#### **Data Warehouse Service**

## **Quick Start**

 Issue
 13

 Date
 2025-03-24





#### Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2025. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Cloud Computing Technologies Co., Ltd.

#### **Trademarks and Permissions**

NUAWEI and other Huawei trademarks are the property of Huawei Technologies Co., Ltd. All other trademarks and trade names mentioned in this document are the property of their respective holders.

#### Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

## **Contents**

1 Quickly Creating a GaussDB(DWS) Cluster and Importing Data for Query	1
2 GaussDB(DWS) Data Development SQL Statements	. 15
3 Getting Started with GaussDB(DWS)	.17

# Quickly Creating a GaussDB(DWS) Cluster and Importing Data for Query

#### Scenario

GaussDB(DWS) is a fully-managed and enterprise-level cloud data warehouse service. It is O&M-free, compatible with the PostgreSQL ecosystem, and supports online cluster scale-out and efficient loading of multiple data sources. It helps enterprises efficiently analyze and monetize massive amounts of data online.

This section describes how to create a GaussDB(DWS) cluster with three nodes and import CSV data from an OBS bucket to perform simple data analysis and querying.

#### NOTICE

The GaussDB(DWS) cluster does not support cross-region access to OBS bucket data. Before creating a GaussDB(DWS) cluster, ensure that the cluster and OBS bucket are in the same region. This document uses CN-Hong Kong as an example.

- Billing Mode: Pay-per-use
- Version: Computing In-Memory(CIM)
- Nodes: 3 nodes
- Cluster connection mode: SQL editor
- Storage capacity: 20 GB/node
- Data source: CSV data in the OBS bucket
- Cluster version: 8.1.3.*x*

#### Procedure

Table 1-1	Procedure
-----------	-----------

Step	Description
Preparations	Register a Huawei account, subscribe to Huawei Cloud, complete real-name authentication, top up the account, and plan and create VPC and subnet resources.
Step 1: Creating a GaussDB(DWS) Cluster	Create a three-node GaussDB(DWS) cluster and configure information such as the cluster node quantity, storage capacity, and cluster version.
Step 2: Uploading Data to OBS	Create an OBS bucket and upload the local CSV data to the bucket.
Step 3: Connecting to the GaussDB(DWS) Cluster and Importing Data for Analysis	Use the SQL editor to connect to the GaussDB(DWS) database, create an OBS foreign table, import OBS bucket data to the GaussDB(DWS) cluster through the OBS foreign table, and perform simple SQL queries.

#### Preparations

- Register a GaussDB(DWS) account, complete the real-name authentication, and check the account status before using GaussDB(DWS). The account cannot be in arrears or frozen.
- Obtain the AK/SK of the account. For details, see Access Keys.
- Create a VPC and subnet resources.

#### Step 1: Creating a GaussDB(DWS) Cluster

Step 1 Log in to the GaussDB(DWS) console.

**Step 2** Configure the parameters according to **Table 1-2**.

Paramete r	Example Value	Description
Region	Select the <b>CN-Hong Kong</b> region. <b>NOTE</b> This guide uses CN-Hong Kong as an example. If you want to select another region, ensure that other cloud services are in the same region.	For lower network latency and quicker resource access, select the nearest region. After a GaussDB(DWS) instance is created, the region cannot be changed. Exercise caution when selecting a region. For more information, see <b>Regions</b> and AZs.
Billing Mode	Pay-per-use	Yearly/Monthly and pay-per-use billing modes are supported. For long-term usage, it is recommended to opt for the yearly or monthly billing mode to avail of a greater discount. For more information, see <b>Billing</b> <b>Overview</b> .
AZ	Choose AZ1 to create a single-AZ cluster. (If the flavor is sold out, select another AZ.)	<ul> <li>Single AZ: All nodes in a cluster are deployed in the same AZ.</li> <li>Multiple AZs: To achieve HA across multiple AZs, deploy cluster nodes in different AZs. For services requiring even higher HA, a multi-AZ setup is recommended.</li> </ul>

#### Table 1-2 Parameter configurations

#### Figure 1-1 Basic configurations

DWS Cluster	
Current Nodes	3
Total Capacity	300
Node Specifications	Standard 3.0   4 vCPUs   16 GB Memory   100 GB Common I/O
Billing Mode	Pay-per-use
★ Plan Type	Periodicity One-time
* Cluster Name	Select a logical cluster
	The logical cluster is mainly used for computing acceleration.
Bind Main Logical Cluster	-Select-
Bind User	-Select-
	After a user is bound, the tables created by the user are still stored in the original logical cluster, and the computing logic of the user is switched to the current cluster for execution
Nodes	-         3         +
Time Range	Start Date Find Date (***)
Time Kange (j	
Period Type	Every Week      Every Month
* Time Configuration(UTC)	Creation Completion Time Sunday × v 01.00 O Deletion Start Time Thursday × v 04.00 O + -
	Creation Completion Time Friday × v 04:00 ⊙ Deletion Start Time Saturday × v 03:00 × + -

#### **Step 3** Table 1-3 lists the required data warehouse parameters.

Paramete	Example Value	Description
Version	Computing In-	Computing In-Memory(CIM): Data is stored     an local disks of compute nodes
	Memory(CIM)	<ul> <li>Decoupled Storage and Compute: Local DN disks are used only for data cache and metadata storage, and user data is stored on OBS.</li> </ul>
Storage Type	Cloud SSD	<ul> <li>Cloud SSD: SSD EVS disks serve as data storage media, offering flexible storage capacity and supporting disk scale-out.</li> </ul>
		• Local SSD: Local disks in ECS flavors serve as data storage media, offering fixed capacity and superior performance. However, disk scale-out is not supported.
		For more information, see What Are the Differences Between Cloud SSDs and Local SSDs?
Deployme nt Mode	Cluster	• <b>Cluster</b> : A cluster contains at least three ECSs. Compute node DNs form a security ring to ensure high reliability in active/ standby mode.
		• <b>Single-node</b> : A single ECS is used. It has a lower cost, but data reliability cannot be guaranteed.
Node Flavor	dwsx2.xlarge NOTE If this flavor is sold out, select other AZs or flavors.	Node specifications measure the computing power of GaussDB(DWS), including vCPUs and memory. When selecting node specifications, consider the service requirements. Typically, higher query performance requires larger node specifications.
		For more information, see <b>Data Warehouse</b> Flavors.
Hot Storage	20 GB per node	Available storage capacity of each node. The page displays the actual storage capacity for the service, such as 20 GB per node with a total of 3 nodes resulting in a total service data capacity of 60 GB. Customers with a large amount of service data should plan their storage capacity accordingly.
Nodes	3	Number of cluster nodes. The number of nodes ranges from 3 to 256.

 Table 1-3 Data warehouse parameters

Data Warehouse Configuration		
Version		
Computing In-Memory(CIM) Decoupled Storage And Compute		
Storage Type		
Cloud SSD Local SSD ③		
Deployment Mode		
Cluster Single-node		
Node Flavor		
Flavor Name	vCPUs   Memory	Hot Storage
dwsx2.xlarge     Trial only. Not for production	4 vCPUs   32 GB	20~2000 GB SSD
dwsx2.h.2xlarge.4.c?n	8 vCPUs   32 GB	100~4000 GB SSD
dwsx2.2xlarge	8 vCPUs   64 GB	100~4000 GB SSD
dwsx2.8xlarge	32 vCPUs   256 GB	100~16000 GB SSD
dwsx2.16xJarge	64 vCPUs   512 GB	100~32000 GB SSD
Hot Storage		
- 20 + GB per node		
Nodes		
- 3 +		
You can still create 1,024.00 nodes. You can apply for a higher quota as needed.		
Total		
dwsx2.xlarge   12 vCPUs   96 GB Memory   60.00 GB hot storage		

Figure 1-2 Data warehouse configuration

#### Step 4 Click Next: Configure Network.

**Step 5** Configure the parameters according to **Table 1-4**.

Paramete r	Example Value	Description
VPC	vpc-default	VPC where GaussDB(DWS) is located.
Subnet	subnet- default(192.168.0. 0/24)	VPC subnet.
Security Group	Automatic creation	You can select <b>Automatic creation</b> or choose a created security group.
		The default option is <b>Automatic creation</b> , which means the system will create a security group automatically if you do not change the setting.
		The default security group has two rules: outbound direction allows all access requests, while inbound direction only opens the database port <b>8000</b> that the user has set for the GaussDB(DWS) cluster.
EIP	Do not use	If the client connected to GaussDB(DWS) and GaussDB(DWS) are not in the same VPC, use an EIP.

**Table 1-4** Network configuration parameters

Paramete r	Example Value	Description
ELB	Do not use	If the private IP address or EIP of a CN is used to connect to a cluster, the failure of this CN will lead to cluster connection failure.
		ELB health checks enable quick forwarding of CN requests to healthy nodes in a cluster, minimizing access faults in case of a faulty CN. This ensures that workload is immediately shifted to a healthy node.
		For more information, see Associating and Disassociating ELB.

#### Figure 1-3 Network configuration

Network					
VPC					
vpc-default	~	Q View VPC 🧿			
Subnet					
subnet-default(192.168.0	).0/24) 🗸	Q 0			
Security Group					
Automatic creation	× )	Q 0			
Automatically enables the p	port's access permission i	n the inbound rule.			
Others					
EIP					
Do not use	Buy now	Specify			
A cluster without an EIP ca	nnot be accessed over the	e Internet, but can be ac	cessed through an ECS (	deployed on a private ne	etwork.
ELB					
Do not use	Use existing				
ELB is a service that auton Clusters that do not use EL	natically distributes incomi .B cannot distribute traffic	ng traffic across multiple and are vulnerable to si	backend cloud servers b Igle points of failure (SP	based on a specified forv OFs). Leam more	warding policy.

## **Step 6** Click **Next: Configure Advanced Settings** to access advanced configurations. **Table 1-5** lists the required parameters.

Table 1-5 Advanced settings	
-----------------------------	--

Paramete r	Example Value	Description
Cluster Name	dws-demo	Cluster name, which is user-defined and can be modified later.
Cluster Version	Use the recommended version.	Each version supports different database functions. By default, the recommended version is selected. To use the functions of a later version, select a later version.
		<b>NOTE</b> On the whitelist customer page, other cluster versions are displayed. Select a cluster version based on the site requirements.
		For details about the functions supported by each cluster version, see <b>Product Notice</b> .
Administr ator Account	dbadmin	Username of the database system administrator. Retain the default value. The database system administrator is assigned the <b>SYSADMIN</b> attribute by default. If the separation of permissions feature is not enabled, the administrator has the highest database permissions. This allows them to plan database objects, create databases and users, and authorize user roles. For more information, see <b>Database Users</b> .
Administr ator Password	-	Password of the database system administrator.
Confirm Password	-	Enter the password again for confirmation.
Database Port	8000	Port for the client to connect to the GaussDB(DWS) database. Retain the default value <b>8000</b> .
Enterprise Project	default	This parameter is displayed only when an enterprise account is used to create a GaussDB(DWS) instance.
		It enables unified management of cloud resources by project.

Paramete r	Example Value	Description	
Advanced Settings	Default	You can select <b>Default</b> or <b>Custom</b> . If you select <b>Custom</b> , the following advanced settings are supported:	
		Backup Device: Choose OBS or NFS.	
CN     wh		• <b>CNs</b> : It is recommended to use multiple CNs when deploying a large-scale cluster.	
		• <b>Tag</b> : In GaussDB(DWS), you can assign tags to items like project names, service types, and background information after creating a cluster. If you use tags in other cloud services, you are advised to create the same tag key-value pairs for cloud resources used by the same business to keep consistency.	
		• Encrypt DataStore: If this function is enabled, Key Management Service (KMS) encrypts the cluster and the cluster's snapshot data.	

#### Figure 1-4 Advanced settings

Management
Cluster Name
dws-demo
Enter 4 to 64 characters starting with a letter or Chinese character, and cannot exceed 64 bytes, one Chinese character is 3 bytes, and other characters are 1 byte. Only Chinese character letters digits hyphens (-) and underscores (_) are allowed.
Cluster Version
Recommended 8.3.0.108
Release Notes Version Policies
Default Database
gaussdb
A designation of Associated
veguminn Fohre fil high characters startinn with a lowernase letter or an underscore ( ) Only lowernase letters dinits and underscore ( ) are allowed
Aministation Association
Confirm Password
Debase Red
a course of the second se
I m 2 one
UIC+08:00 (Beijing, Chongqng, Hong V
Advanced
Enlarprise Project
default V Q (() View enterprise projects.
Advanced Settings
Default Custom

**Step 7** Click **Next: Confirm**, confirm the configuration, and click Buy Now.

Cluster creation takes about 10 minutes. Clusters in the **Available** state are ready for use.

----End

#### Step 2: Uploading Data to OBS

**Step 1** Create an OBS bucket. For details, see **Creating a Bucket**.

Set key parameters as follows and retain the default values for other parameters.

Cross-region access to OBS bucket data is not supported. The OBS bucket must be created in the same region as GaussDB(DWS). Select CN-Hong Kong.

- **Region**: Select the **CN-Hong Kong** region.
- **Bucket Name: dws-obs01** (If a conflict occurs, modify the number in the name to 02, 03, or a higher number)
- Bucket Policy: Private.
- Step 2 After the OBS bucket is created, click the name of the newly created dws-obs01 bucket on the bucket list. In the navigation pane, choose Objects and click Create Folder. Set the folder name to input\_data and click OK.

Figure 1-5 Creating a folder

< 🥌 dws-obs99	Standard   Multi-AZ storage   CN North-Beijing4   Creat	ed Aug 01, 2024 15:47:32 GMT+08:00
Overview	Objects 🗇	Create Folder e
Objects Metrics NEW	Objects Deleted Objects Fragment	Folder Name Input_data Naming rules : - You can create folders with a sincle level or multiple levels.
Basic Configurations	You can use OBS Browser+ to move an object to any Upload Object Create Folder De	<ul> <li>The name of a single-level folder cannot contain the following characters: \: * ? * &lt; &gt;  </li> <li>The name cannot start or end with a period (.) or a slash (/).</li> <li>Use single slashes (/) to separate levels of a folder.</li> <li>The absolute path of the folder cannot exceed 1023 characters.</li> </ul>
Tagging	C Enter an object name prefix.     Name	- Cannot contain two or more consecutive slashes (r).
Permissions		

Step 3 Create three CSV data files on the local PC. The file names (including the file name extension) are product\_info0.csv, product\_info1.csv, and product\_info2.csv.

Figure 1-6 Creating a test.csv file

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

**Step 4** Use a text editor to open the three CSV files and copy the following content to each CSV file.

#### 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-I",2017-09-22,"A","2017 pants 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 new pants 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."

- **Step 5** Use a text editor to convert the encoding format to UTF-8 and save the file.
- **Step 6** Return to the OBS console and click the name of the newly created **dws-obs01** bucket in the bucket list.
- **Step 7** In the navigation pane, choose **Objects** and click the **input\_data** folder.
- Step 8 Click Upload Object, and then click Add File.
- **Step 9** Upload the **product\_info0.csv**, **product\_info1.csv**, and **product\_info2.csv** files to the OBS bucket.

----End

## Step 3: Connecting to the GaussDB(DWS) Cluster and Importing Data for Analysis

- **Step 1** Go to the GaussDB(DWS) console.
- Step 2 In the navigation tree on the left, choose Data > SQL Editor and click Add Data Source.
- **Step 3** Enter the database connection information.
  - **Cluster**: Select the created **dws-demo** cluster.
  - **Database**: Retain the default value **gaussdb**.
  - Data Source: Enter dws-demo-01.
  - Username: Enter dbadmin.
  - **Password**: Enter the password set during GaussDB(DWS) instance creation.

#### **Step 4** Select **Remember password** and click **Test Connection**.

DWS	Q	Data 🖸 Usage Guidelines	Unnamed •	+		
Dashboard		Data Source Scripts (3) Configure	🗋 Import 🔒 Save	Add Data Sou	rce	×
Data		2 + Add Data Source	·-Select ✓	* Cluster	dws-demo 🗸	
SQL Editor		Please enter the data source name Q		+ Database	naussilan X	ā
Data Integration	$\sim$	Data ware Custom		A Database	gaussub	0
Cluster		dws-demo(dbadmin) >		* Data Source	dws-demo-01 ×	
Dedicated Cluster				* Username	dbadmin ×	
Management	~			+ Password		
Wanagement					Remember password	
					This password is only used to create the dat obtain the data source connection, and use editor function.	a source, the WEB-SQL
		<	Line 1, column 1 Runnir	<ul> <li>Connection tes</li> </ul>	t succeeded.	
			SQL execution records		C	ancel OK
			Enter a keyword.			

Figure 1-7 Database connection information

Step 5 Click OK.

The green icon effore the data source name in the following figure indicates that the database is connected. The right pane is the SQL editor window, where you can write and execute SQL statements.

#### Figure 1-8 Successful database connection

DWS	Q	Data 🖸 Usage Guidelines	Unnamed •	+	
Dashboard		Data Source Scripts (3) Configure	Import Save dws-demo-01	O         Running:         O         Stop         Q.         Refresh         I         Image: The store in the st	Execution Plan ~
Data SQL Editor Data Integration Cluster Dedicated Cluster	~ 1	+ Add Data Source Please enter the data source name Q Data ware Custom © dws-demo(dbadmin) > © dws-demo.01 >	•		
Management	~		Line 1, column 1 Running	LINIT: 1000 ~	

#### D NOTE

If the connection session times out or the browser is refreshed,  $\begin{times}{l} \end{times}\end{times}$  will appear grayed out, indicating that the connection has been disconnected. To reconnect to the database, double-click the data source name.

		+ Add Data	a Source
Data		Please enter the data source name	Q
SQL Editor		Data ware Custom	
Data Integration	$\sim$		
Cluster		lws-demo(dbadmin)	>
Dedicated Cluster		😑 dws-demo-01	>

**Step 6** Copy the following SQL statement to the SQL editor window and replace the OBS bucket name, AK value, and SK value with the actual values.

In this document, the OBS bucket name is **dws-obs01**. Obtain the AK and SK from **Access Keys**.

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 bucket name/input\_data/', FORMAT 'CSV', DELIMITER ',', ENCODING 'utf8', HEADER 'false', ACCESS\_KEY 'AK value', SECRET\_ACCESS\_KEY 'SK value', FILL\_MISSING\_FIELDS 'true', IGNORE\_EXTRA\_DATA 'true' ) READ ONLY LOG INTO product\_info\_err PER NODE REJECT LIMIT 'unlimited';

Figure 1-9 Creating an OBS foreign table

Unna	med • +							
는 In	mport 🖻 Save   📀 Running: 💿 Stop 📿 Refresh   📥 Format 🔟 Clear   🗐 SQL diagnostics 🗮 Execution Plan							
😑 dw	s-demo-01 v 🔄 gaussdb v 🖉 🖾 public v 🖓 🖓 💿 UTC v							
1	CREATE FOREIGN TABLE product info ext							
2	(							
3	product_price integer not null,							
4	<pre>product_id char(30) not null,</pre>							
5	product_time date,							
6	<pre>product_level char(10),</pre>							
7	<pre>product_name varchar(200),</pre>							
8	<pre>product_type1 varchar(20),</pre>							
9	<pre>product_type2 char(10),</pre>							
10	product_monthly_sales_cnt integer,							
11	product_comment_time date,							
12	product_comment_num integer,							
13	<pre>product_comment_content varchar(200)</pre>							
14	)							
15	SERVER gsmpp_server							
16	OPTIONS							
17	(							
18	LOCATION 'obs://d000000)/input_data/',							
19	FORMAT 'CSV',							
20	DELIMITER ',',							
21	ENCODING 'utf8',							
22	HEADER 'false',							
23	ACCESS_KEY '							
24	SECRET_ACCESS_KEY '							
25	FILL_MISSING_FIELDS 'true',							
26	IGNORE_EXTRA_DATA 'true'							
27	)							
28	READ ONLY							
29	LOG INTO product_info_err							
30	<pre>PER NODE REJECT LIMIT 'unlimited';</pre>							
Line 18, column	41 Running: LIMIT: 1000 V							

**Step 7** Verify that the SQL statement is correct and click **Running:**. The OBS foreign table is created.

Figure 1-10 Successful creation

Line 18, column 41 Running: LIMIT: 1000 V		Shift+Alt+	F: Format 🕥 overla	lay mode
SQL execution records Notice				
Enter a keyword.				Q
Result Information	SQL Details	Status	Times	
Affected Rows :0	CREATE FOREIGN TABLE product_info_ext ( product_price integer not null, product_id char(30) not null, product	CRun successfully	196ms	

**Step 8** Copy the following SQL statement and run it in the SQL window to create an ordinary table.

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\_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);

**Step 9** Copy the following SQL statement and execute it in the SQL window. Run the **INSERT** statement to import the OBS foreign table data to the GaussDB(DWS) database.

INSERT INTO product\_info SELECT \* FROM product\_info\_ext;

**Step 10** Copy the following SQL statements and run them in the SQL window to perform the **VACUUM** and **ANALYZE** operations.

For details about the syntax, see the related SQL syntax section.

VACUUM FULL product\_info; ANALYZE product\_info;

**Step 11** Copy the following SQL statement and run it in the SQL window. The data details are successfully queried.

SELECT \* FROM product\_info;

#### Figure 1-11 Successful query

Unnamed •	+									
🗋 Import 🖹 Save	Running:	Stop 🔾 Refresh	🛆 Format 🌐 Clea	r 📄 😒 SQL diagnostic	s 🗮 Execution Plan ~				-	Common Europhone
😑 dws-demo-01 🛛 🗸	📜 gaussdb	✓ Se public	~ Q 💽 U	rc v						Common Functions
1 SELECT * FROM	product_info;									
Line 1, column 28 Runn	ing: LIMIT: 100	o ~							Shift+Alt+F: Format	overlay mode
SQL execution records	Notice R	esult1 ×								
table Chart	Export ~						Copy Co	blumn	V Columns produ	ct_pr × 0 ∨
product_price	product_id	product_time	product_level	product_name	product_type1	product_type2	product_monthly_sa	product_comment_t	product_comment	product_com
200	BCQP-E	2017-09-04 00:00:00	В	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	м	85	2017-09-14 00:00:00	22	It's really amazing
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	в	2017 Women's Cotto	red	м	112	(Null)	(Null)	
50	LPEC-N	2017-09-25 00:00:00	в	2017 dress Women	red	м	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	м	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	QQPX-R	2017-09-02 00:00:00	в	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	в	2017 Jeans Women	blue	м	123	2017-09-06 00:00:00	120	good.

----End

## **2** GaussDB(DWS) Data Development SQL Statements

This section describes some basic SQL operations of the GaussDB(DWS) database.

#### Creating, Viewing, and Deleting a Database

- Run the **CREATE DATABASE** statement to create a database. CREATE DATABASE test\_db ENCODING 'UTF8' template = template0;
- Query the database list using the \l meta-command.
- Querying the database list using the **PG\_DATABASE** system catalog SELECT datname FROM pg\_database;
- Run the DROP DATABASE statement to delete a database. DROP DATABASE test\_db;

#### Creating, Viewing, Modifying, and Deleting a Table

- Run the CREATE TABLE statement to create a table.
   CREATE TABLE customer\_t1(id INT, name CHAR(40),age TINYINT);
- Use the PG\_GET\_TABLEDEF() function to view the table creation statement. Replace customer\_t1 with the name of the table to be queried.
   SELECT \* FROM PG\_GET\_TABLEDEF('customer\_t1');
- Run the **ALTER TABLE** statement to modify a table.

Add a column: ALTER TABLE customer t1 ADD (address VARCHAR(100));

Delete a column:

ALTER TABLE customer\_t1 DROP COLUMN address;

Modify a column type:

ALTER TABLE customer\_t1 MODIFY age INTEGER NOT NULL;

• Run the **DROP TABLE** statement to delete a table. DROP TABLE customer\_t1;

#### Creating, Viewing, and Deleting Indexes

• Run the **CREATE INDEX** or **ALTER TABLE** statement to create a common index.

CREATE INDEX c\_id\_index on customer\_t1(id); ALTER TABLE customer\_t1 ADD INDEX c\_id\_index (id);

- Use the **PG\_INDEXES** system catalog to view all indexes in a table. SELECT \* FROM pg\_indexes WHERE tablename = 'customer\_t1';
- Run the **ALTER TABLE** or **DROP INDEX** statement to delete an index. DROP INDEX c\_id\_index; ALTER TABLE customer\_t1 DROP INDEX c\_id\_index;

#### Adding, Deleting, and Modifying Table Data

- Run the INSERT INTO statement to insert table data. INSERT INTO customer\_t1 VALUES(1001,'user1',22);
- Run the SELECT statement to query table data. SELECT \* FROM customer\_t1;
- Run the UPDATE statement to update table data. UPDATE customer\_t1 SET id = 1009 WHERE id = '1001';
- Use the DELETE statement to delete table data. DELETE FROM customer\_t1 WHERE id = '1009';

# **3** Getting Started with GaussDB(DWS)

After creating a cluster, you can use the practice cases provided by GaussDB(DWS) for your services.

Practice		Description
Data Import and Export	Importing Data from OBS to a Cluster	This practice demonstrates how to upload sample data to OBS and import OBS data to the target table on GaussDB(DWS), helping you quickly learn how to import data from OBS to a GaussDB(DWS) cluster.
		CSV, ORC, PARQUET, CARBONDATA, or JSON format from OBS to a GaussDB(DWS) cluster for query.
	Using GDS to Import Data from a Remote Server	This practice demonstrates how to use General Data Service (GDS) to import data from a remote server to GaussDB(DWS).
		GaussDB(DWS) allows you to import data in TXT, CSV, or FIXED format.

#### Table 3-1 Best practices

Practice		Description
	Exporting ORC Data to MRS	GaussDB(DWS) allows you to export ORC data to MRS using an HDFS foreign table. You can specify the export mode and export data format in the foreign table. Data is exported from GaussDB(DWS) in parallel using multiple DNs and stored in HDFS. In this way, the overall export performance is improved.
Data Migration	Migrating Data From Oracle to GaussDB(DWS)	This tutorial demonstrates how to migrate Oracle data to GaussDB(DWS).
	Using a Flink Job of DLI to Synchronize Kafka Data to a GaussDB(DWS) Cluster in Real Time	This practice demonstrates how to use DLI Flink jobs to synchronize consumption data from Kafka to GaussDB(DWS) in real time.
		This practice takes about 90 minutes. The cloud services used in this practice include Virtual Private Cloud (VPC) and subnets, Elastic Cloud Server (ECS), Object Storage Service (OBS), Distributed Message Service (DMS) for Kafka, Data Lake Insight (DLI), and GaussDB(DWS).

Practice		Description
Optimization Table	Table Optimization         Practices	In this practice, you will learn how to optimize the design of your tables. You will start by creating tables without specifying their storage mode, distribution key, distribution mode, or compression mode. Load test data into these tables and test system performance. Then, follow excellent practices to create the tables again using new storage modes, distribution keys, distribution modes, and compression modes. Load the test data and test performance again. Compare the two test results to find out how table design affects the storage space, and the loading and query performance of the tables. Estimated time: 60 minutes
Advanced Features	Best Practices of Hot and Cold Data Management	In massive big data scenarios, with the growing of data, data storage and consumption increase rapidly. The need for data may vary in different time periods, therefore, data is managed in a hierarchical manner, improving data analysis performance and reducing service costs. In some data usage scenarios, data can be classified into hot data and cold data by accessing frequency.

Practice		Description
	Best Practices for Automatic Partition Management	For partition tables whose partition columns are time, the automatic partition management function can be added to automatically create partitions and delete expired partitions, reducing partition table maintenance costs and improving query performance. To facilitate data query and maintenance, the time column is often used as the partition column of a partitioned table that stores time-related data, such as e- commerce order information and real-time IoT data. When the time-related data is imported to a partitioned table, the table should have partitions of the corresponding time ranges. Common partition tables do not automatically create new partitions. Therefore, maintenance personnel need to periodically create new partitions and delete expired partitions, leading to increased O&M costs.
		Addressing this, GaussDB(DWS) introduces the automatic partition management feature. You can set the table-level parameters <b>period</b> and <b>ttl</b> to enable the automatic partition management function, which automatically creates partitions and deletes expired partitions, reducing partition table maintenance costs and improving query performance.

Practice		Description
Database Management	Best Practices of Resource Management	This practice demonstrates how to use GaussDB(DWS) for resource management, helping enterprises eliminate bottlenecks in concurrent query performance. SQL jobs can run smoothly without affecting each other and consume less resources than before.
	Excellent Practices for SQL Queries	Based on a large number of SQL execution mechanisms and practices, we can optimize SQL statements following certain rules to more quickly execute SQL statements and obtain correct results.
	Excellent Practices for Data Skew Queries	<ul> <li>This practice includes the following storage skew cases:</li> <li>Real-Time Detection of Storage Skew During Data Import</li> <li>Quickly Locating the Tables That Cause Data Skew</li> </ul>
	Best Practices of User Management	A GaussDB(DWS) cluster mainly consists of system administrators and common users. This practice describes the permissions of system administrators and common users and describes how to create users and query user information.
	Viewing Table and Database Information	<ul> <li>This practice demonstrates some basic database query cases:</li> <li>Querying Table Information</li> <li>Querying the Table Size</li> <li>Querying Database Information</li> <li>Querying the Database Size</li> </ul>

Practice		Description
Sample Data Analysis	Checkpoint Vehicle Analysis	This practice demonstrates the analysis of passed vehicles at traffic checkpoints. In this practice, 890 million pieces of data from traffic checkpoints are loaded to a single database table on GaussDB(DWS) for performing accurate query and fuzzy query. It is an example of high-performance query of historical data on GaussDB(DWS).
	Supply Chain Requirement Analysis (TPC-H Data Set)	This practice describes how to load the sample data set from OBS to a data warehouse cluster and perform data queries. This example comprises multi-table analysis and theme analysis in the data analysis scenario.
	Operations Status Analysis of a Retail Department Store	In this practice, the daily business data of each retail store is loaded from OBS to the corresponding table in the data warehouse cluster for summarizing and querying KPIs. This data includes store turnover, customer flow, monthly sales ranking, monthly customer flow conversion rate, monthly price-rent ratio, and sales per unit area. This example demonstrates the multidimensional query and analysis of GaussDB(DWS) in the retail scenario.

Practice		Description
Data Security	Encrypting and Decrypting Data Columns	Data encryption is widely used in information systems to prevent unauthorized access and data leakage. As the core of the information system, the GaussDB(DWS) data warehouse also provides data encryption functions, including transparent encryption and encryption using SQL functions. This section describes SQL function encryption.