

第3届全国高校大数据教学研讨会 (BDTS2019)



厦门大学大数据课程与教材体系 建设经验分享

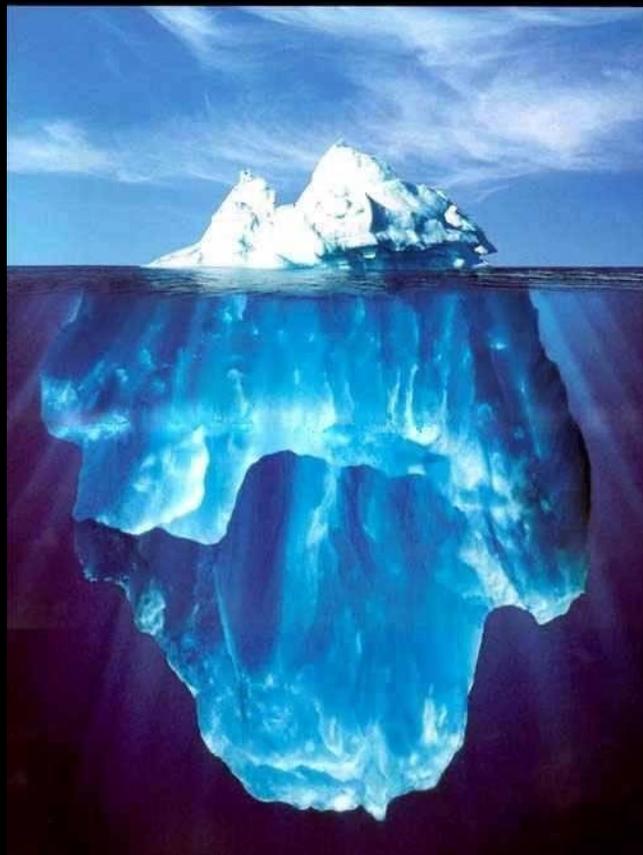
厦门大学 林子雨 博士/助理教授
ziyulin@xmu.edu.cn

2019年5月11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
- 导论课程《大数据技术原理与应用》
- 进阶课程《Spark编程基础》
- 大数据实训课程
- 文科大数据通识导论课程





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://dbllab.xmu.edu.cn/post/8197/>





高校大数据课程公共服务平台

建设周期
六年 (2013-2019)

投入资金
100万+

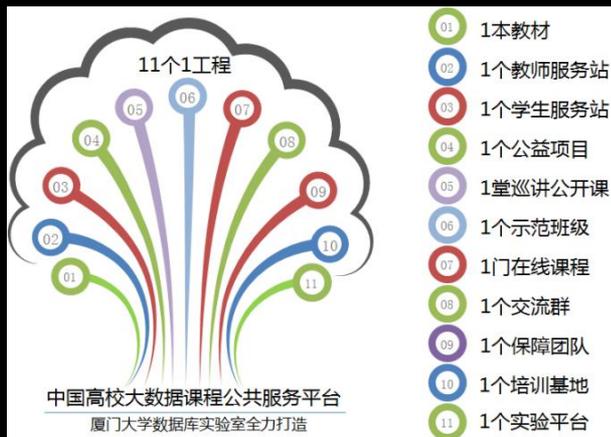




高校大数据课程公共服务平台

打造11大工程

平台每年访问量
超过200万次





高校大数据课程公共服务平台

已经在厦门大学累计举办2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10期大数据师资培训班，为全国200余所高校培养300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1100多名教师加入交流群



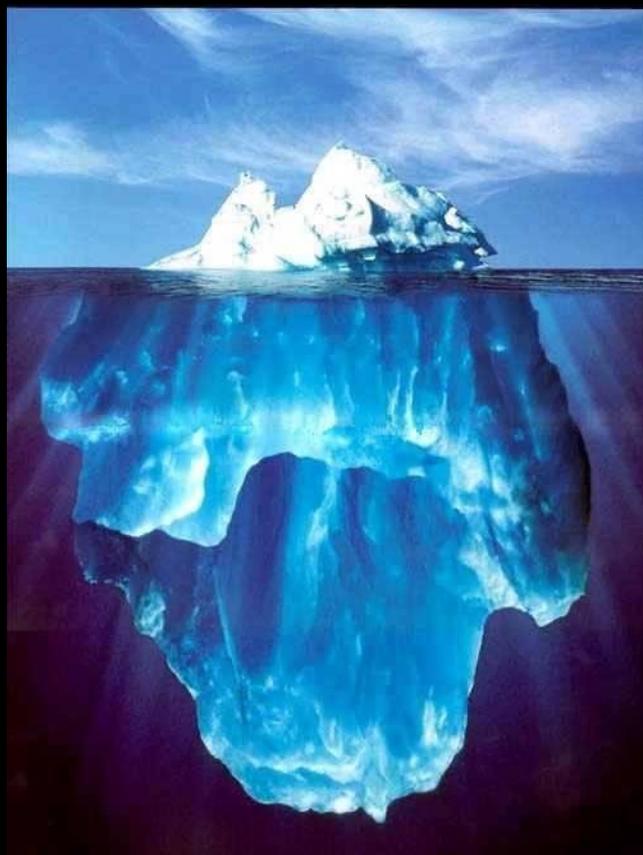
厦门大学、福建师范大学、厦门理工学院、同济大学、浙江财经大学、安徽大学、大连海洋大学、中北大学、河海大学、中山大学、浙江大学、中国农业大学、重庆邮电大学、华中师范大学、武汉理工大学、贵州师范大学、江西财经大学、山西大学、河北经贸大学、东北大学、山东农业大学、海南大学、中国地质大学、武汉大学、中国传媒大学、湖南大学、中国科技大学.....





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
- 导论课程《大数据技术原理与应用》
- 进阶课程《Spark编程基础》
- 大数据实训课程
- 文科大数据通识导论





厦门大学建设的大数据课程体系

文科大数据
通识导论

实训

大数据实习实训案例

进阶

Spark编程基础

导论

大数据技术原理与应用

面向理工科的大数据课程





厦门大学建设的大数据课程体系

厦门大学建设的课程

实训

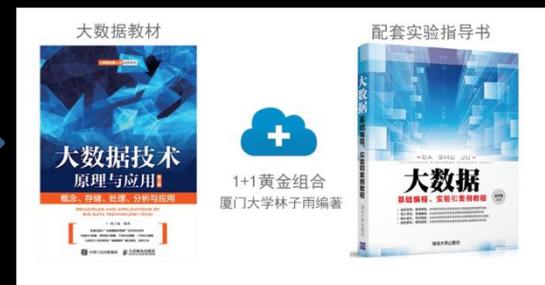
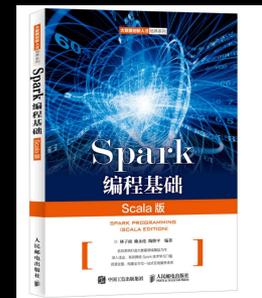
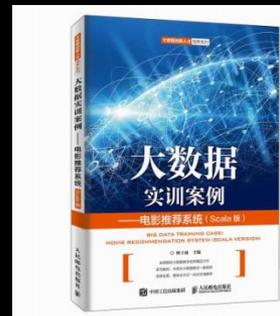
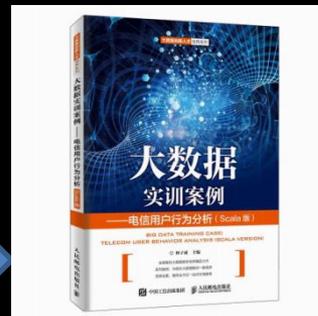
大数据实习实训案例

进阶

Spark编程基础

导论

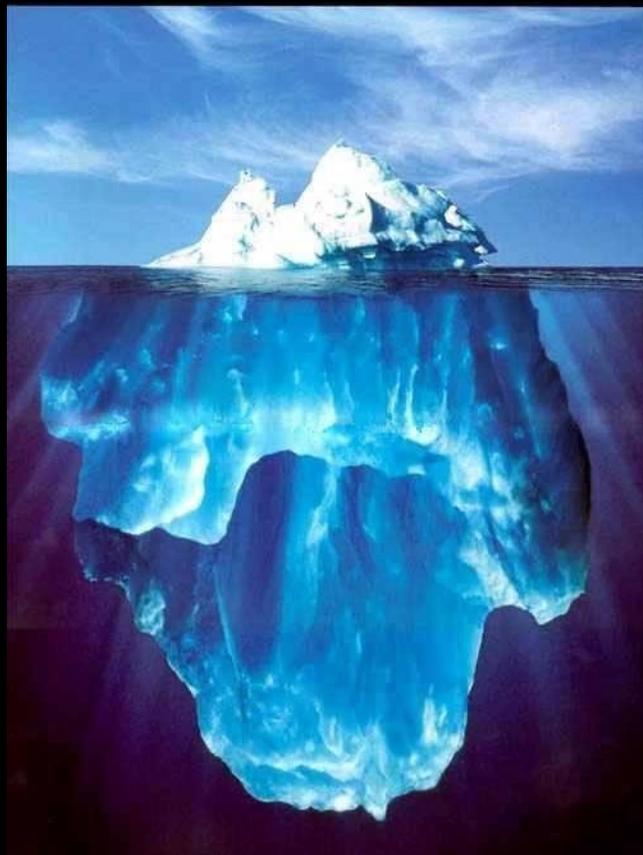
大数据技术原理与应用





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
- 导论课程《大数据技术原理与应用》
- 进阶课程《Spark编程基础》
- 大数据实训课程
- 文科大数据通识导论





导论课程《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

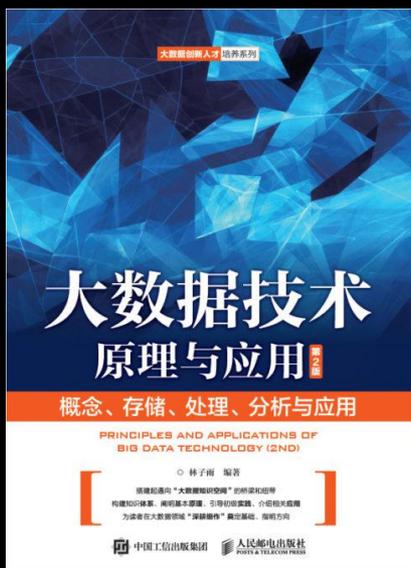




导论课程《大数据技术原理与应用》

课程教材

02



人民邮电出版社
2017年度好书

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

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





导论课程《大数据技术原理与应用》

课程教材

02

大数据教材



大数据 + 实验指导书

1+1黄金组合

厦门大学林子雨编著

配套实验指导书



全力打造大数据精品教材

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





导论课程《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，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/>





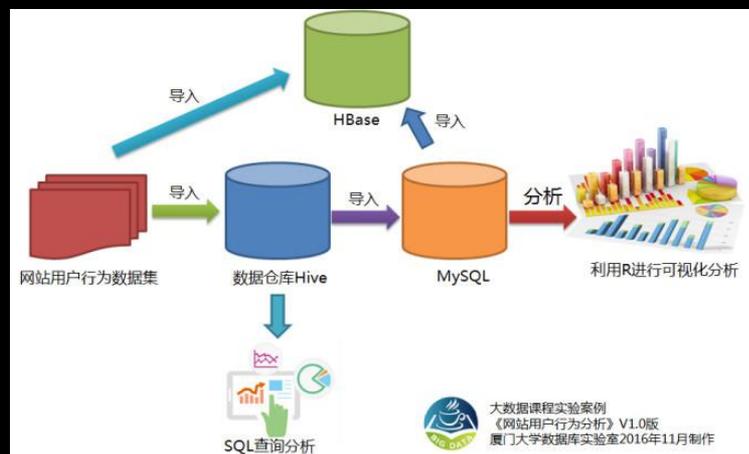
导论课程《大数据技术原理与应用》

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/>





导论课程《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

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

 **廈門大學**
XIAMEN UNIVERSITY

大数据技术原理与应用

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

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 AND APPLICATION
打开大数据之门，遨游大数据世界

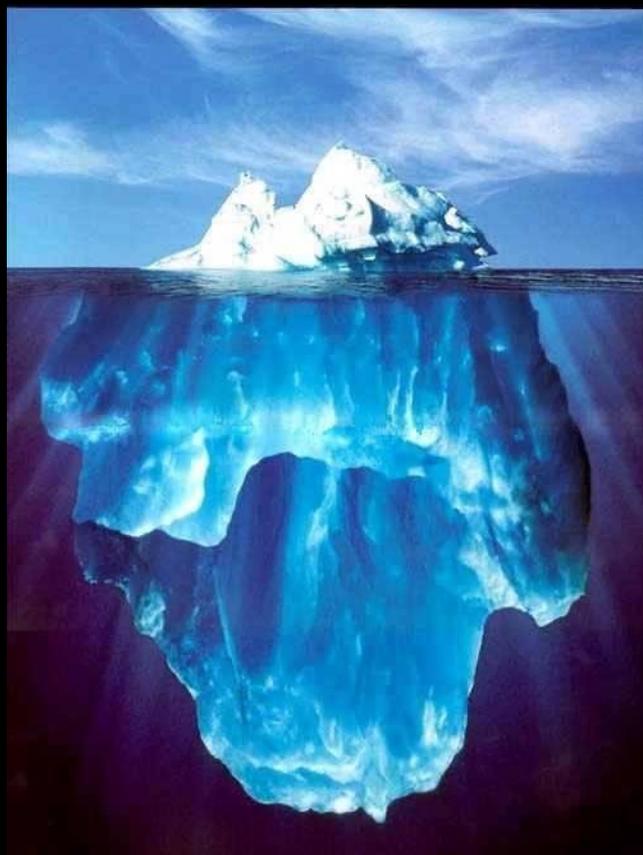
授课视频观看地址：<http://www.icourse163.org/course/XMU-1002335004>





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
- 导论课程《大数据技术原理与应用》
- 进阶课程《Spark编程基础》
- 大数据实训课程
- 文科大数据通识导论





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

6

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

在线免费 《Spark入门教程》

Spark是当前最热门的大数据处理框架，林子雨编著《Spark入门教程》，让初学者零基础零障碍学习Spark。教程采用Scala语言编写Spark应用程序，因此，教程包括Scala入门和Spark入门两个部分的内容



扫一扫手机访问在线教程

免费在线教程：<http://dblab.xmu.edu.cn/blog/spark/>





进阶课程 《Spark编程基础》

课程教材

02

第一部分：快学Scala

第一章 Scala简介

第二章 Scala安装

第三章 Scala基础

声明值和变量、基本数据类型和操作、Range、打印语句、读写文件

第四章 控制结构

if条件表达式、while循环、for循环、数据结构、数组、列表、元组、集、映射、迭代器

第六章 类

第七章 对象

第八章 继承

第九章 特质

第十章 模式匹配

第十一章 函数式编程

函数定义和高阶函数、针对集合的操作、遍历操作、map操作和flatMap操作、filter操作、reduce操作、fold操作、函数式编程实例WordCount





进阶课程 《Spark编程基础》

第二部分：Spark速成 (Spark2.1.0版本)

第1章 Spark的设计与运行原理

Spark简介、Spark运行架构、RDD的设计与运行原理、Spark的部署模式

第2章 Spark的安装与使用

Spark的安装与使用、第一个Spark应用程序：WordCount、使用开发工具IntelliJ IDEA和Eclipse

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

第3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）

第4章 Spark SQL

Spark SQL简介、DataFrame与RDD的区别、DataFrame的创建、从RDD转换得到DataFrame、读取和保存数据（读写Parquet、通过JDBC连接数据库、连接Hive读写数据）

第5章 Spark Streaming

流计算简介、Spark Streaming简介、DStream操作（DStream操作概述、输入源[文件流、套接字流、RDD队列流、Apache Kafka、Apache Flume]、转换操作、输出操作）

第6章 Spark MLlib

Spark MLlib简介、机器学习工作流（机器学习工作流、构建一个机器学习工作流、特征抽取、转化和选择[TF-IDF、Word2Vec、CountVectorizer、标签和索引的转化、卡方选择器]）、分类与回归（逻辑斯蒂回归分类器、决策树分类器）、聚类算法（KMeans聚类算法、高斯混合模型（GMM）聚类算法）、推荐算法（协同过滤算法）





进阶课程 《Spark编程基础》

课程教材

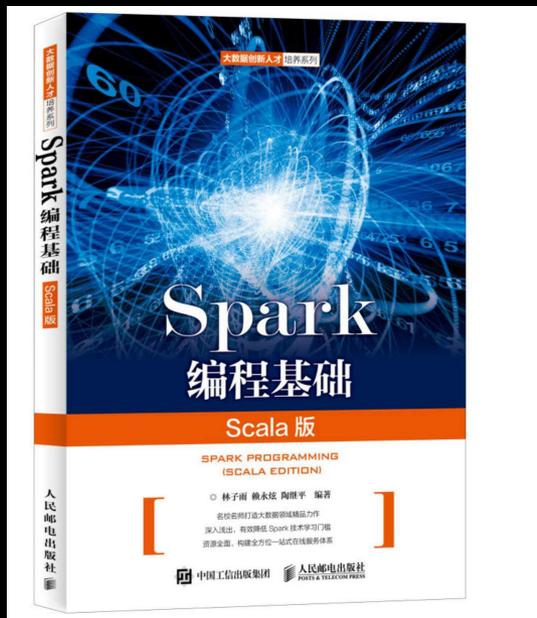
02

《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://dbllab.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

本教材预计在2019年8月上市销售，由人民邮电出版社出版发行





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

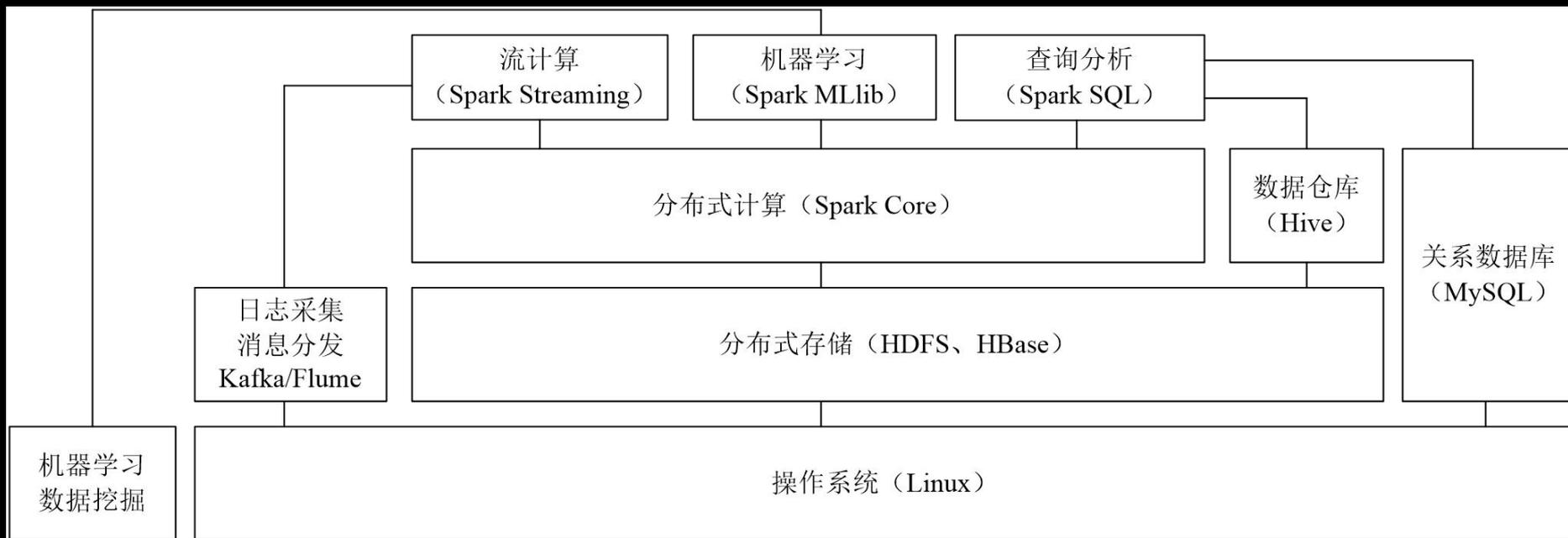




进阶课程 《Spark编程基础》

交叉知识

04





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





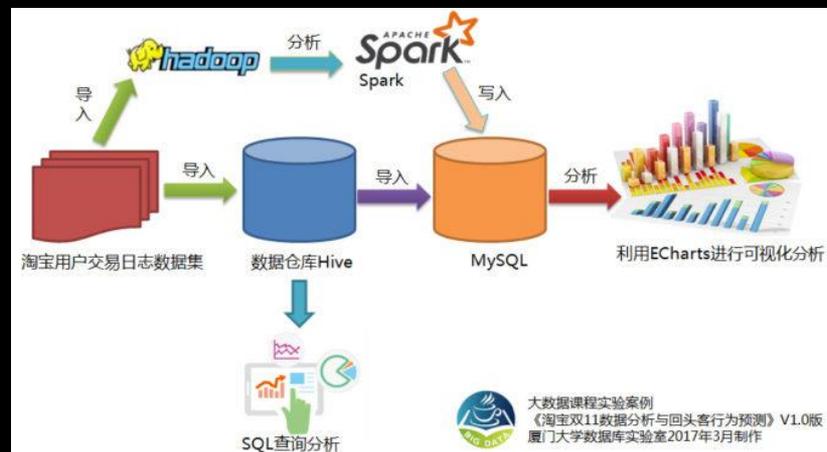
进阶课程 《Spark编程基础》

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://dbl原因.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>

Java

Scala

REPL

RDD

Spark编程基础

林子雨 | 厦门大学计算机科学系

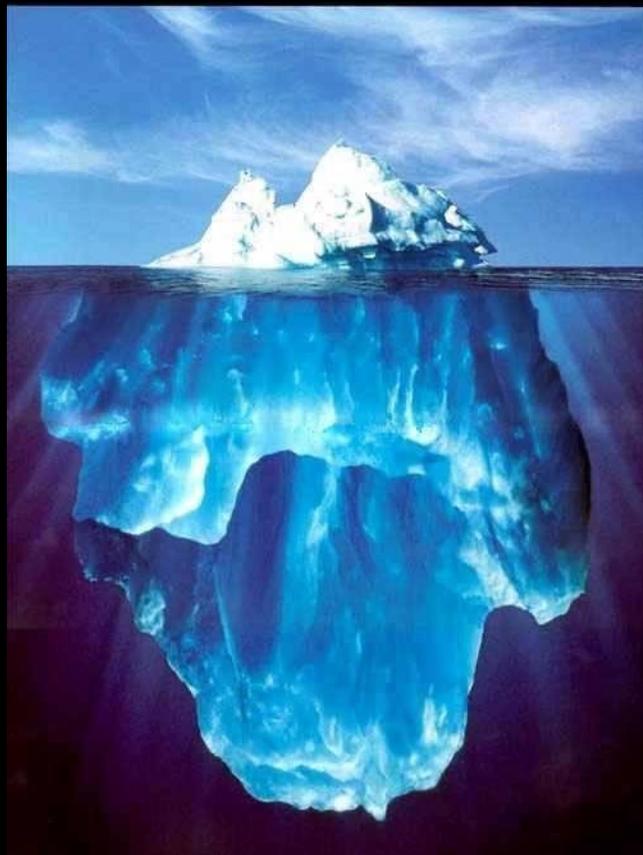
欢迎进入“中国大学MOOC”平台观看课程视频





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
- 导论课程《大数据技术原理与应用》
- 进阶课程《Spark编程基础》
- **大数据实训课程**
- 文科大数据通识导论





大数据实训课程

基本信息

01

目的、任务和要求

02

课程设计内容

03

04

课程设计的
教学方法

05

课程教材

06

授课过程

07

实训成果文档

08

教学管理平台

09

大数据课程
公共服务平台

10

学生作品





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 助理教授
- 6.课程设计学时（周数）：4周/84课时
- 7.实训教材：林子雨编著《高校大数据实训课程系列案例——电影推荐系统》



一、基本信息



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。



一、基本信息

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。





一、基本信息



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





一、基本信息

11. 大数据课程公共服务平台

厦门大学数据库实验室建设的大数据课程公共服务平台，为本次实训课程提供了强有力的支持。学生在实训课程全程可以在线自由免费访问平台上的大量大数据教学资源，解决实训课程的相关大数据技术难题。





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；





二、课程设计的目的、任务与要求

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。





三、课程设计内容

课程设计题目：基于协同过滤算法的电影推荐系统

设计一个基于协同过滤算法的电影推荐算法，采用Scala语言编写Spark MLlib程序实现该算法，利用已有的训练数据集对算法进行训练得到电影推荐模型，并用测试数据集进行电影推荐，最后，将推荐结果进行可视化展现。

基本需求包括：

- (1) 数据ETL：可以把保存在文本文件中的数据集，通过ETL工具，加载到分布式文件系统HDFS中。
- (2) 编写Spark程序实现电影推荐功能：采用Scala语言，编写Spark MLlib程序，对HDFS中的数据进行分析，实现电影推荐算法；
- (3) 可视化：采用可视化技术，把分析结果展现给用户。



四、课程设计的教学方法

课程设计采用“技术教学 + 项目实践”的形式，以技术教学作为项目实践的准备和基础，以项目实践强化技术教学的效果，并训练学生的团队协作、交流沟通以及实施商业项目开发的能力。

技术教学环节采用的教学方法包括：

- 理论讲解：讲解技术原理与编程技巧；
- 案例实践：设计编程案例，让学生动手实现；
- 案例讲解与点评：讲解案例的实现方法，并针对学生在实现过程中产生的问题进行点评。

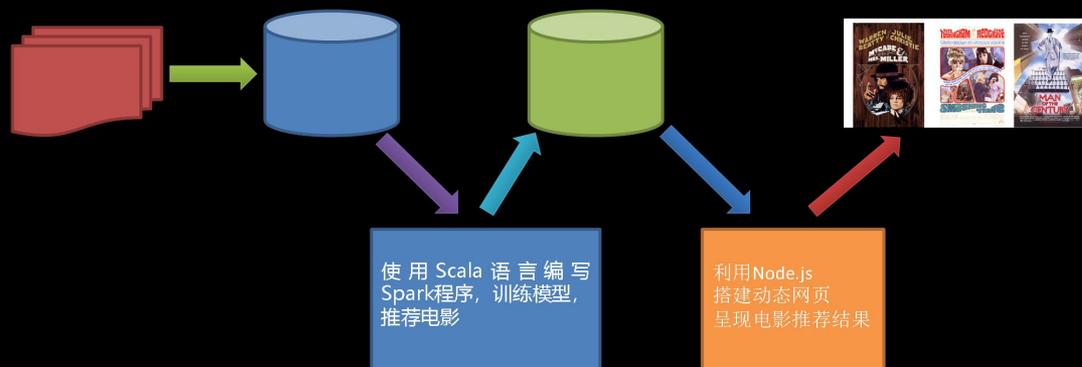
项目实践环节采用的教学方法包括：

- 方法讲解：讲解开发计划、需求分析和系统设计的实施以及文档撰写方法；
- 案例实践：让学生分组协作，完成项目案例的开发文档撰写；
- 反馈修改：对学生提交的开发文档进行评阅，并即时反馈问题和修改意见；
- 指导开发：指导学生针对设计方案进行项目开发。



五、课程教材

林子雨编著《高校大数据实训课程系列案例——电影推荐系统》



系列案例教材，其中两本已经于2019年5月出版发行

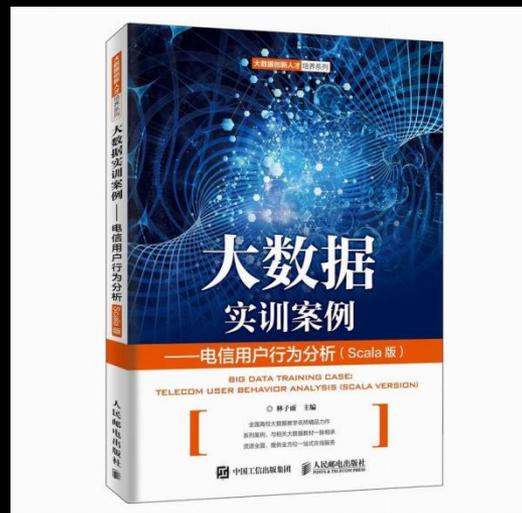
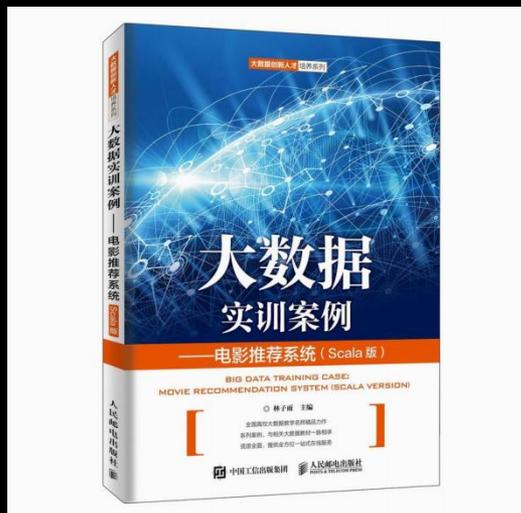
- 《电信用户行为分析》（已出版）
- 《电影推荐系统》（已出版）
- 《实时日志流处理分析》
- 《微博用户情感分析》
- 《互联网广告预测分析》
- 《网站日志处理分析》





五、课程教材

林子雨编著《高校大数据实训课程系列案例》



系列案例教材，其中两本已经于2019年5月出版发行

- 《电信用户行为分析》（已出版）
- 《电影推荐系统》（已出版）
- 《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》





六、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





六、授课过程

表一 (续)

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





六、授课过程

表一 (续)

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





七、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





八、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





八、教学管理平台

授课过程管理

教学过程一体化管理平台 三 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	作业截止日期	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				
	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				
	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	文獻综述	2018-09-17	任务完成情况	未启用评审	课堂记录	编辑 删除
团队						ETL工具Kettle的安装和使用方法	2018-09-17	任务完成情况				
	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				
	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	利用Node.js实现词频统计网页应用程序	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	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
团队						9-12小组会议纪要	2018-09-18	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)	团队	使用Spark SQL读写MySQL数据库	2018-09-18	任务完成情况	未启用评审	课堂记录	编辑 删除
学员						9-14小组会议纪要	2018-09-17	任务完成情况				
	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	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
	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	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
	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	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
	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	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
						无	无	无		查看评审结果	课堂记录	编辑 删除

汇总 总课时: 61





八、教学管理平台

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

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	林浚希	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





八、教学管理平台

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

团队名: 团队5 下载所有附件

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6 下载所有附件

陈文娟 组长 1507052308	苏伟毅 组员 1507052134 设为组长	李家源 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7 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
---------------------	------------------------------	------------------------------	------------------------------	-----------------------------

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
---------------------	------------------------------	------------------------------	------------------------------	-----------------------------





八、教学管理平台

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
							林雨婷		75.30
							龚志松		75.30
							许志峰		75.30
							林诚希		75.30
2	团队2	220.00	44.00	726.00	36.30	80.30	郑雷梅		80.30
							陈颖颖		80.30
							王依婷		80.30
							何李杰		80.30
							吴熙		80.30





八、教学管理平台

教师上传课件
学生上传作业
学生下载作业
教师下载作业



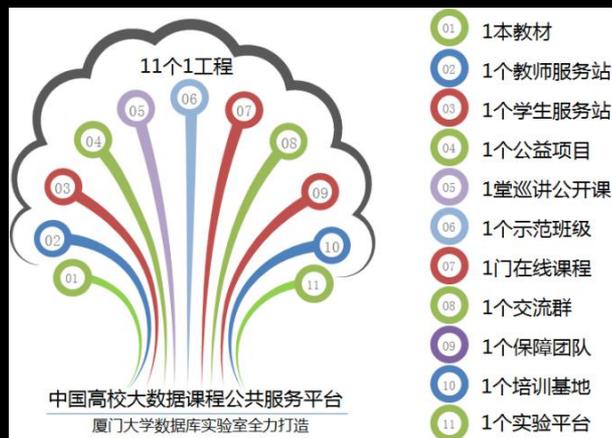
课件上传	
文档类型	附件
作业题目	作业1-Linux系统安装、常用命令以及相关软件用法.docx
	作业2-安装Hadoop.docx
	作业3-HDFS常用Shell命令.docx
	作业4-Spark环境搭建及程序编译运行.docx
	作业5-Scala安装和初级编程实践.docx
	作业6-RDD编程初级实践.docx
	作业7-使用IntelliJ IDEA开发Spark应用程序.docx
	作业8-MySQL数据库的安装和使用方法.docx
	作业9-ETL工具Kettle的安装和使用方法.docx
	作业10-在Spark MLlib中使用ALS算法序.docx
	作业11-利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





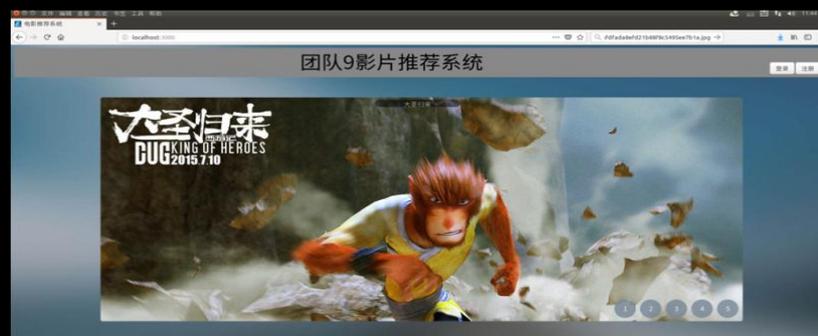
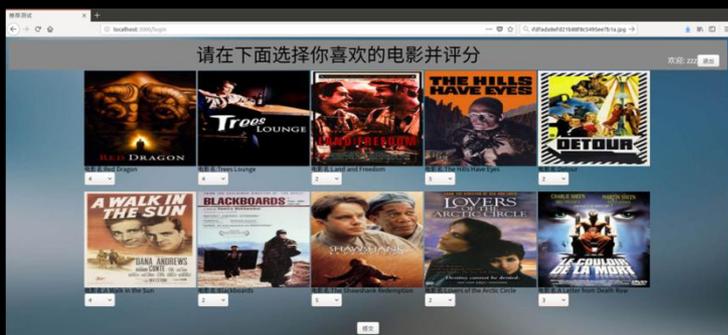
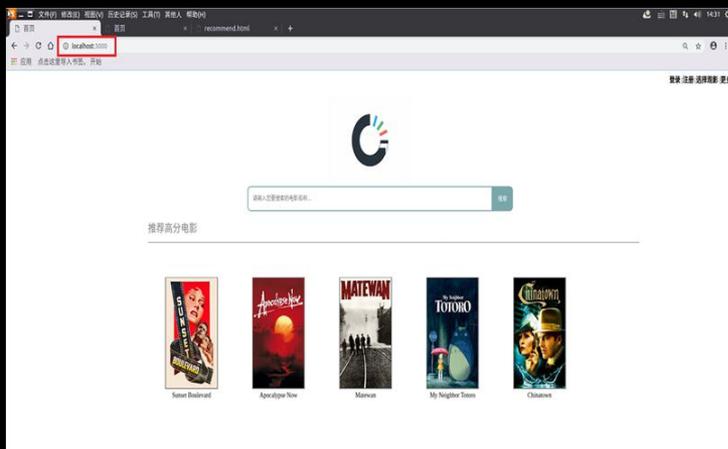
九、大数据课程公共服务平台

厦门大学数据库实验室建设的大数据课程公共服务平台，为本次实训课程提供了强有力的支持。学生在实训课程全程可以在线自由免费访问平台上的大量大数据教学资源，解决实训课程的相关大数据技术难题





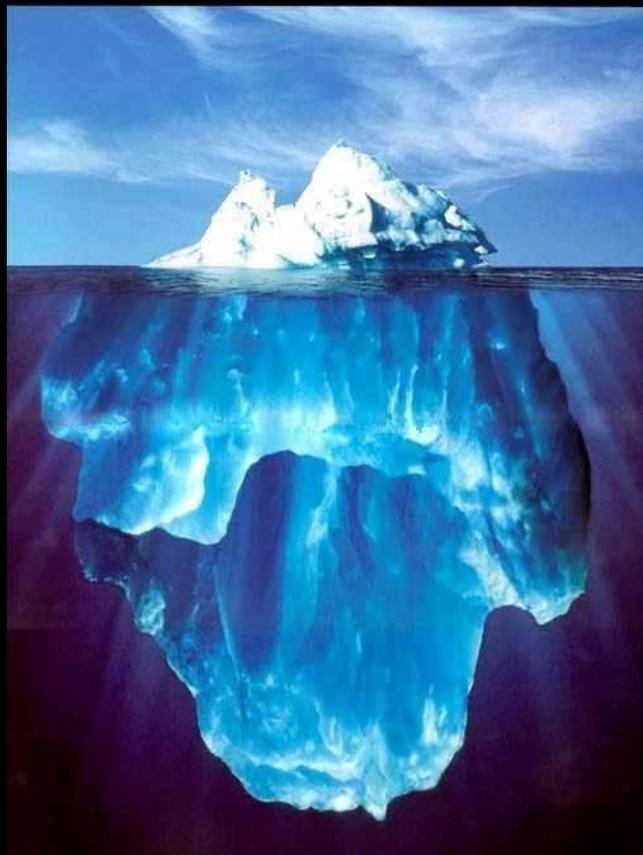
十、学生作品





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
- 导论课程《大数据技术原理与应用》
- 进阶课程《Spark编程基础》
- 大数据实训课程
- 文科大数据通识导论





文科大数据通识导论

课程定位

数据意识

数据伦理

04



01



02



数据思维

03



数据能力





文科大数据通识导论

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





文科大数据通识导论

课程内容

- 大数据时代已经全面开启，带来了信息技术发展的巨大变革，并深刻影响着社会生产和人民生活的方方面面
- 了解大数据概念、具备大数据思维，是新时代对人才的新要求
- 本课程高屋建瓴探讨大数据，内容深入浅出，简单易懂，适合非计算机专业（尤其是文科专业）学生学习。课程内容包括大数据概述，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，大数据交易和大数据治理等





文科大数据通识导论

课程教材

将于2019年7月上市销售

大数据导论

数据思维、数据能力和数据伦理
(通识课版)

林子雨 编著

高等教育出版社





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，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。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年福建省高等教育成果二等奖和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。

