Joyy | DataFun.

基于FLINK+HOLOGRES的实时多维在降本提效的实践

石强　欢聚时代\大数据架构师

公司介绍

欢聚时代 于2005 年 4 月成立，是一家全球领先的社交媒体企业。

欢聚旗下运营多款社交娱乐产品，包括Bigo Live直播、Likee短视频、Hago休闲小游戏社交、即时通讯，电商业务等。

广告


电商


游戏


社交


短视频

目录

Joyy | DataFun.

01 为什么要实时多维化？

现存的问题

两套计算架构，成本高

需要实现离线计算逻辑
+实时计算逻辑

单独的kafka集群，zk集群
4套redis判重，
主从mysql架构

架构复杂，可维护性差

前端为PHP语言，
后端为java，
流式计算层为storm，
缓存为
后端存储为mysql

实时准确性没100%保证

指标在流中做计算，
数据延迟会导致指标不准确

数据修复难度大

明细数据存储在kafka,kafka不支持修改和删除等操作，保存时间短，一旦数据异常，需要人工介入，修正难度大。

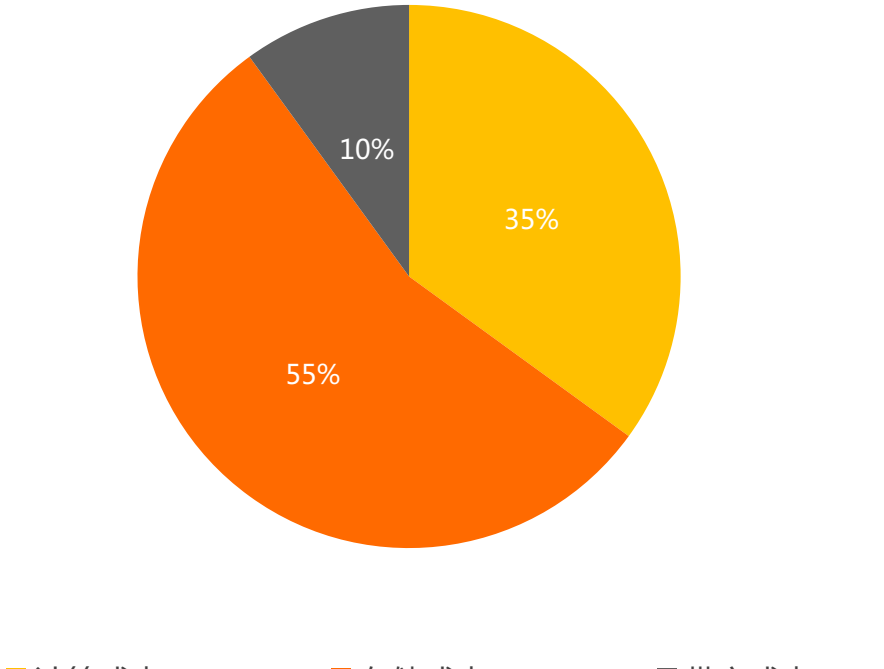
问题排查难

Kafka是append模式，不支持索引等SQL查询，增加了排查问题的复杂度。

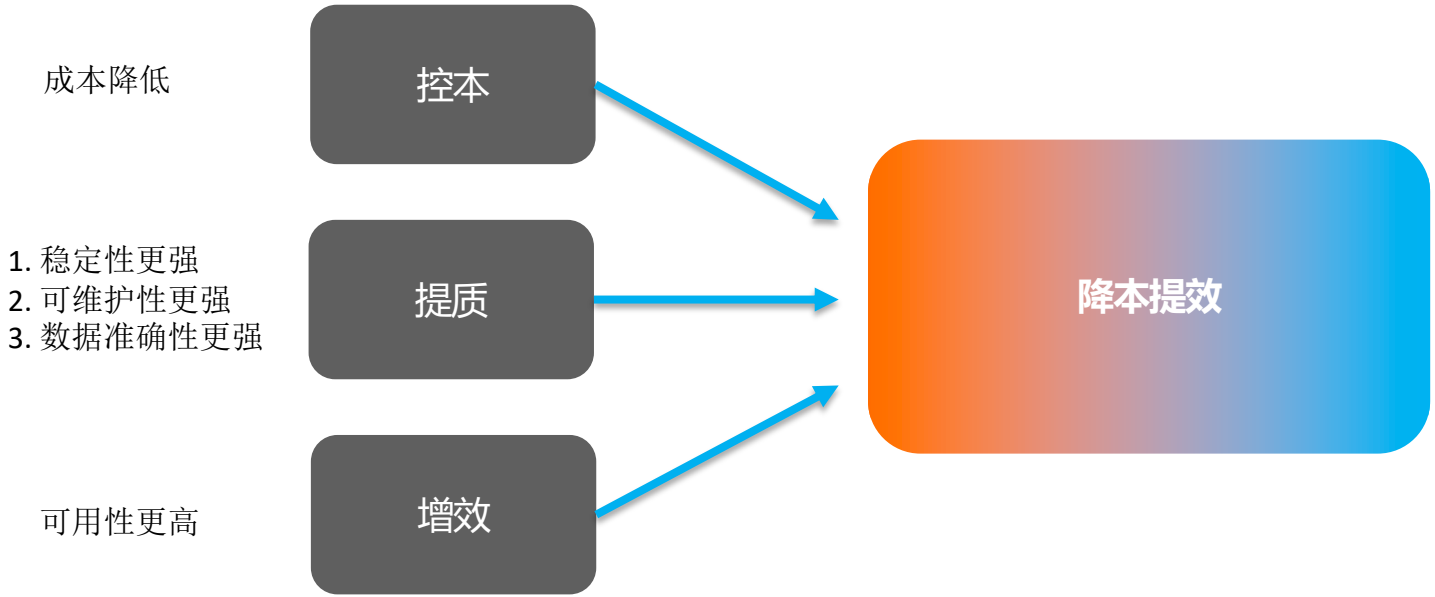
指标固化，资源新增复杂

指标是固化，新增指标需要重新开发
实时资源新增，需要新增storm节点
离线资源新增，需要新增nodemanager节点

成本的构成



数据团队目标



新方案应该具备的能力

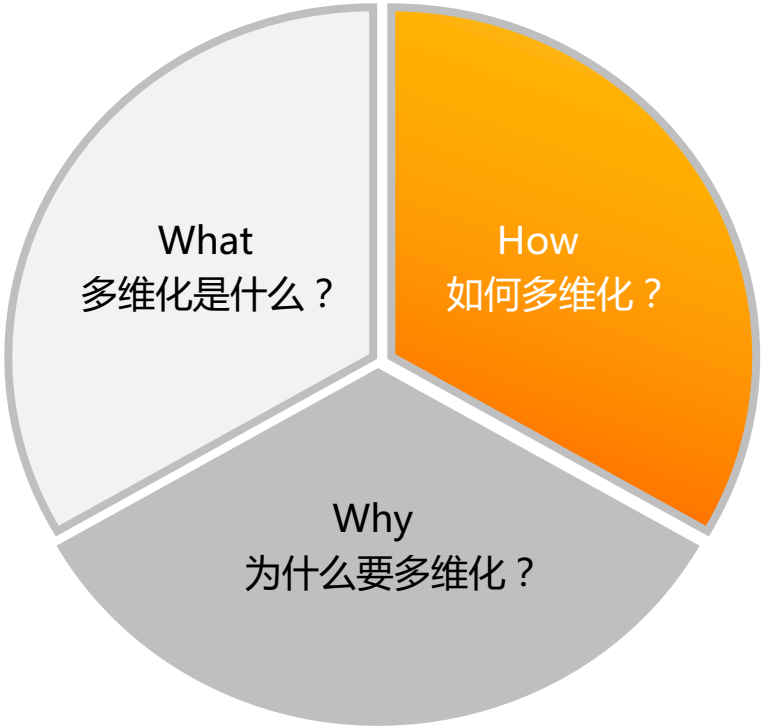
架构上存算分离

计算资源
弹缩方便快捷

运维难度低
使用门槛低

查询速度快

多维分析3W



Joyy | DataFun.

02 多维化的设计与思考

市面上的多维化方案

Clickhouse

Starrocks

Hologres

OLAP技术选型

分类	clickhouse	starrocks	hologres
协议	非标准协议	兼容mysql协议	兼容pg协议
是否开源	是	是	否
存算分离	否	否	是
读写分离	否	否	是
关联hive	否	是	是
分布式	否	是	是
实时更新	支持一般	支持一定程度更新	支持完善
场景匹配度	20%	70%	90%

Hologres测试环境

配置项	配置信息
计算资源	CPU：128 Core，内存：512 GB
存储资源	500 GB（逻辑存储）
Hologres版本	v1.1.68
网络环境	vpc
主要测试表	yy_mbsdkdo_original

测试-单线程场景

场景	子场景	测试表数据量		
		5000万	5亿	10亿
点查		2.6	5.1	3.6
count(*)	无where 条件	51	67	75
	Where +主键与非主键+and/or 自由组合	27	53	70
uniq(column)	无 where 条件	23	21	25
	Where +主键与非主键+and/or 自由组合	46	57	87
分页查询		44	55	61
(维度表)小表 join 大表		208	1417	2445
大表 join 大表		\	\	\

单位：毫秒（平均耗时）

Uniq语义：等同于性能更好的count(distinct column)

测试-多线程场景

场景	子场景	并发数			
		10	100	1000	5000
点查		45	46	304	/
count(*)	Where +主键与非主键+and/or	95	301	349	/
uniq(column)	Where +主键与非主键+and/or	1200	2700	19475	/
分页查询		299	1913	18943	/
(维度表)小表 join 大表		12228	123249	2107169	/

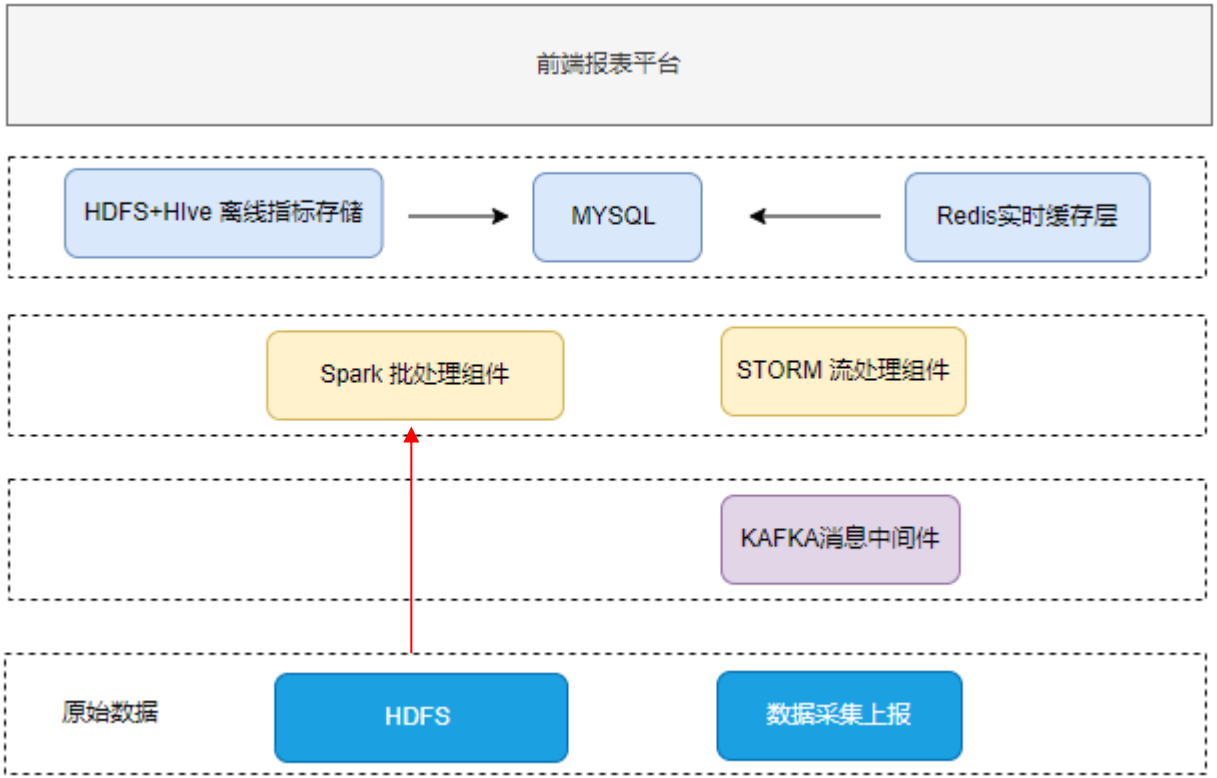
说明：

1. 对数据量为5亿的测试表做并发压测
2. 时间单位为毫秒(平均耗时)
3. holo 最大连接数为128*8(work 节点)=1024

测试结论

1. 大部分场景性能表现不错，点查耗时平均在[2.6ms ,300ms]区间内
2. 聚合统计查询在提高并发度下性能会有所下降
3. 大表之间关联的场景：查询的条件覆盖主键的前提下，10s内出结果，大量的shuffle对CPU和内存压力都很大
4. 数据的实时导入QPS可以达到 400w，符合预期

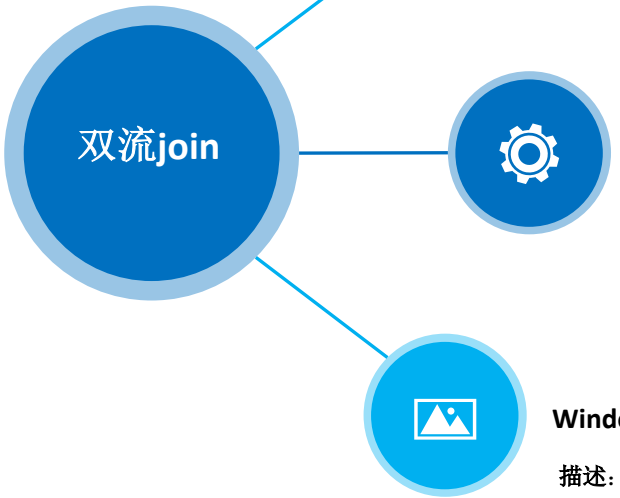
原有技术架构



架构上的优化


前端展示层
QuickBI
存储层
OLAP 存储层 Hologres
计算层
Flink实时计算
传输层
KAFKA消息中间件
DLF元数据
原始数据
数据采集上报
OSS存储系统

技术难点1-双流join



Regular Join

1.**描述**：两个流的输入全部保存在State中。
弊端：状态无限增长，严重依赖外存。

2.**描述**：设置空闲状态保留时长。
弊端：时间跨度长的数据有可能join不上。

Interval Join

描述：间隔连接，支持Process Time和Event Time。
弊端：超出时间范围，乱序数据有可能join不上

Window Join

描述：两个流相同Key，且相同窗口内的数据做Join。
弊端：超出窗口时间join不上。

解决方案：

大量的历史维度关联场景：
将流Join转化为点查

大流join大流：
维度最近数据存在state，形成缓存层，
缓存层查询不到，进入点查

技术难点2

QPS 10w 存明细，数据量大，存储成本高

背景：在移动端统计分析场景，需要进行大量埋点数据的存储，如果存放在hologres,会占用大量的存储容量。

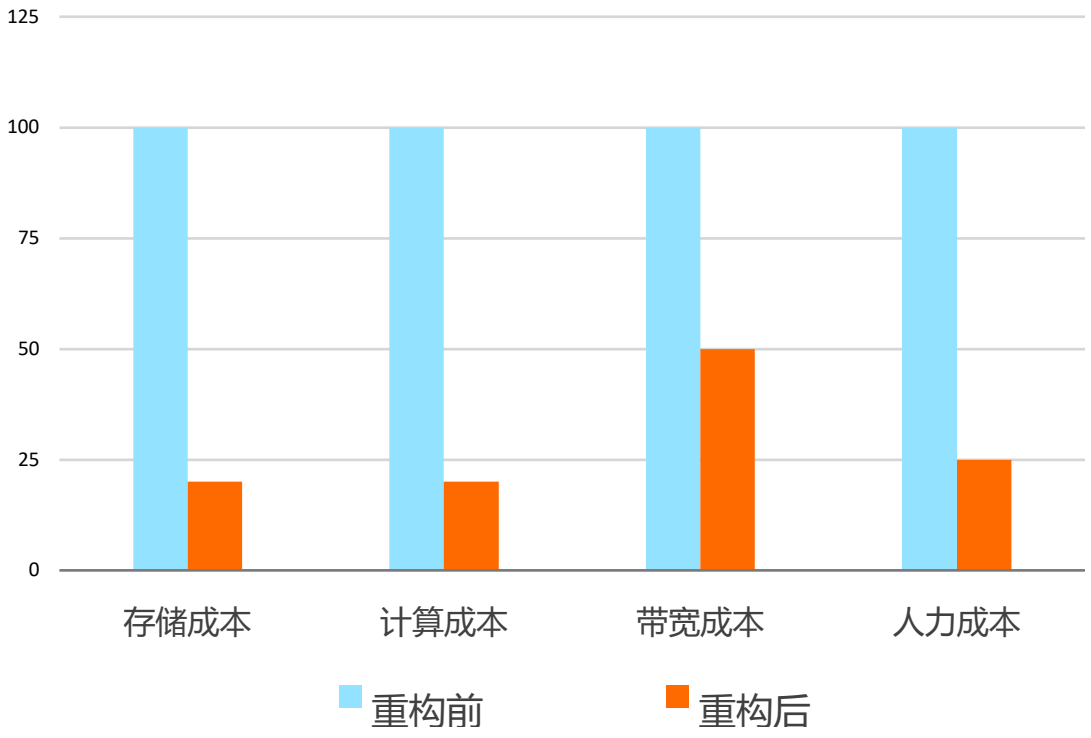
解决方案：
在统计分析场景下，大部分统计指标是UV，PV类型，分析同学更关心的是数据指标的时效性；维度的上卷，下钻；整体的指标趋势，对明细数据相对不太关注，可以使用bitmap方案，将uid，设备id进行bitmap化。

效果：1000w条心跳数据可以压缩为15w条。

Joyy | DataFun.

03 落地的效益

成本效益



存储成本降低：80%

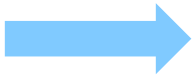
计算成本降低：80%

带宽成本降低：50%

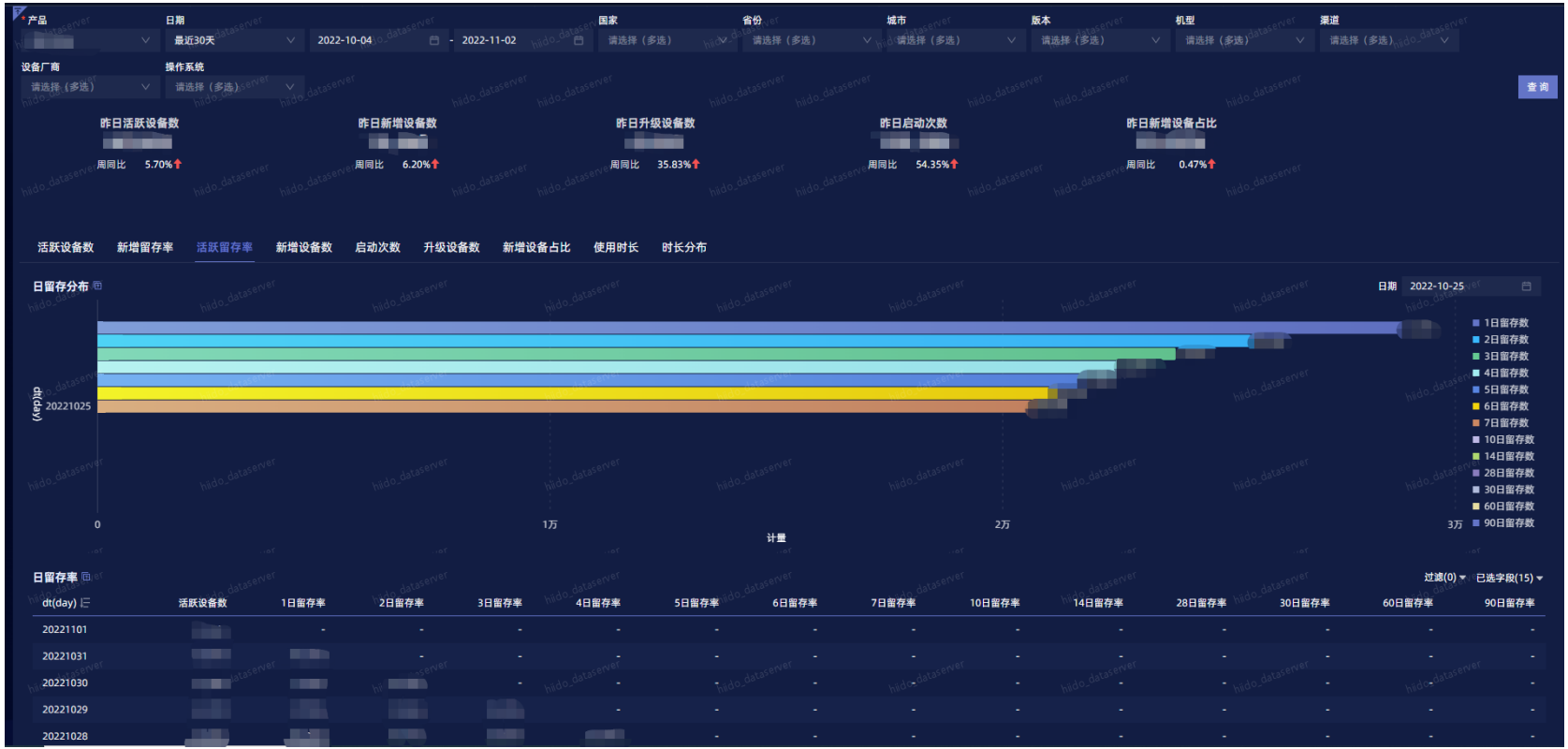
人力成本降低：75%

技术效益

旧统计分析系统	实时多维分析方案
1. 资源扩缩时长（1小时）	1. 存算分离, 扩缩时长（5分钟）
2. 新增统计指标（1天）	2. 新增统计指标（5分钟）
3. 计算引擎（实时，离线2套）	3. 计算引擎（flink 1套）
4. 可维护性（2个人力）	4. 可维护性（半个人力）
5. 指标统计(T+1, 部分小时级别)	5. 指标统计（分钟级别）



新架构的产品形态



Joyy | DataFun.

05 未来与展望

未来的方向

非常感谢您的观看

Joyy | DataFun.