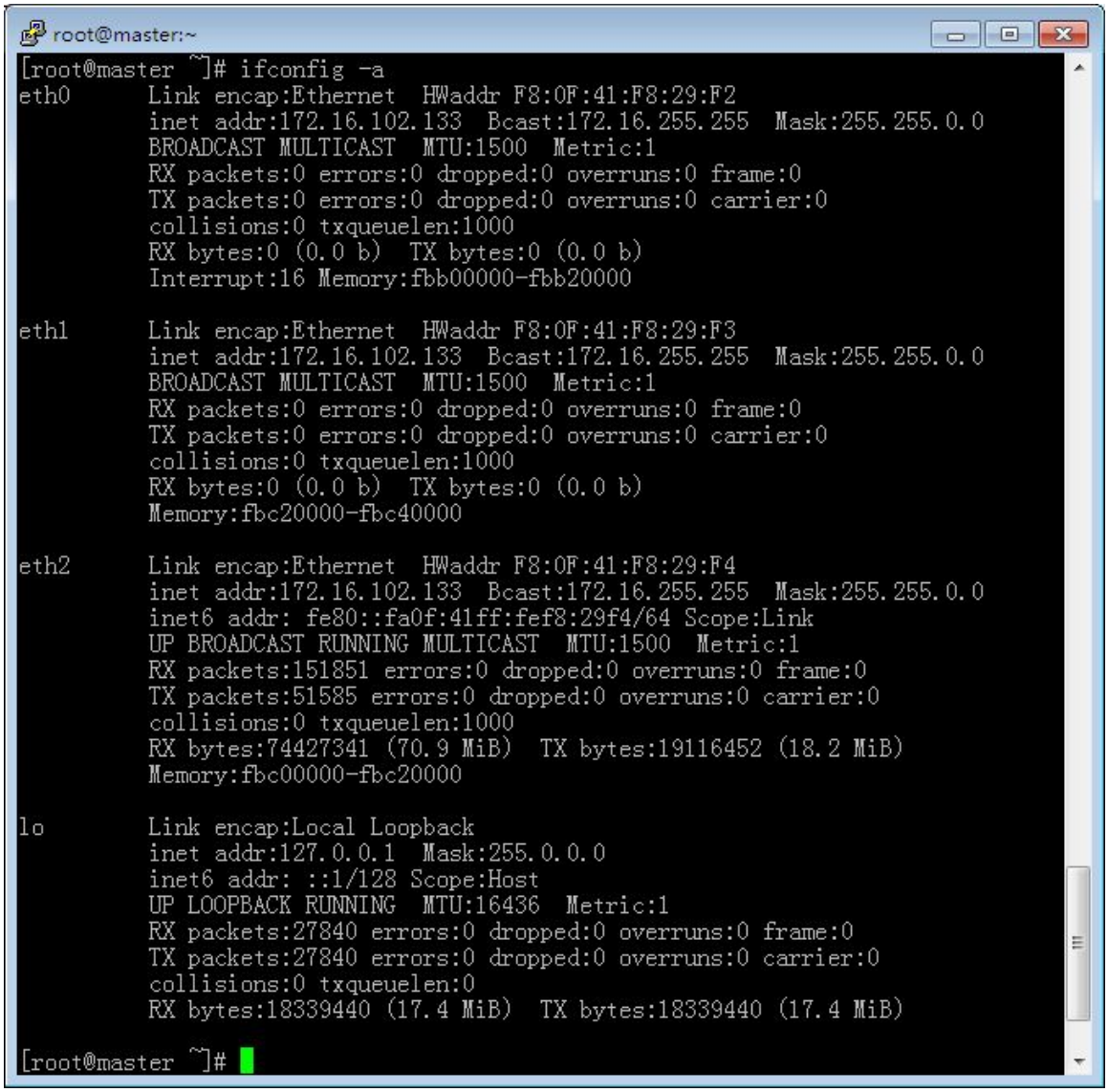
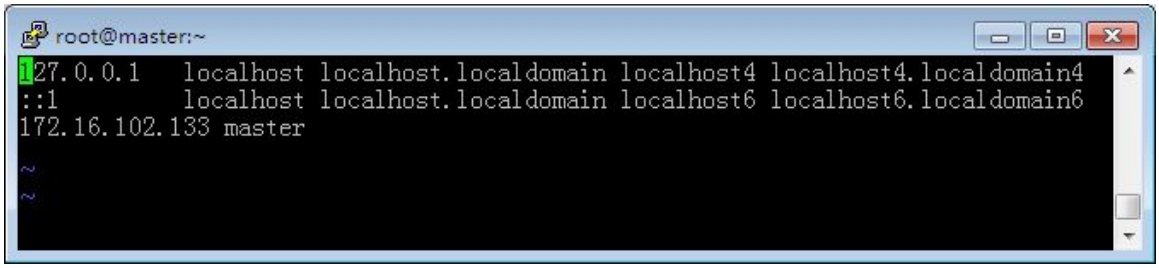
Jdk&hadoop 配置部分

步骤一、配置 hosts 将主机名与 ip 对应

ifconfig -a　　#在 master 和所有虚拟机中运行，获得 master 的虚拟网卡 VMnet8 的 ip 和所有 slaver 的 ip



vi /etc/hosts　　#该文件只需先在 master 上配置，格式：192.168.***.***



步骤二、安装 JDK，添加环境变量

#如果 master 和 slaver 架构不同请分别安装,若相同可以最后上传至 slaver

#下载 jdk-6u45-linux-x64.bin 放在/usr/local 里

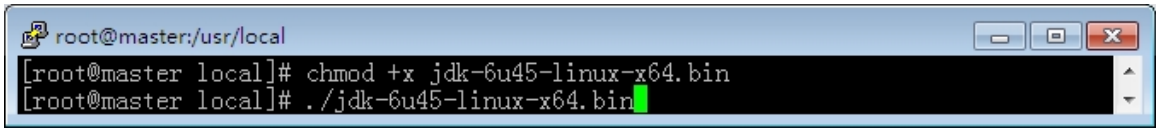
#因为我们虚拟机是 32 位，所以虚拟机中使用 jdk-6u45-linux-i586.bin

chmod +x jdk-6u45-linux-x64.bin

#添加可执行权限

./jdk-6u45-linux-x64.bin

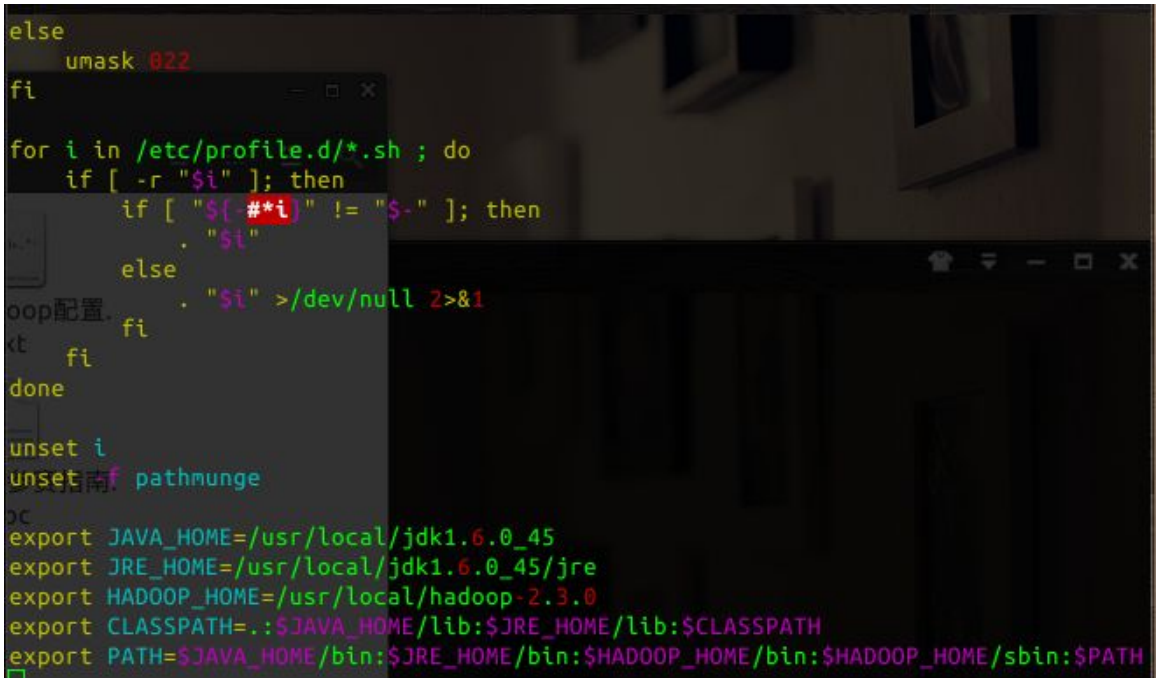
#执行程序，自动解压到当前目录



vi /etc/profile #设置环境变量，在文件最后添加一下内容

```
export JAVA_HOME=/usr/local/java/jdk1.6.0_45
export JRE_HOME=/usr/local/java/jdk1.6.0_45/jre
export CLASSPATH=.:$JAVA_HOME/lib:$JRE_HOME/lib:$CLASSPATH
export PATH=$PATH:$JAVA_HOME/bin:$JRE_HOME/bin:$JAVA_HOME
```

#注：为了以后集群工作的方便，这里建议每台机器的 java 最好一致



source /etc/profile #使 profile 配置文件立即生效（免去了重启）

java -version

javac -version

#验证是否成功,失败可能由于系统预装 openjdk，需要配置默认 jdk

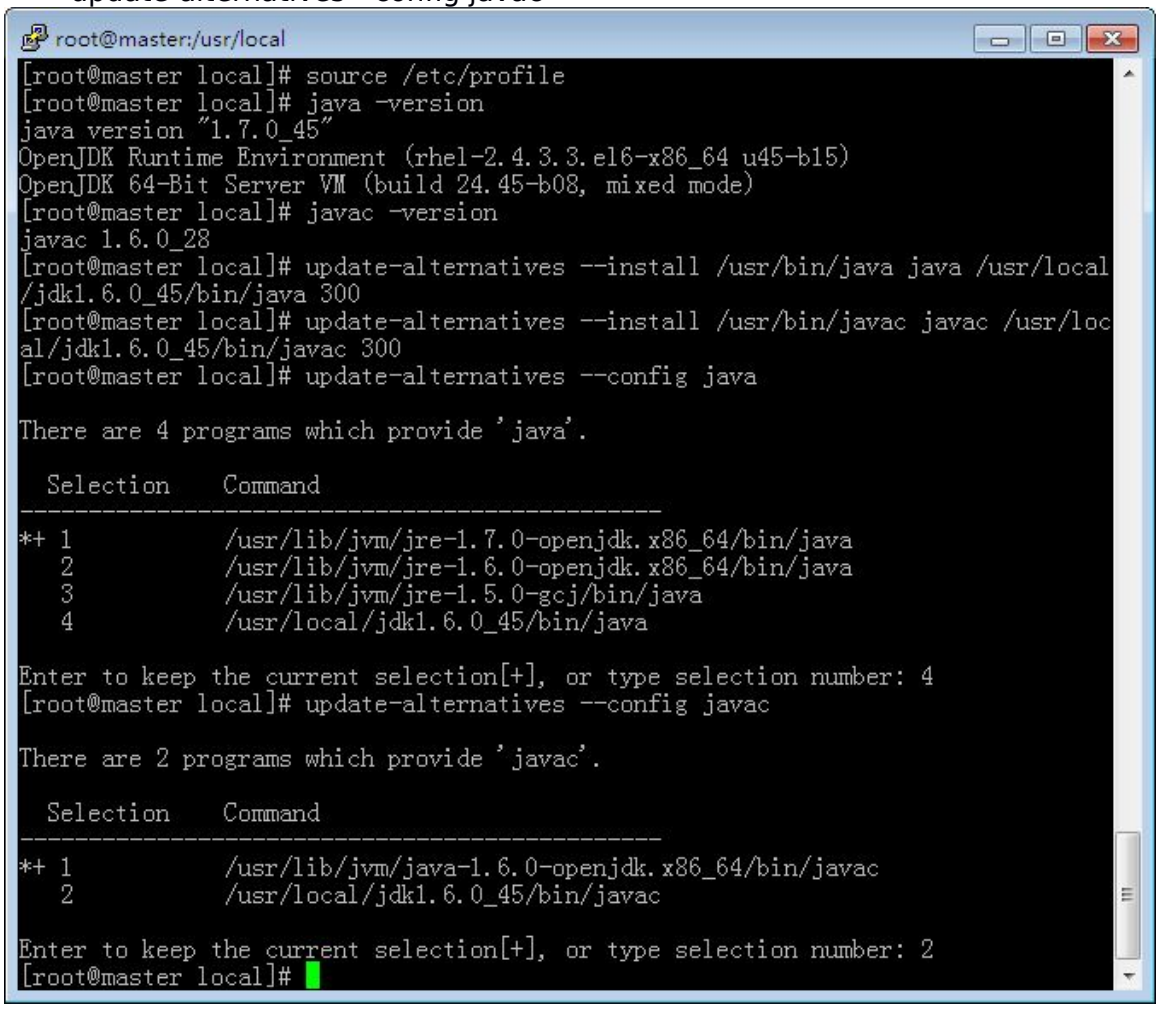
配置默认 jdk：

update-alternatives --install /usr/bin/java java /usr/local/jdk1.6.0_45/bin/java 300

update-alternatives --install /usr/bin/javac javac /usr/local/jdk1.6.0_45/bin/javac 300

update-alternatives --config java

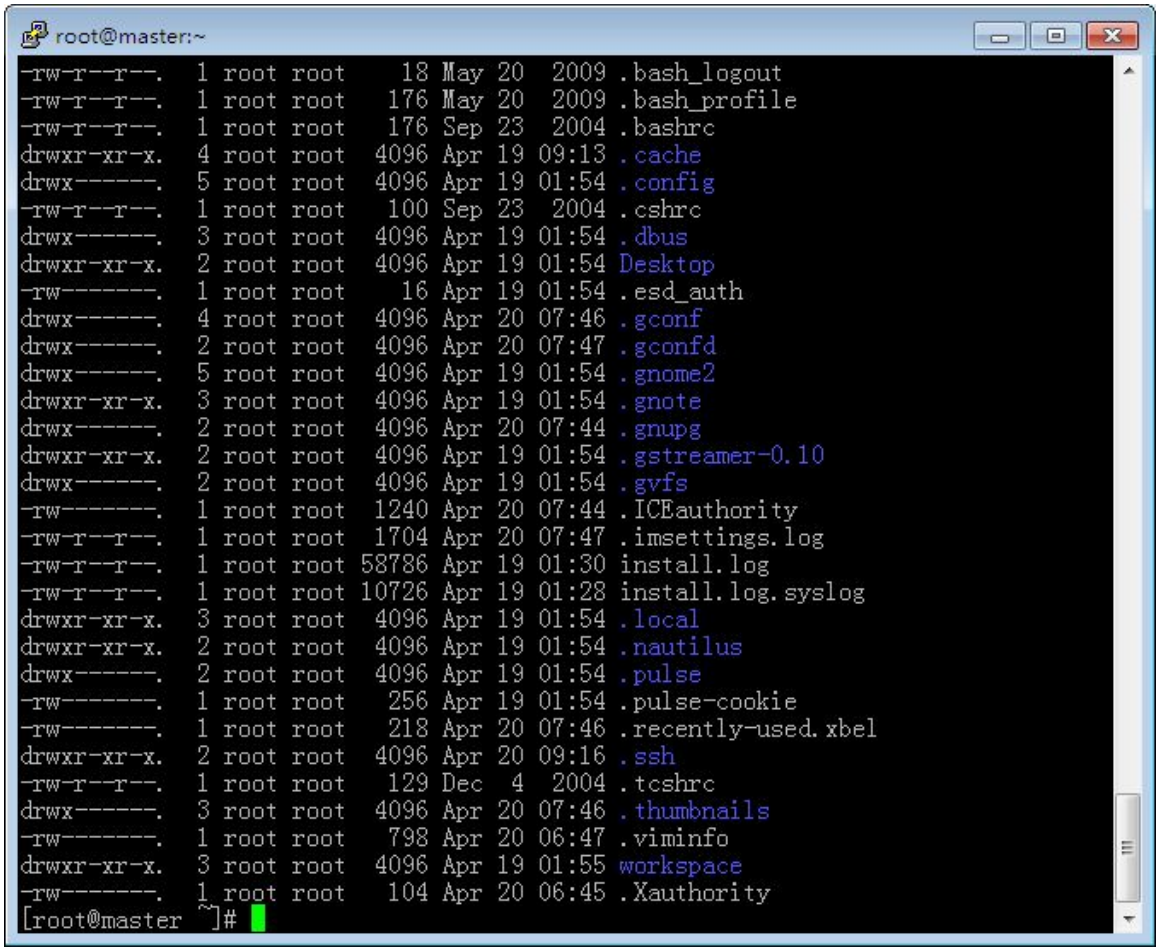
update-alternatives --config javac



步骤三、配置 ssh 无密码登陆（先只在 master 上配置）

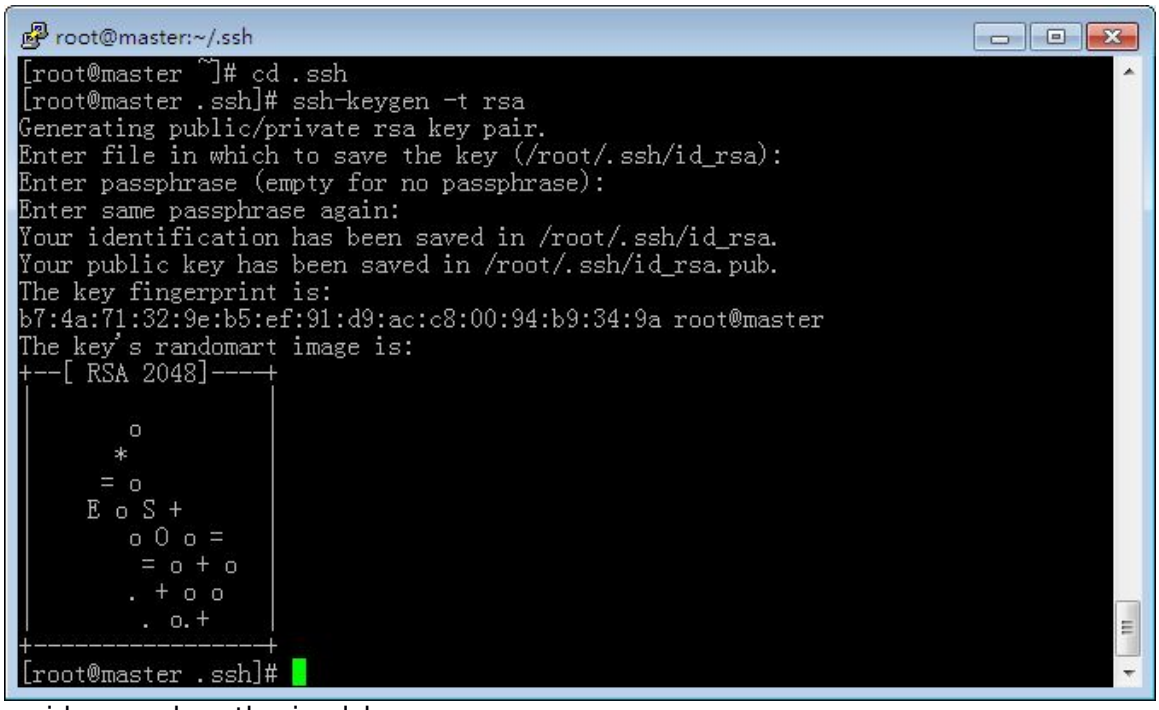
cd　　#进入当前用户主目录（master 的）

ls -al　　#查看有无.ssh 目录，如果没有请 mkdir .ssh 新建该目录



cd .ssh #进入.ssh 目录

ssh-keygen -t rsa #生成密钥，过程中一直<Enter>即可,生成的密钥对保存在 id_rsa 中

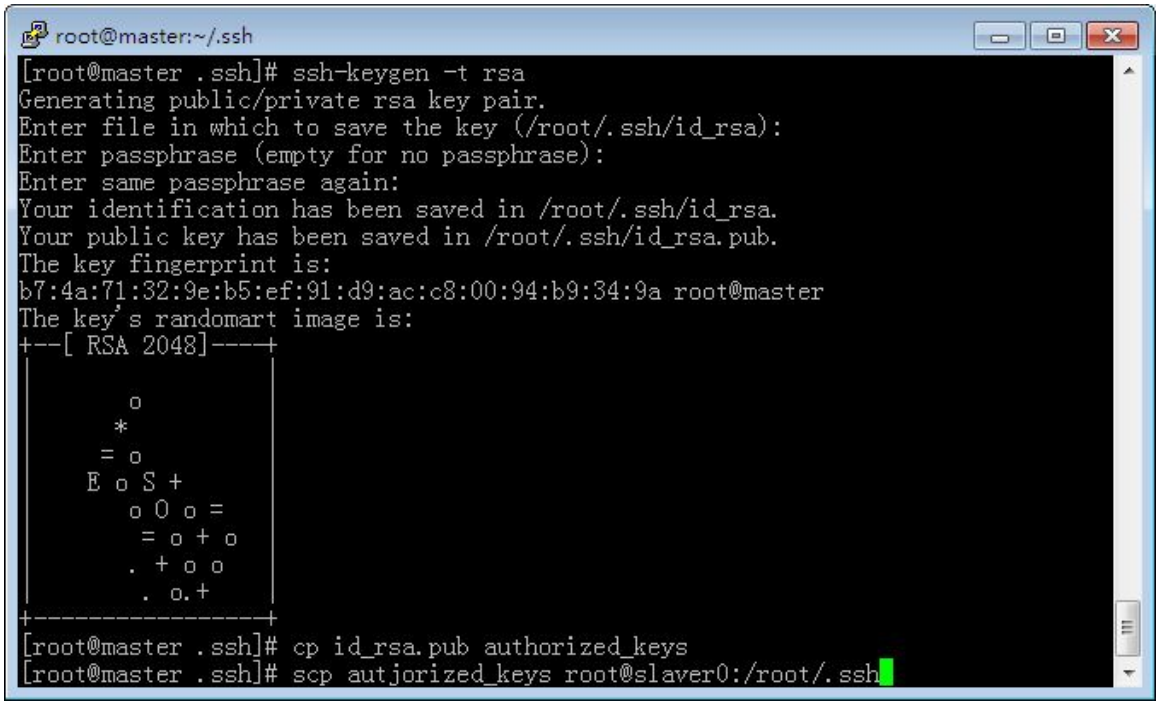


cp id_rsa.pub authorized_keys

#把生成的公钥追加到 authorized_keys 中，用于实现无密码登陆自己

scp authorized_keys root@slaver0:/root/.ssh

#这里对应自己各个 slaver 上运行 hadoop 的用户



#如果上面的方法失败，尝试以下操作来开启 RSA 认证 ：

vi /etc/ssh/sshd_config

RSAAuthentication　　#设为　yes
PubkeyAuthentication　　#设为　yes
AuthorizedKeysFile　　#设为　.ssh/authorized_keys

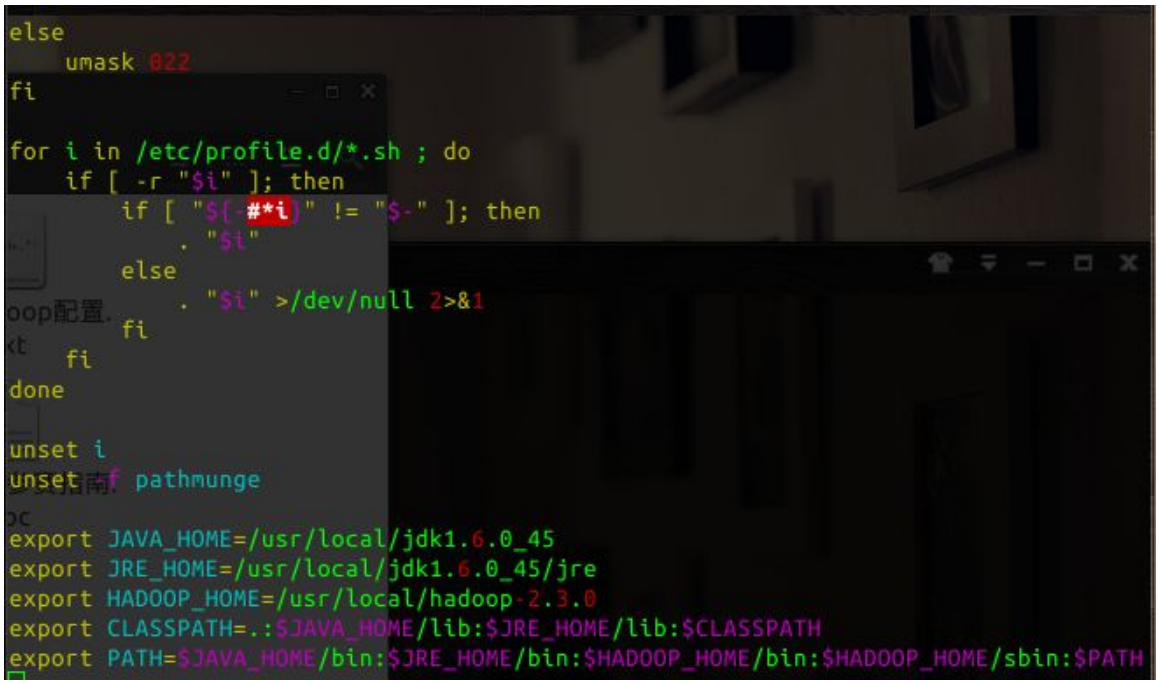
service　sshd restart　　#重启 ssh 服务

步骤四、hadoop 配置

下载 hadoop-2.3.0，解压后放入/usr/local

1.添加环境变量

vi /etc/profile　　#在文件的最后加入以下内容
export　HADOOP_HOME=/usr/local/hadoop-2.3.0
export　PATH=$PATH:$JAVA_HOME/bin:$JRE_HOME/bin:
　　$HADOOP_HOME/bin:$HADOOP_HOME/sbin
　　#PATH 里添加 hadoop 的两项即可



2.配置数个文件　　#配置文件所在目录为/usr/local/hadoop2.x.x/etc/hadoop

文件 1：core-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
　　<property>
```

```
    <!-- 当前集群 NameNode 的 IP 地址（使用 master 代替）和端口号。2.0 前使用 fs.default.name，
    但后续兼容-->
        <name>fs.defaultFS</name>
        <value>hdfs://master:9000</value>
    </property>

    <property>
    <!-- 设置临时文件目录 -->
        <name>hadoop.tmp.dir</name>
        <!-- 当前用户须要对此目录有读写权限。可使用命令 sudo chown -hR [user] /home/hadoop/ -->
        <value>/home/hadoop/hadoop-temp</value>
        <description>Abase for other temporary directories.</description>
    </property>

    <property>
        <name>io.file.buffer.size</name>
        <value>4096</value>
    </property>

</configuration>
```

文件 2：hdfs-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
    <property>
    <!-- SecondaryNamenode 网络地址,这里使用 master 代替-->
        <name>dfs.namenode.secondary.http-address</name>
        <value>master:50090</value>
    </property>

    <property>
    <!-- NameNode 工作目录，须预先存在 -->
        <name>dfs.namenode.name.dir</name>
        <value>file:/home/hadoop/dfs-name</value>
    </property>

    <property>
    <!-- DataNode 工作目录 -->
        <name>dfs.datanode.data.dir</name>
        <value>file:/home/hadoop/dfs-data</value>
    </property>

    <property>
```

```
    <!-- 文件（副本）的存储数量 -->
            <name>dfs.replication</name>
            <!-- 小于或等于附属机数量。默认 3 -->
            <value>3</value>
    </property>

    <property>
    <!-- 可以从网页端监控 hdfs -->
            <name>dfs.webhdfs.enabled</name>
            <value>true</value>
    </property>

    <property>
          <name>dfs.nameservices</name>
          <value>hadoop-cluster1</value>
    </property>

</configuration>
```

文件 3：mapred-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
    <property>
    <!-- map-reduce 运行框架 -->
            <name>mapreduce.framework.name</name>
            <!-- yarn:分布式模式 -->
            <value>yarn</value>
            <final>true</final>
    </property>

    <property>
            <name>mapreduce.jobtracker.http.address</name>
            <value>master:50030</value>
    </property>

    <property>
            <name>mapreduce.jobhistory.address</name>
            <value>master:10020</value>
    </property>

    <property>
            <name>mapreduce.jobhistory.webapp.address</name>
            <value>master:19888</value>
    </property>

    <property>
```

```
                <name>mapred.job.tracker</name>
                <value>http://master:9001</value>
        </property>

</configuration>
```

文件 4：yarn-site.xml

```
<?xml version="1.0"?>
<configuration>
        <!-- Site specific YARN configuration properties -->
        <property>
                <name>yarn.resourcemanager.hostname</name>
                <value>master</value>
        </property>

       <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
       </property>

       <property>
                <name>yarn.resourcemanager.address</name>
                <value>master:8032</value>
       </property>

        <property>
                <name>yarn.resourcemanager.scheduler.address</name>
                <value>master:8030</value>
       </property>

       <property>
                <name>yarn.resourcemanager.resource-tracker.address</name>
                <value>master:8031</value>
       </property>

       <property>
                <name>yarn.resourcemanager.admin.address</name>
                <value>master:8033</value>
       </property>

       <property>
                <name>yarn.resourcemanager.webapp.address</name>
                <value>master:8088</value>
       </property>
</configuration>
```

文件 5：yarn-env.sh

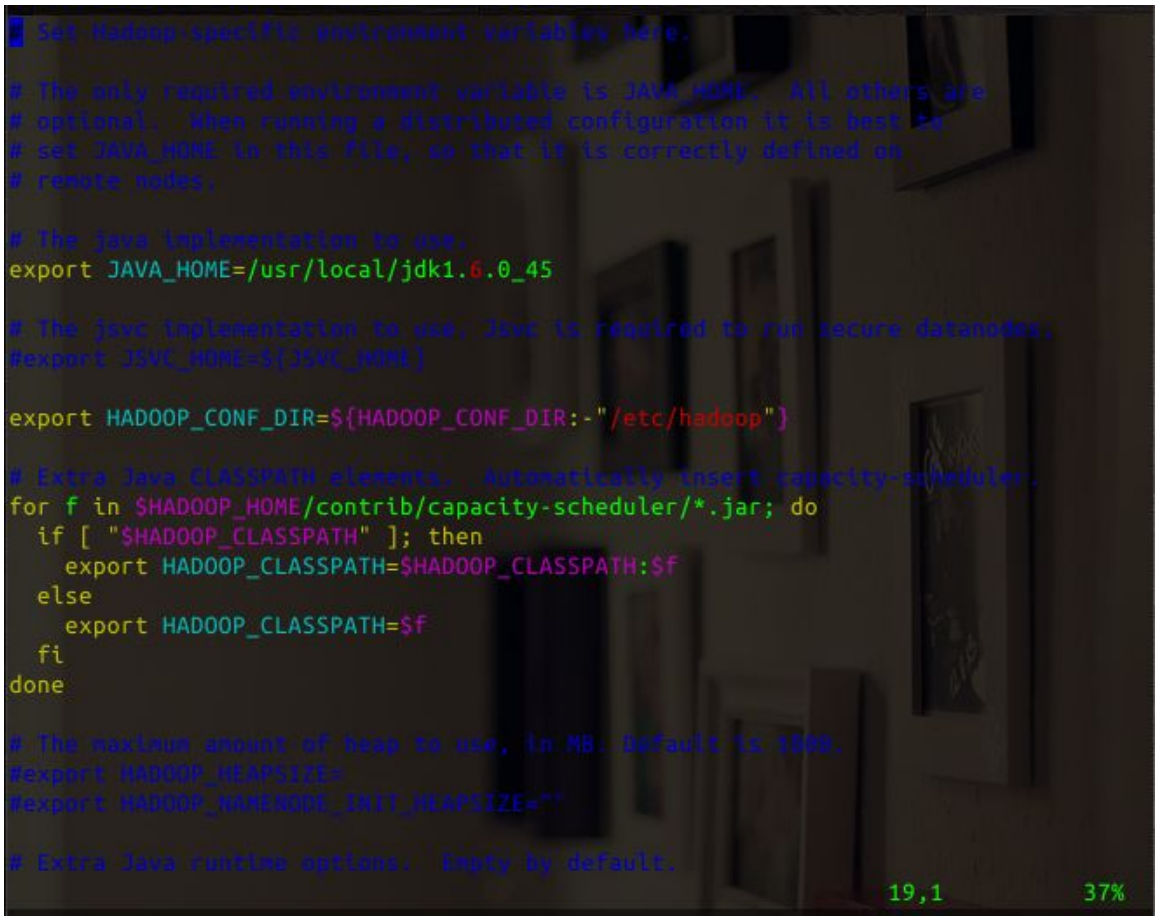
export JAVA_HOME=/usr/local//jdk1.6.0_45

#在文件中“some Java parameters”下面一行把 java 路径写完整

文件 6：hadoop-env.sh

export JAVA_HOME=/usr/local//jdk1.6.0_45

#在文件中“java implementation”下面一行把 java 路径写完整



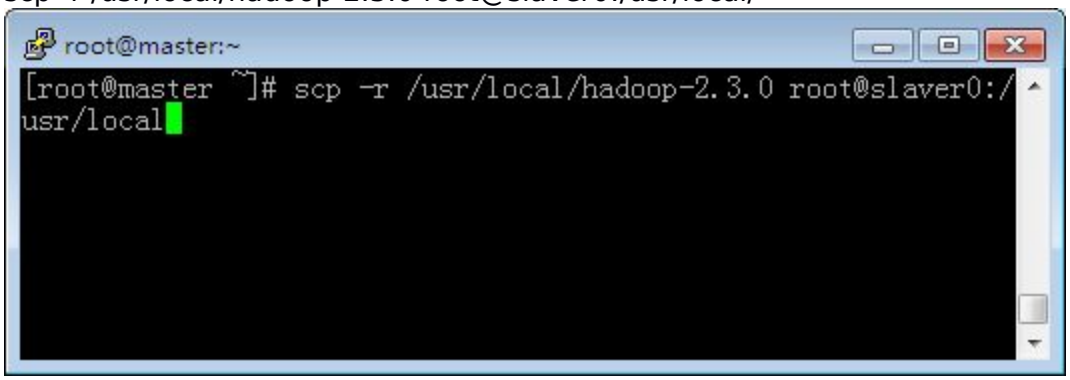
文件 7：slaves

#这里保存的是全部 slaver 的主机名

slaver0
slaver1
slaver2

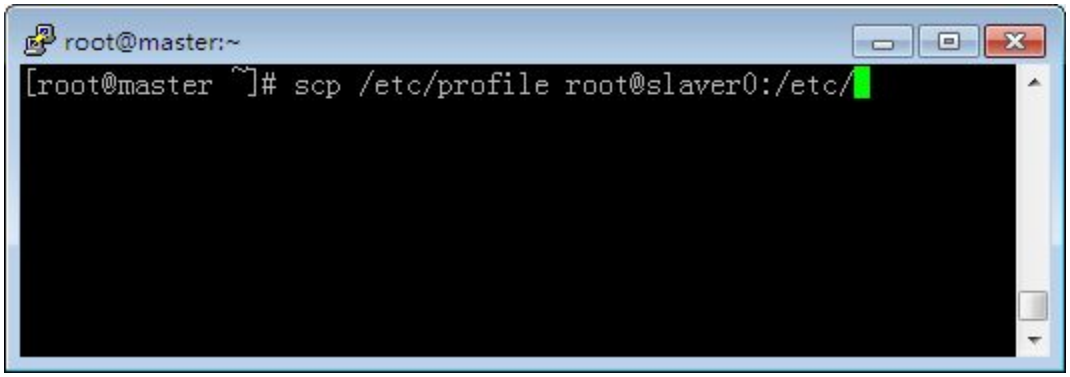
步骤五、上传 hadoop、profile 至 slaver

scp -r /usr/local/hadoop-2.3.0 root@slaver0:/usr/local/



scp /etc/profile root@slaver0:/etc/

#分别上传至多个 slaver

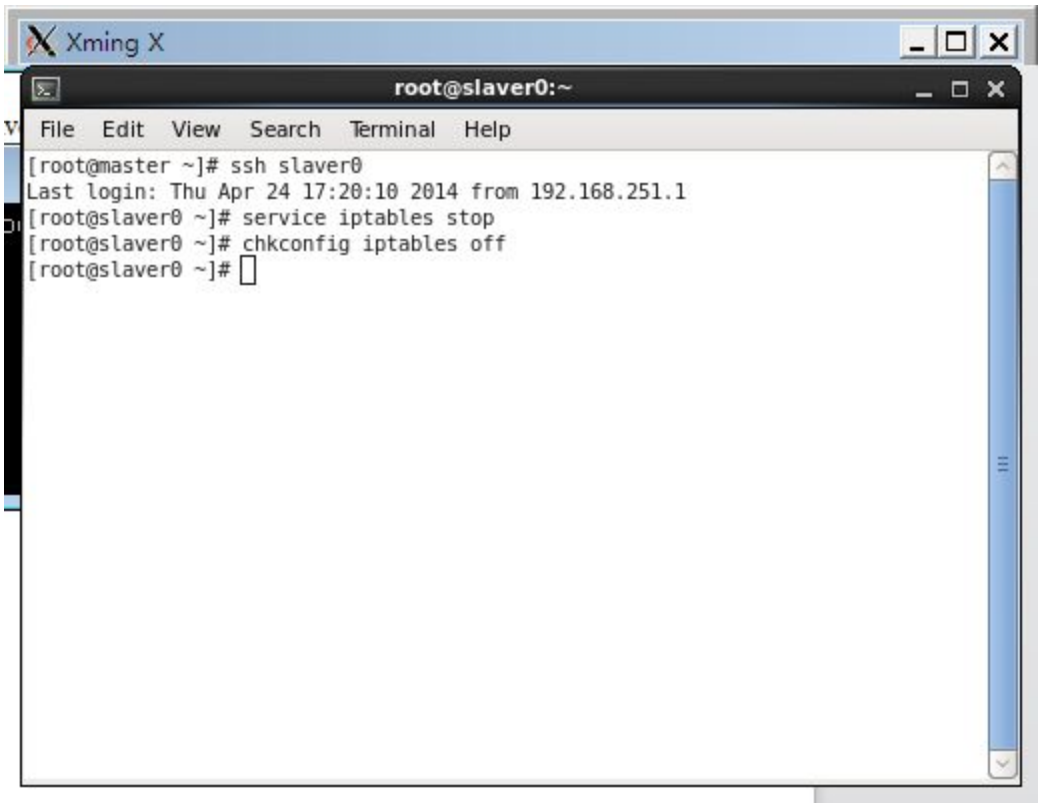


步骤六、开始初始化并运行

1.关闭 master 以及各个 slaver 的防火墙

service iptables stop　　#关闭防火墙

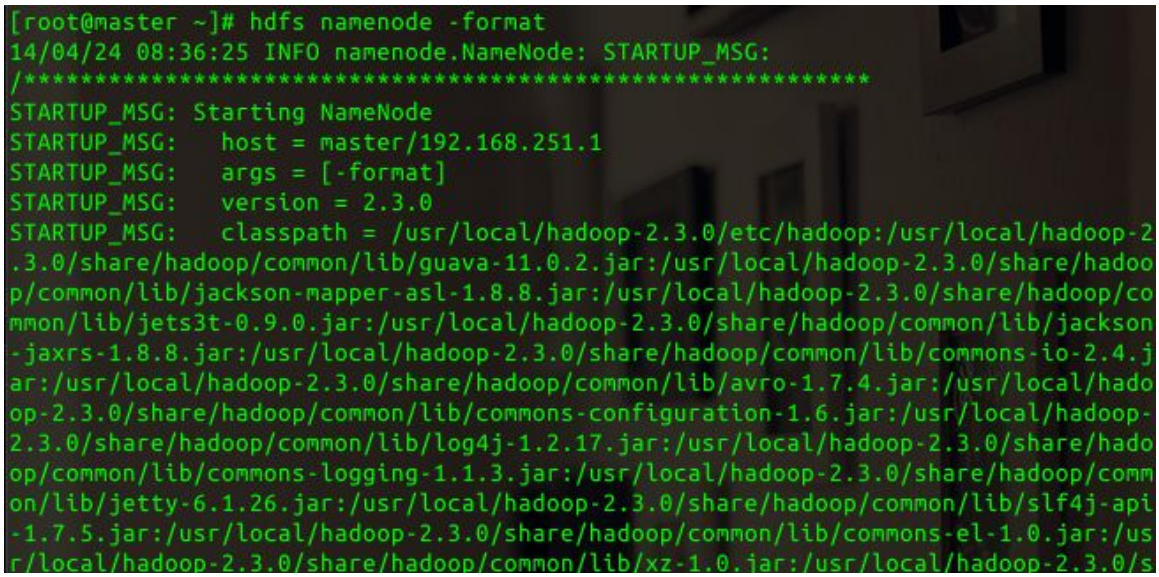
chkconfig iptables off　　#关闭防火墙服务

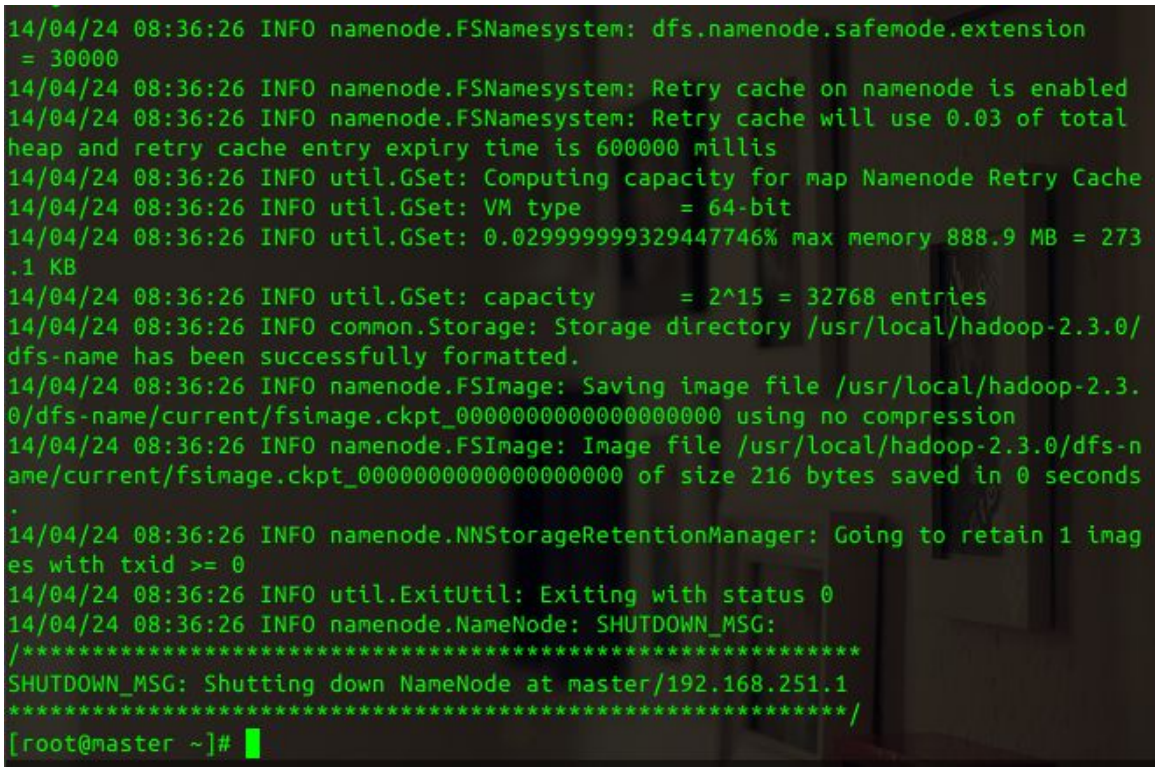


4.格式化 HDFS 系统

hdfs namenode -format 或 hadoop namenode -format

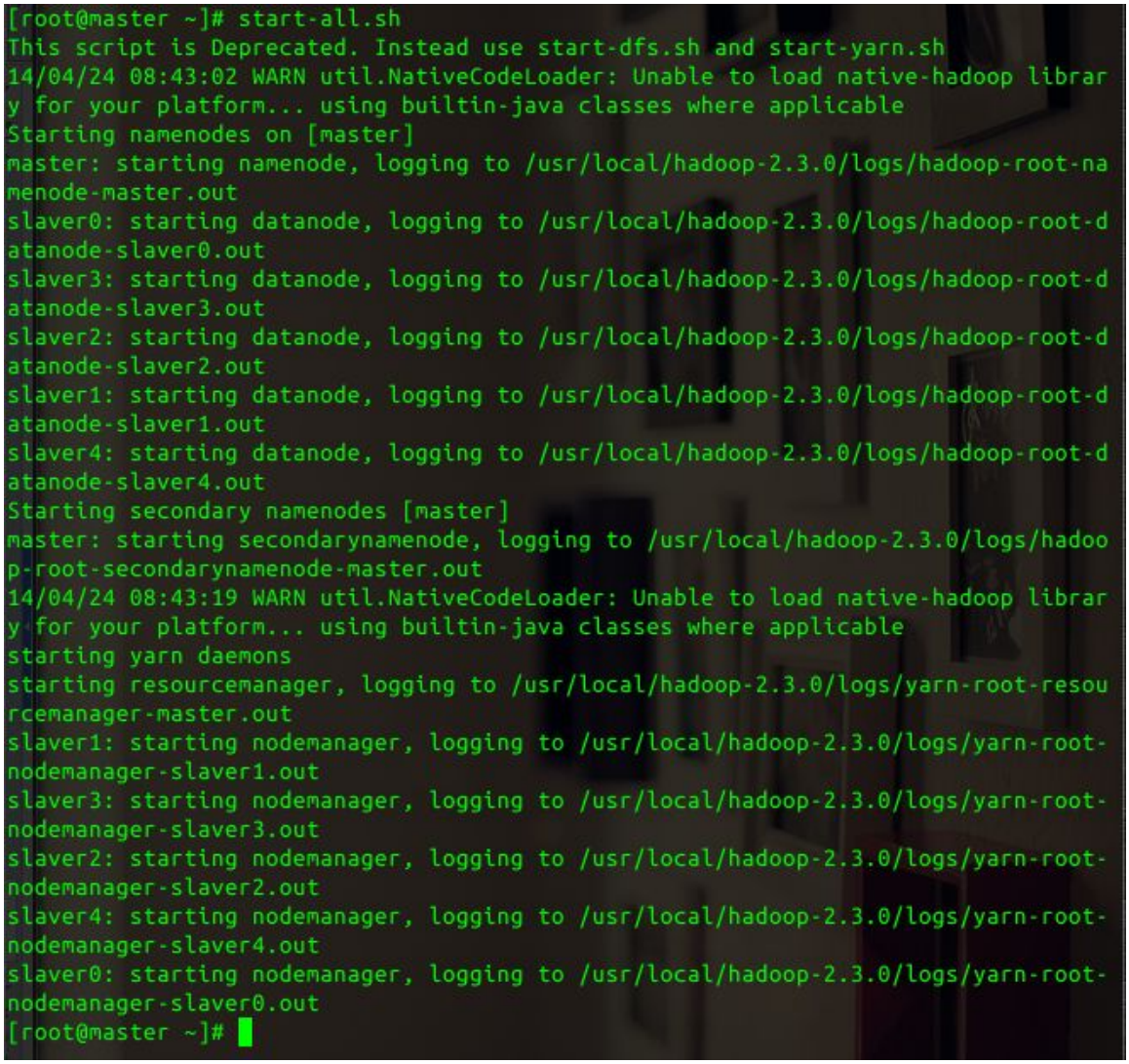
#注：不可以有任何 warning 或 error





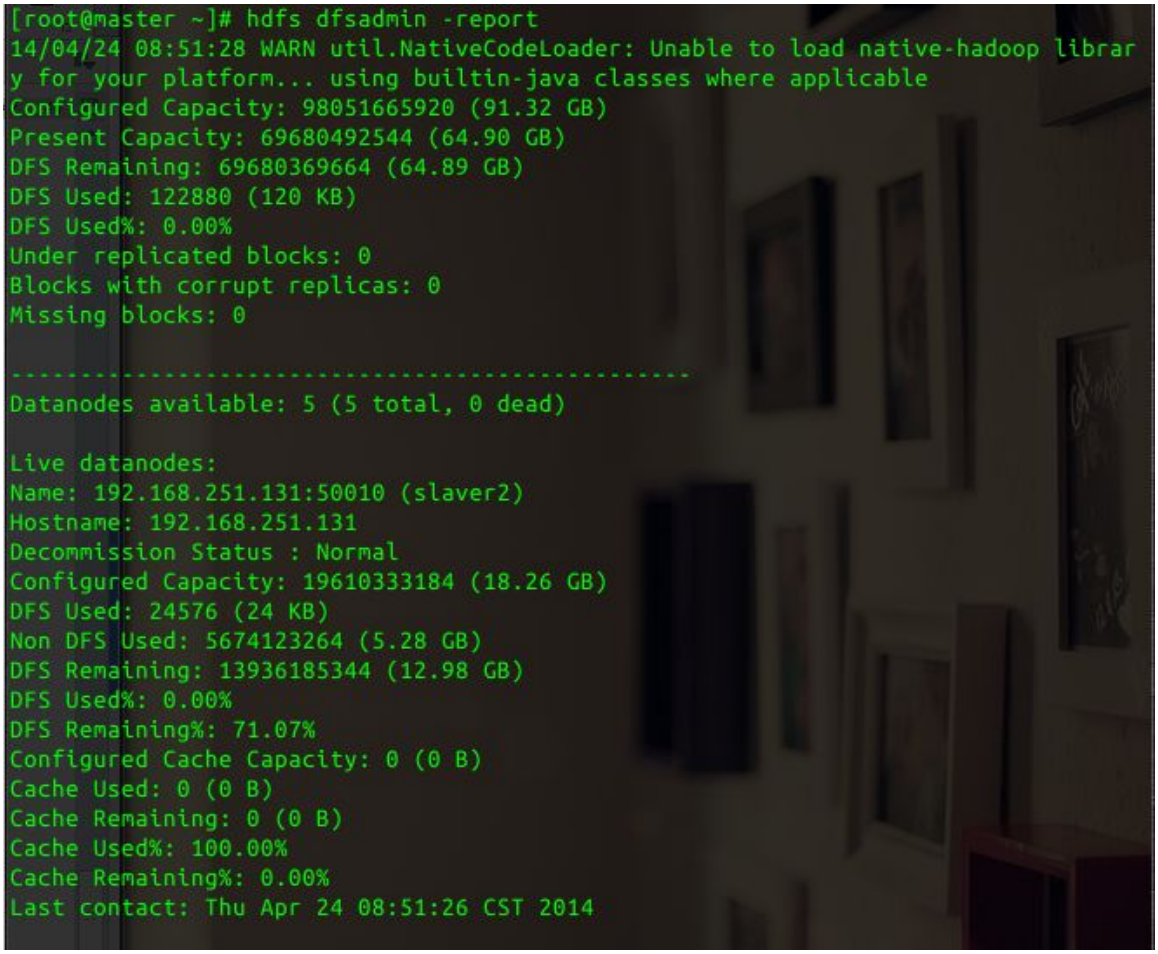
5.启动集群

start-all.sh

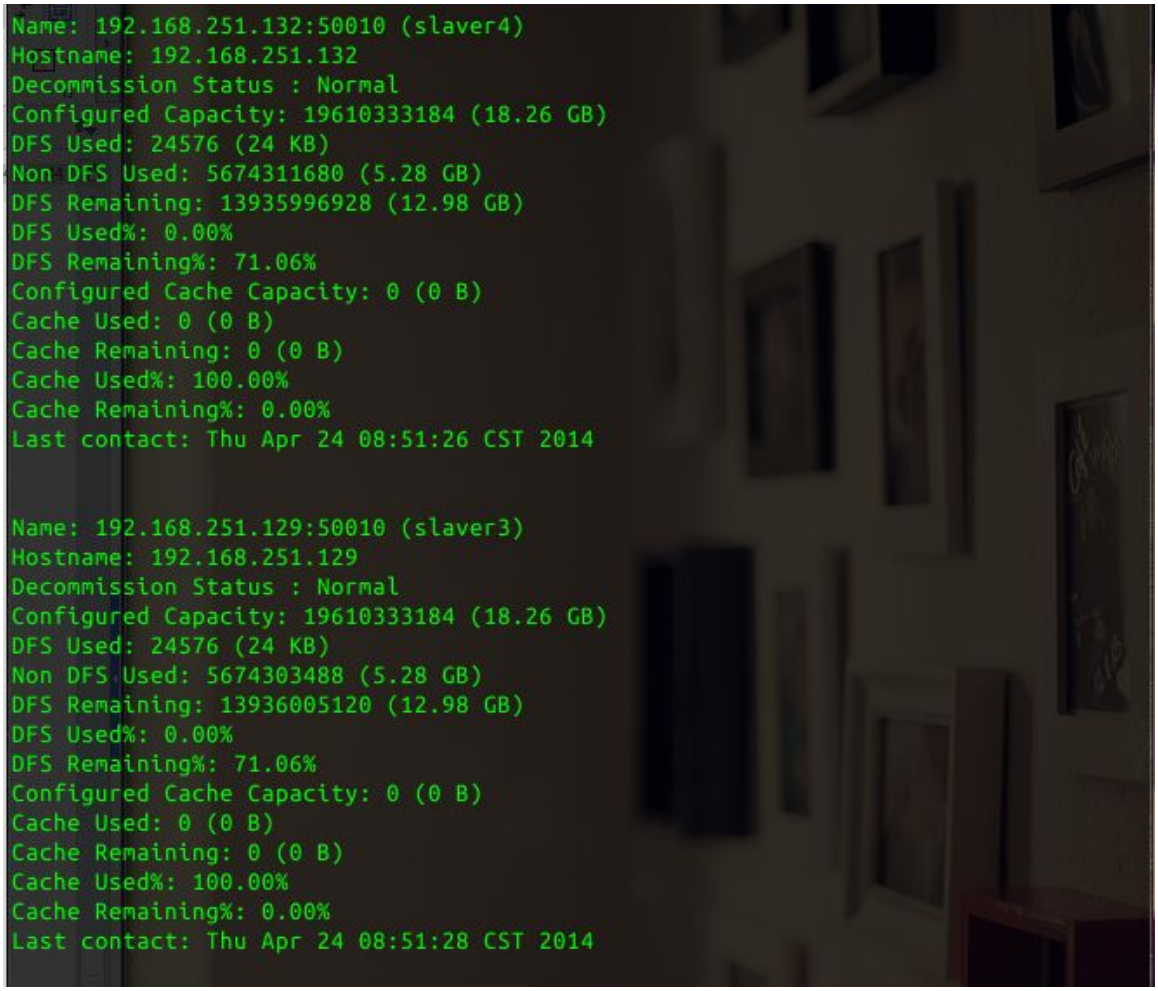


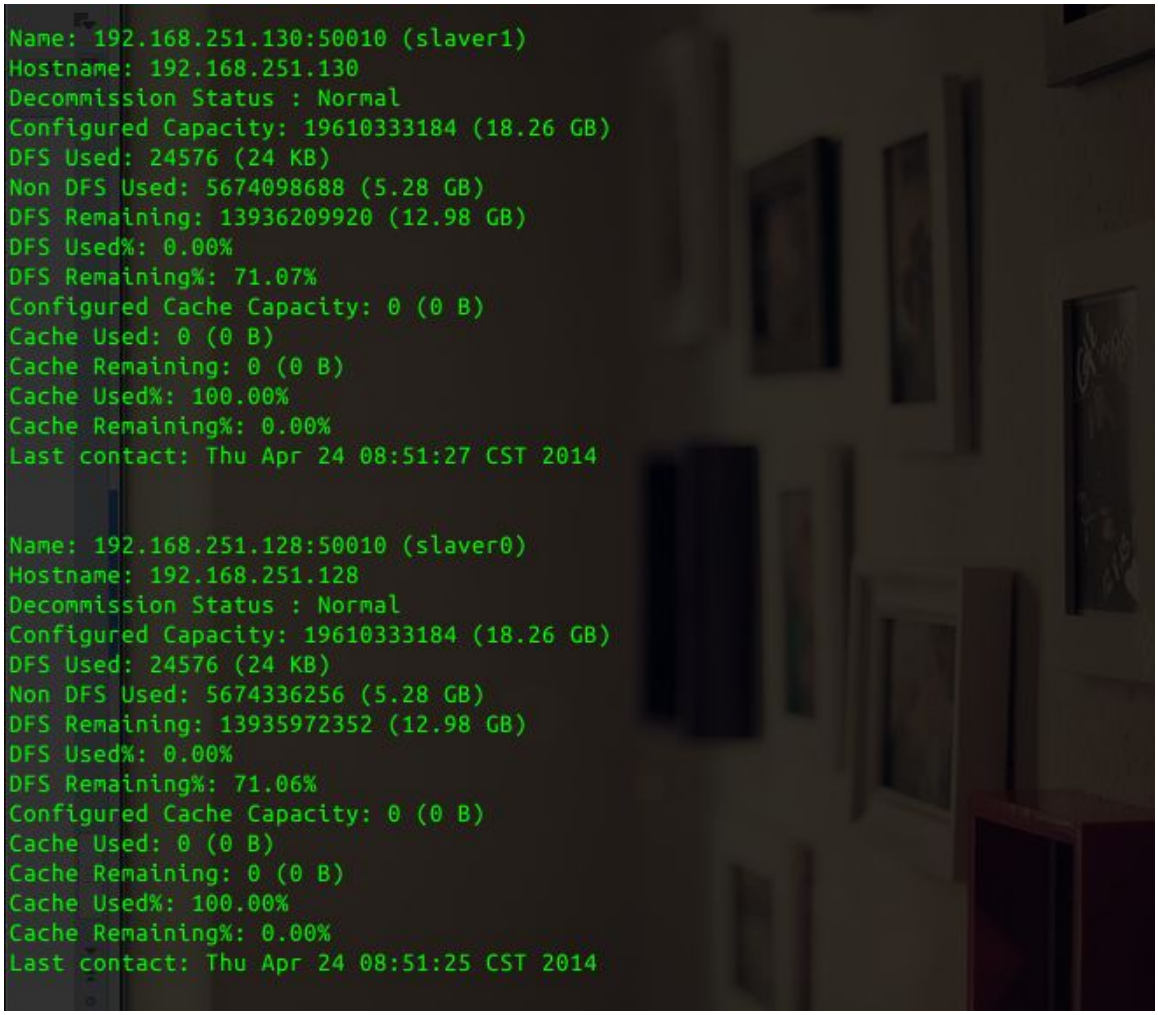
6.监控集群资源

hdfs dfsadmin -report
或者
用 master 登陆网页查看：127.0.0.1:50070
127.0.0.1:8088


[root@master ~]# hdfs dfsadmin -report
14/04/24 08:51:28 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
y for your platform... using builtin-java classes where applicable
Configured Capacity: 98051665920 (91.32 GB)
Present Capacity: 69680492544 (64.90 GB)
DFS Remaining: 69680369664 (64.89 GB)
DFS Used: 122880 (120 KB)
DFS Used%: 0.00%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0

Datanodes available: 5 (5 total, 0 dead)
Live datanodes:
Name: 192.168.251.131:50010 (slaver2)
Hostname: 192.168.251.131
Decommission Status : Normal
Configured Capacity: 19610333184 (18.26 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 5674123264 (5.28 GB)
DFS Remaining: 13936185344 (12.98 GB)
DFS Used%: 0.00%
DFS Remaining%: 71.07%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Last contact: Thu Apr 24 08:51:26 CST 2014


Name: 192.168.251.132:50010 (slaver4)
Hostname: 192.168.251.132
Decommission Status : Normal
Configured Capacity: 19610333184 (18.26 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 5674311680 (5.28 GB)
DFS Remaining: 13935996928 (12.98 GB)
DFS Used%: 0.00%
DFS Remaining%: 71.06%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Last contact: Thu Apr 24 08:51:26 CST 2014
Name: 192.168.251.129:50010 (slaver3)
Hostname: 192.168.251.129
Decommission Status : Normal
Configured Capacity: 19610333184 (18.26 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 5674303488 (5.28 GB)
DFS Remaining: 13936005120 (12.98 GB)
DFS Used%: 0.00%
DFS Remaining%: 71.06%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Last contact: Thu Apr 24 08:51:28 CST 2014



下面从网页监控集群运行情况

Hadoop | Overview | Datanodes | Snapshot | Startup Progress | Utilities

Overview 'master:9000' (active)

Started:	Thu Apr 24 08:51:04 CST 2014
Version:	2.3.0, r1567123
Compiled:	2014-02-11T13:40Z by jenkins from branch-2.3.0
Cluster ID:	CID-82374226-cced-4511-b1d5-f3055f4f7d48
Block Pool ID:	BP-1215179153-192.168.251.1-1398299786222

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 28.09 MB of 237.38 MB Heap Memory. Max Heap Memory is 888.94 MB.

Non Heap Memory used 32.22 MB of 33.5 MB Commited Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	91.32 GB
DFS Used:	120 KB
Non DFS Used:	26.42 GB
DFS Remaining:	64.9 GB
DFS Used%:	0%
DFS Remaining%:	71.07%
Block Pool Used:	120 KB
Block Pool Used%:	0%
DFS Remaining%:	71.07%
Block Pool Used:	120 KB
Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	5 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	0

Namenode Journal Status

Current transaction ID: 10

Journal Manager	State
FileJournalManager(root=/usr/local/hadoop-2.3.0/dfs-name)	EditLogFileOutputStream(/usr/local/hadoop-2.3.0/dfs-name/current/edits_inprogress_0000000000000000010)

NameNode Storage

Storage Directory	Type	State
/usr/local/hadoop-2.3.0/dfs-name	IMAGE_AND_EDITS	Active