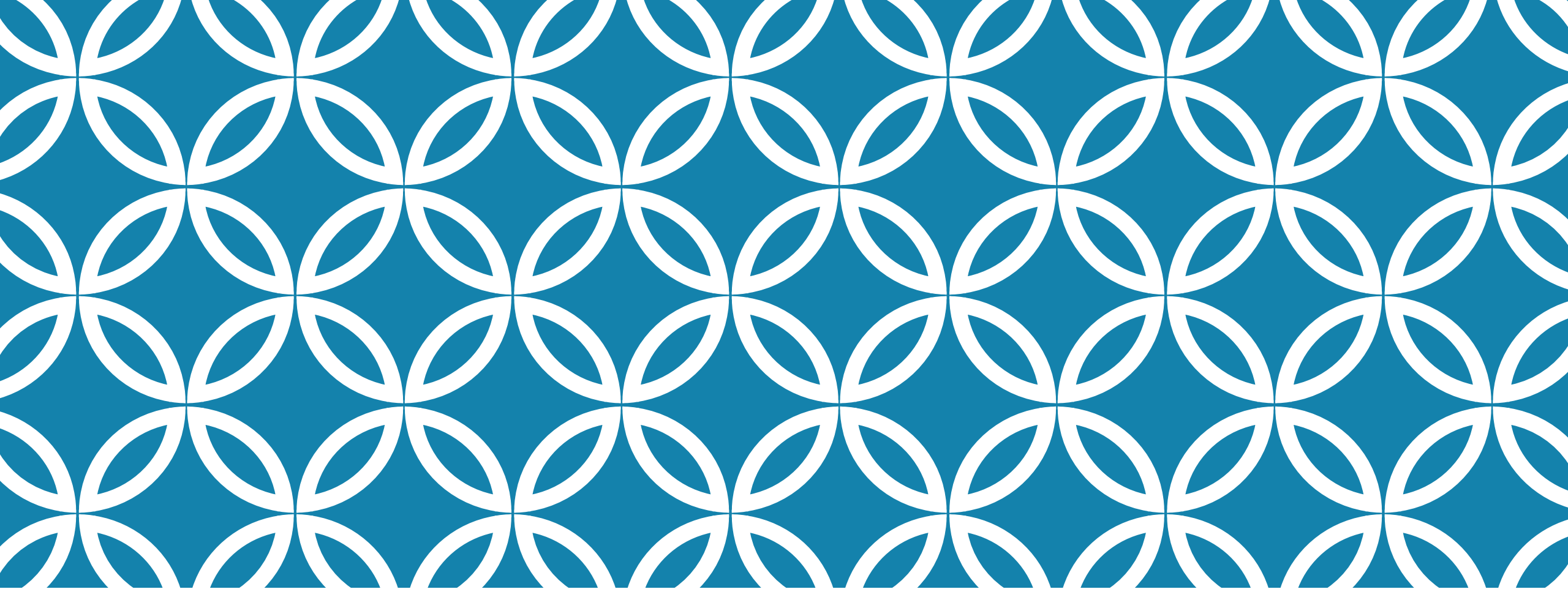




大数据应用

李立志
数据科学家
E2OPEN, INC(USA)

大数据背景

“大数据”的诞生

半个世纪以来，随着计算机和互联网技术全面融入社会生活，信息爆炸已经积累到了一个开始引发变革的程度。它不仅是世界充斥着比以为更多的信息，而且其增长速度也在加快。信息爆炸的学科如天文学和基因学，创造出了“大数据”这个概念。如今，这个概念几乎应用到了所有人类智力与发展的领域当中。



什么是大数据？

大数据或称**海量数据**，指的是所涉及的数据量规模巨大到无法通过传统数据处理方式，在合理时间内处理为人类所能解读的信息。



大数据的4V特性

Volume

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

Value

- 大量的不相关信息
- 对未来趋势与模式的可预测分析
- 深度复杂分析 (机器学习、人工智能Vs传统商务智能)

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、机器数据)
- 无模式或者模式不明显
- 不连贯的语法或句义

Velocity

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

大数据应用

找工作

领英：用大数据颠覆招聘

找对象

大数据男女配对提高成功率

电子商务

阿里巴巴大数据：购买推荐

移动

今日头条，个性化推荐

互联网金融

人人贷，大数据理财

大数据医疗

百度疾病预测



大数据应用实例（1）

阿里音乐：预测未来的流行巨星

大数据如何重塑音乐行业？最近，阿里音乐联合阿里云开展的一项实验尝试解答这个问题——通过将阿里音乐平台上的用户行为数据与整体网络播放、互动、搜索等指数结合，基于阿里云“数加”的强大数据计算能力，能够预测哪些音乐人将成为下一位流行巨星。



大数据应用实例（2）

约会速配

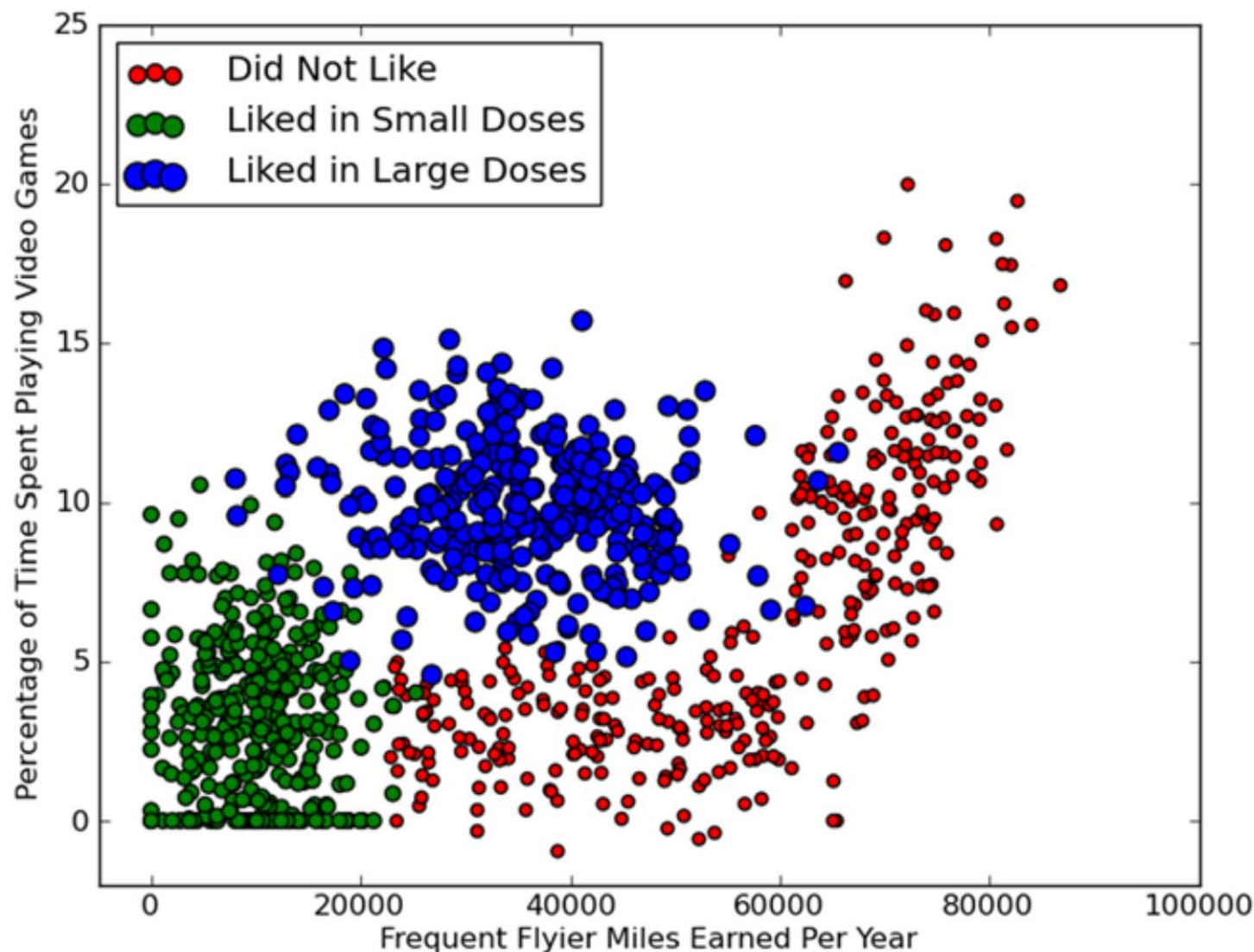


- ❑ 数学家Chris McKinlay是美国加州大学洛杉矶分校在读PhD，在多次相亲后，对于找到另一半的事情毫无起色。
- ❑ McKinlay利用自己的天赋，创建了一个自制的机器人程序从免费约会网站OkCupid上搜集女性大量信息。

大数据应用实例（2）

约会速配

□McKinlay花了三周时间从美国2万女性收集到6万问题和答案，之后利用自己研发的改进型K-Modes算法将这些女网友分成几个在系统上存在差异的集群。



大数据应用实例（2）

约会速配

□通过建立数学模型 计算的方式优化出两组女性，然后进行约会，在约会第88个女网友时，他找到了自己的真爱。

大数据应用实例（3）

网飞公司：纸牌屋

- 作为北美最大的网络影视提供商之一的NETFLIX拥有高达7500万的订阅用户。
- 用户只要登录Netflix，其每一次搜索、点击、播放、暂停甚至看了几分钟就关闭视频的记录，都会被作为数据进入后台分析。



大数据应用实例（3）

网飞公司：纸牌屋

□ 通过大数据分析，Netflix发现：

- 1990 年 BBC 迷你政治剧House of Cards广受大众欢迎。
- 喜欢观看1990版《纸牌屋》的影迷们同时喜欢看导演David Fincher的作品，
- 另外，他们会经常观看奥斯卡影帝Kevin Spacey的作品。



大数据应用实例（3）

网飞公司：纸牌屋

□ 因此新版《纸牌屋》邀请了David Fincher(制作人)和Kevin Spacey(男主演)加盟这部作品的翻拍。



大数据应用实例（3）

网飞公司：纸牌屋

□ 另一个选择在播放形式方面：

- 按照传统连载美剧的习惯，基本都是每周播放一集，
- 而Netflix根据相关数据的分析，更多人不喜欢在固定时刻收看电视剧，而是“攒起来”，直到全集播放完毕再一次性看完。



大数据应用实例（3）

网飞公司：纸牌屋

□ 因此，Netflix这次选择了一次性的播放13集《纸牌屋》。



大数据应用实例（4）

□ 欧洲某领先银行利用安装在在网点进出口、大厅和柜台等位置的高清摄像头，抓取客户的实时表情和细微动作，再加上监视器、传感器、电脑和手持设备等，所有这些产生的大量结构化数据和非结构化数据，通过网络发送并汇总到后端系统中。



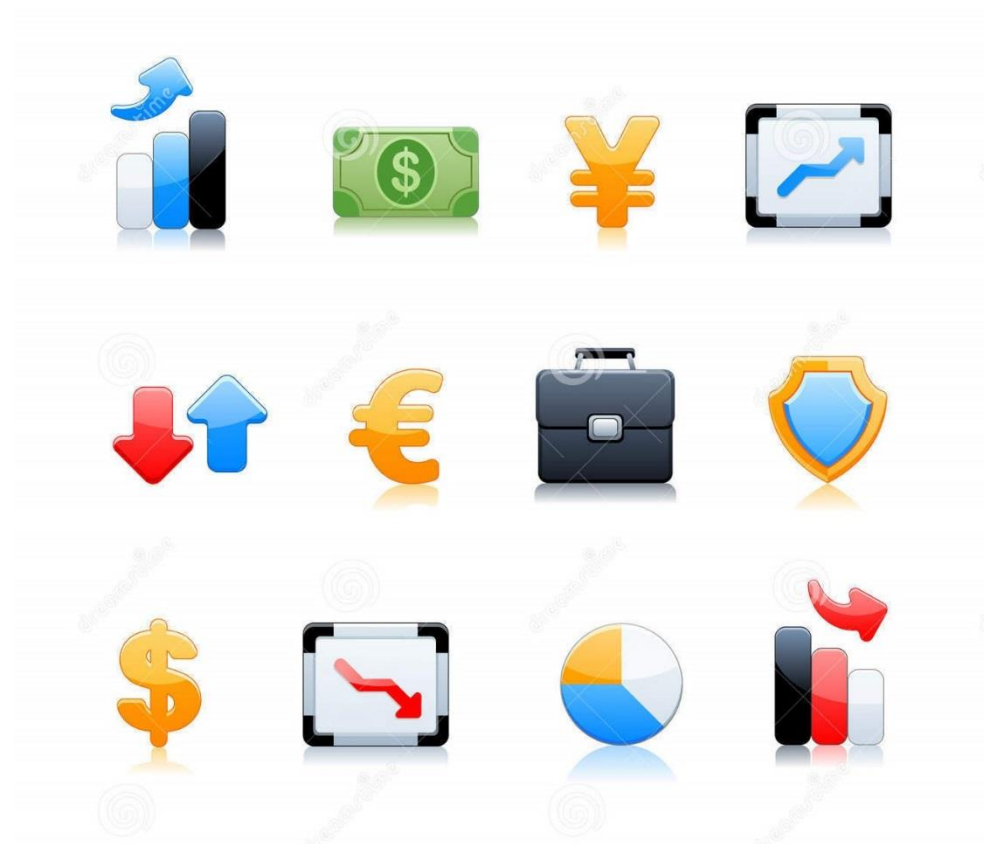
大数据应用实例（4）

□ 在此之前，银行总结出人类常见的16种典型表情，对应于人类的不同情绪，比如喜悦、紧张、沮丧、困惑、焦急、无奈等等.....。



大数据应用实例（4）

□而配合相应的历史资料分析，结合银行这一特定的场景，就可以初步得出相应表情的客户在银行可能进行的业务，比如是大额借贷、还款、金融服务等等。



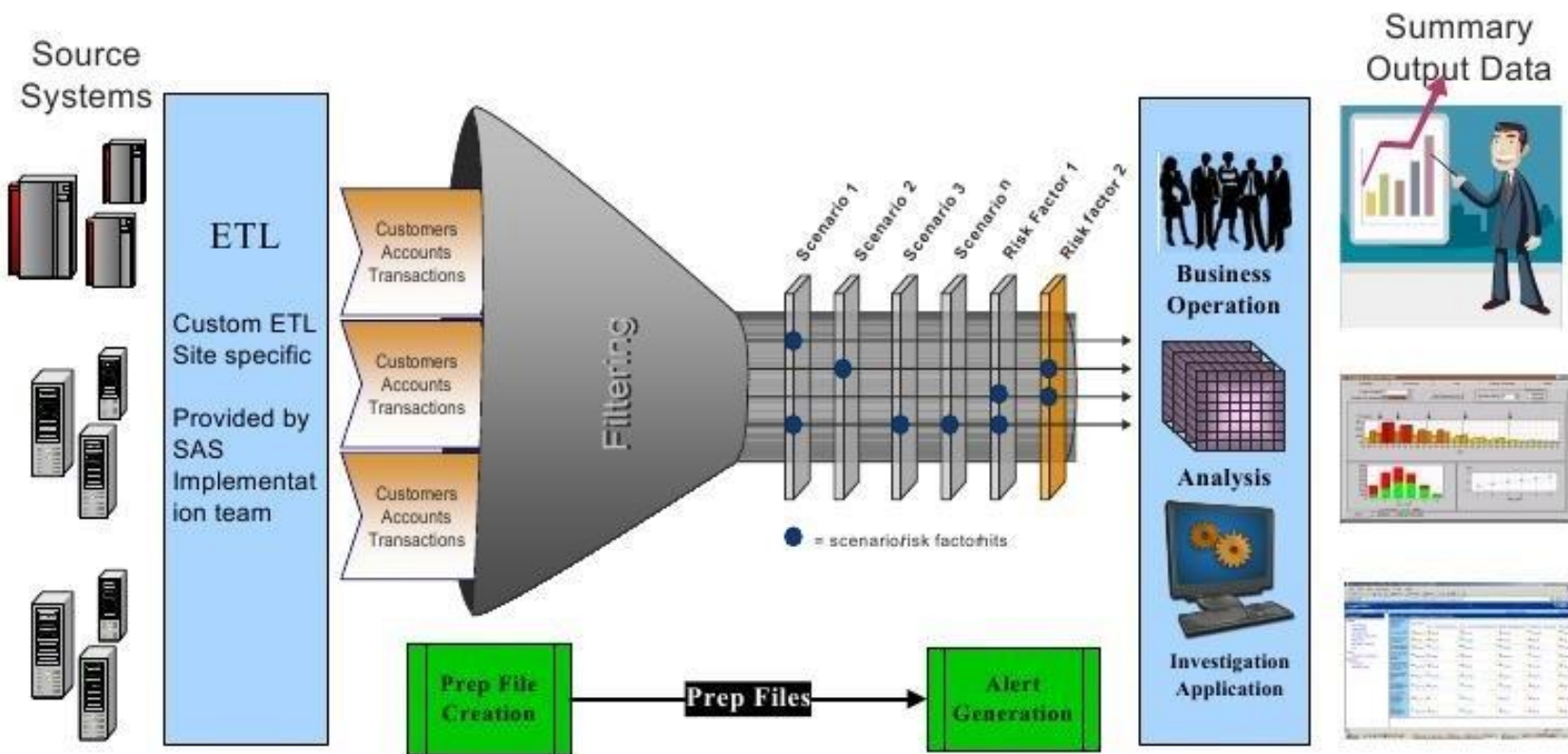
大数据应用实例（4）

- 银行甚至可以分析出相关业务的迫切性，同时配合历史数据的分析，对高危人群进行预警，控制金融风险。
- 最后，在客户走出银行时，也会给用户的情绪进行分析，来判断其对这次银行业务的满意程度，经过倒溯，结合其办理的业务内容、银行相应的接待人员来进行相应的业务优势与员工绩效的辅助参考。

大数据应用实例（4）

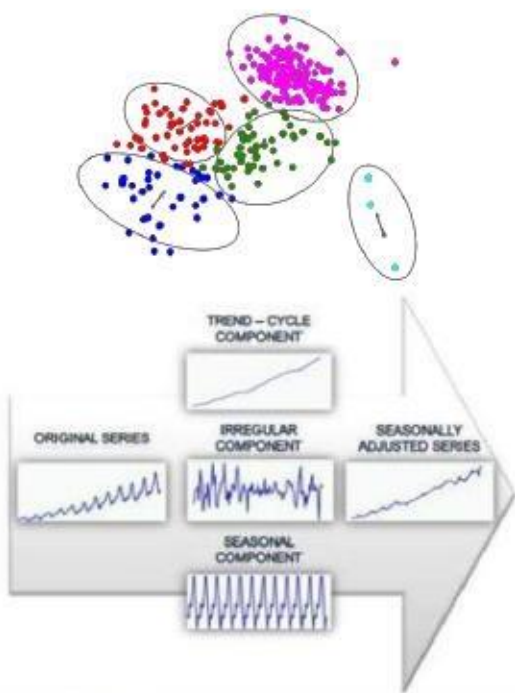
- ❑来自摄像头、传感器和物联网的数据记录了人们的行为和痕迹，从中分析出他们的特征和偏好，判断他们潜在的需求，为其量身定制开发和推销产品，这才是他们真正需要的。
- ❑所以说，从银行认为客户需要什么，转变成根据数据分析出可出真正想要什么，这是银行利用大数据实现的真正创新转型。

大数据应用实例（5） 反洗钱

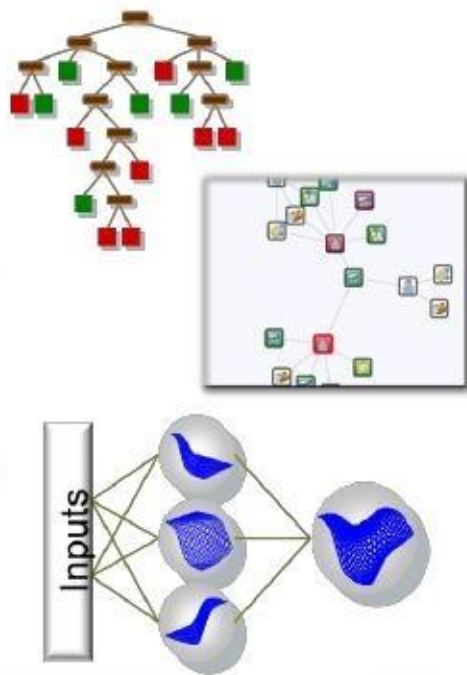


大数据分析三个阶段

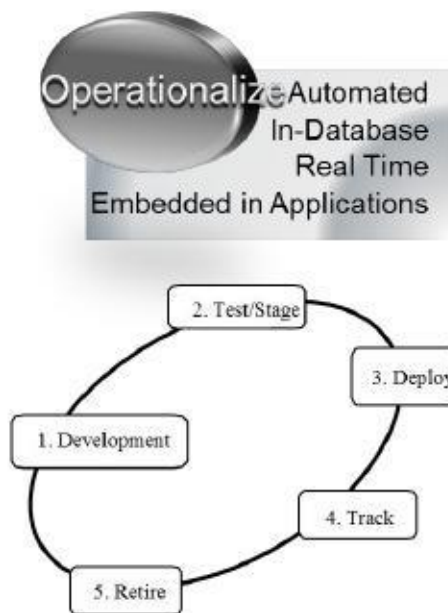
Discovery



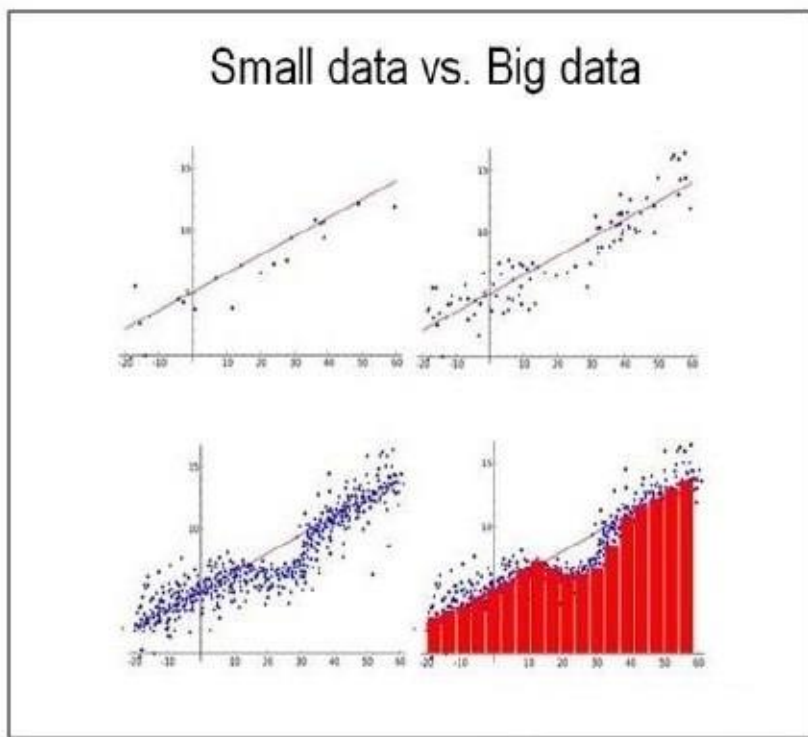
Modeling



Deployment



大数据分析 vs 小数据分析



小数据	大数据
一般性规律	特殊性规律
因果性	关联性
逻辑、推理	非推理
带参数的模型	无参数的模型
可以有效解释	准确率高

大数据常见分析软件

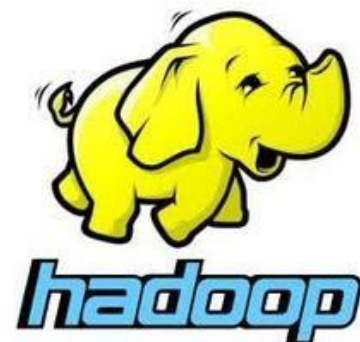
统计分析软件：SAS, R, STATA, ...

传统关联数据库软件：ORACLE, MYSQL, SQL SERVER, ...

大数据流行软件：Hadoop, Sparks, ...



ORACLE®



大数据相关技术



大数据与云计算



大数据的机遇与挑战

机遇：大数据赋予我们洞察未来的能力

- 世界杯预测，股市预测，物价预测etc

挑战：隐私与安全性问题

- CDSN密码泄漏问题， iCloud图片泄露问题etc



THANK YOU

QUESTIONS?