**CHD5集群安装文档**

***ClouderaManager***

# 安装前准备

## 服务器分配

|  |  |  |  |
| --- | --- | --- | --- |
| 系统 | 主机IP | 主机名 | 角色划分 |
| Contos7 | 172.20.100.41 | hadoop1 | NameNode(主节点) |
| Contos7 | 172.20.100.41 | hadoop1 | DataNode(从节点) |
| Contos7 | 172.20.100.41 | hadoop1 | DataNode(从节点) |
| Contos7 | 172.20.100.41 | hadoop1 | DataNode(从节点) |
| Contos7 | 172.20.100.41 | hadoop1 | DataNode(从节点) |

## 服务器配置要求

轻处理型配置：2个16核心CPU 24-64GB内存 8块硬盘（1TB 或 2TB）

均衡型配置： 2个16核心CPU 48-128GB内存 12-16块硬盘（1TB 或 2TB）

重存储型配置：2个16核心CPU 48-96GB内存 16-24块硬盘（2TB 或 4TB）

计算加强型配置：2个16核心CPU 64-512GB内存 8块硬盘（1TB 或 2TB）

NameNode推荐配置

4-6块 1TB 硬盘（1块给操作系统，2块给FS image [RAID 1]，1块给Zookeeper , 一块给Journal Node） （至少100G）

2 个 4/8/16 核心处理器，主频至少为 2-2.5GHz （至少1个8核心处理器）

64 - 128G 内存 （至少16G）

千兆网卡或万兆网卡

DataNode推荐配置

12-24块1-4TB硬盘 （至少100G）

2个 4/8/16核心处理器，主频至少2-2.5GHz （至少1个8核心处理器）

64-512BG 内存 （至少16G）

千兆或万兆网卡（存储密度越高，需要的网络网络吞吐越高）

## 提前需准备资源包

下载CM资源包：cm5.16.1-centos7.tar.gz

下载地址：http://archive-primary.cloudera.com/cm5/repo-as-tarball/5.16.1/

以rpm方式安装postgresql 下载postgresql8.4的包

postgresql84-8.4.22-1PGDG.rhel6.x86\_64.rpm

postgresql84-libs-8.4.22-1PGDG.rhel6.x86\_64.rpm

postgresql84-server-8.4.22-1PGDG.rhel6.x86\_64.rpm

下载地址：https://yum.postgresql.org/8.4/redhat/rhel-6-x86\_64/repoview/

下载：cloudera-manager-installer.bin

下载地址：http://archive-primary.cloudera.com/cm5/installer/5.16.1/

java-mysql连接jar包：mysql-connector-java-6.0.2.jar 下载地址：

http://central.maven.org/maven2/mysql/mysql-connector-java/6.0.2/mysql-connector-java-6.0.2.jar

下载oracle版jdk ：jdk-8u191-linux-x64.tar.gz 下载地址：

https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

# 系统准备工作

## **确认环境（所有节点）**

[root@hadoop1 ~]# more /etc/redhat-release

CentOS Linux release 7.5.1804 (Core)

## **更改主机名（所有节点）**

[root@hadoop1 ]# vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=hadoop1

NTPSERVERARGS=iburst

## **更改hosts（所有节点）**

[root@hadoop1 ~]# vi /etc/hosts

127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4

172.20.100.41 hadoop1

172.20.100.42 hadoop2

172.20.100.43 hadoop3

172.20.100.44 hadoop4

172.20.100.45 hadoop5

## 设置ip（所有节点）

[root@hadoop1 ~]# vi /etc/sysconfig/network-scripts/ifcfg-eth0

TYPE=Ethernet

PROXY\_METHOD=none

BROWSER\_ONLY=no

BOOTPROTO=static

DEFROUTE=yes

IPV4\_FAILURE\_FATAL=no

IPV6INIT=yes

IPV6\_AUTOCONF=yes

IPV6\_DEFROUTE=yes

IPV6\_FAILURE\_FATAL=no

IPV6\_ADDR\_GEN\_MODE=stable-privacy

NAME=eth0

UUID=fff8562b-a05c-4e45-9f33-3765d08acf86

DEVICE=eth0

ONBOOT=yes

IPADDR=172.20.100.41

NETMASK=255.255.254.0

GATEWAY=172.20.100.99

DNS1=172.20.100.211

DNS2=8.8.8.8

重启网卡

[root@hadoop1 ~]#service network restart

## 关闭防火墙（所有节点）

[root@hadoop1 ~]# vim /etc/selinux/config

改为：SELINUX=disabled

# This file controls the state of SELinux on the system.

# SELINUX= can take one of these three values:

# enforcing - SELinux security policy is enforced.

# permissive - SELinux prints warnings instead of enforcing.

# disabled - No SELinux policy is loaded.

SELINUX=disabled

# SELINUXTYPE= can take one of three two values:

# targeted - Targeted processes are protected,

# minimum - Modification of targeted policy. Only selected processes are protected.

# mls - Multi Level Security protection.

SELINUXTYPE=targeted

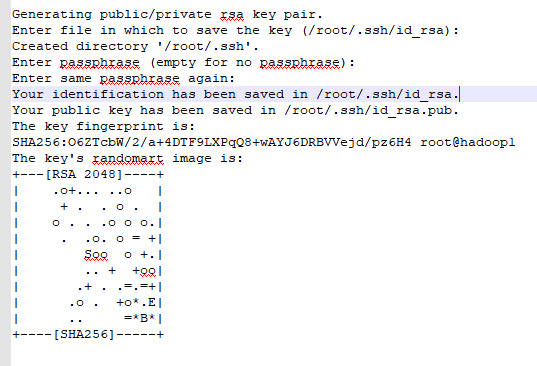
重启

[root@hadoop1 ~]# reboot

## 配置SSH免密码登录（所有节点）

在hadoop1节点生成SSH密钥对

[root@hadoop1 ~]# ssh-keygen -t rsa



将公钥复制到集群所有节点机器上

[root@hadoop1 ~]# ssh-copy-id hadoop2

 通过ssh登录各节点测试是否免密码登录成功

[root@hadoop1 ~]# ssh hadoop2

[root@hadoop2 ~]# exit

## 配置SSH免密码登录（所有节点）(可能会有所差异)

[root@hadoop1 ~]# echo 0 > /proc/sys/vm/swappiness

[root@hadoop1 ~]# echo never > /sys/kernel/mm/transparent\_hugepage/enabled

[root@hadoop1 ~]# echo never > /sys/kernel/mm/transparent\_hugepage/defrag

查看修改后结果

[root@hadoop1 ~]# cat /proc/sys/vm/swappiness

0

[root@hadoop1 ~]# cat /sys/kernel/mm/transparent\_hugepage/enabled

always madvise [never]

[root@hadoop1 ~]# cat /sys/kernel/mm/transparent\_hugepage/defrag

always madvise [never]

开机自启

[root@hadoop1 ~]# vim /etc/rc.local

echo 0 > /proc/sys/vm/swappiness

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

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

## 安装jdk（所有节点）

cloudera manager jdk 默认安装路径 /usr/java/default/ 文件夹不存在请手动创建(版本不同可能会有所调整)

[root@hadoop1 ~]# tar -zxvf jdk-8u191-linux-x64.tar.gz -C /usr/java/default/

配置环境变量

[root@hadoop1 ~]# vim /etc/profile

export JAVA\_HOME=/usr/java/default/jdk1.8.0\_171

export JRE\_HOME=$JAVA\_HOME/jre

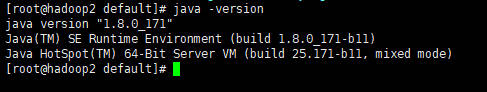
export CLASSPATH=.:$JAVA\_HOME/jre/lib/rt.jar:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar

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

让配置生效

[root@hadoop1 ~]# source /etc/profile

[root@hadoop1 ~]# java -version



## 安装MySQL（只在主节点 hadoop1）

centos7自带的是mariadb，需要先卸载掉

[root@hadoop1 ~]# rpm -qa | grep mariadb

[root@hadoop1 ~]# rpm -e --nodeps mariadb-libs-5.5.60-1.el7\_5.x86\_64

先检查系统是否装有mysql

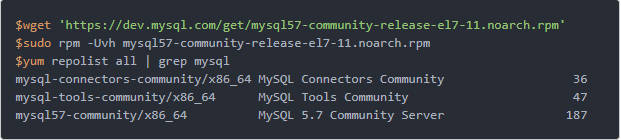
[root@hadoop1 ~]# rpm -qa | grep mysql

添加 MySQL YUM 源

# wget 'https://dev.mysql.com/get/mysql57-community-release-el7-11.noarch.rpm'

# sudo rpm -Uvh mysql57-community-release-el7-11.noarch.rpm

# yum repolist all | grep mysql

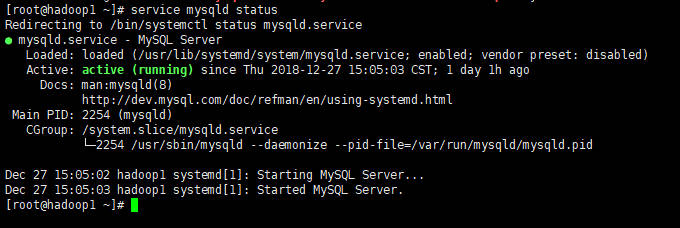


安装最新版本

# sudo yum install mysql-community-server

启动MySQL服务

# service mysqld start



service mysqld stop 关闭命令

service mysqld restart 重启命令

service mysqld status 查看服务状态

cat /etc/my.cnf 查看MySql系统配置

关闭MySQL数据库的SSL功能验证，默认是开启的 修改my.cnf文件，然后重启mysql

[mysqld]

...

skip\_ssl

# disable\_ssl

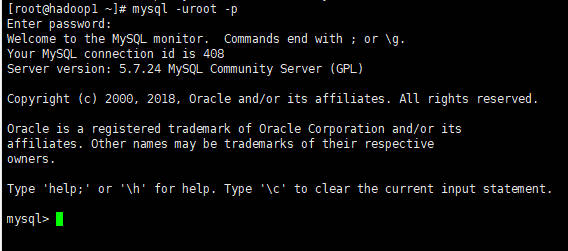
...

MySQL5.7加强了root用户的安全性，因此在第一次安装后会初始化一个随机密码，

以下为查看初始随机密码的方式，执行完该命令后则会看到一组随机字符串为初始密码。

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

# mysql -uroot -p



修改本地登入密码

# ALTER USER 'root'@'localhost' IDENTIFIED BY 'admin123A!';

进入mysql库

# use mysql;

更改库的远程权限

# GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'%' IDENTIFIED BY 'admin123A!' WITH GRANT OPTION;

#GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'localhost' IDENTIFIED BY 'admin123A!' WITH GRANT OPTION;

#GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'hadoop1' IDENTIFIED BY 'admin123A!' WITH GRANT OPTION;

#GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'172.20.100.41' IDENTIFIED BY 'admin123A!' WITH GRANT OPTION;

# select user, host, authentication\_string from user; 查询user表用户、地址、密码

# delete from user where host != '%'; 删除host != '%'的用户

# flush privileges; 刷新MySQL的系统权限

创建hive、hue、oozie 库

# create database hive DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci;

# create database hue DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci;

# create database oozie DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci;

## 安装第三方依赖包(所有节点都要安装)

# yum install chkconfig python bind-utils psmisc libxslt zlib sqlite fuse fuse-libs redhat-lsb cyrus-sasl-plain cyrus-sasl-gssapi

## 安装Apache httpd web服务器

检查httpd服务

# service httpd status

# yum install httpd 安装httpd服务

将httpd加入到启动列表：chkconfig httpd on

启动命令：service httpd start

　　关闭命令：service httpd stop

　　重启命令：service httpd restart

　　查看状态：service httpd status

## 发布CM资源文件

# mkdir -p /var/www/html/cm5/redhat/6/x86\_64

# tar -zxvf cm5.16.1-centos7.tar.gz -C /var/www/html/cm5/redhat/6/x86\_64/

# ll

drwxrwxr-x. 3 1106 592 4096 Jun 11 11:08 cm

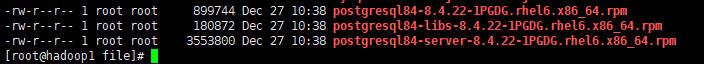
# chmod -R 777 cm

访问http://<hostname>/cm5/redhat/6/x86\_64/cm/ 查看是否发布成功

## 安装PostgerSQL

检查是否安装了postgresql

# rpm -qa|grep postgres

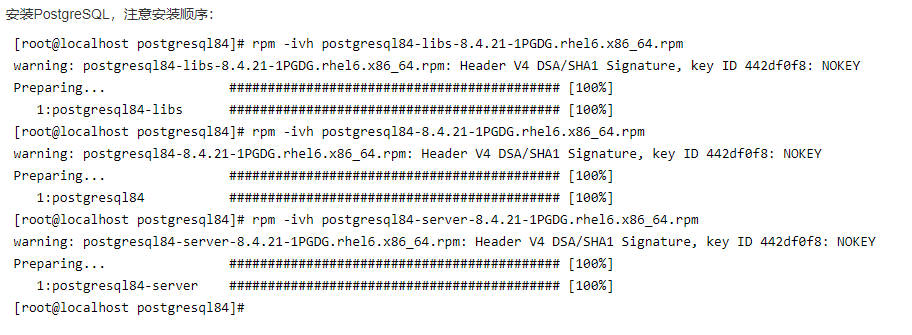


注意顺序安装

# rpm -ivh postgresql84-libs-8.4.22-1PGDG.rhel6.x86\_64.rpm

# rpm -ivh postgresql84-8.4.22-1PGDG.rhel6.x86\_64.rpm

# rpm -ivh postgresql84-server-8.4.22-1PGDG.rhel6.x86\_64.rpm



## 修改客户端配置，使其可以找到资源

# cd /etc/yum.repos.d

创建cloudera-cdh5.repo 、cloudera-manager.repo 源配置文件

cloudera-cdh5.repo内容为:

# vi cloudera-cdh5.repo

[cloudera-cdh5]

# Packages for Cloudera's Distribution for Hadoop, Version 5, on RedHat or CentOS 6 x86\_64

name=Cloudera's Distribution for Hadoop, Version 5

baseurl=https://archive.cloudera.com/cdh5/redhat/6/x86\_64/cdh/5/

gpgkey =https://archive.cloudera.com/cdh5/redhat/6/x86\_64/cdh/RPM-GPG-KEY-cloudera

gpgcheck = 1

cloudera-manager.repo 内容为：

# vi cloudera-manager.repo

[cloudera-manager]

name = Cloudera Manager, Version 5.16.1

baseurl = https://archive.cloudera.com/cm5/redhat/7/x86\_64/cm/5.16.1/

gpgkey = https://archive.cloudera.com/redhat/cdh/RPM-GPG-KEY-cloudera

gpgcheck = 1

# mkdir -p backup

# mv /etc/yum.repos.d/\*.repo ./backup

这些有需要.repo文件留为备用

/etc/yum.repos.d/地址下只留 myrepo.repo 文件和backup文件夹

新建名称为 myrepo.repo 的软件源配置文件，内容如下：

# vi myrepo.repo

[myrepo]

name=myrepo

baseurl=http://<hostname>/cm5/redhat/6/x86\_64/cm/5/

enabled=1

gpgcheck=0

# ll

-rw-r--r-- 1 root root 100 Jun 14 21:14 myrepo.repo

## 拷贝mysql-connector-java-6.0.2.jar(所有节点)

# mkdir -p /usr/share/java

拷贝mysql-connector-java-6.0.2.jar并改名为mysql-connector-java.jar

# cp mysql-connector-java-6.0.2.jar /usr/share/java/mysql-connector-java.jar

## 配置时钟同步（有时间同步可以不做）

集群中所有主机必须保持时间同步，如果时间相差较大会引起各种问题。 具体思路如下：

master节点作为ntp服务器与外界对时中心同步时间，随后对所有datanode节点提供时间同步服务。所有datanode节点以master节点为基础同步时间。

所有节点安装相关组件：

# yum install ntp

完成后，配置开机启动：

# chkconfig ntpd on

检查是否设置成功：

chkconfig --list ntpd

其中2-5为on状态就代表成功。

主节点配置

在配置之前，先使用ntpdate手动同步一下时间，免得本机与对时中心时间差距太大，使得ntpd不能正常同步。这里选用65.55.56.206作为对时中心,

ntpdate -u 65.55.56.206

ntp配置文件

# driftfile /var/lib/ntp/drift

restrict 127.0.0.1

restrict -6 ::1

restrict default nomodify notrap

server 65.55.56.206 prefer

includefile /etc/ntp/crypto/pw

keys /etc/ntp/keys

# service ntpd start 启动

配置ntp客户端（所有datanode节点）

# driftfile /var/lib/ntp/drift

restrict 127.0.0.1

restrict -6 ::1

restrict default kod nomodify notrap nopeer noquery

restrict -6 default kod nomodify notrap nopeer noquery

#这里是主节点的主机名或者ip

server n1

includefile /etc/ntp/crypto/pw

keys /etc/ntp/keys

请求服务器前，请先使用ntpdate手动同步一下时间：ntpdate -u n1 (主节点ntp服务器)

# service ntpd start

# Cloudera Manager安装

所有节点创建文件夹

# mkdir -p /opt/cloudera/parcel-repo

# mkdir -p /opt/cloudera/parcels

# mkdir -p /opt/cloudera/parcel-cache

给 cloudera-manager-installer.bin 添加可执行权限：

# chmod +x cloudera-manager-installer.bin

执行CM5安装开始安装

# ./cloudera-manager-installer.bin

一直下一步 直到安装完成

# 可视化界面配置CDH集群

## 可视化安装界面

<http://172.20.100.41:7180/cmf/home>

## Cloudera产品选择

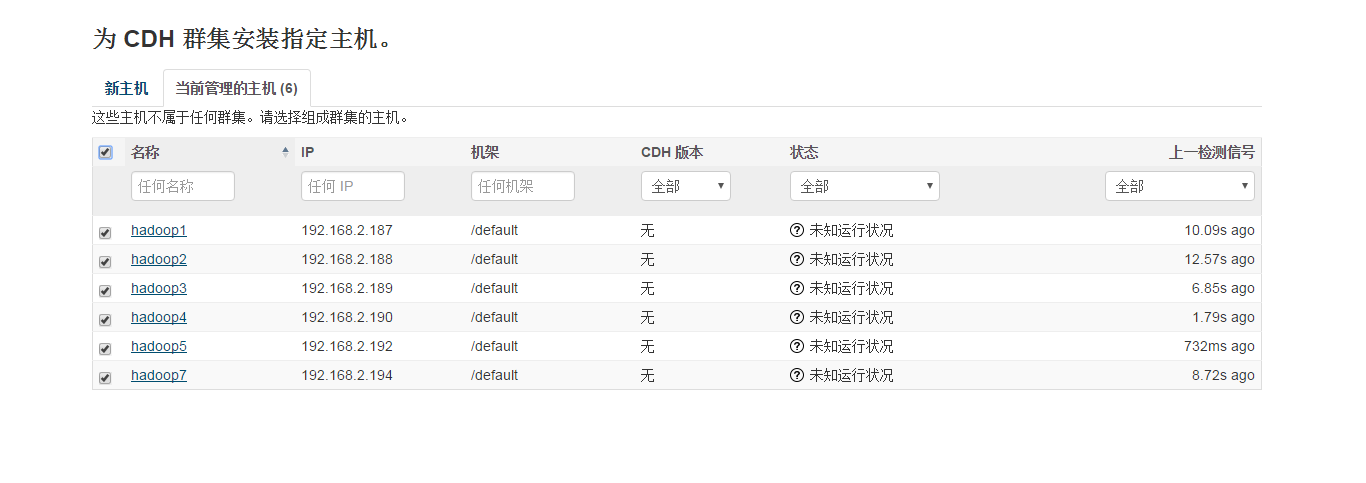
Cloudera产品选择，免费版的对集群节点数量没有限制，老版本的限制 50 个节点。 直接下一步：



直接下一步继续：



## 选择集群节点



## CDH版本选择

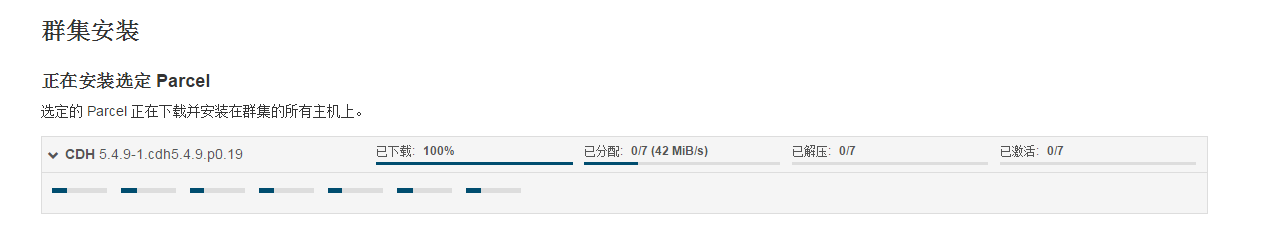
默认选项，无需修改，直接下一步：



## 安装Parcel包

CM自动下载相关的软件包，并分发到其它节点上，并解压，激活。 选择继续：





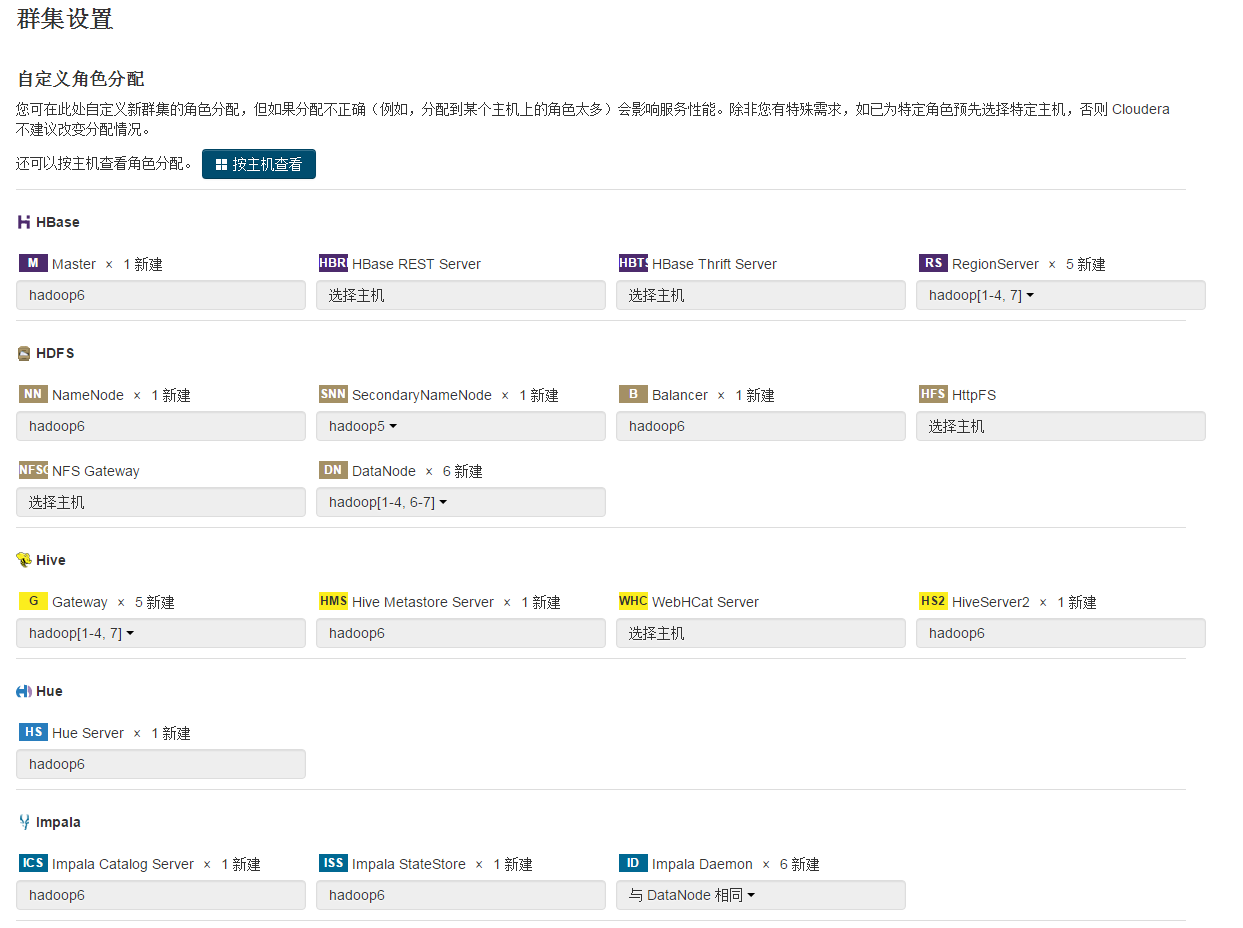
## 主机环境检查

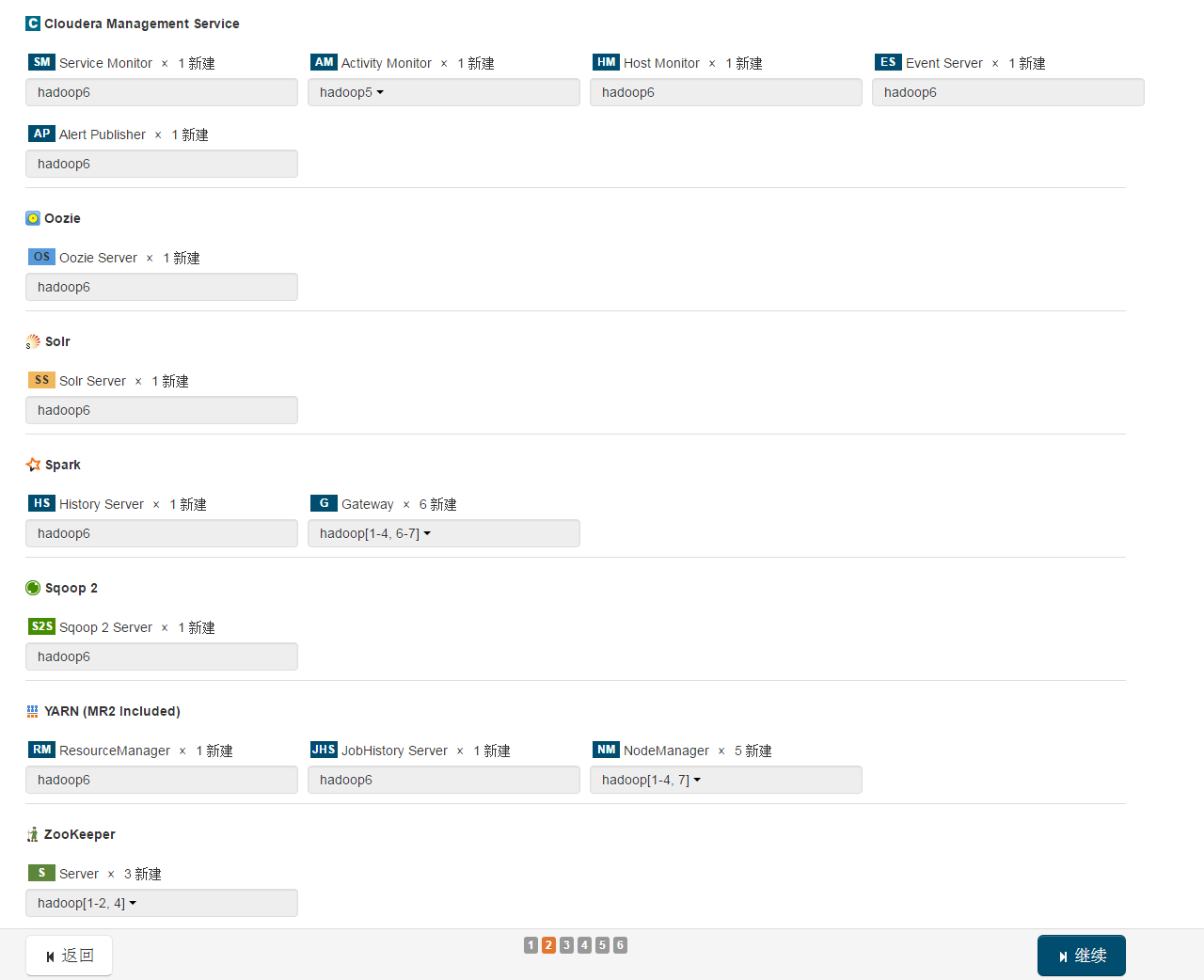
注意事项：

开始检查集群系统环境设置，如果出现报警，根据提示一个一个解决就行，解决完，选择“重新运行”进行重新检测，如果没问题直接选择“完成”。



## 集群设置

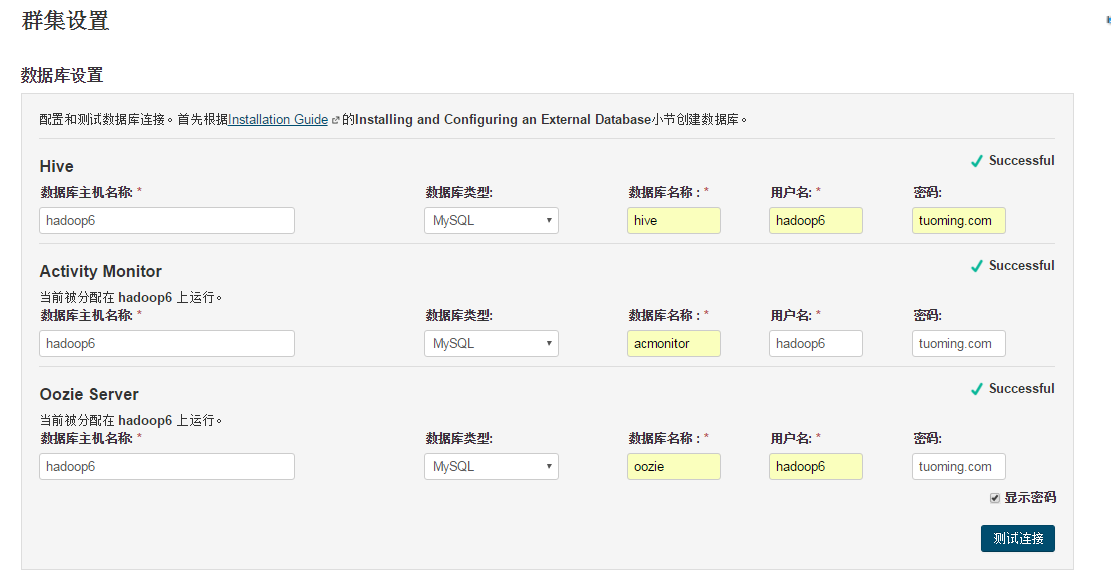




各个节点上的角色信息



## 数据库设置

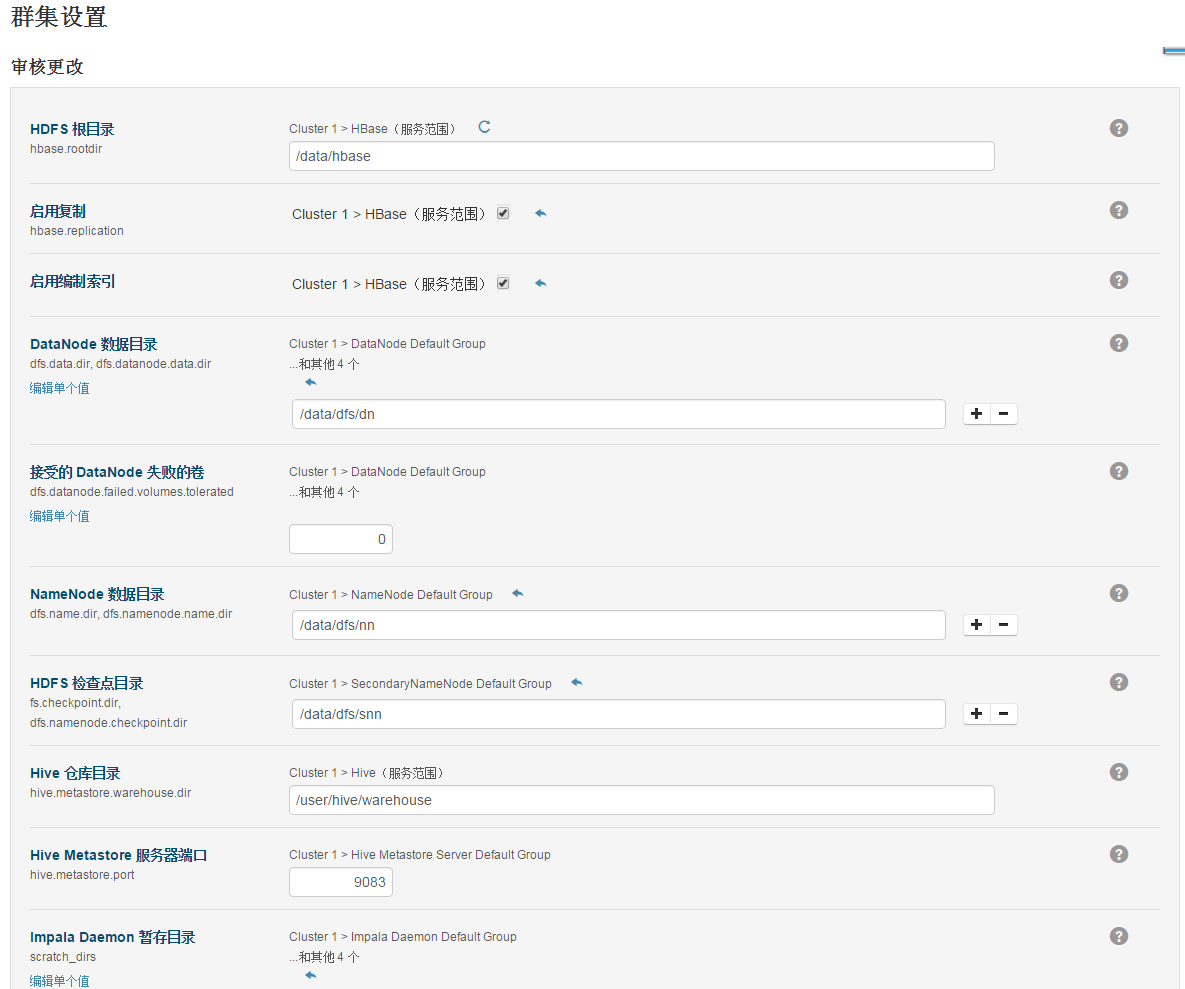


报错解决：

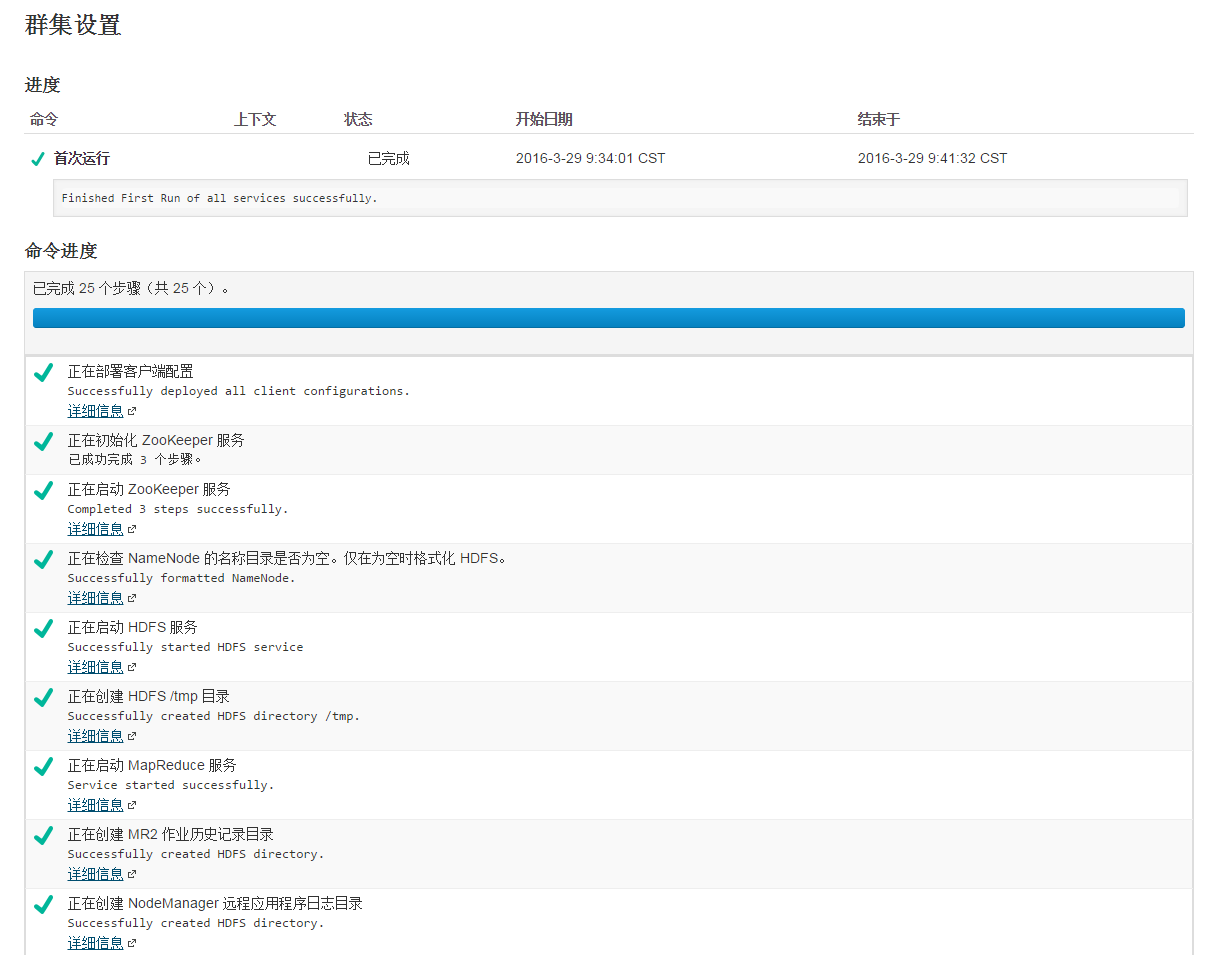
mysql> grant all privileges on \*.\* to 'hadoop6'@'%' identified by password '\*FBEE6903E46AB7D4628426E585F48718DAE951F5';

mysql> flush privileges;

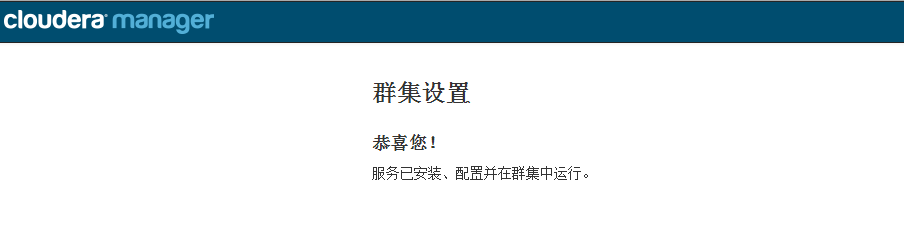
## 存储路径设置



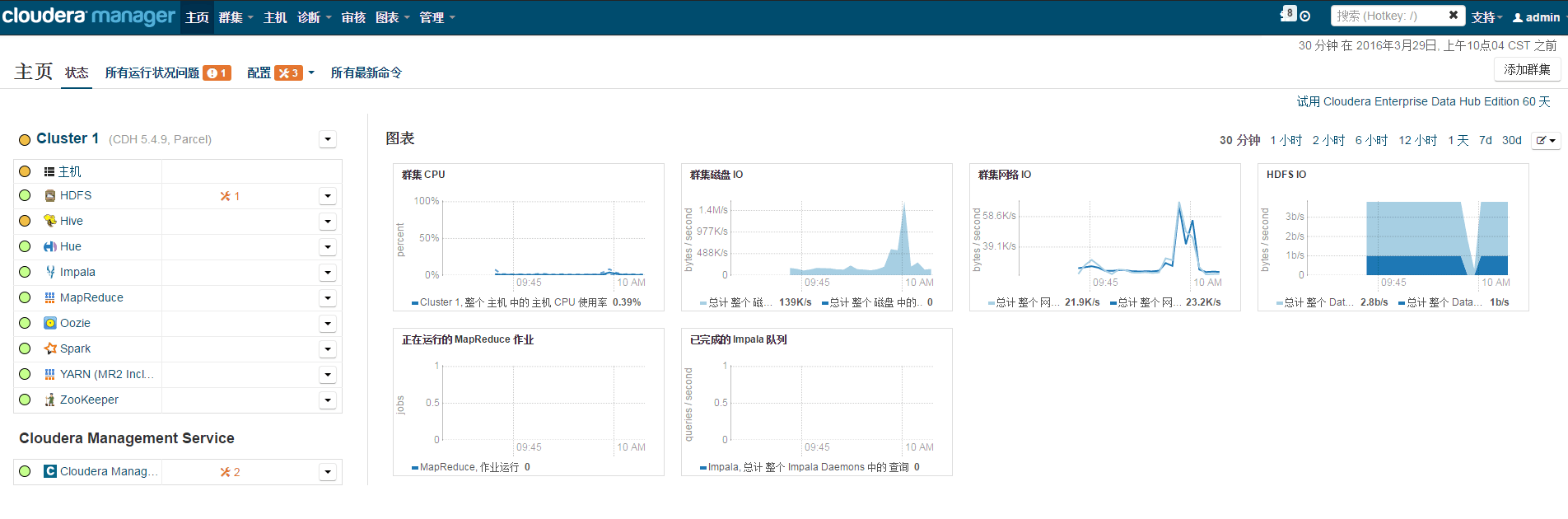
## 集群首次启动



## 安装完成



# 集群可视化管理界面



查看集群各节点运行状态：

