 搭建Hadoop单机伪分布式环境

本次试用的是ubuntu 16.04作为系统环境

使用的是VMware Workstation 12作为虚拟机

本次不介绍CentOS和RedHat搭建Hadoop单机伪分布式环境

hadoop使用的Hadoop 2.6.4

好的，现在开始我们的hadoop搭建环境之旅！！！

第一步：

打开ubuntu终端  如果没有hadoop用户，我们创建一个hadoop用户

       使用命令 sudo useradd -m hadoop -s /bin/bash

这里稍微回顾一下：

用户与组

用户：多用户

用户和组   实现资源分配的核心要素

组是一组容器   可以添加用户  也可以添加权限

用户类型：

管理员  0

系统用户：1--499

运行后台的程序（daemon）

普通用户：500+

登录：交互式访问

组类别：

管理员组：0

系统组：1-499

用户组：500+

Linux:

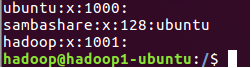
/etc/passwd ：用户的账号信息

/etc/shadow ：用户的密码与相关账户设定

/etc/group :组的账号的信息

/etc/gshaow:组的密码信息

执行完创建用户后 使用sudo cat /etc/group



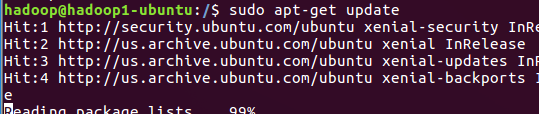
使用sudo passwd hadoop 为hadoop设置密码

Sudo adduser hadoop sudo 为 hadoop 用户增加管理员权限

最后使用reboot命令重启ubuntu 然后使用hadoop用户登录

第二步：

Hadoop用户登录后 更新update 不然有些软件使用不了

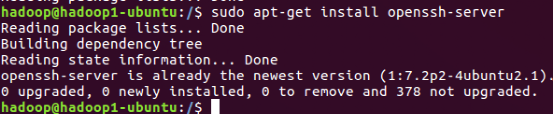


本人使用的是us的镜像源，有的时候更新比较慢，或者你们的hash有问题，说明软件源找不到资源，你们可以更改软件源  建议使用163的。

对于更改软件源，我就不说了，自行解决。

然后安装vim的文本编辑器   文本编辑器有很多种 nano vi vim(vi的增强版) gedit(图形化界面，建议对于vi或者vim不是很熟悉的使用，但是以后还是要学会vi或者vim的，道理都懂的)。 对于刚刚接触linux的我本次使用的gedit，因为后面修改xml配置文件和修改.sh脚本文件方便，加快效率，防止出错。

对于系统有了nano  vi  安装vim就是



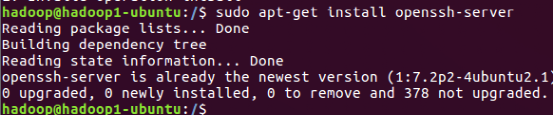
因为我已经装了，所以出现以上画面，没有装的应该是有y/n的选项  直接输入y就可以了。

第三步：

安装ssh 能够实现远程控制

Ubuntu默认装了ssh client  然后我们可以直接安装服务器

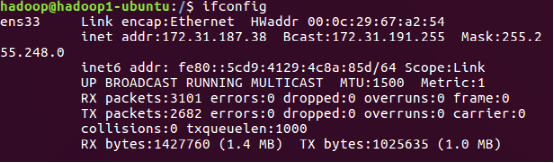
命令为：



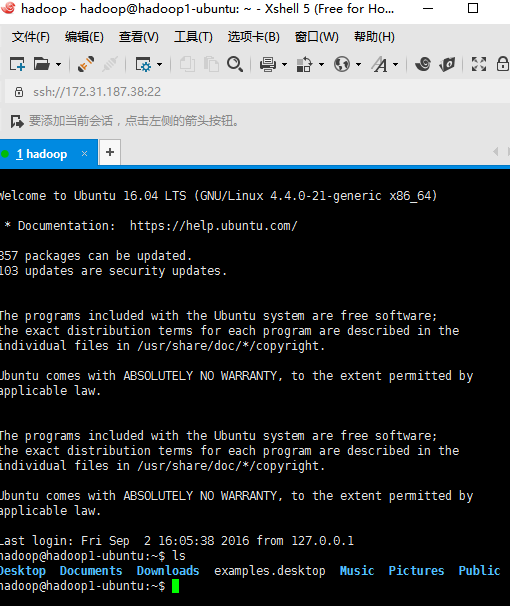
我的已经安装完毕，出现以上的画面

其中远程控制工具有很多中  Xshell   secureCRT

我输入ifconfig 查看ip

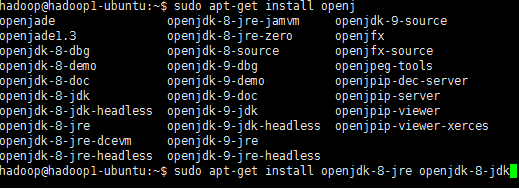


我使用Xshell5连接ubuntu



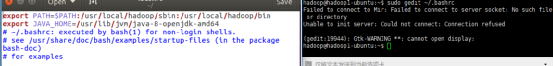
第四步：

安装java  jdk  jre



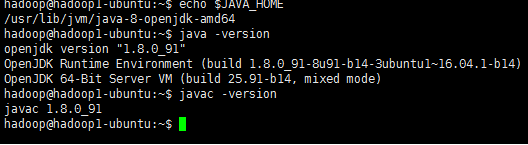
安装好 OpenJDK 后，需要找到相应的安装路径，这个路径是用于配置 JAVA\_HOME 环境变量的。

使用sudo gedit ~/.bashrc 编辑，建立环境变量。



export JAVA\_HOME=JDK安装路径

其中，然后必须使用source  ~/.bashrc  命令使变量生效。

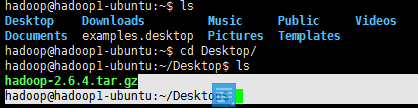


最后使用上述命令检验我们安装的java环境变量是否有效。如果有效，我们继续下一步。否则，请配置好java的环境。不然，无法进行下一步。

第五步：

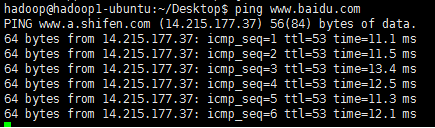
安装hadoop

我使用的是hadoop-2.6.4.tar.gz

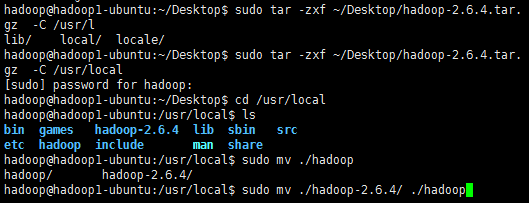


我放在桌面上。我是从window下好的，然后拖进ubuntu的，这个功能的实现，必须开机然后重新安装vmware tools才能使用  (可以建立一个共享的文件夹 可以使用ftp文件传送等等多种方式)

我建议直接在ubuntu中联网下载，然后在下载中找到那个文件。测试能否连接外网，使用ping [www.baidu.com](http://www.baidu.com/) 等方式。



然后，接下来安装hadoop ,我的安装包放在Desktop,建议安装在于与java一样的环境,我的是在/usr/local

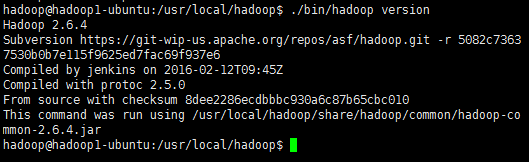


http://b264.photo.store.qq.com/psb?/V11kX7Pb3kyV5m/ZtO1EI*hqtBxuc8qrbHN7sL3.SEy6QFzOI8KcV2482U!/b/dAgBAAAAAAAA&bo=tgE4AAAAAAADAKo!

修改权限

Sudo chown -R hadoop ./hadoop

使用以下的命令检验hadoop的版本



一下稍微写了一下有关权限的介绍：

Chown：改变文件属主  只有管理员才有这个权限

Chown USERNAME file,...  -R  递归修改目录及其子文件的内部的文件的属主

--reference =/path/to/somefile file,...

#chgrp  GRPNAME file...    -R    --reference =/path/to/somefile file,...

Chown USERNAME:GROUP file....

Chown USERANME.GROUP file....

Chmod :修改用户的权限

修改三类用户的权限

Chmod  more  file....

-R

--reference =/path/to/somefile file,...

Rwxr-x---

修改某些用户或者某些类用户的权限

U  g  o  a

Chmod 用户类型 =MORE file....

修改某类的用户某位或者某些位权限

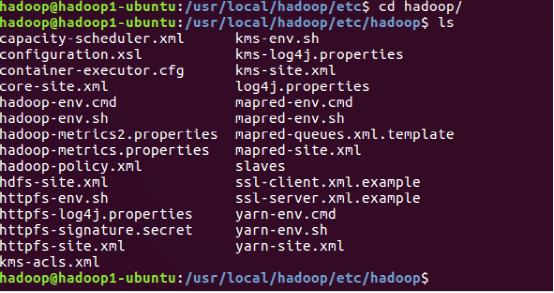
U   g   o  a

Chmod 用户类型 +(-)MORE file....

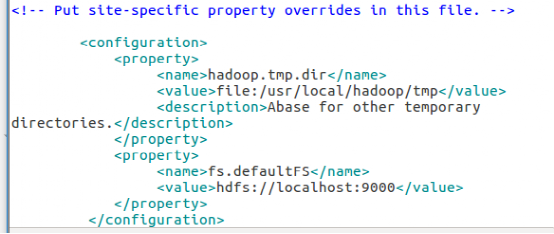
第六步：

Hadoop默认为非分布式的，即java单进程。接下来，我们进行伪分布式的配置

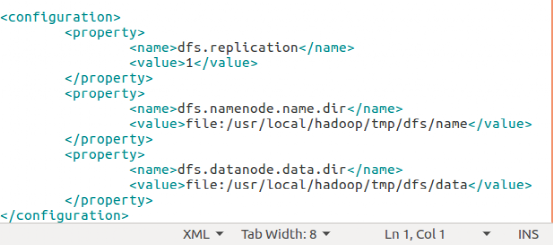
Hadoop 的配置文件位于 /usr/local/hadoop/etc/hadoop/  然后我们进入配置文件，进行修改，需要修改的是core-site.xml和hdfs-site.xml  使用sudo gedit ./etc/hadoop/core-site.xml



修改的core-site.xml文件为下面的截图：

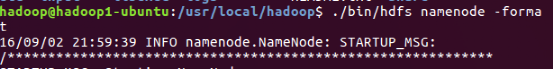


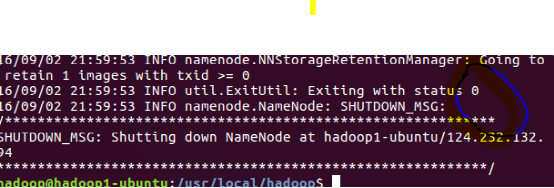
同理：修改hdfs-site.xml的截图如下：



第七步：开启hadoop

./bin/hdfs namenode -format 格式化namenode 记得必须在你安装的hadoop下的bin中使用

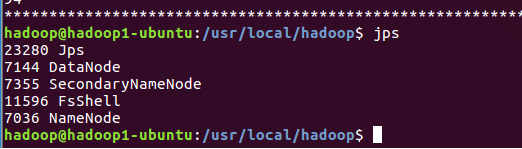




途中的0表示你已经成功了很多了，来继续我们的下面的操作。我们已经快要成功了，走起！！！  如果出错，然后查看有什么错误。

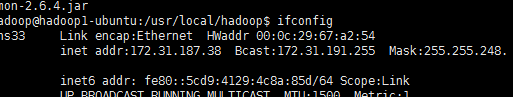
./sbin/start-dfs.sh 开启 NameNode 和 DataNode 守护进程。 第一次，会出现警告，yes即可。

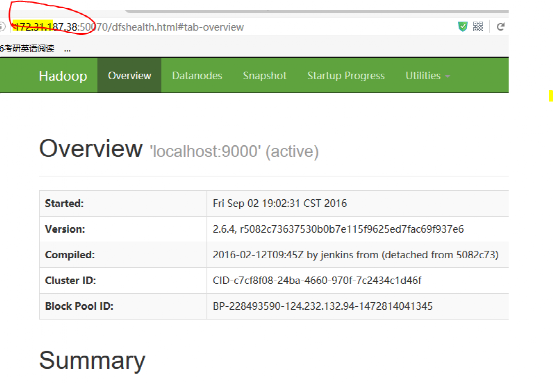
Jps查看是否打开成功。



可能失败的话，删除 log  tmp   使用rm -r ./删除的文件    格式

第八步：成功启动后，可以访问 Web 界面 查看相关信息





红色的使用你的ip地址  然后端口号为50070查看信息

这样已经成功了。哈哈，哈哈，爽吧！！！

第九步：

添加path路径：

因为上述操作都要到安装好了的hadoop下操作，比较麻烦，我们使用path环境变量，使我们在任意的目录可以，开启我们的服务。

修改~/.bashrc文件 与修改JAVA\_HOME设置类似

，亲，最后记得操作source ~/.bashrc 使设置生效。