

华为云
技术
私享会

华为云云上数据仓库解决方案



内容

- 云上数仓需求与挑战
- 场景化解决方案
- 华为公有云关键服务
- Demo

数据分析是必备能力

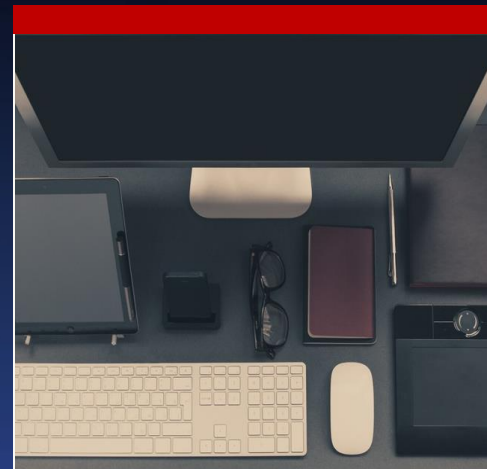


传统行业数字化转型
升级加快

银行+互联网=互联网金融
工业+互联网=工业4.0
商业+互联网=电子商务
教育+互联网=在线教育
医疗+互联网=远程医疗
政务+互联网=电子政务



移动互联网和IoT带
来又一次数据爆发



精细化运营和个性化
服务决定成败



新形势下的业务需求和挑战

对于大数据分析的需求

- 大数据分析可以指导企业经营决策
- 大数据分析可以有效减少运营成本
- 通过大数据分析可以实施精准营销，提供个性化服务



①

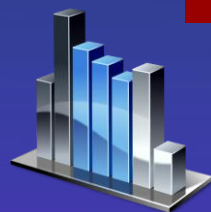
海量数据接入

- 移动互联网业务面对个人市场，用户规模以百万计，大型应用拥有上亿用户
- IoT应用中，每一个终端都可看做一个用户，终端数量庞大



分析平台跟随业务弹性扩展

- 业务往往发展迅速，在几个月甚至几周内，业务规模和数据量会以几何级数增长，需要资源快速响应
- 用户希望只关注业务，而不用关心底层平台的运维
- 需要自动化和智能的运维工具支撑



③

④



②

数据分析实时性要求高

- 企业需要快速构建或迁移数据分析应用
- 分析人员需要在海量数据中快速发掘数据价值
- 数据产生后即可进行分析

传统企业数仓上云

银行 工业 商业 教育 医疗 政务

- 数据量激增需快速扩容
- 系统稳定性、可靠性要求高
- 数据备份、容灾能力
- 需保证业务中立和不触碰数据

OLTP+OLAP

门户 网站 博客 论坛 搜索

- 海量数据，分析性能差
- 大数据分析、业务决策支持
- 打通业务数据库与数据仓库数据通道
- 分析系统快速上线

大数据处理分析

智慧营销 智慧生活 智慧城市

- 数据来源多，集成难
- 数据分析时延低
- 大数据分析、客户画像、精准营销
- 数据处理分析复杂，端到端方案

- 自建数仓费时、费钱、费力
- 业务激增，难以扩容

- 海量数据，无法分析
- 分析系统上线周期短

- 异构数据集成
- 数据实时在线分析

内容

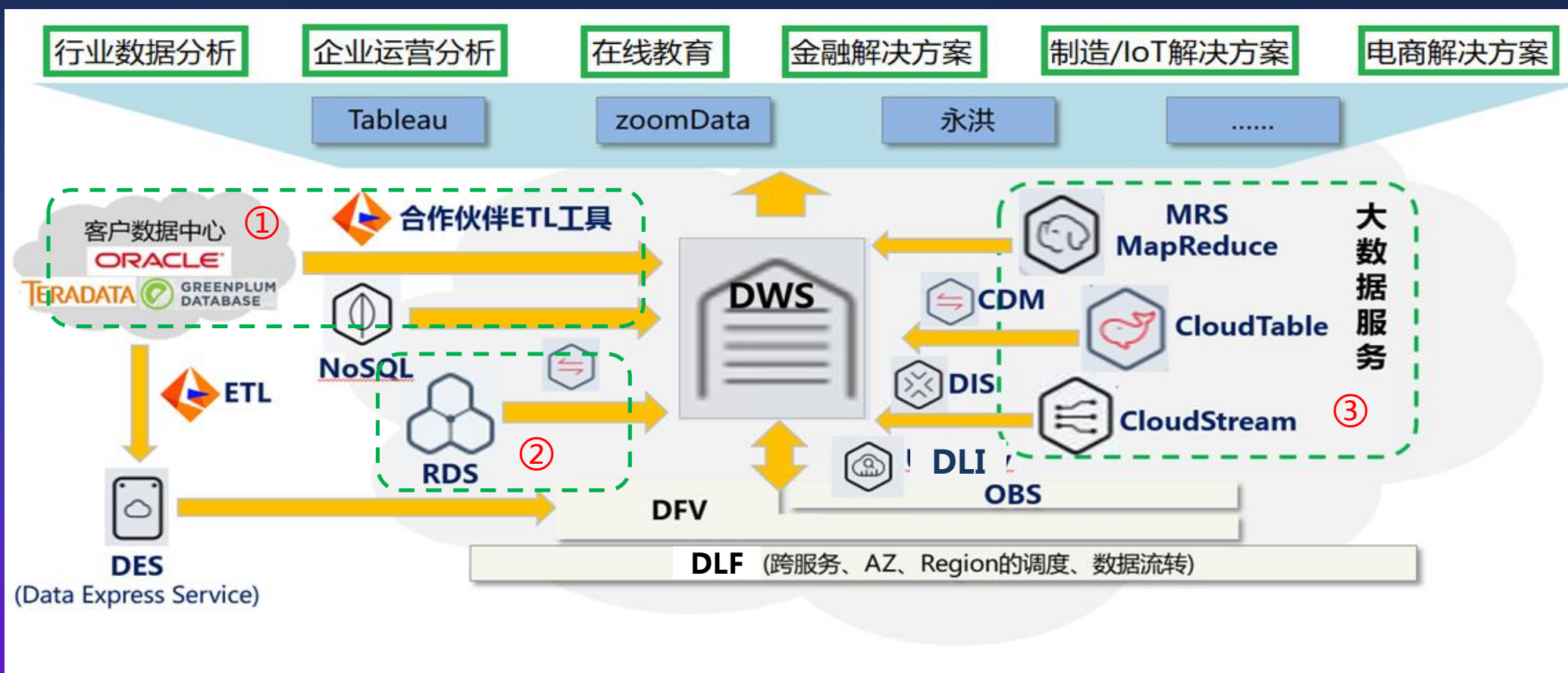
- 云上数仓需求与挑战
- 场景化解决方案
- 华为公有云关键服务
- Demo

云上数据仓库解决方案总览

①传统企业数仓上云

②OLTP+OLAP

③大数据处理分析



传统企业数仓上云

方案描述

提供自研ORACLE/Teradata迁移工具，以及多种数据迁移方案，低成本迁移Oracle、Teradata、Greenplum、MySQL等数仓上公有云。

方案优势

- 应用迁移：支持SQL2003标准，提供应用迁移工具，帮助用户平滑迁移已有业务应用
- 数据迁移：多种数据迁移方式，支持不同环境下不同量级数据迁移
- 开放兼容：全面兼容PG生态，支持对接Tableau、永洪BI、帆软、Superset等第三方BI工具
- 高可靠：全HA架构，无单点；自动数据备份OBS，可在不同可用区实现备份恢复
- 弹性按需：分钟级创建集群，付费方式灵活(按需、套餐)，低成本高效率，保障系统快速上线；资源弹性伸缩、灵活调配，快速响应业务变化
- 极致性能：性能优于友商同类MPP数据库 2-5倍

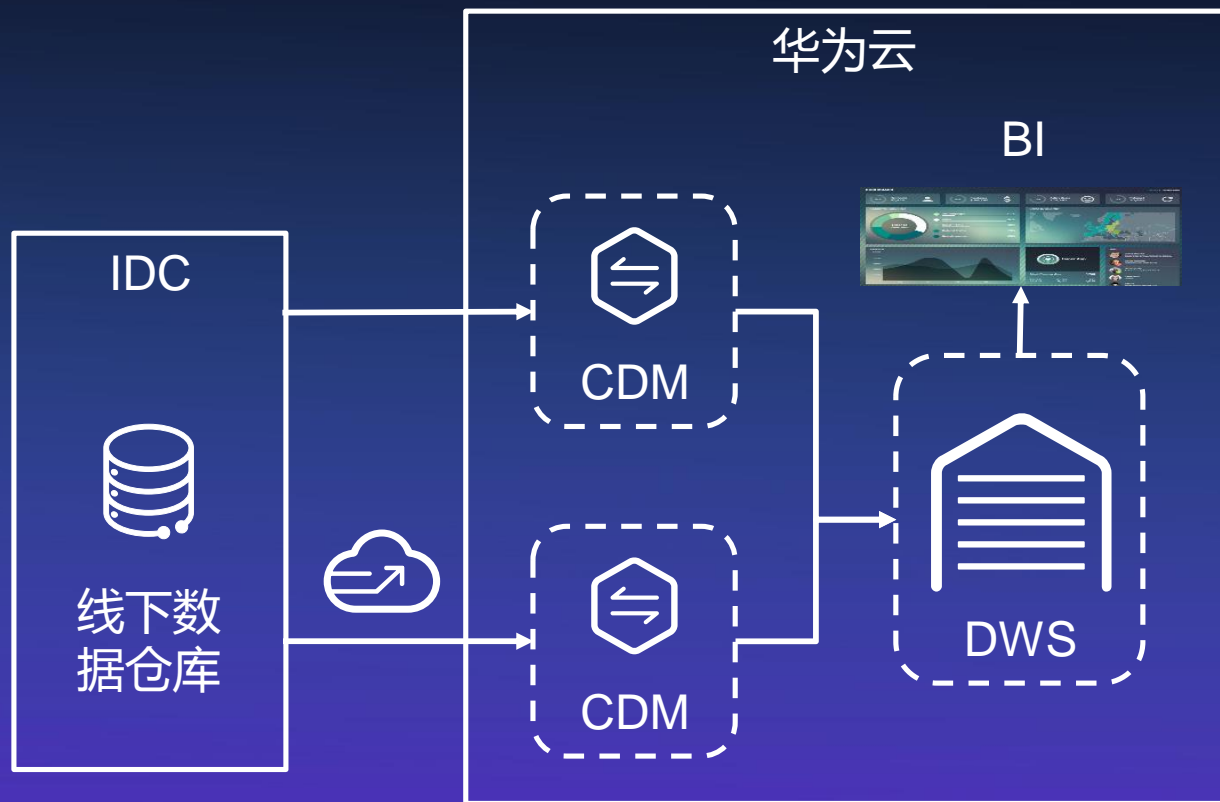
推荐搭配服务



CDM



DES



OLTP+OLAP

方案描述

打通业务数据库与数据仓库之间的数据通道，快速汇集分散的业务数据，构建云上数据分析系统。

方案优势

- 数据迁移：支持存量业务数据迁移和增量数据迁移
- 数据集成：支持异构数据集成，构建数据模型
- 数据处理：支持海量结构化数据加工处理
- 开放兼容：全面兼容PG生态，支持对接Tableau、永洪BI、帆软、Superset等第三方BI工具
- 弹性按需：分钟级创建集群，付费方式灵活(按需、套餐)，低成本高效率，保障系统快速上线；资源弹性伸缩、灵活调配，快速响应业务变化
- 安全可靠：支持基于OBS的数据自动备份和恢复

推荐搭配服务



RDS



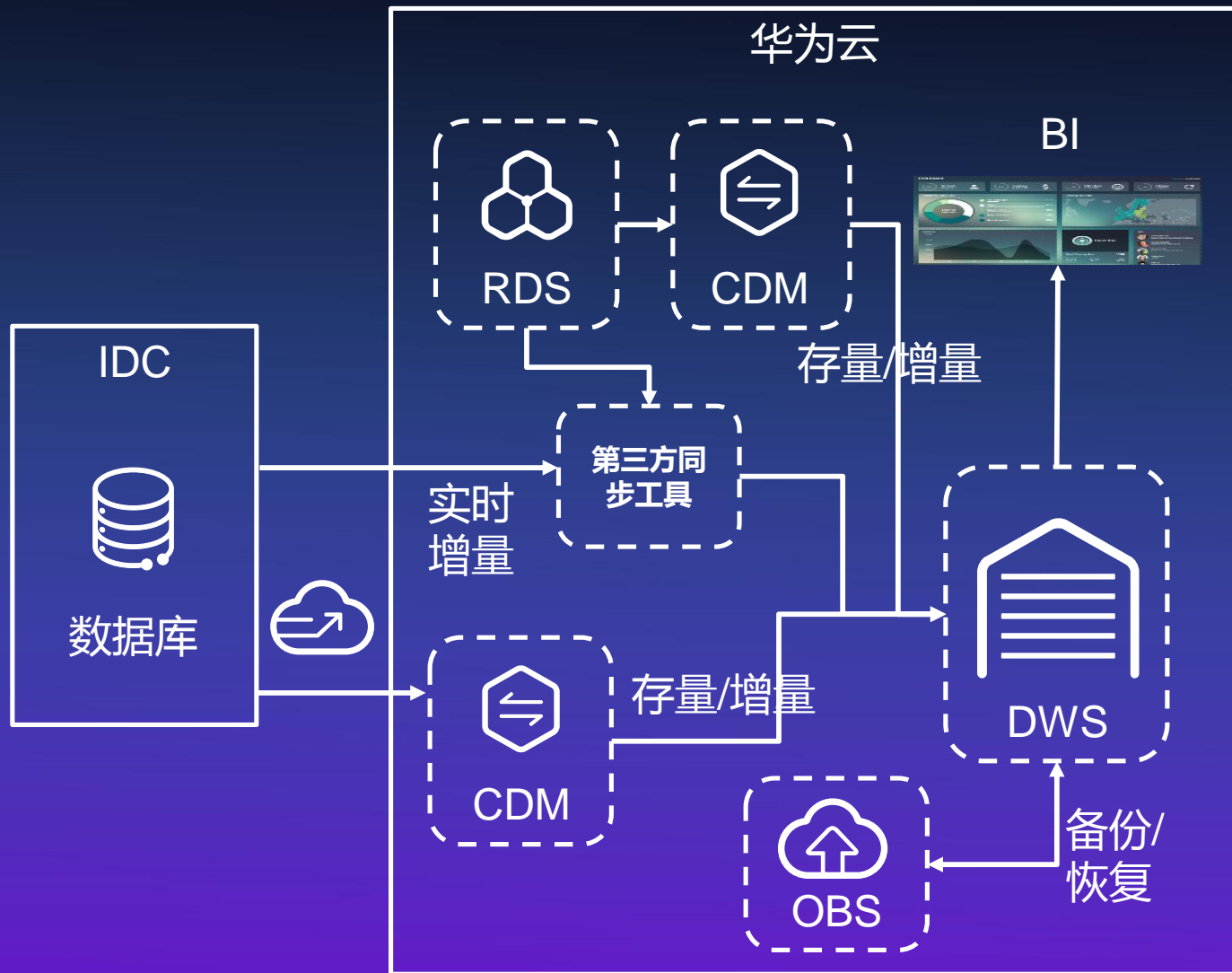
CDM



OBS



DES



大数据处理分析

方案描述

结合其他大数据技术，如实时接入、大数据处理等，实现端到端云上数据分析解决方案。

方案优势

- 端到端方案：数据接入、存储、处理、分析、展示全流程
- 多方式数据接入：异构数据批量与实时接入
- 丰富数据处理：结构化数据与非结构化加工处理
- 多种数据分析能力：标准SQL查询、全文检索、GIS数据分析、JSON数据分析等
- 开放兼容：全面兼容PG生态，支持对接Tableau、永洪BI、帆软、Superset等第三方BI工具
- 一站式数据开发：可视化数据开发、管理、运维

推荐搭配服务



DIS



CDM



DLF



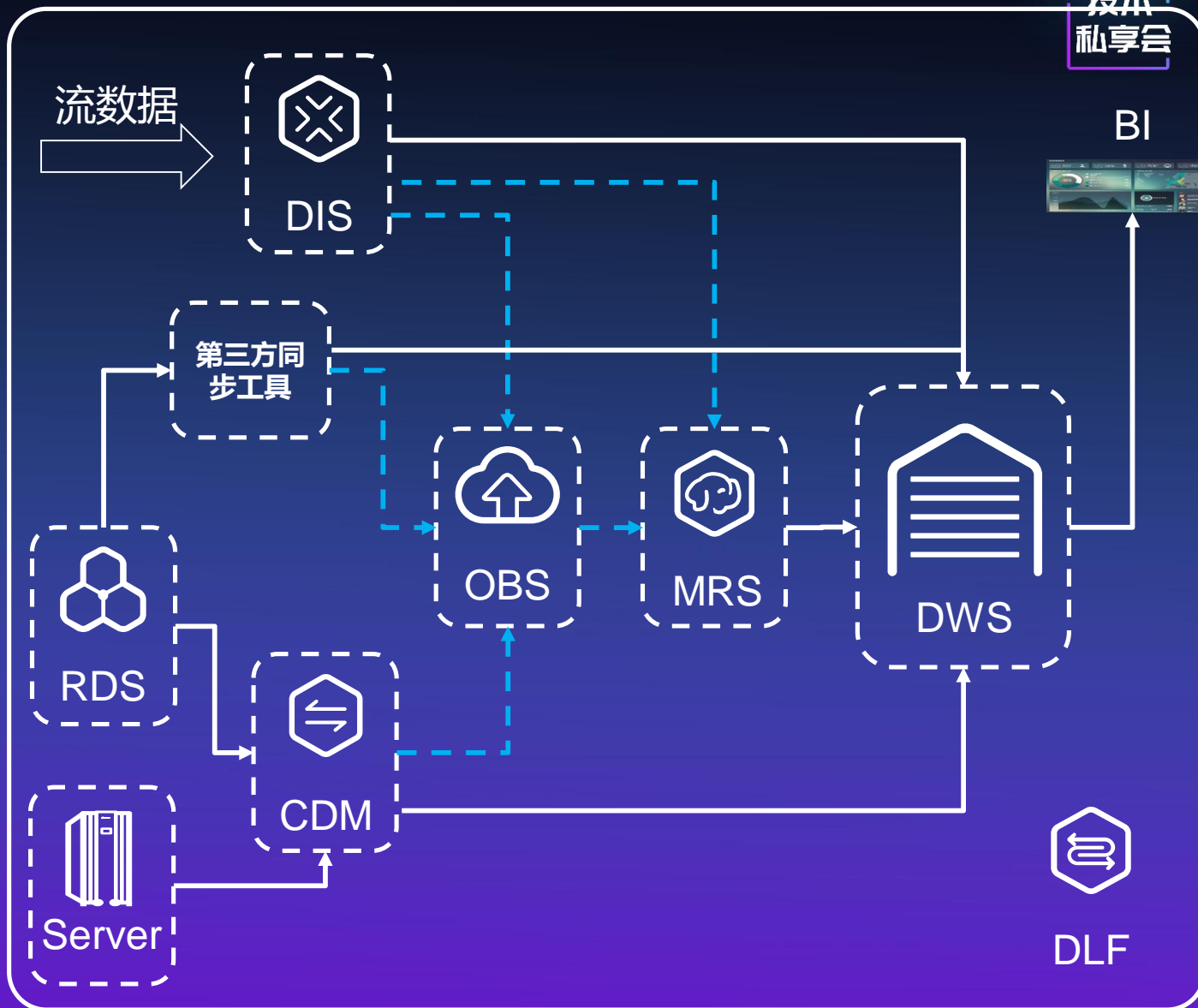
OBS



RDS



MRS



内容

- 云上数仓需求与挑战
- 场景化解决方案
- 华为公有云关键服务
- Demo

数据仓库服务DWS：PB级高性能企业数据仓库

数据仓库服务 (Data Warehouse Service, 简称DWS) 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PostgreSQL生态, 您可基于标准SQL, 结合商业智能(BI)工具, 经济高效地对海量数据进行挖掘和分析。



产品优势



可靠：**全节点**HA, 系统无单点故障



安全：获得欧洲**权威安全标准认证**, 租户隔离保证数据隐私安全



生态丰富：**兼容**PG生态, 支持多源数据**高效入库**



易用：**迁移工具**支持线下数仓轻松迁移到云上, SQL**智能优化**



易运维：**自动**弹性伸缩, 对接CES**监控**集群各项指标



低成本：按需租用, 成本只是传统数仓的**十分之一**

- 完整SQL支持
- 全并行架构
- 向量化引擎
- 拉链更新
- OBS数据源
- 备份恢复
- 高可用性

数据接入服务DIS：数据准实时入库到DWS

数据接入服务（Data Ingestion Service，简称DIS）提供了灵活数据采集、高效数据传输、实时数据分发能力，让您可以轻松将流式数据实时入库到DWS，降低数据分析端到端时延。



产品优势



高效：**百万**并发，**毫秒级**响应，单分片每日可传输数百GB数据



安全可靠：支持数据**加密传输**，租户**资源和操作隔离**，获得欧洲**PSA认证**



易用：端侧多语言**SDK/Agent**对接主流开源采集工具，云侧**自动**对接多种云服务，轻松完成数据采集、传输和分发

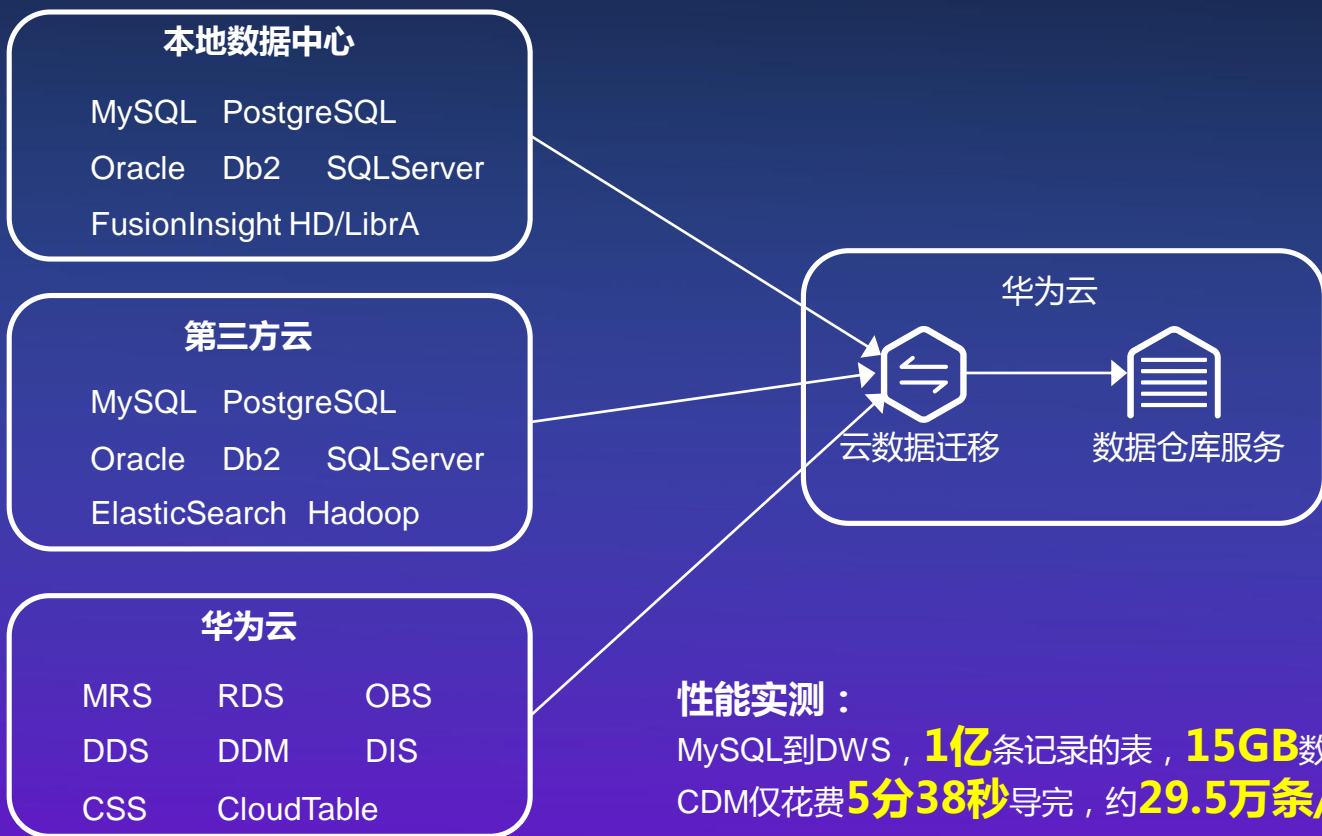


低成本：按需使用，相比基于Flume和Kafka自建系统成本**降低5倍**

云数据迁移服务CDM：异构数据轻松集成到DWS

云数据服务 (Cloud Data Migration , 简称CDM) 提供批量数据迁移服务 , 支持关系数据库 , NoSQL , 大数据 , 文件/对象等30多种数据源迁移到DWS , 无论这些数据源是在客户的本地数据中心还是在云上。

CDM：高效批量数据迁移工具



产品优势

异构数据源迁移：**30多种**自建数据源或云服务之间迁移数据

简单易用：**10分钟**轻松上手，全向导式

安全：租户间**VM强隔离**，确保用户数据安全

低成本高性能：单实例即可实现**每天TB级**数据迁移，与DES结合实现每天**百TB级**数据迁移

数据湖工厂DLF：云数据仓库建设的好搭档

数据湖工厂服务（Data Lake Factory，简称DLF）提供一站式的大数据协同开发平台，支持数据建模、数据集成、脚本开发、作业调度、运维监控等多项任务，帮助用户轻松构建云数据仓库。可实现DWS在线SQL编辑和调试，BI脚本开发和工作流调度，以及与其他服务如机器学习之间的混合作业编排。



数据湖工厂DLF · 一站式协同开发平台

数据管理

数据集成

脚本开发

作业开发

作业调度

运维监控



数据集成



数据存储



分析&计算



业务应用

数据存储&计算分析云服务

OBS

DWS

DLI

CDM

ES

MRS

RDS

.....

产品优势



一站式IDE平台：**一站式**建设云上数仓，无需切换多个工具



数据湖开发：管理**多种**大数据服务，可实现**跨服务**作业编排和调度



简单易用：在线SQL/Shell脚本编辑调试；预设**10多种**任务类型，拖拽式 workflow 编排



调度稳定高效：丰富的调度配置策略，**百万级别**的作业调度能力

内容

- 云上数仓需求与挑战
- 场景化解决方案
- 华为公有云关键服务
- Demo

TPC-DS

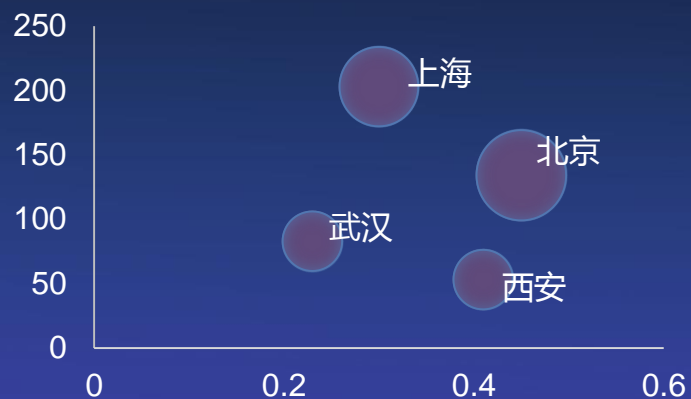
DWS兼容标准SQL1999和SQL2003，TPC-DS的99个SQL不需要修改一次执行通过。在企业中，存在大量的基于关系型数据库构建的数据仓库，数仓中的批处理脚本、应用程序大量使用了SQL。兼容标准SQL将使得传统数仓改造和新建数仓时，更容易与现有应用程序对接，且成本小，人才易获取。

TPC-DS兼容性测试

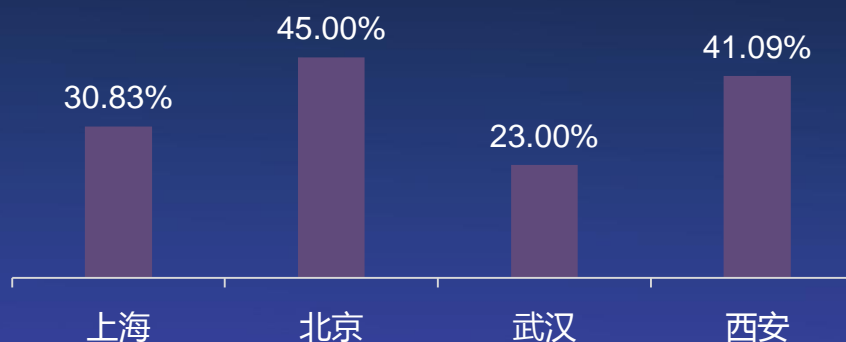
XX百货公司信息展示

$$\text{利润} = \text{客流量} * \text{转换率} * \text{客单价} * \text{利润率}$$

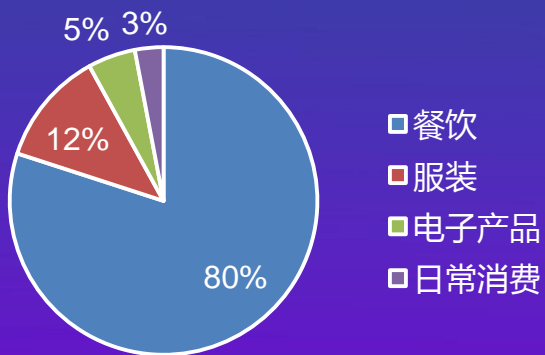
各公司利润分析



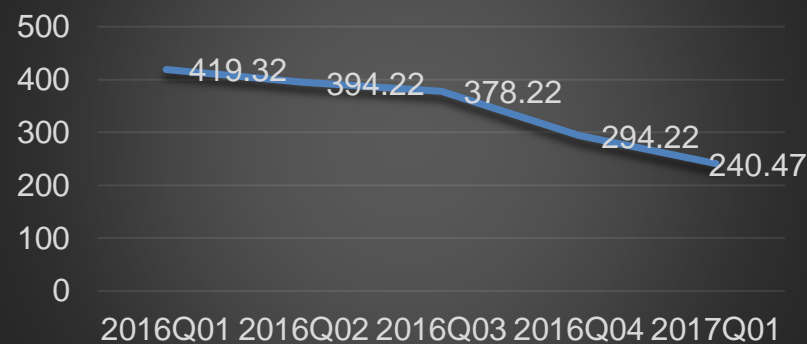
全国各公司提袋率



上海分公司2017Q1销售占比



上海分公司日常消费同比去年销售额分析



XX百货公司BI DEMO

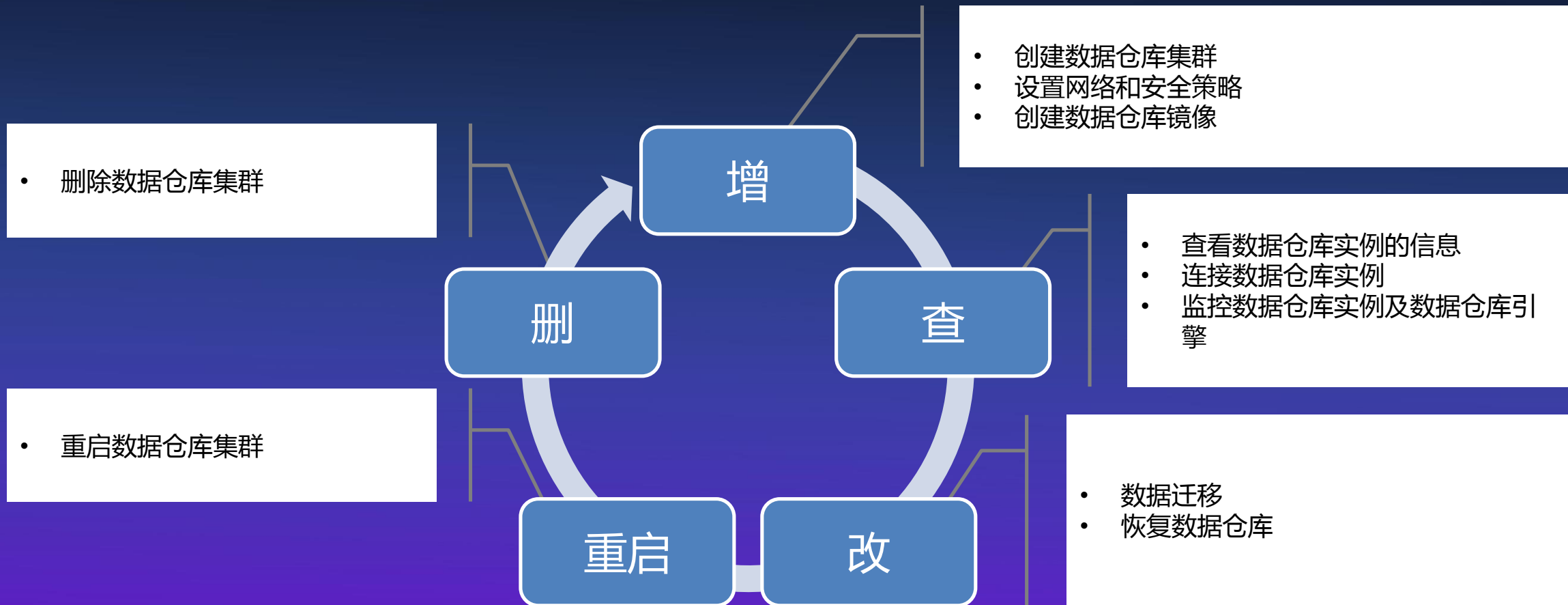
IOT场景DEMO

本示例，将加载8.9亿条交通卡口车辆通行数据到数据仓库单个数据库表中，并进行车辆精确查询和车辆模糊查询，展示DWS对于历史详单数据的高性能查询能力。

[交通卡口场景DEMO](#)

如何使用DWS——数据仓库实例生命周期管理

通过使用 DWS Console，您可以很容易地管理数据仓库实例或集群，因为大多数任务都可以由 DWS 来自动替您完成。



欢迎试用华为云数据仓库服务DWS



Q&A



THANK
YOU

华为云
技术
私享会

