Hadoop full搭建

1. 基础设施：
2. 配置网络

vi /etc/sysconfig/network-scripts/ifcfg-eth0

DEVICE=eth0

TYPE=Ethernet

ONBOOT=yes

NM\_CONTROLLED=yes

BOOTPROTO=static

IPADDR=192.168.244.101

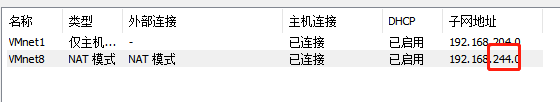
NETMASK=255.255.255.0

GATEWAY=192.168.244.2

DNS1=223.5.5.5

DNS2=114.114.114.114

红色数字和下图中vmware虚拟网络编辑中的VMnet8的第三个网段一样就行



重启网络: service network restart

1. 设置主机名

vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=node01

1. 设置hosts文件

vi /etc/hosts

192.168.244.101 node01

192.168.244.102 node02

192.168.244.103 node03

192.168.244.104 node04

1. 关闭防火墙

关闭防火墙: service iptables stop

设置开机关闭防火墙: chkconfig iptables off

1. 时间同步

安装ntp: yum ntp -y

vi /etc/ntp.conf

server ntp1.aliyum.com

启动时间同步: service ntpd start

设置开机启动时间同步: chkconfig ntpd on

1. 安装JDK

有一些软件只认/usr/java/default

通过rpm安装JDK: rpm -i jdk-8u181-linux-x64.rpm

im /etc/profile

export JAVA\_HOME=/usr/java/default

export PATH=$PATH:$JAVA\_HOME/bin

resource /etc/profile

1. ssh免密

生成密钥: ssh-keygen -t dsa -P '' -f ~/.ssh/id\_dsa

将公钥内容放到自己的authorized\_keys文件中:

cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys

分发id\_dsa.pub文件给别的节点并重命名为node01.pub:

scp ./id\_dsa.pub [root@node02:`pwd`/node01.pub](mailto:root@node02:`pwd`/node01.pub)

scp ./id\_dsa.pub [root@node03:`pwd`/node01.pub](mailto:root@node02:`pwd`/node01.pub)

scp ./id\_dsa.pub [root@node04:`pwd`/node01.pub](mailto:root@node04:`pwd`/node01.pub)

到node02-04节点上将node01.pub内容放到各自的authorized\_keys文件中:

cat ~/.ssh/node01.pub >> ~/.ssh/authorized\_keys

1. Hadoop配置
2. 规划路径：

mkdir /opt/bigdata

tar xzvf /root/soft/hadoop-2.6.5.tar.gz

mv hadoop-2.6.5 /opt/bigdata/

修改hadoop环境变量

vi /etc/profile

export HADOOP\_HOME=/opt/bigdata/hadoop-2.6.5

export PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

source /etc/profile

1. 配置hadoop的角色：

cd $HADOOP\_HOME/etc/hadoop

必须给hadoop配置javahome要不ssh过去找不到

vi hadoop-env.sh

export JAVA\_HOME=/usr/java/default

给出NN角色在哪里启动

vi core-site.xml

<property>

<name>fs.defaultFS</name>

<value>hdfs://node01:9000</value>

</property>

vi hdfs-site.xml

配置hdfs 副本数为2.

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>/var/bigdata/hadoop/full/dfs/name</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>/var/bigdata/hadoop/full/dfs/data</value>

</property>

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>node02:50090</value>

</property>

<property>

<name>dfs.namenode.checkpoint.dir</name>

<value>/var/bigdata/hadoop/full/dfs/secondary</value>

</property>

配置DN这个角色再那里启动

vi slaves

node02

node03

node04

1. 初始化&启动：

hdfs namenode -format

创建目录

并初始化一个空的fsimage

VERSION

CID

start-dfs.sh

第一次：datanode和secondary角色会初始化创建自己的数据目录

http://node01:50070

修改windows： C:\Windows\System32\drivers\etc\hosts

192.168.244.101 node01

192.168.244.102 node02

192.168.244.103 node03

192.168.244.104 node04

1. 简单使用：

hdfs dfs -mkdir /bigdata

hdfs dfs -mkdir -p /user/root

1. 验证知识点：

cd /var/bigdata/hadoop/local/dfs/name/current

观察 editlog的id是不是再fsimage的后边

cd /var/bigdata/hadoop/local/dfs/secondary/current

SNN 只需要从NN拷贝最后时点的FSimage和增量的Editlog

hdfs dfs -put hadoop\*.tar.gz /user/root

cd /var/bigdata/hadoop/local/dfs/data/current/BP-281147636-192.168.150.11-1560691854170/current/finalized/subdir0/subdir0

for i in `seq 100000`;do echo "hello hadoop $i" >> data.txt ;done

hdfs dfs -D dfs.blocksize=1048576 -put data.txt

cd /var/bigdata/hadoop/local/dfs/data/current/BP-281147636-192.168.150.11-1560691854170/current/finalized/subdir0/subdir0

检查data.txt被切割的块，他们数据什么样子