# Hadoop集群搭建

集群规划

主机名 IP 软装的软件 运行程序

ubuntu00 192.168.72.100 jdk1.8、hadoop、zookeeper NM、DM

ubuntu01 192.168.72.101 jkd1.8、hadoop、 RM、

ubuntu02 192.168.72.102 jkd1.8、hadoop、 NM、DM

新增三台虚拟机，

hostname ip

ubuntu03 192.168.72.103

ubuntu04 192.168.72.104

ubuntu05 192.168.72.105

undo /etc/hosts

教程，较为详细： <https://linux.cn/article-8695-1.html>

1、关闭防火墙：

查看当前防火墙状态

# ufw status

inactive状态是防火墙关闭状态

active是开启状态。

关闭防火墙

# ufw disable

2、配置 hostname

# hostnamectl set-hostname namenode ### 为 namenode 节点设置 hostname

# hostnamectl set-hostname datanode1 ### 为 datanode1 节点设置 hostname

# hostnamectl set-hostname datanode2 ### 为 datanode2 节点设置 hostname

Log存储位置

修改 hadoop log 存储位置

hadoop@namenode:/usr/local/hadoop/etc/hadoop$ echo "export HADOOP\_LOG\_DIR=~/log/hadoop" >>hadoop-env.sh

系统粘贴板不能使用：刚刚还是可以使用的

查找原因：

# vim

：reg

是否有 + 寄存器，如果没有 就重启mobaxterm 或者虚拟机

<https://blog.csdn.net/zhangxiao93/article/details/53677764>

无法建立连接：

The authenticity of host 'datanode1 (192.168.72.101)' can't be established.

ECDSA key fingerprint is SHA256:apttw6zRXnZFKX3O0L7ScHlXg2Bnb60oOZyqw/jucsY.

Are you sure you want to continue connecting (yes/no)? The authenticity of host 'datanode2 (192.168.72.102)' can't be established.

ECDSA key fingerprint is SHA256:apttw6zRXnZFKX3O0L7ScHlXg2Bnb60oOZyqw/jucsY.

Are you sure you want to continue connecting (yes/no)? localhost: datanode running as process 2086. Stop it first.

namenode: datanode running as process 2086. Stop it first.

datanode2: Host key verification failed.

datanode1: Host key verification failed.

原因推测：Permanently did not added 'namenode' (ECDSA) to the list of known hosts.即namenode没有添加到.ssh目录下的known\_host中。

解决：使用ssh各自连接一遍，将其添加到host\_knowns

hadoop@namenode:$ ssh namenode

hadoop@namenode:$ ssh datanode1

hadoop@namenode:$ ssh datanode2

hadoop@datanode1:$ ssh namenode

hadoop@datanode1:$ ssh datanode1

hadoop@datanode1:$ ssh datanode2

hadoop@datanode2:$ ssh namenode

hadoop@datanode2:$ ssh datanode1

hadoop@datanode2:$ ssh datanode2

向~/bashrc加入一句话

export PATH=:/usr/local/hadoop/bin:/usr/local/hadoop/sbin

导致系统无法使用ls vim 等基本命令

解决：

/usr/bin/vim ~/.bashrc 将上一句话改为

export PATH=$PATH:/usr/local/hadoop/bin:/usr/local/hadoop/sbin

重启电脑

Namenode没有启动：

原因：可能是tmp导致的文件不一致

解决：改变存放dfs的位置

Vim core-site.xml

<property>

<name>hadoop.tmp.dir</name>

<value>/usr/local/hadoop/hadoop\_tmp</value>

</property>

Mobaxterm 远程ubuntu无法使用ls ，但是ubuntu自己可以使用ls？？？

晚上回家就好了，也没有关机。神奇

linux如何查看端口占用

1、lsof -i:端口号

2、netstat -tunlp|grep 端口号

集群总览：

| **NodeName** | **Web** | **Address** |
| --- | --- | --- |
| namenote | HDFS | <http://192.168.72.100:50070> |
|  | RM | <http://192.168.72.100:8088> |
|  | NM | <http://192.168.72.100:8042> |
| datanode1 | HDFS | <http://192.168.72.101:50075> |
|  | NM | http://192.168. 72.101:8042 |
| datanode2 | HDFS | <http://192.168.72.102:50075> |
|  | NM | http://192.168. 72.102:8042 |

由于容易引起歧义，不便于理解，将hostname改为ubuntu01、ubuntu02、ubuntu03

1. 修改hostname

# hostnamectl set-hostname ubuntu00

# hostnamectl set-hostname ubuntu01

# hostnamectl set-hostname ubuntu02

1. 修改hadoop配置文件slaves，

# cd /usr/local/hadoop/etc/hadoop/

ubuntu00

ubuntu01

ubuntu02

1. 修改/etc/hosts中的域名与ip,域名可以与主机名不相同，但为了便于记忆、理解最好相同

192.168.72.100 ubuntu00

192.168.72.101 ubuntu01

192.168.72.102 ubuntu02

1. 修改win10 hosts中的域名与ip，域名可以与主机名不相同，但为了便于记忆、理解最好相同

C:\Windows\System32\drivers\etc\hosts

192.168.72.100 ubuntu00

192.168.72.101 ubuntu01

192.168.72.102 ubuntu02

1. Ssh

# ssh-copy-id root@ubuntu00 #传递公钥给root@ubuntu00 各个用户+虚拟机都传递一遍

**Hadoop配置文件**

复制默认配置文件,在升级 hadoop 版本时可独立出来。

$ mkdir ~/config/

$ cp -r /usr/local/hadoop-2.8.0/etc/hadoop/ ~/config/

指定 hadoop 运行所使用的 JDK 与配置目录, vim ~/config/hadoop/hadoop-env.sh

$ export JAVA\_HOME=/usr/local/java/jdk1.8.0\_131/

$ export HADOOP\_CONF\_DIR=/home/hadoop/config/hadoop/

### 最好将该环境变量也加入 /etc/profile

# echo "export HADOOP\_CONF\_DIR=/home/hadoop/config/hadoop/" >> /etc/profile

**log 存储位置**

修改 **hadoop log** 存储位置。

$ echo "export HADOOP\_LOG\_DIR=~/log/hadoop" >> ~/config/hadoop/hadoop-env.sh

修改 **YARN log** 存储位置，$ vim config/hadoop/yarn-env.sh。

YARN\_LOG\_DIR="/home/hadoop/log/yarn/"

**各种配置改完要重启电脑才能生效！！！**

建立单机伪集群

由真集群到伪集群只需要修改slavs

真集群slaves有多个元素，单机伪集群slaves只有一个元素

由于我想用其他电脑访问我的虚拟机，所以以前使用的NAT模式不再合适

因为NAT虽然可以通过地址转换连接Internet但是不能实现虚拟机与非宿主机的通信

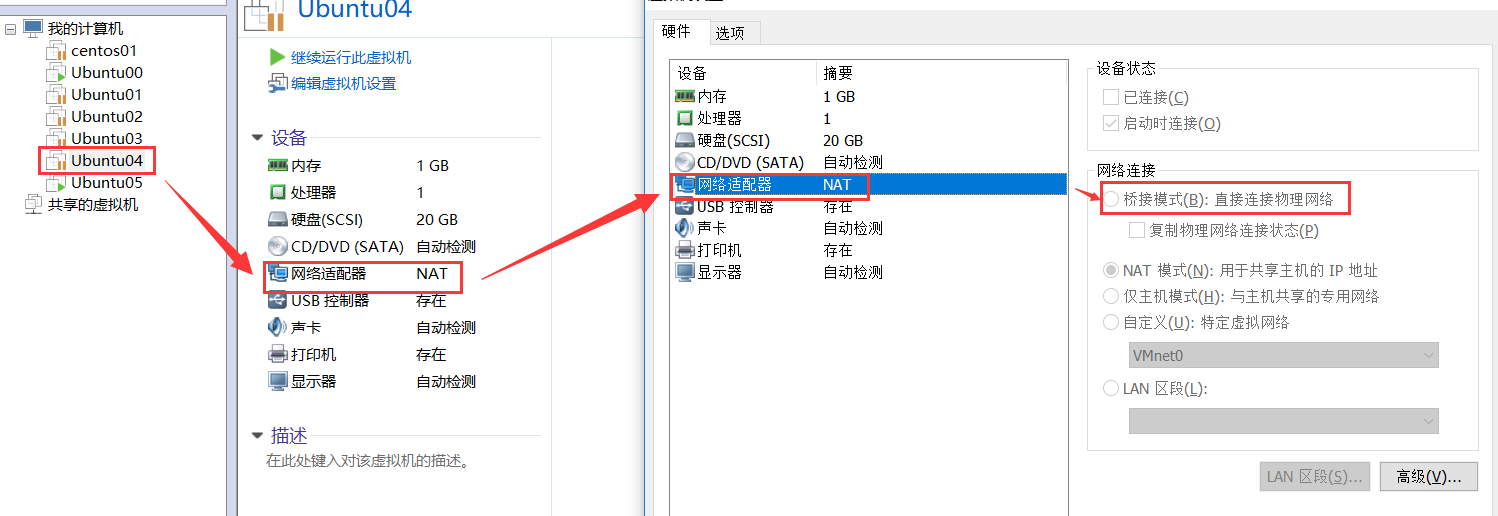
host-only：虚拟机不能连接Internet

桥接：虚拟机作为独立的机器，与宿主机处于同一网段，可以连接外网和其他实体机

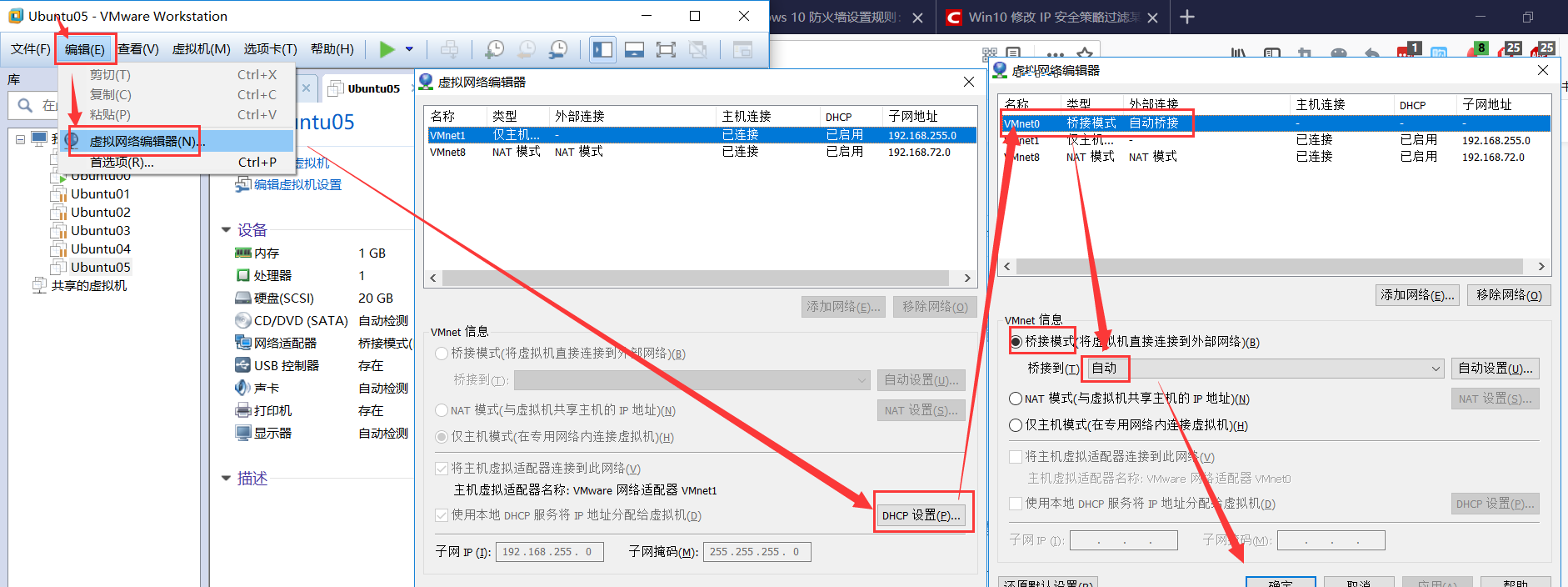
。

桥接的网络设置步骤

1、选择网络适配器：桥接模式（自动）



2、编辑虚拟网络：桥接模式



3、配置静态ip、 dns

开启虚拟机

1. 设置ip：vim 编辑interface

$ sudo vim /etc/network/interfaces

追加几行代码：

# the primary network interface

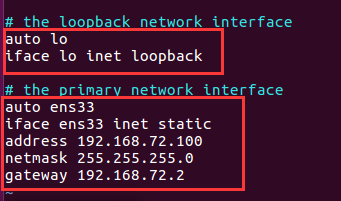
auto ens33 //默认网卡

iface ens33 inet static //静态网络

address 192.168.72.130 //ip地址

netmask 255.255.255.0 //子网掩码

gateway 192.168.72.2 //网关



1. 设置DNS：vim编辑

$ sudo vi /etc/resolvconf/resolv.conf.d/base

追加几行代码：

nameserver 8.8.8.8

nameserver 8.8.4.4

1. 重启网络服务

sudo /etc/init.d/networking restart #使网卡配置生效

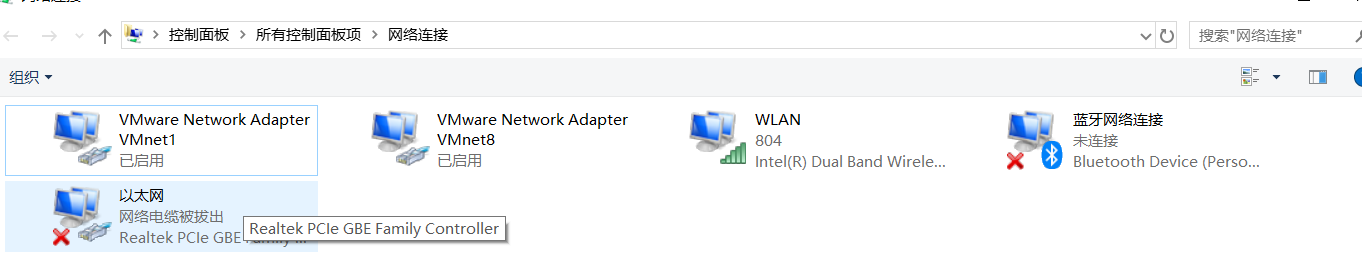
sudo /etc/init.d/resolvconf restart #使DNS生效

如果无效就重启系统

\*注意 自从ubuntu15.04网络管理工具废弃了nm-tool，改为nmcli

nmcli dev show //查看mac ip DNS

win10上没有vmnet0



桥接不是添加一个VMnet0虚拟网卡，而是添加网络服务

vmnet0，其实是你物理网卡上的VMware bridge protocol桥接协议！  
打开本地连接或其他[无线网卡](https://www.baidu.com/s?wd=%E6%97%A0%E7%BA%BF%E7%BD%91%E5%8D%A1&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)的属性，看下是否有这个协议，有的话就有[桥接模式](https://www.baidu.com/s?wd=%E6%A1%A5%E6%8E%A5%E6%A8%A1%E5%BC%8F&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)！  
没有的话，重装VMware 虚拟机软件即可，特殊版本Windows（比如[Win10](https://www.baidu.com/s?wd=Win10&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)测试版个别版本无法添加这个桥接协议，需要更新系统版本）。