



链滴

## 【CDH6】安装 Hive

作者: [k8s](#)

原文链接: <https://ld246.com/article/1608472429325>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

# 安装 Hive

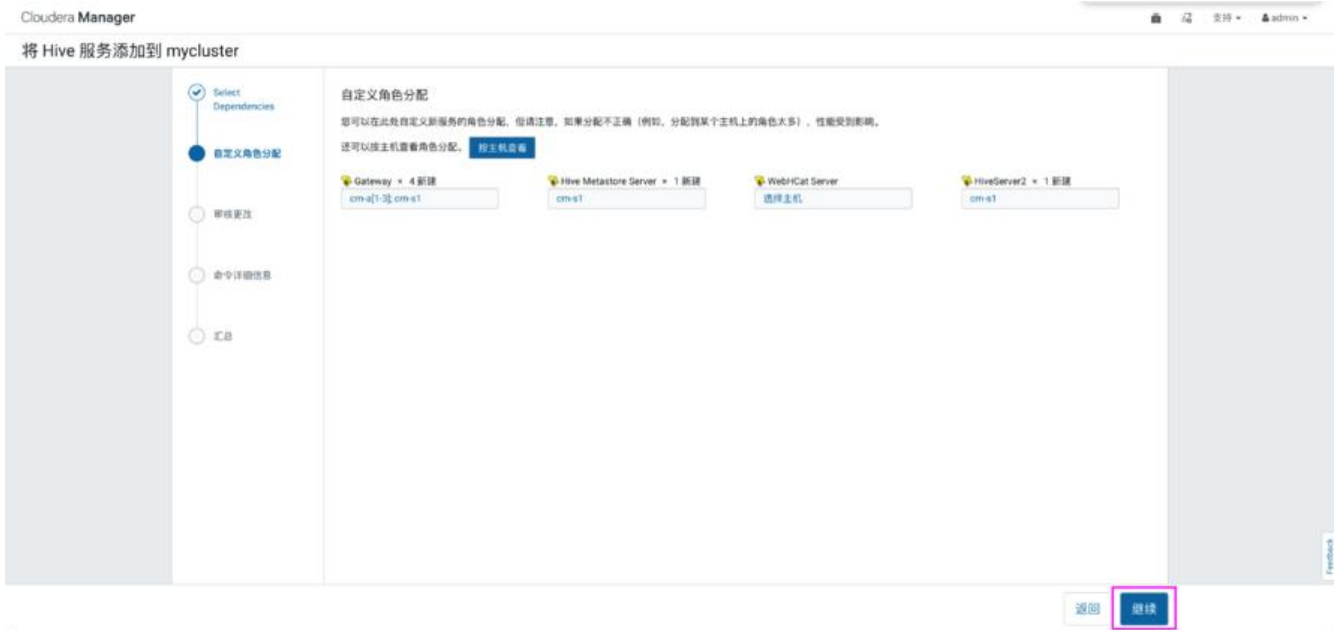
## 1. 选择集群，添加 Hive 服务

The screenshot shows the Cloudera Manager interface. At the top, there are navigation tabs: '主页', '状态', '所有运行状况问题', '配置', and '所有最新命令'. Below this, the cluster 'mycluster' is selected. A dropdown menu is open, showing options like '添加服务', 'Add Hosts', '启动', '停止', '重启', etc. The '添加服务' option is highlighted. Below the menu, a table of available services is shown. The 'Hive' service is selected, and its description is highlighted in a pink box.

服务类型	描述
CDH 6.3.2 Connector	The CDH 6.3.2 Connector service provides key management for accessing AWS S3 (S3) accounts and AWS S3 bucket credentials from CDH services.
Basehadoop	The Apache Hadoop Distributed File System (HDFS) is a robust, scalable, high-performance data storage and retrieval system. This service only works with releases meant to run on top of HDFS.
Flume	Flume 是一个用于收集、移动和写入日志数据的分布式系统 (Distributed System)。
Hive	Apache Hive 是用于在 Hadoop 分布式文件系统 (HDFS) 上执行 SQL 查询的数据库系统。它支持多种数据类型和复杂的查询操作。
Hive Metastore	Hive 是 Apache Hadoop 的元数据存储层，用于存储 Hive 表的元数据。
Ingestor	Ingestor 是用于从 HDFS 中提取数据并将其写入到其他存储系统的工具。它支持多种数据类型和复杂的查询操作。
Kerberos	Kerberos 是一种网络认证协议，用于在分布式系统中提供安全的身份验证。
MapReduce	MapReduce 是一种分布式计算模型，用于在大规模数据集上进行并行计算。
Tez	Tez 是一种分布式计算模型，用于在大规模数据集上进行并行计算。
Yarn	Yarn 是 Hadoop 的分布式资源管理系统，用于在大规模数据集上进行并行计算。
ZooKeeper	ZooKeeper 是一种分布式协调服务，用于在分布式系统中提供可靠的配置管理。

## 2. 添加服务向导

选择依赖（只有一项可供选择时则默认跳过），点击“继续”，选择默认角色配置即可：



点击“继续”之后，需要配置Hive依赖的mysql数据库，需要在cm-s1节点上连接mysql，执行创建数据库及分配权限语句：

```
[root@cm-s1 ~]# mysql -hcm-s1 -pAz123456_ -e "create database hive DEFAULT CHARACTER SET utf8;grant all on hive.* TO 'hive'@'%' IDENTIFIED BY 'Az123456_';flush privileges;"
```

在弹出的页面中选择数据库，填写用户名及密码，点击“测试连接”，测试数据库连接成功后，点击“继续”：



点击继续之后，设置hive的数据目录及端口，默认即可，点击继续之后等待Hive安装完成即可。

### 将 Hive 服务添加到 mycluster

- Select Dependencies
- 自定义角色分配
- 数据库设置
- 审核更改
- 命令详细消息
- 汇总

#### 审核更改

**Hive 仓库目录**  
Hive.metastore.warehouse.dir

Hive (服务范围)

**Hive Metastore 服务器端口**  
Hive.metastore.port

Hive Metastore Server Default Group

### 将 Hive 服务添加到 mycluster

- Select Dependencies
- 自定义角色分配
- 数据库设置
- 审核更改
- 命令详细消息
- 汇总

#### 首次运行 命令

状态 ● 已完成 12月 18, 3:26:35 下午 101.13s

Finished First Run of the following services successfully: Hive.

✓ 已完成 1 个步骤 (共 1 个)。

Show All Steps  Show Only Failed Steps  Show Only Running Steps

✓ Run a set of services for the first time 在服务器 Hive 上成功执行命令 Start	12月 18, 3:26:35 下午 101.13s
<ul style="list-style-type: none"> <li>✓ 依次运行 5 步骤 在服务器 Hive 上成功执行命令 Start</li> <li>▶ Ensuring that the expected software releases are installed on hosts. 12月 18, 3:26:35 下午 5.90s</li> <li>▶ 正在部署客户端配置 <a href="#">mycluster</a> 12月 18, 3:26:40 下午 15.35s</li> <li>▶ 并行运行 2 步骤 12月 18, 3:26:55 下午 30.89s</li> <li>▶ 并行运行 2 步骤 12月 18, 3:27:26 下午 26.05s</li> <li>▶ 启动 Hive <a href="#">Hive</a> 12月 18, 3:27:52 下午 23.68s</li> </ul>	