# 目录

[目录](#_Toc1064917160)

[1. 环境准备（root用户）](#_Toc48059732)

[1.1. 机器hosts修改](#_Toc1053222110)

[1.2. 配置免密登录](#_Toc305304696)

[1.3. 关闭防火墙、SELINUX](#_Toc1511575556)

[1.4. NTP时钟同步](#_Toc1444705583)

[1.5. cloudera-manager安装](#_Toc565354368)

[1.6. 创建cloudera-manager用户](#_Toc251397528)

[2. cloudera-manager安装](#_Toc861455726)

[2.1. 配置mysql数据库连接驱动](#_Toc1391706517)

[2.2. 创建cloudera-manager数据库及用户](#_Toc1057133821)

[2.3. 检查cloudera-manager数据库连接](#_Toc1428122513)

[2.4. 启动cloudera-scm-server](#_Toc854248693)

[2.5. 启动cloudera-scm- agent](#_Toc1497347708)

[3. 大数据平台配置](#_Toc2120035549)

[3.1. 监控安装](#_Toc1706075619)

[4. Web访问地址](#_Toc1090390689)

# 环境准备（root用户）

## 机器hosts修改

修改10.1.20.71的hosts

sudo vi /etc/hosts

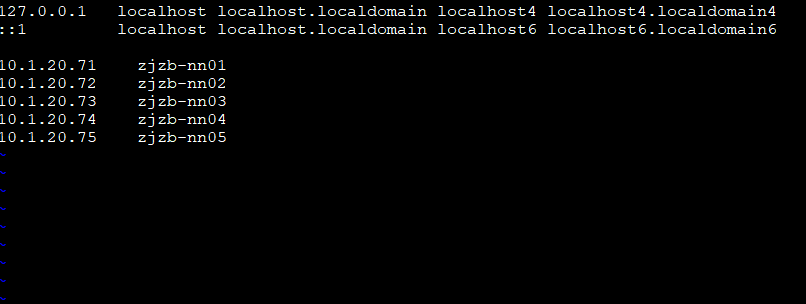
10.1.20.71 zjzb-nn01

10.1.20.72 zjzb-nn02

10.1.20.73 zjzb-nn03

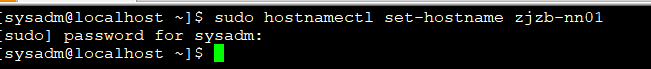
10.1.20.74 zjzb-nn04

10.1.20.75 zjzb-nn05



更新主机名

sudo hostnamectl set-hostname zjzb-nn01



配置10.1.20.72 的hosts

sudo vi /etc/hosts

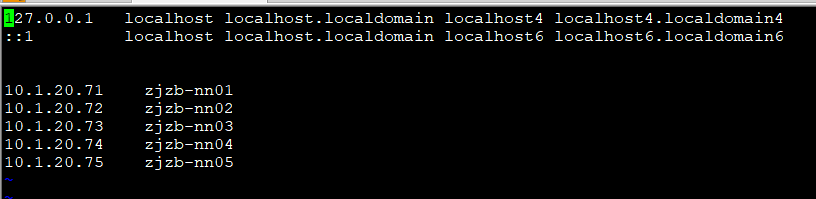
10.1.20.71 zjzb-nn01

10.1.20.72 zjzb-nn02

10.1.20.73 zjzb-nn03

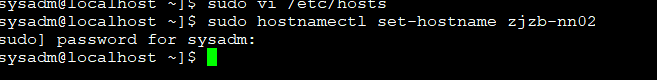
10.1.20.74 zjzb-nn04

10.1.20.75 zjzb-nn05



更新主机名

sudo hostnamectl set-hostname zjzb-nn02



配置10.1.20.73 的hosts

sudo vi /etc/hosts

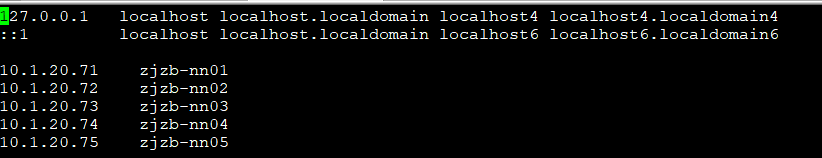
10.1.20.71 zjzb-nn01

10.1.20.72 zjzb-nn02

10.1.20.73 zjzb-nn03

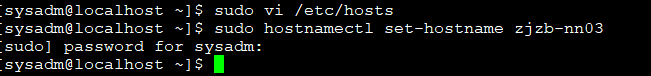
10.1.20.74 zjzb-nn04

10.1.20.75 zjzb-nn05



更新主机名

sudo hostnamectl set-hostname zjzb-nn03



配置10.1.20.74的hosts

sudo vi /etc/hosts

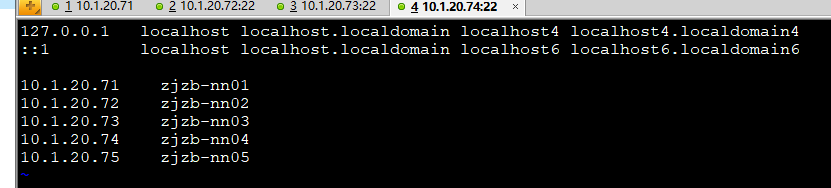
10.1.20.71 zjzb-nn01

10.1.20.72 zjzb-nn02

10.1.20.73 zjzb-nn03

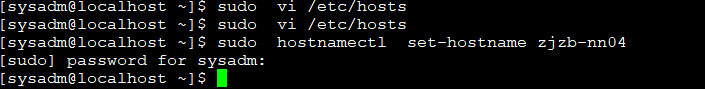
10.1.20.74 zjzb-nn04

10.1.20.75 zjzb-nn05



更新主机名

sudo hostnamectl set-hostname zjzb-nn04



配置10.1.20.75的hosts

sudo vi /etc/hosts

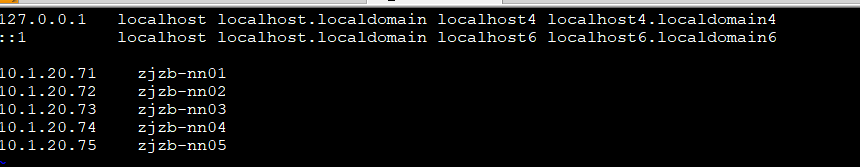
10.1.20.71 zjzb-nn01

10.1.20.72 zjzb-nn02

10.1.20.73 zjzb-nn03

10.1.20.74 zjzb-nn04

10.1.20.75 zjzb-nn05



更新主机名

sudo hostnamectl set-hostname zjzb-nn05

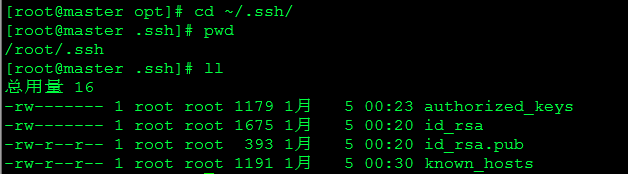


## 配置免密登录

所有节点执行

ssh-keygen

命令执行期间，都直接输入“enter”只到结束



ssh-copy-id root@10.1.20.71

ssh-copy-id root@10.1.20.72

ssh-copy-id root@10.1.20.73

ssh-copy-id root@10.1.20.74

ssh-copy-id [root@10.1.20.75](mailto:root@10.1.20.75)

## 关闭防火墙、SELINUX

sudo systemctl stop firewalld

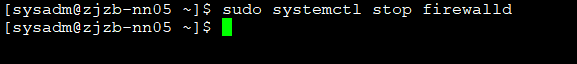
\*\*考虑线上环境，可以开放相关ip或者开放相关端口









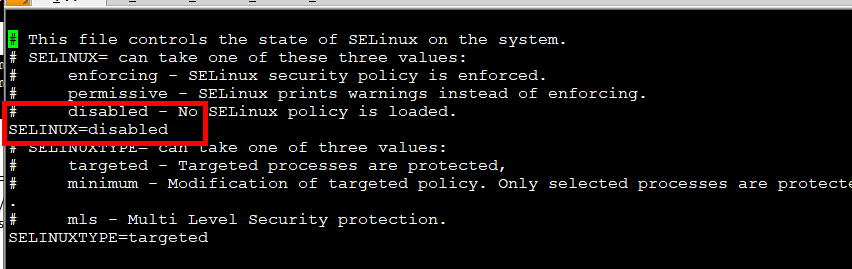


sudo systemctl disable firewalld

以root用户 在10.1.20.71 上执行以下命令

sudo vi /etc/selinux/config

SELINUX=disabled

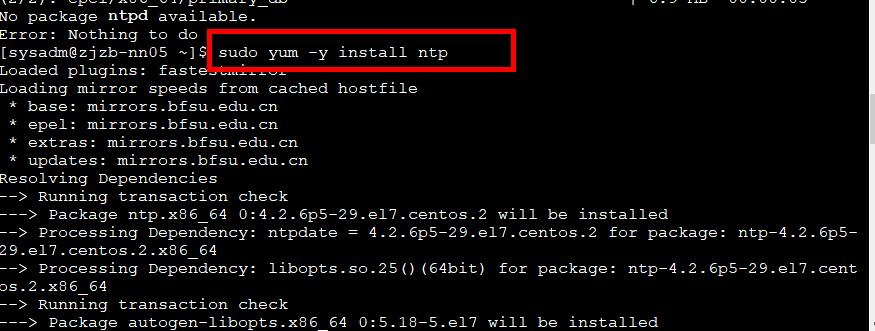


10.1.20.72, 10.1.20.73,10.1.20.74,10.1.20.75 同上

## NTP时钟同步

安装 10.1.20.71的 ntpd服务

sudo yum -y install ntp

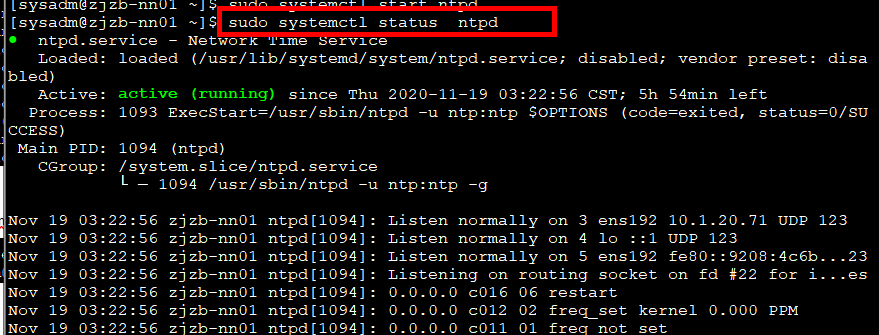


启动

sudo systemctl start ntpd

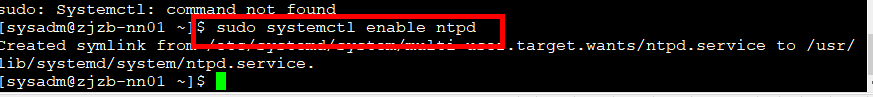


检查 ntpd 服务状态



设置ntpd 开机自启

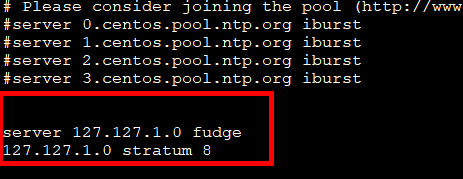
sudo systemctl enable ntpd



10.1.20.72 ,10.1.20.73,10.1.20.74,10.1.20.75 同上

在server端 10.1.20.71 上修改/etc/ntp.conf文件

sudo vi /etc/ntp.conf



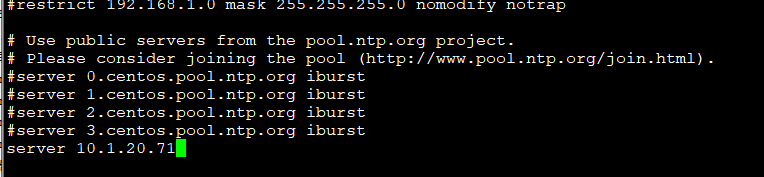
重启 server（10.1.20.71）端 ntpd 服务

sudo systemctl restart ntpd



在10.1.20.71-75机器上， 进行ntp client 配置

sudo su - -c ‘vi /etc/ntp.conf’



sudo su - -c ‘systemctl restart ntpd’

从机 10.1.20.72,10.1.20.73,10.1.20.74,10.1.20.75 手动测试 同步 ntp服务器时间

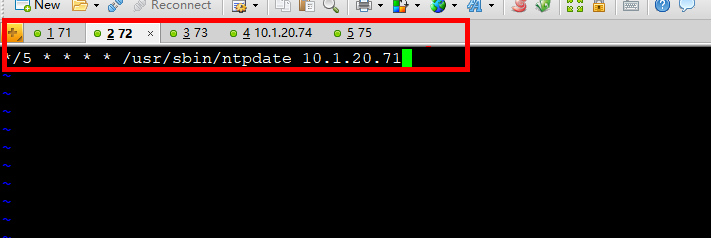
执行：sudo ntpdate -u 10.1.20.71

从机10.1.20.72,10.1.20.73,10.1.20.74,10.1.20.75 配置crontab自动同步

执行：sudo crontab -e

写入

\*/5 \* \* \* \* /usr/sbin/ntpdate 10.1.20.71



## cloudera-manager安装

 解压 安装包

sudo unzip /opt/CDH\_Centos7.zip

移动文件到/data/下

sudo mv /opt/CDH\_Centos7/\* /data

将 cloudera-manager-centos7-cm5.15.1\_x86\_64.tar.gz 拷贝到其他四台机器， 10.1.20.72 ,10.1.20.73,10.1.20.74,10.1.20.75

sudo scp -r cloudera-manager-centos7-cm5.15.1\_x86\_64.tar.gz 10.1.20.72:/data/

sudo scp -r cloudera-manager-centos7-cm5.15.1\_x86\_64.tar.gz 10.1.20.73:/data/

sudo scp -r cloudera-manager-centos7-cm5.15.1\_x86\_64.tar.gz 10.1.20.74:/data/

sudo scp -r cloudera-manager-centos7-cm5.15.1\_x86\_64.tar.gz 10.1.20.75:/data/

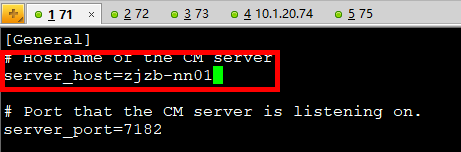
解压 10.1.20.71,10.1.20.72,10.1.20.73,10.1.20.74,10.1.20.75 tar.gz文件

sudo tar -xvf cloudera-manager-centos7-cm5.15.1\_x86\_64.tar.gz

在各个节点执行

sudo vi /data/ cm-5.15.1/etc/clo*udera-scm-agent/config.in*i

修改 server\_host=zjzb-nn01



在zjzb-nn01 上移动软件包

mv /data/CDH-5.15.1-1.cdh5.15.1.p0.4-el7.parcel /data/cloudera/parcel-repo/

mv /data/CDH-5.15.1-1.cdh5.15.1.p0.4-el7.parcel.sha1 /data/cloudera/parcel-repo/CDH-5.15.1-1.cdh5.15.1.p0.4-el7.parcel.sha

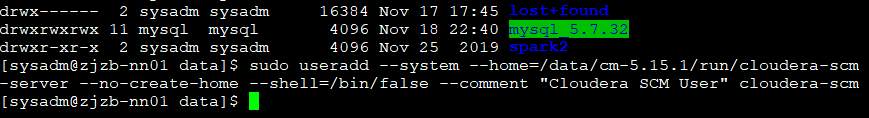
mv /data/manifest.json /data/cloudera/parcel-repo/

## 创建cloudera-manager用户

所有节点执行

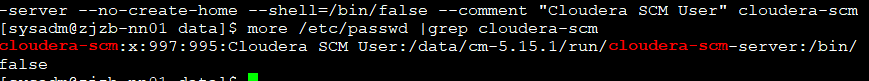
sudo su - -c ‘useradd --system --home=/data/cm-5.15.1/run/cloudera-scm-server --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm’

注：伪用户，cloudera-scm是用来运行cloudera-manager

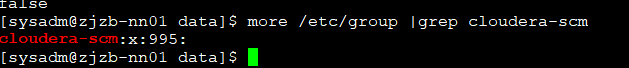


检查

more /etc/passwd |grep cloudera-scm



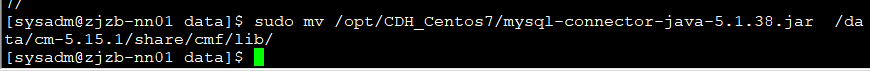
more /etc/group |grep cloudera-scm



# cloudera-manager安装

## 配置mysql数据库连接驱动

sudo mv /opt/CDH\_Centos7/mysql-connector-java-5.1.38.jar /data/cm-5.15.1/share/cmf/lib/



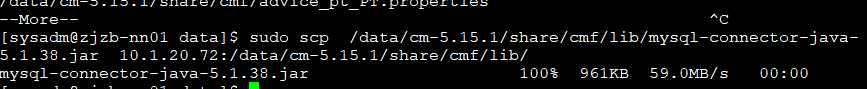
拷贝 mysql 驱动到 72,73,74,75 /data/cm-5.15.1/share/cmf/lib/

sudo scp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar 10.1.20.72:/data/cm-5.15.1/share/cmf/lib/

sudo scp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar 10.1.20.73:/data/cm-5.15.1/share/cmf/lib/

sudo scp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar 10.1.20.74:/data/cm-5.15.1/share/cmf/lib/

sudo scp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar 10.1.20.75:/data/cm-5.15.1/share/cmf/lib/



## 创建cloudera-manager数据库及用户

创建scm后要给cm库授权操作

并编辑vi /etc/my.cnf

修改内容如下

gtid\_mode = ON

enforce\_gtid\_consistency = ON

改为

gtid\_mode = OFF

enforce\_gtid\_consistency = OFF

sudo su - -c '/data/cm-5.15.1/share/cmf/schema/scm\_prepare\_database.sh mysql cm -hzjzb-nn01:3306 -uroot -pDev@123 --scm-host zjzb-nn01 scm scm'

注意：每次执行失败后，一定要删除cm库；

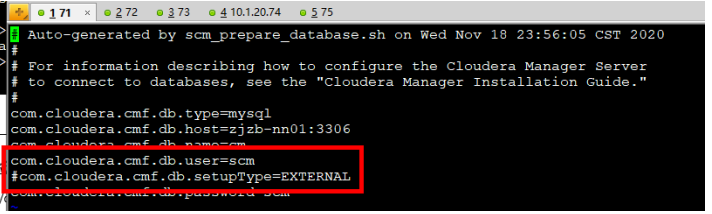
注：

|  |
| --- |
| /opt/cm-5.3.2/share/cmf/schema/scm\_prepare\_database.sh mysql 库名 -h数据库ip:端口 -u数据库用户 -p数据库密码 --scm-host cloudera-manager的ip scm mysql新建的用户 mysql新建的用户的密码 |
|  |
| 查看数据库是否创建  Mysql -uroot -pDev@123 -h zjzb-nn01 |

## 检查cloudera-manager数据库连接

sudo vi /data/cm-5.15.1/etc/cloudera-scm-server/db.properties

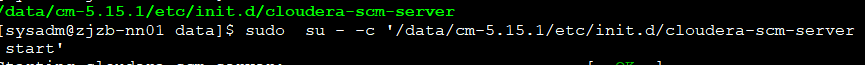
注释掉下图标记的行



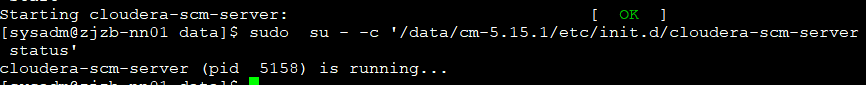
## 启动cloudera-scm-server

在zjzb-nn01上执行

sudo su - -c '/data/cm-5.15.1/etc/init.d/cloudera-scm-server start'

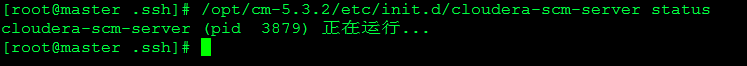


sudo su - -c '/data/cm-5.15.1/etc/init.d/cloudera-scm-server status'



端口检查

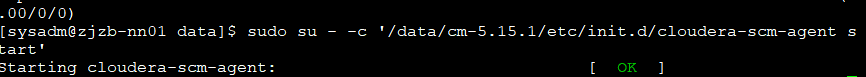
netstat -ano | grep 7180 可能会比较慢



## 启动cloudera-scm- agent

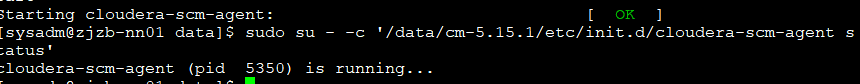
所有节点上执行

sudo su - -c ‘/data/cm-5.15.1/etc/init.d/cloudera-scm-agent start’



状态检查

sudo su - -c '/data/cm-5.15.1/etc/init.d/cloudera-scm-agent status'

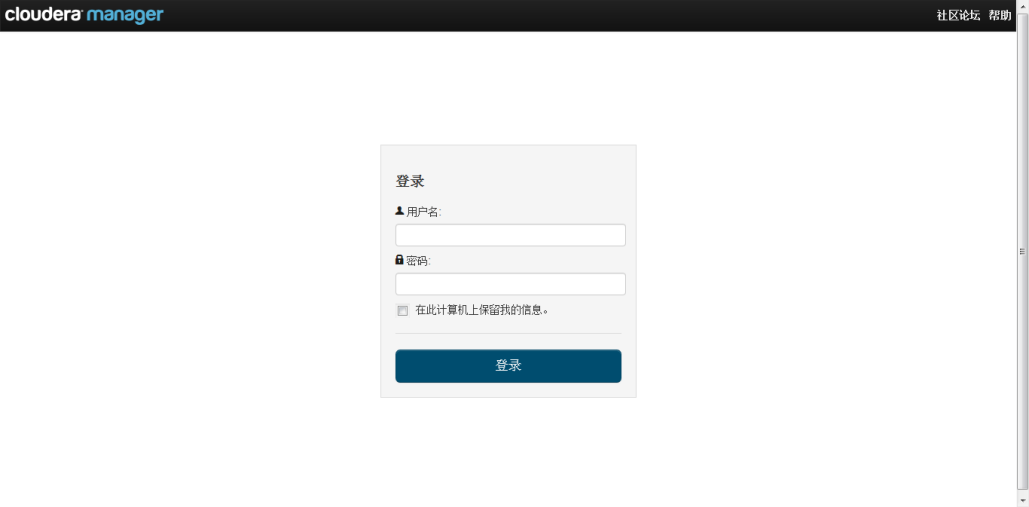


# 大数据平台配置

## 监控安装

浏览器登录

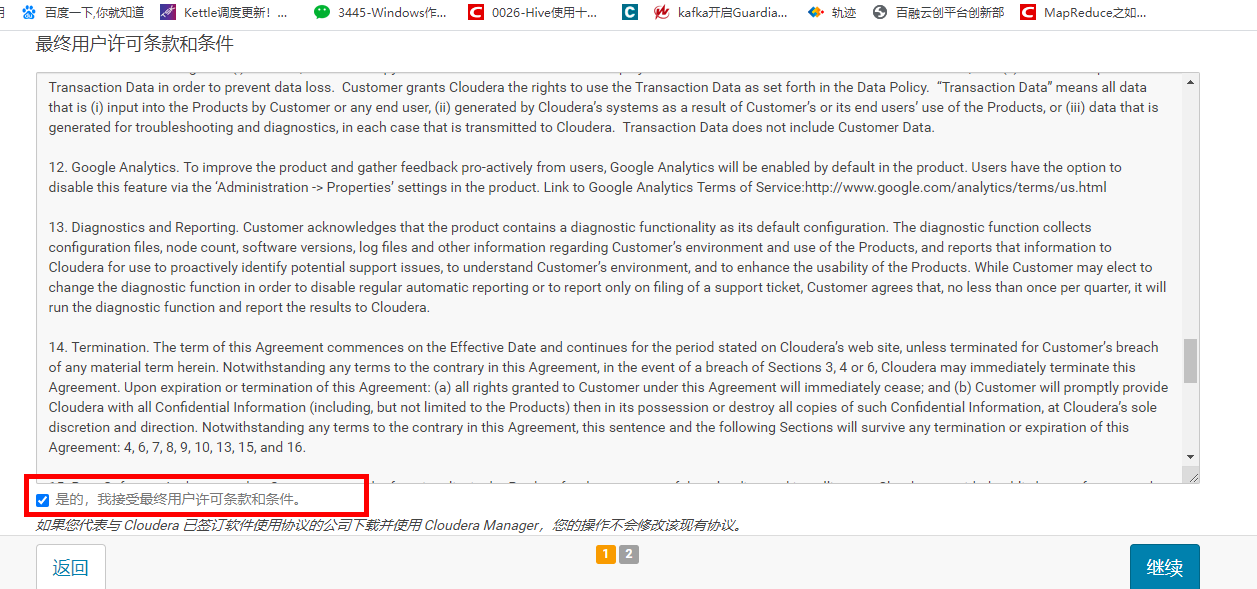
http://10.1.20.71:7180

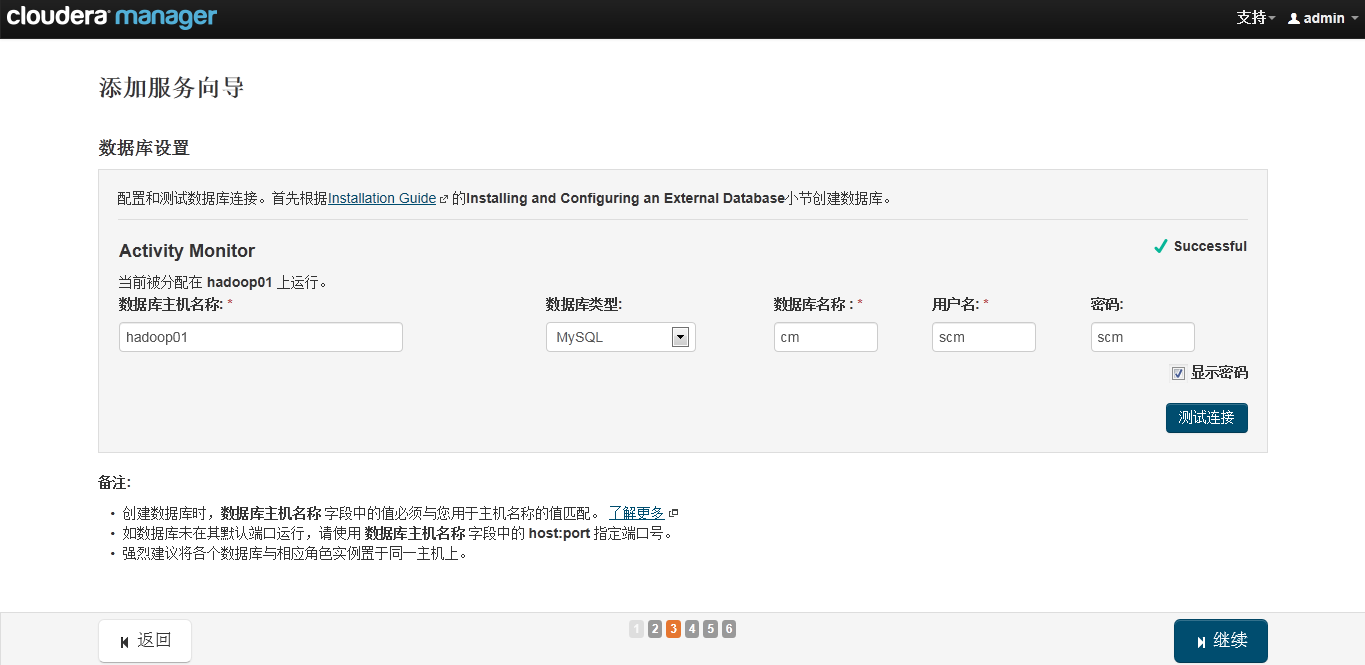


默认

用户名：admin

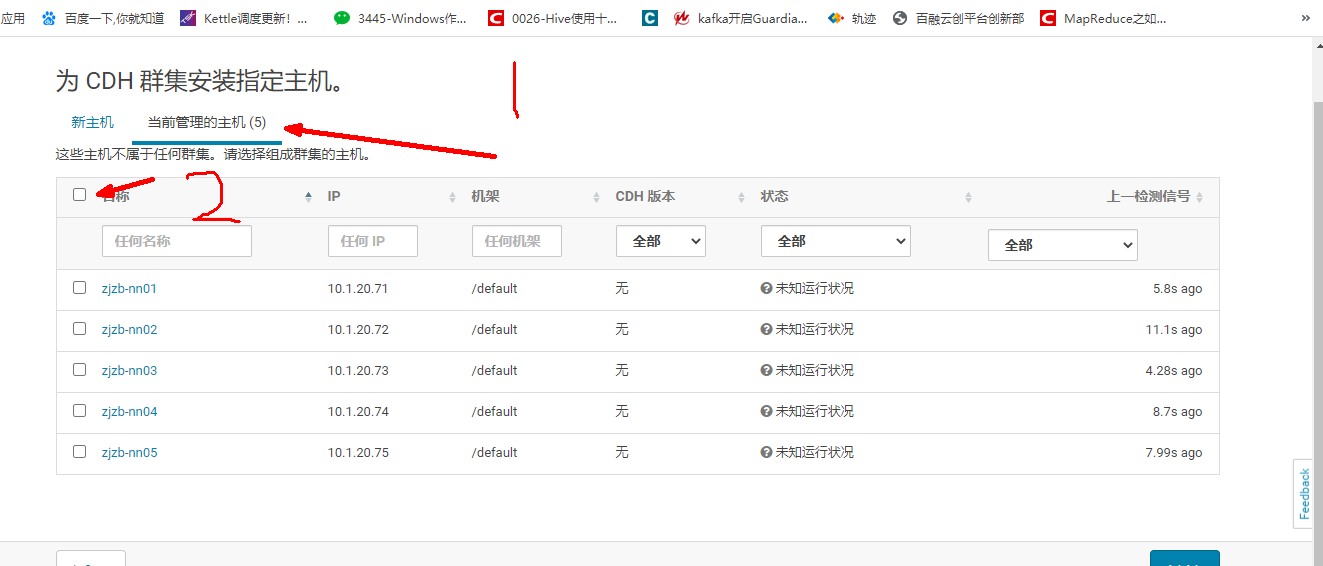
密 码：admin

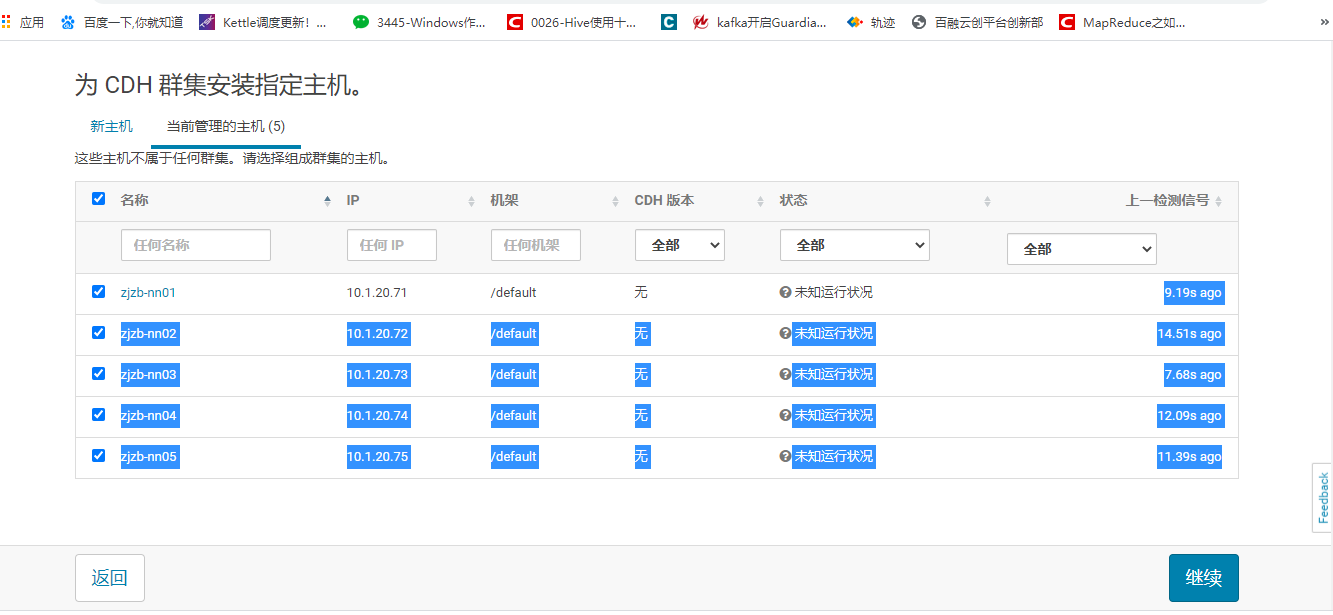






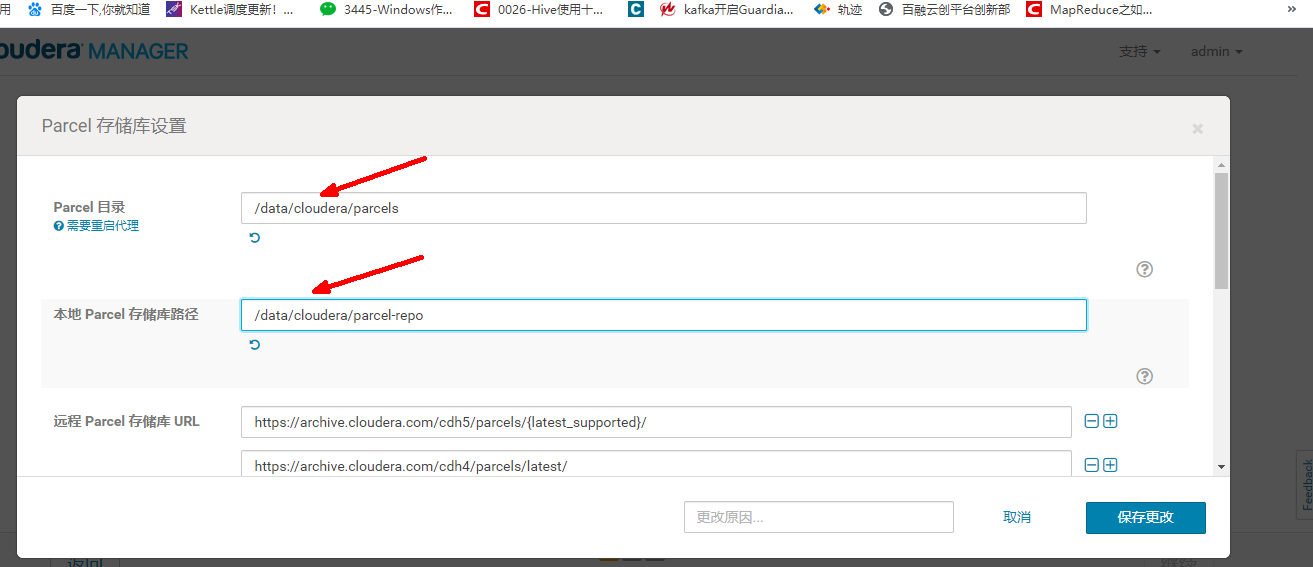
.





点击“继续”





10.1.20.71 重启cloudera-manager server

sudo su - -c ‘/data/cm-5.15.1/etc/init.d/cloudera-scm-server restart’

重复上一步

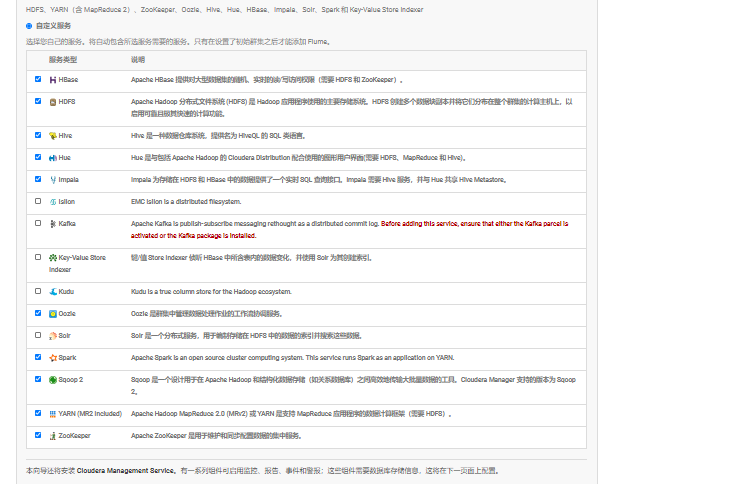


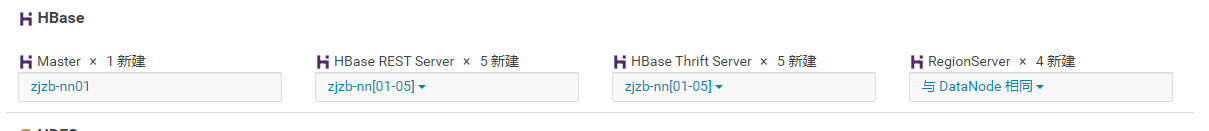


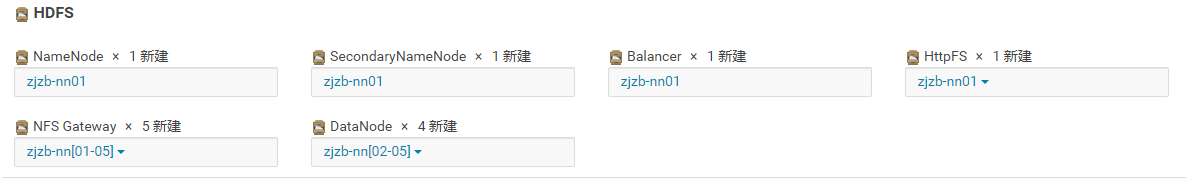


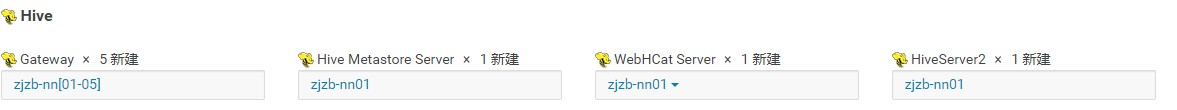
sudo su - -c 'echo never > /sys/kernel/mm/transparent\_hugepage/defrag'

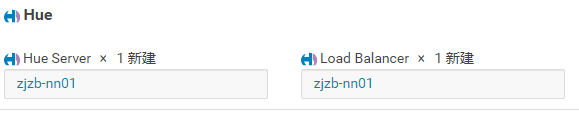
sudo su - -c 'echo never > /sys/kernel/mm/transparent\_hugepage/enabled'

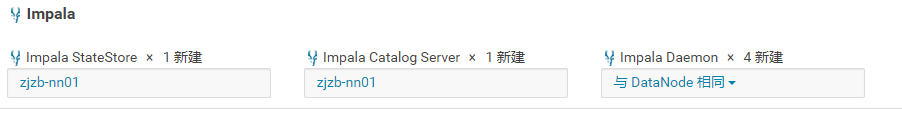


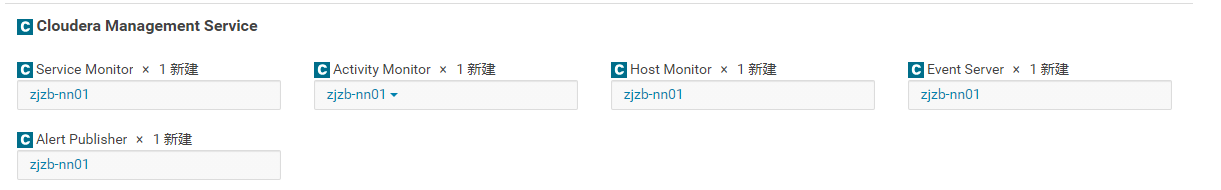




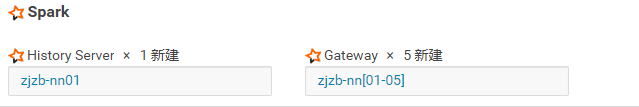




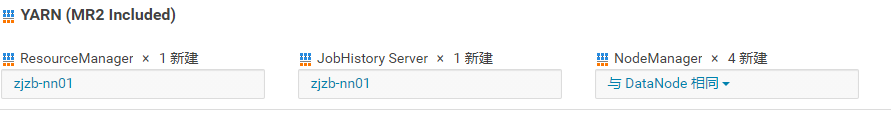




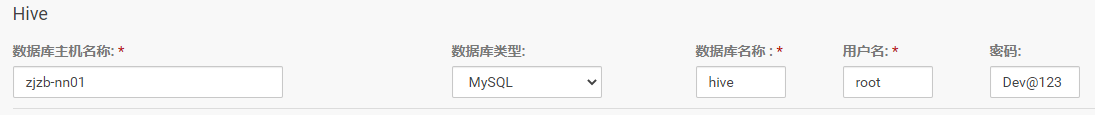


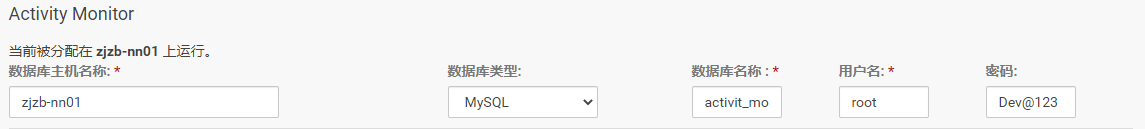




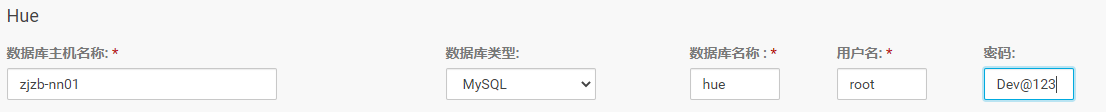












创建数据库

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

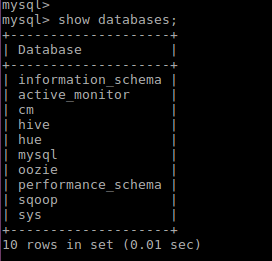
mysql> create database active\_monitor default character set utf8 collate utf8\_general\_ci;

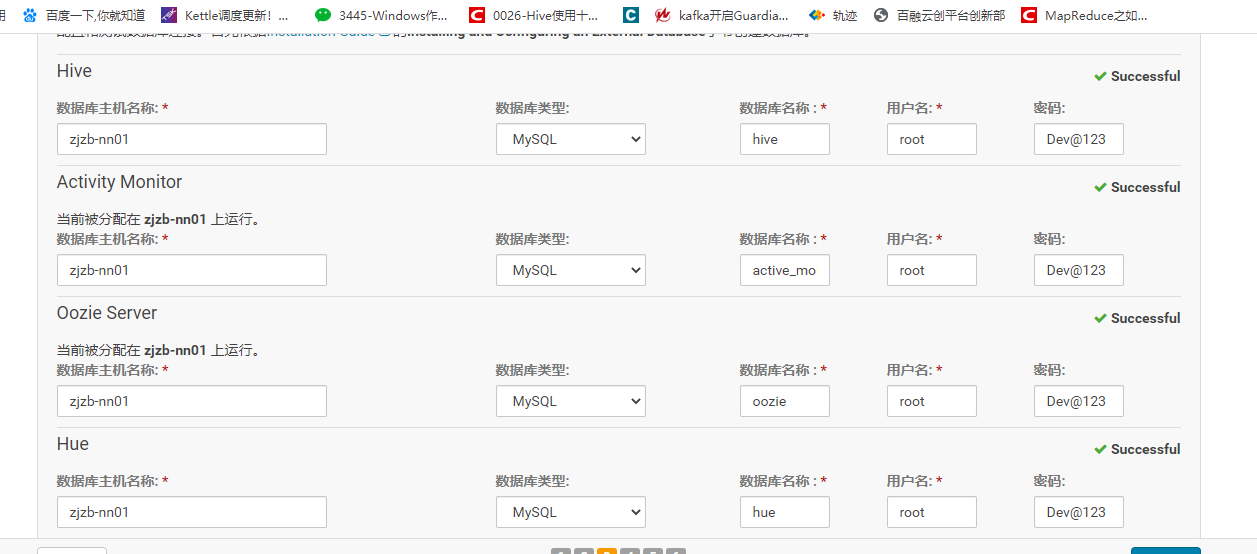
mysql> create database oozie default character set utf8 collate utf8\_general\_ci;

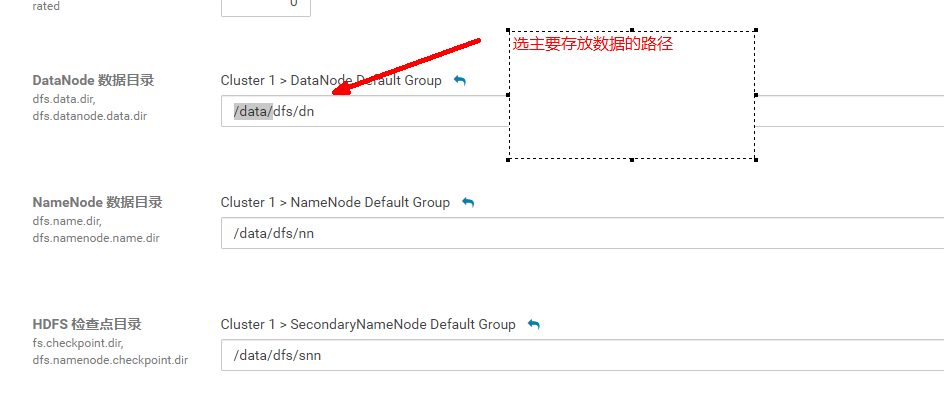
mysql> create database hue default character set utf8 collate utf8\_general\_ci;

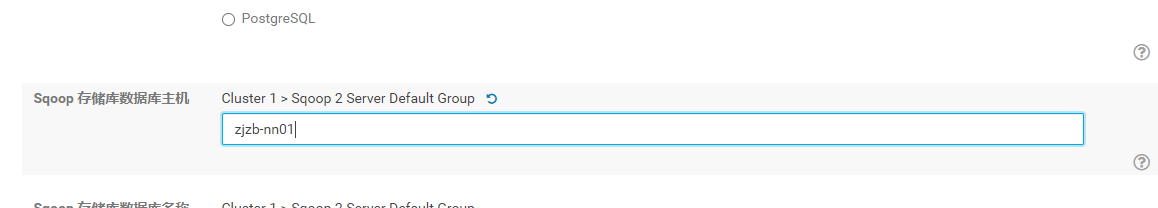
mysql> create database reports\_manager default character set utf8 collate utf8\_general\_ci;

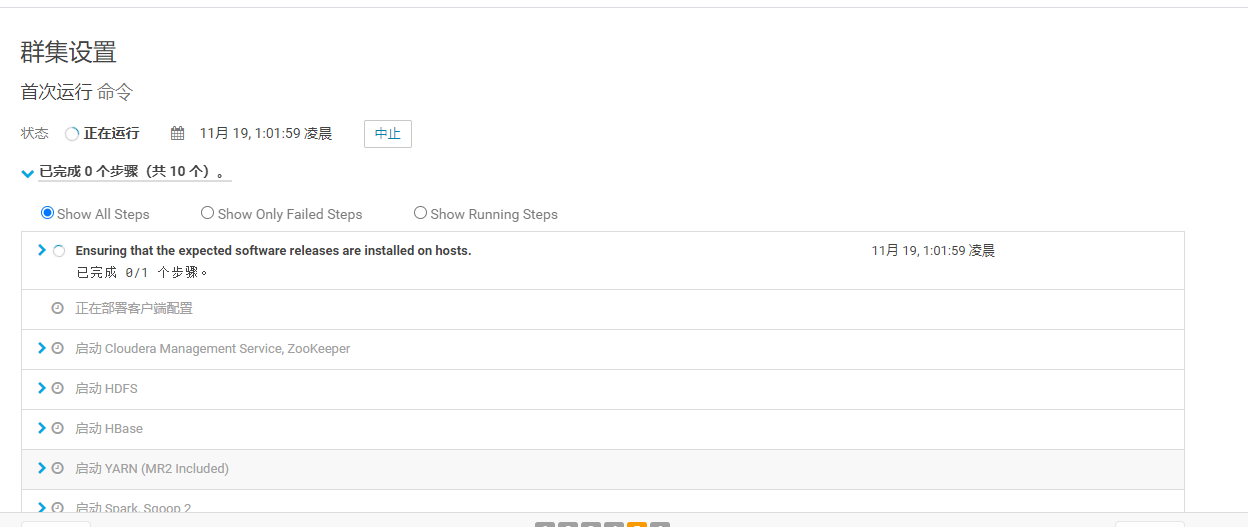
mysql> create database sqoop default character set utf8 collate utf8\_general\_ci;







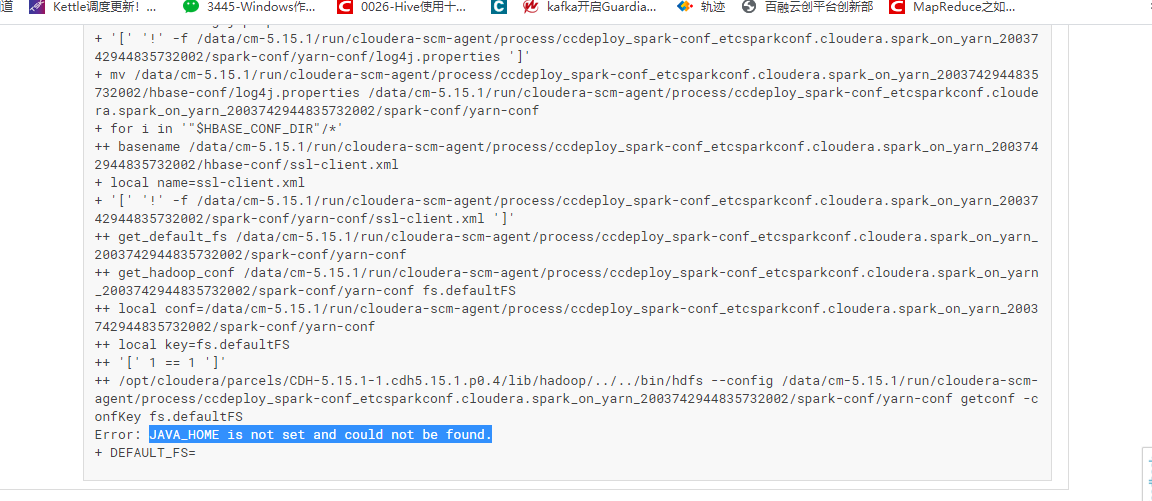




等待启动完成

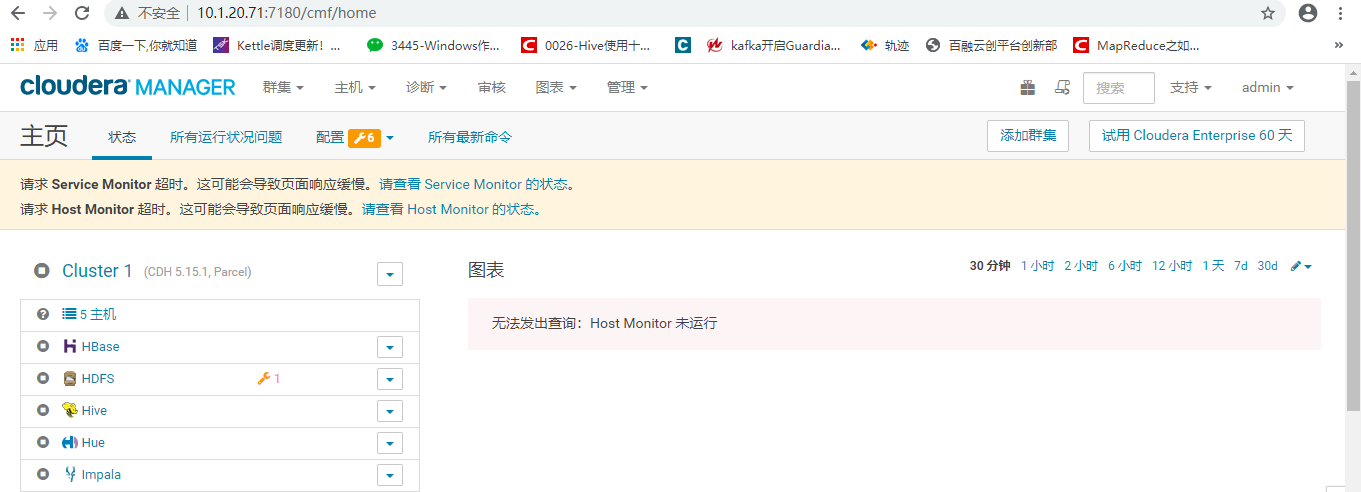
以下为常见问题解决方法：

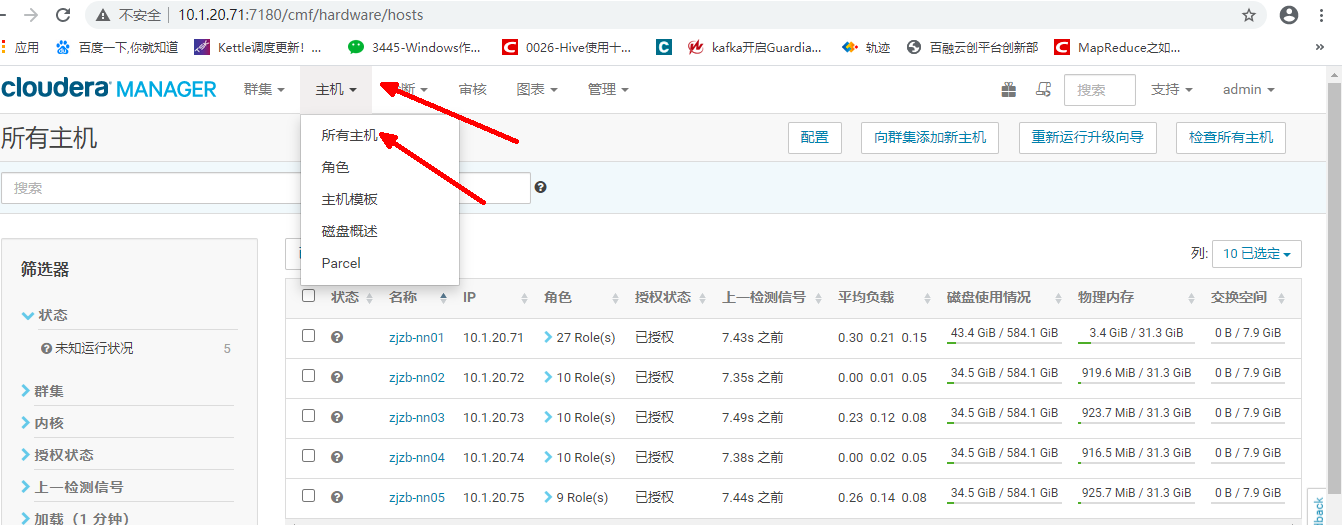
问题一：

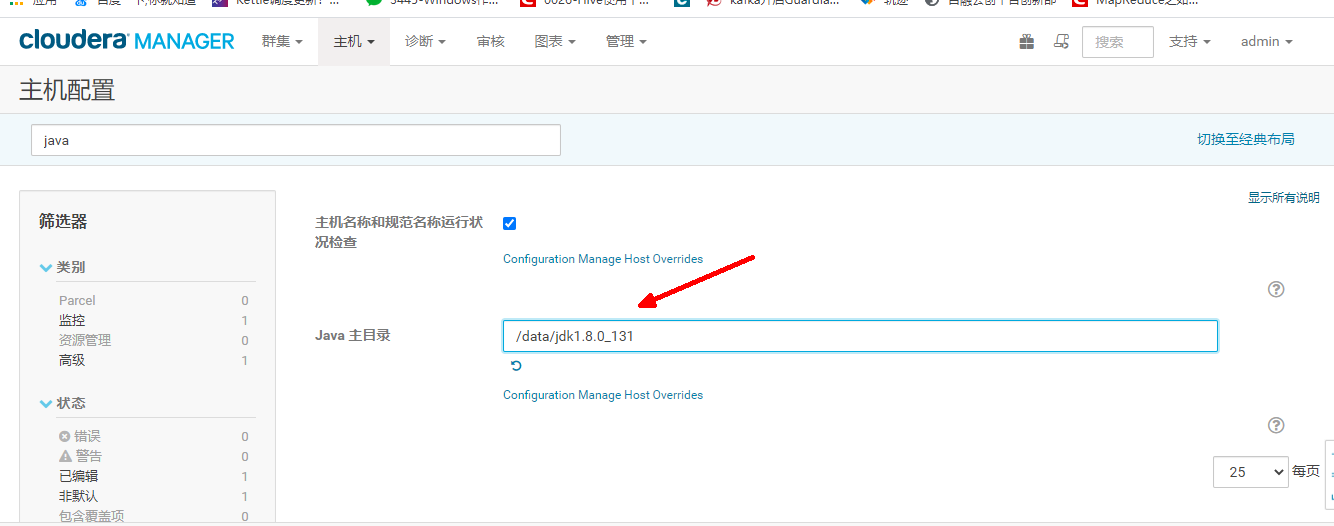


方法一：

打开新的标签页，访问10.1.20.71:7180







问题二：



sudo su - -c 'cp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hive/lib/'

问题三：



sudo su - -c 'cp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar /var/lib/oozie/'

同理解决sqoop数据库连接问题：

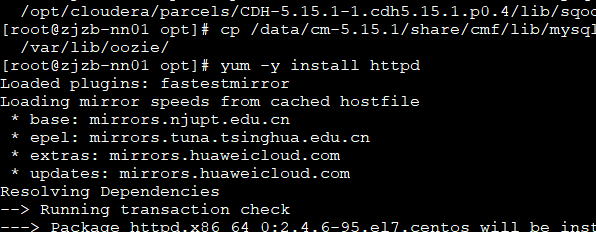
sudo su - -c 'cp /data/cm-5.15.1/share/cmf/lib/mysql-connector-java-5.1.38.jar /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/sqoop/lib/'

等待拷贝完成

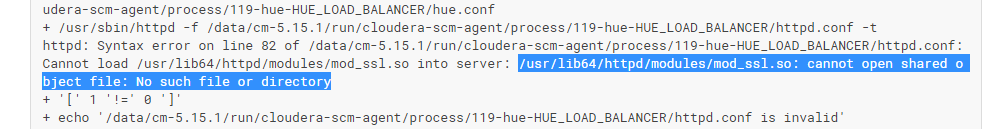
问题四：



对应机器(zjzb-nn01)执行 ： yum -y install httpd

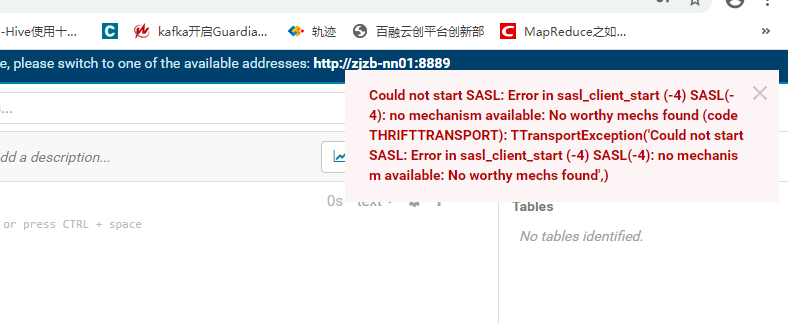


问题五：



对应机器执行： yum -y install mod\_ssl

问题六：



对应机器执行：

yum install cyrus-sasl-plain cyrus-sasl-devel cyrus-sasl-gssapi

# Web访问地址

HDFS namenode   http://10.1.20.71:50070/

HDFS datanode    http://10.1.20.72:50075/

YARN http://10.1.20.71:8088/cluster

HBASE master  http://10.1.20.71:60010/

HBASE regionserver    http://10.1.20.72:60030/

cloudera manager地址：http://10.1.20.71:7180 用户名/密码    admin/admin

Hadoop集群的各部分一般都会使用到多个端口，有些是daemon之间进行交互之用，有些是用于RPC访问以及HTTP访问。而随着Hadoop周边组件的增多，完全记不住哪个端口对应哪个应用，特收集记录如此，以便查询。

这里包含我们使用到的组件：HDFS, YARN, HBase, Hive, ZooKeeper:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 组件 | 节点 | 默认端口 | 配置 | 用途说明 |
| HDFS | DataNode | 50010 | dfs.datanode.address | datanode服务端口，用于数据传输 |
| HDFS | DataNode | 50075 | dfs.datanode.http.address | http服务的端口 |
| HDFS | DataNode | 50475 | dfs.datanode.https.address | https服务的端口 |
| HDFS | DataNode | 50020 | dfs.datanode.ipc.address | ipc服务的端口 |
| HDFS | NameNode | 50070 | dfs.namenode.http-address | http服务的端口 |
| HDFS | NameNode | 50470 | dfs.namenode.https-address | https服务的端口 |
| HDFS | NameNode | 8020 | fs.defaultFS | 接收Client连接的RPC端口，用于获取文件系统metadata信息。 |
| HDFS | journalnode | 8485 | dfs.journalnode.rpc-address | RPC服务 |
| HDFS | journalnode | 8480 | dfs.journalnode.http-address | HTTP服务 |
| HDFS | ZKFC | 8019 | dfs.ha.zkfc.port | ZooKeeper FailoverController，用于NN HA |
| YARN | ResourceManager | 8032 | yarn.resourcemanager.address | RM的applications manager(ASM)端口 |
| YARN | ResourceManager | 8030 | yarn.resourcemanager.scheduler.address | scheduler组件的IPC端口 |
| YARN | ResourceManager | 8031 | yarn.resourcemanager.resource-tracker.address | IPC |
| YARN | ResourceManager | 8033 | yarn.resourcemanager.admin.address | IPC |
| YARN | ResourceManager | 8088 | yarn.resourcemanager.webapp.address | http服务端口 |
| YARN | NodeManager | 8040 | yarn.nodemanager.localizer.address | localizer IPC |
| YARN | NodeManager | 8042 | yarn.nodemanager.webapp.address | http服务端口 |
| YARN | NodeManager | 8041 | yarn.nodemanager.address | NM中container manager的端口 |
| YARN | JobHistory Server | 10020 | mapreduce.jobhistory.address | IPC |
| YARN | JobHistory Server | 19888 | mapreduce.jobhistory.webapp.address | http服务端口 |
| HBase | Master | 60000 | hbase.master.port | IPC |
| HBase | Master | 60010 | hbase.master.info.port | http服务端口 |
| HBase | RegionServer | 60020 | hbase.regionserver.port | IPC |
| HBase | RegionServer | 60030 | hbase.regionserver.info.port | http服务端口 |
| HBase | HQuorumPeer | 2181 | hbase.zookeeper.property.clientPort | HBase-managed ZK mode，使用独立的ZooKeeper集群则不会启用该端口。 |
| HBase | HQuorumPeer | 2888 | hbase.zookeeper.peerport | HBase-managed ZK mode，使用独立的ZooKeeper集群则不会启用该端口。 |
| HBase | HQuorumPeer | 3888 | hbase.zookeeper.leaderport | HBase-managed ZK mode，使用独立的ZooKeeper集群则不会启用该端口。 |
| Hive | Metastore | 9083 | /etc/default/hive-metastore中export PORT=<port>来更新默认端口 |  |
| Hive | HiveServer | 10000 | /etc/hive/conf/hive-env.sh中export HIVE\_SERVER2\_THRIFT\_PORT=<port>来更新默认端口 |  |
| ZooKeeper | Server | 2181 | /etc/zookeeper/conf/zoo.cfg中clientPort=<port> | 对客户端提供服务的端口 |
| ZooKeeper | Server | 2888 | /etc/zookeeper/conf/zoo.cfg中server.x=[hostname]:nnnnn[:nnnnn]，标蓝部分 | follower用来连接到leader，只在leader上监听该端口。 |
| ZooKeeper | Server | 3888 | /etc/zookeeper/conf/zoo.cfg中server.x=[hostname]:nnnnn[:nnnnn]，标蓝部分 | 用于leader选举的。只在electionAlg是1,2或3(默认)时需要。 |

所有端口协议均基于TCP。

对于存在Web UI（HTTP服务）的所有hadoop daemon，有如下url：

**/logs**   
日志文件列表，用于下载和查看

**/logLevel**   
允许你设定log4j的日志记录级别，类似于hadoop daemonlog

**/stacks**   
所有线程的stack trace，对于debug很有帮助

**/jmx**   
服务端的Metrics，以JSON格式输出。

/jmx?qry=Hadoop:\*会返回所有hadoop相关指标。   
/jmx?get=MXBeanName::AttributeName 查询指定bean指定属性的值，例如/jmx?get=Hadoop:service=NameNode,name=NameNodeInfo::ClusterId会返回ClusterId。   
这个请求的处理类：org.apache.hadoop.jmx.JMXJsonServlet

而特定的Daemon又有特定的URL路径特定相应信息。

**NameNode**:http://10.1.20.71:50070/

**/dfshealth.jsp**   
HDFS信息页面，其中有链接可以查看文件系统

/dfsnodelist.jsp?whatNodes=(DEAD|LIVE)   
显示DEAD或LIVE状态的datanode

**/fsck**   
运行fsck命令，不推荐在集群繁忙时使用！

**DataNode**:http://10.1.20.72:50075/

**/blockScannerReport**   
每个datanode都会指定间隔验证块信息