

# 广告系统中的大数据处理

时趣互动 移动广告架构师 宋慧庆

# 个人简介

- 十余年互联网老司机，具有多年互联网广告系统研发经验，负责研发个性化动态创意系统，用户画像标签系统以及大数据治理；对编程语言有浓厚的兴趣。



流量为王 -> 数据为王

# 什么是大数据？

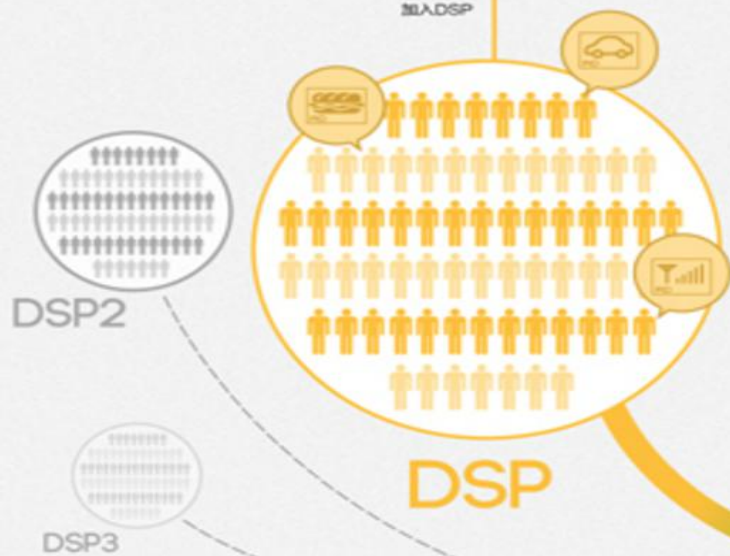
- 从具备4V（Volume，Velocity，Variety，Value）特征的大量数据中挖掘用户的潜在价值



# 精准广告的复杂性

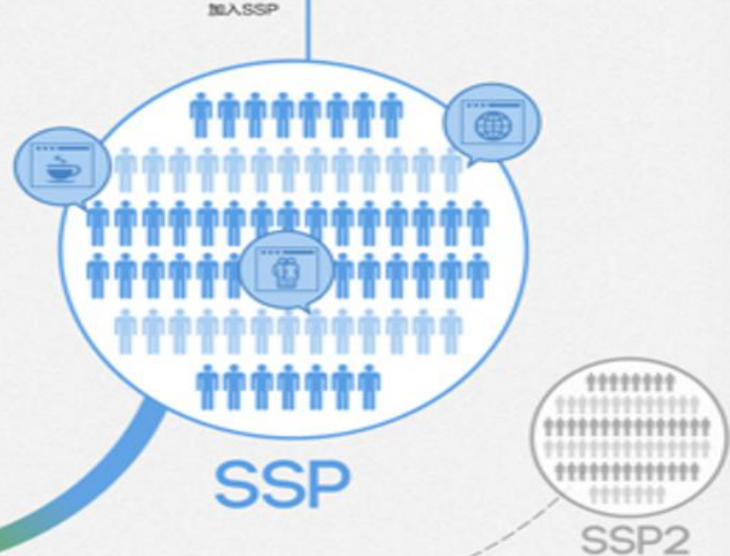
## DSP 需求方平台 Demand Side Platform

互联网里有成千上万的广告主，他们急需推广自己的产品，寻找优质的媒介和精准的目标用户，优化广告投放策略，提高投入产出比。这种情况下，就诞生了为她们提供服务的专业化平台，DSP。简单的讲，DSP就是广告主服务平台，广告主可以在平台上设置广告的目标受众、投放地域、广告出价等等。



## SSP 供应方平台 Supply Side Platform

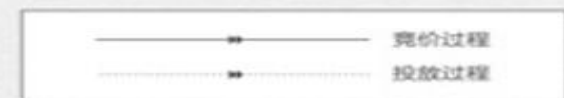
互联网里也有成千上万拥有丰富媒体资源和用户流量的网站，他们急需把庞大的流量变现来发展壮大，同时还希望每一个流量都能够达到最大的收益。简单的讲，SSP就是一个媒体服务平台，站长们可以在ssp上管理自己的广告位，控制广告的展现，设置补余等等。



有了买方，也有了卖方，所以就需要有个中间方联系上下游，AD Exchange为广告主提供下游SSP委托管理的广告位和流量，同时对每个流量都在上游的DSP广告主方进行实时竞价，保证每个广告位的价值都能最大化。



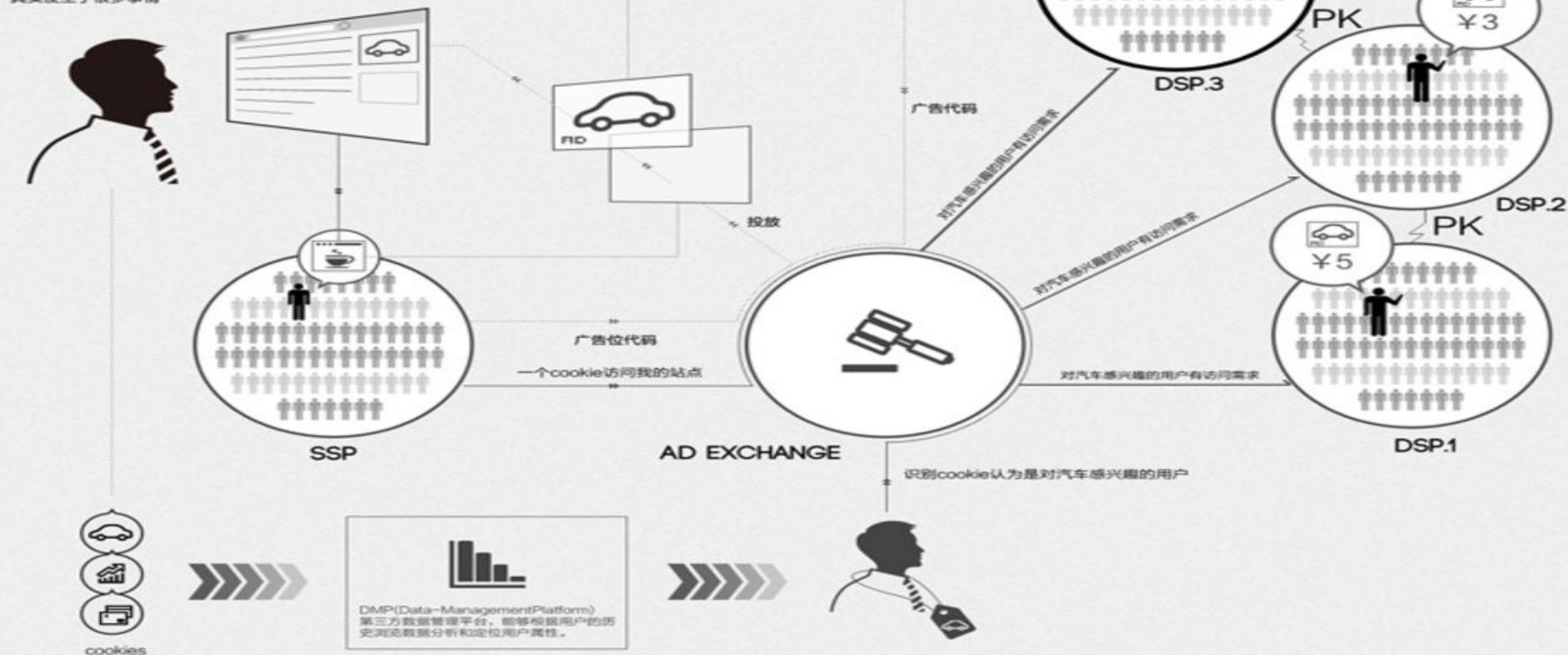
# AD EXCHANGE 模式



当用户浏览一个加入SSP的站点时，  
其实发生了很多事情...

获胜获得展示机会

**RTB** (real time bidding)  
实时竞价。针对每一个访问需求都在DSP  
端进行竞价展现。让每一次展现都实现利益  
最大化。





# 广告系统的数据来源

- DSP监听数据（主要是ssp，adx渠道）
- 广告主数据
- 广告投放过程中收集的数据
- 购买 / 通过交换得到的第三方数据
- 运营商合作数据

# 精准广告的系统特点

- 数据量巨大（广告本身的数据，定向数据，用户特征数据，广告展示环境的特征数据）
- 响应速度要求特别快
- 实时性要求特别高（广告数据，定向数据，用户特征数据，广告展示环境的特征数据等实时性）
- 系统可用性要求特别高（柔性降级，可扩展性）



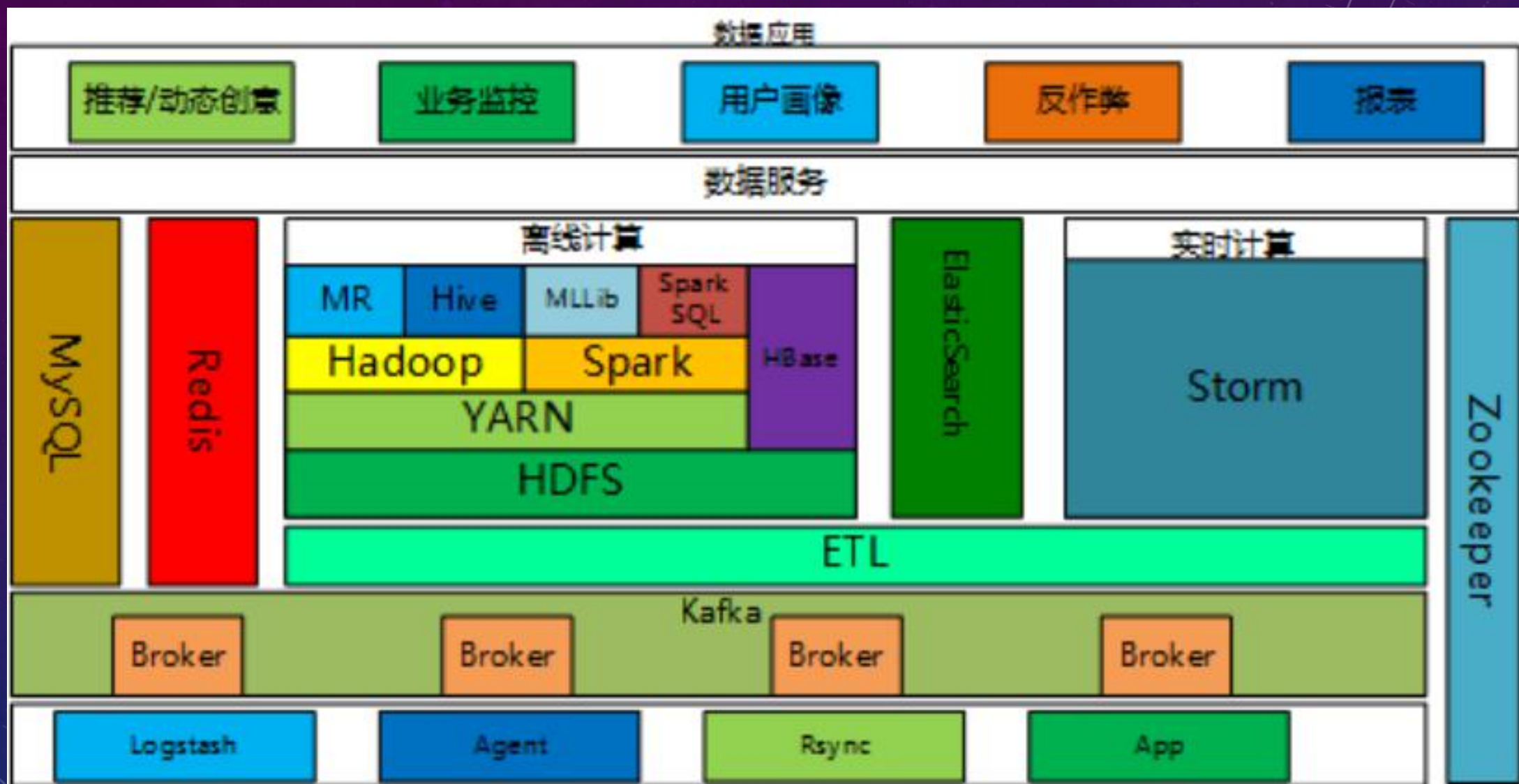
# PC广告与移动广告的区别

- 交互方式的不同
- 使用位置的不同
- 用户标识的不同
- 展现形式的不同
- 网络环境的不同
- 受众定向的不同

# 精准广告中数据的主要应用

- DMP
- 人群 / 用户画像
- 个性化创意 / 推荐引擎

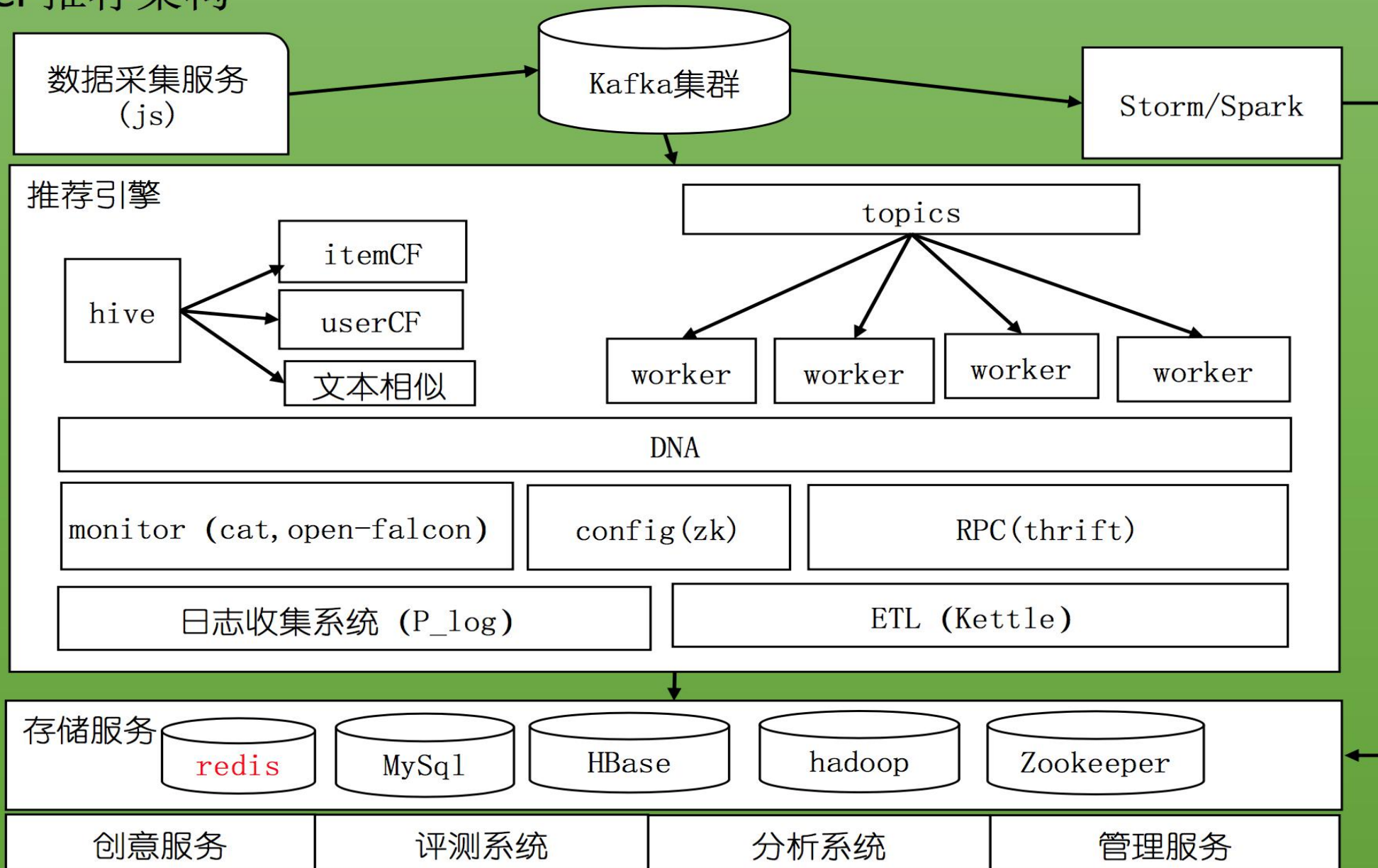
# DMP数据处理架构



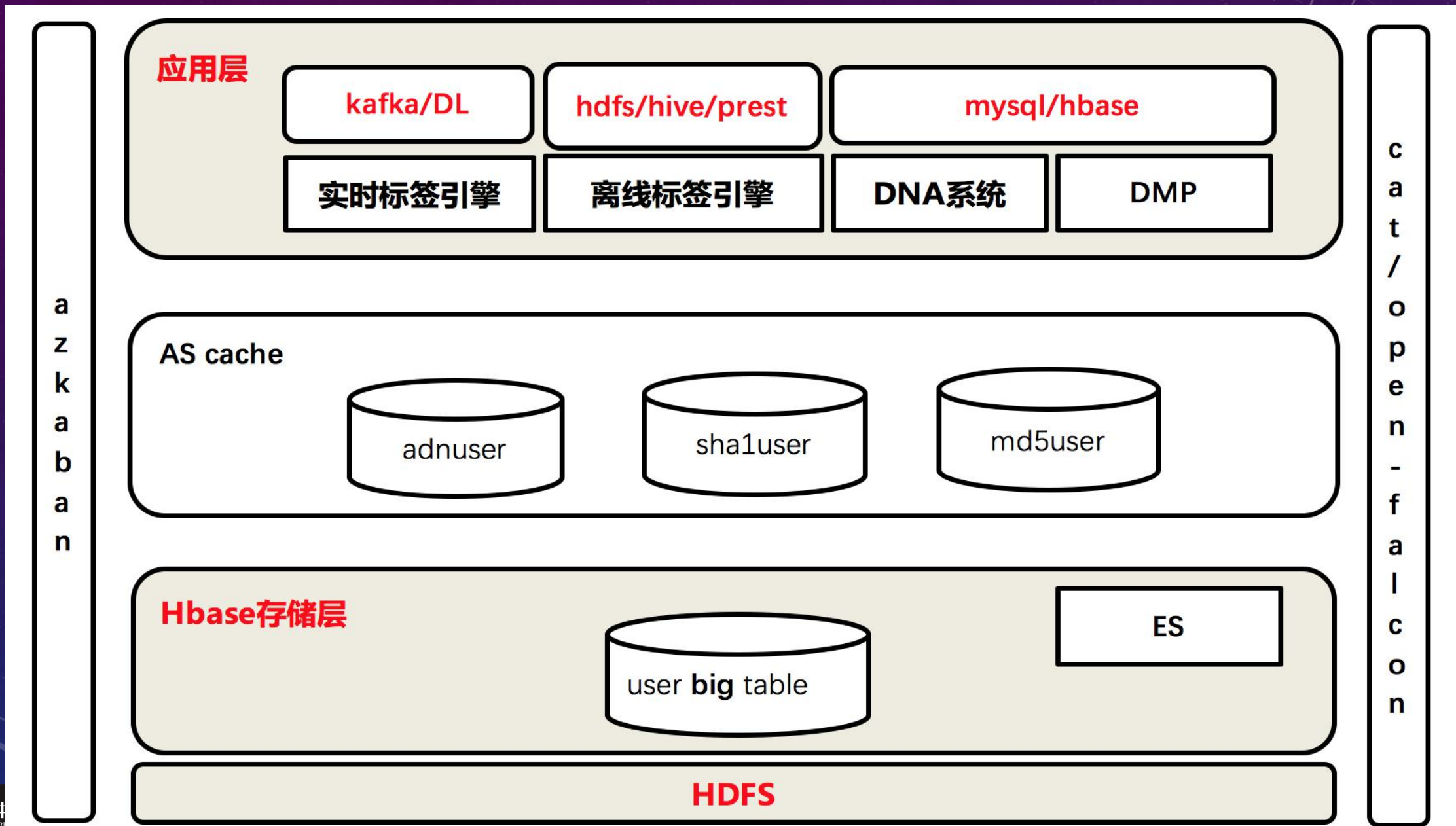


# 个性化创意/推荐引擎架构

## DCP推荐架构



# 人群 / 用户画像架构



# 大数据处理分析技术介绍

- 储存 / 缓存
- 计算
- 算法 / 规则



# 储存 / 缓存

- Tair / redis2 -> ssdb/redis3 -> aerospike
- Kafka / DL
- Mongodb / hbase
- Hadoop
- Mysql

# 计算

- Hive/mr
- Storm
- Spark Streaming

# 算法 / 规则

- 协同过滤（UserCF/ItemCF）
- 文本相似 / 图片相似
- 关联规则（支持度和置信度）
- 逻辑回归模型
- 兴趣时间衰减
- 规则引擎（DNA）



# 心得体会

- 做工程不易，且做且珍惜（技术+管理+工匠精神）
- 建立PDCA效果验证闭环
- 每一个做技术的都应该有崇高的理想，技术只是手段
- 什么是大数据？

# THANKS

谢 | 谢 | 观 | 看