



《大数据导论（通识课版）》

教材官网: <http://dblab.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5章 大数据安全

(PPT版本号: 2019年秋季学期)



扫一扫访问教材官网

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>





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

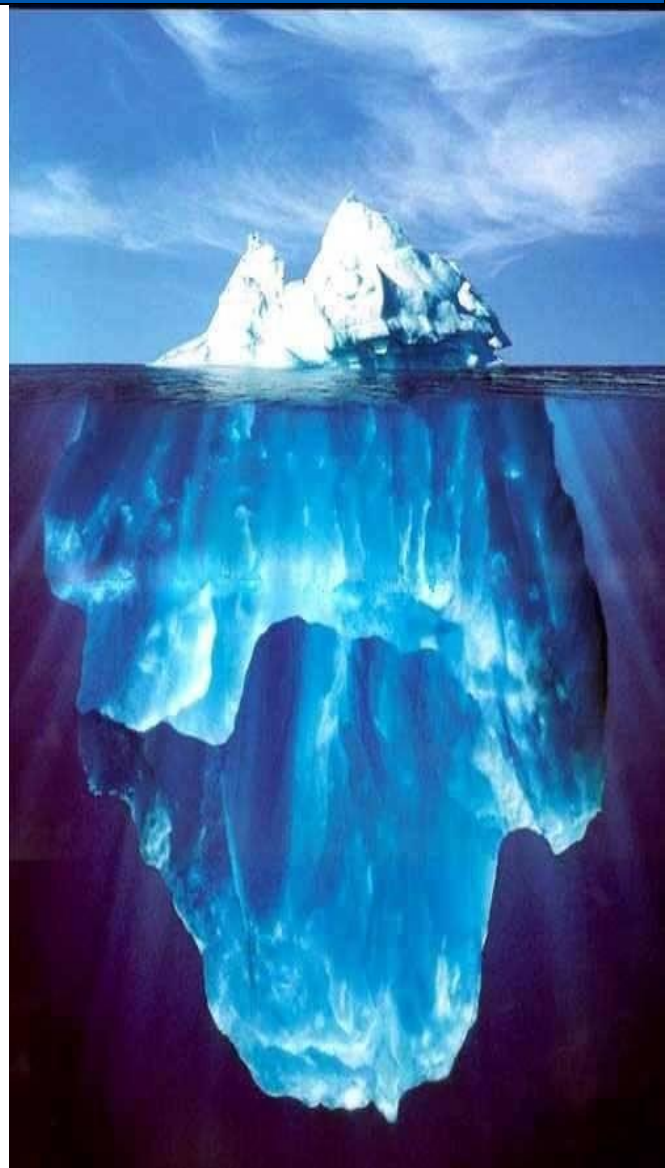
- 5.1 传统数据安全
- 5.2 大数据安全与传统数据安全的不同
- 5.3 大数据安全问题
- 5.4 典型案例
- 5.5 大数据保护的基本原则
- 5.6 大数据时代数据安全与隐私保护的对策
- 5.7 世界各国保护数据安全的实践



高校大数据课程

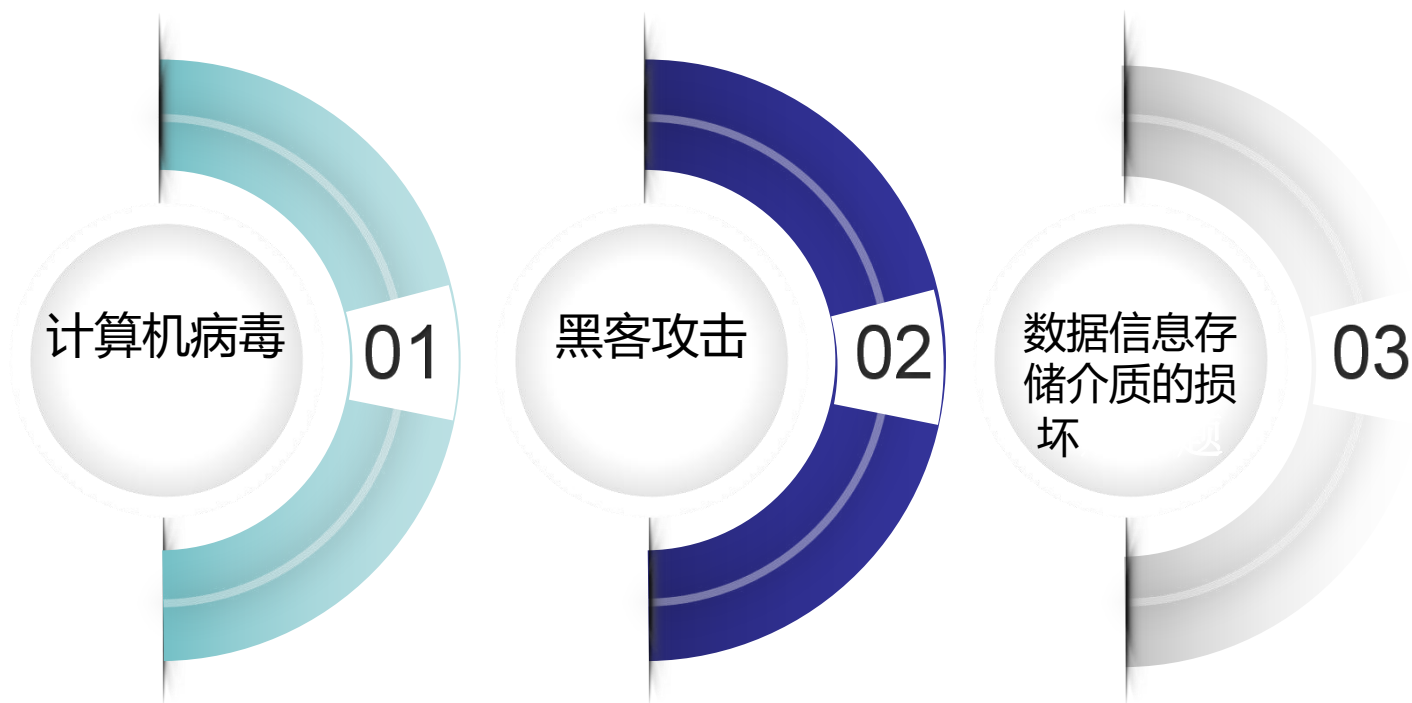
公 共 服 务 平 台

百度搜索厦门大学数据库实验室网站访问平台



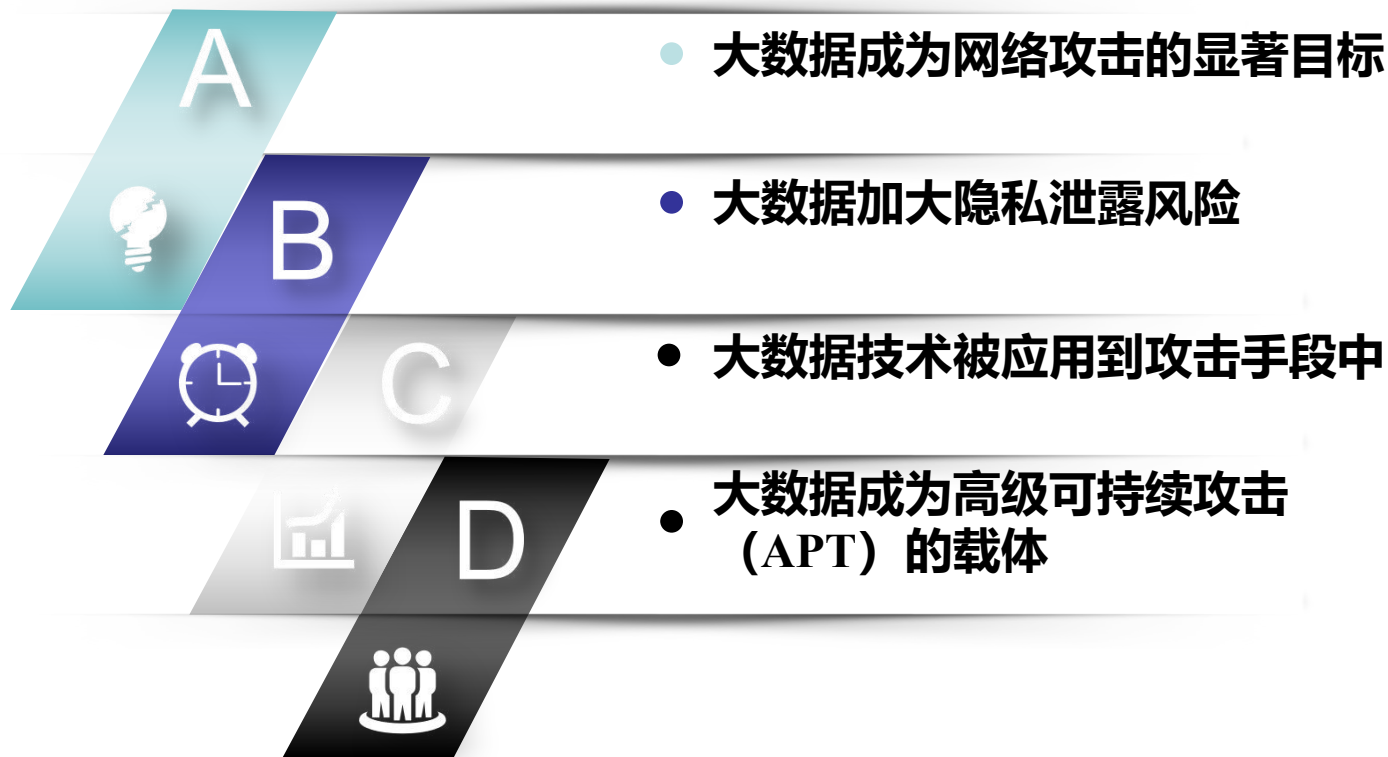


5.1传统数据安全





5.2大数据安全与传统数据安全的不同





5.3大数据安全问题

5.3.1 隐私和个人信息安全问题

5.3.2 国家安全问题



5.3.1 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，都会在各种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。



5.3.1 隐私和个人信息安全问题

- 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。



5.3.1 隐私和个人信息安全问题

- 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**。



5.3.2 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



5.3.2 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



5.4 典型案例

5.4.1 棱镜门事件

5.4.2 维基解密

5.4.3 Facebook数据滥用事件

5.4.4 手机应用软件过度采集个人信息

5.4.5 12306囤票案件

5.4.6 免费WiFi窃取用户信息

5.4.7 收集个人隐私信息的“探针盒子”



5.4.1 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





5.4.2 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





5.4.3 Facebook数据滥用事件

心理学+大数据=颠覆世界



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



5.4.4 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



5.4.5 12306囤票案件

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





5.4.6 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



5.4.7 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





5.5大数据保护的基本原则





5.6大数据时代数据安全与隐私保护的对策



从国家法制层面进行管控

01

02| 从企业端源头进行遏制



提高个人意识，应用安全技术



5.7世界各国保护数据安全的实践

5.7.1 欧盟

5.7.2 美国

5.7.3 英国

5.7.4其他国家

5.7.5中国



5.7.1 欧盟

GDPR是英文“General Data Protection Regulation”的缩写，通常翻译为“通用数据保护条例”，由欧盟于2016年4月推出，并于2018年5月25日正式生效，目的在于遏制个人信息被滥用，保护个人隐私





5.7.2 美国

- 奥巴马政府在**2012年2月**宣布推动《消费者隐私权利法案》的立法程序，这是与大数据最为息息相关的法案，法案中不仅明确且全面地规定了数据的所有权属于用户（即线上线下的使用者），并规定在数据的使用上需对用户有透明性、安全性等更多细节
- **2014年5月**美国发布《大数据：把握机遇，守护价值》白皮书，对美国大数据应用与管理的现状、政策框架和改进建议进行了集中阐述
- **2015年**，美国国防部规定所有为该部门服务的云计算服务提供商须在境内储存数据。**2016年**，美国国家税务局发布规定要求税务信息系统应当位于美国境内



5.7.3 英国

2012年6月英国内阁办公室发布《开放数据白皮书》，推进公共服务数据的开放。英国在《开放数据白皮书》中专门针对个人隐私保护进行规范。

- 一是公共数据开放机构中设立隐私保护专家，确保数据开放过程中及时掌握和普及最新的隐私保护措施，同时还要求各个部门配置隐私保护专家。
- 二是强制要求所有政府部门在处理涉及到个人数据时，都要执行个人隐私影响评估工作制度，制定《个人隐私影响评估手册》。
- 三是要求将开放数据分为大数据和个人数据，制定大数据是政府日常业务过程中收集到的数据，可以对所有人开放，而个人数据仅仅对某条数据所涉及到的个人开放。



5.7.4其他国家

- 德国于2002年通过《联邦数据保护法》，并于2009年进行修订
- 澳大利亚政府于2012年7月发布了《信息安全管理指导方针：整合性信息的管理》，为大数据整合中所涉及到的安全风险提供了最佳管理实践指导
- 印度于2012年批准国家数据共享和开放政策，促进政府拥有的数据和信息得到共享和使用，还拟定一个非共享数据清单，保护国家安全、隐私、机密、商业秘密和知识产权等数据的安全
- 新加坡于2012年公布《个人数据保护法》，旨在防范对国内数据以及源于境外的个人资料的滥用行为
- 日本于2015年4月23日审议《个人信息保护法》和《个人号码法》修正案，以推动并规范大数据的利用



5.7.4其他国家

- 韩国于2013年对个人信息领域的限制做出适当修改，制定了以促进大数据产业发展，并兼顾对个人信息保护的数据共享标准
- 俄罗斯2015年实行新法规定，互联网企业需将收集的俄罗斯公民信息存储在俄罗斯国内
- 巴西于2013年9月出台规定，强制要求所有巴西境内运作的企业，必须将有关巴西人的数据存储在巴西国内，并要求在巴西提供永久网络、电子邮箱及搜索引擎等服务的外国互联网公司，都必须在巴西本土建立数据中心



5.7.5中国

A

加强顶层设计，引领大数据安全发展

B

健全政策法规，防范大数据安全风险

C

构建标准体系，引领大数据规范发展



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年教育部产学研合作育人项目，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

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。

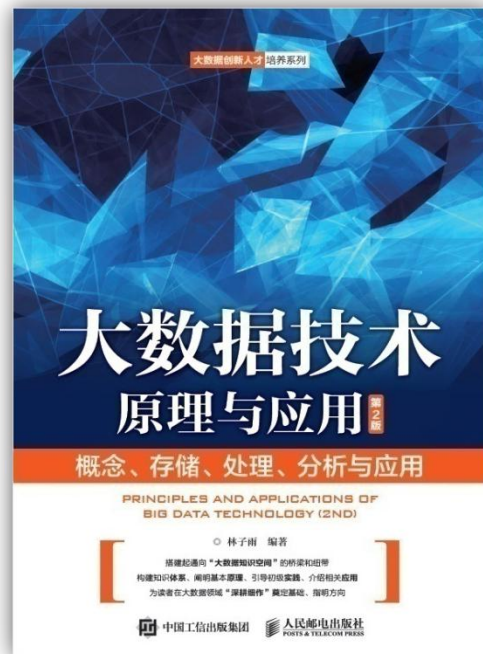
本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

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://dbllab.xmu.edu.cn/post/bigdata>



扫一扫访问教材官网





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://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://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. In the top left, a group of three people is walking. In the top center, a group of about seven people is standing together. In the bottom left, the back of a person's head and shoulders are visible. On the right side, a person is standing with their hand on their chin, appearing to be in deep thought. The overall background is a solid blue color with a subtle gradient.

Thank You!

Department of Computer Science, Xiamen University, 2019