



《大数据导论（通识课版）》

教材官网: <http://dbllab.xmu.edu.cn/post/bigdataintroduction/>

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7章 大数据伦理

(PPT版本号: 2019年秋季学期)

林子雨

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>



扫一扫访问教材官网





课程教材

- 林子雨 编著 《大数据导论——数据思维、数据能力和数据伦理（通识课版）》
- 高等教育出版社，2019年11月



提纲

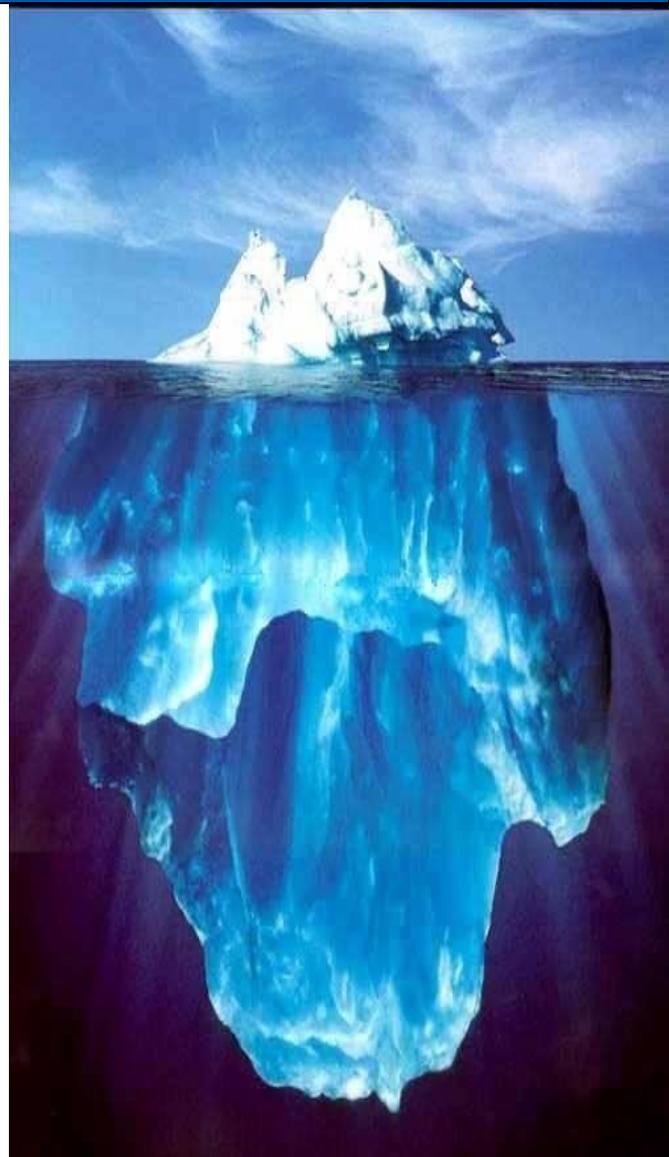
- 7.1 大数据伦理概念
- 7.2 大数据伦理典型案例
- 7.3 大数据的伦理问题
- 7.4 大数据伦理问题产生的原因
- 7.5 大数据伦理问题的治理



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





7.1 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。



7.2 大数据伦理典型案例

- 7.2.1 徐玉玉事件
- 7.2.2 大麦网“撞库”事件
- 7.2.3 大数据“杀熟”
- 7.2.4 隐性偏差问题
- 7.2.5 魏则西事件
- 7.2.6 “信息茧房”问题



7.2.1 徐玉玉事件





7.2.2 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。





7.2.3 大数据“杀熟”





7.2.4 隐性偏差问题



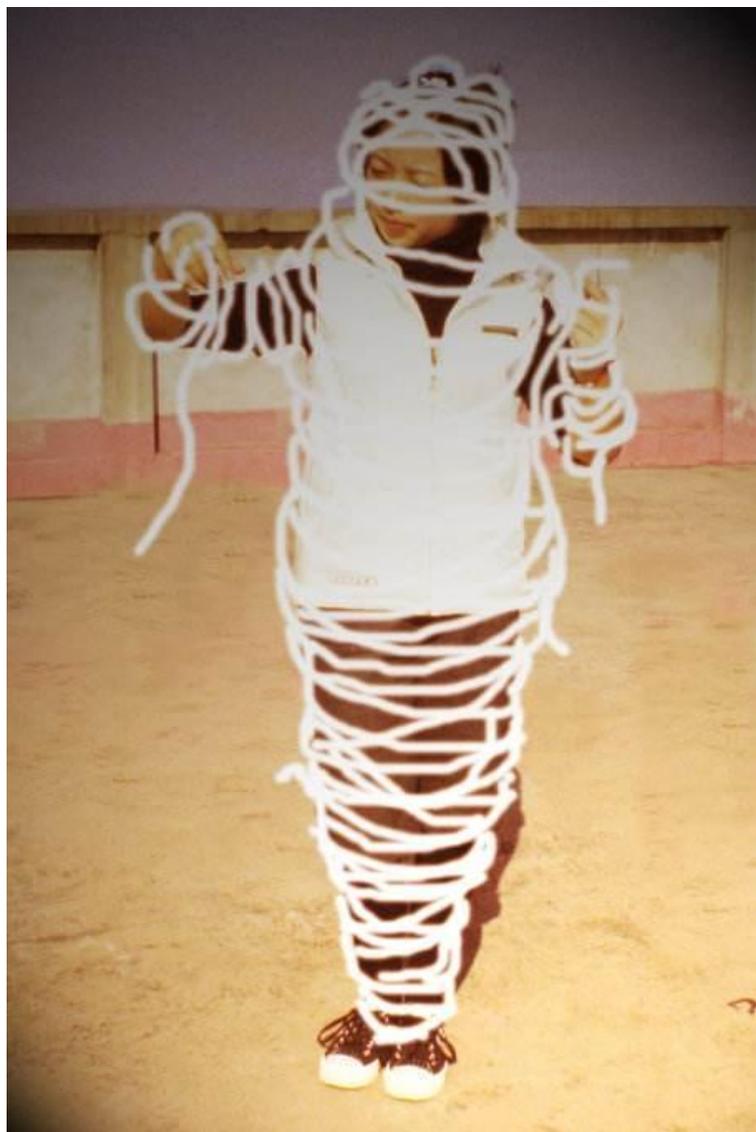


7.2.5 魏则西事件





7.2.6 “信息茧房”问题





7.3 大数据的伦理问题

7.3.1 隐私泄露问题

7.3.2 数据安全问题

7.3.3 数字鸿沟问题

7.3.4 数据独裁问题

7.3.5 数据垄断问题

7.3.6 数据的真实可靠问题

7.3.7 人的主体地位问题



7.3.1 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



7.3.2 数据安全问题

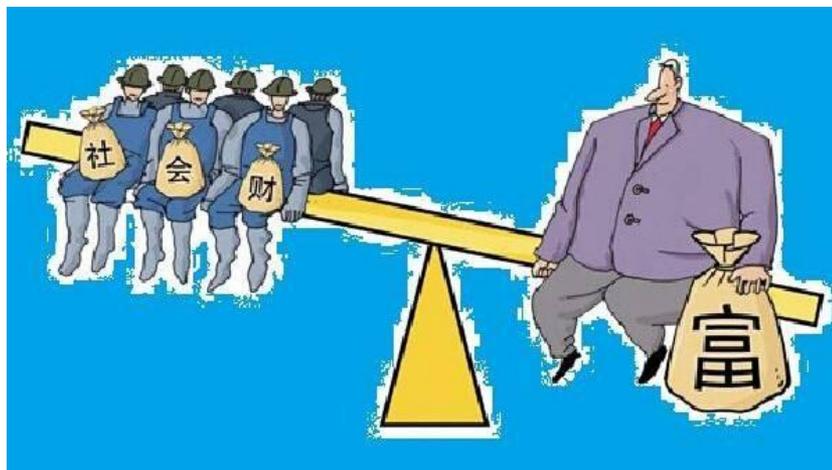
- 一些信息技术本身就存在安全漏洞，可能导致数据泄露、伪造、失真等问题，影响数据安全。
- 智能手机是当今泄漏用户数据的重要途径
- 部分智能家居产品存在安全问题也是不争的事实，给用户的数据安全带来了极大的风险，造成用户隐私的泄露





7.3.3 数字鸿沟问题

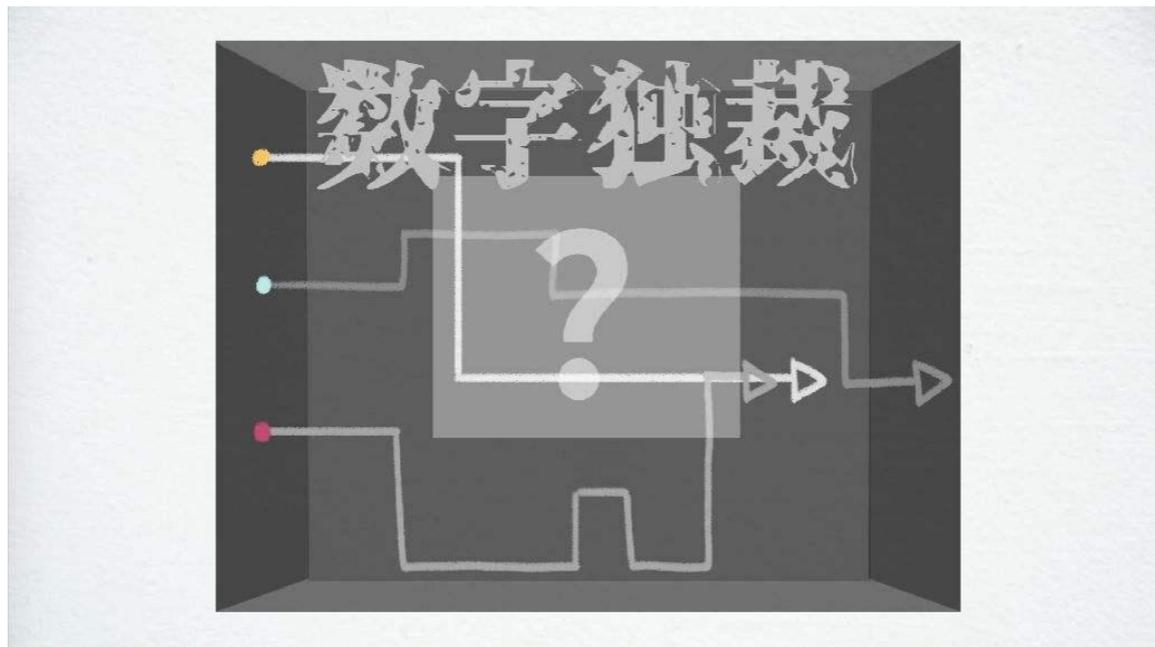
数字鸿沟总是指向信息时代的不公平，尤其在信息基础设施、信息工具以及信息的获取与使用等领域，或者可以认为是信息时代的“马太效应”，即先进技术的成果不能为人公正分享，于是造成“富者越富、穷者越穷”的情况。





7.3.4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。





7.3.5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



7.3.6数据的真实可靠问题





7.3.7 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





7.4 大数据伦理问题产生的原因

- 7.4.1 人类社会价值观的转变
- 7.4.2 数据伦理责任主体不明确
- 7.4.3 相关主体的利益牵涉
- 7.4.4 道德规范的缺失
- 7.4.5 法律体系不健全
- 7.4.6 管理机制不完善
- 7.4.7 技术乌托邦的消极影响
- 7.4.8 大数据技术本身的缺陷



7.4.1 人类社会价值观的转变

- 从总体的发展趋势而言，人类社会的价值观一直朝着更加个性、自由、开放的方向发展。在个人追求自由和社会更加开放的大环境下，人们更加愿意在社会公众层面展示自己的个性化的一面。
- QQ、微博、微信、抖音等新型社交网络媒体的出现，更是给个人的自我展示提供了极大的便利，个人开始热衷于通过智能手机等终端设备向外界展示自己的生产、生活、学习、娱乐等信息。
- 由此，各种社会组织（企业、政府等）能够很容易全方位收集个人生产的海量数据。
- 但是，个人大量分享个性化信息的同时，个人隐私也就随之暴露给社会，从而使自己的身份权、名誉权、自由意志等都有可能受到侵害。



7.4.2 数据伦理责任主体不明确

- 数据在产生、存储、传播和使用过程，都可能存在伦理失范问题。但是，数据权属的不确定性和伦理责任主体的模糊性，都给解决大数据相关的伦理问题增添了难度。
- 由于没有明确数据伦理责任主体，这就意味着数据采集、存储和使用过程中，相关参与方不用对自己的行为负责，都不必遵守相关的伦理规范。



7.4.3 相关主体的利益牵涉

- 虽然技术本身是中性的，但是相关主体的利益牵涉其中，大数据能够给不同的主体带来巨大的价值，这是大数据技术伦理问题形成的重要原因。
- 在利益的驱动下，企业可能有意无意地将法律抛诸脑后，或者巧妙利用法律漏洞，通过各种手段，私自收集公民的个人信息，并向第三方开放共享，甚至肆意买卖公民的个人隐私信息，导致公民的隐私权、知情权受到严重侵害
- 还有一些不法分子通过非法手段肆意窃取公民的个人信息并进行交易，使得网络诈骗等不法行为屡屡得逞



7.4.4 道德规范的缺失

- 大数据时代的开启，引发了一些列新的道德问题，原有的关于“数据观”“隐私权”“网络行为规范”等社会道德规范，无法很好地适应大数据时代的新要求，已经不能有效地引导与制约大数据时代人们的社会价值观与社会行为，而符合大数据时代新要求的社会规范尚未建立，无法形成相应的约束力。
- 由于缺少大数据伦理行业伦理规范，这就给各种组织留下了很大的自由把握空间。



7.4.5 法律体系不健全

- 大数据时代的到来，更是使法律制度建设的滞后性显露无疑。原有的大多数法律大都是为了原子的世界，而不是比特的世界而制定的
- 大数据技术创新导致了与之前迥异的伦理问题，以至于原有的法律法规已无法很好地解决大数据时代所产生的新伦理问题



7.4.6 管理机制不完善

- 大数据技术伦理问题的产生也与社会管理机制建设的缺位密切相关。大数据可以给企业带来巨大的商业价值，企业逐利的本性，使得某些企业在追逐商业利益的同时忘却了基本的技术伦理
- 应该加强在大数据技术的研究、开发和应用阶段建立相应的评估、约束和奖惩机制，有效减少大数据伦理问题的发生。



7.4.7 技术乌托邦的消极影响

- 技术乌托邦认为，人类决定着技术的设计、发展与未来，因此，人类可以按照自身的需求来创新科技，实现科技完全为人类服务的目的。
- 正是在技术乌托邦的影响之下，认为大数据技术是完全正确的，不应加以任何的限制，它所涉及的伦理问题只是小问题，无关乎大数据技术发展。



7.4.8 大数据技术本身的缺陷

- 技术自身也是造成大数据技术伦理问题的一个根源。以数据安全伦理问题为例，日益增长的网络威胁正以指数级速度持续增加，各种网络安全事件层出不穷。
- 据巴黎商学院的统计，**59%**的企业成为了持续恶意攻击的目标；许多大数据企业的IT计划是建立在不够成熟的技术基础上的，很容易出现安全漏洞；有**25%**的组织有明显的安全技能短缺。这些技术的不足，很容易导致数据泄露的危险。



7.5大数据伦理问题的治理

- 7.5.1提高保护个人隐私数据的意识
- 7.5.2加强大数据伦理规约的构建
- 7.5.3努力实现以技术治理大数据
- 7.5.4 完善大数据立法
- 7.5.5 完善大数据伦理管理机制
- 7.5.6引导企业坚持责任与利益并重
- 7.5.7努力弘扬共享精神化解数字鸿沟
- 7.5.8倡导跨行业跨部门合作



7.5.1 提高保护个人隐私数据的意识

- 个人隐私数据信息与公民的利益是紧密相连的，因此，公民要努力提高保护个人数据隐私信息的意识，维护自己的合法权利
- 要加强培养个人的信息权利意识，调整自我的隐私观念，个人作为大数据技术的发展应用过程中的个体参与者，要承担一定的责任，个人要根据道德规则对自身行为可预见的结果负责
- 当遇到侵权行为时，要敢于维护自己的权益，与相关商家及时进行沟通，或者向消费者协会请求帮助，情节严重时，上诉法院，以维护自身的合法权益
- 公民也应当积极履行监督的义务，若发现有企业等机构侵犯隐私信息时，及时向有关部门举报



7.5.2加强大数据伦理规约的构建

首先，大数据应用过程中的个体参与者，需要承担一定的责任。

其次，企业作为大数据应用过程的重要参与者，有责任去保护用户的数据隐私。

最后，政府要履行行政责任。



7.5.3 努力实现以技术治理大数据

- 技术应用过程产生的问题，可以借助技术手段加以解决。加快技术创新，可以有助于规避大数据的各种风险，降低大数据治理成本，提高大数据治理的效率。
- 例如，关于“禁止”的道德规范，在技术上可以通过“防火墙和过滤技术”、“网上监控”、“访问控制”来规避；关于“可控”的道德规范，可以采用“基于数据脱敏技术”、“数字水印技术”、“数据溯源技术”等技术解决。



7.5.4 完善大数据立法

首先，应进一步完善大数据立法。
其次，在法律的基础上制定相关的规章制度，对相关主体的数据采集、存储和使用行为进行规范和约束。
最后，应当通过立法明确公民对个人数据信息的权利。



7.5.5 完善大数据伦理管理机制

首先，加强对专业人士的监管力度和教育。
其次，需要在大数据技术开发阶段建立伦理评估和约束机制。
再次，在大数据技术应用阶段应该建立奖惩机制。
最后，在大数据技术的推广阶段推行安全港模式。



7.5.6 引导企业坚持责任与利益并重

- 追求商业利益，是企业的天然本性，本身无可厚非。但是，当企业的利益和公民个人的利益冲突时，便要进行取舍。因此，大数据企业必须要坚持责任与利益并重的原则，切实承担起属于自己的社会责任，不能唯利是图。
- 企业的责任在于，保护用户数据隐私，避免大数据技术被二次利用。
- 大数据行业伦理规范的形成，也要求企业遵循责任与利益并重的原则



7.5.7 努力弘扬共享精神化解数字鸿沟

- 要相关利益主体都能够公平地参与到数据应用过程，才有可能有效化解数字鸿沟问题。
- 而为了实现大数据利益相关者都能够公平地参与和协作，关键还在于要努力弘扬共享精神。
- 如果无法真正实现大数据的共享，那么必然会导致出现数据割据和数据孤岛现象，就无法从根本上解决数字鸿沟问题



7.5.8 倡导跨行业跨部门合作

- 应该在伦理学家、科学家、社会科学家和技术人员中建立更好的合作，实现跨行业、跨部门协同解决大数据伦理难题
- 在美国国家科学基金会的支持下，大数据应用及研究人员和学术界人士成立了“大数据、伦理和社会委员会”
- 在2016年6月由国家自然科学基金委、复旦大学和清华大学主办召开了“大数据治理与政策”研讨会



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），现为厦门大学计算机科学系助理教授（讲师），曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革委员会副局长。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年度和2017年度厦门大学教学类奖教金获得者，荣获2017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项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金项目(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500万字高价值的研究和教学资料，累计网络访问量超过500万次。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100万次。



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： 《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



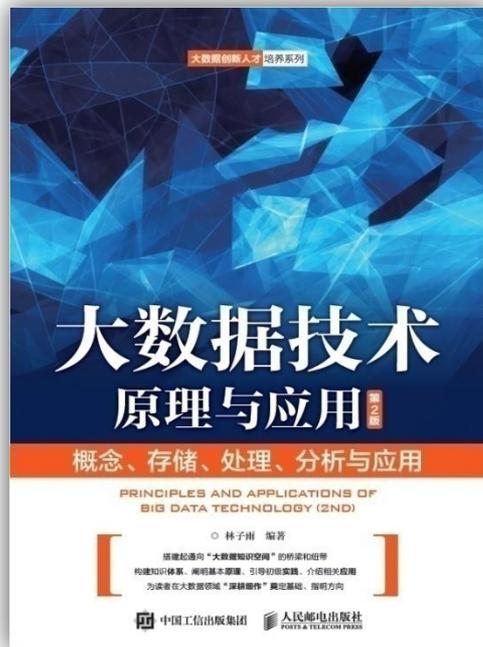
扫一扫访问教材官网

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://dbl原因.xmu.edu.cn/post/bigdata>

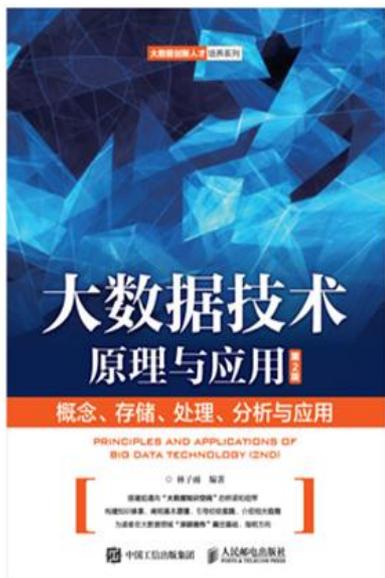




附录D：《大数据基础编程、实验和案例教程》

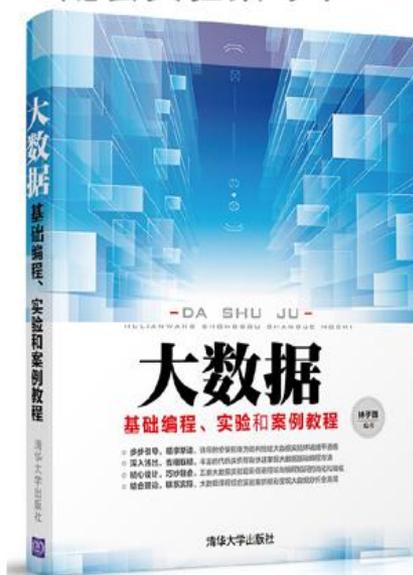
本书是与《大数据技术原理与应用（第2版）》教材配套的唯一指定实验指导书

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

清华大学出版社 ISBN:978-7-302-47209-4 定价：59元



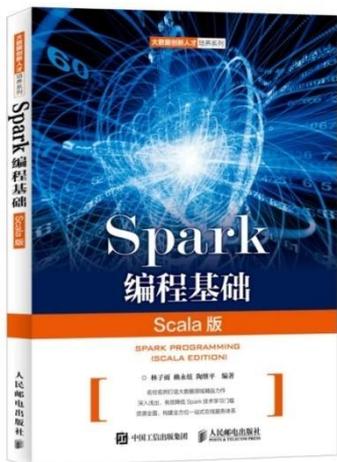
附录E：《Spark编程基础（Scala版）》

《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://dmlab.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F：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://dbllab.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G：高校大数据实训课程系列案例教材

为了更好地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：

《基于协同过滤算法的电影推荐》

《电信用户行为分析》

《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

The background of the slide features several faint, light-blue silhouettes of people. At the top, there are two groups of people standing and holding hands. On the right side, a person is shown in profile, looking towards the center. On the left side, two people are shown in profile, one appearing to be speaking or gesturing towards the other. The overall scene suggests a group of people in a meeting or presentation setting.

Thank You!

Department of Computer Science, Xiamen University, 2019