

# 上汽通用五菱基于 SelectDB 实时数仓实践

韦奔 运营大数据经理

## 分享嘉宾 - 上汽通用五菱



### 韦奔

运营大数据经理

2018年起负责SGMW企业级大数据平台建设，汽车领域数据建设具备丰富经验，先后与团队获得多项市级、区级、国家级数据奖项。

# 目录

01 上汽通用五菱公司简介

02 实时数仓建设背景

03 项目解决方案

04 应用成效与价值收益

01

# 上汽通用五菱公司简介



# 公司基本情况——发展历程



1958年10月  
柳州动力机械厂成立



1985年5月  
柳州微型汽车厂成立  
启用五菱品牌



2005年6月  
上汽通用五菱青岛分公司  
挂牌成立



2010年7月  
宝骏品牌发布



2023年10月  
五菱建厂65周年



1966年1月  
柳州拖拉机厂成立



2002年11月  
上汽通用五菱汽车股份  
有限公司成立



2009年12月  
年度第100万辆汽车下线

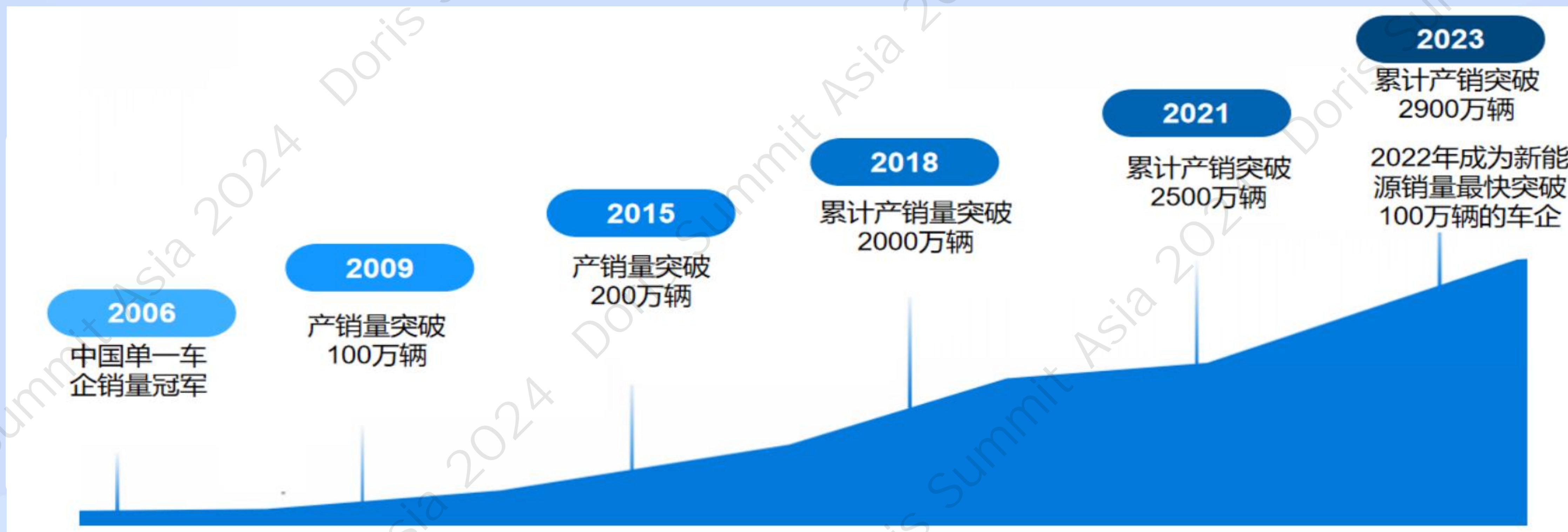


2016年12月  
公司实现年度销售收入  
突破千亿



# 市场表现

- 连续 14 年保持国内单一车企销量冠军，总销量突破 2900 万辆；
- 连续 4 年销售收入突破千亿元，排名广西工业企业第一，累计利税超 1300 亿元。





# 主要产品

 WULING 五菱汽车	 缤果 Bingo	 缤果Plus	 宏光MINI EV	 第三代马卡龙
	 星光S	 星光	 星辰 Asta	 佳辰
 BAOJUN	 悦也PLUS	 宝骏悦也 Baojun Yep	 宝骏云朵	 宝骏云海
	 五菱扬光	 五菱龙卡	 五菱宏光	 荣光小卡



02

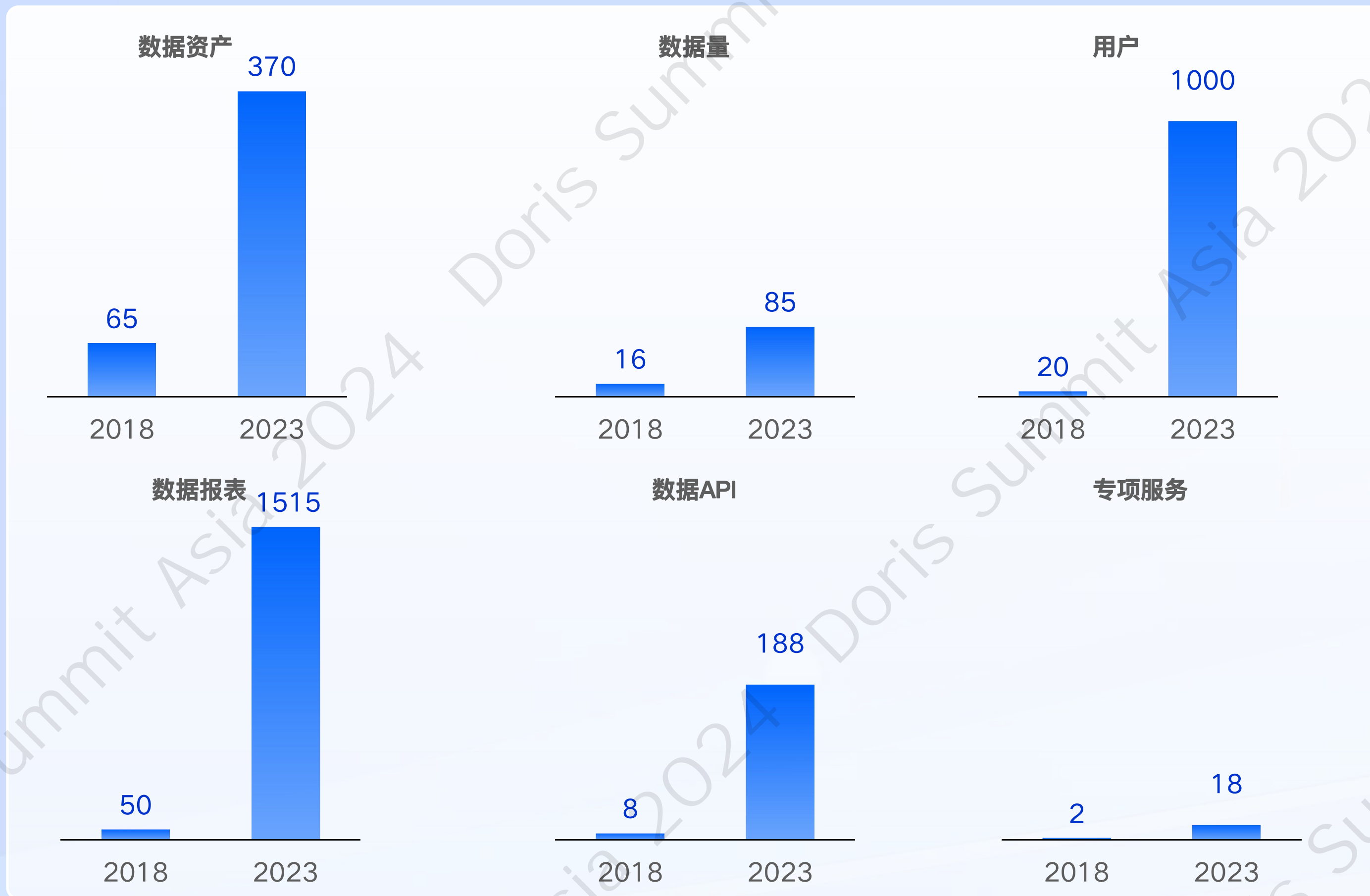
# 实时数仓建设背景



# 大数据工作历程



# 背景和痛点



## 服务背景

从 2018 年开始自主构建大数据平台，支持公司运营分析。历经 4 年运行，数据沉淀和服务积累已发生量和质的变化，覆盖从执行层到决策层的数据服务支持。

## 服务问题

随着数据量及服务增加，大数据平台的处理能力出现瓶颈：

- 仅具备离线批处理能力；
- 数据共享 API 处理超时严重；
- 数据查询效率低，无法满足秒级需求

# 项目目标

## 实时数据处理能力

- 支持 kafka、Flink 等实时数据采集方式；
- 支持快速的数据处理能力
- 支持数据的快速同步、推送
- 支持外表关联查询

## 秒级数据查询能力

- 支持 APP端、报表平台的秒级查询能力
- 支持数据应用的快速响应
- 支持物化视图

## 稳定的数据能力

- 保障数据商店 API 的稳定性
- 支持 MySQL 的快速迁移
- 支持资源隔离
- 支持极致的数据压缩比

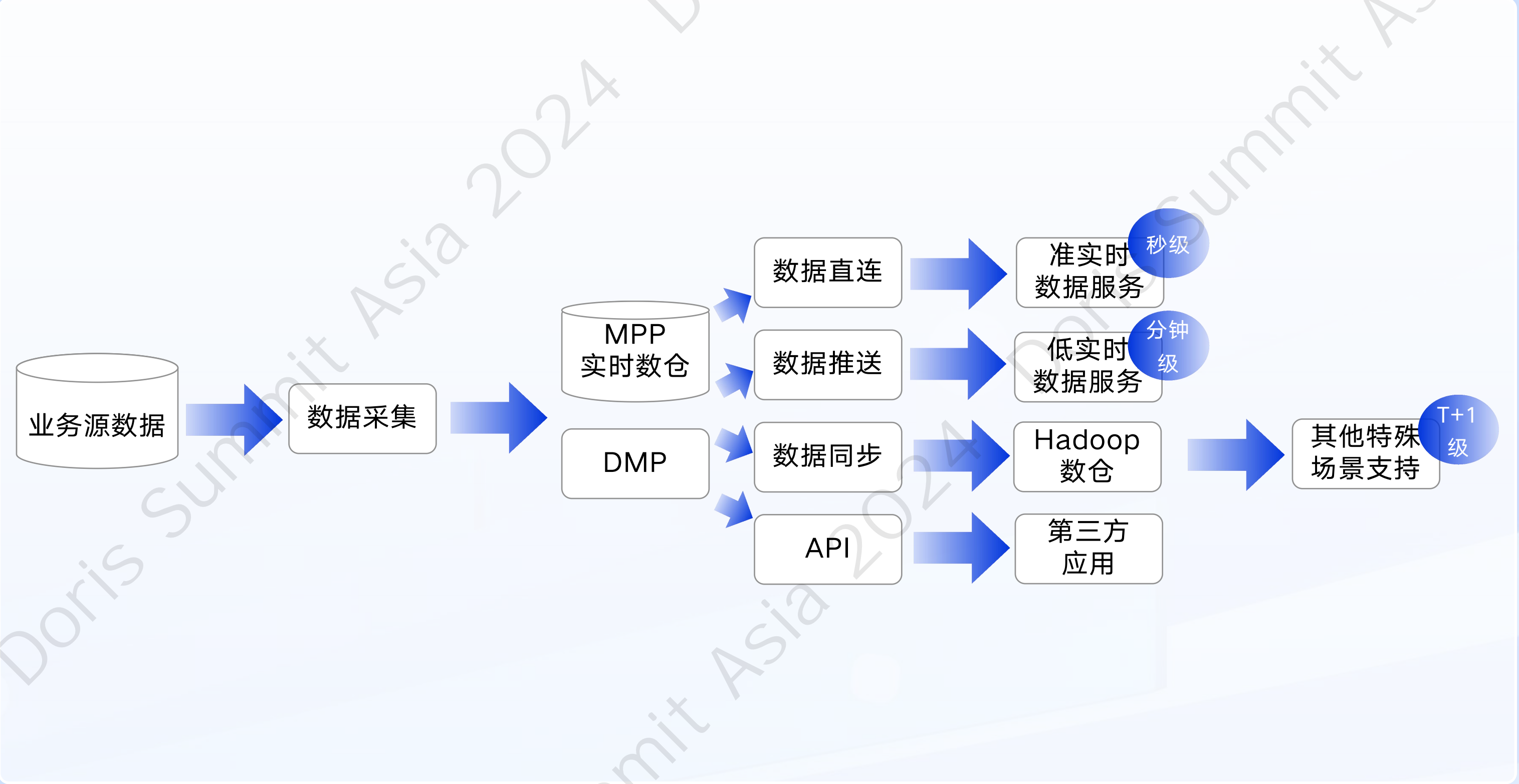
03

# 项目解决方案



# 场景设计

业务场景	业务内容
平台能力	准实时数仓(MPP 类数仓) +批量离线计算(Hadoop)
数据链路	1、JDBC 直连 2、Catalog/DCT 推送 3、Sqoop 同步 4、API 接口
服务时效	实时仓主要提供 T+0 级服务： 1、准实时：T+0 秒级服务 2、低时效：T+0 分钟级服务  离线仓主要提供 T+1 级服务： 1、离线分析：T+1 数据服务 2、历史数据：2 年以上数据归档



# 数仓产品选型

	Apache Doris	某G产品	其他商业产品
是否开源	是，有企业级分发版 SelectDB	是，2024年闭源	否
技术架构	采用MPP 架构的高性能、实时的分析型数据库,支持 MySQL、SQLServer、Oracle、Hadoop等数据库数据同步，并且完全兼容MySQL语法。可以无感迁移MySQL脚本	采用 MPP 架构，以 PostgreSQL 数据库为内核，兼容 PostgreSQL 语法,支持 MySQL、Sqlserver、Oracle、Hadoop 等数据库数据同步	通过优化存储结构、并行计算和高效索引技术，能够满足高并发、大规模数据处理需求。支持标准 SQL 语法和主流数据库接口（如 JDBC、ODBC），便于快速迁移和集成。涵盖 OLTP、OLAP 和云原生场景，提供灵活的解决方案。
扩展性	支持线性扩展节点	支持线性扩展节点	支持线性扩展节点
场景贴合度	插入 更新★ 删除★ 查询★ 数据同步 数据压缩比★ TPC-H★ 复杂脚本★ 非结构化 成本★	插入★ 更新 删除 查询 数据同步★ 数据压缩比 TPC-H 复杂脚本 非结构化★ 成本	POC成本问题，未验证

# 数仓产品选型考虑因素



兼容 MySQL



查询速度快



非单一供应商



满足业务场景

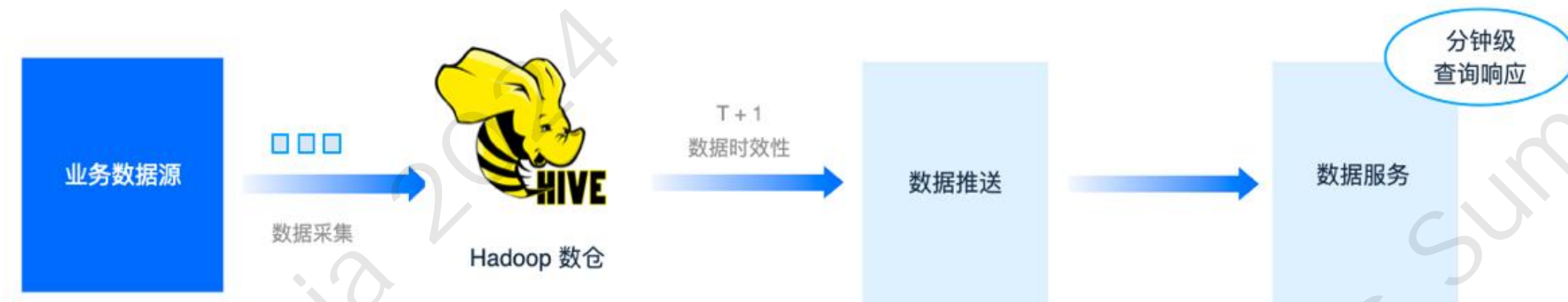
综合以上因素，最终选择 Apache Doris 的企业级分发版 SelectDB 作为我们的实时数仓底座



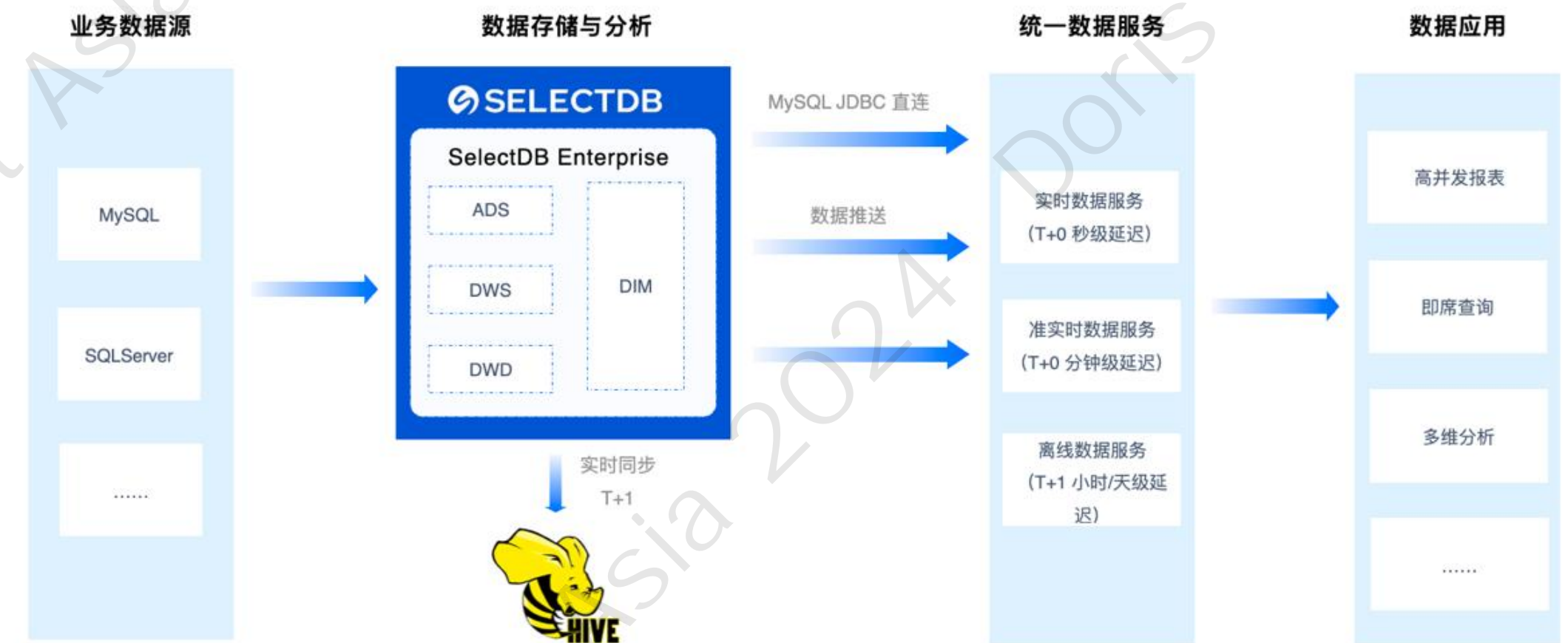
# 数仓架构设计

基于 SelectDB 的平台架构变化

原 Hadoop 平台架构图（示意）



基于 SelectDB 的架构改造图（示意）





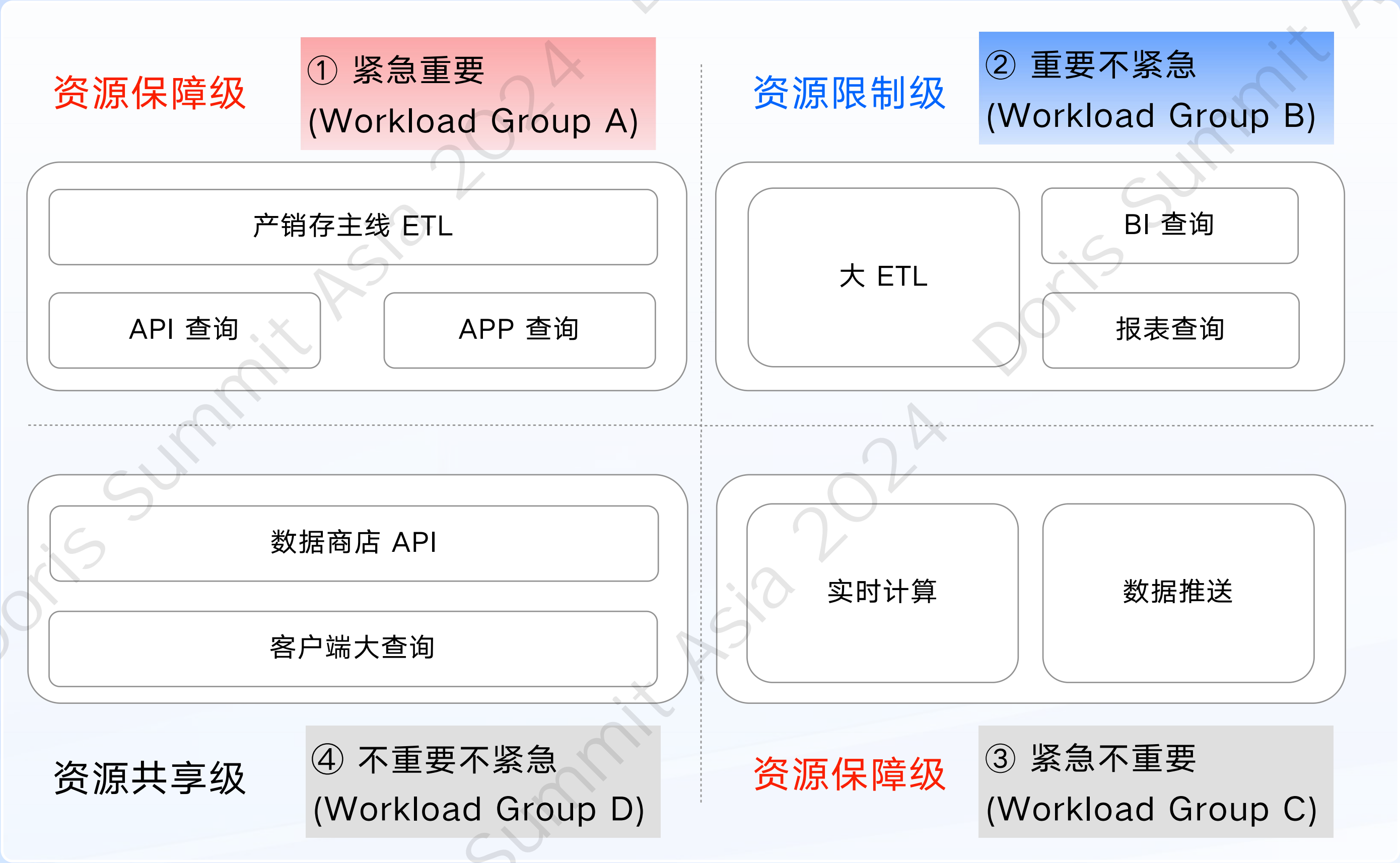
# 资源设计 (Workload Group)

## 资源划分

根据现有使用场景划分，将计算资源分为三个等级：资源保障级、资源限制级、资源共享级。

## 资源限制

单查询资源限制可以在各个等级之间使用，实现资源的二次限制。



04

# 应用成效与价值收益

# 使用场景

## 覆盖场景



## 典型场景

### 大表单表的快速查询

- 物料凭证单表超 20 亿行，查询速度可以达到 80ms
- 库存明细表超 2 亿行，字段超 100 个，查询速度也可以达到 60ms

### 外表关联的快速查询

大数据平台解决了数据汇聚问题，但实际运行过程中，总有一些手工数据或者业务数据不适合入仓。SelectDB 的外表关联查询可以解决异构数据源的关联问题，极大提升了数据服务的灵活性和有效性。



# 业务收益

## 数据服务效率提升

- 数据服务APP平均响应时间2s内，平均查询效率提升 **↑ 80%**；
- 数据商店接口无因数仓问题导致的超时问题；提升接口调用方的使用体验，以更准确的数据实现更大的隐形价值。
- 报表平台查询平均响应时间也缩短至**2s**内；
- 直连 SelectDB 数仓查询，上千万次查询中，最大耗时时间为**2.8s**

## 数据开发效率提升

- 兼容 MySQL，数据开发人员可以快速掌握，总体开发效率也随之提高 **↑ 35%**；
- SelectDB 完全兼容 MySQL 协议，原 MySQL 开发的报表平台或者 BI 分析可无感迁移，节约大量代码迁移工作量；
- 最短任务处理时间由原来的 30min 下降至 **↓ 5min**。



# 平台收益

处理任务	CDH处理耗时	SelectDB处理耗时
实销	4m	13s
批发	3m30s	11s
库存	21m	1m40s
生产	8m	16s
分车	4m	33s

## 算力提升

SelectDB 是基于 MPP 架构的高性能分析型数据库，可提供亚秒级数据响应服务。大数据平台整体计算能力 **92%** 提升。

## 存储能力提升

SelectDB 采用列式存储，具备极高的压缩比，通过减少大量非相关数据的扫描，更加有效利用 IO 和 CPU 资源。使用更少的存储资源，存储更多的数据，存储提升**2.7**倍。

## 湖仓一体能力

SelectDB 支持主流数据湖、数据库的连接访问，可以对数据湖数据进行读取，计算完成的结果入湖。可实现湖仓一体的查询计算能力。

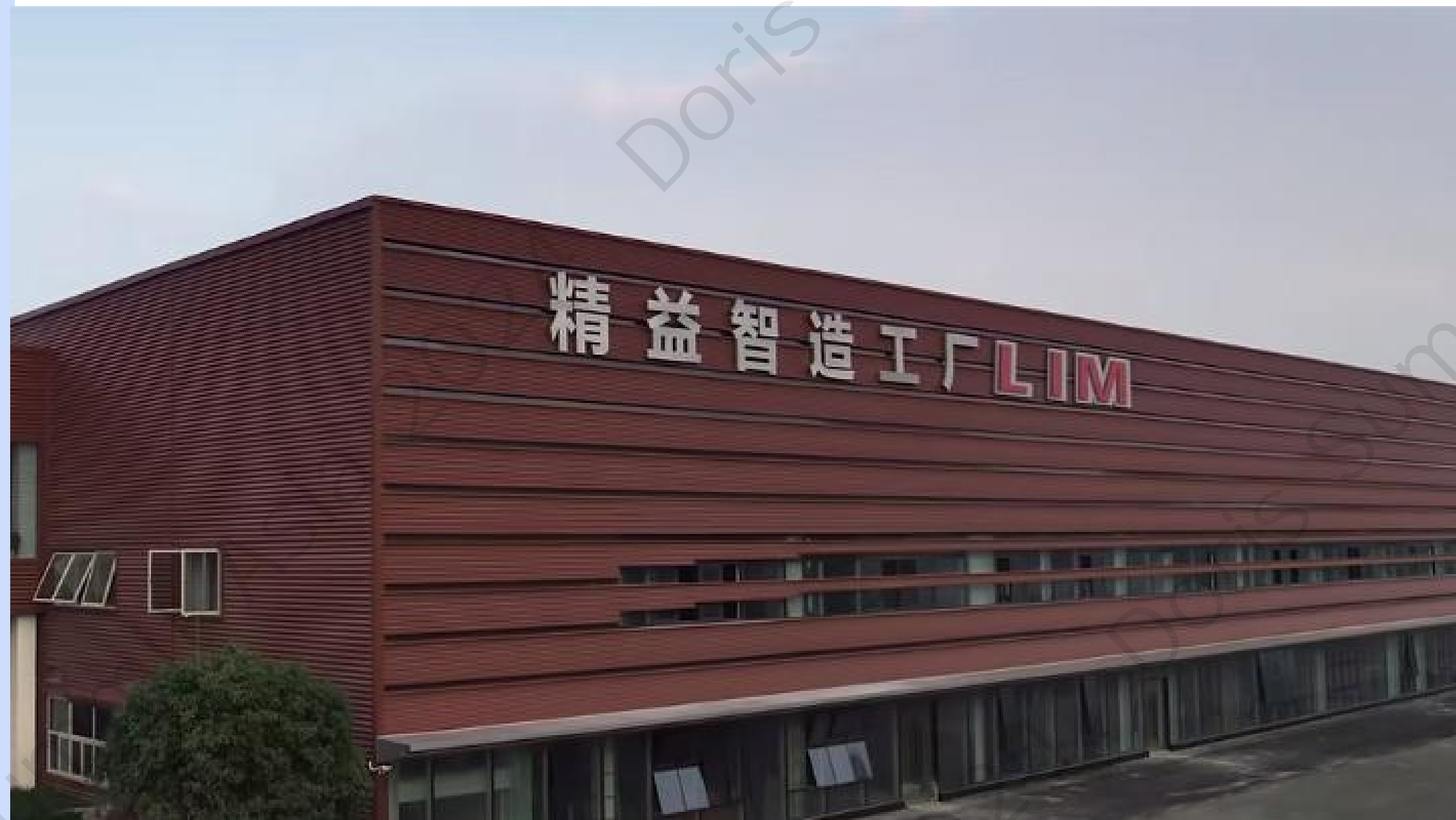
## 应用门槛低

SelectDB 采用 MySQL 协议，高度兼容 MySQL 语法，支持标准 SQL。大多数数据开发人员只要具备 MySQL 能力，就能进行 SelectDB 数仓的数据开发。

# 支持工厂生产

基于 SelectDB 的平台，为工厂生产提供数据算力支持

产线自动化、智能化水平大幅提升



智能岛汽车生产



# Thanks for Watching!