

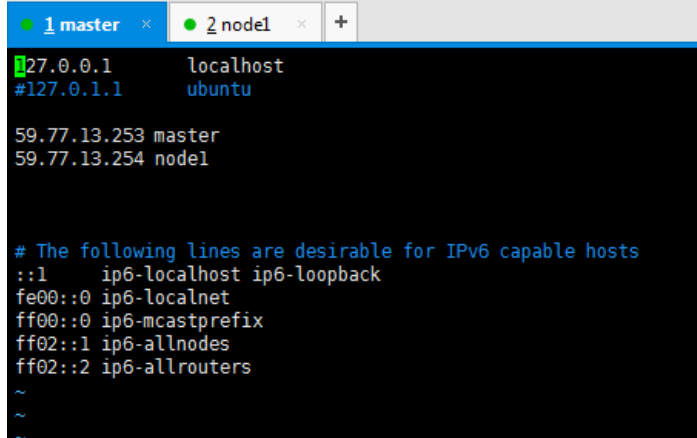
hadoop集群配置

前提条件，主节点和计算节点位于同一网段

一、修改hosts文件

修改主节点和计算节点的hosts文件

vim/etc/hosts



```
1 master x 2 node1 x +
27.0.0.1 localhost
#127.0.1.1 ubuntu

59.77.13.253 master
59.77.13.254 node1

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
~
~
~
```

二、修改hostname文件

vim/etc/hostname

主节点命名为：master 计算节点命名为node1

三、配置java环境

1.安装jdk

apt-get install openjdk-7-jdk

2.配置环境变量

vim/etc/profile

加入如下内容

export JAVA_HOME=/usr/lib/jvm/java ——>对应jdk目录

export JRE_HOME=\${JAVA_HOME}/jre

export CLASSPATH=.:\${JAVA_HOME}/lib:\${JRE_HOME}/lib

export PATH=\${JAVA_HOME}/bin:\$PATH

3.测试

java -version

四、安装ssh无密码登录

1.安装ssh

apt-get update更新apt-get的内容

apt-get install openssh-server (或 apt-get install ssh)

2.查看ssh服务是否启动

ps -e |grep ssh

```
root@master:/home/christy# ps -e |grep ssh
  997 ?        00:00:00 sshd
 7558 ?        00:00:00 sshd
 7618 ?        00:00:00 sshd
```

启动ssh服务

service ssh start

重启ssh服务

service ssh restart

3.修改配置文件"/etc/ssh/sshd_config"

1.

RSAAuthentication yes # 启用 RSA 认证

PubkeyAuthentication yes # 启用公钥私钥配对认证方式

AuthorizedKeysFile %h/.ssh/authorized_keys# 公钥文件路径

PermitRootLogin without-password"加一个"#"号,把它注释掉->

再增加一句"PermitRootLogin yes"->保存, 修改成功。

重启SSH服务: service ssh restart

2.

在主节点和计算节点都进行如下的操作:

以新建用户hadoop身份登录

su -m hadoop

ssh-keygen -t rsa 生成密钥文件, 一切按默认的方式生成, 密钥文件生成后默认存放位置为/home/hadoop(你的用户名)/.ssh

进入该目录

将id_rsa.pub加到授权的key里面去: cat id_rsa.pub >> authorized_keys

```
hadoop@master:/home/christy$ cd /home/hadoop/.ssh/
hadoop@master:~/.ssh$ ls
authorized_keys  id_rsa  id_rsa.pub  known_hosts
hadoop@master:~/.ssh$
```

修改authorized_keys的权限, (这一步很重要不然的话, SSH时仍然需要密码)

chmod 600 authorized_keys

查看本机是否可以SSH无需密码登录: ssh localhost

```
hadoop@master:~/.ssh$ ssh localhost
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.19.0-25-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

311 packages can be updated.
149 updates are security updates.

Last login: Thu Feb 25 22:17:37 2016 from master
$ exit
Connection to localhost closed.
```

3.让主结点(master)能通过SSH免密码登录两个子结点(slave)

为了实现这个功能, slave结点的公钥文件中必须要包含主结点的公钥信息, 这样

当master就可以顺利安全地访问slave结点了。操作过程如下:

有两种方法:

一种是通过 在主节点上面通过sz操作将公钥文件 xxxx.pub文件下载到本地电脑。然后通过rz -e 操作将公钥文件放到节点上

一种是在节点上通过scp操作 (scp 用户名@ip: 目标文件 目标位置+名称)

本人将拷贝过来的公钥命名为master_rsa.pub

```
hadoop@node1: ~/.ssh$ ls
authorized_keys  id_rsa  id_rsa.pub  known_hosts  master_rsa.pub
```

将公钥文件加到授权的key里面去: `cat master_rsa.pub >> authorized_keys`

4.测试

在主节点上进行如下操作:

```
hadoop@master: ~/.ssh$ ssh node1
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.19.0-25-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

302 packages can be updated.
142 updates are security updates.

Last login: Fri Feb 26 04:57:15 2016 from master
$
```

或输入ip地址

```
hadoop@master: ~/.ssh$ ssh 59.77.13.254
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.19.0-25-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

302 packages can be updated.
142 updates are security updates.

Last login: Fri Feb 26 18:40:23 2016 from master
$
```

5.对master自身进行ssh免密码登录测试工作:

表面上看,这两个结点的ssh免密码登录已经配置成功,但是我们还需要对主结点master也要进行上面的同样工作,这一步有点让人困惑,但是这是有原因的,具体原因现在也说不太好,据说是真实物理结点时需要做这项工作,因为jobtracker有可能会分布在其它结点上,jobtracker有不存在master结点上的可能性。

对master自身进行ssh免密码登录测试工作:

```
hadoop@master: ~/.ssh$ ssh master
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.19.0-25-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

311 packages can be updated.
149 updates are security updates.

Last login: Fri Feb 26 18:32:00 2016 from localhost
$
```

至此,SSH免密码登录已经配置成功。

五、下载并解压hadoop安装包

1.下载、解压

这里使用的hadoop的版本为1.2.1

在主节点下进行如下操作:

镜像地址为:

<http://mirror.bit.edu.cn/apache/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz>

通过命令

`wget http://mirror.bit.edu.cn/apache/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz`

将压缩包放入/opt/目录下

```
hadoop@master: /opt$ rm /opt/hadoop-1.2.1.tar.gz /opt/
```

然后解压该文件

`tar -zcvf hadoop-1.2.1.tar.gz`

```
hadoop@master: /opt$ ls
hadoop-1.2.1  hadoop-1.2.1.tar.gz
```

进入hadoop-1.2.1文件中

bin文件是指令文件

conf是配置文件

2.修改配置文件

在1.2.1版本，需要修改的conf文件中的一个sh文件和3个xml文件

hadoop-env.sh修改java路径

在通过echo \$JAVA_HOME的方式获取

```
# Set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.

# The java implementation to use. Required.
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64

# Extra Java CLASSPATH elements. Optional.
# export HADOOP_CLASSPATH=

# The maximum amount of heap to use, in MB. Default is 1000.
# export HADOOP_HEAPSIZE=2000

# Extra Java runtime options. Empty by default.
```

core-site.xml 其中tmp与name文件必须手动在该目录下建立。 hadoop为新建的用户对应的文件夹

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/home/hadoop/tmp</value>
  </property>
  <property>
    <name>dfs.name.dir</name>
    <value>/home/hadoop/name</value>
  </property>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://59.77.13.253:9000</value>
  </property>
</configuration>
~
```

```
<configuration>
<property>
<name>hadoop.tmp.dir</name>
<value>/home/hadoop/tmp</value>
</property>
<property>
<name>dfs.name.dir</name>
<value>/home/hadoop/name</value>
</property>
<property>
<name>fs.default.name</name>
<value>hdfs://59.77.13.253:9000</value>
</property>
</configuration>
```

mapred-site.xml 其中ip填写的是主节点的ip地址

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>59.77.13.253:9002</value>
  </property>
</configuration>
~
~
```

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>59.77.13.253:9002</value>
</property>
</configuration>
```

hdfs-site.xml 这边填写也是主节点的ip, data文件必须手动建

```
<configuration>
<property>
  <name>dfs.http.address</name>
  <value>59.77.13.253:9001</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hadoop/data</value>
</property>
</configuration>
~
~
```

```
<configuration>
<property>
  <name>dfs.http.address</name>
  <value>59.77.13.253:9001</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hadoop/data</value>
</property>
</configuration>
```

3.在计算节点将主节点中的/opt/hadoop-1.2.1整个文件夹复制到自身的/opt/hadoop-1.2.1
scp

4.环境变量配置/etc/profile

在主节点和计算节点进行如下操作

vim /etc/profile

```
# and Bourne compatible shells (dash, bash, ksh, zsh, ...).
export HADOOP_HOME=/opt/hadoop-1.2.1
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
export JRE_HOME=${JAVA_HOME}/jre
export CLASSPATH=.:${JAVA_HOME}/lib:${JRE_HOME}/lib
export PATH=${JAVA_HOME}/bin:${JRE_HOME}/bin:${HADOOP_HOME}/bin:$PATH

if [ "$PS1" ]; then
```

source /etc/profile

5.测试

输入hadoop, 出现下图说明环境配置成功

```

hadoop@master:/opt/hadoop-1.2.1/conf$ hadoop
Warning: $HADOOP_HOME is deprecated.

Usage: hadoop [--config confdir] COMMAND
where COMMAND is one of:
  namenode -format      format the DFS filesystem
  secondarynamenode    run the DFS secondary namenode
  namenode              run the DFS namenode
  datanode              run a DFS datanode
  dfsadmin             run a DFS admin client
  mradmin              run a Map-Reduce admin client
  fsck                 run a DFS filesystem checking utility
  fs                   run a generic filesystem user client
  balancer             run a cluster balancing utility
  oiv                  apply the offline fsimage viewer to an fsimage
  fetchdt              fetch a delegation token from the NameNode
  jobtracker           run the MapReduce job Tracker node
  pipes                run a Pipes job
  tasktracker          run a MapReduce task Tracker node
  historyserver        run job history servers as a standalone daemon
  job                  manipulate MapReduce jobs
  queue               get information regarding JobQueues
  version              print the version
  jar <jar>            run a jar file
  distcp <srcurl> <desturl> copy file or directories recursively
  distcp2 <srcurl> <desturl> DistCp version 2
  archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
  classpath            prints the class path needed to get the
                      Hadoop jar and the required libraries
  daemonlog            get/set the log level for each daemon
or
  CLASSNAME            run the class named CLASSNAME
Most commands print help when invoked w/o parameters.

```

6.初始化节点

在主节点上进行如下操作

```

hadoop@master:/opt/hadoop-1.2.1/conf$ hadoop namenode -format
zhm@master:~$ hadoop namenode -format
12/12/02 22:44:08 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = master/192.168.1.100
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 0.20.2
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branch
es/branch-0.20 -r 911707; compiled by 'chrisdo' on Fri Feb 19 08:07:34 UTC 2
010
*****/
Re-format filesystem in /home/zhm/hadoop/name ? (Y or N) Y
12/12/02 22:44:13 INFO namenode.FSNamesystem: fsOwner=zhm,zhm,adm,dialout,cd
rom,plugdev,lpadmin,admin,sambashare
12/12/02 22:44:13 INFO namenode.FSNamesystem: supergroup=supergroup
12/12/02 22:44:13 INFO namenode.FSNamesystem: isPermissionEnabled=true
12/12/02 22:44:13 INFO common.Storage: Image file of size 93 saved in 0 seco
nds.
12/12/02 22:44:13 INFO common.Storage: Storage directory /home/zhm/hadoop/na
me has been successfully formatted.
12/12/02 22:44:13 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at master/192.168.1.100
*****/

```

7.启动hadoop

在主节点上进行如下操作：

进入bin文件

```

hadoop@master:/opt/hadoop-1.2.1/conf$ cd /opt/hadoop-1.2.1/bin
hadoop@master:/opt/hadoop-1.2.1/bin$ ls
hadoop          hadoop-daemon.sh  rcc      start-all.sh  start-dfs.sh  start-mapred.sh  stop-balancer.sh  stop-jobhistoryserver.sh  task-controller
hadoop-config.sh hadoop-daemons.sh slaves.sh start-balancer.sh start-jobhistoryserver.sh stop-all.sh  stop-dfs.sh  stop-mapred.sh
hadoop@master:/opt/hadoop-1.2.1/bin$

```

通过start-all.sh开启hadoop服务。通过stop-all.sh关闭hadoop服务

8.测试

在主节点上输入jps

```

hadoop@master:/opt/hadoop-1.2.1/bin$ jps
6873 SecondaryNameNode
6673 NameNode
6959 JobTracker
8314 Jps

```

在计算节点上输入jps

```
hadoop@node1:~/.ssh$ jps
3524 TaskTracker
8795 Jps
3398 DataNode
hadoop@node1:~/.ssh$
```

ok已经成功了