

CDH 简介部署方式

1,自动部署方式

直接采用官方源来部署，主要是下载慢

2,parcels 部署方式

自己搭建本地 parcels Repository，适合内部大批量部署，也适合自动化，对系统侵入最小

3,搭建本地源部署 agent

自己搭建本地 yum repository

第一部分：环境配置

➤ 1,卸载系统自带的 jdk

```
yum remove `rpm -qa |grep java`
```

➤ 2,修改主机名为 FQDN

修改/etc/sysconfig/network 文件：

```
NETWORKING=yes
```

```
HOSTNAME=hadoop.one
```

修改 hosts：

```
[root@kali ~]# cat /etc/hosts
```

```
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
```

```
::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
```

```
192.168.3.47 hadoop.one hadoop
```

```
192.168.3.53 dn1.one dn1
```

如果有多台则这里加入多行，比如 datanote 地址

➤ 3,配置 SSH 免密码登录

namenote1 生产密钥：

```
ssh-keygen -t rsa
```

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

将自动应答文件传输到其他机器：

```
scp ~/.ssh/authorized_keys root@datanode1:~/.ssh/
```

➤ 4,关闭防火墙（所有节点）

iptables：

```
service iptables stop (临时关闭)
```

```
chkconfig iptables off (重启后生效)
```

selinux：

```
setenforce 0 (临时生效)
```

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

➤ 5,配置时间同步 ntp

首次同步时间：

```
[root@kali ~]# ntpdate -u 0.centos.pool.ntp.org
```

```
17 Apr 16:06:03 ntpdate[1961]: step time server 202.65.114.202 offset -28800.065175 sec
```

加入 crontab 中：

```
[root@kali ~]# crontab -l
```

```
* */1 * * * ntpdate -u 0.centos.pool.ntp.org
```

➤ 6,安装 jdk

下载安装：

```
rpm -ivh http://archive-primary.cloudera.com/cm5/redhat/7/x86\_64/cm/5.7.0/RPMS/x86\_64/oracle-j2sdk1.7-1.7.0+update67-1.x86\_64.rpm
```

➤ 7,关于系统版本

最好采用 centos6 来进行安装部署，兼容性最好，centos7 也可以，但注意 7.0 不行，7.1 和 7.2 可以，我这里采用 centos6.5

注意 python 版本要是 2.6 或者 2.7

除了免密码配置只需要在 namenode1 节点上配置外，其他的都需要在所有节点上配置

➤ 8,系统 open file 的修改

之所以配置这个是因为 mysql crash 过一次，无论怎么修改都无法启动，这是官方给出的一个建议配置，报错如下：

This could be because you hit a bug. It is also possible that this binary or one of the libraries it was linked against is corrupt, improperly built, or misconfigured. This error can also be caused by malfunctioning hardware. We will try our best to scrape up some info that will hopefully help diagnose the problem, but since we have already crashed, something is definitely wrong and this may fail.

首先执行 `ulimit -n 65535`

然后在 `/etc/security/limits.conf` 下面添加如下的一行

```
* - nofile 65536
```

➤ 9，安装依赖包

```
yum install libxslt readline readline-devel readline-static openssl openssl-devel openssl-static sqlite-devel bzip2-devel bzip2-libs ntp ntpdate -y
```

➤ 10 , 创建用户 cloudera-scm

```
useradd --system --home=/opt/cm-5.7.0/run/cloudera-scm-server/ --no-create-home  
--shell=/bin/false --comment "Cloudera SCM User" cloudera-scm
```

第二部分：自动部署方式

➤ 官方下载 Cloudera Manager 安装脚本

```
$ wget https://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin  
$ chmod u+x cloudera-manager-installer.bin  
$ sudo ./cloudera-manager-installer.bin
```

全部选择接受就可以了，这里会根据你的系统版本在/etc/yum.repo/增加一个官方的源来进行 yum 自动安装

```
[root@hadoop ~]# cat /etc/yum.repos.d/cloudera-manager.repo  
[cloudera-manager]  
name = Cloudera Manager, Version 5.7.0  
baseurl = https://archive.cloudera.com/cm5/redhat/6/x86\_64/cm/5.7.0/  
gpgkey = https://archive.cloudera.com/redhat/cdh/RPM-GPG-KEY-cloudera  
gpgcheck = 1
```

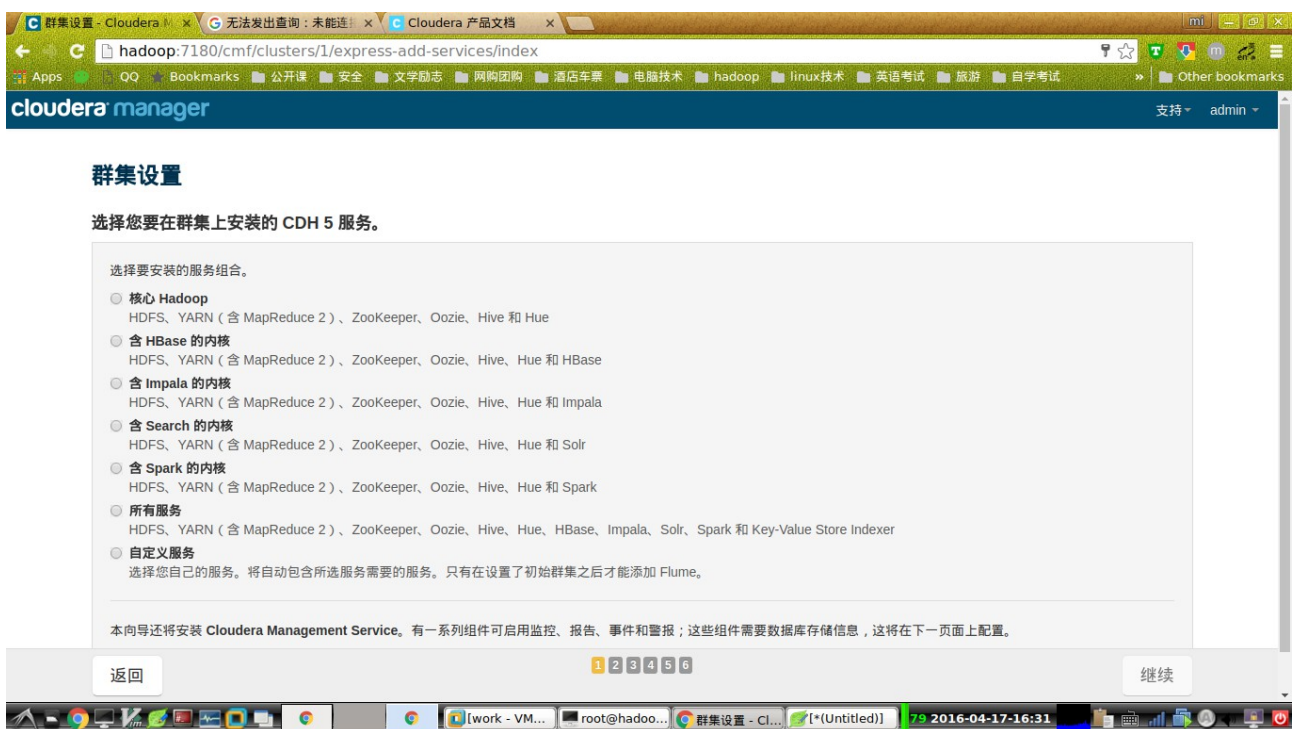
安装完后通过 ip:7180 端口来访问 Cloudera Manager 来进行 Software Installation and Configuration，通过 Cloudera Manager 提供的向导可以安装想安装的组件

首次安装还是一路默认吧，不过注意 FQDN 如果配置出错，会导致安装时的莫名其妙的错误，格式一定要是 xxx.xxx 类似域名的方式，不能只包含一段

这里贴几张安装过程中的图



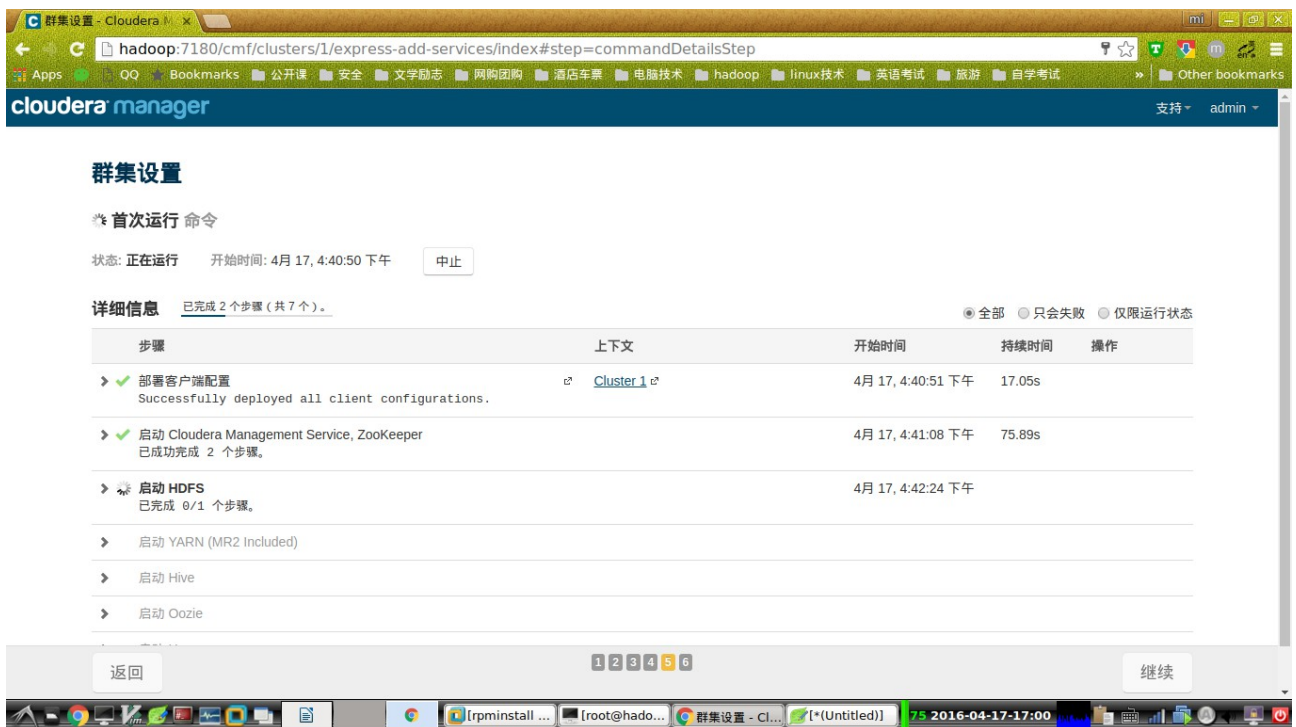
安装完 agent 后的检查结果，如果有错误，只需要按照提示去更新就好了



选择安装的组件



对每台机器分配角色



正在安装组件

第三部分：通过 parcels 包部署方式

环境初始化

参考第一部分环境配置（第一部分）

mysql 的部署

- 1，增加官方源

<http://dev.mysql.com/downloads/repo/yum/>

选择相应的版本下载然后 rpm -ivh 安装

- 2，yum 安装

```
yum install mysql mysql-server -y
```

- 3，启动 mysql

```
service mysqld start
```

我这里安装的是 5.7，启动过程会进行初始化，安装密码插件，并创建初始密码

查看初始密码 `cat /var/log/mysql.log |grep password`

关闭密码插件，否则无法创建简单密码的帐号，在/etc/my.cnf 里面加入

```
validate_password_policy=0
```

```
validate_password_length=0
```

```
validate_password_mixed_case_count=0
```

重启 mysql：service mysqld restart

需要更改初始密码才能创建数据库：mysqladmin -uroot -p'xx&78*(/op' password 'hadoop'

➤ 4，创建后面要用到的数据库

#hive

create database hive DEFAULT CHARSET utf8 COLLATE utf8_general_ci;

#activity monitor

create database amon DEFAULT CHARSET utf8 COLLATE utf8_general_ci;

➤ 5，授权 root 用户拥有所有数据库的访问权限

grant all privileges on *.* to 'root'@'%' identified by 'hadoop' with grant option;

flush privileges;

Cloudera Manage Server 部署

➤ 1，下载 CM (cloudera-manager) 包

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

里面选择对应系统的相应版本









我这里下载：http://archive.cloudera.com/cm5/cm/5/cloudera-manager-el6-cm5.7.0_x86_64.tar.gz

➤ 2，parcel-repo 目录创建，并下载 parcel 包放入其中

官网：<http://archive.cloudera.com/cdh5/parcels/5.7.0/>

下载如下三个文件，根据你的系统去选择

Index of /cdh5/parcels/5.7.0

Name	Last modified	Size	Description
 Parent Directory	-	-	-
 CDH-5.7.0-1.cdh5.7.0.p0.45-el5.parcel	2016-03-31 19:55	1.3G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-el5.parcel.sha1	2016-03-31 19:55	41	
 CDH-5.7.0-1.cdh5.7.0.p0.45-el6.parcel	2016-03-31 19:56	1.3G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-el6.parcel.sha1	2016-03-31 19:56	41	
 CDH-5.7.0-1.cdh5.7.0.p0.45-el7.parcel	2016-03-31 19:56	1.3G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-el7.parcel.sha1	2016-03-31 19:56	41	
 CDH-5.7.0-1.cdh5.7.0.p0.45-precise.parcel	2016-03-31 19:55	1.4G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-precise.parcel.sha1	2016-03-31 19:55	41	
 CDH-5.7.0-1.cdh5.7.0.p0.45-sles11.parcel	2016-03-31 19:54	1.4G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-sles11.parcel.sha1	2016-03-31 19:54	41	
 CDH-5.7.0-1.cdh5.7.0.p0.45-trusty.parcel	2016-03-31 19:54	1.4G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-trusty.parcel.sha1	2016-03-31 19:54	41	
 CDH-5.7.0-1.cdh5.7.0.p0.45-wheezy.parcel	2016-03-31 19:55	1.4G	
 CDH-5.7.0-1.cdh5.7.0.p0.45-wheezy.parcel.sha1	2016-03-31 19:55	41	
 manifest.json	2016-03-31 19:56	56K	

mkdir /opt/cloudera/parcel-repo

将 CDH 的 parcel 包放入此目录中，共三个文件，记得将 CDH-5.7.0-1.cdh5.7.0.p0.45-el6.parcel.sha1 改名为 CDH-5.7.0-1.cdh5.7.0.p0.45-el6.parcel.sha

安装 CDH 将会从这里下发安装

➤ 3，创建用户 cloudera-scm

```
useradd --system --home=/opt/cm-5.7.0/run/cloudera-scm-server --no-create-home  
--shell=/bin/false --comment "Cloudera SCM User" cloudera-scm
```

将相应目录权限改为 cloudera-scm

```
chown cloudera-scm:cloudera-scm /opt/cm-5.7.0
```

```
chown cloudera-scm:cloudera-scm /opt/cloudera
```

➤ 4，下载 jdbc 驱动

官网下载地址：<http://dev.mysql.com/downloads/connector/j/>

<http://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.38.tar.gz>

下载解压后将 mysql-connector-java-5.1.38-bin.jar 放到

mv mysql-connector-java-5.1.38-bin.jar /opt/cm-5.7.0/share/cmfd/lib/

记得更改一下权限：chown cloudera-scm:cloudera-scm /opt/cm-5.7.0

后面 agent 的配置文件 config.ini 中记得将 jdbc 位置改为上面这个位置

```
## Location of JDBC Drivers.
[JDBC]
cloudera_mysql_connector_jar=/opt/cm-5.7.0/share/cmfd/lib/mysql-connector-java-5.1.38-bin.jar
#cloudera_mysql_connector_jar=/usr/share/java/mysql-connector-java.jar
```

➤ 5 , 将 CM 解压到/opt 目录中并执行数据库初始化

tar -zxvf cloudera-manager-el6-cm5.7.0_x86_64.tar.gz -C /opt/

然后通过 scm_prepare_database.sh 脚本创建数据库并进行数据初始化

/usr/share/cmfd/schema/scm_prepare_database.sh mysql -h localhost -uroot -phadoop --scm-host localhost scm scm scm

参数解释：/usr/share/cmfd/schema/scm_prepare_database.sh 数据库类型 数据库 ip 数据库用户名 数据库密码 CM 地址 CM 数据库名称 CM 数据库帐号 CM 数据库密码

表 1. 必要参数

参数	说明
database-type	支持的数据库类型之一：
	MySQL -mysql
	Oracle -oracle
	PostgreSQL -postgresql
database-name	要创建或使用的 Cloudera Manager Server 数据库的名称。
username	要创建或使用的 Cloudera Manager Server 数据库的用户名。
password	要创建或使用的 Cloudera Manager Server 数据库的密码。如果不在命令行中指定密码，脚本会提示您输入它。

表 2. 选项

选项	说明
-h or --host	安装该数据库的主机的 IP 地址或主机名。默认情况下使用本地主机。
-P or --port	用于连接数据库的端口号。默认端口为 3306 (MySQL)、5432 (PostgreSQL) 和 1521 (Oracle)。此选项仅用于远程连接。
-u or --user	数据库应用程序的管理员用户名。对于 -u，选项和提供的值之

选项	说明
	间没有空格。如果提供此选项，脚本会为 Cloudera Manager Server 创建用户和数据库；否则，它会使用您以前创建的用户和数据库。
-p or --password	数据库应用程序的管理员密码。默认为无密码。对于 -p，选项和提供的值之间没有空格。
--scm-host	安装了 Cloudera Manager Server 的主机名。如果 Cloudera Manager Server 和数据库安装在同一台主机上，则予以省略。
--config-path	Cloudera Manager Server 配置文件的路径。默认为/etc/cloudera-scm-server。
--schema-path	Cloudera Manager 架构文件的路径。默认为/usr/share/cmf/schema（脚本的位置）。
-f	出现错误时，脚本将不会停止。
-? or --help	显示帮助。

➤ 6，启动 CM server 端

/opt/cm-5.7.0/etc/init.d/cloudera-scm-server start

查看 log 日志：tail -100f /opt/cm-5.7.0/log/cloudera-scm-server/cloudera-scm-server.log

看是否有报错

启动成功后会开启 7180,7182 端口，此时可以通过 ip:7180 端口去访问看是否出现登录界面

Cloudera Manage Agent 端初始化

➤ 1，解压 CM 压缩包并修改相应文件

mkdir /opt/cdh

tar -zxvf cloudera-manager-el6-cm5.7.0_x86_64.tar.gz -C /opt/cdh/

修改 agent 的 config.ini 文件

vim /opt/cloudera/cm-5.7.0/etc/cloudera-scm-agent/config.ini

主要如下几个地方更改

Hostname of the CM server.

server_host=hadoop

cloudera_mysql_connector_jar=/opt/cm-5.7.0/share/cmf/lib/mysql-connector-java-5.1.38-bin.jar

修改启动脚本

在 pidname 变量下面加入如下两行：

```
USER=cloudera-scm
```

```
GROUP=cloudera-scm
```

在 cm-5.7.0 目录中的 run 目录中创建一个目录

```
mkdir cm-5.7.0/run/cloudera-scm-agent/
```

此目录如果不创建，会在启动 agent 是提示无法创建 pid 文件，因为脚本的默认 pid 文件就在这里面，不知道为什么官方给的包里面没有这个目录，折腾人。。。

也有可能还要创建其他目录，如果启动过程有报错的话

- /var/log/cloudera-scm-headlamp
- /var/log/cloudera-scm-firehose
- /var/log/cloudera-scm-alertpublisher
- /var/log/cloudera-scm-eventserver
- /var/lib/cloudera-scm-headlamp
- /var/lib/cloudera-scm-firehose
- /var/lib/cloudera-scm-alertpublisher
- /var/lib/cloudera-scm-eventserver
- /var/lib/cloudera-scm-server

然后就可以将目录 cm-5.7.0 打包批量批量传输到其他机器上了

➤ 2，启动 agent

```
scp -r cm-5.7.0 dn1:/opt/
```

```
ssh dn1 /opt/cm-5.7.0/etc/init.d/cloudera-scm-agent start
```

当然启动之前还是需要进行环境初始化的，参考第一部分

看 log 是否有报错

```
tail -100f /opt/cm-5.7.0/log/cloudera-scm-agent/cloudera-scm-agent.log
```

如果没有报错，然后就可以访问服务端进行 CDH 安装了

整个安装过程踩了不少坑，像 FQDN 配错、kernel bug(内存不够导致，完整安装至少要 4G 内存)、mysql crash、supervisord 连接不上、parcel 包未加载、hive/hue 的 JDBC 驱动报错等等，总之多 google/baidu 搜索，总有人比你先踩坑的。。。。。

完整安装过程

第三部分：通过搭建本地 YUM 源仓库部署方式

参考：http://www.cloudera.com/content/www/zh-CN/documentation/enterprise/5-3-x/topics/cdh_ig_yumrepo_local_create.html

设置本地 yum 存储库以在群集中的机器上安装 CDH。您可能处于多种原因需要执行此操作，

例如：

- 群集中的计算机无网络访问。您仍通过创建本地 yum 来使用 yum 在这些机器上进行安装。
- 您可能希望保持稳定的本地存储库，以确保任何新安装（或在现有群集成员上重新安装）使用完全相同的位。
- 使用本地存储库可能是向群集成员分配软件的最有效方法。

➤ 1，下载 repo 文件

```
wget http://archive.cloudera.com/cdh5/redhat/6/x86_64/cdh/cloudera-cdh5.repo
```

并将其保存在/etc/yum.repos.d/目录中

对于 **OS** 版本

单击此链接

RHEL/CentOS/Oracle 5

[RHEL/CentOS/Oracle 5 链接](#)

RHEL/CentOS/Oracle 6（64 位）

[RHEL/CentOS/Oracle 6 链接](#)

- 2，在使用 RPM 的机器上安装诸如 apache/lighttpd 之类的 Web 服务器。默认配置应工作。必须允许 HTTP 访问以通过该服务器您与互联网连接之间的任何防火墙。

```
yum install httpd -y
```

- 3，在具有 Web 服务器的服务器上，安装 yum-utils 和 createrepoRPM 软件包（如果其尚未安装）。yum-utils 软件包包含 reposync 命令，需要用其创建逻辑 Yum 存储库。

```
yum install yum-utils createrepo
```

- 4，在执行先前步骤的同一计算机上，将 yum 存储库下载至临时位置。

```
reposync -r cloudera-cdh5
```

或者可以直接下载 <http://archive.cloudera.com/cm5/repo-as-tarball/5.7.0/cm5.7.0-centos6.tar.gz> 使用

- 5，将所有 RPM 置于 Web 服务器使用的目录中，例如/var/www/html/cdh/5/RPMS/noarch/（或 x86_64 或 i386，而不是 noarch）。需要的目录结果为 5/RPMS/noarch。请确保您可以使用与 <http://<yourwebserver>/cdh/5/RPMS/> 相似的 URL 通过 HTTP 远程访问目录中的文件。
- 6，在 Web 服务器上，从 RPM 目录的 5/子目录发出以下命令。

`createrepo .`

这将创建或更新 yum 命令将此目录识别为存储库所需的元数据。该命令将会创建一个名为 repodata 的新目录。必要时，请将整个存储库目录中文件和目录的权限调整为 Web 服务器用户可读。

- 7，编辑您在步骤 1 中下载的存储库文件并使用步骤 5 中的 URL 将以 baseurl=或 mirrorlist=为开头的行替换为
baseurl=http://<yourwebserver>/cdh/5/。将文件保存至/etc/yum.repos.d/。
- 8，断网后，发出以下命令以安装本地 yum 存储库中的 CDH。

`yum update`

`yum install hadoop`

参考网站：

http://www.cloudera.com/content/www/zh-CN/documentation/enterprise/5-3-x/topics/cm_ig_install_path_c.html

http://www.cloudera.com/content/www/zh-CN/documentation/enterprise/5-3-x/topics/cm_ig_install_path_c.html#concept_elj_3sr_5n_unique_1

<http://www.cnblogs.com/jasondan/p/4011153.html>

<http://www.cloudera.com/content/www/zh-CN/documentation/enterprise/5-3-x/topics/introduction.html>