

# 《大数据技术原理与应用》

<https://www.icourse163.org/course/XMU-1002335004>

中国大学MOOC 2018年春季学期

## 课程介绍

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>





中国大学MOOC《大数据技术原理与应用》课程地址：  
<https://www.icourse163.org/course/XMU-1002335004>

 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



# 廈門大學

XIAMEN UNIVERSITY

## 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

播放视频简介



# 课程重要资料



## 大数据软件安装和编程指南

读者在学习《大数据技术原理与应用》MOOC课程时，需要进行安装Linux系统和各种大数据软件，并开展基础编程实践。这个实践过程，如果没有配套的指南，将会耗费读者大量的时间，而且在实践过程中的大量障碍，网络上都没有现成的答案，会给读者带来很大的挫折感，感觉学习大数据是一件“痛苦的事情”。为了帮助读者实现“零基础”学习大数据并顺利完成实验环境搭建和开展基础编程，课程团队建设与本课程配套的《大数据软件安装和基础编程实践指南》，读者即使没有Linux系统知识，没有学习过任何大数据软件的使用，也可以在自己熟悉的Windows操作系统上顺利安装Linux虚拟机和各种大数据软件，顺利完成基础的编程实践，让学习入门级大数据技术变得“相对容易”。课程团队在过去5年时间里建设的在线免费资源，大大降低了大数据技术学习门槛，已经较好地帮助很多大数据初学者顺利完成了大数据基础实践，获得了读者的好评，目前，厦门大学数据库实验室网站上的这些在线免费大数据教学资源每年访问量超过100万次，在国内高校形成了广泛的影响力。

**重要提示：**读者在学习《大数据技术原理与应用》MOOC课程时，在中国大学MOOC课程的栏目中，有一个名称为“大数据软件安装和编程指南”的子栏目，进入这个栏目，可以帮助读者顺利完成大数据上机环境的安装和开展基础编程实践。在观看每个章节的MOOC视频时，可以充分利用该栏目辅助自己完成上机实验操作。



# 欢迎访问教材官网获取教学资源

## 《大数据技术原理与应用——大数据概念、存储、处理、分析与应用》

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

厦门大学 林子雨编著, 人民邮电出版社, 2017年1月第2版  
ISBN:978-7-115-44330-4

- 国内高校第一本系统介绍大数据知识专业教材
- 京东、当当等各大网店畅销书籍
- 大数据入门教材精品
- 国内多所高校采用本教材开课
- 配套目前国内高校最完备的课程公共服务平台
- 福建省精品在线开放课程





# 提纲

- 1.课程特色
- 2.教材介绍
- 2.内容提要
- 3.篇章安排
- 4.班级介绍

附录A：主讲教师林子雨简介

附录B：大数据学习路线图

附录C：《大数据技术原理与应用》教材

附录D：《大数据基础编程、实验和案例教程》

附录E：《Spark编程基础》

附录F：高校大数据课程公共服务平台

本PPT是如下教材的配套讲义：

《大数据技术原理与应用——概念、存储、处理、分析与应用》

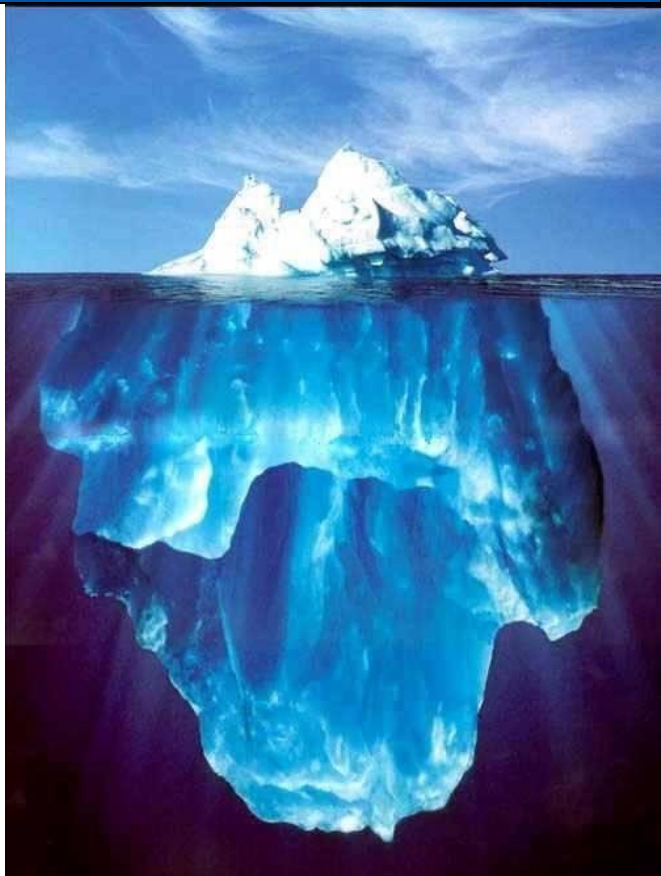
（2017年1月第2版）厦门大学 林子雨 编著，人民邮电出版社

ISBN:978-7-115-44330-4

欢迎访问《大数据技术原理与应用》教材官方网站，

免费获取教材配套资源：

<http://dblab.xmu.edu.cn/post/bigdata>





# 主讲教师



热爱教学，认真科研，关爱学生  
努力打造独具风格的“特产教师”  
我是林子雨，我为自己代言

2013年度厦门大学奖教金获得者  
2017年度厦门大学奖教金获得者

主讲教师：林子雨

中国高校首个“数字教师”提出者和建设者

2009年7月从事教师职业以来

累计免费网络发布超过500万字高价值教学和科研资料

网络浏览量超过500万次





# 大数据之门

为学生在大数据领域“深耕细作”奠定基础、指明方向



# 教材介绍

## 《大数据技术原理与应用——大数据概念、存储、处理、分析与应用》

厦门大学 林子雨编著，人民邮电出版社，2017年1月第2版

ISBN:978-7-115-44330-4

- 国内高校第一本系统介绍大数据知识专业教材
- 京东、当当等各大网店畅销书籍
- 大数据入门教材精品
- 国内多所高校采用本教材开课
- 配套目前国内高校最完备的课程公共服务平台
- 福建省精品在线开放课程







# 内容提要

- 本课程系统介绍了大数据相关知识，共有**13**讲
- 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、MapReduce、Spark等重要章节，安排了入门级的实践操作，让学生更好地学习和掌握大数据关键技术



# MOOC的13讲内容

- 第1讲 大数据概述
- 第2讲 大数据处理架构Hadoop
- 第3讲 分布式文件系统HDFS
- 第4讲 分布式数据库HBase
- 第5讲 NoSQL数据库
- 第6讲 云数据库
- 第7讲 MapReduce
- 第8讲 Hadoop再探讨
- 第9讲 Spark
- 第10讲 流计算
- 第11讲 图计算
- 第12讲 数据可视化
- 第13讲 大数据在不同领域的应用



# MOOC的13讲内容与第2版教材的对应关系

MOOC中的章节	对应的第2版教材章节
第1讲 大数据概述	第1章 大数据概述
第2讲 大数据处理架构Hadoop	第2章 大数据处理架构Hadoop
第3讲 分布式文件系统HDFS	第3章 分布式文件系统HDFS
第4讲 分布式数据库HBase	第4章 分布式数据库HBase
第5讲 NoSQL数据库	第5章 NoSQL数据库
第6讲 云数据库	第6章 云数据库
第7讲 MapReduce	第7章 MapReduce
第8讲 数据仓库Hive	不在第2版教材中
第9讲 Hadoop再探讨	第8章 Hadoop再探讨
第10讲 Spark	第9章 Spark
第11讲 流计算	第10章 流计算
第12讲 图计算	第11章 图计算
第13讲 大数据在不同领域的应用	第13章 大数据在互联网领域的应用 第14章 大数据在生物医学领域的应用 第15章 大数据在其他领域的应用



# 课程配套教学资源



## 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



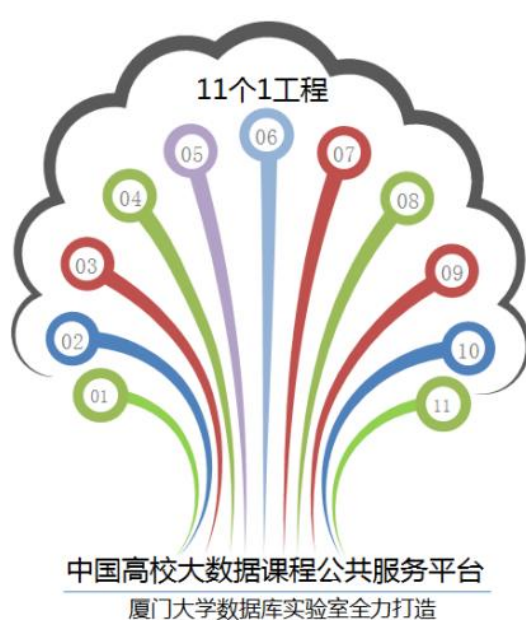
扫一扫观看3分钟FLASH动画宣传片

所有资料全部免费共享  
支持电脑和手机浏览





# 课程配套教学资源



- 01 1本教材
- 02 1个教师服务站
- 03 1个学生服务站
- 04 1个公益项目
- 05 1堂巡讲公开课
- 06 1个示范班级
- 07 1门在线课程
- 08 1个交流群
- 09 1个保障团队
- 10 1个培训基地
- 11 1个实验平台

- 国内高校第一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
- 平台所有资源全部通过网络**免费共享（无需注册）**、采用**WORD**和**PPT**可编辑格式



# 课程配套教学资源——大数据课程学生服务站

**“大数据课程学生服务站”**

**为学生学习大数据课程提供一站式服务**

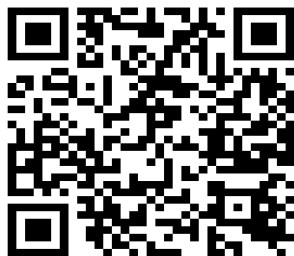


发起人：厦门大学计算机系教师 林子雨 博士  
保障团队：厦门大学数据库实验室  
共建单位：厦门大学云计算与大数据研究中心

为学生学习大数据课程提供课程教材、讲义PPT、技术资料、进阶学习指南等全方位、一站式服务，帮助学生形成知识体系、开展基本编程实践、明确进阶学习方向，提高学习效率，提升学习效果

官网：<http://dblab.xmu.edu.cn/post/4331/>

学生服务站地址：<http://dblab.xmu.edu.cn/post/4331/>



扫一扫访问学生服务站





中国高校大数据课程公共服务平台

# 3分钟FLASH 动画宣传片



# 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://www.cs.xmu.edu.cn/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。中国高校首个“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。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》



材

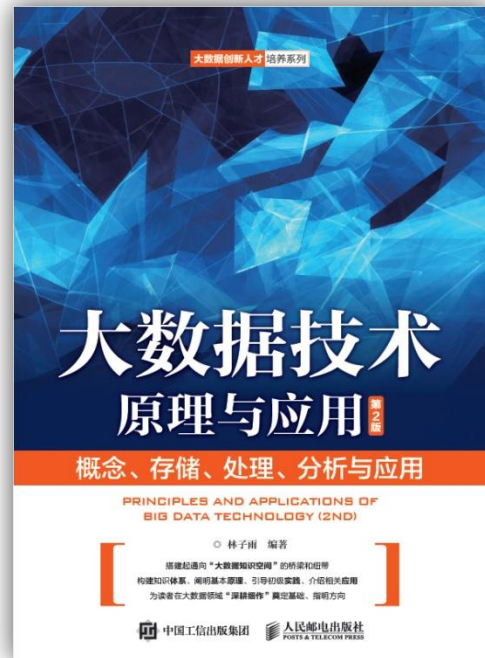
扫一扫访问教材官网

《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，是中国高校第一本系统介绍大数据知识的专业教材。

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>







# 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

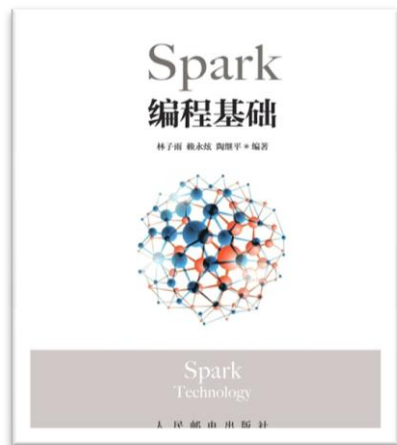
清华大学出版社 ISBN:978-7-302-47209-4



# 附录E：《Spark编程基础》教材

## 《Spark编程基础》

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:978-7-115-47598-5

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

授课视频：<http://dblab.xmu.edu.cn/post/10482/>



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

国内高校首个大数据课程公共服务平台，为全国高校教师和学生提供大数据教学资源一站式“免费”在线服务，包括课程教材、讲义PPT、课程习题、实验指南、学习指南、备课指南、授课视频、技术资料、实验案例、在线教程等，目前平台每年访问量超过100万次，成为全国高校大数据教学知名品牌



## 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

The background is a solid blue color. It features several faint, light-blue silhouettes of people. In the top left, a group of people is holding hands in a circle. In the top center, a group of people is standing in a line, holding hands. In the bottom left, a person is sitting and looking towards the right. In the bottom right, a person is standing and looking towards the left. The text "Thank You!" is centered in the middle of the image.

# **Thank You!**

Department of Computer Science, Xiamen University, 2018