

云智未来⁹_{th}

第九届中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2017

SACC
2017

北京·新云南皇冠假日酒店

IT168.com

ChinaUnix

ITPUB

实时大数据在风控中的实践

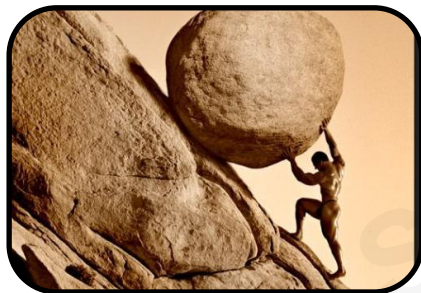
——数美基础架构负责人 关涛



大纲



背景介绍



面临挑战



业务分析



架构设计



经验心得

背景介绍

公司背景介绍

数美致力于构建大数据反欺诈服务，让世界更加“安全、公平、美好”！



个人背景介绍

关 涛

数美技术总监、基础架构负责人

近十年大数据基础架构相关工作

北大计算机硕士



项目背景介绍

主要涉及两个领域：互联网反欺诈、互联网金融风控



+



面临挑战

面临挑战

业务繁杂



数据巨大



时效极高



挑战一：业务繁杂

支撑两大行业

金融

- 银行、P2P、消费金融、信贷、理财等

互联网

- 电商、直播、社交、资讯、论坛、游戏等

提供多种产品

金融风控

- 信贷、中介、逾期、可信度分析

欺诈识别

- 垃圾文本识别、垃圾图片识别
- 设备指纹、注册、登录、薅羊毛

服务千家公司

金融

- 中信银行、宜信、百度金融、360借条、国美金融等

互联网

- 小米、360、爱奇艺、youku、斗鱼等

挑战二：数据巨大

账号数量大

- 黑产掌握了数十亿的账号，所衍生的黑产业年获利超百亿。—— 2017网络安全生态峰会发布《电子商务生态安全白皮书》

从业人数多

- 黑产从业人数超过150万人，市场规模高达千亿级别。——《电子商务生态安全白皮书》

涉及行业广

- 金融、电商、直播、论坛、社交、游戏等。——《2017黑产行业分析报告》

风控数据大

- 对接千家客户风控需求，每天请求量10亿左右，涉及亿级别的设备与账号——《2017数美业务分析报告》

系统请求大

- 每秒请求上万QPS，高峰期十万QPS——《2017数美系统分析报告》

画像数据大

- 亿级别设备、账号画像、手机号库、联系方式库、IP库——《2017数美数据分析研究报告》

挑战三：时效极高



黑产实时变化

- 黑产实时欺诈
- 大量实时攻击
- 黑产实时变策

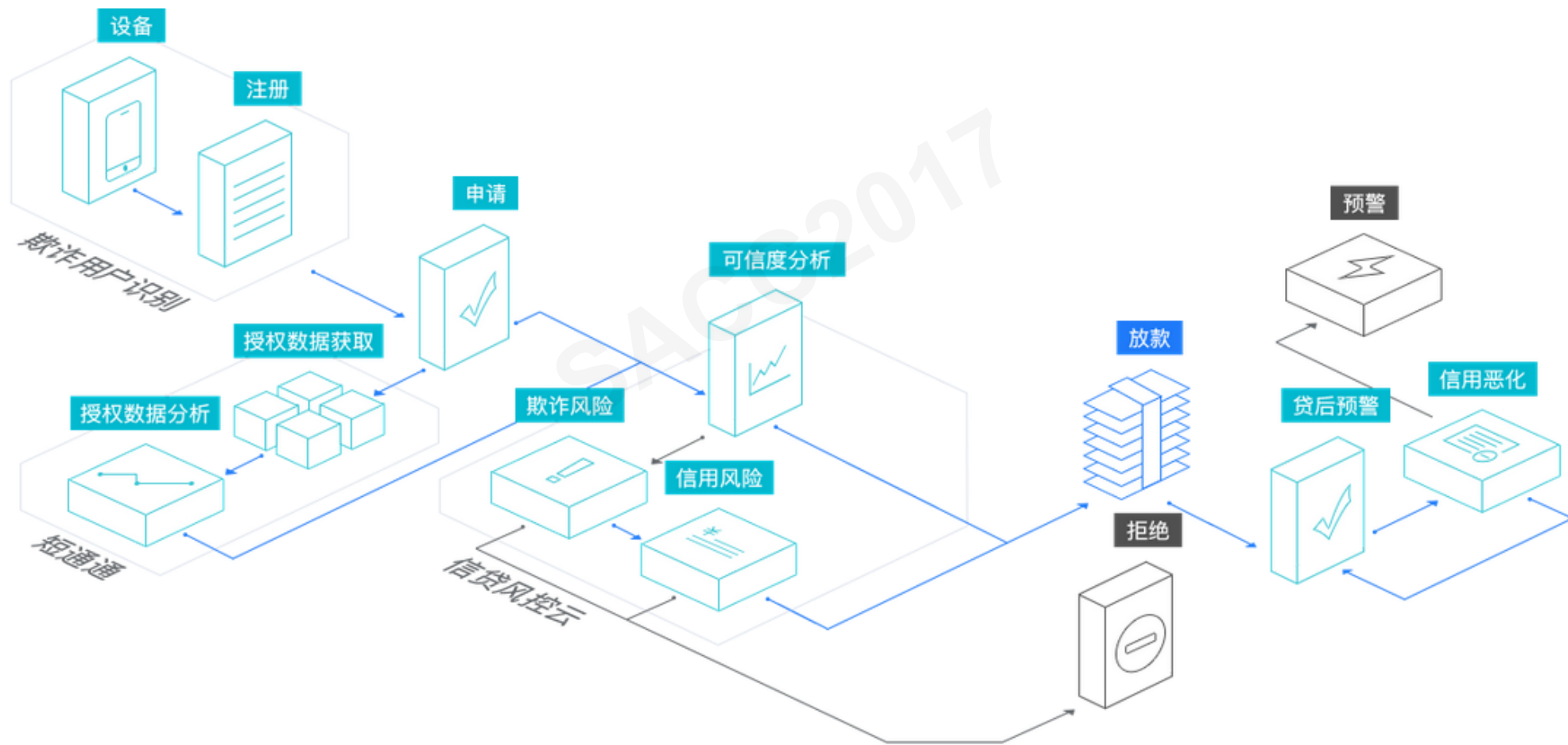
数美实时风控

- 业务需求及时响应
- 策略及时生效上线
- 策略实时动态调整

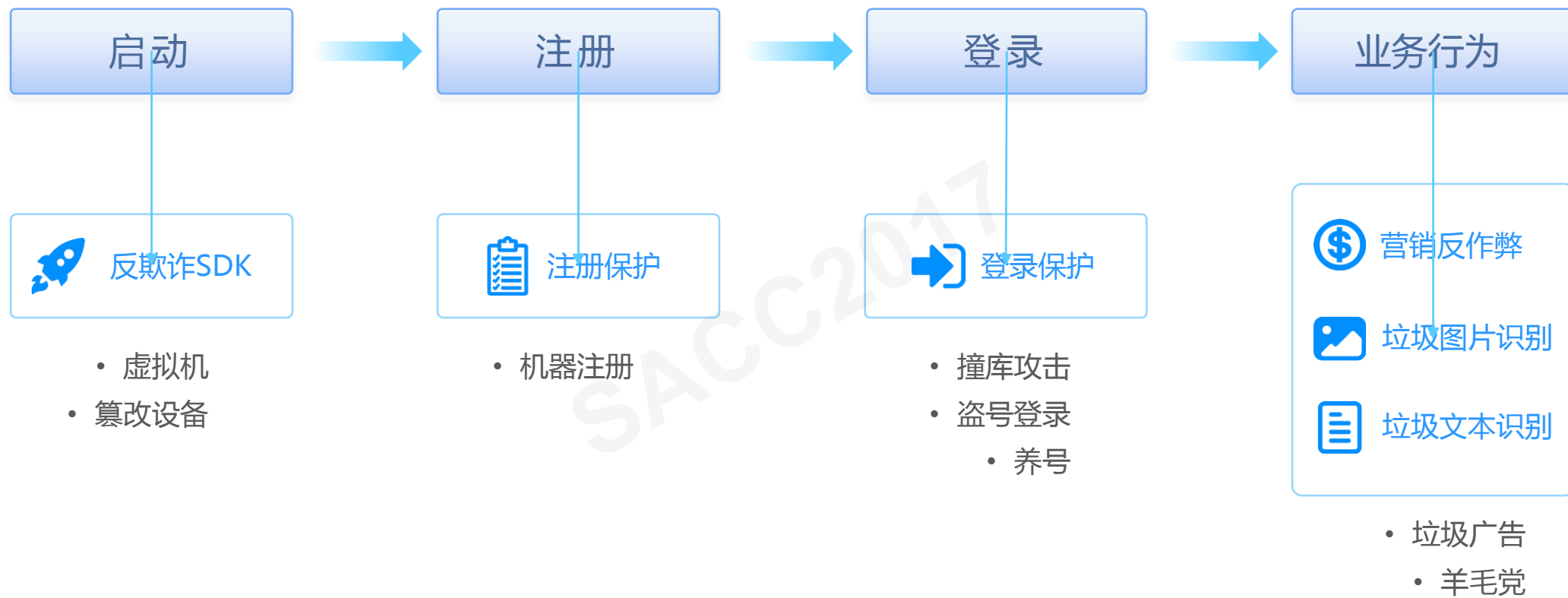
业务分析

金融风控流程分析

全业务流程风控体系



互联网风控流程分析



架构设计

架构设计——设计前的分析

业务繁杂——化繁为简、灵活配置

- 化繁为简：抽象业务共同核心逻辑、集中力量解决核心问题
- 灵活配置：为各业务公司构建灵活个性化可视化配置界面

数据巨大——系统易扩、工具平台

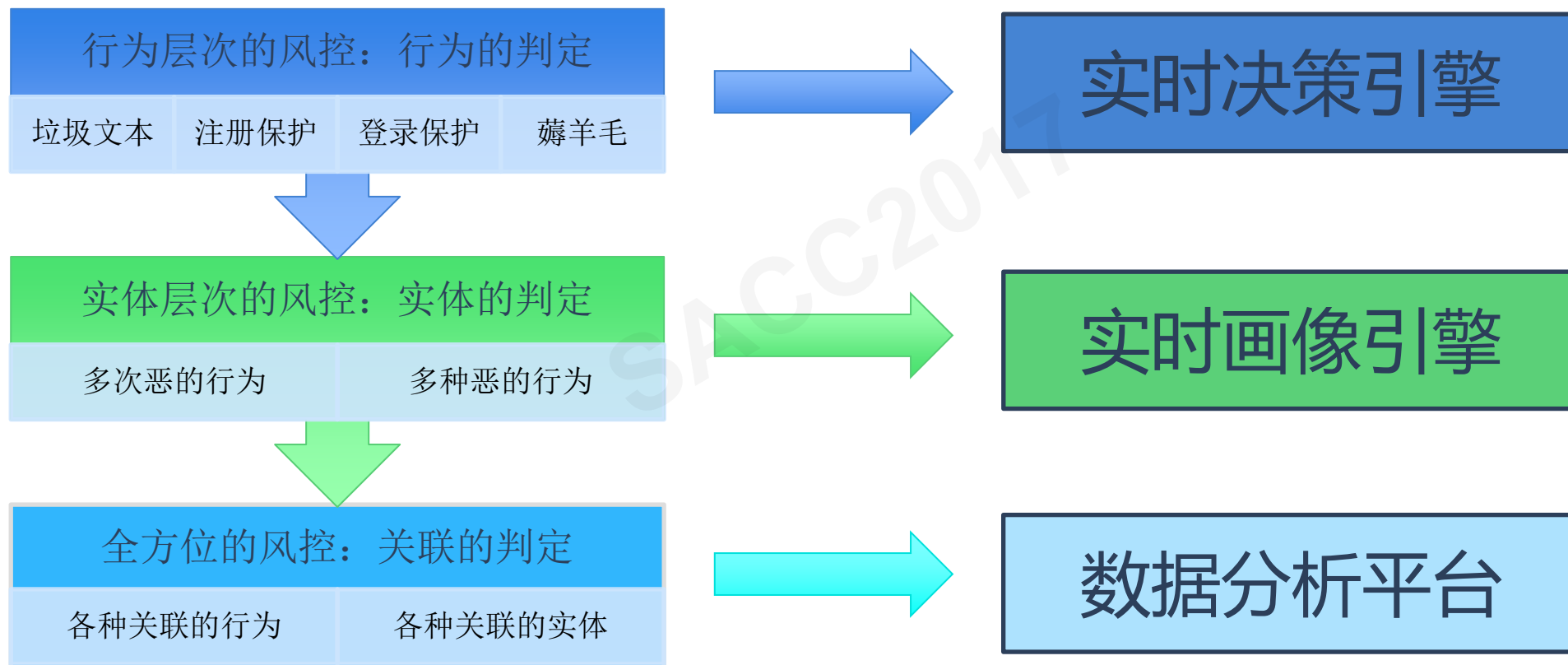
- 系统易扩：解决所有系统的单点问题，支持水平扩展
- 工具平台：利用成熟工具搭建大数据平台，提供各种工具支持业务大数据分析。

时效极高——快、快、快

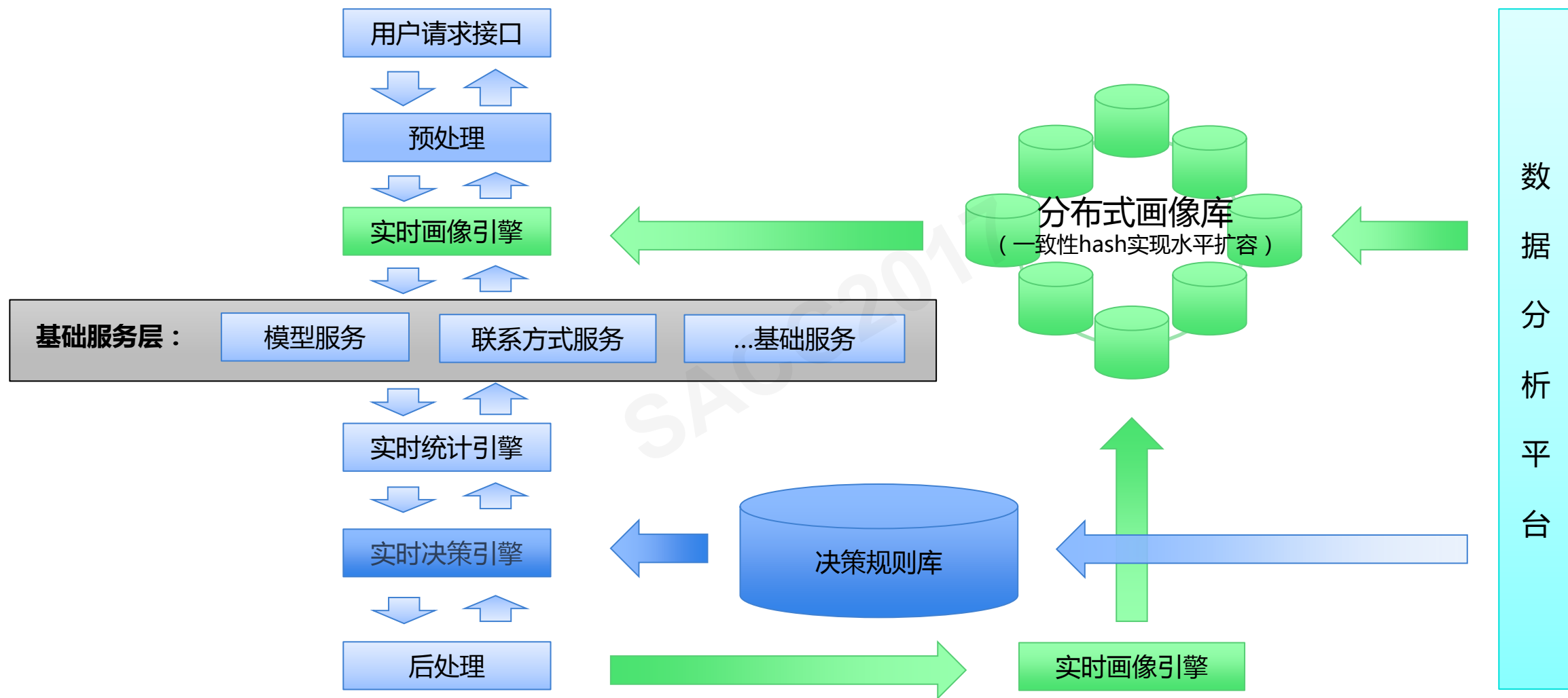
- 策略实时变化：高度可配置的平台、可视化实时分析与配置。
- 业务实时处理：设计实时数据流，实时沉淀画像，系统实时学习。
- 巨量实时请求：构建缓存机制、索引机制、异步机制、容错机制、自降级机制。

架构设计——核心关键

抓住核心关键，即可化繁为简



架构设计——整体设计



实时决策引擎——灵活配置、快速决策

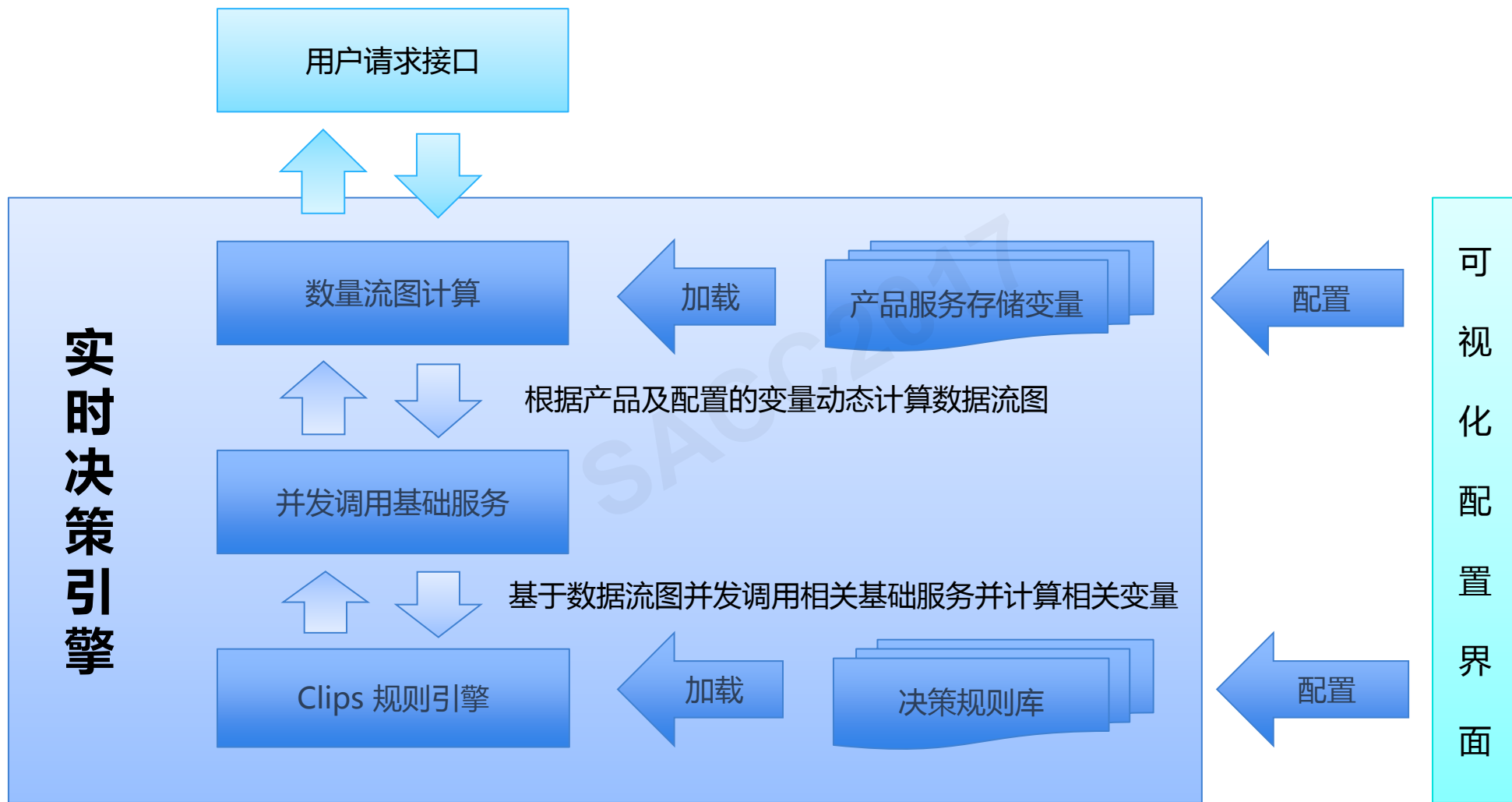
通用性：重要的是灵活

- 不同产品配置不同的数据流图
- 不同场景配置不同的规则策略

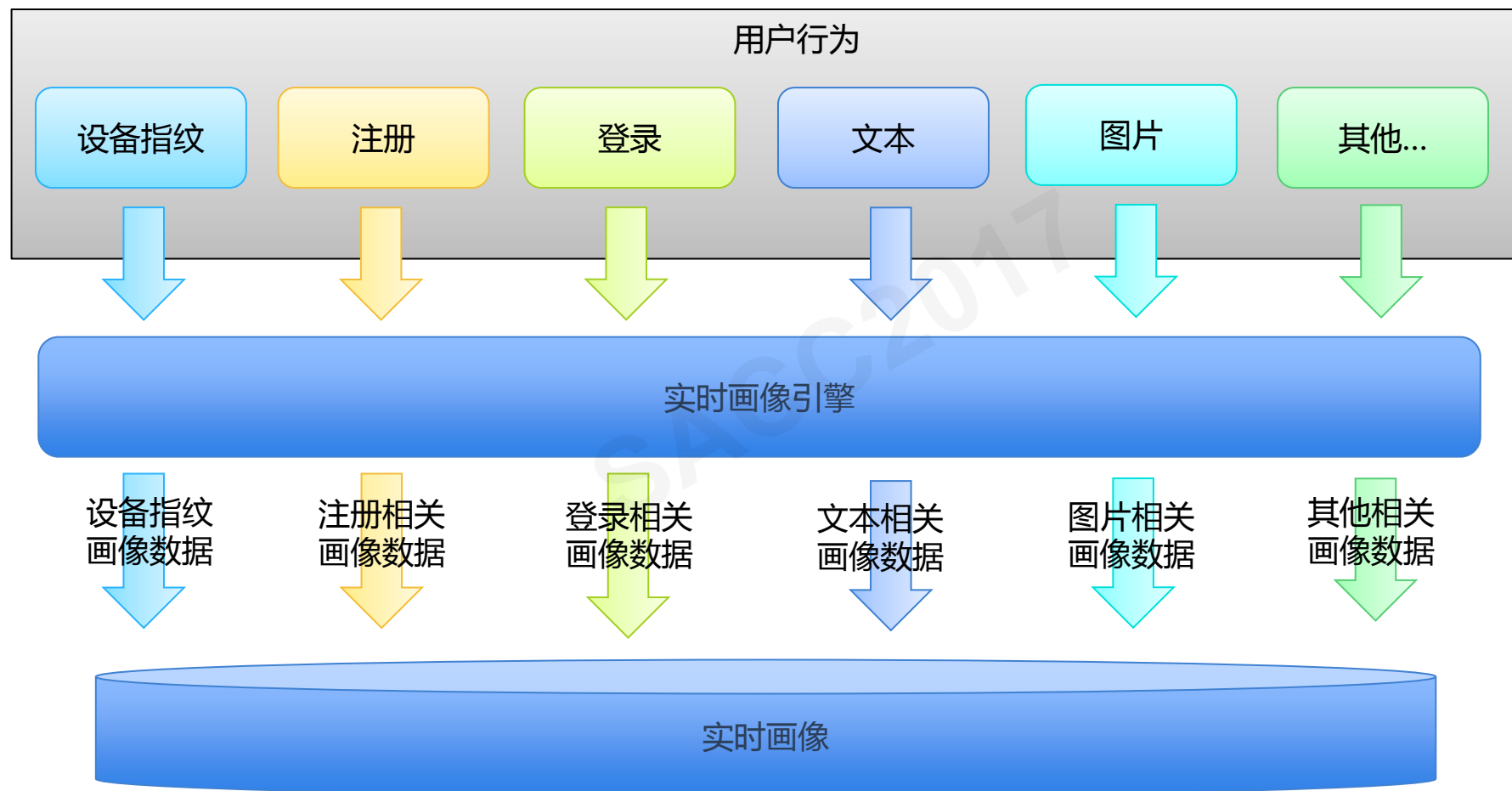
时效性：重要的是快快快

- 系统的决策：实时计算实时决策
- 系统的升级：写代码 => 写规则 => 可视化配置

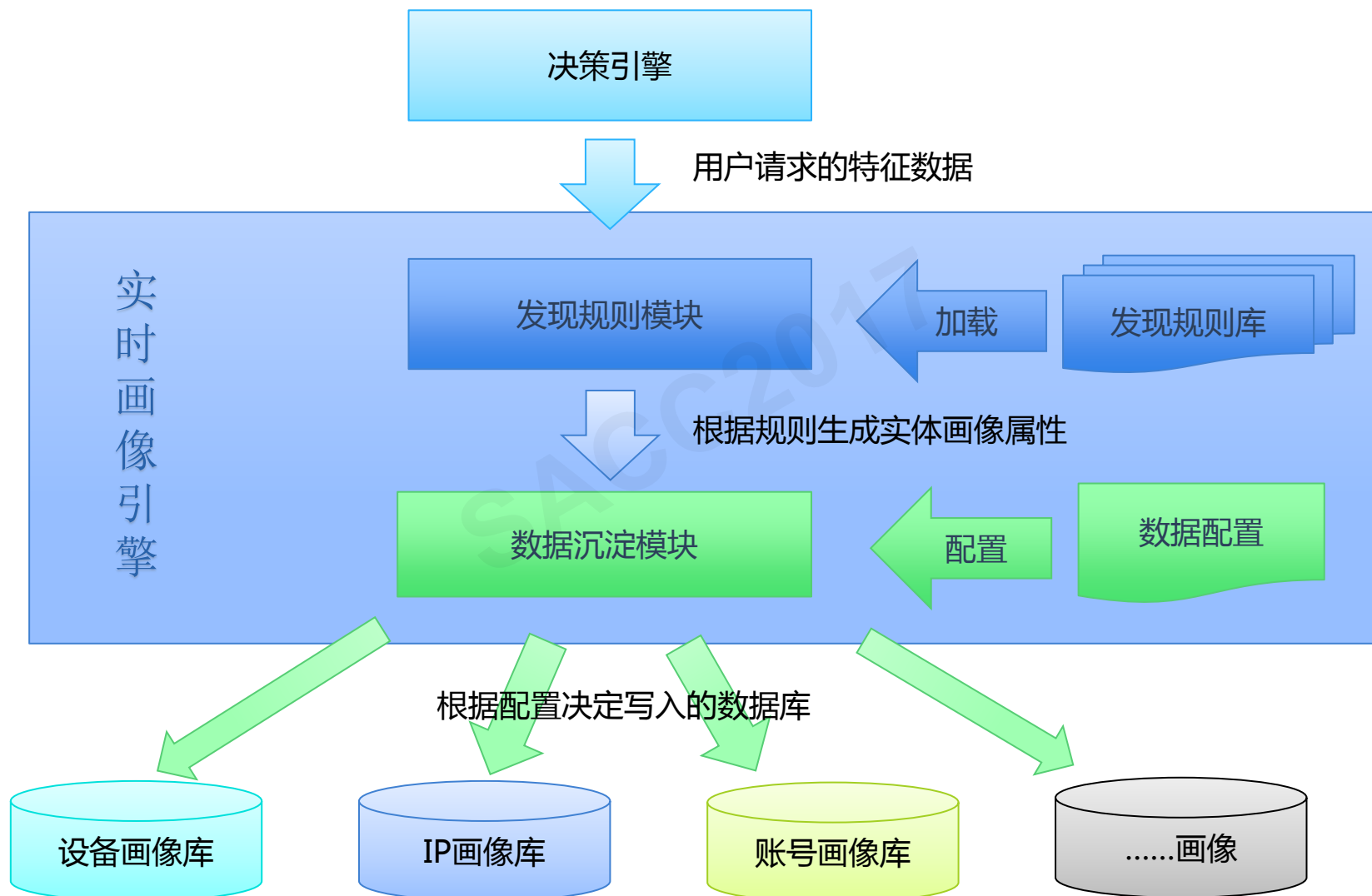
实时决策引擎——详细设计



实时画像引擎——多环节联动联防



实时画像引擎——实时沉淀、构建画像



数据分析平台

业务分析

内容分析

- 广告、色情、涉政

行为分析

- 重复、频度

画像分析

- 静态属性、动态属性、关联属性

系统分析

性能分析

- 分析各服务耗时

稳定性分析

- 分析各服务失败率

超时率分析

- 分析超时率

分析工具

搜索

- 提供搜索工具

聚合

- 关联聚合工具

统计

- 统计工具

数据分析平台——业务分析

构建可视化Saas平台

<input type="checkbox"/>	流水号	渠道	tokenId	设备标识	原因	识别结果	昵称	文本内容	发生时间
<input type="checkbox"/>	4c65a0...	单聊	11383171	201709062355113...	重复发送广告内容	拒绝	娇喘小公举♡	乱伦666	10-07 23:18:05
<input type="checkbox"/>	01e29e...	描述	4078741	20170529231026c...	色情污秽内容	色情污秽内容	哥哥	请求与你复合进进出出操逼呢？	10-07 23:18:05
<input type="checkbox"/>	d44219...	单聊	11516635	201703101801121...	政治敏感内容	拒绝	嘉	我还喜欢习近平	10-07 23:18:04

拉取图片识别 > 历史记录

2017-10-07 00:00:00 至 2017-10-07 23:59:59

<input type="checkbox"/>	流水号	渠道	tokenId	原因
<input type="checkbox"/>	0c0b74e9aa21c27...	聊天图片	10741061	广告
<input type="checkbox"/>	031f179862859d...	聊天图片	11521592	重复发送图片
<input type="checkbox"/>	cf39c650983fdd15...	聊天图片	11516221	广告

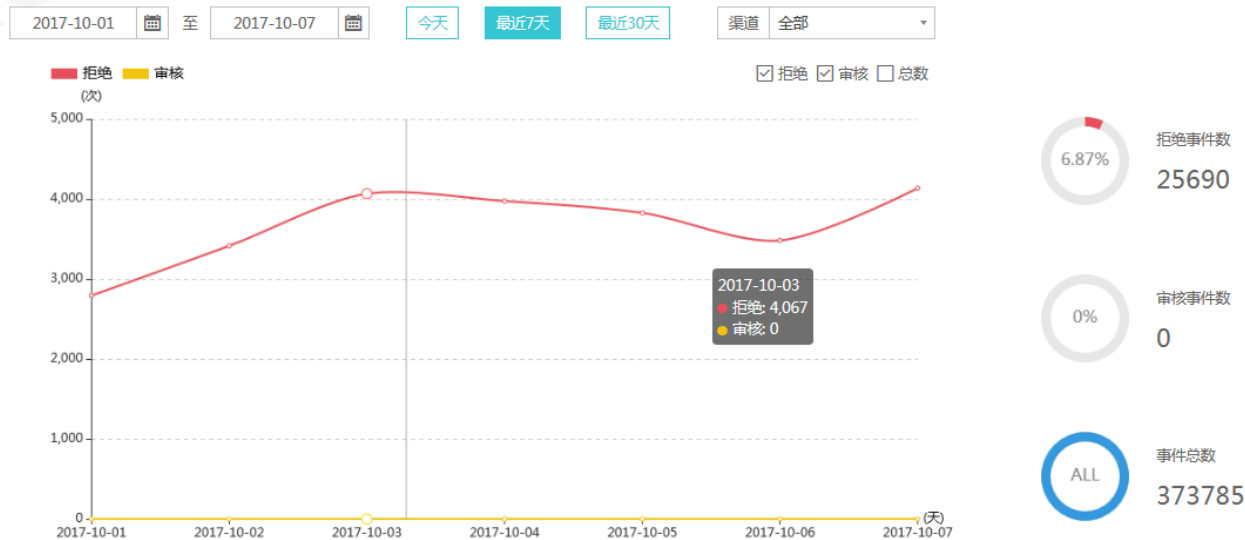
中国移动 22:16

< 列表

全部筛选 今天 22:16 09

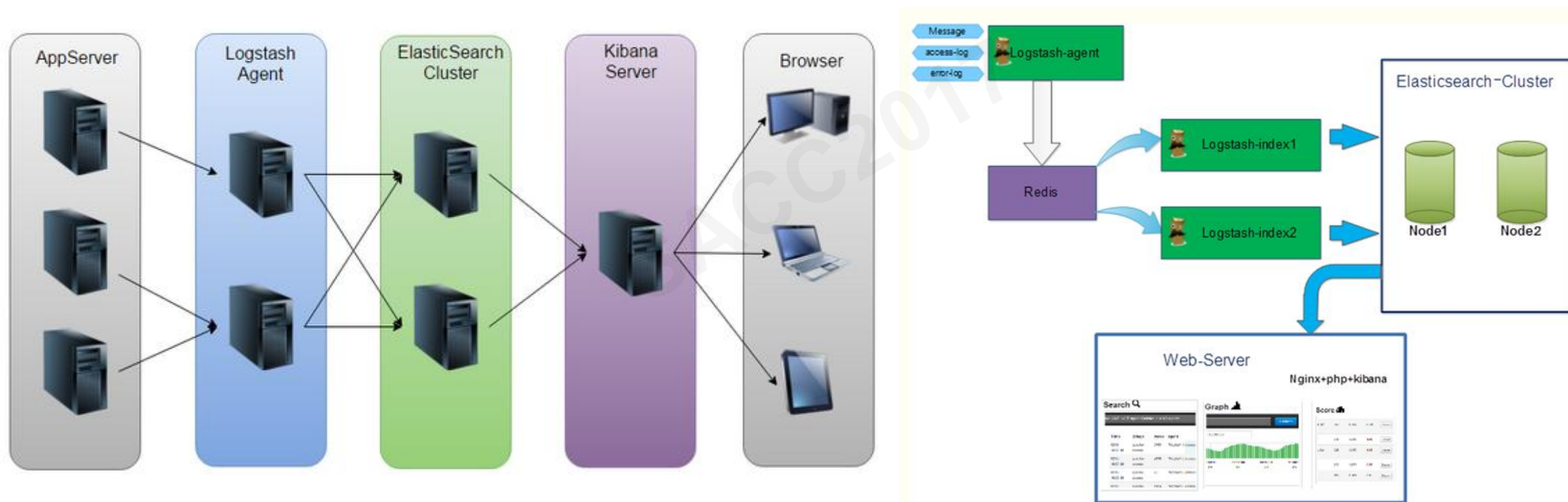
你好，低价小视频一元一部，买三送一，买五送二，另有百度云 天翼云 裸持 月光宝盒 直播盒子等资源出售！有意的朋友加我 QQ3466857498

图片缩略图

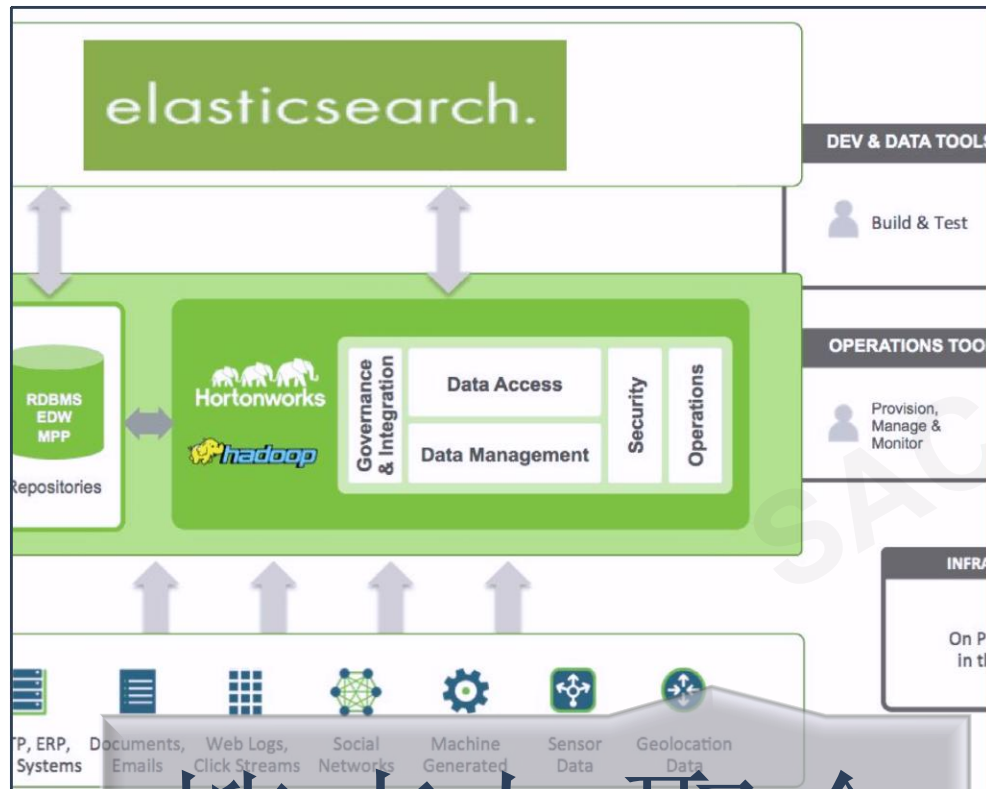


数据分析平台——系统分析

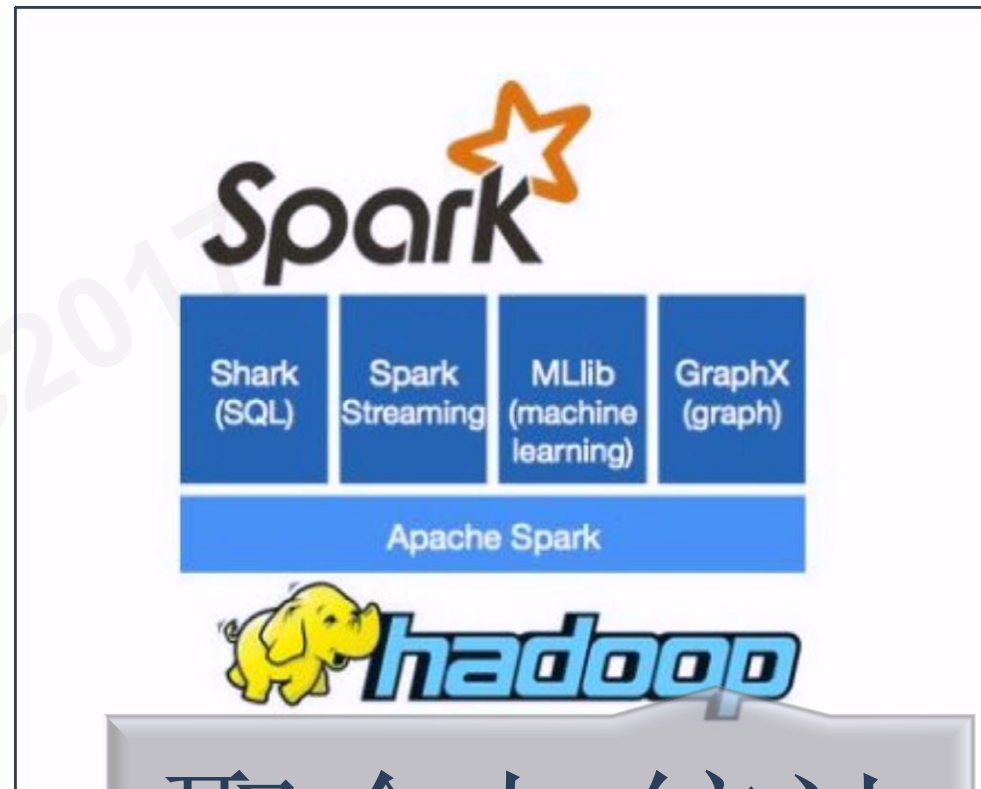
ELK(ElasticSearch, Logstash, Kibana)搭建实时日志分析平台



数据分析平台——分析工具



搜索与聚合



聚合与统计

经验心得

经验与心得



简约设计：化繁为简

- 避免过度：不过早引入额外系统、额外资源
- 灵活可配：可以通过配置的方式实现各种变化
- 扩容简单：消灭结构性单点，可水平线性扩容



优化性能：极速体验

- 尽量缓存：缓存避免重复计算大大提升性能
- 设计索引：合理索引结构可提升服务性能
- 异步优化：避免同步阻塞影响线上服务性能



便捷运维：稳定根基

- 多数据中心：异地多活
- 熔断自降级：高压下系统各个服务自降级避免雪崩
- 多维度监控：可视化工具实时监控



抓住核心：事半功倍

- 优先分析：一定要先分析清楚核心问题，再对症下药
- 不要瞎忙：无关痛痒的问题一律不做
- 聚焦精力：80%的精力要投入到最核心的事情上

/ THANKS /

