

Data Warehouse Service

Quick Start

Edição 01
Data 2026-09-29



Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2026. Todos os direitos reservados.

Nenhuma parte deste documento pode ser reproduzida ou transmitida em qualquer forma ou por qualquer meio sem consentimento prévio por escrito da Huawei Cloud Computing Technologies Co., Ltd.

Marcas registradas e permissões



HUAWEI e outras marcas registradas da Huawei são marcas registradas da Huawei Technologies Co., Ltd. Todas as outras marcas registradas e os nomes registrados mencionados neste documento são propriedade dos seus respectivos detentores.

Aviso

Os produtos, os serviços e as funcionalidades adquiridos são estipulados pelo contrato estabelecido entre a Huawei Cloud e o cliente. Os produtos, os serviços e as funcionalidades descritos neste documento, no todo ou em parte, podem não estar dentro do âmbito de aquisição ou do âmbito de uso. Salvo especificação em contrário no contrato, todas as declarações, informações e recomendações neste documento são fornecidas "TAL COMO ESTÃO" sem garantias ou representações de qualquer tipo, sejam expressas ou implícitas.

As informações contidas neste documento estão sujeitas a alterações sem aviso prévio. Foram feitos todos os esforços na preparação deste documento para assegurar a exatidão do conteúdo, mas todas as declarações, informações e recomendações contidas neste documento não constituem uma garantia de qualquer tipo, expressa ou implícita.

Índice

1 Criação rápida de um cluster de DWS e importação de dados para consulta.....	1
2 Primeiros passos com o SQL de desenvolvimento de dados do DWS.....	15
3 Primeiros passos com o DWS.....	17

1 Criação rápida de um cluster de DWS e importação de dados para consulta

Cenário

DWS é um serviço de armazém de dados em nuvem totalmente gerenciado e de nível empresarial. Ele é livre de O&M, compatível com o ecossistema PostgreSQL e oferece suporte à expansão de cluster online e ao carregamento eficiente de várias fontes de dados. Ele ajuda as empresas a analisar e monetizar com eficiência grandes quantidades de dados online.

Esta seção descreve como criar um cluster de DWS com três nós e importar dados CSV de um bucket do OBS para realizar uma análise e consulta de dados simples.

AVISO

O cluster de DWS não oferece suporte ao acesso entre regiões aos dados do bucket do OBS. Antes de criar um cluster de DWS, verifique se o cluster e o bucket do OBS estão na mesma região. Use CN-Hong Kong como exemplo.

- **Modo de Cobrança:** Pagamento por uso
- **Versão:** Armazenamento e computação acoplados
- **Nós:** 3 nós
- **Modo de conexão do cluster:** Editor SQL
- **Capacidade de armazenamento:** 20 GB/nó
- **Fonte de dados:** dados CSV no bucket do OBS
- **Versão do cluster:** 8.1.3.x

Procedimento

Tabela 1-1 Procedimento

Etapa	Descrição
Preparações	Registre uma conta da Huawei, assine a Huawei Cloud, conclua a autenticação do nome real, recarregue a conta e planeje e crie recursos de VPC e sub-rede.
Passo 1: Criação de um cluster de DWS	Crie um cluster de três nós de DWS e configure informações como a quantidade de nós do cluster, a capacidade de armazenamento e a versão do cluster.
Passo 2: Carregamento de dados para o OBS	Crie um bucket do OBS e carregue os dados CSV locais para o bucket.
Passo 3: Conexão ao cluster de DWS e importação de dados para análise	Use o editor SQL para se conectar ao banco de dados de DWS, criar uma tabela estrangeira do OBS, importar dados do bucket do OBS para o cluster de DWS por meio da tabela estrangeira do OBS e executar consultas SQL simples.

Preparações

- Crie uma conta de DWS, conclua a autenticação do nome real e verifique o status da conta antes de usar o DWS. A conta não pode estar em atraso ou congelada.
- Obtenha o AK/SK da conta. Para obter detalhes, consulte [Chaves de acesso](#).
- Crie uma VPC e recursos de sub-rede.

Passo 1: Criação de um cluster de DWS

Passo 1 Faça login no [console do DWS](#) e vá para a página **Create DWS Cluster**.

Passo 2 Configure os parâmetros de acordo com [Tabela 1-2](#).

Tabela 1-2 Configurações de parâmetros

Parâmetro	Exemplo de valor	Descrição
Região	Selecione a região CN-Hong Kong . NOTA Este guia usa CN-Hong Kong como exemplo. Se você quiser selecionar outra região, certifique-se de que outros serviços em nuvem estejam na mesma região.	Para menor latência de rede e acesso mais rápido aos recursos, selecione a região mais próxima. Depois que um cluster de DWS é criado, a região não pode ser alterada. Tenha cuidado ao selecionar uma região. Para obter mais informações, consulte Regiões e AZs .

Parâmetro	Exemplo de valor	Descrição
Modo de cobrança	Pagamento por uso	Os modos de cobrança anual/mensal e pagamento por uso são suportados. Para uso de longo prazo, é recomendável optar pelo modo de cobrança anual ou mensal para obter um desconto maior. Para obter mais informações, consulte Visão geral de cobrança .
AZ	Escolha AZ1 para criar um cluster de AZ única. (Se o flavor estiver esgotado, selecione outra AZ.)	● AZ única: todos os nós de um cluster são implementados na mesma AZ. ● Várias AZs: para obter HA em várias AZs, implemente nós de cluster em diferentes AZs. Para serviços que exigem HA ainda maior, recomenda-se uma configuração multi-AZ.

Figura 1-1 Configurações básicas

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
The logical cluster is mainly used for computing acceleration.

Bind Main Logical Cluster

Bind User
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

Time Range

Period Type ☒ Every Week ☐ Every Month

* Time Configuration(UTC)

Creation Completion Time	Sunday	01:00	Deletion Start Time	Thursday	04:00
Creation Completion Time	Friday	04:00	Deletion Start Time	Saturday	03:00

Passo 3 A [Tabela 1-3](#) lista os parâmetros de armazém de dados necessários.

Tabela 1-3 Parâmetros de armazém de dados

Parâmetro	Exemplo de valor	Descrição
Version	Armazenamento e computação acoplados	● Armazenamento e computação acoplados: os dados são armazenados em discos locais de nós de computação. ● Armazenamento e computação desacoplados: discos locais de DN são usados apenas para cache de dados e armazenamento de metadados, e os dados do usuário são armazenados no OBS.
Storage Type	SSD em nuvem	● SSD em nuvem: Discos SSD EVS servem como mídia de armazenamento de dados, oferecendo capacidade de armazenamento flexível e suporte a scale-out de disco. ● SSD locais: Discos locais em flavors de ECS servem como mídia de armazenamento de dados, oferecendo capacidade fixa e desempenho superior. No entanto, o dimensionamento horizontal de disco não é suportado. Para obter mais informações, consulte Quais são as diferenças entre SSDs em nuvem e SSDs locais?
Node Flavor	dwsx2.xlarge NOTA Se esse flavor estiver esgotado, selecione outras AZs ou flavors.	Os flavors de nó medem o poder de computação do DWS, incluindo arquitetura de CPU, vCPUs, memória, tipo de computação e tipo de implementação. Você precisa selecionar os flavors de nó conforme necessário. Normalmente, um desempenho de consulta mais alto requer flavors de nó maiores. Para obter mais informações, consulte Flavors do armazenamento de dados.
Hot Storage	20 GB por nó	Capacidade de armazenamento disponível de cada nó. A página exibe a capacidade de armazenamento real para o serviço, como 20 GB por nó com um total de 3 nós resultando em uma capacidade total de dados de serviço de 60 GB. Clientes com uma grande quantidade de dados de serviço devem planejar sua capacidade de armazenamento de acordo.
Nodes	3	Número de nós de cluster. O número de nós varia de 3 a 256.

Figura 1-2 Configuração do armazenamento de Dados

Data Warehouse Configuration

Version

Coupled storage and compute

Decoupled storage and compute

Storage Type

Cloud SSD

Local SSD

Node Flavor

Flavor Name	vCPUs Memory	Hot Storage
☒ dwsx2.xlarge Trial only. Not for production	4 vCPUs 32 GB	20-2,000 GB SSD
☐ dwsx2.h.2xlarge.4.c7n	8 vCPUs 32 GB	100-4,000 GB SSD
☐ dwsx2.2xlarge	8 vCPUs 64 GB	100-4,000 GB SSD
☐ dwsx2.8xlarge	32 vCPUs 256 GB	100-16,000 GB SSD

Hot Storage

-

20

+

 GB per node

Nodes

-

3

+

You can still create 1,024 nodes. You can apply for a higher quota as needed.

Total

dwsx2.xlarge | 12 vCPUs | 96 GB Memory | 60.00 GB hot storage

Passo 4 Configure os parâmetros de acordo com Tabela 1-4.

Tabela 1-4 Parâmetros de configuração de rede

Parâmetro	Exemplo de valor	Descrição
VPC	vpc-default	VPC onde o DWS está localizado.
Subnet	subnet-default(192.168.0.0/24)	Sub-rede da VPC.
Security Group	Automatic creation	Você pode selecionar Automatic creation ou escolher um grupo de segurança criado. A opção padrão é Automatic creation, o que significa que o sistema criará um grupo de segurança automaticamente se você não alterar a configuração. A regra do grupo de segurança padrão é a seguinte: a saída permite todas as solicitações de acesso, enquanto a entrada é aberta apenas para a porta de banco de dados 8000 que você definiu para se conectar ao cluster de DWS.
EIP	Do not use	Se o cliente conectado ao DWS e o cluster de DWS não estiverem na mesma VPC, use um EIP.

Parâmetro	Exemplo de valor	Descrição
ELB	Do not use	Se o endereço IP privado ou EIP de um CN for usado para se conectar a um cluster, a falha desse CN levará à falha de conexão do cluster. As verificações de integridade do ELB permitem o encaminhamento rápido de solicitações de CN para nós íntegros num cluster, minimizando falhas de acesso em caso de um CN com defeito. Isso garante que a carga de trabalho seja imediatamente transferida para um nó íntegro. Para obter mais informações, consulte Associação e Desassociação de ELB.

Figura 1-3 Configuração de rede

Network

VPC

vpc-default

View VPC

Subnet

subnet-default(192.168.0.0/24)

Security Group

Automatic creation

Automatically enables the port's access permission in the inbound rule.

Others

EIP

Do not use

Buy now

Specify

A cluster without an EIP cannot be accessed over the Internet, but can be accessed through an ECS deployed on a private network.

ELB

Do not use

Use existing

ELB is a service that automatically distributes incoming traffic across multiple backend cloud servers based on a specified forwarding policy. Clusters that do not use ELB cannot distribute traffic and are vulnerable to single points of failure (SPOFs). [Learn more](#)

Passo 5 Configure o cluster consultando [Tabela 1-5](#).

Tabela 1-5 Configuração de gerenciamento de cluster

Parâmetro	Exemplo de valor	Descrição
Cluster Name	dws-demo	Nome do cluster, definido pelo usuário e que pode ser modificado posteriormente.

Parâmetro	Exemplo de valor	Descrição
Cluster Version	Use the recommended version	Cada versão oferece suporte a diferentes funções de banco de dados. Por padrão, a versão recomendada é selecionada. Para usar as funções de uma versão posterior, selecione uma versão posterior. NOTA Na página de cliente da lista branca, outras versões de cluster são exibidas. Selecione uma versão de cluster com base nos requisitos do site. Para obter detalhes sobre as funções suportadas por cada versão de cluster, consulte Aviso do produto.
Time Zone	Use default time zone	Fuso horário padrão do cluster.
Enterprise Project	default	Este parâmetro é exibido apenas quando uma conta empresarial é usada para criar um cluster de DWS. Ele permite o gerenciamento unificado de recursos de nuvem por projeto.
Tags	-	No DWS, depois de criar um cluster, você pode adicionar identificadores a itens como o nome do projeto, tipo de serviço e informações de fundo usando tags. Se você usar tags em outros serviços de nuvem, é recomendável criar os mesmos pares de chave-valor de tag para recursos de nuvem usados pela mesma carga de trabalho para manter a consistência.

Figura 1-4 Configurações avançadas

Management

Cluster Name

dwv-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

Administrator Account

dbadmin

Enter 6 to 64 characters starting with a lowercase letter or an underscore (_). Only lowercase letters digits and underscores (_) are allowed.

Administrator Password

Confirm Password

Database Port

8000

Port for clients to connect to DWS. The value ranges from 8000 to 30000.

Time Zone

UTC+08:00 (Beijing, Chongqing, Hong ...)

Advanced

Enterprise Project

default

[View enterprise projects.](#)

Advanced Settings

Default

Custom

Passo 6 Configure o banco de dados consultando [Tabela 1-6](#).

Tabela 1-6 Gerenciamento de banco de dados, criptografia de banco de dados e mais configurações

Parâmetro	Exemplo de valor	Descrição
Administra tor Account	dbadmin	Nome de usuário do administrador do sistema do banco de dados. Mantenha o valor padrão. O administrador do sistema do banco de dados recebe o atributo SYSADMIN por padrão. Se o recurso de separação de permissões não estiver habilitado, o administrador terá as permissões do banco de dados mais altas. Isso permite que eles planejem objetos de banco de dados, criem bancos de dados e usuários e autorizem funções de usuário. Para obter mais informações, consulte Usuários do banco de dados .
Administra tor Password	-	Senha do administrador do sistema do banco de dados.
Confirm Password	-	Insira a senha novamente para confirmar.
Database Port	8000	Porta para o cliente se conectar ao banco de dados do DWS. Mantenha o valor padrão 8000 .

Parâmetro	Exemplo de valor	Descrição
Encrypt DataStore	-	Se essa função estiver habilitada, o Key Management Service (KMS) criptografa o cluster e os dados de snapshot do cluster.
More Configurations	-	● CNs: Em um cluster de grande escala, é recomendável implementar vários CNs. ● IPv6 CIDR Block: Especifique se deseja habilitar a pilha dupla IPv6 para o cluster. Se IPv6 for selecionado, um cliente ou aplicação poderá se conectar ao banco de dados usando um endereço IPv6.

Passo 7 Na área **Configuration Summary** à direita da página, confirme as configurações e clique em **Buy Now**.

A criação do cluster leva cerca de 10 minutos. Clusters no estado **Available** estão prontos para uso.

----Fim

Passo 2: Carregamento de dados para o OBS

Passo 1 Crie um bucket do OBS. Para mais detalhes, consulte [Guia do Usuário do OBS](#).

Defina os parâmetros principais da seguinte forma e mantenha os valores padrão para outros parâmetros.



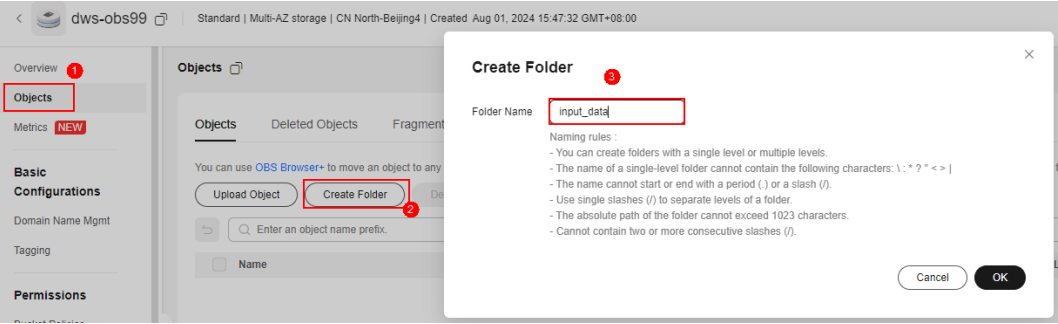
CUIDADO

O acesso entre regiões aos dados do bucket do OBS não é suportado. O bucket do OBS deve ser criado na mesma região que o cluster do DWS. Selecione CN-Hong Kong.

- **Region:** Selecione a **CN-Hong Kong** região.
- **Nome do Bucket:** **dws-obs01** (Se ocorrer um conflito, modifique o número no nome para 02, 03 ou um número maior)
- **Política do Bucket:** **Privado**.

Passo 2 Após a criação do bucket do OBS, clique no nome do bucket do **dws-obs01** recém-criado na lista de buckets. No painel de navegação, escolha **Objects** e clique em **Create Folder**. Defina o nome da pasta como **input_data** e clique em **OK**.

Figura 1-5 Criação de uma pasta



Passo 3 Crie três arquivos de dados CSV no PC local. Os nomes dos arquivos (incluindo a extensão do nome do arquivo) são **product_info0.csv**, **product_info1.csv** e **product_info2.csv**.

Figura 1-6 Criação de um arquivo test.csv

	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

Passo 4 Use um editor de texto para abrir os três arquivos CSV e copiar o seguinte conteúdo para cada arquivo CSV.

- 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"
80,"GKLW-l",2017-09-22,"A","2017 Jeans Men","red","39",58,2017-09-25,72,"Very
comfortable."
30,"HWECL",2017-09-23,"A","2017 shoes
women","red","M",403,2017-09-26,607,"good!"
40,"IQPD-M",2017-09-24,"B","2017 new pants
Women","red","M",35,2017-09-27,52,"very good."
50,"LPEC-N",2017-09-25,"B","2017 dress Women","red","M",29,2017-09-28,47,"not
good at all."
60,"NQAB-O",2017-09-26,"B","2017 jacket
women","red","S",69,2017-09-29,70,"It's beautiful."
70,"HWNB-P",2017-09-27,"B","2017 jacket women","red","L",30,2017-09-30,55,"I
like it so much"
```

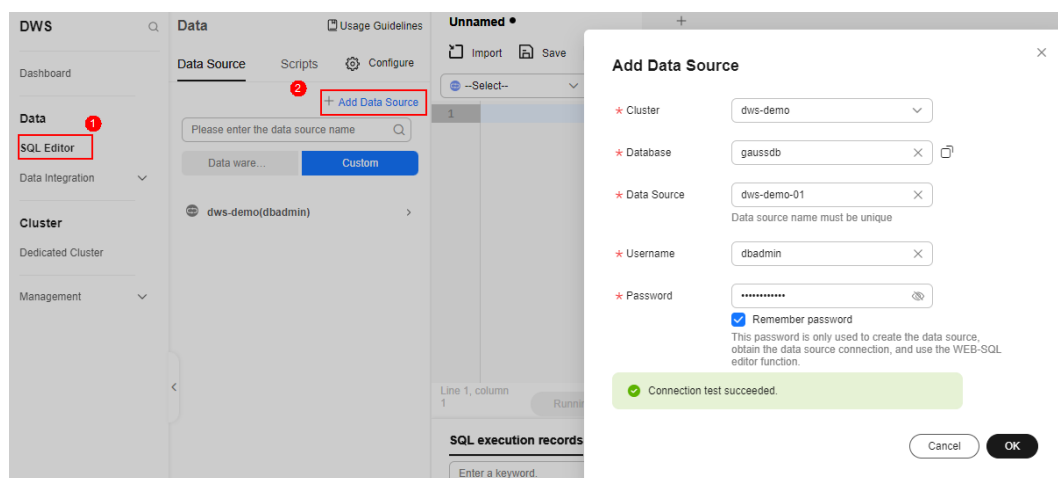
```
80,"JKHU-Q",2017-09-29,"C","2017 T-shirt","red","M",90,2017-10-02,82,"very good."
```

- Passo 5** Use um editor de texto para converter o formato de codificação para UTF-8 e salvar o arquivo.
- Passo 6** Retorne ao console do OBS e clique no nome do bucket do **dws-obs01** recém-criado na lista de buckets.
- Passo 7** No painel de navegação, escolha **Objects** e clique na pasta **input_data**.
- Passo 8** Clique em **Upload Object** e, em seguida, clique em **Add File**.
- Passo 9** Carregue os arquivos **product_info0.csv**, **product_info1.csv** e **product_info2.csv** no bucket do OBS.
- Fim

Passo 3: Conexão ao cluster de DWS e importação de dados para análise

- Passo 1** Acesse o console do DWS.
- Passo 2** Na árvore de navegação à esquerda, escolha **Data > SQL Editor** e clique em **Add Data Source**.
- Passo 3** Insira as informações de conexão do banco de dados.
- **Data Source:** Insira **dws-demo-01**.
 - **Cluster:** Selecione o cluster do **dws-demo** criado.
 - **Database:** Mantenha o valor padrão **gaussdb**.
 - **Username:** Insira **dbadmin**.
 - **Password:** Insira a senha definida durante a criação do DWS no [Passo 1: Criação de um cluster de DWS](#).
- Passo 4** Selecione **Remember password** e clique em **Test Connection**.

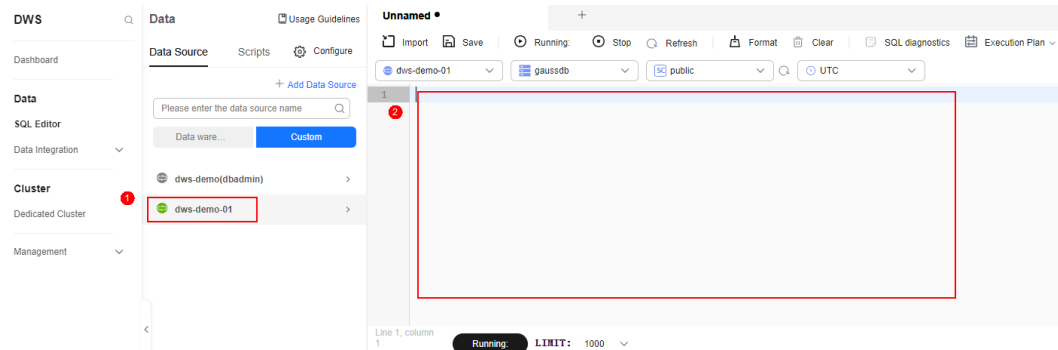
Figura 1-7 Informações de conexão do banco de dados



- Passo 5** Clique em **OK**.

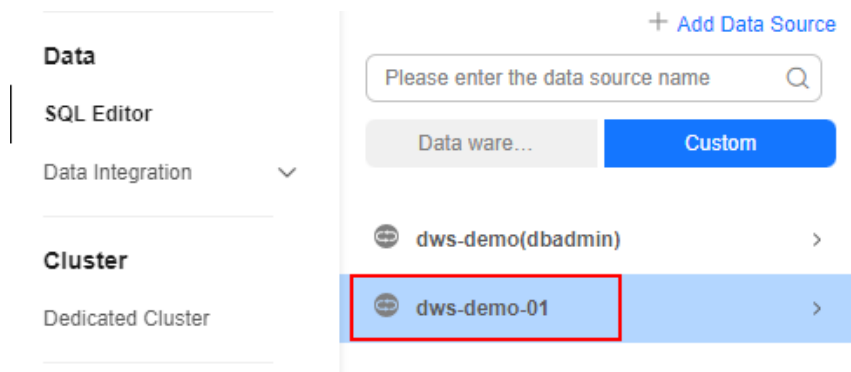
O ícone verde  antes do nome da fonte de dados na figura a seguir indica que o banco de dados está conectado. O painel direito é a janela do editor SQL, onde você pode escrever e executar instruções SQL.

Figura 1-8 Conexão de banco de dados bem-sucedida



NOTA

Se a sessão de conexão expirar ou o navegador for atualizado,  aparecerá em cinza, indicando que a conexão foi desconectada. Para reconectar ao banco de dados, clique duas vezes no nome da fonte de dados.

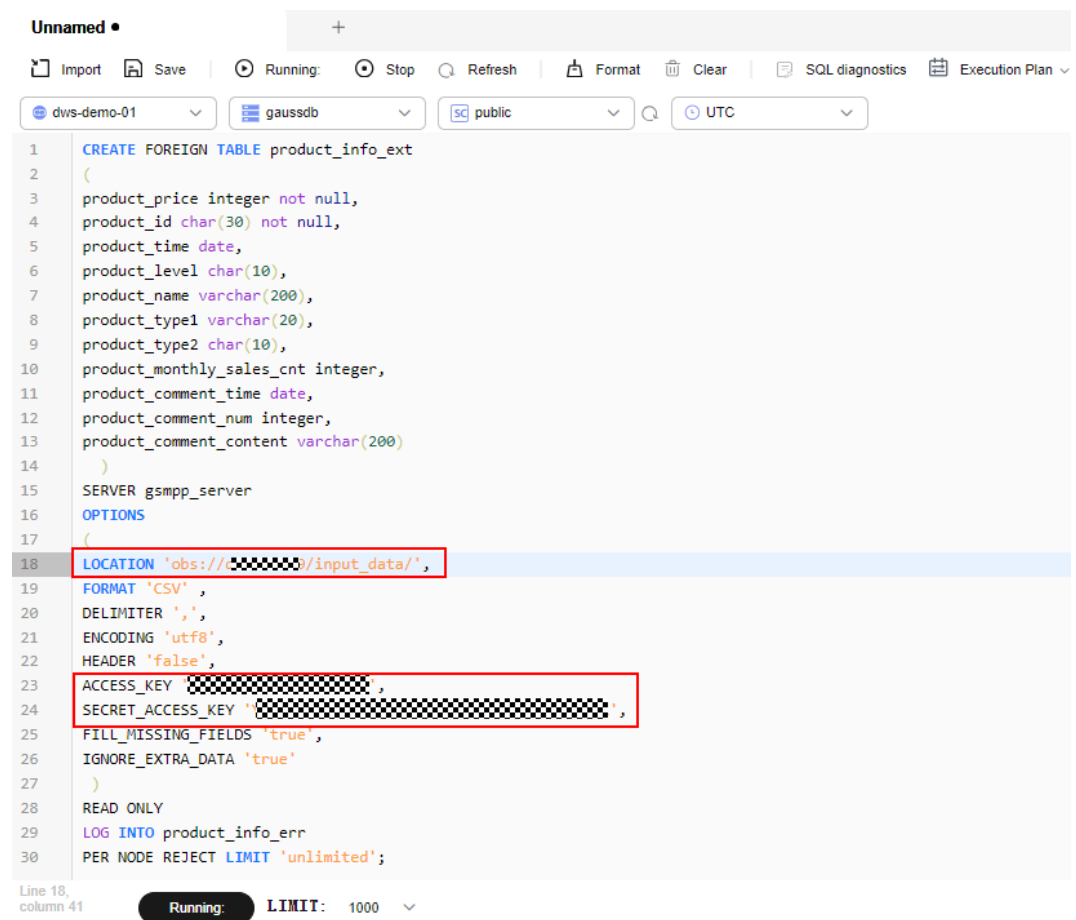


Passo 6 Copie a seguinte instrução SQL para a janela do editor SQL e substitua o nome do bucket do OBS, o valor AK e o valor SK pelos valores reais.

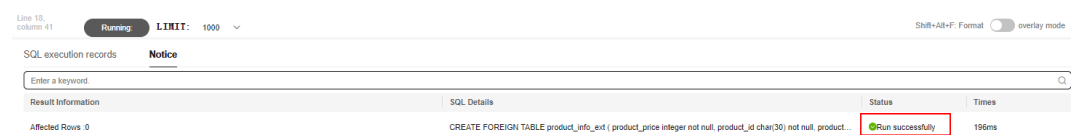
Neste documento, o nome do bucket do OBS é **dws-obs01**. Obtenha o AK e o SK em [Chaves de Acesso](#).

```
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';
```

Figura 1-9 Criação de uma tabela estrangeira do OBS

Passo 7 Verifique se a instrução SQL está correta e clique em **Running**. A tabela estrangeira do OBS é criada.

Figura 1-10 Criação bem-sucedida

Passo 8 Copie a seguinte instrução SQL e execute-a na janela SQL para criar uma tabela comum.

```
CREATE TABLE product_info  
(
```

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

Passo 9 Copie a seguinte instrução SQL e execute-a na janela SQL. Execute a instrução **INSERT** para importar os dados da tabela estrangeira do OBS para o banco de dados do DWS.

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

Passo 10 Copie as seguintes instruções SQL e execute-as na janela SQL para realizar as operações **VACUUM** e **ANALYZE**.

Para obter detalhes sobre a sintaxe, consulte a seção de sintaxe SQL relacionada.

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

Passo 11 Copie a seguinte instrução SQL e execute-a na janela SQL. Os detalhes dos dados são consultados com sucesso.

```
SELECT * FROM product_info;
```

Figura 1-11 Consulta bem-sucedida

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

----Fim

2 Primeiros passos com o SQL de desenvolvimento de dados do DWS

Esta seção descreve algumas operações básicas de SQL no DWS.

Criação, visualização e exclusão de um banco de dados

- Execute a instrução **CREATE DATABASE** para criar um banco de dados.

```
CREATE DATABASE test_db ENCODING 'UTF8' template = template0;
```
- Consulte a lista de banco de dados usando o meta-comando **\l**.

```
\l
```
- Consulta da lista de banco de dados usando o catálogo do sistema **PG_DATABASE**

```
SELECT datname FROM pg_database;
```
- Execute a instrução **DROP DATABASE** para excluir um banco de dados.

```
DROP DATABASE test_db;
```

Criação, visualização, modificação e exclusão de uma tabela

- Execute a instrução **CREATE TABLE** para criar uma tabela.

```
CREATE TABLE customer_t1(id INT, name CHAR(40),age TINYINT);
```
- Use a função **PG_GET_TABLEDEF()** para exibir a instrução de criação da tabela. Substitua **customer_t1** pelo nome da tabela a ser consultada.

```
SELECT * FROM PG_GET_TABLEDEF('customer_t1');
```
- Execute a instrução **ALTER TABLE** para modificar uma tabela.
Adicione uma coluna:

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


Exclua uma coluna:

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


Modifique um tipo de coluna:

```
ALTER TABLE customer_t1 MODIFY age INTEGER NOT NULL;
```
- Execute a instrução **DROP TABLE** para excluir uma tabela.

```
DROP TABLE customer_t1;
```

Criação, exibição e exclusão de índices

- Execute a instrução **CREATE INDEX** ou **ALTER TABLE** para criar um índice comum.

```
CREATE INDEX c_id_index on customer_t1(id);
```

```
ALTER TABLE customer_t1 ADD INDEX c_id_index (id);
```

- Use o catálogo do sistema **PG_INDEXES** para exibir todos os índices em uma tabela.

```
SELECT * FROM pg_indexes WHERE tablename = 'customer_t1';
```
- Execute a instrução **ALTER TABLE** ou **DROP INDEX** para excluir um índice.

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

Adição, exclusão e modificação de dados de tabela

- Execute a instrução **INSERT INTO** para inserir dados de tabela.

```
INSERT INTO customer_t1 VALUES(1001,'user1',22);
```
- Execute a instrução **SELECT** para consultar dados da tabela.

```
SELECT * FROM customer_t1;
```
- Execute a instrução **UPDATE** para atualizar os dados da tabela.

```
UPDATE customer_t1 SET id = 1009 WHERE id = '1001';
```
- Use a instrução **DELETE** para excluir dados da tabela.

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

3 Primeiros passos com o DWS

Depois de criar um cluster de DWS, você pode experimentar alguns Melhores práticas fornecidos pelo DWS para atender aos requisitos da sua carga de trabalho.

Tabela 3-1 Melhores práticas

Prática		Descrição
Importação e exportação de dados	Importação de dados do OBS para um cluster	Esta prática demonstra como carregar dados de exemplo para o OBS e importar dados do OBS para a tabela de destino no DWS, ajudando você a aprender rapidamente como importar dados do OBS para um cluster de DWS. Você pode importar dados no formato TXT, CSV, ORC, PARQUET, CARBONDATA ou JSON do OBS para um cluster de DWS para consulta.
	Usando o GDS para importar dados de um servidor remoto	Esta prática demonstra como usar o General Data Service (GDS) para importar dados de um servidor remoto para o DWS. DWS permite importar dados no formato TXT, CSV ou FIXED.

Prática		Descrição
	Exportação de dados ORC de um cluster de DWS para um cluster de MRS	O DWS permite exportar dados ORC para o MRS usando uma tabela estrangeira HDFS. Você pode especificar o modo de exportação e o formato de dados de exportação na tabela estrangeira. Os dados são exportados do DWS em paralelo usando vários DNs e armazenados no HDFS. Desta forma, o desempenho geral das exportações é melhorado.
Migração de dados	Uso do CDM para migrar dados do Oracle para um cluster de DWS	Esta prática demonstra como migrar dados do Oracle para o DWS.
	Uso de um job do Flink do DLI para sincronizar dados do Kafka com um cluster de DWS em tempo real	Esta prática demonstra como usar jobs do Flink do DLI para sincronizar dados de consumo do Kafka para o DWS em tempo real. Esta prática leva cerca de 90 minutos. Os serviços em nuvem usados nesta prática incluem Virtual Private Cloud (VPC) e sub-redes, Elastic Load Balance (ELB), Elastic Cloud Server (ECS), Object Storage Service (OBS), Distributed Message Service (DMS) para Kafka, Data Lake Insight (DLI) e DWS.

Prática		Descrição
Tabela de otimização	Práticas de otimização de tabelas	Nesta prática, você aprenderá como otimizar o design de suas tabelas. Você começará criando tabelas sem especificar seu modo de armazenamento, chave de distribuição, modo de distribuição ou modo de compactação. Carregar dados de teste nessas tabelas e teste o desempenho do sistema. Em seguida, siga excelentes práticas para criar as tabelas novamente usando novos modos de armazenamento, chaves de distribuição, modos de distribuição e modos de compactação. Carregar os dados de teste e teste o desempenho novamente. Comparar os dois resultados do teste para descobrir como o design da tabela afeta o espaço de armazenamento e o desempenho de carregamento e consulta das tabelas. Tempo estimado: 60 minutos
Recursos avançados	Melhores práticas de gerenciamento de dados quentes e frios	À medida que os dados crescem em cenários de Big Data massivo, as necessidades de armazenamento e o uso de recursos aumentam rapidamente. A necessidade de dados pode variar em diferentes períodos, portanto, os dados são gerenciados de maneira hierárquica, melhorando o desempenho da análise de dados e reduzindo os custos do serviço. Em alguns cenários de uso de dados, os dados podem ser classificados em dados quentes e dados frios acessando a frequência.

Prática		Descrição
	Melhores práticas para gerenciamento automático de partições	Para tabelas de partição cujas colunas de partição são baseadas em tempo, a função de gerenciamento automático de partição pode ser adicionada para criar automaticamente partições e excluir partições expiradas, reduzindo os custos de manutenção da tabela de partição e melhorando o desempenho da consulta. Para facilitar a consulta e a manutenção de dados, a coluna de tempo é frequentemente usada como a coluna de partição de uma tabela particionada que armazena dados relacionados ao tempo, como informações de pedidos de comércio eletrônico e dados de IoT em tempo real. Quando os dados relacionados ao tempo são importados para uma tabela particionada, a tabela deve ter partições dos intervalos de tempo correspondentes. Tabelas de partições comuns não criam automaticamente novas partições nem excluem partições expiradas. Portanto, o pessoal de manutenção precisa criar periodicamente novas partições e excluir partições expiradas, levando ao aumento dos custos de O&M. Para resolver isso, o DWS introduz o recurso de gerenciamento automático de partições. Você pode definir os parâmetros de nível de tabela period e ttl para ativar a função de gerenciamento automático de partições, que cria automaticamente partições e exclui partições expiradas, reduzindo os custos de manutenção de tabelas particionadas e melhorando o desempenho das consultas.

Prática		Descrição
Gerenciamento de banco de dados	Melhores práticas de gerenciamento de recursos	Esta prática demonstra como usar o DWS para gerenciamento de recursos, ajudando as empresas a eliminar gargalos em consultas simultâneas. Os trabalhos de SQL podem ser executados sem problemas sem afetar uns aos outros e consumir menos recursos do que antes.
	Melhores práticas para consultas SQL	Com base num grande número de mecanismos e práticas de execução SQL, podemos otimizar instruções SQL seguindo certas regras para executar instruções SQL mais rapidamente e obter resultados corretos.
	Melhores práticas para consultas de distorção de dados	Esta prática inclui os seguintes casos de distorção de armazenamento: ● Detecção em tempo real de distorção de armazenamento durante a importação de dados ● Localização rápida das tabelas que causam distorção de dados
	Melhores práticas de gerenciamento de usuários	Um cluster do DWS consiste principalmente em administradores de sistema e usuários comuns. Esta prática descreve as permissões de administradores de sistema e usuários comuns e descreve como criar usuários e consultar informações do usuário.
	Exibição de informações de tabelas e banco de dados	Esta prática demonstra alguns casos básicos de consulta de banco de dados: ● Consulta de informações de tabelas ● Consulta do tamanho da tabela ● Consulta de informações do banco de dados ● Consulta do tamanho do banco de dados

Prática		Descrição
Análise de dados de amostra	Análise de veículos em pontos de verificação	Esta prática demonstra como analisar veículos que passaram por pontos de verificação de tráfego. O processo envolve o carregamento de 890 milhões de dados desses pontos de verificação numa única tabela de banco de dados no DWS para consultas precisas e difusas. É um ótimo exemplo de como o DWS pode lidar com consultas de alto desempenho de dados históricos.
	Análise de requisitos da cadeia de suprimentos (conjunto de dados TPC-H)	Esta prática demonstra como carregar o conjunto de dados de amostra do OBS para um cluster de DWS e executar consultas de dados. Ele compreende análise de várias tabelas e análise de temas em cenários de análise de dados.
	Análise do status de operações de uma loja de departamentos de varejo	Nesta prática, os dados de negócios diários de cada loja de varejo são carregados do OBS para a tabela correspondente no cluster de armazém de dados para resumir e consultar KPIs. Esses dados incluem a rotatividade da loja, o fluxo de clientes, a classificação mensal de vendas, a taxa de conversão mensal do fluxo de clientes, índice mensal de preço-aluguel e as vendas por unidade de área. Esta prática demonstra a consulta e análise multidimensional do DWS em cenários de varejo.
Segurança de dados	Criptografia e descriptografia de colunas de dados	A criptografia de dados é amplamente utilizada em sistemas de informação para evitar acesso não autorizado e vazamento de dados. Como o núcleo de um sistema de informação, o armazém de dados do DWS também fornece criptografia transparente e criptografia usando funções SQL. Esta seção descreve a criptografia de função SQL.