

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安装配置简略教程

📅 2015-05-10 (updated: 2015-05-31) 👁 4512

本Hadoop安装教程为简略版本，包含伪分布式和集群的配置，只列出命令与配置，不作详细说明，方便有基础的读者。完整版请浏览Hadoop安装配置教程_伪分布式 (<http://www.powerxing.com/install-hadoop/>)，以及Hadoop集群安装配置教程 (<http://www.powerxing.com/install-hadoop-cluster/>)。

本教程在 Ubutnu 14.04 64位，Hadoop 2.6.0 下验证通过。最新的 Hadoop 2 稳定版可以通过 <http://mirror.bit.edu.cn/apache/hadoop/common/stable2/> (<http://mirror.bit.edu.cn/apache/hadoop/common/stable2/>) 或者

<http://mirrors.cnnic.cn/apache/hadoop/common/stable/>
(<http://mirrors.cnnic.cn/apache/hadoop/common/stable/>) 下载。

Hadoop 伪分布式配置

创建Hadoop用户:

```
$ sudo useradd -m hadoop -s /bin/bash      # 创建hadoop用户
$ sudo passwd hadoop                        # 修改密码
$ sudo adduser hadoop sudo                 # 增加管理员权限
```

注销并使用 Hadoop 用户登录，接着更新 apt 并且安装 vim:

```
$ sudo apt-get update                    # 更新apt
$ sudo apt-get install vim              # 安装vim
```

安装SSH，配置无密码登录:

```
$ sudo apt-get install openssh-server
$ cd ~
$ mkdir .ssh                            # 可能该文件已存在，不影响
$ cd ~/.ssh/
$ ssh-keygen -t rsa                      # 会有提示，都按回车就可以
$ cat id_rsa.pub >> authorized_keys     # 加入授权
```

安装Java环境:

```
$ sudo apt-get install openjdk-7-jre openjdk-7-jdk
$ vim ~/.bashrc                          # 设置JAVA_HOME
```

在文件最前面添加如下单独一行:

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
```

使 JAVA_HOME 变量生效:

```
$ source ~/.bashrc    # 使变量设置生效
```

安装 Hadoop 2:

```
$ cd ~/下载
$ sudo tar -zxvf ./hadoop-2.6.0.tar.gz -C /usr/local # 解压到/usr/local中
$ cd /usr/local/
$ sudo mv ./hadoop-2.6.0/ ./hadoop                # 将文件夹名改为hadoop
$ sudo chown -R hadoop:hadoop ./hadoop            # 修改文件权限
```

进行伪分布式配置:

修改配置文件 **core-site.xml** (`vim /usr/local/hadoop/etc/hadoop/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://localhost:9000</value>
  </property>
</configuration>
```

修改配置文件 **hdfs-site.xml**:

```
<configuration>
  <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>
```

启动 Hadoop:

```
$ cd /usr/local/hadoop
$ bin/hdfs namenode -format          # namenode 格式化
$ sbin/start-dfs.sh                  # 开启守护进程
$ jps                                # 判断是否启动成功
```

若成功启动则会列出如下进程: NameNode 、 DataNode 和 SecondaryNameNode 。

运行 WordCount 实例:

```
$ bin/hdfs dfs -mkdir -p /user/hadoop      # 创建HDFS目录
$ bin/hdfs dfs -mkdir input
$ bin/hdfs dfs -put etc/hadoop/*.xml input  # 将配置文件作为输入
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep input output 'dfs[a-z.]+'
$ bin/hdfs dfs -cat output/*                # 查看输出
```

Hadoop 集群配置

假定有两台机器:

```
Master  192.168.1.121
Slave1   192.168.1.122
```

Hadoop 集群配置过程:

1. 选定一台机器作为 Master，在所有主机上配置网络映射
2. 在 Master 主机上配置hadoop用户、安装SSH server、安装Java环境
3. 在 Master 主机上安装Hadoop，并完成配置
4. 在其他主机上配置hadoop用户、安装SSH server、安装Java环境
5. 将 Master 主机上的Hadoop目录复制到其他主机上
6. 开启、使用 Hadoop

所有主机配置hadoop用户、安装SSH server、安装Java环境：

```
$ sudo useradd -m hadoop -s /bin/bash      # 创建hadoop用户
$ sudo passwd hadoop                       # 修改密码
$ sudo adduser hadoop sudo                 # 增加管理员权限
$ # 注销并使用 Hadoop 用户登录
$ sudo apt-get update                     # 更新apt
$ sudo apt-get install vim                # 安装vim
$ sudo apt-get install openssh-server    # 安装ssh
$ sudo apt-get install openjdk-7-jre openjdk-7-jdk # 安装Java
$ vim ~/.bashrc                           # 设置JAVA_HOME
```

在文件最前面添加如下单独一行：

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
```

使 JAVA_HOME 变量生效：

```
$ source ~/.bashrc      # 使变量设置生效
```

所有主机配置网络映射：

```
$ sudo vim /etc/hostname      # 修改主机名
$ sudo vim /etc/hosts         # 修改主机与 IP 的映射关系
$ sudo reboot                 # 重启，使网络配置生效
```

在 Master 主机上执行：

```
$ cd ~/.ssh
$ ssh-keygen -t rsa # 一直接回车就可以
$ cat ~/id_rsa.pub >> ~/authorized_keys
$ scp ~/.ssh/id_rsa.pub hadoop@Slave1:/home/hadoop/ # 传输公钥到Slave1
```

接着在 Slave1 节点上执行

```
$ cd ~
$ mkdir .ssh
$ cat ~/id_rsa.pub >> ~/.ssh/authorized_keys
```

在Master节点上进行Hadoop集群配置(位于 /usr/local/hadoop/etc/hadoop中):

文件 **slave:**

将原来 localhost 删除, 把所有Slave的主机名写上, 每行一个。

文件 **core-site.xml:**

```
<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>
```

文件 **hdfs-site.xml:**

```
<property>
  <name>dfs.namenode.secondary.http-address</name>
  <value>Master:50090</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>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
```

文件 **mapred-site.xml**(首先需执行 `cp mapred-site.xml.template mapred-site.xml`):

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

文件 **yarn-site.xml**:

```
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>Master</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
```

配置好后, 在 Master 主机上, 将 Hadoop 文件复制到各个节点上:

```
$ cd /usr/local
$ rm -r ./hadoop/tmp # 删除 Hadoop 临时文件
$ sudo tar -zcf ./hadoop.tar.gz ./hadoop
$ scp ./hadoop.tar.gz Slave1:/home/hadoop
```

在 Slave1 上执行:

```
$ sudo tar -zxf ~/hadoop.tar.gz -C /usr/local
$ sudo chown -R hadoop:hadoop /usr/local/hadoop
```

最后在 Master 主机上就可以启动hadoop了:

```
$ cd /usr/local/hadoop/
$ bin/hdfs namenode -format
$ sbin/start-dfs.sh
$ sbin/start-yarn.sh
$ jps                # 判断是否启动成功
```

若成功启动, 则Master节点启动了

NameNode 、 SecondrryNameNode 、 ResourceManager 进程, Slave节点启动了
DataNode 和 NodeManager 进程。

在 Master 主机上执行WordCount实例:

```
$ bin/hdfs dfs -mkdir -p /user/hadoop
$ bin/hdfs dfs -put etc/hadoop input
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-exampl
es-*.jar grep input output 'dfs[a-z.]+'
```

附加教程: 配置PATH环境变量

在 ~/.bashrc 最前面加入如下单独一行:

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

🔗 <http://www.powerxing.com/install-hadoop-simplify/>

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

📖 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/>)

➤ **Hadoop集群安装配置教程_Hadoop2.6.0_Ubuntu/CentOS**

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

➤ **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/>)

5 条评论

最新 最早 最热



benjamin (<http://www.douban.com/people/103654556/>)

(<http://www.douban.com/people/103654556/>)

楼主，请问你看过hadoop权威指南 第三版 那本书吗？里面的
大气数据你用过了没？能否给份数据呢？

2015年11月10日 回复 顶 转发



是这个吧
(<http://folkcat.com/>)

<http://download.csdn.net/detail/u011848525/8762077>

2015年11月14日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

回复 :) 谢谢楼主，以后可以多多交流
(<http://www.douban.com/people/103654556/>)

2015年11月22日 回复 顶 转发



赶紧考完放假廖 (<http://weibo.com/2754778401>)

(<http://weibo.com/2754778401>)

您好学长，我在进行bin/hdfs dfs -mkdir input命令行时提示'input': No such
file or directory
请问这是怎么回事？

2015年9月26日 回复 顶 转发



赶紧考完放假廖 (<http://weibo.com/2754778401>)

已解决，是我的用户不叫做hadoop。。看过那么多hadoop配置教程，还数这个质量最高，非常
感谢！
(<http://weibo.com/2754778401>)

2015年9月26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>)