

# 知乎客户端性能数据分析系统实践



2019-05-25

# 目录

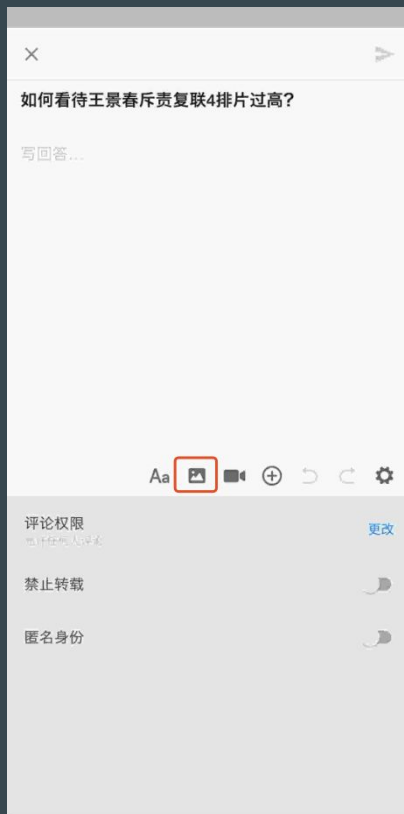
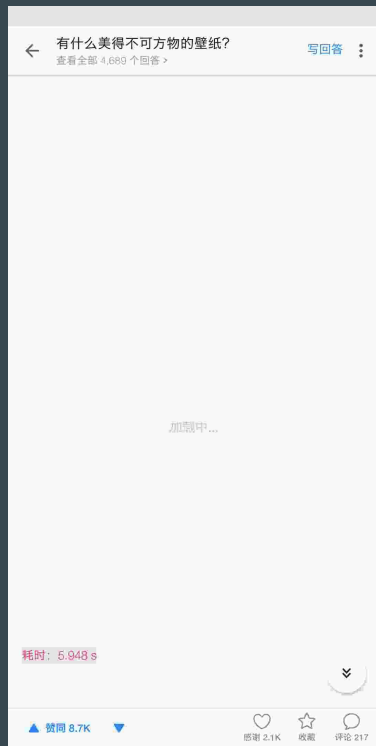
- ❖ 概览和常见问题
- ❖ 项目需求和平台演变
- ❖ 平台功能
- ❖ 设计实现
- ❖ 相关人员和团队介绍
- ❖ QA?

# 概览

为更好的提升客户端质量，实时了解客户端的线上运行情况，发现客户端的低效率代码并提供日志搜索进行优化，提升用户满意度、建立用户信任，积极主动的发现运行问题，建设了一套用于查看客户端线上性能指标运行情况的数据分析系统。

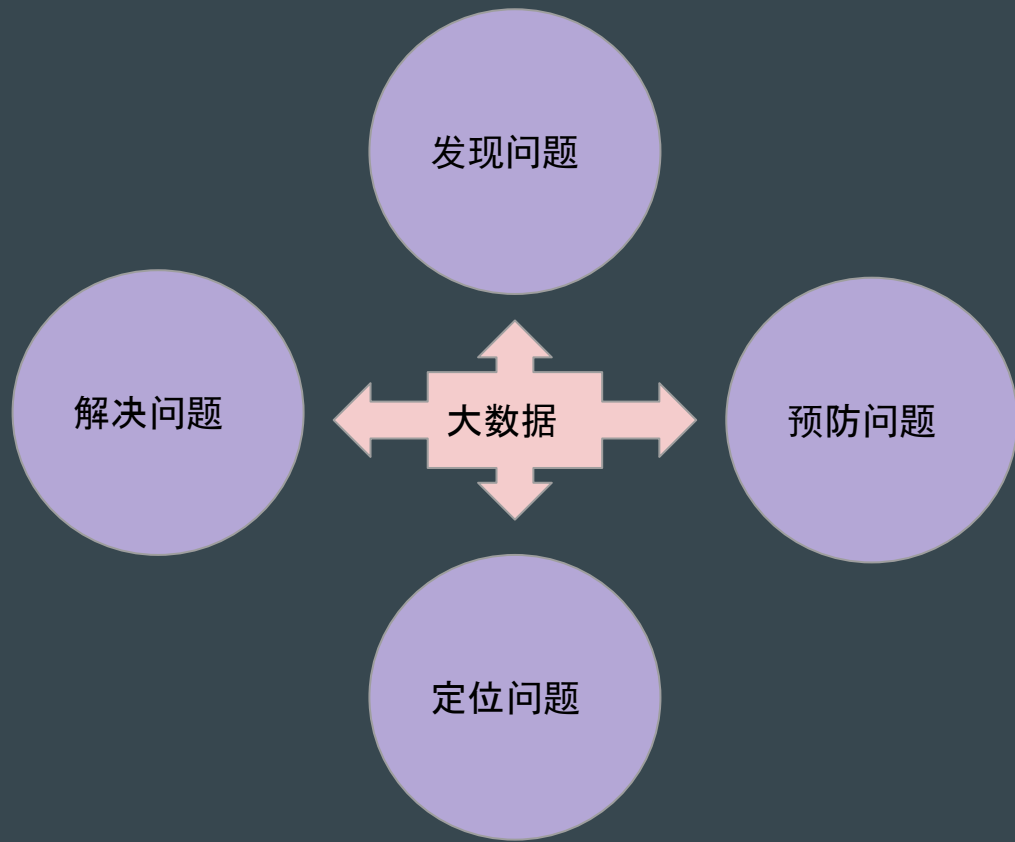
通过大数据技术计算搜集数据的各项指标，如启动时间均值、P90、P50 等，可视化展示记录数据在使用过程中的性能表现，为版本的提测、发布提供质量标准。

# 常见问题



# 项目需求:

- ❑ 客户端线上运行质量数据
- ❑ 疑难问题数据记录和统计
- ❑ APP 质量评定
- ❑ 便捷的性能测试体验
- ❑ APP 性能指标发版质量标准



# 平台演变:

- ❑ 常规数据库
- ❑ 缓存计算分析
- ❑ Hive、hadoop 离线分析
- ❑ 实时分析, 数据落地
- ❑ 听云平台

# 平台功能

# 性能数据分析

## 版本分析

下图中是各个版本的启动时间 P90 图,可以轻松发现版本迭代中启动时间的差异



## 机型分析

上图中是各个机型的启动时间 P90 图,可以轻松发现对应的机型启动时间差异,进行针对性的修改





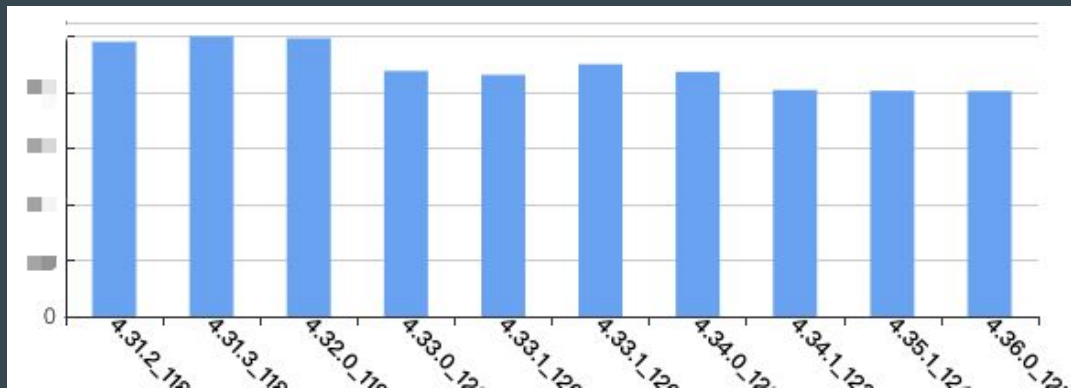
# 趋势分析

## 版本迭代趋势分析

右图中是版本迭代总回答页加载时长

分析说明:

- 对比最近的 10 个迭代版本数据
- 数据随着版本迭代变短
- 可以得出随着版本更新App 回答页面表现变佳



# 日期横向对比

## 日期数据对比分析

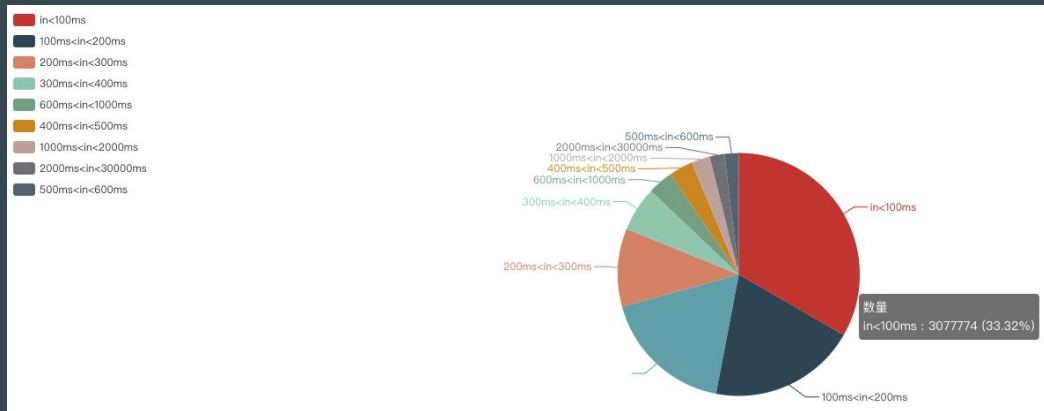
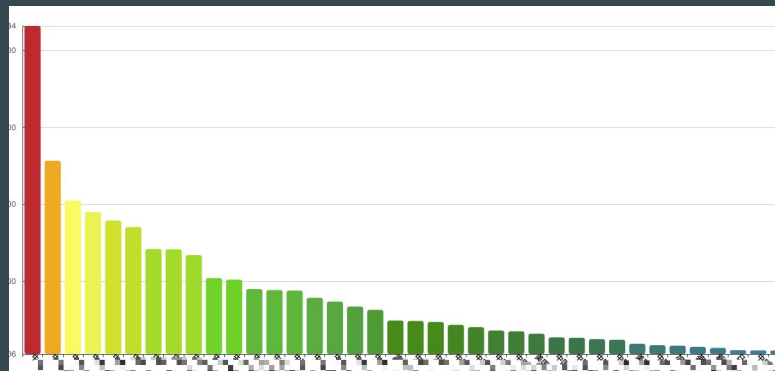
- 对比任意两个时间段的曲线
- 下图是图片请求成功率对比



# 维度饼图/柱图对比

## 同维度数据对比分析

- 网络请求时长的分布区间比例
- 各个区域的请求时长排序柱图



# 地域对比分析

## 地域维度数据对比分析

- 按照地图显示相关性能数据
- 图中是 iOS 端网络请求时长 P50 值

地图



经纬省会	耗时_P50 ↓
台湾	87
新疆	227
香港	07
青海	96
西藏	84
澳门	179
贵州	175
宁夏	75
云南	75
甘肃	157
广西	165
重庆	161
黑龙江	161
辽宁	158
内蒙古	155

# 设计实现

## App 性能数据发生的场景统计

维度:

- 平台
- 版本
- 机型
- 运营商
- 网络
- .....

## App 运行过程中的性能数据统计

指标:

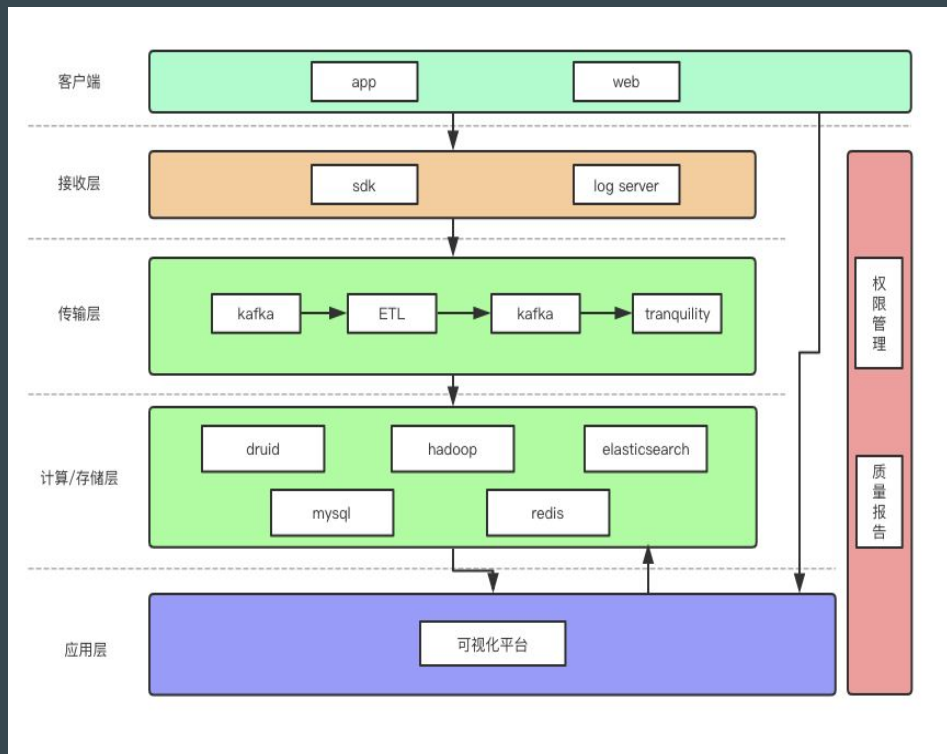
- 页面加载时间
- APP 启动时间
- 系统性能
- 流量统计
- 运行卡顿和异常信息
- .....

# 系统架构图

从上至下数据实时处理

处理步骤:

- 客户端、web 请求数据生成
- sdk 处理数据发送至 kafka
- ETL 实时处理数据并分流
- tranquility、pipeline 处理数据传输
- Druid 入库、elasticsearch 入库
- 可视化平台实时查询分析



# 数据标准化 1

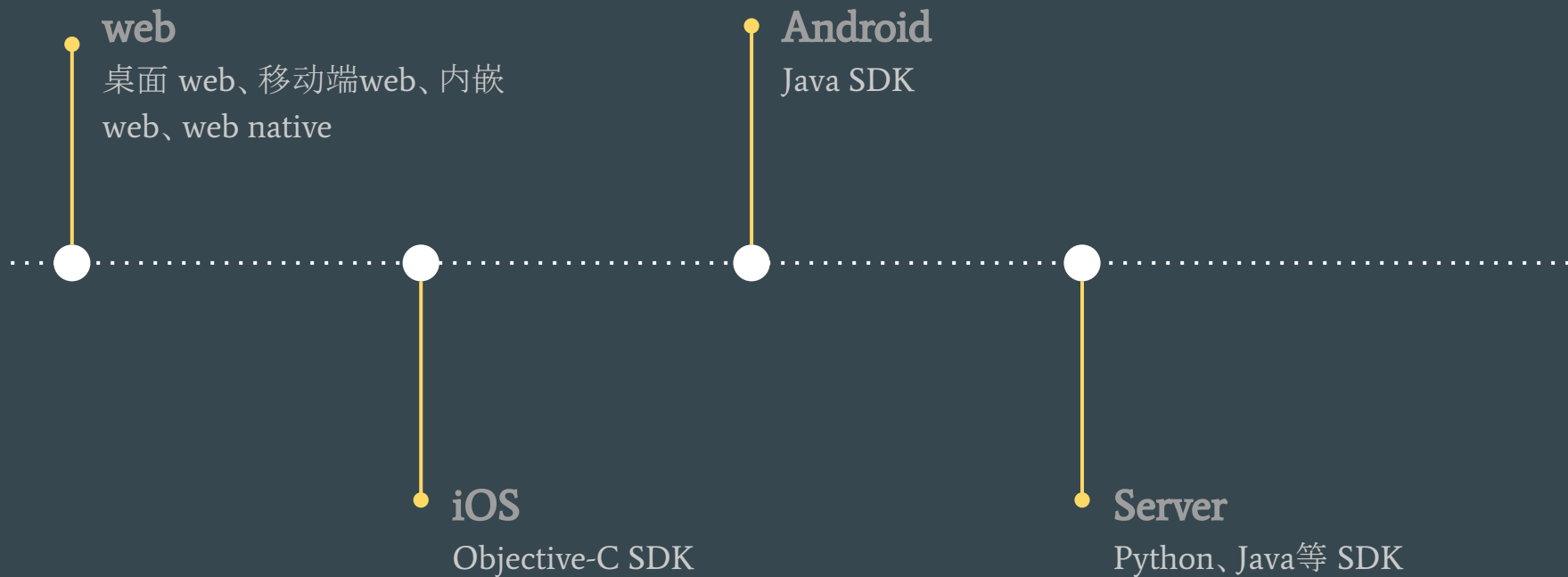
## Protobuf

平台无关、语言无关、可扩展、轻便高效的序列化数据结构的协议，可以用于网络通信和数据存储。

```
message SearchRequest {  
  required string query = 1;  
  optional int32 page_number = 2;  
  optional int32 result_per_page = 3 [default = 10];  
  enum Corpus {  
    UNIVERSAL = 0;  
    WEB = 1;  
    IMAGES = 2;  
    LOCAL = 3;  
    NEWS = 4;  
    PRODUCTS = 5;  
    VIDEO = 6;  
  }  
  optional Corpus corpus = 4 [default = UNIVERSAL];  
}
```



# 数据标准化 2



# 中间技术

## 数据流订阅

Kafka

数据总线、业务数据隔离

## 数据存储

HDFS

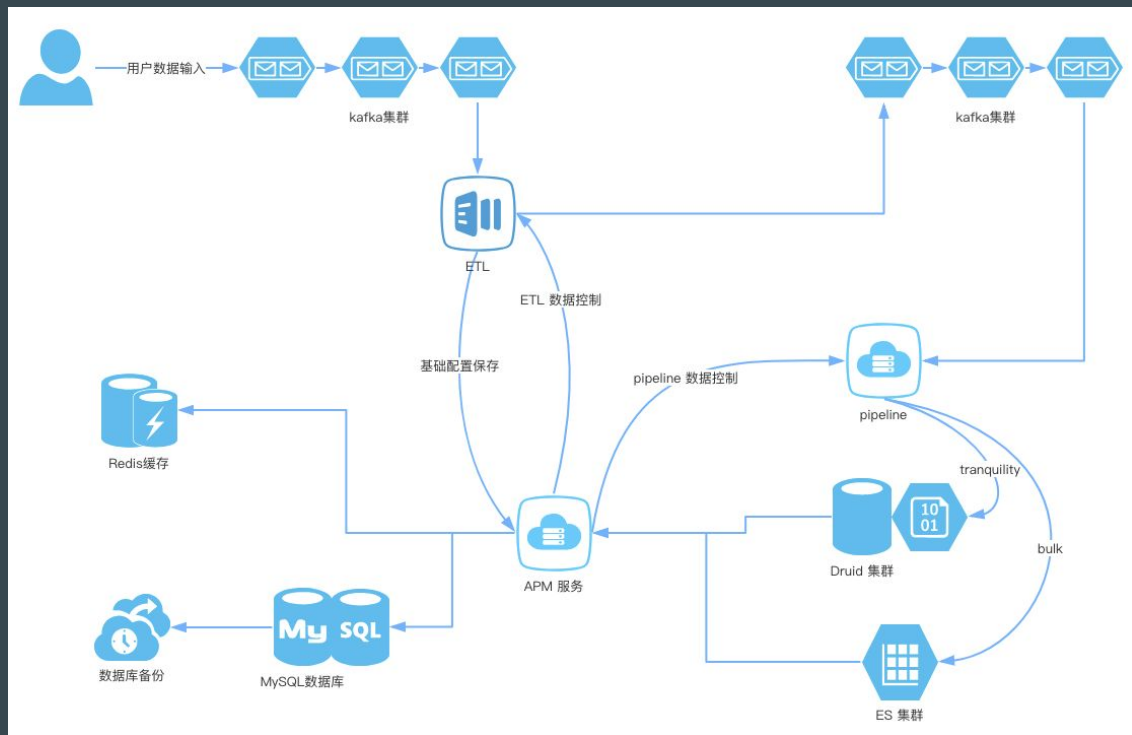
## 实时计算

Spark Streaming、Druid 实时采样、过滤数据

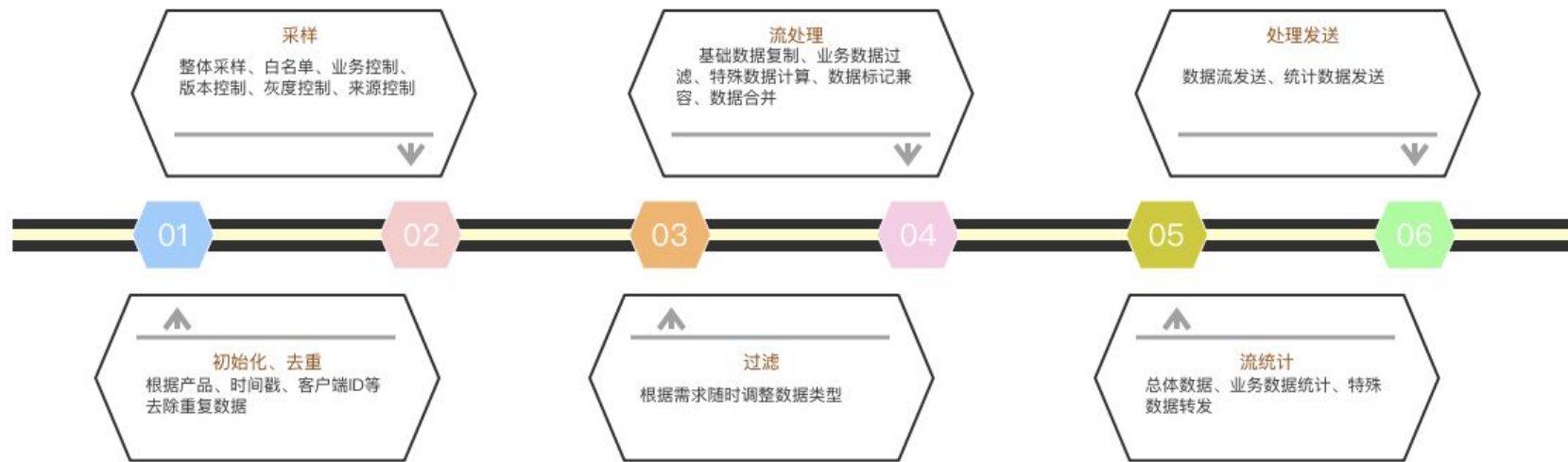
## 数据搜索查询

Elasticsearch 全文数据搜索、秒级查询

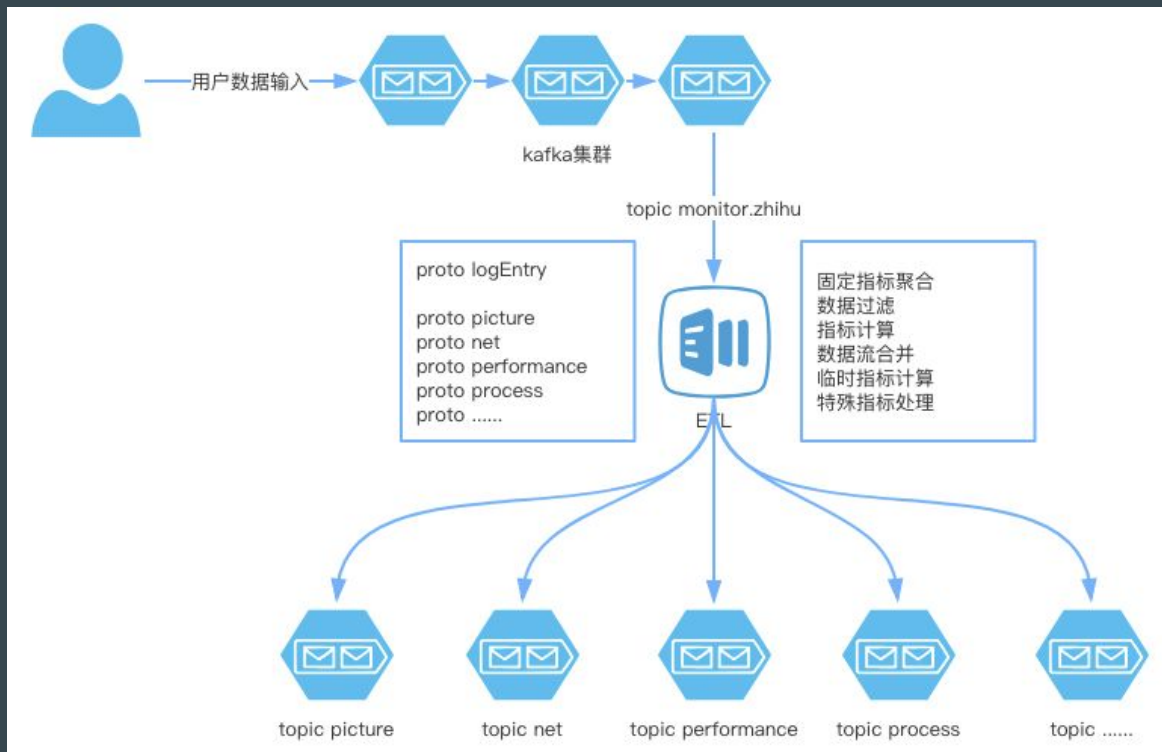
# 数据预处理



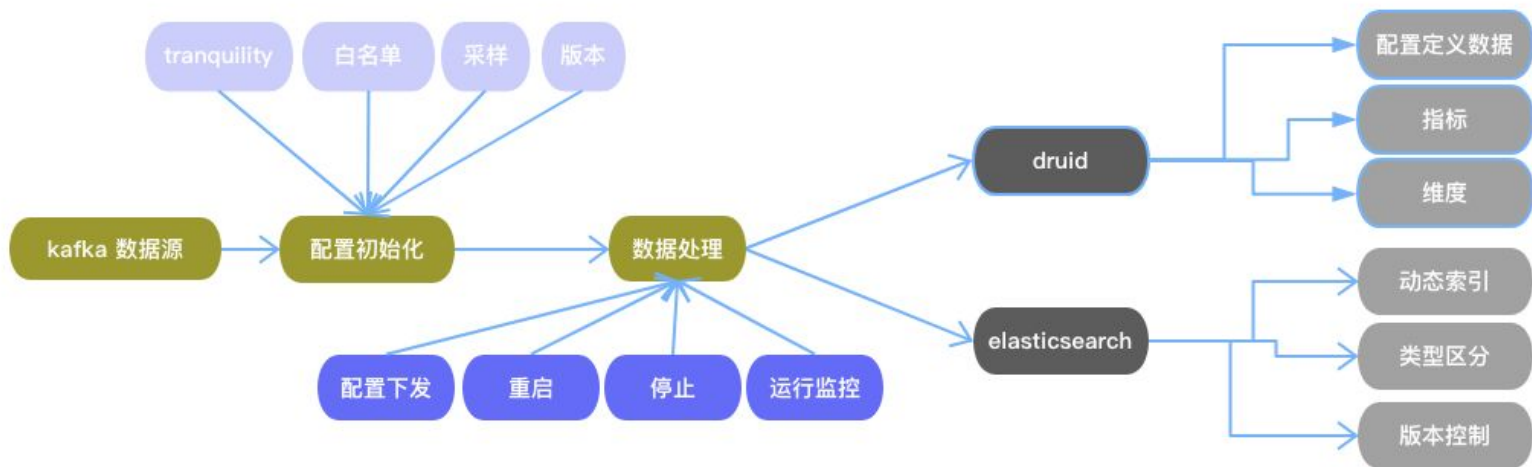
# ETL 数据流处理



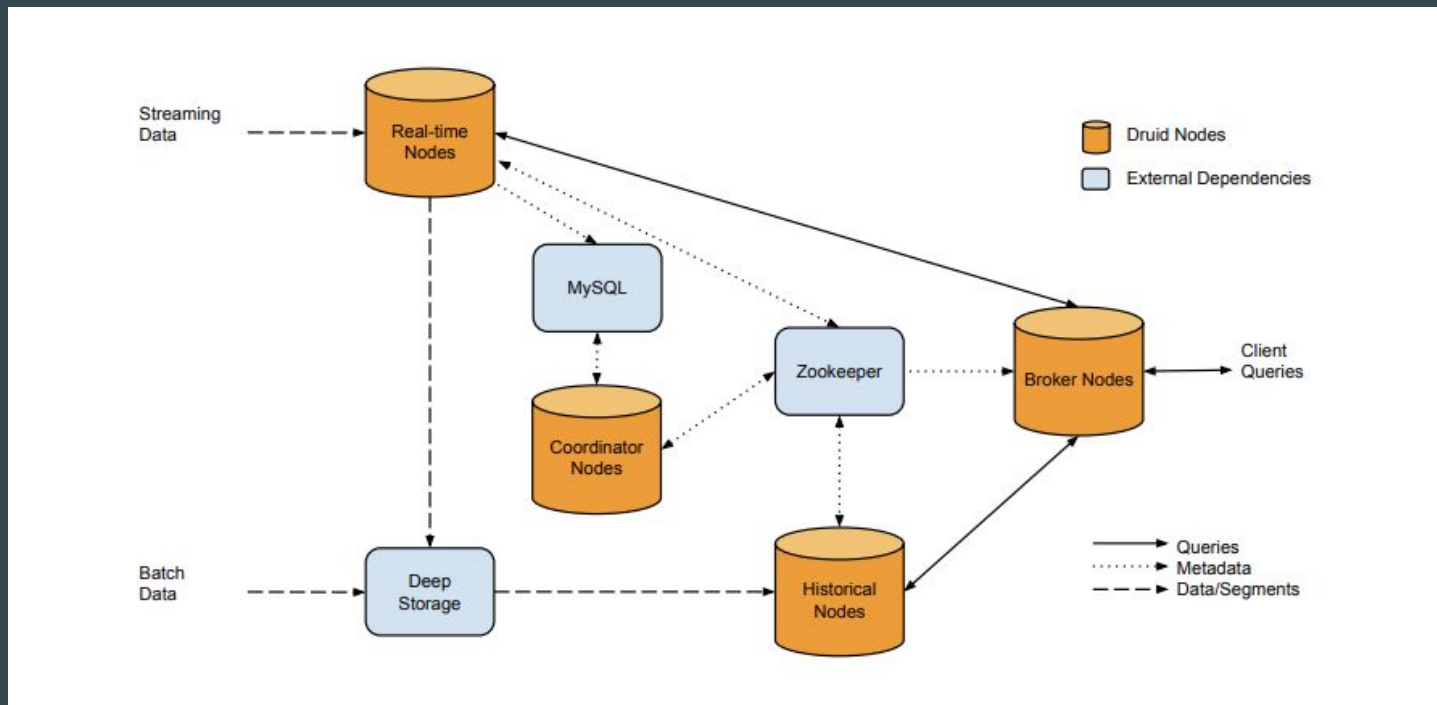
# 数据分流



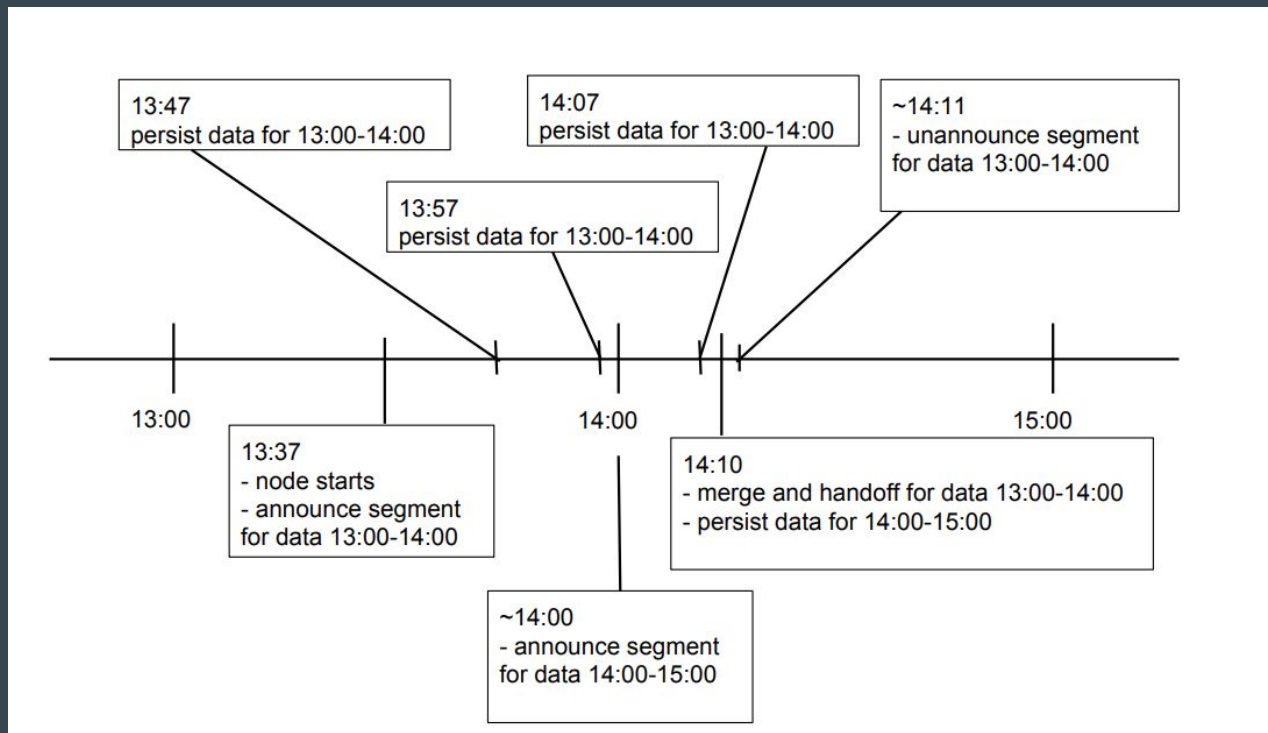
# PIPELINE 数据流处理



# Druid 架构

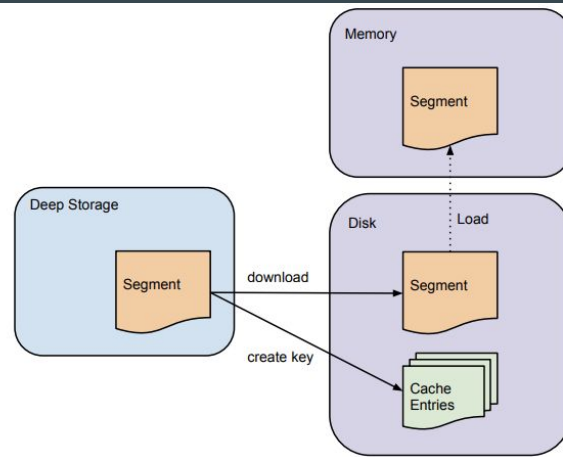
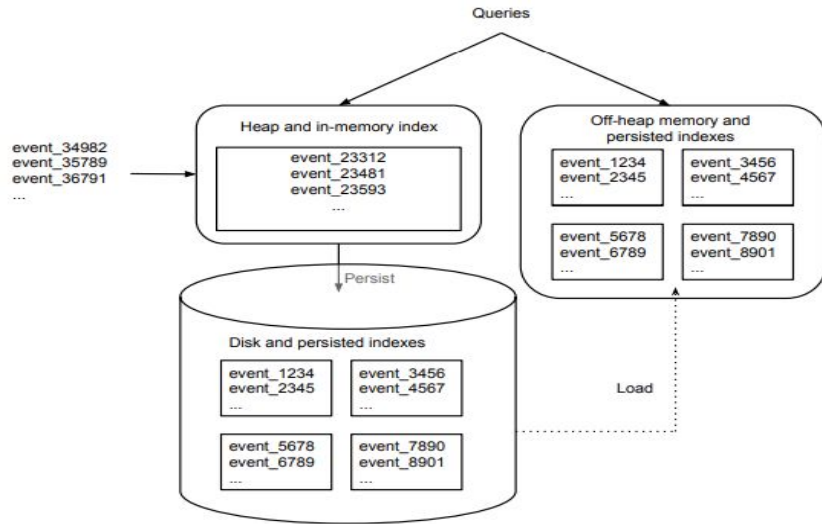


# Segment 存储

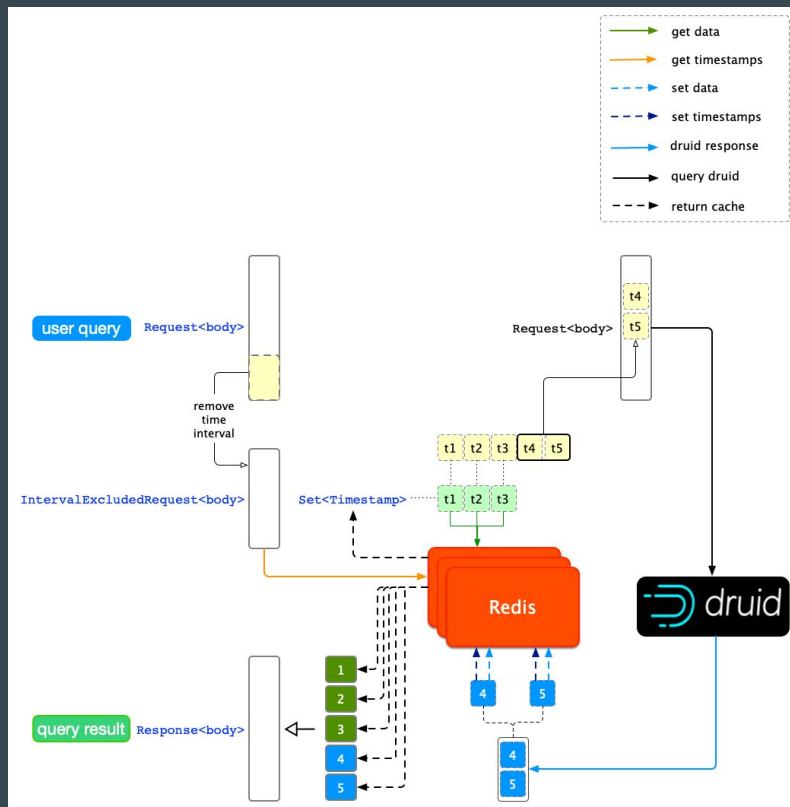
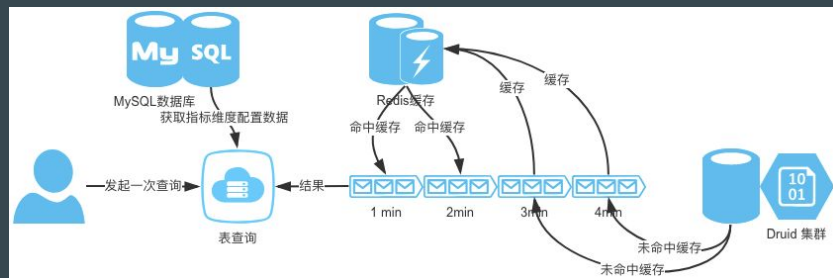




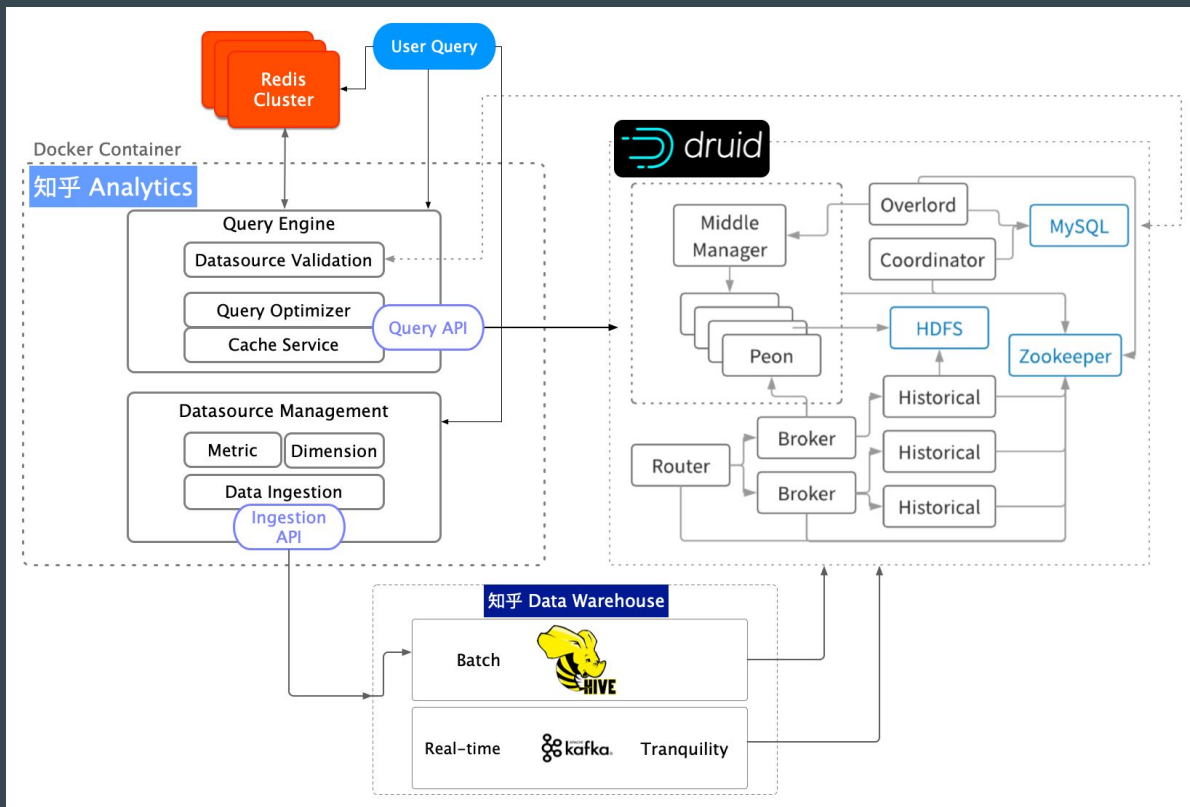
# Druid 数据存取



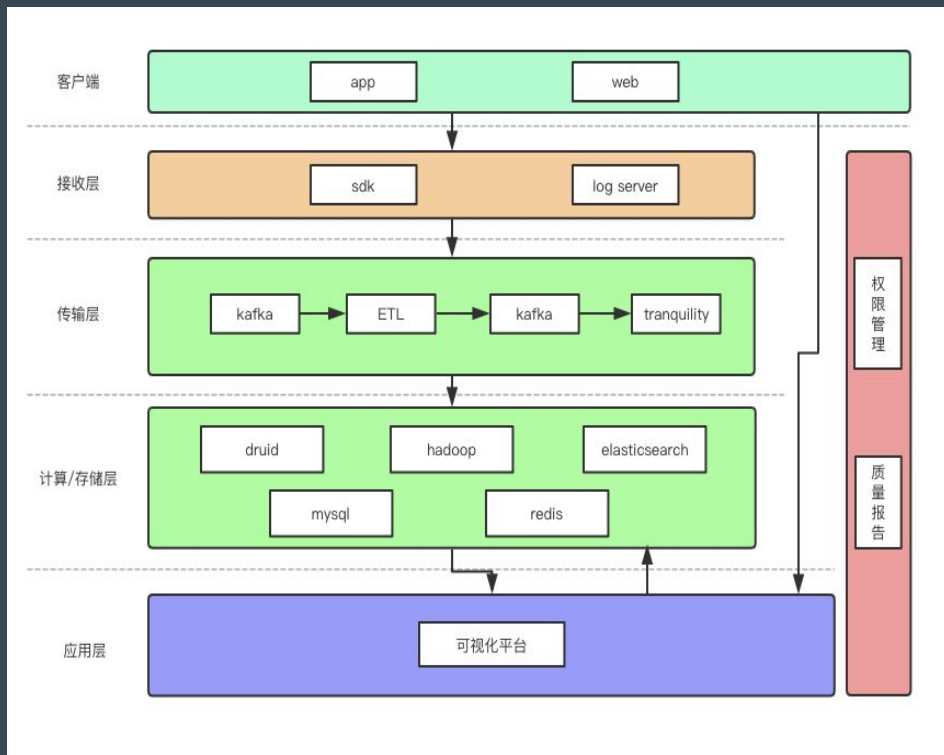
# 分流和缓存设计



# 实时多维分析平台架构



# 系统架构图回顾



从上至下数据实时处理

## 处理步骤:

- 客户端、web 请求数据生成
- sdk 处理数据发送至 kafka
- ETL 实时处理数据并分流
- tranquility、pipeline 处理数据传输
- Druid 入库、elasticsearch 入库
- 可视化平台实时查询分析

# 相关人员和团队介绍

# The Team



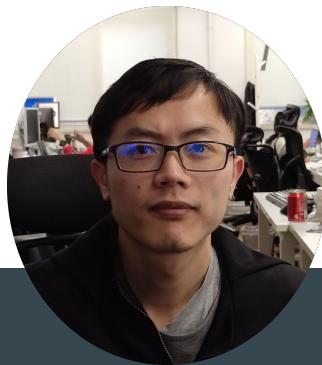
陈康, 高级工程师

负责平台前端开发、数据模型设计和平台架构设计。



王宁珂, 高级工程师

负责平台后端开发、数据平台接入和数据流处理开发。



牛豪飞, 平台工程师

负责平台前后端开发、数据流处理开发和运维。

# 知乎质量保障团队

属于知乎技术中台事业部，承担了公司各产品线的质量保障工作及以下平台的设计开发工作：

- ❑ 移动端云测平台
- ❑ 性能数据分析监控报警平台
- ❑ 移动端持续集成平台
- ❑ 移动端内测及发布平台
- ❑ 知乎质量平台
- ❑ 接口自动化平台
- ❑ ...

欢迎有志于质量保障工作的知友们加入知乎 QA 团队，与知乎一起发现更大的世界。

QA ?