第一步：linux安装阶段

## 1.1基础配置

本文档中涉及到的 hadoop102/hadoop103/hadoop104 分别对应我们集群中的

big-data-01、big-data-02、big-data-03

Cloudera Manager是一个拥有集群自动化安装、中心化管理、集群监控、报警功能的一个工具，使得安装集群从几天的时间缩短在几个小时内，运维人员从数十人降低到几人以内，极大的提高集群管理的效率。



需将以下端口开放

|  |  |
| --- | --- |
| 服务 | 端口 |
| Cloudera Manager Server（WebUI） | 7180 |
| HDFS NameNode（WebUI） | 9870 |
| Yarn ResourceManager（WebUI） | 8088 |
| JobHistory Server（WebUI） | 19888 |
| HBase Master（WebUI） | 16010 |
| HiveServer2 （WebUI） | 10002 |
| HUE Server（WebUI） | 8888/8889 |
| Oozie Server | 11000 |
| Sentry Server | 51000 |
| Spark Master/Worker/History Server | 18080/18081/18088 |
| Kerberos | 88 |
| MySQL | 3306 |
| 本地yum源HTTP端口 | 8900 |

1.1.1 修改hosts文件

[root@big-data-01:/root]# vim /etc/hosts

127.0.0.1 localhost localhost

::1 localhost localhost.localdomain localhost6 localhost6.localdomain6

192.168.32.186 big-data-01 big-data-01

192.168.32.187 big-data-02 big-data-02

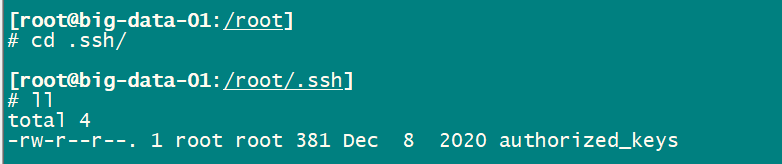
192.168.32.188 big-data-03 big-data-03

注意：这里每个人不一样，ip填写的是私有ip，做完后ping一下。ping big-data-03

### 1.1.2 SSH免密登录

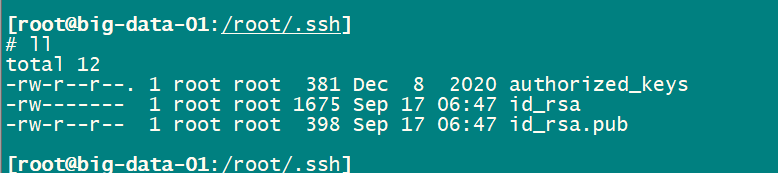
配置big-data-01对big-data-01、big-data-02、big-data-03三台服务器免密登录。CDH服务开启与关闭是通过server和agent来完成的，所以这里不需要配置SSH免密登录，但是为了我们分发文件方便，在这里我们也配置SSH。

1. 生成公钥和私钥：



[root@big-data-01:/root/.ssh]$ ssh-keygen -t rsa

然后敲（三个回车），就会生成两个文件id\_rsa（私钥）、id\_rsa.pub（公钥）



1. 将公钥拷贝到要免密登录的目标机器上

中间要输入root账户的密码

[root@big-data-01:/root/.ssh]$ ssh-copy-id big-data-01

[root@big-data-01:/root/.ssh]$ ssh-copy-id big-data-02

[root@big-data-01:/root/.ssh]$ ssh-copy-id big-data-03

3）重复1和2的操作，配置big-data-02对big-data-01、big-data-02、big-data-03三台服务器免密登录。

### 1.1.3 集群同步脚本

1）在/root目录下创建bin目录，并在bin目录下创建文件xsync，文件内容如下：

[root@big-data-01:/root]$ mkdir bin

[root@big-data-01:/root]$ cd bin/

[root@big-data-01:/root/bin]$ vi xsync

分发同步脚本xsync

2）修改脚本 xsync 具有执行权限

[root@big-data-01:/root/bin]$ chmod 777 xsync

### 1.1.4 集群整体操作脚本

1）在/root/bin目录下创建脚本xcall.sh

[root@big-data-01:/root/bin]$ vim xcall.sh

2）在脚本中编写如下内容

用xcall.sh

3）修改脚本执行权限

[root@big-data-01:/root/bin]$ chmod +x xcall.sh

4）测试

[root@big-data-01:/root/bin]# xcall.sh jps

1.1.5关闭防火墙

阿里云的防火墙默认是关闭的，但是他有web防火墙，我们需要将7180，3306,8888,8889,9870等端口开放。

### 1.1.6 关闭SELINUX

有如下两种关闭方法：

1）临时关闭（不建议使用）

[root@big-data-01:/root]# setenforce 0

但是这种方式只对当次启动有效，重启机器后会失效。

2）永久关闭（建议使用）

修改配置文件/etc/selinux/config

[root@big-data-01:/root]# vim /etc/selinux/config

将SELINUX=enforcing 改为SELINUX=disabled

SELINUX=disabled

3）同步/etc/selinux/config配置文件

[root@big-data-01:/root]# xsync /etc/selinux/config

4）重启big-data-01、big-data-02、big-data-03主机

[root@big-data-01:/root]# reboot

[root@big-data-02:/root]# reboot

[root@big-data-03:/root]# reboot

### 1.1.7 配置NTP时钟同步（暂时没有配置）

1）NTP服务器配置

[root@hadoop102 ~]# vi /etc/ntp.conf

①注释掉所有的restrict开头的配置

②修改#restrict 192.168.1.0 mask 255.255.255.0 nomodify notrap

为restrict 192.168.1.102 mask 255.255.255.0 nomodify notrap

③将所有server配置进行注释

④添加下面两行内容

server 127.127.1.0

fudge 127.127.1.0 stratum 10

2）启动NTP服务 service ntpd start

[root@hadoop102 ~]# service ntpd start

3）NTP客户端配置（在agent主机上进行配置hadoop103,hadoop104）

[root@hadoop103 ~]# vi /etc/ntp.conf

①注释所有restrict和server配置

②添加server 192.168.1.102

4）手动测试

[root@hadoop103~]# ntpdate 192.168.1.102

显示如下内容为成功：

17 Jun 15:34:38 ntpdate[9247]: step time server 192.168.1.102 offset 77556618.173854 sec

如果显示如下内容需要先关闭ntpd：

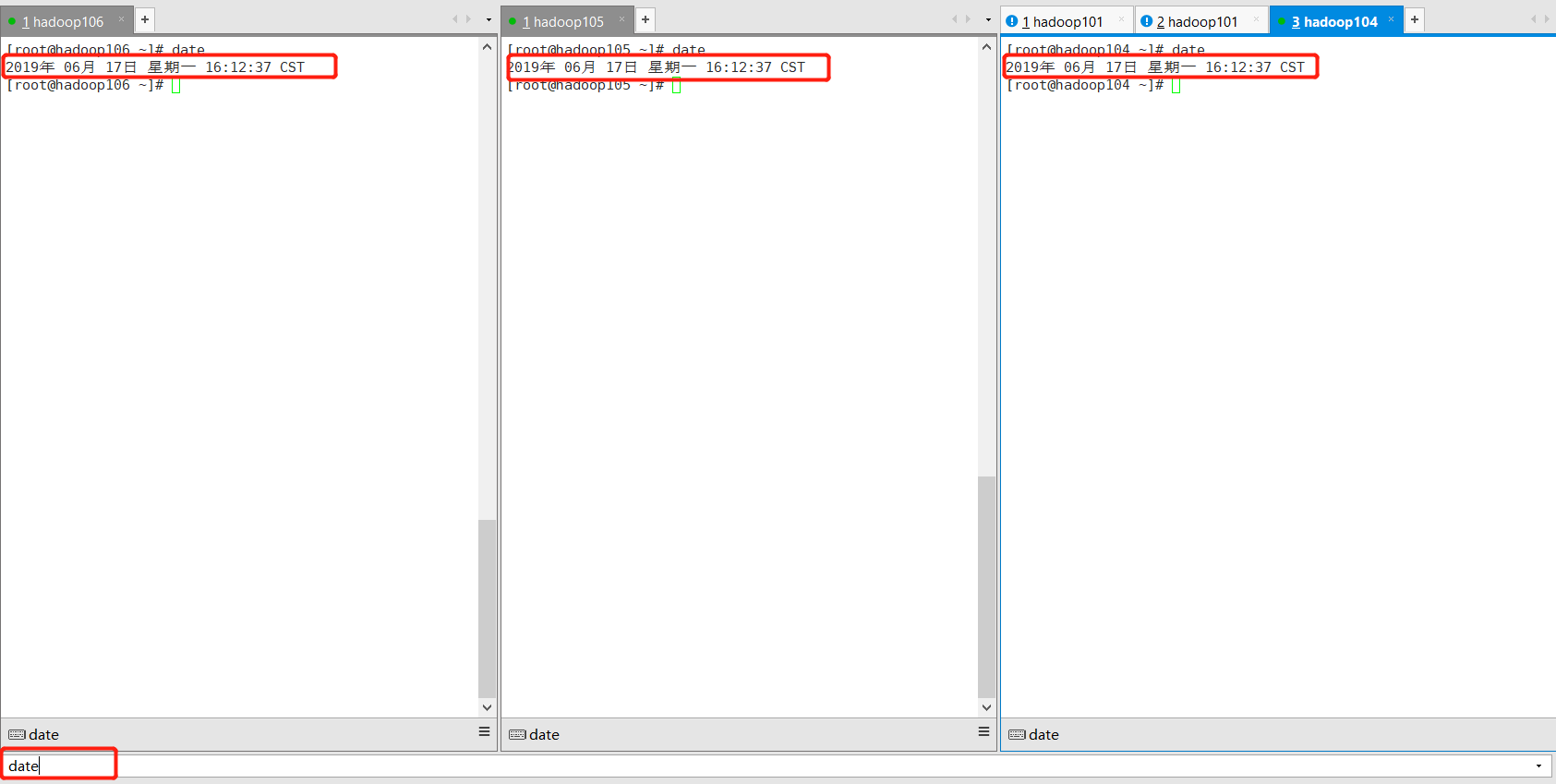
17 Jun 15:25:42 ntpdate[8885]: the NTP socket is in use, exiting

5）启动ntpd并设置为开机自启（每个节点hadoop102，hadoop103，hadoop104）

[root@hadoop103 ~]# chkconfig ntpd on

[root@hadoop103 ~]# service ntpd start

6）使用群发date命令查看结果



4.修改时区

1、查看系统当前的时区

[root@node01 etc]# timedatectl

Local time: Fri 2020-07-24 08:31:38 UTC

Universal time: Fri 2020-07-24 08:31:38 UTC

RTC time: Fri 2020-07-24 08:31:38

Time zone: UTC (UTC, +0000)

NTP enabled: yes

NTP synchronized: yes

RTC in local TZ: no

DST active: n/a

2、修改系统使用的时区

修改为中国时区

timedatectl set-timezone Asia/Shanghai

修改为世界时区

timedatectl set-timezone UTC

3、时间不准确时校准时间

yum install -y ntpdate

ntpdate us.pool.ntp.org

4.将硬件时钟调整为与本地时钟一致, 0 为设置为 UTC 时间(（立即生效）)

timedatectl set-local-rtc 1

5.查看硬件时间

hwclock --show

## 1.2 CM安装部署

### 1.2.1 安装JDK（三台）

1）在hadoop102创建**/usr/java**目录

[root@big-data-01:/opt]# mkdir /usr/java

2）用SecureCRT将jdk-8u144-linux-x64.tar.gz上传至big-data-01，并解压到/usr/java目录下。

[root@big-data-01:/opt]# tar -zxvf jdk-8u144-linux-x64.tar.gz -C /usr/java/

3）配置JAVA\_HOME环境变量

（1）打开/etc/profile文件

[root@big-data-01:/opt/software]$ vim /etc/profile

在profile文件末尾添加JDK路径

#JAVA\_HOME

export JAVA\_HOME=/usr/java/jdk1.8.0\_191-amd64

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

（2）让修改后的文件生效

[root@hadoop102 jdk1.8.0\_144]$ source /etc/profile

4）测试JDK是否安装成功

[root@hadoop102 jdk1.8.0\_144]# java -version

java version "1.8.0\_144"

5）将hadoop102中的JDK和环境变量分发到big-data-02、big-data-03两台主机

[root@big-data-01:/opt]# xsync /usr/java/

[root@big-data-01:/opt]# xsync /etc/profile

分别在big-data-02、big-data-03上source一下

[root@big-data-02:/root]$ source /etc/profile

[root@big-data-03:/root]# source /etc/profile

### 1.2.2 安装MySQL及其驱动（big-data-01上操作）

1）卸载mariadb

rpm -qa | grep mariadb | xargs rpm -e --nodeps

1. 下载msql5.7 yum源

Wget：就是指在外网下载东西。

wget https://dev.mysql.com/get/mysql57-community-release-el7-9.noarch.rpm

3）安装yum源

rpm -ivh mysql57-community-release-el7-9.noarch.rpm

1. 安装mysql

查看是否安装MySQL。（忽略查询的时候的大小写）

rpm -qa|grep -i mysql

如果安装了的话，就先卸载

rpm -e --nodeps mysql-libs-5.1.73-7.el6.x86\_64





yum -y install mysql-server

5）启动mysql

service mysqld start

报错的话就用

systemctl restart mysqld.service

6）查看root用户密码

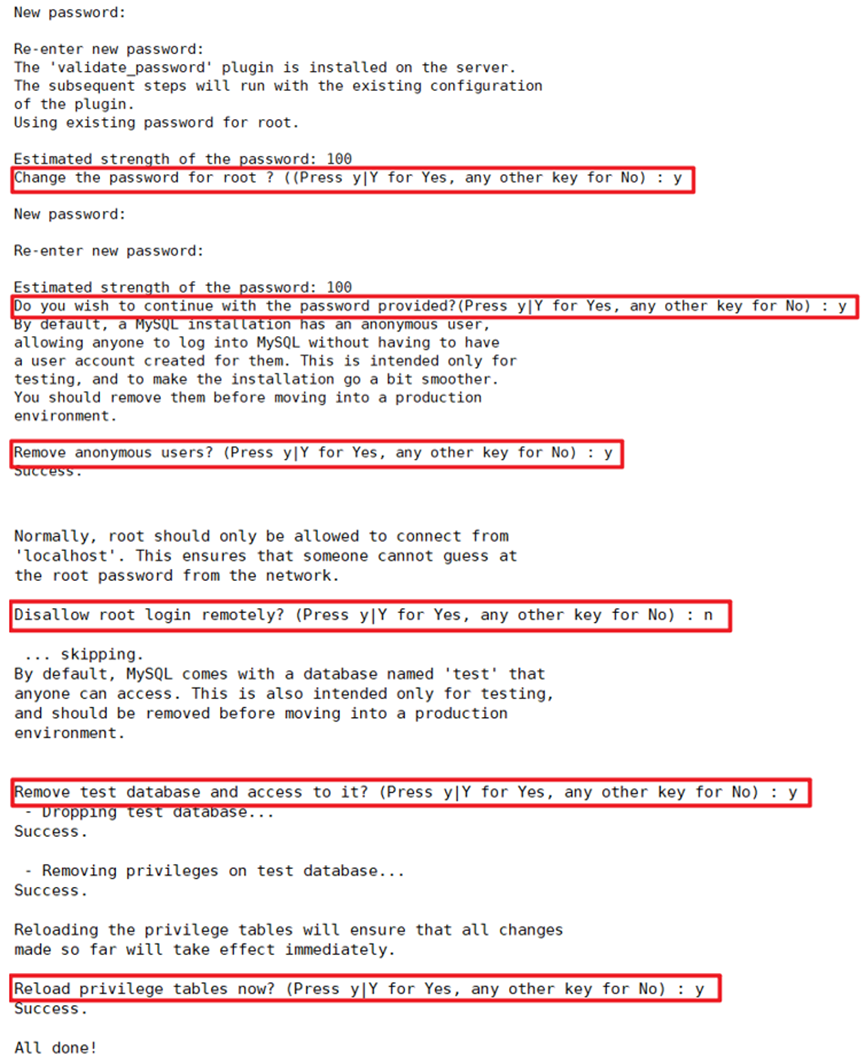
grep 'temporary password' /var/log/mysqld.log

CSqbu,B+Y2dj

7）执行mysql初始化脚本

mysql\_secure\_installation

8）输入新密码（至少12个字符，至少包含一个大写字母有，一个小写字母，一个数字，一个特殊字符）Yzl12345root!



登录MySQL，不要加空格！！！！！！

mysql -rroot -pYzl12345root!

9）配置root用户远程访问权限

mysql> grant all privileges on \*.\* to 'root' @'%' identified by 'Yzl12345root!';

mysql> flush privileges;

10）为CM安装mysql驱动

（1）将mysql-connector-java-5.1.27-bin.jar拷贝到**/usr/share/java**路径下，并**重命名**

[root@big-data-03:/usr/share/java] # mv mysql-connector-java-5.1.27-bin.jar mysql-connector-java.jar

（2）分发驱动

xsync /usr/share/java

### 1.2.3 CM安装

Cloudera Manager默认采用yum安装，对于能够联网的用户，可直接在线安装，十分快捷方便。对于网络不畅的用户，则可搭建本地yum源，进行安装。

1）集群规划

|  |  |  |  |
| --- | --- | --- | --- |
| 节点 | big-data-01 | big-data-02 | big-data-03 |
| 服务 | cloudera-scm-server  cloudera-scm-agent | cloudera-scm-agent | cloudera-scm-agent |

2）下载在线yum源配置文件（在线yum安装）

cd /etc/yum.repos.d

wget https://archive.cloudera.com/cm6/6.2.1/redhat7/yum/cloudera-manager.repo

3）分发在线yum源配置文件（在线yum安装）

xsync /etc/yum.repos.d/cloudera-manager.repo

4）搭建本地yum源（离线yum安装）（**若网络环境畅通，此步可直接跳过**）

（1）将压缩包cloudera-repos.tar.gz拷贝到集群中的一台节点，解压到/var/www/html路径下

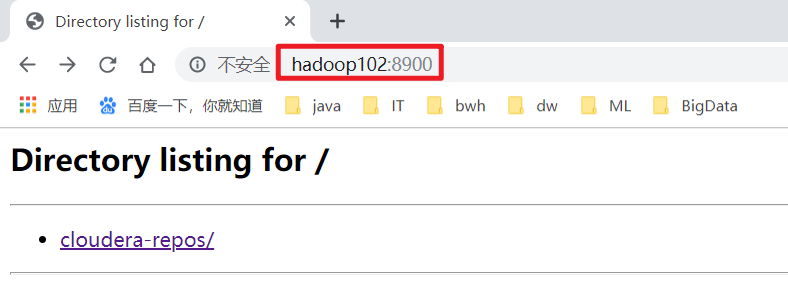
[root@big-data-01:/root]# tar -zxvf cloudera-repos.tar.gz -C /var/www/html

（2）进入/var/www/html/路径，并开启http服务

[root@big-data-01:/root]# cd /var/www/html/

[root@big-data-01:/var/www/html]# python -m SimpleHTTPServer 8900

（3）浏览器访问该节点的8900端口，查看http服务是否正常开启



（4）编辑本地yum源配置文件

vim /etc/yum.repos.d/cloudera-manager.repo

文件内容如下

[cloudera-manager]

name=cloudera-manager

baseurl=http://big-data-01:8900/cloudera-repos/cm6/6.2.1/redhat7/yum/

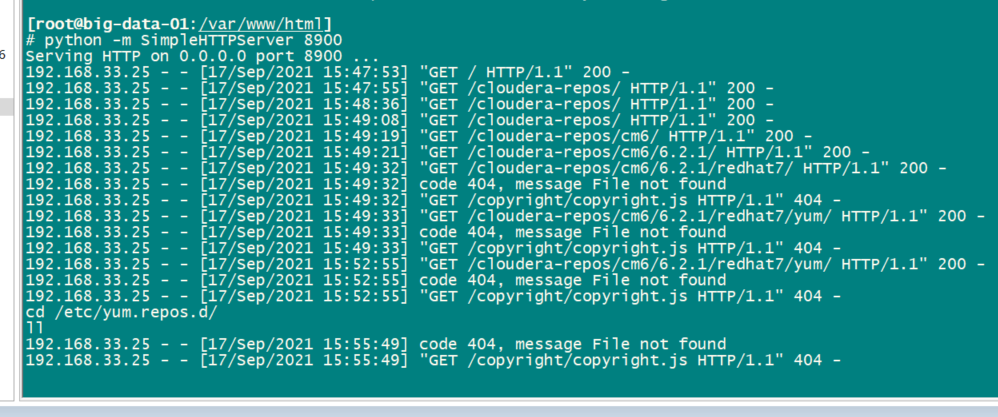
enabled=1

gpgcheck=0

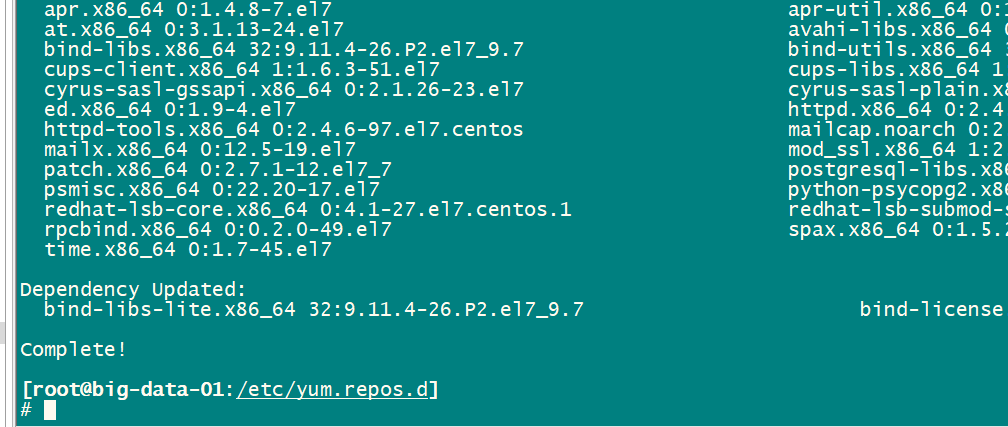
（5）分发该配置文件

xsync /etc/yum.repos.d/cloudera-manager.repo

1. 安装CM server及agent



这个窗口不能关 重开一个窗口执行以下操作。

这才是成功！

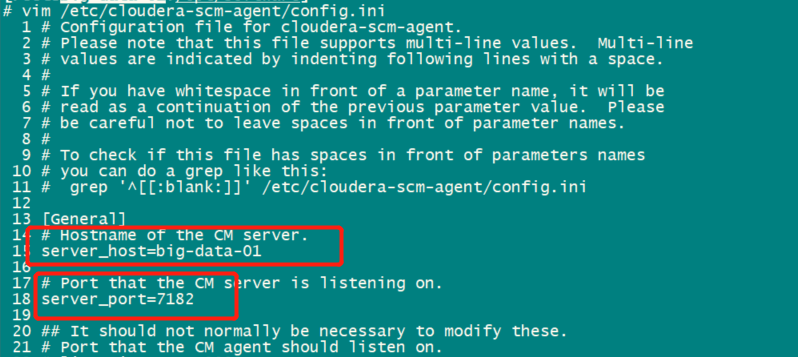
[root@big-data-01:/root]# yum -y install cloudera-manager-daemons cloudera-manager-agent cloudera-manager-server

[root@big-data-02:/root]# yum -y install cloudera-manager-agent cloudera-manager-daemons

[root@big-data-03:/root]# yum -y install cloudera-manager-agent cloudera-manager-daemons

### 1.2.4 修改CM配置文件（三台，不要用xsync同步）

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



3台机器都配置成big-data-01！！！！！一样！！！不一样就报错了！！！！！很麻烦！@！！！！！

### 1.2.5 在MySQL中建库

1）创建各组件需要的数据库(CDH的元数据库)

mysql> CREATE DATABASE scm DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

mysql> CREATE DATABASE amon DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

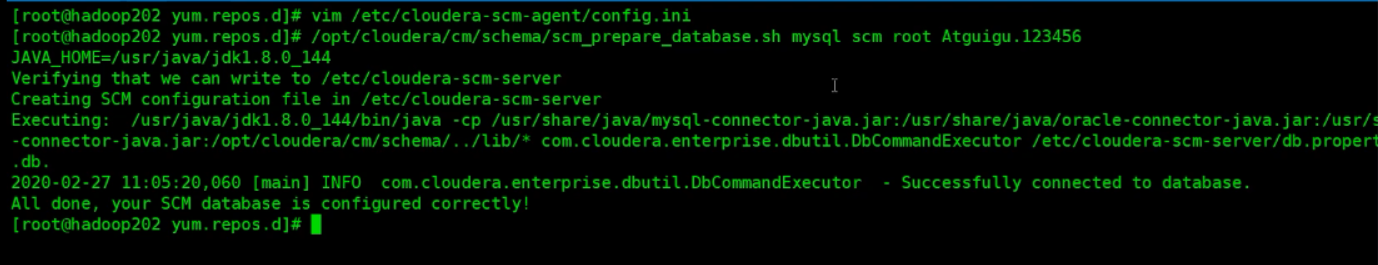
mysql> CREATE DATABASE hue DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

mysql> CREATE DATABASE hive DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

mysql> CREATE DATABASE sentry DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

mysql> CREATE DATABASE oozie DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8\_general\_ci;

2）为CM配置数据库（自带脚本）



/opt/cloudera/cm/schema/scm\_prepare\_database.sh mysql scm root Yzl12345root!

### 1.2.6 启动CM服务

1）启动服务节点：hadoop102

[root@big-data-01:/root]# systemctl start cloudera-scm-server

2）启动工作节点：hadoop102、hadoop103、hadoop104

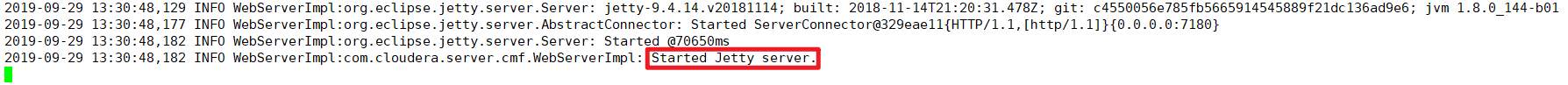
[root@big-data-01:/root]# systemctl start cloudera-scm-agent

[root@big-data-02:/root]# systemctl start cloudera-scm-agent

[root@big-data-03:/root]# systemctl start cloudera-scm-agent

3）查看Server启动日志

[root@big-data-01:/opt/cloudera/cm]# tail -100f /var/log/cloudera-scm-server/cloudera-scm-server.log



出现Started Jetty server字样及表明启动成功。

4）访问[http:// big-data-01:7180](http://hadoop102:7180)（初始用户名、密码均为admin）

CM 就是 CDH 的管家



### 1.2.7 关闭CM服务

1）关闭工作节点：hadoop102、hadoop103、hadoop104

[root@big-data-01:/root]# systemctl stop cloudera-scm-agent

[root@big-data-02:/root]# systemctl stop cloudera-scm-agent

[root@big-data-03:/root]# systemctl stop cloudera-scm-agent

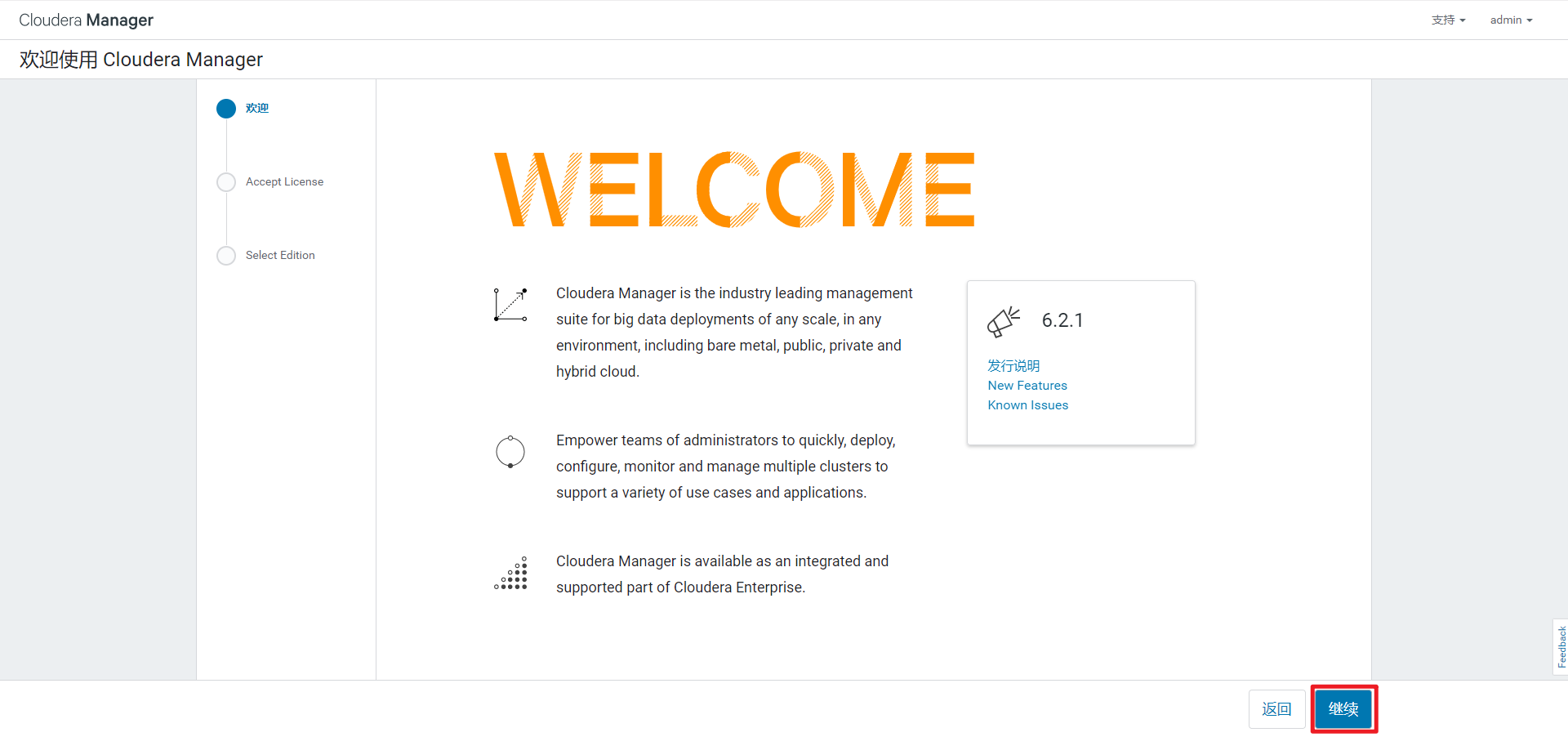
2）关闭服务节点： big-data-01

[root@big-data-01:/root]# systemctl stop cloudera-scm-server

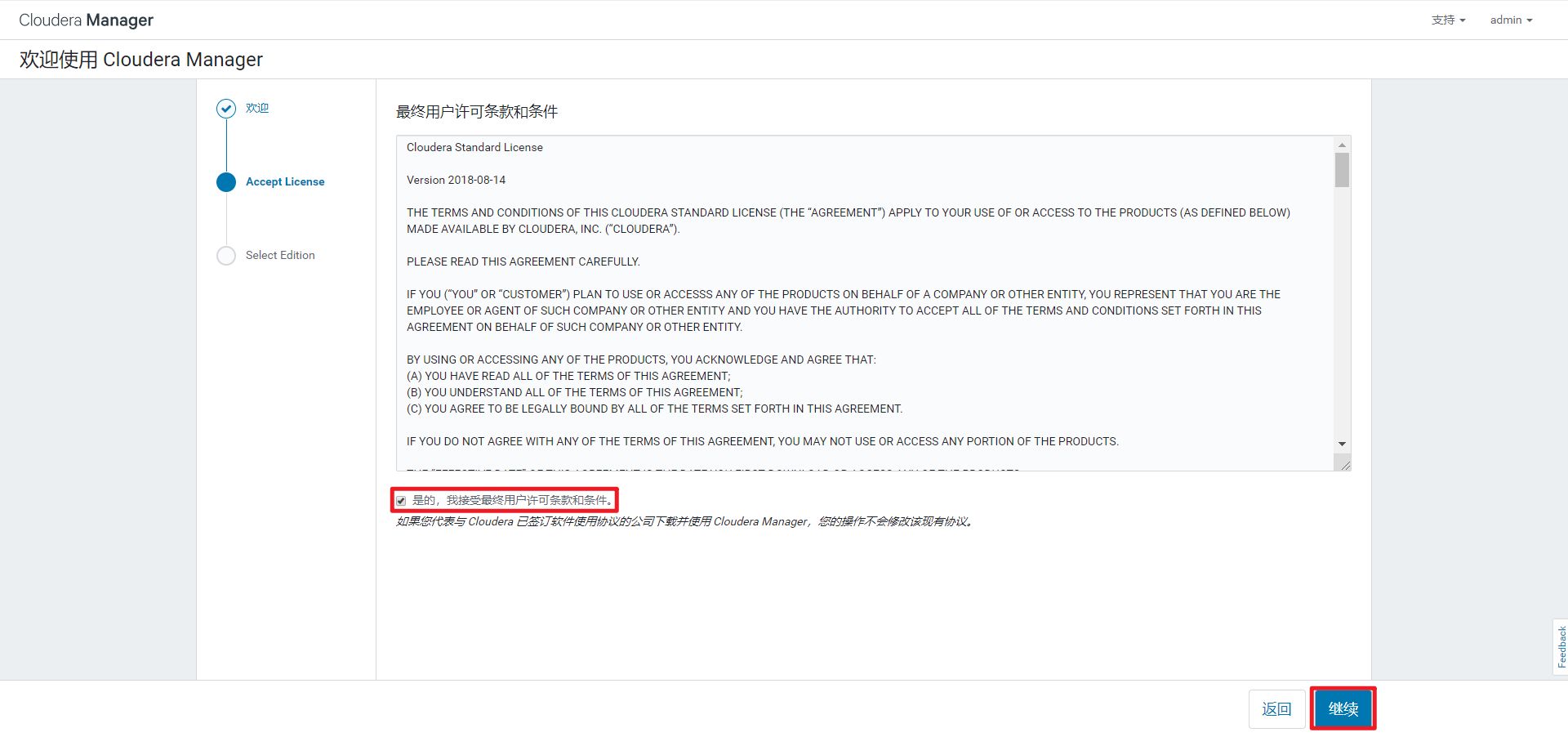
第2步 CDH页面配置

## 2.1 选择商业版本

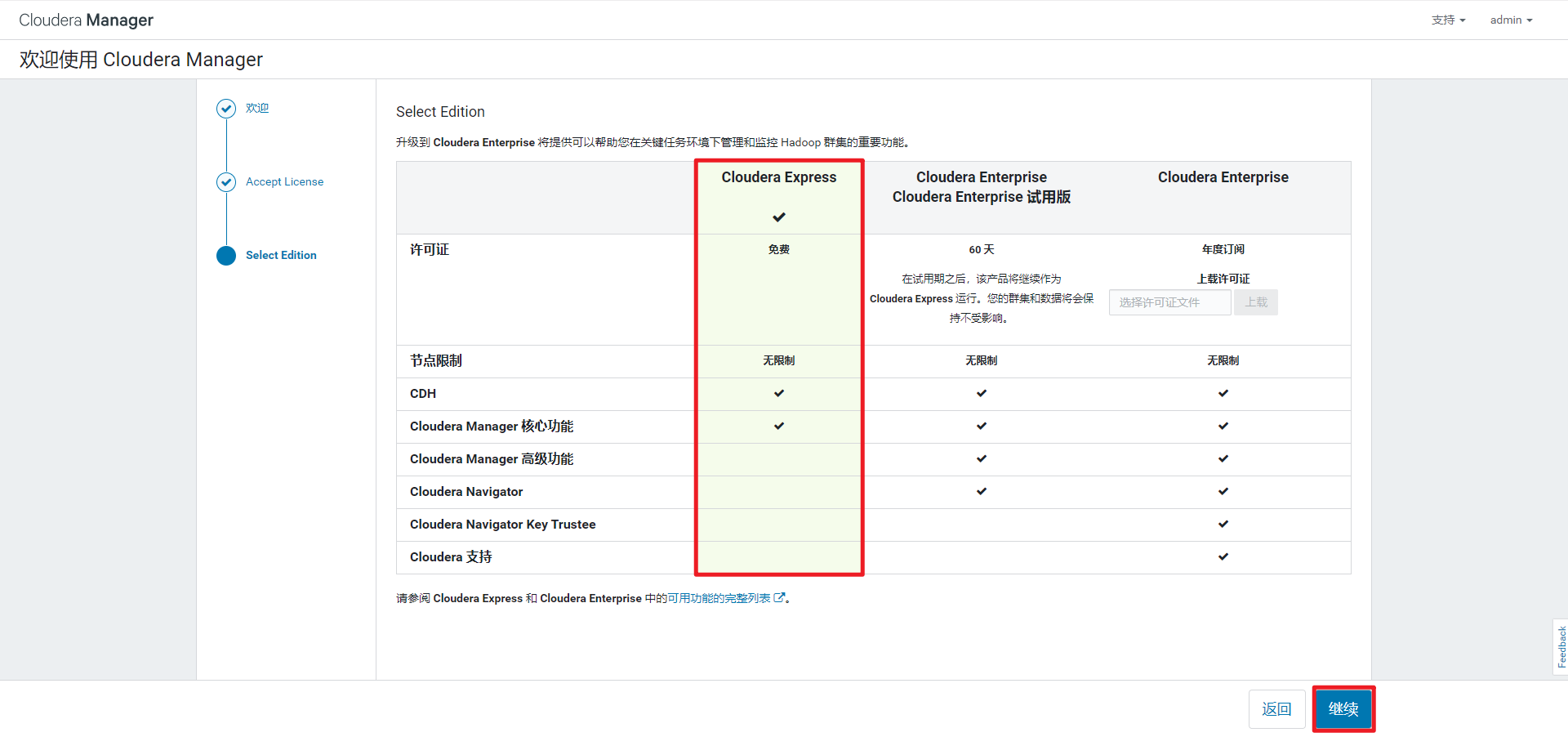
1）欢迎页面



2）用户协议

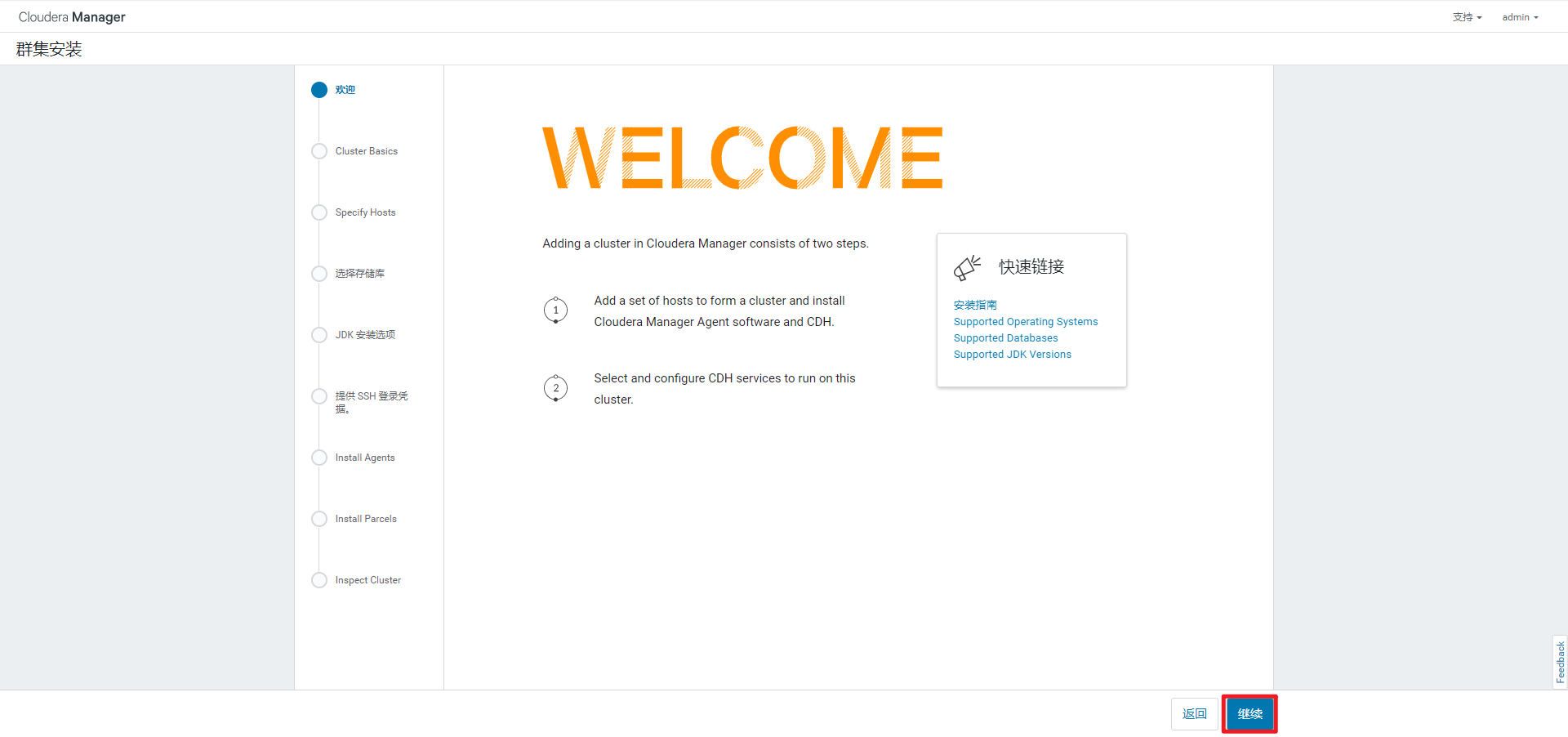


3）选择免费版

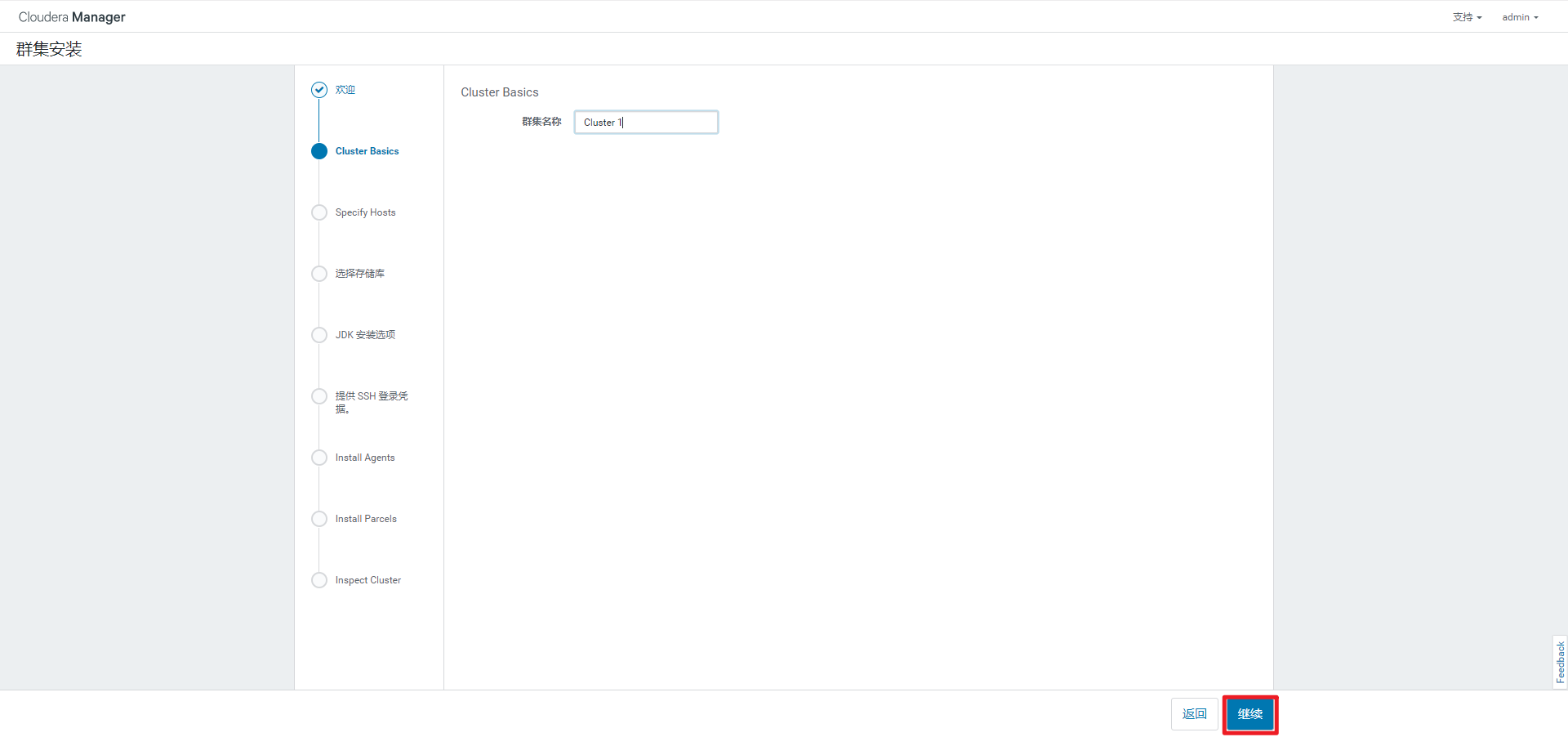


## 2.2 部署CDH集群

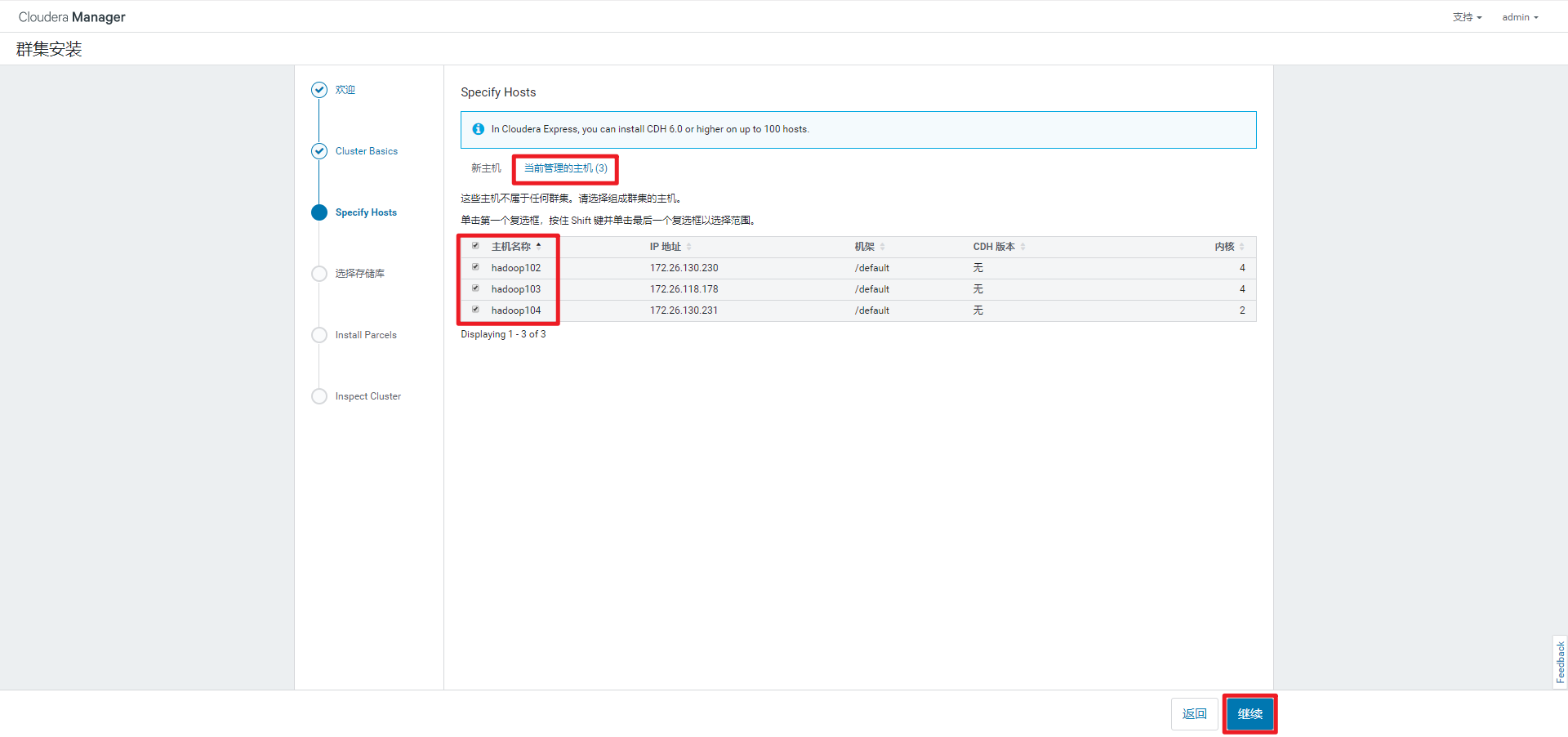
1）欢迎页面



2）集群命名：Cluster 1

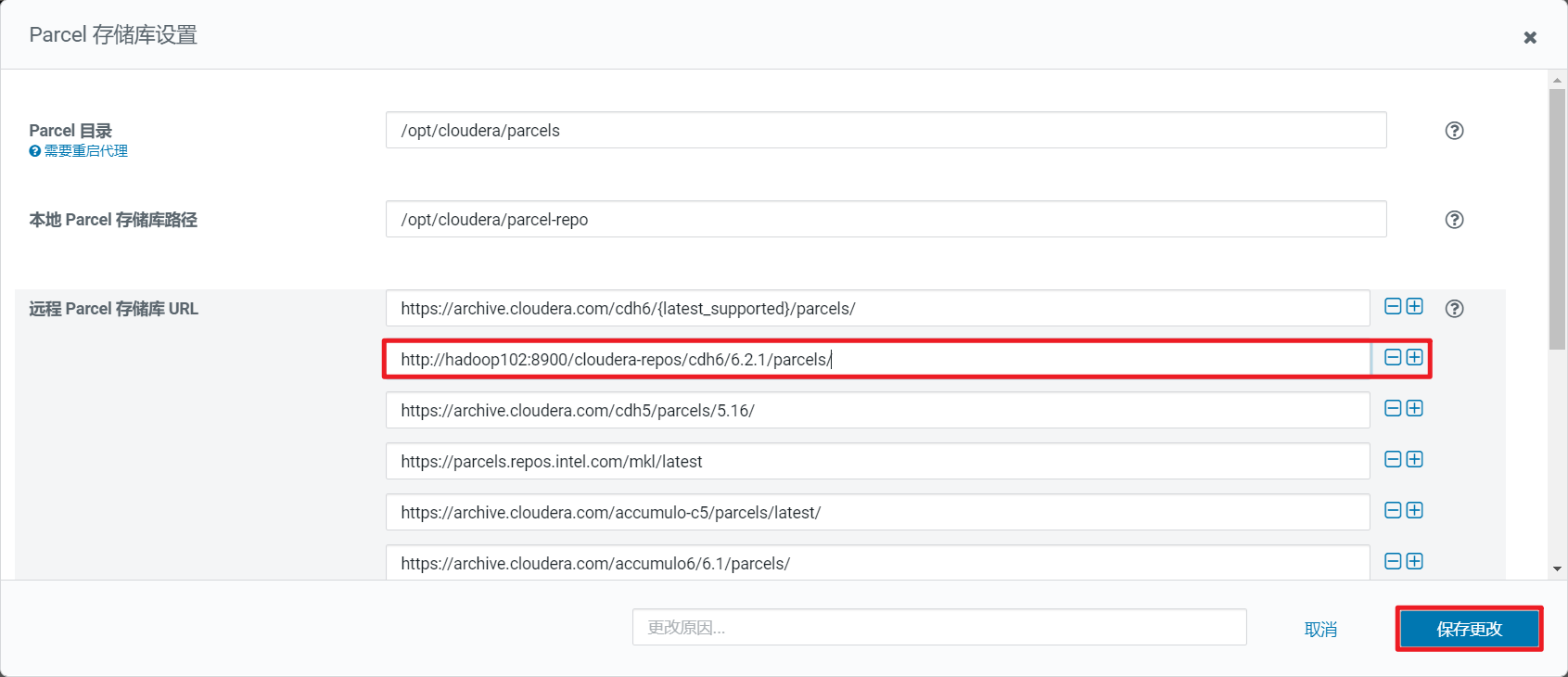


3）选定集群物理节点



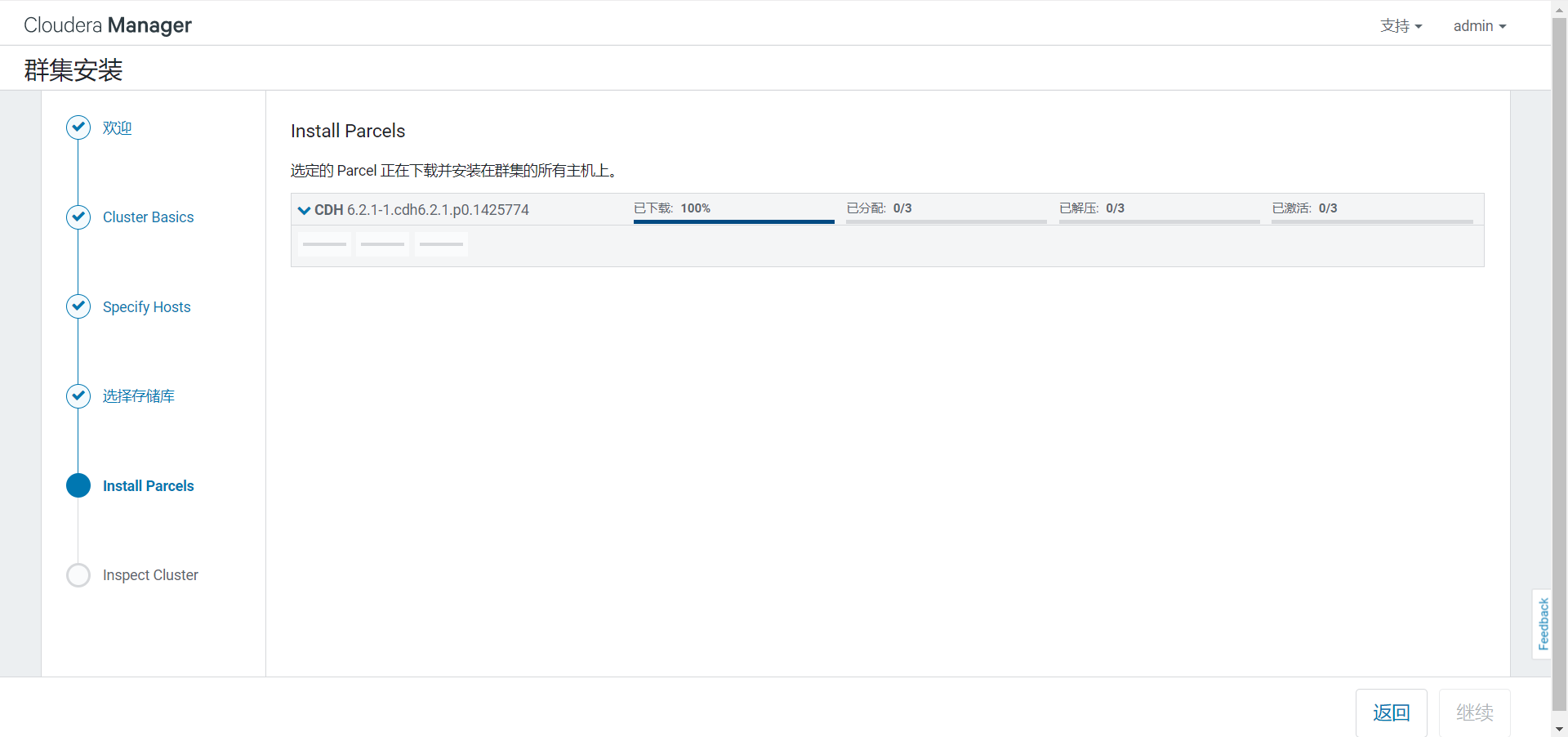
4）添加本地parcel库

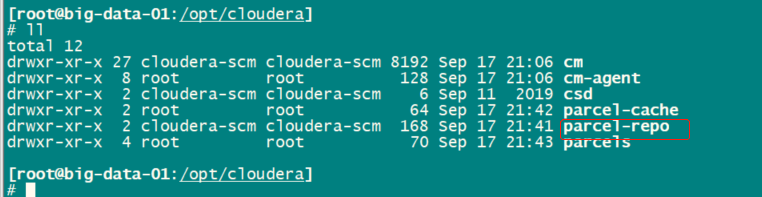




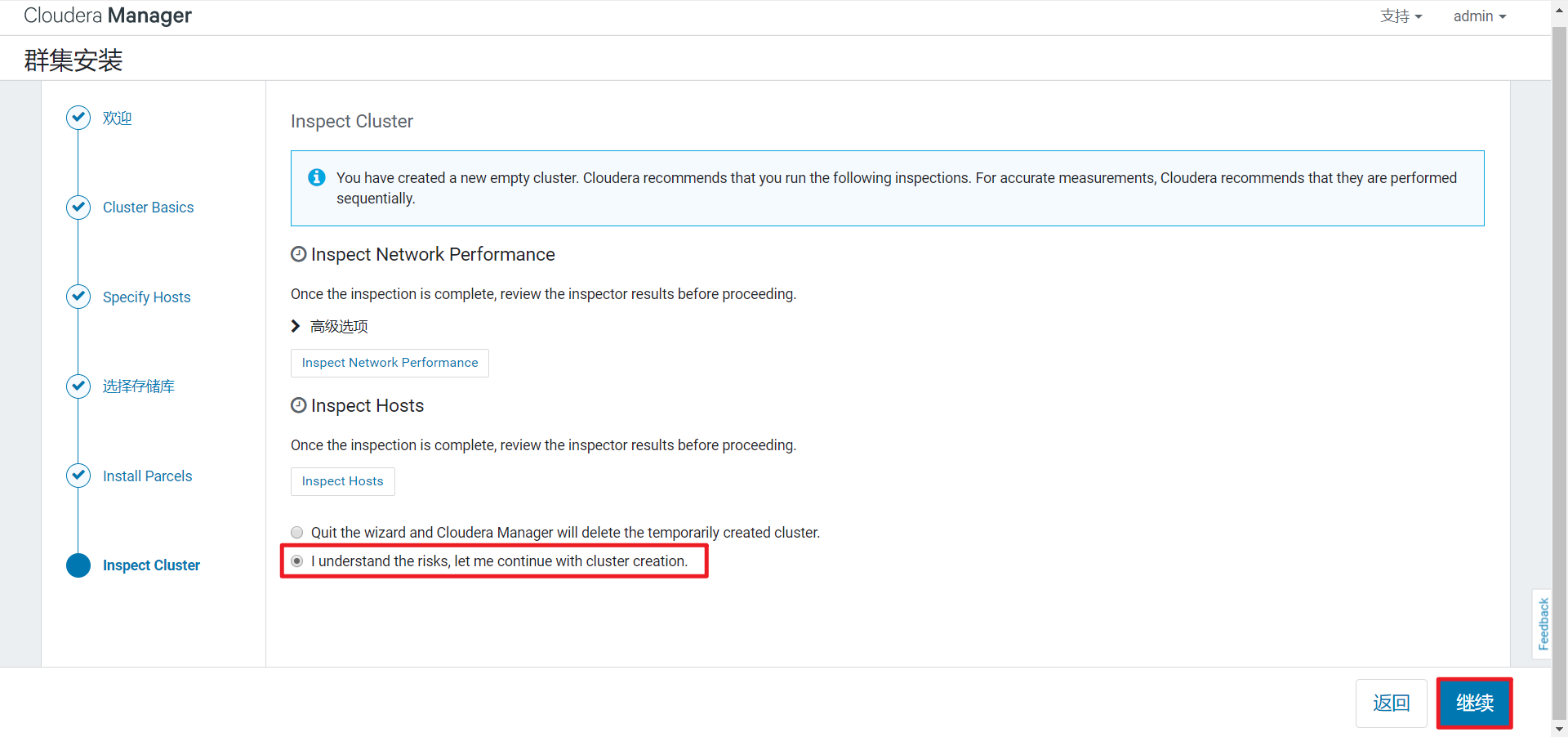


5）等待parcel的下载、分配、解压和激活

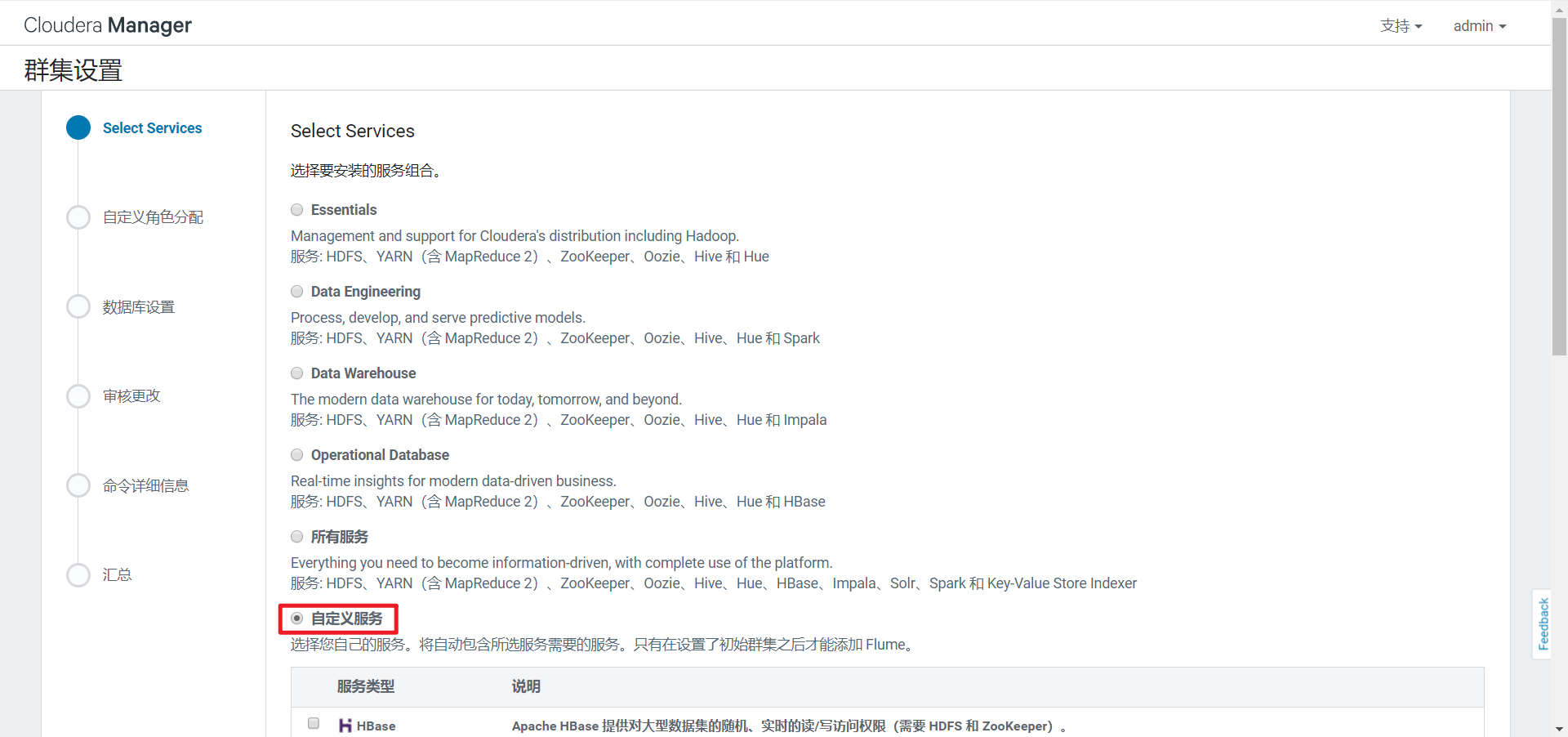


CDH 的本地仓库，parcels就是我们各种服务的

6）检查集群网络环境



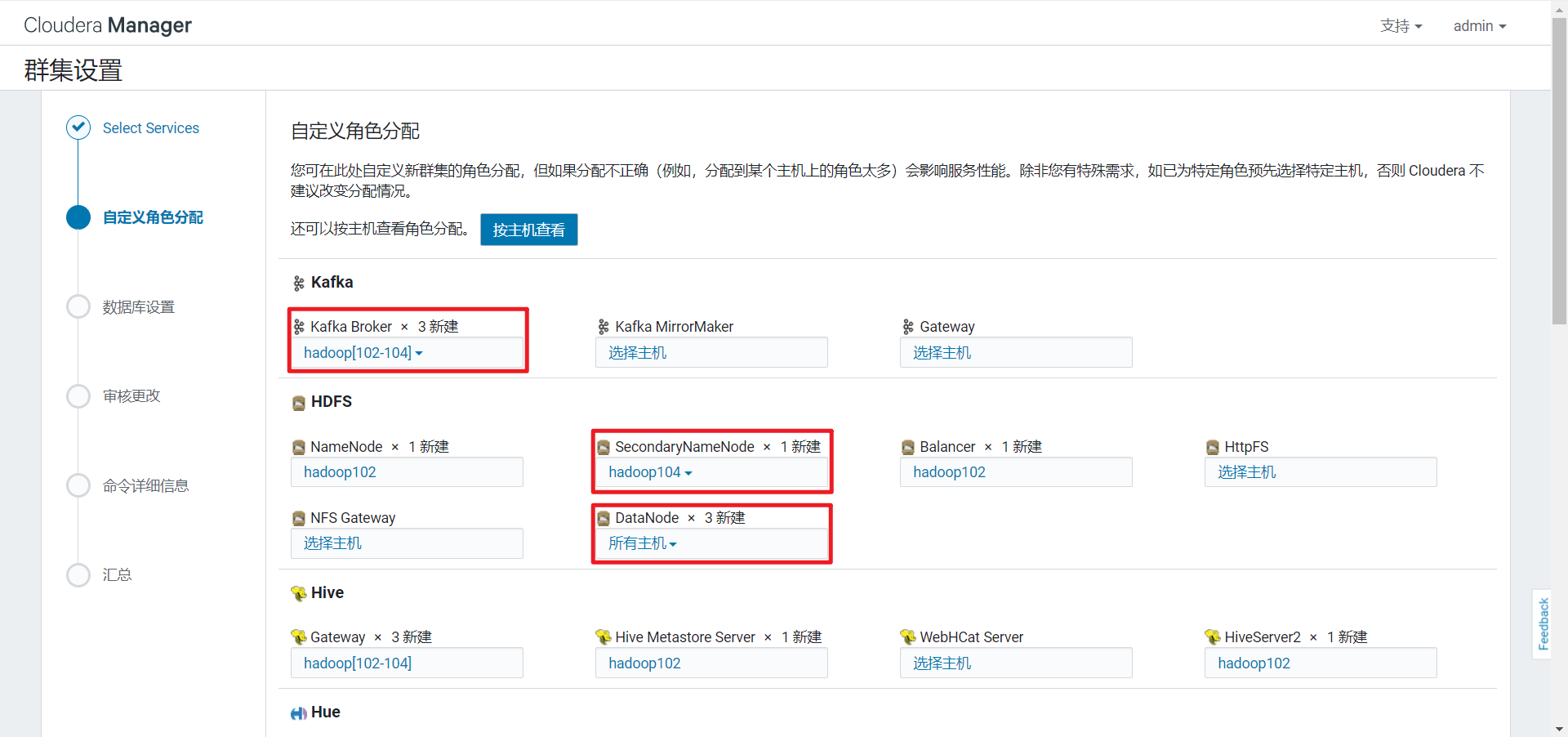
7）选择要安装的CDH组件，选择自定义安装

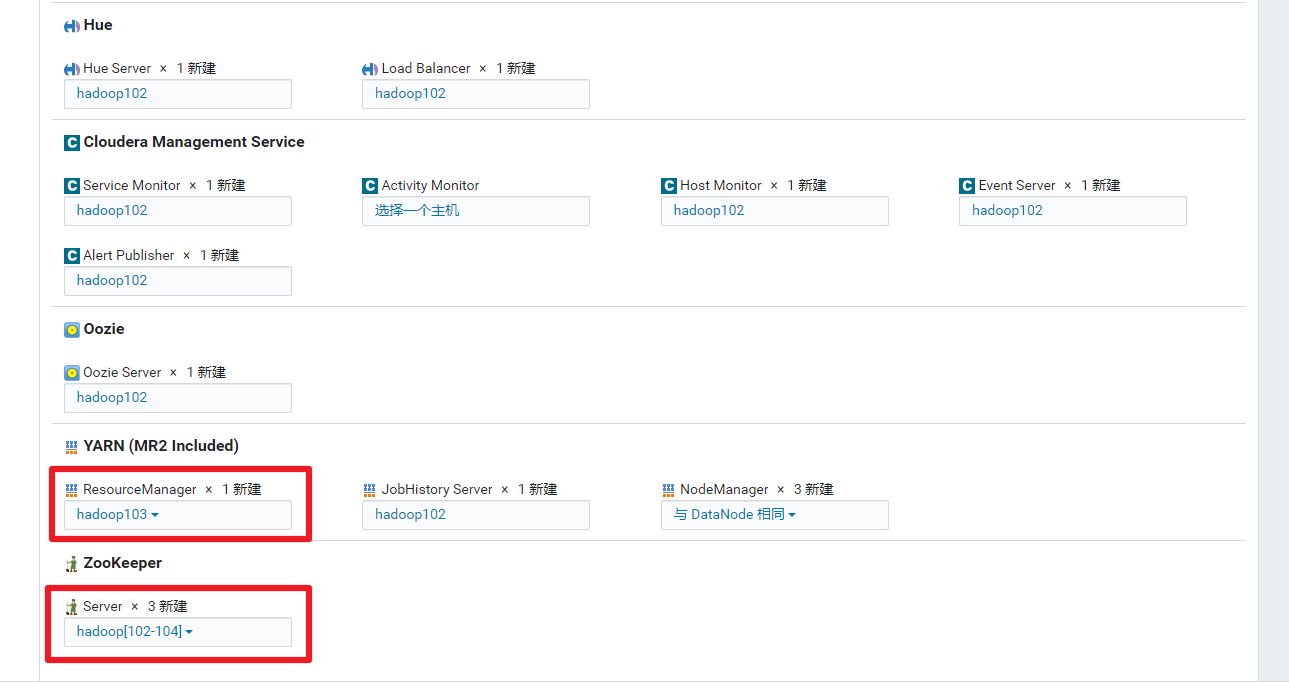


8）选择需要安装的组件，如下

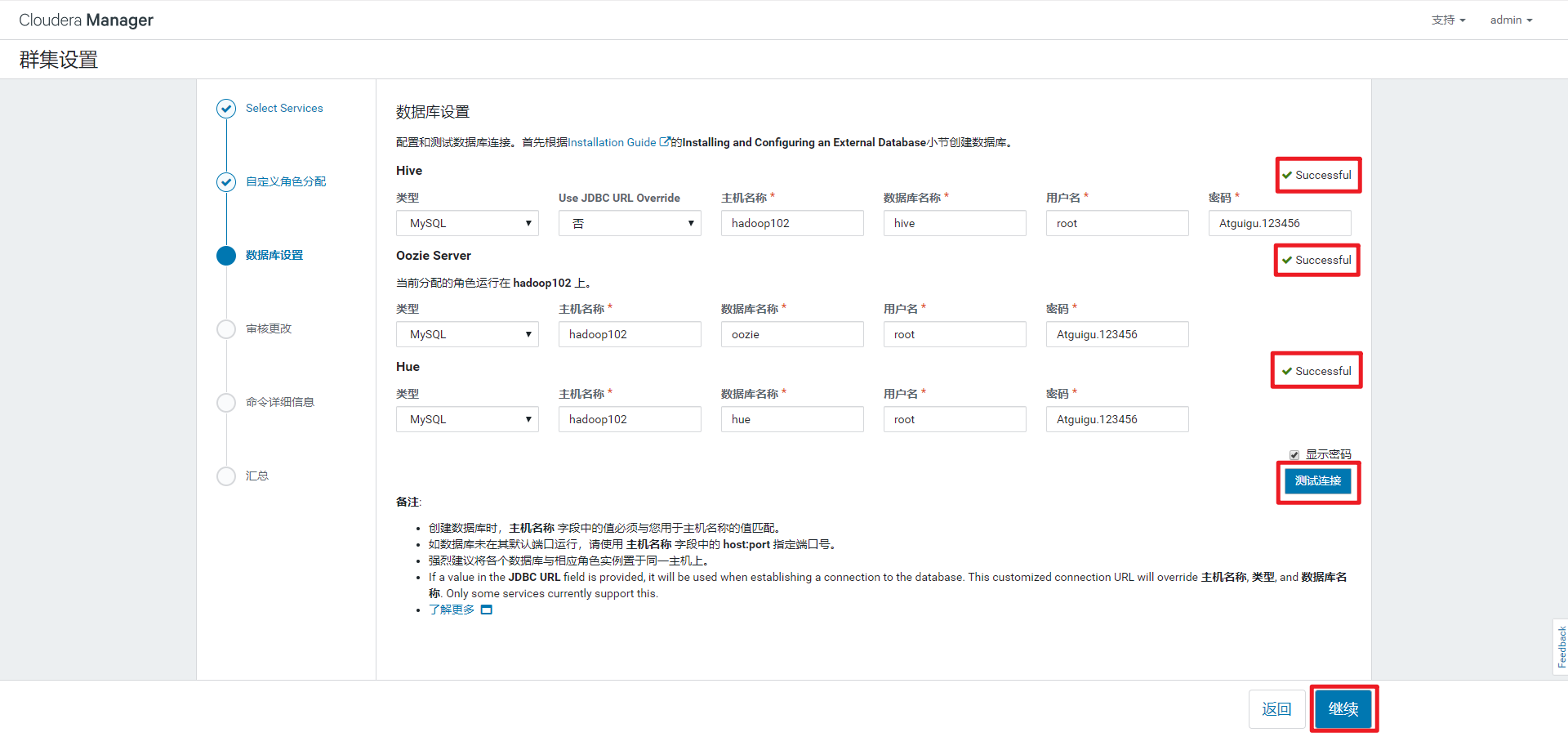


9）CDH各组件角色分布

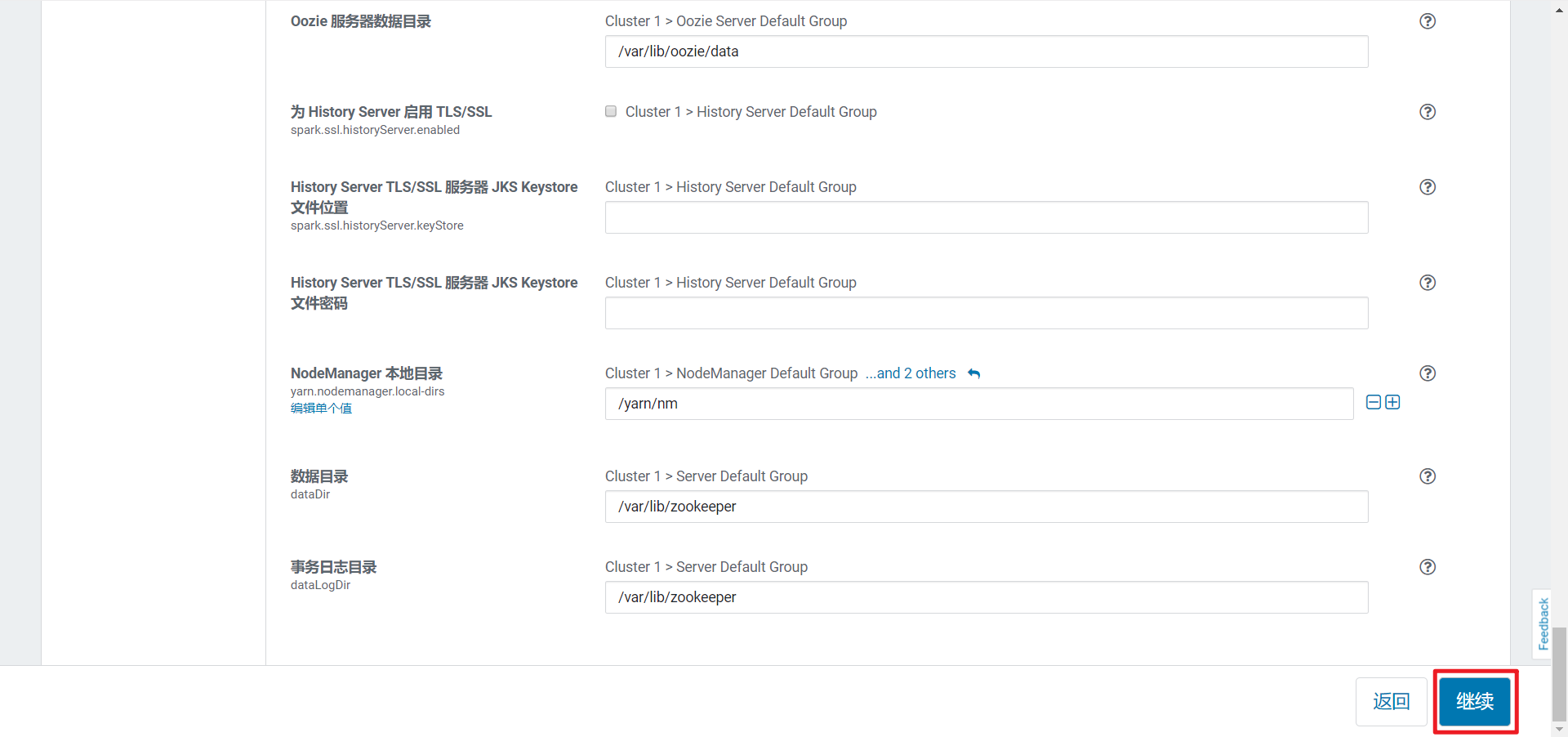




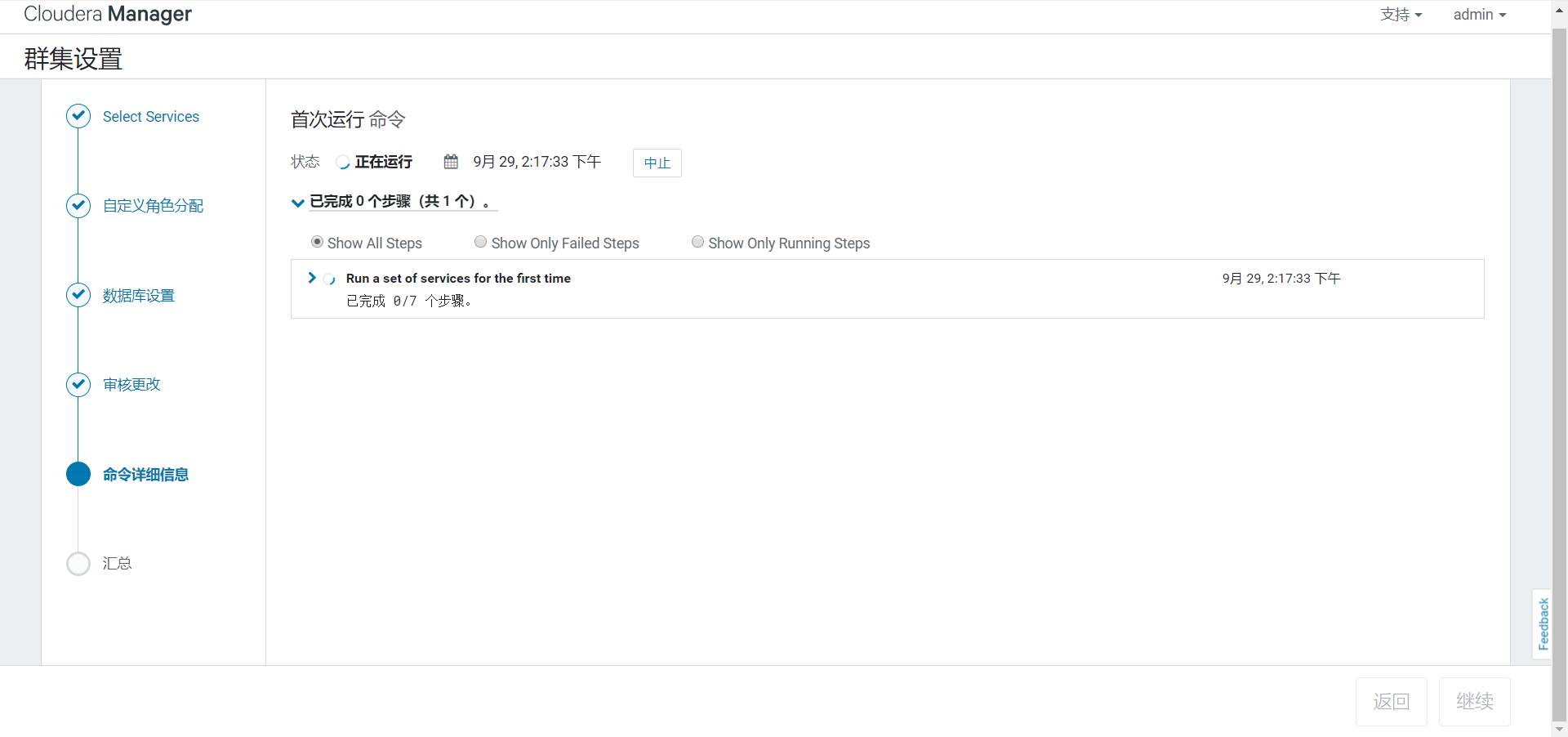
10）数据库连接测试

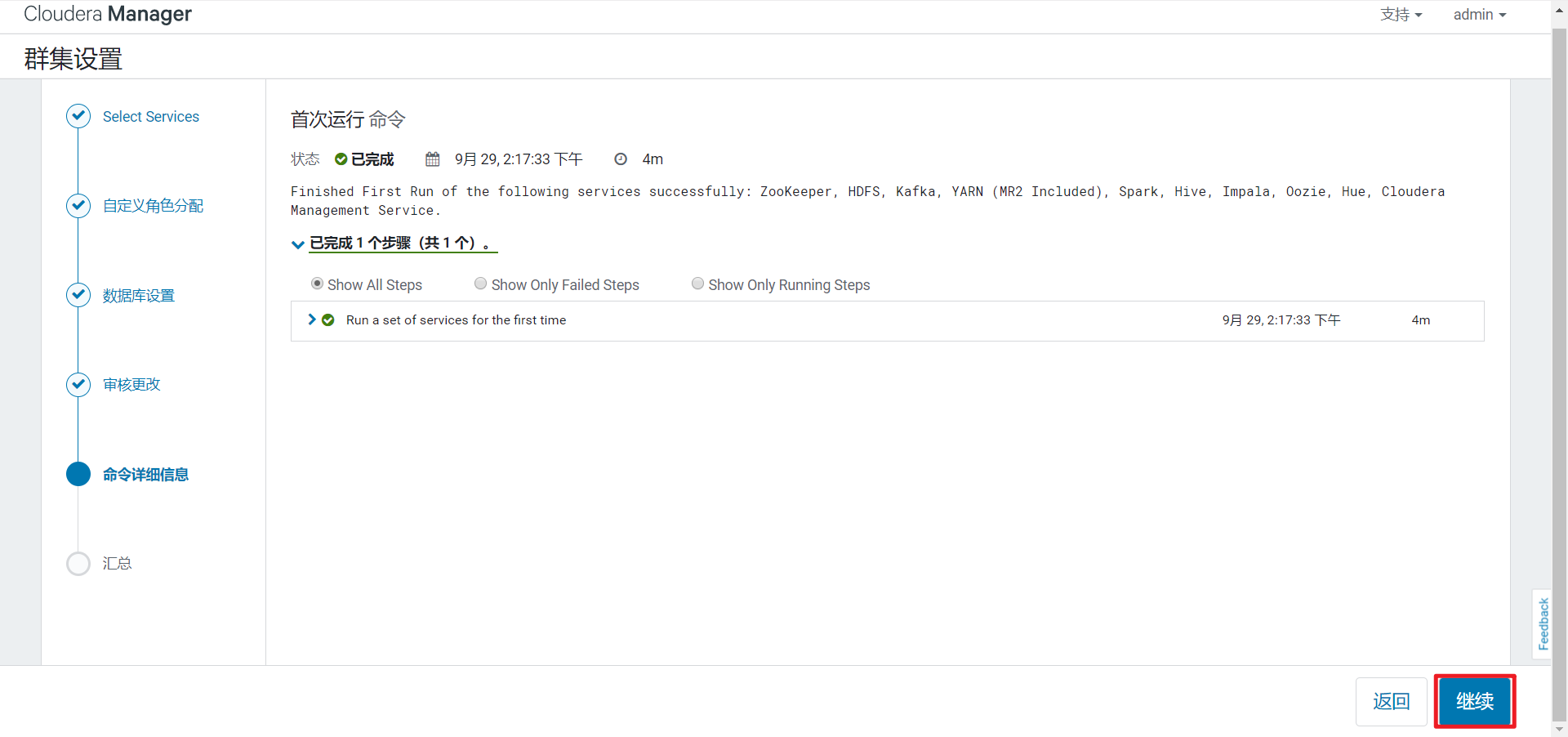


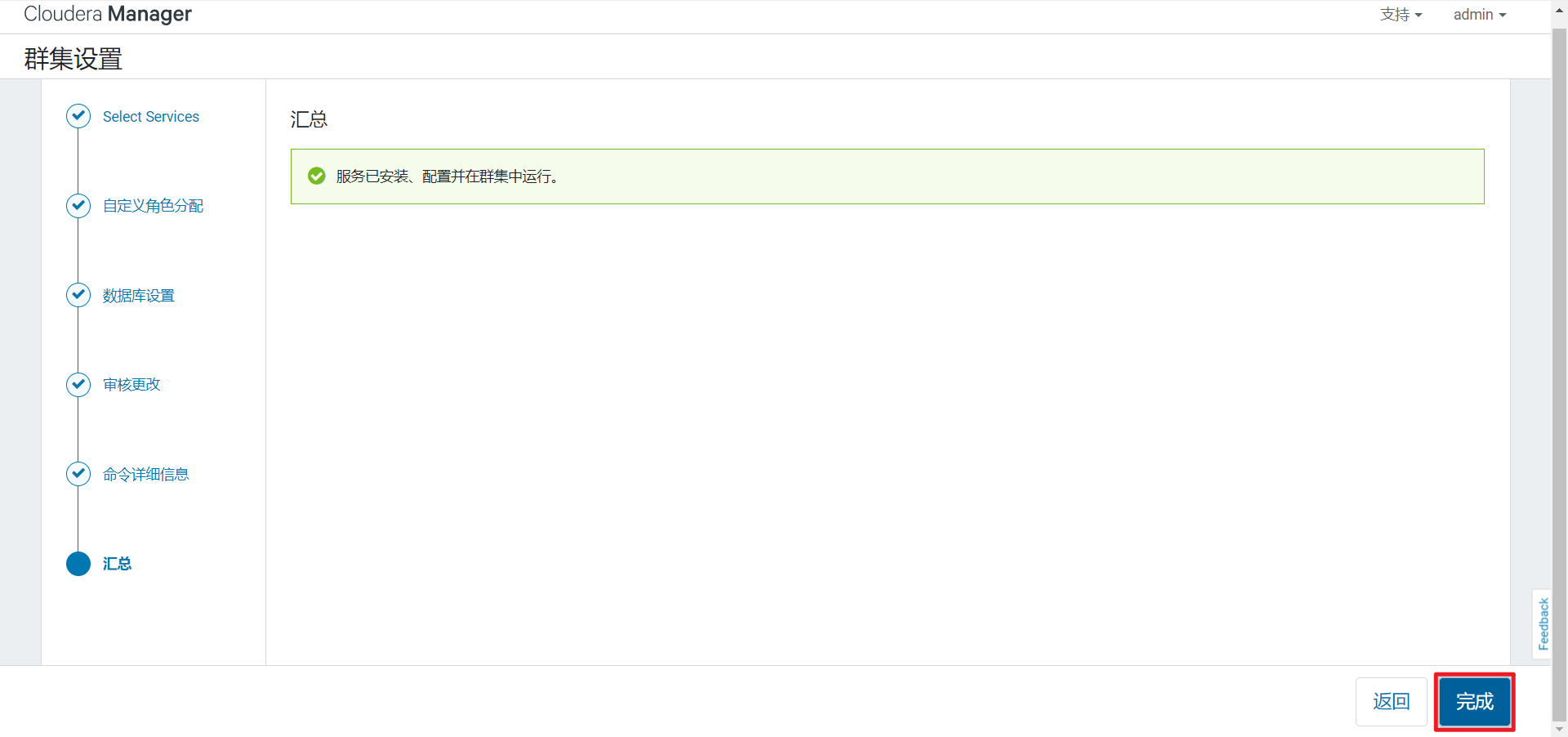
11）各组件基本设置，使用默认即可



12）等待安装部署和启动







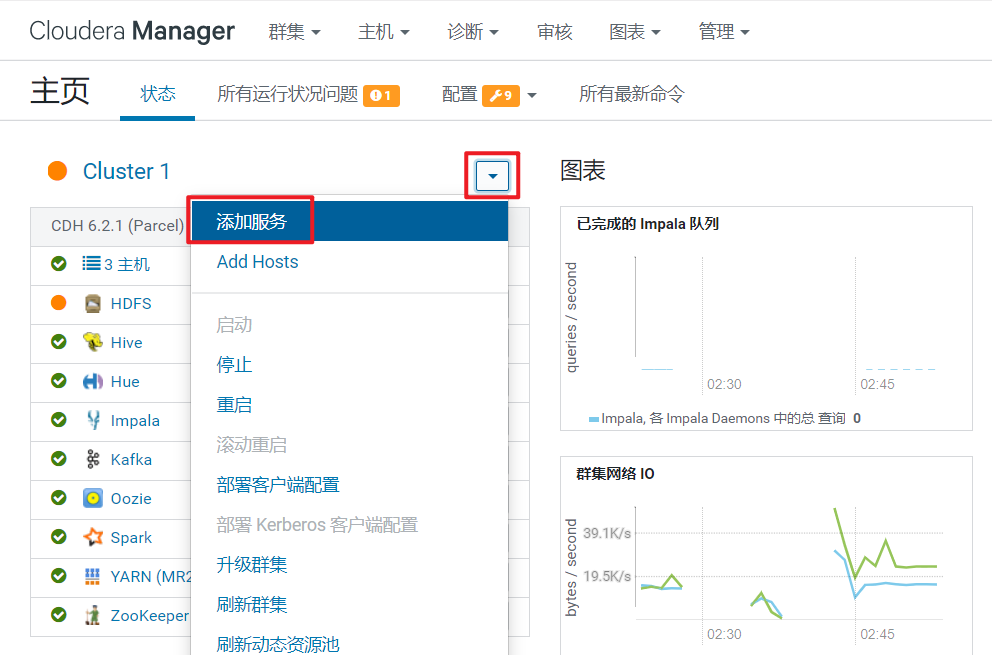
13）启动成功



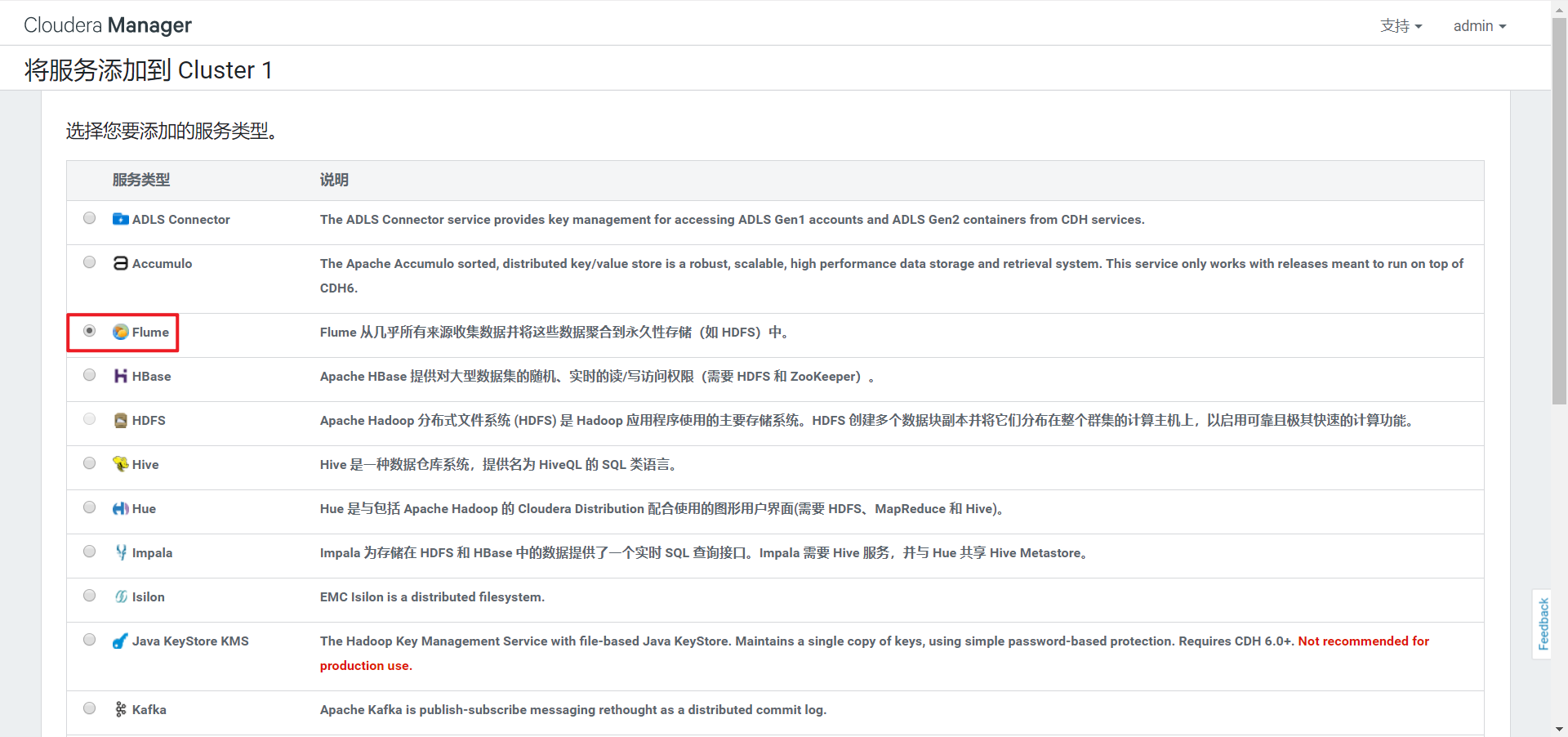
第3步 各组件安装部署

## 3.1 Flume 安装部署

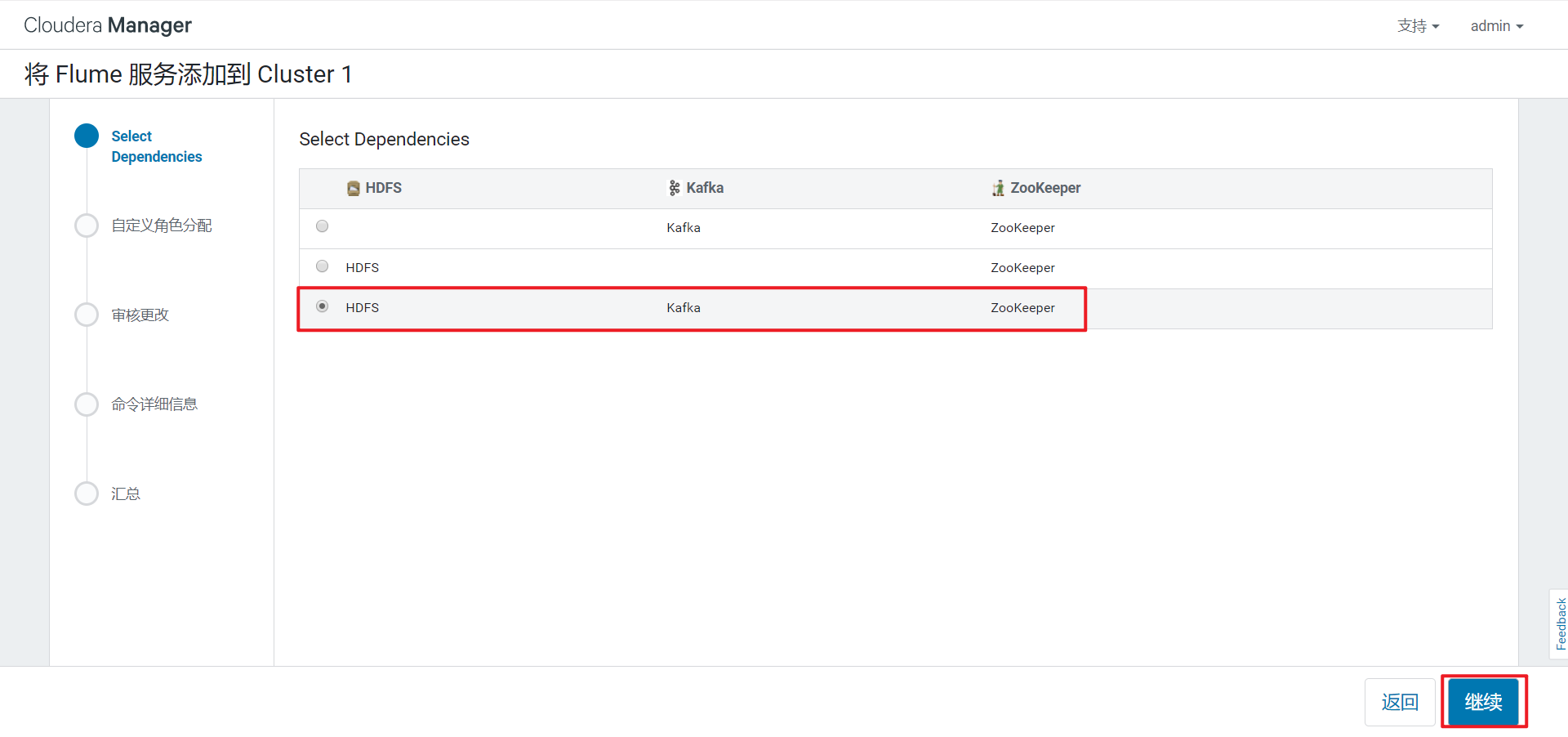
1）添加服务



2）选择Flume



3）选择依赖



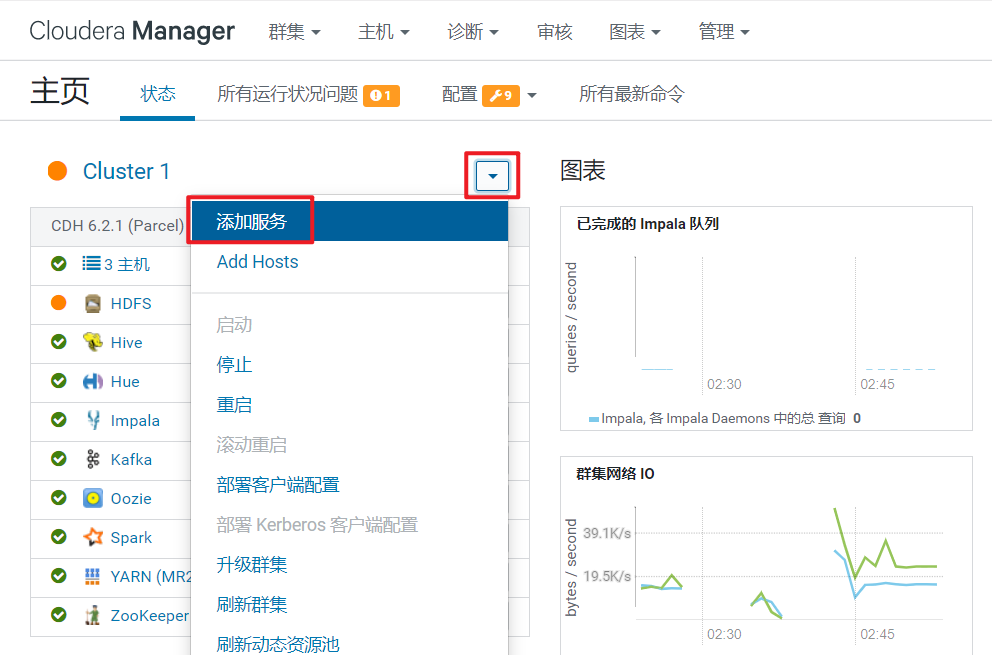
4）选择部署节点



5）完成

## 3.2 Sqoop安装部署

1）添加服务



2）选择Sqoop



3）选择部署节点



4）完成

## 3.3 配置Hadoop支持LZO

1）点击主机，在下拉菜单中点击Parcel



2）点击配置

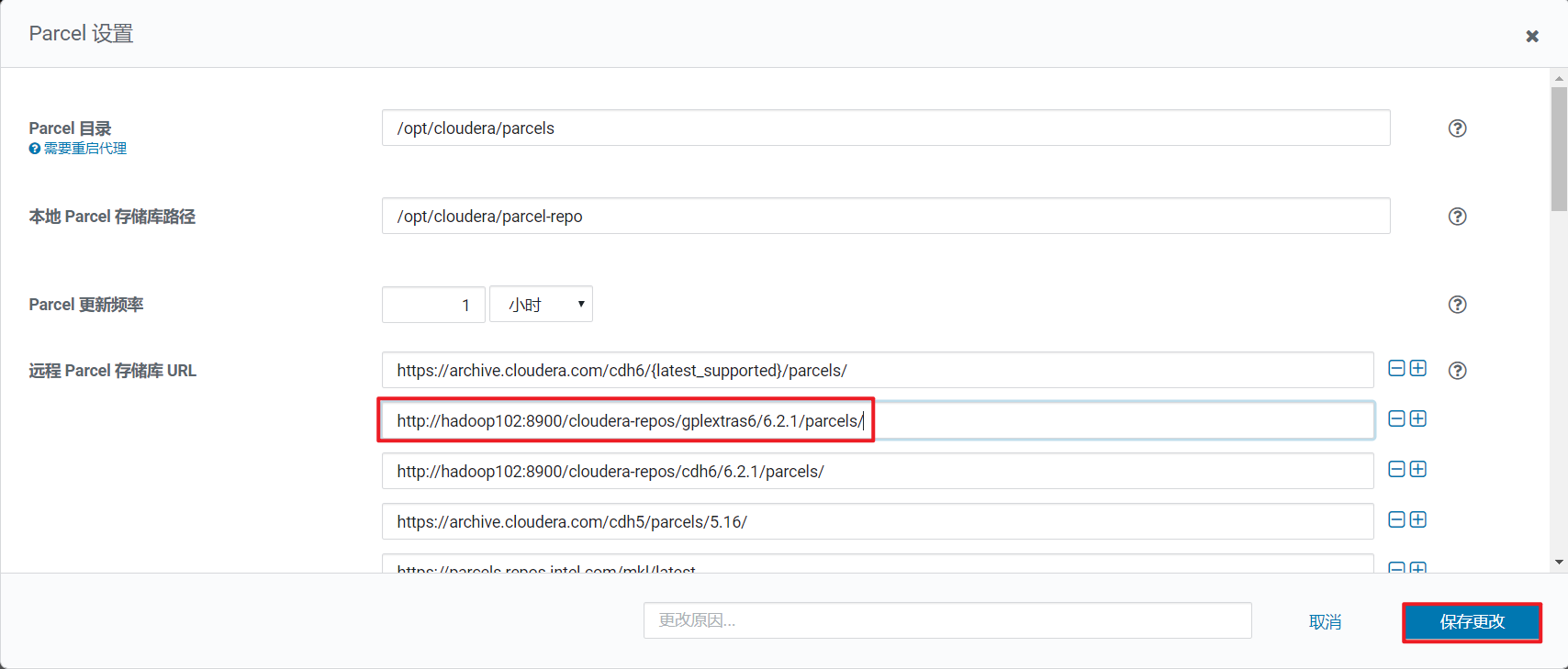


3）加上gplextras parcel库的url

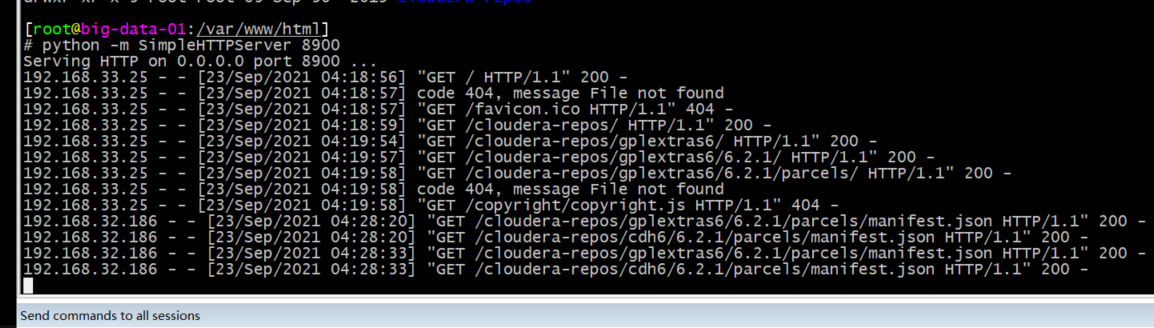
本地url: <http://big-data-01:8900/cloudera-repos/gplextras6/6.2.1/parcels/>

http://big-data-01:8900/cloudera-repos/gplextras6/6.2.1/parcels/

远程url: <https://archive.cloudera.com/gplextras6/6.2.1/parcels/>



下面这个服务一定不用停！！！！

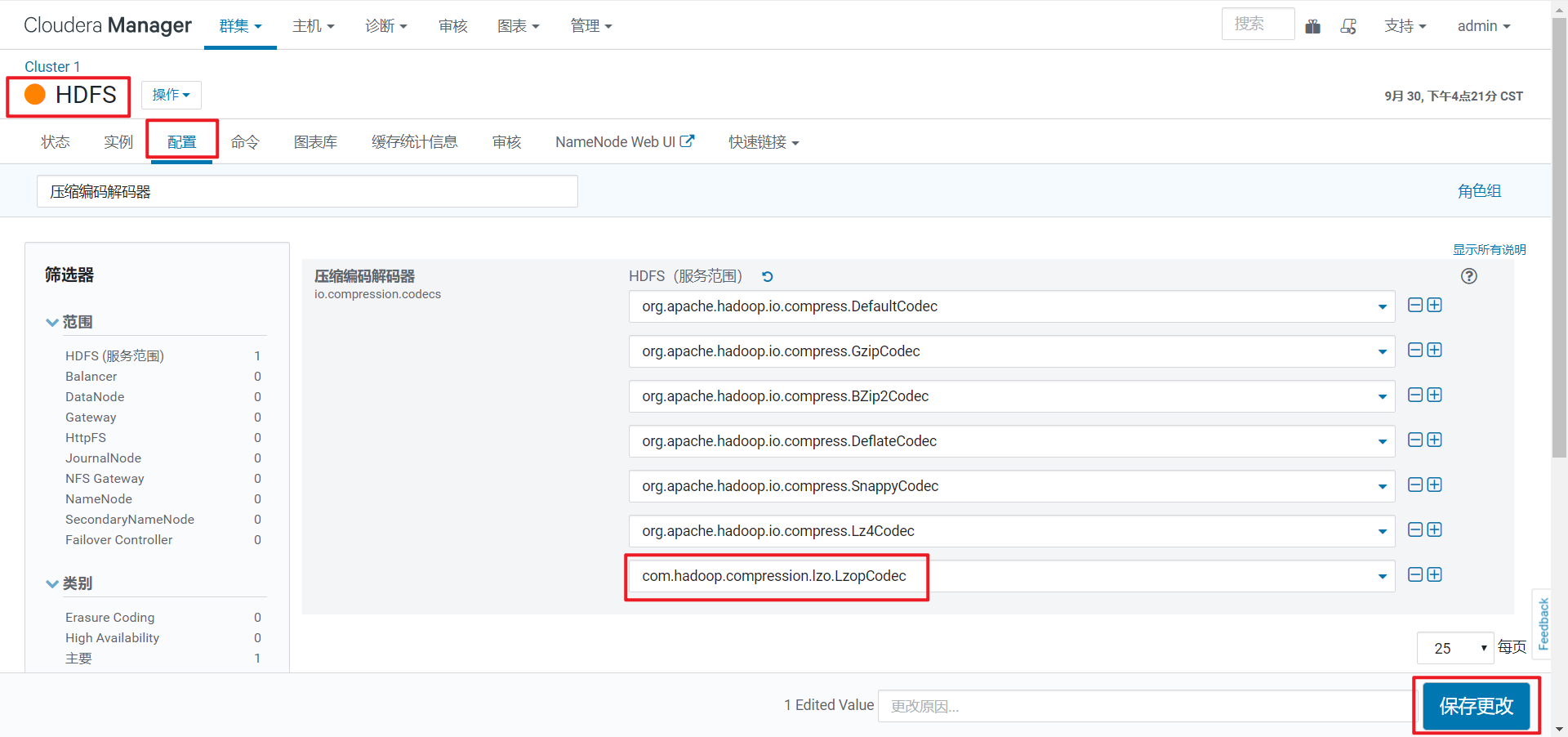


4）静待片刻，Parcel列表中出现了GPLEXTERAS，依次点击下载、分配、激活。



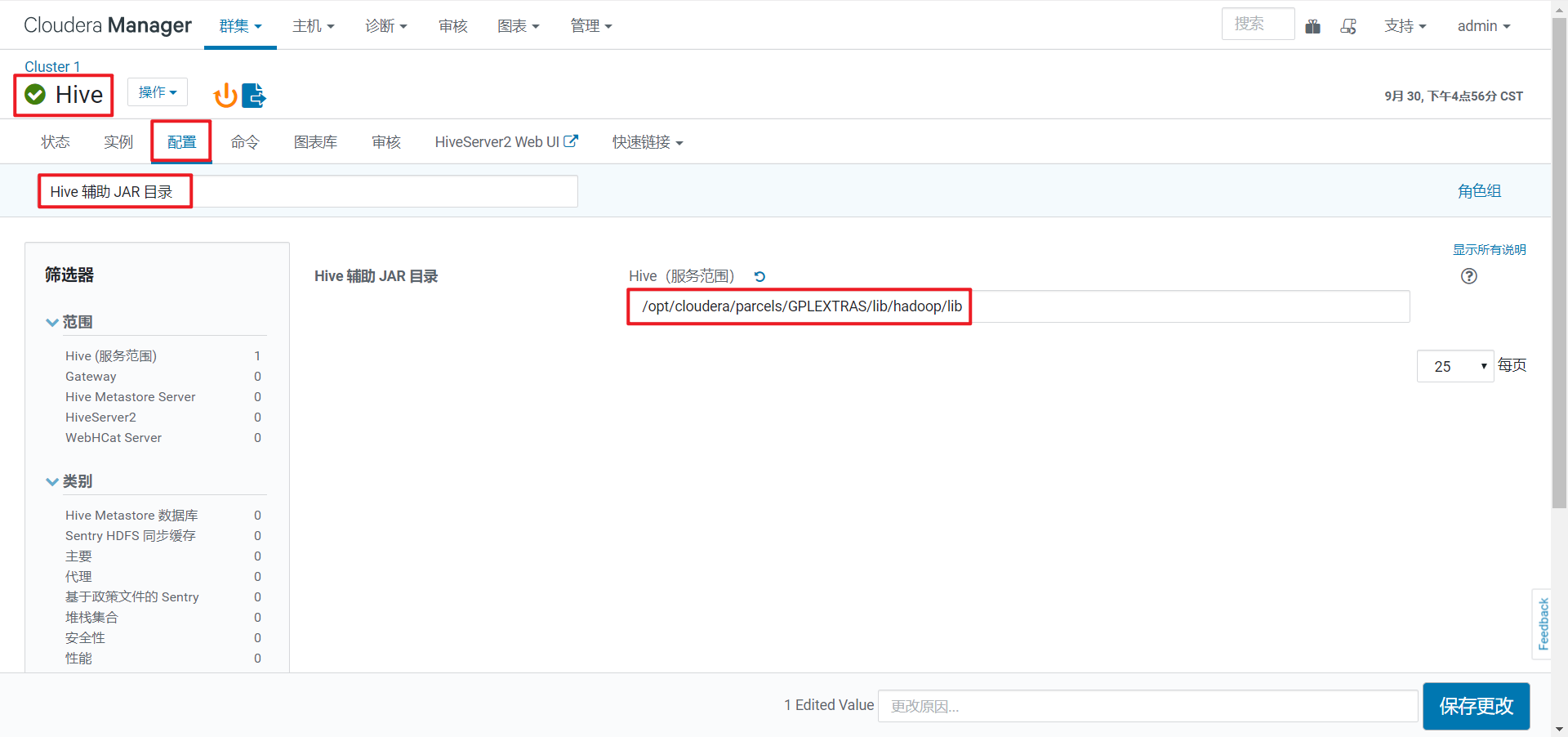
5）修改HDFS配置

在HDFS配置项中搜索“压缩编码解码器”，加入com.hadoop.compression.lzo.LzopCodec



6）修改Hive配置

在Hive配置项中搜索“Hive 辅助 JAR 目录”，加入/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib



7）修改Sqoop配置

在Sqoop的配置项中搜索“sqoop-conf/sqoop-env.sh 的 Sqoop 1 Client 客户端高级配置代码段（安全阀）”，加入以下字段

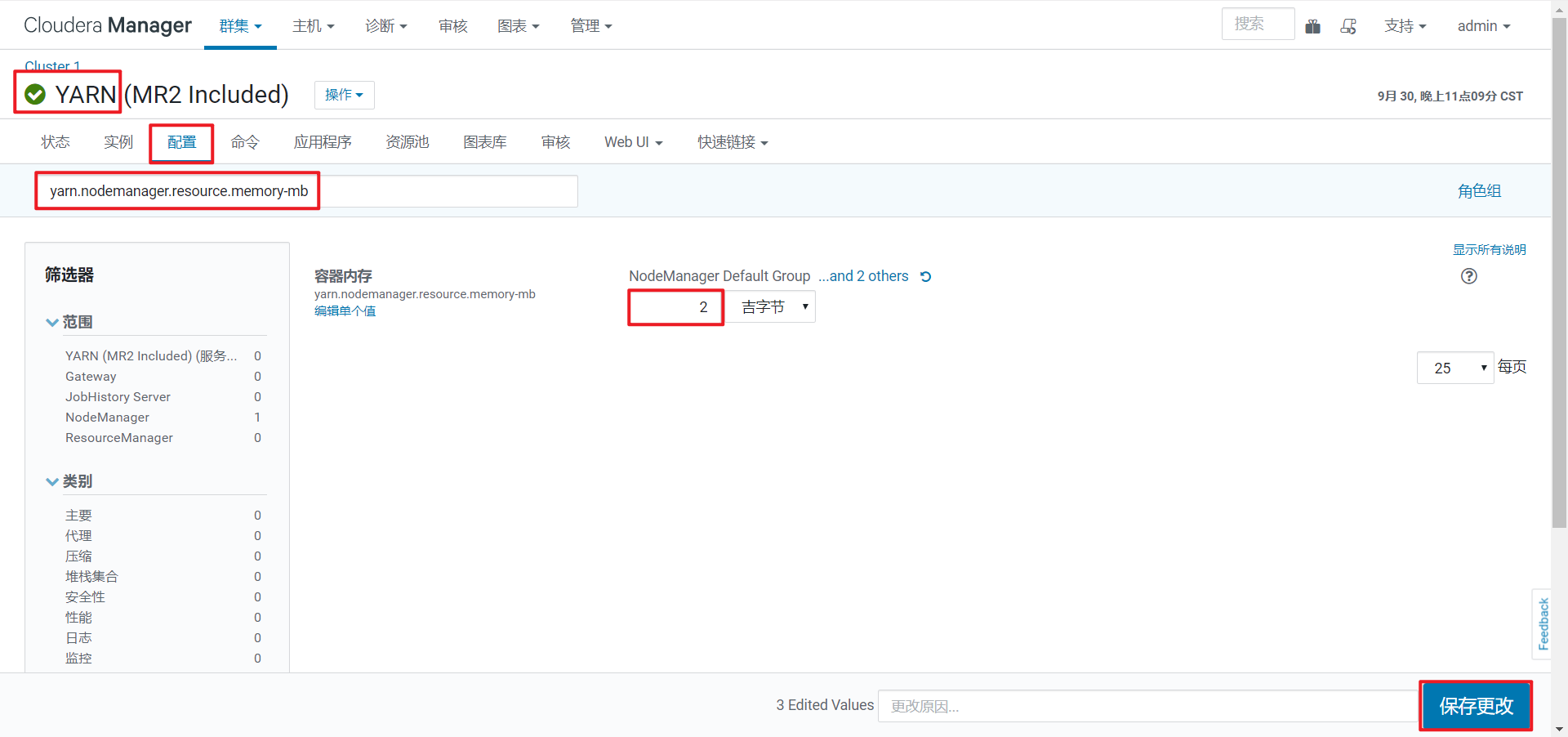
HADOOP\_CLASSPATH=$HADOOP\_CLASSPATH:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/\*

JAVA\_LIBRARY\_PATH=$JAVA\_LIBRARY\_PATH:/opt/cloudera/parcels/GPLEXTRAS/lib/hadoop/lib/native

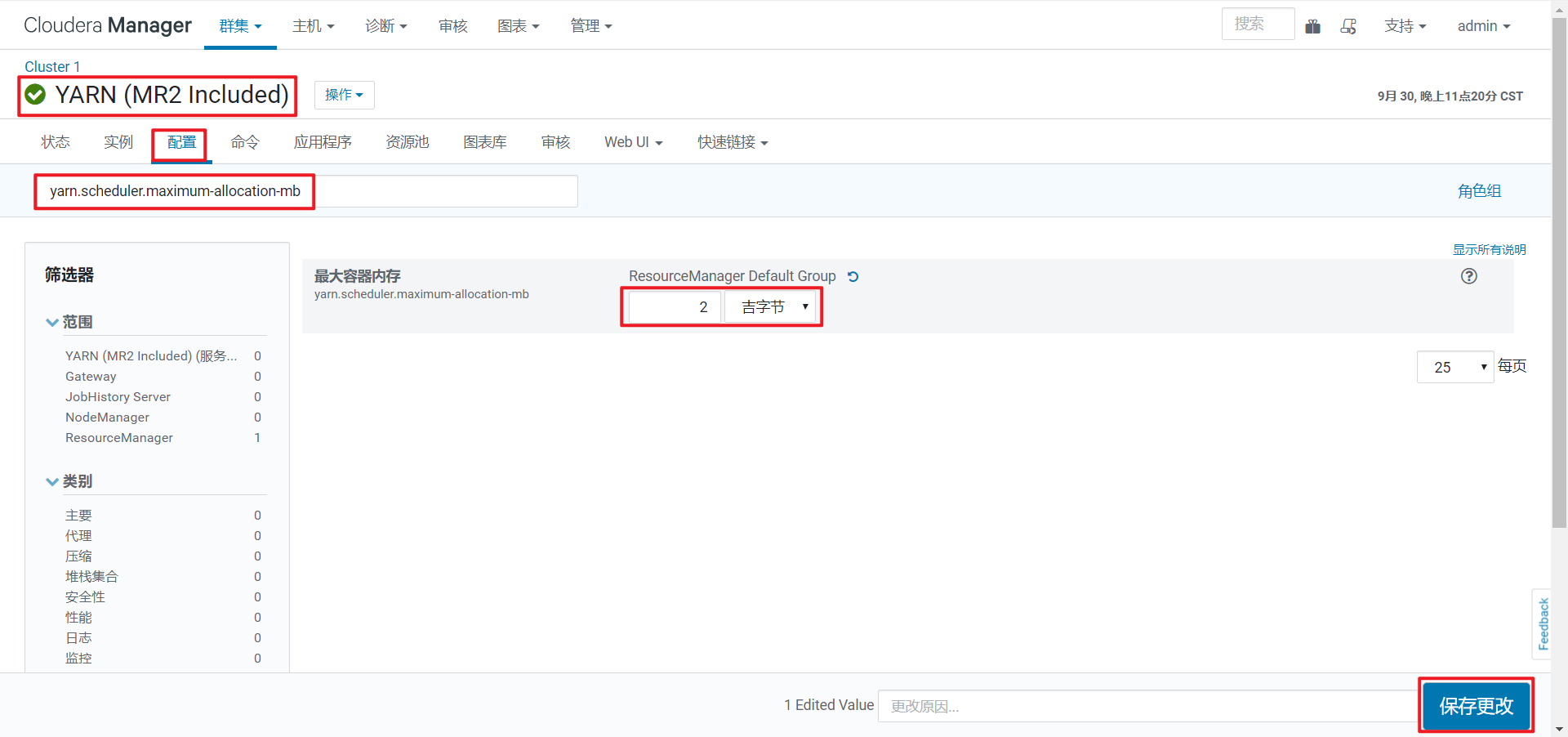


## 3.4 修改yarn配置参数

1）在yarn配置项中搜索“yarn.nodemanager.resource.memory-mb”，修改成4G。



2）在yarn配置项中搜索“yarn.scheduler.maximum-allocation-mb”，修改成2G。



3）重启相关组件



## 3.5 HUE使用概述

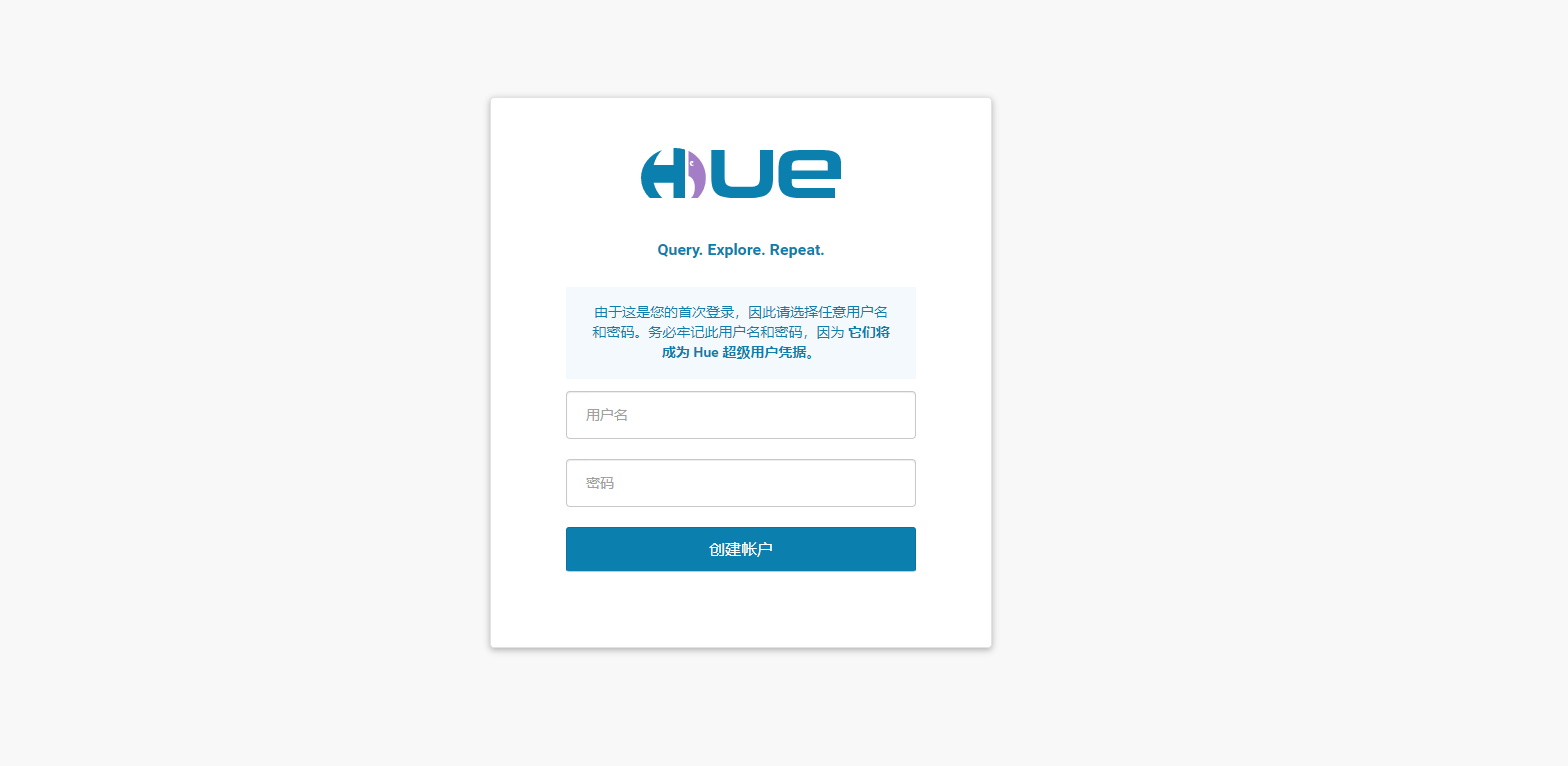
1）HUE官网及使用者

官网网站：<http://gethue.com/>

2）HUE页面

http:// big-data-01:8888（未优化）或http:// big-data-01:8889（优化）

第一次开启HUE会出现以下页面，此时输入的用户名和密码可以随意，之后登录页面以第一次输入的账号密码为依据。例如，用户名：hue 密码：hue123

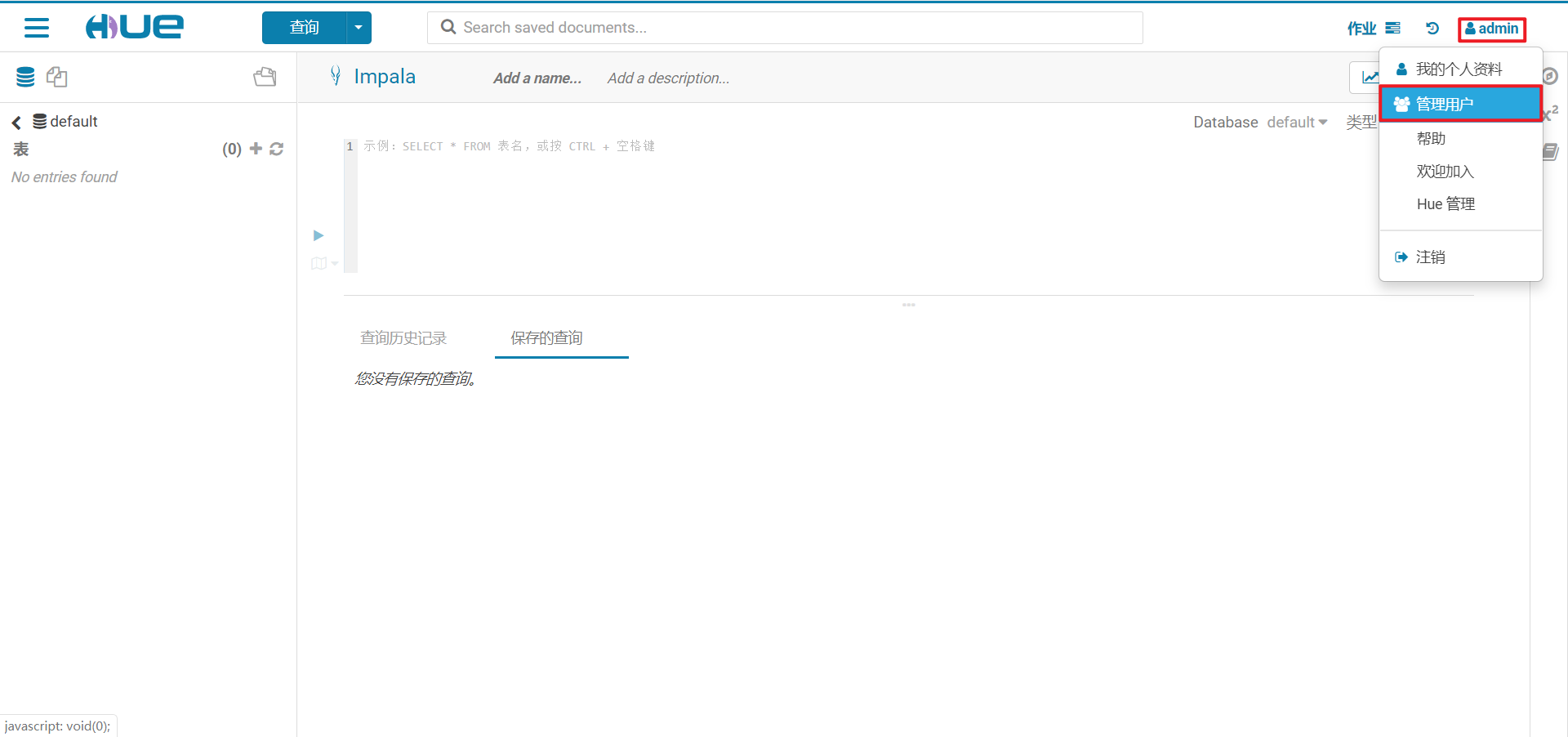


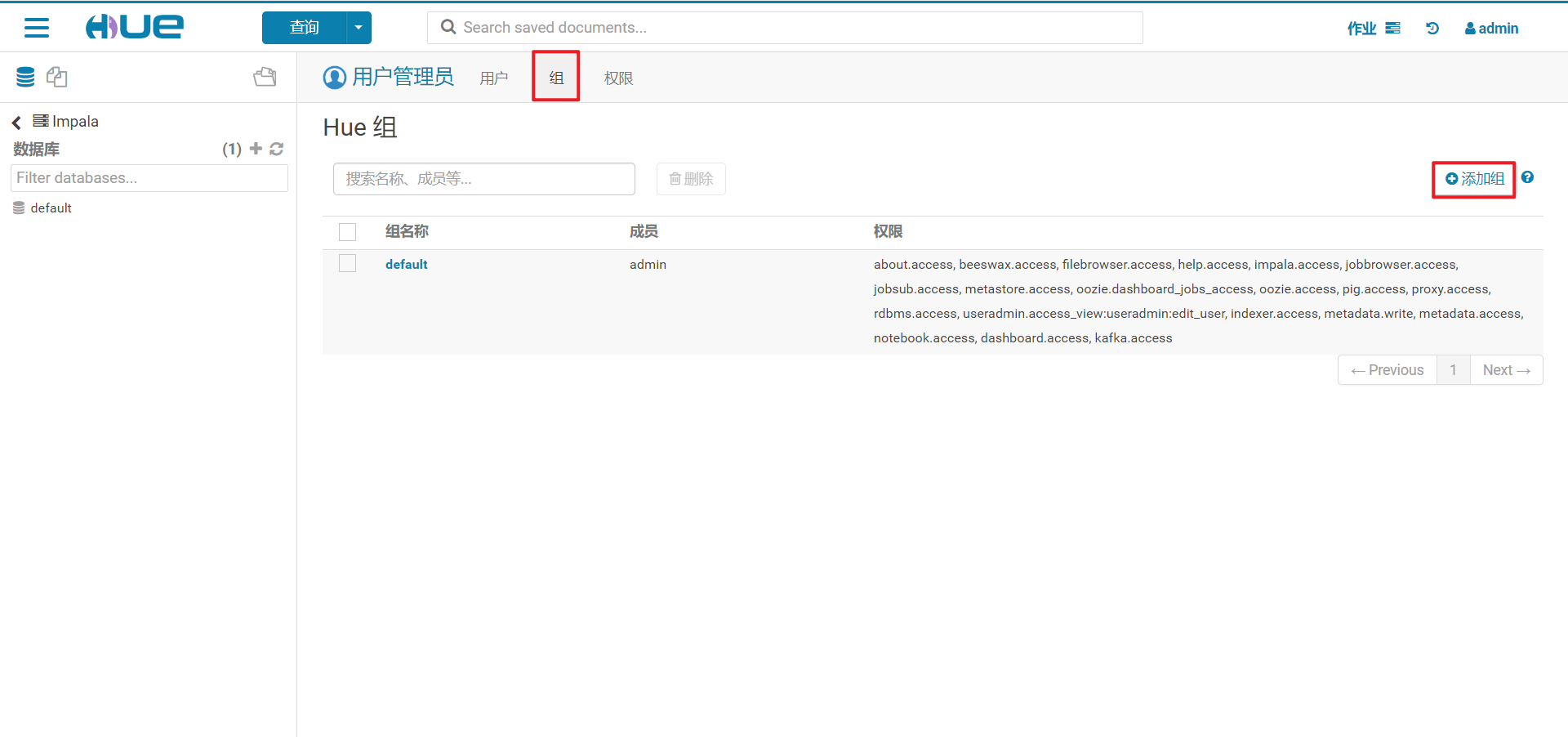
## 3.6 HUE用户管理

HUE的初始管理用户为admin，密码为admin。

1）在HUE中新建一个用户组——hive，并在该组下新建一个用户——hive。

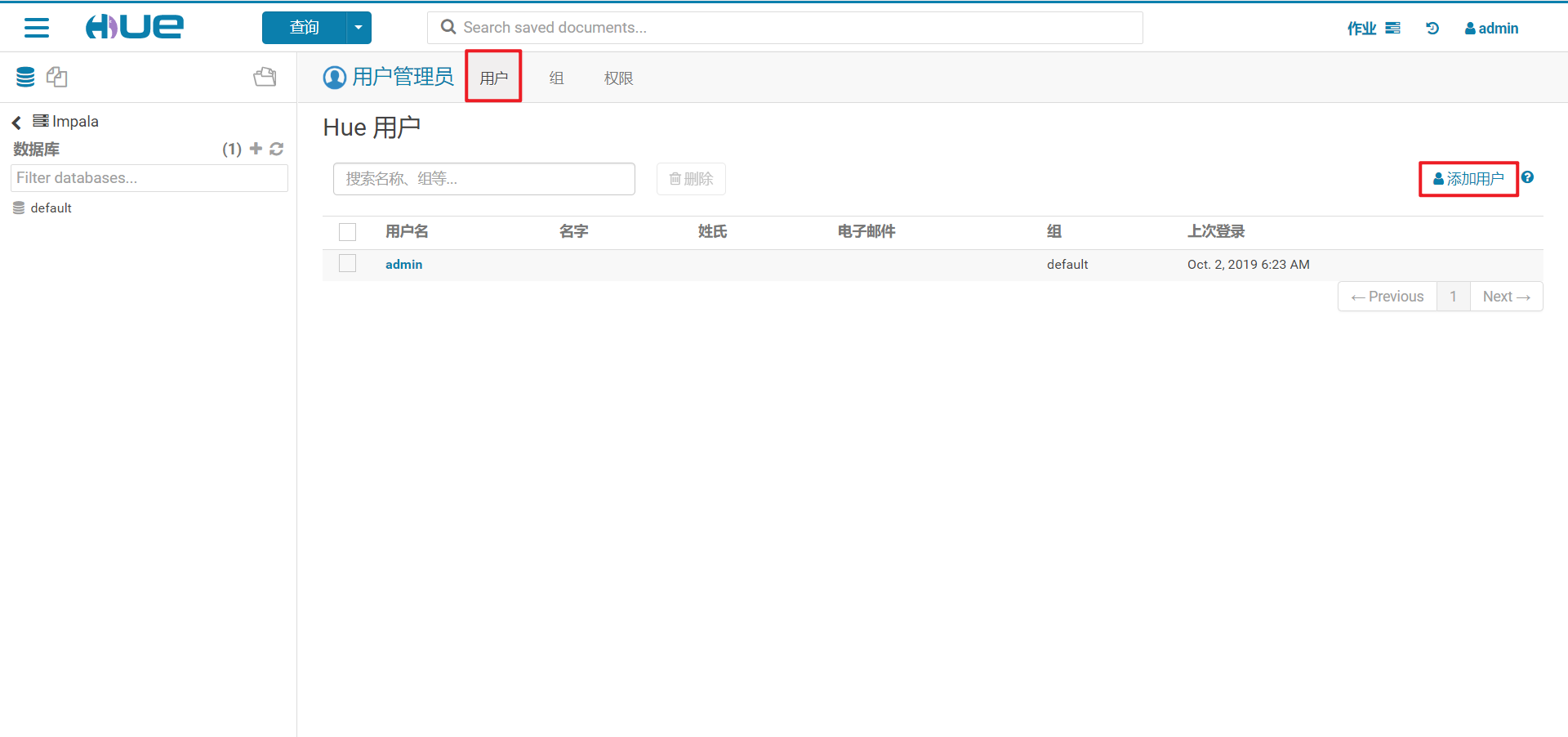
（1）创建hive组



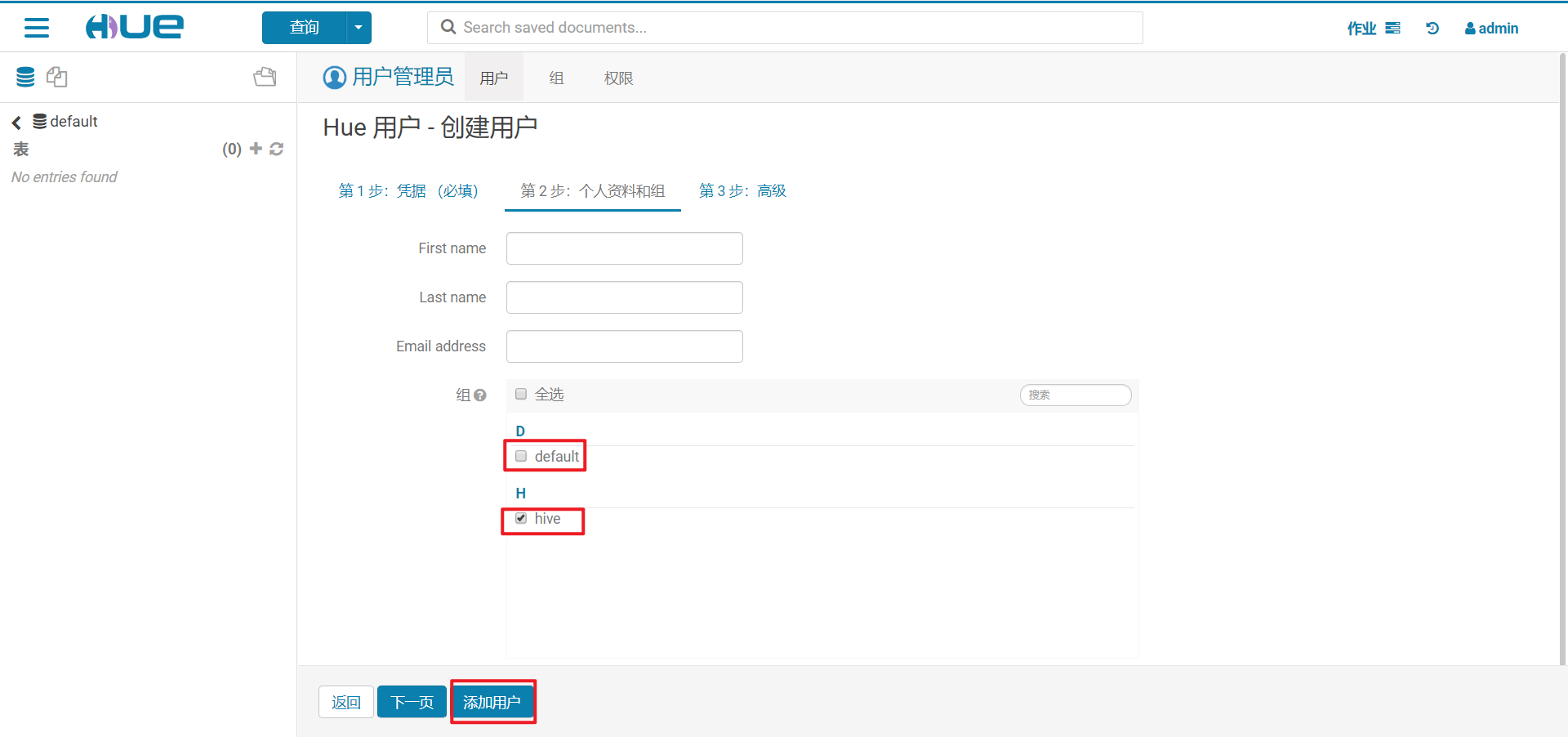




（2）创建hive用户



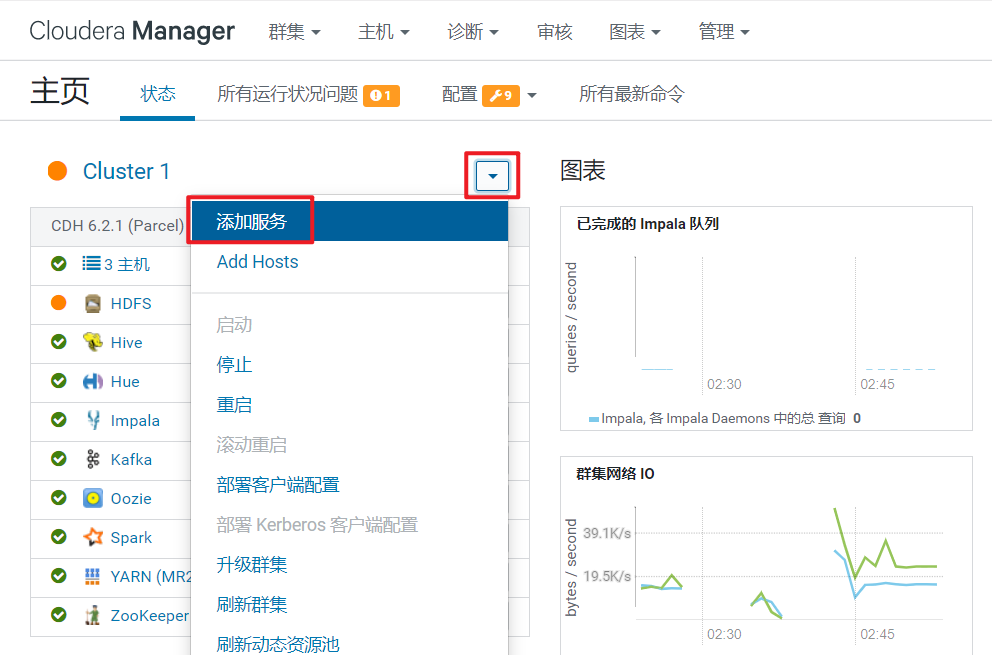




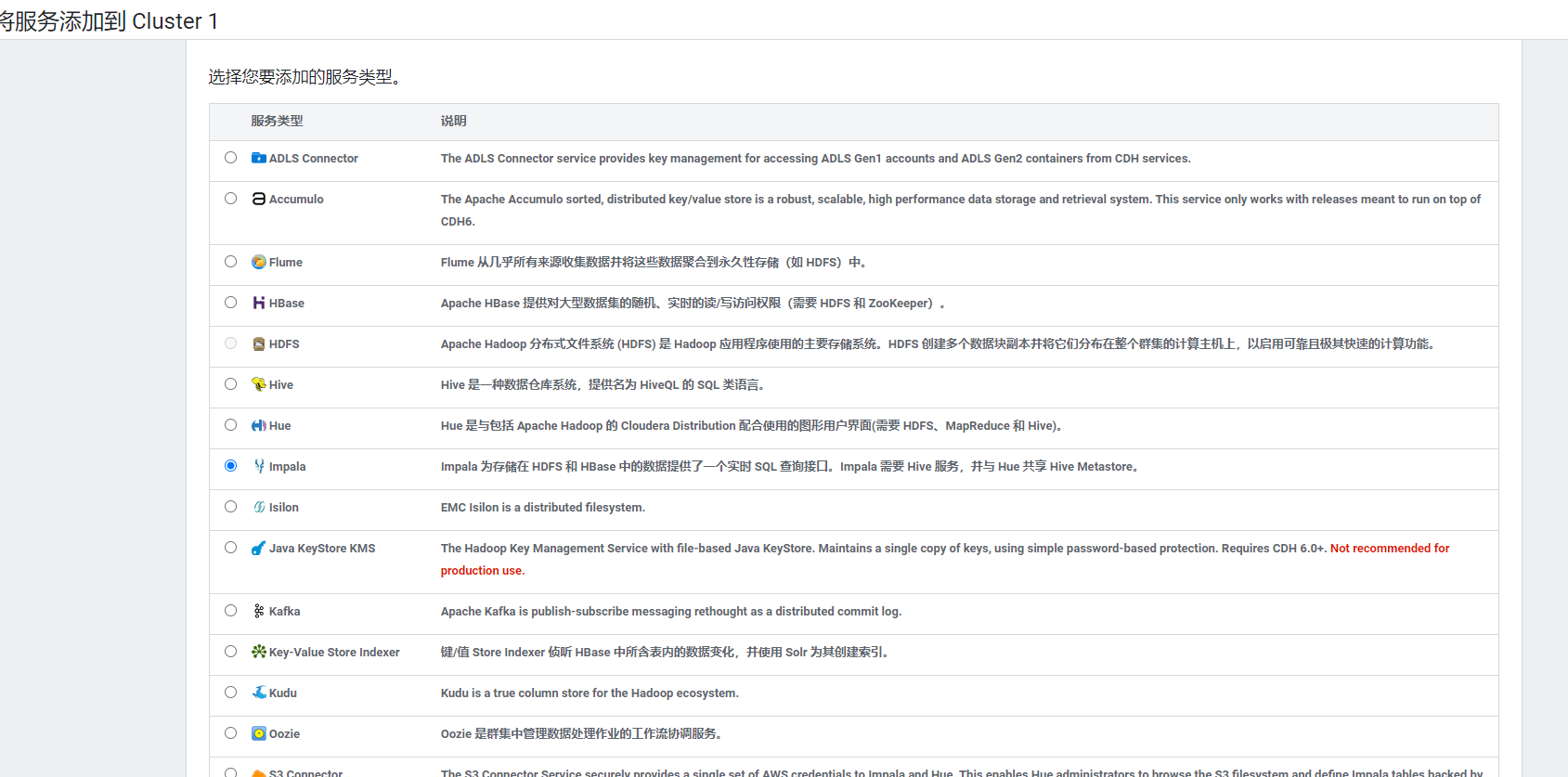
2）切换为hive用户

## 3.7 impala安装部署

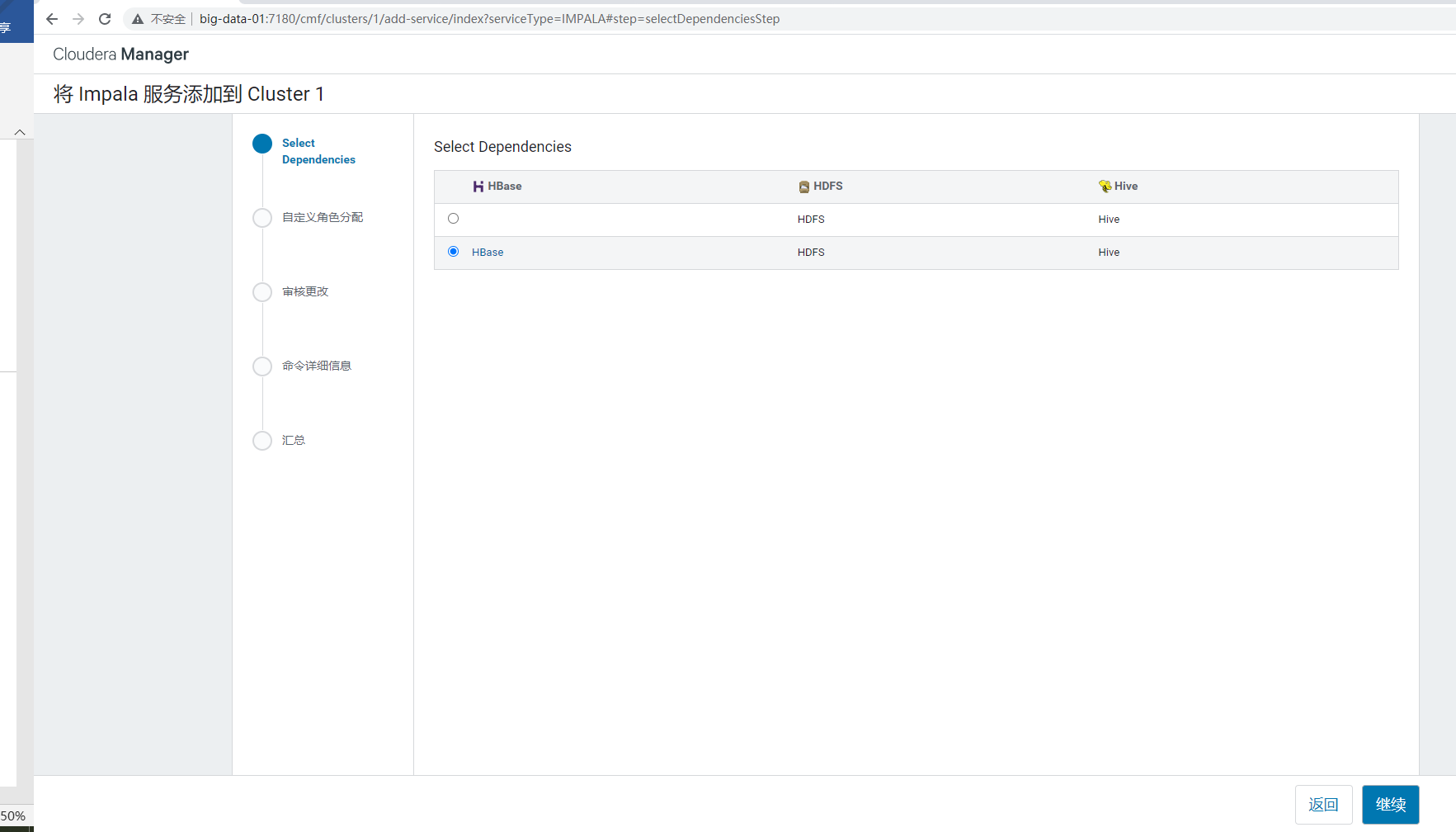
1）添加服务



2）选择impala



3）选择依赖



4）选择部署节点



1. 完成

### 3.7.1 配置Hue支持Impala

1）进入HUE配置页面，搜索“impala”，开启HUE中的impala服务

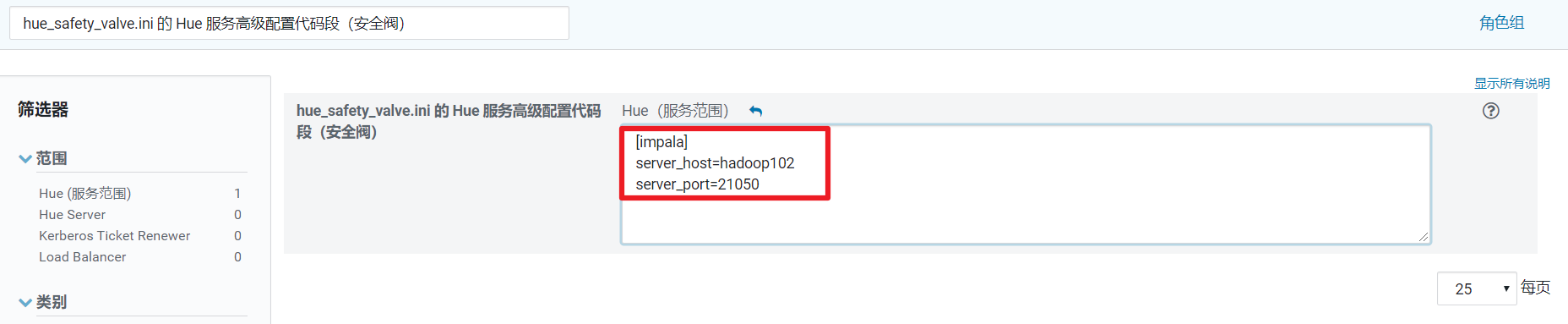


2）搜索“**hue\_safety\_valve.ini 的 Hue 服务高级配置代码段（安全阀）**”，输入以下代码段，确定HUE支持impala搜索引擎

[impala]

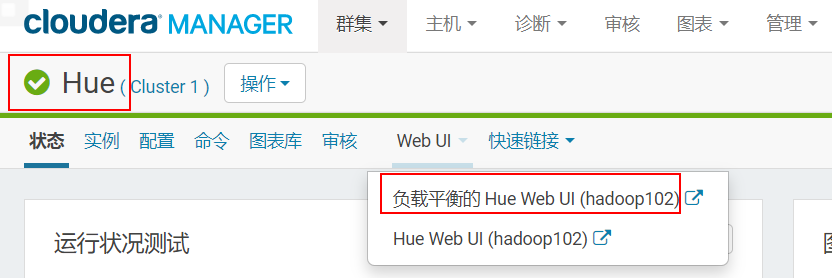
server\_host=big-data-01

server\_port=21050

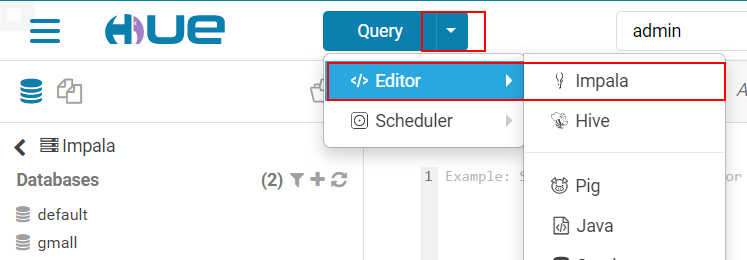


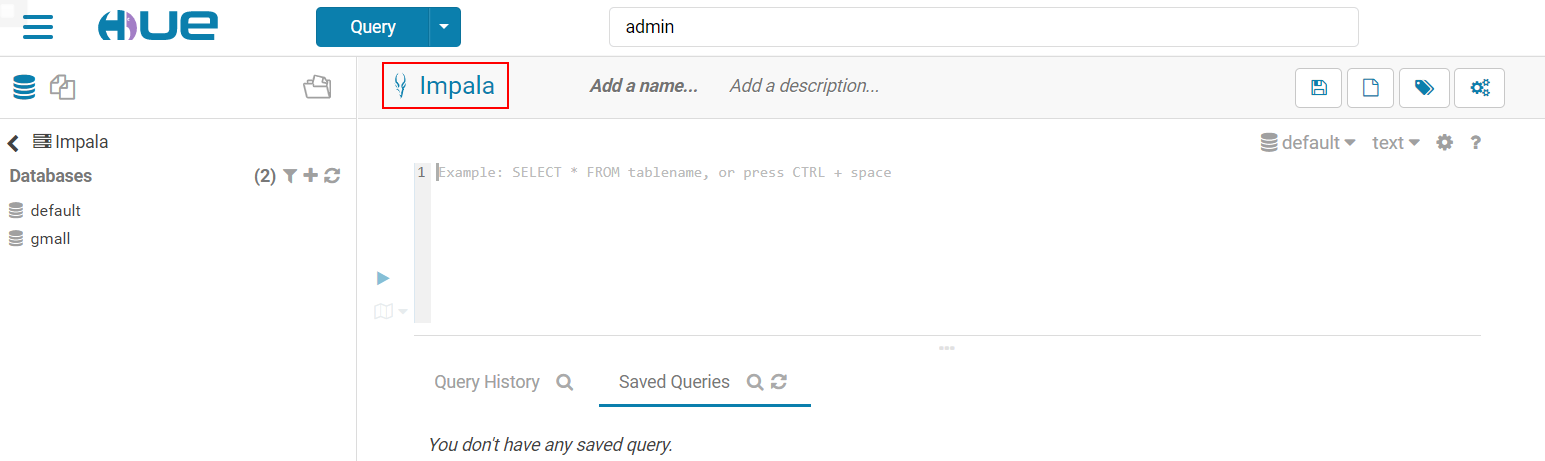
3.8 Impala基于Hue查询

3.8.1 打开Hue



3.8.2 进入impala查询



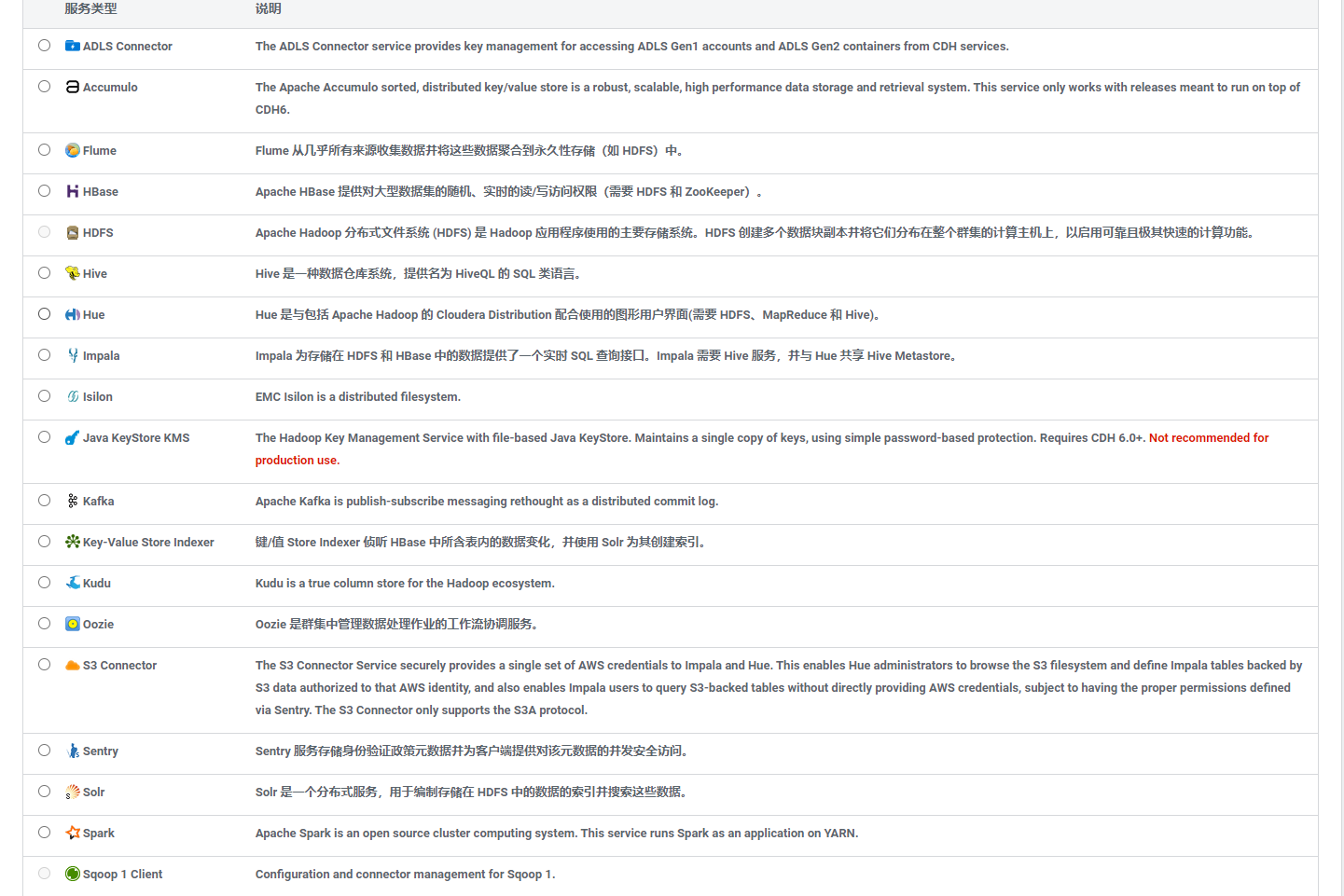


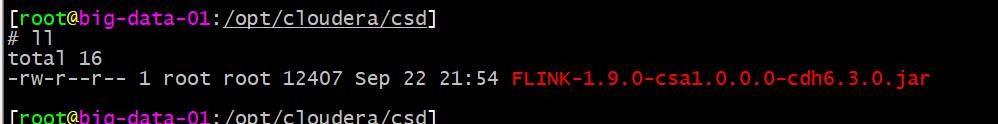
3.8 flink安装部署

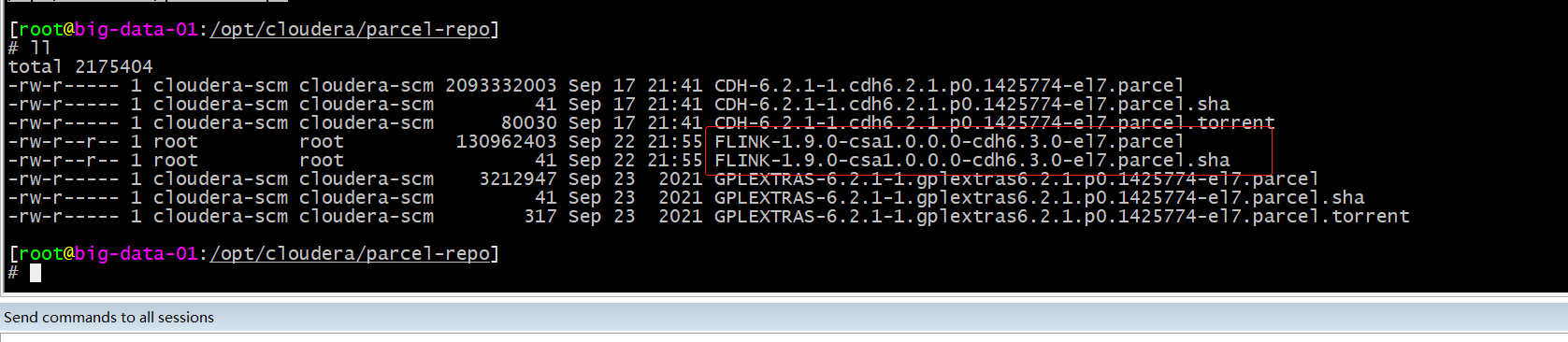
3.8.1 下载flink的各种配置

希望把flink也集成到cm管理平台，方便日常维护，但是cdh6.3后才支持flink服务，不能为了集成flink升级cdh版本。

现有的cdh6.2平台，通过cm管理，无法直接添加flink服务，如下图显示，并没有flink选项。







3.8.2 CM上分配激活

