

华为云
技术
私享会

华为云企业级数据仓库服务 介绍及应用实践



内容

- **DWS的由来**
- DWS优势
- 客户案例

趋势1：移动互联网时代，数据爆炸式增长，数据价值需要挖掘

移动互联网的普及，促使“人”直接参与数据的创造，数据蕴含更多的客户信息。



2000+PB

2018，全球网民1天数据流量



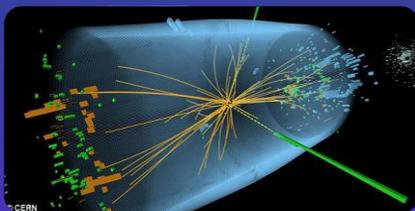
500+亿

2020，IoT连接数



63% GAGR

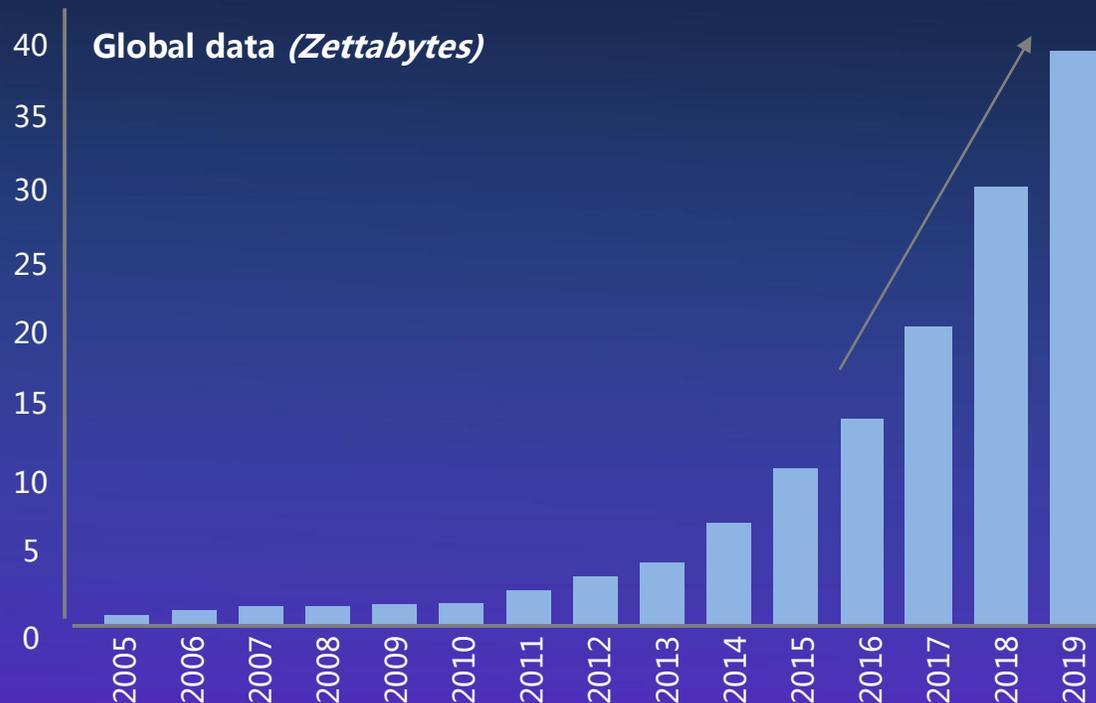
非结构化数据增长率



1PB/S

CERN：核爆产生数据的速度

原有数据分析平台的低扩展性、高扩展成本，成为了客户挖掘数据金矿的阻力。



大数据摩尔定理：全球数据总量大约每24个月会翻一番

趋势2：企业IT向通用化、云化转型

一体机业务迁移



软件与硬件解耦，避免硬件Lock-in

小机替换



X86服务器替代小机，更具扩展性

云化演进



云下向云上迁移

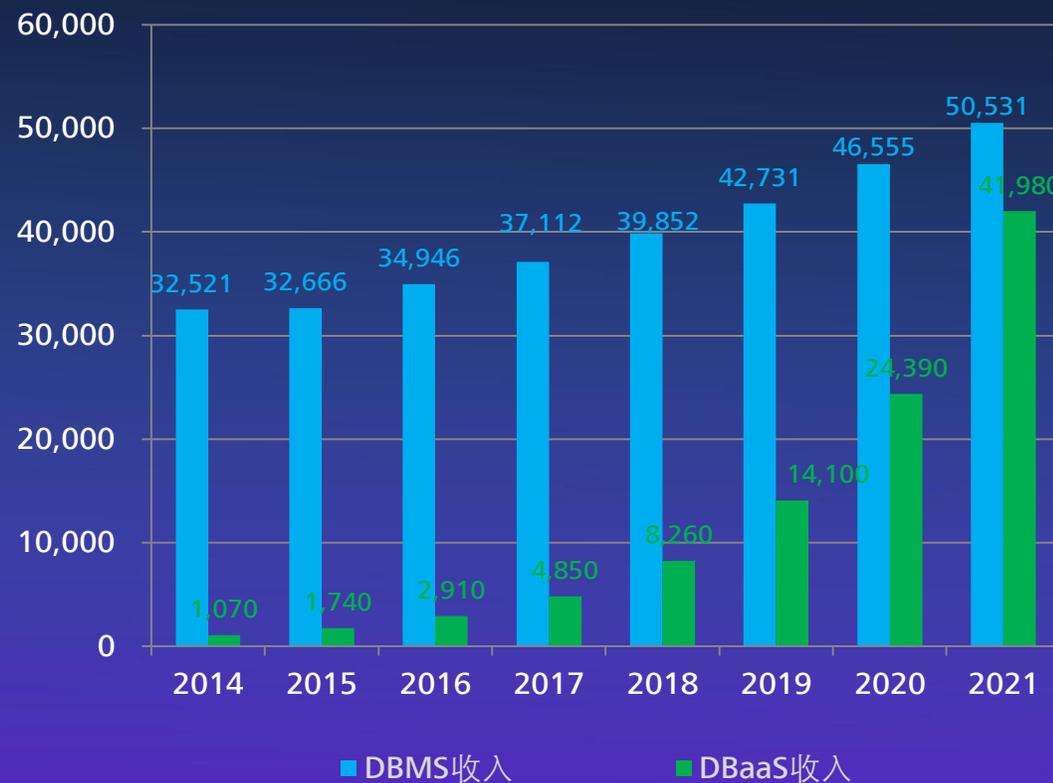
趋势3：企业数据持续加速上云

全球企业应用上云比例（按地域）



数据来源：451 Research, March 8, 2018

企业应用的加速上云，推动了数据加速上云



数据来源：Gartner (DBMS收入及增长情况)、
Research and Markets (DBaaS收入及增长情况)

DWS : 华为云高性能企业级数据仓库服务

核心能力：

(一) 标准SQL，平滑应用迁移

- 兼容标准ANSI SQL92、SQL2003
- 提供Oracle、Teradata语法迁移工具

(二) 开放架构，按需水平扩展

- Shared-Nothing架构+独创的大规模集群通讯技术

(三) 行列混存，极速查询分析

- 全并行计算，行列混存+向量化执行

(四) 全组件HA，无单节点故障

- 集群的协调节点、数据节点等逻辑组件全HA设计

(五) SQL on Anywhere

- 统一SQL访问OBS、HDFS等数据；跨集群协同分析



DWS：与经历6年市场检验的云下产品LibrA同内核

技术演进

行存储
HA机制
分布式框架
2012年~2013年

列存储
自适应压缩
向量化引擎
2014年

大集群通信技术
动态负载管理
图形化工具
2015年

LLVM
SQL on Hadoop
FullText Search
2016年

一键部署
应用平滑迁移
即开即用
2017年

SQL on OBS
自动备份
在线扩容
2018年

产品路标与市场突破

LibrA
原型开发与技术验证

2014年底，首次上市
销售，参与POC测试10+

在**10+**客户成功商用：
国内：G银行一期数仓
项目、贵州GA
海外：沃达丰子网

在**50+**客户成功商用：
国内：G银行二期，
辽宁移动
海外：菲律宾PLDT

华为云发布
DWS

竞争力增强

Gartner大数据分析平台魔力象限

2017年



2018年



内容

- DWS的由来
- **DWS优势**
- 客户案例

DWS vs RDS : 海量数据高性能分析

- RDS 业务数据库中的数据结构是为了**完成交易而设计的**，**不是为了而查询和分析的便利设计的**，对于大量数据的读（一般是复杂的只读类型查询）是支持不足的
- 当用户的**数据及查询的复杂性增加时**，或者，在您要防止报告和分析处理对 OLTP 工作负荷造成干扰时，DWS提供优秀的外扩选项。
- DWS 利用多节点的规模和资源并使用各种优化方法（**列存，向量引擎，分布式框架等**），为传统数据库对**大型数据集**的分析及报告工作负荷提供了数量级改善。

特性	OLTP (RDS)	OLAP (DWS)
面向用户	操作人员，普通业务人员	分析决策者
功能	日常操作处理（频繁增删改）	统计分析、BI
设计	面向应用	面向主题
数据	当前的，最新的细节的，二维的分立的	历史的，聚集的，多维的，集成的，统一的
存取	读/写数十条数据	读上百万条记录
应用	简单的读写	复杂查询
DB大小	0-100GB	TB-PB级别以上

DWS vs 自建数据仓库：简单、易用、免运维

自建数据仓库

- 购买并安装服务器、系统、数据仓库等软硬件
- 租用机房，费用高昂
- 招聘专业DBA,运维人员

V.S

数据仓库 on 云服务器

- 购买并安装数据仓库软件
- 租用云主机
- 招聘专业DBA,运维人员

V.S

华为云数据仓库服务

- 无需购买和安装任何软硬件
- 按需随时租用DDS
- 无需招聘DBA,运维人员

应用优化
数据仓库弹性伸缩
数据库高可用
数据备份与恢复
数据仓库软件版本升级或打补丁
数据仓库软件安装
操作系统版本升级或打补丁
操作系统安装
服务器部署及维护
机架堆叠
机房、电力、空调、网络等

应用优化
数据仓库弹性伸缩
数据仓库高可用
数据备份与恢复
数据仓库软件版本升级或打补丁
数据仓库软件安装
操作系统版本升级或打补丁
操作系统安装
服务器部署及维护
机架堆叠
机房、电力、空调、网络等

应用优化
数据仓库弹性伸缩
数据仓库高可用
数据备份与恢复
数据仓库软件版本升级或打补丁
数据仓库软件安装
操作系统版本升级或打补丁
操作系统安装
服务器部署及维护
机架堆叠
机房、电力、空调、网络等

■ 用户自行解决

■ 云服务提供商解决

华为云DWS优势



高安全：获得欧洲**权威安全标准认证**，租户隔离保证数据隐私安全



高可靠：**全节点**HA，系统无单点故障



高性能：**万亿**数据关联分析**秒级**响应



生态丰富：**兼容**PG生态，支持多源数据**高效**入库



易使用：**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

DWS内置安全能力

DWS当前已在各个维度提供基本安全能力，如口令登录认证、三权分立模型、SSL安全连接、审计日志追溯、安全运维管理和应用通道分离等。

No.	威胁分类	主要威胁	关键相应对策
1	网络访问	-跨网络访问：篡改/泄露	-SSL安全连接 -最大连接数控制
2	安全管理	-非法访问 -日志缺失 -拒绝服务攻击 -越权操作和权限滥用	-RFC5802口令登录认证 -操作审计日志记录 -流量、连接数大小限制 -三权分立、表级权限、列级权限、ACL控制 -防暴力破解机制
3	数据导入导出	-非法访问 -拒绝服务攻击 -仿冒与信息泄露	-流量、连接数大小限制 -外表权限控制
4	数据处理	-敏感数据泄露 -隐私数据泄露	-日志、文件敏感信息mask处理 -隐私数据加密
5	数据存储/备份	-敏感数据泄露 -数据破坏/篡改 -拒绝服务	-列加密、表加密 -数据清除(目录清理) -空间大小处理 -增量备份
6	管理运维	-管理员的恶意行为 -越权操作和权限滥用 -root远程访问	-日志与审计 -系统权限最小化 -脚本管理与权限控制 -主目录锁定

DBSS：数据库防火墙为DWS保驾护航



数据泄露保护 DLP

- 敏感数据发现
- 动态数据脱敏



数据库防火墙 DBF

- SQL注入防御
- 责权分离
- 漏洞检测防御
- 合规检查
- 脱库检测



数据库审计防护 DAP

- 数据库活动监控
- 合规报表
- 日志分析

华为云取得多项国际安全合规认证



可信云
服务认证



CSA STAR



C-STAR



ISO 27001:2013



ISO 20000



ISO 9000



CC EAL3+证书



PCI-DSS



ISCCC 信息
安全资质一级



中国等保



ITSS



云测评证书



中国网络
安全审查



金牌运维



中国国家信息安全
测评信息技术产品
安全测评证书

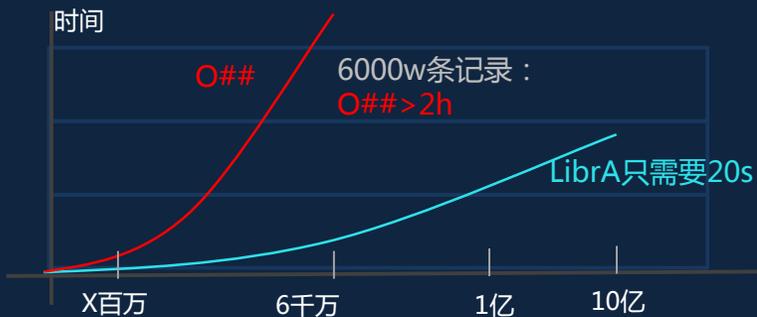
DWS性能：全并行计算引擎，实现万亿数据关联分析秒级响应

传统分析型数据库，无法应对海量数据复杂查询场景下快速响应的诉求

以某市交通套牌车分析为例，6000w条记录，全市1000智能卡口，查询分析计算量大： $6000w \times 6000w \times 1000 \times 1000$



在数据量达到千万级别时，传统数据库的查询响应时间极速增大



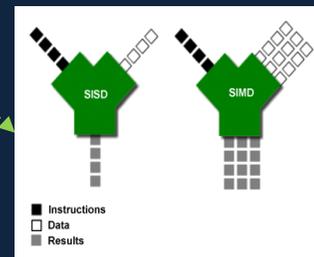
SMP、LLVM、向量化执行等技术带来100倍分析性能的提升

应用程序APP

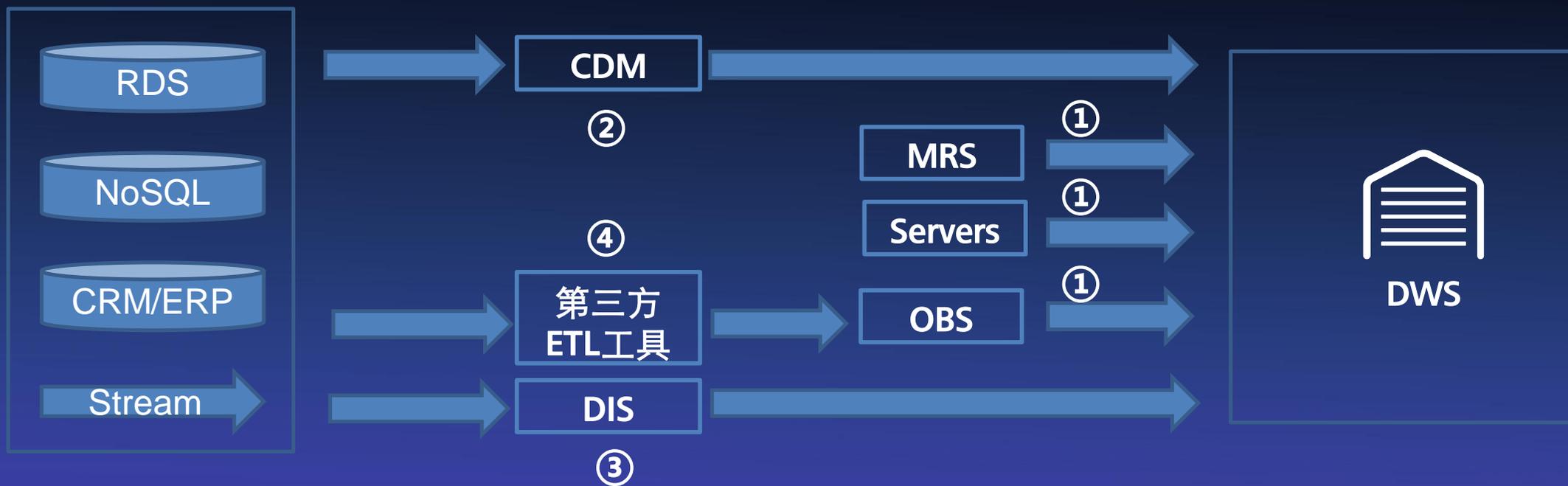


行列混存：适应更多场景，对简单查询、复杂查询、Ad-hoc查询有更好的适应性和更高的性能

向量化执行：优化迭代执行模式为一次处理一批元组



DWS易使用——支持多种数据入库方式



入库方式	数据源	说明
①并行拉取	OBS、MRS (HDFS)、远端服务器	性能好，横向扩展
②CDM	数据库、NoSQL、文件系统、大数据平台	数据源丰富，操作简单
③DIS	流数据	准实时入库
④第三方ETL	数据库、NoSQL、文件系统、大数据平台	通过OBS中转，数据源丰富，数据转换能力强

DWS易使用——丰富的图形化工具



工具分类	工具名称	功能
驱动	JDBC/ODBC	自研数据库连接驱动程序包
客户端	GSQL	自研Linux环境数据库客户端，命令行形式
	Data Studio	自研Windows环境下数据库客户端
	PGAdmin	Web数据库客户端，集成开源
数据开发	DLF	自研一站式数据仓库开发、数据管理、任务调度IDE，Web形式
可视化	SuperSet	BI可视化工具，集成开源
数据加载	GDS	自研数据并行加载工具
应用迁移	Migration Tool	自研数据库应用迁移工具，支持Oracle、Teradata应用迁移

DWS第三方生态工具

厂商	类别
Informatica	ETL
普元ETL	ETL
Kettle	ETL
Attunity	ETL
DataStage	ETL
合众RDS	数据复制
Tableau	BI & Analytics
帆软	BI & Analytics
SuperSet	BI & Analytics
永洪BI	BI & Analytics
Veritas NetBackup	数据存储
IBM Tivoli	监控

持续增加中.....

<https://www.tableau.com/partners/technology>



Huawei Technologies

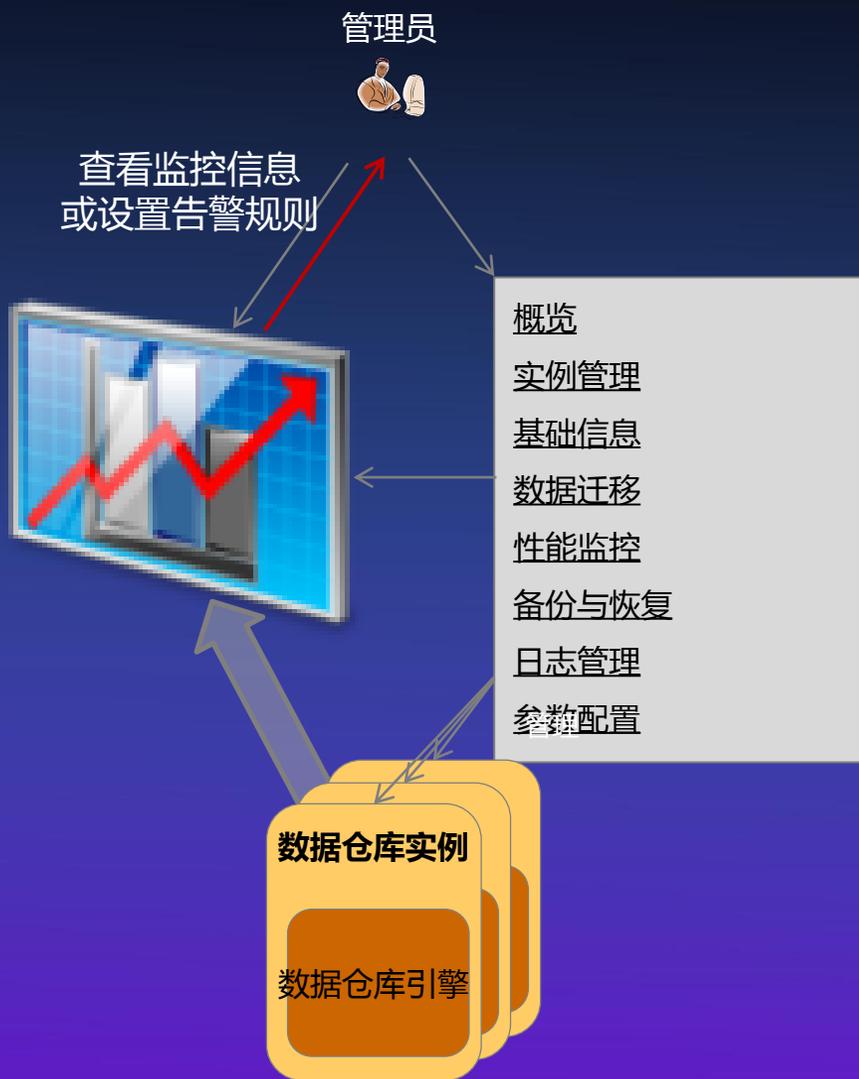
FusionInsight provides a comprehensive Big Data software platform for batch and real-time analytics using open-source Hadoop and Spark technologies. Tableau users can connect to FusionInsight with ODBC Driver.

[less](#)

CHINA

[LEARN MORE](#)

DWS易运维——运维监控能力



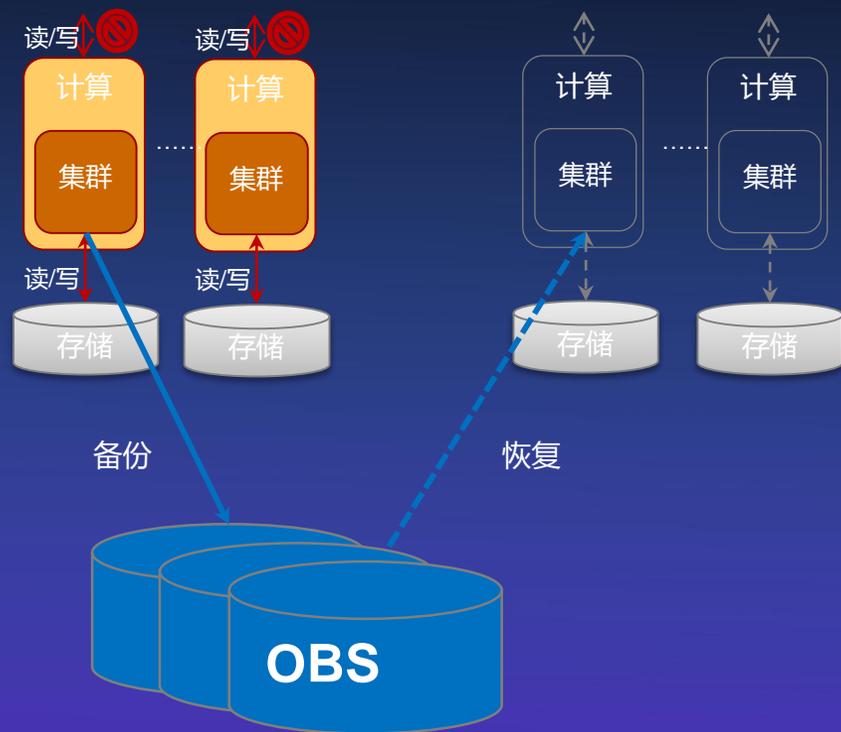
好处

- ✓ 您可以跟踪DWS集群数据仓库实例及数据仓库引擎的性能和健康状况。

做法

- DWS 整合了免费的云监控服务（CES）来呈现数据仓库实例及数据仓库引擎的性能和健康状况。
- DWS监控**系统级**的关键指标，包括磁盘使用率、CPU使用率、内存使用率、IOPS、网络输入吞吐量、网络输出吞吐量等。
- DWS支持监控数据仓库实例及数据仓库引擎的关键**内核级指标**，包共享内存命中率、内排序比率、文件读取次数、文件写入次数、文件读取频率、文件写入频率、数据量大小、活跃SQL数、会话数等。

DWS易运维——自动快照，让数据后顾之忧



好处

- ✓ 如果数据仓库出现错误或损坏，您可以通过备份来恢复它们，从而确保数据可靠性。

做法

- 对于 DWS 集群实例，DWS 使用 OBS 来备份数据并且确保数据一致性。
- 支持自动增量备份

内容

- DWS的由来
- DWS优势
- 客户案例

助力广联达打造云上数据仓库，提升业务能力



客户痛点：

- ◆【存不下】数据增长迅速，已有存量数据5TB，计划存储3年约20TB数据
- ◆【查不出】查询任务在MySQL耗时长，部分跑不出结果，MongoDB数据无法做复杂关联分析
- ◆【管理难】数据存储在不同库，不易进行统一管理和维护

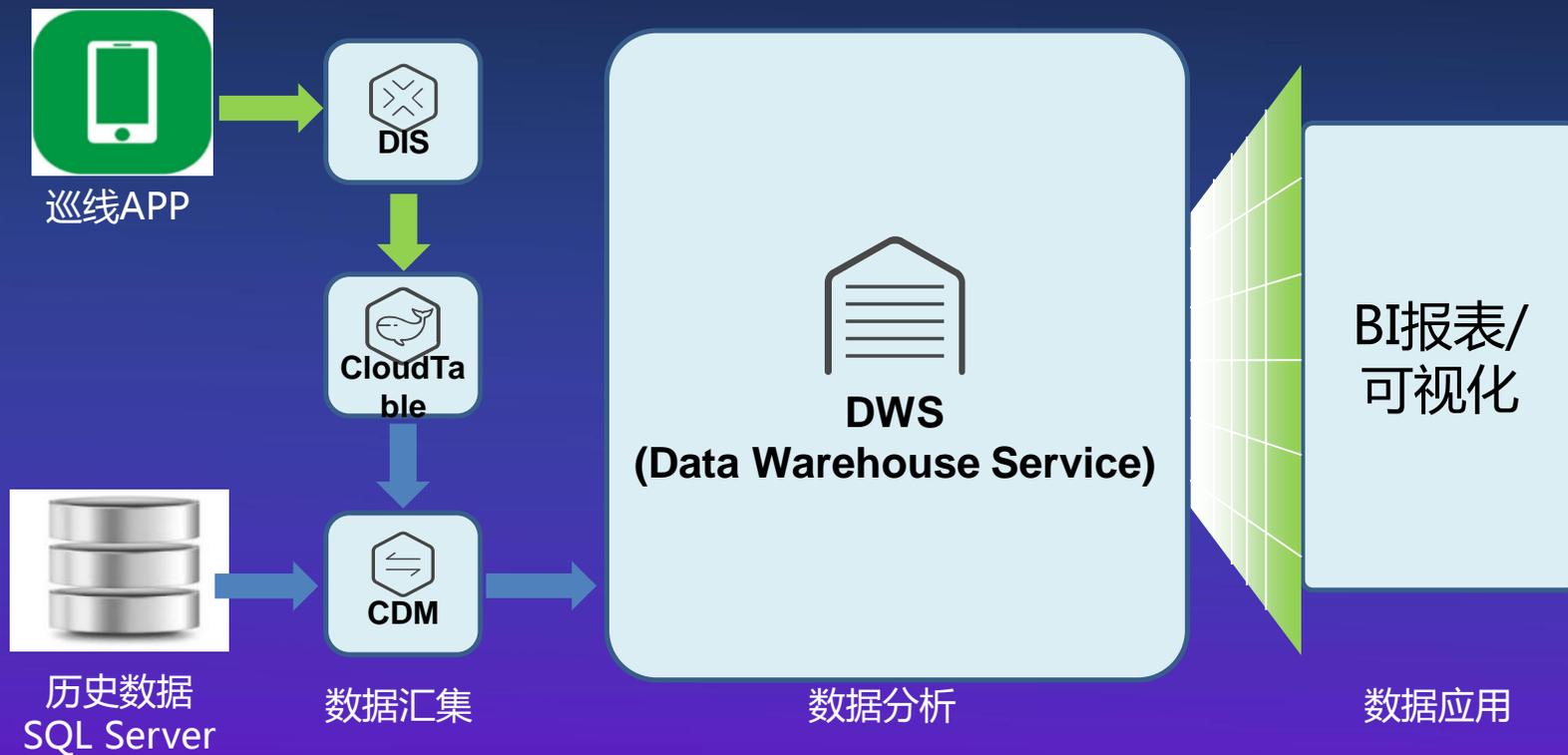
客户价值：

- ◆数据统一分析、管理效果显著，**节约50%**数据处理时间
- ◆降低开发、运维投入，**成本下降50%**
- ◆解决原有系统数据查询慢或查询不出结果的问题

助力某公共事业集团实现大数据业务创新

客户痛点：

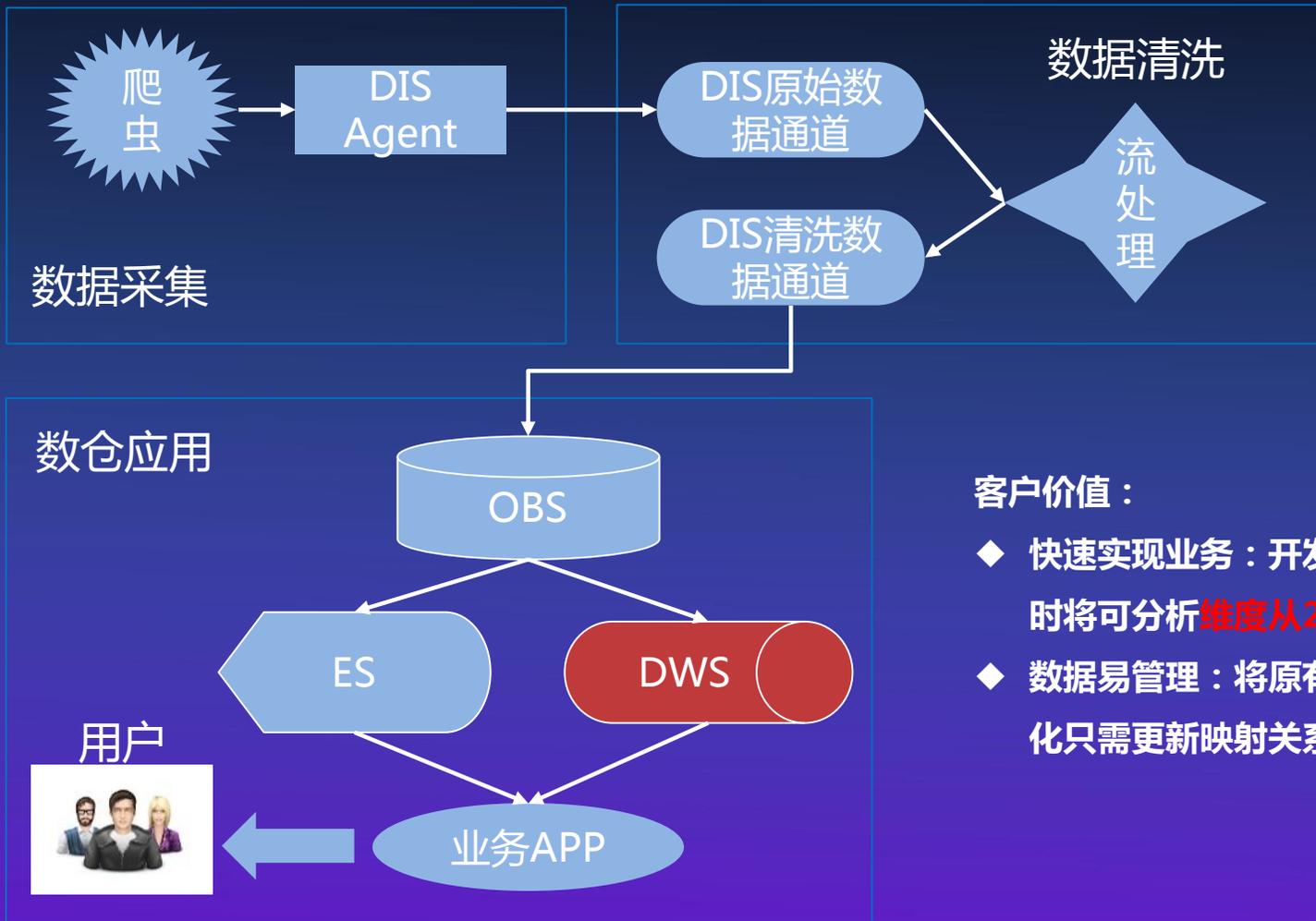
- ◆【数据量大】：计划存储300家分公司1年的历史数据，总计约300TB
- ◆【查询延时低】：查询每个分公司1年的历史数据（1TB），响应<10秒



客户价值：

- ◆ 在线弹性扩容，快速响应业务量变化：支持分阶段业务扩展目标，最终支持集团**300家**分公司数据存储和查询
- ◆ 全脱管服务，易运维：客户业务系统规模持续增长，DWS的运维能力帮助客户**降低运维工作量**
- ◆ 快速实现业务迁移：客户开发人员对新技术掌握能力不足，**标准SQL**符合客户的技术能力，易于上手；原有SQL Server上的业务基本**无改造**实现迁移。

助力某互联网公司提升数据分析能力



客户痛点：

- ◆【业务实现复杂】：现有技术框架无法实现表间关联查询，复杂的业务逻辑实现难度大
- ◆【数据更新难】：数据使用宽表存储，维度数据变化需要更新整个宽表，工作量大（数据量x亿/月，存量数据xx亿）

客户价值：

- ◆ 快速实现业务：开发人员基于熟悉的SQL语言可快速开发分析应用，同时将可分析维度从2-3个扩展为5-10个，扩充业务范围。
- ◆ 数据易管理：将原有宽表拆分为事实表、维度表和映射关系表，维度变化只需更新映射关系表，大大降低了数据更新的工作量。



THANK
YOU

华为云
技术
私享会

