

Apache Doris 在客户精准营销管理方面应用

董生
某大型金融机构 数据架构师



目录

01 背景介绍

02 解决方案

03 实践应用

04 未来展望

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris

联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

背景介绍



金融营销转型的时代背景

- 以产品为中心的营销模式，用户体验差、营销成本高、转化率低。
- 以产品为中心向数据驱动，定制服务转换。理解、满足并超越顾客的期待



用户行为改变

- 更倾向于通过搜索引擎、社交媒体、电商平台等数字化渠道来获取信息和进行购物决策。



科技发展要求

- 提供极速高效准确的数据分析，改进客户体验中，增强品牌认同感、产品依赖感。



业务发展诉求

- 客户挖掘，产品创新，数字化运营、战略决策需要极速统一的数据支撑能力。
- 后置营销向前置营销、全域营销转变

业务诉求



精准客户管理

- 全域信息整合、实现真正的 KYC。知道客户从哪里来，知道客户长什么样，知道客户喜好什么
- 客户数据有效整合
- 构建客户精准画像
- 洞悉客户差分
- 客户关系维护



精准客户营销

- 依托各营销渠道的数据反馈。及时调整营销策略。知道客户在干什么，知道客户为什么流水，知道如何挽回客户。
- 为客户在复杂性、高合规要求和多变的售后情况，选择推荐合适产品。
- 量化、实时对营销活动的效果进行实时监测和评估，以便及时调整和优化营销策略。
- 快速数据分析决策，快速策略，匹配场景，实现交易有效触达。

目录

01 业务背景

02 解决方案

03 实践应用

04 未来展望

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

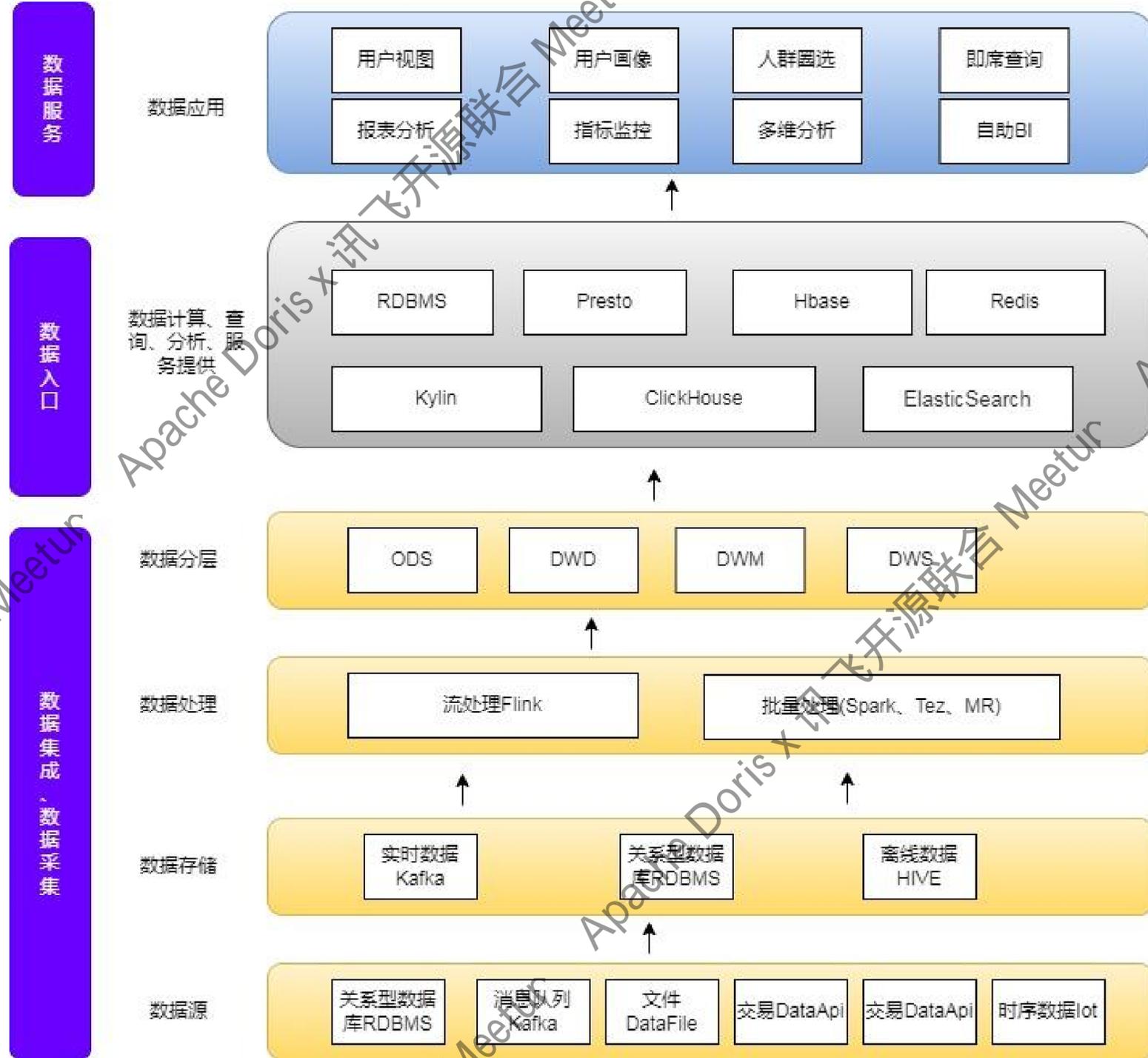
Apache Doris

联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

引入 Doris 之前的数据架构



原架构面临的痛点



研发运维成本高

多种数据处理架构、开发和运维成本较高



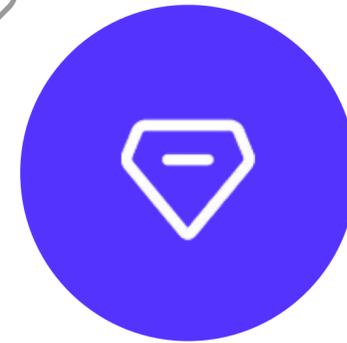
数据时效性不一致

数据库中报表和分析数据是 T+1 的数据，与即时业务数据不符



查询效率低

业务量增长，分库分表依旧无法支撑高并发查询，响应时间超过预期



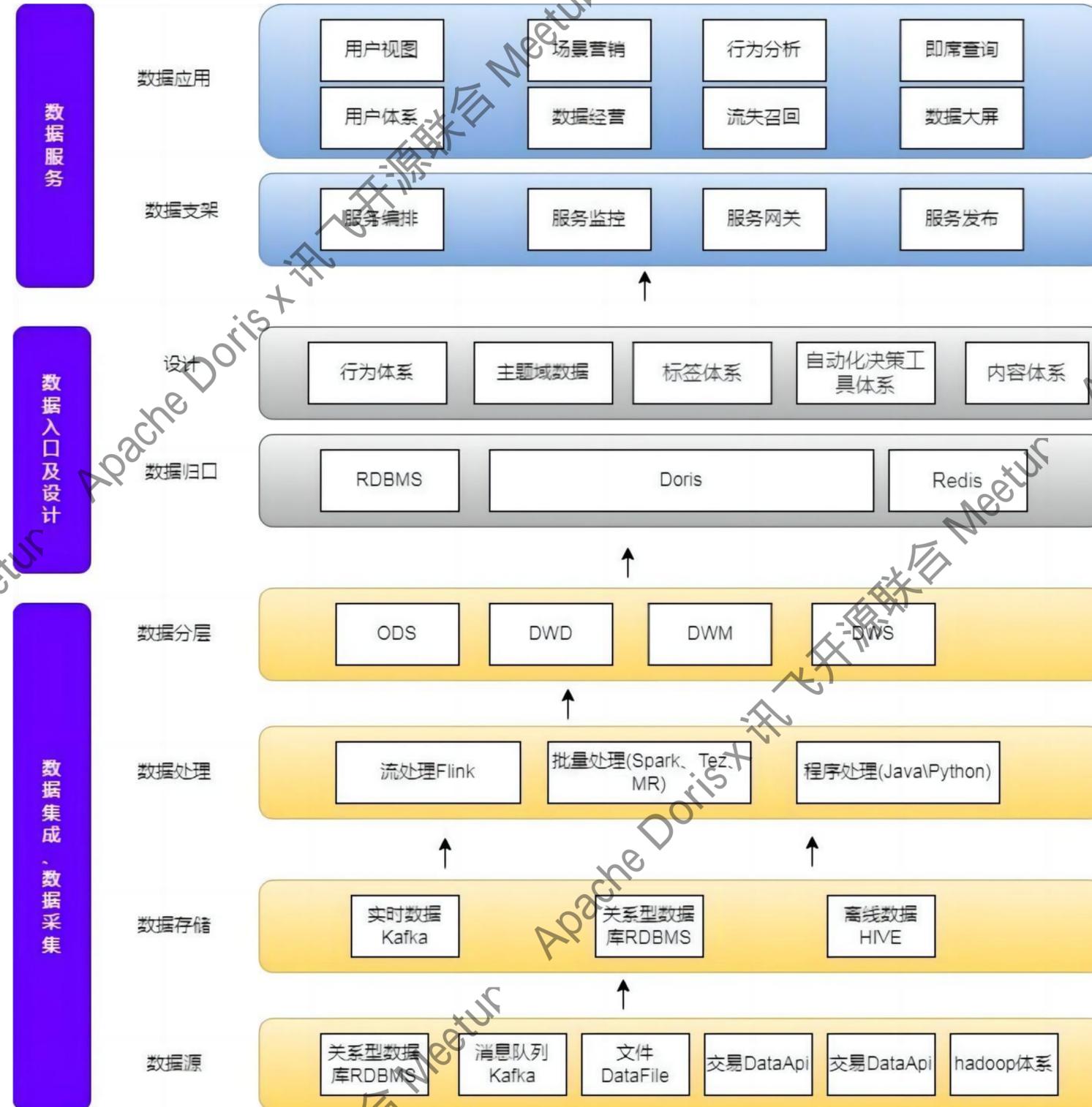
数据分析能力差

OLAP 分析能力较差，无法即时支撑复杂业务，如留存、漏斗、实时标签等

选型因素

关键指标	Apache Doris	现有工具
查询性能	基准测试的单表、宽表、AP 分析性能，满足预期要求。 多表 JOIN 能力强	无法同时兼容 TP 与 AP 的数据查询性能。 JOIN 支持能力十分有限
运维部署	FE/BE 自动扩缩容简单，支持数据自动平衡	依赖第三方软件，扩缩容困难
研发成本	标准 SQL 协议，成本较低	不同引擎之间协议不同，成本较高
数据融合	支持多数据源接入	需要独立开发

引入 Doris 之后的数据架构



目录

01 背景介绍

02 解决方案

03 实践应用

04 未来规划

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

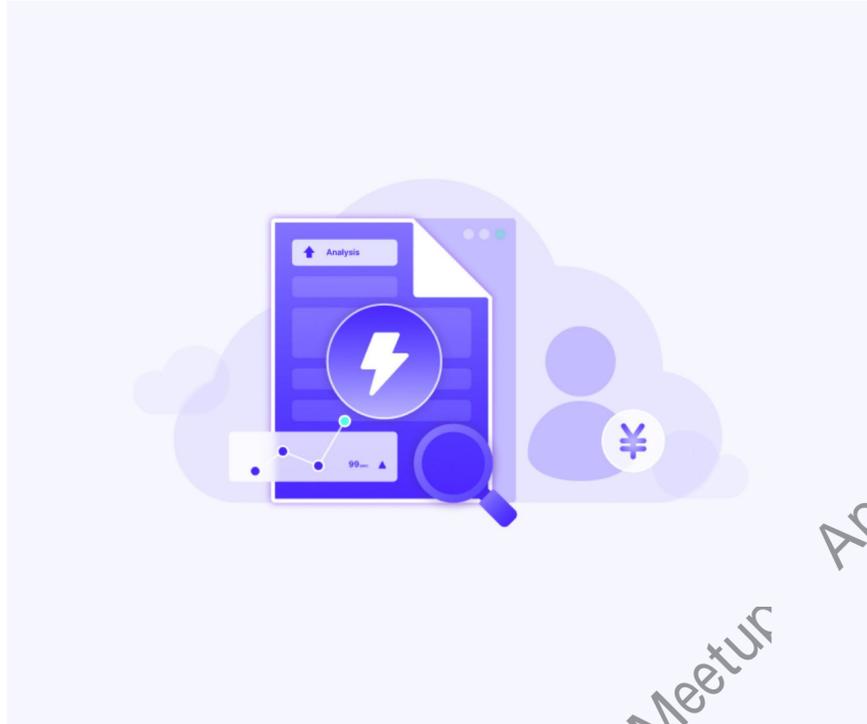
Apache Doris

联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

用户画像



用户明细数据

使用 Unique Key 模型。
按照用户 ID 分桶，每个桶数据量
0.5GB 左右



用户标签

Bitmap 应用



用户行为数据

Aggreagte Key 模型。
通过 `replace_if_not_null` 的方式，
实现部分列更新

用户明细数据

宽表方案:

适用于固定场景, 不存在多表关联, 后续增加属性字段困难。

高表方案:

按照客户相对集中的数据将数据分类拆分存储到不同的 Unique Key 模型, 所有的表的 UNIQUE KEY 都是用户 ID;

按照用户 ID 分桶, 每个桶大小为 0.5G 左右, 便于后续即席查询。

• 查询优化

开启短路径查询

```
"enable_unique_key_merge_on_write"="true";  
"light_schema_change"="true";  
"store_row_column"="true";
```

是否开启行缓存, 默认不开启

```
disable_storage_row_cache = false
```

指定row cache占用内存的百分比, 默认20%

```
row_cache_mem_limit = 40
```

• 分区分桶裁剪



用户标签

标签变动频繁，需要支持实时查询与更新

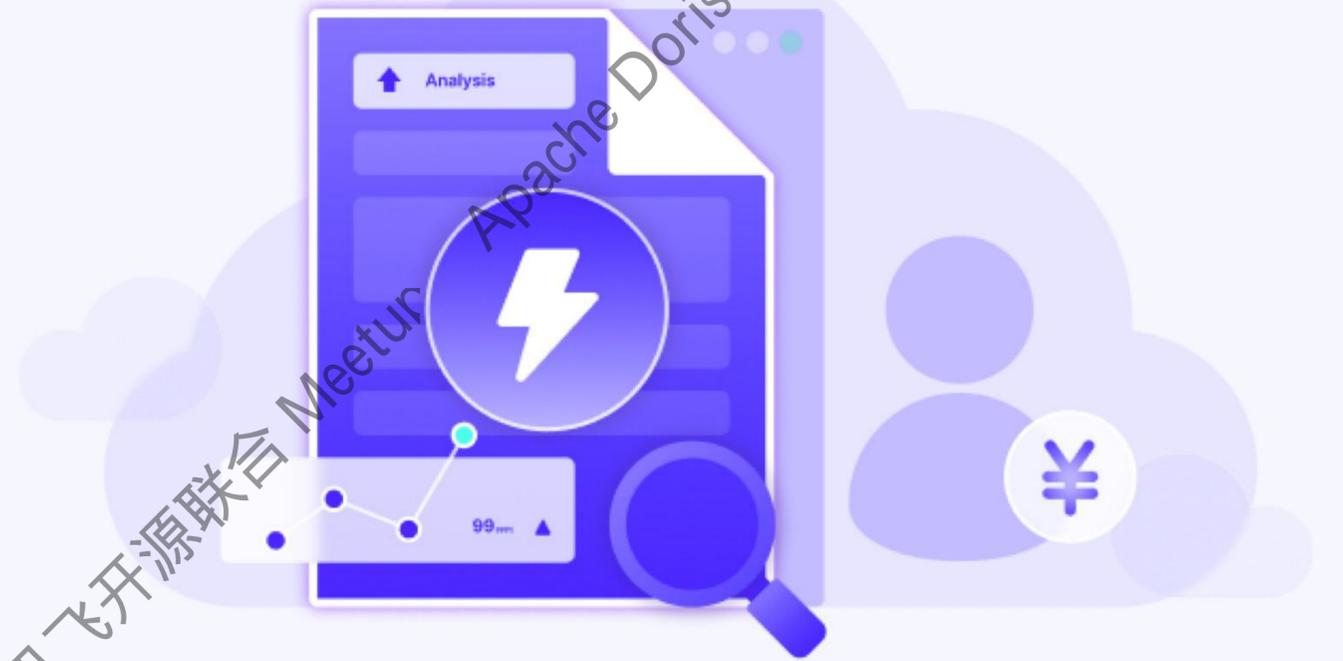
用户群体之间多类标签联合查询

宽表 tag1..N 存储：查询速度慢；占用存储空间；无法支撑高并发。

标签说明表 tag 客户标签 values 标签值

bitmap 存储客户标签

- 新增用户标签：bitmap_or
- 清除用户标签：bitmap_remove
- 统计用户持有标签数量：bitmap_count
- 查询共有标签用户：bitmap_and
- 查询共无标签用户：bitmap_andnot
- 查询共有标签数量：bitmap_and_count
- 查询客户标签列表：bitmap_from_string



bitmap与宽表对比

对比类型	宽表	bitmap
存储占用	60G	10G
查询耗时	10S 以上	1S 以内
并发度 (500)	高并发情况下超时 (15S)	与单次查询相差不多

在数亿用户规模、每位用户存在上千个标签的场景下，使用 Bitmap 索引来查询用户相关数据后，查询时间均能够在 1 秒内响应

用户行为数据

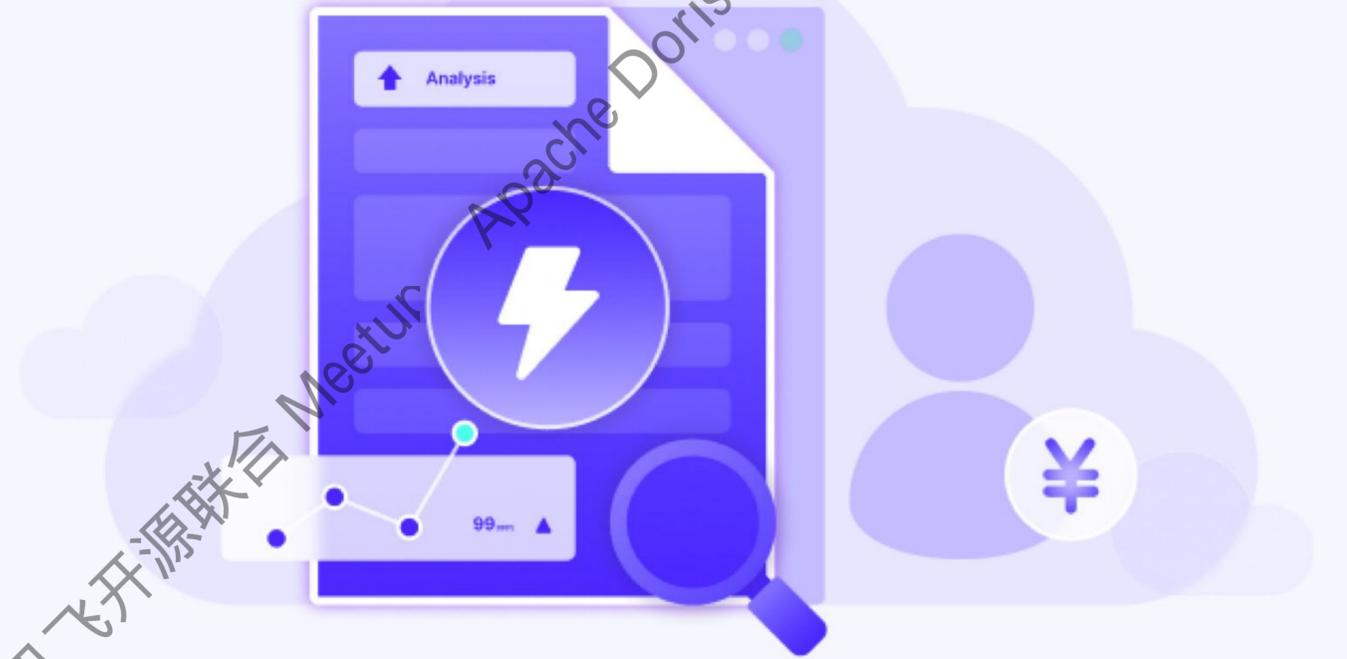
用户行为数据实时采集

特点：采集渠道多，采集属性多，无法全部采集行为

业务系统埋点，先采集到 Kafka 中，由 Flink Doris Connector 写入到 Doris 内。

多流 Join，使用 `replace_if_not_null` 方式实现部分列更新。让不同采集的数据如同一份行为结构。

Exactly-once 精准一次。



用户行为分析



留存分析

retention_info

漏斗分析

window_funnel

路径分析

ROW_NUMBER、LEAD、LAG

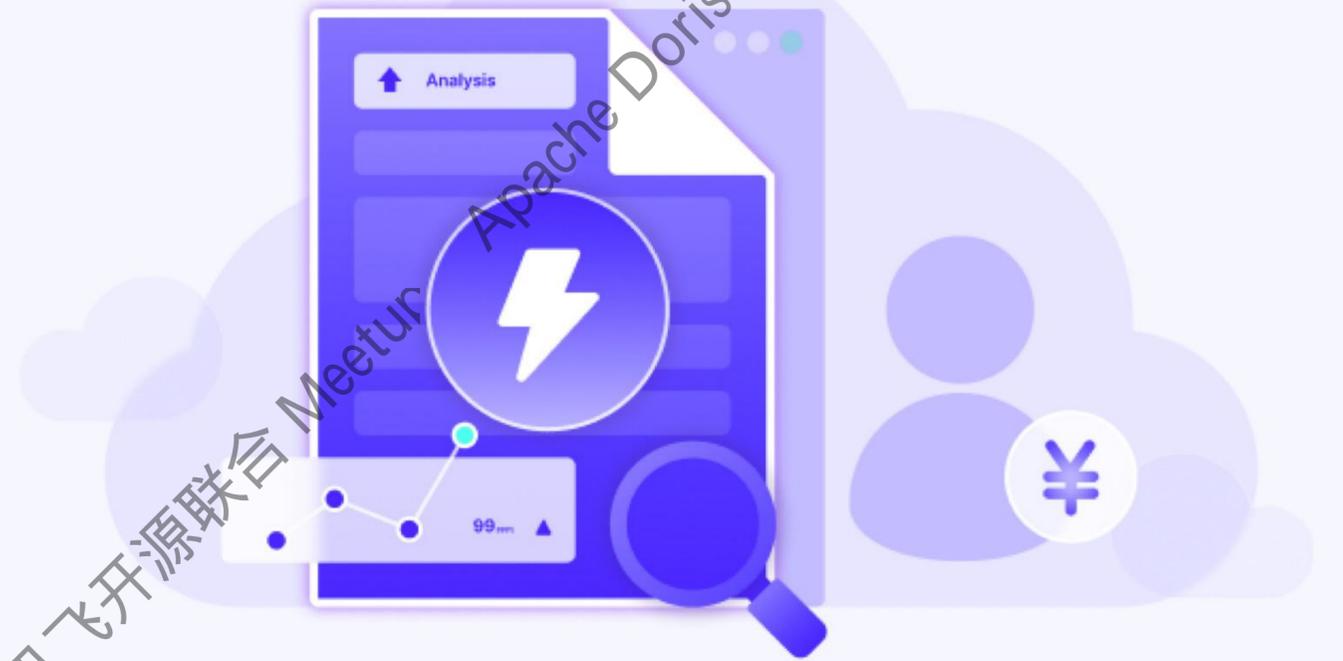
留存分析

用户留存的分析，经常会在多个条件下获取用户的行为状态 retention。

指定多个事件条件，这些条件成对使用。例如，如果第一个和第二个事件都为真，则第二个结果将为真，如果第一个和第三个事件都为真，则第三个结果将为真，以此类推。

场景 1：用于分析用户在一段时间内，是否登录访问过本单位的产品的

场景 2：判断用户联系购买、浏览



漏斗分析

针对用户的转化分析场景，经常需要分析在一定的时间窗口中，用户在一系列连续行为下的相关行为。

从事件链中的第一个条件开始判断。

- 如果数据中包含符合条件的事件，则向计数器加 1，并以此事件对应的时间作为滑动窗口的起始时间。
- 如果未能找到符合第一个条件的数据，则返回为 0。在滑动窗口内，如果事件链中的事件按顺序发生，则计数器递增；如果超出了时间窗口，则计数器不再增加。

```
SELECT
  user_id,
  window_funnel(
    300, // 时间窗口 5 分钟
    "default",
    event_timestamp,
    event_name = '登录',
    event_name = '访问',
    event_name = '下单',
    event_name = '付款'
  ) AS level
FROM events
GROUP BY user_id
```

路径分析

ROW_NUMBER 为每一行唯一排序

场景：为用户每一次行为分配一个唯一的序号

LEAD 问当前行之后的下一行数据。比较当前行与紧接着的下一行的数据，

场景：根据历史预测用户的下一行为；查看哪一步骤用户流失或者增加最多

LAG 访问当前行之前的上一行数据。比较当前行与上一行的数据。

场景：了解用户在采取当前行动前一般会采取什么行动，用于挽回流失用户；发现用户不合常规异常行为。



使用 Apache Doris 后价值提升

DataApi查询

30_{MS}

QPS

1.5_万

报表

300_张

管理用户

数_{亿+}

即席报表时间

6_S

日均查询量

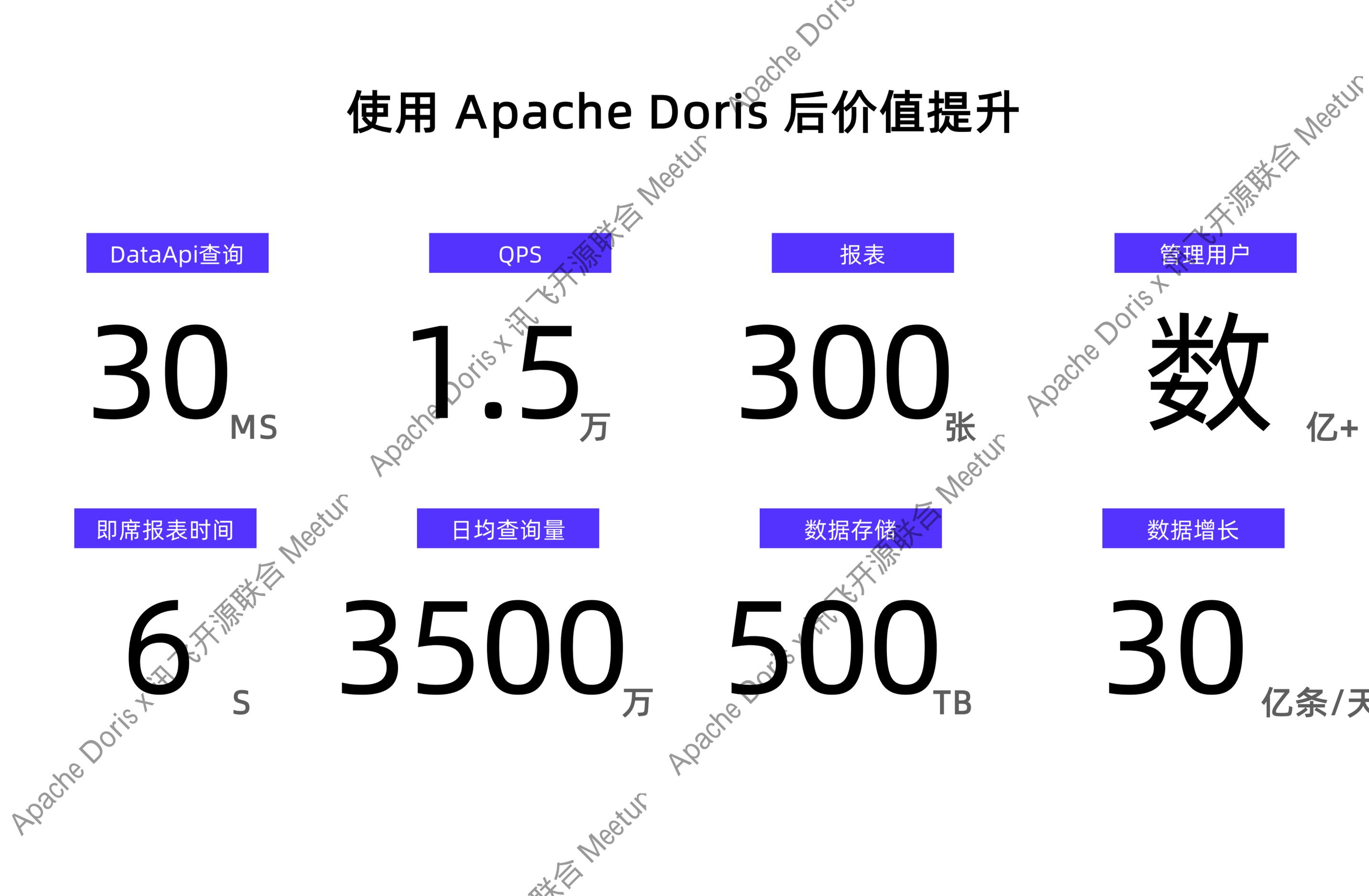
3500_万

数据存储

500_{TB}

数据增长

30_{亿条/天}



目录

01 背景介绍

02 解决方案

03 实践应用

04 未来展望

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris

联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

展望未来

场景优化

深入Doris的研究，特别是OLAP的研究，适应更多的客户管理、营销分析的场景

湖仓一体

跟进最新的技术，探寻基于Doris的湖仓一体方面的应用

存算分离

跟进最新的存算分离，降本增效

与AI结合

text-to-sql 让业务管理用户和分析营销更加简单。提升自助营销分析决策能力



Thanks !



Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris

联合 Meetur

Apache Doris x 讯飞开源联合 Meetur

Apache Doris x 讯飞开源联合 Meetur