

# 一个用户行为分析产品的设计与实现

曹隼@神策数据

客户需求



产品功能

产品需求



技术选型

技术限制



产品实现

- 神策分析：一个可以私有化部署的用户行为分析产品
  - 亿级 DAU，秒级入库，秒级查询
  - 全端埋点/ID-Mapping/跨屏打通
  - 功能灵活易扩充：事件分析、漏斗分析、留存分析、分布分析、用户分群、指标预警、点击分析、自定义查询等
  - 任何分析功能都可以下钻、筛选，维度指标无需预先指定，灵活定义
  - PaaS 平台，数据完全对客户开放



显示

↻

支付订单

▼

的

实际支付金额的总和

▼

和

实际支付金额的人均值

▼

✕

+

↻

任意事件

▼

的

总次数

▼

和

触发用户数

▼

✕

+

↻

搜索商品.触发用户数/任意事件.触发用户数

① 命名为

搜索人数比例

以

百分比

▼

显示

+

按

操作系统

▼

查看

+

按

国家

▼

查看

✕

且

年龄

▼

区间

▼

20

与

30

之间

✕

性别

▼

等于

▼

男

✕

①

✕

显示漏斗

注册流程

按

第 2 步 点击注册 | 注册渠道

查看 ?

创建漏斗

且

用户属性 | 注册来源

等于

融360

第 1 步 启动 App | 是否...

?

为真

+ 筛选条件

2017-7-24 至 2017-7-30

窗口期:  
1天

## 注册流程转化漏斗

趋势

对比

显示设置

总转化率  
6.16%

启动 App

422 人

100%

点击注册

422 人

57.35%

总转化率(%)

14%

12%

10%

8%

6%

4%

2%

0%

## 漏斗步骤

设置关联属性 ☒ ?

1

商品详情页浏览



+ 触发限制条件

操作系统

等于

Android



关联属性

商品ID

2

加入购物车



+ 触发限制条件

关联属性

商品ID

3

提交订单详情



+ 触发限制条件



关联属性

商品ID

4

支付订单详情



+ 触发限制条件



关联属性

无

+ 增加步骤



初始行为是 [o] 登录APP ▾ + 筛选条件

后续行为是 [o] 加购物车 ▾ + 筛选条件

且

商品二级类型 ▾

=

×

商品价格 ▾

=

×

同时显示

[o] 加购物车 ▾

总金额 ▾

×

[o] 分享商品 ▾

次数 ▾

×

+

且用户符合 + 筛选条件

📅 2017-7-17 至 2017-7-22

用户先进行任意事件，后进行任意事件的7天留存分析

显示设置 ▾

7天留存 ▾



初始行为日期 ▾

总人数

第1天

第2天

第3天

第4天

第5天

第6天

第7天

7-17 (一)

👤 5,411

635人  
12200元  
11.74%

在 7-17 (一) 进行登录APP的 5,411 人中，有 635 人随后在次日(7-18(二))进行了加入购物车：加入购物车的总金额为 12200 元。其中有 29 人进行了分享商品，总次数为 86,214 次

0人  
0元  
0%

0人  
0元  
0%

0人  
0元  
0%

7-18 (二)

👤 5,648

635人  
12200元  
11.74%

3.21%

3.20%

0%

0人  
0元  
0%

0人  
0元  
0%

7-19 (三)

👤 5,764

635人  
12200元  
11.74%

241人  
8200元  
3.20%

0人  
0元  
0%

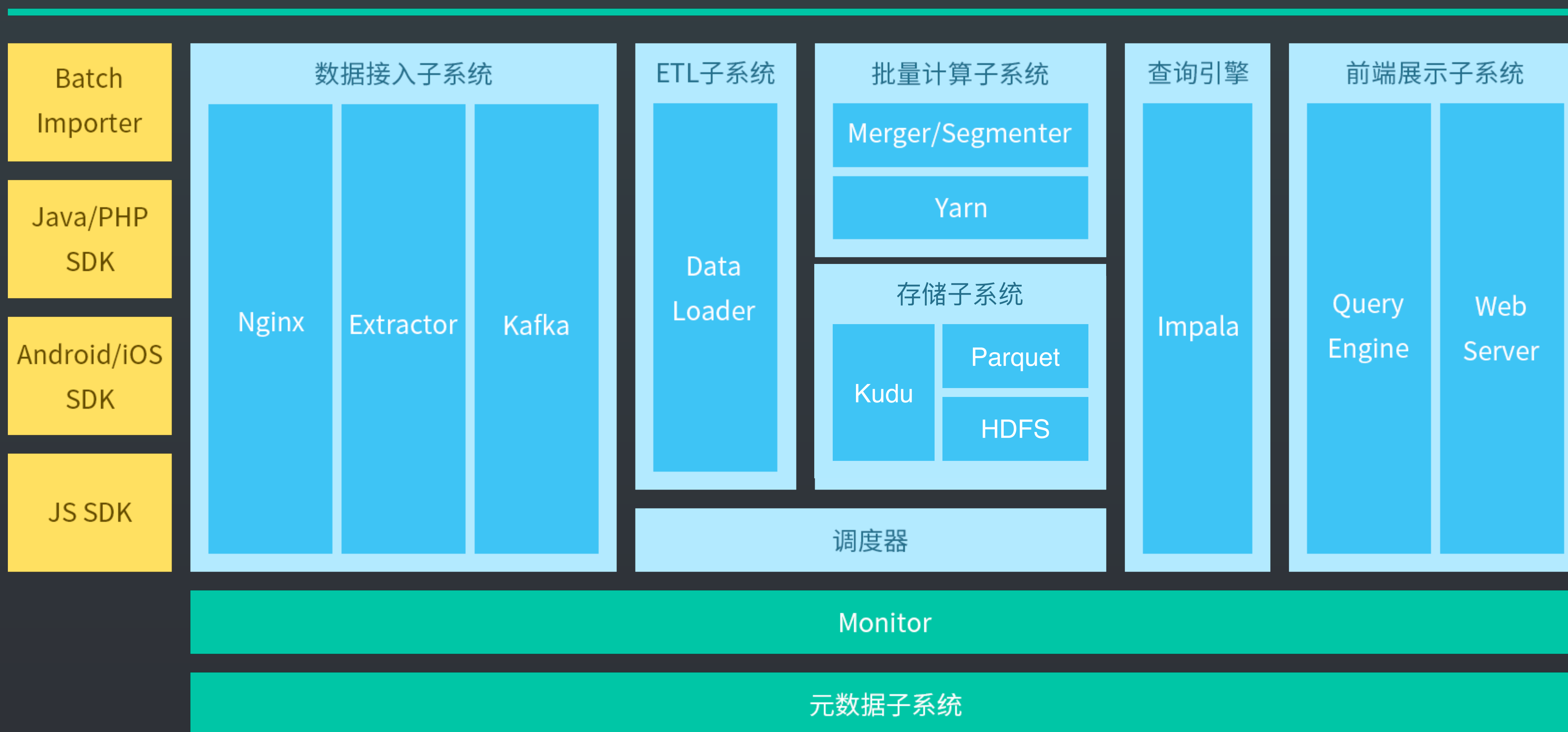
0人  
0元  
0%

0人  
0元  
0%





## 数据流向



# 1

## 前端 操作

全埋点

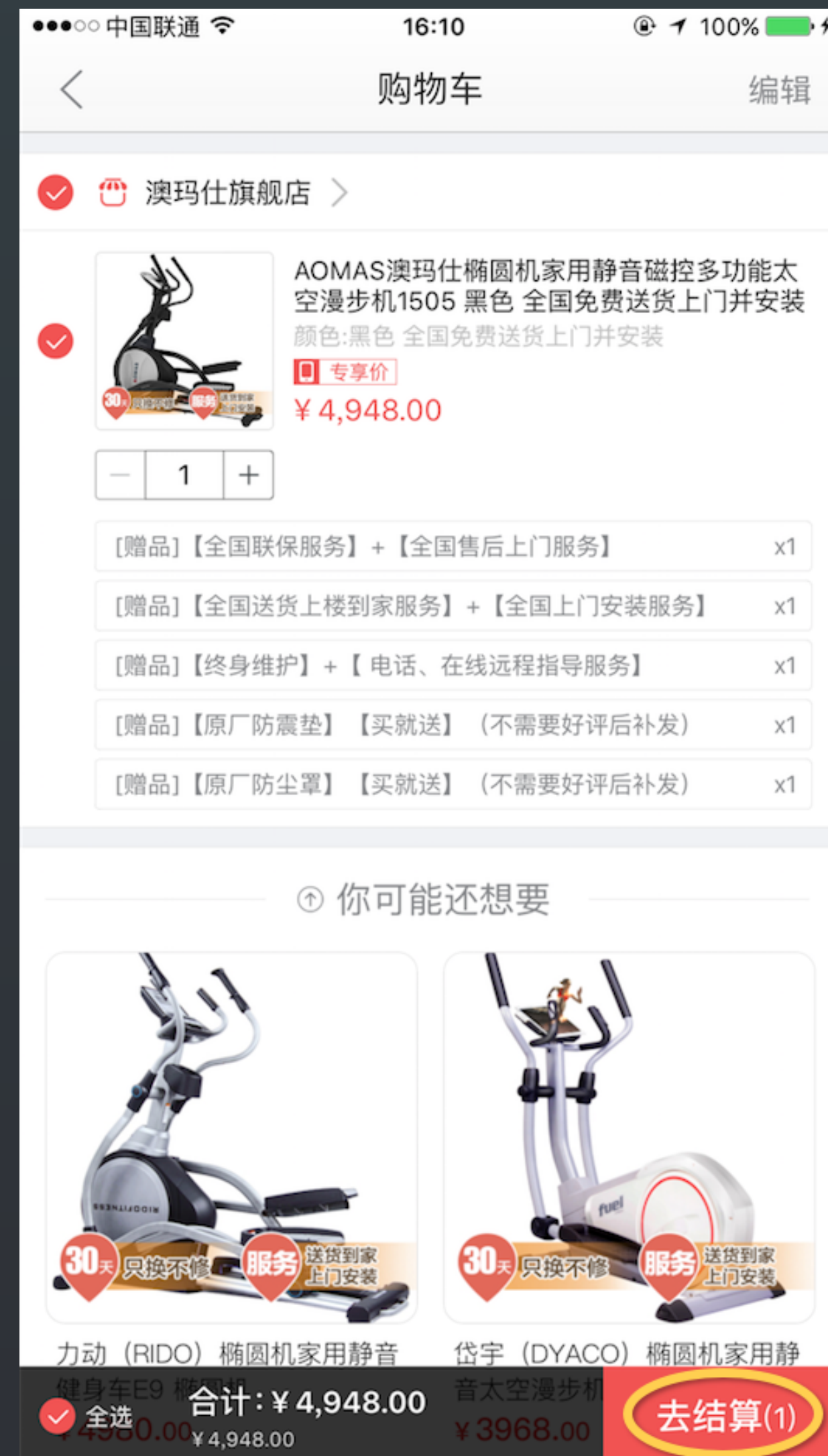
某时某刻某人点击了某个按钮

可视化埋点

某时某刻某人提交了一个订单

代码埋点

订单金额、商品名称、用户级别



- 前端埋点：
  - 全埋点：采集所有的页面浏览、控件操作、App 启动等
  - 可视化埋点：采集指定的控件操作，无法自定义属性
  - 代码埋点：可以自定义属性

- iOS 和安卓 SDK 的数据发送策略：
  - 保守的发送策略，以保证用户体验为核心
  - 数据首先缓存在本地，达到一定规模或间隔时才发送
  - 仅在 3G/4G/Wi-Fi 时发送数据，发送时会打包、压缩
  - 在首条数据、进入后台、退出程序时尝试发送
  - 提供强制发送接口

## 2

# 后端 日志

代码埋点

商品库存、商品成本

导入工具

用户风险级别、用户潜在价值

```
2016-04-18 00:01:17,647 [StreamEventThread-for-partition-0] INFO [com.sens
oader.cachebuffer.VerticaCacheBuffer] - flush buffer of partitionId=0, offs
eventInsertedExecutedCount=22, profileMergeAffectedRows=0, profileDeletedE
63, nextOffset=4257760.
2016-04-18 00:01:17,652 [StreamEventThread-for-partition-0] INFO [com.sens
oader.worker.StreamEventProcessor] - isFlushed=true, loadedToBufferCount=22
ileCacheBufferSize=0, nextOffset=4257760, partitionId=0
2016-04-18 00:01:18,211 [StreamEventThread-for-partition-2] INFO [com.sens
oader.cachebuffer.VerticaCacheBuffer] - flush buffer of partitionId=2, offs
eventInsertedExecutedCount=20, profileMergeAffectedRows=0, profileDeletedE
56, nextOffset=4259532.
2016-04-18 00:01:18,215 [StreamEventThread-for-partition-2] INFO [com.sens
oader.worker.StreamEventProcessor] - isFlushed=true, loadedToBufferCount=20
ileCacheBufferSize=0, nextOffset=4259532, partitionId=2
2016-04-18 00:01:18,417 [StreamEventThread-for-partition-1] INFO [com.sens
oader.cachebuffer.VerticaCacheBuffer] - flush buffer of partitionId=1, offs
eventInsertedExecutedCount=8, profileMergeAffectedRows=0, profileDeletedEx
5, nextOffset=4269790.
2016-04-18 00:01:18,422 [StreamEventThread-for-partition-1] INFO [com.sens
oader.worker.StreamEventProcessor] - isFlushed=true, loadedToBufferCount=8,
```



3

业务  
数据

导入工具

进、销、存  
订单配送状态  
售后服务信息

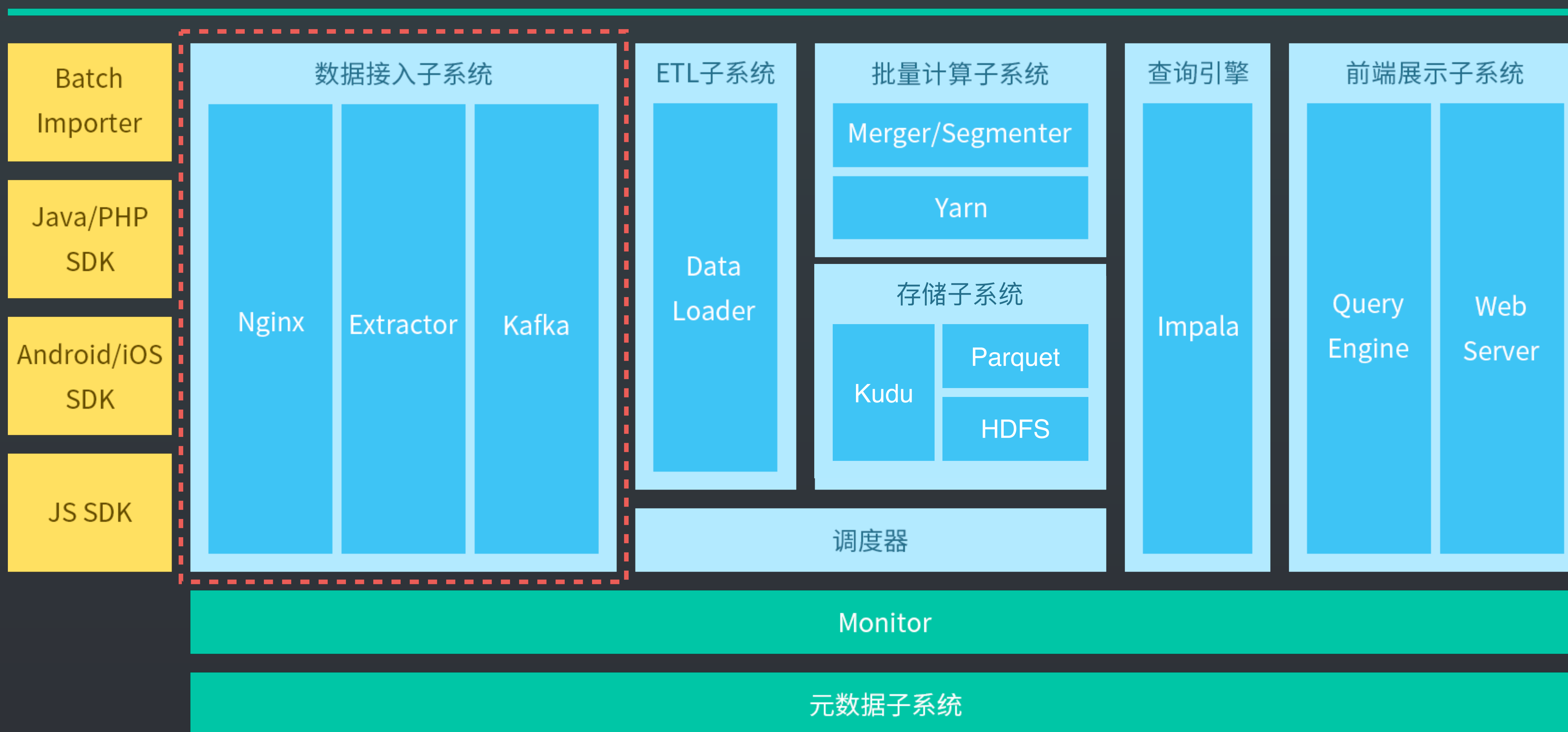
app @fgpbilling (114.112.58.158) - 表													
文件 编辑 查看 窗口 帮助													
导入向导 导出向导 筛选向导 网格查看 表单查看 备注 十六进制 图像 升序排序 降序排序 移除排序 自定义排序													
id	name	type	companyid	appkey	memo	serviceqq	servicetel	serviceemail	analyticsid	notifyurl	authurl	payid1	paykey1
1	系统内部应用	0	30	LzpdGvJ3XQZymRq	系统内部应	123456789	010-66887788	test@test.com	(Null)	(Null)	(Null)	(Null)	(Null)
50	测试Demo	1	30	zmB8wm7sXK9U4or	测试Demo	123456789	010-66887788	test@test.com	3	http://api.issgame.c	http://api.issgame.c	1003050000000110	OUZEQkE5NkM5RTU
100	测试应用	1	30	mI6n3DV98bfIZs4T	测试应用	123456789	010-66887788	test@test.com	20	http://staging-gojh-	(Null)	(Null)	(Null)
300	Demo应用	1	30	K9U4orYzmB8wm7s	掌聚互动D	123456789	010-66887788	test@test.com	3	http://api.issgame.c	http://api.issgame.c	1003050000000110	OUZEQkE5NkM5RTU
6000	豪胆德州扑克	8	1000	rB3uoI7mz682m2Qh	豪胆德州扑	123456789	010-66887788	test@test.com	100	http://www.pokersc	(Null)	(Null)	(Null)
6010	水游骂到	8	1200	blCFsV0r4EHVald8	已终止合作	123456789	010-66887788	test@test.com	101	http://114.66.192.22	(Null)	(Null)	(Null)
6011	水游骂到Test	8	1200	UFcFs4n4EH0Klp1r	已终止合作	123456789	010-66887788	test@test.com	105	http://114.66.192.22	(Null)	(Null)	(Null)
6020	悟空去哪儿	8	1300	Iv2RUUnUhkyTsVhN6	口袋梦幻,+	123456789	010-66887788	test@test.com	102	http://www.gamehe	(Null)	3000183652	MDQ2NDJCQzNGRk
6021	神魔Q卡	8	1300	c03c32125e6ab260	神魔Q卡	152105487887	010-	523039704@qq.con	(Null)	http://www.gamehe	http://www.gamehe	(Null)	(Null)
6030	无双之王	8	1400	m6PD2Cu8xvA7Kwz	无双之王	123456789	010-66887788	test@test.com	103	http://114.112.58.15	(Null)	(Null)	(Null)
6040	天天西游	8	1500	oLI4UnHdkyTswW8c	天天西游	123456789	010-66887788	test@test.com	106	http://xy1.jt-tech.ne	http://api.issgame.c	(Null)	(Null)
6050	葡京梭哈	8	1300	s0Xg4HiU6Wy9kVks	葡京梭哈	123456789	010-66887788	test@test.com	107	http://www.gamehe	http://www.gamehe	(Null)	(Null)
6060	大掌门	8	1600	gY4GwCJ0pVbsE7Rf	大掌门	800049396	4000821680	service@issgame.co	110	http://114.242.187.2	(Null)	(Null)	(Null)
6070	激斗海贼王	8	1300	Rzy4udTDGX8bVVVi	激斗海贼王	800049396	4000821680	service@issgame.co	111	http://www1.gamehe	(Null)	(Null)	(Null)
6071	激斗海贼王内网	8	1300	TJpQ3UhW0axfuPwt	激斗海贼王	800049396	4000821680	service@issgame.co	111	http://www1.gameh	(Null)	(Null)	(Null)
6072	激斗海贼王F	8	1300	kdTzy4udNXG3ebxc	激斗海贼王	800049396	4000821680	service@issgame.co	111	http://www.gamehe	(Null)	(Null)	(Null)
6080	完美武侠	8	1700	zuDUUk02F0jv7Wsa	完美武侠	800049396	4000821680	service@issgame.co	113	http://gojh.api.joytc	(Null)	(Null)	(Null)
6090	妖姬三国	8	1700	We2F2FgwYA9J9vw	妖姬	800049396	4000821680	service@issgame.co	115	http://bppc.56frien	(Null)	(Null)	(Null)
7000	契约2	8	1800	OtGVueb8EcX97xGs	契约2	800049396	4000821680	service@issgame.co	117	http://payment.wist	(Null)	(Null)	(Null)
7010	超级大明皇	8	1900	Lpyq8e1PsmIK3w3E	超级大明皇	800049396	4000821680	service@issgame.co	119	http://g1.xzygame.c	(Null)	(Null)	(Null)
7020	足球	8	2000	bZk5v0tV9adh0CYB	足球	800049396	4000821680	service@issgame.co	(Null)	http://global.gz.125	(Null)	(Null)	(Null)
7030	豪胆黑杰克	8	1300		豪胆黑杰克	123456789			(Null)	http://www.pokersc	(Null)	(Null)	(Null)
8000	锁链战记	8	1300		锁链战记	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8010	佣兵传奇	8	1300		佣兵传奇	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8020	血族	8	1300		血族	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8030	三国志转珠大战	8	1300		三国志转珠	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8040	捕鱼达人2	8	1300		捕鱼达人2	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8050	三国志威力加强版	8	1300		三国志威力	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8060	三剑豪	8	1300		三剑豪	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8070	果宝三国	8	1300		果宝三国	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)
8080	奇幻射击	8	1300		果宝三国	123456789	010-66887788	service@issgame.co	(Null)	http://api.issgame.c	http://api.issgame.c	(Null)	(Null)

www.wyzyto.com

- 后端导入：
  - 代码埋点
  - LogAgent: 实时监控日志，逐条按照 API 发送
  - BatchImporter: 批量处理文件，用于回溯历史数据
  - FormatImporter: 从格式化数据源读取数据，再通过 API 发送
- API:
  - HTTP(S) + JSON

- ID-Mapping:
  - 不同端可以自定义唯一的用户 ID：如设备 ID、Cookie Id、注册 ID 等
  - 系统内部会有唯一的 user\_id
  - 提供一次性的 track\_signup 接口，将两个 ID 贯通起来
    - 不需要回溯数据，但是应用有限制
    - 功能与性能的折衷

## 数据流向



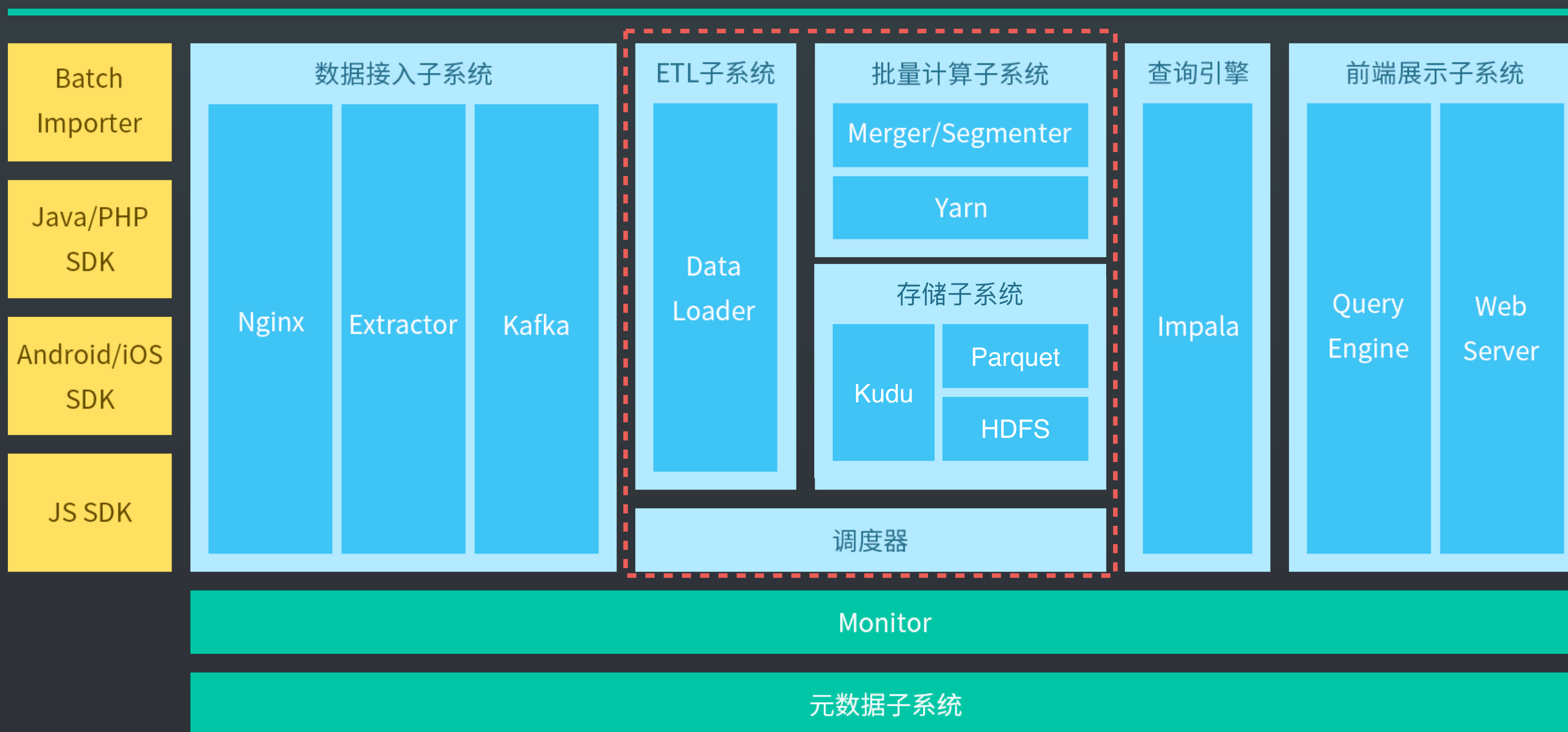
- Nginx:
  - 接收通过 API 发送的数据，并落到日志文件
  - 客户自主选择转发或负载均衡方案
  - 高可靠性
  - 高性能，轻松 5W TPS

- Extractor:
  - 实时读取和处理 Nginx 日志，并将处理结果发布到 Kafka
    - 数据格式的校验
    - 属性类型的识别与相关元数据的操作
    - ID-Mapping 的处理
    - 一些字段的解析工作：IP、UserAgent 等
    - 提供接口，将处理过程开放给客户



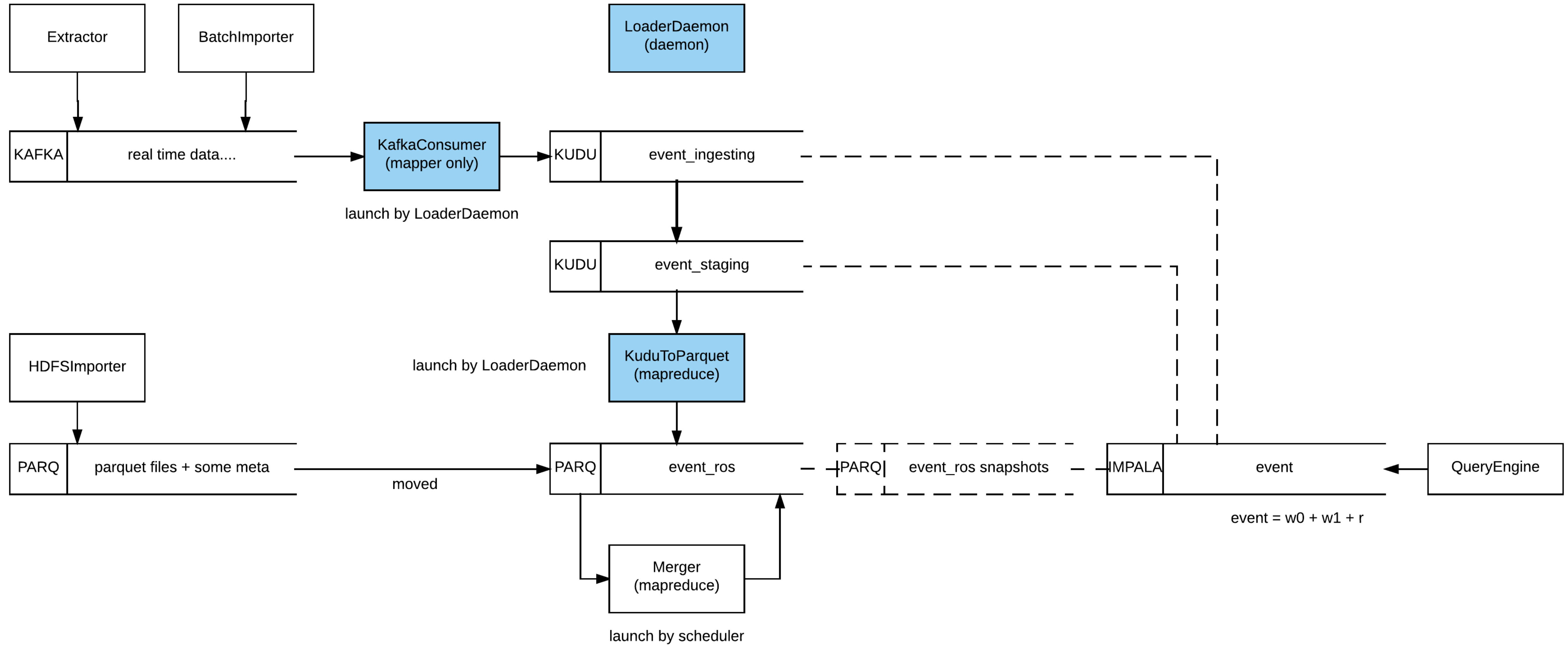
- Kafka:
  - 高可用的分布式消息队列
  - 数据接入与数据处理之间的缓冲
  - 近期数据的一个备份
  - 对外提供访问 API

## 数据流向



- Event:
  - 用户做了什么
  - Event Type + Properties + UserID
  - 追加为主，不可修改，有限的删除能力
- User Profile:
  - 用户是个什么样的人
  - UserID + Properties（年龄、所在地、Tag等）
  - 可以任意修改

- Kudu:
  - 一个新的开源存储引擎，用作 WOS (Write Optimized Store)
  - 写入时会多表逐个切换
- Parquet:
  - 面向分析型业务的列式存储格式，用作 ROS (Read Optimized Store)
  - 按照日期、事件分 Partition
  - 同一个 Partition 内可能会有多个文件，大小保持在 512M 左右
  - 每个文件内部先按照 user\_id 的 hash 有序，再按照 user\_id 有序，最后按照事件时间排序，单独的索引文件记录每个 user\_id 对应行为数据的位置



- KafkaConsumer:
  - 常驻内存的 MapReduce 程序（只有 Mapper）
  - 实时从 Kafka 中订阅数据，并且写入到 Kudu 中
- KuduToParquet:
  - 一个 MapReduce 程序
  - 在 Kudu 单个表写入达到阈值并且不再被写入时，将它转成多个 Parquet 文件，并且移动到对应的 Partition 中
- LoaderDemon:
  - 后台程序，完成一些元数据操作

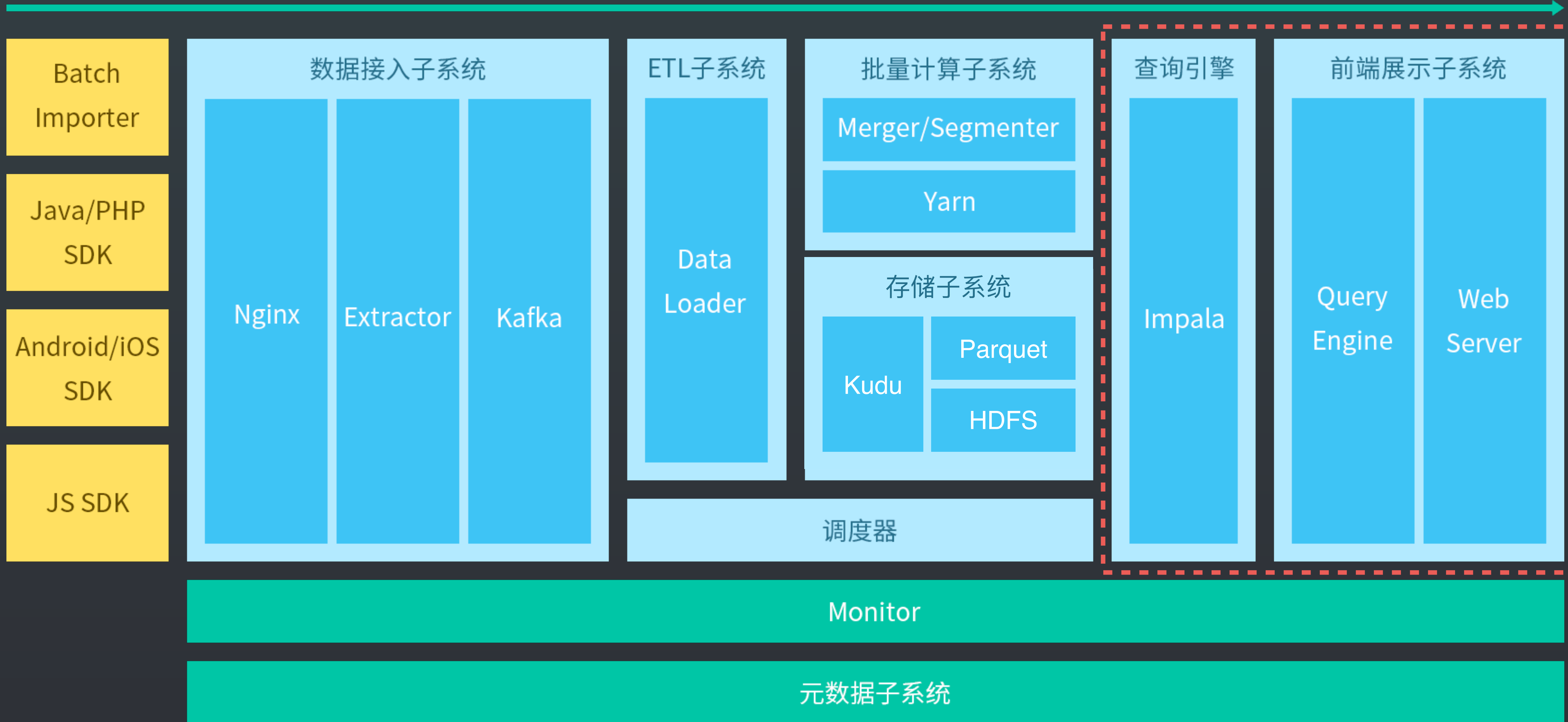


- Merger:
  - 一个定时的 MapReduce 任务
  - 定期合并 Parquet 中每个 Partition 内的碎文件

- Segmenter:
  - 一个临时的或者定时的 MapReduce 任务
  - 完成例行或者一次性的计算：分群、预测、预警等
- 调度器：
  - 负责例行与临时任务的启动、重试、报警等管理工作

- 用户分群：
  - 根据用户以往的行为，给用户打标签
- 用户行为预测：
  - 根据用户过往的行为，预测将来做某个行为的概率

数据流向

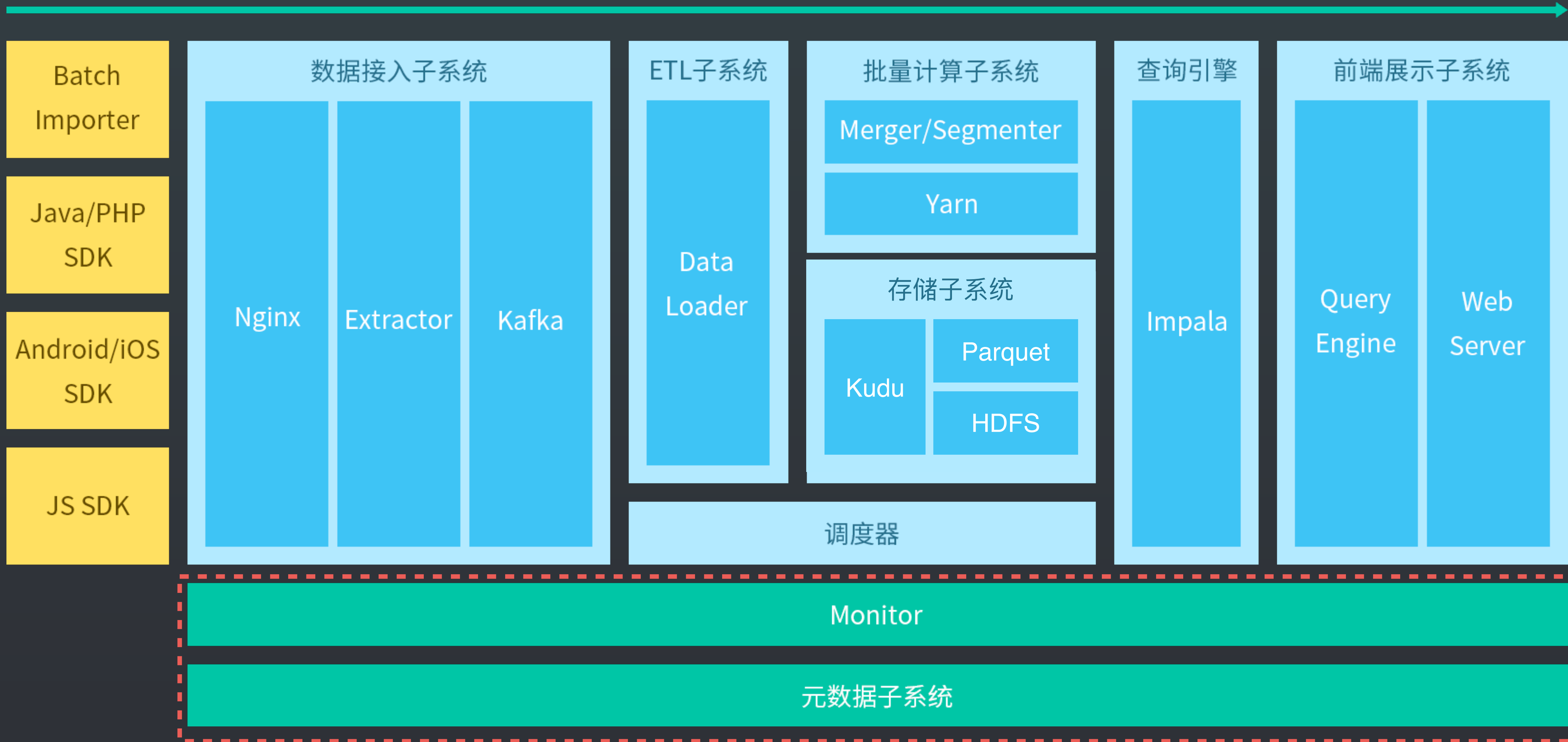


- WebServer:
  - 转发前端的查询请求
  - 提供对外 API
- QueryEngine:
  - 将查询请求翻译成 SQL, 提交给 Impala
- Impala:
  - 访问 Kudu 与 Parquet 的 View
  - 为什么选择 Impala? 为什么不用 SparkSQL?

- 查询优化：
  - UDF/UDAF/UDAnF, 聚合函数替代 Join
  - Partition 与用户分桶, 减少数据扫描量
  - 按用户抽样, 快速尝试, 最终才拿准确结果
  - 缓存, 只对有变化的数据才重查



## 数据流向



- 元数据：
  - MySQL/ZooKeeper/Redis
  - Schema、维度字典
  - 概览、漏斗、分群、预测的配置
  - 任务调度信息
  - 权限信息
  - 应用层缓存

- Monitor:
  - 一个常驻模块
  - 内部模块的语义监控，异常状态的修复
  - 外部组件的监控与报警
  - license 的处理

- 各种运维工具：
  - 一些不常用的功能，不提供界面，而是工具化
  - 数据清理工具
  - 版本升级工具
  - 性能分析工具
  - 多项目管理工具

# 神策分析，帮你实现数据驱动！



# 神策分析，帮你实现数据驱动！

