01

第1章 大数据技术概述

授课教师：[李阳源]
ywlee1212@gmail.com
主页：https://sites.google.com/view/dalguri



第1章 大数据技术概述


BIG
DATA
ANALYTICS
COMPLEX
SOCIAL
INTERNET
TECHNOLOGIES
STORAGE
SYSTEMS
SOFTWARE
INFORMATION
DATABASES
NETWORKS
RECORDS
SEARCH
PARALLEL
SIZE
RESEARCH
PETABYTES
INCLUDE
MANAGEMENT
SETS
SHARED
DISK
EXAMPLES
EVERY
ALSO
TOOLS
TARGET
SENSOR
DEFINITION
MOVING
MANAGE
ARCHIVES
PRACTITIONERS
CAPTURE
BUSINESS
NEEDED
TIME
LARGE
ONE
PERFORMANCE
STORE
LOGS
COST
SAN
MASSIVELY
GROW
DIFFICULTY
GARTNER
WORKING
ELAPSED
TOLERABLE
ABILITY
CASE
TERABYTES
DISTRIBUTED
GENOMICS
CURRENT
MAY
WITHIN
USING
PROCESS
RELATIONAL
CONTINUES
LARGER
TENS
CITATION
BIOLOGICAL
PROCESSING
CAPACITY
HUNDREDS
SET
USE

目录

CONTENT

- 大数据时代
- 大数据概念
- 大数据的影响
- 大数据关键技术
- 大数据计算模式
- 代表性大数据技术



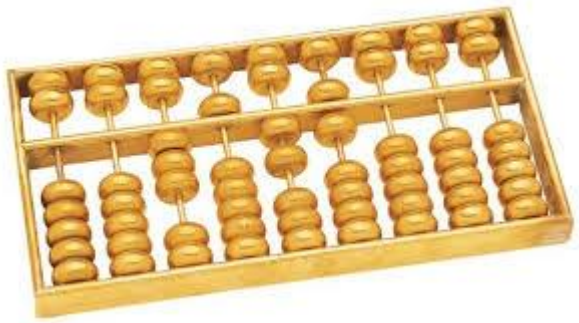
1.3 大数据的影响

提纲

1 大数据对科学研究的影响

2 大数据对思维方式的影响

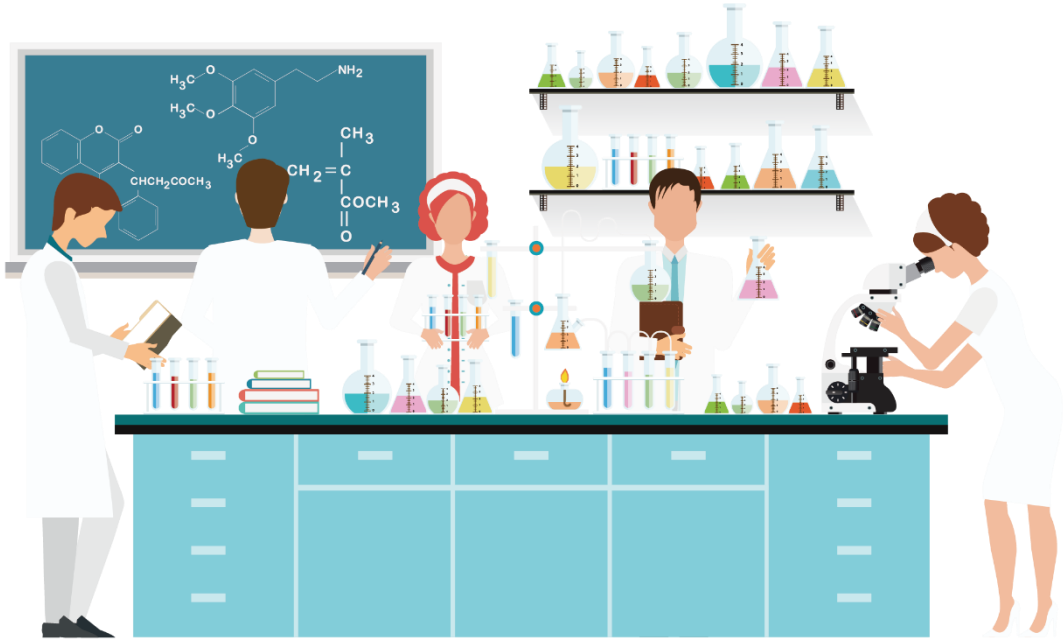
1.3 大数据的影响





大数据时代的开启

1.3 大数据的影响



科学研究进入一个全新的时代

1.3 大数据的影响



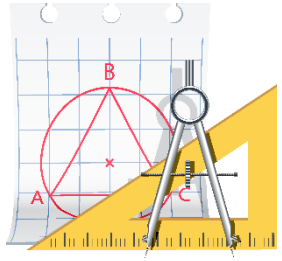


实验

理论

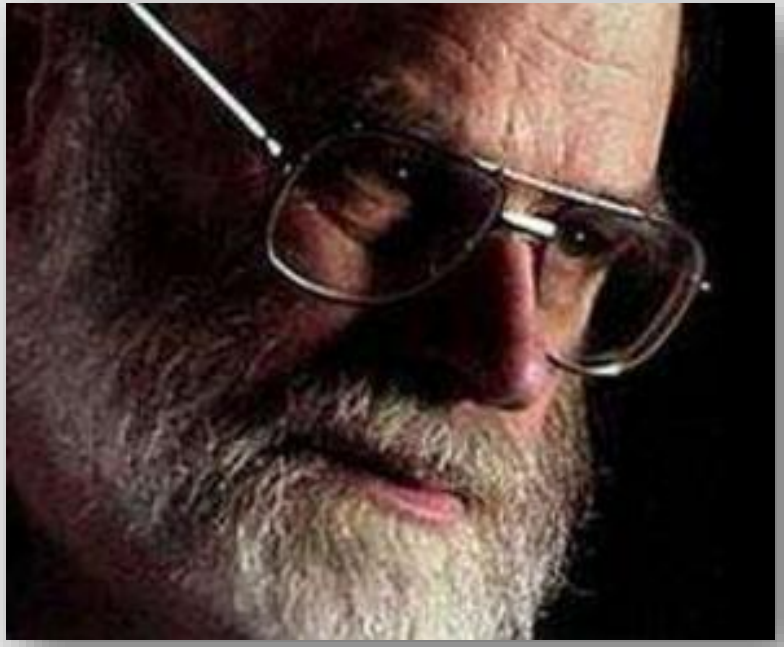
计算

数据





1.3 大数据的影响



数据库专家Jim Gray 博士

进入大数据时代以后

是以数据为中心的研究范式

1.3 大数据的影响

科学研究第一种范式——实验



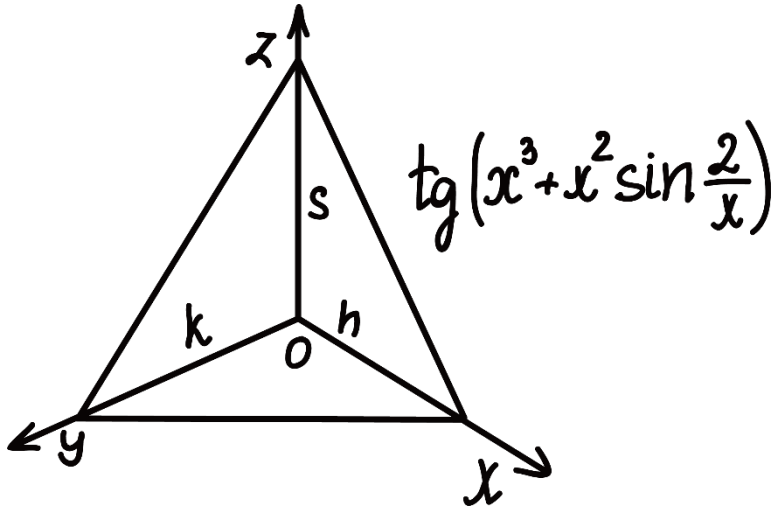
伽利略



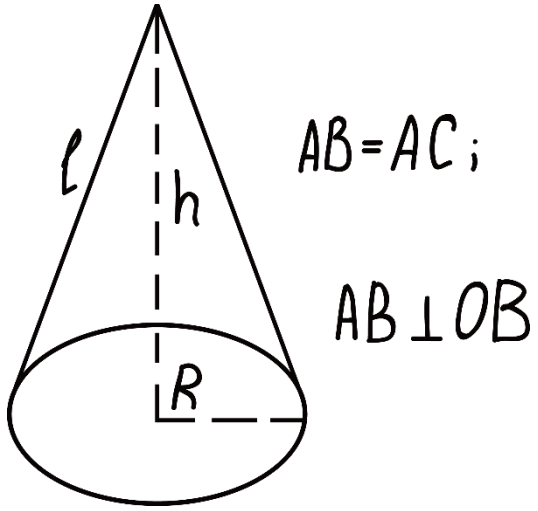
在比萨斜塔做两个铁球同时落地实验

1.3 大数据的影响

科学研究第二种范式——理论



$$tg\left(x^3+x^2\sin\frac{2}{x}\right)$$



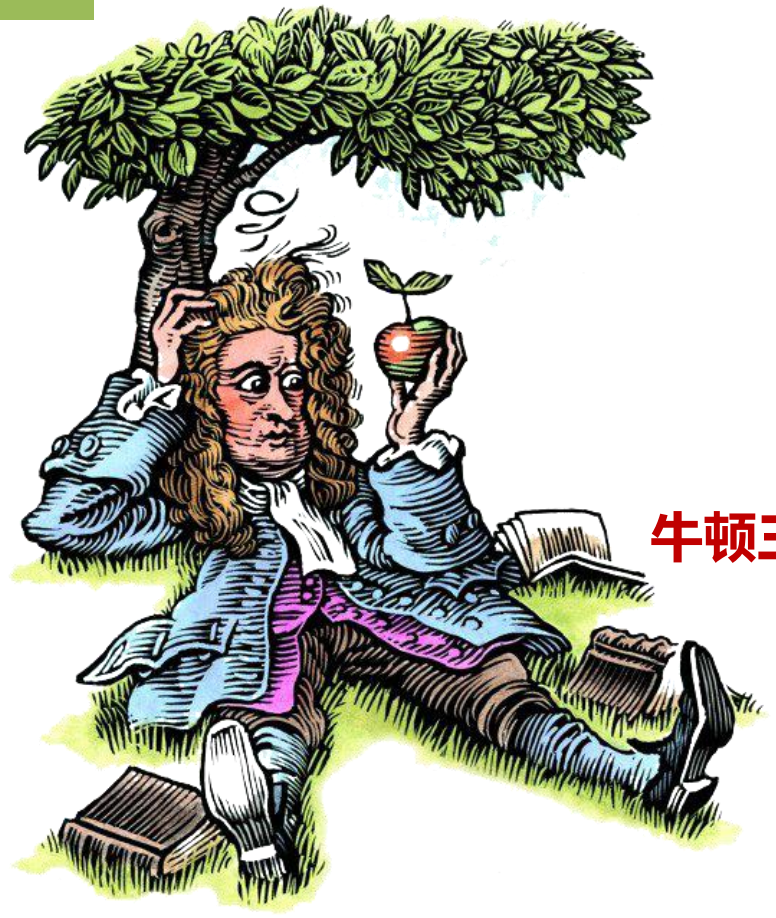
$AB=AC;$

$AB\perp OB$

几何理论

1.3 大数据的影响

科学研究第二种范式——理论



牛顿三大定律

1.3 大数据的影响

科学研究第三种范式——计算



天河二号



互联网

1.3 大数据的影响

科学研究第四种范式——数据



大数据时代以数据为中心

1.3 大数据的影响



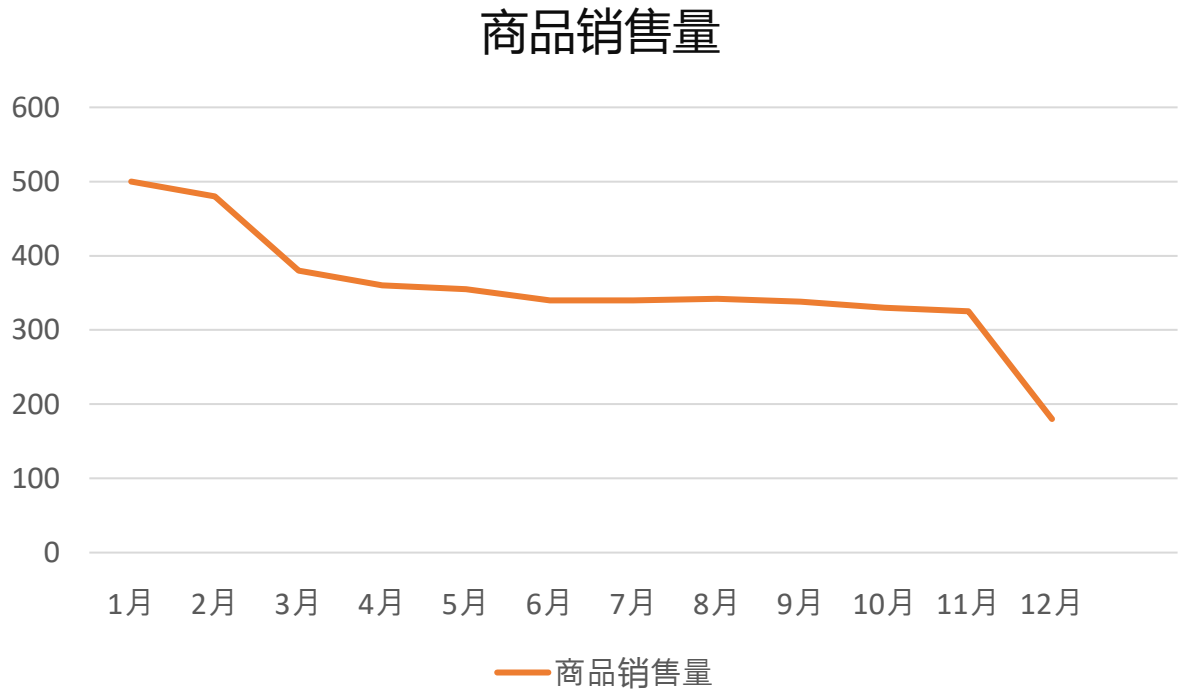
以数据为中心的科学研究时代

差别



以计算为中心的科学研究时代

1.3 大数据的影响



12个月之内某件商品的销量不断下滑

1.3 大数据的影响



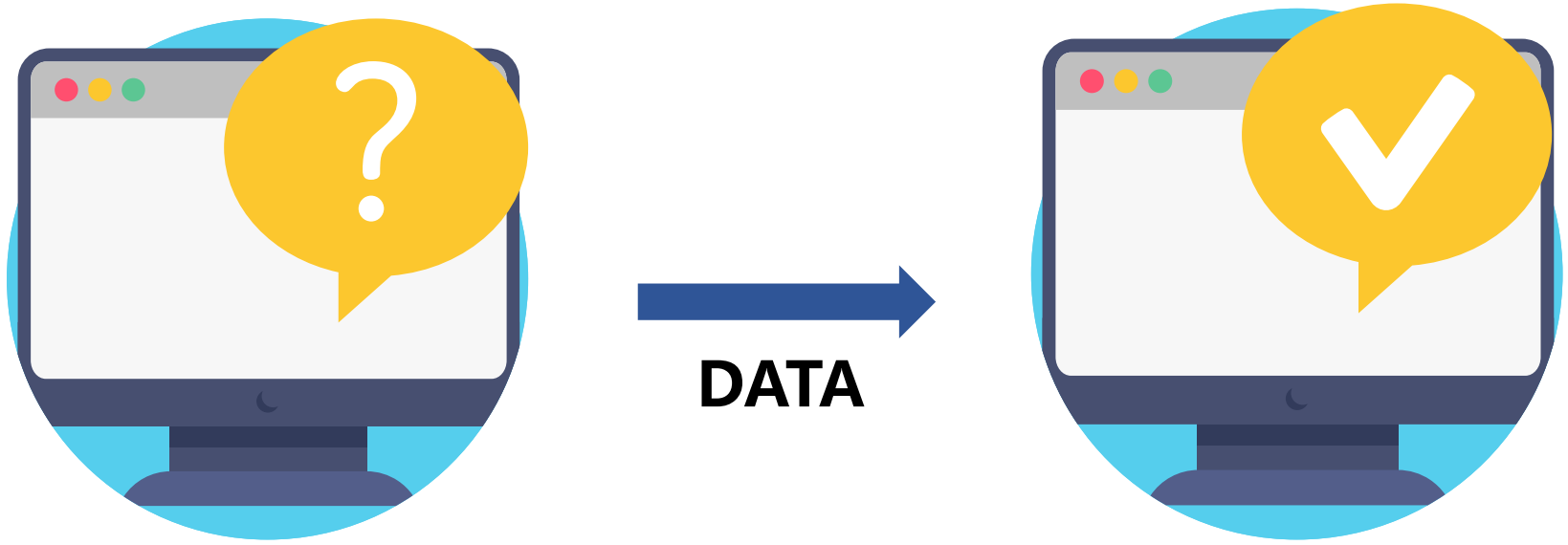
通过编程分析商品销量持续下滑的原因

1.3 大数据的影响



从大量数据当中去发现问题

1.3 大数据的影响



发现问题 解决问题 全部靠数据来驱动

1.3 大数据的影响



大数据时代在思维层面有什么影响呢?

1.3 大数据的影响



《大数据时代》

维克托·迈尔·舍恩伯格 著

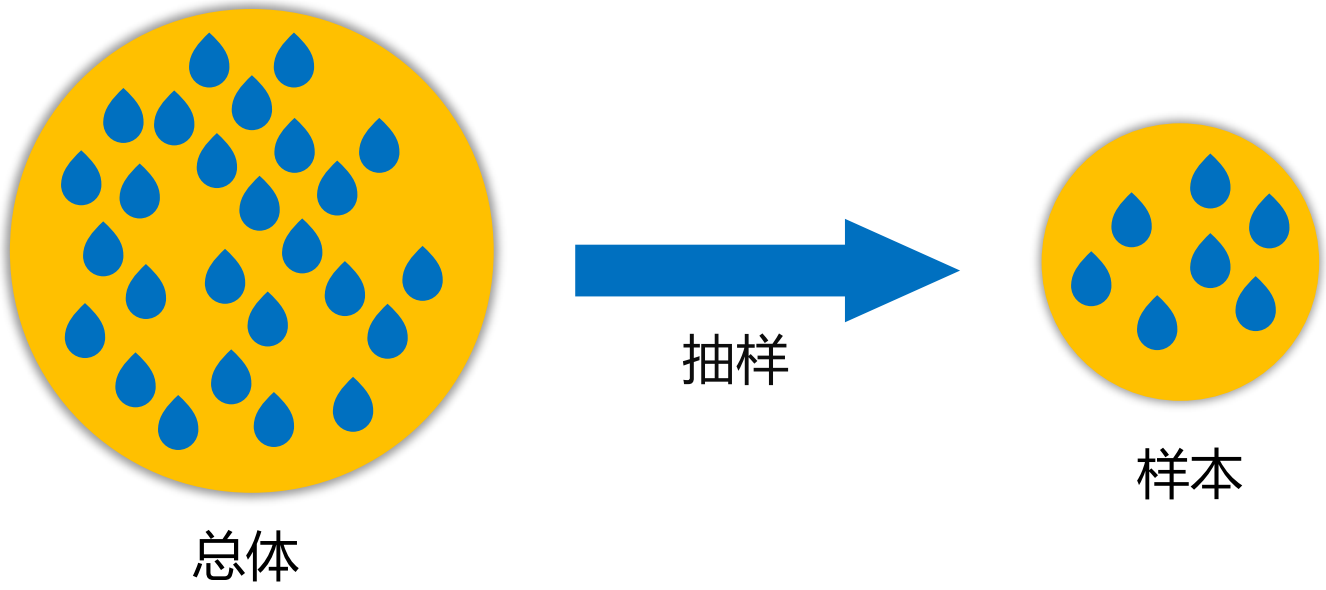


1.3 大数据的影响



1.3 大数据的影响

- 全样而非抽样



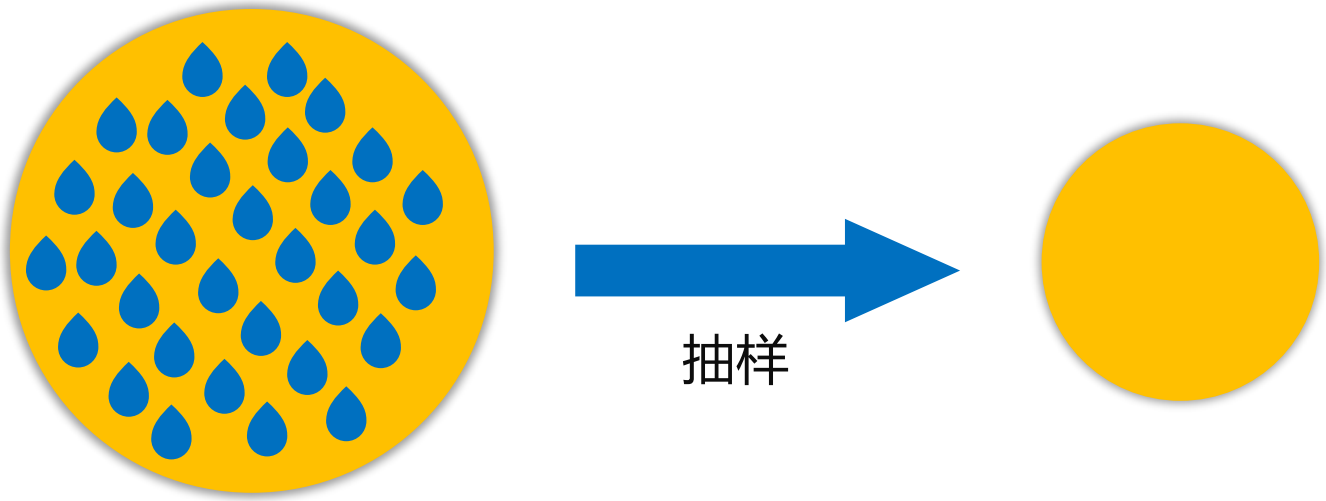
1.3 大数据的影响

采用统计学抽样方法的原因



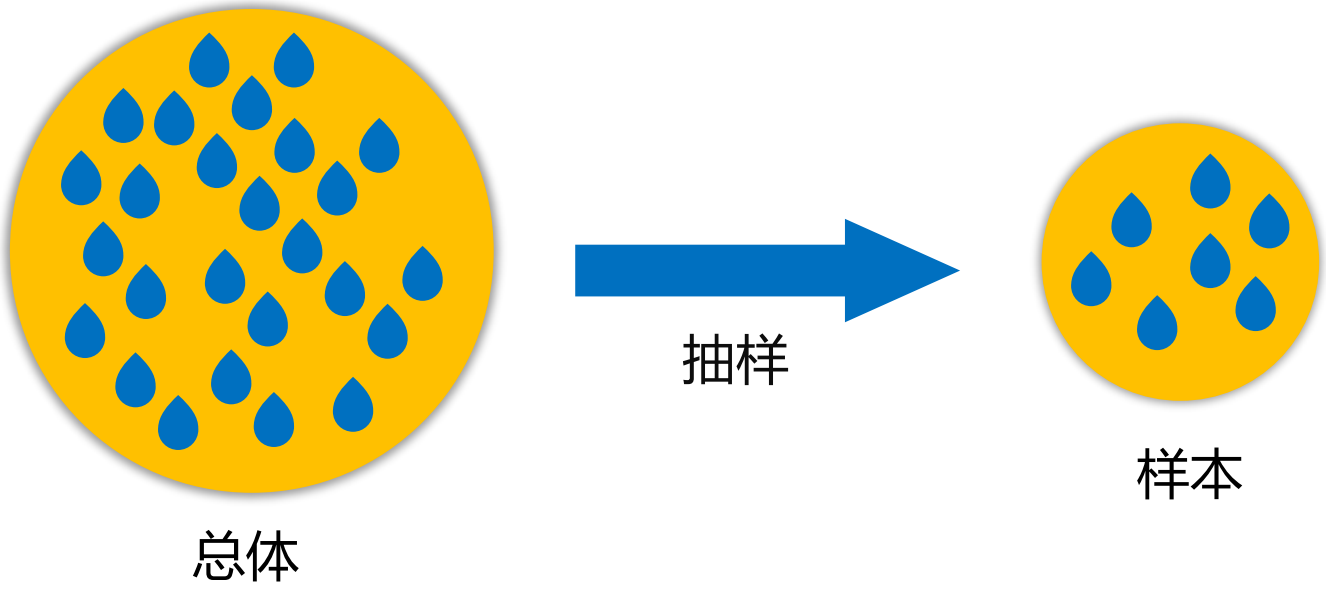
1.3 大数据的影响

- 全样而非抽样



1.3 大数据的影响

- 全样而非抽样



1.3 大数据的影响

- 全样而非抽样



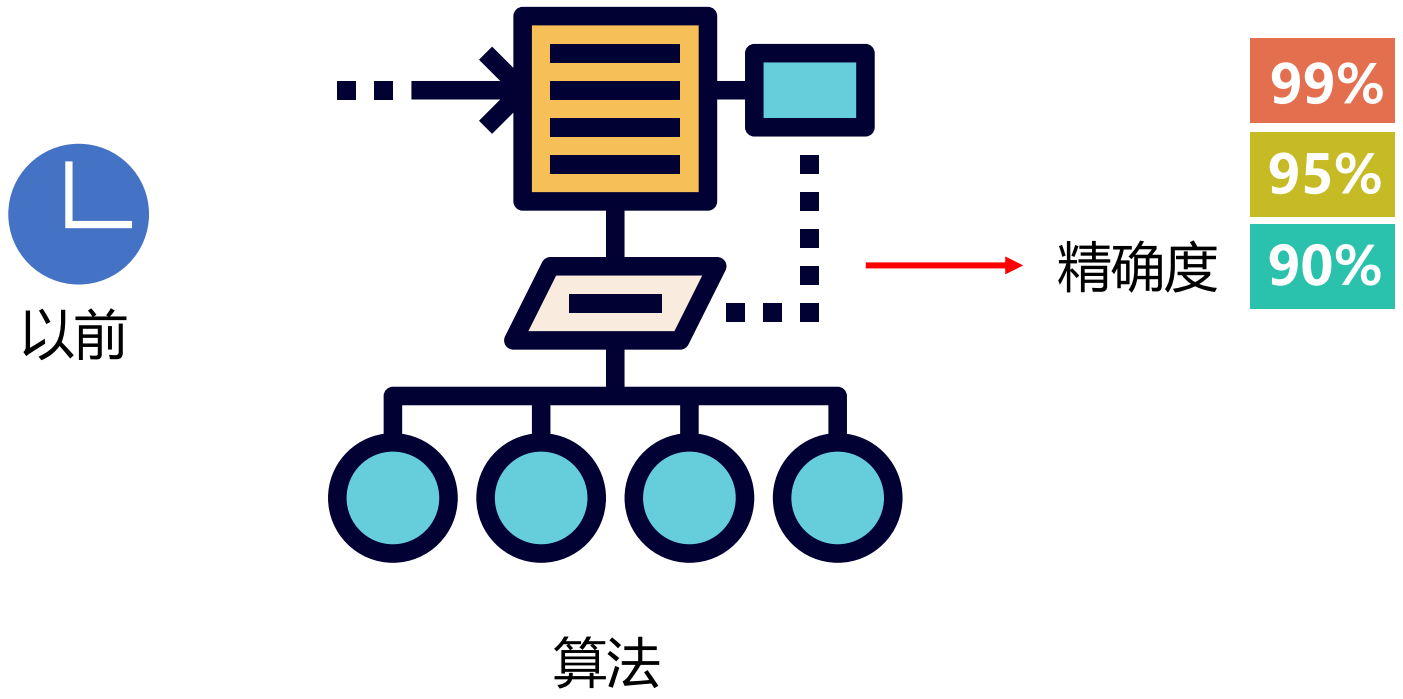
包含众多CPU的服务器集群

1.3 大数据的影响



1.3 大数据的影响

- 效率而非精确



1.3 大数据的影响

- 效率而非精确



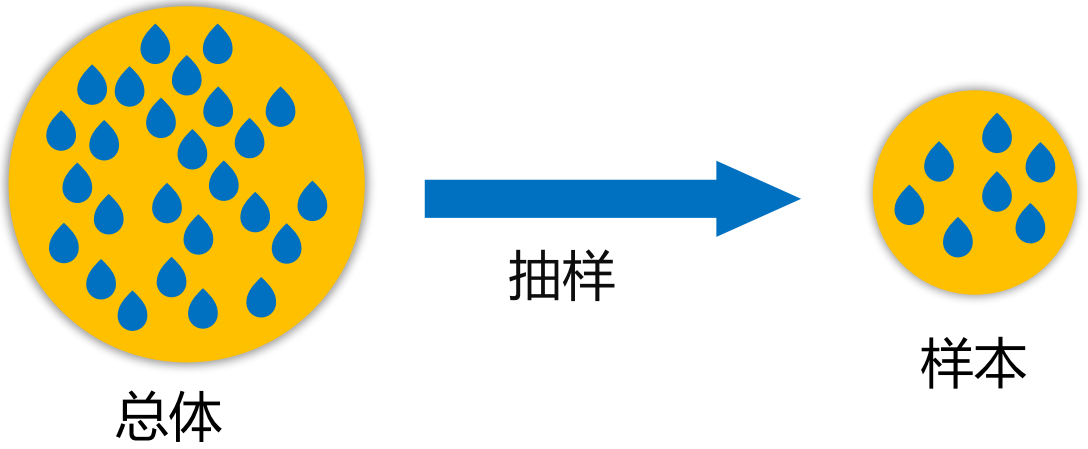
1.3 大数据的影响

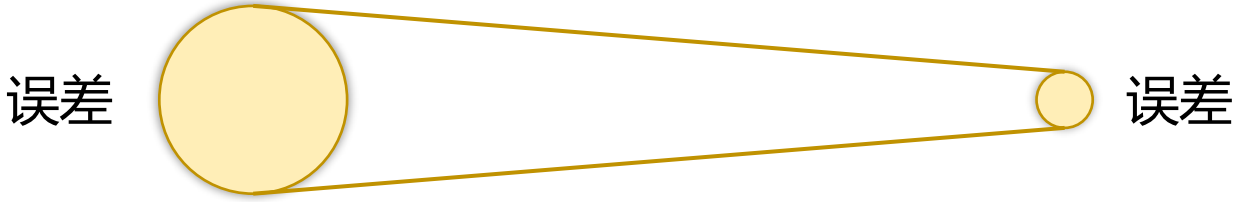
- 效率而非精确



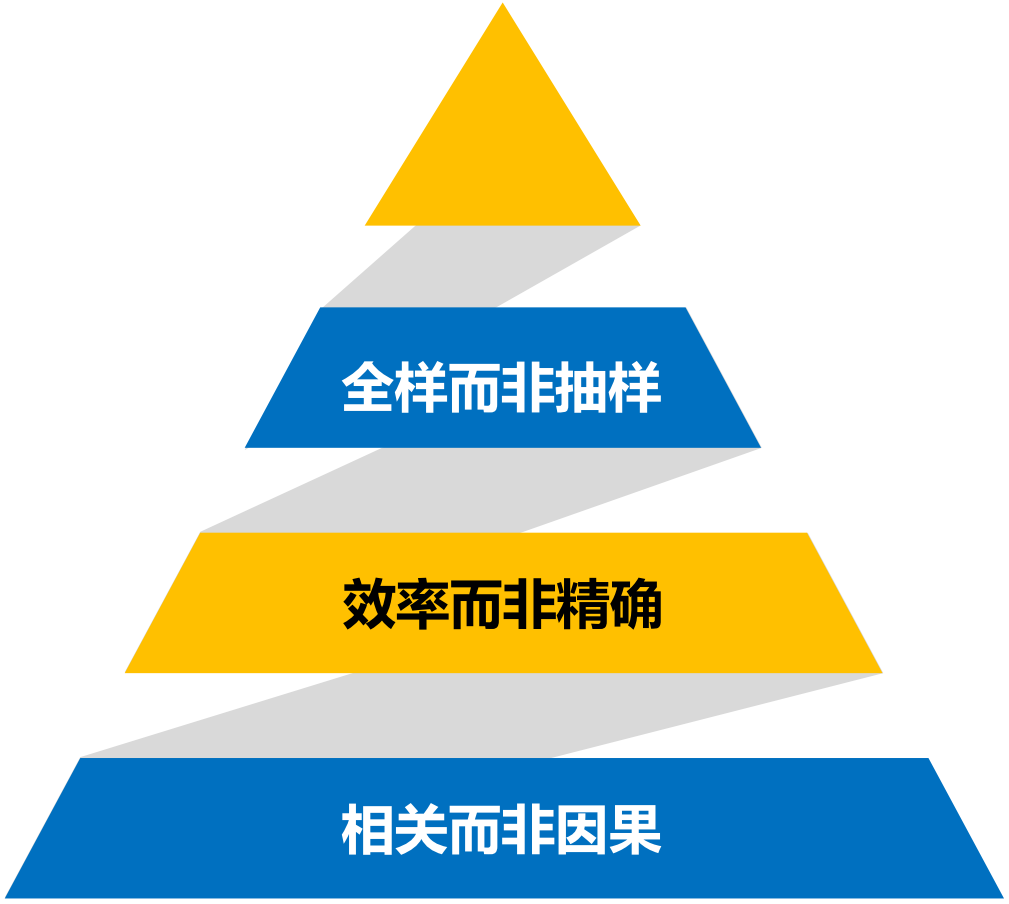
1.3 大数据的影响

- 效率而非精确





1.3 大数据的影响



1.3 大数据的影响

- 相关而非因果



相关性 ✓

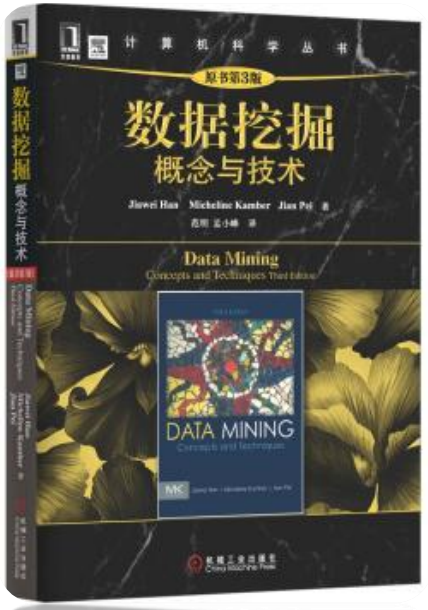


因果性 ✗

大数据时代

1.3 大数据的影响

- 相关而非因果







- 用户在网店购买一本书
- 自动推荐相关的另一本书

1.3 大数据的影响

- 相关而非因果

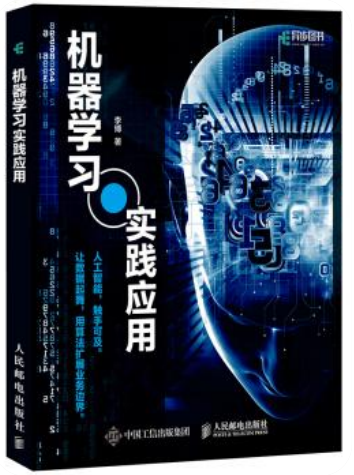


两种物品之间具有相关性

1.3 大数据的影响

- 相关而非因果





推荐相关物品给用户

追求相关性 不是因果性



1.3 大数据的影响

第1题

以下哪一项**不是**“大数据对科学研究的四种范式”之一？

A. 实验
B. 理论
C. 数据
D. 可视化

第2题

以下哪位科学家提出“以数据为中心的研究范式”?

A. 牛顿
B. 伽利略
C. Jim Gray
D. 爱因斯坦

第3题

关于“大数据驱动的科学研究”，以下说法正确的是：

A. 主要依赖手动实验结果
B. 基于假设先行的理论推导
C. 以数据发现为起点，寻找规律
D. 主要采用抽样调查方式

第4题

大数据使得我们在思维方式上更倾向于哪种处理方式?

A. 因果分析优先
B. 精度最大化
C. 抽样调查为主
D. 效率优先，相关性分析

第5题

下列哪一项是“大数据时代强调‘全样而非抽样’”的主要原因？

A. 统计抽样方法过时
B. 当前存储与计算能力提升
C. 数据不再具有代表性
D. 抽样误差太大无法接受

第6题

以下关于“效率而非精确”的说法中，**错误**的是？

A. 更重视大致有效的结论
B. 强调海量数据中快速提取信息
C. 所有数据分析必须精确到个位
D. 容许一定程度的误差存在

第7题

关于“相关而非因果”的分析逻辑，以下说法正确的是?

A. 只要找到因果关系才能进行推荐
B. 相关性强的两个变量必须有逻辑联系
C. 不追求解释性，只追求预测准确率
D. 相关性不能用于商业推荐

第8题

在推荐系统中，当用户购买一本书后被推荐另一本相关书籍，这体现了：

A. 数据冗余
B. 推荐过度
C. 相关性而非因果性
D. 数据采集错误

第9题

哪种说法最能反映大数据对传统科研方法的挑战?

A. 计算能力取代理论能力
B. 抽样技术被完全废弃
C. 数据代替实验成为唯一依据
D. 数据可用于主动发现问题

第10题

大数据思维强调“数据驱动”，这意味着：

A. 不再需要科学假设
B. 先有数据，再形成理论或决策
C. 所有结果必须由专家判断
D. 数据仅用于事后验证



Thank you