

# Hadoop3.0 单机伪分布集群安装

## 准备环境：

- 1 台 linux 虚拟机(Centos6.x 64 位)

ip 为: 192.168.111.100

hostname 为: hadoop100

## 虚拟机基本环境配置：

集群用到的所有机器都参考此流程进行配置。

- 静态 ip 配置

# ip 配置信息仅供参考，此处使用的是 nat 网络模式

```
[root@hadoop100 ~]# vi /etc/sysconfig/network-scripts/ifcfg-eth0
```

```
BOOTPROTO=static
```

```
IPADDR=192.168.111.100
```

```
GATEWAY=192.168.111.2
```

```
NETMASK=255.255.255.0
```

```
DNS1=192.168.111.2
```

# 修改好以后重启网卡，只要看到全部都是 OK 即表示配置成功

```
[root@hadoop100 ~]# service network restart
```

```
Shutting down interface eth0: [ OK ]
```

```
Shutting down loopback interface: [ OK ]
```

```
Bringing up loopback interface: [ OK ]
```

```
Bringing up interface eth0: Determining if ip address 192.168.111.100 is already in use for device eth0...
```

```
[ OK ]
```

- 主机名设置

建议把此机器的主机名设置为 hadoop100

主机名的设置可以分为临时设置和永久设置。

临时设置可以立刻生效，但是不能永久保存，虚拟机重启后就会失效，所以还需要结合永久设置。

# 临时设置主机名

```
[root@hadoop100 ~]# hostname hadoop100
```

# 永久设置主机名

```
[root@hadoop100 ~]# vi /etc/sysconfig/network
```

```
NETWORKING=yes
```

```
HOSTNAME=hadoop100
```

- hosts 文件配置

修改/etc/hosts 文件，把 ip 和 hostname 的映射关系配置进去，此处的 ip 需要和你设置的虚拟机实际 ip 相同

```
[root@hadoop100 ~]# vi /etc/hosts
192.168.111.100 hadoop100
```

- 关闭防火墙

防火墙的关闭也分为临时关闭和永久关闭。

临时关闭可以立刻生效，但是不能永久保存，虚拟机重启后就会失效，所以还需要结合永久设置。

# 临时关闭防火墙

```
[root@hadoop100 ~]# service iptables stop
iptables: Setting chains to policy ACCEPT: filter          [ OK ]
iptables: Flushing firewall rules:                        [ OK ]
iptables: Unloading modules:                              [ OK ]
```

# 永久关闭防火墙

```
[root@hadoop100 soft]# chkconfig iptables off
```

- ssh 免密码登陆

配置 ssh 免密码登陆，在这里先实现免密码登陆本机

# 生成公钥

# 注意：执行这个命令以后，需要连续按 4 次回车键回到 linux 命令行才表示这个操作执行结束，在按回车的时候不需要输入任何内容。

```
[root@hadoop100 ~]# ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /root/.ssh/id_rsa.
Your public key has been saved in /root/.ssh/id_rsa.pub.
The key fingerprint is:
20:d0:89:24:29:1e:c8:50:b3:fb:e0:57:76:94:0c:ea root@hadoop100
```

The key's randomart image is:

```
+--[ RSA 2048]-----+
|*+=0 ..          |
|++.+0. o .       |
|o o o . +        |
|. o . o          |
| o E o S         |
|. o o .          |
|. o              |
|.               |
```

```
|
|
+-----+
# 向本机复制公钥
[root@hadoop100 ~]# cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

# 执行 ssh 命令验证免密码登陆是否配置成功，第一次验证的时候会提示输入 yes 或 no，直接输入 yes 即可，如果输入了 yes 之后按回车可以进入就表示免密码登陆配置成功。
[root@hadoop100 ~]# ssh hadoop100
The authenticity of host 'hadoop100(192.168.111.100)' can't be established.
RSA key fingerprint is 64:87:b9:da:46:c1:85:54:49:6d:65:4a:0f:49:8f:58.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'hadoop100' (RSA) to the list of known hosts.
Last login: Thu Jul 25 09:11:46 2019 from hadoop100
[root@hadoop100 ~]#
```

### ● jdk 安装

注意：在这里建议安装 jdk1.8 版本，jdk-8u181-linux-x64.tar.gz

```
# 上传 jdk 安装包到 linux 的 /data/soft 目录，如果此目录不存在则提前创建
[root@hadoop100 ~]# mkdir -p /data/soft/
[root@hadoop100 ~]# cd /data/soft/

# 解压 jdk
[root@hadoop100 soft]# tar -zxvf jdk-8u181-linux-x64.tar.gz

# 重命名
[root@hadoop100 soft]# mv jdk1.8.0_181/ jdk1.8

# 配置 java 的环境变量，修改/etc/profile 文件，在文件末尾增加下面三行配置
[root@hadoop100 soft]# vi /etc/profile
export JAVA_HOME=/data/soft/jdk1.8
export HADOOP_HOME=/data/soft/hadoop-3.2.0
export PATH=.:$JAVA_HOME/bin:$HADOOP_HOME/bin:$PATH

# 验证环境变量配置是否成功，执行以下命令能看到输入信息即说明配置成功
[root@hadoop100 soft]# java -version
java version "1.8.0_181"
Java(TM) SE Runtime Environment (build 1.8.0_181-b13)
Java HotSpot(TM) 64-Bit Server VM (build 25.181-b13, mixed mode)
```

## Hadoop 伪分布集群安装：

下面开始在 1 台 linux 虚拟机上开始安装 Hadoop3 伪分布环境

在这里我们使用 hadoop3.2.0 版本，hadoop-3.2.0.tar.gz

1：把 hadoop-3.2.0.tar.gz 安装包上传到 linux 机器的 /data/soft 目录下

2：解压 hadoop 安装包

```
[root@hadoop100 ~]# cd /data/soft
```

```
[root@hadoop100 soft]# tar -zxvf hadoop-3.2.0.tar.gz
```

3: 修改 hadoop 相关配置文件

# 进入配置文件所在目录

```
[root@hadoop100 soft]# cd hadoop-3.2.0/etc/hadoop/
```

# 首先修改 hadoop-env.sh 文件, 增加环境变量信息

```
[root@hadoop100 hadoop]# vi hadoop-env.sh
```

```
export JAVA_HOME=/data/soft/jdk1.8
```

```
export HADOOP_LOG_DIR=/data/hadoop_repo/logs/hadoop
```

# 修改 core-site.xml 文件, 注意 fs.defaultFS 属性中的主机名需要和你配置的主机名保持一致

```
[root@hadoop100 hadoop]# vi core-site.xml
```

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop100:9000</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/data/hadoop_repo</value>
  </property>
</configuration>
```

# 修改 hdfs-site.xml 文件, 把 hdfs 中文件副本的数量设置为 1, 因为现在伪分布集群只有一个节点

```
[root@hadoop100 hadoop]# vi hdfs-site.xml
```

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

# 修改 mapred-site.xml, 设置 mapreduce 使用的资源调度框架

```
[root@hadoop100 hadoop]# vi mapred-site.xml
```

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

# 修改 yarn-site.xml, 设置 yarn 上支持运行的服务和环境变量白名单

```
[root@hadoop100 hadoop]# vi yarn-site.xml
```

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
```

```

        <value>mapreduce_shuffle</value>
    </property>
</property>
    <name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>

```

#### 4: 格式化 namenode

```

[root@hadoop100 hadoop]# cd /data/soft/hadoop-3.2.0
[root@hadoop100 hadoop-3.2.0]# bin/hdfs namenode -format
# 如果在后面的日志信息中能看到这一行，则说明 namenode 格式化成功。
common.Storage: Storage directory /data/hadoop_repo/dfs/name has been successfully
formatted.

```

#### 5: 启动 hadoop 集群

```

[root@hadoop100 hadoop-3.2.0]# sbin/start-all.sh
ERROR: Attempting to operate on hdfs namenode as root
ERROR: but there is no HDFS_NAMENODE_USER defined. Aborting operation.
Starting datanodes
ERROR: Attempting to operate on hdfs datanode as root
ERROR: but there is no HDFS_DATANODE_USER defined. Aborting operation.
Starting secondary namenodes [hadoop100]
ERROR: Attempting to operate on hdfs secondarynamenode as root
ERROR: but there is no HDFS_SECONDARYNAMENODE_USER defined. Aborting operation.
2019-07-25 10:04:25,993 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Starting resourcemanager
ERROR: Attempting to operate on yarn resourcemanager as root
ERROR: but there is no YARN_RESOURCEMANAGER_USER defined. Aborting operation.
Starting nodemanagers
ERROR: Attempting to operate on yarn nodemanager as root
ERROR: but there is no YARN_NODEMANAGER_USER defined. Aborting operation.
发现在启动的时候报错，提示缺少 HDFS 和 YARN 的一些用户信息。

```

解决方案如下：

# 修改 start-dfs.sh, stop-dfs.sh 这两个脚本文件，在文件前面增加如下内容

```

[root@hadoop100 hadoop]# cd /data/soft/hadoop-3.2.0/sbin
[root@hadoop100 sbin]# vi start-dfs.sh
HDFS_DATANODE_USER=root
HDFS_DATANODE_SECURE_USER=hdfs
HDFS_NAMENODE_USER=root
HDFS_SECONDARYNAMENODE_USER=root

```

```
[root@hadoop100 sbin]# vi stop-dfs.sh
HDFS_DATANODE_USER=root
HDFS_DATANODE_SECURE_USER=hdfs
HDFS_NAMENODE_USER=root
HDFS_SECONDARYNAMENODE_USER=root
```

# 修改 start-yarn.sh, stop-yarn.sh 这两个脚本文件, 在文件前面增加如下内容

```
[root@hadoop100 sbin]# vi start-yarn.sh
YARN_RESOURCEMANAGER_USER=root
HADOOP_SECURE_DN_USER=yarn
YARN_NODEMANAGER_USER=root
```

```
[root@hadoop100 sbin]# vi stop-yarn.sh
YARN_RESOURCEMANAGER_USER=root
HADOOP_SECURE_DN_USER=yarn
YARN_NODEMANAGER_USER=root
```

# 再重新启动集群

```
[root@hadoop100 sbin]# ./start-all.sh
Starting namenodes on [hadoop100]
Starting datanodes
Starting secondary namenodes [hadoop100]
Starting resourcemanager
Starting nodemanagers
```

#### 6: 验证集群进程信息

# 执行 jps 命令可以查看集群的进程信息, 抛出 Jps 这个进程之外还需要有 5 个进程才说明集群是正常启动的

```
[root@hadoop100 sbin]# jps
2882 ResourceManager
2420 DataNode
3365 Jps
2619 SecondaryNameNode
2315 NameNode
2988 NodeManager
```

# 还可以通过 webui 界面来验证集群服务是否正常

hdfs webui 界面: <http://192.168.111.100:9870>  
yarn webui 界面: <http://192.168.111.100:8088>

#### 7: 停止集群

如果修改了集群的配置文件或者是其它原因要停止集群, 可以使用下面命令。

```
[root@hadoop100 sbin]# ./stop-all.sh
```

## Hadoop3.0 分布式集群安装

### 准备环境：

- 3 台 linux 虚拟机(Centos6.x 64 位)

ip 为：

192.168.111.100

192.168.111.101

192.168.111.102

对应的 hostname 为：

hadoop100

hadoop101

hadoop102

### 虚拟机基本环境配置：

这三台机器的基本配置参考 Hadoop3.0 单机伪分布安装章节中的虚拟机基本环境配置部分。

注意：除了基本配置之外还需要实现主节点能够免密码登陆到所有从节点，因为从节点上面的进程是由主节点通过 ssh 远程启动的。

在 hadoop100 这台机器上执行下面命令，可以实现主节点免密码登陆到所有从节点。

```
[root@hadoop100 ~]# ssh-copy-id -i hadoop101
```

```
[root@hadoop100 ~]# ssh-copy-id -i hadoop102
```

在 hadoop100 机器上使用 ssh 命令确认集群免密码登陆是否设置成功，如果可以免密码登陆进去就说明是没有问题的。

```
[root@hadoop100 ~]# ssh hadoop101
```

```
[root@hadoop100 ~]# ssh hadoop101
```

### Hadoop 分布式集群安装：

注意：在这里我们用到了前面安装伪分布集群的机器 hadoop100，所以需要先把之前的伪分布集群停掉，然后删除/data 目录下的 hadoop\_repo 目录和/data/soft 目录下的 hadoop-3.2.0

目录，恢复此机器的环境。

下面开始安装 **hadoop** 分布式集群。

首先在 **hadoop100** 节点上安装。

1: 把 **hadoop-3.2.0.tar.gz** 安装包上传到 linux 机器的 **/data/soft** 目录下

2: 解压 **hadoop** 安装包

```
[root@hadoop100 ~]# cd /data/soft
```

```
[root@hadoop100 soft]# tar -zxvf hadoop-3.2.0.tar.gz
```

3: 修改 **hadoop** 相关配置文件

# 进入配置文件所在目录

```
[root@hadoop100 soft]# cd hadoop-3.2.0/etc/hadoop/
```

# 首先修改 **hadoop-env.sh** 文件，增加环境变量信息

```
[root@hadoop100 hadoop]# vi hadoop-env.sh
```

```
export JAVA_HOME=/data/soft/jdk1.8
```

```
export HADOOP_LOG_DIR=/data/hadoop_repo/logs/hadoop
```

# 修改 **core-site.xml** 文件，注意 **fs.defaultFS** 属性中的主机名需要和你配置的主机名保持一致

```
[root@hadoop100 hadoop]# vi core-site.xml
```

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop100:9000</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/data/hadoop_repo</value>
  </property>
</configuration>
```

# 修改 **hdfs-site.xml** 文件，把 **hdfs** 中文件副本的数量设置为 1，因为现在伪分布集群只有一个节点

```
[root@hadoop100 hadoop]# vi hdfs-site.xml
```

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>hadoop100:50090</value>
  </property>
</configuration>
```

# 修改 **mapred-site.xml**，设置 **mapreduce** 使用的资源调度框架

```
[root@hadoop100 hadoop]# vi mapred-site.xml
```

```
<configuration>
```



```

    <property>
      <name>mapreduce.framework.name</name>
      <value>yarn</value>
    </property>
  </configuration>
# 修改 yarn-site.xml, 设置 yarn 上支持运行的服务和环境变量白名单
[root@hadoop100 hadoop]# vi yarn-site.xml
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CL
ASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
  </property>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>hadoop100</value>
  </property>
</configuration>

# 修改 workers 文件, 增加所有从节点的主机名, 一个一行
[root@hadoop100 hadoop]# vi workers
hadoop101
hadoop102

# 修改启动脚本
# 修改 start-dfs.sh, stop-dfs.sh 这两个脚本文件, 在文件前面增加如下内容
[root@hadoop100 hadoop]# cd /data/soft/hadoop-3.2.0/sbin
[root@hadoop100 sbin]# vi start-dfs.sh
HDFS_DATANODE_USER=root
HDFS_DATANODE_SECURE_USER=hdfs
HDFS_NAMENODE_USER=root
HDFS_SECONDARYNAMENODE_USER=root

[root@hadoop100 sbin]# vi stop-dfs.sh
HDFS_DATANODE_USER=root
HDFS_DATANODE_SECURE_USER=hdfs
HDFS_NAMENODE_USER=root
HDFS_SECONDARYNAMENODE_USER=root

# 修改 start-yarn.sh, stop-yarn.sh 这两个脚本文件, 在文件前面增加如下内容

```

```
[root@hadoop100 sbin]# vi start-yarn.sh
YARN_RESOURCEMANAGER_USER=root
HADOOP_SECURE_DN_USER=yarn
YARN_NODEMANAGER_USER=root
```

```
[root@hadoop100 sbin]# vi stop-yarn.sh
YARN_RESOURCEMANAGER_USER=root
HADOOP_SECURE_DN_USER=yarn
YARN_NODEMANAGER_USER=root
```

4: 把 hadoop100 节点上修改好配置的安装包拷贝到其他两个从节点

```
[root@hadoop100 hadoop]# cd /data/soft/
[root@hadoop100 soft]# scp -rq hadoop-3.2.0.tar.gz  hadoop101:/data/soft/
[root@hadoop100 soft]# scp -rq hadoop-3.2.0.tar.gz  hadoop102:/data/soft/
```

5: 格式化 namenode

```
[root@hadoop100 hadoop]# cd /data/soft/hadoop-3.2.0
[root@hadoop100 hadoop-3.2.0]# bin/hdfs namenode -format
# 如果在后面的日志信息中能看到这一行，则说明 namenode 格式化成功。
common.Storage: Storage directory /data/hadoop_repo/dfs/name has been successfully
formatted.
```

6: 启动集群，在 hadoop100 节点上执行下面命令

```
[root@hadoop100 hadoop-3.2.0]# sbin/start-all.sh
Starting namenodes on [hadoop100]
Starting datanodes
Starting secondary namenodes [hadoop100]
Starting resourcemanager
Starting nodemanagers
```

7: 验证集群

分别在 3 台机器上执行 jps 命令，进程信息如下所示：

# 在 hadoop100 节点执行

```
[root@hadoop100 hadoop-3.2.0]# jps
8802 ResourceManager
8572 SecondaryNameNode
8334 NameNode
```

# 在 hadoop101 节点执行

```
[root@hadoop101 hadoop-3.2.0]# jps
2229 NodeManager
2124 DataNode
```

```
# 在 hadoop102 节点执行
[root@hadoop102 hadoop-3.2.0]# jps
1857 DataNode
1963 NodeManager
```

至此，hadoop 分布式集群安装成功！