

# ClickHouse Online Summer Meetup China 2022

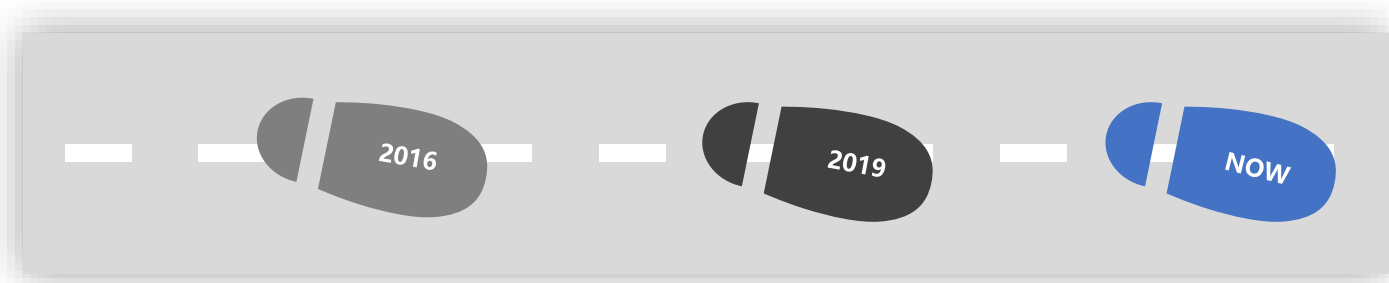
2022.7.2 | 13:30

ClickHouse在零售行业永辉超市的探索与实践  
---永辉 马吉辉

# 自我介绍



马吉辉



公司  
岗位

vargo科技  
(大数据架构)

永辉超市  
(数智研发)

- 2016年开始从事基于Hadoop的大数据开发与架构，精通Hadoop生态圈软件框架
- 对数仓ETL、维度建模、数据分析等有深刻理解
- 0-1构建数据中台，多年大数据平台研发管理运维经验，在大数据架构，数据采集，数据处理，OLAP数据分析领域有着非常丰富的研发经验

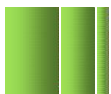
个人技术博客: [https://blog.51cto.com/u\\_12445535](https://blog.51cto.com/u_12445535)

# 目录:

## 本次演讲主要包含6个部分

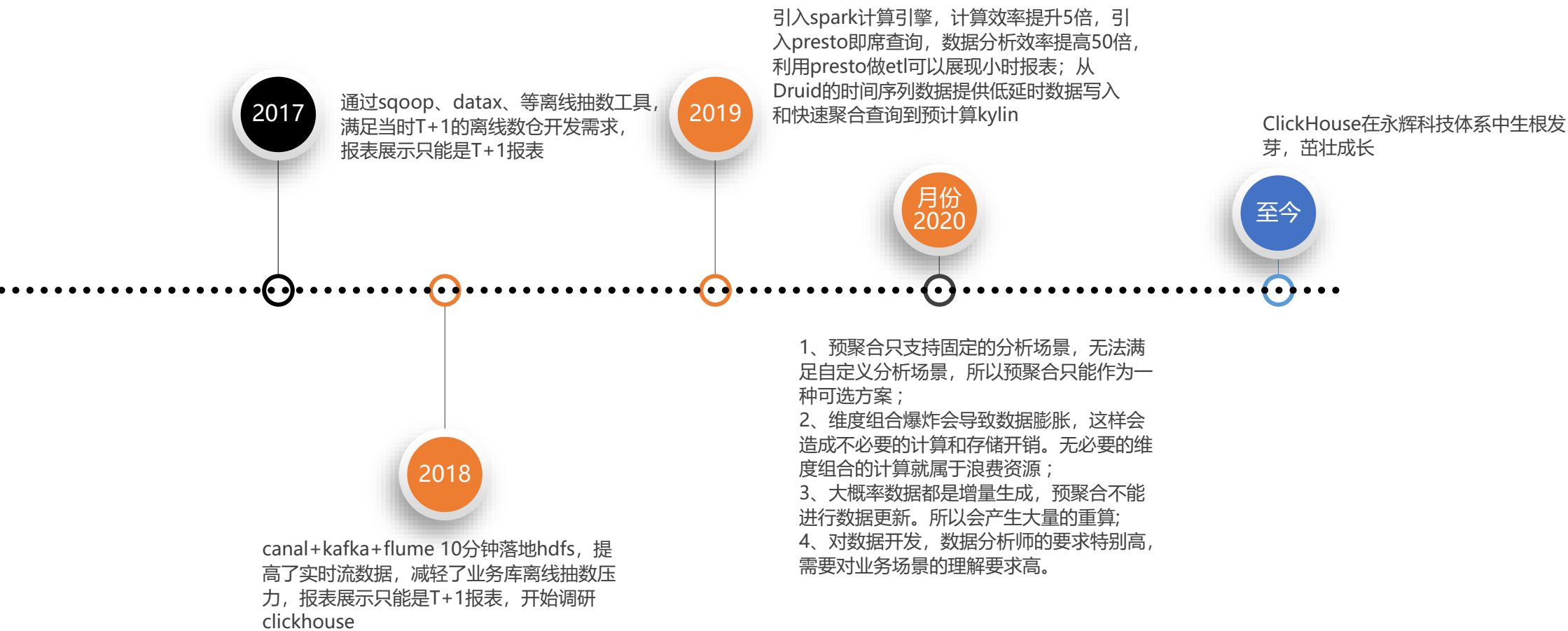
- 1) OLAP 架构在永辉的渐变之路
- 2) ClickHouse集群架构
- 3) ClickHouse在数据中台落地
- 4) ClickHouse业务实践
- 5) 我们遇到的问题
- 6) 未来展望

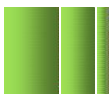




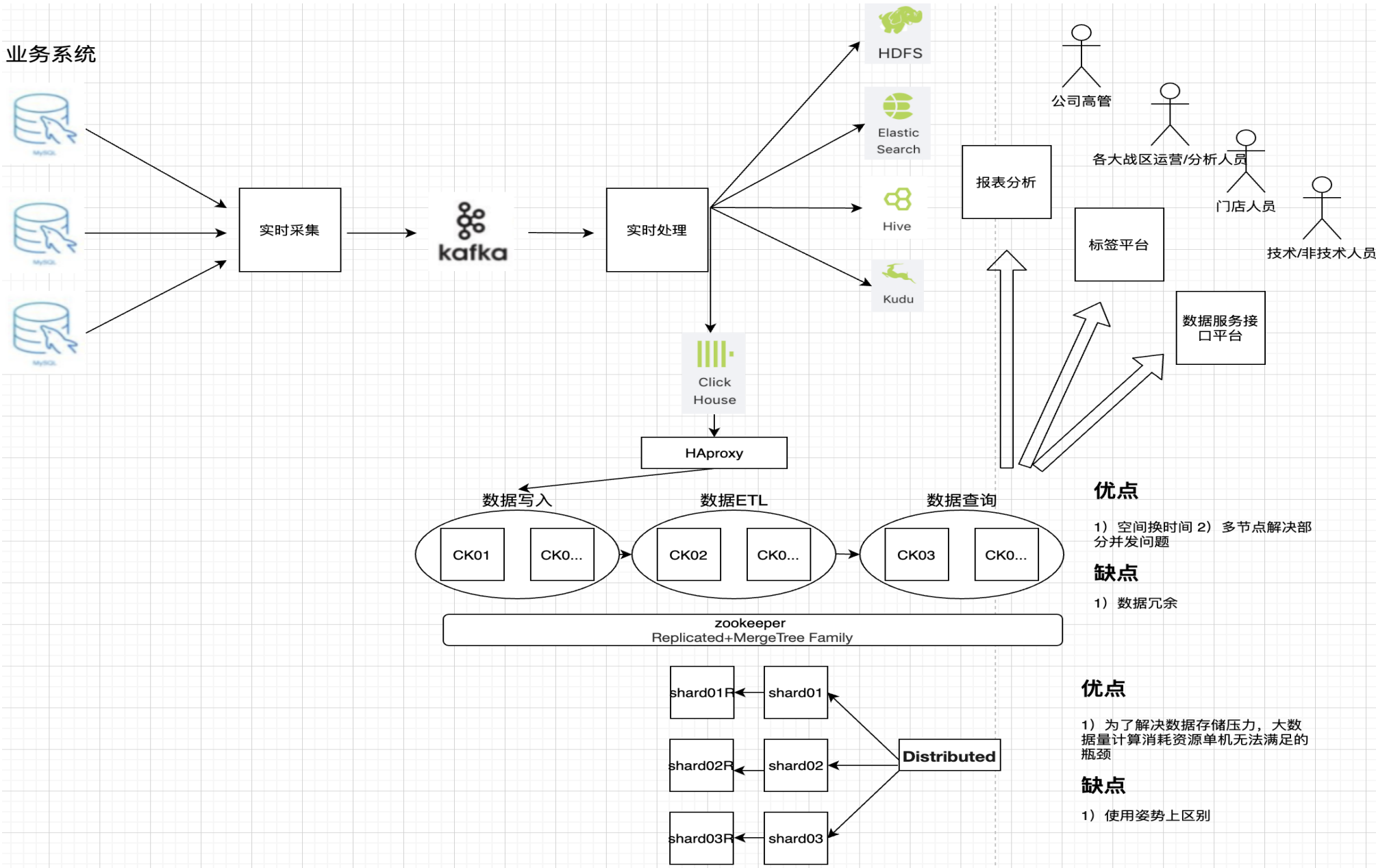
# 一、OLAP架构在永辉的渐变之路

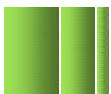
要谈到OLAP架构，那么就要先谈谈离线数仓的架构体系的发展之路，他们之间存在千丝万缕的联系





# 二、ClickHouse集群架构

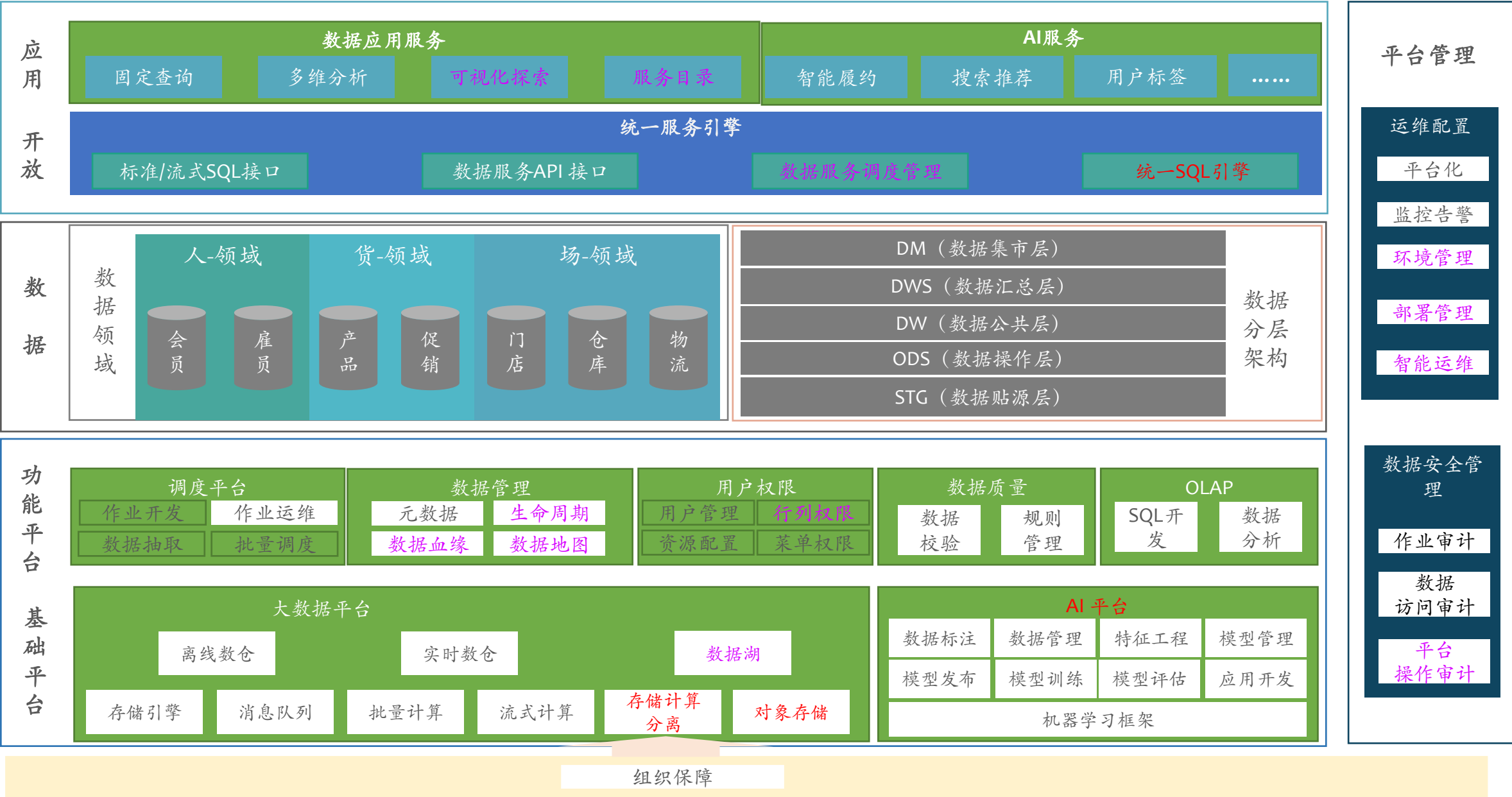




# 三、ClickHouse在数据中台落地

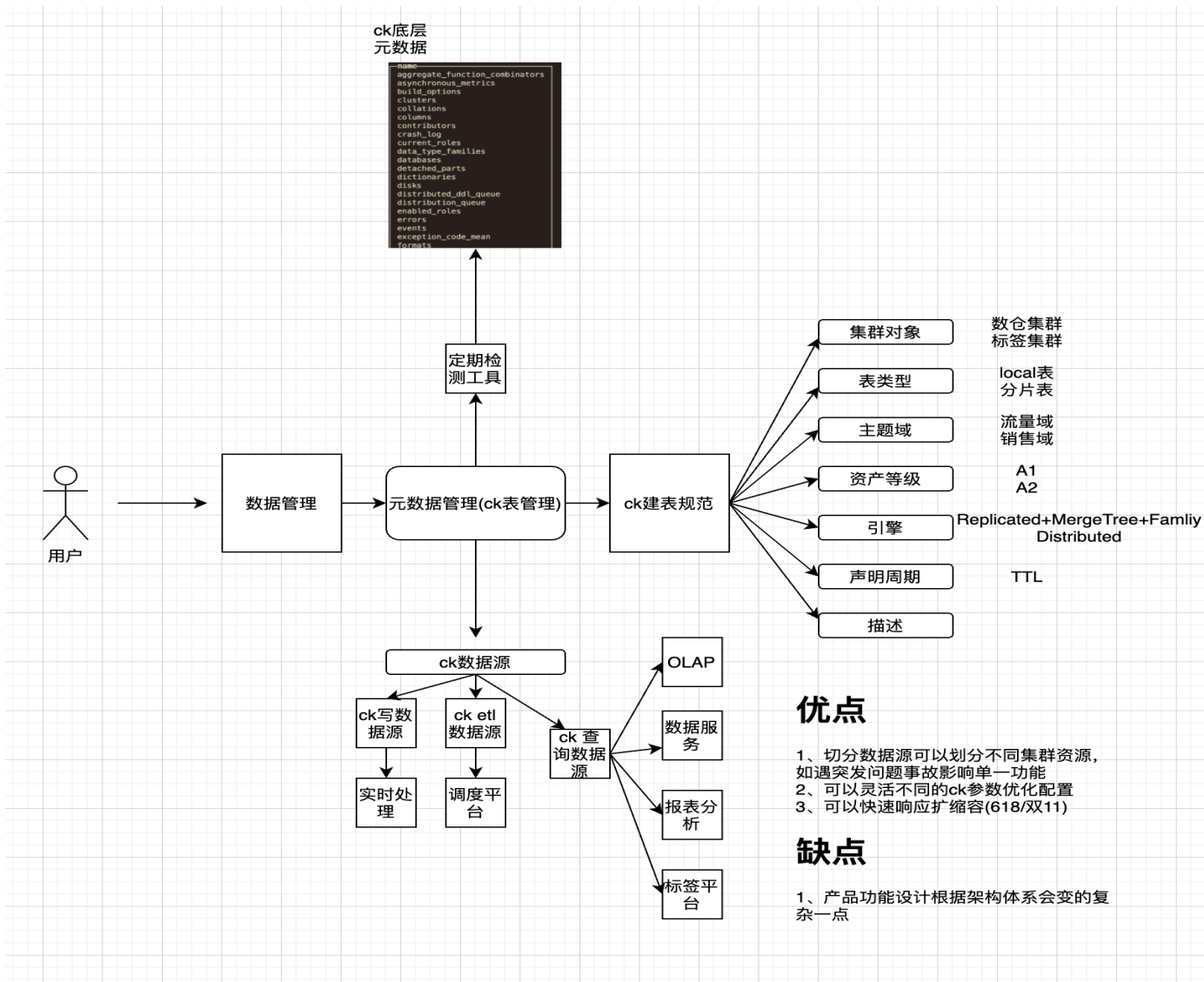
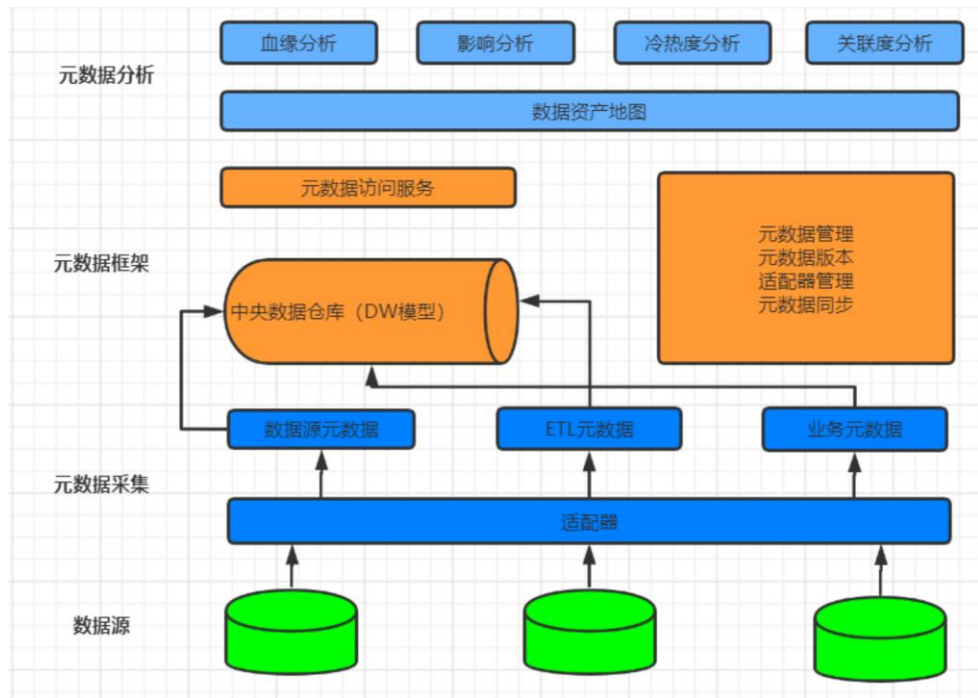
公司2020年开始自研数据中台，助力企业数智化转型，赋能业务全链路数字化

数据  
变  
现



# 三、ClickHouse在数据中台落地

ClickHouse --> 数据管理：数据中台的元数据&数据源管理中心，提供数据治理，资产探查等，赋能数据开发平台和业务数据开放



## 优点

- 1、切分数据源可以划分不同集群资源，如遇突发问题事故影响单一功能
- 2、可以灵活不同的ck参数优化配置
- 3、可以快速响应扩容(618/双11)

## 缺点

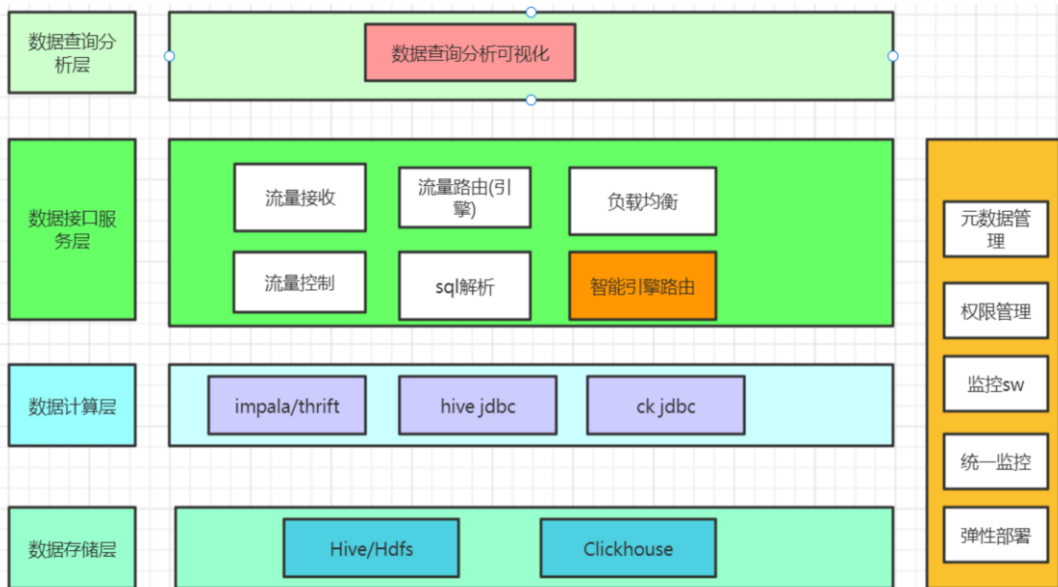
- 1、产品功能设计根据架构体系会变的复杂一点

# 三、ClickHouse在数据中台落地

ClickHouse

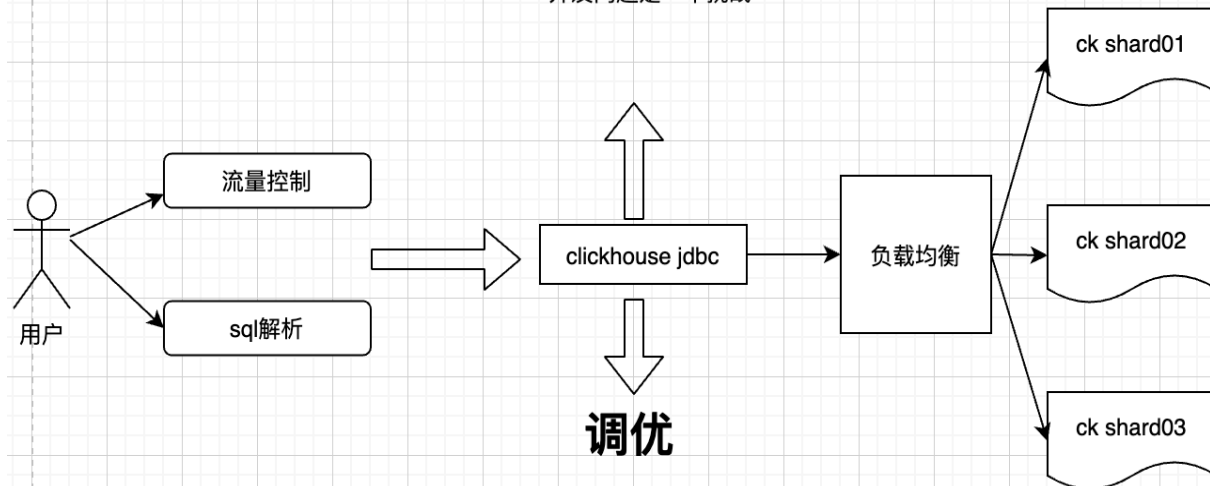
---

OLAP: 高效的自助数据服务套件, 提供各层次门槛的数据提取能力, 便捷的线上自助分析能力



## 挑战

1、OLAP目前主要数据开发数据分析师使用, 未来数据开放越多, 并发问题是一个挑战



## 调优

- 1、OLAP平台数据分页
- 2、用户行为审计日志
- 3、数据脱敏
- 4、帮助用户提高sql质量(比如: 自动识别分区跨段, 3分钟查杀, 内存使用超限预警)





# 三、ClickHouse在数据中台落地

ClickHouse ---> 智能运维(前期) AIOPS 一种全新的数字化运维能力, 且是企业数字化转型的必备能力

数据中台clickhouse治理慢查询与大内存占比sql日报

使用姿势http://.../clickhouse/?page\_name=参数

- 参数1: slow\_query 表示 大于10分钟的慢查询
- 参数2: top\_memory 表示 使用内存大于40G
- 参数3: slow\_query\_and\_top\_memory 表示 使用内存大于20G并且查询时间大于10分钟

说明: 获取的数据为每天T+1日报

|  | type | exception_code | query_id | databases | tables | query_start_time | query_duration_min | query_end_time | read_rows | read_bytes | written_rows |
|--|------|----------------|----------|-----------|--------|------------------|--------------------|----------------|-----------|------------|--------------|
|  |      |                |          |           |        |                  |                    |                |           |            |              |

统计的1天内的成功失败的比例情况 每一个节点的

```
3721 2022-05-23_23:55:01
3722 数据中台ck_ 错误代码汇总提示:
3723 0 832269 正常执行
3724 511 2796 角色相关问题
3725 241 183 (重点关注指标)任务执行失败超过总体ck指定内存限制或单个query内存限制
3726 62 6 语法错误
3727 999 5 (重点关注指标)无法在 ZooKeeper 中分配块号: 协调::异常: 连接丢失
3728 47 4 查询的语句丢失字段
3729 50 3 字段类型错误
3730 493 2 角色方面的错误
3731 1000 1 字段列表中未知的列(引擎名)
3732 6 1 解析语法错误
3733 215 1 某字段不在聚合函数下也不在grouo by 中
3734 386 1 类型没有超类型
3735 349 1 无法将NULL值转换为非Nullable类型
3736 38 1 字段的值取的有问题
3737 497 1 没有足够的权限
3738
```

2022-05-25\_20:51:01

数据中台clickhouse此时状态信息如下:

- 目前 1 负载为32
- 目前 1 正在运行query总内存为201 G
- 目前 1 free内存为5 G
- 目前 1 cache内存为33 G
- 目前 1 available内存为36 G
- 目前 1正在运行query总个数为9
- 目前 1 单个query运行消耗内存大于20G的任务有5个



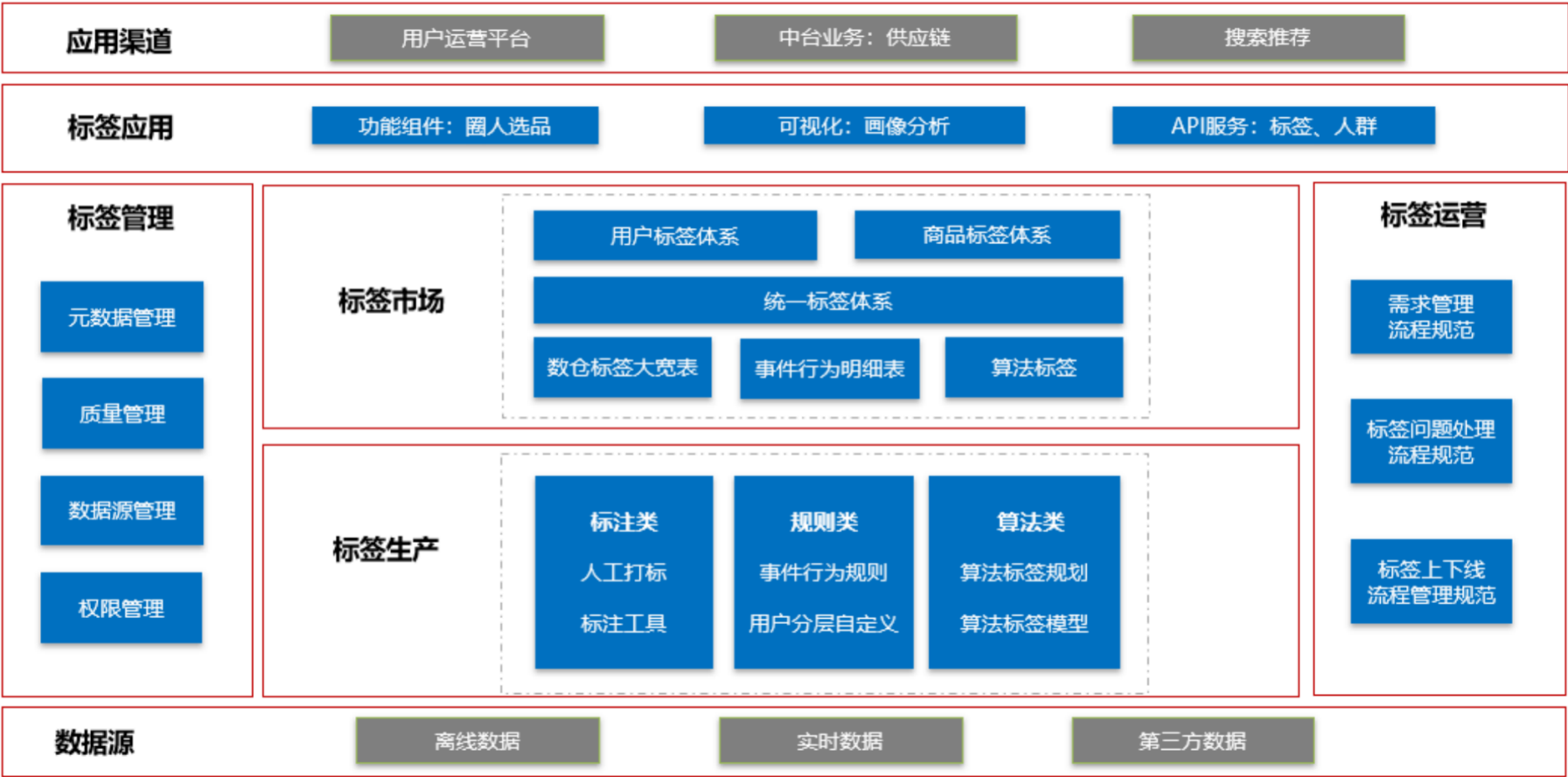


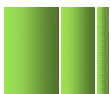
# 四、ClickHouse业务实践

ClickHouse ---->

标签平台：沉淀公司的统一、全面、准确、多维的标签资产，  
并提供完整的标签管理功能，为公司的用户运营  
商品运营以及个性化推荐等场景提供高效的标签服务

标签整体架构图





# 四、ClickHouse业务实践

ClickHouse

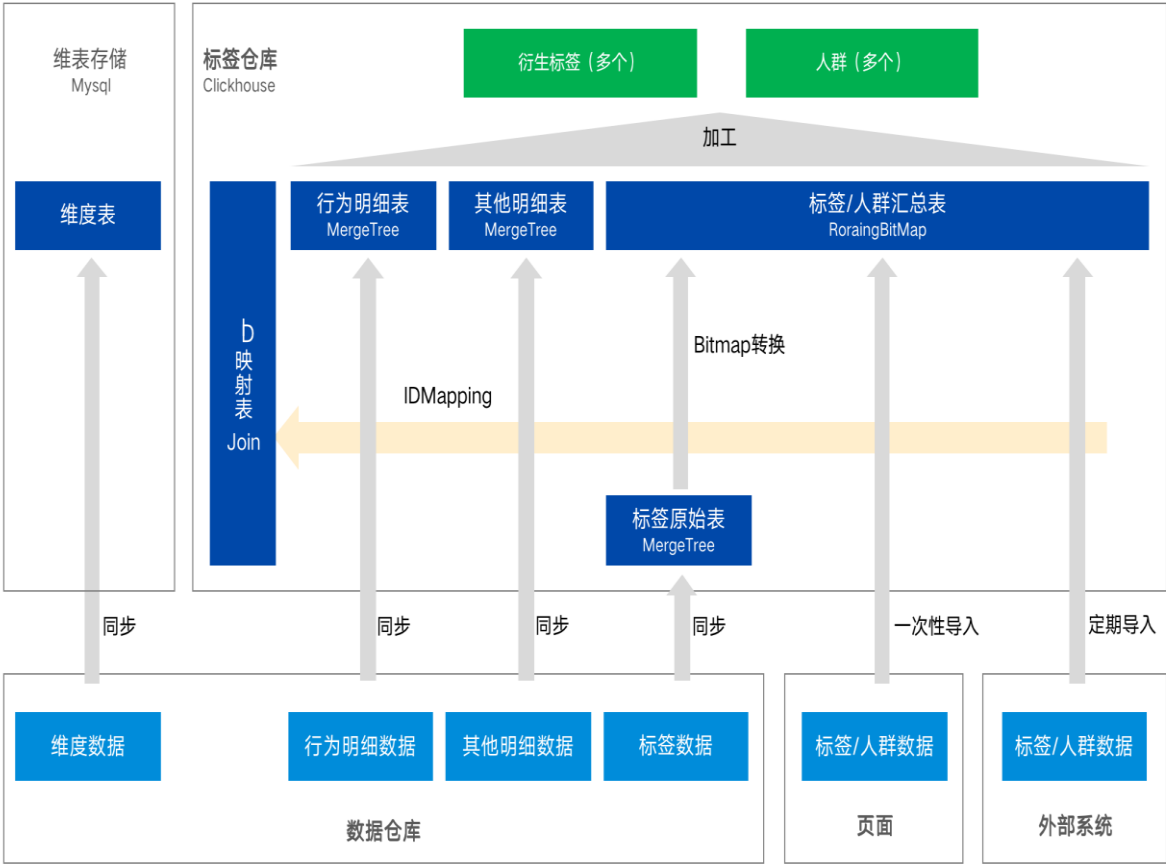
---

标签平台

标签平台设计架构图



标签数据架构图

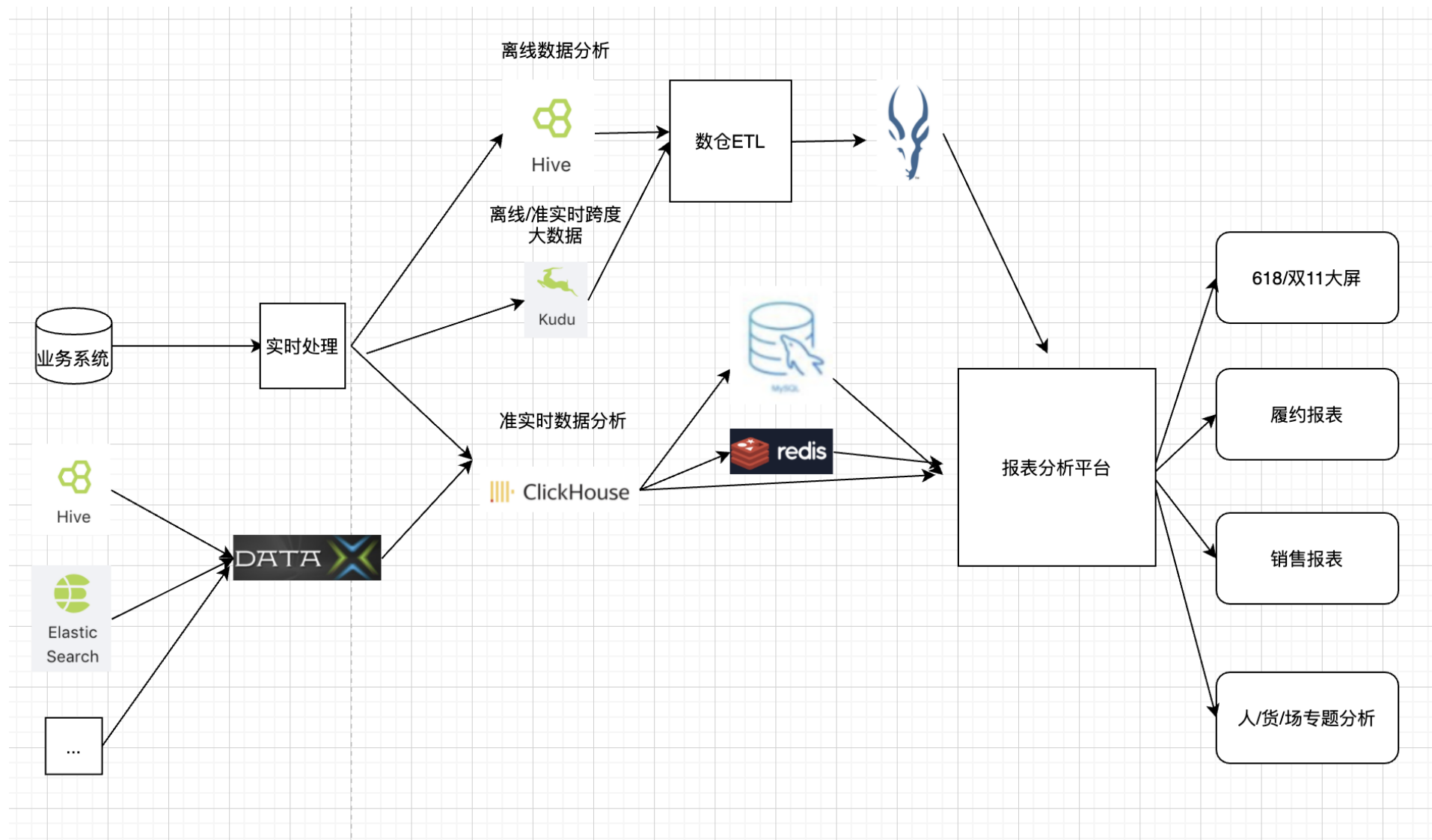


# 四、ClickHouse业务实践

ClickHouse

---

报表分析: 提供多维度、多指标的灵活可视化分析能力, 全面支撑至上而下各个层级的日常数据分析需求, 有效支撑科学化、智慧化决策



## 五、我们遇到的问题



01

CK资源管理不健全，没有类似yarn的统一资源管理器

02

实时报表需求业务夹杂 大量离线数据 同步环比跨度大

03

CK并发问题难支撑高并发  
利用到非replicated引擎，  
数据容灾难度大

04

CK运维成本高/ZK压力随着业务增长压力会越来越大

## 六、未来展望

ClickHouse 智能管理平台





Power Ma 

上海 浦东新区



扫一扫上面的二维码图案，加我微信



永辉生活超市app