

# SACC

卓越 5周年 变迁

SequeMedia  
盛拓传媒

IT168  
www.it168.com

ChinaUnix

ITPUB

## 2013中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2013

大数据下的IT架构变迁

# 腾讯个性化精准推荐平台架构

腾讯公司 洪坤乾

kontenhong@tencent.com

# 大纲

## Part I 精准推荐初识

## Part II 腾讯精准推荐历程

## Part III 系统架构及实现

## Part IV 精准推荐体系介绍

## Part V 对未来的思考

关于内容

**重**系统架构以及要解决的问题

**轻**推荐策略、算法模型

# 大纲

## Part I 精准推荐初识

## Part II 腾讯精准推荐历程

## Part III 系统架构及实现

## Part IV 精准推荐体系介绍

## Part V 对未来的思考

# 精准推荐初识（1）



## 热度推荐

- 简单易行
- 盲目，效果不好





### 个性化推荐

- 有针对性，效果好
- 有一定复杂度
- 缺乏时效性

# 精准推荐初识 ( 3 )



## 个性化实时推荐

- 精准，效果好
- 系统非常复杂

# 行业状况

amazon.com

2012 10亿

facebook

2013Q2 16亿

Tencent 腾讯

2013 Q2 12.9亿



# 大纲

## Part I 精准推荐初识

## Part II 腾讯精准推荐历程

## Part III 系统架构及实现

## Part IV 精准推荐体系介绍

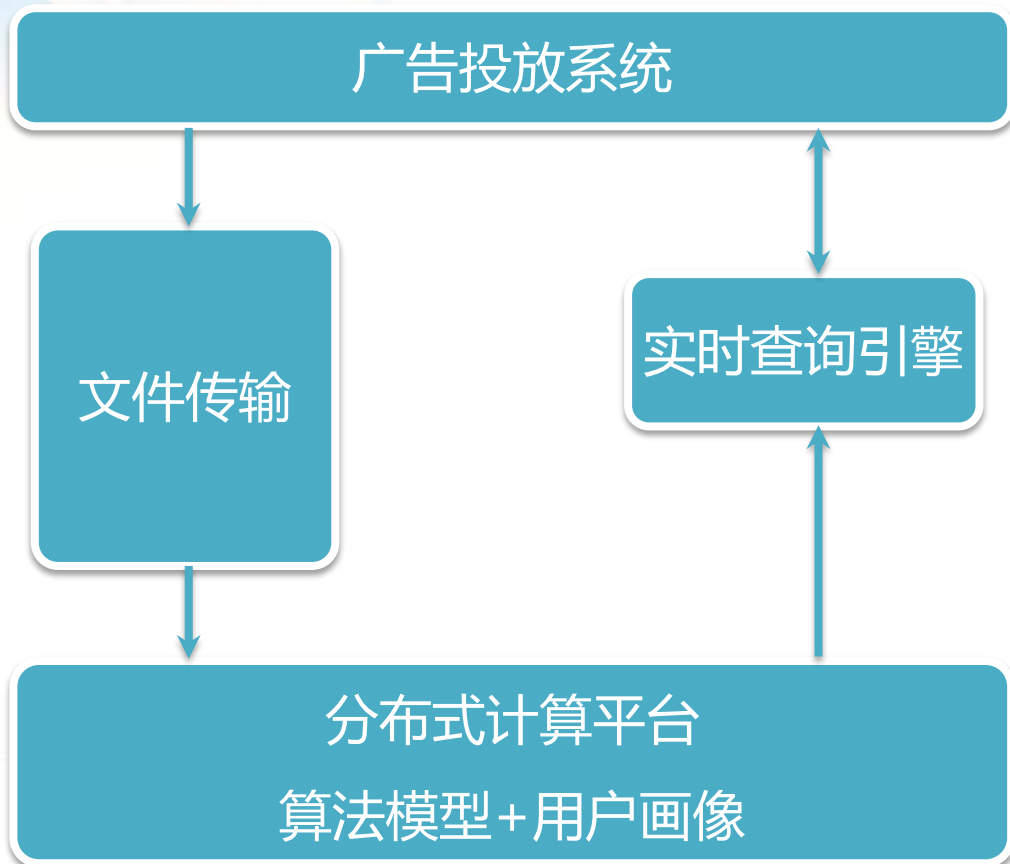
## Part V 对未来的思考

2009年的一天

目标：

**剩余流量变现**

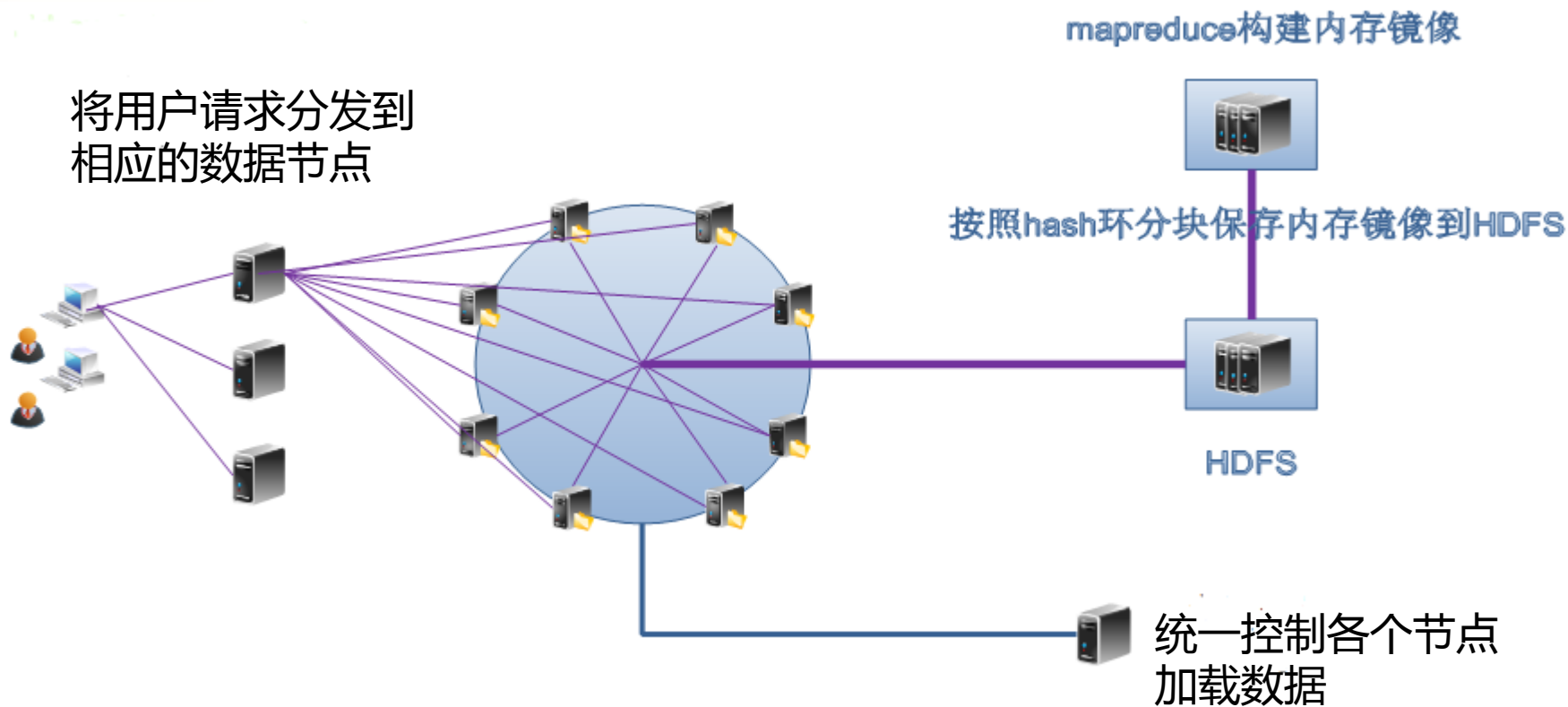
# 第一代架构



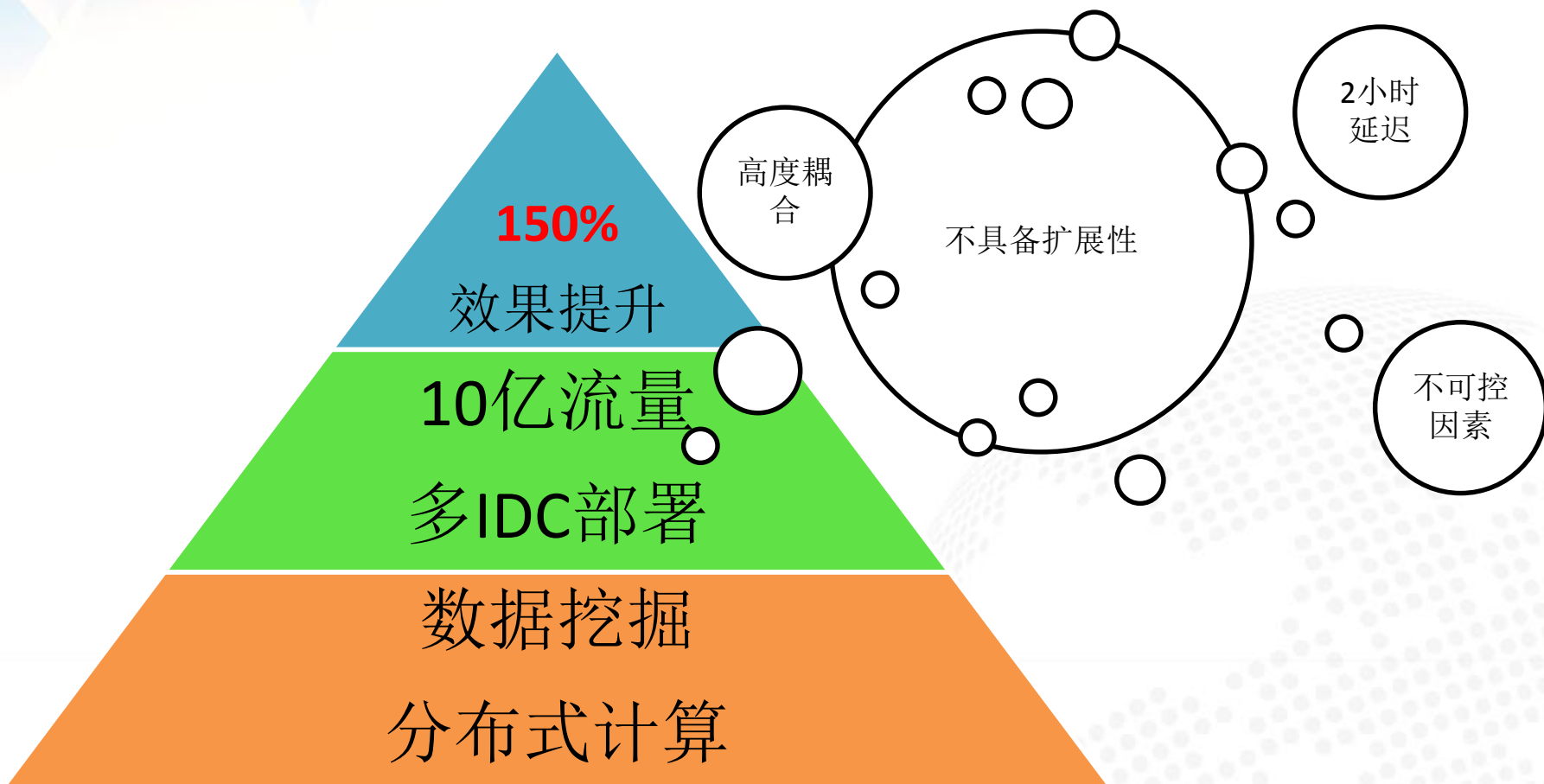
核心特征：

- 离线文件传输
- 批量暴力计算
- 实时匹配查询

# 实时查询引擎



# 效果





# 2010年的一天



腾讯科技 > 互联网报道 > 互联网新闻 > 正文

## 腾讯联姻康盛创想 发力与中小网站开放共赢

2010年08月23日 16:56 腾讯科技 我要评论 (30) 字号: T | T

**腾讯科技讯** 8月23日,中国服务用户最多的互联网企业腾讯公司和中国最大的社区软件平台及服务提供商康盛创想(Comsenz)联合宣布,双方已达成收购协议,康盛创想将成为腾讯的全资子公司,具体交易将在近期完成。

作为中国最专业的社区软件平台开发商之一,康盛创想旗下的Discuz!是全球范围内用户量最大、覆盖面最广的社区软件平台,拥有超过140万家网站用户,也是企业和网站主首选的社区与网站建设平台。在Alexa中国排名前1500家网站中,超过70%的网站选择第三方社区平台时采用了Discuz!的解决方案。

当前：6亿用户×50广告位×500广告

未来：10亿用户×10万广告位×10万广告

60  
万  
倍

# 重构



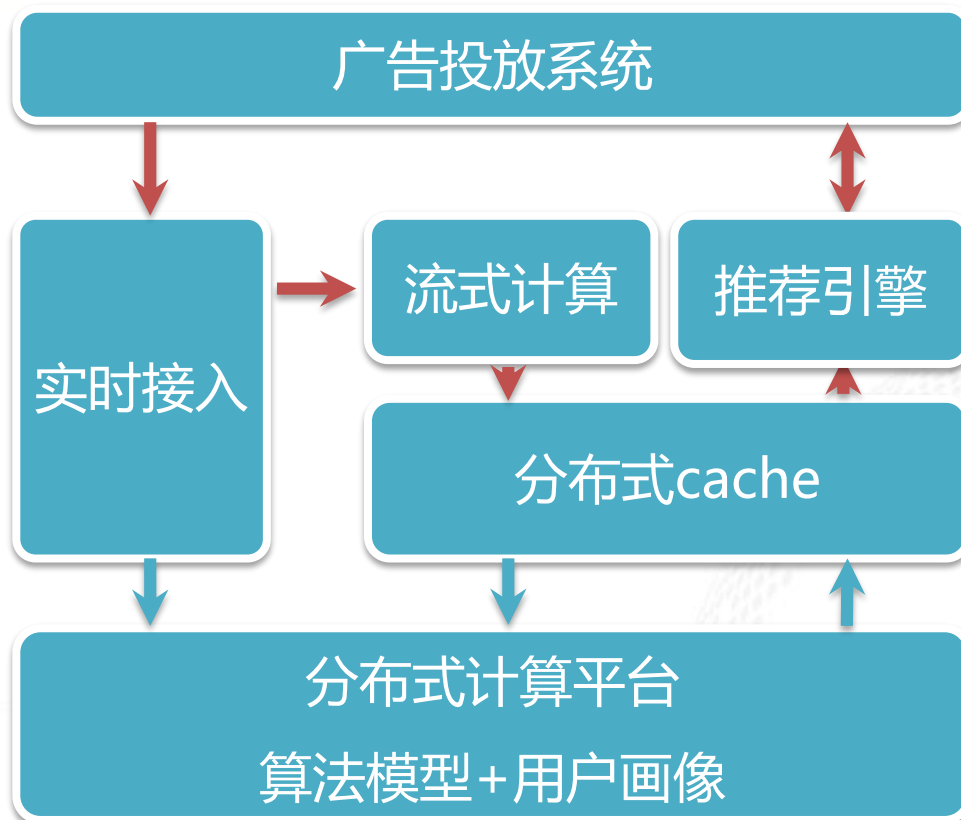
# 平台发展历程



# 第二代架构

特征：

- 实时数据传输
- 在线实时计算



优点：

- 秒级延迟
- 低耦合
- 易扩展

# 2013：持续系统优化

## 业务支撑

- 流量从十亿量级到百亿量级
- 业务种类从1个到10个

## 平台优化

- 分布式cache系统全面优化
- Storm替换S4
- 平台级组件抽象

## 效果

- 推荐效果进一步提升30%以上
- 收入增长



# 大纲

## Part I 精准推荐初识

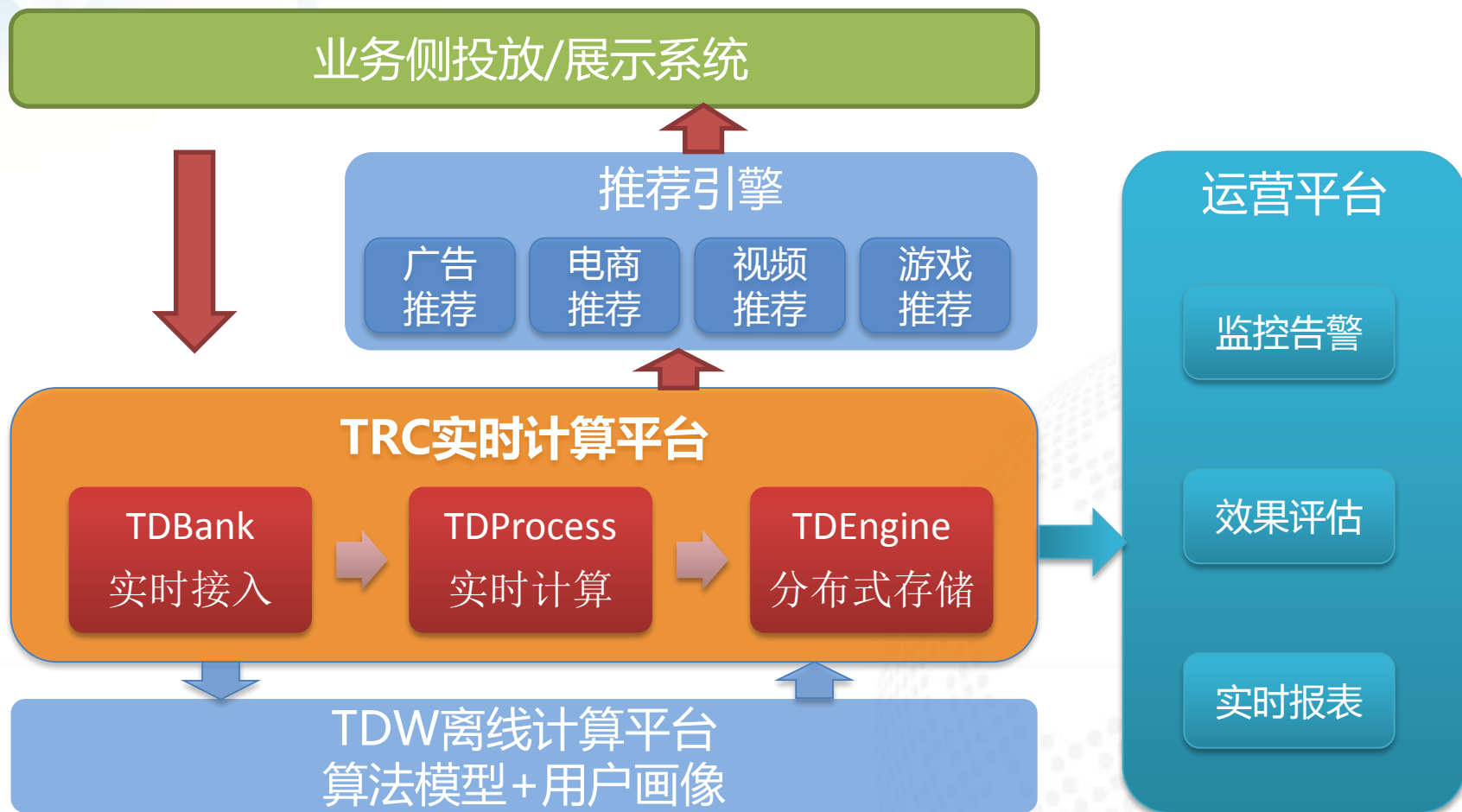
## Part II 腾讯精准推荐历程

## Part III 系统架构及实现

## Part IV 精准推荐体系介绍

## Part V 对未来的思考

# 加强型第二代架构



# 系统指标：工程师的噩梦

300亿：

每天支撑的推荐请求量

15000亿：

每天支撑的多维度交叉计算量

70万：

每秒钟推荐引擎支撑的请求量

7000万：

每秒钟推荐引擎实时预测的广告数量

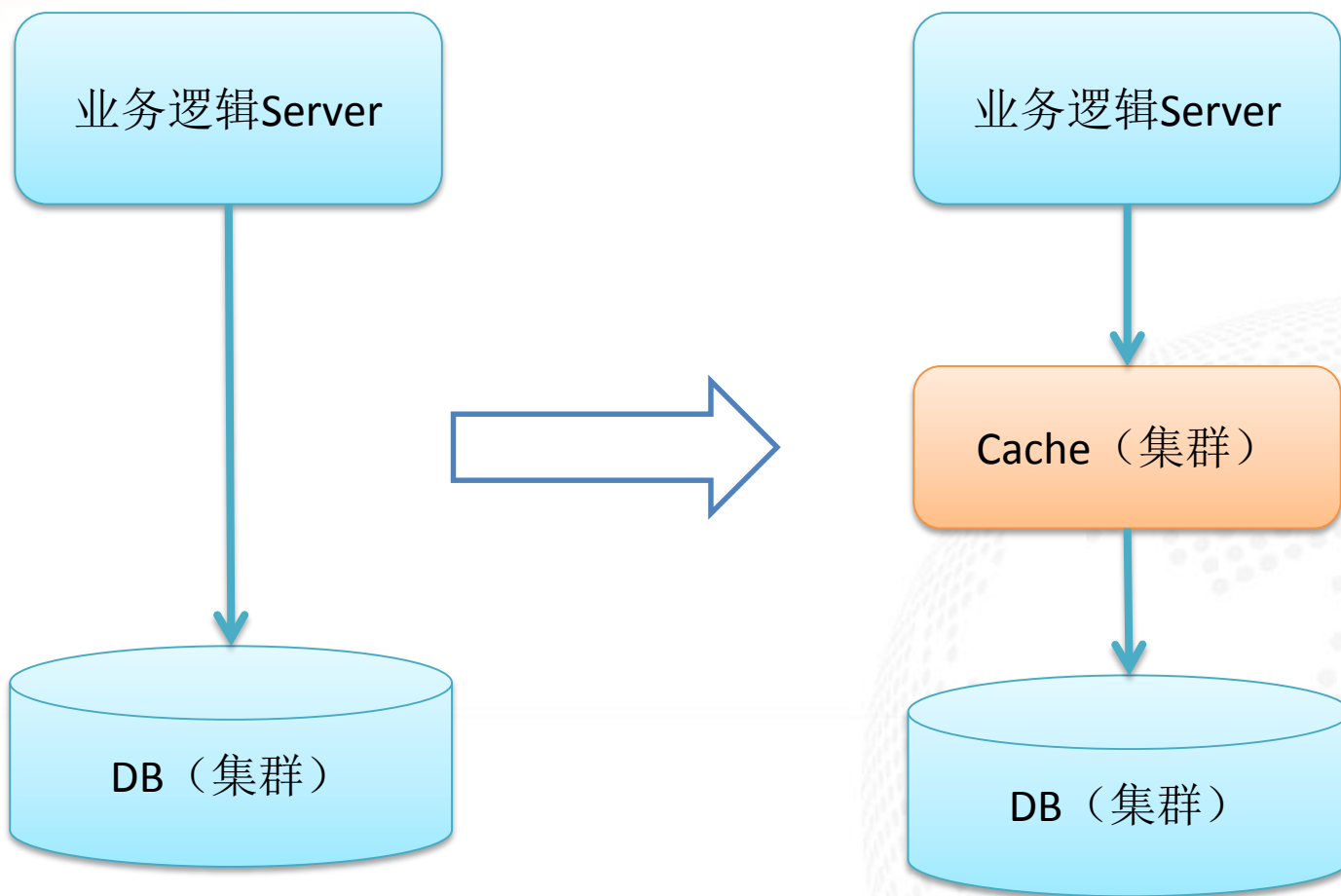
21亿：

每秒钟推荐引擎需要访问的数据次数

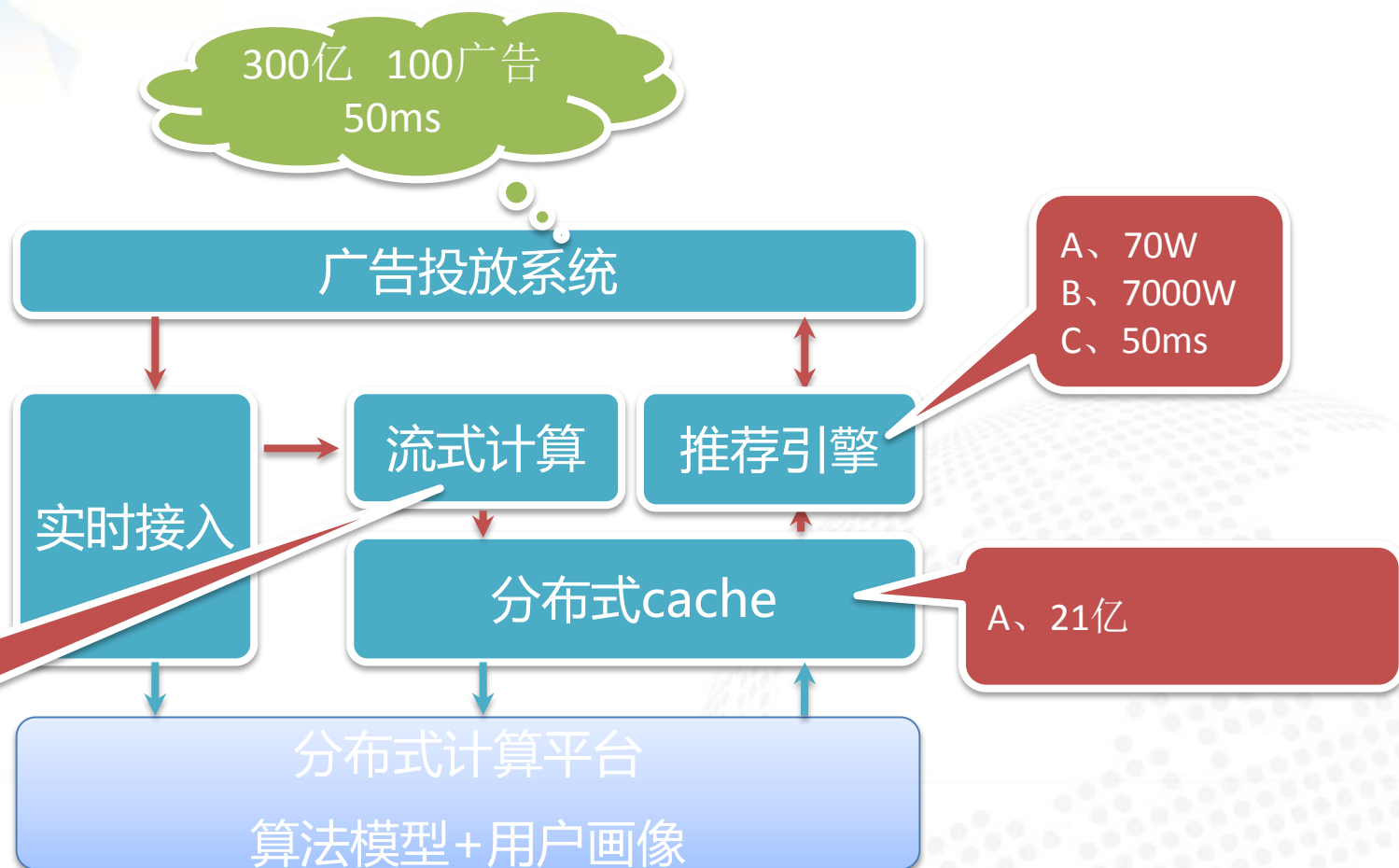
50毫秒：

每个请求的时延

# 实时系统常见架构设计



## 应对巨大的挑战





实时接入

流式计算

存储引擎

推荐引擎

数据源  
接入

在线消息

离线文件

Binlog

数据缓存  
预加工

数据  
传输

格式适配

加密传输

缓存  
分发

消息缓存

订阅分发

数据  
深加工

TRC  
实时计算

TDW  
离线计算

## 核心需求

秒级数据时延

数据管理方便

多种格式适配

## 特色功能

公网传输、加密

超大缓存、订阅

自定义分发

实时接入

流式计算

存储引擎

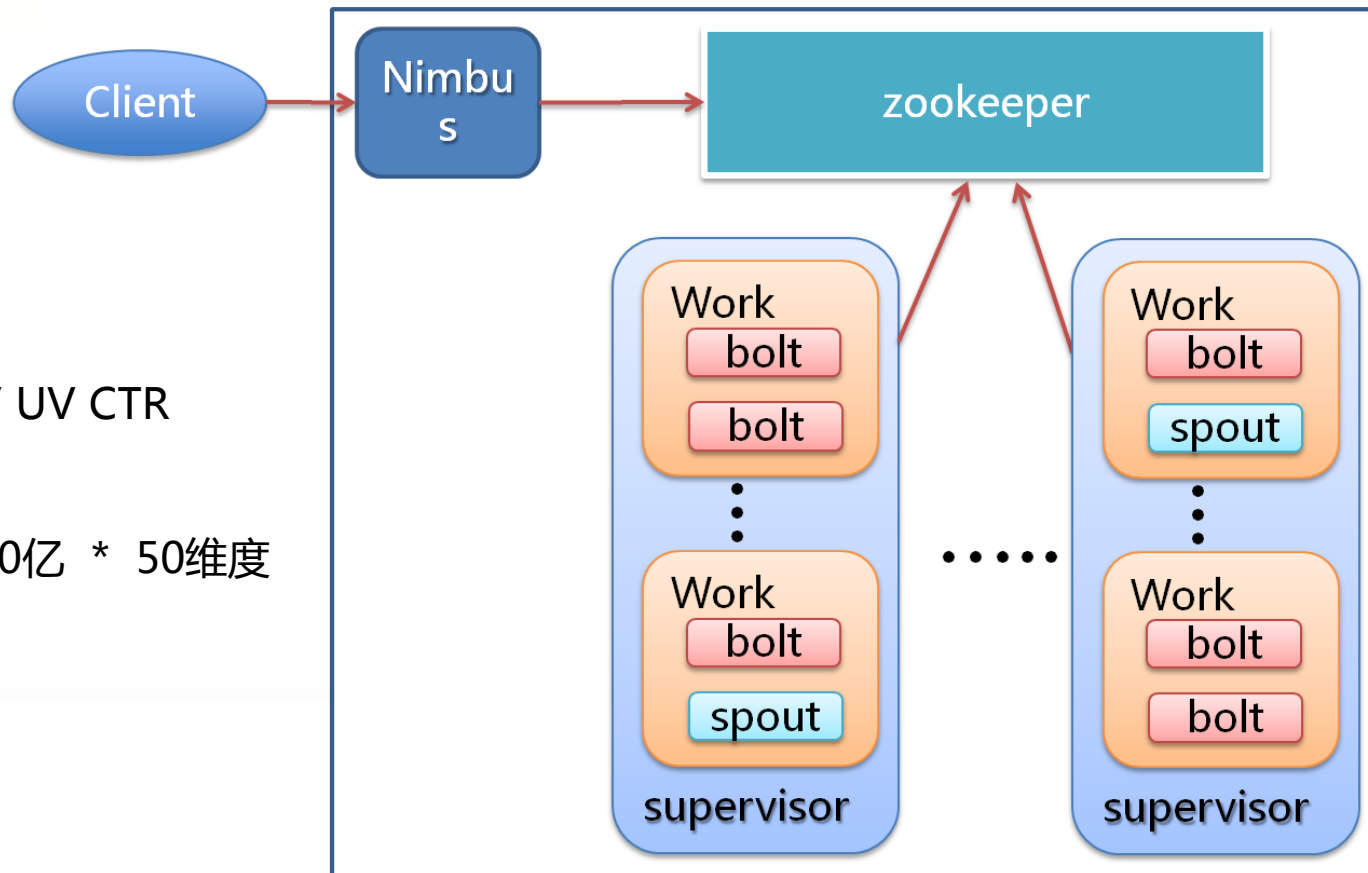
推荐引擎

## 核心需求

连续计算：PV UV CTR

秒级滑动窗口

计算能力：300亿 \* 50维度



实时接入

流式计算

存储引擎

推荐引擎

## 场景1：

当前时刻之前的10秒钟内，年龄在20-30岁、北京地区、男性用户、在QQ空间上 对IP5广告的点击率

## 场景2：

当前时刻用户A对男装类目的感兴趣程度

实时接入

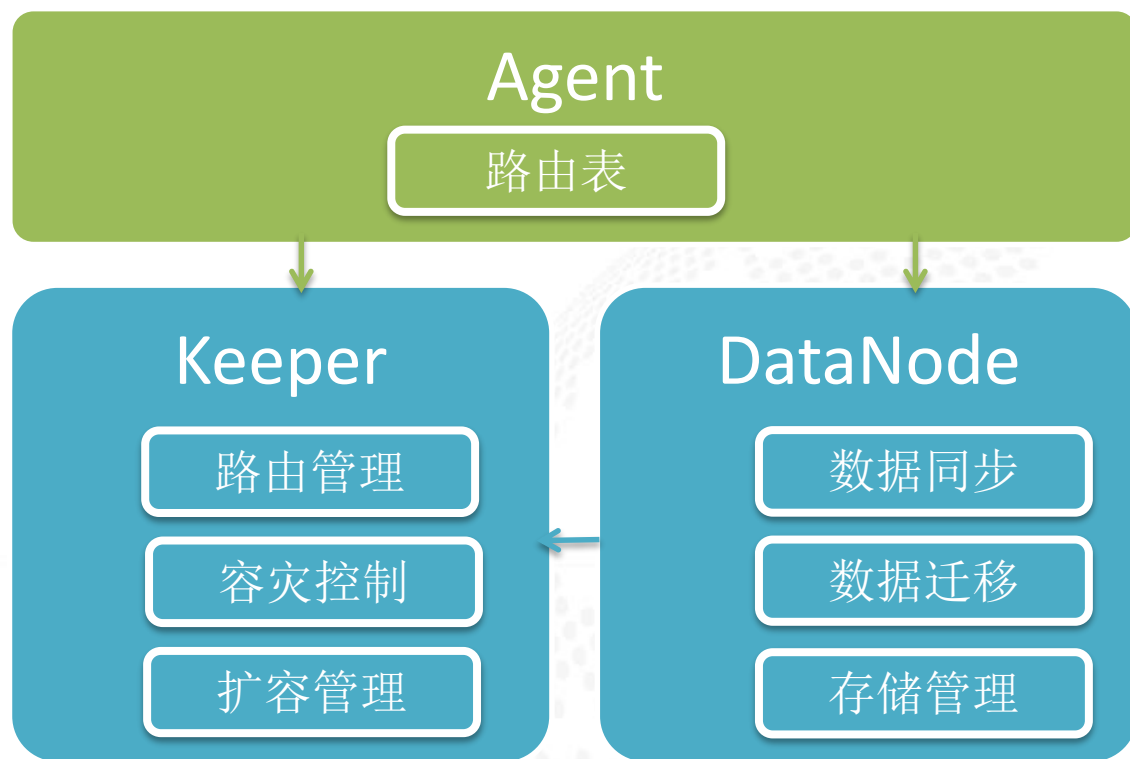
流式计算

存储引擎

推荐引擎

## 核心需求：

高并发、低延时、高可靠  
集群易管理、易扩容部署  
主备机同时服务  
多副本容灾

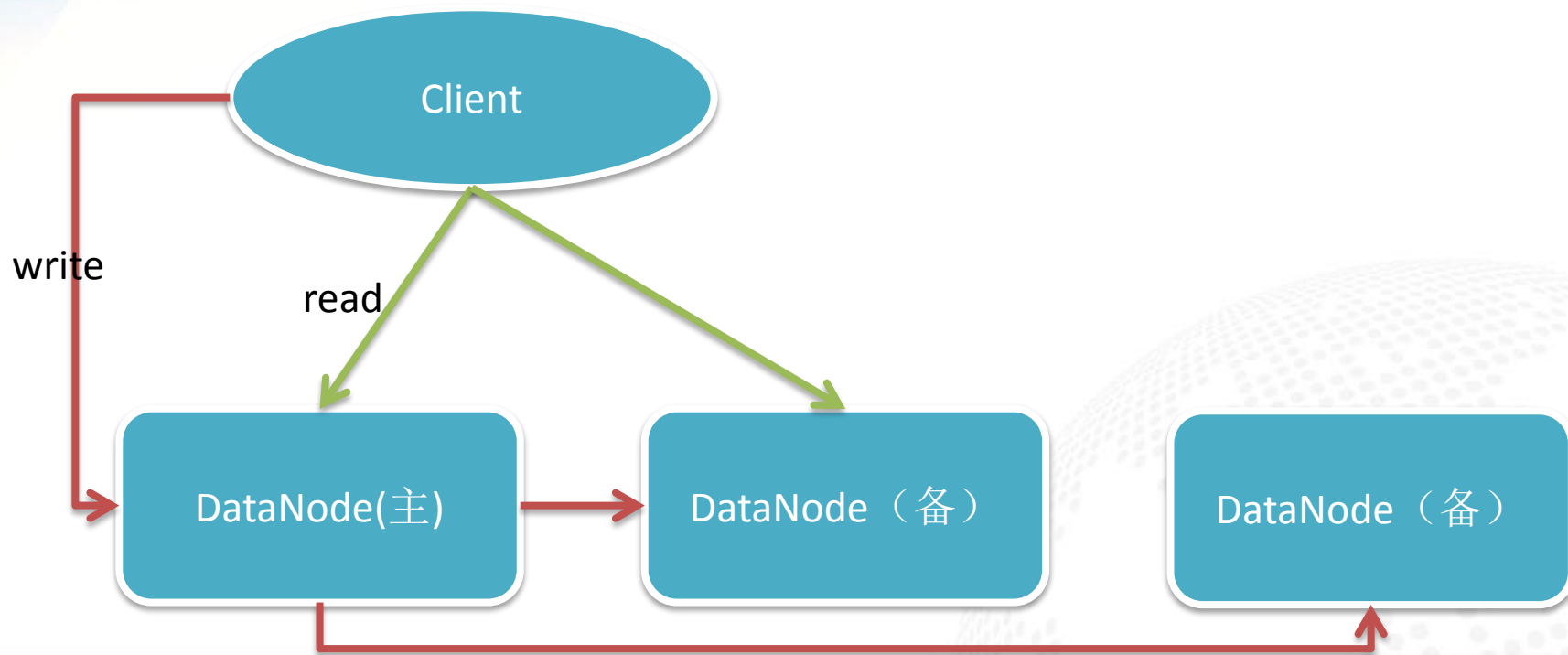


实时接入

流式计算

存储引擎

推荐引擎





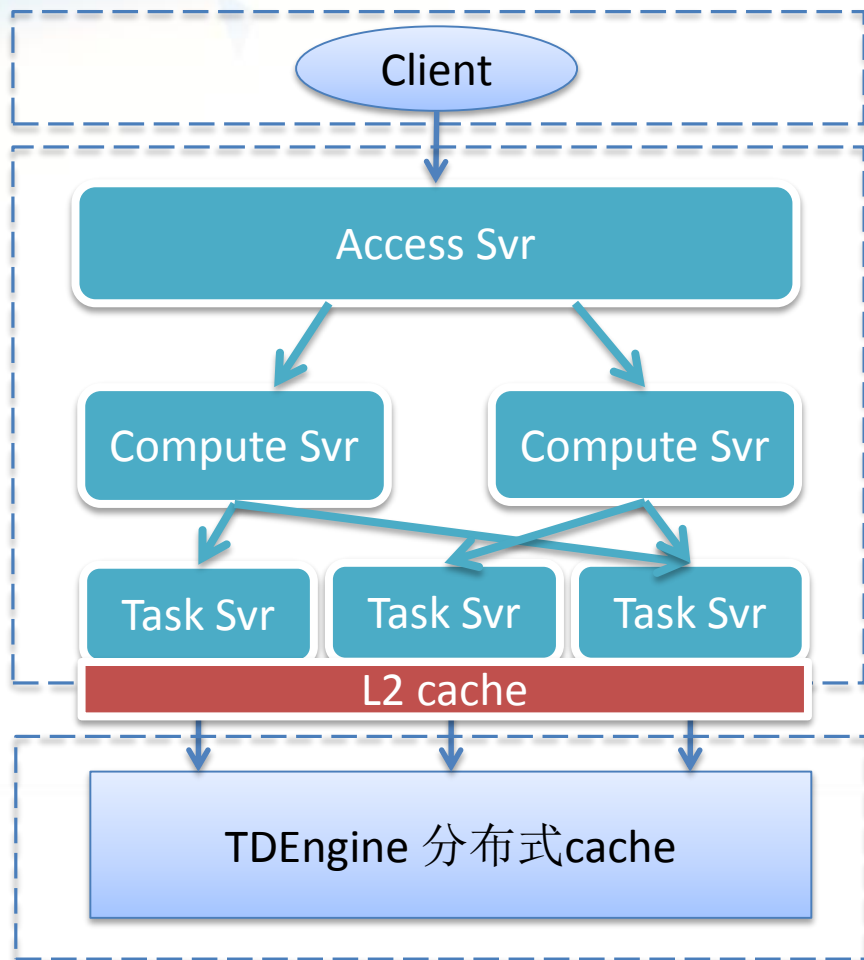
实时接入

流式计算

存储引擎

推荐引擎

应用客户端



**高并发**

300亿推荐流量  
30000亿广告预测

**低延时**

每个请求50ms返回

**重事务**

每个请求同时计算100个广告  
实时算法需要30种不同数据

**海量数据**

Billion级Feature  
T级推荐数据

# 大纲

## Part I 精准推荐初识

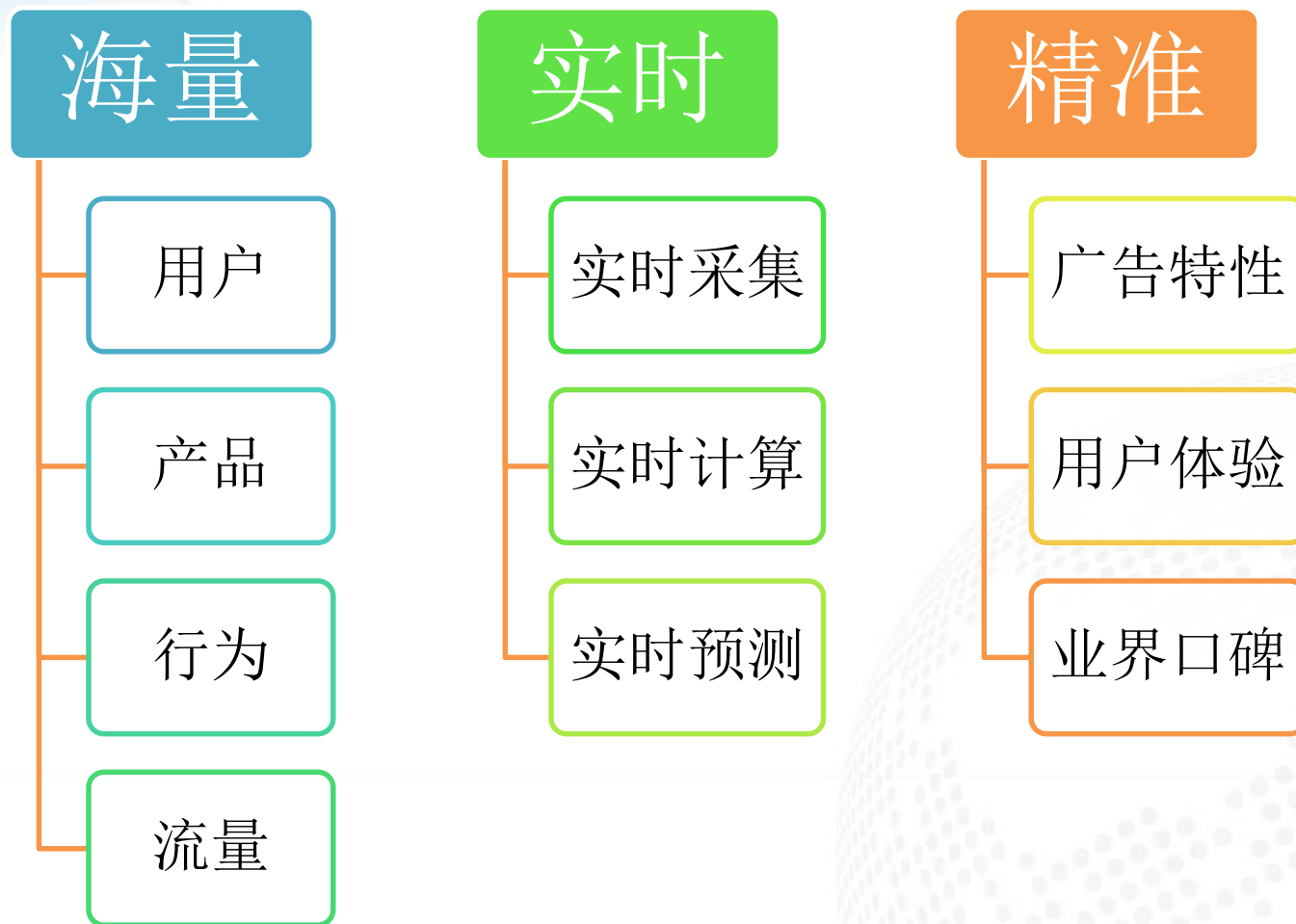
## Part II 腾讯精准推荐历程

## Part III 系统架构及实现

## Part IV 精准推荐体系介绍

## Part V 对未来的思考

# 精准推荐固有的挑战



基于“**数据+算法+系统**”的设计理念，为业务提供“全流程数据驱动的海量、实时、精准的个性化推荐服务”

# 腾讯个性化精准推荐之数据（1）



平均每天约40亿条行为数据

## 腾讯QQ

- 800mil
- 100bn

活跃用户8亿  
在线人际关系链达1000亿

## 微信

- 400mil

注册用户数突破4亿,

## 腾讯微博

- 570mil

注册用户5.7亿

## Qzone空间

- 600mil
- 400mil
- 1bn

活跃用户数达6亿  
日均相册上传超过4亿  
日写操作总数过10亿

## 腾讯视频

- 270mil

月度覆盖用户2.7亿

## 腾讯网

- 2.5 bn
- 160mil

pc日均浏览12亿, 手机 13亿  
日访问用户量1.6亿

...



# 腾讯移动分析 MTA.qq.com

腾讯移动分析是一款专业的移动应用统计分析工具，支持iOS和Android平台。开发者可以方便地通过嵌入统计SDK，实现对移动应用的全面监测，实时掌握产品表现，准确洞察用户行为

有多少人来过超市

新老顾客的比例

来自哪个小区

一般逛多长时间

在选薯片前后  
都看了什么

咨询、办卡、购买  
等您希望他做的事

访问超市



使用APP

有多少用户  
启动应用

新、老用  
户的占比

来自哪些  
地域、渠道

启动应用的  
时长、频率

访问了  
哪些页面

自定义事件

不仅仅是记录，移动APP统计还**分析每个环节**  
**透过现象看本质**



腾讯移动分析拥有六大独有优势，最好用的移动APP统计分析工具：

## 多维分析



多维度交叉分析  
运营有的放矢

## 用户画像



用户的兴趣，年龄，  
性别，职业尽在掌握

## 漏斗模型



结合业务流程  
多步漏斗转化分析

## 秒级实时



秒级实时，即时  
捕捉瞬息万变

## 运维监控



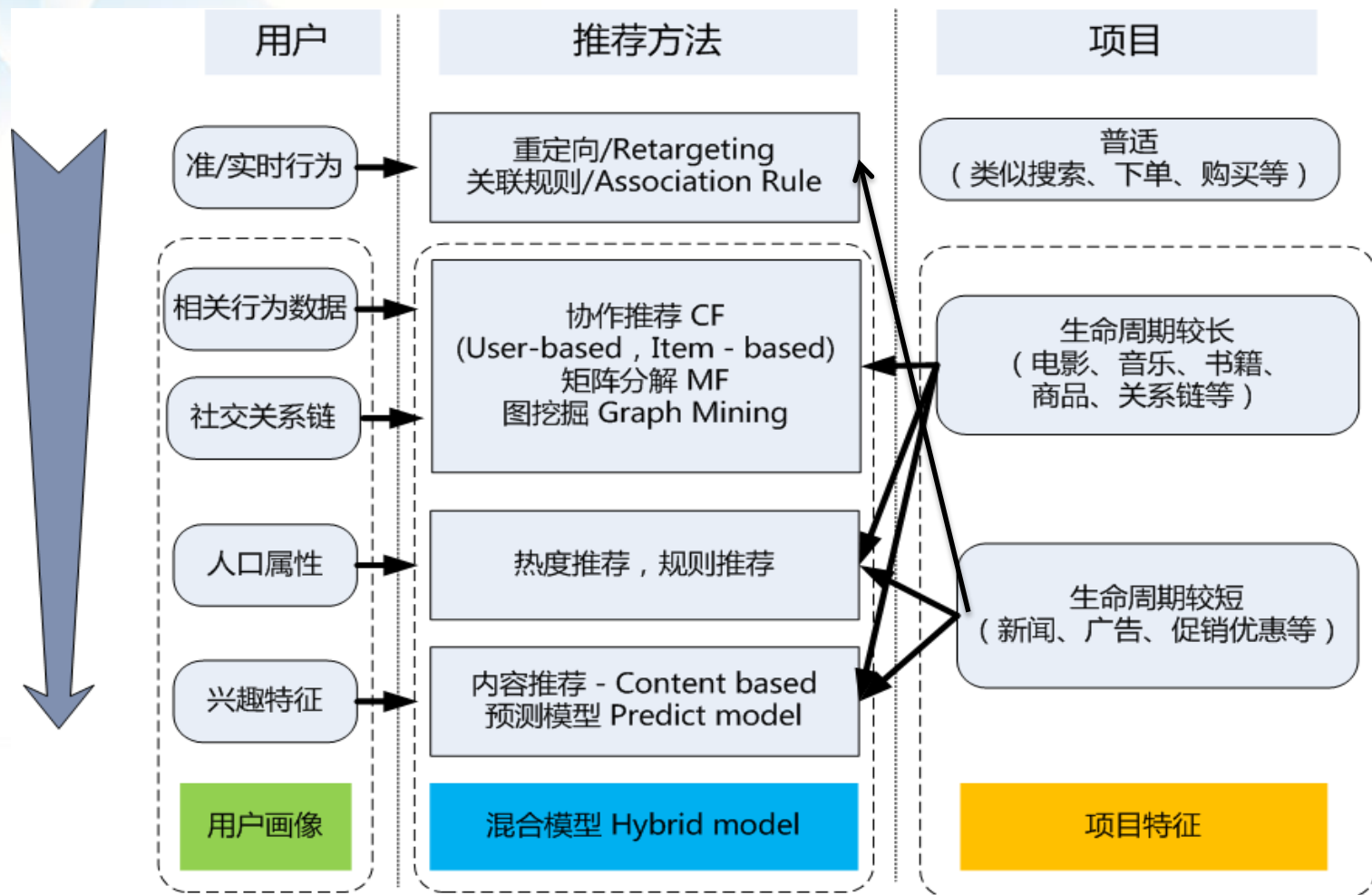
实时监控耗时，错  
误，放心运维

## 情报中心



洞察竞品和行业  
知己知彼百战不殆

# 腾讯个性化精准推荐之算法



上下文环境过滤

# 大纲

## Part I 精准推荐初识

## Part II 腾讯精准推荐历程

## Part III 系统架构及实现

## Part IV 精准推荐体系介绍

## Part V 对未来的思考

# 腾讯个性化精准推荐历程回顾



第一代

2009之前：  
离线定向投放



第二代

2009-2010：  
实时匹配查询



第三代

2011-2013：  
实时算法预测



第四代

2013-2014：  
实时模型训练



第五代

2015-：  
机器智能



# 前方，探索之路



## 更加实时

- 实时用户画像
- 实时模型训练

## 更加精准

- 复杂的算法
- 数据联动优势

## 更加通用

- 多业务支撑
- 移动端
- 第三方开放

# 未来，比你自已更懂自己



纵向  
进化

- 今天的行为，对未来会产生什么影响？

横向  
覆盖

- 你、你的家人、你的朋友喜欢什么？

？ ？



# Thanks

腾讯精准推荐平台:

[kontenhong@tencent.com](mailto:kontenhong@tencent.com)

# Thanks!

SequeMedia  
盛拓传媒

IT168.com  
www.it168.com

ChinaUnix.net

ITPUB