

# 数据与智能定义竞争力

SelectDB 加速汽车智能进化飞轮

# 关于 Apache Doris & SelectDB

Apache Doris 是一个基于 MPP 架构的高性能、实时的分析型数据库，以极速易用的特点被人们所熟知，常用于报表分析、即席查询、实时数仓构建、湖仓一体构建、日志和文件检索分析等使用场景。2013年诞生，2017 年正式对外开源，是全球最活跃的Apache 开源大数据项目之一，全球拥有700+社区开发者，超10000+ 企业用户。**2022年原 Doris 开发团队组建了飞轮科技公司，提供其商业化产品 SelectDB 与服务。**



## 开源 + 商业 双轮驱动

服务全球开发者、用户、客户

致力于发扬光大 Apache Doris 开源项目，基于 Apache Doris 创新企业级产品 SelectDB，实现开源和商业的良性增长飞轮。

- 1 名主席，11 名 PMC 成员，25 名 Committer
- 近三年，Apache Doris 90% 的新增代码由飞轮科技贡献

### 01 开源技术战略

Apache 第 200 个顶级项目 Doris  
国内唯一的全球领先实时数仓开源项目  
全球最活跃的开源大数据项目之一

### 02 云原生战略

基于云原生技术研发的混合云产品体系，国内首家具备完整公有云/私有云产品栈的独立厂商

### 03 开放生态战略

开源和商业互不锁定，用户自由迁移使用  
坚持被集成定位，专注新一代数仓技术产品  
100+ 伙伴集成和合作，生态繁荣

# 目录

## 数据与智能定义竞争力

SelectDB 加速汽车智能进化飞轮

头部企业实践案例

# 2026 年度中国汽车技术趋势

## 趋势1：L3级有条件自动驾驶技术方案逐渐收敛

### 技术类型：实现重大突破技术

#### 趋势解读：

□ 推动行业共识的L3级有条件自动驾驶硬件配置、算法、算力协同方案，是L3相关标准法规制定的重要依据，也是实现产品规模化落地与商业闭环的根本前提。

• L3是“辅助驾驶”与“自动驾驶”之间的关键分水岭。其技术方案的收敛是建立行业统一性能基准、完善法规体系的前提，也将有效推动核心硬件成本的进一步下降，是实现L3从技术“可行”到市场“可用”跨越的核心要素。

□ 2026年，L3级有条件自动驾驶将从当前依靠功能堆砌、配置冗余的探索，迈向技术方案统一、系统效率优化的新阶段。

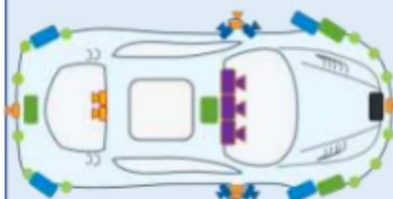
• 截止2025年，比亚迪、长安、宇通等9家车企进入L3级有条件自动驾驶汽车准入试点。

• 2026年，L3技术方案在感知硬件配置、算法范式与算力部署等方面将形成行业共识。感知硬件方面，多传感器融合成为主流，保障系统安全冗余；算法层面，重点开发端到端模型与VLA、世界模型等的融合应用，提升复杂场景的认知与决策能力；为支撑高性能算法的实时运行，单车算力部署向1000-1500 TOPS集中。

### 重点企业



#### 典型L3感知系统传感配置示意



#### 技术评价指标表现



数据来源：中国汽车工程学会，2026年我国汽车技术趋势调查项目

# 2026 年度中国汽车技术趋势

## 趋势2：端到端AI Agent智能座舱即将进入规模化量产阶段

AISI 中国汽车战略院  
CSAE-AISI

技术类型：新量产技术

趋势解读：

□ 智能座舱端到端AI Agent拥有多模态融合感知与上下文长程推理能力，将推动汽车实现从“人适应车”到“车适应人”的范式转型。

- 智能座舱端到端AI Agent可融合舱内外多渠道信息与用户习惯进行自主决策，将显著降低用户交互频次、拓展服务场景，推动智能座舱从“指令执行者”，演进为可主动预判需求、具备情感交互能力的“出行伙伴”。

□ 2026年，智能座舱端到端AI Agent迎来量产元年，向多模态协作、长时空记忆、多场景服务演进。

- 智能座舱端到端AI Agent将按照“两段式”发展演进：首先，依托推理与记忆能力的突破，显著增强智能座舱在理解、规划与决策等方面的认知能力；随后，逐步实现音频、视觉、车辆信号等多模态深度融合，最终达成场景化、个性化的智能服务体验。
- 2026年，蔚来、东风、奇瑞、丰田等车企计划将推进智能座舱端到端AI Agent上车搭载应用。

重点企业



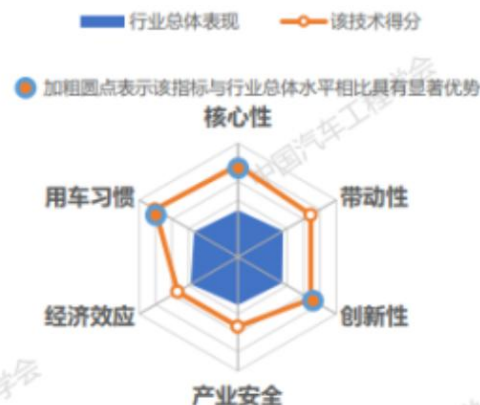
多模态协作

长时空记忆

多场景服务



技术评价指标表现



数据来源：中国汽车工程学会，2026年我国汽车技术趋势调查项目

# 数据与智能定义竞争力

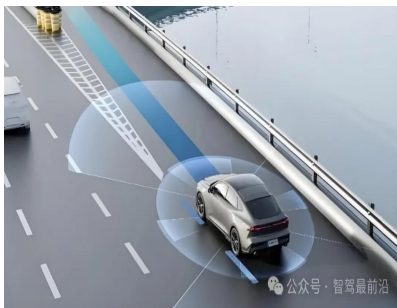
## 2026年智能化会真正成为影响消费决策的核心变量。

2024年智驾的主流车型价格带分布在20万元以上，而截止2025年末，10万以上近一年销量突破10万辆的所有车型，均搭载了智能驾驶系统。

	是否有智驾版	智驾产品		是否有智驾版	智驾产品
Model Y	有	FSD	问界M8	有	华为 ADS
小米SU7	有	Xiaomi Hyper Assisted Driving	宋PLUS DM	有	天神之眼-C
Model 3	有	FSD	银河E5	有	千里浩瀚H3
宋Pro DM	有	天神之眼-C	问界M9	有	华为 ADS
理想L6	有	VLA	小米YU7	有	Xiaomi Hyper Assisted Driving
小鹏 MONA M03	有	AI智能辅助驾驶 XNGP	宋PLUS EV	有	天神之眼-C
元PLUS	有	天神之眼-C	问界M7	有	华为 ADS
宋L DM	有	天神之眼-C	零跑C11	有	Leapmotor Pilot
零跑C10	有	Leapmotor Pilot			

图：截止25年11月销量突破10万热销车型智驾版本，来源：懂车帝、车企官网

# 数据与智能定义竞争力



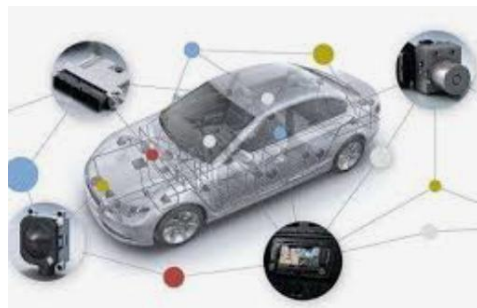
## 自动驾驶

从离线规则筛选转向数据驱动在线仿真，需从PB级数据中快速检索特征片段，要求实时处理与高吞吐。



## 智能座舱

实时分析用户行为数据，动态调整交互策略，实现个性化体验与主动服务，对数据实时性和一致性要求高。



## 车联网

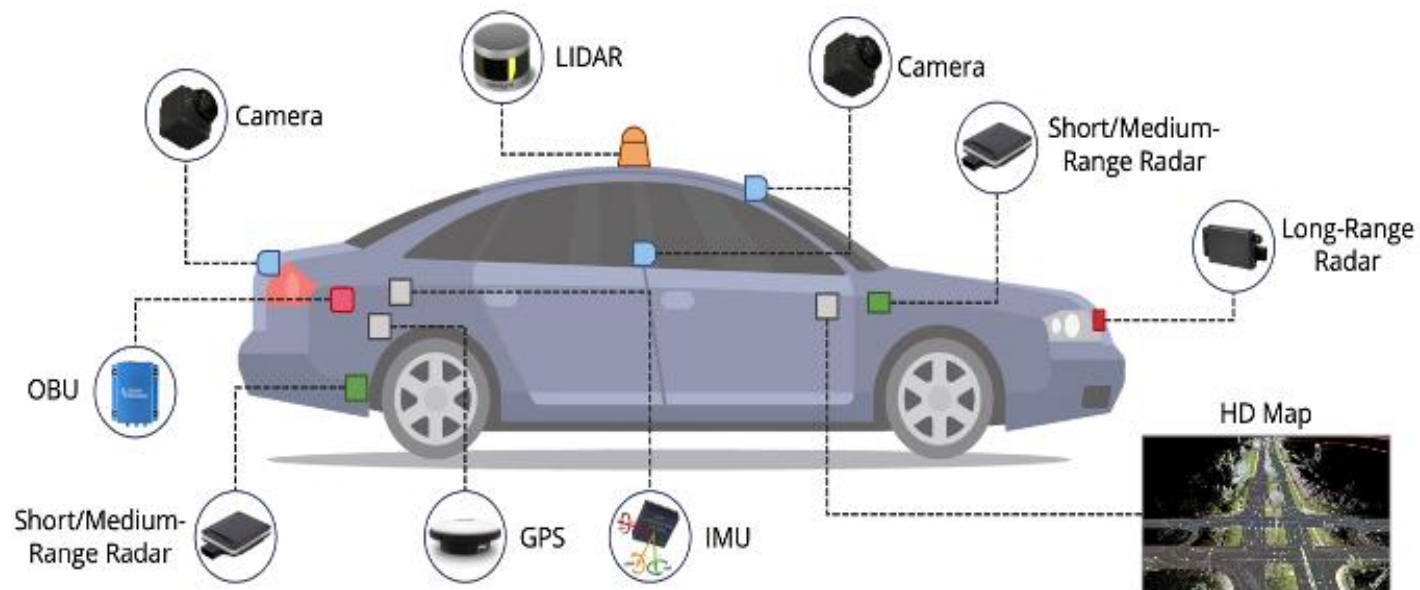
实时监控电池健康与驾驶行为，构建热失控预警模型并干预危险行为，要求高吞吐写入与低延迟响应。



## 智能制造

实时分析生产数据，识别潜在缺陷并预测设备故障，实现过程质量控制和预测性维护，提升产线利用率。

# 以数据闭环为核心的汽车智能进化 – 数据特点



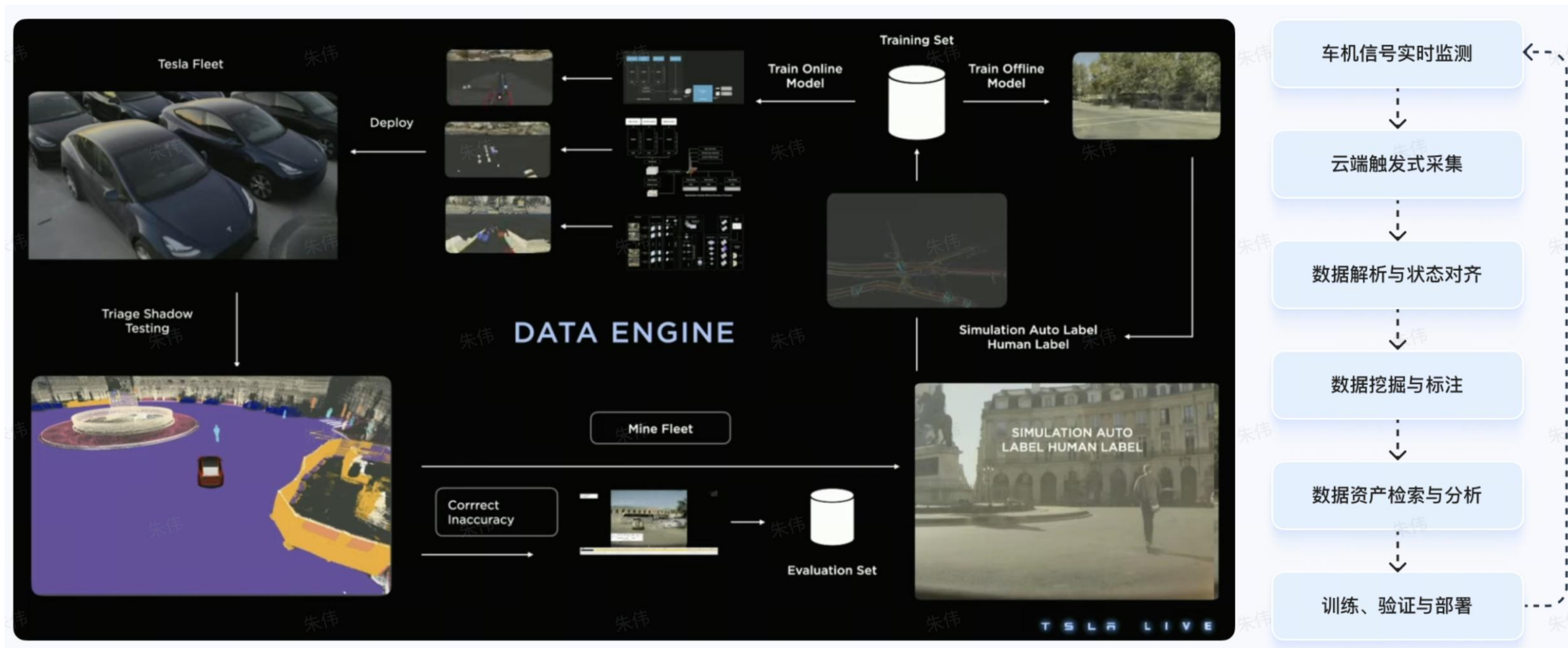
- 数据规模：爆炸性增长
- 数据类型：多模态交织
- 使用方式：在线协同



DORIS

SELECTDB

# 以数据闭环为核心的汽车智能进化



图片来源：2023 Investor Day | Tesla

# 汽车智能进化-数据闭环中的三元模型



## 实时搜索与分析

REAL-TIME ANALYSIS

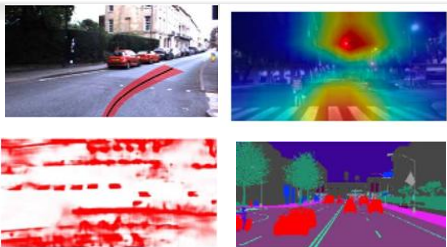
- ✓ 安全与体验：秒级故障预警与状态响应
- ✓ 模型迭代加速：交互式探索而非批量等待



## 世界状态

WORLD STATE

- 🔄 连续演化：车辆是状态载体而非事件生产者
- 🕒 高频采样：7×24小时的稳态数据流
- 📡 高维状态：万级信号维度的瞬时快照



## 语义空间

SEMANTIC SPACE

- ✖ 指数膨胀：认知维度的非线性扩张
- 📦 超大规模：百亿样本与万级标签组合爆炸
- 📐 结构复杂：枚举、嵌套、文本、向量



Traffic Signals



Critical Objects



Motion Intentions



Scene Description

# 超大规模复杂语义空间中的资产挖掘



## 3.1 帧级标签数据

STRUCTURED DATA

- ✓ 已加工为结构化标签（数值 / 枚举），数据最为规范。
- ▼ 查询模式：等值筛选 + 多条件组合 + 聚合统计 (count / group by)。
- ⚡ 算法、质量、运营共用的“秒级数据地板”，访问频率最高。



## 3.2 JSON 资产元数据

SEMI-STRUCTURED

- ✓ 复杂嵌套的 JSON 数据，包含数据质量、设备状态、采集条件等。
- 🔍 需要按 JSON Key / Path 灵活深入地做筛选、统计和分析。
- 🧩 字段结构动态变化，传统数据库难以高效处理。



## 3.3 文本类资产描述

UNSTRUCTURED TEXT

- ✓ 图片/视频生成的 QA 对、场景描述、问题说明、人工标注备注。
- 🔍 需要支持全文检索 (Full-Text Search) 与结构化条件结合使用。
- 💡 利用自然语言语义辅助定位复杂场景。



## 3.4 向量数据

VECTOR EMBEDDINGS

- ✓ 图像、点云、场景 embedding 特征向量数据。
- 🎯 支撑相似性检索 (Similarity Search)，发现相似画面与场景。
- 💡 是 Corner Case 挖掘和数据复用的关键能力，实现“以图搜图”。

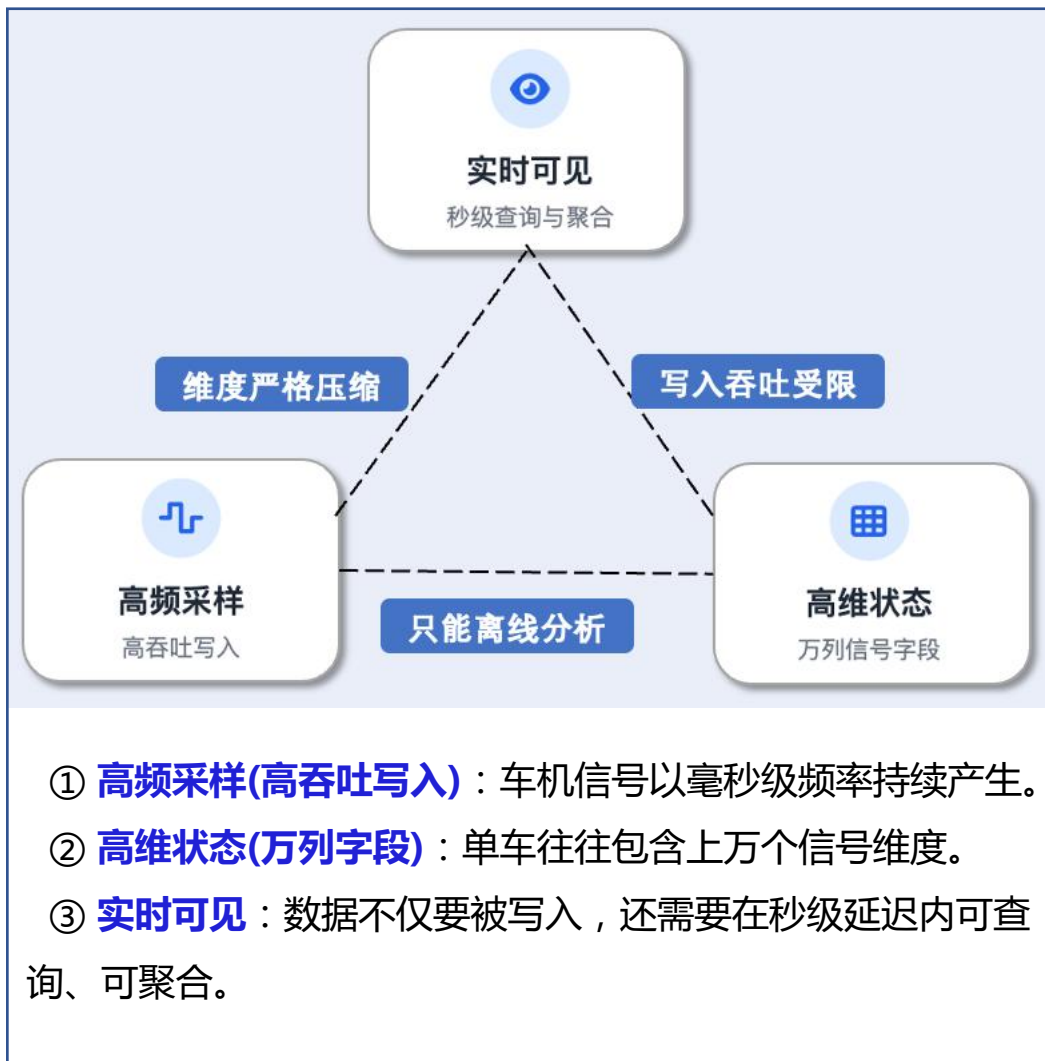


DORIS



SELECTDB

# 数据闭环-两大不可能三角挑战



# 现有数据库系统的局限

现有数据库	不可能三角 - 高频状态可见			不可能三角 - 复杂语义可得			备注
	高吞吐写入	万列字段	实时可见	超大规模	复杂结构	即时可得	
离线大数据系统 ( Hive+Spark )							1. 批处理能力强，数据时效和性能不足，难以满足数据闭环效率
实时OLAP数据库 (Clickhouse)			!				1.实时性是他们的优势，但是在万列规模下高吞入写入的同时，查询性能就有较大的影响 2.他们对复杂结构的数据，全文检索、向量索引、复杂Json处理等方面的能力，非常弱
搜索类数据库 ( ES/MongoDB )							1.他们对复杂结构类型的检索能力较强，但是在高吞吐写入、标签圈选、聚合分析等方面不足。 2.在超大规模情况下，他们的成本非常高

# 目录

数据与智能定义竞争力

SelectDB 加速汽车智能进化飞轮

头部企业实践案例

# SelectDB加速汽车智能进化飞轮



# SelectDB 加速汽车智能进化飞轮



高吞吐  
写入

万列  
字段

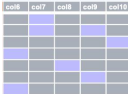
实时  
可见

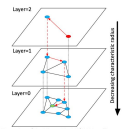
超大  
规模

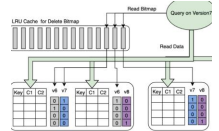
复杂  
结构

即时  
可得



 Variant稀疏列

 全文\向量索引

 Bitmap主键模型

# SelectDB 百亿级数据高性能写入与检索



## 高吞吐写入

面向 PB 级数据的持续、高并发写入与索引构建

离线标签挖掘峰值

**1000** 万 帧 / 秒

数据规模支持

**PB** 级

- ✓ 写入吞吐保障
- ✓ 系统极高稳定性

## ⚡ 多模高性能检索与分析能力



### 标签检索

数值 / 枚举型

查询响应

**< 10s**

百亿级帧数据

并发支持

**200+**

算法工程师

- ▶ 每帧万级标签支持
- ▶ 200组以上Tag组合
- ▶ Count / Group By 分析



### JSON/文本

半结构化 / 非结构化

处理对象

**复杂嵌套 JSON**

按 Key 路径过滤统计

搜索模式

**全文检索**

支持与结构化条件组合

- ▶ 百亿级 JSON 数据
- ▶ 文本资产混合查询
- ▶ 灵活应对Schema变化



### 向量相似性

Embedding 特征

数据规模

**百亿级**

向量数据

核心价值

**Corner Case 挖掘**

图像/场景相似性检索

- ▶ 高性能 ANN 索引
- ▶ 向量 + 标量联合查询
- ▶ 精准定位长尾样本



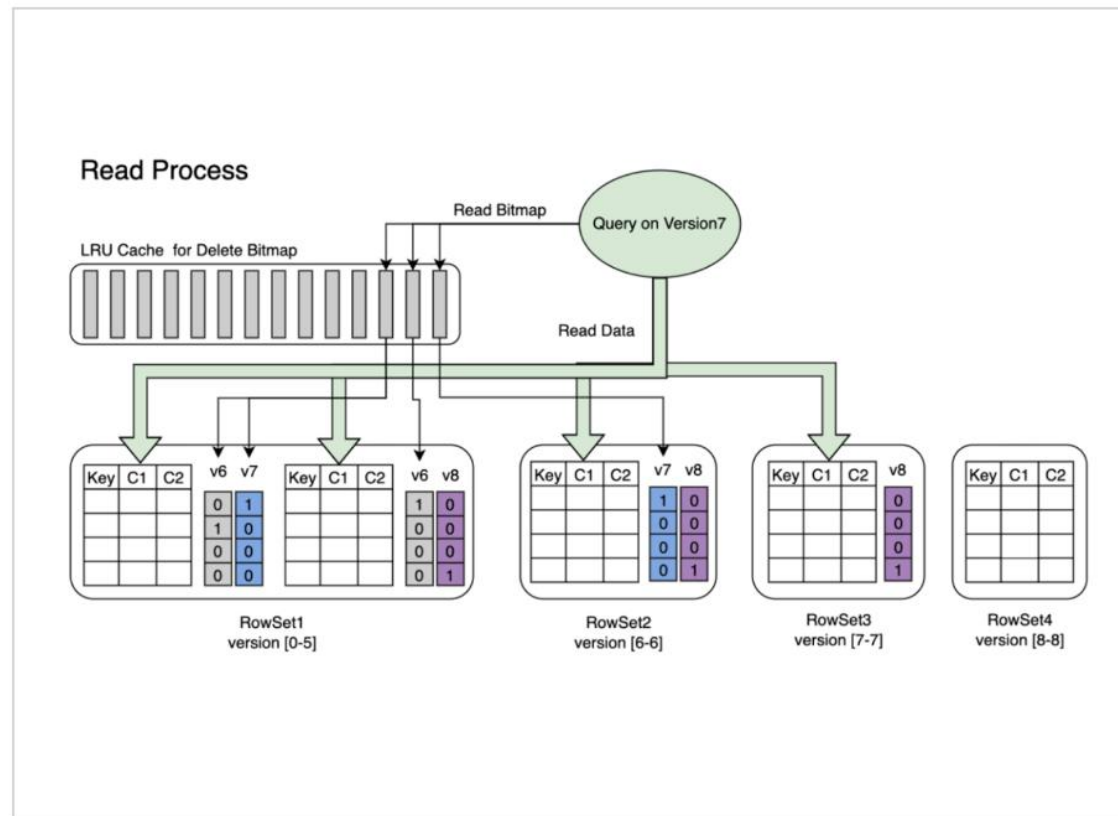
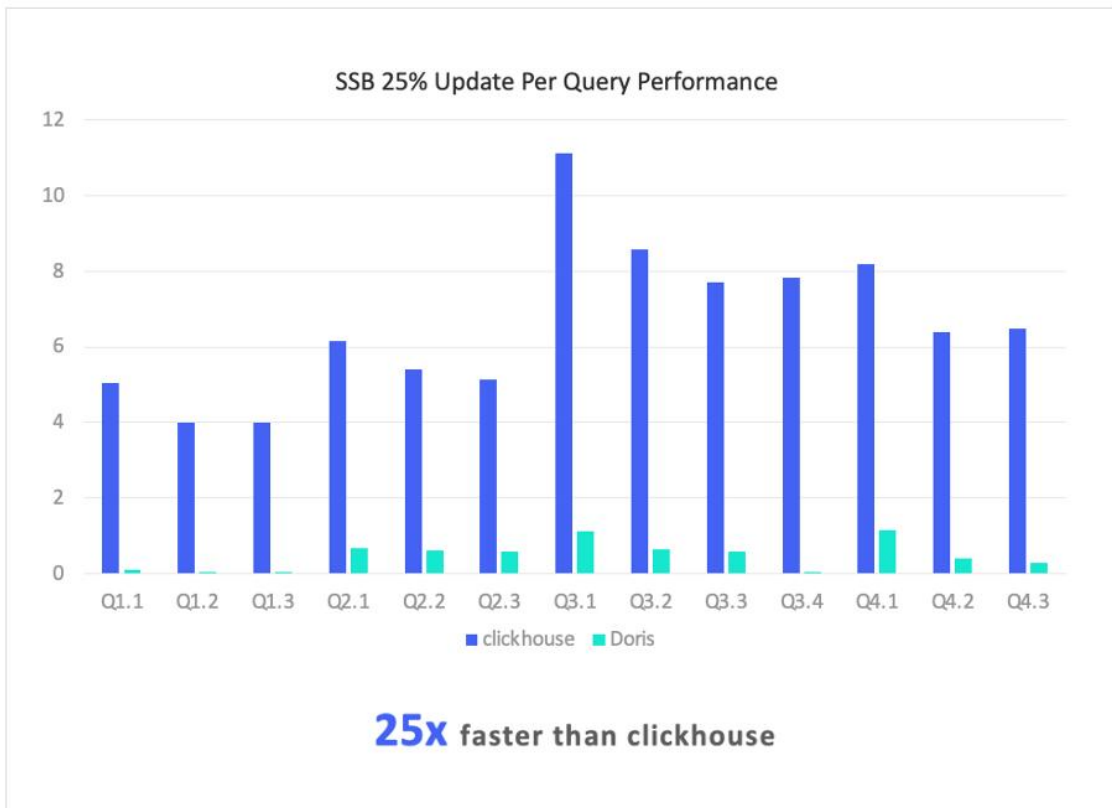
DORIS



SELECTDB

# 基于Index + Bitmap高效更新的主键模型

在高并发写入更新的同时，提供实时查询性能



```
{
  "errCodeList": [
    {
      "Work_Odd": "行车",
      "business_side": "端到端训练",
      "car_type": "CD701EV, CD701REEV",
      "code": "QC004027",
      "description": "定位信息不可靠, 要求sdm_kalman_filter_localization_admtery_eda_cdrtdv-loc_status信号值只能等于4(kFusionFilterLat)或5(kFusionFilter)",
      "errTimestampList": [],
      "error_level": "ERROR",
      "error_type": "其他",
      "interface_version": "2740, 2749",
      "topic": "sdm_kalman_filter_localization_admtery"
    },
    {
      "Work_Odd": "行车",
      "business_side": "端到端训练",
      "car_type": "CD701EV, CD701REEV",
      "code": "QC005006",
      "description": "map的topic时间戳没对齐, 要求pInfo的起始时间戳较dat起始小于3s, 结束时间戳较dat结束小于3s",
      "errTimestampList": [],
      "error_level": "WARNING",
      "error_type": "其他",
      "interface_version": "2740, 2749",
      "topic": "sdm_kalman_filter_localization_admtery"
    }
  ],
  "qc0operatorVersion": "20070106_1000",
  "result": false,
  "type": "e2e_QC",
  "typeName": "端到端采集车质检"
}
```

### Schema Less

- 可以支持任意类型、任意形状的 JSON 格式数据
- 自动动态地处理列增加、类型变更
- 不需要繁琐的 DDL 操作以及 Schema Change 操作

### 高性能

- 自动推断类型进行列式存储
- 与普通字段一样的查询效率
- 通过优化Compaction策略, 可支持 **1万列** 的存储

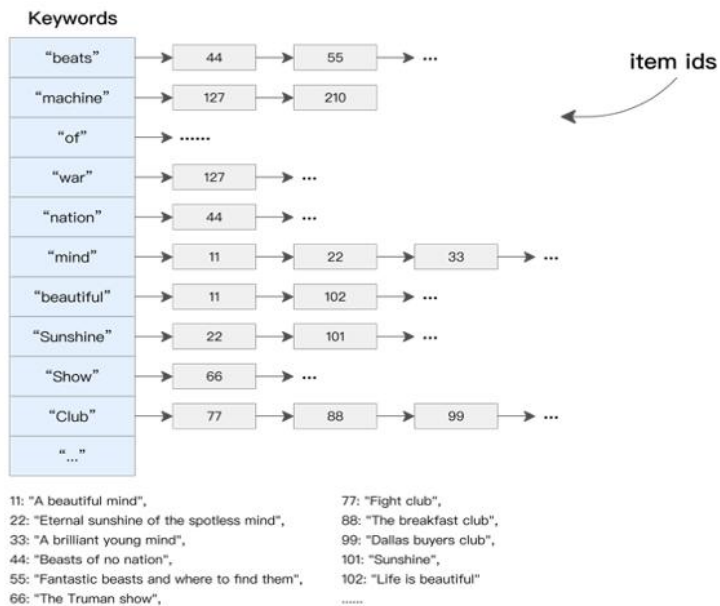
	存储空间	第一次查询	第二次查询	第三次查询
预定义静态列	12.618 GB	233.79 s	86.02 s	83.03 s
Variant 类型	12.718 GB	248.66 s	94.82 s	92.29 s
JSON 类型	35.711 GB	大部分超时	789.24 s	743.69 s



# Full Text Search

## 文本元数据搜索

### Inverted Index



- Equality and set: =, !=, IN, NOT IN
- Range: >, >=, <, <=, BETWEEN
- Null checks: IS NULL, IS NOT NULL
- Arrays: array\_contains, array\_overlaps

### Search Syntax

#### 分词

- Ik 分词
- Jieba 分词
- Pinyin 分词
- ICU 分词
- 自定义分词

#### Match 语法

- MATCH\_ANY
- MATCH\_PHRASE
- MATCH\_ALL

#### Query DSL

```
SELECT id, title
FROM search_test_basic
WHERE SEARCH('category:Technology AND NOT title:Machine');
```



DORIS



SELECTDB

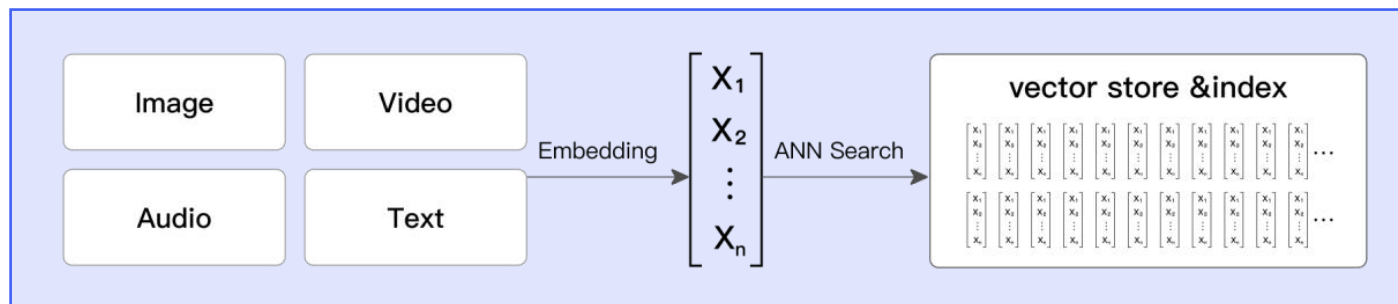
# Vector Search

## 海量多模数据的向量搜索

### ANN

- 算法: HNSW
- 向量类型: Array<float/double>
- 距离计算: L1, L2, COSINE, Product
- 量化方式: SQ4, SQ8, PQ

```
CREATE TABLE sift_1M (  
  id int NOT NULL,  
  embedding array<float> NOT NULL COMMENT "",  
  INDEX ann_index (embedding) USING ANN PROPERTIES(  
    "index_type"="hnsw",  
    "metric_type"="l2_distance",  
    "dim"="128",  
    "quantizer"="pq",      -- Specify using PQ for quant  
    "pq_m"="2",           -- Required when using PQ, in  
    sub-vectors  
    "pq_nbits"="2"        -- Required when using PQ, in  
  )  
) ENGINE=OLAP  
DUPLICATE KEY(id) COMMENT "OLAP"  
DISTRIBUTED BY HASH(id) BUCKETS 1  
PROPERTIES (  
  "replication_num" = "1"  
);
```



### Top N Query

```
SELECT id,  
       L2_distance(  
         embedding,  
         [0,11,77,24,3,0,0,0,28,  
          0,6,92,8,14,73,125,29,  
          50,25,70,64,7,59,18,7,  
        ] AS distance  
FROM sift_1m  
ORDER BY distance  
LIMIT 10;
```

### Range Query

```
SELECT count(*)  
FROM sift_1m  
WHERE l2_distance_approximate(  
  embedding,  
  [0,11,77,24,3,0,0,0,28,70,  
   0,6,92,8,14,73,125,29,7,0,  
   50,25,70,64,7,59,18,7,16,.  
  ] > 300;
```

# Hybrid Search and Analytics Processing

统一搜索与分析：复杂数组嵌套JSON结构的元数据 + 文本 + 向量 + 结构化标签

"Unified Engine, Unified Optimizer"

在同一引擎中统一调度 结构化分析、全文搜索 与 向量搜索，实现真正的协同执行。



# 基于 SelectDB 的 Lakehouse 架构

SelectDB与 Iceberg 等数据湖格式无缝集成，构建强大的湖仓架构，将数据湖的灵活性与分析型数据库的高性能结合起来。



## 原生连接器集成

SelectDB 提供丰富的原生数据源连接器，支持直接查询，无需移动或复制数据。



## 查询分析加速：

在保留 Catalog 数据治理能力的同时，SelectDB 为实时分析增加了 MPP 查询加速与优化。

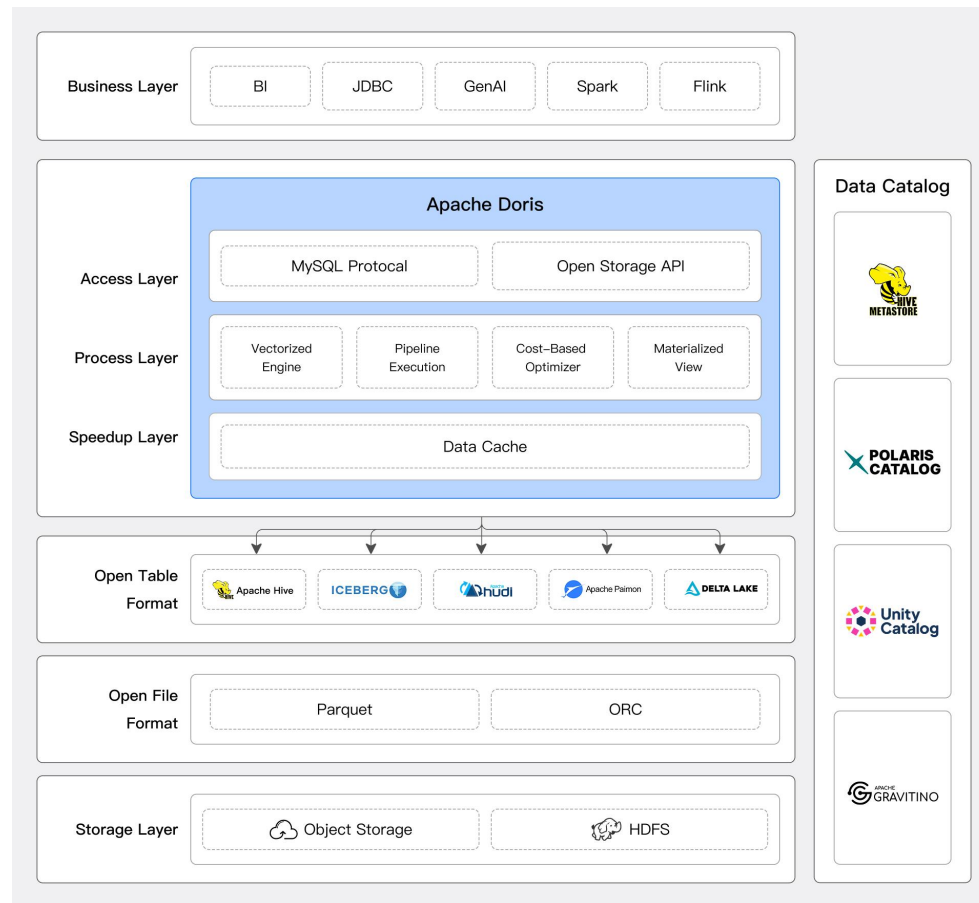
## 应用场景

大规模历史数据的实时仪表盘

批处理与流处理的统一分析

无需 ETL 的数据湖加速

跨数据源的联合查询



# 云原生存算分离架构

## SelectDB on AWS

应用系统 / 客户端

读写

读写

云服务层

实例1 ( Instance 1 )

实例2 ( Instance 2 )

计算集群层

集群1  
( Cluster 1 )

集群2  
( Cluster 2 )

集群3  
( Cluster 3 )

集群4  
( Cluster 4 )

集群5  
( Cluster 5 )

集群6  
( Cluster 6 )

计算(EC2)  
( vCPU , RAM )

计算(EC2)  
( vCPU , RAM )

计算(EC2)  
( vCPU , RAM )

计算(EC2)  
( vCPU , RAM )

计算(EC2)  
( vCPU , RAM )

计算(EC2)  
( vCPU , RAM )

缓存  
( Cache )

缓存  
( Cache )

缓存  
( Cache )

缓存  
( Cache )

缓存  
( Cache )

缓存  
( Cache )

读写

读写

读写

读写

读写

读写

共享存储层

对象存储 S3

对象存储 S3

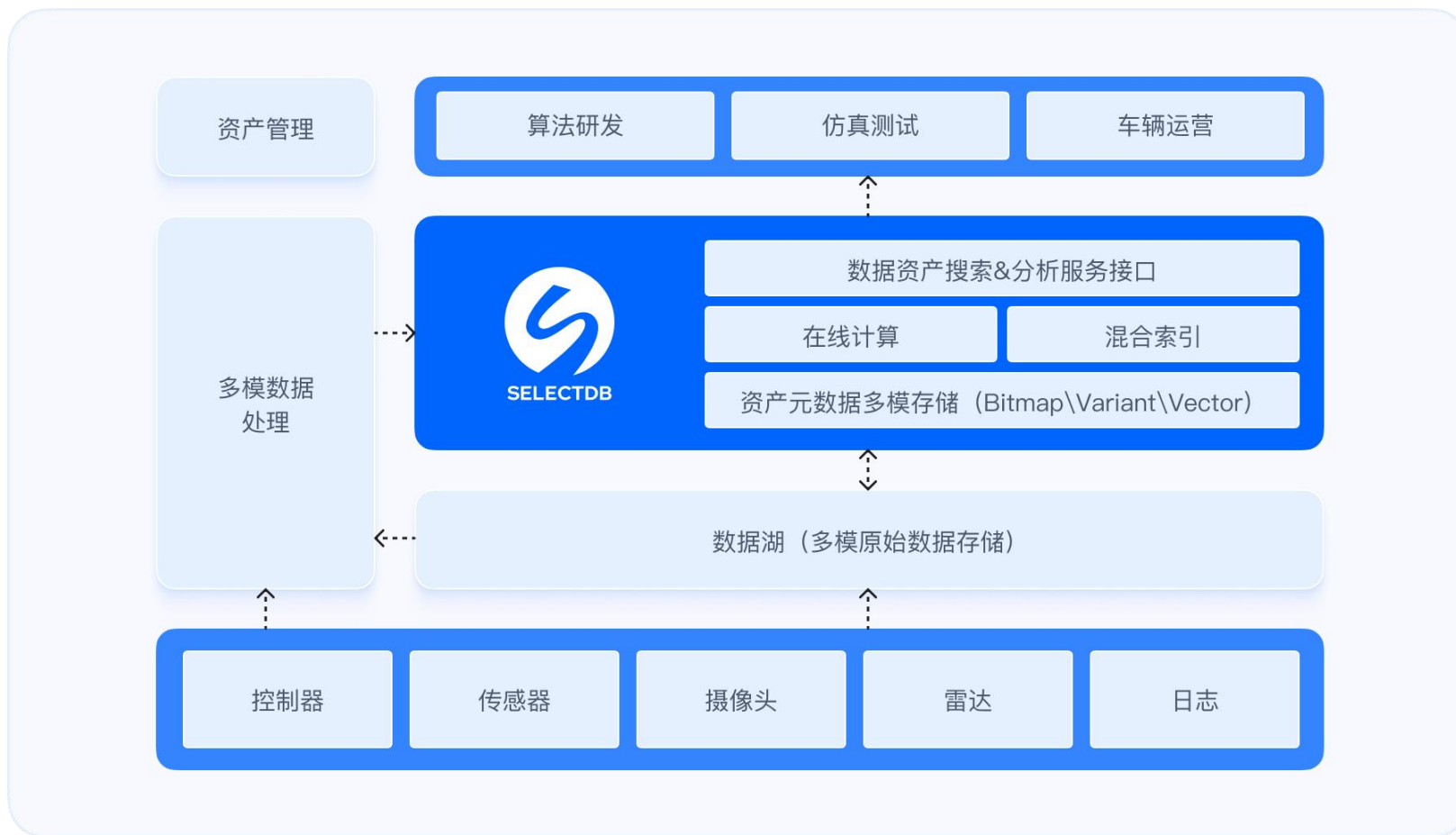
## 存算分离与弹性计算

- 共享对象存储与本地高速缓存
- 多计算集群，实现负载隔离
- 集群计算节点弹性扩缩容

## 极致性价比

- 存储冷热分离，冷数据使用低成本可靠对象存储
- 计算节点通过弹性扩缩容，减少计算资源浪费

# 基于 SelectDB 的新一代自动驾驶数据智能底座



## 统一实时分析底座

原生支持 Variant、无缝接入 JSON，实现“万列”超宽表管理，支持向量检索，替代 Hive、ES 等多套系统。

## 秒级场景挖掘

融合倒排索引与向量索引，支持“雨天+施工+行人闯入”等复杂条件与相似语义联合检索，百亿级数据秒级定位。

## 极致成本优化

热数据在 SelectDB 保证极致性能，高效列式压缩，冷数据湖存储可联邦查询，实现成本与效率最优解。

# 基于SelectDB构建智能座舱数据洞察与体验优化引擎



## 深入理解用户

支持秒级用户画像与分群，灵活处理交互事件，快速分析用户行为，为个性化服务提供精准依据。

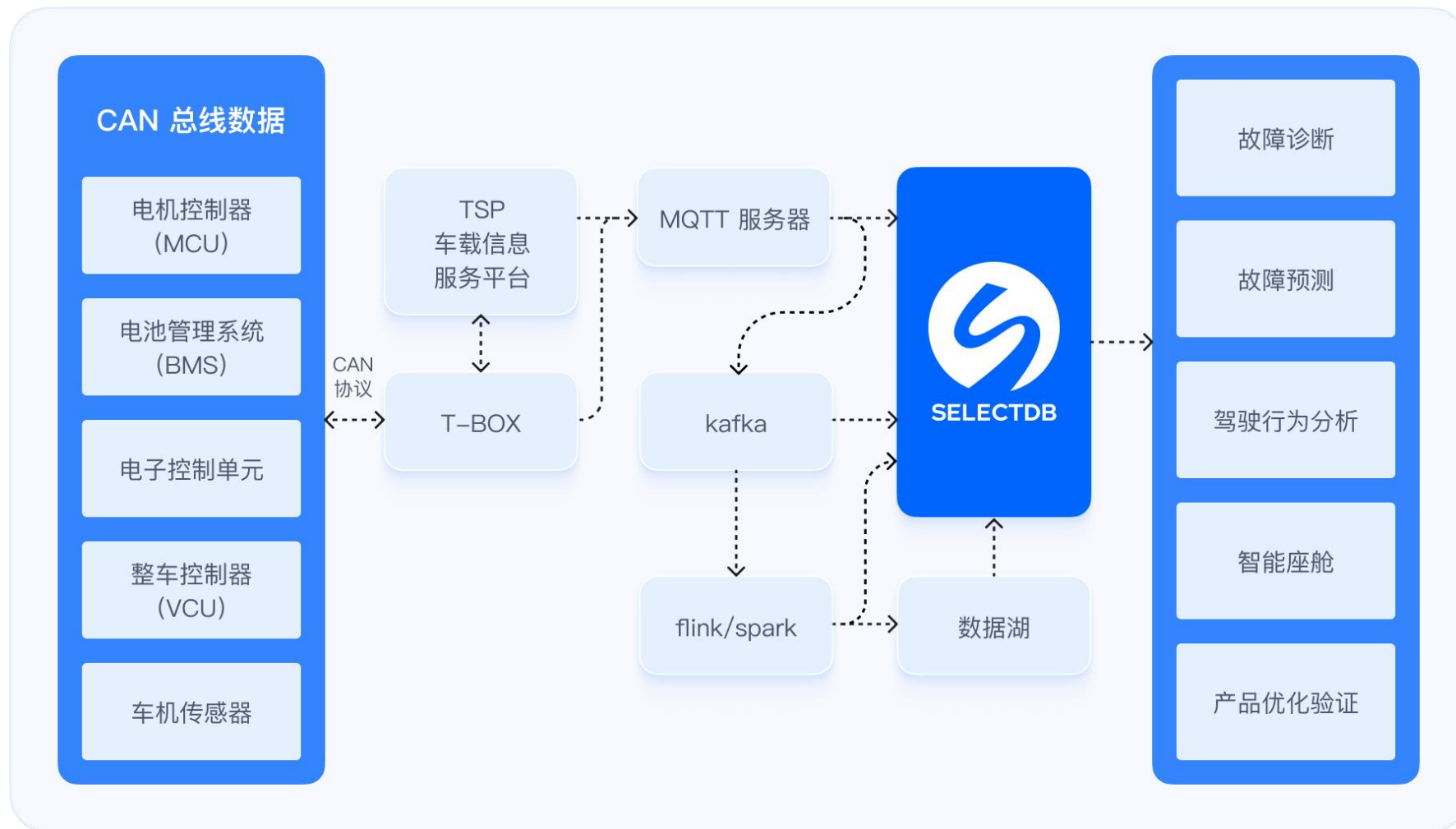
## 优化交互体验

实时分析语音识别准确率、触控响应延迟等指标，支持体验瓶颈发现，让座舱交互更自然流畅。

## 保障系统敏捷

统一存储用户行为、系统日志与标注数据，提供一致数据视图，将迭代周期从数周缩短至天级。

# 基于SelectDB构建新一代车联网实时数据中枢



## 秒级数据延迟

百万级 TPS 写入、秒级端到端延迟，让车辆监控、故障预警与行为分析实时可用。

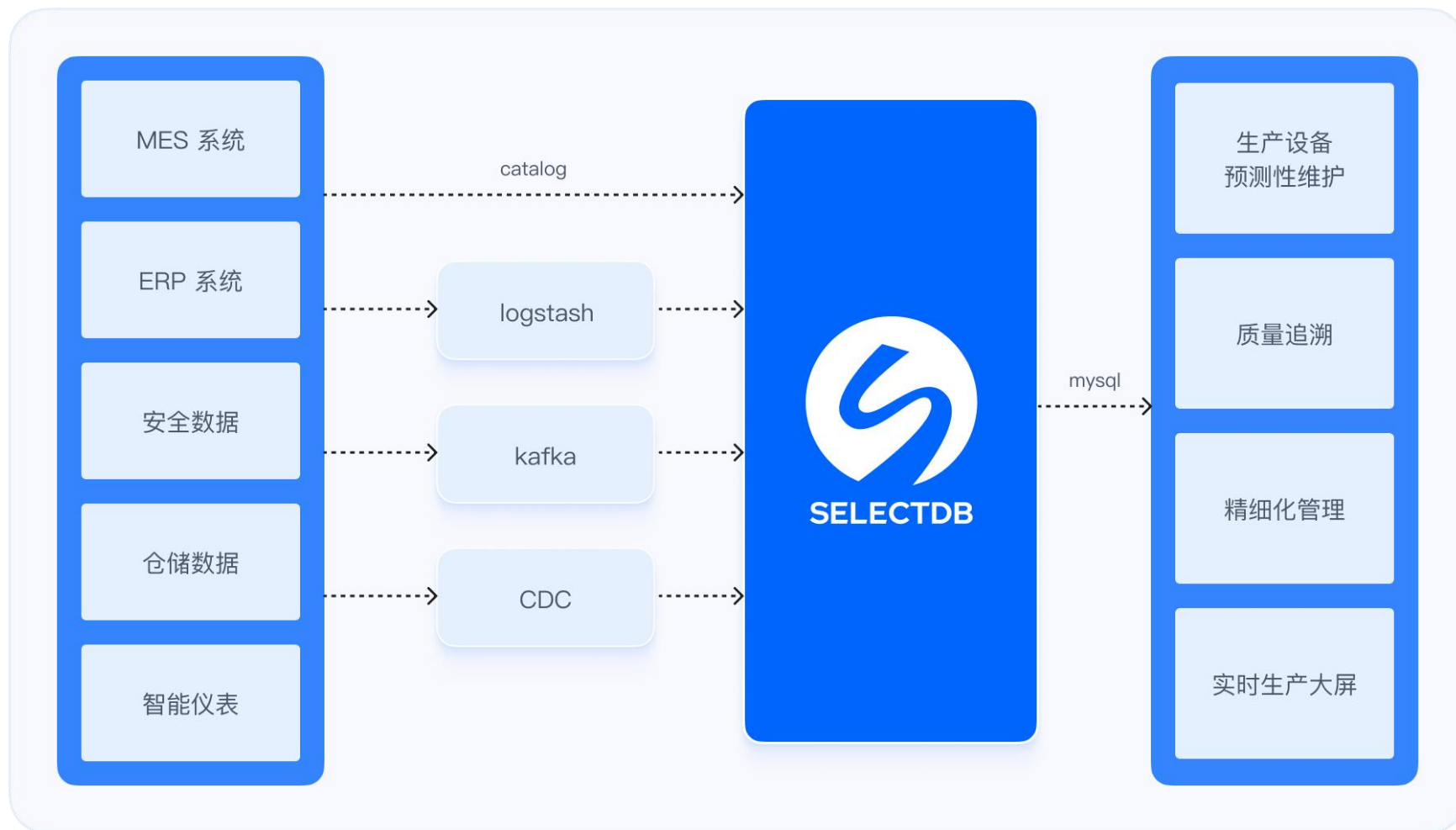
## 海量数据实时分析

在百亿级数据上实现秒级查询，为故障追溯、驾驶行为洞察与产品优化提供强力算力支撑。

## 降低总体成本

轻松应对上万列信号种类，5~20 倍数据压缩，显著降低存储、运维与整体技术成本。

# 基于SelectDB构建智能制造实时数据驱动引擎



## 统一实时分析底座

一体化架构同时支持实时与批量分析，支撑预测性维护、生产指挥与质量追溯等核心业务场景。

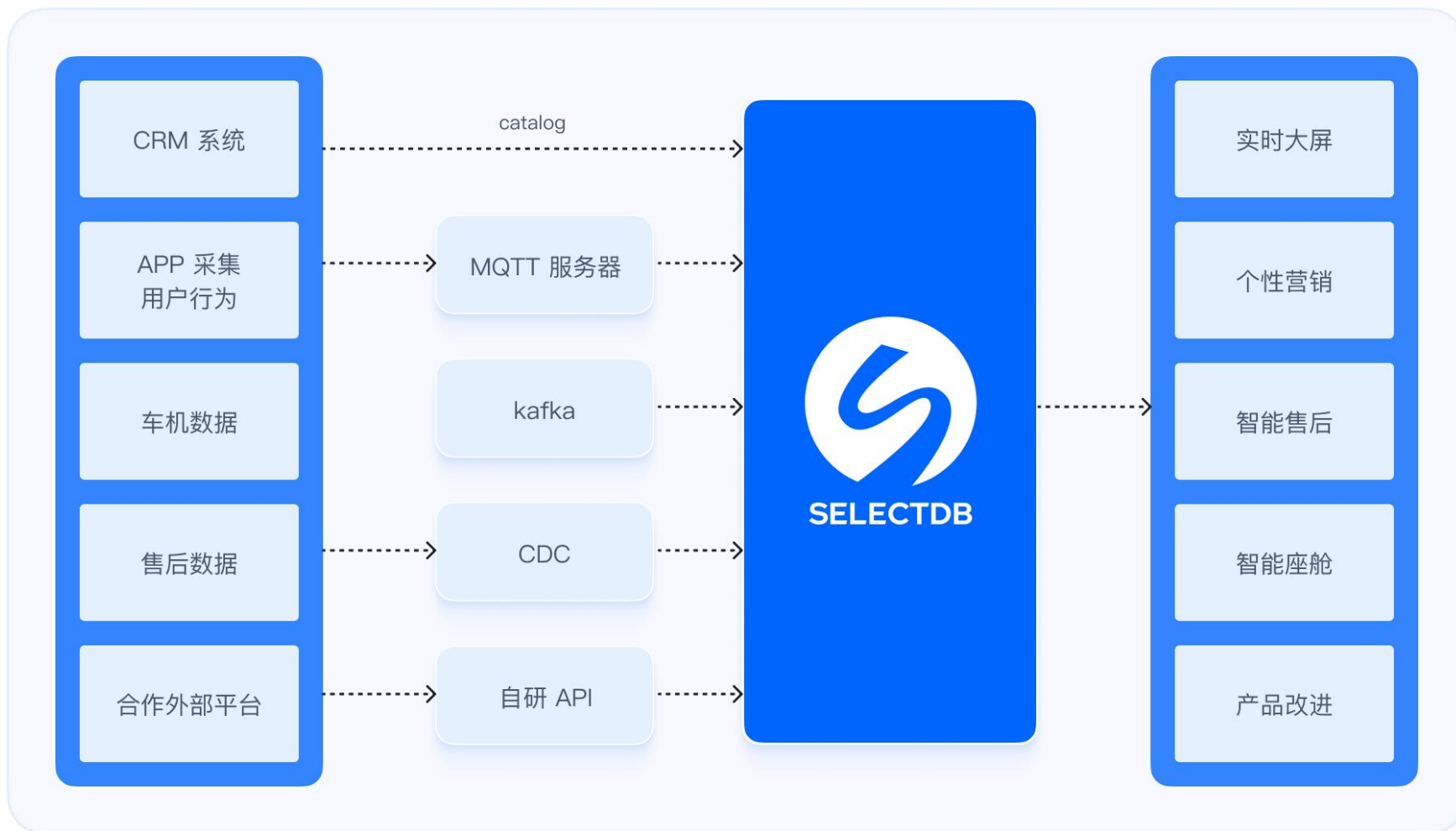
## 极致成本优化

IOT 数据可压缩 20 倍，存算分离节省 60% 存储成本，极大降低海量工业数据长期留存成本。

## 跨源关联分析加速

支持复杂多表关联与跨源联邦查询，轻松整合 MES、质量系统、供应链数据，实现端到端工厂可视化与决策智能化。

# 基于SelectDB构建车企业务运营智能数据引擎



## 实时 360° 用户画像

多源数据统一接入与查询，提供统一实时的用户画像服务，支持精准分群与高效触达。

## 实时运营与营销

GB/s 吞吐与秒级延迟，支持实时大屏、活动监控和快速策略调整，提高转化与运营效率。

## 灵活标签体系

毫秒级表结构变更与高性能分析，让画像、标签、行为模型可随业务快速演进，缩短运营试错和优化周期。



DORIS



SELECTDB

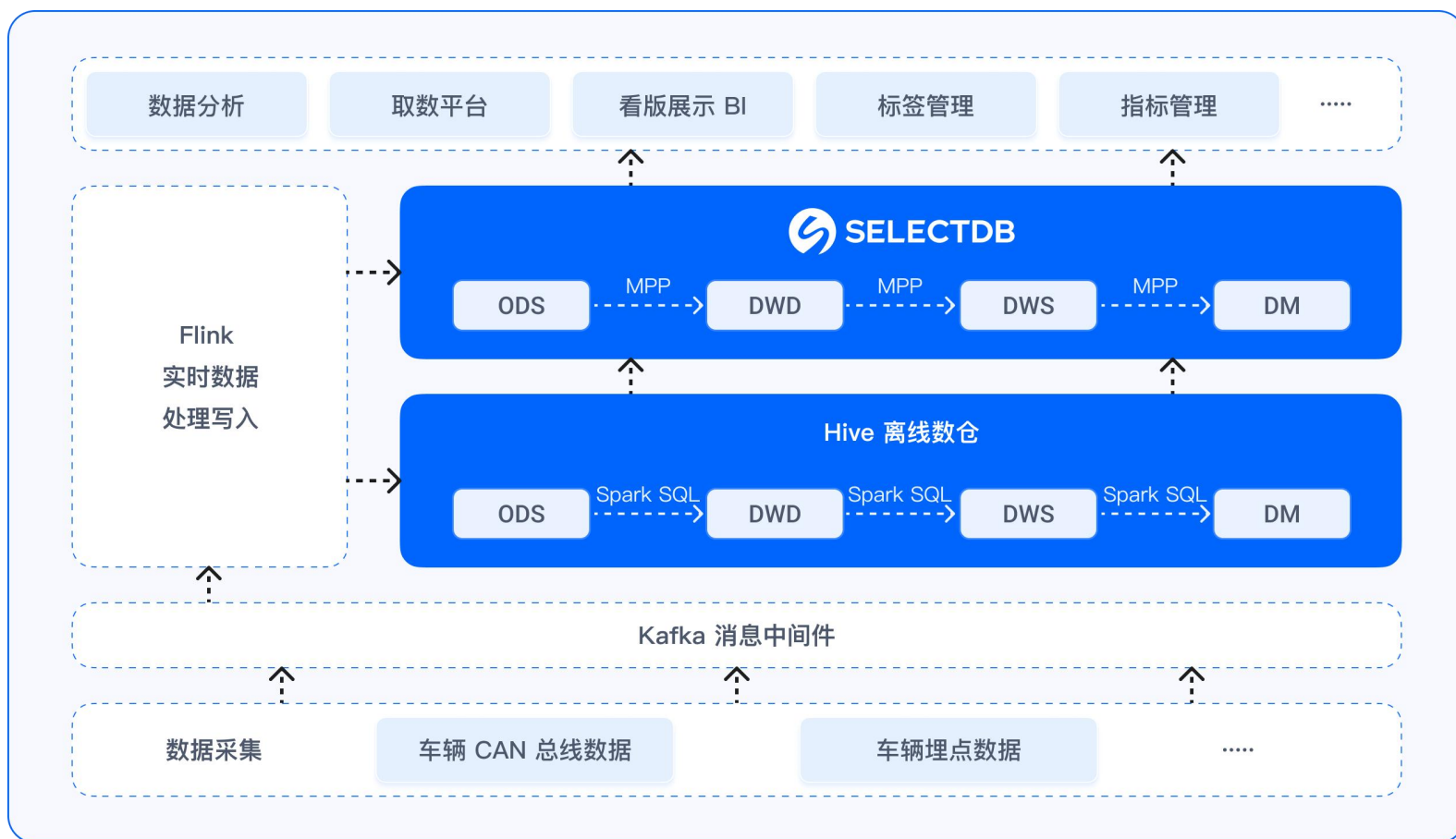
# 目录

数据与智能定义竞争力

SelectDB 加速汽车智能进化飞轮

头部企业实践案例

# 头部车企 - 基于 SelectDB 的车联网数据分析平台



## 业务背景

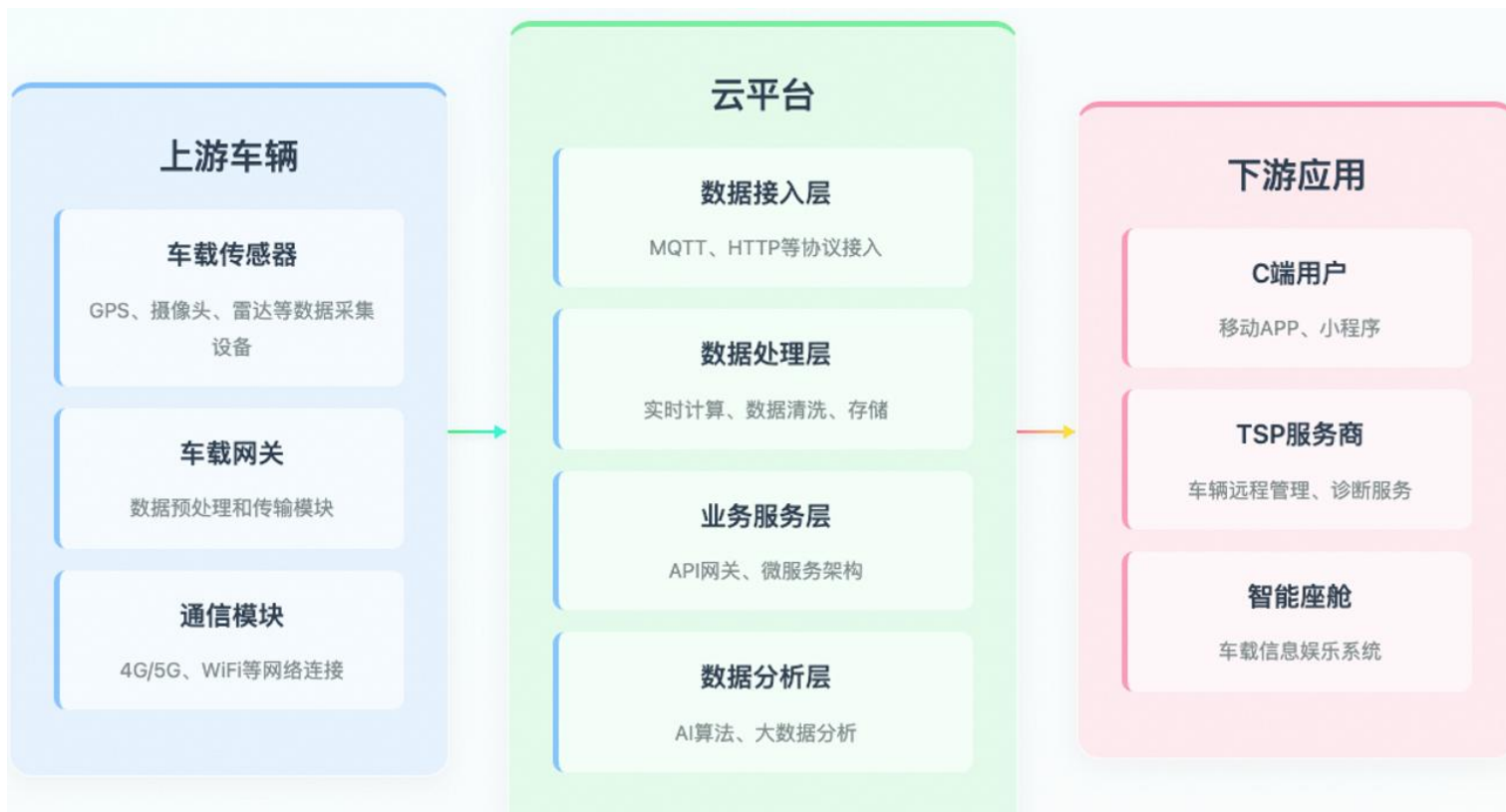
该汽车集团旗下智能化研究院是汽车智能化领域领军企业，需支撑**400万辆**车联网连接，**每日新增数十TB**数据。业务覆盖故障诊断、智慧能耗、智能运营等实时分析场景，要求低延迟查询、主键级更新及统一数据服务。

原架构基于Hive T+1任务，存在数据时效性不足、查询延迟高、组件复杂（Kafka / Flink/ Hive /Spark /MySQL）、存储成本高、无法关联外部配置表等问题，亟需升级。

## 基于SelectDB的方案收益

采用SelectDB统一实时处理与查询：**单日百亿级**数据实时写入，**十亿级查询秒级**响应，实现T+0数据新鲜度；利用ZSTD压缩**降低3-5倍存储成本**；通过Multi-Catalog统一查询Hive、MySQL等数据源，简化架构；兼容MySQL协议，提升开发运维效率；支持多字段宽表构建与跨源关联，满足多业务需求。

# 某新能源汽车 - 车联网云平台



## 业务背景

自研车联网云平台，需支撑**百万级在网车辆**，日处理海量传感器与CAN数据，实现车辆状态实时监控、故障预测、驾驶行为分析及产品迭代优化，为整车安全、用户体验和研发效率提供数据驱动力。原架构基于Impala，存在资源争抢严重、Join查询慢、端到端延迟达T+1、扩容需停机等问题，无法满足毫秒级实时分析需求。

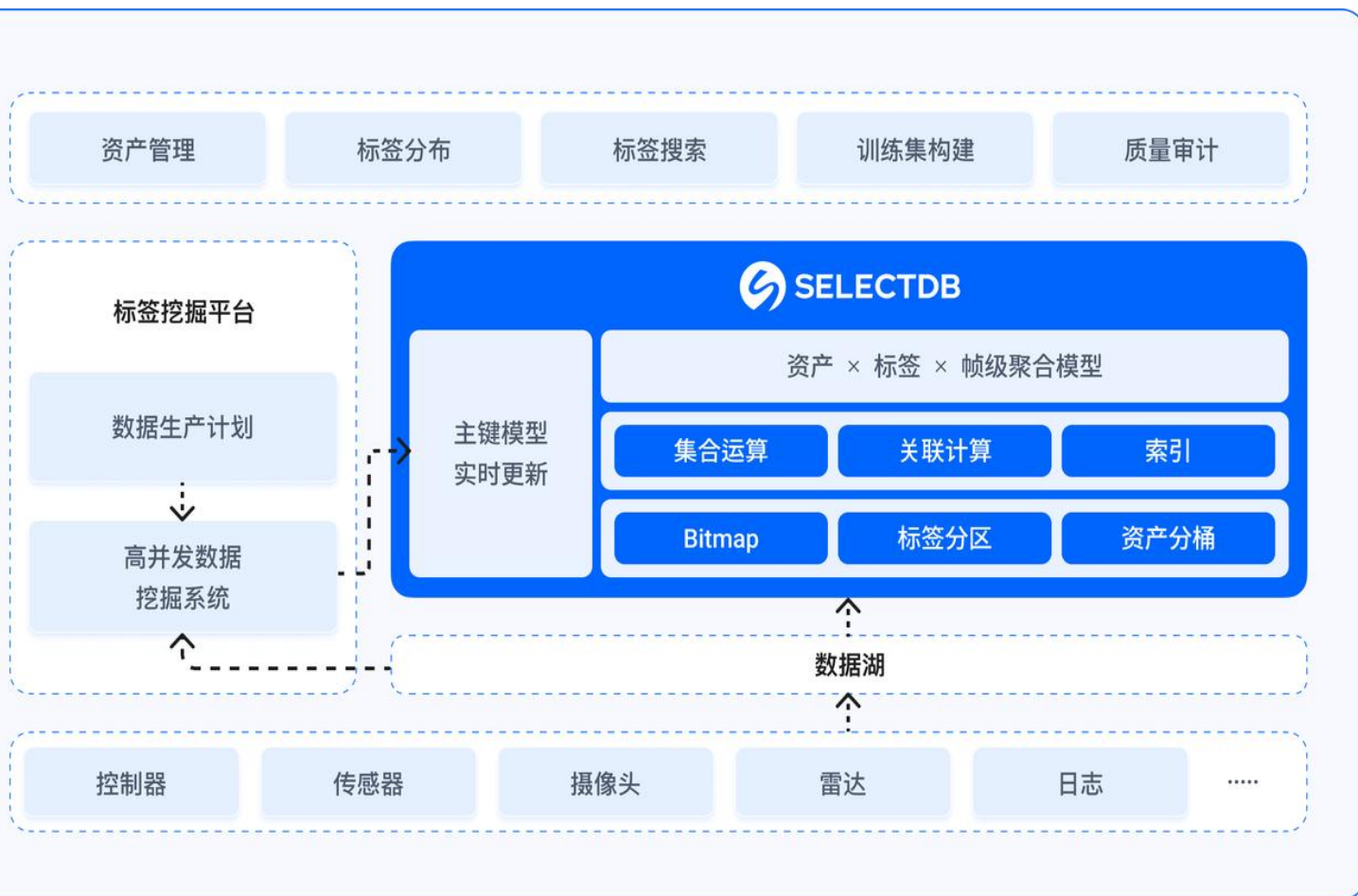
## 基于SelectDB的方案

引入SelectDB构建统一实时数据平台，利用其高性能写入与查询引擎，实现车联网数据实时接入与秒级分析；通过分布式架构和列式存储，保障高并发下的查询稳定性；简化数据处理链路，降低运维复杂度。

## 方案收益

智能驾驶大屏查询P99响应时间从**1.2秒优化至0.3秒**；实现端到端秒级数据实时处理与分析；系统零抖动，高并发下服务持续稳定。

# 某头部自动驾驶企业 - 基于 SelectDB 的实时标签检索



## 业务背景

训练数据资产平台面临数据规模爆炸（峰值**百万帧/秒**实时可见，检索规模**百亿级**）、复杂集合运算密集（bitmap AND/OR/差集、数百标签多维聚合）、严格实时交互要求（高并发下10秒内响应）等挑战，原有方案性能瓶颈制约训练效率提升。

## 基于SelectDB的统一标签分析与训练数据服务平台

采用**Bitmap**原生建模，将帧标签建模为压缩bitmap列，高性能进行集合运算；利用主键模型实现**高吞吐实时写入与秒级可见**；针对训练场景专项优化（按训练资产分桶、标签分区）；**联邦查询**能力统一访问湖仓历史数据。

## 方案收益

并发查询从分钟级降至**秒级**，算法工程师**实时探索**标签组合；数据选择从离线转为实时交互，大幅**缩短模型迭代周期**；统一引擎简化架构，降低运维与开发成本；横向扩展能力支持数据规模持续增长，为数据闭环提供长期基础设施保障，直接提升自动驾驶训练效率。

# THANKS



关注我们



在线咨询



联系我们