Cloudera Manager 离线部署 CDH 文档

目录：

一、说明

操作系统：CentOS 6
JDK 版本：1.7.0_80

所需安装包及版本说明：
   CDH-5.4.0-1.cdh5.4.0.p0.27-el6.parcel
   CDH-5.4.0-1.cdh5.4.0.p0.27-el6.parcel.sha
   manifest.json
   cloudera-manager-el6-cm5.4.3_x86_64.tar.gz

Cloudera Manager 下载目录
http://www.cloudera.com/downloads/manager/5-4-3.html

CDH 下载目录
http://archive.cloudera.com/cdh5/parcels/5.4.0/
CHD5 相关的 Parcel 包放到主节点的/opt/cloudera/parcel-repo/目录中
CDH-5.1.3-1.cdh5.1.3.p0.12-el6.parcel.sha1 重命名为 CDH-5.1.3-1.cdh5.1.3.p0.12-el6.parcel.sha，这点必须注意，否则，系统会重新下载 CDH-5.1.3-1.cdh5.1.3.p0.12-el6.parcel 文件

本文采用**离线安装方式**，在线安装方式请参照官方文档。

二、系统环境搭建

1、网络配置(所有节点)

vi /etc/sysconfig/network 修改 hostname：

```
[root@slave6 cdh]# vi /etc/sysconfig/network

NETWORKING=yes
#HOSTNAME=test-5916-server
HOSTNAME=slave6.hadoop
```

通过 service network restart 重启网络服务生效

vi /etc/hosts ,修改 ip 与主机名的对应关系

```
[root@slave6 cdh]# vi /etc/hosts

127.0.0.1    localhost localhost.localdomain localhost4 localhost4.localdomain4
::1          localhost localhost.localdomain localhost6 localhost6.localdomain6
60.206.137.166 master.hadoop
60.206.137.193 slave1.hadoop
60.206.137.165 slave2.hadoop
60.206.137.214 slave3.hadoop
60.206.137.215 slave4.hadoop
60.206.137.216 slave5.hadoop
60.206.137.217 slave6.hadoop
```

2、SSH 免密码登录

主节点执行：
ssh-keygen -t dsa -P '' -f ~/.ssh/id_dsa
生成无密码密钥对

拷贝公钥到其他节点，执行
cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
测试：主节点 ssh 其他节点……

3、关闭防火墙

临时关闭：
service iptables stop
重启后生效：
chkconfig iptables off

4、关闭 SELINUX

临时关闭：
setenforce 0
修改配置文件/etc/selinux/config（重启生效）：
将 SELINUX=enforcing 改为 SELINUX=disabled

查看 SELINUX 状态：
1、/usr/sbin/sestatus –v
    SELinux status: enabled（enabled：开启；disabled：关闭）
2、使用命令：getenforce

5、安装 JDK

摘自官网：
The Oracle JDK installer is available both as an RPM-based installer for RPM-based systems, and as a binary installer for other systems.

CDH 5.4.x is supported with the versions shown in the following table:

Minimum Supported Version	Recommended Version	Exceptions
1.7.0_55	1.7.0_67 or JDK1.7_75	None
1.8.0_40	1.8.0_40 or higher	None

本文采用 RPM 包安装……执行：
rpm -ivh jdk-7u80-linux-x64.rpm

配置环境变量，修改/etc/profile：
export JAVA_HOME=/usr/java/jdk1.7.0_80
export PATH=$JAVA_HOME/bin:$PATH
export CLASSPATH=.:$JAVA_HOMdE/lib/dt.jar:$JAVA_HOME/lib/tools.jar

生效：
source /etc/profile

查看版本：
[root@slave6 cdh]# java -version
java version "1.7.0_80"
Java(TM) SE Runtime Environment (build 1.7.0_80-b15)
Java HotSpot(TM) 64-Bit Server VM (build 24.80-b11, mixed mode)

6、设置 NTP

所有节点安装 NTP：
yum install ntp

配置开机启动：
chkconfig ntpd on

检查是否设置成功：
chkconfig --list ntpd （2-5 为 on 状态则成功）

设置同步：
ntpdate -u ntp.sjtu.edu.cn（时钟服务器根据实际环境设置、本文采用 210.72.145.44-国家授时中心服务器 IP 地址）

7、安装配置 MySql

MySql 版本选择、摘自官网：

Supported Databases：

Component	MySQL	SQLite	PostgreSQL	Oracle	Derby - see Note 4
Oozie	5.5, 5.6	–	8.4, 9.2, 9.3 See Note 2	11gR2	Default
Flume	–	–	–	–	Default (for the JDBC Channel only)
Hue	5.1, 5.5, 5.6 See Note 6	Default	8.4, 9.2, 9.3 See Note 2	11gR2	–
Hive/Impala	5.5, 5.6 See Note 1	–	8.4, 9.2, 9.3 See Note 2	11gR2	Default
Sentry	5.5, 5.6 See Note 1	–	8.4, 9.2, 9.3 See Note 2	11gR2	–
Sqoop 1	See Note 3	–	See Note 3	See Note 3	–

Component	MySQL	SQLite	PostgreSQL	Oracle	Derby - see Note 4
Sqoop 2	See Note 4	–	See Note 4	See Note 4	Default

■ **Note**:

1. MySQL 5.5 is supported on CDH 5.1. MySQL 5.6 is supported on CDH 5.1 and later. The InnoDB storage engine must be enabled in the MySQL server.

2. PostgreSQL 9.2 is supported on CDH 5.1 and later. PostgreSQL 9.3 is supported on CDH 5.2 and later.

3. For the purposes of transferring data only, Sqoop 1 supports MySQL 5.0 and above, PostgreSQL 8.4 and above, Oracle 10.2 and above, Teradata 13.10 and above, and Netezza TwinFin 5.0 and above. The Sqoop metastore works only with HSQLDB (1.8.0 and higher 1.x versions; the metastore does not work with any HSQLDB 2.x versions).

4. Sqoop 2 can transfer data to and from MySQL 5.0 and above, PostgreSQL 8.4 and above, Oracle 10.2 and above, and Microsoft SQL Server 2012 and above. The Sqoop 2 repository database is supported only on Derby and PostgreSQL.

5. Derby is supported as shown in the table, but not always recommended. See the pages for individual components in the Cloudera Installation and Upgrade guide for recommendations.

6. CDH 5 Hue requires the default MySQL version of the operating system on which it is being installed (which is usually MySQL 5.1, 5.5 or 5.6).

安装过程略……本文采用 MySql 5.5

所需数据库说明，摘自官网：
The **Cloudera Manager Server**, **Oozie Server**, **Sqoop Server**, **Activity Monitor**, **Reports Manager**, **Hive Metastore Server**, **Sentry Server**, **Cloudera Navigator Audit Server**, and **Cloudera Navigator Metadata Server** all require databases. The type of data contained in the databases and their estimated sizes are as follows:

• Cloudera Manager - Contains all the information about services you have configured and their role assignments, all configuration history, commands, users, and running processes. This relatively small database (<100 MB) is the most important to back up.

■ **Important**: When processes restart, the configuration for each of the services is redeployed using information that is saved in the Cloudera Manager database. If this information is not available, your cluster will not start or function correctly. You must therefore schedule and maintain regular backups of the Cloudera Manager database in order to recover the cluster in the event of the loss of this database.

• Oozie Server - Contains Oozie workflow, coordinator, and bundle data. Can grow very large.
• Sqoop Server - Contains entities such as the connector, driver, links and jobs. Relatively small.
• Activity Monitor - Contains information about past activities. In large clusters, this database can grow large. Configuring an Activity Monitor database is only necessary if a MapReduce service is deployed.
• Reports Manager - Tracks disk utilization and processing activities over time. Medium-sized.
• Hive Metastore Server - Contains Hive metadata. Relatively small.
• Sentry Server - Contains authorization metadata. Relatively small.

- Cloudera Navigator Audit Server - Contains auditing information. In large clusters, this database can grow large.
- Cloudera Navigator Metadata Server - Contains authorization, policies, and audit report metadata. Relatively small.

建库操作及脚本参照：步骤三、步骤六

8、下载依赖包

- chkconfig
- python（2.6 required for CDH 5）
- bind-utils
- psmisc
- libxslt
- zlib
- sqlite
- cyrus-sasl-plain
- cyrus-sasl-gssapi
- fuse
- portmap
- fuse-libs
- redhat-lsb

三、Cloudera Manager Server&Agent 安装

1、安装 Cloudera Manager Server&Agent

拷贝 cloudera-manager-el6-cm5.4.3_x86_64.tar.gz 到所有 Server、Agent 节点
创建 cm 目录：
mkdir /opt/cloudera-manager
解压 cm 压缩包：
tar xvzf cloudera-manager*.tar.gz -C /opt/cloudera-manager

2、创建用户 cloudera-scm（所有节点）

cloudera-scm 用户说明，摘自官网：

Cloudera Manager Server and managed services are configured to use the user account cloudera-scm by default, creating a user with this name is the simplest approach. This created user, is used automatically after installation is complete.

执行：

useradd --system --home=/opt/cloudera-manager/cm-5.0/run/cloudera-scm-server --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm

3、配置 CM Agent

修改文件/opt/cloudera-manager/cm-5.4.3/etc/cloudera-scm-agent/config.ini 中 server_host 以及 server_port

4、配置 CM Server 的数据库

将驱动包拷贝到目录下（注意拷贝过去的驱动包名字一定要和下边的一样，否则会报错）：
cp mysql-connector-java-5.1.31/mysql-connector-java-5.1.31-bin.jar /usr/share/java/mysql-connector-java.jar

执行：
mysql> grant all on *.* to 'temp'@'%' identified by 'temp' with grant option;
cd /opt/cloudera-manager/cm-5.4.3/share/cmf/schema
./scm_prepare_database.sh mysql -h myhost1.sf.cloudera.com -utemp -ptemp --scm-host myhost2.sf.cloudera.com scm scm scm
例如：

./scm_prepare_database.sh mysql -h node1 -utemp -ptemp --scm-host node1 scm scm scm
（对应于：数据库类型、数据库服务器、用户名、密码、CMServer 所在节点.......）

```
[root@slave6 schema]# ls
mysql oracle postgresql scm_database_functions.sh scm_prepare_database.sh
[root@slave6 schema]# ./scm_prepare_database.sh mysql -h 60.206.137.193 -utemp -ptemp --scm-host 60.206.137.217 scm scm scm
JAVA_HOME=/home/hadoop/opt/jdk1.7.0_79
Verifying that we can write to /opt/cloudera-manager/cm-5.3.4/etc/cloudera-scm-server
Creating SCM configuration file in /opt/cloudera-manager/cm-5.3.4/etc/cloudera-scm-server
Executing: /home/hadoop/opt/jdk1.7.0_79/bin/java -cp /usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-java.jar:/opt/cloudera-manager/cm-5.3.4/share/cmf/schema/../lib
/* com.cloudera.enterprise.dbutil.DbCommandExecutor /opt/cloudera-manager/cm-5.3.4/etc/cloudera-scm-server/db.properties com.cloudera.cmf.db.
[                    main] DbCommandExecutor              INFO  Successfully connected to database.
All done, your SCM database is configured correctly!
```

mysql> drop user 'temp'@'%';
若上步失败或过程中操作中断，删除所有库、重头来过/(ㄒoㄒ)/~~

若安装 Oozie 等组件可能需要手动创建对应组件所需的数据库，例如：
   create database ooziecm   DEFAULT CHARACTER SET utf8;
   grant all on ooziecm.* TO 'ooziecm'@'%' IDENTIFIED BY 'ooziecm';

其他的建库及删库脚本见步骤五

5、创建 Parcel 目录

Manager 节点创建目录/opt/cloudera/parcel-repo，执行：
mkdir -p /opt/cloudera/parcel-repo
chown cloudera-scm:cloudera-scm /opt/cloudera/parcel-repo
将下载好的文件（CDH-5.4.0-1.cdh5.4.0.p0.27-el6.parcel、 CDH-5.4.0-1.cdh5.4.0.p0.27-el6.parcel.sha、manifest.json）拷贝到该目录下。

Agent 节点创建目录/opt/cloudera/parcels，执行：
mkdir -p /opt/cloudera/parcels
chown cloudera-scm:cloudera-scm /opt/cloudera/parcels

6、启动 CM Manager&Agent 服务

执行：
Manager：/opt/cloudera-manager/cm-5.4.3/etc/init.d/cloudera-scm-server start
Agents：/opt/cloudera-manager/cm-5.4.3/etc/init.d/cloudera-scm-agent start

访问：http://ManagerHost:7180，若可以访问（用户名、密码：admin），则安装成功。
Manager 启动成功需要等待一段时间，过程中会在数据库中创建对应的表需要耗费一些时间。

四、CDH5 安装

CM Manager && Agent 成功启动后，登录前端页面进行 CDH 安装配置。

免费版本的 CM5 已经去除 50 个节点数量的限制。

各个 Agent 节点正常启动后，可以在当前管理的主机列表中看到对应的节点。
选择要安装的节点，点继续。
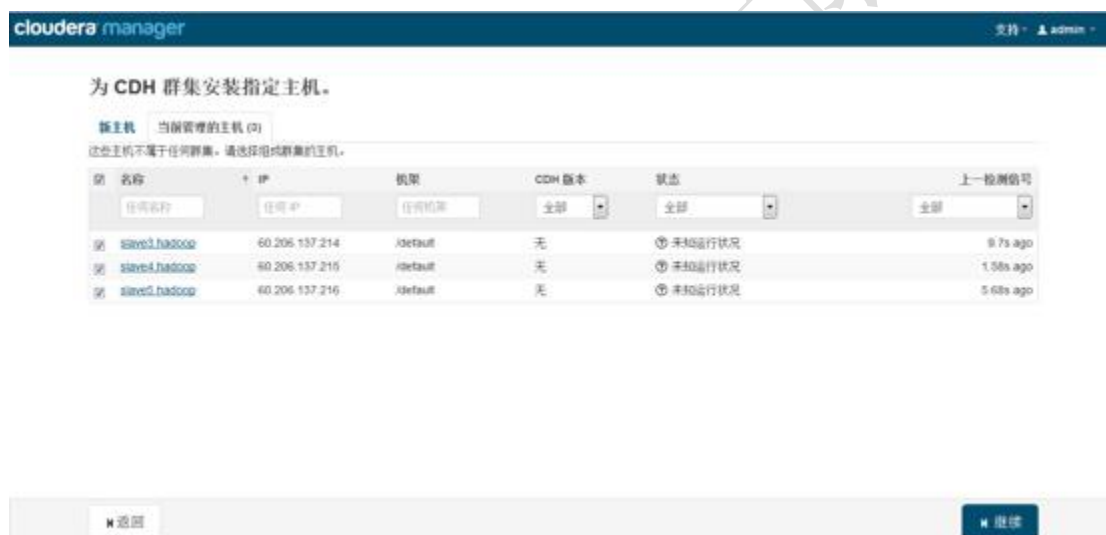
接下来，出现以下包名，说明本地 Parcel 包配置无误，直接点继续就可以了。

点击，继续，**如果配置本地 Parcel 包无误，那么下图中的已下载，应该是瞬间就完成了，**然后就是耐心等待分配过程就行了，大约 10 多分钟吧，取决于内网网速。
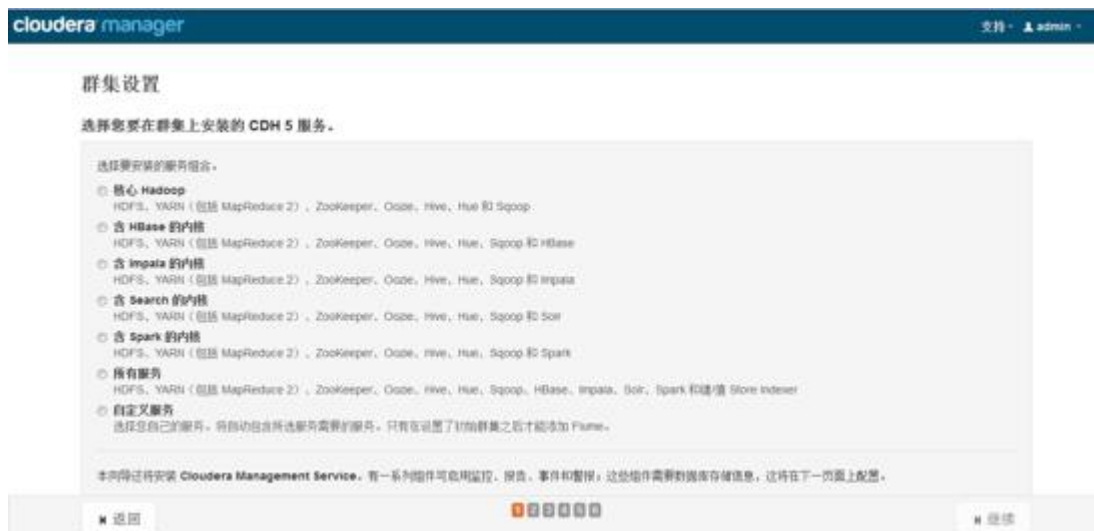
（若本地 Parcel 有问题，重新检查步骤三、5 是否配置正确）

接下来是服务器检查，可能会遇到以下问题：

Cloudera 建议将 /proc/sys/vm/swappiness 设置为 0。当前设置为 60。使用 sysctl 命令在运行时更改该设置并编辑 /etc/sysctl.conf 以在重启后保存该设置。您可以继续进行安装，但可能会遇到问题，Cloudera Manager 报告您的主机由于交换运行状况不佳。以下主机受到影响：

通过 echo 0 > /proc/sys/vm/swappiness 即可解决。

接下来是选择安装服务：

测试采用了 Hadoop 默认，实际按工作环境来定咯~~

服务配置，一般情况下保持默认就可以了（**Cloudera Manager** 会根据机器的配置自动进行配置，如果需要特殊调整，自行进行设置就可以了）：
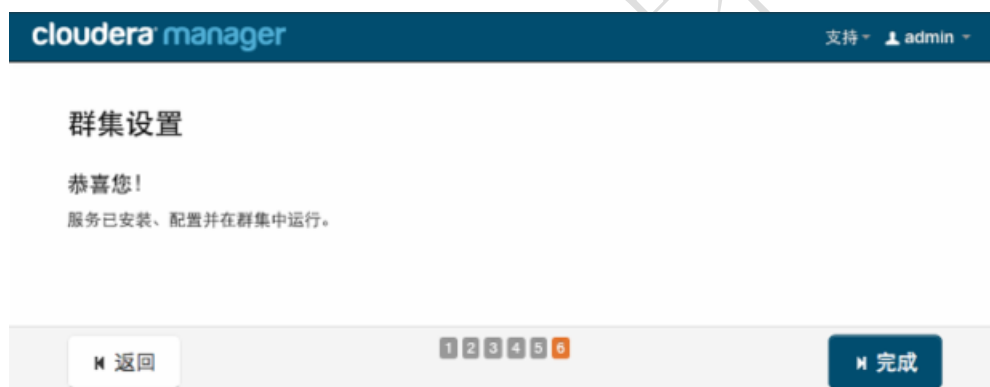

接下来是数据库的设置，检查通过后就可以进行下一步的操作了：

下面是集群设置的审查页面，我这里都是保持默认配置的：

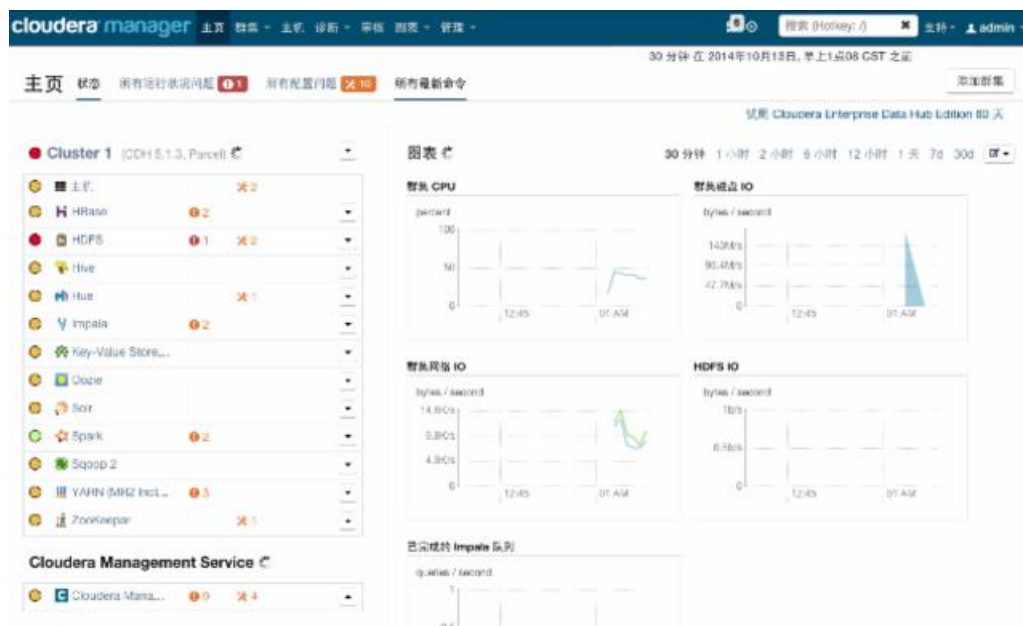终于到安装各个服务的地方了，注意，如果采用其他数据库安装 Hive 等组件的时候报错，检查之前配置 CM Server 数据库时，jar 包拷贝位置及名称是否修改

服务的安装过程大约半小时内就可以完成：

安装完成后，就可以进入集群界面看一下集群的当前状况了。

这里可能会出现无法发出查询：对 Service Monitor 的请求超时的错误提示，如果各个组件安装没有问题，一般是因为服务器比较卡导致的，过一会刷新一下页面就好了：

五、脚本

1、MySql 建库&&删库

Role	Database	User	Password
Activity Monitor（集群监控）	amon	amon	amon
Reports Manager	rman	rman	rman
Hive Metastore Server	metastore	hive	hive
Sentry Server	sentry	sentry	sentry
Cloudera Navigator Audit Server	nav	nav	nav
Cloudera Navigator Metadata Server	navms	navms	navms

```
create database amon        DEFAULT CHARACTER SET utf8;
create database rman        DEFAULT CHARACTER SET utf8;
create database metastore DEFAULT CHARACTER SET utf8;
create database sentry      DEFAULT CHARACTER SET utf8;
create database nav         DEFAULT CHARACTER SET utf8;
create database navms       DEFAULT CHARACTER SET utf8;

grant all on amon.* TO 'amon'@'%' IDENTIFIED BY 'amon';
grant all on rman.* TO 'rman'@'%' IDENTIFIED BY 'rman';
grant all on metastore.* TO 'metastore'@'%' IDENTIFIED BY 'metastore';
grant all on sentry.* TO 'sentry'@'%' IDENTIFIED BY 'sentry';
grant all on nav.* TO 'nav'@'%' IDENTIFIED BY 'nav';
grant all on navms.* TO 'navms'@'%' IDENTIFIED BY 'navms';

##hive
create database hive        DEFAULT CHARACTER SET utf8;
grant all on hive.* TO 'hive'@'%' IDENTIFIED BY 'hive';

##oozie
create database oozie       DEFAULT CHARACTER SET utf8;
grant all on oozie.* TO 'oozie'@'%' IDENTIFIED BY 'oozie';

drop database amon;
drop database rman;
drop database metastore;
drop database sentry;
drop database nav;
```

```
drop database navms;
drop database hive;
drop database oozie;
```

##amon
```
create database amon        DEFAULT CHARACTER SET utf8;
grant all on amon.* TO 'amon'@'%' IDENTIFIED BY 'amon';
```

##hive
```
create database hive        DEFAULT CHARACTER SET utf8;
grant all on hive.* TO 'hive'@'%' IDENTIFIED BY 'hive';
```

##oozie
```
create database oozie       DEFAULT CHARACTER SET utf8;
grant all on oozie.* TO 'oozie'@'%' IDENTIFIED BY 'oozie';
```