



# 第4届全国高校大数据与人工智能教学研讨会

2021.05.14-05.15 中国·厦门

主办单位：教育部高等学校计算机类专业教学指导委员会

承办单位： 厦门大学  
 山东大学  
 华东师范大学  
 清华大学  
 北京大学  
 集美大学  
 湖南师范大学  
 人民邮电出版社

协办单位： 头歌 EduCoder  
 TRANSWARP 星环科技  
 山东省计算中心  
 荔峰科技 LIFENG  
 探知图灵 TOUZHONG TULING  
 美林数据 MERITDATA

第4届全国高校大数据与人工智能教学研讨会 (BDTS2021)



# 厦门大学大数据课程与教材体系

## 建设经验分享

厦门大学 林子雨 博士/副教授

ziyulin@xmu.edu.cn

2021年5月15日 厦门



# 内容提要

- 高校大数据课程公共服务平台
- 厦门大学建设的大数据课程体系
- 各门大数据课程介绍





# 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

为高校提供大数据教学一站式服务

- 大数据专业建设方案
- 系列课程教材
- 讲义PPT、习题、实验、案例
- 教师备课指南
- 学生学习指南
- 授课视频
- 教师培训交流
- 大数据教学研讨会



### 全国高校大数据教学知名品牌

平台构建在厦门大学数据库实验室官网上，在线资源全部**免费**开放

平台访问地址：<http://dblab.xmu.edu.cn/post/8197/>





# 高校大数据课程公共服务平台

建设周期  
八年 (2013-2021)



投入资金  
100万+

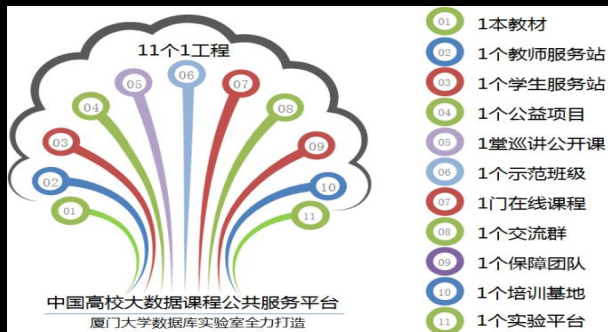




# 高校大数据课程公共服务平台

## 打造11大工程

## 平台累计访问量 超过1200万次





# 高校大数据课程公共服务平台



学习路线图涵盖了高校大数据课程公共服务平台的大量免费大数据学习资源  
 大数据学习路线图访问地址：<http://dbl原因.xmu.edu.cn/post/10164/>





# 高校大数据课程公共服务平台

已经在厦门大学累计举办3届全国高校大数据教学研讨会







# 高校大数据课程公共服务平台

已经累计举办16期大数据师资培训班，为全国400余所高校培养600余名师资





# 高校大数据课程公共服务平台

大数据课程教师交流群 (QQ群号: 461510122)  
促进大数据课程教师之间的沟通和交流

截至目前, 已经有来自全国800多所高校的1800多名教师加入交流群



厦门大学、福建师范大学、厦门理工学院、同济大学、浙江财经大学、安徽大学、大连海洋大学、中北大学、河海大学、中山大学、浙江大学、中国农业大学、重庆邮电大学、华中师范大学、武汉理工大学、贵州师范大学、江西财经大学、山西大学、河北经贸大学、东北大学、山东农业大学、海南大学、中国地质大学、武汉大学、中国传媒大学、湖南大学、中国科技大学.....





# 内容提要

- 高校大数据课程公共服务平台
- 厦门大学建设的大数据课程体系
- 各门大数据课程介绍





# 厦门大学建设的大数据课程体系

面向非计算机专业的大数据课程

大数据  
通识  
导论

实训

大数据实习实训案例

进阶

Spark编程基础  
Flink编程基础

入门

大数据技术原理与应用  
数据采集与预处理  
Python程序设计

导论

大数据导论

面向理工科的大数据课程





# 厦门大学建设的大数据系列教材

【1】林子雨.大数据导论.人民邮电出版社,2020年9月第1版 (用于开设大数据专业的导论课)。

【2】林子雨.大数据导论 (通识课版) .高等教育出版社, 2020年2月. (用于开设全校公共选修课)

【3】林子雨,赵江声,陶继平.Python基础教程.人民邮电出版社,预计2021年9月出版。

【4】林子雨.数据采集与预处理.人民邮电出版社,预计2021年9月出版。

【5】林子雨.大数据技术原理与应用 (第3版) .人民邮电出版社, 2021年1月. (用于开设入门级大数据专业课)

【6】林子雨.大数据基础编程、实验和案例教程 (第2版) .清华大学出版社, 2020年10月. (用于开设入门级大数据专业课)

教材官网提供了全部配套资源 (讲义PPT、习题、实验、视频等) 的下载  
<http://dblab.xmu.edu.cn/post/bigdatabook/>





# 厦门大学建设的大数据系列教材

【7】林子雨, 赖永炫, 陶继平.Spark编程基础 (Scala版).人民邮电出版社, 2018年8月. (用于开设进阶级大数据专业课)

【8】林子雨, 郑海山, 赖永炫.Spark编程基础 (Python版).人民邮电出版社, 2020年3月. (用于开设进阶级大数据专业课)

【9】林子雨,陶继平.Flink编程基础 (Scala版).清华大学出版社, 预计2021年7月出版.

【10】林子雨.大数据实训案例——电影推荐系统.人民邮电出版社, 2019年4月. (用于开设大数据实训课程)

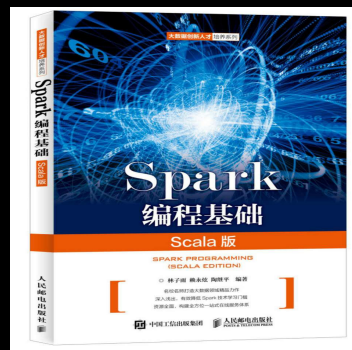
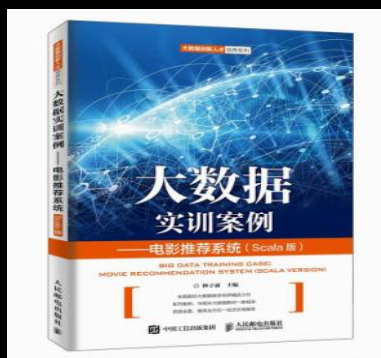
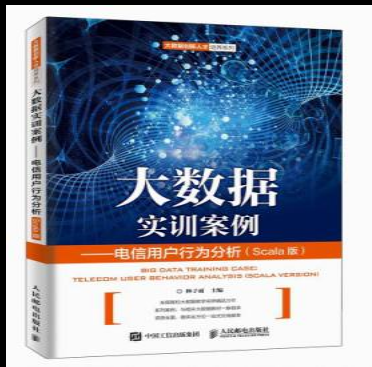
【11】林子雨.大数据实训案例——电信用户行为分析.人民邮电出版社, 2019年4月. (用于开设大数据实训课程)

教材官网提供了全部配套资源 (讲义PPT、习题、实验、视频等) 的下载  
<http://dbl原因.xmu.edu.cn/post/bigdatabook/>





# 厦门大学建设的大数据系列教材



大数据教材

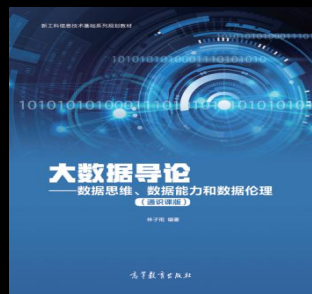
大数据技术  
原理与应用

概念·存储·处理·分析与应用

1+1黄金组合  
厦门大学林子雨编著

大数据  
基础理论·实验化案例教程

配套实验指导书





# 内容提要

- 高校大数据课程公共服务平台
- 厦门大学建设的大数据课程体系
- 各门大数据课程介绍







# 通识导论 《大数据导论》

## 培养目标

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





# 通识导论 《大数据导论》

面向非计算机专业的导论课



- 第1章 大数据概述
- 第2章 大数据与云计算、物联网、人工智能
- 第3章 大数据技术
- 第4章 大数据应用
- 第5章 大数据安全
- 第6章 大数据思维
- 第7章 大数据伦理
- 第8章 数据共享
- 第9章 数据开放
- 第10章 大数据交易
- 第11章 大数据治理

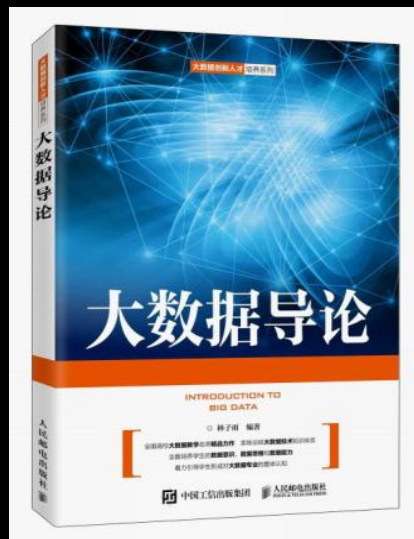
教材官网: <http://dblab.xmu.edu.cn/post/bigdataintroduction/>





# 专业导论课 《大数据导论》

面向大数据专业的导论课



教材官网：<http://dbllab.xmu.edu.cn/post/bigdata-introduction/>





# 专业导论课 《大数据导论》

## 第1章 大数据概述

数据的概念、大数据时代到来的背景、大数据的发展历程、世界各国的大数据发展战略、大数据的概念与影响、大数据的应用以及大数据产业

## 第2章 大数据与其他新兴技术的关系

物联网的概念和应用、云计算的概念和应用、大数据和云计算以及物联网的相互关系、人工智能的概念与关键技术、大数据与人工智能的关系、区块链的概念与技术原理、大数据与区块链的关系

## 第3章 大数据基础知识

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

## 第4章 大数据应用

大数据在各行各业的主要应用





# 专业导论课 《大数据导论》

## 第5章 数据采集与预处理

数据采集、数据清洗、数据转换、数据脱敏

## 第6章 数据存储与管理

传统的数据存储与管理技术、大数据时代的数据存储与管理技术、大数据处理架构Hadoop、分布式文件系统HDFS、NoSQL数据库、云数据库、分布式数据库HBase

## 第7章 数据处理与分析

数据处理与分析的概念、机器学习和数据挖掘算法、大数据处理与分析技术、大数据处理与分析代表性产品

## 第8章 数据可视化

可视化概述、可视化图表、可视化工具、可视化典型案例

## 第9章 大数据分析综合案例

案例任务、系统设计、技术选择、系统实现、案例所需知识和技能





# Python程序设计

第1章 Python语言概述

第2章 基础语法知识

第3章 程序控制结构

第4章 序列

第5章 字符串

第6章 函数

第7章 面向对象程序设计

第8章 模块

第9章 异常处理

第10章 基于文件的持久化

第11章 基于数据库的持久化

第12章 图形用户界面编程

第13章 正则表达式

第14章 网络爬虫

课程教材《Python基础教程》将于2021年9月由人民邮电出版社出版





# 数据采集与预处理

第1章 概述

第2章 大数据实验环境搭建

第3章 网络数据采集

第4章 分布式消息系统Kafka

第5章 日志采集系统Flume

第6章 数据仓库中的数据集成

第7章 ETL工具Kettle

第8章 使用pandas进行数据清洗

课程教材《数据采集与预处理》将于2021年9月由人民邮电出版社出版





# 入门课程《大数据技术原理与应用》

课程定位

01

入门级课程

构建知识体系、阐明基本原理  
引导初级实践、了解相关应用

授课对象：

本科生（计算机、软件工程、数据科学与大数据技术）

知识储备：编程、操作系统、数据库







# 入门课程《大数据技术原理与应用》

课程教材

02

教材定位为大数据技术入门教材  
为读者搭建起通向“大数据知识空间”的桥梁和纽带



人民邮电出版社  
2017年度好书  
2020年畅销书

- 构建知识体系
- 阐明基本原理
- 引导初级实践
- 了解相关应用

• 为读者在大数据领域“深耕细作”奠定基础、指明方向  
• Hadoop、HDFS、HBase、NoSQL、云数据库、MapReduce、流计算、图计算、数据可视化、Spark

教材官网：<http://dbl原因.xmu.edu.cn/post/bigdata3/>





# 入门课程《大数据技术原理与应用》

课程教材

02

大数据教材



大数据技术  
原理与应用

概念、存储、处理、分析与应用

PRINCIPLES AND APPLICATIONS OF  
BIG DATA TECHNOLOGY (2ND)

林子雨 编著

清华大学出版社

1+1黄金组合

厦门大学林子雨编著

配套实验指导书



大数据  
基础编程、实验和案例教程

DA SHU JU

基础编程、实验和案例教程

清华大学出版社

全力打造大数据精品教材

实验指导书官网: <http://dblab.xmu.edu.cn/post/bigdatappractice2/>





# 入门课程《大数据技术原理与应用》

## 03

### 实验内容

- 全套机房上机实验指南，包含题目和答案
- 用于入门级大数据课程的上机实验课
- 每个实验都需要连续4节上机课来完成
- 每个实验的设计，都充分考虑了学生的基础和能力的，力求学生能够在连续4节课的上机时间内，顺利完成课程实验，提交实验报告

实验一：熟悉常用的Linux操作和Hadoop操作

实验二：熟悉常用的HDFS操作

实验三：熟悉常用的HBase操作

实验四：NoSQL和关系数据库的操作比较

实验五：MapReduce初级编程实践

免费在线访问地址：<http://dblab.xmu.edu.cn/post/6131/>





# 入门课程《大数据技术原理与应用》

03

实验内容

## 大数据课程实验案例《网站用户购物行为分析》

- 采用2000万条用户购物数据集
  - 案例涉及数据预处理、存储、查询和可视化分析等数据处理全流程所涉及的各种典型操作
  - 涵盖Linux、MySQL、Hadoop、HBase、Hive、Sqoop、R、Eclipse等系统和软件的安装和使用方法
  - 案例适合高校（高职）大数据教学，可以作为学生学习大数据课程后的综合实践案例
- 免费访问地址：<http://dblab.xmu.edu.cn/post/7499/>





# 入门课程《大数据技术原理与应用》

04

课程视频

2018年国家精品在线开放课程

2020年国家级线上一流本科课程

## 林子雨主讲《大数据技术原理与应用》授课视频

中国大学MOOC 课程 名校 学·问 学校云 考研 新 客户端 搜索感兴趣的课程 登录 | 注册

**厦门大学**  
XIAMEN UNIVERSITY

厦门大学林子雨老师主讲  
《大数据技术原理与应用》  
2017年11月6日正式开课  
欢迎进入中国大学MOOC学习

大数据技术原理与应用

入门级大数据精品课程，适合初学者，完备的课程在线服务体系，可以帮助初学者实现“零基础”学习大数据。课程指导思想是“构建知识体系、阐明基本原理、引导初级实践、了解相关应用”。课程由国内高校知名大数据教师厦门大学林子雨老师主讲。配套的《大数据技术原理与应用》教材已经被众多高校采用。

大数据技术原理与应用  
BIGDATA TECHNOLOGY APPLICATION  
打开大数据之门，遨游大数据世界

授课视频观看地址：<http://www.icourse163.org/course/XMU-1002335004>





# 进阶课程《Spark编程基础》

课程定位

01

课程教材

02

课时安排

03

04

交叉知识

05

实验内容

06

课程视频





# 进阶课程《Spark编程基础》

课程定位

01

大数据技术进阶学习课程

授课对象：本科生、研究生（计算机相关专业）

知识储备：Java编程、数据库、操作系统、Hadoop

先修课程：入门级大数据课程，比如：大数据技术原理与应用





# 进阶课程 《Spark编程基础》

课程教材

02

选择教材时，必须首先确定编程语言

Spark支持多种编程语言：Scala、Java、Python、R

首选语言是Scala，或者采用Python







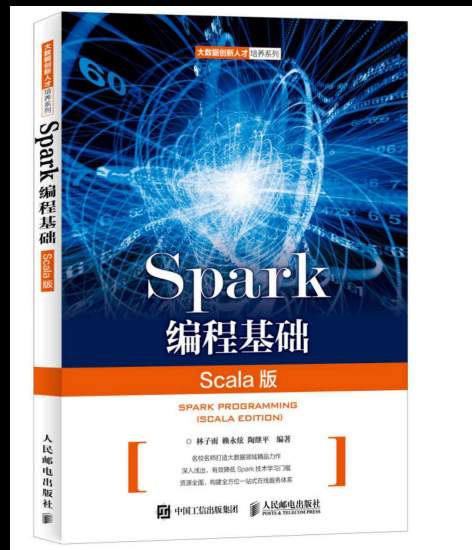
# 进阶课程 《Spark编程基础》

课程教材

02

## 《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/>

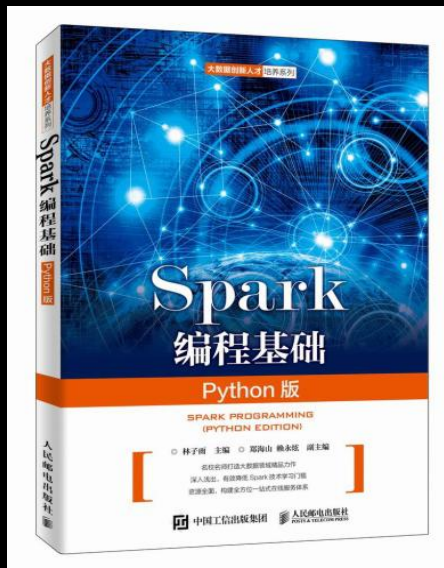




# 进阶课程 《Spark编程基础》

课程教材

02



## 《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

2020年3月由人民邮电出版社出版发行  
教材官网：

<http://dbllab.xmu.edu.cn/post/spark-python/>





# 进阶课程《Spark编程基础》

课时安排

03

理论32学时，可另外增加实验上机学时

章（或节）	主要内容	学时安排
第1章 大数据技术概述	大数据的基本概念、关键技术和代表性软件	2
第2章 Scala语言基础	介绍Scala语言基础语法	6
第3章 Spark设计与运行原理	Spark概述、Spark生态系统、Spark运行架构、Spark的部署和应用方式	3
第4章 Spark安装和使用方法	安装Spark、在Spark Shell中运行代码、编写Spark独立应用程序、第一个Spark应用程序：WordCount、使用开发工具编写Spark应用程序、Spark集群环境搭建、在集群上运行Spark应用程序	3
第5章 Spark编程基础	RDD编程、键值对RDD、数据读写（文件数据读写、读写HBase数据）	4
第6章 Spark SQL	Spark SQL简介、DataFrame、读写Parquet、通过JDBC连接数据库、连接Hive读写数据	2
第7章 Spark Streaming	流计算简介、Spark Streaming简介、DStream操作	4
第8章 Spark MLlib	Spark MLlib简介、机器学习工作流、特征抽取、转化和选择、分类与回归、聚类算法、推荐算法	4
综合案例	Spark综合案例	2
合计		32





# 进阶课程 《Spark编程基础》

05

实验内容

教材配套实验

- 实验1-Linux系统的安装和常用命令
- 实验2-Scala编程初级实践
- 实验3-Spark和Hadoop的安装
- 实验4-RDD编程初级实践
- 实验5-Spark SQL编程初级实践
- 实验6-Spark Streaming编程初级实践
- 实验7-Spark机器学习库MLlib编程实践





# 进阶课程《Spark编程基础》

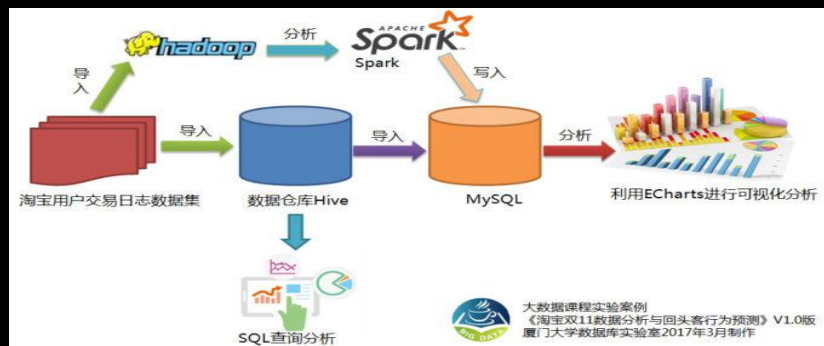
05

实验内容

## Spark课程综合实验案例：淘宝双11数据分析与预测

本案例涉及数据预处理、存储、查询和可视化分析等数据处理全流程所涉及的各种典型操作，涵盖Linux、MySQL、Hadoop、Hive、Sqoop、Eclipse、ECharts、Spark等系统和软件的安装和使用方法

案例访问地址：<http://dblaboratory.xmu.edu.cn/post/8116/>





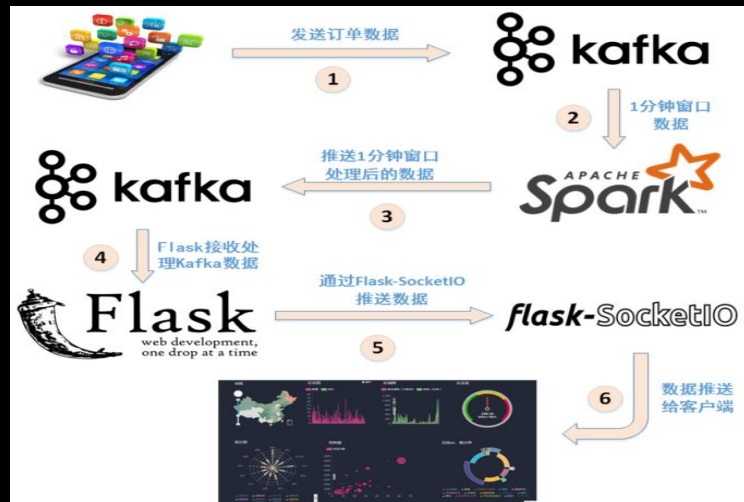
# 进阶课程《Spark编程基础》

05

实验内容

## 大数据课程实验案例：Spark+Kafka构建实时分析Dashboard案例

由厦门大学数据库实验室团队开发，旨在满足全国高校大数据教学对实验案例的迫切需求。本案例涉及数据预处理、消息队列发送和接收消息、数据实时处理、数据实时推送和实时展示等数据处理全流程所涉及的各种典型操作，涵盖Linux、Spark、Kafka、Flask、Flask-SocketIO、Highcharts.js、sockert.io.js、PyCharm等系统和软件的安装和使用方法。案例适合高校（高职）大数据教学，可以作为学生学习大数据课程后的综合实践案例。



免费在线实验案例主页：<http://dbl原因.xmu.edu.cn/post/8274/>





# 进阶课程《Spark编程基础》

06

课程视频

2019年3月11日在中国大学MOOC正式上线

<http://www.icourse163.org/course/XMU-1205811805>

2020年福建省线上一流本科课程

Java

Scala

RDD

REPL

## Spark编程基础

林子雨 | 厦门大学计算机科学系

欢迎进入“中国大学MOOC”平台观看课程视频





# 进阶课程 《Flink编程基础》

第1章 大数据技术概述

第2章 Scala语言基础

第3章 Flink的设计与运行原理

第4章 Flink环境搭建和使用方法

第5章 DataStream API

第6章 DataSet API

第7章 Table API&SQL

第8章 FlinkCEP

教材预计2021年7月由清华大学出版社出版发行







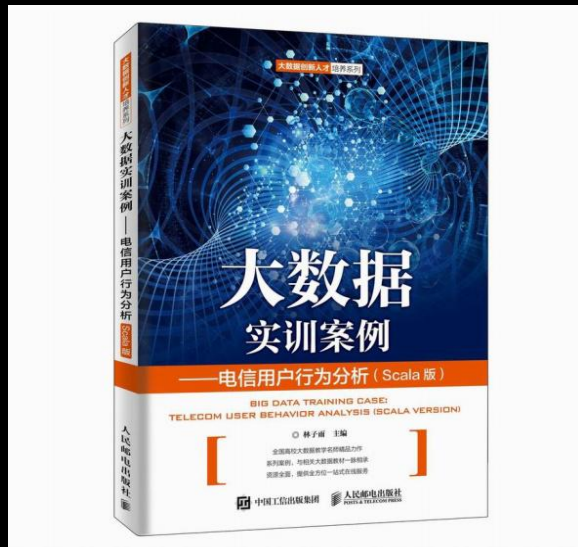
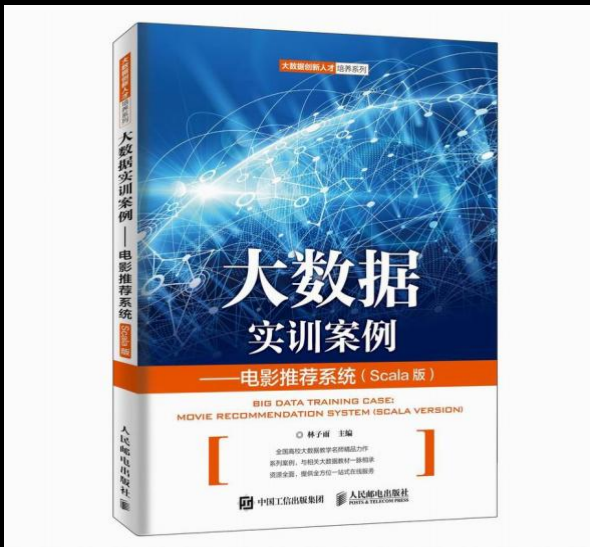
# 大数据实训课程





# 五、课程教材

林子雨编著《高校大数据实训课程系列案例》





# 总结：大数据课程建设模式

- 以大量教学实践推动课程和教材建设
- 以平台思维促进教学资源汇聚和共享
- 以迭代方法不断优化升级教学内容
- 自我造血为课程建设提供稳定资金保障





# THANKS

敬请指正



@林子雨





# 附录：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>

中国高校首个“数字教师”的提出者和建设者

中国高校首个大数据课程公共服务平台建设者



扫一扫访问个人主页

林子雨，男，博士（毕业于北京大学），厦门大学计算机科学系副教授，厦门大学信息学院实验教学中心主任。2013年度、2017年度和2020年度厦门大学教学类奖教金获得者。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，厦门大学数据库实验室负责人，数据中国“百校工程”教育部专家组成员。负责的“大数据技术原理与应用”课程获评“2018年国家精品在线开放课程”和“2020年国家级线上一流本科课程”。负责的“Spark编程基础”课程获评“2020年福建省线上一流本科课程”。国内高校首个“数字教师”的提出者和建设者，编著出版了国内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》，成为国内众多高校开课教材，同时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免费提供全方位、一站式服务，平台累计访问量超过1000万次，成为国内高校大数据教学知名品牌，并获得“2018年福建省教学成果二等奖”和“2018年厦门大学教学成果特等奖”。

