《大数据导论》

教材官网： http://dblab.xmu.edu.cn/post/bigdata-introduction/

温馨提示：编辑幻灯片母版，可以修改每页PPT的厦大校徽和底部文字

第3章 大数据基础知识

（PPT版本号：2020年秋季学期）

扫一扫访问教材官网

林子雨 博士/副教授

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

主页： **http://dblab.xmu.edu.cn/post/linziyu**

课程教材

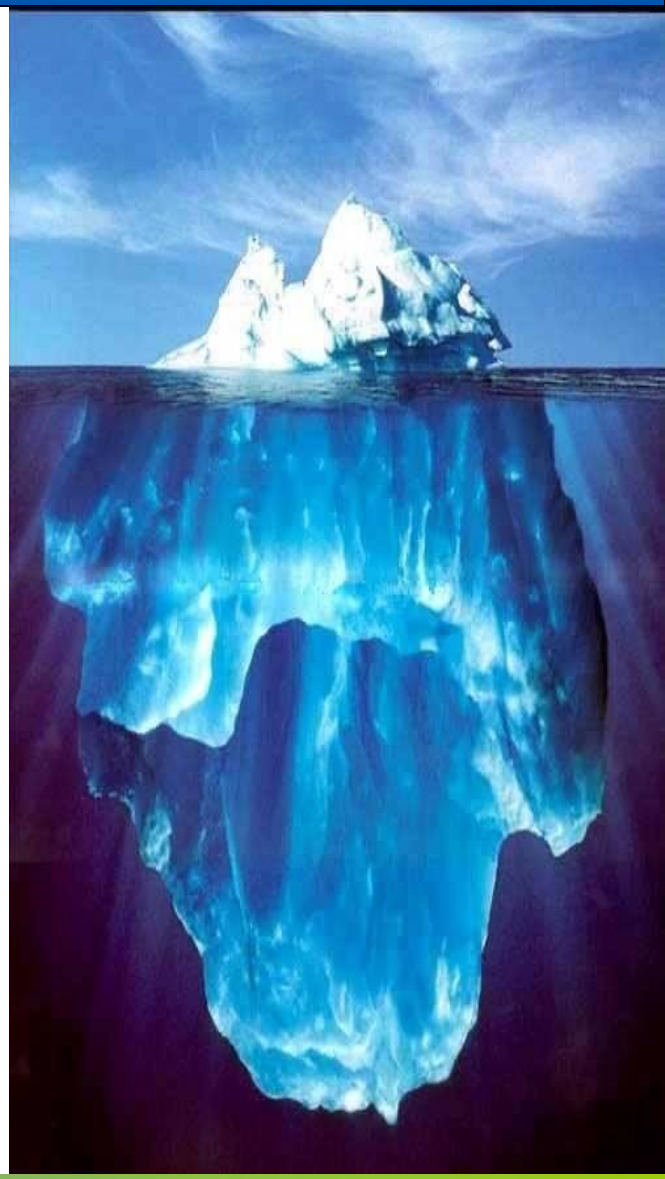
- 林子雨 编著《大数据导论》
- 人民邮电出版社，2020年8月第1版
- ISBN:978-7-115-54446-9 定价：49.80元

教材官网：http://dblab.xmu.edu.cn/post/bigdata-introduction/

扫一扫访问教材官网

提纲

3.1大数据安全
3.2大数据思维
3.3大数据伦理
3.4数据共享
3.5数据开放
3.6大数据交易



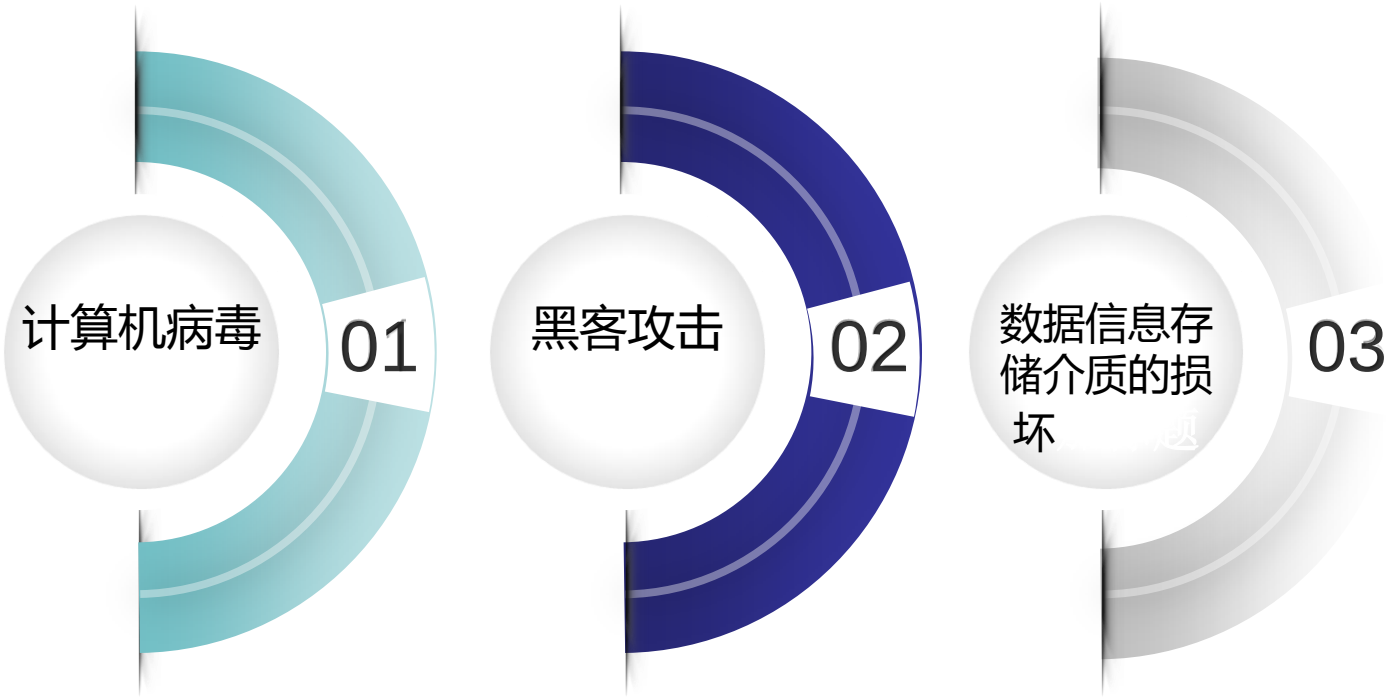
高校大数据课程
公 共 服 务 平 台

百度搜索厦门大学数据库实验室网站访问平台

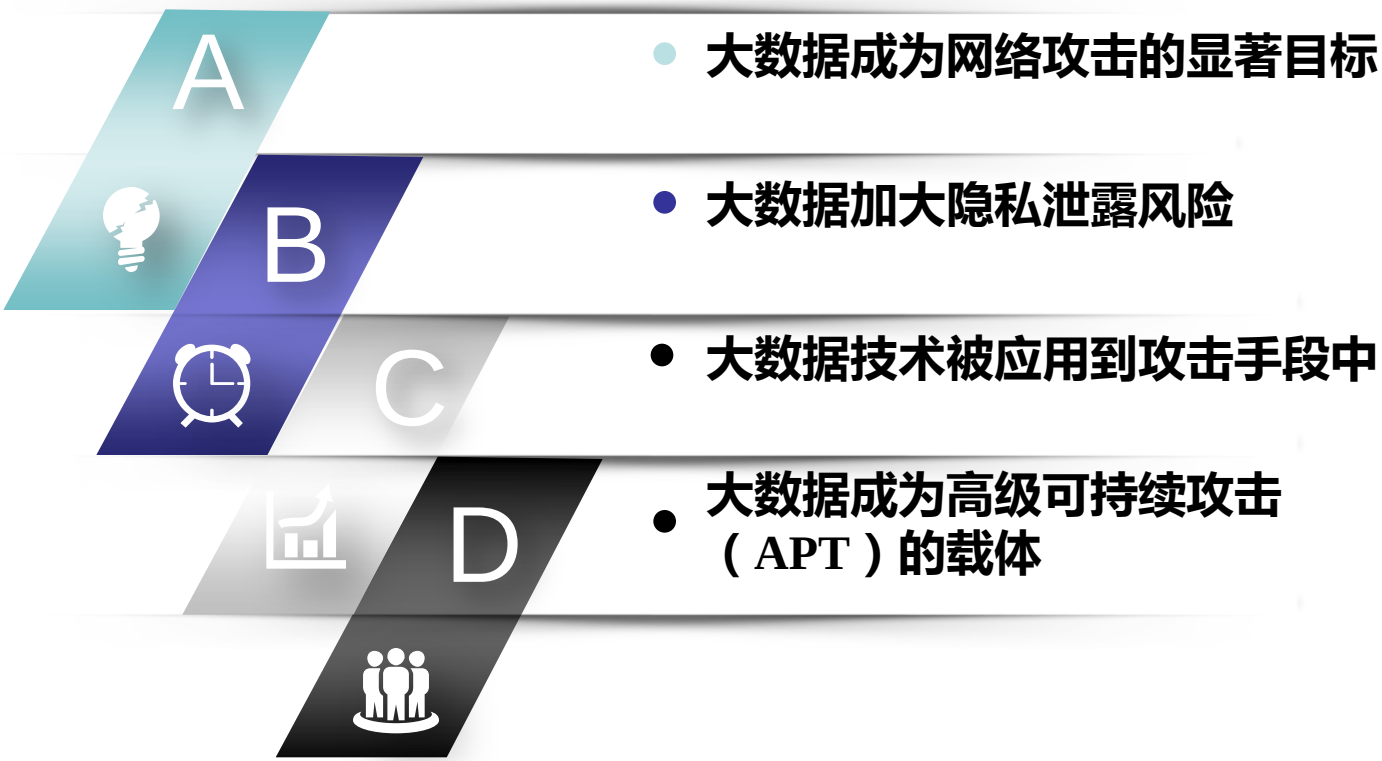
3.1大数据安全

3.1.1传统数据安全
3.1.2大数据安全与传统数据安全的不同
3.1.3大数据安全问题
3.1.4 典型案例

3.1.1传统数据安全



3.1.2大数据安全与传统数据安全的不同



3.1.3大数据安全问题

1. 隐私和个人信息安全问题

•在大数据时代，个人身份、健康状况、个人信用和财产状况以及自己和恋人的亲密过程是隐私；使用设备、位置信息、电子邮件也是隐私；同时上网浏览情况、应用的APP、在网上参加的活动、发表及阅读什么帖子、点赞，也可能成为隐私

•在大数据时代，无论是个人日常购物消费等琐碎小事，还是读书、买房、生儿育女等人生大事，都会在各式各样的数据系统中留下“数据脚印”。就单个系统而言，这些细小数据可能无关痛痒，但一旦将它们通过自动化技术整合后，就会逐渐还原和预测个人生活的轨迹和全貌，使个人隐私无所遁形。

3.1.3大数据安全问题

•据哈佛大学研究显示，只要知道一个人的年龄、性别和邮编，就可以在公开的数据库中识别出此人87%的身份。在模拟和小数据时代，一般只有政府机构才能掌握个人数据，而如今许多企业、社会组织也拥有海量数据，甚至在某些方面超过政府，这些海量数据的汇集使敏感数据暴露的可能性加大，对大数据的收集、处理、保存不当更是会加剧数据信息泄露的风险。

•人类进入大数据时代以来，数据泄密事件时有发生。

3.1.3大数据安全问题

•2017年，京东试用期员工与网络黑客勾结，盗取涉及交通、物流、医疗等个人信息50亿条，在网络黑市贩卖。
•2018年6月，一位ID为“f666666”的用户在暗网上开始兜售圆通10亿条快递数据，该用户表示售卖的数据为2014年下半年的数据，数据信息包括寄（收）件人姓名、电话、地址等信息，10亿条数据已经经过去重处理，数据重复率低于20%，并以1比特币打包出售。
•2018年8月，华住旗下多个连锁酒店开房信息数据正在暗网出售，受到影响的酒店，包括汉庭酒店、美爵、禧玥、漫心、诺富特、美居、CitiGo、桔子、全季、星程、宜必思、怡莱、海友等，泄露数据总数更是近5亿。

3.1.3大数据安全问题

2. 国家安全问题

（1）大数据成为国家之间博弈的新战场

- 大数据意味着海量的数据，也意味着更复杂、更敏感的数据，特别是关系国家安全和利益的数据，如国防建设数据、军事数据、外交数据等，极易成为网络攻击的目标。一旦机密情报被窃取或泄露，就会关系到整个国家的命运。
- “维基解密”（Wikileaks）网站泄露美国军方机密，影响之深远，令美国政府“愤慨”。
- 举世瞩目的“棱镜门”事件，更是昭示着国家安全经历着大数据的严酷挑战。
- 大数据安全已经作为非传统安全因素，受到各国的重视

3.1.3大数据安全问题

2. 国家安全问题

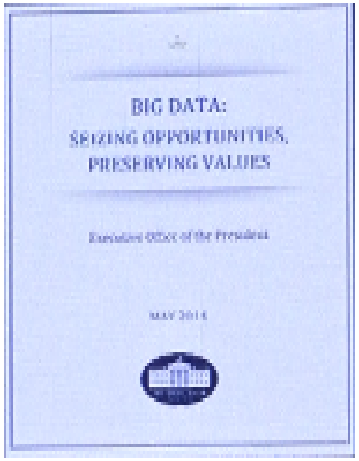
（2）自媒体平台成为影响国家意识形态安全的重要因素

- 自媒体平台包括：博客、微博、微信、抖音、百度官方贴吧、论坛/BBS等网络社区。大数据时代的到来重塑着媒体表达方式,传统媒体不再一枝独秀,自媒体迅速崛起,使得每个人都是自由发声的独立媒体，都有在网络平台有发表自己观点的权力

- 自媒体的发展良莠不齐,一些自媒体平台上垃圾文章、低劣文章层出不穷，甚至一些自媒体为了追求点击率,不惜突破道德底线发布虚假信息,受众群体难以分辨真伪,冲击了主流发布的权威性0

3.1.4 典型案例

1.棱镜门事件



美国政府2014年5月发布的大数据报告：大数据可以极大增强国家安全保证能力

美国前国防部长**拉姆斯菲尔德**多次强调：

一枚导弹没有一条情报

能更有效地应对恐怖活动

3.1.4 典型案例

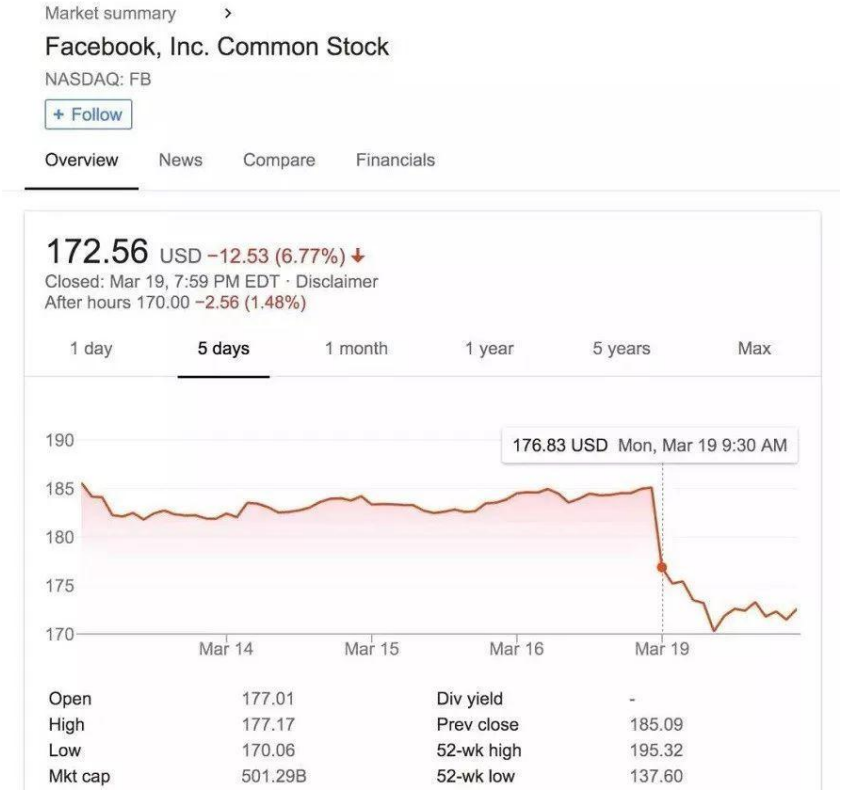
2. 维基解密

•维基解密(WikiLeaks)是一个国际性非营利的媒体组织，专门公开来自匿名来源和网络泄露的文档

•维基解密的目标是发挥最大的政治影响力。维基解密大量发布机密文件的做法使其饱受争议。支持者认为维基解密捍卫了民主和新闻自由，而反对者则认为大量机密文件的泄露威胁了相关国家的国家安全，并影响国际外交

3.1.4 典型案例

3. Facebook数据滥用事件



心理学+大数据=颠覆世界

美国时间2018年3月19日，Facebook股价暴跌7%，一天内市值蒸发近400亿美元

3.1.4 典型案例

4. 手机应用软件过度采集个人信息

- 在我们的日常生活中，部分手机APP往往会“私自窃密”
- 在微信朋友圈广泛传播的各种测试小程序，也可能在窃取用户个人信息

图 在朋友圈传播的测试小程序

3.1.4 典型案例

5. 12306数据泄露

2014年12月25日，12306订票官方网站被指流出约13万用户数据，其中包括姓名、身份证号、手机号、用户名、密码等敏感信息

3.1.4 典型案例

6. 免费WiFi窃取用户信息

曾经有黑客在某网络论坛发帖称，只需要一台电脑、一套无线网络设备和一个网络包分析软件，他就能轻松地搭建出一个不设密码的WiFi网络，而一旦其他用户用移动设备连接上这个WiFi，之后再使用手机浏览器登陆电子邮箱、网络论坛等账号时，他就能很快分析出该用户的各种密码，进而窃取用户的私密信息，甚至利用用户的QQ、微博、微信等通讯工具发布广告诈骗信息

3.1.4 典型案例

7．收集个人隐私信息的“探针盒子”

“探针盒子”是一款自动收集用户隐私的产品。当用户手机无线局域网处于打开状态时，会向周围发出寻找无线网络的信号，探针盒子发现这个信号后，就能迅速识别出用户手机的MAC地址，转换成IMEI号，再转换成手机号码

3.2大数据思维

3.2.1 传统的思维方式
3.2.2大数据时代需要新的思维方式
3.2.3大数据思维方式
3.2.4运用大数据思维的具体实例

3.2.1 传统的思维方式

机械思维

- 第一，世界变化的规律是确定的，这一点从托勒密到牛顿大家都认可。
- 第二，因为有确定性做保障，因此规律不仅是可以被认识的，而且可以用简单的公式或者语言描述清楚。这一点在牛顿之前，大部分人并不认可，而是简单地把规律归结为神的作用。
- 第三，这些规律应该是放之四海而皆准的，可以应用到各种未知领域指导实践，这种认识是在牛顿之后才有的。

3.2.2大数据时代需要新的思维方式

- 不确定性在我们生活的世界里无处不在，由于不确定性是这个世界的重要特征，以至于我们按照传统的方法——机械论的方法，很难做出准确的预测
- 世界的不确定性，折射出在信息时代的方法论：获得更多的信息，有助于消除不确定性，因此，谁掌握了信息，谁就能够获取财富，这就如同在工业时代，谁掌握了资本谁就能获取财富一样。
- 数据学家认为，世界的本质是数据。通过采集、量化、计算、分析各种事物，来重新解释和定义这个世界，并通过数据来消除不确定性，对未来加以预测
- 转变思维方式，努力把身边的事物量化，以数据的形式加以对待，这是实现大数据时代思维方式转变的“核心”

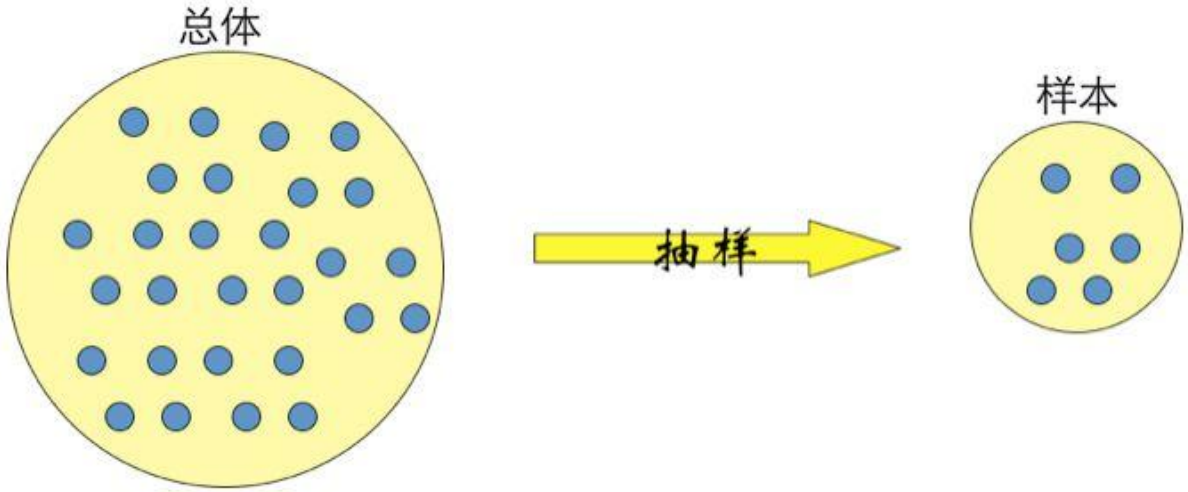
3.2.3大数据思维方式

1.全样而非抽样
2.效率而非精确
3.相关而非因果
4.以数据为中心
5.我为人人，人人为我

3.2.3大数据思维方式

1.全样而非抽样

数据太多，无法保存和分析，统计学采用抽样



3.2.3大数据思维方式

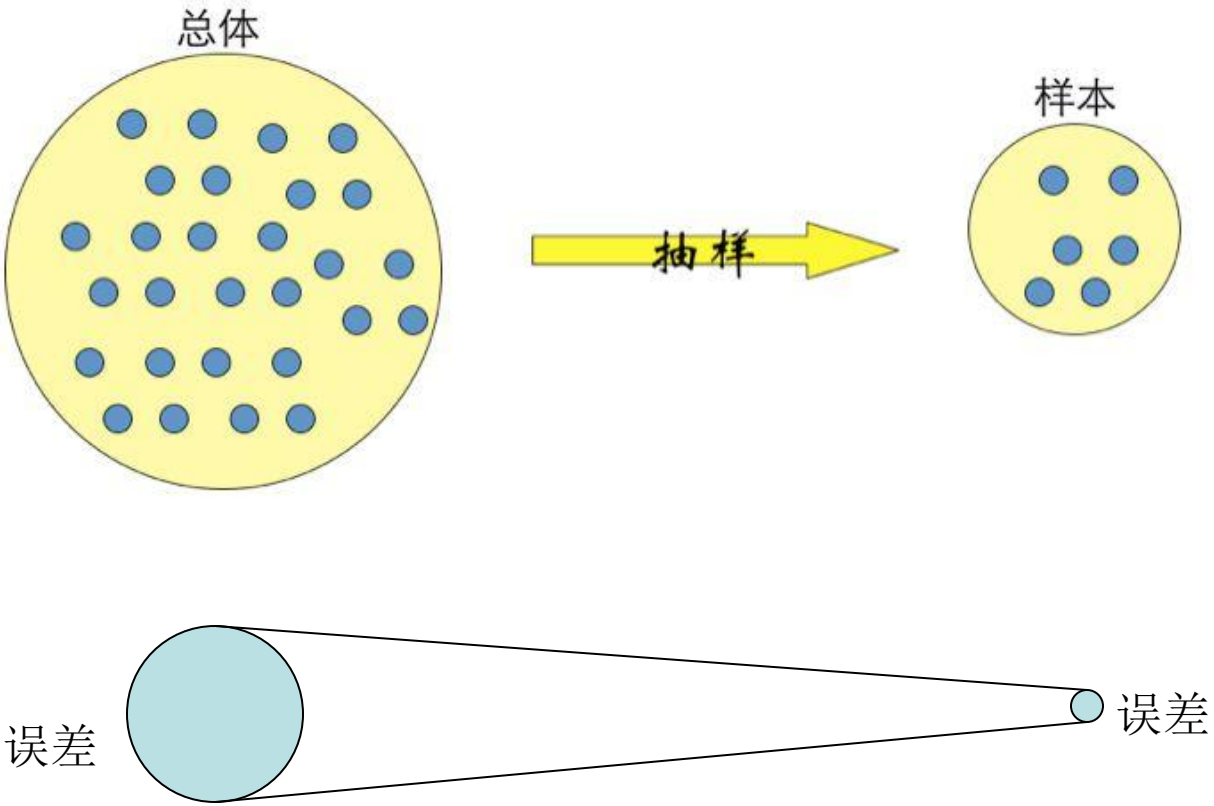
2.效率而非精确

追求效率

不是追求精确

3.2.3大数据思维方式

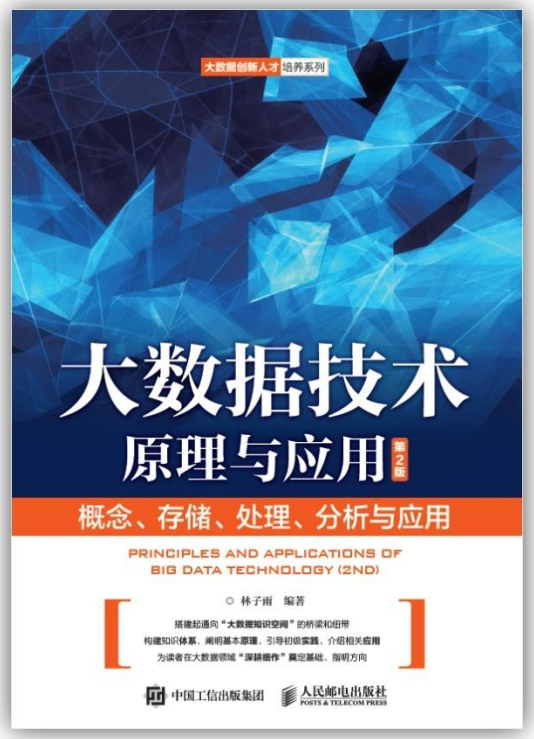
2.效率而非精确



抽样计算的结果误差，放到全样上，会被放大

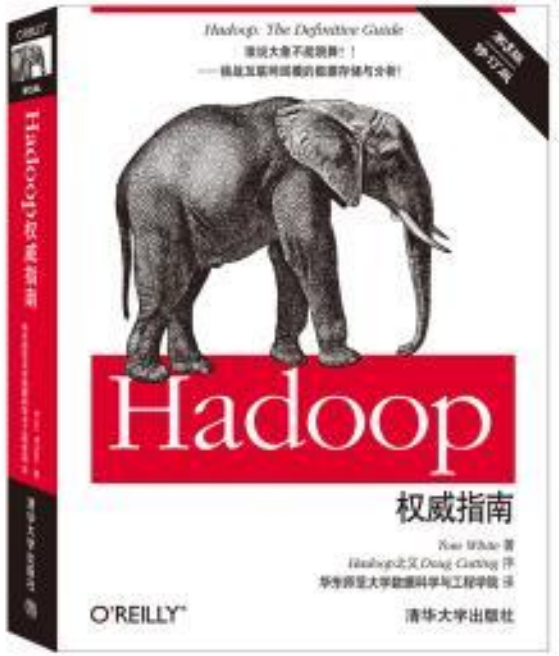
3.2.3大数据思维方式

3.相关而非因果



用户在网店购买一本书

相关性



网店自动推荐相关的另一本书

3.2.3大数据思维方式

4. 以数据为中心

•数据驱动方法从 20 世纪 70 年代开始起步，在八九十年代得到缓慢但稳步的发展。进入 21 世纪后，由于互联网的出现，使得可用的数据量剧增，数据驱动方法的优势越来越明显，最终完成了从量变到质变的飞跃。如今很多需要类似人类智能才能做的事情，计算机已经可以胜任了，这得益于数据量的增加。

•全世界各个领域数据不断向外扩展，渐渐形成了另外一个特点，那就是很多数据开始出现交叉，各个维度的数据从点和线渐渐连成了网，或者说，数据之间的关联性极大地增强，在这样的背景下，就出现了大数据，使得“以数据为中心”的思考解决问题的方式优势逐渐得到显现。

3.2.3大数据思维方式

5. 我为人人，人人为我

每个使用导航软件的智能手机用户，一方面共享自己的实时位置信息给导航软件公司（比如百度地图），使得导航软件公司可以从大量用户那里获得实时的交通路况大数据，另一方面，每个用户又在享受导航软件公司提供的基于交通大数据的实时导航服务。

3.2.4运用大数据思维的具体实例

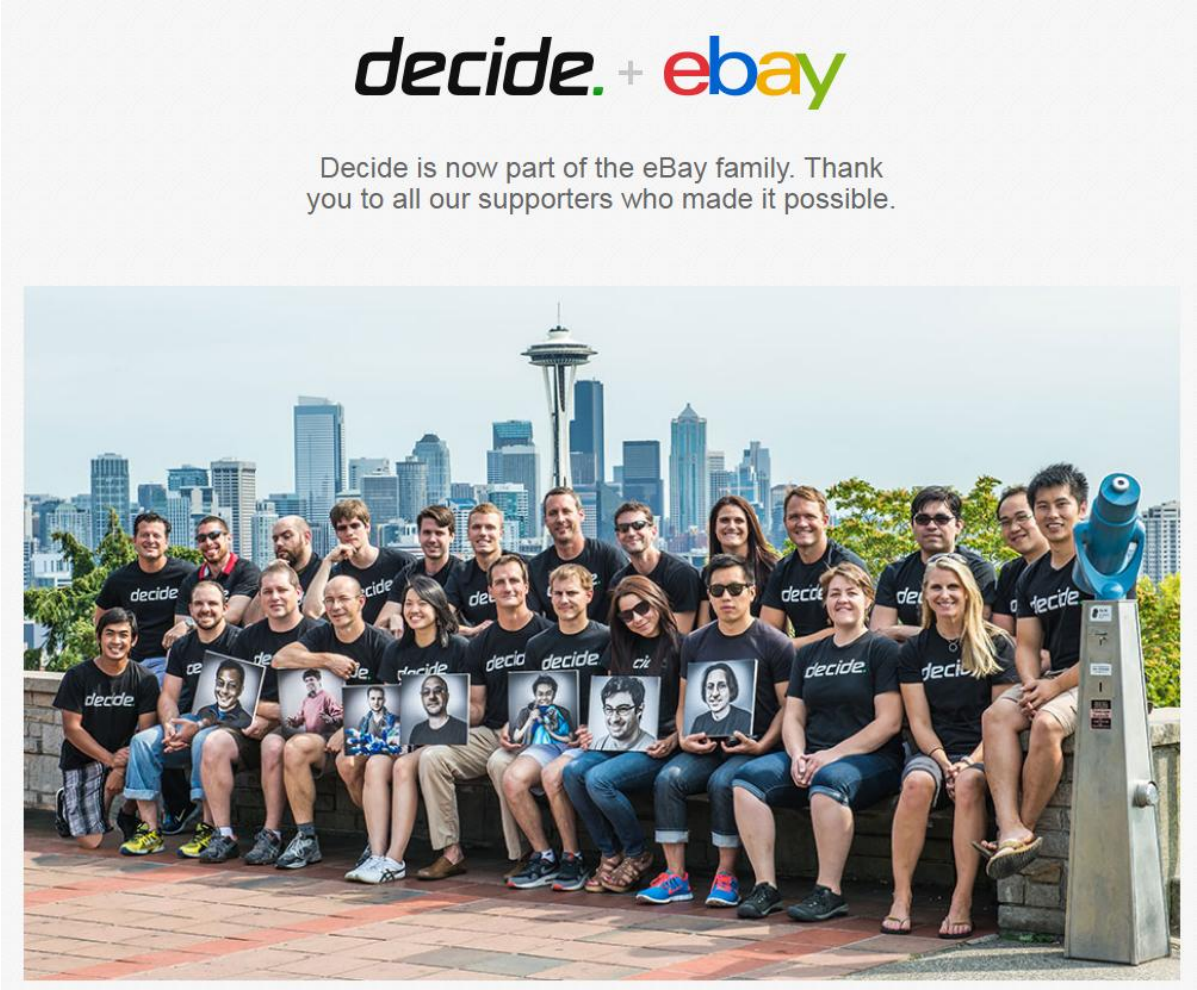
1.商品比价网站Decide.com
2.啤酒与尿布
3.零售商Target的基于大数据的商品营销
4.吸烟有害身体健康的法律诉讼
5.基于大数据的药品研发
6.基于大数据的谷歌广告
7.搜索引擎“点击模型”
8.迪士尼MagicBand手环
9.谷歌流感趋势预测
10.大数据的简单算法比小数据的复杂算法更有效
11.谷歌翻译

3.2.4运用大数据思维的具体实例

思维方式	具体实例
全样而非抽样	商品比价网站Decide.com 谷歌流感趋势预测
效率而非精确	谷歌翻译
相关而非因果	啤酒与尿布 零售商Target的基于大数据的商品营销 吸烟有害身体健康的法律诉讼 基于大数据的药品研发
以数据为中心	基于大数据的谷歌广告 搜索引擎“点击模型” 大数据的简单算法比小数据的复杂算法更有效
我为人人，人人为我	迪士尼MagicBand手环

3.2.4运用大数据思维的具体实例

1.商品比价网站Decide.com



3.2.4运用大数据思维的具体实例

2.啤酒与尿布

交易号	产品
T01	啤酒
T01	尿布
T02	啤酒
T02	尿布
T03	尿布

3.2.4运用大数据思维的具体实例

3.零售商Target的基于大数据的商品营销

3.2.4运用大数据思维的具体实例

4.吸烟有害身体健康的法律诉讼

3.2.4运用大数据思维的具体实例

5.基于大数据的药品研发

3.2.4运用大数据思维的具体实例

6.基于大数据的谷歌广告

3.2.4运用大数据思维的具体实例

7. 搜索引擎“点击模型”

微软

谷歌

3.2.4运用大数据思维的具体实例

8. 迪士尼MagicBand手环

3.2.4运用大数据思维的具体实例

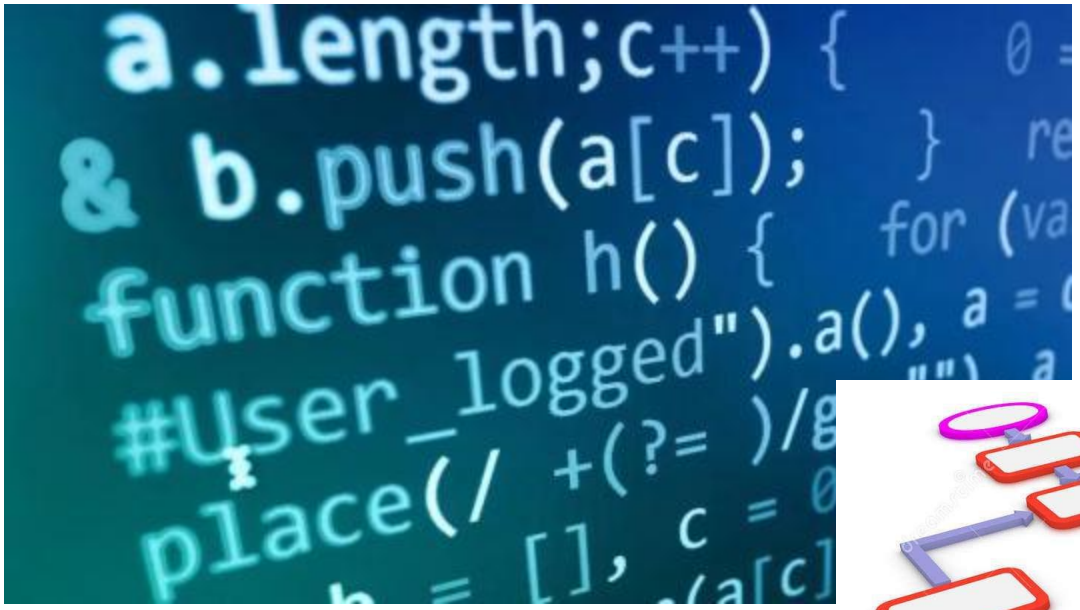
9. 谷歌流感趋势预测

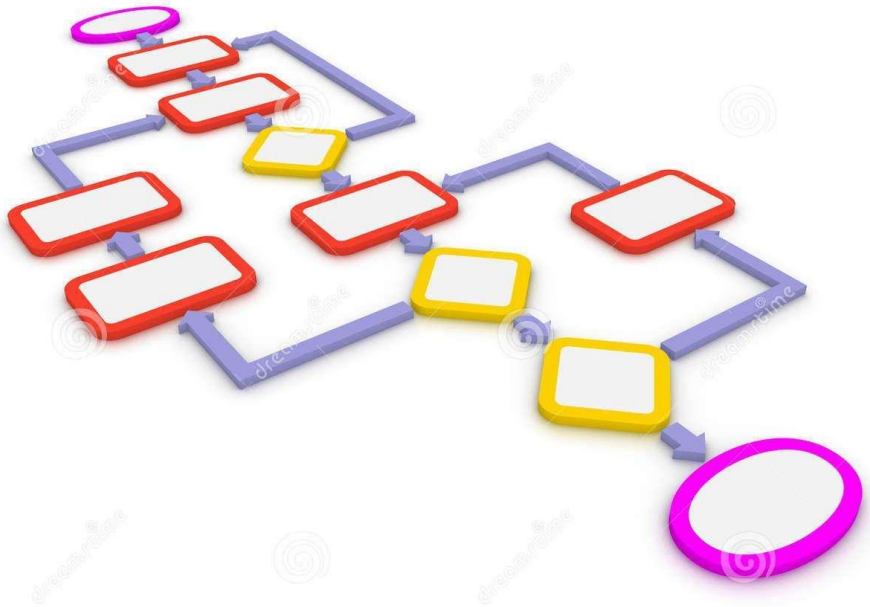
从谷歌流感趋势看大数据的应用价值

“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

3.2.4运用大数据思维的具体实例

10.大数据的简单算法比小数据的复杂算法更有效





3.2.4运用大数据思维的具体实例

11.谷歌翻译

Google 翻译

3.3大数据伦理

3.3.1 大数据伦理概念
3.3.2大数据伦理典型案例
3.3.3大数据的伦理问题

3.3.1 大数据伦理概念

- “伦理”是指一系列指导行为的观念，是从概念角度上对道德现象的哲学思考。它不仅包含着对人与人、人与社会和人与自然之间关系处理中的行为规范，而且也深刻地蕴涵着依照一定原则来规范行为的深刻道理。
- 科技伦理是指科学技术创新与运用活动中的道德标准和行为准则，是一种观念与概念上的道德哲学思考。它规定了科学技术共同体应遵守的价值观、行为规范和社会责任范畴。
- “大数据伦理问题”，就属于科技伦理的范畴，指的是由于大数据技术的产生和使用而引发的社会问题，是集体和人与人之间关系的行为准则问题。

3.3.2大数据伦理典型案例

1.大麦网“撞库”事件

所谓的“撞库”是黑客通过收集互联网已泄露的用户和密码信息，生成对应的字典表，尝试批量登陆其他网站后，得到一系列可以登录的用户。很多用户在不同网站使用的是相同的帐号密码，因此黑客可以通过获取用户在A网站的账户从而尝试登录B网站，这就可以理解为撞库攻击。也就是说撞库简单的理解就是：黑客“凑巧”获取到了一些用户的数据(用户名密码)，再应用到其他网站登录系统。

3.3.2大数据伦理典型案例

2.大数据“杀熟”

3.3.2大数据伦理典型案例

3.隐性偏差问题

3.3.2大数据伦理典型案例

4.“信息茧房”问题

3.3.3大数据的伦理问题

1.隐私泄露问题
2.数据安全问题
3.数字鸿沟问题
4.数据独裁问题
5.数据垄断问题
6.数据的真实可靠问题
7.人的主体地位问题

3.3.3大数据的伦理问题

1.隐私泄露问题

- 大数据时代下的隐私与传统隐私的最大区别在于隐私的数据化，即隐私主要以“个人数据”的形式出现。而在大数据时代，个人数据随时随地可被收集，它的有效保护面临着巨大的挑战
- 进入大数据时代，就进入了一张巨大且隐形的监控网中，我们时刻被暴露在“第三只眼”的监视之下，并留下一条永远存在的“数据足迹”
- 这些直接被采集的数据，已经涉及到个人的很多隐私，此外，针对这些数据的二次使用，还会给个体带来更多的隐私权侵犯

3.3.3大数据的伦理问题

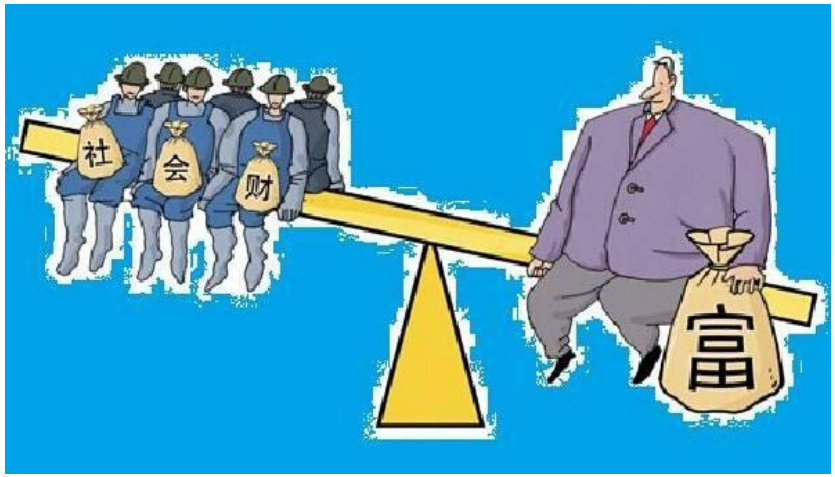
2.数据安全问题

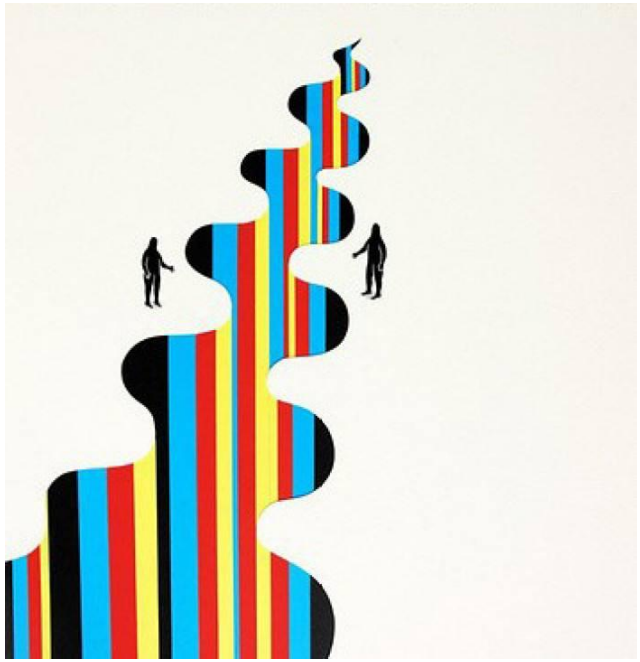
- 一些信息技术本身就存在安全漏洞，可能导致数据泄露、伪造、失真等问题，影响数据安全。
- 智能手机是当今泄漏用户数据的重要途径
- 部分智能家居产品存在安全问题也是不争的事实，给用户的数据安全带来了极大的风险，造成用户隐私的泄露

3.3.3大数据的伦理问题

3.数字鸿沟问题

数字鸿沟总是指向信息时代的不公平，尤其在信息基础设施、信息工具以及信息的获取与使用等领域，或者可以认为是信息时代的“马太效应”，即先进技术的成果不能为人公正分享，于是造成“富者越富、穷者越穷”的情况。

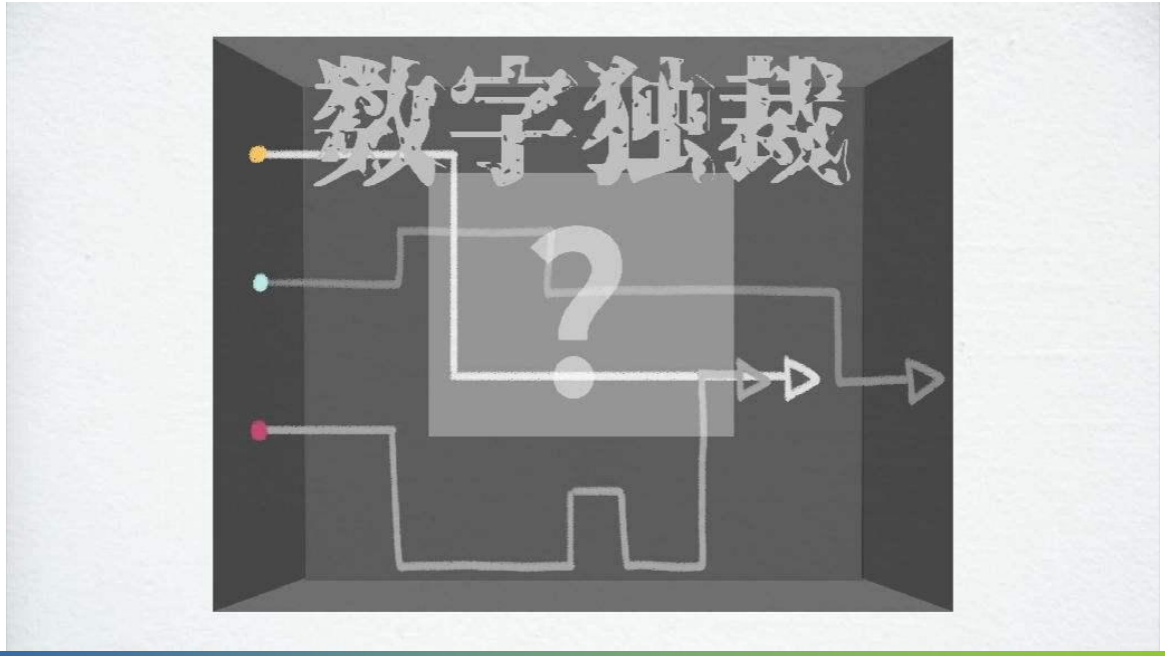




3.3.3大数据的伦理问题

4.数据独裁问题

所谓的“数据独裁”是指在大数据时代，由于数据量的爆炸式增长，导致做出判断和选择的难度徒增，迫使人们必须完全依赖数据的预测和结论才能做出最终的决策。从某个角度来讲，就是让数据统治人类，使人类彻底走向唯数据主义。



3.3.3大数据的伦理问题

5.数据垄断问题

- 企业掌握的数据量越多，越有利于发挥数据的作用，也越有利于最大化消费者福利和社会福利
- 有些企业为了获取更高的经济利益，而故意地不进行数据信息的共享，将所有的数据信息掌握在自己的手中，进行了大数据的垄断
- 因数据而产生的垄断问题，至少包括以下几类：一是数据可能造成进入壁垒或扩张壁垒，二是拥有大数据形成市场支配地位并滥用，三是因数据产品而形成市场支配地位并滥用，四是涉及数据方面的垄断协议，五是数据资产的并购
- 一旦大数据企业形成数据垄断，就会出现消费者在日常生活中被迫地接受服务及提供个人信息的情况

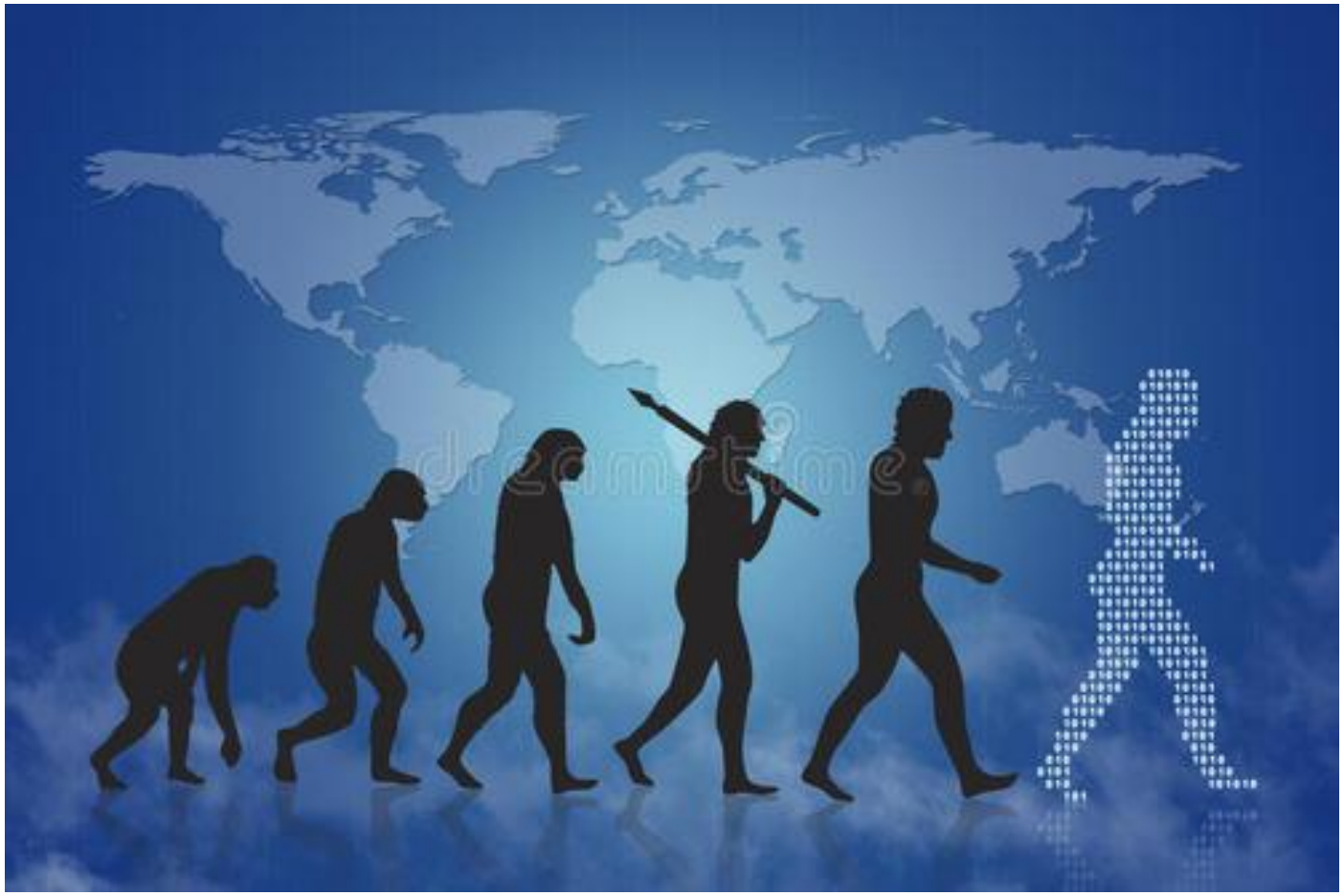
3.3.3大数据的伦理问题

6.数据的真实可靠问题

3.3.3大数据的伦理问题

7.人的主体地位问题

在一切皆数据的条件下，人的主体地位逐渐消失



3.4数据共享

3.4.1数据孤岛问题
3.4.2 数据孤岛问题产生的原因
3.4.3 消除数据孤岛的重要意义
3.4.4 实现数据共享所面临的挑战
3.4.5 推进数据共享开放的举措
3.4.6 数据共享案例

3.4.1数据孤岛问题

1.政府的数据孤岛问题

- 由于各政府部门建设数据库所采用的技术、平台及网络标准不统一，导致政府职能部门之间难以实现数据对接与共享
- 纵向上各级垂直管理部门建设的政府信息系统形成“数据烟囱”，横向上部门间各业务条块则自建系统形成“数据孤岛”，政府公共信息资源的存储彼此独立、管理分散
- 作为政府最重要资产之一的政务数据，因为数据量太大、太散、难以有效融合等问题，严重影响到了数据价值的发挥，大大浪费了各地政府部门在信息化系统建设方面的大量投入

3.4.1数据孤岛问题

2.企业的数据孤岛问题

企业管理职能精细划分，信息系统围绕不同的管理阶段和管理职能展开，如客户管理系统、生产系统、销售系统、采购系统、订单系统、仓储系统和财务系统等，所有数据被封存在各系统中，让完整的业务链上孤岛林立，信息的共享、反馈难，数据孤岛问题是企业信息化建设中的最大难题

3.4.2 数据孤岛问题产生的原因

1. 政府数据孤岛的产生原因

- 有些政府部门错误地将数据资源等同于一般资源，认为占有就是财富，热衷于搜集，但不愿共享；
- 有些部门只盯着自己的数据服务系统，结果因为数据标准、系统接口等技术原因，无法与外单位、外部门联通；
- 还有些地方，对大数据缺乏顶层设计，导致各条线、各部门固有的本位主义作祟，壁垒林立，数据无法流动

3.4.2 数据孤岛问题产生的原因

2. 企业数据孤岛的产生原因

- 不同企业之间，属于不同的经营主体，有着各自的利益，彼此之间数据不共享，产生企业之间的数据孤岛，这种是比较普遍的情况。
- 企业内部也往往会存在大量数据孤岛，这些数据孤岛的形成主要有两个方面的原因：
  - 以功能为标准的部门划分导致数据孤岛
  - 不同类型、不同版本的信息化管理系统导致数据孤岛

3.4.3 消除数据孤岛的重要意义

1. 对于政府的意义

加强政府数据共享开放和大数据服务能力，促进跨领域、跨部门合作，推进数据信息交换，打破部门壁垒，遏制数据孤岛和重复建设，有助于提高行政效率，转变思维观念，推动传统的职能型政府转型为服务型智慧政府。政府数据共享的重要意义表现在以下两个方面：

- 首先，有助于提升资源利用率
- 其次，有助于推动政府转型

3.4.3 消除数据孤岛的重要意义

2. 对于企业的意义

首先，打通企业内部的数据孤岛，实现所有系统数据互通共享，对建立企业自身的大数据平台和企业信息化建设都有重大意义。

其次，打通企业之间的数据孤岛，实现不同企业的数据共享，有利于企业获得更好的经营发展能力。

3.4.4 实现数据共享所面临的挑战

1. 在政府层面的挑战

- A 不愿共享开放
- B 不敢共享开放
- C 不会共享开放
- D 数据中心共享开放作用不强

3.4.4 实现数据共享所面临的挑战

2.在企业层面的挑战

3.4.5 推进数据共享开放的举措

1.在政府层面的举措

- 积极开放政府数据资源，提高政府职能部门之间和具有不同创新资源的主体之间的数据共享广度，促进区域内形成“数据共享池”
- 要改变政府职能部门“数据孤岛”现象，立足于数据资源的共享互换，设定相对明确的数据标准，实现部门之间的数据对接与共享，推进在制度创新方面的系统集成化，为科技创新提供必要条件
- 要促进准确及时的数据信息传递，提高部门条线管理、“一站式”企业网上办事和政府服务项目“一网通办”的网络信息功能，提高数据质量的可靠性、稳定性与权威性，增加相关信息平台的使用覆盖面，让现存数据“连起来”、“用起来”

3.4.5 推进数据共享开放的举措

2.在企业层面的举措

在企业内部，破除“数据孤岛”，推进数据融合

在不同企业之间，建立企业数据共享联盟

3.4.6 数据共享案例

1.案例1：菜鸟物流
2.案例2：政府一站式平台——i厦门
3.案例3：浙江打通政府数据，让群众最多跑一次

3.4.6 数据共享案例

1. 案例1：菜鸟物流

2013年阿里巴巴集团联合多方力量联手共建“中国智能物流骨干网”（又称“菜鸟”），计划在8～10年的时间，建立一张能支撑日均300亿元（年度约10万亿元）网络零售额的智能物流骨干网络，支持数千万家新型企业成长发展，让全中国任何一个地区做到24小时内送货必达。

3.4.6 数据共享案例

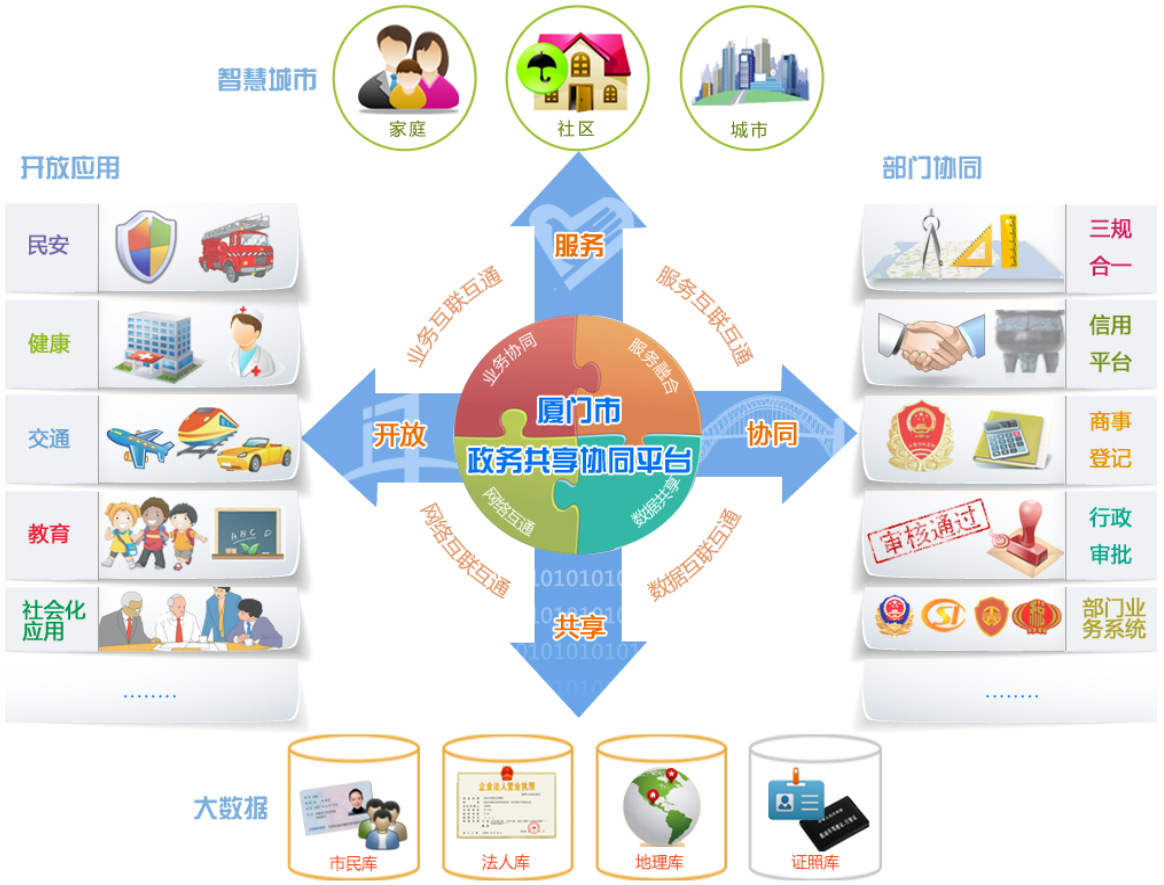
1. 案例1：菜鸟物流

- 菜鸟核心作用的发挥，关键在于其对多方物流数据的有效整合，也就是说，参与菜鸟的相关物流企业，都会把自己企业内部的物流数据（主要是包裹轨迹数据）共享出来，由菜鸟平台对电商和物流数据进行统一整合分析
- 菜鸟平台就相当于中枢协调机构，每个包裹、每家快递从仓库发货就开始接入，揽收、中转、派送信息，整个轨迹都可以显示，这将有利于菜鸟从全局层面帮助快递公司进行运力的统筹调配和规划
- 菜鸟会对后台电商和物流数据进行整合、分析和挖掘，比如根据既往的销售数据来分析预测下一个时期内哪些商品需要提前备货多少量，给予仓储管理商相关的商品陈列建议

3.4.6 数据共享案例

2. 案例2：政府一站式平台——i厦门

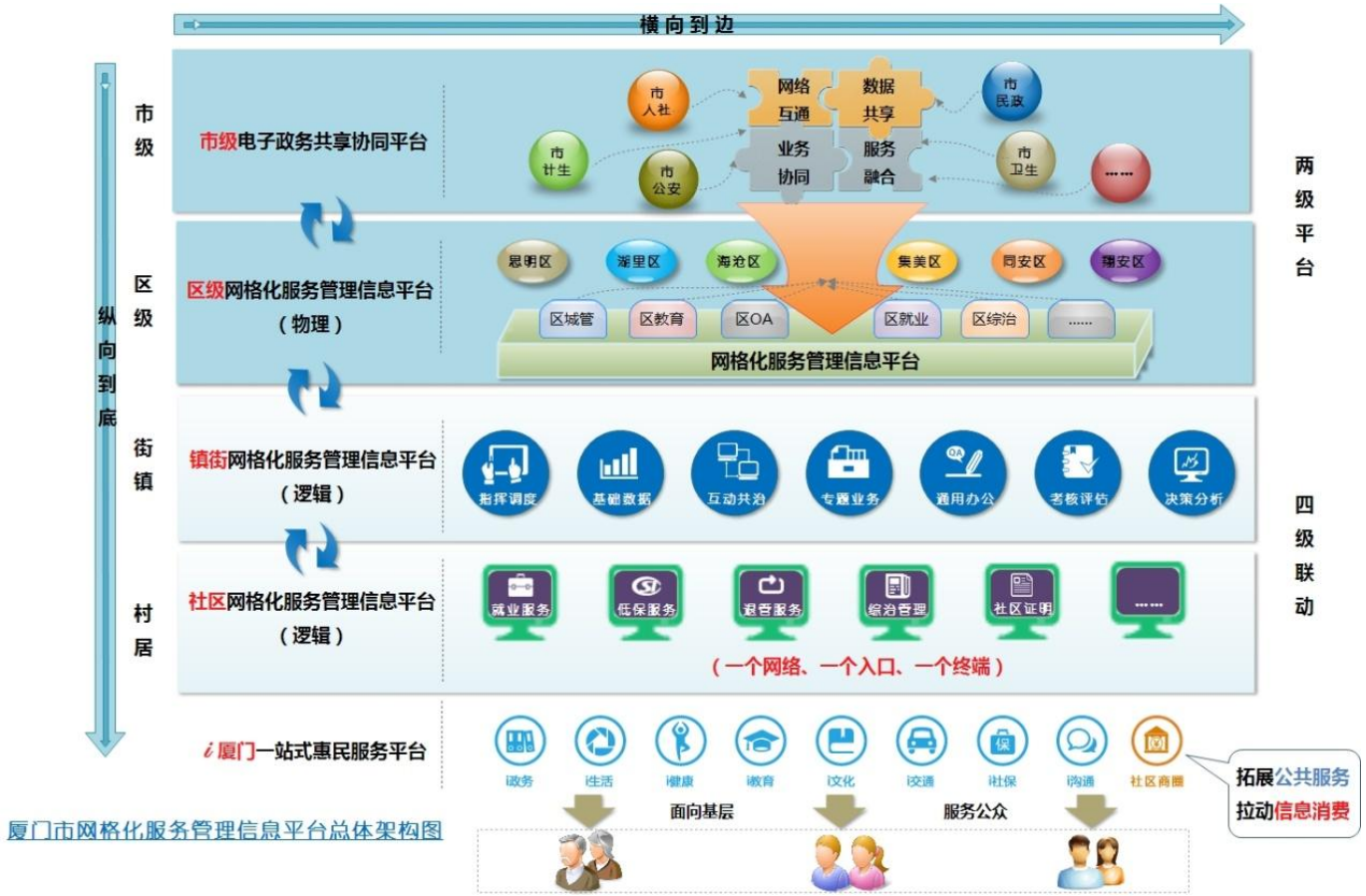
“i厦门”一站式惠民服务平台——服务融合



3.4.6 数据共享案例

2. 案例2：政府一站式平台——i厦门

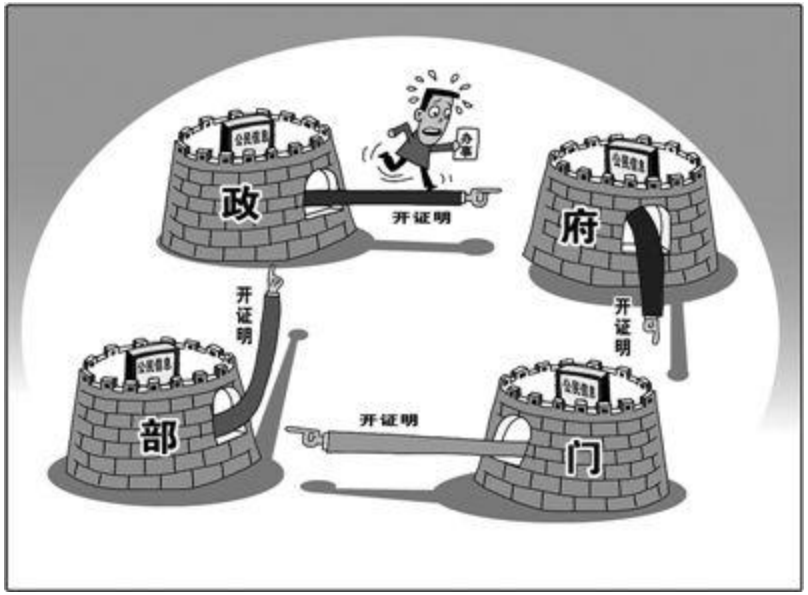
社区网格化服务管理信息平台——业务集成



厦门市网格化服务管理信息平台总体架构图

3.4.6 数据共享案例

3.案例3：浙江打通政府数据，让群众最多跑一次



3.4.6 数据共享案例

3.案例3：浙江打通政府数据，让群众最多跑一次

（1）在社会参保单位办理参保登记方面
（2）在办理不动产登记证方面
（3）在外省就读学生学籍转入方面

3.5数据开放

3.5.1政府开放数据的理论基础
3.5.2政府信息公开与政府数据开放的联系与区别
3.5.3政府数据开放的重要意义

3.5.1政府开放数据的理论基础

1.数据资产理论
2.数据权理论
3.开放政府理论

3.5.1政府开放数据的理论基础

1.数据资产理论

- 在大数据时代，数据已经被当作一种重要的战略资源，也可以成为一种资产。数据资产是无形资产的延伸，是主要以知识形态存在的重要经济资源，是为其所有者或合法使用者提供某种权利、优势和效益的固定资产
- 数据资产的类型有很多，常见的数据资产包括书面技术新材料、数据与文档、技术软件、物理资产（主要指通信协议类）、员工与客户（包括竞争对手）、企业形象和声誉以及服务等
- 同其他资产一样，数据资产也是企业价值创造的工具和资本
- 作为现代企业和政府，拥有数据的规模、活性，以及收集、运用数据的能力，将决定企业和政府的核心竞争力

3.5.1政府开放数据的理论基础

2.数据权理论

•数据权的概念发起于英国，主要将其视为信息社会的一项基本公民权利，让政府所拥有的数据集能够被公众申请和使用，并且按照标准公布数据。因此，早期的数据权理念强调的是公民利用信息的权利

•随着数据的进一步开放，大型网络公司对于历史文献资料的数据化，商业集团对于客户资料的搜集，政府部门对于个人信息的调查与掌握，社会化媒体对于社会交往的渗透与呈现，使国家和政府加强了对数据主权的关注，并将其纳入到数据主权的范畴

•数据主权源于信息主权。信息主权是国家主权在信息活动中的体现，国家对于政权管辖地域内任何信息的制造、传播和交易活动，以及相关的组织和制度拥有最高权力

3.5.1政府开放数据的理论基础

2.数据权理论

- 数据权包括两个方面：数据主权和数据权利
- 数据主权的主体是国家，是一个国家独立自主对本国数据进行管理和利用的权力
- 数据权利的主体是公民，是相对应于公民数据采集义务而形成的对数据利用的权利，这种对数据的利用又是建立在数据主权之下的。只有在数据主权法定框架下，公民才可自由行使数据权利。公民的数据权利，是一项新兴的基本人权，它是信息时代的产物，是公民个人的基本权利。公民数据权的保护，不仅具有正当合理性，而且已经成为一种人权保障的世界性趋势

3.5.1政府开放数据的理论基础

3.开放政府理论

- 开放政府最早出现在20世纪50年代信息自由立法的介绍当中。1957年Park的论文“开放政府原则：依据宪法的知情权”中首次提出开放政府理念，其核心是关于信息自由方面的内容。
- 随着很多国家对信息法案的修订，尤其在2009年奥巴马政府公布了《开放政府指令》后，开放政府的理论又被重新提起。2009年1月21日，在关于政府透明和开放化的备忘录上，奥巴马总统指示美国行政管理预算局局长发布一份《政府开放指令》，开放政府由此提出。
- 自2009年开放政府理念被重新提起后，世界各国都在努力使用信息技术革新政府，并在2011年建立了以美国领导的“开放政府联盟”。

3.5.2政府信息公开与政府数据开放的联系与区别

•政府信息公开与政府数据开放是一对既相互区别又相互联系的概念。

•政府信息公开主要是为了对公众知情权的满足而出现的，信息公开既可以理解为一项制度，又可以理解为一种行为。作为一项制度，主要是指国家和地方制定并用于规范和调整信息公开活动的法规规定；作为一种行为，主要是指掌握信息的主体，即行政机关、单位向不特定的社会对象发布信息，或者向特定的对象提供所掌握的信息的活动

3.5.2政府信息公开与政府数据开放的联系与区别

•政府数据开放是政府信息公开的嬗变必然，将开放对象延伸至原始数据的粒度。政府数据开放强调的是数据的再利用，公众可以分享数据利用创造的经济和社会价值，并且可以根据对数据的分析判断政府的决策是否合理

•政府数据开放强调的是数据的再利用，公众可以分享数据利用创造的经济和社会价值，并且可以根据对数据的分析判断政府的决策是否合理。政府信息公开更侧重对与公众相关信息通过报纸、互联网、电视等媒体的发布，更强调程序公开，正义公开仍是难点

3.5.3政府数据开放的重要意义

1.政府开放数据有利于促进开放透明政府的形成
2.政府开放数据有利于创新创业和经济增长
3.政府开放数据有利于社会治理创新

3.5.3政府数据开放的重要意义

1.政府开放数据有利于促进开放透明政府的形成

•政府开放数据是更高层次的政府信息公开，而政府信息公开也将推动政府民主法治进程

•如果说政府信息公开还是处于起步阶段，那么政府开放数据则是更高层次的政务公开

•数据是政府手中的重要资源，政府开放数据的范围、程度、速度都代表着政府开放的程度

3.5.3政府数据开放的重要意义

2.政府开放数据有利于创新创业和经济增长

•美国是气象灾害频发的国家，为减少气象灾害带来的严重损失，2014年3月，美国白宫宣布：将气象数据发布在Data.gov上，随后，与气象相关的企业服务应运而生，包括各种气象播报、气象顾问、气象保险等，形成了一个新的产业链，创造出了极高的经济价值

•政府数据的再利用，在欧洲也创造出很高的经济价值。2010年欧盟公布的数据显示，欧洲利用政府公开的数据创造出的价值就达到320亿欧元，同时带来了更多的商业和就业机会

3.5.3政府数据开放的重要意义

3.政府开放数据有利于社会治理创新

- 政府数据的开放不仅打破了政府部门对数据的垄断，促进了数据价值的最大发挥，同时也构建起了政府同市场、社会、公众之间互动的平台
- 数据分享和大数据技术应用，不仅可以有效推动政府各部门在公共活动中实现协同治理，提高政府决策的水平，也能够充分调动各方的积极性来完成社会事务，实现社会治理机制的创新，给公众的生活带来便利，比如缓解交通压力、增强食品安全、解决环境污染等

3.6大数据交易

3.6.1概述
3.6.2大数据交易发展现状
3.6.3大数据交易平台

3.6.1概述

- 大数据交易应当是买卖数据的活动，是以货币为交易媒介获取数据这种商品的过程，具有3 种特征：
- 一是标的物受到严格的限制，只有经过处理之后的数据才能交易；
- 二是涉及的主体众多，包括数据提供方、数据购买方、数据平台等；
- 三是交易过程繁琐，涉及大数据的多个产业链，如数据源的获取、数据安全的保障、数据的后续利用等。

3.6.1概述

目前进行数据交易的形式有以下几种：

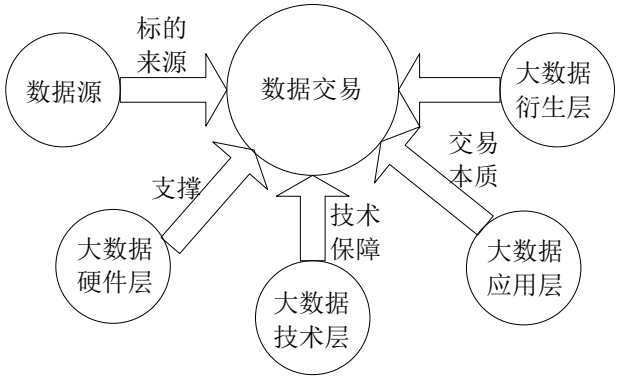
（1）大数据交易公司

（2）数据交易所

（3）API 模式

（4）其他

大数据交易是大数据产业生态系统中的重要一环，与大数据交易相关的其他环节包括数据源、大数据硬件层、大数据技术层、大数据应用层、大数据衍生层等



3.6.2大数据交易发展现状

- 数据交易由来已久，并不是最近几年才出现的新型交易方式
- 进入大数据时代以后，大数据资源愈加丰富
- 庞大的大数据资源为大数据交易的兴起奠定了坚实的基础

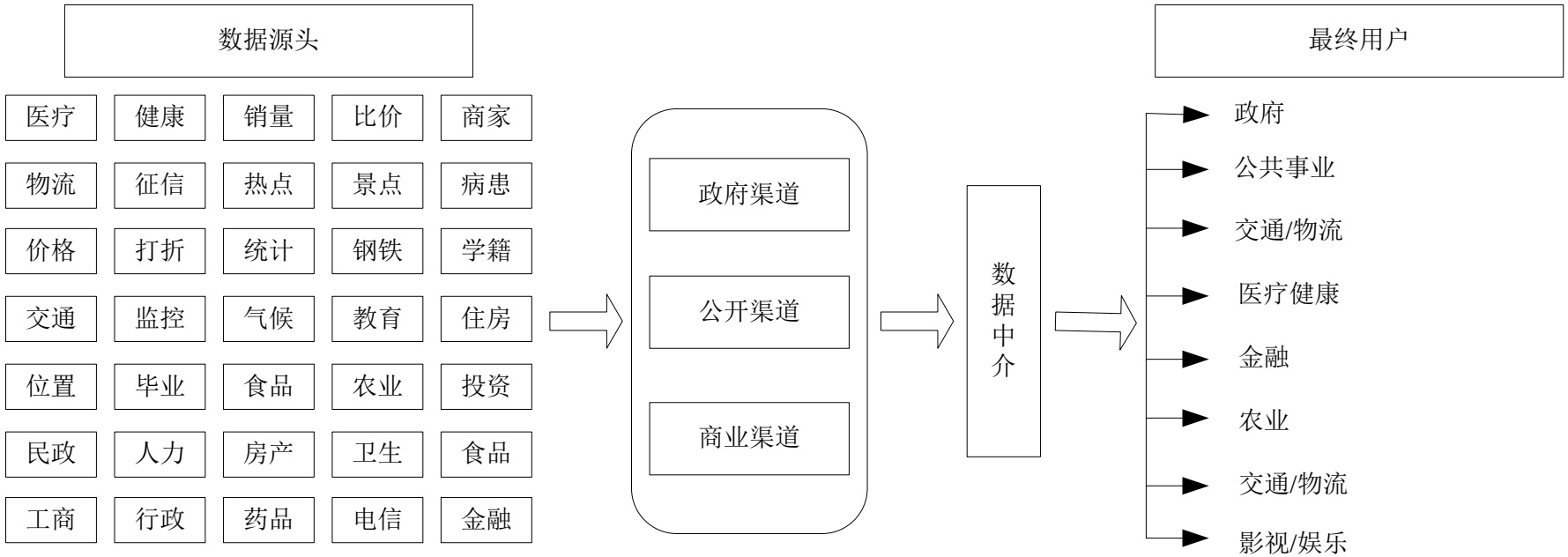


图 数据市场概貌

3.6.2大数据交易发展现状

2014年以来，国内不仅出现了数据堂、京东万象、中关村数海、浪潮卓数、聚合数据等一批数据交易平台，各地方政府也成立了混合所有制形式的数据交易机构，包括贵阳大数据交易所、上海数据交易中心、长江大数据交易中心(武汉)、浙江大数据交易中心等

3.6.3大数据交易平台

1.交易平台的类型
2.交易平台的数据来源
3.交易平台的产品类型
4.交易平台涉及的主要领域
5.平台的交易规则
6.交易平台的运营模式
7.代表性的大数据交易平台

3.6.3大数据交易平台

1.交易平台的类型

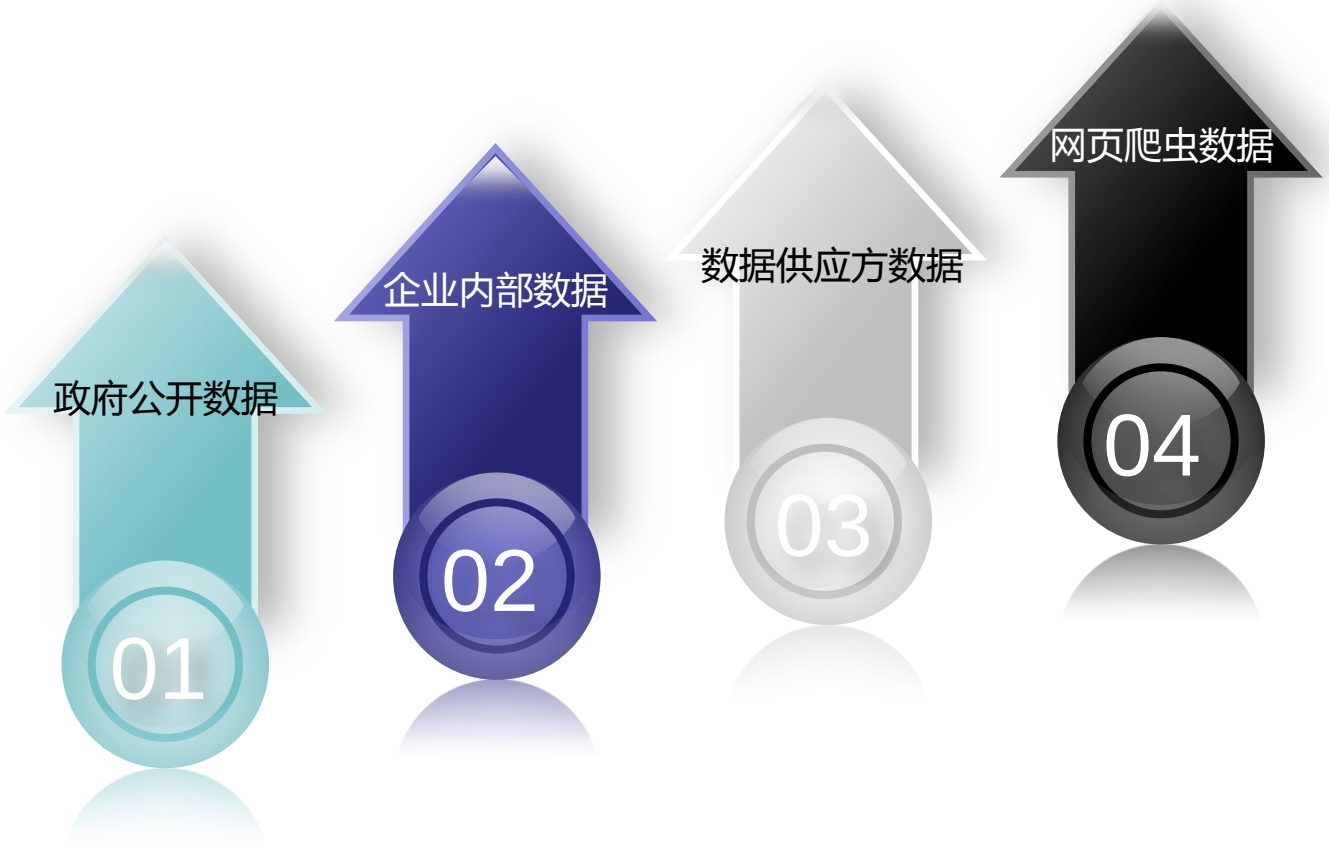
•大数据交易平台主要包括综合数据服务平台和第三方数据交易平台两种。

•综合数据服务平台为用户提供定制化的数据服务，由于需要涉及数据的处理加工，因此，该类型平台的业务相对复杂，国内大数据交易平台大多属于这种类型。

•而第三方数据交易平台业务则相对简单明确，主要负责对交易过程的监管，通常可以提供数据出售、数据购买、数据供应方查询以及数据需求发布等服务。

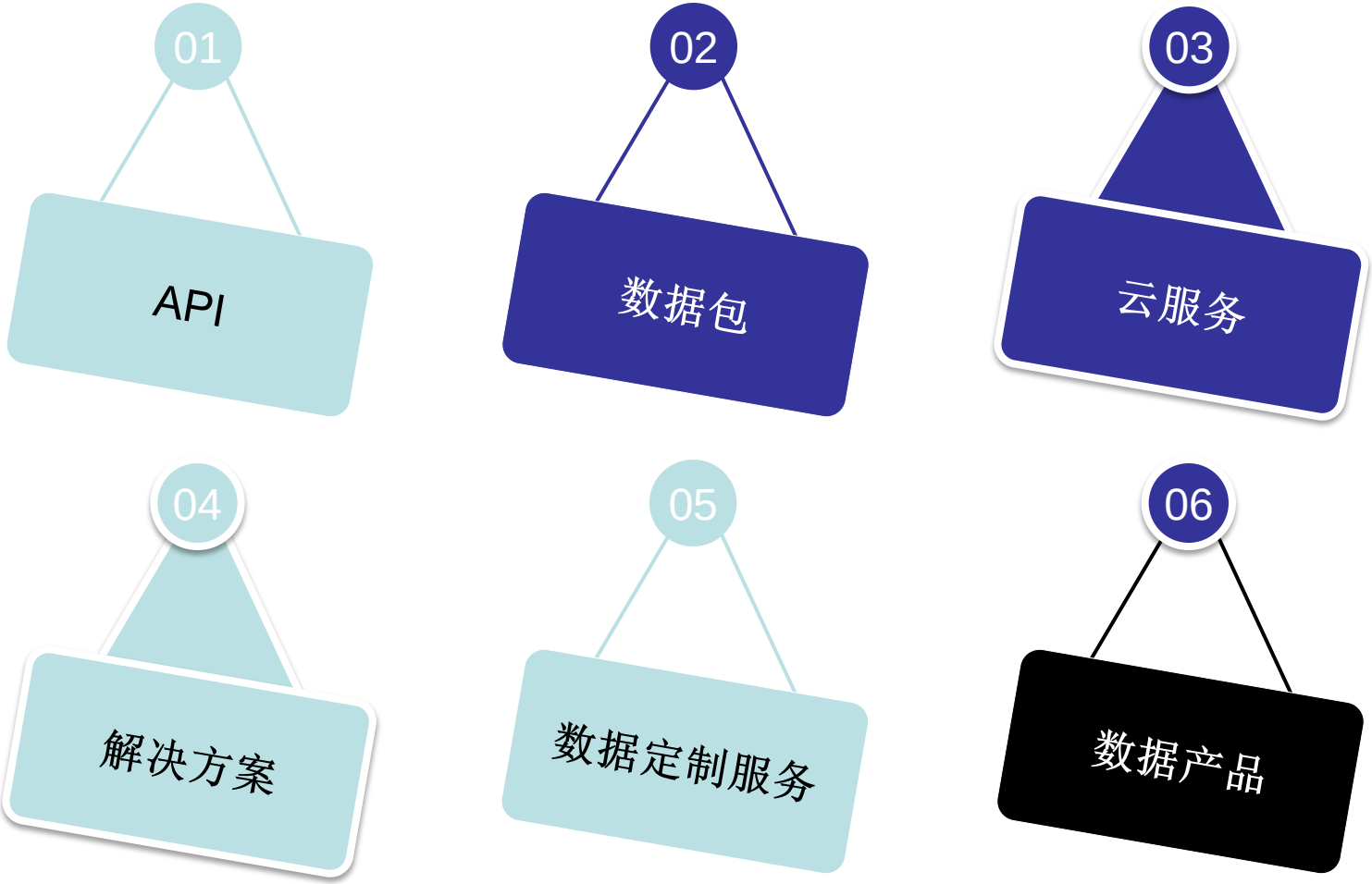
3.6.3大数据交易平台

2.交易平台的数据来源



3.6.3大数据交易平台

3.交易平台的产品类型



3.6.3大数据交易平台

4.交易平台涉及的主要领域

- 国内外大数据交易平台产品涉及的主要领域包括政府、经济、教育、环境、法律、医疗、人文、地理、交通、通信、人工智能、商业、农业、工业等。了解交易平台产品涉及的主要领域，可以帮助用户根据自己的个性化需求有针对性地选择合适的交易平台。
- 国内外交易平台基本上都涉及到多个领域，平台提供的多领域数据，可以较好满足目前广泛存在的用户对跨学科、跨领域数据的需求。

3.6.3大数据交易平台

5.平台的交易规则

- 相对于国外的数据交易公司来说，国内的数据交易平台大多发布了成系统的总体规则，规定更详细，在很多方面也更严格。如《中关村数海大数据交易平台规则》、《贵阳大数据交易所702公约》、《上海数据交易中心（ChinaDEP）数据交易规则》等，以条文的形式对整个平台的运营体系、遵守原则都进行了详细规定，明确了交易主体、交易对象、交易资格、交易品种、交易格式、数据定价、交易融合和交易确权等内容。
- 随着我国数据流通行业的发展，部分企业间已经推出了跨企业的数据交易规则或自律准则。可以说，目前我国建立广泛的数据流通行业自律公约的时机已经相对成熟，行业内部各企业对数据交易自律性协议的需求呼之欲出。

3.6.3大数据交易平台

6.交易平台的运营模式

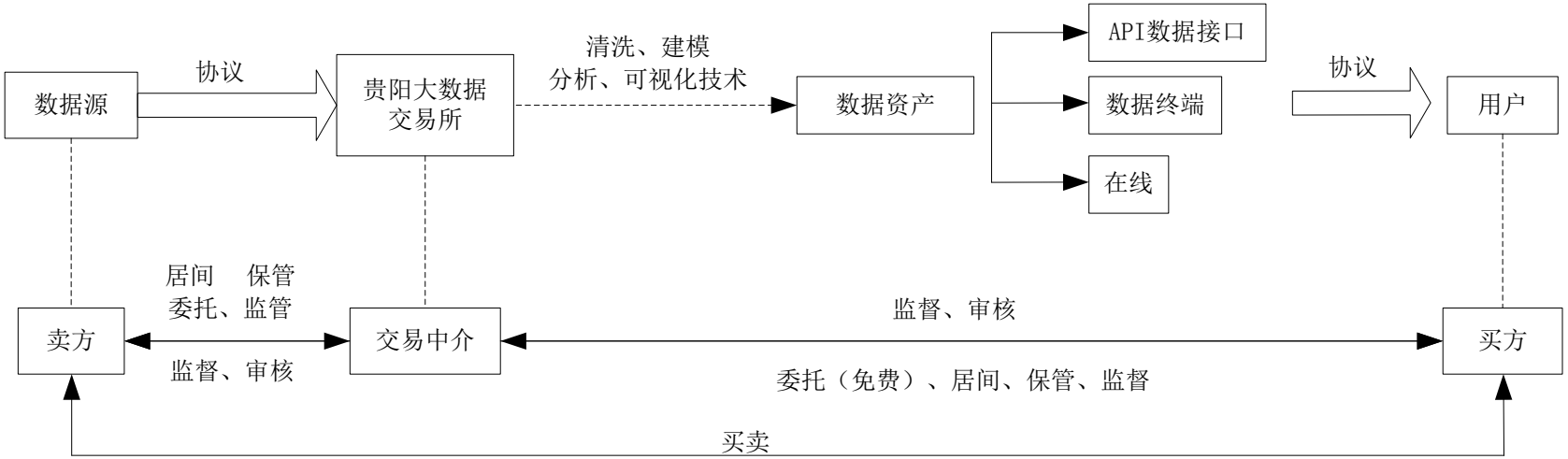


图 贵阳大数据交易所运营模式

3.6.3大数据交易平台

6.交易平台的运营模式

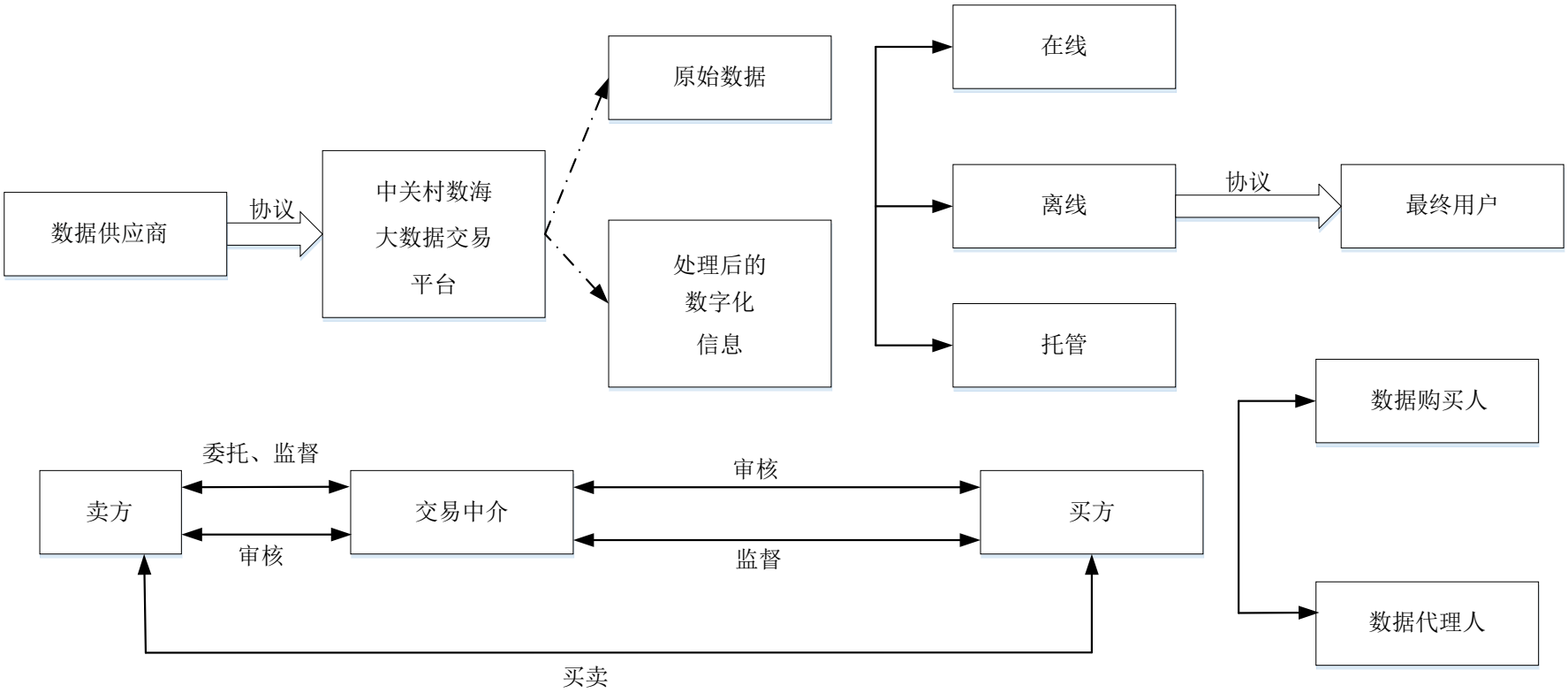


图 中关村数海大数据交易平台运营模式

3.6.3大数据交易平台

7.代表性的大数据交易平台

3.7 本章小结

数据素养教育是大数据专业人才培养的核心内容。国内外高校都十分重视学生数据素养的教育，早在2007年就有国外学者提出培养学生的数据素养，提高他们在21世纪基本的批判性思考能力。数据素养的教育内容不仅包括复杂的大数据专业技能的学习（比如编程语言、操作系统、网络、数据库、大数据处理架构等），还包括大数据基础知识的学习（比如大数据安全、大数据思维、大数据伦理等）。本章就围绕非技术性内容做了大量的论述，详细讨论了大数据安全、大数据思维、大数据伦理、数据共享、数据开放、大数据交易等内容。这些内容的学习，为培养学生的数据素养奠定了坚实的基础。

附录A：主讲教师林子雨简介

主讲教师：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://dblab.xmu.edu.cn/post/linziyu
数据库实验室网站：http://dblab.xmu.edu.cn

扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），全国高校知名大数据教师，现为厦门大学计算机科学系副教授，曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度、2017年度和2020年度厦门大学教学类奖教金获得者，荣获2019年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》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。作为项目负责人主持的科研项目包括1项国家自然科学青年基金项目(No.61303004)、1项福建省自然科学青年基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过1000万字高价值的研究和教学资料，累计网络访问量超过1000万次。打造了中国高校大数据教学知名品牌，编著出版了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》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过200万次，累计访问量超过1000万次。

附录B：大数据学习路线图

大数据学习路线图访问地址：http://dblab.xmu.edu.cn/post/10164/

附录C：林子雨大数据系列教材

了解全部教材信息：http://dblab.xmu.edu.cn/post/bigdatabook/

附录D：《大数据导论（通识课版）》教材

开设全校公共选修课的优质教材

本课程旨在实现以下几个培养目标：

- 引导学生步入大数据时代，积极投身大数据的变革浪潮之中
- 了解大数据概念，培养大数据思维，养成数据安全意识
- 认识大数据伦理，努力使自己的行为符合大数据伦理规范要求
- 熟悉大数据应用，探寻大数据与自己专业的应用结合点
- 激发学生基于大数据的创新创业热情

高等教育出版社 ISBN:978-7-04-053577-8 定价：32元

教材官网：http://dblab.xmu.edu.cn/post/bigdataintroduction/

附录E：《大数据技术原理与应用》教材

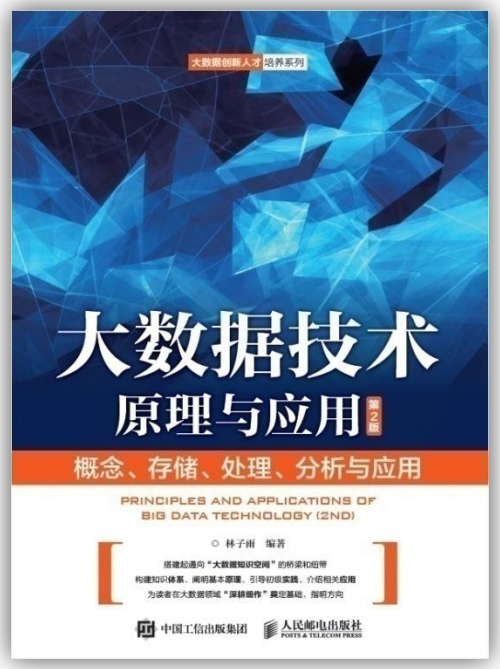
《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社　ISBN:978-7-115-44330-4　定价：49.80元

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据 库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物医学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：
http://dblab.xmu.edu.cn/post/bigdata

扫一扫访问教材官网



附录F：《大数据基础编程、实验和案例教程》

本书是与《大数据技术原理与应用（第2版）》教材配套的唯一指定实验指导书

大数据教材

1+1黄金组合

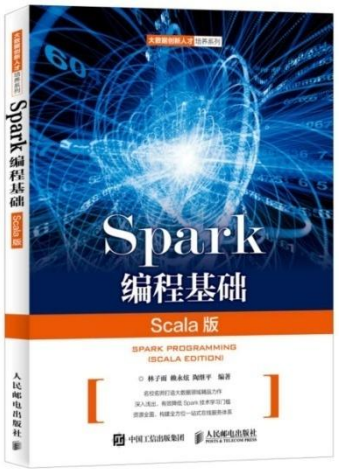
厦门大学林子雨编著

配套实验指导书

- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，五套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

林子雨编著《大数据基础编程、实验和案例教程》
清华大学出版社 ISBN:978-7-302-47209-4 定价：59元

附录G：《Spark编程基础（Scala版）》



《Spark编程基础（Scala版）》

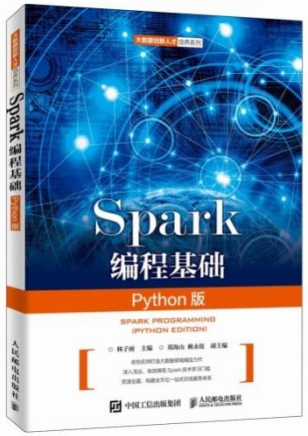
厦门大学 林子雨，赖永炫，陶继平 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-48816-9
教材官网：http://dblab.xmu.edu.cn/post/spark/

本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。

附录H：《Spark编程基础（Python版）》



《Spark编程基础（Python版）》

厦门大学 林子雨，郑海山，赖永炫 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-52439-3

教材官网： http://dblab.xmu.edu.cn/post/spark-python/

本书以Python作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Structured Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、上机实验指南等。

附录I：高校大数据课程公共服务平台

高校大数据课程

公 共 服 务 平 台

http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/

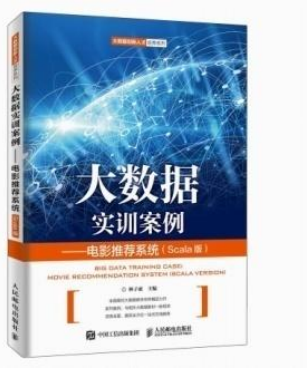
扫一扫访问平台主页

扫一扫观看3分钟FLASH动画宣传片

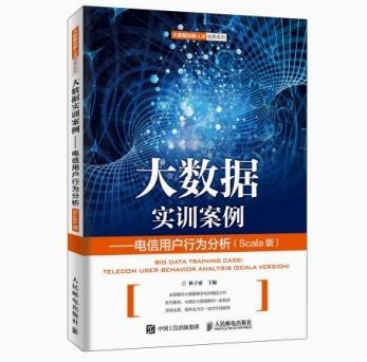
附录J：高校大数据实训课程系列案例教材

为了更好满足高校开设大数据实训课程的教材需求，厦门大学数据库实验室林子雨老师团队联合企业共同开发了《高校大数据实训课程系列案例》，目前已经完成开发的系列案例包括：

《电影推荐系统》（已经于2019年5月出版）
《电信用户行为分析》（已经于2019年5月出版）
《实时日志流处理分析》
《微博用户情感分析》
《互联网广告预测分析》
《网站日志处理分析》



系列案例教材将于2019年陆续出版发行，教材相关信息，敬请关注网页后续更新！
http://dblab.xmu.edu.cn/post/shixunkecheng/



扫一扫访问大数据实训课程系列案例教材主页


Thank You!