

数据仓库服务

快速入门

文档版本 19
发布日期 2024-11-01



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目录

1 免费体验 GaussDB(DWS).....	1
2 快速创建 GaussDB(DWS)集群并导入数据进行查询.....	3
3 GaussDB(DWS)数据开发 SQL 入门操作.....	16
4 GaussDB(DWS)入门实践.....	18

1 免费体验 GaussDB(DWS)

数据分析实验室是华为云官方实验平台，实验环境为华为云现网环境，整个实验过程完全免费。让您在最短时间内体验GaussDB(DWS)真实环境，快速上手。

开发者可通过[表1-1](#)实验手册指导，使用环境中预置的华为云账号，一键创建GaussDB(DWS)实验环境，在云端体验GaussDB(DWS)的数据导入、访问MRS集群、多维度分析、权限管理、JDBC开发、性能调优等场景。

表 1-1 实验列表

场景	实验名称	实验描述	预计时长
导入分析	快速入门	一键式创建GaussDB(DWS)集群，上传csv本地数据到云存储OBS桶，通过创建OBS外表实现数据导入并简单分析。	1h
	零售业经营多维度分析	环境已预置样例数据在OBS桶，通过创建OBS外表导入样例数据，再使用聚合函数、group by、order by、视图进行多维度分析。	1.5h
SQL on Hadoop	导入MRS-Hive数据源	创建数据仓库集群GaussDB(DWS)，并导入MRS的Hive数据，实现跨集群进行大数据融合分析。	1.5h
二次开发	使用Java进行二次开发	使用JDBC驱动连接GaussDB(DWS)进行二次开发，熟悉简单的适配DWS的Java开发用例。	1.5h
数据迁移	从老DWS集群迁移数据到新DWS集群	本实验指导用户创建数据仓库集群GaussDB(DWS)并将老GaussDB(DWS)整库迁移到新的GaussDB(DWS)。同时，针对使用自增序列导致的性能问题场景，指导用户排查原因并提供优化方法。	2~3h
	基于gds实现跨集群数据互联互通	本实验通过部署GDS服务器，使用GDS导入导出的并发能力，实现双DWS集群之间1500万行数据分钟级迁移。	2h

场景	实验名称	实验描述	预计时长
安全管理	权限管理	通过实验创建不同用户，不同Schema，基于权限管理实现数据的隔离和互访，了解用户、角色的关系，了解grant的基本用法，了解基于角色的权限管理(RBAC)。	1.5h
	使用数据脱敏实现卡号等隐私信息屏蔽	本实验通过创建数据仓库服务GaussDB(DWS)并使用DWS的数据脱敏功能，针对不同用户设置部分数据列的屏蔽，实现敏感数据脱敏，确保数据安全。	1h
高级特性	冷热数据管理	指导用户创建数据仓库集群GaussDB(DWS)，并创建冷热分区表实现冷热数据分区管理，不仅可以提高数据分析性能还能降低业务成本。	1h
调优	性能调优	指导用户使用GaussDB(DWS)进行性能调优。通过本实验掌握通过EXPLAIN语句查询执行计划的方法，了解GaussDB(DWS)几种常见的SQL调优手段。	2h
云原生3.0	云原生3.0数仓-存算分离	本实验指导用户创建GaussDB(DWS)新一代Serverless云原生数仓，并体验Serverless存算分离架构下的极致查询。	2h
	云原生3.0数仓-湖仓一体	本实验指导用户创建GaussDB(DWS)新一代Serverless云原生数仓，通过EXTERNAL SCHEMA访问MRS的Hive数据，体验湖仓一体、存算分离等极致查询的高级特性。	2h

2 快速创建 GaussDB(DWS)集群并导入数据进行查询

操作场景

数据仓库服务GaussDB(DWS)是完全托管的企业级云上数据仓库服务，具备免运维、在线扩展、高效的多源数据加载能力，兼容PostgreSQL生态。助力企业经济高效地对海量数据进行在线分析，实现数据快速变现。

本文以创建3节点的GaussDB(DWS)集群为例，通过导入OBS桶上csv格式的数据，完成简单的数据查询。

须知

GaussDB(DWS)集群不支持跨区域访问OBS桶数据，创GaussDB(DWS)集群前需规划好GaussDB(DWS)集群与OBS桶在同一个区域内，本文以“华北-北京四”为例。

- 计费方式：按需
- 数据仓库类型：标准数仓
- 集群节点：3
- 连接集群方式：SQL编辑器
- 存储容量：20G/节点
- 数据源：OBS桶的csv数据
- 集群版本：8.1.3.x

操作流程

表 2-1 操作流程

操作步骤	说明
准备工作	注册华为账号并开通华为云，实名认证，为账户充值以及规划和创建VPC、子网资源。

操作步骤	说明
步骤一：创建 GaussDB(DWS)集群	创建3节点的GaussDB(DWS)集群，配置集群节点、存储容量、集群版本号等信息。
步骤二：上传数据到 OBS桶	创建OBS桶，并将本地的csv格式数据上传到OBS桶。
步骤三：连接 GaussDB(DWS)集群并导入数据进行分析	使用SQL编辑器连接GaussDB(DWS)数据库，并创建OBS外表，通过OBS外表将OBS桶数据导入GaussDB(DWS)集群，并进行简单SQL查询。

准备工作

- 已注册账号并实名认证，且在使用GaussDB(DWS) 前检查账号状态，账号不能处于欠费或冻结状态。
- 获取此账号的“AK/SK”，参见[访问密钥](#)。
- 已创建VPC和子网。

步骤一：创建 GaussDB(DWS)集群

步骤1 登录管理控制台，进入[购买GaussDB\(DWS\)界面](#)。

步骤2 参见[表2-2](#)进行基础配置。

表 2-2 基础配置

参数	示例	说明
区域	选择“华北-北京四”。 说明 本指导以“华北-北京四”为例进行介绍，如果您需要选择其他区域，请确保其他云服务均在同一区域。	请就近选择靠近您业务的区域，可以降低网络时延、提高访问速度。GaussDB(DWS)创建后无法更换区域，请谨慎选择。 更多信息，请参见 区域和可用分区 。
计费模式	按需计费。	支持包年/包月计费、按需计费。如长期使用，建议使用包年/包月，可享受更大折扣。 更多信息，请参见 计费说明 。
可用分区	单AZ，可用区1 （如规格售罄，可选择其他可用区）。	<ul style="list-style-type: none"> ● 单AZ：集群的所有节点部署在同一个AZ内。 ● 多AZ：集群的节点部署在多个AZ内，满足集群跨AZ高可用要求，对高可用要求更高的业务，建议使用多AZ。

图 2-1 基础配置



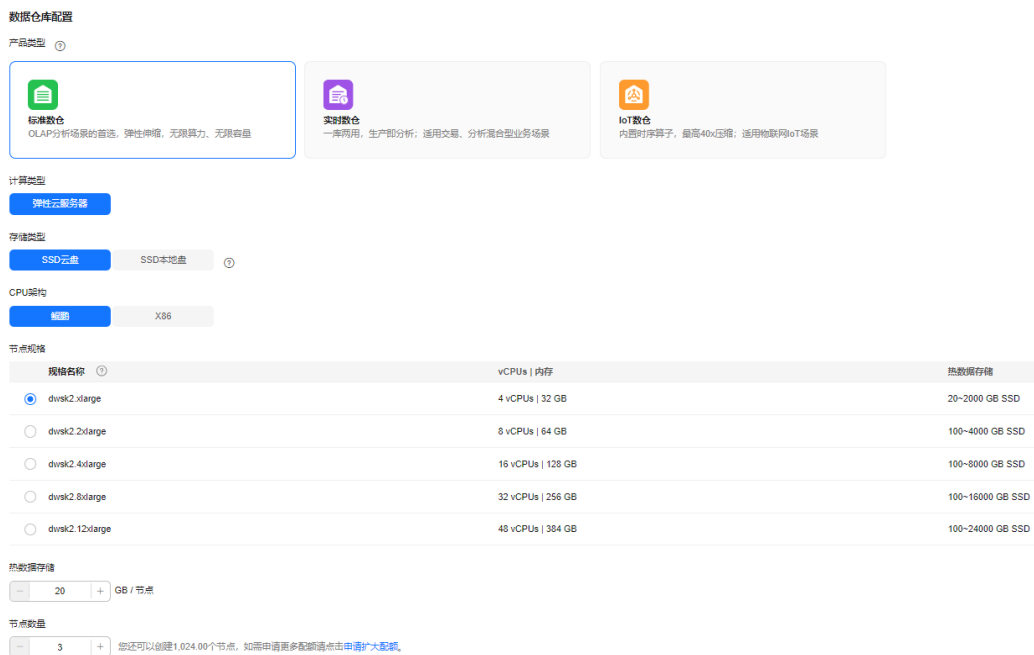
步骤3 参见表2-3，进行数据仓库配置。

表 2-3 数据仓库配置

参数	示例	说明
产品类型	标准数仓	<ul style="list-style-type: none"> 标准数仓：高性价比，支持冷热数据分析，存储、计算弹性伸缩，并按需、按量计价。 实时数仓：在大规模数据查询和分析能力基础上，提供高并发、高性能、低时延、低成本的事务处理能力。适用于HTAP混合负载场景。 更多信息，请参见 数据仓库类型 。
计算类型	弹性云服务器	计算类型包括弹性云服务器和裸金属服务器。当前仅支持弹性云服务器。
存储类型	SSD云盘	<ul style="list-style-type: none"> SSD云盘：使用SSD类型的EVS作为数据存储介质，存储容量更加灵活，支持磁盘扩容。 SSD本地盘：使用ECS规格自带的本地磁盘作为数据存储介质，容量固定，性能更高，不支持磁盘扩容。 更多信息，请参见 SSD云盘和本地盘的区别 。
CPU架构	鲲鹏	集群节点的CPU架构。
节点规格	dwsk2.xlarge 说明 如规格售罄，可选择其他可用区或规格。	节点规格是衡量GaussDB(DWS)算力的一个指标，包含vCPUs和内存，请根据业务需求合理选择，通常对查询性能要求越高，选择的规格越大。更多信息，请参见 数据仓库规格 。

参数	示例	说明
热数据存储	20GB / 节点	每节点可用存储容量。界面上显示的是业务数据实际存储容量，例如20GB/节点，节点总数3，则业务数据总容量为60GB。对于业务数据量较大的客户，请合理规划存储容量。
节点数量	3	集群节点的数量。支持的节点数范围为3~256。

图 2-2 数据仓库配置



步骤4 单击“下一步：网络配置”。

步骤5 参见表2-4进行网络配置。

表 2-4 网络配置

参数	示例	说明
虚拟私有云	vpc-default	GaussDB(DWS)所在的虚拟私有云网络。
子网	subnet-default(192.168.0.0/24)	虚拟私有云的子网。

参数	示例	说明
安全组	自动创建安全组	支持“自动创建安全组”或选择已创建的安全组。 系统默认选中“自动创建安全组”，设置为该选项时，系统将会自动创建一个默认安全组。 默认安全组的规则如下：出方向允许所有访问，入方向只开放用户设置的GaussDB(DWS)集群的数据库端口8000。
公网访问	暂不使用	如果连接GaussDB(DWS)的客户端跟GaussDB(DWS)不在一个虚拟私有云网络内，需要选择使用弹性公网IP。
弹性负载均衡	暂不使用	用户使用客户端连接DWS集群时，如果用户仅连接一个CN节点地址，通过该CN节点内网IP或弹性公网IP连接时，只能连接到固定的CN节点上，存在CN单点问题。 利用ELB健康检查机制可将集群的CN请求快速转发到正常工作的CN节点上，当有CN故障节点时，也能第一时间切换流量到健康节点上，最大化降低集群访问故障。 更多信息，请参见 集群绑定和解绑ELB 。

图 2-3 网络配置

网络

虚拟私有云

vpc-2767
▼
🔍 查看虚拟私有云 ?

子网

subnet-278a(192.168.0.0/24)
▼
🔍 ?

安全组

自动创建安全组
▼
🔍 ?

自动在入规则中打开集群端口的访问权限。

其他

公网访问

暂不使用

现在购买

使用已有

?

不使用弹性IP的集群不能与互联网互通，仅可通过私有网络中已部署的弹性云服务器连接当前集群使用。

弹性负载均衡

暂不使用

使用已有

?

弹性负载均衡是将访问流量根据分配策略分发到后端多台服务器的流量分发控制服务。不使用弹性负载均衡的集群没有流量分发能力，集群存在单点故障的风险。 [了解更多](#)

步骤6 单击“下一步：高级配置”，参见表2-5进行高级配置。

表 2-5 高级配置

参数	示例	说明
集群名称	dws-demo	集群名称，用户自定义，支持后续修改。
集群版本	使用推荐版本	<p>每个版本支持的数据库功能不同，默认选择推荐版本。如需使用高版本功能，请选择更高版本。</p> <p>说明 白名单客户界面，会显示其他集群版本，请根据界面实际情况选择。</p> <p>集群各版本支持功能，请参见产品公告。</p>

参数	示例	说明
管理员用户	dbadmin	指定数据库系统管理员的用户名，默认即可。 数据库系统管理员默认具有SYSADMIN属性，非三权分立模式下，拥有数据库最高权限。可用于规划数据库对象，例如创建数据库、用户、授权用户角色等操作。 更多信息，请参见 数据库用户 。
管理员密码	-	数据库系统管理员密码。
确认密码	-	数据库系统管理员密码。
数据库端口	8000	客户端连接GaussDB(DWS)数据库的端口，默认8000即可。
企业项目	default	仅当使用企业类型的账号创建GaussDB(DWS)时，会显示该参数。 用于按项目统一管理云资源。
高级配置	默认配置	高级配置支持“默认配置”和“自定义”。当选择“自定义”时，支持以下高级配置： <ul style="list-style-type: none"> • 备份设备：OBS或者NFS。 • CN部署量：在大规模集群下，建议部署多个CN。 • 标签：在GaussDB(DWS)中，当用户创建集群后，如果想对集群添加如项目名称、业务类别、背景信息等相关信息的标识，用户可以通过添加标签来实现。如果用户在其他云服务中使用了标签，那么建议用户为同一个业务所使用的云资源创建相同的标签键值对以保持一致性。 • 加密数据库。开启后，GaussDB(DWS)将使用KMS（密钥管理服务）为集群及其快照的数据启用加密功能。

图 2-4 高级配置

The image shows a configuration interface for GaussDB(DWS) clusters. It is divided into two main sections: 'Management' and 'Advanced Configuration'.
In the 'Management' section:
- 'Cluster Name': A text input field containing 'dws-demo'.
- 'Cluster Version': Two buttons for version selection, '8.1.3.330' (highlighted in blue) and '8.3.0.108'.
- 'Default Database': A dropdown menu set to 'gaussdb'.
- 'Admin User': A text input field containing 'dbadmin'.
- 'Admin Password': Two password input fields, one for the password and one for confirmation.
- 'Database Port': A text input field containing '8000'.
- 'Cluster Timezone': A dropdown menu set to 'UTC+08:00 (北京, 重庆, 香港, 乌鲁木齐...)'.
In the 'Advanced Configuration' section:
- 'Enterprise Project': A dropdown menu set to 'default'.
- 'Advanced Configuration': Two buttons, 'Default Configuration' (highlighted in blue) and 'Customize'.
Below the 'Management' section, there are several lines of small text providing constraints for the cluster name, user name, and port.

步骤7 单击“下一步：确认配置”，确认无误后，单击“立即购买”。
等待约10分钟，集群创建成功。
----结束

步骤二：上传数据到 OBS 桶

步骤1 创建OBS桶，参见[OBS用户指南](#)。
关键参数按如下填写，其他参数默认即可。

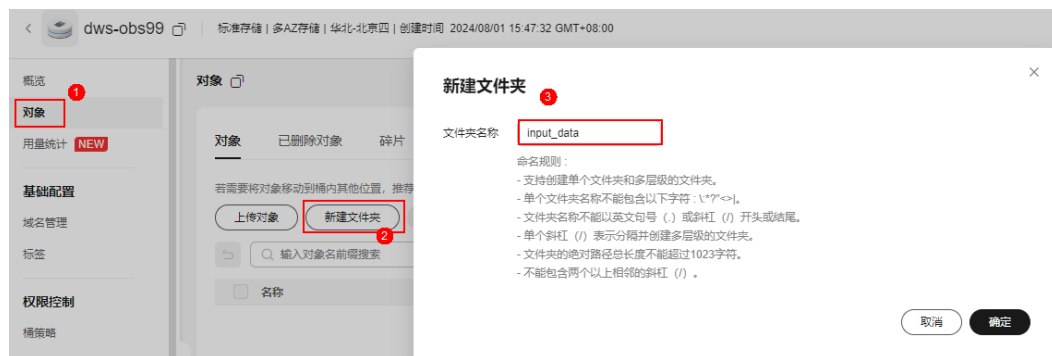
注意

不支持跨区域访问OBS桶数据，OBS桶需跟GaussDB(DWS)创在同一个区域。例如“华北-北京四”。

- **区域**：华北-北京四
- **桶名称**：dws-obs01（如提示冲突，可以依次递增到02、03）
- **桶策略**：私有

步骤2 创建OBS桶成功后，左侧选择“桶列表”，单击新创建的dws-obs01桶名，左侧导航选择“对象”，单击“新建文件夹”，文件夹名称为“input_data”，单击“确定”。

图 2-5 新建文件夹



步骤3 在本地PC创建3个csv数据文件，名字（含格式后缀）分别为**product_info0.csv**、**product_info1.csv**、**product_info2.csv**。

图 2-6 创建 csv 文件

	product_info0.csv	2024/8/1 16:05	Microsoft Excel ...	0 KB
	product_info1.csv	2024/8/1 16:05	Microsoft Excel ...	0 KB
	product_info2.csv	2024/8/1 16:05	Microsoft Excel ...	0 KB

步骤4 使用文本编辑工具分别打开3个csv文件，复制以下内容到每个csv文件中。

- product_info0.csv:**

```
100,XHDK-A,2017-09-01,A,2017 Shirt Women,red,M,328,2017-09-04,715,good!
205,KDKE-B,2017-09-01,A,2017 T-shirt Women,pink,L,584,2017-09-05,40,very good!
300,JODL-X,2017-09-01,A,2017 T-shirt men,red,XL,15,2017-09-03,502,Bad.
310,QQPX-R,2017-09-02,B,2017 jacket women,red,L,411,2017-09-05,436,It's nice.
150,ABEF-C,2017-09-03,B,2017 Jeans Women,blue,M,123,2017-09-06,120,good.
```
- product_info1.csv:**

```
200,BCQP-E,2017-09-04,B,2017 casual pants men,black,L,997,2017-09-10,301,good quality.
250,EABE-D,2017-09-10,A,2017 dress women,black,S,841,2017-09-15,299,This dress fits well.
108,CDXK-F,2017-09-11,A,2017 dress women,red,M,85,2017-09-14,22,It's really amazing to buy.
450,MMCE-H,2017-09-11,A,2017 jacket women,white,M,114,2017-09-14,22,very good.
260,OCDA-G,2017-09-12,B,2017 woolen coat women,red,L,2004,2017-09-15,826,Very comfortable.
```
- product_info2.csv:**

```
980,"ZKDS-J",2017-09-13,"B","2017 Women's Cotton Clothing","red","M",112,,
98,"FKQB-I",2017-09-15,"B","2017 new shoes men","red","M",4345,2017-09-18,5473
50,"DMQY-K",2017-09-21,"A","2017 pants men","red","37",28,2017-09-25,58,"good","good","good"
80,"GKLW-L",2017-09-22,"A","2017 Jeans Men","red","39",58,2017-09-25,72,"Very comfortable."
30,"HWEC-L",2017-09-23,"A","2017 shoes women","red","M",403,2017-09-26,607,"good!"
40,"IQPD-M",2017-09-24,"B","2017 new pants Women","red","M",35,2017-09-27,52,"very good."
50,"LPEC-N",2017-09-25,"B","2017 dress Women","red","M",29,2017-09-28,47,"not good at all."
60,"NQAB-O",2017-09-26,"B","2017 jacket women","red","S",69,2017-09-29,70,"It's beautiful."
70,"HWNB-P",2017-09-27,"B","2017 jacket women","red","L",30,2017-09-30,55,"I like it so much"
80,"JKHU-Q",2017-09-29,"C","2017 T-shirt","red","M",90,2017-10-02,82,"very good."
```

步骤5 使用文本编辑工具转为UTF-8编码，并保存。

步骤6 回到浏览器上的OBS控制台，左侧导航选择“桶列表”，单击新创建的dws-obs01桶名。

步骤7 左侧导航选择“对象”，单击“input_data”文件夹。

步骤8 单击“上传对象”，单击“添加文件”。

步骤9 将product_info0.csv、product_info1.csv、product_info2.csv 3个文件上传到OBS桶。

----结束

步骤三：连接 GaussDB(DWS)集群并导入数据进行分析

步骤1 浏览器切换到GaussDB(DWS)管理控制台。

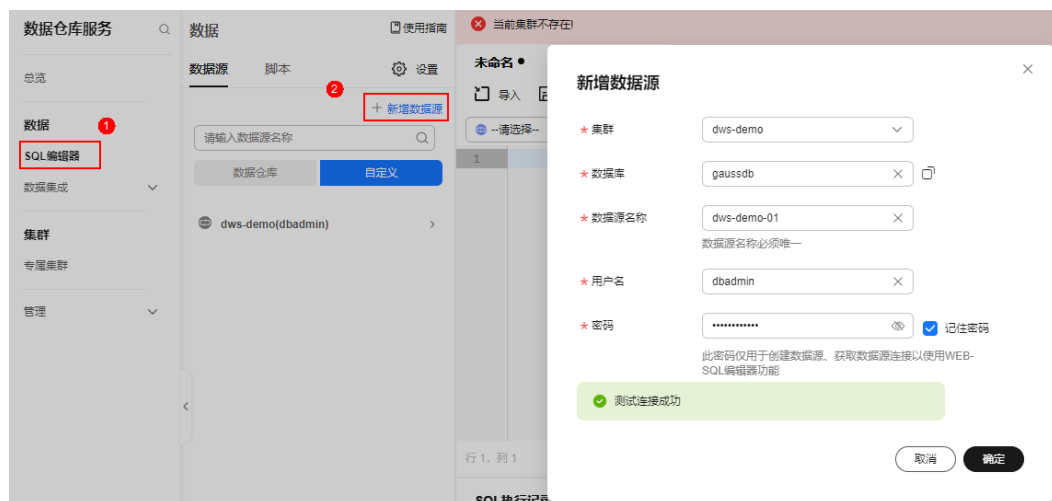
步骤2 左侧导航选择“数据 > SQL编辑器”，单击“新增数据源”。

步骤3 填写数据库连接信息。

- 集群：选择本文创建的dws-demo集群。
- 数据库：保持gaussdb默认。
- 数据源名称：dws-demo-01。
- 用户名：dbadmin
- 密码：[步骤一：创建GaussDB\(DWS\)集群](#)中创建GaussDB(DWS)时设置的密码。

步骤4 勾选“记住密码”，单击“测试连接”。

图 2-7 数据库连接信息



步骤5 单击“确定”。


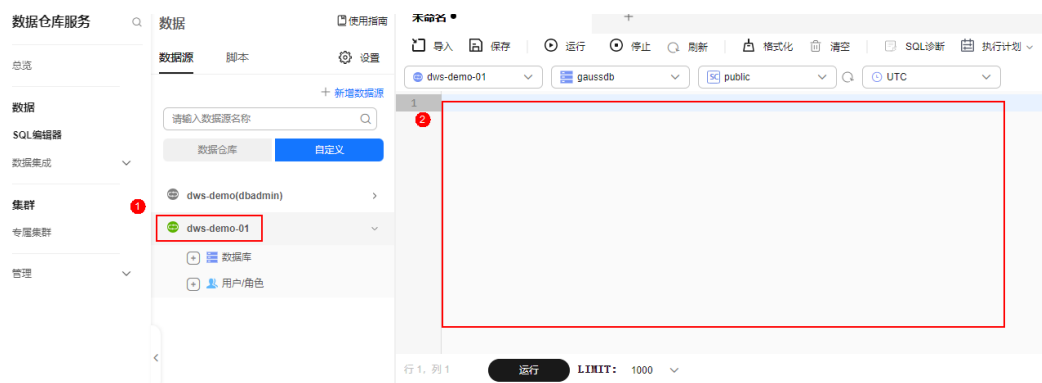

如下图，数据源名称前面图标为绿色，表示数据库连接成功。右侧区域为SQL编辑器窗口，可进行后续的SQL编写并执行SQL语句。

图 2-8 连接数据库成功



说明

后续连接会话超时，或者浏览器重新刷新时， 图标置为灰色，表示连接断开，可以直接双击数据源名称重新连接数据库。



步骤6 将以下SQL语句复制到SQL编辑器窗口中，并将OBS桶名称、AK值、SK值，替换成实际值。

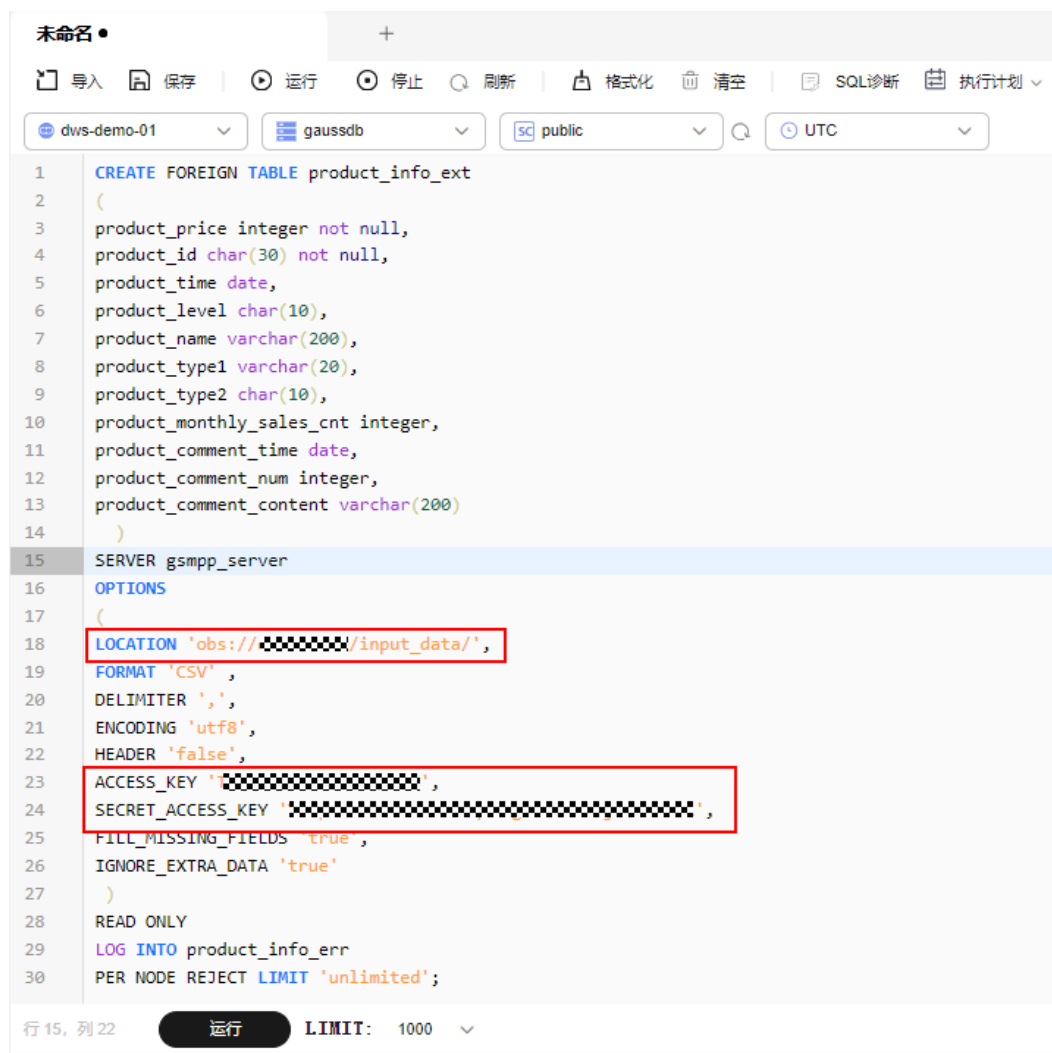
本文OBS桶名为dws-obs01，AK值和SK值从[访问密钥](#)获取。

```
CREATE FOREIGN TABLE product_info_ext
(
  product_price integer not null,
  product_id char(30) not null,
  product_time date,
  product_level char(10),
  product_name varchar(200),
  product_type1 varchar(20),
  product_type2 char(10),
  product_monthly_sales_cnt integer,
  product_comment_time date,
  product_comment_num integer,
  product_comment_content varchar(200)
)
SERVER gsmpp_server
OPTIONS
(
  LOCATION 'obs://OBS桶名称/input_data/',
  FORMAT 'CSV',
  DELIMITER ';;',
  ENCODING 'utf8',
  HEADER 'false',
```



```
ACCESS_KEY 'AK值',
SECRET_ACCESS_KEY 'SK值',
FILL_MISSING_FIELDS 'true',
IGNORE_EXTRA_DATA 'true'
)
READ ONLY
LOG INTO product_info_err
PER NODE REJECT LIMIT 'unlimited';
```

图 2-9 创建 OBS 外表



步骤7 核对SQL语句无误，单击“运行”，创建OBS外表成功。

图 2-10 创建 OBS 外表成功



步骤8 复制以下SQL语句，在SQL窗口中执行，创建普通表。

```
CREATE TABLE product_info
(
product_price integer not null,
```

```
product_id char(30) not null,
product_time date ,
product_level char(10) ,
product_name varchar(200) ,
product_type1 varchar(20) ,
product_type2 char(10) ,
product_monthly_sales_cnt integer ,
product_comment_time date ,
product_comment_num integer ,
product_comment_content varchar(200)
)
WITH
(
orientation = column,
compression=middle
)
DISTRIBUTE BY hash (product_id);
```

步骤9 复制以下SQL语句，在SQL窗口中执行，通过INSERT语句将OBS外表数据导入 GaussDB(DWS)数据库。

```
INSERT INTO product_info SELECT * FROM product_info_ext;
```

步骤10 复制以下SQL语句，在SQL窗口中执行，进行VACUUM和ANALYZE操作。

了解更多语法，请参考相关SQL语法章节。

```
VACUUM FULL product_info;
ANALYZE product_info;
```

步骤11 复制以下SQL语句，在SQL窗口中执行，查询数据明细成功。

```
SELECT * FROM product_info;
```

图 2-11 查询数据成功

product_price	product_id	product_time	product_level	product_name	product_type1	product_type2	product_monthly_sales...	product_comment_t...	product_comment_...	product_com
200	BCOP-E	2017-09-04 00:00:00	B	2017 casual pants men	black	L	997	2017-09-10 00:00:00	301	good quality.
250	EABE-D	2017-09-10 00:00:00	A	2017 dress women	black	S	841	2017-09-15 00:00:00	299	This dress fits we
108	CDXK-F	2017-09-11 00:00:00	A	2017 dress women	red	M	85	2017-09-14 00:00:00	22	It's really amazi
300	JODL-X	2017-09-01 00:00:00	A	2017 T-shirt men	red	XL	15	2017-09-03 00:00:00	502	Bad.
980	ZKDS-J	2017-09-13 00:00:00	B	2017 Women's Collo...	red	M	112	(Null)	(Null)	
50	LPEC-N	2017-09-25 00:00:00	B	2017 dress Women	red	M	29	2017-09-28 00:00:00	47	not good at all.
450	MMCE-H	2017-09-11 00:00:00	A	2017 jacket women	white	M	114	2017-09-14 00:00:00	22	very good.
205	KDKE-B	2017-09-01 00:00:00	A	2017 T-shirt Women	pink	L	584	2017-09-05 00:00:00	40	very good!
310	QOPX-R	2017-09-02 00:00:00	B	2017 jacket women	red	L	411	2017-09-05 00:00:00	436	It's nice.
150	ABEF-C	2017-09-03 00:00:00	B	2017 Jeans Women	blue	M	123	2017-09-06 00:00:00	120	good.

---结束

3 GaussDB(DWS)数据开发 SQL 入门操作

本节主要介绍GaussDB(DWS)数据库的一些SQL基本操作。

创建、查看和删除数据库

- 使用CREATE DATABASE语句创建数据库。

```
CREATE DATABASE test_db ENCODING 'UTF8' template = template0;
```
- 使用\l命令查看数据库系统的数据库列表。

```
\l
```
- 通过系统表PG_DATABASE查询数据库列表。

```
SELECT datname FROM pg_database;
```
- 使用DROP DATABASE语句删除数据库。

```
DROP DATABASE test_db;
```

创建、查看、修改和删除表

- 使用CREATE TABLE语句创建表。

```
CREATE TABLE customer_t1(id INT, name CHAR(40),age TINYINT);
```
- 使用PG_GET_TABLEDEF()函数查看建表语句。

```
SELECT * FROM PG_GET_TABLEDEF('customer_t1');
```
- 使用ALTER TABLE语句修改表。
增加列：

```
ALTER TABLE customer_t1 ADD (address VARCHAR(100));
```


删除列：

```
ALTER TABLE customer_t1 DROP COLUMN address;
```


修改字段类型：

```
ALTER TABLE customer_t1 MODIFY age INTEGER NOT NULL;
```
- 使用DROP TABLE语句删除表。

```
DROP TABLE customer_t1;
```

创建、查看和删除索引

- 使用CREATE INDEX或ALTER TABLE语句创建普通索引。

```
CREATE INDEX c_id_index on customer_t1(id);  
ALTER TABLE customer_t1 ADD INDEX c_id_index (id);
```
- 使用PG_INDEXES系统表查看表内所有索引。

```
SELECT * FROM pg_indexes WHERE tablename = 'customer_t1';
```

- 使用ALTER TABLE或DROP INDEX语句删除索引。

```
DROP INDEX c_id_index;  
ALTER TABLE customer_t1 DROP INDEX c_id_index;
```

增删改查表数据

- 使用INSERT INTO语句插入表数据。

```
INSERT INTO customer_t1 VALUES(1001,'user1',22);
```

- 使用SELECT语句查询表数据。

```
SELECT * FROM customer_t1;
```

- 使用UPDATE更新表数据。

```
UPDATE customer_t1 SET id = 1009 WHERE id = '1001';
```

- 使用DELETE删除表数据。

```
DELETE FROM customer_t1 WHERE id = '1009';
```

4 GaussDB(DWS)入门实践

当用户完成集群创建后，可以根据自身的业务需求使用GaussDB(DWS)提供的一系列常用实践。

表 4-1 常用最佳实践

实践	描述
数据导入导出	从OBS导入数据到集群 本教程旨在通过演示将样例数据上传OBS，并将OBS的数据导入进GaussDB(DWS)上的目标表中，让您快速掌握如何从OBS导入数据到GaussDB(DWS)集群的完整过程。 GaussDB(DWS)支持通过外表将OBS上TXT、CSV、ORC、PARQUET、CARBONDATA以及JSON格式的数据导入到集群进行查询。
	使用GDS从远端服务器导入数据 本教程旨在演示使用GDS (General Data Service) 工具将远端服务器上的数据导入GaussDB(DWS)中的办法，帮助您学习如何通过GDS进行数据导入的方法。 GaussDB(DWS)支持通过GDS外表将TXT、CSV和FIXED格式的数据导入到集群进行查询。

实践		描述
	<p>导入远端DWS数据源</p>	<p>大数据融合分析场景下，支持同一区域内的多套 GaussDB(DWS)集群之间的数据互通互访，本实践将演示通过 Foreign Table 方式从远端 DWS 导入数据到本端 DWS。</p> <p>本实践演示过程为：以 gsql 作为数据库客户端，gsql 安装在 ECS，通过 gsql 连接 DWS，再通过外表方式导入远端 DWS 的数据。</p>
	<p>导出ORC数据到MRS</p>	<p>GaussDB(DWS)数据库支持通过 HDFS 外表导出 ORC 格式数据至 MRS，通过外表设置的导出模式、导出数据格式等信息来指定导出的数据文件，利用多 DN 并行的方式，将数据从 GaussDB(DWS)数据库导出到外部，存放在 HDFS 文件系统中，从而提高整体导出性能。</p>
<p>数据迁移</p>	<p>Oracle迁移到 GaussDB(DWS)实践</p>	<p>本教程演示将 Oracle 业务相关的表数据迁移到 GaussDB(DWS)的数据库的基本过程。</p>
	<p>MySQL表数据实时同步到 GaussDB(DWS)实践</p>	<p>本实践演示通过华为云数据复制服务 DRS 完成 MySQL 数据实时同步到 GaussDB(DWS)的基本过程。</p>
	<p>通过DLI Flink作业将Kafka数据实时写入DWS</p>	<p>本实践演示通过数据湖探索服务 DLI Flink 作业将分布式消息服务 Kafka 的消费数据实时同步至 DWS 数据仓库，实现 Kafka 实时入库到 DWS 的过程。</p> <p>本实践预计时长 90 分钟，实践用到的云服务包括虚拟私有云 VPC 及子网、弹性负载均衡 ELB、弹性云服务器 ECS、对象存储服务 OBS、分布式消息服务 Kafka、数据湖探索 DLI 和数据仓库服务 DWS</p>

实践		描述
调优表	调优表实践	<p>在本实践中，您将学习如何优化表的设计。您首先不指定存储方式、分布键、分布方式和压缩方式创建表，然后为这些表加载测试数据并测试系统性能。接下来，您将应用优秀实践以使用新的存储方式、分布键、分布方式和压缩方式重新创建这些表，并再次为这些表加载测试数据和测试系统性能，以便比较不同的设计对表的加载性能、存储空间和查询性能的影响。</p> <p>估计时间：60 分钟。</p>
高级特性	冷热数据管理优秀实践	<p>海量大数据场景下，随着业务和数据量的不断增长，数据存储与消耗的资源也日益增长。根据业务系统中用户对不同时期数据的不同使用需求，对膨胀的数据进行“冷热”分级管理，不仅可以提高数据分析性能还能降低业务成本。针对数据使用的一些场景，可以将数据按照时间分为：热数据、冷数据。</p>

实践		描述
	<p>分区自动管理优秀实践</p>	<p>对于分区列为时间的分区表，分区自动管理功能可以自动创建新分区和删除过期分区，降低分区表的维护成本，改善查询性能。为了便于查询和维护数据，用户通常使用分区列为时间的分区表来存储时间相关的数据，例如电商的订单信息、物联网采集的实时数据。这些时间相关的数据导入分区表时，需要保证分区表要有对应时间的分区，由于普通的分区表不会自动创建新的分区和删除过期的分区，所以维护人员需要定期创建新分区和删除过期分区，提高了运维成本。</p> <p>为解决上述问题，GaussDB(DWS) 引入了分区自动管理特性。可通过设置表级参数period、ttl开启分区自动管理功能，使分区表可以自动创建新分区和删除过期分区，降低分区表的维护成本，改善查询性能。</p>
<p>数据库管理</p>	<p>资源管理优秀实践</p>	<p>本实践将演示GaussDB(DWS)的资源管理功能，帮助企业客户解决数据分析过程中，多用户查询作业遇到的性能瓶颈，最终实现多用户执行SQL作业互不影响，节省资源消耗。</p>
	<p>SQL查询优秀实践</p>	<p>根据数据库的SQL执行机制以及大量的实践总结发现：通过一定的规则调整SQL语句，在保证结果正确的基础上，能够提高SQL执行效率。</p>
	<p>数据倾斜查询优秀实践</p>	<p>本实践包含以下存储倾斜案例：</p> <ul style="list-style-type: none"> • 导入过程存储倾斜即时检测 • 快速定位查询存储倾斜的表
	<p>用户管理优秀实践</p>	<p>GaussDB(DWS)集群中，常用的用户分别是系统管理员和普通用户。本实践简述了系统管理员和普通用户的权限，如何创建以及如何查询用户相关信息。</p>

实践	描述	
	查看表和数据库的信息	本实践演示了基本数据库查询案例： <ul style="list-style-type: none"> • 查询表信息 • 查询表大小 • 查询数据库 • 查询数据库大小
模拟数据分析	交通卡口通行车辆分析	本实践将演示交通卡口车辆通行分析，将加载8.9亿条交通卡口车辆通行模拟数据到数据仓库单个数据库表中，并进行车辆精确查询和车辆模糊查询，展示GaussDB(DWS) 对于历史详单数据的高性能查询能力。
	供应链需求分析(TPC-H数据集)	本实践将演示从OBS加载样例数据集到GaussDB(DWS) 集群中并查询数据的流程，从而向您展示GaussDB(DWS) 在数据分析场景中的多表分析与主题分析。
	零售业百货公司经营状况分析	本实践将演示以下场景：从OBS加载各个零售商场每日经营的业务数据到数据仓库对应的表中，然后对商铺营业额、客流信息、月度销售排行、月度客流转化率、月度租售比、销售坪效等KPI信息进行汇总和查询。本示例旨在展示在零售业场景中GaussDB(DWS) 数据仓库的多维度查询分析的能力。
数据安全	实现数据列的加解密	数据加密作为有效防止未授权访问和防护数据泄露的技术，在各种信息系统中广泛使用。作为信息系统的核心，GaussDB(DWS)数仓也提供数据加密功能，包括透明加密和使用SQL函数加密。本章节主要讨论SQL函数加密。