

HBase完全分布式安装指南

1. 集群规模(IP地址根据集群实际情况而定)

主机IP地址	主机名址
192.168.0.1	master
192.168.0.2	slave1
192.168.0.3	slave2

PS:如何设置CentOS主机名（如果你在前面没有设置好，供参考）

```
vi /etc/sysconfig/network
```

然后修改HOSTNAME属性，对应每台机器的主机名

2.环境搭建前期准备工作

2.1 java环境安装

参考Spark StandAlone安装文档中第三步中安装java部分（在每台机器上运行）

2.2 修改hosts文件

在前面创建虚拟机和联通网络的时候，每台虚拟机都有了自己的IP地址，不多说了。

1. 执行如下指令（在每个节点上）

```
vi /etc/hosts
```

2. 在文件后面添加如下部分(IP地址根据集群实际情况而定)

```
192.168.0.1    master
192.168.0.2    slave1
192.168.0.3    slave2
```

3. 测试结果

通过 ping + 主机名（不是IP地址）的方式监测host文件是否修改成功

2.3 ssh无密码登录设置

1. 主节点执行如下操作

```
ssh-keygen -t rsa
```

然后一直按回车（需要按多次），到命令执行结束

```
cd ~/.ssh  
cat id_rsa.pub >> authorized_keys  
chmod 600 authorized_keys
```

然后通过如下命令将公钥发送到各个从节点上

PS：此处操作会让你输入一次密码，是你发送的从节点root用户的密码。也可能会让你选择一次yes/no，选择yes。

```
ssh-copy-id root@slave1  
ssh-copy-id root@slave2
```

2. 无密码登录测试（在主节点上执行）

使用命令 `ssh + 主机名` 登录到对应的主机，如：

```
ssh master
```

PS:第一次输入的时候可能会让你选择yes/no，选择yes

连接上以后，可以通过hostname命令查看当前连接的主机名，如果无误，使用exit命令退出连接。

然后使用ssh连接到每台机器上，一定要使用主机名（而不是IP），连接成功后记得exit退出。

执行完这个步骤，表示基本环境搭建已经完成，建议把每个节点都重启一次。

2.4 永久关闭防火墙

在每个虚拟机上都必须执行如下操作

PS：这点一定要做，避免后期出现大问题

```
service iptables stop  
chkconfig iptables off
```

3. Hadoop环境搭建

HBase是基于Hadoop的分布式文件系统（HDFS）实现的，所以在搭建HBase前要先搭建好Hadoop环境

如果前面所有的过程按照教程走到现在，那么不需要做任何修改，将给出的hadoop安装文件，解压到每台机器的/home/cloud目录下即可直接跳转到启动步骤（步骤2）。

若和给出的文档有些许出入，按照一下规则进行调整。

1. 修改配置文件（每台机器相同操作，无论主从）

hadoop配置文件在hadoop目录下hadoop-2.6.4\etc\hadoop中

修改core-site.xml

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://master:9000</value>
    <!--这里用的是master，填写主节点主机名，端口号不要变-->
  </property>
  <property>
    <name>io.file.buffer.size</name>
    <value>131072</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/home/cloud/temp/hadoop-2.6.4</value>
    <description>A base for other temporary directories.</description>
    <!--这个目录最好也不要动-->
  </property>
</configuration>
```

修改hadoop-env.sh

```
export JAVA_HOME=/home/cloud/jdk1.8.0_101
//上面是文件第25行，路径填写你配置的java_home
```

修改hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>master:9001</value>
    <!--这里用的是master，填写主节点主机名，端口号不要变-->
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/home/cloud/hadoop-2.6.4/dfs/name</value>
    <!--这个目录最好也不要动-->
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/home/cloud/hadoop-2.6.4/dfs/data</value>
    <!--这个目录最好也不要动-->
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.webhdfs.enabled</name>
    <value>true</value>
  </property>
</configuration>
```

修改mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>master:10020</value>
    <!-- 这里用的是master，填写主节点主机名，端口号不要变-->
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>master:50030</value>
    <!-- 这里用的是master，填写主节点主机名，端口号不要变-->
  </property>
  <property>
    <name>mapreduce.reduce.memory.mb</name>
    <value>4096</value>
  </property>
</configuration>
```

修改slaves (这个文件表示从节点，只填写主机名)

```
slave1
slave2
```

修改yarn-site.xml

```
export JAVA_HOME=/home/cloud/jdk1.8.0_101
//上面是文件第23行，路径填写你配置的java_home
```

修改yarn-site.xml

```
<configuration>

<!-- 这个文件中设置了很多端口，不要改端口，如果主机名不是master，改成对应的名字 -->
  <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>

  <property>
    <name>yarn.resourcemanager.address</name>
    <value>master:8032</value>
  </property>
```

```
<property>
  <name>yarn.resourcemanager.scheduler.address</name>
  <value>master:8030</value>
</property>

<property>
  <name>yarn.resourcemanager.resource-tracker.address</name>
  <value>master:8031</value>
</property>

<property>
  <name>yarn.resourcemanager.admin.address</name>
  <value>master:8033</value>
</property>

<property>
  <name>yarn.resourcemanager.webapp.address</name>
  <value>master:8088</value>
</property>
</configuration>
```

2. 配置完成后（记得所有虚拟机都要改），启动hadoop

如果是第一次启动Hadoop，需要进行一次格式化主节点操作，首先进入hadoop-2.6.4目录，然后执行

```
bin/hadoop namenode -format
```

以后再启动hadoop，不要格式化主节点了，会丢失数据的。

在主节点中进入hadoop-2.6.4目录，然后按照顺序输入如下命令

```
sbin/start-dfs.sh
sbin/start-yarn.sh
```

然后在主节点上打开浏览器，地址栏输入master:50070查看详情，若安装正确，出现如下界面：

Overview 'hmaster:8020' (active)

Started:	Mon Jul 11 13:05:21 CST 2016
Version:	2.6.4-5082c73637550b0b7e115f9d25ed7fac69937e6
Compiled:	2016-02-12T09:45Z by Jenkins from [detached from 5082c73]
Cluster ID:	CID-7cd9feb0-56c9-4892-b7cf-33dd32a6a2c3
Block Pool ID:	BP-1238311174-192.168.0.41-1498617369068

Summary

Security is off.
 Safemode is off.
 6123 files and directories, 3306 blocks = 6419 total filesystem object(s).
 Heap Memory used 41.08 GB of 118.5 MB Heap Memory. Max Heap Memory is 189 MB.
 Non Heap Memory used 63.03 MB of 64.09 MB Committed Non Heap Memory. Max Non Heap Memory is 1 B.

Configured Capacity:	539.43 GB
DFS Used:	314.01 GB
Non DFS Used:	53.73 GB
DFS Remaining:	171.68 GB
DFS Used%:	58.21%
DFS Remaining%:	41.79%
Block Pool Used:	314.01 GB
Block Pool Used%:	58.21%
DataNodes usage% (Min/Median/Max/stdDev):	43.68% / 64.23% / 68.28% / 2.37%
Live Nodes	4 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	14
Number of Blocks Pending Deletion	0
Block Deletion Start Time	2016/7/31 上午 11:12:52

NameNode Journal Status

Current transaction ID: 3693149

Journal Manager	State
FileJournalManager[root="/home/yun/hadoop-2.6.4/dfs/name"]	EditLogFileOutputStream(/home/yun/hadoop-2.6.4/dfs/name/current/cdbs_nprogress2_0000000003693149)

在Live Nodes项中，你的应该显示数量为2，点击进去，可以看到对应两个从节点的主机名。

3. 关闭Hadoop命令

在主节点中进入hadoop-2.6.4目录，然后按照顺序输入如下命令

PS:如果已经安装了HBase，那么千万记住，在关闭Hadoop之前，必须保证HBase已经正常关闭了

```
sbin/stop-yarn.sh
sbin/stop-dfs.sh
```

4. HBase环境搭建

在进行HBase环境搭建前，必须确保Hadoop已经搭建成功，并可以运行。

同Hadoop搭建一样，如果你是按照文档一步步走下来，都没有问题。那么就可以将给你的hbase文档解压到/home/cloud（所有虚拟机）下，直接跳到步骤2，启动HBase

1. 修改HBase配置文件

HBase配置文件在hbase/conf目录下

修改hbase-env.sh

```
export JAVA_HOME=/home/cloud/jdk1.8.0_101
//以上在文件第27行，修改为java_home目录

export HBASE_MANAGES_ZK=true
//以上在文件第124行，不要动，表示启动自带的zookeeper
```

修改hbase-site.xml

```
<configuration>
  <property>
    <name>hbase.rootdir</name>
    <value>hdfs://master:9000/hbase</value>
    <!-- 端口号不要变，master修改为你的主机名-->
  </property>
  <property>
    <name>hbase.cluster.distributed</name>
    <value>true</value>
  </property>
  <property>
    <name>hbase.zookeeper.quorum</name>
    <value>master,slave1,slave2</value>
    <!-- 这里是zookeeper集群，填写三个节点的主机名-->
  </property>
  <property>
    <name>hbase.tmp.dir</name>
    <value>file:/home/cloud/hbase/tmp</value>
  </property>
  <property>
    <name>hbase.master.port</name>
    <value>60000</value>
  </property>
  <property>
    <name>hbase.master.info.port</name>
    <value>60020</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

修改regionservers文件（填写从节点主机名）

```
slave1
slave2
```

2. 启动HBase

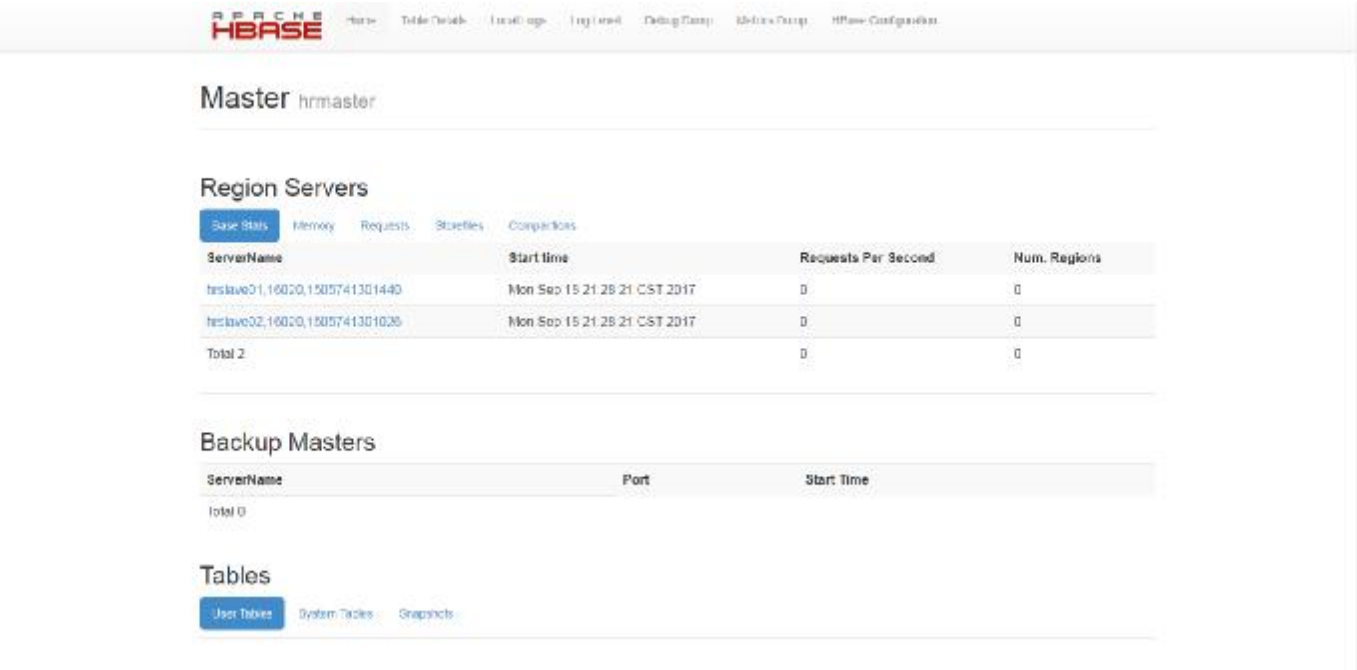
在启动hbase之前，确保Hadoop已经正常启动了，要不然是不会启动成功的

PS：最好在Hadoop启动两分钟以后再启动HBase

进入hbase目录，然后执行如下命令

```
bin/start-hbase.sh
```

在主节点浏览器中输入master:60020,应出现如下界面



Region Servers中会出现两个从节点的主机名，如果第一次打开没出现，刷新一下就好了。

3. 关闭HBase

进入hbase目录，然后执行如下命令

```
bin/stop-hbase.sh
```