



《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11章 大数据治理

(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

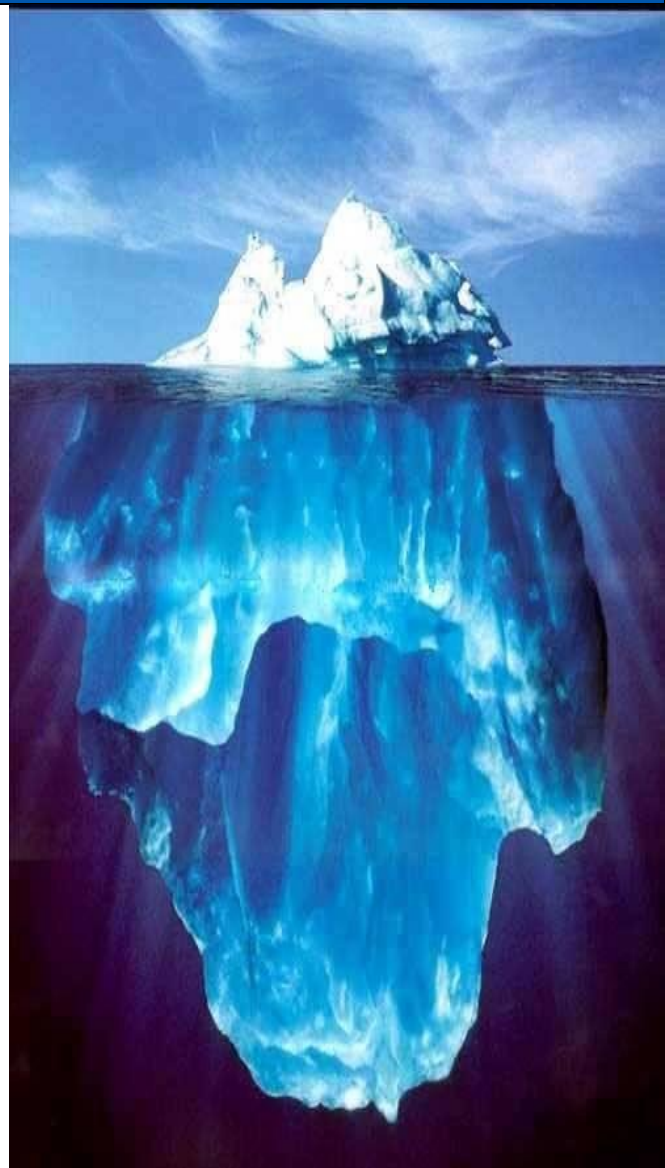
- 11.1 概述
- 11.2 大数据治理要素
- 11.3 大数据治理模型
- 11.4 大数据治理保障机制



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





11.1 概述

11.1.1 为什么需要数据治理

11.1.2 数据治理的基本概念

11.1.3 数据治理与数据管理的关系

11.1.4 大数据治理的基本概念

11.1.5 大数据治理与数据治理的关系

11.1.6 大数据治理的重要意义和作用



11.1.1 为什么需要数据治理

- 企业的信息系统建设烙印着企业规模和信息技术的发展轨迹，普遍存在各系统间数据标准和规范不同、信息相互不通等问题，致使系统的协同性等问题越来越显著
- 数据治理成为解决以上瓶颈的有效手段，为多源、异构、跨界数据应用夯实基础。数据治理可以帮助实现数据资产管理活动始终处于规范、有序、可控的状态，通过多重机制保障基于数据的相关决策是科学的、有效的、前瞻的，以实现资产价值最大化，提升组织的竞争力。



11.1.2 数据治理的基本概念

表“数据治理”代表性观点

机构	定义
DAMA	数据治理是指对数据资产管理行使权力和控制的活动集合（计划、监督和执行）
DGI	数据治理是包含信息相关过程的决策权及责任制的体系，根据基于共识的模型执行，描述谁在何时何种情况下采取什么样的行动、使用什么样的方法
IBM	数据治理是针对数据管理的质量控制规范，它将严密性和纪律性植入企业的数据管理、利用、优化和保护过程中



11.1.2数据治理的基本概念

上述定义较为概括和抽象。为了方便理解，这里从以下4个方面来解释数据治理的概念内涵。

(1)明确数据治理的目标。这里的“目标”是指，在管理数据资产的过程中，确保数据的相关决策始终是正确的、及时、有效和有前瞻性的，确保数据管理活动始终处于规范、有序和可控的状态，确保数据资产得到正确有效的管理，并最终实现数据资产价值的最大化。

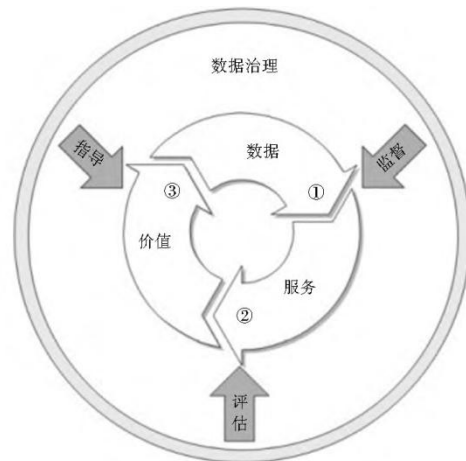
(2)理解数据治理的职能。从决策的角度，数据治理的职能是“决定如何做决定”，因此，数据治理必须回答决策过程中所遇到的问题，即为什么、什么时间、在哪些领域、由谁做决策，以及应该做哪些决策；从具体活动的角度，数据治理的职能是“评估、指导和监督”，即评估数据利益相关者的需求、条件和选择，以达成一致的数据获取和管理的目标，通过优先排序和决策机制来设定数据管理职能的发展方向，然后根据方向和目标来监督数据资产的绩效与是否合规。



11.1.2 数据治理的基本概念

(3) 把握数据治理的核心。数据治理关注的焦点问题是，通过何种机制才能确保所做决策的正确性。决策权分配和职责分工就是确保做出正确有效决策的核心机制，因而也就成为数据治理的核心。

(4) 抓住数据治理的本质。对机构的数据管理和利用进行评估、指导和监督，通过提供不断创新的数据服务，为其创造价值，这是数据治理的本质





11.1.3 数据治理与数据管理的关系

- 治理负责对管理活动进行评估、指导和监督，而管理根据治理所作的决策来具体计划、建设和运营。治理的重点在于，设计一种制度架构，以达到相关利益主体之间的权利、责任和利益的相互制衡，实现效率和公平的合理统一，因此，理性的治理主体通常追求治理效率。
- 而管理则更加关注经营权的分配，强调的是在治理架构下，通过计划、组织、控制、指挥和协同等职能来实现目标，理性的管理主体追求经营效率。从上述论述可以看出，数据治理对数据管理负有领导职能，即指导如何正确履行数据管理职能。



11.1.3 数据治理与数据管理的关系

- 数据治理主要聚焦于宏观层面，它通过明确战略方针、组织架构、政策和过程，并制定相关规则和规范，来评估、指导和监督数据管理活动的执行。
- 相对而言，数据管理会显得更加微观和具体，它负责采相应的行动，即通过计划、建设、运营和监控相关方针、活动和项目，来实现数据治理所做的决策，并把执行结果反馈给数据治理。

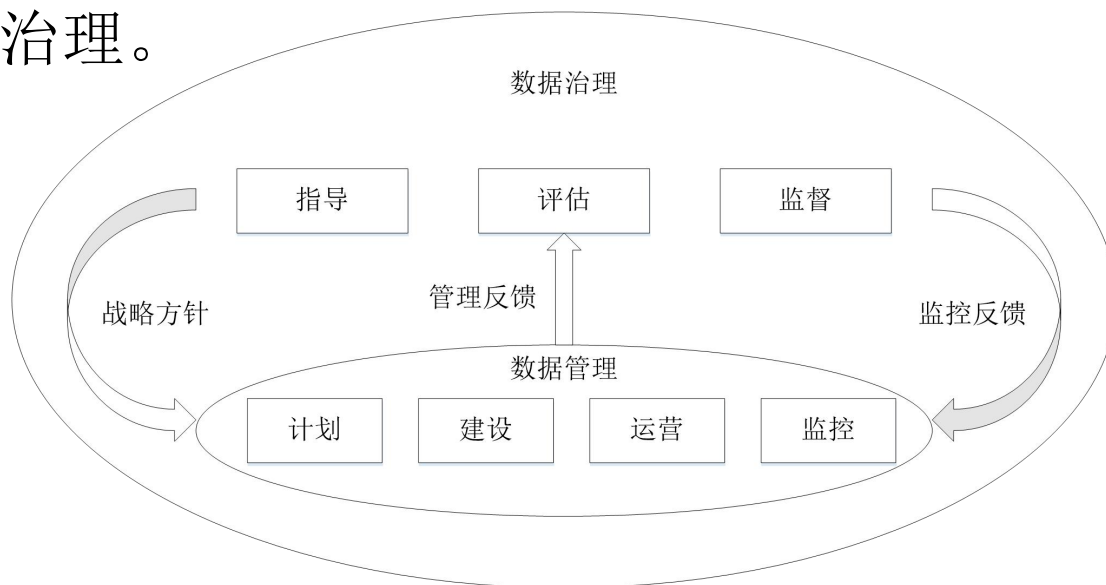


图 数据治理与数据管理的关系



11.1.4 大数据治理的基本概念

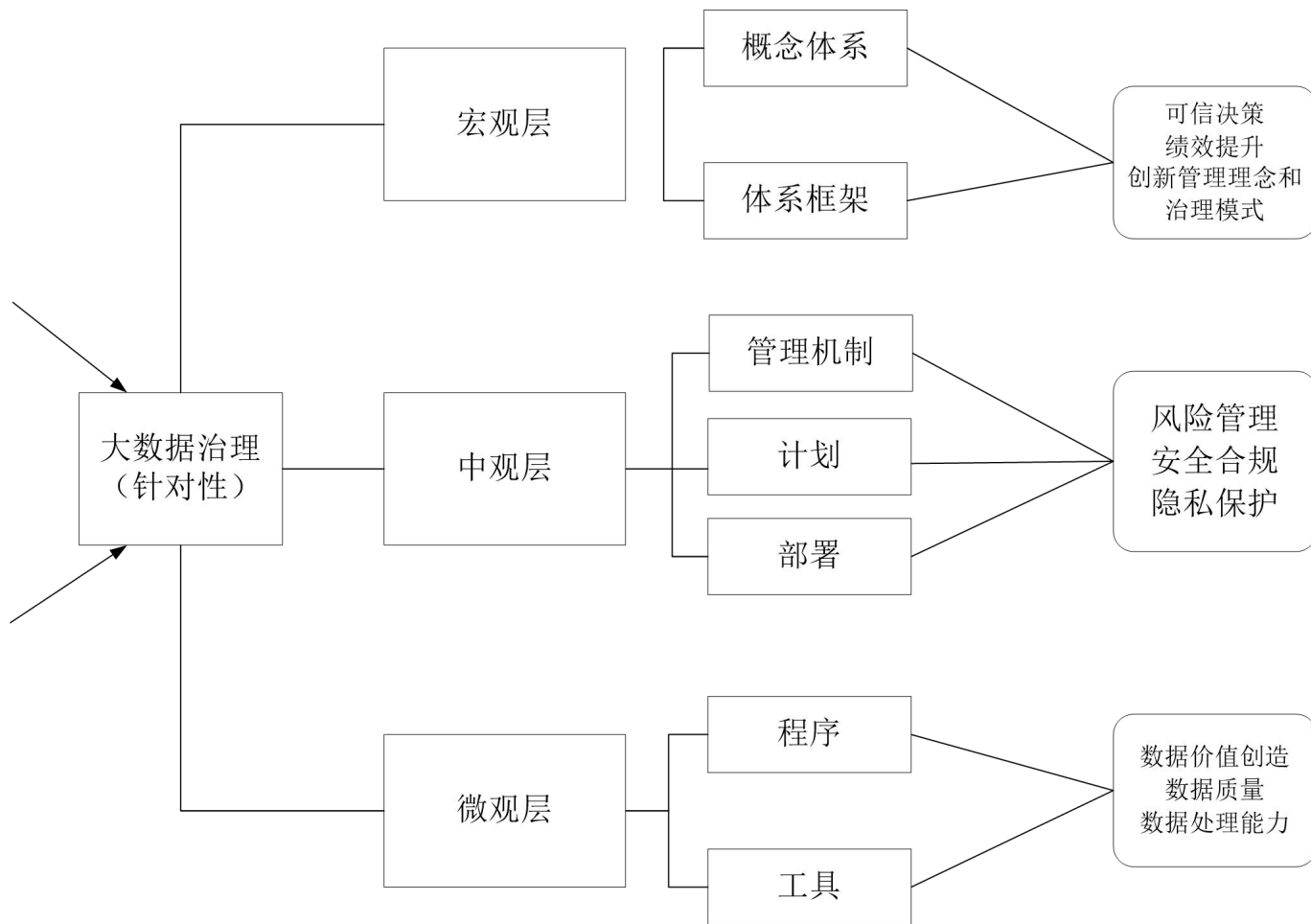


图 大数据治理3大层次



11.1.4 大数据治理的基本概念

- 在宏观层，大数据治理的概念包括两个方面：概念体系和体系框架。
- (1) 概念体系包括明确目标、权力层次、治理对象以及解决问题四个方面。
- (2) 体系框架是实现大数据治理，进行大数据管理、利用、评估、指导和监督的一整套解决方案，其构成要素包括制定战略方针、建立组织架构和明确职责分工等。



11.1.4 大数据治理的基本概念

2. 中观层

在中观层，大数据治理的概念表现在三个层面：

(1) 第一个层面是管理机制。在很大程度上，大数据治理是一种组织行为，完善的管理机制，可以作为数据治理的行动依据和指导方针，为实现“用数据说话、用数据决策、用数据管理、用数据创新”提供一套规范管理的路径。

(2) 第二个层面是信息治理计划，包括新兴的管理方法、技术、流程和实践，能够促成对大量的、有隐私的、有成本效益的结构化和非结构化数据的快速发现，并对其进行收集、运行、分析、存储和可保护性的处理。

(3) 第三个层面是数据全面质量管理的部署。大数据治理全面管理包括数据的可获得性、可用性、完整性和安全性的全生命周期和全面质量管理，尤其关注使用数据时的安全性和数据完整性。



11.1.4 大数据治理的基本概念

3. 微观层

在微观层，大数据治理的概念包括三个层面：

(1) 第一个层面是具体的经济有效的管理策略和过程，包括组织结构上的实践、操作上的实践和相关的实践。组织结构上的实践主要是识别出数据拥有者及其角色和责任；操作上的实践主要是组织执行数据治理的手段；相关的实践主要指改善政策有效性和用户需求之间的联系。

(2) 第二个层面是大数据治理是使用传统的数据质量维度的方法来测评数据质量和数据的可用性，这些维度包括精确性、完整性、一致性、实效性、单值性。

(3) 第三个层面是技术工具应用的大数据治理行为，涉及5个重要因素，包括：以关注人为基础的治理理念、以政府为主体的治理主体、以多种数据为客体的治理客体、以法律和计算机等软硬件为主的治理工具、以对大数据价值为主要发掘对象的治理目标。



11.1.5 大数据治理与数据治理的关系

表 大数据治理与数据治理的概念内涵比较

概念维度	大数据治理概念内涵	数据治理概念内涵
目的	鼓励“实现价值”和“管控风险”期望行为的发生，大数据治理更强调效益实现和管控风险	鼓励“实现价值”和“管控风险”期望行为的发生，数据治理更强调效率提升
权利层次	企业外部的大数据治理强调所有权分配；企业内部的大数据治理强调经营权分配	数据治理强调企业内部经营权分配
对象	权责安排，即决策权归属和责任担当	权责安排，即决策权归属和责任担当
实际问题	有哪些决策；由谁来作决策；如何作出决策；如何对决策进行监控	有哪些决策；由谁来作决策；如何作出决策；如何对决策进行监控



11.1.6 大数据治理的重要意义和作用

促进服务创新
和价值创造

01

提高数据质量，
增强数据可信度，
降低成本

03

02

提升数据管理
和决策水平

04

提高合规监管和
安全控制，降低风险



11.2 大数据治理要素

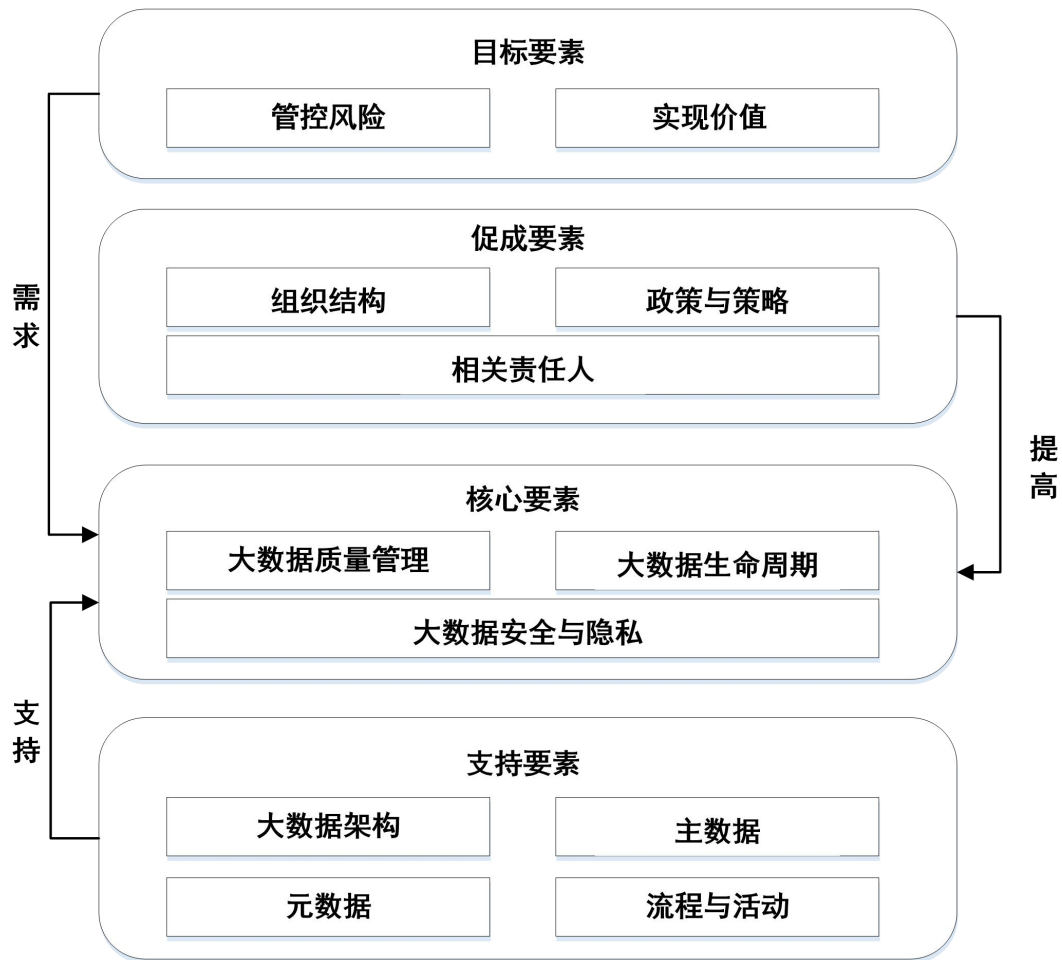


图 大数据治理要素



11.3 大数据治理模型

11.3.1 ISACA数据治理模型

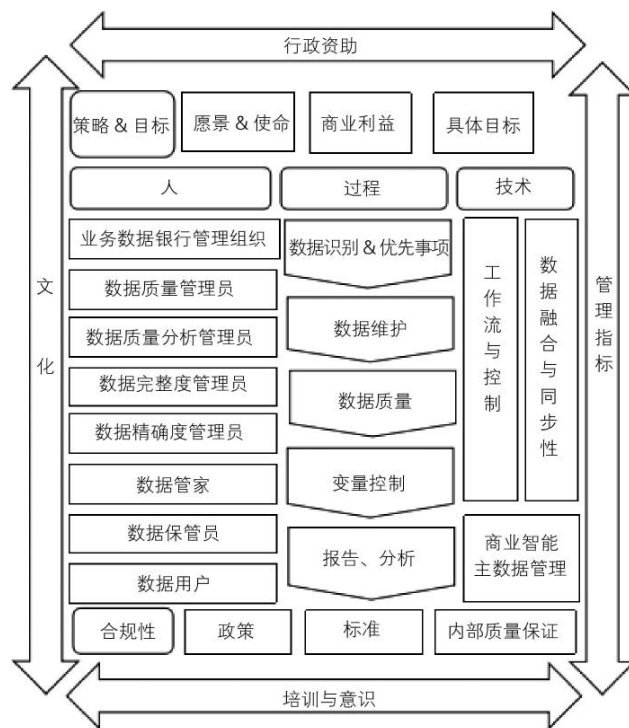
11.3.2 HESA数据治理模型

11.3.3 数据治理螺旋模型



11.3.1 ISACA数据治理模型

国际信息系统审计与控制协会(Information Systems Audit and Control Association, ISACA)是全球公认的信息科技管理、监控领导组织。ISACA从行政资助、文化、管理指标、培训与意识培养四个角度出发,构建了数据治理模型(简“ISACA模型”)



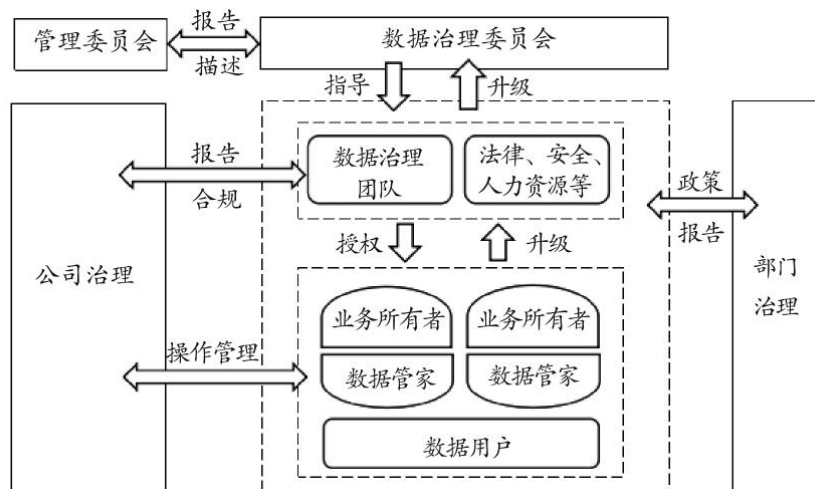


11.3.2 HESA数据治理模型

高等教育统计局是英国收集、分析和传播高等教育定量信息的官方机构，提出了数据治理模型(简称“HESA模型”)。

该模型数据治理的范围包括：

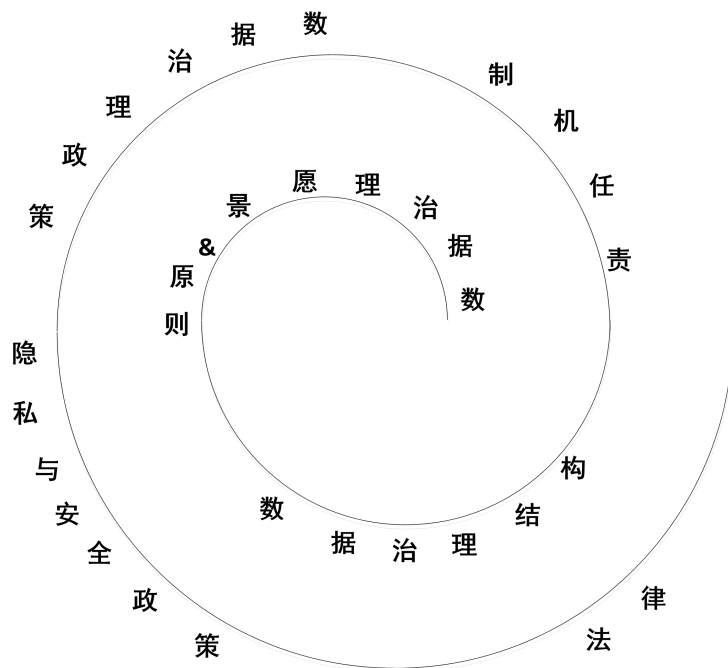
- 确保数据安全，确保组织面临的风险可控；
- 防止和纠正数据错误，从而不断完善数据治理计划；
- 衡量数据质量并提供检测和评估数据质量的改进框架；
- 记录数据及其在组织内的使用情况，作为数据相关问题和具体决策的参考。





11.3.3 数据治理螺旋模型

Mustimuhw Information Solutions是加拿大一家计算机软件公司，该公司研究发现，随着时间的推移，人们的需求和能力会不断变化和发展，治理模型也将随之而扩张和改进，不断迭代循环、发展壮大。因此，该公司认为数据治理应以螺旋模型呈现，以反映模型的动态和不断演变的性质





11.4 大数据治理保障机制

11.4.1 大数据治理战略目标

11.4.2 大数据治理组织

11.4.3 制度章程

11.4.4 流程管理

11.4.5 技术应用



11.4.1 大数据治理战略目标

- 数据治理战略是否与企业发展战略相吻合，是衡量数据治理体系实施是否成熟的一条重要标准。在企业发展战略框架的指导下，应当逐步建立数据治理的企业战略文化，有力推进企业数据治理工作的顺利开展。
- 数据治理的企业战略文化主要包括企业高层领导对数据治理的重视程度、所能提供的资源、重大问题的协调能力，以及对数据治理文化的宣传推广、培训教育等一系列措施。
- 数据治理的战略组成部分主要包括：数据治理的愿景、商业案例摘要、指导原则、长远目标分解、管理措施、实施线路等。在实现数据治理战略目标的过程中，要在集体内部形成统一的认知，让他们充分意识到数据治理工作的必要性和重要意义，由此才能促使企业内部产生统一的行动。



11.4.2 大数据治理组织



图 数据治理组织架构



11.4.3 制度章程

制度章程是一种用来确保对数据治理进行有效实施的认责制度，它包含多种职责，既包括数据治理职能的职责，也包括数据管理职能的职责。数据治理是最高层次的、规划性的数据管理制度活动。

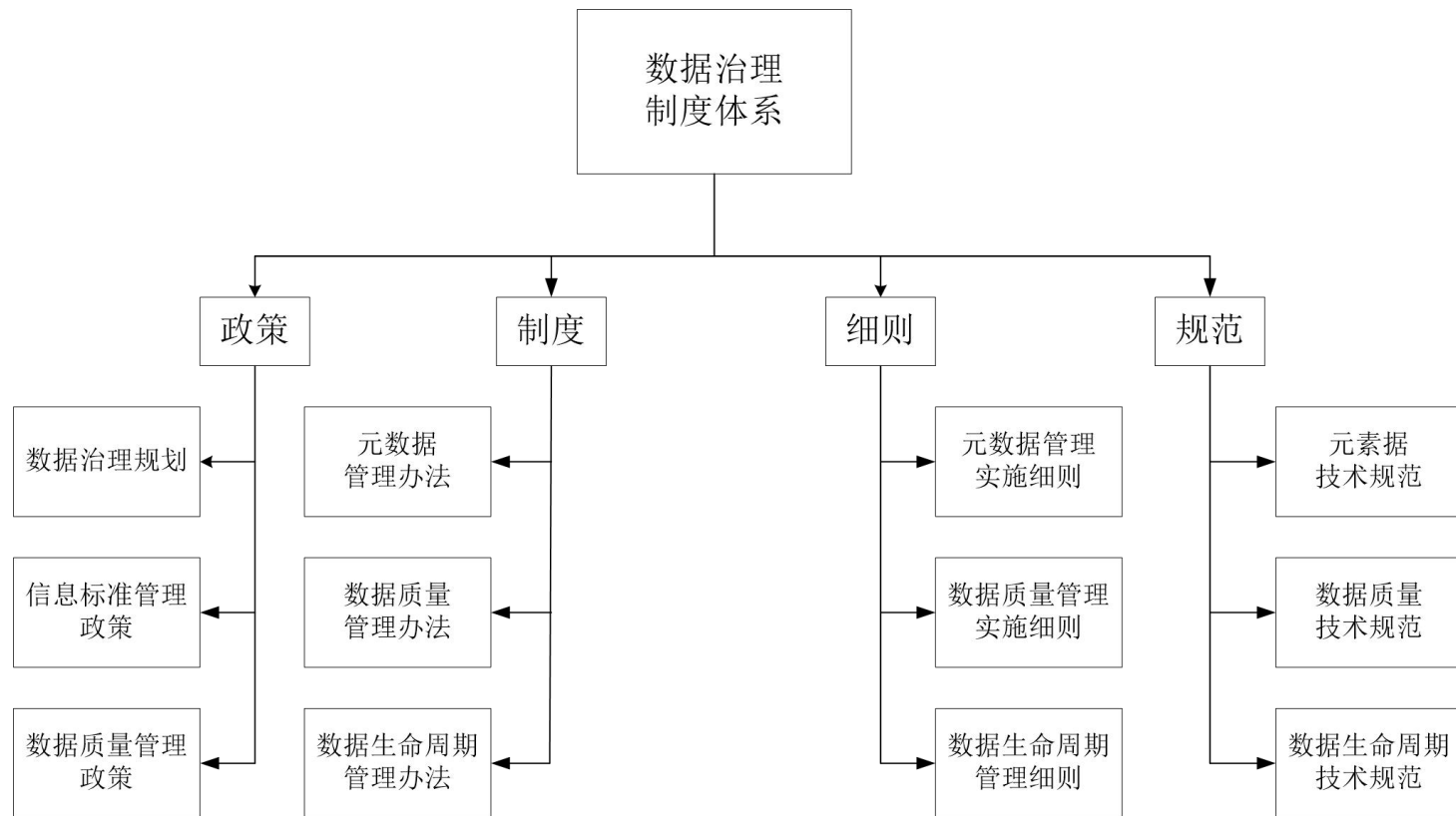


图 数据治理制度体系



11.4.4 流程管理

流程管理包括流程目标、流程任务、流程分级。具体流程的建立，需要以数据治理的内容作为依据，并且要做到严格遵循本单位数据治理的规章制度。

流程管理的具体工作包括：



做好事前预防 | 01

02 | 加强事中监测



进行事后评估和整改



11.4.5技术应用

(1) 建立数据资产管理系统，统一管理数据资产，包括元数据、数据模型、数据标准以及其他重要的数据资产，并提供可视化的数据查询和展示功能，从而支持数据资产的方便快捷查询。

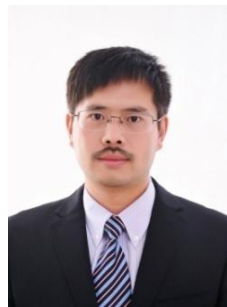
(2) 建立数据质量管理系统，落实数据质量问题的治理工作，实现数据质量问题的发现、跟踪、治理、评价的全流程闭环管理。同时，需要落实数据生命周期管理机制，在必要的时候，可以考虑搭建数据生命周期管理平台。

(3) 建设统一的数据仓库平台，持续整合各个生产系统和业务系统的基础数据，在组织层面提供一个统一的数据视图，满足前台营销、统计分析、决策支持、风险管理和新资本协议等多种需求。

(4) 加强各系统间的互联互通，努力消除信息孤岛，促进数据在各部门、各系统之间的自由流通，充分发挥数据的最大价值。



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年教育部产学研合作育人项目，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



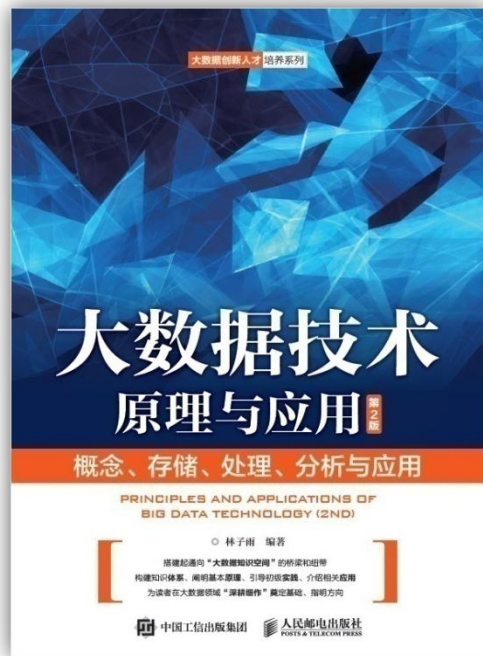
扫一扫访问教材官网

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://dmlab.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 features several faint, light-blue silhouettes of people. At the top, there are two groups of people, one on the left and one on the right, appearing to be in conversation or holding hands. On the right side, there is a larger silhouette of a person standing with their hand to their face. In the bottom left corner, there are silhouettes of two people sitting at a table, possibly in a meeting or discussion. The overall scene is set against a dark blue gradient background.

Thank You!

Department of Computer Science, Xiamen University, 2019