

# 安装虚拟机

安装Vmware或VirtualBox。

Vmware，收费产品，占内存较大。

Oracle的VirtualBox，开源产品，占内存较小，但安装ubuntu过程中，重启会出错。

最好选Vmware。

# 下载安装Linux操作系统

安装Ubuntu 16.04

下载地址：<http://mirrors.163.com/ubuntu-releases/16.04/>

# 下载hadoop3.0和jdk1.8

## 下载hadoop 3.0.0

下载hadoop 3.0.0

下载地址：<https://hadoop.apache.org/releases.html>

## 下载jdk1.8

下载jdk1.8

下载地址：<http://www.oracle.com/technetwork/cn/java/javase/downloads/jdk8-downloads-2133151-zhs.html>

# 安装必备软件

## 在ubuntu上安装一些必备的软件:vim、ssh

#安装vim,执行：

sudo apt-get install vim

#安装ssh,执行：

sudo apt-get install openssh-server

#查看ssh状态,执行：

service ssh status

#如果ssh关闭使用,开启ssh服务,执行：

service ssh start

#安装lrzsz,便于传输文件执行：

apt-get install lrzsz

## 在Windows上安装一些连接linux的软件：X Shell

XShell教育版下载地址：[需用自己的邮箱注册，取到lisence]

<https://www.netsarang.com/download/down_form.html?code=522>

# 修改节点主机名和网络配置

## 分别配置各节点的网络

---**master节点**---

#打开hostname

sudo vim /etc/hostname

#配置如下

-------------------配置内容------------------

master

-----------------------------------------------

---**slave1节点**---

#打开hostname

sudo vim /etc/hostname

#配置如下

-------------------配置内容------------------

slave1

-----------------------------------------------

---**slave2节点**---

#打开hostname

sudo vim /etc/hostname

#配置如下

-------------------配置内容------------------

slave2

-----------------------------------------------

---**slave3节点**---

#打开hostname

sudo vim /etc/hostname

#配置如下

-------------------配置内容------------------

slave3

-----------------------------------------------

#2修改每台电脑的hosts文件,hosts文件和windows上的功能是一样的。存储主机名和ip地址的映射。执行

sudo vim /etc/hosts

#配置如下

-------------------配置内容------------------

127.0.0.1 localhost

192.168.199.120 master

192.168.199.121 slave1

192.168.199.122 slave2

192.168.199.123 slave3

-----------------------------------------------

## 配置ubuntu静态IP

#1.首先用root用户登陆，输入root的密码，命令：

su

#2.编辑interfaces文件，命令：

vim /etc/network/interfaces

#3.添加以下内容进去，命令：

-------------------配置内容------------------

# The primary network interface

auto enp0s3

iface enp0s3inet static

address 192.168.199.120

netmask 255.255.255.0

gateway 192.168.199.1

-----------------------------------------------

#4.配置DNS解析，命令：

vim /etc/resolv.conf

-------------------配置内容------------------

nameserver 202.116.32.254

nameserver 114.114.114.114

-----------------------------------------------

#5.改完上面，如果重启的话，还是会变为原来的样子，所以要让其永久改变，命令：

vim /etc/resolvconf/resolv.conf.d/base

-------------------配置内容------------------

nameserver 202.116.32.254

nameserver 114.114.114.114

-----------------------------------------------

#6.重启网卡,命令：

sudo /etc/init.d/networking restart

如果重启网卡不行，重启机器：reboot

#7.测试

#ping 网关

ping 192.168.199.1

#ping 百度的服务器ip

ping 14.215.177.37

#ping 百度域名

ping [www.baidu.com](http://www.baidu.com)

**如果测试没问题，则按照此法，配置其他slave节点的静态IP。**

# 配置ssh无密码登录

#1.生成密钥

ssh-keygen -t rsa

#2.打开~/.ssh 下面有三个文件

#authorized\_keys(已认证的keys)

#id\_rsa 私钥

#id\_rsa.pub 公钥

#如果没有authorized\_keys，生成authorized\_keys。命令:

touch authorized\_keys

#3.在master上将公钥放到authorized\_keys里。命令：

sudo cat id\_rsa.pub >> authorized\_keys

#4.修改authorized\_keys权限，命令：

chmod 644 authorized\_keys

#5.将master上的公钥放到其他节点的authorized\_keys中。

#6.修改authorized\_keys权限，命令：

chmod 644 authorized\_keys

#7.测试是否成功

ssh slave1

输入用户名密码，然后退出，

再次ssh slave1不用密码，直接进入系统。这就表示成功了。

**测试成功后，按照此法，配置其他slave节点。**

# 安装jdk1.8

1.将前面下载的jdk1.8拷贝到虚拟机中 /opt 的目录

2. 解压JDK

sudo tar zxvf jdk-8u31-linux-x64.tar.gz

3.设置环境变量（全局）

sudo vim /etc/profile

打开profile文件输入

-------------------显示信息------------------

export JAVA\_HOME=/opt/jdk1.8.0\_131

export CLASSPATH=".:$JAVA\_HOME/lib:$CLASSPATH"

export PATH="$JAVA\_HOME/bin:$PATH"

-----------------------------------------------

4.设置系统默认JDK

sudo update-alternatives --install /usr/bin/java java /opt/jdk1.8.0\_131/bin/java 300

sudo update-alternatives --install /usr/bin/javac javac /opt/jdk1.8.0\_131/bin/javac 300

sudo update-alternatives --config java

5. 验证JDK,输入命令

java -version

见到JDK的信息则表示成功。

# 安装hadoop

#1.上传hadoop压缩文件，到/opt下，解压缩

tar -zxvf hadoop3.0.0.tar.gz -C /opt

#2.再重命名一下,这样目录就变成/opt/hadoop3.0

sudo mv hadoop3.0

①修改环境变量，将hadoop加进去[2]

sudo vim /etc/profile

# -------------------config content------------------

export HADOOP\_HOME=/opt/hadoop3.0

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

# ---------------------------------------------------

#修改完后，用source /etc/profile让配置文件生效。

source /etc/profile

②修改/opt/hadoop3.0/etc/hadoop下配置文件(也可直接使用附件里hadoopconf.tar.gz的文件替换)

sudo vim /opt/hadoop3.0/etc/hadoop/hadoop-env.sh

# -------------------配置内容------------------

export JAVA\_HOME=/opt/jdk1.8.0\_131

# -----------------------------------------------

sudo vim /opt/hadoop3.0/etc/hadoop/core-site.xml

# -------------------配置内容------------------

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

</configuration>

# ---------------------------------------------------

sudo vim /opt/hadoop3.0/etc/hadoop/hdfs-site.xml

# -------------------配置内容------------------

<configuration>

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>/data/hd3/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>/data/hd3/datanode</value>

</property>

</configuration>

# ---------------------------------------------------

sudo vim /opt/hadoop3.0/etc/hadoop/mapred-site.xml

# -------------------配置内容------------------

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

</configuration>

# ---------------------------------------------------

#修改yarn-site文件

sudo vim /opt/hadoop3.0/etc/hadoop/yarn-site.xml

# -------------------配置内容------------------

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

<name>yarn.resourcemanager.hostname</name>

<value>master</value>

</property>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.env-whitelist</name>

<value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PREPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME</value>

</property>

<property>

<name>yarn.nodemanager.vmem-check-enabled</name>

<value>false</value>

</property>

<property>

<name>yarn.nodemanager.resource.memory-mb</name>

<value>49152</value>

</property>

<property>

<name>yarn.scheduler.maximum-allocation-mb</name>

<value>49152</value>

</property>

</configuration>

# ---------------------------------------------------

#修改workers文件

sudo vim /opt/hadoop3.0/etc/hadoop/workers

# -------------------配置内容-------------------------

slave1

slave2

slave3

# ---------------------------------------------------

#上面的hadoop-env.sh，core-site.xml，mapred-site.xml，hdfs-site.xml，yarn-site.xml，workers几个文件，在四台linux中都是一样的。

#配置完一台电脑后，可以将hadoop包，直接拷贝到其他电脑上。

③将hadoop的用户加进去

#创建组 hadoop ,并将用户 lpf加入hadoop组中

sudo groupadd -g 888 hadoop

sudo gpasswd -a lpf hadoop

#④在master节点上格式化namenode

hadoop namenode -format

#⑤启动hadoop程序,切换到 /opt/hadoop3.0/sbin/下

/opt/hadoop3.0/sbin/start-all.sh

#⑥查看进程，是否启动

jps

#master节点配置内容：

-------------------配置内容-------------------------

11290 Jps

11195 JobTracker

10924 NameNode

11119 SecondaryNameNode

---------------------------------------------------

slave1节点配置内容：

-------------------配置内容-------------------------

2440 TaskTracker

2345 DataNode

3149 Jps

---------------------------------------------------

**成功后，同样的方法配置slave2、slave3，显示结果与slave1相同。**

**其他使用**

#启动hadoop

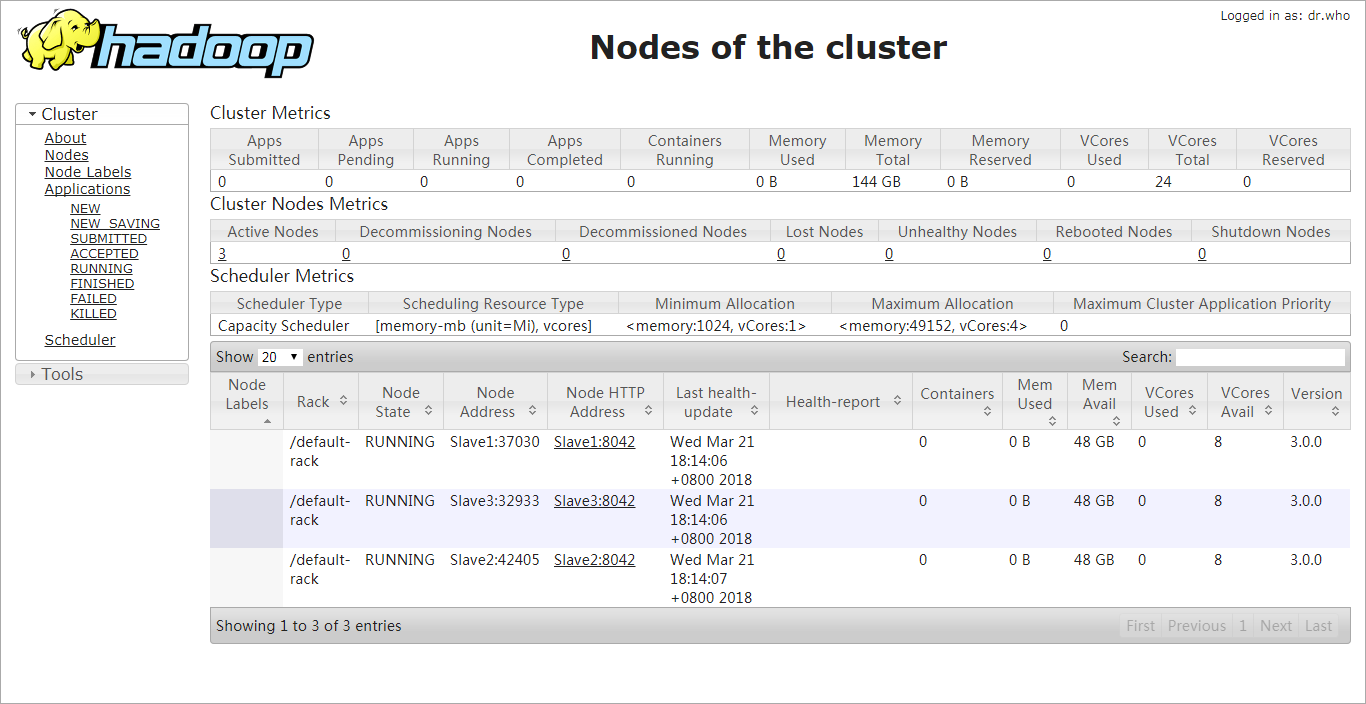
/opt/hadoop3.0/bin/start-all.sh

#停止hadoop

/opt/hadoop3.0/bin/stop-all.sh

在浏览器中输入：<http://192.168.199.120:8088/>

可以看到集群的监控。

******

# *参考*

1. Hadoop相关软件. <http://archive.apache.org/dist/hadoop/>
2. hadoop-3.0.0集群环境搭建、配置. http://blog.csdn.net/cc1949/article/details/78836141