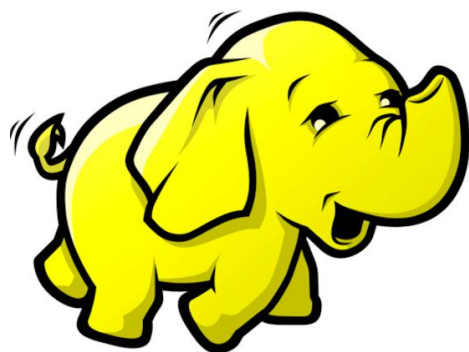


物联网云计算与大数据应用培训班

2016年8月26日 泉州市数控一代科技创新中心



大数据技术及其企业化应用

林子雨 博士/助理教授

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

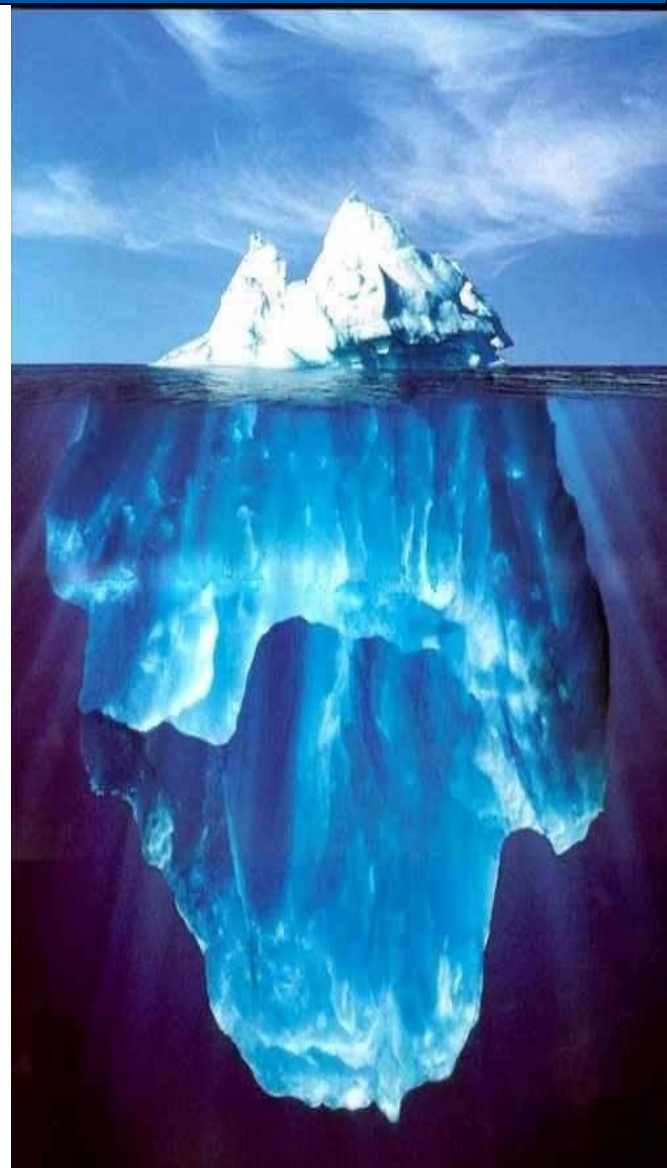
主页: <http://www.cs.xmu.edu.cn/linziyu>





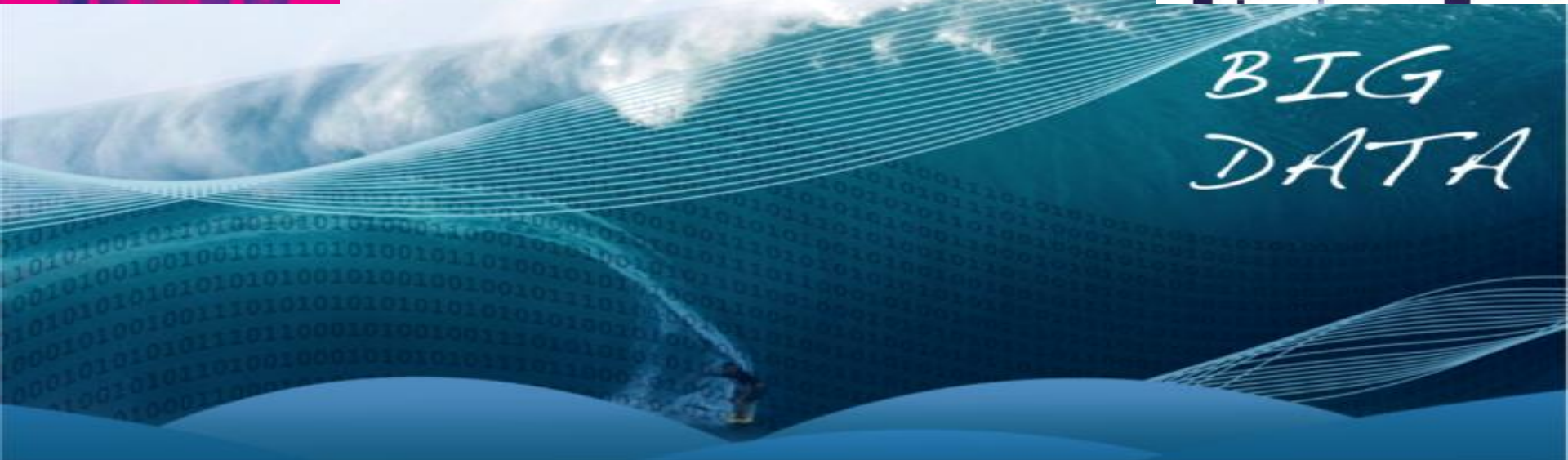
提纲

- 大数据
- 大数据应用
- 大数据的影响
- 大数据的四个特征
- 大数据产业化应用案例





1.1 大数据时代





1.1.1 第三次信息化浪潮

- 根据IBM前首席执行官郭士纳的观点，IT领域每隔十五年就会迎来一次重大变革

表1-1 三次信息化浪潮

信息化浪潮	发生时间	标志	解决问题	代表企业
第一次浪潮	1980年前后	个人计算机	信息处理	Intel、AMD、IBM、苹果、微软、联想、戴尔、惠普等
第二次浪潮	1995年前后	互联网	信息传输	雅虎、谷歌、阿里巴巴、百度、腾讯等
第三次浪潮	2010年前后	物联网、云计算和大数据	信息爆炸	将涌现出一批新的市场标杆企业



1.1.2 信息科技为大数据时代提供技术支撑

1. 存储设备容量不断增加

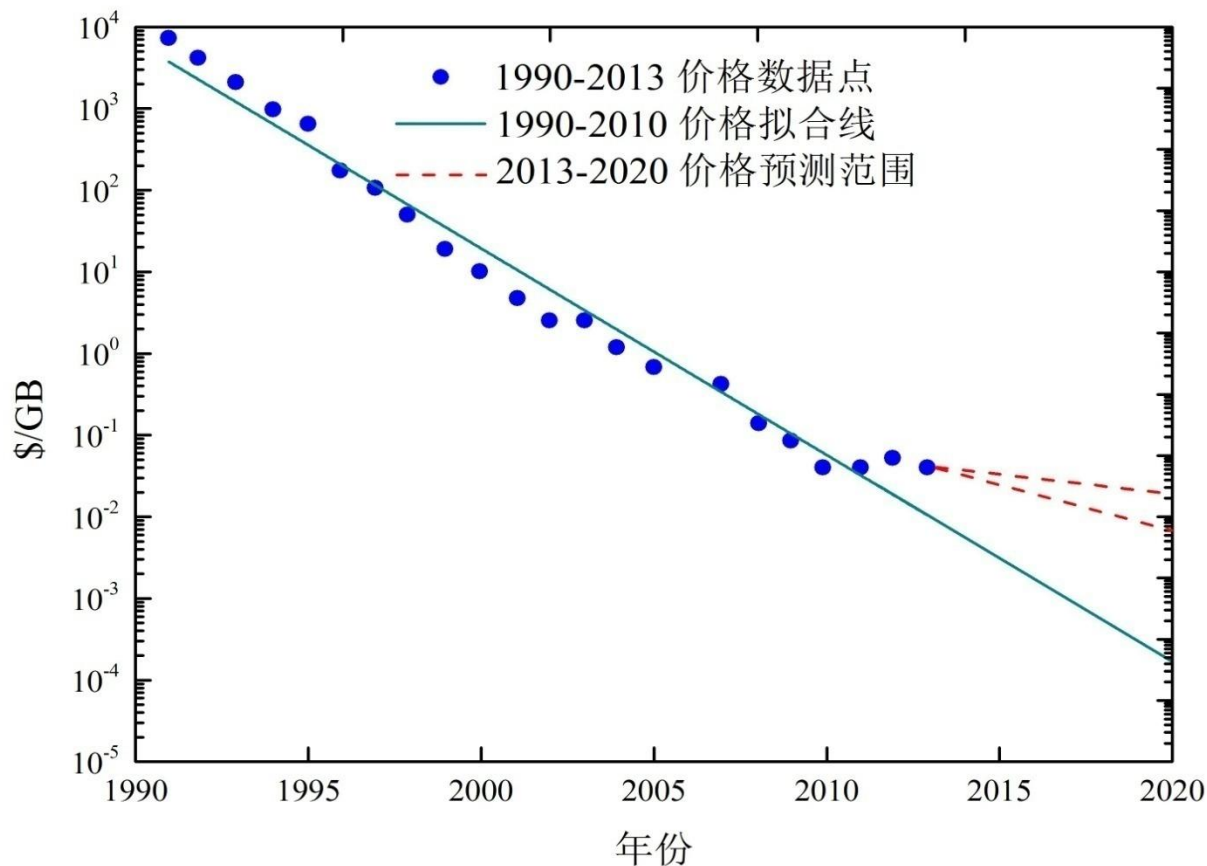


图1-1 存储价格随时间变化情况



1.2 信息科技为大数据时代提供技术支撑



来自斯威本科技大学（Swinburne University of Technology）的研究团队，在2013年6月29日刊出的《自然通讯（Nature Communications）》杂志的文章中，描述了一种全新的数据存储方式，可将1PB（1024TB）的数据存储到一张仅DVD大小的聚合物碟片上。



1.1.2 信息科技为大数据时代提供技术支撑

2. CPU处理能力大幅提升

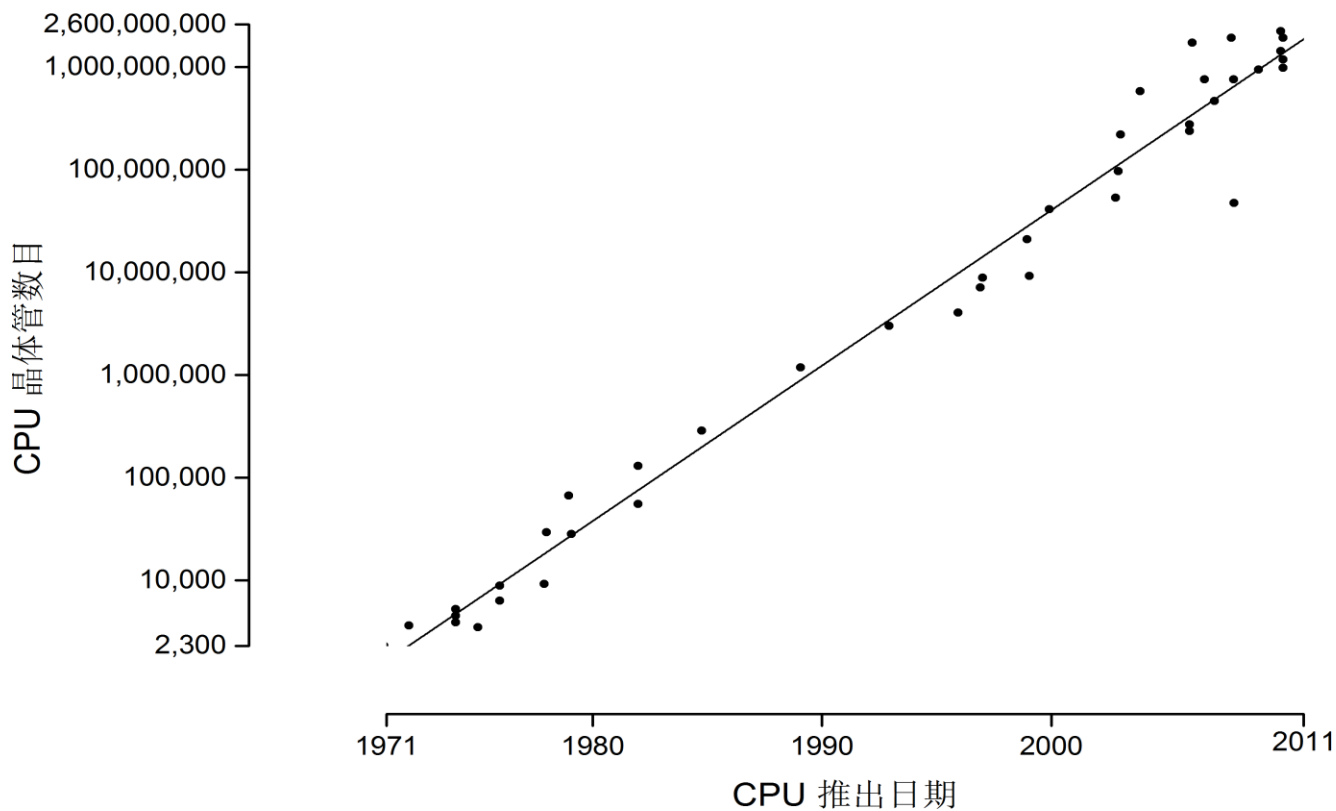


图1-3 CPU晶体管数目随时间变化情况



1.1.2 信息科技为大数据时代提供技术支撑

3. 网络带宽不断增加

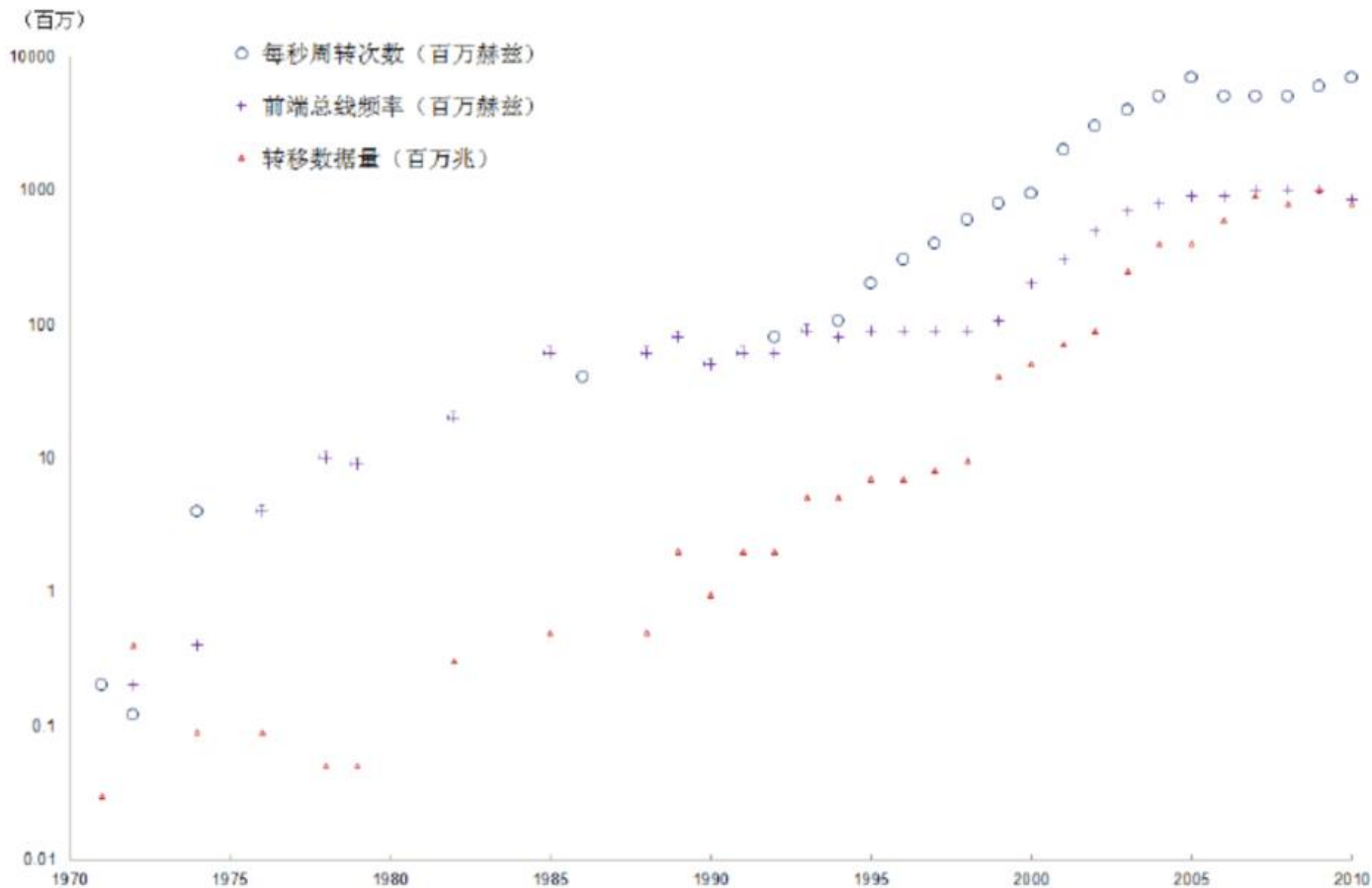


图1-4 网络带宽随时间变化情况



1.1.3数据产生方式的变革促成大数据时代的来临

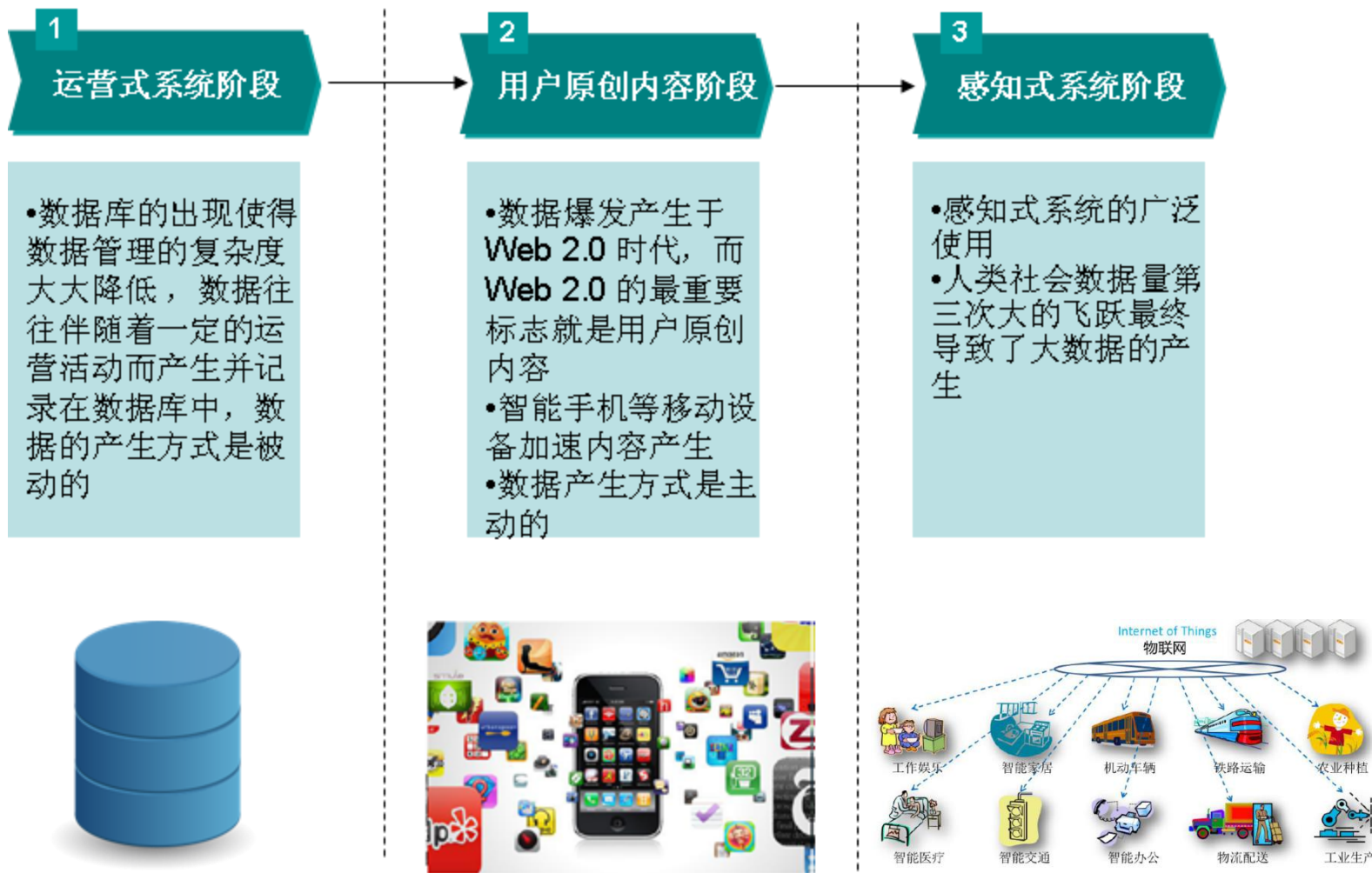


图1-5 数据产生方式的变革



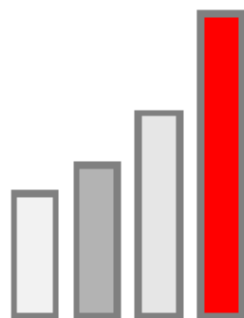
1.1.4 大数据的发展历程

表1-2 大数据发展的三个阶段

阶段	时间	内容
第一阶段：萌芽期	上世纪90年代至本世纪初	随着数据挖掘理论和数据库技术的逐步成熟，一批商业智能工具和知识管理技术开始被应用，如数据仓库、专家系统、知识管理系统等。
第二阶段：成熟期	本世纪前十年	Web2.0应用迅猛发展，非结构化数据大量产生，传统处理方法难以应对，带动了大数据技术的快速突破，大数据解决方案逐渐走向成熟，形成了并行计算与分布式系统两大核心技术，谷歌的GFS和MapReduce等大数据技术受到追捧，Hadoop平台开始大行其道
第三阶段：大规模应用期	2010年以后	大数据应用渗透各行各业，数据驱动决策，信息社会智能化程度大幅提高



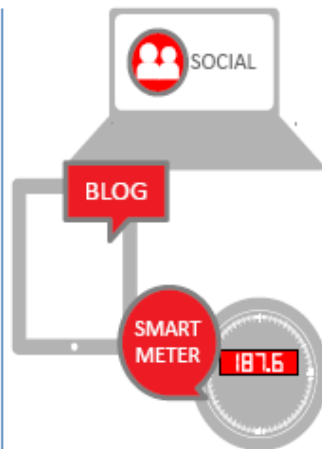
1.2 大数据概念



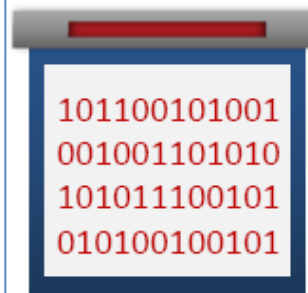
VOLUME
大量化



VELOCITY
快速化



VARIETY
多样化



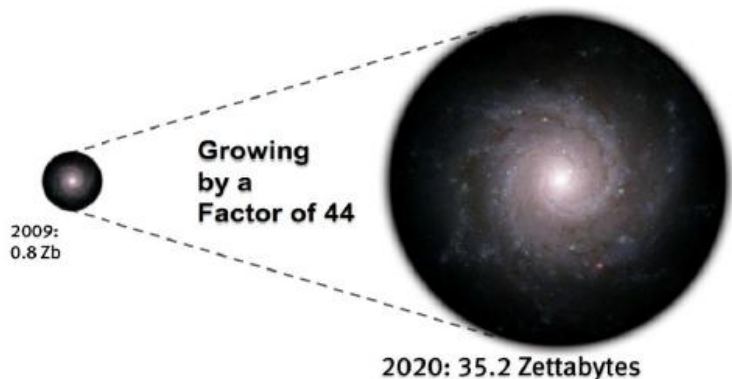
VALUE

大数据不仅仅是数据的“大量化”，而是包含“快速化”、“多样化”和“价值化”等多重属性。



1.2.1 数据量大

- 根据IDC作出的估测，数据一直都在以每年50%的速度增长，也就是说每两年就增长一倍（大数据摩尔定律）
- 人类在最近两年产生的数据量相当于之前产生的全部数据量
- 预计到2020年，全球将总共拥有35ZB的数据量，相较于2010年，数据量将增长近30倍

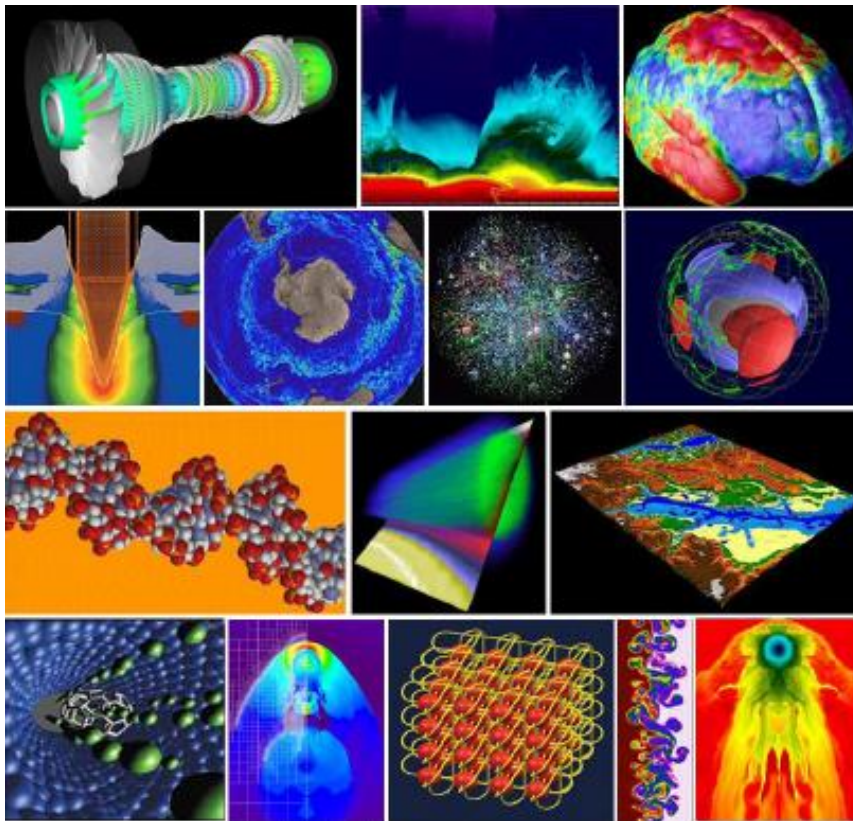


TERABYTE	10 的 12 次方	一块 1TB 硬盘		200,000 照片或 mp3 歌曲
PETABYTE	10 的 15 次方	两个数据中心机柜		16 个 Blackblaze pod 存储单元
EXABYTE	10 的 18 次方	2,000 个机柜		占据一个街区的 4 层数据中心
ZETTABYTE	10 的 21 次方	1000 个数据中心		纽约曼哈顿的 1/5 区域
YOTTABYTE	10 的 24 次方	一百万个数据中心		特拉华州和罗德岛州



1.2.2 数据类型繁多

- 大数据是由结构化和非结构化数据组成的
 - 10%的结构化数据，存储在数据库中
 - 90%的非结构化数据，它们与人类信息密切相关



- 科学研究
 - 基因组
 - LHC 加速器
 - 地球与空间探测
- 企业应用
 - Email、文档、文件
 - 应用日志
 - 交易记录
- Web 1.0数据
 - 文本
 - 图像
 - 视频
- Web 2.0数据
 - 查询日志/点击流
 - Twitter/ Blog / SNS
 - Wiki



1.2.3处理速度快

- 从数据的生成到消耗，时间窗口非常小，可用于生成决策的时间非常少
- 1秒定律：这一点也是和传统的数据挖掘技术有着本质的不同

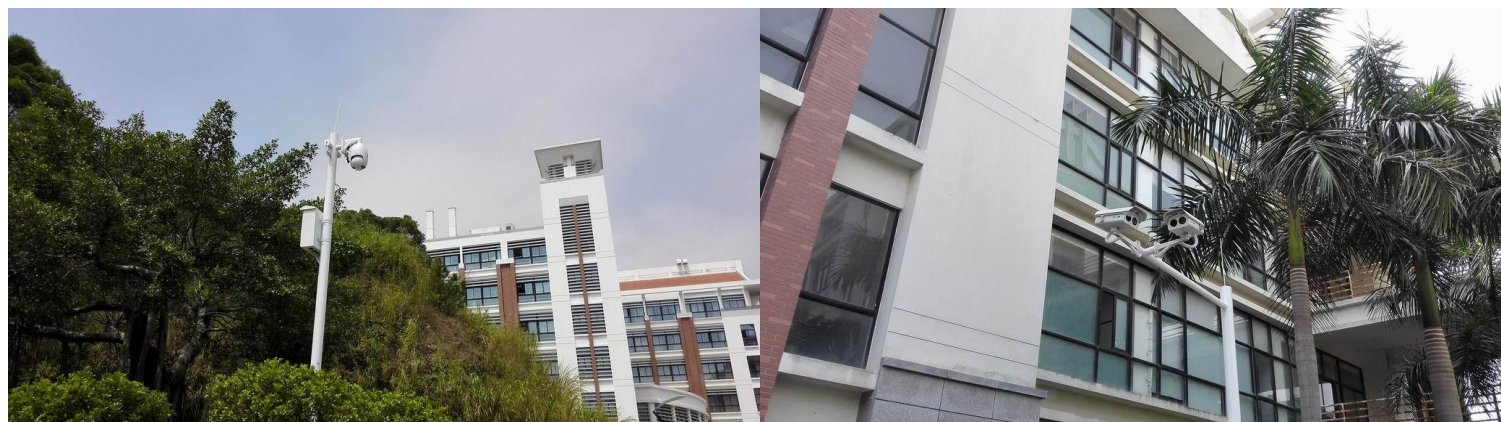




1.2.4 价值密度低

价值密度低，商业价值高

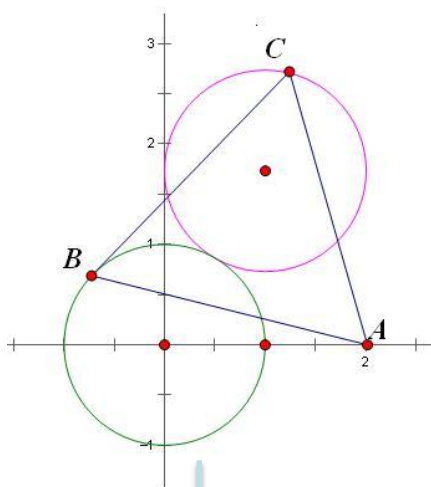
以视频为例，连续不间断监控过程中，可能有用的数据仅仅有一两秒，但是具有很高的商业价值





1.3大数据的影响

图灵奖获得者、著名数据库专家Jim Gray 博士观察并总结人类自古以来，在科学研究上，先后历经了实验、理论、计算和数据四种范式



实验

理论

计算

数据



1.3大数据的影响

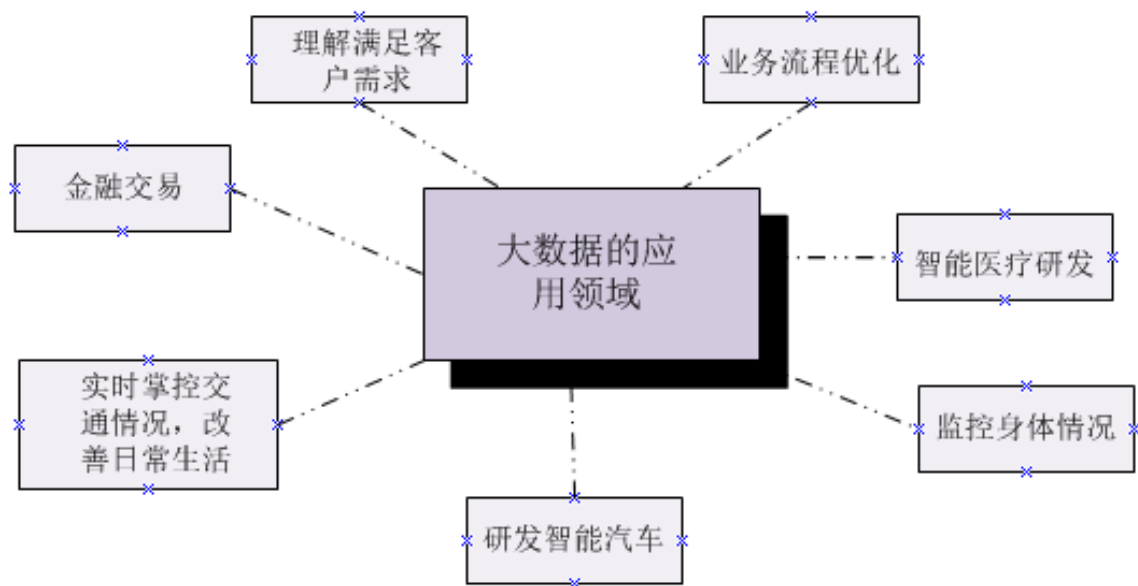
- 在思维方式方面，大数据完全颠覆了传统的思维方式：
 - 全样而非抽样
 - 效率而非精确
 - 相关而非因果





1.4大数据的应用

- 大数据无处不在，包括金融、汽车、零售、餐饮、电信、能源、政务、医疗、体育、娱乐等在内的社会各行各业都已经融入了大数据的印迹





1.4大数据的应用



商品销售



股票投资



UPS快递的最佳行车路径



VISA信用卡与商户推荐



汽车保险



阿里信用贷款和淘宝数据魔方



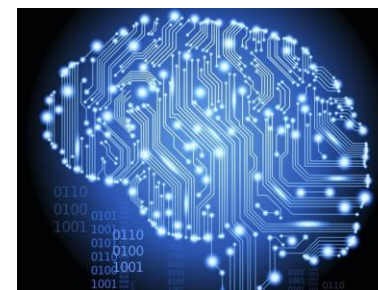
智能电表



广告投放



机器翻译



人的数字化



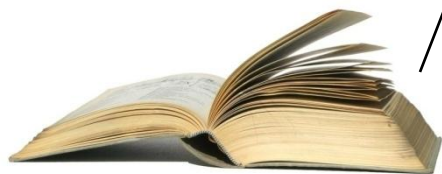
1.4大数据的应用



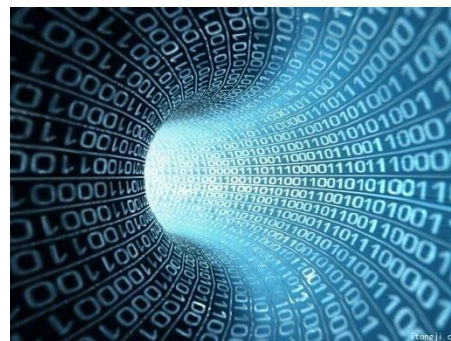
Kevin Spacey



David Fincher



英国同名小说《纸牌屋》



大数据分析



风靡全球的美剧《纸牌屋》



1.4大数据的应用



从谷歌流感趋势看大数据的应用价值

“谷歌流感趋势”，通过跟踪搜索词相关数据来判断全美地区的流感情况

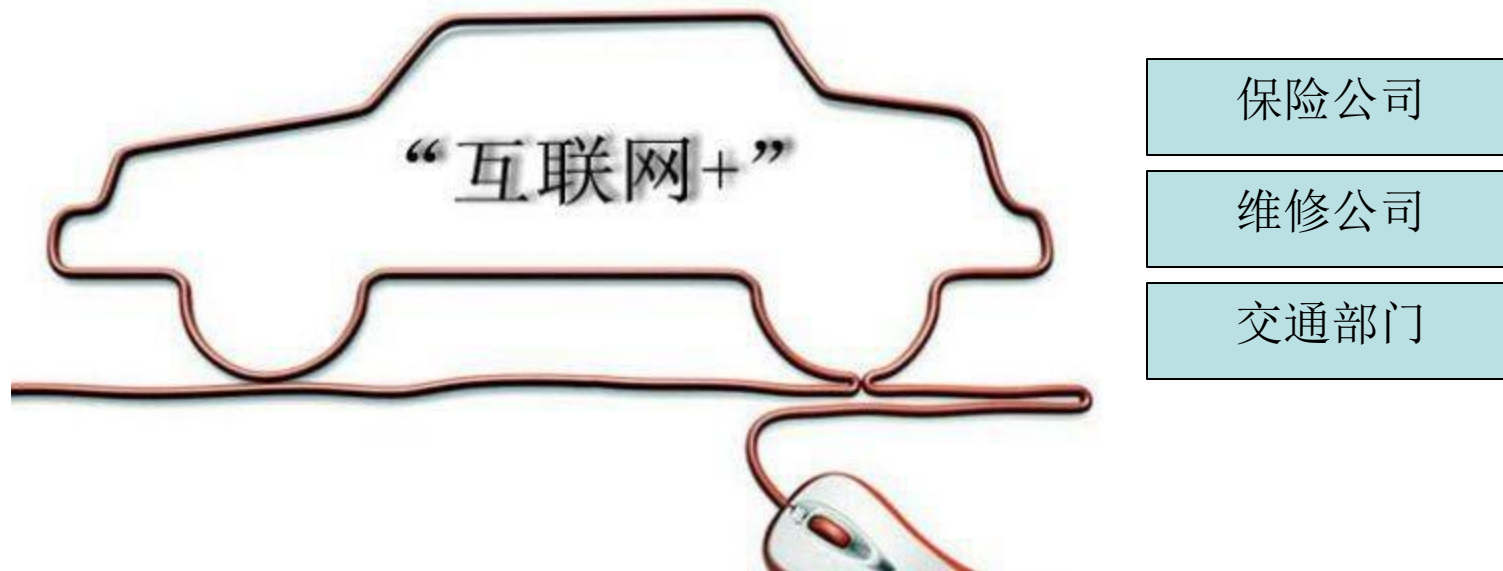
图:美国某地区历年来的流感发病率



数据来源: 谷歌趋势, 美国各地疾病预防控制中心



1.4大数据的应用

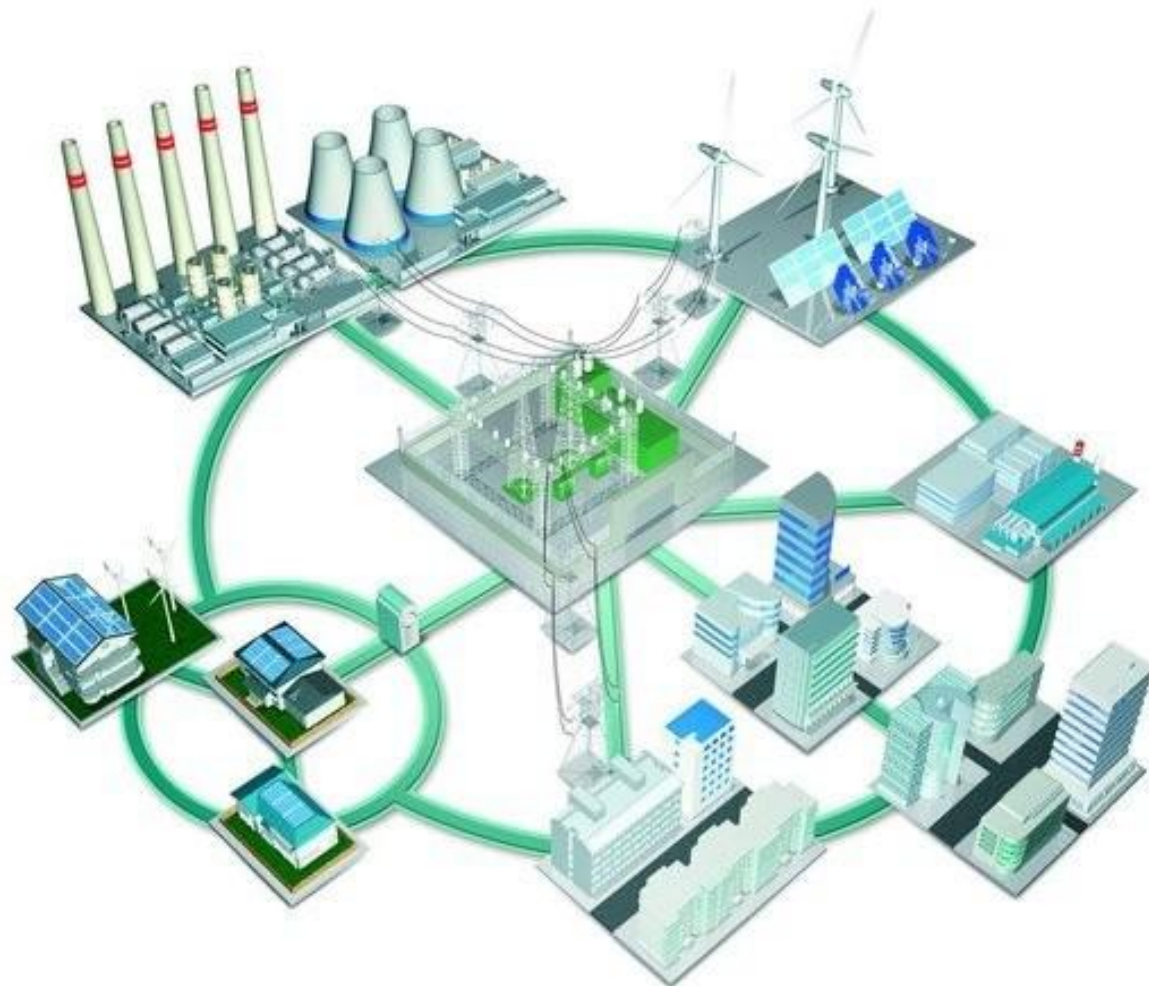


“互联网+汽车”就是基于汽车生产企业的工业大数据，综合人、车、路、环境、社会之间关系



1.4大数据的应用

智能电网





1.4大数据的应用



无人机实现大数据精准农业



1.9 产业化应用案例分享

案例一：在物流行业的应用



物流行业应用案例：智能物流

智能物流集成商案例：阿里巴巴的中国智能物流骨干网（地网）



中国智能物流骨干网

“菜鸟”将物流资源重组，欲将运力变得更集中、高效

菜鸟网络到底是什么？

- 中国智能物流骨干网，又名“菜鸟”
- 菜鸟网络计划在5到8年内，打造一个全国性的超级物流网。
- 这个网络能在24小时内将货物运抵国内任何地区，能支撑日均300亿元(年度约10万亿元)的巨量网络零售额。

1000亿元投资物流基础设施 强强联手共建智能骨干网络
物流信息系统向所有的制造商、网商、快递公司、第三方物流公司完全开放



阿里物流体系

天网

天猫牵头负责与各大物流快递公司对接的数据平台

地网

即“菜鸟”，又称“中国智能物流骨干网 (CSN)”



中国智能物流骨干网——菜鸟网络

依托阿里巴巴集团旗下多个电商平台为核心的大数据平台（**天网**），即掌握的网络购物物流需求数据、电商货源数据、货流量及分布数据、以及消费者长期购买习惯数据，优化仓储选址、干线物流基础设施建设、以及物流体系建设

关键举措一：智能化建立物流集散中心（基础设施平台），搭建骨干网框架

关键举措三：应用智能化技术，补足物流行业仓储环节短板

采用自动分拣、自动传输、自动出库、自动补货等手段建立智能实体仓库，在减少库存积压的基础上提升效率，同时建立虚拟仓库，实现信息与数据对接的信息化管理

建立统一的仓储及调度体系，整合和集中管理原本各快递公司自建的物流体系

关键举措二：整合所有服务商信息系统，实现骨干网内部信息统一

关键举措四：构建开放数据应用平台，向物流生态系统内各种群提供服务

构建向“电子商务企业、物流公司、仓储企业、第三方物流服务商以及供应链服务商”开放的数据应用平台



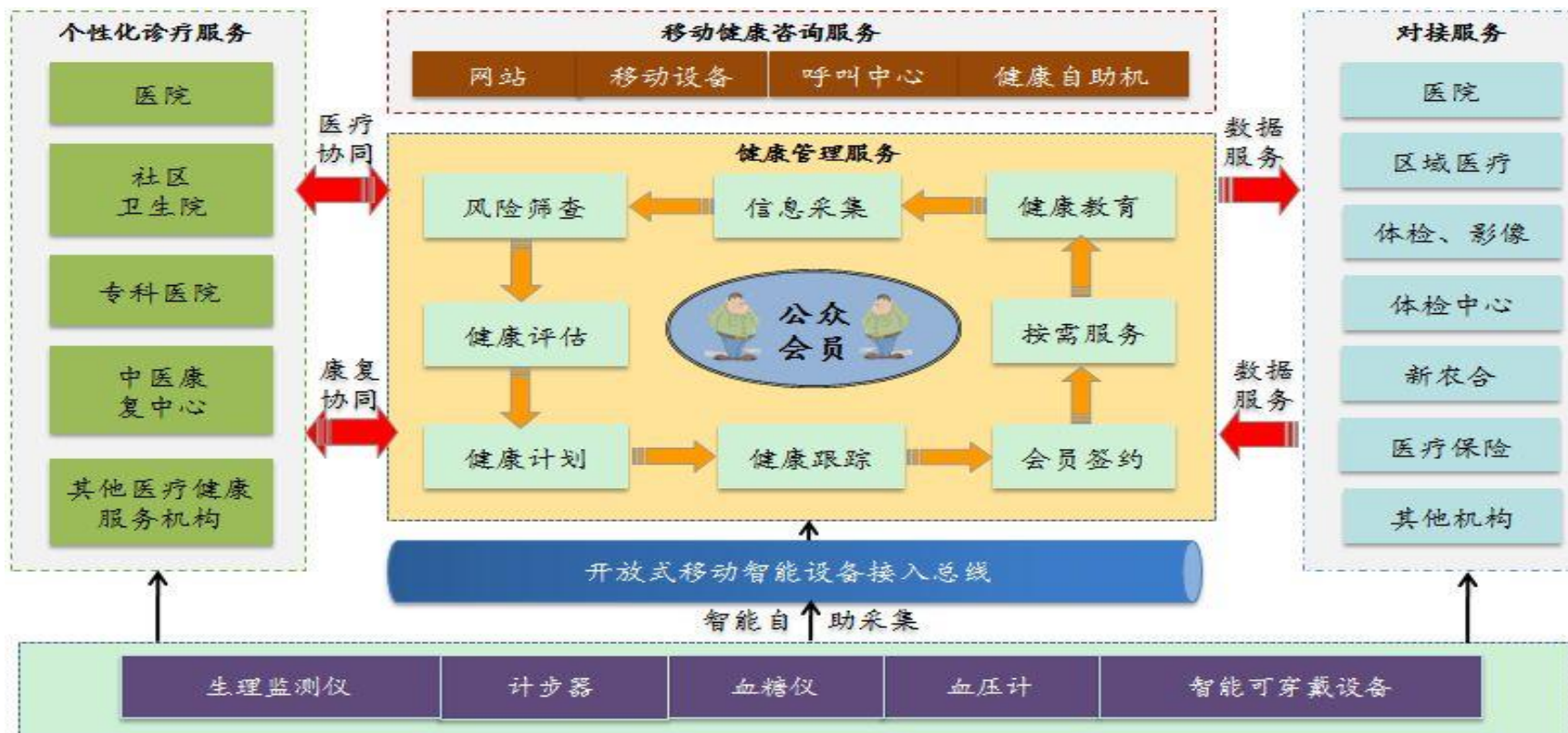
1.9 产业化应用案例分享

案例二：在医疗健康行业的应用



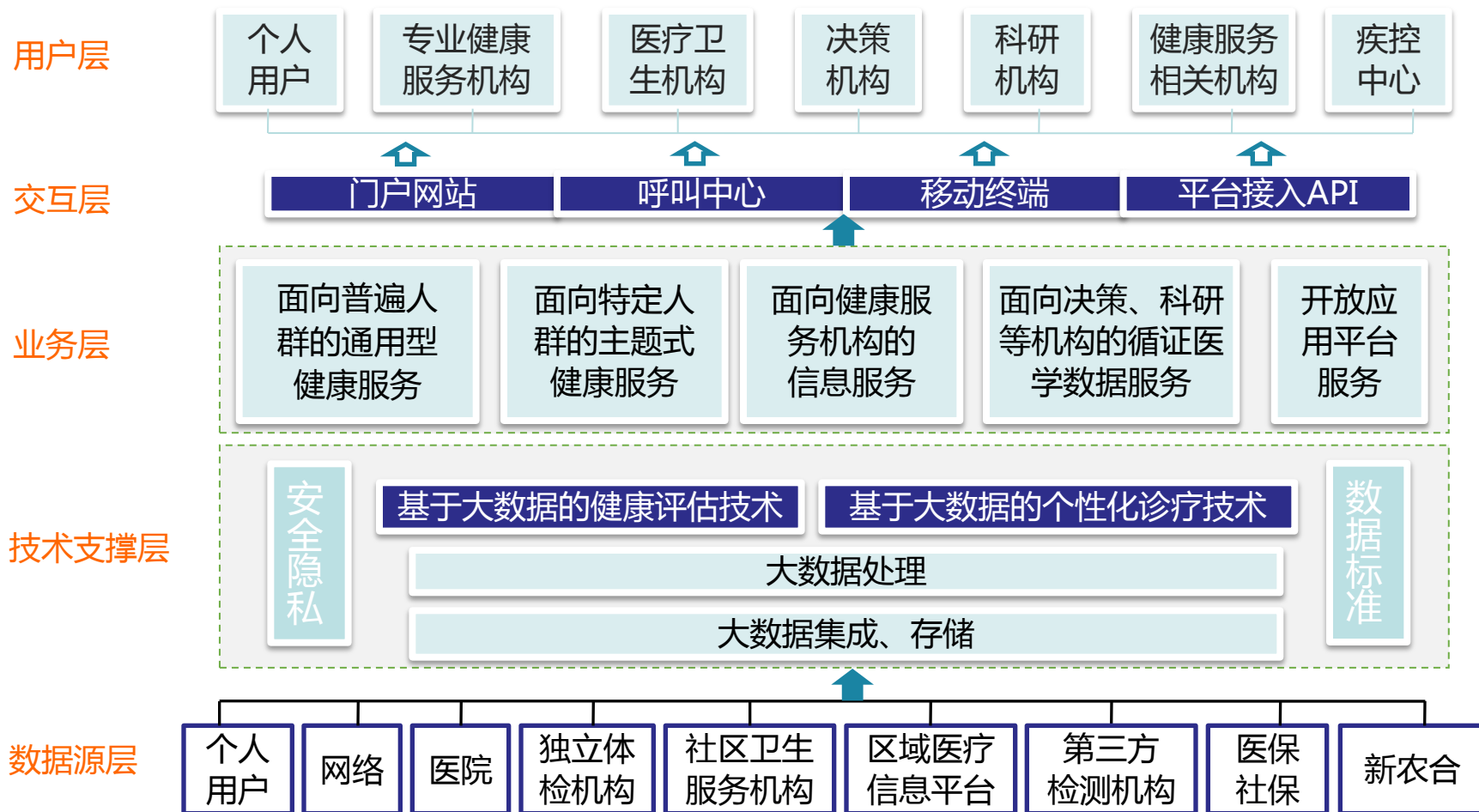
医疗健康行业应用：综合健康服务平台

建设目标：构建覆盖全生命周期、内涵丰富、结构合理的以人为本全面连续的综合健康服务体系，利用大数据技术和智能设备技术，提供线上线下相结合的公众健康服务，实现“未病先防、已病早治、既病防变、愈后防复”，满足社会公众多层次、多方位的健康服务需求，提升人民群众的身心健康水平。





医疗健康行业应用：综合健康服务平台





总结

- 大数据无处不在，是变革经济的力量
- 大数据是多种技术的集合，不只是数据
- 大数据与云计算相辅相成
- Google公司在大数据时代引领技术
- Hadoop成为事实上的大数据处理标准
- 大数据可以改变企业传统管理和经营方式



附件：林子雨简介



林子雨

单位：厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

个人网页: <http://www.cs.xmu.edu.cn/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>




扫一扫访问林子雨个人主页

简介：林子雨，男，1978年出生，北京大学博士，现为厦门大学计算机科学系教师，获得2013年度厦门大学教学类奖教金。研究领域包括数据库、数据仓库、大数据、云计算和物联网。主持或参与了包括国家自然科学基金、863计划在内的多项国家重点科研项目，并以第一作者身份在《计算机学报》《软件学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文，并编著出版中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，建设了国内高校首个大数据课程公共服务平台。

社会服务：面向企业和政府部门，提供科研支持、软件开发、企业信息化培训、科技项目申请书与产业规划文档撰写等服务。

项目经历：作为项目负责人主持完成的项目主要包括《城市信息化顶层设计框架研究报告》、《2015年泉州市互联网经济调研报告》、《厦门市云计算产业发展技术路线图》、《石狮市物流园区建设方案研究报告》、《晋江市海洋生物科技园建设方案建议书》、《国家物联网重大应用示范工程重点项目福建省物联网科学院平台实施方案》、《国家物联网重大应用示范工程区域试点泉州市总体工作方案》、《基于地面远程控制的新一代智能塔吊系统实施方案》、《基于大数据的综合健康服务平台2015科技部支撑计划项目申报书》等。



服务政府
服务企业



科学分析
科学决策



厦门大学数据库实验室

厦门大学云计算与大数据研究中心 | 厦门大学数据库实验室

地址：福建省厦门大学厦门大学海韵园科研2号楼

电话：(0595)2580033

传真：(0595)2580033

邮编：361005 E-mail: ziyulin@xmu.edu.cn

网址：<http://dblab.xmu.edu.cn>