目录

[Hadoop-CDH集群安装部署 1](#_Toc18074238)

[1. CDH介绍 1](#_Toc18074239)

[2. 系统硬件要求 1](#_Toc18074240)

[3. 软件包准备（所有节点） 1](#_Toc18074241)

[4. CDH安装方式 3](#_Toc18074242)

[5. 安装包准备 3](#_Toc18074243)

[5.1. Cloudera Manager 3](#_Toc18074244)

[5.2. CDH安装包 4](#_Toc18074245)

[5.3. 上传 6](#_Toc18074246)

[6. 系统环境设置 6](#_Toc18074247)

[6.1. 修改网络主机名 6](#_Toc18074248)

[6.2. 创建hadoop用户hadoop组 7](#_Toc18074249)

[6.3. 修改主机名及主机映射(所有节点) 7](#_Toc18074250)

[6.4. 配置hadoop、root的ssh免密 8](#_Toc18074251)

[6.5. 安装jdk(所有节点) 9](#_Toc18074252)

[6.6. 关闭防火墙 12](#_Toc18074253)

[6.7. 关闭SELinux 13](#_Toc18074254)

[6.8. 设置open files 13](#_Toc18074255)

[6.9. 配置NTP时间同步服务 14](#_Toc18074256)

[6.10. 设置swap空间 16](#_Toc18074257)

[6.11. 关闭大页面压缩 16](#_Toc18074258)

[6.12. 安装mysql（只在管理节点或者其中一台机器上） 16](#_Toc18074259)

[6.13. 创建cdh 组件所需数据库 21](#_Toc18074260)

[7. 安装Cloudera Manager Server和 Agent 22](#_Toc18074261)

[7.1. 解压 22](#_Toc18074262)

[7.2. 初始化CM 22](#_Toc18074263)

[7.3. Agent配置 23](#_Toc18074264)

[7.4. cloudera-scm用户创建 24](#_Toc18074265)

[7.5. 启动scm-server和scm-agent 24](#_Toc18074266)

[8. WEB界面进行安装CDH 25](#_Toc18074267)

[8.1. 拷贝mysql驱动包至oozie和hive 25](#_Toc18074276)

[8.2. 登录web 26](#_Toc18074277)

[8.3. 组件服务选择 30](#_Toc18074278)

[8.4. 服务主机选择 31](#_Toc18074279)

[8.5. 测试元数据库 32](#_Toc18074280)

[8.6. 集群设置 34](#_Toc18074281)

# Hadoop-CDH集群安装部署

## CDH介绍

CDH (Cloudera's Distribution, including Apache Hadoop)，是Hadoop众多分支中的一种，由Cloudera维护，基于稳定版本的Apache Hadoop构建，并集成了很多补丁，可直接用于生产环境。并且提供了Hadoop的核心：可扩展存储和分布式计算，另外，CDH是基于web的用户界面操作，用户可以在CM页面上进行组件的添加删除，各服务的启动、停止，以及各组件的服务配置修改等。

## 系统硬件要求

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **节点类型** | **OS(系统)** | **CPU** | **内存(G)** | **OS盘大小(G)** | **网络** | **最小数据分区盘容量** |
| 管理节点 | centos 7.0 | 4cores | 16 | 64 | 1Gbps | / |
| Master节点 | centos 7.0 | 4cores | 16 | 128 | 1Gbps | 600 |
| Slave节点 | centos 7.0 | 4cores | 8 | 64 | 1Gbps | 2000 |

注：以上配置均为最低配置，如果硬件配置不达要求的话，安装好之后服务将启动不起来。如果服务器数量比较少的话，那管理节点与主节点部署同一台服务器上。同时，主节点也作为分析节点来使用（即slave节点），总的服务器节点数量不低于3台。

## 软件包准备（所有节点）

对于新的环境服务器，请检查以下服务或者工具包是否安装，如果没有安装则安装：

* yum -y install net-tools
* yum -y install unzip zip
* yum -y install psmisc

（pstree命令用）

* yum -y install httpd
* yum -y install mod\_ssl
* yum -y install python-lxml

（解决 hue是连不上的，会报错：Unexpected error. Unable to verify database connection.）

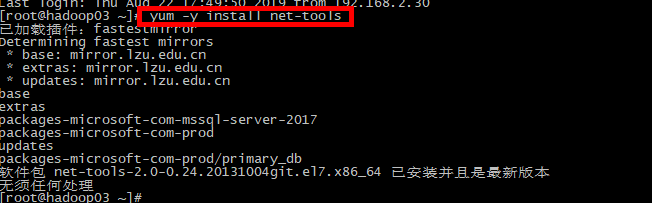
* yum -y install perl

(perl --version 查看版本)

* yum -y install gcc python-devel
* yum -y install cyrus-sasl\*

(impala 服务需要，然后重启集群的agent和集群服务)

* yum -y install telnet-server
* yum -y install telnet
* yum -y install snappy-devel
* yum -y install lzo-devel



注：以上安装要求服务器上开通外网环境，用于yum源的下载以及安装插件

## CDH安装方式

官方共给出了3种安装方式，地址：[https://www.cloudera.com/documentation/enterprise/latest/topics/installation.html](http://www.cloudera.com/content/cloudera/en/documentation/cloudera-manager/v5-latest/Cloudera-Manager-Installation-Guide/cm5ig_install_path_C.html)

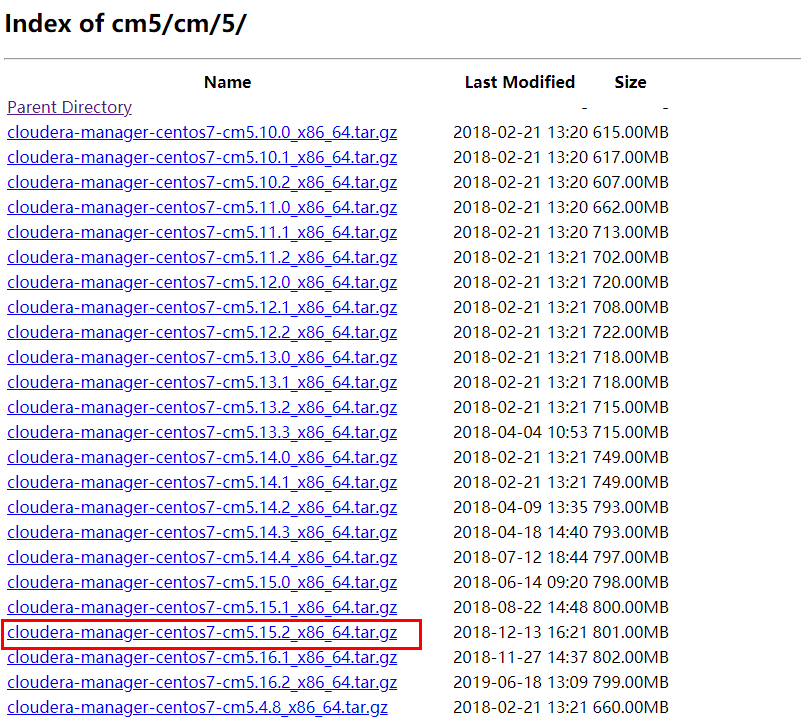
* 方法一：必须要求所有机器都能连网，由于各种国外的网站被墙的厉害，尝试了几次各种超时错误，耽误时间不说，一旦失败，重装非常痛苦。
* 方法二：下载很多包，不方便容易迷路。
* 方法三：对系统侵入性最小,可实现全离线安装，而且重装非常方便。后期的集群统一包升级也非常好。（本篇采用该方法安装）

## 安装包准备

本次安装**5.15.2**版本

### Cloudera Manager

<http://archive.cloudera.com/cm5/cm/5/cloudera-manager-centos7-cm5.15.2_x86_64.tar.gz>

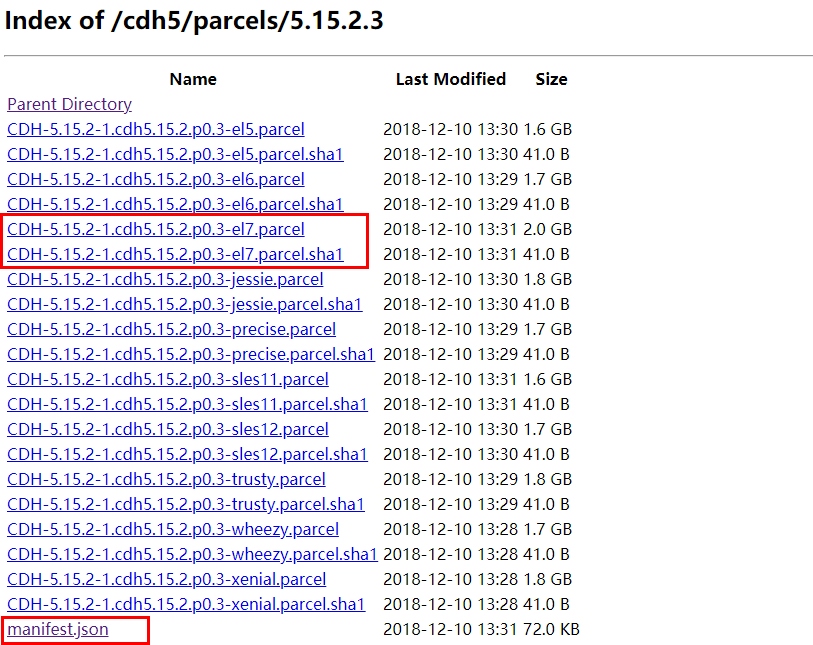


### CDH安装包

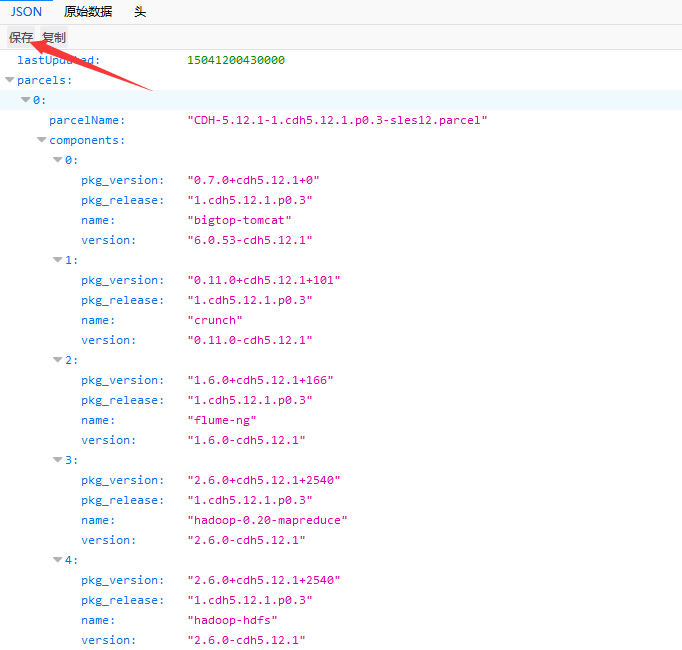
需下载3个文件（CDH-5.15.2-1.cdh5.15.2.p0.3-el7.parcel、[CDH-5.15.2-1.cdh5.15.2.p0.3-el7.parcel.sha1](http://archive.cloudera.com/cdh5/parcels/5.15.2/CDH-5.15.2-1.cdh5.15.2.p0.3-el7.parcel.sha1)、manifest.json、）

地址如下：

[http://archive.cloudera.com/cdh5/parcels/5.15.2/](http://archive.cloudera.com/cdh5/parcels/5.15.2/CDH-5.15.2-1.cdh5.15.2.p0.3-el7.parcel)



manifest.json不能直接下载，需打开复制里面的内容，另存为manifest.json 文件。



### 上传

上传cloudera-manager-centos7-cm5.15.2\_x86\_64.tar.gz、CDH-5.15.2-1.cdh5.15.2.p0.3-el7.parcel、CDH-5.15.2-1.cdh5.15.2.p0.3-el7.parcel.sha1、manifest.json 这4个文件至主节点 hadoop01 的 /opt/

## 系统环境设置

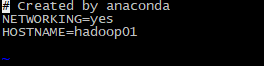
**以下操作全部为root用户执行，而且所有节点都需配置**

### 修改网络主机名

vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=hadoop01



scp /etc/sysconfig/network 发送到其他的节点

重启网络服务生效：service network restart

### 创建hadoop用户hadoop组

之后所有的操作的基于hadoop用户去执行

groupadd hadoop

useradd -g hadoop -d /home/hadoop hadoop

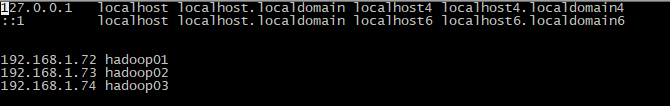
passwd hadoop

### 修改主机名及主机映射(所有节点)

vi /etc/hostname



vi /etc/hosts

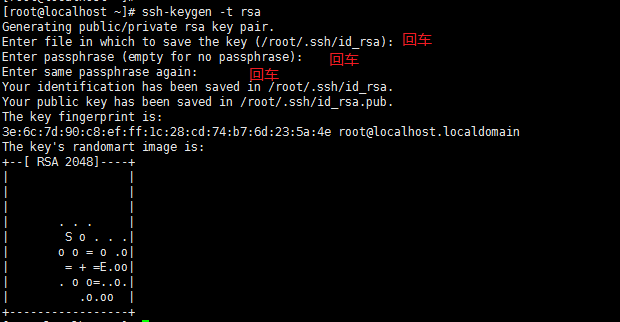


除hadoop01以外，其他的节点对应修改

### 配置hadoop、root的ssh免密

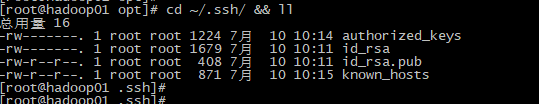
1. root免密

ssh-keygen -t rsa （生成公钥和私钥）



生成之后会在用户的根目录生成一个 “.ssh”的文件夹

cd ~/.ssh （会生成以下文件）



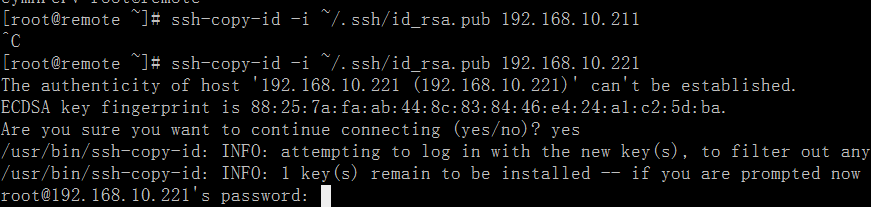
分发公钥至各服务器，每台节点执行（包括本节点）：

ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop01

ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop02

ssh-copy-id -i ~/.ssh/id\_rsa.pub hadoop03

……



第一次需要输入密码

1. hadoop用户免密

切换至hadoop用户

su – Hadoop



接下来方法同上root用户

### 安装jdk(所有节点)

首先卸载自带的OpenJdk，有些linux会默认自带较低版本的jdk如Centos。

查找java相关的包：

rpm -qa | grep java



卸载低版本java包：

rpm -e -nodeps \*java\*

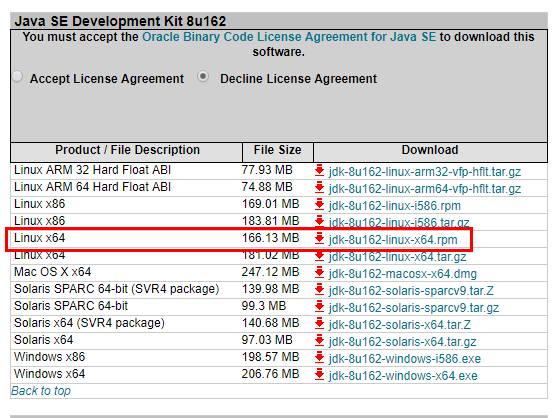
如果rpm -qa | grep java什么都没有则没有默认安装的jdk，直接进行下一步即可

去Oracle的官网下载jdk的rpm安装包，并使用rpm -ivh 包名安装之。

下载地址：

<https://www.oracle.com/technetwork/java/javase/downloads/java-archive-javase8-2177648.html>

不建议使用最新版本的jdk，因为oracle官网更新的jdk版本要高于CDH官方推荐的最高版本，会出现兼容性警告。



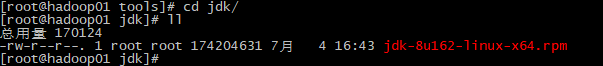
这里我们选择小版本162的rpm安装包进行安装。选择rpm软件包的安装方式好处有二：

第一：rpm安装的默认路径即是CDH安装时会去寻找jdk的路径，否则在CM的安装中会报错。

注：如果用tar包手动的方式安装可以创建软连接 进入/usr/java目录[没有就手动创建之]：ln -s你的jdk安装目录 /usr/java/default,运用此方法别忘记手动配置环境变量

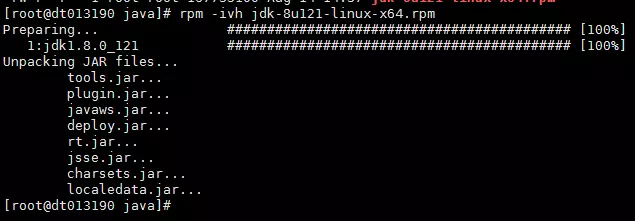
第二：rpm安装会自动创建环境变量。

将我们下载好的软件包上传至管理节点，利用rsync命令，同步其他节点



使用rpm -ivh 包名 安装之

rpm -ivh jdk-8u162-linux-x64.rpm

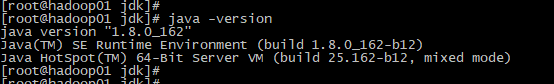


由于是rpm包并不需要我们来配置环境变量，我们只需要配置一个全局的JAVA\_HOME变量即可，执行命令：

echo "JAVA\_HOME=/usr/java/latest/" >> /etc/environment

source /etc/environment

最后执行命令java -version出现如下画面即安装成功

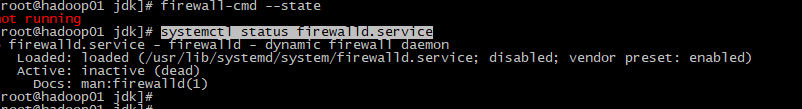


最后，重复如上动作，给剩余其他节点都安装好jdk。

### 关闭防火墙

查看防火墙状态：

firewall-cmd –state 或者 systemctl status firewalld.service



停止firewall：

systemctl stop firewalld.service

禁止firewall开机启动：

systemctl disable firewalld.service



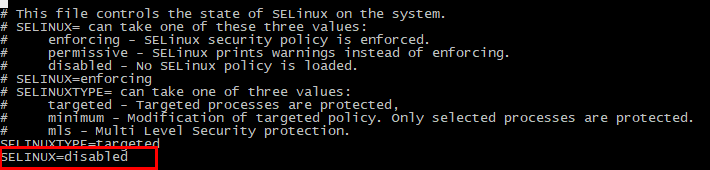
注：在所有的节点上执行，因为涉及到的端口太多，临时关闭防火墙是为了安装起来更方便。安装完毕后可以根据需要设置防火墙策略，保证集群安全。

### 关闭SELinux

注：需重启机器生效

vi /etc/selinux/config

SELINUX=enforcing 修改为 SELINUX=disabled

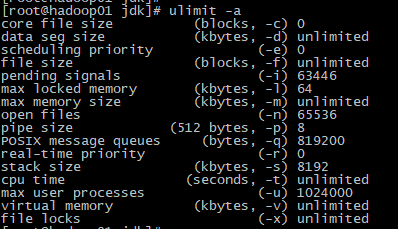


重启后使用如下命令检查一下：sestatus -v



### 设置open files

ulimit -a，查看当前系统的配置的上限，unlimited为无上限



修改参数 （重启生效）（所有节点）

vi /etc/security/limits.conf

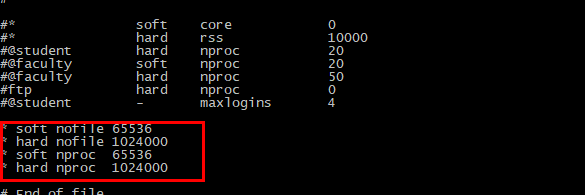
最后面添加：

\* soft nofile 65536

\* hard nofile 1024000

\* soft nproc 65536

\* hard nproc 1024000



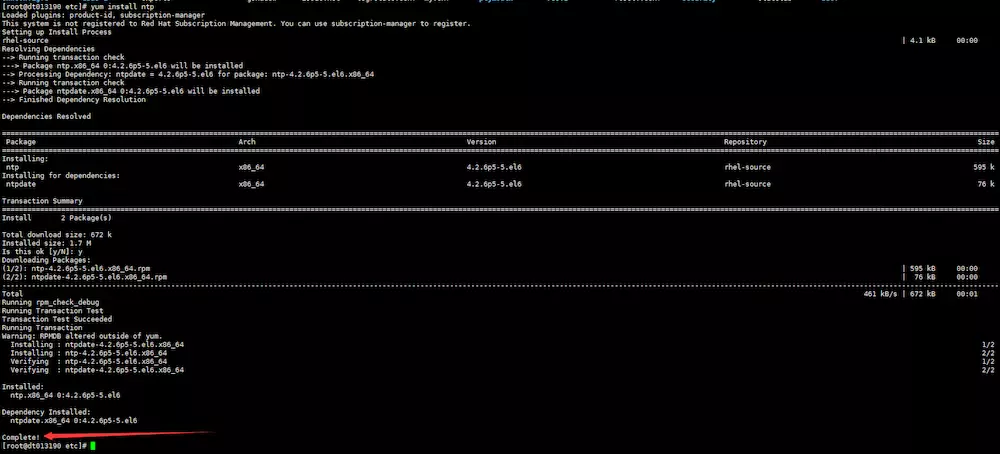
注：提示：\*代表所有用户，@表示对某个用户组生效，直接写用户名表示对某个用户生效

### 配置NTP时间同步服务

注：这一步暂时可以跳过，但是在集群启动以后，会提示时间不同步。集群中所有主机必须保持时间同步，如果时间相差较大会引起各种问题。

安装NTP（每个节点）

yum install ntp



根据提示输入Y后待到出现Complete！字样则表示为成功 。

6.2 配置NTP

先将原有的ntp.conf配置文件进行备份 cp /etc/ntp.conf /etc/ntp.conf.backup（所有节点）

修改ntp.conf文件如下

vi /etc/ntp.conf

maste配置：

restrict 192.168.1.0 mask 255.255.255.0 nomodify notrapserver 127.127.1.0fudge 127.127.1.0 stratum 10

slave配置：server master（此处为master节点的hostname）

执行命令（所有节点）

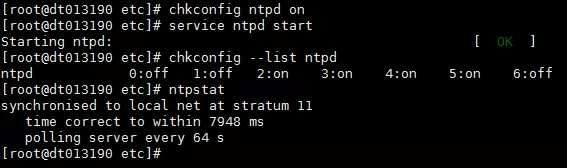
设置开机启动:

chkconfig ntpd on

启动ntp：service ntpd start

检查是否设置成功：chkconfig --list ntpd

其中2-5为on状态就代表成功。



### 设置swap空间

Cloudera官方文档：建议将交换空间设置为0，过多的交换空间会引起GC耗时的激增，所以还是关闭的为好。

执行命令 （所有节点）

echo "vm.swappiness = 0" >> /etc/sysctl.conf

### 关闭大页面压缩

执行命令 （所有节点）

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

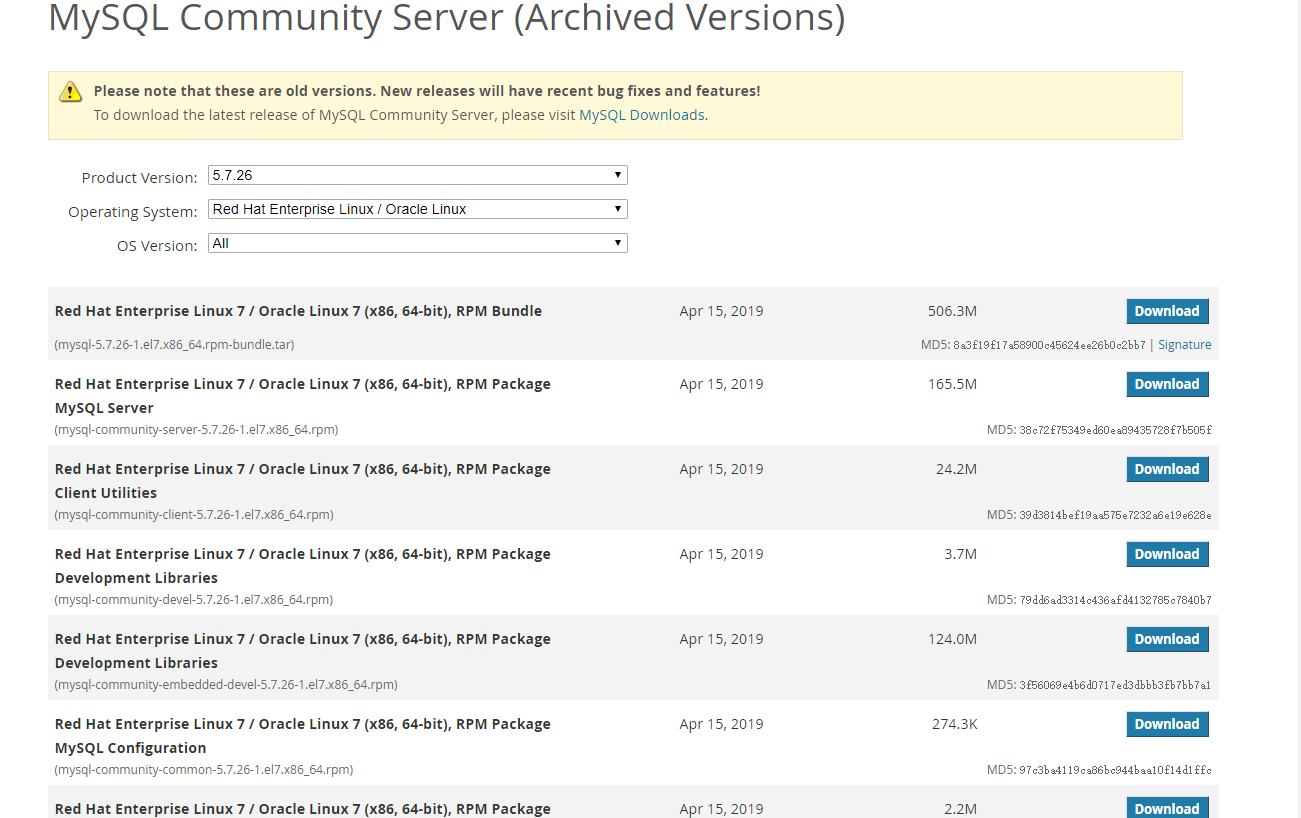
echo never > /sys/kernel/mm/transparent\_hugepage/enabled

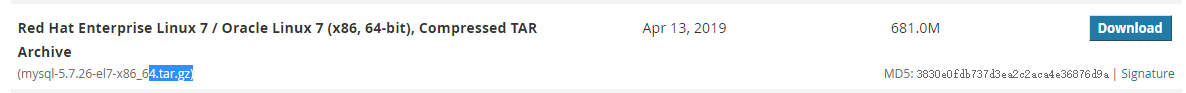
### 安装mysql（只在管理节点或者其中一台机器上）

下载mysql

最新版本地址：<https://dev.mysql.com/downloads/mysql/>

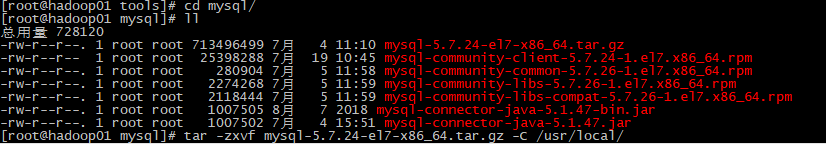
历史版本地址：<https://downloads.mysql.com/archives/community/>





上传tar包到我们建好的安装目录下，解压缩到/usr/local/ 目录下

tar -zxvf mysql-5.7.24-el7-x86\_64.tar.gz -C /usr/local/



cd /usr/local/ && mv mysql-5.7.24-el7-x86\_64/ mysql

创建mysql database目录

mkdir -p /usr/local/mysql/data

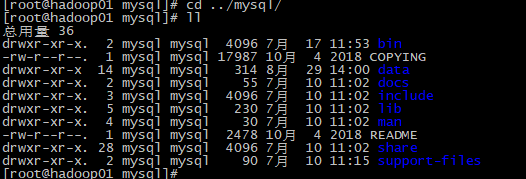
创建mysql用户用户组及目录

groupadd mysql

useradd -r -s /sbin/nologin -g mysql mysql -d /usr/local/mysql

修改mysql的用户目录权限

chown -R mysql:mysql /usr/local/mysql



初始化mysql

/usr/local/mysql/bin/mysqld --initialize --user=mysql --basedir=/usr/local/mysql --datadir=/usr/local/mysql/data

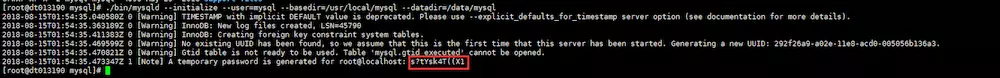
部分人的linux系统会出现如下错误

这是因为缺少安装包libaio和libaio-devel导致的

yum install libaio\*



再次执行初始化，并且记录初始化时生成的密码，如下图红色标注：



修改配置文件: vi /etc/my.cnf

添加以下内容：

[mysql]

default-character-set=utf8

[mysqld]

port = 3306

basedir = /usr/local/mysql

datadir = /usr/local/mysql/data

max\_connections = 1000

max\_allowed\_packet = 64M

sort\_buffer\_size = 32M

join\_buffer\_size = 32M

thread\_cache\_size = 16

query\_cache\_size = 128M

query\_cache\_limit = 4M

max\_heap\_table\_size = 128M

lower\_case\_table\_names = 1

table\_open\_cache = 4096

character-set-server=utf8

init\_connect='SET NAMES utf8'

[mysqld\_safe]

sql\_mode=NO\_ENGINE\_SUBSTITUTION,STRICT\_TRANS\_TABLES

# include all files from the config directory

!includedir /etc/my.cnf.d

启动mysql

/usr/local/mysql/support-files/mysql.server start

进入mysql，初始化生成的密码

mysql -uroot -p

修改密码并且设置权限：

SET password for root@localhost = password('root');

update mysql.user set host='%' where user='root';

grant all privileges on \*.\* to 'root'@'%' identified by 'root' with grant option;

grant all privileges on \*.\* to 'hadoop'@'%' identified by 'hadoop' with grant option;

flush privileges;

quit;

设置开机自启动

cp /usr/local/mysql/support-files/mysql.server /etc/init.d/mysql

或者

ln -s /usr/local/mysql/support-files/mysql.server /etc/init.d/mysql

chmod +x /etc/init.d/mysql

chkconfig --add mysql --将MySQL添加为受chkconfig管理的服务

chkconfig --level 345 mysql on --在级别3和5上设定服务为”on”

chkconfig --list | grep mysql



mysql命令加入到环境

ln -s /usr/local/mysql/bin/mysql /usr/bin

重启mysql生效

service mysql restart

### 创建cdh 组件所需数据库

create database hive default charset utf8 collate utf8\_general\_ci;

create database amon default charset utf8 collate utf8\_general\_ci ;

create database hue default charset utf8 collate utf8\_general\_ci;

create database oozie default charset utf8 collate utf8\_general\_ci;

## 安装Cloudera Manager Server和 Agent

### 解压

将我们下载好的CM安装包(cloudera-manager-centos7-cm5.15.2\_x86\_64.tar.gz)上传至服务器并解压安装(所有节点)

cloudera manager的目录默认在/opt下，cdh5的源会默认在/opt/cloudera/parcel-repo寻找，所以不能解压到其他地方，使用命令：

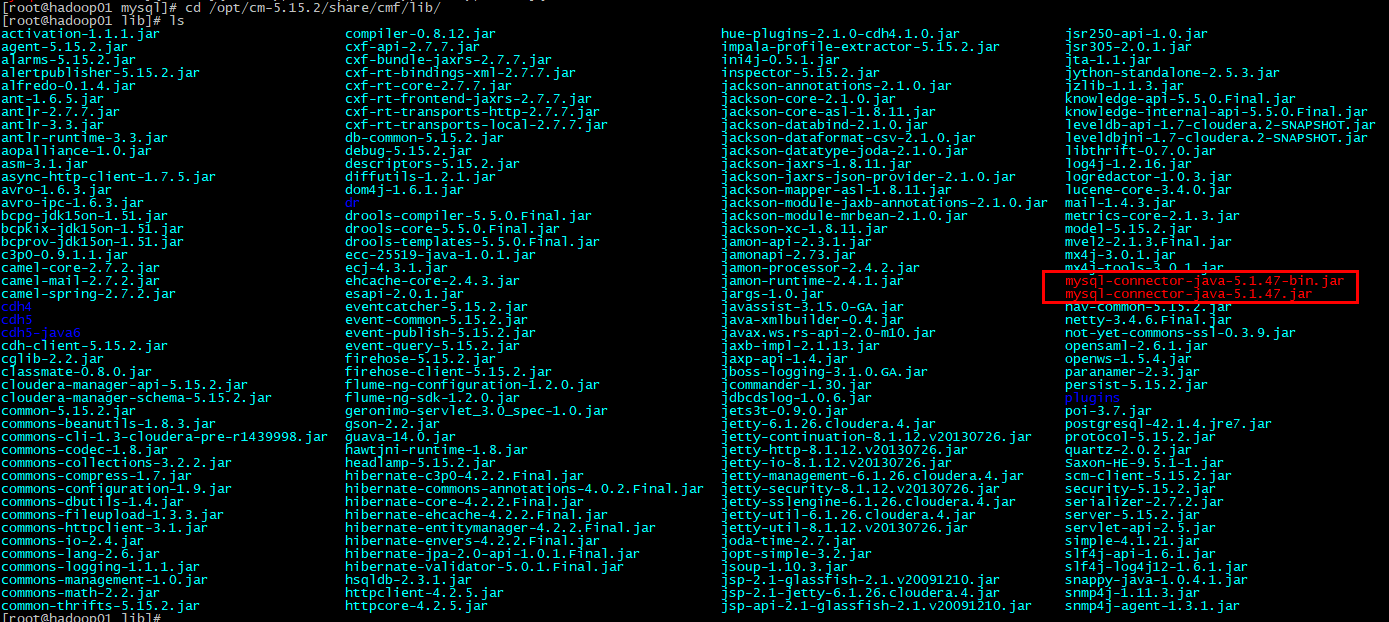
tar -zxvf cloudera-manager\*.tar.gz

mv cloudera /opt/

mv cm-5.15.2 /opt/

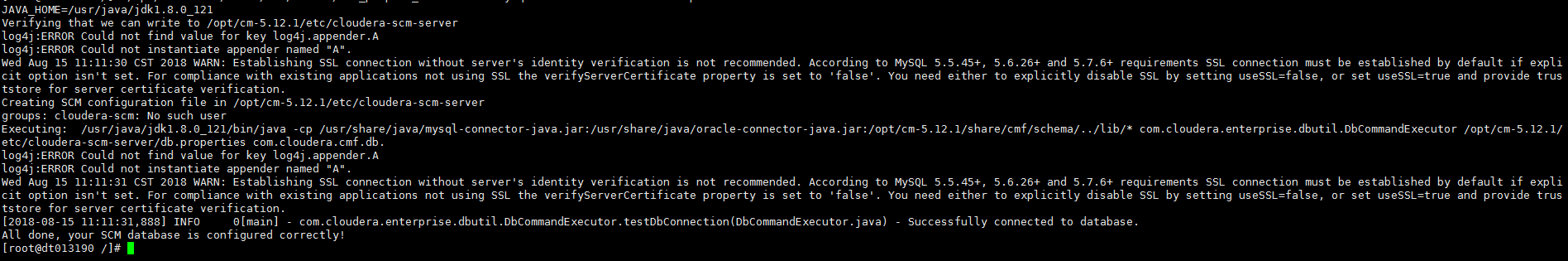
### 初始化CM

接下来的安装，cm需要与mysql进行交互，所以首先需要去MySql的官网下载JDBC驱动，http://dev.mysql.com/downloads/connector/j/，解压后找到mysql-connector-java-5.1.\*\*-bin.jar，放到/opt/cm-5.15.2/share/cmf/lib/中。



在主节点初始化CM5的数据库：

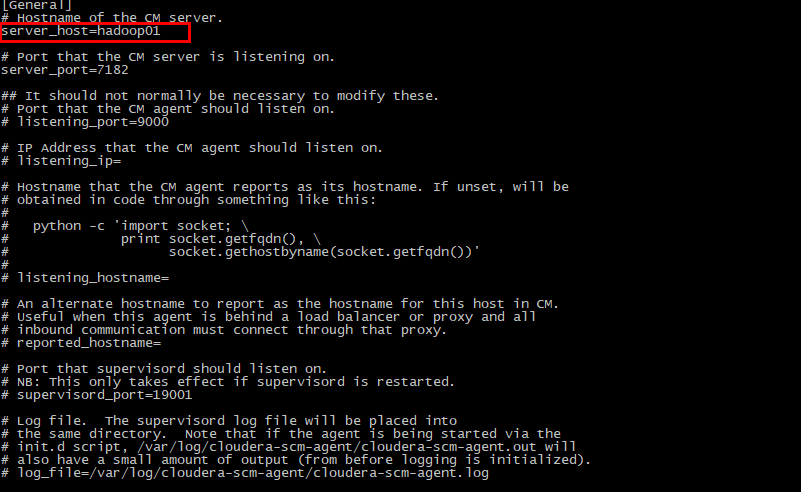
/opt/cm-5.15.2/share/cmf/schema/scm\_prepare\_database.sh mysql cm -hhadoop01 -uroot -proot --scm-host hadoop01 scm scm scm



### Agent配置

/opt/cm-5.15.2/etc/cloudera-scm-agent/config.ini中的server\_host为主节点的主机名为hadoop01。

vi /opt/cm-5.15.2/etc/cloudera-scm-agent/config.ini



同步Agent到其他节点

scp /opt/cm-5.15.2/etc/cloudera-scm-agent/config.ini hadoop02: /opt/cm-5.15.2/etc/cloudera-scm-agent/

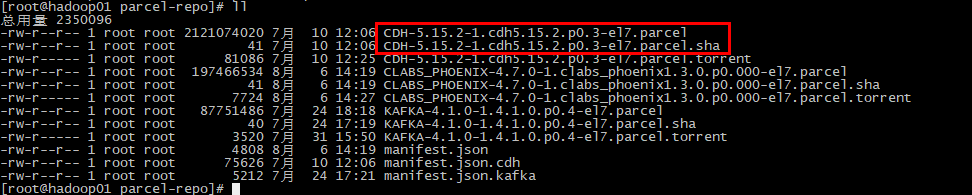
scp /opt/cm-5.15.2/etc/cloudera-scm-agent/config.ini hadoop03: /opt/cm-5.15.2/etc/cloudera-scm-agent/

### cloudera-scm用户创建

在所有节点创建cloudera-scm用户

useradd --system --home=/opt/cm-5.15.2/run/cloudera-scm-server --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm

将我们提前下载好的Parcels文件上传到主节点的/opt/cloudera/parcel-repo/目录中（如果没有parcel-repo需要手动创建，这里不需要分发到所有节点）。



注意！！最后一定要将CDH-5.12.1-1.cdh5.12.1.p0.3-el6.parcel.sha1，重命名为CDH-5.12.1-1.cdh5.12.1.p0.3-el6.parcel.sha，去掉末尾的1，否则系统会重新下载CDH-5.12.1-1.cdh5.12.1.p0.3-el6.parcel文件，也就不是纯离线安装了。

### 启动scm-server和scm-agent

通过启动服务端。

/opt/cm-5.15.2/etc/init.d/cloudera-scm-server restart

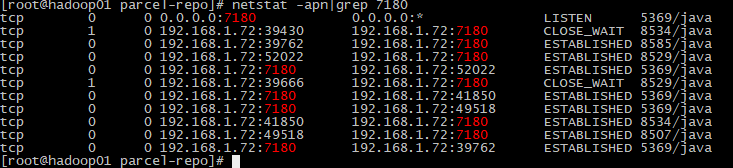
通过启动Agent服务（所有Agent节点）

/opt/cm-5.15.2/etc/init.d/cloudera-scm-agent restart

以上我们是通过脚本来启动的，所以需要停止服务的时候，只需要把以上的start参数改为stop就可以啦。

service服务启动需要几分钟，取决于服务器的性能，此时我们可以使用以下命令查看启动进程：

netstat -apn|grep 7180



查询端口被占用后，就可以用浏览器打开http://hadoop01:7180/进行访问了，登录用户名与密码分别为admin。

## WEB界面进行安装CDH



### 拷贝mysql驱动包至oozie和hive

在页面执行之前 copy mysql 的 jdbc 驱动到hive和oozie

cp /home/tools/mysql/mysql-connector-java-5.1.47-bin.jar /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hive/lib/

scp /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hive/lib/mysql-connector-java-5.1.47-bin.jar hadoop02:/opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hive/lib

scp /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hive/lib/mysql-connector-java-5.1.47-bin.jar hadoop03:/opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hive/lib

cp /home/tools/mysql/mysql-connector-java-5.1.47-bin.jar /var/lib/oozie/mysql-connector-java.jar

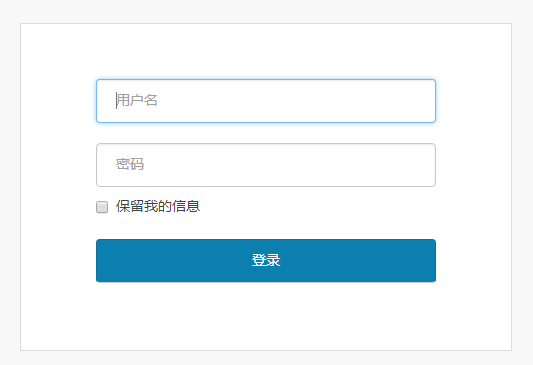
scp /var/lib/oozie/mysql-connector-java.jar hadoop02:/var/lib/oozie

scp /var/lib/oozie/mysql-connector-java.jar hadoop03:/var/lib/oozie

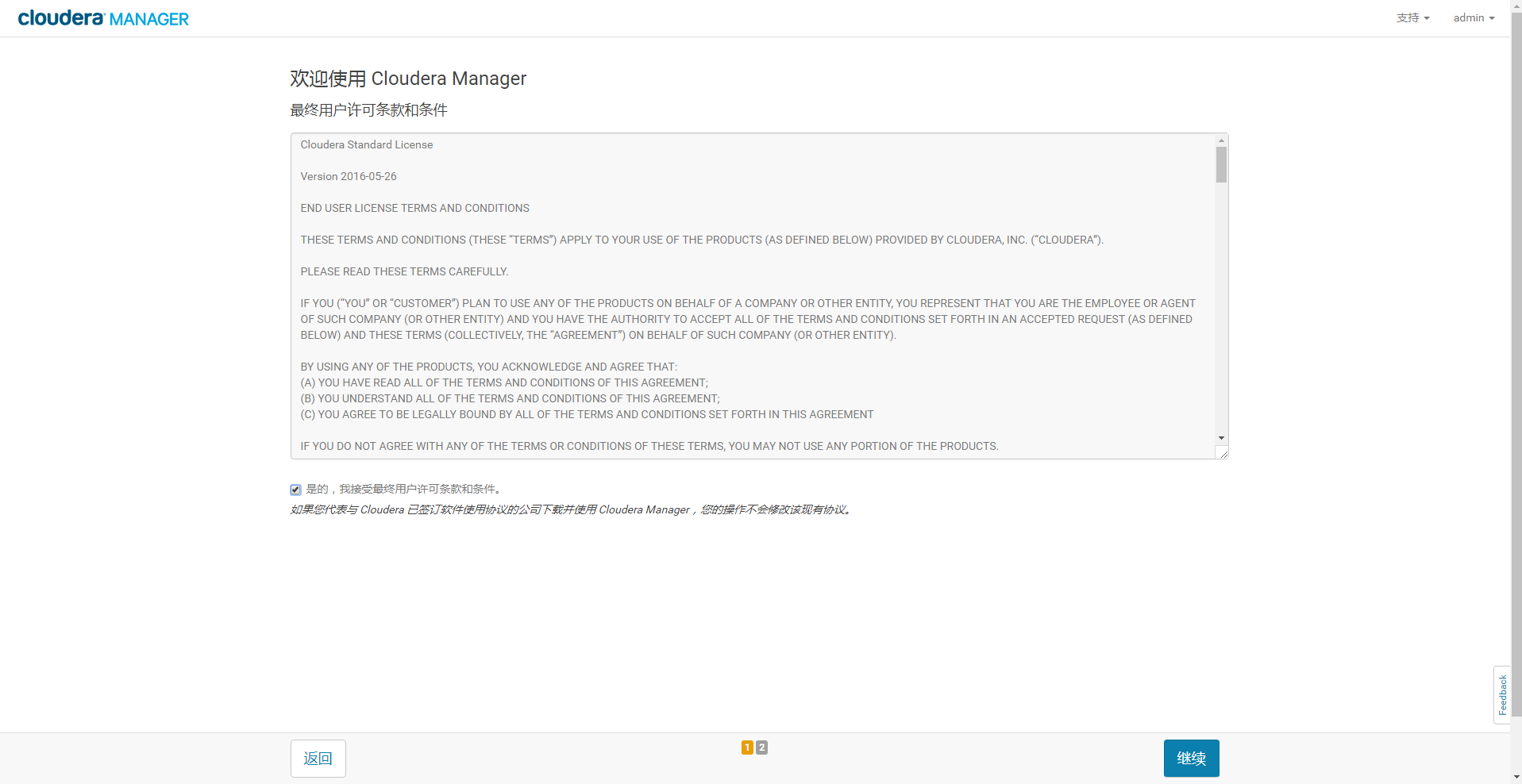
### 登录web

登录页面：<http://hadoop01:7180>

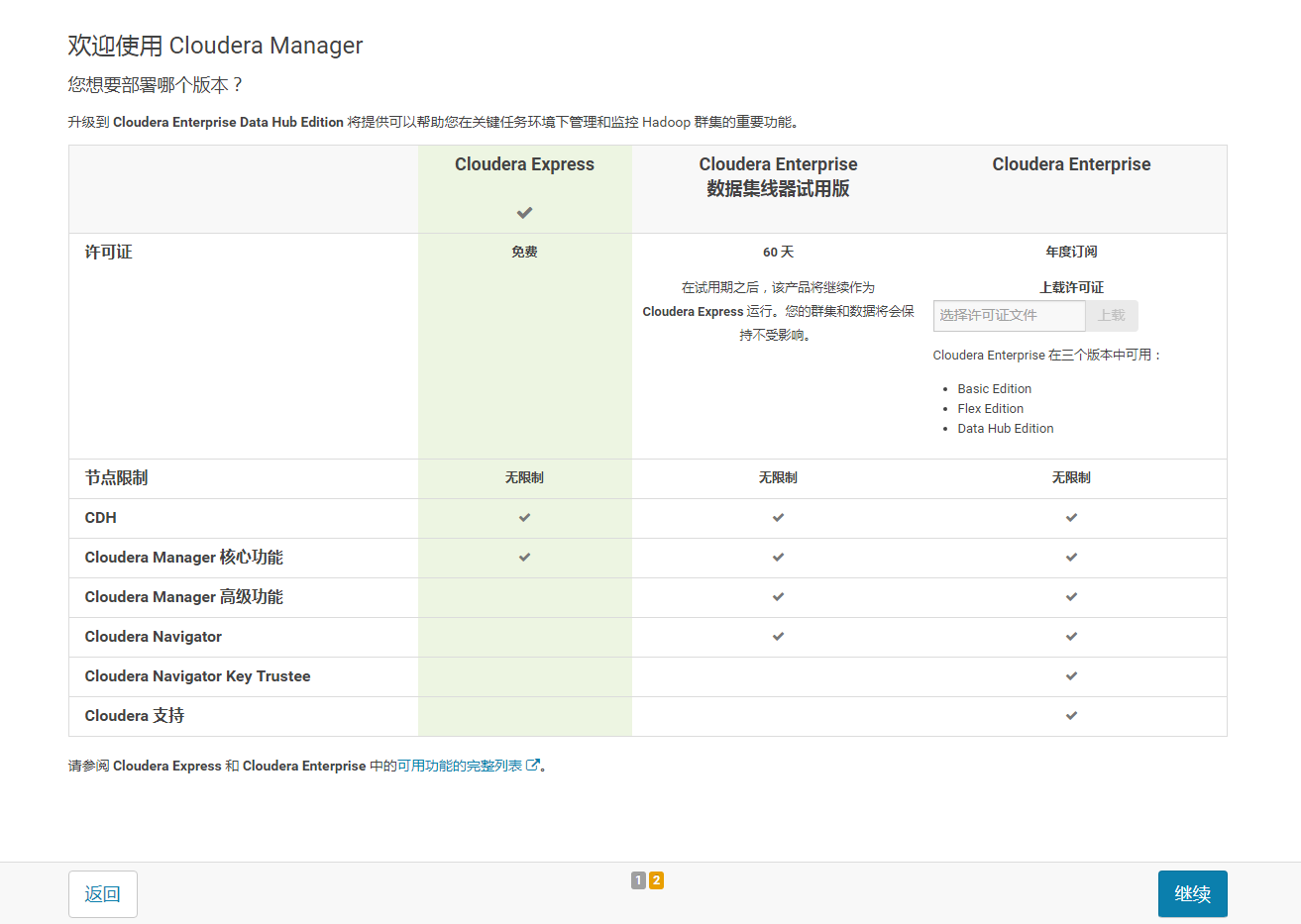
用户名：admin 密码：admin



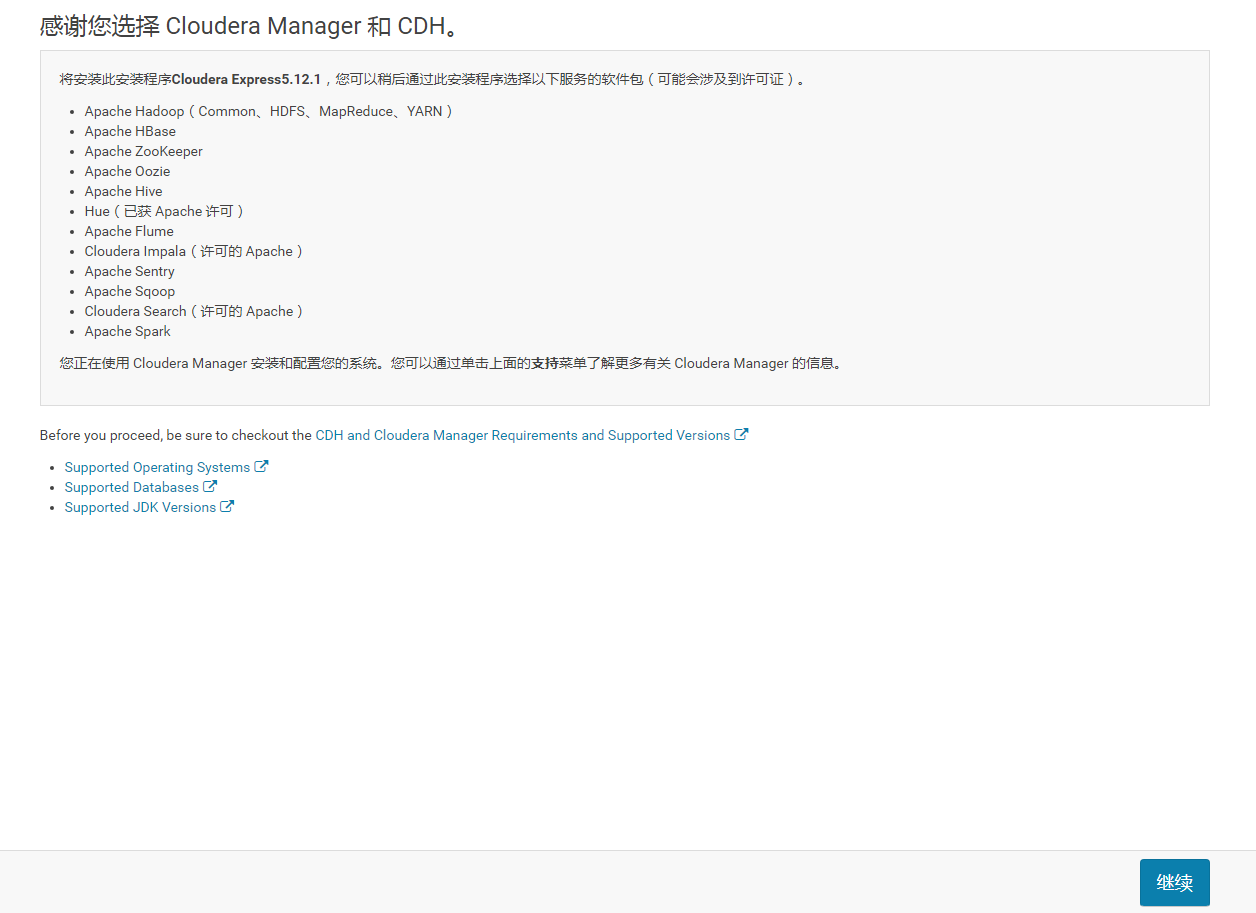
勾选同意，点击继续



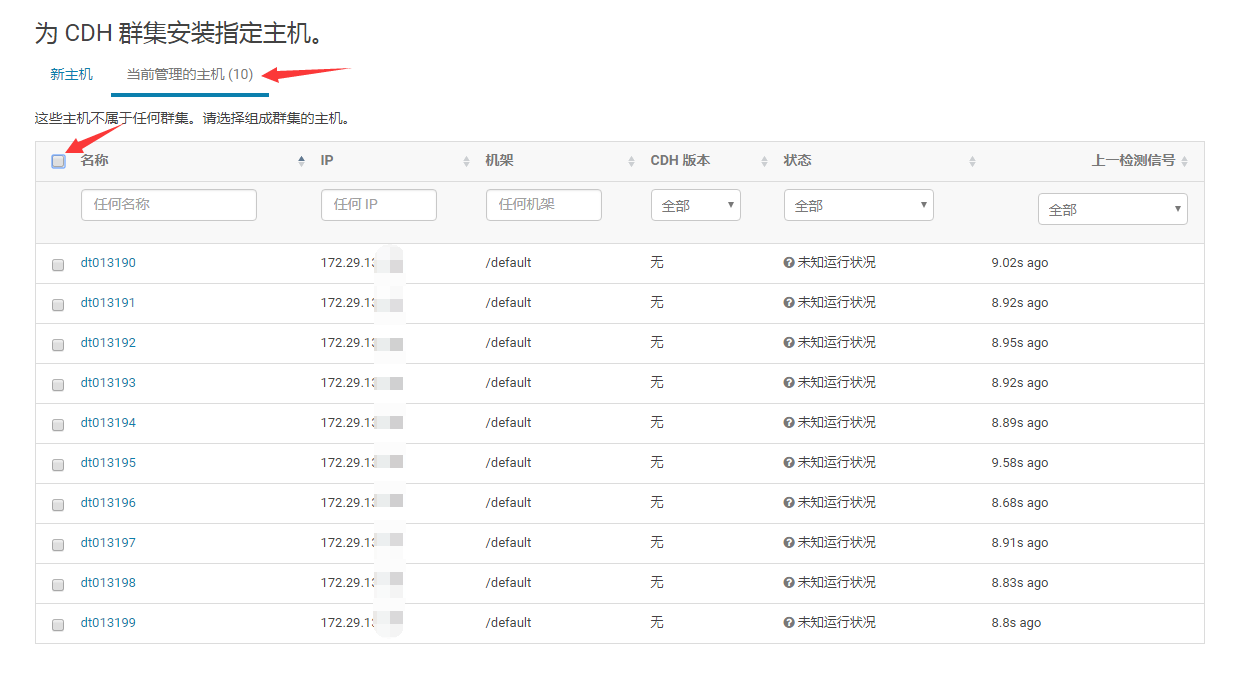
选择免费版本，继续



继续



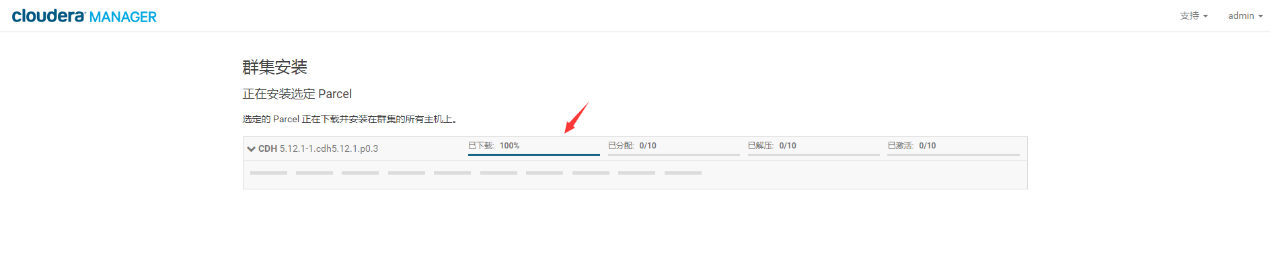
各个Agent节点正常启动后，可以在当前管理的主机列表中看到对应的节点。选择要安装的节点，这里我选择全部节点（之前已经安装过了，此图片从网上获取的，节点原本只有3个，为hadoop01、hadoop02、hadoop03），继续。

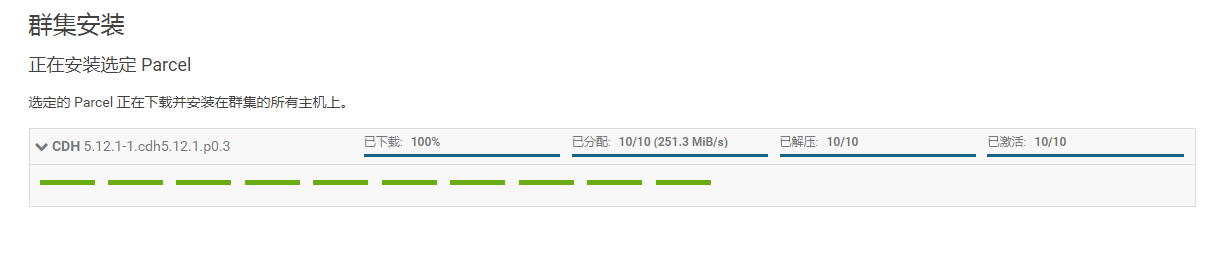


出现以下版本的包，说明本地Parcel包配置无误，直接点继续。

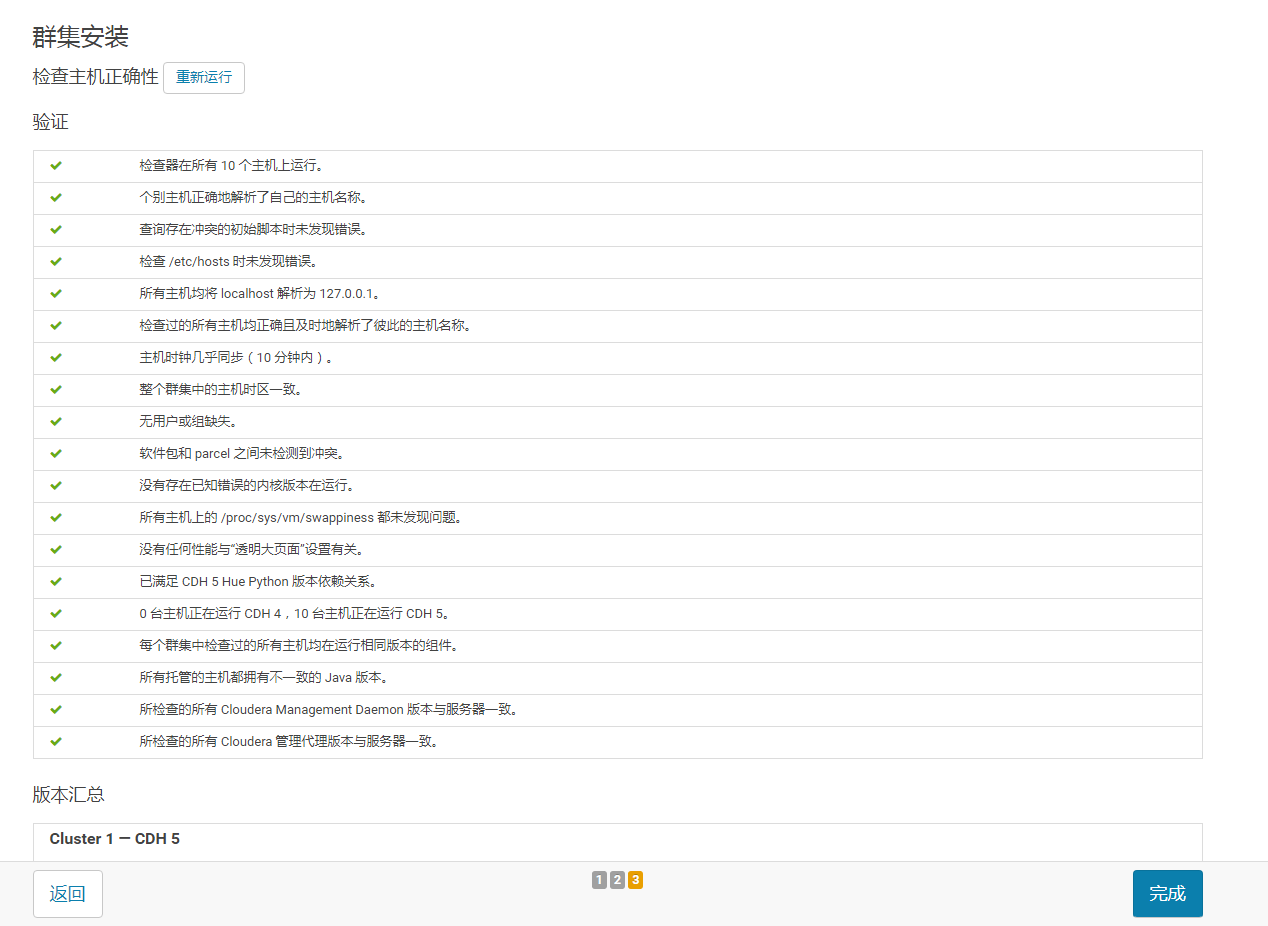


如果配置本地Parcel包无误，那么已下载是瞬间就完成的，因为不需要下载，然后耐心等待分配过程就行了，速度由内网网速决定。完成后点击继续。



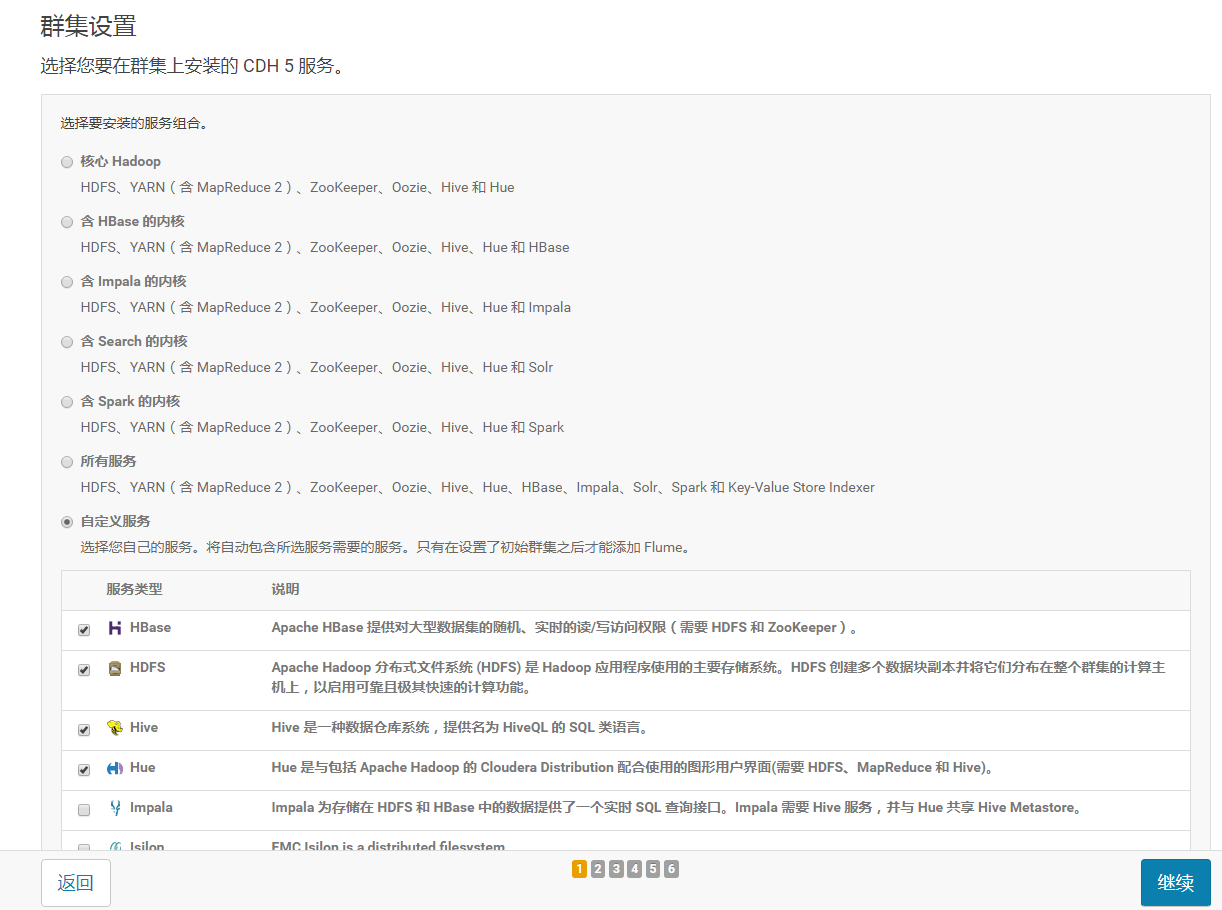


接下来是服务器检查，如果前面严格按照步骤一步步做下来这里应该是不会出现任何问题的，全绿通过。点击完成。



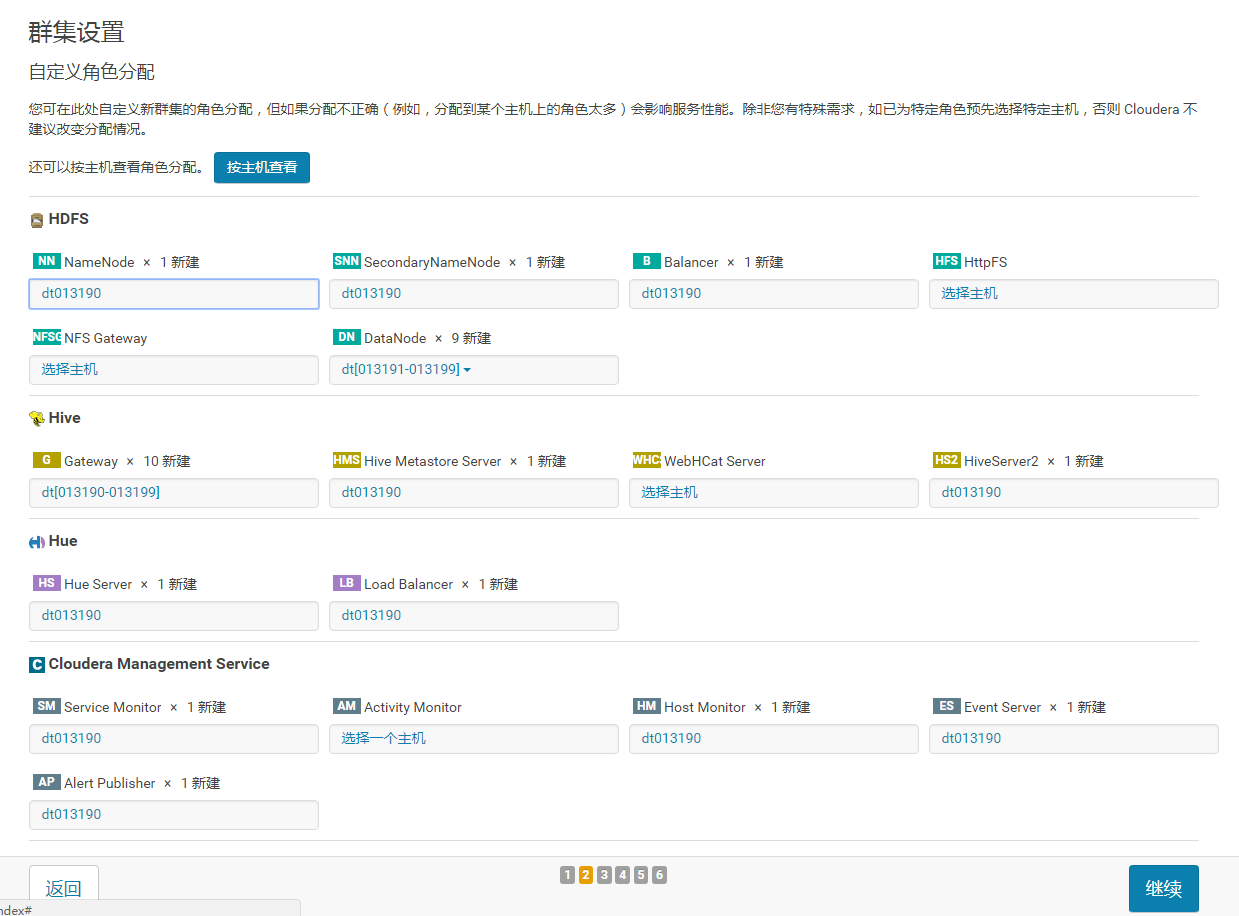
### 组件服务选择

接下来是选择安装服务，根据需要按需选择就好，点击继续。



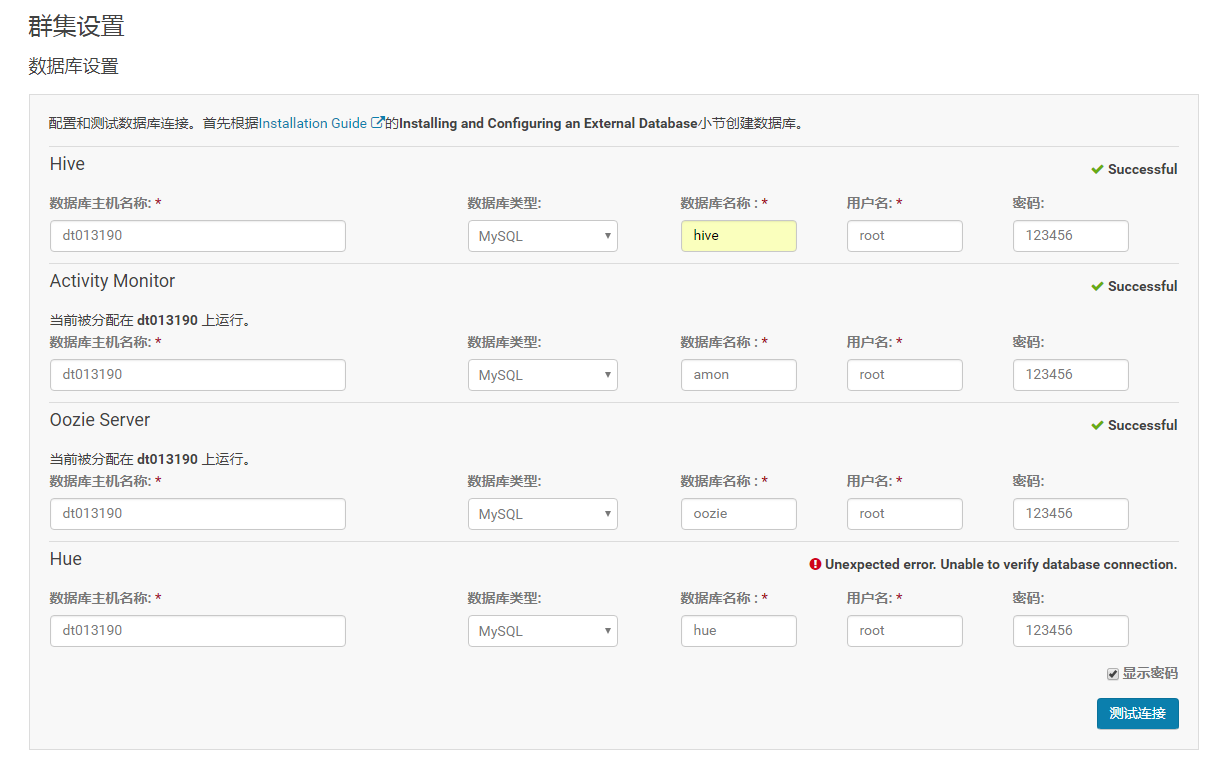
### 服务主机选择

服务配置，一般情况下保持默认就可以了，也可以自己手动根据节点情况进行分配，点击继续



### 测试元数据库

接下来是数据库的设置，部分系统会通过测试，部分可能错误

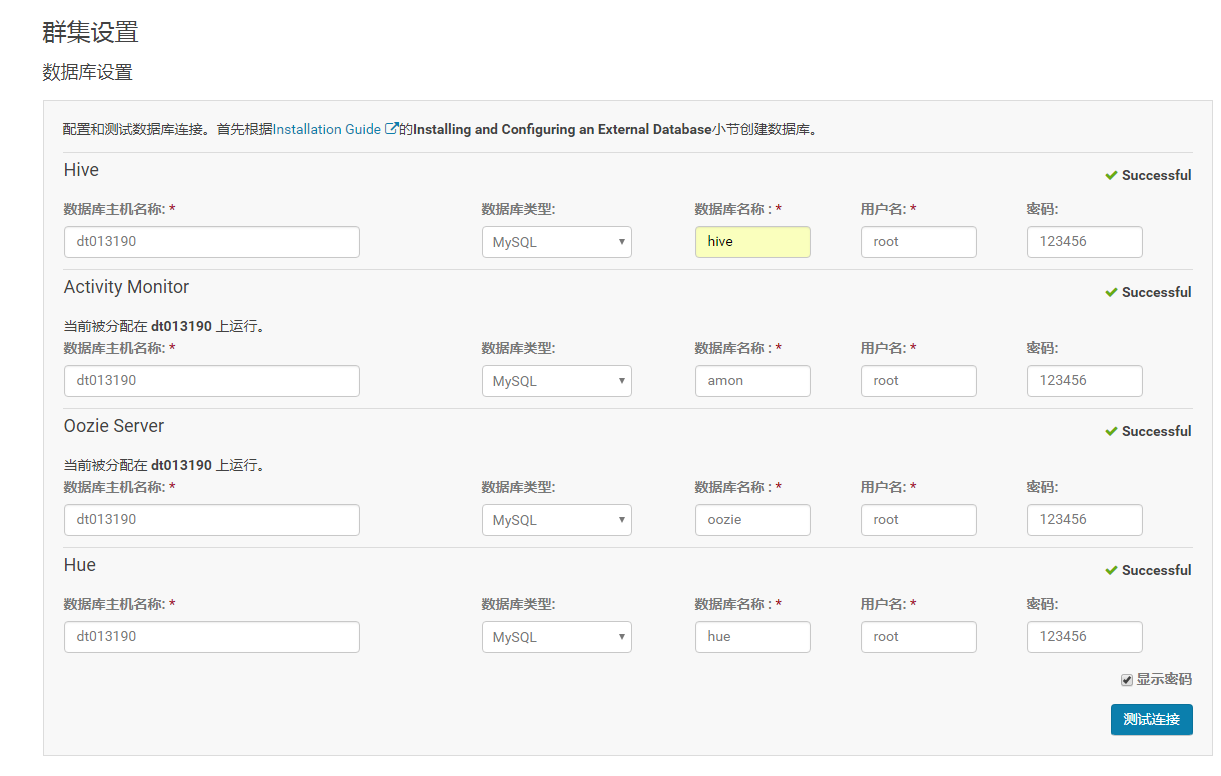


以上Hue测试不通过是因为，hue要存元数据库，CM的服务端是用Java编写的，而CM的客户端是Python编写的，这里测试连接时是用Python去连接数据库。

yum安装python-lxml 包：

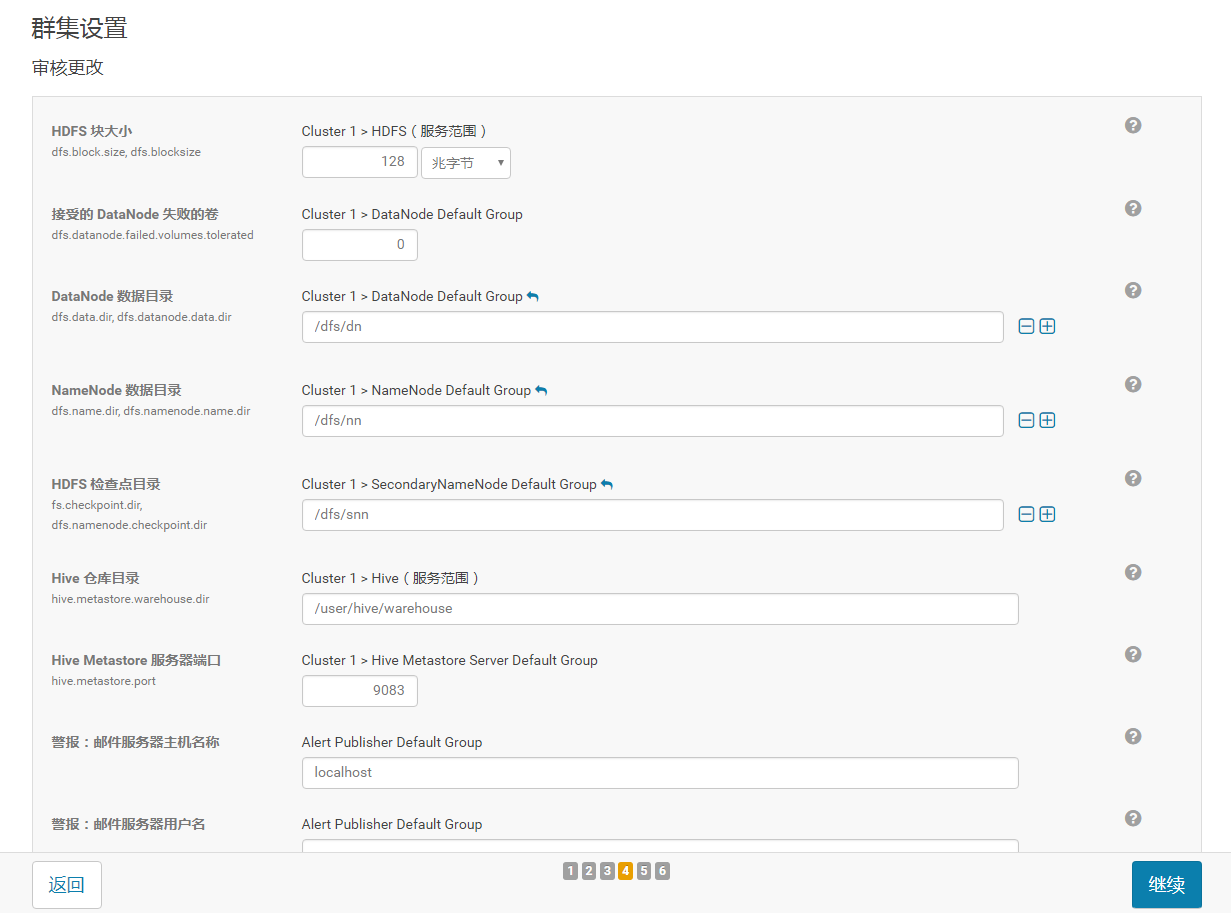
yum -y install python-lxml

再次点击测试连接，成功，点击继续

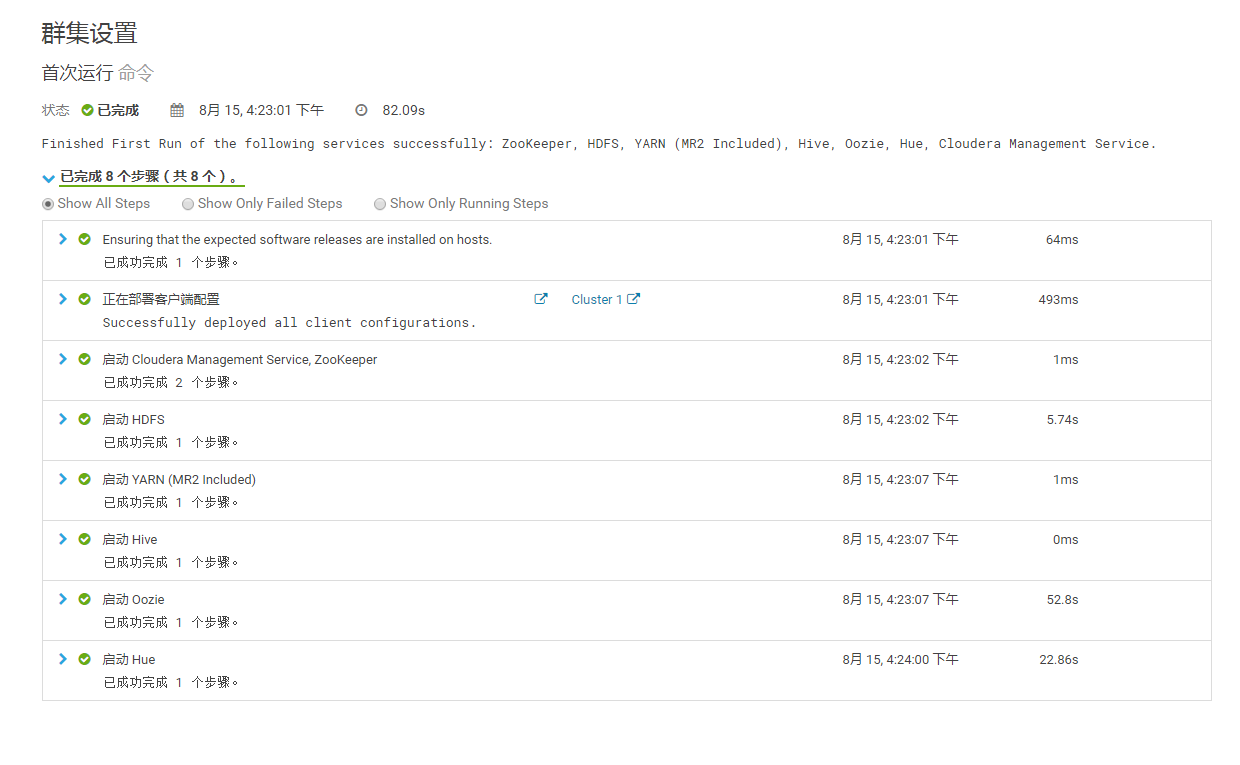


下面是集群设置的审查页面，保持默认就可以，因为后面还需要更多详细的改动，不建议这里完成，我们首要的任务是先将集群建立起来。所以点击继续。

### 集群设置



全部绿色通过，然后点击继续。



到此为止CDH集群就算部署完毕了。