**1.软件包准备**

Hadoop-1.1.2 <http://hadoop.apache.org/common/releases.html>

jdk-8u40-linux-i586

**2.配置2台ubuntu虚拟机**

**#2-1 给每台虚拟机创建hadoop账户和组**

sudo addgroup hadoop

sudo adduser --ingroup hadoop hadoop

**#2-2 给hadoop赋予root权限**

sudo gedit /etc/sudoers

**给打开的文件赋予，如下**

# User privilege specification

root ALL=(ALL:ALL) ALL

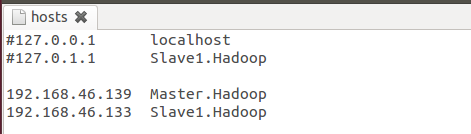
hadoop ALL=(ALL:ALL) ALL

**#2-3修改2台虚拟机的主机名和IP地址，便于互相通信**

**分别对2台虚拟机修改主机名：sudo gedit /etc/hostname**

**弹出的文件分别输入各自的文件名：Master.Hadoop和Slave1.Hadoop**

**分别对2台虚拟机修改IP地址：**

1. **首先输入ifconfig查看虚拟机的静态ip地址。**
2. **本例子中ip地址分别为：192.168.46.139/192.168.46.133**
3. **sudo gedit /etc/hosts**
4. **弹出的文件分别按下图修改，这两条命令加在弹出文件的前面**
5. **配置完成后，用ping 192.168.46.139看是否连通。**

**3.配置SSH无密码验证，使得2台ubuntu虚拟机能互相访问**

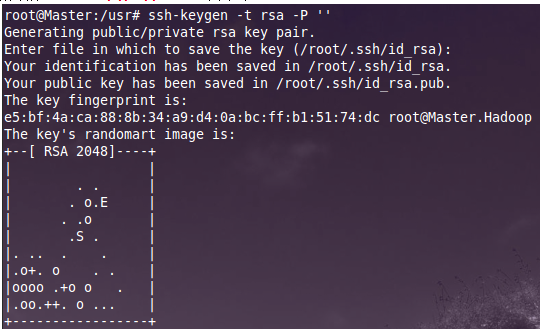
**#3-1 首先安装ssh在虚拟机上(建议联网安装)**

sudo apt-get install ssh

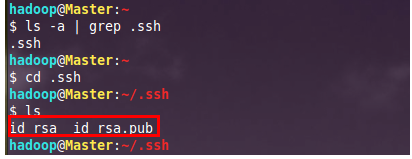
ps：如果想离线安装，请查看其他网上教程。**安装完成后分别在两台ubuntu上进行如下操作**

**#3-2 创建一个无密码密钥对，操作如下，按默认回车，得到下图结果。**

ssh-keygen -t rsa -P ""

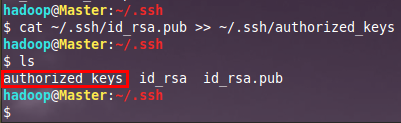


**#3-3 查看生成是否成功生成密钥。**

****

**#3-4 把生成的id\_rsa.pub追加到授权的key里面去。**

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

[](http://images.cnitblog.com/blog/390583/201401/182210564089.png)

利用如下命令设置该文件的权限：

chmod 600 authorized\_keys

**#3-5用root用户登录修改SSH配置文件"/etc/ssh/sshd\_config"的下列内容**

sudo su #获取root权限

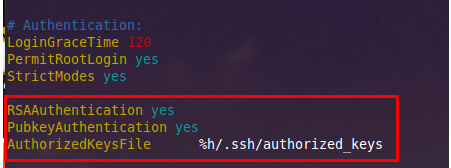
sudo gedit /etc/ssh/sshd\_config

**检查下面几行前面”#”注释是否取消掉：**

RSAAuthentication yes # 启用 RSA 认证

PubkeyAuthentication yes # 启用公钥私钥配对认证方式

AuthorizedKeysFile  %h/.ssh/authorized\_keys # 公钥文件路径



**设置完之后记得重启SSH服务，才能使刚才设置有效。**

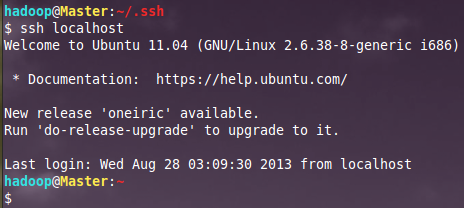
**[clip_image014[12]](http://images.cnitblog.com/blog/390583/201401/182211004864.png)**

**#3-6退出root登录，使用普通用户验证是否设置成功。**

exit

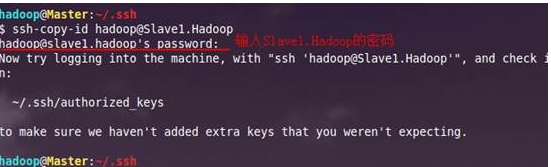
cd /.ssh

ssh localhost

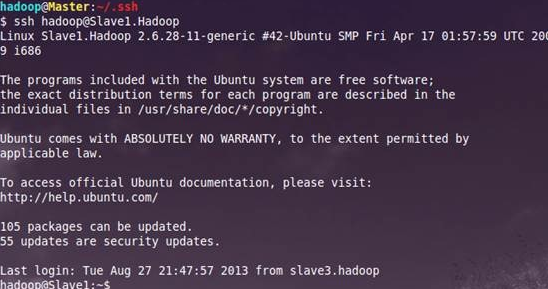


**#3-7从上图中得知无密码登录本级已经设置完毕，接下来的是把公钥复制所有的Slave机器上。**

ssh-copy-id hadoop@Slave1.Hadoop



**#3-8测试是否无密码登录Slave机器上。**



**到此已经实现了从“Master.Hadoop”到“Slave1.Hadoop”的SSH无密码登录。**

**#3-9 Slave1.Hadoop到Master.Hadoop的配置略微有些不同**

**在Slave1上执行3-1到3-6，到3-7时按如下操作：**

scp ~/.ssh/id\_rsa.pub hadoop@Master.Hadoop:/~

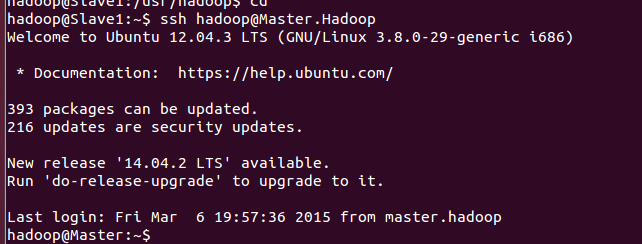
(此步骤是将Slave1上的公共密钥发送到主机Master.Hadoop上)

cat ~/id\_rsa.pub >> ~/.ssh/authorized\_keys

(此步骤是将接受到的公共密钥附加到授权密钥中)

**#3-10 检验slave1是否能无密码访问Master**

ssh hadoop@Master.Hadoop



**4.Java环境安装**

**#4-1首先用root身份登录**

sudo su

**#4-2 创建"/usr/local"下创建"java"文件夹**

sudo mkdir /usr/local/java

**#4-3 将" jdk-8u40-linux-i586.gz"复制到"/usr/local/java"文件夹中.**

**#4-4 解压" jdk-8u40-linux-i586.gz"**

cd /usr/local/java

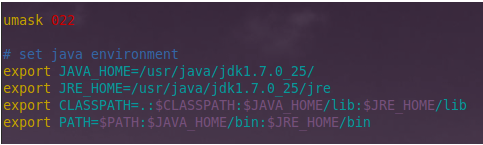
sudo tar xvzf jdk-8u40-linux-i586.gz

(可用于删除的命令 rm -r /usr/local/java)

**#4-5 配置/etc/profile**

cd

sudo gedit /etc/profile

* **If you are installing the JDK then Type/Copy/Paste:**  
  JAVA\_HOME=/usr/local/java/jdk1.8.0\_40  
  JRE\_HOME=$JAVA\_HOME/jre  
  PATH=$PATH:$JAVA\_HOME/bin:$JRE\_HOME/bin  
  export JAVA\_HOME  
  export JRE\_HOME  
  export PATH
* **或者用下图代码**
* 

**#4-6配置完成后重新加载/etc/profile使其生效**

source /etc/profile 或 . /etc/profile

**#4-7 告诉linux系统你的java在哪**

**If you are installing the JDK then Type/Copy/Paste:**

sudo update-alternatives --install "/usr/bin/java" "java" "/usr/local/java/ jdk1.8.0\_40/jre/bin/java" 1

* **Only if you are installing the JDK then Type/Copy/Paste:**

sudo update-alternatives --install "/usr/bin/javac" "javac" "/usr/local/java/ jdk1.8.0\_40/bin/javac" 1

**(更详细的的安装参考**[**http://www.wikihow.com/Install-Oracle-Java-on-Ubuntu-Linux**](http://www.wikihow.com/Install-Oracle-Java-on-Ubuntu-Linux)**)**

**#4-8 验证是否安装成功**

java -version

**#4-9 在Slave1上也安装Java**

**可以直接将Master上的java文件夹复制到Slave1上，由于权限问题无法直接复制到Slave1的local文件夹下，因此先发送到Slave1的～/Downloads文件夹下**

sudo scp -r /usr/local/java seed@Slave1.Master:~/Downloads

**#4-10 Slave1上的java安装和master上的配置相同**

**5.Hadoop 安装**

**#5-1用root权限登录Master.Hadoop机器**

sudo su

**#5-2将下载好的"**hadoop-1.1.2.tar.gz**"复制到**/usr**目录下**

cd /usr

**#5-3解压**

sudo tar –xzvf hadoop-1.1.2.tar.gz

**#5-4 移动到hadoop文件夹**

sudo mv Hadoop-1.1.2 hadoop

**#5-5 赋予hadoop用户关于hadoop文件夹的完全权限**

chown –R hadoop:hadoop Hadoop

**#5-6删除hadoop压缩包**

rm -rf hadoop-1.1.2.tar.gz

(ps:有时候命令无法执行可尝试在前面加入 sudo, e.g., sudo rm -rf hadoop-1.1.2.tar.gz)

**#5-7 在/usr/hadoop下创建tmp文件夹**

cd /usr/hadoop

sudo mkdir tmp

**#5-8 将hadoop的安装路径添加到/etc/profile**

cd

sudo gedit /etc/profile



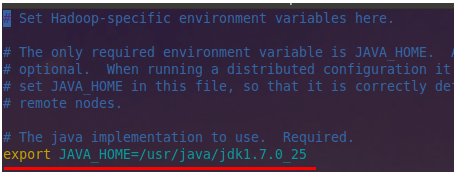
**#5-9配置/usr/hadoop/conf下的文件**

**(1)配置hadoop-env.sh**

cd /usr/hadoop/conf

sudo gedit hadoop-env.sh

加入”export JAVA\_HOME=/usr/local/java/jdk1.8.0\_40”参考这个图



**(2)配置core-site.xml**

修改Hadoop核心配置文件core-site.xml，这里配置的是HDFS master（namenode）的地址和端口号。

**sudo gedit core-site.xml**

<configuration>

    <property>

        <name>hadoop.tmp.dir</name>

        <value>/usr/hadoop/tmp</value>

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

    </property>

<!-- file system properties -->

    <property>

        <name>fs.default.name</name>

        <value>hdfs:// **192.168.46.139**:**9000**</value>

    </property>

</configuration>

（**备注：**请先在 /usr/hadoop 目录下建立 tmp 文件夹）

**(3)配置hdfs-site.xml**

**sudo gedit hdfs-site.xml**

<configuration>

    <property>

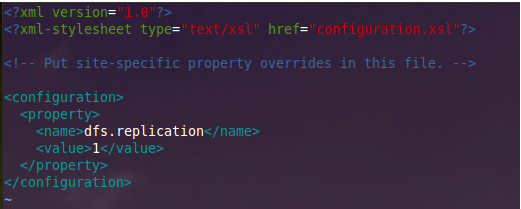
        <name>dfs.replication</name>

        <value>**1**</value>

        (**备注：**replication 是数据副本数量，默认为3，salve少于3台就会报错)

    </property>

<configuration>



**(4)配置mapred-site.xml**

sudo gedit mapred-site.xml

<configuration>

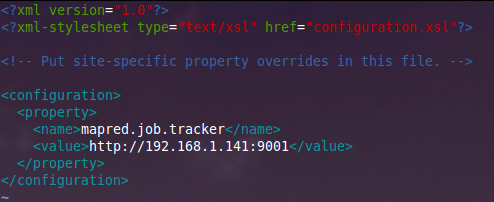
    <property>

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

        <value>http:// **192.168.46.139**:**9001**</value>

    </property>

</configuration>

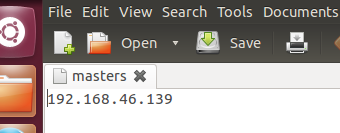


**(5)配置masters**

sudo gedit masters

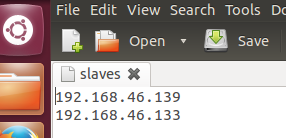
第一种 修改localhost为Master.Hadoop

第二种 去掉"localhost"，加入Master机器的IP：**192.168.46.139** 为保险起见，启用第二种，因为万一忘记配置"/etc/hosts"局域网的DNS失效，这样就会出现意想不到的错误,但一旦IP配对，网络畅通，能通过IP找到相应主机。



**(6)配置slaves**

sudo gedit slaves



**(7)注意在slave机器上不用配置slaves文件，只需要配置masters文件，并把master指向hadoop@Master.Hadoop(192.168.46.139)**

**6.Hadoop安装验证是否成功**

**#6-1 在master上格式化HDFS文件系统**

cd /usr/hadoop

bin/hadoop namenode –format

**#6-2启动hadoop**

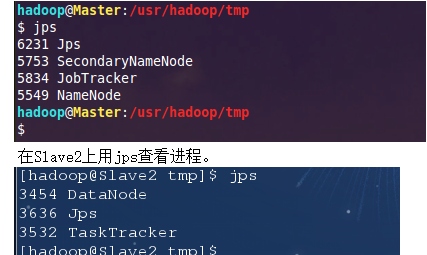
bin/start-all.sh

**ps：先启动master机器，再启动slave机器**

**#6-3 验证hadoop**

**（1）验证方法一**：用"jps"命令

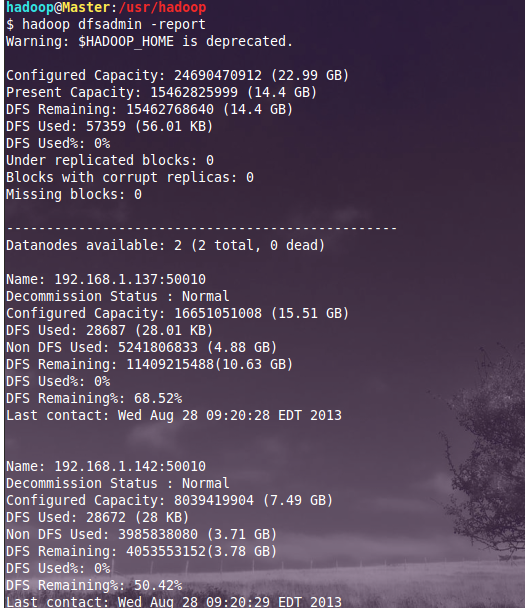
在Master上用 java自带的小工具**jps**查看进程。

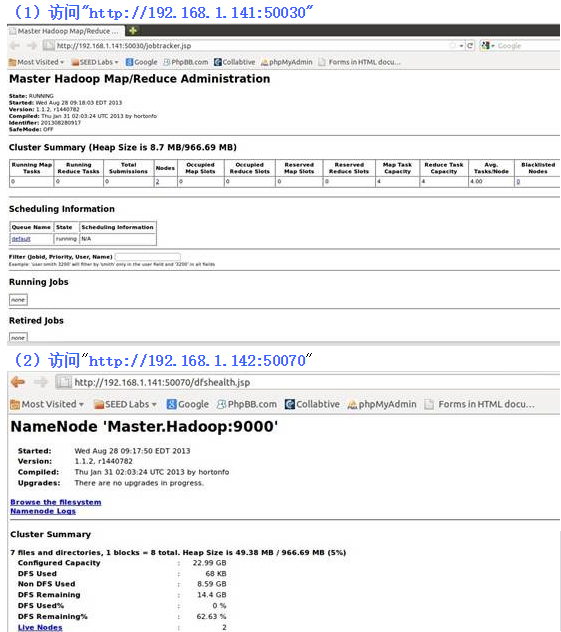


如果在查看Slave机器中发现"DataNode"和"TaskTracker"没有起来时，先查看一下日志的，如果是"namespaceID"不一致问题，采用"常见问题FAQ6.2"进行解决，如果是"No route to host"问题，采用"常见问题FAQ6.3"进行解决。

**（2）验证方式二：**用"hadoop dfsadmin -report"

用这个命令可以查看Hadoop集群的状态。





**http://www.cnblogs.com/lanxuezaipiao/p/3525554.html**