

字节跳动A/B测试平台演进历史及实践

王珂

字节跳动 A/B 测试平台技术负责人

目录

1. Experiment @ ByteDance
2. 发展历程
3. 经验沉淀
4. 未来展望

一个真实的🍉 - 西瓜视频名字怎么来的？

① 实验背景

- 为了让“头条视频”APP能在短视频行业拥有更清晰的品牌辨识度，团队希望为其更改一个更好的名字。
- 经过用户调研+多轮征集筛选后，最终选定了4个待选名字进行AB测试。

③ 实验设计

实验版本	测试名称	实验指标效果
		点击率
对照组V0	头条视频	/
★ V1	西瓜视频	+5.22%
V2	奇妙视频	+5.08%
V3	筷子视频	+4.87%
V4	阳光视频	+4.36%

② 实验目的

- 验证哪一个应用名称能更好地提升“头条视频”APP在应用商店的点击率。

④ 实验结论

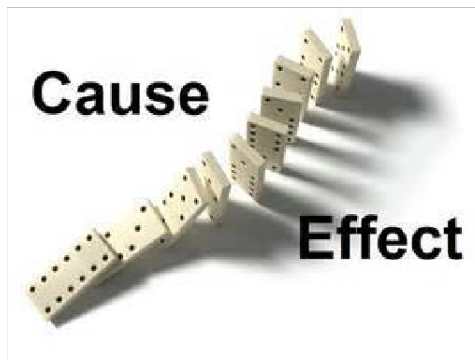
- 西瓜视频和奇妙视频的点击率位列前二，且差距不显著
- 结合用户调性等因素的综合考量后，最终决定：头条视频正式更名为**西瓜视频**。

A/B 测试能给我们带来什么？

风险控制



因果推断



复利效应

$$1.01^{365}=37.8$$

$$0.99^{365}=0.03$$

字节跳动A/B测试规模

1500+



日新增试验

1W+



同时运行试验

500+



服务业务线

40W+



近一年试验数

无处不在的数据驱动理念

- 无实验，不上线
- A/B不是万能的；没有A/B是万万不能的

Talk is cheap, show me your A/B test metrics.



能力覆盖

效率提升

目录

1. Experiment @ ByteDance
2. 发展历程
3. 经验沉淀
4. 未来展望

发展历程 - 工具

2012年

- 字节跳动成立，开始使用A/B测试协助决策制定
- “今日头条”这个产品的命名也来自A/B测试

2013年

- 第一个较为完整的A/B测试平台构建完成
- 主要支撑推荐算法迭代
- 分流服务、平台门户、数据建设三部分功能基本成型

2015年

- 协同客户端配置发布，完成第一版客户端实验能力建设

发展历程 - 产品

2016年

- A/B测试平台重构，正式定名Libra（天秤座），寓意客观公平
- 广告场景接入 A/B 测试平台
- 推送场景接入 A/B 测试平台

测试总数破 1W

2017年

- 互娱社区业务完整接入A/B测试平台
- 应对高涨的访问量，分流服务重构，资源占用下降67%，延迟缩减80%+

2018年

- 搜索业务接入 A/B 测试平台
- Interleaving 实验方案正式上线
- 基于曝光逻辑的客户端实验方案正式上线

测试总数破 10W

发展历程 - 解决方案

2019年

- 直播业务接入 A/B 测试平台
- 指标平台孵化成功，定向解决大盘指标和A/B指标二义性问题
- 服务端配置中心孵化完成

测试总数破 50W

2020年

- 电商业务接入 A/B 测试平台
- 自动调参平台孵化完成
- 客户端配置中心孵化完成

2021年

- 运营类场景高效支持
- 人工合成对照组能力支持
- MAB 能力上线

目录

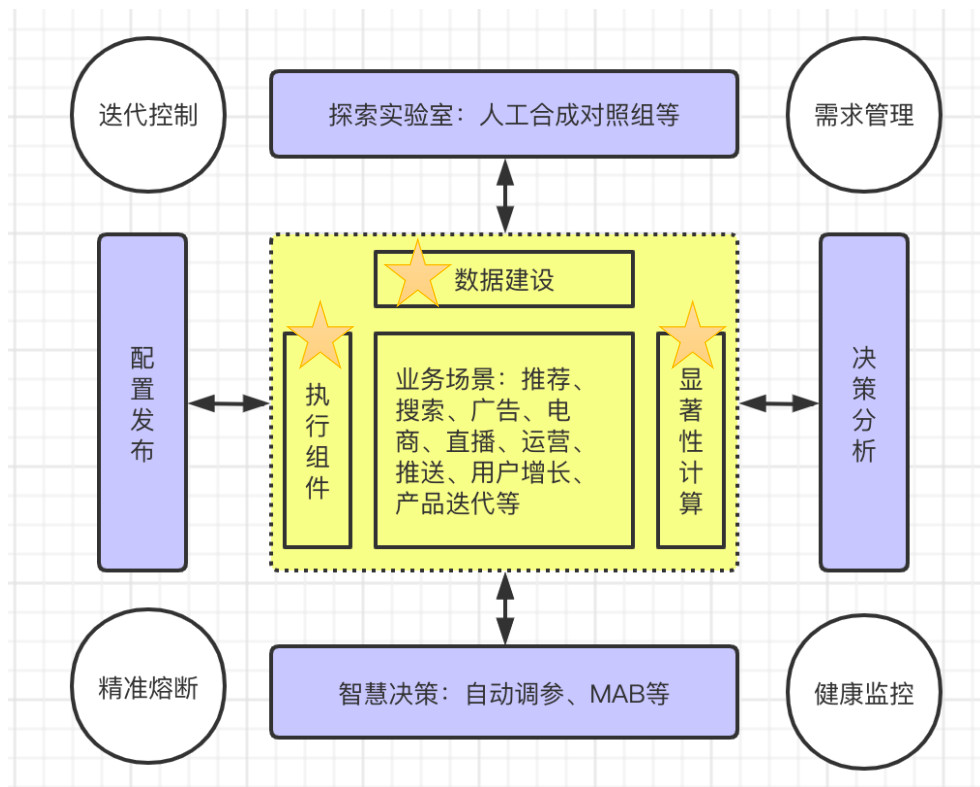
1. Experiment @ ByteDance
2. 发展历程
3. 经验沉淀
4. 未来展望



经验沉淀

- 能力图谱
- 统计分析篇
 - 假设检验及其修正
 - Pre-Experiment Data Utilization
- 测试执行篇
 - 流量的复用与隔离
 - 分流能力建设
- 数据建设篇
 - 离线处理 vs. 即席查询

经验沉淀 - 能力图谱



经验沉淀 - 统计分析篇

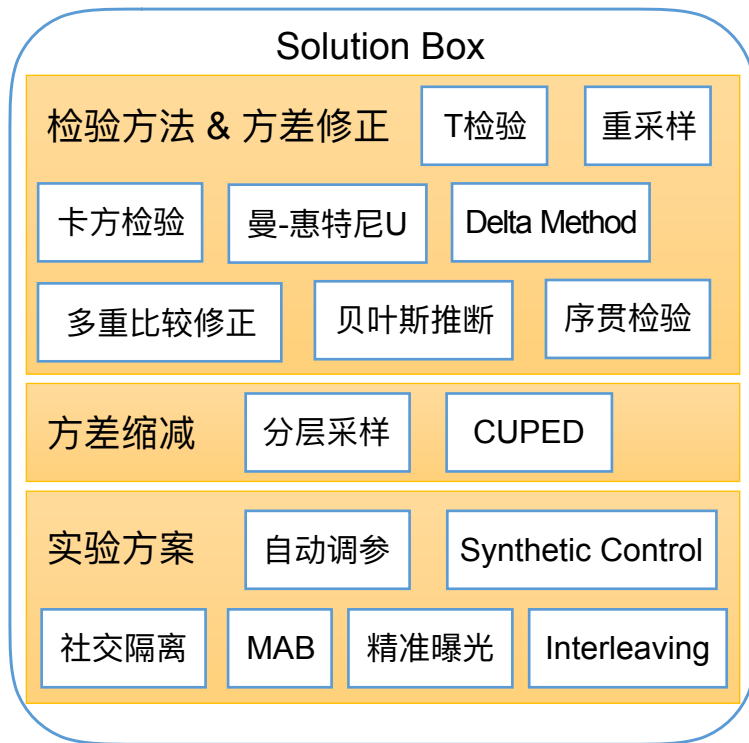
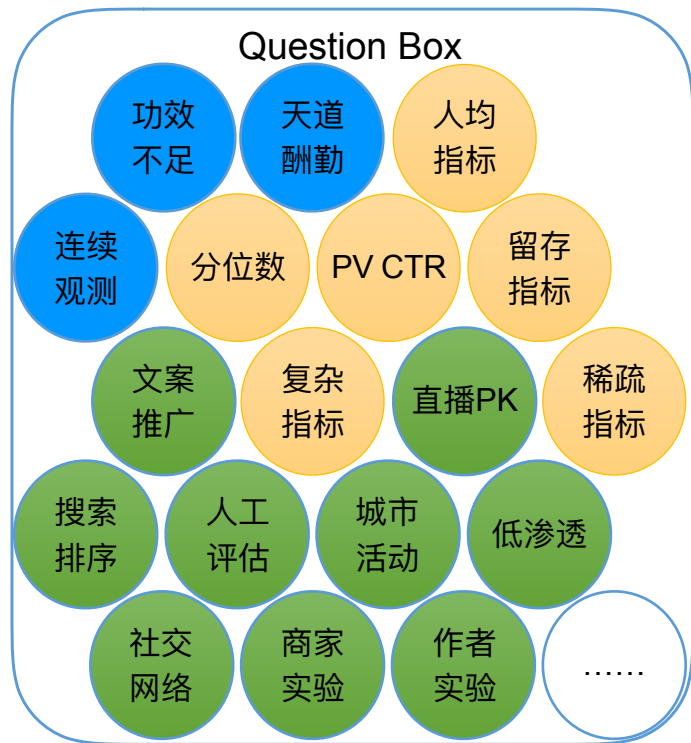


指标微涨，是波动还是策略收益？

答：假设检验(hypothesis testing)，是用来判断样本与样本、样本与总体的差异是由抽样误差引起还是本质差别造成的统计推断方法。

一类错误 vs 二类错误 vs $\alpha = 5\%$

经验沉淀 - 统计分析篇



经验沉淀 - 统计分析篇

- 设置显著性水平为5%，表示：对任意指标而言，开启100次AA测试（对照组和策略组行为一致的A/B测试），期望有 5 次AA测试该指标出现显著
- 当线上A/B测试数量很多，产品迭代进入精细化阶段时，**假阳性**问题带来的用户体感将变得尤为严重

小X：这个A/B测试的策略组用户是不是在这段时间里天生就比对照组的用户更活跃呢？

小Y：再进一步，我们是不是能够根据一个A/B测试里命中用户的实验前行为来做些什么呢？毕竟用户前后行为统计上看总是有些相关性的吧？

经验沉淀 - 统计分析篇

Pre-Experiment Data Utilization

SeedFinder - 以及变体后的 PreAA 校验

双重差分法 (Diff in Diff)

CUPED (Controlled-experiment Using Pre-Experiment Data)

经验沉淀 - 测试执行篇

- 流量的复用与隔离
 - Basic: 两次哈希保证层内流量互斥, 层间流量正交
 - 互斥域: 第三次哈希, 强制层隔离
 - 父子测试: 强制A/B测试的血缘关系
 - 共享对照组: 样本实体为用户请求时, 退化至一次哈希 - 勤俭节约
 - 流量轮转: 定期强制更新层上哈希种子 - 用户有害的A/B测试

经验沉淀 - 测试执行篇

- 分流能力建设
 - 从未停止的**高速迭代**
 - **多样化**的接入场景、功能拼图和服务端技术栈
 - 苛刻的**性能要求**

RPC

简单、灵活、可控
最主要的A/B 测试接入方式，高度性能优化

伴生进程

高效、稳定、可控
C++核心封装
Unix Domain Socket

SDK

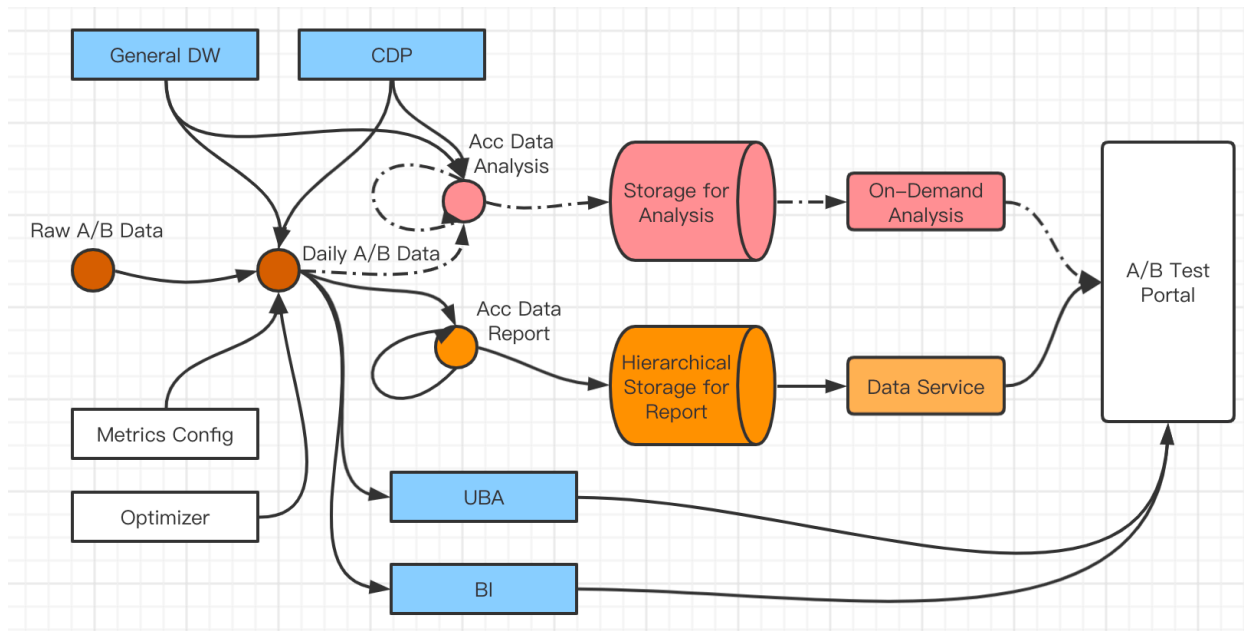
性能极限
有限的功能覆盖
部分语言提供

UDF

离线、流式处理

经验沉淀 - 数据建设篇

- 运算
 - 离线计算 - Report
 - 即席查询 - Analysis
 - 流式处理 - Monitor
- Rolling 累计指标运算
- 数据管道建设





经验沉淀 - 数据建设篇

- 多维数据处理
 - 报告向
 - 明细数据 - HIVE - 非核心维度交叉即席查询
 - 预聚合数据 - ByteSQL - 核心维度交叉即席查询
 - 指标数据 - ABase(KV) - 核心维度非交叉预结算查询
 - 分析向
 - 面向单个实验的明细数据抽取
 - Rolling to Accumulate Data

目录

1. Experiment @ ByteDance
2. 发展历程
3. 经验沉淀
4. 未来展望

未来展望

- 进一步提升A/B测试的效率
 - 配置中心 - 打穿迭代过程中工程师的开发测试和发布
 - 提升效率，降低风险，多职能协同
 - 智能分析 - 降低分析师日常分析成本
 - 维度拆解，异常数据监测，测试结论自动化
 - 场景化支持
 - 合适的才是最好的。
 - 对于典型场景，由通用产品走向解决方案



未来展望

- 能力上覆盖更广泛的业务场景
 - 跟随业务发展的脚步
 - 方法论探索
 - 社交网络场景下更好的解决方案



未来展望

- 能力闭环与平台开放
 - A/B测试设计
 - 迭代经验积累
 - 业务中台互联互通

未来展望

- ToB 能力进一步提升



The screenshot shows the top navigation bar of the 火山引擎 (Huoshan Engine) website. The navigation bar includes the logo, links for 最新活动 (Latest Activities), 产品 (Products), 解决方案 (Solutions), 文档中心 (Documentation Center), 合作与生态 (Partnerships and Ecosystem), 支持与服务 (Support and Services), and 关于我们 (About Us). There is also a search bar with the placeholder text "请输入关键词" (Please enter keywords) and buttons for 管理控制台 (Management Console), 登录 (Login), and 立即注册 (Register Now).

The main content area features a large section for A/B testing. The title "A/B测试" (A/B Testing) is prominently displayed. Below the title, the text reads: "原数据优化师，摆脱猜测，用科学的实验衡量决策收益，打造更好的产品，让业务的每一步都通往增长" (Original data optimizer, get rid of guessing, use scientific experiments to measure decision-making benefits, create better products, let every step of the business lead to growth). There are two buttons: "体验Demo" (Experience Demo) and "说明文档" (Documentation).

On the right side of the main content area, there is a large, glowing blue 3D graphic of a cube with an upward-pointing arrow inside it.

At the bottom of the page, there are two promotional banners. The left banner is titled "面向小微企业限时免费" (Limited time free for small and medium enterprises) and includes the text "MAU≤5万、事件量≤2亿情况下，最高可享12个月免费" (Under the conditions of MAU ≤ 50,000 and event volume ≤ 200 million, up to 12 months of free service). The right banner is titled "A/B测试大讲堂" (A/B Testing Lecture) and includes the text "A/B测试何以备受追捧，企业又该如何进行实践" (Why is A/B testing so popular, and how should enterprises practice it?).



THANKS



ByteDance 字节跳动