## 准备环境

1 台 linux 虚拟机(Centos6.x64 位)

ip 为：192.168.229.130

hostname 为：hadoop130

## 1.1 虚拟机环境准备

1. 克隆虚拟机

2. 修改克隆虚拟机的静态IP

[root@hadoop130~]#vi/etc/sysconfig/network-scripts/ifcfg-eth0

3. 修改主机名

vi /etc/sysconfig/network

vi /etc/hosts

4. 关闭防火墙

查看防火墙状态

service iptables status

临时关闭防火墙

service iptables stop

永久关闭防火墙

chkconfig iptables off

5. 创建用户

useradd hadoop

6. 配置用户具有root权限

（1）修改/etc/sudoers文件,找到下面一行。

## Allow root to run any commands anywhere  
hadoop    ALL=(ALL)    ALL

（2） 然后修改用户，使其属于root组（wheel），命令如下：

#usermod -g root hadoop

7. ssh免密登录

（1）生成公钥和私钥

ssh-keygen -t rsa

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

（2）将公钥拷贝到要免密登录的目标机器上

[hadoop@hadoop130 hadoop]$ ssh-copy-id hadoop131

[hadoop@hadoop130 hadoop]$ ssh-copy-id hadoop132

8．在/data目录下创建文件夹

（1）在/data目录下创建module、software文件夹

[hadoop@hadoop data]$ sudo mkdir module

[hadoop@hadoop data]$ sudo mkdir software

（2）修改module、software文件夹的所有者cd

[hadoop@hadoop data]$ sudo chown hadoop:hadoop module/ software/

## 1.2 安装JDK

1. 卸载现有JDK

（1）查询是否安装Java软件：

[hadoop@hadoop data]$ rpm -qa | grep java

（2）如果安装的版本低于1.7，卸载该JDK：

[hadoop@hadoop data]$ sudo rpm -e 软件包

（3）查看JDK安装路径：

[hadoop@hadoop data]$ which java

2. 用xftp工具将JDK导入到data目录下面的software文件夹下面

3. 解压JDK到/data /module目录下

[hadoop@hadoop130 software]$ tar -zxvf jdk-8u144-linux-x64.tar.gz -C /data/module/

4. 配置JDK环境变量

（1）先获取JDK路径

[hadoop@hadoop130 jdk1.8.0\_144]$ pwd

/data/module/jdk1.8.0\_144

（2）打开/etc/profile文件

[hadoop@hadoop130 jdk1.8.0\_144]$ sudo vi /etc/profile

在profile文件末尾添加JDK路径

#JAVA\_HOME

export JAVA\_HOME=/data/module/jdk1.8.0\_144

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

（3）保存后退出

:wq

（4）让修改后的文件生效

[hadoop@hadoop130 jdk1.8.0\_144]$ source /etc/profile

5. 测试JDK是否安装成功

[hadoop@hadoop130 jdk1.8.0\_144]# java -version

java version "1.8.0\_144"

注意：重启（如果java -version可以用就不用重启）

[hadoop@hadoop130 jdk1.8.0\_144]$ sync

[hadoop@hadoop130 jdk1.8.0\_144]$ sudo reboot

## 1.3 安装Hadoop

0. Hadoop下载地址：

<https://archive.apache.org/dist/hadoop/common/>

1. 解压安装文件到/data/module下面

[hadoop@hadoop130 software]$ tar -zxvf hadoop-3.1.3.tar.gz -C /data/module/

1. 查看是否解压成功

[hadoop@hadoop130 software]$ ls /data/module/hadoop-3.1.3

5. 将Hadoop添加到环境变量

（1）获取Hadoop安装路径

[hadoop@hadoop130 hadoop-3.1.3]$ pwd

/data/module/hadoop-3.1.3

（2）打开/etc/profile文件

[hadoop@hadoop130 hadoop-3.1.3]$ sudo vi /etc/profile

在profile文件末尾添加JDK路径：（shitf+g）

##HADOOP\_HOME

export HADOOP\_HOME=/data/module/hadoop-3.1.3

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

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

（3）保存后退出

:wq

（4）让修改后的文件生效

[hadoop@hadoop130 hadoop-3.1.3]$ source /etc/profile

6. 测试是否安装成功

[hadoop@hadoop130 hadoop-3.1.3]$ hadoop version

Hadoop 3.1.3

7. 重启(如果Hadoop命令不能用再重启)

[hadoop@hadoop130 hadoop-3.1.3]$ sync

[hadoop@hadoop130 hadoop-3.1.3]$ sudo reboot

## 1.4 Hadoop目录结构

1、查看Hadoop目录结构

[atguigu@hadoop101 hadoop-3.1.3]$ ll

总用量 200

drwxr-xr-x. 2 hadoop root 4096 9月 12 2019 bin

drwxr-xr-x. 3 hadoop root 4096 9月 12 2019 etc

drwxr-xr-x. 2 hadoop root 4096 9月 12 2019 include

drwxr-xr-x. 3 hadoop root 4096 9月 12 2019 lib

drwxr-xr-x. 4 hadoop root 4096 9月 12 2019 libexec

-rw-r--r--. 1 hadoop root 147145 9月 4 2019 LICENSE.txt

-rw-r--r--. 1 hadoop root 21867 9月 4 2019 NOTICE.txt

-rw-r--r--. 1 hadoop root 1366 9月 4 2019 README.txt

drwxr-xr-x. 3 hadoop root 4096 9月 12 2019 sbin

drwxr-xr-x. 4 hadoop root 4096 9月 12 2019 share

2、重要目录

（1）bin目录：存放对Hadoop相关服务（HDFS,YARN）进行操作的脚本

（2）etc目录：Hadoop的配置文件目录，存放Hadoop的配置文件

（3）lib目录：存放Hadoop的本地库（对数据进行压缩解压缩功能）

（4）sbin目录：存放启动或停止Hadoop相关服务的脚本

（5）share目录：存放Hadoop的依赖jar包、文档、和官方案例

## 1.5 修改配置文件

（1）配置集群

（a）配置：hadoop-env.sh

Linux系统中获取JDK的安装路径：

[[hadoop@hadoop130 hadoop]# echo $JAVA\_HOME

/opt/module/jdk1.8.0\_144

修改JAVA\_HOME 路径：

export JAVA\_HOME=/opt/module/jdk1.8.0\_144

（b）配置：core-site.xml

|  |
| --- |
| <!-- 指定HDFS中NameNode的地址 -->  <property>  <name>fs.defaultFS</name>  <value>hdfs://hadoop130:9000/</value>  </property>  <!-- 指定Hadoop运行时产生文件的存储目录 -->  <property>  <name>hadoop.tmp.dir</name>  <value>/data/module/hadoop\_repo</value>  </property>  <!-- io.file.buffer.size的默认值4096.这时读写sequence file的buffer size，可减少I/O次数 -->  <property>  <name>io.file.buffer.size</name>  <value>13107</value>  </property> |

（c）配置：hdfs-site.xml

|  |
| --- |
| <!-- 指定HDFS副本的数量 -->  <property>  <name>dfs.replication</name>  <value>2</value>  </property>  <!-- 指定Hadoop辅助名称节点主机配置 -->  <property>  <name>dfs.namenode.secondary.http-address</name>  <value>hadoop130:9001</value>  </property>  <property>  <name>dfs.namenode.name.dir</name>  <value>file:/data/module/dfs/name</value>  </property>  <property>  <name>dfs.datanode.data.dir</name>  <value>file:/data/module/dfs/data</value>  </property>  <property>  <name>dfs.webhdfs.enabled</name>  <value>true</value>  </property> |

（d）配置：mapred-site.xml

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<!-- 历史服务器端地址 -->

<property>

<name>mapreduce.jobhistory.address</name>

<value>hadoop130:10020</value>

</property>

<!-- 历史服务器web端地址 -->

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>hadoop130:19888</value>

</property>

（d）配置：yarn-site.xml

<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>hadoop130</value>

</property>

2.在workers配置文件中指定DataNode节点的主机名

[hadoop@hadoop130 hadoop]$ vim workers

hadoop131

hadoop132

3.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@hadoop130 hadoop]# cd /data/soft/hadoop-3.2.0/sbin

[root@hadoop130 sbin]# vi start-dfs.sh

HDFS\_DATANODE\_USER=root

HDFS\_DATANODE\_SECURE\_USER=hdfs

HDFS\_NAMENODE\_USER=root

HDFS\_SECONDARYNAMENODE\_USER=root

[root@hadoop130 sbin]# vi stop-dfs.sh

YARN\_RESOURCEMANAGER\_USER=root

HADOOP\_SECURE\_DN\_USER=yarn

YARN\_NODEMANAGER\_USER=root

## 1.6把Hadoop安装软件复制到其他DataNode节点：hadoop131和hadoop132

[hadoop@hadoop130 ~]$ scp -r /data/module/hadoop-3.1.3 hadoop@192.168.229.131:/data/module

[hadoop@hadoop130 ~]$ scp -r /data/module/hadoop-3.1.3 hadoop@192.168.229.132:/data/module

[hadoop@hadoop130 ~]$ scp -r /data/module/jdk1.8.0\_144 hadoop@192.168.229.131:/data/module

[hadoop@hadoop130 ~]$ scp -r /data/module/jdk1.8.0\_144 hadoop@192.168.229.132:/data/module

## 1.7在Hadoop130上启动Hadoop

（a）**格式化NameNode**（第一次启动时格式化，以后就不要总格式化）

[hadoop@hadoop130 hadoop-3.1.3]$ bin/hdfs namenode -format

//（b）启动hdfs（单起HDFS）

[root@hadoop130 hadoop-3.1.3]# sbin/start-dfs.sh

（b）启动全部集群

[root@hadoop130 hadoop-3.1.3]# sbin/start-all.sh

Starting namenodes on [hadoop130]

Starting datanodes

Starting secondary namenodes [hadoop130]

2020-06-17 02:25:45,602 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Starting resourcemanager

Starting nodemanagers

（c）验证集群（分别在各个机器上执行）

[root@hadoop130 hadoop-3.1.3]# jps

6370 NameNode

6646 SecondaryNameNode

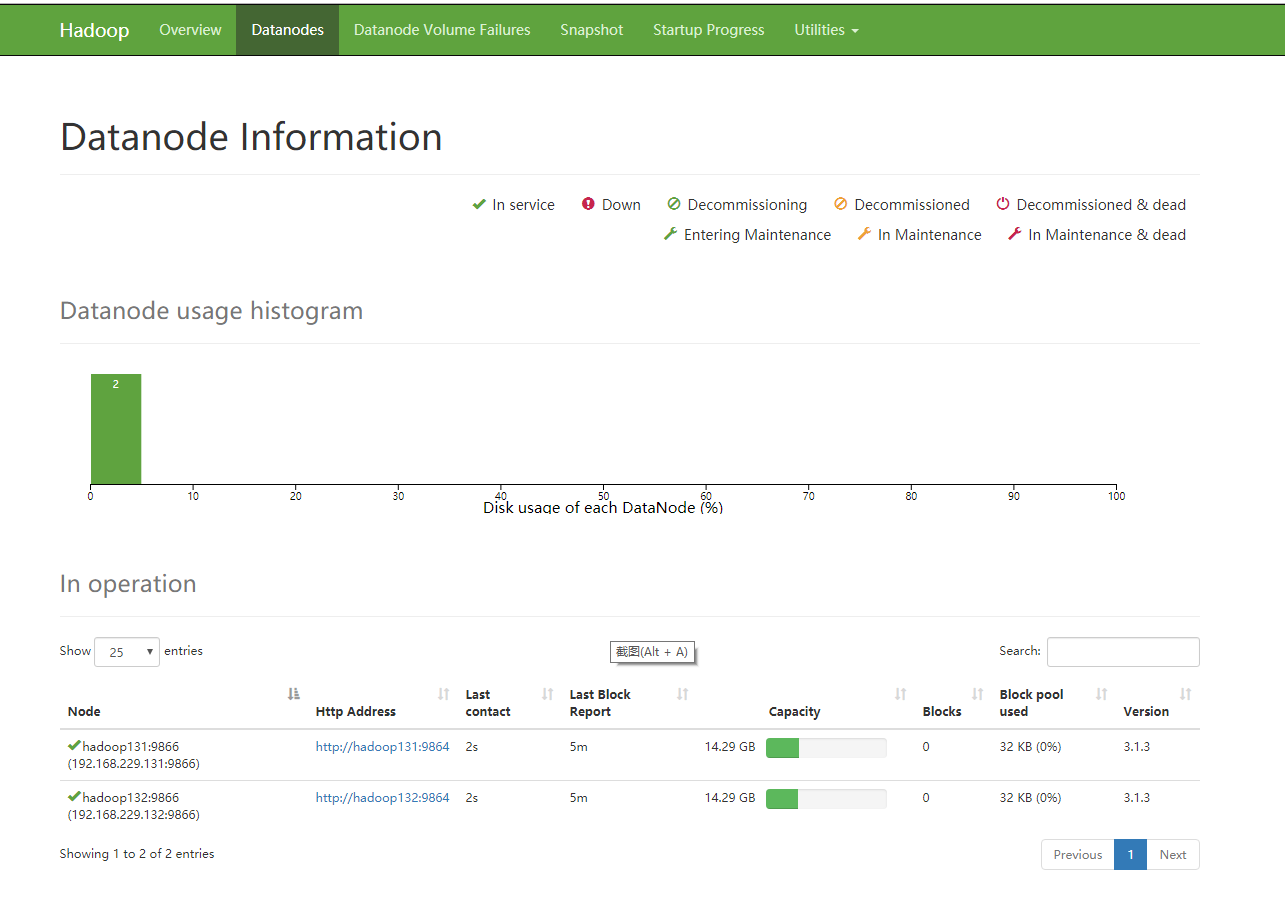
6921 ResourceManager

7257 Jps

至此，hadoop 分布式集群安装成功

## 1.8登录web界面查看hdfs状态和查看文件款组成部分

<http://192.168.229.130:9870/dfshealth.html#tab-datanode>



## 1.9通过web查看Hadoop集群状态

<http://192.168.229.130:8088>