**Hadoop实验教程**

**一、安装虚拟机软件VirtualBox （Vmware Workstation也可）**

**二、使用VirtualBox新建虚拟机并安装Ubuntu系统**

提示：硬盘不要分太小，后续计算有出问题的可能性

Ubuntu镜像下载地址 <https://ubuntu.com/download/desktop>

**三、创建一个新账户**

**1、使用CTRL+ALT+T打开命令行终端，输入命令创建新用户**

　　　　sudo useradd -m hadoop -s /bin/bash

　　　　说明：-m表示自动创建用户目录（/home/hadoop），hadoop是用户名，-s表示使用/bin/bash作为shell (**注意用户名能体现是自己做的**)

**2、给新用户设置密码**

　　　　sudo passwd hadoop

　　　　说明：输入两次新密码

**3、给新用户增加管理员权限**

　　　　sudo adduser hadoop sudo

**4、注销当前用户，并以新用户登录（登录新建的hadoop账户）**

**四、在Ubuntu中挂载共享文件夹**

**VirtualBox：**

**1、“设备”->“共享文件夹”，记住共享文件夹的共享名**

**2、“设备”->“安装增强功能“**

　　　　说明：安装好了，桌面上会出现光盘图标VBox\_GAs\_5.2.18和一个虚拟盘sf\_VirtualBoxShare

**3、进入目录/media/hadoop/VBox\_GAs\_5.2.18，执行命令./VBoxLinuxAddtion.run**

**4、将共享文件夹挂载到Ubuntu的一个文件夹**

　　　　（1）创建挂载的目标文件夹

　　　　　　mkdir  ~/share

　　　　（2）挂载

　　　　　　格式：mount -t vboxsf 共享名 本地目录

　　　　　　mount -t vboxsf VirtualBoxShare  ~/share

　　　　　　说明：VirtualBox是共享文件夹的共享名，share目录需要事先在家目录（~，即/home/hadoop） 下创建

**5、设置vbox共享粘贴板**

　　　　“设备”->“共享粘贴板”->“双向“

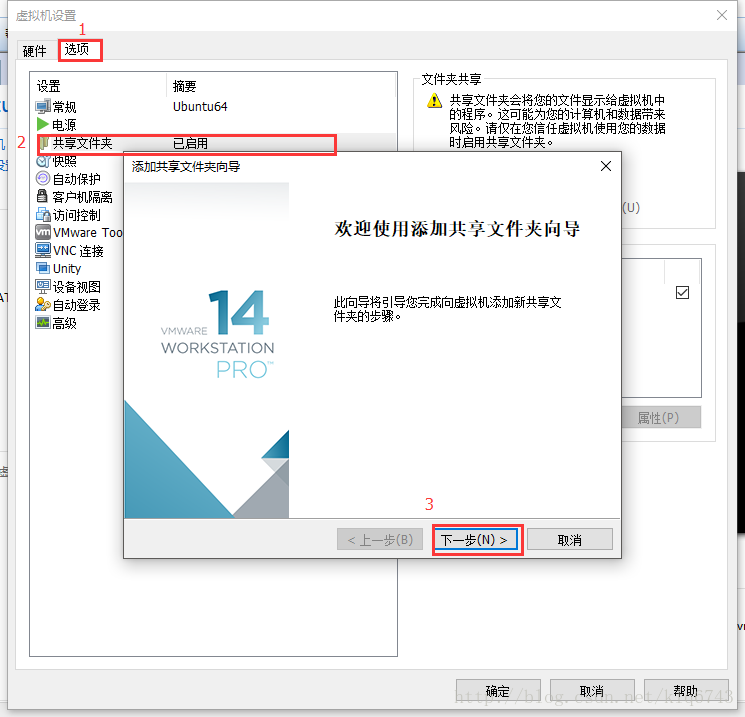
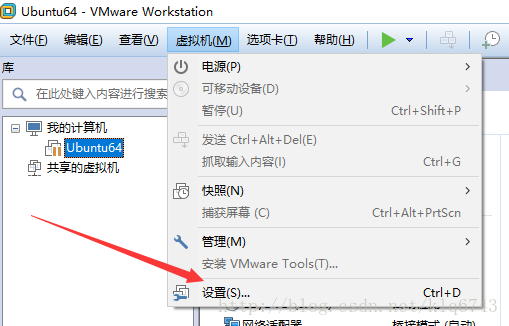
　　　　说明：这样在真机和虚拟机之间可以任意的复制和粘贴。

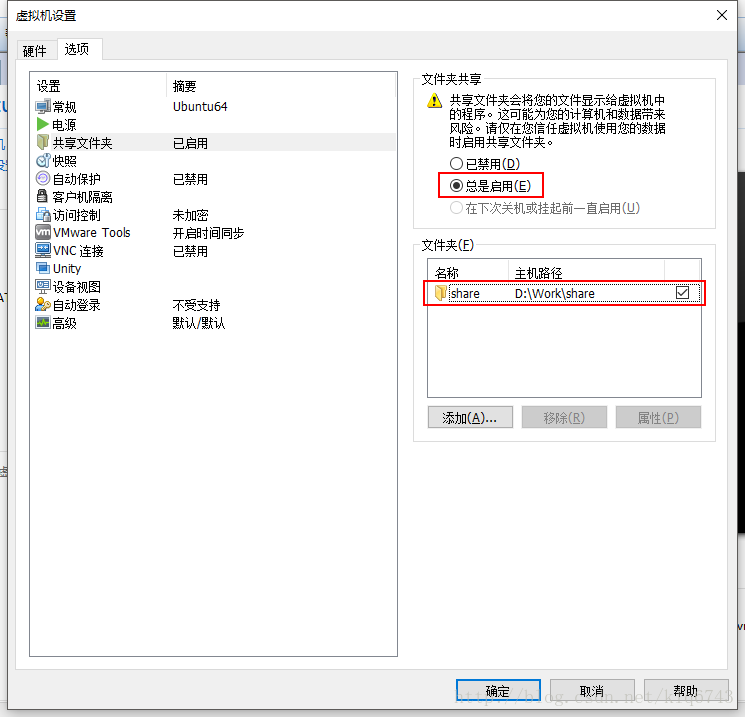
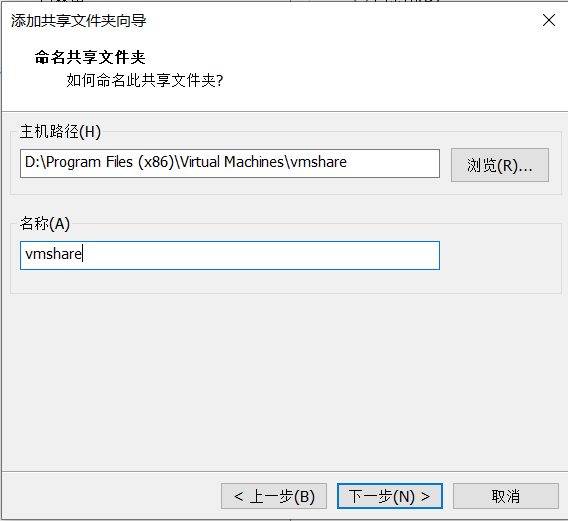
**Vmware Workstation：**

**1.依次点击VMWare窗口左上方的“虚拟机”、“设置”**

**2.弹出的对话框中依次点击“选项”、“共享文件夹”、“下一步”**

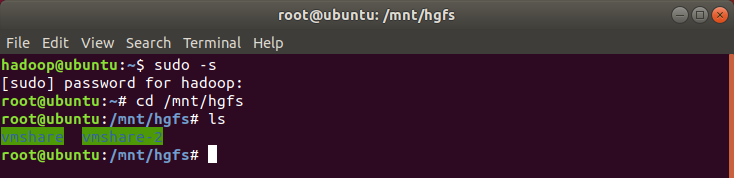
**3.选择要共享的文件夹路径，并填写名称。**





**4.进入虚拟机，安装VMware Tools：依次点击“虚拟机”、“重新安装VMware Tools”（1. 具体安装方法自行查找2. 如果该选项是灰色，说明已经安装好，无需再次安装）**

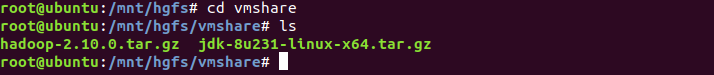
**5.至此，在/mnt/hgfs/下应该就能看到刚才在虚拟机中设置的windows下的共享文件夹vmshare了**

****

**6.如果到上面一步时/mnt/hgfs/下没有共享文件夹,再加一条命令即可：**

**vmhgfs-fuse .host:/ /mnt/hgfs/**

**这时把你在windows下的文件放入共享文件夹vmshare中，再回到Ubuntu中，cd 到/mnt/hgfs/vmshare，就可以看到这些共享的文件了。**

****

**五、安装一些必要的工具**

**1、更新apt-get**

　　　　sudo apt-get update (**将用户模式切换到根模式**)

　　　　说明：apt-get是用来安装软件的工具

**2、安装ssh服务**

　　　　apt-get install openssh-server

　　　　说明： ssh是机器之间远程登录的服务

**3、安装vim (如果不习惯使用vim编辑器可不安装)**

　　　　apt-get install vim

**六、配置免密登录**

**1、禁用防火墙**

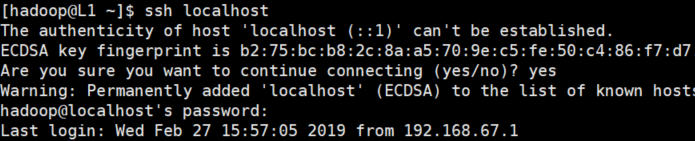
　　　　ufw disable

　　　　ufw status //查看防火墙状态

**2、生成密钥对**

ssh localhost（**测试一下 SSH 是否可用，要在用户模式下登录**）

exit



cd ~/.ssh

　　　　ssh-keygen  -t  rsa

　　　　说明：一路回车，在~/.ssh目录会生成一个两个文件，id\_rsa和id\_rsa.pub

**3、将公钥复制到localhost**

　　　　ssh-copy-id localhost

　　　　说明：也可cat  ./id\_rsa.pub >> ./authorized\_keys，这样~/.ssh目录又会多一个文件authorized\_keys

**4、登录测试**

　　　　ssh localhost （**看是否需要输入密码**）

**5、退出与localhost的连接**

　　　　exit

**七、安装并配置jdk**

**1、在家目录（~）下新建/app 文件夹**

　　　　mkdir  ~/app

 　　　　说明：后面的jdk、hadoop等都安装在此

**2、在真机中将jdk、hadoop的安装包复制到真机的共享文件夹中**

**jdk、hadoop安装包下载地址（尽量下载最新的适用版本）**

<https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

<https://mirrors.cnnic.cn/apache/hadoop/common/hadoop-2.10.1/>

**3、安装jdk**

tar -zxvf /mnt/hgfs/vmshare/jdk-8u181-linux-x64.tar.gz -C ~/app

　　　　mv ~/app/jdk1.8.0\_181 ~/app/jdk (**注意文件名匹配**)

**4、配置jdk**

　　（1）编辑配置文件.bashrc

　　　　vim ~/.bashrc （**不习惯用vim编辑器可改用 gedit ~/.bashrc，下同**）

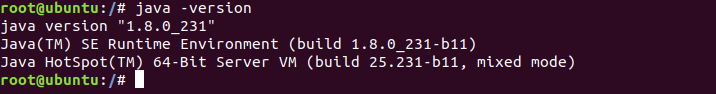
　　　　　export JAVA\_HOME=/home/hadoop/app/jdk  
　　　　　export JRE\_HOME=$JAVA\_HOME/jre  
　　　　　export CLASSPATH=.:$CLASSPATH:$JAVA\_HOME/lib:$JRE\_HOME/lib  
　　　　　export PATH=$PATH:$JAVA\_HOME/bin:$JRE\_HOME/bin

　　（2）编译配置文件

　　　　source ~/.bashrc

　　（3）测试jdk

　　　　java -version



**八、安装hadoop**

**1、安装hadoop**

　　　tar -zxvf /mnt/hgfs/vmshare/hadoop-2.7.3.tar.gz -C ~/app //解压

　　　mv ~/app/hadoop-2.7.3 ~/app/hadoop //改名 (**注意文件名匹配**)

sudo chown -R hadoop ./hadoop //修改hadoop权限

**2、配置hadoop**

　　　　（1）配置环境变量

　　　　　　vim ~/.bashrc，添加或修改以下语句：

export HADOOP\_HOME=/app/hadoop

export PATH=$PATH:$JAVA\_HOME/bin:$JRE\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

　　　　　　source ~/.bashrc //编译生效

　　　　（2）配置hadoop-env.sh

　　　　　　vim ~/app/hadoop/etc/hadoop/hadoop-env.sh

　　　　　　找到export JAVA\_HOME=${JAVA\_HOME}这一行，将其修改为：  
　　　　　　export JAVA\_HOME=/home/hadoop/app/jdk

　　　　（3）配置core-site.xml (**均位于~/app/hadoop/etc/hadoop/，下同**)

　　　　　　<configuration>  
　　　　　　　　<property>  
　　　　　　　　　　<name>fs.defaultFS</name>  
　　　　　　　　　　<value>hdfs://localhost:9000</value>  
　　　　　　　　</property>  
　　　　　　　　<property>  
　　　　　　　　　　<name>hadoop.tmp.dir</name>  
　　　　　　　　　　<value>/home/hadoop/app/hadoop/tmp</value> <!--该目录需要手动创建 -->   
　　　　　　　　</property>  
　　　　　　</configuration>

　　　　（4）配置hdfs-site.xml

　　　　　　<configuration>  
　　　　　　　　<property>

　　　　　　　　　　<!--配置块的副本数 -->  
　　　　　　　　　　<name>dfs.replication</name>  
　　　　　　　　　　<value>1</value>  
　　　　　　　　</property>  
　　　　　　</configuration>

　　　　（5）配置mapred-site.xml

　　　　　　说明：需将mapred-site.xml.template复制成mapred-site.xml

　　　　　　<configuration>  
　　　　　　　　<property>  
　　　　　　　　　　<name>mapreduce.framework.name</name>  
　　　　　　　　　　<value>yarn</value>  
　　　　　　　　</property>  
　　　　　　</configuration>

　　　　（6）配置yarn-site.xml

　　　　　　<configuration>

　　　　　　　　<property>  
　　　　　　　　　　<name>yarn.nodemanager.aux-services</name>  
　　　　　　　　　　<value>mapreduce\_shuffle</value>  
　　　　　　　　</property>  
　　　　　　</configuration>

**九、运行hadoop**

**1、格式化namenode**

　　　　cd ~/app/hadoop/bin

./hadoop namenode -format

　　　　说明：首次运行须格式化，以后运行就不需要这一步。格式化后会在core-site.xml中hadoop.tmp.dir所指定的目录中生成name文件夹。

**2、运行hadoop (要在用户模式下)**

　　　　cd ~/app/hadoop/sbin

　　　　./start-all.sh

**3、查看进程：**

　　　　jps

　　　　　　14262 NameNode  
　　　　　　15976 NodeManager  
　　　　　　14385 DataNode  
　　　　　　15884 ResourceManager  
　　　　　　14601 SecondaryNameNode  
　　　　　　16120 Jps  
　　　　如果出现以上进程，说明hadoop环境搭建成功，分布式系统启动完成。

**4、webUI查看**

<http://localhost:50070> //查看hdfs

<http://localhost:8088> //查看mapreduce作业情况

**十、测试wordcount程序**

**1、生成单词文件**

　　　　mkdir ~/tmp  
　　　　echo 'hello world hello hadoop' > ~/tmp/word1.txt  
　　　　echo 'hive run on hadoop' > ~/tmp/word2.txt

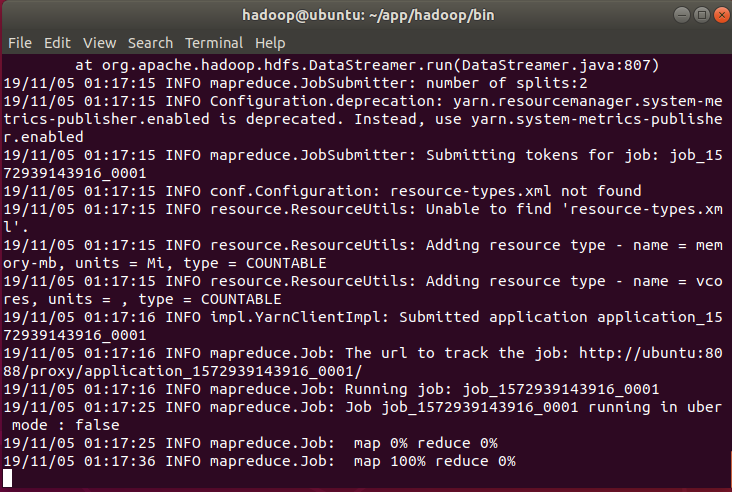
**2、上传至hdfs**

　　　　（1）在hdfs上新建输入数据的目录  
　　　　　　./hdfs dfs -mkdir /input  
　　　　（2）上传文件 (**~/app/hadoop/bin**)

　　　　　　./hdfs dfs -put  ~/tmp/word\*.txt  /input

**3、运行hadoop自带的单词计数程序**

　　　　./hadoop jar ~/app/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /input output



**4、查看结果**

hdfs dfs -cat /user/hadoop/output/part-r-00000

　　　　可以看到每个单词出现的次数

**十一、实现并测试矩阵相乘程序（必做）**

https://blog.csdn.net/catharryy/article/details/9186757