



| DataFun.

# DORIS在蔚来中的应用

---

唐怀东 数据团队负责人



# 目录 CONTENT

**01** OLAP在蔚来的发展

**03** Doris在运营平台上的实践

**02** Doris作为统一OLAP数仓

**04** 经验总结

# 01

## OLAP在蔚来年的发展



## OLAP在蔚来的发展



2017年引入Druid



2021年引入Doris

2019年引入TiDB



# 为什么选择Doris

OLAP	优点	缺点	技术使用成本	协议	性能	运维成本
Druid	<ul style="list-style-type: none"><li>实时+离线数据摄入</li><li>列存储</li><li>高并发</li></ul>	<ul style="list-style-type: none"><li>使用门槛高</li><li>基本上不支持Join</li><li>unique操作效率低</li></ul>	<ul style="list-style-type: none"><li>高</li></ul>	<ul style="list-style-type: none"><li>非标准协议</li></ul>	<ul style="list-style-type: none"><li>分场景</li></ul>	<ul style="list-style-type: none"><li>高</li></ul>
TiDB	<ul style="list-style-type: none"><li>HTAP(OLTP + OLAP)</li><li>支持更新</li><li>支持明细和聚合</li></ul>	<ul style="list-style-type: none"><li>列式存储需要单独占存储和计算资源</li><li>OLAP能力不足</li></ul>	<ul style="list-style-type: none"><li>低</li></ul>	<ul style="list-style-type: none"><li>SQL标准</li></ul>	<ul style="list-style-type: none"><li>分场景</li></ul>	<ul style="list-style-type: none"><li>低</li></ul>
Doris	<ul style="list-style-type: none"><li>高并发和高吞吐</li><li>实时+离线数据摄入</li><li>支持明细和聚合</li><li>支持一定程度更新</li><li>物化视图</li><li>多表JOIN</li><li>schema在线变更</li><li>向量化引擎</li></ul>	<ul style="list-style-type: none"><li>成熟度有待进一步提高</li><li>应用广泛程度不够</li></ul>	<ul style="list-style-type: none"><li>低</li></ul>	<ul style="list-style-type: none"><li>兼容MySQL协议</li></ul>	<ul style="list-style-type: none"><li>满足</li></ul>	<ul style="list-style-type: none"><li>低</li></ul>
ClickHouse	<ul style="list-style-type: none"><li>列存储</li><li>单机性能强</li><li>向量化引擎</li><li>保留明细</li></ul>	<ul style="list-style-type: none"><li>多表JOIN支持差</li><li>并发度低</li><li>实时更新支持一般</li></ul>	<ul style="list-style-type: none"><li>高</li></ul>	<ul style="list-style-type: none"><li>非标准协议</li></ul>	<ul style="list-style-type: none"><li>满足</li></ul>	<ul style="list-style-type: none"><li>高</li></ul>

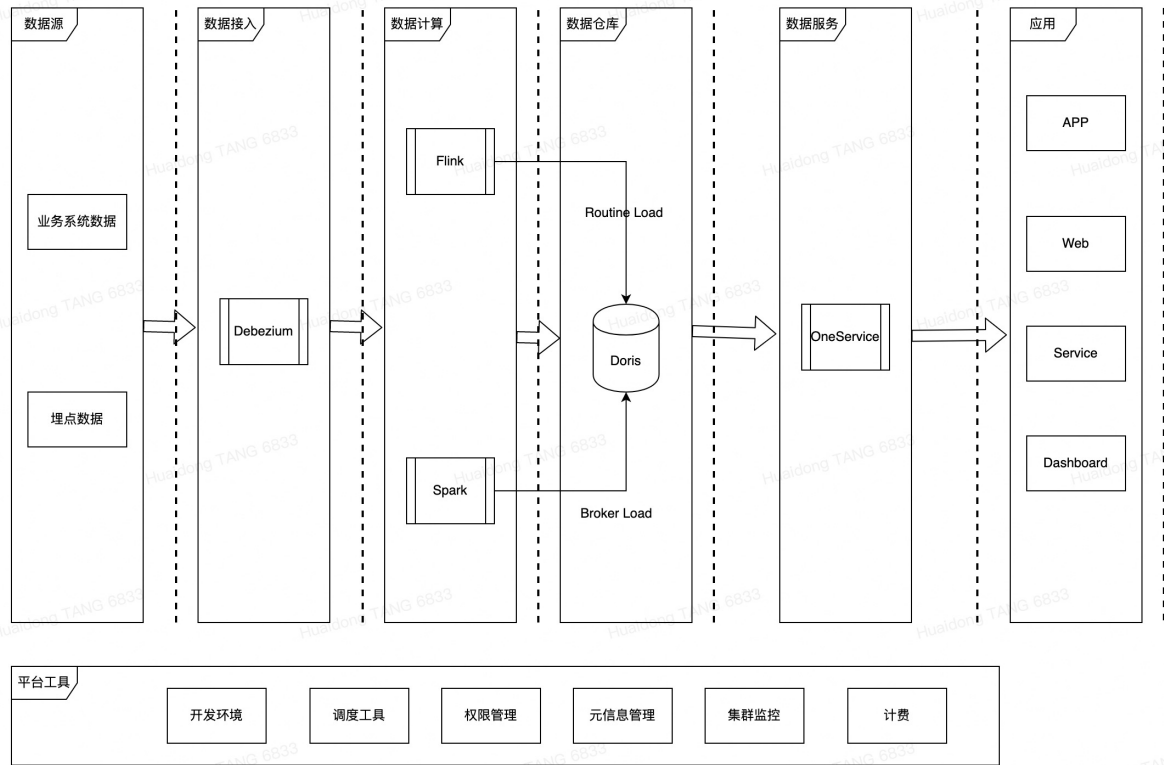


# 02

## Doris作为统一OLAP 数仓



# 统一OLAP数仓



# 03

## Doris在运营平台上的 实践





# 架构图



## CDP存储选型的考量点



实时离线统一

高效圈选

高效聚合

多表关联

联邦查询

## 离线实时统一

场景	需求	Doris功能点
实时标签	数据的实时更新	Routine load
离线标签	高效的大批量导入	Broker load
流批统一	实时离线数据存储统一	Routine load和broker load更新同一张表的不同列

## 高效圈选

场景	需求	Doris功能点
复杂条件圈选	高效的支持多条件圈选	SIMD优化

## 高效聚合

场景	需求	Doris功能点
标签值的分布	每天都需要更新所有标签的分布值，需要快速高效统计	<ul style="list-style-type: none"><li>• 数据分片(Partition)，减少数据传输和计算</li><li>• 存算统一，每个节点先聚合</li><li>• SIMD提速</li></ul>
群体的分布	同上	
效果分析的统计值	同上	

## 多表关联

场景	需求	Doris功能点
群体的特征分布	统计群体在某个特征下的分布	多表关联
单体的标签	展示不同单体的标签	

## 联邦查询

场景	需求	Doris功能点
效果分析关联任务执行明细	Doris数据关联Tidb数据	关联外表进行查询
人群标签关联行为聚合数据	Doris数据关联Elasticsearch数据	

# 04

## 经验总结





## 经验总结



Bitmap



ES外表



分批更新列



在线服务

## Bitmap聚合(计算UV场景)

- Roaring bitmap
- Id全集小于1000万，bitmap聚合和直接使用join，存储和查询效率差别不大
- Id全集大于5000万，可以考虑bitmap聚合

## ES外表

- 适合单表查询，支持下推
- 不适合聚合查询，会把ES明细全读到doris，然后聚合
- 不合适多表关联，读取大量数据，会引发ES频繁GC

## 分批更新列

- 聚合模型单独更新某些列，需要使用REPLACE\_IF\_NOT\_NULL
- 如果需要用null替换原来的非null值，则做不到
- 可以将所有null值替换成有意义的默认值，比如unknown

## 在线服务

- 同一份数据，同时服务在线离线场景，容易互相影响
- 尽量用另外单独的技术方案提供在线服务

# 非常感谢您的观看

---

