



扫一扫下载PPT



高校大数据专业建设若干问题探讨

林子雨 博士/副教授

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://dbllab.xmu.edu.cn/post/linziyu>





提纲

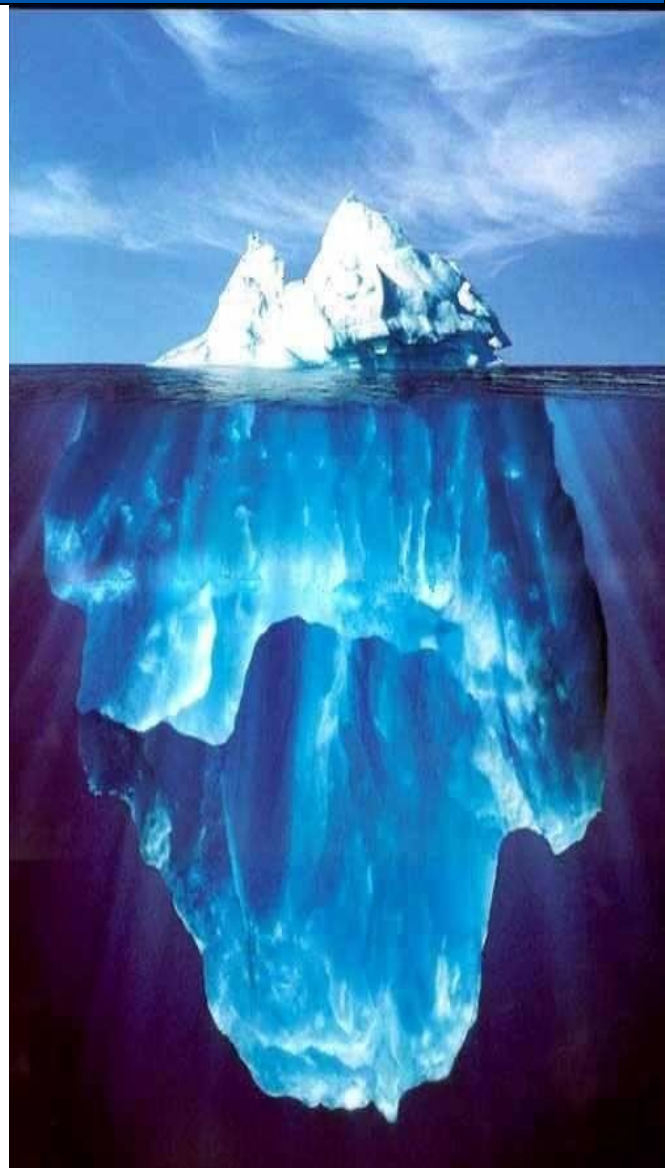
- 一、高校大数据专业
- 二、大数据导论课
- 三、大数据实验环境
- 四、大数据实训课



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





主讲教师



2020年度厦门大学奖教金获得者
2018年国家精品在线开放课程

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**1000万字**高价值教学和科研资料

网络浏览量超过**1000万次**



数字教师LOGO



一、高校大数据专业

- 1.1 大数据专业的人才培养目标
- 1.2 毕业生就业岗位
- 1.3 大数据专业知识体系
- 1.4 大数据专业课程体系
- 1.5 大数据专业的编程语言
- 1.6 厦门大学建设的大数据课程体系

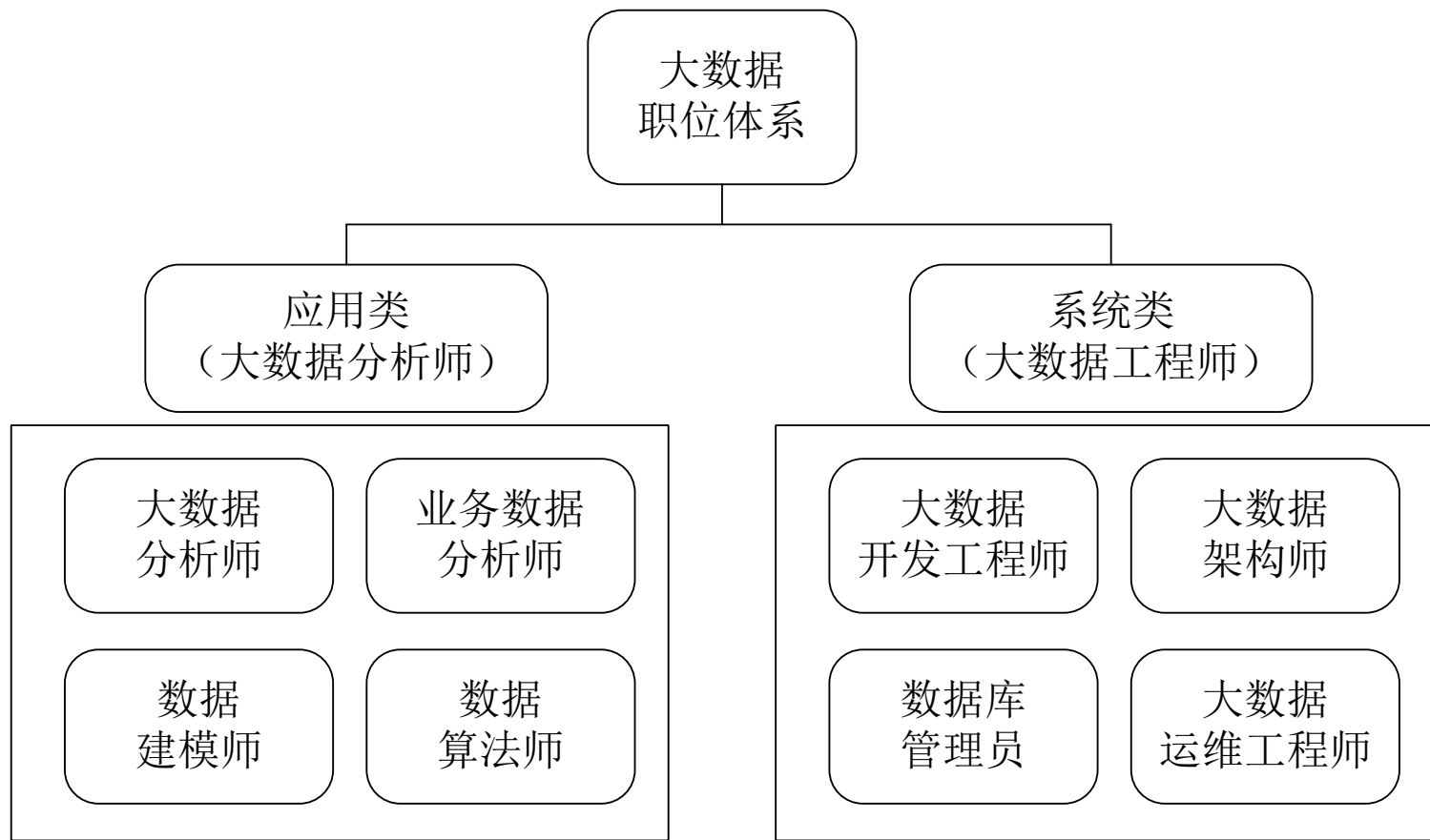


1.1 大数据专业的人才培养目标

大数据专业致力于培养符合国家战略及大数据产业发展需求，具备较好的数据素养和数理基础、扎实的编程基础以及大数据基础知识与技能，熟练掌握大数据采集、预处理、存储、处理、分析、应用技术，能够运用大数据思维、模型和工具解决实际问题的高级复合型人才。大数据专业的毕业生能在互联网企业、金融机构、科研院所、高等院校等从事大数据分析、挖掘、处理、服务、应用和研究工作，亦可从事各行业大数据系统的集成、设计、开发、管理、维护等工作，也适合在高等院校及科研院所的相关交叉学科继续深造。



1.2 毕业生就业就业岗位

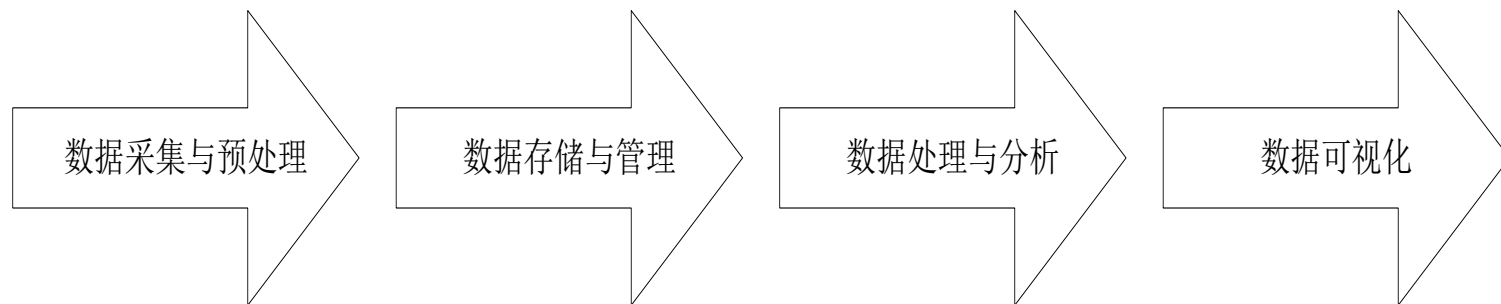




1.3 大数据专业知识体系

从学科角度而言，大数据可以理解为一个跨多学科领域的，从数据中获取知识的科学方法、技术和系统的集合。因此，大数据专业知识体系涵盖了计算机、数学、统计学等多个学科领域，结合了诸多领域中的理论和技术，包括应用数学、统计学、模式识别、机器学习、人工智能、深度学习、数据可视化、数据挖掘、数据仓库、分布式计算、云计算、系统架构设计等。

从大数据分析角度而言（如图所示），典型的大数据分析过程包括：数据采集与预处理、数据存储与管理、数据处理与分析、数据可视化等。因此，大数据专业知识体系涵盖了数据采集与预处理技术、数据存储与管理技术、数据处理与分析技术、数据可视化技术等。同时，在分析过程中，对商业领域的业务知识也需要一定的理解。





1.4 大数据专业课程体系

大数据专业课程体系涵盖通识教育课、学科基础课、专业基础课、专业核心课和专业课，具体如下：

- (1) 通识教育课：思政类课程、军体类课程、外语课、创新创业课等；
- (2) 学科基础课：高等数学（数学分析）、线性代数、概率论与数理统计、离散数学等；
- (3) 专业基础课：程序设计、计算机系统基础及组成原理、计算机网络、算法与数据结构、数据库系统、操作系统、软件工程等；
- (4) 专业核心课：大数据导论、网络爬虫与数据采集、数据清洗、NoSQL数据库、数据可视化、分布式并行编程、机器学习等；
- (5) 专业课：云计算、数据安全、数据仓库、数据挖掘等。



讨论1：各个高校大数据专业的课程设置

北京大学：

1) 公共与基础课程：40-46学分

大学英语系列课程（2-8学分），政治系列课程（14学分），计算机系列课程（6学分），体育系列课程（4学分），数学分析(14学分)

2) 核心课程：29学分

高等代数 I (5学分)，高等代数 II（4学分），几何学（5学分），抽象代数（3学分），复变函数(3学分)，常微分方程（3学分），数学模型（3学分），概率论（3学分）



讨论1：各个高校大数据专业的课程设置

3) 限选课程：32学分

a) 专业必选课程（6学分）：数理统计(3学分)，应用多元统计分析(3学分)

b) 限选课程: 在下列课程中选修15学分

实变函数（3学分）/实函与泛函（4学分），应用回归分析（3学分），贝叶斯理论与算法（3学分），应用时间序列分析（3学分），统计计算（3学分），统计机器学习（3学分），程序设计实习（3学分），数据结构与算法（3学分），分布与并行计算（3学分），算法设计与分析（3学分），数据库概论（3学分），自然语言处理导论（3学分），数值与计算方法（3学分），人工智能（3学分），最优化方法（3学分），深度学习（3学分），数据科学导论（3学分）



讨论1：各个高校大数据专业的课程设置

中国人民大学：

基础课程(38学分)：数学分析、高等代数、普通物理数学与信息科学概论、数据结构、数据科学导论、程序设计导论、程序设计实践。

必修课(37学分)：离散数学、概率与统计、算法分析与设计、数据计算智能、数据库系统概论、计算机系统基础、并行体系结构与编程、非结构化大数据分析。

选修课：数据科学算法导论、数据科学专题、数据科学实践、互联网实用开发技术、抽样技术、统计学习、回归分析、随机过程。



讨论2：大数据课程与云计算课程相互关系

云计算课程知识要点

- 云计算概念、云计算体系架构、数据中心、虚拟化技术（平台虚拟化、资源虚拟化、虚拟机的动态迁移、云操作系统）、SOA架构及开发技术、云数据中心设计与测试、云数据中心维护与管理、云安全架构、桌面云、PaaS应用开发平台、开源的云计算管理平台Openstack、Docker容器、大数据存储与管理

大数据课程知识要点

- 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
- 流计算
- 图计算
- Spark



讨论3：大数据专业课程中的数学知识

最核心的数学知识

一、矩阵和向量空间

- 1.标量、向量与矩阵
- 2.特殊矩阵（单位矩阵、对角矩阵、三角矩阵）
- 3.矩阵运算（矩阵的加减法、矩阵的乘法、矩阵的除法、矩阵的转置）
- 4.向量空间

二、概率：量化随机

- 1.定义概率：事件和概率空间
- 2.条件概率：信息的价值
- 3.随机变量：两种不同的随机
- 4.正态分布
- 5.P值

三、微积分

- 1.导数和积分：位置、速度
- 2.极限：变化的终点
- 3.复合函数：链式法则
- 4.多元函数：偏导数
- 5.极值与最值：最优选择



讨论3：大数据专业课程中的数学知识

- 离散数学
- 概率论与数理统计
- 高等数学（数学分析）
- 线性代数



1.5大数据专业的编程语言

- 1.C语言
- 2.C++
- 3.Java
4. Python
- 5.Scala
- 6.R语言



节选自林子雨编著《大数据导论》

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



1.5 大数据专业的编程语言

1. C语言

C语言是一门面向过程的计算机编程语言，与C++、Java等面向对象编程语言有所不同。C语言的设计目标是提供一种能以简易的方式编译、处理低级存储器、仅产生少量的机器码以及不需要任何运行环境支持便能运行的编程语言。C语言描述问题比汇编语言迅速、工作量小、可读性好、易于调试、修改和移植，而代码质量与汇编语言相当。C语言一般只比汇编语言代码生成的目标程序效率低10%~20%。因此，C语言可以编写系统软件。C语言在一些编程语言排行榜中长期排在第一的位置。



1.5 大数据专业的编程语言

1. C语言

C语言具有很多优点，主要如下：

(1) 它具有现代高级程序设计语言的基本语法特征，并且是编写操作系统的首选语言，与计算机硬件打交道时灵巧且高效，目前几乎所有的操作系统（如Windows、Unix和Linux等）均是由C语言编写的；

(2) 常用的面向对象程序设计语言（例如C++和Java），其基本语法源于C语言。C语言甚至是其它编程语言的母语言，比如Java语言就是用C语言编写的。

(3) 简洁紧凑，灵活方便。C语言一共只有32个关键字，9种控制语句，程序书写自由，主要用小写字母表示，它把高级语言的基本结构和语句与低级语言的实用性结合了起来。

C语言一般作为学习计算机程序设计语言的入门语言。



1.5 大数据专业的编程语言

2. C++

C++是C语言的继承，是一门以C为基础发展而来的、面向对象的高级程序设计语言，它既可以进行C语言的过程化程序设计，又可以进行以继承和多态为特点的面向对象的程序设计。C++不仅拥有计算机高效运行的实用性特征，同时还致力于提高大规模程序的编程质量与程序设计语言的问题描述能力。

C++的优点主要包括：

- (1) 实现了面向对象程序设计，处理运行速度非常快，大部分的游戏软件都是由C++来编写的。
- (2) 语言非常灵活，功能非常强大。
- (3) 非常严谨、精确和数理化，标准定义很细致。
- (4) 语言的语法思路层次分明。



1.5大数据专业的编程语言

3.Java

Java是目前最热门的编程语言之一，在一些编程语言排行榜中长期排在前三名。虽然Java没有和 R、Python一样好的可视化功能，也不是统计建模的最佳工具，但是，如果需要建立一个庞大的应用系统，那么Java通常会是比较理想的选择。由于Java具有简单、面向对象、分布式、鲁棒、安全、体系结构中立、可移植、高性能、多线程以及动态性等诸多优良特性，因此，被大量应用于企业大型系统开发中，企业对于Java人才的需求一直比较旺盛。



1.5大数据专业的编程语言

4. Python

Python是目前国内外很多大学里流行的入门语言，学习门槛低，简单易用，开发人员可以使用Python来构建桌面应用程序和Web应用程序，此外，Python在学术界备受欢迎，常被用于科学计算、数据分析和生物信息学等领域。Python是最近几年发展最为迅速的编程语言，在一些编程语言排行榜当中甚至已经进入了前三名。



1.5 大数据专业的编程语言

Python的主要优点如下：

(1) 可以使用多种执行方式。可以直接在命令行执行相关命令，也可以用函数的方式执行相关命令，或者也可以用面向对象的方式执行相关命令。

(2) 语法简洁，且强制缩格，程序具有很好的可读性。

(3) 跨平台。支持多种开发平台，如Windows、Linux、Mac OS X、Solaris等。

(4) 面向对象。Python既支持面向过程，又支持面向对象，这使得其编程更加灵活。

(5) 丰富的第三方库。Python有丰富且强大的库，而且由于Python的开源特性，第三方库非常多，如Web开发、爬虫、科学计算等。

在数据分析领域，Python是广受欢迎的编程语言，网络数据采集（比如网络爬虫）、数据清洗、数据分析与挖掘、数据可视化等环节，通常都使用Python语言编写程序。



1.5 大数据专业的编程语言

5. Scala

Scala是一门类似Java的多范式语言，它整合了面向对象编程和函数式编程的最佳特性，具有诸多优点，主要包括以下几个方面：

- (1) 具备强大的并发性，支持函数式编程，可以更好地支持分布式系统；
- (2) Scala兼容Java，可以与Java互操作；
- (3) Scala代码简洁优雅；
- (4) Scala支持高效的交互式编程；
- (5) Scala是Spark的开发语言。



1.5 大数据专业的编程语言

6. R语言

R是专门为统计和数据分析开发的语言，具有数据建模、统计分析和可视化等功能，简单易上手。R语言主要具有如下优点：

(1) 免费开源。R的源代码可以自由下载使用，也有已编译的可执行文件版本可以下载。

(2) 简单易学。虽然R与其他程序设计语言相比结构相对松散，使用变量前不需要明确定义变量类型等，但是，仍然保留了程序设计语言的基础逻辑与自然的语言风格。

(3) 几乎兼容全部平台。除了支持OS X、Linux、Windows之外，甚至可以在iOS设备上编辑和运行R程序，还可以在iPhone等移动设备上安装R程序。

(4) 多领域的统计资源。学者和数据分析师开发了很多R语言包，涉及到统计的各个方面，资源很丰富。

(5) 出色的图形统计功能。除了基本统计直方图、折线图等，还可以绘制一些高级的图形，而这些是SPSS这类软件所不能匹敌的。



1.6 厦门大学建设的大数据课程体系

大数据导论（通识课版）
（全校公共选修课）



面向理工科的大数据课程



1.6 厦门大学建设的大数据课程体系

- **【1】** 林子雨.大数据导论.人民邮电出版社, 预计**2020年9月**出版.(入门级, 大数据专业的导论课教材)
- **【2】** 林子雨.大数据导论(通识课版).高等教育出版社, **2020年2月**.(入门级, 公共选修课教材)
- **【3】** 林子雨.大数据技术原理与应用(第2版).人民邮电出版社, **2017年1月**.(入门级, 专业课教材)
- **【4】** 林子雨.大数据基础编程、实验和案例教程.清华大学出版社, **2017年8月**.(入门级, 专业课教材)
- **【5】** 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**.(进阶级, 专业课教材)
- **【6】** 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**.(进阶级, 专业课教材)
- **【7】** 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**.(实训课, 专业课教材)
- **【8】** 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**.(实训课, 专业课教材)
- 关于这些教材的配套资源(讲义PPT、视频、实验、案例等), 请见网页:
<http://dbl原因.xmu.edu.cn/post/bigdatabook/>





1.6 厦门大学建设的大数据课程体系



林子雨大数据系列教材

用于导论课、专业课、实训课、公共课

系列教材官网：<http://dblab.xmu.edu.cn/post/bigdatabook/>

扫一扫查看所有教材及其配套资源





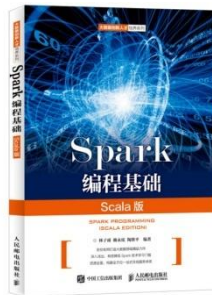
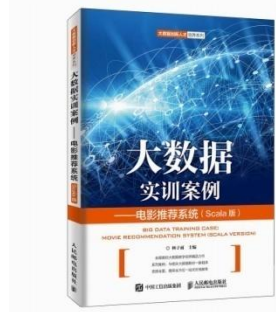
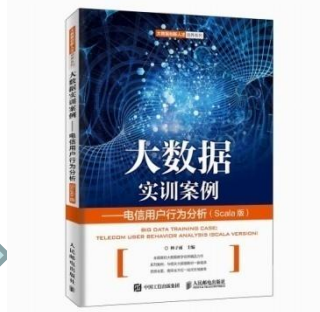
1.6 厦门大学建设的大数据课程体系



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



1.6 厦门大学建设的大数据课程体系



1+1黄金组合
厦门大学林子雨编著





2.大数据导论课

- 《大数据导论（通识课版）》（面向全校非计算机专业）
- 《大数据导论（专业课版）》（面向大数据专业）



2.大数据导论课

- 《大数据导论（通识课版）》（面向全校非计算机专业）

- 关于通识课

一般而言，通识教育课程是指除专业教育之外的基础教育课程。如果说专业教育旨在培养学生在某一知识领域的专业技能和谋生手段，那么通识课程则要通过知识的基础性、整体性、综合性、广博性，使学生拓宽视野、避免偏狭，培养独立思考与判断能力、社会责任感和健全人格



2.大数据导论课

- 《大数据导论（通识课版）》（面向全校非计算机专业）
- 一般而言，我国大学的本科课程体系主要由“公共基础课+学科基础课+专业课”三部分构成。其中，“公共基础课”部分均应视作“通识课程”，主要包括思想政治理论课、英语、军训与体育、通识教育选修课等，理工科大学还包括数理化等自然科学基础课。



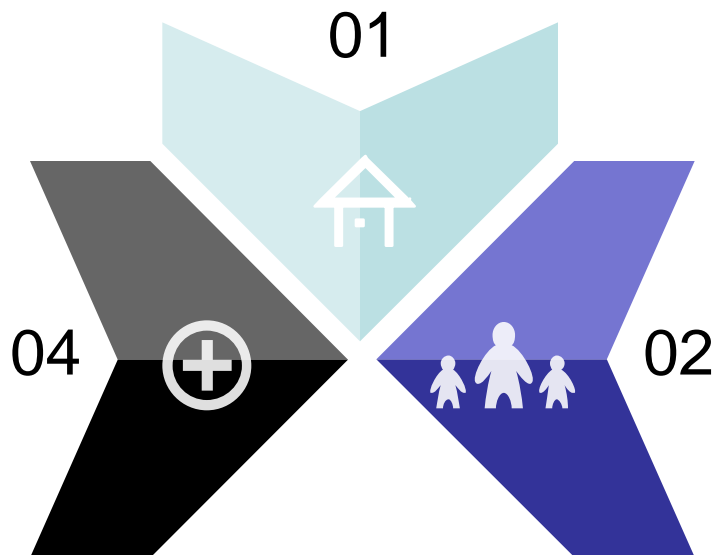
2.大数据导论课

- 《大数据导论（通识课版）》（面向全校非计算机专业）

课程定位

数据意识

数据伦理



数据思维

数据能力



2.大数据导论课

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



2.大数据导论课

课程教材



扫一扫访问
教材官网

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



2.大数据导论课

配套慕课



已经录制完成
2020年秋季上线
中国大学MOOC



2.大数据导论课

《大数据导论（通识课版）》讲课视频已经在网易云课堂上线
<https://study.163.com/course/introduction/1209607806.htm>





2.大数据导论课

课程内容

- 大数据概述
- 大数据与云计算、物联网和人工智能
- 大数据技术
- 大数据应用
- 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
- 大数据治理



2.大数据导论课

学时安排 (32学时)

章 (或节)	学时安排
第1章 大数据概述	2
第2章 大数据和云计算、物联网的关系	2
第1次讨论课	2
第3章 大数据技术	2
第4章 大数据应用	4
第5章 大数据安全	2
第6章 大数据思维	2
第7章 大数据伦理	2
第2次讨论课	2

第8章 数据共享	2
第9章 数据开放	2
第3次讨论课	2
第10章 大数据交易	2
第11章 大数据治理	2
第4次讨论课	2



2.大数据导论课

《大数据导论》（大数据专业导论课）

- 第1章 大数据概述
- 第2章 大数据与其他新兴技术的关系
- 第3章 大数据基础知识
- 第4章 大数据应用
- 第5章 数据采集与预处理
- 第6章 数据存储与管理
- 第7章 数据处理与分析
- 第8章 数据可视化
- 第9章 大数据分析综合案例

《大数据导论》教材预计2020年9月上市销售
讲义PPT、授课视频、课程习题正在制作中.....

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





3.大数据实验环境





3.大数据实验环境

- 单机构建实验环境
 - 本地：虚拟机、8GB内存
 - 单机模拟集群：Docker容器、多个虚拟机
 - 参考：<http://dblab.xmu.edu.cn/blog/337-2/>
- 实验室多机构建分布式环境
 - 参考：<http://dblab.xmu.edu.cn/blog/install-hadoop-cluster/>
- 构建云端大数据环境：比如阿里云
 - 参考：在阿里云中搭建大数据实验环境
<http://dblab.xmu.edu.cn/blog/1952-2/>
- 统一大数据实验机房
 - 一台服务器推送云桌面到多台终端机器
 - 多台物理机器构建分布式环境
 - 云端



4.大数据实训课



一、基本信息

- 1.实训题目：基于协同过滤算法的电影推荐
- 2.实训时间：2018年9月3日-28日
- 3.实训学生：厦门理工学院计算机与信息工程学院软件工程专业2015级本科生（大四），44人
- 4.实训类型：软件工程综合实践（大数据方向）
- 5.授课教师：厦门大学计算机科学系 林子雨 老师
- 6.课程设计学时（周数）：4周/84课时
- 7.实训教材：

林子雨编著《大数据实训案例——电影推荐系统》



扫一扫访问
实训班级主页



4.大数据实训课



8.实训课程需要达到的专业工程认证毕业要求

- (1) 指标点4.1：针对软件开发特定需求，了解信息学科发展动态；
- (2) 指标点9.1：正确理解个人与团队的关系，理解团队合作的重要性，具备在多学科背景下团队合作的意识和能力；
- (3) 指标点9.2：在多学科背景的团队下，理解团队成员的不同角色在团队中的作用，能够作为个体或团队成员完成所承担的任务；
- (4) 指标点10.2：具有书面表达与沟通能力、科技论文与技术报告写作能力，能够撰写报告和设计文稿；
- (5) 指标点11.2：了解市场、用户的需求变化以及技术发展，提出技术改造、系统更新、效能改进的方案。



4.大数据实训课

9.实训教室：数字内容与信息智能处理实验室

实验室位于厦门理工学院精工园1号楼416教室，建于2008年6月，实验室面积约120平方米，设备总额约46万元。实验室配备具有较强的数据运算与图形图像处理能力的计算机，能满足图形图像处理、遥感图像处理、数据挖掘、大数据等课程的需求。

主要仪器设备：Dell OptiPlex7040 计算机61台，主要技术参数：CPU: Intel 酷睿i5-6500；内存：16GB；硬盘：1TB*2；显示器：23寸 LCD。





4.大数据实训课



10.实训管理系统

本次实训课程采用“教学过程一体化管理平台”，对教学全过程进行高效的信息化管理。该系统具备课程资源管理、学生管理、授课过程管理、授课质量分析、成绩自动统计等功能。

实训管理系统由厦门大学数据库实验室和厦门理工学院团队联合研发，各项功能根据一线教师的教学需求“量身定制”，在一线教学环境中经过大量实践应用

教学过程一体化管理平台

我的面板

在授课课程

序号	课程名	开始时间	结束时间
1	软件工程综合实践	2018-09-03	2018-09-28

我的历史课程成绩

综合成绩报表

团队成绩排名	学员成绩排名
1 团队1 0.00	1 陈德翰 0.00
2 团队2 0.00	2 郑神奥 0.00
3 团队3 0.00	3 陈卓 0.00
4 团队4 0.00	4 方向旭 0.00
5 团队5 0.00	5 李家豪 0.00
6 团队6 0.00	6 林斌彬 0.00
7 团队7 0.00	7 林城 0.00
8 团队8 0.00	8 吴照 0.00
9 团队9 0.00	9 许志峰 0.00
	10 张亚辉 0.00



4.大数据实训课

二、课程设计的目的、任务与要求

1.课程设计目的

- 了解并掌握软件项目开发过程；
- 训练独立或协同实施软件项目开发各环节的能力；
- 培养良好的表达、沟通和团队协作能力；
- 培养自主学习能力，能够独立地分析问题和解决问题；
- 了解开发文档的作用及撰写规范，培养撰写文档的意识与能力；
- 了解并掌握企业化的编程规范；
- 积累一定的编码量，进一步提高编码能力；
- 了解大数据软件项目的架构；
- 了解大数据应用的开发需求，能够搭建及配置符合需求的开发平台与环境；
- 掌握关系数据库的原理和操作方法；
- 掌握Hadoop分布式文件系统HDFS的原理和编程操作方法；
- 掌握大数据处理架构Spark的原理和Spark MLlib程序编写方法；
- 掌握数据ETL工具Kettle的使用方法；
- 了解数据挖掘的步骤和挖掘方法，学会基于协同过滤算法的Spark程序编写；
- 掌握数据可视化的技术；





4.大数据实训课

2. 课程设计任务与要求

本课程设计项目涉及数据的预处理、数据的存储、数据查看、算法编写、算法应用和大数据结果展现等全流程所涉及的各种典型操作，涵盖Linux、MySQL、Hadoop、Spark、SparkMlib、IntelliJ IDEA（简称IDEA）、Node.js等系统和软件的安装和使用方法。通过本项目，将有助于学生综合运用大数据课程知识以及各种工具软件，实现数据全流程操作。

本课程设计项目需要完成以下任务：

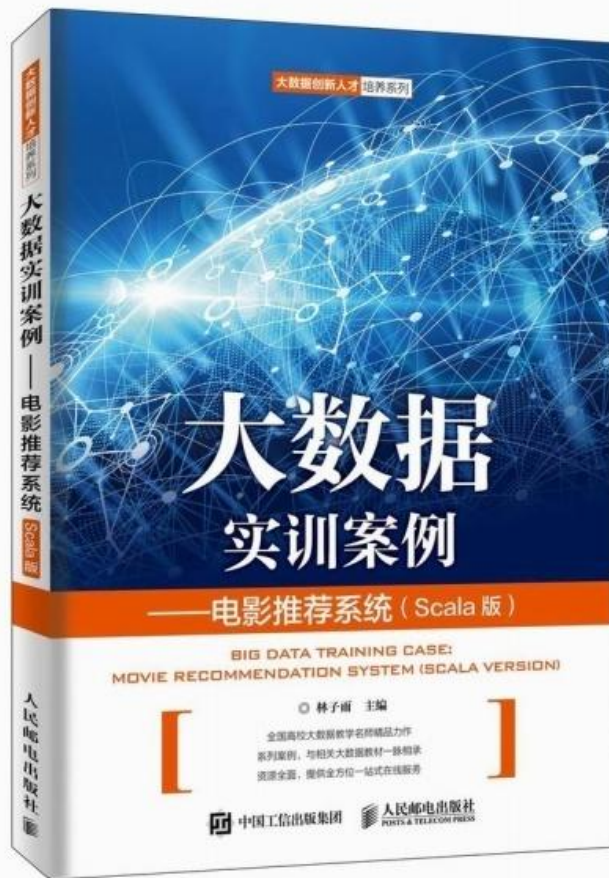
- 安装Ubuntu操作系统；
- 安装关系数据库MySQL；
- 安装大数据处理框架Hadoop；
- 对Hadoop进行伪分布式环境配置；
- 安装服务端语言js；
- 安装ETL工具Kettle；
- 对数据文件形式的原始数据集进行预处理；
- 把数据文件的数据集ETL到HDFS中；
- 使用Kettle将数据文件从Ubuntu本地文件导入HDFS；
- 应用IDEA工具开发Scala程序；
- 编写Spark程序实现电影推荐功能；
- 使用协同过滤算法实现电影的推荐；
- 使用Node.js搭建Web动态网页获取推荐结果。





4.大数据实训课

三、课程教材



扫一扫访问教材官网

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



4.大数据实训课

四、授课过程

授课时间：2018年9月3日到28日，共84学时

表一

号	课程主题	授课日期	授课时间	学时	作业
1	Linux系统安装及使用方法	2018-09-03	08:20 ~ 11:50	4	作业1：Linux安装、常用命令以及相关软件用法
2	Hadoop的安装和使用方法	2018-09-03	14:20 ~ 16:50	3	作业2：安装Hadoop
3	分布式文件系统HDFS介绍及其常用Shell命令	2018-09-04	08:20 ~ 11:50	4	作业3：HDFS常用Shell命令
4	Spark框架介绍及环境搭建	2018-09-04	14:20 ~ 16:50	3	作业4：Spark环境搭建及程序编译运行
5	Scala语言基础	2018-09-05	08:20 ~ 11:50	4	作业5：Scala安装和初级编程实践 团队建设文档 团队奖惩制度 9-5小组会议纪要
6	Spark编程基础	2018-09-07	08:20 ~ 11:50	4	作业6：RDD编程初级实践
7	开发工具IntelliJ IDEA的安装和使用方法	2018-09-07	14:20 ~ 16:50	3	作业7：使用IntelliJ IDEA开发Spark应用程序 周记报告（第1周）
8	MySQL数据库和ETL工具Kettle的安装和使用方法	2018-09-10	08:20 ~ 11:50	4	作业8：MySQL数据库的安装和使用方法 作业9：ETL工具Kettle的安装和使用方法



4.大数据实训课

表一 (续1)

9	推荐系统与协同过滤算法	2018-09-10	14:20 ~ 16:50	3	文献综述 9-10小组会议纪要
10	Spark MLlib程序编写方法	2018-09-11	08:20 ~ 11:50	4	作业10: 在Spark MLlib中使用ALS算法
11	Node.js的安装和使用方法	2018-09-11	14:20 ~ 16:50	3	作业11: 利用Node.js实现词频统计网页应用程序 9-11小组会议纪要
12	Spark读写MySQL数据库	2018-09-12	08:20 ~ 11:50	4	作业12: 使用Spark SQL读写MySQL数据库 9-12小组会议纪要
13	项目开发计划	2018-09-14	08:20 ~ 11:50	4	撰写项目开发计划
14	撰写需求分析说明书	2018-09-14	14:20 ~ 16:50	3	9-14小组会议纪要 周记报告 (第2周)
15	撰写需求分析说明书	2018-09-17	08:20 ~ 11:50	4	撰写需求分析说明书
16	撰写需求分析说明书	2018-09-17	14:20 ~ 16:50	3	提交需求分析说明书和答辩PPT
17	需求分析说明书评审	2018-09-18	08:20 ~ 11:50	4	需求分析说明书评审



4.大数据实训课

表一 (续2)

18	系统设计说明书撰写	2018-09-18	14:20 ~ 16:50	3	系统设计说明书撰写
19	系统设计说明书撰写	2018-09-19	08:20 ~ 11:50	4	系统设计说明书撰写
20	系统设计说明书撰写	2018-09-21	08:20 ~ 11:50	4	系统设计说明书撰写
21	项目开发	2018-09-21	14:20 ~ 16:50	3	项目开发
22	项目开发	2018-09-25	08:20 ~ 11:50	4	项目开发
23	项目开发和用户手册撰写	2018-09-25	14:20 ~ 16:50	3	项目开发和用户手册撰写
24	项目答辩	2018-09-26	08:20 ~ 11:50	4	项目答辩



4.大数据实训课

五、实训成果文档

实训成果文档包括：

- 团队建设和奖惩制度
- 文献综述
- 项目开发计划
- 需求分析说明书
- 系统设计说明书
- 项目代码
- 项目答辩PPT
- 用户手册

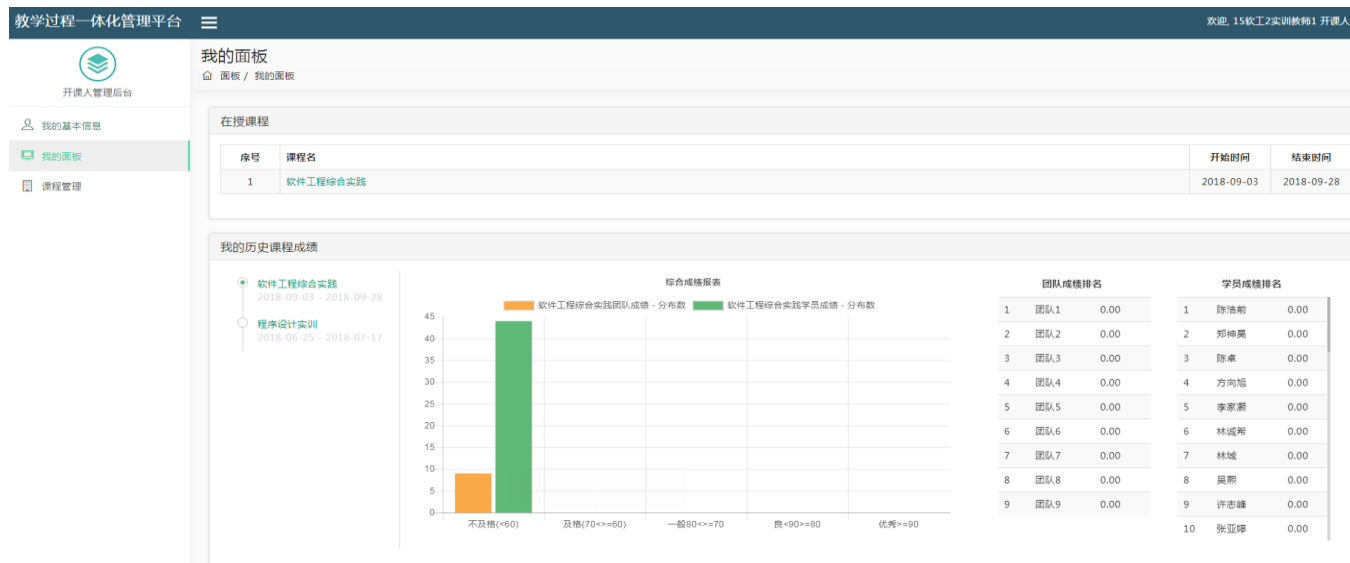




4.大数据实训课

六、教学管理平台

- 本次实训课程采用“教学过程一体化管理平台”，对教学全过程进行高效的信息化管理。该系统具备课程资源管理、学生管理、授课过程管理、授课质量分析、成绩自动统计等功能
- 实训管理系统由厦门大学数据库实验室和厦门理工学院团队联合研发，各项功能根据一线教师的教学需求“量身定制”，在一线教学环境中经过大量实践应用





4.大数据实训课

授课过程管理

教学过程一体化管理平台 白 欢迎, 15软工2实训教师1 开课人

新增

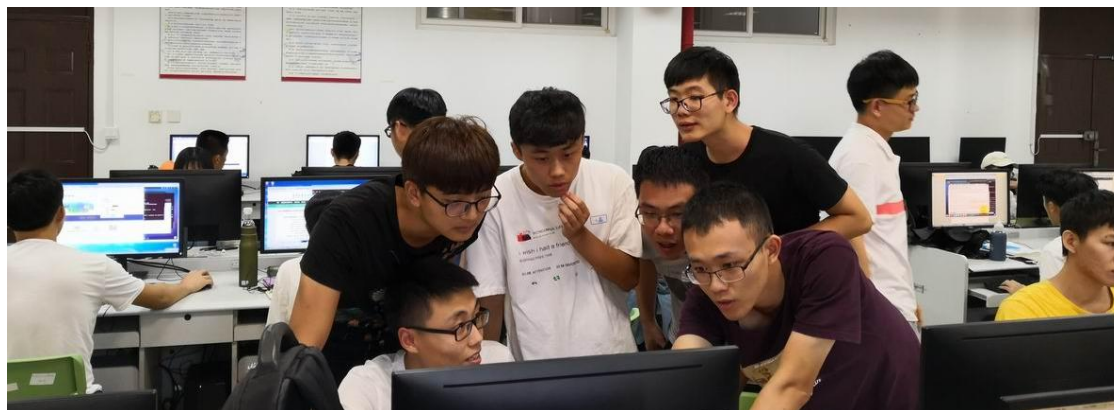
序号	授课主题	课时	授课日期	签到情况(未签到/总人数)	作业类型	作业标题	作业截至日期	任务完成情况	评审情况	课堂记录	操作
1	Linux系统安装及使用方法	4	2018-09-03 上午开始时间 ~ 上午结束时间	查看详情 (1 / 44)	团队	Linux安装、常用命令以及相关软件用法	2018-09-10	任务完成情况	未启用评审	课堂记录	编辑 删除
2	Hadoop的安装和使用方法	3	2018-09-03 下午开始时间 ~ 下午结束时间	查看详情 (0 / 44)	团队	安装Hadoop	2018-09-10	任务完成情况	未启用评审	课堂记录	编辑 删除
3	分布式文件系统HDFS介绍及其常用Shell命令	4	2018-09-04 上午开始时间 ~ 上午结束时间	查看详情 (3 / 44)	团队	HDFS常用Shell命令	2018-09-10	任务完成情况	未启用评审	课堂记录	编辑 删除
4	Spark框架介绍及环境搭建	3	2018-09-04 下午开始时间 ~ 下午结束时间	查看详情 (1 / 44)	团队	Spark环境搭建及程序编译运行	2018-09-10	任务完成情况	未启用评审	课堂记录	编辑 删除
5	Scala语言基础	4	2018-09-05 上午开始时间 ~ 上午结束时间	查看详情 (1 / 44)	团队	9-5小组会议纪要	2018-09-10	任务完成情况	未启用评审	课堂记录	编辑 删除
					学员	Scala安装和初级编程实践	2018-09-10	任务完成情况			
					团队	团队建设文档	2018-09-10	任务完成情况			
团队	团队奖惩制度	2018-09-10	任务完成情况								
6	Spark编程基础	4	2018-09-07 上午开始时间 ~ 上午结束时间	查看详情 (0 / 44)	团队	RDD编程初级实践	2018-09-12	任务完成情况	未启用评审	课堂记录	编辑 删除
7	开发工具IntelliJ IDEA的安装和使用方法	3	2018-09-07 下午开始时间 ~ 下午结束时间	查看详情 (1 / 44)	团队	使用IntelliJ IDEA开发Spark应用程序	2018-09-12	任务完成情况	未启用评审	课堂记录	编辑 删除
8	推荐系统与协同过滤算法	3	2018-09-10 下午开始时间 ~ 下午结束时间	查看详情 (0 / 44)	学员	周记报告 (第1周)	2018-09-12	任务完成情况	未启用评审	课堂记录	编辑 删除
					团队	9-10小组会议纪要	2018-09-12	任务完成情况			
团队	文献综述	2018-09-17	任务完成情况								
9	MySQL数据库和ETL工具Kettle的安装和使用方法	4	2018-09-10 上午开始时间 ~ 上午结束时间	查看详情 (0 / 44)	团队	ETL工具Kettle的安装和使用方法	2018-09-17	任务完成情况	未启用评审	课堂记录	编辑 删除
10	Node.js的安装和使用方法	3	2018-09-11 下午开始时间 ~ 下午结束时间	查看详情 (1 / 44)	团队	MySQL数据库的安装和使用方法	2018-09-17	任务完成情况	未启用评审	课堂记录	编辑 删除
					团队	9-11小组会议纪要	2018-09-17	任务完成情况			
					团队	利用Node.js实现视频统计网页应用程序	2018-09-18	任务完成情况			
11	Spark MLlib程序编写方法	4	2018-09-11 上午开始时间 ~ 上午结束时间	查看详情 (0 / 44)	团队	在Spark MLlib中使用ALS算法	2018-09-18	任务完成情况	未启用评审	课堂记录	编辑 删除
12	Spark读写MySQL数据库	4	2018-09-12 上午开始时间 ~ 上午结束时间	查看详情 (4 / 44)	团队	9-12小组会议纪要	2018-09-18	任务完成情况	未启用评审	课堂记录	编辑 删除
					团队	使用Spark SQL读写MySQL数据库	2018-09-18	任务完成情况			
团队	9-14小组会议纪要	2018-09-17	任务完成情况								
13	需求分析说明书撰写	3	2018-09-14 下午开始时间 ~ 下午结束时间	查看详情 (2 / 44)	学员	周记报告 (第2周)	2018-09-18	任务完成情况	未启用评审	课堂记录	编辑 删除
14	项目开发计划	4	2018-09-14 上午开始时间 ~ 上午结束时间	查看详情 (3 / 44)	团队	撰写项目开发计划	2018-09-18	任务完成情况	未启用评审	课堂记录	编辑 删除
15	需求分析说明书撰写	3	2018-09-17 下午开始时间 ~ 下午结束时间	课程未开始	团队	9-17小组会议纪要	2018-09-23	未完成	未启用评审	课堂记录	编辑 删除
16	需求分析说明书撰写	4	2018-09-17 上午开始时间 ~ 上午结束时间	课程未开始	团队	撰写需求分析说明书	2018-09-18	未完成	未启用评审	课堂记录	编辑 删除
17	需求分析说明书评审	4	2018-09-18 上午开始时间 ~ 上午结束时间	课程未开始	无	无	无		查看评审结果	课堂记录	编辑 删除

汇总 总课时: 61



4.大数据实训课

自动签到统计



授课计划签到统计

面板 / 授课计划管理 / 授课计划签到统计

授课计划签到统计：分布式文件系统HDFS介绍及其常用Shell命令

准点：39人；旷课：0人；迟到：2人；请假：3人

批量更改 全部 准点 旷课 迟到 请假

序号	组名	姓名	签到状态	签到时间	操作
1	团队1	廖建超 (组长)	准点	08:18	编辑
2	团队1	林诚希	准点	08:17	编辑
3	团队1	许志峰	准点	08:17	编辑
4	团队1	林雨婷	准点	08:19	编辑
5	团队1	龚志松	准点	08:19	编辑
6	团队2	郑晋博 (组长)	准点	08:18	编辑
7	团队2	何宇杰	准点	08:17	编辑
8	团队2	吴熙	准点	08:18	编辑
9	团队2	王依婷	准点	08:17	编辑
10	团队2	陈颖颖	准点	11:45	编辑

当前第 1 页共 5 页/每页 10 条 共 44 条

上一页 1 跳转 下一页



4.大数据实训课

自动分组组建团队

教学过程一体化管理平台

- 软件工程综合实践...
- 我的主菜单
- 我的课程面板
- 课程资料
- 学员管理
 - 学员信息管理
 - 学员分组管理
 - 分组测试题
 - 分组测试考试结果
 - 学员智能分组
 - 学生作业下载
- 授课过程
- 教学质量分析
- 个人成绩排名
- 团队成绩排名

陈浩前 组长 1507052102	许富翔 组员 1507052138 设为组长	薛伟强 组员 1507052139 设为组长	吕鸿鹏 组员 1507052132 设为组长	杨一郎 组员 1507052141 设为组长
----------------------	------------------------------	------------------------------	------------------------------	------------------------------

团队名: 团队5 下载所有附件

陈文娟 组长 1507052308	苏伟毅 组员 1507052134 设为组长	李家灏 组员 1507052316 设为组长	苏镇峰 组员 1507052135 设为组长	李鸿利 组员 1507052125 设为组长
----------------------	------------------------------	------------------------------	------------------------------	------------------------------

团队名: 团队6 下载所有附件

陈涛 组长 1507052104	张表铭 组员 1507052143 设为组长	刘凌志 组员 1507052131 设为组长	李闯浩 组员 1507052220 设为组长	林城 组员 1507052323 设为组长
---------------------	------------------------------	------------------------------	------------------------------	-----------------------------

团队名: 团队7 下载所有附件

陈卓 组长 1507052309	蒋东阳 组员 1507052122 设为组长	蔡燕银 组员 1507052403 设为组长	郑时凯 组员 1507052340 设为组长	何蕾 组员 1507052114 设为组长
---------------------	------------------------------	------------------------------	------------------------------	-----------------------------



4.大数据实训课

评审系统自动统计分数



教学过程一体化管理平台

评审结果

需求分析说明书评审

评审专家人数: 3

专家比重: 60

团队比重: 40

序号	被评审团队	专家评审总分	专家评审最终得分	团队互评总分	团队互评最终得分	评审最终成绩	组员	贡献率	评审个人成绩
1	团队1	195.00	39.00	726.00	36.30	75.30	廖建超		75.30
							林莉婷		75.30
							龚志松		75.30
							许志雄		75.30
							林成希		75.30
2	团队2	220.00	44.00	726.00	36.30	80.30	郑露梅		80.30
							陈颖颖		80.30
							王依婷		80.30
							何李杰		80.30
							吴熙		80.30



4.大数据实训课

教师上传课件
学生上传作业
学生下载作业
教师下载作业

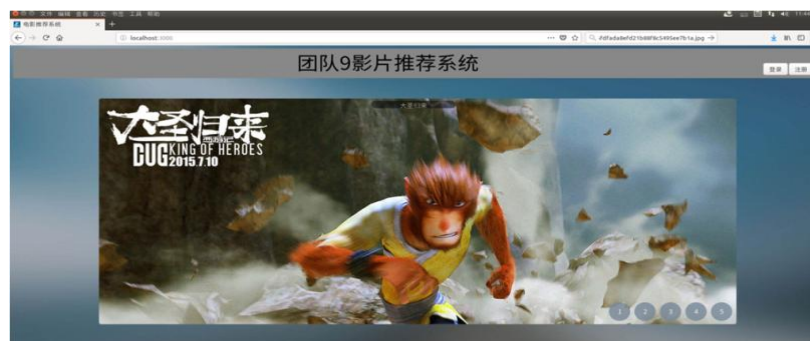
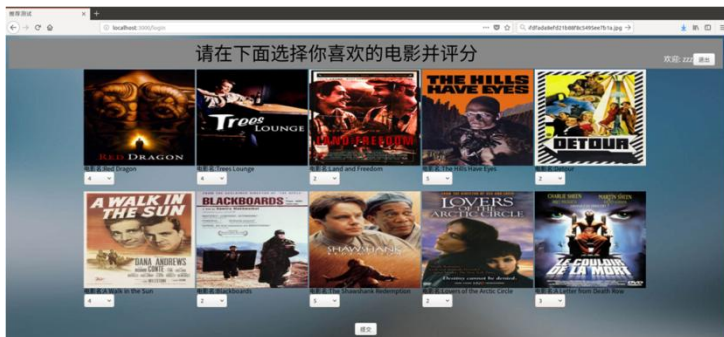
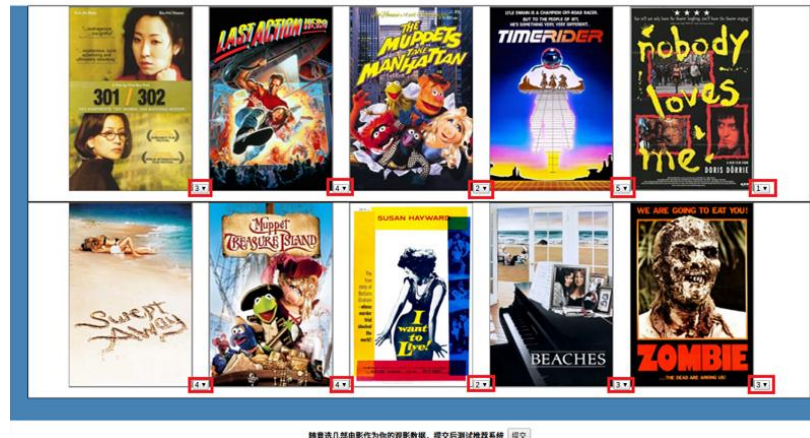
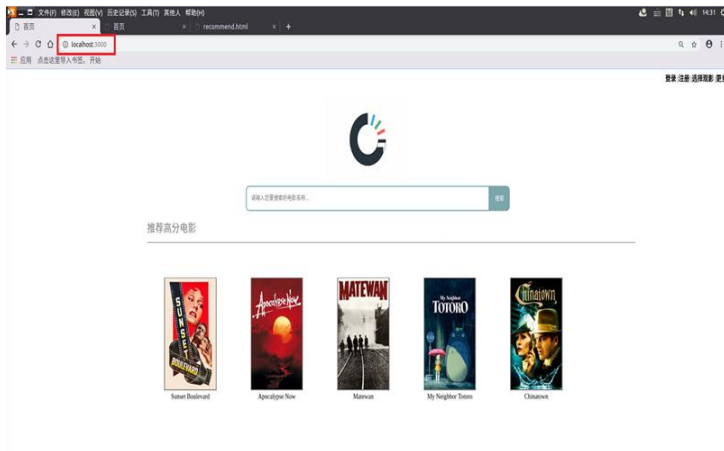


课件上传	
文档类型	附件
作业题目	作业1-Linux系统安装、常用命令以及相关软件用法.docx
	作业2-安装Hadoop.docx
	作业3-HDFS常用Shell命令.docx
	作业4-Spark环境搭建及程序编译运行.docx
	作业5-Scala安装和初级编程实践.docx
	作业6-RDD编程初级实践.docx
	作业7-使用IntelliJ IDEA开发Spark应用程序.docx
	作业8-MySQL数据库的安装和使用方法.docx
	作业9-ETL工具Kettle的安装和使用方法.docx
	作业10-在Spark MLlib中使用ALS算法序.docx
	作业11-利用Node.js实现词频统计网页应用程序.docx
	作业12-使用Spark SQL读写MySQL数据库.docx
	作业13-撰写项目开发计划.docx



4.大数据实训课

七、学生作品





敬请批评指正



高校大数据课程

公共服务平台



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），全国高校知名大数据教师，现为厦门大学计算机科学系副教授，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。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、2017年度和2020年度厦门大学教学类奖教金获得者，荣获2019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青年基金项目(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1000万次。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200万次，累计访问量超过1000万次。



附录B：大数据学习路线图



大数据学习路线图访问地址：<http://dbllab.xmu.edu.cn/post/10164/>



附录C：林子雨大数据系列教材



林子雨大数据系列教材

用于导论课、专业课、实训课、公共课

了解全部教材信息：<http://dbllab.xmu.edu.cn/post/bigdatabook/>



附录D：《大数据导论（通识课版）》教材

开设全校公共选修课的优质教材



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

高等教育出版社 ISBN:978-7-04-053577-8 定价：32元

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



附录E：《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



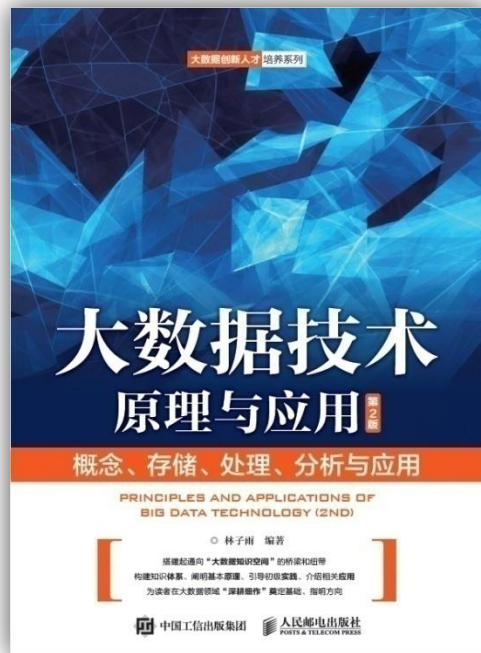
扫一扫访问教材官网

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://dbl原因.xmu.edu.cn/post/bigdata>





附录F：《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

林子雨编著《大数据基础编程、实验和案例教程》
清华大学出版社 ISBN:978-7-302-47209-4 定价：59元



附录G：《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://dblalab.xmu.edu.cn/post/spark/>

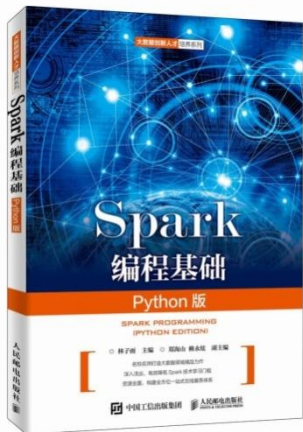


本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录H：《Spark编程基础（Python版）》

《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

人民邮电出版社出版发行，ISBN:978-7-115-52439-3

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



本书以Python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、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、Structured 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、上机实验指南等。



附录I：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J：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：

《电影推荐系统》（已经于2019年5月出版）

《电信用户行为分析》（已经于2019年5月出版）

《实时日志流处理分析》

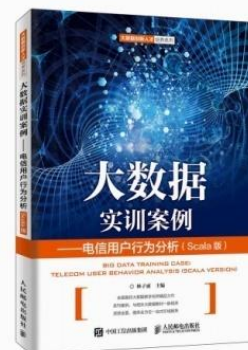
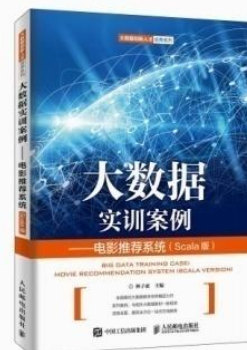
《微博用户情感分析》

《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://dbllab.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

A group of silhouettes of people standing in a circle, holding hands, positioned at the top of the slide.

Thank You!

A group of silhouettes of people standing in a circle, holding hands, positioned at the bottom of the slide.

Department of Computer Science, Xiamen University, 2020