

# Hadoop 2.x 集群系统安装

侯在钱

## 目 录

1. Hadoop 集群系统安装.....	1
1.1. 上传 Hadoop 安装文件，并解压.....	1
1.2. 将安装目录权限赋予给 Hadoop 用户.....	2
1.3. 修改 <code>hadoop-env.sh</code> 配置文件.....	2
1.4. 修改 <code>slaves</code> 配置文件.....	2
1.5. 修改 <code>core-site.xml</code> 配置文件.....	2
1.6. 修改 <code>hdfs-site.xml</code> 配置文件.....	3
1.7. 修改 <code>yarn-site.xml</code> 配置文件.....	3
1.8. 修改 <code>mapred-site.xml</code> 配置文件.....	4
1.9. 将 Master 配置好的 Hadoop 系统拷贝到所有 Slave 上.....	5
1.10. 在 Master 上设置 Hadoop 的环境变量.....	5
2. Hadoop 格式化.....	5
3. Hadoop 启动.....	6

本文使用的版本是：`hadoop-2.6.0.tar.gz`。

## 1. Hadoop 集群系统安装

以下工作都在 Master 主机上执行配置，配置好后，在 1.9 节把此配置复制到其他 Slave 设备上。

### 1.1. 上传 Hadoop 安装文件，并解压

新建 `hadoop` 安装目录，命令如下：

```
mkdir /usr/local/hadoop
```

把 `hadoop-2.6.0.tar.gz` 安装文件上传到 `/usr/local/hadoop/` 目录下，然后使用如下命令解压：

```
tar -xvzf hadoop-2.6.0.tar.gz
```

## 1.2. 将安装目录权限赋予给 Hadoop 用户

在每一台机器上执行如下命令：

```
chown -R hadoop.hadoop /usr/local/hadoop/
```

## 1.3. 修改 hadoop-env.sh 配置文件

配置文件位置：/usr/local/hadoop/hadoop-2.6.0/etc/hadoop/hadoop-env.sh

找到如下的内容：

```
export JAVA_HOME=${JAVA_HOME}
```

修改成如下的内容：

```
export JAVA_HOME=/usr/java/jdk1.7.0_04/
```

特别注意的是，红色部分为 Java 的实际 HOME 目录，需要根据自己安装的 Java 的 HOME 路径确定。可 cd 到 /usr/java/ 目录下查看实际目录名称。

## 1.4. 修改 slaves 配置文件

配置文件位置：/usr/local/hadoop/hadoop-2.6.0/etc/hadoop/slaves

在这个配置文件中配置有哪些从服务器。我们把 hadoop1 和 hadoop2 作为 Slave 从服务器，hadoop0 作为 Master 主服务器。

修改后的内容：

```
hadoop1
```

```
hadoop2
```

## 1.5. 修改 core-site.xml 配置文件

配置文件位置：/usr/local/hadoop/hadoop-2.6.0/etc/hadoop/core-site.xml

修改后的内容：

```
<configuration>
```

```
  <property>
```

```
<name>fs.defaultFS</name>
<value>hdfs://hadoop0:9000</value>
</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/usr/local/hadoop/hadoop-tmp</value>
</property>
</configuration>
```

## 1.6. 修改 hdfs-site.xml 配置文件

配置文件位置：/usr/local/hadoop/hadoop-2.6.0/etc/hadoop/hdfs-site.xml

修改后的内容：

```
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///usr/local/hadoop/hadoop-dfs/name</value>
  </property>
  <property>
    <name>dfs.namenode.data.dir</name>
    <value>file:///usr/local/hadoop/hadoop-dfs/data</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
</configuration>
```

## 1.7. 修改 yarn-site.xml 配置文件

配置文件位置：/usr/local/hadoop/hadoop-2.6.0/etc/hadoop/yarn-site.xml

修改后的内容：

```
<configuration>
  <property>
    <name>yarn.resourcemanager.address</name>
    <value>hadoop0:8032</value>
  </property>
  <property>
    <name>yarn.resourcemanager.scheduler.address</name>
    <value>hadoop0:8030</value>
  </property>
</configuration>
```

```
<property>
  <name>yarn.resourcemanager.resource-tracker.address</name>
  <value>hadoop0:8031</value>
</property>
<property>
  <name>yarn.resourcemanager.admin.address</name>
  <value>hadoop0:8033</value>
</property>
<property>
  <name>yarn.resourcemanager.webapp.address</name>
  <value>hadoop0:8088</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>
```

## 1.8. 修改 mapred-site.xml 配置文件

配置文件位置：/usr/local/hadoop/hadoop-2.6.0/etc/hadoop/mapred-site.xml

修改后的内容：

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>hadoop0:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>hadoop0:19888</value>
  </property>
</configuration>
```

## 1.9. 将 Master 配置好的 Hadoop 系统拷贝到所有 Slave 上

分别在 `hadoop1` 和 `hadoop2` 上执行如下命令创建目录：

```
mkdir -p /usr/local/hadoop
chmod 777 /usr/local/hadoop
```

在 Master 上执行如下两条命令，将 Master 配置好的 Hadoop 系统拷贝到所有 Slave（`hadoop1` 和 `hadoop2`）上：

```
scp -r /usr/local/hadoop/hadoop-2.6.0/ hadoop@hadoop1:/usr/local/hadoop/
scp -r /usr/local/hadoop/hadoop-2.6.0/ hadoop@hadoop2:/usr/local/hadoop/
```

注意：执行命令后，系统需要输入拷贝目标主机上名为 `hadoop` 用户的密码，本文在所有 Slave 机器上创建了 `hadoop` 用户，密码设置的也是 `hadoop`。

## 1.10. 在 Master 上设置 Hadoop 的环境变量

在 `/etc/profile` 文件的末尾处，增加如下的环境变量：

```
export HADOOP_HOME=/usr/local/hadoop/hadoop-2.6.0
export PATH=$HADOOP_HOME/bin:$PATH
```

注意：红色部分需要根据你的 Hadoop 实际存放的目录确定。

修改完后，需要通过如下的命令使得新配置的环境变量起作用：

```
# source /etc/profile
```

## 2. Hadoop 格式化

配置好 Hadoop 后，需要格式化 Hadoop 的 HDFS 文件系统。

`cd` 到 `/usr/local/hadoop/hadoop-2.6.0/bin/` 目录下，然后执行如下命令：

```
# ./hdfs namenode -format
```

（说明，2.x 中使用的是 `hdfs` 命令格式化文件系统，而 Hadoop1.x 中使用的是 `hadoop` 命令）

### 3. Hadoop 启动

cd 到/usr/local/hadoop/hadoop-2.6.0/**sbin**/（此处是 **sbin**，不是 **bin**）目录下，然后执行如下命令启动服务：

```
./start-dfs.sh
```

```
./start-yarn.sh
```

启动后，在 **hadoop0** 主机上使用 **jps** 命令可以查看到如下几个进程：

23050 NameNode
23655 Jps
23400 ResourceManager
23234 SecondaryNameNode

在每一台 **Slave** 机器上使用 **jps** 命令可以查看到如下几个进程：

13525 Jps
13496 NodeManager
13364 DataNode

如果能查到进程都在运行，则说明系统已经安装成功。

另外可通过如下的方法查看系统的运行状态：

查看 HDFS: <http://192.168.0.100:50070/>

查看 ResourceManager: <http://192.168.0.100:8088/>

如果关闭 HDFS 或 YARN 的服务，则可执行如下命令：

```
stop-dfs.sh #停止 HDFS 服务
```

```
stop-yarn.sh #停止 YARN 服务
```