# **linux中Cloudera Manager 5和CDH5离线安装**

www.111cn.net 更新:2016-02-27 编辑:yahoo 来源:转载

Cloudera可以简化Hadoop的安装配置过程,自动在集群节点上安装hadoop相关组件,创建用户。所有机器上一定要安装好Perl包,下面我们来看一篇关于linux中Cloudera Manager 5和CDH5离线安装的例子吧。

一、相关软件准备及规划

1、相关软件及下载地址：

Cloudera Manager：http://archive-primary.cloudera.com/cm5/cm/5/  
CDH安装包地址：http://archive.cloudera.com/cdh5/parcels/latest/  
JAVA官方下载（需登陆）：http://www.oracle.com/technetwork/java/archive-139210.html  
JAVA各版本归档下载（无需登陆）：htt[ps](http://www.111cn.net/fw/photo.html" \t "http://www.111cn.net/sys/linux/_blank)://www.reucon.com/cdn/java/  
MySql的JDBC驱动jar包：http://dev.mysql.com/downloads/connector/j/

2、下载后所得文件如下：

cloudera-manager-el6-cm5.5.3\_x86\_64.tar.gz  
CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel  
CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel.sha1  
manifest.json  
mysql-connector-java-5.1.38.tar.gz（解压后里面有相当的jar包）  
java下载版本建议大于等于1.7以上版本  
CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel 该包比较大，后面离线安装时是将本包的内容解压后上传到各个节点上的，CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel.sha1该文件后续使用时需要改下名字，这个一定需要注意。

3、主机规划

我这里选取了5台主机安装，在测试时一般建议主机数量一定要大于2台以下。5台主机命名如下：

nn1.hadoop.com        192.168.0.10  
nn2.hadoop.com        192.168.0.11  
dn1.hadoop.com        192.168.0.12  
dn2.hadoop.com        192.168.0.13  
dn3.hadoop.com        192.168.0.14  
二、主机环境配置

主机配置分IP地址、主机名、java环境、句柄数优化几项，具体如下：

1、IP配置

# cat /etc/sysconfig/network-scripts/ifcfg-eth0  
DEVICE=eth0  
TYPE=Ethernet  
ONBOOT=yes  
NM\_CONTROLLED=yes  
BOOTPROTO=none  
IPADDR=192.168.0.10  
PREFIX=24  
GATEWAY=192.168.0.1  
DEFROUTE=yes  
IPV4\_FAILURE\_FATAL=yes  
IPV6INIT=no  
2、主机名配置

# cat /etc/sysconfig/network  
HOSTNAME=nn1.hadoop.com  
并在hosts文件里做下几台主机的IP指向。如果内部搭建的有DNS服务器，/etc/hosts指定可以省略。

3、JAVA环境变量配置

这部分可以参看另一篇博文 －－－ jDK的安装

4、句柄数优化

修改/etc/security/limits.conf文件，增加如下内容：

\*    hard    nofile   65535  
\*    soft    nofile   65535  
5、ssh密钥打通

这一个是非必须的，cloudera-manager对各个节点安装hdfs、flume、hive等应用时不依赖ssh进行传输的，其基于agent程序通信、传输、安装文件。如果想要ssh密钥打通可以参看我的另一篇博文－－－linux配置ssh公钥认证 。

6、iptables与selinux

service iptables stop (临时关闭)  
chkconfig iptables off (重启后生效)  
setenforce 0 (临时生效)  
修改/etc/selinux/config 下的模式为 SELINUX=disabled （重启后生效）。

7、ntp对时配置

选取其中一个节点服务器向外网对时，其他主机对该主机对时。由于ntp服务配置相当简单，这里也略过。

同样，按上述方法，对上例中的五台主机做下修改。

三、Cloudera Manager安装

先将下载好的Cloudera Manager包上传到其中一台服务器。并在/opt下解包－－－强烈建议到/opt下，因为默认路径是在/opt下，如果解压到其他路径下应用，配置文件会涉及到多项修改。

1、mysql 服务的配置

下载安装mysql 并启动

[root@nn1 opt]# yum -y install mysql-server  
[root@nn1 opt]# /etc/init.d/mysqld start  
[root@nn1 opt]# /usr/bin/mysqladmin -u root password 'hadoop'  
[root@nn1 opt]# chkconfig mysqld on  
创建hive和监控服务mysql 库

mysql> create database hive DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci;  
mysql> create database monitor DEFAULT CHARSET utf8 COLLATE utf8\_general\_ci;  
mysql> grant all on \*.\* TO 'root'@'%' IDENTIFIED BY 'hadoop';  
a、在标准安装中hive库是必须的，monitor监控[数据库](http://www.111cn.net/database/database.html" \t "http://www.111cn.net/sys/linux/_blank)为可选库－－－如果在安装时不选择安装hive服务的话，也可以不创建hive库。

b、给root用户授权所有主机可以登陆是因为hive和监控会涉及到多个服务，当服务不是安装在同一台主机上时，会出现需要通过其他主机连接上面创建的数据库的问题，这样就会造成失败。如果出于安全考虑，可以给mysql 用户授权时，只指定hadoop所在的网段。

2、导入数据库

mysql jar包放入lib库

将下面好的mysql jdbc 包里的jar文件取出放到/opt/cm-x.x.x/share/cmf/lib/中，其中x.x.x为cloudera manager的版本号。由于目前mysql是甲骨文公司下的产品，所以没有开源授权，所以其他产品在进行mysql支持时，都需要从mysql官站上自行下载放入lib包。而cloudera manager支持的postgresql和oracle没有这个问题（有点不明白为什么oracle不存在这个问题，NND的甲骨文）。

创建用户

创建cloudera-scm用户

useradd --system --home=/opt/cm-5.1.3/run/cloudera-scm-server/ --no-create-home --shell=/bin/false --comment "Cloudera SCM User" cloudera-scm  
上面这个用户需要在五台主机上都创建，不然在后面web界面管理安装时，在check项会错，提示不存在cloudera-scm用户。

导入数据

/opt/cm-5.5.3/share/cmf/schema/scm\_prepare\_database.sh mysql cm -hlocalhost -uroot -pxxxx --scm-host localhost scm scm scm  
3、agent配置

修改/opt/cm-5.5.3/etc/cloudera-scm-agent/config.ini中的server\_host为主节点的主机名－－－也可以使用IP，但建议使用主机名，这样涉及到IP变动时，直接在DNS或hosts里做下IP指向修改即可。这个配置文件里也会一些其他项的设置，有兴趣的可以自行查看下。

并通过scp或其他工具将/opt/cm-5.5.3目录复制到其他agent主机上的opt目录下。agent复制一定要在启动前复制，启动后会产生uuid，复制到其主机再启动，会报UUID相关的报错。

4、Parcels离线包设置

将CHD5相关的Parcel包放到主节点的/opt/cloudera/parcel-repo/目录中（parcel-repo不存时即手动创建）。将之前下载好的如下三个文件复制进该目录：

CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel  
CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel.sha1  
manifest.json  
需要将CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel.sha1文件重命名为CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel.sha，否则系统会重新下载CDH-5.3.9-1.cdh5.3.9.p0.8-el6.parcel包。这个包有1个多G，hadoop常用的生态用应都包含在这个包里。

5、启动manager和agent服务

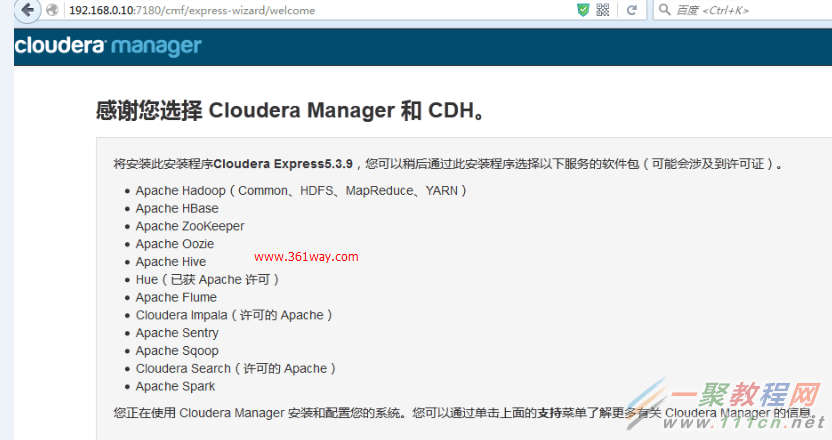
通过/opt/cm-5.5.3/etc/init.d/cloudera-scm-server start启动服务端；  
通过/opt/cm-5.5.3/etc/init.d/cloudera-scm-agent start启动Agent服务。  
同理，也可以stop和restart 。

四、通过cloudera manager安装CDH5

cloudera manager默认使用的端口有两个－－－7180是web界面使用的端口，7182是和agent通信使用的端口。可以通过浏览器打开：http://cloudera manager ip:7180 打开管理界面，默认用户名密码都是admin 。



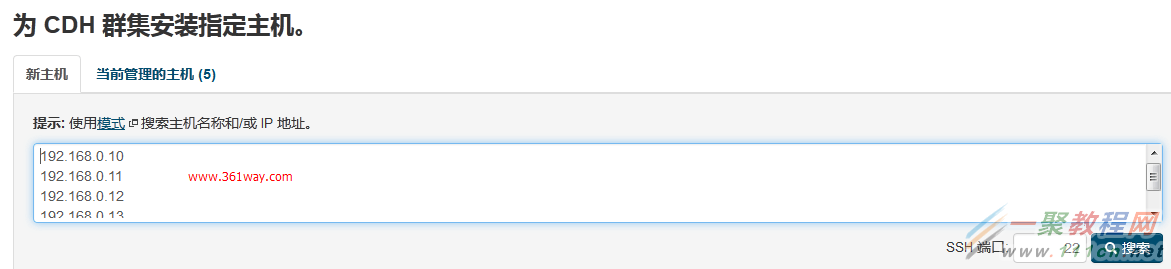
登陆完成后，会让我们选取版本，我们选择免费的cloudera express，在很早之前的版本中express版只允许50个节点，目前已经没有这个限制。选择完成后，会对express版本倚重个简单的内容介绍，如下：



由于agent启动前，我们通过配置文件指定了manger主机的IP，所以这里可以发现manager已经接管了5台主机



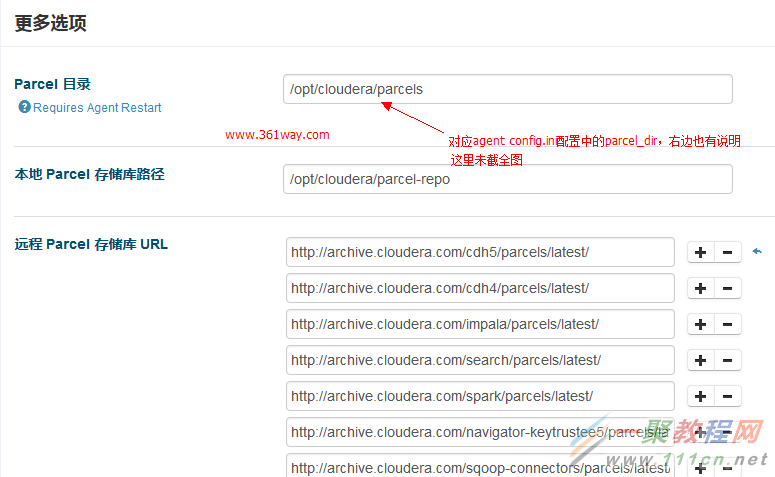
增加主机这项也可以通过新主机项中输入IP后自动搜索



这里的安装我们由于是离线安装，选择parcel 安装



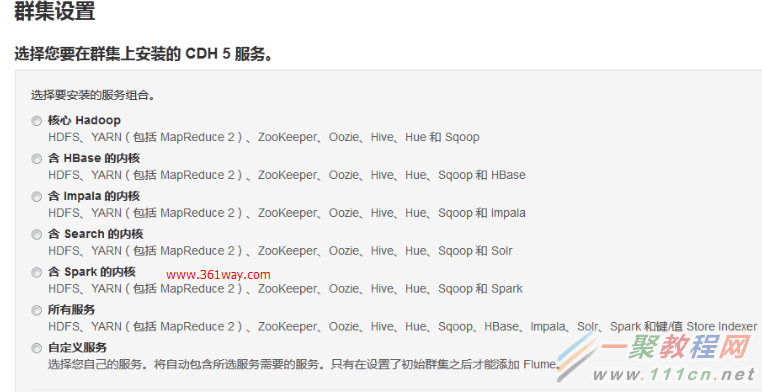
这里会有一些选择设置，之所以选择opt下就是由于默认是在opt下的。向下拉还会有更多的配置，由于版面原因，其他部分我示截取



在不增加cloudera-scm用户或未将swappiness设置为0时，会有警告信息，修改后，可以重新check一次



这里是选择需要安装的服务，我这里选择的核心hadoop，这里可以根据自己需要进行选择



下面这个图是本篇的主题部分，之所以开篇未选择其他安装方式，就是因为这个图表将hadoop主要生态圈的应用和功能做了一个详细的概述



默认会根据配置随机指定每台主机安装的服务，不通这下图这里可以在每个组件对应的功能模块上指定相应的主机



数据库安装部分，这里也可以使用postgresql 和 oracle，这里既然选择mysql，同样需要/opt/cloudera/parcels/CDH-5.3.9-1.cdh5.3.9.p0.8-el6/lib/hive/lib/下 ，由于这里很容易出错，所以上面建库后建议先在其他主机上通过mysql -h -u连接测试下



后面的安装进度就略过了，无非是一段时间的等待后进入安装完成界面。安装完成后通过查看上面的功能模块可以发现，尽管是express版本的cloudera manager，功能还是十分强大的。



在manager上我们可以新增节点，配置kerberos或ldap认证，一般的监控，数据的查询和展示等