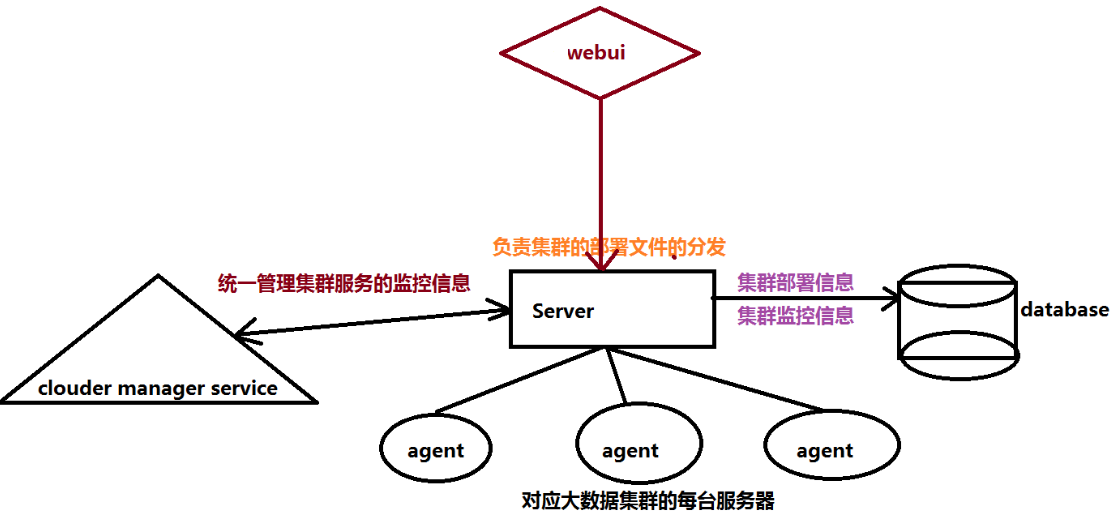
大数据平台

官方操作文档：

<http://www.cloudera.com/documentation/enterprise/5-3-x/topics/installation_installation.html>



内存分配？

hdfs :

namenode(读写元数据在内存中往复进行):

内存要求配置要高点

经验分享：1000M--->存储100W条元数据-->100W文件

datanode(主要磁盘进行数据存储):

对内存要求不大 对磁盘要求配置高点

服务器

卡槽: 磁盘矩阵 数据冗余(不需要支持RAID)

常规 服务器: 磁盘 16TB~18TB

中小型网站:访问量20W~50W 日志访问量

每次访问量: 平均5条日志

每条日志大小: 100~500bytes

每天日志访问量: 1G~5G

50W\*5\*500/1024/1024=1192M=2G

一般需要存储一年的历史数据:

1G~5G\*365=365G~1825G\*3 大致在5TB

服务器 磁盘\*80%

yarn:

ResourceManager

网络带宽要求高点

与namenode 部署在一起

NodeManager

对内存有一定的要求

与datanode 部署在一起

zoookeeper:

内存要求不高

磁盘读写要求高(SSD 固态硬盘)

网络带宽要求非常高(HA 是不是要与 zookeeper 通信)

hbase:

HMaster:

网络带宽要求配置高点

Regionserver: 内存要求配置高点

memstore

blockcache

# 环境准备

## 基础环境

### 操作系统版本:CentOS-6.5-x86\_64-minimal.iso

### IP 地址

尽可能在同一网段，

交换机\ 机架Rack

[root@cdh-1 yum.repos.d]# cat /etc/hosts

127.0.0.1 localhost.localdomain localhost

10.105.49.231 cdh-1 虚拟机5G内存

10.105.49.232 cdh-2 虚拟机2G内存

10.105.49.233 cdh-3 虚拟机2G内存

### 禁用IPv6

# echo "alias net-pf-10 off" >> /etc/modprobe.d/dist.conf

# echo "alias ipv6 off" >> /etc/modprobe.d/dist.conf

### 创建普通用户

用于安装软件

# adduser hadoop

# passwd hadoop

#### 配置普通用户的sudo权限

/etc/sudoers

写权限 # chmod u+w /etc/sudoers

>>> 添加 # vi /etc/sudoers 内容：

hadoop ALL=(root)NOPASSWD:ALL

>>> 收回权限

# chmod u-w /etc/sudoers

### 关闭防火墙

$ sudo service iptables stop

$ sudo chkconfig iptables off

### Selinux 禁用

$ sudo vi /etc/sysconfig/selinux

内容：

SELINUX=disabled

### 卸载JDK

$ sudo rpm -qa|grep java

$ sudo rpm -e --nodeps xxx yyy zzz

### 设置文件打开数量和用户最大进程数

>> 文件打开数量

$ ulimit -a

>> 用户最大进程数

$ ulimit -u

>> 设置

$ sudo vi /etc/security/limits.conf

内容：

\* soft nofile 65535

\* hard nofile 65535

\* soft nproc 32000

\* hard nproc 32000

## 软件环境

### 集群时间同步

·找一台机器作为【时间服务器】 cdh-1

yum –y install ntp

vi /etc/ntp.conf

# Hosts on local network are less restricted.

restrict 10.105.49.0 mask 255.255.255.0 nomodify notrap

# Use public servers from the pool.ntp.org project.

# Please consider joining the pool (http://www.pool.ntp.org/join.html).

#server 0.centos.pool.ntp.org iburst

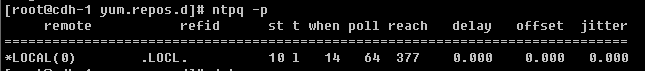
#server 1.centos.pool.ntp.org iburst

#server 2.centos.pool.ntp.org iburst

#server 3.centos.pool.ntp.org iburst

server 127.127.1.0

fudge 127.127.1.0 stratum 10



·客户端同步 定时任务



### 安装jdk软件及配置jdk环境(每台机器都执行) (可选的)

[root@master soft]# cd /root/soft/

#JDK下载:http://www.oracle.com/technetwork/java/javase/downloads/java-archive-downloads-javase7-521261.html#jdk-7u80-oth-JPR/jdk-7u80-linux-x64.rpm

[root@master soft]# rpm -ihv jdk-7u80-linux-x64.rpm

[root@master soft]# vim /etc/profile

#文件末尾增加下面内容

export JAVA\_HOME=/usr/java/default

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

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

#刷新环境变量

[root@master soft]# source /etc/profile && java –version

3、ssh无密码登录

#主机:cd-1执行命令

[hadoop@master ~]# ssh-keygen -t rsa

[hadoop@master ~]# ssh-copy-id -i ~/.ssh/id\_rsa.pub root@cdh-1

[hadoop@master ~]# ssh-copy-id -i ~/.ssh/id\_rsa.pub root@cdh-2

[hadoop@master ~]# ssh-copy-id -i ~/.ssh/id\_rsa.pub root@cdh-3

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

ssh 免密码登录(每台机器 本机+主机名)

#NOTE: 切记不要忘记ssh localhost

0) rpm -qa|grep ssh

a) 生成公钥(每台机器)

# ssh-keygen -t rsa

b) 分发公钥

# ssh-copy-id localhost

# ssh-copy-id bigdata-spark01.ibeifeng.com

# ssh-copy-id bigdata-spark02.ibeifeng.com

# ssh-copy-id bigdata-spark03.ibeifeng.com

# ssh-copy-id bigdata-spark04.ibeifeng.com

# ssh-copy-id bigdata-spark05.ibeifeng.com

c) ssh localhost

# ssh bigdata-spark01.ibeifeng.com

# ssh bigdata-spark02.ibeifeng.com

# ssh bigdata-spark03.ibeifeng.com

# ssh bigdata-spark04.ibeifeng.com

# ssh bigdata-spark05.ibeifeng.com

d) exit

Permission denied (publickey,gssapi-keyex,gssapi-with-mic,password).

ssh 可同时支持 publickey和password两种方式 默认 publickey 不开启 需要配置为yes

如果 客户端不存在 publickey 则使用password 授权;若publickey 存在则使用 publickey，如果publickey 授权失败 依然会继续使用password 授权

## cat /etc/ssh/sshd\_config 进行修改

ssh-copy-id -i ~/.ssh/id\_rsa.pub localhost

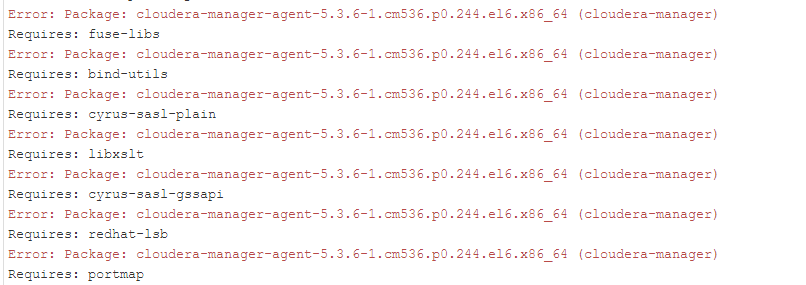
## [-I pkcs11] [-i identity\_file]

### 安装perl

yum install perl

### Cloudera Agent 所依赖的软件包（需要提前安装）：

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

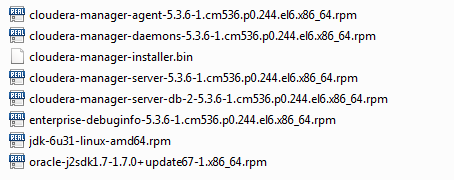


yum install fuse-libs bind-utils cyrus-sasl-plain libxslt cyrus-sasl-gssapi redhat-lab portmap

# 软件安装

## 安装 CM 5.3.x

rpm+本地yum源安装 软件下载： <http://archive.cloudera.com/cm5>



### yum本地源

1) Apache 服务器

[beifeng@bigdata-cdh01 ~]$ sudo service httpd status

httpd is stopped

[beifeng@bigdata-cdh01 ~]$ sudo service httpd start

Starting httpd: [ OK ]

[beifeng@bigdata-cdh01 ~]$ sudo chkconfig httpd on

2) 搭建本地源

[beifeng@bigdata-cdh01 html]$ pwd

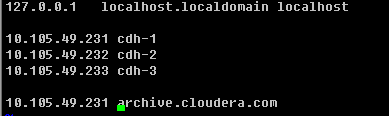
/var/www/html

[beifeng@bigdata-cdh01 html]$ sudo mkdir -p cm5/redhat/6/x86\_64/cm/5.3.6/RPMS/x86\_64/

[beifeng@bigdata-cdh01 x86\_64]$ pwd

/var/www/html/cm5/redhat/6/x86\_64/cm/5.3.6/RPMS/x86\_64

[beifeng@bigdata-cdh01 x86\_64]$ sudo mv /opt/softwares/cm-5.3.6/\* .



[root@cdh-1 yum.repos.d]# vi CentOS-Base.repo

[cloudera-manager]

# Packages for Cloudera Manager, Version 5, on RedHat or CentOS 6 x86\_64

name=Cloudera Manager

baseurl=http://archive.cloudera.com/cm5/redhat/6/x86\_64/cm/5.3.6/

gpgkey =http://archive.cloudera.com/cm5/redhat/6/x86\_64/cm/RPM-GPG-KEY-cloudera

gpgcheck = 0

### 安装依赖包 postgresql\*

cd /opt/softwares/postgresq-libs

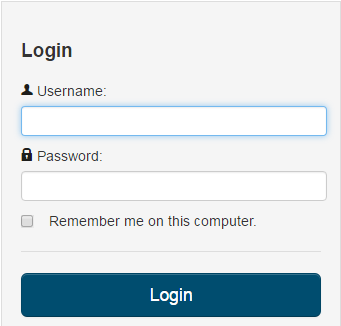
[hadoop@cdh-1 postgresq-libs]$ sudo rpm -ivh \*.rpm

## 安装cm

[hadoop@cdh-1 cm-5.3.6]$ sudo ./cloudera-manager-installer.bin

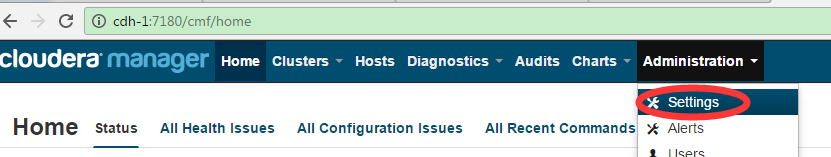
安装成功后默认用户名 admin /admin

webUI : <http://cdh-1:7180/cmf/login>

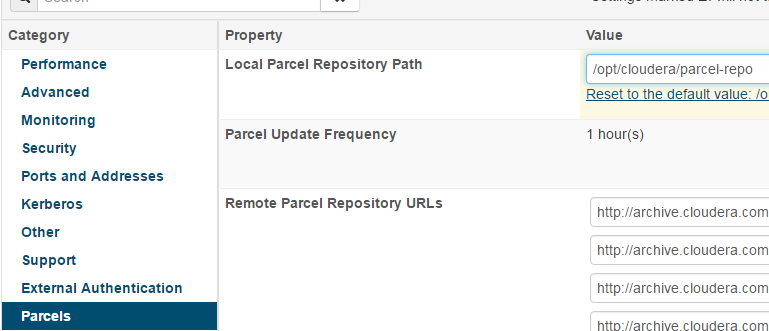


## 设置本地parcel文件路径

登录成功后，点击 Administration – Settings



找到 parcels



将parcel 文件移动到目标位置：

[root@cdh-1 parcel-repo]# mv /opt/softwares/cdh-5.3.6/\* .

[root@cdh-1 parcel-repo]# ll

total 1473856

-rw-r--r-- 1 hadoop hadoop 1509217191 Oct 18 2015 CDH-5.3.6-1.cdh5.3.6.p0.11-el6.parcel.parcel

-rw-r--r-- 1 hadoop hadoop 41 Oct 17 2015 CDH-5.3.6-1.cdh5.3.6.p0.11-el6.parcel.sha1

修改文件名 ：

mv CDH-5.3.6-1.cdh5.3.6.p0.11-el6.parcel.parcel CDH-5.3.6-1.cdh5.3.6.p0.11-el6.parcel

**[root@cdh-1 parcel-repo]# mv CDH-5.3.6-1.cdh5.3.6.p0.11-el6.parcel.sha1 CDH-5.3.6-1.cdh5.3.6.p0.11-el6.parcel.sha**

## 查看和重启CM 服务

[root@cdh-1 parcel-repo]# service cloudera-scm-server status

cloudera-scm-server (pid 1614) is running...

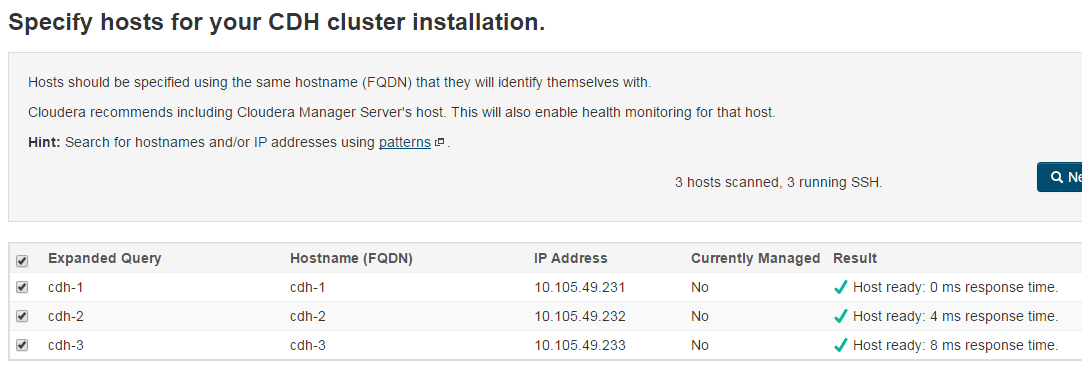
[root@cdh-1 parcel-repo]# service cloudera-scm-server restart

Stopping cloudera-scm-server: [ OK ]

Starting cloudera-scm-server: [ OK ]

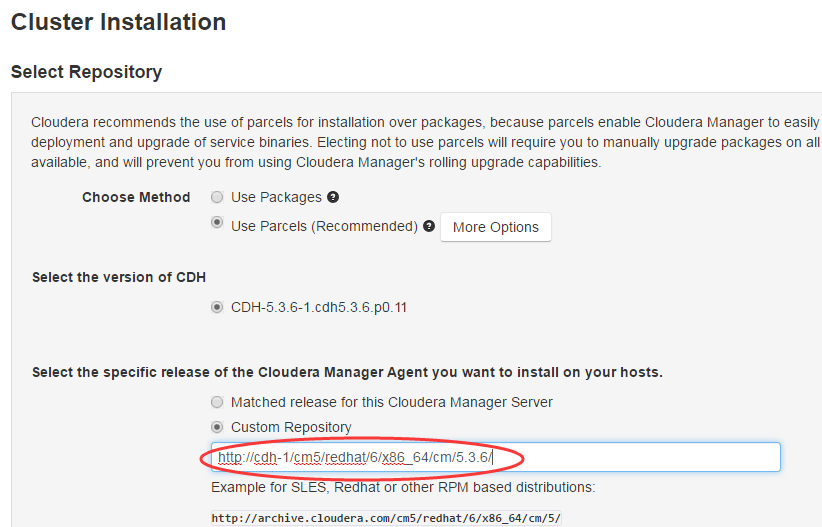
## 为集群添加主机

cdh-[1-3]

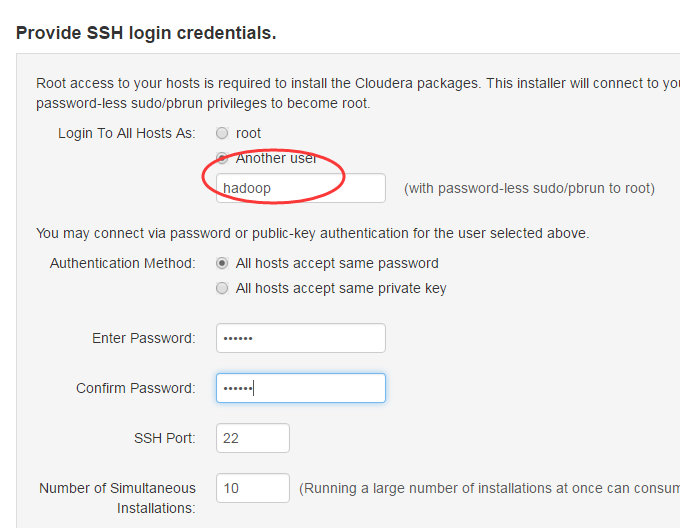


## 安装Cloudera Agent

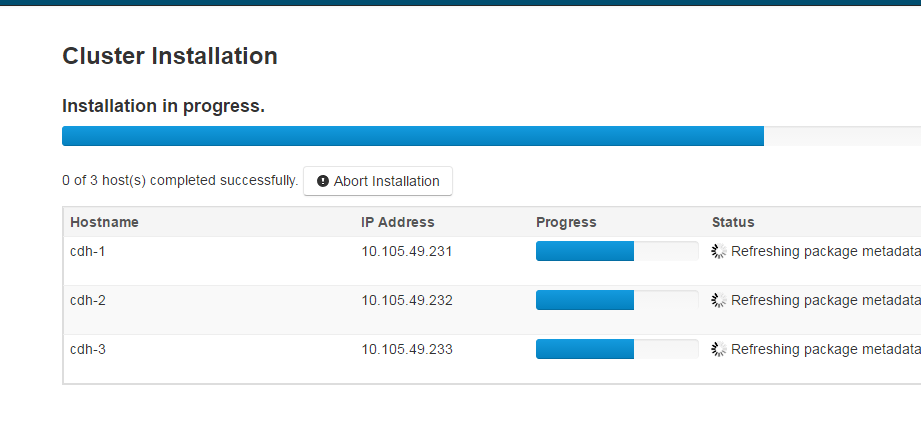
三台机器都需要安装 agent.



设置安装用户 ：



开始安装：

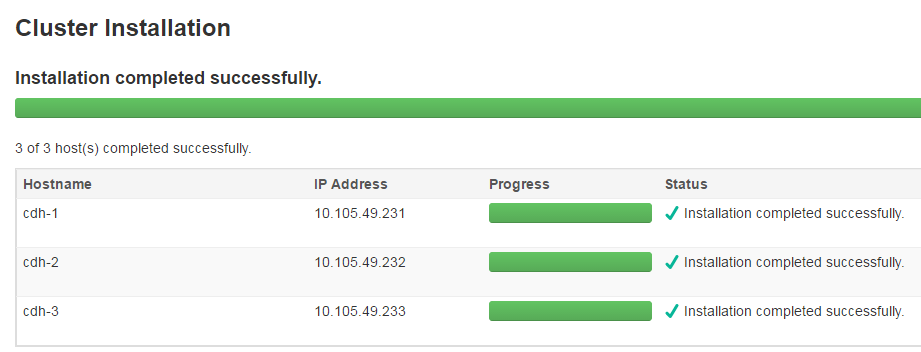
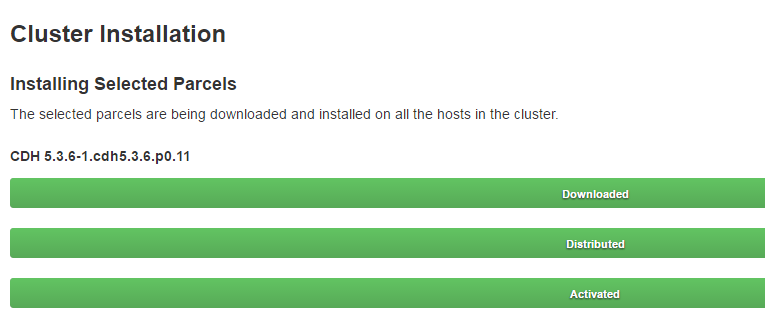


service cloudera-scm-agent restart

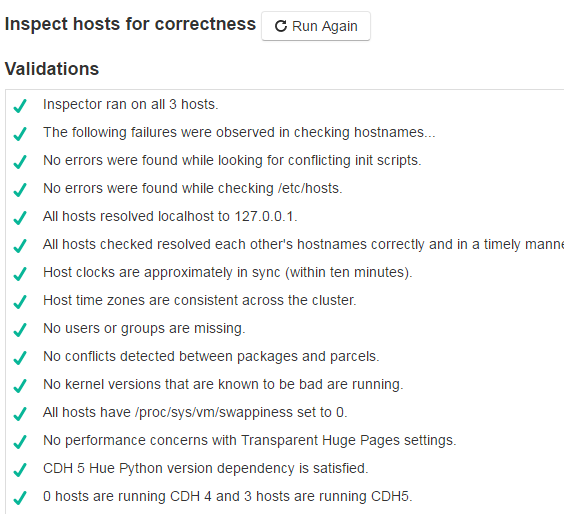
#### 问题1：



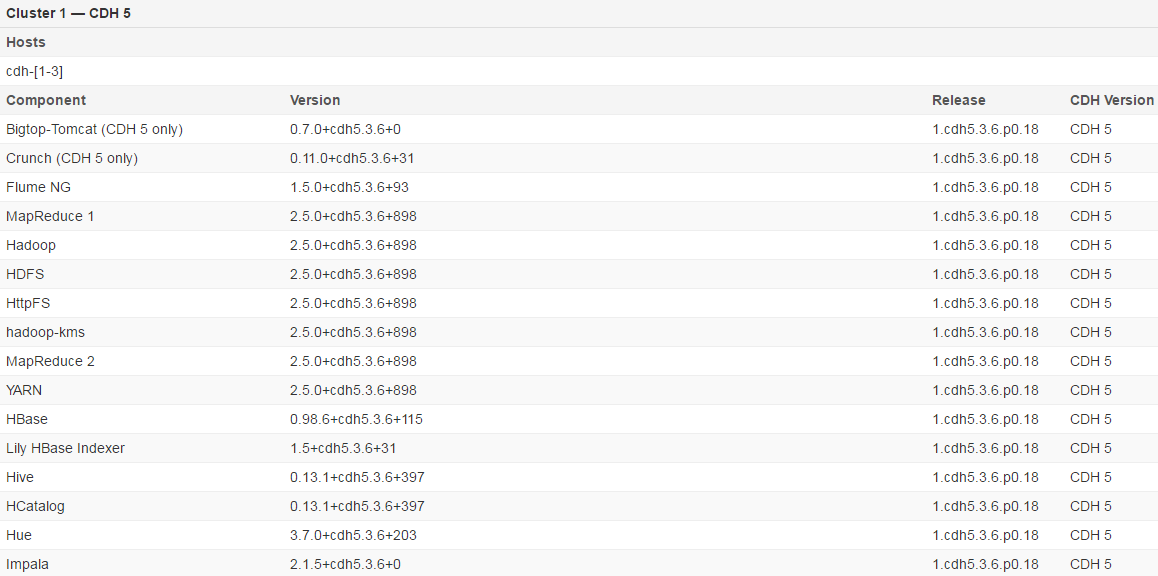
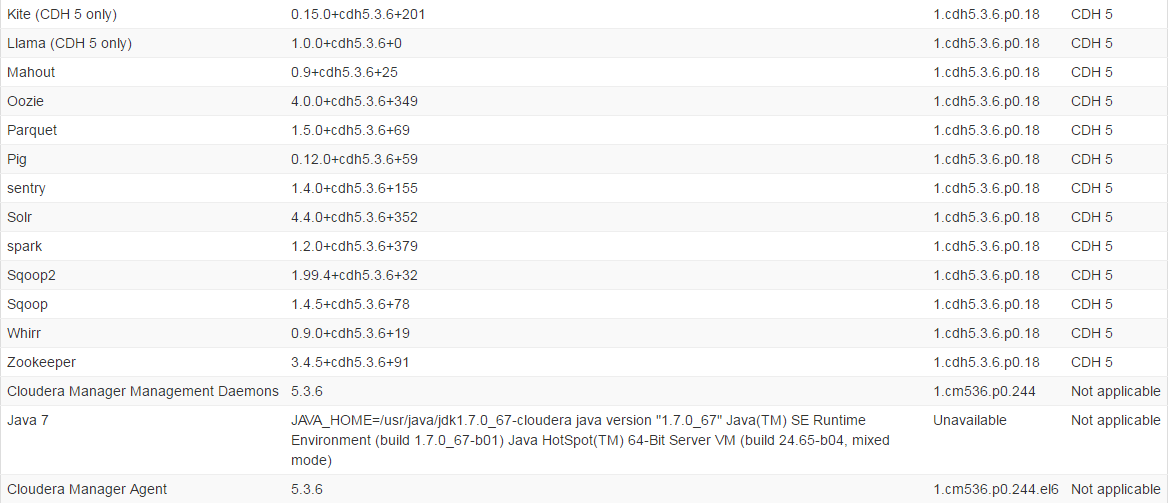
安装成功：

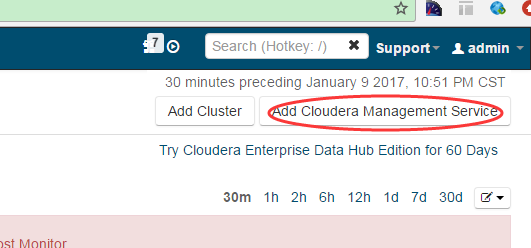
检查项：



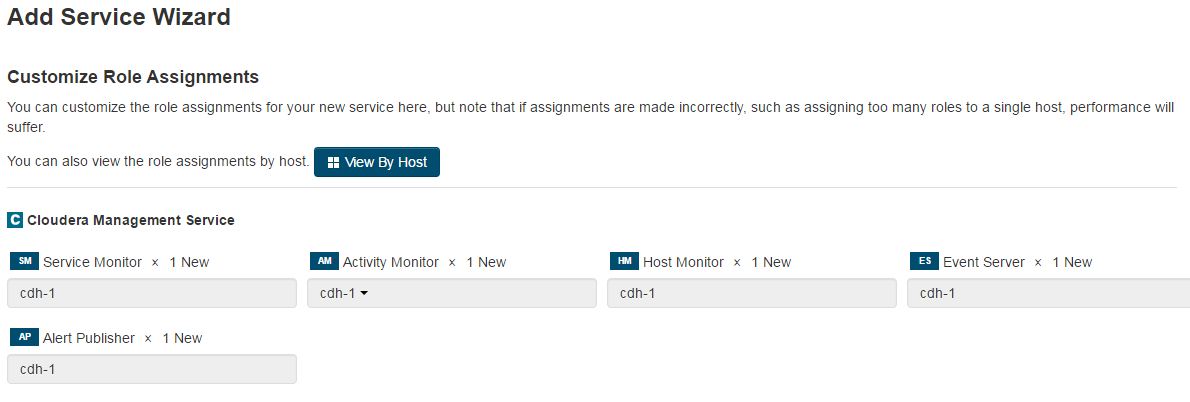
摘要信息：

## 安装集群监控服务

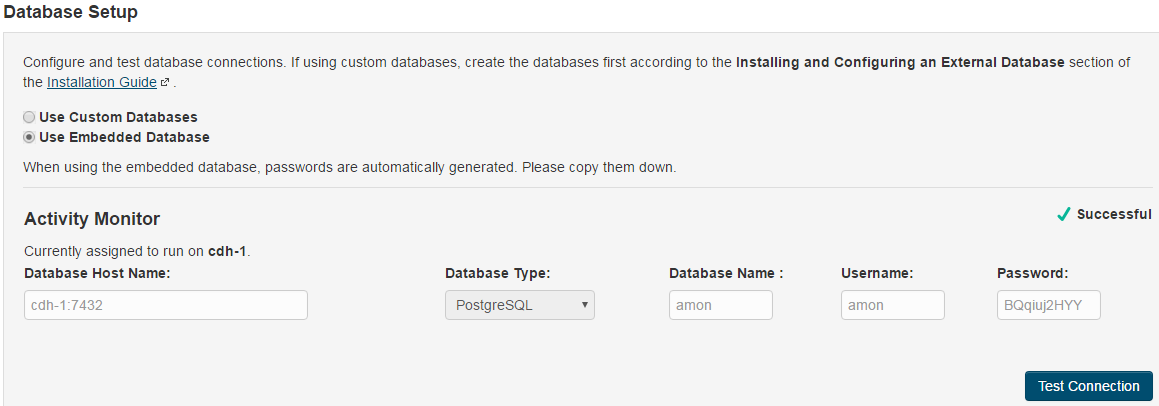


### 选择服务安装主机

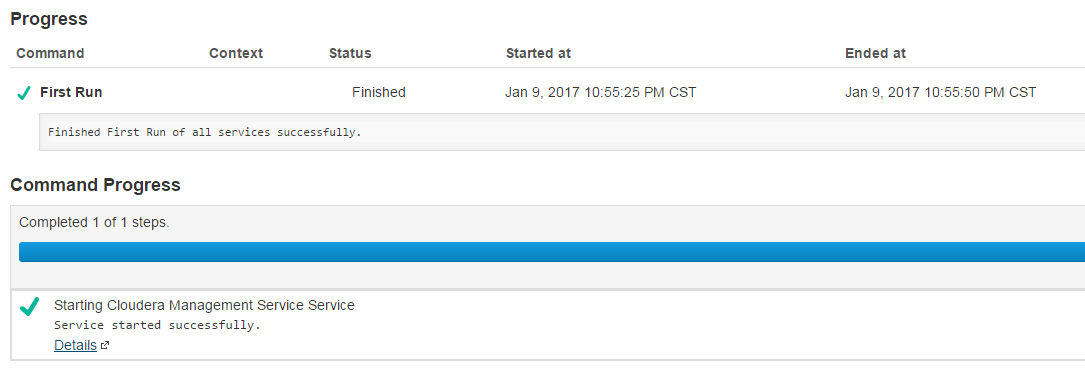


这里选择放在第一台主机上

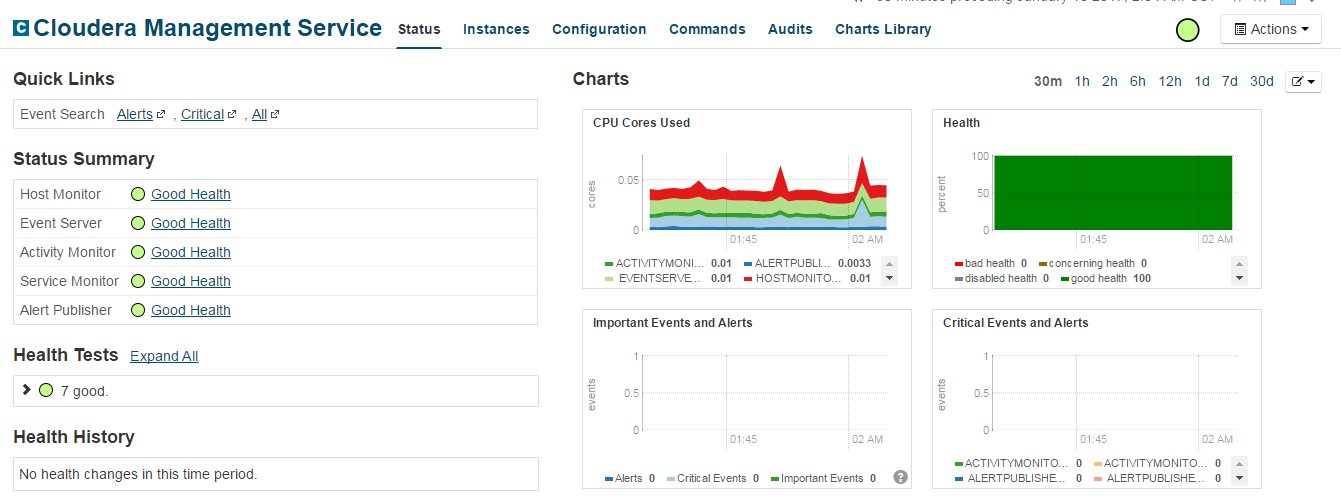
### 选择数据库



### 安装完成等待服务启动



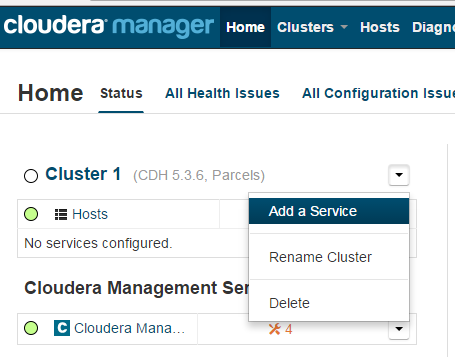
完成后示意图：



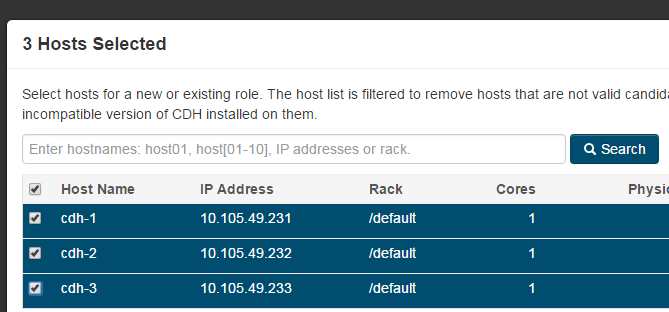
# 应用组件安装

## Zookeeper 安装及测试

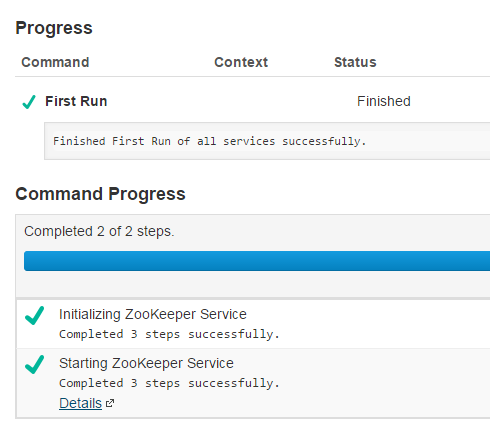
### 添加服务



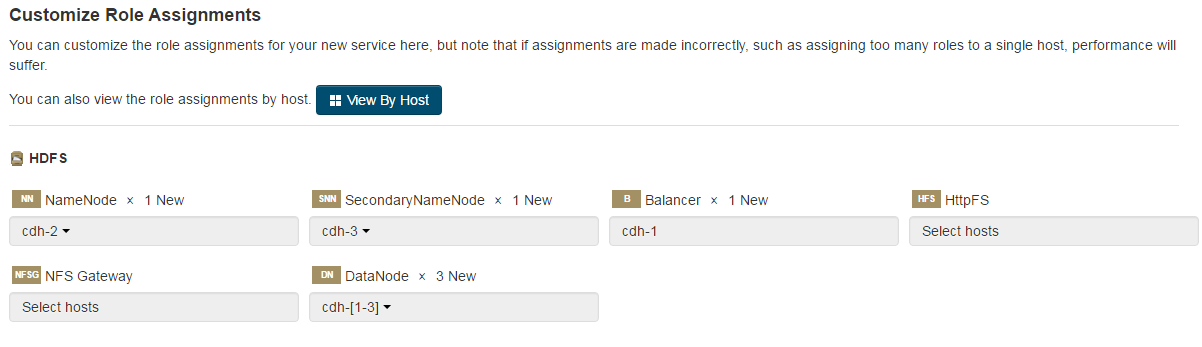
选择zookeeper 安装到哪些主机：

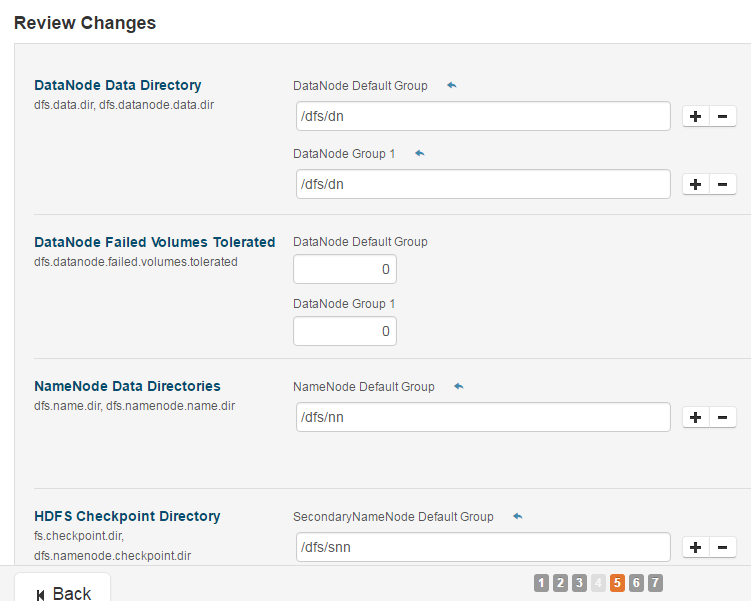


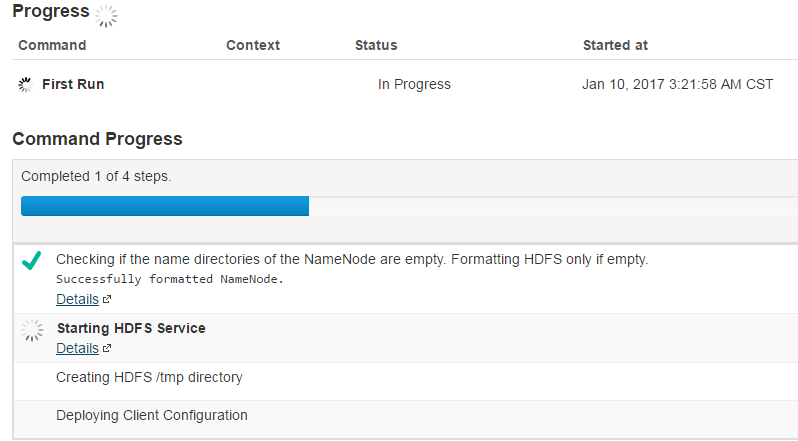
安装完成：



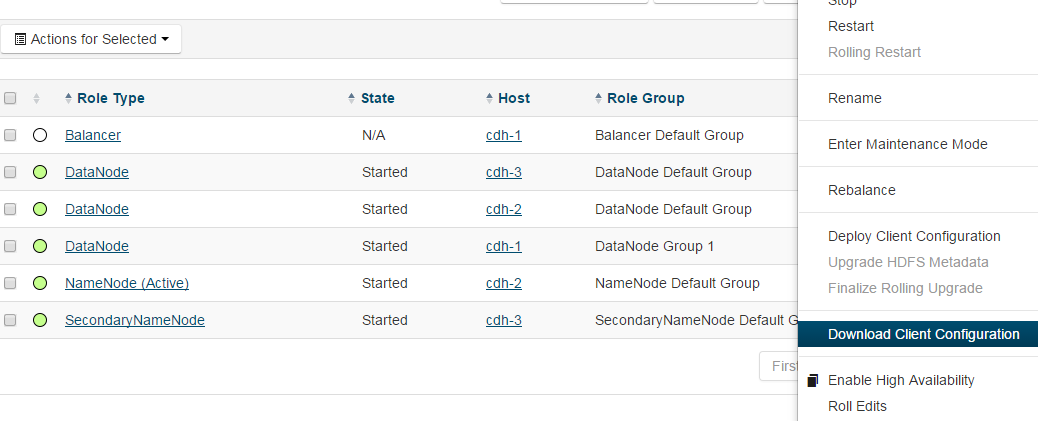
## HDFS



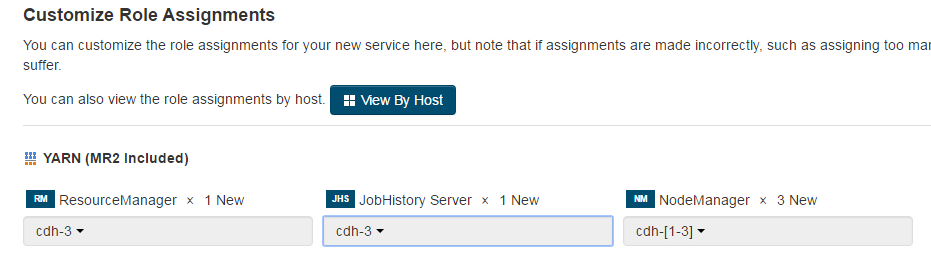




### 下载客户端配置：



## YARN



# 卸载CM

卸载之前，在管理界面停止所有服务

**卸载CM Server**

如果用cloudera-manager-installer.bin方式安装，通过如下卸载：

$ sudo /usr/share/cmf/uninstall-cloudera-manager.sh

如果没用cloudera-manager-installer.bin方式安装，通过如下：

1. 停止Cloudera Manager Server 和db服务:

$ sudo service cloudera-scm-server stop

$ sudo service cloudera-scm-server-db stop

2. 卸载Cloudera Manager Server 和db，如果采用了内嵌的 PostgreSQL database：

$ sudo yum remove cloudera-manager-server

$ sudo yum remove cloudera-manager-server-db-2

停止CM agent服务：

$ sudo service cloudera-scm-agent hard\_stop

卸载所有组件：

$ sudo yum remove 'cloudera-manager-\*' \

avro-tools crunch flume-ng \

hadoop-hdfs-fuse hadoop-hdfs-nfs3 \

hadoop-httpfs hbase-solr hive-hbase \

hive-webhcat hue-beeswax hue-hbase \

hue-impala hue-pig hue-plugins hue-rdbms \

hue-search hue-spark hue-sqoop \

hue-zookeeper impala impala-shell kite \

llama mahout oozie pig pig-udf-datafu \

search sentry solr-mapreduce \

spark-python sqoop sqoop2 whirr

清空文件及yum缓存：

sudo rm -Rf /usr/share/cmf /var/lib/cloudera\* /var/cache/yum/cloudera\*

$ sudo yum clean all

# 附录节点划分参考

