

# 第6-7讲、

## 电子商务系统 用户行为模型

郑小林 **XLZHENG@ZJU.EDU.CN**

# OPEN CASE: 《纸牌屋》的成功背后！



A NETFLIX ORIGINAL SERIES  
**HOUSE** of CARDS



# 《纸牌屋》的成功背后！

□ Netflix的自制剧《纸牌屋》在美国和其他40多个国家引起了观剧热潮，作为从DVD租赁转型到在线影视剧播放的公司，Netflix首次试水自制剧，就取得了令人吃惊的成功，它的秘诀是——通过大数据分析用户行为！



## 客户行为分析——》精准推荐

- Netflix早就意识到数据的重要性，它记录Netflix用户的位置、设备等信息，以及收藏、推荐、回放、暂停等动作，加上400万个评分、300万次搜索请求，等等一系列数据。
- 早些年，这些数据被Netflix用来进行DVD租赁业务的精准推荐。

# 客户偏好挖掘——个性化定制电影 MINING CUSTOMER PREFERENCE

Netflix基于其3000万北美用户观看视频时留下的行为数据，预测出“凯文·史派西”、“大卫·芬奇”和“BBC出品”三种元素结合在一起的电视剧产品将会大火特火，由此大胆的在纸牌屋的拍摄、发布方式上做了一系列革新。



用户能对影片做出1 - 5的评级，产生超过了几十亿条的大数据，Netflix使用推荐算法和软件来标识具有相似品味的观众对影片可能做出的评级。

# 用户偏好挖掘——个性化定制电影

相比传统收视率统计只抽取数千个样本户，《纸牌屋》的数据库包含了3000万用户的收视选择、400万条评论、300万次主题搜索。

拍什么、谁来拍、谁来演、怎么播，都由数千万观众的客观喜好统计决定。

从受众洞察、受众定位、受众接触到受众转化，每一步都由精准细致高效经济的数据引导，从而实现大众创造的C2B，即由用户需求决定生产。

# 用户偏好挖掘——个性化定制电影

- Netflix根据数据技术推导出《纸牌屋》的关键要素，喜欢BBC剧集的用户、大卫·芬奇和凯文·史派西存在交集。
- 1亿美元买下版权、要求大卫·芬奇担任导演、凯文·史派西担任男主角，这些都是Netflix利用数据挖掘得出的结论。

## 2014年各大视频网站自制内容举措

引进海外版权，  
打造三档大型综  
艺节目；上线15  
季以上自制剧

推出700集自  
制剧，不间断午间剧  
场；推出黄健翔、  
刘欣综艺节目

投入3亿发展自  
制剧，优酷主打“全”，土  
豆主打“最”



郭涛、王岳伦执  
导自制剧上线；郭  
德纲《人瑞儿》推  
出

韩国团队打造综  
艺节目；推出贯穿  
全年的喜剧播出  
带“周三笑场”

# 案例总结 CONCLUSION

- 启示：光收集信息远远不够。  
    成败关键在于根据客户制定出  
    个体客户了解客户嘛？
- 了解客户需求非常重要
- 了解客户消费习惯非常重要
- 了解客户消费能力非常重要
- .....

# **OUTLINE**

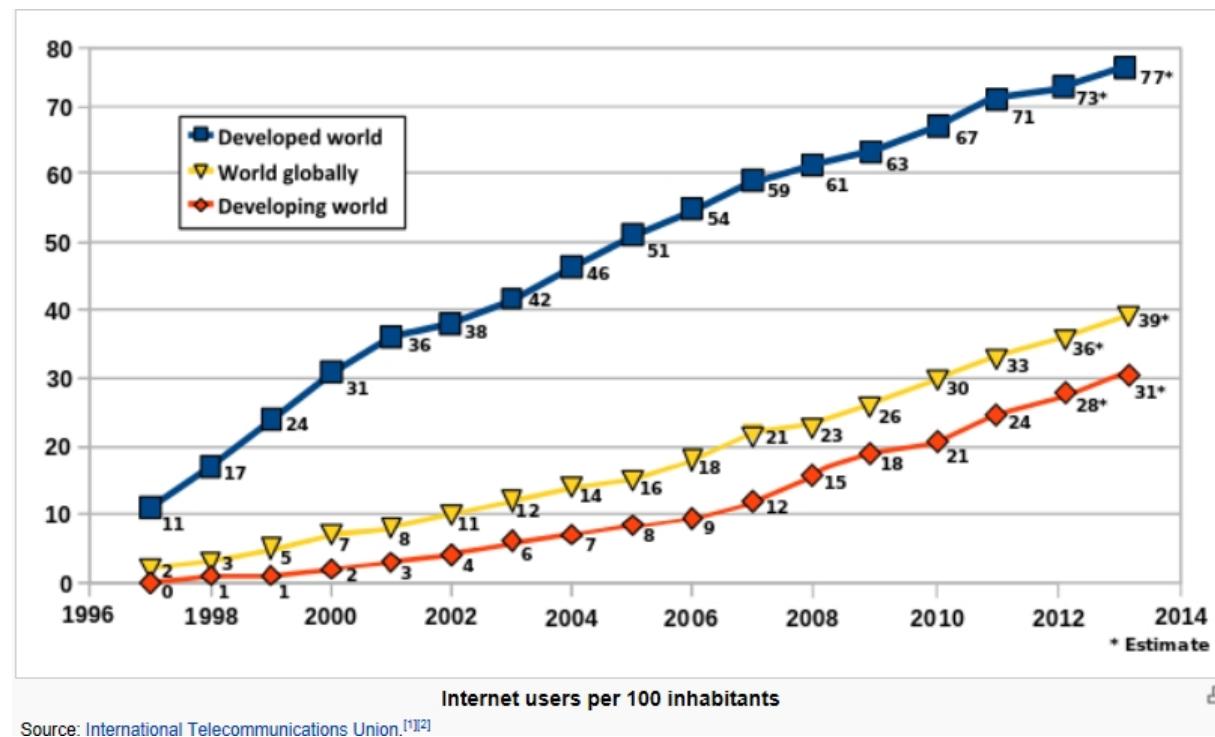
- 1. The Internet Audience and Consumer Behavior**
- 2. Customer Behavior Model**
- 3. Customer Visiting Model**
- 4. Customer/Server Interactive Diagram**
- 5. Case Study**
- 6. Mining of Web Site Logs**
- 7. Related Applications**

# 1. THE INTERNET AUDIENCE AND CONSUMER BEHAVIOR

## 1.1 Internet

### Audience

- Around 160 million Americans (56% of total population) have Internet access mid-2003
- Growth rate has slowed to less than 10% a year
- Intensity and scope of use both increasing



Source: International Telecommunications Union.<sup>[1][2]</sup>

# TOP 10 MOST POPULAR INTERNET ACTIVITIES (2002)

TABLE 7.1

TOP 10 MOST POPULAR INTERNET ACTIVITIES (2002)

ACTIVITY	PERCENTAGE OF USERS
E-mail and instant messaging	87.9%
Web surfing/browsing	76.0%
Reading news	51.9%
Accessing entertainment information	46.4%
Shopping and buying online	44.5%
Hobbies	43.7%
Travel information	36.2%
Medical information	35.5%
Games	26.5%
Tracking credit cards	24.2%

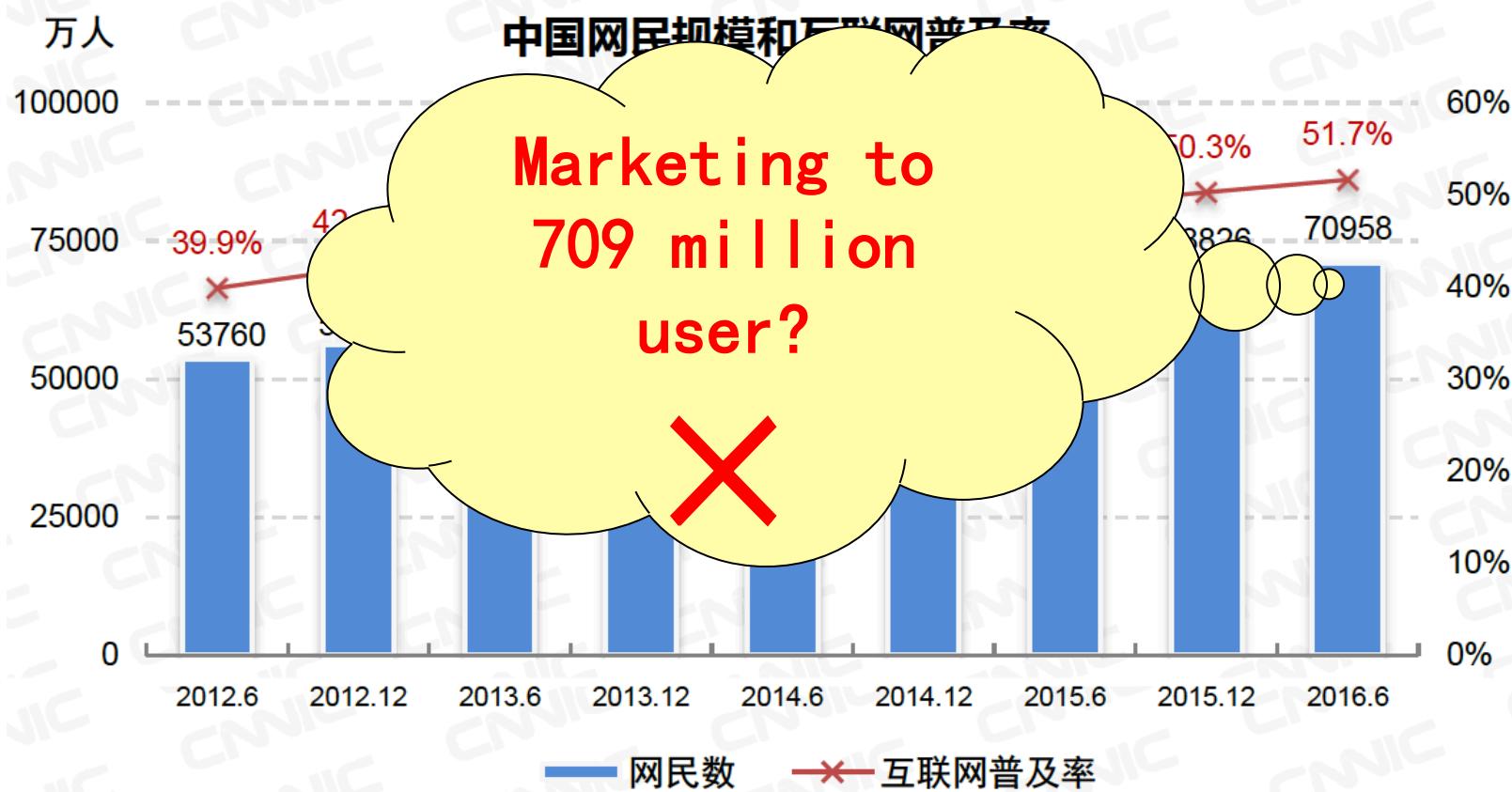
SOURCE: UCLA Center for Communication Policy, 2003.

表 5 2015.12-2016.6 中国网民各类互联网应用的使用率

	2016.6		2015.12		
应用	用户规模(万)	网民使用率	用户规模(万)	网民使用率	半年增长率
即时通信	64177	90.4%	62408	90.7%	2.8%
搜索引擎	59258	83.5%	56623	82.3%	4.7%
网络新闻	57927	81.6%	56440	82.0%	2.6%
网络视频	51391	72.4%	50391	73.2%	2.0%
网络音乐	50214	70.8%	50137	72.8%	0.2%
网上支付	45476	64.1%	41618	60.5%	9.3%
网络购物	44772	63.1%	41325	60.0%	8.3%
网络游戏	39108	55.1%	39148	56.9%	-0.1%
网上银行	34057	48.0%	33639	48.9%	1.2%
网络文学	30759	43.3%	29674	43.1%	3.7%
旅行预订 <sup>6</sup>	26361	37.1%	25955	37.7%	1.6%
电子邮件	26143	36.8%	25847	37.6%	1.1%

## 1.2 MARKETING OBJECT

### THE INTERNET AUDIENCE AND CONSUMER BEHAVIOR



来源 : CNNIC 中国互联网络发展状况统计调查

2016.6

# INTERNET AUDIENCE AND CONSUMER BEHAVIOR

- Demographics and access: some demographic groups have much higher percentages of online usage than other groups.
- Demographics to examine include:
  - Income
  - Age
  - Ethnicity
  - Education
  - Gender

# CHANGING DEMOGRAPHIC DIFFERENCES IN INTERNET ACCESS

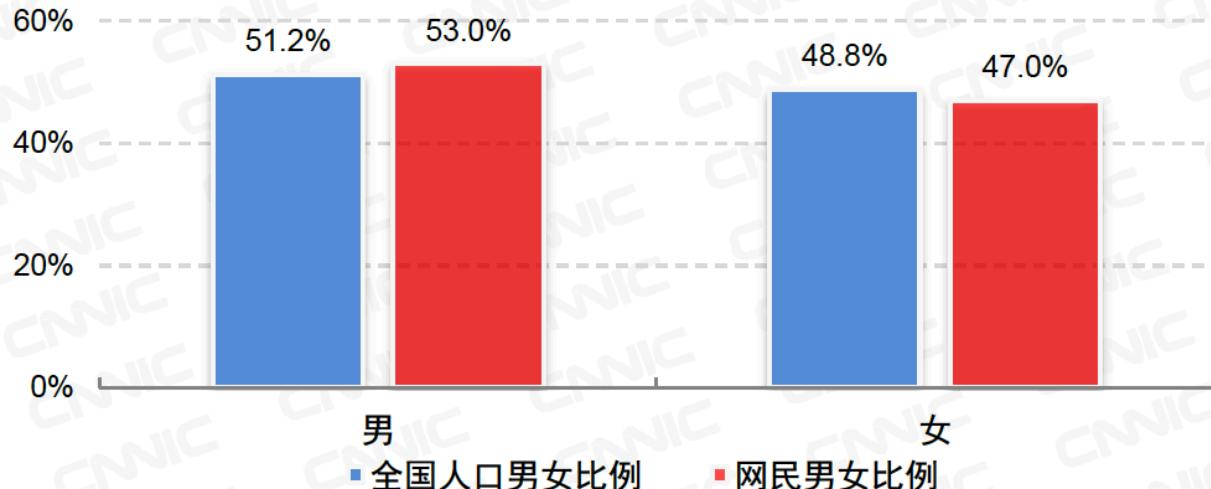
TABLE 7.3

## CHANGING DEMOGRAPHIC DIFFERENCES IN INTERNET ACCESS, THE PERCENT OF EACH GROUP ONLINE

	2 0 0 0	2 0 0 2
<i>All Adults</i>	50%	50%
Men	51%	60%
Women	46%	56%
Whites	50%	60%
Blacks	34%	45%
Hispanics	43%	54%
<i>Age Cohorts</i>		
18–29	69%	74%
30–49	60%	67%
50–64	45%	52%
65+	14%	18%
<i>Income Brackets</i>		
Under \$30,000	31%	38%
\$30,000–\$50,000	52%	65%
\$50,000–\$75,000	67%	74%
\$75,000+	78%	86%
<i>Educational Attainment</i>		
High school or less	34%	45%
Some college	63%	72%
College degree or more	75%	82%

SOURCE: Lenhart, et al., 2003.

## 中国网民性别结构

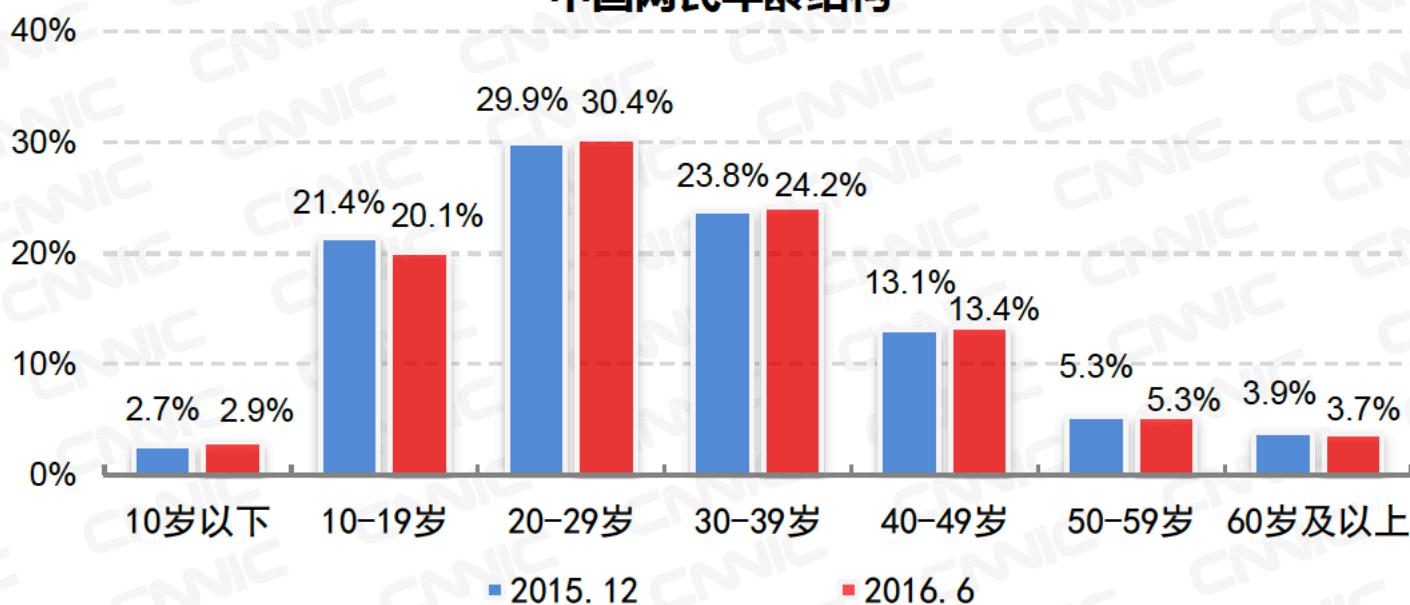


来源: CNNIC

中国互联网络发展状况统计调查

2016.6

## 中国网民年龄结构

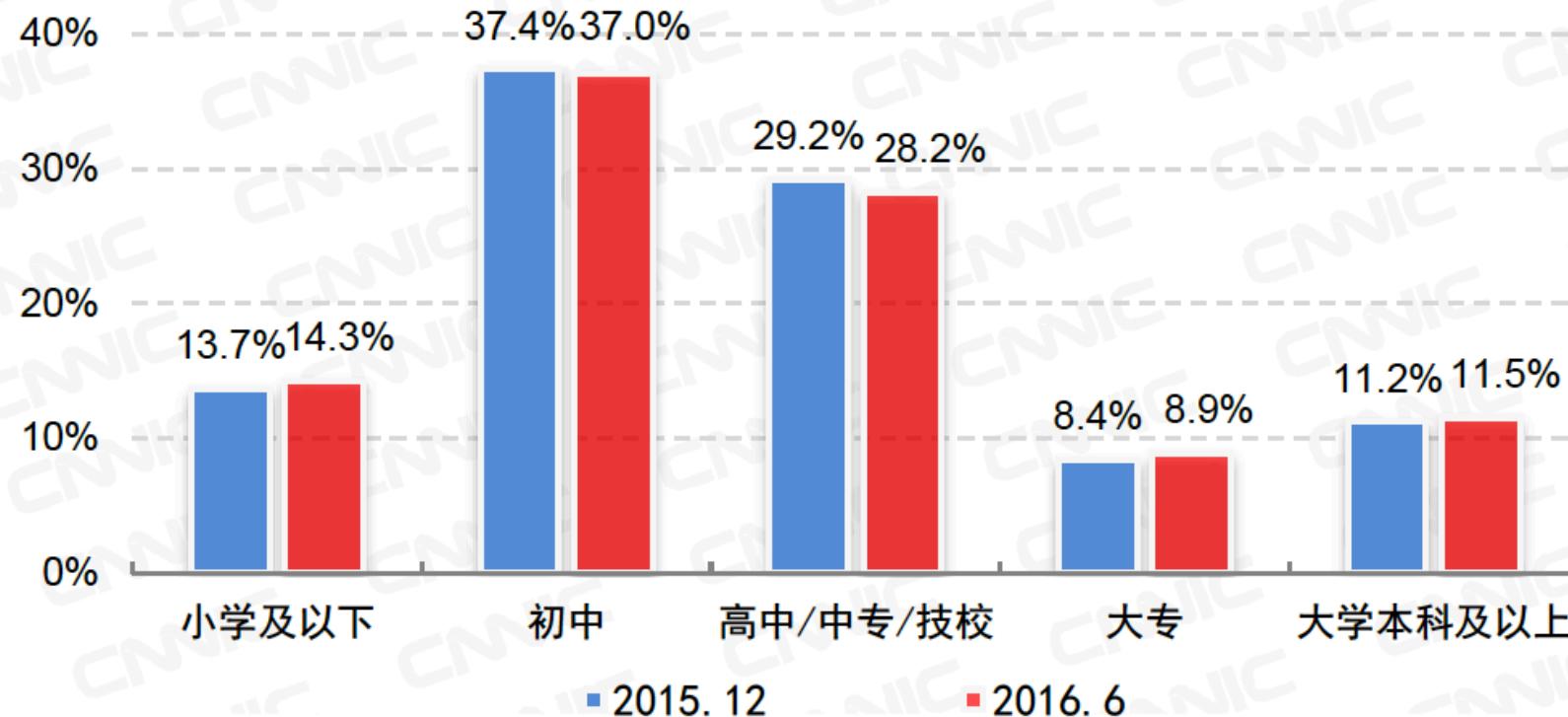


来源: CNNIC

中国互联网络发展状况统计调查

2016.6

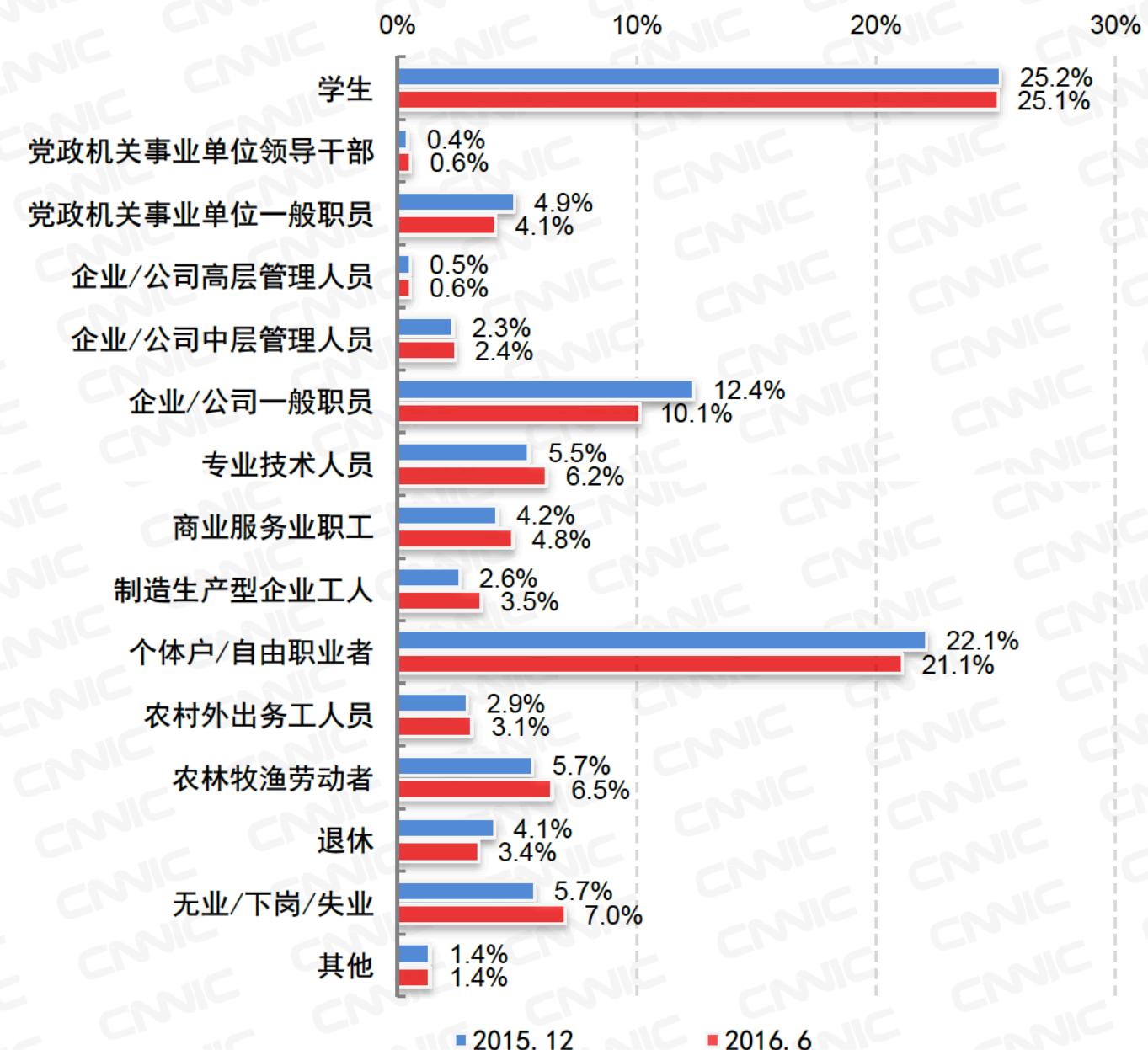
## 中国网民学历结构



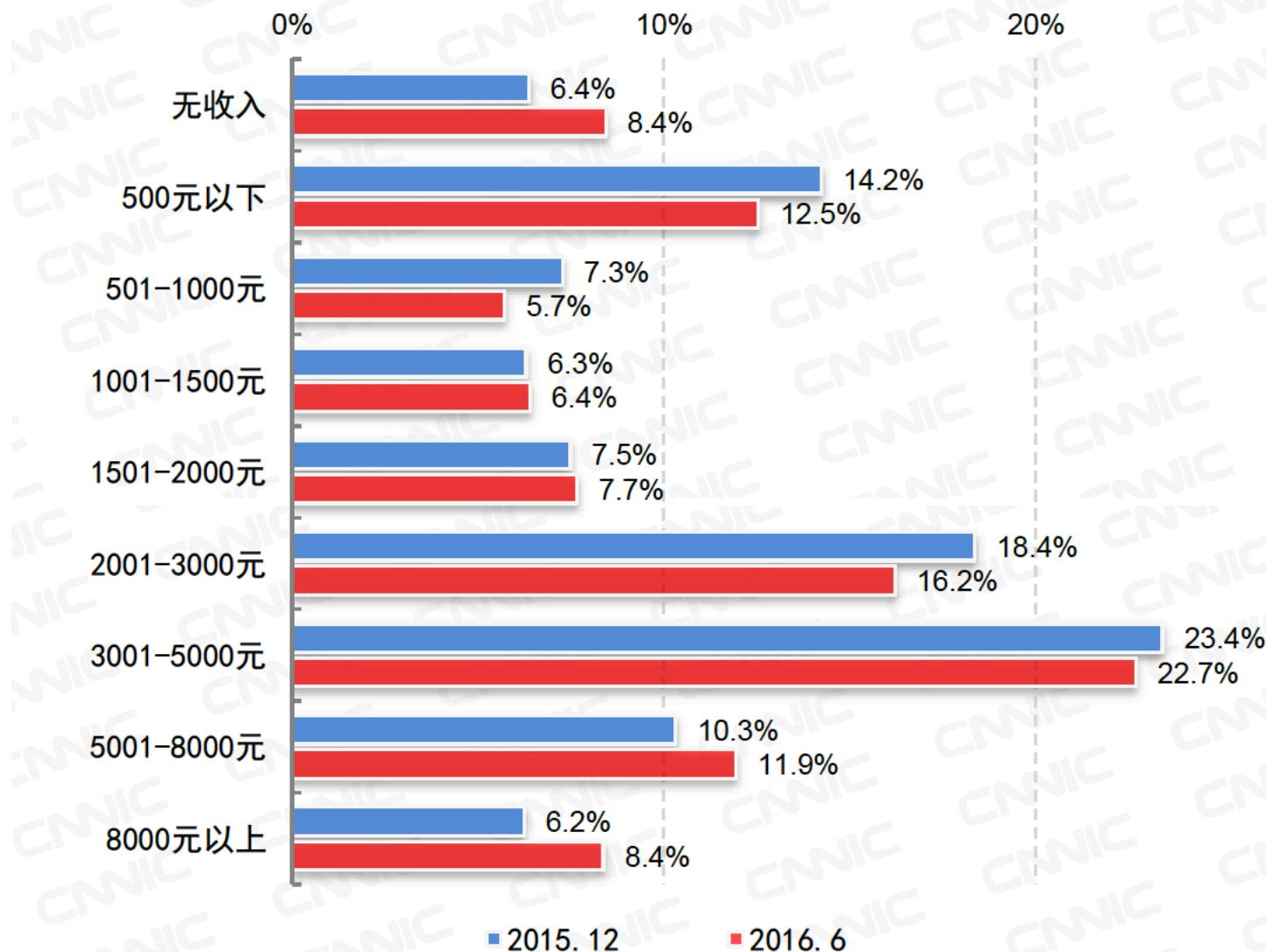
来源：CNNIC 中国互联网络发展状况统计调查

2016.6

## 中国网民职业结构



## 中国网民个人月收入结构



## 1.3 CONSUMER BEHAVIOR MODELS

- Attempt to predict/explain **what** consumers purchase and **where, when, how much and why** they buy. （试图模拟和理解人们在市场环境中的行为方式的社会科学理论）；
- Consumer behavior models based on background demographic factors and other intervening, more immediate variables （消费行为模式：试图在消费者的**背景因素**、一系列的**干扰因素**以及许多影响消费者最终决策的**直接变量**的基础上，对消费者作出的各种决策进行预测。）

# **BACKGROUND DEMOGRAPHIC FACTORS**

(人口统计学因素)

- **Culture:** Shapes basic human values, wants perceptions and behaviors
- **Subculture:** Subset of culture; forms around major social differences such as ethnicity, age, lifestyle, geography
- **Direct reference group:** Include one's family, profession/occupation, religion, neighborhood, schools
- **Indirect reference group:** Includes one's life-cycle state, social class and lifestyle group
- **Opinion leaders (viral influencers):** Influence the behavior of others through their personality, skills or other factors, such as in twitter.

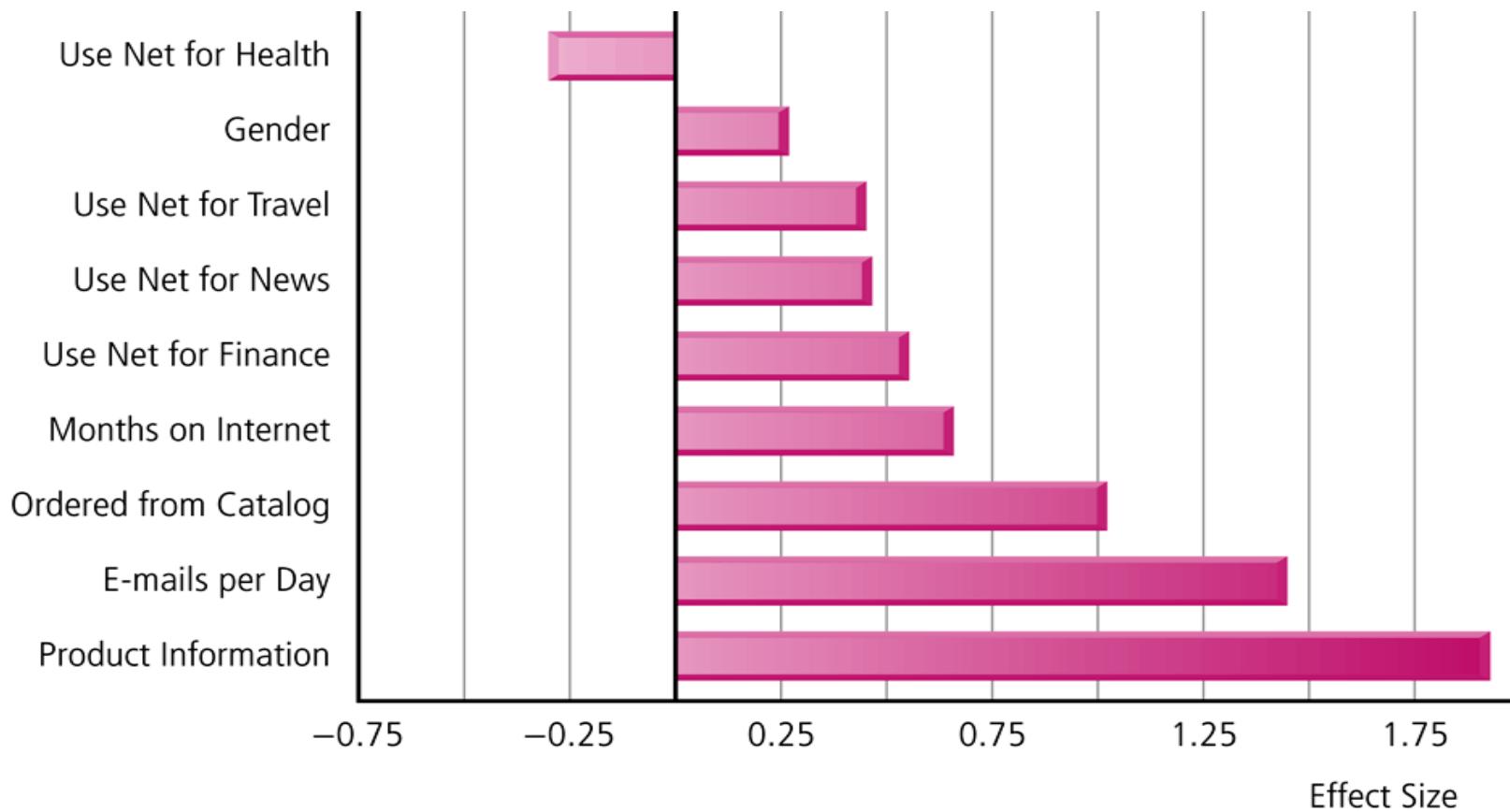
# **BACKGROUND DEMOGRAPHIC FACTORS (CONT'D)**

- Psychological profile: set of needs, drives, motivations, perceptions and learned behaviors
- Divides market into different groups based on demographic and psychological data

# A GENERAL MODEL OF CONSUMER BEHAVIOR



# FACTORS THAT PREDICT ONLINE BUYING BEHAVIOR

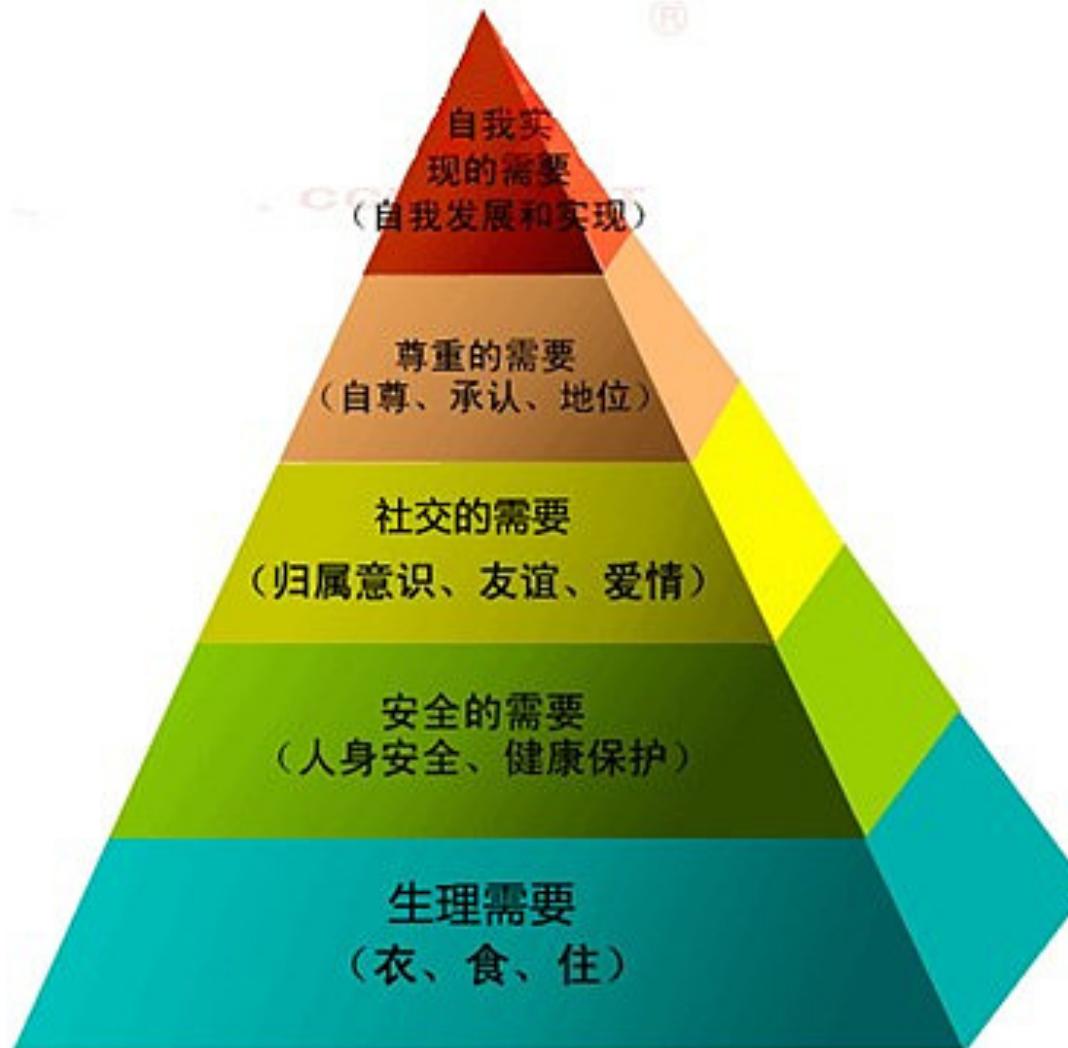


# **1.4 THE PURCHASING DECISION**

**□ Five stages in the consumer decision process:**

- Awareness of need**
- Search for more information**
- Evaluation of alternatives**
- Actual purchase decision**
- Post-purchase contact with firm**

# Awareness of need



马斯洛需要层次理论图

# THE CONSUMER DECISION PROCESS AND SUPPORTING COMMUNICATIONS

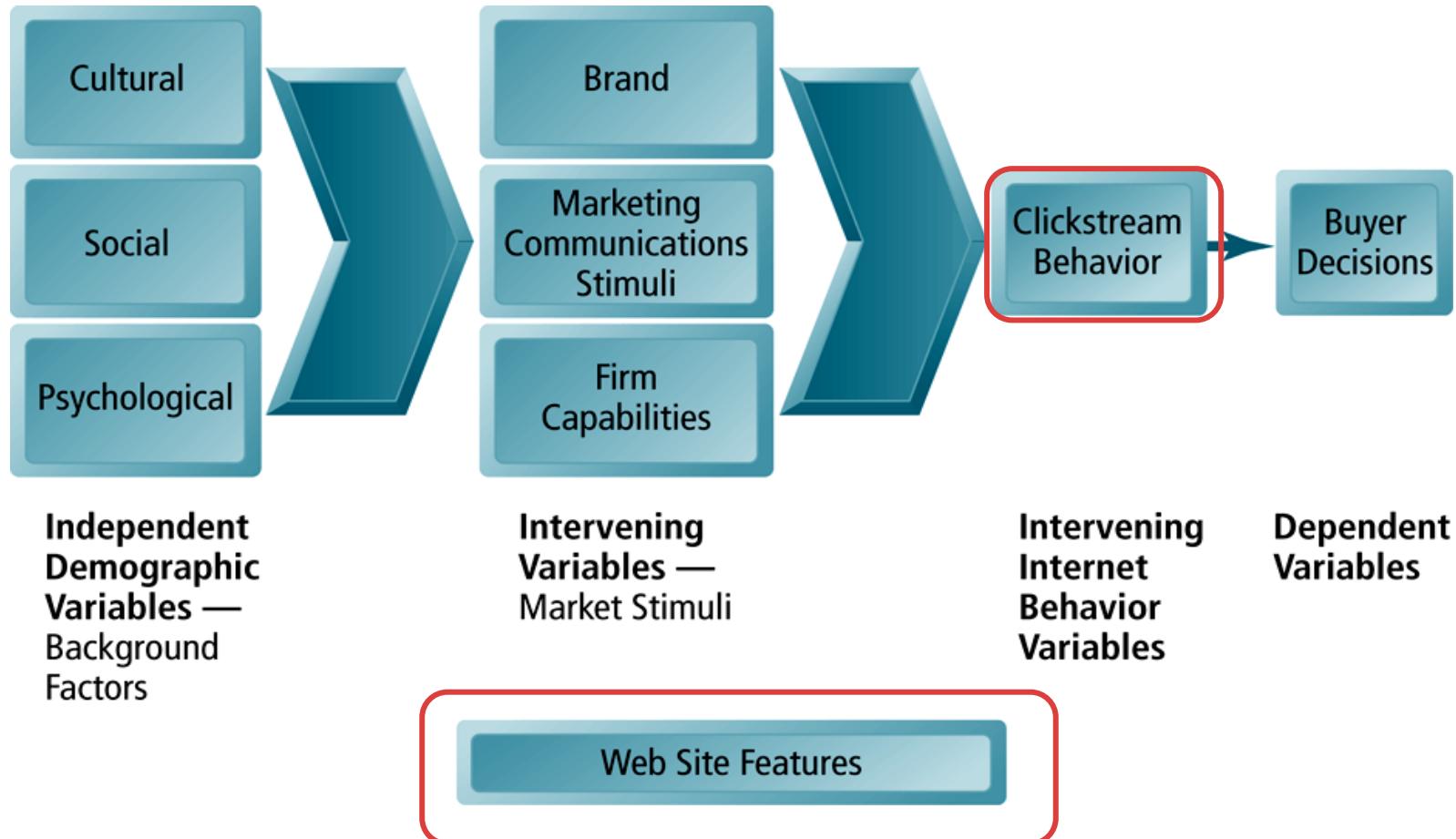
MARKET COMMUNICATIONS	Awareness—Need Recognition	Search	Evaluation of Alternatives	Purchase	Post-purchase Behavior—Loyalty
Offline Communications	Mass media TV, Radio Print media	Catalogs Print ads Mass media Sales people Product raters Store visits	Reference groups Opinion leaders Mass media Product raters Store visits	Promotions Direct mail Mass media Print media	Warranties Service calls Parts and Repair Consumer groups
Online Communications	Targeted banner ads Interstitials Targeted event promotions	Search engines Online catalogs Site visits Targeted e-mail	Search engines Online catalogs Site visits Product reviews User evaluations	Online promotions Lotteries Discounts Targeted e-mail	Communities of consumption Newsletters Customer e-mail Online updates

# **1.5 A MODEL OF ONLINE CONSUMER BEHAVIOR**

## **□ Adds two new factors:**

- Web site capabilities** – the content, design and functionality of a site
- Consumer clickstream behavior** – the transaction log that consumers establish as they move about the Web

# A MODEL OF ONLINE CONSUMER BEHAVIOR



# SEVEN TYPES OF ONLINE SESSIONS

ONLINE BEHAVIOR	SESSION LENGTH (MIN)	TIME PER PAGE (MIN)	SITE FAMILIARITY*	CATEGORY CONCENTRATION**
Quickies (finding sports, reading e-mail)	1	0.25	90%	90%
Just the Facts (finding and evaluating information from related sites)	9	0.5	88%	47%
Single Mission (going to unfamiliar sites in same category, e.g., sports to real estate to find what they need)	10	1.5	11%	85%
Do It Again (going to favorite sites, e.g., banking and chat rooms)	14	2	95%	87%
Loitering (leisurely visiting favorite sites, news, gaming)	33	2	90%	87%
Information Please (researching all aspects of a topic from many sites)	37	1	14%	41%
Surfing (exploring attention-grabbing sites such as news and shopping)	70	1	14%	26%

The researchers discovered that purchasing behavior was highest in Loitering, Information Please, and Surfing sessions. These types of sessions were the longest: Customers spent more time on each Web page, and thus were more receptive to advertising.

\*Site Familiarity: the percent of time in a session that users spent at familiar sites they had visited four or more times.

\*\*Category Concentration: the percent of time in a session users spent at sites that belong to the most frequented category. For example, if a user spent 10 minutes in a session, 5 minutes in sports, 3 minutes in news, and 2 minutes in entertainment, then the user had a 50% concentration in that session.

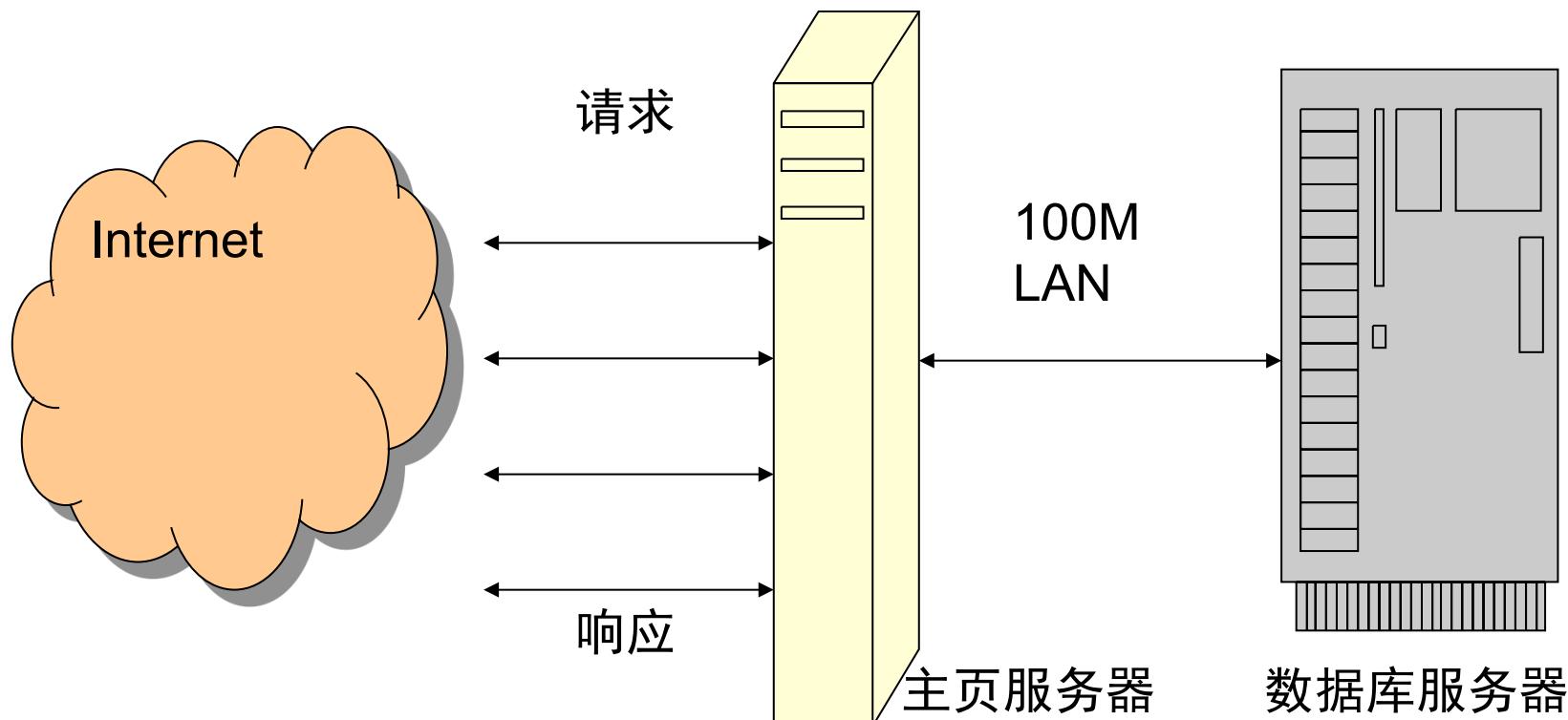
SOURCE: Rozanski, Bollman, and Lipman, 2001.

## **2. CUSTOMER BEHAVIOR MODEL**

- 概述
- 客户行为模型图
- 站点测量指标

# 一个简单电子商务服务器的性能分析

在线玩具商店：



# 在线玩具商店基本情况

- 商店提供了多个商务功能，诸如放入购物车，检索，浏览，欢迎，注册和付款。
- 为了简单起见，让我们假设客户只有一个导航模式(Navigate model)：他们访问主页，接着检索目录来查找他们喜爱的玩具 .

# 销售活动

- 业务管理层想增加客户的数量，决定在一些流行的面向家庭的杂志上开始一个广告活动。
- 为了给技术管理层提供系统利用率的信息，WEB站点的管理者在高峰时段内监测站点，并且提供了测量数据。

数据类型	值
已经结束的客户会话数目	35,000
网络服务器总的忙碌时间	1,200秒
数据库服务器总的忙碌时间	2,100秒
测量时段	3,600秒

# 技术人员关心：在高峰时段内的访问者的增长

- 如果在高峰时段内客户会话量增长了一倍的话，这个在线站点的性能会怎样？
- 如果使用分页功能，那么目录页的大文件利用率会怎样？
- 如果客户改变了他们的导航模式并且目录检索增加了50%的话，对性能有什么影响？

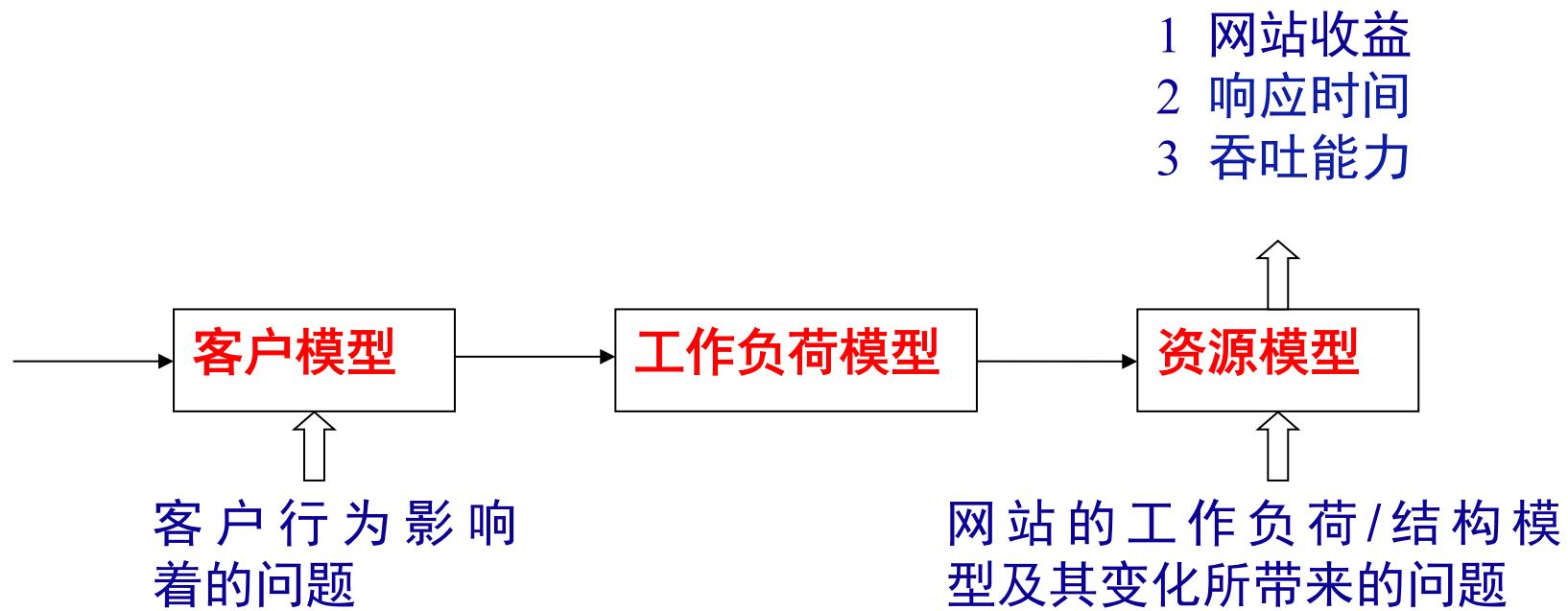


客户行为  
模型

# 客户行为模型

- **会话session**: 一个电子商务站点的客户通过单次访问期间发出的一系列连贯的和相关的请求与站点进行交互。
- 不同类型的客户在一个会话期间有不同的遍历电子商务网站的模式，所以就有不同的方式和频率来调用电子商务系统的不同功能。
- **客户行为模型**: 模拟不同类型客户在会话中的行为

# 客户模型、工作负荷模型以及资源模型之间的联系



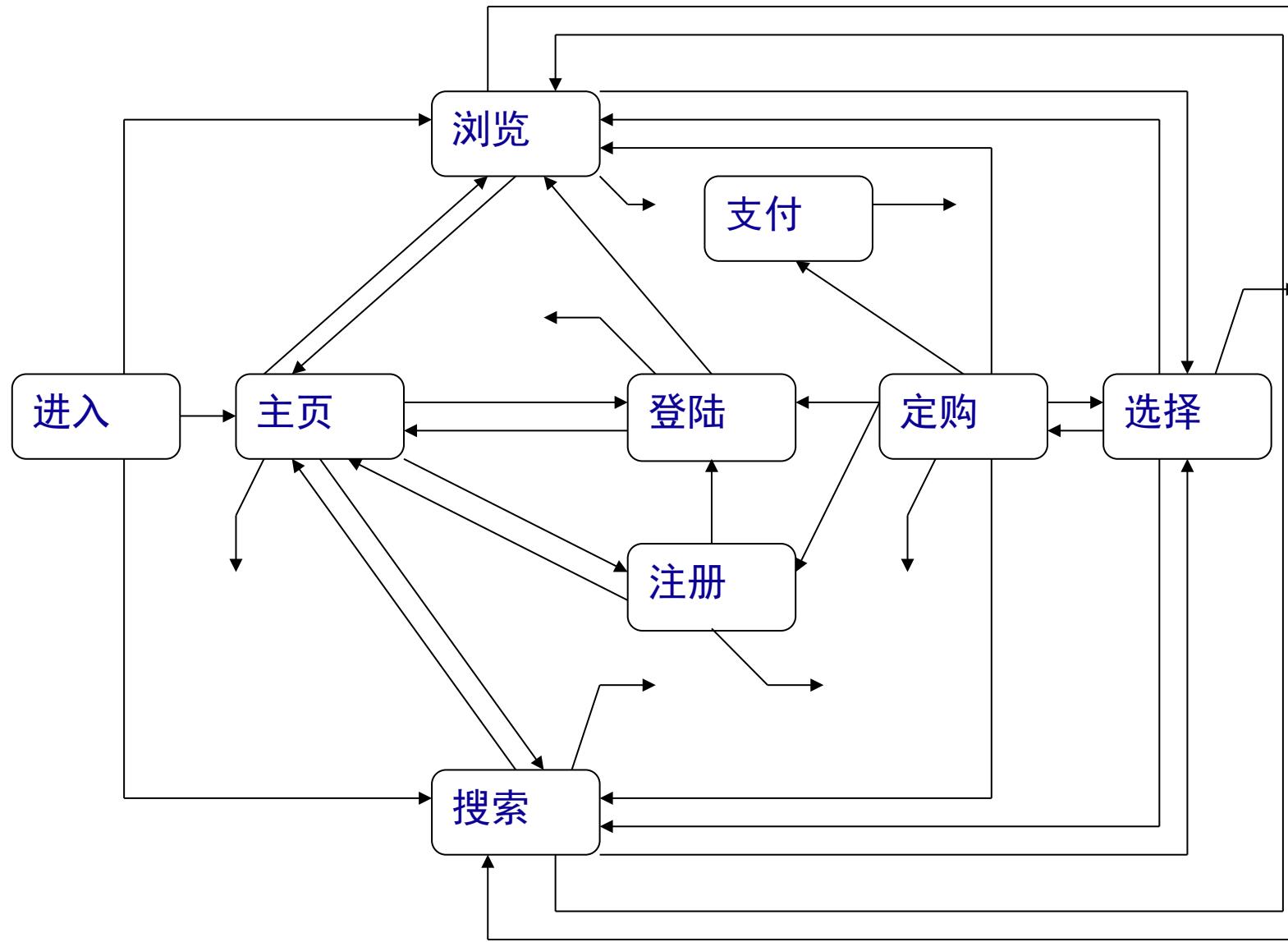
# 客户行为模型图(CBMG)

- 静态、动态
- 建立客户行为模型图
  - 确定电子商务站点向客户提供的功能集合
  - 根据资源消耗来精简功能集合
  - 确定状态间的转换

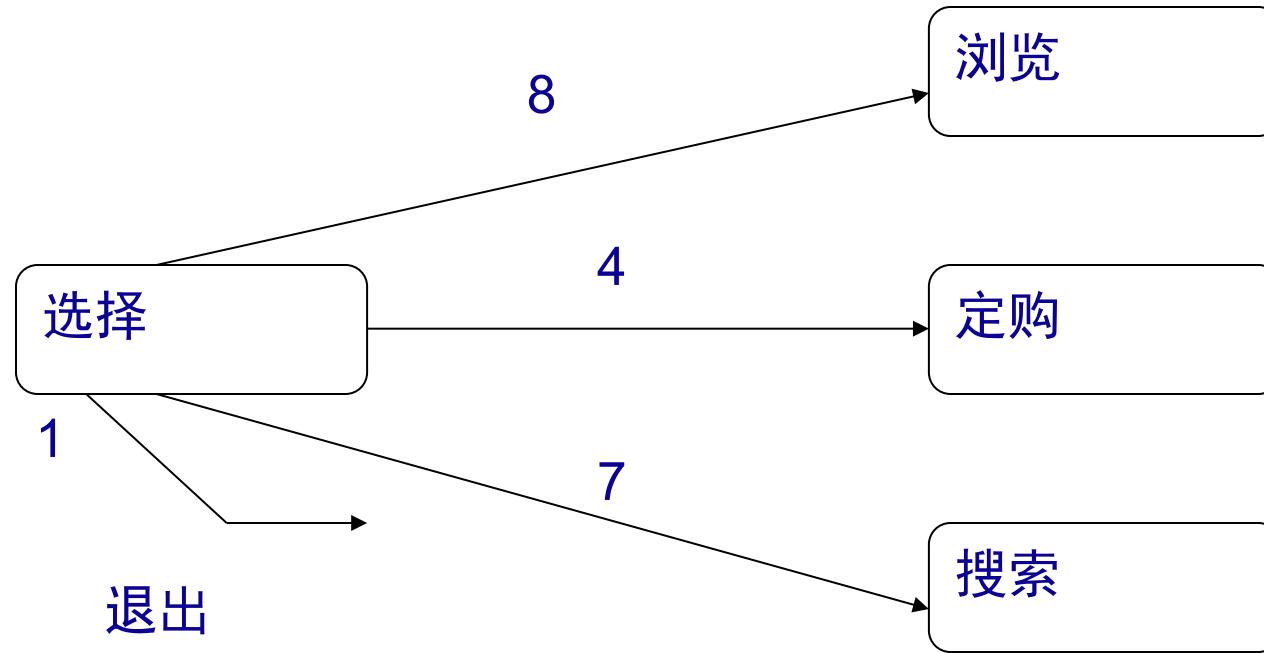
# 确定电子商务站点向客户提供的功能集合——网上商店

功能模型类别	功能	功能描述
常规	登录 注册 搜索 选择 浏览	登录站点 新用户注册 搜索站点商品数据库 查看搜索结果的具体信息 由连接浏览站点
零售	订购商品 取消订购 查看订购商品 注册新用户 订购新用户的商品 检查状态 付款	将商品放入购物车 将商品移出购物车 查看购物车中商品信息 创建一个新注册 将商品放入用户的购物车 检查前一状态 购买购物车中的商品
商贸	开户 获得报价 获得公司运行报告 检查成交价格 查看货物信息 进行交易 资产规划	为商贸交易双方开通帐号 获得实时或者曾经的报价 获得交易公司的运行状态 查看完成交易的价格信息 查看交易商品的信息 双方进行资金和物品交易 资产完成重新配置
信息	下载信息 订阅信息 在线收听 在线收视	下载有关的各种信息 订阅有关商贸信息 在线收听有关音频节目 在线观看有关视频节目

# 实例：网上购物

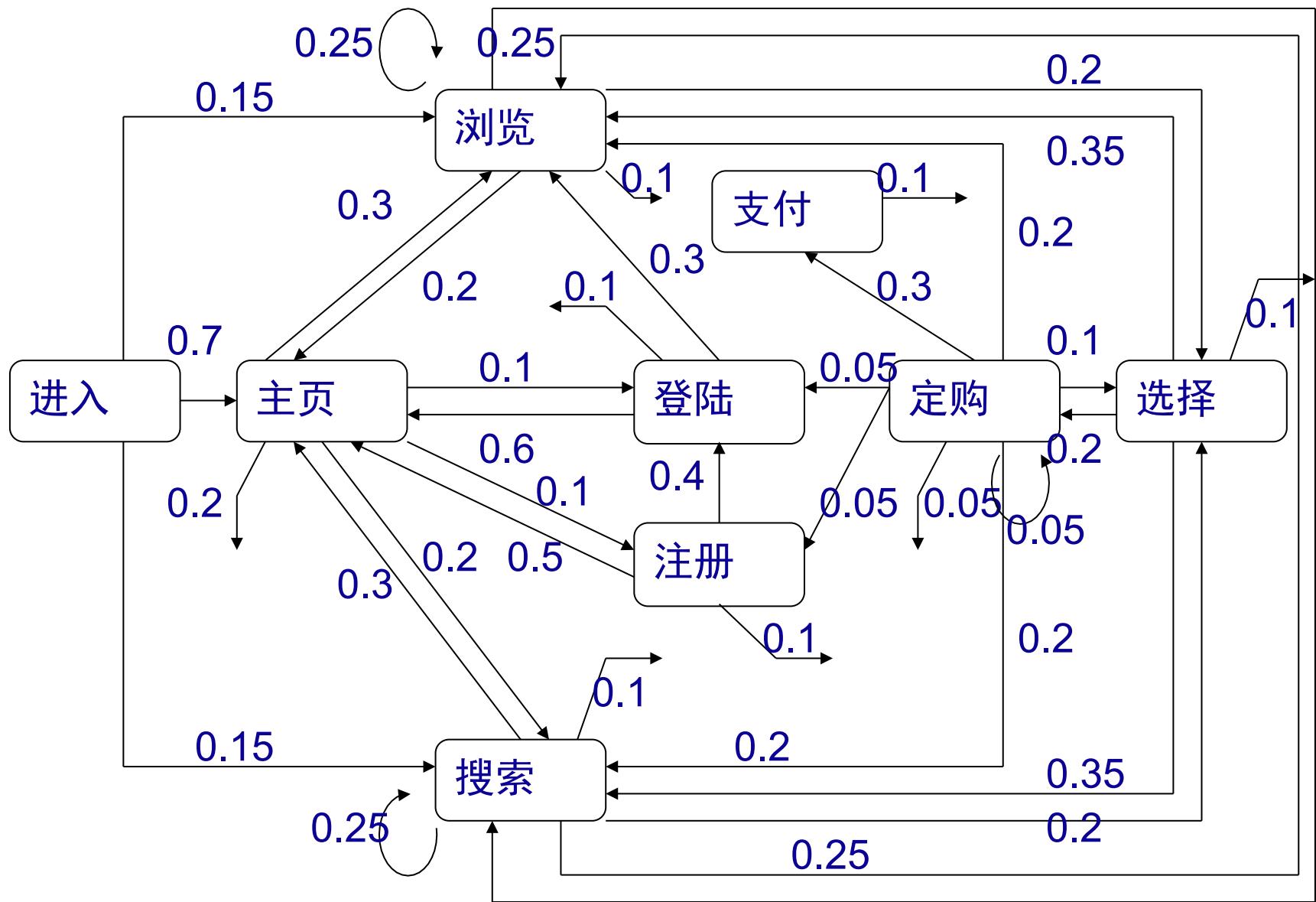


# 确定状态间的转换



# 状态转换矩阵 $P = [P_{i,j}]$

# 完整的CBMG



# 网络广告站点的度量指标

- **点击率**: 测量一个站点每秒钟内请求服务对象的次数。
  - 不过点击率往往不能很真实的反应用户的请求数。一个页面通常由一个HTML文件和几个其他的图像文件的链接组成，用户看一次页面往往web收到的请求数是好几个。网页里面的图片链接数量越多，点击率就越不准确。
- **日平均页面访问数**: 反映了每天页面被调用的总次数。
  - 这比点击率更加准确。这个数值越高，对于在该网页上发布付费广告的公司来说就意味着它的广告受众面越广。
- **点击进入百分率(Click-throughs)**: 测量观看在线广告并点击进入广告页面的用户的百分率。这个度量标准更多的定位于估计在线广告的影响。

# 电子商务站点度量指标

- **单一的客户数**：指出在一个特定期间内有多少不同的客户访问站点。
- **站点收益率 (Revenue Throughput)**：是用于测量一个电子商务站点每秒钟销售获得的货币数量的面向商业的度量标准。
- **潜在损失(Potential Loss)**：是测量由于低劣的性能或其它原因使得客户离开站点，使得客户购物车中的货币没有转化为销售额的另一个面向商业的度量标准。

# 从客户行为模型图获得度量标准

希望客户行为模型图能够解决哪些问题？

- 每次访问电子商务站点期间各个电子商务功能（如搜索、浏览、选择、添加）被调用的平均次数是多少？
- 平均起来，客户每次访问电子商务站点时多长时间购物一次？
- 在访问电子商务站点期间，电子商务功能执行客户请求的平均次数是多少？

对于CBMG图中的状态  $i$ , 假设其平均访问次数为  $V_i$  , 那么有

$$V_i = \sum_{k=1}^n V_k \times P_{k,i}$$

其中  $P_{k,i}$  表示一个客户从状态  $k$  转换到状态  $i$  的概率。

# 例1

考虑前面客户行为模型图。  
客户行为模型图中每个状态的平均访问次数是多少？

访问状态	平均访问次数
进入	1.000
主页	1.862
浏览	2.303
搜索	2.193
登录	0.274
付款	0.058
注册	0.196
订购	0.193
选择	0.919
离开	1.000

## 例2

- 电子商务站点的一个重要度量标准是**购买与访问的比例（BV）**，它被定义为每次访问站点期间销售事务的平均数。
- 在该客户行为模型图中，购买与访问的比率可以从例1中得到：**0.058**

## 例3

如果转换概率作如下变动：

$P_{\text{select, Browse}} = 0.4$ ,  $P_{\text{Select, AddtoCart}} = 0.1$ ,

$P_{\text{Select, Search}} = 0.4$ , 新的购买访问比率是多少？

新的购买访问比率值是支付状态的平均访问次数，为0.030。也就是说，访问站点的客户中只有3.0%购买而不是前一个例子中的5.8%。

### 3. 客户访问模型

- 客户访问模型是一组向量，它说明了电子商务站点提供的功能被执行的次数
- 简化了CBMG

# 客户访问模型表

访问状态	会话1	会话2	会话3
主页	1	2	3
浏览	4	8	4
搜索	5	5	3
登录	0	1	1
付款	0	0	1
注册	0	0	1
订购	0	2	1
选择	3	3	2

# CVM图说明

上表展示了使用各个状态访问次数描述的3个会话例子。注意，在客户访问模型（CVM）中没有表示进入和退出状态，因为这两个状态的访问次数总是为1。

- 会话1表示了一个浏览站点的客户，他只做少量的搜索，但并不登录和购买任何东西。
- 会话2中客户登录但并不需要注册，因为他已经是一个注册用户。这个客户在支付前放弃了这个站点，虽然他已经把两个项目放入了购物车。
- 会话3表示了一个在这个站点注册的新客户，他把一个项目放入购物车，并支付。

## 例 4

考虑表中的3个会话。假设每个搜索请求平均执行3个包含项目目录索引的磁盘输入/输出，每个选择功能平均执行2个包含产品描述的磁盘输入/输出。这3个会话在两个磁盘上平均各产生多少次输入/输出？

- 这三个会话总共执行 $5+5+3=13$ 次搜索和 $3+3+2=8$ 次选择，因此，对索引磁盘的输入/输出平均次数是 $13\times 3=39$ ，对目录磁盘的平均输入/输出次数是 $8\times 2=16$ 。
- CVM可被用于获取用于资源模型的工作负荷信息。

# 4. 客户端/服务器端交互图(CSID)

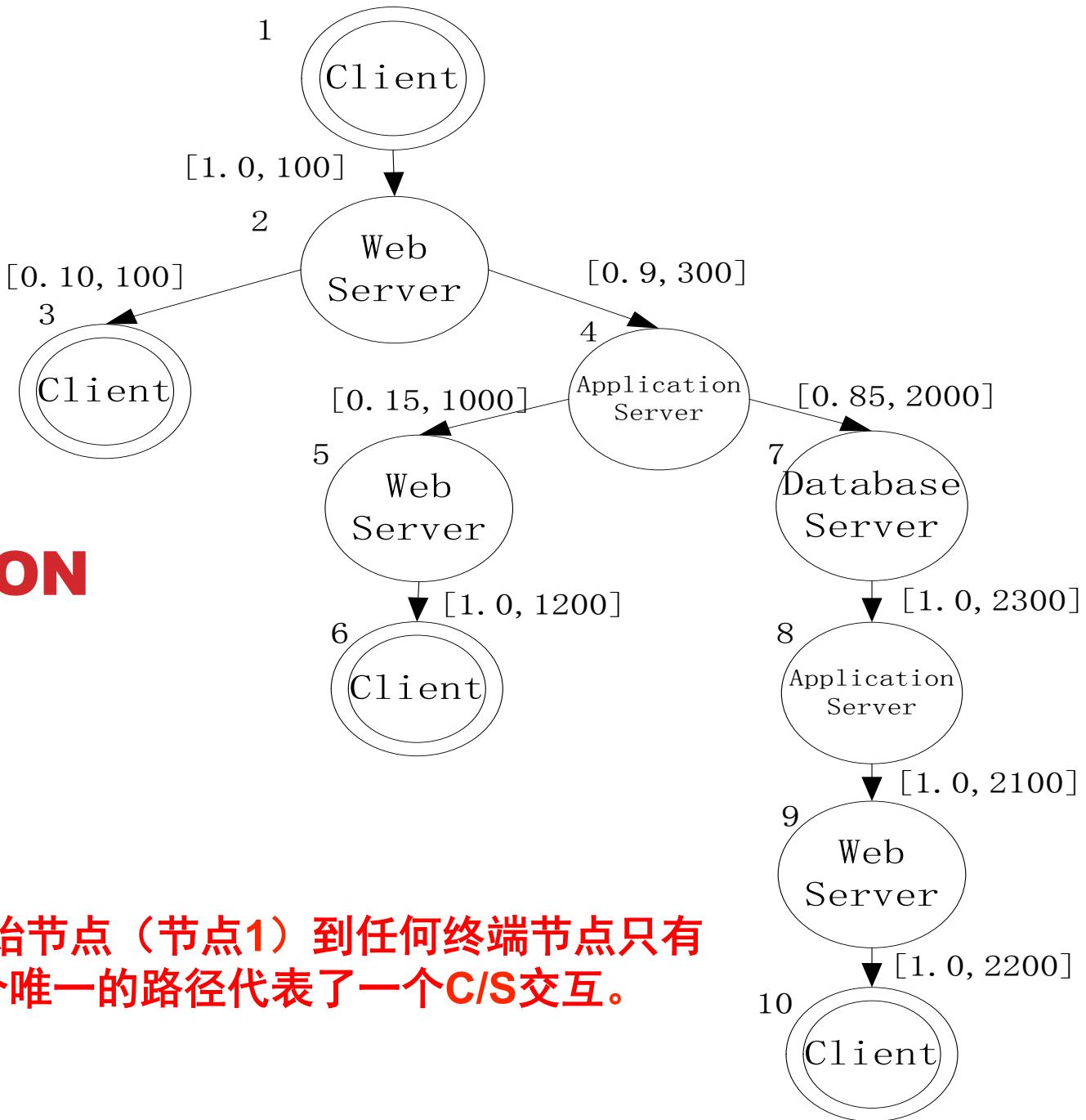
通过CSID中的节点表示的客户和服务器的各种资源的消耗，在对同一个客户端或服务器端的不同访问中可能是不同的。

- 一个CSID包括节点（矩形和圆圈）和连接这些节点的箭头。CSID的节点表示在执行电子商务功能期间对客户端和/或服务器端的访问。
- 圆形节点通常代表执行电子商务功能期间涉及的服务器，圆圈节点内部的标注（如WS、AS、DB）指出了服务器的名称。
- CSID中的弧线用

[p, m]

类型的对来标注，其中p是消息发送的概率，m是消息的字节大小。

# CLIENT SERVER INTERACTION DIAGRAM (CSID)

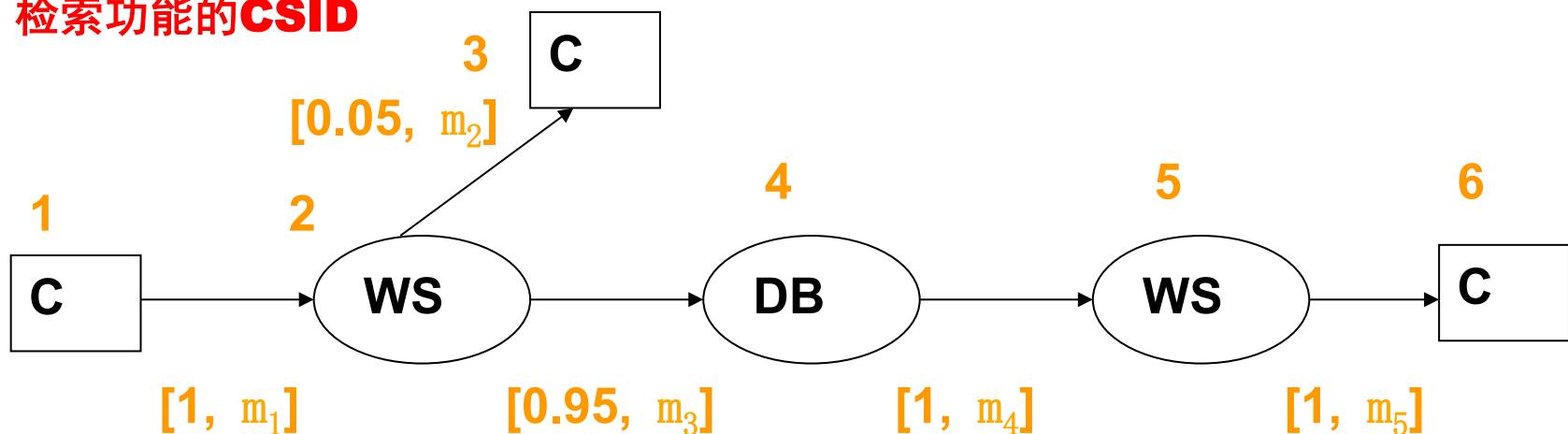


从CSID中的起始节点（节点1）到任何终端节点只有一条途径。这个唯一的路径代表了一个C/S交互。

# 在线玩具商店 (1)



检索功能的CSID



# 根据CSID分析

- 现在我们能够知道在执行一个电子商务功能的时候，是如何使用Web服务器和数据库服务器的：只有当沿着 $1 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 6$ 事务序列时，才使用数据库服务器。
- 网络服务器在这个序列中出现两次，
- 网络服务器在序列 $1 \rightarrow 2 \rightarrow 3$ 中出现一次，这个序列出现的概率是0.5。
- 所以网络服务器处于运行状态的平均次数是 $2 \times 0.95 + 1 \times 0.05 = 1.95$ ，而数据库服务器处于运行状态的平均次数是 $1 \times 0.95 = 0.95$ 。

# 分析结果

数据类型	值
已经结束的客户会话数目	35000
网络服务器总的忙碌时间	1200
数据库服务器总的忙碌时间	2100
测量时段	3600
每个客户会话的检索次数	2.5
每次检索时访问网络服务器的次数	1.95
每次检索时访问数据库服务器的次数	0.95

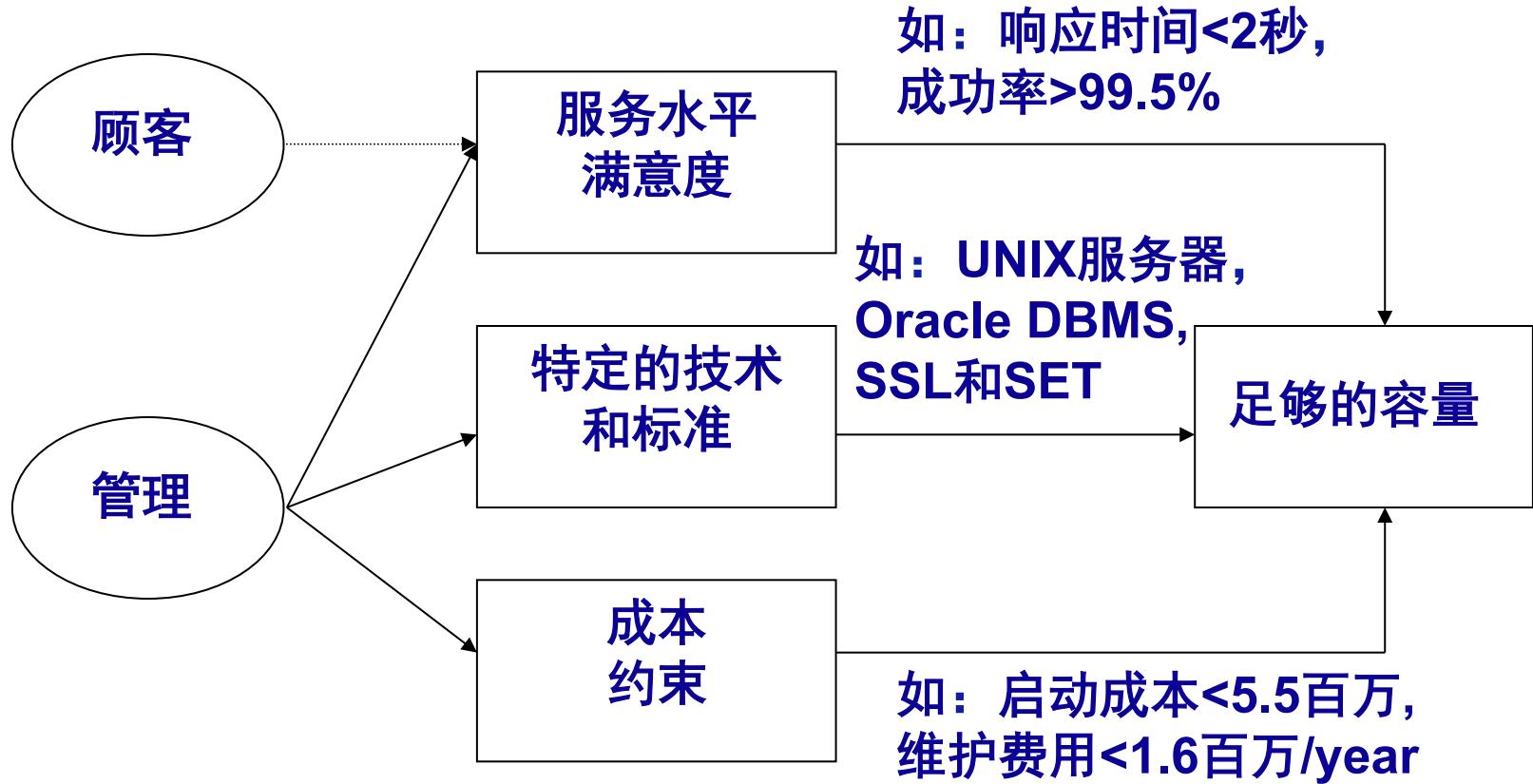
# 进一步参考

Menasce D A, Almeida V A F. Scaling for E-Business: technologies, models, performance, and capacity planning.  
Prentice Hall, 2000

### 3. 电子商务系统容量规划

- 预测未来负载水平何时会使系统饱和以及确定一个尽可能延迟系统饱和的**最经济方法**的过程
- 未来负载水平
  - 现有工作负载的发展
  - 新的应用和服务的采用
  - 客户行为的变化

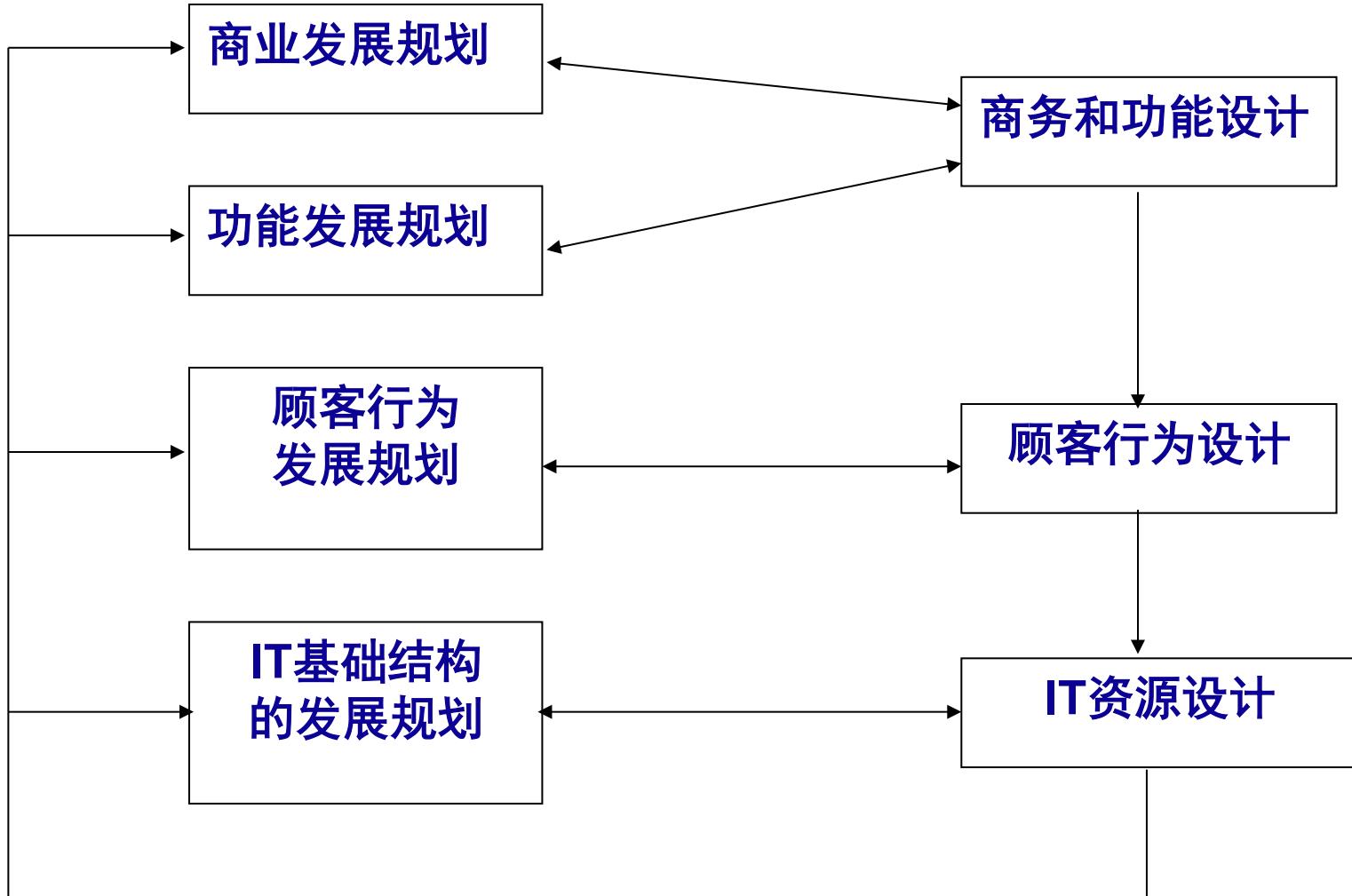
# 电子商务系统容量规划



# 足够容量的目标

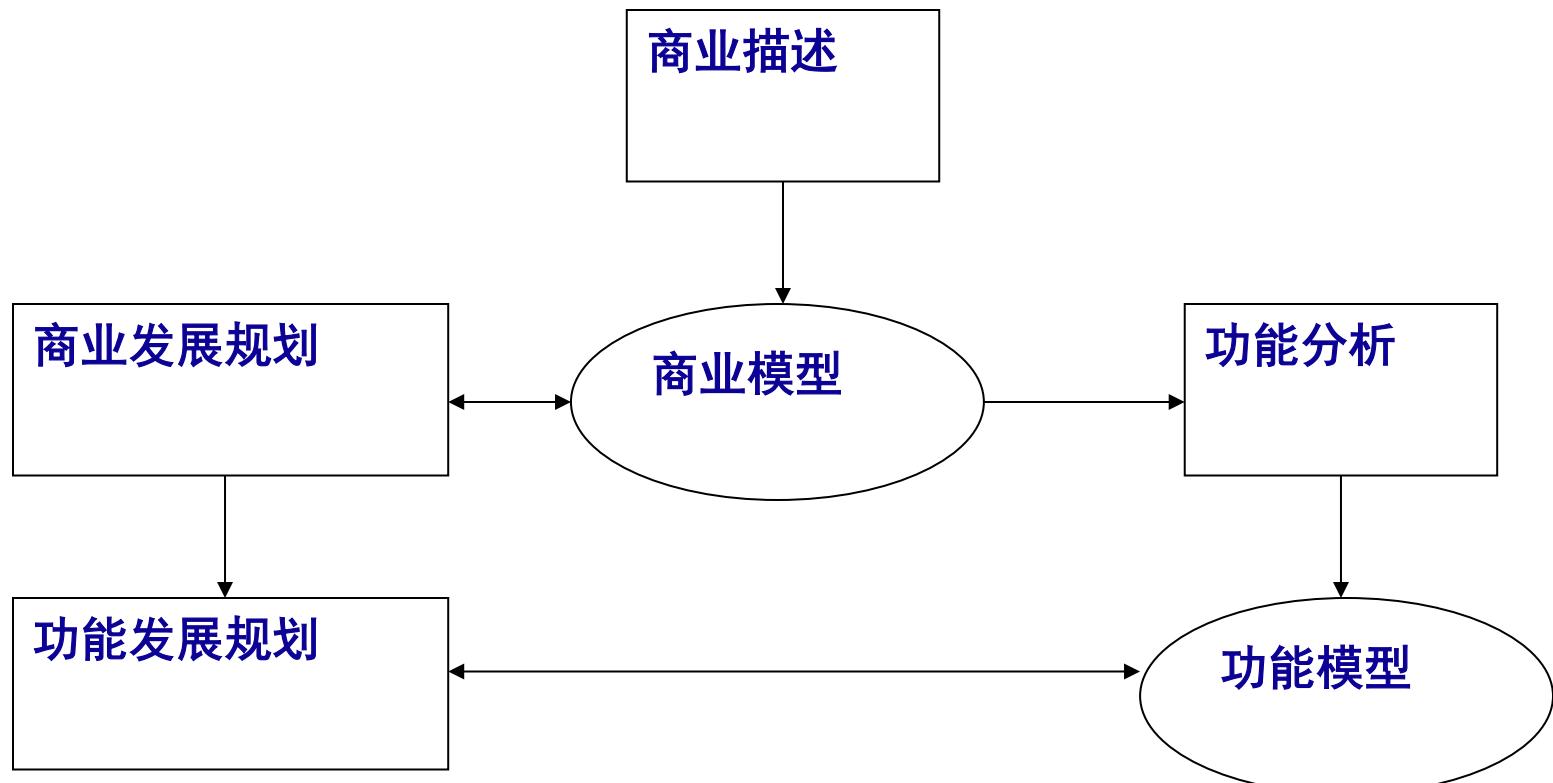
- 降低停机时间和网络瓶颈现象
- 可用性、可扩展性、快速且高效
- .....

# 容量规划方法论



# 商务层

- 商务和功能设计包括：描绘商务的特征和分析电子商务站点提供的功能

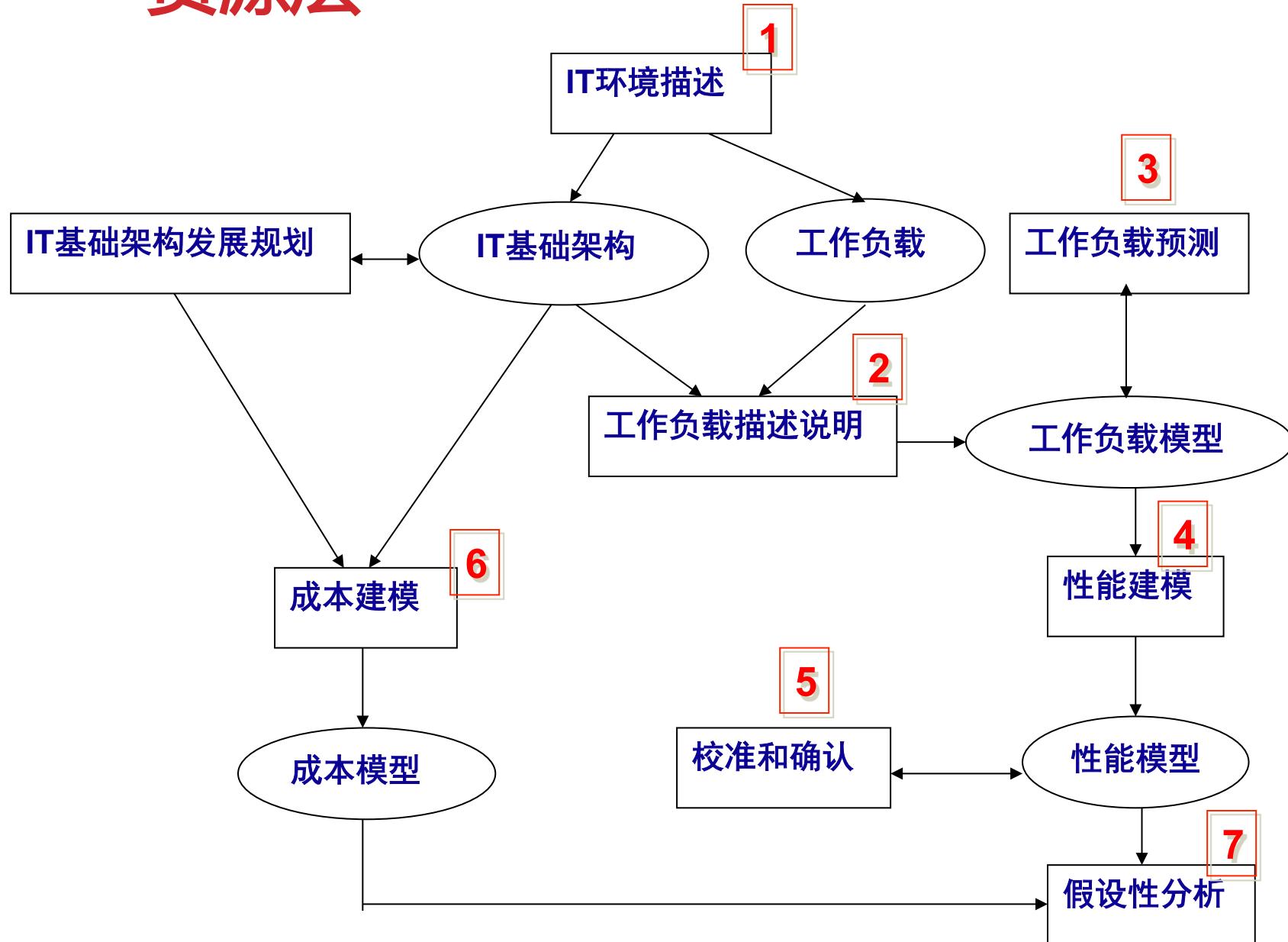


# 客户行为层

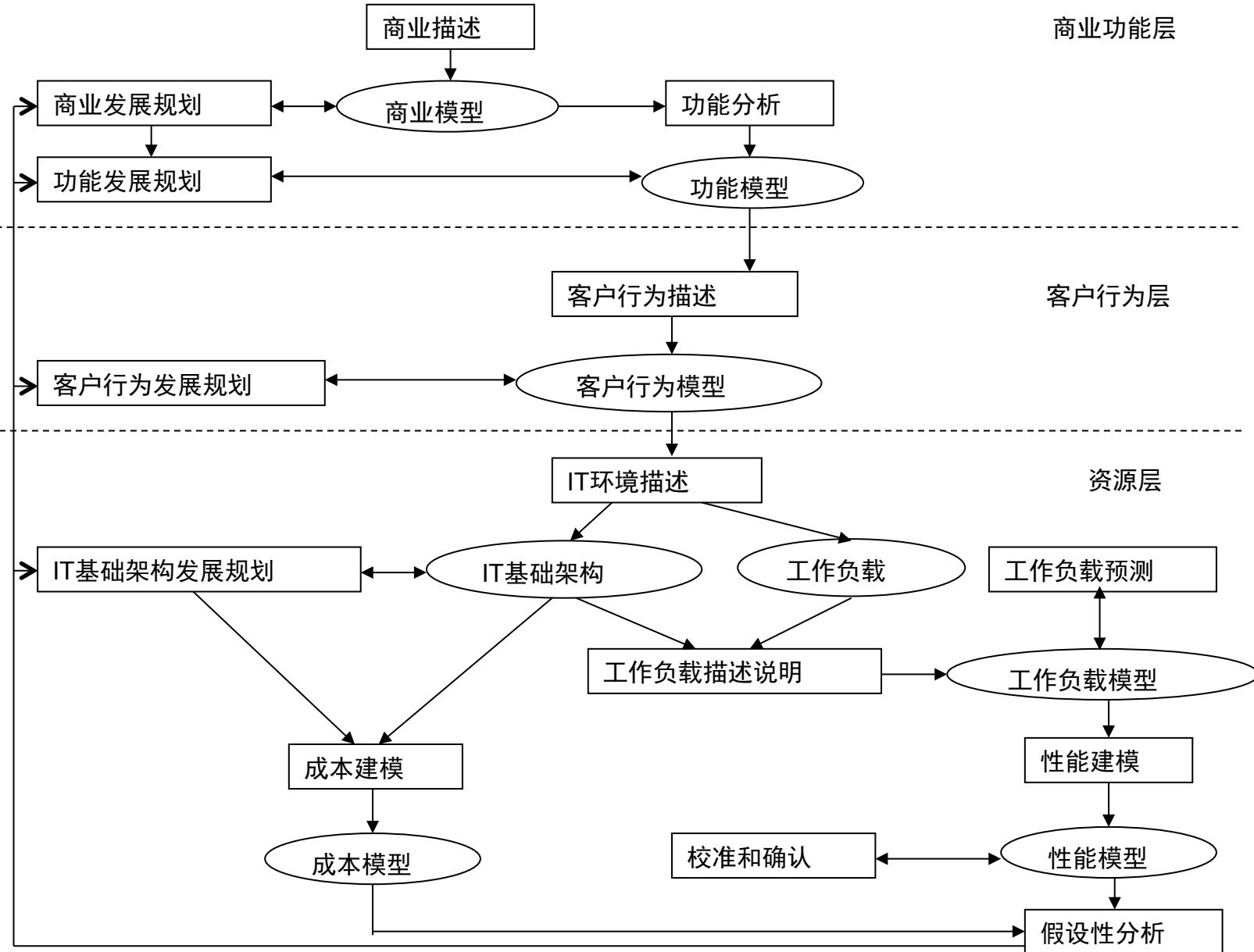
- 这个层次的主要步骤是**客户行为描述**和**客户行为发展规划**。
  - **客户行为描述**生成一个**CBMG**或者**CVM**形式的客户行为模型，如果只有静态的客户行为模型也已足够。
  - **客户行为发展规划**由对站点布局规划的有计划的调整产生，布局规划的调整可能会改变可行的导航模式。
  - 每一次客户行为的变化，都应该有一个修改过的客户行为模型与之对应。



# 资源层



# 电子商务系统容量规划



# 容量规划操作步骤

1. 了解环境
2. 对工作负载进行分类
3. 确定影响最大的组件
4. 选择要应用的度量技术
5. 应用度量技术
6. 重新评估

# **RIGHT-SIZING A WEB SITE? USE A SIMULATOR**

- Web site simulators such as IBM's High Volume Web Site (HVWS) Simulator can help answer such questions as**
  - How many **Web servers** does your site require
  - How many **CPUs** should each server have
  - How powerful does the site's **database server** need to be
  - What kind of **connection speed** do you need to the Internet
- The HVWS Simulator uses a queuing model that estimates the performance and capacity of a Web site based on workload patterns, performance objectives and specific hardware and software**

# RIGHT-SIZING A WEB SITE? USE A SIMULATOR

High Volume Web Sites Simulator Version 4.0.3 for WebSphere - project1

File Window Help

IBM WebSphere

## High Volume Web Sites Simulator

Business Pattern | Objectives | Hardware Topology | Software

Online Activities

Business Pattern

- Online Shopping
- Online Trading(Trivial)
- Online Banking
- Reservation system
- Inventory management
- Online Brokerage(Complex)
- Online Auction
- WP 2.1
- WP 4.1
- Business to Business
- PERVASIVE - WP 4.1**
- USER DEFINED**



Interactions per User Session

	Percentage(%)	Page View	Think time(min)
Browse Only	83	4	0.5
Product Search	15	2	0.5
Buy	2	12	0.5

Burst to Peak Ratio

Total page views per user visit(session) **3.86**

Total think time per user visit(session) **1.93 minutes**

Pattern Diagram Definition

Calculate Results Graph Results Reset Help

For Help, press F1

# **USER DEFINED SCENARIO**

**Definition for Scenario 1**

Measured Scenario Information

Use SSL      % Transaction using SSL:

Transactions are:  Session Persistent  Non-Persistent      Session Size:

WebSphere Version:

Name of the Scenario:

Description:

Measured Data per Page View

Measured Configuration:  Single Tier  Two Tiers  Three Tiers  Disable DB

 CPU Serv. Time

Measured Server Brand:	<input type="text" value="pSeries(RS/6000)"/>	Web/App Server	Backend Server
Measured Server Model:	<input type="text" value="44P-270 2-way 450"/>	<input type="text" value="44P-270 2-way 450"/>	
Measured Service Time(sec.)/page view:	<input type="text" value="0.1"/>	<input type="text" value="0.1"/>	
Measured Disk IO/page view:	<input type="text" value="0.1"/>	<input type="text" value="0.1"/>	
Nodes per Tier/LPAR per System:	<input type="text" value="1"/>	<input type="text" value="1"/>	

**OK**

# **SETTING OBJECTIVES FOR PERFORMANCE SIMULATOR**

High Volume Web Sites Simulator Version 4.0 for WebSphere - project1

File Window Help

IBM WebSphere High Volume Web Sites Simulator

Business Pattern Objectives Hardware Topology Software

Performance Target:

Specific arrival rate:  User visits per second

Page view rate:  Page Views per Second

Response time per page view:  seconds

Average user session time:  seconds

Number of concurrent users:

Processor utilization:  %

Percent Contingency Factor:  %

Response Time Calculations:

Average response time  
 90th Percentile response time  
 95th Percentile response time

Configuration Estimator:

Single Tier  Two Tiers  Three Tiers

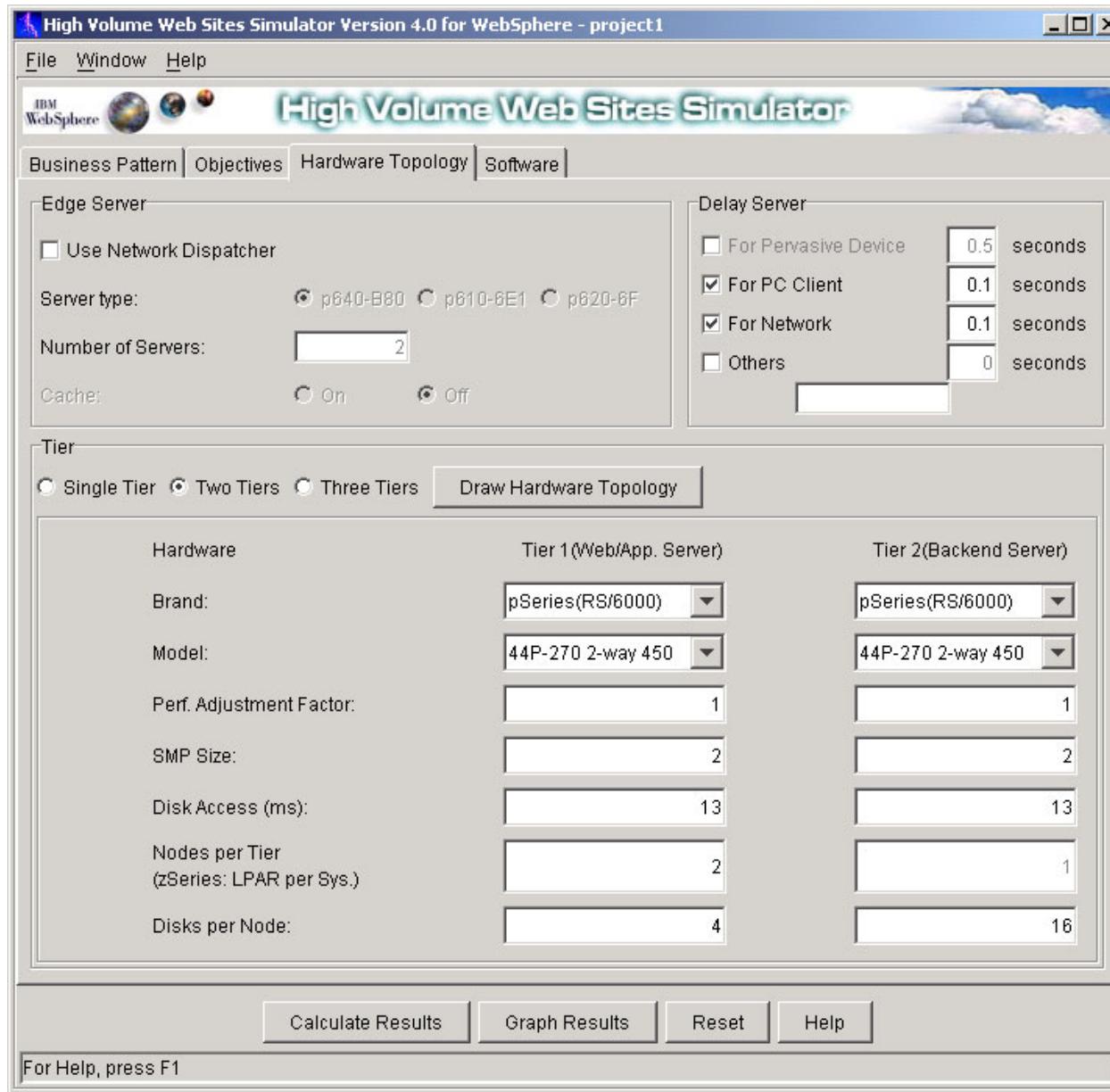
Tier 1 Server Brand:

Tier 2 Server Brand:

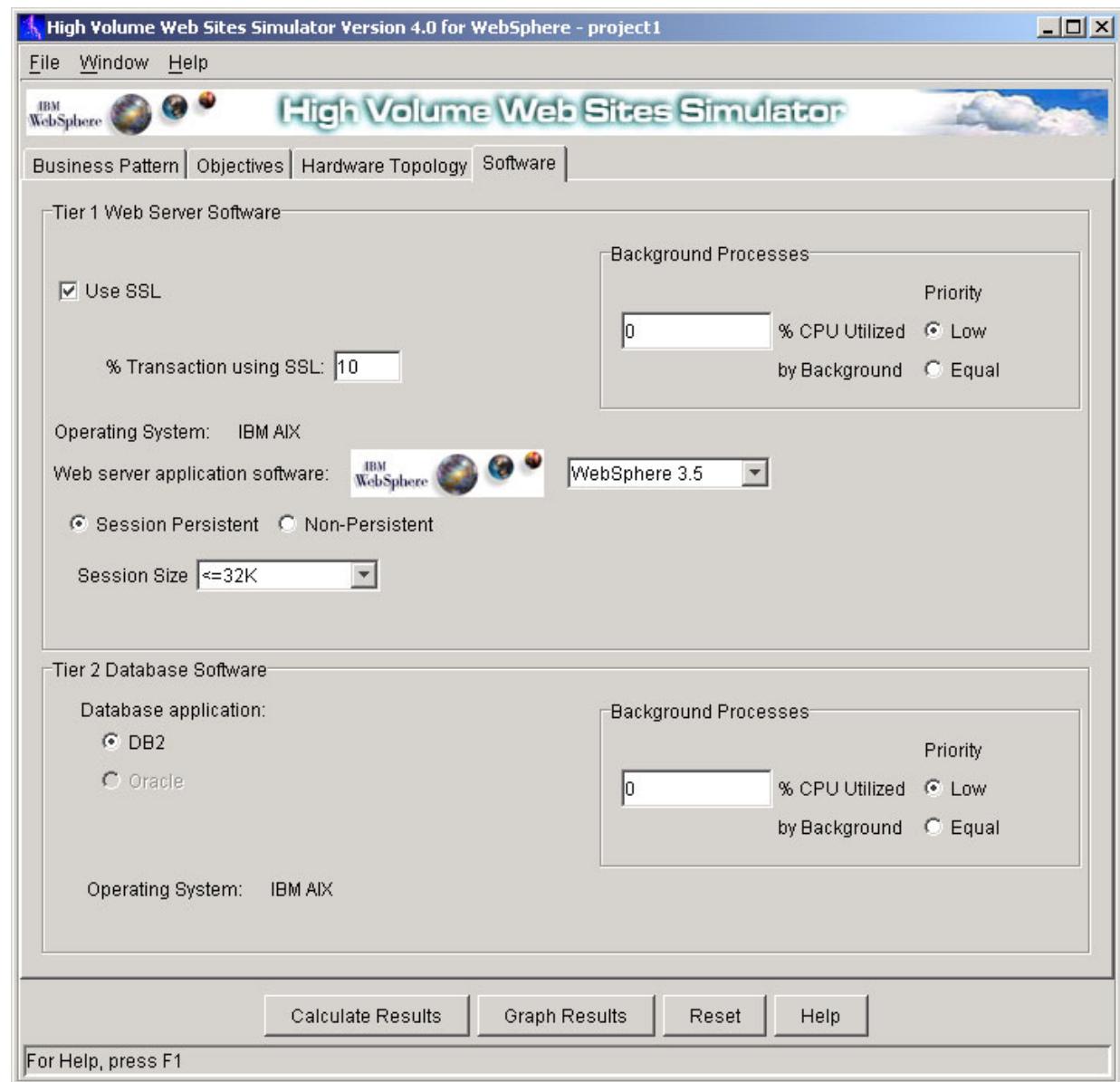
Calculate Results Graph Results Reset Help

For Help, press F1

# HARDWARE TOPOLOGY SPECIFICATION



# SPECIFY THE SOFTWARE COMPONENTS USED OR PROJECTED TO BE USED



# RESULTS

**Calculated Results**

Over All | Min Response Time | Utilization | Memory

	Base Plus Contingency	Base
Arrival Rate(user visits per second)	7.00	7.00
Response Time per page view(sec)	0.306	0.273
User Session Time	117	117
Concurrent Users	819	818
Page Views per Second	27.0	27.0

**Calculated Results**

Over All | Min Response Time | Utilization | Memory

	Base Plus Contingency	Base
Arrival Rate(user visits per second)	0.100	0.100
Response Time per page view(sec)	0.279	0.261
User Session Time	117	117
Concurrent Users	11.7	11.7
Page Views per Second	0.386	0.386

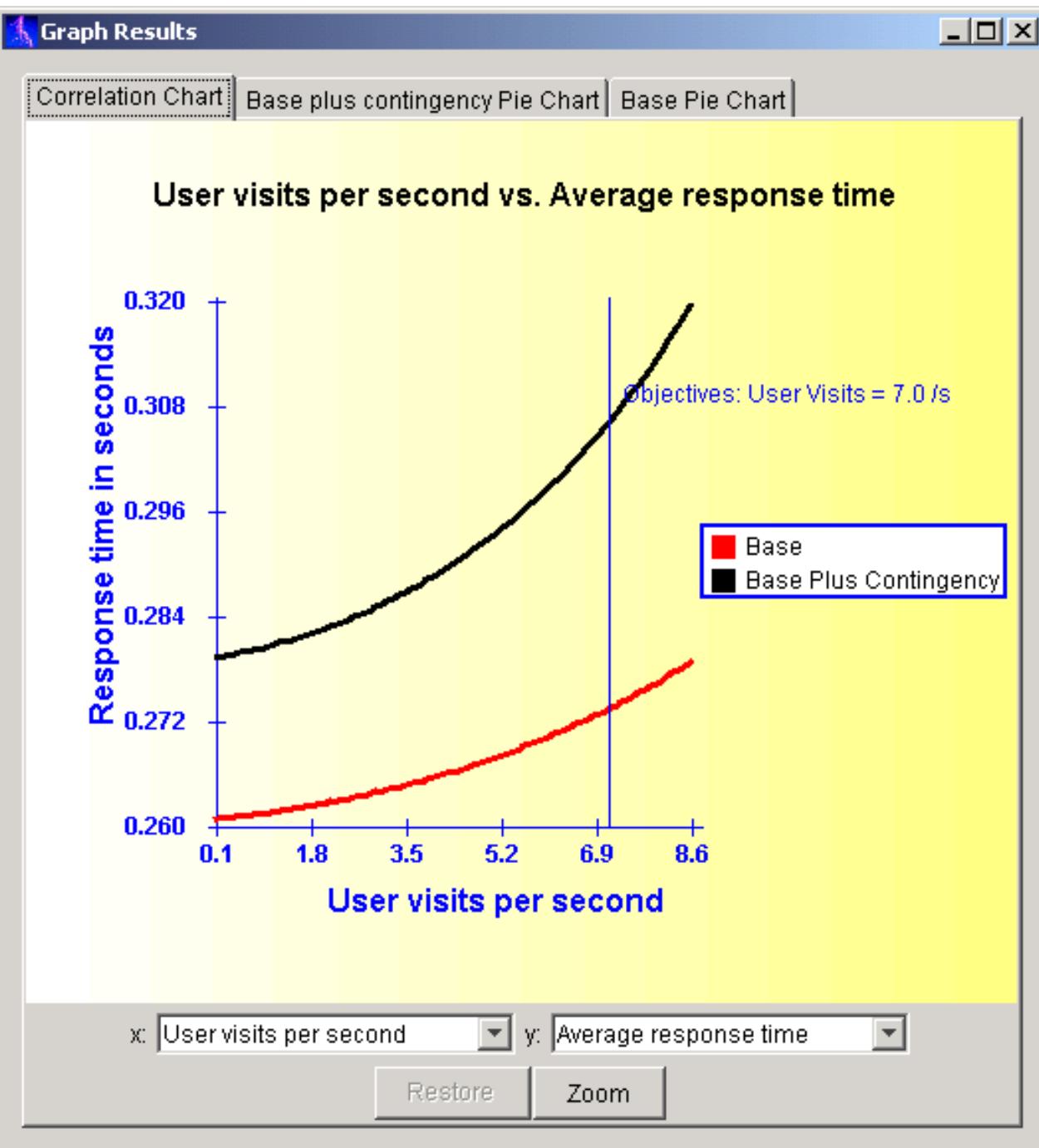
# RESULTS

Processor utilization		
	Base Plus Contingency	Base
Edge Server	0.00%	0.00%
Tier 1	43.59%	33.53%
Tier 2	10.48%	8.06%
Tier 3	0.00%	0.00%

Disk Utilization:		
	Base Plus Contingency	Base
Tier 1	0.75%	0.58%
Tier 2	0.80%	0.61%
Tier 3	0.00%	0.00%

Calculated Results	
Over All	Min Response Time
Utilization	Memory
WebSphere Nodes Memory Requirement (MB)	Recommended Value 1020



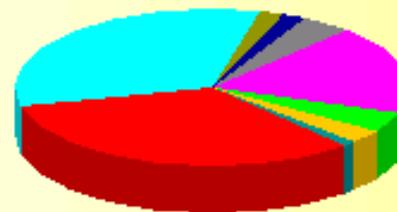
Correlation Chart

Base plus contingency Pie Chart

Base Pie Chart

## Components of performance

Base plus contingency calculations:



Web Presentation Tier CPU service	3%
Web Presentation Tier CPU wait	4%
Web Presentation Tier disk service	0%
Web Presentation Tier disk wait	0%
Web App Tier CPU service	18%
Web App Tier CPU wait	4%
Web App Tier disk service	0%
Web App Tier disk wait	0%
DB Tier CPU service	2%
DB Tier CPU wait	0%
DB Tier disk service	2%
DB Tier disk wait	0%
Client Delay	32%
Network Delay	32%
Edge Server Service	1%
Edge Server wait	0%
Other Delay	0%

# 6. MINING OF WEB SITE LOGS

□ 网站事务日志  
建立在网站服  
务器软件上，  
记录用户在网  
站上的活动

FIGURE 7.11

FOUR SECONDS FROM THE WEB TRANSACTION LOG OF AZIMUTH-INTERACTIVE.COM

```
64.88.16.67 - - [13/May/2003:12:51:53 -0400] "GET /images/ebook.gif HTTP/1.1" 304 1313
"http://www.azimuth-interactive.com/landingpage_access.php?item=50018&source=
overturemsa&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; Q312461)"

64.88.16.67 - - [13/May/2003:12:51:54 -0400] "GET /images/bookmarkusart.gif HTTP/1.1" 200
363 "http://www.azimuth-interactive.com/landingpage_access.php?item=50018&source=
overturemsa&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; Q312461)"

64.88.16.67 - - [13/May/2003:12:51:54 -0400] "GET /images/smallazimuthlogo.gif HTTP/1.1" 200
2695 "http://www.azimuth-interactive.com/landingpage_access.php?item=50018&source=
overturemsa&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; Q312461)"

64.88.16.67 - - [13/May/2003:12:51:54 -0400] "GET /images/trustseal_small.gif HTTP/1.1" 200
937 "http://www.azimuth-interactive.com/landingpage_access.php?item=50018&source=
overturemsa&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; Q312461)"

64.88.16.67 - - [13/May/2003:12:51:54 -0400] "GET /images/OrderByPhone.gif HTTP/1.1" 200
2017 "http://www.azimuth-interactive.com/landingpage_access.php?item=50018&source=
overturemsa&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; Q312461)"

64.88.16.67 - - [13/May/2003:12:51:54 -0400] "GET /images/credit.jpg HTTP/1.1" 200 12177
"http://www.azimuth-interactive.com/landingpage_access.php?item=50018&source=
overturemsa&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; Q312461)"

ddsl-216-68-138-118.fuse.net - - [13/May/2003:12:51:56 -0400] "GET /onlinetesting.html
HTTP/1.1" 200 1603 "http://www.azimuth-interactive.com/masthead.cgi?url=viewproduct.
php&item=60018&source=googlest&" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0; .NET
CLR 1.0.3705)"

ddsl-216-68-138-118.fuse.net - - [13/May/2003:12:51:56 -0400] "GET /masthead.cgi?page=
onlinetesting HTTP/1.1" 200 3621 "http://www.azimuth-interactive.com/onlinetesting.html"
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0; .NET CLR 1.0.3705)"

ddsl-216-68-138-118.fuse.net - - [13/May/2003:12:51:57 -0400] "GET /onlinetestingtext.php
HTTP/1.1" 200 48318 "http://www.azimuth-interactive.com/onlinetesting.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT
```

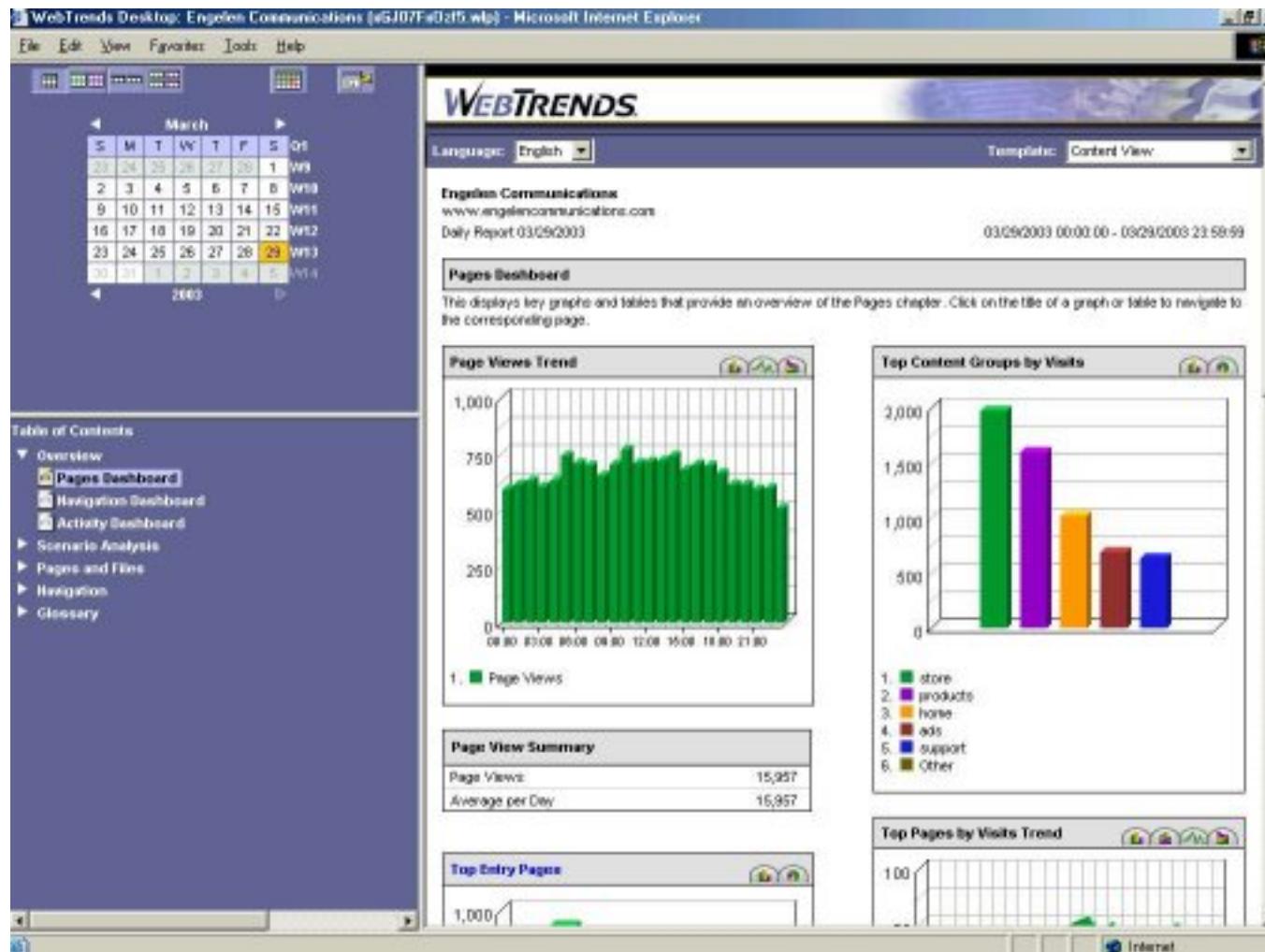
# 网站的事务日志数据对营销的作用

数据元素	在营销中的作用
访问者的IP地址: <b>64.212.128.3</b>	当用户使用固定网络而不是拨号 <b>Modem</b> 时，该地址可以用来发送关于营销信息的电子邮件。拨号 <b>Modem</b> 使用临时的 <b>IP地址</b> ，所以不能用来发送电子邮件
日期与时间戳: <b>[12/May/2014:15:01:18-0500]</b>	用来了解用户网上活动时间
所请求和访问的网页对象(“Get”语句): <b>GET/images/traininglibraryart.gif</b>	用来了解这个特定用户对网上什么内容感兴趣（点击流），以后可以用来发送“个性化”的信息，“定制化”的产品或者简单地用来发送相关产品信息的电子邮件
发送的网页的大小(信息的字节数): <b>1996</b>	用来了解对服务器和通信连接能力的要求
用户来自的网站名或网页名: <b>“http://www.azimuth-interactive.com/office97text.htm”</b>	用来了解用户是如何进入网站的，和进入网站以后的行为方式

# 网站的事务日志数据对营销的作用(续)

数据元素	在营销中的作用
用户使用的浏览器名和版本: <b>Mozilla/5.0 compatible; MSIE6.0</b>	有助于了解客户的浏览器，确保网站能够和客户所使用的浏览器兼容
用户所使用的操作系统名和版本: <b>Windows XP</b>	有助于了解客户机的性能，较新版本的操作系统意味着客户使用的是新机器，或者该客户熟悉技术
在访问网站的这一段时间内，用户浏览的所有网页和对象的历史信息	用来建立用户的个人资料，分析其在网站上的活动，了解哪些网页和资源最受欢迎

# WEBTRENDS



- WebTrends 是一款网站日志分析工具，在全球网站分析工具市场占有率排名第一。
- WebTrends 通过对网站数据的分析，可以预测网站的成长及未来发展需求，并通过这些数据通常用来判定网站成效或证明网站之存在价值。

# COOKIES

- **Cookies**: 是一个小文本文件，当浏览器每次访问网站时，网站服务器就会在浏览者的计算机硬盘上创建一个**cookie**文件，作为一个特别文件在浏览者访问网站时使用；
- **cookie**在WiKi上的解释是：**某些网站为了辨别用户身份而储存存在用户本地终端（Client Side）上的数据（通常经过加密）**。
- 利用**Cookies**文件，网络营销人员能够快速了解顾客以及他或她上一次在网站上的行为.

# A TYPICAL NETSCAPE COOKIE FILE

The screenshot shows a Windows Notepad window with the title bar 'cookies - Notepad'. The menu bar includes 'File', 'Edit', 'Search', and 'Help'. The content area displays a list of cookies in a tabular format. The columns represent the domain, value, path, and various cookie parameters like name, value, and expiration date.

Domain	Value	Path	Name	Value	Expiration
.amazon.com	TRUE	/	FALSE	2082794688	ubid-main 077-5856495-5508330
.yahoo.com	TRUE	/	FALSE	1271369085	B b5n4p1gtmh2ep&b=2
.b2bmarketingbiz.com	TRUE	/	FALSE	2137630415	PREVENT_POPUP 1
.google.com	TRUE	/	FALSE	2147375933	PREF ID=31a7b20218407e37:TM=996890867:LM=996890867
.lbfast.com	TRUE	/	FALSE	1628387455	UID 2 4980464748 20210807
.doubleclick.net	TRUE	/	FALSE	1920510222	id 8000000bd25becd
.advertising.com	TRUE	/	FALSE	1154930584	ACID ee040009971597910016!
.hitbox.com	TRUE	/	FALSE	1028775502	WSS_GW V1AQ@
www.thestandard.com	FALSE	/	FALSE	1312563906	aid 129.37.115.166.686997192825307
www.ecommercetimes.com	FALSE	/	FALSE	1028757090	Apache 166.72.133.237.10673997210009776
.internet.com	TRUE	/	FALSE	1293851081	RMID a64885ed3b7037f0
.adobe.com	TRUE	/	FALSE	1312591094	AWID 129.37.76.22.2659997220007941
.zdnet.com	TRUE	/	FALSE	1041321881	cgversion 4
.avenuea.com	TRUE	/	FALSE	1312599903	AA002 997240055-67392125/998449754
www.business2.com	FALSE	/	FALSE	1028788635	wtCookie 8710223371548699
.timeinc.net	TRUE	/	FALSE	2051233482	PFUID cdbcee543b70b2d80Fbc1000FFFFF9d
.pathfinder.com	TRUE	/	FALSE	2051233482	PFUID cdbcee423b70b2e114761002FFFFF9d
cookies.cmpnet.com	FALSE	/	FALSE	1043910042	Apache 32.103.44.65.15006997242961885
.wired.com	TRUE	/	FALSE	2145927881	p_uniqid 7AXuU3NLY91AJx79CC
.salon.com	TRUE	/	FALSE	1293851080	RMID 20672c413b70bc30
.marketwatch.com	TRUE	/	FALSE	1293851080	RMID 812574833b717ab0
.spinbox.net	TRUE	/	FALSE	1060385851	SBID E32B672002AC28888&S=472480CC
.office.com	TRUE	/	FALSE	1155022982	UTC UT997332236702198904
.mediaplex.com	TRUE	/	FALSE	1245628800	svid 9973322184954002892166704995
www.businessweek.com	FALSE	/	FALSE	2145812682	NGUserID a646515-13599-99733030-5
.cnet.com	TRUE	/	FALSE	2145841482	aid D8C8F7813B721E14000976F200006503
.msn.com	TRUE	/	FALSE	1065305100	MC1 U=2&GUID=4AA7FD1069A44E5EADAE679A5BC12A3
.expedia.com	TRUE	/	FALSE	1065305102	MC1 U=2&GUID=4AA7FD1069A44E5EADAE679A5BC12A3
.mycereal.com	TRUE	/	FALSE	2051233759	SITESERVER ID=ed12eF6F2b870ef998511Fe29b1b3a0F
.bcg.com	TRUE	/	FALSE	2051233304	SITESERVER ID=187afb7d865a997bf07846b1179a7bbb

# COOKIE 的种类

□Cookie的生命周期有两种，一种是整个会话的，一种是永久的。

- 一种是临时性的Cookie（“会话Cookie”），用户关掉浏览器，这个Cookie也就失效了。
- 一种是永久的Cookie，可以持续存在的。一般网站分析工具 判断Unique Visitor使用的是后者。

# COOKIE 的种类

## □ 第一方和第三方COOKIE

- “第一方Cookie”，指的是来自当前正在访问的网站，储存了一定信息的Cookie；
- “第三方Cookie”，指的是来自当前访问网站以外的站点，最常见的就是那些在被访问站点放置广告的第三方站点，这第三方站点可能正在使用Cookie

# GOOGLE ANALYTICS COOKIE



Google Analytics

- Google Analytics 使用第一方 Cookie 来跟踪访问者的互动情况。
- 系统会使用这些 Cookie 来存储信息，例如当前访问发生的时间、访问者以前是否来过网站以及访问者是通过哪一个网站来到这个网页。
- 浏览器不会跨多个域共享第一方 Cookie。

# GOOGLE ANALYTICS



Google Analytics

- Google Analytics设置了5个不同的cookie，\_utma, \_utmb, \_utmc, \_utmz和\_utmv。分别用来存储不同的信息。
- 所有的cookie统一使用\_u-t-m开头，所以我们看到这三个字母就知道是google的cookie。

站点	Cookie 名称
bluewhale.cc	_utmb
bluewhale.cc	_utmv
bluewhale.cc	_utmc
bluewhale.cc	_utma
bluewhale.cc	_utmz

# UTMA : 识别唯一身份访客

名称: \_\_utma

内容: 127635166.1360367272.1264374807.1264374807.1264374807.1

域: .bluewhale.cc

路径: /

发送条件: 任意类型的连接

过期时间: 2012年1月24日 15:13:26

- utma的生存周期为2年。其中第二组的随机唯一ID和第三组的时间戳联合组成了访问者ID，Google Analytics通过这个ID来辨别网站的唯一访问者。
- utma Cookie存储的内容：
  - 第一组数字被叫做“域哈希”，是GA表示这个域的唯一代码。
  - 第二组数字是一个随机产生的唯一ID。
  - 第三，四，五组数字是时间戳
    - 第三组数字表示初次访问的时间
    - 第四组数字表示上一次访问的时间
    - 第五组数字表示本次访问开始的时间。
  - 第六组数字是访问次数计数器。这个数字随着访问次数的增加而增加。
- PS: 上面的三个时间戳数字相同，并且最后的访问次数计数器是1，表示这是第一次访问。

# **\_UTMB : 和\_UTMC一起决定访客的 SESSION**

- \_utmb的生存周期为30分钟，当访问者在你的网站持续30分钟静止时，utmb将被删除。**
- Google Analytics使用\_utmb 和\_utmc一起辨别一个session。**

名称：	<u>_utmb</u>
内容：	127635166.2.10.1264374807
域：	.bluewhale.cc
路径：	/
发送条件：	任意类型的连接
过期时间：	2010年1月24日 15:43:26

# **\_UTMC: 和\_UTMB一起决定访客的SESSION**

名称：	__utmc
内容：	127635166
域：	.bluewhale.cc
路径：	/
发送条件：	任意类型的连接
过期时间：	当会话结束时

- utmc是一个临时cookie，当用户关闭浏览器时\_utmc将一起被删除**
- \_\_utmc和\_\_utmb一起来识别一个session，当用户访问一个网站时，Google Analytics会检查这两个cookie，如果缺少其中任何一个，Google Analytics都将认为这是一个新的session。**

# \_utmz：存储流量来源信息和链接标记的变量值

名称： \_\_utmz  
内容： 127635166.1264374807.1.1.utmcsrc=(direct)|utmccn=(direct)|utmcmd=(none)  
域： .bluewhale.cc  
路径： /  
发送条件： 任意类型的连接  
过期时间： 2010年7月26日 4:13:26

- \_utmz的生存周期是6个月， \_utmz中存储了所有流量的来源信息。
- \_utmz的内容中一共有4组数字：
  - 第一组数字是“域哈希”。
  - 第二组数字是时间戳。
  - 第三组数字是session number。
  - 第四组数字是campaign number 记录通过不同来源访问网站的次数。
  - utmcsrc=(direct)|utmccn=(direct)|utmcmd=(none)

这些信息代表流量的来源，如果是输入域名直接访问的，则来源和媒介都是direct

# **\_UTMV : 自定义访问者的属性**

名称: \_\_utmv  
内容: 127635166.user  
域: .bluewhale.cc  
路径: /

发送条件: 任意类型的连接

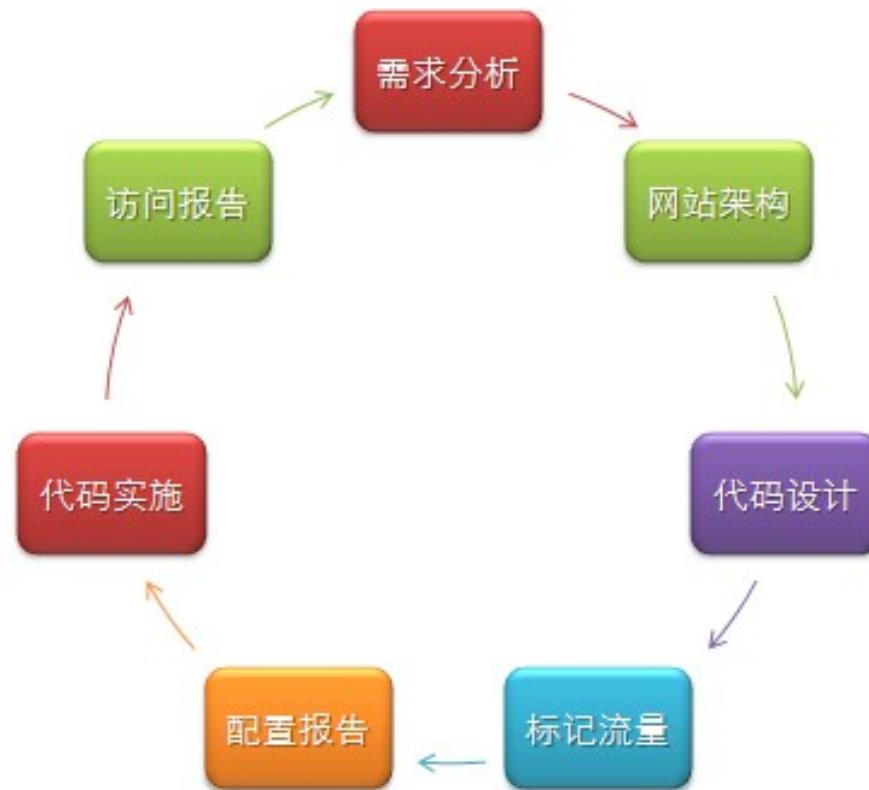
过期时间: 2012年1月24日 15:13:26

- \_\_utmv的生存周期是2年，存储通过\_setVar() 自定义用户属性。
- \_\_utmv的内容: 127635166.user
  - 第一组数字是“域哈希”。
  - 第二个值user是通过\_setVar() 设置的用户属性。
- PS: \_\_utmv只有在调用了\_setVar() 时才会出现。



# Google Analytics实施七步曲

1. 明确网站需求及目标
2. 详细了解网站结构
3. 设计个性化追踪代码
4. 标记所有关注的流量
5. 配置个性化报告
6. 代码实施及检查
7. 访问并检查报告



# Google Analytics账户的建立，以及跟踪代码的配置

---

- <script type="text/javascript">
- var \_gaq = \_gaq || [];
  - \_gaq.push(['\_setAccount', 'UA-32520222-1']);
  - \_gaq.push(['\_setDomainName', '120fd.com']);
  - \_gaq.push(['\_addOrganic', 'soso', 'w']);
  - \_gaq.push(['\_addOrganic', 'sogou', 'query']);
  - \_gaq.push(['\_addOrganic', 'youdao', 'q']);
  - \_gaq.push(['\_addOrganic', 'baidu', 'word']);
  - \_gaq.push(['\_addOrganic', 'baidu', 'wd']);
  - \_gaq.push(['\_addOrganic', 'baidu', 'q1']);
  - \_gaq.push(['\_trackPageview']);
  - \_gaq.push(['\_trackPageLoadTime']);
- (function() {
  - var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
  - ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-analytics.com/ga.js';
  - var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
- })();
- </script>

[Analytics Settings](#)[View Reports:](#)

Click to select Website Profile



## Dashboard

[Saved Reports](#)

## Visitors

## Traffic Sources

## Content

## Goals

## Ecommerce

## Settings

[Email](#)

## Site Usage

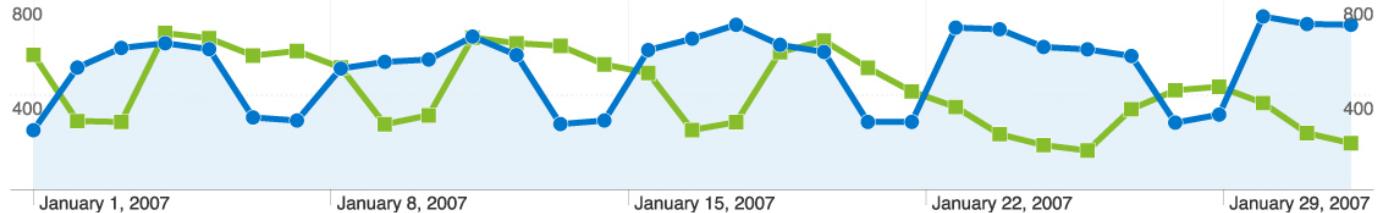
- [About this Report](#)
- [Conversion University](#)
- [Common Questions](#)
- [Report Finder](#)

## Dashboard

[Export](#)

Jan 1, 2007 - Jan 31, 2007

Dec 1, 2006 - Dec 31, 2006    Jan 1, 2007 - Jan 31, 2007    Visits



## Site Usage

**16.107** Visits

Dec 1, 2006 - Dec 31, 2006: 13,209 (21.94%)

**3.86** Pages/Visit

Dec 1, 2006 - Dec 31, 2006: 4.08 (-5.37%)

**36.81%** Bounce Rate

Dec 1, 2006 - Dec 31, 2006: 35.98% (2.30%)

**62.142** Pageviews

Dec 1, 2006 - Dec 31, 2006: 53,855 (15.39%)

**00:03:22** Avg. Time on Site

Dec 1, 2006 - Dec 31, 2006: 00:03:31 (-4.27%)

**50.44%** % New Visits

Dec 1, 2006 - Dec 31, 2006: 49.73% (1.42%)

## Visitors Overview



## Map Overlay world



# 百度统计



官方贴吧 | 联系我们



二维码 ▾

首页

产品介绍

带你入门

成功案例

帮助中心

登录

注册

## 百度统计转化詳解

教您跟踪好 推广效果 的最后一环

重要按钮 »

活动页面 »

咨询工具 »

移动咨询控件 »

## 统计系统



### 百度统计-网站统计

专业的网站流量  
分析工具

[演示](#) [注册](#) [登录](#)



### 百度统计-推广分析

百度推广效果一站式  
评估工具

[演示](#) [注册](#) [登录](#)



### 百度统计-移动统计

免费移动应用统计  
分析工具

[演示](#) [注册](#) [登录](#)



### 百度统计-开放平台

数据收集与导出的  
开放API工具

[开始使用](#)

## 最新升级

- 百度推荐广告变现功能强势上线！ 15-05-19
- 百度游戏统计上线啦！ 14-12-19
- 百度统计电商分析正式上线了！ 14-10-29
- 百度统计应用中心全新上线。 14-05-29
- 百度统计武汉站交流会成功举办 14-05-23

## 经验分享

[更多]

百度统计真是太棒了，不但数据类型齐全，界面友好，数据及时性强，对门户和社区...

-- 杭州19楼总经理 程伟锋

百度统计不仅功能强大易用，更为重要的是提供安全稳定的服务和准确可靠的数据，...

-- 7k7k小游戏 COO 刘学

百度统计帮助了中国站长从不同角度去理解网站运营，将过去简单的“看流量”的习...

-- 站长之家 北京分公司总经理 邱松

## 网站服务



百度推荐  
专业的网站内容  
推荐工具

[开始使用](#)



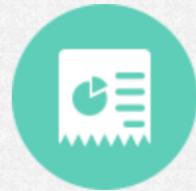
流量研究院  
中国互联网权威的数据  
资讯中心

[开始使用](#)



百度站长平台  
免费的SEO工具及站长  
交流平台

[开始使用](#)



数据市场  
百度大数据产品  
整合中心

[开始使用](#)

## 典型客户

[更多]



# WEB BUGS

- 是隐藏在电子邮件信息和网站上的小图形文件(1×1象素);
- 自动将用户信息和其访问的网页信息传递给监视服务器

# 社会透视: WEB BUGS 是否非法?

- 营销人员认为他们利用**Web bugs**只是用来帮助收集统计数据; **Privacy advocates say, if so, why are they hidden?**
- **Web bugs**类型包括透明 **GIF**,可执行 **bugs**, 以及 **script-based executable bugs**
- **Privacy Foundation guidelines for Web bug usage:**
  - 应该是能在屏幕上显示的可视图标;
  - 必须注明其功能以及在网页上放置该**web bugs**的公司名;
  - 点击该**web bugs**时, 应该能够告知用户正在收集什么数据, 哪些公司将会拥有这些数据, 他们将如何使用这些数据?
  - 应该可以提供“禁用”选择
- 目前没有相关法律

## 7. RELATED APPLICATIONS

- 精准网络营销
- 电子商务个性化推荐系统
- 客户关系管理
- .....

# 7.1 精准网络营销

# PRECISION INTERNET MARKETING

US Online Advertising Spending, 2006-2012 (billions)

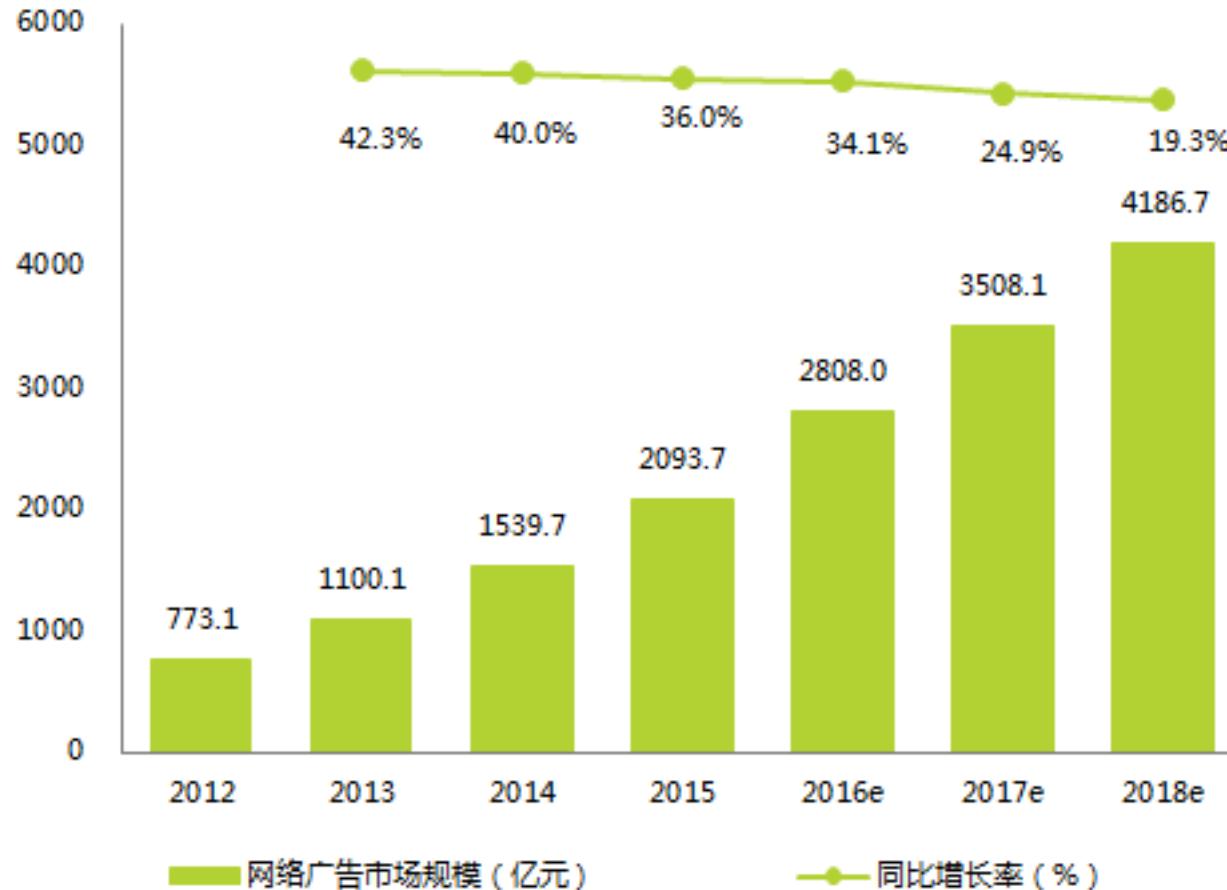


Note: eMarketer benchmarks its US online advertising spending projections against the Interactive Advertising Bureau (IAB)/PricewaterhouseCoopers (PwC) data, for which the last full year measured was 2007; online ad data includes categories as defined by IAB/PwC benchmark—display ads (such as banners), search ads (including paid listings, contextual text links and paid inclusion), rich media (including video), classified ads, sponsorships, lead generation (referrals) and e-mail (embedded ads only); excludes mobile ad spending

Source: eMarketer, March 2008

- When
- Who
- Where
- What
- How

## 2012-2018年中国网络广告市场规模及预测



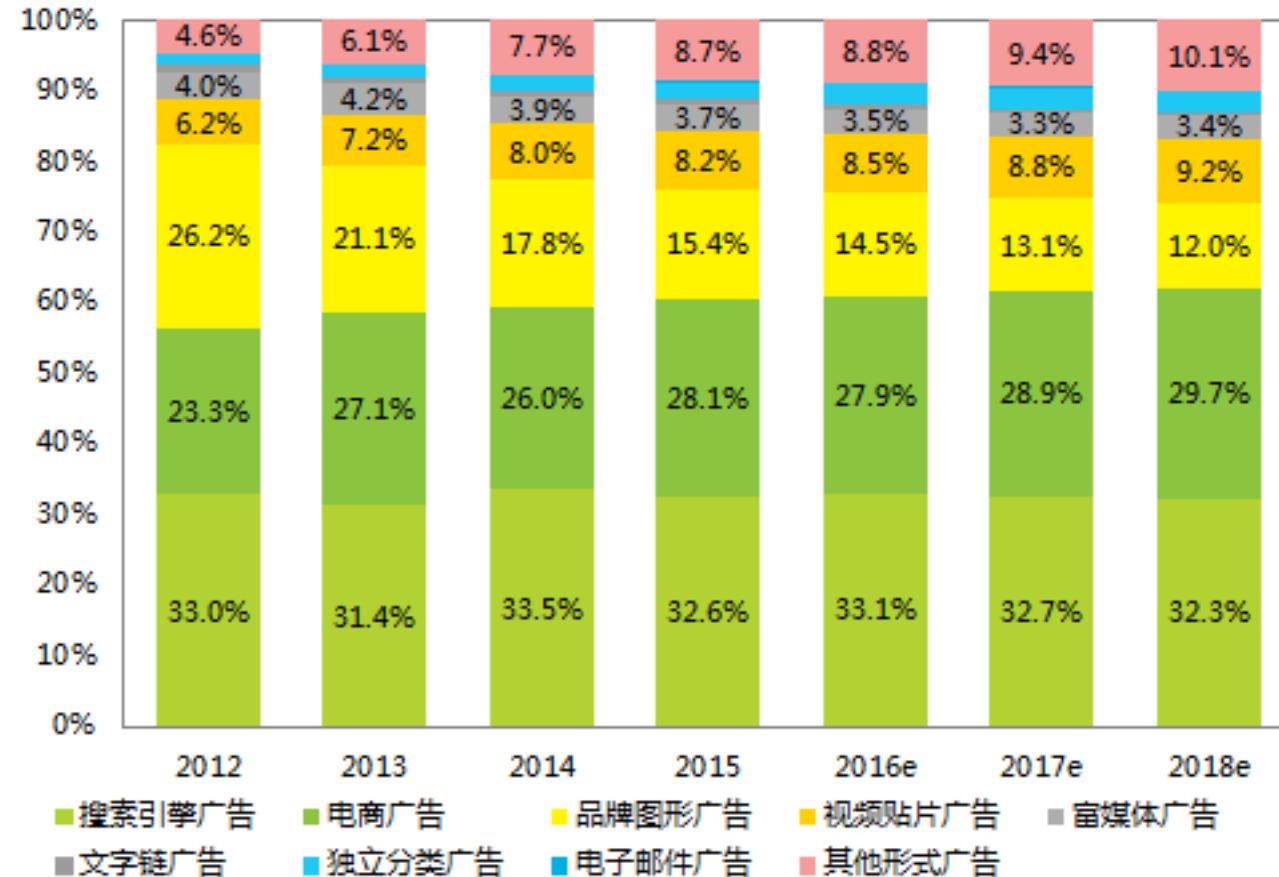
2015年中国网络广告  
市场规模为**2093.7亿**  
元，同比增长**36.0%**。

注释：1.互联网广告市场规模按照媒体收入作为统计依据，不包括渠道代理商收入；2.此次统计数据包含搜索联盟的联盟广告收入，也包含搜索联盟向其他媒体网站的广告分成。

来源：根据企业公开财报、行业访谈及艾瑞统计预测模型估算。

## 2012-2018年中国不同形式网络广告市场份额及预测

- 2015年，搜索广告依旧是份额占比最大的广告类型，占比为32.6%，较2014年占比略有下降。
- 电商广告份额排名第二，占比达28.1%，比2014年增长2个百分点。
- 品牌图形广告市场份额持续受到挤压，位居第三，占比为15.4%。视频贴片广告份额继续增大，占比为8.2%。
- 其他广告形式份额增长迅速，占比达8.7%，主要包括导航广告和门户社交媒体中的信息流广告等

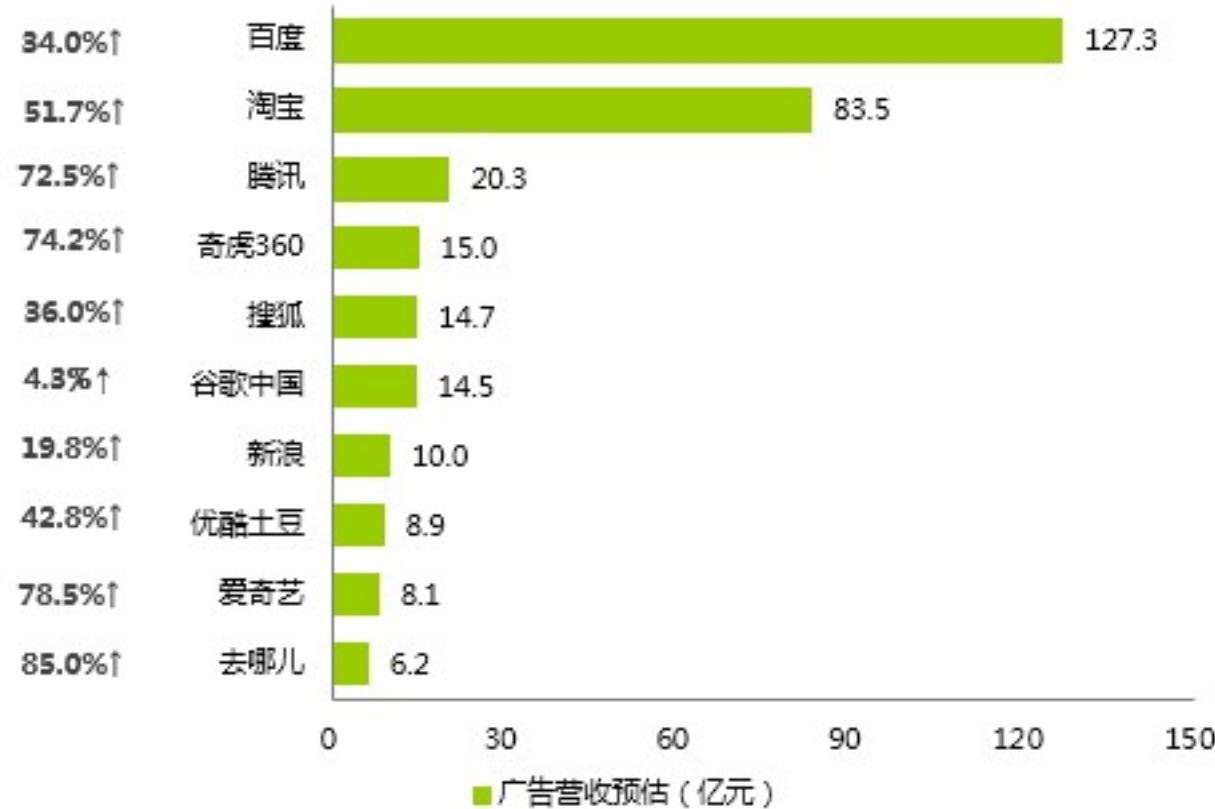


注释：1. 搜索广告包括搜索关键字广告及联盟广告，搜索引擎广告>搜索广告>搜索关键词广告；2. 电商广告包括垂直搜索类广告以及展示类广告，例如淘宝、京东、去哪儿；3. 独立分类广告从2014年开始核算，仅包括58同城、赶集网等分类网站的广告营收，不包含搜房等垂直网站的分类广告营收；4. 其他形式广告包括导航和门户及社交媒体中的效果类广告。

来源：根据企业公开财报、行业访谈及艾瑞统计预测模型估算。

## 2015Q1中国网络广告市场媒体营收规模预估Top10

同比增长率(%)



注释：1.各上市企业广告营收统计标准以其财务报表中公布的广告营收数字为准，不考虑因税收和返点引起的统计口径差异；2.搜狐广告营收包括门户和搜狗的广告营收；3.淘宝广告营收由财报及其他公开信息结合艾瑞咨询集团推算模型估算，淘宝广告营收为中国商业零售业务中的核心收入来源，其广告营收不包含佣金收入及其他店铺费用。

来源：根据企业公开财报、行业访谈、iAdTracker监测数据及艾瑞统计预测模型估算，仅供参考。

- 根2015Q1，百度广告营收位列第一，为127.3亿元，同比增长34.0%。
- 淘宝预估广告营收为83.5亿元，同比增长51.7%。
- 腾讯预估广告营收为20.3亿元，同比增长为72.5%。

# 精准广告的基础—用户行为分析

- 是指在获得网站访问量基本数据的情况下，对有关数据进行统计、分析，从中发现用户访问网站的规律；
- 将这些规律与网络营销策略等相结合，从而发现目前网络营销活动中可能存在的问题；
- 并为进一步修正或重新制定网络营销策略提供依据。

# 用户行为分析

口 用户行为分析应该包含以下重点分析数据：

- 用户在网站的停留时间、跳出率、回访者、新访问者、回访次数、回访相隔天数；
- 注册用户和非注册用户，分析两者之间的浏览习惯；
- 用户所使用的搜索引擎、关键词、关联关键词和站内关键字；
- 用户选择什么样的入口形式（广告或者网站入口链接）更为有效；
- 用户访问网站流程，用来分析页面结构设计是否合理；
- 用户在页面上的网页热点图分布数据和网页覆盖图数据；
- 用户在不同时段的访问量情况等。

# 精准营销—主要技术应用

## □ 1. 广告监控

- 内涵：包括广告展示、点击、跳转、关键性页面的到达。
- 明星公司：doubleclick



# 精准营销—主要技术应用

## □ 2. 广告发布和媒体广告位管理

- 内涵：包括广告位的定义，各种尺寸的素材的支持，各种结算方式的支持（CPM和CPA）以及广告的轮流显示等。
- 明星公司：包括doubleclick, openx等。

# Google Ad Exchange

## 美国的网络营销市场

2007年4月

Google以31亿美元收购网络营销公司DoubleClick

The screenshot shows a news article from Sina Tech Times. The headline reads "谷歌拟 31 亿美元购 DoubleClick". Below the headline, there is a photo of a road sign for Doubleclick. The main text of the article discusses Google's acquisition of Doubleclick for \$31 billion. A sidebar on the right contains a poll about the impact of the acquisition on Microsoft.

**科技时代** | 新浪首页 | 导航 | 搜索 | **Google** 网页 | 新闻

## 谷歌拟 31 亿美元购 DoubleClick

**31 亿美元现金**

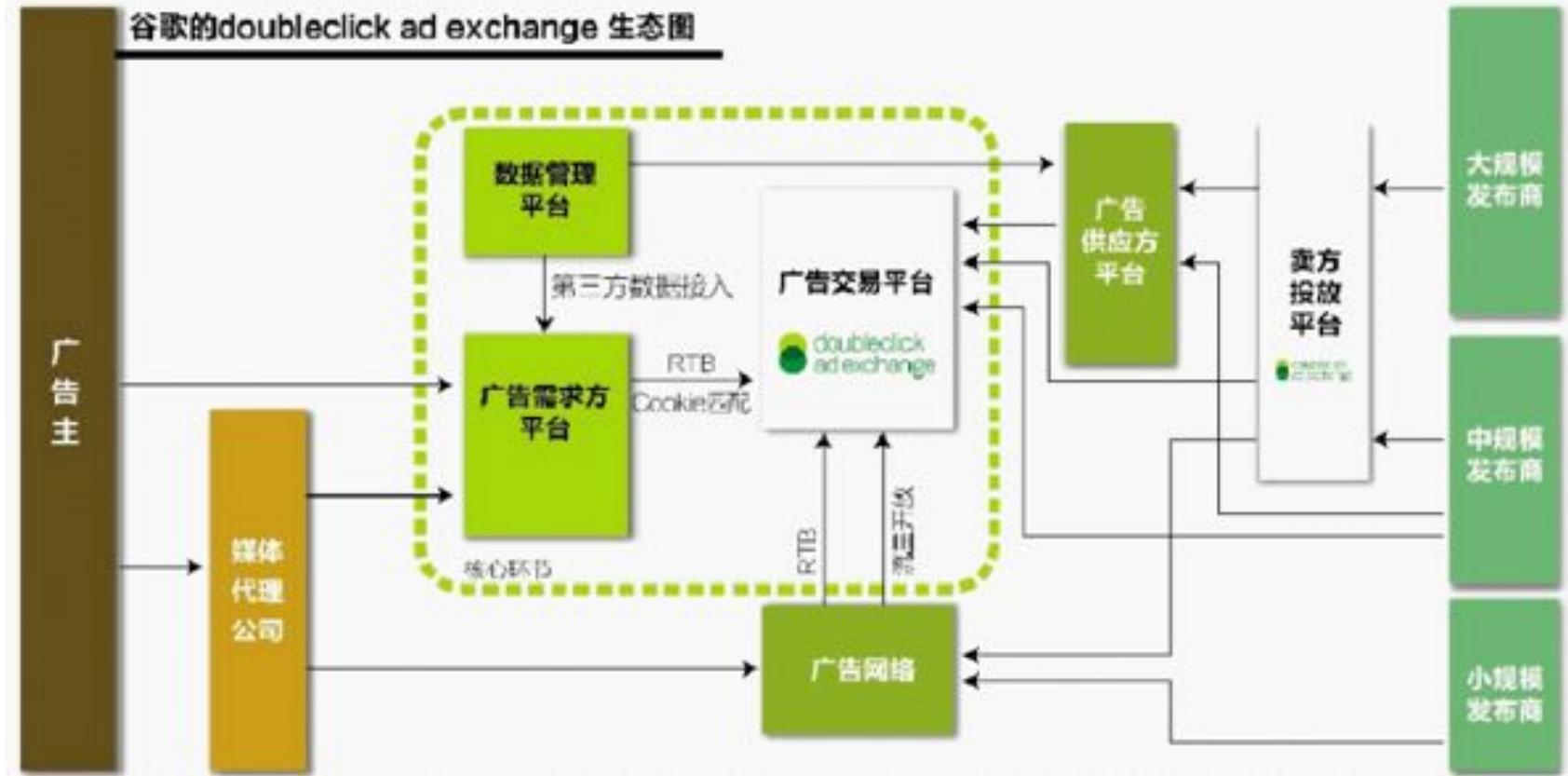
**DoubleClick**

**你怎么看谷歌收购 DoubleClick后对雅虎、微软形成的影响**

形成很大威胁  
 没什么影响  
 说不好

**提交** **查看**

## 谷歌的doubleclick ad exchange 生态圈



# 美国的网络营销市场

2007年5月18日

微软以**60亿美元**收购网络营销公司aQuantive

## WebLeOn's Blog

webleon blogging the internet

blog profile links gallery

---

### Microsoft收购aQuantive

2007年5月19日 星期六

Microsoft即将收购数字营销服务商aQuantive，收购价格是令人咂舌的60亿美元。收购方式将全部以现金购买股票的方式，并与Microsoft的2008财年的上半年完成。

**aQuantive**

```
graph TD; aQuantive --> DMS[Digital Marketing Services]; aQuantive --> DMT[Digital Marketing Technologies]; aQuantive --> DPM[Digital Performance Media];  
DMS --- razorfish[avenue a  
razorfish];  
DMT --- atlas[atlas  
Software for Marketing];  
DPM --- DRIVE[DRIVE  
PERFORMANCE MEDIA]
```

Google Custom Search 站内搜索

About me

webleon blogging the internet, more at: [Blogger Profile](#)  
contact me:  
[WebLeOn@Gmail.com](mailto:WebLeOn@Gmail.com)

Google talk twit this 分享到 饭否

---

### MyBlogLog

You! Join My Community



# 美国的网络营销市场

2007年5月下旬

WPP以6.49亿美元的价格成功收购网络营销公司24/7RealMedia



TOM首页 > TOM科技 > 互联网 > 正文

WPP 收

## WPP已收购24/7RealMedia 微软收购战再败一阵

<http://tech.tom.com> 2007年05月18日 07时17分 TOM科技 孤行 编译

**关键字:** WPP 收购 24/7RealMedia 微软

**TOM科技讯** 美国东部时间5月17日（北京时间5月18日）据外电的最新报道称，世界第二大广告公司WPP集团周四宣布，该公司将以6.49亿美元的价格收购网络广告公司24/7 Real Media。此前曾有消息称，微软计划出价10亿美元收购24/7 Real Media。

# 美国的网络营销市场

2007年7月初

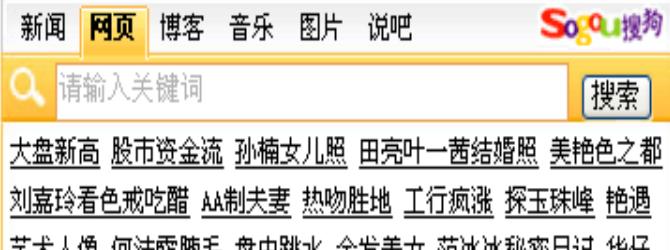
AOL拟斥资3亿美元收购网络营销公司Tacoda

搜狐IT | IT频道 > 互联网 > 国际互联网

AOL收购Tacoda在线广告 美反垄断机构绿灯放行

2007年08月28日08:19

[我来说两句] [字号: 大 中 小]



# 美国的网络营销市场

2007年7月27

## 微软收购网络营销交易平台AdECN

用户名:  密码:  登录 新用户注册

天极传媒: 天极网 | 比特网 | IT专家网 | Impress中文站

**即时新闻** **要闻** **深度报道** **专题** **博客|论坛|评论** **视频**

简体版 | 繁体版 热点推荐: 天极传媒集团 新品牌暨新战略发布会 滚动SOA频道 索尼专区

天极搜索

**天极网站群联合招聘** 

首页 互联网 比特网 > 互联网 > 微软收购网络广告公司AdECN 对抗Google 戴尔高端商务本，国庆惊喜价¥ 5799！ 800-858-0

业界 互联网 门户·搜索

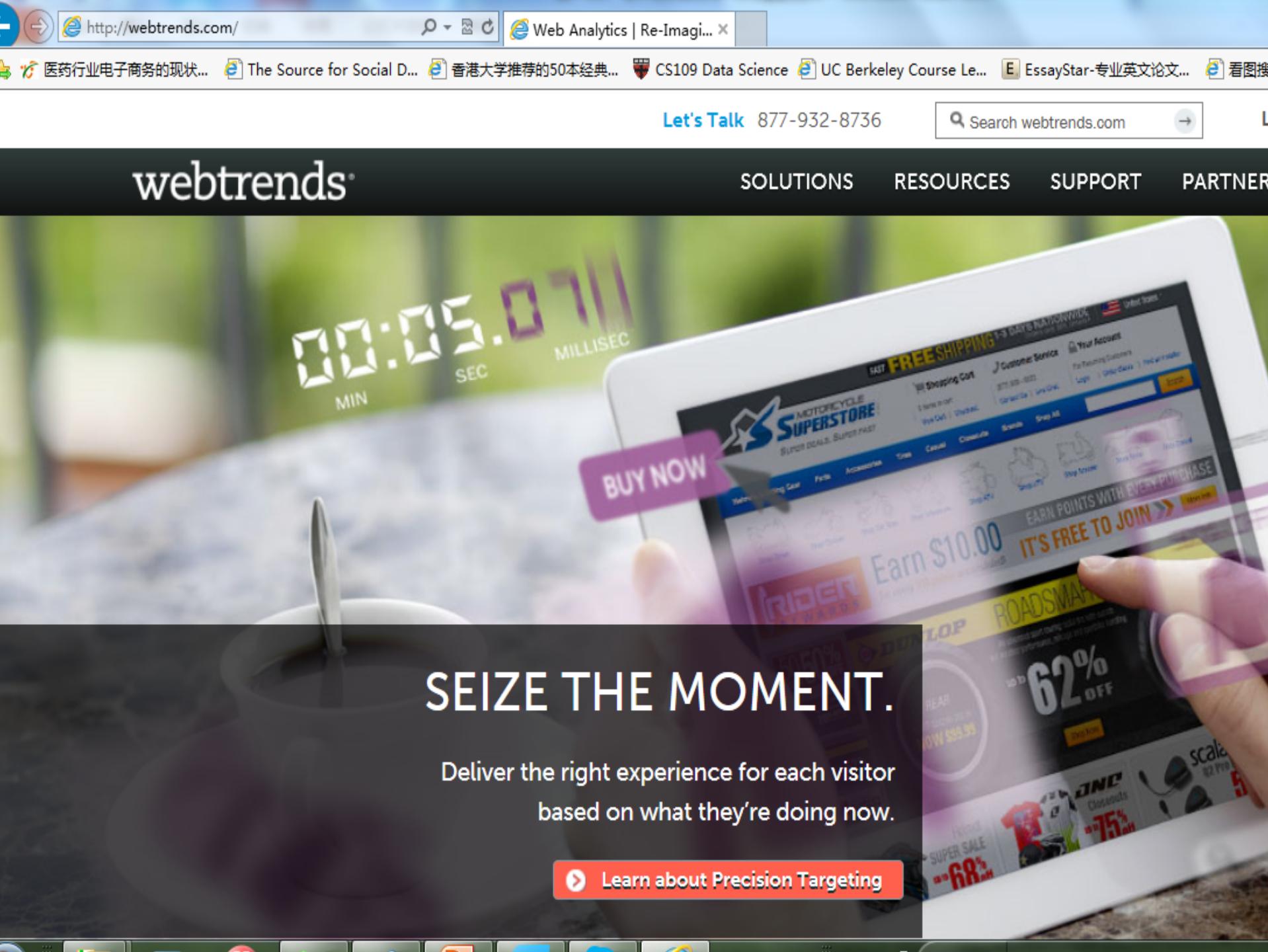
本文关键字: [微软\(9959\)](#) [Google\(4889\)](#) [雅虎\(2064\)](#)

微软收购网络广告公司AdECN 对抗Google  
出处: 天极ChinaByte 作者: 老沈 2007-07-27 07:50 评论

# 网络广告主要技术应用

## □ 3. 网站用户行为跟踪和分析

- 内涵：用户在网站关键性页面的到达，就是网站用户行为跟踪的一个指标，其他指标还包括网站的跳出率，用户停留时间，用户的回访率，新旧用户比例等几十个指标。
- 明星公司产品：包括Webtrends，Coremetric，Google analytics等。



Let's Talk 877-932-8736

Search webtrends.com

webtrends®

SOLUTIONS

RESOURCES

SUPPORT

PARTNER

## SEIZE THE MOMENT.

Deliver the right experience for each visitor  
based on what they're doing now.

[Learn about Precision Targeting](#)

[Let's Talk](#) 877-932-8736

 Search webtrends.com

[Login](#)

[SOLUTIONS](#)
[RESOURCES](#)
[SUPPORT](#)
[PARTNERS](#)
[ABC](#)

[Digital  
Measurement](#)
[Multi-Channel  
Measurement](#)
[Heatmaps](#)
[Webtrends Streams™](#)
[Website Measurement](#)
[SAP Integration](#)
[Collaboration  
Measurement](#)
[Social Measurement](#)
[Mobile Measurement](#)

[Conversion  
Optimization](#)
[Mobile & Social  
Optimization](#)
[Online Testing](#)
[Targeting &  
Segmentation](#)
[Optimization for Retail](#)
[Optimization for Travel](#)
[Optimization for Finance](#)

[Campaign  
Optimization](#)
[Facebook Campaigns](#)
[Email Remarketing](#)
[Facebook Advertising](#)
[Ad Spend Optimization](#)

[Collaboration  
Optimization](#)
[SharePoint for Intranet](#)
[SharePoint for Internet](#)
[CRM Exchange](#)

[Services](#)
[Implementation Services](#)
[Data Visualization](#)
[Experts on Retainer](#)
[Training Courses](#)

Follow along as we share our own marketing tests—and reveal what works and what doesn't.


[Get Backstage Access](#)

# 网络广告主要技术应用

## □ 4. 广告用户定向

□ 内涵：包括地域定向、人口统计学信息定向，  
用户兴趣定向，上下文定向。

□ 明星公司产品：包括Google adsense，  
Audience science等。

# 网络广告主要技术应用

## □ 5. 用户抽样和分析

- 内涵：他们基于浏览器插件数据和ISP服务商业数据，对用户进行抽样分析。
- 明星公司：包括Nielsen/netratings, ComScore, Quantcast等公司。



English | Chinese

China

Search

Home

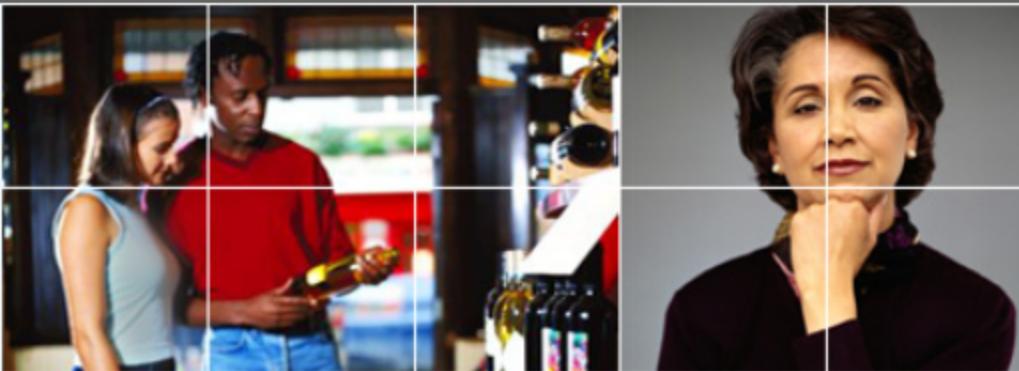
Company

Solutions

News

Careers

Trends & Insights



## Integrating Information

Nielsen gives clients a more complete view of their markets and a better understanding of their consumers through our combined insights, experiences, knowledge, market intelligence and advanced technologies.

Find out more...

Select your business issue ▾

### Nielsen worldwide

Select a market ▾

### Nielsen News

16 January 2014

Nielsen Survey: Health Issues, Unexpected Household Emergencies And Children's Future Top 3 Priorities For Current Saving Plan Among Chinese Consumers

2 January 2014

Nielsen: ComRatings Author Joint Whitepaper on the Chinese Online Cosmetics Shopper

26 December 2013

Nielsen: More Than 86 Percent of Chinese Consumers Show

## Trends & Insights

CONSUMER 360

### 尼尔森中国消费者360论坛

时间: 2013/05/22

地点: 上海浦东嘉里大酒店

### Nielsen Answers

Email

Password

 >

Forgot password?

nielsenwire  
Latest news on the wire, from nielsenwire.com  
Click to read more

Nielsen China Market Beat  
Quarterly Market Insights click here for details



+ 加关注

The Nielsen Company, China  
t.sina.com.cn/nielsenchina

State of The Media: The Social Media Report 2012



互联网广告领域的[大数据管理](#)和[优化平台](#)



[客户登录](#)

## 互联网广告的大数据管理和优化平台

*Internet Audience Data Marketplace*



### 关于我们

ComRatings（讯实科技）是一家基于创新技术的互联网大数据公司。

我们致力于成为领先的第三方互联网广告大数据管理和优化平台，为广告主和营销客户提供以目标用户为中心的程序化洞察，从而帮助互联网营销和



### 新闻动态 News & Activities

#### ■ 2015年11月上海 | 母婴产品网购行为分析白皮书——ComRatings重磅发现，网购的黄金17分钟！

陪女人购物一直是一件让男人苦恼的事情，因为时间真的很！漫！长！但...也有例外！ComRatings中国母婴人群网络行为数据分析白皮书的最新研究发现：年轻妈妈们平均网购时间只有**17分钟**！17分钟是什么概念？大概是听4首歌的时间，大约是坐5站地铁的时间，正是在这么短的时间内，年轻妈妈们完成了一次网购旅程。

电商铺货应打组合拳：“母”+“婴”用品齐上阵

数据发现，在单次购物中，有八成的妈

# 网络广告主要技术应用

## □ 7. 与广告创意相关的技术

□ 内涵：包括富媒体、流媒体技术，也包括线上自主制作广告创意的技术。

□ 明星公司：Smartmedia（好耶）,

Rightmedia

[我们的服务](#)[关于我们](#)[我们的作品](#)[工作在好耶](#)[联系我们](#)[微信二维码](#)[关注我们](#)

## 我们的服务

# 1 品牌数字整合营销

### ① 品牌数字推广策略制定

通过对数字化品牌的诊断以及目标用户网络行为的分析，制定基于用户触点的品牌数字整合营销策略和系列活动方案。

### ② 创意发展与制作

将网络技术与创新能力相结合，在互联网空间中生成品牌体验。通过不断尝试国内外最前沿的互联网技术，从优化用户体验角度实现最佳传播效果。

### ③ 媒介计划与投放

通过对数字媒体的研究，探寻赞助广告、视频广告和富媒体广告等各种不同广告形式的机会与价值，为广告主制定更高ROI的媒介计划与执行方案。



# 中国的网络营销市场

2007年3月初

分众传媒3.25亿美元收购网络营销公司好耶

 科技时代

科技时代 | 新浪首页 | 爱问 (iAsk.com) | 请输关键词 | 搜

## 分众传媒3.25亿美元合并聚众传媒



来源：1月8日消息，分众传媒宣布合并聚众传媒，分众以3.25亿美元的价格取得聚众传媒100%股权。[\[全文\]](#) [\[评论\]](#)

**聚众传媒CEO虞锋谈分众聚众合并**  
1月10日19:00，聚众传媒CEO虞锋做客新浪聊天室，谈与聚众传媒合并的细节。[\[全文\]](#) [\[评论\]](#)

**江南春做客谈合并内情 不会大提价 发布会图文**  
今日12点45分，分众传媒董事长、CEO江南春做客新浪上海聊天室披露合并聚众细节；14点半双方在上海举行发布会共同宣

**专题调查**

你看好的是两者合并吗？

- 好
- 不好
- 不好说

合并会否带来垄断？

- 不会
- 会

# 美国的网络营销市场

2007年5月初

雅虎6.8亿美金收购网络营销公司**RightMedia**20%股份

---

搜狐IT | [IT频道](#) > [互联网](#) > [国际互联网](#)

雅虎6.8亿美元收购网络广告中介Right Media

2007年04月30日22:10

[[我来说两句\(1\)](#)] [[字号：大](#) [中](#) [小](#)]



您需要又便宜又精准的流量么？ 您还在为老用户不再回头而苦恼么？ 您还在买某宝里的昂贵流量么？

# 摸象大数据推荐引擎

—— 史上最高ROI的全网直通车

RTB实时竞价，超低CPC起价，  
无限量免费曝光，灵活预算控制，  
摸象大数据为您的事业加速！

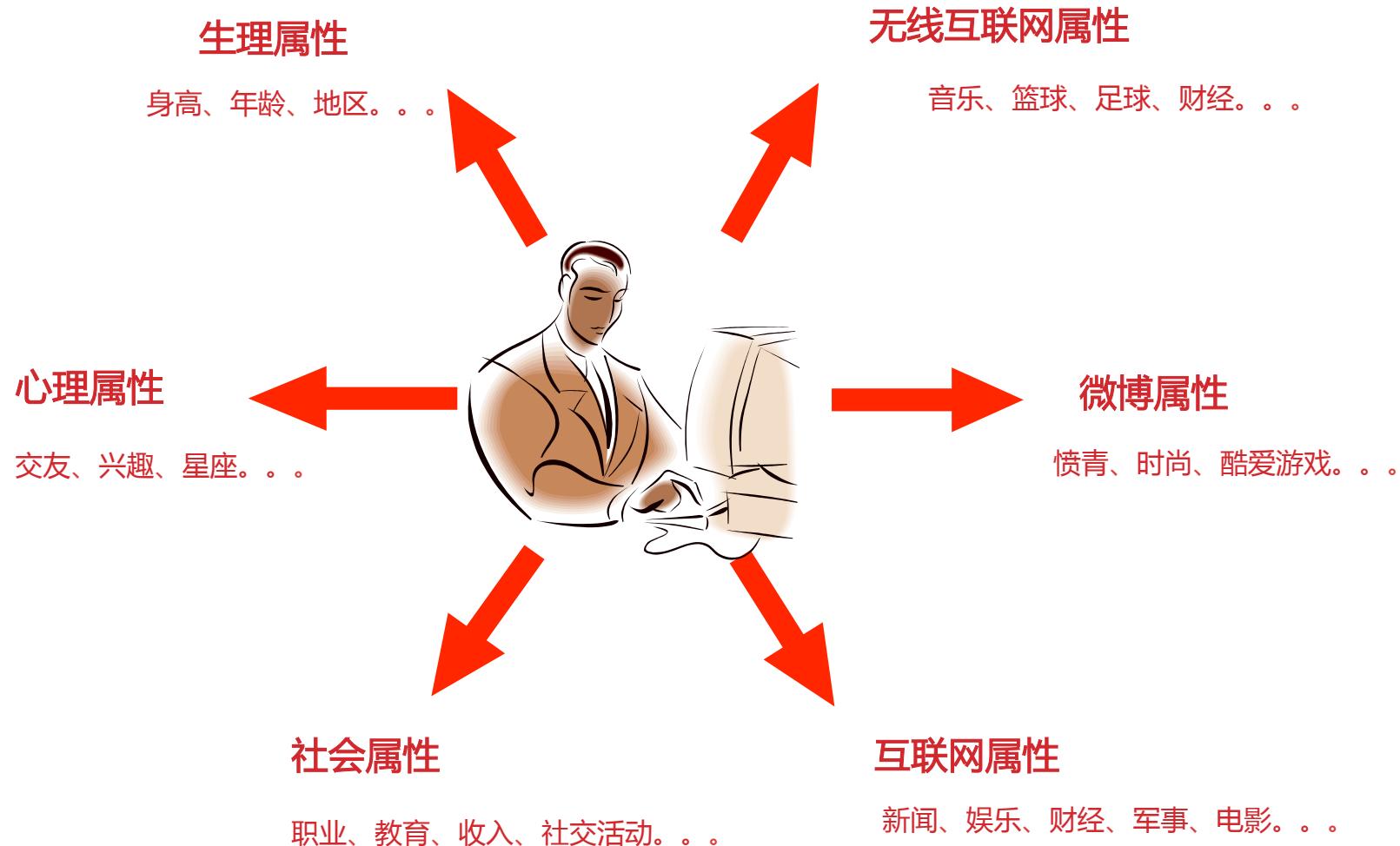
# 摸象互动—第三代互联网广告平台

从“购买广告位”到“购买目标用户”的革命，打造中国最领先的DSP平台

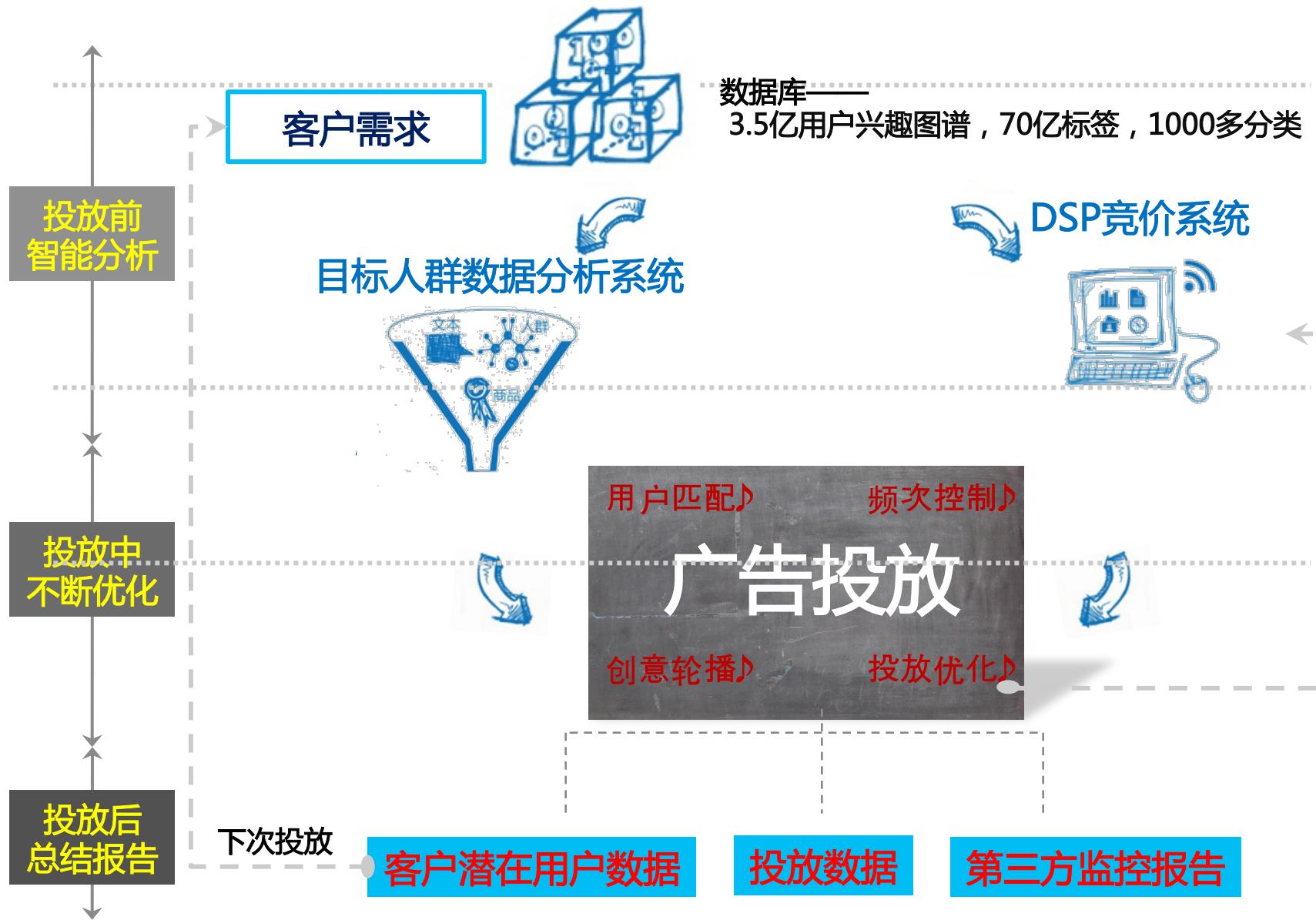


# 摸象互动个人兴趣图谱建立

人群画像库——3.5亿用户兴趣图谱，70亿标签，1000多分类



# 摸象互动投放原理图



# 案例：途牛网广告精准投放

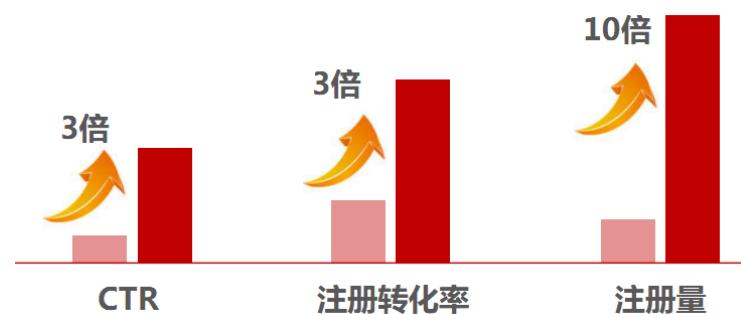
途牛网——国内Top3旅游类网络服务平台，对于ROI(Return On Investment)有着极高的要求。摸象互动通过对数据的分析和投放优化，大幅提升了效果。

**总体目标：**迅速降低广告的营销成本，有效覆盖目标人群，提升注册用户。



**最终结果：**

- 1、CTR提升3倍！
- 2、注册转化率提升3倍！
- 3、注册量提升10倍！



## 7.2 电子商务个性化推荐系统

- 个性化服务 (Personalization)
  - 尽可能使得自己的每个用户在浏览该商业网站时都有他就是该网站的唯一用户的感觉
  - 尽可能地迎合每个用户的浏览兴趣并且不断调整自己来适应用户浏览兴趣的变化

# 个性化服务的表现形式

- 推荐超链接列表
- 推荐商品列表（电子商务）
- 推荐广告列表
- 经裁剪的文本或图像列表

# 实现的步骤

## (1) 构建用户档案

- 收集用户的评分、评价行为等；
- 进行数据清理、转换，最终形成用户对各种项目的评价矩阵。

$R_{ij}$ 代表第*i*个用户对商品*j*的评分。

## (2) 最近邻居搜索

- 计算目标用户与数据集中各个用户的相似度；
- 寻找相似度最高的作为最近邻居集

	Item <sub>1</sub>	Item <sub>2</sub>	.....	Item <sub>n</sub>
User <sub>1</sub>	4	5		4
.....	...	...	$R_{ij}$	...
User <sub>n</sub>	...	4	...	3

## (3) 推荐产生

- 通过一定的推荐算法，根据最近邻居集的评价值，产生推荐

# 相似度的度量

$sim(Item2, Item4)$ ?

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	8	1	?	2	7
User 2	2	?	5	7	5
User 3	5	4	7	4	7
User 4	7	1	7	3	8
User 5	1	7	4	6	5
User 6	8	3	8	3	7

# Problem Definition

Objects: a user set, an item set, a rating matrix

**Target in industry:** Predict the users' unknown actions on other items

The tendency of a user's action on an item can be indicated by a real-valued number— rating

**Target in research:** predict unknown user-item ratings

# Recommendation Metrics

- Prediction accuracy
- Coverage
- Novelty
- Diversity
- Privacy
- Scalability

# Recommendation Strategies

- Popularity
- Association role
- Content based
- Collaborative filtering
- Hybrid

# Recommendation Strategies

## POPULARITY

热度推荐，  
就是把最火的ITEM列在  
前面！

电影榜单 (32593个) :

类型: 全部 爱情 喜剧 动作 剧情 科幻 恐怖 动画 惊悚 犯罪

地区: 全部 美国 大陆 香港 台湾 韩国 日本 泰国 英国 其它

年代: 全部 2016 2015 2014 2013 2012 00年代 90年代 80年代

最热电影 最新电影 用户好评



微微一笑很倾城

豆瓣评分:5.4



激战

豆瓣评分:7.8



三傻大闹宝莱坞

豆瓣评分:9.2



疯狂动物城

豆瓣评分:9.3



我的战争

豆瓣评分:5.2



极限挑战之皇家宝藏

豆瓣评分:4.5



僵尸世界大战

豆瓣评分:7.2



火锅英雄

豆瓣评分:7.4

# Recommendation Strategies

## ASSOCIATION ROLE



关联规则的目的就在一个数据集中找出项与项之间的关系，也被称为购物蓝分析 (Market Basket analysis)，因为“购物蓝分析”很贴切的表达了适用该算法情景中的一个子集。

# Recommendation Strategies

## ASSOCIATION ROLE — APRIORI ALGORITHM (先验算法)

Apriori算法[1] 是一种最有影响的挖掘布尔关联规则频繁项集的算法。其核心是基于两阶段频集思想的递推算法。在这里，所有支持度大于最小支持度的项集称为频繁项集，简称频集。

Let the database of transactions consist of following itemsets:

Itemsets
{1,2,3,4}
{1,2,4}
{1,2}
{2,3,4}
{2,3}
{3,4}
{2,4}

We will use Apriori to determine the frequent item sets of this database. To do so, we will say that an item set is frequent if it appears in at least 3 transactions of the database: the value 3 is the **support threshold** (最小支持度).

Itemsets
{1,2,3,4}
{1,2,4}
{1,2}
{2,3,4}
{2,3}
{3,4}
{2,4}

# Recommendation Strategies

## ASSOCIATION ROLE — APRIORI ALGORITHM

The first step of Apriori is to count up the number of occurrences, called the support(支持度), of each member item separately, by scanning the database a first time. We obtain the following result:

Item	Support
{1}	3
{2}	6
{3}	4
{4}	5

All the itemsets of size 1 have a support of at least 3, so they are all frequent.

**支持度 (Support) :** 定义为  $\text{supp}(X) = \text{occur}(X) / \text{count}(D) = P(X)$

例如：在100个人去超市买东西的，其中买苹果的有9个人，那就是说苹果在这里的支持度是 9, 9/100;

# Recommendation Strategies

## ASSOCIATION ROLE – APRIORI ALGORITHM

The next step is to generate a list of all pairs of the frequent items:

Item	Support
{1,2}	3
{1,3}	1
{1,4}	2
{2,3}	3
{2,4}	4
{3,4}	3

The pairs {1,2}, {2,3}, {2,4}, and {3,4} all meet or exceed the minimum support of 3, so they are frequent.

Item	Support
{2,3,4}	2

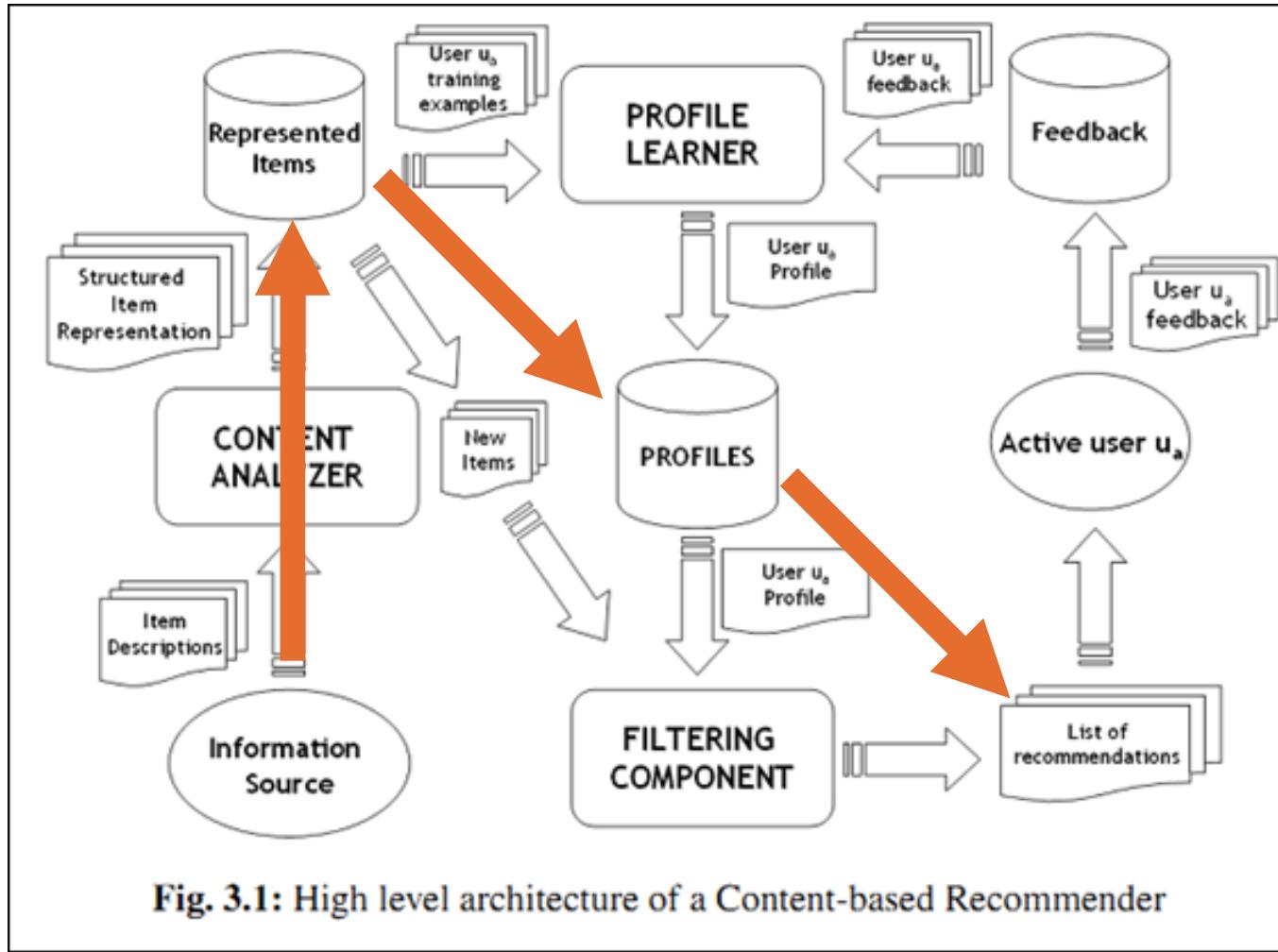
In the example, there are no frequent triplets

# ASSOCIATION ROLE — APRIORI ALGORITHM

- 置信度 (Confidence/Strength)：
  - $\text{conf}(X \rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X) = P(Y|X)$ 。
  - 在历史数据中，已经买了某某（例如：A、B）的支持度和经过挖掘的某规则（例如： $A \Rightarrow B$ ）中A的支持度的比例，也就是说买了A和B的人和已经买了 A的人的比例，这就是对A推荐B的置信度 ( $A \Rightarrow B$  的置信度)
  - 经过关联规则分析后，针对某些人推销（根据某规则）比盲目推销（一般来说是整个数据）的比率，这个比率越高越好，我们称这个规则为**强规则**

# Recommendation Strategies

## CONTENT-BASED



CB应该算是最早被使用的推荐方法，它根据用户过去喜欢的产品（本文统称为 **item**），为用户推荐和他过去喜欢的产品相似的产品。

# 相似度的度量

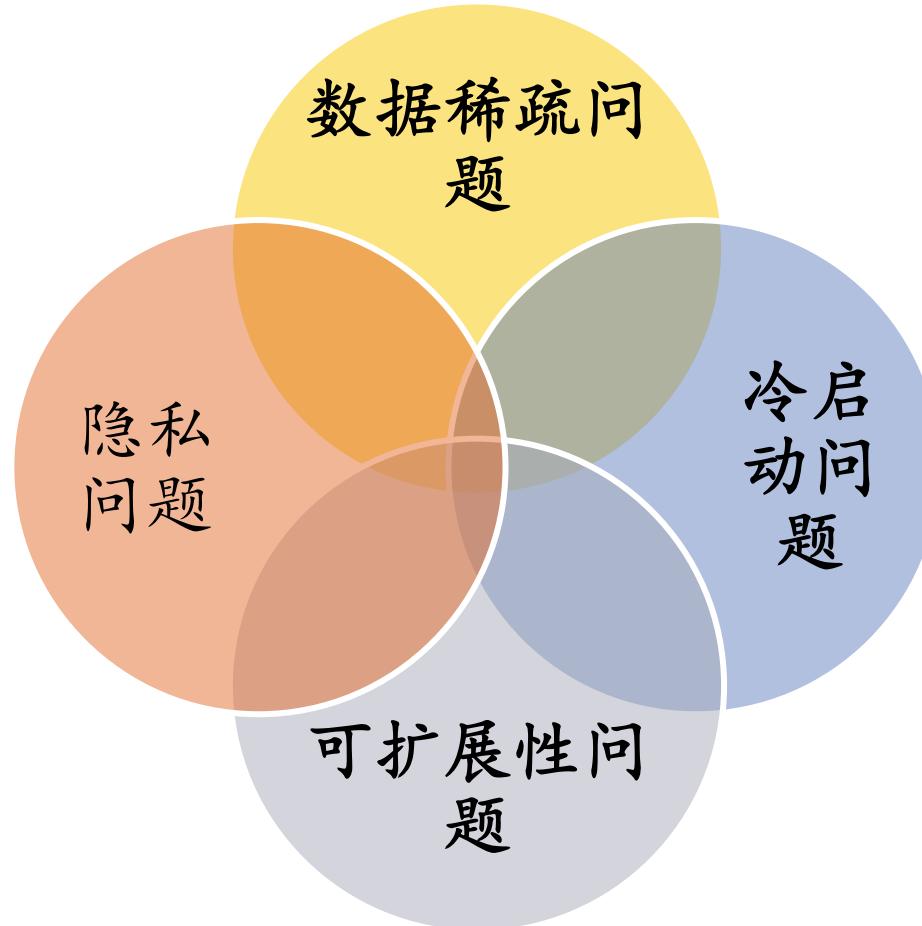
$sim(Item2, Item4)$ ?

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	8	1	?	2	7
User 2	2	?	5	7	5
User 3	5	4	7	4	7
User 4	7	1	7	3	8
User 5	1	7	4	6	5
User 6	8	3	8	3	7

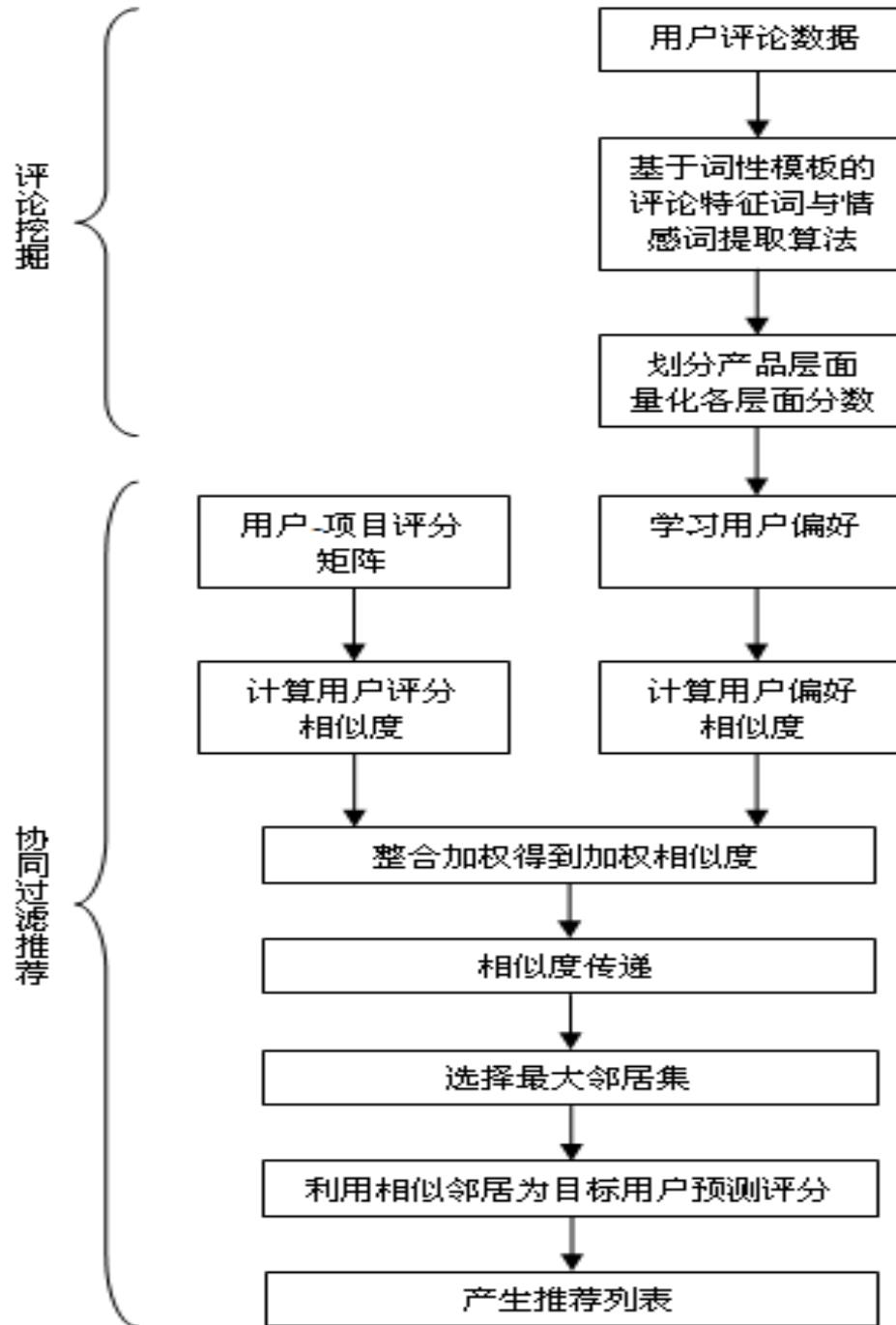
# CONTENT-BASED

1. **Item Representation**: 为每个item抽取出一些特征（也就是item的content了）来表示此item（例如一个商品、一篇文章）；
2. **Profile Learning**: 利用一个用户过去喜欢（及不喜欢）的item的特征数据，来学习出此用户的喜好特征（profile）；
3. **Recommendation Generation**: 通过比较上一步得到的用户profile与候选item的特征，为此用户推荐一组相关性最大的item。

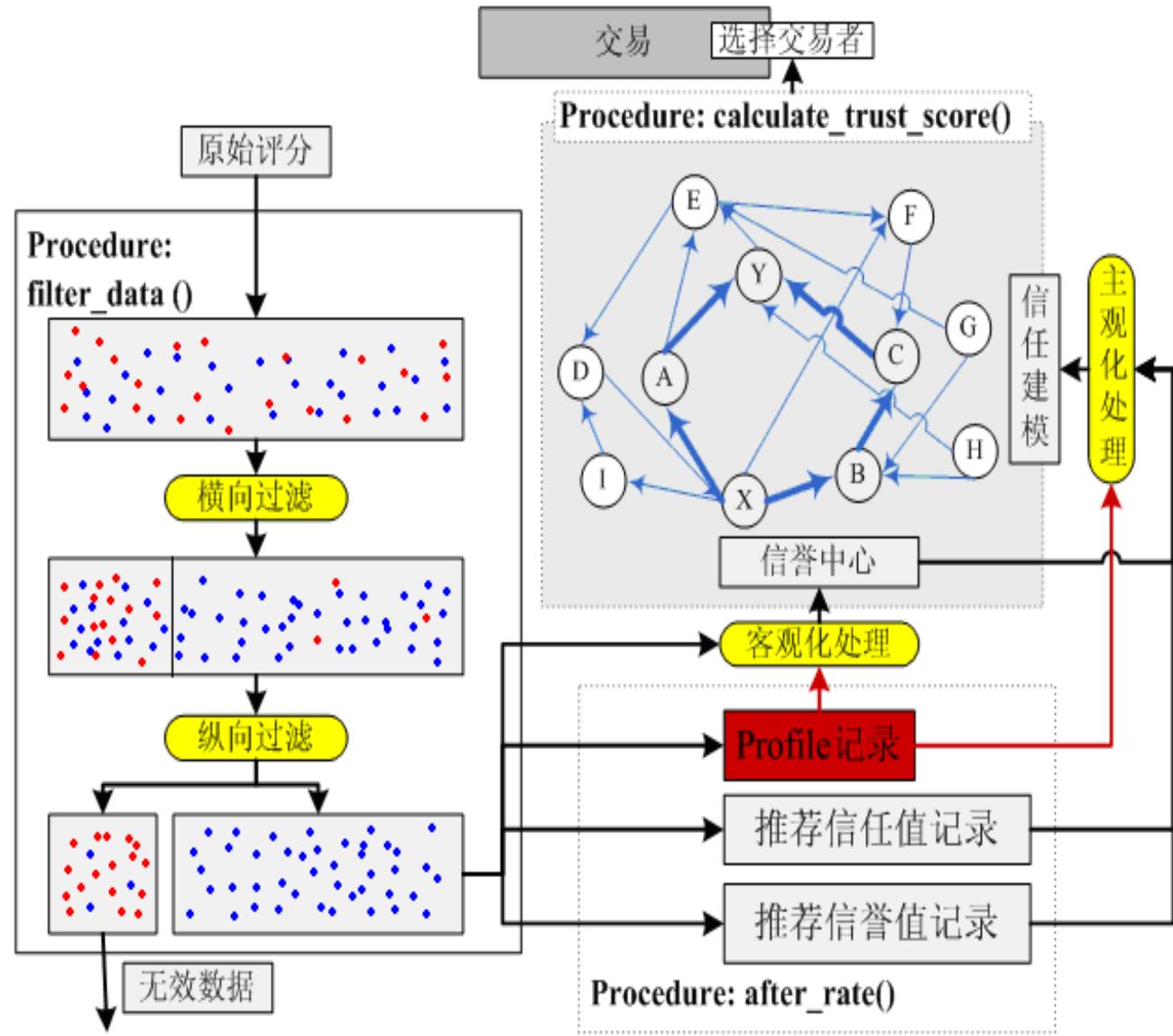
# 面临的问题



# 从用户评论 信息中挖掘 出用户偏好

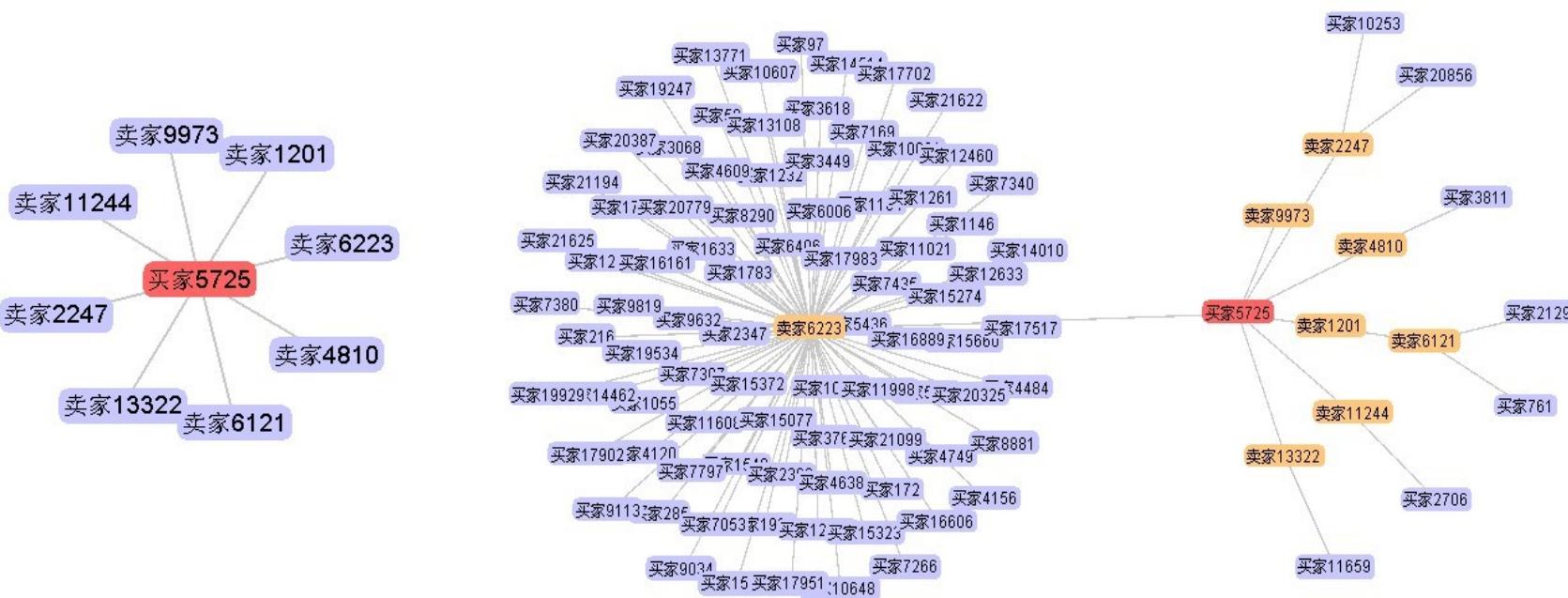


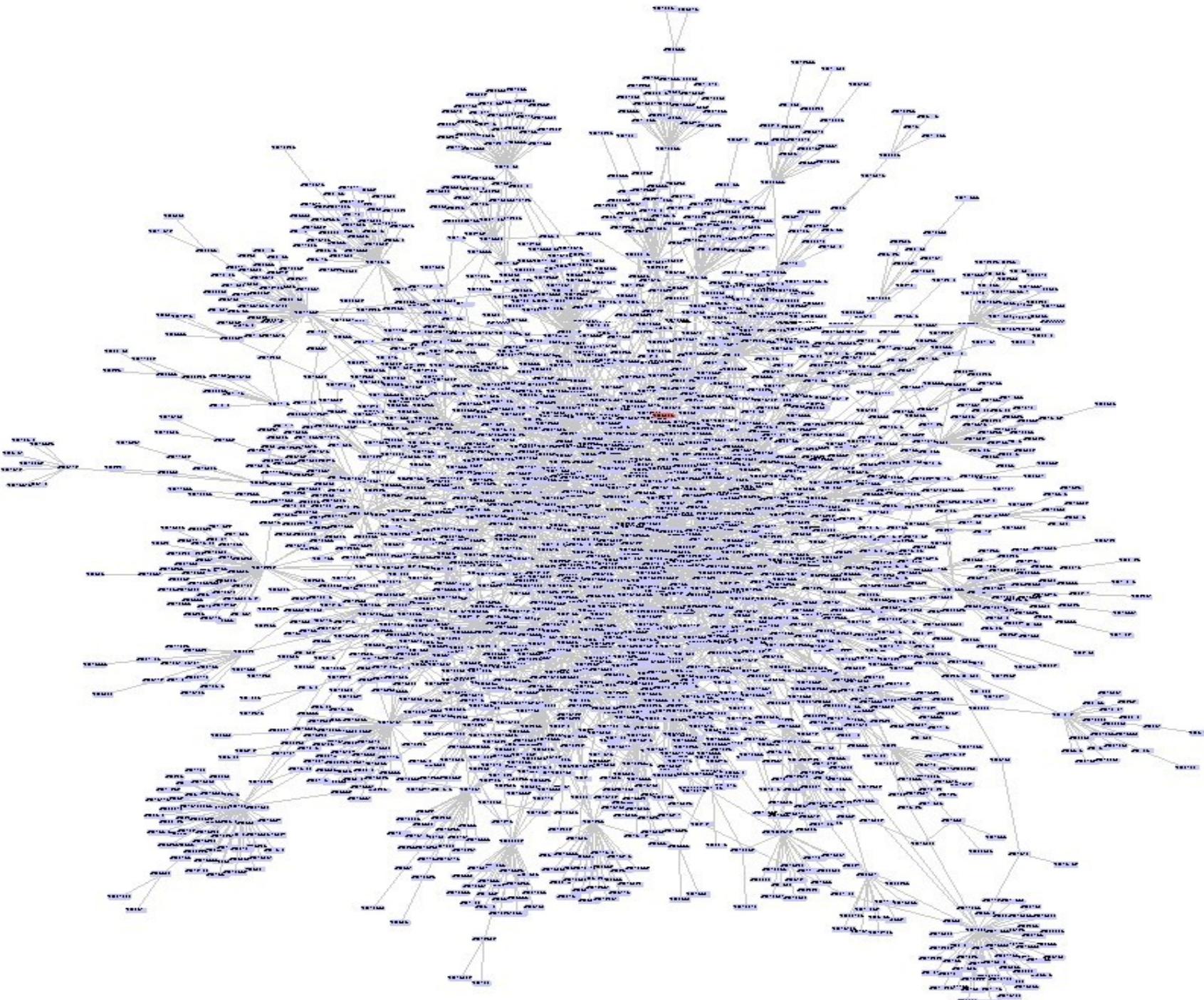
# 构建用户间的可信社会网络



# 淘宝网数据实证分析

- 原始交易数据展现出的社会网络如下左图所示，网络极为稀疏（以买家5725为例）。
- 为了通过与买家5725相似的买家计算更多卖家的信任度，我们首先需要找到与买家5725相似的买家列表。如下图所示：





# 十大机器学习问题

—BY IBM WATSON LAB

- 客户流失(churn)预测
- 情感(sentiment)分析
- 真实性(truth/veracity)分析
- 推荐(recommendation)
- 在线广告
- 新闻聚合
- (算法)扩展性
- 内容发现/搜索
- 智能学习(让机器能像人一样看到、听到)
- 机器学习在医疗中的应用



http://102.alibaba.com/competition/addDiscover

微博综合搜索 - 阿里巴巴 天池...

大赛首页

业电子商务的现状... The Source for Social D... 香港大学推荐的50本经典... CS109 Data Science UC Berkeley Course Le... EssayStar-专业英文论文...

天池  
大数据科研平台

首页

比赛介绍

赛题和数据

排行榜

社区

FAQ

登录

ALIDATA DISCOVERY

# 阿里巴巴大数据竞赛

进入校招直通车 入围“阿里星”

# 最高奖金 百万!

和阿里工程师一起 备战双11 挑战天猫算法

阿里巴巴大数据竞赛  
赛 ) 精彩回顾▶

2013年年底，阿里巴巴大数据竞赛  
赛，参赛选手high翻了，“玩着玩

http://102.alibaba.com/competition/addDiscover

大数据科研平台 比赛介绍-比赛介绍

医药行业电子商务的现状... The Source for Social D... 香港大学推荐的50本经典... CS109 Data Science UC Berkeley Course Le... EssayStar-专业英文论文... 看图

# 天池

大数据科研平台

首页 | 比赛介绍 | 赛题和数据 | 排行榜 | 社区 | FAQ | 登录

## 大赛简介

阿里巴巴大数据竞赛是阿里巴巴集团主办，在阿里巴巴大数据科研平台——“天池”上开展的，基于天猫海量真实用户的访问数据。本次比赛的目的是让广大的高校同学在大数据时代的背景下，有机会学习和利用阿里巴巴的大数据和云计算平台，尝试解决真实的

整个比赛分为三季（ season ）：

Season 1：2014年3月10日至4月20日，参赛者可以下载少量天猫的真实行为数据，并在本地调试算法，提交结果。

Season 2：2014年4月25日至7月30日，参赛者须使用“天池”平台（阿里巴巴自主研发的分布式计算平台），访问海量的天猫数据，使用 Map&Reduce、SQL 及各种平台集成的机器学习算法包调试模型、提交结果。

Season 3（即线上赛，实战双11）：2014年10月至11月，Season 2 的 top 选手可以访问天猫，参与2014年双11的真实算法设计，争夺大奖。

## 报名规则

## 参赛对象

全日制高校（研究院所）本科、硕士、博士均可参赛

## 组织规则

## 赛题介绍

本赛题适用于season1 和season 2的比赛，season 3 为线上赛（ season 2 赛后公布题目）。

在天猫，每天都会有数千万的用户通过品牌发现自己喜欢的商品，品牌是联接消费者与商品最重要的纽带。本届赛题的任务就是根据用户的行为日志，建立用户的品牌偏好，并预测他们在将来一个月内对品牌下商品的购买行为。

我们会开放如下数据类型：

字段	字段说明	提取说明
user_id	用户标记	抽样&字段加密
Time	行为时间	精度到天级别&隐藏年份
action_type	用户对品牌的行为类型	包括点击、购买、加入购物车、收藏4种行为 (点击：0 购买：1 收藏：2 购物车：3 )
brand_id	品牌数字ID	抽样&字段加密

用户对任意商品的行为都会映射为一行数据。其中所有商品ID都已汇总为商品对应的品牌ID。用户和品牌都分别做了一定程度的数据抽样，都做了加密。所有行为的时间都精确到天级别(隐藏年份)。

## 评估指标

## 评估指标

我们希望参赛队的预测，预测的品牌准确率越高越好，也希望覆盖的用户和品牌越多越好，所以用最常用的准确率与召回率作为排行榜的评价指标。

$$\text{准确率: precision} = \frac{\sum_i^N \text{hitBrands}_i}{\sum_i^N \text{pBrands}_i}$$

注：

N 为参赛队预测的用户数

pBrands<sub>i</sub> 为对用户 i 预测他(她)会购买的品牌列表个数

hitBrands<sub>i</sub> 对用户 i 预测的品牌列表与用户 i 真实购买的品牌交集的个数

$$\text{召回率: Recall} = \frac{\sum_i^M \text{hitBrands}_i}{\sum_i^M \text{bBrands}_i}$$

注：

M 为实际产生成交的用户数量

bBrands<sub>i</sub> 为用户 i 真实购买的品牌个数

hitBrands<sub>i</sub> 预测的品牌列表与用户 i 真实购买的品牌交集的个数

最后我们用 F1-Score 来拟合准确率与召回率，并且大赛最终的比赛成绩排名以 F1 得分为准。

$$F_1 = \frac{2 * P * R}{P + R}$$

## 不同赛季使用数据说明

### Season 1 赛季数据



首 页

产 品

行 业

案 例

大 数据 时 代

证 书 查 询

关 于 我 们



# 最权威的 个性化推荐引擎 技术平台

MOST AUTHORITATIVE PERSONALIZED RECOMMENDATION  
ENGINE TECHNOLOGY PLATFORM

超越 算法、模型的技术构架  
实时 响应, 抓住第一商机



▶ 个性化推荐引擎  
PERSONALIZED RECOMMENDATION ENGINE

▶ 快消品案例  
FAST MOVING CONSUMPTION PROPOSAL

▶ 百分点智库  
BAIFENDIAN LIBRARY



# 2016 CCF大数据与计算智能大赛

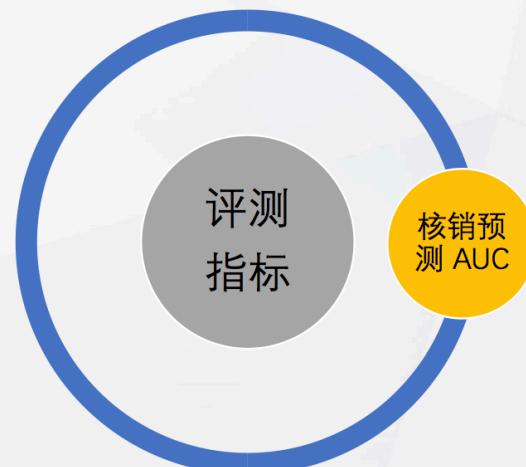


用户行为预测

机器学习

推荐系统

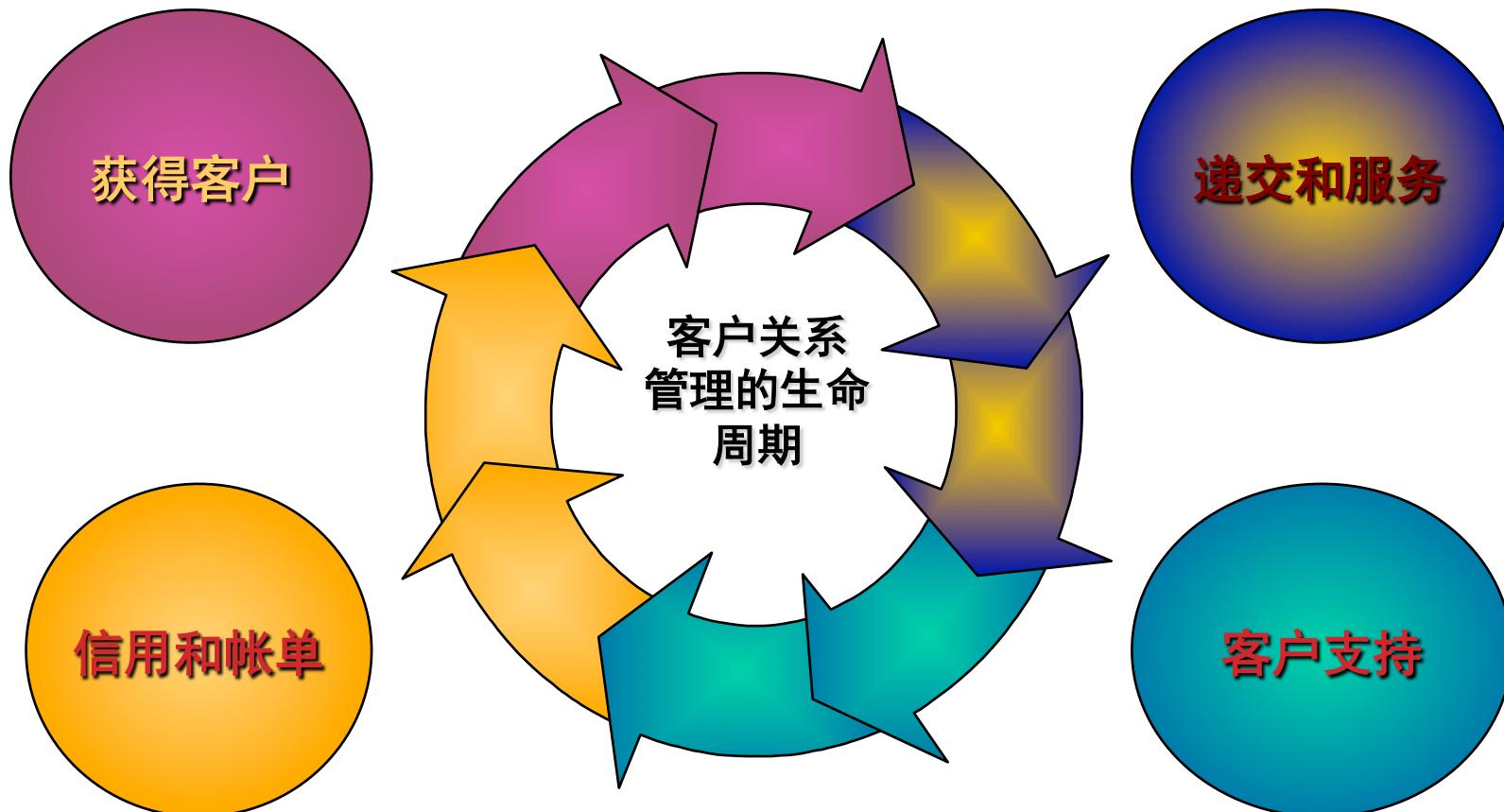
用户画像



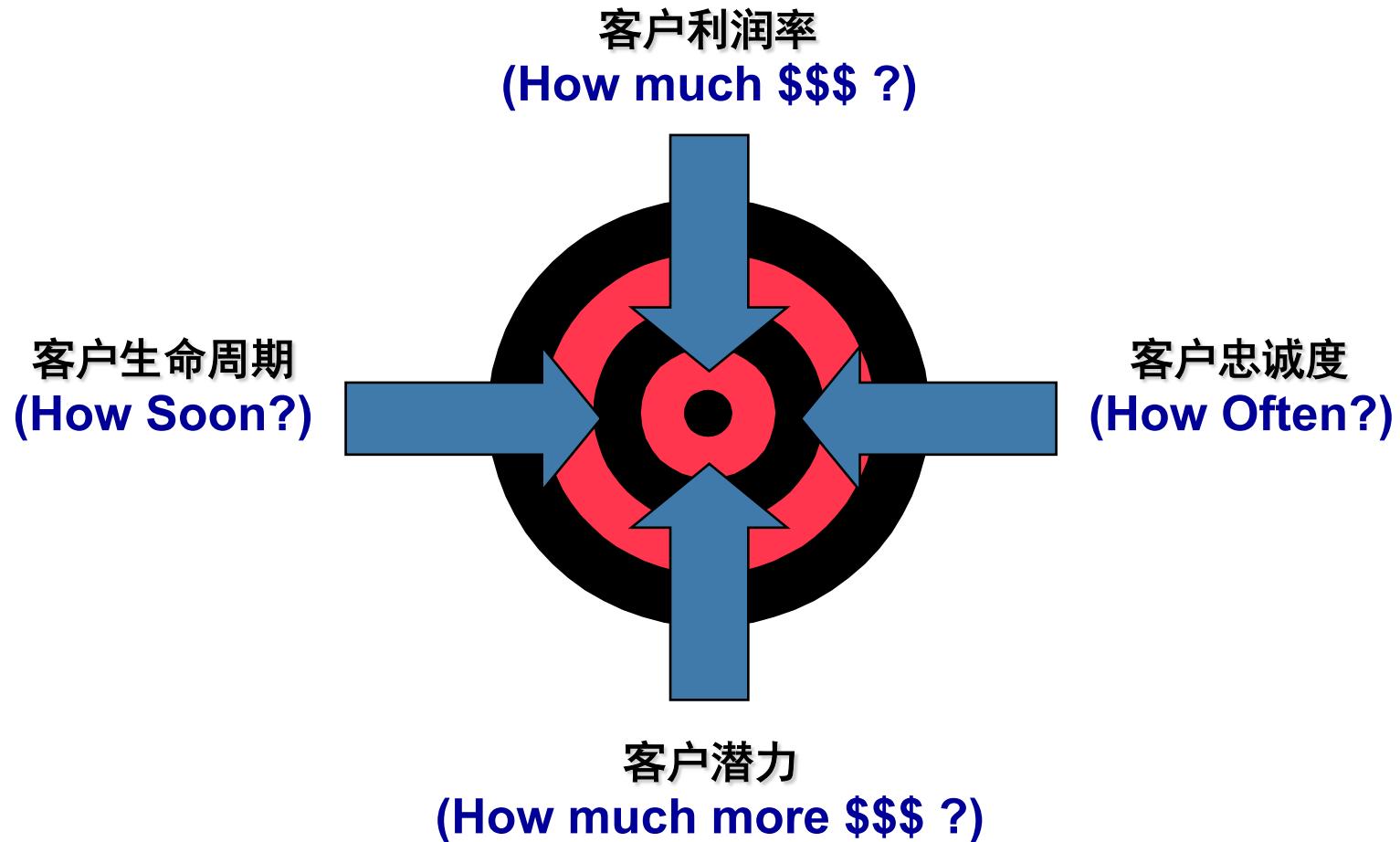
## 7.3 客户关系管理

- 客户关系管理中的“接触点”
- 客户关系的衡量
- 客户关系管理的定义
- 客户关系管理的内涵
- 客户关系管理中的常用术语
- 客户关系生命周期

# (1) 客户关系管理中的“接触点”



## (2) 我们如何衡量客户关系



### (3) 客户关系管理的定义

- ① **Gartnet Group**认为：客户关系管理就是为企业提供全方位的管理视角，赋予企业更完善的客户交流能力，最大化客户的收益率。
- ② **Carlson Marketing Group**把客户关系管理定义为：通过培养公司的每一个员工，经销商或客户对该公司更积极的偏爱或偏好，留住他们并以此提高公司业绩的一种营销策略。

# 客户关系管理的定义（续）

- ③ **Hurwitz group**（霍尔维茨组织）认为：客户关系管理的焦点是自动化并改善与销售、市场营销、客户服务和支持等领域的客户关系有关的商业流程。
- ④ **IBM**：客户关系管理包括企业识别、挑选、获取、发展和保持客户的整个商业过程。**IBM**把客户关系管理分为三类：关系管理、流程管理和接入管理。

## (4) 客户关系管理的内涵



- 理念是**CRM**成功的关键，是**CRM**实施应用的基础和土壤；
- 信息系统、**IT**技术是**CRM**成功实施的手段和方法；
- 实施是决定**CRM**成功与否、效果如何的直接因素。

## ( 5 ) 客户关系管理中的术语

1对1市场

客户潜力

市场反映时间

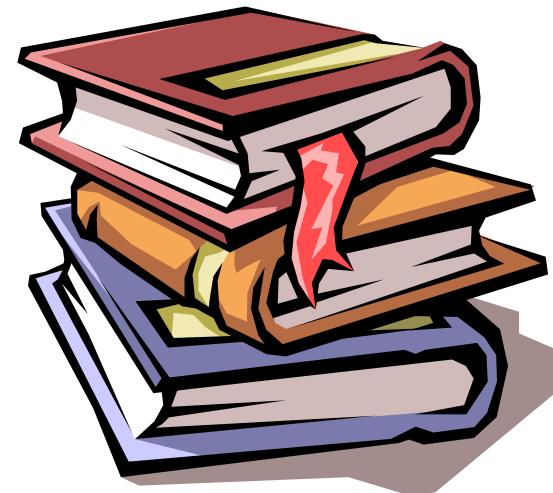
风险管理

价值转换

非居间化 (**Disintermediation**)

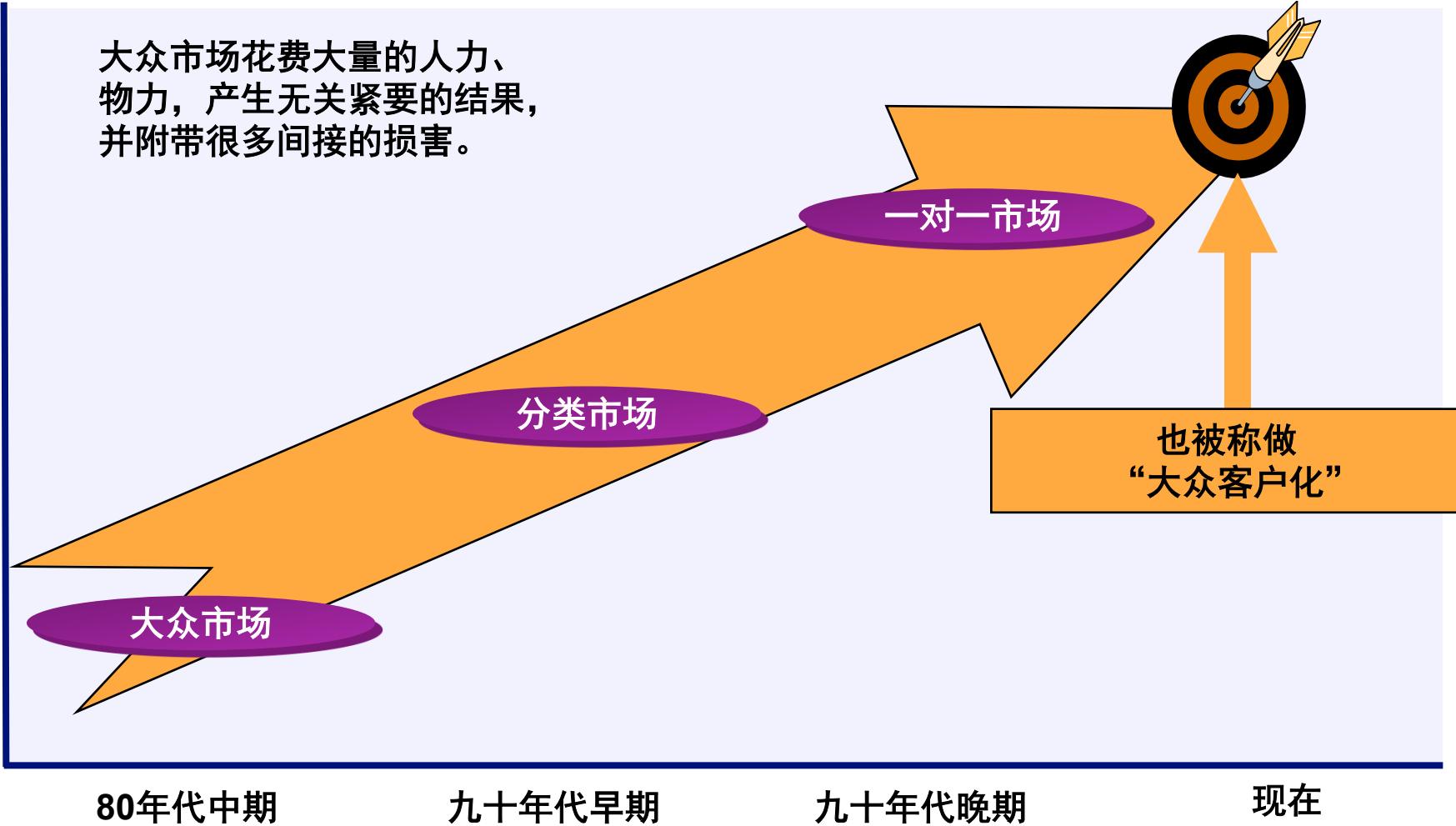
面向事件的探测

客户描述



# 一对一市场

大众市场花费大量的人力、  
物力，产生无关紧要的结果，  
并附带很多间接的损害。



# 客户潜力

“未开发的潜力” = 利润

利润率的关键

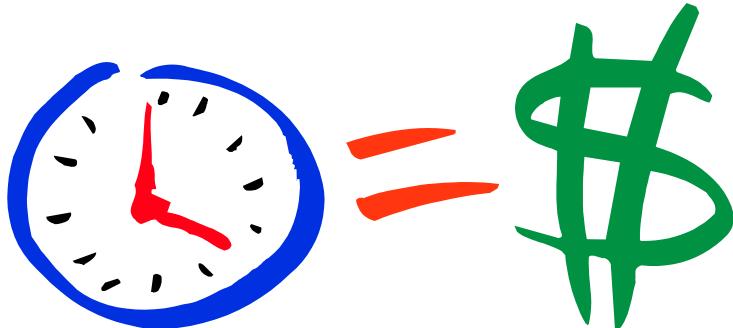
- 为提高市场份额的战斗是非常昂贵、冲突性极强的业务策略
- 增强已占有的客户市场，成本较低、利润更高和竞争的冲突性较不明显



# 对市场的响应时间

## 发布速度至关重要

- 了解你的客户使得你能用最快的速度响应他们的需求....
- 而且，让你在竞争中处于领先地位



# 风险管理

了解你客户的购买模式帮助你：

- 发现欺诈行为
- 改善客户服务
- 弥补产品的不足



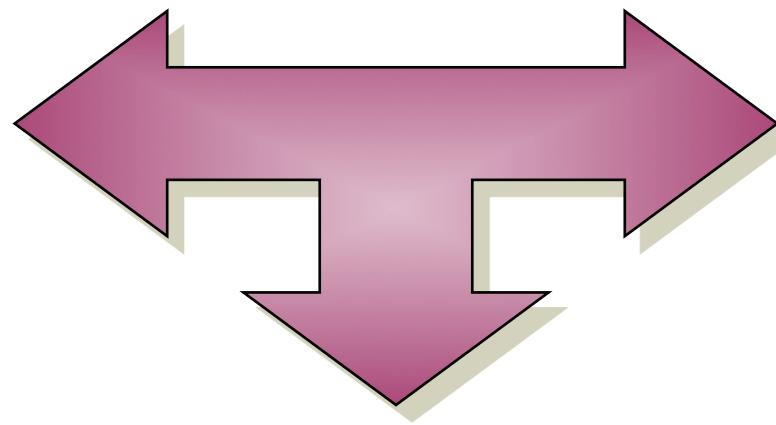
同时，了解客户帮助你：

- 辨别和巩固你的风险最小，利润率最高的客户

# 价值转换

## 广度 (获得客户)

- 确定新的客户分类
- 找到隐藏的市场
- 获得新的客户



## 深度 (潜伏性)

- 向已有的客户销售更多的东西
  - 交叉销售 / 升级销售
- 了解购买倾向

## 长度 (忠诚度)

- 维持较久的客户关系
- 降低流失
- 提高转换成本
- 提高可见的价值

# 非居间化

(DISINTERMEDIATION = “CUTTING OUT THE MIDDLEMAN”)

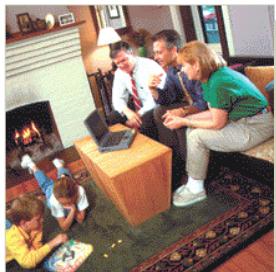
截取掉中间部分...

客户

经纪人

生产商

销售商



消除中间环节让你：

- 降低渠道成本
- 增加渠道利润率

Internet, Direct Mail, Call Centers etc.



# 面向事件的探测

## EVENT-ORIENTED PROSPECTING



- 面向事件探测（EOP）的关键是了解客户生活中可能与你的业务有联系的事件，并马上采取有针对性的市场措施响应之。
  - 地址的变化
  - 结婚
  - 换工作
  - 小孩的出生
  - 退休

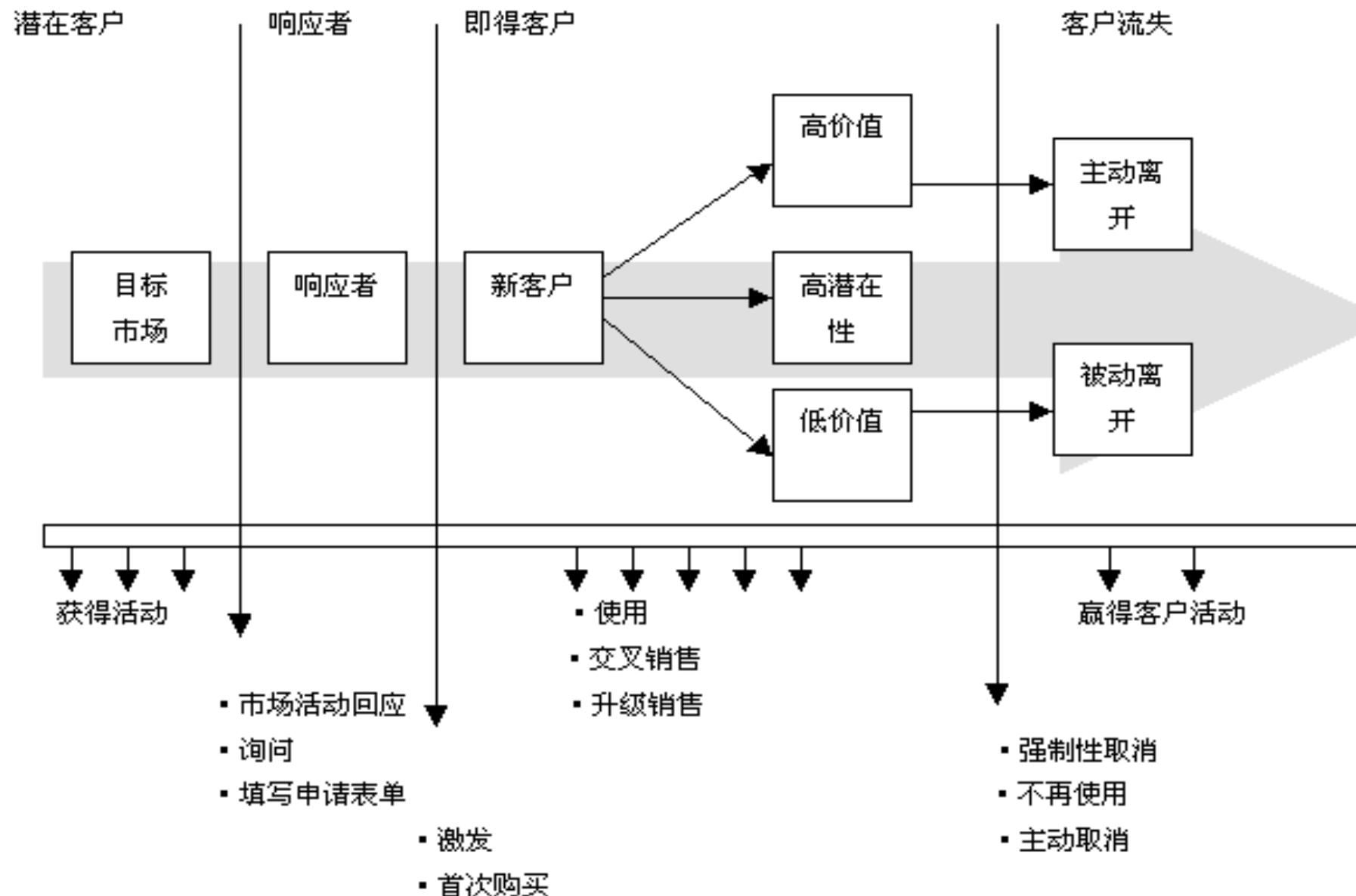
# 客户描述



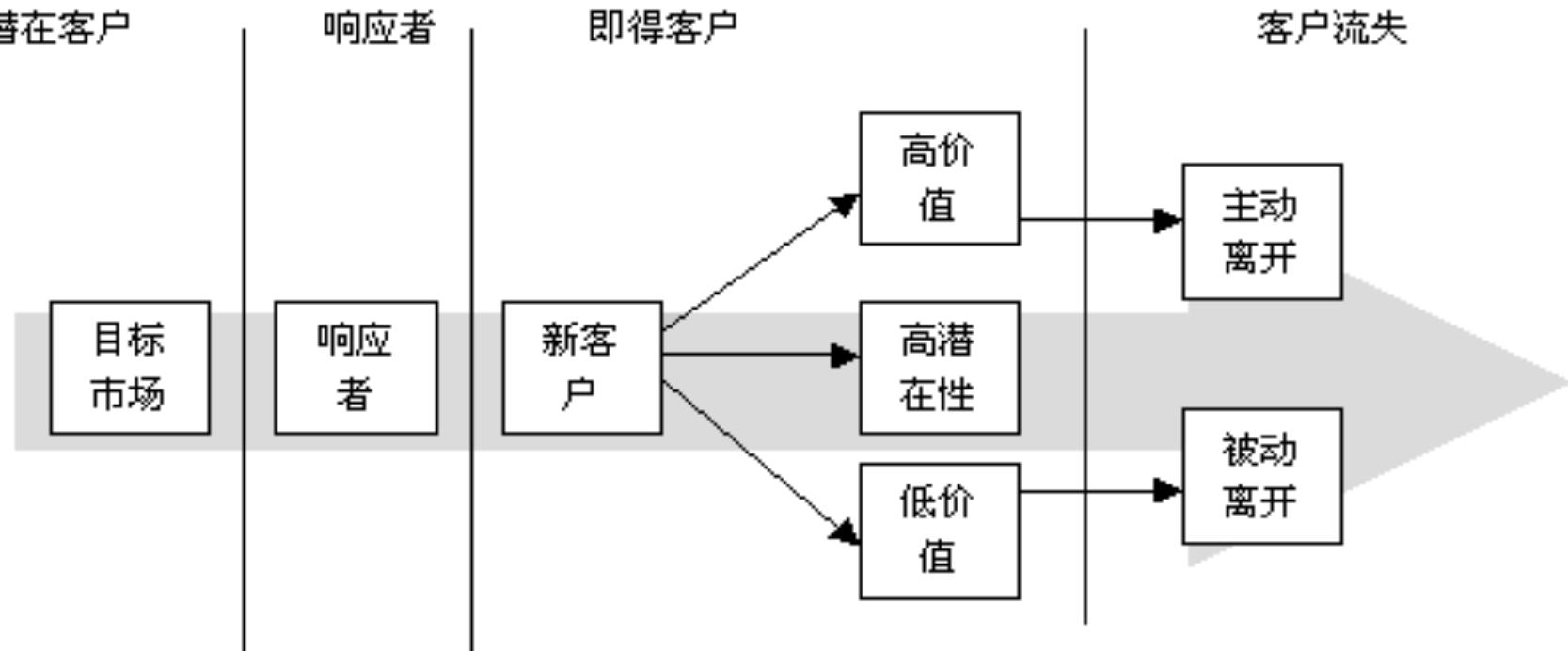
完整客户描述的10大要素：

- **1.** 客户的外界关系
- **2.** 客户行为
- **3.** 客户的人口统计学数据
- **4.** 家庭成员和亲戚关系
- **5.** 客户风险
- **6.** 嗜好
- **7.** 购买倾向
- **8.** 客户收益率
- **9.** 生活事件
- **10.** 客户满意度级别

# (6) 客户生命周期分析



# 不同客户生命周期阶段出现的数据



- 活动历史
- 购买统计
- 市场活动回应

- 信用报告
- 自动报告的信息
- 其它

- 产品使用
- 付帐历史
- 活动响应
- 渠道偏爱
- 其它

- 终止原因

# 典型应用：移动用户管理

- 通过电信数据的多维分析，识别和比较数据通信情况、系统负载、资源使用、用户组行为、利润等；
- 通过多维分析、聚类分析和孤立点分析进行盗用模式分析和异常模式识别；
- 通过多维关联和序列模式分析进行电信服务组合和个性化服务；

## 集团客户管理

## 集团客户资料管理

-  新增集团客户
-  集团客户管理
-  集团合并
-  非在线集团管理
-  集团关键人物管理
-  集团统计信息
-  集团成员管理
-  批量修改管理
-  集团合同管理
-  语音专线业务记录
-  集团资料库管理

按集团客户统计    按客户经理统计

集团编号

起始月份  [YYYYMM]

结束月份  [YYYYMM]

查询

第1/1页, 共10/10条

« < 1 > »

每页 600 条, 跳至  页 Go

## 集团客户话费结构统计

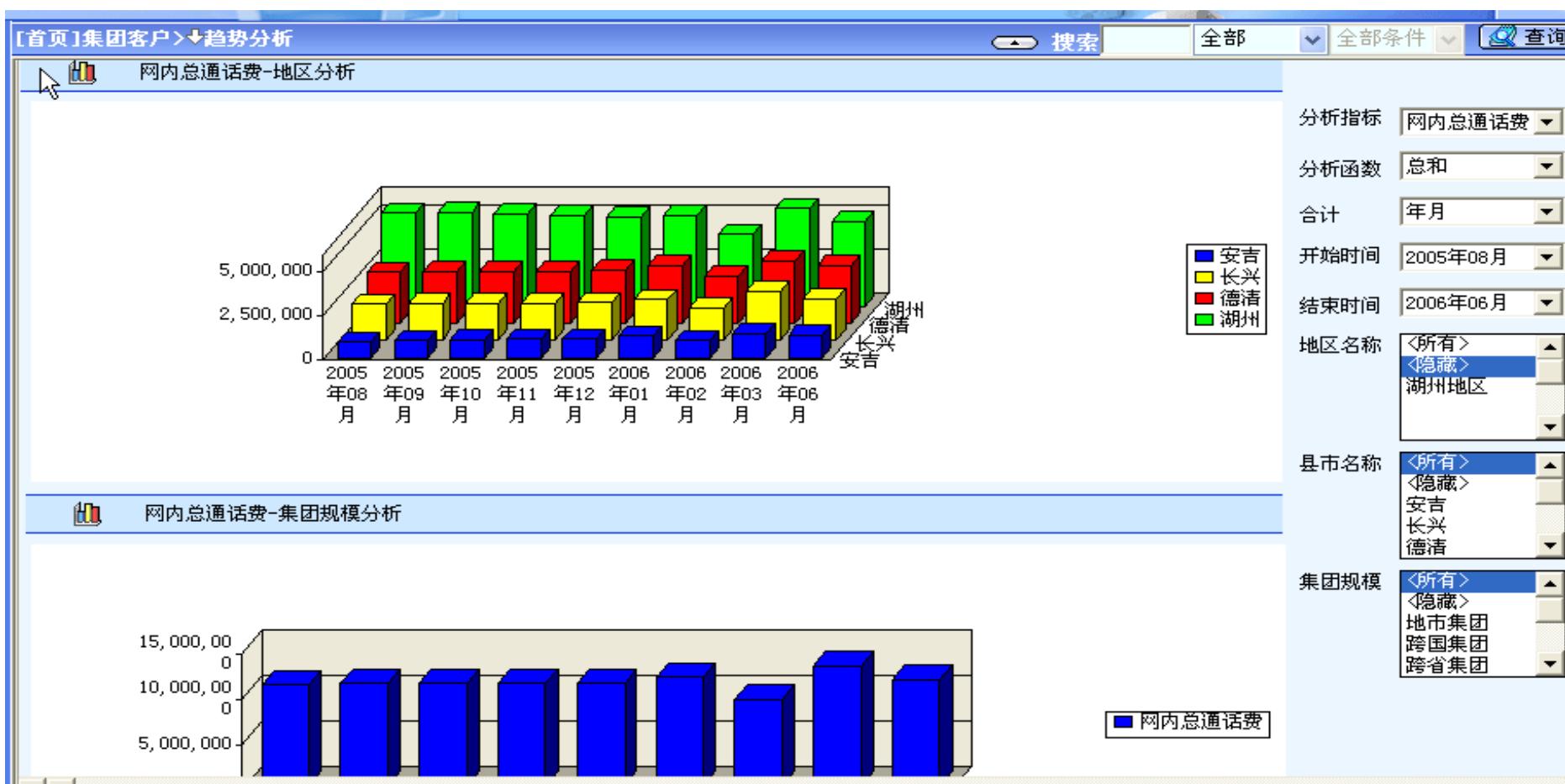
## 集团大客户统计

总话费	网内总话费	网外总话费	基本通话费	国内长途费	国际长途费	漫游通话费	漫游国内长途费	漫游国际长途费	大客户数	大客户总话费	IP费
6049.74	2936.94	3657.78	4808.49	946.43	0.00	839.80	465.57	0.00	17	3057.77	0.0
6503.71	2854.64	3515.23	4060.60	1238.12	0.00	1071.15	618.80	0.00	15	3388.00	0.0
6531.86	3754.14	4019.20	5664.20	1310.60	0.00	798.54	450.64	0.00	15	3577.74	0.0
5944.76	5401.94	5306.62	7301.59	1962.02	0.00	1444.95	783.98	0.00	22	4688.53	0.0
9923.07	9631.45	0.00	5483.34	3757.01	0.00	391.10	1457.92	0.00	17	4232.64	0.0
5277.54	19715.71	0.00	10156.60	8851.11	0.00	708.00	3353.29	0.00	21	5719.07	0.0
4631.23	18891.93	0.00	9137.25	8970.68	18.00	766.00	4862.01	0.00	20	5630.02	0.0
4466.24	18276.70	0.00	8329.30	9056.20	4.00	887.20	5367.01	0.00	19	5690.29	0.0
4311.43	18875.99	0.00	7484.50	9908.59	203.00	1279.90	7017.27	33.60	18	5289.90	0.0
4896.82	21888.87	0.00	8017.70	11950.52	137.00	1783.65	8469.03	99.50	18	5541.51	0.0

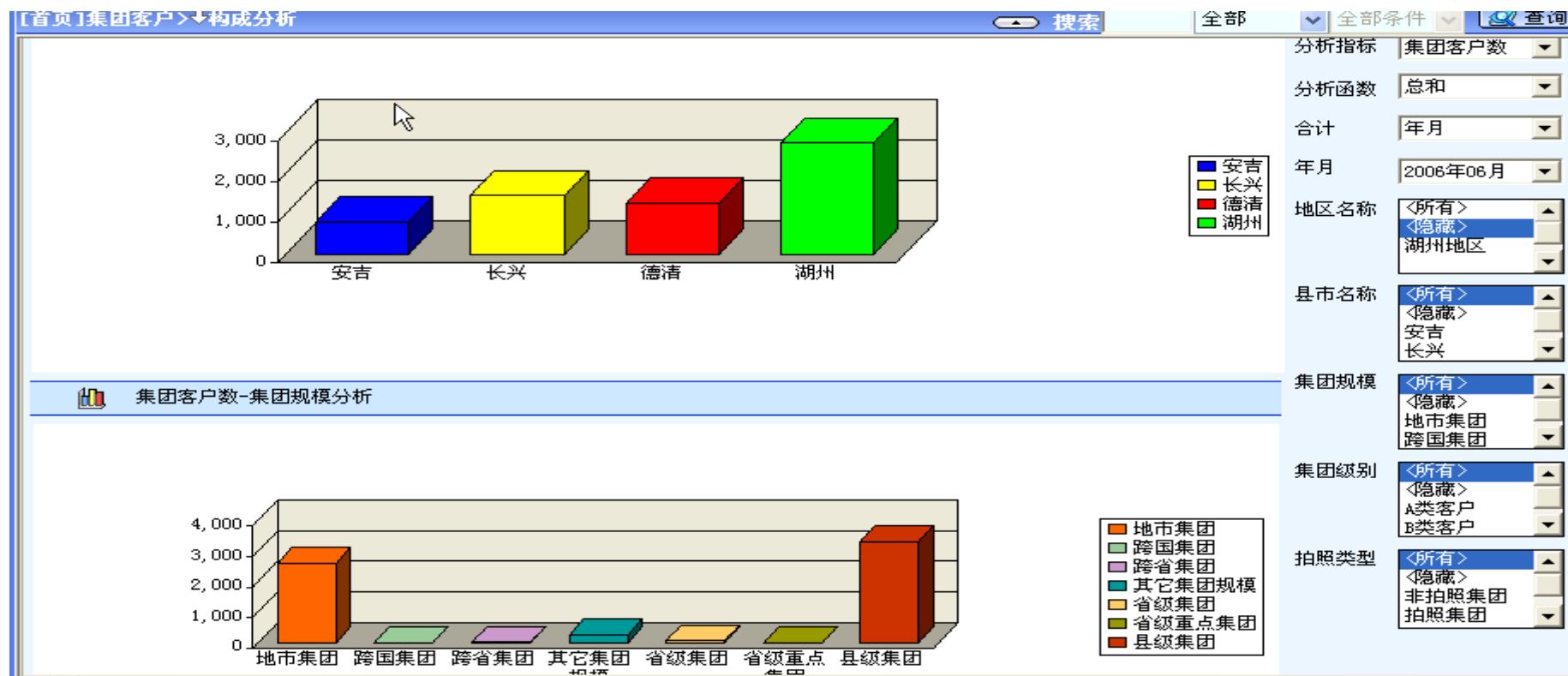
导出

- 大客户管理
- 日常工作
- 客户服务管理
- 客户经理管理
- 统计报表
- 行业应用码号管理
- 系统管理

# 集团客户收入分析



# 集团客户构成分析



# 套餐分析

-  终端捆绑业务
-  资费预演
-  套餐分析
-  彩铃分析
-  高网分析
-  重入网
-  元数据
-  报表

品牌：  请输入通话用户数：

综合实力套餐：ARPU/综合单价/用户保持率 降序排列  
 发展实力套餐：通话用户数/用户增长率/用户保持率 降序排列  
 问题套餐：通话用户数/用户保持率/综合单价 升序排列

虚拟网外单价= (出帐收入 - 虚拟网收入) / (通话时长分钟 - 虚拟网时长分钟);  
用户保持率=本月套餐保持用户数 / (保持用户数+流失用户数)  
对新入网用户吸引力=新入网用户数 / (新入网用户数+转网新增用户数)  
对转网用户的吸引力=网内用户转入新增 / (网外用户新增+网内用户转入新增)

## 套餐列表

套餐名称	品牌	通话用户数	通话用户数变化	用户保持率(%)	用户增长率(%)	对新入网用户吸引	对转网用户的吸引	综合单价(元)	综合单价
10071036 _全球通话费150全球通	全球通	10971	0.13	95.64	4.63	16.53	83.46	0.26	
10111001 _全球通	全球通	31770	0.20	90.56	8.94	59.67	40.32	0.33	
10071046 _全球通话费100全球通	全球通	22276	-0.60	96.03	3.29	25.20	74.80	0.24	
10071159 _全球通手机100全球通	全球通	11237	0.47	95.39	5.04	17.86	82.13	0.23	
10070244 _神州行畅听卡2神州行本卡	神州行	33519	1.41	95.91	5.54	40.58	59.41	0.22	
10071513 _全球通自由套餐全球通	全球通	18569	-2.97	97.39	0.01	100.00	0.00	0.28	
10010069 _动感地带龙卷风动感地带	动感地带	155412	2.71	93.12	10.19	76.17	23.82	0.31	
10010014 _神州行畅听卡1神州行本卡	神州行	148310	2.63	96.28	7.25	77.09	22.90	0.32	
10070835 _神州行长话卡	金卡神州行	58172	6.25	95.11	12.36	99.97	0.02	0.54	



套餐盈利能力分析										
月份	套餐编号	区域	本月变更用户数	本月网外新增用户数	本月流失用户数	在网用户数	用户保持率	对新增用户吸引力	对网内其它套餐用户吸引力	ARPU变化
1	10070253		0.	0.	1.	77.	99%	0%	0%	-5%
1	10070253		0.	0.	1.	23.	96%	0%	0%	20%
1	10070428		0.	0.	185.	4,579.	96%	0%	0%	14%
1	10070428		0.	0.	165.	3,341.	95%	0%	0%	20%
1	10070797		0.	0.	2.	5.	71%	0%	0%	-100%
1	10070835		3.	476.	166.	2,240.	91%	99%	1%	10%
1	10070835		3.	6,677.	1,032.	23,983.	94%	100%	0%	10%
2	10070253		0.	0.	0.	5.	100%	0%	0%	-31%
2	10070253		0.	0.	0.	23.	100%	0%	0%	22%
2	10070428		0.	0.	69.	1,855.	96%	0%	0%	-11%
2	10070428		0.	0.	113.	3,227.	97%	0%	0%	-9%
2	10070797		0.	0.	2.	3.	60%	0%	0%	0%
2	10070835		3.	445.	111.	2,579.	95%	99%	1%	-4%
2	10070835		0.	6,438.	992.	29,464.	96%	100%	0%	-9%
3	10070253		0.	0.	3.	72.	96%	0%	0%	-10%
3	10070253		0.	0.	1.	4.	80%	0%	0%	-23%
3	10070428		0.	0.	155.	4,271.	96%	0%	0%	-1%
3	10070428		0.	0.	162.	3,067.	95%	0%	0%	-4%
3	10070797		0.	0.	0.	1.	100%	0%	0%	-100%
3	10070797		0.	0.	0.	3.	100%	0%	0%	-100%
3	10070835		2.	290.	186.	2,686.	93%	99%	1%	2%
3	10070835		0.	6,376.	1,697.	34,181.	94%	100%	0%	6%

谢 谢 !