# CDH搭建

参考

https://blog.csdn.net/sinat\_28007043/article/details/92840860

https://www.cnblogs.com/mylovelulu/p/10384732.html

## 一 、安装centos

下载centos7 iso镜像，使用Ultraiso或者其他启动盘制作软件，将u盘制作成启动盘，然后安装。

需要注意的问题：启动u盘的名称需要和iso文件中默认设定的名称一致

但是FAT32格式的u盘盘符名称长度有限制，可能没法输入那么长的名称，会导致安装时找不到安装文件。这时需要在centos安装时选择安装文件时按tab或e，修改名称和u盘的一致。

## 二、固定ip、hosts和hostname

主机的Ip地址不能变动。在路由器里绑定主机MAC地址和ip地址，让每台主机的ip不变。

然后在每台主机上进行如下操作。

修改网卡配置文件：

vim /etc/sysconfig/network-scripts/ifcfg-ens33

BOOTPROTO=static #开机协议，有dhcp及static；

ONBOOT=yes #设置为开机启动；

DNS1=114.114.114.114 #这个是国内的DNS地址，是固定的；

IPADDR=192.168.1.150 #在路由器上为每台主机固定好的IP地址；

NETMASK=255.255.255.0 #子网掩码，不需要修改；

GATEWAY=192.168.1.1 #网关

修改hosts文件，添加集群每台主机的IP和名称

vim /etc/hosts

修改hostname，即主机名称。

vim /etc/hostname

重启网络服务

service network restart

## 三、关闭防火墙和SSH免密连接(所有节点)

查看防火墙状态

firewall-cmd --state

关闭防火墙

systemctl stop firewalld.service

关闭防火墙自动启动

systemctl disable firewalld.service

关闭selinux

vim /etc/selinux/config —> SELINUX=disabled (修改)

设置SSH免密连接

生成密钥

ssh-keygen -t rsa

将本主机的公钥添加到其他主机的认证密匙中

scp ~/.ssh/id\_rsa.pub root@node1:~/.ssh/authorized\_keys

## 四、时间同步

(网上说centos7默认是用chrony同步时间，不是ntp)

参考：https://blog.csdn.net/weixin\_42652946/article/details/83935097

主节点端

1)vi /etc/chrony.conf

# Allow NTP client access from local network.

该行下添加： allow 192.168.1.157 （需要同步时间的服务器的所有ip）

解除下面一行注释：local stratum 10

2）启动chrony

systemctl start chronyd.service

3）设置开机同步时间

systemctl enable chronyd.service

node节点

1) vi /etc/chrony.conf

# Please consider joining the pool (http://www.pool.ntp.org/join.html).

注释掉该行下所有以 server开头的，并添加主节点ip iburst(例如：master iburst)。

2）启动chrony

systemctl start chronyd.service

3）设置开机同步时间

systemctl enable chronyd.service

所有主机刷新一下同步情况：

chronyc sources

（若时间同步软件是ntp）

vim /etc/ntp.conf

添加 server ntp.aliyun.com 注释其他。

## 五、mysql(仅master节点)和jdk(所有主机)和jdbc驱动(所有节点)

参考：https://blog.csdn.net/qq\_36582604/article/details/80526287

wget -i -c http://dev.mysql.com/get/mysql57-community-release-el7-10.noarch.rpm

yum -y install mysql57-community-release-el7-10.noarch.rpm

yum -y install mysql-community-server

systemctl start mysqld

systemctl enable mysqld

grep 'temporary password' /var/log/mysqld.log(获得l密码)

ALTER USER 'root'@'localhost' IDENTIFIED BY 'fuzadeMIMA1.';

创建初始数据库

GRANT ALL ON \*.\* TO 'cdh'@'%' IDENTIFIED BY 'fuzadeMIMA1.';

CREATE DATABASE cdh DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

安装JDK(提前下好了rpm包)

安装

yum localinstall jdk-8u212-linux-x64.rpm

配置环境文件

vim /etc/profile

#java environment

export JAVA\_HOME=/usr/java/jdk1.8.0\_212-amd64

export CLASSPATH=.:${JAVA\_HOME}/jre/lib/rt.jar:${JAVA\_HOME}/lib/dt.jar:${JAVA\_HOME}/lib/tools.jar

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

source /etc/profile

jdbc驱动

jdbc改好名字放到 usr/share/java 路径下面

cp mysql-connector-java-8.0.16.jar /usr/share/java/mysql-connector-java.jar

更新psycopg(cdh6需要2.5.4版本以上)

yum install postgresql-devel python-devel gcc #psycopg依赖包

yum install epel-release #安装pip工具的库

yum install python-pip #安装pip

pip install --upgrade Psycopg2

(若无法安装)pip install Psycopg2 --upgrade --ignore-installed Psycopg2

## 六、swappiness参数和透明大页面(所有节点)

修改当前参数为10

sysctl vm.swappiness=10

查看参数数值

cat /proc/sys/vm/swappiness

永久修改参数

echo 'vm.swappiness=10'>> /etc/sysctl.conf（在/etc/sysctl.conf 文件里添加如下参数：vm.swappiness=10）

临时修改透明大页面

echo never > /sys/kernel/mm/transparent\_hugepage/defrag

echo never > /sys/kernel/mm/transparent\_hugepage/enabled

永久生效

在/etc/rc.local 添加上面两行命令

## 七、配置CDH本地库,并安装(离线安装方式，仅主机设置)

yum install httpd

systemctl start httpd

systemctl enable httpd

yum install createrepo

将cm安装包放到 /var/www/html/cloudera-repos/ 下

├── allkeys.asc

├── cloudera-manager-agent-6.0.1-610811.el7.x86\_64.rpm

├── cloudera-manager-daemons-6.0.1-610811.el7.x86\_64.rpm

├── cloudera-manager-server-6.0.1-610811.el7.x86\_64.rpm

└── cloudera-manager-server-db-2-6.0.1-610811.el7.x86\_64.rpm

创建库

cd /var/www/html/cloudera-repos

createrepo .

配置本地库的源(主节点和所有分节点都设置)

cd /etc/yum.repos.d

vim cloudera-manager.repo

[cloudera-manager]

name=Cloudera Manager 6.0.1

baseurl=http://master/cloudera-repos/

gpgcheck=0

enabled=1

刷新源(主节点和所有分节点都设置)

yum clean all

yum makecache

安装(主节点，需要安装cloudera-manager-server)

yum install cloudera-manager-daemons cloudera-manager-agent cloudera-manager-server

安装(分节点，不需要安装cloudera-manager-server)

yum install cloudera-manager-daemons cloudera-manager-agent

设置那个节点为集群主节点(所有节点设置)

vi /etc/cloudera-scm-agent/config.ini

找到选项，并将主节点ip修改为当前的主节点ip。

设置cloudera-manager-server的数据库(仅主节点)

cd /etc/cloudera-scm-server

vim db.properties

将cdh的安装包放到指定路径(仅主节点)

CDH-6.0.1-1.cdh6.0.1.p0.590678-el7.parcel和manifest.json

cd /opt/cloudera/parcel-repo

生成sha文件

sha1sum CDH-6.0.1-1.cdh6.0.1.p0.590678-el7.parcel | awk '{ print $1 }' > CDH-6.0.1-1.cdh6.0.1.p0.590678-el7.parcel.sha

修改文件权限，可能不需要

chown -R cloudera-scm:cloudera-scm /opt/cloudera/parcel-repo/\* （可能不需要）

主节点启动服务器(仅主节点)

systemctl start cloudera-scm-server //启动服务器

tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log //查看启动过程

启动客户端(所有节点)

systemctl start cloudera-scm-agent //启动客户端

启动好访问主节点7180端口。

Master:7180

在页面中设置安装CDH时，应当设置parsel库为本地，不然在线下载可能有点慢。

CDH集群提示副本不足的块的问题

参考：https://blog.csdn.net/xjping0794/article/details/77848001

hdfs fsck /hbase (查看文件块的数量等

推测是主机数量不够

修改复制因子从3到2，发现没用(原来是这个参数只对以后存进来的文件有效)

sudo -u hdfs hadoop dfs -setrep -w 2 -R /(修改现在的所有文件的复制因子为2，使用hdfs用户是因为，文件的权限是755，必须要所有者才能操作)