

CDH5.6.0单节点安装文档

集群节点数量1个,IP为192.168.4.22。

一、安装CentOS6.7（64位）并搭建基本环境

- 1.主机参数优化
- 2.主机/etc/hosts文件修改

二、下载clouder-manger的安装包及parcel文件

- 1.下载cloudera Manager地址：

下载地址:http://archive.cloudera.com/cm5/cm/5/

下载文件：cloudera-manager-el6-cm5.6.0.x8664.tar.gz

- 2.CDH安装包地址：

下载地址：http://archive-primary.cloudera.com/cdh5/parcels/5.6.0/

下载文件：CDH-5.6.0-1.cdh5.6.0.p0.45-el6.parcel，CDH-5.6.0-1.cdh5.6.0.p0.45-el6.parcel.sha1，manifest.json

- 3.JDK下载

下载地址：http://archive-primary.cloudera.com/cm5/redhat/6/x8664/cm/5.2.1/RPMS/x8664/

下载文件：jdk-6u31-linux-amd64.rpm，oracle-j2sdk1.7-1.7.0+update67-1.x86_64.rpm

- 4.JDBC驱动

下载地址：http://download.softagency.net/MySQL/Downloads/Connector-J/

下载文件：mysql-connector-java-5.1.38.tar.gz

三、安装JDK环境

```
$ rpm -Uvh jdk-6u31-linux-amd64.rpm
$ rpm -Uvh oracle-j2sdk1.7-1.7.0+update67-1.x86_64.rpm
```

四、部署CDM和CDH

- 1.CDM部署

上传cloudera-manager-el6-cm5.2.1.x8664.tar.gz到主机/opt目录下

```
$ cd /opt/
$ tar xf cloudera-manager-el6-cm5.6.0.x86_64.tar.gz
```

- 2.CDH部署

上传CDH-5.6.0-1.cdh5.6.0.p0.45-el6.parcel、CDH-5.6.0-1.cdh5.6.0.p0.45-el6.parcel.sha1、manifest.json 到/opt/cloudera/parcel-repo目录下

```
$ mv CDH-5.6.0-1.cdh5.6.0.p0.45-el6.parcel.sha1 CDH-5.6.0-1.cdh5.6.0.p0.45-el6.parcel.sha
```

如果不重命名CDH会识别不了版本

```
$ chown -R root.root /opt/*
```

- 3.JDBC驱动部署

拷贝mysql-connector-java-5.1.7-bin.jar 到 /usr/share/java下并重命名mysql-connector-java.jar

```
$ mkdir -p /usr/share/java && cd /usr/share/java
$ mv mysql-connector-java-5.1.38-bin.jar mysql-connector-java.jar
```

五、mysql 数据库配置

- 1.创建数据库

创建cloudera数据库:

```
mysql >create database cloudera56 character set utf8;
mysql >grant all privileges on cloudera56.* to cloudera56@localhost identified by '123456';
mysql >grant all privileges on cloudera56.* to cloudera56@'%' identified by '123456';
mysql >flush privileges;
```

创建Hive数据库:

```
mysql >create database hive56 default character set latin1;
```

字符必须为拉丁(latin1),如果utf8可能会报错Specified key was too long; max key length is 767 bytes

```
mysql >grant all privileges on hive56 .* to hive56@localhost identified by '123456';
mysql >grant all privileges on hive56 .* to hive56@'%' identified by '123456';
mysql >flush privileges;
```

创建report数据库

```
mysql >create database report56 character set utf8;
mysql >grant all privileges on report56.* to report56@localhost identified by '123456';
mysql >grant all privileges on report56.* to report56@'%' identified by '123456';
mysql >flush privileges;
```

创建oozie数据库

```
mysql >create database oozie56 character set utf8;
mysql >grant all privileges on oozie56.* to oozie56@localhost identified by '123456';
mysql >grant all privileges on oozie56.* to oozie56@'%' identified by '123456';
mysql >flush privileges;
```

- 2.添加系统用户并初始化数据库

添加用户

```
useradd --system --home=/opt/cm-5.6.0/run/cloudera-scm-server/ --no-create-home
--shell=/bin/false --comment "Cloudera SCM User"
```

初始化数据库

```
cloudera-scm /opt/cm-5.6.0/share/cmf/schema/scm_prepare_database.sh -h <IP of MySQL> mysql cloudera56 cloudera56 123456
```

```
[root@cdh ~]# /opt/cm-5.6.0/share/cmf/schema/scm_prepa
JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera
Verifying that we can write to /opt/cm-5.6.0/etc/cloud
Creating SCM configuration file in /opt/cm-5.6.0/etc/c
groups: cloudera-scm: No such user
Executing: /usr/java/jdk1.7.0_67-cloudera/bin/java -c
tor-java.jar:/opt/cm-5.6.0/share/cmf/schema/./lib/* c
ra-scm-server/db.properties com.cloudera.cmf.db.
[ main] DbCommandExecutor
All done, your SCM database is configured correctly!
[root@cdh ~]#
```

六、启动服务器

启动cm server :

```
$ /opt/cm-5.6.0/etc/init.d/cloudera-scm-server start
$ /opt/cm-5.6.0/etc/init.d/cloudera-scm-server status
```

```
[root@cdh ~]# /opt/cm-5.6.0/etc/init.d/cloudera-scm-server start
Starting cloudera-scm-server: [ OK ]
[root@cdh ~]# /opt/cm-5.6.0/etc/init.d/cloudera-scm-server status
cloudera-scm-server (pid 1746) is running...
[root@cdh ~]#
```

七、启动cm agent配置与启动

修改客户端配置文件config.ini

```
$ vi /opt/cm-5.6.0/etc/cloudera-scm-agent/config.ini
```

修改host=后面的值为cdh

取消parcel_dir屏蔽

```
# Parcel directory. Unpacked parcels will be stored in this directory.
# Downloaded parcels will be stored in <parcel_dir>/../parcel-cache
parcel_dir=/opt/cloudera/parcels
```

启动客户端

```
$ /opt/cm-5.6.0/etc/init.d/cloudera-scm-agent start
$ /opt/cm-5.6.0/etc/init.d/cloudera-scm-agent status
```

```
[root@cdh ~]# /opt/cm-5.6.0/etc/init.d/cloudera-scm-agent start
Starting cloudera-scm-agent: [ OK ]
[root@cdh ~]# /opt/cm-5.6.0/etc/init.d/cloudera-scm-agent status
cloudera-scm-agent (pid 1824) is running...
[root@cdh ~]#
```

八、验证CDH安装是否正常

Server启动1-2分钟后, 才可能看到结果

1.查看端口

7180是CDH server监听端口

```
$ netstat -anlp|grep 7180
```

```
[root@cdh ~]# netstat -anlp|grep 7180
tcp        0      0 0.0.0.0:*                0.0.0.0:*                LISTEN      1920/java
tcp        0      0 192.168.4.220:7180      192.168.3.44:59634      ESTABLISHED 1920/java
tcp        0      0 192.168.4.220:7180      192.168.3.44:59635      ESTABLISHED 1920/java
tcp        0      0 192.168.4.220:7180      192.168.3.44:59640      ESTABLISHED 1920/java
tcp        0      0 192.168.4.220:7180      192.168.3.44:59638      ESTABLISHED 1920/java
tcp        0      0 192.168.4.220:7180      192.168.3.44:59641      ESTABLISHED 1920/java
tcp        0      0 192.168.4.220:7180      192.168.3.44:59639      ESTABLISHED 1920/java
```

7182是CDH Agent监听端口

```
[root@cdh ~]# netstat -anlp|grep 7182
tcp        0      0 0.0.0.0:*                0.0.0.0:*                LISTEN      1706/java
tcp        0      0 192.168.4.220:64451     192.168.4.220:7182      ESTABLISHED 1753/python
tcp        0      0 192.168.4.220:7182     192.168.4.220:64451     ESTABLISHED 1706/java
```

2.查看log

```
$ cd /opt/cm-5.6.0/log/
```

```
[root@cdh ~]# cd /opt/cm-5.6.0/log/
[root@cdh log]# ll
total 12
drwxr-xr-x 2 root root 4096 Mar 16 11:39 cloudera-scm-agent
drwxr-xr-x 2 root root 4096 Mar 16 11:38 cloudera-scm-server
drwxr--r-- 2 root root 4096 Feb 12 12:05 cmf
```

九、启动浏览器，开始集群web安装配置

访问http://<ip address>:7180,账号:admin 密码:admin

cloudera manager

支持 admin

Welcome to Cloudera Manager

End User License Terms and Conditions

Cloudera Standard License

Version 2015-08-06

END USER LICENSE TERMS AND CONDITIONS

THESE TERMS AND CONDITIONS (THESE "TERMS") APPLY TO YOUR USE OF THE PRODUCTS (AS DEFINED BELOW) PROVIDED BY CLOUDERA, INC. ("CLOUDERA").

PLEASE READ THESE TERMS CAREFULLY.

IF YOU ("YOU" OR "CUSTOMER") PLAN TO USE ANY OF THE PRODUCTS ON BEHALF OF A COMPANY OR OTHER ENTITY, YOU REPRESENT THAT YOU ARE THE EMPLOYEE OR AGENT OF SUCH COMPANY (OR OTHER ENTITY) AND YOU HAVE THE AUTHORITY TO ACCEPT ALL OF THE TERMS AND CONDITIONS SET FORTH IN AN ACCEPTED REQUEST (AS DEFINED BELOW) AND THESE TERMS (COLLECTIVELY, THE "AGREEMENT") ON BEHALF OF SUCH COMPANY (OR OTHER ENTITY).

BY USING ANY OF THE PRODUCTS, YOU ACKNOWLEDGE AND AGREE THAT:

(A) YOU HAVE READ ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT;

(B) YOU UNDERSTAND ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT;

(C) YOU AGREE TO BE LEGALLY BOUND BY ALL OF THE TERMS AND CONDITIONS SET FORTH IN THIS AGREEMENT

IF YOU DO NOT AGREE WITH ANY OF THE TERMS OR CONDITIONS OF THESE TERMS, YOU MAY NOT USE ANY PORTION OF THE PRODUCTS.

☒ Yes, I accept the End User License Terms and Conditions.

1 2

继续

Welcome to Cloudera Manager

Which edition do you want to deploy?

升级到 **Cloudera Enterprise Data Hub Edition** 将提供可以帮助您在关键任务环境下管理和监控 Hadoop 群集的重要功能。

Cloudera Express		Cloudera Enterprise 数据集线器试用版	Cloudera Enterprise
许可证	免费	60 天 在试用期之后，该产品将继续作为 Cloudera Express 运行。您的群集和数据将会保持不受影响。	年度订阅 上传许可证 选择许可证文件 上传 Cloudera Enterprise 在三个版本中可用： <ul style="list-style-type: none">Basic EditionFlex EditionData Hub Edition
节点限制	无限制	无限制	无限制
CDH	✓	✓	✓
Cloudera Manager 核心功能	✓	✓	✓
Cloudera Manager 高级功能		✓	✓
Cloudera Navigator		✓	✓
Cloudera Navigator Key Trustee			✓

返回

12

继续

感谢您选择 Cloudera Manager 和 CDH。

将安装此安装程序**Cloudera Enterprise Data Hub Edition 试用版5.6.0**，您可以稍后通过此安装程序选择以下服务的软件包（可能会涉及到许可证）。

- Apache Hadoop（Common、HDFS、MapReduce、YARN）
- Apache HBase
- Apache ZooKeeper
- Apache Oozie
- Apache Hive
- Hue（已获 Apache 许可）
- Apache Flume
- Cloudera Impala（许可的 Apache）
- Apache Sentry
- Apache Sqoop
- Cloudera Search（许可的 Apache）
- Apache Spark

您正在使用 Cloudera Manager 安装和配置您的系统。您可以通过单击上面的**支持**菜单了解更多有关 Cloudera Manager 的信息。

继续

为 CDH 群集安装指定主机。

新主机

当前管理的主机 (1)

提示: 使用模式 搜索主机名称和/或 IP 地址。

cdh

SSH 端口: 22 搜索

为 CDH 群集安装指定主机。

新主机

当前管理的主机 (1)

提示: 使用[模式](#) 或搜索主机名称和/或 IP 地址。

已扫描 1 个主机, 其中 1 个正在运行 SSH。

新搜索

<input type="checkbox"/> 已扩展查询	主机名称 (FQDN)	IP 地址	当前受管	结果
<input type="checkbox"/> cdh	cdh	192.168.4.220	是	✔ 主机准备就绪: 1 毫秒响应时间。

为 CDH 群集安装指定主机。

新主机

当前管理的主机 (1)

这些主机不属于任何群集。请选择组成群集的主机。

<input type="checkbox"/>	名称	IP	机架	CDH 版本	状态	上一检测信号
	<input type="text" value="任何名称"/>	<input type="text" value="任何 IP"/>	<input type="text" value="任何机架"/>	<div>全部</div>	<div>全部</div>	<div>全部</div>
<input checked="" type="checkbox"/>	cdh	192.168.4.220	/default	无	⌚ 未知运行状况	13.75s ago

返回

继续

群集安装

选择存储库

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

选择 CDH 的版本

☒ CDH-5.6.0-1.cdh5.6.0.p0.45

☐ CDH-4.7.1-1.cdh4.7.1.p0.47

对于此 Cloudera Manager 版本 (5.6.0) 太新的 CDH 版本不会显示。

其他 Parcel

☐ ACCUMULO-1.6.0-1.cdh5.1.4.p0.116

☐ ACCUMULO-1.4.4-1.cdh4.5.0.p0.65

☒ 无

☐ KAFKA-2.0.0-1.kafka2.0.0.p0.12

☒ 无

返回

1234

继续

更多选项

代理用户

?

本地 Parcel 存储库路径

/opt/cloudera/parcel-repo

?

Parcel 目录

/opt/cloudera/parcels

?

需要重启代理

远程 Parcel 存储库 URL

https://archive.cloudera.com/cdh5/parcels/{latest_supported}/

+

-

←

?

https://archive.cloudera.com/cdh4/parcels/latest/

+

-

https://archive.cloudera.com/impala/parcels/latest/

+

-

https://archive.cloudera.com/search/parcels/latest/

+

-

https://archive.cloudera.com/accumulo/parcels/1.4/

+

-

确定

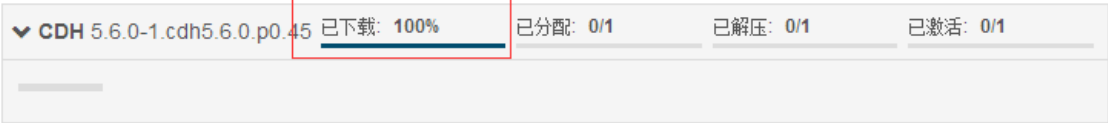
取消

只要文件路径配置正确,已下载就是100%,如果出现进度条1%，必须检查文件和客户端配置

群集安装

正在安装选定 Parcel

选定的 Parcel 正在下载并安装在群集的所有主机上。



群集安装

正在安装选定 Parcel

选定的 Parcel 正在下载并安装在群集的所有主机上。



群集安装


正在安装选定 Parcel

选定的 Parcel 正在下载并安装在群集的所有主机上。



群集安装

检查主机正确性

正在检查主机...这可能需要一分钟时间。 

⏮ 跳过主机检查器

群集安装

检查主机正确性 🔄 重新运行

验证

- ✓ 检查器在所有 1 个主机上运行。
- ✓ 检查主机名称时发现以下失败结果...
- ✓ 查询存在冲突的初始脚本时未发现错误。
- ✓ 检查 `/etc/hosts` 时未发现错误。
- ✓ 所有主机均将 `localhost` 解析为 `127.0.0.1`。
- ✓ 检查过的所有主机均正确且及时地解析了彼此的主机名称。
- ✓ 主机时钟几乎同步（10 分钟内）。
- ✓ 整个群集中的主机时区一致。
- ✓ 无用户或组缺失。
- ✓ 软件包和 `parcel` 之间未检测到冲突。
- ✓ 没有存在已知错误的内核版本在运行。

⏮ 返回

1 2 3 4

⏭ 完成

群集设置

选择您要在群集上安装**CDH 5** 服务。

选择要安装的服务组合。

- ☒ **核心 Hadoop**
HDFS、YARN (包括 MapReduce 2)、ZooKeeper、Oozie、Hive、Hue 和 Sqoop
- ☐ **含 HBase 的内核**
HDFS、YARN (包括 MapReduce 2)、ZooKeeper、Oozie、Hive、Hue、Sqoop 和 HBase
- ☐ **含 Impala 的内核**
HDFS、YARN (包括 MapReduce 2)、ZooKeeper、Oozie、Hive、Hue、Sqoop 和 Impala
- ☐ **含 Search 的内核**
HDFS、YARN (包括 MapReduce 2)、ZooKeeper、Oozie、Hive、Hue、Sqoop 和 Solr
- ☐ **含 Spark 的内核**
HDFS、YARN (包括 MapReduce 2)、ZooKeeper、Oozie、Hive、Hue、Sqoop 和 Spark
- ☐ **所有服务**
HDFS、YARN (包括 MapReduce 2)、ZooKeeper、Oozie、Hive、Hue、Sqoop、HBase、Impala、Solr、Spark 和键/值 Store Indexer
- ☐ **自定义服务**
选择您自己的服务。将自动包含所选服务需要的服务。只有在设置了初始群集之后才能添加 Flume。

本向导还将安装 **Cloudera Management Service**。有一系列组件可启用监控、报告、事件和警报；这些组件需要数据库存储信息，这将在下一页面上配置。

[返回](#)

1 2 3 4 5 6

▶ 继续

群集設置

自定义角色分配

您可在此处自定义新群集的角色分配，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。除非您有特殊需求，如已为特定角色预先选择特定主机，否则 **Cloudera** 不建议改变分配情况。

还可以按主机查看角色分配。

[按主机查看](#) **HDFS**

NN NameNode × 1 新建

与 `DataRow` 相同

SNN SecondaryNameNode

与 `DataRow` 相同

B Balancer x 1 新建

与 DataNode 相同

HFS HttpFS

选择主机

NFS Gateway

选择主机

DN DataNode × 1 新建

cdh ▼

Hive

G Gateway × 1 新建

与 `DataNode` 相同

HMS Hive Metastore Server

与 DataNode 相同

WHC WebHCat Server

选择主机

HS2 HiveServer2 x 1 新建

与 DataNode 相同

群集设置

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: *

192.168.4.126

数据库类型:

MySQL

数据库名称: *

hive56

用户名: *

hive56

密码:

123456

Successful

Reports Manager

当前被分配在 **cdh** 上运行。

数据库主机名称: *

192.168.4.126

数据库类型:

MySQL

数据库名称: *

report56

用户名: *

report56

密码:

123456

Successful

Oozie Server

当前被分配在 **cdh** 上运行。

数据库主机名称: *

192.168.4.126

数据库类型:

MySQL

数据库名称: *

oozie56

用户名: *

oozie56

密码:

123456

Successful

☒ 显示密码

测试连接

备注:

- 创建数据库时，**数据库主机名称** 字段中的值必须与您用于主机名称的值匹配。[了解更多](#)
- 如数据库未在其默认端口运行，请使用 **数据库主机名称** 字段中的 **host:port** 指定端口号。
- 强烈建议将各个数据库与相应角色实例置于同一主机上。

群集设置

审核更改

HDFS 块大小

dfs.block.size, dfs.blocksize

Cluster 1 > HDFS (服务范围)

128

兆字节

?

接受的 DataNode 失败的卷

dfs.datanode.failed.volumes.tolerated

Cluster 1 > DataNode Default Group

0

?

DataNode 数据目录

dfs.data.dir, dfs.datanode.data.dir

Cluster 1 > DataNode Default Group

/dfs/dn

+ -

?

NameNode 数据目录

dfs.name.dir, dfs.namenode.name.dir

Cluster 1 > NameNode Default Group

/dfs/nn

+ -

?

HDFS 检查点目录

fs.checkpoint.dir,

Cluster 1 > SecondaryNameNode Default Group

?

返回

123456

继续

群集设置

首次运行 命令

状态: Running 开始时间: 3月 16, 11:48:07 中午 中止

详细信息 Completed 0 of 7 step(s). ☒ 全部 ☐ 只会失败 ☐ Running Only

Step	上下文	开始时间
> 部署客户端配置	Cluster 1	3月 16, 11:48:07 中午
> Start Cloudera Management Service, ZooKeeper		
> 正在启动 HDFS 服务		
> 正在启动 YARN (MR2 Included) 服务		
> 正在启动 Hive 服务		
> 正在启动 Oozie 服务		
> 正在启动 Hue 服务		

返回

123456

继续

首次运行 命令

状态: Running 开始时间: 3月 16, 11:02:50 上午 中止

详细信息 Completed 1 of 7 step(s). ☒ 全部 ☐ 只会失败 ☐ Running Only

Step	上下文	开始时间
> 部署客户端配置 Successfully deployed all client configurations.	Cluster 2	3月 16, 11:02:51 上午
> Start Cloudera Management Service, ZooKeeper 仅完成 0/2 个步骤。正在等待其他步骤完成。		3月 16, 11:03:06 上午
> 正在启动 HDFS 服务		
> 正在启动 YARN (MR2 Included) 服务		
> 正在启动 Hive 服务		
> 正在启动 Oozie 服务		

返回

123456

继续

安装完成后进入如下界面

主页

30 分钟 在 2016年3月16日, 下午2点25 CST 之前

状态 所有运行状况问题 **1** 配置 **5** ▾ 所有最新命令

添加群集

Cluster 1 (CDH 5.6.0, Parcel) ▾

● 主机		✖ 1	
● HDFS	! 1	✖ 2	▾
● Hive			▾
● Oozie			▾
● YARN (MR2 Incl...			▾
● ZooKeeper		✖ 1	▾

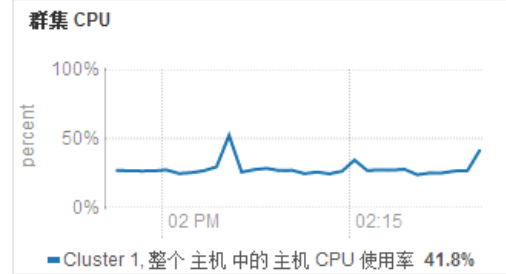
Cloudera Management Service

● Clouder...	🔌	▾
--------------	---	---

30 分钟 1 小时 2 小时 6 小时 12 小时 1 天 7d 30d



图表



群集磁盘 IO

集群安装完成。