主要参考：

<http://dblab.xmu.edu.cn/blog/install-hadoop/>

<http://dblab.xmu.edu.cn/blog/install-hadoop-cluster/>

以及文件夹中的两个PDF

使用的是VM12，Ubuntu14.04，jdk1.8.0-112，hadoop2.7.3

1. 安装虚拟机

安装虚拟机时，内存给大点，运行的时候不容易卡；

虚拟机都使用同一个用户名，当后面使用ssh无密码登录时，要求这样。否则需要新建一个统一的用户名；

虚拟机的网络需要使用桥接模式。

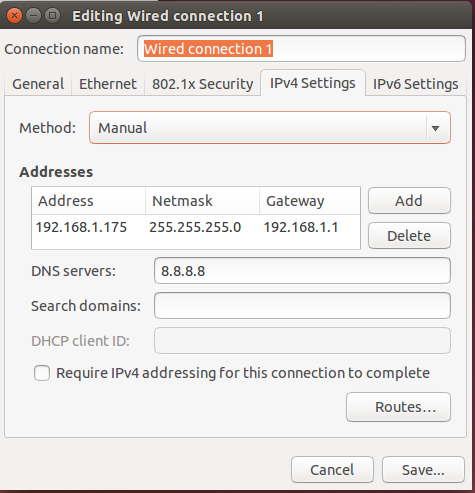
1. 安装需要的软件：vim、ssh、openssh-server等。

安装java并配置路径等。

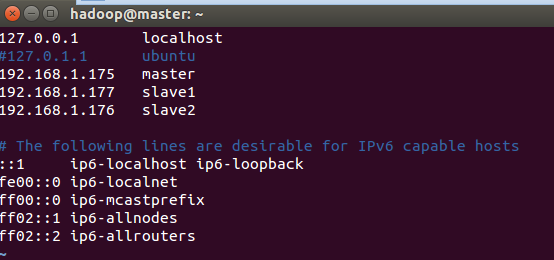
Java配置路径方法：http://dblab.xmu.edu.cn/blog/linux-environment-variable/

更新软件。

1. 设置静态IP（从网络设置界面设置），网络需要设为桥接模式。



1. 编辑修改主机名，编辑修改/etc/hosts,/etc/hostname两个文件，hosts截图：



修改主机名之后需要重启才能够生效。

注意区分开主机名、用户名、虚拟机名等，在hosts、hostname文件中的是主机名。

1. 实现主机之间的无密码登录：需要使用统一的用户名

若在建立虚拟机的时候没有使用统一的用户名，可以使用root作为统一的用户名，我是用的在所有虚拟机上新建一个同样的用户。

ssh生成的公钥和私钥与主机名和用户名有关，当修改过主机名之后，需要重新生成；当使用了一个新的用户名之后，ssh也要重新生成当前用户、主机名的公钥和私钥。

（同一个主机名和用户名，两次生成的公钥和私钥不知道会不会相同，猜测不会相同。）（我曾经删除掉一个虚拟机，然后重新建立一个用户名、主机名一样的虚拟机，IP也一样，之前那个虚拟机已经实现了与另一台主机的免密登录，新建立的主机与另一台主机的ssh会出错，说明公钥和私钥都改变了。）当与某一台主机建立过ssh连接后，在~/.ssh目录中会生成一个known\_hosts文件，在该文件中会记录该主机的某些信息，当删掉该虚拟机，再重新建一个一样的虚拟机再连接时，由于与known\_hosts文件中的某些内容不匹配了，会导致连接失败，要删掉known\_hosts文件中的相关记录，按照出错信息的提示来，输入指令应该就好了。要实现免密登录的话，还要重新生成公钥和私钥。

Ssh 用户名@主机名 以用户名为“用户名”的身份登录到“主机名”在hosts文件中对应的IP的主机上。

Ssh 主机名 用当前的用户名登录到“主机名”在hosts文件中对应的IP的主机上。

我们要实现的就是使用ssh 主机名这种方法的免密登陆，所以需要在所有的机器上有统一的用户名。

实现免密登陆，只要在.ssh文件夹中的authorized\_keys文件中有各个主机的ssh公钥就好了。

使用scp传输文件，scp传输文件时，总是出现permission denied错误，有时候是因为对目标文件夹没有写权限，换一个文件夹就好了。写入/home文件夹失败，但写入/home/hadoop文件夹就成功了。

1. 安装hadoop，直接解压就可以使用了。能够看到版本号，就证明按装好了。注意hadoop的路径，很多地方都会用到。特别是修改配置文件时。
2. 修改hadoop的配置文件，注意需要修改hadoop的路径，master和slave的写法（大小写）,等。
3. 有时候会出现Java\_home找不到或出错的问题，在~/hadoop/hadoop-2.7.3/etc/hadoop/hadoop-env.sh中，将其中的Java\_home直接改为JAVA的路径就好了。

echo $JAVA\_HOME 可以输出java的路径。

1. 按照教程中的指令运行hadoop。按装openjdk-7-jdk，运行jps指令，就可以看到正在运行的hadoop进程。
2. 运行实例时，将运行结果写在具体的目录下，否则找不到运行结果。

按照教程跑例子时，将教程中的输出目录由output 改为~/hadoop/hadoop-2.7.3/output 或者./output ,否则找不到结果文件。