



大数据是什么



啤酒与尿不湿

啤酒与各类小零食

调料与一次性桌布和胶皮手套

饼干与湿纸巾

Walmart 
Save money. Live better.





亲！你喜欢吗？

生活研究所 发现属于你的生活方式



家有萌娃
宝贝快快长大

♥ 人气1196514



萌宝学院
老师早上好

♥ 人气176994



华丽上班族
职场升职记

♥ 人气593378



吃货的后裔
翻滚吧美食君

♥ 人气397931



DIY达人
就爱动手折腾

♥ 人气158202



海派妈咪
囤货第一波

♥ 人气240407

有好货 与品质生活不期而遇

换一换

爱逛街 献给聪明可爱的你

更多 >



新款近视偏光太阳镜
此款太阳镜针对男性睿智、

☺ 58 人说好



优力优儿童成长自律表
培养孩子好习惯，增强自律

☺ 0 人说好



小涂涂儿童生活学习记
为孩子记录下成长的点点滴滴

☺ 1 人说好



“ 伊利 无菌砖纯牛奶
250ml*24盒

潮流女孩girl



“ 伊利原味优酸乳
250ml*24盒

爱逛***01



“ 590片有背胶小红花奖励
贴 幼儿园儿童表扬贴海绵

爱逛***01



Facebook之选





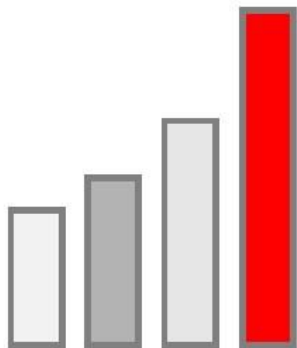
提问

“大数据” 是什么？

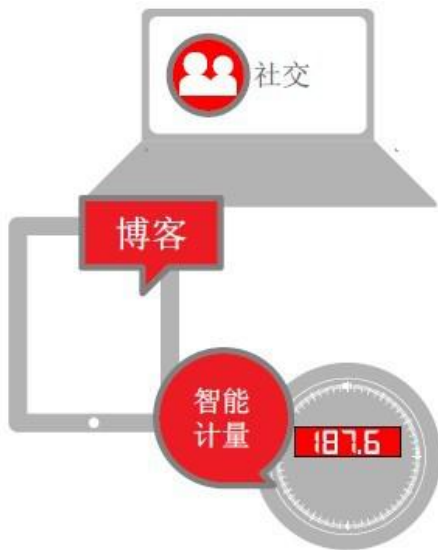
- 目前对大数据尚未有一个公认的学术定义
- 不同的机构基本上是从大数据的特征出发给出定义
- 大数据很抽象，表示数据规模的庞大，但不只是“大”



大数据4V特征



数据量
Volume



多样性
Variety



速度
Velocity



值
Value



1、数据体量巨大 Volume

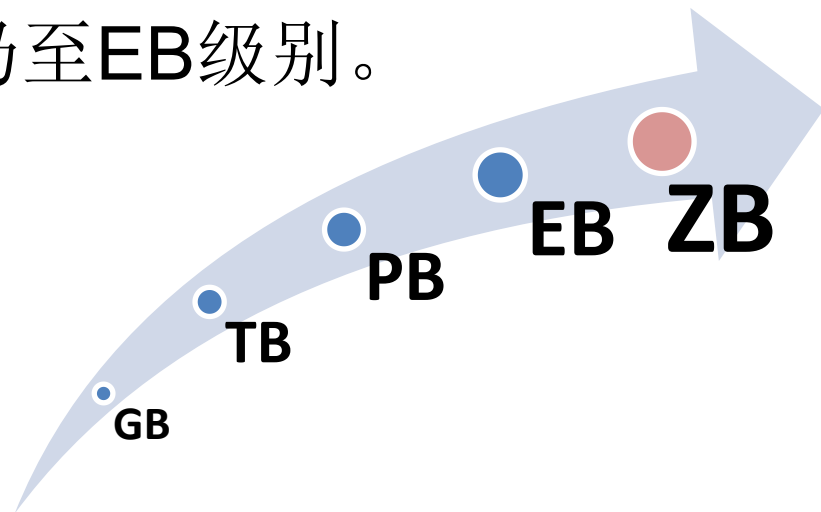
- 从TB级别跃升到PB乃至EB级别。

1TB=1024GB

1PB=1024TB

1EB=1024PB

1ZB=1024EB



每天交易数据

20_{PB} = 20971520[↑]_{GB}



2、数据类型繁多 Variety

- 数据类型多样
非结构化数据为主 85%
- 新型数据
网络日志
音频
视频
图片
地理位置信息
.....





3、产生和处理速度快 Velocity

- 数据产生快：预计**2020**年全球数据使用量将达到**35.2ZB**
- 数据处理快：**1秒定律**，在秒级时间范围内给出分析结果





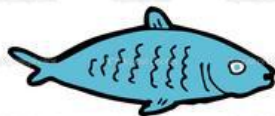
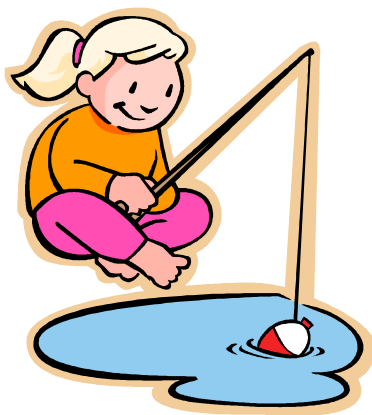
4、价值密度低 Value

- 价值密度的高低与数据总量的大小成反比
- 迅速地完成数据的价值“提纯”是目前大数据汹涌背景下亟待解决的难题

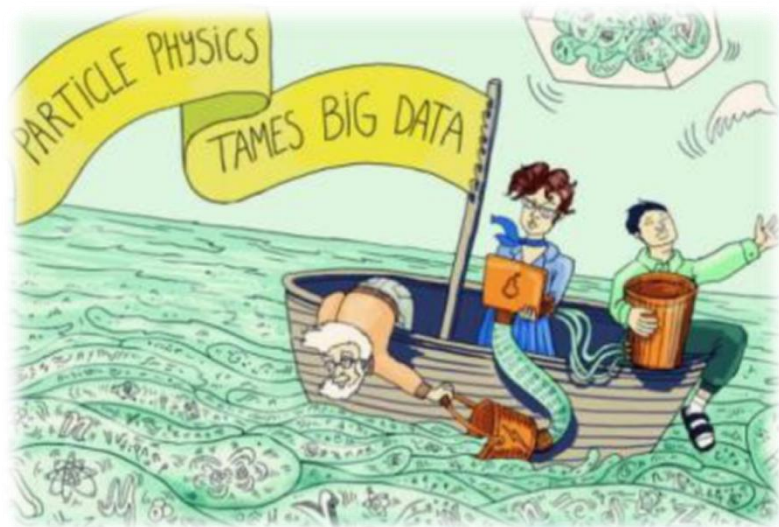




怎样的数据才能称为“大数据”？



待处理数据



数据库 (database, DB)
鱼塘捕鱼

VS

大数据 (big data, BD)
大海捕鱼

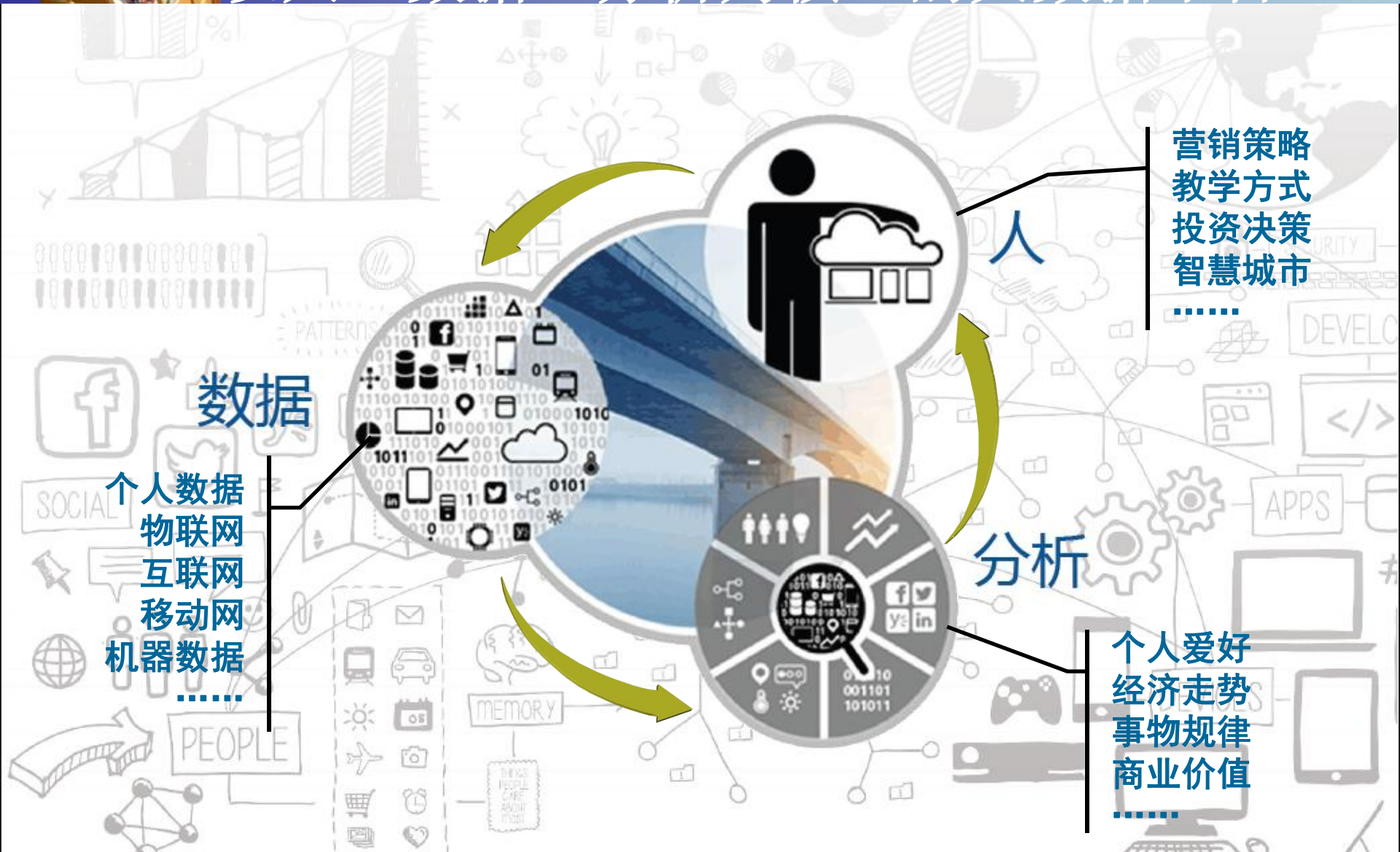
处理工具

先有模式后有数据
(先有池塘后有鱼)

先有数据后有模式
(先有大海后有鱼)



以人、数据、分析为核心的大数据平台





回答

“大数据”是什么？

维基百科给出的定义：

大数据是指利用常用软件工具捕获、管理和处理数据所耗时间超过可容忍时间的数据集。

大数据是下一个自然资源

“我们首次拥有一个基于关键资源（信息）的经济体制，不仅是可再生的而且是自发生成的。耗尽它没关系，但不要淹没在其中。”

——John Naisbitt



人类应该如何对待大数据



机遇

- 大数据技术促进国家和社会发展
- 大数据成为企业竞争的新焦点
- 大数据时代呼唤创新型人才



挑战

- 大数据技术的运用仍有困难
- 大数据给信息安全带来新挑战



机遇：国家战略

自2014年以来，“大数据”连续六年进入国务院政府工作报告，上升为国家战略。

实施国家大数据战略 建设数字中国

光明日报
百家号 | 01-28 09:03

2017年12月8日下午，中共中央政治局就实施国家大数据战略进行第二次集体学习。中共中央总书记习近平在主持学习时强调，大数据发展日新月异，我们应该审时度势、精心谋划、超前布局、力争主动，深入了解大数据发展现状和趋势及其对经济社会发展的影响，分析我国大数据发展取得的成绩和存在的问题，推动实施国家大数据战略，加快完善数字基础设施，推进数据资源整合和开放共享，保障数据安全，加快建设数字中国，更好服务我国经济社会发展和人民生活改善。

李克强对大数据在经济政策中发挥的作用的 **最** 明晰表述





挑战：棱镜门引爆大数据时代争议

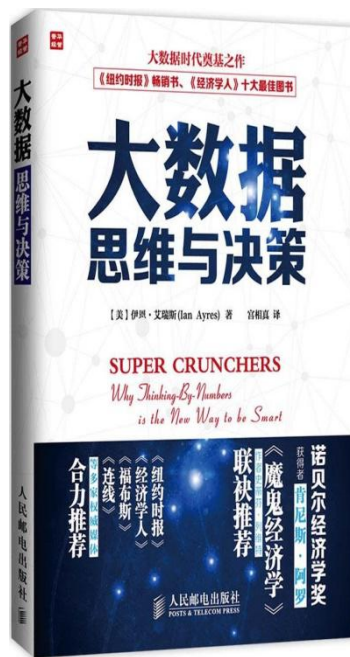
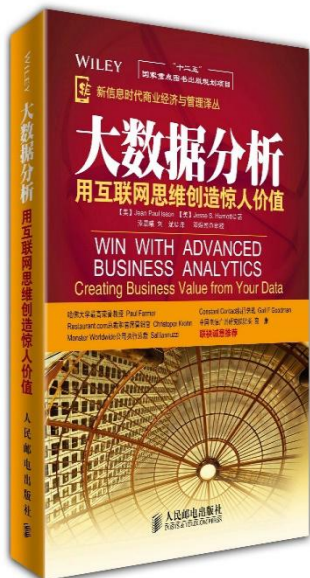
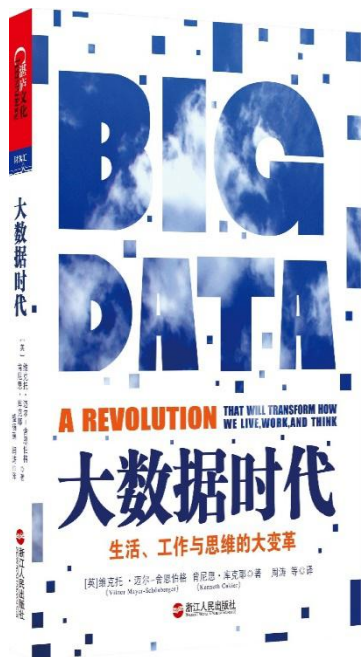


这是一个最好的时代，这是一个最坏的时代
不仅仅是隐私的泄露，还有被预知的可能性

大数据能预测我们可能生病、拖欠还款和犯罪的算法会让我们无法购买保险、无法贷款、甚至在实施犯罪前就被预先逮捕。统计把大数据放在了首位，但即便如此，个人意志是否应该凌驾于大数据之上呢？利用大数据的同时，**遵守工程伦理。**



与大数据相关的学习资料——图书





与大数据相关的学习资料——微博

数据分析故事



数思网



**中国互联网
研究院**



数据魅力



**社会网络与
数据挖掘**





与大数据相关的学习资料——微信公众号

网络大数据



行业大数据分析



大数据有大智慧



电商大数据

