

<http://dxmall.com.cn>

Hadoop集群配置

NOTES

- 进行下一步之前要先检查上一步的目标是否达成
- 以下操作未特殊说明的均在主节点进行
- 参考文档仅作参考，根据实际情况进行配置
- 进行删改操作之前建议对配置文件进行备份
- 设计其他详细命令及操作可参考其他文档

所有服务器在同一网段下（能ping通）

ping不通可能是不在同一网段下，或者防火墙等问题

- ☐ 目标：所有服务器在同一网段下（能ping通）

修改hostname

```
vim /etc/hostname
```

给各台服务器起一个名字。例 ubuntu1 ubuntu2 ...

修改完文件中的hostname之后要重启才生效，如果不想重启可以使用

```
hostname ubuntu1
```

- ☐ 目标：输入hostname 显示修改过后的hostname（对每台服务器都要进行的操作，每台服务器不同的名字）

修改hosts文件

```
vim /etc/hosts
```

```
root@ubuntu1:/etc
127.0.0.1      localhost
127.0.1.1      ubuntu
192.168.1.136  ubuntu1
192.168.1.117  ubuntu2
192.168.1.137  ubuntu3

# The following lines are desirable for IPv6 capable hosts
::1          ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
~
~
~
~
~
~
~
~
~
~/etc/hosts" 12L, 290C
```

```
scp /etc/hosts ubuntu2:/etc
```

关于hosts文件功能见百度

```
ssh-keygen
//一直回车

ssh-copy-id ubuntu1 //发给自己
ssh-copy-id ubuntu2
ssh-copy-id ubuntu3
```

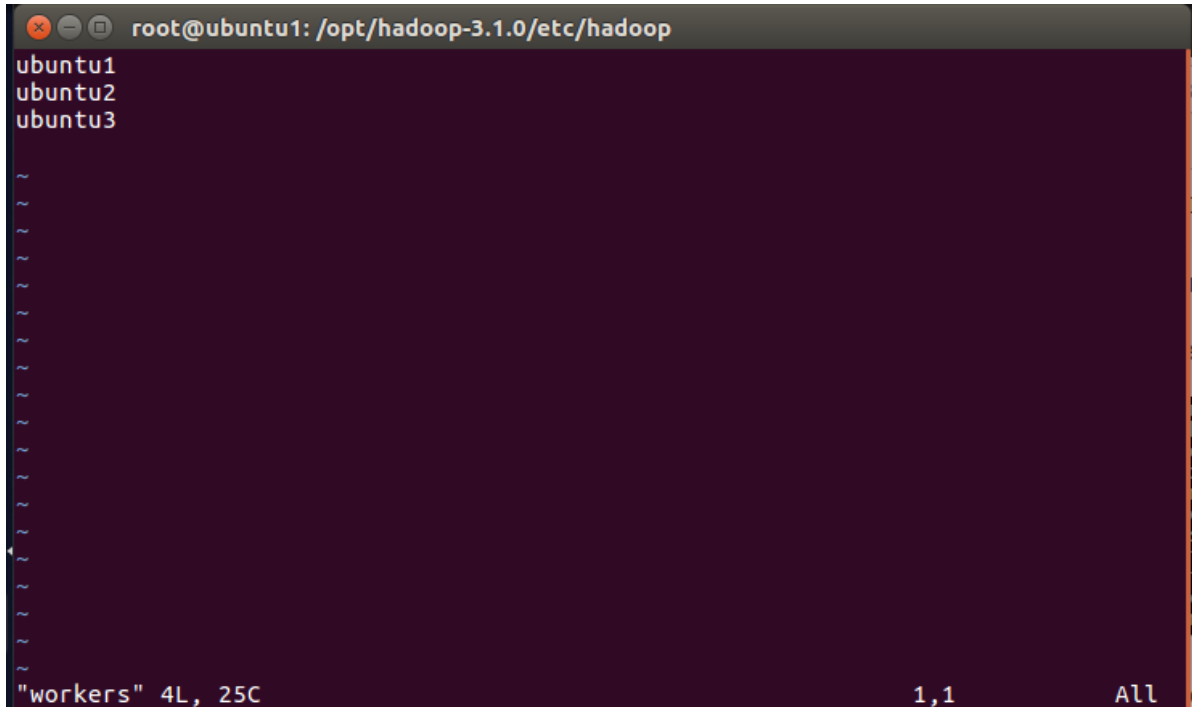
参考: https://blog.csdn.net/qg_25615395/article/details/89018208

```
cd /opt/hadoop-3.1.0/etc/hadoop/  
vim core-site.xml  
vim hdfs-site.xml  
vim mapred-site.xml  
vim yarn-site.xml
```

注意根据自己集群的情况修改配置文件，不要照抄。

把节点添加到集群中

```
vim workers
//以前叫slaves
```



Hadoop根据这个文件来确定把谁加到集群中

修改好主节点的配置文件之后，使用scp分发给其他节点

```
scp -r /opt/hadoop-3.1.0/etc/hadoop/ ubuntu2:/opt/hadoop-3.1.0/etc
```

同理发送给3

☐ 目标：检查各个节点配置文件是否修改正确

参考：https://blog.csdn.net/code_online/article/details/80178032

清理

第一次配置才要format，以后有数据时不要轻易format

清空 /opt/hadoop-3.1.0 目录下的tmp 和logs文件夹（每个节点都清空）

对主节点执行

```
hadoop namenode -format

//format时要关闭hadoop服务
```

参考：<https://blog.csdn.net/yijichangkong/article/details/45438931>

启动

在主节点执行

```
source /etc/profile
start-all.sh
```

查看集群状态

```
hdfs dfsadmin -report
```

如果启动正常就应该出现alive nodes(n) n为集群数量，以及各个节点的状态

```
-----
Live datanodes (3):

Name: 192.168.1.117:9866 (ubuntu2)
Hostname: ubuntu2
Decommission Status : Normal
Configured Capacity: 59050815488 (55.00 GB)
DFS Used: 32768 (32 KB)
Non DFS Used: 19114082304 (17.80 GB)
DFS Remaining: 36913500160 (34.38 GB)
DFS Used%: 0.00%
```

访问ubuntu1:50070查看集群状态

未启动成功见问题及解决

其它

集群配置的最终状态

1. 主节点可以免密登录其他节点（ssh时不用输入密码）
2. 每个节点有相同的hosts文件，包含所有节点的IP->别名之间的映射。
3. 每个节点的hadoop文件夹下的配置文件相同
4. 各个节点的logs和tmp文件夹都清空
5. 对主节点进行format

```
jps
查看相关进程
```

如果未启动应该是只有一个JPS进程

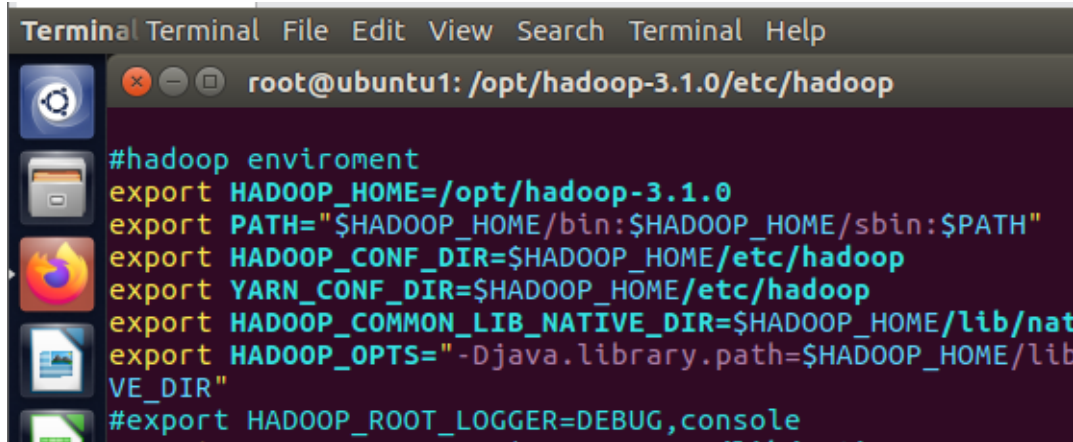
如果启动成功各个节点都应出现相应的进程如NameNode、DataNode...

例：

```
root@ubuntu1:/opt/hadoop-3.1.0/etc/hadoop# jps
5328 NameNode
6019 ResourceManager
6631 Jps
5735 SecondaryNameNode
5483 DataNode
6174 NodeManager
root@ubuntu1:/opt/hadoop-3.1.0/etc/hadoop#
```

```
Last login: Sat Nov 14 19:33:48 2020 from 192.168.1.136
root@ubuntu3:~# jps
3970 NodeManager
3814 DataNode
4154 Jps
root@ubuntu3:~#
```

1. 可以在一台服务器上配置好后，把整个hadoop-3.1.0文件夹scp给其他节点
2. workers文件是Hadoop决定把谁加到集群中
3. 环境变量要注意（本实验环境变量原来已经配置好，只做简单描述）
 1. Linux系统的环境变量都存在 /etc/profile中
 2. 例如Hadoop的安装目录



```
Terminal Terminal File Edit View Search Terminal Help
root@ubuntu1: /opt/hadoop-3.1.0/etc/hadoop

#hadoop enviroment
export HADOOP_HOME=/opt/hadoop-3.1.0
export PATH="$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH"
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export YARN_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
#export HADOOP_ROOT_LOGGER=DEBUG,console
```

3. 包括Java的环境变量

问题及解决

- 如果出现各个节点都启动起来了但是report出现connection什么什么错误
 1. stop-all.sh 使用jps看进程是否都结束了
 2. 清空tmp和logs文件夹
 3. 在主节点执行 hdfs datanode -format
- 涉及到修改文件时最好用tab补全，避免打错导致新建一个文件。（vim命令后面的文件如果不存在默认新建一个）这样所有的操作都在新建的文件上进行，导致错误。
- 在每次启动失败之后，要stop-all.sh，然后清空logs和tmp文件夹，以及格式化节点之后再尝试启动。
- 出现命令未找到的错误执行

```
source /etc/profile
```

参考文档：

https://blog.csdn.net/code_online/article/details/80178032

https://blog.csdn.net/qg_16069927/article/details/82954965

<https://blog.csdn.net/yjjichangkong/article/details/45438931>

免密登录：https://blog.csdn.net/qg_25615395/article/details/89018208?utm_medium=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-4.channel_param&depth_1-utm_source=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-4.channel_param

一键免密脚本：<https://itbook.com/article/1550832877722>

v2:

- 更新免密登录
- 其它错误修改