

# 给力星 (<http://www.powerxing.com/>)

追逐内心的平和

首页 (<http://www.powerxing.com/>)

■ 笔记 (<http://www.powerxing.com/notes/>)

搜藏 (<http://www.powerxing.com/articles/>)

代码 (<http://www.powerxing.com/snippets/>)

音乐 (<http://www.powerxing.com/music/>)

关于 (<http://www.powerxing.com/about/>)

搜索

## Hadoop集群安装配置教程 \_Hadoop2.6.0\_Ubuntu/CentOS

📅 2014-08-09 (updated: 2016-01-06) 👁 32303



厦门大学林子雨编著《大数据技术原理与应用》  
人民邮电出版社出版发行 21世纪高等教育计算机规划教材  
中国高校第一本系统介绍大数据知识专业教材  
京东、当当网等各大网店畅销书籍  
最完备的大数据课程公共服务体系

➡ 点击进入  
教材官网

(<http://dblab.xmu.edu.cn/post/bigdata/>)

本教程讲述如何配置 Hadoop 集群，默认读者已经掌握了 Hadoop 的单机伪分布式配置，否则请先查看Hadoop安装教程\_单机/伪分布式配置 (<http://www.powerxing.com/install-hadoop/>) 或 CentOS安装Hadoop\_单机/伪分布式配置 (<http://www.powerxing.com/install-hadoop-in-centos/>)。

本教程由厦门大学数据库实验室 (<http://dblab.xmu.edu.cn>)出品，转载请注明。本教程适合于原生 Hadoop 2，包括 Hadoop 2.6.0, Hadoop 2.7.1 等版本，主要参考了官方安装教程 (<http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ClusterSetup.html>)，步骤详细，辅以适当说明，保证按照步骤来，都能顺利安装并运行 **Hadoop**。另外有Hadoop安装配置简略版 (<http://www.powerxing.com/install-hadoop-simplify/>)方便有基础的读者快速完成安装。

为了方便新手入门，我们准备了两篇不同系统的 **Hadoop** 伪分布式配置教程。但其他 **Hadoop** 教程我们将不再区分，可同时适用于 **Ubuntu** 和 **CentOS/RedHat** 系统。例如本教程以 **Ubuntu** 系统为主要演示环境，但对 **Ubuntu/CentOS** 的不同配置之处、**CentOS 6.x** 与 **CentOS 7** 的操作区别等都会尽量给出注明。

## 环境

本教程使用 **Ubuntu 14.04 64位** 作为系统环境，基于原生 Hadoop 2，在 **Hadoop 2.6.0 (stable)** 版本下验证通过，可适合任何 Hadoop 2.x.y 版本，例如 Hadoop 2.7.1，Hadoop 2.4.1 等。

本教程简单的使用两个节点作为集群环境: 一个作为 Master 节点，局域网 IP 为 192.168.1.121；另一个作为 Slave 节点，局域网 IP 为 192.168.1.122。

## 准备工作

Hadoop 集群的安装配置大致为如下流程:

1. 选定一台机器作为 Master
2. 在 Master 节点上配置 hadoop 用户、安装 SSH server、安装 Java 环境
3. 在 Master 节点上安装 Hadoop，并完成配置
4. 在其他 Slave 节点上配置 hadoop 用户、安装 SSH server、安装 Java 环境
5. 将 Master 节点上的 /usr/local/hadoop 目录复制到其他 Slave 节点上
6. 在 Master 节点上开启 Hadoop

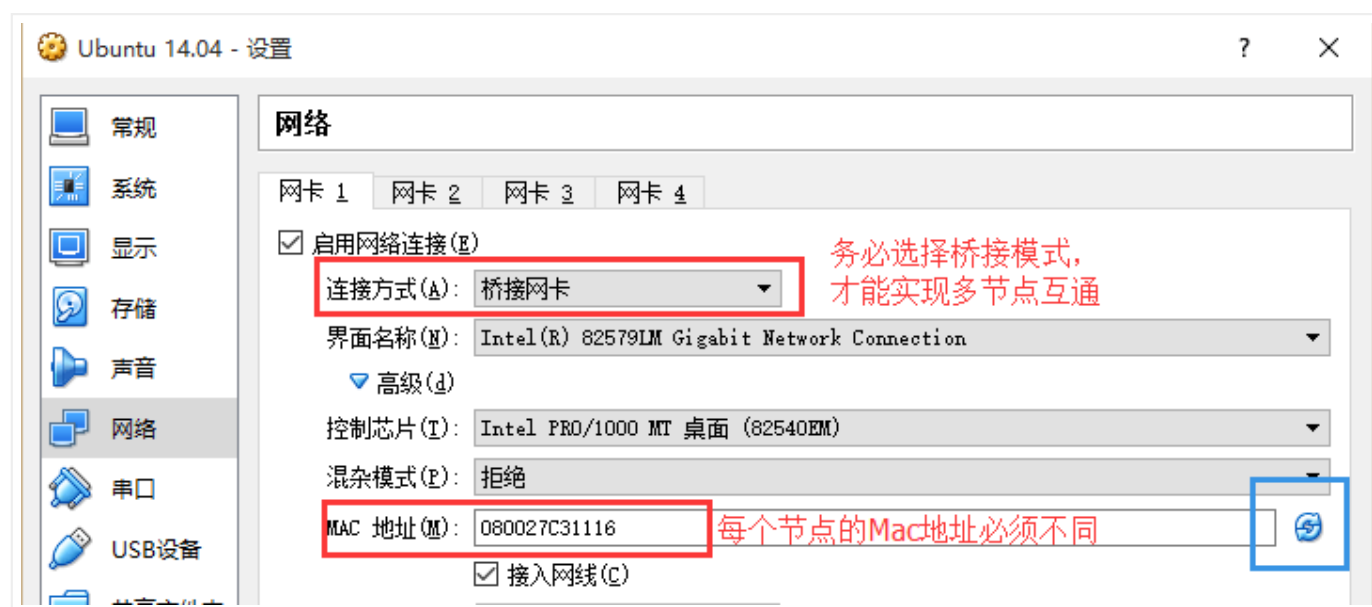
配置 hadoop 用户、安装 SSH server、安装 Java 环境、安装 Hadoop 等过程已经在 Hadoop安装教程\_单机/伪分布式配置 (<http://www.powerxing.com/install-hadoop/>) 或 CentOS安装Hadoop\_单机/伪分布式配置 (<http://www.powerxing.com/install-hadoop-in-centos/>)中有详细介绍，请前往查看，不再重复叙述。

继续下一步配置前，请先完成上述流程的前 4 个步骤。

## 网络配置

假设集群所用的节点都位于同一个局域网。

如果使用的是虚拟机安装的系统，那么需要更改网络连接方式为桥接（Bridge）模式，才能实现多个节点互连，例如在 VirtualBox 中的设置如下图。此外，如果节点的系统是在虚拟机中直接复制的，要确保各个节点的 Mac 地址不同（可以点右边的按钮随机生成 MAC 地址，否则 IP 会冲突）：



VirtualBox中节点的网络设置

Linux 中查看节点 IP 地址的命令为 `ifconfig`，即下图所示的 inet 地址（注意虚拟机安装的 **CentoS** 不会自动联网，需要点右上角连上网络才能看到 **IP** 地址）：

```
hadoop@DBLab-XMU: ~  
hadoop@DBLab-XMU:~$ ifconfig  
eth0      link encap:以太网  硬件地址 08:00:27:ea:52:a5  
          inet 地址:192.168.1.107  广播:192.168.1.255 掩码:255.255.255.0  
          inet6 地址: fe80::a00:27ff:feea:52a5/64 Scope:Link  
          UP BROADCAST RUNNING MULTICAST  MTU:1500  跃点数:1  
          接收数据包:3960 错误:0 丢弃:0 过载:0 帧数:0  
          发送数据包:1739 错误:0 丢弃:0 过载:0 载波:0  
          碰撞:0 发送队列长度:1000  
          接收字节:4930262 (4.9 MB)  发送字节:151797 (151.7 KB)
```

### Linux查看IP命令

首先在 Master 节点上完成准备工作，并关闭 Hadoop ( /usr/local/hadoop/sbin/stop-dfs.sh )，再进行后续集群配置。

为了便于区分，可以修改各个节点的主机名（在终端标题、命令行中可以看到主机名，以便区分）。在 Ubuntu/CentOS 7 中，我们在 Master 节点上执行如下命令修改主机名（即改为 Master，注意是区分大小写的）：

```
$ sudo vim /etc/hostname
```

如果是用 CentOS 6.x 系统，则是修改 /etc/sysconfig/network 文件，改为 HOSTNAME=Master，如下图所示：

```
hadoop@dblab:~  
文件(E) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)  
NETWORKING=yes  
HOSTNAME=Master  
CentOS 6.x 的主机名配置文件与 Ubuntu 有所不同  
但 CentOS 7 的主机名配置文件也是 /etc/hostname
```

### CentOS中hostname设置

然后执行如下命令修改自己所用节点的IP映射：

```
$ sudo vim /etc/hosts
```

例如本教程使用两个节点的名称与对应的 IP 关系如下：

```
192.168.1.121    Master  
192.168.1.122    Slave1
```

我们在 `/etc/hosts` 中将该映射关系填写上去即可，如下图所示（一般该文件中只有一个 `127.0.0.1`，其对应名为 `localhost`，如果有多余的应删除，特别是不能有“`127.0.0.1 Master`”这样的记录）：



```
hadoop@Master: ~  
127.0.0.1    localhost  
192.168.1.121 Master  
192.168.1.122 Slave1  
  
# The following lines are desirable for IPv6 capable hosts  
::1        ip6-localhost ip6-loopback  
fe00::0    ip6-localnet  
ff00::0    ip6-mcastfix
```

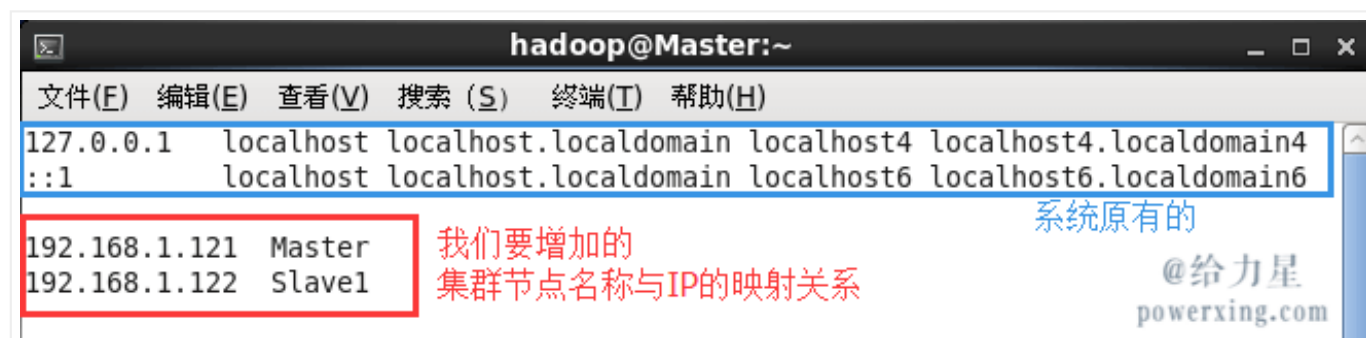
只有一个 127.0.0.1 对应 localhost

集群节点的主机名与IP地址映射关系

@给力星  
powerxing.com

Hadoop中的hosts设置

CentOS 中的 `/etc/hosts` 配置则如下图所示：



```
hadoop@Master:~  
文件(E) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)  
127.0.0.1    localhost localhost.localdomain localhost4 localhost4.localdomain4  
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6  
192.168.1.121 Master  
192.168.1.122 Slave1
```

系统原有的

我们要增加的  
集群节点名称与IP的映射关系

@给力星  
powerxing.com

CentOS中的hosts设置

修改完成后需要重启一下，重启后在终端中才会看到机器名的变化。接下来的教程中请注意区分 Master 节点与 Slave 节点的操作。

## 需要在所有节点上完成网络配置

如上面讲的是 Master 节点的配置，而在其他的 Slave 节点上，也要对 `/etc/hostname`（修改为 `Slave1`、`Slave2` 等）和 `/etc/hosts`（跟 Master 的配置一样）这两个文件进行修改！

配置好后需要在各个节点上执行如下命令，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：

```
$ ping Master -c 3    # 只ping 3次，否则要按 Ctrl+c 中断  
$ ping Slave1 -c 3
```

例如我在 Master 节点上 ping Slave1，ping 通的话会显示 time，显示的结果如下图所示：

```
hadoop@Master: ~  
hadoop@Master:~$ ping Slave1 -c 3  
PING Slave1 (192.168.1.122) 56(84) bytes of data.  
64 bytes from Slave1 (192.168.1.122): icmp_seq=1 ttl=64 time=0.315 ms  
64 bytes from Slave1 (192.168.1.122): icmp_seq=2 ttl=64 time=0.427 ms  
64 bytes from Slave1 (192.168.1.122): icmp_seq=3 ttl=64 time=0.338 ms  
  
--- Slave1 ping statistics ---  
3 packets transmitted, 3 received, 0% packet loss, time 1999ms  
rtt min/avg/max/mdev = 0.315/0.360/0.427/0.048 ms  
hadoop@Master:~$
```

检查是否ping得通

继续下一步配置前，请先完成所有节点的网络配置，修改过主机名的话需重启才能生效。

## SSH无密码登陆节点

这个操作是要让 Master 节点可以无密码 SSH 登陆到各个 Slave 节点上。

首先生成 Master 节点的公匙，在 Master 节点的终端中执行（因为改过主机名，所以还需要删掉原有的再重新生成一次）：

```
$ cd ~/.ssh          # 如果没有该目录，先执行一次ssh localhost  
$ rm ./id_rsa*       # 删除之前生成的公匙（如果有）  
$ ssh-keygen -t rsa  # 一直按回车就可以
```

让 Master 节点需能无密码 SSH 本机，在 Master 节点上执行：

```
$ cat ./id_rsa.pub >> ./authorized_keys
```

完成后可执行 ssh Master 验证一下（可能需要输入 yes，成功后执行 exit 返回原来的终端）。接着在 Master 节点将上公匙传输到 Slave1 节点：

```
$ scp ~/.ssh/id_rsa.pub hadoop@Slave1:/home/hadoop/
```



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(hadoop)，输入完成后会提示传输完毕，如下图所示：

```
hadoop@Master: ~/.ssh
hadoop@Master:~/.ssh$ scp ~/.ssh/id_rsa.pub hadoop@Slave1:/home/hadoop/
The authenticity of host 'slave1 (192.168.1.122)' can't be established.
ECDSA key fingerprint is e3:40:14:58:1c:37:4d:21:a0:24:bf:00:e6:a0:fb:2f.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'slave1,192.168.1.122' (ECDSA) to the list of known hosts.
hadoop@slave1's password:
id_rsa.pub 100% 395 0.4KB/s 00:00
hadoop@Master:~/.ssh$
```

通过scp向远程主机拷贝文件

接着在 Slave1 节点上，将 ssh 公匙加入授权：

```
$ mkdir ~/.ssh # 如果不存在该文件夹需先创建，若已存在则忽略
$ cat ~/id_rsa.pub >> ~/.ssh/authorized_keys
$ rm ~/id_rsa.pub # 用完就可以删掉了
```

如果有其他 Slave 节点，也要执行将 Master 公匙传输到 Slave 节点、在 Slave 节点上加入授权这两步。

这样，在 Master 节点上就可以无密码 SSH 到各个 Slave 节点了，可在 Master 节点上执行如下命令进行检验，如下图所示：

```
$ ssh Slave1
```

```
hadoop@Slave1: ~
hadoop@Master:~/.ssh$ ssh Slave1 注意我是在 Master 上执行的 ssh
Welcome to Ubuntu 14.04.1 LTS (GNU/Linux 3.13.0-32-generic x86_64)

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

549 packages can be updated.
245 updates are security updates.

Last login: Sat Dec 19 19:09:57 2015 from master
hadoop@Slave1 ~$
hadoop@Slave1:~$
hadoop@Slave1:~$
hadoop@Slave1:~$
```

ssh登录后，终端标题以及命令符变为 Slave1  
此时执行的命令等同于在 Slave1 节点上执行  
(可执行 exit 退回到原来的 Master 终端)

@给力星  
powerxing.com

在Master节点中ssh到Slave节点

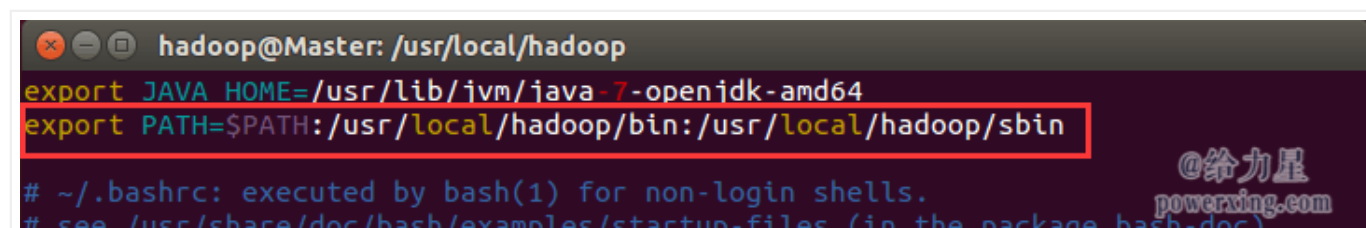
# 配置PATH变量

(CentOS 单机配置 Hadoop 的教程中有配置这一项了，这一步可以跳过)

在单机伪分布式配置教程的最后，说到可以将 Hadoop 安装目录加入 PATH 变量中，这样就可以在任意目录中直接使用 hadoo、hdfs 等命令了，如果还没有配置的，需要在 Master 节点上进行配置。首先执行 `vim ~/.bashrc`，加入一行：

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

如下图所示：



配置PATH变量

保存后执行 `source ~/.bashrc` 使配置生效。

## 配置集群/分布式环境

集群/分布式模式需要修改 `/usr/local/hadoop/etc/hadoop` 中的5个配置文件，更多设置项可点击查看官方说明，这里仅设置了正常启动所必须的设置项：`slaves`、`core-site.xml` (<http://hadoop.apache.org/docs/r2.6.0/hadoop-project-dist/hadoop-common/core-default.xml>)、`hdfs-site.xml` (<http://hadoop.apache.org/docs/r2.6.0/hadoop-project-dist/hadoop-hdfs/hdfs-default.xml>)、`mapred-site.xml` (<http://hadoop.apache.org/docs/r2.6.0/hadoop-mapreduce-client/hadoop-mapreduce-client-core/mapred-default.xml>)、`yarn-site.xml` (<http://hadoop.apache.org/docs/r2.6.0/hadoop-yarn/hadoop-yarn-common/yarn-default.xml>)。

1, 文件 **slaves**，将作为 DataNode 的主机名写入该文件，每行一个，默认为 localhost，所以在伪分布式配置时，节点即作为 NameNode 也作为 DataNode。分布式配置可以保留 localhost，也可以删掉，让 Master 节点仅作为 NameNode 使用。



本教程让 Master 节点仅作为 NameNode 使用，因此将文件中原来的 localhost 删除，只添加一行内容：Slave1。

2, 文件 **core-site.xml** 改为下面的配置：

```
<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.</de
scription>
  </property>
</configuration>
```

3, 文件 **hdfs-site.xml**，dfs.replication 一般设为 3，但我们只有一个 Slave 节点，所以 dfs.replication 的值还是设为 1：

```
<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, 文件 **mapred-site.xml**（可能需要先重命名，默认文件名为 mapred-site.xml.template），然后配置修改如下：

```
<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>
</configuration>
```

## 5, 文件 **yarn-site.xml**:

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>Master</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</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。因为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 节点上执行:

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

在 Slave1 节点上执行:

```
$ sudo rm -r /usr/local/hadoop    # 删掉旧的 (如果存在)
$ sudo tar -zxvf ~/hadoop.master.tar.gz -C /usr/local
$ sudo chown -R hadoop:hadoop /usr/local/hadoop
```

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、在 Slave 节点解压文件的操作。

首次启动需要先在 Master 节点执行 NameNode 的格式化：

```
$ hdfs namenode -format          # 首次运行需要执行初始化，之后不需要
```

## CentOS系统需要关闭防火墙

CentOS系统默认开启了防火墙，在开启 Hadoop 集群之前，需要关闭集群中每个节点的防火墙。有防火墙会导致 ping 得通但 telnet 端口不通，从而导致 DataNode 启动了，但 Live datanodes 为 0 的情况。

在 CentOS 6.x 中，可以通过如下命令关闭防火墙：

```
$ sudo service iptables stop    # 关闭防火墙服务
$ sudo chkconfig iptables off   # 禁止防火墙开机自启，就不用手动关闭了
```

若用是 CentOS 7，需通过如下命令关闭（防火墙服务改成了 firewall）：

```
$ systemctl stop firewalld.service    # 关闭firewall
$ systemctl disable firewalld.service # 禁止firewall开机启动
```

如下图，是在 CentOS 6.x 中关闭防火墙：

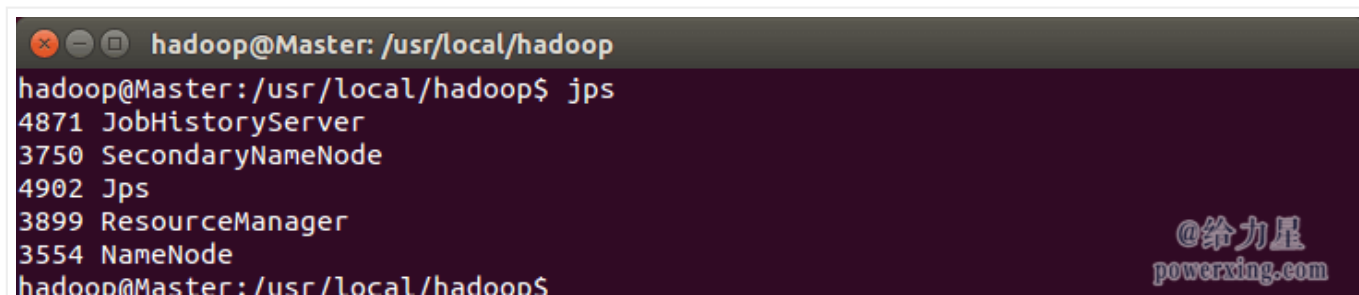


## CentOS6.x系统关闭防火墙

接着可以启动 hadoop 了，启动需要在 Master 节点上进行：

```
$ start-dfs.sh
$ start-yarn.sh
$ mr-jobhistory-daemon.sh start historyserver
```

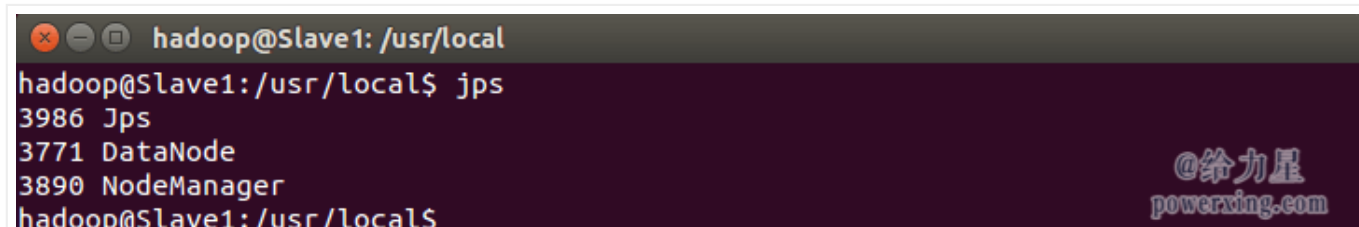
通过命令 `jps` 可以查看各个节点所启动的进程。正确的话，在 Master 节点上可以看到 NameNode、ResourceManager、SecondaryNameNode、JobHistoryServer 进程，如下图所示：

A terminal window titled 'hadoop@Master: /usr/local/hadoop' showing the output of the 'jps' command. The output lists five processes: JobHistoryServer (PID 4871), SecondaryNameNode (PID 3750), Jps (PID 4902), ResourceManager (PID 3899), and NameNode (PID 3554). A watermark '@给力星 powerxing.com' is visible in the bottom right corner.

```
hadoop@Master: /usr/local/hadoop$ jps
4871 JobHistoryServer
3750 SecondaryNameNode
4902 Jps
3899 ResourceManager
3554 NameNode
hadoop@Master: /usr/local/hadoop$
```

通过jps查看Master的Hadoop进程

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

A terminal window titled 'hadoop@Slave1: /usr/local' showing the output of the 'jps' command. The output lists three processes: Jps (PID 3986), DataNode (PID 3771), and NodeManager (PID 3890). A watermark '@给力星 powerxing.com' is visible in the bottom right corner.

```
hadoop@Slave1: /usr/local$ jps
3986 Jps
3771 DataNode
3890 NodeManager
hadoop@Slave1: /usr/local$
```

通过jps查看Slave的Hadoop进程

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 DataNode 是否正常启动，如果 Live datanodes 不为 0，则说明集群启动成功。例如我这边一共有 1 个 Datanodes：

```
hadoop@Master: /usr/local/hadoop
-----
Live datanodes (1):
Name: 192.168.1.122:50010 (Slave1)
Hostname: Slave1
Decommission Status : Normal
Configured Capacity: 7262953472 (6.76 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 5365833728 (5.00 GB)
DFS Remaining: 1897095168 (1.77 GB)
DFS Used%: 0.00%
DFS Remaining%: 26.12%
Configured Cache Capacity: 0 (0 B)
```

通过dfsadmin查看DataNode的状态

也可以通过 Web 页面看到查看 DataNode 和 NameNode 的状态：<http://master:50070/> (<http://master:50070/>)。如果不成功，可以通过启动日志排查原因。

## 伪分布式、分布式配置切换时的注意事项

1. 从分布式切换到伪分布式时，不要忘记修改 `slaves` 配置文件；
2. 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 DataNode 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`，再次启动试试。

## 执行分布式实例

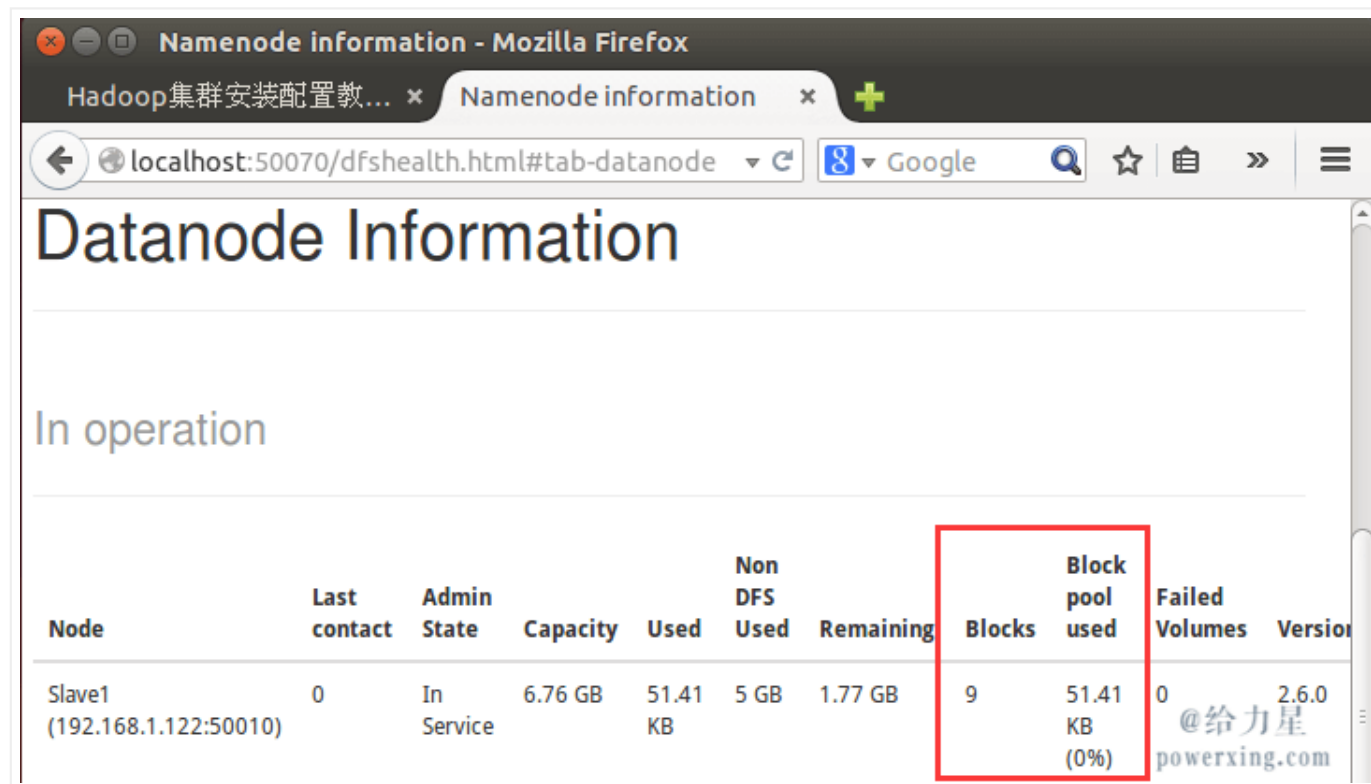
执行分布式实例过程与伪分布式模式一样，首先创建 HDFS 上的用户目录：

```
$ hdfs dfs -mkdir -p /user/hadoop
```

将 `/usr/local/hadoop/etc/hadoop` 中的配置文件作为输入文件复制到分布式文件系统中：

```
$ hdfs dfs -mkdir input
$ hdfs dfs -put /usr/local/hadoop/etc/hadoop/*.xml input
```

通过查看 DataNode 的状态（占用大小有改变），输入文件确实复制到了 DataNode 中，如下图所示：



Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
Slave1 (192.168.1.122:50010)	0	In Service	6.76 GB	51.41 KB	5 GB	1.77 GB	9	51.41 KB (0%)	0	2.6.0

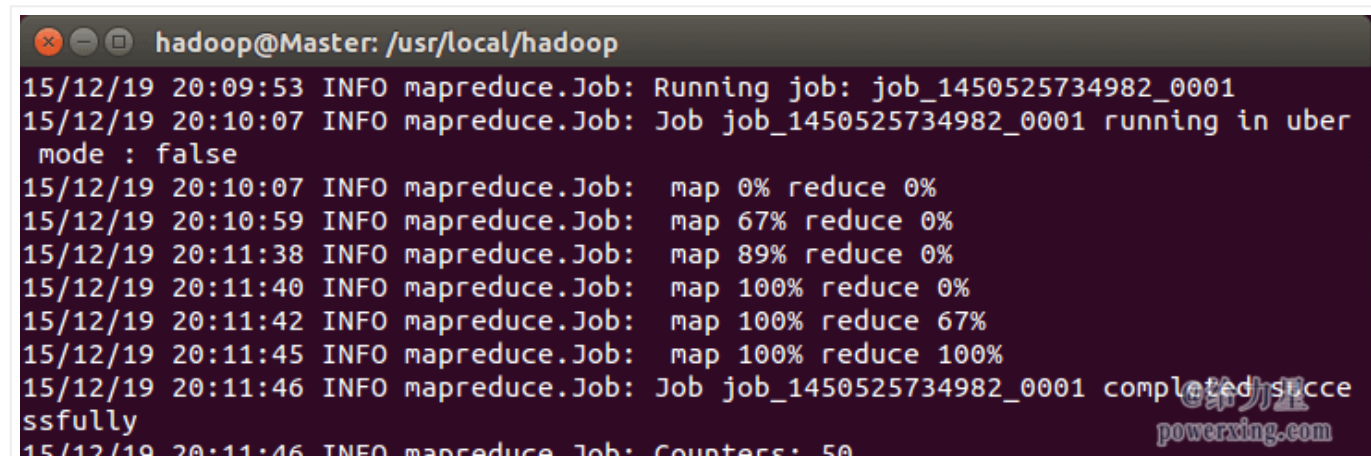
通过Web页面查看DataNode的状态

接着就可以运行 MapReduce 作业了：

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

运行时的输出信息与伪分布式类似，会显示 Job 的进度。

可能会有点慢，但如果迟迟没有进度，比如 5 分钟都没看到进度，那不妨重启 Hadoop 再试试。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 的内存配置解决。



```
hadoop@Master: /usr/local/hadoop
15/12/19 20:09:53 INFO mapreduce.Job: Running job: job_1450525734982_0001
15/12/19 20:10:07 INFO mapreduce.Job: Job job_1450525734982_0001 running in uber
mode : false
15/12/19 20:10:07 INFO mapreduce.Job: map 0% reduce 0%
15/12/19 20:10:59 INFO mapreduce.Job: map 67% reduce 0%
15/12/19 20:11:38 INFO mapreduce.Job: map 89% reduce 0%
15/12/19 20:11:40 INFO mapreduce.Job: map 100% reduce 0%
15/12/19 20:11:42 INFO mapreduce.Job: map 100% reduce 67%
15/12/19 20:11:45 INFO mapreduce.Job: map 100% reduce 100%
15/12/19 20:11:46 INFO mapreduce.Job: Job job_1450525734982_0001 completed succe
ssfully
15/12/19 20:11:46 INFO mapreduce.Job: Counters: 50
```



## 显示MapReduce Job的进度

同样可以通过 Web 界面查看任务进度 <http://master:8088/cluster>

(<http://master:8088/cluster>), 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:

Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
	default	Sun, 20 Dec 2015 05:59:42 GMT	Sun, 20 Dec 2015 06:00:12 GMT	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>
	default	Sun, 20 Dec	Sun, 20 Dec	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>

- Application
- Job
  - Overview
  - Counters
  - Configuration
  - Map tasks
  - Reduce tasks
- Tools

Job Overview

**Job Name:** grep-sort

**User Name:** hadoop

**Queue:** default

**State:** SUCCEEDED

**Uberized:** false

**Submitted:** Sun Dec 20 13:59:42 CST 2015

**Started:** Sun Dec 20 13:59:54 CST 2015

**Finished:** Sun Dec 20 14:00:12 CST 2015

**Elapsed:** 17sec

通过Web页面查看集群和MapReduce作业的信息

执行完毕后的输出结果:

```
hadoop@Master: /usr/local/hadoop
      WRONG_LENGTH=0
      WRONG_MAP=0
      WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=263
File Output Format Counters
  Bytes Written=107
hadoop@Master:/usr/local/hadoop$ ./bin/hdfs dfs -cat output/*
1      dfsadmin
1      dfs.replication
1      dfs.namenode.secondary.http
1      dfs.namenode.name.dir
1      dfs.datanode.data.dir
hadoop@Master:/usr/local/hadoop$
```

MapReduce作业的输出结果

关闭 Hadoop 集群也是在 Master 节点上执行的:

```
$ stop-yarn.sh
$ stop-dfs.sh
$ mr-jobhistory-daemon.sh stop historyserver
```

此外，同伪分布式一样，也可以不启动 YARN，但要记得改掉 `mapred-site.xml` 的文件名。

自此，你就掌握了 Hadoop 的集群搭建与基本使用了。

## 相关教程

- 使用Eclipse编译运行MapReduce程序 (<http://www.powerxing.com/hadoop-build-project-using-eclipse/>): 用文本编辑器写 Java 程序是不靠谱的，还是用 Eclipse 比较方便。
- 使用命令行编译打包运行自己的MapReduce程序 (<http://www.powerxing.com/hadoop-build-project-by-shell/>): 有时候需要直接通过命令来编译 MapReduce 程序。

## 参考资料

- <http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ClusterSetup.html> (<http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ClusterSetup.html>)

🔗 <http://www.powerxing.com/install-hadoop-cluster/> (<http://www.powerxing.com/install-hadoop-cluster/>)

📖 笔记 (<http://www.powerxing.com/notes/>) 💡

Hadoop (<http://www.powerxing.com/tag/hadoop/>)

## 相关文章

➤ **Hadoop安装教程\_单机/伪分布式配置\_Hadoop2.6.0/Ubuntu14.04**

(<http://www.powerxing.com/install-hadoop/>)

➤ **使用命令行编译打包运行自己的MapReduce程序 Hadoop2.4.1**

(<http://www.powerxing.com/hadoop-build-project-by-shell/>)

➤ **使用Eclipse编译运行MapReduce程序 Hadoop2.6.0\_Ubuntu/CentOS**

(<http://www.powerxing.com/hadoop-build-project-using-eclipse/>)

## ➤ Hadoop-in-Practice第四章MapReduce-Join代码运行问题

(<http://www.powerxing.com/hadoop-in-practice-joins/>)

## ➤ Hadoop安装配置简略教程 (<http://www.powerxing.com/install-hadoop-simplify/>)

## ➤ Hadoop安装教程\_单机/伪分布式配置\_CentOS6.4/Hadoop2.6.0

(<http://www.powerxing.com/install-hadoop-in-centos/>)

80 条评论

最新 最早 最热



Stardustsky (<http://weibo.com/2705083914>)

(<http://weibo.com/2705083914>)

2015年12月23日

感谢，之前一直在web控制台看不到节点，搞了好久都没解决，今天看到你的才知道没关防火墙，太感谢了！

回复 顶 转发



彭思瑞 (<http://t.qq.com/Ly291913010>)

(<http://t.qq.com/Ly291913010>)

2015年12月21日

教程写的很棒，不过启动hadoop后，在master和两个从节点上都启动的相应的进程，但在网址中显示存活的节点数为0，是什么原因引起的？谢谢了

回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

一般是由于防火墙的原因，虽然进程开启了，但相互之间的端口连接不了，所以活动节点数为0。

关闭防火墙再重启 Hadoop 就可以了（文中有给出 CentOS 的关闭防火墙命令）

2015年12月21日

回复 顶 转发



彭思瑞 (<http://t.qq.com/Ly291913010>)

(<http://t.qq.com/Ly291913010>)

回复 给力星：我们是在ubuntu系统下执行的，找到原因了，是hosts文件里面127.0.0.1对应的除localhost外，还有master

不过你的教程还是写的非常仔细，谢谢你！

2015年12月21日

回复 顶 转发



杨宁 (<http://t.qq.com/ab881661xy>)

(<http://t.qq.com/ab881661xy>)

2015年12月7日

楼主写的很赞，我是刚学习hadoop的，在ping slave和ping master 的时候无法ping 通，但却可以ping 通其他的地址，在master和slave两边的主机名以及hosts都做了相应的修改，向您咨询下这种情况怎么办，谢谢啊！

回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

2015年12月21日

回复 顶 转发

ping不通的话一般是因为master和slave不在同一个网段之中？



潇锋

顶顶顶

2015年12月4日

回复 顶 转发



ymj

你好，教程很好，但是我成功运行grep后，无法打开8088界面查看任务状况。该怎么解决。

2015年11月28日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

无法打开localhost:8088界面一般是因为没有开启 YARN 。

(<http://www.powerxing.com/>)

2015年12月21日    回复    顶    转发



thinks

您好，我根据您的教程已经安装完成并且单机版的例子也能运行。但是在分布式的情况下，我运行sbin/start-dfs.sh后在master输入jps，显示

```
hadoop@master:/usr/local/hadoop$ jps
```

```
8412 NameNode
```

```
8758 Jps
```

```
8629 SecondaryNameNode
```

没有ResourceManager。

在slave上输入jps显示只有DataNode没有nodemanager

然后我运行例子，提示我一直在连接ResourceManager。

这是提示信息：

```
15/11/18 10:58:55 INFO client.RMProxy: Connecting to ResourceManager at
master/10.103.241.212:8032
```

```
15/11/18 10:58:56 INFO ipc.Client: Retrying connect to server: master/10.103.241.212:8032.
```

```
Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10,
sleepTime=1000 MILLISECONDS)
```

2015年11月18日    回复    顶    转发



thinks

哈，我找到问题了。是我自己的问题。

在我的/etc/hosts文件中，我多加了几个127.0.0.1

-----  
错误的hosts文件：

```
127.0.0.1 localhost
```

```
127.0.0.1 slave
```

```
127.0.0.1 master
```

-----  
修改方法：

把hosts中的

```
127.0.0.1 slave
```

```
127.0.0.1 master
```

去掉。只留下127.0.0.1 localhost一个。

博主在教程中也说过，我没有注意。大家引以为鉴。

最后谢谢博主。

2015年11月18日    回复    顶    转发



thinks

回复 thinks: 另外博主的博客系统好赞啊。是自己写的还是用的框架？能给我一份吗？

2015年11月18日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

回复 [thinks: wordpress](http://www.powerxing.com/) 搭建的, 主题自己设计的。

2015年11月30日 回复 顶 转发



张思航 (<http://t.qq.com/zhangsihang789>)

可以在root用户下执行bin/hdfs dfs -mkdir -p /user/hadoop? 如果能执行, 文件夹创建在哪里了?

2015年11月11日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

文件夹是在hdfs上创建的, 由hdfs控制存储在各个datanode节点中, 最终的存储位置是hdfs-site.xml中的dfs.datanode.data.dir项指定的位置。

2015年12月21日 回复 顶 转发



Kopo (<http://t.qq.com/KopoDidream>)

楼主, 其他都对, 就是slaves启动后没有datanode, 只有nodemanage, 这是怎么回事?

2015年11月7日 回复 顶 转发



风尘

你好, 我也出现过这个问题, 最后是把临时文件tmp删除后, 重启hadoop, 就可以啦, 文章中有提到

2015年11月22日 回复 顶 转发



王旭

hi ,按照您的教程最后我的namenode开启不了。。

2015年11月5日 回复 顶 转发



月亮忘记了AA (<http://weibo.com/2970678631>)

刚刚开始学习hadoop, 单机的, 分布式的也是按照你的教程来, 给了我很大的帮助, 谢谢你哈! 不过, 在集群下最后一步执行mapreduce successfully之后, 又多了一句redirecting connect server什么的, 导致后面报错, 结果也输出不出来不知道为什么

2015年10月22日 回复 顶 转发



深海

您好, 我请问下, 我用http://master: 50070登录HDFSweb页面, 在其中的Utilities — Browse Directory中, 我想下载我上传的文件, 但我点击download, 就提示找不到页面, 请问这是是什么原因, 谢谢

2015年10月12日 回复 顶 转发



zxdzxdxghxgh

你好, 有问题请教。按照教程上的设置后, 运行Eclipse的时候出现Call From Master/192.168.1.121 to localhost:9000 failed on connection exception: java.net.ConnectException: 拒绝连接; For more details

see:<http://wiki.apache.org/hadoop/ConnectionRefused>这个错误，是什么原因啊？

2015年10月4日    回复    顶    转发

---



谢谢楼主

bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar grep input output 'dfs[a-z.]+' 这条命令是什么意思？还有'dfs[a-z.]+'这个又是什么意思？ 还有文中将etc/hadoop中的文件作为输入文件复制到分布式文件系统中：bin/hdfs dfs -put etc/hadoop input。假如我需要更换输入文件，换成我自己的输入文件，我应该如何操作？

2015年9月24日    回复    顶    转发

---



寒风大侠

命令是在input中抓取正则表达式dfs[a-z.] +（这个表达式的意思是dfs开头+多个小写字母和.构成的字符串）并将结果放入output文件夹中

这里的input和output都是hdfs里面的文件夹

可以用bin/hdfs dfs -ls命令查看

2015年10月27日    回复    顶    转发

---



小明滚

到 bin/hdfs dfs -put etc/hadoop input 那步：

put: File /user/peiming/input/hadoop/yarn-site.xml.\_COPYING\_ could only be replicated to 0 nodes instead of minReplication (=1). There are 0 datanode(s) running and no node(s) are excluded in this operation.

但是jps查看Slave1是有启动datanode的啊。这是为什么？

2015年8月27日    回复    顶    转发

---



寒风大侠

有可能是防火墙或者ssh没有配置好

2015年10月27日    回复    顶    转发

---



伤我

mapred-site.xml 应该要配置成

mapreduce.jobhistory.addressMaster.Hadoop:10020

要不然 bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar grep input output 'dfs[a-z.]+' 一直报错java.net.ConnectException: to 0.0.0.0:10020 failed错误

2015年8月26日    回复    顶    转发

---



heng

您好，我执行例程的时候总是卡在15/08/24 19:51:24 INFO mapreduce.Job: map 0% reduce 0% 这里，不知道为什么。好无语。请问您知道怎么解决吗

2015年8月25日    回复    顶    转发

---



给力星 (<http://www.powerxing.com/>)

在本页面 Ctrl+F 搜索一下 wuyang 的留言，看看是否一样的情况。  
(<http://www.powerxing.com/>)

2015年8月25日    回复    顶    转发

---





唐轶男 (<http://t.qq.com/tyn513>)

您好，我执行了分布式实例后 在reduce未完成时发生错误：

(<http://t.qq.com/tyn513>) error:java.io.FileNotFoundException: Path is not a file: /usr/hadoop/input/hadoop  
不知道如何解决，是不是配置出错了？

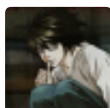
2015年8月17日    回复    顶    转发



keketi

同样有这个错误，请问解决了吗？

2015年11月21日    回复    顶    转发



用户名什么最讨厌了 (<http://weibo.com/3535325540>)

lz你好，如果slave节点hadoop安装目录不是/usr/local/hadoop应该怎么配置？

(<http://weibo.com/3535325540>)

2015年8月3日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 这个没研究过。建议都安装在统一的目录下，避免无谓的麻烦。

2015年8月4日    回复    顶    转发



用户名什么最讨厌了 (<http://weibo.com/3535325540>)

(<http://weibo.com/3535325540>) 回复 给力星: 另外，因为slave节点在服务器上，没有服务器的root权限，没办法修改/etc/hosts，这时，master和slave节点都应该怎么配置？

2015年8月4日    回复    顶    转发



伤我

回复 用户名什么最讨厌了: 使用sudo vim 来修改

2015年8月26日    回复    顶    转发



用户名什么最讨厌了 (<http://weibo.com/3535325540>)

(<http://weibo.com/3535325540>) 回复 伤我: 没有root权限啊

2015年10月26日    回复    顶    转发



citizen2047

楼主求助，目前我两个虚拟机的datanode进程都启动了，master跟slave的jps看起来都没有问题，但是用hdfs dfsadmin -report看的时候却没有node，网页上看的时候live node跟dead node都是0，好奇怪啊，是我哪里配错了么？

2015年8月2日    回复    顶    转发



citizen2047

看了slave的log，貌似一直在连接失败，都是这样的信息

2015-08-02 15:27:52,657 INFO org.apache.hadoop.ipc.Client: Retrying connect to server: localhost/127.0.0.1:9000. Already tried 3 time(s); retry policy is

RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)

是不是我hosts设置不对呢？但是重复检查了好几遍，并没有找出问题所在。。

2015年8月2日    回复    顶    转发



citizen2047

回复 citizen2047: 好吧是core-site.xml没配对。。

2015年8月2日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

回复 citizen2047: 可以了就好哈  
(<http://www.powerxing.com/>)

2015年8月4日    回复    顶    转发



loveyy

回复 citizen2047: 你好，我想问下怎么改的，我也是这个问题

2015年8月25日    回复    顶    转发



y'y'y

回复 citizen2047: 请问您是怎么改的，我也是这个问题。

2015年11月22日    回复    顶    转发



yyy

回复 loveyy: 请问最后你是改的，我也是这个问题。

2015年11月22日    回复    顶    转发



hadoop学习

大神，我在Master节点上启动hadoop，通过命令jps查看各个节点所启动的进程，发现 master 多了nodemanager和datanode，而slave01只有jps，slave02也是，这是什么问题呢

2015年7月28日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

应该是配置错误，配置正确的话，master 不会有 datanode 进程的。检查一下  
(<http://www.powerxing.com/>)  
/usr/local/hadoop/etc/hadoop/slaves 这个文件的配置，或者查看一下启动日志。

2015年7月28日    回复    顶    转发



小明

ssh Slave1那里，应该要加上Slave1主机的用户名吧：ssh B@Slave1（假设我Master的用户名是A，Slave1的用户名是B）

2015年7月26日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

ssh的时候，如果没有指定用户名，则是默认使用当前的用户名。之前写教程的时候用的都是  
(<http://www.powerxing.com/>)  
hadoop用户名，所以没有这个问题，改天我补充一下。

2015年7月27日    回复    顶    转发



小明

回复 给力星: 真的是这样，是我太菜了。感谢楼主！

2015年7月27日    回复    顶    转发



小明

求问大神：执行到Master无密码登录Slave1那一步。在Master终端上输入ssh Slave1,结果提示要输入A@slave1's password。奇怪的是A是Master的用户名，而且slave1的s小写了。

2015年7月26日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) ip的名称映射是不区分大小写的，就像域名也不区分大小写。

2015年7月27日    回复    顶    转发



wuyang

到最后一步运行Job之前都没有什么问题，但是运行

```
bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.1.jar grep /input output 'dfs[a-z.]+'
```

的时候，一直在这一步没有反应15/07/20 22:11:06 INFO mapreduce.Job: Running job:

job\_1437444490434\_0001

按你提示重启了也没有改变。用hadoop job -status 查看工作进程

Job: job\_1437444490434\_0001

Job File: /tmp/hadoop-yarn/staging/root/.staging/job\_1437444490434\_0001/job.xml

Job Tracking URL : http://node1-1.sb5.orbit-

lab.org:8088/proxy/application\_1437444490434\_0001/

Uber job : false

Number of maps: 0

Number of reduces: 0

map() completion: 0.0

reduce() completion: 0.0

Job state: PREP

retired: false

reason for failure:

Counters: 0

状态是PREP 一直在等待状态。请问有什么办法解决吗

2015年7月21日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 1. 我看你的输入文件夹是 /input，这是hdfs根目录下的input，但输出结果目录是 /user/hadoop/output，你确认一下。

2. 检查一下 datanode 的状态: bin/hdfs dfsadmin -report

3. 我搜索了一下，网上也没多少说法，有一个说法是"原因是job得不到执行所需的内存资源，解决办法是调整yarn-site.xml中和分配内存相关的几个参数，这个取决于你具体执行的job所需内存以及你各个节点的物理可用内存"，你不妨试一下。

2015年7月21日    回复    顶    转发



wuyang

回复 给力星: 谢谢lz答复！确实调整一下内存就解决啦

2015年7月22日    回复    顶    转发



PageNotFound

回复 wuyang: 请问你是把内存调到了多大, 设置了哪些参数, 我也是这个问题, 我是1台master, 2台slave, 每台都是1core, 1G的vm, 也遇到了这个问题

2015年10月12日    回复    顶    转发

---



yyy

您好, 请问您是怎么解决的这个问题的。我也遇到了同样的问题, 该如何调整yarn-site.xml的内存呢? 请您指教, 我已经被这个问题困扰好几天了, 谢谢了。

2015年12月9日    回复    顶    转发

---



yyy

您好, 请问您是怎么解决的这个问题的。我也遇到了同样的问题, 该如何调整yarn-site.xml的内存呢? 请您指教, 我已经被这个问题困扰好几天了, 谢谢了。

2015年12月9日    回复    顶    转发

---



mm

回复 PageNotFound: 请问您这个问题解决了吗? 我也遇到了同样的问题, 请指教, 我已经被这个问题困扰好几天了, 希望能得到您的帮助, 谢谢!

2015年12月9日    回复    顶    转发

---



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

回复 yyy: <http://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-common/yarn-default.xml> 我不太清楚需要改哪些参数, 我没遇到这问题。不过用关键字 memory 搜索, 例如可以找到 yarn.scheduler.minimum-allocation-mb 参数, 默认为1024mb, 这是最小分配内容, 把这个参数改小一些应该就可以了。

2015年12月11日    回复    顶    转发

---



Zhisong

教程很实用, 多谢多谢

2015年7月12日    回复    顶    转发

---



徐文涛 (<http://t.qq.com/Reborn0126>)

(<http://t.qq.com/Reborn0126>) scp ~/.ssh/id\_rsa.pub hadoop@Slave1:/home/hadoop/。。。我ping Slave1是可以的。为什么使用这个命令后没有反应, 没有提示输入密码。一直没反应, 接着就提示Connection time out

2015年7月11日    回复    顶    转发

---



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

看样子应该是 Slave1 主机上没有安装 SSH server?

2015年7月11日    回复    顶    转发

---



dong

你好, 请教个问题, 我在虚拟机上, 两台机器, 建了个master, 建了一个slave, 是完全分布式模式, 在master启动hdfs后, 通过jps看都正常, 但是就是没有datanode, 大概是因为什么原因啊?

2015年7月9日    回复    顶    转发

---

给力星 (<http://www.powerxing.com/>)



查看一下启动日志 `/usr/local/hadoop/logs/hadoop-hadoop-datanode-*.out`，看下最后面有带 Fatal、Error 或者 Java Exception 的信息。  
(<http://www.powerxing.com/>)

2015年7月10日    回复    顶    转发



(<http://t.qq.com/wlq591111>)

龙奇 (<http://t.qq.com/wlq591111>)

朋友 我是一个hadoop新手 那个hadoop集群完全分布配置 我第一步就卡住了 怎么建立两个主机啊

2015年7月9日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)  
(<http://www.powerxing.com/>)

集群一般是多台机器（电脑），我教程中所说的就是两台机器。如果条件不允许，可以使用虚拟机，同时开启两个虚拟机，也可以搭集群。

2015年7月9日    回复    顶    转发



龙奇 (<http://t.qq.com/wlq591111>)  
(<http://t.qq.com/wlq591111>)

回复 给力星：我可以加你的qq么 好想认识你啊 我建立完虚拟机 让他们ping 结果就From bogon (192.168.1.120) icmp\_seq=2 Destination Host Unreachable 这个192.168.1.120也没有设置过啊

2015年7月9日    回复    顶    转发



龙奇 (<http://t.qq.com/wlq591111>)  
(<http://t.qq.com/wlq591111>)

回复 给力星：我建立了一个虚拟机，又克隆了这个虚拟机。在两个虚拟机中分别\$`sudo vim /etc/hostname` # 修改了主机名为Master和Slave1

2.\$`sudo vim /etc/hosts` # 在此处修改主机Ip: 192.168.1.121 和 分机IP为192.168.1.122 。

之后我在主机里ping Slave1 出现了

From bogon (192.168.1.120) icmp\_seq=2 Destination Host Unreachable

From bogon (192.168.1.120) icmp\_seq=5 Destination Host Unreachable

2015年7月9日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)  
(<http://www.powerxing.com/>)

回复 龙奇：这个问题可以自行搜索一下“两个虚拟机 ping不通”，例如这篇文章有讲到原因与解决方法 [http://blog.csdn.net/jerry\\_cool/article/details/7065698](http://blog.csdn.net/jerry_cool/article/details/7065698)

2015年7月10日    回复    顶    转发



罹离  
赞

2015年6月18日    回复    顶    转发



df  
bucuo de jiaocheng

2015年5月18日    回复    顶    转发



Kartty



2015年3月13日    回复    顶    转发



给力星 (<http://www.powerxing.com/>)

谢谢认可~  
(<http://www.powerxing.com/>)

2015年3月15日    回复    顶    转发

---



Hadoop

非常详细的教程，一次就成功~

2015年3月13日    回复    顶    转发

---



给力星 (<http://www.powerxing.com/>)

谢谢认可~  
(<http://www.powerxing.com/>)

2015年3月15日    回复    顶    转发

---



Hadoop

将etc/hadoop中的文件作为输入文件复制到分布式文件系统中。

`bin/hdfs dfs -put etc/hadoop input`

会提示，input不存在。是不是得建一个，那么前面的user/hadoop 建了又是干啥的呢？？这南非大裂谷般的跳跃实在超出我的理解范围了

2015年3月4日    回复    顶    转发

---



给力星 (<http://www.powerxing.com/>)

这个问题在那篇单机教材中有提及。  
(<http://www.powerxing.com/>)

2015年3月15日    回复    顶    转发

---



小明

我也遇到这个问题了！！请问层主解决了没有？同样疑问“前面的user/hadoop 建了又是干啥的呢？”看单机篇没有提及啊

2015年7月26日    回复    顶    转发

---



寒风大侠

回复 小明: 感觉是在hadoop下的文件系统里建立的文件夹  
用来存储hadoop运行过程中的一些文件

2015年10月27日    回复    顶    转发

---



Hadoop

教程很详细，非常感谢。

2015年2月18日    回复    顶    转发

---



给力星 (<http://www.powerxing.com/>)

谢谢认可~  
(<http://www.powerxing.com/>)

2015年3月15日    回复    顶    转发

---

社交帐号登录:    微信    微博    QQ    人人    更多»





说点什么吧...	
	发布

多说 (<http://duoshuo.com>)

© 2014 给力星 (<http://www.powerxing.com>)