**大数据基础平台**

**软件安装手册**

**文件建立/修改记录**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 版本 | 建立或修改 | 建立/修改人  日期 | 审核人  日期 | 批准人  日期 |
| 1 | 1.0 | 建立 | 罗达志  2016年11月18日 |  |  |
| 2 | 2.0 | 修改 | 吴延飞  2019年8月26日 |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

目录

[第零步 简介 8](#_Toc19799136)

[1目的 8](#_Toc19799137)

[2前提要求 8](#_Toc19799138)

[3典型安装环境 8](#_Toc19799139)

[4术语列表 9](#_Toc19799140)

[5整体安装步骤 9](#_Toc19799141)

[第一步 检查生产环境 10](#_Toc19799142)

[1确认硬件信息 10](#_Toc19799143)

[2确认系统信息 10](#_Toc19799144)

[3确认网络 10](#_Toc19799145)

[第二步 优化、修改系统配置 10](#_Toc19799146)

[1系统配置修改 10](#_Toc19799147)

[1.1检查并修改主机名(可选) 10](#_Toc19799148)

[1.2修改/etc/hosts文件 11](#_Toc19799149)

[1.3配置各机器免密登录 12](#_Toc19799150)

[1.4禁用防火墙 12](#_Toc19799151)

[1.5禁用selinux 12](#_Toc19799152)

[1.6 nn1上节点配置本地yum源（file协议） 12](#_Toc19799153)

[2系统配置优化 14](#_Toc19799154)

[2.1优化系统内核参数 14](#_Toc19799155)

[2.2禁用透明大页面优化参数 14](#_Toc19799156)

[第三步 基础服务安装 15](#_Toc19799157)

[1 安装httpd服务 15](#_Toc19799158)

[1.1.1 重新配置yum源 15](#_Toc19799159)

[2安装ftp服务 16](#_Toc19799160)

[2.1安装vsftpd组件 16](#_Toc19799161)

[2.2安装ftp客户端组件（用来验证vsftpd是否正常工作） 16](#_Toc19799162)

[2.3取消匿名登录 17](#_Toc19799163)

[2.4开启root用户ftp使用权限 17](#_Toc19799164)

[3安装NTP服务 17](#_Toc19799165)

[3.1安装ntp server 17](#_Toc19799166)

[3.2安装ntp client 18](#_Toc19799167)

[3.3检查时间同步情况 18](#_Toc19799168)

[4安装Mysql服务 19](#_Toc19799169)

[4.1卸载原有MySQL服务 19](#_Toc19799170)

[4.2上传MySQL RPM安装介质 19](#_Toc19799171)

[4.3安装MySQL 19](#_Toc19799172)

[4.3初始化MySQL 20](#_Toc19799173)

[5安装JDK 21](#_Toc19799174)

[5.1上传jdk1.8的tar包 21](#_Toc19799175)

[5.2 配置java环境变量 22](#_Toc19799176)

[第四步 服务、空间规划 22](#_Toc19799177)

[第五步 安装Cloudera Manager 23](#_Toc19799178)

[1上传安装文件 23](#_Toc19799179)

[2 CM6.2安装 23](#_Toc19799180)

[3 CDH6.2安装 24](#_Toc19799181)

[4安装CFM 29](#_Toc19799182)

[第六步 安装CDH组件 31](#_Toc19799183)

[1安装ZOOKEEPER 31](#_Toc19799184)

[1.2 Zookerper服务添加 31](#_Toc19799185)

[1.3 Zookeeper服务验证 32](#_Toc19799186)

[2安装HDFS并启用HA 32](#_Toc19799187)

[2.1 Hdfs服务添加 32](#_Toc19799188)

[2.2 Hdfs服务验证 38](#_Toc19799189)

[3安装YARN 39](#_Toc19799190)

[3.1 Yarn服务添加 39](#_Toc19799191)

[3.2 Yarn服务验证 41](#_Toc19799192)

[4安装HIVE 41](#_Toc19799193)

[4.1 Hive服务添加 41](#_Toc19799194)

[4.2 Hive服务验证 43](#_Toc19799195)

[5安装IMPALA 45](#_Toc19799196)

[5.1 Impala服务添加 45](#_Toc19799197)

[5.2 Impala服务验证 47](#_Toc19799198)

[6安装HBASE 47](#_Toc19799199)

[6.1 Hbase服务添加 47](#_Toc19799200)

[6.2 Hbase使用验证 50](#_Toc19799201)

[7安装PHOENIX（非必须） 50](#_Toc19799202)

[7.1 Phoenix安装 50](#_Toc19799203)

[7.2 Phoenix验证 52](#_Toc19799204)

[8安装Oozie 54](#_Toc19799205)

[8.1 Oozie服务添加 54](#_Toc19799206)

[9安装HUE 54](#_Toc19799207)

[9.1 hue服务添加 54](#_Toc19799208)

[10安装SPARK2（非必须） 55](#_Toc19799209)

[10.1安装CDH自带版本 55](#_Toc19799210)

[10.2替换CDH的spark版本 56](#_Toc19799211)

[10.3验证spark的使用 58](#_Toc19799212)

[11安装kerberos（非必须） 58](#_Toc19799213)

[11.1安装kdc server 58](#_Toc19799214)

[11.2修改配置文件 59](#_Toc19799215)

[11.3同步配置文件 61](#_Toc19799216)

[11.4创建数据库 61](#_Toc19799217)

[11.5启动服务 62](#_Toc19799218)

[11.6创建Kerberos管理员 62](#_Toc19799219)

[11.7测试Kerberos 62](#_Toc19799220)

[12 CDH集群启用Kerberos（非必须） 65](#_Toc19799221)

[12.1 CDH启动kerberos 65](#_Toc19799222)

[12.2验证各组件服务的使用 68](#_Toc19799223)

[13 HBASE权限修改（非必须） 73](#_Toc19799224)

[13.1 Hbase配置修改 73](#_Toc19799225)

[13.2 hbase登录验证 73](#_Toc19799226)

[13.3 phoenix登录验证 73](#_Toc19799227)

[14 CDH启用sentry服务 74](#_Toc19799228)

[14.1添加sentry服务 74](#_Toc19799229)

[14.2修改sentry配置 74](#_Toc19799230)

[15 CDH启用ACL服务 76](#_Toc19799231)

[16安装LDAP认证（非必须） 76](#_Toc19799232)

[1.1.2 安装LDAP服务 76](#_Toc19799233)

[1.1.3 IMPALA集成LDAP 78](#_Toc19799234)

[1.1.4 Impala验证 78](#_Toc19799235)

[17安装spark集群（on yarn模式独立安装） 79](#_Toc19799236)

[17.1部署环境说明 79](#_Toc19799237)

[17.2安装spark 80](#_Toc19799238)

[17.3 Spark验证 81](#_Toc19799239)

[18安装REDIS集群 83](#_Toc19799240)

[18.1安装redis服务 83](#_Toc19799241)

[18.2 Redis集群配置 85](#_Toc19799242)

[18.3 Redis集群创建 88](#_Toc19799243)

[18.4 Redis集群重启 90](#_Toc19799244)

[19安装zookeeper集群（独立部署） 90](#_Toc19799245)

[19.1部署环境说明 90](#_Toc19799246)

[19.2安装zookeeper集群 91](#_Toc19799247)

[20安装KAFKA（CDH集成版，非必须） 93](#_Toc19799248)

[20.1部署环境说明 93](#_Toc19799249)

[20.2 CDH安装kafka 93](#_Toc19799250)

[21安装KAFKA（独立部署，非kerberos） 96](#_Toc19799251)

[21.1部署环境说明 96](#_Toc19799252)

[21.2安装kafka 96](#_Toc19799253)

[22安装STORM 98](#_Toc19799254)

[22.1部署环境说明 98](#_Toc19799255)

[22.2安装Storm 98](#_Toc19799256)

[23安装FLUME 102](#_Toc19799257)

[23.1部署环境说明 102](#_Toc19799258)

[23.2安装Flume 102](#_Toc19799259)

[问题记录 103](#_Toc19799260)

# 第零步 简介

## 1目的

该手册旨在标准化公司在各地部署大数据CDH集群的流程，提高利用CM部署集群的效率和减少出现问题的可能性，帮助现场部署人员更好，更快地完成部署任务。

## 2前提要求

该手册要求，现场部署环境已经安装好操作系统，建议centos7.4，且网络已经配置完毕（不要求连互联网），磁盘空间达到TB级别，最好按照磁盘的格式，进行目录划分, 如 /data1-N，根目录空间至少500G。

## 3典型安装环境

一般情况下，给到我们大数据的是几台安装好操作系统的裸机，典型的配置如下：

操作系统版本：centos7.4 / 64位

内存：128G

磁盘：4.8T， 根目录500G，四个增加的磁盘分别为1.1T,分别挂载到 /data1-4

网络环境：专网（局域网），无法连通互联网，因此全部需要离线安装

系统核数：28核

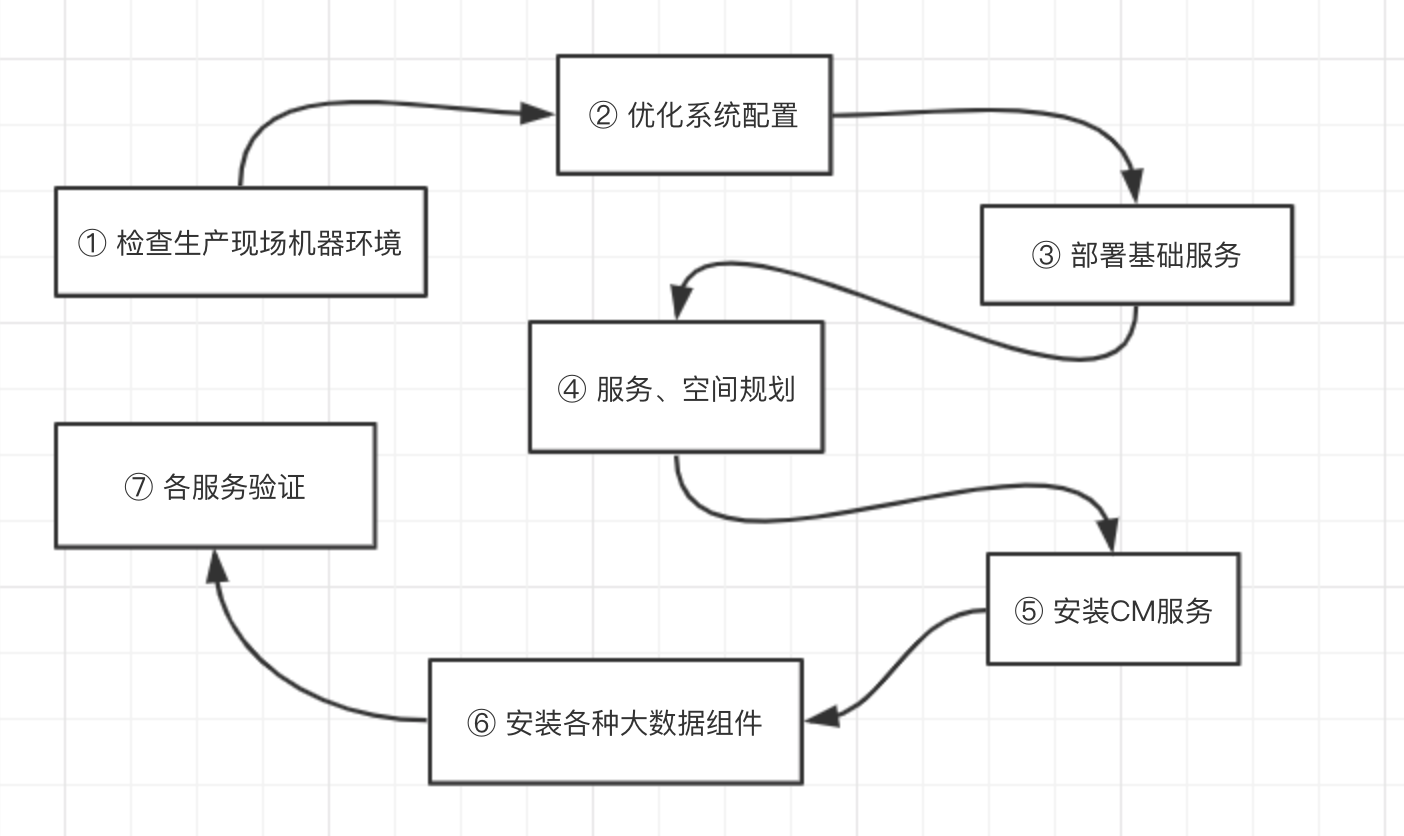
## 4术语列表

Cloudera Manager（CM）：Cloudera公司的Hadoop系统组件的安装管理工具

CDH：Cloudera's Distribution Hadoop，是Cloudera公司发布的Hadoop版本

## 5整体安装步骤

得到安装好centos系统，并且网络正常的集群机器后，安装步骤分为七步：



# 第一步 检查生产环境

## 1确认硬件信息

硬件信息一般装系统的同学会提供，且集群的机器也会按照我们的要求进行配置，检验几个核心的系统和硬件参数：

内存、磁盘空间和目录划分、系统核数

可用如下命令查看：

free -g、df -h、cat /proc/cpuinfo | grep "processor" | sort | uniq | wc -l

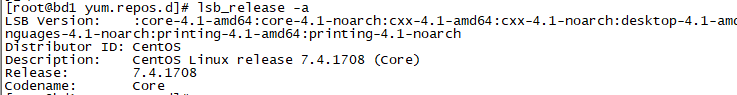
## 2确认系统信息

保证操作系统是要求的centos系列，最低要求7.4，位数建议64位

请先安装 redhat-lsb （需要配置yum源，见第二步1.6）

yum install –y redhat-lsb

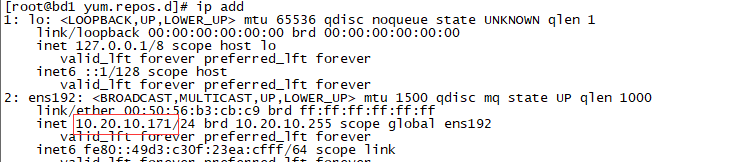
然后使用 命令lsb\_release –a查看Linux版本信息



## 3确认网络

保证各机器已经分配内网IP，且各机器之间互通

利用 ip add命令查看IP，用ping命令验证网络互通



# 第二步 优化、修改系统配置

## 1系统配置修改

### 1.1检查并修改主机名(可选)

检查机器名，需要调整为小写（大写在使用一些工具中会有问题）

vi /etc/sysconfig/network

然后将HOSTNAME后面的值改为想要设置的主机名。



*示例：bd1.bcht, bd2.bcht, bd3.bcht*

**主机名命名规则：【节点类型】 + 【编号】 + 【.】 + 【部门简写】**

例如广州交管局的机器主机名为：

dn1.gdjgj \dn2.gdjgj\dn1.gdjgj\dn2.gdjgj\dn3.gdjgj

dn=datanode

nn=namenode

### 1.2修改/etc/hosts文件

选择一台机器，设置本机的主机名和IP映射关系

加上模拟archive.cloudera.com这个域名的IP（如果不联网，则最好加上）

例如加上如下内容：

44.5.10.195 nn1.gdjgj nn1

44.5.10.196 nn2.gdjgj nn2

44.5.10.197 dn1.gdjgj dn1

44.5.10.198 dn2.gdjgj dn2

44.5.10.199 dn3.gdjgj dn3

44.5.10.195 archive.cloudera.com （IP为yum源所在的机器IP地址）

将上述内容同步至集群中所有机器的 /etc/hosts文件中。

**实际部署，请将上述IP替换为集群机器所分配的IP地址！**

为简单起见，在后面的文档中已ip地址的均以主机名简称来称呼，例如44.5.10.195

简称nn1主机。其他主机以此类推。

### 1.3配置各机器免密登录

在所有namenode节点（此处两台主机 nn1\nn2) 都执行以下命令：

ssh-copy-id root @44.5.10.195

ssh-copy-id root @44.5.10.196

ssh-copy-id root @44.5.10.197

ssh-copy-id root @44.5.10.198

ssh-copy-id root @44.5.10.199

一共需要执行 2（台机器）\* 5（条命令）=10次命令

### 1.4禁用防火墙

systemctl disable firewalld

### 1.5禁用selinux

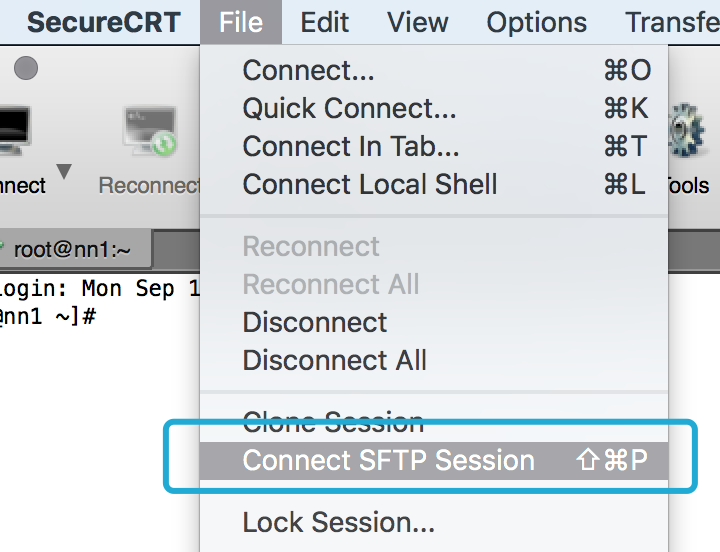
sed -i ‘s/SELINUX=enforce/SELINUX=disabled/g’ /etc/selinux/config

### 1.6 nn1上节点配置本地yum源（file协议）

此步骤是为了安装httpd服务，因为默认的裸操作系统没有httpd服务，但此服务又是必须的，在没有联网的情况下，想要通过yum安装httpd,必须配置本地的yum源。

#### 1.6.1 上传ISO镜像文件

利用secureCRT的SFTP session将镜像文件 CentOS-7-x86\_64-DVD-1810.iso 上传至nn1的/root目录下，默认不带lrzsz工具，且无法通过yum安装。



put 本地ISO镜像文件地址

通过put命令上传。

#### 1.6.2 挂载镜像文件

创建空目录：

mkdir /root/centos7.4

挂载ISO文件到该空目录下：

mount -o loop -t iso9660 /root/CentOS-7-x86\_64-DVD-1810.iso /root/centos7.4

#### 1.6.3 备份yum.repos.d目录

mv /etc/yum.repos.d /etc/yum.repos.d\_bak

mkdir /etc/yum.repos.d

#### 1.6.4 新建配置文件将其指向刚挂载的ISO文件目录

vi /etc/yum.repos.d/centos7.4.repo 添加如下内容

[centos7.4]

name=centos7.4

baseurl=file:///root/centos7.4

gpgcheck=0

enabled=1

:wq保存即可

#### 1.6.5 重建yum缓存

yum clean all

yum makecache

#### 1.6.6 测试yum正确配置

yum install -y gcc\*

## 2系统配置优化

### 2.1优化系统内核参数

vi /etc/sysctl.conf 添加下面这一行

vm.swappiness=1 #使用到达99%，才开始用Swap虚拟内存

sysctl –p #使配置生效

### 2.2禁用透明大页面优化参数

vi /etc/rc.local 文件并添加以下两行

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

# 第三步 基础服务安装

## 1 安装httpd服务

默认情况下是没有安装的，通过以下命令安装：(因为前面我们已经配置了本地yum源）

yum install -y httpd

然后启动并设置开机启动：

systemctl start httpd

systemctl enable httpd  
**注意: 只需要在nn1节点上安装即可 !**

### 重新配置yum源

之前的yum源是file协议的，这样需要每台机器都进行ISO文件上传、挂载等繁琐操作，所以通过前面我们安装的httpd服务，我们将yum源改为http协议，集群中所有的机器均访问nn1这台机器的ISO镜像文件夹即可。具体操作步骤如下：（在nn1上操作）

#### 1.1.1 重新挂载ISO镜像文件

umount /root/cento7.4

mkdir /var/www/html/centos7.4

mount -o loop -t iso9660 /root/CentOS-7-x86\_64-DVD-1810.iso /var/www/html/centos7.4

#### 1.1.2 修改yum配置，改为http协议

vi /etc/yum.repos.d/centos7.4.repo 内容修改为：

[centos7.4]

name=centos7.4

baseurl=http://44.5.10.195/centos7.4

gpgcheck=0

enabled=1

#### 1.1.3 将nn1的yum配置同步到所有的机器上

将除nn1之外所有机器的yum配置清除，每天机器执行:

rm -rf /etc/yum.repos.d/\*

然后通过nn1将配置同步到所有其他的机器上

scp /etc/yum.repos.d/centoos7.4.repo root@nn2.gdjgj:/etc/yum.repos.d/

scp /etc/yum.repos.d/centoos7.4.repo root@dn1.gdjgj:/etc/yum.repos.d/

scp /etc/yum.repos.d/centoos7.4.repo root@dn2.gdjgj:/etc/yum.repos.d/

scp /etc/yum.repos.d/centoos7.4.repo root@dn3.gdjgj:/etc/yum.repos.d/

每台机器重建yum缓存

yum clean all

yum makecache

## 2安装NTP服务

### 2.1安装ntp server

选择一台机器作为NTP主服务器，例如以下是使用nn1主机作为ntp主服务器

首先安装NTP服务

yum -y install ntp

然后编辑 vi /etc/ntp.conf 添加以下内容

restrict 44.5.10.0 mask 255.255.255.0 nomodify notrap

server 127.127.1.0

fudge 127.127.1.0 stratum 8

启动ntp服务

service ntpd status

service ntpd start

设置开机自启动：

systemctl enable ntpd

### 2.2安装ntp client

除server所在的主机外，其他主机全部作为client，安装步骤如下：

首先安装ntp

yum -y install ntp

修改ntp配置 vi /etc/ntp.conf

server 44.5.10.195 #指向ntp服务器nn1主机

fudge 44.5.10.195 stratum 8

启动NTP服务：systemctl start ntpd

设置开机自启动：system enable ntpd

### 2.3检查时间同步情况

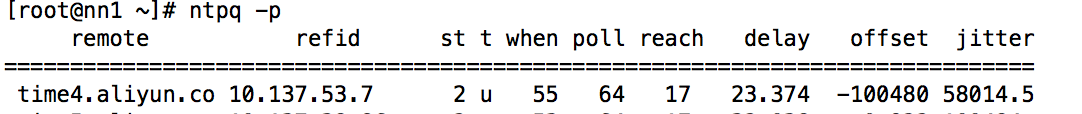
查看状态:

ntpstat



查看偏移量：

ntpq -p



## 3安装Mysql服务

### 3.1卸载原有MySQL服务

mysql安装前必须保证系统中原有的MySQL或者mariadb被卸载：

查询命令: rpm -qa |grep mariadb

卸载命令: rpm -e --nodeps mariadb\*

查询命令: rpm -qa |grep mysql

卸载命令: rpm -e --nodeps mysql\*

### 3.2上传MySQL RPM安装介质

通过SFTP session将文件 mysql-5.7.20-1.el7.x86\_64.rpm-bundle.tar 上传至nn1主机的/root目录下，进行解压。

tar -xvzf mysql-5.7.20-1.el7.x86\_64.rpm-bundle.tar

### 3.3安装MySQL

安装Mysql需要一些依赖的服务，所以要提前安装好，命令如下:

yum -y install numactl

yum –y install net-tools

然后依次安装MySQL相关的rpm包，**按顺序安装！**

rpm -ivh mysql-community-common-5.7.20-1.el7.x86\_64.rpm

rpm -ivh mysql-community-libs-5.7.20-1.el7.x86\_64.rpm

rpm -ivh mysql-community-libs-compat-5.7.20-1.el7.x86\_64.rpm

rpm -ivh mysql-community-devel-5.7.20-1.el7.x86\_64.rpm

rpm -ivh mysql-community-client-5.7.20-1.el7.x86\_64.rpm

rpm -ivh mysql-community-server-5.7.20-1.el7.x86\_64.rpm

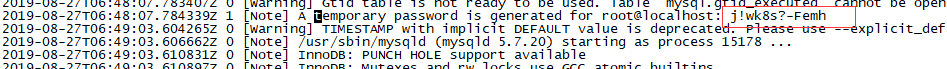
上传JDBC驱动包mysql-connector-java.jar到/opt目录 并将该jar包拷贝到/usr/share/java目录下。

然后启动MySQL

systemctl start mysqld

### 3.3初始化MySQL

#### 3.3.1 从启动日志中拿到Mysql安装成功后随机生成的密码



#### 3.3.2 用root用户和随机密码登录MySQL

mysql -uroot -pj!Wk8s?-Femh #默认链接locahost的3306端口

#### 3.3.3 重置密码

登录之后必须要进行重置密码才能进行其他操作

set password=password(‘mima’); # 请根据实际情况，自行设定较为复杂的密码, 如果提示密码不符合校验规则，设置校验规则

set global validate\_password\_policy=LOW;

set global validate\_password\_length=2;

#### 3.3.4 修改默认编码

**此步骤非常重要，必须要将默认编码改为utf8，然后重启MySQL服务**

vi /etc/my.cnf

#在[mysqld]下面添加utf8配置

character-set-server=utf8

collation-server=utf8\_general\_ci

lower\_case\_table\_names=1 #设置不mysql不区分大小写

#在[client]下面添加utf8配置

default-character-set=utf8

保存文件后，重启MySQL即可生效。

systemctl stop mysqld

systemctl start mysqld

systemctl enable mysqld

#### 3.3.5添加数据所需的库，用户以及权限

给root用户开启远程访问权限:

# 修改mysql的密码的校验规则，否则无法设置简单密码

set global validate\_password\_policy=LOW;

set global validate\_password\_length=2;

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

flush privileges;

创建用户、库以及权限，列表如下：

hive、oozie、scm、hue

创建命令如下:

# 修改mysql的密码的校验规则，否则无法设置简单密码

set global validate\_password\_policy=LOW;

set global validate\_password\_length=2;

# 添加库，用户，权限

CREATE DATABASE hive;

CREATE USER 'hive'@'%' IDENTIFIED BY 'hive';

GRANT ALL PRIVILEGES ON hive.\* TO 'hive'@'%';

flush privileges;

# 请依次将上述红色字体换为 oozie\cmserver\hue执行

## 4安装JDK

### 4.1上传jdk1.8的tar包

通过SFTP session上传 jdk1.8.0\_191.tar.gz 到 /opt 目录下，然后解压

tar -xvzf jdk1.8.0\_191.tar.gz

### 4.2 配置java环境变量

vi /etc/profile #添加以下内容

export JAV\_HOME=/opt/jdk1.8.0\_191

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

执行source /etc/profile 以使配置生效

**注意: 所有的机器都需进行上述操作，安装JDK，tar包可以通过scp命令互传，不需要每台机器都进行上传！**

**至此，所有的所需的环境配置和基础服务已全部完成，但是有些配置需要重启才会生效，所以需要重启一遍所有机器，命令reboot !**

# 第四步 服务、空间规划

针对给定的所有主机，需要按照下述表格进行服务和空间规划：

(请根据实际的机器列表情况进行规划，最好经过讨论，确定最终方案)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 服务组 | 服务 | **nn1** | **nn2** | **dn1** | **dn2** | **dn3** |
| Zookeeper | Zookeeper Server |  |  |  |  |  |
| HDFS | NameNode |  |  |  |  |  |
| DataNode |  |  |  |  |  |
| JournalNode |  |  |  |  |  |
| Failover Controller |  |  |  |  |  |
| YARN | ResourceManager |  |  |  |  |  |
| NodeManager |  |  |  |  |  |
| JobHistory Server |  |  |  |  |  |
| Hive | HiveServer2 |  |  |  |  |  |
| Hive Metastore Server |  |  |  |  |  |
| Hbase | Hmaster |  |  |  |  |  |
| RegionServer |  |  |  |  |  |
| Spark (Standalone) | Master |  |  |  |  |  |
| Worker |  |  |  |  |  |
| Impala | Impala Catalog Server |  |  |  |  |  |
| Impala StateStore |  |  |  |  |  |
| Impala Daemon |  |  |  |  |  |
| Oozie | Oozie Server |  |  |  |  |  |
| ElasticSearch | ElasticSearch |  |  |  |  |  |
| Hue | Hue Server |  |  |  |  |  |
| Algorithm Engine | ApiApplicationServer |  |  |  |  |  |
| LoggerServer |  |  |  |  |  |
| AlertServer |  |  |  |  |  |
| MasterServer |  |  |  |  |  |
| WorkServer |  |  |  |  |  |

# 第五步 安装Cloudera Manager

## 1上传安装文件

上传CM6和CDH6.2.0的安装文件。在配置了yum的主机上（nn1主机）上传安装文件，从安装介质中找到cm6、parcel-repo两个目录，并上传至nn1主机/var/www/html目录下。

并且需要再之前的yum配置文件**追加**如下配置，否则无法通过yum安装cloudera manager 的服务：

[cloudera-manager]

name=cloudera-manager

baseurl=http://44.5.10.195/cm6

gpgcheck=0

enabled=1

## 2 CM6.2安装

在主节点上进行Cloudera Manager Server的安装：

sudo yum install cloudera-manager-daemons cloudera-manager-agent cloudera-manager-server

在其余节点上进行Cloudera Manager Agent的安装：

sudo yum install cloudera-manager-daemons cloudera-manager-agent

修改配置： /etc/cloudera-scm-agent/config.ini

更改server\_host属性为对应的域名既可

校验数据库，执行：

# cd /opt/cloudera/cm/schema/

# ./scm\_prepare\_database.sh -h 主机名 -P 3306 mysql scm 账号 密码

启动server和agent：

再server所在机器执行：

systemctl start cloudera-scm-server;

systemctl enable cloudera-scm-server;

在所有节点执行：

systemctl start cloudera-scm-agent;

systemctl enable cloudera-scm-agent;

Cloudera Manager安装完成后，会在主节点（Server节点）的机器上会打开7180端口，这时在浏览器中访问这个地址，即可进入Cloudera Hadoop的管理界面。

<http://44.5.10.195:7180>

默认的账号和密码：admin/admin

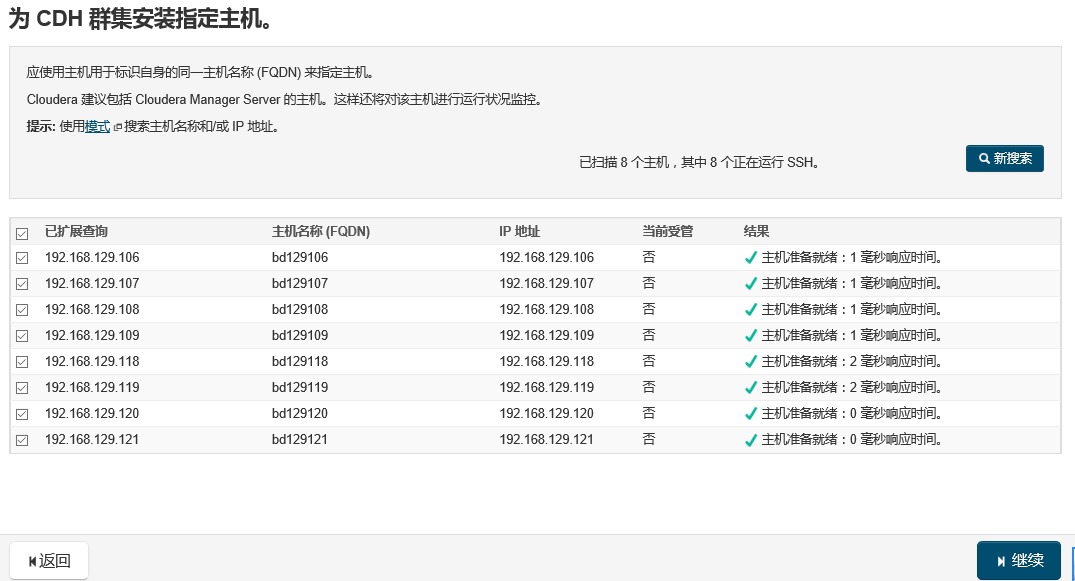


以上CM6.2安装完成，这里选择免费版，点击继续，使用CM6.2来安装CDH6.2.0及相关服务组件：

## 3 CDH6.2安装

(截图较旧，但原理相同)

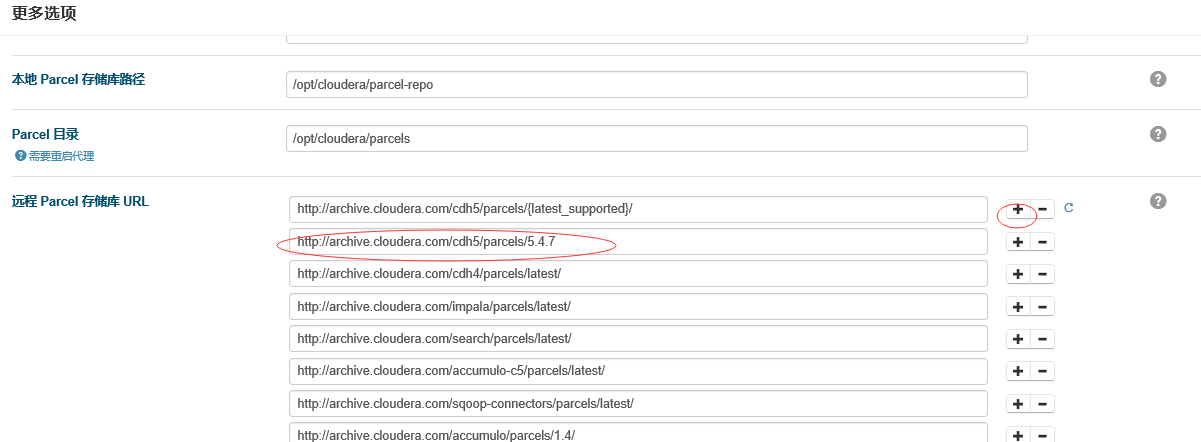




继续下一步：

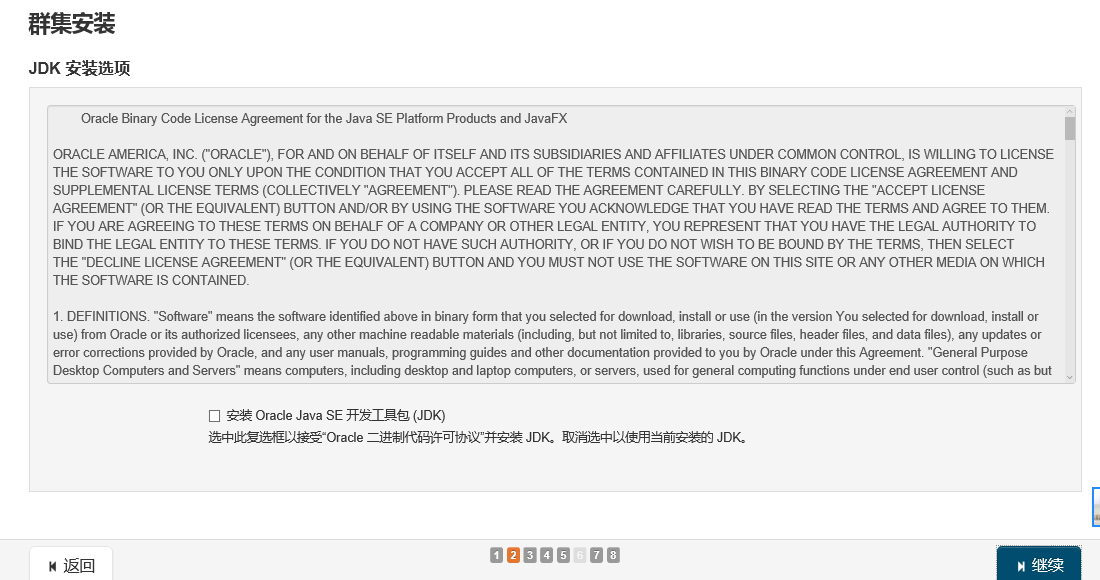


点击更多选项，添加cdh6的安装包地址：





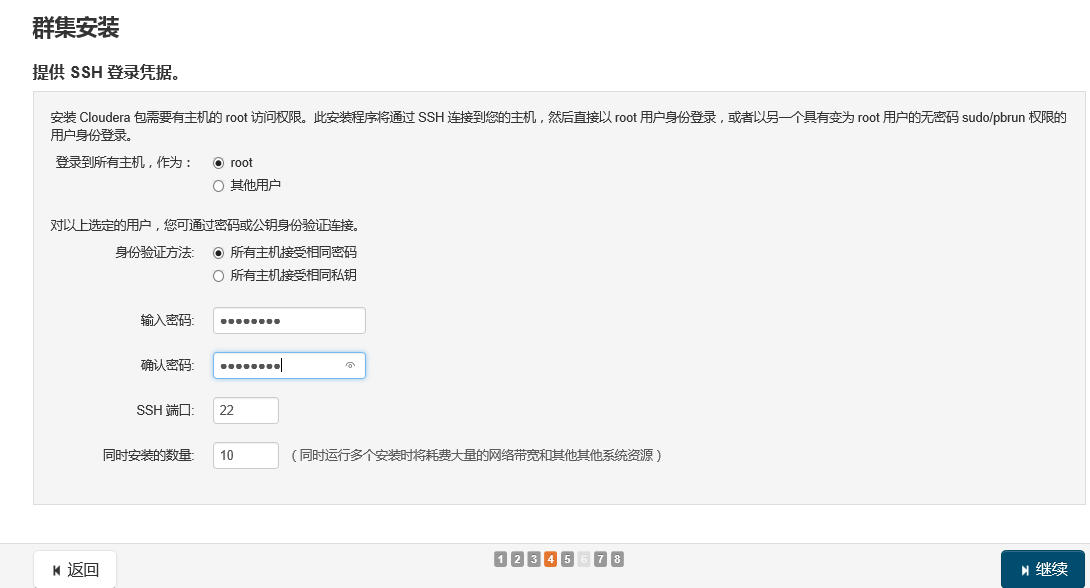
点击确定后，可以看到cm已经搜索到了cdh的安装版本，点击继续



点击继续：



点击继续：



输入5台主机的root帐号密码，点击继续：



点击继续：

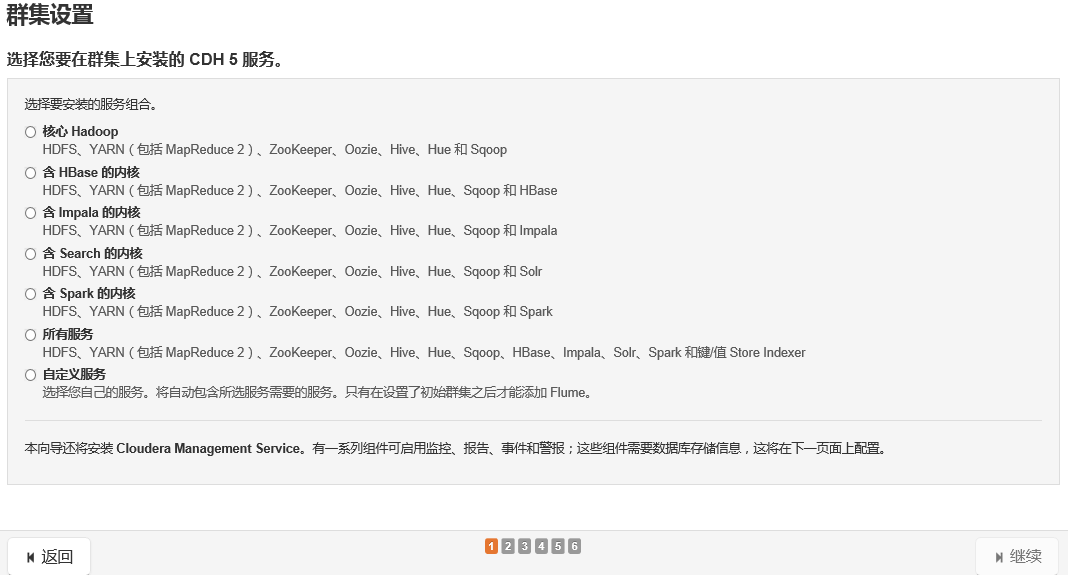


点击继续：



检查告警无影响，点击继续：

下面开始进行CDH6.2.0相关服务组件的安装：



选择自定义服务，然后往下按顺序进行服务的添加：

## 4安装CFM

（此处为了安装NIFI）想要在CM中使用CFM，需要安装CSD文件

将下载好的三个Nifi相关的Jar包文件上传至nn1的 /opt/cloudera/csd目录下：

三个包名分别为：

NIFI-1.9.0.1.0.0.0-90.jar

NIFICA-1.9.0.1.0.0.0-90.jar

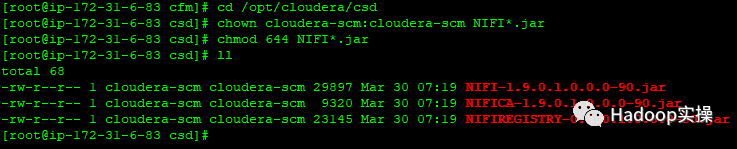
NIFIREGISTRY-0.3.0.1.0.0.0-90.jar

进行如下操作

ls /opt/cloudera/csd

chown cloudera-scm:cloudera-scm NIFI\*.jar

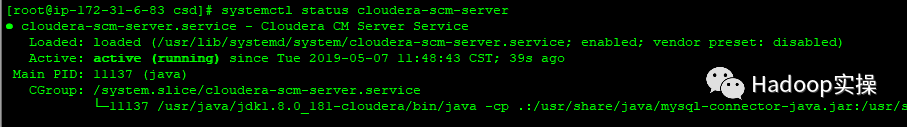
chmod 644 NIFI\*.jar



重启cloudera-scm-server服务

systemctl restart cloudera-scm-server

systemctl status cloudera-scm-server



此时，便可以通过CM安装NIFI相关组件了。

# 第六步 安装CDH组件

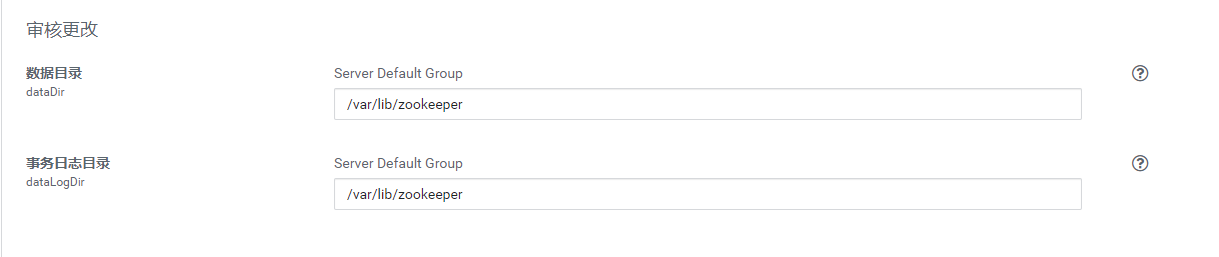
## 1添加Zookeeper服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择zookerper

分配主机和角色：



这里选择dn1.gdjgj，dn2.gdjgj和dn3.gdjgj三台主机作zookeeper服务器，点击继续：



采用默认配置即可，点击继续，完成Zookeeper的安装既可！

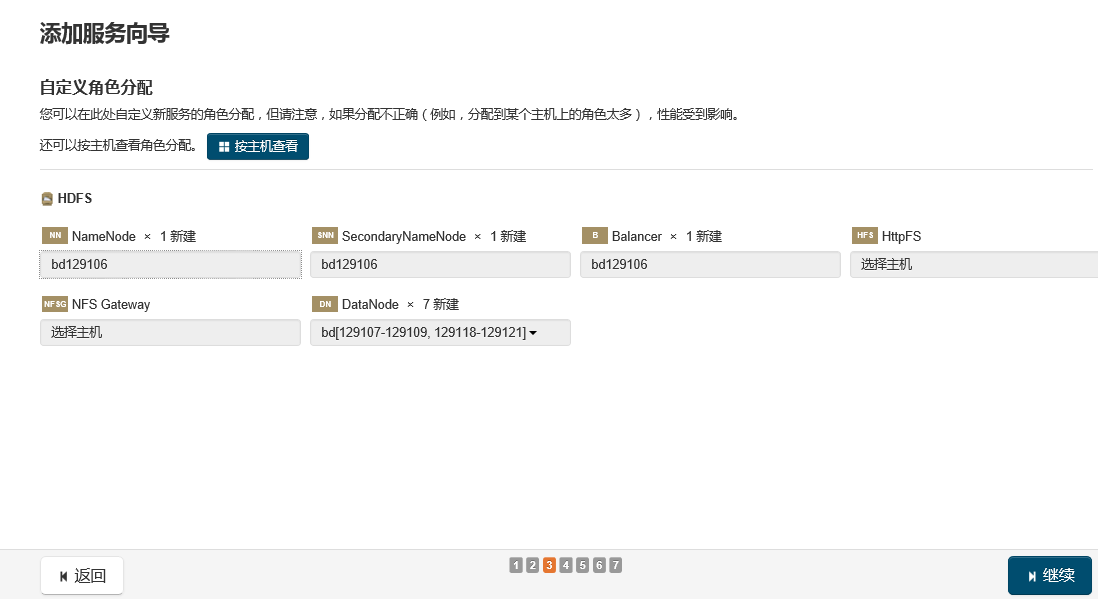
2、Zookeeper服务验证

执行：/opt/cloudera/parcels/CDH-6.2.0-1.cdh6.2.0.p0.967373/lib/zookeeper/bin/zkCli.sh -server dn1.gdjgj:2181

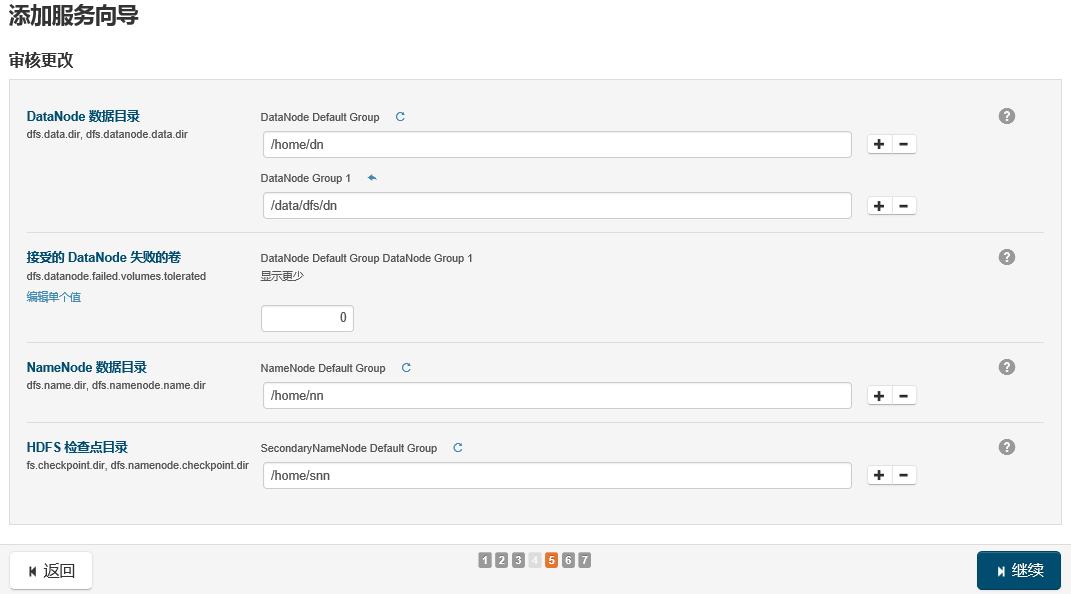
## 2添加HDFS服务，并启用HA

1. 登录CM界面，添加服务――下拉“操作”――添加服务――选择hdfs

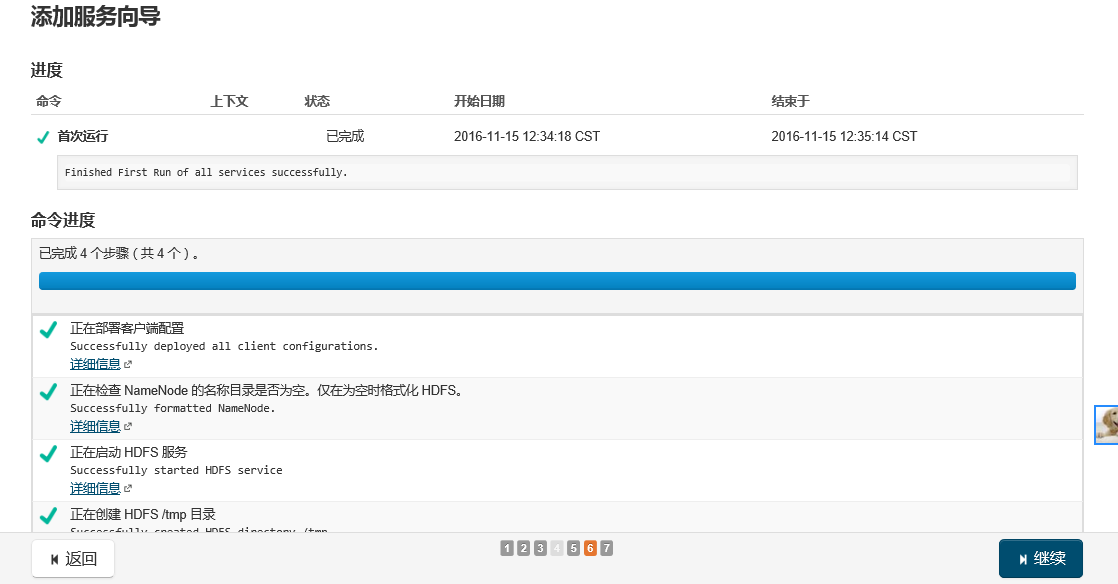
分配主机和角色



点击继续：



修改namenode和datanode目录到/data1-N下，多目录的情况下要配置多目录，点击继续：

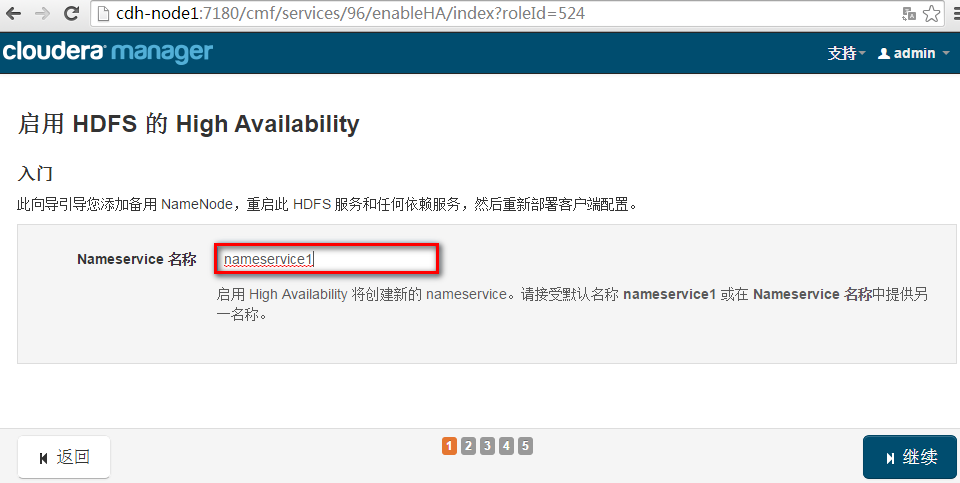


**2、配置Namenode 的HA**

**进入HDFS界面，点击“启用High Availability”**



**输入NameService名称，这里设置为：nameservice1，点击继续按钮。**



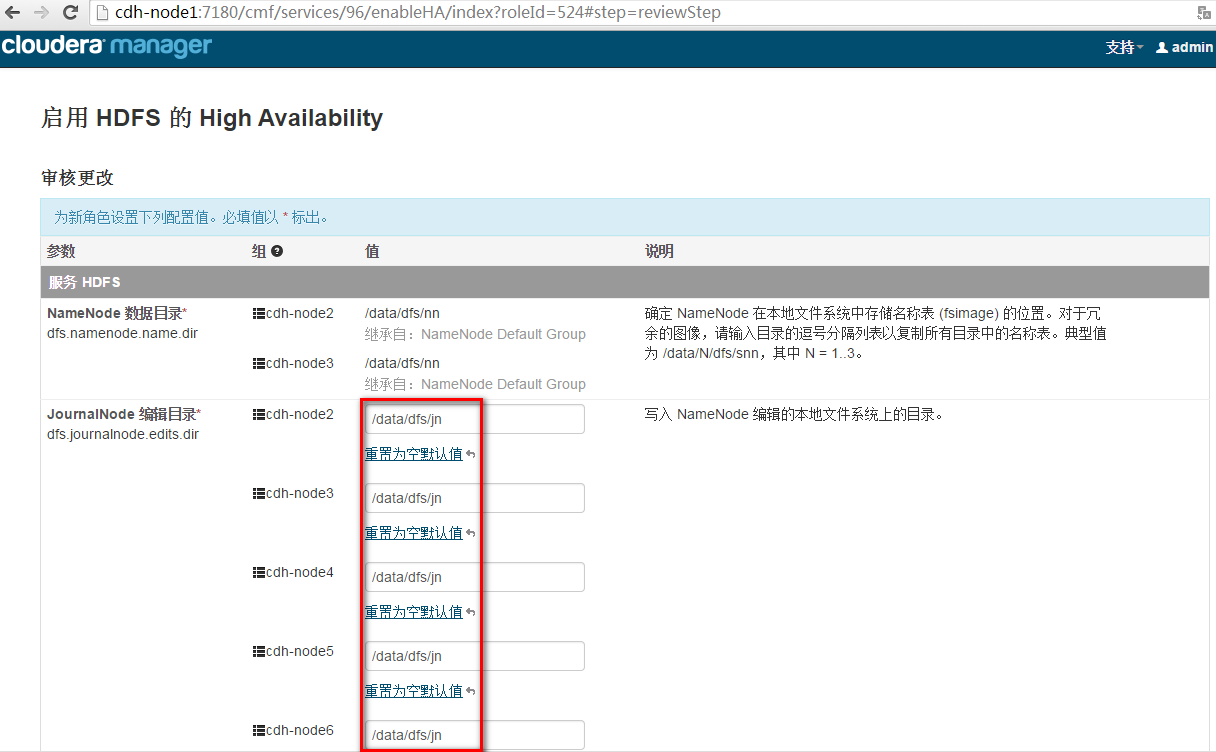
**设置另一个NameNode节点，这里设置为：nn1.gdjgj**

**设置JournalNode节点，这里设置为：dn1.gdjgj，dn2.gdjgj和dn3.gdjgj一共3个节点。**

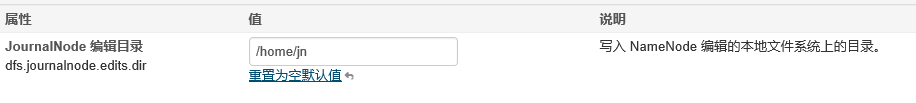
**注意：JournalNode必须设置>=3个节点。**



**设置JournalNode目录，这里全部设置为：/data1/jn。**







**启用HDFS的High Availability。**





如果发现有如下错误信息，可以忽略。



**成功启用HA**





3、验证HDFS服务

查看hdfs目录: hadoop fs -ls /

创建测试目录：hadoop fs -mkdir /tmp/test

上传测试文件：hadoop fs -put test.txt /tmp/test

查看上传文件：hadoop fs -cat /tmp/test/\*

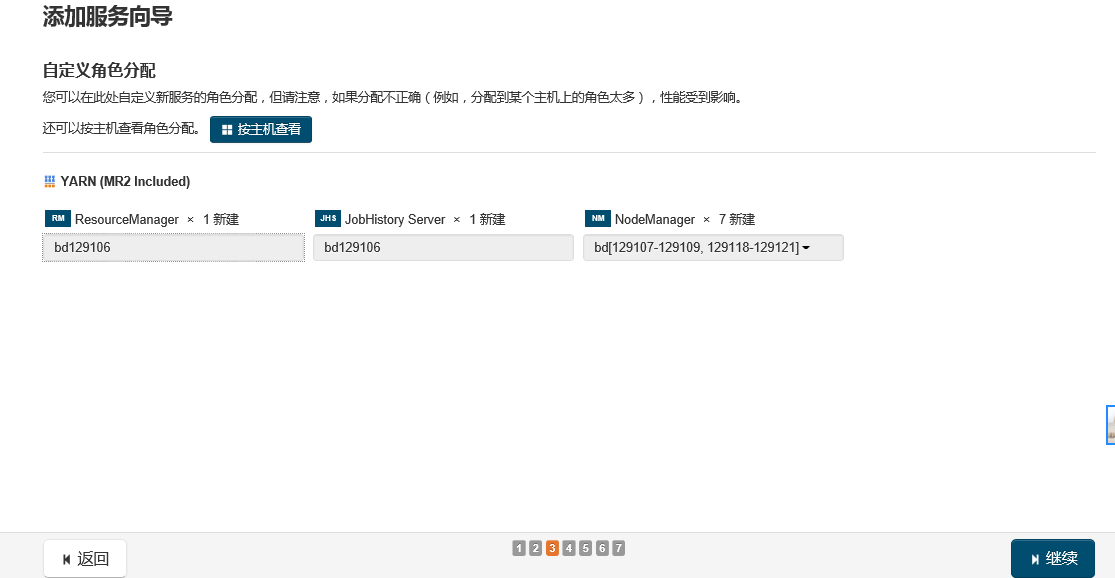
测试下载文件：先删除主机上的test.txt

hadoop fs -get /tmp/test/\*

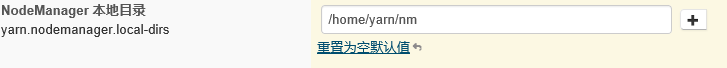
## 3添加Yarn服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择yarn

分配主机和角色：

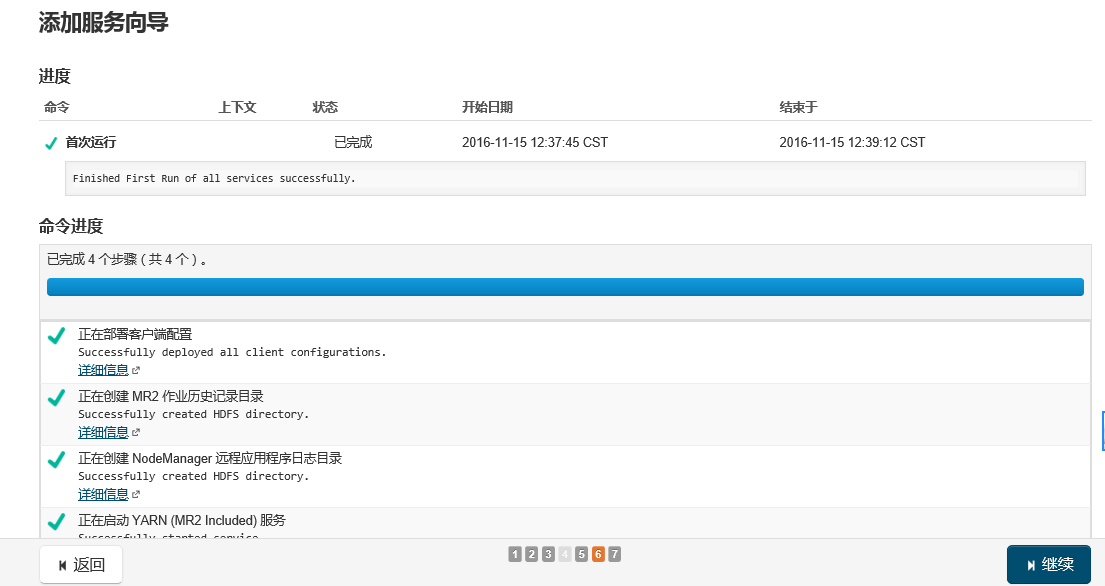


点击继续：



修改NodeManager本地目录到/data/yarn/nm 下面

点击继续：



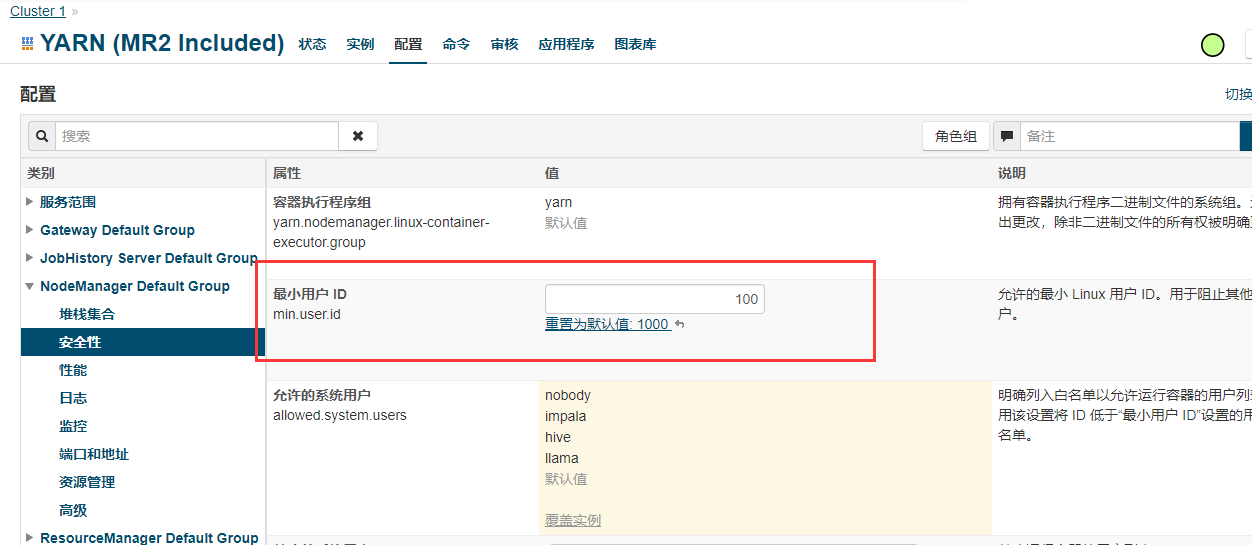
修改YARN资源管理配置：



需要根据实际分配的主机CPU和内存进行调整：

一般最大容器内存分配建议为每个节点总内存的60%， cpu建议分配每个节点CPU总内核数的50~60%

userid 选项默认1000 需要修改为100



2、Yarn服务验证

su – hdfs

hadoop fs -mkdir /user/root

hadoop fs -chown root:root /user/root

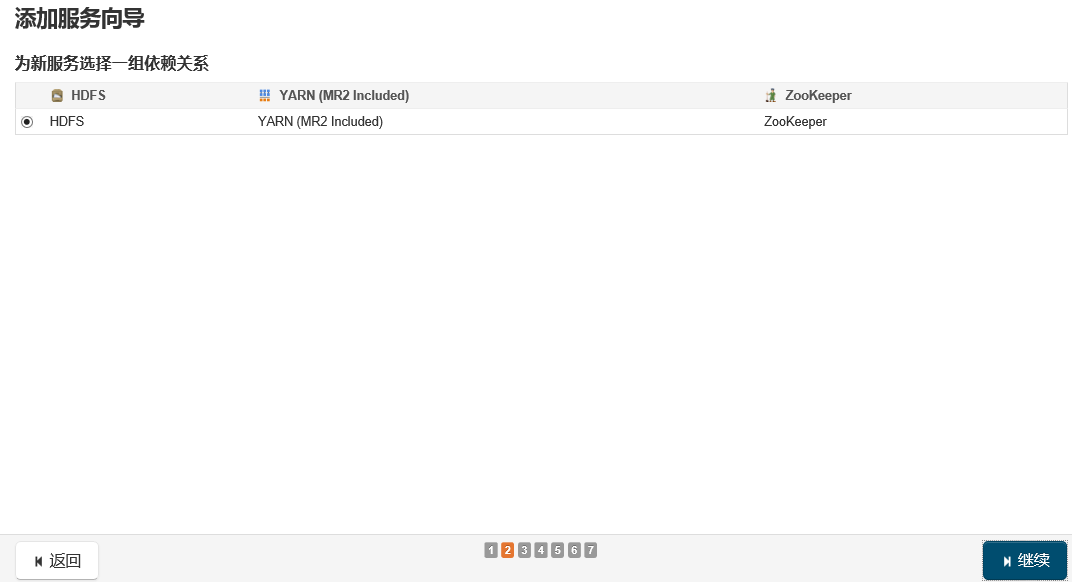
exit;

hadoop jar /opt/cloudera/parcels/CDH/jars/hadoop-examples.jar pi 4 4

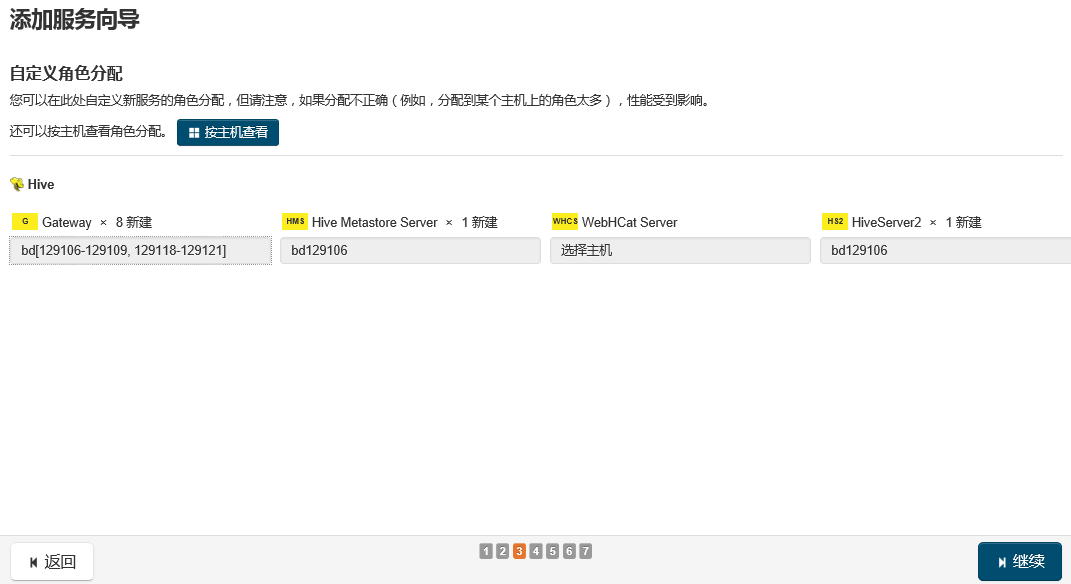
## 4添加Hive服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择hive

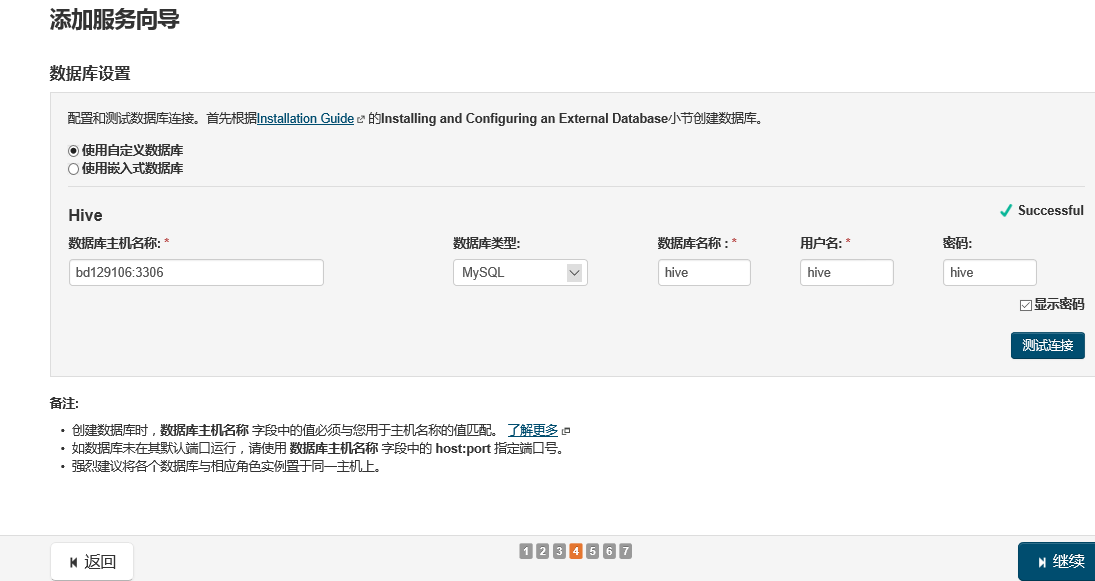
添加依赖关系：



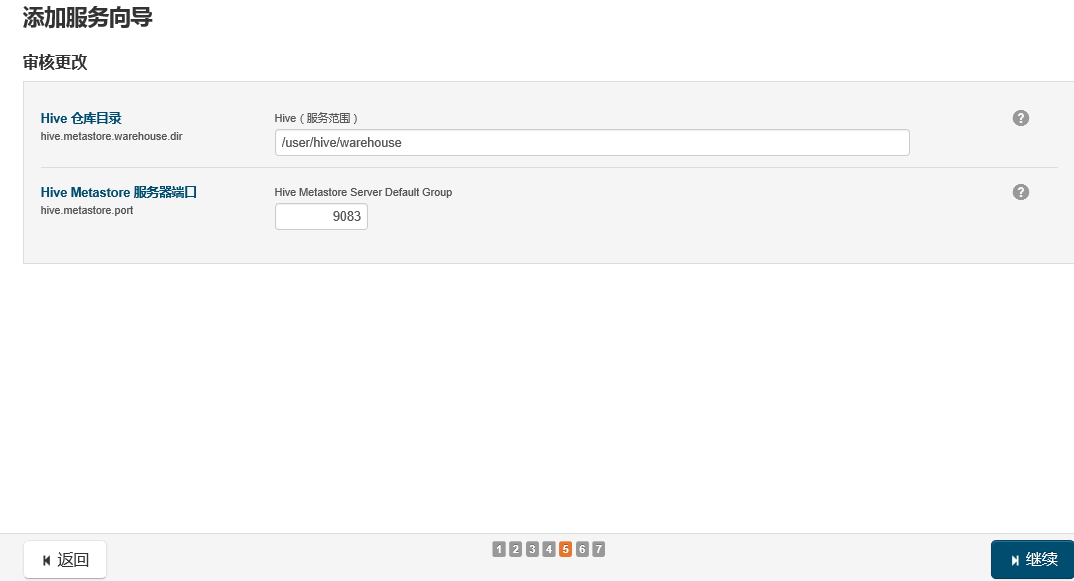
点击继续：



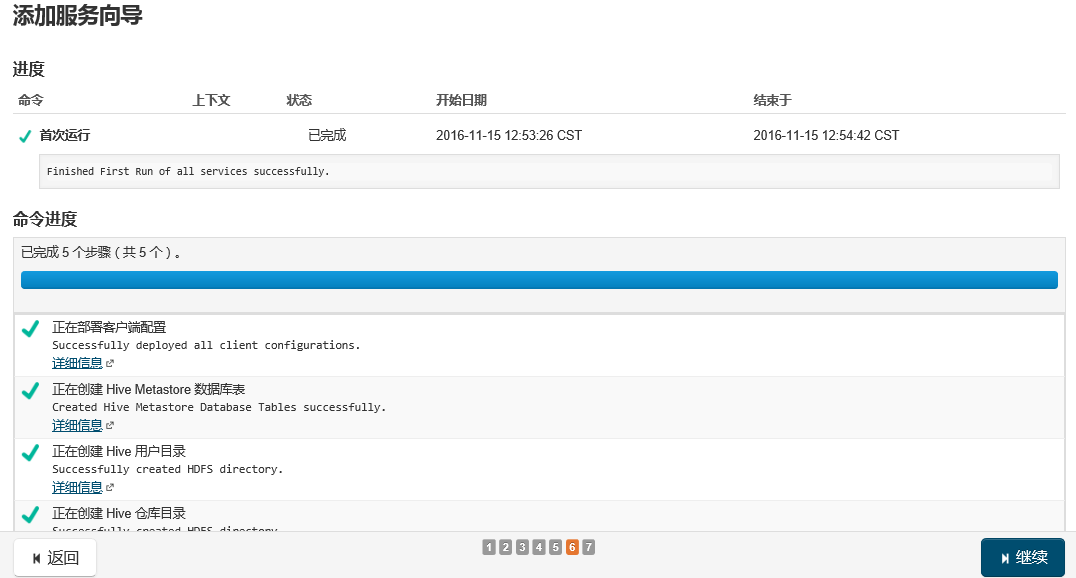
点击继续：



选择使用自定义的数据库，就是前面第3章节安装的mysql数据库，测试连接，点击继续：



点击继续：



1. Hive服务验证

首先准备测试数据：

groupadd eda

useradd eda -g eda

su - eda

vi test.txt

1

2

3

hadoop fs -mkdir /user/eda/test

hadoop fs -put test.txt /user/eda/test

使用beeline连接hive：

/opt/cloudera/parcels/CDH/lib/hive/bin/beeline -u "jdbc:hive2://192.168.129.106:10000/default" -nhive -phive；

查看数据库：show databases;

创建数据库：create database db\_test;

切换数据库：use db\_test;

创建测试表：create table test(id int);

查看表清单：show tables;

加载测试数据：load data inpath '/user/eda/test/test.txt' into table test;

测试MR任务：select count(1) from test;

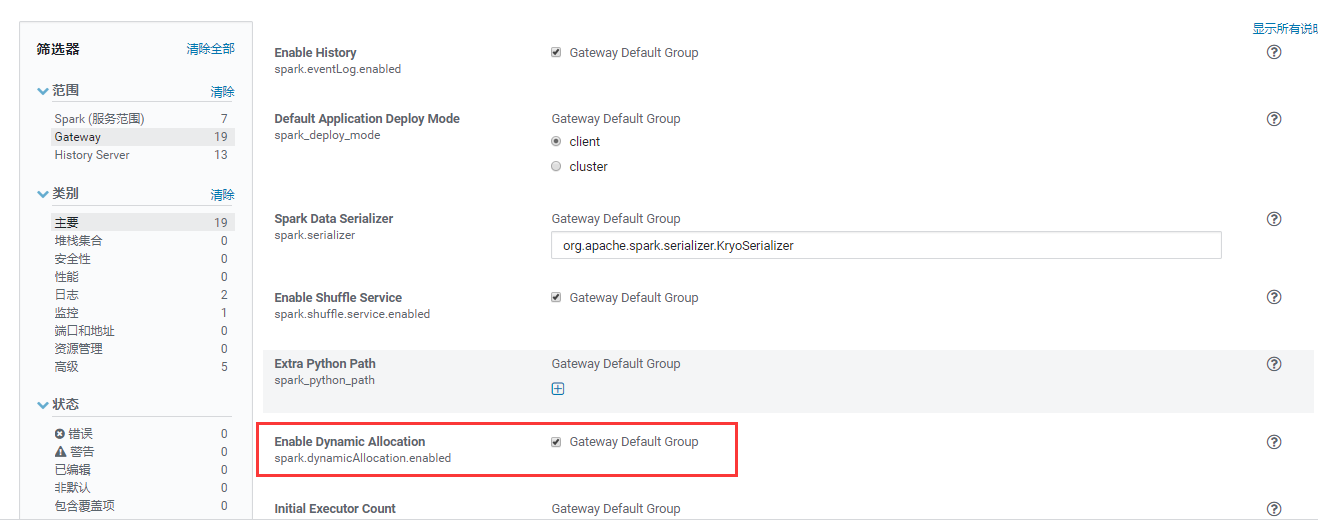
测试插入数据：create table test2(id int);

insert into test2 select \* from test;

select \* from test2;

测试删除表：drop table test;

修改HiveServer2的动态分配配置：

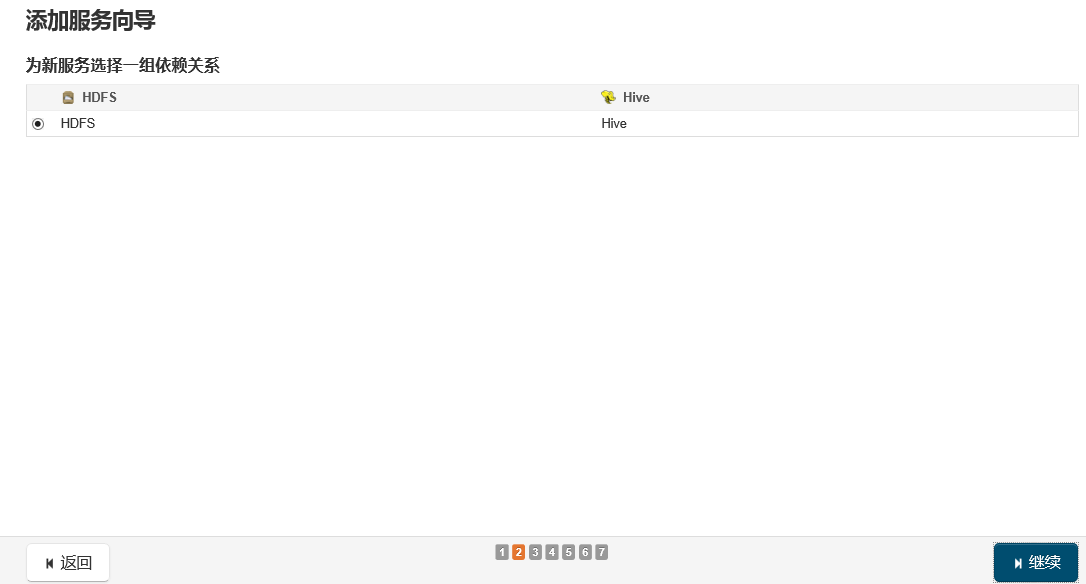


将该选项去掉

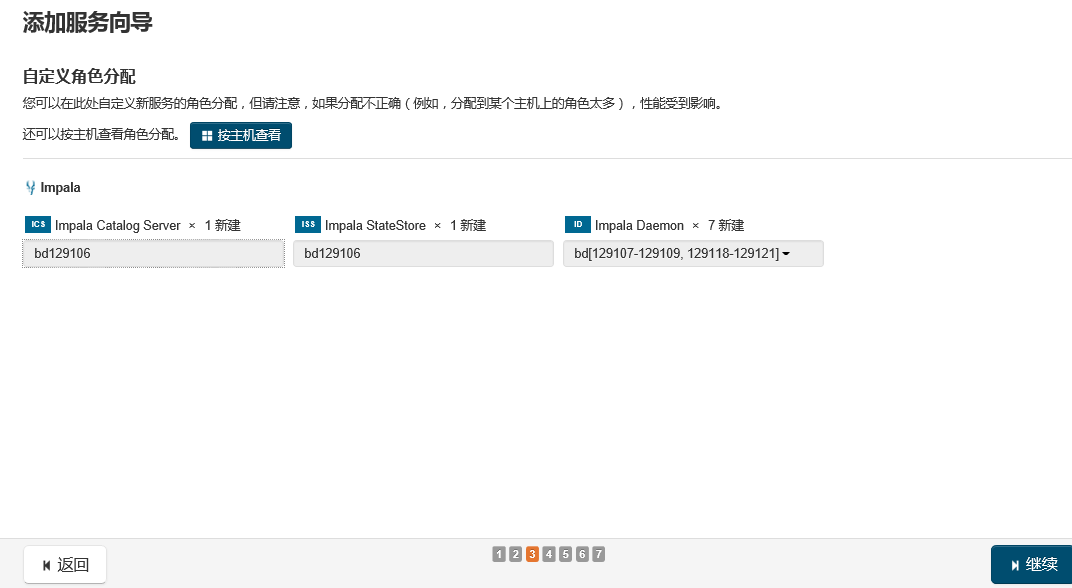
## 5添加Impala服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择impala

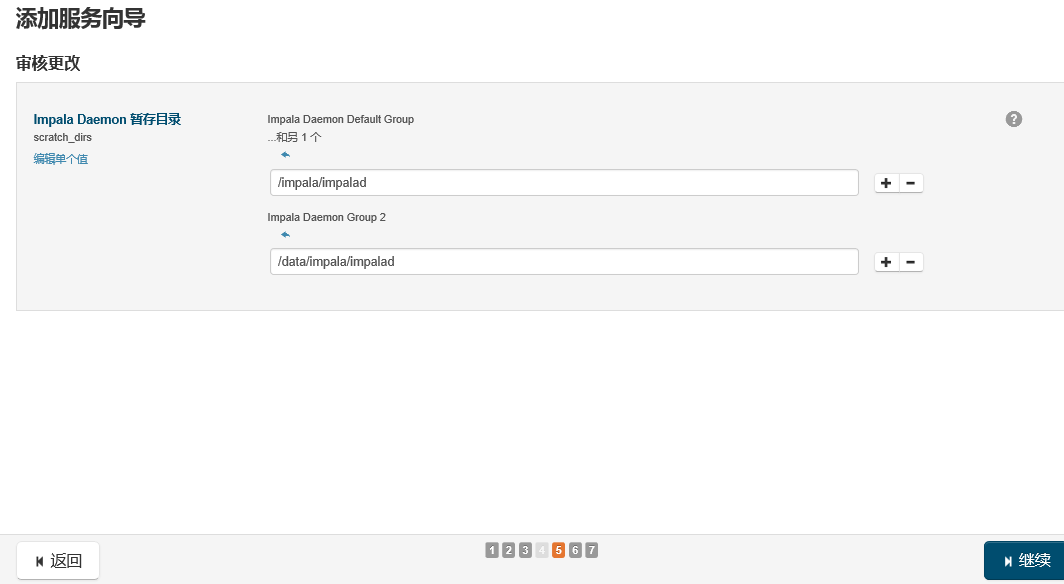
添加依赖关系：

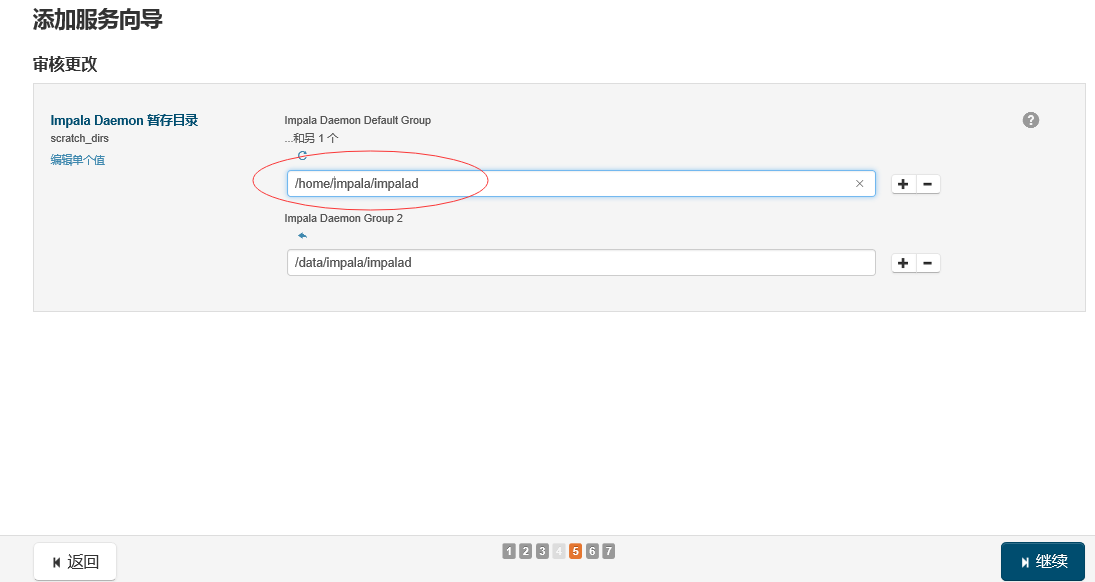


点击继续：

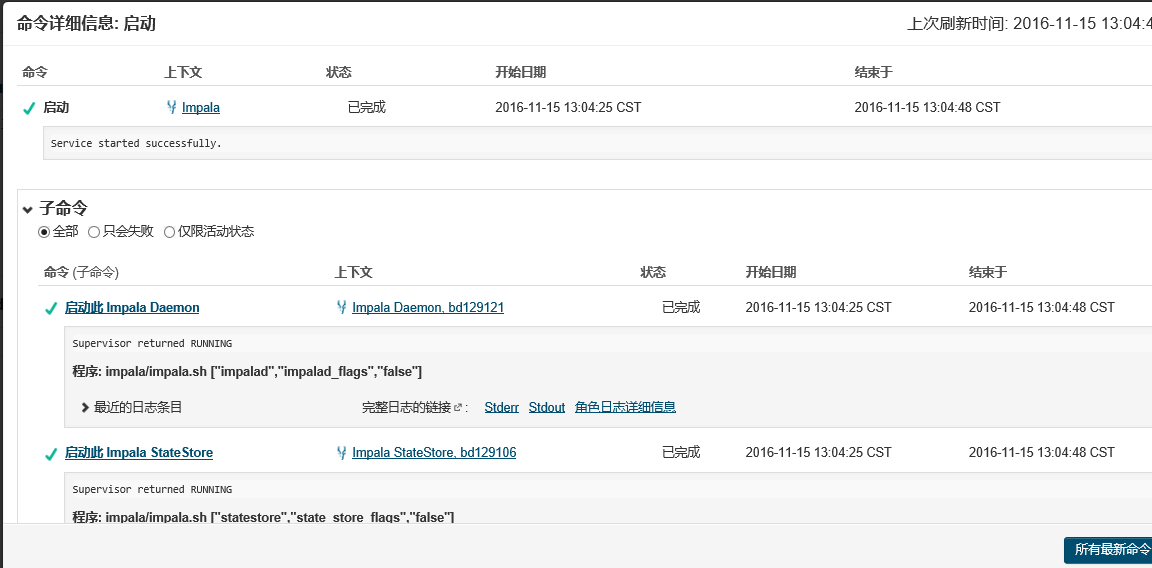


点击继续：





修改impala数据目录到home下面，点击继续：



修改impala的内存配置（根据主机实际内存调整）：



1. Impala服务验证

impala-shell

connect bd129107:21000;

invalidate metadata;

show databases;

use db\_test;

create table test(id int);

show tables;

select count(1) from test;

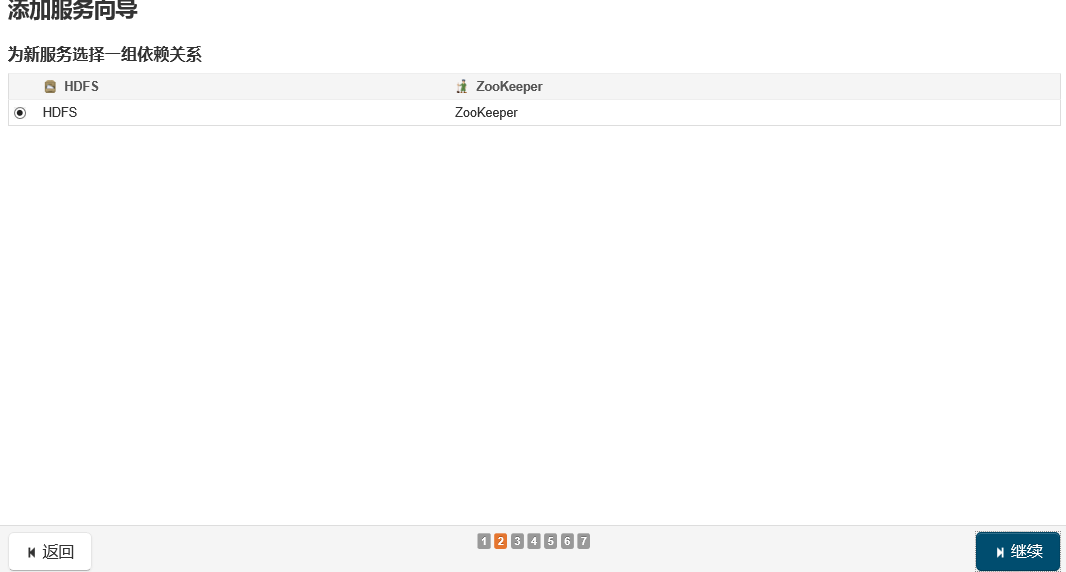
drop table test;

exit;

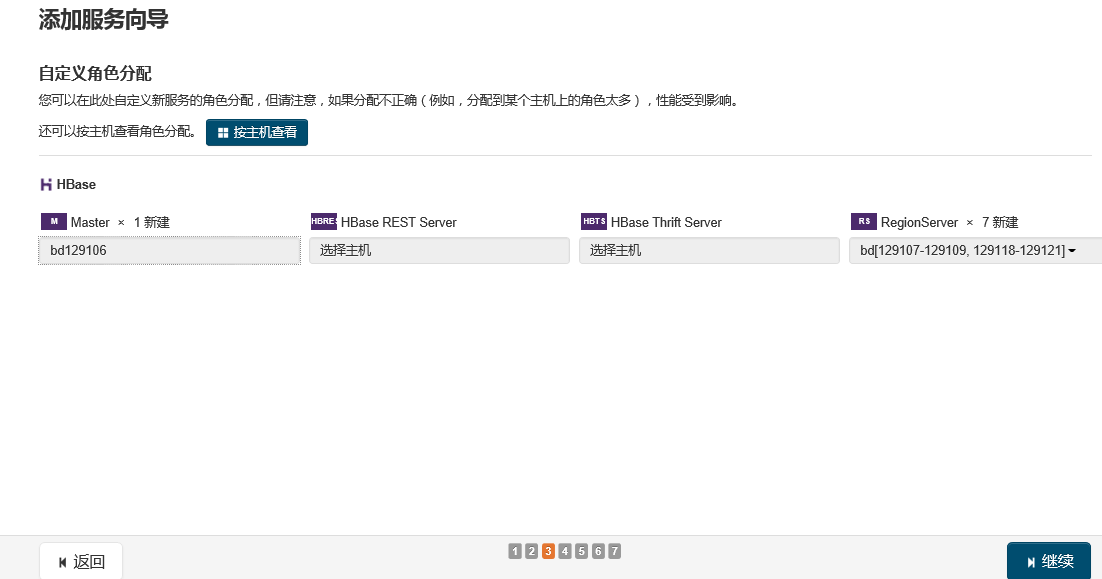
## 6添加Hbase服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择hbase

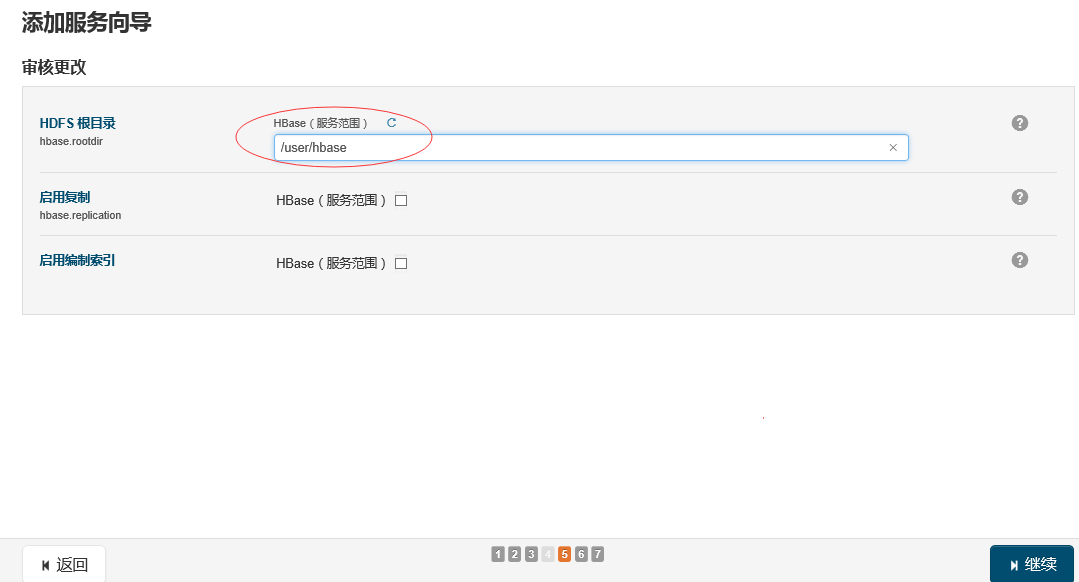
添加依赖关系：



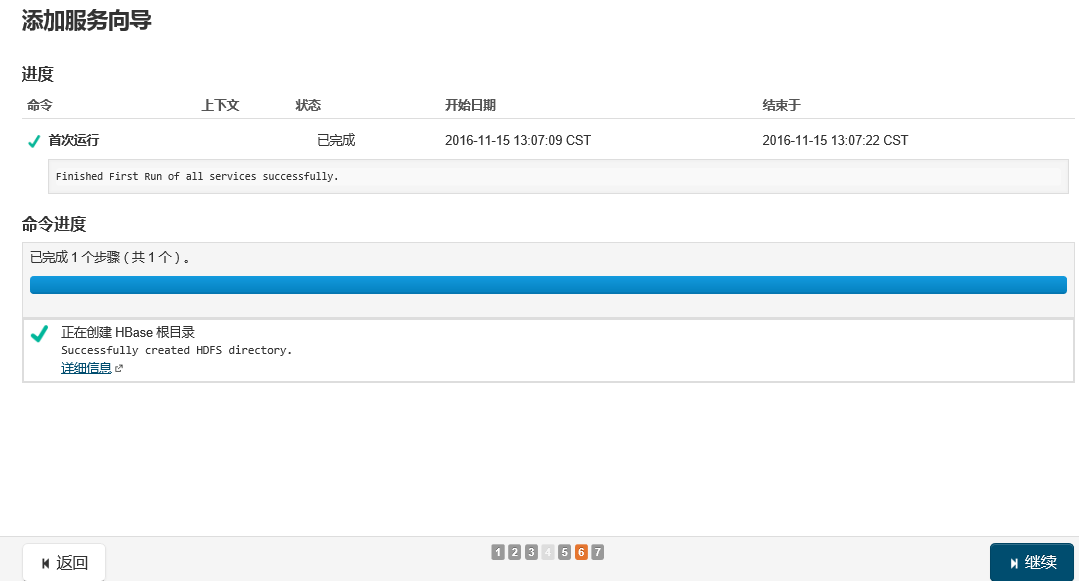
点击继续：



点击继续：



修改hbase根目录到/user/hbase，点击继续：



修改hbase regionserver的内存配置（根据主机实际内存调整）：



1. Hbase服务验证

hbase shell

list

create 'test','a'

put 'test','00001','a:v01','11111'

scan 'test'

## 7添加Oozie服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择Oozie添加安装即可

安装过程中需要输入前面mysql配置的Oozie数据库用户名和密码

## 8添加HUE服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择hue

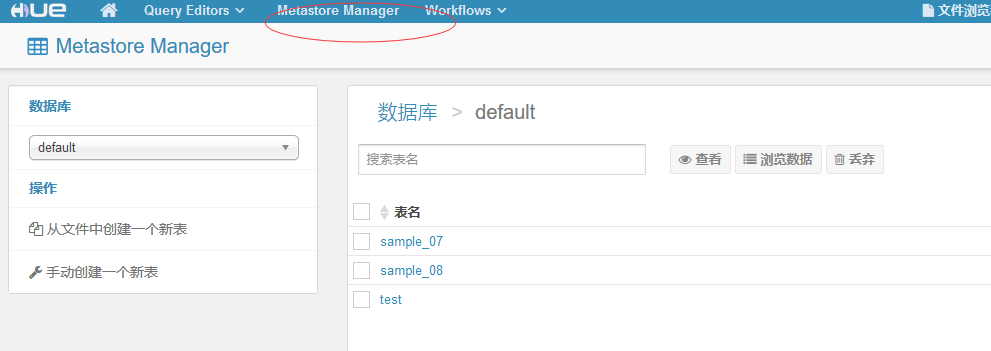
2、登录hue界面创建超级管理员帐号：



这里我们创建超级管理员帐号和密码为：hive/hive

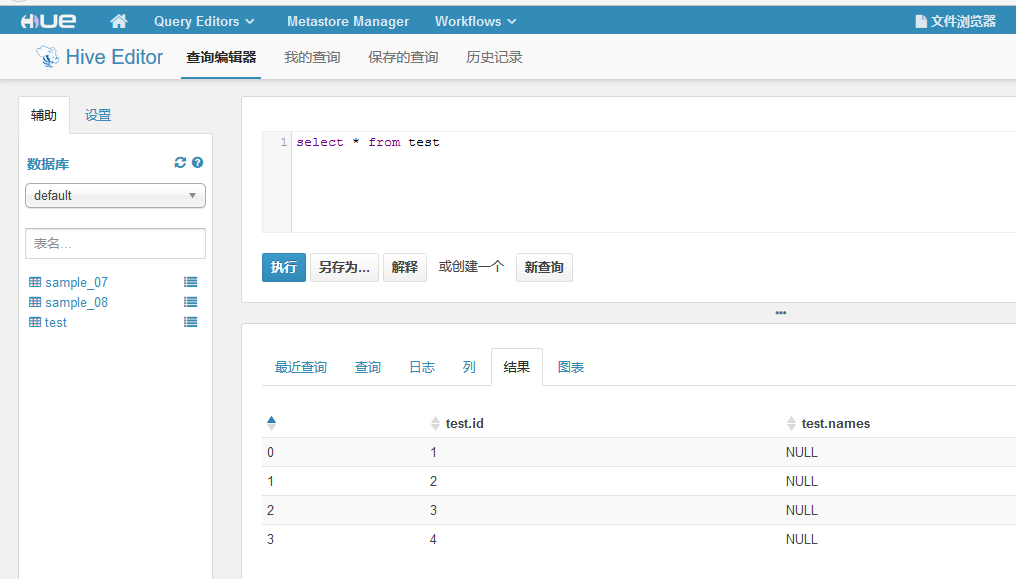


点击metastore manager可以查看表信息



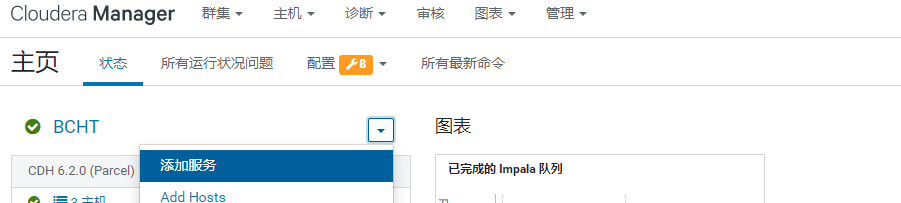
点击query editors可以输入sql语句查询hive表



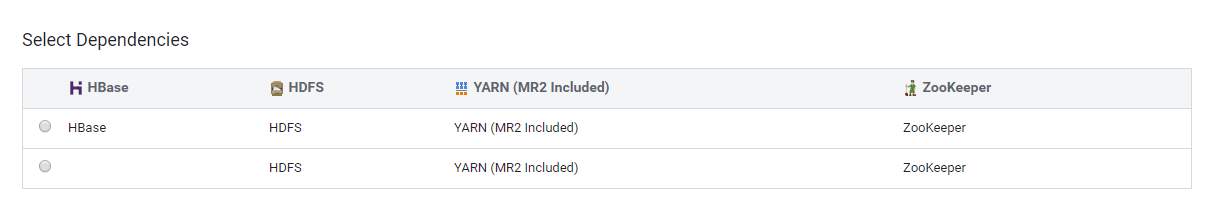


## 9添加Spark服务

1、登录CM界面，添加服务――下拉“操作”――添加服务――选择spark



选择spark服务，然后配置相应的服务节点



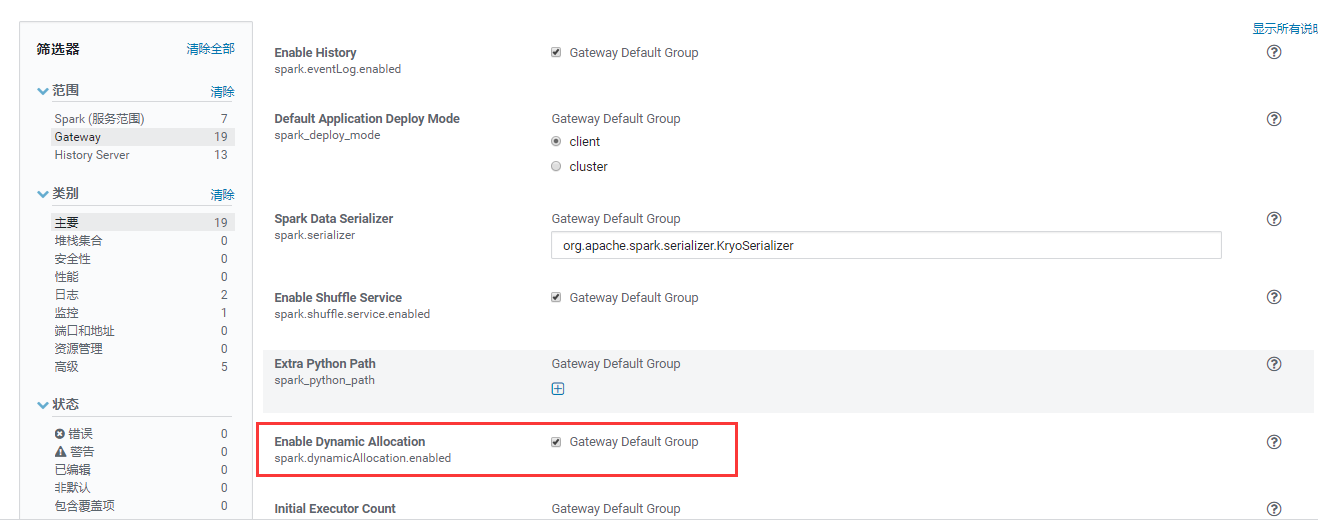
Spark服务要依赖CDH的Zookeeper、HDFS、YARN

确定后，一路继续下去既可，等待安装完成既可！

2、修改CDH6.2中Spark的动态资源分配配置（Hive默认用Spark作为引擎时）

spark.dynamicallocation.enabled=true 这个配置会导致，spark-submit提交的任务中，资源配置失效，因为会配置的是自动进行分配，所以要将配置改为false

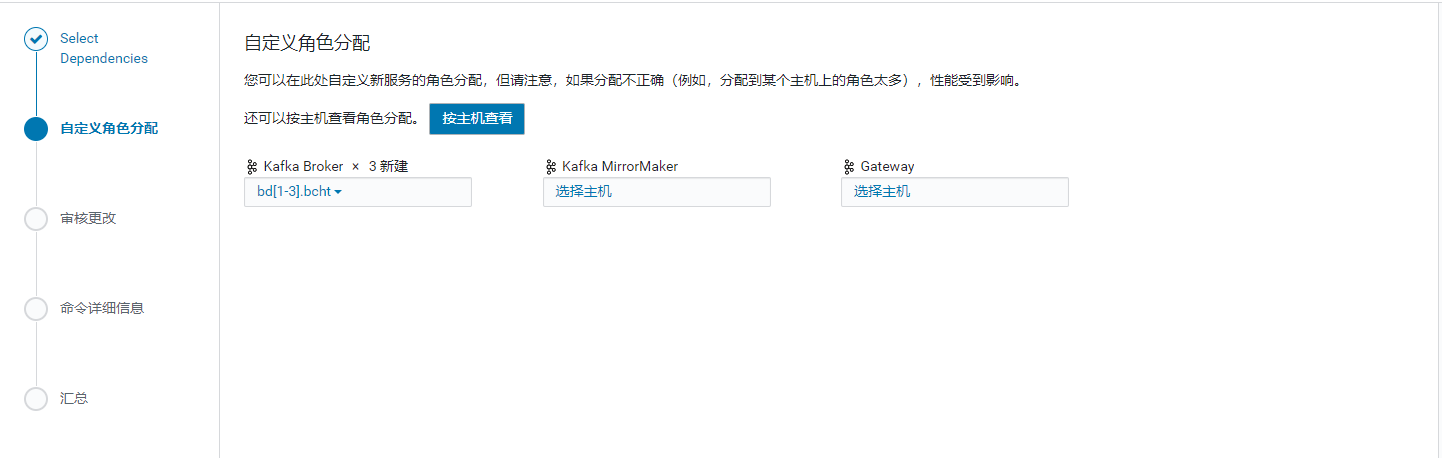
该配置位于Hive的HiveServer2服务中



去掉勾既可！

## 10添加Kafka服务

1. 登录CM界面，添加服务――下拉“操作”――添加服务――选择spark
2. 确定服务角色分配



配置使用默认配置既可，点击继续完成安装。

# 问题记录

yarn 问题记录

1、userid 选项默认1000 需要修改为100

2、yarn运行MR任务报错问题：

本地目录 ，在启用kerberos前如果运行过MR任务，则启用kerberos后，目录权限要求就不一样了，这时可以删除本地目录，让cm重新自动创建即可。

3、nifi目前采用独立安装的方式，不与CM集成，所以CFM安装步骤可以省略。

<https://blog.csdn.net/qq_41979143/article/details/107850970>

[https://www.cnblogs.com/swordfall/p/10816797.html#auto\_id\_21](https://www.cnblogs.com/swordfall/p/10816797.html" \l "auto_id_21) 主要

<https://blog.csdn.net/weixin_34205076/article/details/91893636>

装http服务，把rpm包放上去，其他节点使用这个yum源

agent页面安装失败，用yum命令安装

安装过程注意磁盘空间充足

安装parcel，设置url，使用本地http地址，官网地址下载太慢

装agent可以跳过

Scm服务启动比较慢