

# Hadoop 安装需要的 Linux 相关操作

## 目 录

1. 设置 Linux 启动后不启动图形桌面 .....	1
2. VMWare 设置 Linux 虚拟机的内存 .....	1
3. 配置 IP 地址 .....	2
4. 更改主机名 .....	5
5. IP 与主机名映射 .....	5
6. SSH 免密码公钥认证设置 .....	5
7. 使用 VMWare 克隆出几个 Linux .....	9
8. 把文件追加到另外一个文件尾部 .....	9
9. SCP 远程拷贝 .....	9
10. 关闭防火墙 .....	10
11. 关闭 SELinux .....	10

在安装 Hadoop 时需要用到很多 Linux 的操作, 本文对这些内容作详细说明。

## 1. 设置 Linux 启动后不启动图形桌面

Linux 系统启动后直接进入图形桌面, 会消耗较多的内存。如果在同一台计算机上运行多个 Linux 虚拟机, 系统会非常慢。所以要设置 Linux 系统启动后直接 Shell 界面, 而不要启动图形界面。

设置方法如下: 编辑 `/etc/inittab` 文件。找到 `id:5:initdefault:` 这一行, 将 5 改为 3, 即将它改为 `id:3:initdefault:`。设置完毕后重新启动系统即可。

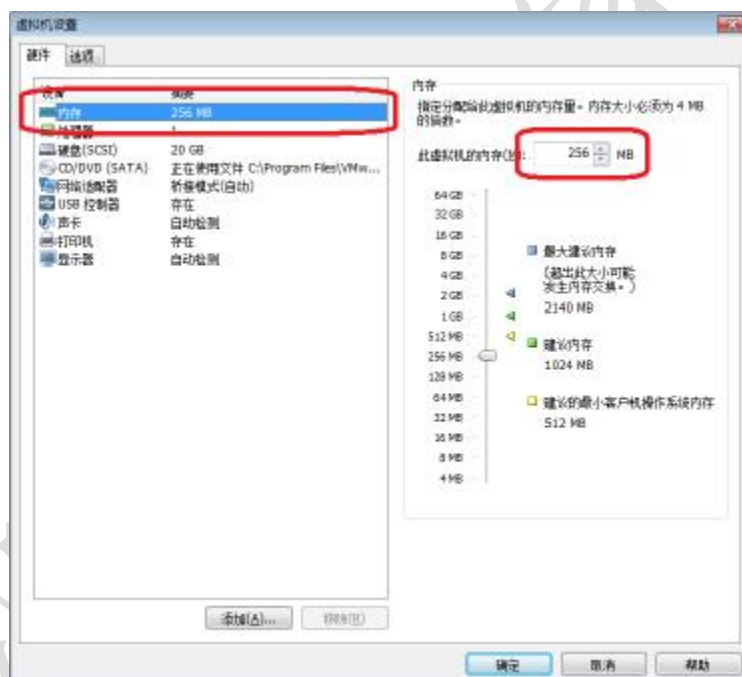
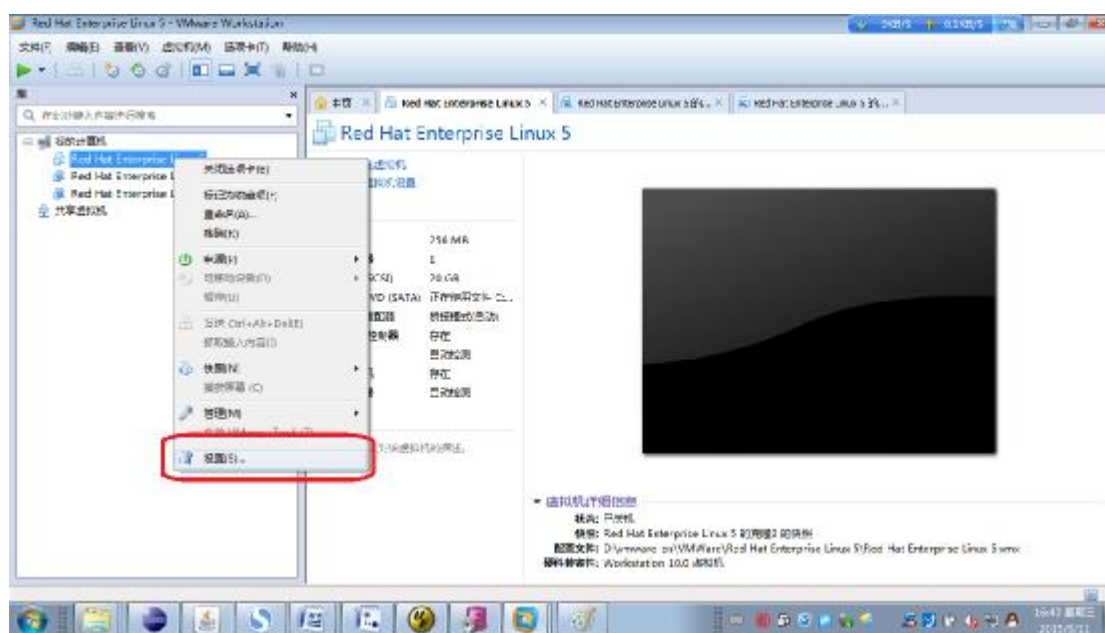
如果在 Shell 控制台想进入图形桌面, 只要执行命令: `startx`。

## 2. VMWare 设置 Linux 虚拟机的内存

如果 Linux 是使用虚拟机创建, 则给 Linux 虚拟机分配 256M 内存即可。不然多个 Linux 系统无法启动。当然, 如果你的计算机内存比较多, 可适当调高此处的值。

VMWare 中的设置方法: 右键点击左侧的虚拟机, 选择弹出菜单“设置”打

开如下的设置窗口。注意：只能在 Linux 未启动的情况下设置。



### 3. 配置 IP 地址

以下列举修改 IP 的两种方法，只要使用其中一种即可。

#### (1) 第 1 种修改 IP 永久生效方法

使用图形化工具配置 IP 地址，菜单：系统->管理->网络。然后双击 eth0 设备，打开如下界面，选中“静态设置的 IP 地址”，然后在下方设置 IP，设置好后，点击确定。

常规(G) 路由(R) 硬件设备(H)

别名(N) :

☒ 当计算机启动时激活设备(A)

☐ 允许所有用户启用和禁用该设备(U)

☐ 为此界面启用 IPv6 配置

☐ 自动获取 IP 地址设置使用 :

DHCP 设置

主机名(可选)(O) :

☒ 自动从提供商处获取 DNS 信息(D)

☒ 静态设置的 IP 地址 :

手工设置 IP 地址

地址(d) :

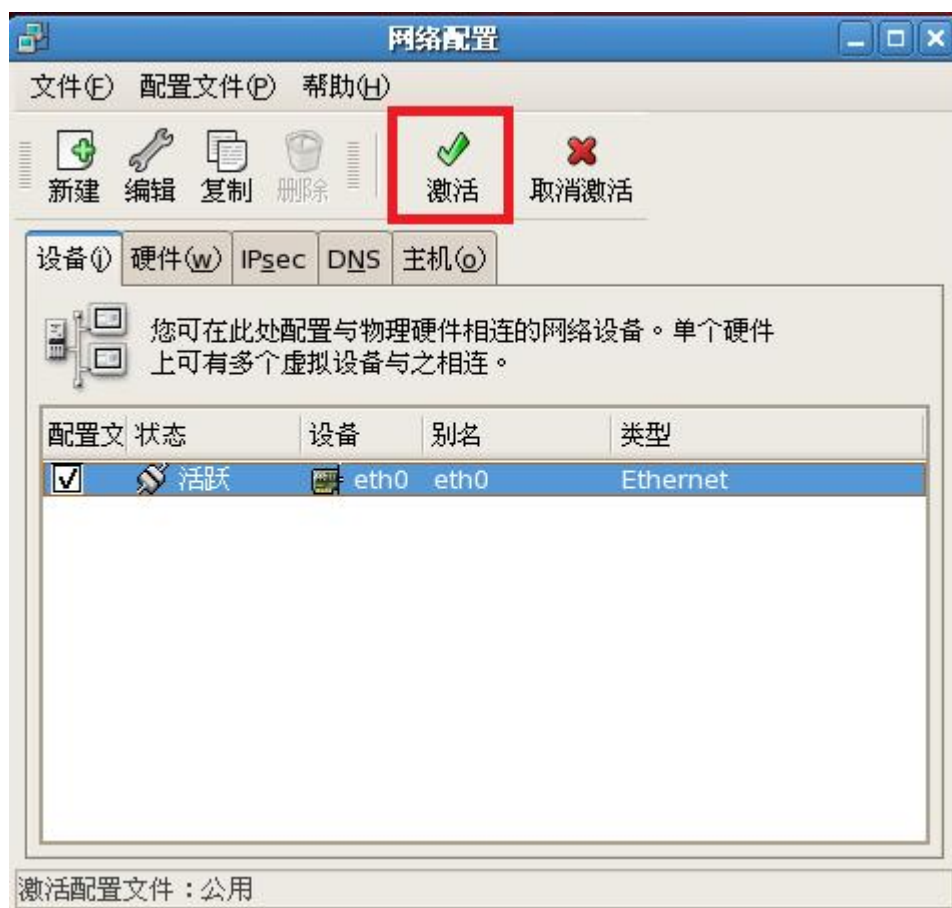
子网掩码(S) :

默认网关(G)地址 :

☐ 设置 MTU 为 :

☐ Set MRU to:

设置 IP 后，一定要激活网卡，方法：选中 eth0 设备，再点击“激活”按钮。



## (2) 第 2 种修改 IP 永久生效方法

直接修改 IP 的配置文件：/etc/sysconfig/network-scripts/ifcfg-eth0。通过 vi 命令编辑：

```
vi /etc/sysconfig/network-scripts/ifcfg-eth0
```

说明：此处的 eth0 表示第一块网卡，如果是第二块则为 eth1。

按如下修改 IP：

```
DEVICE=eth0（如果是第二块则为 eth1）  
BOOTPROTO=static  
IPADDR=192.168.0.11(改成要设置的 IP)  
NETMASK=255.255.255.0 (子网掩码)  
GATEWAY=192.168.0.1(网关)  
ONBOOT=yes(指明在系统启动时是否激活网卡)
```

然后使用如下命令重启：

```
service network restart
```

IP 地址生效，重启系统后不变。

## 4. 更改主机名

修改/etc/sysconfig/network 里面的 hostname，改成指定的主机名。

修改后需要重启机器才会生效。

## 5. IP 与主机名映射

在/etc/hosts 文件中增加映射，每行为一个主机，每行由三部份组成，每个部份由空格隔开。其中“#”符号是注释语句。

第一部份：网络 IP 地址；

第二部份：主机名.域名，注意主机名和域名之间有个半角的点，比如 localhost.localdomain；

第三部份：主机名(主机名别名)，其实就是主机名；

当然每行也可以是两部份，就是主机 IP 地址和主机名；比如 192.168.1.195 debian。

例如：

```
192.168.1.100  hadoop0 hadoop0
```

```
192.168.1.101  hadoop1 hadoop1
```

```
192.168.1.102  hadoop2 hadoop2
```

修改不用重启即生效。

## 6. SSH 免密码公钥认证设置

### (1) 工作原理

在 Hadoop 启动以后，Namenode 是通过 SSH 来启动和停止各个节点上的各种守护进程的，这就需要在节点之间执行指令的时候是不需要输入密码的方式，所以我们需要配置 SSH 使用无密码公钥认证的方式登录。

首先要保证每台机器上都装了 SSH 服务器，且都正常启动。实际中我们用的都是 OpenSSH，这是 SSH 协议的一个免费开源实现。现在的 Linux 发行版一般都

默认安装了 SSH 的，开机的时候你可以观察下是否有名字为 sshd 的守护程序启动，如果没有，则需要安装 OpenSSH 软件。

以本文中的 3 台机器为例，现在 hadoop0 是主节点，它需要主动发起 SSH 连接到 hadoop1 和 hadoop2，对于 SSH 服务来说，hadoop0 就是 SSH 客户端，而 hadoop1，hadoop2 则是 SSH 服务端，因此在 hadoop1，hadoop2 上需要确定 sshd 服务已经启动。简单的说，在 hadoop0 上需要生成一个密钥对，即一个私钥，一个公钥。然后将公钥拷贝到 hadoop1 上，这样，比如当 hadoop0 向 hadoop1 发起 ssh 连接的时候，hadoop1 上就会生成一个随机数并用 hadoop0 的公钥对这个随机数进行加密，并发送给 hadoop0，hadoop0 收到这个加密的数以后用私钥进行解密，并将解密后的数发送回 hadoop1，hadoop1 确认解密的数无误后就允许 hadoop0 进行连接了。这就完成了一次公钥认证过程。

## (2) 配置 SSH 面密码登录

### (1) 需要关闭 SELinux

参见《关闭 SELinux》一节。

### (2) 配置 SSH 的无密码登录认证

如果需要 hadoop0 机器能够 SSH 无密码登录到 hadoop1 和 hadoop2，则需要先在 hadoop0 上生成密钥对，然后把公钥字符串加到 hadoop1 和 hadoop2 机器的 `/root/.ssh/authorized_keys` 文件中。

以下是配置 SSH 的无密码认证的过程。对于本文中的 3 台机器，首先在 hadoop0 上生成密钥对，命令如下(在命令中直接指定参数值)：

```
[root@hadoop0 ~]# ssh-keygen -t rsa -P "" -f /root/.ssh/id_rsa
```

参数说明：

-t: 指定密钥算法，选项内容有：RSA 或者 DSA。

-P: 密码。

-f: 生成的文件。如果不设置此参数，缺省的是 `/root/.ssh/id_rsa`。

命令第 1 个 `rsa` 表示采用 `rsa` 算法，最后的 `rsa` 只是个代号，你可以取你自己喜欢的名字。这个命令将为 hadoop0 上的当前用户 `root` 生成其密钥对，密钥对

的保存路径是 `/root/.ssh/id_rsa`。这样生成的证书以及公钥将存储在 `/root/.ssh` 目录，形成两个文件 `id_rsa` 和 `id_rsa.pub`。

然后将 `id_rsa.pub` 文件的内容复制到每一台机器(包括 `hadoop0`)的 `/root/.ssh/authorized_keys` 文件的尾部。若机器上不存在这个 `authorized_keys` 文件，可以自行创建一个。请注意 `id_rsa.pub` 文件的内容是长长的一行，复制时需注意，不要遗漏字符或混入了多余换行符。可以使用 `cat` 命令把一个文件写到另一个文件的末尾，如果目标文件不存在，会自动创建，命令如下：

```
$ cat /root/.ssh/id_rsa.pub >> /root/.ssh/authorized_keys
```

修改 `authorized_keys` 的读写权限：

```
[root @hadoop1:~]# chmod 640 /root/.ssh/authorized_keys
```

必须保证 `authorized_keys` 只对其所有者有读写权限，其他人不允许有写的权限，否则 SSH 是不会工作的。

### (3) SSH 连接测试

接下来可以做一下 SSH 连接测试，从 `hadoop0` 分别向 `hadoop0`, `hadoop1`, `hadoop2`, 发起 SSH 连接请求，确保不需要输入密码就能 SSH 连接成功。SSH 登录到另外一台服务器的命令格式如下：

```
[root@hadoop0 ~]# ssh 服务器 IP
```

注意第一次 SSH 连接时会出现如下提示信息：

```
The authenticity of host [homer06] can't be established. The key fingerprint is:
74:32:91:f2:9c:dc:2e:80:48:73:d4:53:ab:e4:d3:1a
Are you sure you want to continue connecting (yes/no)?
```

请输入 `yes`，这样 OpenSSH 会把连接过来的这台主机的信息自动加到 `/root/.ssh/known_hosts` 文件中，第二次再连接时，就不会有这样的提示信息了。

配置好 SSH 免密码认证方式后，从一台机器登录到另一台机器，不再需要输入密码就可以建立 SSH 连接。

注意事项:

这里提供一种将 `id_rsa.pub` 的内容复制到 `/root/.ssh/authorized_keys` 文件中的方法:

```
[root @ hadoop0:~ssh]# cat /root/.ssh/id_rsa.pub >> /root/.ssh/authorized_keys
```

远程拷贝:

```
[root@hadoop0 ~]# scp /root/.ssh/authorized_keys hadoop1:/root/.ssh/
```

```
[root@hadoop0 ~]# scp /root/.ssh/authorized_keys hadoop2:/root/.ssh/
```

此处的 `scp` 就是通过 `ssh` 进行远程 `copy`, 此处需要输入远程主机的密码, 即 `hadoop1`, `hadoop2` 机器上 `hadoop` 帐户的密码, 当然, 你也可以用其他方法将 `authorized_keys` 文件拷贝到其他机器上。

然后分别登录到 `hadoop1` 和 `hadoop2` 上修改 `authorized_keys` 文件的读写权限:

```
[root @hadoop1:~ssh]# chmod 640 /root/.ssh/authorized_keys
```

```
[root @hadoop2:~ssh]# chmod 640 /root/.ssh/authorized_keys
```

这一步非常关键, 必须保证 `authorized_keys` 只对其所有者有读写权限, 其他人不允许有写的权限, 否则 `SSH` 是不会工作的。

在 `hadoop0` 上执行的代码如下:

```
[root@hadoop0 ~]# ssh-keygen -t rsa -P "" -f /root/.ssh/id_rsa
[root@hadoop0 ~]# cat /root/.ssh/id_rsa.pub >> /root/.ssh/authorized_keys
[root@hadoop0 ~]# chmod 640 /root/.ssh/authorized_keys
[root@hadoop0 ~]#
[root@hadoop0 ~]# scp /root/.ssh/authorized_keys hadoop1:/root/.ssh/
[root@hadoop0 ~]# scp /root/.ssh/authorized_keys hadoop2:/root/.ssh/
```

然后分别登录到 `hadoop1` 和 `hadoop2` 上修改 `authorized_keys` 文件的读写权限:

```
[root @hadoop1 ~]# chmod 640 /root/.ssh/authorized_keys
```

`hadoop2` 上修改 `authorized_keys` 文件的读写权限:

```
[root @hadoop2 ~]# chmod 640 /root/.ssh/authorized_keys
```



## 7. 使用 VMWare 克隆出几个 Linux

如果使用 VMware Workstation 安装 3 个 RHEL 5.4 系统, 可以先安装好一个 Linux, 并且安装好 JDK, 然后再利用 VMWare Workstation 的克隆功能完成另外两个的安装。

克隆方法如下: 选择 VMWare 菜单“虚拟机->管理->克隆”。克隆方式使用“创建完整克隆”, 而不是“创建链接克隆”, 链接克隆的 Linux 启动特别慢。其他选项默认即可。克隆出的 Linux, 要配置新的 IP 地址。

克隆后, 一般不需要重新生成 MAC 网卡地址, 只有在 VMWare 老版本(如 VMWare7)中需要重新生成每一台克隆出来的 MAC 网卡地址, 否则克隆出来的 MAC 地址和原来的机器的 MAC 地址一样, 就不能联网。如果克隆出的 Linux 配置了 IP 也不能联网时, 可以重新生成一下克隆出来的 Linux 的 MAC 网卡地址。

重新生成 MAC 地址的方法: 在 VMWare 中右键点击要设置的虚拟机, 选择弹出菜单中的“设置”, 选择“网络适配器”, 在点击“高级”按钮, 然后在 MAC 地址处, 点击“生成”按钮。最后重启机器。

## 8. 把文件追加到另外一个文件尾部

使用 cat 命令:

```
$ cat file1 file2 >> file3
```

如果文件 file3 不存在, 则新建 file3。

例如:

```
$ cat /root/.ssh/id_rsa.pub >> /root/.ssh/authorized_keys
```

## 9. SCP 远程拷贝

scp 命令可以在两个 Linux 之间拷贝文件。

scp 有两个参数, 第 1 个参数为本地需要拷贝的文件, 第 2 个参数指定拷贝到什么地方去。

拷贝文件:

```
scp /home/hadoop/conf/core-site.xml root@192.168.1.100: /home/hadoop/
```

拷贝目录:

```
scp -r /home/hadoop root@192.168.1.100: /home/
```

## 10. 关闭防火墙

执行 `setup` 命令，在打开的界面中设置。

或者使用 `root` 权限关闭防火墙，执行 `/etc/init.d/iptables stop`，运行命令 `/etc/init.d/iptables status` 查看防火墙状态。此方法为临时停用防火墙，重启系统后，防火墙会自动启动，所以最好使用第 1 种方法。

## 11. 关闭 SELinux

关闭 SELinux 的方法：修改 `/etc/selinux/config` 文件中的 `SELINUX=""` 为 `disabled`，然后重启。如果不想重启系统，使用命令 `setenforce 0`。

说明：

`setenforce 1` 设置 SELinux 成为 enforcing 模式。

`setenforce 0` 设置 SELinux 成为 permissive 模式。