

数据仓库服务

产品介绍

文档版本 01
发布日期 2024-04-16



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

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

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

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

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

目录

1 什么是数据仓库服务	1
2 数据仓库类型	6
3 数据仓库规格	10
4 产品优势	19
5 应用场景	21
6 产品功能	25
7 基本概念	31
8 与其他云服务的关系	33
9 安全	36
9.1 责任共担.....	36
9.2 身份认证与访问控制.....	37
9.2.1 资源访问控制（IAM 权限控制）.....	37
9.2.2 数据库访问实现权限分层.....	38
9.2.3 GRANT 授权和 REVOKE 撤销授权.....	39
9.2.4 行级访问控制.....	39
9.3 数据保护技术.....	40
9.3.1 数据备份.....	40
9.3.2 透明加密.....	40
9.3.3 SSL 传输加密.....	42
9.3.4 数据脱敏.....	42
9.3.5 使用函数进行加密.....	42
9.4 审计与日志.....	43
9.5 服务韧性.....	44
9.6 监控安全风险.....	45
9.7 认证证书.....	45
9.8 安全公告.....	46
9.8.1 漏洞修复说明.....	47
10 GaussDB(DWS)权限管理	48
11 如何访问 GaussDB(DWS)	52

12 计费说明.....	54
13 使用限制.....	57
14 配额说明.....	61
15 技术指标.....	62
16 版本说明.....	64

1 什么是数据仓库服务

数据仓库服务 GaussDB (DWS) 是一种基于华为云基础架构和平台的在线数据分析处理数据库，提供即开即用、可扩展且完全托管的分析型数据库服务，兼容 ANSI/ISO 标准的 SQL92、SQL99 和 SQL 2003 语法，同时兼容 PostgreSQL/Oracle/Teradata/MySQL 等数据库生态，为各行业 PB 级海量大数据分析提供有竞争力的解决方案。

DWS 提供标准数仓 (DWS 2.0)、标准数仓 (DWS 3.0)、IoT 数仓和实时数仓等多种产品形态，围绕企业级内核、实时分析、协同计算、融合分析、云原生五大方向构筑业界第一数据仓库。详情请参见[数据仓库类型](#)。

- **标准数仓 (DWS 2.0)**：面向数据分析场景，为用户提供高性能、高扩展、高可靠、高安全、易运维的企业级数仓服务，支持 2048 节点、20PB 级超大规模数据分析能力。支持冷热数据分析，存储、计算弹性伸缩，并按需、按量计价，为用户提供弹性灵活、极致性价比的体验。适用于“库、仓、市、湖”一体化的融合分析业务。
- **标准数仓 (DWS 3.0)**：采用存算分离云原生架构，极致性价比，支持冷热数据分析，存储、计算弹性伸缩，无限算力、无限容量，并按需、按量计价，适用 OLAP 分析场景。
- **IoT 数仓**：在标准数仓基础上，通过融合流、时序引擎，为用户提供实时数据接入和高并发实时分析能力。适用于物联网、IoT 等实时分析场景。
- **实时数仓**：在大规模数据查询和分析能力基础上，提供高并发、高性能、低时延的事务处理能力。适用于 HTAP 混合负载场景，“一库两用，生产即分析”。

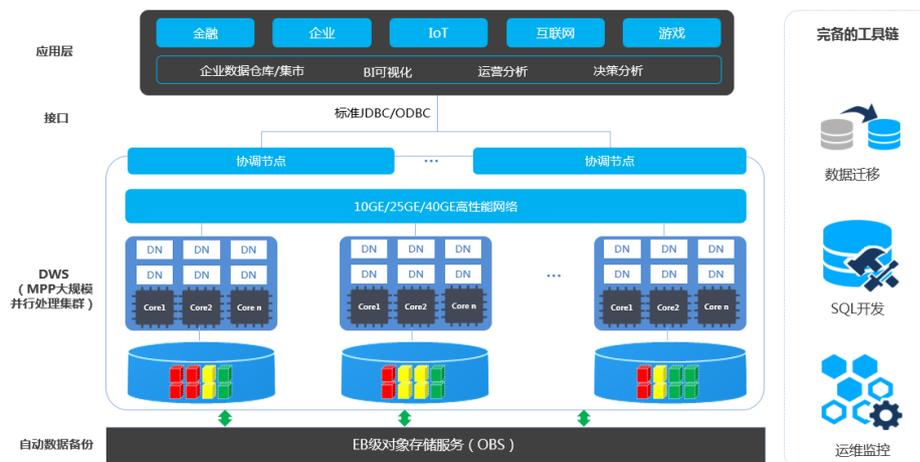
GaussDB (DWS) 可广泛应用于金融、车联网、政企、电商、能源、电信等多个领域，已连续两年入选 Gartner 发布的数据管理解决方案魔力象限，相比传统数据仓库，性价比提升数倍，具备大规模扩展能力和企业级可靠性。

此外，GaussDB (DWS) 也支持物理机部署模式，详情请参见[物理机部署文档](#)。

产品架构

GaussDB (DWS) 基于 Shared-nothing 分布式架构，具备 MPP (Massively Parallel Processing) 大规模并行处理引擎，由众多拥有独立且互不共享的 CPU、内存、存储等系统资源的逻辑节点组成。在这样的系统架构中，业务数据被分散存储在多个节点上，数据分析任务被推送到数据所在位置就近执行，并行地完成大规模的数据处理工作，实现对数据处理的快速响应。

图 1-1 产品架构



应用层

数据加载工具、ETL (Extract-Transform-Load) 工具、以及商业智能BI工具、数据挖掘和分析工具，均可以通过标准接口与GaussDB(DWS)集成。GaussDB(DWS)兼容PostgreSQL生态，且SQL语法进行了兼容MySQL、Oracle和Teradata的处理。应用只需做少量改动即可向GaussDB(DWS)平滑迁移。

接口

支持应用程序通过标准JDBC和ODBC连接GaussDB(DWS)。

GaussDB(DWS)

一个GaussDB(DWS)集群由多个在相同子网中的相同规格的节点组成，共同提供服务。集群的每个DN负责存储数据，其存储介质是磁盘。协调节点 (Coordinator) 负责接收来自应用的访问请求，并向客户端返回执行结果，此外，协调节点还负责分解任务，并调度任务分片在各DN上并行执行。

自动数据备份

支持将集群快照自动备份到EB级对象存储服务OBS (Object Storage Service) 中，方便利用业务空闲期对集群做周期备份以保证集群异常后的数据恢复。

快照是GaussDB(DWS)集群在某一时间点的完整备份，记录了这一时刻指定集群的所有配置数据和业务数据。

工具链

提供了数据并行加载工具GDS (General Data Service)、SQL语法迁移工具DSC (Database Schema Converter)、SQL开发工具Data Studio、迁移工具GDS-Kafka，并支持通过控制台对集群进行运维监控。

集群逻辑架构

GaussDB(DWS)集群逻辑架构如图1-2所示。实例的详细介绍请参见表1-1。

图 1-2 集群逻辑架构图

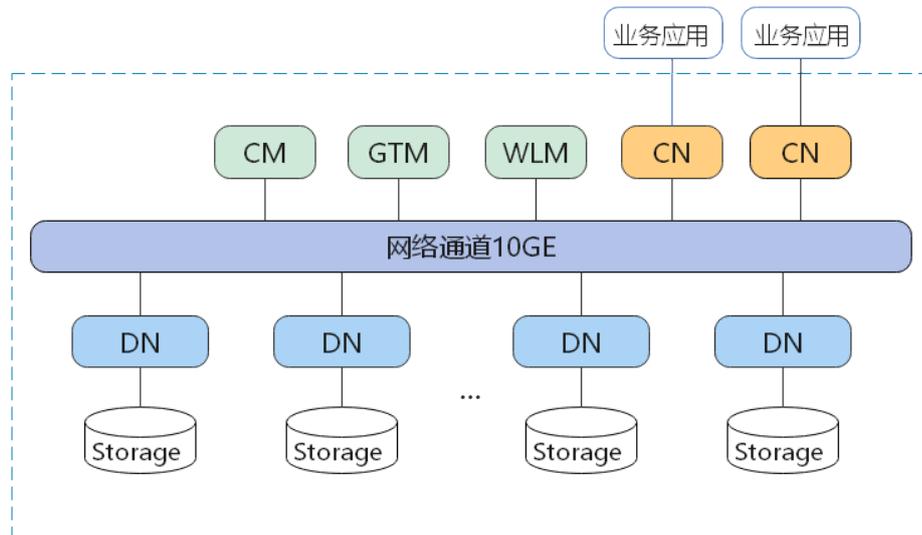


表 1-1 集群架构说明

名称	描述	说明
CM	集群管理模块（Cluster Manager）。管理和监控分布式系统中各个功能单元和物理资源的运行情况，确保整个系统的稳定运行。	<p>CM由CM Agent、OM Monitor和CM Server组成。</p> <ul style="list-style-type: none"> CM Agent：负责监控所在主机上主备 GTM、CN、主备DN的运行状态并将状态上报给CM Server。同时负责执行CM Server下发的仲裁指令。集群的每台主机上均有CM Agent进程。 OM Monitor：看护CM Agent的定时任务，其唯一的任务是在CM Agent停止的情况下将CM Agent重启。如果CM Agent重启不了，则整个主机不可用，需要人工干预。 <p>说明 CM Agent重启的情况很少发生，如果出现可能是因为系统资源不够用导致无法启动新进程。</p> <ul style="list-style-type: none"> CM Server：根据CM Agent上报的实例状态判定当前状态是否正常，是否需要修复，并下发指令给CM Agent执行。 <p>GaussDB(DWS)提供了CM Server的主备实例方案，以保证集群管理系统本身的高可用性。正常情况下，CM Agent连接主CM Server，在主CM Server发生故障的情况下，备CM Server会主动升为主CM Server，避免出现CM Server单点故障。</p>

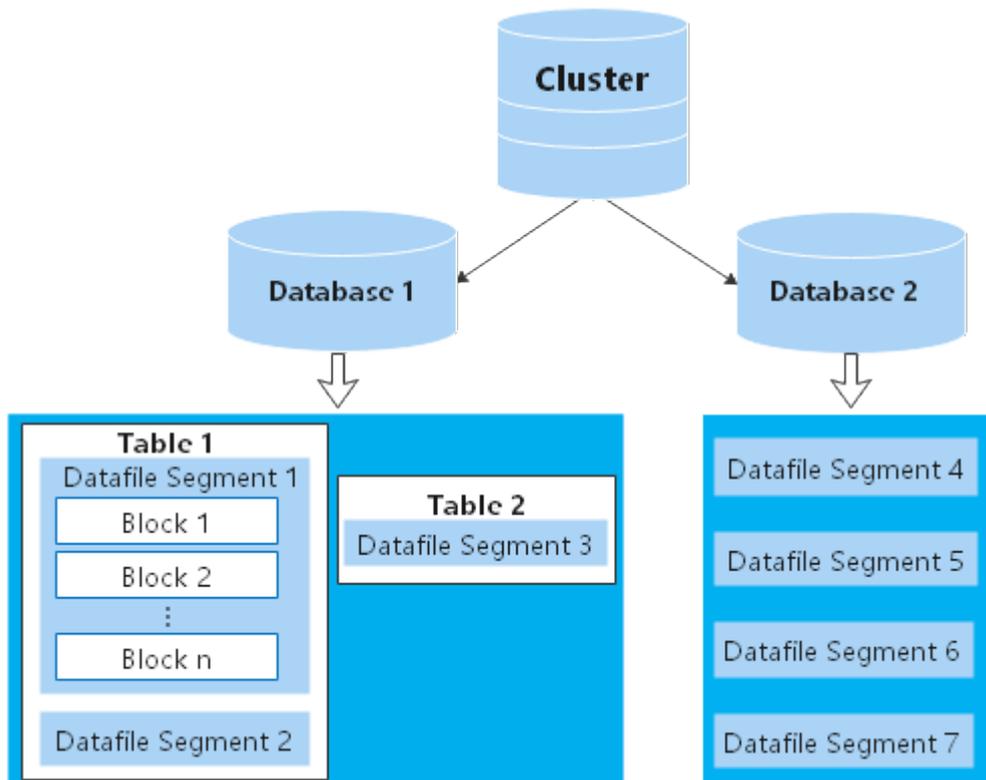
名称	描述	说明
GTM	全局事务管理器（Global Transaction Manager），负责生成和维护全局事务ID、事务快照、时间戳等全局唯一的信息。	整个集群只有一组GTM：主、备GTM各一个。
WLM	工作负载管理器（Workload Manager）。控制系统资源的分配，防止过量业务负载对系统的冲击而导致业务拥塞和系统崩溃。	不同于集群中的实例（GTM、CM、CN、DN）模块，不需要在安装过程中指定主机名称。安装程序会自动在各主机上安装此模块。
CN	协调节点（Coordinator）。负责接收来自应用的访问请求，并向客户端返回执行结果；负责分解任务，并调度任务分片在各DN上并行执行。	<p>集群中，CN有多个并且CN的角色是对等的（执行DML语句时连接到任何一个CN都可以得到一致的结果）。只需要在CN和应用程序之间增加一个负载均衡器，使得CN对应用是透明的。CN故障时，由负载均衡自动路由连接到另外一个CN，请参见集群绑定和解绑ELB。</p> <p>当前分布式事务框架下无法避免CN之间的互连，为了减少GTM上线程过多导致负载过大，建议CN配置数目≤10个。</p> <p>GaussDB(DWS)通过CCN（Central Coordinator）负责集群内的资源全局负载控制，以实现自适应的动态负载管理。CM在第一次集群启动时，通过集群部署形式，选择编号最小的CN作为CCN。若CCN故障之后，由CM选择新的CCN进行替换。</p>
DN	数据节点（Datanode）。负责存储业务数据（支持行存、列存、混合存储）、执行数据查询任务以及向CN返回执行结果。	<p>在集群中，DN有多个。每个DN存储了一部分数据。GaussDB(DWS)对DN提供了高可用方案：主DN、备DN、从备DN。三者的工作原理如下：</p> <ul style="list-style-type: none"> 主、备DN同步数据期间，如果主DN突然故障不可用，备DN会升为主DN。 在原主DN恢复前，新升为主的DN会将数据日志同步到从备DN。 原主DN恢复后将成为备DN，并且会使用“从备DN”上的数据日志恢复异常期间的数据。 <p>也就是说从备DN永远只作为从备使用，不会因为主DN或备DN故障而升级为主DN或备DN，从备DN只存放原主DN故障时，新升为主的DN同步到从备DN的Xlog数据和数据通道复制产生的数据。因此从备DN不额外占用存储资源，相比传统三副本节约了三分之一的存储空间。</p>
Storage	服务器的本地存储资源，持久化存储数据。	-

集群的每个DN上负责存储数据，其存储介质也是磁盘。图1-3从逻辑上介绍了每个DN上都有哪些对象，以及这些对象之间的关系，其中：

- Database，即数据库，用于管理各类数据对象，各数据库间相互隔离。
- Datafile Segment，即数据文件，通常每张表只对应一个数据文件。如果某张表的数据大于1GB，则会分为多个数据文件存储。
- Table，即表，每张表只能属于一个数据库。
- Block，即数据块，是数据库管理的基本单位，默认大小为8KB。

数据有三种分布方式，可以在建表的时候指定：REPLICATION、ROUNDROBIN、HASH。

图 1-3 数据库逻辑结构图



2 数据仓库类型

产品类型概述

- **标准数仓 (DWS 2.0)**：高性价比，支持冷热数据分析，存储、计算弹性伸缩，无限算力、无限容量，并按需、按量计价。适用于“库、仓、市、湖”一体化的融合分析业务，是OLAP分析场景的首选。
- **标准数仓 (DWS 3.0)**：采用存算分离云原生架构，极致性价比，支持冷热数据分析，存储、计算弹性伸缩，无限算力、无限容量，并按需、按量计价，适用OLAP分析场景。
- **实时数仓**：在大规模数据查询和分析能力基础上，提供高并发、高性能、低时延、低成本的事务处理能力。适用于HTAP混合负载场景，“一库两用，生产即分析”，支持单机部署和集群部署两种部署方式。
- **IoT数仓**：在标准数仓基础上，提供高效的时序计算和IoT分析能力，支持实时和历史数据关联，内置时序算子，最高40x压缩。适用于物联网IoT等实时分析场景。

📖 说明

GaussDB(DWS)数仓之间暂不支持互相访问，用户可通过创建的OBS外表，将两个数据库关联在同一个数据目录下进行数据查询。

功能特性对比

表 2-1 功能特性对比

功能模块	功能模块	标准数仓 (DWS 2.0)	标准数仓 (DWS 3.0)	IoT数仓	实时数仓 (集群部署)	实时数仓 (单机部署)
总览	资源	支持	支持	支持	支持	支持
	告警	支持	支持	支持	支持	支持
	近期事件	支持	支持	支持	支持	支持
	集群监控指标 (DMS)	支持	支持	支持	支持	不支持

功能模块	功能模块	标准数仓 (DWS 2.0)	标准数仓 (DWS 3.0)	IoT数仓	实时数仓 (集群部署)	实时数仓 (单机部署)
集群管理	DAS登录	支持	支持	支持	支持	支持
	监控面板 (DMS)	支持	支持	支持	支持	不支持
	查看监控指标 (Cloud Eye)	支持	支持	支持	支持	不支持
	重启	支持	支持	支持	支持	支持
	启动	不支持	不支持	不支持	不支持	不支持
	停止	不支持	不支持	不支持	不支持	不支持
	扩容	支持	支持	支持	支持	不支持
	缩容	支持	支持	支持	支持	不支持
	重分布	支持	支持	支持	支持	不支持
	查看重分布详情	支持	支持	支持	支持	不支持
	弹性变更规格	支持	支持	支持	支持	支持
	经典变更规格	支持	不支持	不支持	支持	不支持
	重置密码	支持	支持	支持	支持	支持
	创建快照	支持	支持	支持	支持	不支持
	解除只读	支持	支持	支持	支持	支持
	删除	支持	支持	支持	支持	支持
	管理CN节点	支持	支持	支持	支持	不支持
	磁盘扩容	支持	支持	支持	支持	支持
集群详情	基本信息	支持	支持	支持	支持	支持
	ELB负载均衡	支持	支持	支持	支持	不支持
	资源管理	支持	支持	支持	支持	不支持
	智能运维	支持	支持	支持	支持	不支持
	逻辑集群	支持	支持	支持	支持	不支持

功能模块	功能模块	标准数仓 (DWS 2.0)	标准数仓 (DWS 3.0)	IoT数仓	实时数仓 (集群部署)	实时数仓 (单机部署)
	快照	支持	支持	支持	支持	不支持
	参数修改	支持	支持	支持	支持	支持
	安全设置	支持	支持	支持	支持	支持
	MRS数据源	支持	支持	支持	支持	不支持
	标签	支持	支持	支持	支持	支持
	节点管理	支持	支持	支持	支持	不支持
	升级管理	支持	支持	支持	支持	支持
	日志服务	支持	支持	支持	支持	支持
	用户管理	支持	支持	支持	支持	支持
集成	数据迁移	支持	支持	支持	支持	支持
容灾管理	容灾管理	支持	不支持	支持	支持	不支持
快照管理	恢复	支持	支持	支持	支持	不支持
	删除	支持	支持	支持	支持	不支持
	复制	支持	支持	支持	支持	不支持
事件管理	事件管理 (通用)	支持	支持	支持	支持	支持
告警管理	告警管理	支持	支持	支持	支持	支持
连接客户端	连接客户端	支持	支持	支持	支持	支持
其他模块	巡检	支持	支持	支持	支持	不支持
	智能运维	支持	支持	支持	支持	不支持
	节点修复	支持	支持	支持	支持	不支持
	租户侧温备	支持	支持	支持	支持	不支持
	OpenApi	支持	不支持	支持	支持	不支持

说明

- 实时数仓（单机部署）不支持分布式模式，因此不支持扩缩容、逻辑集群、资源管理等操作。
- 实时数仓（单机部署）不支持快照、监控面板、MRS数据源等功能。
- 实时数仓（单机部署）中，原有指定分布列语法将被忽略，此外不支持job定时任务、SEQUENCE类型、HDFS/OBS外表导入导出（其中OBS外表导入导出8.2.0及以上集群版本支持）、多温存储、自增分区表等功能。
- DWS 3.0集群仅9.0.2及以上集群版本支持快照功能。

3 数据仓库规格

GaussDB(DWS)的规格按照产品类型分为标准数仓、实时数仓和IoT数仓。其中实时数仓还包含单机版模式。各产品类型的不同差异，详情请参见[数据仓库类型](#)。

标准数仓（DWS 2.0）规格

- 标准数仓（DWS 2.0）云盘规格，该规格弹性伸缩，无限算力、无限容量，规格详情请参见[表3-1](#)。
- 标准数仓（DWS 2.0）本地盘规格，该规格存储容量固定，不能够进行磁盘扩容，只能进行节点扩容，规格详情请参见[表3-2](#)。

📖 说明

步长指在集群变配过程中增大或减小磁盘大小的间隔大小。用户在进行操作时需要按照对应规格的存储步长来选择。

表 3-1 标准数仓（DWS 2.0）云盘规格

规格名称	CP U 架 构	vC P U	内 存 (G B)	单 节 点 存 储 容 量	默 认 存 储	步 长 (G B)	建 议 存 储	DN 数 量	使用场景
dwsx2.x large.m 7	X86	4	32	20GB ~ 2000G B	100	10	800	1	DWS的入门规格，一般用于测试、学习环境或者小型分析系统。
dwsk2. xlarge	AR M	4	32	20GB ~ 2000G B	100	10	800	1	
dwsx2.x large.m 7n	X86	4	32	20GB ~ 2000G B	100	10	800	1	

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	默认存储	步长 (GB)	建议存储	DN数量	使用场景
dwsx2.2xlarge.m7	X86	8	64	100GB ~ 4000GB	200	100	1600	1	适用于中小型企业内部数据仓库构建和报表分析。
dwsk2.2xlarge	ARM	8	64	100GB ~ 4000GB	200	100	1600	1	
dwsx2.2xlarge.m7n	X86	8	64	100GB ~ 4000GB	200	100	1600	1	
dwsx2.4xlarge.m7	X86	16	128	100GB ~ 8000GB	400	100	3200	1	
dwsk2.4xlarge	ARM	16	128	100GB ~ 8000GB	400	100	3200	1	
dwsx2.8xlarge.m7	X86	32	256	100GB ~ 16000GB	800	100	6400	2	推荐在生产环境下使用，适用于绝大部分企业大数据量OLAP分析系统，BI报表，可视化大屏场景。
dwsk2.8xlarge	ARM	32	256	100GB ~ 16000GB	800	100	6400	2	
dwsx2.8xlarge.m7n	X86	32	256	100GB ~ 16000GB	800	100	6400	2	
dwsk2.12xlarge	ARM	48	384	100GB ~ 24000GB	1200	100	9600	4	有着极致的性能，适用于高吞吐数仓加工，高并发在线查询生产环境。
dwsx2.16xlarge.m7	X86	64	512	100GB ~ 32000GB	1600	100	12800	4	

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	默认存储	步长 (GB)	建议存储	DN数量	使用场景
dwsx2.16xlarge.m7n	X86	64	512	100GB ~ 32000GB	1600	100	12800	4	

表 3-2 标准数仓 (DWS 2.0) 本地盘规格

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	DN数量	使用场景
dws2.olap.4xlarge.i3	X86	16	128	1490GB	1	推荐在生产环境下使用，适用于绝大部分企业大数据量OLAP分析系统，BI报表，可视化大屏等场景。 有着极致的性能，适用于高吞吐数仓加工，高并发在线查询生产环境。
dws2.olap.4xlarge.ki1	ARM	16	64	2980GB	1	
dws2.olap.8xlarge.i3	X86	32	256	2980GB	2	
dws2.olap.8xlarge.ki1	ARM	32	128	5960GB	2	
dws2.olap.16xlarge.i3	X86	64	512	5960GB	4	
dws2.olap.16xlarge.ki1	ARM	64	228	11921GB	4	

标准数仓 (DWS 3.0) 规格

- 标准数仓 (DWS 3.0) 云盘规格，该规格弹性伸缩，无限算力、无限容量，规格详情请参见[表3-3](#)。
- 标准数仓 (DWS 3.0) 本地盘规格，该规格存储容量固定，不能够进行磁盘扩容和规格变更，只能进行节点扩容，规格详情请参见[表3-4](#)。

表 3-3 标准数仓（DWS 3.0）云盘规格

规格名称	CPU架构	vCPU	内存(GB)	单节点存储容量	步长(GB)	DN数量	使用场景
dwsx3.4U 16G.4DP U	X86	4	16	20GB~200 0GB	10	1	DWS的入门规格，一般用于测试、学习环境或者小型分析系统。
dwsk3.4U 16G.4DP U	ARM	4	16	20GB~200 0GB	10	1	
dwsx3.8U 32G.8DP U	X86	8	32	100GB~40 00GB	100	1	适用于中小企规模企业内部数据仓库构建和报表分析。
dwsk3.8U 32G.8DP U	ARM	8	32	100GB~40 00GB	100	1	
dwsx3.16 U64G.16 DPU	X86	16	64	100GB~80 00GB	100	1	推荐在生产环境下使用，适用于绝大部分企业大数据量OLAP分析系统，BI报表，可视化大屏场景。
dwsk3.16 U64G.16 DPU	ARM	16	64	100GB~80 00GB	100	1	
dwsx3.32 U128G.32 DPU	X86	32	128	100GB~16 000GB	100	2	
dwsk3.32 U128G.32 DPU	ARM	32	128	100GB~16 000GB	100	2	
dwsk3.48 U192G.48 DPU	ARM	48	192	200GB~24 000GB	100	4	
dwsx3.64 U256G.64 DPU	X86	64	256	200GB~32 000GB	100	4	

表 3-4 标准数仓（DWS 3.0）本地盘规格

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	DN 数量	使用场景
dws3.16U1 28G.i7.16DP U	X86	16	128	2980GB	1	推荐在生产环境下使用，适用于绝大部分企业大数据量 OLAP分析系统，BI报表，可视化大屏等场景。 有着极致的性能，适用于高吞吐数仓加工，高并发在线查询生产环境。
dws3.16U6 4G.ki.16DP U	ARM	16	64	5960GB	1	
dws3.32U2 56G.i7.32DP U	X86	32	256	5960GB	2	
dws3.32U1 28G.ki.1.32D PU	ARM	32	128	11920GB	2	
dws3.64U5 12G.i7.64DP U	X86	64	512	11920GB	4	
dws3.64U2 28G.ki.1.64D PU	ARM	64	228	23840GB	4	

实时数仓规格

- 创建实时数仓时，用户可根据需求选择集群部署或单机部署：
 - 集群部署：创建集群时可选择带有h的节点规格，规格详情请参见表3-5，此时实时数仓支持集群部署，支持节点扩容、资源池等功能。
 - 单机部署：创建集群时可选择带有h1的节点规格，规格详情请参见表3-6，此时实时数仓只支持单机部署，单机形态不提供高可用服务，因此存储成本可减半，单机模式服务可用性通过ECS自动重建实现，数据可靠性通过EVS多副本机制保证。单机形态性价比更高，建议用于轻量化业务。

表 3-5 实时数仓（集群形态）规格

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	步长 (GB)	DN 数量	使用场景
dwsx2.h. xlarge.4.c 7	X86	4	16	20GB ~ 2000GB	20	1	DWS的入门规格，一般用于测试、学习环境或者小型分析系统。
dwsk2.h. xlarge.4.k c1	ARM	4	16	20GB ~ 2000GB	20	1	

规格名称	CPU架构	vCPU	内存(GB)	单节点存储容量	步长(GB)	DN数量	使用场景
dwsx2.h.xlarge.4.c7n	X86	4	16	20GB ~ 2000GB	20	1	适用于中小企规模企业内部数据仓库构建和报表分析。
dwsx2.h.2xlarge.4.c6	X86	8	32	100GB ~ 4000GB	100	1	
dwsx2.h.2xlarge.4.c7	X86	8	32	100GB ~ 4000GB	100	1	
dwsk2.h.2xlarge.4.kc1	ARM	8	32	100GB ~ 4000GB	100	1	
dwsx2.h.2xlarge.4.c7n	X86	8	32	100GB ~ 4000GB	100	1	
dwsx2.h.4xlarge.4.c7	X86	16	64	100GB ~ 8000GB	100	1	推荐在生产环境下使用，适用于绝大部分企业大数据量OLAP分析系统，BI报表，可视化大屏等场景。
dwsk2.h.4xlarge.4.kc1	ARM	16	64	100GB ~ 8000GB	100	1	
dwsx2.h.4xlarge.4.c7	X86	16	64	100GB ~ 8000GB	100	1	
dwsx2.h.8xlarge.4.c7	X86	32	128	100GB ~ 16000GB	100	2	
dwsk2.h.8xlarge.4.kc1	ARM	32	128	100GB ~ 16000GB	100	2	
dwsx2.h.8xlarge.4.c7n	X86	32	128	100GB ~ 16000GB	100	2	有着极致的性能，适用于高吞吐数仓加工，高并发在线查询生产环境。
dwsx2.h.12xlarge.4.c7	X86	48	192	100GB ~ 24000GB	100	4	
dwsk2.h.12xlarge.4.kc1	ARM	48	192	100GB ~ 24000GB	100	4	
dwsx2.h.16xlarge.4.c7	X86	64	256	100GB ~ 32000GB	100	4	
dwsk2.h.16xlarge.4.kc1	ARM	64	256	100GB ~ 32000GB	100	4	

规格名称	CPU架构	vCPU	内存(GB)	单节点存储容量	步长(GB)	DN数量	使用场景
dwsx2.h16xlarge.4.c7n	X86	64	256	100GB ~ 32000GB	100	4	

表 3-6 实时数仓（单机形态）规格

规格名称	CPU架构	vCPU	内存(GB)	单节点存储容量	步长(GB)	DN数量	使用场景
dwsx2.h1.xlarge.2.c7	X86	4	8	20GB ~ 2000GB	20	1	DWS的入门规格，一般用于测试、学习环境或者小型分析系统。
dwsx2.h1.xlarge.2.kc1	ARM	4	8	20GB ~ 2000GB	20	1	
dwsx2.h1.xlarge.2.c7n	X86	4	8	20GB ~ 2000GB	20	1	
dwsx2.h1.2xlarge.4.c7	X86	8	32	100GB ~ 4000GB	100	1	适用于中小企规模企业内部数据仓库构建和报表分析。
dwsx2.h1.2xlarge.4.kc1	ARM	8	32	100GB ~ 4000GB	100	1	
dwsx2.h1.2xlarge.4.c7n	X86	8	32	100GB ~ 4000GB	100	1	
dwsx2.h1.4xlarge.4.c7	X86	16	64	100GB ~ 8000GB	100	1	推荐在生产环境下使用，适用于绝大部分企业大数据量OLAP分析系统，BI报表，可视化大屏场景。
dwsx2.h1.4xlarge.4.kc1	ARM	16	64	100GB ~ 8000GB	100	1	
dwsx2.h1.4xlarge.4.c7n	X86	16	64	100GB ~ 8000GB	100	1	
dwsx2.h1.8xlarge.4.c7	X86	32	128	100GB ~ 16000GB	100	2	

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	步长 (GB)	DN数量	使用场景
dwsk2.h1.8xlarge.4.kc1	ARM	32	128	100GB ~ 16000GB	100	2	有着极致的性能，适用于高吞吐数仓加工，高并发在线查询生产环境。
dwsx2.h1.8xlarge.4.c7n	X86	32	128	100GB ~ 16000GB	100	2	
dwsk2.h1.12xlarge.4.kc1	ARM	48	192	100GB ~ 24000GB	100	4	
dwsx2.h1.16xlarge.4.c7	X86	64	256	100GB ~ 32000GB	100	4	
dwsx2.h1.16xlarge.4.c7n	X86	64	256	100GB~32000GB	100	4	

IoT 数仓规格

表 3-7 IoT 数仓规格

规格名称	CPU架构	vCPU	内存 (GB)	单节点存储容量	步长 (GB)	DN数量	使用场景
dwsx2.rt.xlarge.m7	X86	4	32	20GB ~ 2000GB	20	1	DWS的入门规格，一般用于测试、学习环境或者小型分析系统。
dwsk2.rt.xlarge.km1	ARM	4	32	20GB ~ 2000GB	20	1	
dwsx2.rt.xlarge.m7n	X86	4	32	20GB ~ 2000GB	20	1	
dwsx2.rt.2xlarge.m7	X86	8	64	100GB ~ 4000GB	100	1	适用于中小企规模企业内部数据仓库构建和报表分析。
dwsk2.rt.2xlarge.km1	ARM	8	64	100GB ~ 4000GB	100	1	

规格名称	CPU架构	vCPU	内存(GB)	单节点存储容量	步长(GB)	DN数量	使用场景
dwsx2.rt.2xlarge.m7n	X86	8	64	100GB ~ 4000GB	100	1	
dwsx2.rt.4xlarge.m7	X86	16	128	200GB ~ 8000GB	100	1	
dwsk2.rt.4xlarge.km1	ARM	16	128	200GB ~ 8000GB	100	1	
dwsx2.rt.8xlarge.m7	X86	32	256	100GB ~ 16000GB	100	2	推荐在生产环境下使用，适用于绝大部分企业大数据量OLAP分析系统，BI报表，可视化大屏等场景。
dwsk2.rt.8xlarge.km1	ARM	32	256	100GB ~ 16000GB	100	2	
dwsx2.rt.8xlarge.m7n	X86	32	256	100GB ~ 16000GB	100	2	
dwsk2.rt.12xlarge.km1	ARM	48	384	100GB ~ 24000GB	100	4	有着极致的性能，适用于高吞吐数仓加工，高并发在线查询生产环境。
dwsx2.rt.16xlarge.m7	X86	64	512	100GB ~ 32000GB	100	4	
dwsx2.rt.16xlarge.m7n	X86	64	512	100GB ~ 32000GB	100	4	

4 产品优势

GaussDB(DWS)数据库内核自主研发，兼容PostgreSQL 9.2.4的数据库内核引擎，从单机OLTP数据库改造为企业级MPP（大规模并行处理）架构的OLAP分布式数据库，其主要面向海量数据分析场景。

GaussDB(DWS)与传统数据仓库相比，主要有以下特点与显著优势，可解决多行业超大规模数据处理与通用平台管理问题：

易使用

- 一站式可视化便捷管理

GaussDB(DWS)让您能够轻松完成从项目概念到生产部署的整个过程。通过使用GaussDB(DWS)管理控制台，您不需要安装数据仓库软件，也不需要部署数据仓库服务器，就可以在几分钟之内获得高性能、高可靠的企业级数据仓库集群。

您只需单击几下鼠标，就可以轻松完成应用程序与数据仓库的连接、数据备份、数据恢复、数据仓库资源和性能监控等运维管理工作。

- 与大数据无缝集成

您可以使用标准SQL查询HDFS、对象存储服务（Object Storage Service，OBS）上的数据，数据无需搬迁。

- 提供一键式异构数据库迁移工具

GaussDB(DWS)提供配套的迁移工具，可支持Oracle和Teradata的SQL脚本迁移到GaussDB(DWS)。

高性能

- 云化分布式架构

GaussDB(DWS)采用全并行的MPP架构数据库，业务数据被分散存储在多个节点上，数据分析任务被推送到数据所在位置就近执行，并行地完成大规模的数据处理工作，实现对数据处理的快速响应。

- 查询高性能，万亿数据秒级响应

GaussDB(DWS)通过算子并行执行、向量化执行引擎实现指令在寄存器并行执行，以及LLVM动态编译减少查询时冗余的条件逻辑判断，助力数据查询性能提升。

GaussDB(DWS)支持行列混合存储，可以同时为用户提供更优的数据压缩比（列存）、更好的索引性能（列存）、更好的点更新和点查询（行存）性能。

- 数据加载快

GaussDB(DWS)提供了GDS极速并行大规模数据加载工具。

易扩展

- 按需扩展：Shared-Nothing开放架构，可随时根据业务情况增加节点，扩展系统的数据存储能力和查询分析性能。
- 扩容后性能线性提升：容量和性能随集群规模线性提升，线性比0.8。
- 扩容不中断业务：扩容过程中支持数据增、删、改、查，及DDL操作(Drop/Truncate/Alter table)，表级别在线扩容技术，扩容期间业务不中断、无感知。

高可靠

- ACID
支持分布式事务ACID (Atomicity, Consistency, Isolation, Durability)，数据强一致保证。
- 全方位HA设计
GaussDB(DWS)所有的软件进程均有主备保证，集群的协调节点 (CN)、数据节点 (DN) 等逻辑组件全部有主备保证，能够保证在任意单点物理故障的情况下系统依然能够保证数据可靠、一致，同时还能对外提供服务。
- 安全
GaussDB(DWS)支持数据透明加密，同时可与数据库安全服务 (DBSS) 对接，基于网络隔离及安全组规则，保护系统和用户隐私及数据安全。GaussDB(DWS)还支持自动数据全量、增量备份，提升数据可靠性。

低成本

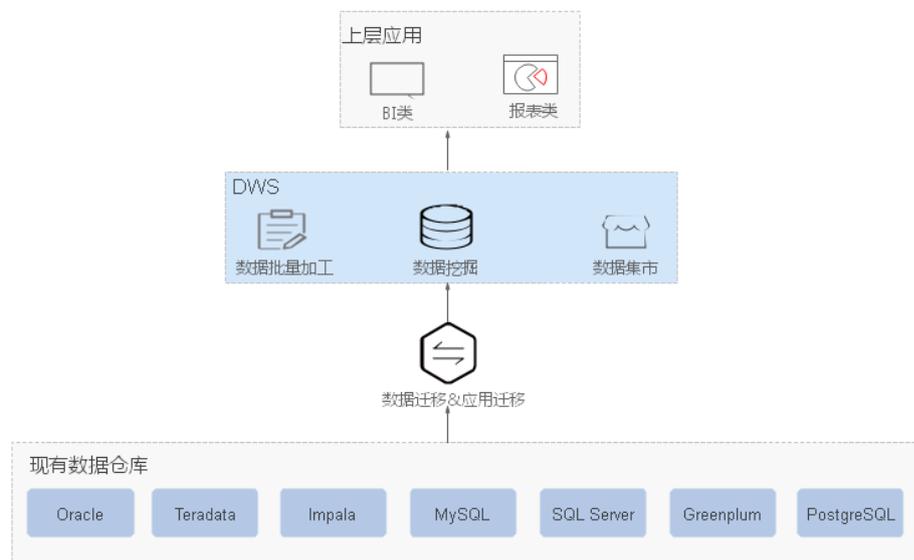
- 按需付费：GaussDB(DWS)按实际使用量和使用时长计费。您需要支付的费率很低，只需为实际消耗的资源付费。
- 门槛低：您无需前期投入较多固定成本，可以从低规格的数据仓库实例起步，后续随时根据业务情况弹性伸缩所需资源，按需开支。

5 应用场景

数据仓库迁移

数据仓库是企业的重要数据分析系统，随着业务量的增长，自建数仓性能逐渐不能满足实际要求，同时扩展性差、成本高，也使扩容极为困难。GaussDB(DWS)作为云上企业级数据仓库，具备高性能、低成本、易扩展等特性，满足大数据时代企业数据仓库业务诉求。

图 5-1 数据仓库迁移



优势

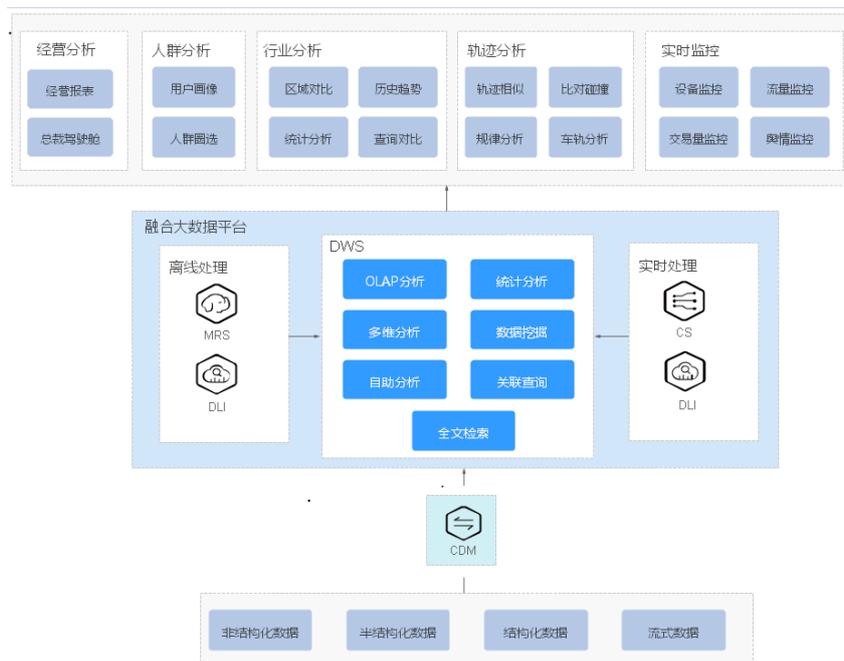
- 平滑迁移
GaussDB(DWS)提供配套的迁移工具，可支持TeraData、Oracle、MySQL、SQL Server、PostgreSQL、Greenplum、Impala等常用数据分析系统的平滑迁移。
- 兼容传统数据仓库
GaussDB(DWS)支持SQL 2003标准，兼容Oracle的部分语法和数据结构，支持存储过程，可与常用BI（business intelligence）工具无缝对接，业务迁移修改量极小。

- 安全可靠
GaussDB(DWS)支持数据加密，同时可与数据库安全服务对接，保证云上数据安全。同时GaussDB(DWS)支持自动数据全量、增量备份，提升数据可靠性。

大数据融合分析

随着信息技术的发展和进步，数据资源已经成为企业的核心资源。整合数据资源，构建大数据平台，发现数据价值，成为企业经营的新趋势和迫切诉求。而如何从海量数据中快速挖掘“价值”，成为助力客户实现预测性分析的关键要素。

图 5-2 大数据融合分析



优势

- 统一分析入口
以GaussDB(DWS)的SQL作为上层应用的统一入口，应用开发人员使用熟悉的SQL语言即可访问所有数据。
- 实时交互分析
针对即时的分析需求，分析人员可实时从大数据平台上获取信息。
- 弹性伸缩
增加节点，即可扩展系统的数据存储能力和查询分析的性能，可支持PB级数据的存储和计算。

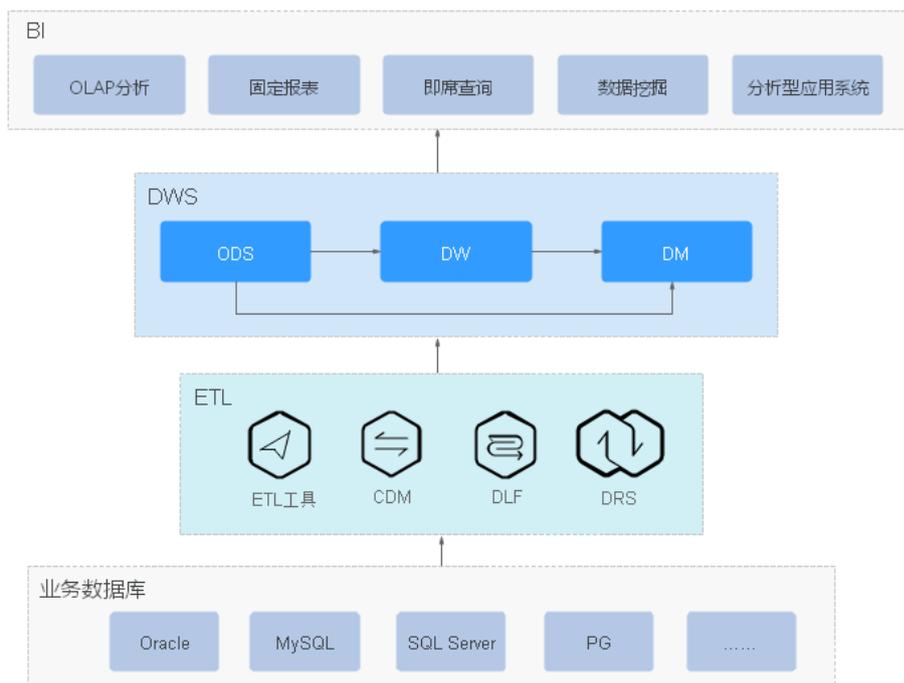
增强型 ETL 和实时 BI 分析

数据仓库在整个BI系统中起到了支柱的角色，更是海量数据收集、存储、分析的核心。为IoT（Internet of things）、金融、教育、移动互联网、O2O（Online to Offline）等行业提供强大的商业决策分析支持。

优势

- 数据迁移
多数据源，高效批量、实时数据导入。
- 高性能
PB级数据低成本的存储与万亿级数据关联分析秒级响应。
- 实时
业务数据流实时整合，及时对经营决策进行优化与调整。

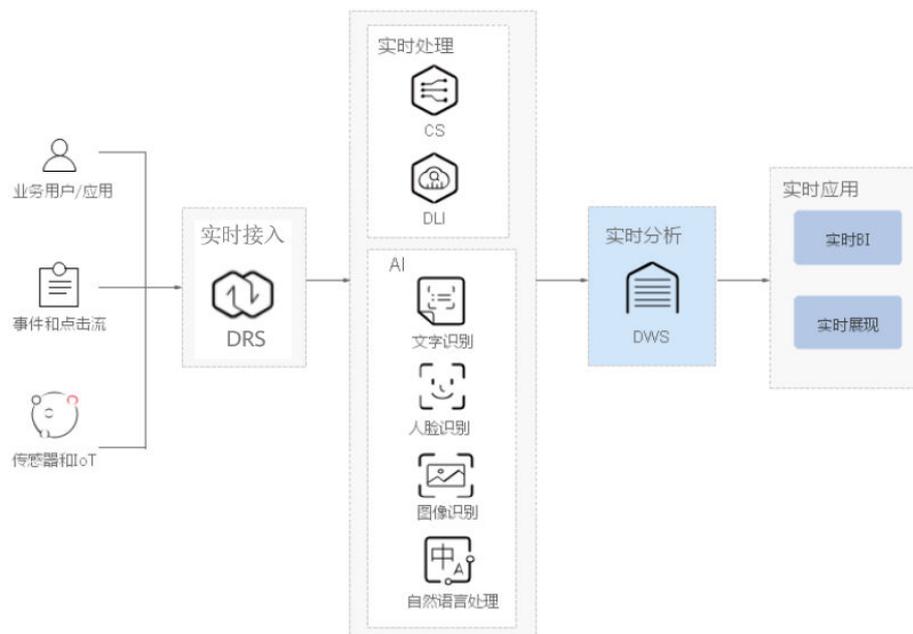
图 5-3 增强型 ETL+实时 BI 分析



实时数据分析

移动互联网、IoT场景下会产生大量实时数据，为了快速获取数据价值，需要对数据进行实时分析，GaussDB(DWS)的快速入库和查询能力可支持实时数据分析。

图 5-4 实时数据分析



优势

- 流式数据实时入库

IoT、互联网等数据经过流计算及AI服务处理后，可实时写入GaussDB(DWS)。
- 实时监控与预测

围绕数据进行分析和预测，对设备进行监控，对行为进行预测，实现控制和优化。
- AI融合分析

AI服务对图像、文本等数据的分析结果可在GaussDB(DWS)中与其他业务数据进行关联分析，实现融合数据分析。
- 电商场景

华为云GaussDB(DWS)是支撑华为商城Vmall进行数据分析的数据仓库引擎。电商的数据分析，主要用于支撑营销推荐、运营分析、全文检索、客户分析等业务。

优势如下：

 - 分析维度广：可从商品、用户、运营、地区等多个维度进行复杂的分析。
 - 随业务增长扩容：可根据业务增长，随时进行集群规模扩展。
 - 可靠性高：GaussDB(DWS)的高可靠性支撑电商系统的长时稳定运行。
- IoT场景

物联网（IoT）所产生的数据，通过构建GaussDB(DWS)，围绕海量的数据进行实时分析并进行反馈优化。应用在工业IoT、O2O业务系统、车联网等解决方案。

优势如下：

 - 设备监控与预测：围绕数据进行分析和预测，对设备进行监视，对行为进行预测，实现控制和优化及自我诊断和修复。
 - 信息推荐：结合在围绕人周围的联网设备汇集的数据，为用户进行定向的信息推荐。

6 产品功能

GaussDB(DWS)为用户提供GaussDB(DWS)管理控制台、GaussDB(DWS)客户端、REST API等多样化的使用方式。本章节为您介绍GaussDB(DWS)的主要功能。

企业级数据仓库和标准 SQL 的支持

当创建好GaussDB(DWS)集群后，用户就可以使用SQL客户端工具连接集群，然后执行创建数据库、管理数据库、导入/导出数据以及查询数据的操作。

GaussDB(DWS)为用户提供了PB（petabyte）级高性能数据库，主要体现在：

- MPP大规模并行处理框架，支持行列混存、向量化执行，实现万亿数据关联分析秒级响应。
- 内存计算，基于Bloom Filter的Hash Join优化，性能提升2~10倍。
- 支持全对称分布式的Active-Active多节点集群架构，系统无单点故障。
- 基于电信技术优化的大规模集群通信，提升计算节点间的传输效率。
- 基于代价的智能优化器，根据集群规模、数据量、生成最优计划，提高执行效率。

GaussDB(DWS)具有完备的SQL能力：

- 兼容ANSI/ISO标准的SQL92、SQL99和SQL 2003语法标准，支持存储过程，支持GBK和UTF-8字符集，支持SQL标准函数与OLAP分析函数。
- 兼容PostgreSQL/Oracle/Teradata/MySQL数据库生态，与主流第三方数据库ETL，BI（business intelligence）工具厂商对接认证。
- 支持位图roaring bitmap数据类型及对应常见函数，在互联网、零售、教育、游戏等行业，提取用户特征、用户画像等场景有广泛应用。
- 在现有范围分区（Range Partitioning）基础上，新增支持list分区方式：PARTITION BY LIST (partition_key,[...])，满足用户使用习惯。
- HDFS/OBS外表READ ONLY外表支持JSON文件格式。
- 支持系统表赋权给普通用户，Vacuum能够单独赋权，以及支持可扩展的预定义权限角色功能。
 - 新增表级权限Alter/Drop/Vacuum。
 - 新增Schema级权限Alter/Drop。
 - 新增预置角色role_signal_backend、role_read_all_stats。

详细的SQL语法和数据库操作指导，请参见《[数据仓库服务数据库开发指南](#)》。

集群管理

一个GaussDB(DWS)集群由多个在相同子网中的相同规格的节点组成，共同提供服务。GaussDB(DWS)为用户提供了简单易用的Web管理控制台，让用户可以快速申请集群，轻松执行数据仓库管理任务，专注于数据和业务。

集群管理的主要功能如下：

- 创建集群

如果用户需要在云上环境中使用数据仓库服务，首先应创建一个GaussDB(DWS)集群。用户可根据业务需求选择相应的产品规格和节点规格快速创建集群。您也可以先购买包年/包月，然后再创建集群。

- 管理快照

快照是GaussDB(DWS)集群在某一时间点的完整备份，记录了这一时刻指定集群的所有配置数据和业务数据，快照可用于还原某一时刻的集群。用户可以为集群手动创建快照，也可以开启定时创建自动快照。自动快照有保留天数限制，用户可以对自动快照进行复制，生成手动快照以便长期保留。

当您从快照恢复集群时，系统支持将快照数据恢复到新集群或者原集群中。

对于不再需要的快照，可以选择在控制台删除快照（自动快照不支持手动删除），以释放存储空间。

- 管理节点

用户可查看所创集群的节点列表，在节点管理页面可以清晰的查询各节点的状态、节点规格、是否已使用等信息。若用户需要进行大规模扩容操作时，可通过节点管理功能提前分批次添加准备好用于扩容的节点。例如需要新扩容180个BMS节点，可分3批各添加60个，如果其中有一部分添加失败，可再次添加失败数量的节点，等180个节点添加成功后，再使用这些添加好的节点进行扩容，添加节点过程中不影响集群业务。

- 扩容集群

随着业务的增长，现有集群规模可能无法满足业务需要，此时，用户可以扩容集群，为集群增加计算节点。扩容时业务不中断。进行扩容操作时，用户可根据需求选择是否在线扩容和自动重分布。

- 管理重分布

默认情况下，在扩容之后将自动调起重分布任务，为了增强扩容重分布整个流程的可靠性，可以选择在扩容时关闭自动重分布功能，在扩容成功之后再手动使用重分布功能执行重分布任务，数据重分布后将大大提升业务响应速率。当前重分布支持离线重分布、在线重分布两种模式，默认情况下，提交重分布任务时将选择离线重分布模式。

- 磁盘扩容

随着客户业务的发展，磁盘空间往往最先出现资源瓶颈，在其他资源尚且充足的情况下，执行传统扩容操作不仅耗时久，还伴随着资源浪费问题。通过磁盘扩容可快速缓解存储资源瓶颈现象，操作过程中无需暂停业务，并且不会造成CPU、内存等资源浪费。用户可在没有其他业务情况下选择磁盘扩容操作，扩容成功后如果磁盘空间不够可以继续磁盘扩容，若扩容失败用户可尝试重新进行磁盘扩容操作。

- 资源管理

当您有多个数据库用户同时查询作业时，一些复杂查询可能会长时间占用集群资源，从而影响其他查询的性能。例如一组数据库用户不断提交复杂、耗时的查

询，而另一组用户经常提交短查询。在这种情况下，短时查询可能不得不在队列中等待耗时查询完成。为了提高效率，GaussDB(DWS)提供了资源管理功能，GaussDB(DWS)资源管理以资源池为资源承载，对于不同的业务类型可以创建不同的资源池，为这些资源池配置不同的资源占比，并将数据库用户添加至对应的资源池中，以此来限制这些数据库用户的资源使用。

- 逻辑集群

逻辑集群是基于Node Group机制来划分物理节点的一种集群模式，从节点层次将大集群进行划分，和数据库形成交叉。一个数据库中的表可以按逻辑集群来分配到不同的物理节点，而一个逻辑集群也可以包含多个数据库的表。

- 重启集群

重启集群将有可能导致正在运行中的业务数据丢失，如果需要执行重启操作，请确定不存在正在运行的业务，所有数据都已经保存。

- 删除集群

当用户不再需要集群时，可选择删除集群。此操作为高危操作，删除集群可能导致数据丢失，请谨慎操作。

GaussDB(DWS)为用户提供了以下两种方式管理集群：

- 管理控制台方式

使用管理控制台方式访问GaussDB(DWS)集群。用户在注册后，可直接登录管理控制台，选择“数据仓库服务”。

有关集群管理的更多内容，请参见[管理集群](#)。

- REST API方式

支持使用GaussDB(DWS)提供的REST API接口以编程的方式管理集群。如果用户需要将GaussDB(DWS)集成到第三方系统，用于二次开发，请使用API方式访问。

详细内容，请参见《[数据仓库服务API参考](#)》。

多样化的数据导入方式

GaussDB(DWS)支持多数据源高效入库，典型的入库方式如下所示。详细指导请参见[迁移数据到GaussDB\(DWS\)](#)章节。

- 从OBS并行导入数据
- 使用GDS从远端服务器导入数据
- 从MRS导入数据到集群
- 从GaussDB(DWS)集群导入数据到新集群
- 使用gsql元命令\COPY导入数据
- 使用COPY FROM STDIN导入数据
- 使用DLI将数据导入GaussDB(DWS)
- 使用DRS将数据导入GaussDB(DWS)
- 使用CDM迁移数据到GaussDB(DWS)
- 使用DSC工具迁移SQL脚本
- 使用gs_dump和gs_dumpall命令导出元数据
- 使用gs_restore导入数据

应用程序接口

支持调用JDBC (Java database connectivity)、ODBC (Open Database Connectivity) 等标准接口开发应用程序访问GaussDB(DWS)集群中的数据库。

详细内容请参见[使用JDBC连接数集群](#)、[使用ODBC连接数集群](#)。

高可靠性

- 通过实例冗余、数据冗余，实现整个系统无单点故障。
- 数据多副本，且所有数据可手动备份至OBS。
- 自动隔离有故障的节点，利用副本重新恢复数据，并在必要时替换节点。
- 自动快照结合OBS存储，实现Region内容灾，若用户生产集群所处的地理位置发生自然灾害，或者集群内部出现了故障从而导致生产集群无法正常对外提供读写服务，那么灾备集群可以切换为生产集群，从而保障业务的连续性。
- 当集群状态为“非均衡”时会出现某些节点主实例增多，从而负载压力较大，用户可在业务低峰期进行集群主备恢复操作。
- 用户使用客户端连接DWS集群时，如果用户仅连接一个CN节点地址，通过该CN节点内网IP或弹性公网IP连接时，只能连接到固定的CN节点上，存在CN单点问题。因此GaussDB(DWS)引入了弹性负载均衡服务，解决集群访问的单点问题。弹性负载均衡 (ELB) 是将访问流量根据转发策略分发到后端多台弹性云服务器的流量分发控制服务，可以通过流量分发扩展应用系统对外的服务能力，提高应用程序的容错能力。
- 当用户集群创建后，实际需要的CN数量会随着业务需求而发生变化，因此GaussDB(DWS)提供了增删CN节点功能，实现用户可以根据实际需求动态调整集群CN数量的要求。

安全管理

- 基于网络隔离、安全组规则以及一系列安全加固项，实现租户隔离和访问权限控制，保护系统和用户的隐私及数据安全。
 - 支持SSL安全网络连接、用户权限管理、密码管理等功能，保证数据库在网络层、管理层、应用层和系统层的安全性。
- 详细内容请参见[设置SSL连接](#)和[设置三权分立](#)。

监控与审计

- 监控集群
GaussDB(DWS)与云监控服务集成，使您能够对集群中的计算节点和数据库进行实时监控。详情请参见[监控集群](#)。
- 数据库监控
数据库监控 (DMS) 是一个为GaussDB(DWS)数据库提供多维度监控服务的系统，为客户数据库的快速、稳定运行提供保驾护航的能力。该功能对业务数据库使用的磁盘、网络、OS指标数据以及集群运行关键性能指标数据进行收集、监控、分析。通过综合收集到的多种类型指标，对数据库主机、实例、业务SQL进行诊断，及时暴露数据库中关键故障及性能问题，指导客户进行优化解决。详情请参见[数据库监控](#)。
- 告警管理
告警管理包含查看告警规则、告警规则配置与告警信息订阅功能。其中，告警规则可以提供过去一周的告警信息统计与告警信息明细，方便用户自行查看租户下

的告警。该特性除了以默认值的形式提供一套GaussDB(DWS)告警最佳实践外，还允许用户根据自己的业务特点，个性化修改告警阈值。详情请参见[告警管理](#)。

- 事件通知

GaussDB(DWS)与消息通知服务对接，使您能够订阅事件并查看触发的各类事件。详情请参见[事件通知](#)。

- 审计日志

- GaussDB(DWS)与云审计服务集成，使您能够对所有的管理控制台操作及API调用进行审计。详情请参见[查看管理控制台关键操作审计日志](#)。
- GaussDB(DWS)数据库还会记录所有的SQL操作，包括连接尝试、查询和数据库的变动。详情请参见[设置数据库审计日志](#)。

支持多种数据库工具

GaussDB(DWS)提供了以下几款自研工具，用户可以在GaussDB(DWS)管理控制台下载相关的工具包。有关工具的详细指导请参见《[数据仓库服务工具指南](#)》。

- gsql工具

它是一款运行在Linux操作系统的命令行SQL客户端工具，用于连接GaussDB(DWS)集群中的数据库，并对数据库进行操作和维护。

- Data Studio工具

它是一款运行在Windows操作系统上的图形界面SQL客户端工具，用于连接GaussDB(DWS)集群中的数据库、管理数据库和数据库对象，编辑、运行、调试SQL脚本，查看执行计划等。

- GDS工具

它是GaussDB(DWS)提供的数据服务工具，通过和外表机制的配合，实现数据的高速导入导出。

GDS工具包需要安装在数据源文件所在的服务器上，数据源文件所在的服务器称为数据服务器，也称为GDS服务器。

- DSC SQL语法迁移工具

DSC (Database Schema Convertor) 是一款运行在Linux或Windows操作系统上的命令行工具，致力于向客户提供简单、快速、可靠的应用程序SQL脚本迁移服务，通过内置的语法迁移逻辑解析源数据库应用程序SQL脚本，并迁移为适用于GaussDB(DWS)数据库的应用程序SQL脚本。

DSC支持迁移Teradata、Oracle、Netezza、MySQL和DB2数据库的SQL脚本。

- 数据管理服务 (Data Admin Service, 简称DAS)

GaussDB(DWS)支持页面登录功能 (WebSQL)，该功能依赖DAS，目前具体支持“库管理”和“SQL操作”。使用时，用户需填写数据库登录名和密码方可连接集群，查看元数据及执行SQL等操作。

- gs_dump和gs_dumpall

gs_dump支持导出单个数据库或其内的对象，而gs_dumpall支持导出集群中所有数据库或各库的公共全局对象。

通过导入工具将导出的元数据信息导入至需要的数据库，可以完成数据库信息的迁移。

- gs_restore

在数据库迁移场景下，支持使用gs_restore工具将事先使用gs_dump工具导出的文件格式，重新导入GaussDB(DWS)集群，实现表定义、数据库对象定义等元数据的导入。

7 基本概念

GaussDB(DWS)的管理概念

- 集群
表示由多个节点组成的服务器群组。GaussDB(DWS)以集群为单位进行组织，一个GaussDB(DWS)集群由多个在相同子网中的相同规格的节点组成，共同提供服务。
- 节点
每个GaussDB(DWS)集群部署节点取值范围为3~256，其中实时数仓（单机模式）取值为1，每个节点都支持存储与分析数据。详情请参见[技术指标](#)。
- 规格
用户创建GaussDB(DWS)集群时需要指定集群节点的规格，不同规格节点使用的CPU、内存和存储资源不同。
- 快照
快照是用户根据需要创建的GaussDB(DWS)集群备份，用于还原创建快照时的集群数据。快照会一直保存，直到用户使用管理控制台将此快照删除（自动快照不支持手动删除）。快照会占用用户购买使用的OBS配额。
- Project
Project用于将OpenStack的资源（计算资源、存储资源和网络资源）进行分组和隔离。Project可以是一个部门或者一个项目组。一个账户中可以创建多个Project。

GaussDB(DWS) 的数据库概念

- 数据库
一个GaussDB(DWS) 集群实际上是一个支持联机分析处理方式的数据库平台，属于分析性的关系型数据库。
- 联机分析处理
联机分析处理（Online Analytical Processing，简称OLAP），是GaussDB(DWS) 集群的主要工作方式，支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。
- 大规模并行处理
大规模并行处理（Massive Parallel Processing，简称MPP，也可称为无共享架构），是指在GaussDB(DWS)集群的每个节点中，都有独立的内存计算和磁盘存

储系统，GaussDB(DWS) 根据数据库模型和应用特点，将业务数据划分到不同节点上，通过节点间的网络彼此连接协同计算，使集群作为整体提供数据库服务以满足业务需要。

- 无共享架构

无共享架构（Shared Nothing Architecture）是一种分布式计算架构。这种架构中的每一个节点都是独立、自给的，使整个系统中不出现单点竞争的情况，以提高工作效率。

- 数据库版本

每个GaussDB(DWS)集群对应一个数据库版本，用户在创建GaussDB(DWS)集群时确认此集群的版本。

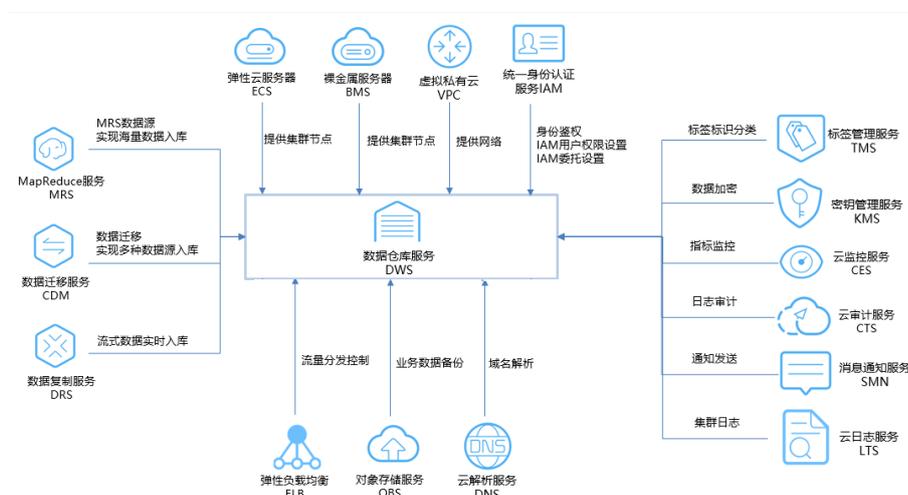
- 数据库连接

用户通过客户端连接GaussDB(DWS)集群的方式，实现数据库的连接。支持在华为云平台以及互联网环境上使用客户端进行连接。

- 数据库用户

GaussDB(DWS)集群支持不同人员连接并使用数据库，为了方便管理，数据库中支持添加新用户，并通过为用户分配不同权限达到权限控制的目的。默认的数据用户为数据库管理员，在创建GaussDB(DWS)集群时生成。

8 与其他云服务的关系



与统一身份认证服务的关系

数据仓库服务使用统一身份认证服务（Identity and Access Management，简称IAM）实现认证和鉴权功能。

需要拥有DWS Administrator权限的用户才能完整使用数据仓库服务。如需开通该权限，请联系拥有Security Administrator权限的用户或者申请新的具有DWS Administrator权限的用户。拥有DWS Database Access权限的用户，可以基于IAM用户生成临时数据库用户凭证以连接GaussDB(DWS) 集群数据库。

与弹性云服务器的关系

数据仓库服务使用弹性云服务器（Elastic Cloud Server，简称ECS）作为集群的节点，每个弹性云服务器是集群中的一个节点。

与裸金属服务器的关系

数据仓库服务使用裸金属服务器（Bare Metal Server，简称BMS）作为集群的节点，每个裸金属服务器是集群中的一个节点。

与虚拟私有云的关系

数据仓库服务使用虚拟私有云（Virtual Private Cloud，简称VPC）为集群提供网络拓扑，实现多个不同集群互相隔离并控制访问。

与对象存储服务的关系

数据仓库服务使用对象存储服务（Object Storage Service，简称OBS）作为集群数据与外部数据互相转化的一个方法，实现安全、高可靠和低成本存储需求。

与 MapReduce 服务的关系

数据仓库服务使用MapReduce服务（MapReduce Service，简称MRS）将数据从MRS迁移到GaussDB(DWS) 集群，实现海量数据通过Hadoop处理后使用GaussDB(DWS) 进行分析查询。

与云数据迁移的关系

您可使用云数据迁移（Cloud Data Migration，简称CDM）实现多种数据源数据到GaussDB(DWS) 的迁移。

与数据复制服务的关系

您可使用数据复制服务（Data Replication Service，简称DRS）将流式数据实时同步至GaussDB(DWS) 。

与云监控的关系

数据仓库服务使用云监控（Cloud Eye Service，简称CES）监控集群中的多项性能指标，从而集中高效地呈现状态信息。云监控支持发送自定义告警，用户可以即时获取异常通知。

与云审计服务的关系

数据仓库服务使用云审计服务（Cloud Trace Service，简称CTS）审计用户在管理控制台页面的非查询操作，可用于检视是否存在非法或越权操作，完善服务安全管理。

与云日志服务的关系

数据仓库服务用户可以在云日志服务（Log Tank Service，简称LTS）查看采集的集群日志或进行日志转储。

与消息通知服务的关系

数据仓库服务使用消息通知服务（Simple Message Notification，简称SMN）依据用户的事件订阅需求主动推送通知消息，使用户可以在事件发生（如集群关键操作）时能立即接收到通知。

与标签管理服务的关系

数据仓库服务使用标签管理服务（Tag Management Service，简称TMS）为用户提供跨区域、跨服务的集中标签管理和资源分类功能，让用户可以对资源进行自定义标记，实现资源的分类，然后可以对资源按标签进行搜索。

与云解析服务的关系

数据仓库服务使用云解析服务（Domain Name Service, DNS）可以提供域名映射到 GaussDB(DWS) 集群的IP地址。

与弹性负载均衡的关系

数据仓库服务利用弹性负载均衡（Elastic Load Balance, 简称ELB）健康检查机制可将集群的CN请求快速转发到正常工作的CN节点上，当有CN故障节点时，也能第一时间切换流量到健康节点上，最大化降低集群访问故障。

9 安全

9.1 责任共担

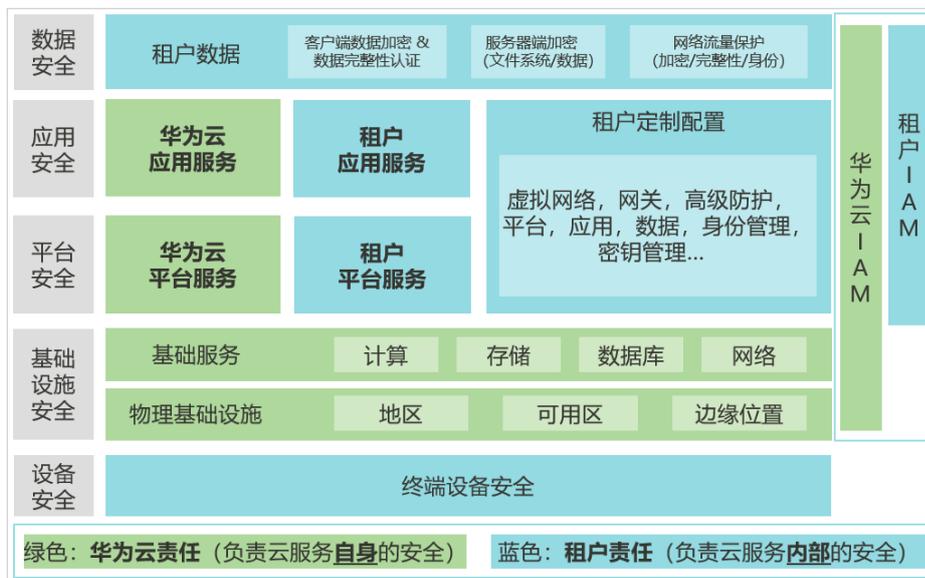
华为云秉承“将对网络和业务安全性保障的责任置于公司的商业利益之上”。针对层出不穷的云安全挑战和无孔不入的云安全威胁与攻击，华为云在遵从法律法规业界标准的基础上，以安全生态圈为护城河，依托华为独有的软硬件优势，构建面向不同区域和行业的完善云服务安全保障体系。

安全性是华为云与您的共同责任，如图9-1所示。

- **华为云**：负责云服务自身的安全，提供安全的云。华为云的安全责任在于保障其所提供的 IaaS、PaaS 和 SaaS 类云服务自身的安全，涵盖华为云数据中心的物理环境设施和运行其上的基础服务、平台服务、应用服务等。这不仅包括华为云基础设施和各项云服务技术的安全功能和性能本身，也包括运维运营安全，以及更广义的安全合规遵从。
- **租户**：负责云服务内部的安全，安全地使用云。华为云租户的安全责任在于对使用的 IaaS、PaaS 和 SaaS 类云服务内部的安全以及对租户定制配置进行安全有效的管理，包括但不限于虚拟网络、虚拟主机和访客虚拟机的操作系统，虚拟防火墙、API 网关和高级安全服务，各项云服务，租户数据，以及身份账号和密钥管理等方面的安全配置。

《[华为云安全白皮书](#)》详细介绍华为云安全性的构建思路与措施，包括云安全战略、责任共担模型、合规与隐私、安全组织与人员、基础设施安全、租户服务与租户安全、工程安全、运维运营安全、生态安全。

图 9-1 华为云安全责任共担模型



9.2 身份认证与访问控制

9.2.1 资源访问控制 (IAM 权限控制)

如果您需要对华为云上的 GaussDB(DWS) 资源, 为企业中的员工设置不同的访问权限, 以达到不同员工之间的权限隔离, 您可以使用统一身份认证服务 (Identity and Access Management, 简称 IAM) 进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能, 可以帮助您安全的控制华为云资源的访问。通过 IAM, 您可以在华为云账号中给员工创建 IAM 用户, 并授权控制他们对华为云资源的访问范围。

- **场景一:** 您的员工中有负责软件开发的人员, 您希望他们拥有 GaussDB(DWS) 的使用权限, 但是不希望他们拥有删除集群等高危操作的权限, 那么您可以使用 IAM 为开发人员创建用户, 通过授予仅能使用 GaussDB(DWS), 但是不允许删除集群的权限, 控制他们对 GaussDB(DWS) 资源的使用范围。
- **场景二:** 您希望您的员工只有 GaussDB(DWS) 的资源使用权限, 不希望拥有其他云资源的权限, 以防止资源滥用。例如只开通 GaussDB(DWS) 的操作权限, 不能使用其他云服务。

通过 IAM 权限控制, 有效达到云资源访问控制, 避免云资源误操作。了解更多请参见 [创建用户并授权使用 GaussDB\(DWS\)](#)。



9.2.2 数据库访问实现权限分层

GaussDB(DWS)中可以使用Database和Schema实现业务的隔离，区别在于：

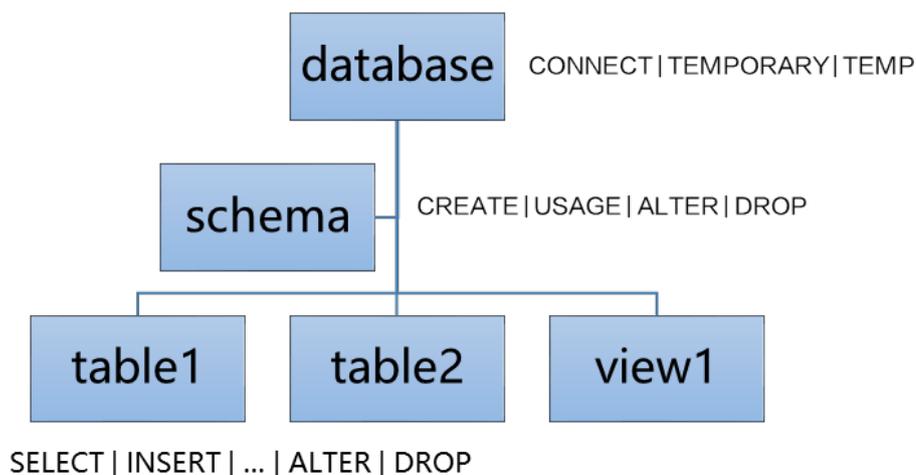
- Database之间无法直接互访，通过连接隔离实现彻底的权限隔离。各个Database之间共享资源极少，可实现连接隔离、权限隔离等。
- Schema隔离的方式共用资源较多，可以通过GRANT与REVOKE语法便捷地控制不同用户对各Schema及其下属对象的权限，从而赋给业务更多的灵活性。

从便捷性和资源共享效率上考虑，推荐使用Schema进行业务隔离。建议系统管理员创建Schema和Database，再赋予相关用户对应的权限。

1. 每个数据库包括一个或多个Schema。每个Schema包含表、视图、函数等其他类型的对象。
2. 用户要访问底层的对象，必须先赋予上层对象的权限。
3. 用户要创建或者删除Schema，需要首先被授予Database的CREATE权限。
4. 用户要访问包含在Schema中的table1，需要首先被授予Database的CONNECT权限，再被授予Schema的USAGE权限，最后授予table1的SELECT权限。

了解更多请参见[GaussDB\(DWS\)如何实现业务隔离](#)。

图 9-2 权限分层



9.2.3 GRANT 授权和 REVOKE 撤销授权

授权

GaussDB(DWS)通过GRANT语法对角色和用户进行授权操作，普通用户如果要访问某张表，如果没有系统管理员dbadmin或者表的owner进行GRANT授权，是无法访问的，此默认机制有效地控制用户对数据的访问，防止数据泄露。

使用GRANT命令进行用户授权包括以下三种场景：

- **将系统权限授权给角色或用户**
系统权限又称为用户属性，包括SYSADMIN、CREATEDB、CREATEROLE、AUDITADMIN和LOGIN。
系统权限一般通过CREATE/ALTER ROLE语法来指定。其中，SYSADMIN权限可以通过GRANT/REVOKE ALL PRIVILEGE授予或撤销。但系统权限无法通过ROLE和USER的权限被继承，也无法授予PUBLIC。
- **将数据库对象授权给角色或用户**
将数据库对象（表和视图、指定字段、数据库、函数、模式等）的相关权限授予特定角色或用户；
GRANT命令将数据库对象的特定权限授予一个或多个角色。这些权限会追加到已有的权限上。
GaussDB(DWS)会将某些类型的对象上的权限授予PUBLIC。默认情况下，对表、表字段、序列、外部数据源、外部服务器、模式或表空间对象的权限不会授予PUBLIC，而以下这些对象的权限会授予PUBLIC：数据库的CONNECT权限和CREATE TEMP TABLE权限、函数的EXECUTE特权、语言和数据类型（包括域）的USAGE特权。当然，对象拥有者可以撤销默认授予PUBLIC的权限并专门授予权限给其他用户。为了更安全，建议在同一个事务中创建对象并设置权限，这样其他用户就没有时间窗口使用该对象。另外，这些初始的默认权限可以使用ALTER DEFAULT PRIVILEGES命令修改。
- **将角色或用户的权限授权给其他角色或用户**
将一个角色或用户的权限授予一个或多个其他角色或用户。在这种情况下，每个角色或用户都可视为拥有一个或多个数据库权限的集合。
当声明了WITH ADMIN OPTION，被授权的用户可以将该权限再次授予其他角色或用户，以及撤销所有由该角色或用户继承到的权限。当授权的角色或用户发生变更或被撤销时，所有继承该角色或用户权限的用户拥有的权限都会随之发生变更。
数据库系统管理员可以给任何角色或用户授予/撤销任何权限。拥有CREATEROLE权限的角色可以赋予或者撤销任何非系统管理员角色的权限。

了解更多请参见[GRANT](#)。

撤销授权

某个用户被授权获得相应的数据库对象权限后，如果后期不再需要该对象权限或由于权限管控，需要撤销授权，可通过REVOKE语法实现。

了解更多请参见[REVOKE](#)。

9.2.4 行级访问控制

在业务开发过程中，存在多个用户共同访问和维护同一张表的场景，需要针对不同用户设置不同行数据的访问权限。例如只允许用户A查看跟自己相关的行数据，即相对于

表的管理人员能看到全部表数据而言，用户A执行SELECT * FROM table_name的时候，只能看到部分行数据，不能看到所有，以行级进行数据访问控制，对此 GaussDB(DWS)行级访问控制特性实现了这一功能。将数据库访问控制精确到数据表行级别，使数据库达到行级访问控制的能力。不同用户执行相同的SQL查询操作，读取到的结果是不同的。

ID	Name	Card	Time	Email	City
12456	Lily	458975653	2002-08-20 14:20:18	122445@hotmail.com	Shenzhen
23566	Liming	422975653	2002-09-20 11:20:18	134645@hotmail.com	Shanghai
...
23456	Lucy	422974583	2002-01-25 10:20:23	135666@hotmail.com	Shanghai
27556	Yuyi	422945683	2003-02-25 09:18:23	134576@hotmail.com	Nanjing

用户可以在数据表创建行访问控制(Row Level Security)策略，该策略是指针对特定数据库用户、特定SQL操作生效的表达式。当数据库用户对数据表访问时，若SQL满足数据表特定的Row Level Security策略，在查询优化阶段将满足条件的表达式，按照属性(PERMISSIVE | RESTRICTIVE)类型，通过AND或OR方式拼接，应用到执行计划上。

行级访问控制的目的是控制表中行级数据可见性，通过在数据表上预定义Filter，在查询优化阶段将满足条件的表达式应用到执行计划上，影响最终的执行结果。当前受影响的SQL语句包括SELECT，UPDATE，DELETE。

了解更多请参见[行级访问控制](#)。

9.3 数据保护技术

9.3.1 数据备份

GaussDB(DWS)业务数据备份与恢复通过快照功能实现。快照是对GaussDB(DWS)集群在某一时间点的一次全量数据和增量数据的备份，记录了当前数据库的数据以及集群的相关信息，其中包括节点数量、节点规格和数据库管理员用户名称等。GaussDB(DWS)可通过管理控制台进行快照的手动创建，也支持自定义自动快照策略，以便在计划时间内自动创建快照，以防止数据丢失。

当前GaussDB(DWS) 数据库支持基于OBS介质的备份与恢复方式，详情请参见[快照概述](#)。

9.3.2 透明加密

GaussDB(DWS)支持透明数据加密 (Transparent Data Encryption简称TDE或透明加密)，对数据文件进行实时加密和解密，保护用户数据隐私。

功能介绍

透明数据加密 (TDE) 是指加密GaussDB(DWS)的数据文件。通常在数据库的安全防护措施里面，可以采取一些消减措施来帮助保护数据安全。例如，设计一个安全系统、加密机密资产以及在数据库服务器的周围构建防火墙。但是，如果遇到物理介质 (如硬盘) 被黑客或者内部人员盗取的情况，恶意破坏方只需还原或附加数据库即可浏览用户数据。有一种解决方案是加密数据库中的敏感数据，并保护加密数据的密

钥，该方案可以防止任何没有密钥的人使用这些数据，但这种保护必须事先计划。GaussDB(DWS)提供了完整的解决方案TDE。

TDE可对数据实时I/O加密和解密，只要打开透明加密开关，对正常使用的用户是无感知的。这种加密使用数据库加密密钥 (DEK)，该密钥不会直接存储在数据库系统中。DEK是使用存储在KMS服务器的集群密钥 (CEK) 保护的对称密钥，数据库服务器仅仅保存其密文。在数据库启动阶段，数据库连接KMS服务器，并且解密DEK密文，从而获取到密钥明文，缓存在内存中。一旦机器下电或者集群关闭，密钥将会被清理。因此，需要保护好集群中的密钥文件，因为一旦丢失，则会造成不可恢复的危害。

使用场景

传统数据库集群中，用户数据明文保存在行存/列存文件中，集群的维护人员或者恶意攻击者可在OS层面绕过数据库的权限控制机制或者窃取磁盘直接访问用户数据。GaussDB(DWS)通过对接华为云数据加密服务的密钥管理KMS，可实现数据的透明加密，保障用户数据安全。

GaussDB(DWS)数据库级透明加密，每个GaussDB(DWS)集群有一个CEK，每个数据库单独配置DEK加密保护，DEK使用CEK加密保护，保存在GaussDB(DWS)集群侧。密钥通过KMS服务申请和加解密，加密算法通过配置项统一配置。目前支持AES、SM4算法。

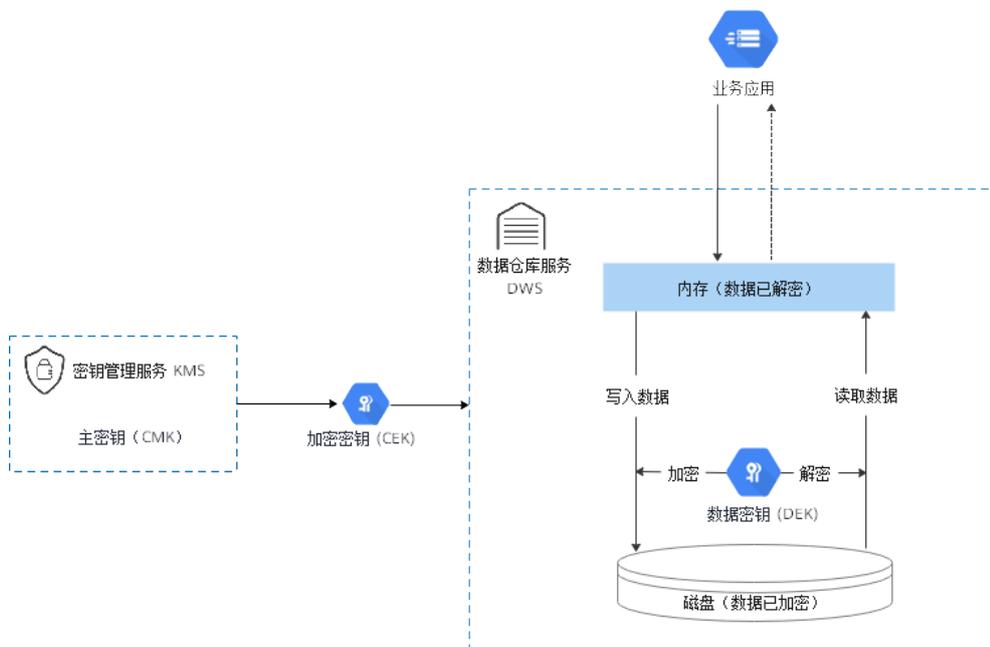
当前支持数据库级别的透明加密，在创建集群的时候进行加密配置。

了解更多请参见[数据库加密简介](#)。

TDE 加解密原理

透明加密所有加密解密操作均在内存中进行，内存中的数据是明文，磁盘中的数据是密文。数据库的使用方式保持不变，透明加密不会影响数据加工和SQL执行。

图 9-3 TDE 加解密原理



9.3.3 SSL 传输加密

GaussDB(DWS)支持SSL标准协议，SSL协议是安全性更高的协议标准，它们加入了数字签名和数字证书来实现客户端和服务器的双向身份验证，保证了通信双方更加安全的数据传输。为支持SSL连接方式，GaussDB(DWS)已经从CA认证中心申请到正式的服务器、客户端的证书和密钥（假设服务器的私钥为server.key，证书为server.crt，客户端的私钥为client.key，证书为client.crt，CA根证书名称为cacert.pem）。

SSL连接方式的安全性高于普通模式，集群默认开启SSL功能允许来自客户端的SSL连接或非SSL连接，从安全性考虑，建议用户在客户端使用SSL连接方式。并且GaussDB(DWS)服务器端的证书、私钥以及根证书已经默认配置完成。

了解更多请参见[使用SSL进行安全的TCP/IP连接](#)。

9.3.4 数据脱敏

GaussDB(DWS)提供列级别的数据脱敏(Dynamic Data Masking)功能。针对某些敏感信息（如身份证号、手机号、银行卡号等），通过应用脱敏函数进行原始数据的变形改写，实现敏感隐私数据的可靠保护，从而增强产品在数据安全和隐私保护方面的能力。



了解更多请参见[数据脱敏](#)。

9.3.5 使用函数进行加密

数据加密作为有效防止未经授权访问和防护数据泄露的技术，在各种信息系统中广泛使用。作为信息系统的核心，GaussDB(DWS)数仓也提供数据加密功能，包括透明加密和使用SQL函数加密。

GaussDB(DWS)主要提供了哈希函数和对称密码算法来实现对数据列的加解密。哈希函数支持sha256，sha384，sha512和国密sm3。对称密码算法支持aes128，aes192，aes256和国密sm4。

- 哈希函数
 - md5(string)
将string使用MD5加密，并以16进制数作为返回值。MD5的安全性较低，不建议使用。
 - gs_hash(hashstr, hashmethod)

以hashmethod算法对hashstr字符串进行信息摘要，返回信息摘要字符串。
支持的hashmethod：sha256, sha384, sha512, sm3。

- 对称密码算法
 - gs_encrypt(encryptstr, keystr, cryptotype, cryptomode, hashmethod)
采用cryptotype和cryptomode组成的加密算法以及hashmethod指定的HMAC算法，以keystr为密钥对encryptstr字符串进行加密，返回加密后的字符串。
 - gs_decrypt(decryptstr, keystr, cryptotype, cryptomode, hashmethod)
采用cryptotype和cryptomode组成的加密算法以及hashmethod指定的HMAC算法，以keystr为密钥对decryptstr字符串进行解密，返回解密后的字符串。解密使用的keystr必须保证与加密时使用的keystr一致才能正常解密。
 - gs_encrypt_aes128(encryptstr,keystr)
以keystr为密钥对encryptstr字符串进行加密，返回加密后的字符串。keystr的长度范围为1~16字节。
 - gs_decrypt_aes128(decryptstr,keystr)
以keystr为密钥对decryptstr字符串进行解密，返回解密后的字符串。解密使用的keystr必须保证与加密时使用的keystr一致才能正常解密。keystr不得为空。

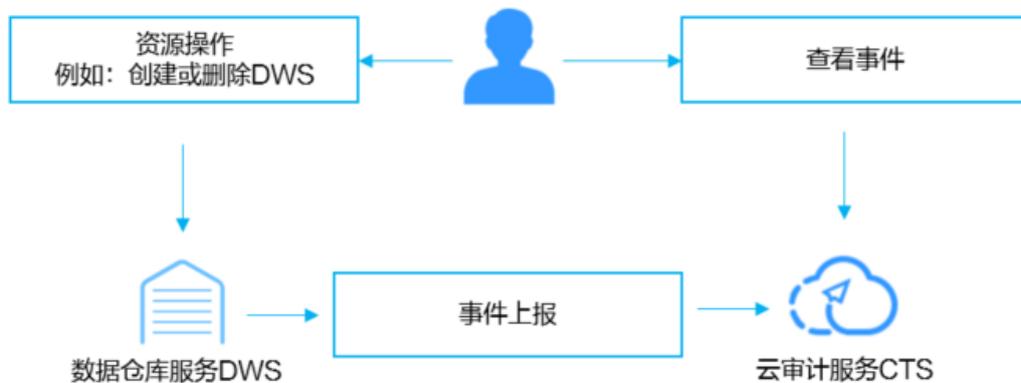
了解更多请参见[实现数据列的加解密](#)。

9.4 审计与日志

管理控制台的操作日志

GaussDB(DWS)在**管理控制台上的操作日志**，例如创建或删除DWS集群的日志记录，通过云审计服务（Cloud Trace Service, CTS）实现。CTS是华为云安全解决方案中专业的日志审计服务，提供对各种云资源操作记录的收集、存储和查询功能，可用于支撑安全分析、合规审计、资源跟踪和问题定位等常见应用场景。

用户开通云审计服务并创建和配置追踪器后，CTS可记录DWS的管理事件和数据事件用于审计。



了解更多请参见[管理控制台审计日志](#)。

数据库层级的审计日志

GaussDB(DWS)的**数据库层级日志**，包括数据库运维日志（CN节点、DN节点和操作系统日志）和DDL/DML数据库审计日志记录，通过云日志服务（Log Tank Service，简称LTS）实现。LTS用于收集来自主机和云服务的日志数据，通过海量日志数据的分析与处理，可以将云服务和应用程序的可用性和性能最大化，为您提供实时、高效、安全的日志处理能力，帮助您快速高效地进行实时决策分析、设备运维管理、用户业务趋势分析等。

当前GaussDB(DWS)的**CN节点、DN节点、操作系统messages日志以及数据库的DML/DDl审计日志**可通过开启LTS日志进行查看。

此外，为了保证日志的保留时长，数据库审计日志可通过OBS进行转储。如何设置和查看请参见[数据库审计日志](#)。

9.5 服务韧性

管理面安全加固

- Tomcat加固：在DWS管理面容器镜像中，针对Tomcat基于开源做了功能增强。
- JRE加固：
 - 升级HuaweiJre8内核版本为1.8.0_262(这里可能比真实版本低，以实际版本号为准)。
 - 将JRE的PATH设置在原来的PATH后面，规避本地越权问题（PATH=\$PATH:\$JAVA_HOME/bin）。
- 系统资源加固：GaussDB(DWS)在底层虚拟机已预置了安全相关参数，确保ECS/BMS底层的OS安全环境。

数据库与外部网络隔离

GaussDB(DWS)在单独的虚拟私有云VPC网络中，各VPC之间相互隔离。数据库作为一个组件，按照防火墙安全区域的划分是属于内部用户接口网络区域内，即受信类区域，不涉及其他区域的划分。客户端和对外服务节点CN之间的通信（命令行工具、GUI Tool及基于客户端库开发的应用程序）均支持SSL进行加密。集群内部节点运行于安全内网中，与外网隔离。

数据库集群高可用

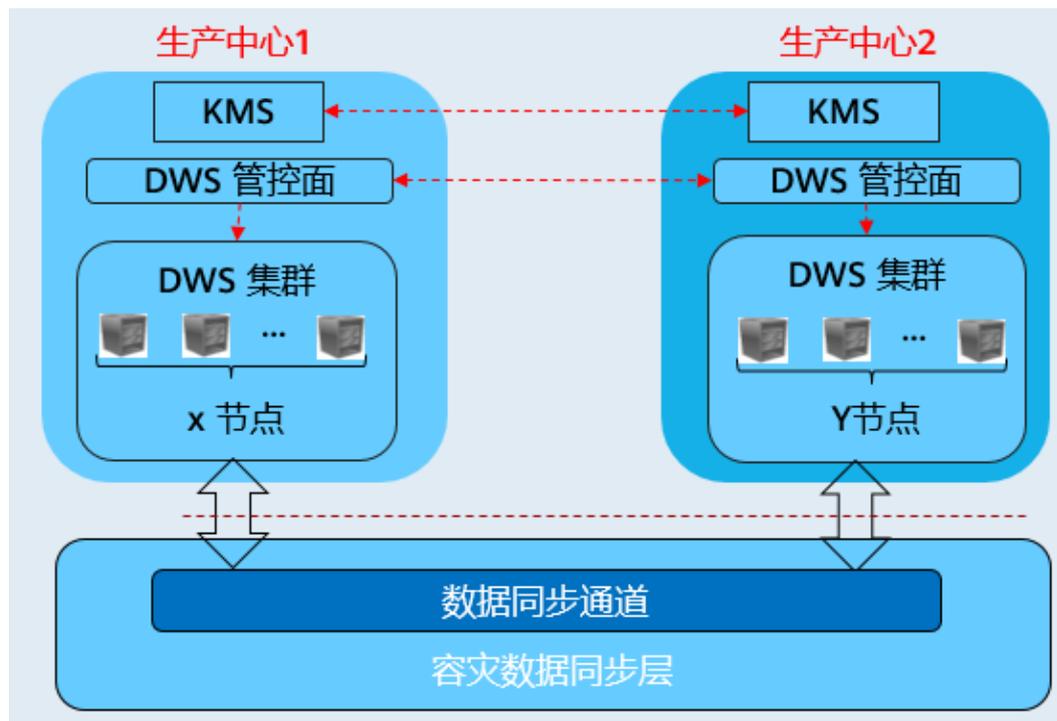
集群高可用(HA)是指数据库实践Write Ahead Logging(WAL)思想，通过各组件的主备数据同步、倒换、重建等机制，使得数据库实例具备故障恢复及自愈的能力，保证数据库遭遇Crash后数据的可靠性和完整性，最终实现集群对外业务连续性的过程。

Region 内容灾部署能力

GaussDB(DWS)提供双集群Region内容灾能力，即在另一个可用区（Region内）部署一个同构的GaussDB(DWS)灾备集群，如果生产集群所处的地理位置发生自然灾害，或者集群内部出现了故障从而导致生产集群无法正常对外提供读写服务，那么灾备集群可以切换为生产集群，从而保障业务连续性。

双集群容灾框架基于Roach备份恢复，在两套集群之间实现周期性的同步机制。该同步机制，自由灵活，两套集群之间，可分可合，无相互影响。可以实现小时级别的RTO和RPO，并且在非恢复期间，备集群是一个热备的状态，可以提供只读服务。

了解更多请参见[集群容灾](#)。



9.6 监控安全风险

GaussDB(DWS)提供基于云监控服务CES的监控能力，帮助用户监控账号下的DWS服务，执行自动实时监控、告警和通知操作。用户可以实时掌握DWS集群的各项指标、健康度。

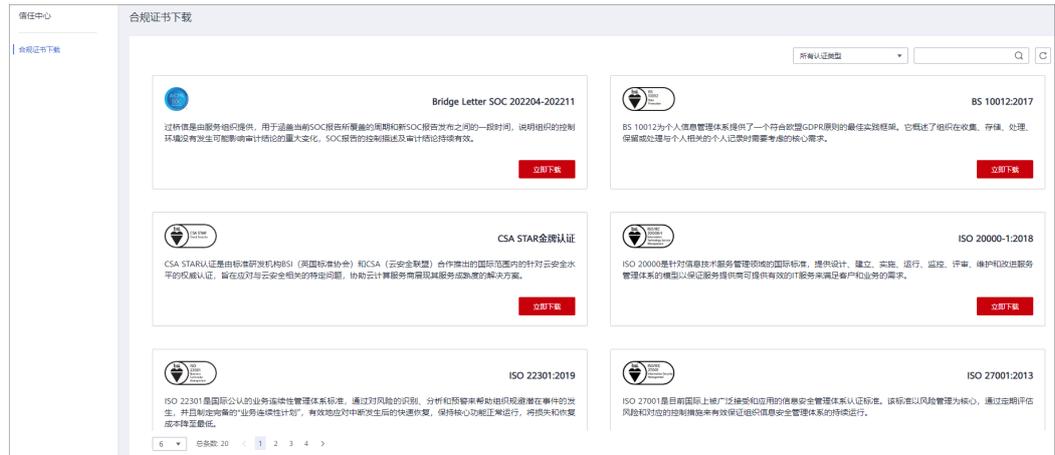
了解更多请参见[Cloud Eye监控集群](#)。

9.7 认证证书

合规证书

华为云服务及平台通过了多项国内外权威机构（ISO/SOC/PCI等）的安全合规认证，用户可自行[申请下载](#)合规资质证书。

图 9-4 合规证书下载



资源中心

华为云还提供以下资源来帮助用户满足合规性要求，具体请查看[资源中心](#)。

图 9-5 资源中心



9.8 安全公告

9.8.1 漏洞修复说明

表 9-1 已修补的开源及第三方软件漏洞列表

软件名称	软件版本	CVE编号	CS S得分	漏洞描述	受影响版本	解决版本
log4j	2.1 3.2	CVE-2021-44228	9.8	Apache Log4j2 2.0-beta9 through 2.15.0 (excluding security releases 2.12.2, 2.12.3, and 2.3.1) JNDI features used in configuration, log messages, and parameters do not protect against attacker controlled LDAP and other JNDI related endpoints. An attacker who can control log messages or log message parameters can execute arbitrary code loaded from LDAP servers when message lookup substitution is enabled. From log4j 2.15.0, this behavior has been disabled by default. From version 2.16.0 (along with 2.12.2, 2.12.3, and 2.3.1), this functionality has been completely removed. Note that this vulnerability is specific to log4j-core and does not affect log4net, log4cxx, or other Apache Logging Services projects.	8.0.0~8.1.2	8.1.3

10 GaussDB(DWS)权限管理

如果您需要对华为云上的GaussDB(DWS)资源，为企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制华为云资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并授权控制他们对华为云资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有GaussDB(DWS)的使用权限，但是不希望他们拥有删除集群等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用GaussDB(DWS)，但是不允许删除集群的权限，控制他们对GaussDB(DWS)资源的使用范围。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户进行权限管理，您可以跳过本章节，不影响您使用GaussDB(DWS)服务的其它功能。

IAM是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见[《IAM产品介绍》](#)。

GaussDB(DWS) 权限

默认情况下，IAM管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。

GaussDB(DWS) 部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域对应的项目中设置相关权限，并且该权限仅对此项目生效；如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问GaussDB(DWS)时，需要先切换至授权区域。

- **角色：** IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于华为云各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。
- **策略：** IAM最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。例如：针对GaussDB(DWS) 服务，IAM管理员能够控制IAM用户仅能对某一类资源进行指定的管理操作。

多数细粒度策略以API接口为粒度进行权限拆分，GaussDB(DWS) 支持的API授权项请参见[权限策略和授权项](#)。

细粒度策略的创建方法，请参见[GaussDB\(DWS\)自定义策略](#)。

如表10-1所示，包括了GaussDB(DWS)的所有系统权限。

表 10-1 GaussDB(DWS) 系统权限

系统角色/策略名称	描述	类别	依赖关系
DWS ReadOnlyAccess	数据仓库服务只读权限，拥有该权限的用户仅能查看数据仓库服务资源。	系统策略	无
DWS FullAccess	数据仓库服务数据库管理员权限，拥有数据仓库服务所有权限。	系统策略	无
DWS Administrator	DWS数据库管理员权限，拥有对数据仓库服务的所有执行权限。 <ul style="list-style-type: none"> 拥有VPC Administrator权限的GaussDB(DWS)用户可以创建VPC或子网。 拥有云监控Administrator权限的GaussDB(DWS)用户，可以查看GaussDB(DWS)集群的监控指标信息。 	系统角色	依赖Tenant Guest和Server Administrator策略，需要在同项目中勾选依赖的策略。
DWS Database Access	GaussDB(DWS)数据库访问权限，拥有该权限的用户，可以基于IAM用户生成临时数据库用户凭证以连接GaussDB(DWS)集群数据库。	系统角色	依赖DWS Administrator，需要在同项目中勾选依赖的策略。

表10-2列出了GaussDB(DWS)常用操作与系统权限的授权关系，您可以参照该表选择合适的系统策略。

📖 说明

- 每个区域的每个项目首次使用弹性IP绑定功能时，系统将提示创建名称为“DWSAccessVPC”委托以授权GaussDB(DWS)访问VPC。授权成功后，GaussDB(DWS)可以在绑定弹性IP的虚拟机故障时切换至健康虚拟机。
- 在实际业务中，除了具备策略权限外还需要给不同角色的用户授予不同的资源操作权限。例如创建快照、重启集群等操作，详情请参见[策略语法：细粒度策略](#)。
- 默认情况下，只有华为云账号或拥有Security Administrator权限的用户才具备查询委托和创建委托的权限。账号中的IAM用户，默认没有查询委托和创建委托的权限，在使用弹性IP绑定功能时页面会屏蔽绑定按钮，此时需联系有“DWS Administrator”权限的用户在当前页面完成对GaussDB(DWS)的委托授权。

表 10-2 GaussDB(DWS) 常用操作与系统权限的关系

操作	DWS FullAccess	DWS ReadOnlyAccess	DWS Administrator	DWS Database Access
创建/恢复集群	√	x	√	x
获取集群列表	√	√	√	x
获取单个集群详情	√	√	√	x
设置自动快照	√	x	√	x
设置安全参数/参数组	√	x	√	x
重启集群	√	x	√	x
扩容集群	√	x	√	x
重置密码	√	x	√	x
删除集群	√	x	√	x
设置可维护时间段	√	x	√	x
绑定EIP	x	x	√	x
解绑EIP	x	x	√	x
创建DNS域名	√	x	√	x
释放DNS域名	√	x	√	x
修改DNS域名	√	x	√	x
创建MRS连接	√	x	√	x
更新MRS连接	√	x	√	x
删除MRS连接	√	x	√	x
添加/删除标签	√	x	√	x
编辑标签	√	x	√	x
创建快照	√	x	√	x
获取快照列表	√	√	√	√
删除快照	√	x	√	x
复制快照	√	x	√	x

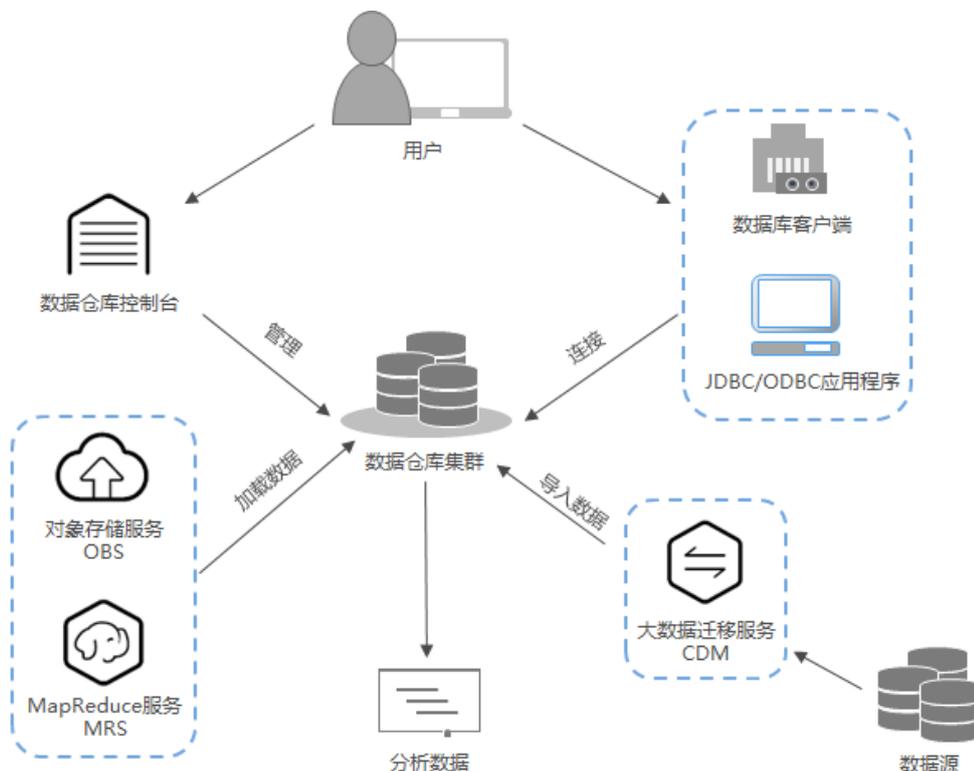
相关链接

- [IAM产品介绍](#)
- [创建用户并授权使用GaussDB\(DWS\)](#)
- [策略支持的授权项](#)

11 如何访问 GaussDB(DWS)

数据仓库服务的基本使用流程如下图所示：

图 11-1 数据仓库服务基本使用流程



访问集群

GaussDB(DWS) 提供了Web化的服务管理平台，即管理控制台，还提供了基于HTTPS请求的API (Application programming interface) 方式管理GaussDB(DWS) 集群。

说明

集群部署时，如果单个节点故障，访问GaussDB(DWS)时会自动屏蔽异常节点，但会影响集群性能。

访问集群数据库

GaussDB(DWS) 支持如下方式访问集群数据库：

- GaussDB(DWS) 客户端方式
用户通过GaussDB(DWS) 客户端访问集群数据库，请参见[使用gsqll命令行客户端连接集群](#)、[使用Data Studio图形界面客户端连接集群](#)和[使用DAS连接集群](#)。
- JDBC、ODBC等接口调用方式
支持调用JDBC (Java database connectivity)、ODBC (Open Database Connectivity) 等标准接口开发应用程序访问GaussDB(DWS) 集群中的数据库。详细内容请参见[使用JDBC连接数集群](#)、[使用ODBC连接数集群](#)。
- psycopg2、PyGreSQL第三方驱动程序
用户在创建好数据仓库集群后使用psycopg2、PyGreSQL第三方库连接到集群，则可以使用Python访问GaussDB(DWS)，并进行数据表的各种操作。详细内容请参见[使用Python第三方库psycopg2连接集群](#)、[使用Python第三方库PyGreSQL连接集群](#)。

📖 说明

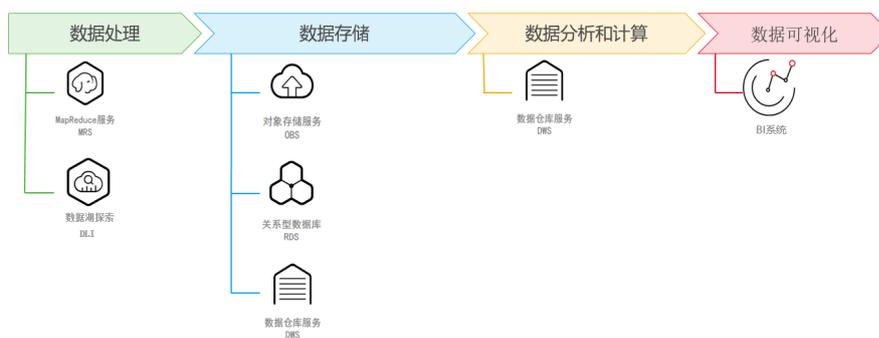
GaussDB(DWS)暂不支持跨库访问，可以通过schema进行资源隔离。详情请参见[CREATE SCHEMA](#)。

端到端数据分析流程

数据仓库服务与华为云平台其他一系列服务已无缝集成，以使用户能够快速部署端到端的数据分析和解决方案。

端到端数据分析流程如下图所示，该流程展示了每个过程可以使用的云服务。

图 11-2 端到端数据分析流程



12 计费说明

数据仓库服务 GaussDB(DWS) 支持包年/包月计费、按需计费和折扣套餐计费模式。为了便于您便捷的下单购买，在控制台购买界面中已经为您计算好了整个 GaussDB(DWS) 集群的价格，您可一键完成整个配置在购买。

计费项

华为云 GaussDB(DWS) 根据您选择的数据仓库节点规格及使用的相关资源进行计费。了解 GaussDB(DWS) 价格详情，请参见[GaussDB\(DWS\)价格详情](#)。您可以通过 GaussDB(DWS) 提供的价格计算器，选择您需要的集群节点规格，来快速计算出购买 GaussDB(DWS) 集群的参考价格。

表 12-1 GaussDB(DWS) 计费项说明

计费项	计费说明
数据仓库节点	对您选择的节点规格计费，计费详情请参见 GaussDB(DWS)价格详情 。针对节点规格提供了以下计费模式： <ul style="list-style-type: none">包年/包月（推荐购买）按需付费（小时） 上述计费模式的详细说明请参见 计费模式 。
快照存储空间	GaussDB(DWS) 提供了部分免费存储空间，用于存放您的快照数据，免费空间内不计费。当快照数据存储空间超过免费空间大小时，超出部分按照对象存储服务（OBS）的计费规则进行计费，参见 OBS价格详情 。 免费空间大小是您的集群的总存储空间大小，即单节点存储空间大小 x 节点数。
标准数仓存储	用户在可选范围内自定义集群使用的存储类型与存储容量，采用存储资源独立计费的方式。计费详情请参见 GaussDB(DWS)价格详情 。
OBS热数据存储	用户在可选范围内自定义集群使用的存储类型与存储容量，采用存储资源折扣套餐独立计费的方式。
冷数据存储	冷数据采用OBS存储性价比高，按需集群使用按需计费，热数据套餐可用于抵扣冷数据按需套餐费用。

计费项	计费说明
公网IP及带宽 (可选)	<ul style="list-style-type: none"> GaussDB(DWS) 集群支持绑定弹性公网IP，所需费用按照虚拟私有云（VPC）服务的弹性公网IP计费规则进行计费。计费详情请参见弹性公网IP价格详情。 GaussDB(DWS) 集群在华为云内部网络产生的流量不计费。
数据库加密密钥 (可选)	在创建GaussDB(DWS) 集群时，如果开启“加密数据库”的功能，需使用KMS(密钥管理服务)的密钥加解密数据库。密钥按照数据加密服务密钥的计费规则进行计费，详情请参见 数据加密服务价格详情 。

计费模式

GaussDB(DWS) 提供以下计费模式供您选择：包年/包月、按需计费和折扣套餐计费。供您灵活选择，使用越久越便宜。

- 包年/包月：**客户可通过预付费，以包年包月模式购买云服务，这种购买方式相对于按需付费提供更大的折扣，对于长期使用者，推荐该计费模式。购买的云服务和开通/使用的云服务实例一一对应和绑定。购买的使用期限到期、进入宽限期/保留期，保留期过后系统会自动释放资源，客户还可通过“续费”来继续使用云服务。

说明

包年/包月计费模式仅8.1.3.100及以上集群版本支持。

- 按需付费（小时）：**购买方式比较灵活，可以即开即停。使用时按实际使用时长计费（精确到分钟），每个自然小时出账单并从您的账户余额中扣减费用，删除集群即停止计费。该方式适合需要做前期运行试验、概念验证和短期使用的客户。
- 折扣套餐：**按月或按年预先支付OBS热数据存储费用，相比按需计费提供大额折扣，适合计划长期使用DWS 3.0场景的客户。

变更配置

- 规格：**在购买GaussDB(DWS) 集群时，有多种节点规格供您选择，您可根据业务需要选择合适的节点规格。当集群创建成功后，随着业务量的增长，您可以对集群进行扩容，从而增加集群节点，详情请参见[扩容集群](#)章节。若扩容的变更配置方式不满足您的要求，您也可以重新创建集群，然后通过数据迁移的方式实现集群配置的变更。
- 计费模式：**
 - 按需变更为包年/包月：按需变更为包年/包月，会生成新的订单，用户支付订单后，包年/包月资源立即生效。
 - 包年/包月变更为按需：包年/包月转按需，需包年/包月资费模式到期后，按需的资费模式才会生效。

续费

目前GaussDB(DWS) 提供包年/包月、按需和折扣购买模式，“包年/包月”是您在购买时一次性付费，使用过程中不会再另外扣费，只要您的账户上有足够余额，则不会

影响您的使用。按需是每小时扣费，如果余额不足导致欠费。如果您未能续费，华为云不会立即停止您的业务，订单转入保留期，此时集群将终止服务，数据仍然保留。

保留期的时长由客户等级而定。

如需续费，请进入[续费管理](#)页面进行续费操作。

到期

- 包年/包月的集群如果在保留期结束时您没有续费，实例将终止服务，系统中的数据也将被永久删除。
- 按需购买的集群，没有到期。
- 折扣套餐的集群时长到期后，将自动转为按需付费方式。

欠费

欠费后，可以查看欠费详情。为防止相关资源不被停止或者释放，请及时进行充值。按需购买的集群是按每小时扣费，当余额不足，无法对上一个小时的费用进行扣费，就会导致集群欠费。您续费后解冻集群，可继续正常使用，请注意在保留期进行的续费，是以原到期时间作为生效时间，您应当支付从进入保留期开始到续费时的服务费用。

说明

- 欠费后将导致资源被冻结。其中集群删除、重启、DAS登录等操作会受限。需客户账户充值后自动解除。
- 折扣集群，没有欠费的概念。

13 使用限制

本文介绍使用GaussDB(DWS)产品重点功能限制以及注意事项。

创建GaussDB(DWS)集群后，您不需要做数据库的基础运维（例如高可用、打安全补丁等），但需要重点关注如下事项：

表 13-1 GaussDB(DWS)约束限制

事项	说明
创建集群	<ul style="list-style-type: none"> 设置的节点数量不能大于用户可使用的节点配额或256个。如果可使用的节点配额不足，用户可以单击“申请扩大配额”以提工单的形式申请更多节点配额。 集群创建成功后，“产品类型”不能再进行更改，请参考数据仓库类型了解不同产品类型功能上的具体差异。 推荐使用包年/包月方式购买集群，包月时长到期之后，集群会自动释放。
连接集群	<ul style="list-style-type: none"> 用户需要在与集群相同的虚拟私有云子网中使用客户端连接GaussDB(DWS)集群。 数据仓库服务只支持用户管理集群，不支持直接访问集群节点。用户或应用程序使用数据库对应的集群访问IP地址和端口，访问数据仓库。
SQL语法	<ul style="list-style-type: none"> 实时数仓（单机部署）暂不支持全文检索、OBS和HDFS外表导入导出、自动创建分区、SEQUENCE及相关函数。 实时数仓（单机部署）由于只有单DN，因此分布规则会被忽略，也不支持针对分布规则的修改。 从文档直接复制命令到执行环境中粘贴时可能自动换行，导致运行出错，请注意删除换行。 在实际业务中，根据需要创建新的Database，不建议直接使用集群默认的gaussdb数据库。

事项	说明
规格变更	<ul style="list-style-type: none"> ● 弹性变更规格仅8.1.1.300及以上集群版本支持。 ● 弹性变更规格当前仅支持离线变更规格，变更时间大约需要10分钟左右。 ● 若创建本地盘或者存算一体形态的集群，则创建好以后不支持修改规格，如果需要使用更高规格的节点，请重新创建一个新的集群。弹性变更规格目前仅支持SSD云盘形态的标准数仓或实时数仓集群。 ● 包年/包月计费模式集群进行弹性变更规格操作时，默认按照包年包月计费。 ● 如果当前集群是包年/包月计费模式，将不支持经典变更规格功能。 ● 在经典变更规格期间新集群是不计费的，当变更成功后开始计费。变更完成后老集群处于保留期，不再计费。 ● 经典变更规格功能目前仅支持标准数仓和实时数仓（单节点部署不支持）。逻辑集群不支持经典变更规格。 ● 经典变更规格最大支持调整到240节点，且原节点数量加调整后节点数量不大于480节点。 ● 磁盘扩容功能仅8.1.1.203及以上版本支持，并且创建集群规格需要为标准数仓SSD云盘、实时数仓或IoT数仓类型。 ● 集群状态为“可用”、“待重启”、“只读”、“非均衡”、“节点故障”或者“不可用”的情况下才能下发磁盘扩容。
节点扩容	<ul style="list-style-type: none"> ● 如果当前集群是包年/包月计费模式，那么新增节点也是按照包年包月模式进行计费。 ● 标准数仓集群扩容时使用与当前集群相同的存储规格。 ● 实时数仓（单机部署）不支持扩容集群。 ● 重分布功能8.1.1.200及以上集群版本支持。 ● 只有在扩容之后，集群任务信息为“待重分布”状态时才能手动使用“重分布”功能，其他时段该功能不可使用。 ● 缩容集群仅8.1.1.300及以上版本支持，包周期计费模式的集群仅8.2.1及以上版本支持。 ● 缩容的节点默认是按节点数计费。 ● 标准数仓集群缩容时只支持缩容与当前集群相同的存储规格。 ● 实时数仓（集群部署）不支持缩容至单机部署集群。

事项	说明
备份集群	<ul style="list-style-type: none"> ● GaussDB(DWS) 根据快照创建的新集群与生成快照的原始集群具有相同的配置，即节点的数量和规格、内存、磁盘与原集群要一致。 ● 根据快照创建新集群时，如果没有指定其他值，则参数默认与生成快照时的备份信息保持一致。 ● 实时数仓（单机部署）暂不支持快照功能。 ● 快照生成期间，请避免进行Vacuum Full操作，否则可能会导致集群只读。 ● 创建快照时因备份数据会降低磁盘I/O性能，建议您在业务相对空闲的时期进行快照操作。 ● 快照期间会保留一些中间文件，需额外占用磁盘空间，因此请避开业务高峰期并保证磁盘容量在70%以下。 ● 支持恢复快照到当前集群，但不支持恢复逻辑集群与资源池到当前集群。
版本升级	<ul style="list-style-type: none"> ● 8.1.1及以上集群版本支持用户自助通过console下发集群升级操作。 ● 如果节点故障/系统升级导致集群长时间中断，请联系技术支持人员进行处理。
数据迁移	<ul style="list-style-type: none"> ● OBS导入导出数据时，不支持中文路径。 ● OBS导入导出数据时，暂不支持跨Region进行OBS数据导入导出，必须确保OBS和DWS集群在同一个Region中。
故障切换	<ul style="list-style-type: none"> ● 容灾创建后，生产集群快照功能正常使用，但是灾备集群禁用快照功能，同时生产、灾备集群均禁用恢复功能。 ● 容灾不支持同步配置的外部数据源数据。 ● 容灾管理为同一租户下的双集群容灾。 ● 创建容灾需要生产集群与灾备集群类型、版本号一致且逻辑同构。 ● Region内容灾生产集群和灾备集群需要在同一个VPC内。 ● Region内容灾在生产集群与容灾集群切换后，原生产集群绑定的弹性负载均衡（ELB）会自动切换到新的生产集群，过程中会有短暂的连接中断，请注意在变更时间窗内尽量避免业务语句写入。 ● Region内容灾在生产集群与容灾集群，原生产集群所在的EIP和内网域名，连接IP地址不会自动切换，业务系统中如果有使用EIP，域名或者IP方式连接，需要切换到新集群。

事项	说明
冷热数据管理	<ul style="list-style-type: none"> ● 实时数仓（单机部署）暂不支持冷热分区切换功能。 ● 冷热表只支持列存2.0版本的分区表，外表不支持冷热分区。 ● 仅支持从热数据切换为冷数据，不支持从冷数据切换为热数据。 ● 对于同一分区在同一DN只会存在冷或热的一种情况，对于同一分区在不同DN可能存在部分DN为热数据，部分DN为冷数据。 ● 只支持修改冷热表的冷热切换策略，不支持修改冷热表的冷数据的表空间。
IoT数仓	<ul style="list-style-type: none"> ● 时序表不支持：UPDATE，UPSERT，主键，pck。 ● 创建时序表的用户需要有schema cstore的USAGE权限。 ● 修改时序表存储参数enable_delta时，不能与其他ALTER操作同时进行。

14 配额说明

为防止资源滥用，平台限定了各服务资源的配额，对用户的资源数量和容量做了限制。

数据仓库服务的默认用户配额如表14-1所示。关于如何查看配额以及扩大配额的详细信息，请参见[关于配额](#)。

表 14-1 用户配额

资源类型	总配额
节点数	256

15 技术指标

GaussDB(DWS)不同版本的技术指标如下所示。

表 15-1 GaussDB(DWS)技术指标 8.1.3-8.3.0 版本

技术指标	8.1.3最大值	8.2.0最大值	8.2.1最大值	8.3.0最大值
集群节点数	2048	2048	2048	2048
并发连接数	分钟级复杂查询并发数：80 秒级短查询并发数：500 毫秒级短事务并发数：5000	分钟级复杂查询并发数：80 秒级短查询并发数：500 毫秒级短事务并发数：5000	分钟级复杂查询并发数：80 秒级短查询并发数：500 毫秒级短事务并发数：5000	分钟级复杂查询并发数：80 秒级短查询并发数：500 毫秒级短事务并发数：5000
集群数据容量	20PB	20PB	20PB	20PB
单表大小	1PB	1PB	1PB	1PB
单行数据大小	1GB	1GB	1GB	1GB
单表列数（除Hudi表外）	1600	1600	1600	1600
Hudi表列数	-	-	5000	5000
分区表的分区个数	32768	32768	32768	32768
集群内单点故障RTO	60s	60s	60s	60s
集群内单点故障RPO	0	0	0	0
集群故障容灾切换RTO	60min	60min	60min	60min
集群故障容灾切换RPO	60min	60min	60min	60min

表 15-2 GaussDB(DWS)技术指标 8.0.x-8.1.1 版本

技术指标	8.0.x最大值	8.1.0最大值	8.1.1最大值
数据容量	10PB	10PB	20PB
集群节点数	256	256	2048
单表大小	1PB	1PB	1PB
单行数据大小	1GB	1GB	1GB
每条记录单个字段的大小	1GB	1GB	1GB
单表记录数	2 ⁵⁵	2 ⁵⁵	2 ⁵⁵
单表列数	1600	1600	1600
单表中的索引个数	无限制	无限制	无限制
单表索引包含列数	32	32	32
单表约束个数	无限制	无限制	无限制
并发连接数	分钟级复杂查询并发数：60 毫秒级短事务并发数：5000	分钟级复杂查询并发数：60 毫秒级短事务并发数：5000	分钟级复杂查询并发数：80 毫秒级短事务并发数：5000
分区表的分区个数	32768	32768	32768
分区表的单个分区大小	1PB	1PB	1PB
分区表的单个分区记录数	2 ⁵⁵	2 ⁵⁵	2 ⁵⁵

📖 说明

并发连接数的最大值基于数仓规格48 vCPU和64 vCPU。例如，标准数仓云盘规格 dwsk.12xlarge (48 vCPU | 384GB | 24000GB SSD) 或dwsx2.16xlarge.m7 (64 vCPU | 512GB | 32000GB SSD)。

16 版本说明

本章节为您介绍GaussDB(DWS) 集群版本的更新记录。

V9.0.2 版本

发布时间：2023年11月15日

配套内核版本：9.0.2

新增云原生数仓DWS 3.0，其采用计算存储分离架构，解决了计算存储必须等比例缩放的问题。赋能用户面向业务峰谷时，对计算能力进行快速且独立的扩缩要求，同时保证存储无限扩展、按需付费，做到快速、敏捷的响应业务变化，同时具有更高的性价比，进一步助力企业降本增效。

V8.2.0.107 版本

发布时间：2023年10月7日

配套内核版本：8.2.0

1. DWS支持用户权限管理，可以在控制台创建、删除、更新数据库用户并对其进行权限管理。
2. 转储数据库审计日志优化：
 - 审计日志压缩类型由zip改为gzip。
 - 审计日志文件格式支持csv格式。
 - 老版本集群升级到新版本，console有开关控制兼容。

V8.2.0.106 版本

发布时间：2023年8月14日

配套内核版本：8.2.0

1. 集群详情界面增加“运维账户”管理模块，通过授权账户控制运维人员访问集群的操作权限与影响范围。
2. 集群创建MRS数据源连接增加普通用户权限配置方式。

V8.2.0.101 版本

发布时间：2023年4月27日

配套内核版本：8.2.0

1. 支持用户自助通过控制台下发集群升级操作。
2. WebSql支持操作审计。
3. DMS性能监控面板节点维度监控视图支持多主机指标展示。
4. 支持在指定节点上添加CN。

V8.2.0.100 版本

发布时间：2023年2月15日

配套内核版本：8.2.0

1. 原“云数仓”产品类型变更为“标准数仓”。
2. 资源管理页面重构。
3. 弹性伸缩入口整合，调整大小改名为经典变更规格，变更规格改名为弹性变更规格。其中经典变更规格、弹性变更规格、磁盘扩容入口整合至集群列表所在行操作列“更多>规格变更”下；扩容、缩容、重分布入口整合至集群列表所在行操作列“更多>节点变更”下。
4. 参数修改支持查看历史版本和回退，如果多次修改后无法确认哪些参数已经生效，可通过参数修改历史记录查验。
5. 新增快照参数配置功能，可指定创建快照和快照恢复的参数信息，以便对快照功能进行优化调整。
6. DMS新增对锁的监控和告警。

V8.1.3.300 版本

发布时间：2022年12月20日

配套内核版本：8.1.3

1. 新增集群拓扑图页面，向用户展示了集群内所有节点的拓扑结构，方便用户查看节点状态、节点进程、节点IP地址等信息。
2. 集群绑定ELB解除同VPC/同企业项目的限制，当前支持同VPC绑定ELB和跨VPC绑定ELB两种类型操作。
3. 增删CN支持在线增删CN及多CN并行增加。

V8.2.0 版本

发布时间：2022年11月23日

配套内核版本：8.2.0

1. 支持以委托方式访问客户OBS上的数据，客户通过创建DWS云服务委托并授予相应权限，然后再创建OBS数据源绑定该委托，便可通过OBS外表的方式访问OBS上的数据。
2. 工作负载管理名称变更为资源管理，并增强其易用性。
3. 资源池异常规则优化，新增“单DN算子下盘大小”和“DN平均消耗CPU占比”规则参数。

4. DMS新增表诊断页面，对于集群中数据表统计信息进行监控分析，可查看倾斜率高于5%、脏页率高于50%且表大小TOP50的表信息。

V8.1.3.200 版本

发布时间：2022年11月15日

配套内核版本：8.1.3

1. 新增调整大小功能，用户可对当前集群规格进行调整或进行集群拓扑结构变更。
2. 支持快照恢复到当前集群。
3. 支持手动停止快照。
4. DMS新增主机级别的趋势。
5. 节点管理支持对节点别名进行修改。
6. 数据库内核版本8.1.3功能增强，详情参见[8.1.3新增数据库特性](#)。

V8.1.3.101 版本

发布时间：2022年8月23日

配套内核版本：8.1.3

1. DWS数仓数据库支持基于DAS进行管理连接集群。
2. 逻辑集群支持资源管理。
3. 支持使用Windows gsql上客户端连接集群。
4. DWS支持使用Python第三方库psycopg2、PyGreSQL连接集群。
5. 新增数据库内核版本8.1.3功能：只做除零返回NULL：参数behavior_compat_options新增兼容配置项enable_division_by_zero_mysql。

V8.1.3.100 版本

发布时间：2022年6月20日

配套内核版本：8.1.3

1. 工作负载管控面和内核CPU共享配额描述统一。

V8.1.1.500 版本

发布时间：2022年6月20日

配套内核版本：8.1.1

1. 解决更新现网关键问题，不涉及刷新资料。

V8.1.3 版本

发布时间：2022年4月15日

配套内核版本：8.1.3

1. 新增智能运维功能，可帮助用户智能执行运维任务，通过集群负载情况，选择合理时间窗、并发度完成用户指定的任务。

2. 集群重分布新增调度重分布模式，在配置的时间窗口内做表的重分布，时间窗口外将处于重分布暂停状态。
3. 新增数据库内核版本8.1.3功能，详情参见[8.1.3新增数据库特性](#)。

V8.1.1.300 版本

发布时间：2022年4月15日

配套内核版本：8.1.1

1. 扩容操作中，重分布支持设置并发数量。
2. 支持创建MRS3.0.x自定义类型集群数据源连接。
3. 新增集群缩容功能，对已有集群进行缩容操作，以便充分利用GaussDB(DWS) 提供的计算资源和存储资源。
4. 新增逻辑集群在线扩容功能。
5. 新增日志服务功能，可在LTS云日志服务查看采集的集群日志或进行日志转储。
6. 支持快照容量大小监控及展示，支持冷数据容量展示。
7. 数据库监控（DMS）更新及新增以下功能：
 - a. 集群概览页面更新，为客户展示了集群的状态，实时资源消耗，TOP SQL，集群资源消耗，数据库主要指标等信息。
 - b. 查询监控变更为实时查询，展示了当前运行在集群中的所有查询和会话的实时信息。
 - c. 会话监控变更为历史查询，展示了当前集群中的所有历史查询信息。
 - d. 新增DDL审核功能，DDL审核是SQL审核范畴，用于对DDL元数据的规范性检测，方便用户对潜在的表定义问题提前感知。
 - e. 新增SQL探针功能，提供SQL探针上传，校验，一键执行和定时执行探针任务等功能。
 - f. 新增负荷诊断报告(WDR)功能，为数据库提供性能数据收集和分析，用户可通过创建负荷信息快照记录指定时间段集群的负荷信息数据。
8. 告警页面新增资源池队列阻塞告警。

V8.1.1.205 版本

发布时间：2022年3月30日

配套内核版本：8.1.1

1. 解决更新现网关键问题，不涉及刷新资料。

V8.1.1.203 版本

发布时间：2022年3月18日

配套内核版本：8.1.1

1. 控制台页面新增“磁盘扩容”功能，通过磁盘扩容可快速缓解存储资源瓶颈现象。
2. “实时数仓”支持分布式部署，支持能力和标准数仓一致，单机部署支持能力不变。

V8.1.1.202 版本

发布时间：2022年2月24日

配套内核版本：8.1.1

1. 新增“实时数仓”集群类型，适用于HTAP混合负载场景，满足“一库两用，生产即分析”场景。
2. 控制台页面新增“主备恢复”功能，将“非均衡”状态的集群实例进行主备切换，恢复集群性能。
3. IoT数仓支持绑定弹性负载均衡（ELB），解决IoT数仓的CN单点故障问题。
4. 控制台页面支持的GUC参数范围扩展增强：新增GUC参数支持在管理控制台上进行修改。

V8.1.1.200 版本

发布时间：2021年12月9日

配套内核版本：8.1.1

1. 新增数据库监控（DMS）功能，支持数据库细粒度监控，包括节点、性能、数据库、会话、查询、实例和负载等指标。
2. 新增告警管理功能，支持在控制台查看DMS触发的告警信息。
3. 支持分段扩容功能，包括添加节点、集群扩容和数据库重分布三个阶段。
4. 新增部分GUC参数支持在控制台进行设置。

V8.1.1.100 版本

发布时间：2021年7月30日

配套内核版本：8.1.1

1. 新增管理CN节点功能，支持CN最大个数由5调整为20，并支持在管理控制台进行增删CN操作。
2. 新增在线扩容功能，在集群扩容过程中，支持数据持续入库。
3. 支持敏感操作保护功能，当用户在控制台进行敏感操作时，则需要输入证明身份的凭证，身份验证通过后方可进行相关操作。
4. 新增双集群容灾功能，即在另一个可用区（Region内）部署一个同构的GaussDB(DWS)灾备集群，保障业务连续性。
5. 数据库内核8.1.1上线：优化详情请参见[最新动态](#)。

V8.1.0.100 版本

发布时间：2021年5月15日

配套内核版本：8.1.0

1. 新增逻辑集群功能，数据库中的表可按逻辑集群来分配到不同的物理节点，而一个逻辑集群也可以包含多个数据库的表。
2. 创建集群中默认数据库名更改为gaussdb。
3. 数据库内核8.1.0上线：优化详情请参见[最新动态](#)。

V8.0.1.100 版本

发布时间：2020年7月31日

配套内核版本：8.0.0.1

1. 新增IoT数仓集群类型，提供时序计算和时序压缩服务。
2. 新增资源管理功能，实现两种作业同时执行互不影响。
3. 新增gs_dump，gs_dumpall数据导出工具。

V1.7.1 版本

发布时间：2020年3月20日

1. 数据库内核新增内容：
 - UPDATE支持别名以及表名引用。
 - 支持创建索引时指定schema，指定的schema需与基表所在schema一致。
 - 新增alter table modify null/not null语法。
 - 支持alter sequence maxvalue操作。
 - 对于临时表空间和算子落盘空间进行限额管控。
 - TopSQL中去除对于子语句的记录。
 - 支持对于CN轻量化语句的记录。
 - 对于历史TopSQLinfo表中数据老化处理。
 - CN Retry支持存储过程、函数语句。

V1.5.200 版本

发布时间：2019年8月8日

- 优化功能
 - a. 提升自动快照的备份频率，默认每天三次自动快照，从00:00（UTC）时间开始。快照执行周期可以在创建集群时进行设置，也可以对集群已有自动快照策略进行修改。
 - b. 在集群列表的“更多”操作中，可以对该集群执行手动创建快照的操作。