

数据仓库服务

# 产品介绍

文档版本 01  
发布日期 2025-09-02



版权所有 © 华为云计算技术有限公司 2025。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

<b>1 什么是数据仓库服务</b> .....	<b>1</b>
<b>2 数据仓库类型</b> .....	<b>9</b>
<b>3 数据仓库规格</b> .....	<b>12</b>
<b>4 产品优势</b> .....	<b>22</b>
<b>5 应用场景</b> .....	<b>24</b>
<b>6 产品功能</b> .....	<b>30</b>
<b>7 基本概念</b> .....	<b>35</b>
<b>8 与其他云服务的关系</b> .....	<b>38</b>
<b>9 安全</b> .....	<b>41</b>
9.1 责任共担.....	41
9.2 身份认证与访问控制.....	42
9.2.1 资源访问控制（IAM 权限控制）.....	42
9.2.2 数据库访问实现权限分层.....	43
9.2.3 GRANT 授权和 REVOKE 撤销授权.....	44
9.2.4 行级访问控制.....	45
9.3 网络安全.....	45
9.4 数据保护技术.....	48
9.4.1 数据备份.....	48
9.4.2 透明加密.....	49
9.4.3 SSL 传输加密.....	50
9.4.4 数据脱敏.....	50
9.4.5 使用函数进行加密.....	51
9.5 审计与日志.....	52
9.6 服务韧性.....	52
9.7 监控安全风险.....	53
9.8 认证证书.....	54
9.9 安全公告.....	55
9.9.1 漏洞修复说明.....	56
<b>10 DWS 权限管理</b> .....	<b>57</b>

---

11 如何访问 DWS.....	61
12 使用限制.....	63
13 技术支持.....	68
14 配额说明.....	69
15 DWS 技术指标.....	70

# 1 什么是数据仓库服务

数据仓库服务 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提供存算一体、存算分离多种产品形态，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](#)。

- **存算一体**：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，适用于“库、仓、市、湖”一体化的融合分析业务。
- **存算分离**：采用存算分离云原生架构，计算、存储分层弹性伸缩，极致性价比，采用多逻辑集群（Virtual Warehouse，以下简称VW）共享存储技术，实现不同负载的计算隔离和并发扩展，适用于OLAP分析场景。

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。

## 集群逻辑架构

DWS集群逻辑架构如[图1-1](#)所示。实例的详细介绍请参见[表1-1](#)。

图 1-1 集群逻辑架构图

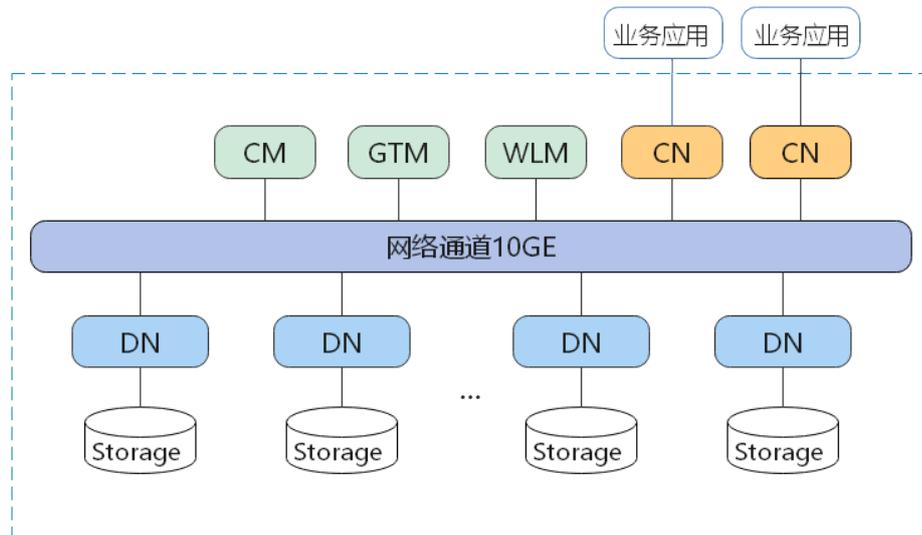


表 1-1 集群架构说明

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"> <li>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。</li> <li>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。</li> </ul> <p><b>说明</b> 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"> <li>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。</li> </ul> <p>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>

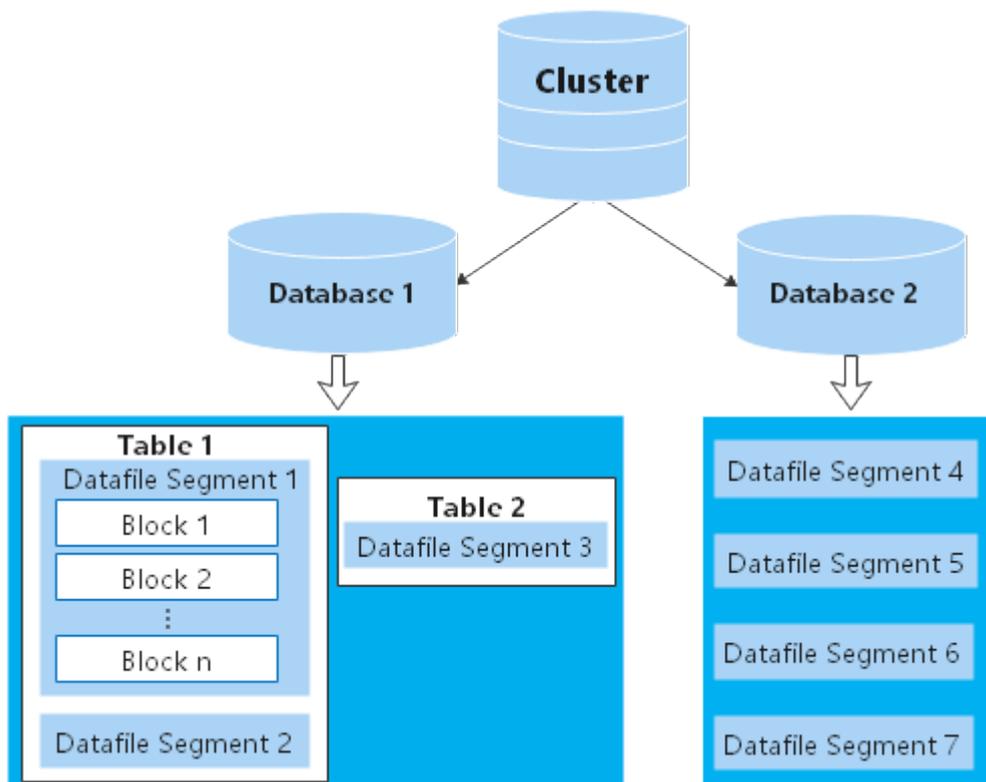
名称	描述	说明
GTM	全局事务管理器（Global Transaction Manager），负责生成和维护全局事务ID、事务快照、时间戳等全局唯一的信息。	整个集群只有一组GTM：主、备GTM各一个。
WLM	工作负载管理器（Workload Manager）。控制系统资源的分配，防止过量业务负载对系统的冲击而导致业务拥塞和系统崩溃。	不同于集群中的实例（GTM、CM、CN、DN）模块，不需要在安装过程中指定主机名称。安装程序会自动在各主机上安装此模块。
CN	协调节点（Coordinator）。负责接收来自应用的访问请求，并向客户端返回执行结果；负责分解任务，并调度任务分片在各DN上并行执行。	<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一个CN，请参见<a href="#">集群绑定和解绑ELB</a>。</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>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。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"> <li>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。</li> <li>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。</li> <li>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。</li> </ul> <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1-2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 1-2 数据库逻辑结构图

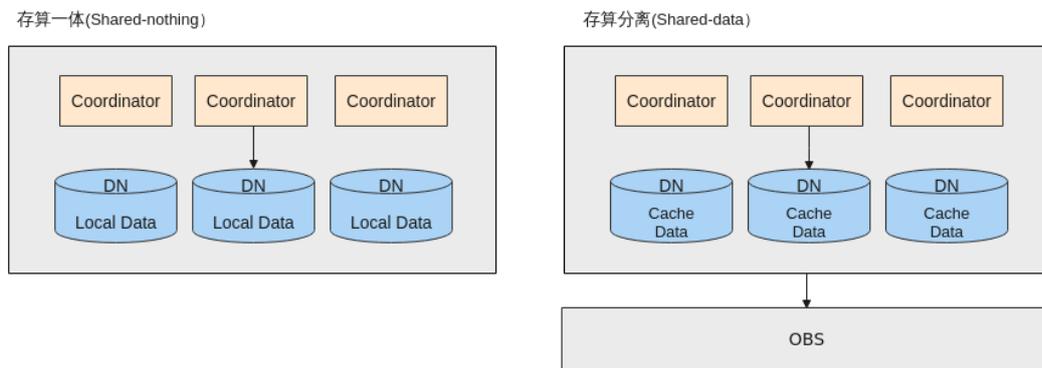


## 集群物理架构

DWS支持存算一体架构和存算分离架构。

其中，存算一体架构，数据存储在DN本地盘上。存算分离架构，DN本地盘仅做数据缓存和存储元数据，用户数据存储在OBS对象存储上。您可以根据需要选择相应的架构。

图 1-3 架构选择



## 存算一体架构

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 1-4 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DWS集成。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DWS平滑迁移。
- 接口**  
 支持应用程序通过标准JDBC和ODBC连接DWS。
- DWS**  
 一个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DWS集群在某一时间点的完整备份，记录了该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。

## 存算分离架构

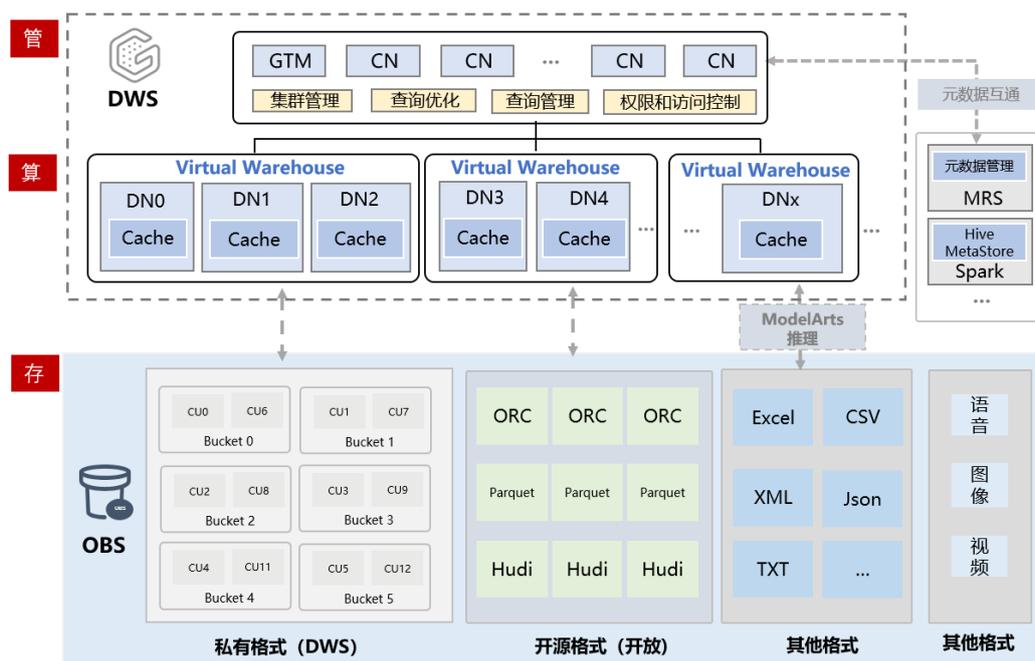
DWS全新推出云原生数仓存算分离版本，利用云基础设施提供的资源池化和海量存储能力，结合MPP数据库技术，采用计算存储分离架构，实现了极致弹性、实时入库、数据实时共享和湖仓一体等特性。

云原生数仓存算分离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，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。

存算分离具有以下优势：

- **湖仓一体**：提供简单、易维护的湖仓一体体验，无缝对接DLI，支持元数据自动导入、外部表查询加速、内外表关联查询，支持数据湖格式读写，简化数据入湖入仓。
- **实时写入**：提供H-Store存储引擎，对实时写入场景进行了设计优化，支持高吞吐实时写入与更新，同时支持大批量写入场景。
- **极致弹性**：计算资源快速伸缩，存储空间按需使用，同时大幅度降低存储成本。历史数据无需再迁移到其他存储介质上，让数据分析更简单，一站式解决金融、互联网等行业快速增长的数据分析需求。
- **数据共享**：一份数据承载多样负载，数据实时共享，多写多读的使用模式，在支持不同业务数据快速共享的同时，具备良好的计算资源隔离能力。

图 1-5 存算分离架构



- 极致弹性
  - 逻辑集群（Virtual Warehouse）随业务需求并发扩展。
  - 多VW间数据实时共享，一份数据承载多样负载，无需拷贝。
  - 通过多VW实现吞吐/并发的线性提升，同时具备良好的读写分离、负载隔离能力。
- 湖仓一体
  - 数据湖与数据仓库数据无缝混合查询。
  - 数据湖分析体验数仓的极致性能和精准管控制度。

### 存算一体与存算分离产品形态对比

表 1-2 存算分离与存算一体差异

数仓类型	存算一体	存算分离
------	------	------

<p><b>存储介质</b></p>	<p>数据存储在计算节点的本地磁盘（根据选择的存储类型决定存储在SSD云盘或SSD本地盘上）。</p> <ul style="list-style-type: none"> <li>SSD云盘使用SSD类型的EVS作为数据存储介质，存储容量更加灵活，云盘支持后期进行磁盘扩容，推荐您使用SSD云盘。</li> <li>SSD本地盘使用ECS规格自带的本地磁盘作为数据存储介质，容量固定，但本地盘不支持磁盘扩容及规格调整，在容量不足的情况下，只能通过节点扩容方式解决。</li> </ul>	<p>列存数据存储在华为云对象存储（OBS），计算节点本地磁盘主要作为OBS数据的查询缓存，行存仍然存储在计算节点本地磁盘。</p>
<p><b>产品优势</b></p>	<p>数据存储在计算节点本地磁盘，性能高。</p>	<p>存算分离，计算、存储分层弹性，存储按需使用，计算快速伸缩，无限算力、无限容量。</p> <p>数据存储在对象存储上，存储成本更低，多VW支持的并发更高。</p> <p>支持数据共享，支持湖仓一体。</p>

# 2 数据仓库类型

## 产品类型概述

- **存算一体**：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、低时延、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，适用于“库、仓、市、湖”一体化的融合分析业务。
- **存算分离**：采用存算分离云原生架构，计算、存储分层弹性伸缩，极致性价比，采用多逻辑集群（Virtual Warehouse，以下简称VW）共享存储技术，实现不同负载的计算隔离和并发扩展，适用于OLAP分析场景。

### 📖 说明

- 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	存算一体	存算分离
总览	资源	支持	支持
	告警	支持	支持
	近期事件	支持	支持
	集群监控指标（DMS）	支持	支持
数据	-	支持	支持
集群管理	SQL编辑器	支持	支持
	监控面板（DMS）	支持	支持
	查看监控指标（Cloud Eye）	支持	支持
	重启	支持	支持

功能模块	功能模块	存算一体	存算分离
	启动	支持	支持
	停止	支持	支持
	扩容	支持	支持
	缩容	支持	支持
	重分布	支持	支持, 备注1
	查看重分布详情	支持	支持
	弹性变更规格	支持	支持
	经典变更规格	支持	不支持
	重置密码	支持	支持
	创建快照	支持	支持
	解除只读	支持	支持
	删除	支持	支持
	管理CN节点	支持	支持
	磁盘扩容	支持	支持
	集群详情	基本信息	支持
ELB负载均衡		支持	支持
资源管理		支持	支持
智能运维		支持	支持
逻辑集群		支持	支持
快照		支持	支持
参数修改		支持	支持
安全设置		支持	支持
MRS数据源		支持	支持
标签		支持	支持
节点管理		支持	支持
升级管理		支持	支持
日志服务		支持	支持
用户管理	支持	支持	
集成	数据迁移	支持	支持
容灾管理	容灾管理	支持	不支持

功能模块	功能模块	存算一体	存算分离
快照管理	恢复	支持	支持
	删除	支持	支持
	复制	支持	支持
事件管理	事件管理（通用）	支持	支持
告警管理	告警管理	支持	支持
连接客户端	连接客户端	支持	支持
其他模块	巡检	支持	支持
	智能运维	支持	支持
	节点修复	支持	支持
	租户侧温备	支持	支持

#### 说明

- 备注1：存算分离表数据存储在OBS上，无需重分布，但是元数据和索引存储在本地，仍然需要进行重分布。存算分离表在重分布时，表只支持读，元数据的重分布时间一般比较短，但是，如果表上创建了索引，索引会影响重分布的性能，重分布完成时间与索引的数据量成正比关系，在此期间，表只支持读。
- 存算分离集群仅9.0.2及以上集群版本支持快照功能。

# 3 数据仓库规格

DWS的规格按照产品类型分为存算一体和存算分离。各产品类型的不同差异，详情请参见[数据仓库类型](#)。

## 说明

低配置集群，如内存16G、vCPU4核及以下的规格，建议不要用于生产环境，可能会导致资源过载风险。

## 存算一体规格

- 存算一体1:8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](#)。
- 存算一体1:4云盘规格，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，规格详情请参见[表3-2](#)。
- 存算一体本地盘规格，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3-3](#)。

## 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 存算一体 1:8 云盘规格

规格名称	CP U架 构	vC PU	内 存 ( GB )	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.large.m7	X86	4	32	20GB ~ 2000GB	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。

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
dwsk2.xlarge	ARM	4	32	20GB ~ 2000GB	100	10	800	1	
dwsx2.xlarge.m7n	X86	4	32	20GB ~ 2000GB	100	10	800	1	
dwsk2.xlarge.km2	ARM	4	32	20GB ~ 2000GB	100	10	800	1	
dwsx2.2xlarge.m7	X86	8	64	100GB ~ 4000GB	200	100	1600	1	适用于中小规模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	
dwsk2.2xlarge.km2	ARM	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	
dwsk2.4xlarge.km2	ARM	16	128	100GB ~ 8000GB	400	100	3200	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
dwsx2.8xlarge.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	
dwsk2.8xlarge.km2	ARM	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	
dwsx2.16xlarge.m7n	X86	64	512	100GB ~ 32000GB	1600	100	12800	4	
dwsx2.16xlarge.m7	X86	64	512	100GB ~ 32000GB	1600	100	12800	4	
dwsk2.16xlarge	ARM	64	512	100GB ~ 32000GB	1600	100	12800	4	
dwsx2.24xlarge.m7	X86	96	768	100GB ~ 48000GB	2400	100	19200	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
dwsk2.24xlarge	ARM	96	768	100GB ~ 48000GB	2400	100	19200	4	
dwsx2.32xlarge.m7	X86	128	1024	100GB ~ 48000GB	3200	100	25600	4	

表 3-2 存算一体 1:4 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
dwsx2.h.xlarge.4.c7	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.kc1	ARM	4	16	20GB ~ 2000GB	20	1	
dwsk2.h.xlarge.kc2	ARM	4	16	20GB ~ 2000GB	20	1	
dwsx2.h.xlarge.4.c7n	X86	4	16	20GB ~ 2000GB	20	1	
dwsx2.h.2xlarge.4.c6	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.h.2xlarge.4.c7	X86	8	32	100GB ~ 4000GB	100	1	
dwsk2.h.2xlarge.4.kc1	ARM	8	32	100GB ~ 4000GB	100	1	
dwsk2.h.2xlarge.kc2	ARM	8	32	100GB ~ 4000GB	100	1	
dwsx2.h.2xlarge.4.c7n	X86	8	32	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.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	
dwsk2.h.4xlarge.kc2	ARM	16	64	100GB ~ 8000GB	100	1	
dwsx2.h.4xlarge.4.c7n	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	
dwsk2.h.8xlarge.kc2	ARM	32	128	100GB ~ 16000GB	100	2	
dwsx2.h.8xlarge.4.c7n	X86	32	128	100GB ~ 16000GB	100	2	
dwsx2.h.8xlarge.4.c7n	X86	32	128	100GB ~ 16000GB	100	2	
dwsk2.h.12xlarge.4.kc1	ARM	48	192	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。
dwsk2.h.12xlarge.kc2	ARM	48	192	100GB ~ 24000GB	100	4	
dwsx2.h.16xlarge.4.c7	X86	64	256	100GB ~ 32000GB	100	4	
dwsx2.h.16xlarge.4.c7n	X86	64	256	100GB ~ 32000GB	100	4	
dwsk2.h.16xlarge	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
dwsk2.h.2 4xlarge	ARM	96	384	100GB ~ 48000GB	100	4	
dwsk2.h.3 2xlarge	ARM	128	512	100GB ~ 64000GB	100	4	

表 3-3 存算一体本地盘规格

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	

## 存算分离规格

- 存算分离云盘规格，该规格弹性伸缩，无限算力、无限容量，规格详情请参见表 3-4。
- 存算分离本地盘规格，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 3-5。

### 说明

创建存算分离集群时规格仅显示后半部分（例如4U16G.4DPU），下列规格列表中前缀（dwsx3/dwsax3/dwsk3）代表存算分离对应的CPU架构。

表 3-4 存算分离云盘规格

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。
dwsx3.4U 16G.4DP U	ARM	4	16	20GB~200 0GB	10	1	
dwsax3.4 U16G.4D PU	X86	4	16	20GB~200 0GB	10	1	
dwsax3.4 U32G.4D PU	X86	4	32	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。
dwsx3.8U 32G.8DP U	ARM	8	32	100GB~40 00GB	100	1	
dwsax3.8 U32G.8D PU	X86	8	32	100GB~40 00GB	100	1	
dwsax3.8 U64G.8D PU	X86	8	64	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。
dwsx3.16 U64G.16 DPU	ARM	16	64	100GB~80 00GB	100	1	
dwsax3.1 6U64G.16 DPU	X86	16	64	100GB~80 00GB	100	1	
dwsax3.1 6U128G.1 6DPU	X86	16	128	100GB~80 00GB	100	1	
dwsx3.32 U128G.32 DPU	X86	32	128	100GB~16 000GB	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
dwsk3.32 U128G.32 DPU	ARM	32	128	100GB~1600GB	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。
dwsax3.3 2U128G.3 2DPU	X86	32	128	100GB~1600GB	100	2	
dwsax3.3 2U256G.3 2DPU	X86	32	256	100GB~1600GB	100	2	
dwsk3.48 U192G.48 DPU	ARM	48	192	200GB~2400GB	100	4	
dwsx3.64 U256G.64 DPU	X86	64	256	200GB~3200GB	100	4	
dwsk3.64 U256G.64 DPU	ARM	64	256	100GB~3200GB	100	4	
dwsax3.6 4U256G.6 4DPU	X86	64	256	100GB~3200GB	100	4	
dwsax3.6 4U512G.6 4DPU	X86	64	512	100GB~3200GB	100	4	
dwsx3.96 U768G.96 DPU	X86	96	768	100GB~4800GB	100	4	
dwsk3.96 U384G.96 DPU	ARM	96	384	100GB~4800GB	100	4	
dwsax3.9 6U384G.9 6DPU	X86	96	384	100GB~4800GB	100	4	
dwsax3.9 6U768G.9 6DPU	X86	96	768	100GB~4800GB	100	4	
dwsx3.12 8U1024G. 128DPU	X86	128	1024	100GB~6400GB	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
dwsk3.12 8U512G.1 28DPU	ARM	128	512	100GB~64 000GB	100	4	
dwsax3.1 28U512G. 128DPU	X86	128	512	100GB~64 000GB	100	4	
dwsax3.1 28U1024 G.128DP U	X86	128	1024	100GB~64 000GB	100	4	

表 3-5 存算分离本地盘规格

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.ki1.16DP U	ARM	16	64	5960GB	1	
dws3.32U2 56G.i7.32DP U	X86	32	256	5960GB	2	
dws3.32U1 28G.ki1.32D PU	ARM	32	128	11920GB	2	
dws3.64U5 12G.i7.64DP U	X86	64	512	11920GB	4	
dws3.64U2 28G.ki1.64D PU	ARM	64	228	23840GB	4	

## DEC 场景节点规格

表 3-6 规格说明（仅在专属云 DEC 场景下支持）

节点规格	vCPU 核数	内存大小 (GB)	架构	规格类型
dwsx2.2xlarge.dec	8	64	X86	ECS/存算分离EVS
dwsx2.8xlarge.dec	32	256	X86	ECS/存算分离EVS
dwsx2.16xlarge.dec	64	512	X86	ECS/存算分离EVS
dwsk2.2xlarge.dec	8	64	ARM	ECS/存算分离EVS
dwsk2.8xlarge.dec	32	256	ARM	ECS/存算分离EVS
dwsk2.12xlarge.dec	48	384	ARM	ECS/存算分离EVS
dws2.olap.4xlarge.i3. dec	16	128	X86	ECS/本地直通
dws2.olap.8xlarge.i3. dec	32	256	X86	ECS/本地直通
dws2.olap.16xlarge.i 3.dec	64	512	X86	ECS/本地直通
dws2.olap.4xlarge.ki 1.dec	16	64	ARM	ECS/本地直通
dws2.olap.8xlarge.ki 1.dec	32	128	ARM	ECS/本地直通
dws2.olap.16xlarge.k i1.dec	64	228	ARM	ECS/本地直通

# 4 产品优势

数据仓库服务 DWS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

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

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

## 高性能

- 云化分布式架构

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

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。

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。

此外，DWS基于传统列存执行引擎的性能瓶颈点做出了独创性突破，推出Turbo引擎，相比原列存执行引擎，对字符串、numeric类型做了内存和磁盘存储格式优

化，且对常用sort/agg/join/scan等算子做了极致性能优化，使得执行器整体性能提升1倍左右，可显著降低业务计算成本。

- 数据加载快

DWS提供了GDS极速并行大规模数据加载工具。

- 列存下的数据压缩

对于非活跃的早期数据可以通过压缩来减少空间占用，降低采购和运维成本。

DWS列存储压缩支持Delta Value Encoding、Dictionary、RLE、LZ4、ZLIB等压缩算法，且能够根据数据特征自适应的选择压缩算法，平均压缩比7:1。压缩数据可直接访问，对业务透明，极大缩短历史数据访问的准备时间。

## 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)，表级别扩容技术，扩容期间业务不中断、无感知。
- 支持在线升级：8.1.1及以上源版本支持大版本在线升级，8.1.3及以上源版本支持补丁在线升级，升级期间用户无需停止业务，业务存在闪级秒断。

## 高可靠

- 事务管理

- 支持事务块，用户可以通过start transaction命令显式启动一个事务块。
- 支持单语句事务，用户不显式启动事务，则单条语句就是一个事务。
- 分布式事务管理。支持全局事务信息管理，包括gxid、snapshot、timestamp的管理，分布式事务状态管理，gxid溢出的处理。
- 分布式事务支持ACID特性（Atomicity, Consistency, Isolation, Durability），数据强一致保证。
- 支持分布式死锁预防，保证在出现死锁时自动解锁或者预防死锁。

- 全方位HA设计

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

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。DWS还支持自动数据全量、增量备份，提升数据可靠性。

## 低成本

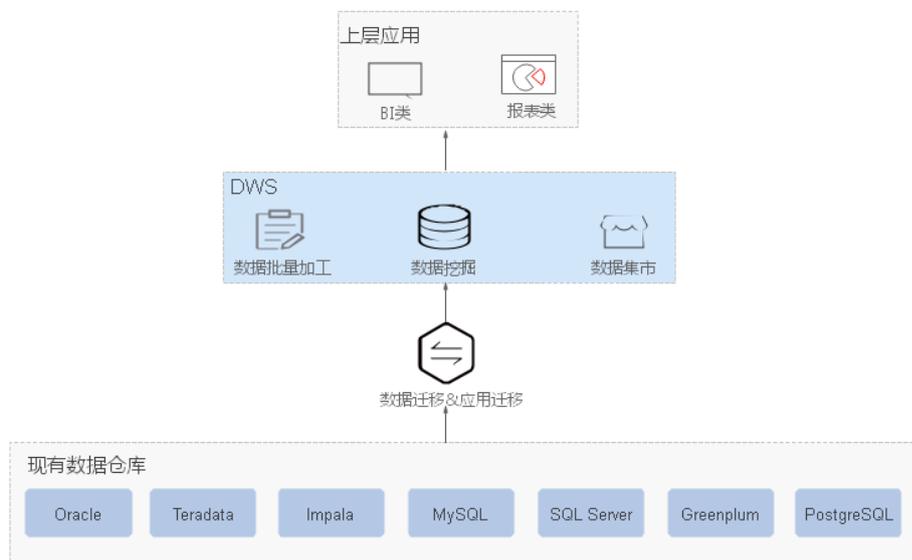
- 按需付费：DWS按实际使用量和使用时长计费。您需要支付的费率很低，只需为实际消耗的资源付费。
- 门槛低：您无需前期投入较多固定成本，可以从低规格的数据仓库实例起步，后续随时根据业务情况弹性伸缩所需资源，按需开支。

# 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。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

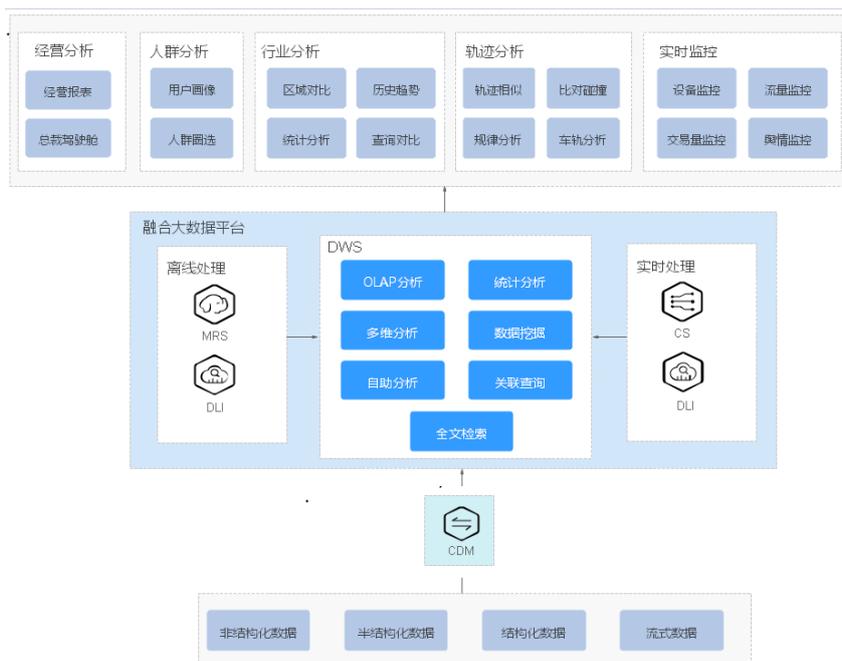
- 平滑迁移  
DWS提供配套的迁移工具，可支持TeraData、Oracle、MySQL、SQL Server、PostgreSQL、Greenplum、Impala等常用数据分析系统的平滑迁移。
- 兼容传统数据仓库  
DWS支持SQL 2003标准，兼容Oracle的部分语法和数据结构，支持存储过程，可与常用BI（business intelligence）工具无缝对接，业务迁移修改量极小。
- 安全可靠

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DWS支持数据自动全量、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”，成为助力用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

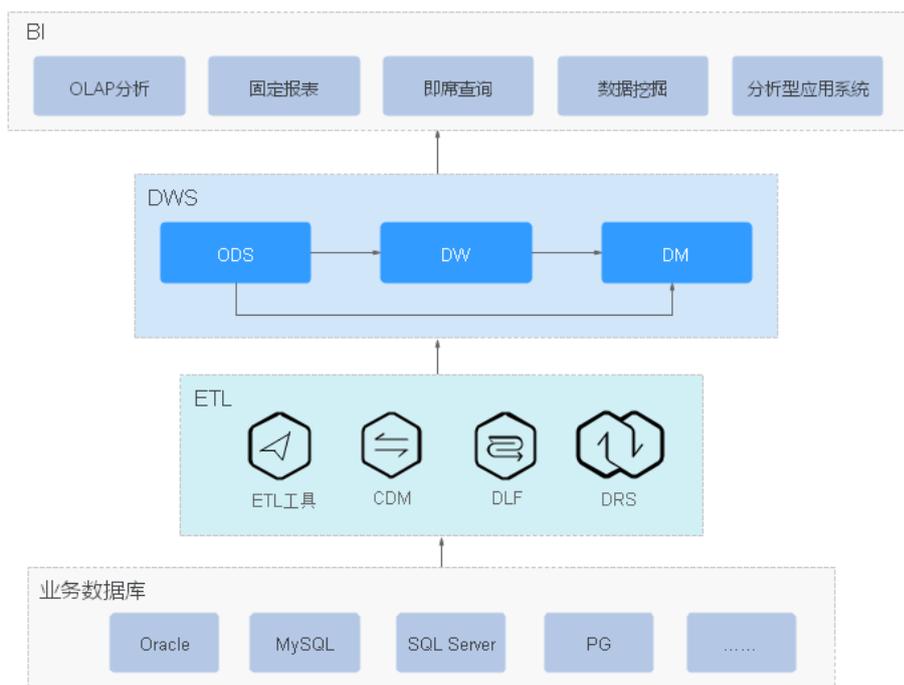
数据仓库在整个BI系统中起到了支柱的角色，更是海量数据收集、存储、分析的核心。为金融、教育、移动互联网、O2O（Online to Offline）等行业提供强大的商业决策分析支持。

### 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。

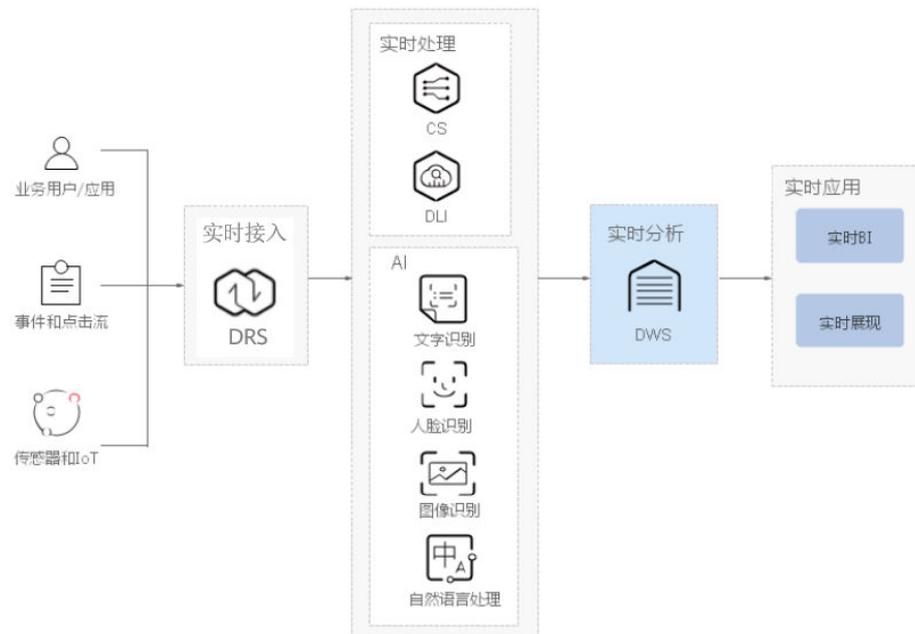
图 5-3 增强型 ETL+实时 BI 分析



## 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，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

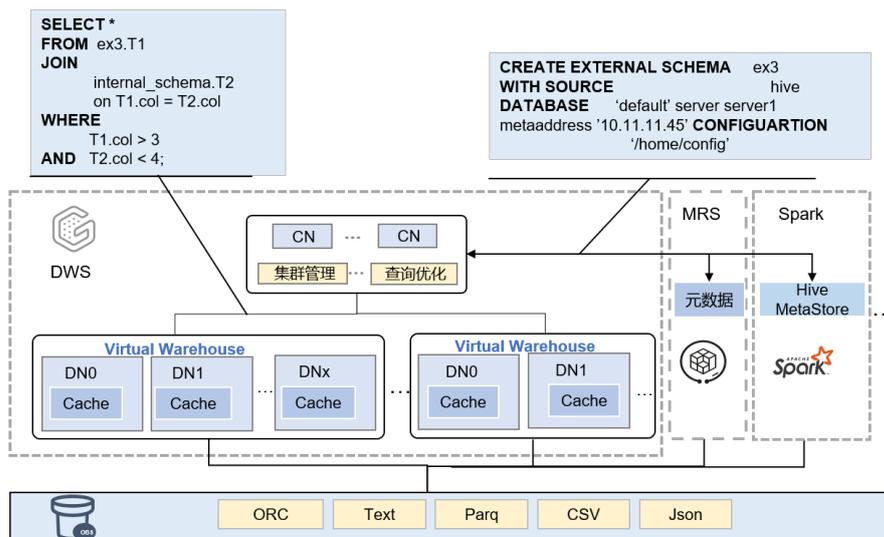


### 优势

- 流式数据实时入库  
互联网等数据经过流计算及AI服务处理后，可实时写入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DWS中与其他业务数据进行关联分析，实现融合数据分析。

### 湖仓一体

- 无缝访问数据湖
  - 对接Hive Metastore元数据管理，直接访问数据湖的数据表定义，无需用户创建外表，只需创建external schema即可。
  - 支持主要数据格式：ORC，Parquet。
- 融合查询
  - 混合查询数据湖和仓内的任意数据。
  - 查询一步到位输出到仓内/数据湖，无需额外数据中转拷贝。
- 极致查询性能
  - 使用数仓高质量的查询计划和高效的执行引擎。
  - 使用数仓的负载管理手段，精准控制。

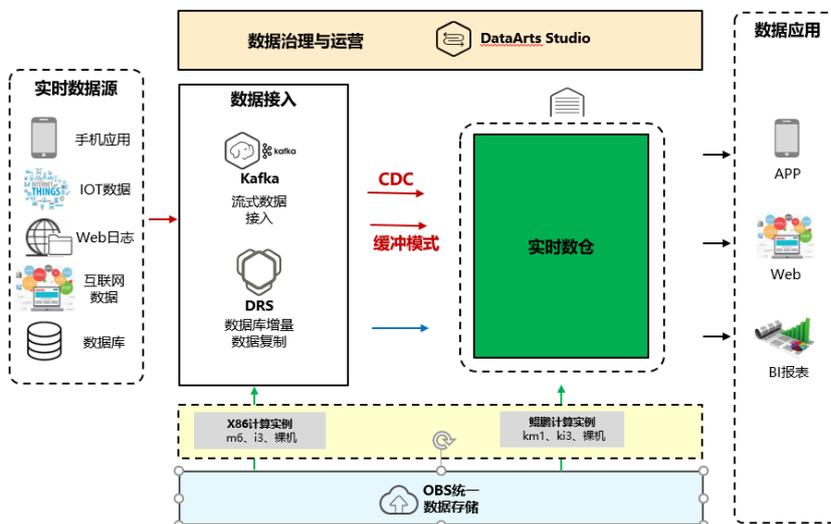


## 实时写入

提供H-Store存储引擎，微批数据存放在本地，周期合并到OBS存储，支持高吞吐实时写入与更新，同时支持大批量写入场景。

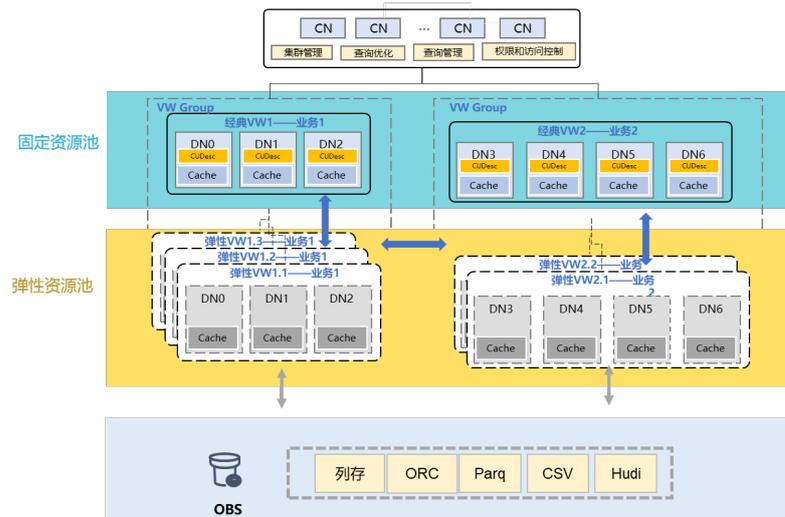
数据实时写入并经实时计算之后，可以用于实时大屏、实时分析、实时监控、实时风控、实时推荐。

## 实时数据接入，实时分析



## 多计算组支持业务隔离和极致弹性（存算分离架构）

- 通过多计算组（Virtual Warehouse，简称VW）实现业务的负载隔离，相比资源管理的软隔离，基于VW的虚拟机级别的硬隔离可以最大化的降低业务间的互相影响。
- 支持多经典VW和多弹性VW。



- 通过经典VW实现业务隔离：
  - 根据业务需要，可部署多个VW，不同业务绑定不同的VW，经典VW允许建表。
  - VW之间资源隔离，实现业务之间互不影响。
  - VW之间数据实时共享。
  - 固定VW的大小在MPP架构下决定了单SQL性能上限。
  - 固定VW适合承载实时接入、实时加工等稳定负载/低时延作业，根据业务变化提前规划大小。
- 通过弹性VW实现并发扩展：
  - VW在高并发场景下业务出现排队时，动态弹出弹性VW，弹性VW支持读写，不支持建表。
  - 排队查询自动路由到弹性VW。
  - 通过弹性VW，减少查询排队，提升业务整体并发度。
  - 业务并发下降后，自动回收弹性VW。
  - 弹性VW实现计算资源按需供给，用户设定弹性上限。
  - 弹性VW规格与固定VW相同，在弹性VW上运行的SQL，性能与固定VW基本相同。
  - 弹性VW按资源使用时长计费。
  - 弹性VW适合承载突增、周期性的查询、加工负载。

比如：客户有多个业务部门，那么可以为每个业务部门绑定一个经典VW，实现不同业务之间的资源隔离，如果业务1使用一个3节点VW，业务2使用一个4节点VW。如果业务1，在每天的10:00-12:00有业务高峰，那么可以为业务1配置弹性VW，在高峰期弹出弹性VW，在高峰期过后销毁。

# 6 产品功能

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

## 企业级数据仓库和标准 SQL 的支持

当创建好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。

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。

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。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，DWS提供了资源管理功能，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

逻辑集群是基于节点组（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。

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

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DWS
- 使用CDM迁移数据到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DWS集群三权分立](#)。

## 监控与审计

- 监控集群  
DWS与云监控（Cloud Eye）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）是为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（DMS）](#)。
- 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。该功能以默认告警规则阈值的方式提供常用的DWS告警监控，还允许用户根据自己的业务特点，自定义告警阈值。详情请参见[告警管理](#)。

- 事件通知  
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
  - 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](#)。
  - 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

DWS提供了以下几款自研工具，用户可以在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](#)》。

- SQL编辑器  
DWS线上SQL编辑器提供了一站式数据开发、接入、处理等功能，支持直接在DWS管理控制台上连接集群数据库进行SQL编辑执行，以使用户能更好的使用DWS。
- 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 Converter ) 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。
- 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

## DWS 的管理概念

- 集群  
表示由多个节点组成的服务器群组。DWS以集群为单位进行组织，一个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DWS集群部署节点取值范围为3~256，每个节点都支持存储与分析数据。详情请参见[DWS技术指标](#)。
- 规格  
用户创建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。

## DWS 的数据库概念

- 数据库  
数据库用于管理各类数据对象，与其他数据库隔离。创建数据库时可以指定对应的表空间，如果不指定相应的表空间，相关的对象会默认保存在PG\_DEFAULT空间中。数据库管理的对象可分布在多个表空间上。
- 联机分析处理  
联机分析处理（Online Analytical Processing，简称OLAP），是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，

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DWS集群对应一个数据库版本，用户在创建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。

- 数据库用户和角色

DWS使用用户和角色来控制对数据库的访问。根据角色自身的设置不同，一个角色可以看做是一个数据库用户，或者一组数据库用户。在DWS中角色和用户之间的区别只在于角色默认是没有LOGIN权限的。在DWS中一个用户唯一对应一个角色，不过可以使用角色叠加来更灵活地进行管理。

- 实例

实例在DWS中是运行在内存中的一组数据库进程，一个实例可以管理一个或多个数据库，这些数据库组成一个集簇。集簇是存储磁盘上的一个区域，这个区域在安装时初始化并由一个目录组成，所有数据都存储在这个目录中，这个目录被称为数据目录，使用initdb创建。理论上来说一个服务器上可以在不同的端口启动多个实例，但是DWS一次只能管理一个实例，启动和停止都是依赖于具体的数据目录。以后由于兼容的需要不排除引入实例名这个概念的可能。

- 表空间

在DWS中，表空间是一个目录，可以存在多个，里面存储的是它所包含的数据库的各种物理文件。由于表空间是一个目录，仅是起到了物理隔离的作用，其管理功能依赖于文件系统。

- 模式

DWS的模式是对数据库做一个逻辑分割。所有的数据库对象都建立在模式下面。DWS的模式和用户是弱绑定的，所谓的弱绑定是指虽然创建用户的同时会自动创建一个同名模式，但用户也可以单独创建模式，并且为用户指定其他的模式。

- V2表

V2表特指建表环节时，执行CREATE TABLE语法中定义的colversion取值为2.0的表，表示列存表的每列合并存储在一个文件中，文件名以relfilenode.C1.0命名，数据存储在本地盘。存算一体场景下，不指定colversion取值时，用户创建的列存表默认为V2表。

- V3表

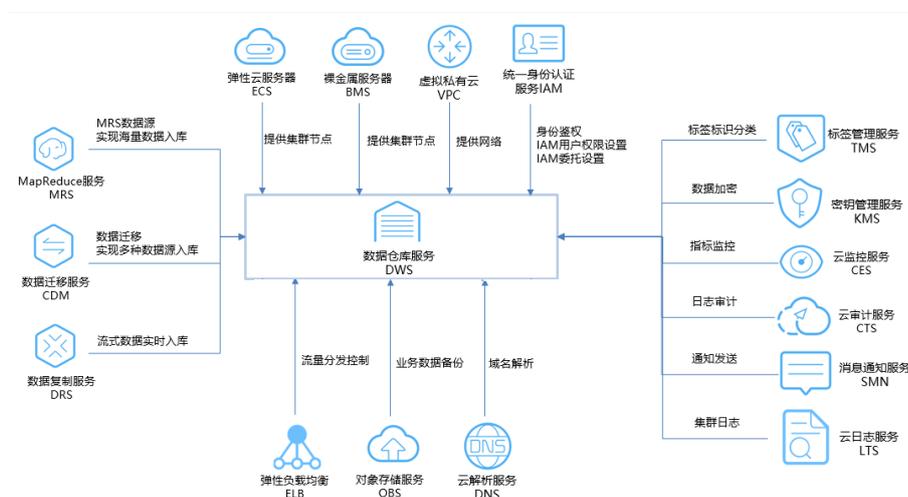
V3表特指建表环节时，执行CREATE TABLE语法中定义的colversion取值为3.0的表，即存算分离表，表示列存表的每列合并存储在一个文件中，文件名以C1\_field.0命名，数据存储在OBS文件系统。存算分离场景下，不指定colversion取值时，用户创建的列存表默认为V3表。

- 事务管理

在事务管理上，DWS采取了MVCC（多版本并发控制）结合两阶段锁的方式，其特点是读写之间不阻塞。DWS的MVCC没有将历史版本数据统一存放，而是和当前元组的版本放在了一起。DWS没有回滚段的概念，但是为了定期清除历史版本数据DWS引入了一个VACUUM进程。一般情况下用户不用关注它，除非要做性能调优。此外，DWS是自动提交事务。

- Node Group  
节点组，这是数据库内核的一个基本概念，即每个集群节点就是一个Node，若干个Node组成一个Node Group，一个物理集群包含多个Node Group，每个Node Group封装成用户感知的形态，就是一个逻辑集群（Logical Cluster）。
- Logical Cluster  
逻辑集群，存算一体形态下，每个物理集群下划分成多个节点组（Node Group），每个节点组对应一个逻辑集群。
- Virtual Warehouse  
计算组，即存算分离形态下的逻辑集群叫法，当前统一叫计算组。  
以上三者的关系：Node Group是数据库内核的叫法，封装成产品，存算一体集群中，叫逻辑集群；存算分离集群中，叫计算组。

# 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DWS集群，实现海量数据通过Hadoop处理后使用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DWS的迁移。

## 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。

与传统的本地数据中心相比，云计算的运营方和使用方分离，提供了更好的灵活性和控制力，有效降低了客户的运营负担。正因如此，云的安全性无法由一方完全承担，云安全工作需要华为云与您共同努力，如图9-1所示。

- **华为云：**无论在任何云服务类别下，华为云都会承担基础设施的安全责任，包括安全性、合规性。该基础设施由华为云提供的物理数据中心（计算、存储、网络等）、虚拟化平台及云服务组成。在PaaS、SaaS场景下，华为云也会基于控制原则承担所提供服务或组件的安全配置、漏洞修复、安全防护和入侵检测等职责。
- **客户：**无论在任何云服务类别下，客户数据资产的所有权和控制权都不会转移。在未经授权的情况，华为云承诺不触碰客户数据，客户的内容数据、身份和权限都需要客户自身看护，这包括确保云上内容的合法合规，使用安全的凭证（如强口令、多因子认证）并妥善管理，同时监控内容安全事件和账号异常行为并及时响应。

图 9-1 华为云安全责任共担模型



云安全责任基于控制权，以可见、可用作为前提。在客户上云的过程中，资产（例如设备、硬件、软件、介质、虚拟机、操作系统、数据等）由客户完全控制向客户与华为云共同控制转变，这也就意味着客户需要承担的责任取决于客户所选取的云服务。如图9-1所示，客户可以基于自身的业务需求选择不同的云服务类别（例如IaaS、PaaS、SaaS服务）。不同的云服务类别中，每个组件的控制权不同，这也导致了华为云与客户的责任关系不同。

- 在On-prem场景下，由于客户享有对硬件、软件和数据等资产的全部控制权，因此客户应当对所有组件的安全性负责。
- 在IaaS场景下，客户控制着除基础设施外的所有组件，因此客户需要做好除基础设施外的所有组件的安全工作，例如应用自身的合法合规性、开发设计安全，以及相关组件（如中间件、数据库和操作系统）的漏洞修复、配置安全、安全防护方案等。
- 在PaaS场景下，客户除了对自身部署的应用负责，也要做好自身控制的中间件、数据库、网络控制的安全配置和策略工作。
- 在SaaS场景下，客户对客户内容、账号和权限具有控制权，客户需要做好自身内容的保护以及合法合规、账号和权限的配置和保护等。

## 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DWS，但是不允许删除集群的权限，控制他们对DWS资源的使用范围。
- **场景二：**您希望您的员工只有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DWS](#)。



## 9.2.2 数据库访问实现权限分层

DWS中可以使用Database和Schema实现业务的隔离，区别在于：

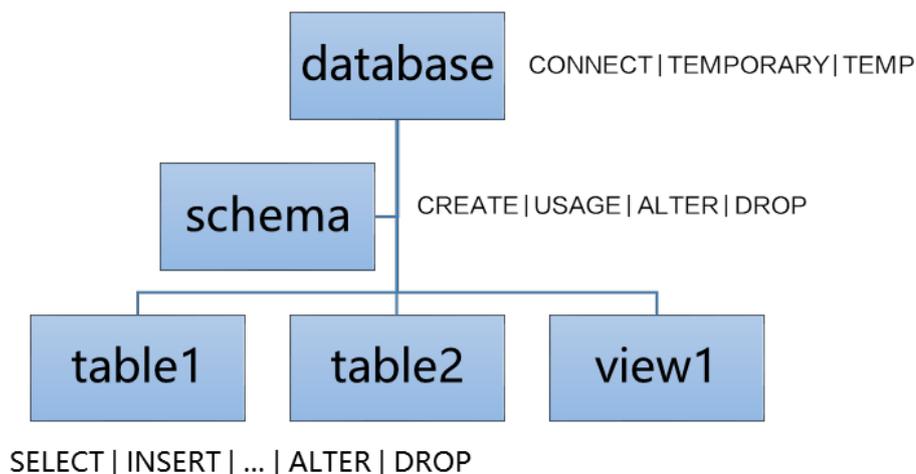
- Database之间无法直接互访，通过连接隔离实现彻底的权限隔离。各个Database之间共享资源极少，可实现连接隔离、权限隔离等。
- Schema隔离的方式共用资源较多，可以通过GRANT与REVOKE语法便捷地控制不同用户对各Schema及其下属对象的权限，从而赋给业务更多的灵活性。

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[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

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。  
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	45897565 3	2002-08-20 14:20:18	122445@hotmail.com	Shenzhen
23566	Liming	42297565 3	2002-09-20 11:20:18	134645@hotmail.com	Shanghai
...	...	...	...	...	...
23456	Lucy	42297458 3	2002-01-25 10:20:23	135666@hotmail.com	Shanghai
27556	Yuyi	42294568 3	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 网络安全

### 数据库安全设置

#### 账户说明

您在创建DWS集群时，系统会自动为集群创建如下账户，用于给数据库提供完善的后台运维管理服务。

- dbadmin: 系统管理员账户, 首次登录DWS数据库使用, 用于业务数据库创建、普通用户创建、权限分配等。
- 如果开通了三权分立, 则除了系统管理员dbadmin, 还会生成安全管理员、审计管理员账户, 账户名称由用户自定义, 详情参见[开启DWS数据库用户三权分立](#)。
- Ruby: 系统默认运维账户, 非云服务提供商运维人员不可使用。
- om\_user\_集群ID前8位: 其他运维账户, 且系统赋予了gs\_role\_analyze\_any, gs\_role\_vacuum\_any, gs\_role\_read\_all\_stats, gs\_role\_signal\_backend等预置权限, 用于故障定位, 非云服务提供商运维人员不可使用, 详情参见[授权DWS集群运维账号](#)。

### 账户密码复杂度设置

- 数据仓库服务的数据库系统管理员dbadmin密码在创建DWS集群时进行设定, 且DWS控制台为dbadmin用户提供了密码复杂度校验, 校验不通过, 无法设定密码。  
其复杂度要求请参见[创建DWS集群](#)中描述的密码复杂度要求。
- 其他普通数据库用户也相应设置了密码安全策略, 系统默认安全策略如下:
  - 长度为8~32个字符。
  - 至少包含大写字母、小写字母、数字或特殊字符中三种的组合。
  - 不能是用户名和用户名反序, 此条要求为非大小写敏感。
  - 不能是当前密码、当前密码的反序。

#### 须知

为避免恶意人员轻易破解您的数据库密码, 请按照界面提示和要求设置足够复杂度密码, 并定期修改。

### 自定义密码策略

以上默认安全策略支持用户修改, 详情参见[自定义密码策略](#)。

### 账户锁定及设置密码过期策略

- 数据库用户密码有效期: 通过password\_effect\_time控制, 一旦密码过了有效期, 则系统会提醒修改密码, 修改此参数请联系技术支持。
- 数据库用户密码输入错误次数: 通过failed\_login\_attempts控制, 当用户输错次数超过该设定值, 账户会自动锁定, 需要系统管理员进行解锁, 该参数可在DWS管理控制台进行设定, 设定方法参见[修改DWS集群GUC参数](#)。
- 账户锁定后自动解锁时间: 通过password\_lock\_time控制, 锁定时间超过设定值, 账户会自行解锁, 修改此参数请联系技术支持。

### 重置账户密码

- 系统管理员dbadmin账户被锁定或忘记密码时, 可通过DWS控制台进行重置密码操作。
- 普通用户密码重置可通过系统管理员dbadmin在后台通过SQL命令重置, 重置命令示例如下。参见[密码设置和修改](#)。

```
ALTER USER joe IDENTIFIED BY 'password';
```

## 配置安全组

安全组是一个逻辑上的分组，为具有相同安全保护需求并相互信任的云服务器、云容器、云数据库等实例提供访问策略。安全组创建后，用户可以在安全组中定义各种访问规则，当实例加入该安全组后，即受到这些访问规则的保护。

安全组中包括入方向规则和出方向规则，您可以针对每条入方向规则指定来源、端口和协议，针对出方向规则指定目的地、端口和协议，用来控制安全组内实例入方向和出方向的网络流量。

数据库开放弹性公网EIP后，如果公网上的恶意人员获取到您的EIP DNS和数据库端口，那么便可尝试破解您的数据库并进行进一步破坏。因此，强烈建议您保护好EIP DNS、数据库端口、数据库账号和密码等信息，并通过DWS的安全组限定源IP，保障只允许可信源连接数据库。

安全组在创建DWS集群时配置，您也可以选择系统默认配置的安全组，默认只放开8000的端口，集群创建后，支持更改安全组规则，也支持切换其他安全组。

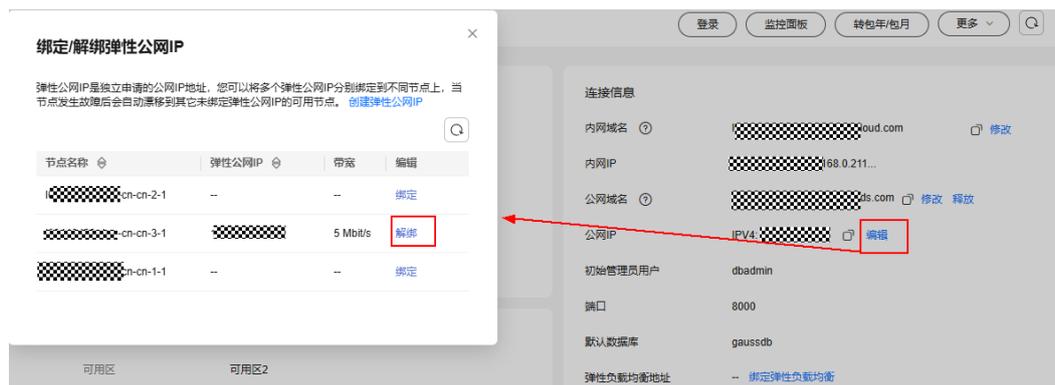
## 解绑弹性公网 IP 外部链接

弹性公网IP（Elastic IP，简称EIP）可以提供独立的公网IP资源，包括公网IP地址和公网出口带宽服务，方便外部访问，但同时增加了受到全网攻击的风险。使用弹性公网IP会受到外部的Dos/DDos攻击。

如果没有公网访问的特殊要求，建议将数据库作为内部组件，通过内部IP进行访问，建议解绑弹性公网IP。

### 解绑步骤：

- 步骤1** 登录DWS管理控制台，左侧选择“专属集群 > 集群列表”。
- 步骤2** 在右侧集群列表中，单击待解绑EIP的集群名称，进入集群详情页面。
- 步骤3** 单击“连接信息”区域的“编辑”，在弹出的页面，单击“解绑”。



### 说明

因高可用需求，如果集群绑定了弹性负载均衡（ELB），一般情况下弹性公网IP绑定到ELB端，只需要解绑ELB端对应的弹性公网IP即可，详情参考ELB用户手册。

### ----结束

## SSL 传输加密

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，DWS已经从CA认证中心申请到正式的服务器、客户端的证书和私钥（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](#)。

## 使用数据安全服务 DBSS（建议）

数据库安全服务（Database Security Service，DBSS）是一个智能的数据库安全服务，基于机器学习机制和大数据分析技术，提供数据库审计，SQL注入攻击检测，风险操作识别等功能，保障云上数据库的安全。

建议使用DBSS来提供扩展的数据安全能力，详情请参考[数据库安全服务](#)。

### 优势

- 助力企业满足等保合规要求。
  - 满足等保测评数据库审计需求。
  - 满足国内外安全法案合规需求，提供满足数据安全标准（例如Sarbanes-Oxley）的合规报告。
- 支持备份和恢复数据库审计日志，满足审计数据保存期限要求。
- 支持风险分布、会话统计、会话分布、SQL分布的实时监控能力。
- 提供风险行为和攻击行为实时告警能力，及时响应数据库攻击。
- 帮助您对内部违规和不正当操作进行定位追责，保障数据资产安全。

数据库安全审计采用数据库旁路部署方式，在不影响用户业务的前提下，可以对数据库进行灵活的审计。

- 基于数据库风险操作，监视数据库登录、操作类型（数据定义、数据操作和数据控制）和操作对象，有效对数据库进行审计。
- 从风险、会话、SQL注入等多个维度进行分析，帮助您及时了解数据库状况。
- 提供审计报告模板库，可以生成日报、周报或月报审计报告（可设置报表生成频率）。同时，支持发送报表生成的实时告警通知，帮助您及时获取审计报告。

## 9.4 数据保护技术

### 9.4.1 数据备份

DWS业务数据备份与恢复通过快照功能实现。快照是对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。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](#)。

## 9.4.2 透明加密

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。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。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。

DWS数据库级透明加密，每个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。

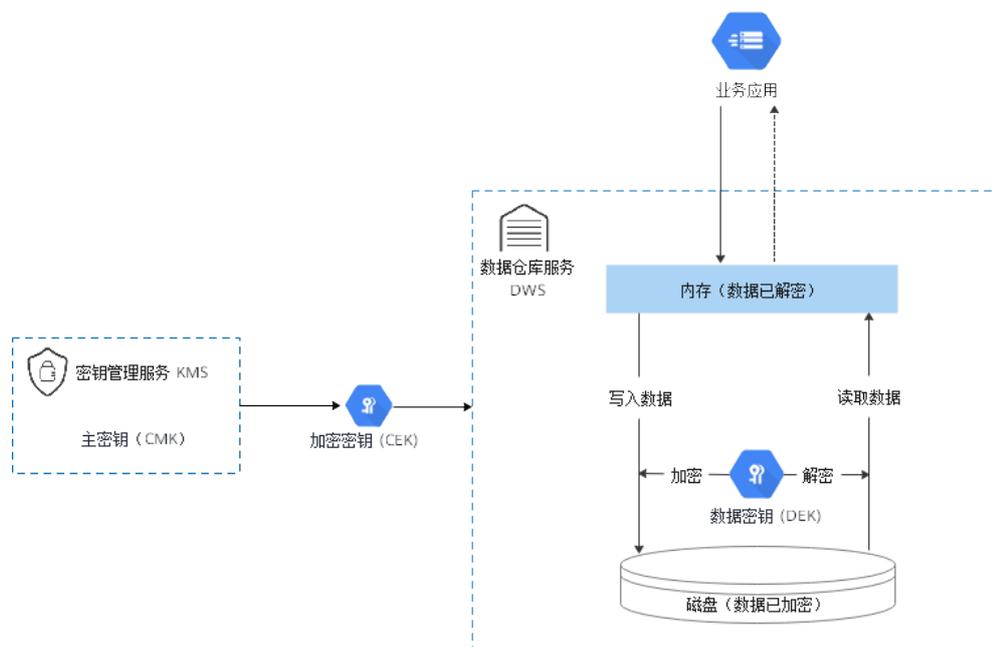
当前支持数据库级别的透明加密，在创建集群的时候进行加密配置。

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



### 9.4.3 SSL 传输加密

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，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](#)。

### 9.4.4 数据脱敏

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](#)。

## 9.4.5 使用函数进行加密

数据加密作为有效防止未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，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。

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。

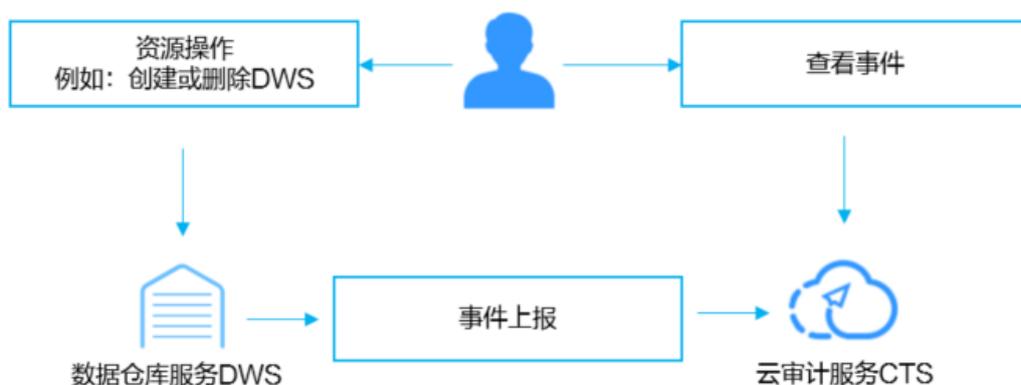
了解更多请参见[实现数据列的加解密](#)。

## 9.5 审计与日志

### 管理控制台的操作日志

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

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](#)。

## 9.6 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：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

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。

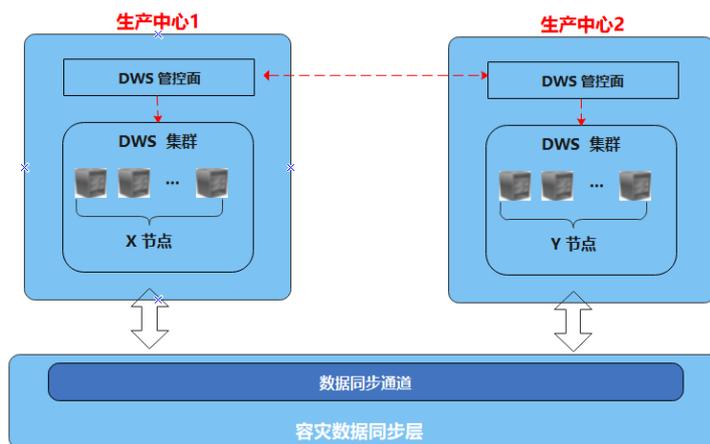
## Region 内容灾部署能力

DWS提供双集群Region内容灾能力，即在另一个可用分区（Region内）部署一个同构的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](#)。

图 9-4 容灾架构图



## 9.7 监控安全风险

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](#)。

## 9.8 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 9-5 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 9-6 资源中心



## 合规资质证书

华为云安全服务提供了网络安全专用产品安全检测证书、软件著作权等证书，供用户下载和参考。具体请查看[合规资质证书](#)。

图 9-7 网络安全专用产品安全检测证书&软件著作权证书



## 9.9 安全公告

## 9.9.1 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 DWS 权限管理

如果您需要对华为云上的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DWS，但是不允许删除集群的权限，控制他们对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](#)。

## 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。

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（如华北-北京四）对应的项目（cn-north-4）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，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[DWS自定义策略](#)。

如表10-1所示，包括了DWS的所有系统权限。

表 10-1 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"> <li>• 拥有VPC Administrator权限的DWS用户可以创建VPC或子网。</li> <li>• 拥有云监控Administrator权限的DWS用户，可以查看DWS集群的监控指标信息。</li> <li>• 使用此权限场景下，如果使用到创建委托场景，还需配置对应Security Administrator权限。</li> </ul>	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	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DWS访问VPC。授权成功后，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DWS的委托授权，详情请参见[委托DWS管理资源](#)。

表 10-2 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

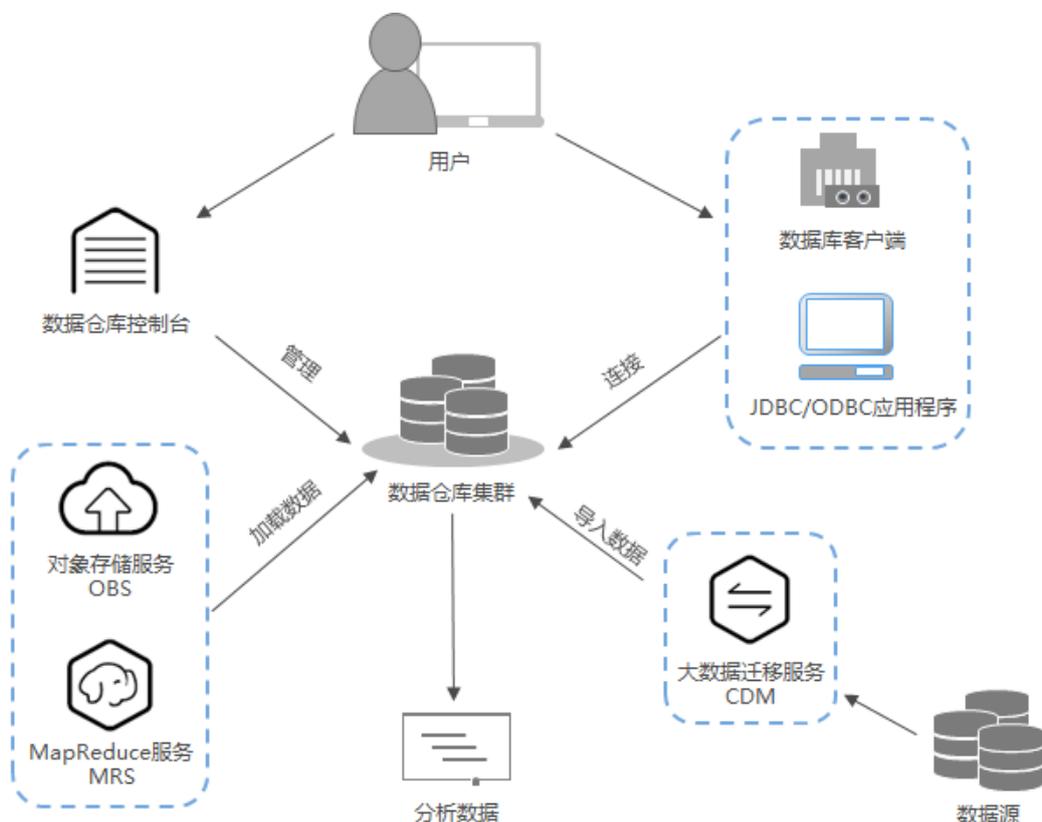
## 相关链接

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

# 11 如何访问 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

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DWS集群。

### 说明

集群部署时，如果单个节点故障，访问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

DWS支持如下方式访问集群数据库：

- DWS客户端方式  
用户使用DWS客户端访问集群数据库，请参见[使用gsq命令客户端连接集群](#)、[使用Data Studio图形界面客户端连接集群](#)和[使用SQL编辑器连接集群](#)。
- 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DWS，并进行数据表各类操作。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

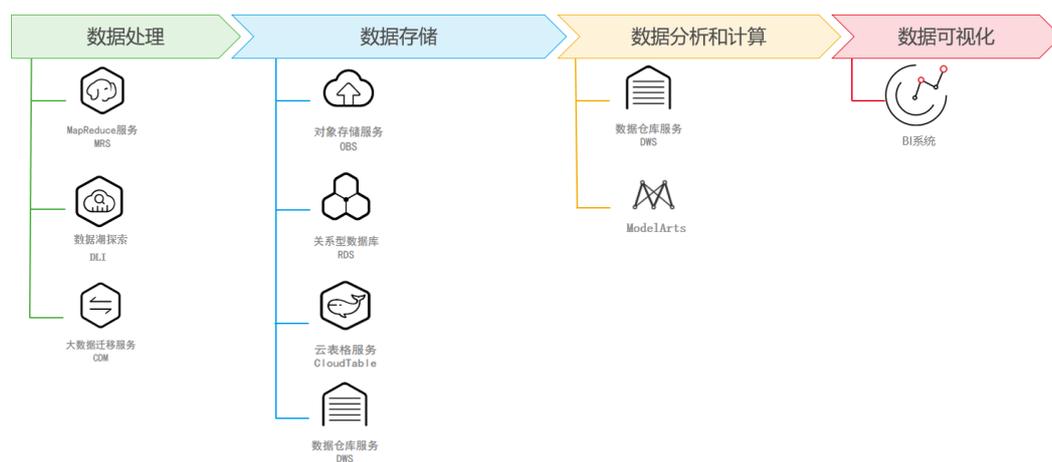
DWS暂不支持跨库访问，可以通过schema进行资源隔离。详情请参见[CREATE SCHEMA](#)。

## 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 使用限制

本文介绍使用DWS产品重点功能限制以及注意事项。

创建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。

## 规格与性能限制

表 12-1 规格与性能限制

资源类型	限制	说明
多AZ可用分区	3	<ul style="list-style-type: none"><li>只有当前选择的区域可用分区数量<math>\geq 3</math>时会展示“多AZ”选项，不满足该条件只能创建单AZ集群。</li><li>多AZ集群目前只支持同时选择3个AZ，主机节点均匀分配在选择的3个AZ中。</li><li>多AZ集群节点数量必须满足3的倍数。</li><li>多AZ集群只支持单节点DN数量<math>\leq 2</math>的规格。</li></ul>
存储空间	<ul style="list-style-type: none"><li>对于SSD云盘，如果您想提高存储空间扩容上限，请联系客服申请。</li><li>SSD本地盘存储空间范围与所选数据仓库规格有关。</li></ul>	<ul style="list-style-type: none"><li><b>SSD云盘</b>：高性价比的云硬盘，适合中等性能诉求的企业应用。</li><li><b>SSD本地盘</b>：SSD本地盘不支持磁盘扩容。</li></ul>
CN部署量	取值范围为3~集群节点数，最大值为20，默认值为3。	在大规模集群下，建议部署多个CN。

资源类型	限制	说明
最大连接数	<ul style="list-style-type: none"> <li>• CN连接数：100 ~ 262,143</li> <li>• DN连接数：100 ~ 262,143</li> </ul>	CN及DN概述详情请参见 <a href="#">集群逻辑架构</a> 。

## 配额限制

表 12-2 配额限制

资源类型	限制	说明
节点数	256	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。
标签	每个集群最多可以添加20个标签。	更多信息请参见 <a href="#">标签简介</a> 。
快照存储空间	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 <a href="#">OBS价格详情</a> 。 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（副本）x 节点数。
自动备份保留天数	可设置为1~31天，默认为7天	系统会在保留期结束时删除到期快照。

## 命名限制

表 12-3 命名限制

限制项	说明
集群名称	集群名称长度为4到64个字符，且不超过64个字节，1个中文字符为3个字节，其他字符为1个字节，不区分大小写，必须以字母或中文开头，可以包含中文、字母、数字、中划线或者下划线，不能包含其他的特殊字符。

限制项	说明
管理员用户	<ul style="list-style-type: none"> <li>只能由小写字母、数字或下划线组成。</li> <li>必须以小写字母或下划线开头。</li> <li>长度为6~64个字符。</li> <li>用户名不能为DWS数据库的关键字。DWS数据库的关键字，具体请参见《SQL语法参考》中<a href="#">关键字</a>章节。</li> </ul>
角色名称	以字母开头，可以包含字母、数字、下划线，长度不超过63个字符。
用户名称	以字母开头，可以包含字母、数字、下划线，长度不超过63个字符。
快照名称	快照名称长度为4~64个字符，必须以字母开头，不区分大小写，可以包含字母、数字、中划线或者下划线，不能包含其他的特殊字符。
快照策略名称	策略名称要求在4位到92位之间，必须以字母开头，不区分大小写，可以包含字母、数字、中划线或者下划线，不能包含其他特殊字符，并且名称唯一。
告警规则名称	规则名称长度为6到64个字符，且只能是首位非数字的中文、英文、数字、斜线组合。
容灾名称	容灾名称长度为4到64个字符，不区分大小写，必须以字母开头，可以包含字母、数字、中划线或者下划线，不能包含其他的特殊字符。

## 集群基础操作限制

表 12-4 集群基础操作限制

限制项	说明
绑定负载均衡	<ul style="list-style-type: none"> <li>绑定ELB时，需创建和DWS集群相同的区域、VPC以及企业项目的ELB进行绑定。</li> <li>DWS的ELB功能仅支持独享型ELB规格，不支持共享型ELB绑定操作。</li> <li>解绑ELB时，DWS侧会清除ELB中集群相关信息，但并不会删除用户ELB，请注意ELB本身的计费影响，防止不必要的成本支出。</li> </ul>
增删CN节点	<ul style="list-style-type: none"> <li>资源发放时默认CN数量为3，用户可以根据实际发放节点数量调整CN数量，范围为2~20。</li> <li>删除CN节点过程中，如果只有一个节点异常，只能删除这个异常的节点；如果有两个及以上节点异常，不能删除CN节点，需要先处理恢复异常的节点后，才能删除CN节点。</li> </ul>

限制项	说明
资源负载管理	<ul style="list-style-type: none"> <li>离线扩容过程中不支持资源管理功能，如开启了资源管理计划，请在进行离线扩容操作之前停止。</li> </ul>
逻辑集群管理	<ul style="list-style-type: none"> <li>在实际业务场景中，建议用户尽可能将同一个数据库的表创建到同一个逻辑集群中。</li> <li>物理集群切换逻辑集群时，如果原物理集群有数据则不支持切换。用户请确保切换时原物理集群数据为空。</li> </ul>
重启集群	<p>重启期间集群将无法提供服务。因此，在重启前，请确定集群中没有正在运行的任务，并且所有数据都已经保存。如果集群正在处理业务数据，如导入数据、查询数据、创建快照或恢复快照时，一旦重启集群，有可能导致文件损坏或重启失败。因此，建议停止所有集群任务后，再重启集群。</p>
启动/停止集群	<p>集群停止后实例关机，ECS场景下基础资源（vCPU、内存）不再保留，当再次启动云服务时，可能由于资源不足无法正常开机，请耐心等待，稍后再试。</p>

## 集群运维操作限制

表 12-5 集群运维操作限制

限制项	说明
扩容集群	<ul style="list-style-type: none"> <li>扩容期间集群会有闪断，请谨慎操作。</li> <li>正在扩容的集群将禁用重启集群、停止集群、启动集群、规格变更、增删CN、扩容集群、缩容集群、创建快照、重置数据库管理员密码等功能。</li> <li>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。</li> <li>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。</li> </ul>
缩容集群	<ul style="list-style-type: none"> <li>正在缩容的集群禁用重启集群、扩容集群、创建快照、节点管理、智能运维、资源管理、参数修改、安全设置、日志服务、重置数据库管理员密码和删除集群的功能。</li> </ul>
弹性规格变更	<ul style="list-style-type: none"> <li>弹性规格变更仅支持ECS+EVS云盘的存算一体及存算分离集群，ECS本地盘的集群不支持。</li> <li>弹性变更规格需要停止虚拟机，仅支持离线变更，变更时间大约需要5~10分钟。</li> </ul>

限制项	说明
经典规格变更	<ul style="list-style-type: none"> <li>● 如果当前集群是包年/包月计费模式，将不支持该功能。</li> <li>● 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。</li> <li>● 逻辑集群不支持经典变更规格。</li> </ul>
备份恢复集群	<ul style="list-style-type: none"> <li>● 对于有数据可靠性要求的用户，尤其是针对业务方无法通过上游重新导入进行数据恢复的场景，集群需要备份操作，避免人为或其他因素造成数据丢失后无法恢复。</li> <li>● 正在创建快照的集群，暂时无法提供完整服务，如无法执行重启、扩容、重置密码、修改集群配置信息等操作。</li> </ul>
升级集群	<ul style="list-style-type: none"> <li>● 8.1.3以前集群版本下发升级任务后，完成集群升级前不支持回滚或提交操作。</li> <li>● 集群安装热补丁后无法建立容灾。</li> </ul>
容灾管理	<ul style="list-style-type: none"> <li>● 存算分离以及多AZ集群不支持容灾功能。</li> <li>● 灾备集群在容灾任务停止或者异常但灾备集群正常的情况下，可以提供读服务，灾备切换成功后可以提供读写服务。</li> <li>● 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，以及生产、灾备集群均禁用恢复功能。</li> <li>● 不支持逻辑集群和资源池。</li> </ul>
日志管理	<ul style="list-style-type: none"> <li>● 非OBS场景下，此功能不可用。</li> <li>● 经典变更规格、增删CN等变更CN场景会导致数据丢失，建议在执行变更期间关闭审计日志转储。</li> </ul>

# 13 技术支持

数据仓库服务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。

## 维护策略声明

DWS集群资源属于数据仓库服务，DWS提供基于该资源的全托管云服务能力，用户拥有对集群的完全控制权。默认情况下，云服务仅对客户集群提供相关监控、告警能力，无权限对客户集群进行操作，集群日常运维由用户负责。如果在运行过程中遇到相关技术问题，可以联系技术支持团队获得帮助，该技术支持仅协助分析处理DWS云服务相关求助，不包含云服务以外的求助，例如用户基于DWS构建的应用系统等。

## 技术支持范围

### • 支持的服务

DWS云服务管理控制台提供的相关功能：

- 集群的创建、删除、扩容、缩容、规格调整、升级、补丁、备份恢复等操作
- 集群的监控告警管理
- IAM用户委托管理
- 对外API接口管理

### • 不支持的服务

- 不负责基于DWS集群之上的客户业务应用开发问题答疑和处理，例如业务设计、代码开发、作业性能调优和业务迁移等。如需支持，可咨询购买对应专家服务。
- 在DWS集群组件服务无明显异常或明确产品质量缺陷的情况下，不负责作业运行异常问题的排查分析。

## 数据仓库专家服务

如需获得更多支持，可咨询购买对应的数据仓库专家服务：<https://www.huaweicloud.com/service/dwssupport.html>。

# 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 DWS 技术指标

DWS不同版本的技术指标如下所示。

表 15-1 DWS 技术指标 8.1.3-9.1.0 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	9.1.0最大值
集群节点数	2048	2048	2048	2048	<ul style="list-style-type: none"><li>存算一体：2048</li><li>存算分离：采用多VW技术，最大支持256个VW，每个VW 1024个DN，建议VW不超过32个，每个VW不超过128个DN，所有VW的DN节点总数不超过2048。</li></ul>

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	9.1.0最大值
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	<ul style="list-style-type: none"> <li>存算一体: 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</li> <li>存算分离: 多VW技术, 可以实现并发数目的叠加, 随着VW个数增加, 支撑的并发数也可以相应的扩展, 集群总的并发数目受GTM/CCN排队影响, 建议不超过8192。</li> </ul>			
集群数据容量	20PB	20PB	20PB	20PB	<ul style="list-style-type: none"> <li>存算一体: 20PB</li> <li>存算分离: 数据存储在OBS上, 理论上容量可以无限扩展。</li> </ul>
单表大小	1PB	1PB	1PB	1PB	1PB
单行数据大小	1GB	1GB	1GB	1GB	1GB
单表列数 (除Hudi表外)	1600	1600	1600	1600	<ul style="list-style-type: none"> <li>行存: 1600</li> <li>列存: 1600</li> <li>HStore: 1600</li> </ul>
Hudi表列数	-	-	5000	5000	5000

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	9.1.0最大值
分区表的分区个数	32768	32768	32768	32768	最大32768, 建议不超过1000
集群内单点故障RTO	60s	60s	60s	60s	60s
集群内单点故障RPO	0	0	0	0	0
集群故障容灾切换RTO	60min	60min	60min	60min	60min
集群故障容灾切换RPO	60min	60min	60min	60min	60min

 说明

VW(Virtual Warehouse): 也称计算组, DWS存算分离支持把一个物理集群拆分成多个VW, 不同业务可以绑定不同VW, 实现业务负载隔离和并发数目扩展。

表 15-2 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 <sup>55</sup>	2 <sup>55</sup>	2 <sup>55</sup>
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

技术指标	8.0.x最大值	8.1.0最大值	8.1.1最大值
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 <sup>55</sup>	2 <sup>55</sup>	2 <sup>55</sup>

### 📖 说明

并发连接数的最大值基于数仓规格48 vCPU和64 vCPU。例如，存算一体云盘规格 dwsk.12xlarge ( 48 vCPU | 384GB | 24000GB SSD ) 或dwsx2.16xlarge.m7 ( 64 vCPU | 512GB | 32000GB SSD )。