**Hadoop集群配置安装文档**

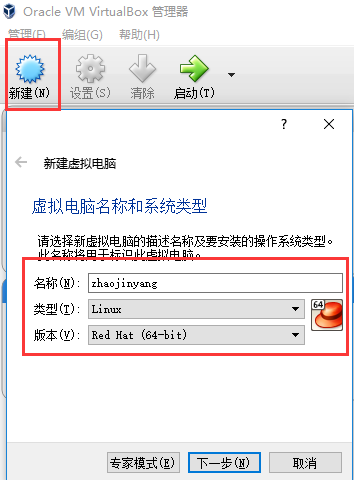
# 一、配置大数据平台宿主操作系统

## 1.1、安装虚拟机

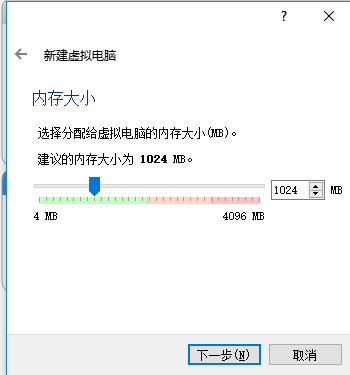
安装步骤：开始安装VirtualBox-5.1.18-114002-Win，然后一直点下一步，没有特殊要注意的地方。如下图所示。



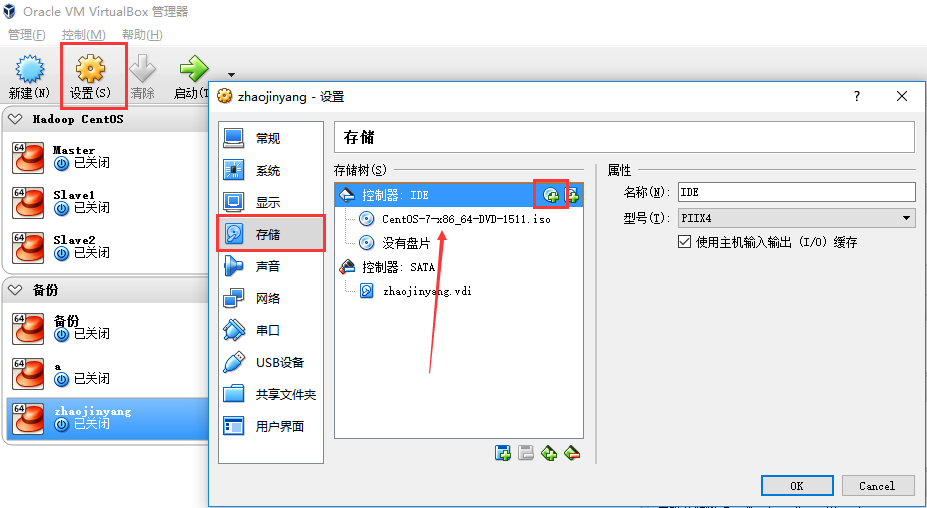
1、打开虚拟机virtualBox点击新建，配置系统的名字、操作系统的版本及类型，进行下一步。如下图所示。

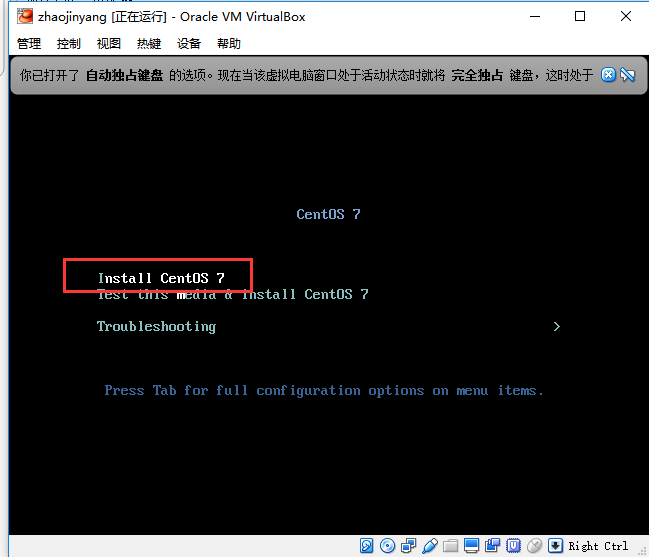


2、选择分配虚拟机大小，根据自己电脑的实际配置情况选择虚拟机内存大小的分配。因为后边要部署hadoop集群，要建一个master和两个slave，所以这里我的虚拟机内存分配大小选择512M。如下图所示。

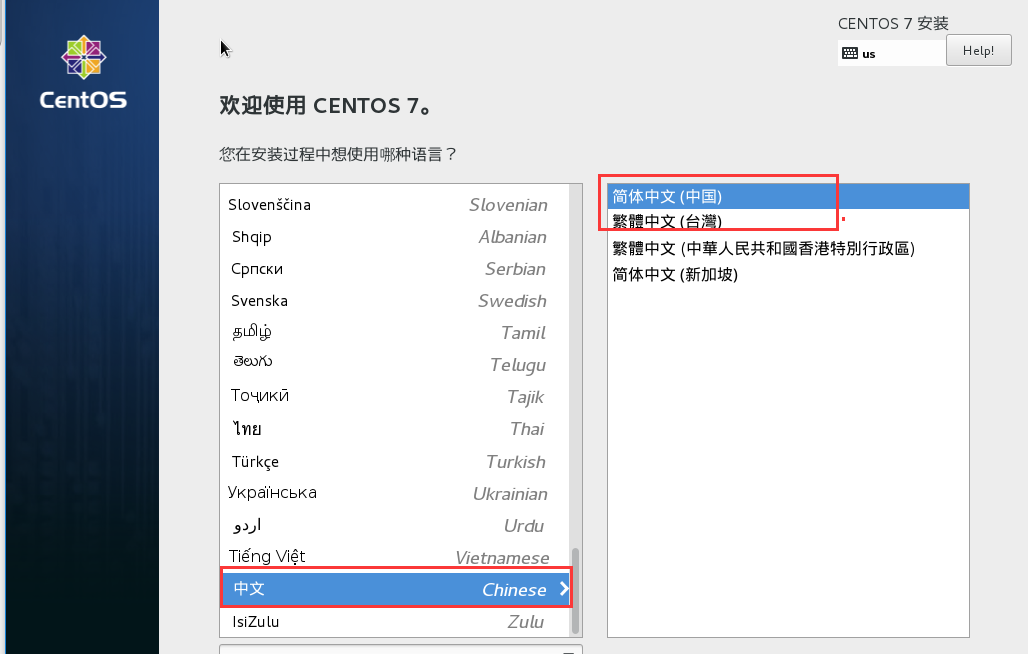


3、为虚拟机做了基本的配置后，开始安装操作系统，打开设置中存储，添加CentOS-7镜像进行安装。如下图所示。CentOS-7-x86\_64-DVD-1511



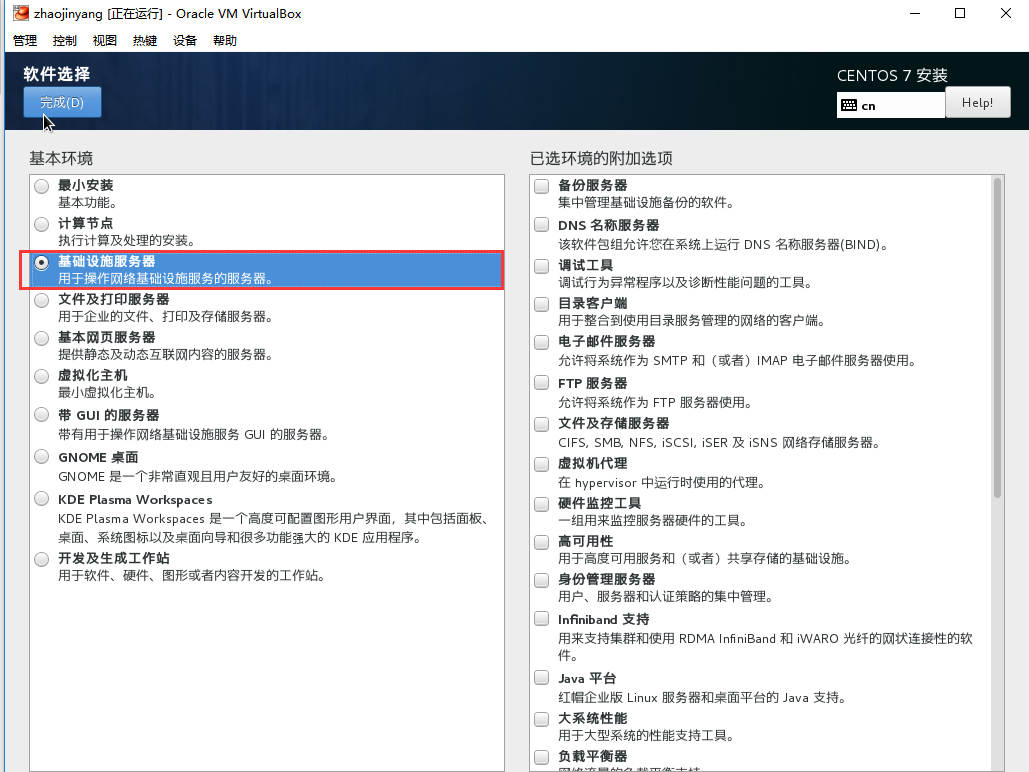


4、按照需求为虚拟机系统选择语言，如下图所示。



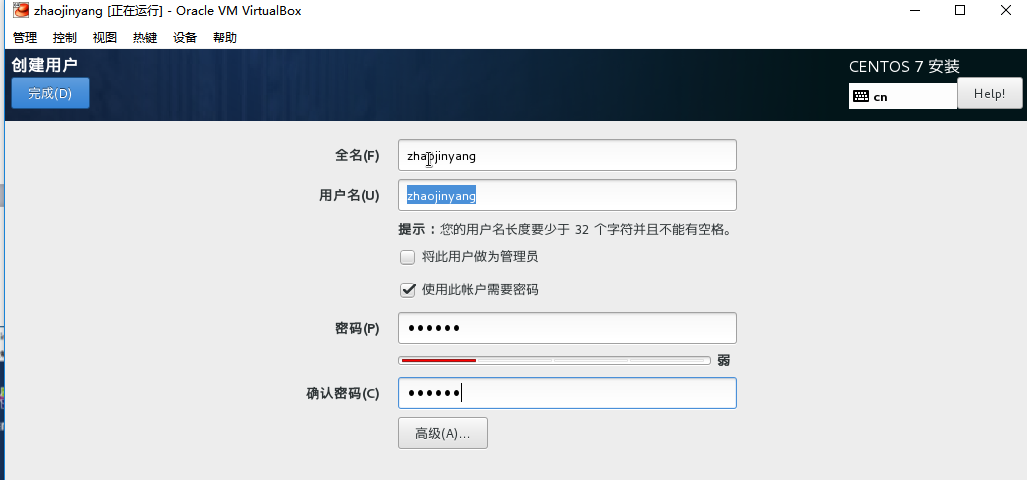
1. 按照需求为虚拟机系统选择时间日期、语言，在这里的软件选择需要特殊注意，以为后边需要配置hadoop集群做服务器，这里点开软件选择将最小安装更改为基础设施服务器，此处需要特殊注意。如下图所示。





6、点击下一步后为虚拟机系统创建用户及root密码。如下图所示。







1. 做完上述步骤后，等待虚拟机的安装，重启虚拟机，完成虚拟机的安装。

# 二、大数据HDFS平台的搭建

## 2.1、软件准备

Virtualbox、CentOS、Hadoop、JDK、Xshell、Xftp

## **2.2、安装虚拟机**

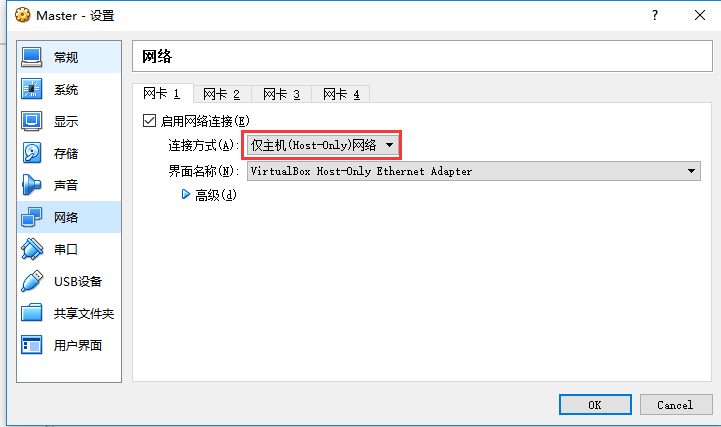
1. 、保留一个干净的虚拟机，随时可以提供复制。
2. 、在安装CentOs时，注意将其设置为基础服务器。

**2.3、查看主机的IP**



## **2.3、在虚拟机中设置网络**

（1）、在虚拟机中选用host-only模式。



（2）、设置虚拟机的网络

vi /etc/sysconfig/network

# Created by anaconda

NETWORKING=yes

GATEWAY=192.168.56.1



1. 、在虚拟机中配置主机ip

vi /etc/sysconfig/network-scripts/ifcfg-enp0s3

TYPE=Ethernet

IPADDR=192.168.56.100

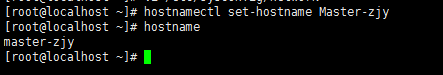
NETMASK=255.255.255.0



1. 、重启网络设置并将虚拟机主机名更改

service network restart

hostnamectl set-hostname master

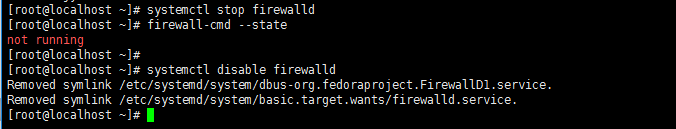


1. 、关闭windows和linux的防火墙。

systemctl stop firewalld

firewall-cmd --state

systemctl disable firewalld



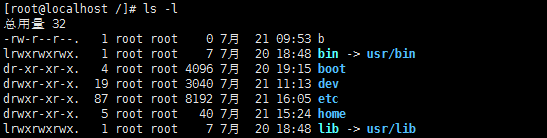
# 三、Linux文件操作部分命令使用

## **3.1、Ls命令**

ls命令查看文件夹包含的文件，后边跟一些参数，可以查看文件权限等信息。

ls -a列出目录所有文件。

ls -la列出包含隐藏文件



## **3.2、pwd查看当前路径**



## **3.3、Mkdir创建、删除文件夹**

mkdir file 建立文件夹file

rmdir file 删除文件夹file

mv file newfile mv移动文件或修改文件名，将file文件夹修改为newfile

## **3.4、cp命令**

将源文件复制至目标文件，或将多个源文件复制至目标目录。如果不加-i参数，则不会提示，而是直接覆盖。

cp a.txt b.txt 将文件a.txt复制给b.txt

## **3.5、cat命令**

         cat主要功能：

  1.显示整个文件 cat a.txt

## **3.6、more命令**

功能类似于cat, more会以一页一页的显示方便使用者逐页阅读，而最基本的指令就是按空白键（space）就往下一页显示，按 b 键就会往回（back）一页显示。

cat a.txt | more

## **3.7、less命令**

less 与 more 类似，但使用 less 可以随意浏览文件，而 more 仅能向前移动，却不能向后移动，而且 less 在查看之前不会加载整个文件。

## **3.8、用户和用户组**

添加用户

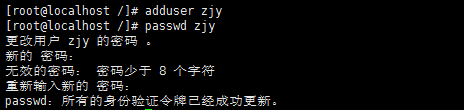
adduser zjy //添加一个名为zjy的用户

userdel zjy //删除用户

passwd zjy //修改用户密码

cat /etc/passwd //查看所有用户信息

Whoami //查看当前用户



ll -d /home/zjy //显示创建用户名的初始权限为700

1

添加、删除、修改、查看用户和用户组

groupadd test //创建组test

groupdel test //删除组test

groupmod -n test2 test //将test名改为test2

cat /etc/group //查看组

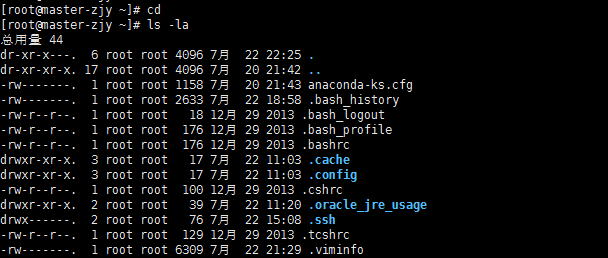
usermod -G group1 zjy //将zjy添加到用户组group中

cat /etc/passwd //查看用户是否可以登录， nologin 标志的话，那么该用户就是无法登录进系统的

## **3.9、配置ssh免密码登录**

（1）、执行命令cd退回根目录：cd

（2）、运行ls -la (隐藏文件夹）

（3）、在根目录下运行ssh-keygen -t rsa连续四下回车生成密钥对私钥id\_rus和公钥id\_rsa.pub,默认存储在/root/.ssh中

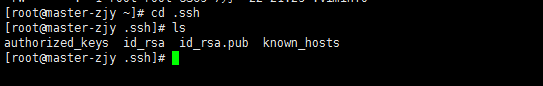
（4）、将公钥分别发给Slava1-zjy和Slave2-zjy和自己Master-zjy

ssh-copy-id Slave1-zjy

ssh-copy-id Slave2-zjy

ssh-copy-id Master-zjy

（5）、查看私钥和公钥



（6）、3台虚拟机分别执行命令：ssh 主机名或ip，如果都能登录则配置成功。

ssh Master-zjy

ssh Slave1-zjy

ssh Slave2-zjy

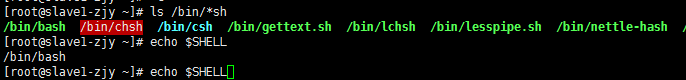


3.10、SHELL的使用

ls /bin/\*sh //查看系统安装类型

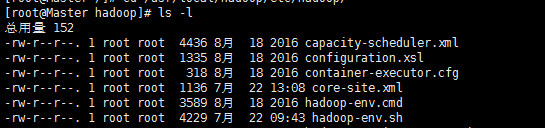
echo $SHELL //查看系统当前默认shell

Exit //退出当前shell使用



3.11、文件操作

Ls -l 查看文件类型 r可读 w可写 x可执行



-：表示该文件是普通文件。

d：表示该文件是一个目录。

l：表示该文件是一个链接文件。

b：表示该文件是块设备文件，块设备通常为硬盘、光驱等传输速度较快的物理设备。

c：表示键盘、鼠标等传输速度较慢的设备文件。

File //查看文件类型

wc yarn-site.xml //查看文件大小

Cat/more/less/head /tail yarn-site.xml //不同查看文件的命令

touch a.txt //创建文件命令

IMG_256

ll /usr/\* >mylist.txt //使用重定向，将/usr/目录下的文件输入到mylist.txt文件中 其中是对源文件的覆盖，>>作用是对存在文件追加。

| 管道是将前一个命令的标准输出所为后一个命令的标准输入。

find /usr/local/ -name h\* //通配符使用

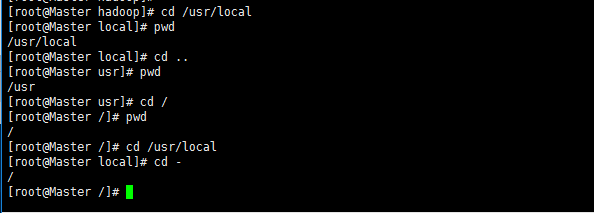
Man find //帮助文档，查阅find的使用

Info find //帮助文档

，查阅find的使用

find --help //帮助文档，查阅find的使用

cat hadoop-env.sh | more //管道的使用，使用q可以退出



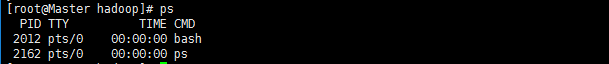
cd /usr/local //切换到根目录下的usr/bin子目录

cd .. //返回上一级目录

cd / //切换根目录

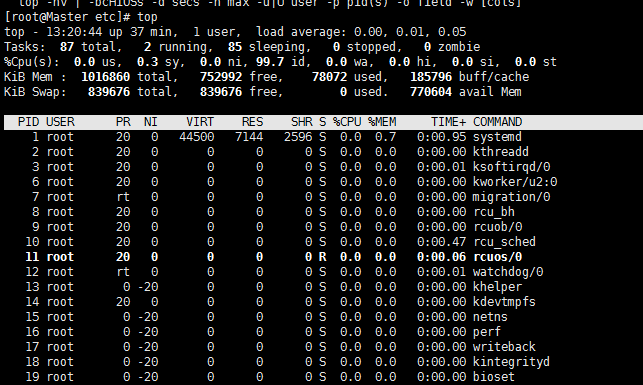
cd //回到当前用户的主目录

cd - //回到当前目录之前所在的目录



Ps //查看当前系统运行的进程

Top //显示各个进程所占资源



Kill //关门某个进程

# 心得体会

本次作业的主要任务就是完成安装虚拟机，配置大数据平台宿主操作系统CentOS和Linux文件操作命令使用，在安装操作系统的时候，注意内存大小的分配，和操作系统的版本，我第一次安装CentOS的时候，总是失败，后来我上网查找原因，发现是版本的问题，我就到官网上下载CentOS-7，总算是解决了问题。

Linux基本命令的使用，由于长时间不使用linux,一些命令需要查找帮助文档才能使用，本次我在完成作业之后，还把hadoop1.0的平台给搭建了。