Data Warehouse Service (DWS)
8.1.3.333

Service Overview

Issue	01
Date	2024-08-09

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


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 What Is GaussDB(DWS)?..... 1
2 Data Warehouse Types..... 9
3 Data Warehouse Flavors..... 13
3.1 Supported ECS Flavors..... 13
3.2 Supported BMS Flavors..... 15
4 Advantages..... 22
5 Application Scenarios..... 24
6 Functions..... 27
7 Concepts..... 33
8 GaussDB(DWS) Access..... 35
9 Restrictions..... 37
10 Restricted Functions..... 41
11 Technical Specifications..... 43

1 What Is GaussDB(DWS)?

GaussDB(DWS) is an online data analysis and processing database built on the cloud infrastructure and platform. It offers scalable, ready-to-use, and fully managed analytical database services, and is compatible with ANSI/ISO SQL92, SQL99, and SQL 2003 syntax. Additionally, GaussDB(DWS) is interoperable with other database ecosystems such as PostgreSQL, Oracle, Teradata, and MySQL. This makes it a competitive option for petabyte-scale big data analytics across diverse industries.

- Standard Data Warehouse (DWS 2.0): Oriented to data analysis scenarios, DWS 2.0 provides enterprise-level data warehouse services with high performance, high scalability, high reliability, high security, and easy O&M. It is capable of data analysis at a scale of 2048 nodes and 20 petabytes of data. It supports hot and cold data analysis, elastic scaling of storage and computing resources, and on-demand and pay-per-use pricing, providing users with elastic, flexible, and cost-effective experience. It is applicable to converged analysis services that integrate libraries, warehouses, cities, and lakes.
- **Stream data warehouse**: The stream data warehouse is built on top of the standard data warehouse. It integrates stream and time sequence engines to provide real-time data ingestion and high-concurrency real-time analysis capabilities. It can be used for IoT real-time analysis.
- **Hybrid data warehouse**: It provides high-concurrency, high-performance, and low-latency transaction processing capabilities based on large-scale data query and analysis capabilities. It is suitable for hybrid transaction/analytical processing (HTAP). A database can be used for both production and analysis.

In addition, GaussDB(DWS) can be deployed on physical machines. For details, see **Physical Machine Deployment**.

Version Form

When GaussDB(DWS) is installed, the following types of clusters are provided: Elastic Cloud Server (ECS) and Bare Metal Server (BMS) clusters installed using images, and physical machine clusters managed by ManageOne.

NOTE

Existing GaussDB(DWS) clusters of 8.0.0 need to be upgraded to 8.1.1 or later (The version of 6.5.1 needs to be upgraded to 8.0.0 first and then to 8.1.1 or later) for management.

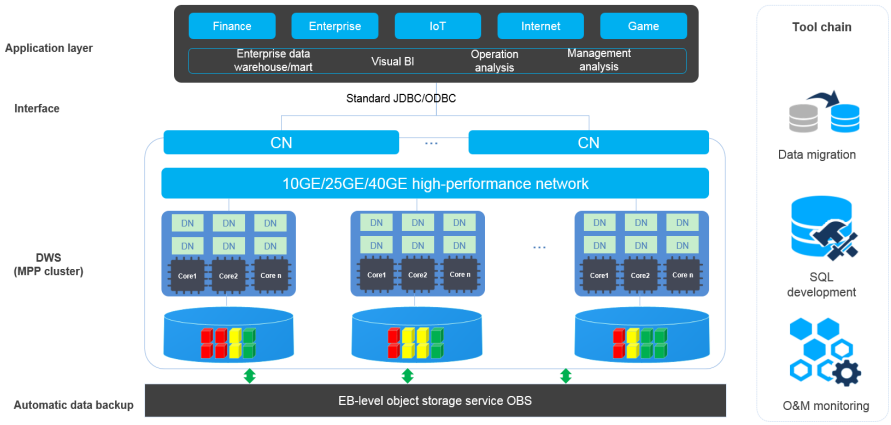
Table 1-1 GaussDB(DWS) cluster types

Cluster Type	Service Version	Cluster Provisioning Mode
ECS	8.1.3.333	When DWS Console and the corresponding GaussDB(DWS) image are installed, create an ECS GaussDB(DWS) cluster on the console.
BMS	8.1.3.333	When DWS Console and the corresponding GaussDB(DWS) image are installed, create a BMS GaussDB(DWS) cluster on the console.
Physical machine	• MPPDB service: 8.1.0 or later • FusionInsight Manager: 6.5.1.7 and later, and 8.0.2.1 • FusionInsight Base: 6.5.1.7 and later, and 8.0.2.1	• **Existing physical machines**: – 6.5.1 and earlier: Upgrade the cluster to 8.0 and then to 8.1.1 or later and manage the cluster on the GaussDB(DWS) console. – 8.0: Upgrade the cluster to 8.1.1 or later and manage the cluster on the GaussDB(DWS) console. • **Upgrade of managed physical machines**: For the 8.1.0 physical machine cluster that has been managed on the GaussDB(DWS) console, if the upgrade function is required, cancel the management, upgrade the cluster to 8.1.1 or later, and then manage the cluster again.

Architecture

GaussDB(DWS) employs the shared-nothing architecture and the massively parallel processing (MPP) engine, and consists of numerous independent logical nodes that do not share the system resources such as CPUs, memory, and storage. In such a system architecture, service data is separately stored on numerous nodes. Data analysis tasks are executed in parallel on the nodes where data is stored. The massively parallel data processing significantly improves response speed.

Figure 1-1 Architecture



- **Application layer**

  Data loading tools, extract, transform, and load (ETL) tools, business intelligence (BI) tools, as well as data mining and analysis tools, can be integrated with GaussDB(DWS) through standard APIs. GaussDB(DWS) is compatible with the PostgreSQL ecosystem, and the SQL syntax is compatible with Oracle and Teradata. Applications can be smoothly migrated to GaussDB(DWS) with few changes.

- **API**

  Applications can connect to GaussDB(DWS) through the standard Java Database Connectivity (JDBC) 4.0 and Open Database Connectivity (ODBC) 3.5.

- **GaussDB(DWS)**

  A GaussDB(DWS) cluster contains nodes of the same flavor in the same subnet. These nodes jointly provide services. Datanodes (DNs) in a cluster store data on disks. Coordinators (CNs) receive access requests from applications and return the execution results to clients. In addition, a CN splits and distributes tasks to the DNs for parallel processing.

- **Automatic data backup**

  Cluster snapshots can be automatically backed up to the EB-level Object Storage Service (OBS), which facilitates periodic backup of the cluster during off-peak hours, ensuring data recovery after a cluster exception occurs.

  A snapshot is a complete backup of GaussDB(DWS) at a specific time point, including the configuration data and service data of a cluster.

- **Tool chain**

  The parallel data loading tool General Data Service (GDS), SQL syntax migration tool Database Schema Convertor (DSC), and SQL development tool Data Studio are provided. The cluster O&M can be monitored on a console.

Logical Cluster Architecture

Figure 1-2 shows the logical architecture of a GaussDB(DWS) cluster. For details about instances, see Table 1-2.

Figure 1-2 Logical cluster architecture

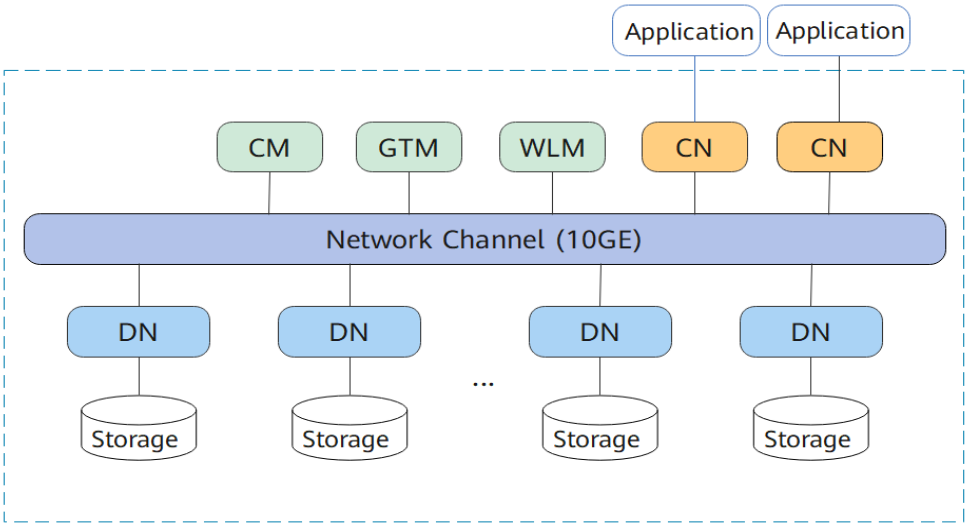


Table 1-2 Cluster architecture description

Name	Function	Description
Cluster Manager (CM)	Cluster Manager. It manages and monitors the running status of functional units and physical resources in the distributed system, ensuring system stability.	The CM consists of CM Agent, OM Monitor, and CM Server. • CM Agent monitors the running status of primary and standby GTMs, CNs, and primary and standby DNs on the host, and reports the status to CM Server. In addition, it executes the arbitration instruction delivered by CM Server. A CM Agent process runs on each host. • OM Monitor monitors scheduled tasks of CM Agent and restarts CM Agent when CM Agent stops. If CM Agent cannot be restarted, the host cannot be used. In this case, manually rectify this fault. **NOTE** CM Agent cannot be restarted probably because of insufficient system resources, which is not a common situation. • CM Server checks whether the current system is normal according to the instance status reported by CM Agent. In the case of exceptions, CM Server delivers recovery commands to CM Agent. GaussDB(DWS) provides the primary/standby CM Server solution to ensure system HA. CM Agent connects to the primary CM Server. If the primary CM Server is faulty, the standby CM Server is promoted to primary to prevent a single point of failure (SPOF).
Global Transaction Manager (GTM)	Generates and maintains the globally unique information, such as the transaction ID, transaction snapshot, and timestamp.	The cluster includes only one pair of GTMs: one primary GTM and one standby GTM.
Workload Manager (WLM)	Workload Manager. It controls allocation of system resources to prevent service congestion and system crash resulting from excessive workload.	You do not need to specify names of hosts where WLMs are to be deployed, because the installation program automatically installs a WLM on each host.

Name	Function	Description
Coordinator (CN)	A CN receives access requests from applications, and returns execution results to the client; splits tasks and allocates task fragments to different DNs for parallel processing.	CNs in a cluster have equivalent roles and return the same result for the same DML statement. Load balancers can be added between CNs and applications to ensure that CNs are transparent to applications. If a CN is faulty, the load balancer automatically connects the application to the other CN. For details, see section "Associating and Disassociating ELB". CNs need to connect to each other in the distributed transaction architecture. To reduce heavy load caused by excessive threads on GTMs, no more than 10 CNs should be configured in a cluster. GaussDB(DWS) handles the global resource load in a cluster using the Central Coordinator (CCN) for adaptive dynamic load management. When the cluster is started for the first time, the CM selects the CN with the smallest ID as the CCN. If the CCN is faulty, CM replaces it with a new one.

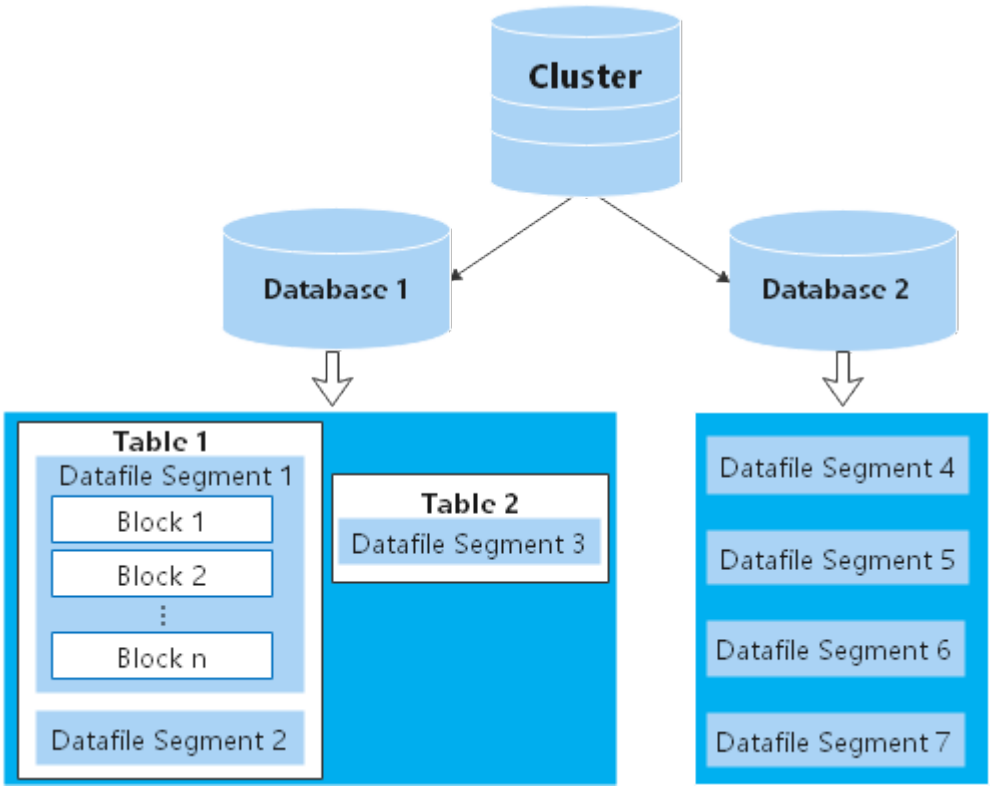
Name	Function	Description
Datanode (DN)	A DN stores service data by column or row or in the hybrid mode, executes data query tasks, and returns execution results to CNs.	A cluster consists of multiple DNs and each DN stores part of data. GaussDB(DWS) provides DN high availability: active DN, standby DN, and secondary DN. The working principles of the three are as follows: • During data synchronization, if the active DN suddenly becomes faulty, the standby DN is switched to the active state. • Before the faulty active DN recovers, the new active DN synchronizes data logs to the secondary DN. • After the faulty active DN recovers, it becomes the standby DN and uses data logs stored on the secondary DN to restore data generated during its faulty period. The secondary DN serves exclusively as a backup, never ascending to active or standby status in case of faults. It conserves storage by only holding Xlog data transferred from the new active DN and data replicated during original active DN failures. This efficient approach saves one-third of the storage space compared to conventional tri-backup methods.
Storage	Functions as the server's local storage resources to store data permanently.	-

DNs in a cluster store data on disks. **Figure 1-3** describes the objects on each DN and the relationships among them logically.

- A database manages various data objects and is isolated from other databases.
- A datafile segment stores data in only one table. A table containing more than 1 GB of data is stored in multiple data file segments.
- A table belongs only to one database.
- A block is the basic unit of database management, with a default size of 8 KB.

Data can be distributed in replication, round-robin, or hash mode. You can specify the distribution mode during table creation.

Figure 1-3 Logical database architecture



2 Data Warehouse Types

Product Type Overview

- **Standard data warehouse (DWS 2.0)**: It can analyze hot and cold data and is highly cost-effective. Its storage and computing resources are not limited, and can be elastically scaled and billed per use. It is suitable for the converged analysis that requires integrated databases, warehouses, marts, and lakes. It is most suitable for OLAP workloads.
- **Hybrid data warehouse**: It provides high-concurrency, high-performance, and low-latency transaction processing capabilities at low costs based on large-scale data query and analysis capabilities. It is suitable for hybrid transaction/analytical processing (HTAP). A database can be used for both production and analysis.
- **Stream data warehouse**: It provides efficient time series computing and IoT analysis capabilities based on the standard data warehouse and supports correlation between real-time and historical data. The compression ratio can reach 40:1. It can be used for IoT real-time analysis.

  NOTE

  GaussDB(DWS) data warehouses cannot access each other. You can create an OBS foreign table to associate two databases in the same data directory for data query.

Features

Table 2-1 Features

Module	Module	Standard Data Warehouse (DWS 2.0)	IoT	Hybrid (Cluster)
Navigation menu	Dashboard	Yes	Yes	Yes
	Cluster management	Yes	Yes	Yes
	DR Management	Yes	Yes	Yes

Module	Module	Standard Data Warehouse (DWS 2.0)	IoT	Hybrid (Cluster)
	Snapshot management	Yes	Yes	Yes
	Parameters	Yes	Yes	Yes
	Incident management	Yes	Yes	Yes
	Alarms	Yes	Yes	Yes
	Client connections	Yes	Yes	Yes
Dashboard	Resources	Yes	Yes	Yes
	Alarms	Yes	Yes	Yes
	Recent Events	Yes	Yes	Yes
	Cluster monitoring metrics (DMS)	Yes	Yes	Yes
Cluster management	Monitoring panel (DMS)	Yes	Yes	Yes
	Monitoring metrics (Cloud Eye)	Yes	Yes	Yes
	Restart	Yes	Yes	Yes
	Start	Yes	Yes	Yes
	Stop	Yes	Yes	Yes
	Scaling	Yes	Yes	Yes
	Scale-in	Yes	Yes	Yes
	Redistributing data	Yes	Yes	Yes
	Viewing redistribution details	Yes	Yes	Yes
	Changing node flavor	Yes	Yes	Yes
	Changing all specifications	Yes	No	Yes

Module	Module	Standard Data Warehouse (DWS 2.0)	IoT	Hybrid (Cluster)
	Resetting passwords	Yes	Yes	Yes
	Creating a snapshot	Yes	Yes	Yes
	Canceling read-only status	Yes	Yes	Yes
	Deletion	Yes	Yes	Yes
	CNs	Yes	Yes	Yes
	Adding disk capacity	Yes	Yes	Yes
Cluster information	Basic information	Yes	Yes	Yes
	ELB	Yes	Yes	Yes
	Resource pool	Yes	Yes	Yes
	Logical cluster	Yes	Yes	Yes
	Snapshot	Yes	Yes	Yes
	Parameter modifications	Yes	Yes	Yes
	Security settings	Yes	Yes	Yes
	Monitoring panel	Yes	Yes	Yes
	Tags	Yes	Yes	Yes
	Node management	Yes	Yes	Yes
DR Management	DR management	Yes	Yes	Yes
Snapshot management	Restoration	Yes	Yes	Yes
	Deletion	Yes	Yes	Yes
	Copy	Yes	Yes	Yes
Incident management	Event management (general)	Yes	Yes	Yes

Module	Module	Standard Data Warehouse (DWS 2.0)	IoT	Hybrid (Cluster)
Alarms	Alarms	Yes	Yes	Yes
Client connections	Client connections	Yes	Yes	Yes
Others	Inspection	Yes	Yes	Yes
	Intelligent O&M	Yes	Yes	Yes
	Node restoration	Yes	Yes	Yes
	Warm backup on the tenant side	Yes	Yes	Yes
	OpenApi	Yes	Yes	Yes

3 Data Warehouse Flavors

3.1 Supported ECS Flavors

GaussDB(DWS) provides standard, hybrid, and stream data warehouses. For details about the differences between them, see **Data Warehouse Types**.

ECS+Local Passthrough Flavors

Table 3-1 ECS/Local passthrough flavors

Resource Feature	Flavor	Type	vCPUs	Memory (GB)	Single Disk Capacity (GB)	Physical Drives	Primary DNs	Disk Type
Ultra-high I/O	dws.i3.4xlarge.6	x86	12	72	3,840	2	1	SSD
	dws.i3.8xlarge.6	x86	24	144	3,840	4	2	SSD
	dws.i3.12xlarge.6	x86	48	288	3,840	8	4	SSD
	dws.ki1.4xlarge.4	Arm	16	64	3,840	2	1	SSD
	dws.ki1.8xlarge.4	Arm	32	128	3,840	4	2	SSD
	dws.ki1.16xlarge.4	Arm	64	228	3,840	8	4	SSD

ECS+EVS VM Flavors

Table 3-2 ECS/EVS VM flavors

Product Form	CPU Architecture	Flavor	vCPUs	Memory (GB)	Storage (GB SSD)
Cloud data warehouse	X86	dwsx2.xlarge	4	32	100 ~ 2000
	ARM	dwsk2.xlarge	4	32	100 ~ 2000
	X86	dwsx2.2xlarge	8	64	200 ~ 4000
	ARM	dwsk2.2xlarge	8	64	200 ~ 4000
	X86	dwsx2.4xlarge	16	128	400 ~ 8000
	ARM	dwsk2.4xlarge	16	128	400 ~ 8000
	X86	dwsx2.8xlarge	32	256	800 ~ 16000
	ARM	dwsk2.8xlarge	32	256	800 ~ 16000
	ARM	dwsk2.12xlarge	48	384	1200 ~ 24000
	X86	dwsx2.16xlarge	64	512	1600 ~ 32000
Hybrid data warehouse	ARM	dwsk2.h.xlarge.4.kc1	4	16	100 ~ 2000
	X86	dwsx2.h.xlarge.4.c6	4	16	100 ~ 2000
	ARM	dwsk2.h.2xlarge.4.kc1	8	32	200 ~ 4000
	X86	dwsx2.h.2xlarge.4.c6	8	32	200 ~ 4000
	ARM	dwsk2.h.4xlarge.4.kc1	16	64	400 ~ 8000
	X86	dwsx2.h.4xlarge.4.c6	16	64	400 ~ 8000
	ARM	dwsk2.h.8xlarge.4.kc1	32	128	800 ~ 16000
	X86	dwsx2.h.8xlarge.4.c6	32	128	800 ~ 16000
	ARM	dwsk2.h.12xlarge.4.kc1	48	192	1200 ~ 24000

Product Form	CPU Architecture	Flavor	vCPUs	Memory (GB)	Storage (GB SSD)
	X86	dwsx2.h.16xlarge.4.c6	64	256	1600 ~ 32000
Stream data warehouse	ARM	dwsk2.rt.xlarge	4	32	100 ~ 2000
	X86	dwsx2.rt.xlarge	4	32	100 ~ 2000
	ARM	dwsk2.rt.2xlarge.km1	8	64	200 ~ 4000
	X86	dwsx2.rt.2xlarge.m6	8	64	200 ~ 4000
	ARM	dwsk2.rt.8xlarge.km1	32	256	800 ~ 16000
	X86	dwsx2.rt.8xlarge.m6	32	256	800 ~ 16000
	ARM	dwsk2.rt.12xlarge.km1	48	384	1200 ~ 24000
	X86	dwsx2.rt.16xlarge.m6	64	512	1600 ~ 32000

3.2 Supported BMS Flavors

The following table lists the BMS flavors supported by GaussDB(DWS).

NOTE

The system disk size, which is generally 2 x 480 GB or 2 x 960 GB, is not listed in the following table. The system disk is used as RAID 1 and the first RAID logical disk. Set the boot mode of all BMSs to UEFI.

Table 3-3 List of supported BMS flavors

No.	Service Type	Flavor	Type	Number of vCPUs	Memory (GB)	Single Disk Capacity (GB)	Number of Physical Disks	RAID Group Member Disks	Disk Type
1	Standard data warehouse	physical.d2.24xlarge.5	x86	96	512	1800	24/25, where 25 indicates that one hot standby disk is added.	4	SAS
2		physical.i3.24xlarge.5	x86	96	512	3840	12	2	SSD
3		physical.i3.40xlarge.6	x86	160	1024	3840	21	4	SSD
4		physical.i3.24xlarge.5.jbod	x86	96	512	8000	4	2	SSD
5		physical.cw.24xlarge.5	x86	96	512	1920	12	2	SSD
6		physical.d2.26xlarge.5	x86	104	512	1800	24/25, where 25 indicates that one hot standby disk is added.	4	SAS
7		physical.i3.26xlarge.10	x86	104	1024	960	24	4	SSD

No.	Service Type	Flavor	Type	Number of vCPUs	Memory (GB)	Single Disk Capacity (GB)	Number of Physical Disks	RAID Group Member Disks	Disk Type
8		physical.i3.24xlarge.10	x86	96	1024	960	24	4	SSD
9		physical.i3.12xlarge.5.24t	x86	48	256	960	24	4	SSD
10		physical.i3.16xlarge.8.24t	x86	64	512	3840	24	4	SSD
11		physical.i3.16xlarge.4.12t	x86	64	256	3840	12	2	SSD
12		physical.i3.26xlarge.7.12t	X86	104	768	3840	12	2	SSD
13		physical.hd.24xlarge.6	X86	96	512	1800	24	4	SAS
14		physical.hi.32xlarge.6	x86	128	768	3840	12	2	SSD
15		physical.i6.26xlarge.8	x86	104	768	960	24	4	SSD
16		physical.i6.24xlarge.8	x86	96	768	960	24	4	SSD
17		physical.i3.48xlarge.2.12t	X86	192	512	3840	12	2	SSD
18		physical.i3.48xlarge.2.24t	X86	72	512	3840	24	4	SSD
19		physical.ki.24xlarge.5.b	ARM	96	512	3200	24	4	SSD
20		physical.kd2.24xlarge.4	ARM	96	384	1800	24	4	SAS

No.	Service Type	Flavor	Type	Number of vCPUs	Memory (GB)	Single Disk Capacity (GB)	Number of Physical Disks	RAID Group Member Disks	Disk Type
21		physical.kd2.24xlarge.5	ARM	96	512	1800	24/25, where 25 indicates that one hot standby disk is added.	4	SAS
22		physical.ki.24xlarge.5	ARM	96	512	3800	12	2	SSD
23		physical.kd2.32xlarge.4	ARM	128	512	1800	24	4	SAS
24		physical.ki.32xlarge.4	ARM	128	512	3800	12	2	SSD
25		physical.kd2.16xlarge.4	ARM	64	256	1800	24	4	SAS
26		physical.ki.24xlarge.5.a	ARM	96	512	960	12	2	SSD
27		physical.ki.16xlarge.8	ARM	64	512	3840	12	2	SSD
28		physical.ki.16xlarge.8.b	ARM	64	512	960	24	4	SSD
29		physical.ki.16xlarge.4	ARM	64	256	1920	12	2	SSD

No.	Service Type	Flavor	Type	Number of vCPUs	Memory (GB)	Single Disk Capacity (GB)	Number of Physical Disks	RAID Group Member Disks	Disk Type
30		physical.fi.32xlarge.4	ARM	128	512	3840	12	2	SSD
31		physical.fd.32xlarge.4	ARM	128	512	1800	24	4	SAS
32		physical.ki.32xlarge.8	ARM	128	1024	960	24	4	SSD
33	Stream data warehouse	stream.physical.i3.24xlarge.5	x86	96	512	19200	2	2	SSD
34		stream.physical.d2.24xlarge.5	x86	96	512	9000	4	4	SAS
35		stream.physical.i3.26xlarge.7.12t	X86	104	768	3840	12	2	SSD
36		stream.physical.ki.16xlarge.8	ARM	64	512	19200	2	2	SSD
37		stream.physical.ki.32xlarge.4	ARM	128	512	19000	2	2	SSD
38		stream.physical.kd2.16xlarge.4	ARM	64	256	9000	4	4	SAS
39		stream.physical.kd2.32xlarge.4	ARM	128	512	9000	4	4	SAS
40		stream.physical.i3.48xlarge.2.12t	X86	192	512	3840	12	2	SSD
41		stream.physical.i3.48xlarge.2.24t	X86	72	512	3840	24	4	SSD
42		stream.physical.i6.26xlarge.8	x86	104	768	960	24	4	SSD

No.	Service Type	Flavor	Type	Number of vCPUs	Memory (GB)	Single Disk Capacity (GB)	Number of Physical Disks	RAID Group Member Disks	Disk Type
43		stream.physical.i6.24xlarge.8	x86	96	768	960	24	4	SSD
44		stream.physical.hd.24xlarge.6	X86	96	512	1800	24	4	SAS
45		stream.physical.hi.32xlarge.6	x86	128	768	3840	12	2	SSD
46		stream.physical.fd.32xlarge.4	ARM	128	512	1800	24	4	SAS
47		stream.physical.fi.32xlarge.4	ARM	128	512	3840	12	2	SSD
48	Hybrid data warehouse	hybrid.physical.i3.26xlarge.7.12t	X86	104	768	3840	12	2	SSD
49		hybrid.physical.hd.24xlarge.6	x86	96	512	9000	4	4	SAS
50		hybrid.physical.i6.24xlarge.8	x86	96	768	4800	4	4	SSD
51		hybrid.physical.i6.26xlarge.8	x86	104	768	4800	4	4	SSD
52		hybrid.physical.hi.32xlarge.6	x86	128	768	19200	2	2	SSD
53		hybrid.physical.i3.24xlarge.5	x86	96	512	19200	2	2	SSD
54		hybrid.physical.d2.24xlarge.5	x86	96	512	9000	4	4	SAS
55		hybrid.physical.fd.32xlarge.4	ARM	128	512	9000	4	4	SAS
56		hybrid.physical.fi.32xlarge.4	ARM	128	512	19200	2	2	SSD

No.	Service Type	Flavor	Type	Number of vCPUs	Memory (GB)	Single Disk Capacity (GB)	Number of Physical Disks	RAID Group Member Disks	Disk Type
57		hybrid.physical.ki.16xlarge.8	ARM	64	512	19200	2	2	SSD
58		hybrid.physical.ki.32xlarge.4	ARM	128	512	19000	2	2	SSD
59		hybrid.physical.kd2.16xlarge.4	ARM	64	256	9000	4	4	SAS
60		hybrid.physical.kd2.32xlarge.4	ARM	128	512	9000	4	4	SAS
61		hybrid.physical.i3.48xlarge.2.12t	X86	192	512	3840	12	2	SSD
62		hybrid.physical.i3.48xlarge.2.24t	X86	72	512	3840	24	4	SSD

4 Advantages

GaussDB(DWS) uses the GaussDB database kernel and is compatible with PostgreSQL 9.2.4. It transforms from a single OLTP database to an enterprise-level distributed OLAP database oriented to massive data analysis based on the massively parallel processing (MPP) architecture.

Unlike conventional data warehouses, GaussDB(DWS) excels in massive data processing and general platform management with the following benefits:

Ease of use

- Visualized one-stop management

  GaussDB(DWS) allows you to easily complete the entire process from project concept to production deployment. With the GaussDB(DWS) management console, you can obtain a high-performance and highly available enterprise-level data warehouse cluster within several minutes. Data warehouse software or data warehouse servers are not required.

  With just a few clicks, you can easily connect applications to the data warehouse, back up data, restore data, and monitor data warehouse resources and performance.

- Seamless integration with big data

  Without the need to migrate data, you can use standard SQL statements to directly query data on HDFS and OBS.

- Heterogeneous database migration tools

  GaussDB(DWS) provides various migration tools to migrate SQL scripts of Oracle and Teradata to GaussDB(DWS).

High performance

- Cloud-based distributed architecture

  GaussDB(DWS) adopts the MPP-based database so that service data is separately stored on numerous nodes. Data analysis tasks are executed in parallel on the nodes where data is stored. The massively parallel data processing significantly improves response speed.

- Query response to trillions of data records within seconds

  GaussDB(DWS) improves data query performance by executing multi-thread operators in parallel, running commands in registers in parallel with the

vectorized computing engine, and reducing redundant judgment conditions using LLVM.

GaussDB(DWS) provides you with a better data compression ratio (column-store), better indexing (column-store), and higher point update and query (row-store) performance.

- Fast data loading

  GDS is a tool that helps you with high-speed massively parallel data loading.

Robust reliability

- ACID

  Support for the atomicity, consistency, isolation, and durability (ACID) feature, which ensures strong data consistency for distributed transactions.

- Comprehensive HA design

  All software processes of GaussDB(DWS) are in active/standby mode. Logical components such as the CNs and DNs of each cluster also work in active/standby mode. This ensures data reliability and consistency when any single point of failure (SPOF) occurs.

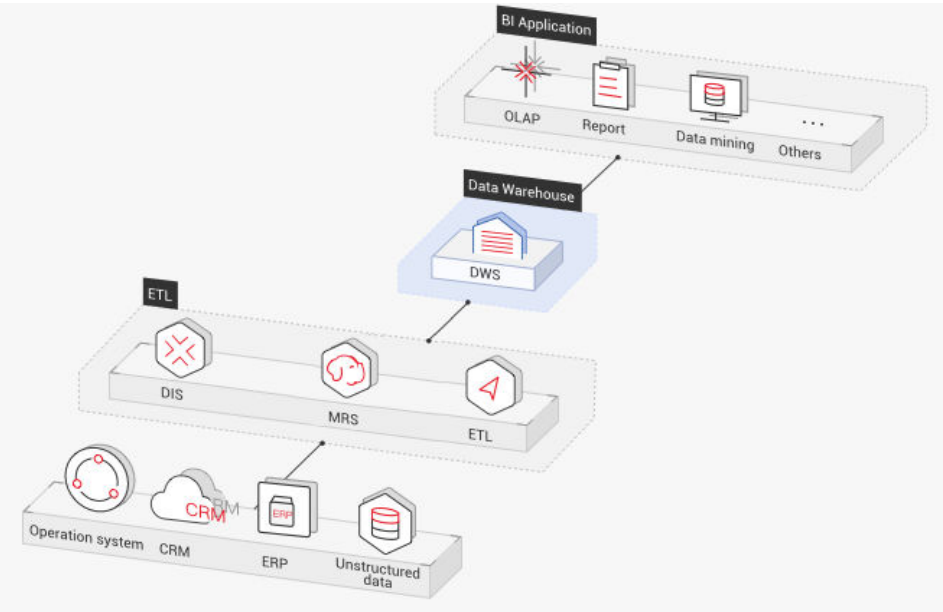
- High security

  GaussDB(DWS) supports transparent data encryption and can interconnect with the Database Security Service (DBSS) to better protect user privacy and data security with network isolation and security group rule setting options. In addition, GaussDB(DWS) supports automatic full and incremental backup of data, improving data reliability.

5 Application Scenarios

- Enhanced ETL + Real-time BI analysis

  Figure 5-1 ETL+BI analysis

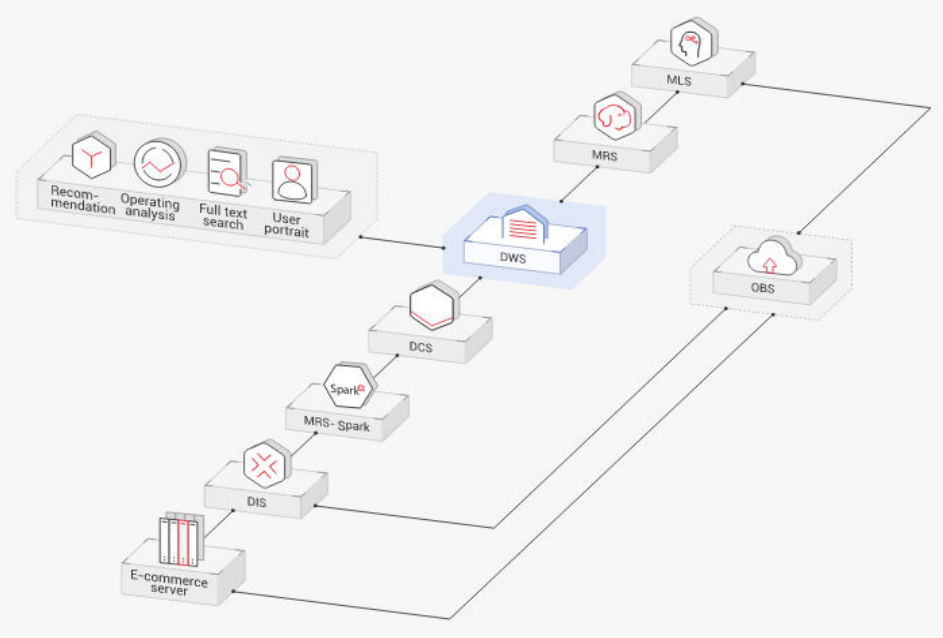
  The data warehouse is the pillar of the Business Intelligence (BI) system for collecting, storing, and analyzing massive amounts of data. It provides powerful business analysis support for IoT, mobile Internet, gaming, and Online to Offline (O2O) industries.

  Advantages of GaussDB(DWS) are as follows:

  - **Data migration**: efficient and real-time data import in batches from multiple data sources
  - **High performance**: cost-effective PB-level data storage and second-level response to correlation analysis of trillions of data records
  - **Real-time**: real-time consolidation of service data for timely optimization and adjustment of operation decision-making

- **E-commerce**

Figure 5-2 E-commerce



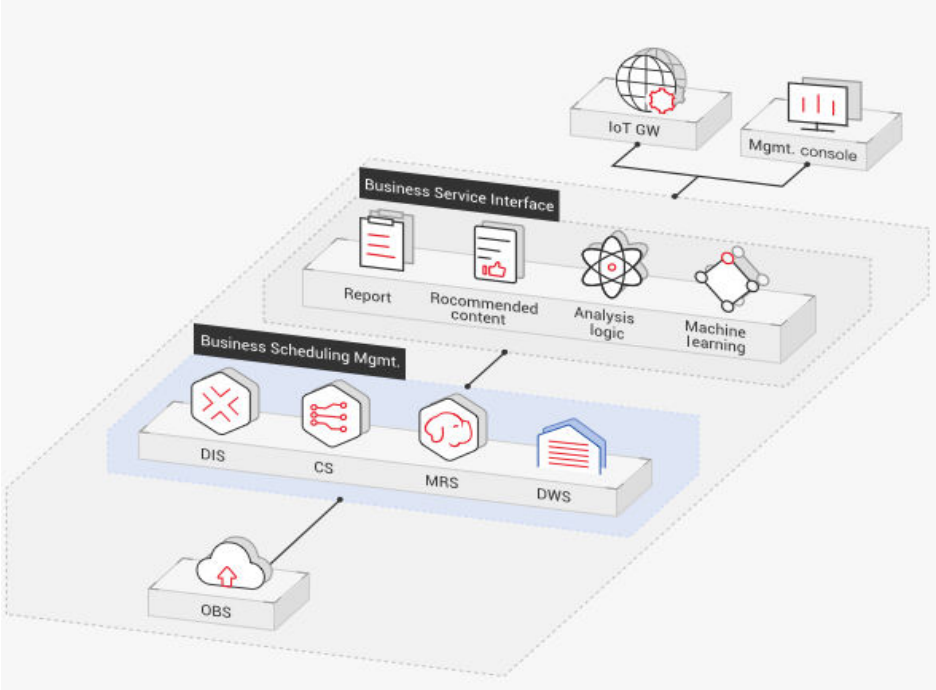
Data of online retailers is mainly used for marketing recommendation, operating and customer analysis, and full text search.

Advantages of GaussDB(DWS) are as follows:

- **Multi-dimensional analysis**: analysis from products, users, operation, regions, and more
- **Scale-out as the business grows**: on-demand cluster scale-out as the business grows
- **High reliability**: long-term stable running of the e-commerce system

- **IoT**

Figure 5-3 IoT



GaussDB(DWS) helps you analyze massive amounts of data from Internet of Things (IoT) in real time and perform optimization based on the results. It is widely used in industrial IoT, O2O service system, and IoV solutions.

Advantages of GaussDB(DWS) are as follows:

- **Device monitoring and prediction**: control, optimization, self-diagnosis, and self-healing based on data analysis, device monitoring, and behavior prediction
- **Information recommendation**: tailed recommendation based on data of users' connected devices

6 Functions

GaussDB(DWS) enables you to use this service through various methods, such as the GaussDB(DWS) management console, GaussDB(DWS) client, and REST APIs. This section describes the main functions of GaussDB(DWS).

Enterprise-Level Data Warehouses and Compatibility with Standard SQL

After a data warehouse cluster is created, you can use the SQL client to connect to the cluster and perform operations such as creating a database, managing the database, importing and exporting data, and querying data.

GaussDB(DWS) provides petabyte-level (PB-level) high-performance databases with the following features:

- MPP computing framework, hybrid row-column storage, and vectorized execution, enabling response to billion-level data correlation analysis within seconds
- Optimized in-memory computing based on Hash Join of Bloom Filter, improving the performance by 2 to 10 times
- Supports the symmetrically distributed, active-active multi-node cluster architecture, ensuring no SPOFs.
- Optimized communication between large-scale clusters based on telecommunication technologies, improving data transmission efficiency between compute nodes
- Cost-based intelligent optimizers, helping generate the optimal plan based on the cluster scale and data volume to improve execution efficiency

GaussDB(DWS) has comprehensive SQL capabilities:

- Supports ANSI/ISO SQL 92, SQL99, and SQL 2003 standards, stored procedures, GBK and UTF-8 character sets, and SQL standard functions and OLAP analysis functions.
- Compatible with the PostgreSQL/Oracle/Teradata/MySQL ecosystem and supports interconnection with mainstream database ETL and BI tools provided by third-party vendors.
- Supports roaring bitmaps and common functions used with them, which are widely used for user feature extraction, user profiling, and more applications in the Internet, retail, education, and gaming industries.

- List partitioning (**PARTITION BY LIST** *(partition_key,[...])*) and range partitioning are supported.
- Read-only HDFS and OBS foreign tables in JSON file format are supported.
- Permissions on system catalogs can be granted to common users. The **VACUUM** permission can be granted separately. Roles with predefined, extensible permissions are supported, including:
  - **ALTER**, **DROP** and **VACUUM** permissions at table level
  - **ALTER** and **DROP** permissions at schema level
  - Preset roles **role_signal_backend** and **role_read_all_stats**

Cluster Management

A data warehouse cluster contains nodes with the same flavor in the same subnet. These nodes jointly provide services. GaussDB(DWS) provides a professional, efficient, and centralized management console, allowing you to quickly apply for clusters, easily manage data warehouses, and focus on data and services.

Main functions of cluster management are described as follows:

- Creating Clusters

  To use data warehouse services on the cloud, create a GaussDB(DWS) cluster first. You can select product and node specifications to quickly create a cluster.

- Managing Snapshots

  A snapshot is a complete backup that records point-in-time configuration data and service data of a GaussDB(DWS) cluster. A snapshot can be used to restore a cluster at a certain time. You can manually create snapshots for a cluster or enable automated snapshot creation (periodic). Automated snapshots have a limited retention period. You can copy automatic snapshots for long-term retention.

  When you restore a cluster from a snapshot, the system can restore the snapshot data to a new cluster or the original cluster.

  You can delete snapshots that are no longer needed on the console to release storage space. Automated snapshots cannot be manually deleted.

- Managing nodes

  You can check the nodes in a cluster, including the status, specifications, and usage of each node. To prepare for a large scale-out, you can add nodes in batches. For example, if 180 more BMS nodes are needed, add them in three batches (60 for each batch). If some nodes fail to be added, add them again. After all the 180 nodes are successfully added, use the nodes for cluster scale-out. Adding nodes does not affect cluster services.

- Scaling out clusters

  As the service volume increases, the current scale of a cluster may not meet service requirements. In this case, you can scale out the cluster by adding compute nodes to it. Services are not interrupted during the scale-out. You can enable online scale-out and automatic redistribution if necessary.

- Managing redistribution

  By default, redistribution is automatically started after cluster scale-out. For enhanced reliability, disable the automatic redistribution function and manually start a redistribution task after the scale-out is successful. Data

redistribution can accelerate service response. Currently, offline redistribution, online redistribution, and offline scheduling are supported. The default mode is offline redistribution.

- Resource management

  When multiple database users query jobs at the same time, some complex queries may occupy cluster resources for a long time, affecting the performance of other queries. For example, a group of database users continuously submit complex and time-consuming queries, while another group of users frequently submit short queries. In this case, short queries may have to wait in the queue for the time-consuming queries to complete. To improve efficiency, you can use the GaussDB(DWS) resource management function to handle such problems. You can create different resource pools for different types of services, and configure different resource ratios for these pools. Then, add database users to the corresponding pools to restrict their resource usages.

- Logical cluster

  A physical cluster can be divided into logical clusters that use the node-group mechanism. Tables in a database can be allocated to different physical nodes by logical cluster. A logical cluster can contain tables from multiple databases.

- Restarting clusters

  Restarting a cluster may cause data loss in running services. If you have to restart a cluster, ensure that there is no running service and all data has been saved.

- Deleting Clusters

  You can delete a cluster when you do not need it. Deleting a cluster is risky and may cause data loss. Therefore, exercise caution when performing this operation.

GaussDB(DWS) allows you to manage clusters in either of the following ways:

- Management console

  Use the management console to access GaussDB(DWS) clusters. When you have registered an account, log in to the management console and choose **Data Warehouse Service**.

- REST APIs

  Use REST APIs provided by GaussDB(DWS) to manage clusters. In addition, if you need to integrate GaussDB(DWS) into a third-party system for secondary development, use APIs to access the service.

Diverse Data Import Modes

GaussDB(DWS) supports efficient data import from multiple data sources. The following lists typical data import modes. For details, see "Data Migration to GaussDB(DWS)" in *Data Warehouse Service (DWS) Developer Guide*.

- Importing data from OBS in parallel
- Using GDS to import data from a remote server
- Importing data from one GaussDB(DWS) cluster to another
- Using the gsql meta-command **\COPY** to import data

- Running the **COPY FROM STDIN** statement to import data
- Using Database Schema Convertor (DSC) to migrate SQL scripts
- Using **gs_dump** and **gs_dumpall** to export metadata
- Using **gs_restore** to import data

APIs

You can call standard APIs, such as JDBC and ODBC, to access databases in GaussDB(DWS) clusters.

For details, see "Using the JDBC and ODBC Drivers to Connect to a Cluster" in the *Data Warehouse Service (DWS) User Guide*.

High Reliability

- Supports instance and data redundancy, ensuring zero single points of failure (SPOF) in the entire system.
- Supports multiple data backups, and all data can be manually backed up to OBS.
- Automatically isolates the faulty node, uses the backup to restore data, and replaces the faulty node when necessary.
- Automatic snapshots work with OBS to implement intra-region disaster recovery (DR). If the production cluster fails to provide read and write services due to natural disasters in the specified region or cluster internal faults, the DR cluster becomes the production cluster to ensure service continuity.
- In the **Unbalanced** state, the number of primary instances on some nodes increases. As a result, the load pressure is high. In this case, you can perform a primary/standby switchback for the cluster during off-peak hours to improve performance.
- If the internal 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. To avoid single-CN failures, GaussDB(DWS) uses Elastic Load Balance (ELB). An ELB distributes access traffic to multiple ECSs for traffic control based on forwarding policies. It improves the fault tolerance capability of application programs.
- After a cluster is created, the number of required CNs varies with service requirements. GaussDB(DWS) allows you to add or delete CNs as needed.

Security Management

- Isolates tenants and controls access permissions to protect the privacy and data security of systems and users based on the network isolation and security group rules, as well as security hardening measures.
- Supports SSL network connections, user permission management, and password management, ensuring data security at the network, management, application, and system layers.

  For details, see "Configuring SSL Connection" and "Configuring Separation of Permissions" in the *Data Warehouse Service (DWS) User Guide*.

Monitoring and Auditing

- Monitoring Clusters

  GaussDB(DWS) integrates with Cloud Eye, allowing you to monitor compute nodes and databases in the cluster in real time. For details, see "Monitoring Clusters" in the *Data Warehouse Service (DWS) User Guide*.

- Database Monitoring

  DMS is provided by GaussDB(DWS) to ensure the fast and stable running of databases. It collects, monitors, and analyzes the disk, network, and OS metric data used by the service database, as well as key performance metric data of cluster running. It also diagnoses database hosts, instances, and service SQL statements based on the collected metrics to expose key faults and performance problems in a database in a timely manner, and guides customers to optimize and resolve the problems. For details, see "Database Monitoring" in the *Data Warehouse Service (DWS) User Guide*.

- Alarms

  Alarm management includes viewing and configuring alarm rules and subscribing to alarm information. Alarm rules display alarm statistics and details of the past week for users to view tenant alarms. In addition to providing a set of default GaussDB(DWS) alarm rules, this feature allows you to modify alarm thresholds based on your own services. For details, see "Alarm Management" in the *Data Warehouse Service (DWS) User Guide*.

- Notifying Events

  GaussDB(DWS) interconnects with Simple Message Notification (SMN) so that you can subscribe to events and view events that are triggered. For details, see "Event Notifications" in the *Data Warehouse Service (DWS) User Guide*.

- Audit Logs
  - GaussDB(DWS) records all SQL operations, including connection attempts, query attempts, and database changes. For details, see "Setting Database Audit Logs" in the *Data Warehouse Service (DWS) User Guide*.

Multiple Database Tools

GaussDB(DWS) provides the following self-developed tools. You can download the tool packages on the GaussDB(DWS) management console. For details about the tools, see the *Data Warehouse Service (DWS) Tool Guide*.

- gsql

  gsql is a command line SQL client tool running on the Linux operating system. It helps connect to, operate, and maintain the database in a data warehouse cluster.

- Data Studio

  Data Studio is a Graphical User Interface (GUI) SQL client tool running on the Windows operating system. It is used to connect to the database in a data warehouse cluster, manage the database and database objects, edit, run, and debug SQL scripts, and view the execution plans.

- GDS

  GDS is a data service tool provided by GaussDB(DWS). It works with the foreign table mechanism to implement high-speed data import and export.

The GDS tool package needs to be installed on the server where the data source file is located. This server is called the data server or the GDS server.

- DSC SQL syntax migration tool

  The DSC is a command-line tool running on the Linux or Windows OS. It is dedicated to providing customers with simple, fast, reliable application SQL script migration services. It parses SQL scripts of source database applications by using the built-in syntax migration logic, and migrates them to be applicable to GaussDB(DWS) databases.

  The DSC can migrate SQL scripts of Teradata, Oracle, Netezza, MySQL, and DB2 databases.

- **gs_dump** and **gs_dumpall**

  gs_dump exports a single database or its objects. **gs_dumpall** exports all databases or global objects in a cluster.

  To migrate database information, you can use a tool to import the exported metadata to a target database.

- gs_restore

  During database migration, you can export files using **gs_dump tool** and import them to GaussDB(DWS) by using **gs_restore**. In this way, metadata, such as table definitions and database object definitions, can be imported.

7 Concepts

GaussDB(DWS) Management Concepts

- Cluster

  A cluster is a server group that consists of multiple nodes. GaussDB(DWS) is organized using clusters. A data warehouse cluster contains nodes with the same flavor in the same subnet. These nodes work together to provide services.

- Node

  A GaussDB(DWS) cluster can have 3 to 256 nodes. A hybrid data warehouse (standalone) can only have one node. Each node can store and analyze data.

- Type

  You need to specify the node flavors when you create a data warehouse cluster. CPU, memory, and storage resources vary depending on node flavors.

- Snapshot

  You can create snapshots to back up GaussDB(DWS) cluster data. A snapshot is retained until you delete it on the management console. Automated snapshots cannot be manually deleted. Snapshots will occupy your OBS quotas.

- Project

  Projects are used to group and isolate OpenStack resources (computing resources, storage resources, and network resources). A project can be a department or a project team. Multiple projects can be created for one account.

GaussDB(DWS) Database Concepts

- Databases

  A data warehouse cluster is an analysis-oriented relational database platform that supports online analysis.

- OLAP

  OLAP is a major function of data warehouse clusters. It supports complex analysis, provides decision-making support tailored to analysis results, and delivers intuitive query results.

- MPP

On each node in the data warehouse cluster, memory computing and disk storage systems are independent from each other. With MPP, GaussDB(DWS) distributes service data to different nodes based on the database model and application characteristics. Nodes are connected through the network and collaboratively process computing tasks as a cluster and provide database services that meet service needs.

- Shared-Nothing Architecture

  The shared-nothing architecture is a distributed computing architecture. Each node is independent so that nodes do not compete for resources, which improves work efficiency.

- Database Version

  Each data warehouse cluster has a specific database version. You can check the version when creating a data warehouse cluster.

- Database Connections

  You can use a client to connect to the GaussDB(DWS) cluster. The client can be used for connection on the cloud platform and over the Internet.

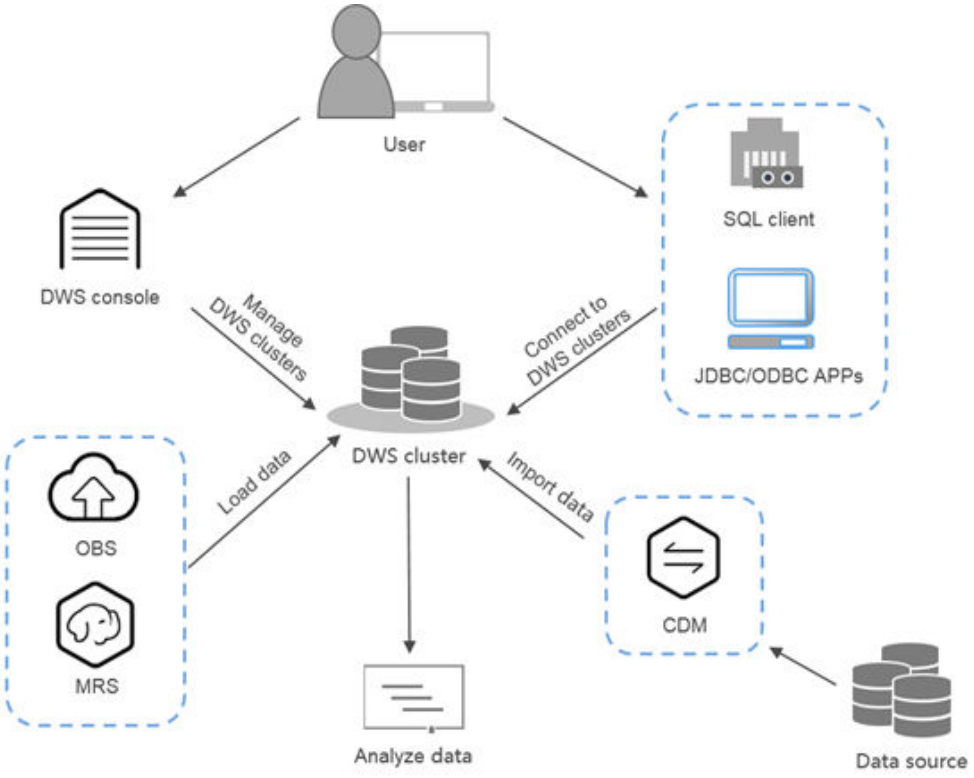
- Database User

  You can add and control users who can access the database of a data warehouse cluster by assigning specific permissions to them. The database administrator generated when you create a cluster is the default database user.

8 GaussDB(DWS) Access

The following figure shows how to use GaussDB(DWS).

Figure 8-1 Process for using GaussDB(DWS)



Accessing a Cluster

GaussDB(DWS) provides a web-based management console and HTTPS-compliant APIs for you to manage data warehouse clusters.

NOTE

In cluster deployment, if a single node is faulty, the abnormal node is automatically skipped when GaussDB(DWS) is accessed. However, the cluster performance will be affected.

Accessing the Database in a Cluster

GaussDB(DWS) supports database access using the following methods:

- GaussDB(DWS) clients

  Access the cluster database using GaussDB(DWS) clients. For details, see "Connecting to a Cluster" in the *Data Warehouse Service (DWS) User Guide*.

- JDBC and ODBC API calling

  You can call standard APIs, such as JDBC and ODBC, to access databases in clusters.

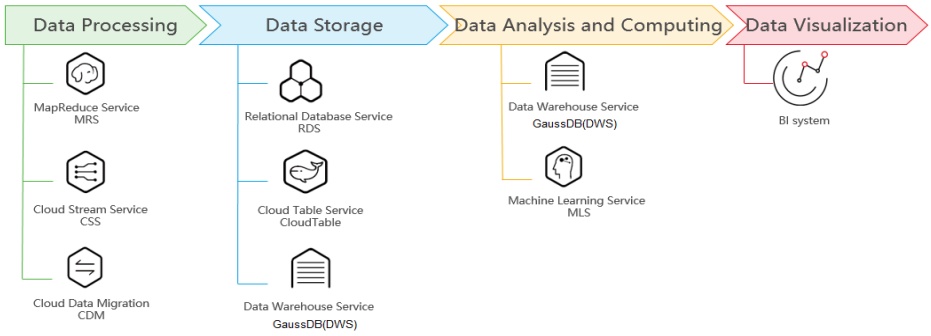
  For details, see "Using the JDBC and ODBC Drivers to Connect to a Cluster" in the *Data Warehouse Service (DWS) User Guide*.

End-to-End Data Analysis Process

GaussDB(DWS) has been seamlessly integrated with other services on cloud, helping you rapidly deploy end-to-end data analysis solutions.

The following figure shows the end-to-end data analysis process. Services in use during each process are also displayed.

Figure 8-2 End-to-end data analysis process



9 Restrictions

This document describes the constraints and precautions of using the key functions of GaussDB(DWS).

After creating a GaussDB(DWS) cluster, you do not need to perform basic database O&M operations, such as HA and security patch installation. However, you need to pay attention to the following:

Table 9-1 GaussDB (DWS) constraints

Item	Constraint
Creating a Cluster	• After a cluster is created, its type cannot be changed. For details about the differences between product types, see "Data Warehouse Types".
Connecting to a Cluster	• If you use a client to connect to a cluster, its VPC subnet must be the same as that of the cluster. • You can manage clusters only and cannot directly access nodes in a cluster. You can use a cluster's IP address and port to access the database in the cluster.
SQL Syntax Changes	• If you copy commands from the document to the operating environment, the text wraps automatically, causing command execution failures. To solve the problem, delete the line break. • You are advised to create databases as required. Do not use the default gaussdb database of a cluster.

Item	Constraint
Changing Flavor	• Only clusters of version 8.1.1.300 or later support flavor change. • Currently, only offline flavor chage is supported. The change takes about 10 minutes. • If your cluster is created using local disks or compute-storage integration, you cannot change the flavor of the cluster. If you need nodes with a higher flavor, create a new cluster. Currently, you can only change the flavors of cloud data warehouse clusters that use SSD cloud disks or hybrid data warehouse clusters. • Currently, changing all specifications can only be performed for standard data warehouses and hybrid data warehouses (except single-node deployment). The **Change all specifications** option does not support logical clusters. • A cluster can have up to 240 nodes. The old and new clusters can have up to 480 nodes in total. • Disk capacity expansion can be performed only for standard data warehouses using SSD, hybrid data warehouses, or stream data warehouses. Only version 8.1.1.203 and later are supported. • Disk capacity can be expanded only if the cluster is in **Available**, **To be restarted**, **Read-only**, or **Node fault**, **Unbalanced** state.
Scaling Out	• When you scale out the standard data warehouse cluster, use the same storage specifications as the cluster. • The cluster redistribution function is supported in 8.1.1.5 or later cluster versions. • This function can be manually enabled only when the cluster task information displays **To be redistributed** after scale-out. • Scale-in is supported only by clusters of version 8.1.1.300 and later. • When you scale in a standard data warehouse cluster, you can only modify the same storage specifications as used by the cluster.

Item	Constraint
Backing Up a Cluster	• The new GaussDB(DWS) cluster created based on the snapshot must have the same configurations as the original cluster. That is, the number and specifications of nodes, memory, and disks in the new cluster must be the same as those in the original cluster. • If you create a new cluster based on a snapshot without modifying parameters, the parameters of the new cluster will be the same as those of the snapshot. • During snapshot creation, do not perform the VACUUM FULL operation, or the cluster may become read-only. • Snapshot creation affects disk I/O performance. You are advised to create snapshots during off-peak hours. • During the snapshot creation, some intermediate files are retained, which occupy extra disk space. Therefore, avoid peak hours and ensure that the disk capacity usage is less than 70%. • Snapshots can be restored to the current cluster, but logical clusters and resource pools cannot be restored to the current cluster.
Version Upgrade	• Cluster 8.1.1 and later versions allow users to deliver cluster upgrade operations on the console. • If the cluster is interrupted for a long time due to a node fault or system upgrade, contact technical support.
Data Migration	• Ensure that no Chinese characters are contained in paths used for importing data to or exporting data from OBS. • Data cannot be imported to or exported from OBS across regions. Ensure that OBS and the GaussDB(DWS) cluster are in the same region.

Item	Constraint
Failover	• When the DR task is created, the snapshot function of the production cluster is normal, but that of the DR cluster is disabled. Besides, snapshot restoration of both clusters is disabled. • DR does not synchronize data from external sources. • DR management refers to dual-cluster DR under the same tenant. • The DR cluster and the production cluster must be logically homogeneous and in the same type and version. • The production cluster and DR cluster used for intra-region DR must be in the same VPC. • In intra-region DR, after services are switched over from the production cluster to the DR cluster, the bound ELB is automatically switched to the new production cluster. During the switchover, the connection is interrupted for a short period of time. Do not run service statements to write data during the switchover. • During intra-region DR, the EIP, intranet domain name, and connection IP address of the original production cluster are not automatically switched with the cluster switchover. The EIP, domain name, or IP address used for connection in the service system need to be switched to the new cluster.
Hot and Cold Data Management	• The hybrid data warehouse (standalone) does not support cold and hot partition switchover. • Currently, cold and hot tables support only column-store partitioned tables of version 2.0. Foreign tables do not support cold and hot partitions. • Only hot data can be switched to cold data. Cold data cannot be switched to hot data. • A partition on a DN is either hot or cold. For a partition across DNs, its data on some DNs may be hot, and some may be cold. • Only the cold and hot switchover policies can be modified. The tablespace of cold data in cold and hot tables cannot be modified.
Stream Data Warehouse	• Time series tables do not support UPDATE, UPSERT, primary keys, or PCKs. • To create a time series table, you must have the **USAGE** permission on schema cstore. • When you modify the **enable_delta** parameter of a time series table, other ALTER operations cannot be performed.

10 Restricted Functions

Context

GaussDB(DWS) depends on services such as Elastic Load Balance (ELB) and Object Storage Service (OBS). This section describes the constraints on using DWS without ELB or OBS.

Restricted Functions in the Non-OBS Scenario

Table 10-1 Restricted functions in the non-OBS scenario

Function	Support
Cluster snapshot	Users cannot manually create snapshots, configure automatic snapshot policies, or restore snapshots.
Audit log storage	Users cannot record the audit logs of specific operations, involving audit log retention policies, unauthorized access, as well as DML, DDL, **SELECT** and **COPY** operations performed on stored procedures and database objects. Key operations, such as cluster creation and restart, cannot be recorded on the management console.
Load snapshot	Users cannot create load snapshots to record the cluster load data in a specified period.

Restricted Functions in the Non-ELB Scenario

If the internal 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.

An ELB distributes access traffic to multiple ECSs for traffic control based on forwarding policies. It improves the fault tolerance capability of application programs.

With ELB health checks, CN requests of a cluster can be quickly forwarded to normal CNs. If a CN is faulty, the workload can be immediately shifted to a healthy node, minimizing cluster access faults.

ECS Cluster Constraints

GaussDB(DWS) clusters provisioned by ECS can be used only for non-production environments.

11 Technical Specifications

This section describes the technical specifications of GaussDB(DWS) in different versions.

Table 11-1 Technical specifications of GaussDB(DWS) 8.1.3 – 8.3.0

Technical Specifications	Maximum Value of 8.1.3	Maximum Value of 8.2.0	Maximum Value of 8.2.1	Maximum Value of 8.3.0
Number of cluster nodes	2048	2048	2048	2048
Number of concurrent connections	Number of concurrent complex queries in minutes: 80 Number of short queries in seconds: 500 Number of concurrent short transactions in milliseconds: 5000	Number of concurrent complex queries in minutes: 80 Number of short queries in seconds: 500 Number of concurrent short transactions in milliseconds: 5000	Number of concurrent complex queries in minutes: 80 Number of short queries in seconds: 500 Number of concurrent short transactions in milliseconds: 5000	Number of concurrent complex queries in minutes: 80 Number of short queries in seconds: 500 Number of concurrent short transactions in milliseconds: 5000
Cluster data capacity	20 PB	20 PB	20 PB	20 PB
Size of a single table	1 PB	1 PB	1 PB	1 PB
Size of data in each row	1 GB	1 GB	1 GB	1 GB

Technical Specifications	Maximum Value of 8.1.3	Maximum Value of 8.2.0	Maximum Value of 8.2.1	Maximum Value of 8.3.0
Number of columns in a single table: (excluding Hudi tables)	1600	1600	1600	1600
Number of columns in a Hudi table	-	-	5000	5000
Number of partitions of the partitioned table	32,768	32,768	32,768	32,768
RTO after a SPOF	60s	60s	60s	60s
RPO after a SPOF	0	0	0	0
RTO after cluster DR switchover	60min	60min	60min	60min
RPO after cluster DR switchover	60min	60min	60min	60min

Table 11-2 Technical specifications of GaussDB(DWS) 8.0.x-8.1.2

Technical Specifications	Maximum Value of 8.0.*x*	Maximum Value of 8.1.0	Maximum Value of 8.1.1	Maximum Value of 8.1.2
Data capacity	10 PB	10 PB	20 PB	20 PB
Number of cluster nodes	256	256	2048	2048
Size of a single table	1 PB	1 PB	1 PB	1 PB
Size of data in each row	1 GB	1 GB	1 GB	1 GB
Size of a single column in each record	1 GB	1 GB	1 GB	1 GB

Technical Specifications	Maximum Value of 8.0.*x*	Maximum Value of 8.1.0	Maximum Value of 8.1.1	Maximum Value of 8.1.2
Number of records in each table	2^{55}	2^{55}	2^{55}	2^{55}
Number of columns in each table	1600	1600	1600	1600
Number of indexes in each table	Unlimited	Unlimited	Unlimited	Unlimited
Number of columns in the index of each table	32	32	32	32
Number of constraints in each table	Unlimited	Unlimited	Unlimited	Unlimited
Number of concurrent connections	Number of concurrent complex queries in minutes: 60 Number of concurrent short transactions in milliseconds: 5000	Number of concurrent complex queries in minutes: 60 Number of concurrent short transactions in milliseconds: 5000	Number of concurrent complex queries in minutes: 80 Number of concurrent short transactions in milliseconds: 5000	Number of concurrent complex queries in minutes: 80 Number of short queries in seconds: 500 Number of concurrent short transactions in milliseconds: 5000
Number of partitions in a partitioned table	32,768	32,768	32,768	32,768
Size of each partition in a partitioned table	1 PB	1 PB	1 PB	1 PB

Technical Specifications	Maximum Value of 8.0.*x*	Maximum Value of 8.1.0	Maximum Value of 8.1.1	Maximum Value of 8.1.2
Number of records in each partition in a partitioned table	2^{55}	2^{55}	2^{55}	2^{55}

NOTE

The maximum number of concurrent connections is based on the data warehouse with the cloud disk flavor of 48 vCPUs or 64 vCPUs. For example, **dwsk.12xlarge (48 vCPU | 384GB | 24000GB SSD)** or **dwsx2.16xlarge.m7 (64 vCPU | 512GB | 32000GB SSD)** of a standard data warehouse.

