



第十一届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2020

架构革新 高效可控



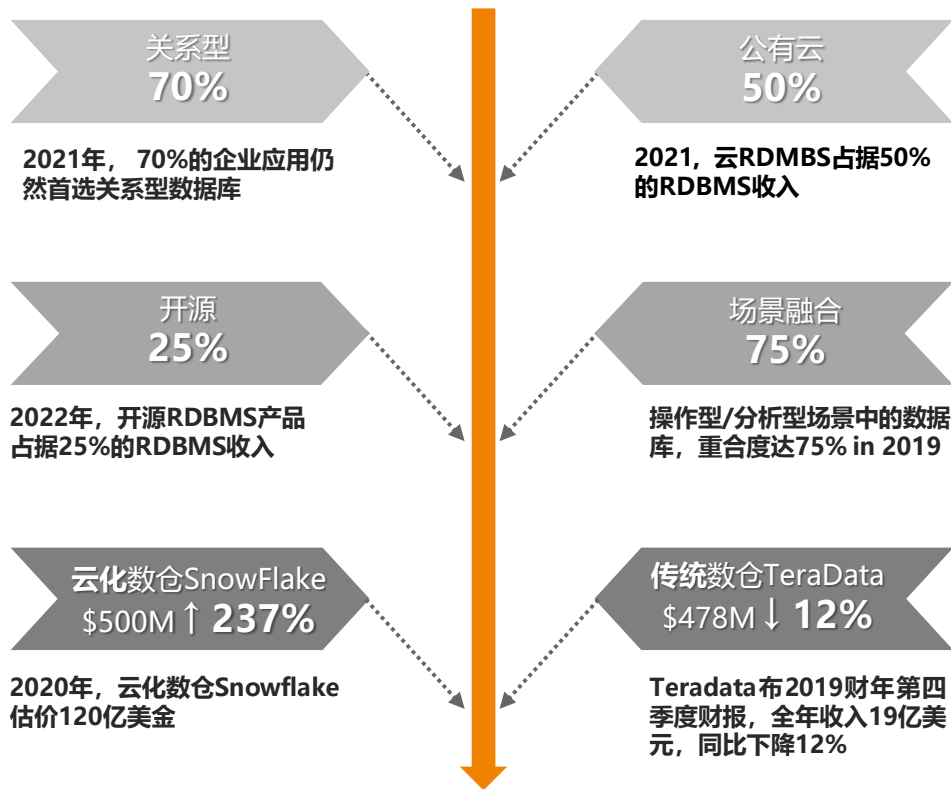
北京国际会议中心 | 2020/12/21-12/23

华为集团IT数据库上云实践

黄兆勤 | 华为HIS大数据平台首席架构师
2020年12月

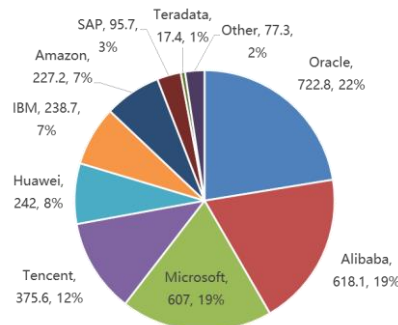


企业数据库市场以及上云趋势分析



观点:

1. 数据库总体市场空间依然强劲增长, 中国数据库市场增长迅速, 远超全球市场
2. 中国数据库市场其中45%以上是云服务形态, 云数据库增速远超整个数据库市场增速。
3. OLTP和OLAP有融合的趋势, 大数据和数据库走向一体化



数据来源: Gartner

预计到2025年中国区数据库市场空间总计约104亿美金 (以Gartner为基数), 其中OLTP约62亿、OLAP约24亿、NoSQL约18亿

流程IT数据库上云进展

GaussDB：基于统一的DFV技术架构打造的自研数据库服务

关系型数据库服务

GaussDB(openGauss)
(华为开源生态)

GaussDB(for MySQL/PostgreSQL)
(开放生态)

非关系型数据库服务

GaussDB(for Mongo)
(开放生态)

GaussDB(for Cassandra)
(开放生态)

GaussDB(for Influx)
(开放生态)

GaussDB(for Redis)
(开放生态)

数据仓库服务

GaussDB(DWS)

RDS：开源数据库服务

RDS for MySQL/PostgreSQL

公有云

HCSO

HCS

华为集团内部的数据库业务从自建转向直接使用华为云服务

华为云MySQL：

类别	华为云-MySQL
实例数	5K+
CPU	6W U+
内存	TB级
存储	4PB+

华为云PostgreSQL：

类别	华为云-PG
实例数	400
CPU	5K U+
内存	TB级
存储	TB级

华为云 MongoDB：

类别	华为云-Mongo
实例数	4K+
CPU	3W U+
内存	TB级
存储	PB级

华为云DWS（高斯A）：

类别	华为云-PG
集群数量	27
节点数	500
数据量	PB级

上述为集团内部迁移到华为云的部分数据

企业去O上云的复杂度 $O(n^3)$ 交易+分析+集成

1. 应用复杂度

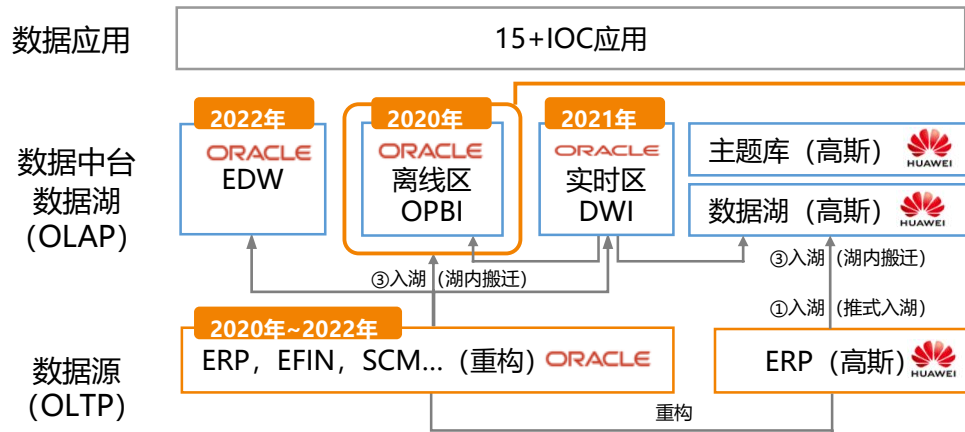
数仓 20W+作业依赖 (A) , 主要是(B) ETL作业, 或者(C) 存储过程作业

2. 交易和分析的依赖

数据/服务的上下游依赖, 业务对象重构带来模型变化

3. 数据库层面的挑战

云数据库和传统非云数据库能力有区别 (分布式), 不能简单直接搬迁。



A.调度

调度: 复杂JOB依赖

B.ETL

JOB: 单个ETL任务逻辑

C.存储过程

存储过程: 1000~5000行

华为集团去O上云三部曲，应对 $O(n^3)$ 复杂性

STEP1: 清理解耦

识别低流量无用数据资产

OLTP和OLAP解耦

老系统解耦

存量系统逐步封版本

STEP2: 新能力

+运营（面向场景的运营服务）

+治理（元数据驱动）

+AI（基于AI提升效率）

+云（使用华为云云原生能力）

STEP3: 迁移

领域迁移策略（核心->非核心）

场景迁移策略（简单->复杂）

上下游协同，关联影响分析

平滑迁移+并行系统保障

场景服务

自研服务

自研单体

软件包

电商应用

经营分析

运营分析

实时分析

自助分析

数据底座

数据感知

数据建模

数据加工

数据联接

数据共享

华为云数据基础设施

RDBMS

KV Database

Document
Database

Graph Database

In-Memory
Database

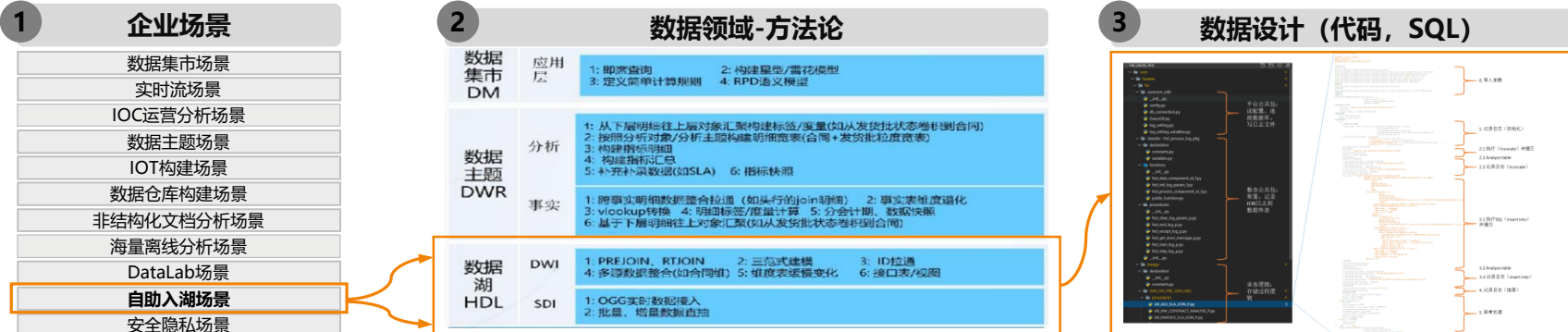
Data warehouse

TimeSeries
Database

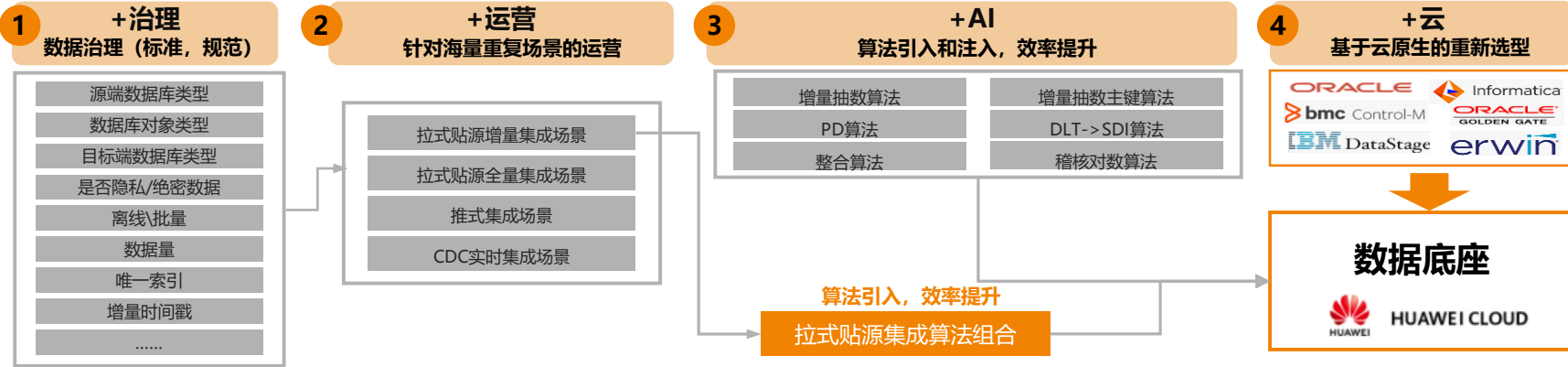
大数据

成果： +治理 +运营 +AI +云的数据数据库迁移方案

场景



工具链



数据库上云需要大量的投入和积累

数据库上云战场

解决方案专题

DDD领域建模专题	分布式访问专题	分布式事务专题
云数据实时同步专题	数据主权专题	分布式访问中间件专题
融合数据共享专题	数据比对/修复专题	数据归档清理专题
数据库高性能专题	实时数仓计算专题	数据切换策略专题

工具链专题

数据集成	数据融合	数据迁移	ETL	数据稽核
CDC	任务调度	实时计算	数据服务	分布式中间件

数据库专题

RDBMS	KV Database	Document Database	Graph Database
In-Memory Database	DWS	TimeSeries Database	大数据

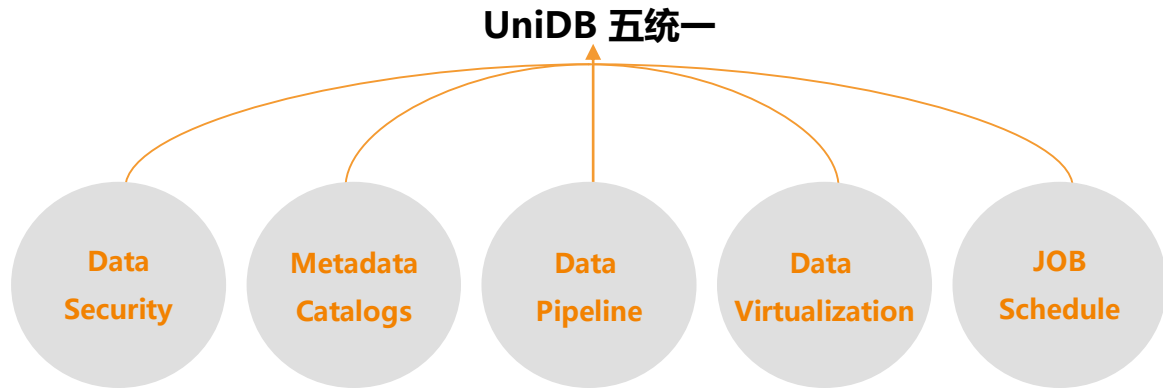
分类	工作项
整体架构设计	数据域架构整体梳理
	DDD领域模型设计
切换策略	数据库选型
	软件包的去O方案
前端应用改造	具体切换策略和计划
	前端功能解耦
后端服务改造	前端交互体验优化
	中台服务化
	数据处理代码实现逻辑
	SQL语法改造
	跨AZ的读写和高可用
数据库设计	性能调优
	事务一致性
	表结构重构
	表关联关系重构
	同库授权方式调整
	存储过程逻辑回归中台
	视图逻辑回归中台
	程序调度作业方式重构
	数据中间件的职责分配
	读写分离
分布式方案	数据库分布式架构
上线方案	数据库分布式同步方式
高可用	灰度方案/新旧并存方案
	容灾、主备方案调整
数据工具	跨AZ的读写和高可用
	DBLINK集成重构
	CDC集成重构
	数据集成
	数据归档、数据迁移
	推式入湖
	任务调度工具的重构

✓ 统一数据库平台 UniDB



企业面对数据库上云的问题1：多种不同类型数据库的融合

UniDB是华为内部
的统一数据库平台

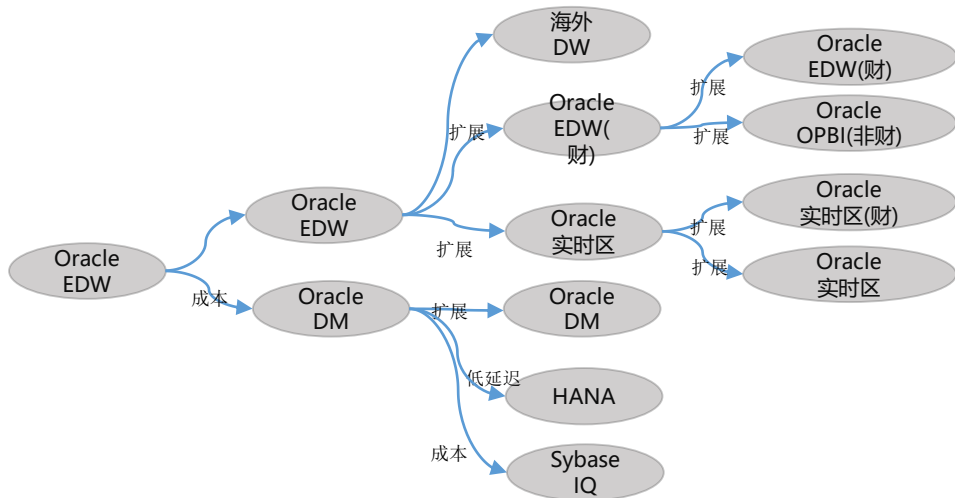


数据基础设施



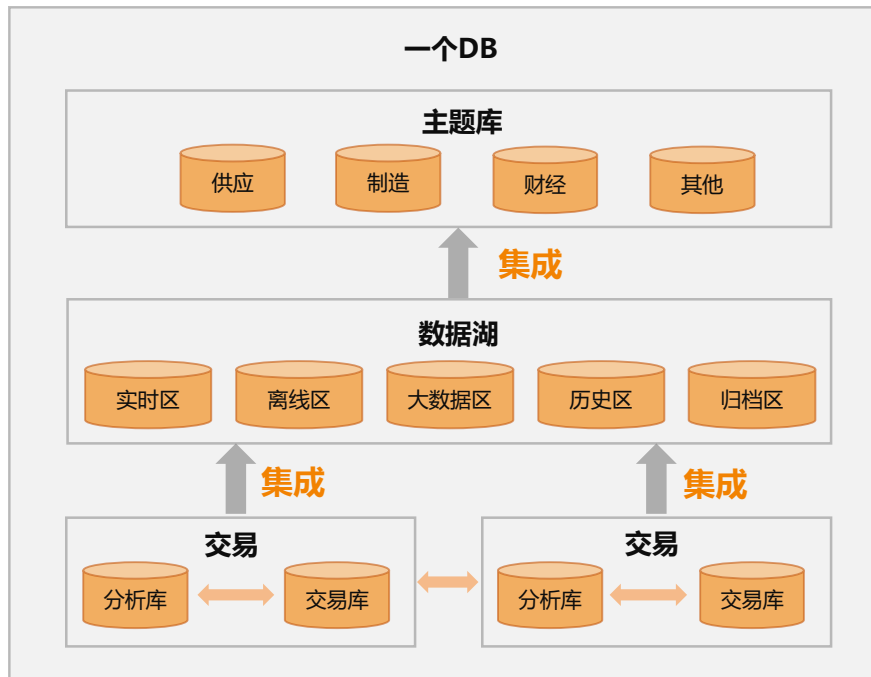
企业面对数据库上云的问题2：OLTP/OLAP融合趋势

痛点：早期Oracle切换高斯前的数据分散



- 问题1) 数据库分散，OLTP分表分库，OLAP拆库
- 问题2) 数据库多地分布，带来的数据汇聚问题
- 问题3) 交易库和分析库的边界逐步模糊，交易库承载分析行为
- 问题4) 企业领域烟囱，引入大量的技术栈

基于华为云的数据库、数仓、大数据，进行有效管理和融合



华为企业数据底座

列的革命，是企业面向外部场景多样性的需求驱动下，演进出来的一种多维度融合的数据能力，是平台的革命。

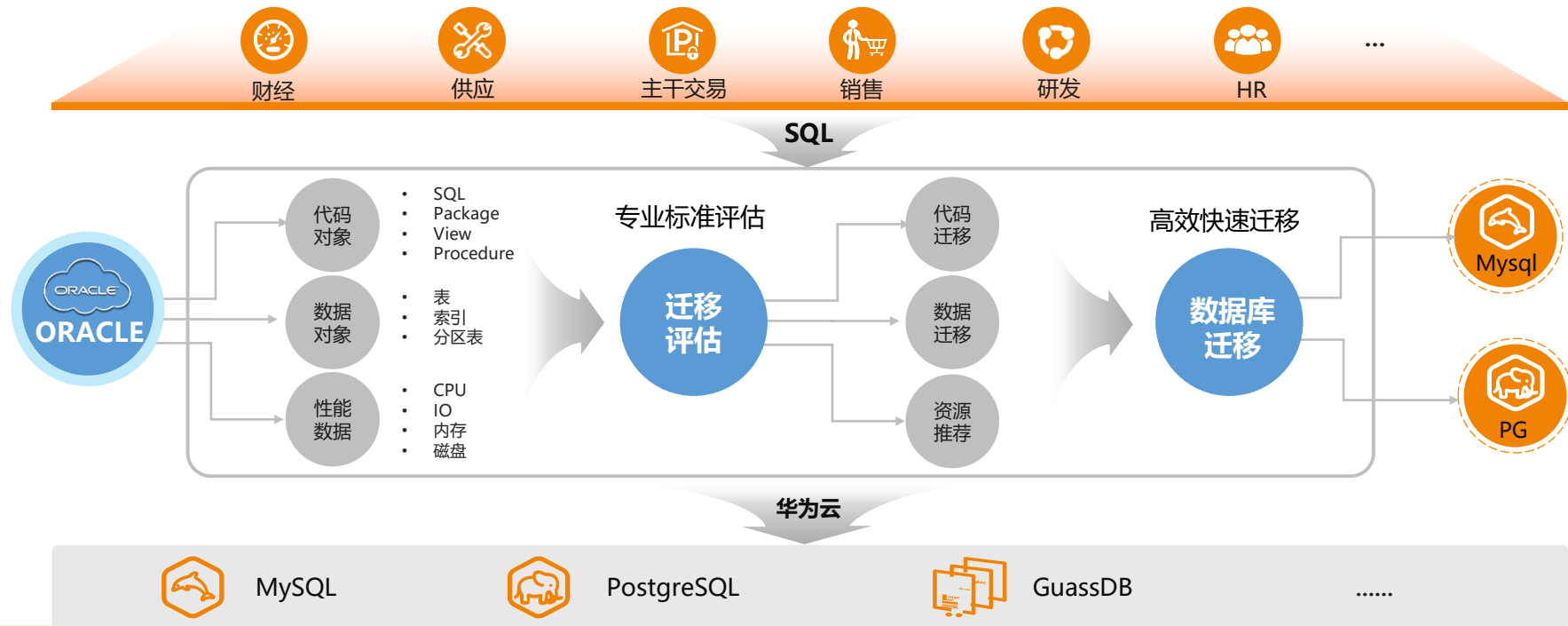


✓ 数据库迁移解决方案



数据库工具，助力企业迁移到基于华为云的数据库生态

集DBA专家经验，为产品团队提供数据库引擎推荐、资源评估和迁移难度分析，其中资源评估数据已成为容量小组评审的标准；评估效率从月至天，总体可节省750人天（按全网迁移1500个应用计算）



传统数据库上云的常见问题

传统数据库迁移面临的问题

1

源库的运行状态如何，容量和复杂度如何，库表之间关系如何？

2

如何选择合适的目标库及其规格，改造风险和工作量有多大？

3

从源库迁移到目标库，有哪些对象不兼容，在目标库是否有替代方案？

4

对于应用程序改造，如何动态或者静态抓取到应用改造的SQL？

5

人力进行语法改造，成本巨大，如何做到迁移过程半自动化或者自动化？

6

改造过程涉及业务重构时，如何为用户提供合适的改造方案？

7

源库对象较多，如何保障源库对象顺利迁移到目标库？

8

如何保证结构迁移和数据迁移时，不影响源库/目标库正在运行的业务？

9

源端按业务模块迁移时，源和目标如何进行持续的增量数据同步？

10

语法改造后，如何保证性能符合预期，如何对改造后的对象进行优化？

11

迁移完成时，如何保障迁移结果正确，包括结构的准确性与数据的准确性？

12

迁移完成后，如何保障云上数据库的开发运维与性能的持续优化？

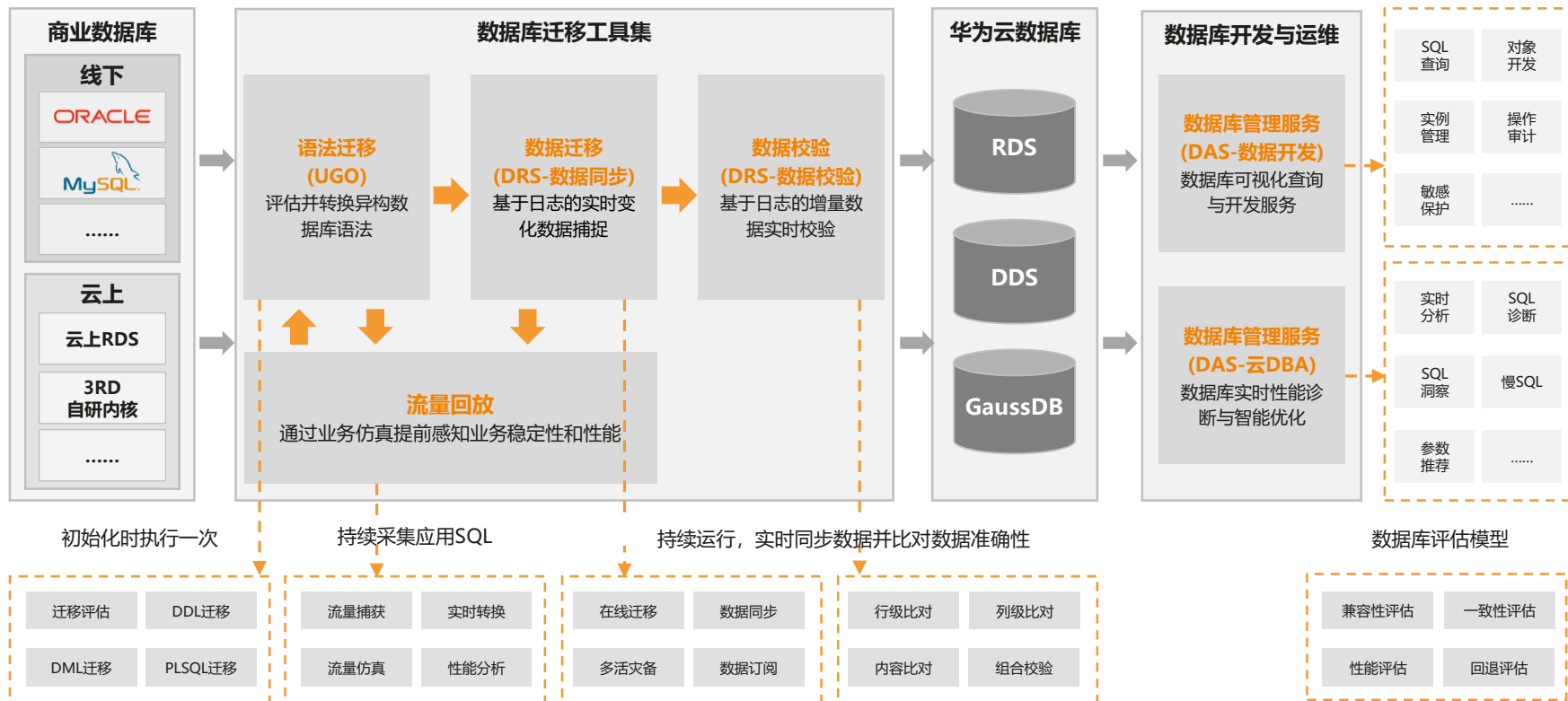
兼容性

一致性

性能

回滚

华为云数据库迁移技术解决方案



THANKS

