

汽车之家数据平台架构

DTCC

2015中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2015

大数据技术探索和价值发现



小思考：

互联网企业数据仓库构建是采用
自下而上的方式，还是自上而下的方式？

如果你是一个数据部门的架构师，
你怎么去规划数据仓库呢？



架构师必须要面对的两个方面



该怎么办？



一切以实现数据价值为目标

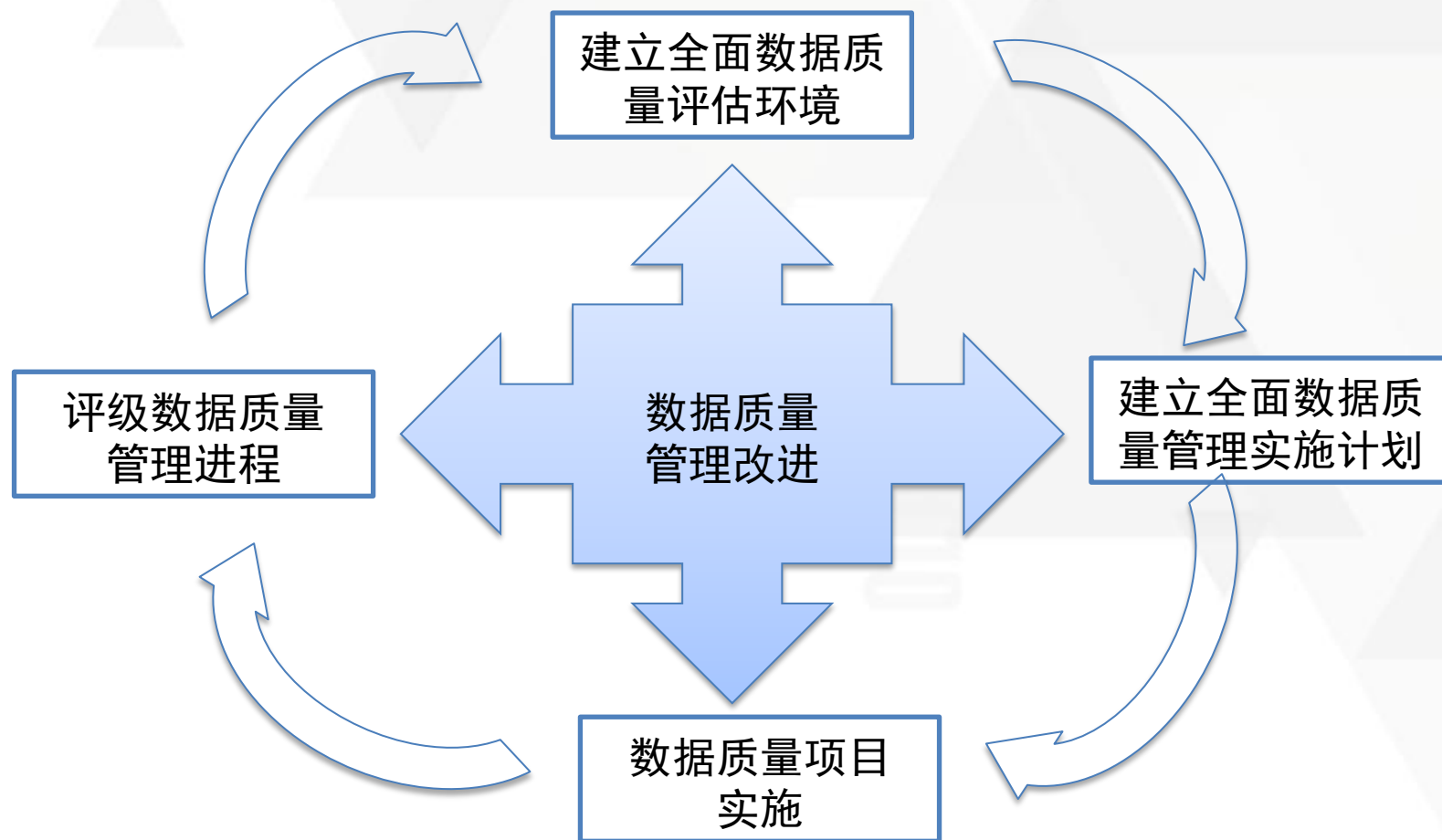


如何实现数据价值

数据价值



数据价值的保障-数据质量



数据质量委员会

数据标准管理

数据标准规范管理

业务数据
标准规范技术数据
标准规范

数据标准流程管理

流程
定义流程
变更

数据标准执行

标准
执行标准
监控

系统管理

版本
管理标准
查询标准
维护

数据质量系统

数据质量检核

检核定义

检核引擎

检核分析

质量过程管理

需求管理

知识管理

问题管理

数据质量监控

数据监控

安全监控

系统监控

数据质量报告

流量数据报告

业务数据报告

元数据系统

元数据基础管理

技术元数据管理

报表
元数据ETL元数
据源系统元
数据数据仓库
元数据

业务元数据管理

业务模型
管理业务指标
管理元数据映
射

计算规则

主数据

版本管理

视图管理

变更管理

生命周期管理

元数据访问管理

元数据查询

元数据采集

人工录入

自动采集

模板导入

元数据分析

核心数据分析

影响分析

血缘分析

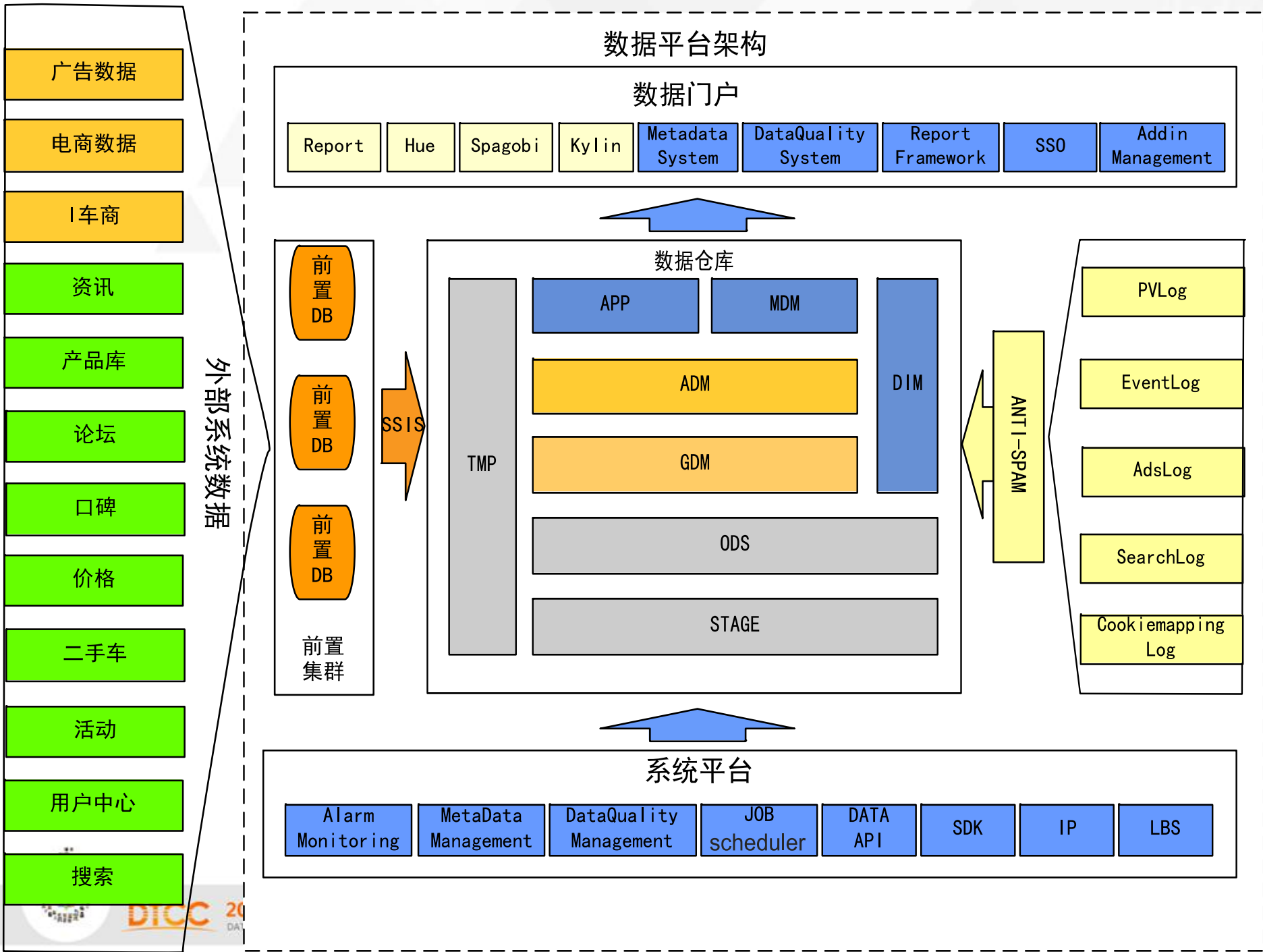
活力分析

元数据检核



实现数据价值的必经之路-平台化





系统架构

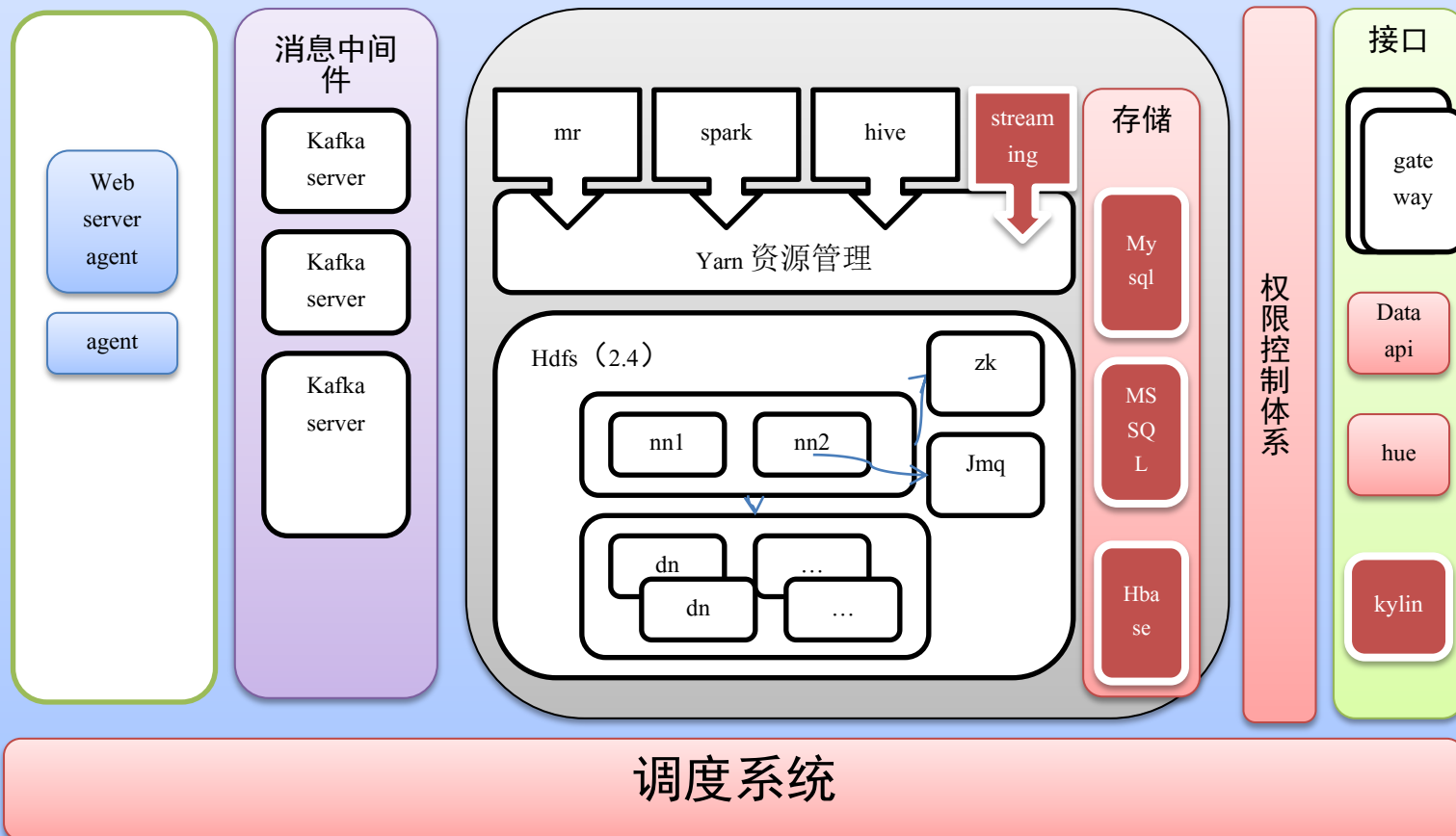
收集

采集

存储计算

应用

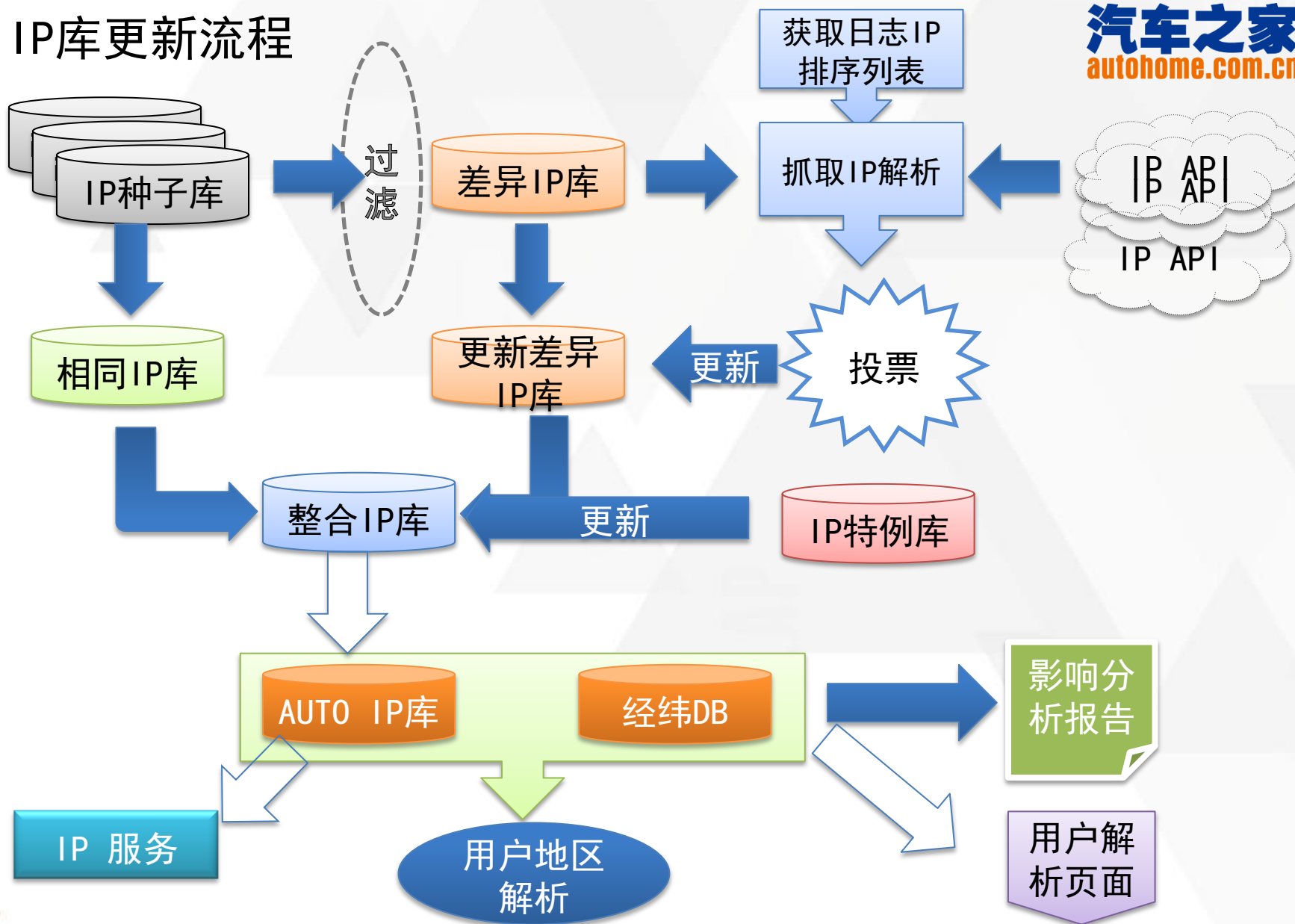
监控系统



个案分享



IP库更新流程



Cookie 技巧

Cookie_用户标识ID = UID (36位) + First Datetime + First ReferDomain
标识用户；标示用户第一次访问时间；标示用户第一次访问来源；
提升统计分析效率；

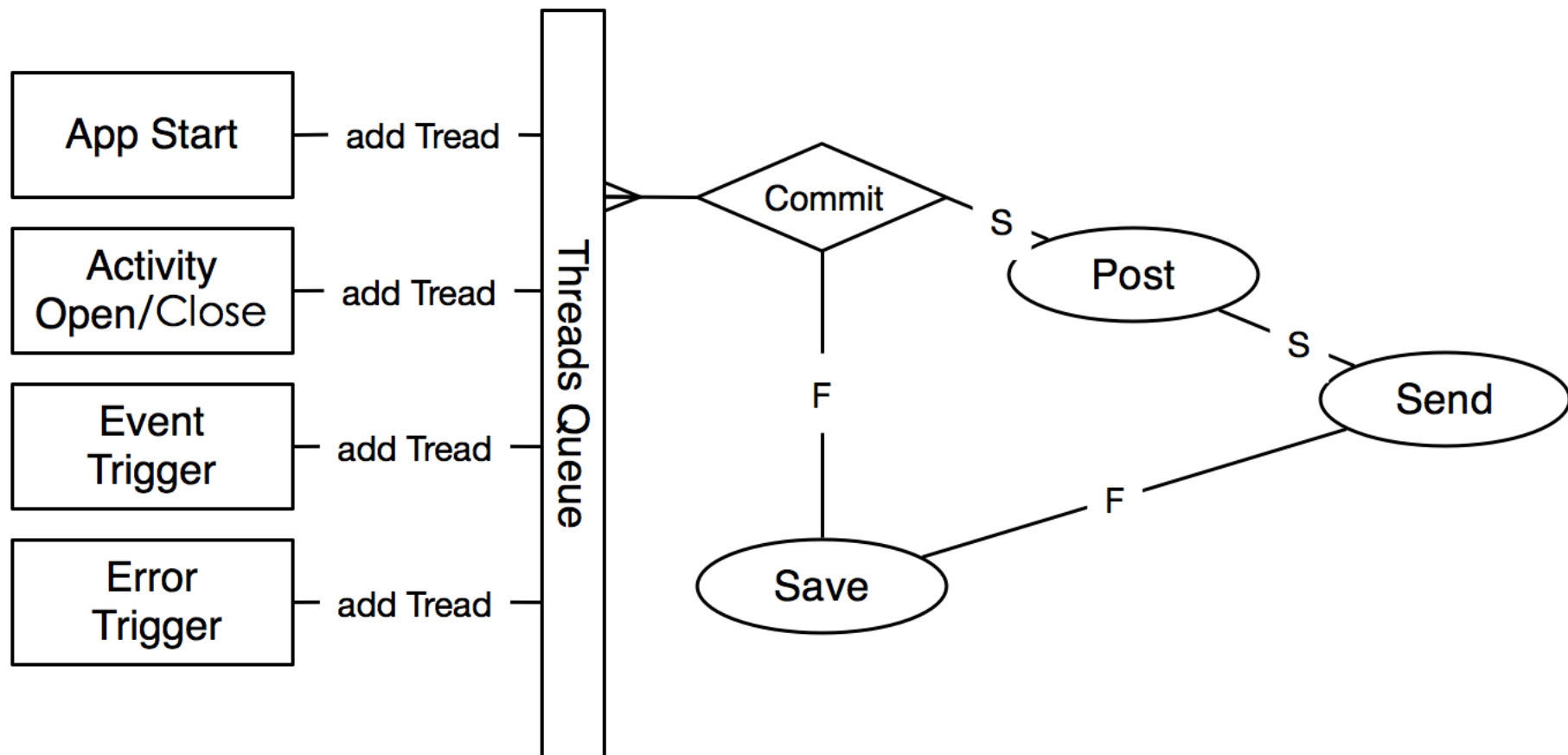
Cookie_Evercookie = Cookie_用户标示ID
恢复用户标识ID；
使用户跟踪更精准；

Cookie_fingerprint = number
评估用户规模；防作弊；

Cookie_ref = 上次访问来源 + 本次访问来源
用于提升统计分析效率



APP统计SDK日志采集模型

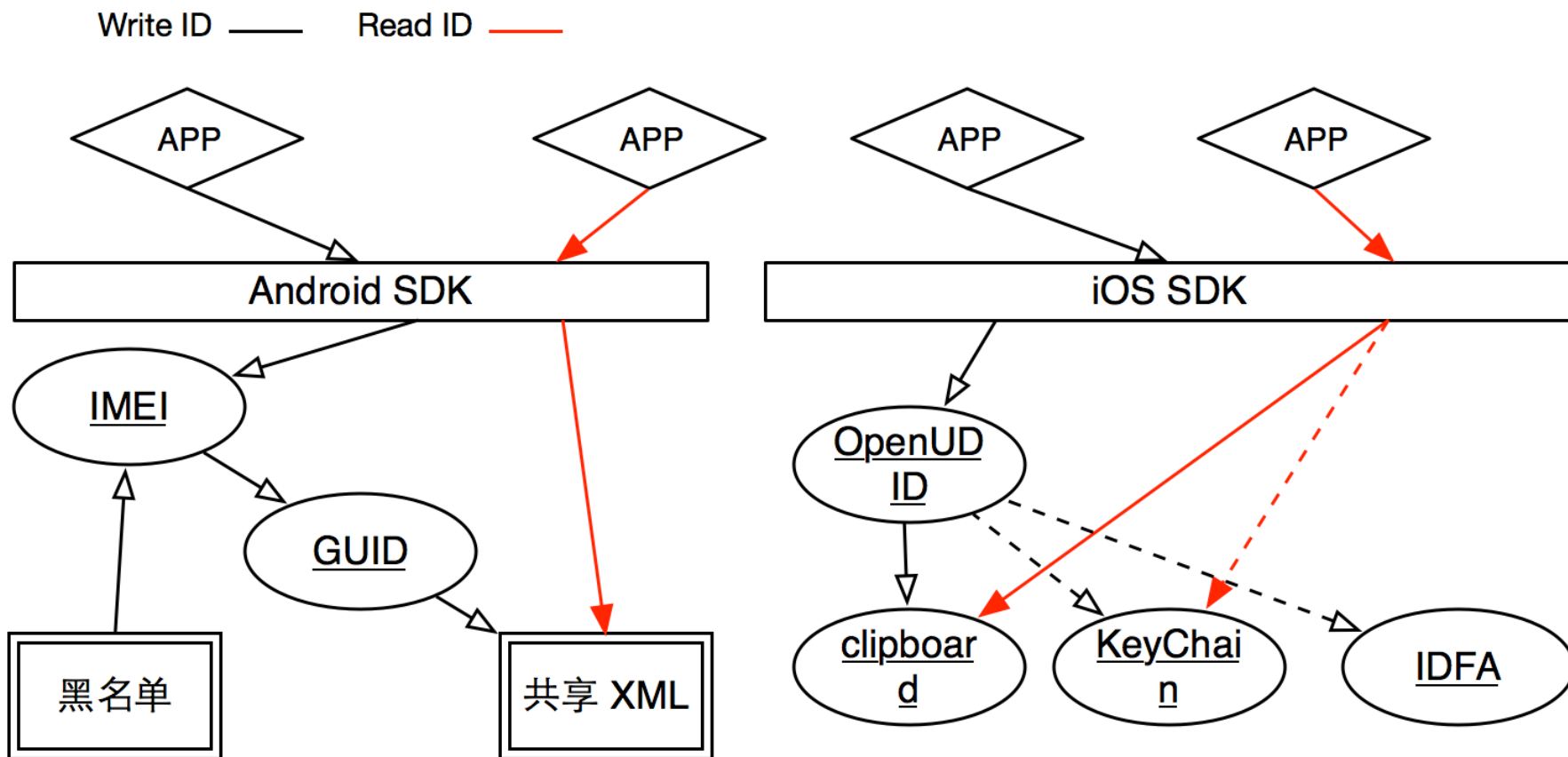


移动设备唯一标识问题

- 冲突与漂移：
 - 冲突——多个设备获取的UUID相同
 - 漂移——单个设备产生多个UUID
- Android IMEI 缺陷
 - (1) 山寨
 - (2) 权限限制
 - (3) 如何多app共享
- iOS OpenUDID 缺陷
 - (1) 剪贴板共享被限制，如果证书前缀一致，可以共享
 - (2) Keychain要求证书前缀一致
 - (3) IDFA可能被用户禁用，存在被被app store拒风险
 - (4) MAC已被禁用 (iOS7)



解决冲突和漂移的方案



欢迎加入汽车之家

高红锋 数据平台部门经理

email:

gaohongfeng@autohome.com.cn

微信: sdkf2014122



IT168... ChinaUnix IT PUB
THANKS