**Ubuntu16.04上搭建CDH5.14集群**

2018年03月16日 10:46:40 [岁月饶过谁](https://me.csdn.net/q1370992706) 阅读数：849 标签： [hadoop](http://so.csdn.net/so/search/s.do?q=hadoop&t=blog)[CDH](http://so.csdn.net/so/search/s.do?q=CDH&t=blog)[Cloudera Manager](http://so.csdn.net/so/search/s.do?q=Cloudera%20Manager&t=blog) 更多

个人分类： [大数据](https://blog.csdn.net/q1370992706/article/category/7512353)

**Ubuntu16.04上搭建CDH5.14集群**

最近实验室开始搞大数据的项目，基于Hadoop进行大数据管理并使用机器学习算法分析数据，于是花了几天时间倒腾了一下ubuntu16.04下Hadoop集群的环境搭建。

创建cloudera-scm用户（所有节点）：

root@master:~# useradd --system --home=/opt/cloudera-manager/cm-5.14.2/run/cloudera-scm-server --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm

在介绍具体的硬件、软件配置之前，需要强调一点，所有的命令都是在root用户下执行的。使用如下的命令可以切换到root用户

sudo su

在教程中使用的编辑器是vim，如果没有安装，请使用下面的命令安装，并自行查找相关的使用资料。

apt-get install vim

**硬件平台介绍**

集群总共包括三台电脑，一台主机两台从机，电脑名称和名称对应关系如下：

hadoop-master 大电脑

hadoop-slave1 一体机

hadoop-slave2 台式机

各个主机通过路由器组件局域网，通常路由器会给每个主机分配固定的IP地址，在Ubuntu终端下，可以使用ifconfig查看机器的IP地址。如下为集群的主机和IP对应关系（需要根据自己的情况修改）：

192.168.1.129 hadoop-master

192.168.1.78 hadoop-slave1

192.168.1.53 hadoop-slave2

**Hostname、hosts和防火墙设置**

首先在此强调所有的命令都是在root用户下执行的。

* hostname设置   
  首先设置所有主机的hostname，使用下面的命令

vim /etc/hostname

* + 1

将xx-PC类似的名字改成hadoop-master（主机）或者是hadoop-slave1、hadoop-slave2（从机）。

* hosts配置   
  修改所有主机的hosts配置，使用如下的命令：

vim /etc/hosts

* + 1

hosts文件修改后的内容如下所示

127.0.0.1 localhost

#127.0.0.1 xx-pc 这一行一定要删除掉，否则后面配置CDH的时候会有问题

192.168.1.129 hadoop-master

192.168.1.178 hadoop-slave1

192.168.1.53 hadoop-slave2

#ipv6相关的配置不需要更改

* 关闭防火墙   
  使用下面的命令关闭所有主机的防火墙
* iptables-save > /root/firewal.rules #保存防火墙的规则

serviece ufw stop #关闭防火墙

* + 1
  + 2

如果不关闭防火墙，在后续配置集群的时候主节点会无法检测到其他的节点。   
在执行完上面的所有配置以后，重启所有电脑。

* 测试配置结果   
  依次在所有节点的终端上，执行下面的命令
* ping hadoop-master
* ping hadoop-slave1

ping hadoop-slave2

如果都ping成功了，说明上面的配置就没有问题了。

**ssh服务配置**

再次强调，所有的指令都是在root命令下执行的。   
为了能够让机器远程ssh登录到root账号，需要进行两个配置：修改ssh的配置，允许远程登录到root用户；拷贝公钥。

* 修改ssh配置

首先确认在所有的节点上都安装了ssh，测试的方法为在终端输入ssh会弹出提示信息，否则显 示没有这个可执行程序。   
如果没有安装ssh，则使用下面的命令安装

apt-get install openssh-\*

* + 1

然后使用下面的命令打开ssh配置

vim /etc/ssh/sshd\_config

* + 1

修改ssh的root用户远程设置

#PermitRootLogin prohibit-password #注释掉下面这一行，有时候为without-password

PermitRootLogin yes

退出vim并保存。   
重新启动ssh

service ssh restart

* 生成并拷贝ssh公钥   
  在**所有的节点**上，执行下面的命令生成ssh的公钥

ssh-keygen -t rsa

上面的命令会在当前用户目录（也就是root用户）下生成.ssh文件，里面存放了公钥和私钥，需要将公钥添加到hadoop-master和hadoop-slave[1-2]上

ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop-master

ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop-slave1

ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop-slave2

上面的命令会将公钥添加到远程机器相同账号（root用户）的~/.ssh/authorized\_keys文件夹中。   
【注意】在所有的机器上进行完上面的命令以后，再进行测试。

* 测试配置结果   
  在终端下，输入下面的命令测试
* ssh hadoop-master #如果成功，请输入`exit`命令退出远程登录
* ssh hadoop-slave1 #如果成功，请输入`exit`命令退出远程登录

ssh hadoop-slave2 #如果成功，请输入`exit`命令退出远程登录

ssh成功后，会进入到远程机器上，请使用exit命令退出远程登录再测试其他的机器。否则，会出现Access Denied的提示信息。

**配置JDK环境和MySQL**

下面介绍Cloudera Manager的安装需要的java和MySQL环境配置

* JDK安装   
  在Ubunt16.04上默认的jdk版本就已经够了，如果在系统上没有进行过修改， 输入java -version 命令应该就会直接显示java的版本。
* 安装MySQL

在hadoop-master节点上，使用下面的命令安装MySQL

apt-get install mysql-server mysql-client

* + 1

**部署Cloudera Manager**

为了安装Cloudera Manager，需要执行如下的几个步骤

* 软件准备   
  （这个步骤只需要在hadoop-master上执行）首先需要准备四个文件cloudera-manager-xenial-cm5.14.0\_amd64.tar.gz，CDH-5.14.0-1.cdh5.14.0.p0.24-xenial.parcel，CDH-5.14.0-1.cdh5.14.0.p0.24-xenial.parcel.sha1和manifest.json文件。如果没有这四个文件，可以使用下面的命令安装
* # 默认下载在/root/backup文件夹下
* mkdir -p /root/backup
* cd /root/backup
* # 下载cloudera-manager
* wget -c http://archive.cloudera.com/cm5/cm/5/cloudera-manager-xenial-cm5.14.0\_amd64.tar.gz
* # 下载cdh的parcel文件
* wget -c http://archive.cloudera.com/cdh5/parcels/latest/CDH-5.14.0-1.cdh5.14.0.p0.24-xenial.parcel
* # 下载cdh的parcel.sha文件
* wget -c http://archive.cloudera.com/cdh5/parcels/latest/CDH-5.14.0-1.cdh5.14.0.p0.24-xenial.parcel.sha1
* # 下载manifest文件

wget -c http://archive.cloudera.com/cdh5/parcels/latest/manifest.json

下载了四个文件以后，需要将CDH-5.14.0-1.cdh5.14.0.p0.24-xenial-parcel.sha1改名为CDH-5.14.0-1.cdh5.14.0.p0.24-xenial-parcel.sha，使用下面的命令

cd /root/backup

mv CDH-5.14.0-1.cdh5.14.0.p0.24-xenial-parcel.sha1 CDH-5.14.0-1.cdh5.14.0.p0.24-xenial-parcel.sha

然后建立如下所示的文件树

| --/opt

|--/cloudera

|--/parcels

|--/parcel-repo

|--/CDH-5.14.0-1.cdh5.14.0.p0.24-xenial.parcel

|--/CDH-5.14.0-1.cdh5.14.0.p0.24-xenial.parcel.sha

|--/manifest.json

|--/cm-5.14.0

* + 1

具体的操作命令如下

cd /opt

mkdir cloudera #创建cloudera文件夹

cd cloudera

mkdir parcels #创建parcels文件夹

mkdir parcel-repo #创建parcel-rep文件夹

cd parcel-repo

cp /root/backup/\* ./ #将三个文件拷打percel-repo文件夹下

tar -zxf cloudera-manager-xenial-cm5.14.0\_amd64.tar.gz -C /opt #解压cm

* 为Cloudera Manager连接MySQL   
  首先在所有节点上安装mysql-connector-java软件包

apt-get install libmysql-java

在hadoop-master上链接mysql连接库到cm

ln -s /usr/share/java/mysql-connector-java.jar /opt/cm-5.14.0/share/cmf/lib/mysql-connector-java.jar

在hadoop-master上配置cm5的数据库

cd /opt/cm-5.14.0/share/cmf/schema

./scm\_prepare\_database.sh mysql -uroot -p123456 scm scm scm --force

在命令中加--force是为了防止重新配置时，因为scm数据库已经存在而终止执行命令。

* 配置Agent   
  在hadoop-master上修改配置文件的server
* vim /opt/cm-5.14.0/etc/cloudera-scm-agent/config.ini #打开config.ini文件

server\_host=hadoop-master #修改server\_host为hadoop-master

将hadoop-master上的cm-5.14.0复制到其他的节点上

scp -r /opt/cm-5.14.0 hadoop-slave1:/opt/

scp -r /opt/cm-5.14.0 hadoop-slave2:/opt/

* 启动Cloudera Manager的Server和Agent   
  在hadoop-master节点上启动Server进程和Agent进程
* cd /opt/cm-5.14.0/etc/init.d
* ./cloudera-scm-server start

./cloudera-scm-agent start

在hadoop-slave[1-2]节点上启动Agent进程

cd /opt/cm-5.14.0/etc/init.d

./cloudera-scm-agent start

**正式部署-可视化集群配置**

在等待Cloudera Manager Server和Agent启动完毕后，就可以使用CDH5的可视化界面了，在集群的机器上浏览器输入链接http://hadoop-master/7180，登录管理界面。如果访问失败，可能需要稍微等几十秒，不停的刷新浏览器。如果还是不行，可能是前面的配置有问题，请确认前面的配置没有错误。

* 登录   
  如果配置没有问题，登录以后的界面为如下所示，用户名和密码默认为admin   
  + 同意条款
  + 选择版本   
    选择中间的使用版本，在试用期结束后会默认进入免费版本。
  + 选择集群机器   
    注意，一定要选择红色椭圆那个选项，如果红色的方框内主机数目和IP地址，如果数目不对，很有可能是没有关闭防火墙，注意所有的节点都需要关闭防火墙；如果是IP地址不对，那么需要检查Hosts文件。
  + 选择集群安装方式   
    请选择Parcel方式，如果CDH的版本没有5.14.0，那么请确认文件树是对的，并且xx-parcel.sha1文件名字修改为了xx-parcel.sha。
  + 分发Parcel到各个节点
  + 检测安装环境   
    如果出现如下的警告，在所有的节点上输入echo 10 > /proc/sys/vm/swapiness，然后刷新下浏览器。
  + 选择安装服务   
    仅选择了核心服务，如果后续需要增加，可以通过管理界面操作。
  + 角色配置
  + 数据库配置   
    数据库配置是最容易出错的地方，配置数据库之前，首先需要创建数据库，只需要在hadoop-master节点上配置数据，包括四个数据库（hive,rman,ozzie,hue）和四个用户名（hive,rman,ozzie,hue），配置的命令如下
  + mysql -uroot -p123456
  + #创建数据库
  + create database hive DEFAULT CHARSET utf8;
  + create database rman DEFAULT CHARSET utf8;
  + create database oozie DEFAULT CHARSET utf8;
  + create database hue DEFAULT CHARSET utf8;
  + #创建用户名
  + grant all on hive.\* TO 'hive'@'%' IDENTIFIED BY '123456';
  + grant all on rman.\* TO 'hive'@'%' IDENTIFIED BY '123456';
  + grant all on oozie.\* TO 'hive'@'%' IDENTIFIED BY '123456';

grant all on hue.\* TO 'hue'@'%' IDENTIFIED BY '123456';

注意，上面的每一个命令后都会有分号，在输完一行命令以后都会在下面显示OK!   
  
然后点击测试连接，如果测试失败，可以试试将hadoop-master改成localhost试试。

* + 进行安装
  + 安装成功