

**Data Warehouse Service**

# **Pasos iniciales**

**Edición**            01  
**Fecha**             2023-11-30



**Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2023. Todos los derechos reservados.**

Quedan terminantemente prohibidas la reproducción y/o la divulgación totales y/o parciales del presente documento de cualquier forma y/o por cualquier medio sin la previa autorización por escrito de Huawei Cloud Computing Technologies Co., Ltd.

## **Marcas registradas y permisos**



El logotipo  y otras marcas registradas de Huawei pertenecen a Huawei Technologies Co., Ltd. Todas las demás marcas registradas y los otros nombres comerciales mencionados en este documento son propiedad de sus respectivos titulares.

## **Aviso**

Es posible que la totalidad o parte de los productos, las funcionalidades y/o los servicios que figuran en el presente documento no se encuentren dentro del alcance de un contrato vigente entre Huawei Cloud y el cliente. Las funcionalidades, los productos y los servicios adquiridos se limitan a los estipulados en el respectivo contrato. A menos que un contrato especifique lo contrario, ninguna de las afirmaciones, informaciones ni recomendaciones contenidas en el presente documento constituye garantía alguna, ni expresa ni implícita.

Huawei está permanentemente preocupada por la calidad de los contenidos de este documento; sin embargo, ninguna declaración, información ni recomendación aquí contenida constituye garantía alguna, ni expresa ni implícita. La información contenida en este documento se encuentra sujeta a cambios sin previo aviso.

---

# Índice

---

<b>1 Paso 1: Iniciar los preparativos.....</b>	<b>1</b>
<b>2 Paso 2: Creación de un clúster.....</b>	<b>3</b>
<b>3 Paso 3: Conexión a un clúster.....</b>	<b>7</b>
<b>4 Paso 4: Importar datos de muestra y realizar consultas.....</b>	<b>13</b>
4.1 Análisis de Vehículos Pasados en Puntos de Control de Tráfico.....	13
4.2 Análisis de requisitos de la cadena de suministro de una empresa.....	17
4.3 Análisis del estado de operaciones de una tienda departamental minorista.....	25
<b>5 Paso 5: Visualización de otros documentos y eliminación de recursos.....</b>	<b>34</b>

# 1 Paso 1: Iniciar los preparativos

---

Esta guía es un tutorial introductorio que muestra cómo crear un clúster de GaussDB de muestra (DWS), conectarse a la base de datos del clúster, importar los datos de muestra de OBS y analizar los datos de muestra. Puede utilizar este tutorial para evaluar GaussDB(DWS).

Antes de crear un clúster de GaussDB(DWS), asegúrese de que se cumplen los siguientes requisitos previos:

- [Registro de una cuenta de nube pública y finalización de la autenticación con nombre real](#)
- [Determinación de los puertos del clúster](#)

## Registro de una cuenta de nube pública y finalización de la autenticación con nombre real

Si no tiene una cuenta de nube pública, registre una. Si ya tiene una cuenta con autenticación de nombre real, omita este paso y use la cuenta.

1. Abra el sitio web oficial de la nube pública (<https://www.huaweicloud.com/intl/es-us/>) y haga clic en **Register** en la esquina superior derecha. Se muestra la página de registro.
2. Rellene la información del usuario según las instrucciones para completar el registro.
3. Haga clic en el nombre de usuario en la esquina superior derecha para ingresar a la página **Account Info**. A continuación, haga clic en **Real-Name Authentication** en el panel de navegación izquierdo.
4. Realice la autenticación de nombre real según se le solicite.

### NOTA

Se debe llevar a cabo la autenticación de nombre real antes de habilitar los servicios en la nube.

## Determinación de los puertos del clúster

- Al crear un clúster de GaussDB(DWS), debe especificar un puerto para que los clientes o aplicaciones de SQL accedan al clúster.
- Si su cliente está detrás de un firewall, necesita un puerto disponible para conectarse al clúster y realizar consultas y análisis desde la herramienta cliente de SQL.

- Si no conoce un puerto disponible, póngase en contacto con el administrador de la red para especificar un puerto abierto en el firewall. Los puertos soportados por GaussDB(DWS) oscilan entre 8000 y 30000.
- Después de crear un clúster, no se puede cambiar su número de puerto. Asegúrese de que el puerto especificado esté disponible.

# 2 Paso 2: Creación de un clúster

Antes de utilizar GaussDB(DWS) para analizar datos, cree un clúster. Un clúster consta de varios nodos en la misma subred. Estos nodos juntos proporcionan servicios. Esta sección describe cómo crear un clúster de GaussDB(DWS).

## Creación de un clúster

- Paso 1** Inicie sesión en la consola de gestión de GaussDB(DWS).
- Paso 2** Seleccione **Clusters** en el panel de navegación de la izquierda.
- Paso 3** En la página de **Clusters**, haga clic en **Create Cluster** en la esquina superior derecha.
- Paso 4** Seleccione la región a la que pertenece el clúster que se va a crear.
- **Region:** Seleccione **CN-Hong Kong**.
  - **AZ:** Conservar el valor predeterminado.
- Paso 5** Configurar parámetros de nodo.
- **Resource:** Por ejemplo, **Estándar**.
  - **CPU Architecture:** seleccione una arquitectura de CPU basada en sus requisitos, por ejemplo, **x86**.
  - **Node Flavor:** conserva el valor predeterminado.
  - **Nodos:** conserva el valor predeterminado. Se requieren al menos **3** nodos.

Figura 2-1 Configuración de parámetros de nodo

The screenshot displays the configuration interface for a GaussDB(DWS) cluster. It includes several sections:

- Resource:** Tabs for Cloud, Standard, Open Beta Test, Stream, and Hybrid.
- CPU Architecture:** Options for Kunpeng and x86.
- Node Flavor:** A table with columns for Flavor Name, vCPUs | Memory, and Application Scenario.

Flavor Name	vCPUs   Memory	Application Scenario
<input checked="" type="radio"/> dwsx2.2xlarge	8 vCPUs   64GB	Production environment
<input type="radio"/> dwsx2.8xlarge	32 vCPUs   256GB	Production environment
<input type="radio"/> dwsx2.16xlarge (Sold Out)	64 vCPUs   512GB	Production environment
- Storage Type:** Ultra-high I/O.
- Available Storage for Each Node:** A slider set to 100 GB, with a total capacity of 4,000 GB.
- Nodes:** A numeric input field set to 3, with a note: "You can create 799999975 more nodes. Increase quota".
- Total Capacity:** 300 GB.
- Discount Nodes:** A note: "You have not purchased a discount package containing nodes of flavor dwsx2.2xlarge." with links for "Buy Discount Package" and "View Order".

**Paso 6** Configurar los parámetros del clúster.

- **Cluster Name:** Ingrese **dws-demo**.
- **Cluster Version:** se muestra la versión actual del clúster y no se puede cambiar.
- **Default Database:** el valor es **gaussdb**, que no se puede cambiar.
- **Administrator Account:** el valor predeterminado es **dbadmin**. Utilice el valor predeterminado. Después de crear un clúster, el cliente utiliza esta cuenta de administrador y su contraseña para conectarse a la base de datos del clúster.
- **Administrator Password:** Introduzca la contraseña.
- **Confirm Password:** Ingrese de nuevo la contraseña del administrador.
- **Database Port:** Utilice el número de puerto predeterminado. Este puerto es utilizado por el cliente o la aplicación para conectarse a la base de datos del clúster.

**Figura 2-2** Configuración del clúster

The image shows a configuration form for a GaussDB(DWS) cluster. The fields and their values are as follows:

Field	Value
Cluster Name	
Cluster Version	Two radio buttons, one selected.
Default Database	gaussdb
Administrator Account	dbadmin
Administrator Password	
Confirm Password	
Database Port	8000

**Paso 7** Configurar parámetros de red.

- **VPC:** Puede seleccionar una VPC existente en la lista desplegable. Si no se ha configurado ninguna VPC, haga clic en **View VPC** para entrar en la consola de administración de VPC para crear una, por ejemplo, **vpc-dws**. A continuación, vuelva a la página para crear un clúster en la consola de gestión de GaussDB(DWS), haga clic



junto a la lista desplegable de **VPC** y seleccione la nueva VPC.

- **Subnet:** Cuando se crea una VPC, se crea una subred de forma predeterminada. Puede seleccionar la subred correspondiente.
- **Security Group:** Seleccione **Automatic creation**.

El grupo de seguridad creado automáticamente se denomina **GaussDB(DWS)-<Cluster name>-<GaussDB(DWS) cluster database port>**. El saliente permite todas las solicitudes de acceso, mientras que el entrante solo habilita el **Database Port** para las solicitudes de acceso de clientes o aplicaciones.

Si selecciona un grupo de seguridad personalizado, agregue una regla de entrada para habilitar **Database Port** para que los hosts cliente accedan a GaussDB(DWS). [Tabla 2-1](#) muestra un ejemplo. Para obtener más información sobre cómo agregar una regla de

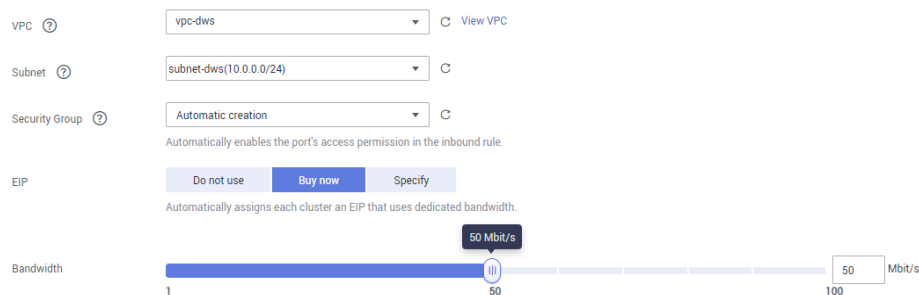
grupo de seguridad, consulte [Adición de una regla de grupo de seguridad](#) en la *Guía del usuario de la nube privada virtual*.

**Tabla 2-1** Ejemplo de regla entrante

Parámetro	Valor de ejemplo
Protocolo/Aplicación	TCP
Puerto	8000 <b>NOTA</b> Introduzca el valor del conjunto de <b>Database Port</b> al crear el clúster GaussDB(DWS). Este puerto se utiliza para recibir conexiones de cliente a GaussDB(DWS). El número de puerto predeterminado es 8000.
Fuente	Seleccione <b>IP address</b> e introduzca la dirección IP y la máscara de subred del host de cliente que accede a GaussDB(DWS), por ejemplo, <b>192.168.0.10/16</b> .

- **EIP**: Seleccione **Buy now** para comprar un EIP para el clúster como dirección de IP de red pública del clúster y establezca su **Bandwidth**.

**Figura 2-3** Configuración de la red



**Paso 8** Seleccione **Default** para **Advanced Settings** en este ejemplo.

- **Default**: indica que las siguientes configuraciones avanzadas utilizan las configuraciones predeterminadas.
  - **Automated Snapshot**: De forma predeterminada, la política para crear instantáneas de clúster automáticamente está habilitada.
  - **CNs**: Se despliegan dos CNs por defecto.
  - **Parameter Template**: la plantilla de parámetro de base de datos predeterminada está asociada con el clúster.
  - **Tag**: De forma predeterminada, no se agrega ninguna etiqueta al clúster.
  - **Encrypt DataStore**: Este parámetro está deshabilitado de forma predeterminada, lo que indica que la base de datos no está cifrada.
- **Custom**: Seleccione esta opción para configurar los siguientes parámetros avanzados: **Automated Snapshot**, **CNs**, **Parameter Template**, **Tag**, **Encrypt DataStore**

**Paso 9** Haga clic en **Next**. Se muestra la página de **Confirm**.

**Paso 10** Haga clic en **Submit**.



Después de que el envío es exitoso, la creación comienza. Haga clic en **Back to Cluster List**. Se muestra la página **Clusters**. El estado inicial del clúster es **Creating**. La creación de clústeres lleva algún tiempo. Espere un rato. Los clústeres en el estado **Available** están listos para su uso.

---**Fin**

# 3 Paso 3: Conexión a un clúster

---

## Escenario


Esta sección describe cómo utilizar un cliente de base de datos para conectarse a la base de datos en un clúster de GaussDB(DWS). En el siguiente ejemplo, la herramienta de cliente de Data Studio se utiliza para la conexión a través de la dirección de red pública. También puede utilizar otros clientes de SQL para conectarse al clúster. Para obtener más métodos de conexión, consulte "Conectar a un clúster" en la *Guía de administración de Data Warehouse Service (DWS)*.

1. Obtenga el nombre, nombre de usuario y contraseña de la base de datos que se va a conectar.  
Si utiliza el cliente para conectarse al clúster por primera vez, utilice el nombre de usuario y la contraseña del administrador establecidos en **Paso 2: Creación de un clúster** para conectarse a la base de datos predeterminada **gaussdb**.
2. **Obtención de la dirección de red pública del clúster**: Conéctese a la base de datos del clúster mediante la dirección de red pública.
3. **Conexión a la base de datos de clúster mediante Data Studio**: Descargue y configure el cliente de Data Studio y conéctese a la base de datos del clúster.

## Obtención de la dirección de red pública del clúster

**Paso 1** Inicie sesión en la consola de gestión de GaussDB(DWS).

**Paso 2** En el árbol de navegación de la izquierda, haga clic en **Clusters**.

**Paso 3** En la lista de clústeres, seleccione un clúster creado (por ejemplo, **dws-demo**) y haga clic en  junto al nombre del clúster para obtener la dirección de red pública.

La dirección de red pública se utilizará en **Conexión a la base de datos de clúster mediante Data Studio**

**Figura 3-1** Gestión de clústeres

Cluster Name	Cluster Status	Task Informatio...	Node Flavor	Rece...	Operation
dws-demo	Available	--	dws.m3.xl...	2	View Metric   Restart   More ▾

Private Network Address	[Redacted] .com		
Public Network Address	[Redacted] .com		
Region	Shanghai2	VPC	vpc-dws
AZ	AZ1	Security Group	dws-dws-demo-8000
Subnet	subnet-dws(10.0.0.0/24)	Created	Jan 21, 2020 09:37:04 GMT+08:00
Cluster Version	1.5.600	Last Snapshot Created	--
Nodes	3		
Tag	--		

---Fin

## Conexión a la base de datos de clúster mediante Data Studio

**Paso 1** GaussDB(DWS) proporciona un cliente de GUI de Data Studio basado en Windows. La herramienta depende de JDK, por lo que debe instalar Java 1.8.0\_141 o posterior en el host cliente.

En el sistema operativo Windows, puede descargar la versión requerida de JDK desde el [sitio web oficial de JDK](#), e instalarla siguiendo la guía de instalación.

**Paso 2** Inicie sesión en la consola de gestión de GaussDB(DWS).

**Paso 3** Haga clic en **Conexiones**.

**Paso 4** En la página de **Download Client and Driver**, descargue **Data Studio GUI Client**.

- Seleccione **Windows x86** o **Windows x64** según el tipo de sistema operativo y haga clic en **Download** para descargar la herramienta Data Studio que coincida con la versión actual del clúster.

Si hay clústeres de diferentes versiones disponibles, descargará la herramienta Data Studio que coincida con la versión más antigua del clúster después de hacer clic en **Download**. Si no hay clúster, descargará la herramienta Data Studio de la primera versión después de hacer clic en **Download**. Los clústeres GaussDB(DWS) son compatibles con versiones anteriores de las herramientas de Data Studio.

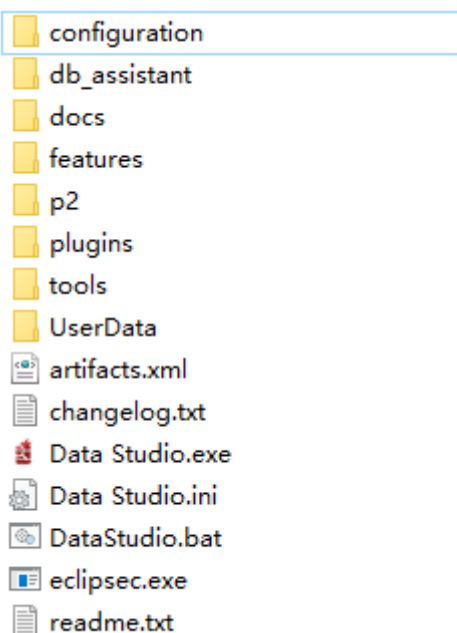
- Haga clic en **Historical Version** para descargar la versión de Data Studio correspondiente. Se recomienda descargar el Data Studio basado en la versión del clúster.

Si tiene clústeres de diferentes versiones, el sistema muestra un cuadro de diálogo en el que se le solicita que seleccione la versión del clúster y descargue el cliente correspondiente. En la lista de clústeres de la página de **Clusters**, haga clic en el nombre del clúster especificado y haga clic en la ficha **Basic Information** para ver la versión del clúster.

**Paso 5** Descomprima el paquete de software cliente descargado (32 bits o 64 bits) en el directorio de instalación.

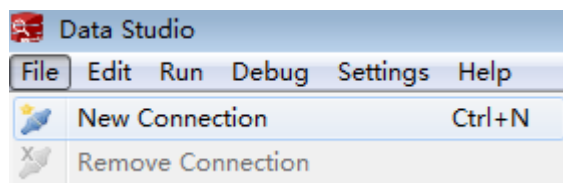
**Paso 6** Abra el directorio de instalación y haga doble clic en **Data Studio.exe** para iniciar el cliente de Data Studio. Consulte [Figura 3-2](#).

Figura 3-2 Iniciar el cliente



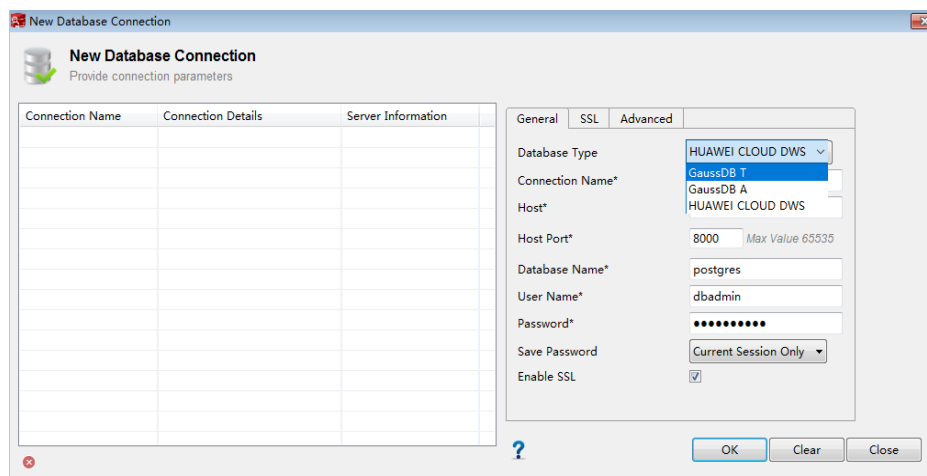
Paso 7 Seleccione **File > New Connection** desde el menú principal. Consulte [Figura 3-3](#).

Figura 3-3 Cómo crear una conexión



Paso 8 En la ventana **New Database Connection** mostrada, introduzca los parámetros de conexión.

Figura 3-4 Configuración de parámetros de conexión



**Tabla 3-1** Parámetros de conexión

Parámetro	Descripción	Ejemplo
Tipo de base de datos	Seleccione <b>HUAWEI CLOUD DWS</b> .	HUAWEI CLOUD DWS
Nombre de la conexión	Nombre de una conexión	dws-demo
Host	Dirección IP (IPv4) o nombre de dominio del clúster a conectar	-
Puerto de host	Puerto de la base de datos	8000
Nombre de la base de datos	Nombre de la base de datos	gaussdb
Nombre de usuario	Nombre de usuario para conectarse a la base de datos	-
Contraseña	Contraseña para iniciar sesión en la base de datos que se va a conectar	-
Guardar contraseña	Seleccione una opción de la lista desplegable: <ul style="list-style-type: none"> <li>● <b>Current Session Only</b>: la contraseña se guarda solo en la sesión actual.</li> <li>● <b>Do Not Save</b>: la contraseña no se guarda.</li> </ul>	-
Habilitar SSL	Si se selecciona <b>Enable SSL</b> , el cliente puede usar SSL para cifrar conexiones. El modo SSL es más seguro que los modos comunes, por lo que se recomienda habilitar la conexión SSL.	-

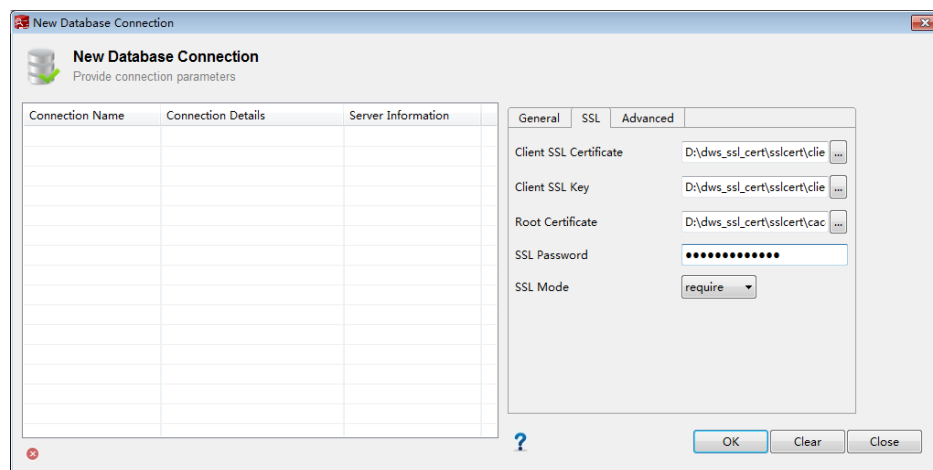
Cuando se selecciona **Enable SSL**, descargue y descomprima el certificado SSL. Para obtener más información, consulte [Descargar el certificado SSL](#). A continuación, haga clic en la ficha de **SSL** en la ventana que se muestra en [Figura 3-4](#) y configure los parámetros que se muestran en la siguiente tabla.

**Tabla 3-2** Configuración de parámetros de SSL

Parámetro	Descripción
Certificado de SSL de cliente	Seleccione el archivo de <code>sslcert\client.crt</code> en el directorio de certificados SSL descomprimido.

Parámetro	Descripción
Clave SSL de cliente	Solo se admite el formato PK8. Seleccione el archivo de <b>sslcert\client.key.pk8</b> en el directorio donde se descomprime el certificado SSL.
Certificado raíz	Cuando el modo <b>SSL Mode</b> se establece en <b>verify-ca</b> , se debe configurar el certificado root. Seleccione el archivo de <b>sslcert\cacert.pem</b> en el directorio de certificados SSL descomprimido.
Contraseña de SSL	Establezca la contraseña para la clave de SSL del cliente en formato de PK8.
Modo SSL	Los modos SSL compatibles incluyen: <ul style="list-style-type: none"> <li>● require</li> <li>● verify-ca</li> </ul> GaussDB(DWS) no soporta el modo de <b>verify-full</b> .

**Figura 3-5** Configuración de parámetros de SSL

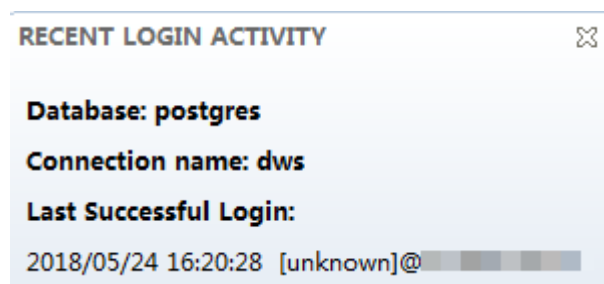


**Paso 9** Haga clic en **OK** para establecer la conexión a la base de datos.

Si SSL está habilitado, haga clic en **Continue** en el cuadro de diálogo **Connection Security Alert** que se muestra.

Una vez que el inicio de sesión se realiza correctamente, se muestra el cuadro de diálogo **RECENT LOGIN ACTIVITY**, que indica que Data Studio está conectado a la base de datos. Puede ejecutar la instrucción SQL en la ventana de **SQL Terminal** de la página Data Studio.

**Figura 3-6** Inicio de sesión exitoso



Para obtener más información sobre cómo utilizar otras funciones de Data Studio, presione **F1** para ver el manual de usuario de Data Studio.

----**Fin**

# 4 Paso 4: Importar datos de muestra y realizar consultas

Esta sección proporciona tres ejemplos independientes para que usted experimente GaussDB(DWS).

## 4.1 Análisis de Vehículos Pasados en Puntos de Control de Tráfico

Esta práctica demuestra el análisis de los vehículos pasados en los puntos de control de tráfico. En esta práctica, 890 millones de piezas de datos de puntos de control de tráfico se cargan en una única tabla de base de datos en GaussDB(DWS) para realizar una consulta precisa y una consulta difusa. Es un ejemplo de consulta de alto rendimiento de datos históricos en GaussDB(DWS).

### NOTA

Los datos de muestra se han cargado a la carpeta de **traffic-data** en un bucket de OBS, y a todas las cuentas de Huawei Cloud se les ha concedido el permiso de solo lectura para acceder al depósito OBS.

Esta práctica dura unos 40 minutos. El proceso básico es el siguiente:


1. **Preparación**
2. **Paso 1: Creación de un clúster**
3. **Paso 2: Usar Data Studio para conectarse a un clúster**
4. **Paso 3: Importación de datos de muestra**
5. **Paso 4: Realizar el análisis del vehículo**



### Preparación

- Registre una Huawei Cloud cuenta y compruebe el estado de la cuenta antes de usar GaussDB(DWS). La cuenta no puede estar en mora o congelada.
- Usted ha obtenido el AK y el SK de la cuenta.



## Paso 1: Creación de un clúster

- Paso 1** Inicie sesión en la consola de gestión de Huawei Cloud.
- Paso 2** Seleccione **Service List > Analytics > GaussDB(DWS)**.
- Paso 3** En el panel de navegación de la izquierda, elija **Clusters**. En la página mostrada, haga clic en **Create Cluster** en la esquina superior derecha.
- Paso 4** Para obtener más información, consulte [Creación de un clúster](#).
- Paso 5** Confirme la información, haga clic en **Next**, and then click **Submit**.
- Paso 6** Espere unos 6 minutos. Después de crear el clúster, haga clic en  junto al nombre del clúster. En la página de información del clúster mostrada, registre el valor de **Public Network Address**, por ejemplo, **dws-demov.dws.huaweicloud.com**.

Cluster Name	Cluster Status
 dws-demo	 Available
Private Network Address	dws-demo.dws.myhuaweiclouds.com
Public Network Address	<span style="border: 2px solid red; padding: 2px;">dws-demov.dws.huaweicloud.com</span>

----Fin

## Paso 2: Usar Data Studio para conectarse a un clúster

- Paso 1** Asegúrese de que JDK 1.8.0 o posterior se ha instalado en el host de cliente. Seleccione **PC > Properties > Advanced System Settings > Environment Variables** y establezca **JAVA\_HOME** (por ejemplo, **C:\Program Files\Java\jdk1.8.0\_191**). Agregue ; **%JAVA\_HOME%\bin** a la **path** de acceso de la variable.
- Paso 2** En la página **Connections** de la consola GaussDB(DWS), descargue el cliente GUI de Data Studio.
- Paso 3** Descomprima el paquete de software de Data Studio descargado, vaya al directorio descomprimido y haga doble clic en **Data Studio.exe** para iniciar el cliente.
- Paso 4** En el menú principal de Data Studio, seleccione **File > New Connection**. En el cuadro de diálogo que se muestra, configure la conexión basada en [Tabla 4-1](#).

**Tabla 4-1** Configuración del software Data Studio

Parámetro	Configuración
Tipo de base de datos	GaussDB(DWS)

Parámetro	Configuración
Nombre de la conexión	dws-demo
Host	dws-demov.dws.huaweicloud.com El valor de este parámetro debe ser el mismo que el valor de <b>Public Network Address</b> consultada en <b>Paso 1: Creación de un clúster</b> .
Puerto de host	8000
Nombre de la base de datos	gaussdb
Nombre de usuario	dbadmin
Contraseña	-
Habilitar SSL	Deshabilitar

**Paso 5** Haz clic en **OK**.

----**Fin**

### Paso 3: Importación de datos de muestra

Después de conectarse al clúster mediante la herramienta cliente SQL, realice las siguientes operaciones en la herramienta cliente SQL para importar los datos de ejemplo de los puntos de control de tráfico y realizar consultas de datos.

**Paso 1** Ejecute la siguiente instrucción para crear la base de datos de **traffic**:

```
create database traffic encoding 'utf8' template template0;
```

**Paso 2** Realice los siguientes pasos para cambiar a la nueva base de datos:

1. En la ventana **Object Browser** del cliente de Data Studio, haga clic con el botón derecho en la conexión de base de datos y elija **Refresh** en el menú contextual. A continuación, se muestra la nueva base de datos.
2. Haga clic con el botón derecho en el nombre del nuevo **traffic** de base de datos y elija **Connect to DB** en el menú contextual.
3. Haga clic con el botón derecho en el nombre de la nueva base de datos de **traffic** y elija **Open Terminal** en el menú contextual. Se muestra la ventana de comandos SQL para conectarse a la base de datos especificada. Realice los siguientes pasos en la ventana.

**Paso 3** Ejecute las siguientes instrucciones para crear una tabla de base de datos para almacenar información del vehículo desde puntos de control de tráfico:

```
create schema traffic_data;
set current_schema= traffic_data;
drop table if exists GCJL;
CREATE TABLE GCJL
(
    kkbh    VARCHAR(20),
    hphm    VARCHAR(20),
    gcsj    DATE,
    cplx    VARCHAR(8),
```

```

        cllx  VARCHAR(8),
        csys  VARCHAR(8)
    )
with (orientation = column, COMPRESSION=MIDDLE)
distribute by hash(hphm);

```

**Paso 4** Cree una tabla externa, que se utiliza para identificar y asociar los datos de origen en OBS.

#### AVISO

- `<obs_bucket_name>` indica el nombre del bucket OBS. Solo algunas regiones son compatibles. Los clústeres de GaussDB(DWS) no admiten el acceso entre regiones a los datos del bucket de OBS.
- En esta práctica, se utiliza como ejemplo **CN-Hong Kong** Escriba **dws-demo-ap-southeast-1** y reemplace `<Access_Key_Id>` y `<Secret_Access_Key>` por el valor obtenido en [Preparación](#).
- Si se muestra el mensaje "ERROR: schema "xxx" does not exist Position" al crear una tabla externa, el esquema no existe. Realice el paso anterior para crear un esquema.

```

create schema tpchobs;
set current_schema = 'tpchobs';
drop FOREIGN table if exists GCJL_OBS;
CREATE FOREIGN TABLE GCJL_OBS
(
    like traffic_data.GCJL
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/traffic-data/gcxx',
    format 'text',
    delimiter ',',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);

```

**Paso 5** Ejecute la siguiente instrucción para importar datos de la tabla externa a la tabla de base de datos:

```
insert into traffic_data.GCJL select * from tpchobs.GCJL_OBS;
```

Se necesita algún tiempo para importar datos.

----Fin

## Paso 4: Realizar el análisis del vehículo

### 1. Realización de ANALYZE

Esta instrucción recopila estadísticas relacionadas con tablas ordinarias en bases de datos. Las estadísticas se guardan en el catálogo del sistema **PG\_STATISTIC**. Cuando ejecuta el planificador, las estadísticas le ayudan a desarrollar un plan de ejecución de consultas eficiente.

Ejecute la siguiente instrucción para generar las estadísticas de la tabla:

```
Analyze;
```

### 2. Consultar el volumen de datos de la tabla de datos

Ejecute la siguiente instrucción para consultar el número de registros de datos cargados:

```
set current_schema= traffic_data;  
Select count(*) from traffic_data.gcjl;
```

### 3. Consulta precisa del vehículo

Ejecute las siguientes instrucciones para consultar la ruta de conducción de un vehículo por el número de matrícula y el segmento de tiempo. GaussDB(DWS) responde a la solicitud en segundos.

```
set current_schema= traffic_data;  
select hphm, kkbh, gcsj  
from traffic_data.gcjl  
where hphm = 'YD38641'  
and gcsj between '2016-01-06' and '2016-01-07'  
order by gcsj desc;
```

### 4. Consulta de vehículo difuso

Ejecute las siguientes instrucciones para consultar la ruta de conducción de un vehículo por el número de matrícula y el segmento de tiempo. GaussDB(DWS) responde a la solicitud en segundos.

```
set current_schema= traffic_data;  
select hphm, kkbh, gcsj  
from traffic_data.gcjl  
where hphm like 'YA23F%'  
and kkbh in('508', '1125', '2120')  
and gcsj between '2016-01-01' and '2016-01-07'  
order by hphm, gcsj desc;
```

## 4.2 Análisis de requisitos de la cadena de suministro de una empresa

Esta práctica describe cómo cargar el conjunto de datos de ejemplo de OBS a un clúster de almacén de datos y realizar consultas de datos. Este ejemplo comprende análisis de múltiples tablas y análisis de temas en el escenario de análisis de datos.

### NOTA

En este ejemplo, se ha generado un conjunto de datos TPC-H-1x estándar de 1 GB de tamaño en GaussDB (DWS), y se ha cargado a la carpeta **tpch** de un bucket OBS. A todas las cuentas de Huawei Cloud se les ha concedido el permiso de solo lectura para acceder al bucket OBS. Los usuarios pueden importar fácilmente el conjunto de datos usando sus cuentas.

Esta práctica dura unos 60 minutos. El proceso básico es el siguiente:

1. [Preparación](#)
2. [Paso 1: Importación de datos de muestra](#)
3. [Paso 2: Realización de análisis de múltiples tablas y análisis de temas](#)

## Descripción del escenario

Finalidad: Comprender las funciones básicas de GaussDB(DWS) y cómo importar datos. Analizar los datos de pedidos de una empresa y sus proveedores de la siguiente manera:

1. Analizar los ingresos aportados por los proveedores de una región a la empresa. Las estadísticas se pueden usar para determinar si es necesario establecer un centro de asignación local en una región dada.
2. Analizar la relación entre piezas y proveedores para obtener el número de proveedores de piezas en función de las condiciones de contribución especificadas. La información se

puede utilizar para determinar si los proveedores son suficientes para grandes cantidades de pedidos cuando la tarea es urgente.

3. Analizar la pérdida de ingresos de pequeños pedidos. Puede consultar la pérdida de ingresos promedio anual si no hay pedidos pequeños. Filtrar pedidos pequeños que son inferiores al 20% del volumen de suministro promedio, y calcular la cantidad total de esos pedidos pequeños para calcular la pérdida de ingresos anuales promedio.

## Preparación

- Registre una Huawei Cloud cuenta y compruebe el estado de la cuenta antes de usar GaussDB(DWS). La cuenta no puede estar en mora o congelada.
- Usted ha obtenido el AK y el SK de la cuenta.
- Se ha creado un clúster y se ha conectado mediante Data Studio. Para más detalles, consulte [Análisis de Vehículos Pasados en Puntos de Control de Tráfico](#).

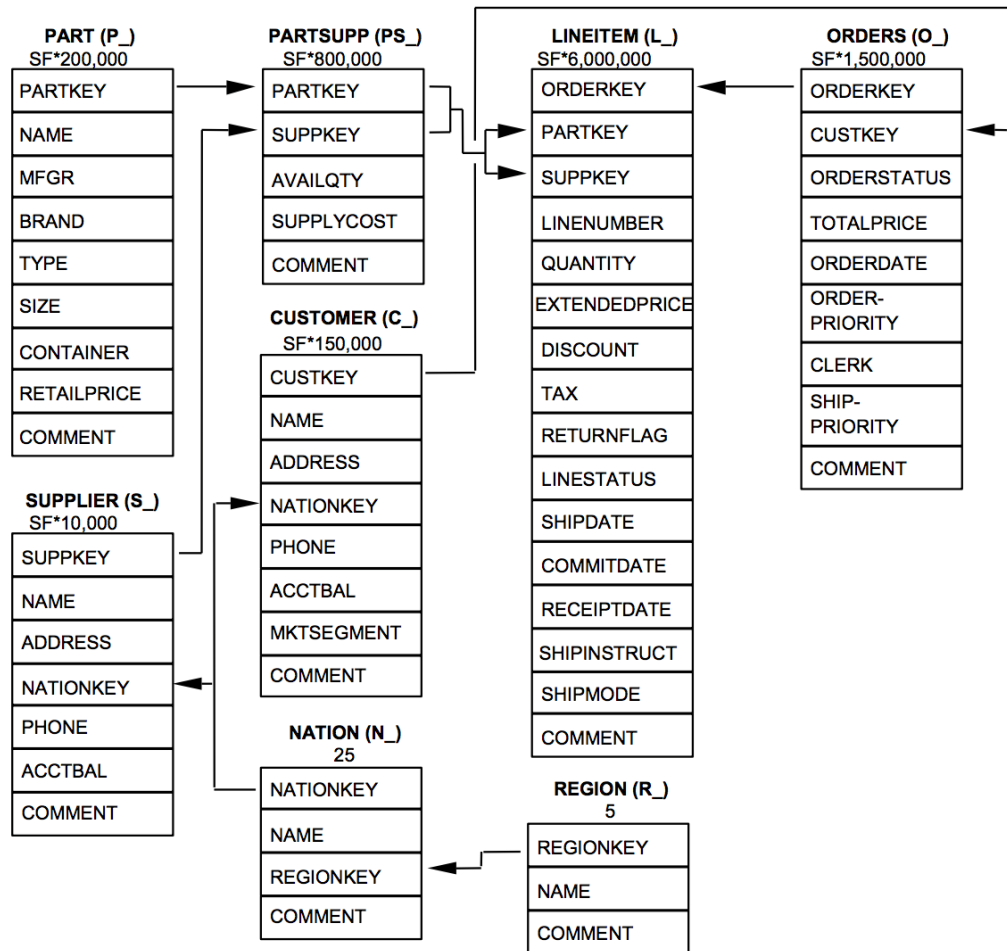
## Paso 1: Importación de datos de muestra

Después de conectarse al clúster mediante la herramienta cliente SQL, realice las siguientes operaciones en la herramienta cliente SQL para importar los datos de muestra de TPC-H y realizar consultas de datos.

### **Paso 1** Crear una tabla de base de datos.

Los datos de muestra de TPC-H consisten en ocho tablas de base de datos cuyas asociaciones se muestran en [Figura 4-1](#).

Figura 4-1 Tablas de datos TPC-H



Ejecute las siguientes instrucciones para crear tablas en la **gaussdb** base de datos.

```
CREATE schema tpch;
set current_schema = tpch;

drop table if exists region;
CREATE TABLE REGION
(
    R_REGIONKEY INT NOT NULL ,
    R_NAME      CHAR(25) NOT NULL ,
    R_COMMENT   VARCHAR(152)
)
with (orientation = column, COMPRESSION=MIDDLE)
distribute by replication;

drop table if exists nation;
CREATE TABLE NATION
(
    N_NATIONKEY INT NOT NULL,
    N_NAME      CHAR(25) NOT NULL,
    N_REGIONKEY INT NOT NULL,
    N_COMMENT   VARCHAR(152)
)
with (orientation = column, COMPRESSION=MIDDLE)
distribute by replication;

drop table if exists supplier;
CREATE TABLE SUPPLIER
```

```
(
    S_SUPPKEY      BIGINT NOT NULL,
    S_NAME         CHAR(25) NOT NULL,
    S_ADDRESS      VARCHAR(40) NOT NULL,
    S_NATIONKEY    INT NOT NULL,
    S_PHONE        CHAR(15) NOT NULL,
    S_ACCTBAL      DECIMAL(15,2) NOT NULL,
    S_COMMENT      VARCHAR(101) NOT NULL
)
with (orientation = column,COMPRESSION=MIDDLE)
distribute by hash(S_SUPPKEY);

drop table if exists customer;
CREATE TABLE CUSTOMER
(
    C_CUSTKEY      BIGINT NOT NULL,
    C_NAME         VARCHAR(25) NOT NULL,
    C_ADDRESS      VARCHAR(40) NOT NULL,
    C_NATIONKEY    INT NOT NULL,
    C_PHONE        CHAR(15) NOT NULL,
    C_ACCTBAL      DECIMAL(15,2) NOT NULL,
    C_MKTSEGMENT  CHAR(10) NOT NULL,
    C_COMMENT      VARCHAR(117) NOT NULL
)
with (orientation = column,COMPRESSION=MIDDLE)
distribute by hash(C_CUSTKEY);

drop table if exists part;
CREATE TABLE PART
(
    P_PARTKEY      BIGINT NOT NULL,
    P_NAME         VARCHAR(55) NOT NULL,
    P_MFGR         CHAR(25) NOT NULL,
    P_BRAND        CHAR(10) NOT NULL,
    P_TYPE         VARCHAR(25) NOT NULL,
    P_SIZE         BIGINT NOT NULL,
    P_CONTAINER    CHAR(10) NOT NULL,
    P_RETAILPRICE DECIMAL(15,2) NOT NULL,
    P_COMMENT      VARCHAR(23) NOT NULL
)
with (orientation = column,COMPRESSION=MIDDLE)
distribute by hash(P_PARTKEY);

drop table if exists partsupp;
CREATE TABLE PARTSUPP
(
    PS_PARTKEY     BIGINT NOT NULL,
    PS_SUPPKEY     BIGINT NOT NULL,
    PS_AVAILQTY    BIGINT NOT NULL,
    PS_SUPPLYCOST DECIMAL(15,2) NOT NULL,
    PS_COMMENT     VARCHAR(199) NOT NULL
)
with (orientation = column,COMPRESSION=MIDDLE)
distribute by hash(PS_PARTKEY);

drop table if exists orders;
CREATE TABLE ORDERS
(
    O_ORDERKEY     BIGINT NOT NULL,
    O_CUSTKEY      BIGINT NOT NULL,
    O_ORDERSTATUS  CHAR(1) NOT NULL,
    O_TOTALPRICE   DECIMAL(15,2) NOT NULL,
    O_ORDERDATE    DATE NOT NULL ,
    O_ORDERPRIORITY CHAR(15) NOT NULL,
    O_CLERK        CHAR(15) NOT NULL ,
    O_SHIPPRIORITY BIGINT NOT NULL,
    O_COMMENT      VARCHAR(79) NOT NULL
)
with (orientation = column,COMPRESSION=MIDDLE)
```

```
distribute by hash(O_ORDERKEY);

drop table if exists lineitem;
CREATE TABLE LINEITEM
(
    L_ORDERKEY    BIGINT NOT NULL,
    L_PARTKEY     BIGINT NOT NULL,
    L_SUPPKEY     BIGINT NOT NULL,
    L_LINENUMBER  BIGINT NOT NULL,
    L_QUANTITY    DECIMAL(15,2) NOT NULL,
    L_EXTENDEDPRICE DECIMAL(15,2) NOT NULL,
    L_DISCOUNT   DECIMAL(15,2) NOT NULL,
    L_TAX         DECIMAL(15,2) NOT NULL,
    L_RETURNFLAG  CHAR(1) NOT NULL,
    L_LINESTATUS  CHAR(1) NOT NULL,
    L_SHIPDATE    DATE NOT NULL,
    L_COMMITDATE  DATE NOT NULL,
    L_RECEIPTDATE DATE NOT NULL,
    L_SHIPINSTRUCT CHAR(25) NOT NULL,
    L_SHIPMODE    CHAR(10) NOT NULL,
    L_COMMENT     VARCHAR(44) NOT NULL
)
with (orientation = column,COMPRESSION=MIDDLE)
distribute by hash(L_ORDERKEY);
```

**Paso 2** Crear una tabla externa, que se utiliza para identificar y asociar los datos de origen en OBS.

#### AVISO

- `<obs_bucket_name>` indica el nombre del bucket OBS. Solo algunas regiones son compatibles. Los clústeres de GaussDB(DWS) no admiten el acceso entre regiones a los datos del bucket de OBS.
- En esta práctica, se utiliza como ejemplo **CN-Hong Kong** Escriba **dws-demo-ap-southeast-1** y reemplace `<Access_Key_Id>` y `<Secret_Access_Key>` por el valor obtenido en **Preparación**.
- Si se muestra el mensaje "ERROR: schema "xxx" does not exist Position" al crear una tabla externa, el esquema no existe. Realice el paso anterior para crear un esquema.

```
CREATE schema tpchobs;
set current_schema='tpchobs';
drop FOREIGN table if exists region;
CREATE FOREIGN TABLE REGION
(
    like tpch.region
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/region.tbl',
    format 'text',
    delimiter '|',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
drop FOREIGN table if exists nation;
CREATE FOREIGN TABLE NATION
(
    like tpch.nation
)
SERVER gsmpp_server
OPTIONS (
```



```
        encoding 'utf8',
        location 'obs://<obs_bucket_name>/tpch/nation.tbl',
        format 'text',
        delimiter '|',
        access_key '<Access_Key_Id>',
        secret_access_key '<Secret_Access_Key>',
        chunksize '64',
        IGNORE_EXTRA_DATA 'on'
    );
drop FOREIGN table if exists supplier;
CREATE FOREIGN TABLE SUPPLIER
(
    like tpch.supplier
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/supplier.tbl',
    format 'text',
    delimiter '|',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
drop FOREIGN table if exists customer;
CREATE FOREIGN TABLE CUSTOMER
(
    like tpch.customer
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/customer.tbl',
    format 'text',
    delimiter '|',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
drop FOREIGN table if exists part;
CREATE FOREIGN TABLE PART
(
    like tpch.part
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/part.tbl',
    format 'text',
    delimiter '|',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
drop FOREIGN table if exists partsupp;
CREATE FOREIGN TABLE PARTSUPP
(
    like tpch.partsupp
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/partsupp.tbl',
    format 'text',
    delimiter '|',
```

```
        access_key '<Access_Key_Id>',
        secret_access_key '<Secret_Access_Key>',
        chunksize '64',
        IGNORE_EXTRA_DATA 'on'
    );
drop FOREIGN table if exists orders;
CREATE FOREIGN TABLE ORDERS
(
    like tpch.orders
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/orders.tbl',
    format 'text',
    delimiter '|',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
drop FOREIGN table if exists lineitem;
CREATE FOREIGN TABLE LINEITEM
(
    like tpch.lineitem
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/tpch/lineitem.tbl',
    format 'text',
    delimiter '|',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on'
);
```

**Paso 3** Copie y ejecute las siguientes instrucciones para importar los datos de la tabla externa a la tabla de base de datos correspondiente.

Ejecute el comando **insert** para importar los datos de la tabla externa de OBS a la tabla de base de datos GaussDB(DWS). El núcleo de la base de datos importa simultáneamente los datos OBS a alta velocidad a GaussDB(DWS).

```
insert into tpch.lineitem select * from tpchobs.lineitem;
insert into tpch.part select * from tpchobs.part;
insert into tpch.partsupp select * from tpchobs.partsupp;
insert into tpch.customer select * from tpchobs.customer;
insert into tpch.supplier select * from tpchobs.supplier;
insert into tpch.nation select * from tpchobs.nation;
insert into tpch.region select * from tpchobs.region;
insert into tpch.orders select * from tpchobs.orders;
```

Se tarda 10 minutos en importar datos.

----Fin

## Paso 2: Realización de análisis de múltiples tablas y análisis de temas

A continuación se utiliza la consulta TPC-H estándar como ejemplo para demostrar cómo realizar una consulta de datos básica en GaussDB(DWS).

Antes de consultar datos, ejecute el comando **Analyze** para generar estadísticas relacionadas con la tabla de la base de datos. Los datos estadísticos se almacenan en la tabla del sistema PG\_STATISTIC y son útiles cuando se ejecuta el planificador, que le proporciona un plan de ejecución de consultas eficiente.

Los siguientes son ejemplos de consulta:

- **Querying revenue of a supplier in a region (TPCH-Q5)**

Al ejecutar la instrucción de consulta TPCH-Q5, puede consultar las estadísticas de ingresos de un proveedor de piezas de repuesto en una región. Los ingresos se calculan en base a la **sum(l\_extendedprice \* (1 - l\_discount))**. Las estadísticas se pueden usar para determinar si es necesario establecer un centro de asignación local en una región dada.

Copie y ejecute la siguiente instrucción TPCH-Q5 para la consulta. Esta instrucción presenta una consulta de unión de varias tablas con **GROUP BY**, **ORDER BY**, y **AGGREGATE**.

```
set current_schema='tpch';
Select
n_name,
sum(l_extendedprice * (1 - l_discount)) as revenue
from
customer,
orders,
lineitem,
supplier,
nation,
region
where
c_custkey = o_custkey
and l_orderkey = o_orderkey
and l_suppkey = s_suppkey
and c_nationkey = s_nationkey
and s_nationkey = n_nationkey
and n_regionkey = r_regionkey
and r_name = 'ASIA'
and o_orderdate >= '1994-01-01'::date
and o_orderdate < '1994-01-01'::date + interval '1 year'
group by
n_name
order by
revenue desc;
```

- **Querying relationships between spare parts and suppliers (TPCH-Q16)**

Al ejecutar la instrucción de consulta TPCH-Q16, puede obtener el número de proveedores que pueden suministrar piezas de repuesto con las condiciones de contribución especificadas. Esta información se puede utilizar para determinar si hay suficientes proveedores cuando la cantidad de pedido es grande y la tarea es urgente.

Copie y ejecute la siguiente instrucción TPCH-Q16 para la consulta. La instrucción presenta operaciones de conexión de varias tablas con subconsulta agrupar por, ordenar por, agregar, deduplicar y NOT IN.

```
set current_schema='tpch';
select
p_brand,
p_type,
p_size,
count(distinct ps_suppkey) as supplier_cnt
from
partsupp,
part
where
p_partkey = ps_partkey
and p_brand <> 'Brand#45'
and p_type not like 'MEDIUM POLISHED%'
and p_size in (49, 14, 23, 45, 19, 3, 36, 9)
and ps_suppkey not in (
    select
    s_suppkey
    from
```

```
supplier
where
  s_comment like '%Customer%Complaints%'
)
group by
  p_brand,
  p_type,
  p_size
order by
  supplier_cnt desc,
  p_brand,
  p_type,
  p_size
limit 100;
```

- **Querying revenue loss of small orders (TPCH-Q17)**

Puede consultar la pérdida de ingresos promedio anual si no hay pedidos pequeños. Filtrar pedidos pequeños que son inferiores al 20% del volumen de suministro promedio, y calcular la cantidad total de esos pedidos pequeños para calcular la pérdida de ingresos anuales promedio.

Copie y ejecute la siguiente instrucción TPCH-Q17 para la consulta. La instrucción presenta operaciones de conexión de varias tablas con subconsulta agregada y agregada.

```
set current_schema='tpch';
select
  sum(l_extendedprice) / 7.0 as avg_yearly
from
  lineitem,
  part
where
  p_partkey = l_partkey
  and p_brand = 'Brand#23'
  and p_container = 'MED BOX'
  and l_quantity < (
    select 0.2 * avg(l_quantity)
    from lineitem
    where l_partkey = p_partkey
  );
```

## 4.3 Análisis del estado de operaciones de una tienda departamental minorista

### Fondo

En esta práctica, los datos empresariales diarios de cada tienda minorista se cargan desde OBS a la tabla correspondiente en el clúster de almacén de datos para resumir y consultar KPI. Estos datos incluyen la rotación de la tienda, el flujo de clientes, la clasificación de ventas mensuales, la tasa de conversión mensual del flujo de clientes, la relación precio-alquiler mensual y las ventas por unidad de área. Este ejemplo muestra la consulta y el análisis multidimensional de GaussDB(DWS) en el escenario minorista.

#### NOTA

Los datos de muestra se han cargado a la carpeta de **traffic-data** en un bucket de OBS, y a todas las cuentas de Huawei Cloud se les ha concedido el permiso de solo lectura para acceder al depósito OBS.

Esta práctica dura unos 60 minutos. El proceso básico es el siguiente:

1. **Preparativos**

2. [Paso 1: Importación de datos de muestra desde la tienda departamental minorista](#)
3. [Paso 2: Realización del análisis del estado de las operaciones](#)

## Preparativos

- Usted ha registrado una Huawei Cloud cuenta, y la cuenta no está en mora o congelada.
- Usted ha obtenido el AK y el SK de la cuenta.
- Se ha creado un clúster y se ha conectado mediante Data Studio. Para obtener más información, consulte [Paso 1: Creación de un clúster](#) y [Paso 2: Usar Data Studio para conectarse a un clúster](#).

## Paso 1: Importación de datos de muestra desde la tienda departamental minorista

Después de conectarse al clúster mediante la herramienta cliente SQL, realice las siguientes operaciones en la herramienta cliente SQL para importar los datos de ejemplo de los grandes almacenes minoristas y realizar consultas.

**Paso 1** Ejecute la siguiente instrucción para crear la base de datos **minorista**:

```
create database retail encoding 'utf8' template template0;
```

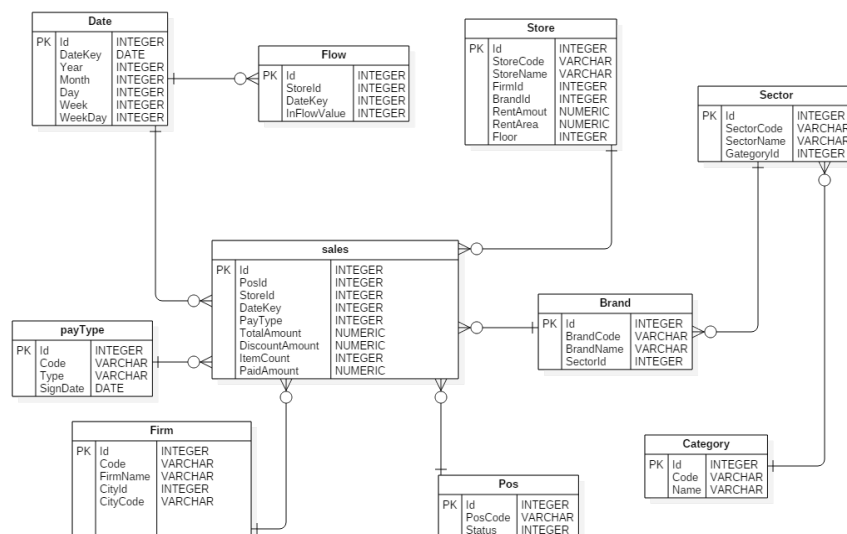
**Paso 2** Realice los siguientes pasos para cambiar a la nueva base de datos:

1. En la ventana **Object Browser** del cliente de Data Studio, haga clic con el botón derecho en la conexión de base de datos y elija **Refresh** en el menú contextual. A continuación, se muestra la nueva base de datos.
2. Haga clic con el botón derecho en el nombre de la nueva base de datos **retail** y elija **Connect to DB** a la base de datos en el menú contextual.
3. Haga clic con el botón derecho en el nombre de la nueva base de datos **retail** y elija **Open Terminal** en el menú contextual. Se muestra la ventana de comandos SQL para conectarse a la base de datos especificada. Realice los siguientes pasos en la ventana.

**Paso 3** Crear una tabla de base de datos.

Los datos de muestra consisten en 10 tablas de base de datos cuyas asociaciones se muestran en [Figura 4-2](#).

**Figura 4-2** Ejemplos de tablas de datos de grandes almacenes minoristas



Copie y ejecute las siguientes instrucciones para cambiar a crear una tabla de base de datos de información de tiendas por departamentos.

```
create schema retail_data;
set current_schema='retail_data';

DROP TABLE IF EXISTS STORE;
CREATE TABLE STORE (
    ID INT,
    STORECODE VARCHAR(10),
    STORENAME VARCHAR(100),
    FIRMID INT,
    FLOOR INT,
    BRANDID INT,
    RENTAMOUNT NUMERIC(18,2),
    RENTAREA NUMERIC(18,2)
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS POS;
CREATE TABLE POS (
    ID INT,
    POSCODE VARCHAR(20),
    STATUS INT,
    MODIFICATIONDATE DATE
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS BRAND;
CREATE TABLE BRAND (
    ID INT,
    BRANDCODE VARCHAR(10),
    BRANDNAME VARCHAR(100),
    SECTORID INT
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS SECTOR;
CREATE TABLE SECTOR(
    ID INT,
    SECTORCODE VARCHAR(10),
    SECTORNAME VARCHAR(20),
    CATEGORYID INT
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS CATEGORY;
CREATE TABLE CATEGORY(
    ID INT,
    CODE VARCHAR(10),
    NAME VARCHAR(20)
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS FIRM;
CREATE TABLE FIRM(
    ID INT,
    CODE VARCHAR(4),
    NAME VARCHAR(40),
    CITYID INT,
    CITYNAME VARCHAR(10),
    CITYCODE VARCHAR(20)
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS DATE;
CREATE TABLE DATE(
    ID INT,
    DATEKEY DATE,
    YEAR INT,
```

```

        MONTH INT,
        DAY INT,
        WEEK INT,
        WEEKDAY INT
    )
    WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS PAYTYPE;
CREATE TABLE PAYTYPE(
    ID INT,
    CODE VARCHAR(10),
    TYPE VARCHAR(10),
    SIGNDATE DATE
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY REPLICATION;

DROP TABLE IF EXISTS SALES;
CREATE TABLE SALES(
    ID INT,
    POSID INT,
    STOREID INT,
    DATEKEY INT,
    PAYTYPE INT,
    TOTALAMOUNT NUMERIC(18,2),
    DISCOUNTAMOUNT NUMERIC(18,2),
    ITEMCOUNT INT,
    PAIDAMOUNT NUMERIC(18,2)
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY HASH(ID);

DROP TABLE IF EXISTS FLOW;
CREATE TABLE FLOW (
    ID INT,
    STOREID INT,
    DATEKEY INT,
    INFLOWVALUE INT
)
WITH (ORIENTATION = COLUMN, COMPRESSION=MIDDLE) DISTRIBUTE BY HASH(ID);

```

**Paso 4** Cree una tabla externa, que se utiliza para identificar y asociar los datos de origen en OBS.

#### AVISO

- `<obs_bucket_name>` indica el nombre del bucket OBS. Solo algunas regiones son compatibles. Los clústeres de GaussDB(DWS) no admiten el acceso entre regiones a los datos del bucket de OBS.
- En esta práctica, se utiliza como ejemplo **CN-Hong Kong** Escriba **dws-demo-ap-southeast-1** y reemplace `<Access_Key_Id>` y `<Secret_Access_Key>` por el valor obtenido en [Preparación](#).
- Si se muestra el mensaje "ERROR: schema "xxx" does not exist Position" al crear una tabla externa, el esquema no existe. Realice el paso anterior para crear un esquema.

```

create schema retail_obs_data;
set current_schema='retail_obs_data';
drop FOREIGN table if exists SALES_OBS;
CREATE FOREIGN TABLE SALES_OBS
(
    like retail_data.SALES
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/retail-data/sales',

```

```

        format 'csv',
        delimiter ',',
        access_key '<Access_Key_Id>',
        secret_access_key '<Secret_Access_Key>',
        chunksize '64',
        IGNORE_EXTRA_DATA 'on',
        header 'on'
    );

drop FOREIGN table if exists FLOW_OBS;
CREATE FOREIGN TABLE FLOW_OBS
(
    like retail_data.flow
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/retail-data/flow',
    format 'csv',
    delimiter ',',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on',
    header 'on'
);

drop FOREIGN table if exists BRAND_OBS;
CREATE FOREIGN TABLE BRAND_OBS
(
    like retail_data.brand
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/retail-data/brand',
    format 'csv',
    delimiter ',',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on',
    header 'on'
);

drop FOREIGN table if exists CATEGORY_OBS;
CREATE FOREIGN TABLE CATEGORY_OBS
(
    like retail_data.category
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/retail-data/category',
    format 'csv',
    delimiter ',',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on',
    header 'on'
);

drop FOREIGN table if exists DATE_OBS;
CREATE FOREIGN TABLE DATE_OBS
(
    like retail_data.date
)

```



```
SERVER gsmpp_server
OPTIONS (
  encoding 'utf8',
  location 'obs://<obs_bucket_name>/retail-data/date',
  format 'csv',
  delimiter ',',
  access_key '<Access_Key_Id>',
  secret_access_key '<Secret_Access_Key>',
  chunksize '64',
  IGNORE_EXTRA_DATA 'on',
  header 'on'
);

drop FOREIGN table if exists FIRM_OBS;
CREATE FOREIGN TABLE FIRM_OBS
(
  like retail_data.firm
)
SERVER gsmpp_server
OPTIONS (
  encoding 'utf8',
  location 'obs://<obs_bucket_name>/retail-data/firm',
  format 'csv',
  delimiter ',',
  access_key '<Access_Key_Id>',
  secret_access_key '<Secret_Access_Key>',
  chunksize '64',
  IGNORE_EXTRA_DATA 'on',
  header 'on'
);

drop FOREIGN table if exists PAYTYPE_OBS;
CREATE FOREIGN TABLE PAYTYPE_OBS
(
  like retail_data.paytype
)
SERVER gsmpp_server
OPTIONS (
  encoding 'utf8',
  location 'obs://<obs_bucket_name>/retail-data/paytype',
  format 'csv',
  delimiter ',',
  access_key '<Access_Key_Id>',
  secret_access_key '<Secret_Access_Key>',
  chunksize '64',
  IGNORE_EXTRA_DATA 'on',
  header 'on'
);

drop FOREIGN table if exists POS_OBS;
CREATE FOREIGN TABLE POS_OBS
(
  like retail_data.pos
)
SERVER gsmpp_server
OPTIONS (
  encoding 'utf8',
  location 'obs://<obs_bucket_name>/retail-data/pos',
  format 'csv',
  delimiter ',',
  access_key '<Access_Key_Id>',
  secret_access_key '<Secret_Access_Key>',
  chunksize '64',
  IGNORE_EXTRA_DATA 'on',
  header 'on'
);
```

```
drop FOREIGN table if exists SECTOR_OBS;
CREATE FOREIGN TABLE SECTOR_OBS
(
    like retail_data.sector
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/retail-data/sector',
    format 'csv',
    delimiter ',',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on',
    header 'on'
);

drop FOREIGN table if exists STORE_OBS;
CREATE FOREIGN TABLE STORE_OBS
(
    like retail_data.store
)
SERVER gsmpp_server
OPTIONS (
    encoding 'utf8',
    location 'obs://<obs_bucket_name>/retail-data/store',
    format 'csv',
    delimiter ',',
    access_key '<Access_Key_Id>',
    secret_access_key '<Secret_Access_Key>',
    chunksize '64',
    IGNORE_EXTRA_DATA 'on',
    header 'on'
);
```

**Paso 5** Copy and execute the following statements to import the foreign table data to the cluster:

```
insert into retail_data.store select * from retail_obs_data.STORE_OBS;
insert into retail_data.sector select * from retail_obs_data.SECTOR_OBS;
insert into retail_data.paytype select * from retail_obs_data.PAYTYPE_OBS;
insert into retail_data.firm select * from retail_obs_data.FIRM_OBS;
insert into retail_data.flow select * from retail_obs_data.FLOW_OBS;
insert into retail_data.category select * from retail_obs_data.CATEGORY_OBS;
insert into retail_data.date select * from retail_obs_data.DATE_OBS;
insert into retail_data.pos select * from retail_obs_data.POS_OBS;
insert into retail_data.brand select * from retail_obs_data.BRAND_OBS;
insert into retail_data.sales select * from retail_obs_data.SALES_OBS;
```

Se necesita algún tiempo para importar datos.

**Paso 6** Copie y ejecute la siguiente instrucción para crear la vista `v_sales_flow_details`:

```
set current_schema='retail_data';
CREATE VIEW v_sales_flow_details AS
SELECT
    FIRM.ID FIR MID, FIRM.NAME FIRNAME, FIRM. CITYCODE,
    CATEGORY.ID CATEGORYID, CATEGORY.NAME CATEGORYNAME,
    SECTOR.ID SECTORID, SECTOR.SECTORNAME,
    BRAND.ID BRANDID, BRAND.BRANDNAME,
    STORE.ID STOREID, STORE.STORENAME, STORE.RENTAMOUNT, STORE.RENTAREA,
    DATE.DATEKEY, SALES.TOTALAMOUNT, DISCOUNTAMOUNT, ITEMCOUNT, PAIDAMOUNT,
    INFLOWVALUE
FROM SALES
INNER JOIN STORE ON SALES.STOREID = STORE.ID
INNER JOIN FIRM ON STORE.FIR MID = FIRM.ID
INNER JOIN BRAND ON STORE.BRANDID = BRAND.ID
INNER JOIN SECTOR ON BRAND.SECTORID = SECTOR.ID
INNER JOIN CATEGORY ON SECTOR.CATEGORYID = CATEGORY.ID
```

```
INNER JOIN DATE ON SALES.DATEKEY = DATE.ID  
INNER JOIN FLOW ON FLOW.DATEKEY = DATE.ID AND FLOW.STOREID = STORE.ID;
```

----Fin

## Paso 2: Realización del análisis del estado de las operaciones

A continuación se utiliza la consulta estándar de información de venta minorista de grandes almacenes como ejemplo para demostrar cómo realizar una consulta de datos básica en GaussDB(DWS).

Antes de consultar datos, ejecute el comando **Analyze** para generar estadísticas relacionadas con la tabla de la base de datos. Los datos estadísticos se almacenan en la tabla del sistema PG\_STATISTIC y son útiles cuando se ejecuta el planificador, que le proporciona un plan de ejecución de consultas eficiente.

Los siguientes son ejemplos de consulta:

- **Consultar los ingresos totales por ventas de cada tienda**

Copie y ejecute las siguientes instrucciones para consultar el volumen de negocios total de cada tienda:

```
set current_schema='retail_data';  
SELECT DATE_TRUNC('month',datekey)  
AT TIME ZONE 'UTC' AS __timestamp,  
SUM(paidamount)  
AS sum__paidamount  
FROM v_sales_flow_details  
GROUP BY DATE_TRUNC('month',datekey) AT TIME ZONE 'UTC'  
ORDER BY SUM(paidamount) DESC;
```

- **Consultar los ingresos por ventas y la relación precio-alquiler de cada tienda**

Copie y ejecute la siguiente instrucción para consultar los ingresos por ventas y la relación precio-alquiler de cada tienda:

```
set current_schema='retail_data';  
SELECT firname AS firname,  
storename AS storename,  
SUM(paidamount)  
AS sum__paidamount,  
AVG(RENTAMOUNT)/SUM(PAIDAMOUNT)  
AS rentamount_sales_rate  
FROM v_sales_flow_details  
GROUP BY firname, storename  
ORDER BY SUM(paidamount) DESC;
```

- **Analizar los ingresos por ventas de cada provincia**

Copie y ejecute la siguiente declaración para analizar y consultar los ingresos por ventas de todas las provincias:

```
set current_schema='retail_data';  
SELECT citycode AS citycode,  
SUM(paidamount)  
AS sum__paidamount  
FROM v_sales_flow_details  
GROUP BY citycode  
ORDER BY SUM(paidamount) DESC;
```

- **Analizar y comparar la relación precio-alquiler y la tasa de conversión del flujo de clientes de cada tienda**

```
set current_schema='retail_data';  
SELECT brandname AS brandname,  
firname AS firname,  
SUM(PAIDAMOUNT)/AVG(RENTAREA) AS sales_rentarea_rate,  
SUM(ITEMCOUNT)/SUM(INFLOWVALUE) AS poscount_flow_rate,  
AVG(RENTAMOUNT)/SUM(PAIDAMOUNT) AS rentamount_sales_rate
```

```
FROM v_sales_flow_details  
GROUP BY brandname, firname  
ORDER BY sales_rentarea_rate DESC;
```

- **Análisis de marcas en la industria minorista**

```
set current_schema='retail_data';  
SELECT categoryname AS categoryname,  
brandname AS brandname,  
SUM(paidamount) AS sum__paidamount  
FROM v_sales_flow_details  
GROUP BY categoryname,  
brandname  
ORDER BY sum__paidamount DESC;
```

- **Consultar información de ventas diaria de cada marca**

```
set current_schema='retail_data';  
SELECT brandname AS brandname,  
DATE_TRUNC('day', datekey) AT TIME ZONE 'UTC' AS __timestamp,  
SUM(paidamount) AS sum__paidamount  
FROM v_sales_flow_details  
WHERE datekey >= '2016-01-01 00:00:00'  
AND datekey <= '2016-01-30 00:00:00'  
GROUP BY brandname,  
DATE_TRUNC('day', datekey) AT TIME ZONE 'UTC'  
ORDER BY sum__paidamount ASC  
LIMIT 50000;
```

# 5 Paso 5: Visualización de otros documentos y eliminación de recursos

---

## Visualización de otros documentos relevantes

Después de realizar los pasos de las secciones anteriores, puede consultar la documentación listada de la siguiente manera para obtener más información acerca de GaussDB(DWS):

- **Guía de administración del servicio de almacén de datos (DWS):** Esta guía proporciona información detallada sobre los conceptos y operaciones relacionados con la creación, administración, supervisión y conexión de clústeres.
- **Guía del desarrollador de Data Warehouse Service (DWS):** Esta guía proporciona información completa y detallada sobre cómo crear, administrar y consultar bases de datos de GaussDB(DWS), que abarca la sintaxis SQL, la gestión de usuarios y la importación y exportación de datos.

## Eliminación de recursos

Después de realizar los pasos de "Preparación", si no necesita utilizar los datos de muestra, clústeres, ECS o VPC, elimine los recursos para que las cuotas de servicio no se desperdicien ni se ocupen.

### Paso 1 Eliminar un clúster de GaussDB(DWS).

En la consola de gestión de GaussDB(DWS), haga clic en **Clusters**, busque la fila que contiene **dws-demo** en la lista de clústeres y elija **More > Delete**. En el cuadro de diálogo que se muestra, seleccione **Release the EIP bound with the cluster** y haga clic en **OK**.

Si el clúster que se va a eliminar utiliza un grupo de seguridad creado automáticamente que no lo utilizan otros clústeres, el grupo de seguridad se elimina automáticamente cuando se elimina el clúster.

### Paso 2 Eliminar una subred. Antes de eliminar la subred, asegúrese de que no está vinculada a otros recursos.

Inicie sesión en la consola de gestión de VPC. En el árbol de navegación de la izquierda, haz clic en **Virtual Private Cloud**. En la lista VPC, haga clic en **vpc-dws**. En la lista de subred, busque la fila que contiene **subnet-dws** y haga clic en **Delete**.

### Paso 3 Eliminar una VPC. Antes de eliminar la VPC, asegúrese de que no está vinculada a otros recursos.

Inicie sesión en la consola de gestión de VPC, busque la fila que contiene **vpc-dws** en la lista de VPC y haga clic en **Delete**.

Para obtener más información, consulte "VPC y subred > Eliminar una VPC" en la *Guía del usuario de la nube privada virtual*.

----**Fin**