Nn1:

yum groupinstall "Desktop"

yum groupinstall "X Window System"

yum groupinstall "Chinese Support"

create database amon DEFAULT CHARACTER SET utf8;

grant all on amon.\* TO 'amon'@'%' IDENTIFIED BY 'amon\_password';

create database rman DEFAULT CHARACTER SET utf8;

grant all on rman.\* TO 'rman'@'%' IDENTIFIED BY 'rman\_password';

create database metastore DEFAULT CHARACTER SET utf8;

grant all on metastore.\* TO 'hive'@'%' IDENTIFIED BY 'hive\_password';

create database sentry DEFAULT CHARACTER SET utf8;

grant all on sentry.\* TO 'sentry'@'%' IDENTIFIED BY 'sentry\_password';

create database nav DEFAULT CHARACTER SET utf8;

grant all on nav.\* TO 'nav'@'%' IDENTIFIED BY 'nav\_password';

create database scm DEFAULT CHARACTER SET utf8;

grant all on scm.\* TO 'scm'@'%' IDENTIFIED BY 'scm\_password';

mkdir /usr/share/mysql/java/

mv mysql-connector-java-5.1.25-bin.jar /usr/share/mysql/java/

mysqldump -h192.168.152.181 -uroot -phadoop mysql > /data/database-backup.sql

yum install cloudera-manager-server-db-2

vim /etc/cloudera-scm-server/db.properties

# The database type

# Currently 'mysql', 'postgresql' and 'oracle' are valid databases.

com.cloudera.cmf.db.type=mysql

# The database host

# If a non standard port is needed, use 'hostname:port'

com.cloudera.cmf.db.host=192.168.152.181:3306

# The database name

com.cloudera.cmf.db.name=cmf

# The database user

com.cloudera.cmf.db.user=cmf

# The database user's password

com.cloudera.cmf.db.password=scm\_password

service cloudera-scm-server-db start

service cloudera-scm-server start

chkconfig cloudera-scm-server on

/usr/share/cmf/schema/scm\_prepare\_database.sh mysql scm scm scm\_password

/usr/share/cmf/schema/scm\_prepare\_database.sh mysql -h 192.168.152.181 -ucm -pcm --scm-host 192.168.152.181 scm scm scm\_password

/usr/share/cmf/schema/scm\_prepare\_database.sh mysql -h 192.168.152.181 -u root -phadoop --scm-host 192.168.152.181 scm scm scm\_password

wget <http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin>

yum --enablerepo=cloudera-manager clean metadata

./cloudera-manager-installer.bin --skip\_repo\_package=1

删除软件：

<http://www.cloudera.com/content/cloudera/en/documentation/cloudera-manager/v5-latest/Cloudera-Manager-Installation-Guide/cm5ig_uninstall_cm.html#cmig_topic_18>

卸载 Cloudera Manager 5.1.x.和 相关软件【官网翻译】

<http://www.douban.com/note/413871846/>

Jar包：

cd /usr/share/cmf/lib

cp ~/mysql-connector-java-5.1.25-bin.jar .

service cloudera-scm-server start

chkconfig cloudera-scm-server on

agent节点安装:

yum install cloudera-manager-agent cloudera-manager-daemons

vim /etc/cloudera-scm-agent/config.ini

server\_host Name of host where the Cloudera Manager Server is running.

server\_port Port on host where the Cloudera Manager Server is running.

Cloudera5占用的端口：

<http://www.cloudera.com/content/cloudera/en/documentation/cloudera-manager/v5-latest/Cloudera-Manager-Installation-Guide/cm5ig_ports_cm.html>

service cloudera-scm-agent start

chkconfig cloudera-scm-agent on

mv local-hadoop-cm5.repo cloudera-manager.repo

# 修改hosts文件（也是必须的）

必须严格按照以下这种格式 ip aaa.bbb.ccc aaa

例如：116.255.234.198a234-198.hadoop.com a234-198

192.168.152.181 a152-181.hadoop.com a152-181

192.168.152.182 a152-182.hadoop.com a152-182

192.168.152.183 a152-183.hadoop.com a152-183

修改主机名（必须的，所有主机全部需要修改,而且必须遵守类似这种格式 a234-198.hadoop.com）

[root@a234-198 yum.repos.d]# vi /etc/sysconfig/network

NETWORKING=yes

HOSTNAME=a234-198.hadoop.com

安装zookeeper

mkdir -p /data/data/zookeeper

mkdir -p /data/log/zookeeper

mkdir /var/lib/zookeeper/version-2

hown zookeeper.zookeeper /var/lib/zookeeper/version-2

mkdir -p /dfs/dn /data/dfs/dn /dfs/nn /data/dfs/nn /dfs/snn

chown -R hdfs.hdfs /dfs/dn /data/dfs/dn /dfs/nn /data/dfs/nn /dfs/snn

mkdir -p /yarn/nm /data/yarn/nm

chown -R yarn.yarn /yarn/nm /data/yarn/nm

mkdir -p /mapred/local /data/mapred/local /mapred/jt

chown -R mapred.mapred /mapred/local /data/mapred/local /mapred/jt

mkdir -p /impala/impalad /data/impala/impalad

chown -R impala.impala /impala/impalad /data/impala/impalad

useradd zookeeper

yum install -y bigtop\*

yum install zookeeper\* -y

hdfs:

yum install \*hdfs\* -y

mkdir -p /data/dn/

mkdir -p /data/nns/nn/

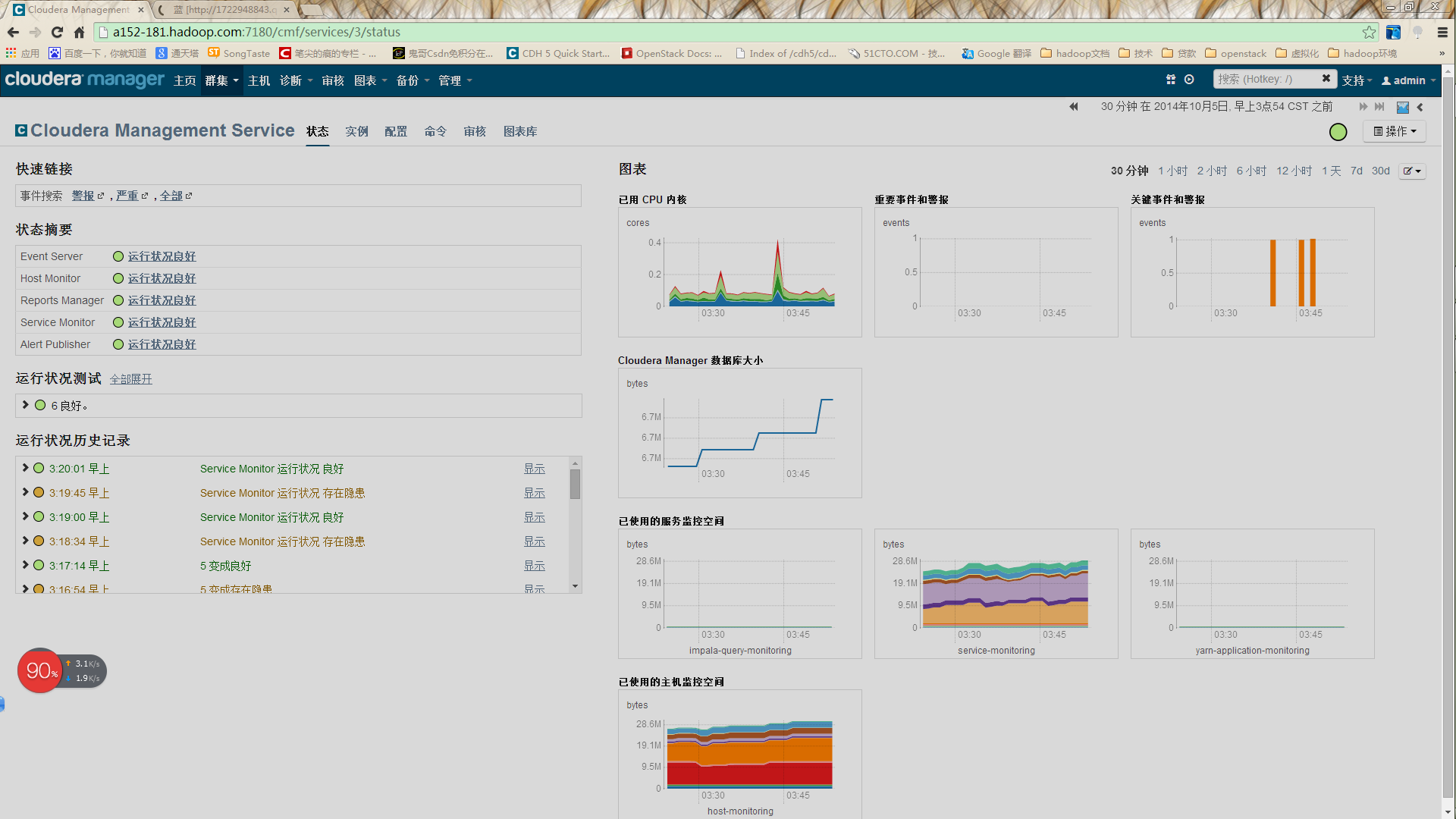
mkdir -p /data/snn/

chown -R hdfs.hdfs /data/nns/

chown -R hdfs.hdfs /data/snn/

chown -R hdfs.hdfs /data/dn/

yum install 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



service cloudera-scm-server stop

service cloudera-scm-agent stop

service zookeeper-server stop

chkconfig zookeeper-server on

chkconfig hadoop-hdfs-datanode off

chkconfig hadoop-hdfs-journalnode off

chkconfig hadoop-hdfs-namenode off

chkconfig hadoop-hdfs-nfs3 off

chkconfig hadoop-hdfs-secondarynamenode off

chkconfig hadoop-hdfs-zkfc off

chkconfig hadoop-httpfs off

service hadoop-hdfs-datanode stop

service hadoop-hdfs-journalnode stop

service hadoop-hdfs-namenode stop

service hadoop-hdfs-nfs3 stop

service hadoop-hdfs-secondarynamenode stop

service hadoop-hdfs-zkfc stop

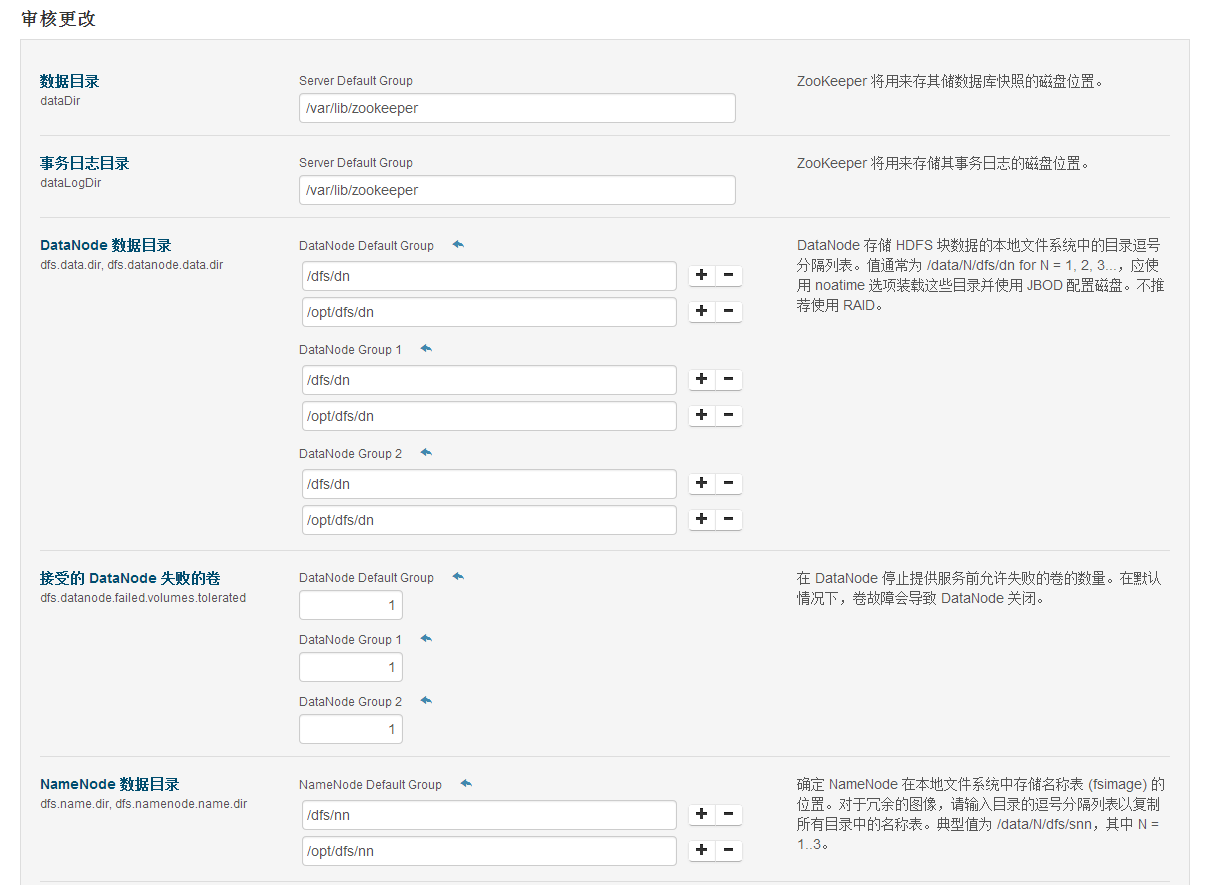
service hadoop-httpfs stop

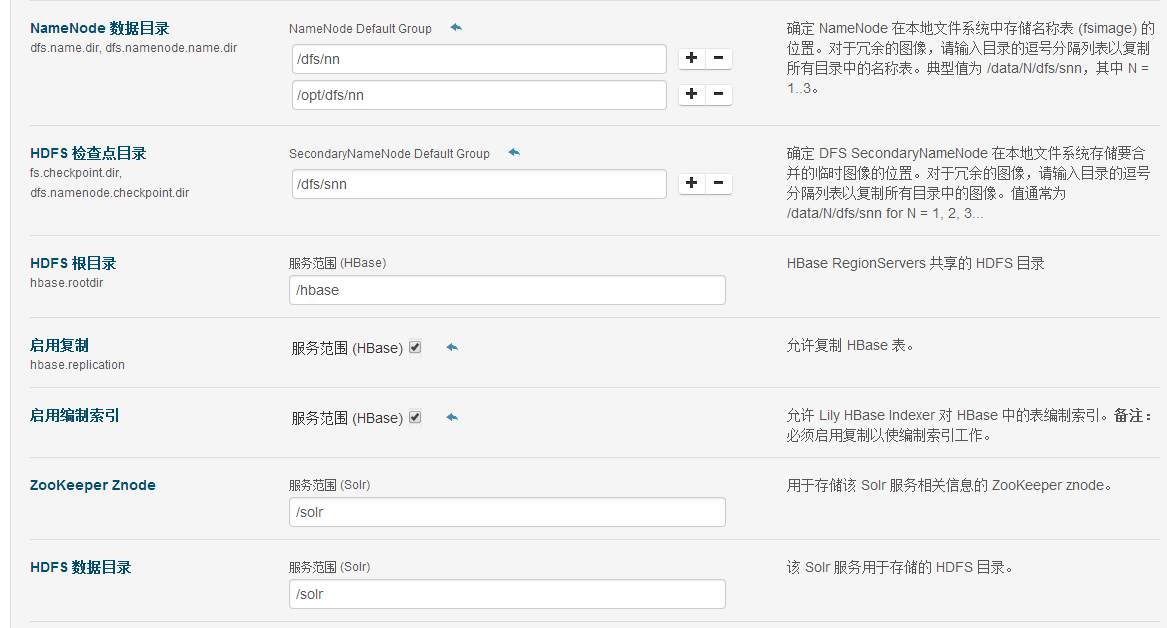
oozie配置：

yum install unzip

unzip ext-2.2.zip

mv ext-2.2 /var/lib/oozie/





<http://archive-primary.cloudera.com/cm5/cm/5/>

<http://archive-primary.cloudera.com/cdh5/parcels/5.3.0/>

安装教程：

<http://www.wangyongkui.com/hadoop-cdh5/>

# Cloudera Manager 5 和 CDH5 本地（离线）安装指南（Hadoop集群搭建）

[6条回复](http://www.wangyongkui.com/hadoop-cdh5/#comments)

转载请注明出处：http://www.wangyongkui.com/hadoop-cdh5。

# 系统环境

1. 4台联想R510服务器，其中master节点4核心、8GB，其它节点双核心、4GB。
2. 网卡：100M。
3. 共有硬盘6TB。
4. 网络环境内网。
5. Centos6.5 x64（安装系统时尽量把开发包安装齐全，另master节点需要Mysql可以在安装系统时勾选）。

# 准备工作

### 卸载系统自带OPEN-JDK（所有节点）

安装好的Centos系统有时会自动安装OpenJdk，用命令java -version查看：

|  |  |
| --- | --- |
| 1 | java version "1.6.0" |
| 2 | OpenJDK Runtime Environment (build 1.6.0-b09) | |

|  |  |
| --- | --- |
| 3 | OpenJDK 64-Bit Server VM (build 1.6.0-b09, mixed mode) |

如有上述显示，说明系统里已经有OpenJdk，执行以下命令查看系统中有哪些OpenJdk相关包：

|  |  |
| --- | --- |
| 1 | rpm -qa | grep java |

其中有如下包必须卸载，根据系统版本不同，各个包版本号会有所差异： java-1.5.0-gcj-1.5.0.0-29.1.el6.x86\_64 java-1.6.0-openjdk-1.6.0.0-1.66.1.13.0.el6.x86\_64 java-1.6.0-openjdk-devel-1.6.0.0-1.66.1.13.0.el6.x86\_64 执行以下命令，卸载：

|  |  |
| --- | --- |
| 1 | rpm -e --nodeps java-1.5.0-gcj-1.5.0.0-29.1.el6.x86\_64 |
| 2 | rpm -e --nodeps java-1.6.0-openjdk-1.6.0.0-1.66.1.13.0.el6.x86\_64 | |

|  |  |
| --- | --- |
| 3 | rpm -e --nodeps java-1.6.0-openjdk-devel-1.6.0.0-1.66.1.13.0.el6.x86\_64 |

### 安装JDK（所有节点）

从官方网站上下载rpm包，本次使用版本1.7.0\_55-b13（CDH5可能支持1.7之前的版本，具体情况未经测试），执行命令：

|  |  |
| --- | --- |
| 1 | rpm -ivh jdk-7u55-linux-x64.rpm |

由于是rpm包并不需要我们来配置环境变量，我们只需要配置一个全局的JAVA\_HOME变量即可，执行命令：

|  |  |
| --- | --- |
| 1 | echo "JAVA\_HOME=/usr/java/latest/" >> /etc/environment |

执行命令，查看Jdk是否安装正确

|  |  |
| --- | --- |
| 1 | java -version |
| 2 | javac -version | |

### 修改主机名

修改/etc/sysconfig/network文件：

|  |  |
| --- | --- |
| 1 | NETWORKING=yes |
| 2 | HOSTNAME=master.hadoop | |

|  |  |
| --- | --- |
| 3 | NETWORKING\_IPV6=no |
| 4 | GATEWAY=192.168.64.1 | |

其中HOSTNAME与主机名一致。主机名如果跟系统安装时不一致请执行hostname命令让其立即生效，否则会影响各节点互相访问。修改/etc/hosts文件，添加：

|  |  |  |
| --- | --- | --- |
| 1 | 192.168.1.101 master.hadoop | |
| 2 | 192.168.1.102 datanode1 |

|  |  |
| --- | --- |
| 3 | 192.168.1.103 datanode2 |
| 4 | 192.168.1.104 datanode3 |

执行命令：

|  |  |
| --- | --- |
| 1 | service network restart |

### 打通SSH（所有节点）

这里master节点和其他datanode节点会略有区别。 首先说有节点均执行如下命令，遇到提示一路回车即可：

|  |  |
| --- | --- |
| 1 | ssh-keygen -t rsa |

然后在主节点执行以下命令：

|  |  |
| --- | --- |
| 1 | cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys |

scp文件到所有datenode节点：

|  |  |
| --- | --- |
| 1 | scp ~/.ssh/authorized\_keys root@datanode1:~/.ssh/ |

然后输入密码即可，以后再到其他机器就不需要密码额。

### 关闭防火墙（所有节点）

防火墙会引起hadoop相关组件通讯的各种异常。 防火墙：

|  |  |
| --- | --- |
| 1 | service iptables stop (临时关闭) |
| 2 | chkconfig iptables off (重启后生效) | |

SELINUX：

|  |  |
| --- | --- |
| 1 | setenforce 0 (临时生效) |

修改/etc/selinux/config 下的 SELINUX=disabled （重启后生效）。

### 安装NTP服务（所有节点）

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：

1. 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。
2. 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：ntp与ntpdate。按顺序安装即可，完成后，配置开机启动：

|  |  |
| --- | --- |
| 1 | chkconfig ntpd on |

检查是否设置成功：

|  |  |
| --- | --- |
| 1 | chkconfig --list ntpd |

其中2-5为on状态就代表成功。

##### 配置内网NTP服务器（master节点）

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。

|  |  |
| --- | --- |
| 1 | ntdate -u 65.55.56.206 |

ntp服务只有一个配置文件，配置好了就OK。 这里只给出有用的配置，不需要的配置都用#注掉，这里就不在给出：

|  |  |  |
| --- | --- | --- |
| 1 | driftfile /var/lib/ntp/drift | |
| 2 | restrict 127.0.0.1 |

|  |  |
| --- | --- |
| 3 | restrict -6 ::1 |
| 4 | restrict default nomodify notrap | |

|  |  |
| --- | --- |
| 5 | server 65.55.56.206 prefer |
| 6 | includefile /etc/ntp/crypto/pw | |

|  |  |
| --- | --- |
| 7 | keys /etc/ntp/keys |

配置文件完成，保存退出，启动服务，执行如下命令：

|  |  |
| --- | --- |
| 1 | service ntpd start |
| 2 | chkconfig ntpd on (设置开机启动) | |

检查是否成功，用ntpstart命令查看同步状态，出现以下状态代表启动成功：

|  |  |  |
| --- | --- | --- |
| 1 | synchronised to NTP server () at stratum 2 | |
| 2 | time correct to within 74 ms |

|  |  |
| --- | --- |
| 3 | polling server every 128 s |

如果出现异常请等待几分钟，一般等待5-10分钟才能同步。

##### 配置ntp客户端（所有datanode节点）

|  |  |  |
| --- | --- | --- |
| 1 | driftfile /var/lib/ntp/drift | |
| 2 | restrict 127.0.0.1 |

|  |  |
| --- | --- |
| 3 | restrict -6 ::1 |
| 4 | restrict default kod nomodify notrap nopeer noquery | |

|  |  |  |
| --- | --- | --- |
| 5 | restrict -6 default kod nomodify notrap nopeer noquery | |
| 6 | server 192.168.1.101 |

|  |  |  |
| --- | --- | --- |
| 7 | includefile /etc/ntp/crypto/pw | |
| 8 | keys /etc/ntp/keys |

ok保存退出，请求服务器前，请先使用ntpdate手动同步一下时间：

|  |  |
| --- | --- |
| 1 | ntdate -u 192.168.1.101 (内网ntp服务器) |

这里可能出现同步失败的情况，请不要着急，一般是本地的ntp服务器还没有正常启动，一般需要等待5-10分钟才可以正常同步。启动服务：

|  |  |  |
| --- | --- | --- |
| 1 | service ntpd start | |
| 2 | chkconfig ntpd on |

因为是连接内网，这次启动等待的时间会比master节点快一些，但是也需要耐心等待一会儿。

### Mysql配置（master节点）

只需要配置master节点的Mysql即可。（这里不再介绍如何安装Mysql） 将mysql设置为开机启动：

|  |  |
| --- | --- |
| 1 | chkconfig mysqld on |

本次安装需要创建如下数据库（不包括Cloudera Manager的数据库，Cloudera Manager数据库有相关脚本创建，后面会有说明）

|  |  |
| --- | --- |
| 1 | --hive数据库 |
| 2 | create database hive DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci | |

|  |  |
| --- | --- |
| 3 | --集群监控数据库 |
| 4 | create database amon DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci | |

|  |  |
| --- | --- |
| 5 | --hue数据库 |
| 6 | create database hue DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci | |

以上数据库，可能会根据安装组件的不同略有差别。 给用户授权（这里密码设为hadoop）

|  |  |
| --- | --- |
| 1 | grant all on \*.\* to root@"%" Identified by "hadoop"; |

# 正式开工

### 安装Cloudera Manager 5（cm5）

下载地址<http://archive-primary.cloudera.com/cm5/cm/5/>，根据自己的系统选择相应的版本，本次安装选用的是cloudera-manager-el6-cm5.0.0\_x86\_64.tar.gz。下载完成后只上传到master节点即可。然后解压到/opt目录下，不能解压到其他地方，因为cdh5的源会默认在/opt/cloudera/parcel-repo寻找，怎么制作cdh5的本地源文件会在之后介绍。给所有节点添加cloudera-scm用户：

|  |  |
| --- | --- |
| 1 | useradd --system --home=/opt/cm-5.0.0/run/cloudera-scm-server --no-createhome --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm |

修改/opt/cm-5.0.0/etc/cloudera-scm-agent/config.ini 下面的serer\_host

|  |  |
| --- | --- |
| 1 | server\_host=master.hadoop |

为Cloudera Manager 5建立数据库：

|  |  |
| --- | --- |
| 1 | /opt/cm-5.0.0/share/cmf/schema/scm\_prepare\_database.sh mysql cm -hlocalhost -uroot -phadoop --scm-host localhost scm scm scm |

格式是:scm\_prepare\_database.sh 数据库类型  数据库 服务器 用户名 密码  –scm-host  Cloudera\_Manager\_Server所在的机器，后面那三个不知道代表什么，直接照抄官网的了。 因为我们用的是Mysql数据库，所以我们需要下载Mysql的JDBC驱动，本次从官网上下载最新稳定版：mysql-connector-java-5.1.30.tar.gz，解压之后找到mysql-connector-java-5.1.30-bin.jar放到/opt/cm-5.0.0/share/cmf/lib/目录下。 开启Cloudera Manager 5 Server端：

|  |  |
| --- | --- |
| 1 | /opt/cm-5.0.0/etc/init.d/cloudera-scm-server start |

注意server首次启动不要立即关闭或重启，因为首次启动会自动创建相关表以及数据，如果因为特殊原因中途退出，请先删除所有表以及数据之后再次启动，否则将会出现启动不成功的情况。 开启Cloudera Manager 5 Agents端。 先scp /opt/cm-5.0.0到所有datanode节点上，然后在每台机器上开启Agents端：

|  |  |
| --- | --- |
| 1 | scp -r  /opt/cm-5.0.0 root@datanode1:/opt/cm-5.0.0 |

等待拷贝成功，在所有datanode节点上启动：（注意必须以管理员权限启动）

|  |  |
| --- | --- |
| 1 | sudo /opt/cm-5.0.0/etc/init.d/cloudera-scm-agent start |

浏览器启动Cloudera Manager 5 控制台（默认端口号是7180），启动成功就会看到登陆页面。

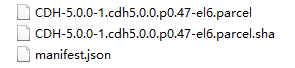
### 安装CDH5

先下载到本地<http://archive-primary.cloudera.com/cdh5/parcels/5.0.0/>，这里需要下载两样东西，首先是与自己系统版本相对应的parcel包，然后是manifest.json文件。下载完成后将这两个文件放到master节点的/opt/cloudera/parcel-repo下（目录在安装Cloudera Manager 5时已经生成），注意目录一个字都不能错，接下来打开manifest.json文件，里面是json格式的配置，我们需要的就是与我们系统版本相对应的hash码，因为我们用的是Centos6.5，所以找到如下位置：  
[无标题](http://www.wangyongkui.com/wp-content/uploads/2014/07/%E6%97%A0%E6%A0%87%E9%A2%98.png)

在这个大括号的最下面找到“hash”所对应的值。

[无标题1](http://www.wangyongkui.com/wp-content/uploads/2014/07/%E6%97%A0%E6%A0%87%E9%A2%981.png)

将“hash”的值复制下来，然后创建一个文件，文件名与你的parel包名一致，并加上.sha后缀：

[](http://www.wangyongkui.com/wp-content/uploads/2014/07/%E6%97%A0%E6%A0%87%E9%A2%982.png)

这样你的目录下将有这3个文件，将“hash”的值复制到新建的sha文件中，并保存，好了，我们的本地源制作完成了。这样基本大功告成了，再之后的操作就是控制台按照步骤安装即可。

打开http://192.168.1.101:7180，登陆控制台，默认账户和密码都是admin，安装时选择免费版，之后由于cm5对中文的支持很强大，按照提示安装即可，如果系统配置有什么问题在安装过程中会有提示，根据提示给系统安装组件就可以了。

如果在安装时选择了安装Hive，可能会遇到安装失败的问题，查看一下日志发现时安装Hive时需要安装JDBC驱动，所以同样我们将Mysql的驱动包拷贝到/opt/cloudera/parcels/CDH-5.0.0-1.cdh5.0.0.p0.47/lib/hive/lib/目录下，之后再继续安装就不会遇到问题了。

参考文献：<http://www.cnblogs.com/thinkCoding/p/3567408.html>                <http://www.byywee.com/page/M0/S877/877056.html>

GRANT ALL PRIVILEGES ON \*.\* TO 'amon'@'%' IDENTIFIED BY 'amon' WITH GRANT OPTION;

GRANT ALL PRIVILEGES ON \*.\* TO 'hive'@'%' IDENTIFIED BY 'hive' WITH GRANT OPTION;

rm -rf /dfs/dn /opt/dfs/dn /dfs/nn /opt/dfs/nn /dfs/snn /yarn/nm /opt/yarn/nm /yarn/nm \

/var/lib/sqoop2 /impala/impalad /opt/impala/impalad /var/lib/oozie/data /var/lib/cloudera-host-monitor \

/var/lib/cloudera-service-monitor