Presto使用指南

# 购买ECS服务器

* 基础配置

根据自身需要选择合适的区域，规格等，在此镜像已经默认选择为该镜像，而后在网络配置中选择网络功能。



* 高级配置（重要）



**#!/bin/bash**

**echo 'root:${password}' | chpasswd**

**bash /home/init.sh**

**${password}需要替换成您希望设置的密码。**

而后即可确定购买！完成后可以使用XShell或其他登录方式完成登录并前往指定目录修改配置信息。

# 配置并启动服务

ECS启动成功后，使用root账号在ECS控制台或者远程连接工具进入终端，进入成功后，需要进行以下配置。

## 配置并启动Hive

1. 配置MySql数据库信息和OBS访问信息：

cd /opt/software/hive-3.1.3/conf

vim hive-site.xml

# 需要配置MySQL数据库信息，否则hive metastore无法启动，该ECS已安装MySQL，您可以使用已有的MySQL数据库信息，也可以在该ECS上搭建新的MySQL数据库。

# 需要配置OBS信息，否则hive无法访问OBS，OBS区域对应的Endpoint可以在华为云官网：[https://developer.huaweicloud.com/endpoint?OBS](https://console.huaweicloud.com/apiexplorer/%22%20%5Cl%20%22/endpoint) 中查询。



1. 登录mysql，创建metastore数据库：

create database metastore;

1. 初始化元数据库

cd /opt/software/hive-3.1.3

bin/schematool -dbType mysql -initSchema

1. 启动Hive：

hive --service metastore

## 配置并启动Presto

1. 编辑hiveobs.properties文件：

cd /opt/software/presto-server-0.283/etc/catalog

vim hiveobs.properties

1. hiveobs.properties文件添加如下内容：

############ 配置 connector ############

connector.name=hive-obs

hive.metastore.uri=thrift://127.0.0.1:9083

# 是否授予删表的权限

hive.allow-drop-table=true

# 是否开启向Hive的外部表写入数据

hive.non-managed-table-writes-enabled=true

############ Metastore配置属性 ############

# Hive元存储请求超时,默认10s，单位：秒

hive.metastore-timeout=10s

# 缓存的Metastore数据应被视为有效的持续时间。单位：秒

hive.metastore-cache-ttl=0s

# HiveMetastore缓存最大大小。

hive.metastore-cache-maximum-size=10000

# 访问后异步刷新缓存的Metastore数据，如果它比这更旧但尚未过期，则允许后续访问查看新数据。单位：秒

hive.metastore-refresh-interval=0s

# 用于刷新缓存的元存储数据的最大线程数。

hive.metastore-refresh-max-threads=100

############ OBS配置 ############

# 要使用的默认OBS访问密钥。

hive.obs.access-key=华为云OBS AK

# 要使用的默认OBS密钥。

hive.obs.secret-key=华为云OBS SK

# 华为云obs endpoint，本例写的是华北-北京四区域节点，实际请修改为桶所在区域节点

hive.obs.endpoint=https://obs.cn-north-4.myhuaweicloud.com

# 临时AK/SK对应的securityToken,可不填

hive.obs.security-token=

# 写入数据时使用的OBS存储类。目前仅支持STANDARD和INTELLIGENT\_TIERING存储类。默认存储类别是STANDARD

hive.obs.storage-class=STANDARD

# 判断是否使用路径访问方式访问OBS服务，true使用路径访问方式，false使用虚拟主机访问方式，默认值：false

hive.obs.path-style-access=false

# 写入OBS的数据的本地暂存目录。这默认为JVM系统属性指定的Java临时目录java.io.tmpdir。

#hive.obs.staging-directory=

# 使用HTTPS与OBS API通信（默认为true）。

hive.obs.ssl.enabled=true

# 使用OBS服务器端加密（默认为false）。

hive.obs.sse.enabled=false

# OBS服务器端加密的密钥管理类型；开启服务端加密时 KMS/C 二选一。

#hive.obs.sse.type=KMS

# 用于使用KMS管理的密钥进行OBS服务器端加密的KMS密钥ID。如果未设置，则使用默认密钥。

hive.obs.sse.kms-key-id=按需填写

# SSE-C方式的加密密钥

hive.obs.sse.c-key-base64=按需填写

# 将文件上传到OBS时使用的预制ACL（默认为Private）。

hive.obs.upload-acl-type=private

# 是否启用流式传输模式。默认为true。

hive.obs.streaming.enabled=true

# 设置流式传输的分段大小。默认16MB。

hive.obs.streaming.part-size=16MB

# 忽略Glacier对象而不是使查询失败。这将跳过可能是表或分区的一部分的数据。默认为false.

#hive.obs.skip-glacier-objects=

# 与OBS通信时遇到错误时尝试重新操作（如读取、写入等）的最大次数；默认5次。

hive.obs.max-error-retries=5

# 重试的最大读取尝试次数。

hive.obs.max-client-retries=5

# 与OBS通信时，连续的重试之间允许的最大等待时间

hive.obs.max-backoff-time=10m

# 重试与OBS通信的最长时间

hive.obs.max-retry-time=10m

# TCP连接超时

hive.obs.connect-timeout=5s

# TCP套接字读取超时

hive.obs.socket-timeout=5s

# 与OBS的最大连接数

hive.obs.max-connections=500

### 启动Presto

启动presto前需启动hive metastore。

1. 启动Presto

cd /opt/software/presto-server-0.283/bin

./launcher start

1. 可查看启动日志，日志中看到：SERVER STARTED，即表示presto启动成功。

cd /opt/software/presto-server-0.283/data/var/log

tail -f server.log



## 使用Presto

### 启动客户端

cd /opt/software

./presto-cli.jar --server 127.0.0.1:7670 --catalog hiveobs --schema default

更多Presto的配置和使用可以参考官方文档

<https://gitee.com/HuaweiCloudDeveloper/huaweicloud-presto-plugins>