# CDH安装

## 环境准备

###### 系统环境:centos7

###### 安装包概览

cloudera-manager-centos7-cm5.13.0\_x86\_64.tar.gz

CDH相关:

CDH-5.13.0-1.cdh5.13.0.p0.29-el7.parcel

CDH-5.13.0-1.cdh5.13.0.p0.29-el7.parcel.sha1

manifest.json

jdk-8u202-linux-x64.tar.gz

mysql-5.7.24-rpm.tar.gz

mysql-connector-java-5.1.47-bin.jar

spark2.2.cl2.zip

## 环境配置:root用户操作

###### 基本配置(需修改所有节点)

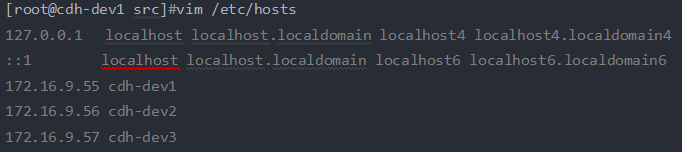
1.1修改/etc/hostname

vi /etc/hostname

cdh-dev1

1.2修改/etc/hosts:注修改此项不要更改行首为主机名称,否则识别不出名称和ip

vi /etc/hosts



1.3修改 /etc/sysconfig/network

vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=cdh-dev1

1.4关闭防火墙 #centos 7 默认使用的是firewall，不是iptables

systemctl stop firewalld.service

systemctl mask firewalld.service #注销服务,不会开机启动

1.5关闭SELinux

vi /etc/selinux/config #设置SELINUX=disabled

SELINUX=disabled

1.6所有节点上创建用户

useradd --system --home=/opt/cm-5.13.0/run/cloudera-scm-server/ --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm

**1.7设置swap空间**

echo "vm.swappiness = 0" >> /etc/sysctl.conf

1.8关闭大页面压缩

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

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

###### Java环境(需修改所有节点)

2.1查看java版本,满足需求即可,不满足需重新安装

java -version



2.2若为系统自带jdk,卸载自带jdk

rpm -qa | grep -i java #查看安装信息

rpm -e java\* --nodeps #根据查找到的软件包信息卸载

2.3安装jdk,配置环境变量

mkdir /usr/java

tar xzvf jdk-8u202-linux-x64.tar.gz -C /usr/java

vi /etc/profile #文件末尾添加下面三行

export JAVA\_HOME=/usr/java/jdk1.8.0\_202

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

export CLASSPATH=.:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar

source /etc/profile #使文件立即生效

java -version #查看版本是否为安装版本

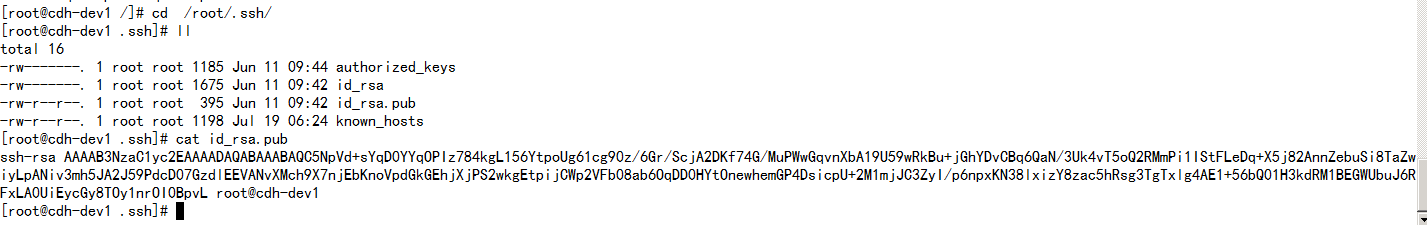
###### SSH免密登录(配置集群内节点间互相免密登录)

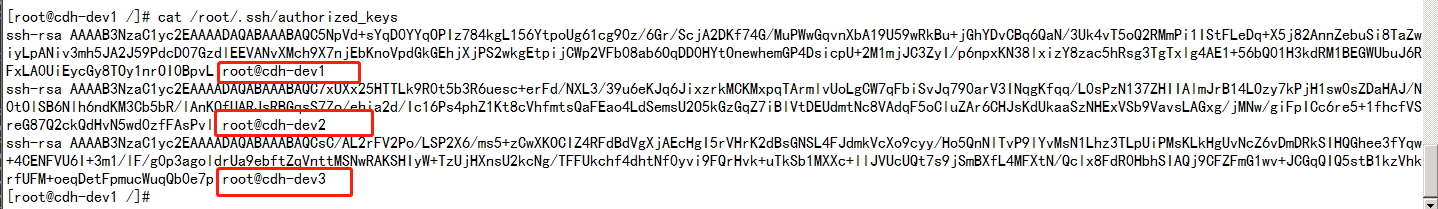
3.1创建公钥(需在所有节点内创建)

cd /root/.ssh

ssh-keygen -t rsa #直接回车

3.2合并所有节点id\_rsa.pub内容为authorized\_keys文件(就是复制所有节点的id\_rsa.pub文件内容,粘贴到文件authorized\_keys中)





3.3更改权限并分发到全部节点

chmod 600 authorized\_keys

scp authorized\_keys [root@cdh-dev2:~/.ssh](mailto:root@cdh-dev2:~/.ssh)

scp authorized\_keys [root@cdh-dev3:~/.ssh](mailto:root@cdh-dev3:~/.ssh)

###### 安装MySQL(主节点安装即可,用于存储CDH相关信息)

4.1卸载mariadb(centos7默认安装mariadb,不卸载会报错mariadb-libs is obsoleted by )

rpm -qa | grep mariadb #查看mariadb安装信息,未安装可跳过一步

rpm -e mariadb-libs\* --nodeps #根据查找到的软件包信息卸载

4.2解压安装MySQL(严格按照顺序安装)

tar zxvf mysql-5.7.24-rpm.tar.gz && cd mysql-5.7.24-rpm

rpm -ivh mysql-community-common-5.7.24-1.el7.x86\_64.rpm --nodeps

rpm -ivh mysql-community-libs-5.7.24-1.el7.x86\_64.rpm --nodeps

rpm -ivh mysql-community-client-5.7.24-1.el7.x86\_64.rpm --nodeps

rpm -ivh mysql-community-server-5.7.24-1.el7.x86\_64.rpm --nodeps

4.3MySQL的基本设置

service mysqld start #启动服务

cat /var/log/mysqld.log |grep password #查看初始密码

mysql -uroot -p #登录,密码为刚查看的初始密码

mysql> set password = password('Upchina#3#2#1'); #设置密码

#授权用户root使用密码从任意主机连接到mysql服务器

mysql> grant all privileges on \*.\* to 'root'@'%' identified by 'Upchina#3#2#1' with grant option;

mysql> flush privileges;

4.4相关数据库的创建(用于存储CDH相关组件的信息)

mysql> create database hive default charset utf8 collate utf8\_general\_ci;

mysql> create database oozie default charset utf8 collate utf8\_general\_ci;

mysql> create database hue default charset utf8 collate utf8\_general\_ci;

###### 配置ntp服务(需设置所有节点)

5.1安装ntp

yum install ntp -y #安装

chkconfig ntpd on #设置开机启动

如为内网环境且未配置yum源可下载rpm包使用rpm安装,需先安装依赖

rpm -ivh autogen-libopts-5.18-5.el7.x86\_64.rpm

rpm -ivh openssl-libs-1.0.2k-16.el7.x86\_64.rpm

rpm -ivh ntp-4.2.6p5-28.el7.centos.x86\_64.rpm

5.2主节点配置为时间服务器(CM安装节点)

vi /etc/ntp.conf #按以下内容修改

server 127.127.1.0 #无外网，配置为自身

fudge 127.127.1.0 stratum 8 #无外网，配置为自身

server ntp1.aliyun.com iburst #有外网，阿里云时钟服务器

server ntp2.aliyun.com iburst

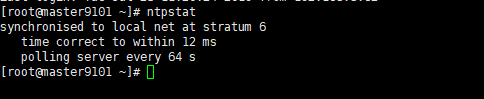
server ntp3.aliyun.com iburst

server ntp4.aliyun.com iburst

server ntp5.aliyun.com iburst

service ntpd start #启动服务

ntpstat #查看状态,如下则成功



vi /etc/ntp.conf #设置其他节点的时间服务器为主节点

server cdh-dev1 prefer

5.3安装ntpdate手动同步时间工具

yum install ntpdate #yum安装或使用如下rpm离线安装

#rpm -ivh ntpdate-4.2.6p5-28.el7.centos.x86\_64.rpm

## 安装Cloudera Manager Server 和Agent

1. 将下载好的cloudera-manager-centos7-cm5.13.0\_x86\_64.tar.gz解压到主节点的/opt/,解压出来的文件夹为cloudera与cm-5.13.0

tar -zxvf cloudera-manager-centos7-cm5.13.0\_x86\_64.tar.gz -C /opt/

1. 为Cloudera Manager建立数据库，须使用mysql jar包

cp mysql-connector-java-5.1.47-bin.jar /opt/cm-5.13.0/share/cmf/lib/

/opt/cm-5.13.0/share/cmf/schema/scm\_prepare\_database.sh mysql cm -h(MySql-IP) -uroot -pUpchina#3#2#1 --scm-host (本机IP) scm Scm#3#2#1

1. 配置agent

vi /opt/cm-5.13.0/etc/cloudera-scm-agent/config.ini

server\_host=cdh-dev1

#拷贝到其他节点

scp -r /opt/cm-5.13.0 root@cdh-dev2:/opt/

scp -r /opt/cm-5.13.0 root@cdh-dev3:/opt/

1. 配置parcels 将CDH相关文件拷贝到主节点/opt/cloudera/parcel-repo/

#相关文件

CDH-5.13.0-1.cdh5.13.0.p0.29-el7.parcel

CDH-5.13.0-1.cdh5.13.0.p0.29-el7.parcel.sha1

manifest.json

#操作

cp CDH-5.13.0-1.cdh5.13.0.p0.29-el7.parce\* /opt/cloudera/parcel-repo/

cp manifest.json /opt/cloudera/parcel-repo/

1. 启动cm

#所有节点 也可离线安装

yum install psmisc perl -y

#主节点

/opt/cm-5.13.0/etc/init.d/cloudera-scm-server start

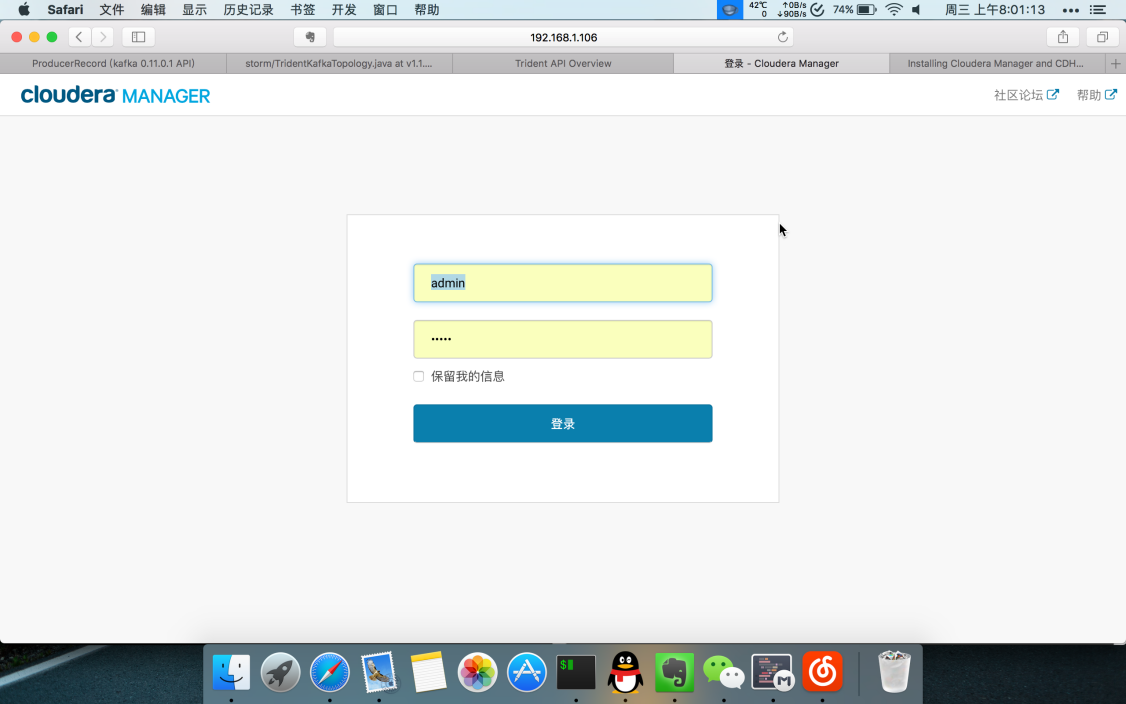
#主节点及其他节点

/opt/cm-5.13.0/etc/init.d/cloudera-scm-agent start

## 安装CDH及相关功能组件

1. 登录CM管理界面安装CDH

访问主节点的7180端口。http://cdh-dev1:7180/ 默认登录用户名和密码都是admin

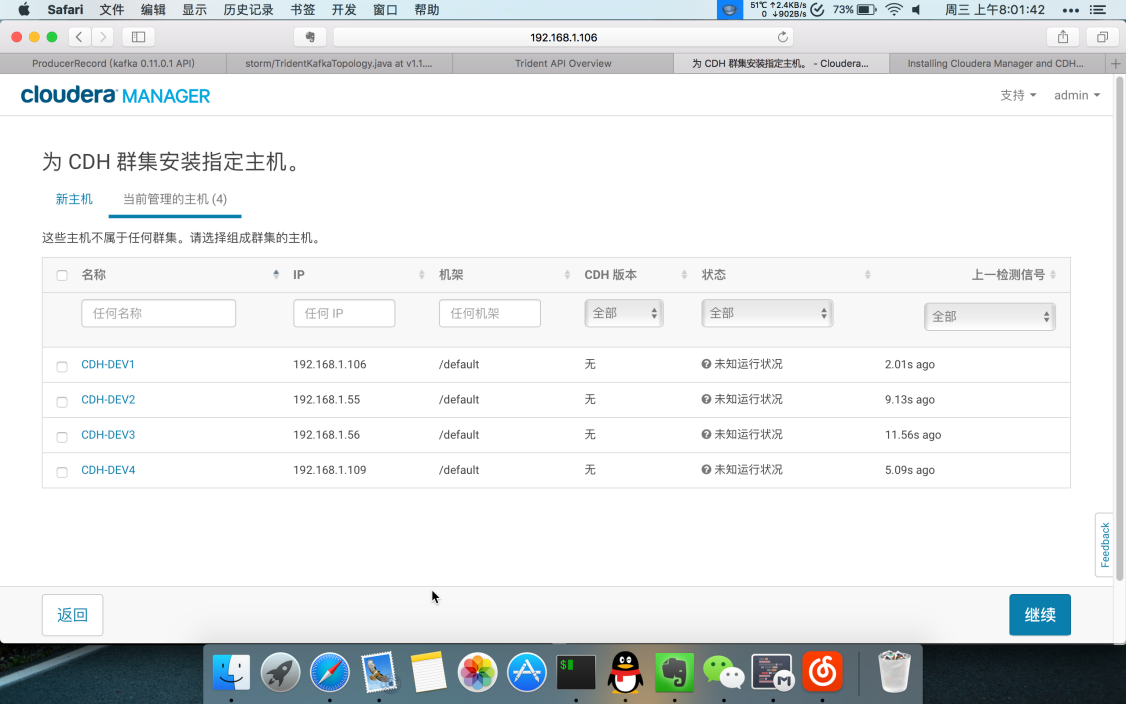


1. 选择安装版本



1. 为CDH群集安装指定主机

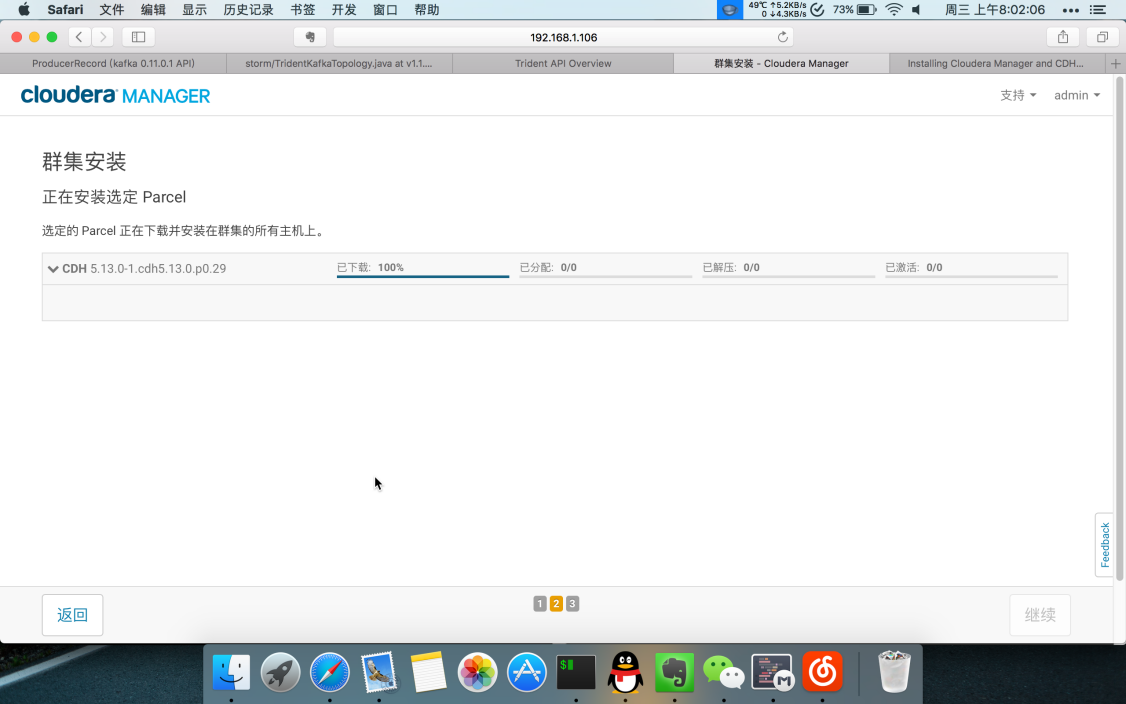
在当前管理的主机里选择所有节点，主机信息一定要正确

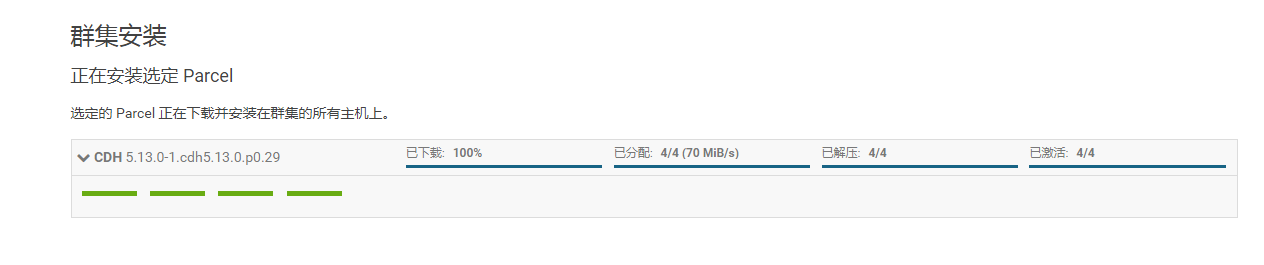


1. 集群安装

安装指定的parcel,在本地配置的是正确的Parcel包,那么下载是瞬间完成的,因为不需要下载,之后通过内网进行分配,分配完成点击继续即可







1. 检查主机正确性

若出现以下提示:



第一个警告直接在所有节点上执行echo 10 > /proc/sys/vm/swappiness

第二个警告按照上面的说明，在所有节点上执行这两条命令

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

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

并加入/etc/rc.d/rc.local中

1. 集群设置:根据业务需求添加相关功能组件

群集设置--自定义服务--根据需要按需选择

这里的服务一般是先添加zookeeper,再添加hdfs以及yarn,之后添加其他依赖于hdfs的服务,如hive,hbase,impala



群集设置--自定义角色分配--默认即可(也可按需分配:根据集群节点性能分配角色)



群集设置--审核更改--默认即可

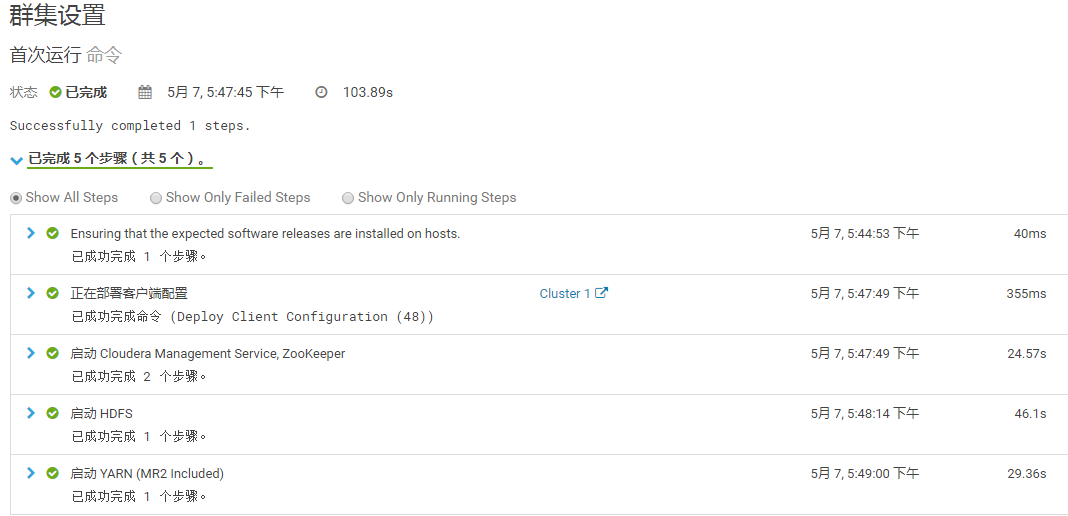
群集设置--首次运行--默认即可

群集设置--数据库设置



群集设置--审查--默认即可

群集设置--首次运行--默认即可



1. 关于角色分配

HDFS

[NameNode](http://172.16.9.101:7180/cmf/services/9/instances/337/status):元数据管理,角色分配为主节点

[DataNode](http://172.16.9.101:7180/cmf/services/9/instances/548/status):文件存储管理的服务,角色分配为相关大磁盘存储节点

SecondaryNameNode:备份元数据,分配为任一非NameNode即可

[HttpFS](http://172.16.9.101:7180/cmf/services/9/instances/466/status):hdfs的Rest Api,配置为主节点即可

Gateway:client代理,会使用到hdfs的节点需配置

[Balancer](http://172.16.9.101:7180/cmf/services/9/instances/532/status):用于平衡hdfs各节点数据,分配为任一节点亦可

HIVE

[Hive Metastore Server](http://172.16.9.101:7180/cmf/services/17/instances/174/status):hive元数据管理,配置为主节点

[HiveServer2](http://172.16.9.101:7180/cmf/services/17/instances/175/status):支持远程客户端调用hive查询,任一节点

Gateway:client代理,会使用到hive的节点需配置

YARN

[ResourceManager :](http://172.16.9.101:7180/cmf/services/15/instances/159/status)yarn节点及资源管理,配置为主节点

[NodeManager](http://172.16.9.101:7180/cmf/services/15/instances/152/status):yarn的单节点代理,进程监控,配置为hdfs 的DataNode节点

[JobHistory Server](http://172.16.9.101:7180/cmf/services/15/instances/148/status):历史程序管理,任一节点

Gateway:client代理,会使用到yarn的节点需配置

KAFKA

Kafka Broker:kafaka主要服务节点,配置为划分的kafka节点即可

Gateway:client代理,会使用到kafka的节点需配置

KUDU

[Master](http://172.16.9.101:7180/cmf/services/31/instances/542/status):kudu主节点,用于管理元数据,任一节点

[Tablet Server](http://172.16.9.101:7180/cmf/services/31/instances/544/status):实际存储数据节点,任意节点,当需使用副本时,建议为奇数个节点

SPARK2

History Server:历史任务管理,可配置为任一节点

GateWay:client代理,会使用到spark 2的节点需配置

SQOOP 1

Gateway:client代理,会使用到sqoop的节点需配置

HBASE

[Master](http://172.16.9.101:7180/cmf/services/13/instances/131/status) :主节点,管理员数据等信息,配置为主节点

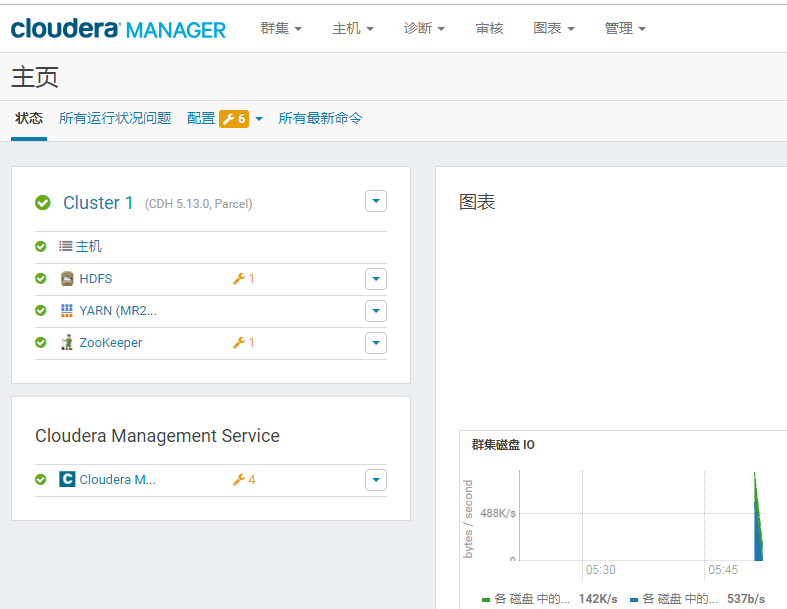
[RegionServer](http://172.16.9.101:7180/cmf/services/13/instances/130/status):负责实际数据的存储,配置为划定的hbase数据节点

[HBase Thrift Server](http://172.16.9.101:7180/cmf/services/13/instances/132/status):Thrift Api服务街店,配置为任意节点,通常为1-2个节点

HBase REST Server:REST Api服务街店,配置为任意节点,通常为1-2个节点

Gateway:client代理,会使用到hbase的节点需配置

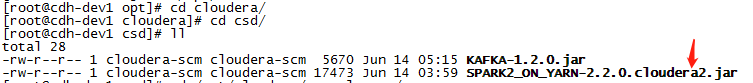
1. 完成安装



## 安装spark2

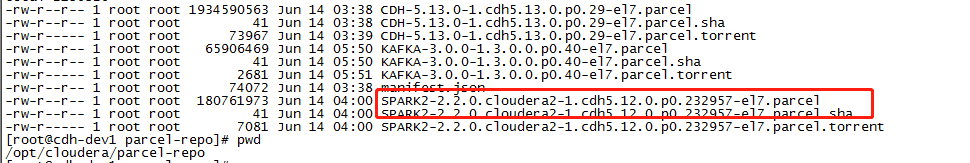
spark2不是CDH5.13.0默认组件,需要配置parcel之后安装,kafka同样如此,依样安装

1. 上传jar包到CM节点的/opt/cloudera/csd目录并更改权限



chown cloudera-scm:cloudera-scm /opt/cloudera/csd/SPARK2\_ON\_YARN-2.2.0.cloudera2.jar

1. 上传parcel的2个包到CM的/opt/cloudera/parcel-repo目录下



1. 重启CM和集群

/opt/cm-5.13.0/etc/init.d/cloudera-scm-server restart #主节点

/opt/cm-5.13.0/etc/init.d/cloudera-scm-agent restart #所有从节点

1. 通过CM页面安装spark2

进入CM-主机-Parcel界面,点击分配并激活SPARK2的parcel

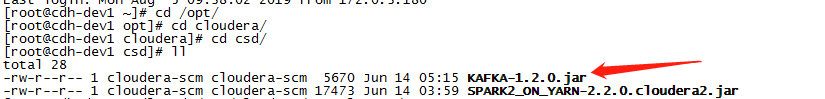


1. 通过添加服务安装SPARK2



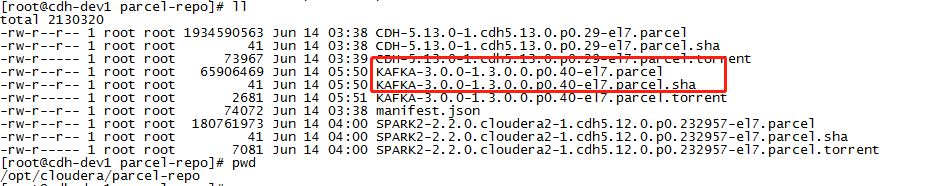
## 安装kafka

1.上传jar包到CM节点的/opt/cloudera/csd目录并更改权限



chown cloudera-scm:cloudera-scm /opt/cloudera/csd/KAFKA-1.2.0.jar

2.上传parcel的2个包到CM的/opt/cloudera/parcel-repo目录下



3.重启CM和集群

/opt/cm-5.13.0/etc/init.d/cloudera-scm-server restart #主节点

/opt/cm-5.13.0/etc/init.d/cloudera-scm-agent restart #所有从节点

4.通过CM页面安装kafka

进入CM-主机-Parcel界面,点击分配并激活kafka的parcel



5.通过添加服务安装kafka

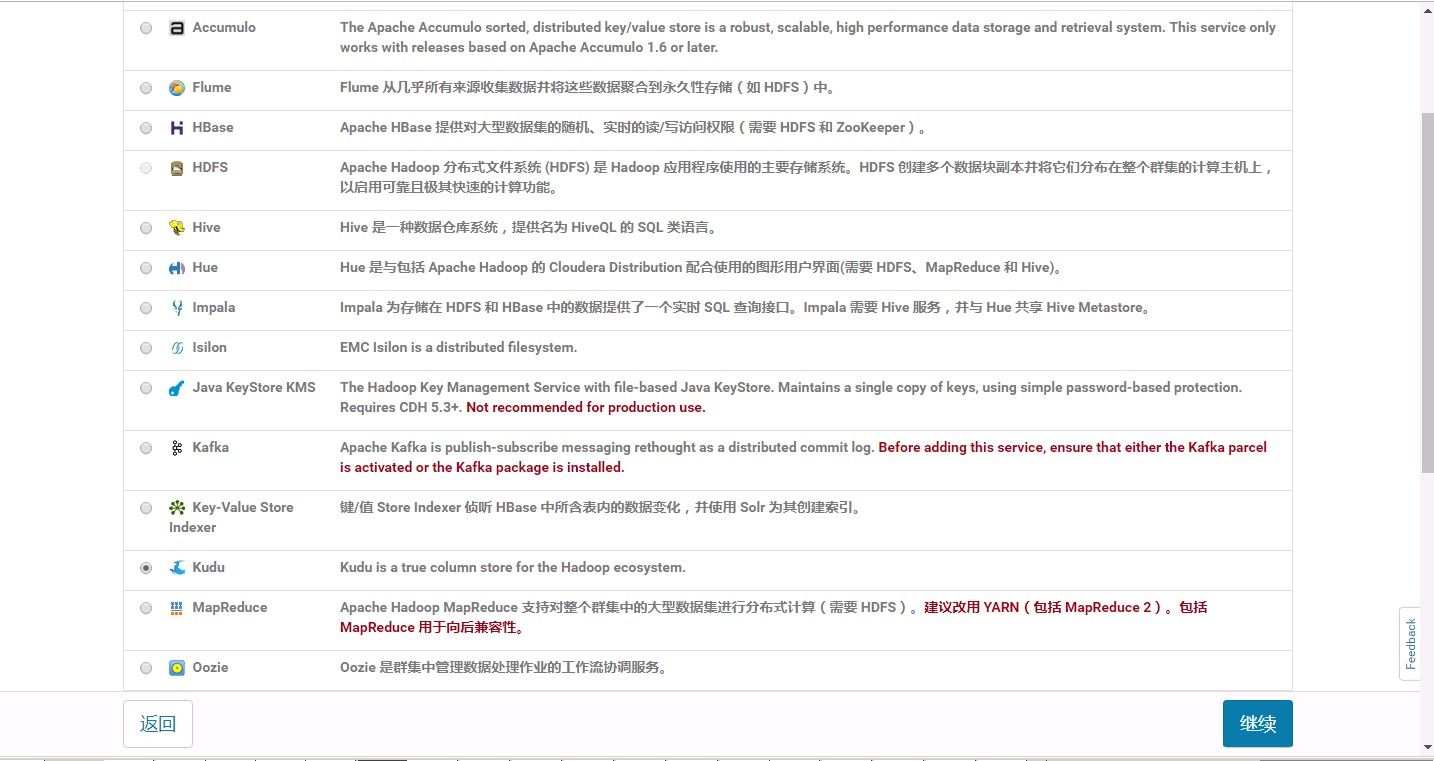


## 安装kudu

1.通过添加服务安装kudu



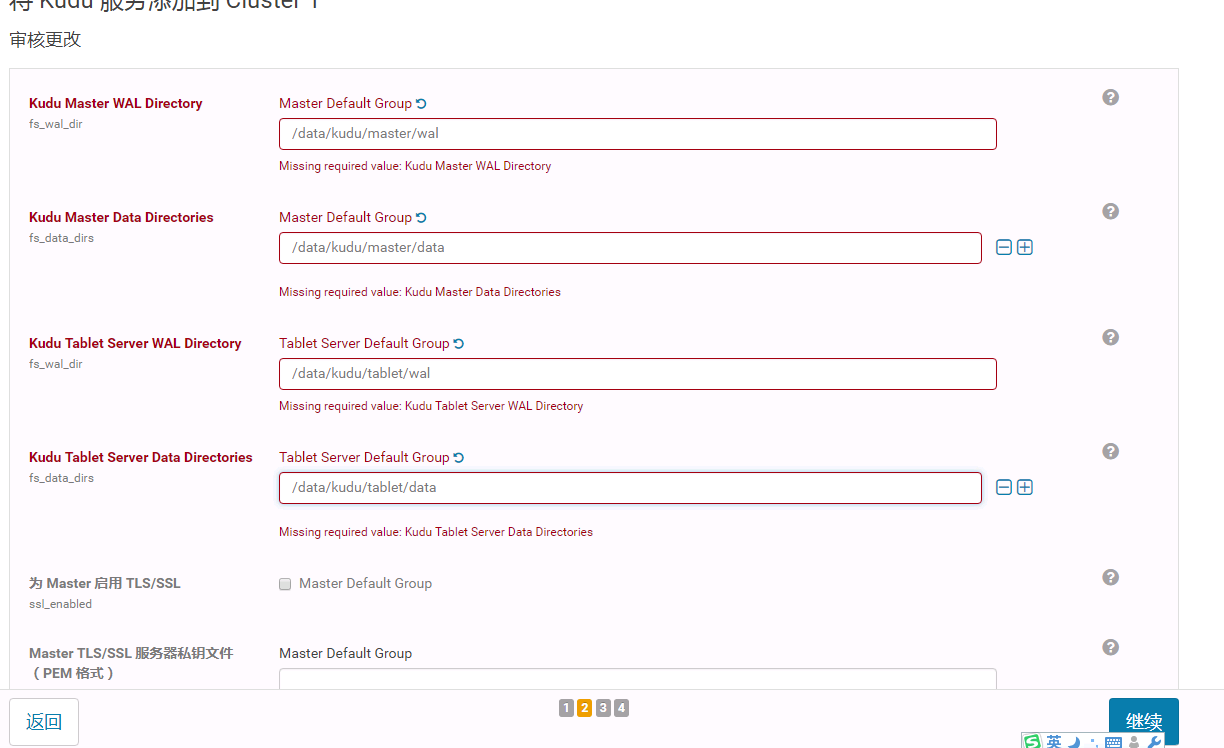
2.选择kudu,点击继续



1. 按需分配角色



1. 添加配置,继续完成安装

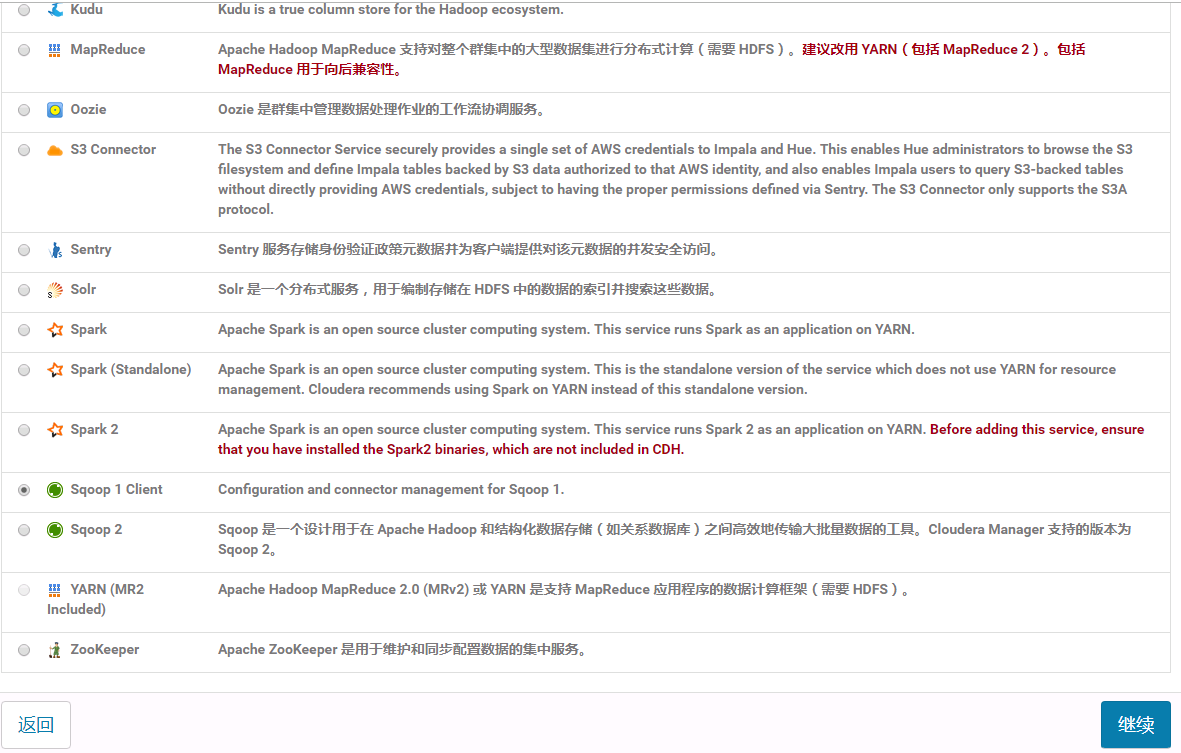


## 安装sqoop

1.通过添加服务安装sqoop



2.选择sqoop 1 Client



3.按需分配角色,需要用到sqoop的节点均分配为GateWay,点击继续,完成安装



## 集群添加节点

1. 计算平台调度spark,hive等相关服务,首先需要将计算平台的taf服务器添加为CDH相关服务的节点,集群扩容也需要添加节点(如下为计算平台taf节点)



首先需要按照上文一,二对系统环境进行配置,尤其是主节点需要配置新增节点的免密登录以及hostname

1. 解压安装CM及配置CM主节点,将cm解压安装到/opt/

tar -zxvf cloudera-manager-centos7-cm5.13.0\_x86\_64.tar.gz -C /opt/

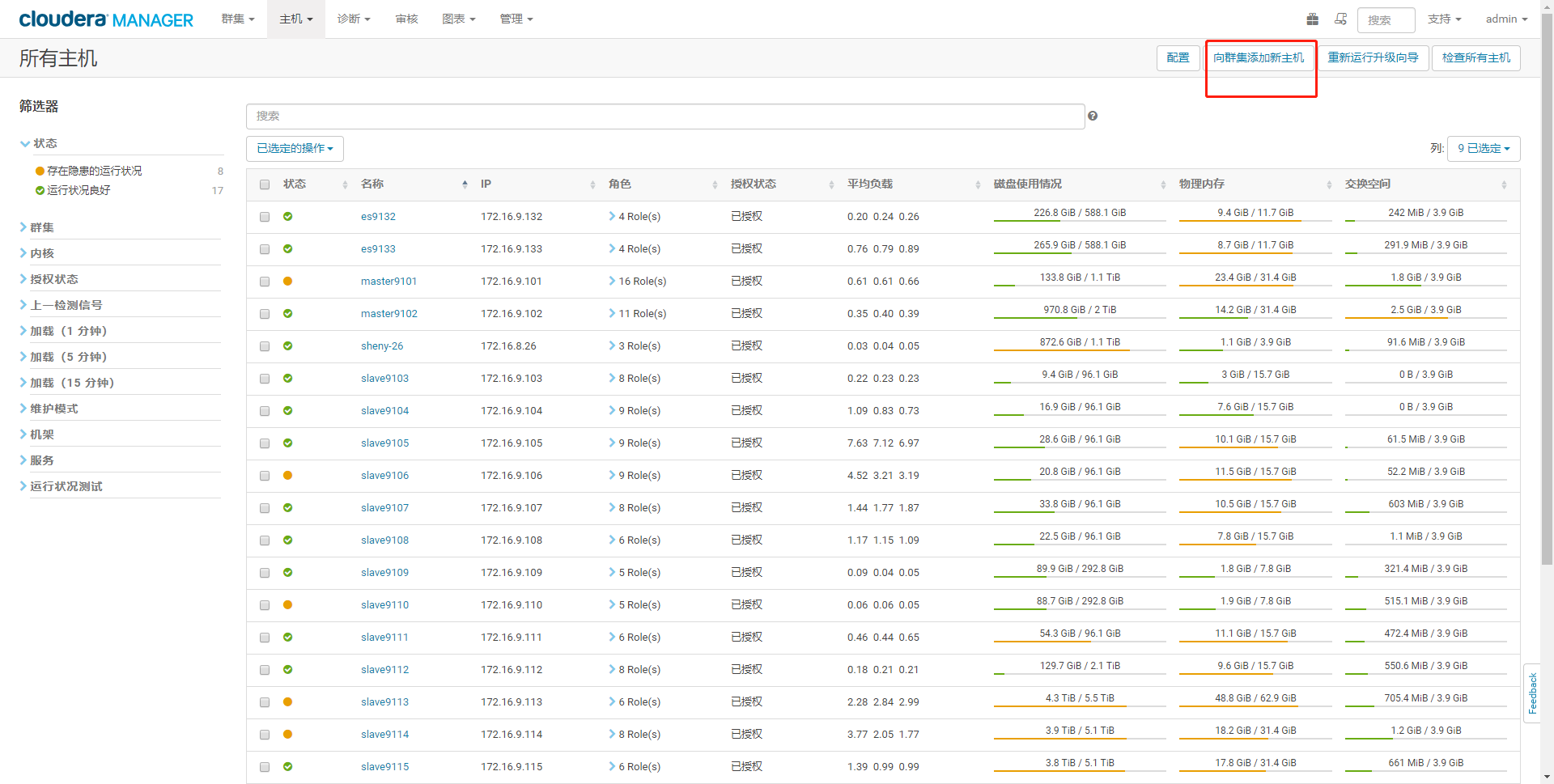
vi /opt/cm-5.13.0/etc/cloudera-scm-agent/config.ini

server\_host=cdh-dev1

1. 启动cm agent

/opt/cm-5.13.0/etc/init.d/cloudera-scm-agent start

启动之后再CM管理页面可以看见新添加的设备



1. 添加新主机并分配相关角色

点击向集群中添加新主机之后按照步骤分配角色,如果是计算平台taf节点需要分配hdfs,spark,hive等gatway角色顺序安装,之后重新部署配置即可

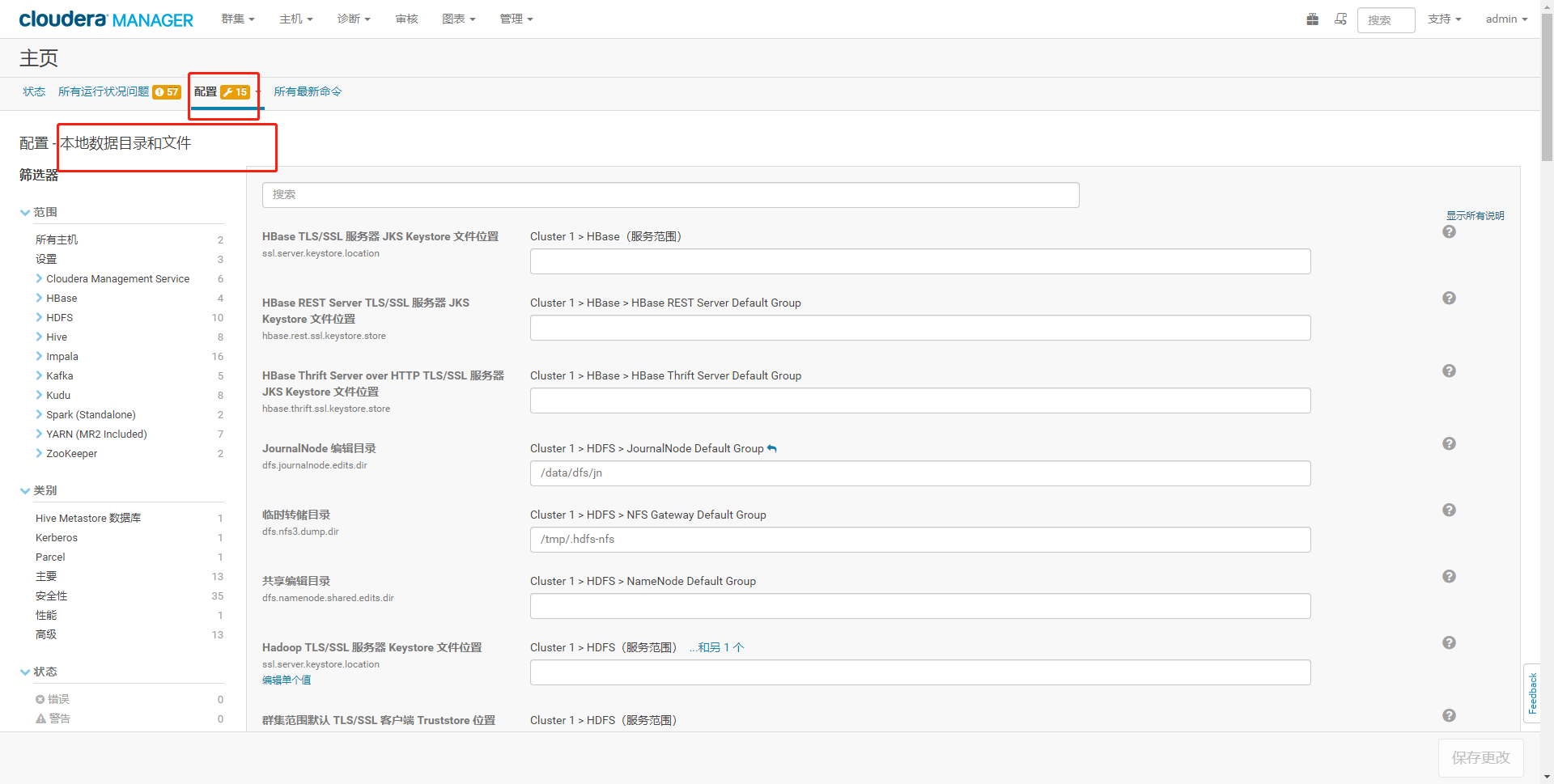
## 修改部分配置,优化集群

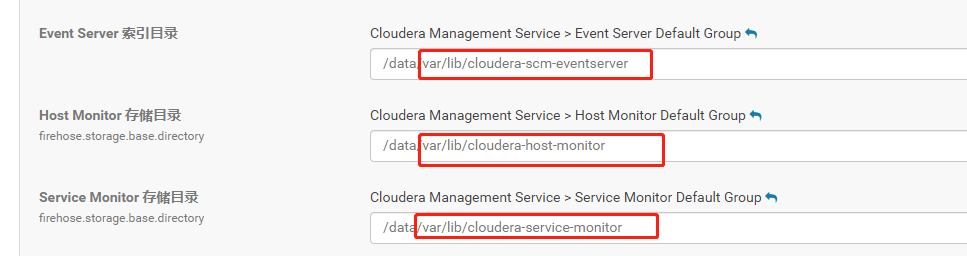
1. 更改EventServer索引目录,HostMonitor存储目录,ServiceMonitor存储目录

以上三个目录的默认存储位置为/var/lib 而且其数据量较大,占据较多系统盘空间,易造成系统盘数据空间不足,导致CDH报错,因此在CM所在节点系统盘磁盘空间不大的情况下,迁移其位置很有必要

点击配置-本地数据目录和文件

找到相应的存储设置,默认为标红部分,将其修改为非系统盘磁盘目录文件夹,如下:





mkdir -p /data/var/lib

重新部署设置即可

## 相关下载

##### 1.相关parcel下载

CDH:<http://archive.cloudera.com/cdh5/parcels/5.13.0/>

Spark2:<http://archive.cloudera.com/spark2/parcels/>

Kafka:<http://archive.cloudera.com/kafka/parcels/3.0.0/>

##### 相关jar包下载

Kafka:<http://archive.cloudera.com/csds/kafka/>

Spark2:<http://archive.cloudera.com/spark2/csd/>