Hadoop1.0集群搭建详细教程

制作人：八斗大数据三期一班黄海峰

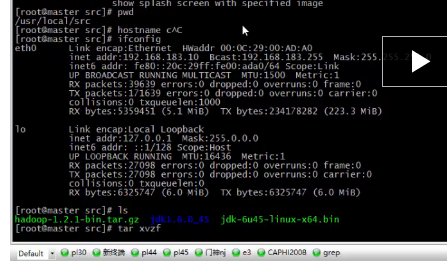
看到这份文档的，我想必大家都把linux下把java给安装和配置完毕了！

那么接下来我们开始搭建hadoop1.0集群了！！

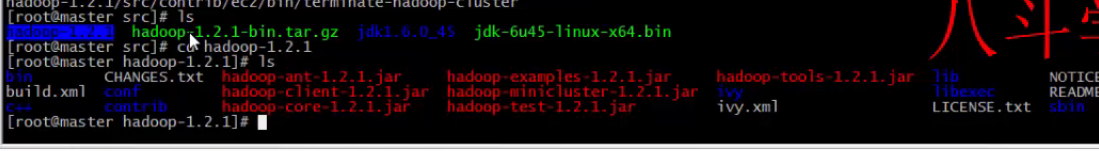
我写在下面的所有linux的命令可以直接复制到终端linux下进行运行！（那样方便我们以后的再次搭建！） 注意：红色字体的为重点字和相关命令。

1. 在主节点的 /usr/local/src/下进行对hadoop压缩包进行解压

命令行输入：tar xvzf hadoop-1.2.1-bin.tar.gz



解压完成以后，用ls命令进行查看，然后我们进入这个hadoop-1.2.1这个目录



然后在hadoop-1.2.1这个目录下创建一个tmp目录,用来存放之后的一些临时文件

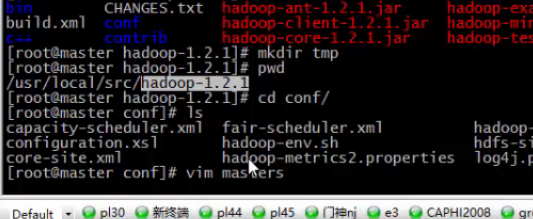
命令行输入：mkdir tmp

然后我们进入一个配置文件

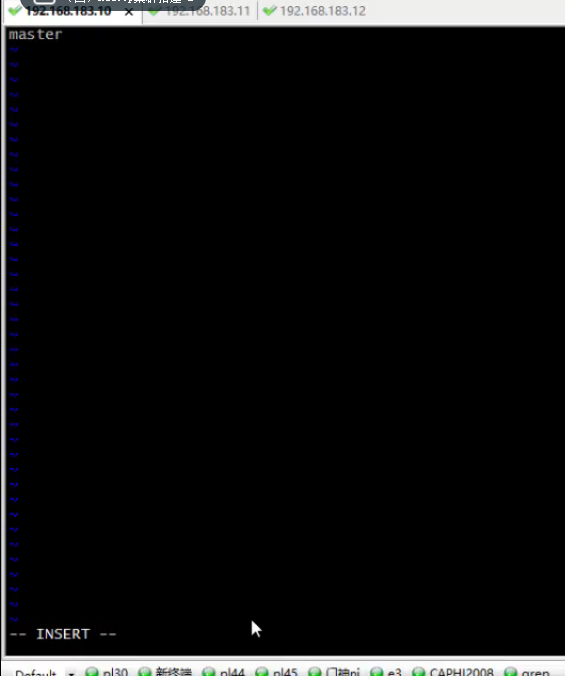
命令行输入：cd conf/

然后修改hadoop里面的masters文件

命令行输入：vim masters

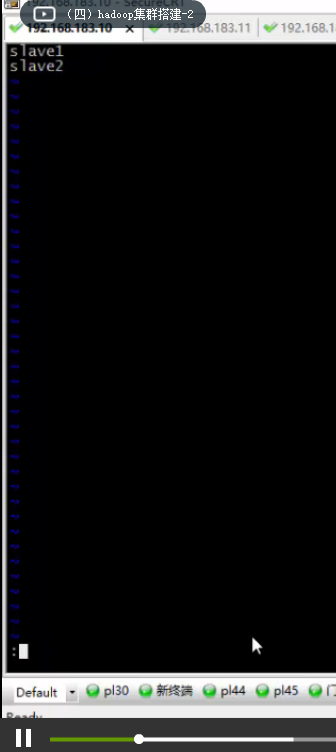


然后在masters里面写入master并且保存退出！



然后我们再修改slaves文件把子节点的名称：slave1和slave2写入slaves文件中

命令行输入：vim slaves



然后我们修改一下core-site.xml这个文件

命令行输入：vim core-site.xml

把我以下写的代码复制到core-site.xml这个文件中去然后保存退出

<property>

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

<value>/usr/local/src/hadoop-1.2.1/tmp</value>

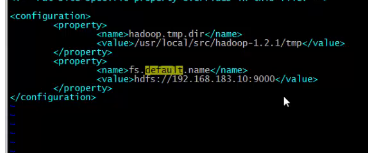
</property>

<property>

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

<value>hdfs://192.168.183.10:9000</value>

</property>



然后我们接下来修改mapred-site.xml这个文件

命令行输入：vim mapred-site.xml

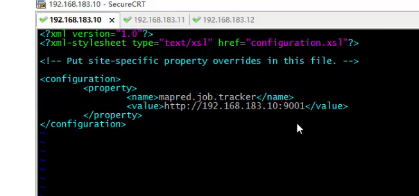
把我以下写的代码复制到vim mapred-site.xml这个文件中去然后保存退出

<property>

<name>mapred.job.tracker</name>

<value>http://192.168.183.10:9001</value>

</property>



然后我们紧接着配置修改hdfs-site.xml这个文件

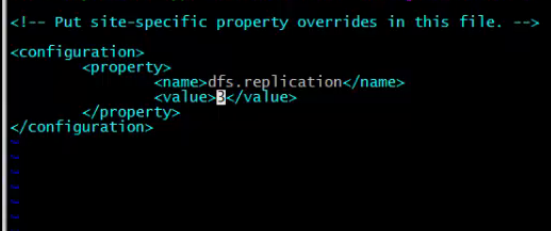
把我以下写的代码复制到hdfs-site.xml这个文件中去然后保存退出

<property>

<name>dfs.replication</name>

<value>3</value>

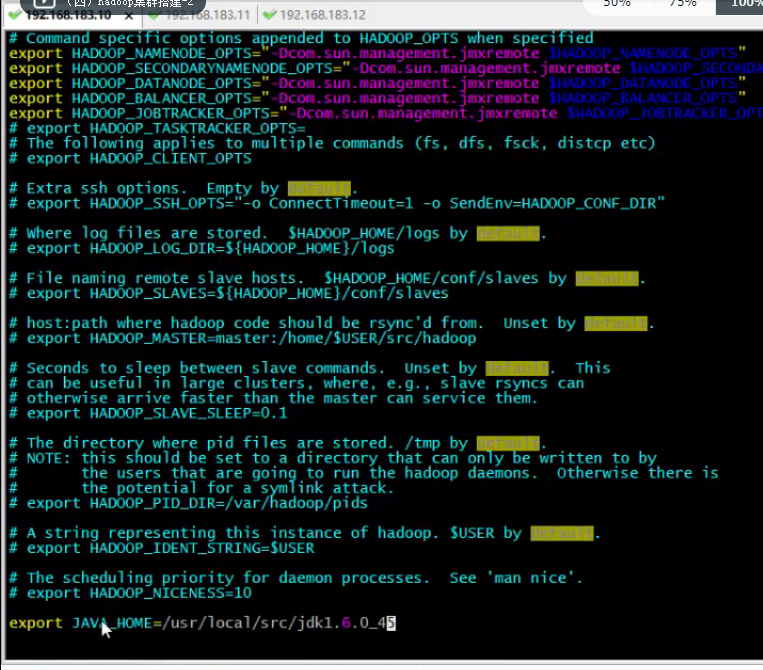
</property>



然后我们再修改 hadoop-env.sh这个文件（在这个文件里的最后面写入以下配置代码）

命令行输入:vim hadoop-env.sh

输入：export JAVA\_HOME=/usr/local/src/jdk1.6.0\_45



好了我们以上已经配置了6个文件了，在这里我写的手都酸了哈哈

1. 我们开始配置本地网络配置(修改hosts文件)

还是在conf下进行本地网络配置

命令行输入：vim /etc/hosts

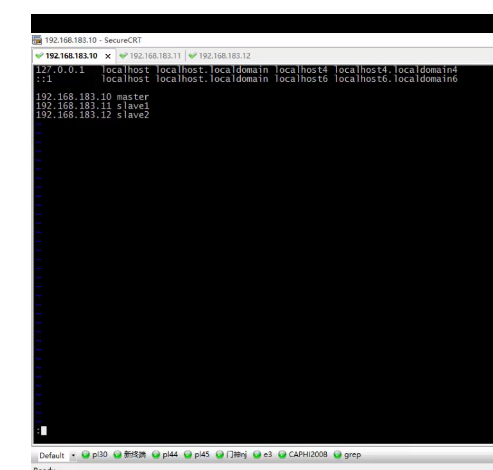


然后在hosts这个文件里面把主节点的ip地址写入进去

192.168.183.10 master

192.168.183.11 slave1

192.168.183.12 slave2

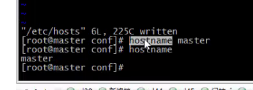


然后保存退出

但是host还没有生效，需要生效

在命令行输入：hostname master

然后再输入：hostname 查看

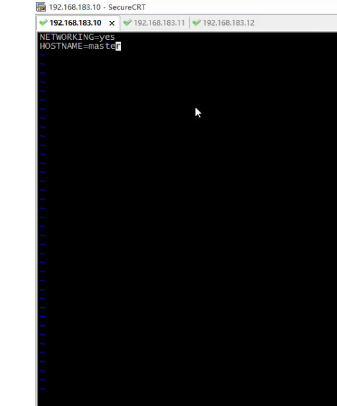


但是这个生效只是临时生效，我们需要改为永久生效：

在命令行输入：vim /etc/sysconfig/network

然后原先network里面的是:HOSTNAME=localhost

我们改为：HOSTNAME=master



然后再重新加载生效：命令行输入：hostname master

然后保存退出

现在主机节点已经配置完毕

那么接下来我们给剩下两个子节点进行配置

1. 在主节点命令行里面进入到src下面

命令行输入：cd /usr/local/src/

再输入：ls 将hadoop-1.2.1分别远程复制到其他两个节点当中去



在主节点命令行输入：scp -rp hadoop-1.2.1 192.168.183.11:/usr/local/src/

在主节点命令行输入：scp -rp hadoop-1.2.1 192.168.183.12:/usr/local/src/

然后切换到slave1，然后进入/usr/local/src/目录下面查看是否有hadoop-1.2.1

命令行输入：cd /usr/local/src/

然后给slave1进行配置hosts

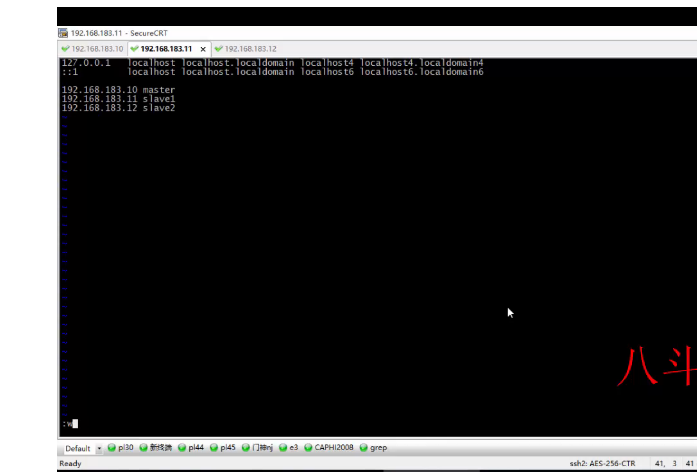
命令行输入：vim /etc/hosts

然后把下列代码写入到slave1中的hosts里面去

192.168.183.10 master

192.168.183.11 slave1

192.168.183.12 slave2



然后slave2也是一样把上面的hosts里面写入一下代码：  
192.168.183.10 master

192.168.183.11 slave1

192.168.183.12 slave2

然后再给slave1配置修改network

在命令行输入：vim /etc/sysconfig/network

然后原先network里面的是:HOSTNAME=localhost

我们改为：HOSTNAME=slave1

然后再重新加载生效：命令行输入：hostname slave1

然后再给slave2配置修改network

在命令行输入：vim /etc/sysconfig/network

然后原先network里面的是:HOSTNAME=localhost

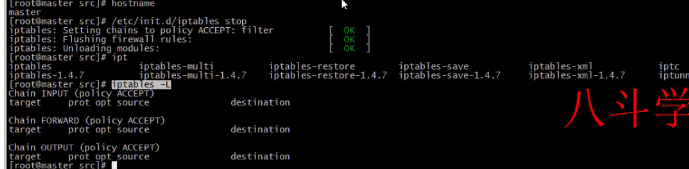
我们改为：HOSTNAME=slave2

然后再重新加载生效：命令行输入：hostname slave2

4.关闭防火墙(在三台机器上或者更多的机器上执行一下命令)在src目录下

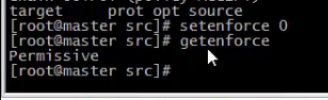
在命令行输入：/ect/init.d/iptables stop

然后检查一下是否关闭成功：在命令行输入：chkconfig --level 35 iptables off



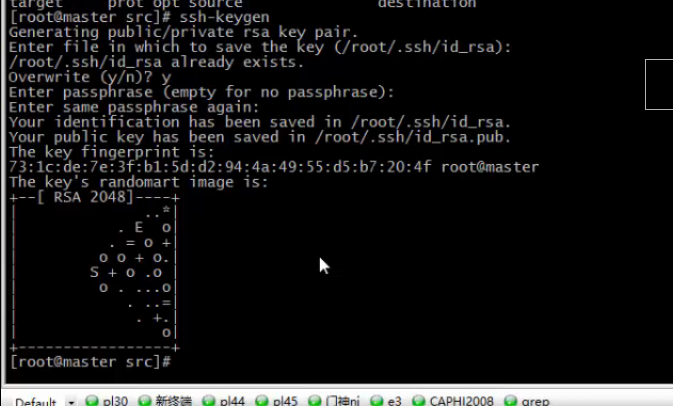
然后再从命令行输入：setenforce 0 (在每台机器都输入此命令)

然后再从命令行输入: getenforce 查看是否关闭了 (在每台机器都输入此命令)



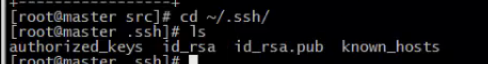
1. 建立每台机器的互信关系(这样方便我们日后一台机器对另外一台机器的操作)

在命令行输入:ssh-keygen 然后回车然后回车（回两次车）



然后进入隐藏目录

在命令行输入：cd ~/.ssh/ 然后ls



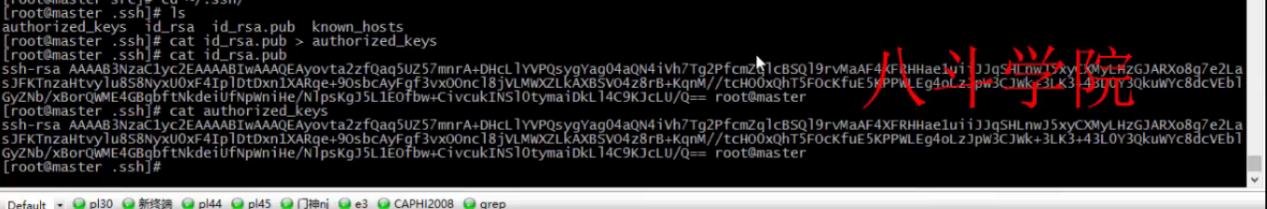
Id\_rsa.pub是共钥文件，id\_rsa是密钥文件

我们要对id\_rsa.pub这个公钥文件的内容拷贝给authorized\_keys这个文件（在此说明：如果没有authorized\_keys这个文件，请自己创建！！）

我们看一下id\_rsa.pub这个公钥文件的内容

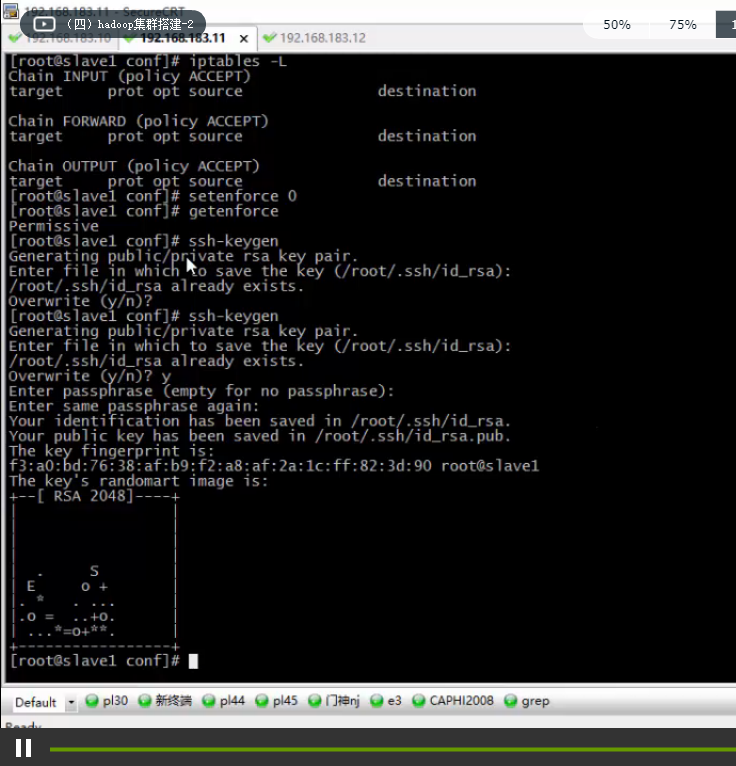
输入命令行：cat id\_rsa.pub

然后再输入命令行：cat authorized\_keys 看看这两个文件的公钥是否一样



然后我们分别给其他两个子节点设置公钥

在slave1命令行输入:ssh-keygen 然后回车然后回车（回两次车）



然后进入隐藏目录

在命令行输入：cd ~/.ssh/ 然后ls

我们看一下id\_rsa.pub这个公钥文件的内容

输入命令行：cat id\_rsa.pub

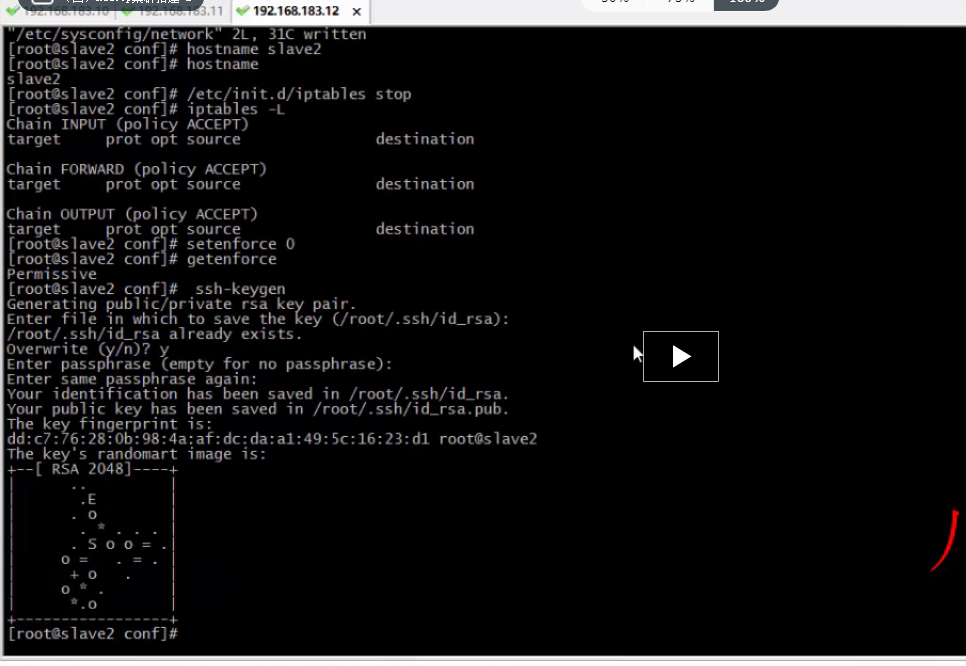
然后把slave1的公钥复制到主机节点master的authorized\_keys上

在master命令行输入：vim authorized\_keys

然后接下来把然后把slave1的公钥复制到这里



在slave2命令行输入:ssh-keygen 然后回车然后回车（回两次车）



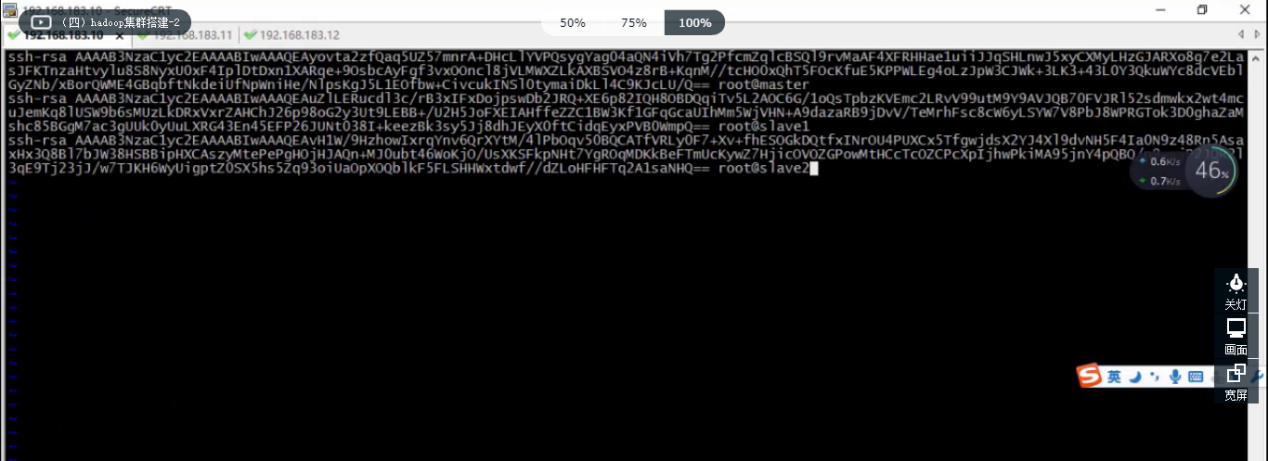
我们看一下id\_rsa.pub这个公钥文件的内容

输入命令行：cat id\_rsa.pub

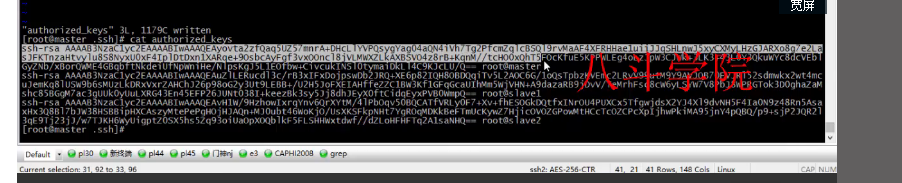
然后把slave2的公钥复制到主机节点master的authorized\_keys上

在master命令行输入：vim authorized\_keys

然后接下来把然后把slave2的公钥复制到这里



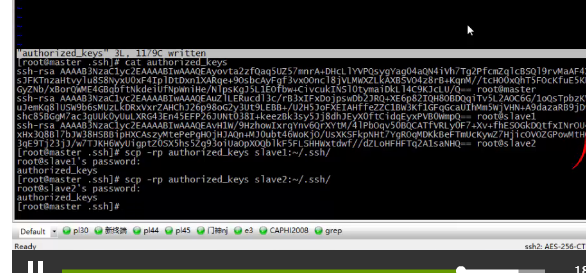
然后在主节点上输入：cat authorized\_keys 进行查看



然后下一步操作就是把主节点上的authorized\_keys分别复制给slave1和slave2这两个节点上

命令行输入：scp -rp authorized\_keys slave1:~/.ssh/

命令行输入：scp -rp authorized\_keys slave2:~/.ssh/



然后分别给slave1和slave2检查一下是否有主节点传过来的公钥文件

在slave1和slave2的命令行上分别输入：cat authorized\_keys

然后我们就可以给三台机器互相登录而且不需要密码了

那么我们试一下：在主节点上命令行上输入：ssh slave1 如果要退出的话就是输入：exit

然后我们在主机节点上登录：ssh slave2

1. 再进入主机点上进行操作：  
   我们开始安装hadoop1.0

命令行输入：cd hadoop-1.2.1

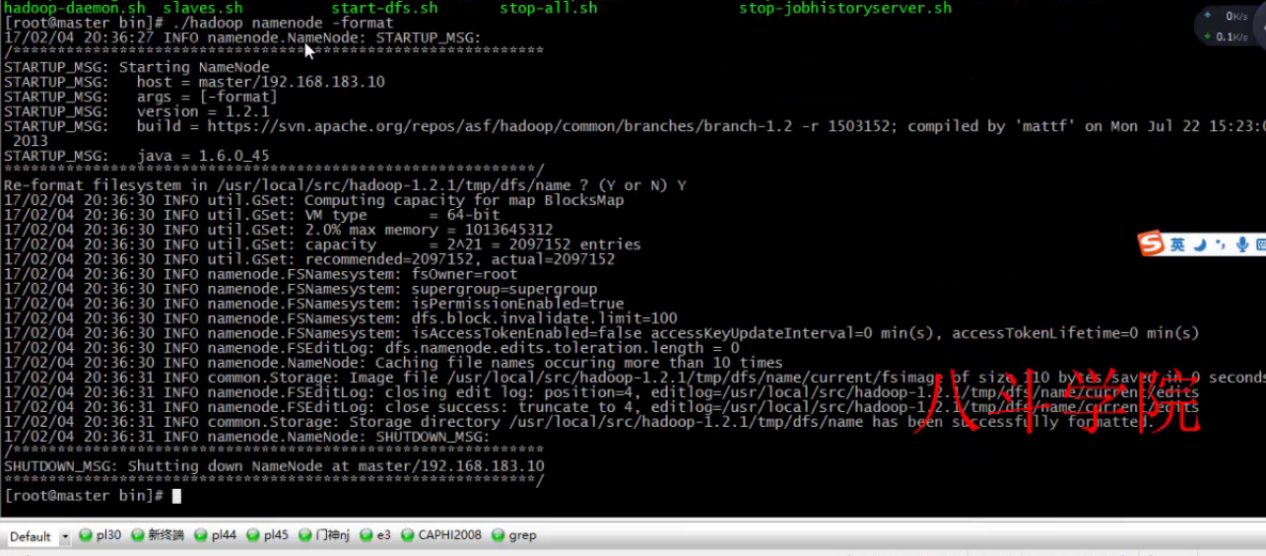
命令行再输入：cd bin

然后开始启动我们的hadoop集群

我们第一次启动的时候需要格式化

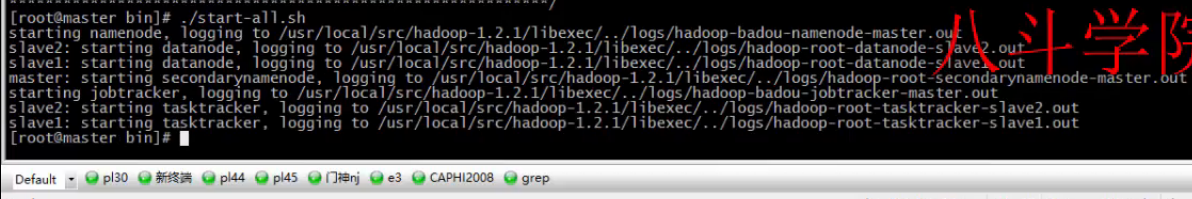
命令行输入: ./hadoop namenode -format





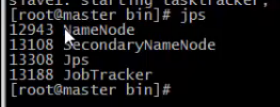
格式化以后我们再把hadoop集群启动：

命令行输入：./start-all.sh



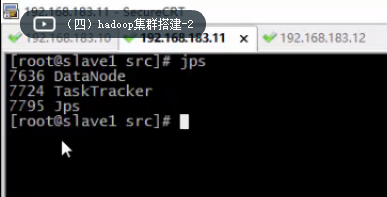
然后我们查看这个集群是否有问题

命令行输入：jps



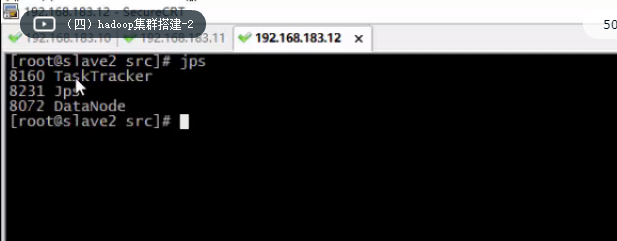
然后再去看看slave1子节点上是否运行集群成功了呢？

在slave1子节点上输入：jps



然后再去看看slave2子节点上是否运行集群成功了呢？

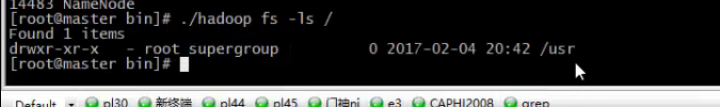
在slave2子节点上输入：jps



那么我们测试一下hadoop能不能用呢？

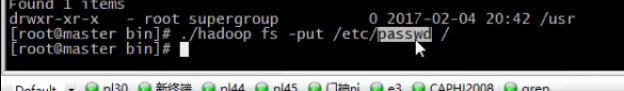
随便敲一个hadoop命令

在主节点输入：./hadoop fs -ls /



那么我们上传一个文件试试

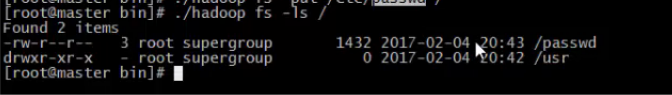
在主节点输入：./hadoop fs -put /etc/passwd /



那么我们看看能不能把这个文件读出来呢？

在主节点输入：./hadoop fs -ls /

在主机点输入：./hadoop fs -cat /passwd





好了，在这里我们就已经完成了hadoop搭建了！！然后就可以进行我们后续的学习了！！！