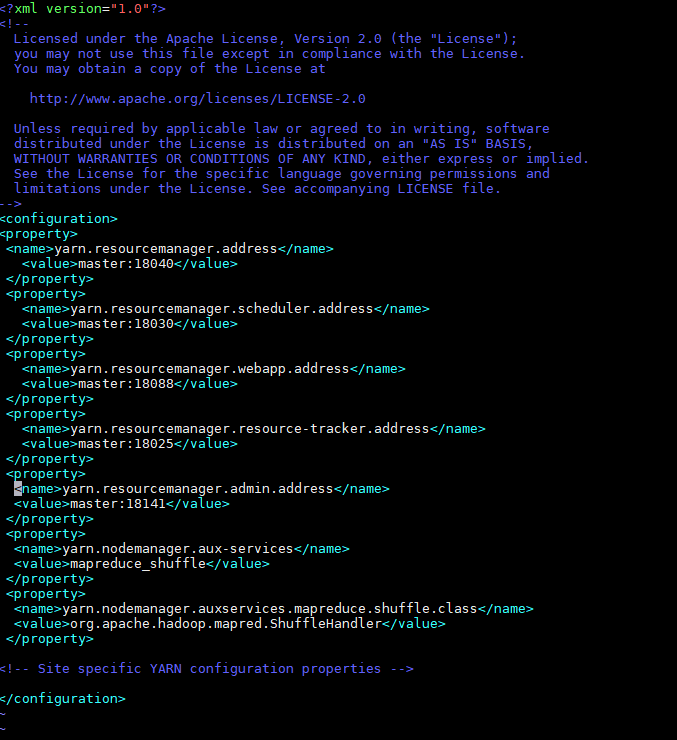
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

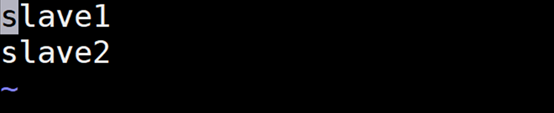
</property>

<!-- Site specific YARN configuration properties -->

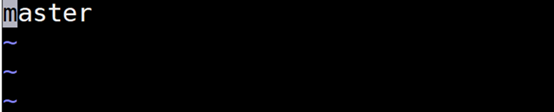
</configuration>



　　5.使用vim命令编写slaves文件，并在其中加入以下内容：



　　6.使用vim命令编写master文件，并在其中加入以下内容：



　　7. 使用vim命令编辑hdfs-site.xml文件，并在配置文件中加入以下内容：

<configuration>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/hadoop/hadoop-2.7.3/hdfs/name</value>

<final>true</final>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/usr/hadoop/hadoop-2.7.3/hdfs/data</value>

<final>true</final>

</property>

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>master:9001</value>

</property>

<property>

<name>dfs.webhdfs.enabled</name>

<value>true</value>

</property>

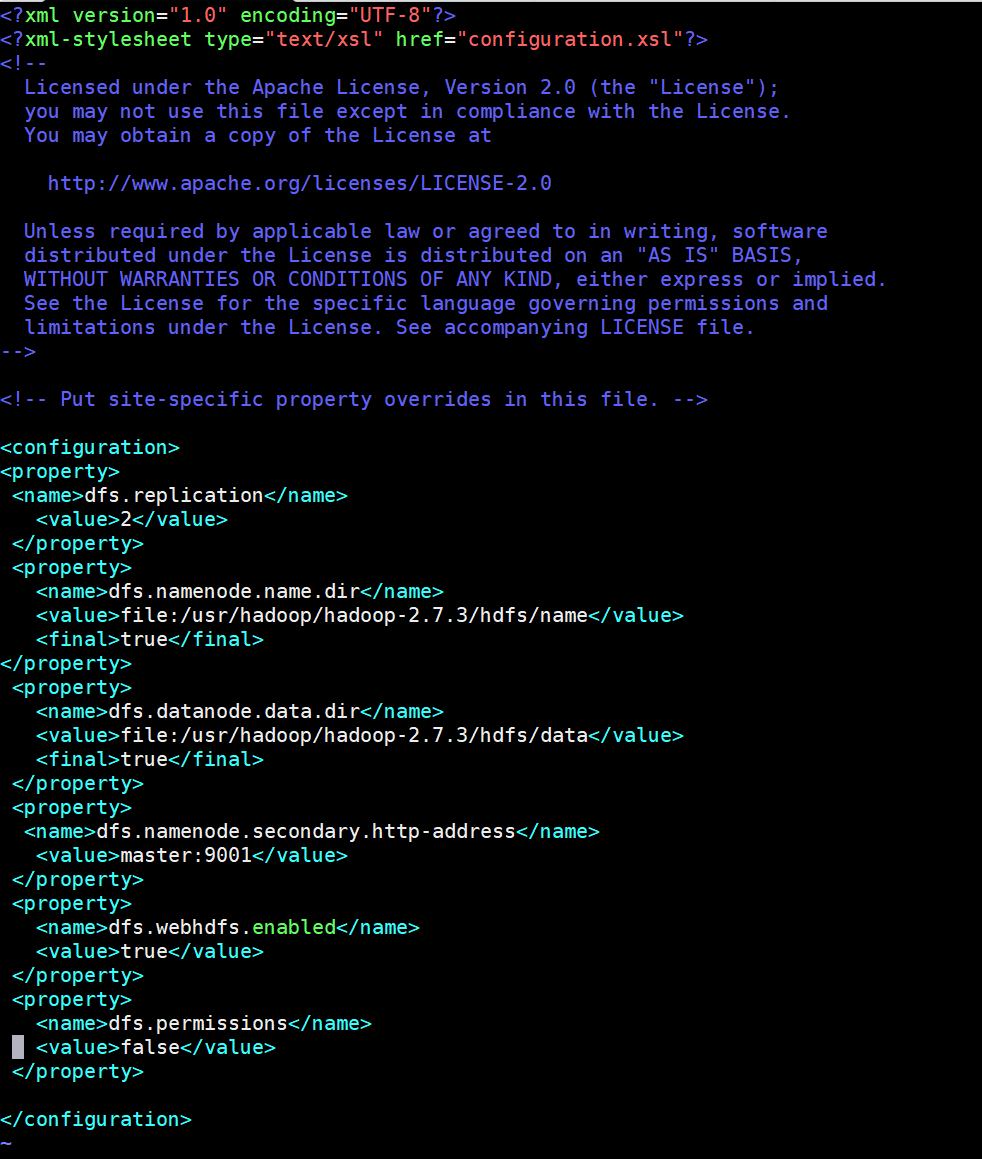
<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>



　　8. 使用vim命令编辑mapred-site.xmll文件，并在配置文件中加入以下内容：  
　　 但是文件夹内并没有mapred-site.xml这个文件所以我们需要使用cp命令将mapred-site.xml.template 复制为 mapred-site.xml：  
cp mapred-site.xml.template mapred-site.xml  
然后对其进行编辑，并在配置文件中加入以下内容：

http://bigdata.cloudlab.honyasec.com/storage/images/20180929003224zxJ.png

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

上一步

**分发hadoop**

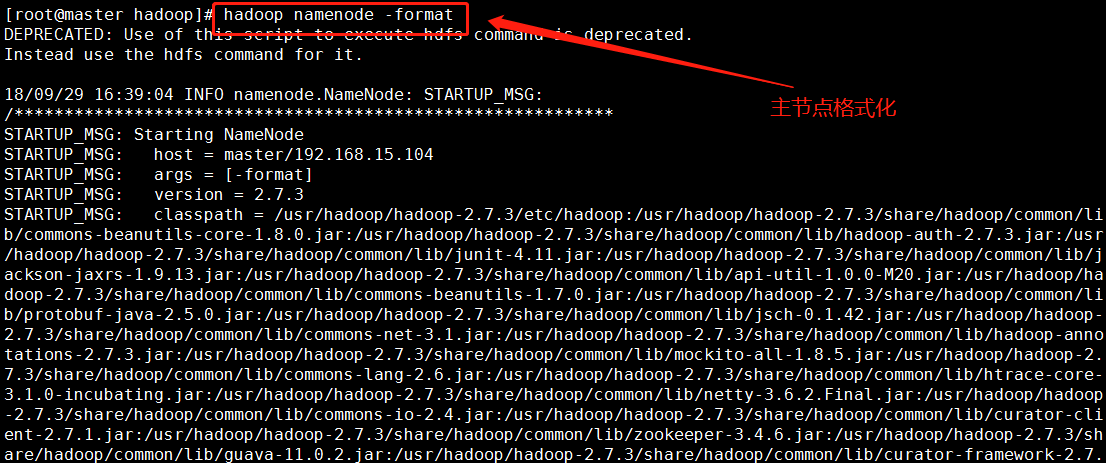
* scp -r /usr/hadoop root@slave1:/usr/
* scp -r /usr/hadoop root@slave2:/usr/

http://bigdata.cloudlab.honyasec.com/storage/images/20180929003338pED.png

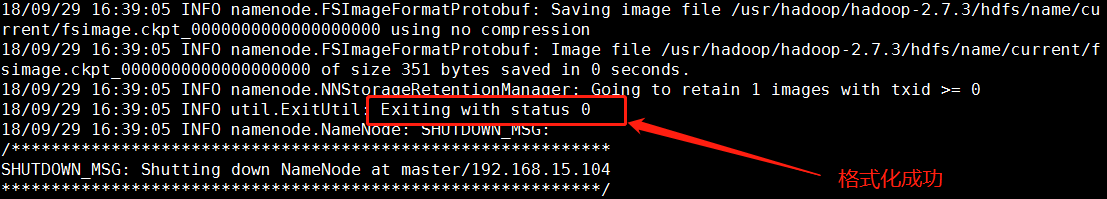
* 注意：slave1和slave2节点上都需要配置环境变量，参考hadoop中第二个步骤，同样是向/etc/hosts文件中的添加hadoop 的环境变量。

**格式化Hadoop并开启集群**

10.master中格式化hadoop  
输入hadoop namenode -format命令进行hadoop的格式化操作，如下图所示：

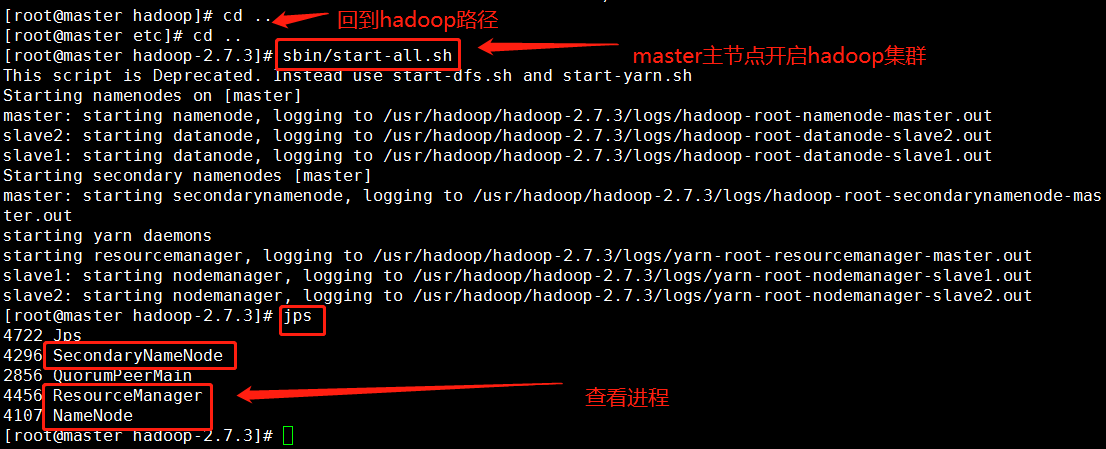


格式化成功如下图所示：



11.在格式化hadoop之后，退回到/usr/hadoop/hadoop-2.7.3目录下，然后使用sbin/start-all.sh命令开启hadoop集群：

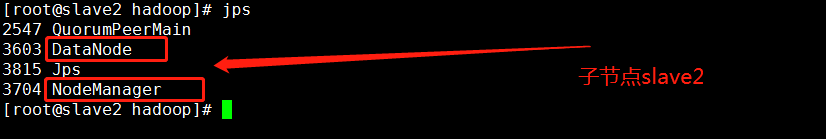
在master节点输入jps查看进程：



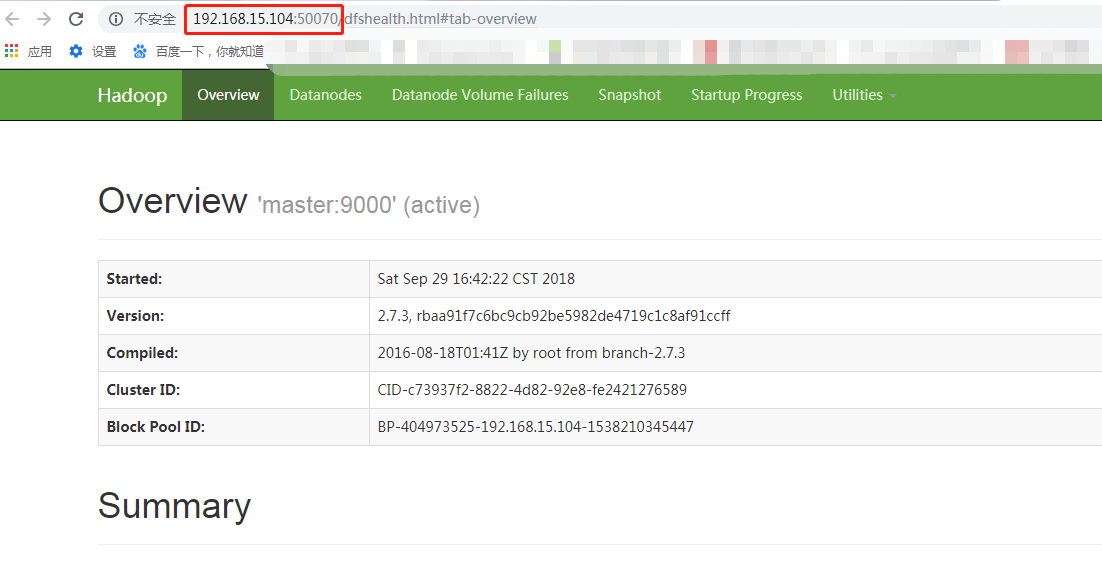
在slave1节点输入jps查看进程：



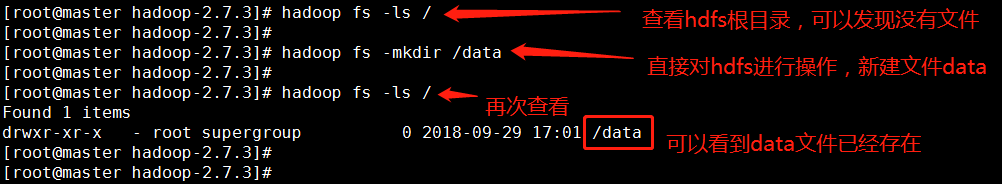
在slave2节点输入jps查看进程：



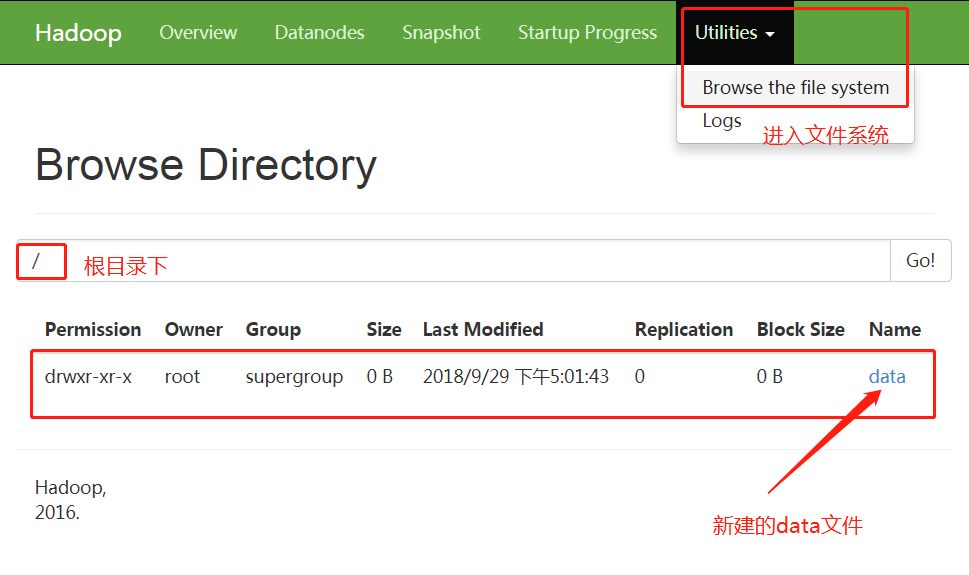
访问主节点master：50070



1. hadoop脚本命令练习  
   查看dfs根目录文件：hadoop fs –ls /  
   在hdfs上创建文件data ：hadoop fs –mkdir /data  
   再次进行查看：hadoop fs –ls /



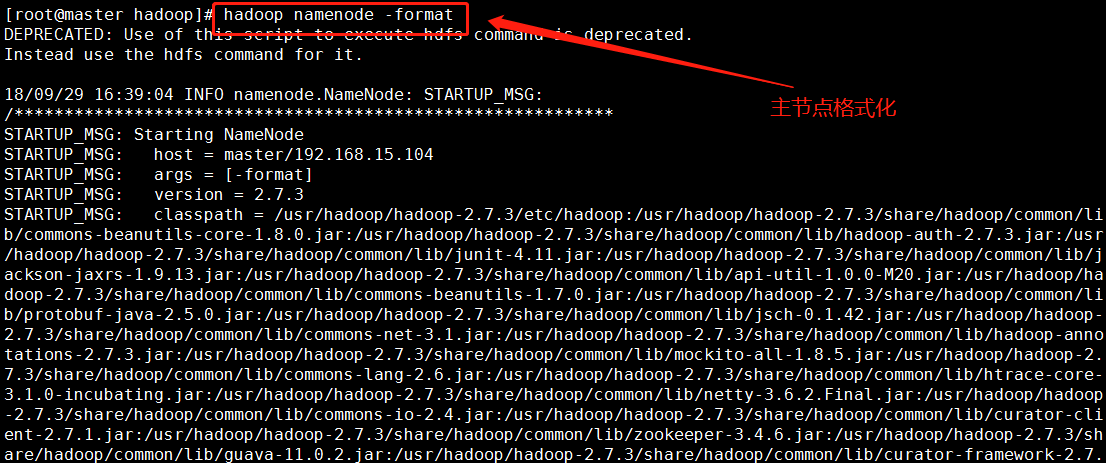
13.也可以使用浏览器对集群进行查看。依次进入“Utilities”->“Browse the file system”



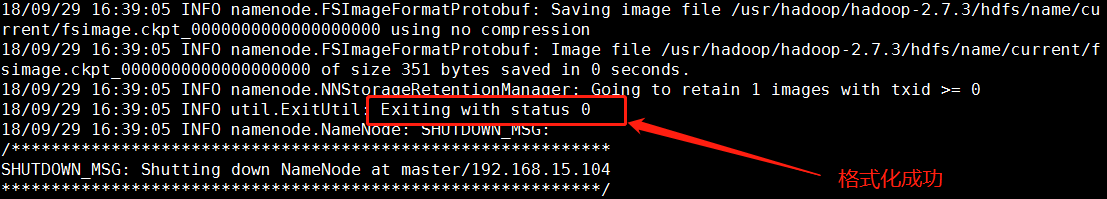
**实验总结**

实验总结空空的

10.master中格式化hadoop  
输入hadoop namenode -format命令进行hadoop的格式化操作，如下图所示：

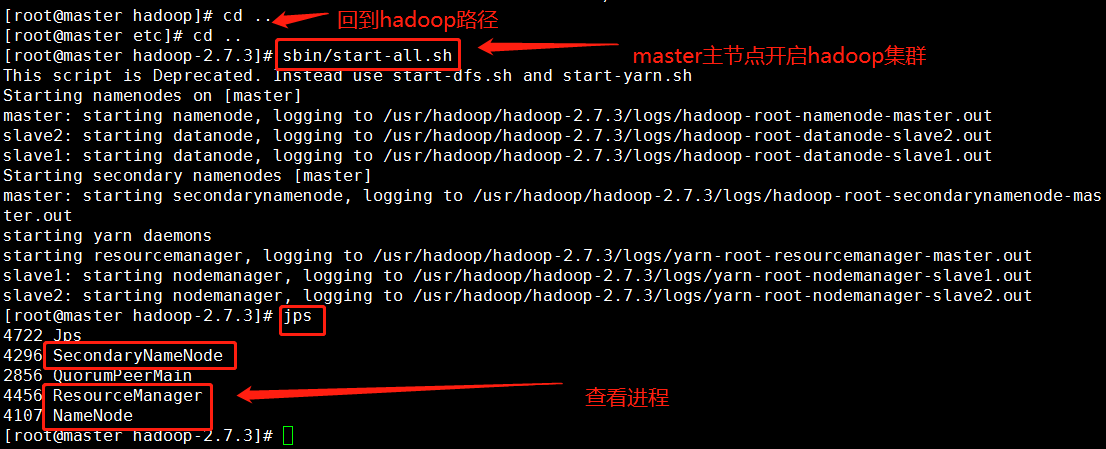


格式化成功如下图所示：



11.在格式化hadoop之后，退回到/usr/hadoop/hadoop-2.7.3目录下，然后使用sbin/start-all.sh命令开启hadoop集群：

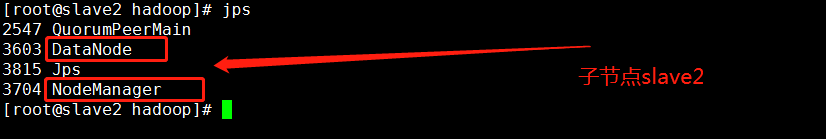
在master节点输入jps查看进程：



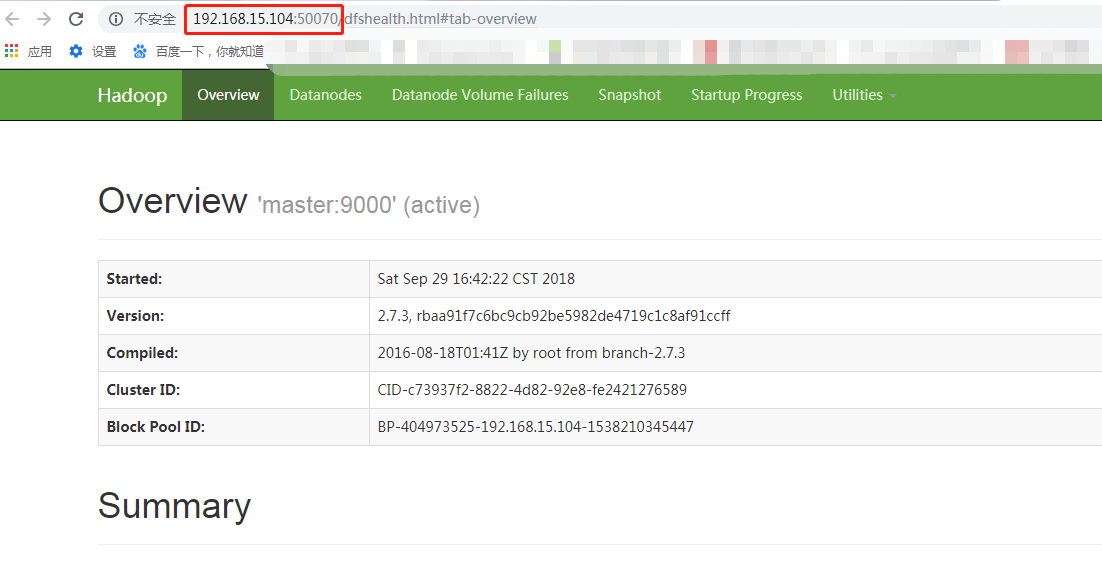
在slave1节点输入jps查看进程：



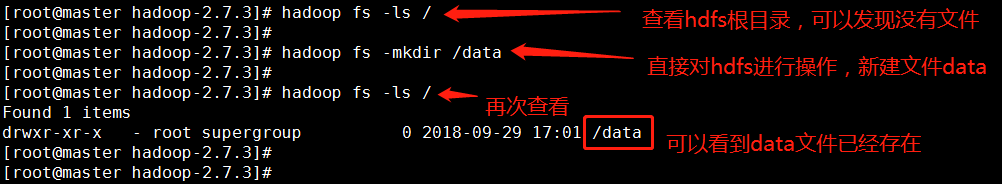
在slave2节点输入jps查看进程：



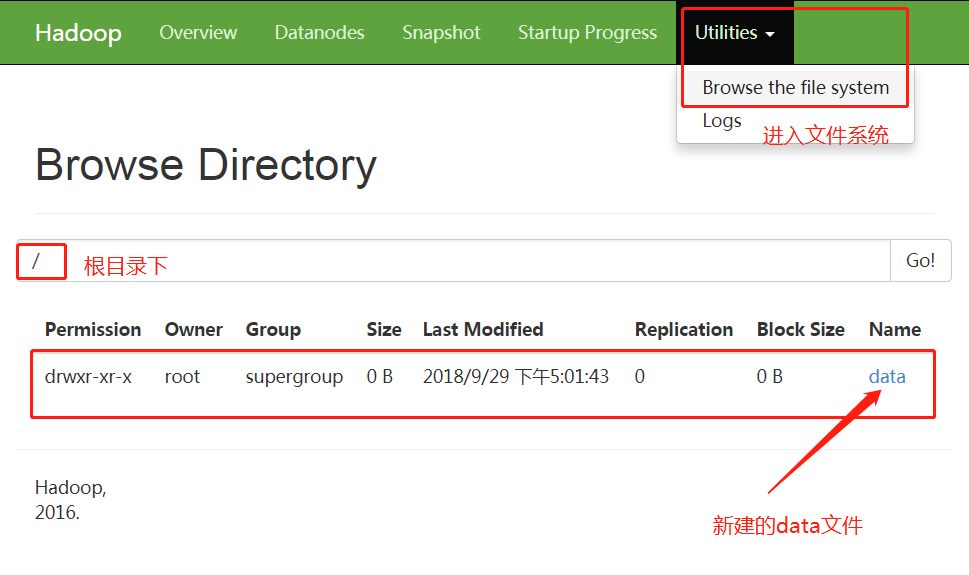
访问主节点master：50070



1. hadoop脚本命令练习  
   查看dfs根目录文件：hadoop fs –ls /  
   在hdfs上创建文件data ：hadoop fs –mkdir /data  
   再次进行查看：hadoop fs –ls /



13.也可以使用浏览器对集群进行查看。依次进入“Utilities”->“Browse the file system”



上一步