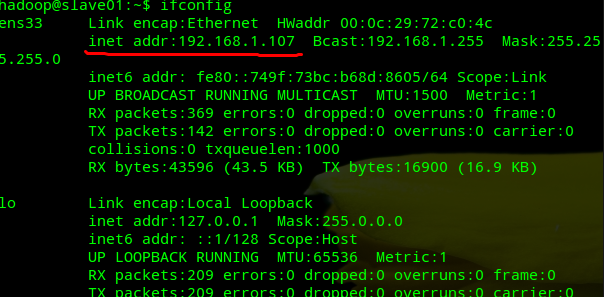
Hadoop是一个能够让用户轻松架构和使用的分布式计算平台。为了更好演示集群分布，本文没有使用一台电脑上构建多个虚拟机的方法来模拟集群，而是使用三台电脑来搭建一个小型分布式集群环境安装。本文记录如何搭建并配置Hadoop分布式集群环境。

**集群机器**

一台Ubuntu主机系统作Master,一台Ubuntu主机系统做slave01,一台Ubuntu主机系统做slave02。三台主机机器处于同一局域网下。  
这里使用三台主机搭建分布式集群环境，更多台机器同样可以使用如下配置。  
IP在不同局域网环境下有可能不同，可以用ifconfig命令查看当前主机打IP。

1. ifconfig

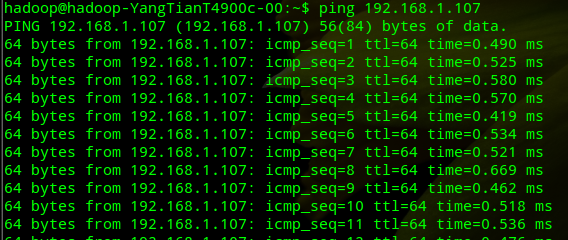
Shell 命令

即可获得当前主机的IP在局域网地址，如下图：  
  
三台机器的名称和IP如下,

| **主机名称** | **IP地址** |
| --- | --- |
| master | 192.168.1.104 |
| slave01 | 192.168.1.107 |
| slave02 | 192.168.1.108 |

三台电脑主机的用户名均为hadoop.  
三台机器可以ping双方的ip来测试三台电脑的连通性。  
在master节点主机上的Shell中运行如下命令,测试能否连接到slave01节点主机

ping 192.168.1.107

如果出现如下图，说明连接成功  
  
为了更好的在Shell中区分三台主机，修改其显示的主机名，执行如下命令

1. sudo vim /etc/hostname

Shell 命令

master的/etc/hostname添加如下配置：

master

同样slave01的/etc/hostname添加如下配置：

slave01

同样slave02的/etc/hostname添加如下配置：

slave02

重启三台电脑，重启后在终端Shell中才会看到机器名的变化,如下图：  
20161205_003  
修改三台机器的/etc/hosts文件,添加同样的配置：

1. sudo vim /etc/hosts

Shell 命令

配置如下：

127.0.0.1 localhost

192.168.1.104 master

192.168.1.107 slave01

192.168.1.108 slave02

# 配置ssh无密码登录本机和访问集群机器

三台主机电脑分别运行如下命令，测试能否连接到本地localhost

1. ssh localhost

Shell 命令

登录成功会显示如下结果：

Last login: Mon Feb 29 18:29:55 2016 from ::1

如果不能登录本地，请运行如下命令，安装openssh-server,并生成ssh公钥。

1. sudo apt-get openssh-server
2. ssh-keygen -t rsa -P ""
3. cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys

Shell 命令

在保证了三台主机电脑都能连接到本地localhost后，还需要让master主机免密码登录slave01和slave02主机。在master执行如下命令，将master的id\_rsa.pub传送给两台slave主机。

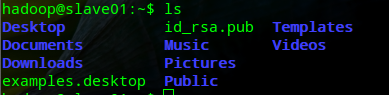
1. scp ~/.ssh/id\_rsa.pub hadoop@slave01:/home/hadoop/
2. scp ~/.ssh/id\_rsa.pub hadoop@slave02:/home/hadoop/

Shell 命令

在slave01,slave02主机上分别运行ls命令

1. ls ~

Shell 命令

可以看到slave01、slave02主机分别接收到id\_rsa.pub文件  
  
接着在slave01、slave02主机上将master的公钥加入各自的节点上,在slave01和slave02执行如下命令:

1. cat ~/id\_rsa.pub >> ~/.ssh/authorized\_keys
2. rm ~/id\_rsa.pub

Shell 命令

如果master主机和slave01、slave02主机的用户名一样，那么在master主机上直接执行如下测试命令，即可让master主机免密码登录slave01、slave02主机。

1. ssh slave01

Shell 命令

如果master主机和slave01主机的用户名不一样，还需要在master修改~/.ssh/config文件，如果没有此文件，自己创建文件。

Host master

user Ruanrc

Host slave01

user hadoop

然后master主机再执行免密码登录：

1. ssh slave01

Shell 命令

# JDK和Hadoop安装配置

分别在master主机和slave01、slave02主机上安装JDK和Hadoop，并加入环境变量。

## 安装JDK

分别在master主机和slave01,slave02主机上执行安装JDK的操作

1. sudo apt-get install default-jdk

Shell 命令

编辑~/.bashrc文件，添加如下内容：

export JAVA\_HOME=/usr/lib/jvm/default-java

接着让环境变量生效，执行如下代码：

1. source ~/.bashrc

Shell 命令

## 安装Hadoop

先在master主机上做安装Hadoop，暂时不需要在slave01,slave02主机上安装Hadoop.稍后会把master配置好的Hadoop发送给slave01,slave02.  
在master主机执行如下操作：

1. sudo tar -zxf ~/下载/hadoop-2.7.3.tar.gz -C /usr/local *# 解压到/usr/local中*
2. cd /usr/local/
3. sudo mv ./hadoop-2.7.3/ ./hadoop *# 将文件夹名改为hadoop*
4. sudo chown -R hadoop ./hadoop *# 修改文件权限*

Shell 命令

编辑~/.bashrc文件，添加如下内容：

export HADOOP\_HOME=/usr/local/hadoop

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

接着让环境变量生效，执行如下代码：

1. source ~/.bashrc

Shell 命令

# Hadoop集群配置

修改master主机修改Hadoop如下配置文件，这些配置文件都位于/usr/local/hadoop/etc/hadoop目录下。  
修改slaves：  
这里把DataNode的主机名写入该文件，每行一个。这里让master节点主机仅作为NameNode使用。

slave01

slave02

修改core-site.xml

<configuration>

<property>

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

<value>file:/usr/local/hadoop/tmp</value>

<description>Abase for other temporary directories.</description>

</property>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

修改hdfs-site.xml：

<configuration>

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>master:50090</value>

</property>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<property>

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

<value>file:/usr/local/hadoop/tmp/dfs/name</value>

</property>

<property>

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

<value>file:/usr/local/hadoop/tmp/dfs/data</value>

</property>

</configuration>

修改mapred-site.xml(复制mapred-site.xml.template,再修改文件名)

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

修改yarn-site.xml

<configuration>

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

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

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

<value>master</value>

</property>

</configuration>

配置好后，将 master 上的 /usr/local/Hadoop 文件夹复制到各个节点上。之前有跑过伪分布式模式，建议在切换到集群模式前先删除之前的临时文件。在 master 节点主机上执行：

1. cd /usr/local/
2. rm -rf ./hadoop/tmp *# 删除临时文件*
3. rm -rf ./hadoop/logs/\* *# 删除日志文件*
4. tar -zcf ~/hadoop.master.tar.gz ./hadoop
5. cd ~
6. scp ./hadoop.master.tar.gz slave01:/home/hadoop
7. scp ./hadoop.master.tar.gz slave02:/home/hadoop

Shell 命令

在slave01,slave02节点上执行：

1. sudo rm -rf /usr/local/hadoop/
2. sudo tar -zxf ~/hadoop.master.tar.gz -C /usr/local
3. sudo chown -R hadoop /usr/local/hadoop

Shell 命令

# 启动hadoop集群

在master主机上执行如下命令：

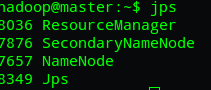
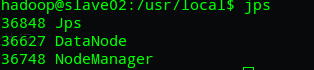
1. cd /usr/local/hadoop
2. bin/hdfs namenode -format
3. sbin/start-all.sh

Shell 命令

运行后，在master，slave01,slave02运行jps命令，查看：

1. jps

Shell 命令

master运行jps后，如下图：  
  
slave01、slave02运行jps，如下图：  
  


bin/hdfs dfsadmin -report