

1、JDK 安装,解压到/usr/local/jdk1.6,并配置/etc/profile 环境

```
export JAVA_HOME=/usr/local/jdk1.6
```

```
export PATH=$JAVA_HOME/bin:$JAVA_HOME/jre:$PATH
```

```
export CLASSPATH=$JAVA_HOME/lib:$JAVA_HOME/lib/tools.jar
```

2、下载 Hadoop 解压到每台服务器的/data/soft 目录下，最好目录都相同。

```
tar zxvf hadoop-2.20.1.tar.gz
```

配置

配置说明

具体配置选项说明详见

http://www.hadoop.org.cn/document/cn/r0.18.2/cluster_setup.html

1、修改 **conf/hadoop-env.sh**，添加 **jdk** 支持

```
export JAVA_HOME=/usr/local/jdk1.6
```

2、**cd/data/soft/hadoop-2.20.1** 修改 **hadoop** 目录下的 **conf/core-site.xml** 文件，增加以下内容

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://ws:54310</value> //你的 namenode 的配置，机器名加端口
  </property>
  <property>
    <name>dfs.name.dir</name>
    <value>/data/hdfs/hdfs1</value> //NameNode 持久存储名字空间及事务日志的本地文件系统路径。
  </property>
```

```
<property>
<name>dfs.data.dir</name>  //DataNode 存放块数据的本地文件系统路
径，逗号分割的列表。
```

```
<value>/data/hdfs/hdfsdata1</value>
```

```
</property>
```

```
<property>
```

```
<name>hadoop.tmp.dir</name>> //Hadoop 的默认临时路径，这个最好配置，然后在
新增节点或者其他情况下莫名其妙的 DataNode 启动不了，就删除此文件中的 tmp 目录
即可。不过如果删除了 NameNode 机器的此目录，那么就需要重新执行 NameNode 格式化
的命令了。
```

```
<value>/data/hdfs/hdfstmp</value>
```

```
<description>A base for other temporary directories.</description>
```

```
</property>
```

```
<property>
```

```
<name>dfs.replication</name>  //数据需要备份的数量，默认是三
```

```
<value>1</value>
```

```
</property>
```

```
</configuration>
```

3、修改 **conf/mapred-site.xml** 增加如下内容

```
<property>
```

```
<name>mapred.job.tracker</name> //JobTracker 的主机（或者 IP）
和端口。
```

```
<value>ws:54311</value>
```

```
</property>
```

4、配置 **/masters**,修改 **conf / masters**

```
master-dfs
```

修改 hosts 文件 vi/etc/hosts

192.168.1.11 master-dfs

192.168.1.12 data-dfs

192.168.1.13 client-dfs

5、配置 **slaves**,修改 **conf/slaves**

data-dfs

复制 **hadoop-2.20.1** 目录以及所有文件到 **DataNode** 和 **Client**

```
scp -rp /data/soft/hadoop-2.20.1 192.168.1.12:/data/soft
```

```
scp -rp /data/soft/hadoop-2.20.1 192.168.1.13:/data/soft
```

6、格式化 **Hdfs** 文件系统 **namenode**,进入 **hadoop** 目录

```
[master@hadoop]$bin/hadoop namenode -format
```

输入 **Y**, (注意区分大小写.这里一定要输入大写的 **Y**,否则不会成功 **format** 文件系统)

不出意外，应该会提示格式化成功。如果不成功，就去 **hadoop/logs/** 目录下去查看日志文件。

7、启动 **Hdfs** 分布式文件系统

进入 **hadoop** 目录,在 **bin/**下面有很多启动脚本，可以根据自己的需要来启动。

* **start-all.sh** 启动所有的 Hadoop 守护。包括 **namenode**, **datanode**, **jobtracker**, **tasktrack**

* **stop-all.sh** 停止所有的 Hadoop

* **start-mapred.sh** 启动 Map/Reduce 守护。包括 **Jobtracker** 和 **Tasktrack**

* **stop-mapred.sh** 停止 Map/Reduce 守护

* `start-dfs.sh` 启动 Hadoop DFS 守护.Namenode 和 Datanode

* `stop-dfs.sh` 停止 DFS 守护

启动所有守护

`[master@hadoop]bin/start-all.sh` //会自动找到 `conf/slaves` 文件里的 IP 或者机器名, 启动相应的 `slaves` 服务端

同样, 如果要停止 `hadoop`, 则

`[master@hadoop]$bin/stop-all.sh`

8、HDFS 操作

进入 Hadoop 目录,运行 `bin/`目录的 `hadoop` 命令, 可以查看 Hadoop 所有支持的操作及其用法, 这里以几个简单的操作为例。

建立目录

`[master@hadoop]$bin/hadoop dfs -mkdir testdir`

在 HDFS 中建立一个名为 `testdir` 的目录

复制文件

`[master@hadoop]$bin/hadoop dfs -put /home/dbrg/large.zip
testfile.zip`

把本地文件 `large.zip` 拷贝到 HDFS 的根目录/`user/dbrg/`下, 文件名为
`testfile.zip`

查看现有文件

`[master@hadoop]$bin/hadoop dfs -ls`

运行 `hadoop` 程序时, 中途我把它终止了, 然后再向 `hdfs` 加文件或删除文件时, 出现

Name node is in safe mode 错误:

```
rmr: org.apache.hadoop.dfs.SafeModeException: Cannot delete  
/user/hadoop/input. Name node is in safe mode
```

解决的命令:

```
bin/hadoop dfsadmin -safemode leave #关闭 safe mode
```

本文转载自: <http://cqfish.blog.51cto.com/622299/207766>