# Hadoop 伪分布操作的安装

本文档描述了如何设置和配置单节点Hadoop安装，以便您可以使用Hadoop MapReduce和Hadoop分布式文件系统(HDFS)快速执行简单的操作。

## 一、环境准备

在安装Hadoop之前需要在机器上提前安装JDK和机器SSH免密码登录

**1、安装JDK**

如果不知如何安装查看

http://note.youdao.com/noteshare?id=7156f51a2a763353828175c0db4e310b&sub=2767F793CFB8441FAEDABFEB287C1FA7

**2、SSH免密码登录**

如果不知如何安装查看

http://note.youdao.com/noteshare?id=b9db1e3e8a5b28f495c4e0b2ae830342&sub=438F4B9021E74319A68DF91819B8B001

## 二、下载安装包

在写本文时Hadoop最新的稳定版为hadoop-2.9.0，作为学习的目的则是用最新稳定版，下载链接为：

http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.9.0/hadoop-2.9.0.tar.gz

## 三、伪分布安装操作

**1、解压安装包**

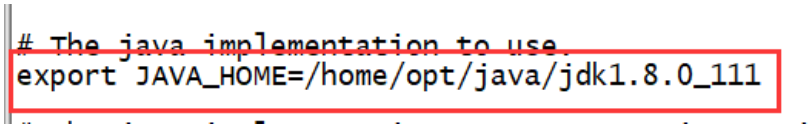
```
tar -zxf /home/tar/hadoop-2.9.0.tar.gz -C /home/opt/hadoop/
```

**2、编辑配置文件**

**2.1、etc/hadoop/hadoop-env.sh**

修改配置文件中的JAVA_HOME属性值为JDK安装路径

export JAVA_HOME=${JAVA_HOME}
为
export JAVA_HOME=/home/opt/java/jdk1.8.0_111




**2.2、etc/hadoop/core-site.xml**

在配置文件中配置日下两个属性

fs.defaultFS：指定HDFS namenode的地址
hadoop.tmp.dir：指定Hadoop的运行时产生的文件的存放目录

```
<configuration>
```

```
<property>
 <name>fs.defaultFS</name>
 <value>hdfs://myj01:9000</value>
</property>
<property>
 <name>hadoop.tmp.dir</name>
 <value>/home/opt/hadoop/hadoop-2.9.0/tmp</value>
</property>
</configuration>
```

```
<configuration>
        <property>
                <name>fs.defaultFS</name>
                <value>hdfs://myj01:9000</value>
        </property>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/home/opt/hadoop/hadoop-2.9.0/tmp</value>
        </property>
</configuration>
```

**2.3、etc/hadoop/hdfs-site.xml**

配置指定hdfs的保存数据副本的数量为1

```
<configuration>
 <property>
  <name>dfs.replication</name>
  <value>1</value>
 </property>
</configuration>
```

```
<configuration>
        <property>
                <name>dfs.replication</name>
                <value>1</value>
        </property>
</configuration>
```

**2.4、etc/hadoop/mapred-site.xml**

先复制配置文件mapred-site.xml.template为mapred-site.xml
cd /home/opt/hadoop/hadoop-2.9.0/etc/hadoop
cp mapred-site.xml.template  mapred-site.xml
在配置文件中设置mapreduce运行在yarn上

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

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

**2.5、etc/hadoop/yarn-site.xml**

yarn.nodemanager.aux-services：nodemanager获取数据的方式是shuffle
yarn.resourcemanager.hostname：指定yarn的ResourceManager的主机名

```
<configuration>
 <property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
 </property>
 <property>
  <name>yarn.resourcemanager.hostname</name>
  <value>myj01</value>
 </property>
</configuration>
```

```
<configuration>
        <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
        </property>
        <property>
                <name>yarn.resourcemanager.hostname</name>
                <value>myj01</value>
        </property>
</configuration>
[root@myj01 hadoop]#
```

# 四、配置Hadoop环境变量

通过上面的配置，Hadoop的基本配置工作已经完成，下面设置Hadoop的环境变量
vim /etc/profile
在文件尾部追加如下两行

```
export HADOOP_HOME=/home/opt/hadoop/hadoop-2.9.0
export PATH=$PATH:$HADOOP_HOME/bin
```

刷新配置文件

```
source /etc/profile
```

# 五、格式化HDFS并启动HDFS和yarn

## 1、格式化文件系统

执行如下命令

```
/home/opt/hadoop/hadoop-2.9.0/bin/hdfs namenode -format
```

命令执行结束后如果看到如下输出说明格式化成功

```
18/01/22 16:04:17 INFO util.GSet: Computing capacity for map NameNodeRetryCache
18/01/22 16:04:17 INFO util.GSet: VM type       = 64-bit
18/01/22 16:04:17 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 297.0 KB
18/01/22 16:04:17 INFO util.GSet: capacity      = 2^15 = 32768 entries
18/01/22 16:04:17 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1243767849-192.168.88.128-1516608257600
18/01/22 16:04:17 INFO common.Storage: Storage directory /home/opt/hadoop/hadoop-2.9.0/tmp/dfs/name has been successfully
 formatted.
18/01/22 16:04:17 INFO namenode.FSImageFormatProtobuf: Saving image file /home/opt/hadoop/hadoop-2.9.0/tmp/dfs/name/curre
nt/fsimage.ckpt_0000000000000000000 using no compression
18/01/22 16:04:17 INFO namenode.FSImageFormatProtobuf: Image file /home/opt/hadoop/hadoop-2.9.0/tmp/dfs/name/current/fsim
age.ckpt_0000000000000000000 of size 321 bytes saved in 0 seconds.
18/01/22 16:04:17 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
18/01/22 16:04:17 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at myj01/192.168.88.128
************************************************************/
```

### 2、启动 NameNode节点和DataNode节点

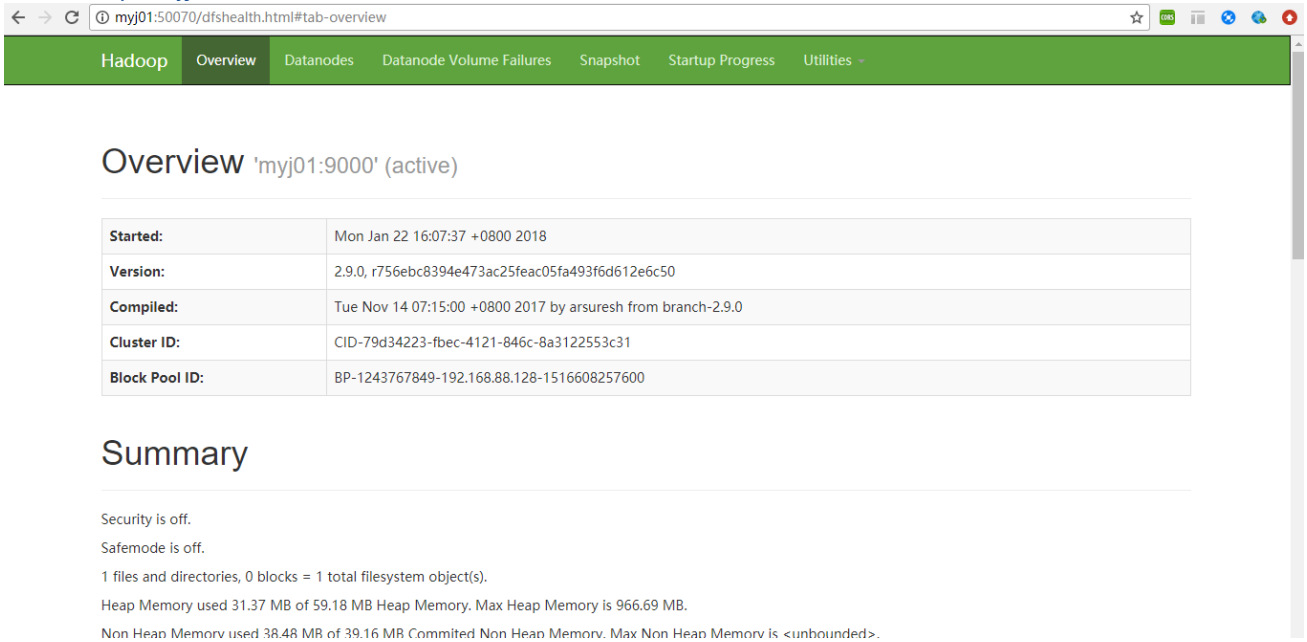
执行如下命令

/home/opt/hadoop/hadoop-2.9.0/sbin/start-dfs.sh

### 3、浏览器打开NameNode的Web界面

通过浏览器打开NameNode的Web界面接口，如果能打开则NameNode启动成功

http://myj01:50070/




### 4、创建HDFS目录并执行MapReduce作业

在HDFS上创建如下命令目录

/home/opt/hadoop/hadoop-2.9.0/bin/hdfs dfs -mkdir /user

/home/opt/hadoop/hadoop-2.9.0/bin/hdfs dfs -mkdir /user/test

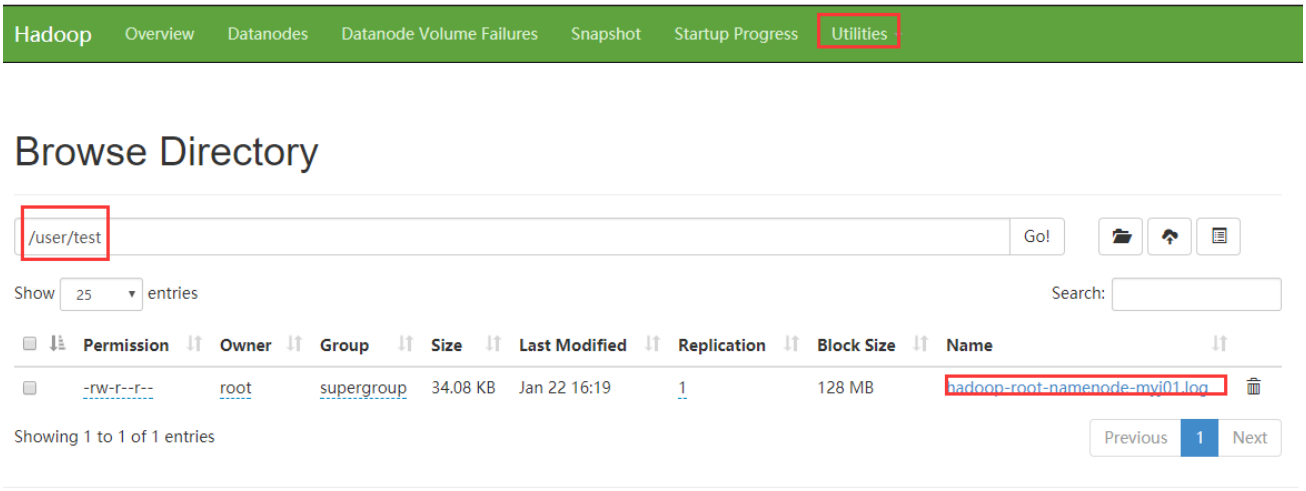
```
[root@myj01 ~]# /home/opt/hadoop/hadoop-2.9.0/bin/hdfs dfs -mkdir /user
[root@myj01 ~]# /home/opt/hadoop/hadoop-2.9.0/bin/hdfs dfs -mkdir /user/test
```

### 5、复制一个文件到HDFS

执行如下目录复制一个文件到HDFS

/home/opt/hadoop/hadoop-2.9.0/bin/hdfs dfs -put hadoop-root-namenode-myj01.log /user/test/hadoop-root-namenode-myj01.log

上传后可以在浏览器中看到：




### 6、启动 yarn

执行如下命令

/home/opt/hadoop/hadoop-2.9.0/sbin/start-yarn.sh

```
[root@myj01 ~]# /home/opt/hadoop/hadoop-2.9.0/sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/opt/hadoop/hadoop-2.9.0/logs/yarn-root-resourcemanager-myj01.out
localhost: starting nodemanager, logging to /home/opt/hadoop/hadoop-2.9.0/logs/yarn-root-nodemanager-myj01.out
[root@myj01 ~]# jps
2947 NameNode
3077 DataNode
4101 NodeManager
4377 Jps
3276 SecondaryNameNode
3997 ResourceManager
[root@myj01 ~]#
```

**到此Hadoop为分布式已成功安装。**