



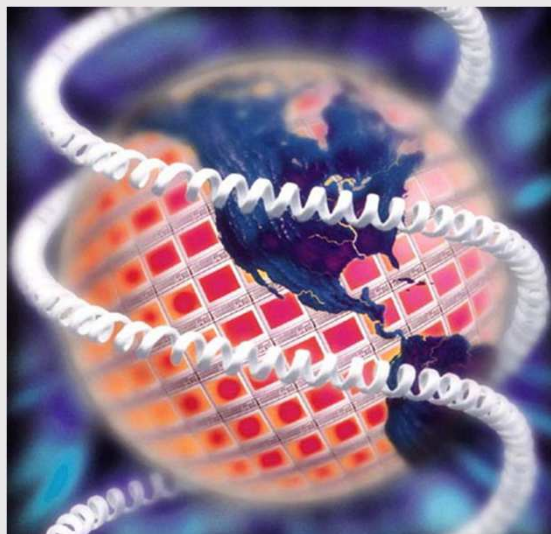
# 大数据与电子商务

哈尔滨工业大学计算机科学与技术学院

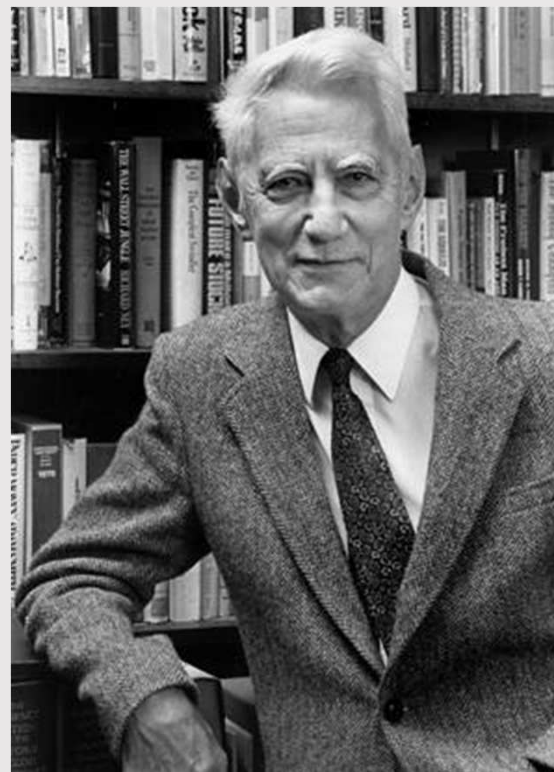
王轩

# 人类社会的发展



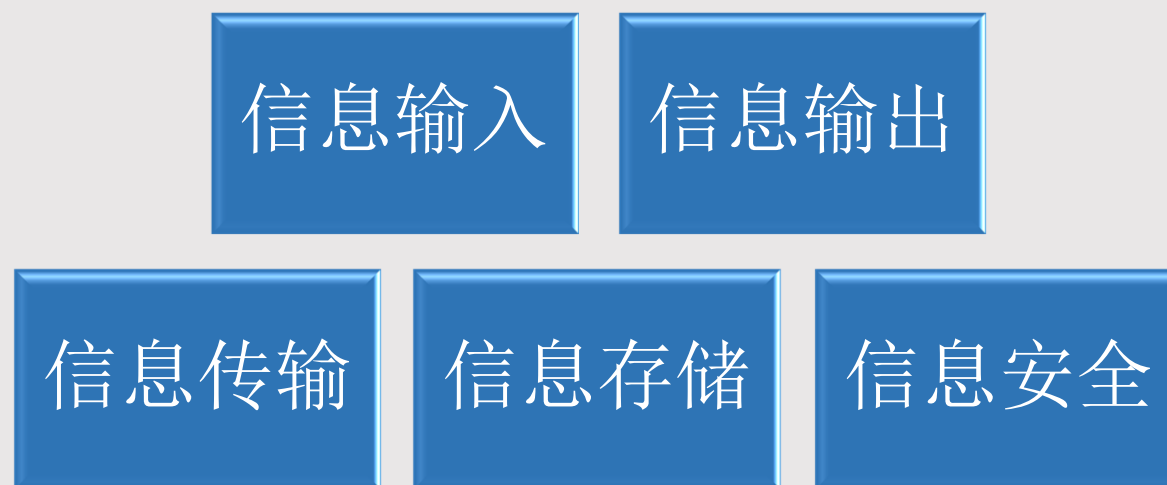


信息泛指人类社会传播的一切内容。人通过获得、识别自然界和社会的不同信息来区别不同事物，得以认识和改造世界。在一切通讯和控制系统中，信息是一种普遍联系的形式。



克劳德·艾尔伍德·香农（Claude Elwood Shannon，1916年4月30日—2001年2月26日）美国数学家、信息论的创始人。

# 信息技术的范畴



信息技术（英语：Information Technology，缩写：IT），是人类在生产斗争和科学实验中认识自然和改造自然过程中所积累起来的获取信息，传递信息，存储信息，处理信息以及使信息标准化的经验、知识、技能和体现这些经验、知识、技能的劳动资料有目的的结合过程。

## 常见的输入设备

字符输入设备：键盘 光学阅读设备：光学标记阅读机 图形输入设备：鼠标、操纵杆



模拟输入设备：语言模数转换识别系统 图像输入设备：摄像机、扫描仪

# 句级智能汉字输入



I

以语句为单位的汉字输入方法，汉字只需要输入部分信息，缺省信息产生的歧义由计算机根据语句上下文自动处理。

## 常见的输出设备



显示器



打印机



绘图仪



影像输出系统



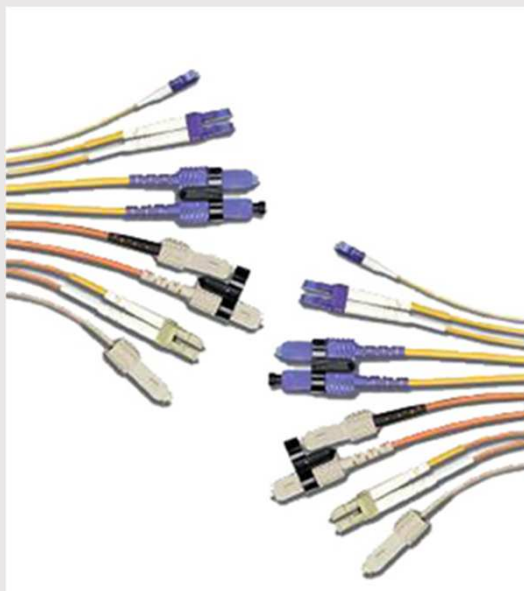
语音输出系统



磁记录设备

# 信息传输

信息传输是从一端将命令或状态信息经信道传送到另一端，并被对方所接收。包括传送和接收。传输介质分有线和无线两种，有线为电话线或专用电缆；无线是利用电台、微波及卫星技术等。



有线传输



无线传输

## 常见的存储介质



纸张



磁带



磁盘



光盘



记忆卡



闪存盘

## 信息安全的范围

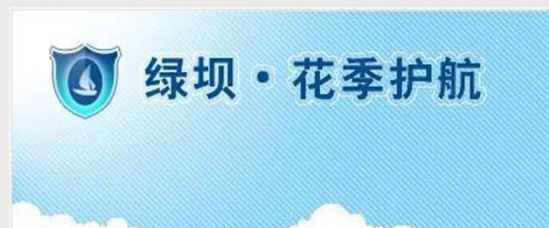
保护国家军事政治等机密安全



防范商业企业机密泄露



防范青少年对不良信息的浏览



个人信息的泄露



# 什么是数字技术

数字技术（Digital Technology）是一项与信息技术相伴相生的科学技术，它是指借助一定的设备将各种信息，包括图、文、声、像等转化为电子计算机能识别的二进制数字“0”和“1”后进行运算、加工、存储、传送、传播、还原的技术。

由于在运算、存储等环节中要借助计算机对信息进行编码、压缩、解码等，因此也称为数码技术、计算机数字技术等。



# 字技术中的数制

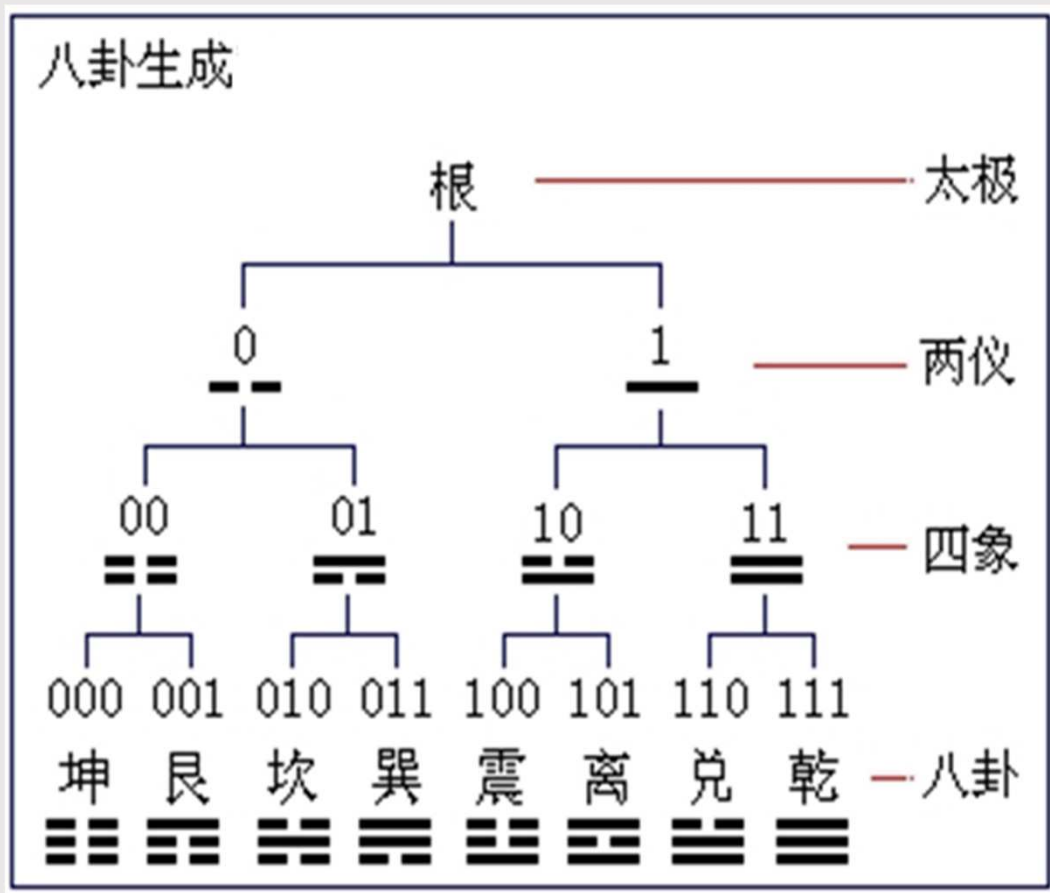
十进制D(decimal)日常生活中最熟悉的进位计数制。在十进制中，数用0, 1, 2, 3, 4, 5, 6, 7, 8, 9这十个符号来描述。计数规则是逢十进一。

二进制B(binary)在计算机系统中采用的进位计数制。在二进制中，数用0和1两个符号来描述。计数规则是逢二进一。

十六进制H(hexadecimal)在计算机指令代码和数据的书写中经常使用的数制。在十六进制中，数用0, 1, ..., 9和A,B,..., F (或a, b, ..., f) 16个符号来描述。计数规则是逢十六进一。

八进制O(octal)在早期的计算机系统中很常见，采用0, 1, 2, 3, 4, 5, 6, 7八个数字，逢八进位，并且开头一定要以数字0开头。

# 进制与易经



太极生两仪，两仪生四象，四象生八卦，这正是二进制数的体现。

# 进制的优点

可行性：电子元件易实现

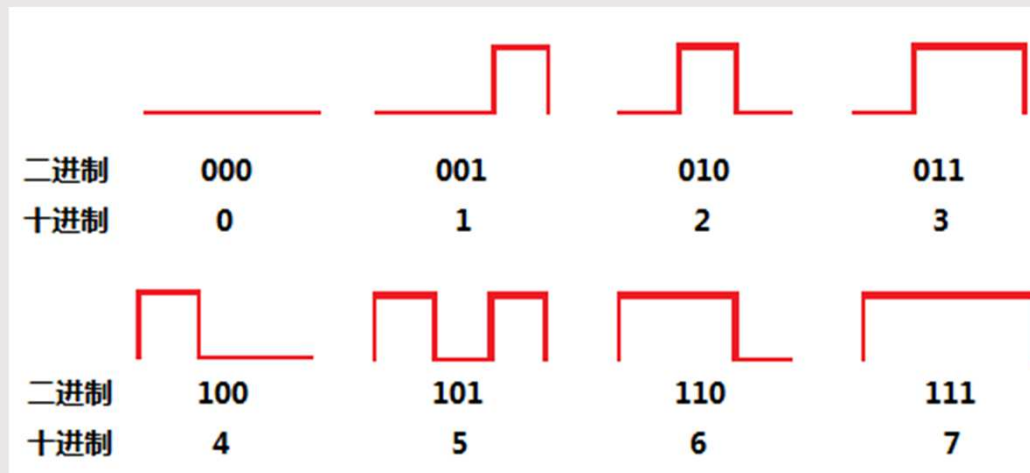
可靠性：状态稳定

简易性：算法简单

求和法则  $0+0=0$  ,  $0+1=1+0=1$  ,  $1+1=10$

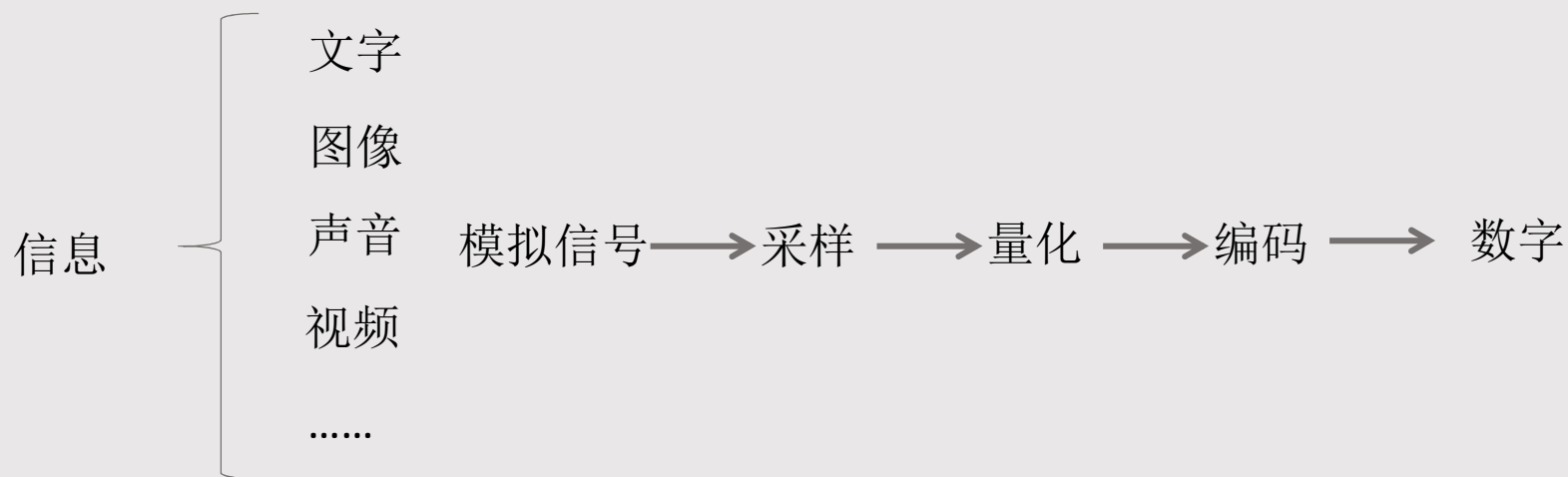
求积法则  $0*0=0$  ,  $0*1=1*0=0$  ,  $1*1=1$

逻辑性：可用进制的0, 1直接代表逻辑代数中的“假”和“真”。



# 信息的数字化

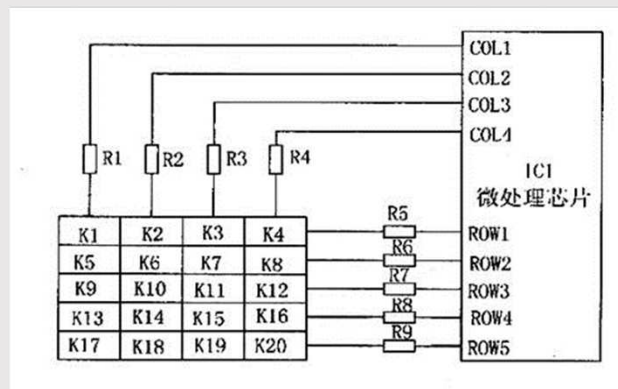
数字化就是将许多复杂多变的信息转变为可以度量的数字、数据，再以这些数字、数据建立起适当的数字化模型，把它们转变为一系列二进制代码，引入计算机内部，进行统一处理。



# 数字化



文字从键盘输入

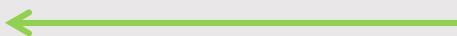


芯片处理



01100001

编码



a

字符

ASCII是“美国信息交换标准码”的英文缩写，用来制订计算机中每个符号对应的代码，这也叫做计算机的内码(code)。

一个字节的低7位表示，共有128个编码。

如：A ----- 65 (1000001)

a ----- 97 (1100001)

汉字字符的编码技术相对复杂。具体有GB2312汉字编码标准，简称国标码。GB2312收录简化汉字及符号、字母、日文假名等共7445个图形字符，每个汉字占两个字节，共6763个。

区码+位码：一个汉字用两个字节存储。

“啊” ----- 16 01

# 2312码表和ASCII表

code	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
B0A0		啊	阿	埃	挨	哎	唉	哀	皑	癌	蔼	矮	艾	碍	爱	隘
B0B0	鞍	氨	安	俺	按	暗	岸	胺	案	肮	昂	盎	凹	敖	熬	翱
B0C0	袄	傲	奥	懊	澳	芭	捌	扒	叭	吧	笆	八	疤	巴	拔	跋
B0D0	靶	把	耙	坝	霸	罢	爸	白	柏	百	摆	佰	败	拜	裨	斑
B0E0	班	搬	扳	般	颁	板	版	扮	拌	伴	瓣	半	办	绊	邦	帮
B0F0	梆	榜	膀	绑	棒	磅	蚌	镑	傍	谤	苞	胞	包	褒	剥	

.....

code	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
D7A0		住	注	祝	驻	抓	爪	拽	专	砖	转	撰	赚	篆	桩	庄
D7B0	装	妆	撞	壮	状	椎	锥	追	赘	坠	缀	淳	准	捉	拙	卓
D7C0	桌	琢	茁	酌	啄	着	灼	浊	兹	咨	资	姿	滋	淄	孜	紫
D7D0	仔	籽	滓	子	自	渍	字	鬃	棕	踪	宗	综	总	纵	邹	走
D7E0	奏	揍	租	足	卒	族	祖	诅	阻	组	钻	篡	嘴	醉	最	罪
D7F0	尊	遵	昨	左	佐	柞	做	作	坐	座	□	□	□	□		

code	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
D8A0		于	丌	兀	丐	廿	卅	丕	亘	丞	鬲	孛	噩	丨	禺	丿
D8B0	匕	乇	夭	爻	厄	氏	凶	胤	虺	鞞	鞞	龔	龔	、	亟	龔
D8C0	乚	兀	半	孛	嗇	仄	厓	厓	厓	厓	厓	厓	厓	厓	厓	厓
D8D0	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚	匚
D8E0	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌	劌
D8F0	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞	仞

ASCII值	控制字符	ASCII值	控制字符	ASCII值	控制字符	ASCII值	控制字符
0	NUL	32	(space)	64	@	96	`
1	SOH	33	!	65	A	97	a
2	STX	34	"	66	B	98	b
3	ETX	35	#	67	C	99	c
4	EOT	36	\$	68	D	100	d
5	ENQ	37	%	69	E	101	e
6	ACK	38	&	70	F	102	f
7	BEL	39	,	71	G	103	g
8	BS	40	(	72	H	104	h
9	HT	41	)	73	I	105	i
10	LF	42	*	74	J	106	j
11	VT	43	+	75	K	107	k
12	FF	44	,	76	L	108	l
13	CR	45	-	77	M	109	m
14	SO	46	.	78	N	110	n
15	SI	47	/	79	O	111	o
16	DLE	48	0	80	P	112	p
17	DC1	49	1	81	Q	113	q
18	DC2	50	2	82	R	114	r
19	DC3	51	3	83	X	115	s
20	DC4	52	4	84	T	116	t
21	NAK	53	5	85	U	117	u
22	SYN	54	6	86	V	118	v
23	TB	55	7	87	W	119	w

# 片数字化

将图像上每个点的颜色信息按某种规律编成一系列二进制编码，即用0和1表示图像信息的过程就是图像数字化。

非数字化图像



采样



数字化图像

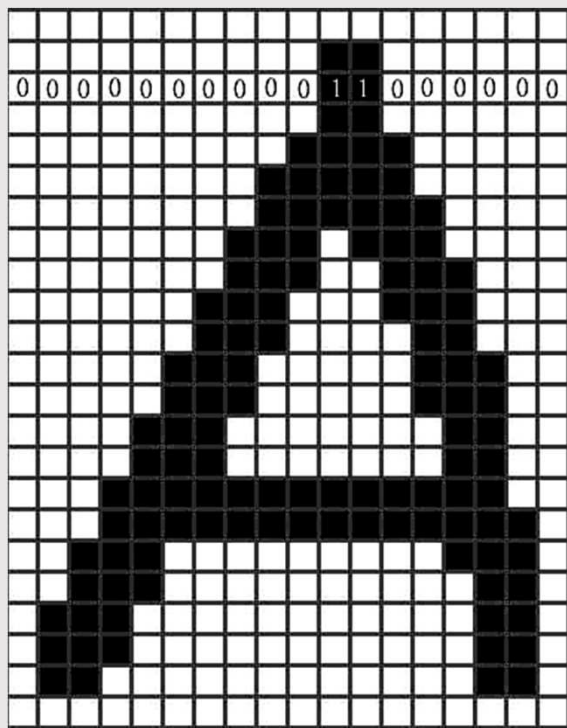


量化

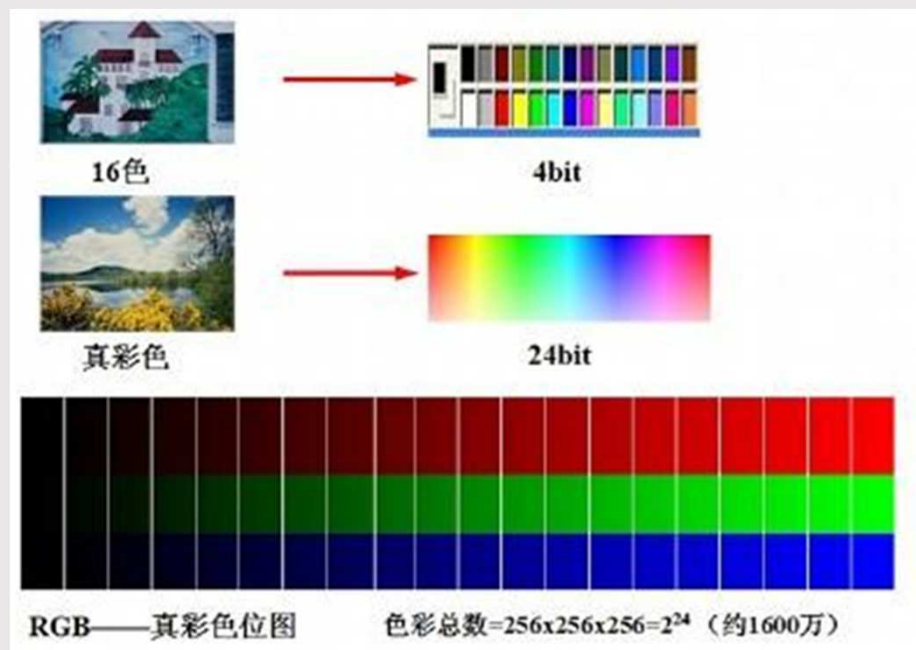


北京奥运会开幕式中，通过一个个方块构造的图案，正是图像数字化的体现。每个方块相当于一个像素点，通过人为表示该点来对其量化，最后组合成了数字化的图像。

# 片数字化



“黑”——“1”  
“白”——“0”



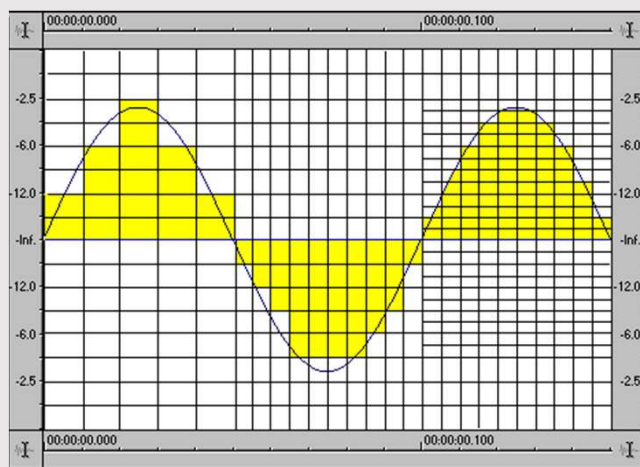
彩色图片的表示方法

# 音数字化

对于声波，通用是根据声波在指定点位于0线之上或之下的量，用一个二进制数记录。



模拟信号



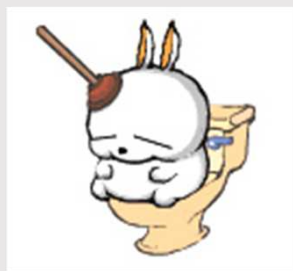
数字音频

在1秒钟内对声音的波形进行多次采样

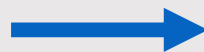
# 视频数字化

视频是由一连串相关的静止图象组成，我们将一幅图象称为一个帧。

数字视频:以一定的采样频率对模拟视频信号进行捕获、处理并生成数字视频信号的过程。数字视频信号可以在无损的情况下无限次地复制和长时间地保存，方便进行编辑和加工。



第一帧



第二帧

# 字技术的应用

电子商务---网上购物，就是通过互联网检索商品信息，并通过电子订购单发出购物请求，然后填上私人支票帐号或信用卡的号码，厂商通过邮购的方式发货，或是通过快递公司送货上门。



网上购物



常用的购物网站

# 什么是商务？

商务是广义概念，是指一切与买卖商品相关的商业服务。

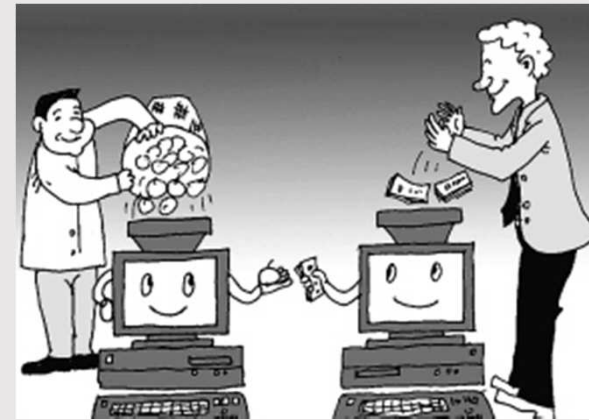
狭义的商务概念指商业或者贸易

商业活动，是指企业为实现生产经营目的而从事的各类有关资源、知识、信息交易等活动的总称。



# 电子商务的产生

- 电子商务产生的基础
  - 电子商务最早产生于20世纪60年代，发展于20世纪90年代。
- 产生和发展的社会基础
  - 政府的支持与推动
  - 计算机的广泛应用
  - 网络的普及和成熟
  - 完善的网络服务



# 电子商务的发展

## 电子数据交换 (EDI)

电子商务意味着利用电子化的手段，将商业买卖活动简化。基于EDI和电子货币转账的电子商务开始在上世纪70年代末应用。最典型的是将订单和发票等文档用电子数据发送。

## 信息分析技术应用于电子商务

信息分析技术应用于电子商务，80年代信用卡、自动柜员机成为电子贸易组成部分。90年代企业资源计划(ERP)数据挖掘和数据仓库成为电子商务一部分

## 互联网时代的电子商务

互联网时代到来电子商务有了新的组成部分“网络贸易”，客户在数据加密传输技术支持下，利用网上商店，通过互联网完成商品和服务的采购

# 中国电子商务的发展

起步期  
1990-1993

雏形期  
1993-1997

发展期  
1998-2000

成熟期  
2009-至今

稳定期  
2000-2009

1998年3月中国第一笔网上交易成功；

1999年10月国家经贸委

B2B的电子商务  
逐渐替代传统产业，电子商务进  
图可持续发展稳定  
定期

# 电子商务与传统商务的对比

1.交易前准备：电子商务模式下的交易中所需的供求信息是通过交易双方的网址和网页完成的。

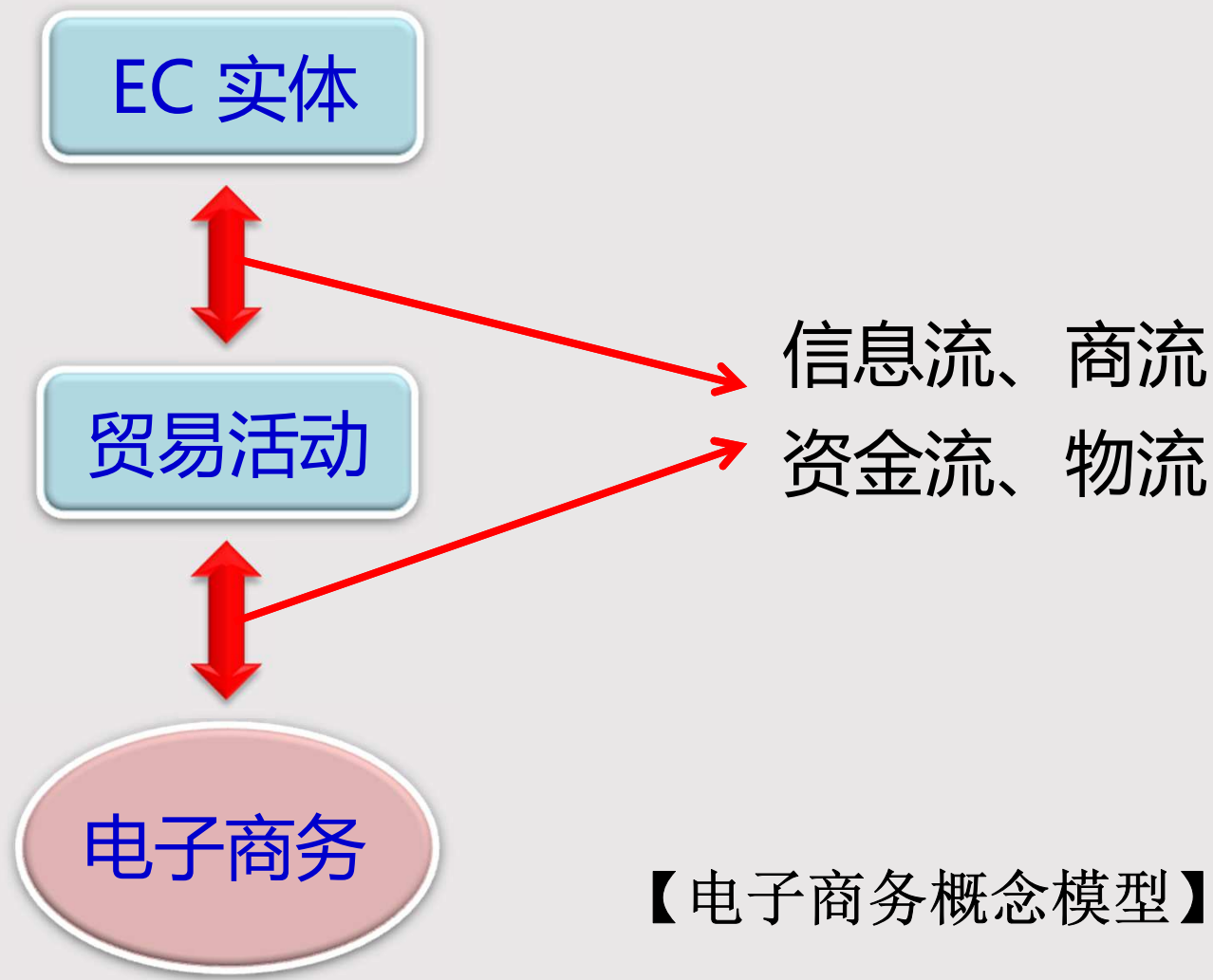
2.贸易磋商：电子商务模式下的磋商是通过电子记录替代纸张单据，通过文件和电子文档，及时通信等在网络上传递交易双方的信息。

## 交易过程

3.合同的签订与执行：电子商务环境下的网络协议和电子商务应用系统保障交易双方贸易磋商文件的安全性和可靠性。并且在第3方授权的情况下这些文件有法律效力，可以作为交易双方纠纷的法律仲裁依据。

4资金的支付：电子商务交易中加以双方的资金支付采用信用卡，电子支票，电子现金，和电子钱包等形式，以网上支付方式进行。

# 什么是电子商务—电子商务概念模型



## 什么是电子商务—广义和狭义的电子商务

商务：

**电子商务即商务，网络即生活**

**广义：**使用各种电子工具从事商务活动。

**狭义：**主要利用互联网从事商务活动。

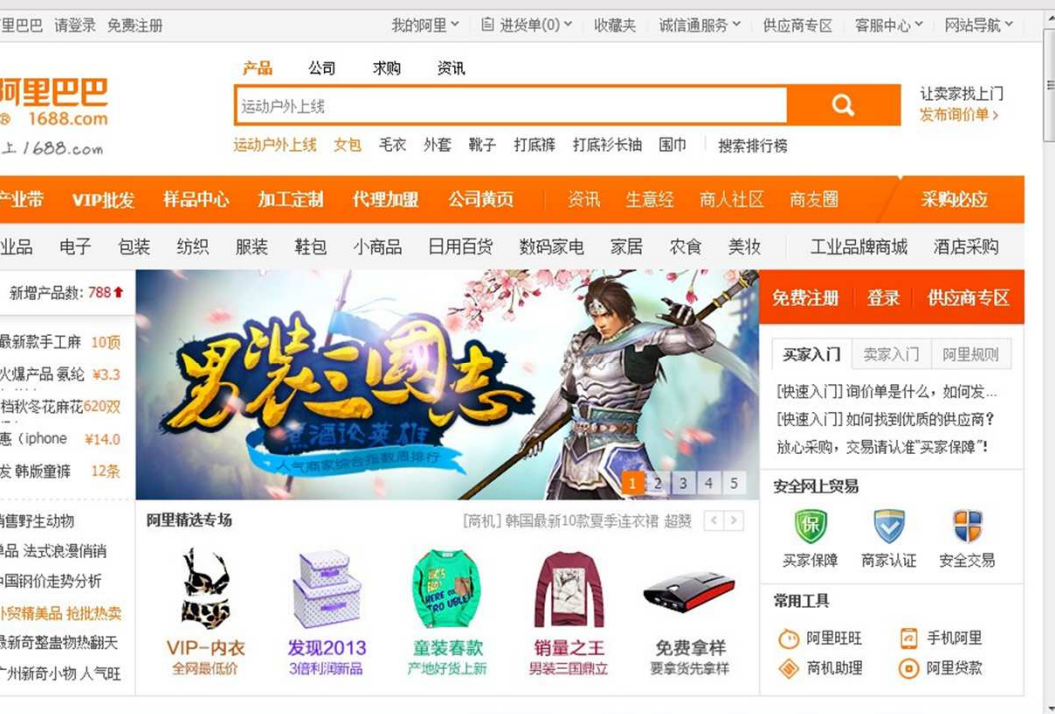
## 什么是电子商务—浅层次的电子商务

**浅层次的电子商务只是通过使用电子工具来完成贸易活动的部分流程和环节，比如：**

- 1.在网页上发布电子商情。
- 2.采用电子方式取代纸质凭据来做生意的电子贸易。
- 3.通过电子邮件和电子数据交换来明确合作方之间责权利关系的电子合同等等。

# 电子商务—阿里巴巴

## 网站首页

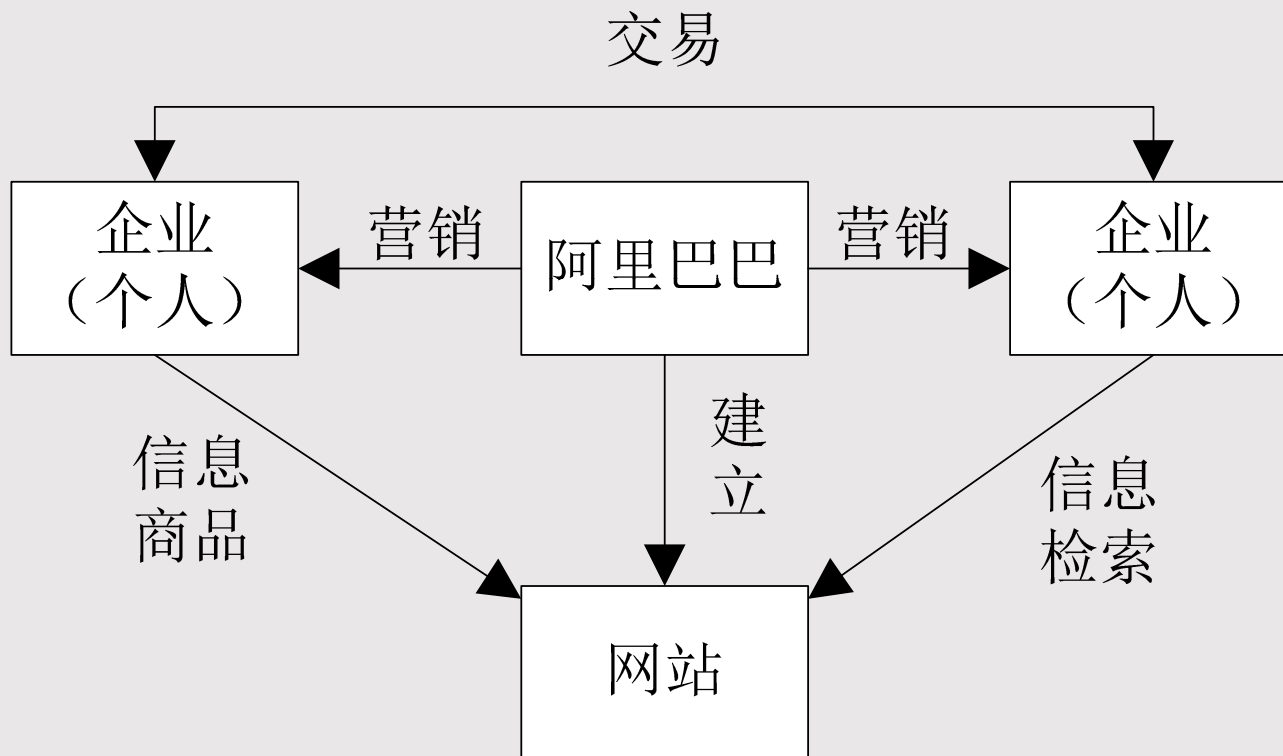


## 企业框架

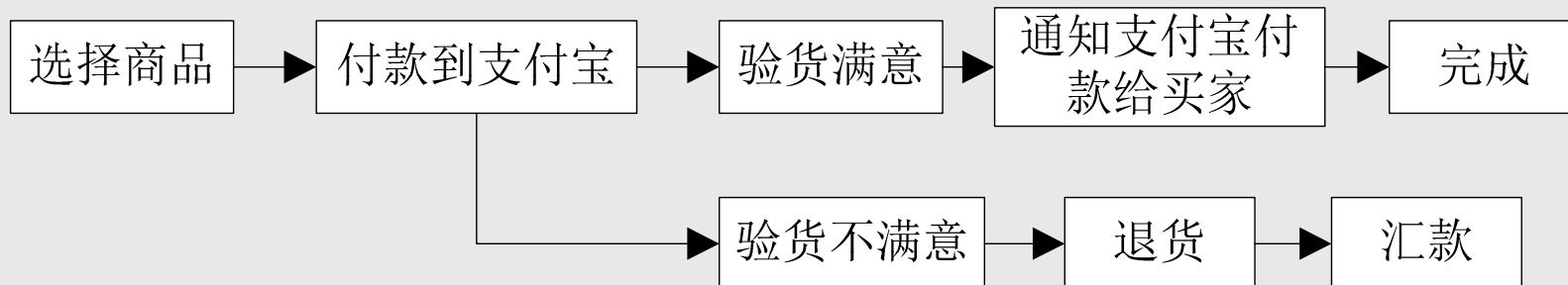


# 电子商务—阿里巴巴

阿里巴巴企业构架



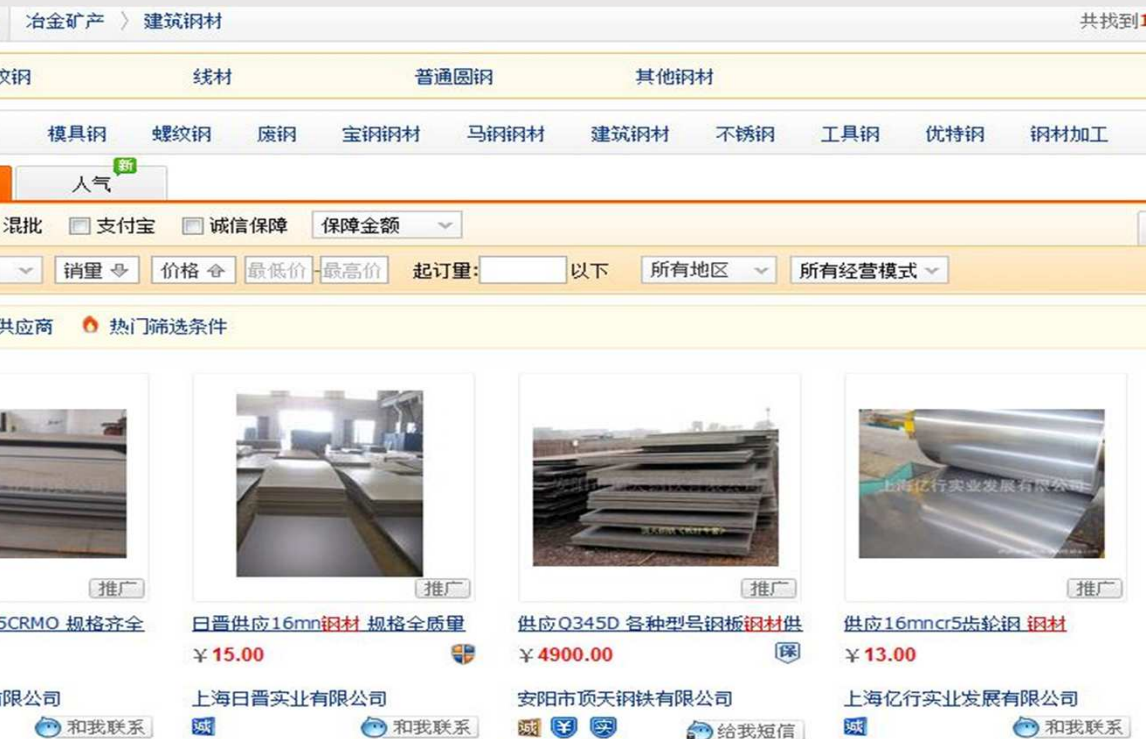
阿里巴巴的购买流程



# 电子商务—阿里巴巴功能

**商业机会与产品展示：**有各个行业大量产品分类的商业机会供查阅，并提供大约几十万供求信息。并按产品分类陈列展示阿里巴巴会上各类图文并茂的产品信息库。

**2、公司全库：**公司网站大全，目前已经汇聚4万多家公司网页。用户可以通过搜索寻找贸易伙伴，了解公司详细资讯。



# 电子商务—阿里巴巴功能

**行业资讯：**按各类行业分类发布最新动态信息，会员可以订阅最新信息，直接通过电子邮件接受。还可以按行提供企业最新报价和市场价格动态信息。

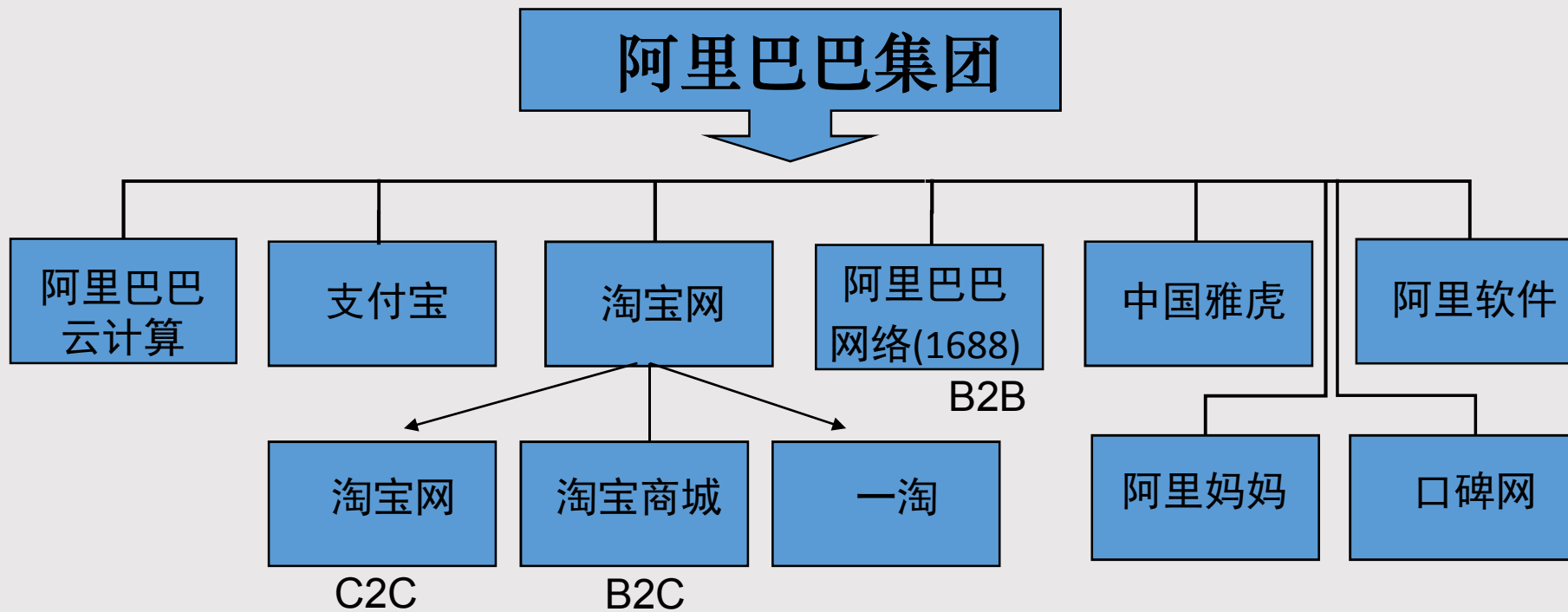
	<p><b>现货供应卡通全棉印花帆布</b></p> <p>最低起批量:1000米 成分及含量:全棉 纱支:10安 密度:10*10 幅宽:150 克重:色织 织物组织:平纹 染整工艺:印花 主...</p> <p>[浙江 绍兴市] <a href="#">绍兴县中国轻纺城爱妮布行</a></p> <p>减 第3年</p>	¥ 9.20
	<p><b>供应全棉/涤棉C060印花过胶帆布</b></p> <p>最低起批量:100码 成分及含量:CVC,80/20 纱支:— 密度:— 幅宽:55 克重:— 织物组织:平纹 染整工艺:...</p> <p>[广东 广州市] <a href="#">广州市花都区狮岭福龙达布业商行</a></p> <p>减 第6年</p>	¥ 19.50
	<p><b>供应2*2 120涂料印花帆布</b></p> <p>最低起批量:5000码 成分及含量:100%全棉帆布印花 纱支:21 密度:2*2 幅宽:152 克重:180 产地:湖北 主要用途:...</p> <p>[广东 博罗县] <a href="#">博罗县园洲镇奕胜印花厂</a></p> <p>减 第1年</p>	¥ 25.00
	<p><b>10安全棉帆布印花10安印花豹纹斑马纹12安印花迷彩印花布料印小花</b></p> <p>最低起批量:1码 成分及含量:LDZQ294#厂家大量现货¥18/码 起宽度58"=147cm厚度0.8,《匹匹验货检查合格;客户...</p> <p>[广东 东莞市] <a href="#">陈福建</a></p>	¥ 16.00

4、商人社区：在这里会员交流行业见解，谈天说地。

商机市场	商业资讯	<b>商人社区</b>
论坛: 商周刊   小企业大民主   明星网商   明星版主   新手上路		
生意经   云计划   博客   视		
<p><b>热点聚焦</b> <span style="float:right">更多&gt;</span></p> <p><b>最美厂妹的一天：旺铺主管的一天</b></p> <p>论坛与旺铺同步，qq旺旺齐开!这就是我每日电脑前的生活，方寸屏幕之前，是我缤纷多彩的网络工作生活。忙并快乐着...文/☆木喜☆</p> <p><b>【最美厂妹的一天】：1个帖子20张图片赢取6万财富值</b> 36人回复</p> <p><b>【最美厂妹的一天】：电脑保护套小店老板娘的一天</b> 136人回复</p> <p>21岁打工仔做游戏机赚百万，被哥哥7个月搞倒闭 13人回复</p> <p>生产26万本书，因管理不到位，补纸用了1.7万张 6人回复</p> <p>高龄职场人士的尴尬：43岁，工作14年，月入2千5 20人回复</p>		
<p><b>通明众美女亮相，缘在这里</b></p> <p>口号：让我们一起过好生活! 1 2 3 4</p>		
<p>社区服务</p> <p>阿里公告   社区共建   论坛帮助   版主之家</p>		
<p>综合商务</p> <p>管理杂谈   商业模式   创业   国内贸易</p> <p>采购俱乐部   网商故事   职场生涯   外贸</p>		
<p>行业商经</p> <p>能源矿产   五金工具   化工   机械设备</p> <p>家居   橡塑   纺织   建筑建材   农业</p>		
<p>海阔天空</p> <p>休闲娱乐   谈天贴图   情感空间   单身同盟</p>		
<p>城市文明</p>		
<p><b>贸易实战</b> <span style="float:right">更多&gt;</span></p> <p><b>13人团队创业1年，月销售额130万</b></p> <p>杭州赫品贸易有限公司阿里网店成于2010年，仅仅一年多的时间，创造出月销售额130W的耀人业绩，这个只有13人的网络团队...文/诚信服务批发商</p> <p><b>[电商]分期付+降价，贱卖来抢单，要尊严还是要生意?</b> 文/姜☆峰☆</p> <p><b>[转载]为什么有些农民跑去 弄弄十把的线线却不原主睡?</b> 文/老根's</p>		

# 电子商务—阿里巴巴

阿里巴巴集团公司有10家子公司，分别是阿里巴巴、淘宝、支付宝、阿里软件、阿里妈妈、口碑网、阿里云、中国雅虎、一淘网、淘宝商城。



# 什么是电子商务—深层次的电子商务



信息流



商流



资金流



物流



寻找客户



在线洽谈



下单购买



支付收款



在线发货

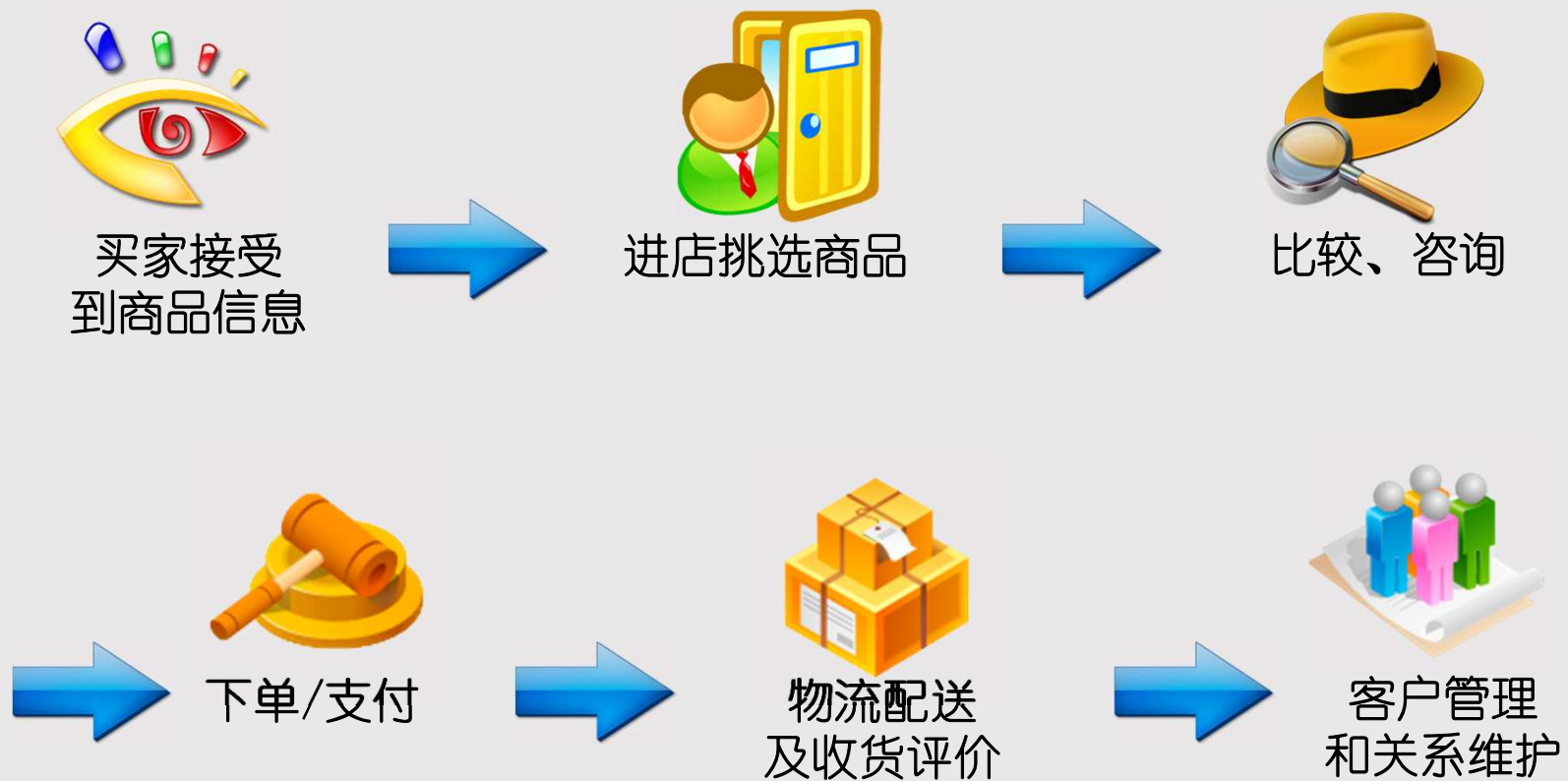


电子报关

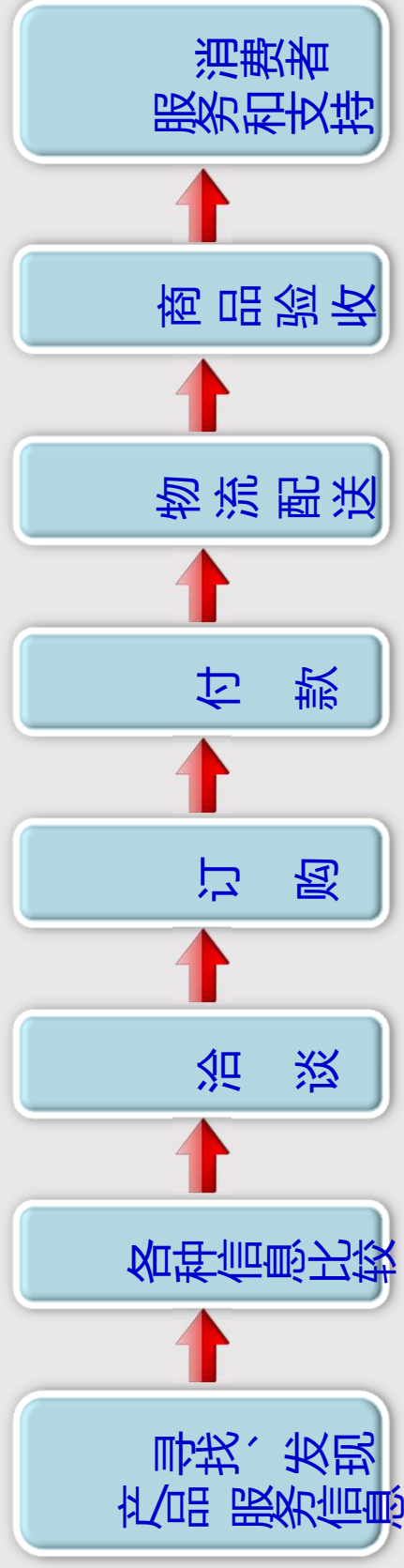


电子纳税

# 什么是电子商务—深层次的电子商务



# 电子商务的运作流程



## 【电子商务的一般流程】

# 电子商务的特点和运用



降低成本，提高效率



没有时间和地域的限制



公平享用信息资源



减少了交易的中间环节



增强了买卖双方之间的互动性



快捷的信息传递速度

## 电子商务主要表现形式

### **B2B** (商家对商家)

企业与企业之间通过互联网进行产品、服务及信息的交换。

### **B2C** (商家对个人)

企业通过互联网为消费者提供一个新型的购物环境——网上商店。

### **C2C** (个人对个人)

用户对用户的模式，C2C商务平台就是通过为买卖双方提供一个在线交易平台，使卖方可以主动提供商品上网拍卖，而买方可以自行选择商品进行竞价。

# 电子商务B2C案例



当当网

卓越网



B2C的先行者推进了网购概念的普及，培养了大家的网购习惯。

# 电子商务B2B&C2C案例



B2B

C2C



# 电子商务中数据分析

数字是很有说服力的东西，而且也可以更加深入的掌握不同变量之间的逻辑关系。而针对电子商务中的商家来说，通过**数据分析**，采取正确的策略，是竞争的关键！

通过分析用户的注册时间，购买次数，购买金额，购买频率，最后一次购买时间等数值，对我们都有重要的参考和分析意义，只有细致分析，才能精准营销。

## 案例：塔吉特超市未卜先知怀孕案例

曾经有一位男性顾客到一家塔吉特超市店中投诉，商店竟然给他还在读书的女儿寄婴儿用品的优惠券。这家全美第二大零售商，会搞出如此大的乌龙？但经过这位父亲与女儿进一步沟通，才发现自己女儿真的已经怀孕了。

为什么塔吉特能知道这个用户怀孕了？必须有哪几个关键环节

A: 用户数据收集 B: 怀孕特征库 C: 怀孕潜在用户筛选

塔吉特在和顾客沟通过程中采用了哪种营销方式

A: 电子邮件 B: 直邮 C: 电话营销 D: 数据库营销



## 关键环节一：数据信息记录

家零售商是如何比一位女孩的亲生父亲更早得知其怀孕消息的呢？每位顾客初次到塔吉特刷卡消费时，都会得一组**顾客识别编号**，内含顾客姓名、信用卡卡号及电子邮件等个人资料。日后凡是顾客在塔吉特消费，计算机系统就会自动记录消费内容、时间等信息。再加上从其他管道取得的统计资料，塔吉特便能形成一个庞大数据库，运用于分析顾客喜好与需求。

个ID号还会对号入座的记录下你的人口统计信息：年龄、是否已婚、是否有子女、所住市区、住址离Target的路程、薪水情况、最近是否搬过家、钱包里的信用卡情况、常访问的网址等等。Target还可以从其他相关机构**购买你的其他信息：种族、就业史、喜欢读的杂志、破产记录、婚姻史、购房记录、求学记录、阅读习惯等**。乍一看，你会觉得这些数据毫无意义，但在Andrew Pole和顾客数据分析部的手里，这些看似无用的数据便发了前述强劲的威力

## 关键环节二：数据模型建立

Andrew Pole想到了Target有一个迎婴聚会(baby shower)的登记表。Andrew Pole开始对这些登记表里的顾客的消费数据进行建模分析，不久就发现了许多非常有用的数据模式。

模型发现，许多孕妇在第2个妊娠期的开始会买许多大包装的无香味护手霜;在怀孕的最初20周大量购买补充钙、锌的善存片之类的保健品。最后Andrew Pole选出了25种典型商品的消费数据构建了“怀孕预测指数”，通过这个指数，Target能够在很小的误差范围内预测到顾客的怀孕情况，因此Target就能早早地把孕妇优惠广告寄给顾客。

## 关键环节三：建立和用户沟通渠道

那么，顾客收到这样的广告会不会吓坏了呢？Target很聪明地避免了这种情况，它把孕妇用品的优惠广告夹杂在其他一大堆与怀孕不相关的商品优惠广告当中，这样顾客就不知道Target知道她怀孕了

## get取得的成就

根据Andrew Pole的大数据模型,Target制订了全新的广告营销方案,结果Target的孕期用品销售呈现了爆炸性的增长。Andrew Pole的大数据分析技术从孕妇这个细分顾客群开始向其他各种细分客户群推广,从Andrew Pole加入Target的2002年到2010年间, Target的销售额从440亿美元增长到了670亿美元。

# 风&蛋挞 And 啤酒&尿布

沃尔玛在上世纪创造的啤酒与尿布的相关性销售是一个营销的经典案例，但是类似的不只有这一个，飓风和啤酒也是沃尔玛基于大数据分析之后的一个又很成功的营销案例。



啤酒和尿布、星期五、男人、小孩

飓风和蛋挞、躲在家里、吃蛋挞

# 信息化发展步入大数据时代



数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来，**大数据正成为最得关注的领域之一**。“大数据”在互联网行业通常表现为互联网公司在日常运营中生成、的用户网络行为数据。

- 每一秒：全球发送290万封电子邮件；
- 每一分钟：微博推特上新发的数据量超过10万；社交网络Facebook的浏览量超过600万；
- 每一天：全球上传2.88万小时视频至Youtube；
- 每个月：网民在Facebook上花费7千亿分钟；
- 上一年：人们制造并使用的数据达1.8ZB。

世界公民？

# 数据时代 → 数据爆炸式增长 (每分钟.....)

100+个iPhone  
应用下载

Skype上  
37万+分钟的语音  
通话

Twitter上发布  
98000+新微博

6600张新照  
片到flickr

发出1.68亿+条  
Email

Facebook上更新  
69.5万+条新状态

Tube上上传  
100+新视频

淘宝光棍节  
10680+个新订单

12306出票  
1840+张



# 我们正在 面临一个 “大”问题

在互联网带来的“大”问题（**大数据、大用户、大系统、大平台、大统一、大优化.....**）压力下，我们需要全新的思想，来重新定义计算资源的使用方式、务的提供方式，以及社会化大生产的协作过程。现代企业必须通过自身信息资源的优化整合，实现存储容量和可用性的最大化，实施灵活性设计以支持不断变的业务需求等相关措施来迎接大数据（BIG DATA）时代带来的机遇和挑战。



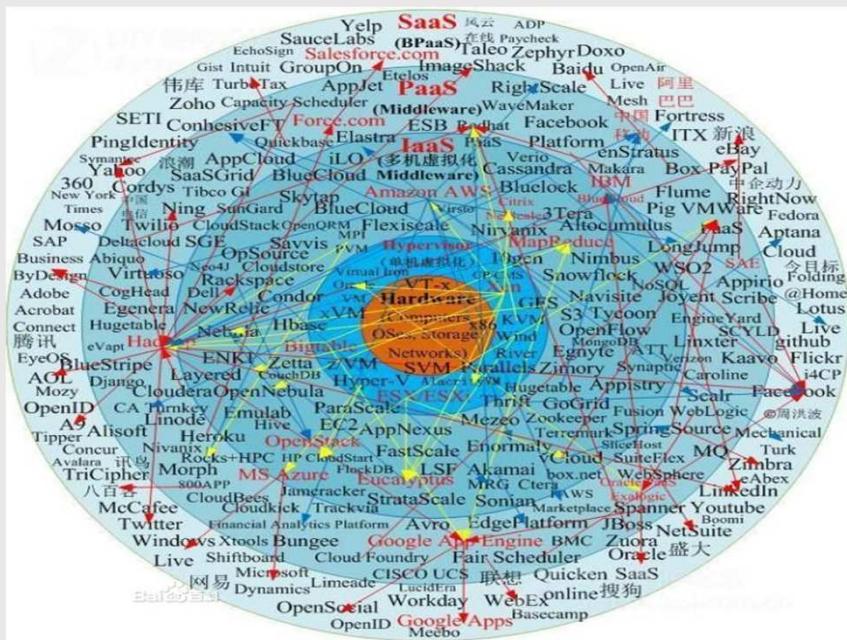
核心问题

# 大数据 BIG DATA



# 数据是什么？

大数据 = 海量数据 + 分析软件 + 挖掘过程



软件是大数据的驱动力、引擎，软件改变世界

# 数据给社会带来巨大价值

大数据为我们更深入准确地认识和把握事物发展的内在规律提供了信息基础，蕴涵着非常大的潜在价值。这一点不仅是许多业内机构的共识，更已经在商业应用中得到了体现。

据美国麦肯锡公司预测，大数据为美国医疗服务业每年带来**3000亿美元**的潜在增加值，为欧洲的公共管理每年带来**2500亿欧元**的潜在价值，为位置服务产业带来**6000亿美元**的潜在年收入。

零售商充分利用大数据可实现运营**利润增长60%**，制造业充分利用大数据可降低设备装配成本50%。

经合组织（OECD）的一项最新研究成果还对互联网数据的市场价值进行了估计，佐证了大数据的巨大潜在价值。



## 市场价值

2013年      \$181亿

2017年      \$470亿



# 数据成为现代企业核心竞争力

大数据正在“吞噬”和重构很多行业，广泛运用数据分析手段管理和优化运营的公司其实质都是一个数据公司。麦当劳、肯德基以及苹果旗舰店专卖店的位置都是建立在数据分析基础之上的精准选址。在零售业中，数据分析的技术与手段更是得到广泛的应用，传统企业如沃尔玛开始挖掘重塑并优化供应链。而电商大鳄如卓越亚马逊、淘宝等则通过对海量数据的掌握和分析，为用户提供更加专业化和个性化的服务。

**自身服务提升：**亚马逊每天生成630万份订单，大数据帮助亚马逊提高对客户的洞察力，随时跟踪用户需求变化。

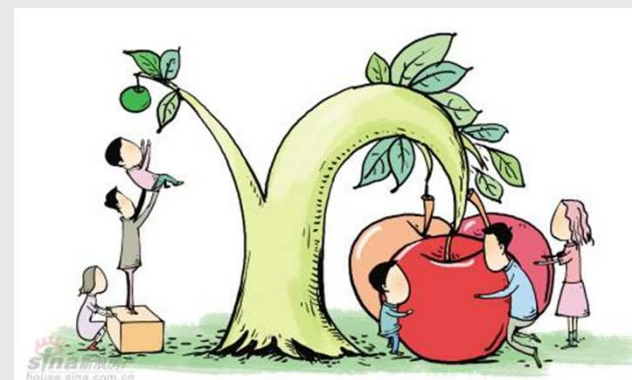
**对外提供业务：**淘宝每天活跃数据量50TB，提供各种行业分析报告（如顾客的特征、什么最好卖、行业的竞争对手、广告投放的效果、用户的购买行为等）



识别客户

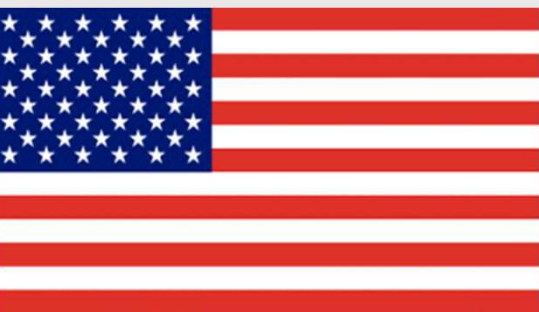


区分客户



快速适配

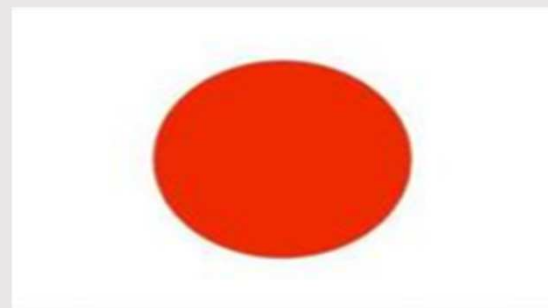
## 数据时代 → 大数据规划战略



2012年3月，美国政府启动“大数据研究和发展倡议”计划。这是继“信息高速公路”后的又一重大科技战略部署。美国政府将大数据视为“未来的新石油”并将对其研究上升为国家意志。



2010年1月，英国政府的数据开放网站正式出台，2010年5月卡梅伦上台之后正式提出“数据权”的概念，2011年4月主要部门宣布“我的数据”新项目，旨在落实和强化数据权。



继美英之后，国际上德国、法国、日本、加拿大和新西兰等国家也开始了各自国家的大数据发展战略规划。

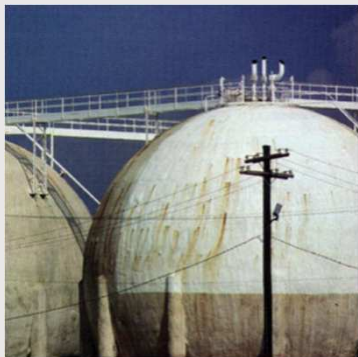
# 数据时代 → 数据是可再生资源，本质是生产资料



仅供开采162年

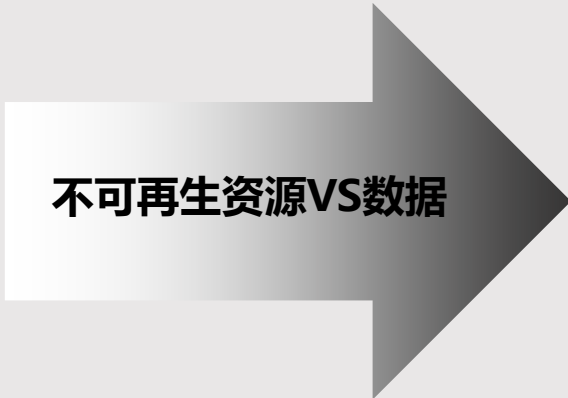


仅供开采45年



仅供开采60年

数据不再是社会生产的“副产物”，而是可被二次乃至多次加工的原料，从中可以探索更大价值，它变成了生产资料。



过去3年数据总量被以往4万年还多

2013年,10分钟的信息总量将达1.8ZB

2010年全球数据总量1.2ZB，年增长50%



# 大数据 → 大数据的4V特征（更普遍意义）

## Volume

结构化数据的超大规模和增长  
数据量的80~90%  
结构化数据增长快10倍到50倍  
传统数据仓库的10倍到50倍

## Value

大量的不相关信息  
来自趋势与模式的可预测分析  
复杂分析（机器学习、人工智能Vs传统  
智能）

# Big Data 大数据



## Variety

- 大数据的异构和多样性
- 很多不同形式（文本、图像、视频、音频、传感器数据等）
- 无模式或者模式不明显
- 不连贯的语法或句义

## Velocity

- 数据输入、处理与丢弃
- 立竿见影而非事后见效
- 实时分析而非批量式分析

# 大数据 → 大数据的4V特征 ( Volume )



一般情况下，大数据是以PB、EB、ZB为单位进行计量的

1PB相当于50%的全美学术研究图书馆藏书信息内容

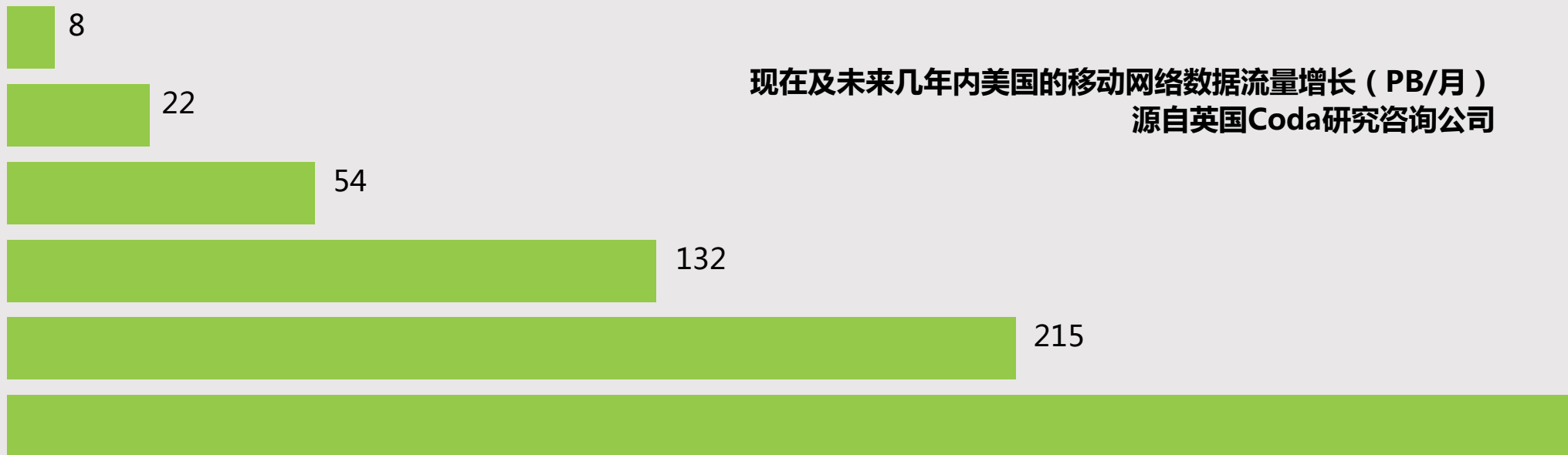
5EB相当于至今全世界人类所讲过的话语

1ZB如同全世界海滩上的沙子数量总和

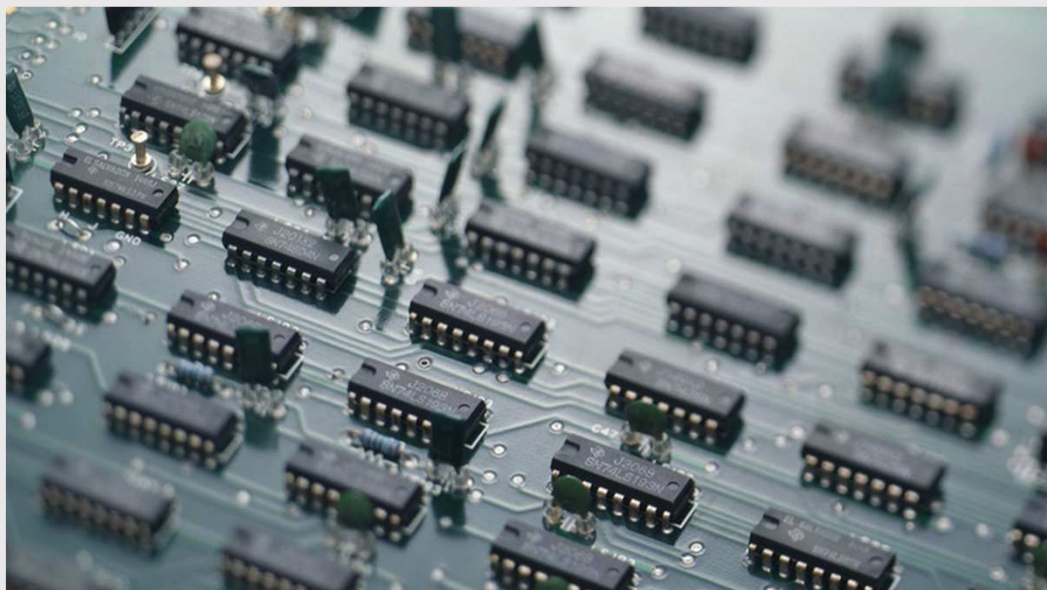
1YB相当于7000位人类体内的微细胞总和

# 大数据 → 大数据的4V特征 ( Velocity )

大数据的增长速度快

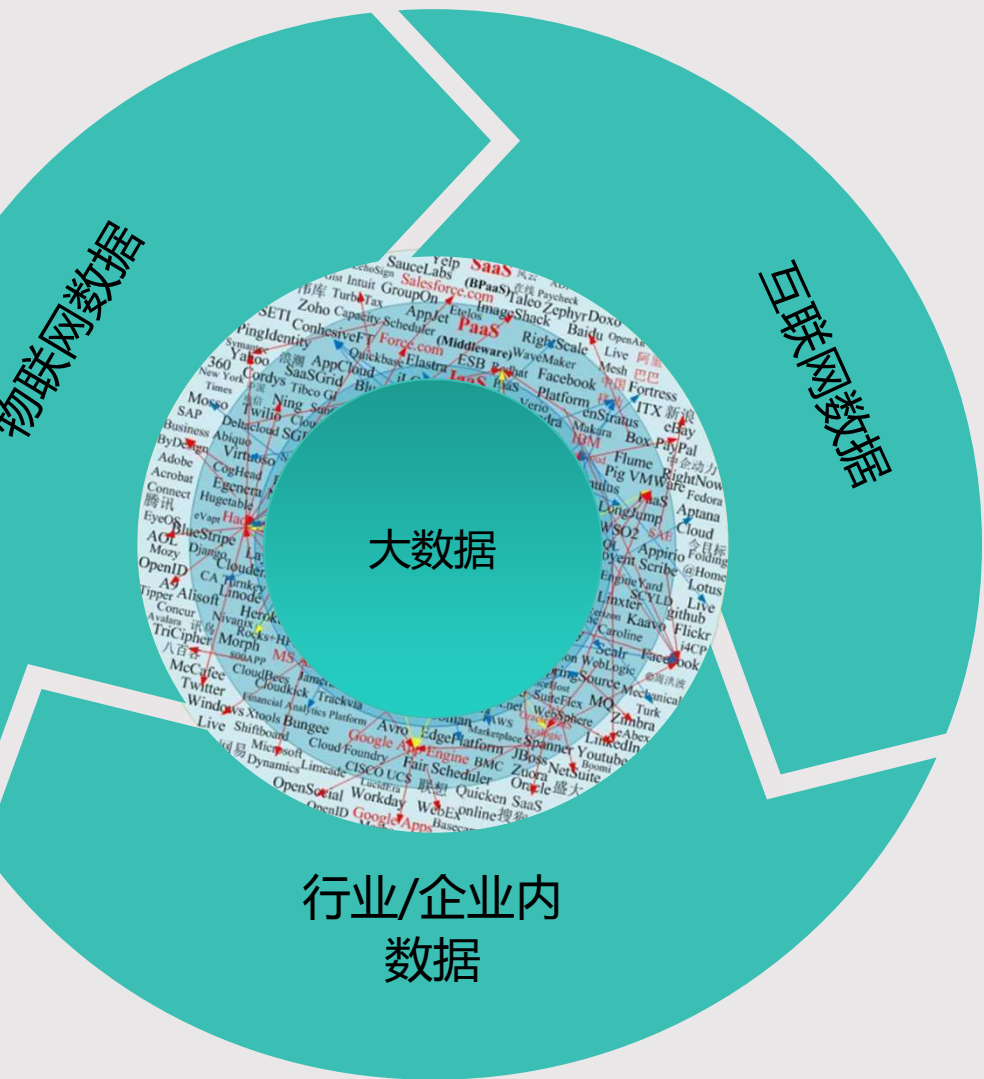


大数据的处理速度快



- 实时数据流处理的要求，是区别大数据引用和传统数据仓库技术，BI技术的关键差别之一；
- 1s 是临界点，对于大数据应用而言，必须要在1秒钟内形成答案，否则处理结果就是过时和无效的；

# 大数据 → 大数据的4V特征 (Variety)



## 数据来源多

企业内部多个应用系统的数据、互联网和物联网的兴起，带来了微博、社交网站、传感器等多种来源。

## 数据类型多

保存在关系数据库中的结构化数据只占少数，70~80%的数据是如图片、音频、视频、树状连接信息、文档等非结构化和半结构化数据。

## 关联性强

数据之间频繁交互，比如游客在旅行途中上传的照片和日志，就与游客的位置、行程等有了很强的关联性。

## 大数据 → 大数据的4V特征 ( Value )

- 挖掘大数据的价值类似沙里淘金，从海量数据中挖掘稀疏但珍贵的信息；
- 价值密度低，是大数据的一个典型特征；



# 大数据 → 大数据带来的思维变革

**更杂**

不是精确性而是混杂性

**更多**

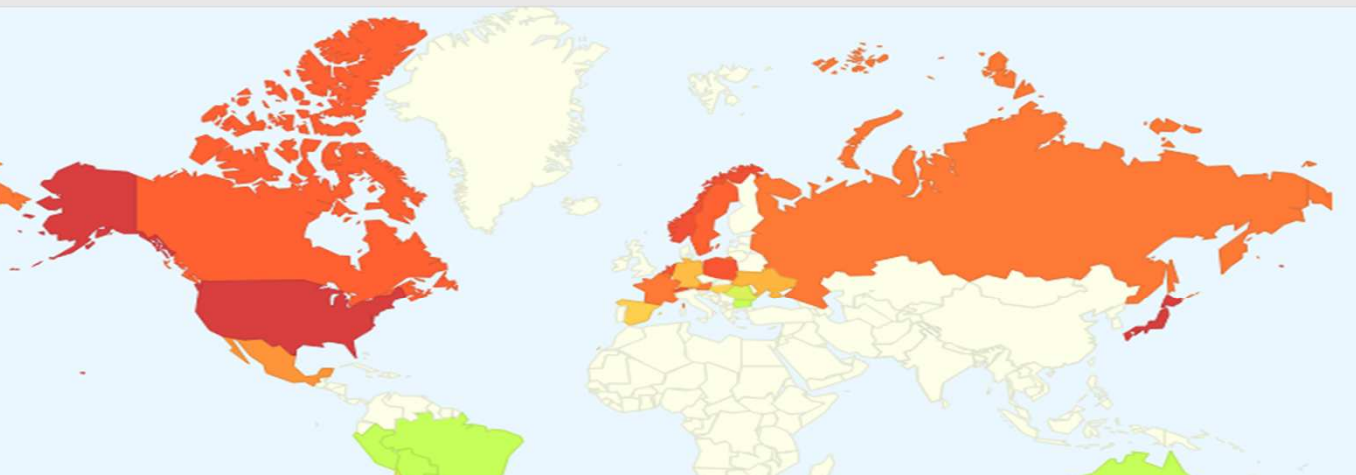
不是随机样本而是全部数据

**更好**

不是因果关系而是相关关系



# 大数据 → 大数据带来的思维变革



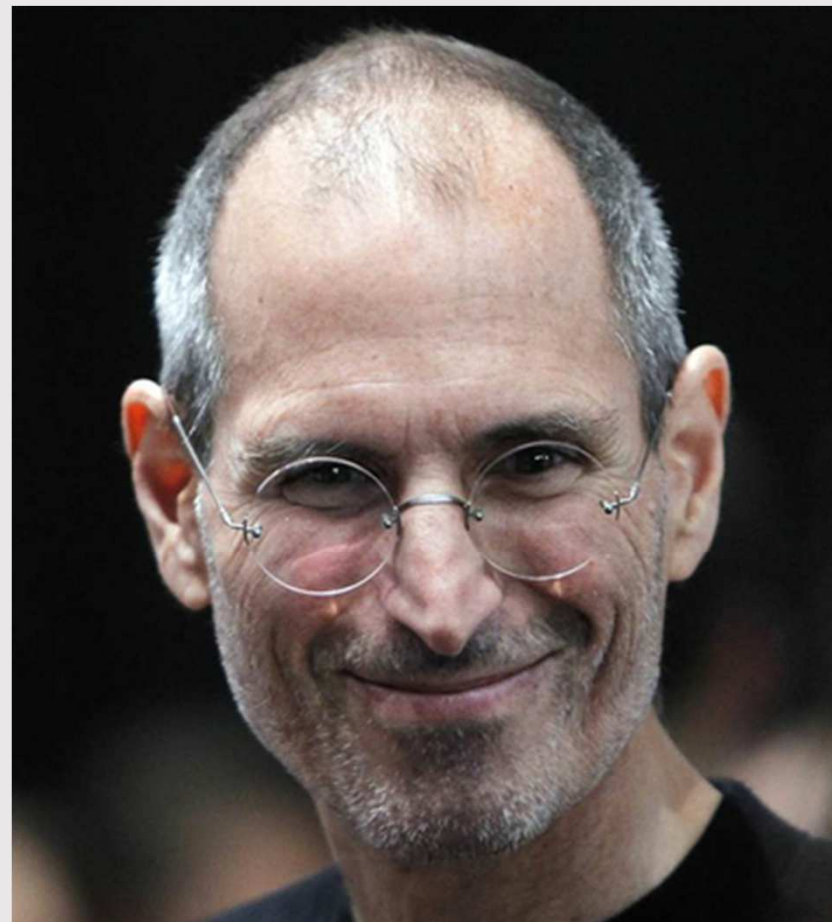
Google利用网络大数据预测流感



于64000场比赛进行全数据相扑比赛的作弊分析



埃齐奥尼的Forecast有10万亿条数据预测机票价格



乔布斯的癌症抗争，自身所有DNA和肿瘤DNA排序



- “2008年初,阿里巴巴平台上整个买家询盘数急剧下滑, 欧美对中国采购在下滑。海关是卖了货,出去以后再获得数据;而我们提前半年时间从询盘上推断出世界贸易发生了变化了。”

■ **马云对未来的预测, 是建立在对用户行文分析的基础上。**通常而言, 买家在采购商品前, 会比较多家供应商的产品, 反映到阿里巴巴网站统计数据中, 就是查询点击的数量和购买点击的数量会保持一个相对的数值, 综合各个维度的数据可建立用户行为模型。因为数据样本巨大, 保证用户行为模型的准确性。因此在这个案例中, 询盘数据的下降, 自然导致买盘的下降。

■ 腾讯在天津投资建立亚洲最大的数据中心; 百度也在投资建立大数据处理中心; 新浪推出企业微博产品, 提供精准的数据分析服务。

# 智慧城市-典型的大数据应用集合



机器生成数据

7x24不间断

数据量大

产生速度快

实时分析

流式分析

模式挖掘

# 算、物联网变成实用工程

国内各地制定云计算“十二五”规划



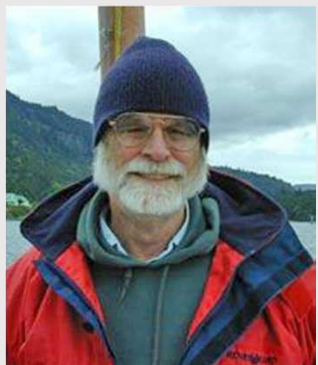
云计算、物联网园区



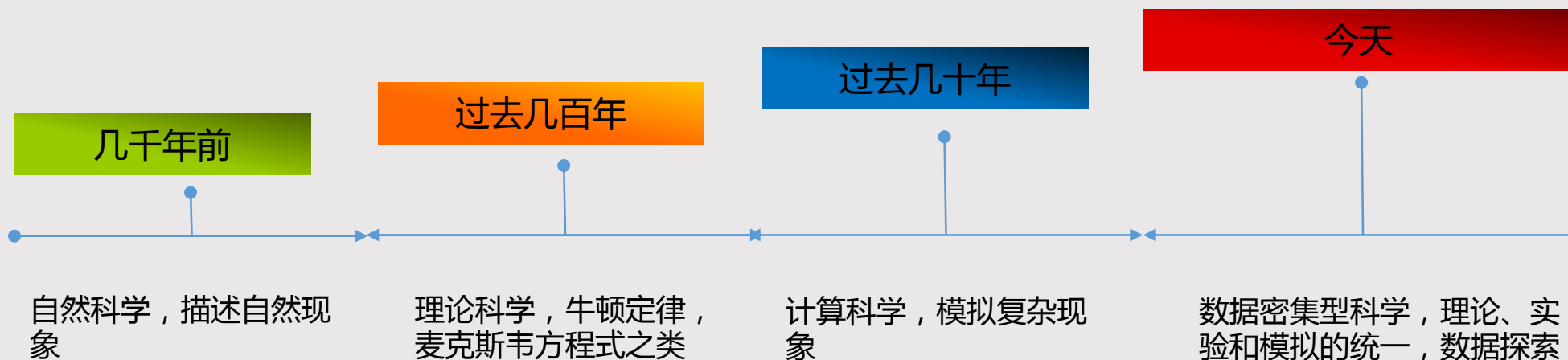
中国各地制定或公布了云计算、物联网等产业规划；这些工程的初始着眼点在房地产，政绩工程居多，大数据作为核心内容端，使得政绩工程变为使用工程。

云计算、物联网、社交媒体、GIS为大数据提供了丰富的数据来源。因此大数据中包括的每个用户的身份、地点、时间、喜好、厌恶、社会关系等等大量的信息。伴随数据挖掘和分析的技术发展，我们即将步入基于大数据的智能化时代。

# 数据是世界的未来



2007年图灵奖得主Jim Gray在美国国家科学研究委员会发表演讲，指出科学研究已经在几千年前的凭经验、几百年前的靠理论模型、几十年前的科学仿真之后，进入了第四个阶段——数据探索！



## 数据时代 → 业界对大数据的关注持续升温

早在1980年，著名未来学家阿尔文·托夫勒便在《第三次浪潮》一书中，就将大数据热情地赞颂为“第三次浪潮的华丽乐章”。

2012年世界经济论坛发布了“Big data , Big impact”的报告

从金融服务、健康、教育、农业、医疗等多个领域阐述了大数据为国际发展带来的机会。

麦肯锡于2011年5月发布《下一个前言：创新、竞争和生产力》报告

大数据将成为竞争的关键，大数据将引发新一轮的生产力增长与创新。

Big Data Insight Group和EMC完成的《解读大数据的商业价值和战略意义》

对大数据在英国企业中的应用趋势进行了调查和解读，发现大数据还处于发展初级阶段。

# 数据时代 → 大数据在国内日益受重视



## 大数据事件

- 2012年5月，中国云计算国际论坛在北京召开；
- 2012年6月，中国计算机协会成立“CCF大数据专家委员会”；
- 2012年7月，阿里巴巴宣布设立首席数据官，推动“数据分享平台”战略；
- 2012年7月，第二届大数据世界论坛召开，首届中国大数据论坛同年召开；



## 专业报告

- 2013年1月，华为《超越信息和通信，迎接新一轮数字化变革》；
- 2012年4月，ETIRI《大数据行业应用现状与未来应用热点》；
- 2011年12月，中信证券《大数据（Big Data）专题研究报告》；
- 2011年12月，国金证券《“大数据”专题分析报告》；
- 2012年6月，《中国互联网市场洞见：互联网大数据技术创新研究》；



## 他们如此说

- 随着大数据产品的逐步退出与完善，2013年国内大数据市场将进入主要培育期；光大证券
- 随着解决方案增多以及用户需求的明确，未来大数据市场将进入高速增长期；计世资讯
- 大数据对国家治理模式、企业决策、组织和流程、个人生活都将产生巨大影响；国金证券
- 在这场呼之欲来的产业革命中，我国并没有落在世界的后面。事实上，全世界都还没有准备好迎接这场革命；周涛

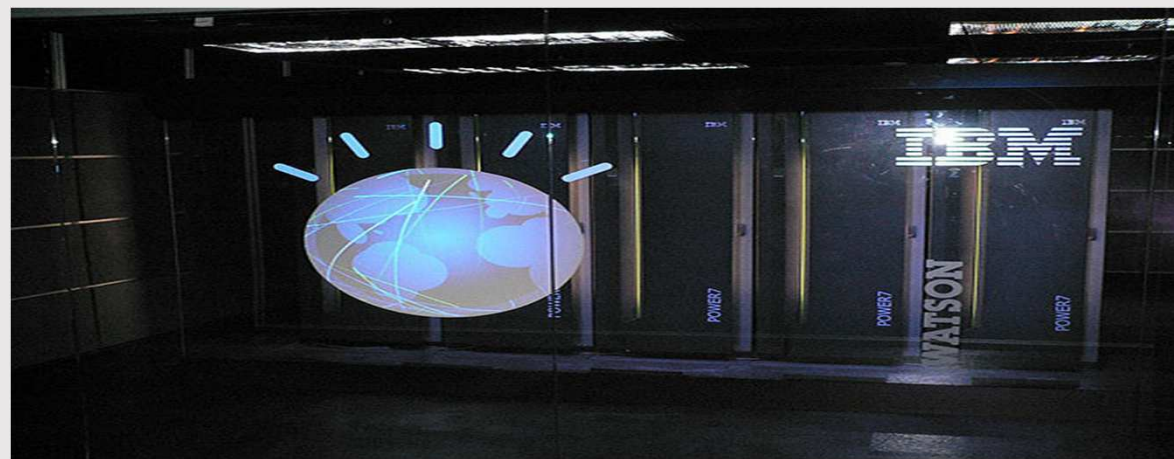
# 已经开始投资应用



美国国务院采用大数据技术开发新的美国护照系统。



美国IT公司开始意识到大数据技术能够为公司创造价值；



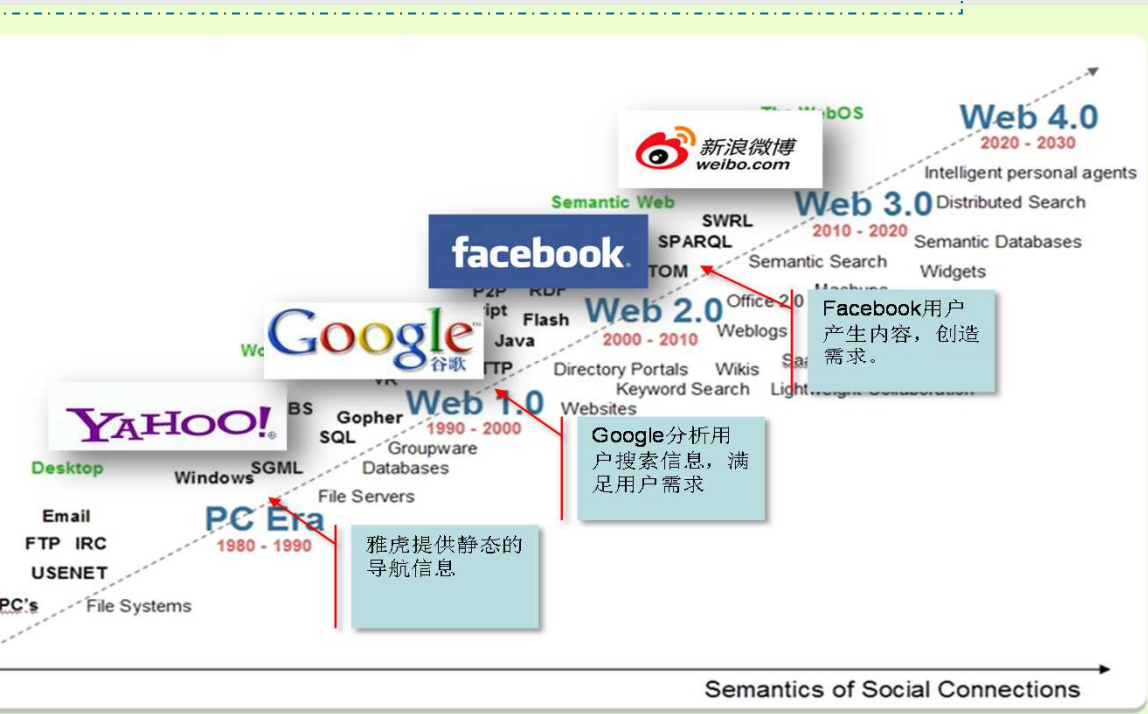
■ IBM宣布投资1亿美元用于大数据研究；



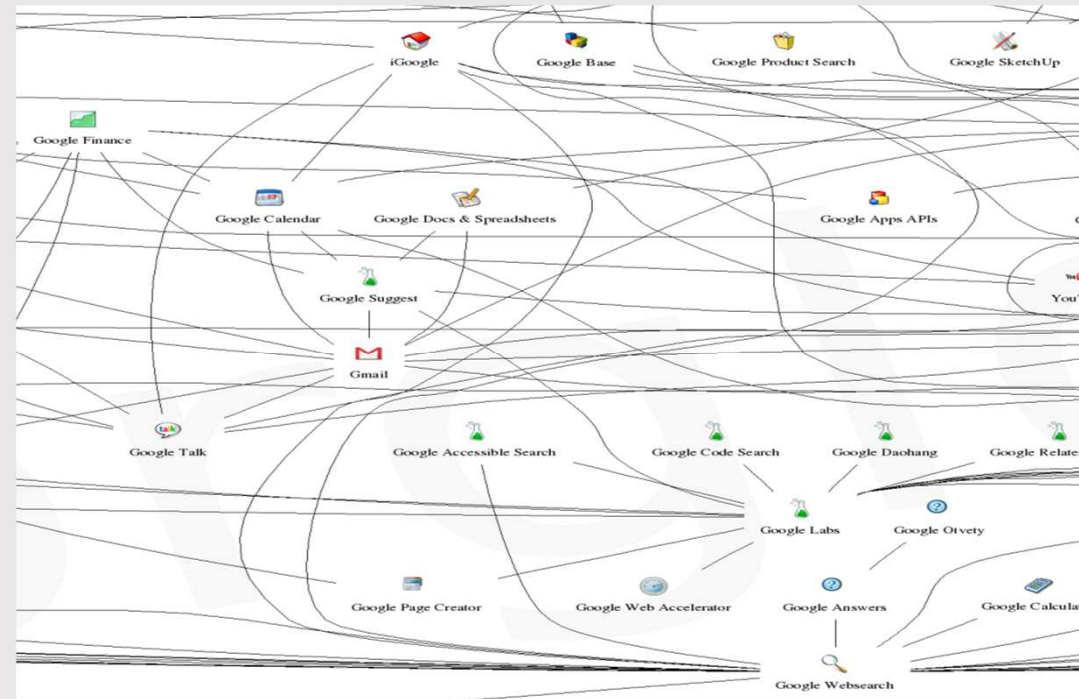
■ 大数据公司引入汽车行业高管人员扩展营销业务；

# 孙、Google等互联网公司

互联网越来越智能



Google精确掌握用户行为、获取需求



前瞻来看, 随着互联网对网民的理解, 网民对网络的反作用, 互联网将变得越来越智能。它在满足你需求的同时, 也在创造新的需求。前者的代表是Google, 后者的典型则是Facebook。

谷歌的盈利在于所有的软件应用都是在线的。用户在免费使用这些产品的同时, 把个人的行为、喜好等信息也免费的送了Google。因此Google的产品线越丰富, 他对用户的理解就越深入, 他的广告就越精准。广告的价值就越高。

这是正向的循环, 谷歌好用的、免费得软件产品, 换取对用户的理解; 通过精准的广告, 找到生财之道。颠覆了微软卖软件拷贝赚钱的模式。成为互联网的巨擘。

# 大数据市场分析

## 2011年-2016年中国大数据市场规模

●计世资讯认为，2011年是中国大数据市场元年，一些大数据产品已经推出，部分行业也有大数据应用案例的产生。2012年-2016年，将迎来大数据市场的飞速发展。

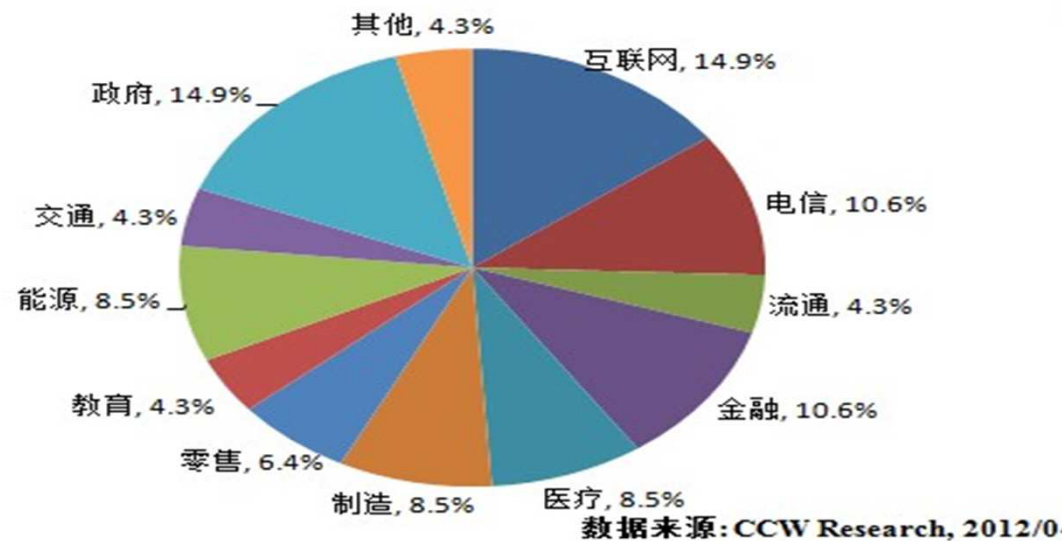
●计世资讯预测，2012年中国大数据市场规模将达到4.7亿元，2013年大数据市场将迎来增速为138.3%的飞跃，到2016年，整个市场规模逼近百亿。



## 2012年各行业大数据市场规模

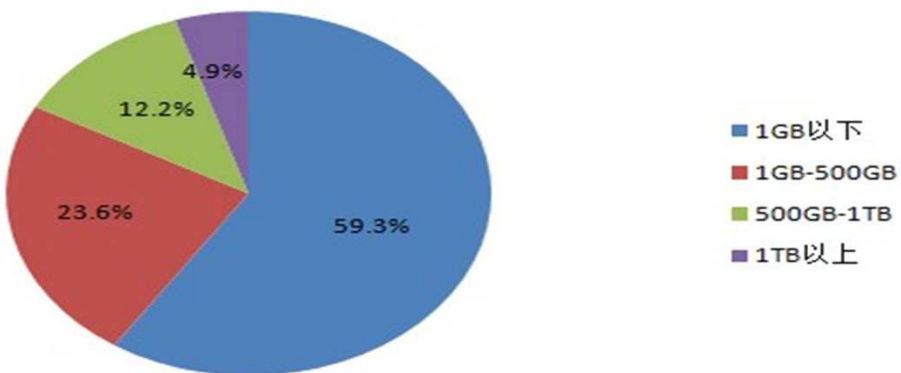
●计世资讯预测，2012年政府、互联网、电信、金融的大数据市场规模较大，四个行业将占据一半市场份额。

●由于各个行业都存在大数据应用需求，潜在市场空间非常可观。

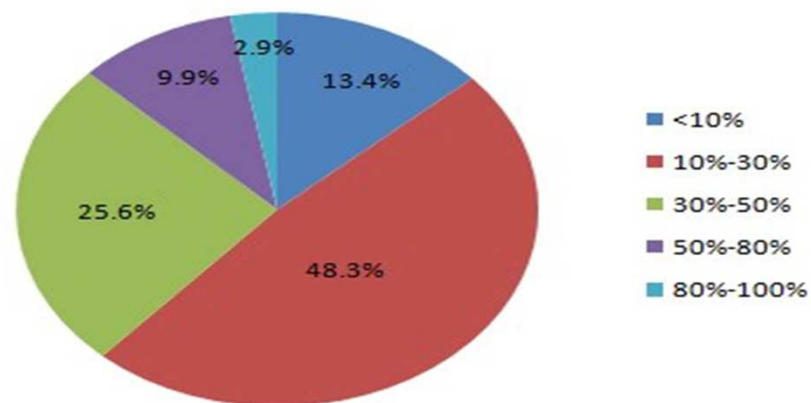


# 企业大数据现状

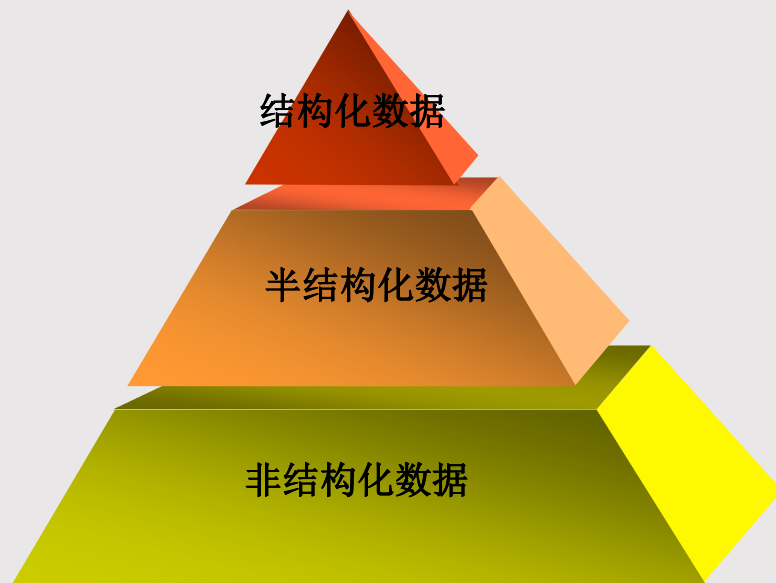
## 中国500强企业日数据生成量



## 中国500强企业数据中心数据年增长率



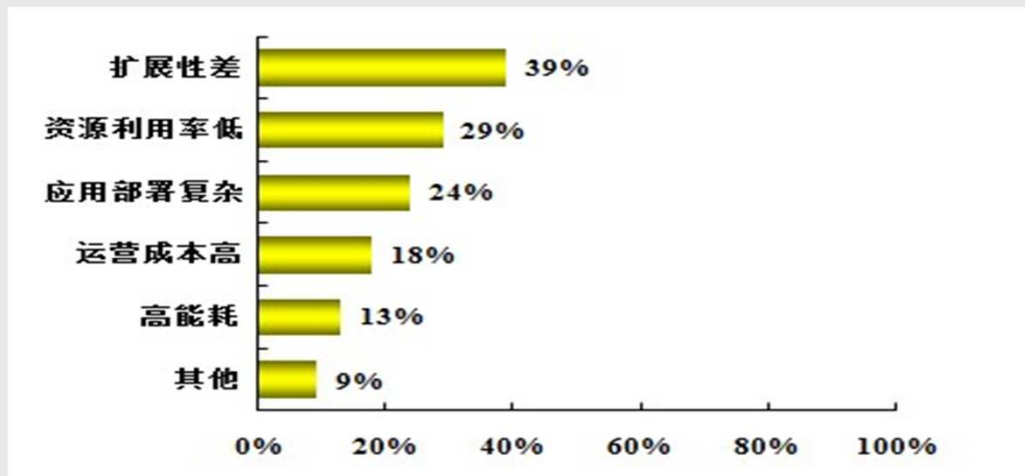
## 企业非结构化数据越来越多



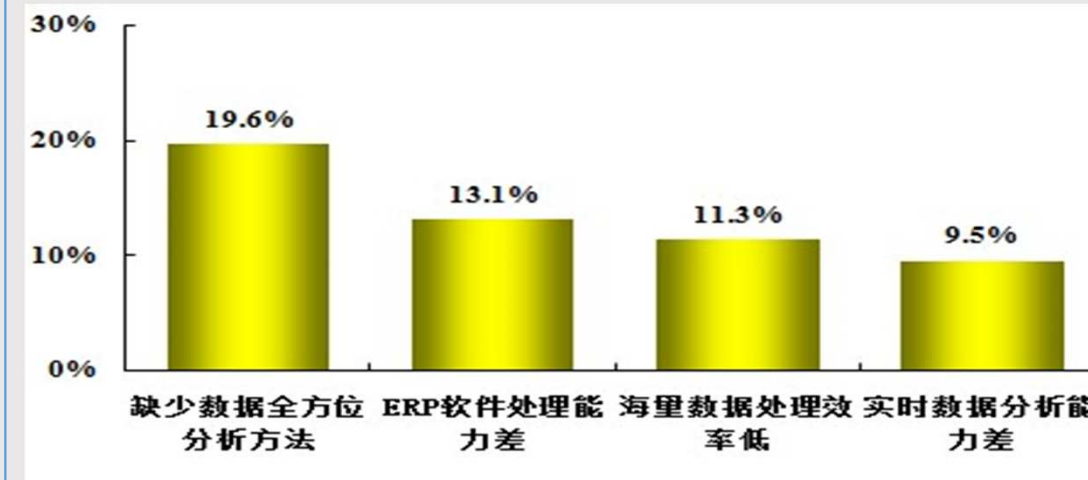
中国企业级数据中心数据存储量正在快速增长，非结构化数据呈指数倍增长，如果能有效的处理和分析，非结构数据中也富含了对企业非常有价值的信息。

# 大数据应用需求分析

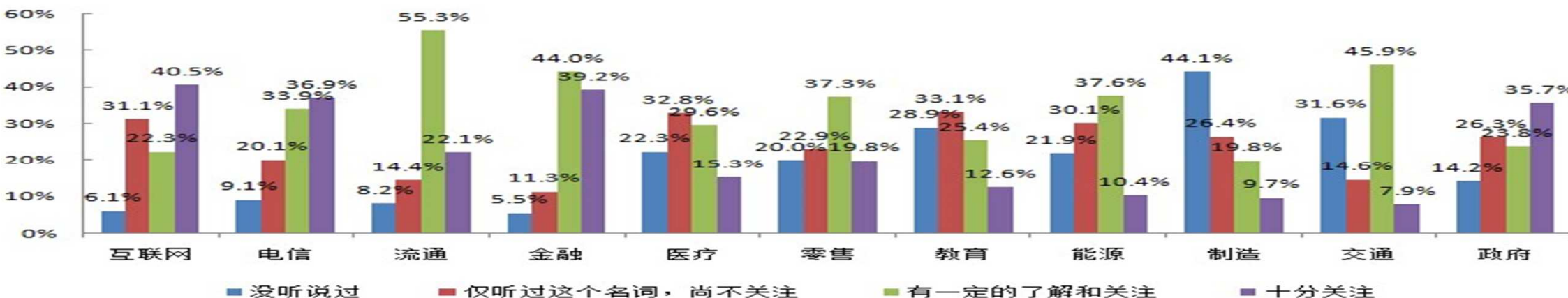
目前企业的数据系统架构存在问题



2 目前企业数据分析处理面临的问题

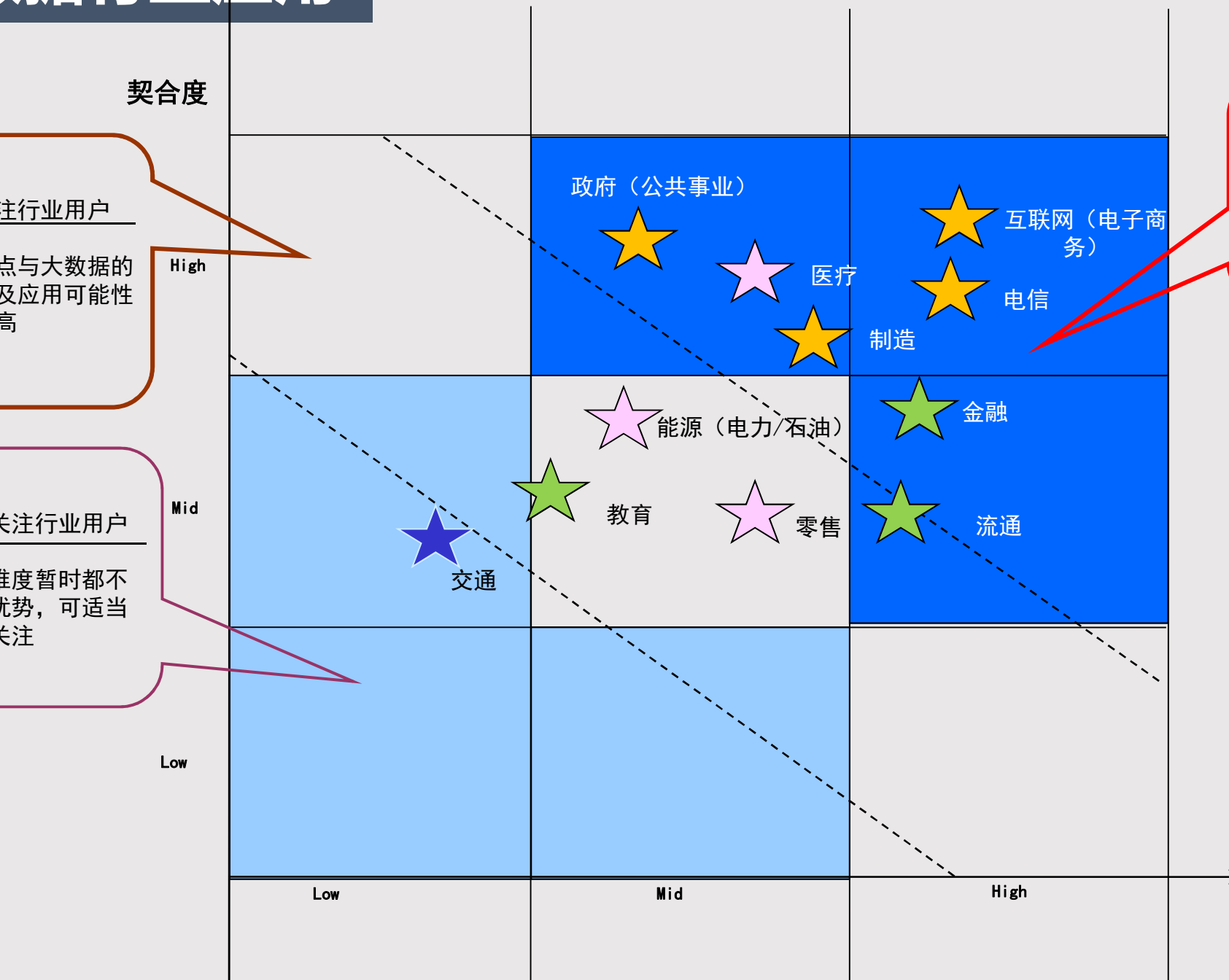


各行业企业对大数据的关注程度



数据来源: CCW Research, 2012/04

# 数据行业应用



优先关注行业用户

应用特点与大数据技术有较高的契合度，在主客观条件上也有较高的应用可能性。

• **纵轴契合度：**

表示该用户的IT应用特点与大数据特性的契合程度；

• **横轴应用可能性：**表示该用户出于主客观因素在短期内投资大数据的可能性；

• **注：**

该位置为分析师访谈的综合印象，为定性分析，图中位置不代表具体数值

关注行业用户

应用特点与大数据的契合度及应用可能性高

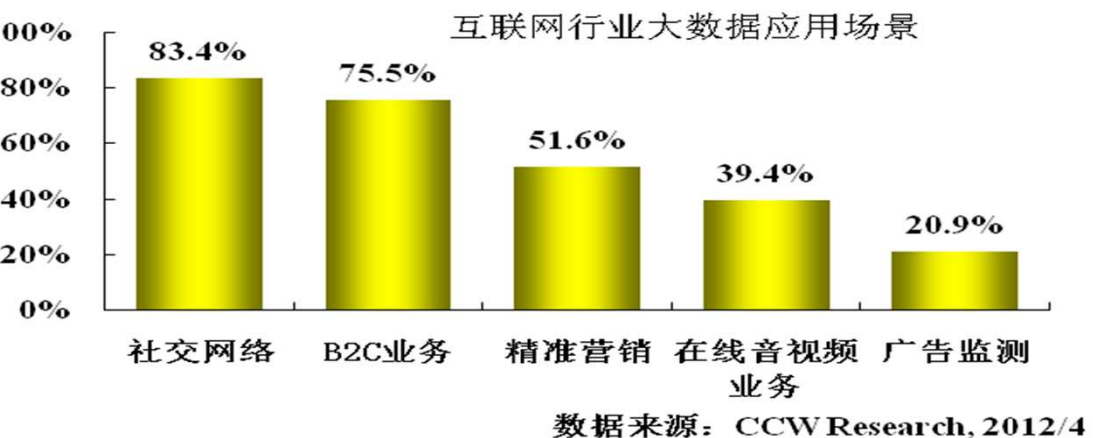
关注行业用户

契合度暂时都不占优势，可适当关注

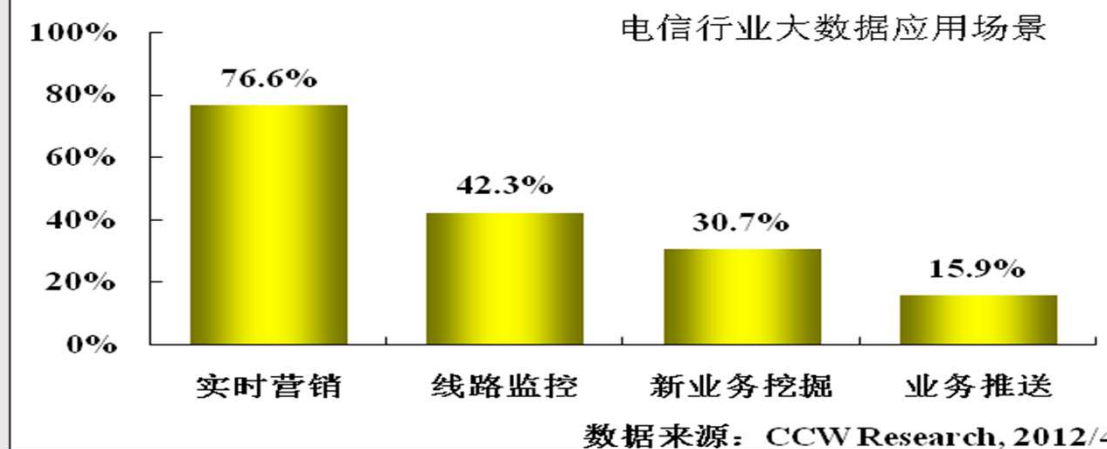
应用可能性

# 数据潜在应用

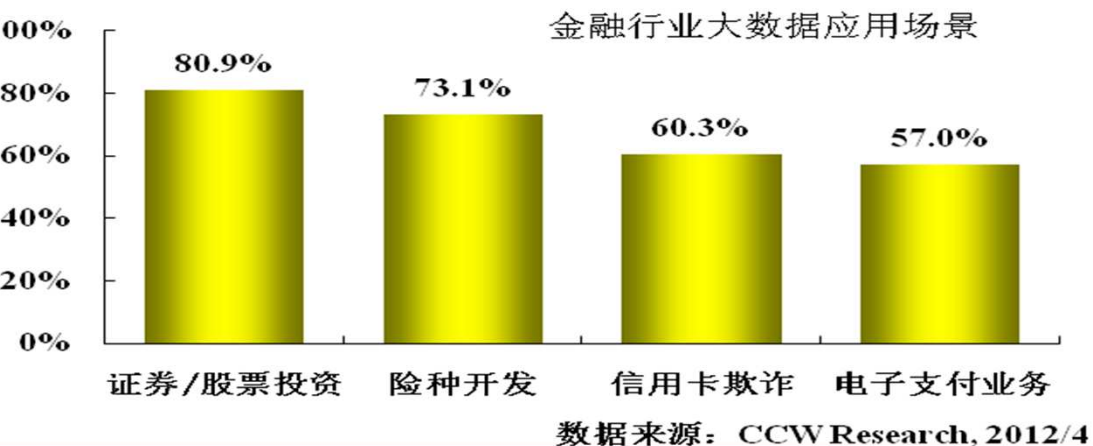
互联网行业大数据主要应用在社交和网购方面



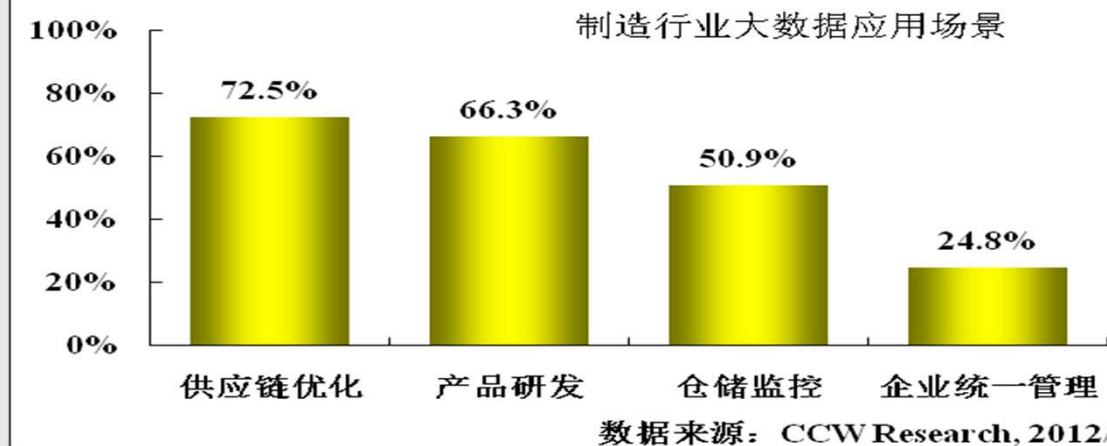
2 结合位置数据、消费数据进行实时营销信息推送是电信行业大数据应用主要场景



金融行业大数据应用场景主要集中在投资方面



4 制造行业具有多环节、多地域特色, 各个环节的优化是制造行业最关注的大数据应用场景



# 数据应用案例（中信银行信用卡中心）

## 大数据挑战

- 发卡量增长迅速：2008年发卡约500万张，2010年增加了一倍。
- 业务数据增长迅速：随着业务的迅猛增长，业务数据规模也线性膨胀。
- 数据存储、系统维护、数据有效利用都面临巨大压力。

## 需求

### 可扩展、高性能的数据仓库解决方案

能够实现业务数据的集中和整合；可以支持多样化和复杂化数据分析提升信用卡中心的业务效率；通过从数据仓库提取数据，改进和推动有针对性的营销活动。

EMC  
Green-  
plum

## 采用大数据方案后价值体现

### 实时的商业智能

可以结合实时、历史数据进行全局分析,风险管理部门现在可以每天评估客户的行为，并决定对客户的信用额度在同一天进行调整；原有内部系统、模型整体性能显著提高

### 秒级营销

Greenplum数据仓库解决方案提供了统一的客户视图，更有针对的进行营销。2011年，中信银行信用卡中心通过其数据库营销平台进行了1286个宣传活动，每个营销活动配置平均时间从2周缩短到2-3天。

# 数据应用案例（农夫山泉）

## 大数据挑战

- 农夫山泉数据量变得越来越大，分销表中数据基数大，增速快，数据展现速度越来越慢；
- 数据运算速度越来越慢，已经让人无法忍受，影响业务的正常进行；
- 数据更新慢，采用传统的ETL(数据抽取、转换、装载)，农夫山泉的分析系统数据基本上一天才能更新一次。

## 需求

能够应对海量数据的挑战，实现高效的逻辑运算、实时的数据分析以及快速的数据展现的解决方案。

SAP  
HANA

## 采用大数据方案后价值体现

**实现了快速的数据展现**

与原有商业智能报表展现方案相比，新方案数据展现速度快25-30倍；

**形成了强大逻辑计算能力**

测试了120多张已经上线的报表，基本上速度提升100~150倍；SAP HANA和Business Objects 4.0组合只用了46秒就完成原来需要24小时才能完成的逻辑计算；

**实现了数据的实时、同步**

HANA使得数据从业务系统中转换到HANA中时基本上没有任何延迟。

# 数据应用案例（“数字黄河”）

## 大数据挑战

- 数据激增，IT系统负担加重；
- 地域分隔，信息孤岛拉低效能；
- 无法共享，数据同步成为难题；
- 标准各异，数据规范有待统一。

## 需求

制定短期和长期技术规划，以适应未来信息系统的发展。

IBM  
Info-Sphere

## 采用大数据方案后价值体现

### 解决跨平台异构应用系统的数据共享与集成问题

黄河水利委员会各部门随时获取其权限范围内的最新数据，而无须将其存储在本部门系统中；  
**消除信息孤岛，实现数据统一管理**

有效消除了各业务系统和各组织结构之间的信息孤岛，简单获取黄河数据资源的单一视图，并确保了数据的完整性、及时性、准确性和一致性，同时首次实现元数据的可视化统一管理

## 京东利用大数据做用户需求分析

**买2000元手机的是什么样的人呢？**

他们是最懂手机的一群人，对手机的软硬件配置要求高。

手机芯片绝对**要求最新的四核芯片**；

屏幕尺寸绝不是越大越好，**5英寸屏幕加超窄边框**是他们的最佳选择；

屏幕分辨率必须要达到**1080P级别**，并且要是最好的夏普屏幕；

拍照方面除了像素最基本要达到**前置500万+后置1300万像素**的配置以外，更追求摄像头的品牌.....

**在大数据分析下，JDPHONE与中兴进行C2B订制，产生国母手机！**

所谓数字图书馆，就是对有价值的图像、文本、语音、影视、软件和科学数据等多媒体信息进行收集、组织和规范再加工，通过网络提供高速横向跨库连接的多媒体信息存取服务，促进社会各类信息高效、经济地传递，从而极大地方便人们的学习、交流和生活；

数字图书馆是具有服务功能的整理过的信息收藏，其中信息以数字化格式存储并可通过网络存取。该定义的关键在于信息是整理过的。

美国著名数字图书馆专家 **William Arms**

数字图书馆产生与发展主要源于两种力量的推动

一是图书馆自身资源共享的内在要求；

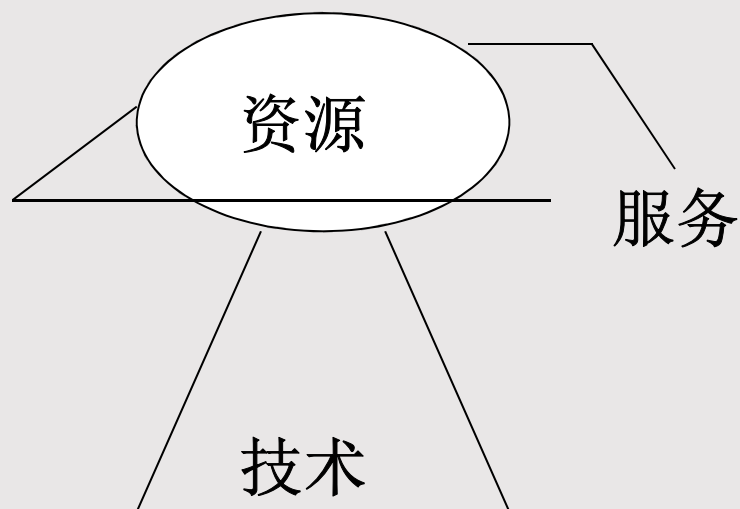
二是计算机互联网络的发展对数字信息进行有序化、结构化组织的要求。

**A、数字图书馆产生的内在因素**

**B、数字图书馆产生的外在因素**

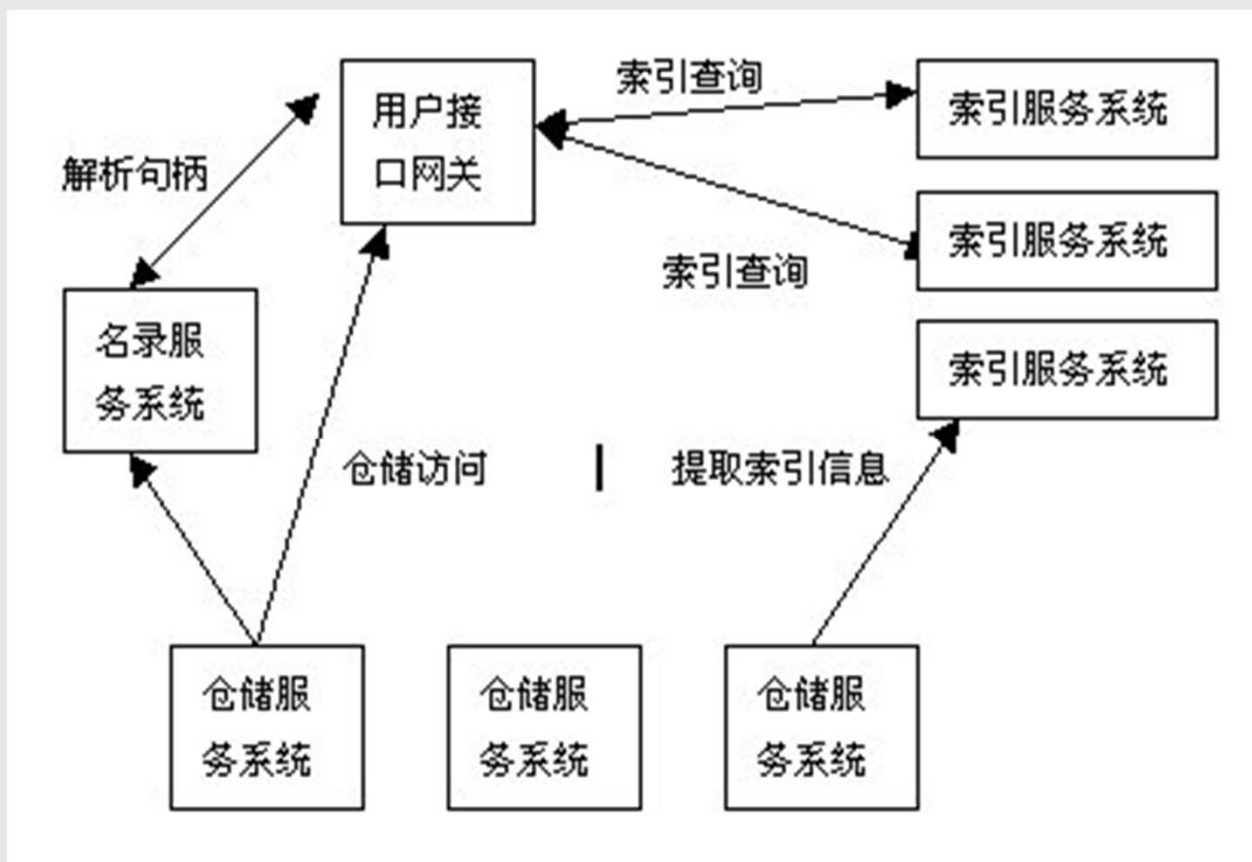
**C、数字图书馆发展的社会背景**

# 数字图书馆的形象模型

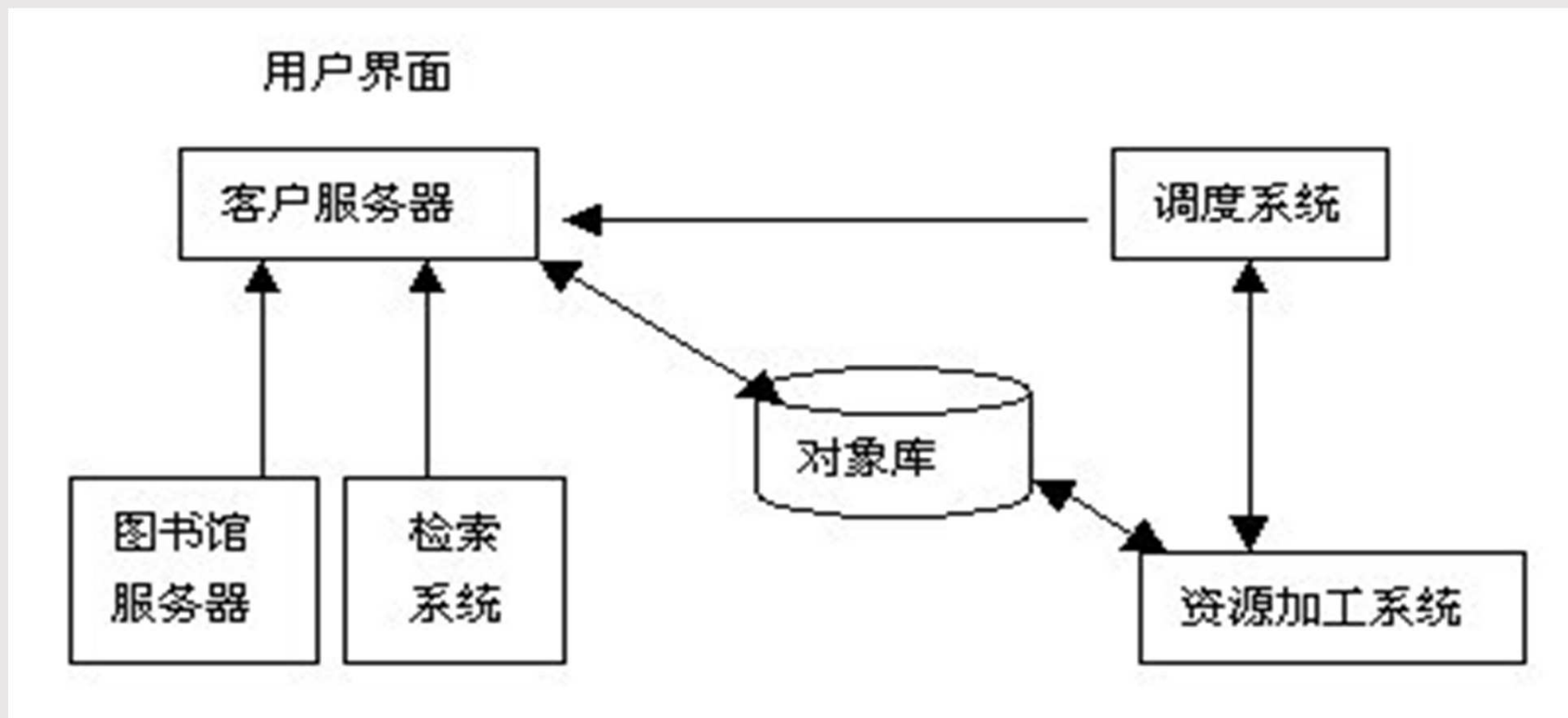


资源是数字图书馆赖以存在的基础，技术是数字图书馆得以成立的条件，服务是数字图书馆努力追求的目标，这是一个以资源为核心、技术为支撑、服务为目的的数字图书馆理论形象模型。

# 图书馆的信息体系结构



# 图书馆系统的结构模型



这个面向对象的、分布式的网络结构模式是由用户界面、预处理系统（调度系统）、查询系统、图书馆服务器和对象库五个部分组成的数字图书馆系统



# 大学城图书管理系统



深圳市科技图书馆  
深圳大学城图书馆  
SHENZHEN SCIENCE & TECHNOLOGY LIBRARY /  
UNIVERSITY TOWN LIBRARY OF SHENZHEN



中文

English

欢迎登陆深圳大学城图书馆

用户名:

密 码:

记住用户名和密码

登 录

超星入口

# 大学城图书管理系统



## 信息资源

数据库列表  
有任何问题请[联系我们](#)

当前位置: 数据库列表 (有任何问题请[联系我们](#))

- ACS (美国化学学会)
- ASCE Online Research Library
- arXiv预印本
- ASME (美国机械工程师协会)
- 北大法宝
- BNA数据库
- CALIS中文现刊目次库
- CNKI《中国学术网络出版总库》
- Derwent Innovation Index (DII)
- EBSCO (ASP、BSP等)
- Ei Village
- Ebrary 电子图书
- Elsevier ScienceDirect OnSite (SDOL)
- 方正Apabi电子图书
- 国研网
- HeinOnline法律数据库
- IEEE/IET Electronic Library (IEL)
- Investment Arbitration Reporter (IA Reporter)
- APS (美国物理学会)
- ASTM电子图书数据库
- Annual Reviews数据库
- APA PsycINFO & PsycARTICLES
- Berkeley Electronic Press 10种电子期刊
- 北大法宝英文版 (ChinaLawInfo)
- Cambridge Scientific Abstracts (CSA)
- Cambridge University Press (CUP)
- 读秀图书搜索
- Encyclopedia Britannica Online
- Emerald
- EMIS
- Wiley InterScience 在线工具书
- Gale
- H. W. Wilson法律期刊索引全文数据库
- IWA国际水协会期刊
- IMechE英国机械工程师协会数据库
- Investment Claims

# 如何应对大数据时代？



## PC互联网

## 移动互联网

手机与社交网络从时空和社会情境两个维度对我们进行了双重锁定。你不是不是一条狗，网络比你更知道。

社交网络让我们越来越多地从数据中观察到人类社会的复杂行为。

# 隐私：它什么都知道

## 你的什么大数据都知道

你最近手头不宽裕

你可能会得高血压

你两年后可能要买车

邻居上高中的女儿怀孕了

特别讨厌某个明星

喜欢日本爱情动作片

是一个爱国的屌丝

旧手机坏了想换新的

女朋友有了一个新手提包



你在这个世界上留下了  
太多的痕迹！！！！



铁路客服中心



国家电网  
STATE GRID

服务热线 95598



淘宝网  
Taobao.com

Google

拉手网  
Lashou.com

京东商城  
360buy.com



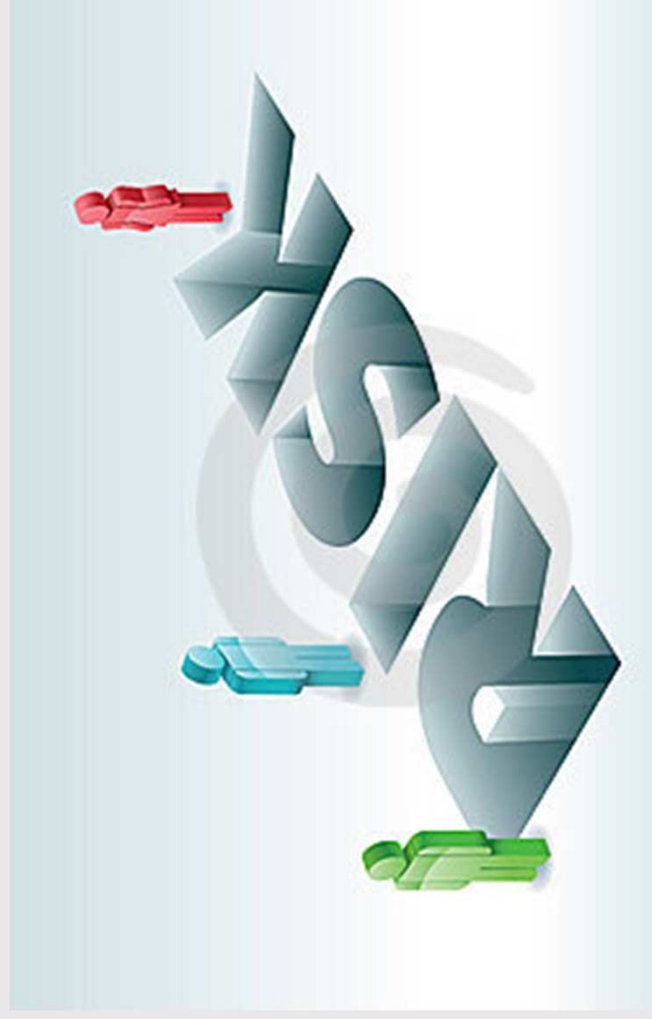
中国移动通信  
CHINA MOBILE



ICBC

中国工商银行

# 大数据带来的风险!



輿情监测、人肉搜索  
网络暴民！

## “风险” —— 让数据主宰一切的隐忧

我们时刻都暴露在“第三只眼”之下：亚马逊监视着我们的购物习惯，谷歌监视着我们的网页浏览习惯，而微博似乎什么都知道，不仅窃听到了我们心中的“TA”，还有我们的社交关系网。

- 无处不在的“第三只眼”
- 我们的隐私被二次利用了
- 预测与惩罚，不是因为“所做”，而是因为“将做”
- 数据独裁
- 挣脱大数据的困境

## “掌控” —— 责任与自由并举的信息管理

当世界开始迈向大数据时代时，社会也将经历类似的地壳运动。在改变人类基本的生活与思考方式的同时，大数据早已在推动人类信息管理准则上重新定位。然而，不同于印刷革命，我们没有几个世纪的时间去适应，我们也许只有几年时间。

- **管理变革1：个人隐私保护，从个人许可到让数据使用者承担责任**
- **管理变革2：个人动因VS预测分析**
- **管理变革3：击碎黑盒子，大数据程序员的崛起**
- **管理变革4：反数据垄断大亨**

## “精细化”——完美掌控资源下的执行力

当世界开始迈向大数据时代时，企业拥有和掌控的资源更多，更透明。需要改变的是企业的执行力，如何将精准的分析人性化执行到位！

- **精细化1：企业客户行为的精准分析**
- **精细化2：企业员工执行力、行为规范、现场管理**
- **精细化3：社会公共管理、文化习惯**
- **精细化4：行之有效的流程管理、法治建设**
- **拒绝：粗放式管理、拍脑袋决策、权责利不清！**

谢谢观赏！