

# CDH 安装手册

## 一、前言

### 1.1. 什么是 CDH

Apache Hadoop 作为目前最主流、应用范围最广的分布式应用架构,根据 Google 公司发表的 MapReduce 和 Google 档案系统的论文自行实作而成,提供在 x86 服务器上构建大型应用集群的能力,它采用 Apache 2.0 许可协议发布开源协议,官方版本也称为社区版 Hadoop。

因为 Hadoop 采用 Apache 开源协议,用户可以免费地任意使用和修改 Hadoop,市面上就出现了很多 Hadoop 版本。其中有很多厂家在 Apache Hadoop 的基础上开发自己的 Hadoop 产品,比如 Cloudera 的 CDH (Cloudera's Distribution Including Apache Hadoop), Hortonworks 的 HDP, MapR 的 MapR 产品等。

### 1.2. 为什么使用 CDH

社区版本的 Hadoop 具备很多的优点,例如:完全的开源免费,活跃的社区,文档资料齐全。但由于 Hadoop 的生态圈过于复杂,包括 Hive、Hbase、Sqoop、Flume、Spark、Hue、Oozie 等,需要考虑版本和组件的兼容性;同时集群部署、安装、配置较为复杂,需要手工调整配置文件后,对每台服务器的分发配置分发操作,较为容易出错;同时缺少配套的运行监控和运维工具,需要结合 ganglia、nagios 等实现运行监控,运维成本较高。

而 Cloudera 的 CDH 版本为目前最成型的发行版本,拥有最多的部署案例。通过 CDH 提供更为稳定商用的 Hadoop 版本;提供强大的部

署、管理和监控工具，通过统一的可视化管理后台，实现集群的动态监控，大大提高了集群部署的效率；同时 CDH Express 版本完全免费，不涉及昂贵的商业授权费用。



1.3. 环境说明

机器名	IP 地址	操作系统
BigDataServer1	10.68.128.215	CentOS 7
BigDataServer2	10.68.128.216	CentOS 7
BigDataServer3	10.68.128.217	CentOS 7

二、环境准备

2.1. 关闭防火墙

```
停止 firewall
# systemctl stop firewalld.service
禁止 firewall 开机启动
```

```
# systemctl disable firewalld.service
```

## 2.2. 修改 hosts 文件

增加机器、IP 映射

```
# vi /etc/hosts
10.68.128.215 BigDataServer1
10.68.128.216 BigDataServer2
10.68.128.217 BigDataServer3
:wq
```

## 2.3. 设置 SSH 免登录

在 BigDataServer1 上执行：

```
# ssh-keygen -t rsa
# ssh-copy-id -i BigDataServer2
# ssh-copy-id -i BigDataServer3
```

## 2.4. 设置时间同步

CentOS 7 默认安装的时间同步软件为 Chrony，不再是 NTP，应该需要调整 Chrony 的配置。如需要启动 NTP，需要先删除 Chrony，再安装 NTP 软件

在 BigDataServer1 上增加外网时间服务器地址

```
# vi /etc/chrony.conf
server 10.100.2.5 iburst
:wq
# systemctl start chronyd.service
# systemctl enable chronyd.service
```

在 BigDataServer2、BigDataServer3 上分别设置

```
# vi /etc/chrony.conf
server BigDataServer1 iburst
:wq
# systemctl start chronyd.service
# systemctl enable chronyd.service
```

参考：<http://linux.cn/article-4764-1.html>

## 2.5. 系统优化

禁用交换分区

```
# sysctl -w vm.swappiness=0
# echo 0 > /proc/sys/vm/swappiness
```

禁用透明大页面

```
# echo never > /sys/kernel/mm/transparent_hugepage/defrag
# echo "echo never > /sys/kernel/mm/transparent_hugepage/defrag" >> /etc/rc.local
```

## 2.6. JDK 安装

从 [http://java.com/zh\\_CN/download/](http://java.com/zh_CN/download/) 的网站上下载 JDK8 的 Linux 安装程序，并上传到服务器（三台服务器均需要进行安装）。

因在后续的安装中，需要安装 Presto 进行学习，Presto 要求采用 JDK8 以上版本，因此在本安装示例采用的是 JDK8u65 的版本，jdk-8u65-linux-x64.tar.gz。

### 2.6.1. 删除自带 OpenJDK

```
# yum list installed | grep jdk
# yum -y remove java-1.6.0-openjdk.x86_64
# yum -y remove java-1.7.0-openjdk.x86_64
```

### 2.6.2. 解压安装包

解压到/opt 安装目录下

```
# tar -zxvf jdk-8u65-linux-x64.tar.gz -C /opt
# ln -s /opt/jdk1.8.0_65 /opt/jdk
# mkdir /usr/java
# ln -s /opt/jdk /usr/java/default
```

*备注：其中/usr/java/default 软链接的创建尤为重要，否则 Spark 将无法安装完成。*

### 2.6.3. 修改环境变量

修改/etc/profile 文件

```
# vi /etc/profile
```

```
export JAVA_HOME=/opt/jdk
export PATH="$JAVA_HOME/bin:$PATH"
:wq
```

启用配置

```
# source /etc/profile
```

## 2.6.4. 验证安装

```
# java -version
```

## 2.7. MySQL 安装

CentOS7 默认安装 MariaDB（MySQL 的开源分支），为确保数据库的稳定性，依据采用 MySQL 官方社区版本。

### 2.7.1 删除 MariaDB

```
# rpm -qa | grep mariadb
# rpm -e mariadb-libs-5.5.44-2.el7.centos.x86_64
```

### 2.7.2. 安装 MySQL

从 MySQL 官网（<http://dev.mysql.com/downloads/mysql/>）上下载 Red Hat Enterprise Linux 7 / Oracle Linux 7 (x86, 64-bit), RPM Bundle 版本的安装包，具体下载文件为 mysql-5.7.11-1.el7.x86\_64.rpm-bundle。

将下载的文件，上传到 BigDataServer1，并进行解压。

```
# tar -xvf mysql-5.7.11-1.el7.x86_64.rpm-bundle.tar
```

安装 MySQL 包

```
# rpm -ivh mysql-community-common-5.7.11-1.el7.x86_64.rpm
# rpm -ivh mysql-community-libs-5.7.11-1.el7.x86_64.rpm
# rpm -ivh mysql-community-client-5.7.11-1.el7.x86_64.rpm
# rpm -ivh mysql-community-server-5.7.11-1.el7.x86_64.rpm
```

启动服务

```
# systemctl start mysqld.service
# systemctl enable mysqld.service
```

*备注：本示例仅仅讲解最基础的 MySQL 安装，未涉及配置调优过程，同时为确保 MySQL 的高可用性，应配置 MySQL 的主主双活，具体参见《MySQL 主备双活安装说明》中的配置说明。*

### 2.7.3. 设置 root 管理员密码

CentOS 在安装 MySQL 过程中，会自动生成临时密码，查看临时密码的方式为：

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

```
[root@BigDataServer1 ~]# grep 'temporary password' /var/log/mysqld.log
2016-02-18T07:12:06.752154Z 1 [Note] A temporary password is generated for root@localhost: D/BjdIa<u657
[root@BigDataServer1 ~]#
```

红色圈选部分则为随机生成的密码。

修改 MySQL 的 root 密码

```
# mysql -uroot -p
```

输入随机生成的密码，然后修改密码

```
mysql> SET PASSWORD FOR 'root'@'localhost' = PASSWORD('Whcyit123!@#');
```

备注：因 CentOS 安装 MySQL 后，启用了密码增强插件，新密码必须包括大写字母、小写字母、数字、特殊字符，否则无法进行密码修改。

### 2.7.4. 创建数据库

为 Hive、Oozie 创建数据库

```
mysql> CREATE DATABASE hive DEFAULT CHARSET utf8 COLLATE utf8_general_ci;
mysql> CREATE DATABASE oozie DEFAULT CHARSET utf8 COLLATE utf8_general_ci;
```

授权给 chd 用户

```
mysql> grant all privileges on *.* to 'cdh'@'%' identified by 'Whcyit123!@#' with
grant option;
mysql> flush privileges;
```

## 2.8. 创建 SCM 用户

所有服务器均需要执行

```
# useradd --system --home=/opt/cm-5.5.1/run/cloudera-scm-server/
--no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm
```

## 2.9. 需要 CDH 目录

在所有的服务器上创建以下目录（首次安装时，不需要手动创建，程序会自动进行创建，如安装出现问题，需要重新安装时，需要执行下面命令）：

```
# rm-rf /dfs
# rm-rf /var/lib/zookeeper
# rm-rf /opt/cloudera-manager/cm-5.5.0/lib/cloudera-scm-agent/*
# rm-rf /var/lib/cloudera-scm-headlamp
# rm-rf /var/lib/cloudera-scm-firehose
# rm-rf /var/lib/cloudera-scm-alertpublisher
# rm-rf /var/lib/cloudera-scm-eventserver
# rm-rf /var/lib/cloudera-scm-server

# rm-rf /var/log/cloudera-scm-headlamp
# rm-rf /var/log/cloudera-scm-firehose
# rm-rf /var/log/cloudera-scm-alertpublisher
# rm-rf /var/log/cloudera-scm-eventserver
# rm-rf /var/log/cloudera-scm-server
# rm-rf /hbase

# mkdir -p /var/lib/cloudera-scm-headlamp
# mkdir -p /var/lib/cloudera-scm-firehose
# mkdir -p /var/lib/cloudera-scm-alertpublisher
# mkdir -p /var/lib/cloudera-scm-eventserver
# mkdir -p /var/lib/cloudera-scm-server

# mkdir -p /var/log/cloudera-scm-headlamp
# mkdir -p /var/log/cloudera-scm-firehose
# mkdir -p /var/log/cloudera-scm-alertpublisher
# mkdir -p /var/log/cloudera-scm-eventserver
# mkdir -p /var/log/cloudera-scm-server
# mkdir -p /opt/cloudera-manager
# mkdir -p /opt/cloudera/parcel-repo
# mkdir /hbase

# chown cloudera-scm:cloudera-scm /var/lib/cloudera-scm-headlamp
# chown cloudera-scm:cloudera-scm /var/lib/cloudera-scm-firehose
# chown cloudera-scm:cloudera-scm /var/lib/cloudera-scm-alertpublisher
# chown cloudera-scm:cloudera-scm /var/lib/cloudera-scm-eventserver
# chown cloudera-scm:cloudera-scm /var/lib/cloudera-scm-server
```

```
# chown cloudera-scm:cloudera-scm /var/log/cloudera-scm-headlamp
# chown cloudera-scm:cloudera-scm /var/log/cloudera-scm-firehose
# chown cloudera-scm:cloudera-scm /var/log/cloudera-scm-alertpublisher
# chown cloudera-scm:cloudera-scm /var/log/cloudera-scm-eventserver
# chown cloudera-scm:cloudera-scm /var/log/cloudera-scm-server
# chown -R cloudera-scm:cloudera-scm /opt/cloudera-manager
# chown cloudera-scm:cloudera-scm /hbase/

# chown -R root:root /opt/cloudera/parcel-repo
```

同时在 **BigDataServer1** 上，执行以下命令：

```
# rm-rf /user/hive/warehouse
# mkdir -p /user/hive/warehouse
# chown cloudera-scm:cloudera-scm /user/hive/warehouse

# rm-rf /var/lib/cloudera-host-monitor
# mkdir -p /var/lib/cloudera-host-monitor
# chown cloudera-scm:cloudera-scm /var/lib/cloudera-host-monitor

# rm-rf /var/lib/cloudera-service-monitor
# mkdir -p /var/lib/cloudera-service-monitor
# chown cloudera-scm:cloudera-scm /var/lib/cloudera-service-monitor
```

## 2.10. 缺失依赖包安装

在实际安装过程中，CDH 对部分包存在依赖（依据环境差异，具体安装过程中可能存在一定的区别，注意分析查看日志文件）。在本次安装过程中，发现 CentOS 由于 `psmisc`、`libxslt`、`libxslt-python` 三个包未安装，在程序安装过程中出现报错。

如服务器可以连接互联网，直接采用 `yum` 安装即可。

```
# yum install -y psmisc libxslt libxslt-python
```

由于本次安装过程中，服务器不能访问外网，需要手动下载对应 RPM 包，上传到服务器后进行安装。RPM 的下载地址为：

[http://mirror.centos.org/centos/7.2.1511/os/x86\\_64/Packages/](http://mirror.centos.org/centos/7.2.1511/os/x86_64/Packages/)（如果安

装的操作系统为 CentOS 6.5 的版本，RPM 包的下载地址为



[http://vault.centos.org/6.5/os/x86\\_64/Packages/](http://vault.centos.org/6.5/os/x86_64/Packages/)，对于其他版本的系统，请自行百度对应 RPM 包下载地址）。

依次下载 psmisc-22.20-9.el7.x86\_64.rpm、libxslt-1.1.28-5.el7.x86\_64.rpm、libxslt-python-1.1.28-5.el7.x86\_64.rpm 等三个文件，并进行安装。

```
# rpm -ivh psmisc-22.20-9.el7.x86_64.rpm
# rpm -ivh libxslt-1.1.28-5.el7.x86_64.rpm
# rpm -ivh libxslt-python-1.1.28-5.el7.x86_64.rpm
```

## 三、CDH 安装

### 3.1. 下载安装介质

CDH 常用安装方式一般包括：Cloudera Manager 在线安装、Parcel 安装、YUM 安装以及 RPM 安装。因本次服务器无法连接外网，同时官方文档推荐使用 Parcel 模式进行安装，故本次安装采用 Parcel 方式进行。

CDH 安装需要下载以下介质（**下载过程中，需要注意依据操作系统及版本，选择正确的安装包**）：

#### ■ Cloudera Manager

下载地址为 <http://archive.cloudera.com/cm5/cm/5/>，本示例下载的为 cloudera-manager-centos7-cm5.5.3\_x86\_64.tar.gz。

#### ■ CDH Parcel 包

下载地址为 <http://archive.cloudera.com/cdh5/parcels/>，本示例下载的为三个文件：CDH-5.5.2-1.cdh5.5.2.p0.4-el7.parcel、CDH-5.5.2-1.cdh5.5.2.p0.4-el7.parcel.sha1、manifest.json。

#### ■ Kafka CSD 包

下载地址为 <http://archive.cloudera.com/csds/kafka/>，本示例下载的为 KAFKA-1.2.0.jar。

### ■ Kafka Parcel 包

下载地址为 <http://archive.cloudera.com/kafka/parcels/latest/>，本示例下载的为 KAFKA-2.0.0-1.kafka2.0.0.p0.12-el7.parcel 、KAFKA-2.0.0-1.kafka2.0.0.p0.12-el7.parcel.sha1。

### ■ MySQL JDBC 驱动

从 MySQL 官方网站上下载最新的 MySQL JDBC 驱动程序，下载地址为 <http://dev.mysql.com/downloads/connector/j/>。本次示例下载的文件为 mysql-connector-java-5.1.36.jar。

## 3.2. 安装 Cloudera Manager

### 3.2.1. 解压 CM 安装程序

将 cloudera-manager-centos7-cm5.5.3\_x86\_64.tar.gz 上传到主服务器（BigDataServer1）上，并进行解压操作。

```
# tar -xvf cloudera-manager-centos7-cm5.5.3_x86_64.tar.gz -C /opt
```

### 3.2.2. 初始化 CM 数据库

同时将 mysql-connector-java-5.1.36.jar 上传到服务器中，并拷贝到 cm/share/cmfd/lib/ 目录中。

```
# cp mysql-connector-java-5.1.36.jar /opt/cm-5.5.3/share/cmfd/lib/
```

初始化 Cloudera Manager 数据库

```
# /opt/cm-5.5.3/share/cmfd/schema/scm_prepare_database.sh mysql cm  
-hlocalhost -ucdh -p'Whcyit123!@#' scm 'Scm123!@#'
```

备注：创建的数据库名为 cm，对应的用户名、密码为 scm 和 Scm123!@#。

### 3.2.3. 修改配置

```
# vi /opt/cm-5.5.3/etc/cloudera-scm-agent/config.ini
server_host=10.68.128.215
```

### 3.2.4. 同步 CM 到其他节点

```
# scp -r /opt/cm-5.5.3 BigDataServer2:/opt/
# scp -r /opt/cm-5.5.3 BigDataServer3:/opt/
```

### 3.2.5. 上传 Parcel 到服务器

将 manifest.json、CDH-5.3.9-1.cdh5.3.9.p0.8-trusty.parcel、CDH-5.3.9-1.cdh5.3.9.p0.8-trusty.parcel.sha1 三个文件上传到 BigDataServer1 的/opt/cloudera/parcel-repo/目录中，同时将 CDH-5.3.9-1.cdh5.3.9.p0.8-trusty.parcel.sha1 复制为 CDH-5.3.9-1.cdh5.3.9.p0.8-trusty.parcel.sha。

```
# cp CDH-5.3.9-1.cdh5.3.9.p0.8-trusty.parcel.sha1
CDH-5.3.9-1.cdh5.3.9.p0.8-trusty.parcel.sha
```

### 3.2.6. 启动 CM 服务

在 BigDataServer1 上启动 CM 服务和代理。

```
# /opt/cm-5.5.3/etc/init.d/cloudera-scm-server start
# /opt/cm-5.5.3/etc/init.d/cloudera-scm-agent start
```

在 BigDataServer2、BigDataServer3 启动 CM 代理。

```
# /opt/cm-5.5.3/etc/init.d/cloudera-scm-agent start
```

*备注: 如果在将 CM 文件夹同步到其他节点之前, 启动过代理程序, 需要删除所有服务器上的/opt/cm-5.5.3/lib/cloudera-scm-agent 中生成 response.avro 和 uuid 两个文件, 并重启代理程序, 否则服务程序将无法正确检查到 agent。*

## 3.3. 安装 Hadoop

### 3.3.1. 登录 CM 管理控制台

使用浏览器打开 <http://10.68.128.215:7180> 登录 Cloudera Manager

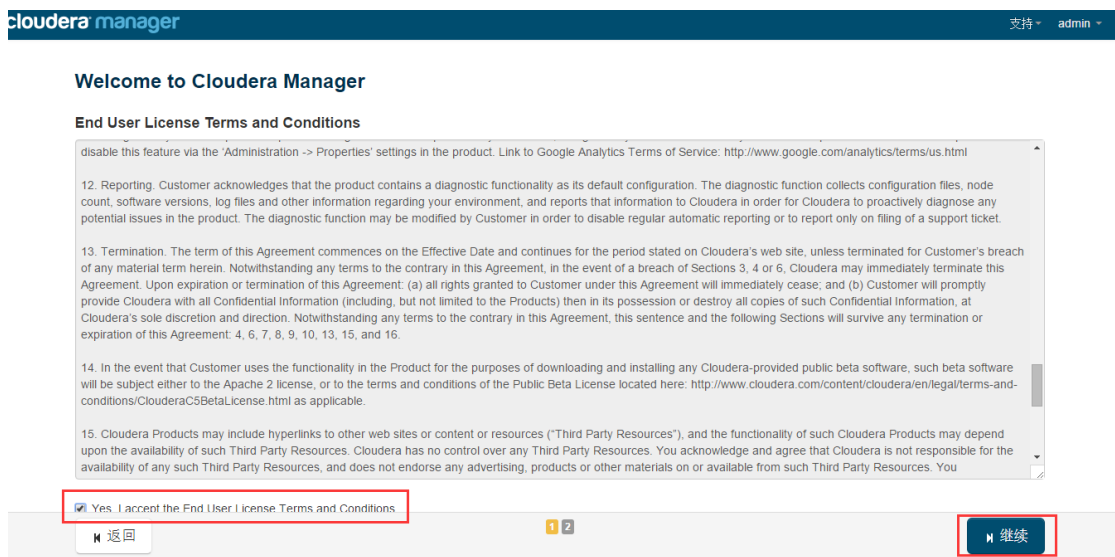
的 Web 管理控制台。（默认的用户名和密码为 admin/admin）



The image shows the Cloudera Manager login interface. At the top, there is a dark blue header with the 'cloudera manager' logo on the left and '社区论坛' (Community Forum) and '帮助' (Help) links on the right. The main content area is white and contains a login form titled '登录' (Login). The form has two input fields: '用户名:' (Username) with the value 'admin' and '密码:' (Password) with masked characters '\*\*\*\*\*'. Below these fields is a checkbox labeled '在此计算机上保留我的信息。' (Remember me on this computer). At the bottom of the form is a large blue button labeled '登录' (Login). Red rectangular boxes highlight the username field, the password field, the 'Remember me' checkbox, and the login button.

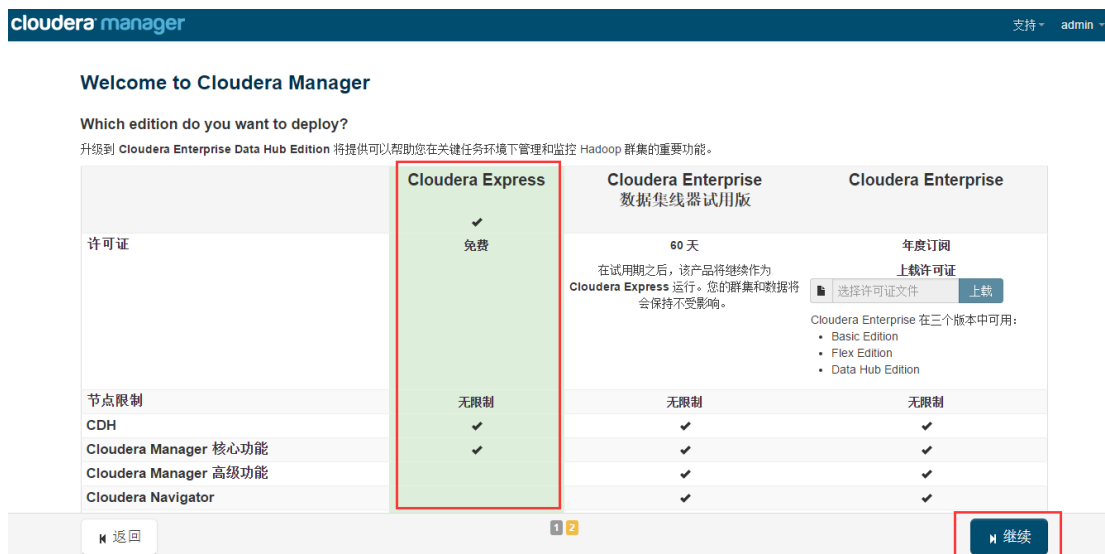
### 3.3.2. 使用向导进行安装

第一次进入时，会自动启动安装向导进行配置安装。



The image shows the 'Welcome to Cloudera Manager' screen. At the top, there is a dark blue header with the 'cloudera manager' logo on the left and '支持' (Support) and 'admin' links on the right. The main content area is white and contains the text 'Welcome to Cloudera Manager' and 'End User License Terms and Conditions'. Below this text is a scrollable area containing the license terms. At the bottom of the screen, there is a checkbox labeled 'Yes, I accept the End User License Terms and Conditions.' which is checked. To the left of this checkbox is a button labeled '返回' (Back). To the right of the checkbox is a button labeled '继续' (Continue). Red rectangular boxes highlight the 'Yes, I accept the End User License Terms and Conditions.' checkbox and the '继续' (Continue) button.

选择接受用户授权协议，点击“继续”按钮进入下一步。



选择要安装的 CDH 版本，安装选择 Cloudera Express，点击“继续”，进入下一步。



直接点击“继续”。

为 CDH 群集安装指定主机。

新主机 当前管理的主机 (3)

这些主机不属于任何群集。请选择组成群集的主机。

<input checked="" type="checkbox"/>	名称	IP	机架	CDH 版本	状态	上一检测信号
	<input type="text" value="任何名称"/>	<input type="text" value="任何 IP"/>	<input type="text" value="任何机架"/>	<input type="text" value="全部"/>	<input type="text" value="全部"/>	<input type="text" value="全部"/>
<input checked="" type="checkbox"/>	BigDataServer1	10.68.128.215	/default	无	未知运行状况	7.18s ago
<input checked="" type="checkbox"/>	BigDataServer2	10.68.128.216	/default	无	未知运行状况	1.83s ago
<input checked="" type="checkbox"/>	BigDataServer3	10.68.128.217	/default	无	未知运行状况	4.6s ago

返回

继续

切换到“当前管理的主机”标签，勾选所有的服务器，点击“继续”。

群集安装

选择存储库

Cloudera 建议使用 parcel 来代替软件包进行安装，因为 parcel 可以使服务二进制文件的部署和升级自动化，让 Cloudera Manager 轻松地管理群集上的软件。如果选择不使用 parcel，当有软件更新可用时，将需要您手动升级群集中所有主机上的包，并会阻止您使用 Cloudera Manager 的滚动升级功能。

选择方法
 ☐ 使用数据包
 ☒ 使用 Parcel (建议)
 [更多选项](#)

选择 CDH 的版本
 

☒ CDH-5.5.2-1.cdh5.5.2.p0.4
 

对于此 Cloudera Manager 版本 (5.5.3) 太新的 CDH 版本不会显示。

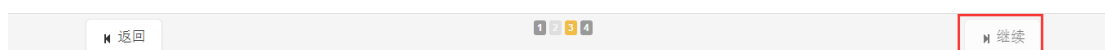
其他 Parcel
 ☐ KAFKA-2.0.0-1.kafka2.0.0.p0.12
 ☒ 无

返回

1 2 3 4

继续

因当前只有一个 Parcel，直接点击“继续”进入下一步。



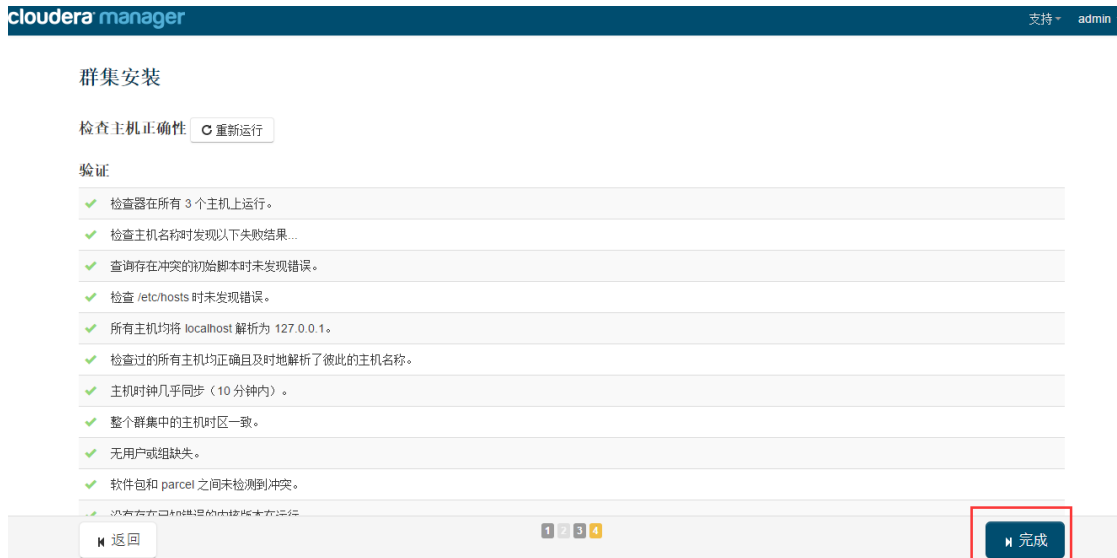
本步骤会将 Parcel 分发到三台服务器，并进行解压、激活操作，该步骤会花费一定的时间，等待操作完成后，点击“继续”即可。

*备注: 在激活完成后, HIVE 以及 Oozie 安装需要使用到 MySQL JDBC 驱动, 需要在所有节点上执行以下操作。(本步骤也可以在系统安装出错时进行执行)*

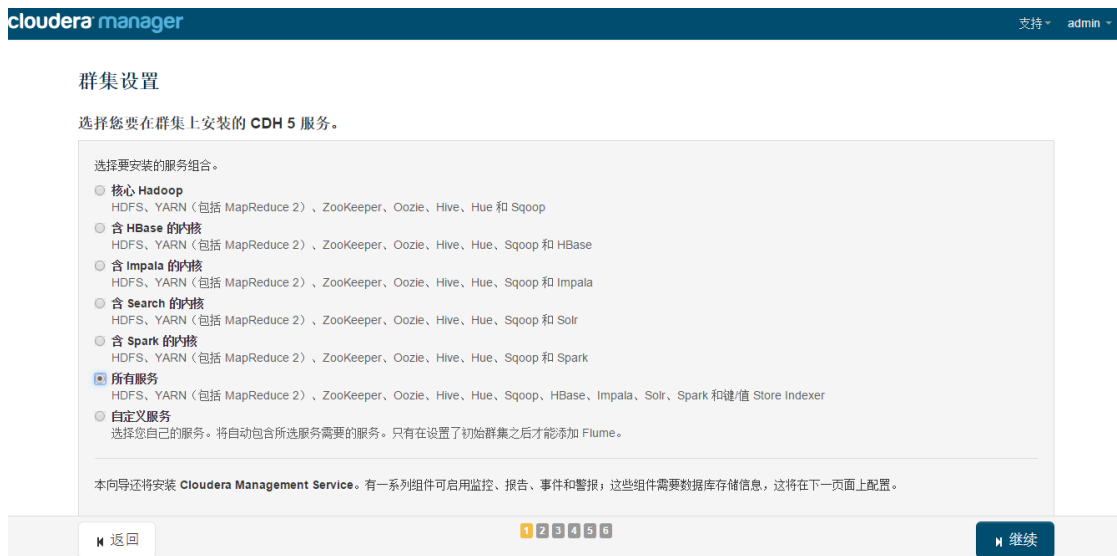
```
#cp mysql-connector-java-5.1.36.jar /opt/cloudera/parcels/CDH/lib/hive/lib/  
#cp mysql-connector-java-5.1.36.jar /opt/cloudera/parcels/CDH/lib/hadoop  
#cp mysql-connector-java-5.1.36.jar /var/lib/oozie
```



本步骤对所有服务器的环境进行检查，等待一分钟左右，会自动显示检查结果。

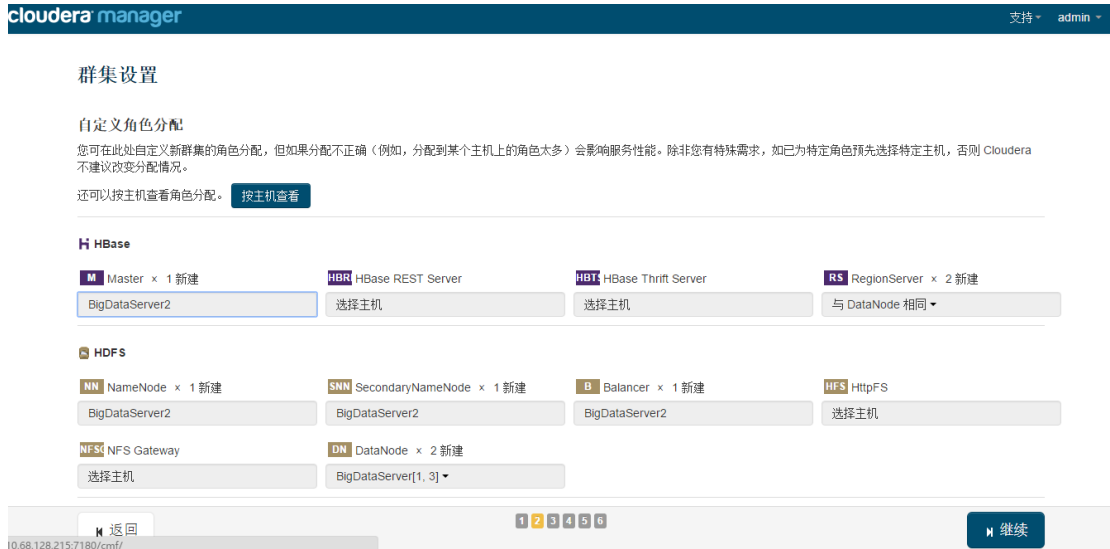


如安装过程依据环境准备完成所有的准备工作，将无提示信息，如果有警告提示内容，安装提示进行修正即可，完成后进行重新检测，全部修正后，点击“完成”进入正式安装过程。

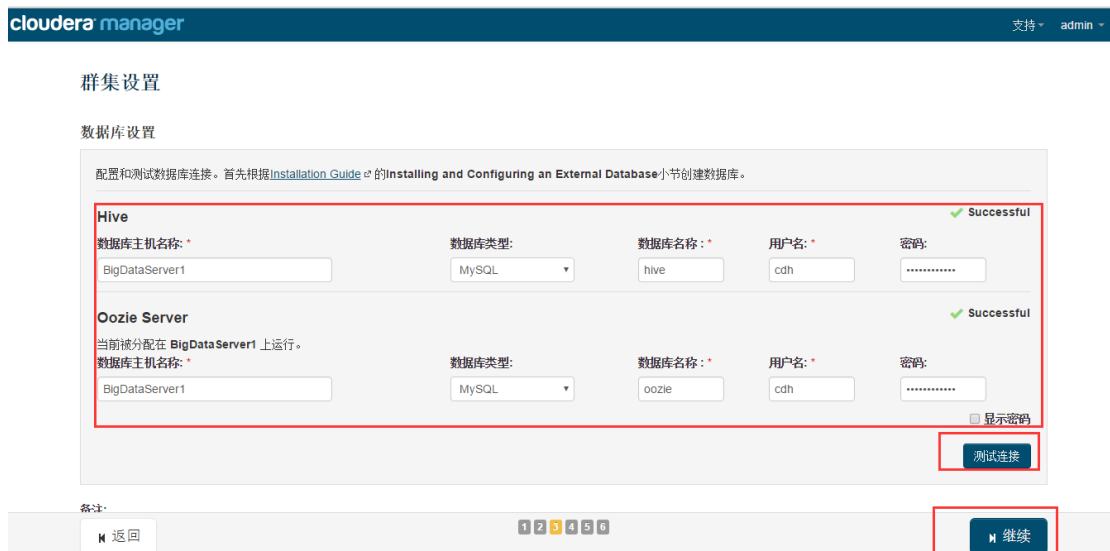


本步骤选择需要安装 CDH5 的组件，可依据实际情况选择对应的服务组合，也可选择“核心 Hadoop”完成基础安装，在后续过程中，依据使用情况，再进行组件服务的增加。





将不同服务分配到不同服务器上，点击“继续”进入下一步。



输入 Hive、Oozie 的数据源配置，点击“测试连接”，测试通过后，点击“继续”按钮。

cloudera manager

支持admin

群集设置

审核更改

HDFS 根目录

hbase.rootdir

Cluster 1 > HBase (服务范围)

/hbase

?

启用编制索引

Cluster 1 > HBase (服务范围) ☒

?

启用复制

hbase.replication

Cluster 1 > HBase (服务范围) ☒

?

接受的 DataNode 失败的卷

dfs.datanode.failed.volumes.tolerated

Cluster 1 > DataNode Default Group

和其他 2 个

编辑单个值

0

?

DataNode 数据目录

dfs.data.dir, dfs.datanode.data.dir

Cluster 1 &gt; DataNode Default Group

和其他 2 个

编辑单个值

?

返回

123456

继续

依据要求，调整系统配置参数，调整后，点击“继续”按钮。

cloudera manager

支持admin

群集设置

首次运行 命令

状态: Running 开始时间: 2月 25, 10:49:57 上午 中止

详细信息 Completed 0 of 9 step(s).

全部 只会失败 Running Only

Step	上下文	开始时间	持续时间	操作
部署客户端配置	Cluster 1	2月 25, 10:49:57 上午		中止
Start Cloudera Management Service, ZooKeeper				
正在启动 HDFS 服务				
Start Solr, HBase				
Start Key-Value Store Indexer, YARN (MR2 Included)				
正在启动 Spark 服务				
正在启动 Hive 服务				
Start Oozie, Impala				

返回

123456

继续

CM 将依据前面做的配置内容，对环境进行部署，并启动服务。

按照向导完成安装即可。服务安装完成后，所有服务会自动启动。

如本步骤出现问题，注意到 CM 的运行日志中，查找对应的日志信息，具体目录为/opt/cm-5.5.3/run/cloudera-scm-agent/process。

备注: 在安装中，如安装选型选择了 Spark，暂无法进行安装成功，系统提示 JAVA\_HOME 未设置，强制调整 JAVA\_HOME 后，依然无法正常安装。

## 群集设置

恭喜您！

服务已安装、配置并在群集中运行。

[返回](#)

1 2 3 4 5 6

[完成](#)

安装完毕。

### 3.4. 安装 Kafka

#### 3.4.1. 上传介质到 CM

将 KAFKA-1.2.0.jar 上传到服务器的/opt/cloudera/csd 目录中。

将 KAFKA-2.0.0-1.kafka2.0.0.p0.12-el7.parcel、  
KAFKA-2.0.0-1.kafka2.0.0.p0.12-el7.parcel.sha1 上传到服务器的  
/opt/cloudera/parcel-repo 目录中。

同时将 KAFKA-2.0.0-1.kafka2.0.0.p0.12-el7.parcel.sha1 复制为  
KAFKA-2.0.0-1.kafka2.0.0.p0.12-el7.parcel.sha。

#### 3.4.2. 分配激活 Parcel

进入 CM 管理控制台，进入“主机”的“Parcel”界面，检查新的  
Parcel 后，分配并激活 KAFKA 对应的 Parcel。



### 3.4.3. 使用向导进行安装

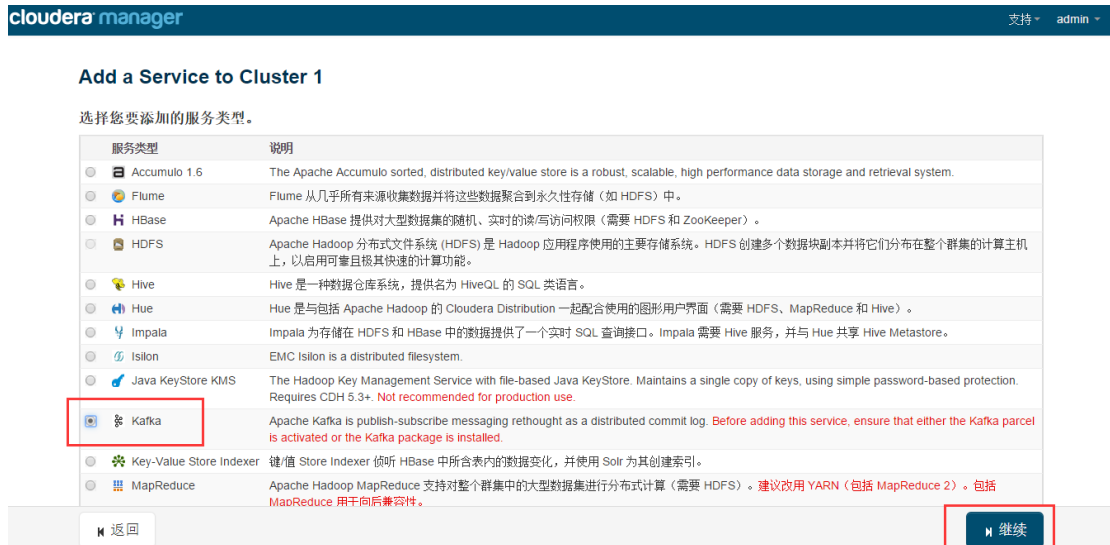
登录 CM 管理控制台。



选择需要增加 Kafka 服务的集群。



选择右侧的“操作”下拉按钮中的“添加服务”菜单项。



选择 Kafka 服务，点击“继续”按钮。



选择要安装 Kafka 服务的主机，点击“继续”按钮。

cloudera manager

支持admin

Add a Kafka Service to Cluster 1

审核更改

ZooKeeper Root  
zookeeper.chroot

Kafka-3 (服务范围)

?

Enable Kerberos Authentication  
kerberos.auth.enable

Kafka-3 (服务范围)

☐

?

Topic Auto Creation  
auto.create.topics.enable

Kafka-3 (服务范围)

☒

?

Default Replication Factor  
default.replication.factor

Kafka-3 (服务范围)

1

?

Offset Commit Topic Number of  
Partitions  
offsets.topic.num.partitions

Kafka-3 (服务范围)

50

?

Offset Commit Topic Replication

Kafka-3 (服务范围)

?

返回

123456

继续

调整 Kafka 的配置，点击“继续”按钮。

cloudera manager

支持admin

Add a Kafka Service to Cluster 1

首次运行 命令

状态: Running 开始时间: 2月 25, 8:40:45 早上 中止

详细信息 Completed 0 of 1 step(s).

全部 只会失败 Running Only

Step	上下文	开始时间	持续时间	操作
启动	Kafka-3	2月 25, 8:40:45 早上		中止
Starting 3 roles on service 0/3 start commands completed.		2月 25, 8:40:45 早上		中止
启动此 Kafka Broker	Kafka Broker (bigdataserver3)	2月 25, 8:40:45 早上		中止
启动此 Kafka Broker	Kafka Broker (bigdataserver1)	2月 25, 8:40:45 早上		中止
启动此 Kafka Broker	Kafka Broker (bigdataserver2)	2月 25, 8:40:45 早上		中止

返回

123456

继续

等待 Kafka 服务增加完成，点击“继续”按钮，并依据向导完成后续确认操作即可。

### 3.5. 其他服务增加

后续服务的新增，可以参看 3.4.安装 Kafka 的安装过程，安装过程类似，唯一的区别是不需要删除新的 Parcel 到服务器并进行激活。

## 四、备注

#### 3.5.1. Hue 切换到 MySQL

在 MySQL 创建 hue 数据库

```
mysql> CREATE DATABASE hue DEFAULT CHARSET utf8 COLLATE utf8_general_ci;
```

进入 CM 管理控制台，切换到 Hue 的“配置”一栏。



做左侧选择“数据库”，将数据库类型调整为“MySQL”，同时设置数据库的 IP、端口、用户名、密码以及数据库名称，点击“保存更改”按钮。

最后重启“Hue”服务。

### 3.5.2. 参考学习资料

- Alex 的 Hadoop 菜鸟教程

<http://blog.csdn.net/nsrainbow/article/category/2738493>

- 官方文档

<http://www.cloudera.com/documentation/enterprise/latest.html>