

9 大数据时代：从知识回到数据

9.1 大数据

大数据

- 一种大量而复杂的数据集合
- 在可承受的范围内，无法使用传统数据库系统和常规软件工具对内容进行获取、存储、管理和分析
- **特征**
 - 容量：数据量巨大
 - 种类：数据类型复杂多样
 - 速度：快速甚至实时地采集、处理数据并做出正确反馈
 - 价值：价值密度低
 - 真实性：数据判断真实可靠

大数据的发展趋势

- **数据的资源化**
 - 大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点
- **与云计算的深度结合**
 - 大数据离不开云处理，云处理为大数据提供了弹性可拓展的基础设备，是产生大数据的平台之一
- **科学理论的突破**
 - 随着大数据的快速发展，其很有可能是新一轮的技术革命
 - 随之兴起的数据挖掘、机器学习和人工智能等相关技术，可能会改变数据世界里的很多算法和基础理论，实现科学技术上的突破
- **数据科学和数据联盟的成立**
 - 未来，数据科学将成为一门专门的学科，被越来越多的人所认知
- **数据泄露泛滥**
 - 未来几年数据泄露事件的增长率也许会达到 100%，除非数据在其源头就能够得到安全保障
 - 企业需要从新的角度来确保自身以及客户数据，所有数据在创建之初便需要获得安全保障，仅仅加强数据保存的安全措施已被证明于事无补
- **数据管理成为核心竞争力**
 - 数据管理成为核心竞争力，直接影响财务表现
 - 当“数据资产是企业核心资产”的概念深入人心之后，企业对于数据管理便有了更清晰的界定，将数据管理作为企业核心竞争力
 - 持续发展，战略性规划与运用数据资产成为企业数据管理的核心
 - 数据资产管理效率与主营业务收入增长率、销售收入增长率显著正相关
 - 对于具有互联网思维的企业而言，数据资产的管理效果将直接影响企业的财务表现

9.2 DIKW 模型

9.2.1 DIKW 联系

数据：事实、信号或符号的集合；数据的价值有限

数据满足一定条件具有价值：

- 可以迅速获得
- 简洁、经过很好的组织，而且是相关的
- 具有基于经验的含义和上下文
- 是多个数据源的一种聚合

信息：按一致的方式整理和排序的数据集合；信息形式的数据更有价值——因为很容易存储和检索

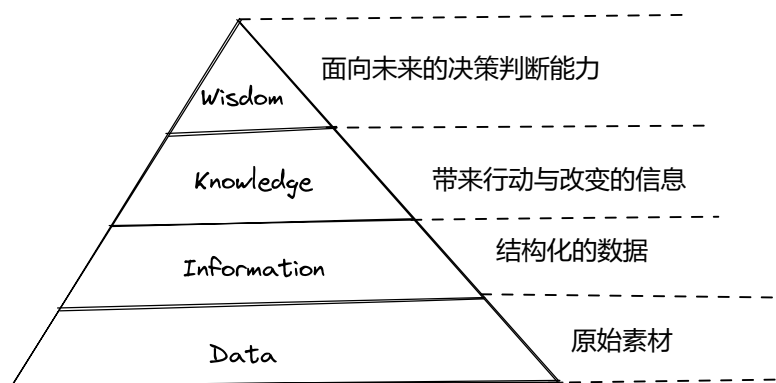
知识：信息及其相关上下文的集合，上下文表现为不断收集的信息集之间的关系；知识是处理一些信息的经验结果

智慧：根据知识来选择达到目标结果的最佳方式的能力；对达到某个成功结果的早期尝试的经验结果或知识

人类通过经验和知识来获取智慧，智慧来自：

- 培养对问题解决方法的理解决
- 通过分析给定上下文的数据和信息来开发洞察
- 从解决相同问题的其他人那里收集情报

9.2.2 DIKW 模型图



大数据技术的战略意义不仅仅在于掌握庞大的数据信息，还在于对这些含有意义的数据进行专业化处理

智慧：怎么做最好

- 前瞻性看法与想法
- eg：针对特定个体，预测最合适的治疗方法

知识：怎样做，为什么

- 信息、文化背景和经验的组合

- eg: 这类病症使用这个方法都会有帮助
信息: 是什么, 怎么样
- 经过筛选、整理与分析的资料
- eg: 用了这套疗法的效果不错
数据: 原始素材, 基本描述
- 离散、不相关的事实、文字、数字或符号
- eg: 手术后住院记录