



# 《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6章 大数据思维

(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](mailto: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

- 6.1 传统的思维方式
- 6.2 大数据时代需要新的思维方式
- 6.3 大数据思维方式
- 6.4 运用大数据思维的具体实例



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





# 6.1 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。



## 6.2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”



## 6.3 大数据思维方式

6.3.1 全样而非抽样

6.3.2 效率而非精确

6.3.3 相关而非因果

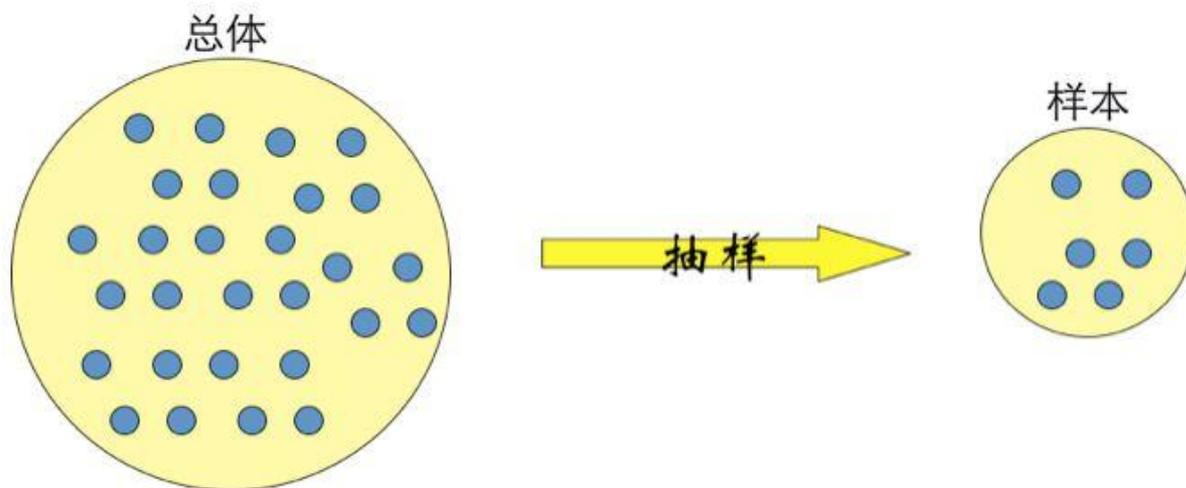
6.3.4 以数据为中心

6.3.5 我为人人，人人为我



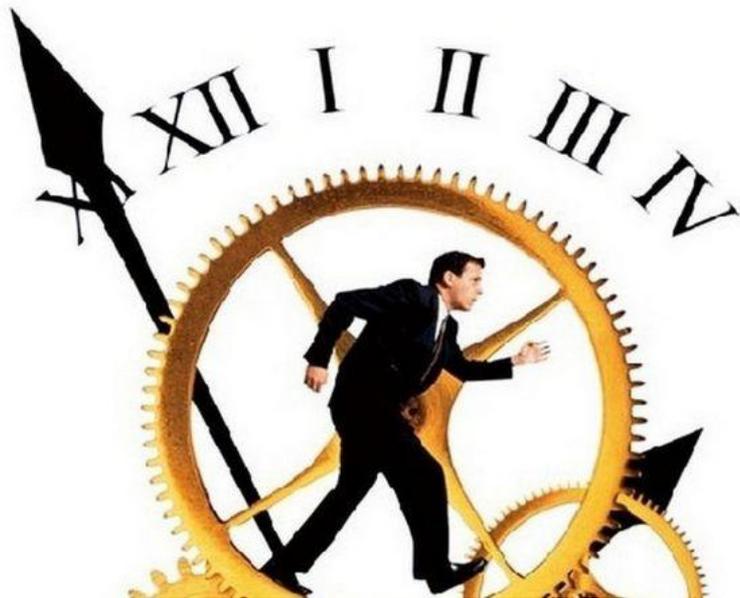
## 6.3.1 全样而非抽样

数据太多，无法保存和分析，统计学采用抽样





## 6.3.2效率而非精确



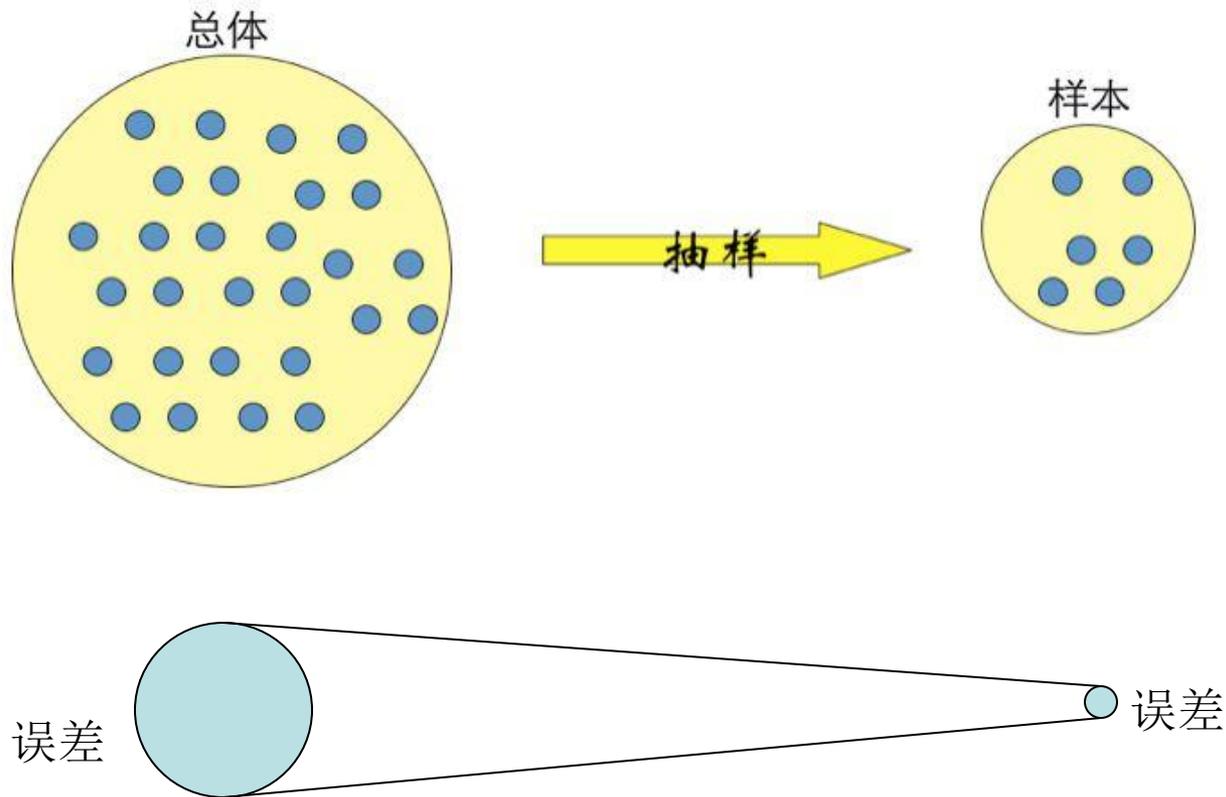
追求效率



不是追求精确



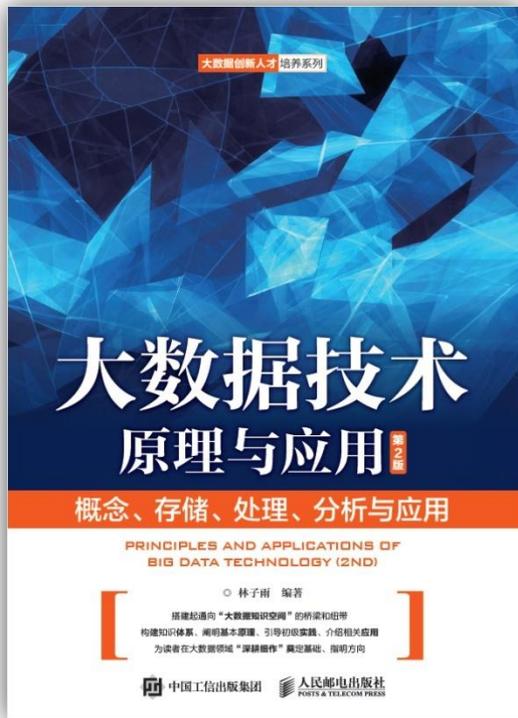
## 6.3.2效率而非精确



抽样计算的结果误差，放到全样上，会被放大

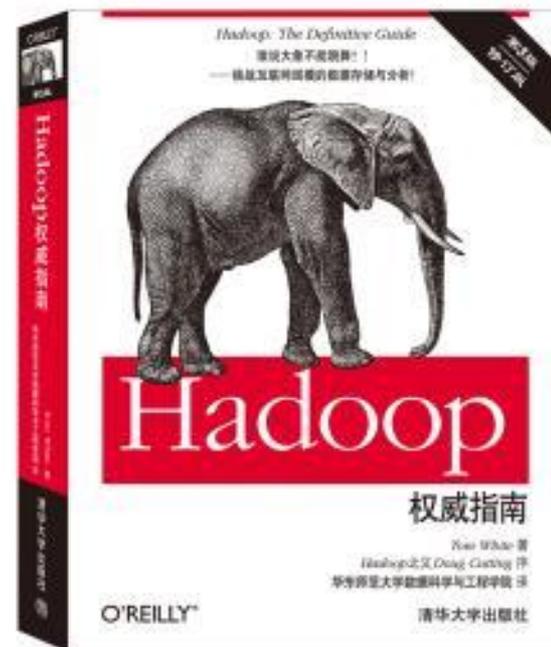
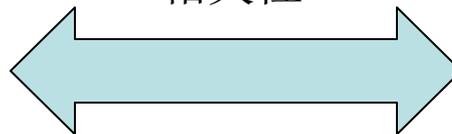


## 6.3.3 相关而非因果



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



## 6.3.4 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。



## 6.3.5 我为人人，人人为我

每个使用导航软件智能手机用户，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。





## 6.4运用大数据思维的具体实例

6.4.1商品比价网站Decide.com

6.4.2啤酒与尿布

6.4.3零售商Target的基于大数据的商品营销

6.4.4吸烟有害身体健康的法律诉讼

6.4.5基于大数据的药品研发

6.4.6基于大数据的谷歌广告

6.4.7 搜索引擎“点击模型”

6.4.8 迪士尼MagicBand手环

6.4.9 谷歌流感趋势预测

6.4.10大数据的简单算法比小数据的复杂算法更有效

6.4.11谷歌翻译



## 6.4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



## 6.4.1 商品比价网站Decide.com

**decide.** + **ebay**

Decide is now part of the eBay family. Thank you to all our supporters who made it possible.





## 6.4.2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





## 6.4.3零售商Target的基于大数据的商品营销





# 6.4.4吸烟有害身体健康的法律诉讼





## 6.4.5 基于大数据的药品研发





## 6.4.6 基于大数据的谷歌广告



Google AdWords



## 6.4.7 搜索引擎“点击模型”

必应 bing™

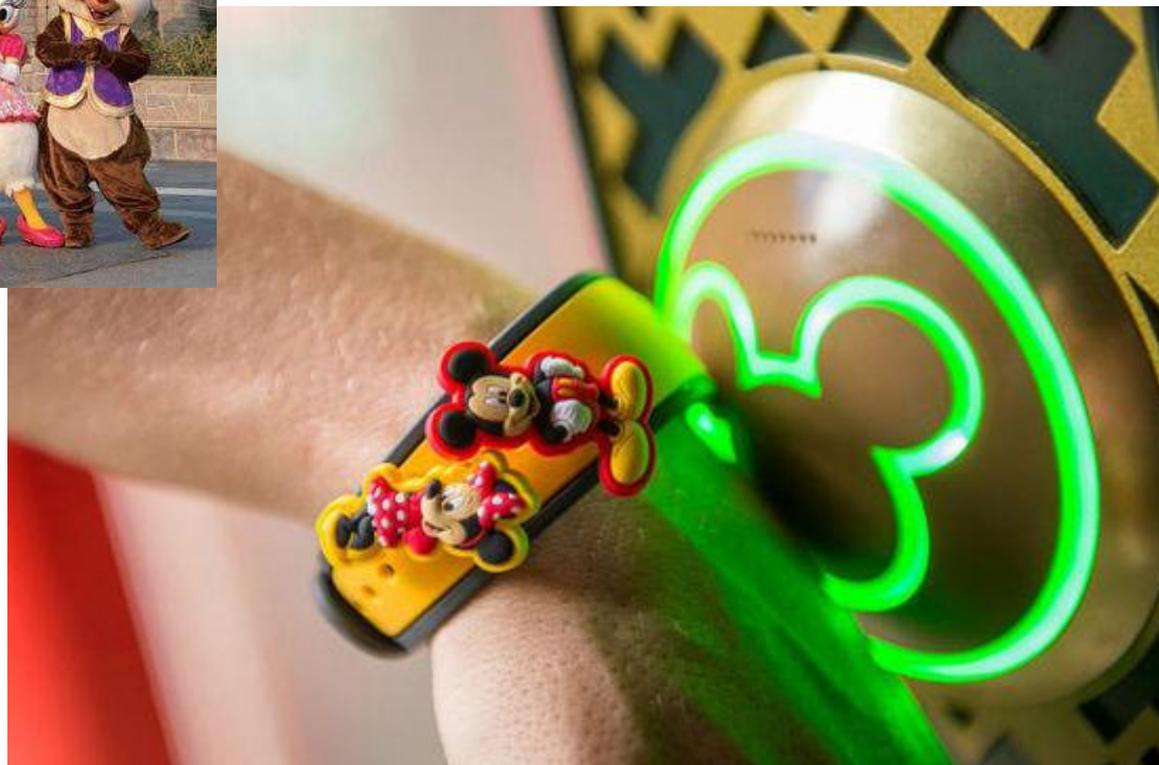
微软

Google

谷歌



## 6.4.8 迪士尼MagicBand手环





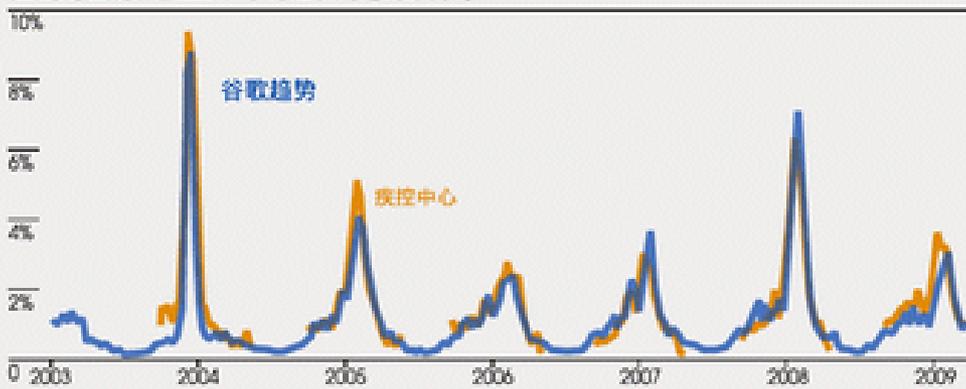
## 6.4.9 谷歌流感趋势预测



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

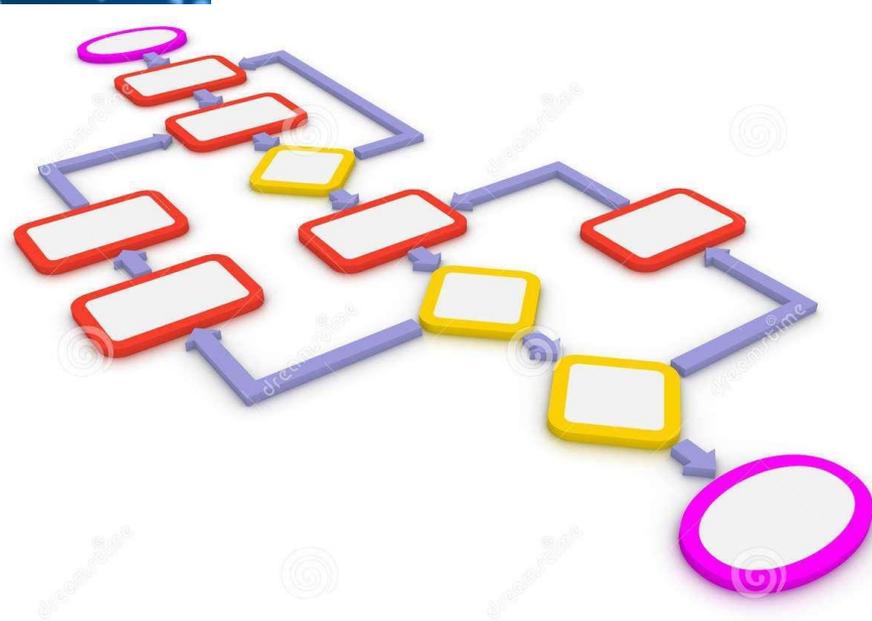
图:美国某地区历年来的流感发病率



数据来源: 谷歌趋势, 美国各地疾病预防控制中心



## 6.4.10 大数据的简单算法比小数据的复杂算法更有效





## 6.4.11 谷歌翻译



Google  
翻译



# 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



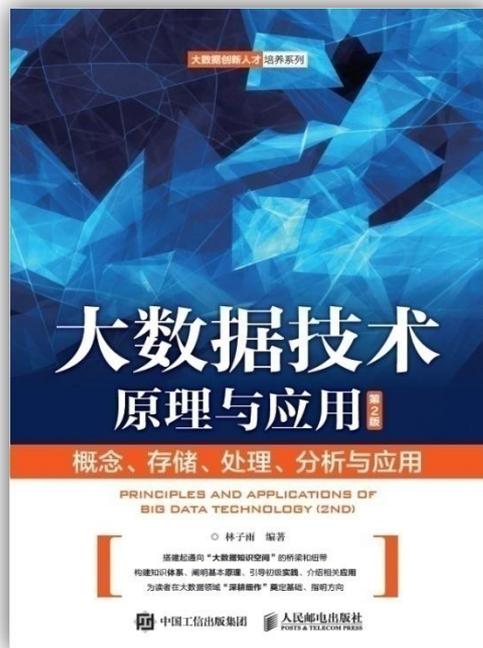
扫一扫访问教材官网

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



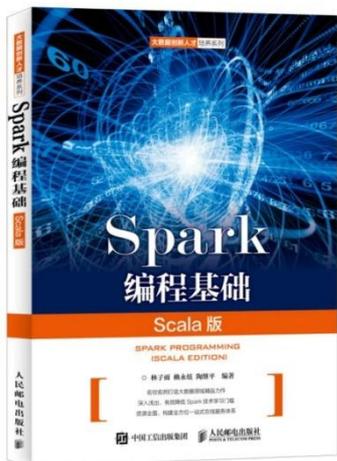
# 附录E：《Spark编程基础（Scala版）》

## 《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**