

CDH6.0.0详细安装教程及所遇到的问题

关于CDH和Cloudera Manager

CDH (Cloudera's Distribution, including Apache Hadoop), 是Hadoop众多分支中的一种, 由Cloudera维护, 基于稳定版本的Apache Hadoop构建, 并集成了很多补丁, 可直接用于生产环境。Cloudera Manager则是为了便于在集群中进行Hadoop等大数据处理相关的服务安装和监控管理的组件, 对集群中主机、Hadoop、Hive、Spark等服务的安装配置管理做了极大简化。

官方参考文档

<https://www.cloudera.com/documentation/enterprise/6/6.0/topics/installation.html>

系统环境

- 操作系统 : 3台 CentOS 7.6 x64 (8C/32G/500G)
- Cloudera Manager : 6.2.0
- CDH: 6.2.0

注 : 10G以下内存建议还是别折腾了, 保证卡的让你失去耐心。我们正式环境都是7台主机*64G内存。

相关包的下载地址

Cloudera Manager下载地址 :

https://archive.cloudera.com/cm6/6.0.0/redhat7/yum/RPMS/x86_64/

- cloudera-manager-agent-6.0.0-530873.el7.x86_64.rpm
- cloudera-manager-daemons-6.0.0-530873.el7.x86_64.rpm
- cloudera-manager-server-6.0.0-530873.el7.x86_64.rpm
- cloudera-manager-server-db-2-6.0.0-530873.el7.x86_64.rpm
- oracle-j2sdk1.8-1.8.0+update141-1.x86_64.rpm

CDH安装包地址 : <https://archive.cloudera.com/cdh6/6.0.0/parcels/> , 由于我们的操作系统为CentOS7.3 , 需要下载以下文件 :

- [CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel](#)
- [CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel.sha256](#)
- [manifest.json](#)

准备工作 : 系统环境搭建

以下操作均用root用户操作。

1.网络配置(所有节点)

在所有节点上把IP和主机名的对应关系写入。

```
# vi /etc/hosts
```

```
192.168.10.73 cdh-master.test.com
```

```
192.168.10.74 cdh-slave.test.com
```

在相应的节点主机上修改主机名。

```
# vi /etc/sysconfig/network
```

修改或者添加HOSTNAME= cdh-master.test.com

执行命令

hostname cdh-master.test.com

CentOS7要多执行以下这步：

执行以下命令：

hostnamectl set-hostname cdh-master.test.com

注意：这里需要将每台机器的ip及主机名对应关系都写进去，本机的也要写进去，否则启动Agent的时候会提示hostname解析错误。

2.打通SSH，设置ssh无密码登陆（所有节点）

在主节点上执行`ssh-keygen -t rsa`一路回车，生成无密码的密钥对。

将公钥添加到认证文件中：

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

并设置authorized_keys的访问权限

```
chmod 600 ~/.ssh/authorized_keys。
```

scp文件到所有datanode节点：

```
scp ~/.ssh/authorized_keys root@cdh-slave.test.com:~/.ssh/
```

3.关闭防火墙和SELinux

注意：需要在所有的节点上执行，因为涉及到的端口太多了，临时关闭防火墙是为了安装起来更方便，安装完毕后可以根据需要设置防火墙策略，保证集群安全。

关闭防火墙并关闭自启动：

```
systemctl stop firewalld
```

```
Systemctl disable firewalld
```

关闭SELINUX（实际安装过程中发现没有关闭也是可以的，不知道会不会有问题，还需进一步进行验证）：

```
setenforce 0（临时生效）
```

```
修改 /etc/selinux/config 下的 SELINUX=disabled（重启后永久生效）
```

4.所有节点配置NTP服务

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：

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。

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：`yum install ntp。`

启动服务：`systemctl start ntpd`

配置开机启动：`systemctl enable ntpd`

主节点配置

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，`ntpdate -u ntp1.aliyun.com`

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：

```
driftfile /var/lib/ntp/drift
restrict 127.0.0.1
restrict -6 ::1
restrict default nomodify notrap
server ntp1.aliyun.com prefer
minpoll 6
includefile /etc/ntp/crypto/pw
keys /etc/ntp/keys
```

配置文件完成，保存退出，启动服务，执行如下命令：`service ntpd start`

检查是否成功，用ntpstat命令查看同步状态，出现以下状态代表启动成功：

```
synchronised to NTP server () at stratum 2
time correct to within 74 ms
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）

```
driftfile /var/lib/ntp/drift
restrict 127.0.0.1
restrict -6 ::1
restrict default kod nomodify notrap nopeer noquery
restrict -6 default kod nomodify notrap nopeer noquery
#这里是主节点的主机名或者ip
server cdh-master.test.com
minpoll 6
includefile /etc/ntp/crypto/pw
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：ntpdate -u cdh-master.test.com (主节点ntp服务器).

5.准备Parcels，用以安装CDH6

将CHD6相关的Parcel包放到主节点的/opt/cloudera/parcel-repo/目录中，如果没有此目录，可以自己创建。

- [CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel](#)
- [CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel.sha256](#)
- [manifest.json](#)

注意：最后将• CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel.sha256，重命名为• CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel.sha，这点必须注意否则，系统会重新下载• CDH-6.0.0-1.cdh6.0.0.p0.537114-el7.parcel文件。

正式开始安装

Step 1: Configure a Repository

1.1安装repo

wget <https://archive.cloudera.com/cm6/6.0.0/redhat7/yum/cloudera-manager.repo> -P /etc/yum/repos.d/

1.2导入GPG key

```
rpm --import https://archive.cloudera.com/cm6/6.0.0/redhat7/yum/RPM-GPG-KEY-cloudera
```

Step 2: Install JDK

```
yum install oracle-j2sdk1.8
```

Step 3: Install Cloudera Manager Server

带删除线的不用执行。直接略过。

```
yum install cloudera-manager-daemons-cloudera-manager-agent-cloudera-manager-server
```

3.1 yum安装CM

```
yum install cloudera-manager-server(前期只需要安装这个)
```

If you are using an Oracle database for Cloudera Manager Server, edit the /etc/default/cloudera-scm-server file on the Cloudera Manager server host. Locate the line that begins with export CM_JAVA_OPTS and change the -Xmx2G option to -Xmx4G.

3.2 开启https

开启https支持,由于开启后配置稍微复杂，暂时不配置。

```
JAVA_HOME=/usr/java/jdk1.8.0_141-cloudera /opt/cloudera/cm-agent/bin/certmanager setup --configure-services
```

注：后面的章节会重点介绍这部分

Step 4: Install Databases

4.1 Installing the MySQL Server

```
wget http://repo.mysql.com/mysql-community-release-el7-5.noarch.rpm
```

```
rpm -ivh mysql-community-release-el7-5.noarch.rpm
```

```
yum update
```

```
yum install mysql-server
```

```
systemctl start mysqld
```

```
systemctl enable mysqld
```

Cloudera Manager必须使用innodb引擎。使用MyISAM服务无法启动。

初始化Mysql

```
/usr/bin/mysql_secure_installation
```

```
[...]  
Enter current password for root (enter for none):  
OK, successfully used password, moving on...  
[...]  
Set root password? [Y/n] Y  
New password:  
Re-enter new password:  
Remove anonymous users? [Y/n] Y  
[...]  
Disallow root login remotely? [Y/n] N  
[...]  
Remove test database and access to it [Y/n] Y  
[...]  
Reload privilege tables now? [Y/n] Y  
All done!
```

官方推荐配置文件

```
[mysqld]  
datadir=/var/lib/mysql  
socket=/var/lib/mysql/mysql.sock  
transaction-isolation = READ-COMMITTED  
# Disabling symbolic-links is recommended to prevent assorted security risks;  
# to do so, uncomment this line:  
symbolic-links = 0  
key_buffer_size = 32M  
max_allowed_packet = 32M  
thread_stack = 256K  
thread_cache_size = 64  
query_cache_limit = 8M  
query_cache_size = 64M  
query_cache_type = 1  
max_connections = 550  
#expire_logs_days = 10  
#max_binlog_size = 100M  
#log_bin should be on a disk with enough free space.  
#Replace '/var/lib/mysql/mysql_binary_log' with an appropriate path for your  
#system and chown the specified folder to the mysql user.  
log_bin=/var/lib/mysql/mysql_binary_log  
#In later versions of MySQL, if you enable the binary log and do not set  
#a server_id, MySQL will not start. The server_id must be unique within  
#the replicating group.  
server_id=1  
binlog_format = mixed  
read_buffer_size = 2M  
read_rnd_buffer_size = 16M
```

```

sort_buffer_size = 8M
join_buffer_size = 8M
# InnoDB settings
innodb_file_per_table = 1
innodb_flush_log_at_trx_commit = 2
innodb_log_buffer_size = 64M
innodb_buffer_pool_size = 4G
innodb_thread_concurrency = 8
innodb_flush_method = O_DIRECT
innodb_log_file_size = 512M
[mysqld_safe]
log_error=/var/log/mysql.log
pid_file=/var/run/mysql/mysql.pid
sql_mode=STRICT_ALL_TABLES

```

4.2 Installing the MySQL JDBC Driver

```

wget https://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.46.tar.gz
tar zxvf mysql-connector-java-5.1.46.tar.gz
mkdir -p /usr/share/java/
cd mysql-connector-java-5.1.46
cp mysql-connector-java-5.1.46-bin.jar /usr/share/java/mysql-connector-java.jar

```

Databases for Cloudera Software

Service	Database	User
Cloudera Manager Server	scm	scm
Activity Monitor	amon	amon
Reports Manager	rman	rman
Hue	hue	hue
Hive Metastore Server	metastore	hive
Sentry Server	sentry	sentry
Cloudera Navigator Audit Server	nav	nav
Cloudera Navigator Metadata Server	navms	navms
Oozie	oozie	oozie

@51CTO博客

4.3 Creating Databases for Cloudera Software

需要建的库有

scm、amon、rman、hue、metastore、sentry、nav、navms、oozie

```

CREATE DATABASE scm DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON scm.* TO 'scm'@'%' IDENTIFIED BY 'scm@123';
CREATE DATABASE amon DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON amon.* TO 'amon'@'%' IDENTIFIED BY 'amon@123';
CREATE DATABASE rman DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON rman.* TO 'rman'@'%' IDENTIFIED BY 'rman@123';
CREATE DATABASE hue DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON hue.* TO 'hue'@'%' IDENTIFIED BY 'hue@123';
CREATE DATABASE metastore DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON metastore.* TO 'metastore'@'%' IDENTIFIED BY 'metastore@123';

```

```
CREATE DATABASE sentry DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON sentry.* TO 'sentry'@'%' IDENTIFIED BY 'sentry@123';

CREATE DATABASE nav DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON nav.* TO 'nav'@'%' IDENTIFIED BY 'nav@123';

CREATE DATABASE navms DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON navms.* TO 'navms'@'%' IDENTIFIED BY 'navms@123';

CREATE DATABASE oozie DEFAULT CHARACTER SET utf8 DEFAULT COLLATE utf8_general_ci;
GRANT ALL ON oozie.* TO 'oozie'@'%' IDENTIFIED BY 'oozie@123';
```

Step 5: Set up the Cloudera Manager Database

5.1语法

/opt/cloudera/cm/schema/scm_prepare_database.sh \

[options] <databaseType> <databaseName> <databaseUser> <password>

运行/opt/cloudera/cm/schema/scm_prepare_database.sh 可以查看参数

Parameters

Parameter (Required in bold)	Description
<databaseType>	One of the supported database types: <ul style="list-style-type: none"> • MariaDB: <code>mysql</code> • MySQL: <code>mysql</code> • Oracle: <code>oracle</code> • PostgreSQL: <code>postgresql</code>
<databaseName>	The name of the Cloudera Manager Server database to use. For MySQL, MariaDB, and PostgreSQL databases, the script can create the specified database if you specify the <code>-u</code> and <code>-p</code> options with the credentials of a user that has privileges to create databases and grant privileges. The default database name provided in the Cloudera Manager configuration settings is <code>scm</code> , but you are not required to use it.
<databaseUser>	The username for the Cloudera Manager Server database to create or use. The default username provided in the Cloudera Manager configuration settings is <code>scm</code> , but you are not required to use it.
<password>	The password for the <databaseUser> to create or use. If you do not want the password visible on the screen or stored in the command history, do not specify the password, and you are prompted to enter it as follows: <div>Enter SCM password:</div>

@51CTO博客

5.2执行

5.2.1 如果数据库和CM在一台服务器上

```
/opt/cloudera/cm/schema/scm_prepare_database.sh mysql scm scm
Enter SCM password:
JAVA_HOME=/usr/java/jdk1.8.0_141-cloudera
Verifying that we can write to /etc/cloudera-scm-server
Creating SCM configuration file in /etc/cloudera-scm-server
Executing: /usr/java/jdk1.8.0_141-cloudera/bin/java -cp /usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-
java.jar:/usr/share/java/postgresql-connector-java.jar:/opt/cloudera/cm/schema/../lib/* com.cloudera.enterprise.dbutil.DbCommandExecutor /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!
```

5.2.2 如果数据库和CM不在一台服务器上

```
/opt/cloudera/cm/schema/scm_prepare_database.sh mysql -h db01.example.com --scm-host cm01.example.com scm scm
Enter database password:
```

```

JAVA_HOME=/usr/java/jdk1.8.0_141-cloudera
Verifying that we can write to /etc/cloudera-scm-server
Creating SCM configuration file in /etc/cloudera-scm-server
Executing: /usr/java/jdk1.8.0_141-cloudera/bin/java -cp /usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-
java.jar:/usr/share/java/postgresql-connector-java.jar:/opt/cloudera/cm/schema/../lib/* com.cloudera.enterprise.dbutil.DbCommandExecutor /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!

```

5.2.3 如果你的数据库是oracle

```

/opt/cloudera/cm/schema/scm_prepare_database.sh -h cm-oracle.example.com oracle orcl sample_user sample_pass
JAVA_HOME=/usr/java/jdk1.8.0_141-cloudera
Verifying that we can write to /etc/cloudera-scm-server
Creating SCM configuration file in /etc/cloudera-scm-server
Executing: /usr/java/jdk1.8.0_141-cloudera/bin/java -cp /usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-
java.jar:/usr/share/java/postgresql-connector-java.jar:/opt/cloudera/cm/schema/../lib/* cloudera.enterprise.dbutil.DbCommandExecutor /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!

```

Step 6: Install CDH and Other Software

6.1 启动cm服务

```
systemctl start cloudera-scm-server
```

6.2 查看日志

```
tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log
```

显示如下表示服务已经正常起来。

```
INFO WebServerImpl:com.cloudera.server.cmf.WebServerImpl: Started Jetty server.
```

6.3 登录cm WEB界面

http://<server_host>:7180

如果你配置了auto-TLS，可以通过https登录。

https://<server_host>:7183

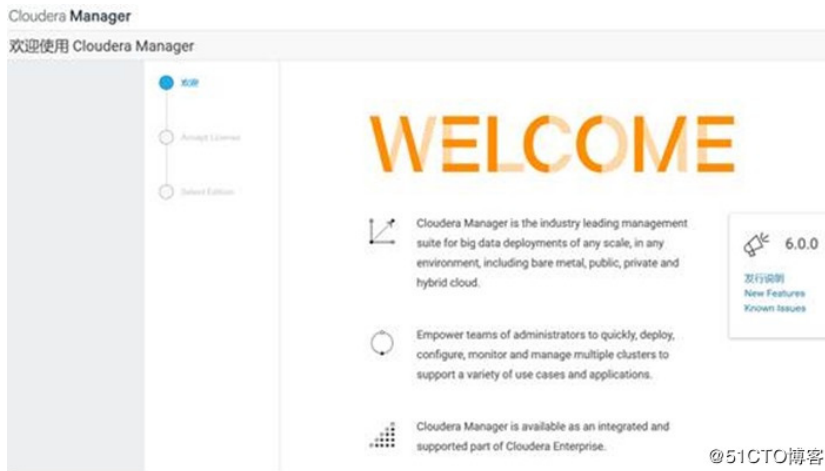
用户名admin

密码admin

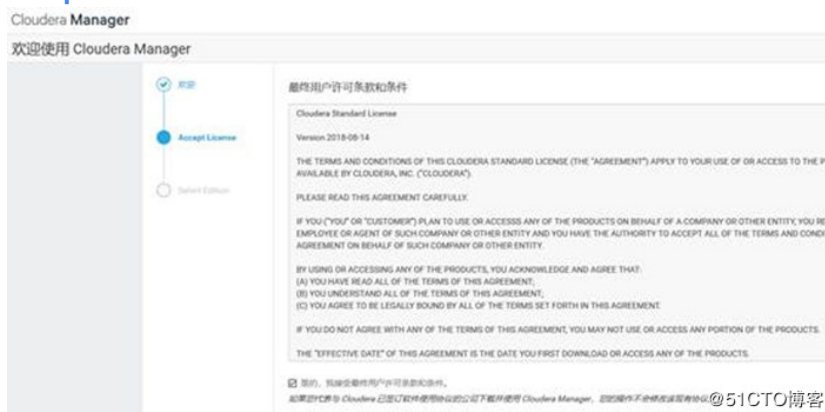


6.4 具体安装步骤

WELCOME



Accept License



Select Edition

版本选择免费版, 已经够用。



Welcome (Add Cluster - Installation)

安装

Specify Hosts

选择存储库

JDK 安装选项

提供 SSH 登录凭据

Install Agents

WELCOME

Adding a cluster in Cloudera Manager consists of two steps.

1

Add a set of hosts to form a cluster and install Cloudera Manager Agent software and CDH.

2

Select and configure CDH services to run on this cluster.

快速链接

安装指南

Supported Operating Systems

Supported Databases

Supported JDK Versions

@51CTO博客

Specify Hosts

Specify Hosts

应使用主机用于标识自身的同一主机名称 (FQDN) 来指定主机。Cloudera 建议包括 Cloudera Manager Server 的主机。这还将启用该主机的运行状况监控。

主机名称

cdh-master.rongyi.com
cdh-slave.rongyi.com

提示: 使用模式 ☐ 指定主机名称或 IP 地址。

SSH 端口:

搜索

已扫描 2 个主机, 其中 2 个正在运行 SSH。

<input checked="" type="checkbox"/> 已扫描主机	主机名称 (FQDN)	IP 地址	当前状态	结果
<input checked="" type="checkbox"/>	cdh-master.rongyi.com	192.168.10.73	否	已成功扫描主机。
<input checked="" type="checkbox"/>	cdh-slave.rongyi.com	192.168.10.74	否	已成功扫描主机。

Displaying 1 - 2 of 2

@51CTO博客

Select Repository

选择存储库

Cloudera Manager Agent

Cloudera Manager Agent 6.0.0 (#530873) needs to be installed on all new hosts.

Repository Location

☒ Public Cloudera Repository

Ensure the above version is listed in <https://archive.cloudera.com/cm6/>. Requires direct Internet access on all hosts.

☐ 自定义存储库

CDH and other software

Cloudera 建议使用 parcel 来代替软件包进行安装, 因为 parcel 可以使服务二进制文件的部署和升级自动化。让 Cloudera Manager 轻松地管理群集上的软件。如果选择不使用 parcel, 当有软件更新可用时, 将需要您手动升级群集中所有主机上的包, 并会阻止您使用 Cloudera Manager 的滚动升级功能。

选择方法

☐ 使用数据包

☒ 使用 Parcel (建议)

更多选项

代理设置

CDH 版本

☒ CDH 6.0.0-1.cdh6.0.0.p0.537114

☐ CDH 5.14.4-1.cdh5.14.4.p0.3

对于此 Cloudera Manager 版本 (6.0.0) 太新的 CDH 版本不会显示。

选择 Parcel

☐ ACCUMULO-1.7.2-5.5.0.ACCUMULO.05.5.0.p0.8

☒ 无

☐ KAFKA-3.1.0-1.3.1.0.p0.35

☒ 无

☐ KUDU-1.4.0-1.cdh5.12.2.p0.8

☒ 无

☐ SPARK-0.9.0-1.cdh4.6.0.p0.98

☒ 无

☐ SQOOP_NETEZZA_CONNECTOR-1.5.1e6

☐ SQOOP_NETEZZA_CONNECTOR-1.5.1e5

@51CTO博客

JDK 安装选项



要允许Cloudera管理器在集群主机上自动安装Oracle JDK，请阅读JDK许可并选中标有install Oracle Java SE Development Kit (JDK)的复选框(如果您接受这些条件)。

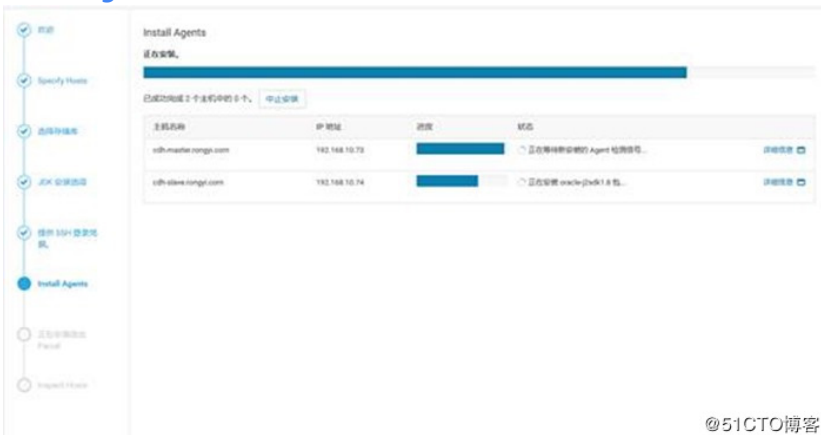
如果您在步骤2中安装了自己的Oracle JDK版本:安装Java开发工具包，请不要选中该框。

如果您允许Cloudera管理器安装JDK，则会出现第二个复选框，该复选框标有install Java Unlimited Strength Encryption Policy文件。在JDK版本中，要启用小于1.8u161的AES-256加密，需要这些策略文件。JDK 1.8u161及更高版本默认支持无限强度加密，并且不需要策略文件。在阅读许可条款并检查适用框之后，单击Continue。

Enter Login Credentials



Install Agents



Install Parcels



Inspect Hosts



6.5 遇到的问题

服务器检查，可能会遇到以下问题：

6.5.1 虚拟内存设置

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：

临时解决

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

永久解决

```
sysctl -w vm.swappiness=0
```

```
echo vm.swappiness = 0 >> /etc/sysctl.conf
```

6.5.2 大内存页设置

大内存页禁用

```
echo never>/sys/kernel/mm/transparent_hugepage/defrag
```

```
echo never>/sys/kernel/mm/transparent_hugepage/enabled
```

6.5.3 升级软件依赖版本

Starting with CDH 6, PostgreSQL-backed Hue requires the Psycopg2 version to be at least 2.5.4, see the documentation for more information. This warning can be ignored if hosts will not run CDH 6, or will not run Hue with PostgreSQL. The following hosts have an incompatible Psycopg2 version of '2.5.1':

解决方法：可以忽略。

```
yum install python-pip
```

```
pip install --upgrade psycopg2
```

6.5.4 无法正常安装agent



原因:由于我一开始起用了Https支持,所以安装agent时无法和master进行加密通信。其实agent是已经安装好了。最后把CM卸载重新安装,不启用https。

下面附卸载命令:

```
service cloudera-scm-agent stop
```

```
service supervisord stop
```

```
systemctl stop cloudera-scm-server
```

```
yum remove cloudera-manager-daemons cloudera-manager-agent cloudera-manager-server
```

```
rm /opt/cloudera/*
```

数据库部分: drop database scm.

6.5.5 安装Parcel提示主机运行状况不良

Cloudera 安装Parcel提示主机运行状况不良



解决方法

删除agent目录下面的cm_guid文件,并重启失败节点的agent服务恢复。

```
[root@esgyn001 ~]# cd /var/lib/cloudera-scm-agent/
```

```
cm_guid response.avro uuid
```

```
[root@esgyn001 cloudera-scm-agent]# rm -rf cm_guid
```

```
[root@esgyn001 cloudera-scm-agent]# service cloudera-scm-agent restart
```

```
Stopping cloudera-scm-agent: [ OK ]
```

```
Starting cloudera-scm-agent:
```

Step 7: Set Up a Cluster

7.1 具体步骤

Select Services

Select Services

自定义角色分配

数据仓库设置

审核配置

安全策略配置

Summary

Select Services

选择要安装的服务组合。

Essentials

Management and support for Cloudera's distribution including Hadoop.

服务: HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, and Hue

Data Engineering

Process, develop, and serve predictive models.

服务: HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Spark

Data Warehouse

The modern data warehouse for today, tomorrow, and beyond.

服务: HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Impala

Operational Database

Real-time insights for modern data-driven business.

服务: HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and HBase

所有服务

Everything you need to become information-driven, with complete use of the platform.

服务: HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, HBase, Impala, Solr, Spark, and Key-Value Store Indexer

自定义服务

选择您自己的服务。将自动包含的服务所需的组件。只有在安装了初始部署之后才能添加 Flume。

本向导将对安装 Cloudera Management Service, 有一系列组件可应用监控、报告、事件和警报; 这些组件需要数据库存储信息。

在下一页上安装

自定义角色分配

选择要安装角色到那台服务器上。

自定义角色分配

您可在自定义部署的角色分配, 但如果分配不正确 (例如, 分配到某个主机上的角色太多) 会影响服务性能。除非有特殊情况, 如已为特定角色预先选择特定主机, 否则 Cloudera 不建议自定义分配。

还可以按主机查看角色分配。

按主机查看

HBase

H Master x 1 新建

cdh-master.rongyi.com

HBase REST Server

选择主机

HBase Thrift Server

选择主机

H RegionServer x 1 新建

与 DataNode 相同

HDFS

NameNode x 1 新建

cdh-master.rongyi.com

SecondaryNameNode x 1 新建

cdh-master.rongyi.com

Balancer x 1 新建

cdh-master.rongyi.com

HttpFS

选择主机

NFS Gateway

选择主机

DataNode x 1 新建

cdh-slave.rongyi.com

Hive

Gateway x 2 新建

cdh-master.rongyi.com, cdh-slave.rongyi.com

Hive Metastore Server x 1 新建

cdh-master.rongyi.com

WebHcat Server

选择主机

HiveServer2 x 1 新建

cdh-master.rongyi.com

Hue

Hue Server x 1 新建

cdh-master.rongyi.com

Load Balancer x 1 新建

cdh-master.rongyi.com

角色分配官方推荐：

https://www.cloudera.com/documentation/enterprise/6/6.0/topics/cm_ig_host_allocations.html#host_role_assignment

3 - 10 Worker Hosts without High Availability			
Master Hosts	Utility Hosts	Gateway Hosts	Worker Hosts
Master Host 1: <ul style="list-style-type: none"> NameNode YARN ResourceManager JobHistory Server ZooKeeper Kudu master Spark History Server 	One host for all Utility and Gateway roles: <ul style="list-style-type: none"> Secondary NameNode Cloudera Manager Cloudera Manager Management Service Hive Metastore HiveServer2 Impala Catalog Server Impala StateStore Hue Oozie Flume Gateway configuration 		3 - 10 Worker Hosts: <ul style="list-style-type: none"> DataNode NodeManager Impalad Kudu tablet server

数据库设置

数据库设置

配置和测试数据库连接。首先根据[Installation Guide](#)的[Installing and Configuring an External Database](#)小节创建数据库。

Hive

数据库主机名称: * 数据库类型: 数据库名称: * 用户名: * 密码:

Oozie Server

当前被分配在 cdh-master.rongyi.com 上运行。
数据库主机名称: * 数据库类型: 数据库名称: * 用户名: * 密码:

Hue

数据库主机名称: * 数据库类型: 数据库名称: * 用户名: * 密码:

☐ 显示密码

@51CTO博客

审核更改

mkdir -p /data/

审核更改

HDFS 根目录 Cluster 1 > HBase (服务范围) ⓘ

应用程序路径 Cluster 1 > HBase (服务范围) ⓘ

应用程序 Cluster 1 > HBase (服务范围) ⓘ

HDFS 块大小 兆字节 ⓘ

修复的 DataNode 失败的卷 Cluster 1 > DataNode Default Group ⓘ

DataNode 数据目录 Cluster 1 > DataNode Default Group ⓘ

NameNode 数据目录 Cluster 1 > NameNode Default Group ⓘ

HDFS 检查点目录 Cluster 1 > SecondaryNameNode Default Group ⓘ

@51CTO博客

需要修改的选项如下：（不修改也可以，根据自己需要）

HDFS 根目录 /data/hbase

DataNode 数据目录 /data/dfs/dn

NameNode 数据目录 /data/dfs/nn

HDFS 检查点目录 /data/dfs/snn

Hive 仓库目录 /data/user/hive/warehouse

ShareLib 根目录 /data/user/oozie

NodeManager 本地目录 /data/yarn/nm

注意：关于/solr的目录不要修改，不然后面会报错。

命令详细信息

首次运行命令

状态: ☐ 正在运行 ☒ 已失败 9月 5, 7:21:34 晚上 重试

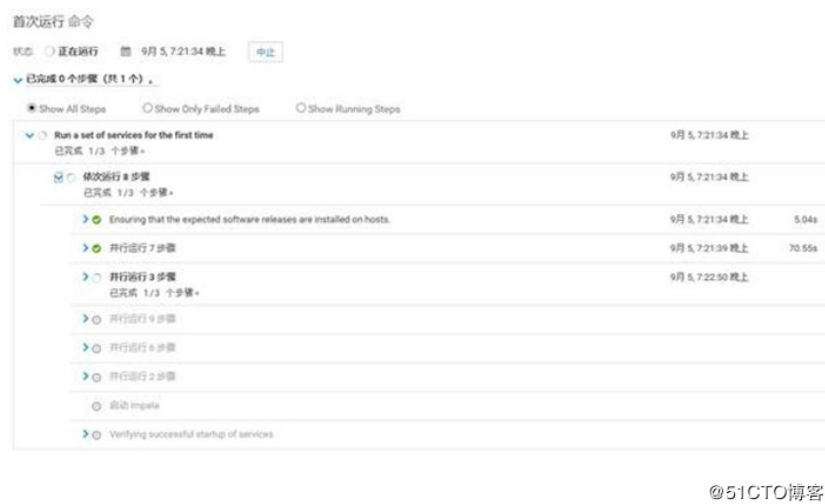
已失败 0 个步骤 (共 1 个) ⓘ

显示所有步骤 ☒ 显示仅失败的步骤 ☐ 显示正在运行的步骤

Run a set of services for the first time 9月 5, 7:21:34 晚上
已失败 5/7 个步骤 ⓘ

继续执行 8 步骤 9月 5, 7:21:34 晚上
已失败 5/7 个步骤 ⓘ

@51CTO博客



Solr服务初始化失败

如果安装过程中，遇到如下错误安装过程中，我修改了默认的/solr的目录为/data/solr,报错无法执行初始化操作。

解决方法

退到上一步，重新修改目录为默认目录。不要修改默认目录。默认是/solr



Summary



点击完成，安装完成。

7.2 完成界面及警告处理

界面如下，红色和×××的部分是警告，根据提示来做相应处理。



此处遇到的问题

问题1. NFS GATEWAY ?

No portmap or rpcbind service is running on this host. Please start portmap or rpcbind service before attempting to start the NFS Gateway role on this host.

解决方法：

安装yum install portmap

[root@my-cdh-01 hue-httpd]# service rpcbind start

[root@my-cdh-01 hue-httpd]# systemctl enable rpcbind

问题2. q-bi-07: [Memory Overcommit Validation Threshold Suppress...](#) (此问题是生产环境遇到的)

主机 q-bi-07 上的内存被调拨过度。总内存分配额是 53.5 GiB 个字节，但是 RAM 只有 62.8 GiB 个字节（其中的 12.6 GiB 个字节是保留给系统使用的）。如需获得分配详细信息，请访问“主机”页面上的“资源”选项卡。重新配置主机上的角色以降低总内存分配额。请注意：Java 最大堆大小乘以 1.3 等于近似的 JVM 开销。

答：当为该节点上的服务分配的内存大于该节点可用的总内存时(注意，默认情况下，20%的内存是为系统保留的)，这个错误就支持了我们转到每个主机上的参考资料页，查看每个服务使用了多少内存。您可以通过转到configuration选项卡(用于每个服务)并更改它(在配置页面的搜索栏中使用“memory”或“heap”来打开这些属性)来降低每个服务使用的内存数量这样做可以避免这个错误，但这不是推荐的修复方法。

首先出现这个错误的原因是您可能在集群上运行了太多的服务。试着去掉一些你不用的东西。您总是可以稍后添加它们(当您添加更多内存时)。

选择主机—所有主机—选择报警的主机—资源（检查相应角色的内存分配，然后选择服务—配置—搜索“memory”或“heap”进行相应的更改，并重启相应服务，生效）

内存			
服务	实例	限制	单位
H-HBase	RegionServer	HBase RegionServer 的 Java 堆栈大小 (字节) (> 30% 开销)	2.6 吉字节
H-HBase	RegionServer	BulkWriteCache 大小	0 字节
H-HDFS	Datanode	Datanode 的 Java 堆栈大小 (字节) (> 30% 开销)	1.3 吉字节
H-HDFS	Datanode	用于备份的总内存	4.0 吉字节
I-Impala	Impala Daemon	Impala Daemon 内存限制	14.2 吉字节
K-Kudu	Tablet Server	Kudu Tablet Server 内存限制	4.0 吉字节
N-NAMN (MR2 Included)	NodeManager	NodeManager 的 Java 堆栈大小 (字节) (> 30% 开销)	1.3 吉字节
N-NAMN (MR2 Included)	NodeManager	容器内存	

Step8安全:加密传输数据

为Cloudera管理器配置TLS加密。传输层安全(TLS)是一套工业标准的加密协议，用于保护网络上的通信安全。TLS是从安全套接字层(SSL)发展而来的。由于SSL术语仍然广泛使用，Cloudera软件和文档将TLS称为TLS/SSL，但实际使用的协议是TLS。在Cloudera软件中没有使用SSL。

除了TLS/SSL加密之外，使用远程过程调用(rpc)的HDFS和HBase传输数据。为了保护这个传输，您必须启用RPC加密。有关启用TLS/SSL和RPC加密的说明，请查看官方文档。

8.1 配置自动TLS加密为Cloudera管理器和CDH

自动TLS说明

对于新安装，Cloudera建议将自动tls作为安装过程的一部分。

自动TLS极大地简化了在集群上启用和管理TLS加密的过程。它自动创建内部证书颁发机构(CA)并跨所有集群主机部署证书。它还可以自动分发现有证书，例如由公共CA签名的证书。在启用自动tls的集群中添加新的集群主机或服务，将自动创建和部署所需的证书。

有关启用自动TLS的说明，请参阅使用自动TLS安装Cloudera管理器、CDH和托管服务以及为Cloudera管理器和CDH配置TLS加密。

8.1.1 允许自动tls自动创建证书

在Cloudera管理器服务器主机上安装了Cloudera管理器代理之后，在启动Cloudera管理器之前，按照如下方式启用自动tls。

在Cloudera管理器服务器主机上，初始化证书管理器并配置Cloudera管理器服务器，以通过运行以下命令启用自动tls:

```
JAVA_HOME=/usr/java/jdk1.8.0_141 /opt/cloudera/cm-agent/bin/certmanager setup --configure-services
```

用您环境的正确路径替换JDK版本。如果希望将文件存储在默认目录(/var/lib/cloudera-scm-server/certmanager)以外的目录中，请添加--location选项如下:

```
JAVA_HOME=/usr/java/jdk1.8.0_141 /opt/cloudera/cm-agent/bin/certmanager --location/opt/cloudera/CMCAsetup --configure-services
```

用要使用的目录路径替换/opt/cloudera/CMCA。这个目录是自动创建的，在运行命令之前不存在。检查/var/log/cloudera-scm-agent/certmanager.log 日志文件，以确认目录已创建。

如果Cloudera管理器服务器正在运行，重新启动它:

```
systemctl restart cloudera-scm-server
```

8.1.2 使用现有证书启用自动-TLS

在启用自动tls之前，请确保您拥有所有集群主机的主机证书。在Cloudera管理器服务器主机上安装了Cloudera管理器代理之后，在启动Cloudera管理器之前，按照如下方式启用自动tls:

a、初始化证书管理器并配置Cloudera管理器服务器，以使用自定义证书目录启用自动tls:

```
JAVA_HOME=/usr/java/jdk1.8.0_141 /opt/cloudera/cm-agent/bin/certmanager --location/opt/cloudera/CMCAsetup_custom_certdir --configure-services --host-cert/path/to/hostcert.pem--host-key/path/to/hostkey.pem--ca-cert/path/to/cacert.pem--keystore-pw-file/path/to/keystore_pw_file--truststore-pw-file/path/to/truststore_pw_file
```

用适合您环境的正确路径替换JDK版本及其它相关路径。如：/path/to/hostcert.pem，/path/to/hostkey.pem，观察日志是否有报错：
/var/log/cloudera-scm-agent/certmanager.log

b、对于每个集群主机(具有唯一的证书和密钥)，运行以下命令:

```
JAVA_HOME=/usr/java/jdk1.8.0_141 /opt/cloudera/cm-agent/bin/certmanager --location/opt/cloudera/CMCA add_custom_cert --host-cert <host_cert> --host-key <host_key> <hostname>
```

用适合您环境的正确路径替换JDK版本。

c、如果Cloudera Manager Server是运行的状态，请重启。

```
systemctl restart cloudera-scm-server
```

8.1.3 自动tls代理文件位置

由自动tls生成的证书、密钥存储库和密码文件存储在每个Cloudera管理器代理上的
/var/lib/Cloudera -scm- Agent/Agent-cert中。文件名如下：

Auto-TLS Agent Files	
Filename	Description
cm-auto-global_cacerts.pem	CA证书和其他PEM格式的受信任证书
cm-auto-global_truststore.jks	CA证书和JKS格式的其他受信任证书
cm-auto-in_cluster_ca_cert.pem	PEM格式的CA证书
cm-auto-in_cluster_truststore.jks	JKS格式的CA证书
cm-auto-host_key_cert_chain.pem	代理主机证书和PEM格式的私钥
cm-auto-host_cert_chain.pem	PEM格式的代理主机证书
cm-auto-host_key.pem	代理主机私钥的PEM格式

cm-auto-host_key.pem	代理主机私钥的PEM格式
cm-auto-host_keystore.jks	代理主机私钥的JKS格式
cm-auto-host_key.pw	代理主机私钥密码文件

8.2 Cloudera Manager配置TLS加密

现有集群不支持Auto-TLS。要在现有集群上启用TLS（在已经安装好的集群上启用TLS），请遵循下面的说明。

https://www.cloudera.com/documentation/enterprise/6/6.0/topics/how_to_configure_cm_tls.html#concept_wk4_jl

1. [Generate TLS Certificates](#)
2. [Configure TLS for the Cloudera Manager Admin Console](#)
3. [Configure TLS for Cloudera Manager Agents](#)
4. [Enable Server Certificate Verification on Cloudera Manager Agents](#)
5. [Configure Agent Certificate Authentication](#)