



欢迎参加

第 1 届全国高校大数据教学研讨会

2017年5月12日-13日 厦门大学



第1届全国高校大数据教学研讨会 (BDTS2017)
大会特邀报告



2017年5月12日至13日，第1届全国高校大数据教学研讨会（BDTS2017）在厦门大学科艺中心音乐厅隆重举行。本届研讨会由教育部高等学校计算机类专业教育指导委员会主办，厦门大学、厦门理工学院、贵州师范大学、人民邮电出版社联合承办，旨在搭建专业的大数据教学交流平台，汇聚全国高校大数据教学精英力量，共同探讨大数据专业和课程体系建设，为加快推进全国高校大数据教学发展贡献力量。来自全国300多所院校的400余名教师参加了本次研讨会。

厦门大学谭绍滨校长助理、人民邮电出版社教育中心营销部肖稳副主任，北京大学、中国科学院、厦门大学、华东师范大学、同济大学等重点院校的6位大数据教学知名专家，以及来自国内知名大数据企业的3名业界专家出席会议并做特邀大会报告。厦门大学林子雨助理教授主持会议。

更多内容请访问大会官网：<http://dbl原因lab.xmu.edu.cn/post/bigdata2017/>



厦门大学 林子雨 助理教授 在做大会特邀报告

大会报告全文网页版：<http://dbllab.xmu.edu.cn/post/8284/>

第1届全国高校大数据教学研讨会 (BDTS2017) 特邀报告



高校大数据课程 资源建设和教学经验分享

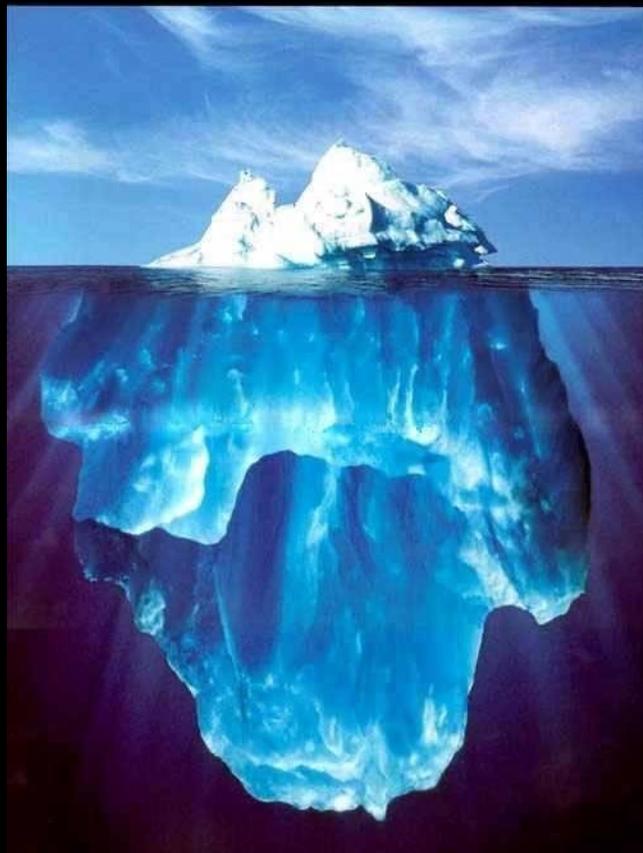
厦门大学 林子雨 博士/助理教授
ziyulin@xmu.edu.cn

2017年5月13日 中国.厦门



内容提要

- 大数据课程公共服务平台
- 大数据技术原理与应用课程建设经验
- 大数据处理技术Spark课程建设经验





大数据课程公共服务平台



中国高校大数据课程 公共服务平台





大数据课程公共服务平台

建设周期
四年 (2013-2017)

投入资金
100万+

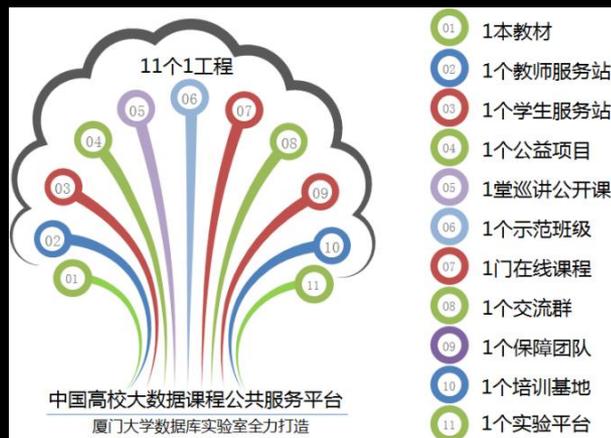




大数据课程公共服务平台

打造11大工程

平台每年访问量
超过100万次





大数据课程公共服务平台

1份精美的平台宣传册





大数据课程公共服务平台

中国高校大数据课程公共服务平台

3分钟FLASH 动画宣传片





大数据课程公共服务平台



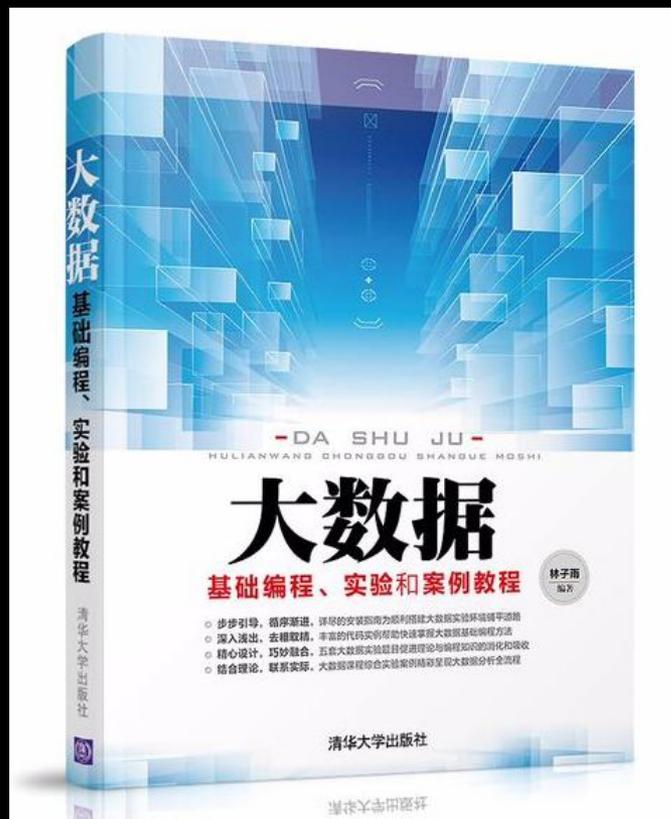
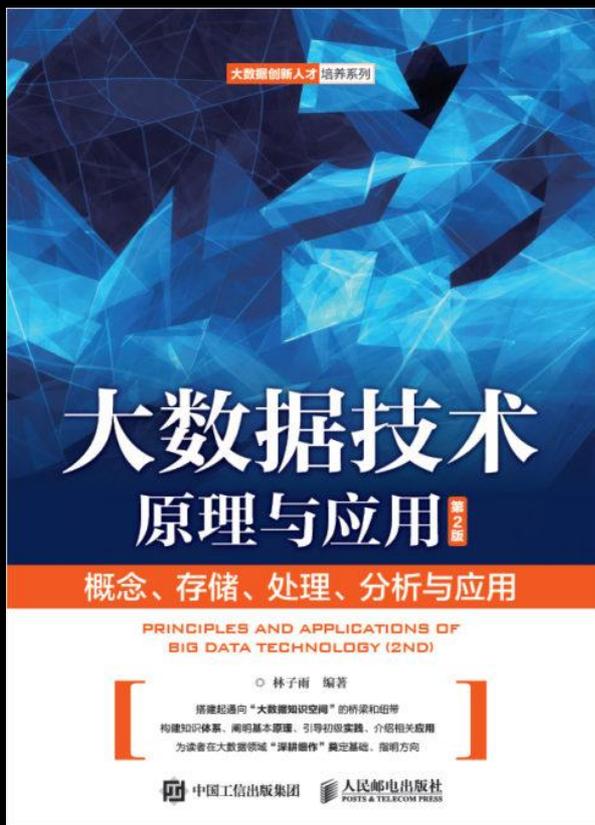
中国高校大数据课程教师培训交流基地

制作5分钟培训基地宣传片





大数据课程公共服务平台





大数据课程公共服务平台

开展大数据公开课全国高校巡讲计划

巡讲8个省、10所高校、1个科研院所、12场公开课，累计听众人数超过1500人





大数据课程公共服务平台

全力打造示范班级

- 大数据技术基础2013班级
- 大数据技术基础2016班级
- 大数据处理技术Spark2017班级
- 大数据技术原理与应用2017班级





大数据课程公共服务平台

入门级大数据在线课程《大数据技术原理与应用》，开辟大数据课程网络讲授平台





大数据课程公共服务平台

- 2016年3月28日林子雨主讲《大数据技术原理与应用》在网易云课堂正式上线
- 2017年5月12日学习人数超过3万2千人
- 99%为5星级最高评价
- 长期稳居热门排行榜第一名
- 被众多网友评为“经典课程”





大数据课程公共服务平台

建设1个大数据课程教师培训交流基地



中国高校大数据课程教师培训交流班（2016暑期班）

时间：2016.8.1-8.7 地点：中国高校大数据课程教师培训交流基地
位于中国国际信息技术（福建）产业园

专注奉献决定卓越品质

同心同德汇聚菁英力量

厦门大学 林子雨 主讲

- 国内一流的大数据课程教师培训交流基地
- 国内高校大数据教学知名团队
- 投资百亿的福建省首个大数据产业园区
- 国内领先的大数据实验平台供应商
- 助力高校大数据教学发展的共同理念
- 具有一线丰富大数据教学经验的培训讲师
- 为高校教师顺利开课量身定制的培训内容
- 低收费高品质的30人小班制教学
- 细致入微的全程贴心周到服务
- 提供全方位一站式免费教学资源





大数据课程公共服务平台

累计开展4期公益培训

2016年8月1日-7日
第1期



2016年8月14日-20日
第2期



2016年9月2日-4日
第3期



2017年1月16日-22日
第4期





大数据课程公共服务平台

举办4次大数据课程教师研讨会





大数据课程公共服务平台

承办3次全国高校教学研讨会



第五届全国高等学校计算机程序设计课程研讨会参会代表合影

2015年12月4日-6日 中国·厦门





大数据课程公共服务平台

承办NDBC2016全国高校大数据教学论坛





大数据课程公共服务平台



欢迎参加

第 1 届全国高校大数据教学研讨会

2017年5月12日-13日 厦门大学





大数据课程公共服务平台

- 2013年9月，发布在线免费电子书《大数据技术基础》
- 2014-2016，发布并持续更新在线教程《大数据软件安装和基础编程指南》
- 2016年10月，发布免费在线教程《Spark入门教程》



厦门大学林子雨



子雨大数据之Spark入门教程

披荆斩棘，在大数据丛林中开辟学习捷径

欢迎选修厦门大学计算机学院研究生课程

《大数据技术基础》

2013全新改版

主讲教师：林子雨
<http://www.cs.xmu.edu.cn/linziyu>

激情·活力 成长·收获



BIGDATA2013

带你一起体验年轻的课堂..... 



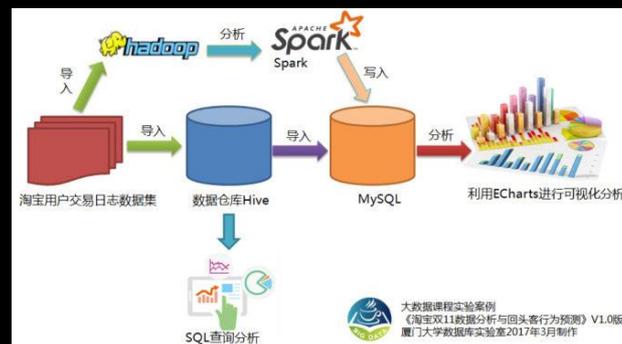


大数据课程公共服务平台

大数据课程实验案例：网站用户购物行为分析

Spark课程综合实验案例：淘宝双11数据分析与预测

Spark课程实验案例：Spark+Kafka构建实时分析Dashboard





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241所高校的349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.....





大数据课程公共服务平台

微信公众号名称：**云谷大数据**

公众号定位：

- 跟踪行业热点
- 汇集百家观点
- 分享教学经验
- 推荐经典书籍



子雨大数据

微信 | 公众平台

跟踪行业热点 汇集百家观点

分享教学经验 推荐经典书籍



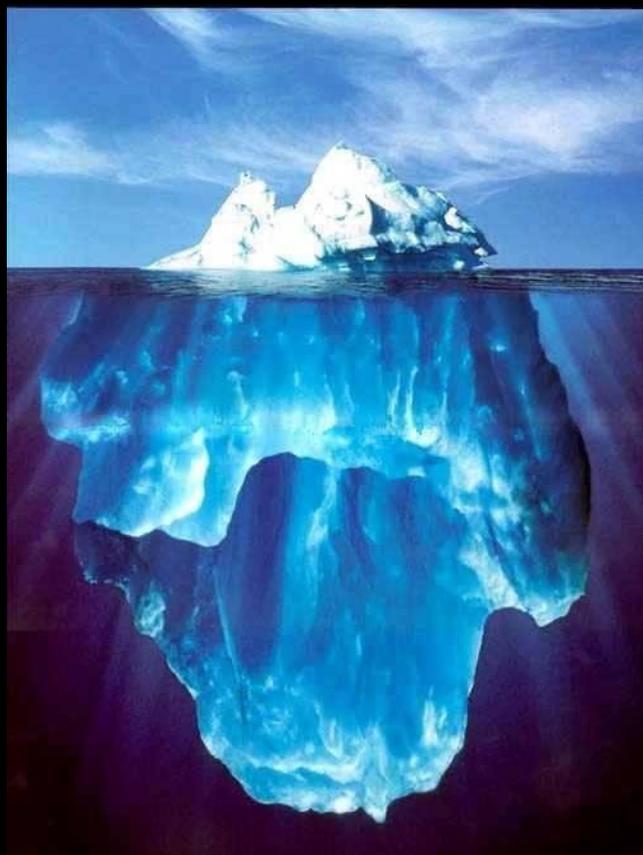
扫一扫关注微信公众号





内容提要

- 大数据课程公共服务平台
- 大数据技术原理与应用课程建设经验
- 大数据处理技术Spark课程建设经验





大数据技术原理与应用课程建设经验

课程定位

01

教材选择

02

课时安排

03

04

实验内容

05

考核方法

06

课程资源





大数据技术原理与应用课程建设经验

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（计算机相关专业）

知识储备：编程、操作系统、数据库

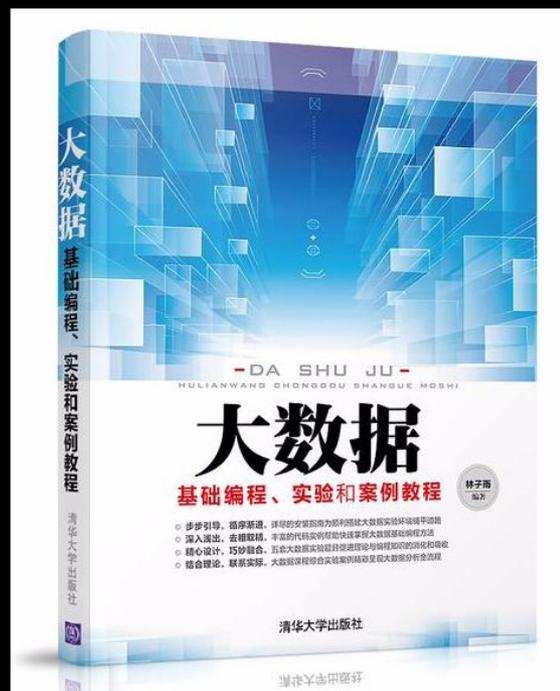
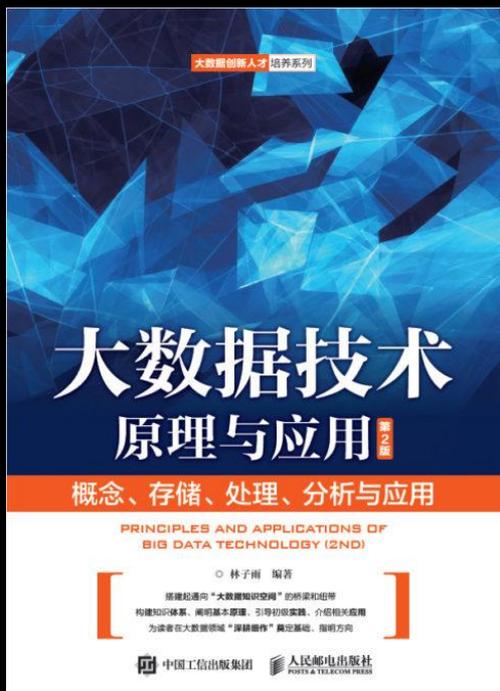




大数据技术原理与应用课程建设经验

教材选择

02





大数据技术原理与应用课程建设经验

教材选择

02

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





大数据技术原理与应用课程建设经验

课时安排

03

章或节	主要内容	学时安排
1	介绍大数据的基本概念和应用领域，并阐述大数据、云计算和物联网的相互关系	2
2	介绍大数据处理架构Hadoop	2
3	分布式文件系统HDFS的基本原理和使用方法	4
4	分布式数据库HBase的基本原理和使用方法	4
5	NoSQL数据库的概念和基本原理	2
6	云数据库的概念和基本原理	2
7	分布式并行编程模型MapReduce原理和使用方法	6
8	流计算基本原理	2
9	图计算基本原理	4
10	数据可视化概念、工具和案例	2

11 大数据在互联网领域的典型应用：推荐系统

高校大数据课程资源建设和教学经验分享 | 厦门大学 林子雨



中国高校大数据课程
公共服务平台



大数据技术原理与应用课程建设经验

04

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/>





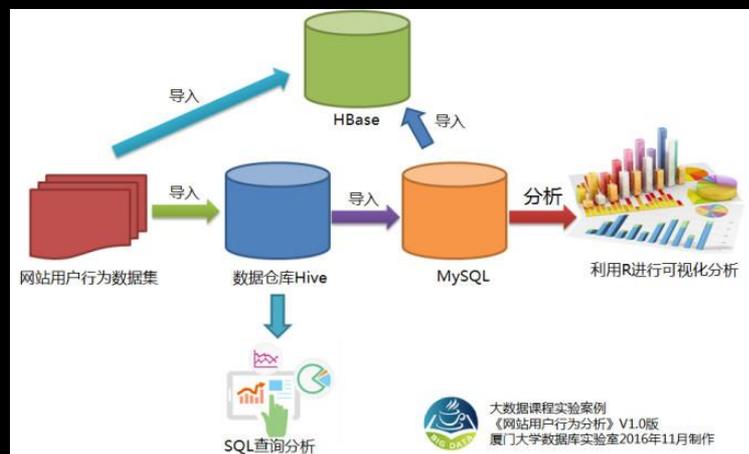
大数据技术原理与应用课程建设经验

04

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/>





大数据技术原理与应用课程建设经验

05

考核方法

- 平时签到考勤10%
- 上机实验报告20%
- 期末大实验20%
- 期末笔试成绩50%





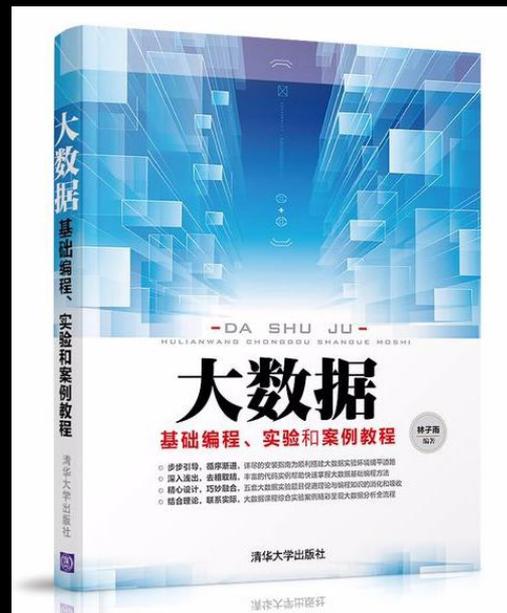
大数据技术原理与应用课程建设经验

06

课程资源

大数据软件安装和编程实践指南

详细学习如何安装运行各种大数据软件以及如何进行初级编程实践，包括Hadoop、HDFS、HBase、MapReduce、Spark、MongoDB等安装、操作、编程指南



访问地址：<http://dmlab.xmu.edu.cn/post/5663/>





大数据技术原理与应用课程建设经验

06

课程资源

《大数据技术原理与应用》备课指南

详细说明了教师如何备课，包括教学大纲、讲义PPT、授课视频、课后习题、上机题目等

访问地址：<http://dbllab.xmu.edu.cn/post/5637/>





大数据技术原理与应用课程建设经验

06

课程资源

《大数据技术原理与应用》授课视频

林子雨主讲的全套大数据课程视频，供老师上课参考。课程内容涵盖大数据、云计算和物联网概念及其相互关系、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、图计算、流计算、基于内存的大数据处理框架Spark、基于Hadoop的数据仓库Hive、大数据在不同领域的应用等

网易云课堂



访问地址：<http://dblab.xmu.edu.cn/post/bigdata-online-course/>





大数据技术原理与应用课程建设经验

06

课程资源

《大数据技术原理与应用》实验指南

用于机房统一上机，包含题目和答案

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

访问地址：<http://dblab.xmu.edu.cn/post/6131/>





大数据技术原理与应用课程建设经验

06

课程资源

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241所高校的349名教师加入交流群



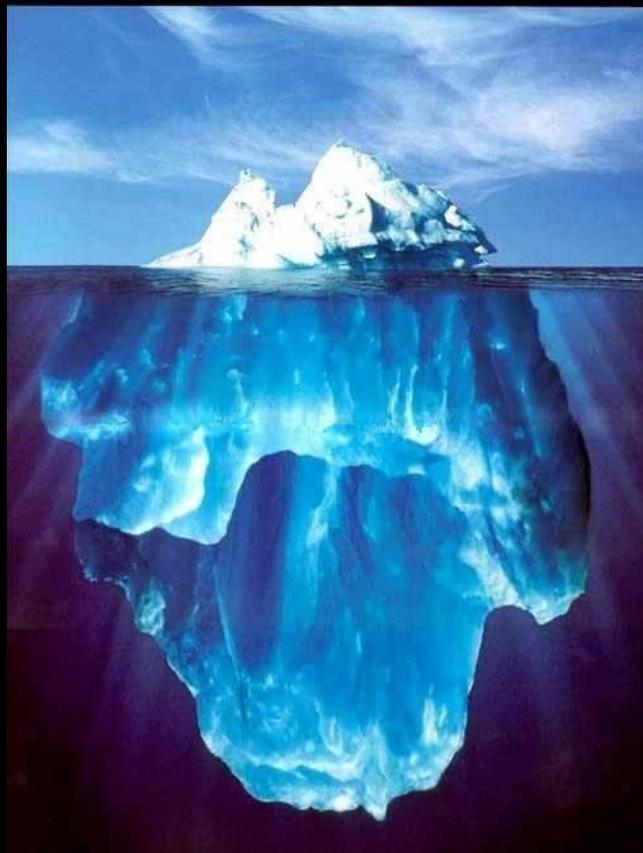
厦门大学、福建师范大学、厦门理工学院、同济大学、浙江财经大学、安徽大学、大连海洋大学、中北大学、河海大学、中山大学、浙江大学、中国农业大学、重庆邮电大学、华中师范大学、武汉理工大学、贵州师范大学、江西财经大学、山西大学、河北经贸大学、东北大学、山东农业大学、海南大学、中国地质大学、武汉大学、中国传媒大学、湖南大学、中国科技大学.....





内容提要

- 大数据课程公共服务平台
- 大数据技术原理与应用课程建设经验
- 大数据处理技术Spark课程建设经验





大数据处理技术Spark课程建设经验

课程定位

01

教材选择

02

课时安排

03

04

实验内容

05

考核方法





大数据处理技术Spark课程建设经验

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





大数据处理技术Spark课程建设经验

教材选择

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/>





大数据处理技术Spark课程建设经验

教材选择

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





大数据处理技术Spark课程建设经验

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、共享变量、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）





大数据处理技术Spark课程建设经验

课时安排

03

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、共享变量、数据读写（文件数据读写、读写HBase数据）	6
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	3
第7章 Spark Streaming	流计算简介、Spark Streaming简介、DStream操作	3
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	6
合计		32





大数据处理技术Spark课程建设经验

04

实验内容

实验1：Linux系统基本命令和Hadoop使用方法

实验2：RDD基本操作

实验3：迭代式算法编程实践

实验4：自定义分区、排序、合并

实验5：利用DataFrame实现数据库的读写

实验6：利用Spark Streaming实现流数据处理

备注：将放入《Spark入门教程》教材一起出版



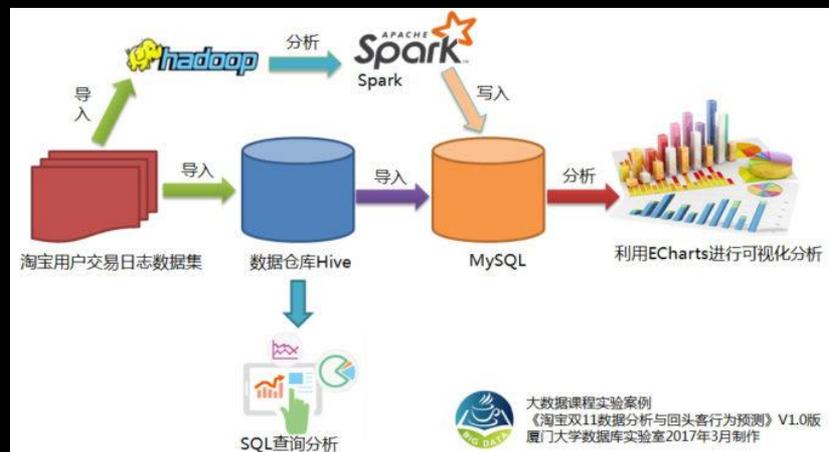
大数据处理技术Spark课程建设经验

04

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://dblab.xmu.edu.cn/post/8116/>





大数据处理技术Spark课程建设经验

04

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/>





大数据处理技术Spark课程建设经验

05

考核方法

- 平时签到考勤10%
- 上机实验报告20%
- 期末大实验20%
- 期末笔试成绩50%





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委会委员，中国计算机学会信息系统专委会委员。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年度厦门大学奖教金（教学类）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)；主持的教学课题包括1项福建省教改课题和1项教育部产学合作育人项目课题；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。编著出版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》并成为畅销书籍；建设了国内高校首个大数据课程公共服务平台，成为国内高校大数据教学知名品牌，平台每年访问量超过100万次；建设了国内高校首个大数据课程教师培训交流基地，已经累计为全国100余所高校培养100多名大数据课程教师；主讲厦门大学本科生课程《数据库系统原理》《大数据技术原理与应用》和研究生课程《分布式数据库》《大数据处理技术Spark》。具有丰富的政府和企业信息化培训经验，厦门大学管理学院EDP中心、浙江大学管理学院EDP中心、厦门大学国际关系学院、中国人民大学等特邀培训讲师，曾先后给中国移动通信集团公司、福州马尾区政府、温州市国资委、厦门市物流协会、福建龙岩卷烟厂、贵阳市政府等多家单位和企业开展信息化培训，累计培训人数达3000人以上。

