



高校大数据课程建设中的关键问题探讨

厦门大学 林子雨 博士/助理教授

ziyulin@xmu.edu.cn

<http://dbllab.xmu.edu.cn>

2017年10月15日 深圳



提纲

- 如何设置大数据专业课程？
- 如何搭建大数据实验平台？
- 如何解决云计算与大数据课程的知识交叉？
- 如何建设优质的大数据教学资源？

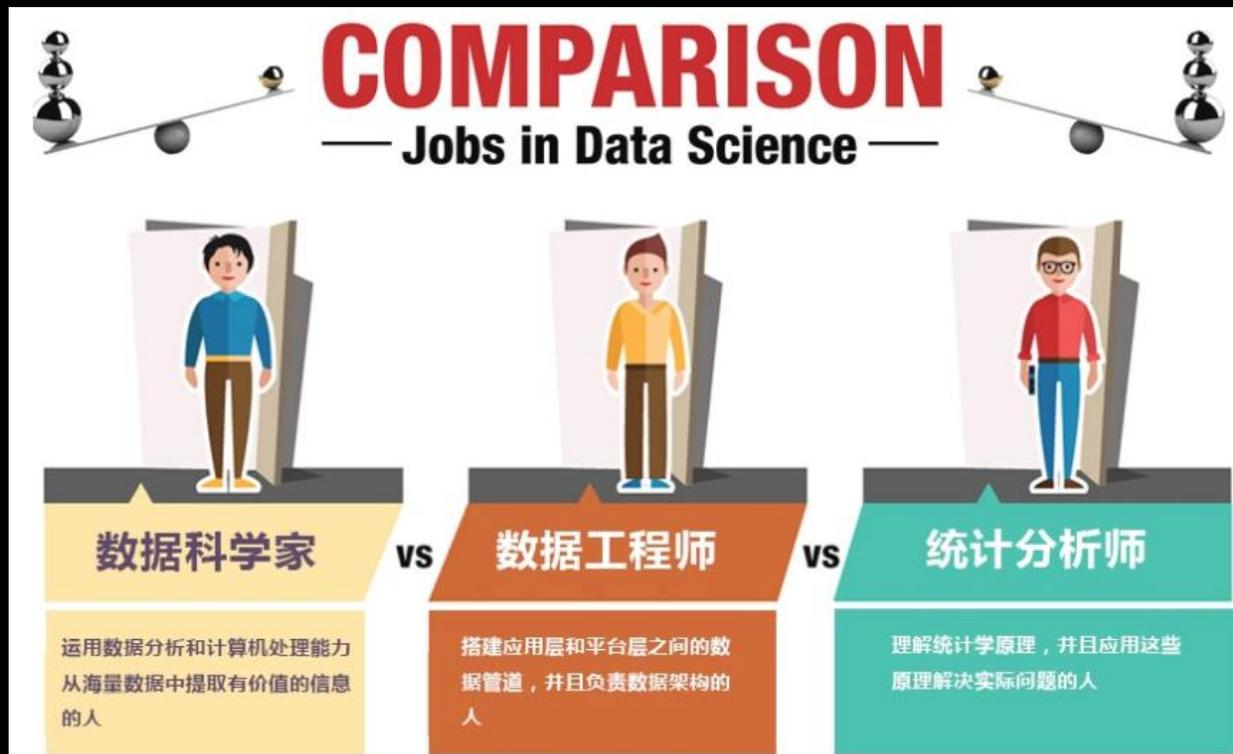




如何设置大数据专业课程?

•培养目标

- 数据科学家
- 数据工程师
- 统计分析师





如何设置大数据专业课程?

职 责 RESPONSIBILITIES

- 1、根据商业需要规划和实施数据分析项目；
- 2、致力于数据挖掘架构、模型标准、数据报告、数据分析方法；
- 3、与利益相关人合作在现有的数据系统中集成数据挖掘结果；
- 4、监督数据挖掘系统性能并实施优化和改进。

数据科学家

- 1、设计、搭建、安装、测试和维护大规模数据管理系统；
- 2、改进数据基础设施、业务流程和数据标准；
- 3、在现有体系架构中集成新的数据管理技术和软件工具；
- 4、开发用户侧软件和数据分析应用程序；

数据工程师

- 1、应用统计学原理和方法解决企业实际问题；
- 2、决定数据采集和数据发现的方法；
- 3、分析和解释数据，承担数据分析师一职；
- 4、提供数据分析结果报告；

数据分析师





如何设置大数据专业课程?

技能 SKILLS

编程、数学、统计学、商业理解、数据可视化、机器学习、关注细节的能力

数据库设计、程序开发、数据采集、数据存储、数据转换、与数据打交道的能力

计算机处理能力、数学、调研、写作能力，分析、建模和解释数据的能力，用简单方式理解数据的敏锐洞察能力

数据科学家

数据工程师

数据分析师

“独角兽”

据美国的一项研究表明，数据科学家来自不同的研究领域，他们当中有28%学习数学、统计学，18%来自工程学科，17%来自计算机科学，16%来自自然科学，还有4-7%的学生来自社会科学、经济学和商科等专业。





如何设置大数据专业课程?



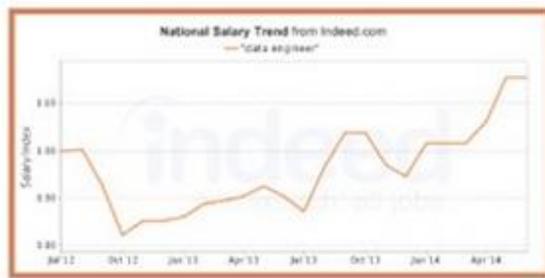


如何设置大数据专业课程?

薪水 SALARY(US)



Mean \$122,000 (as on 17th Oct' 15)



Mean \$115,000 (as on 17th Oct' 15)



Mean \$92,000 (as on 17th Oct' 15)

数据科学家

数据工程师

数据分析师





如何设置大数据专业课程?

职业方向 CAREER PATH

↑

- 首席数据科学家
- 高级数据科学家
- 数据科学家
- 初级数据科学

数据科学家

↑

- 数据架构师
- BI架构师
- 高级数据工程师
- 数据工程师

数据工程师

↑

- 私人统计顾问
- 统计经理
- 高级统计师
- 应用统计学家
- 统计分析

数据分析师





如何设置大数据专业课程?

- 统计学
- 数学
- 计算机

- Java
- C++
- Scala
- Matlab





如何搭建大数据实验平台？

- 单机构建实验环境（虚拟机、8GB内存）
- 实验室多机构建分布式环境
- 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





如何解决云计算与大数据课程的知识交叉？

云计算和大数据的相互关系

- 云计算侧重资源获取的模式
- 大数据侧重于数据的分布式存储和处理

• 云计算的教学重点

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、Openstack、Docker容器、**大数据存储与管理**

• 大数据的教学重点

系统论述大数据的基本概念、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、大数据处理架构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





如何建设优质的大数据教学资源?

- 高校大数据课程建设紧迫需求
- 大数据庞杂的知识体系
- 大数据知识的快速更新
- 大数据教学资源稀缺





建立高校大数据课程公共服务体系的重要性

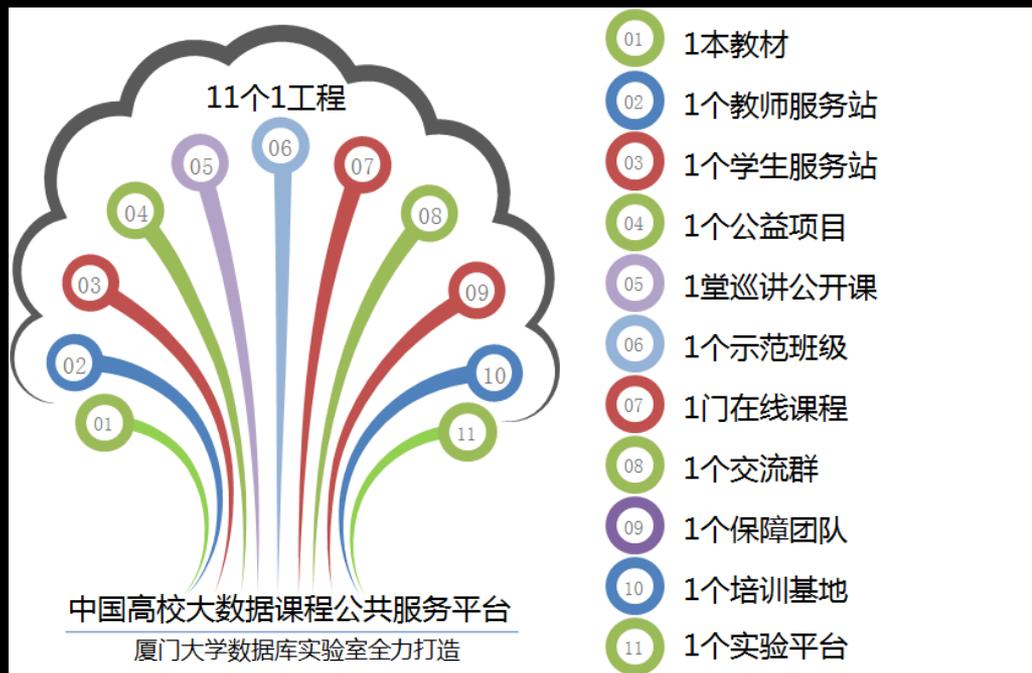
能够解决什么问题？

- 提供丰富的教学资源
- 降低大数据课程开课门槛
- 提升学生学习效果
- 加快高校大数据课程建设进程
- 不断提升高校大数据教学水平





案例：高校大数据课程公共服务平台



访问厦门大学数据库实验室网站，即可**免费**访问所有教学资源





高校开设大数据课程所需7大黄金资源（免费）

- 1.《大数据技术原理与应用》教材
- 2.大数据软件安装和编程实践指南
- 3.教师备课指南
- 4.授课视频
- 5.实验指南
- 6.大数据课程实验案例《网站用户购物行为分析》
7. Spark入门教程



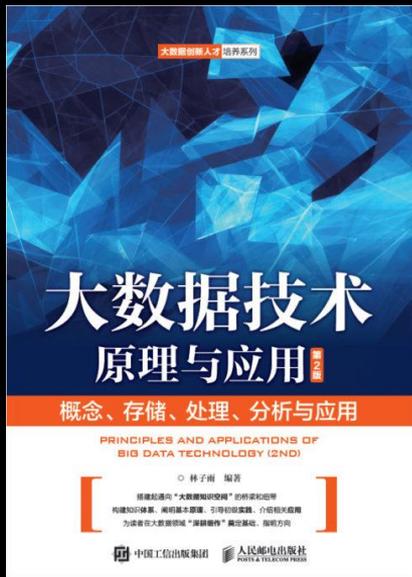
扫一扫免费获取
7大黄金教学资源

免费获取地址：<http://dbllab.xmu.edu.cn/post/7553/>





《大数据技术原理与应用》教材



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

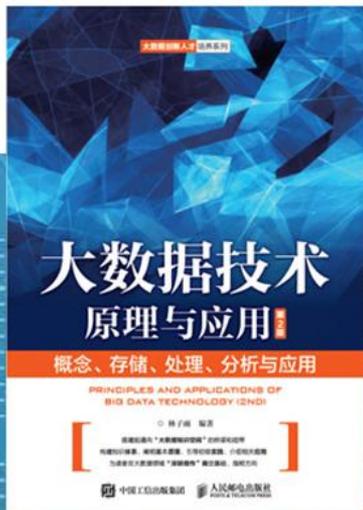
• 为读者在大数据领域“深耕细作”奠定基础、指明方向
• Hadoop、HDFS、HBase、NoSQL、云数据库、MapReduce、流计算、图计算、数据可视化、Spark





《大数据基础编程、实验和案例教程》

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



全力打造大数据精品教材





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
- 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



扫一扫免费访问
大数据软件安装和编程实践指南

免费在线访问地址：<http://dbllab.xmu.edu.cn/post/5663/>





教师备课指南

- 根据林子雨老师实际上课经验撰写
- 详细说明了教师如何备课《大数据技术原理与应用》
- 包括教学大纲、讲义PPT、授课视频、课后习题、上机题目等
- 给出了讲解每个章节时的知识难点和注意事项



扫一扫访问
教师备课指南

免费在线访问地址：<http://dbllab.xmu.edu.cn/post/5637/>





实验指南

- 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



扫一扫访问
实验指南

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





《大数据技术原理与应用》在线课程

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





《大数据技术原理与应用》在线课程

免费在线视频：<http://dbllab.xmu.edu.cn/post/bigdata-online-course/>

□ 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10月14日学习人数超过4万
人6千





《大数据技术原理与应用》在线课程

免费在线视频：<http://dbllab.xmu.edu.cn/post/bigdata-online-course/>

- ❑ 99%为5星级最高评价
- ❑ 15天登顶热门排行榜第一名
- ❑ 被众多网友评为“经典课程”
- ❑ 获评“福建省精品在线开放课程”

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





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



扫一扫免费访问
实验案例

免费访问地址：<http://dblab.xmu.edu.cn/post/7499/>





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》，免费在线访问
- 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

免费在线访问地址：

<http://dbllab.xmu.edu.cn/blog/spark/>

《Spark编程基础》教材将于2018年1月出版发行



扫一扫免费访问
Spark入门教程





Spark教学综合案例

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





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

是一个开放的平台

不断进步提升的平台

热忱欢迎国内高校热爱大数据教学的开拓创新者加入平台

为平台建设添砖加瓦

共同推进中国高校大数据教学事业不断迈上新的台阶





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://dmlab.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，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，2013年度和2017年度厦门大学奖教金获得者。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)，并主持1项福建省教改课题，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，编著并免费网络发布40余万字中国高校第一本闪存数据库研究专著《闪存数据库概念与技术》；主讲厦门大学计算机系本科生课程《数据库系统原理》和研究生课程《分布式数据库》《大数据技术基础》。

