



《大数据导论（通识课版）》

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

温馨提示: 编辑幻灯片母版, 可以修改每页PPT的厦大校徽和底部文字

第8章 数据共享

(PPT版本号: 2019年秋季学期)



扫一扫访问教材官网

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://www.cs.xmu.edu.cn/linziyu>





课程教材

- 林子雨 编著 《大数据导论——数据思维、数据能力和数据伦理（通识课版）》
- 高等教育出版社，2019年11月



提纲

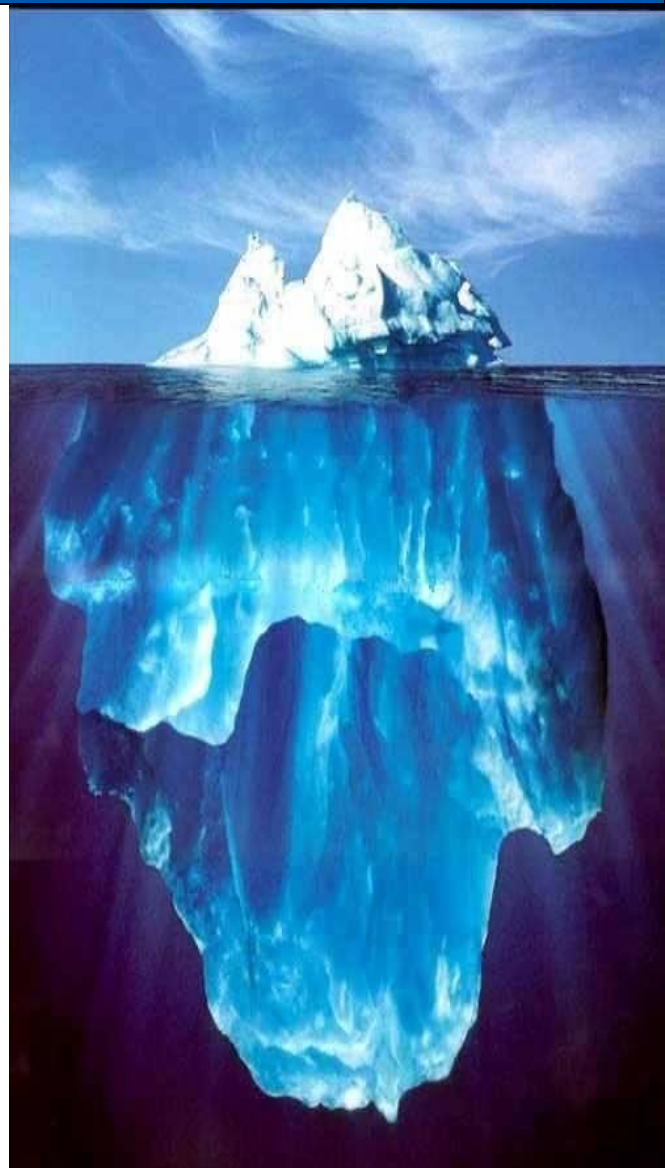
- 8.1 数据孤岛问题
- 8.2 数据孤岛问题产生的原因
- 8.3 消除数据孤岛的重要意义
- 8.4 实现数据共享所面临的挑战
- 8.5 推进数据共享开放的举措
- 8.6 数据共享案例



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





8.1 数据孤岛问题

8.1.1 政府的数据孤岛问题

8.1.2 企业的数据孤岛问题



8.1.1 政府的数据孤岛问题

- 由于各政府部门建设数据库所采用的技术、平台及网络标准不统一，导致政府职能部门之间难以实现数据对接与共享
- 纵向上各级垂直管理部门建设的政府信息系统形成“数据烟囱”，横向上部门间各业务条块则自建系统形成“数据孤岛”，政府公共信息资源的存储彼此独立、管理分散
- 作为政府最重要资产之一的政务数据，因为数据量太大、太散、难以有效融合等问题，严重影响到了数据价值的发挥，大大浪费了各地政府部门在信息化系统建设方面的大量投入



8.1.2 企业的数据孤岛问题

企业管理职能精细划分，信息系统围绕不同的管理阶段和管理职能展开，如客户管理系统、生产系统、销售系统、采购系统、订单系统、仓储系统和财务系统等，所有数据被封存在各系统中，让完整的业务链上孤岛林立，信息的共享、反馈难，数据孤岛问题是企业信息化建设中的最大难题





8.2 数据孤岛问题产生的原因

8.2.1 政府数据孤岛的产生原因

8.2.2 企业数据孤岛的产生原因



8.2.1 政府数据孤岛的产生原因

- 有些政府部门错误地将数据资源等同于一般资源，认为占有就是财富，热衷于搜集，但不愿共享；
- 有些部门只盯着自己的数据服务系统，结果因为数据标准、系统接口等技术原因，无法与外单位、外部门联通；
- 还有些地方，对大数据缺乏顶层设计，导致各条线、各部门固有的本位主义作祟，壁垒林立，数据无法流动



8.2.2 企业数据孤岛的产生原因

- 不同企业之间，属于不同的经营主体，有着各自的利益，彼此之间数据不共享，产生企业之间的数据孤岛，这种情况是比较普遍的情况。
- 企业内部也往往会存在大量数据孤岛，这些数据孤岛的形成主要有两个方面的原因：
 - 以功能为标准的部门划分导致数据孤岛
 - 不同类型、不同版本的信息化管理系统导致数据孤岛





8.3 消除数据孤岛的重要意义

8.3.1 对于政府的意义

8.3.2 对于企业的意义



8.3.1 对于政府的意义

加强政府数据共享开放和大数据服务能力，促进跨领域、跨部门合作，推进数据信息交换，打破部门壁垒，遏制数据孤岛和重复建设，有助于提高行政效率，转变思维观念，推动传统的职能型政府转型为服务型智慧政府。政府数据共享的重要意义表现在以下两个方面：

- 首先，有助于提升资源利用率
- 其次，有助于推动政府转型



8.3.2 对于企业的意义

首先，打通企业内部的数据孤岛，实现所有系统数据互通共享，对建立企业自身的大数据平台和企业信息化建设都有重大意义。

其次，打通企业之间的数据孤岛，实现不同企业的数据共享，有利于企业获得更好的经营发展能力。



8.4 实现数据共享所面临的挑战

8.4.1 在政府层面的挑战

8.4.2 在企业层面的挑战



8.4.1 在政府层面的挑战

A

不愿共享开放

B

不敢共享开放

C

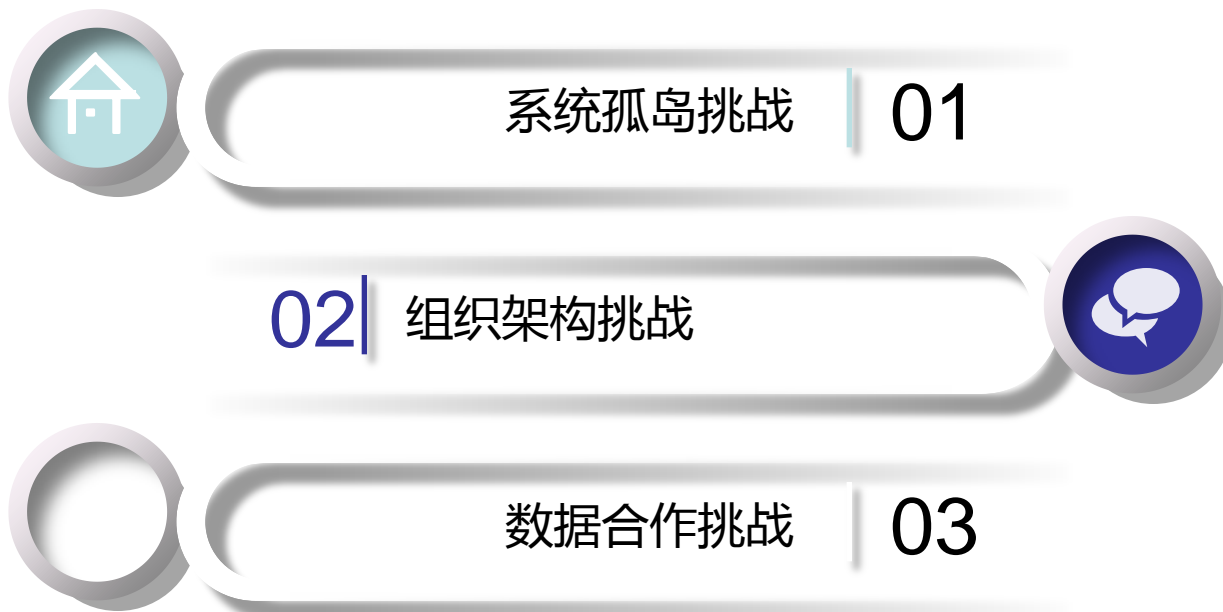
不会共享开放

D

数据中心共享开放作用不强



8.4.2 在企业层面的挑战





8.5 推进数据共享开放的举措

8.5.1 在政府层面的举措

8.5.2 在企业层面的举措



8.5.1 在政府层面的举措

- 积极开放政府数据资源，提高政府职能部门之间和具有不同创新资源的主体之间的数据共享广度，促进区域内形成“数据共享池”
- 要改变政府职能部门“数据孤岛”现象，立足于数据资源的共享互换，设定相对明确的数据标准，实现部门之间的数据对接与共享，推进在制度创新方面的系统集成化，为科技创新提供必要条件
- 要促进准确及时的数据信息传递，提高部门条线管理、“一站式”企业网上办事和政府服务项目“一网通办”的网络信息功能，提高数据质量的可靠性、稳定性与权威性，增加相关信息平台的使用覆盖面，让现存数据“连起来”、“用起来”



8.5.2 在企业层面的举措

在企业内部，破除“数据孤岛”，推进数据融合

在不同企业之间，建立企业数据共享联盟



8.6 数据共享案例

8.6.1 案例1：菜鸟物流

8.6.2 案例2：政府一站式平台——i厦门

8.6.3 案例3：浙江打通政府数据，让群众最多跑一次



8.6.1 案例1：菜鸟物流

2013年阿里巴巴集团联合多方力量联手共建“中国智能物流骨干网”（又称“菜鸟”），计划在8~10年的时间，建立一张能支撑日均300亿元（年度约10万亿元）网络零售额的智能物流骨干网络，支持数千万家新型企业成长发展，让全中国任何一个地区做到24小时内送货必达。





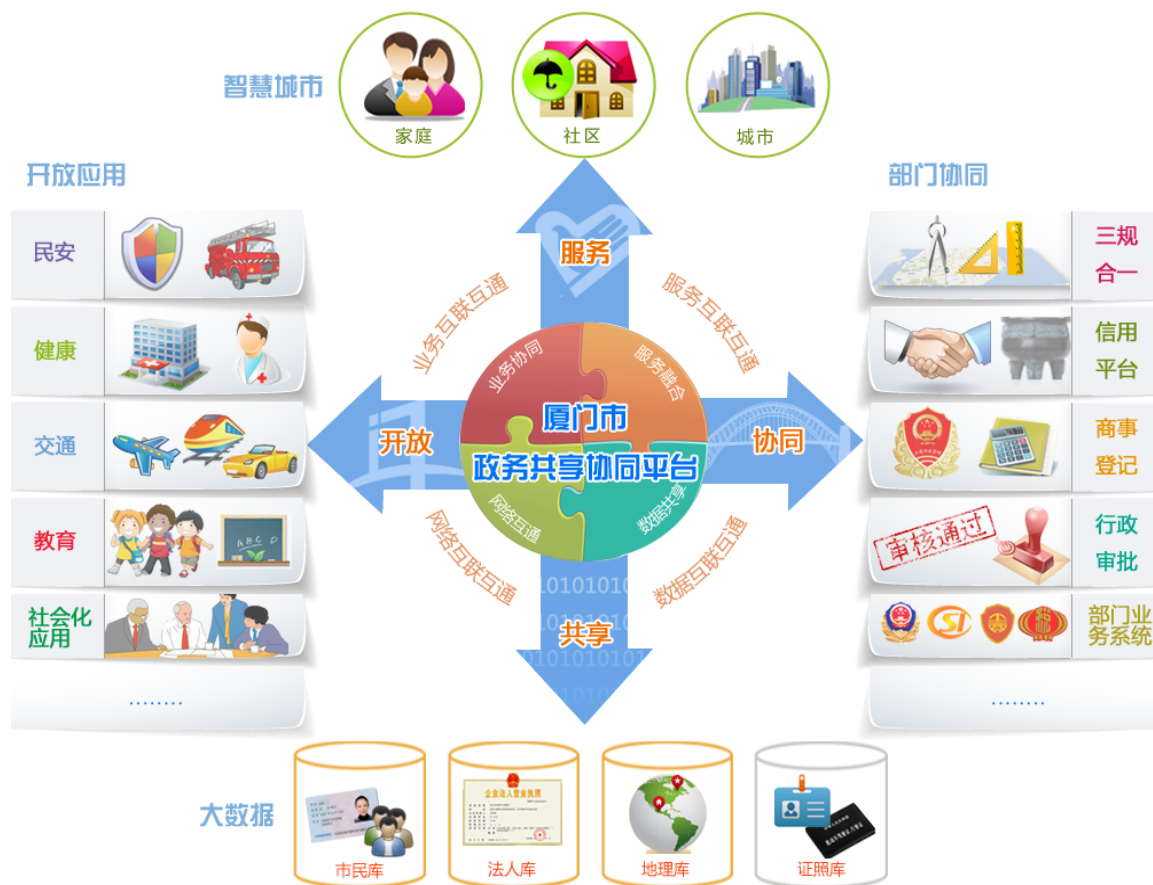
8.6.1 案例1：菜鸟物流

- 菜鸟核心作用的发挥，关键在于其对多方物流数据的有效整合，也就是说，参与菜鸟的相关物流企业，都会把自己企业内部的物流数据（主要是包裹轨迹数据）共享出来，由菜鸟平台对电商和物流数据进行统一整合分析
- 菜鸟平台就相当于中枢协调机构，每个包裹、每家快递从仓库发货就开始接入，揽收、中转、派送信息，整个轨迹都可以显示，这将有利于菜鸟从全局层面帮助快递公司进行运力的统筹调配和规划
- 菜鸟会对后台电商和物流数据进行整合、分析和挖掘，比如根据既往的销售数据来分析预测下一个时期内哪些商品需要提前备货多少量，给予仓储管理商相关的商品陈列建议



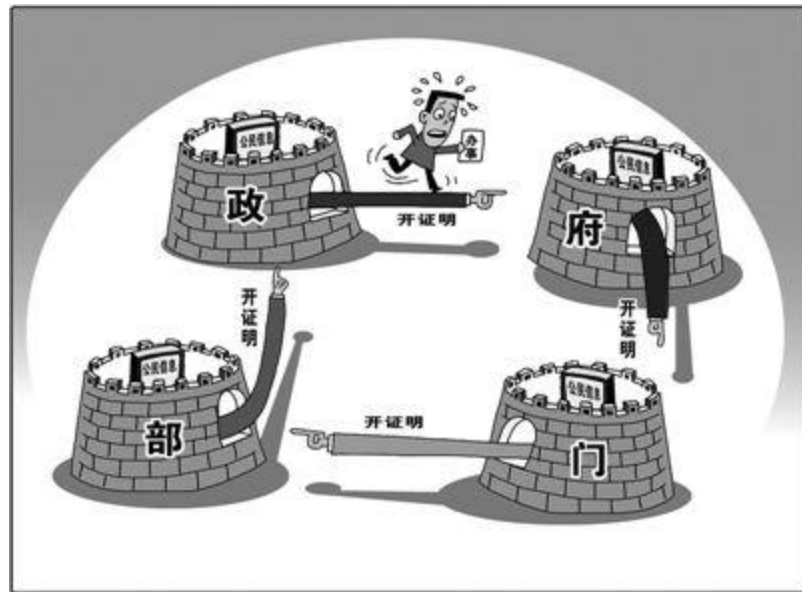
8.6.2 案例2：政府一站式平台——i厦门

“i厦门”一站式惠民服务平台——服务融合





8.6.3案例3：浙江打通政府数据，让群众最多跑一次





8.6.3案例3：浙江打通政府数据，让群众最多跑一次



- (1) 在社会参保单位办理参保登记方面
- (2) 在办理不动产登记证方面
- (3) 在外省就读学生学籍转入方面



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），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、晋江市发展和改革委员会副局长。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年教育部产学研协作育人项目，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》 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



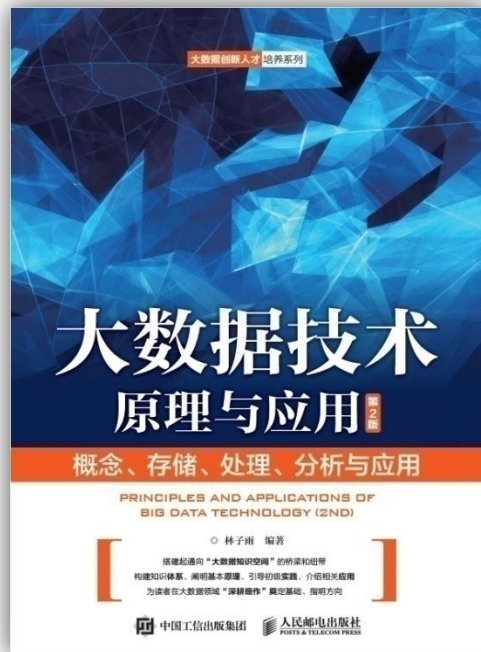
扫一扫访问教材官网

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 定价：59元



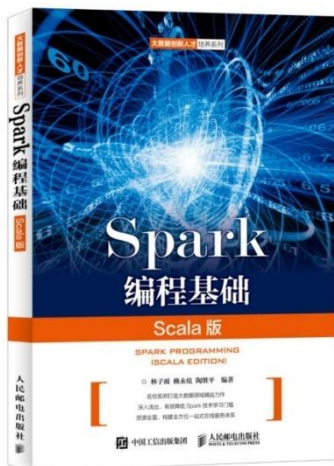
附录E：《Spark编程基础（Scala版）》

《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://dblab.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F：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://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



附录G：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：

《基于协同过滤算法的电影推荐》

《电信用户行为分析》

《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://dblab.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

The background of the slide features several faint, light-blue silhouettes of people. At the top, there are two groups of people standing and holding hands. On the right side, a person is shown in profile, looking towards the center. On the left side, two people are shown in profile, one appearing to be speaking or gesturing towards the other. The overall scene suggests a group of people in a meeting or presentation setting.

Thank You!

Department of Computer Science, Xiamen University, 2019