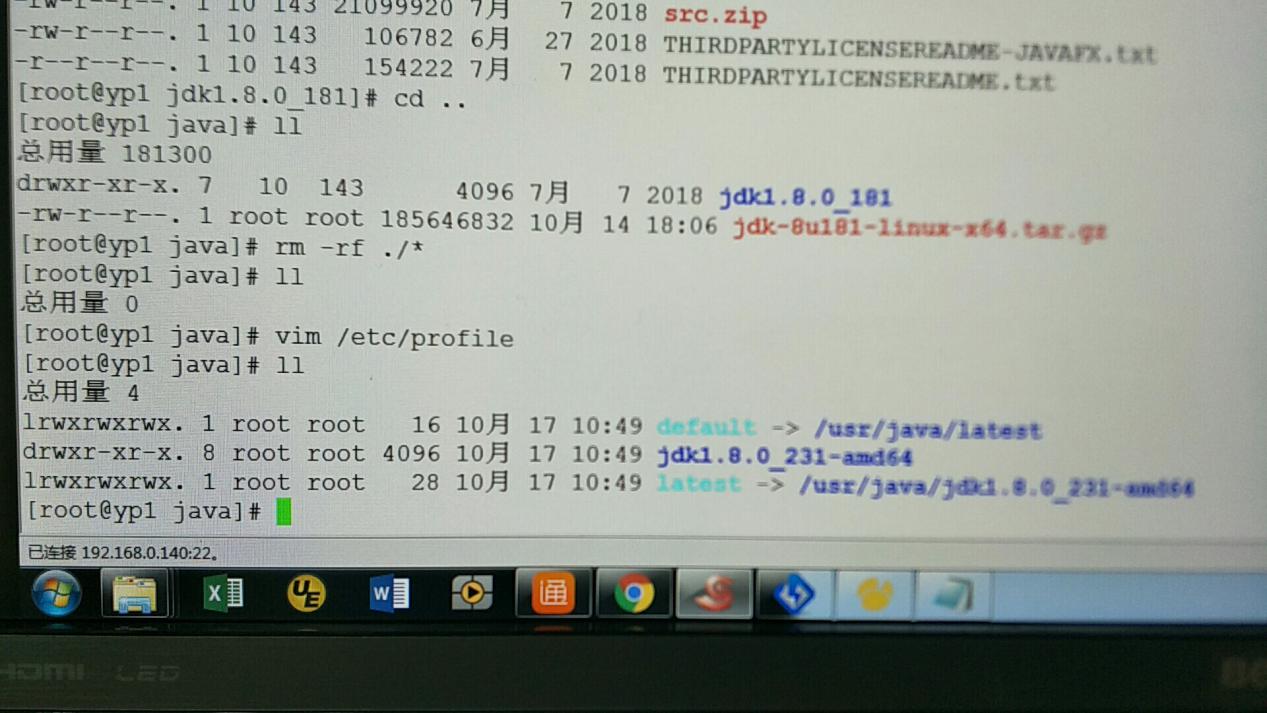
# CDH集群搭建

1. **准备基础环境**

**（一）安装jdk**

1. 在192.168.0.140机器上，/home/banshuai下，有一个jdk的rpm包，scp都要安装的机器上，通过rpm -ivh jdkXX.rpm命令安装

安装成功后，会在/usr/java目录下出现一些目录：



确保java -version的时候显示的是刚刚安装的版本，如果有其他版本，将/etc/profile下的关于java的配置都清除掉，并用yum -y remove jdk（具体命令可以网上找到）类似的命令卸载。

通过rpm -qa|grep jdk来验证是否还有别的版本的jdk，确保只有一个刚刚安装的版本。

1. 将140服务器上的/etc/yum.repos.d目录分发到所需服务器的相同目录下，可供使用相同的yum源
2. 修改hostname：执行hostnameclt hostname-set yp1来修改hostname，通过命令hostname来查看当前hostname，修改后会变为新的hostname。确保每一台都修改后，将所有的hostname与对应的ip追加到/etc/hosts文件中，然后reboot命令重启服务器。

**（二）安装ntp**

1、安装ntp，通过yum -y install ntp\*，安装完成后，修改/etc/ntp.conf，修改参照192.168.0.140机器（即yp1机器）。启动ntp服务器，每台机器通过ntpdate yp1来将时间与yp1对准

1. **安装mysql**

注意，一般CDH的cm在哪台机器上，mysql就装到哪台机器上。

1、将安装mysql所需要的bundle包统一放入到/var/www/html/mysql-5.7.27下，（对于dx项目已经将这些包放入到140机器的该目录下，来自于班帅的mysql-5.7.27-1.el7.x86\_64.rpm-bundle.tar中，该tar中绑定了mysql所有的依赖）。

注：由于140机器上已经做了mysql的yum源，所以接下来几个步骤不需要做。只需执行第（一）、2步骤即可。

2、卸载机器上的mariadb，通过yum -y remove mariadb\*命令

3、vi /etc/yum.repos.d/mysql-community.repo（如果没有则新增），增加如下内容：

[mysql57-community]

name=MySQL 5.7 Community Server

baseurl=http://192.168.0.140/mysql-5.7.27/

enabled=1

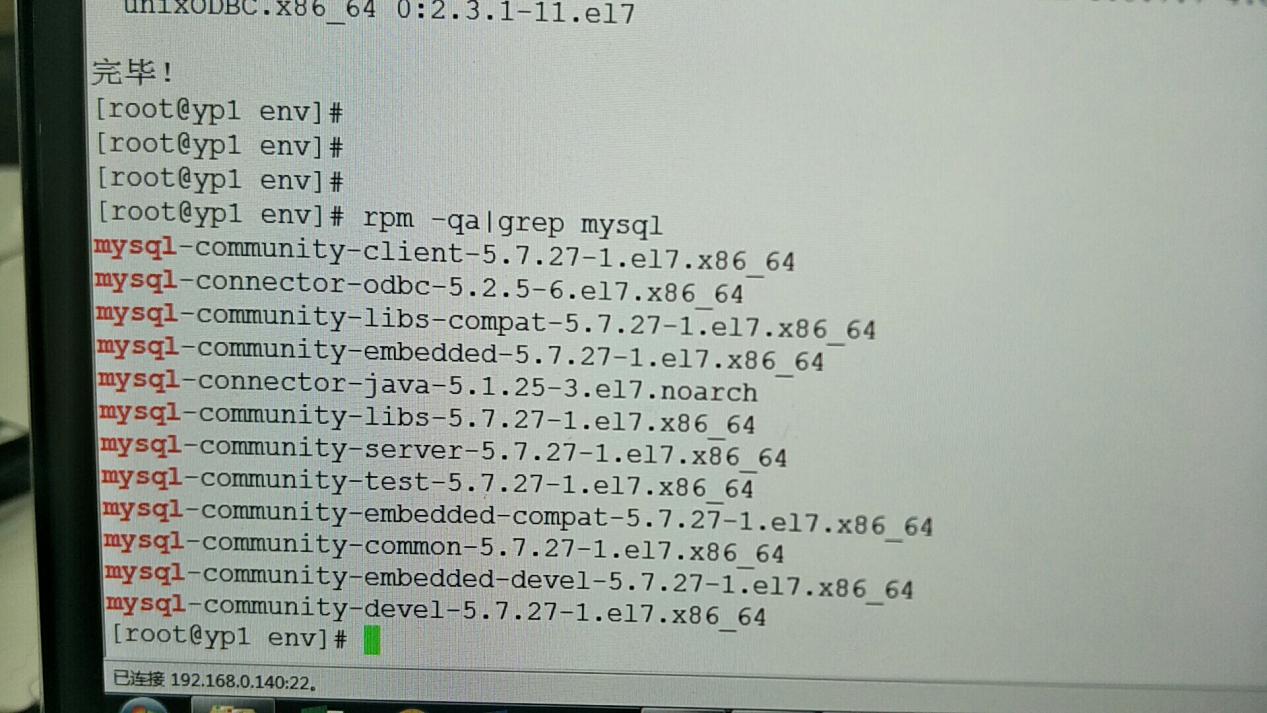
gpgcheck=0

gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-mysql

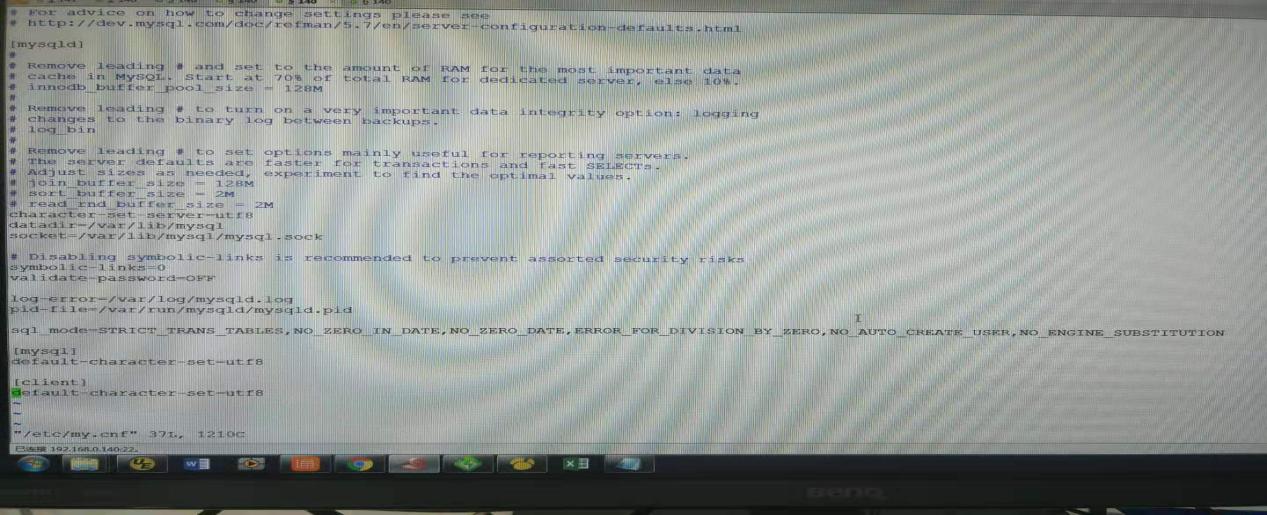
注：baseurl即是rpm包所放置的位置，由于前面做成了http请求，所以这样就可以定位到包的位置。

4、createrepo /var/www/html/mysql-5.7.27，为这些rpm包创建索引。

5、yum -y install mysql\* 安装mysql，确保下图中的包都已安装



1. 修改mysql配置文件：vim /etc/my.conf



8、启动mysql：server mysqld start

9、修改mysql密码，并给指定的机器授权。

**（三）准备CDH需要的包**

1、将安装CDH需要的cloudera-manager-el6-cm5.12.2\_x86\_64.tar.gz包放到/opt目录下（可在<http://archive.cloudera.com/cm5/cm/5/>中下载对应版本），并解压。

2、在

<https://archive.cloudera.com/cm5/redhat/7/x86_64/cm/5.12.2/RPMS/x86_64/>下载centos7对应的安装包



（针对dx项目，由于之前已经安装，所以可以不用载，已经放到了/var/www/html/cm5下作为一个yum源。如果没有安装新下载的，则继续看下面的步骤）

1. mkdir /var/www/html/cm5,将下载的7个包放到该目录下，执行

createrepo /var/www/html/cm5建立rpm索引

1. vi /etc/yum.repos.d/cm5.repo（如果没有该文件新增） 增加如下内容：

[cm5]

name=cm5

baseurl=http://192.168.0.140/cm5

enabled=1

gpgcheck=0

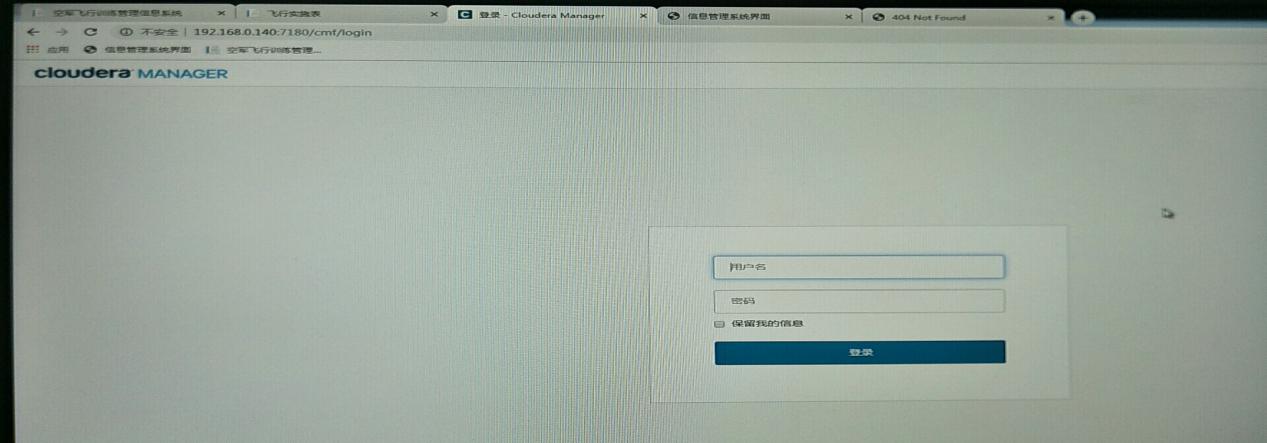
1. 将该文件分发到各个机器的相同目录下，如果之前已有（可能在（一）、2的时候已经做了），检查内容是否一致。
2. 在官网<http://archive.cloudera.com/cdh5/parcels/5.12/>处下载对应版本的[parcel](http://archive.cloudera.com/cdh5/parcels/5.12/CDH-5.12.2-1.cdh5.12.2.p0.4-el7.parcel)文件，由于使用的centos7，所以要下载el7的文件（对于DX项目，该文件在140机器的/opt/cloudera/parcel-rpo下，只需将文件放到相同目录下），



1. **安装CDH**

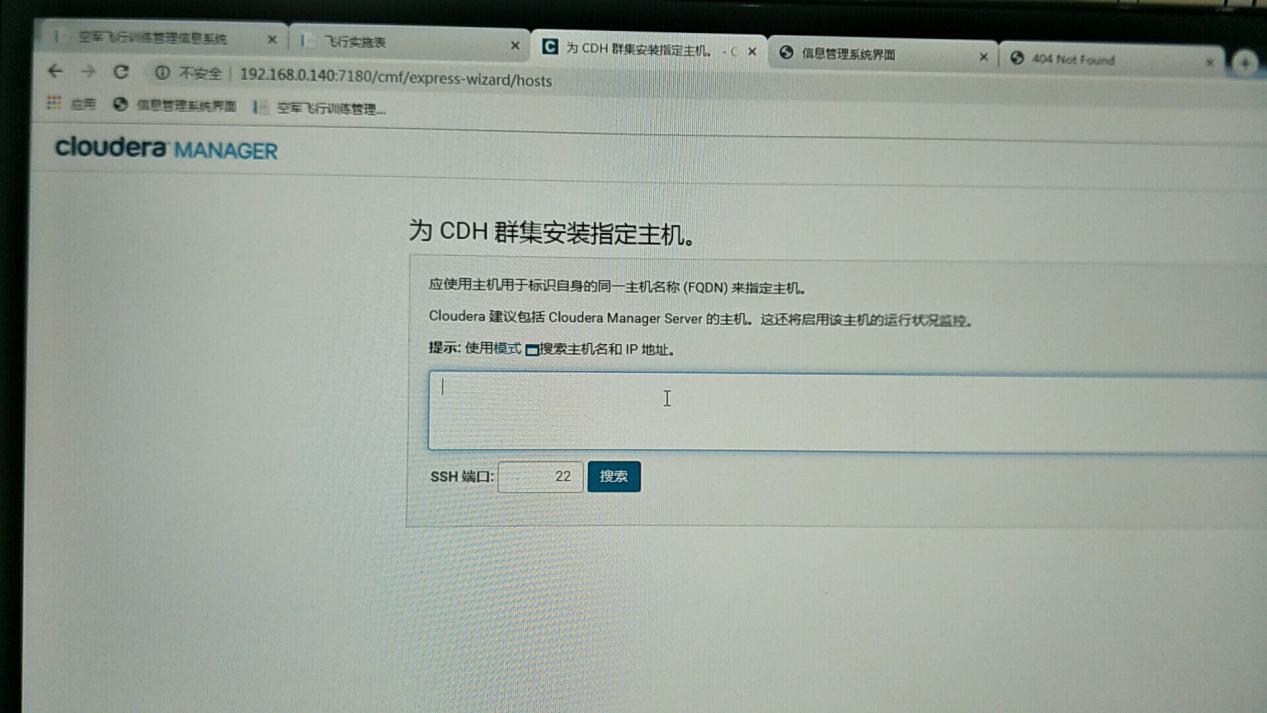
**（一）安装CDH服务**

1. 在cm机器上安装server，执行命令：yum -y install cloudera-manager-server\*
2. 安装后启动server：/opt/cm-5.12.2/etc/init.d/cloudera-scm-server restart，启动成功后页面上可以放问CDH，ip:7180，用户名和密码都是admin



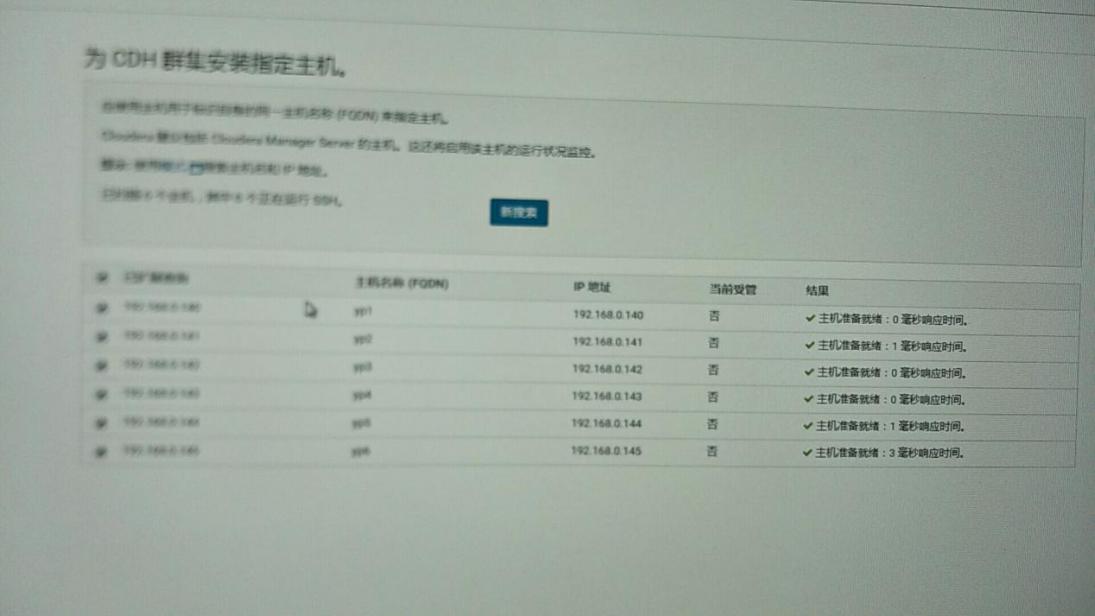
注：服务启动后执行（二）中数据库初始化

1. 进入后会要求同意协议，同意后点击左上角的logo进入首页。
2. 点击“主机”->“所有主机”->“向集群添加主机”
3. 点击“继续”跳到选择主机界面：



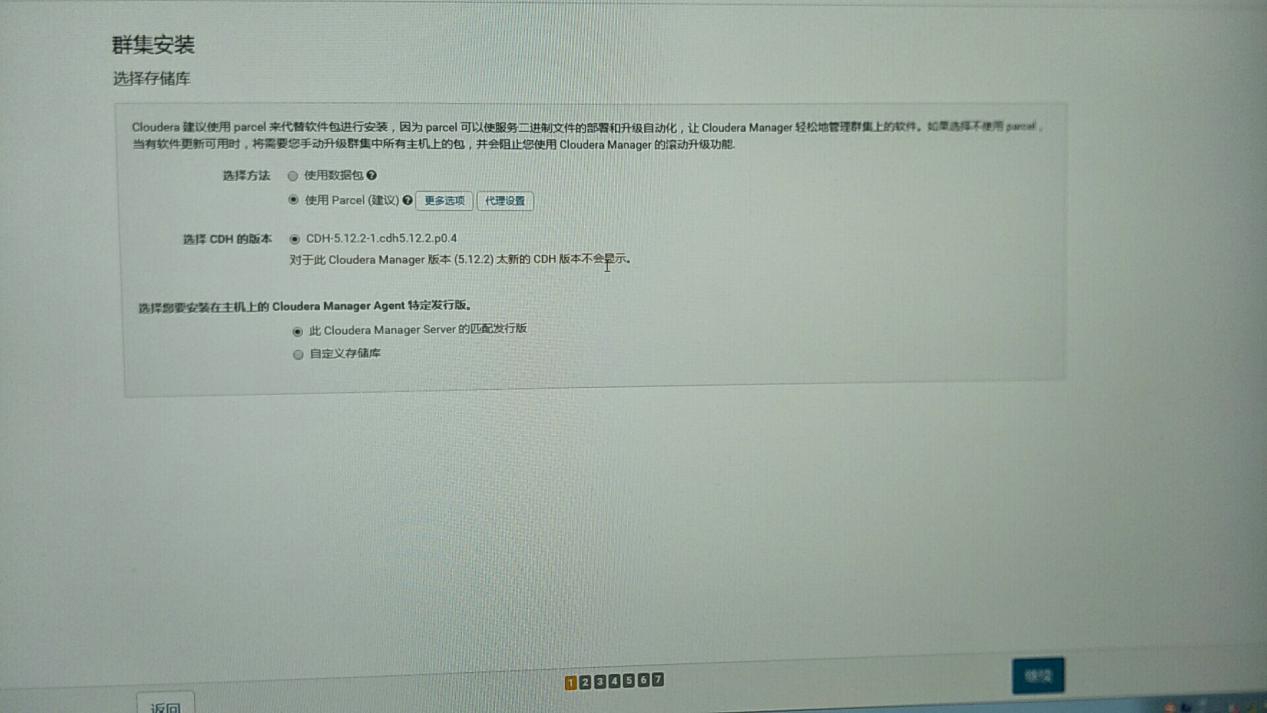
此处填写规则可以是网段型（例如：192.168.0.[140-145]）的，也可以是每行一个ip，填写完毕后点击搜索。

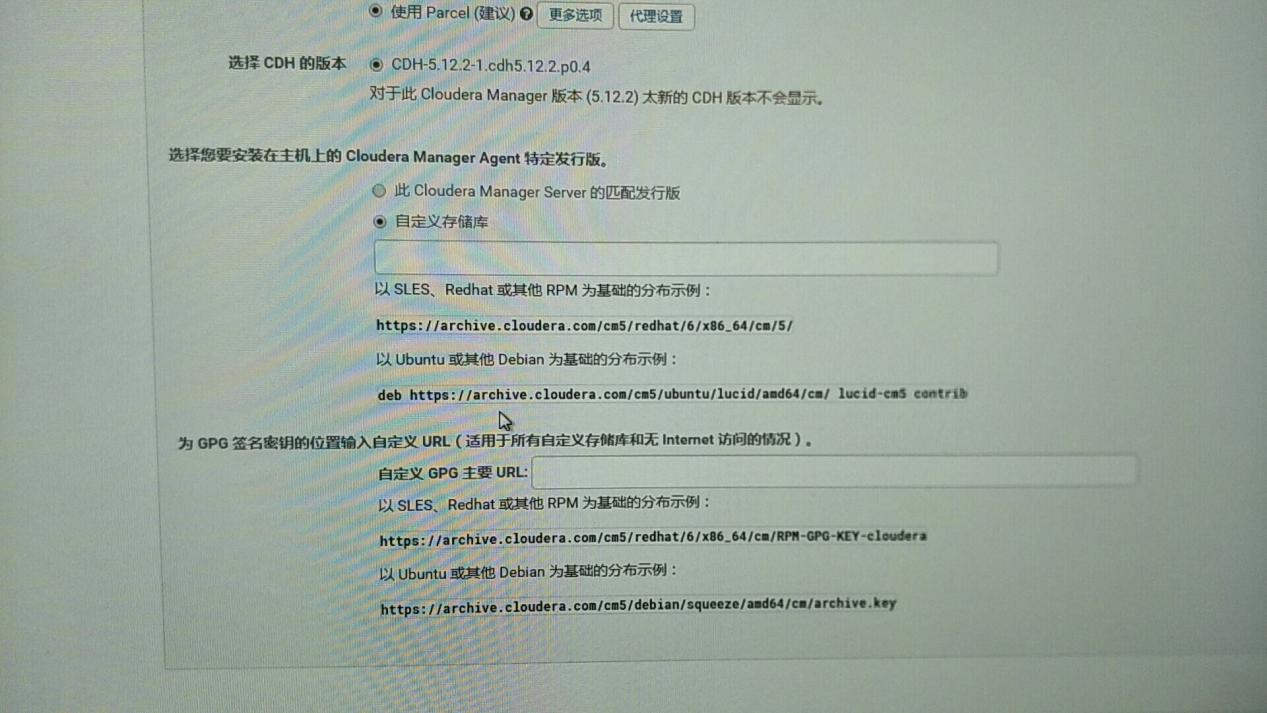
搜索完成后会出现被搜索到的ip，



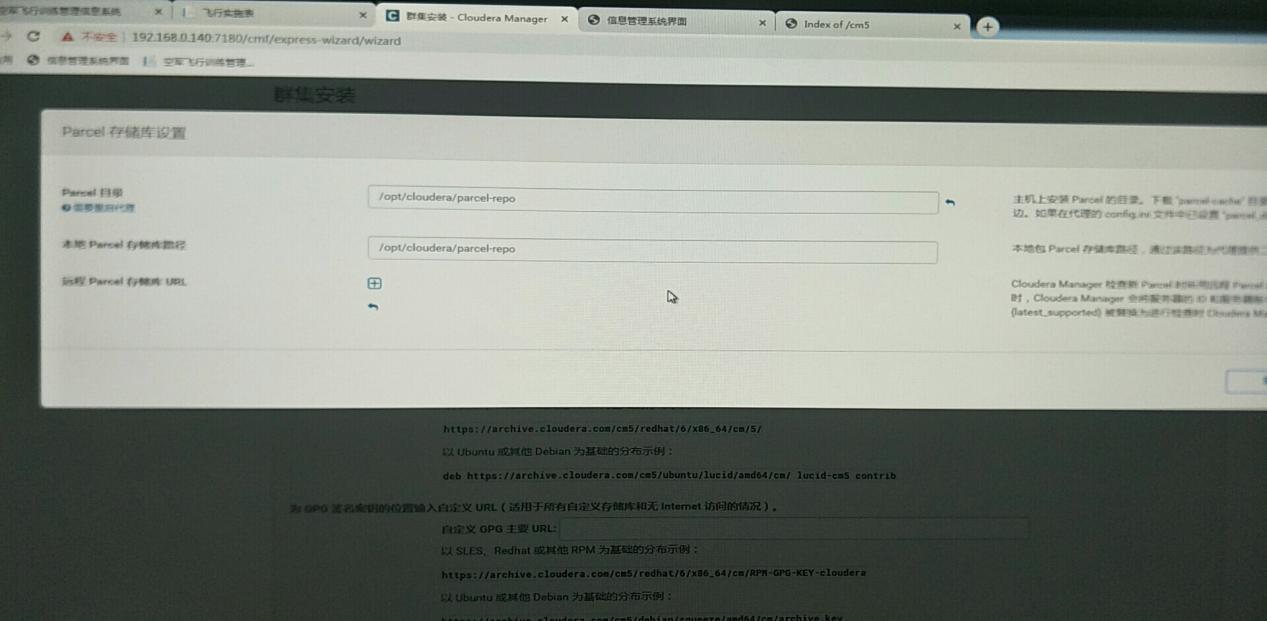
将搜索出来的ip勾选，点击“继续”进入下一步

1. 在集群安装目录下，“选择方法”选择“使用Parcel安装”
2. “特定发行本”处选择“自定义存储库”

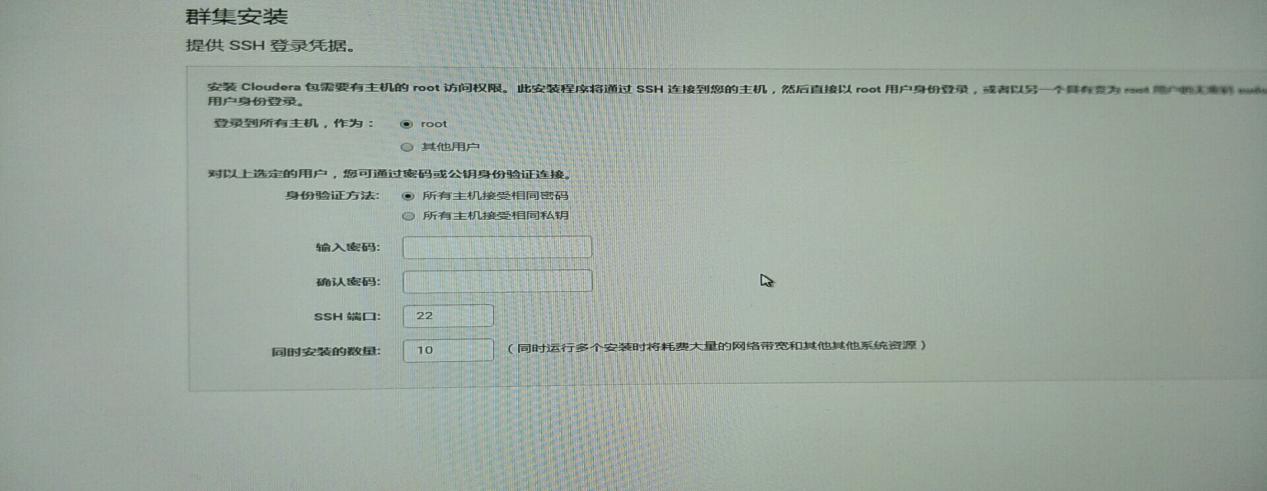




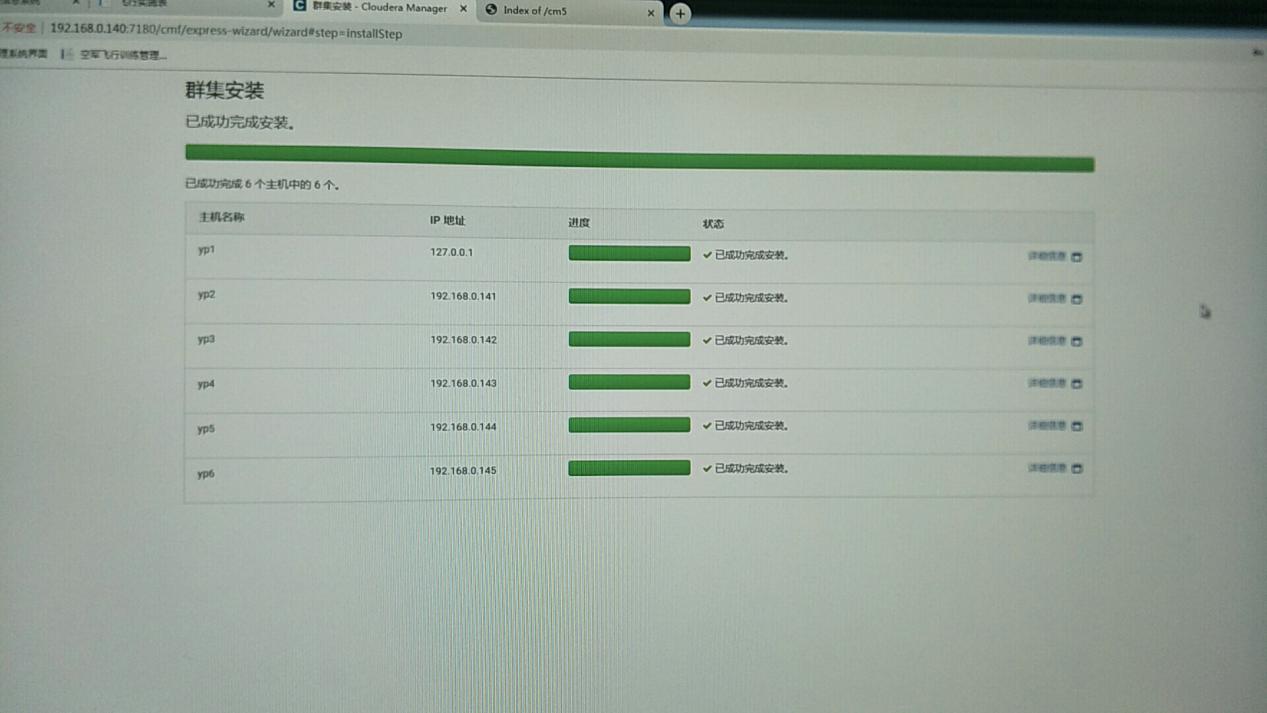
1. “使用Parcel”处“更多选项”，写入如下内容（注：已将 一、（三）、6 步骤中的parcel包放入到该目录下），删除其他自定义目录连接。保留内容如下



1. “自定义存储库”填“ip/cm5”（例如http://192.168.0.140/cm5/）
2. 下面的GPG的url不用管，“继续”下一步，会跳转到jdk安装，由于前面已经安装过jdk，且CDH自带JDK版本过低，所以不勾选安装，直接“继续”下一步，进入“ssh登录凭据”页面

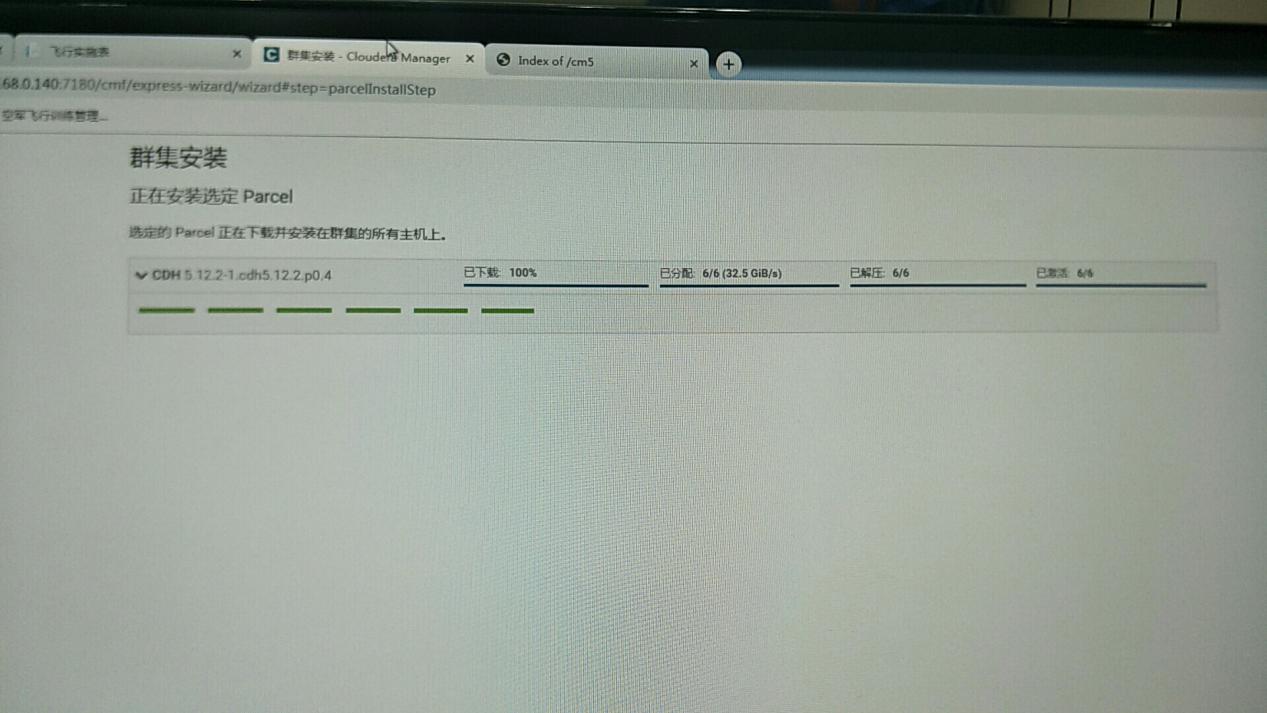


1. “登录到所有主机，作为”选择“root”
2. “身份验证方法”选择“所有主机接受相同的密码”，下面输入密码填入服务器密码（注：确保所有机器密码都是一样）
3. 确定后继续下一步开始给各个节点安装agent。安装完成后都是绿条：



注：该图中yp1的ip地址不对，原因是由于hosts文件配置问题，所以一定要将hosts文件配置正确。如果出现这种状况，退回第一步，修改好hosts文件后，通过yum -y remove cloudera-manager-agent\*卸载所有机器上的agent，然后再从头开始安装。直到ip正确且安装都已经通过为止。

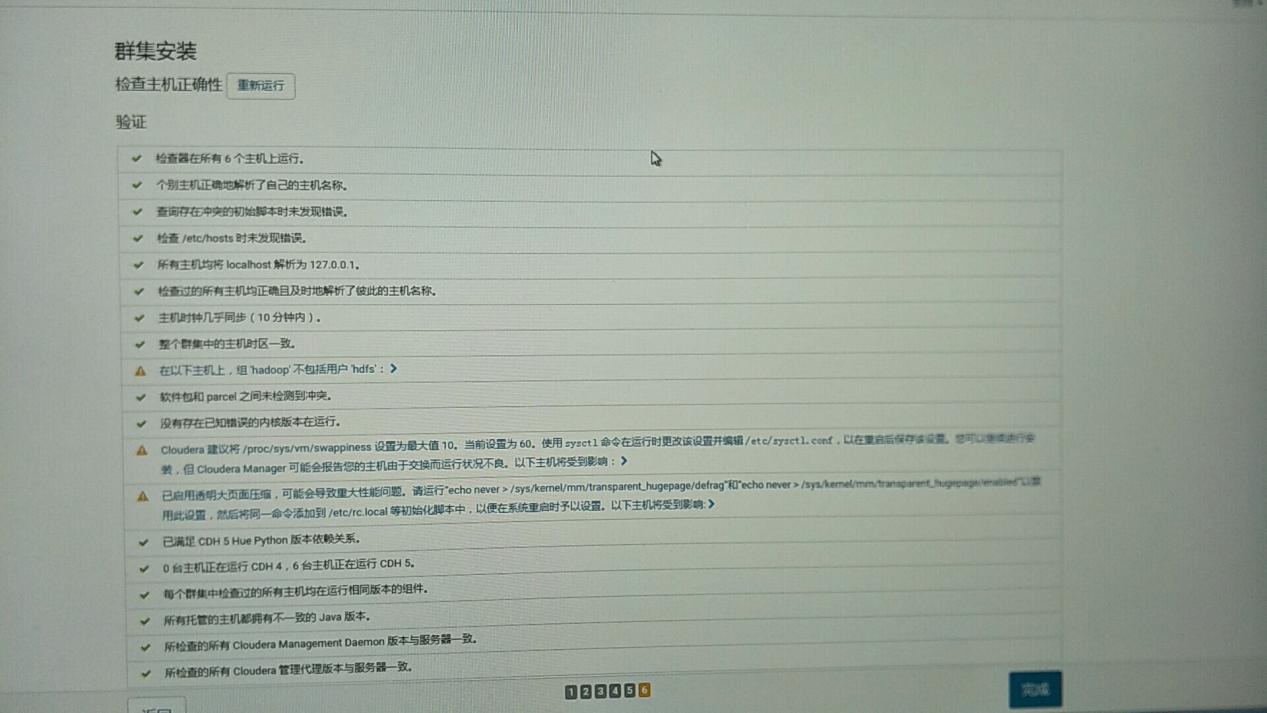
1. agent安装通过后，将是分发parcel文件：



注：分发时出现“运行状况不良”时，参考博客：

<https://blog.csdn.net/u010766519/article/details/90633887>

分发条都变绿代表分发完成，一般3-5分钟完成。然后“继续”下一步，出现下面的界面代表cdh安装完成。

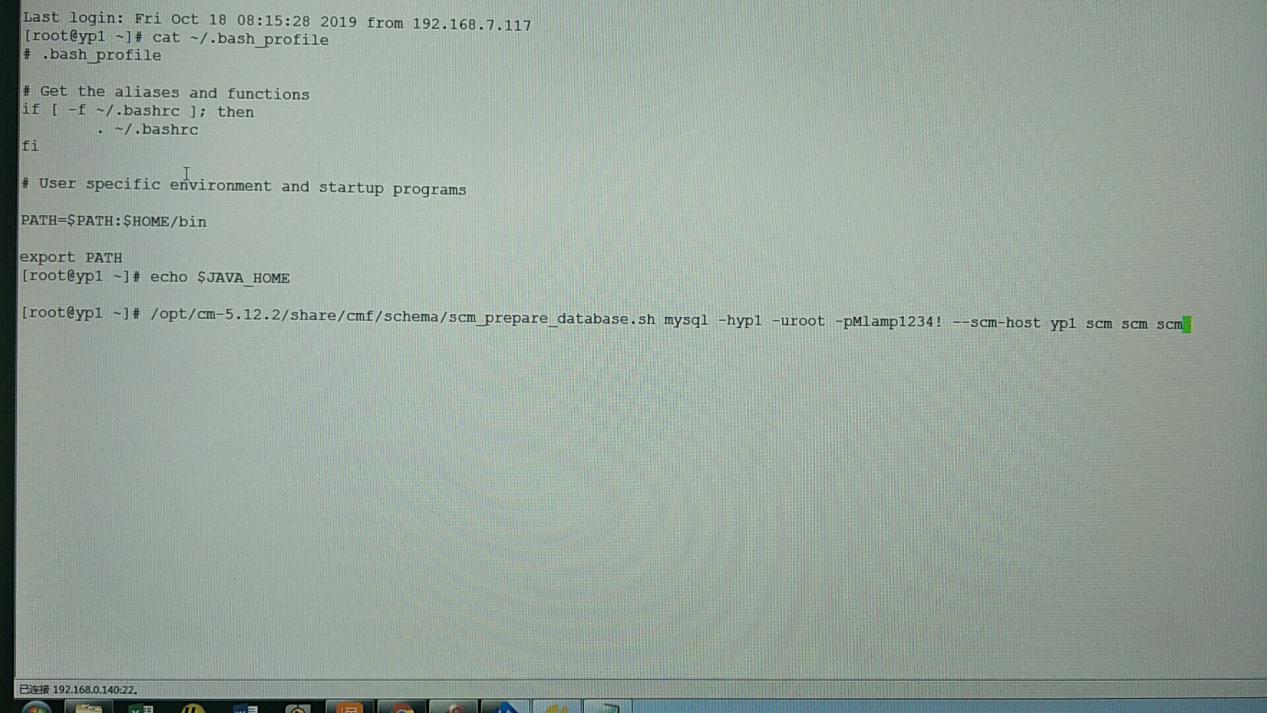


点击“完成”，可进行机器的服务安装。

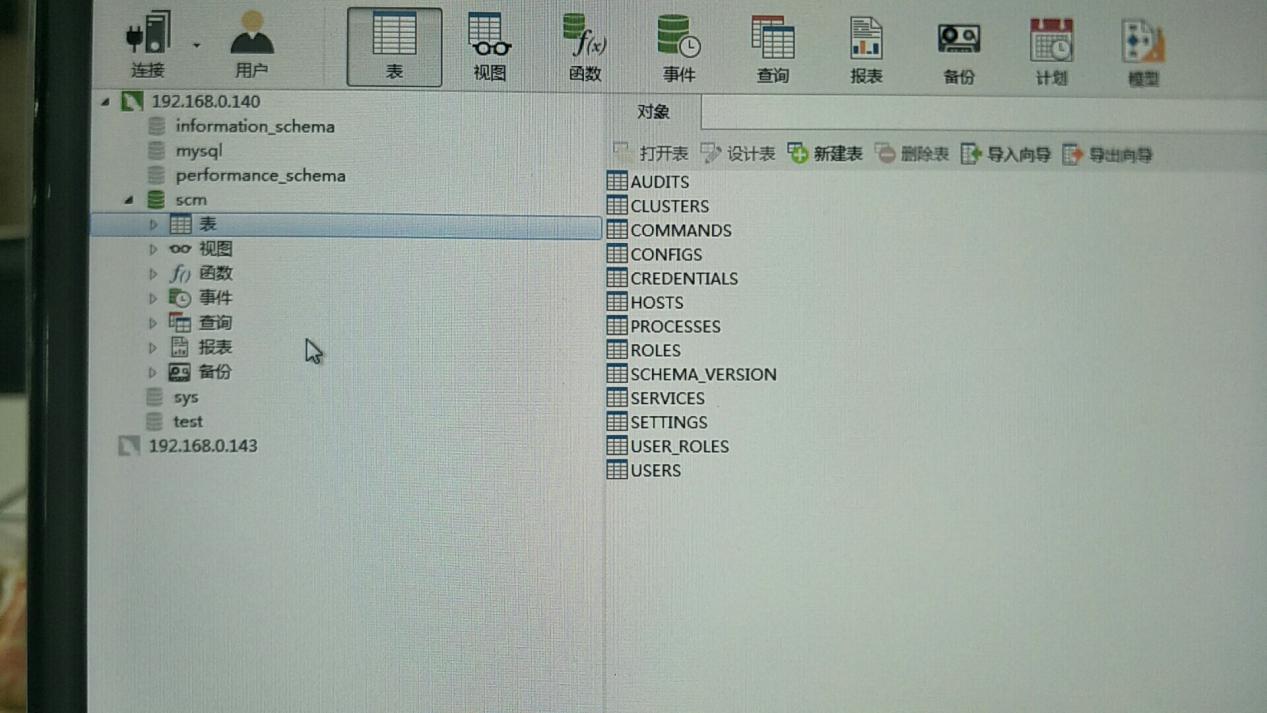
1. 初始化数据库（如果是安装集群而不是给集群中追加集群，这需要执行该步骤）

1、将mysql的jdbc驱动包放入到/opt/cm-5.12.2/share/cmf/lib/目录下

2、执行初始化数据库命令，初始化数据库命令如图：



如果初始化成功，则在该mysql下出现scm数据库，里面有少量的几张表，然后重启server

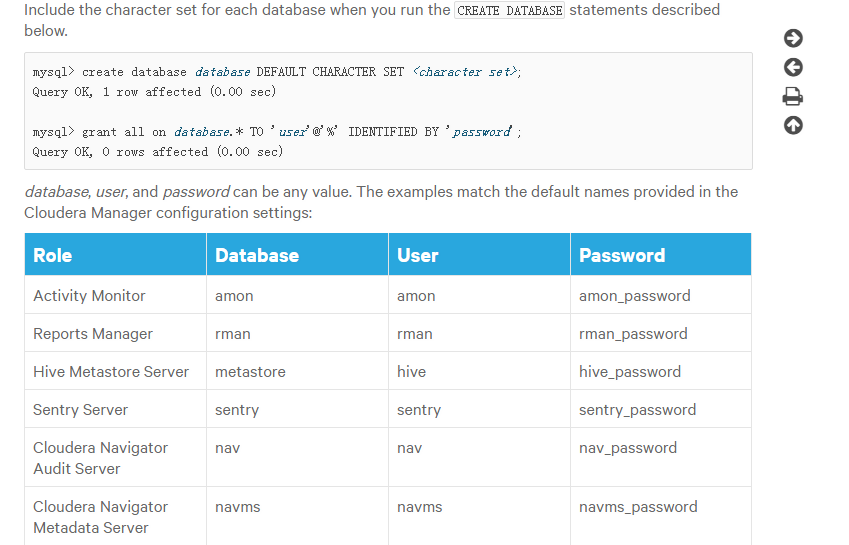


重启server后，scm库下的表会增多。

1. 创建组件服务所需要的数据库（要创建的数据库可参考官网：

https://docs.cloudera.com/documentation/enterprise/5-12-x/topics/cm\_ig\_mysql.html#concept\_dsg\_3mq\_bl）

要建立的库如图（来自于官网）：



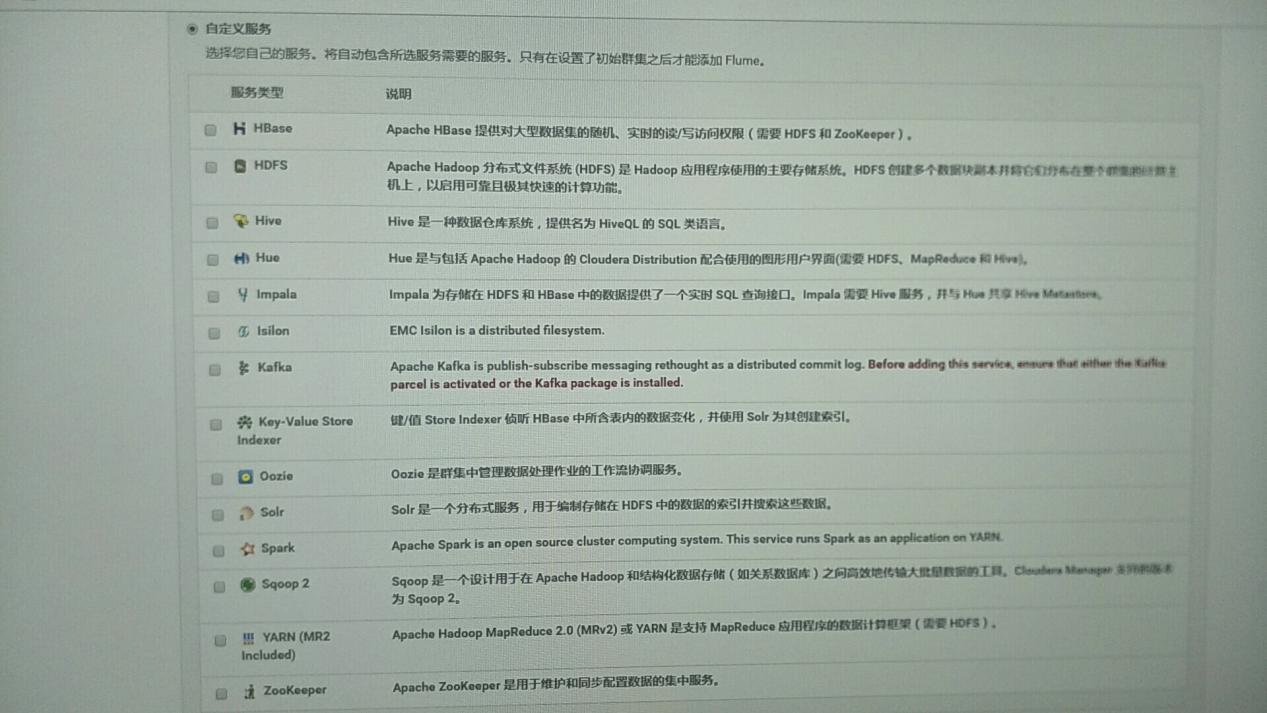
参考mysql命令也在官网中

1. 选择组件服务
2. 当parcel文件分发完成后，进入选择组件界面。

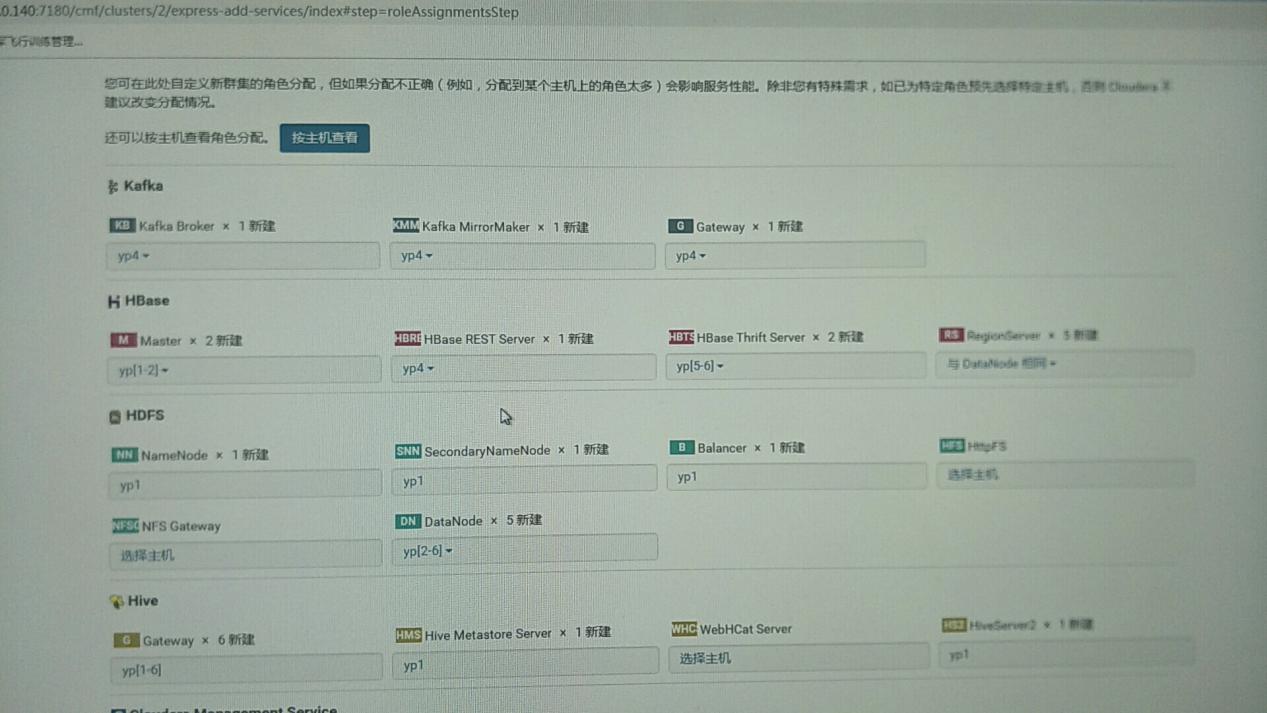


这里选择“自定义选择”，下一步

1. 出现服务选择界面，需要什么服务选择什么服务，按照事先规划好的来勾选。

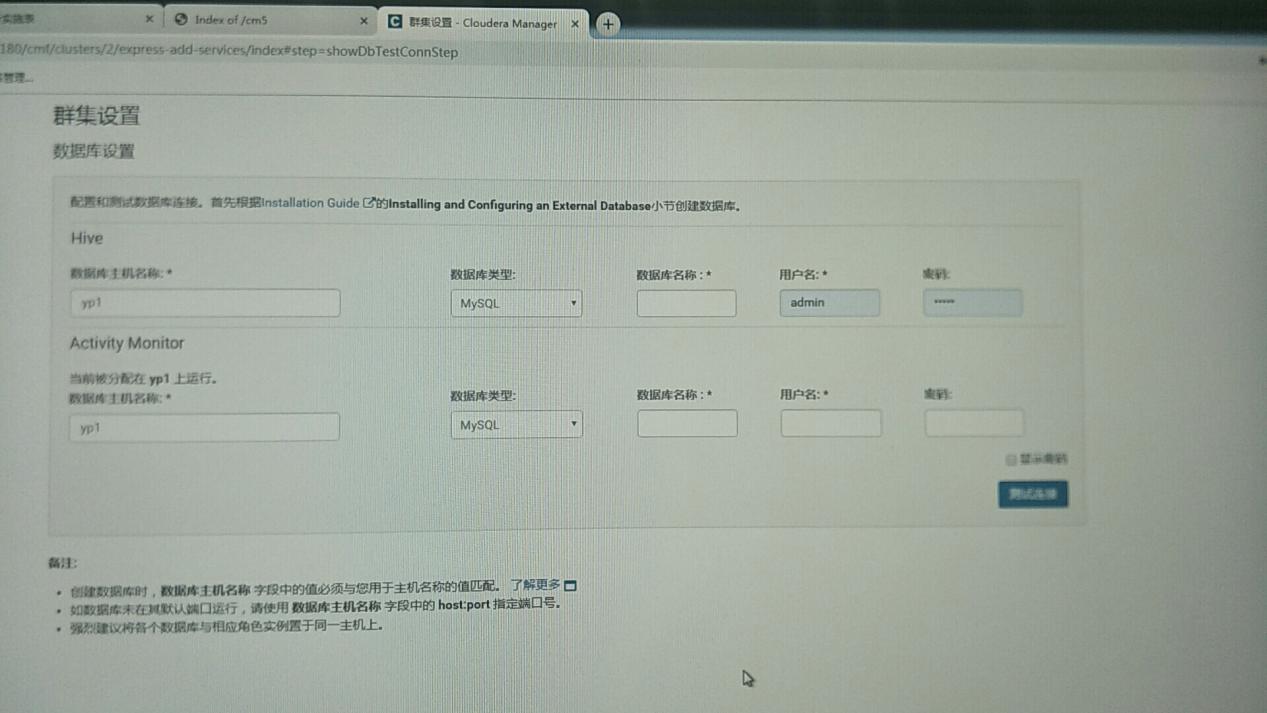


1. 勾选了对应服务后，进入到服务角色分配界面：



对应角色分配好对应机器。

1. 下一步进入数据库对应界面：



将之前建立的数据库与这两个对应，“测试连接”没问题后下一步进入到集群配置界面，按需进行配置。

配置完成后开始安装，如果安装过程中出现错误，查看角色日志，定位问题，然后网上找答案。

集群安装成功后，回到首页，将出现安装的各个组件，如果正常，则都是绿色。

