

# AB测试和灰度发布平台 架构设计和实践

1688技术部 欧舟（翰承）

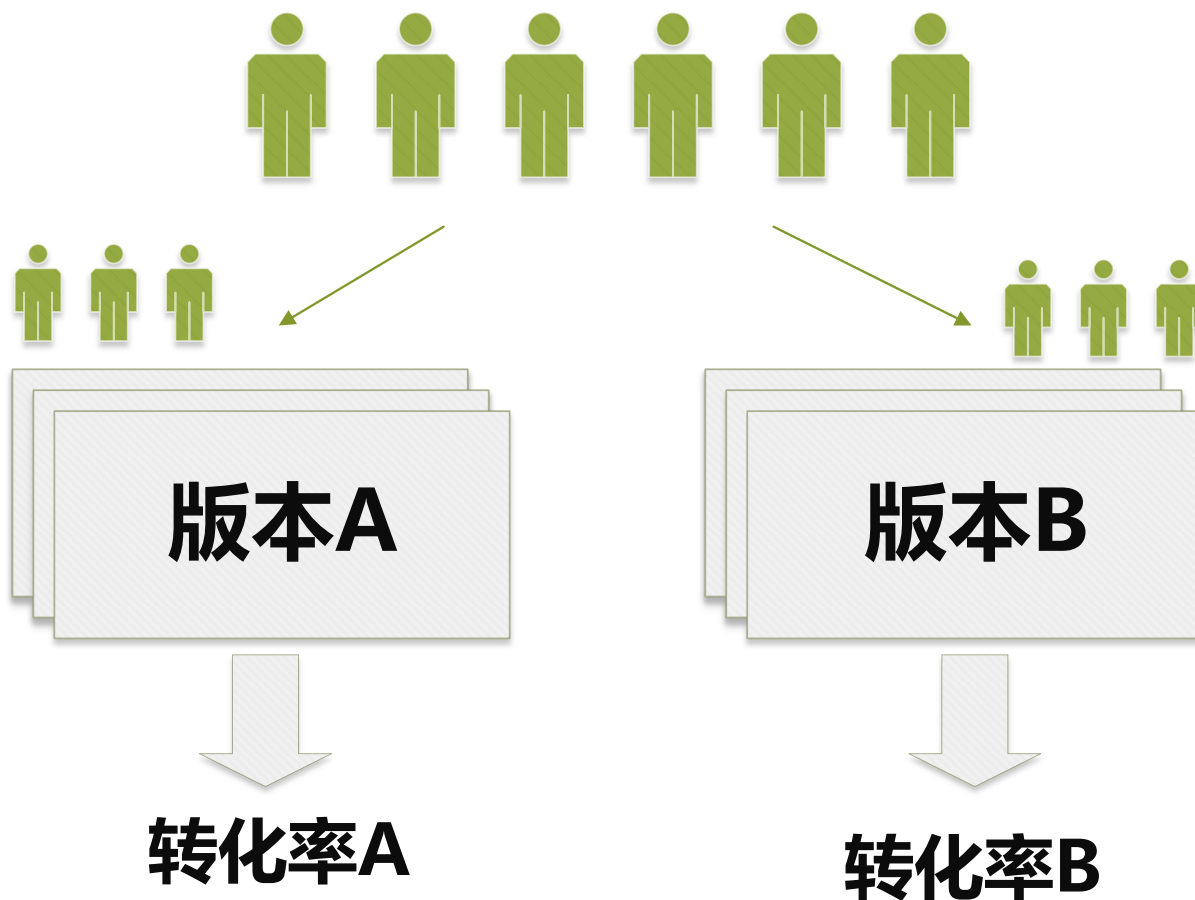
2013.7.12

# 大纲

- 业务价值
- 系统设计
- 架构优化
- Q&A

# AB测试介绍

AB测试 (A/B Test) , 也称Bucket Test



# 业务价值

- 缩短测试周期
- 减少测试干扰，提高测试准确性和价值
- 减少用户伤害，降低测试影响

页面优化及改版方案评估决策

后台算法，引擎及业务逻辑优化

# 还有哪些公司在使用A/B测试？

- Google

- 百度

- 微软

- 易趣

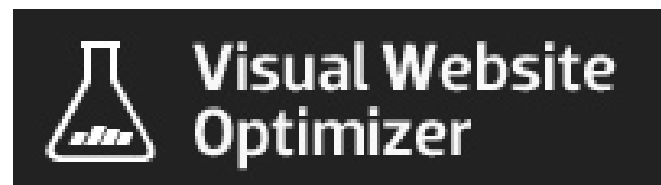
- Amazon

- BBC

- 迪斯尼

- ○ ○ ○

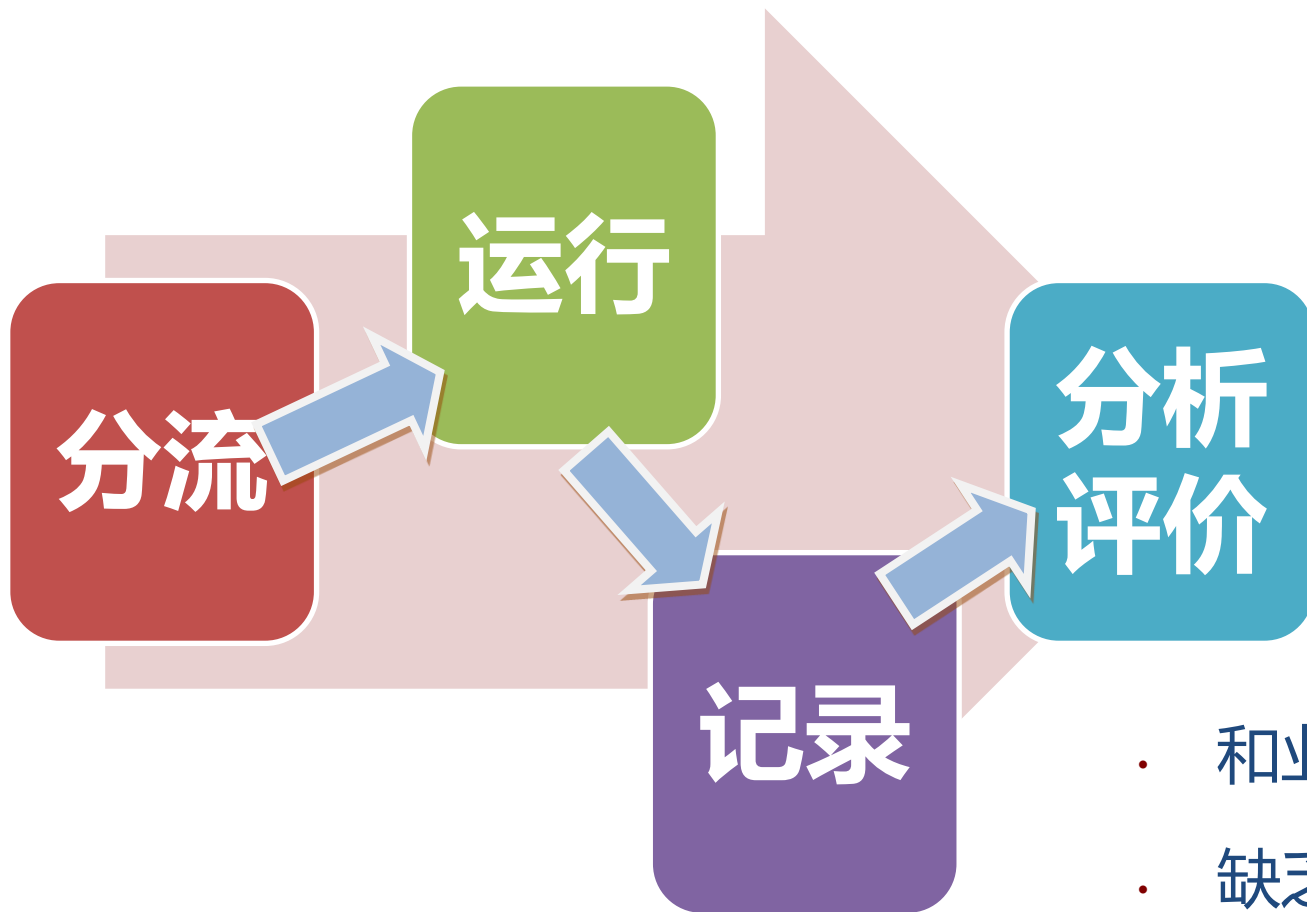
Website Optimizer



# 大纲

- 业务价值
- **系统设计**
- 架构优化
- Q&A

# 核心环节



- 和业务系统高度耦合
- 缺乏通用性
- 过程繁琐

# BucketTest系统介绍

- 阿里巴巴集团最主要的AB测试和灰度发布平台
- 中文站、国际站、阿里金融、天猫，淘宝，支付宝（开源）、一淘（开源）等多个事业部
- 30多条产品线的200多个项目
- 多项核心技术



# BucketTest系统案例

## 中文站举例

- Exodus系统会员登录
- 中文站首页改版
- 人气搜索
- 321大促
- 云归项目
- 919大促
- SearchWeb2.0项目
- 服装和社区首页改版
- 1122大促活动
- 物流项目
- 服装行业年度盘点活动
- 搜索排序测量模型测试
- offerdetail页面测试

## 国际站举例

- 排序算法mlr升级
- magellan样式切换及改版
- p4p广告算法升级及样式改版
- 搜索样式、按地域实验
- PPC右侧样式实验
- ae\_style实验
- 勃兰项目
- winner项目
- Supplier频道
- compare分流实验
- minOrder分流实验
- 首页推荐算法

## 阿里金融举例

- 直投贷款
- 老客户利率
- 订单贷款提前收款
- 账务明细白名单名单
- 天猫订单未续签不准入
- 天猫个人担保名单
- 集市信贷等额本金
- 集市信贷合并还款日
- 集市短期贷款测试名单
- 订单息费模式白名单

## 一淘举例

- P4P精美广告投放
- P4P114首页改版升级

# 主要模块

## 工具端

技术角度

ibatis

Spring Context

Roma

Sqlmap

Model

Service

Controller

Velocity

业务角度

定制应用  
(应用描述文件)

配置测试  
(创建测试, 修改测试, 删除测试)

即时更新

数据缓存  
(实时性, 周期性更新)

数据服务  
(提供测试数据)

报表服务  
(表格和图形)

后台管理  
(应用管理  
用户管理  
权限管理)

应用集成方

每隔5秒同步

## 应用集成方

集成方业务模块

分流服务

注册与监听

客户端 API  
(解析配置信息)

分流计算  
Selector

配置监听服务  
(解析配置信息)

配置缓存区  
(缓存配置信息)

用户信息缓存区  
(ThreadLocal)

解析

解析模块  
(解析配置信息)

Bucket TestFilter  
(提取用户信息)

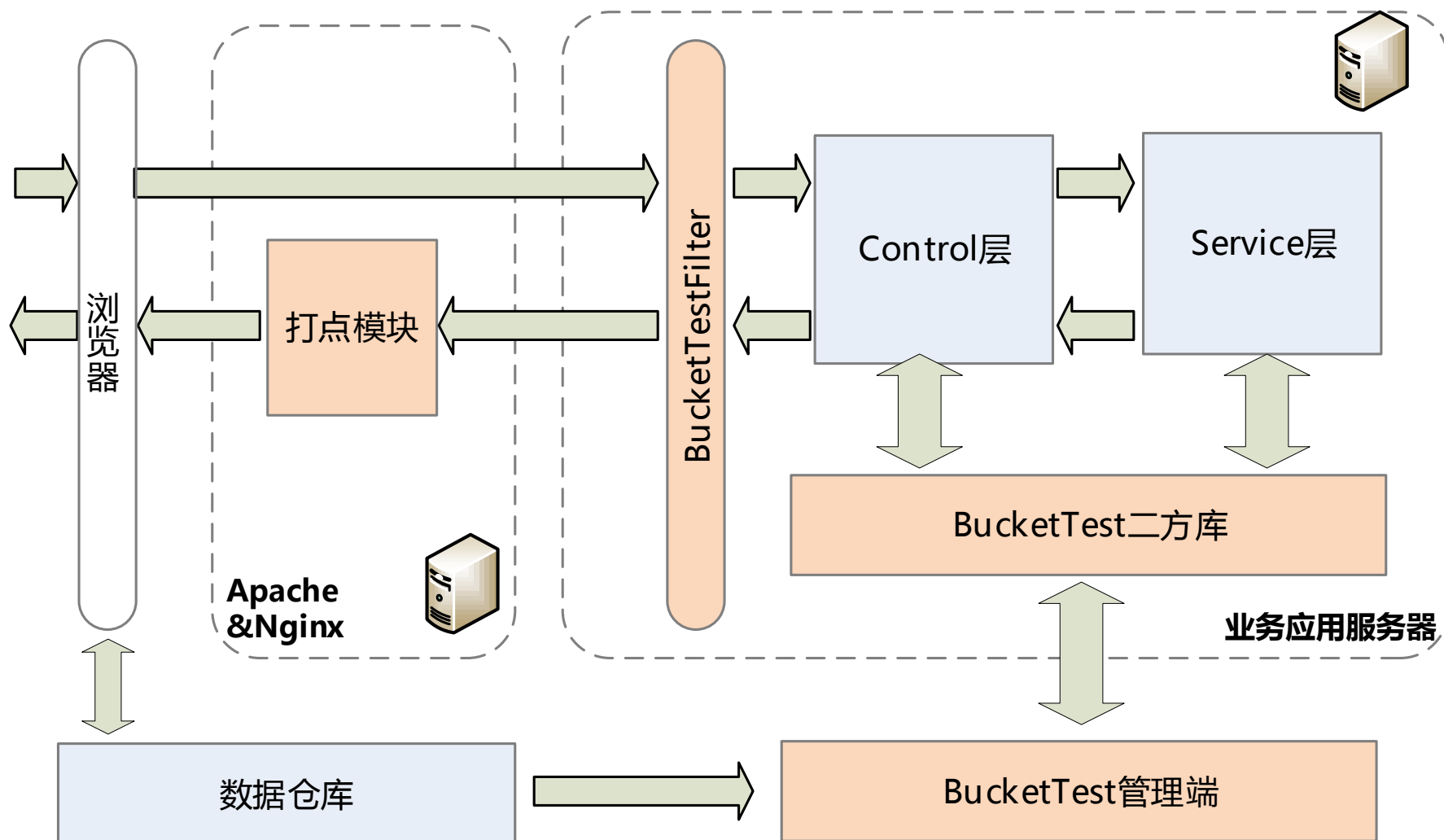
同步配置  
(同步配置信息)

自动打点  
(向DB打点)

通知

BucketTest二方库

# 系统设计



# BucketTest二方库

- **BucketTest二方库同步模块**：向配置端同步用户配置信息
- **BucketTestFilter**：提取用户信息并缓存，以及在返回结果页面时封装打点信息
- **BucketTestClientAPI**：提供实际分流计算和服务
- **Apache打点模块**：生成javascript脚本，支撑打点服务

- 无兼容性风险
- 极低性能风险（内存运算）
- 分布式设计，易于风险控制
- 自省能力（分流自监控）

# 分流与记录（打点）过程



# 一站式报表

全站视角指标

P4P指标

点击指标

页面指标

Offer指标

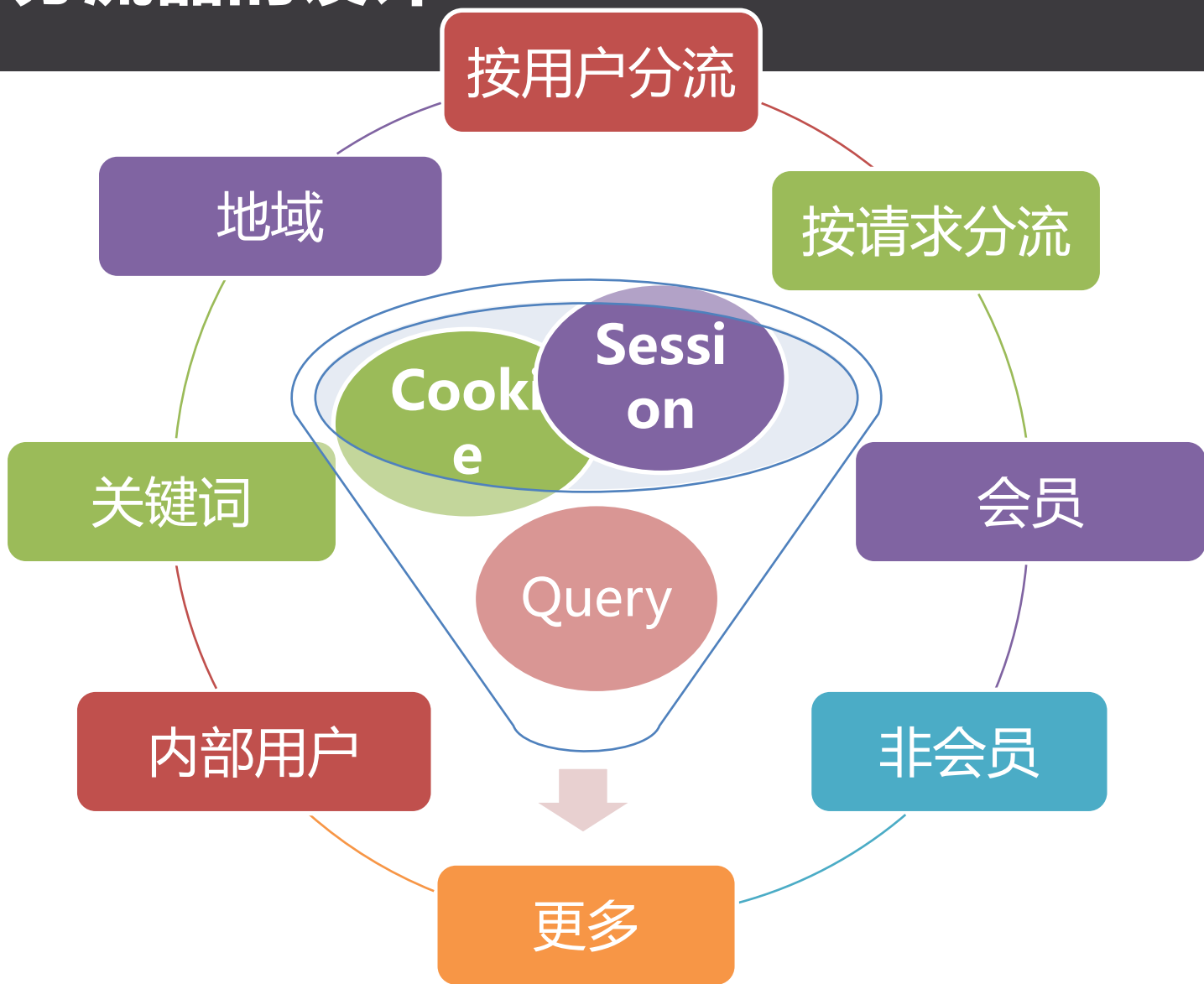
会员指标

关键词指标

产品线：  来源：  Cookie类型：  省份：  城市：

版本	用户类型	区域点击数	区域UV转化率	平均访问深度	平均停留
新版本1	新版本	576519	3.2	4.19	79.78
老版本	等量对比	497462	2.76	3.77	79.92
老版本	全量	567261	2.76	3.77	79.89
新版本1	新版本	827513	3.29	4.54	73.34
老版本	等量对比	709508	2.82	4.04	73.63
老版本	全量	830702	2.82	4.04	73.76
新版本1	新版本	179424	2.76	4.42	61.73
老版本	等量对比	155169	2.38	3.94	62.91
老版本	全量	181905	2.38	3.94	62.88

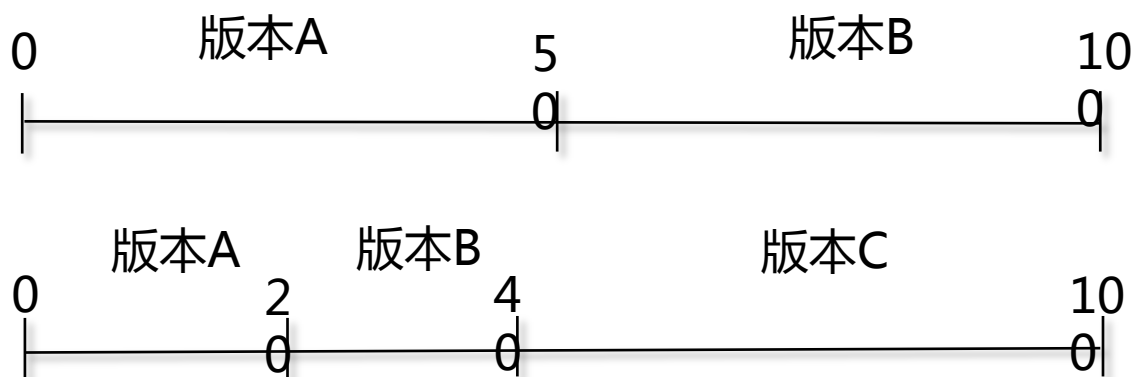
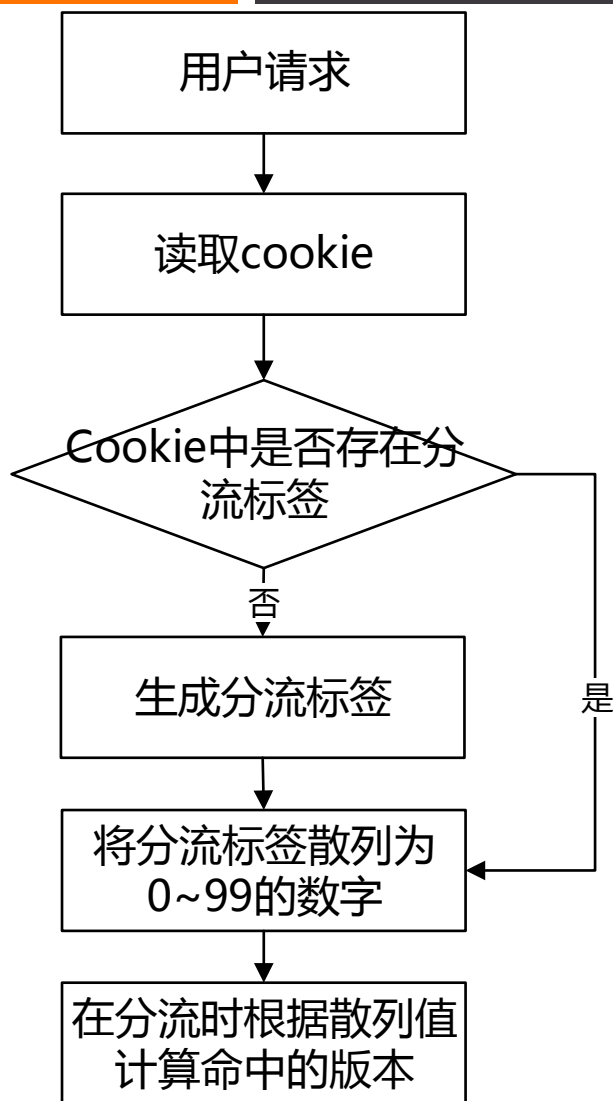
# 分流器的设计



# 分流（按用户百分比分流）

**【分流标签生成策略】**  
基于用户机器的IP地址  
+ 用户第一次访问的时间  
+ 一个随机数

如121.0.29.199.1335163135361.3





# 大纲

- 业务价值
- 系统设计
- **架构优化**
- Q&A

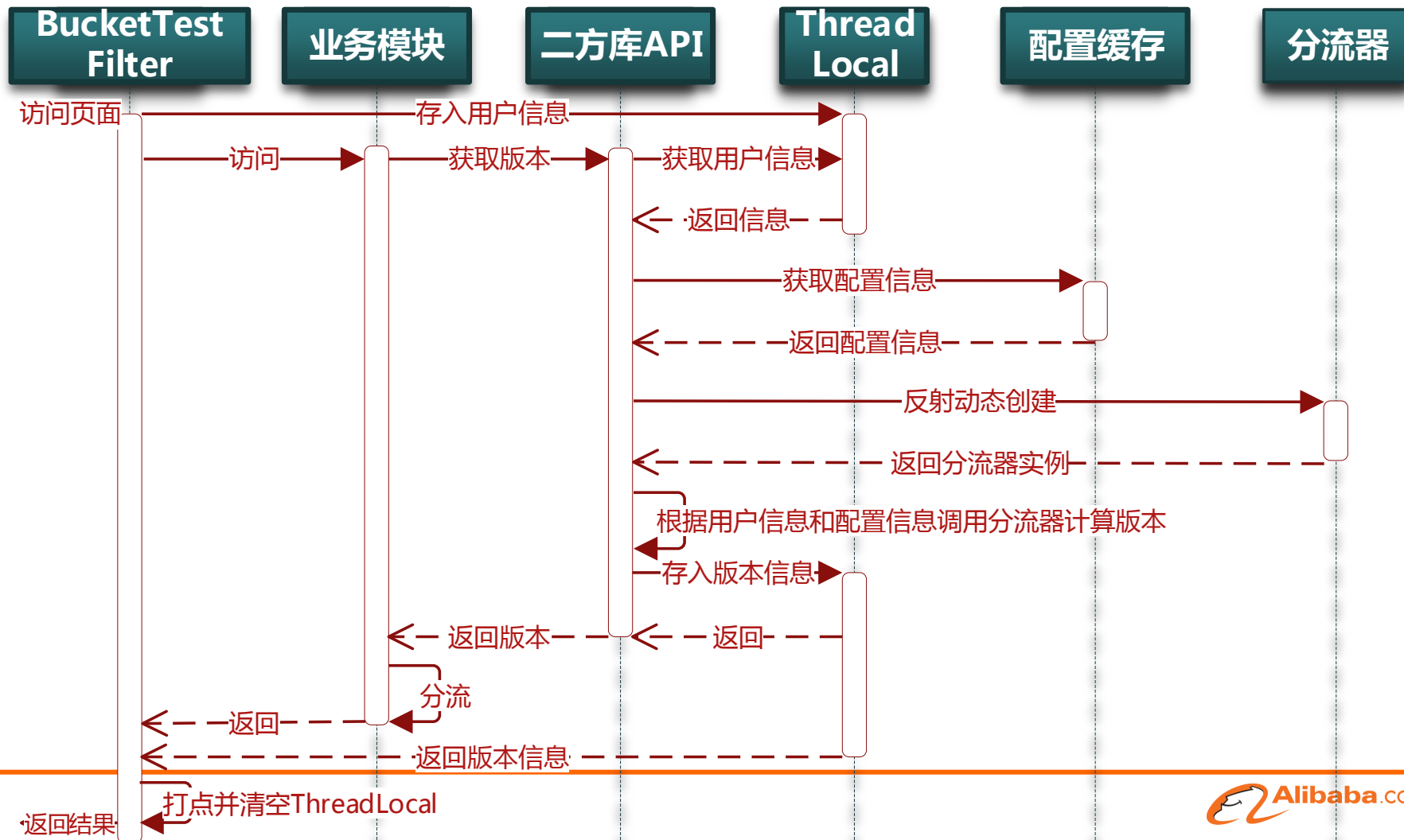
# 自动透明化打点

- 问题背景
- 解决方案

向http头写入打点信息

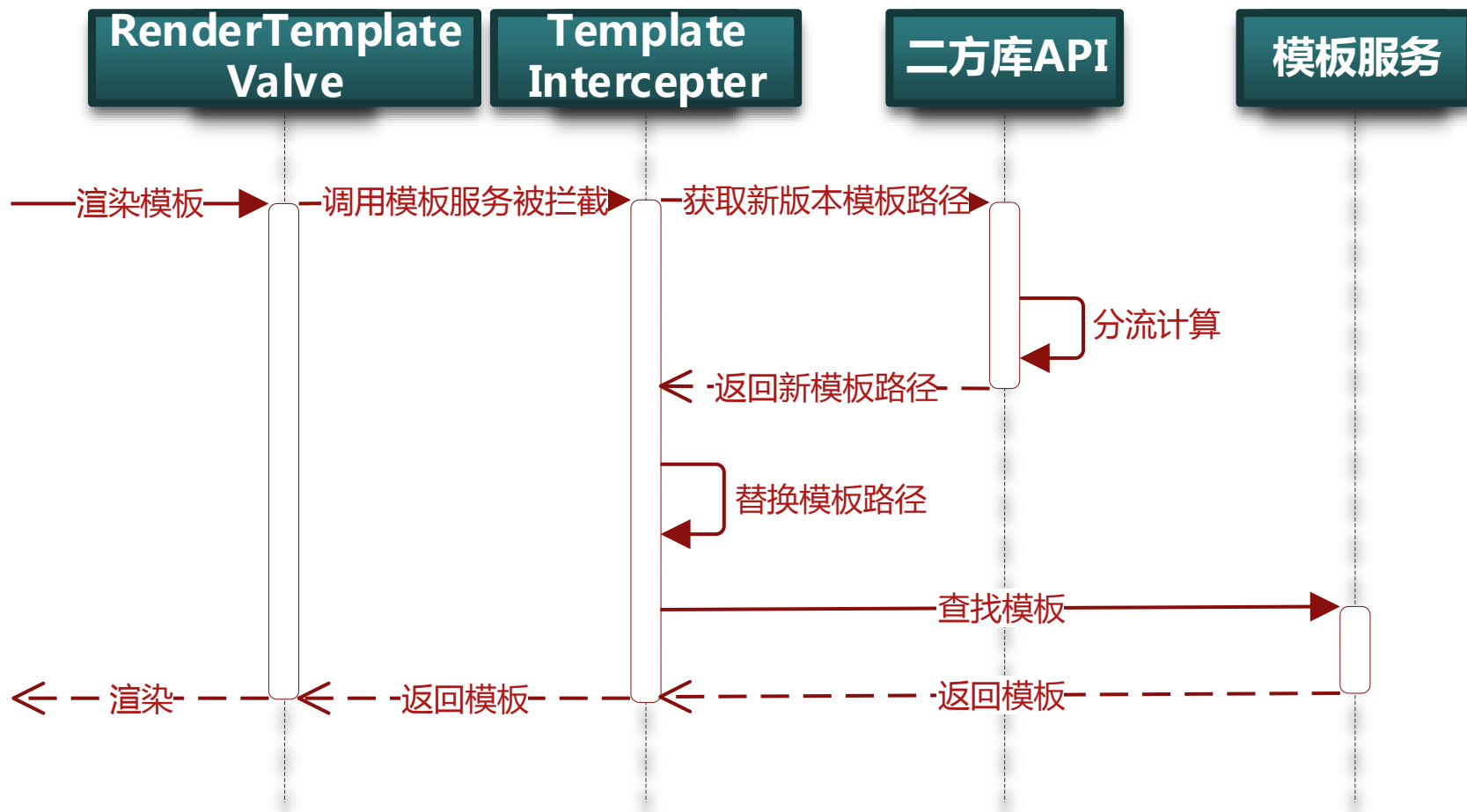
ApacheModule  
打点信息转为JS

页面加载时打点



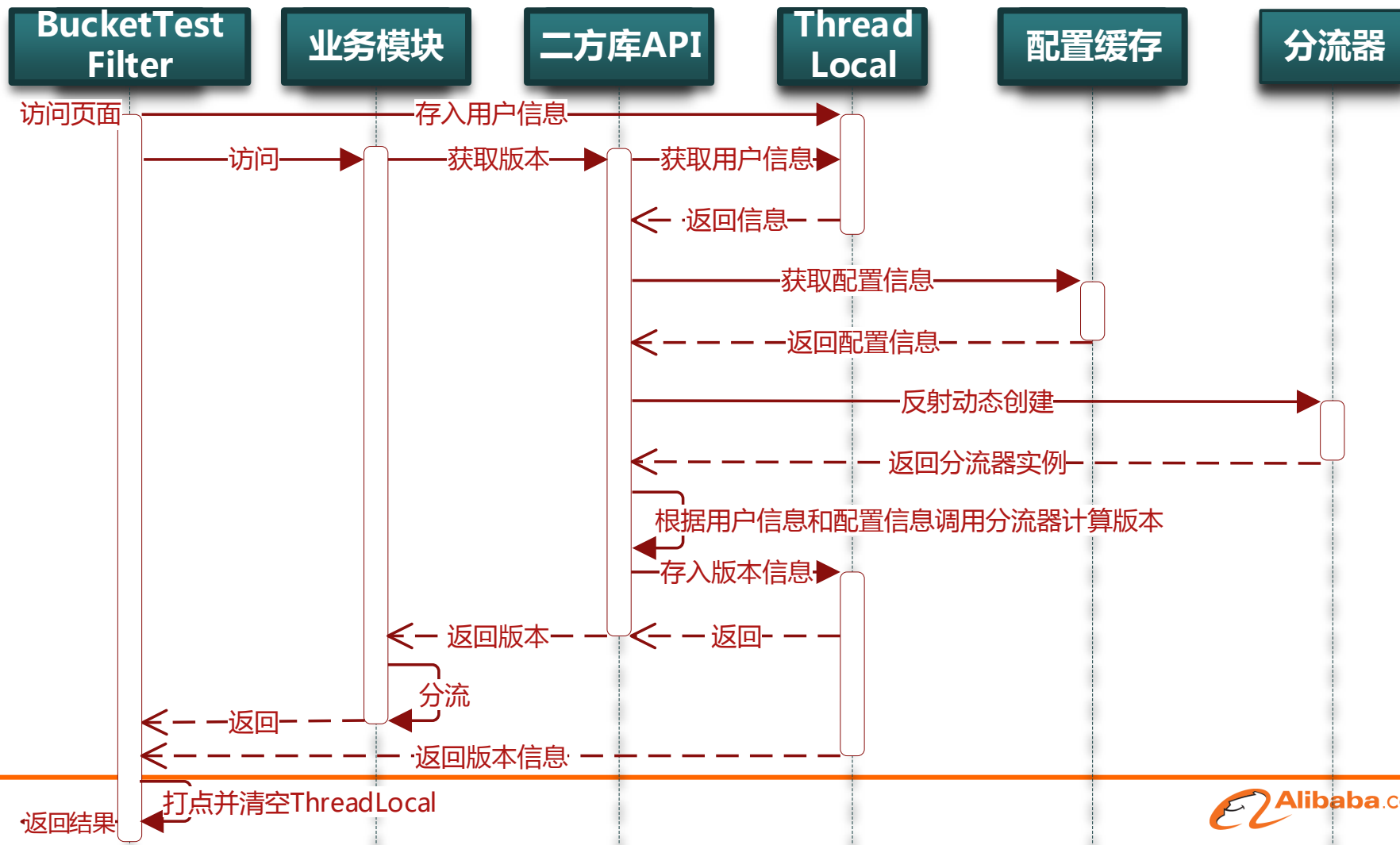
# 动态页面测试无侵入方案

- 问题背景
- 解决方案

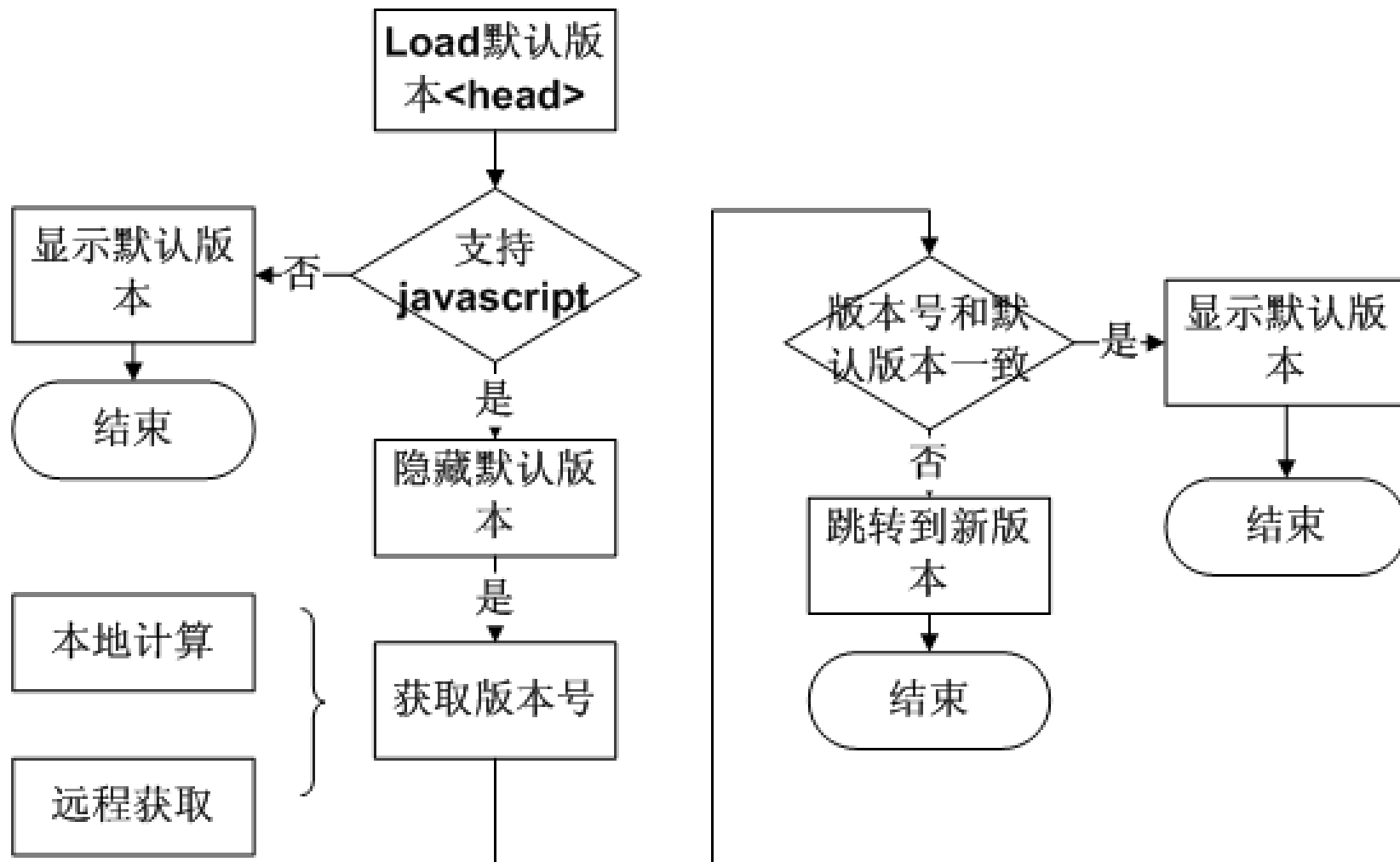


# 优化用户信息获取

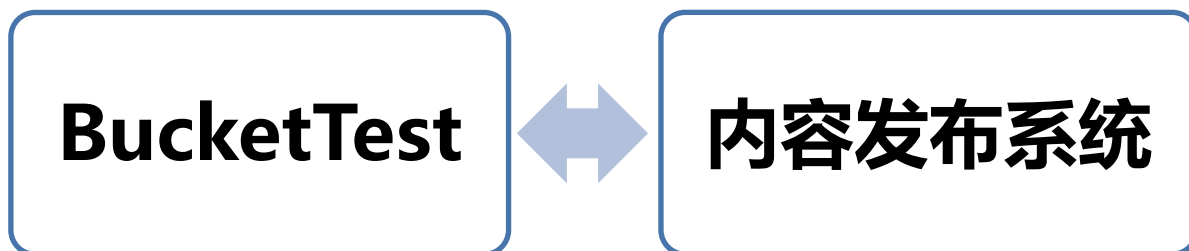
- 问题背景
- 解决方案



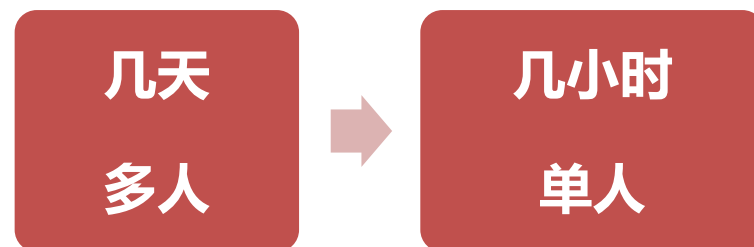
# 静态分流技术



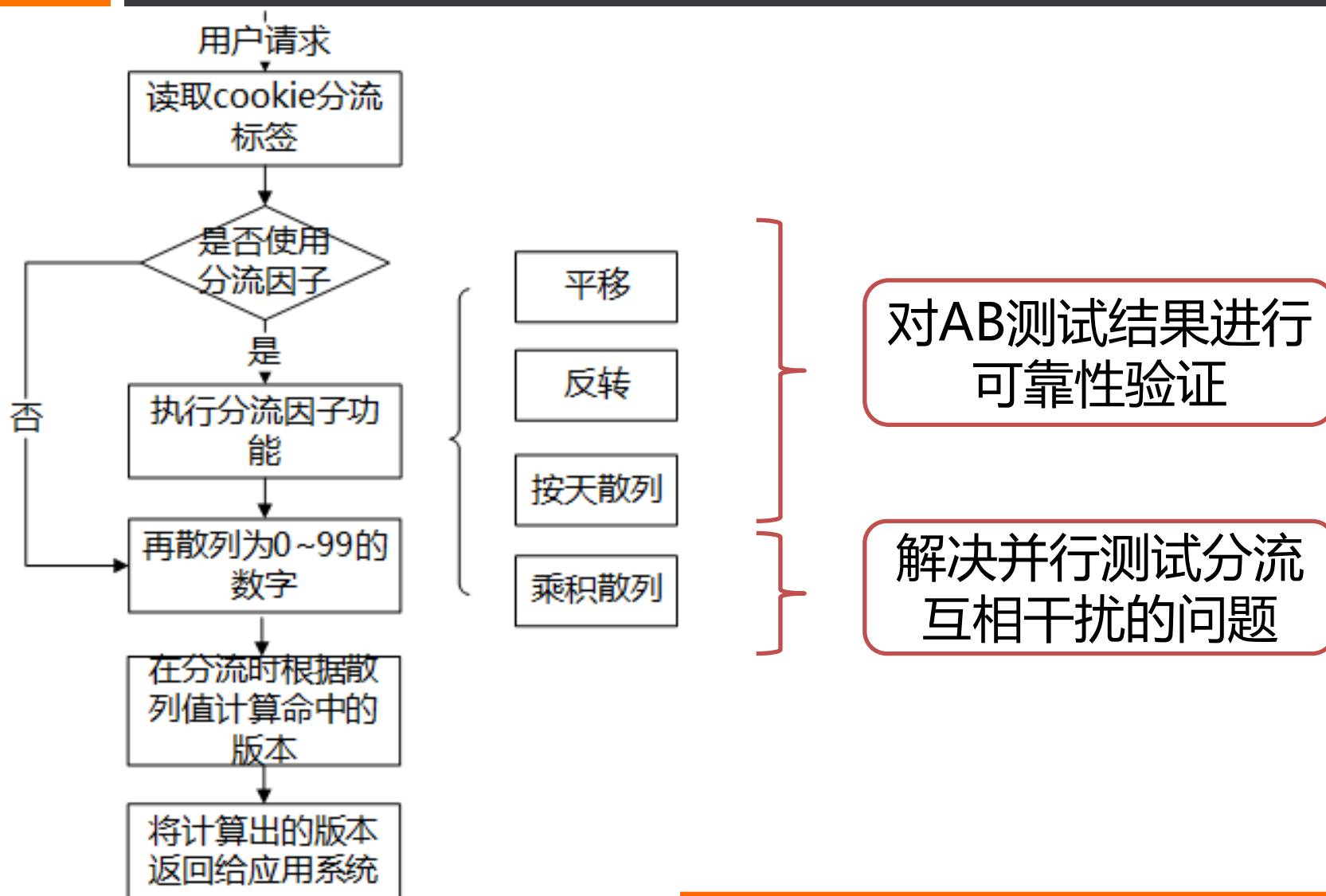
# 静态分流技术



- 不需要申请开发资源
- 测试准备周期大大缩短
- 脚本模板复用（性能，异常处理，兼容性，细节处理，反复验证）



# 分流因子功能



# 解决并行测试干扰问题





# 分流扩展性

**Selector接口->getVersion方法  
( DefaultSelector )**

- 性别分流
- 先关键词分流，再Cookie分流
- 对五年以上的会员分流一个版本
- 可以方便实现自己的分流意图
- 分流只有想不到，没有做不到

# 表单界面配置自动生成

配置该产品支持什么样的分流器，如Cookie分流，地域分流等等

- `<selector name="cookie" displayName="Cookie选择方式"`
- `impl="com.alibaba.asc.buckettestclient.selector.CookieSelector">`
- `<param name="percent" displayName="百分比 "`
- `type="int" displayType="input" use="required" />`
- `</selector>`

配置该产品支持什么样的参数，即版本A对应哪几个参数及参数类型

- `<group name="sale_offergb" displayName="供应线的isearch"`  
`bidPrefix="sale_isearch" maxBucket="5">`
- `<param name="search" displayName="展示页面模板" type="string"`  
`displayType="input " use="required" />`
- `</group>`

# 表单界面配置自动生成

时间和打点设置：

选择器信息：

+ 选择

Group

测试

Group信息：

测试组： company\_companyModel

版本1

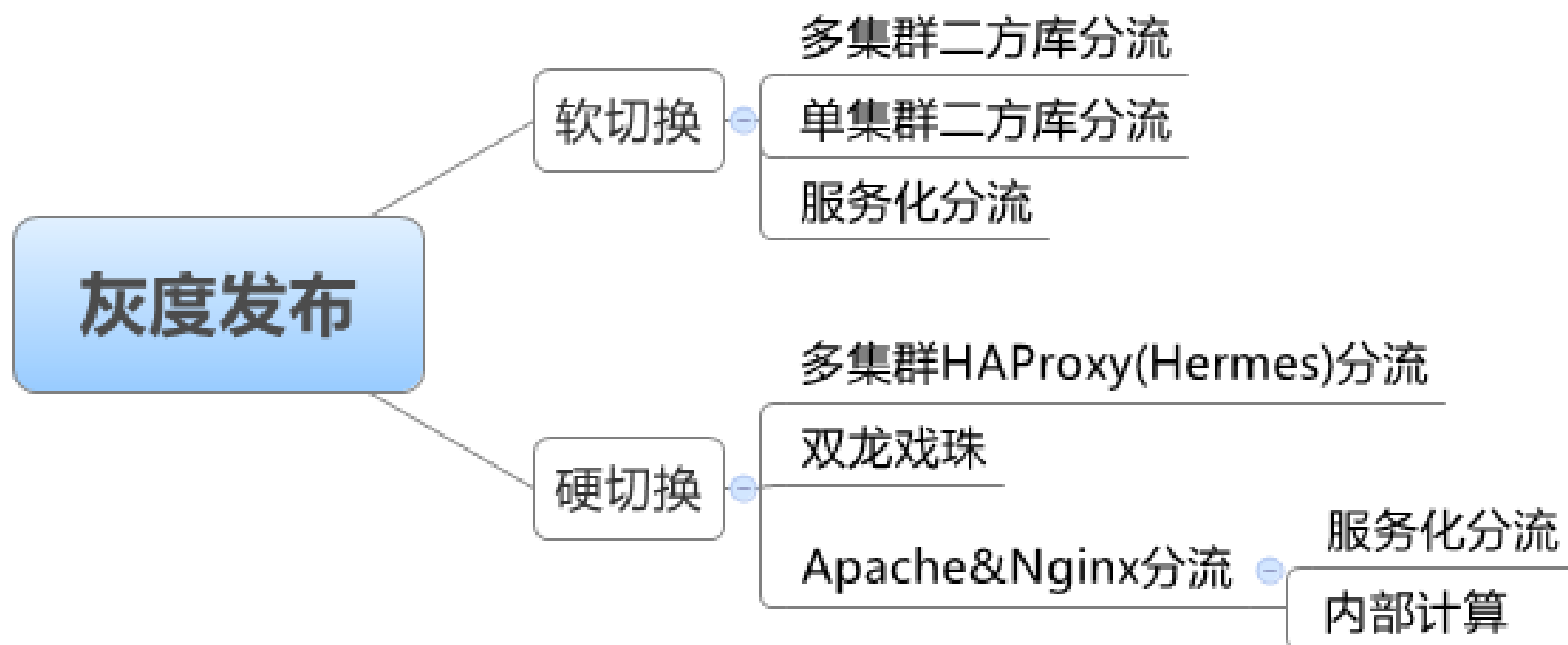
后台服务地址( address )：

复杂参数操作( query )：

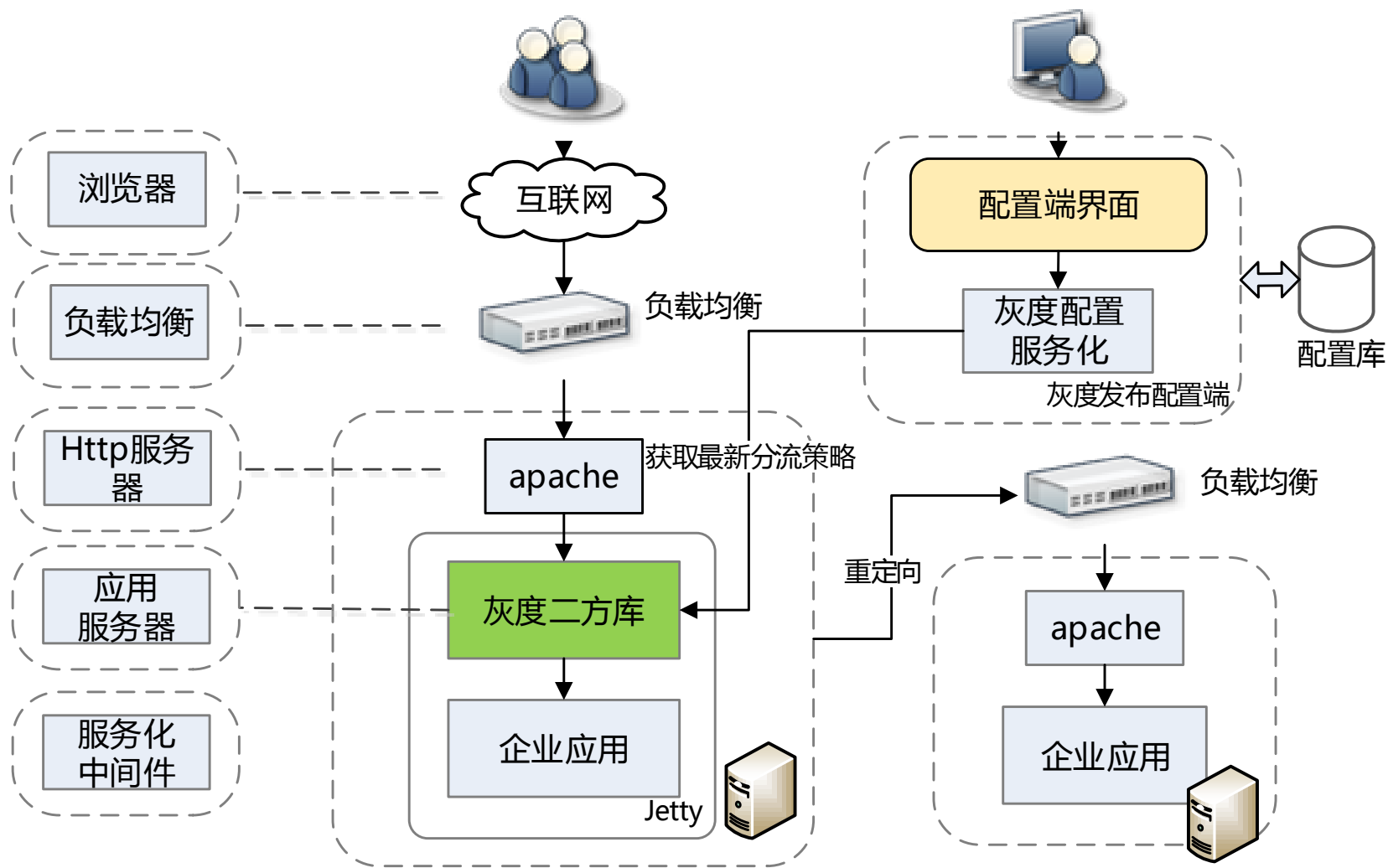
SmartQuery标识( sq )：

百分比

# 灰度发布解决方案



# 灰度发布解决方案



# 假设检验在AB测试中的应用

- 什么是假设检验？
- 基本原理（小概率原则）

# AB测试结果验证-假设检验

A版本			
6	9	10	5
7	7	3	4
4	9	3	6
9	5	9	4
4	7	8	4
8	7	8	8
4	5	9	8
9	6	8	3
6	8	10	3
7	6	5	4
8	8	8	5
5	2	9	8
5	4		

B版本			
10	8	10	6
9	8	10	10
3	9	4	8
5	7	5	7
3	7	3	9
6	9	7	9
5	5	7	9
3	9	4	8
2	10	5	1
7	6	6	7
6	3	8	8
8	8	6	8
5	10		

版本	A版本	B版本
平均值	6.34	6.72
标准差	2.162859	2.373665
方差	4.677959	5.634286

- 一个是随机因素
- 一个是系统性因素

# AB测试检验方法-Z测试

- 提出假设（原假设H0，没有差异，那么备择假设H1就是有差异）
- 规定显著性水平，如0.05。
- 计算Z测试值
- 比较计算所得Z值与理论Z值，推断发生的概率，采用双尾比较

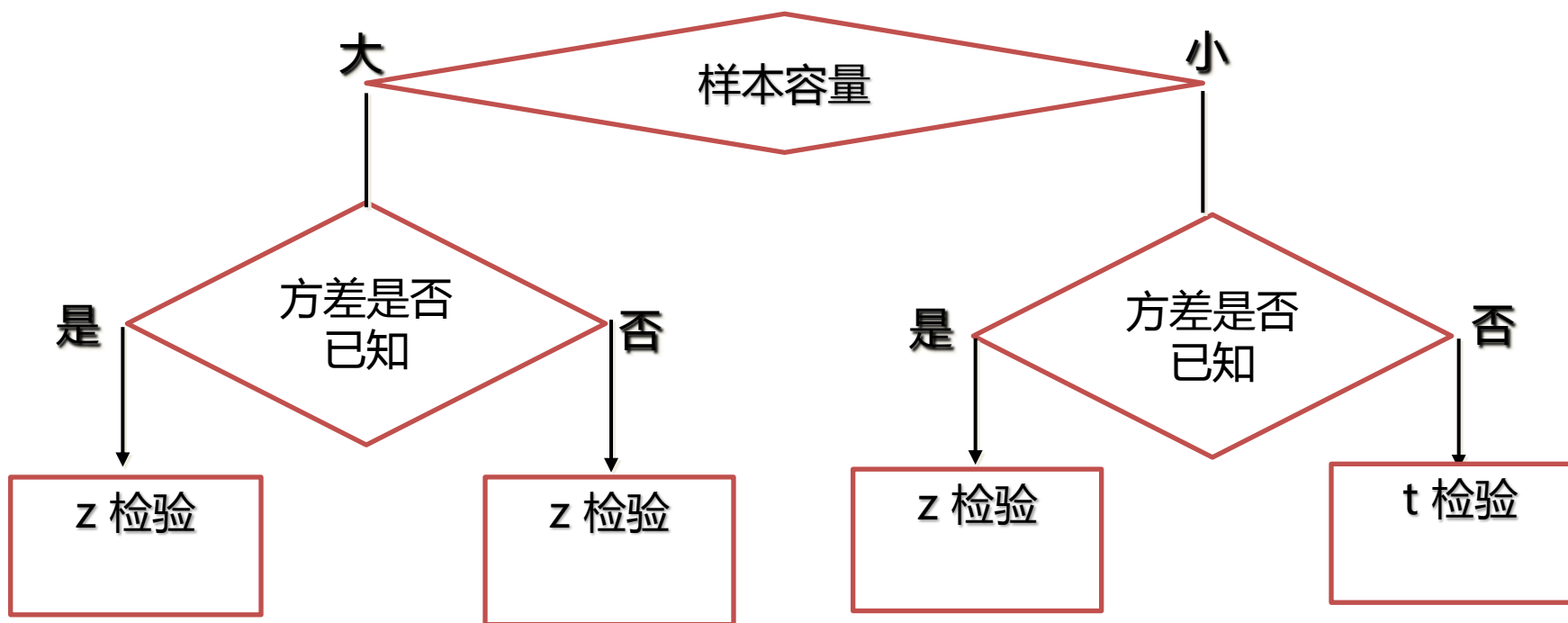
$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1}{n_1} + \frac{S_2}{n_2}}}$$

$$|-0.84| < 1.96$$

Z	P值	差异程度
$\geq 2.58$	$\leq 0.01$	非常显著
$\geq 1.96$	$\leq 0.05$	显著
$< 1.96$	$> 0.05$	不显著



# AB测试结果验证-假设检验



**Z检验、T检验、G检验、秩和检验、卡方检验等**

**谢谢大家**

**欧舟（翰承）  
工号：44063  
分机：30016  
旺旺：ozkorea@hotmail.com**