# 伪分布式安装配置

## 集群环境准备

* 准备虚拟机，内存要求1024mb。

### 关闭防火墙

service iptables start;#立即开启防火墙，但是重启后失效。

service iptables stop;#立即关闭防火墙，但是重启后失效。

=============================================================

chkconfig iptables on;#开启防火墙，重启后生效

chkconfig iptables off;#关闭防火墙，重启后生效

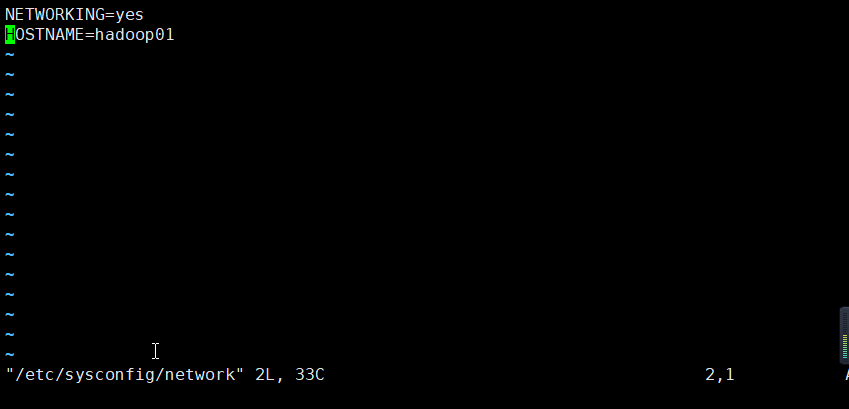
* 两种方式配合使用，可以实现不重启关闭防火墙并且重启后仍然生效。

### 配置主机名

注意安装hadoop的集群主机名不能有下划线!!不然会找不到主机!无法启动!

配置主机名

vim /etc/sysconfig/network



由于该配置需要重启才能生效，所以配合以下命令一起使用。

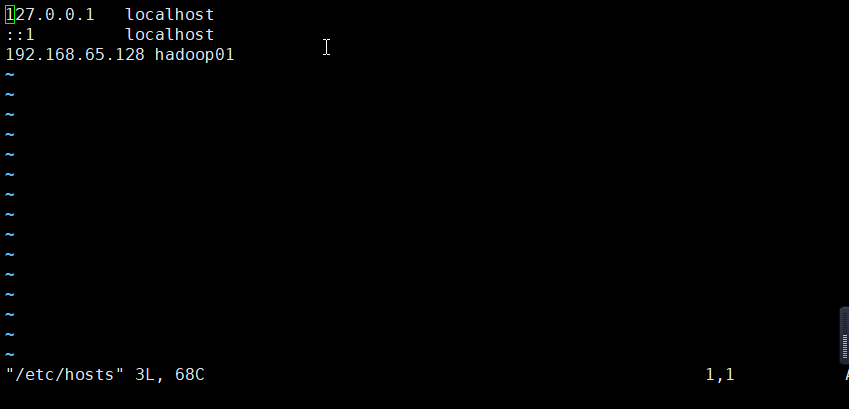
hostname hadoop01

hostname //查看主机名

### 配置Hosts

vim /etc/hosts

* 修改文件内容如下，其中第三行ip地址改为本机ip。



### OS7还需要配置（OS6不需要）

修改/etc/hostname，vim /etc/hostname

### 配置免密码互通

* 生成自己的公钥和私钥,生成的公私钥将自动存放在/root/.ssh目录下。

ssh-keygen

出现输入项后直接回车，三次回车后秘钥生成成功。

* 把生成的公钥copy到远程机器上

ssh-copy-id [user]@[host]

此时在远程主机的/root/.ssh/authorized\_keys文件中保存了公钥,在known\_hosts中保存了已知主机信息，当再次访问的时候就不需要输入密码了。

* 通过此命令远程连接，检验是否可以不需密码连接，未输入密码直接登录证明成功。

ssh [host]

* 退出登录

exit

## Hadoop安装

### 安装JDK

* 将jdk安装包上传
* 解压安装包

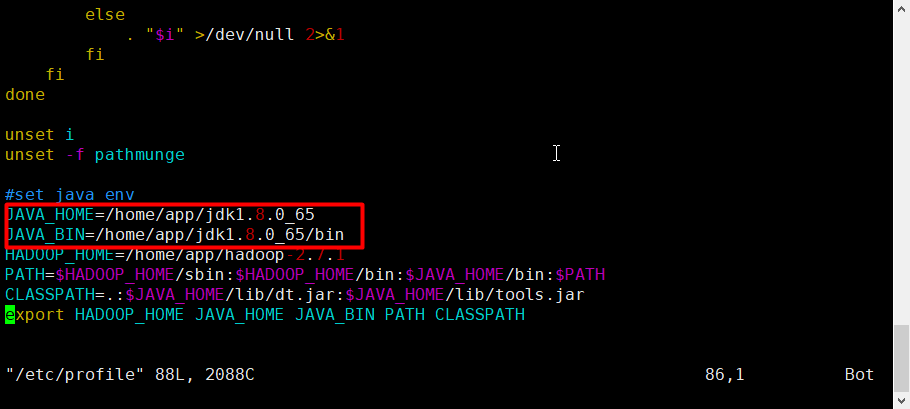
tar -xvf [jdk安装包位置]

* 配置环境变量

修改profile -- 这个文件是每个用户登录时都会运行的环境变量设置，当用户第一次登录时,该文件被执行. 并从/etc/profile.d目录的配置文件中搜集shell的设置。

vim /etc/profile

* 在文件行尾加入配置后保存退出



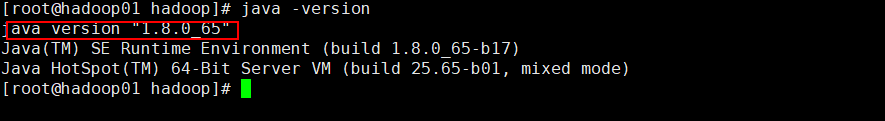
* 重新加载profile使配置生效

source /etc/profile

* 环境变量配置完成，测试环境变量是否生效

java -version

出现版本号证明成功



### 安装hadoop

* 将hadoop安装包上传到linux
* 解压安装包

tar -xvf [hadoop安装包位置]

### 配置hadoop

**1.修改 hadoop-env.sh**

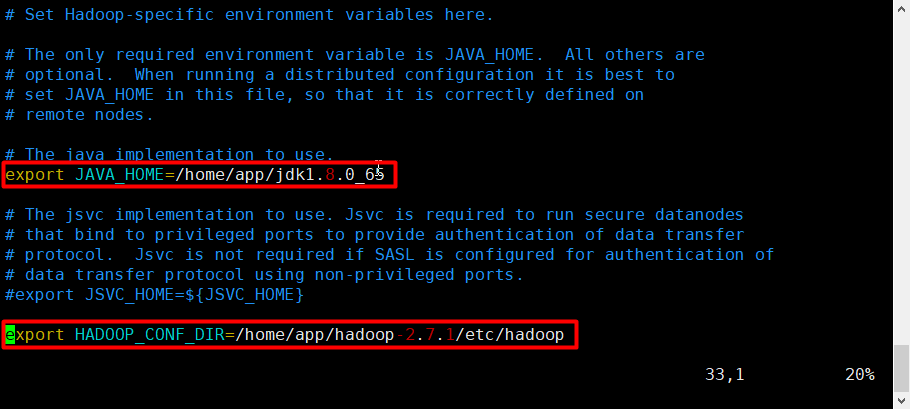
* 通过vim打开

vim [hadoop]/etc/hadoop/hadoop-env.sh

主要是修改java\_home的路径

在hadoop-env.sh的第25行,把export JAVA\_HOME=${JAVA\_HOME}修改成具体的路径

在33行，修改HADOOP\_CONF\_DIR为具体的路径



* 重新加载使修改生效

source hadoop-env.sh

**2.修改 core-site.xml**

* 通过vim打开

vim [hadoop]/etc/hadoop/core-site.xml

* 增加namenode配置、文件存储位置配置：粘贴高亮部分到<configuration>标签内

<configuration>

<property>

<!--用来指定hdfs的老大，namenode的地址-->

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

<value>hdfs://hadoop01:9000</value>

</property>

<property>

<!--用来指定hadoop运行时产生文件的存放目录-->

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

<value>/home/software/hadoop-2.7.7/tmp</value>

</property>

</configuration>

**3.修改 hdfs-site.xml**

* 通过vim打开

vim [hadoop]/etc/hadoop/hdfs-site.xml

* 配置包括自身在内的备份副本数量：粘贴高亮部分到<configuration>标签内

<configuration>

<property>

<!--指定hdfs保存数据副本的数量，包括自己，默认为3-->

<!--伪分布式模式，此值必须为1-->

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<!--设置hdfs操作权限，false表示任何用户都可以在hdfs上操作文件-->

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>

**4.修改 mapred-site.xml**

说明：在/etc/hadoop的目录下，只有一个mapred-site.xml.template文件，复制一个。

cp mapred-site.xml.template mapred-site.xml

* 通过vim打开

vim [hadoop]/etc/hadoop/mapred-site.xml

* 配置mapreduce运行在yarn上：粘贴高亮部分到<configuration>标签内

<configuration>

<property>

<!--指定mapreduce运行在yarn上-->

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

<value>yarn</value>

</property>

</configuration>

**5.修改 yarn-site.xml**

* 通过vim打开

vim [hadoop]/etc/hadoop/yarn-site.xml

* 配置：粘贴高亮部分到<configuration>标签内

<configuration>

<property>

<!--指定yarn的老大resourcemanager的地址-->

<name>yarn.resourcemanager.hostname</name>

<value>hadoop01</value>

</property>

<property>

<!--NodeManager获取数据的方式-->

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

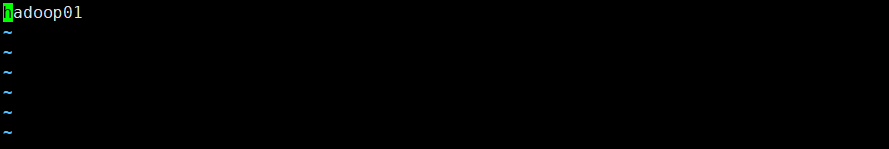
<value>mapreduce\_shuffle</value>

</property>

</configuration>

1. **修改 slaves**

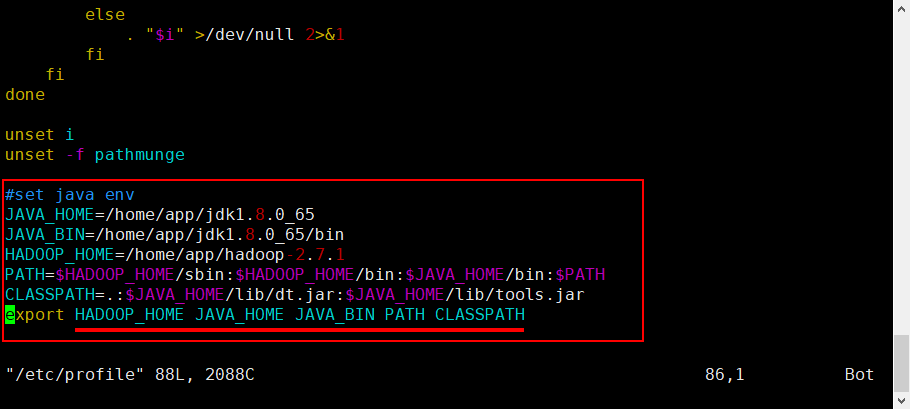
vim slaves



**7.配置hadoop的环境变量**

vim /etc/profile

最终结果如下图：



* 重新加载profile使配置生效

source /etc/profile

* 环境变量配置完成，测试环境变量是否生效

echo $HADOOP\_HOME

1. **创建tmp文件夹**

cd /home/software/hadoop-2.7.7

mkdir tmp

**9.格式化namenode**

* 进入 hadoop/bin 输入命令格式化namenode

hadoop namenode -format

在格式化的时候，会有这样的输出证明成功：

Storage directory /tmp/hadoop-root/dfs/name has been successfully formatted

## 测试hadoop

### 启动hadoop

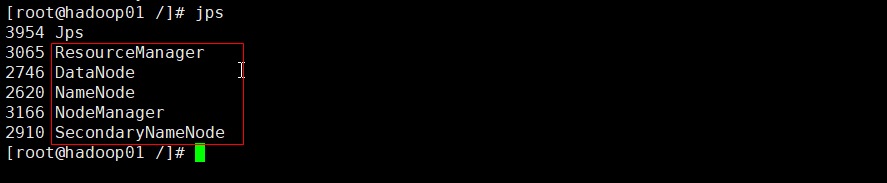
在/home/software/hadoop-2.7.7/sbin目录下执行下面的命令 （由于配置了hadoop环境变量，所以关于hadoop的命令也可以在其他目录下执行）

start-all.sh

### 验证hadoop是否安装启动成功

执行如下命令出现图中五个进程证明hadoop安装并启动成功：

jps



### 关闭hadoop

stop-all.sh

### 通过浏览器访问hadoop管理页面

http://[server\_ip]:50070

