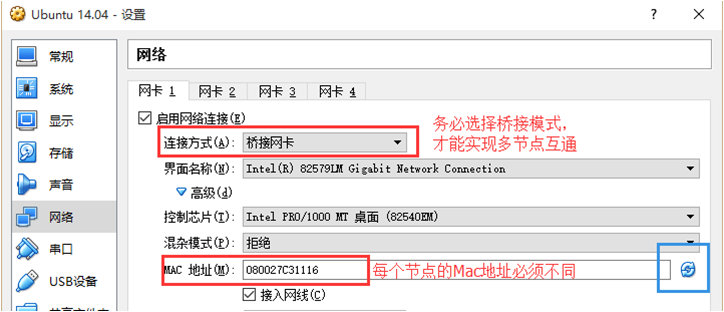
当Hadoop采用分布式模式部署和运行时，存储采用分布式文件系统HDFS，而且，HDFS的名称节点和数据节点位于不同机器上。这时，数据就可以分布到多个节点上，不同数据节点上的数据计算可以并行执行，这时的MapReduce分布式计算能力才能真正发挥作用。  
为了降低分布式模式部署难度，本教程简单使用两个节点（两台物理机器）来搭建集群环境，一台机器作为 Master节点，局域网IP地址为192.168.1.121，另一台机器作为 Slave 节点，局域网 IP 地址为192.168.1.122。由三个以上节点构成的集群，也可以采用类似的方法完成安装部署。  
Hadoop 集群的安装配置大致包括以下步骤：  
（1）步骤1：选定一台机器作为 Master；  
（2）步骤2：在Master节点上创建hadoop用户、安装SSH服务端、安装Java环境；  
（3）步骤3：在Master节点上安装Hadoop，并完成配置；  
（4）步骤4：在其他Slave节点上创建hadoop用户、安装SSH服务端、安装Java环境；  
（5）步骤5：将Master节点上的“/usr/local/hadoop”目录复制到其他Slave节点上；  
（6）步骤6：在Master节点上开启Hadoop；  
上述这些步骤中，关于如何创建hadoop用户、安装SSH服务端、安装Java环境、安装Hadoop等过程，已经在前面介绍伪分布式安装的时候做了详细介绍（查看[Hadoop安装教程\_单机/伪分布式配置教程](http://dblab.xmu.edu.cn/blog/2441-2/)），请按照之前介绍的伪分布式安装方法完成步骤1到步骤4，这里不再赘述。在完成步骤1到步骤4的操作以后，才可以继续进行下面的操作。

**1. 网络配置**

假设集群所用的两个节点（机器）都位于同一个局域网内。如果两个节点使用的是虚拟机安装的Linux系统，那么两者都需要更改网络连接方式为“桥接网卡”模式，才能实现多个节点互连，如下图所示。此外，一定要确保各个节点的Mac地址不能相同，否则会出现 IP冲突。如果是采用导入虚拟机镜像文件的方式安装Linux系统，则有可能出现两台机器的MAC地址是相同的，因为一台机器复制了另一台机器的配置，因此，需要改变机器的MAC地址，如下图所示，可以点击界面右边的“刷新”按钮随机生成 MAC 地址，这样就可以让两台机器的MAC地址不同了。  


网络配置完成以后，可以查看一下机器的IP地址，可以使用ifconfig命令查看。本教程在同一个局域网内部的两台机器的IP地址分别是192.168.1.121和192.168.1.122。  
由于集群中有两台机器需要设置，所以，在接下来的操作中，一定要注意区分Master节点和Slave节点。为了便于区分Master节点和Slave节点，可以修改各个节点的主机名，这样，在Linux系统中打开一个终端以后，在终端窗口的标题和命令行中都可以看到主机名，就比较容易区分当前是对哪台机器进行操作。在Ubuntu中，我们在 Master 节点上执行如下命令修改主机名：

1. sudo vim /etc/hostname

Shell 命令

执行上面命令后，就打开了“/etc/hostname”这个文件，这个文件里面记录了主机名，比如，假设安装Ubuntu系统时，设置的主机名是“dblab-VirtualBox”，因此，打开这个文件以后，里面就只有“dblab-VirtualBox”这一行内容，可以直接删除，并修改为“Master”（注意是区分大小写的），然后，保存退出vim编辑器，这样就完成了主机名的修改，需要重启Linux系统才能看到主机名的变化。  
要注意观察主机名修改前后的变化。在修改主机名之前，如果用hadoop登录Linux系统，打开终端，进入Shell命令提示符状态，会显示如下内容：

hadoop@ dblab-VirtualBox:~$

修改主机名并且重启系统之后，用hadoop登录Linux系统，打开终端，进入Shell命令提示符状态，会显示如下内容：

hadoop@ Master:~$

可以看出，这时就很容易辨认出当前是处于Master节点上进行操作，不会和Slave节点产生混淆。  
然后，在Master节点中执行如下命令打开并修改Master节点中的“/etc/hosts”文件：

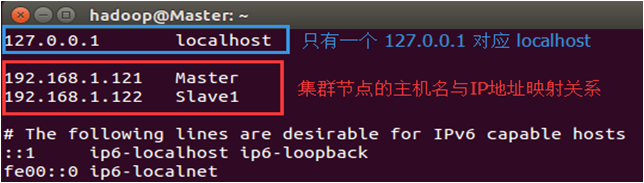
1. sudo vim /etc/hosts

Shell 命令

可以在hosts文件中增加如下两条IP和主机名映射关系：

192.168.1.121 Master

192.168.1.122 Slave1

修改后的效果如下图所示。  


需要注意的是，一般hosts文件中只能有一个127.0.0.1，其对应主机名为localhost，如果有多余127.0.0.1映射，应删除，特别是不能存在“127.0.0.1 Master”这样的映射记录。修改后需要重启Linux系统。  
上面完成了Master节点的配置，接下来要继续完成对其他Slave节点的配置修改。本教程只有一个Slave节点，主机名为Slave1。请参照上面的方法，把Slave节点上的“/etc/hostname”文件中的主机名修改为“Slave1”，同时，修改“/etc/hosts”的内容，在hosts文件中增加如下两条IP和主机名映射关系：

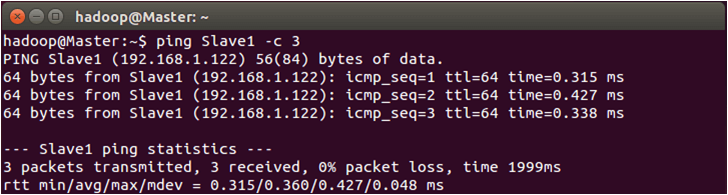
192.168.1.121 Master

192.168.1.122 Slave1

修改完成以后，请重新启动Slave节点的Linux系统。  
这样就完成了Master节点和Slave节点的配置，然后，需要在各个节点上都执行如下命令，测试是否相互ping得通，如果ping不通，后面就无法顺利配置成功：

1. ping Master -c 3 *# 只ping 3次就会停止，否则要按Ctrl+c中断ping命令*
2. ping Slave1 -c 3

Shell 命令

例如，在Master节点上ping Slave1，如果ping通的话，会显示如下图所示的结果。  


**2. SSH无密码登录节点**

必须要让Master节点可以SSH无密码登录到各个Slave节点上。首先，生成Master节点的公匙，如果之前已经生成过公钥，必须要删除原来生成的公钥，重新生成一次，因为前面我们对主机名进行了修改。具体命令如下：

1. cd ~/.ssh *# 如果没有该目录，先执行一次ssh localhost*
2. rm ./id\_rsa\* *# 删除之前生成的公匙（如果已经存在）*
3. ssh-keygen -t rsa *# 执行该命令后，遇到提示信息，一直按回车就可以*

Shell 命令

为了让Master节点能够无密码SSH登录本机，需要在Master节点上执行如下命令：

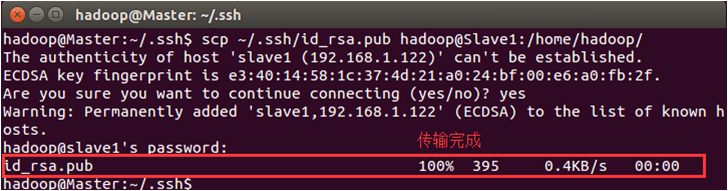
1. cat ./id\_rsa.pub >> ./authorized\_keys

Shell 命令

完成后可以执行命令“ssh Master”来验证一下，可能会遇到提示信息，只要输入yes即可，测试成功后，请执行“exit”命令返回原来的终端。  
接下来，在Master节点将上公匙传输到Slave1节点：

1. scp ~/.ssh/id\_rsa.pub hadoop@Slave1:/home/hadoop/

Shell 命令

上面的命令中，scp是secure copy的简写，用于在 Linux下进行远程拷贝文件，类似于cp命令，不过，cp只能在本机中拷贝。执行scp时会要求输入Slave1上hadoop用户的密码，输入完成后会提示传输完毕，如下图所示。  


接着在Slave1节点上，将SSH公匙加入授权：

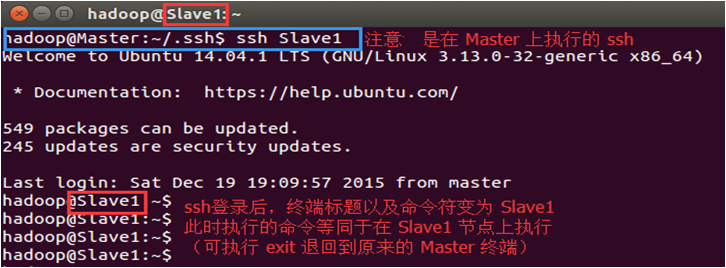
1. mkdir ~/.ssh *# 如果不存在该文件夹需先创建，若已存在，则忽略本命令*
2. cat ~/id\_rsa.pub >> ~/.ssh/authorized\_keys
3. rm ~/id\_rsa.pub *# 用完以后就可以删掉*

Shell 命令

如果有其他Slave节点，也要执行将Master公匙传输到Slave节点以及在Slave节点上加入授权这两步操作。  
这样，在Master节点上就可以无密码SSH登录到各个Slave节点了，可在Master节点上执行如下命令进行检验：

1. ssh Slave1

Shell 命令

执行该命令的效果如下图所示。  


**3. 配置PATH变量**

在前面的伪分布式安装内容中，已经介绍过PATH变量的配置方法。可以按照同样的方法进行配置，这样就可以在任意目录中直接使用hadoop、hdfs等命令了。如果还没有配置PATH变量，那么需要在Master节点上进行配置。 首先执行命令“vim ~/.bashrc”，也就是使用vim编辑器打开“~/.bashrc”文件，然后，在该文件最上面的位置加入下面一行内容：

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

保存后执行命令“source ~/.bashrc”，使配置生效。

**4. 配置集群/分布式环境**

在配置集群/分布式模式时，需要修改“/usr/local/hadoop/etc/hadoop”目录下的配置文件，这里仅设置正常启动所必须的设置项，包括workers 、core-site.xml、hdfs-site.xml、mapred-site.xml、yarn-site.xml共5个文件，更多设置项可查看官方说明。

**（1）修改文件workers**

需要把所有数据节点的主机名写入该文件，每行一个，默认为 localhost（即把本机作为数据节点），所以，在伪分布式配置时，就采用了这种默认的配置，使得节点既作为名称节点也作为数据节点。在进行分布式配置时，可以保留localhost，让Master节点同时充当名称节点和数据节点，或者也可以删掉localhost这行，让Master节点仅作为名称节点使用。  
本教程让Master节点仅作为名称节点使用，因此将workers文件中原来的localhost删除，只添加如下一行内容：

Slave1

**（2）修改文件core-site.xml**

请把core-site.xml文件修改为如下内容：

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://Master:9000</value>

</property>

<property>

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

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

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

</property>

</configuration>

各个配置项的含义可以参考前面伪分布式模式时的介绍，这里不再赘述。

**（3）修改文件hdfs-site.xml**

对于Hadoop的分布式文件系统HDFS而言，一般都是采用冗余存储，冗余因子通常为3，也就是说，一份数据保存三份副本。但是，本教程只有一个Slave节点作为数据节点，即集群中只有一个数据节点，数据只能保存一份，所以 ，dfs.replication的值还是设置为 1。hdfs-site.xml具体内容如下：

<configuration>

<property>

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

<value>Master:50090</value>

</property>

<property>

<name>dfs.replication</name>

<value>1</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>

**（4）修改文件mapred-site.xml**

“/usr/local/hadoop/etc/hadoop”目录下有一个mapred-site.xml.template，需要修改文件名称，把它重命名为mapred-site.xml，然后，把mapred-site.xml文件配置成如下内容：

<configuration>

<property>

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

<value>yarn</value>

</property>

<property>

<name>mapreduce.jobhistory.address</name>

<value>Master:10020</value>

</property>

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>Master:19888</value>

</property>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP\_MAPRED\_HOME=/usr/local/hadoop</value>

</property>

<property>

<name>mapreduce.map.env</name>

<value>HADOOP\_MAPRED\_HOME=/usr/local/hadoop</value>

</property>

<property>

<name>mapreduce.reduce.env</name>

<value>HADOOP\_MAPRED\_HOME=/usr/local/hadoop</value>

</property>

</configuration>

**（5）修改文件 yarn-site.xml**

请把yarn-site.xml文件配置成如下内容：

<configuration>

<property>

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

<value>Master</value>

</property>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>

上述5个文件全部配置完成以后，需要把Master节点上的“/usr/local/hadoop”文件夹复制到各个节点上。如果之前已经运行过伪分布式模式，建议在切换到集群模式之前首先删除之前在伪分布式模式下生成的临时文件。具体来说，需要首先在Master节点上执行如下命令：

1. cd /usr/local
2. sudo rm -r ./hadoop/tmp *# 删除 Hadoop 临时文件*
3. sudo rm -r ./hadoop/logs/\* *# 删除日志文件*
4. tar -zcf ~/hadoop.master.tar.gz ./hadoop *# 先压缩再复制*
5. cd ~
6. scp ./hadoop.master.tar.gz Slave1:/home/hadoop

Shell 命令

然后在Slave1节点上执行如下命令：

1. sudo rm -r /usr/local/hadoop *# 删掉旧的（如果存在）*
2. sudo tar -zxf ~/hadoop.master.tar.gz -C /usr/local
3. sudo chown -R hadoop /usr/local/hadoop

Shell 命令

同样，如果有其他Slave节点，也要执行将hadoop.master.tar.gz传输到Slave节点以及在Slave节点解压文件的操作。  
首次启动Hadoop集群时，需要先在Master节点执行名称节点的格式化（只需要执行这一次，后面再启动Hadoop时，不要再次格式化名称节点），命令如下：

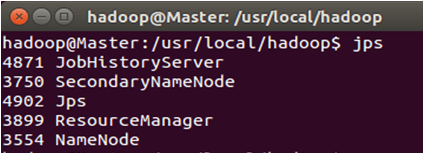
1. hdfs namenode -format

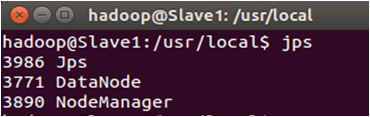
Shell 命令

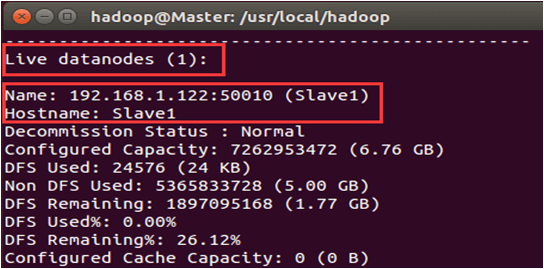
现在就可以启动Hadoop了，启动需要在Master节点上进行，执行如下命令：

1. start-dfs.sh
2. start-yarn.sh
3. mr-jobhistory-daemon.sh start historyserver

Shell 命令

通过命令jps可以查看各个节点所启动的进程。如果已经正确启动，则在Master节点上可以看到NameNode、ResourceManager、SecondrryNameNode和JobHistoryServer进程，如下图所示。  


在Slave节点可以看到DataNode和NodeManager进程，如下图所示。  


缺少任一进程都表示出错。另外还需要在Master节点上通过命令“hdfs dfsadmin -report”查看数据节点是否正常启动，如果屏幕信息中的“Live datanodes”不为 0 ，则说明集群启动成功。由于本教程只有1个Slave节点充当数据节点，因此，数据节点启动成功以后，会显示如下图所示信息。  


也可以在Linux系统的浏览器中输入地址“http://master:9870/”，通过 Web 页面看到查看名称节点和数据节点的状态。如果不成功，可以通过启动日志排查原因。  
这里再次强调，伪分布式模式和分布式模式切换时需要注意以下事项：  
（a）从分布式切换到伪分布式时，不要忘记修改slaves配置文件；  
（b）在两者之间切换时，若遇到无法正常启动的情况，可以删除所涉及节点的临时文件夹，这样虽然之前的数据会被删掉，但能保证集群正确启动。所以，如果集群以前能启动，但后来启动不了，特别是数据节点无法启动，不妨试着删除所有节点（包括Slave节点）上的“/usr/local/hadoop/tmp”文件夹，再重新执行一次“hdfs namenode -format”，再次启动即可。

**5. 执行分布式实例**

执行分布式实例过程与伪分布式模式一样，首先创建HDFS上的用户目录，命令如下：

1. hdfs dfs -mkdir -p /user/hadoop

Shell 命令

然后，在HDFS中创建一个input目录，并把“/usr/local/hadoop/etc/hadoop”目录中的配置文件作为输入文件复制到input目录中，命令如下：

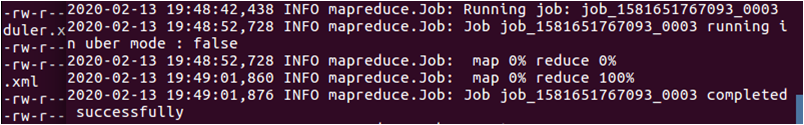
1. hdfs dfs -mkdir input
2. hdfs dfs -put /usr/local/hadoop/etc/hadoop/\*.xml input

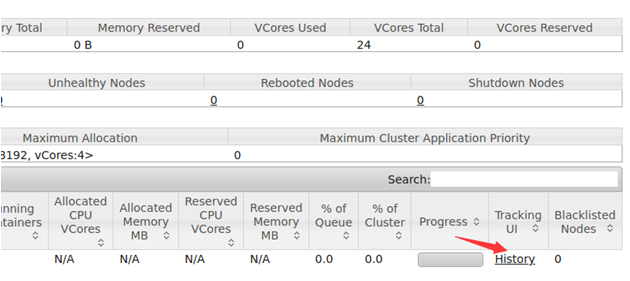
Shell 命令

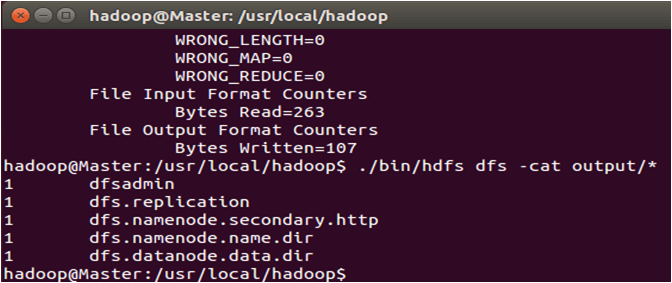
接着就可以运行 MapReduce 作业了，命令如下：

1. hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.3.jar grep input output 'dfs[a-z.]+'

Shell 命令

运行时的输出信息与伪分布式类似，会显示MapReduce作业的进度，如下图所示。  


执行过程可能会有点慢，但是，如果迟迟没有进度，比如5分钟都没看到进度变化，那么不妨重启Hadoop再次测试。若重启还不行，则很有可能是内存不足引起，建议增大虚拟机的内存，或者通过更改YARN的内存配置来解决。  
在执行过程中，可以在Linux系统中打开浏览器，在地址栏输入“http://master:8088/cluster”，通过Web界面查看任务进度，在Web界面点击 “Tracking UI” 这一列的History连接，可以看到任务的运行信息，如下图所示。  


执行完毕后的输出结果如下图所示。  


最后，关闭Hadoop集群，需要在Master节点执行如下命令：

1. stop-yarn.sh
2. stop-dfs.sh
3. mr-jobhistory-daemon.sh stop historyserver

Shell 命令

至此，就顺利完成了Hadoop集群搭建。