

数据仓库服务

# 产品介绍

文档版本

08

发布日期

2020-07-31



华为技术有限公司



版权所有 © 华为技术有限公司 2020。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，华为公司对本文档内容不做任何明示或默示的声明或保证。

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 产品优势.....	3
3 应用场景.....	5
4 产品功能.....	9
5 基本概念.....	13
5.1 DWS 的管理概念.....	13
5.2 DWS 的数据库概念.....	13
6 与其他云服务的关系.....	15
7 DWS 权限管理.....	17
8 如何访问 DWS.....	21
9 计费说明.....	23
10 使用限制.....	26
11 配额说明.....	27
12 版本说明.....	28

# 1 什么是数据仓库服务

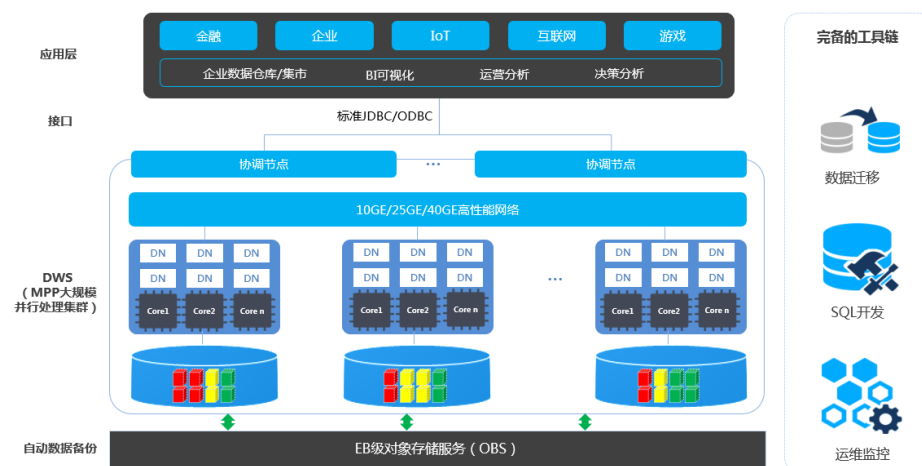
数据仓库服务（Data Warehouse Service，简称DWS）是一种基于公有云基础架构和平台的在线数据处理数据库，提供即开即用、可扩展且完全托管的分析型数据库服务。DWS是基于华为融合数据仓库GaussDB产品的云原生服务，兼容标准ANSI SQL 99和SQL 2003，同时兼容PostgreSQL/Oracle数据库生态，为各行业PB级海量大数据分析提供有竞争力的解决方案。

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。

## 产品架构

DWS基于Shared-nothing分布式架构，具备MPP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 4.0和ODBC 3.5连接DWS。

- **DWS（MPP大规模并行处理集群）**

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 Node）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、SQL开发工具Data Studio，并支持通过控制台对集群进行运维监控。

# 2 产品优势

DWS数据库内核使用华为自主研发的GaussDB数据库，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。

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。

您只需点击几下鼠标，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、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后台还通过算子多线程并行执行、向量化计算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。

- 数据加载快

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  
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

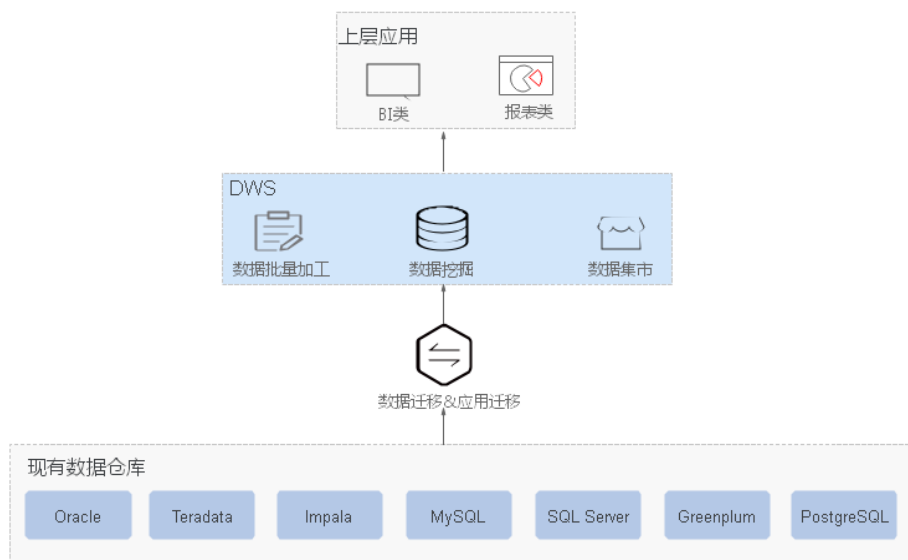
- 按需付费：DWS按实际使用量和使用时长计费。您需要支付的费率很低，只需为实际消耗的资源付费。
- 门槛低：您无需前期投入较多固定成本，可以从低规格的数据仓库实例起步，以后随时根据业务情况弹性伸缩所需资源，按需开支。

# 3 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 3-1 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

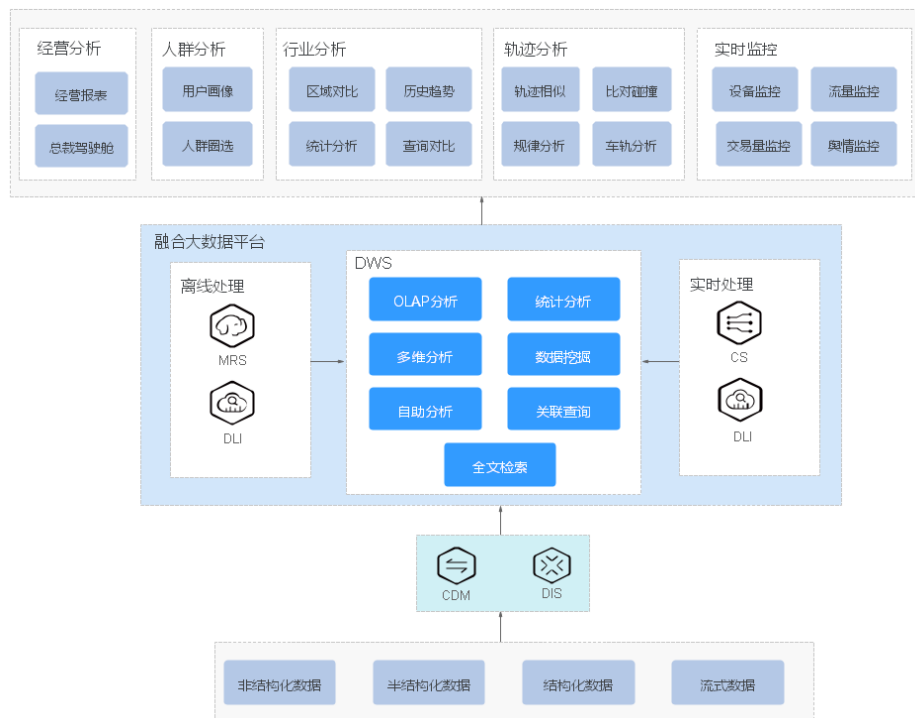


DWS支持数据加密，同时可与数据库安全服务对接，保证云上数据安全。同时DWS支持自动数据全量、增量备份，提升数据可靠性。

## 大数据融合分析

随着IT、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 3-2 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中获取信息。
- 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。

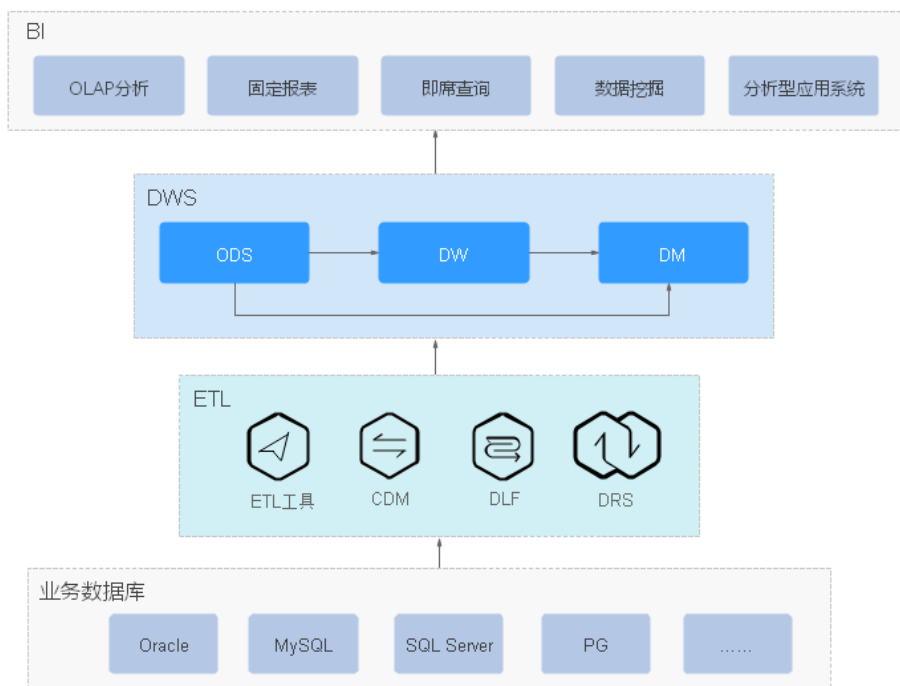
## 增强型 ETL 和实时 BI 分析

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。

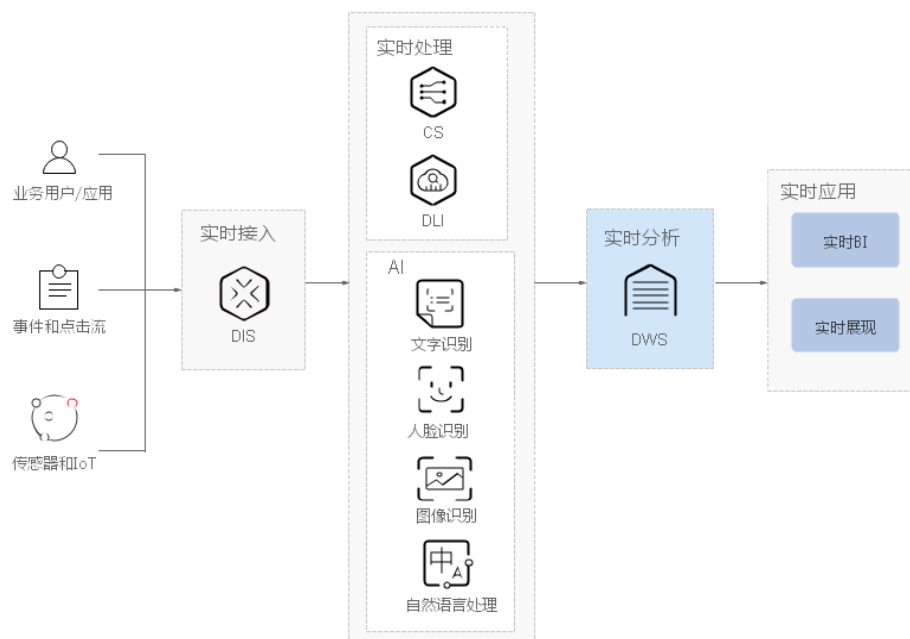
图 3-3 增强型 ETL+实时 BI 分析



## 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，DWS的快速入库和查询能力可支持实时数据分析。

图 3-4 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。

# 4 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。

## 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**  
用户可以根据业务需要选择节点规格和节点数量快速创建集群。您也可以先购买包年包月套餐，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，生成手工快照以便长期保留。  
当您从快照恢复集群时，系统将为您创建一个与原始集群规格相同、节点数也相同的新集群，并导入快照数据。  
对于不再需要的快照，可以选择删除快照，以释放存储空间。
- **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。
- **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为用户提供了以下方式管理集群和快照：

- **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](#)》。

## 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。
- 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：

- 支持SQL 92、SQL 2003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生态，与主流第三方数据库ETL，BI（business intelligence）工具厂商对接认证。

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](#)》。

## 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《[数据仓库服务数据库开发指南](#)》中的[导入方式](#)章节。

- 从OBS（对象存储服务）并行导入数据
- 使用GDS（General Data Service）从远端服务器导入数据
- 通过INSERT语句直接写入数据
- 使用COPY FROM STDIN导入数据
- 使用gsq命令导入数据
- 从MRS（MapReduce服务）导入数据到集群
- 从DIS（数据接入服务）导入流式数据到DWS
- 使用CDM（云数据迁移服务）迁移数据到DWS
- 使用DSC工具迁移SQL脚本

此外，DWS还支持主流第三方ETL工具进行数据导入。

## 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，且所有数据可手工备份至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，实现跨AZ容灾。

## 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  
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](#)。
- 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](#)》。

- 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 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。

# 5 基本概念

## 5.1 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集群至少包含3个节点。每个节点都支持存储与分析数据。

### 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。

## 5.2 DWS 的数据库概念

### 数据库

一个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。支持在公有云以及互联网环境上使用客户端进行连接。

## 数据库用户

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。默认的数据库用户为数据库管理员，在创建DWS集群时生成。

# 6 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。

## 与数据接入服务的关系

您可使用数据接入服务（Data Ingestion Service，简称DIS）将流式数据实时接入DWS。

## 与云监控的关系

数据仓库服务使用云监控（Cloud Eye）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。

# 7 DWS 权限管理

如果您需要对华为云上购买的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，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-北京1）对应的项目（cn-north-1）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，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7-1所示，包括了DWS的所有系统权限。

表 7-1 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> </ul>	系统角色	依赖 Tenant Guest和 Tenant 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7-2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。

表 7-2 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	√	√	√	√
获取单个集群详情	√	√	√	√
设置自动快照	√	x	√	x

操作	DWS FullAccess	DWS ReadOnlyAccess	DWS Administrator	DWS Database Access
设置安全参数/参数组	√	x	√	x
重启集群	√	x	√	x
调整大小	√	x	√	x
重置密码	√	x	√	x
应用参数模板	√	x	√	x
删除集群	√	x	√	x
设置可维护时间段	√	x	√	x
绑定EIP	√	x	√	x
解绑EIP	√	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
创建参数模板	√	x	√	x
删除参数模板	√	x	√	x
修改参数模板	√	x	√	x

## 相关链接

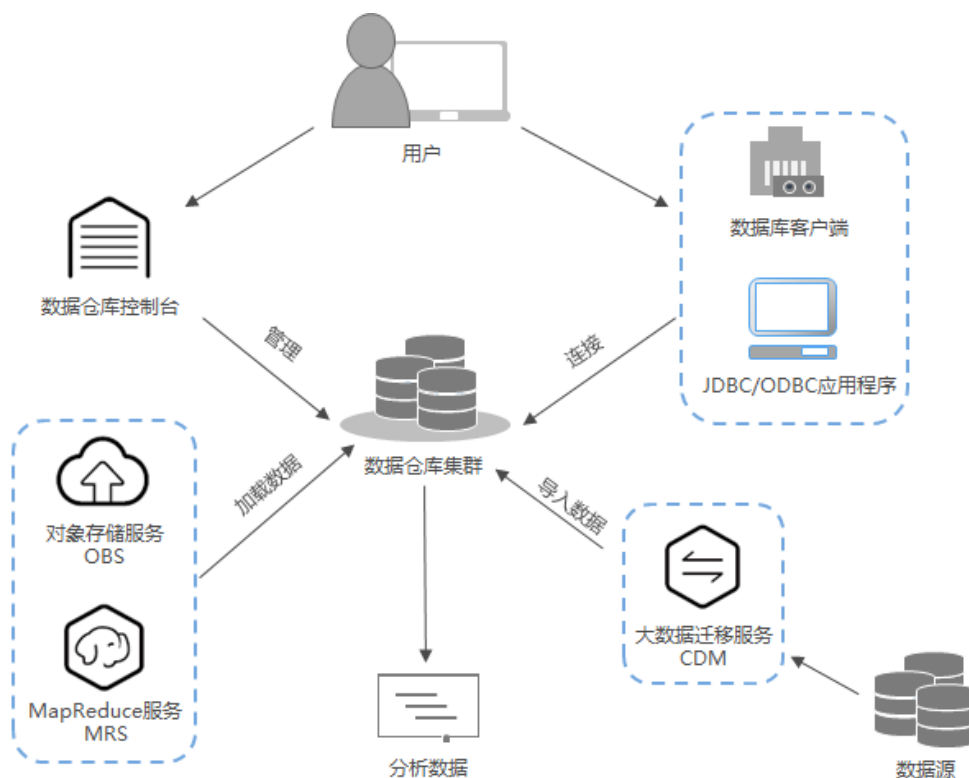
- [IAM产品介绍](#)
- [创建用户并授权使用DWS](#)

- [策略支持的授权项](#)

# 8 如何访问 DWS

数据仓库服务的基本使用流程如下图所示：

图 8-1 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。

## 访问集群数据库

DWS支持如下方式访问集群数据库：



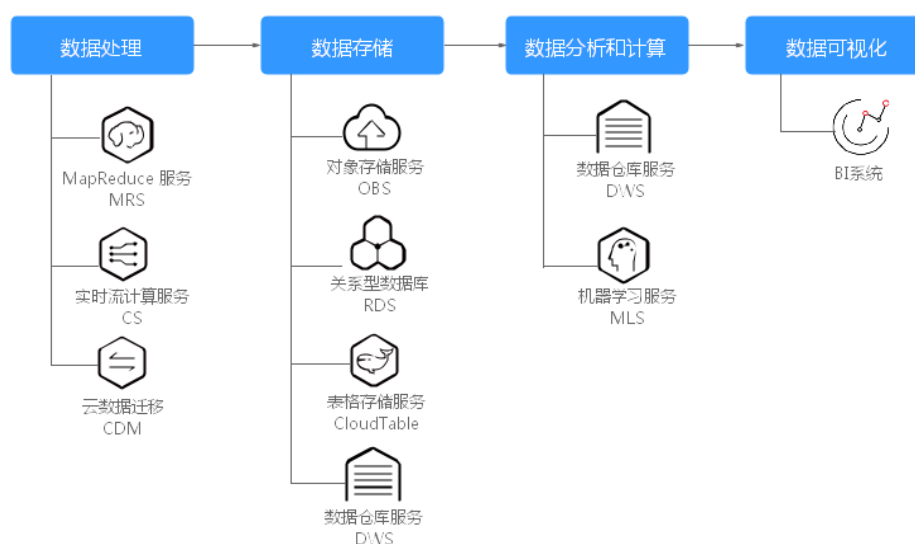
- DWS客户端方式  
用户使用DWS客户端访问集群数据库，请参见[使用gsql命令行客户端连接集群](#)和[使用Data Studio图形界面客户端连接集群](#)。
- 开源PostgreSQL客户端方式  
兼容的开源客户端有：
  - PostgreSQL的psql 9.2.4或更高版本  
有关使用psql的信息，请参见：<https://www.postgresql.org/>。
  - pgadmin  
有关使用pgadmin的信息，请参见：<https://www.pgadmin.org/>
  - dbeaver  
有关使用dbeaver的信息，请参见：<https://dbeaver.jkiss.org/download/>。
- 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](#)。

## 端到端数据分析流程

数据仓库服务与公有云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 8-2 端到端数据分析流程



# 9 计费说明

数据仓库服务（DWS）支持按需计费，同时您也可以选择更经济的包年包月计费模式。为了便于您便捷的下单购买，我们在控制台购买界面中已经为您计算好了整个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DWS价格详情，请参见[DWS价格详情](#)。您可以通过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DWS集群的参考价格。

表 9-1 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 <a href="#">DWS价格详情</a> 。针对节点规格提供了以下两种计费模式： <ul style="list-style-type: none"><li>• 按需付费（小时）</li><li>• 包年包月套餐</li></ul> 上述两种计费模式的详细说明请参见 <a href="#">计费模式</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 * 节点数。
Express加速集群数据扫描量	Express使用按需计费模式，即根据Express实际扫描的字节数收费，数据定义语言（DDL）语句（如 CREATE/ALTER/DROP TABLE）、管理分区的语句以及失败的查询无需付费。计费详情请参见 <a href="#">DWS价格详情</a> 。
公网IP及带宽（可选）	<ul style="list-style-type: none"><li>• 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<a href="#">弹性公网IP价格详情</a>。</li><li>• DWS集群在华为云内部网络产生的流量不计费。</li></ul>

计费项	计费说明
数据库加密密钥 (可选)	在创建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 <a href="#">数据加密服务价格详情</a> 。

## 计费模式

DWS提供两种计费模式供您选择：按需计费和包年包月套餐计费。两种方式供您灵活选择，使用越久越便宜。

- **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。该方式适合需要做前期运行试验和概念验证、短期使用的客户。
- **包年包月套餐**：用户通过承诺使用DWS更长时间并在购买时一次性支付费用，可获得相对于按需付费提供更大的折扣，节省35%至45%的费用。该方式适合在试验和概念验证结束后计划长期使用的客户。

包年包月套餐是一种计费理念，可以理解是一种优惠券，购买包年包月套餐和创建并使用集群是独立的过程。当您在使用集群时，计费系统会自动比较您已购买的有效期内的包年包月套餐节点数量和您使用的相同区域、相同节点规格的集群节点数量，包年包月节点数量内的节点不再收费（因为您前期已经支付过费用），超过的部分按需计费。包年包月套餐所购买的时长到期后，将自动转为按需付费方式。

为了进一步说明包年包月套餐节点和集群不存在绑定关系，举例如下：

- a. 购买包年包月套餐后不会自动创建数据仓库集群。如您尚未运行集群，需要前往[数据仓库服务](#)界面自行创建对应区域、节点类型和节点数量的集群。
- b. 如果您的集群目前是按需使用，在您购买对应区域、对应规格和数量的包年包月套餐节点后，计费系统在出账单时会自动识别并完成优惠减免，您不用再做任何关联操作。
- c. 如果您的集群需要扩容，且您在该区域下存在有效的包年包月套餐节点，则计费系统在出账单时会自动识别并完成优惠减免，您不用再做任何关联操作。如果您不存在可用的包年包月套餐节点，则新扩容的节点会按需计费。
- d. 删除了老集群后重新创建了新集群，包年包月套餐仍可以在相同区域、相同节点类型的新集群上使用。

### 须知

包年包月套餐注意事项：

- 包年包月套餐节点和具体的区域绑定，购买的包年包月套餐节点只能在绑定的区域使用，不能在非绑定区域使用。
- 包年包月套餐节点和具体的节点规格绑定，购买的包年包月套餐节点只能使用在对应规格上，不同规格的节点有不同的包年包月套餐价格。
- 包年包月套餐节点从购买之日起生效，到期自动结束，将自动转为按需付费方式。例如您1月1日购买了3节点套餐，购买时长1年，则该套餐在该年12月31日会自动结束，即使您该年内未使用数据仓库集群，该节点也不会延期，且无法退还费用。为了节省您的费用，建议您购买包年包月套餐节点后立即创建集群，或创建完集群后再购买包年包月套餐。

## 变更配置

在购买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[调整集群大小](#)。

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。

## 续费

目前DWS提供按需和包年包月购买模式，按需是每小时扣费，如果余额不足导致欠费，而包年包月是集群时长用完前需续费。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

- 按需购买的集群，没有到期。
- 包年包月集群到期后无法在DWS管理控制台进行该集群的操作，相关接口也无法调用，自动化监控或告警等运维也会停止。如果在保留期结束时您没有续费，集群将终止服务，系统中的数据也将被永久删除。

## 欠费

- 包年包月集群，没有欠费的概念。
- 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。

# 10 使用限制

---

- 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。
- 集群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。
- 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DWS集群。
- 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。

# 11 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11-1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 11-1 用户配额

资源类型	总配额
节点数	32

# 12 版本说明

本章节为您介绍DWS集群版本的更新记录。

## 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. 新增实时数仓集群类型，提供时序计算和时序压缩服务。
2. 新增工作负载管理功能，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。

## V1.7.2 版本

发布时间：2020年4月28日

配套内核版本：8.0.0

1. 支持细粒度资源权限控制。
2. 支持在页面解除集群只读。

## V1.7.1 版本

发布时间：2020年3月20日

1. 新增鲲鹏kc,km系列规格支持。
2. 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，也可以对已有集群修改自动快照策略。
  - b. 在集群列表的“更多”操作中，可以对该集群执行创建快照的操作。