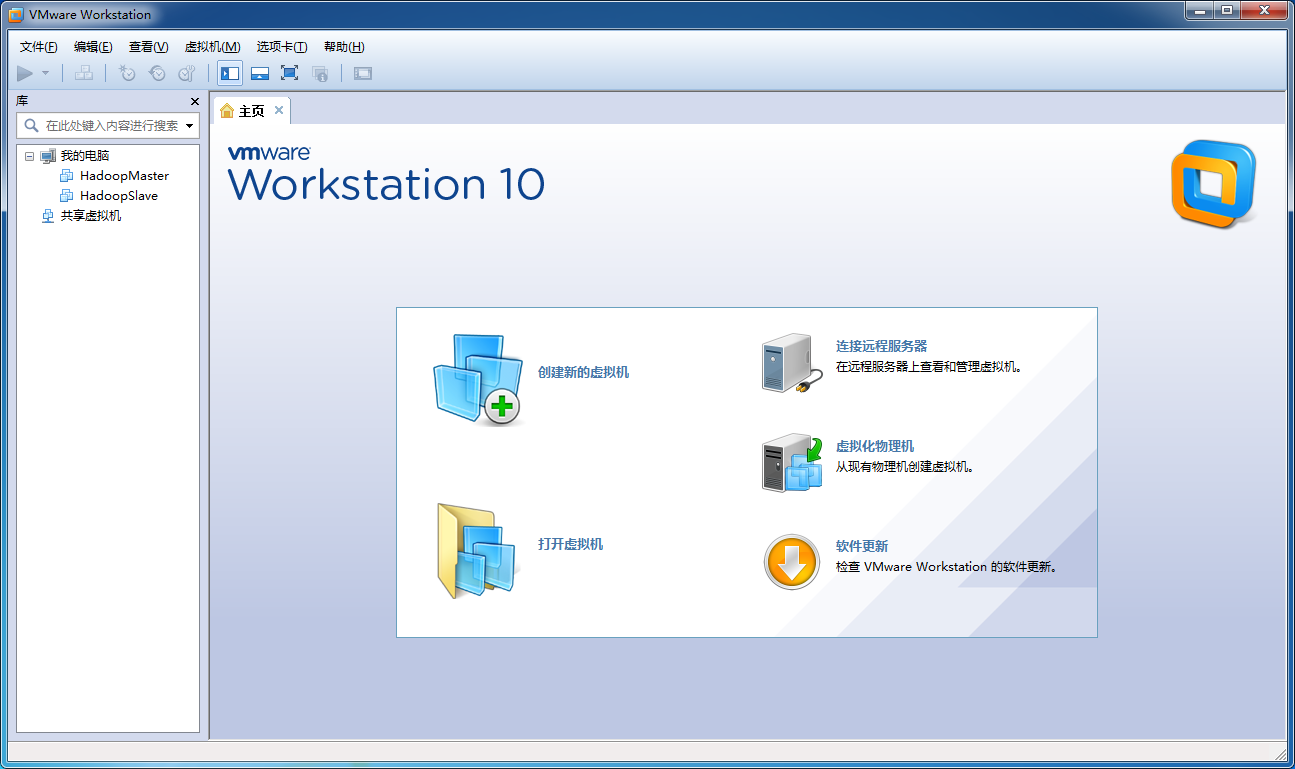
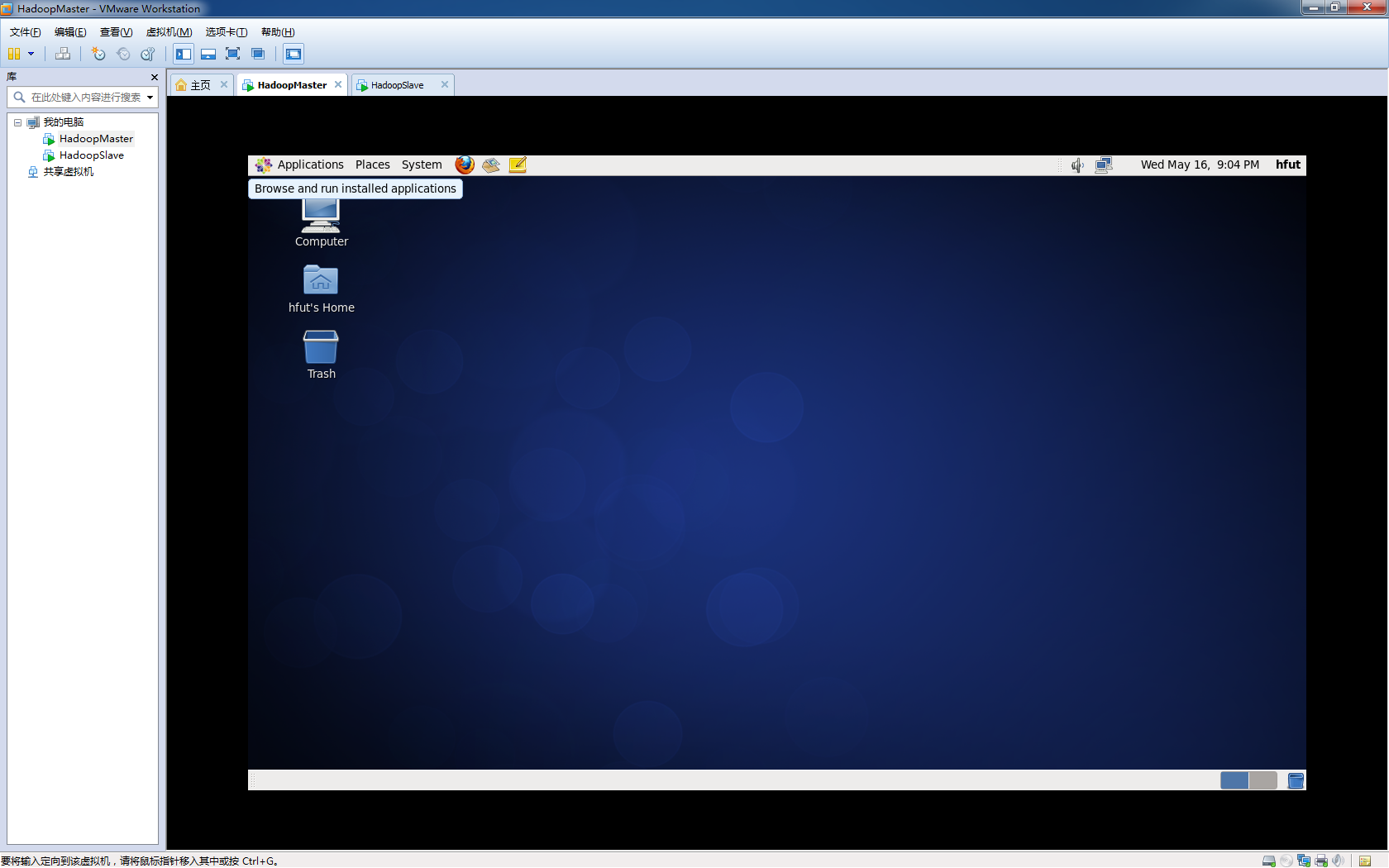
**4. CentOS 6上安装Hadoop**

**4.1 启动两台虚拟客户机**

打开VMware Workstation 10



打开之前已经安装好的虚拟机：点击“开启此虚拟机”，分别打开HadoopMaster和HadoopSlave



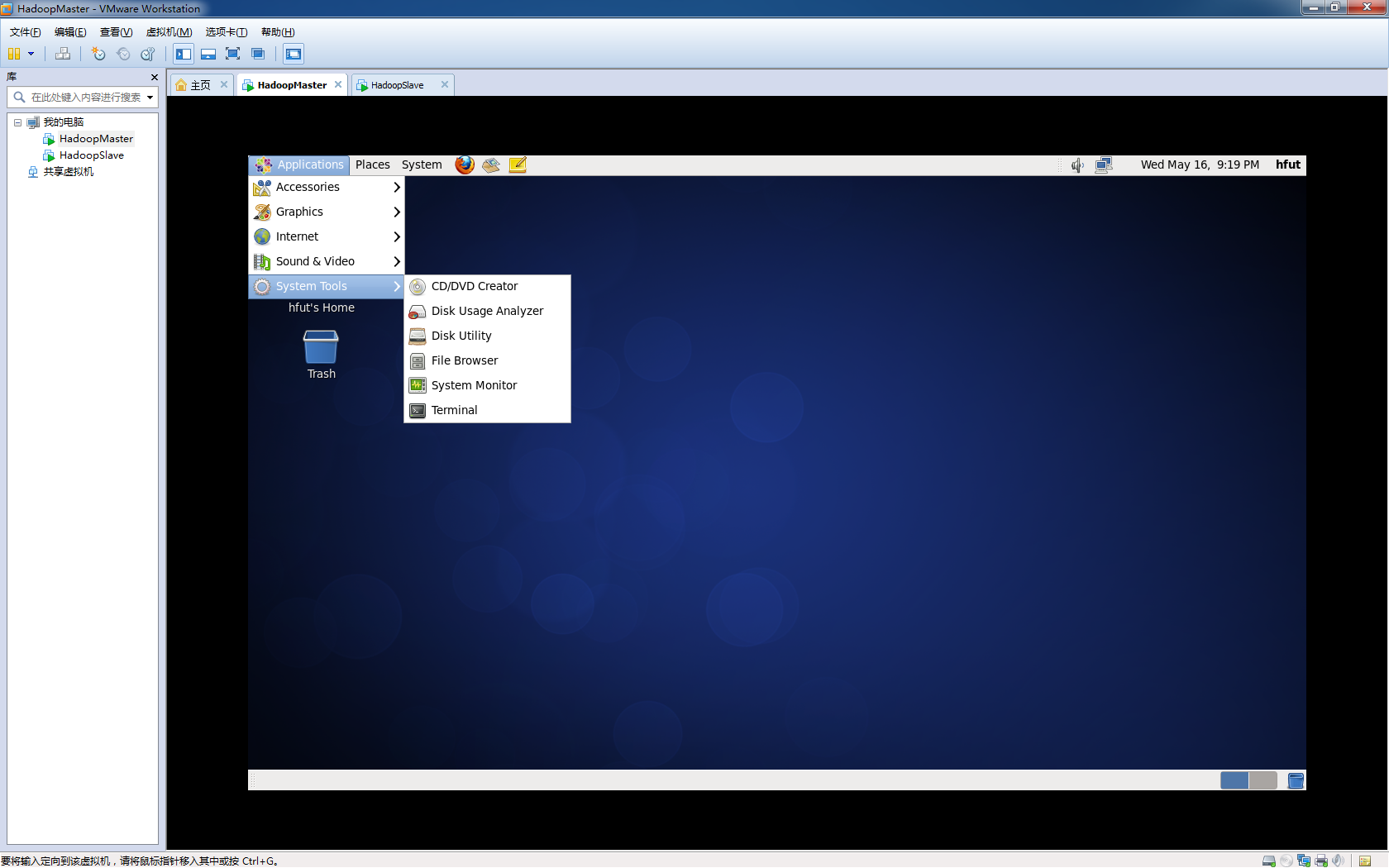
**4.2 Linux系统配置**

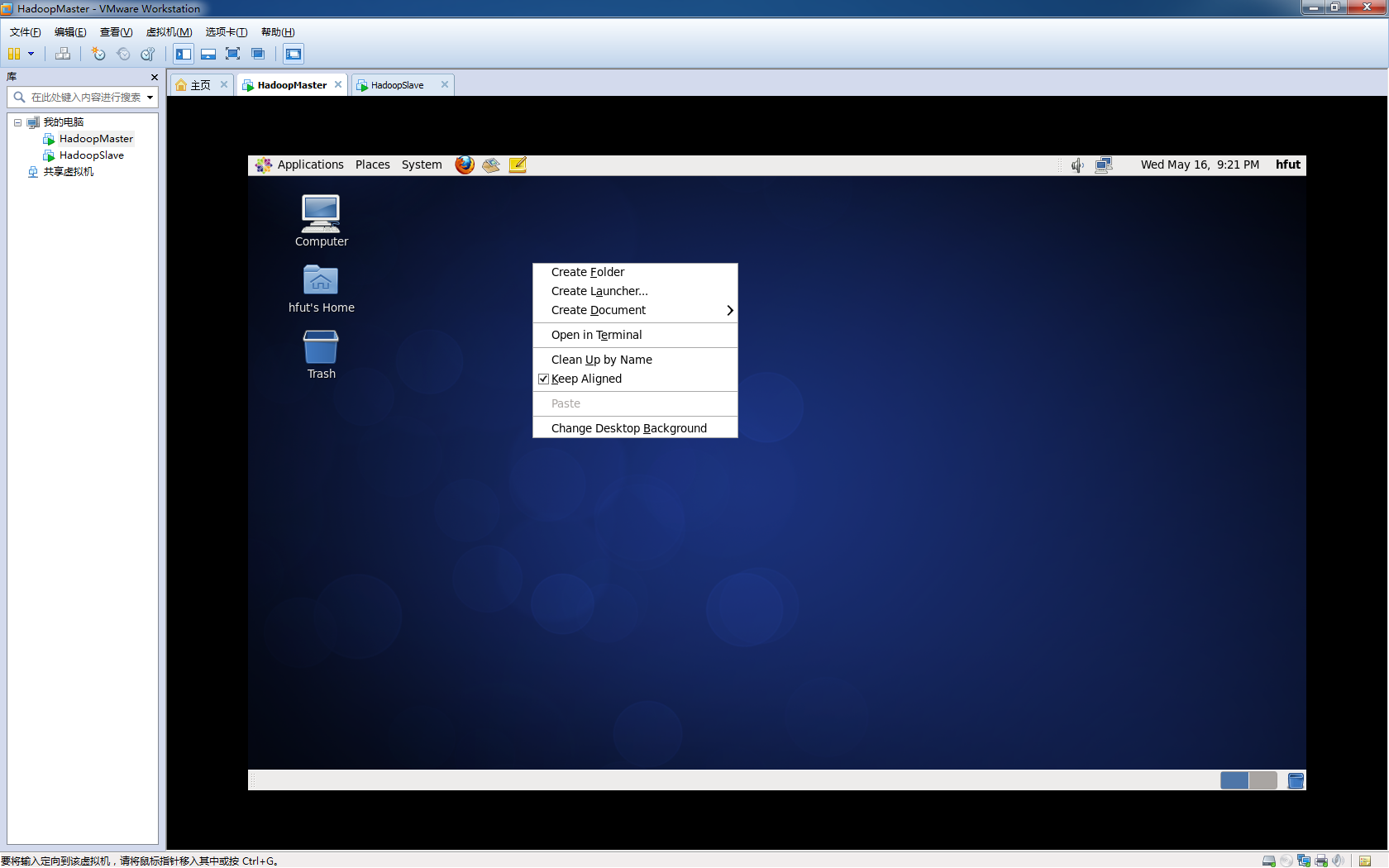
**以下操作步骤需要在HadoopMaster和HadoopSlave节点上分别完整操作，都使用root用户，从当前用户切换到root用户的命令如下：**

[hfut@localhost ~]$ su root

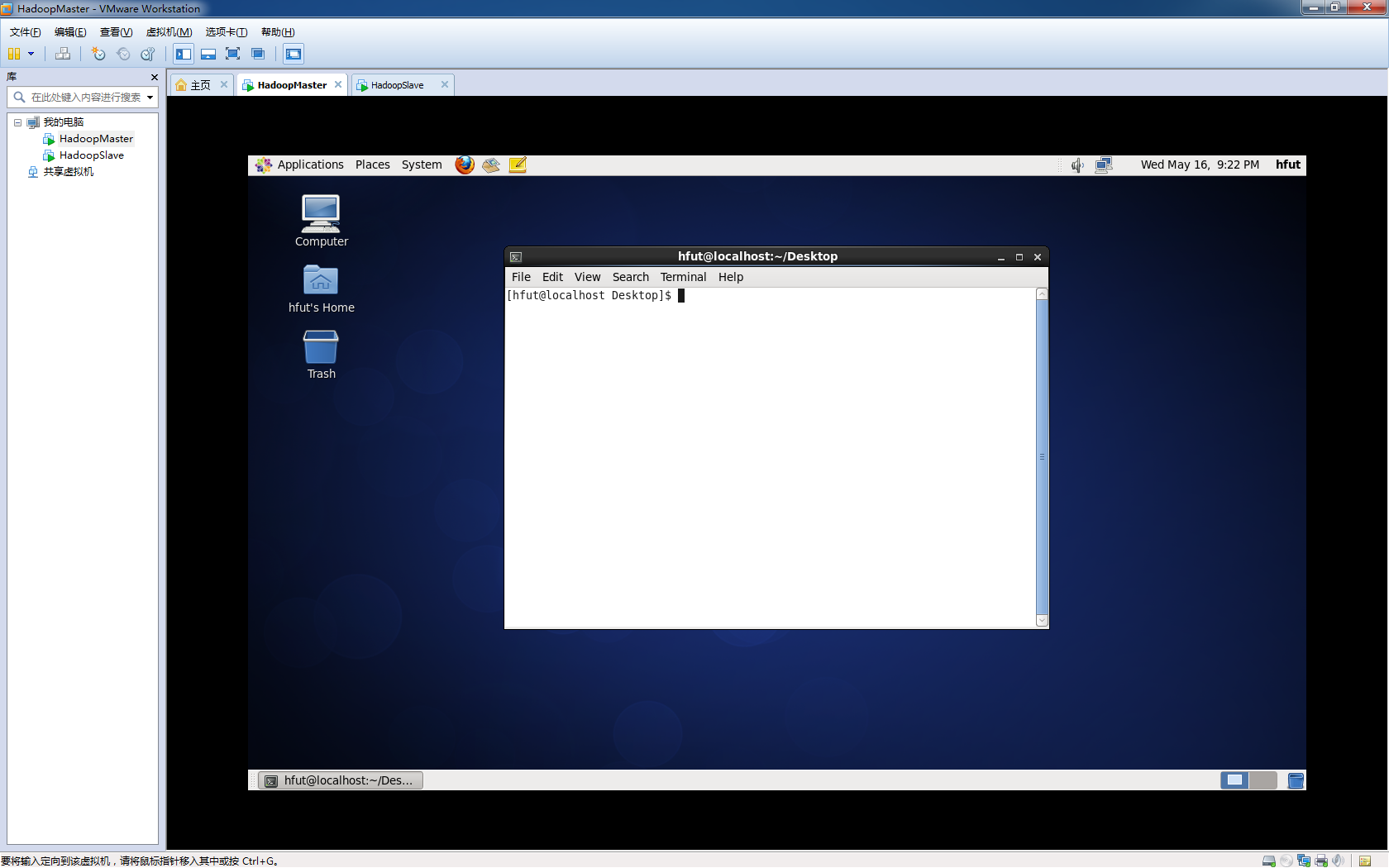
输入密码：hfut

本节所有的命令操作都在终端环境，打开终端的过程如下图的Applications->System Tools->Terminal菜单，或使用鼠标右键菜单的“Open in Terminal”





终端打开后如下图中命令行窗口所示。



**4.2.1 软件包和数据包说明**

大数据处理技术实验教学资源链接：链接: https://pan.baidu.com/s/1P4vDgBEj\_eOSakabM93rew 提取码: eujp

**将完整的软件包上传到“/home/hfut/resources/software”。**

**4.2.2 配置时钟同步**

（1）配置自动时钟同步

该项同时需要在HadoopSlave节点配置。

使用Linux命令配置

[root@localhost hfut]# crontab -e

// 编辑crontab服务

该命令是vi编辑命令，按i进入插入模式，按Esc，然后键入:wq保存退出。

键入下面的一行代码，输入i，进入插入模式（星号之间和前后都有空格）

0 1 \* \* \* /usr/sbin/ntpdate cn.pool.ntp.org

//分 小时 日 月 每周的第几天 命令

（2）修改时区

[root@localhost hfut]# vi /etc/sysconfig/clock

输入如下内容：

ZONE="Asia/Shanghai"

连接到上海时区文件

[root@localhost hfut]# rm /etc/localtime

[root@localhost hfut]# ln -sf /usr/share/zoneinfo/Asia/Shanghai /etc/localtime

[root@localhost hfut]# reboot

（3）手动同步时间

直接在Terminal运行下面的命令：

[root@localhost hfut]# /usr/sbin/ntpdate cn.pool.ntp.org

**4.2.3 配置主机名**

（1）HadoopMaster节点

使用gedit编辑主机名，如果不可以使用gedit，请直接使用vi编辑器（后面用到gedit的地方也同此处一致）。

[root@localhost hfut]# vi /etc/sysconfig/network

配置信息如下，将HadoopMaster节点的主机名改为master，即下面代码的第2行所示。

NETWORKING=yes

HOSTNAME=master

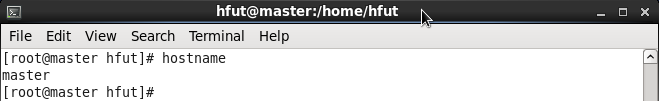
确认修改生效命令

[root@localhost hfut]# hostname master

检测主机名是否修改成功命令如下，在操作之前需要关闭当前终端，重新打开一个终端：

[root@localhost hfut]# hostname

执行完命令，会看到下图的打印输出：



（2）HadoopSlave节点

使用vi编辑主机名。

[root@localhost hfut]# vi /etc/sysconfig/network

配置信息如下，将HadoopMaster节点的主机名改为slave，即下面代码的第2行所示。

NETWORKING=yes

HOSTNAME=slave

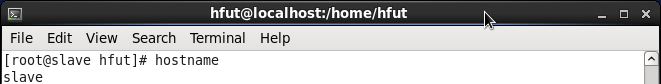
确认修改生效命令

[root@localhost hfut]# hostname slave

检测主机名是否修改成功命令如下，在操作之前需要关闭当前终端，重新打开一个终端：

[root@localhost hfut]# hostname

执行完命令，会看到下图的打印输出：



**4.2.4 使用setup命令配置网络环境**

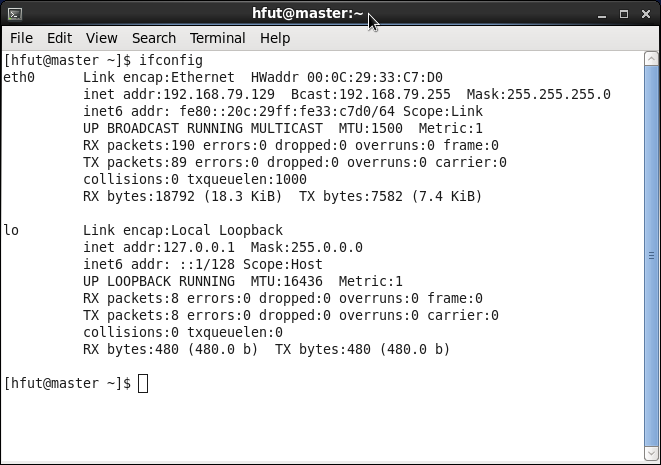
**也可在图形界面上配置**

**该项也需要在HadoopSlave节点配置**

在终端执行下面的命令：

[hfut@master ~]$ ifconfig

如果看到下面的打印输出

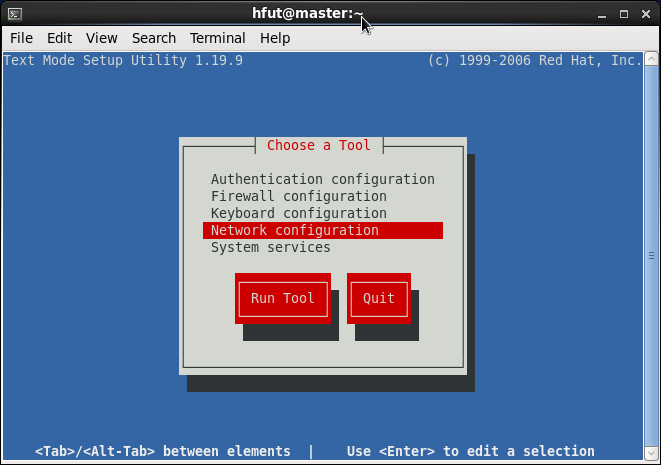


如果看到出现红线标注部分出现，即存在内网IP、广播地址、子网掩码，说明该节点不需要配置网络，否则进行下面的步骤。

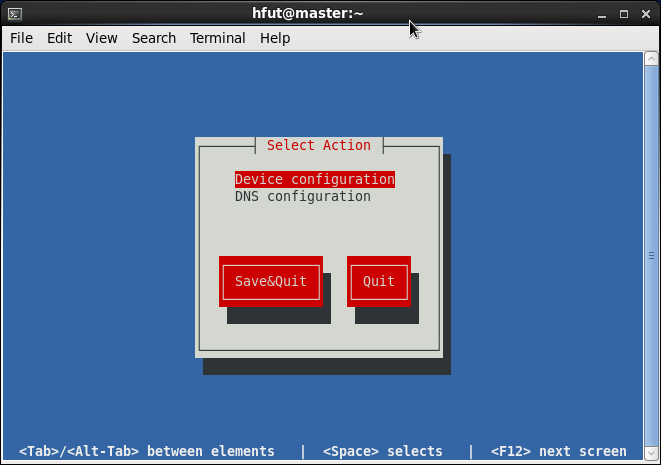
在终端执行下面的命令：

[hfut@master ~]$ setup

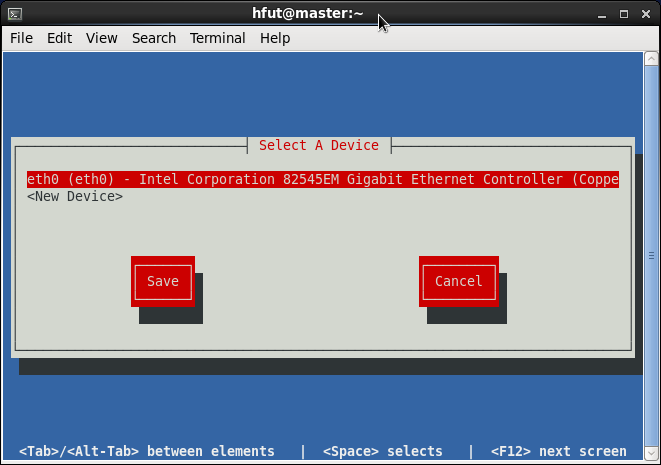
会出现下图中的内容：



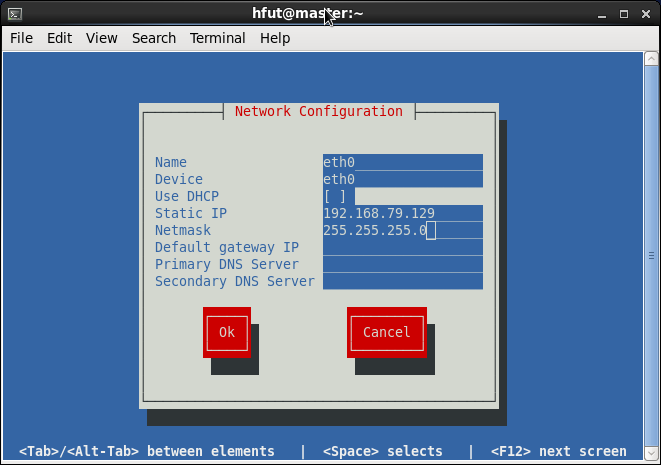
使用光标键移动选择“network configuration”，回车进入该项。



选“Device configuration”，回车



选eth0，回车



输入各项内容

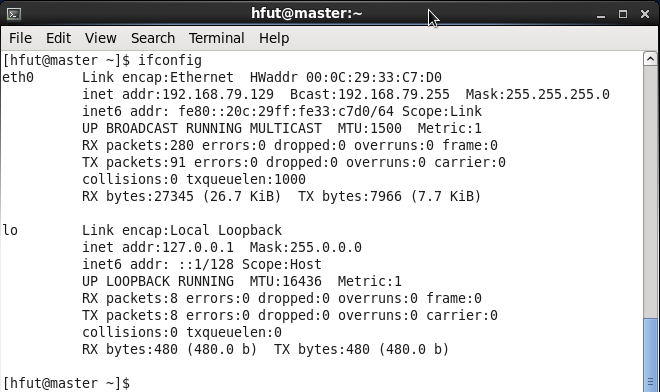
重启网络服务

[hfut@master ~]$ /sbin/service network restart

检查是否修改成功

[hfut@master ~]$ ifconfig

看到如下图的内容（IP不一定和下图相同，根据你之前的配置出现相应的内容），说明配置成功，请特别关注红线部分。



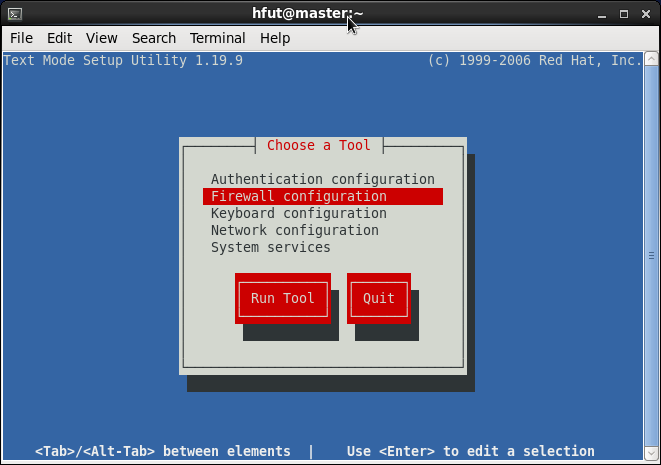
**4.2.5 关闭防火墙**

**该项也需要在HadoopSlave节点配置**

在终端执行下面的命令：

[hfut@master ~]$ setup

会出现下图中的内容

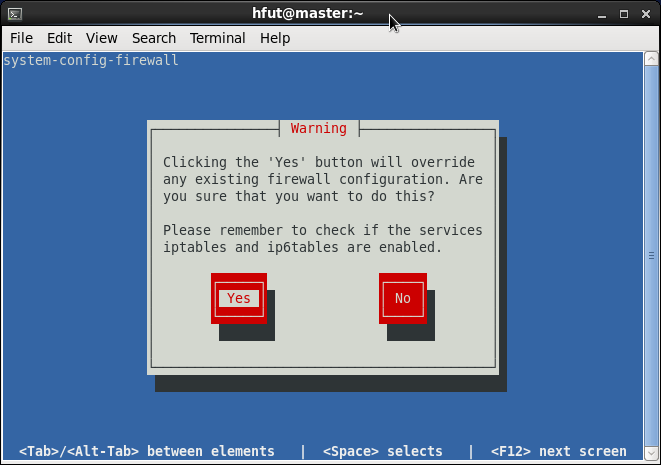


光标移动选择“Firewall configuration”选项，回车进入选项

如果该项前面有“\*”标记，则按一下空格键关闭防火墙，如下图所示，然后移动光标选择“OK”保存修改内容。



选择“OK”



选择“Yes”

**4.2.6 配置hosts列表**

**该项也需要在HadoopSlave节点配置**

需要在root用户下（使用su命令），编辑主机名列表的命令：

[root@master hfut]# vi /etc/hosts

将下面两行添加到/etc/hosts文件中：

192.168.79.129 master

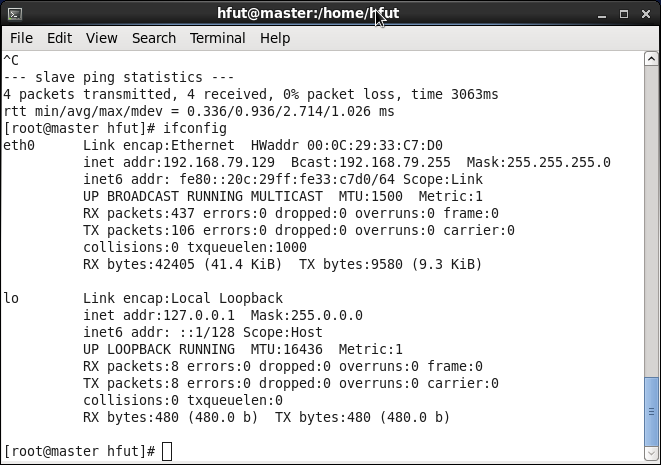
192.168.79.130 slave

**注意：这里master节点对应IP地址是192.168.79.129，slave对应的IP是192.168.79.130，而自己在做配置是，需要将这两个IP地址改成你的master和slave对应的IP地址。**

查看master的IP地址使用下面的命令

[root@master hfut]# ifconfig

Master节点的IP是下图中红线标注的内容。



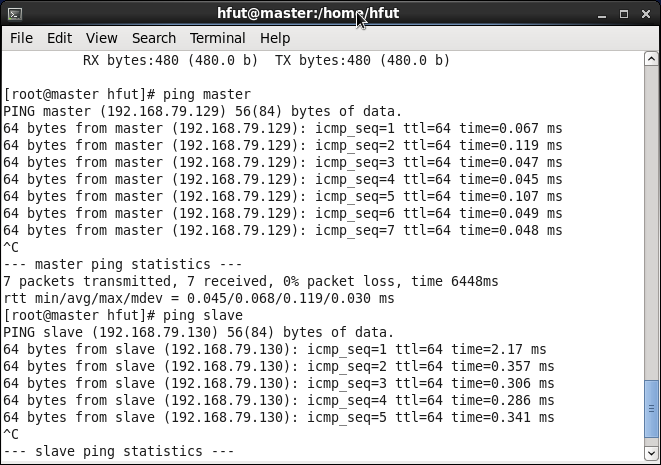
slave的IP地址也是这样查看。

验证是否配置成功的命令是：

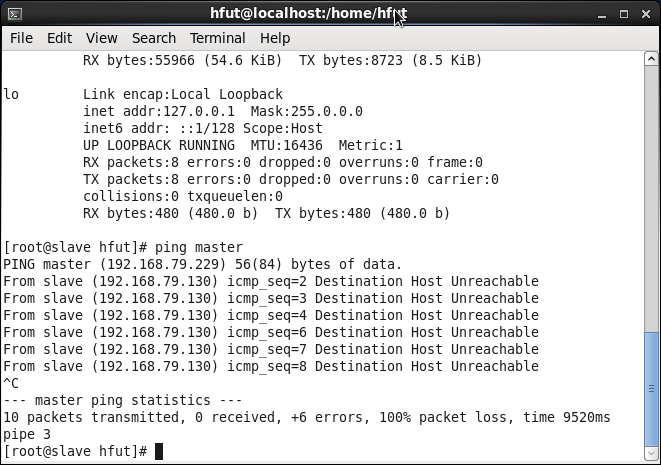
[root@master hfut]# ping master

[root@master hfut]# ping slave

如果出现下图的信息表示配置成功：



如果出现下图的内容，表示配置失败：



**4.2.7 安装JDK**

**该项也需要在HadoopSlave节点配置**

将JDK文件解压，放在/usr/java目录下

[root@master hfut]# cd /home/hfut/resources/software/jdk

[root@master jdk]# mkdir /usr/java

[root@master jdk]# cp jdk-7u71-linux-x64.gz /usr/java

[root@master jdk]# cd /usr/java

[root@master java]# tar -zxvf jdk-7u71-linux-x64.gz

[root@master java]# rm jdk-7u71-linux-x64.gz

使用gedit配置环境变量

[root@master hfut]# su hfut

[hfut@master java]$ gedit /home/hfut/.bash\_profile

复制粘贴以下内容添加到.bash\_profile

export JAVA\_HOME=/usr/java/jdk1.7.0\_71/

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

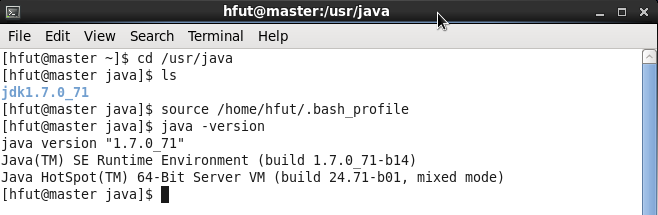
使改动生效命令

[hfut@mater java]$ source /home/hfut/.bash\_profile

测试配置：

[hfut@mater java]$ java -version

如果出现下图的信息，表示JDK安装成功：



**4.2.8 免密匙登录配置**

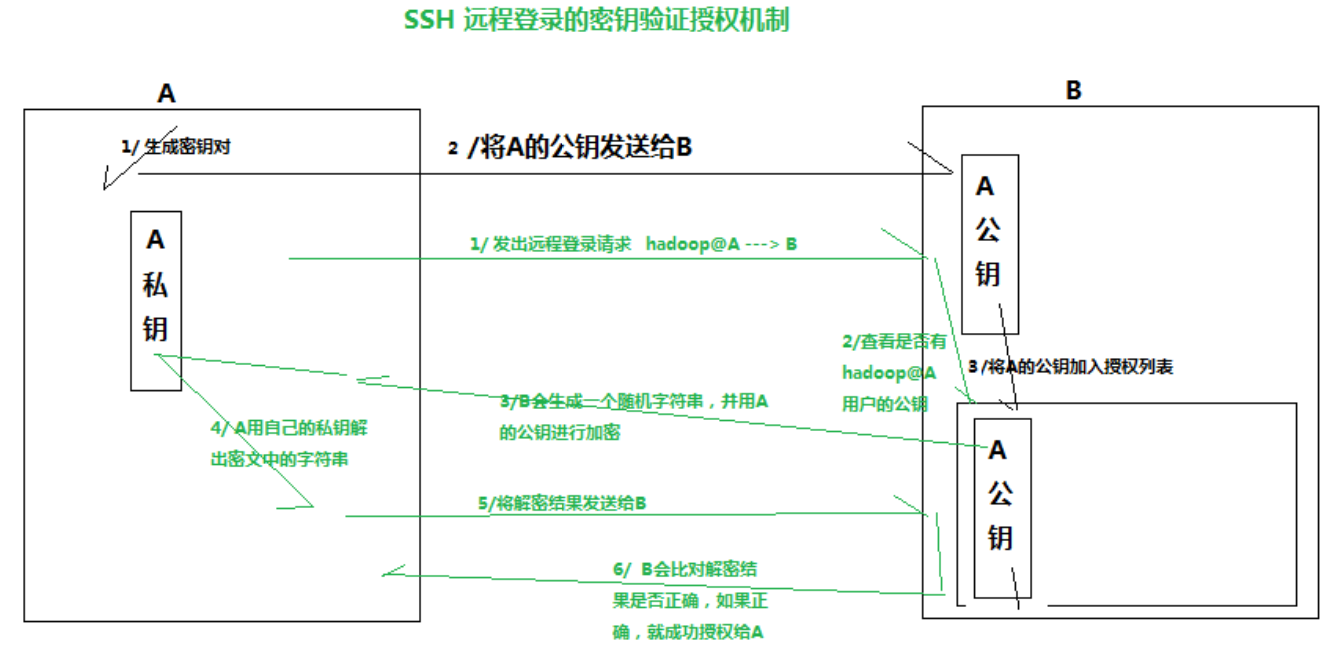
**该部分所有的操作都要在hfut用户下，切换到hfut用户的命令是：**

**su hfut**

**密码是hfut**

（1）SSH免密匙登录的原理与方法

**SSH免密匙登录的原理：**



**SSH免密匙登录的方法：**

配置主机A免密登录到主机B（方法一）

① 在主机A生产密钥对: ssh-keygen -t rsa， 会在.ssh目录下产生密钥文件

② 拷贝主机A的公钥到主机B: scp id\_rsa.pub

③ 将主机A的公钥加到主机B的授权列表.ssh/authorized\_keys（若不存在，手动创建）: cat id\_rsa.pub >> authorized\_keys

④ 授权列表authorized\_keys的权限必须是600，chmod 600 authorized\_keys（权限的设置非常重要，因为不安全的设置安全设置，会让你不能使用RSA功能）。另一件事是用root用户设置"/etc/ssh/sshd\_config"的内容。使其无密码登录有效。

（1）SSH免密匙登录配置

**HadoopMaster节点**

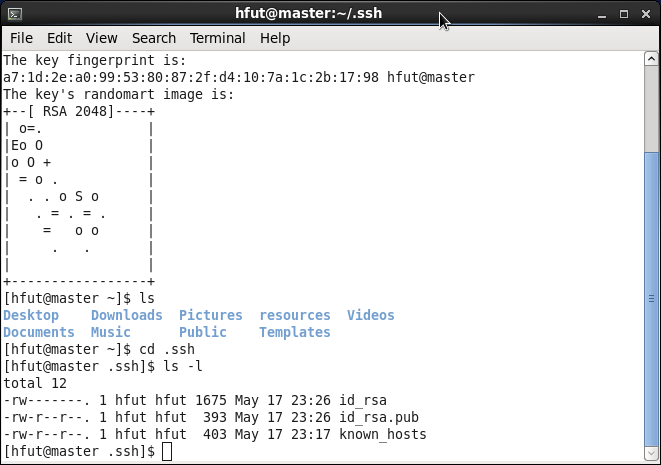
在终端生成密匙，命令如下（一路点击回车生成密匙）

[hfut@master ~]$ ssh-keygen -t rsa

生成的密匙在.ssh目录下，如下图所示：

[hfut@master ~]$ cd .ssh

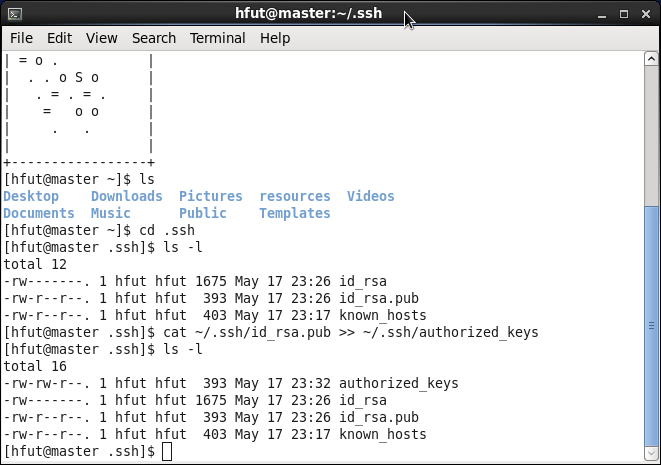
[hfut@master .ssh]$ ls -l



将公钥追加到授权列表

[hfut@master .ssh]$ cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

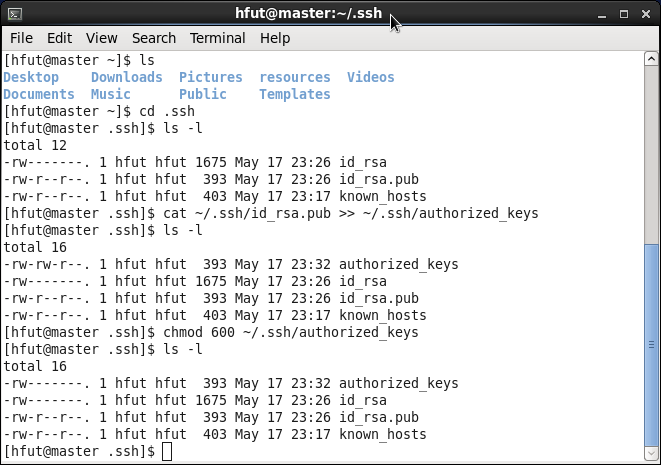
执行ls -l命令后会看到下图的文件列表：



修改authorized\_keys文件的权限，命令如下：

[hfut@master .ssh]$ chmod 600 ~/.ssh/authorized\_keys

修改完权限后，文件列表情况如下：



**HadoopMaster节点**

将authorized\_keys文件复制到slave节点，命令如下：

[hfut@master .ssh]$ scp ~/.ssh/authorized\_keys hfut@slave:~/.ssh

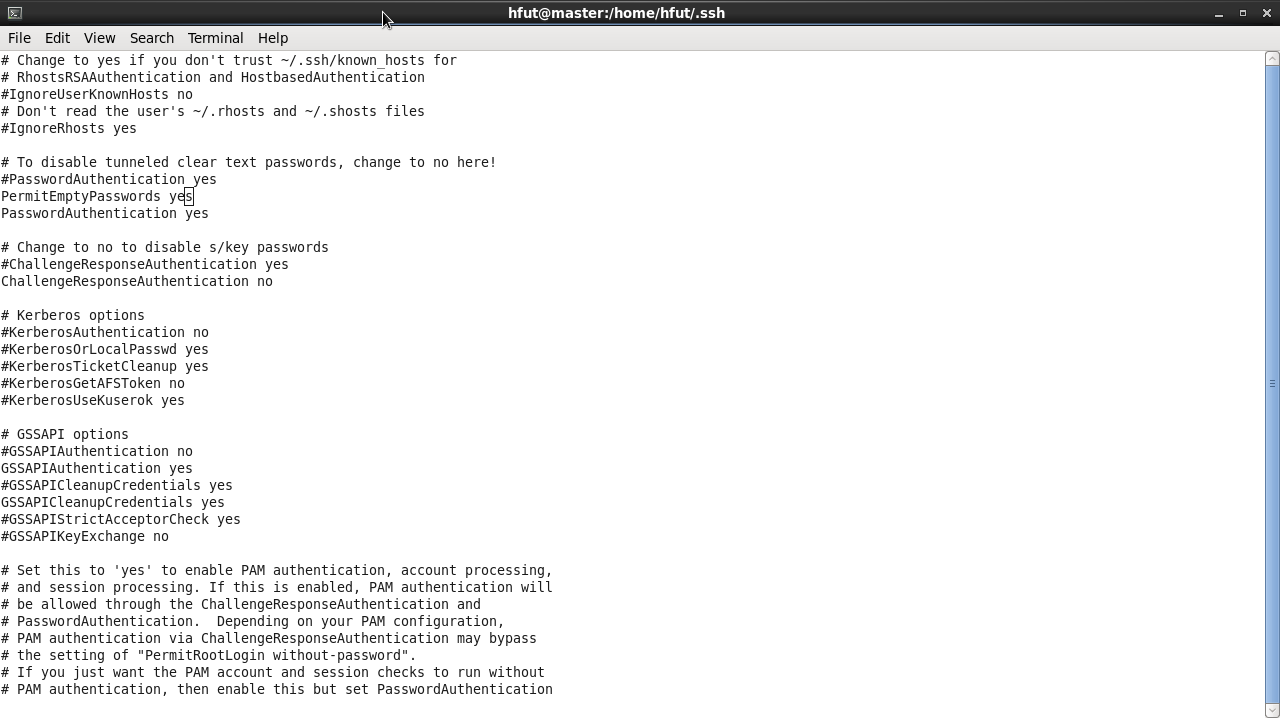
**修改HadoopMaster节点的SSH配置**

[hfut@master .ssh]$ su root

[root@master .ssh]# vi /etc/ssh/sshd\_config

设置PermitEmptyPasswords配置为

PermitEmptyPasswords yes



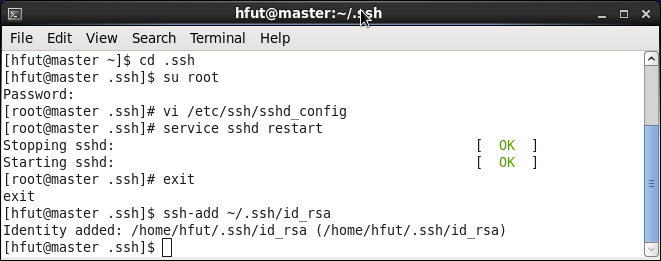
使配置生效

[root@master .ssh]# service sshd restart

使用 ssh-add指令将私钥加进来，解决“Agent admitted failure to sign using the key.”错误问题。

[root@master .ssh]# exit

[hfut@master .ssh]$ ssh-add ~/.ssh/id\_rsa



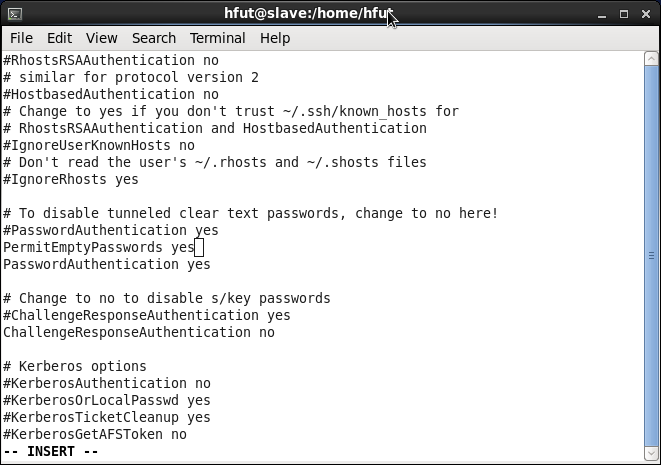
**修改HadoopSlave节点的SSHD配置**

[hfut@slave ~]$ su root

[root@slave .ssh]# vi /etc/ssh/sshd\_config

设置PermitEmptyPasswords配置为

PermitEmptyPasswords yes



使配置生效

[root@slave .ssh]# service sshd restart

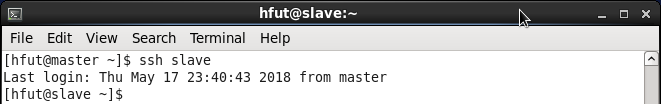


**验证免密匙登录**

在HadoopMaster机器上执行下面的命令：

[hfut@master ~]$ ssh slave

如果出现下图的内容表示免密匙配置成功：



**4.3 Hadoop配置部署**

每个节点上的Hadoop配置基本相同，在HadoopMaster节点操作，然后完成复制到其它节点。

下面所有的操作都要在hfut用户下，切换到hfut用户的命令是：

su hfut

密码是hfut

**4.3.1 Hadoop安装包解压**

复制并解压Hadoop软件包，命令如下：

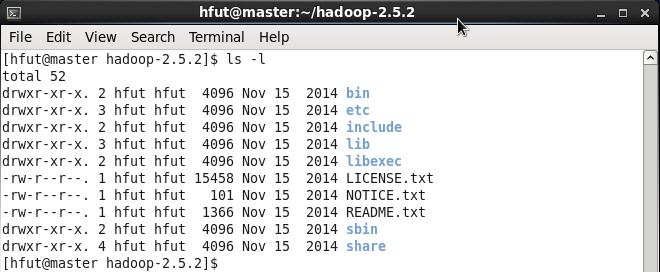
[hfut@master ~]$ cp ~/resources/software/hadoop/apache/hadoop-2.5.2.tar.gz ~/

[hfut@master ~]$ tar -zxvf hadoop-2.5.2.tar.gz

[hfut@master ~]$ rm hadoop-2.5.2.tar.gz

[hfut@master ~]$ cd hadoop-2.5.2

ls -l看到如下图的内容，表示解压成功：



**4.3.2 配置环境变量hadoop-env.sh**

环境变量文件中，只需要配置JDK的路径。

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/hadoop-env.sh

在文件靠前的部分找到下面的一行代码：

export JAVA\_HOME=${JAVA\_HOME}

将这行代码修改为下面的代码：

export JAVA\_HOME=/usr/java/jdk1.7.0\_71/

然后保存文件。

**4.3.3 配置环境变量yarn-env.sh**

环境变量文件中，只需要配置JDK的路径。

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/yarn-env.sh

在文件靠前的部分找到下面的一行代码：

# export JAVA\_HOME=/home/y/libexec/jdk1.6.0/

将这行代码修改为下面的代码（将#号也去掉）：

export JAVA\_HOME=/usr/java/jdk1.7.0\_71/

然后保存文件。

**4.3.4 配置核心组件core-site.xml**

使用gedit编辑：

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/core-site.xml

用下面的代码替换core-site.xml中的内容：

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

<property>

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

<value>/home/hfut/hadoopdata</value>

</property>

</configuration>

**4.3.5 配置文件系统hdfs-site.xml**

使用gedit编辑：

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/hdfs-site.xml

用下面的代码替换hdfs-site.xml中的内容：

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

**4.3.6 配置文件系统yarn-site.xml**

使用gedit编辑：

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/yarn-site.xml

用下面的代码替换yarn-site.xml中的内容：

<?xml version="1.0"?>

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<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.resource-tracker.address</name>

<value>master:18025</value>

</property>

<property>

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

<value>master:18141</value>

</property>

<property>

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

<value>master:18088</value>

</property>

</configuration>

**4.3.7 配置计算框架mapred-site.xml**

复制mapred-site.xml-template.xml文件：

[hfut@master Hadoop-2.5.2]$ cp ~/hadoop-2.5.2/etc/hadoop/mapred-site.xml.template ~/hadoop-2.5.2/etc/hadoop/mapred-site.xml

使用gedit编辑：

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/mapred-site.xml

用下面的代码替换mapred-site.xml中的内容：

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!--

Licensed under the Apache License, Version 2.0 (the "License");

you may not use this file except in compliance with the License.

You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software

distributed under the License is distributed on an "AS IS" BASIS,

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and

limitations under the License. See accompanying LICENSE file.

-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

**4.3.8 在master节点配置slaves文件**

使用gedit编辑：

[hfut@master Hadoop-2.5.2]$ gedit ~/hadoop-2.5.2/etc/hadoop/slaves

用下面的代码替换slaves中的内容：

slave

**4.3.9 复制到从节点**

使用下面的命令将已经配置完成的Hadoop复制到从节点HadoopSlave上：

[hfut@master Hadoop-2.5.2]$ cd

[hfut@master Hadoop-2.5.2]$ scp -r hadoop-2.5.2/ hfut@slave:~/

**4.4 启动Hadoop集群**

下面所有的操作都要在hfut用户下，切换到hfut用户的命令是：

su hfut

密码是hfut

**4.4.1 配置Hadoop启动的系统环境变量**

该节的配置需要同时在两个节点（HadoopMaster和HadoopSlave）上进行操作，操作命令如下：

[hfut@master Hadoop-2.5.2]$ cd

[hfut@master ~]$ gedit ~/.bash\_profile

将下面的代码追加到.bash\_profile末尾：

#HADOOP

export HADOOP\_HOME=/home/hfut/hadoop-2.5.2

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

然后执行命令：

[hfut@master ~]$ source ~/.bash\_profile

**4.4.2 创建数据目录**

该节的配置需要同时在两个节点（HadoopMaster和HadoopSlave）上进行操作，操作命令如下：

在hfut的用户主目录下，创建数据目录，命令如下：

[hfut@master ~]$ mkdir /home/hfut/hadoopdata

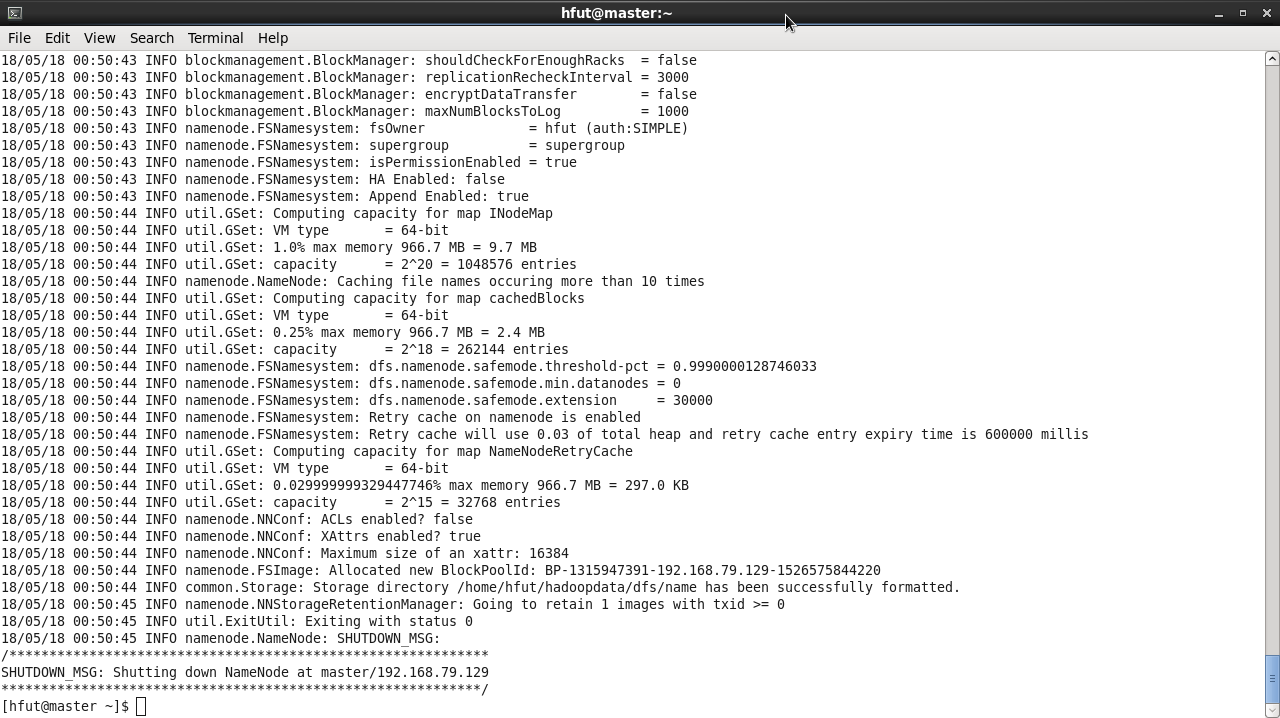
**4.4.3 启动Hadoop集群**

（1）格式化文件系统

格式化命令如下，该操作需要在HadoopMaster节点上执行：

[hfut@master ~]$ hdfs namenode -format

看到下图的打印信息表示格式化成功，如果出现Exception/Error，则表示出现问题。



（2）启动Hadoop

使用start-all.sh启动Hadoop集群，首先进入Hadoop安装主目录，然后执行启动命令：

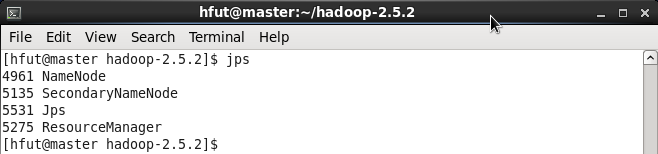
[hfut@master ~]$ cd ~/hadoop-2.5.2

[hfut@master ~]$ sbin/start-all.sh

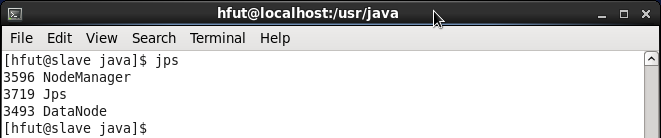
执行命令后，提示输入yes/no时，输入yes。

（3）查看进程是否启动

在HadoopMaster的终端执行jps命令，在打印结果中会看到4个进程，分别是ResourceManager、Jps、NameNode和SecondaryNameNode，如下图所示。如果出现这4个进程表示主节点进程启动成功。

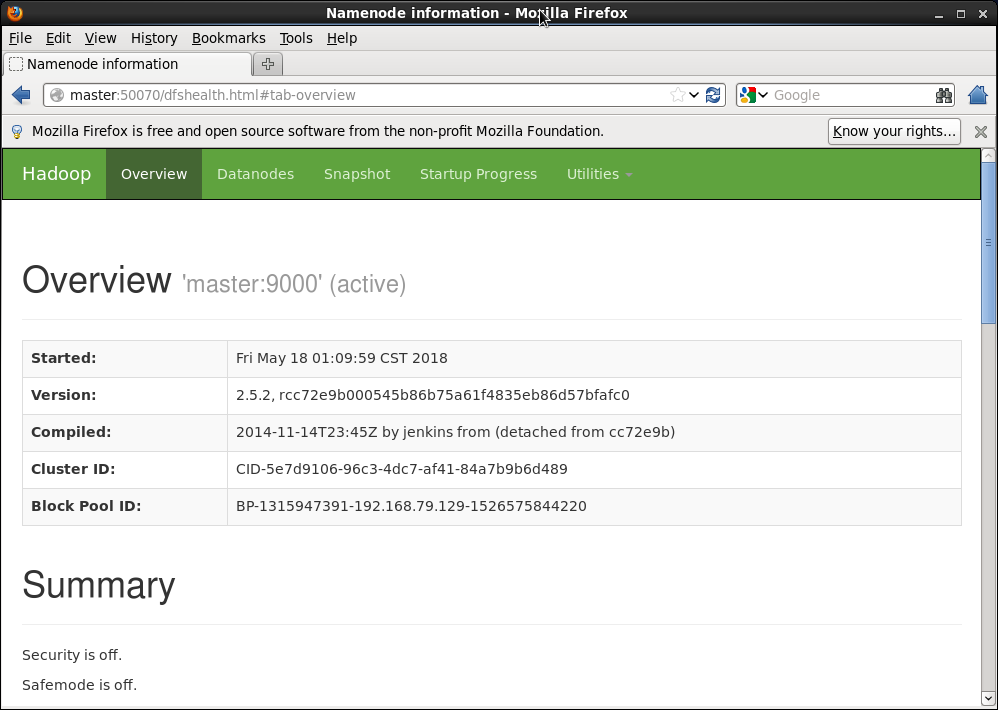


在HadoopSlave的终端执行jps命令，在打印结果中会看到3个进程，分别是Nodemanager、Datanode和Jps，如下图所示。如果出现这3个进程表示主节点进程启动成功。

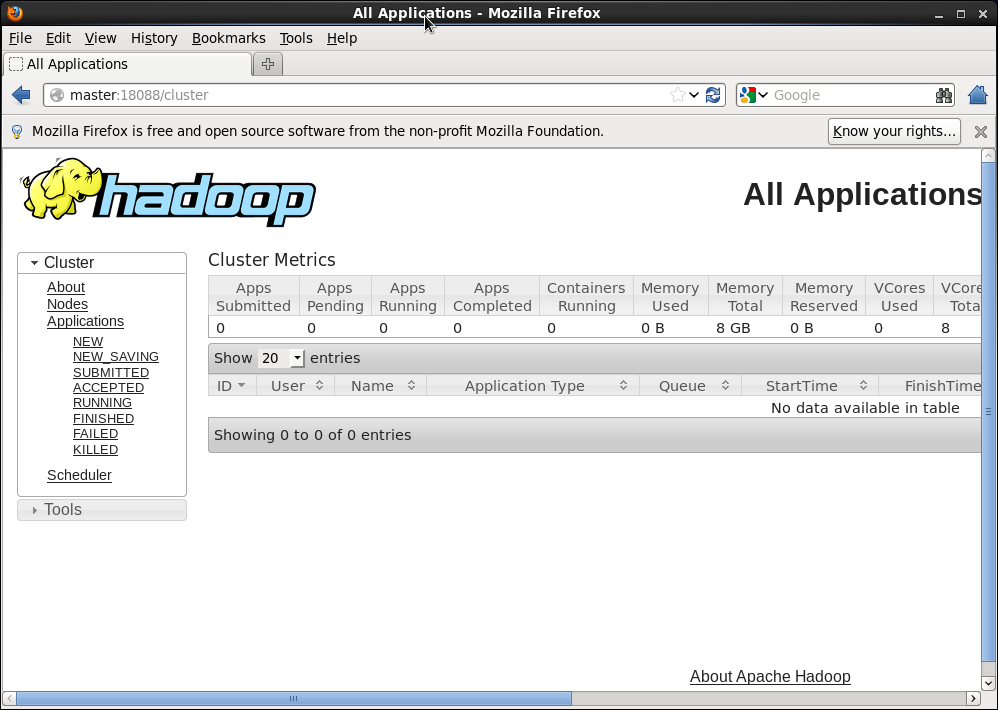


（4）Web UI查看集群是否成功启动

在HadoopMaster上启动Firefox浏览器，在浏览器地址栏中输入<http://master:50070/>，检查namenode和datanode是否正常。UI页面如下图所示。



在HadoopMaster上启动Firefox浏览器，在浏览器地址栏中输入<http://master:18088/>，检查Yarn是否正常。页面如下图所示。



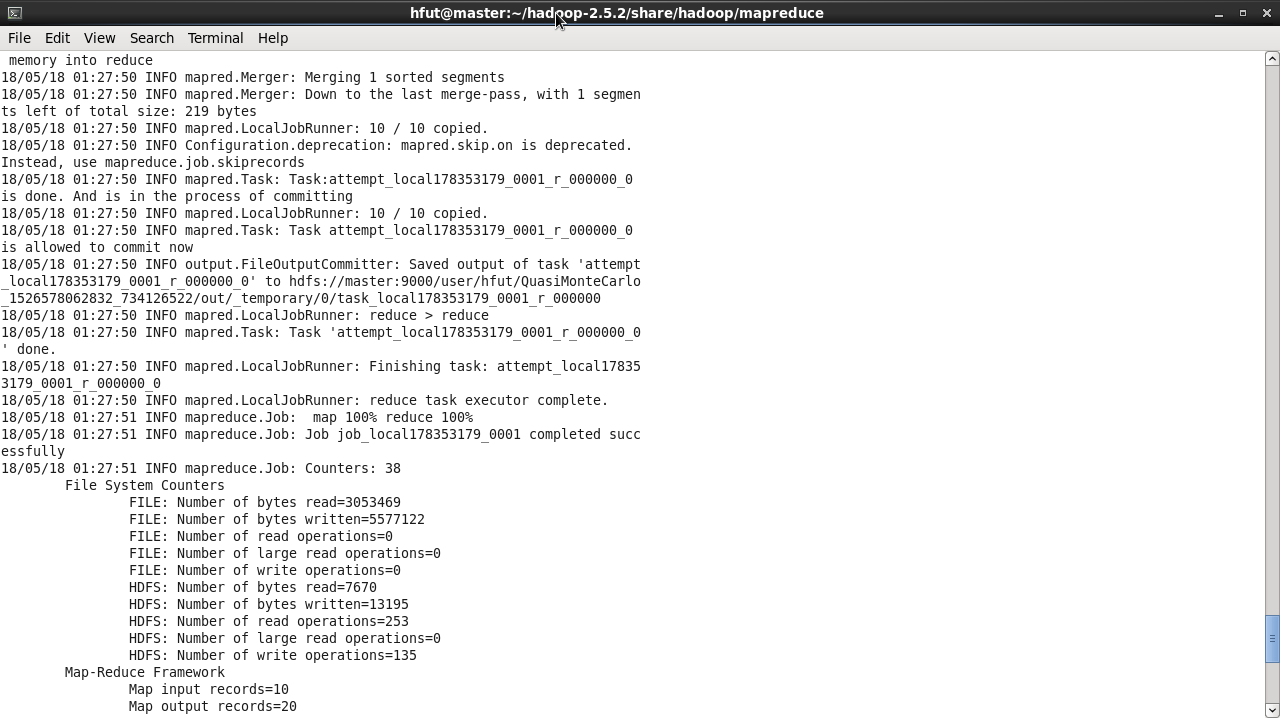
（5）运行PI实例检查Hadoop集群是否搭建成功

进入Hadoop安装主目录，执行下面的命令：

[hfut@master ~]$ cd ~/hadoop-2.5.2/share/hadoop/mapreduce/

[hfut@master mapreduce]$ hadoop jar ~/hadoop-2.5.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.2.jar pi 10 10

会看到如下的执行结果：



最后输出：

Estimated value of Pi is 3.20000000000000000000

如果以上的3个验证步骤都没有问题，说明集群正常启动。