

Hive数据仓库使用指南

1 商品说明

Hive 是基于 Hadoop 构建的分布式数据仓库工具，可将存储在 HDFS 的结构化数据映射为数据库表，并提供类 SQL（HiveQL）查询功能，底层自动转换为 MapReduce/Spark 等分布式任务执行

本商品通过鲲鹏服务器+Euler2.0 部署。

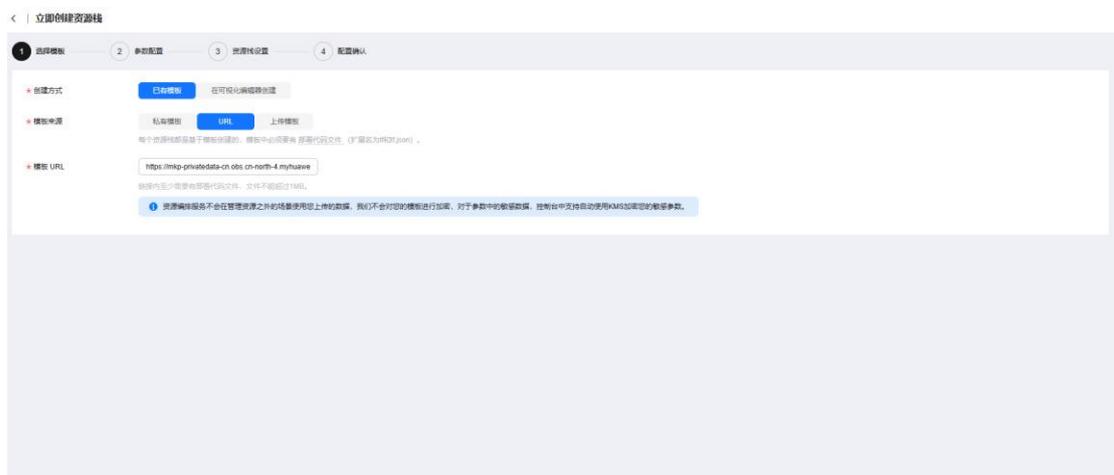
2 商品购买

您可以在云商店搜索“Hive数据仓库”。

其中，地域、规格、推荐配置使用默认，购买方式根据您的需求选择按需/按月/按年，短期使用推荐按需，长期使用推荐按月/按年，确认配置后点击“立即购买”。

2.1 商品支持自定义 ECS 购买，具体见章节 3.1.1

2.2 使用 RFS 模板直接部署



必填项填写后，点击 下一步

基本信息 资源 输出 事件 模板 执行计划

执行计划名称ID	状态	操作	创建时间	描述	操作
executionPlan_20250324_1957_4e9a	创建成功, 待部署	查看部署详情	2025/03/24 10:58:08 GMT+08:00	-	部署

如下图“Apply required resource success.”即为资源创建完成

基本信息 资源 输出 事件 模板 执行计划

事件时间	事件类型	事件描述	资源名称/类型	关联资源ID
2025/03/24 11:00:06 GMT+08:00	日志	Apply required resource success	-	-
2025/03/24 11:00:01 GMT+08:00	生成完成	module.ecs.huaweicloud_compute_instance.ecs[0]: Creation complete after 57s [id=aa08d928-dc2e-4095-b388-8b333ec8b44]	ecs ECS	aa08d928-dc2e-4095-b388-8b333ec8b44
2025/03/24 11:00:01 GMT+08:00	完成概要	Apply complete! Resources: 0 added, 0 changed, 0 destroyed.	-	-
2025/03/24 10:59:54 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [50s elapsed]	ecs ECS	-
2025/03/24 10:59:44 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [40s elapsed]	ecs ECS	-
2025/03/24 10:59:34 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [30s elapsed]	ecs ECS	-
2025/03/24 10:59:24 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [20s elapsed]	ecs ECS	-
2025/03/24 10:59:14 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [10s elapsed]	ecs ECS	-
2025/03/24 10:59:04 GMT+08:00	生成完成	module.vpc.huaweicloud_vpc_subnet.subnet: Creation complete after 9s [id=fb3ec9ef-a475-4a7d-9e7c-00000316c3763]	subnet Subnet	4b3aceef-a475-4a7d-9e7c-00000316c3763
2025/03/24 10:59:04 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Creating...	ecs ECS	-

3 商品资源配置

商品支持ECS控制台配置，下面对资源配置的方式进行介绍。

3.1 ECS 控制台配置

3.1.1 准备工作

在使用ECS控制台配置前，需要您提前配置好安全组规则。

安全组规则的配置如下：

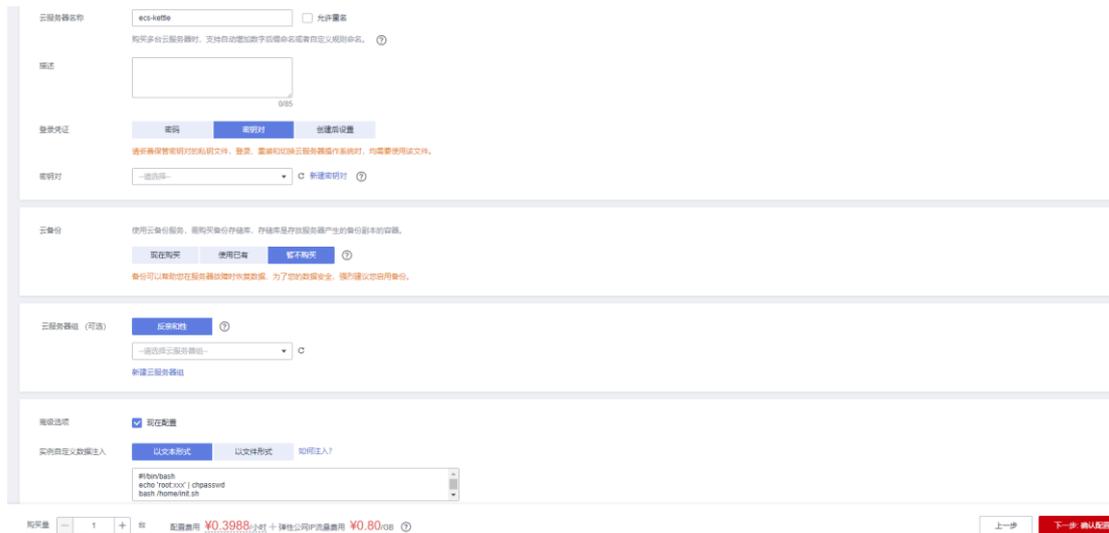
- 入方向规则放通CloudShell连接实例使用的端口22，以便在控制台登录调试。
- 出方向规则一键放通

3.1.2 创建 ECS

前提工作准备好后，选择ECS控制台配置跳转到购买ECS页面，ECS资源的配置如下图所示：

The screenshot displays the ECS console configuration interface. Key sections include:

- Basic Information:** Region (北京-北四区), Billing Mode (按量计费), Availability Zone (可用区1).
- Instance Type Selection:** CPU Architecture (x86_64), Instance Type (s6.large.2), vCPUs (2), Memory (4GB). A table lists instance types with columns: 实例类型, 规格名称, vCPU核数, 内存, CPU, 基准/最大带宽, 内网带宽, 规格参考价.
- Network Configuration:** VPC (vpc_kettle_app), Subnet (subnet_kettle_app), and IP address.
- Security Group:** A security group named 'secgroup_kettle_app' is selected, showing a table of rules with columns: 安全组名称, 优先级, 策略, 协议端口, 类型, 源地址, 描述.
- Public IP:** '现在购买' (Buy Now) is selected.
- Billing:** Billing mode (按量计费) and bandwidth (5 Mbps) are configured. The total price is 0.3988 RMB per hour.



值得注意的是：

- VPC您可以自行创建
- 安全组选择3.1.1章节中配置的安全组
- 弹性公网IP选择现在购买，推荐选择“按流量计费”，带宽大小可设置为5Mbit/s
- 其余默认或按规则填写即可。

4 商品使用

4.1 Hive 使用

启动 **Hive CLI** ，创建测试表

hive命令直接进入命令行，无需启动额外服务

```
【CREATE TABLE demo (id INT, name STRING) ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';】
```

```
[root@hive ~]# hive
which: no hbase in (/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/opt/hadoop/bin:/opt/hadoop/sbin:/opt/hive/bin)
Hive Session ID = e8cfd200-6f90-4226-afe9-4e3b4ae9b712

Logging initialized using configuration in jar:file:/opt/hive/lib/hive-common-3.1.3.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Hive Session ID = 9b318ad07-b4e0-41e0-a1be-abb412635740
hive> CREATE TABLE demo (id INT, name STRING)
  > ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
OK
Time taken: 1.38 seconds
```

插入数据并查询：

```
【INSERT INTO demo VALUES (1, 'Hive');SELECT * FROM demo;】
```

```
hive> INSERT INTO demo VALUES (1, 'Hive');
Query ID = root_20250425171216_6d52a86d-b24a-4859-a9be-0d93a2fc8892
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (Local Hadoop)
2025-04-25 17:12:23,061 Stage-1 map = 0%, reduce = 0%
2025-04-25 17:12:25,073 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local270810207_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/demo/.hive-staging_hive_2025-04-25_17-12-16_748_4763468506853506818-1/-ext-10000
Loading data to table default.demo
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 0 HDFS Write: 150 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
col1      col2
Time taken: 9.629 seconds
hive> SELECT * FROM demo;
OK
demo.id demo.name
1       Hive
Time taken: 0.188 seconds, Fetched: 1 row(s)
```

4.2 参考文档

- [hive官网](#)