**Cloudera部署手册**

概述

Cloudera 企业级数据中心的安装主要分为4 个步骤：

1. 集群服务器配置，包括安装操作系统、关闭防火墙、同步服务器时钟等；

2. 外部数据库安装

3. 安装Cloudera 管理器；

4. 安装CDH 集群；

5. 集群完整性检查，包括HDFS文件系统、MapReduce、Hive 等是否可以正常运行。

这篇文档将着重介绍Cloudera 管理器与CDH 的安装，并基于以下假设：

1. 操作系统版本：CentOS 6/RHEL 6

2. CM 版本：CM 5.x

3. CDH 版本：CDH 5.x

4. 采用root 对集群进行部署

5. 您已经下载CDH 和CM 的安装包

下载地址

CDH 下载地址：<http://archive-primary.cloudera.com/cdh5/parcels/5.7.0.45/>

CDH-5.7.0-1.cdh5.7.0.p0.45-el6.parcel

CDH-5.7.0-1.cdh5.7.0.p0.45-el6.parcel.sha1

manifest.json

CM 下载地址：

<http://archive-primary.cloudera.com/cm5/redhat/6/x86_64/cm/5.7.0/RPMS/x86_64/>

cloudera-manager-agent-5.7.0-1.cm570.p0.76.el6.x86\_64.rpm

cloudera-manager-daemons-5.7.0-1.cm570.p0.76.el6.x86\_64.rpm

cloudera-manager-server-5.7.0-1.cm570.p0.76.el6.x86\_64.rpm

cloudera-manager-server-db-2-5.7.0-1.cm570.p0.76.el6.x86\_64.rpm

enterprise-debuginfo-5.7.0-1.cm570.p0.76.el6.x86\_64.rpm

jdk-6u31-linux-amd64.rpm

oracle-j2sdk1.7-1.7.0+update67-1.x86\_64.rpm

Kafka下载地址：<http://archive-primary.cloudera.com/kafka/parcels/latest/>

KAFKA-2.0.1-1.2.0.1.p0.5-el6.parcel

KAFKA-2.0.1-1.2.0.1.p0.5-el6.parcel.sha1

manifest.json

**一、服务器配置**

1. 安装操作系统，建议对操作系统盘做RAID 1

2. 如果不能连接互联网，先创建OS的repository，以便yum 或zypper 可以访问OS 镜像

3. 为了使集群中各个节点之间能互相通信，需要静态配置节点的IP 地址，请正确配置各节点的IP ，并在/etc/hosts 配置所有节点的静态DNS 解析，并修改主机名。

vi /etc/sysconfig/network -scripts/ifcfg -eth0

DEVICE=eth0

ONBOOT=yes

BOOTPROTO=static

IPADDR=192.168.204.32

GATEWAY=192.168.204.254

NETMASK=255.255.255.0

vi /etc/hosts

192.168.204.32 osp-c01

192.168.204.42 osp-c02

192.168.204.46 osp-c03

192.168.204.47 osp-c04

192.168.204.48 osp-c05

192.168.204.49 osp-c06

192.168.204.50 osp-c07

192.168.204.53 osp-c08

vi /etc/sysconfig/network

HOSTNAME= osp-c01 // 将hostname修改为对应的机器名，机器名中不能有下划线

4. 如果机器配置有双网卡，可以做双网卡绑定；

5. 关闭并禁用iptables

$> service iptables stop

$> chkconfig iptables off

6. 关闭SELinux

$> echo "SELINUX=disabled" > /etc/sysconfig/selinux ; //永久关闭，需要重启

$> setenforce 0 //临时关闭

7. 重启网络服务，并初始化网络

$> service network restart

8. 修改transparent \_hugepage 参数，这一参数默认值可能会导致CDH 性能下降

RHEL/CentOS

# 在/etc/init.d/rc.local中增加一行：

$> echo never > /sys/kernel/mm/redhat\_transpa rent\_hugepage/defrag

9. 禁止交换（可选），内存页面交换在某些情况下会导致CDH 性能下降，建议在kerne 在

2.6.32 -303 之前的版本中将其关闭。

$> vi /etc/sysctl.conf

增加一行：vm.swappiness=0

$> sysctl –p //立即生效

1. （可选）修改/etc/security/limits.conf 下增加相应的配置文件，

vi /etc/security/limits.conf

\* soft nofile 65536

\* hard nofile 65536

\* soft nproc 102400

\* hard nproc 102400

vi /etc/pam.d/login

session required /lib64/security/pam\_limits.so

注销该用户，重新登录使用ulimit -a查看

11. ntp对时服务

集群所有机器修改 /etc/ntp.conf

vi /etc/ntp.conf

服务端

restrict default kod nomodify

server 127.127.1.0 # local clock

fudge 127.127.1.0 stratum 10

执行命令

$>/etc/init.d/ntpd start //打开ntp服务

$>chkconfig ntpd on //设置开机自动启动

客户端

server 192.168.204.32

restrict 192.168.204.32 mask 255.255.255.0 nomodify notrap

在ntp服务端启动几分钟以后首先执行以下命令手动对时一把

$>ntpdate -u 192.168.204.32

所有ntp客服端执行如下命令

$>/etc/init.d/ntpd start //打开ntp服务

$>chkconfig ntpd on //设置开机自动启动

12. 从linux安装光盘的根目录下面的Packages目录中找到如下rpm包：

fuse-2.8.3-4.el6.x86\_64.rpm

fuse-libs-2.8.3-4.el6.x86\_64.rpm

deltarpm-3.5-0.5.20090913git.el6.x86\_64.rpm python-deltarpm-3.5-0.5.20090913git.el6.x86\_64.rpm

createrepo-0.9.9-22.el6.noarch.rpm

httpd-tools-2.2.15-45.el6.x86\_64.rpm

apr-util-ldap-1.3.9-3.el6\_0.1.x86\_64.rpm

httpd-2.2.15-45.el6.x86\_64.rpm

mysql-server-5.1.73-5.el6\_6.x86\_64.rpm

mysql-5.1.73-5.el6\_6.x86\_64.rpm

mysql-connector-java-5.1.17-6.el6.noarch.rpm

13. 在需要作为Repo库的节点上安装必要的软件，包含HTTP服务和Repo创建工具。

所有机器安装fuse和fuse-libs

$>rpm -ivh fuse-2.8.3-4.el6.x86\_64.rpm fuse-libs-2.8.3-4.el6.x86\_64.rpm

作为源的机器安装 createrepo

$>rpm -ivh deltarpm-3.5-0.5.20090913git.el6.x86\_64.rpm python-deltarpm-3.5-0.5.20090913git.el6.x86\_64.rpm createrepo-0.9.9-22.el6.noarch.rpm

作为源的机器安装 httpd

$>rpm -ivh httpd-tools-2.2.15-45.el6.x86\_64.rpm apr-util-ldap-1.3.9-3.el6\_0.1.x86\_64.rpm httpd-2.2.15-45.el6.x86\_64.rpm

$>service httpd start //启动http服务

$>chkconfig httpd on //设置开机自启动

14. 外部数据库安装并配置，Cloudera 5 推荐安装MySQL。

$>rpm -ivh mysql-server-5.1.73-5.el6\_6.x86\_64.rpm mysql-5.1.73-5.el6\_6.x86\_64.rpm

$>chkconfig mysqld on //启动mysql服务

$>service mysqld start //设置开机自启动

mysql root用户配置

$ /usr/bin/mysql\_secure\_installation

[...]

Enter current password for root (enter for none):

OK, successfully used password, moving on...

[...]

Set root password? [Y/n] y

New password: 123456

Re- enter new password:

Remove anonymous users? [Y/n] Y

[...]

Disallow root login remotely? [Y/n] N

[...]

Remove test database and access to it [Y/n] Y

[...]

Reload privilege tables now? [Y/n] Y

All done!

为cloudera manager创建用户，在终端中执行如下命令：

mysql -u root --password='123456' -e "create user 'hive'@'%' identified by '123456'"

mysql -u root --password='123456' -e 'create database metastore default character set utf8;'

mysql -u root --password='123456' -e "grant all privileges on metastore.\* to 'hive'@'%'"

mysql -u root --password='123456' -e "create user 'amon'@'%' identified by '123456'"

mysql -u root --password='123456' -e 'create database amon default character set utf8'

mysql -u root --password='123456' -e "grant all privileges on amon.\* to 'amon'@'%'"

mysql -u root --password='123456' -e "create user 'rman'@'%' identified by '123456'"

mysql -u root --password='123456' -e 'create database rman default character set utf8'

mysql -u root --password='123456' -e "grant all privileges on rman.\* to 'rman'@'%'"

mysql -u root --password='123456' -e "create user 'sentry'@'%' identified by '123456'"

mysql -u root --password='123456' -e 'create database sentry default character set utf8'

mysql -u root --password='123456' -e "grant all privileges on sentry.\* to 'sentry'@'%'"

mysql -u root --password='123456' -e "create user 'nav'@'%' identified by '123456'"

mysql -u root --password='123456' -e 'create database nav default character set utf8'

mysql -u root --password='123456' -e "grant all privileges on nav.\* to 'nav'@'%'"

mysql -u root --password='123456' -e "create user 'navms'@'%' identified by '123456'"

mysql -u root --password='123456' -e 'create database navms default character set utf8'

mysql -u root --password='123456' -e "grant all privileges on navms.\* to 'navms'@'%'"

15. 在作为软件源的机器上做如下操作。

$>mkdir -p /var/www/html/cdh5/parcels/5.7.0/

将CDH 的三个文件放到/var/www/html/cdh5/parcels/5.7.0/cdh目录中

将Kafka的三个文件放到/var/www/html/cdh5/parcels/5.7.0/kafka

将CM的七个文件和RedHat的Packages目录中找到rpm包放到/var/www/html/cm

$>chmod -R ugo+rX /var/www/html

在/var/www/html/cm目录中执行createrepo .

集群中所有节点设置myrepo.repo

$>vi /etc/yum.repos.d/myrepo.repo

[myrepo]

name=repo

baseurl=http://192.168.204.32/cm/

enabled=true

gpgcheck=false

16. 所有机器全部重启一遍，便于机器名的设置生效。

**二、Cloudera Manager安装**

1. 在作为cm服务端的机器上安装cloudera-manager-server。

$>yum clean all

$>yum makecache

$>yum install java cloudera-manager-daemons cloudera-manager-server mysql-connector-java

//所有机器安装mysql的jdbc驱动

$>yum install mysql-connector-java

//所有机器安装java

$>yum install oracle-j2sdk1.7

update-alternatives --install /usr/bin/java java /usr/java/jdk1.7.0\_67-cloudera/bin/java 60

update-alternatives --config java

There are 4 programs which provide 'java'.

Selection Command

-----------------------------------------------

1 /usr/lib/jvm/jre-1.5.0-gcj/bin/java

2 /usr/lib/jvm/jre-1.8.0-openjdk.x86\_64/bin/java

+ 3 /usr/java/jdk1.7.0\_67-cloudera/bin/java

\* 4 /usr/lib/jvm/jre-1.7.0-openjdk.x86\_64/bin/java

Enter to keep the current selection[+], or type selection number:

选择jdk1.7.0\_67-cloudera作为默认jdk。

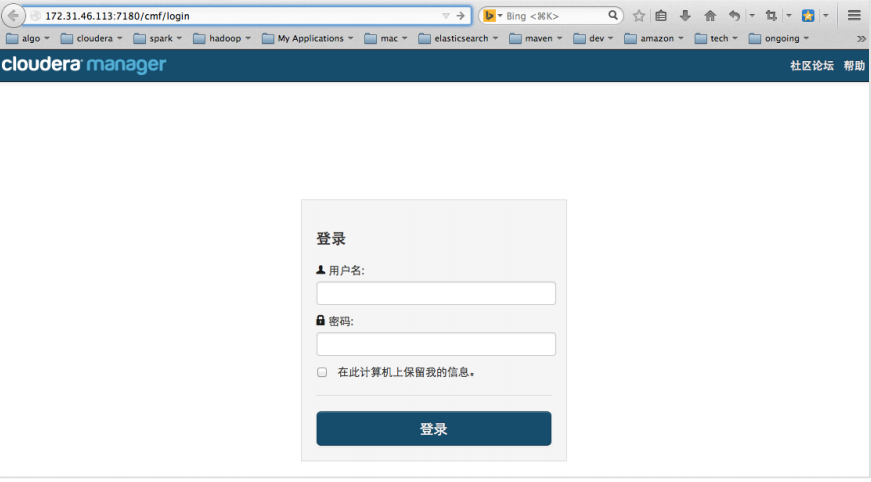
2. 为cloudera manager准备数据库的环境

$>/usr/share/cmf/schema/scm\_prepare\_database.sh mysql cm -hlocalhost -uroot -p123456 --scm-host localhost scm scm scm

3. 启动cloudera-scm-server

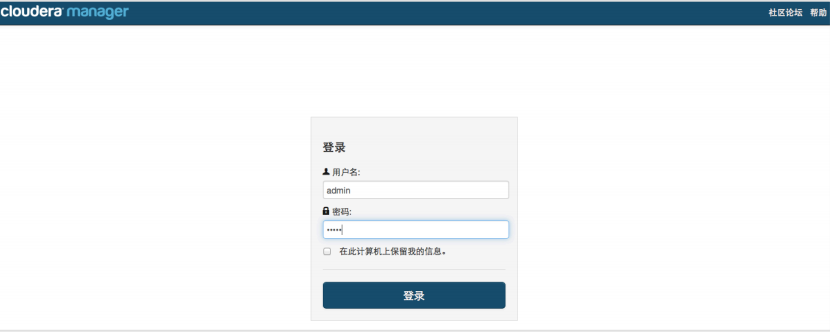
$>service cloudera-scm-server start

执行启动命令过几分钟后在浏览器中输入<http://192.168.204.32:7180>， 该ip地址为cloudera-scm-server安装的机器的ip地址，一直刷新直到见到如下界面：



**三、CDH安装**

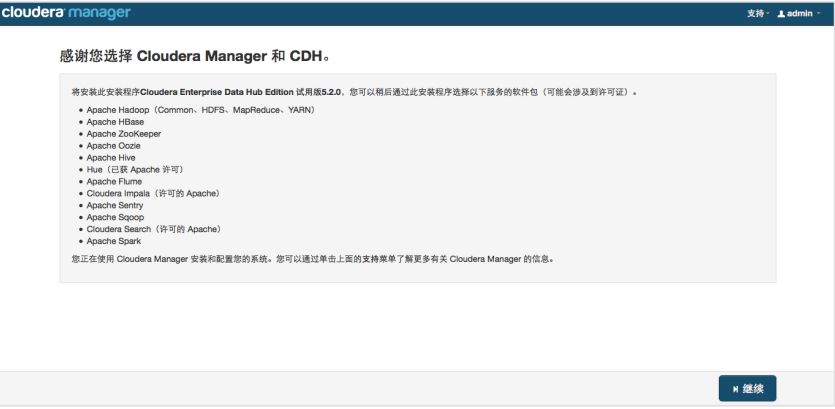
1. 输入账户密码admin/admin ，点击“登录”



2. 选择要安装的集群版本（在这里我们选择试用版），点击“继续”



3. 了解CDH 支持的Hadoop组件信息，点击“继续”

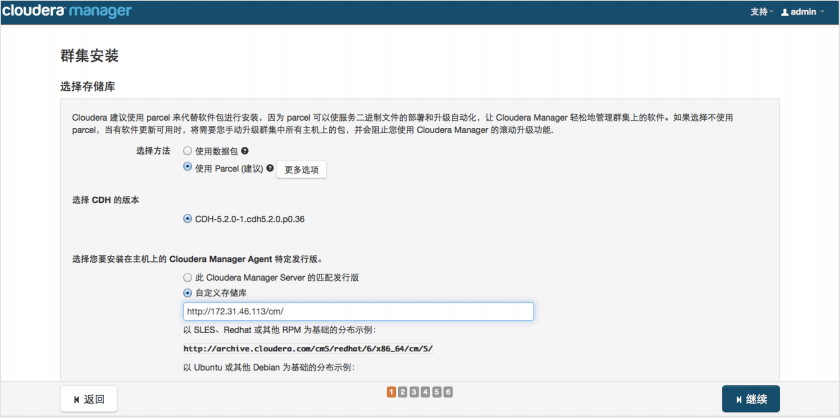


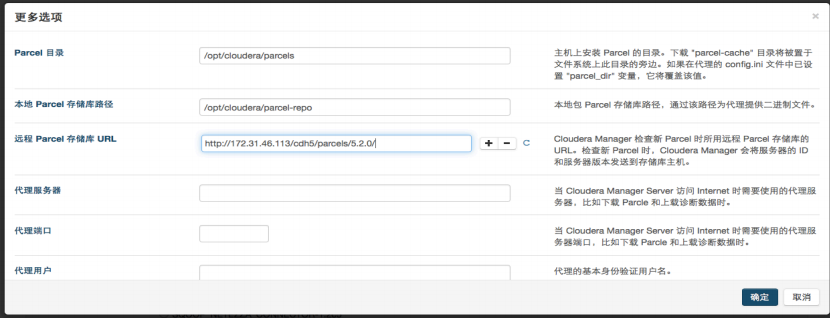
4. 查找并选择需要安装CDH 的机器，填写ip地址，ip地址之间以分号作为分隔符

点击“继续”



5. 点击“使用Parcels （建议）”右侧的“更多选项”按钮，在弹出框中设置CDH Parcel包的URL，点击“确定”



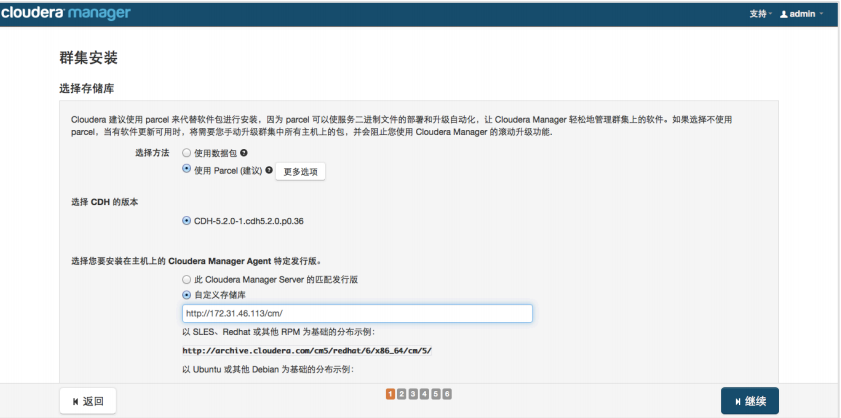


远程Parecel存储库URL改为

http:// 192.168.204.32/cdh5/parcels/5.7.0/cdh/

http:// 192.168.204.32/cdh5/parcels/5.7.0/kafka/

6. 选择CDH -5.2.0-1.cdh5.2.0.p0.36 （根据实际下载CDH版本选择），在“自定义存储库”填写依赖包的URL，点击“继续”

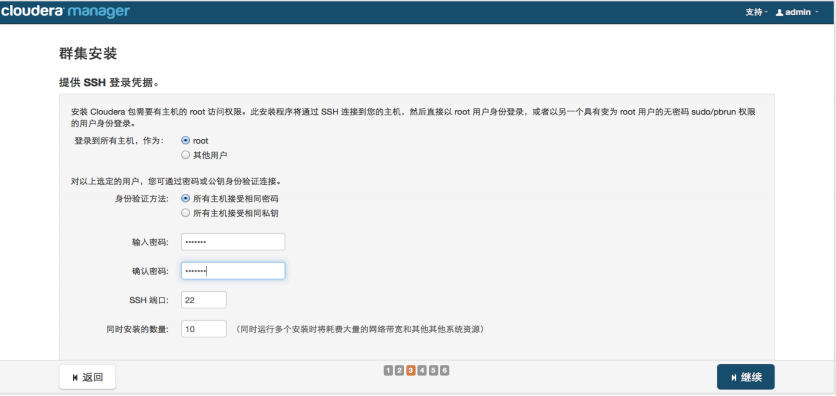


自定义存储库URL为http:// 192.168.204.32/cm/

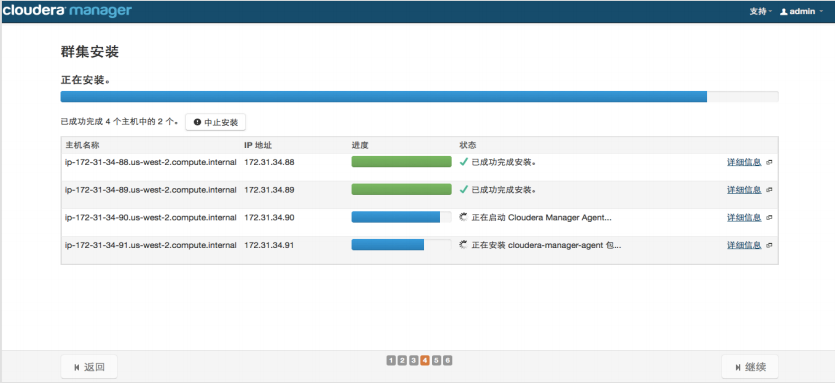
7. 选择需要的JDK ，点击“继续”



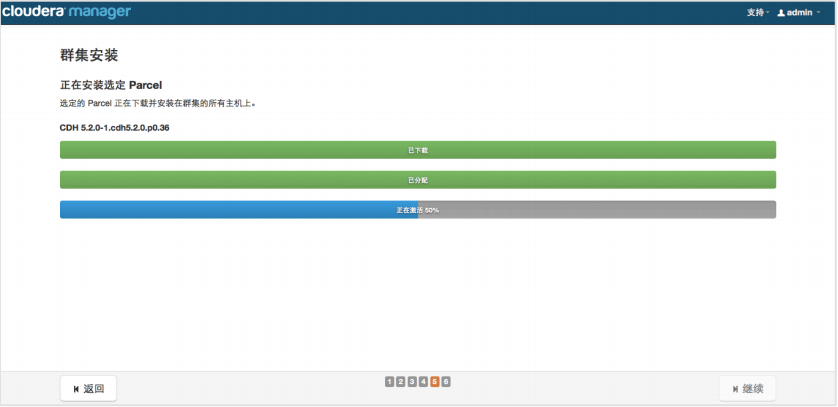
8. 输入集群机器的登录密码，点击“继续”



9. 集群依赖包安装，安装完后点击“继续”



10. Parcel 包安装，安装完后点击“继续”

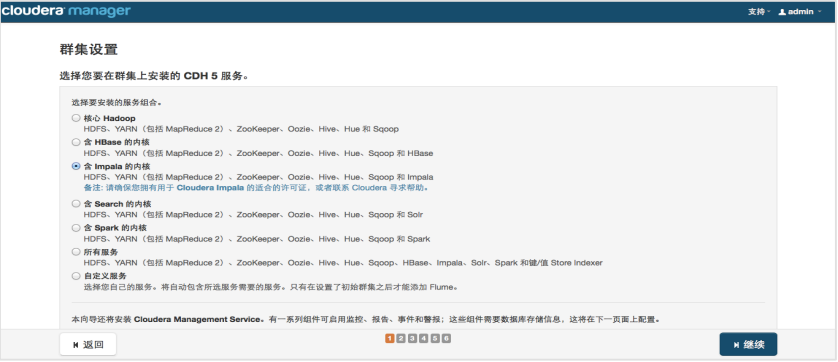


11. 检查主机正确性，如果检查出现任何潜在问题，你可以到集群中进行修复，修复后点

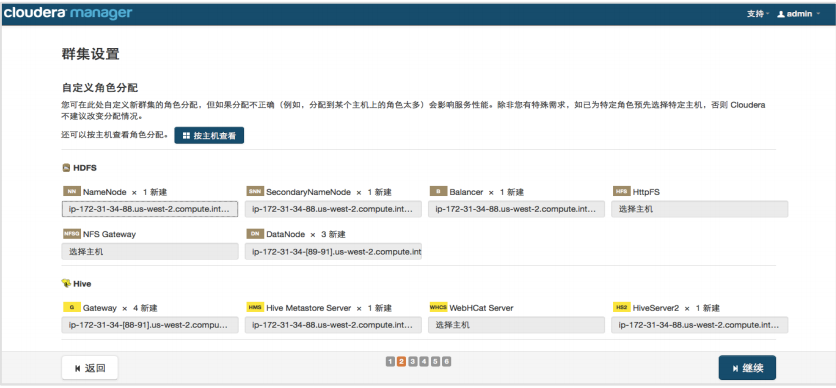
击“重新运行”重新检查。解决所有问题后，点击“完成”



12. 选择要安装的服务套装，点击“继续”

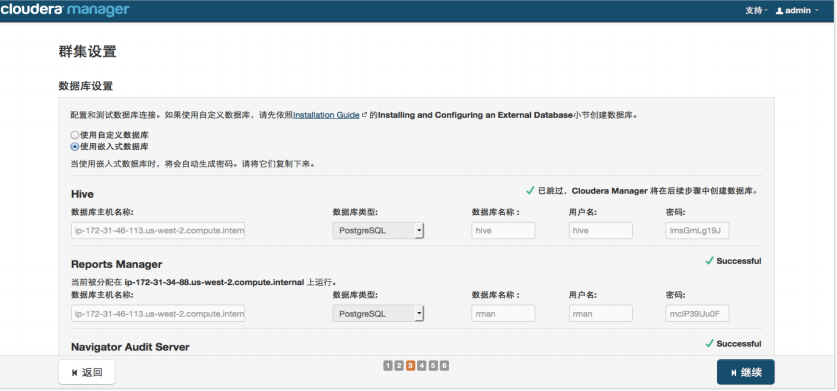


13. 选择具体的角色分配，我们建议你直接使用Cloudera 设置的默认值，点击“继续”

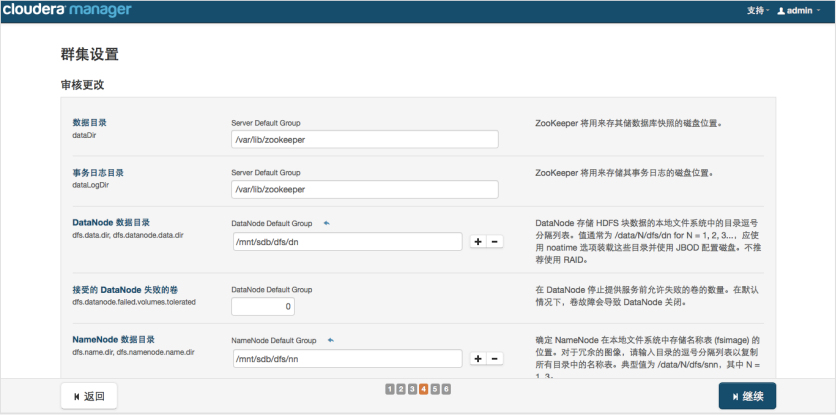


尽量将各个服务分散到集群中，不能将服务集中到某台机器中。

14. 设置数据库，设置完毕后点击“测试连接”，测试全部通过后点击“继续”



15. 配置集群组件的相关参数，点击“继续”



16. 启动集群，完成后点击“继续”



17. 点击“完成”



18. 访问集群



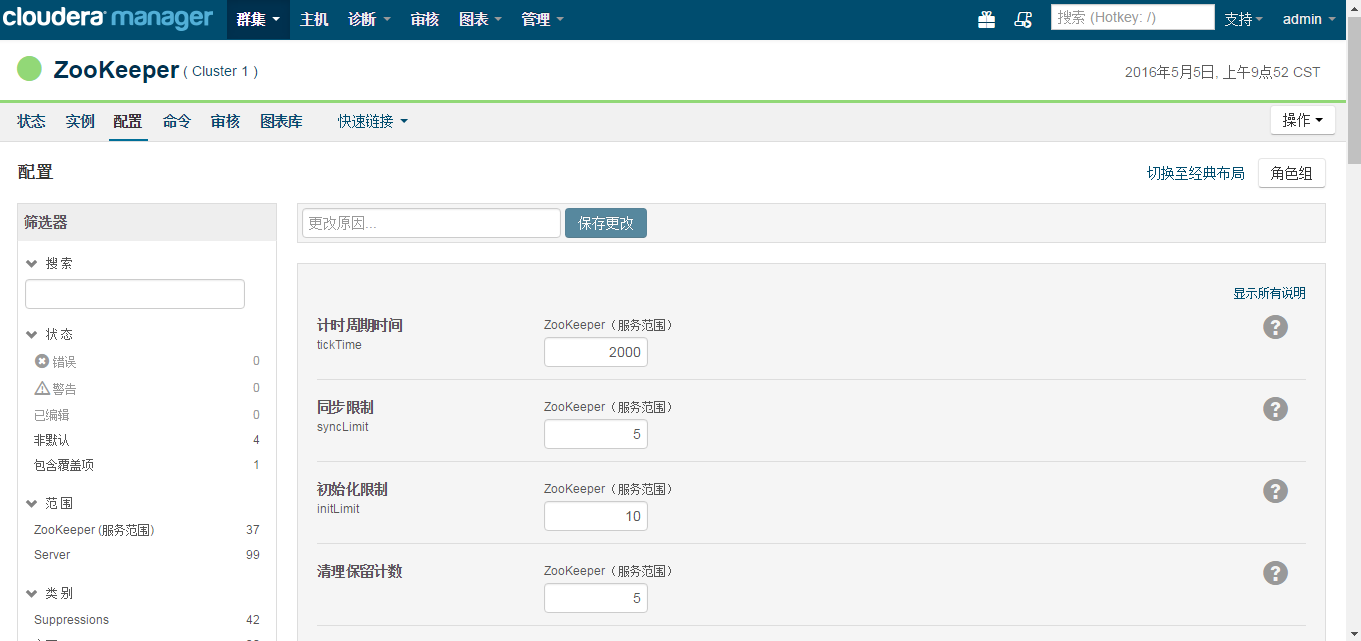
19. 配置集群

HDFS配置



dfs.permissions 设置为false

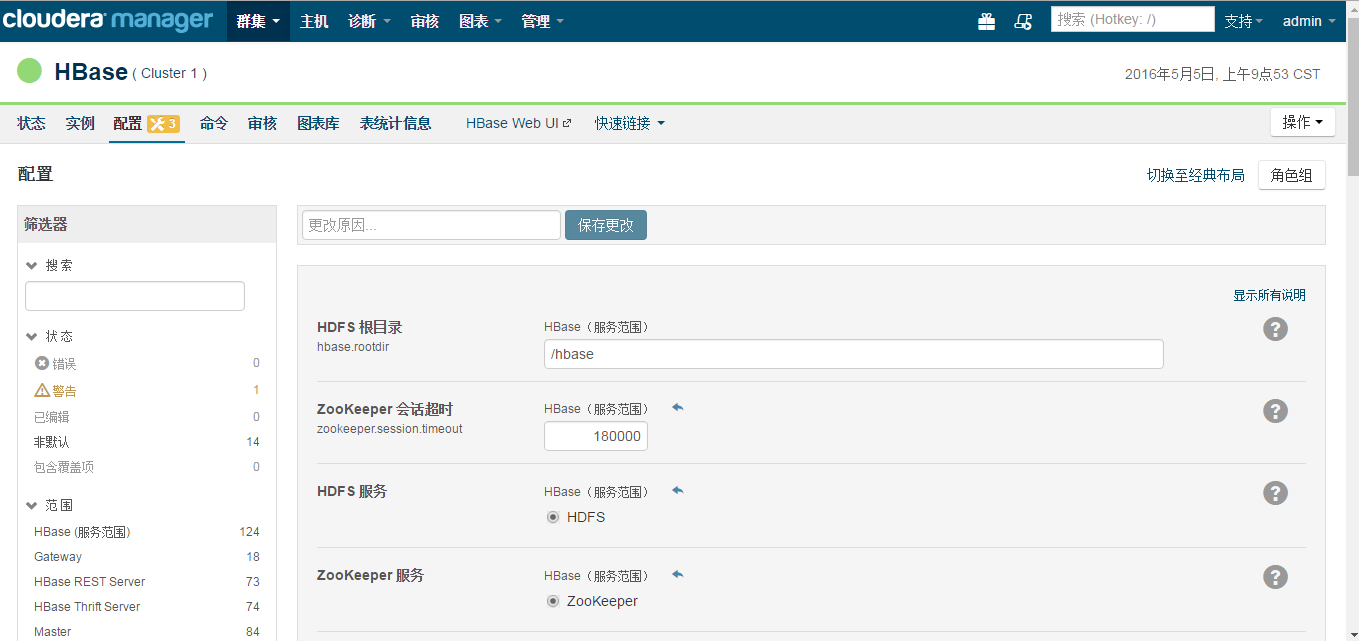
Zookeeper配置



maxClientCnxns 2000

maxSessionTimeout 180000

Hbase配置



zookeeper.session.timeout 180000

hbase.client.write.buffer 32MB

hbase.master.handler.count 200

hbase.regionserver.handler.count 100

hbase.hregion.memstore.block.multiplier 4

hbase.hregion.max.filesize 100GB

hbase.hregion.majorcompaction 0天

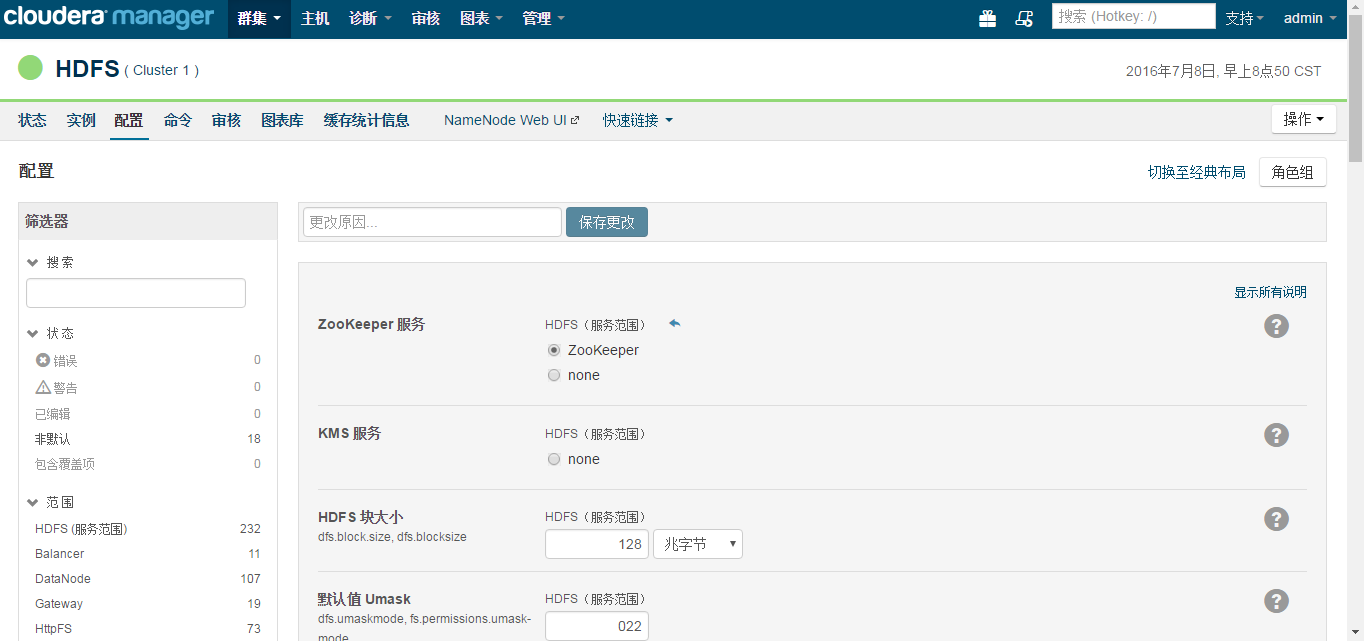
HBase Thrift Server 的 Java 堆栈大小（字节） 100MB

HBase Master 的 Java 堆栈大小（字节） 1GB

HBase RegionServer 的 Java 堆栈大小（字节） 8GB

Hbase.hregion.memstore.flush.size 256MB

HDFS配置



dfs.datanode.max.xcievers, dfs.datanode.max.transfer.threads ?40960

dfs.namenode.handler.count 180

dfs.namenode.service.handler.count 180

dfs.permissions false

配置完成，进行部署客户端配置，并将集群重新启动。





1. Hbase建表

进入一套机器，输入hbase shell，执行如下创建表的语句进行建表操作。

create 'monitor1', 'unit1', {NUMREGIONS => 70, SPLITALGO => 'HexStringSplit', COMPRESSION => 'snappy'}

alter 'monitor1',NAME => 'unit1',COMPRESSION => 'snappy'

disable 'monitor1'

major\_compact 'monitor1'

enable 'monitor1'