



链滴

CDH 6.3.2 部署

作者: [k8s](#)

原文链接: <https://ld246.com/article/1608271790792>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

Cloudera Manager安装

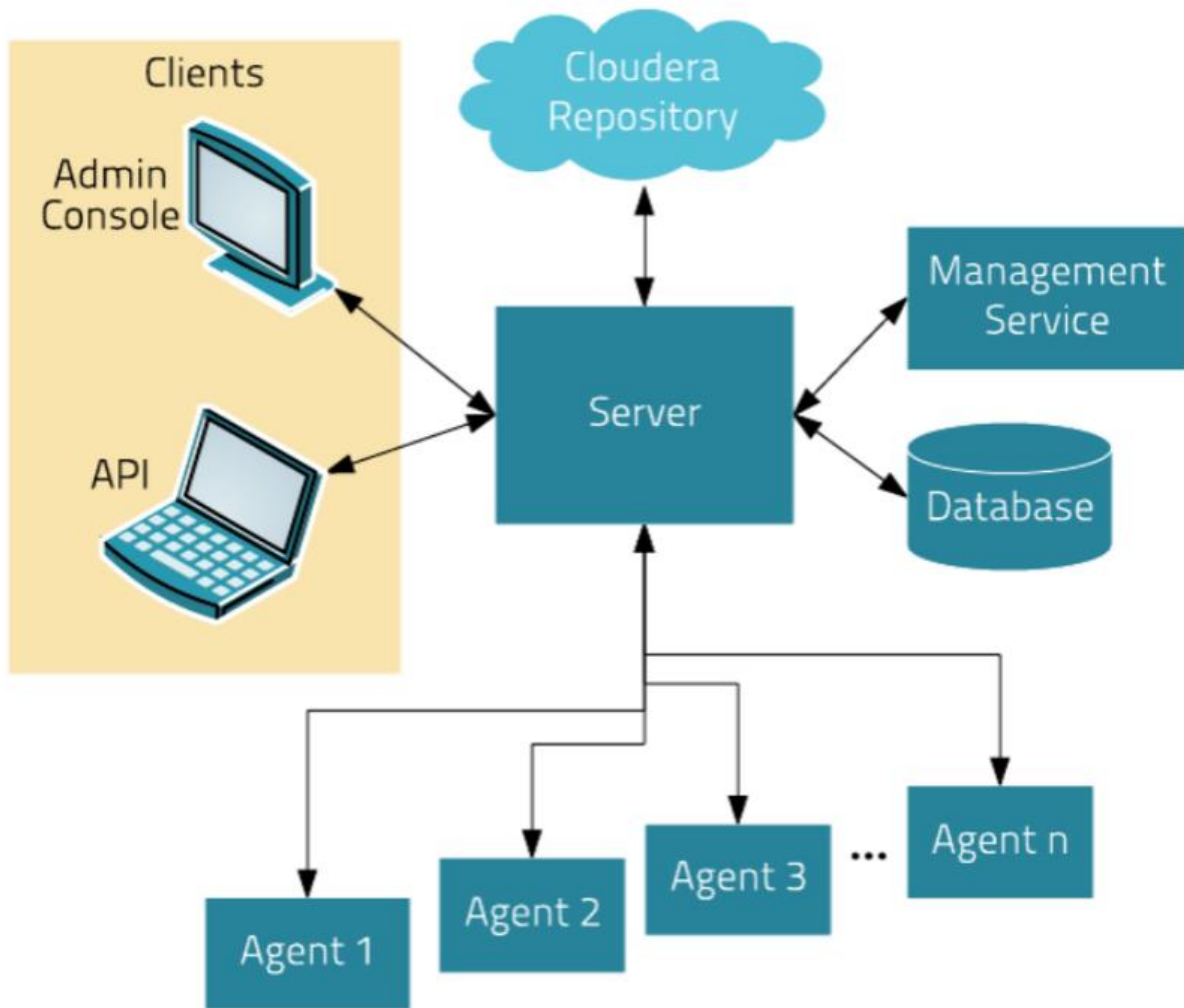
系统环境准备，安装基础环境

1. 选择四台已经安装 CentOS7 Linux系统的节点，分配资源。

安装CDH节点推荐内存为64G，大部分内存被Cloudera Management Service 占用，因为做了大量数据分析和整合。这里，划分四台节点如下：

主机名 件资源	IP地址	节点角色
cm-s1 C/32G/100G	192.168.0.50	Server,Agent
cm-a1 C/8G/100G	192.168.0.51	Agent
cm-a2 C/8G/100G	192.168.0.52	Agent
cm-a3 C/8G/100G	192.168.0.53	Agent

服务架构如下：



● Cloudera Manager Server是Cloudera Manager 的核心是，Server 管理控制台服务和托管应用程序逻辑，负责软件的安装、配置、服务的启动与关闭及管理集群。

- Agent安装在每台主机上，负责进程的启动和停止，解压配置，触发安装及监控主机。
- Management Service是由一组角色组成的服务，这些角色执行各种监视，警报和报告功能。
- DataBase用来存储配置及监控信息。
- ClouderaRepository是Cloudera Manager 分发软件的存储库。
- Clients分为AdminConsole（管理员web界面版）和API（用于开发者创建Cloudera Manager 序）两种形式。

2. 初始化节点

- 更改主机名分别为 cm-s1、cm-a1、cm-a2、cm-a3
- 配置 SSH 密钥验证，达到任一节点无需密码验证登录其他站点 root 用户的效果
- 配置/etc/hosts 文件，使任一节点通过 cm-s1、cm-a1、cm-a2、cm-a3 即可域名解析到对应点 IP
- 关闭SELinux和firewalld.service

3. 配置时间同步服务

在每个节点上安装chrony服务，并配置：

```
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i yum -y install chrony;done
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i wget https://sto.17ker.top/sources/chrony.conf -O /etc/chrony.conf;done
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i systemctl start chronyd;done
```

4. 安装JDK

给每台节点安装jdk，这里我们安装的CDH版本为6.3.2，在[官网](#)描述中这里需要使用jdk8，且对小版有要求，

Oracle JDK 8 versions that are tested and recommended

Oracle JDK Version	Notes
1.8u181	Recommended / Latest version tested
1.8u162	Recommended
1.8u141	Recommended
1.8u131	Recommended
1.8u121	Recommended
1.8u111	Recommended
1.8u102	Recommended
1.8u91	Recommended
1.8u74	Recommended
1.8u31	Minimum required

这里我们选择jdk8版本中的1.8u181版本安装。

```
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i rpm -ivh https://sto.17ker.top/sources/jdk-8u181-linux-x64.rpm;done
```

在每台节点配置jdk的环境变量：

```
[root@cm-s1 ~]# tail -n5 /etc/profile
#Java Environment variables
export JAVA_HOME=/usr/java/default
export CLASSPATH=.:$CLASSPATH:$JAVA_HOME/lib
export PATH=$PATH:$JAVA_HOME/bin
[root@cm-s1 ~]# source /etc/profile
```

5. 具体如下：

6. 安装MySQL数据库

在cm-s1节点安装mysql5.7

- 添加 mysql5.7 仓库

```
[root@cm-s1 ~]# rpm -ivh https://dev.mysql.com/get/mysql57-community-release-el7-11.no
rch.rpm
```

```
[root@cm-s1 ~]# yum makecache fast
```

- 安装MySQL

```
[root@cm-s1 ~]# yum -y install mysql-community-server
```

- 启动

```
[root@cm-s1 ~]# systemctl start mysqld
```

- 安全访问设置

- 查看默认 root 密码:

```
[root@cm-s1 ~]# cat /var/log/mysqld.log | grep -i 'temporary password'
```

- 更改 root 密码, 移除匿名用户:

```
[root@cm-s1 ~]# mysql_secure_installation
```

- 移除 root 用户远程登录限制:

```
[root@cm-s1 ~]# mysql -hlocalhost -pAz123456_ -e "UPDATE mysql.user SET host='192.168.0
%' WHERE user='root';FLUSH PRIVILEGES;"
```

7. 安装第三方依赖包

在每台节点上安装ClouderManager需要的第三方依赖包, 每台节点执行如下命令:

```
yum install -y chkconfig bind-utils psmisc cyrus-sasl-plain cyrus-sasl-gssapi portmap /lib/lsb/
nit-functions httpd mod_ssl openssl-devel python python-psycopg2 MySQL-python libxslt zli
sqlite fuse fuse-libs redhat-lsb
```

至此, 安装Cloudera Manager的基础环境准备完成。

Cloudera Manager 组件安装

1. 添加yum源

```
$ sudo wget https://archive.cloudera.com/cm6/6.3.1/redhat7/yum/cloudera-manager.repo -P
/etc/yum.repos.d/
```

```
$ sudo rpm --import https://archive.cloudera.com/cm6/6.3.1/redhat7/yum/RPM-GPG-KEY-cl
udera
```

```
$ sudo yum makecache fast
```

2. 在cm-s1节点安装 CM Server

```
[root@cm-s1 ~]# yum install -y cloudera-manager-daemons cloudera-manager-server cloude
a-manager-agent
```

3. 在cm-a1、cm-a2、cm-a3节点安装daemons、agent RPM包

```
[root@cm-a1 ~]# yum install -y cloudera-manager-daemons cloudera-manager-agent
```

当完成安装agent后, 会在当前节点生成 "/opt/cloudera/cm-agent" 目录。

4. 配置CM Agent的server host

在所有的节点上，修改Agent的配置文件 `/etc/cloudera-scm-agent/config.ini`，修改 `server_host=m-s1`，指定所有Agent节点的Server节点为cm-s1。

```
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i sed -i 's/server_host=localhost/server_host=cm-s1/' /etc/cloudera-scm-agent/config.ini;done
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i grep server_host /etc/cloudera-scm-agent/config.ini;done
```

5. 配置CM Server 数据库

在cm-s1节点上，修改 `/etc/cloudera-scm-server/db.properties`文件，配置内容如下：

```
com.cloudera.cmf.db.type=mysql
com.cloudera.cmf.db.host=cm-s1
com.cloudera.cmf.db.name=cmf
com.cloudera.cmf.db.user=cmf
com.cloudera.cmf.db.password=Az123456_
com.cloudera.cmf.db.setupType=EXTERNAL
```

上传 `mysql-connector-java-5.1.48.jar`文件到cm-s1节点的`/usr/share/java`目录下，如果没有目录要创建，并且软链接到 `mysql-connector-java.jar`

```
[root@cm-s1 ~]# mkdir -p /usr/share/java
[root@cm-s1 ~]# wget https://sto.17ker.top/sources/mysql-connector-java-5.1.48.jar -P /usr/share/java
[root@cm-s1 ~]# cd /usr/share/java/
[root@cm-s1 java]# ln -s mysql-connector-java-5.1.48.jar mysql-connector-java.jar
```

为MySQL服务器创建cmf用户

```
[root@cm-s1 ~]# mysql -uroot -hcm-s1 -pAz123456_ -e "grant all on *.* to 'cmf'@'%' identified by 'Az123456_' with grant option;"
```

在cm-s1节点上执行如下命令初始化数据库：

```
[root@cm-s1 ~]# cd /opt/cloudera/cm/schema/
# -u用户， -p密码， 数据库类型， 数据库， 用户， 密码
[root@cm-s1 schema]# ./scm_prepare_database.sh -ucmf -pAz123456_ mysql cmf cmf Az123456_
...
All done, your SCM database is configured correctly!
```

6. 准备 CDH Parcels 本地源

在cm-s1节点上创建 `/opt/cloudera/parcel-repo`目录，并下载如下文件：

```
[root@cm-s1 ~]# mkdir -p /opt/cloudera/parcel-repo
[root@cm-s1 ~]# wget -P /opt/cloudera/parcel-repo https://archive.cloudera.com/cdh6/6.3.2/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554-el7.parcel
[root@cm-s1 ~]# wget -O /opt/cloudera/parcel-repo/CDH-6.3.2-1.cdh6.3.2.p0.1605554-el7.parcel.sha https://archive.cloudera.com/cdh6/6.3.2/parcels/CDH-6.3.2-1.cdh6.3.2.p0.1605554-el7.parcel.sha1
[root@cm-s1 ~]# wget -P /opt/cloudera/parcel-repo https://archive.cloudera.com/cdh6/6.3.2/parcels/manifest.json
```

7. 启动 CM Server、Agent

在Server节点启动ClouderaManager Server，分别在Agent节点启动ClouderaManager Agent：

```
[root@cm-s1 ~]# systemctl start cloudera-scm-server.service
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i systemctl start cloudera-scm-agent.service;done
```

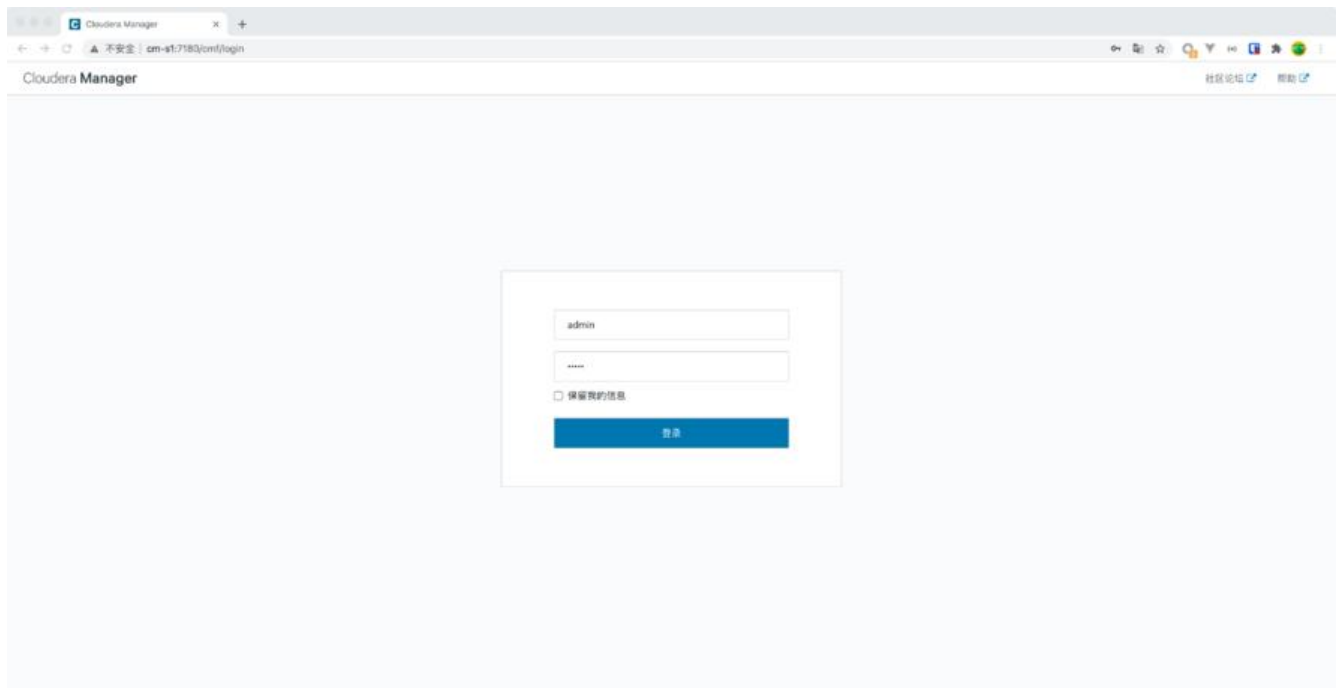
Server启动比较慢，需要等待一会才能访问对应的web界面，可以查看启动日志，Sever启动的日志认在路径*/var/log/cloudera-scm-server/cloudera-scm-server.log下。Agent启动的日志默认在路径var/log/cloudera-scm-agent/cloudera-scm-agent.log*下。可以通过以上日志来检查启动中是否错误。

Sever首次启动会自动创建表以及数据，不要立即关闭或重启，否则需要删除所有表及数据重新安装。至此，Cloudera Manager的安装完成。

CDH安装

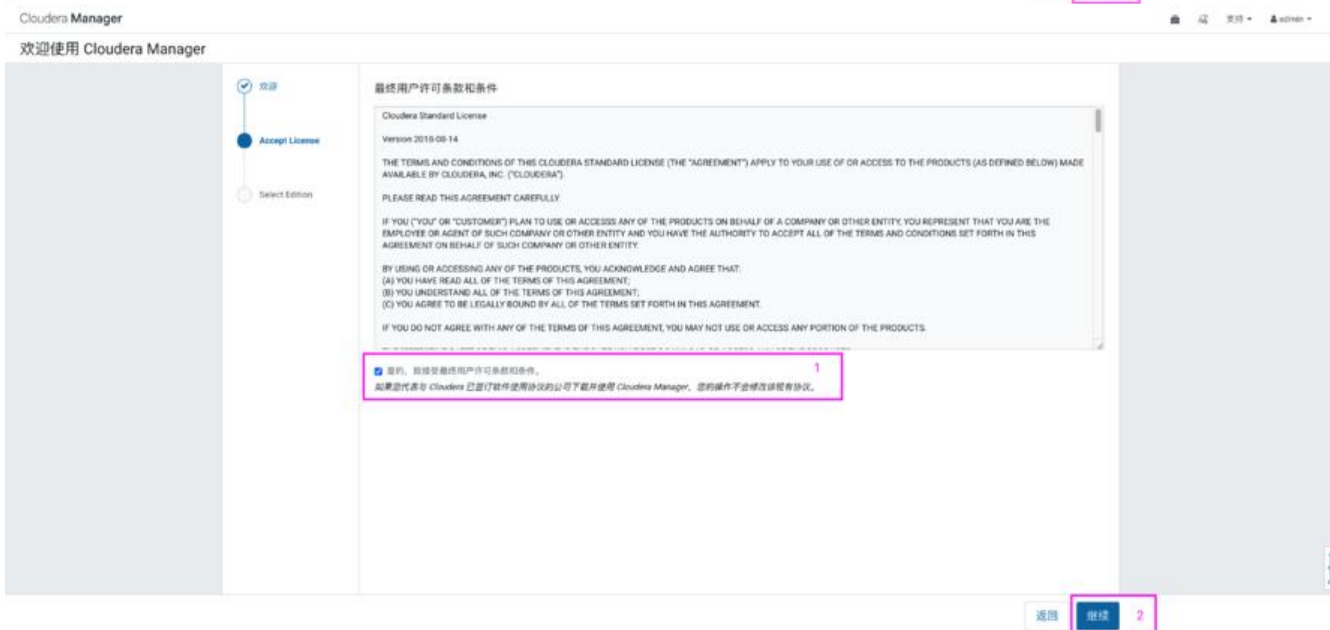
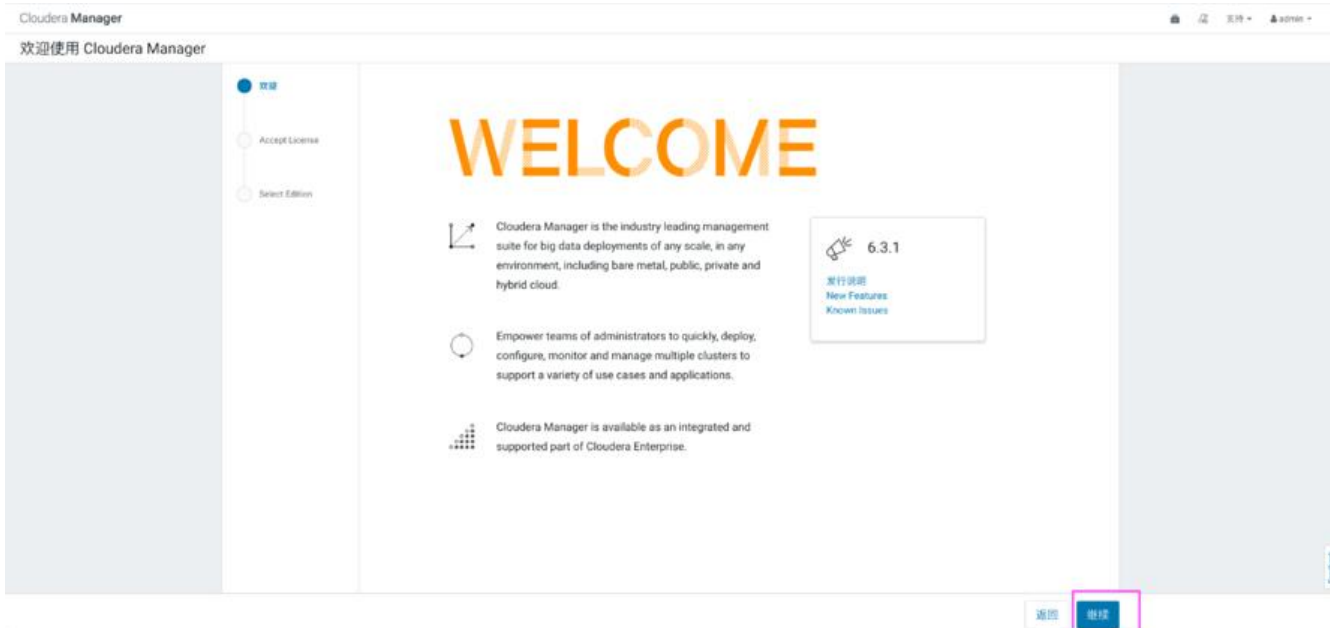
1. 登录 Cloudera Manager WebUI界面

Cloudera Manager的WebUI界面登录地址为 <http://cm-s1:7180>，默认的用户名和密码都是 admin。输入完成后，点击登录：

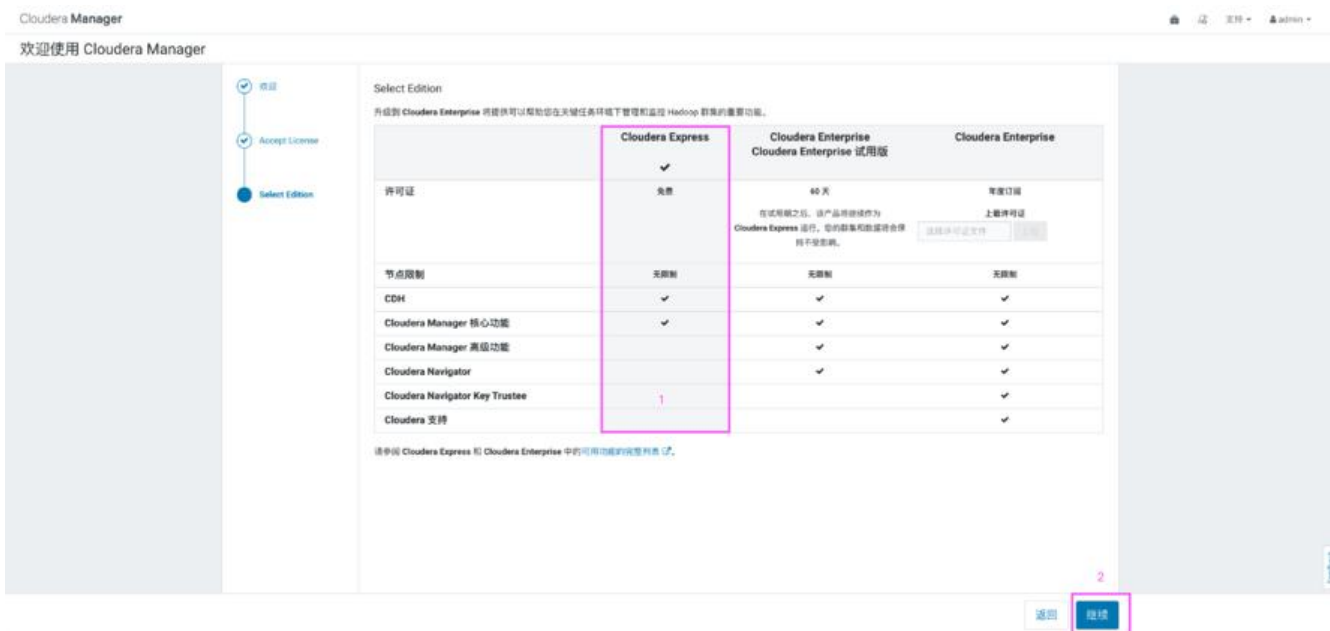


2. 选择安装 CDH 的节点

登录之后，一直选择“继续”即可



选择Cloudera Express，点击“继续”，弹出页面后继续点击“继续”：



群集安装

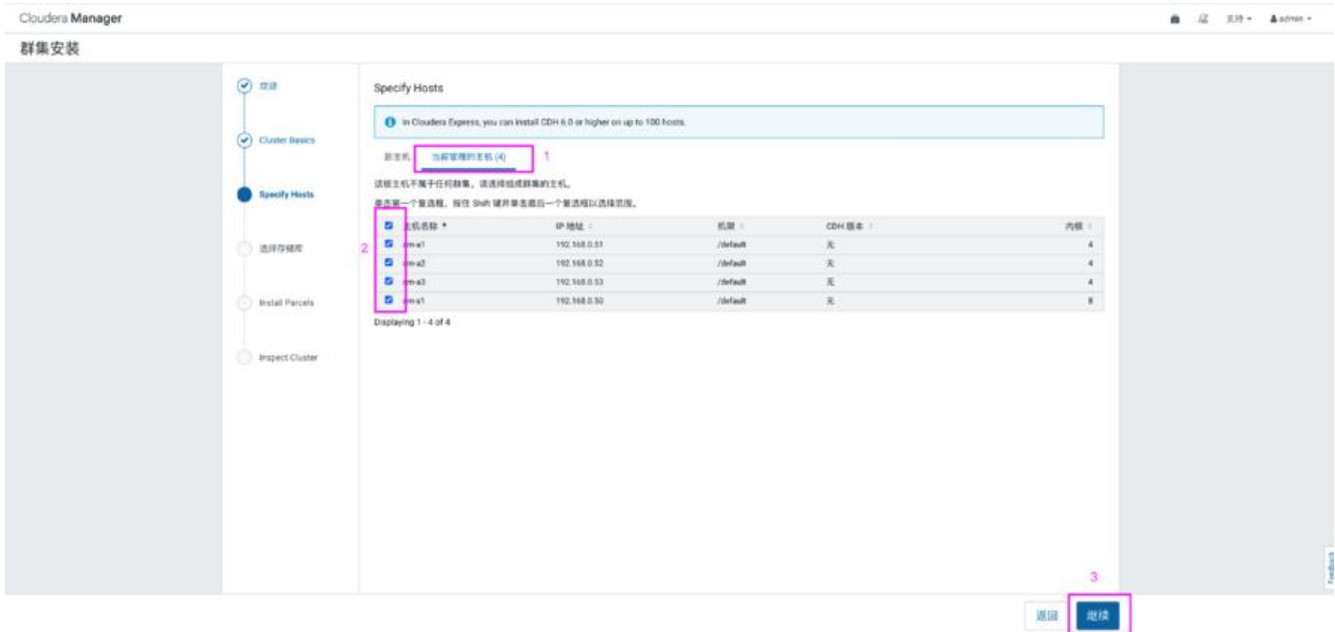
The screenshot shows the 'Welcome' page in Cloudera Manager. On the left is a vertical progress bar with steps: 欢迎 (Welcome), Cluster Basics, Specify Hosts, 选择存储库 (Select Repositories), JDK 安装选项 (JDK Installation Options), 提供 SSH 登录凭据 (Provide SSH Login Credentials), Install Agents, Install Parcels, and Inspect Cluster. The 'Cluster Basics' step is currently active. The main content area features a large 'WELCOME' heading, followed by a two-step process description. Step 1 states that Cloudera Manager is a leading management suite for big data. Step 2 emphasizes empowering administrators to quickly deploy and manage multiple clusters. A '快速链接' (Quick Links) box on the right provides links for the Install Guide, Operating System Requirements, Database Requirements, and JDK Requirements. At the bottom right, there are '返回' (Back) and '继续' (Continue) buttons, with '继续' highlighted by a pink box.

群集安装

The screenshot shows the 'Cluster Basics' configuration page. The progress bar on the left now has 'Cluster Basics' selected. The main content area is titled 'Cluster Basics' and contains a single text input field labeled '群集名称' (Cluster Name) with the value 'mycluster' entered. The '继续' (Continue) button at the bottom right is highlighted with a pink box.

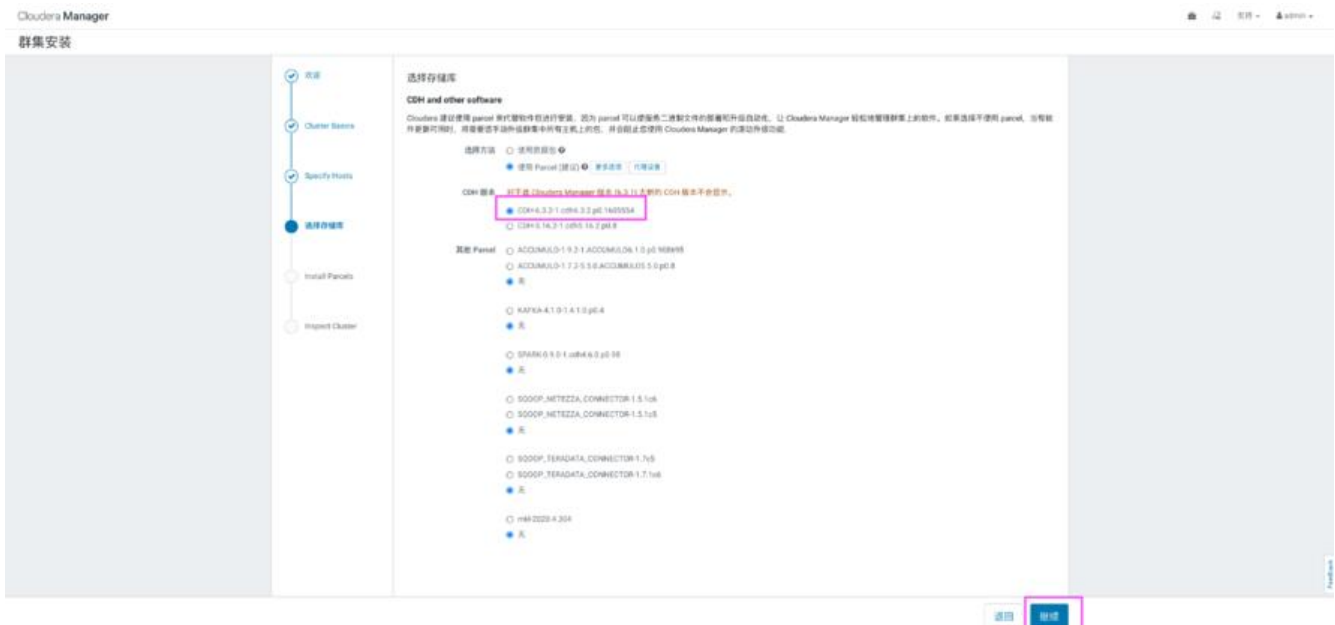
弹出如下页面，该页面是为CDH指定主机，可以使用“模式”通配来选择主机，也可以选择“当前管理的主机”，这里“当前管理的主机”中有节点内容的原因正是由于之前我们在这三台节点配置过agent。如果未来集群增加机器，可以在新主机中搜索添加，后期会自动将agent安装到选中的新节点。

这里我们选择“当前管理的主机”中的所有节点，点击“继续”

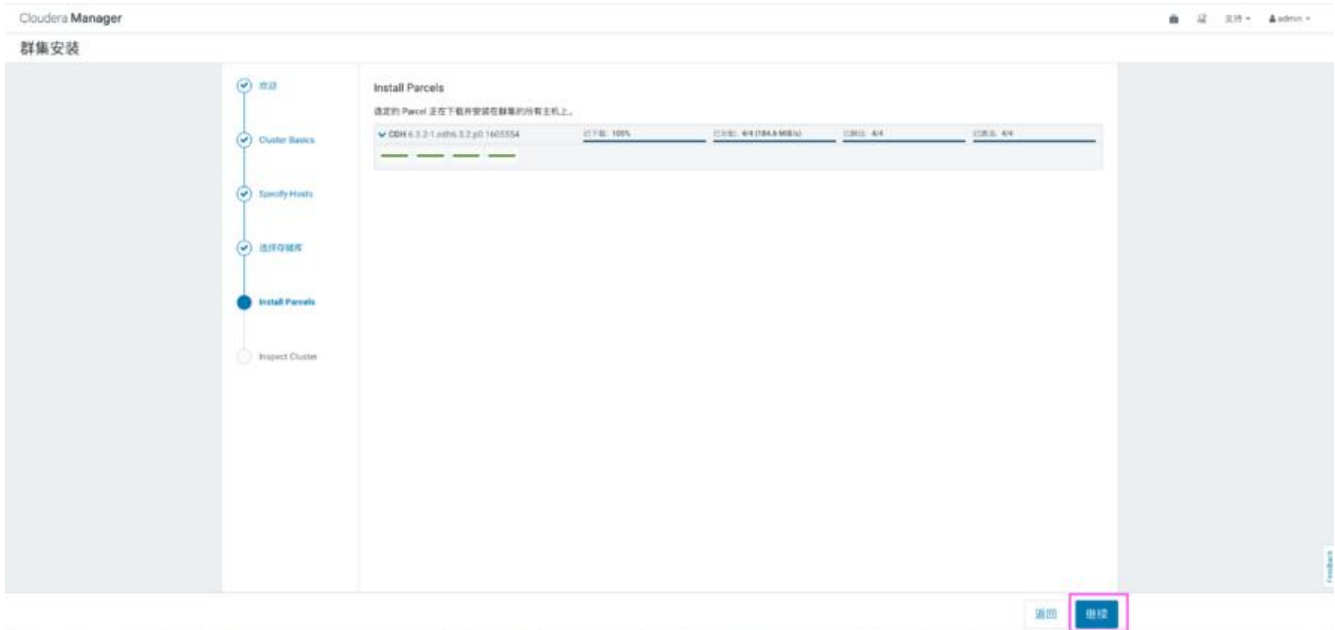


3. 集群安装

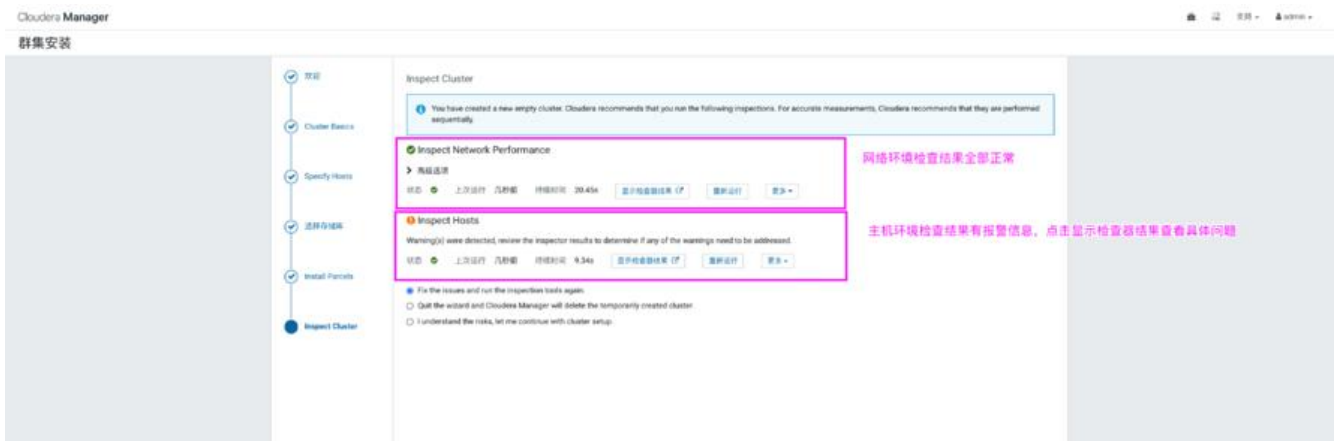
选择 CDH 版本为 CDH 6.3.2，点击“继续”，如下图：



点击“继续”之后，Cloudera Manager主节点会将下载好的CDH分发到各个Agent节点，这时可以在三个Agent节点上看到路径/opt/cloudera/parcel-cache目录中有Cloudera Manager主节点传过来CDH安装包，同时在完成传输之后，每个Agent节点还会将CDH安装包解压到/opt/cloudera/parcel路径下，此时，界面显示如下：



点击“继续”，显示正在检查主机健康状况，稍等片刻，会显示检查结果：



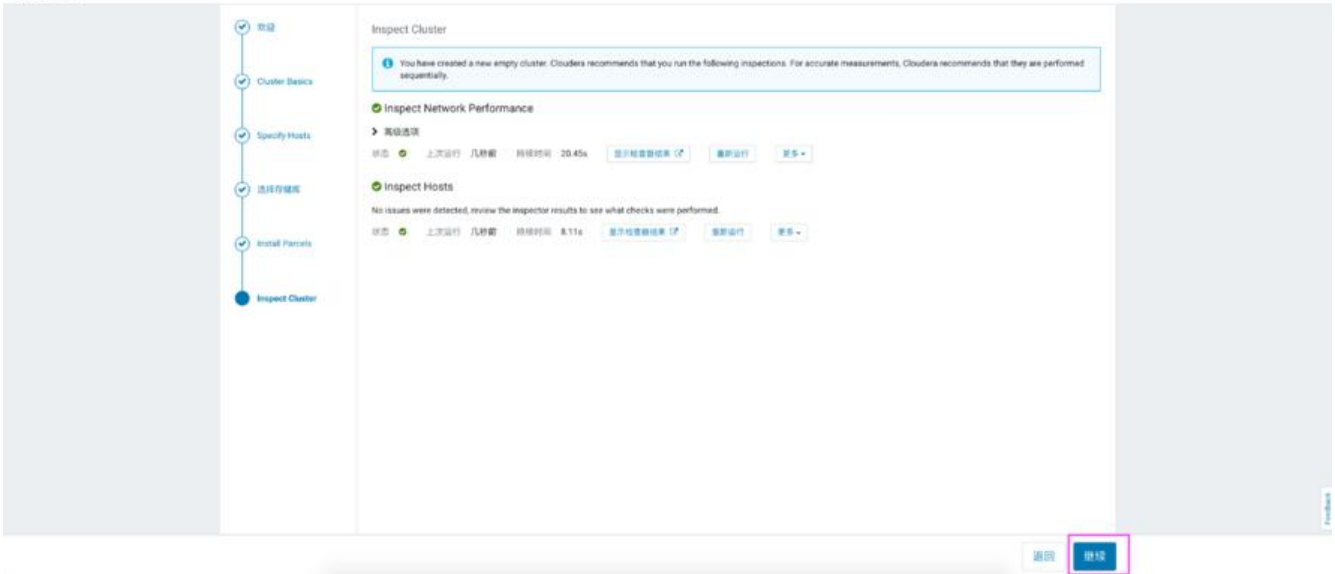
点击检查结果，查看检查主机出现的问题：

- ✔ 没有存在已知错误的内核版本在运行。
- ⚠ Cloudera 建议将 `/proc/sys/vm/swappiness` 设置为最大值 10。当前设置为 30。使用 `sysctl` 命令在运行时更改该设置并编辑 `/etc/sysctl.conf`，以在重启后保存该设置。您可以继续进行安装，但 Cloudera Manager 可能会报告您的主机由于交换而运行状况不良。以下主机将受到影响：
 - ▼ 查看详细信息
 - cm.a[1-3]
- ⚠ 已启用透明大页面压缩，可能会导致重大性能问题。请运行 `echo never > /sys/kernel/mm/transparent_hugepage/defrag` 和 `echo never > /sys/kernel/mm/transparent_hugepage/enabled` 以禁用此设置，然后将同一命令添加到 `/etc/rc.local` 等初始化脚本中，以便在系统重启时予以设置。以下主机将受到影响：
 - ▼ 查看详细信息
 - cm.a[1-3]
- ✔ 已满足 CDH 5 Hue Python 版本依赖关系。

这里建议 `swappiness=10` 的时候表示最大限度使用物理内存。由于透明超大页面已知会导致意外的点重新启动，可以在每台节点中执行如下命令，在检查主机正确性后面点击“重新运行”即可解决。令如下：

```
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i sed -i '\$a/vm.swappiness\ =\ 10 /etc/sysctl.conf;done
[root@cm-s1 ~]# for i in `seq 0 3`;do ssh root@192.168.0.5$i sysctl -p;done
# 在每个节点执行如下命令，并写入到/etc/rc.local中，注意软链接源文件/etc/rc.d/rc.local文件的执行性
echo never > /sys/kernel/mm/transparent_hugepage/defrag
echo never > /sys/kernel/mm/transparent_hugepage/enabled
```

群集安装



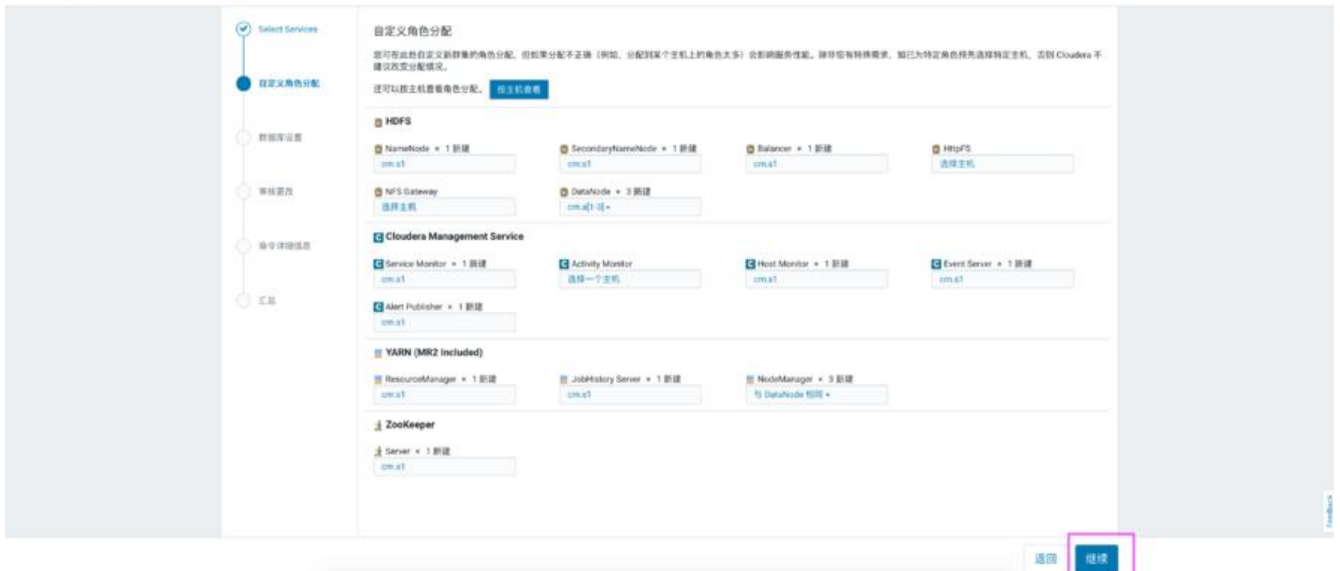
最后，点击“继续”，进入“集群设置”

4. 集群设置

在进入到集群设置界面中，点击“自定义服务”，选择“HDFS”、“YARN”、“Zookeeper”进安装，注意：这里提示还会安装Cloudera Management Service，用于后期的监控、报告、事件和报。

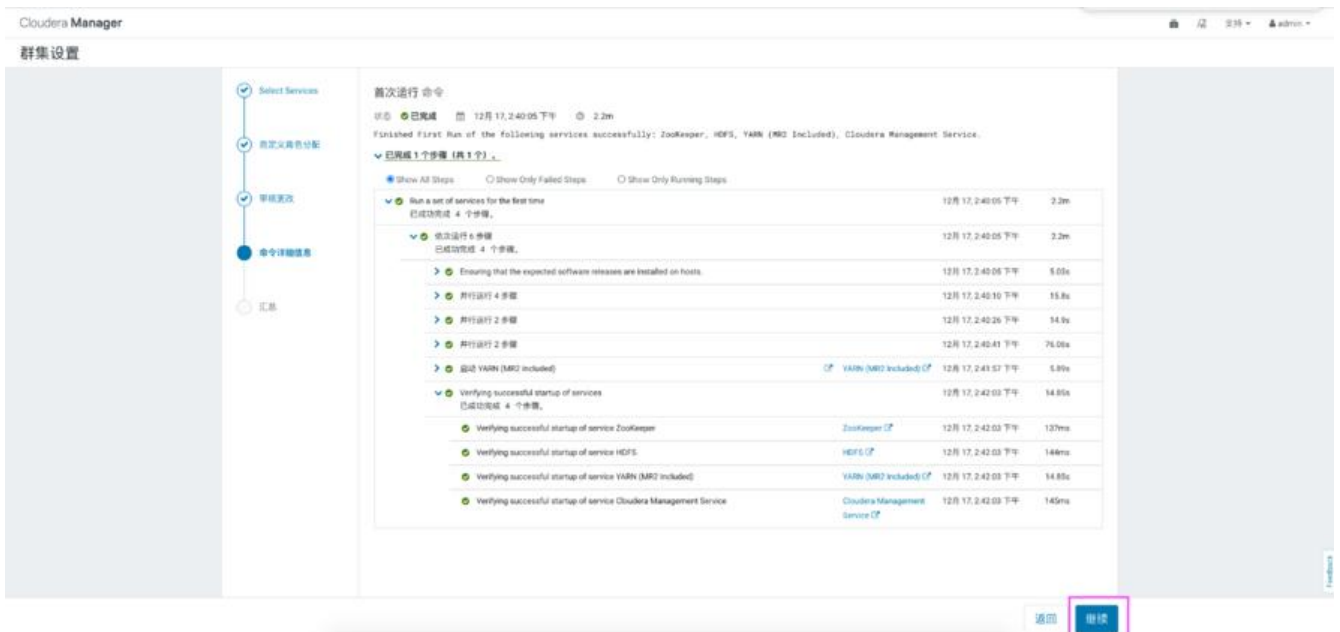


以上点击“继续”，进入“角色分配”页面，显示的是Cloudera Manager自动划分的集群分配，当这里也可以自己分配节点。如下图所示：



继续点击“继续”，进入“审核更改页面”，在这里可以手动修改一些配置文件，取默认配置。

继续点击“继续”，进入执行集群命令，启动各个组件页面，当所有组件启动完成之后，如下图所示：



点击“继续”，完成集群设置。

群集设置

- Select Services
- 自定义角色分配
- 审核更改
- 命令详细视图
- 汇总**

汇总

服务已安装，配置并在群集中运行。

返回 完成

5. CDH主页面

Cloudera Manager 群集 - 主页 - 群集 - 审核 - 图表 - 管理 -

主页 状态 所有运行状况问题 配置 所有最新命令

mycluster

CDH 5.3.2 (Parcels)

- 4 主机
- HDP
- YARN (AM2 L)
- ZooKeeper

Cloudera Management Service

- Cloudera Ma...

图表

- 群集 CPU: mycluster 群集中主机的平均 CPU 使用率 1.3%
- 群集数据 IO: 群集中主机 8440N 的总数据 IO 量 0
- 群集网络 IO: 群集中主机 17.86% 的总网络 IO 量 5.16%
- HDFS IO: 群集中主机 2.86% 的总 HDFS IO 量 18%