第6届全国高校大数据与人工智能教学研讨会

2023.05.12–2023.05.13　中国·厦门

主办单位：教育部高等学校计算机类专业教学指导委员会

承办单位：厦门大学 XIAMEN UNIVERSITY　浙江大学 ZHEJIANG UNIVERSITY　中山大学 SUN YAT-SEN UNIVERSITY　中南大学 CENTRAL SOUTH UNIVERSITY　湖南大学 HUNAN UNIVERSITY　郑州大学 ZHENGZHOU UNIVERSITY　华侨大学 HUAQIAO UNIVERSITY　集美大学　人民邮电出版社 POSTS & TELECOM PRESS

协办单位：美林数据 MERITDATA　头歌 在线实践 educoder.net　普开数据 PUKAI DATA　山东省计算中心（国家超级计算济南中心）　荔峰科技 LIFENG　Tedu.cn 达内教育　高慧强学 HIGHSTRONG.CN

哈尔滨工业大学的
数据科学与大数据技术专业建设

哈尔滨工业大学 王宏志
wangzh@hit.edu.cn
http://homepage.hit.edu.cn/pages/wang

专业内涵

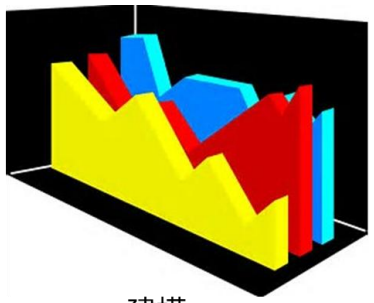
大数据是通过**传统数据库技术**和数据处理工具不能处理的庞大而复杂的数据集合。

数据科学通过融合多学科交叉技术实现从数据中发现**有价值的信息或规律**。

大数据与计算机

- 人工处理规模大、变化快的数据难度越来越大
- 大数据需要“沙里淘金”

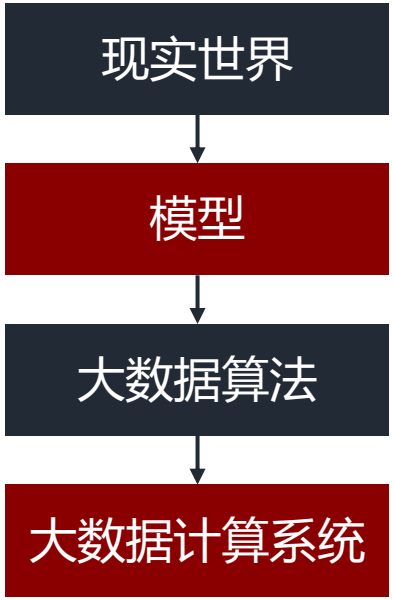
系统



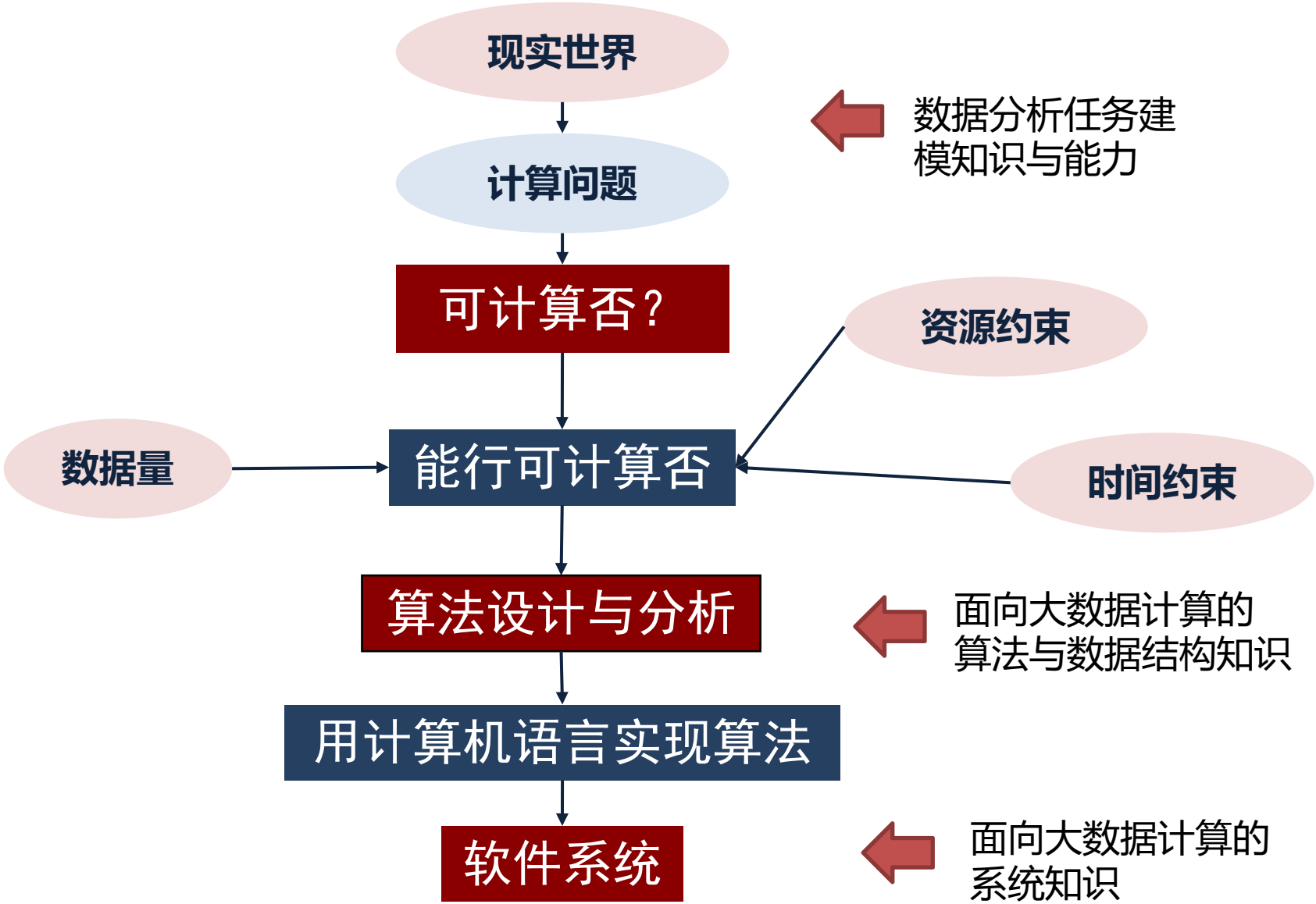
建模

实现

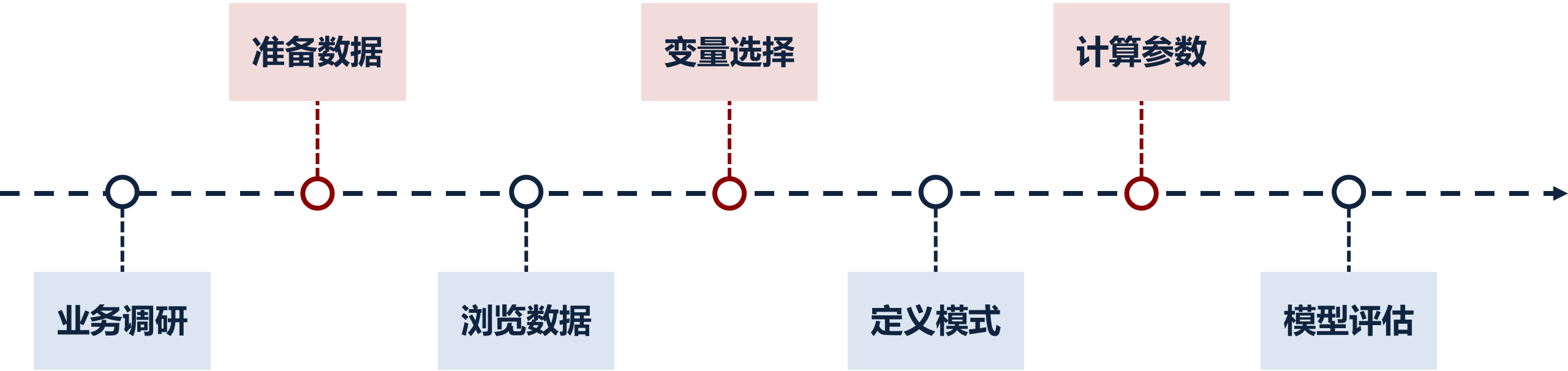
评估



求解大数据计算问题的过程



“现实”到“模型”



数据驱动业务的逻辑思维能力

面向大数据计算的算法知识

- 资源限制
  - 内存不足
    - 外存算法
    - 空间亚线性算法
  - 处理器计算能力不足
    - 并行算法
- 实时性要求
  - 问题计算复杂度下界难以满足要求
    - 时间亚线性算法

“好算法”与“好系统”

- ✓ 面向大数据计算的分布式系统知识
- ✓ 面向大数据计算的数据管理知识
- ✓ 面向大数据计算的程序设计知识

适合的大数据计算软硬件平台

设计高效的大数据存取结构

- 数据存储结构
- 数据分布策略
- 数据索引方法

编写适用于大数据的“好程序”

- 避免使用系统垃圾回收机制
- 减少内存拷贝
- 减少数据重分布次数
- 减小重分布数据量

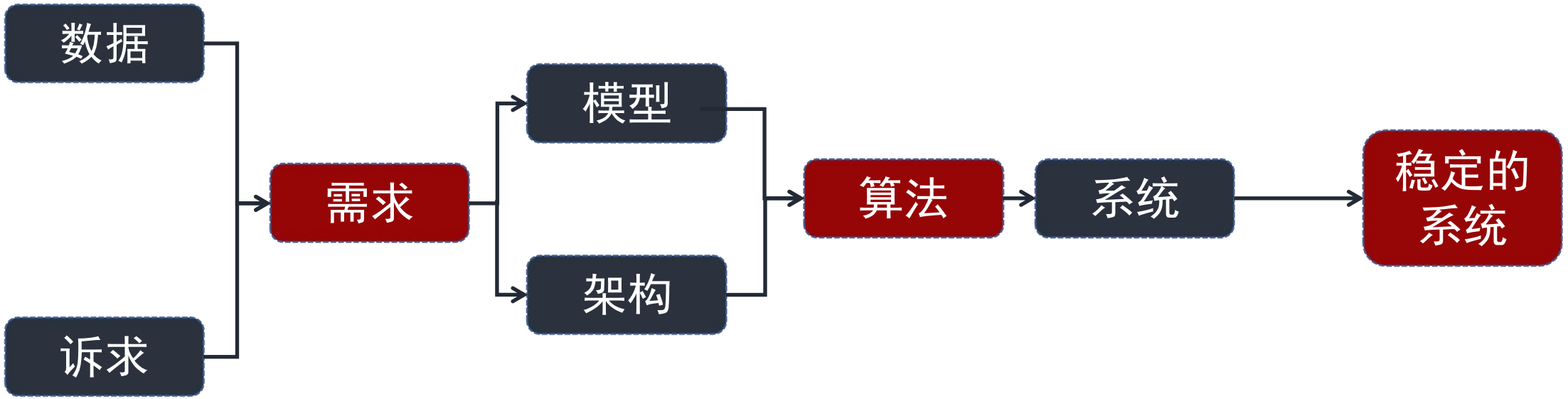
专业区分度

- **独特的学科基础和内涵**
  - 大数据表达理论、大数据计算理论与技术、大数据应用基础理论
  - 不同于计算机科学与技术、软件工程等学科

  不仅要掌握计算方法和工具，还要认知数据本身的现象和规律、数据管理和处理的基础理论、全生命周期的数据管理方法和系统等专门知识
  - 不同于商业智能和统计学

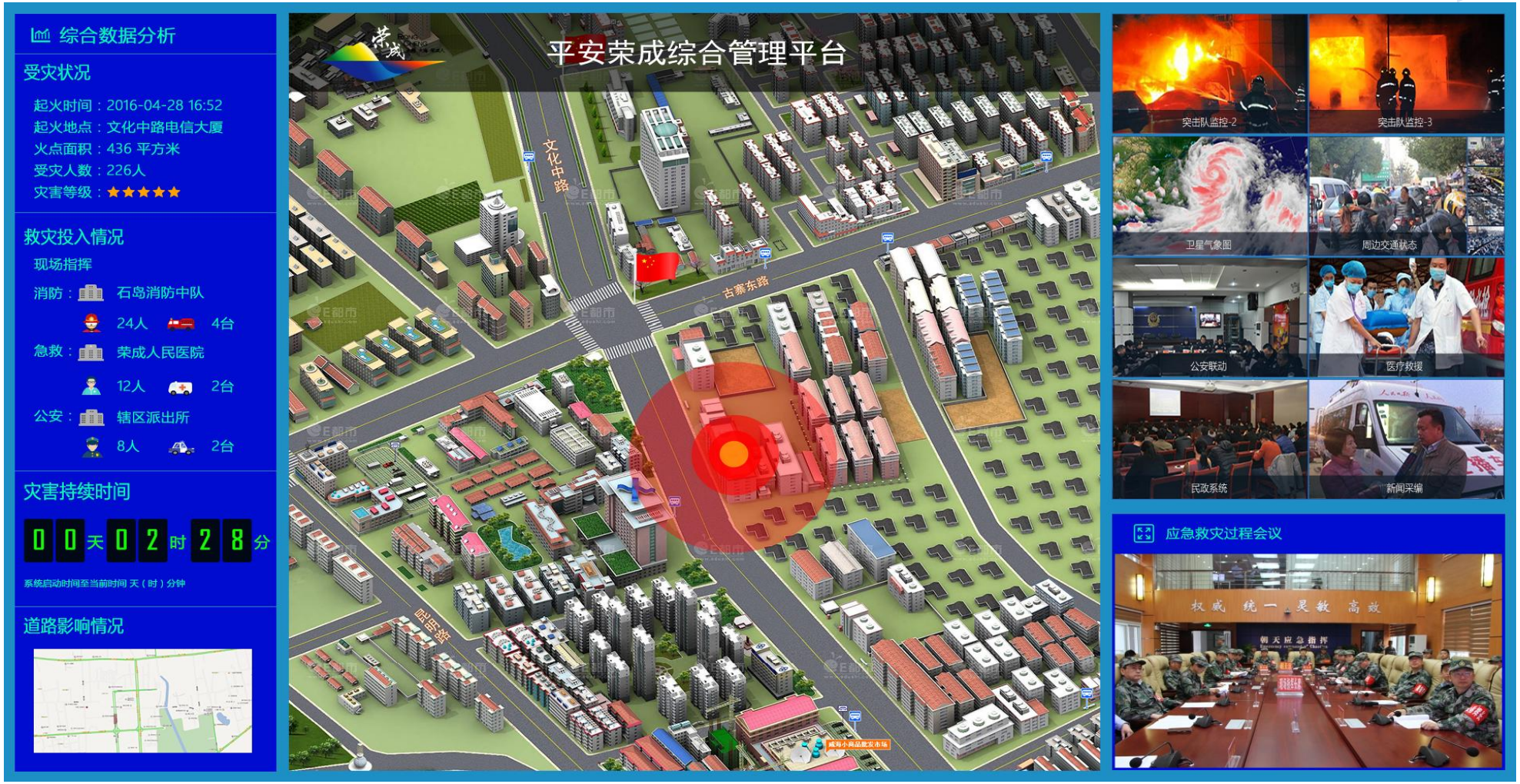
  不仅包含统计和商业智能的方法和模型，还包括算法设计与分析和计算系统的设计、研发、运维、评测、优化、应用等
- **独特的专业课程设置**
  - 突出数据科学基础课程教学
  - 裁剪传统计算机统类课程
  - 强调大数据管理与处理的全生命周期
  - 充分结合行业，突出实用性
- **独特的能力要求**
  - 数据密集型计算系统的设计、构建、运维及应用的能力
  - 数据密集型计算平台的开发及应用能力
  - 大数据理论、系统及应用的创新能力
  - 面向数据密集型问题，将现实问题抽象为数据计算模型的能力
  - 建立由多源异构数据到全面智能应用的建模及求解算法能力

大数据开发项目的过程



从大数据项目看大数据人才

某智慧城市项目



从数据和诉求到需求

平安（人民生产生活的底限需求）

幸福（能让百姓安居乐业的基础）

繁荣（新常态下经济转型

绿色（城市生态文明发展

高效（深化改革政府转型的结果）

“全生态”之城——智慧家园

地理信息公共平台

智能公交信息管理平台

出租车后台定位管理系统

XX市市山洪灾害监测平台

国家地名数据库管理系统

民心网

理平台

公共卫生管理系统

数据产品经理

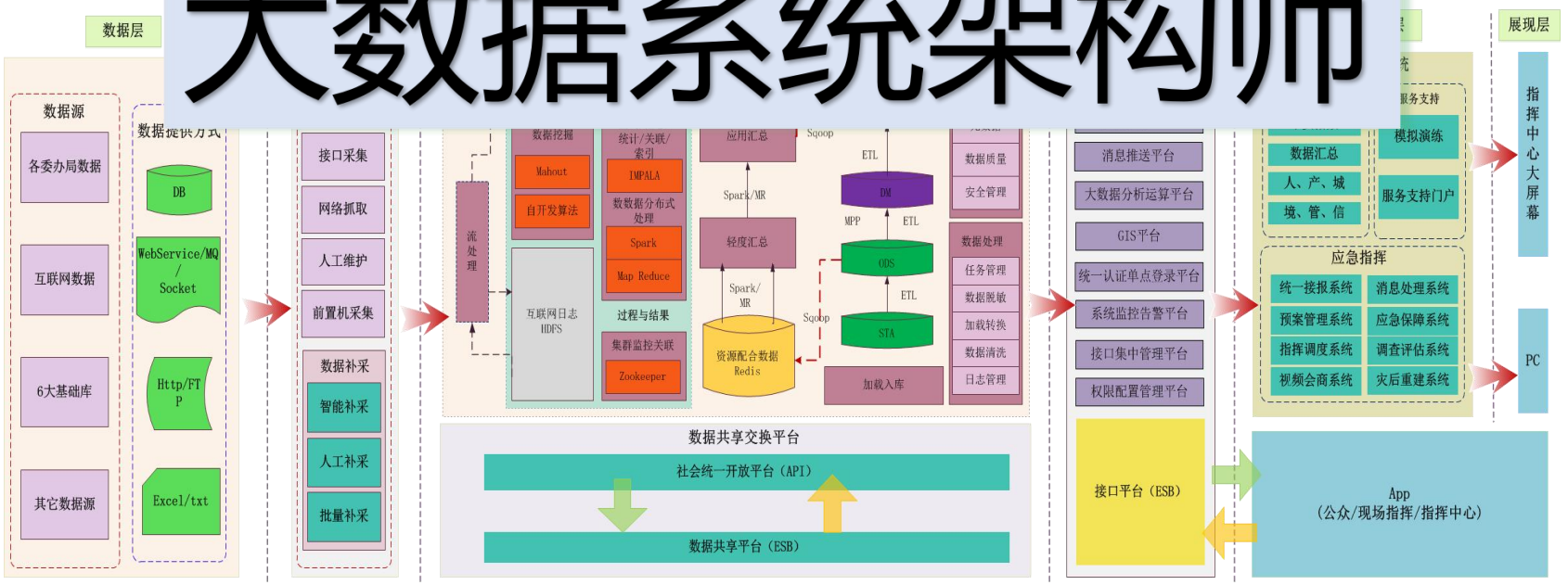
境 管 信 城

从需求到模型

数据分析师/数据科学家

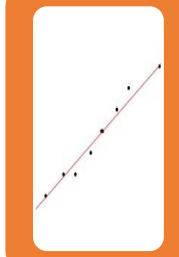
从需求到架构

大数据系统架构师



从模型和架构到算法

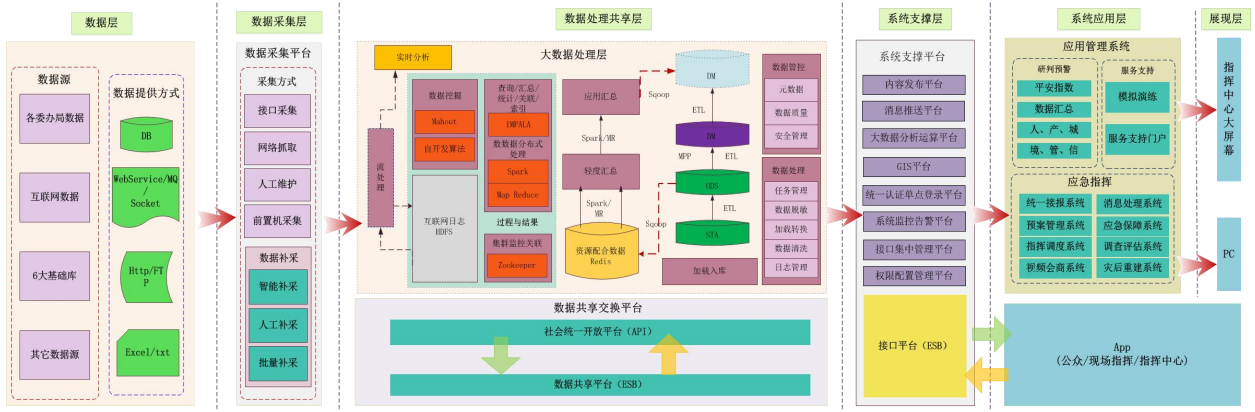
回归算法



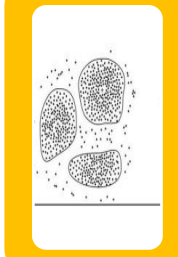
- 最小二乘
- Lasso回归
- ……

算法科学家/算法工程师

- …

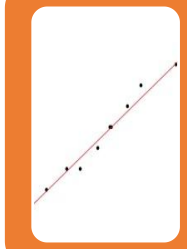


聚类算法



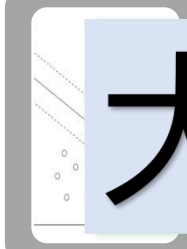
- Kmeans
- DBSCAN
- ……

从算法到系统



回归算法

- 最小二乘
- Lasso回归
- ……



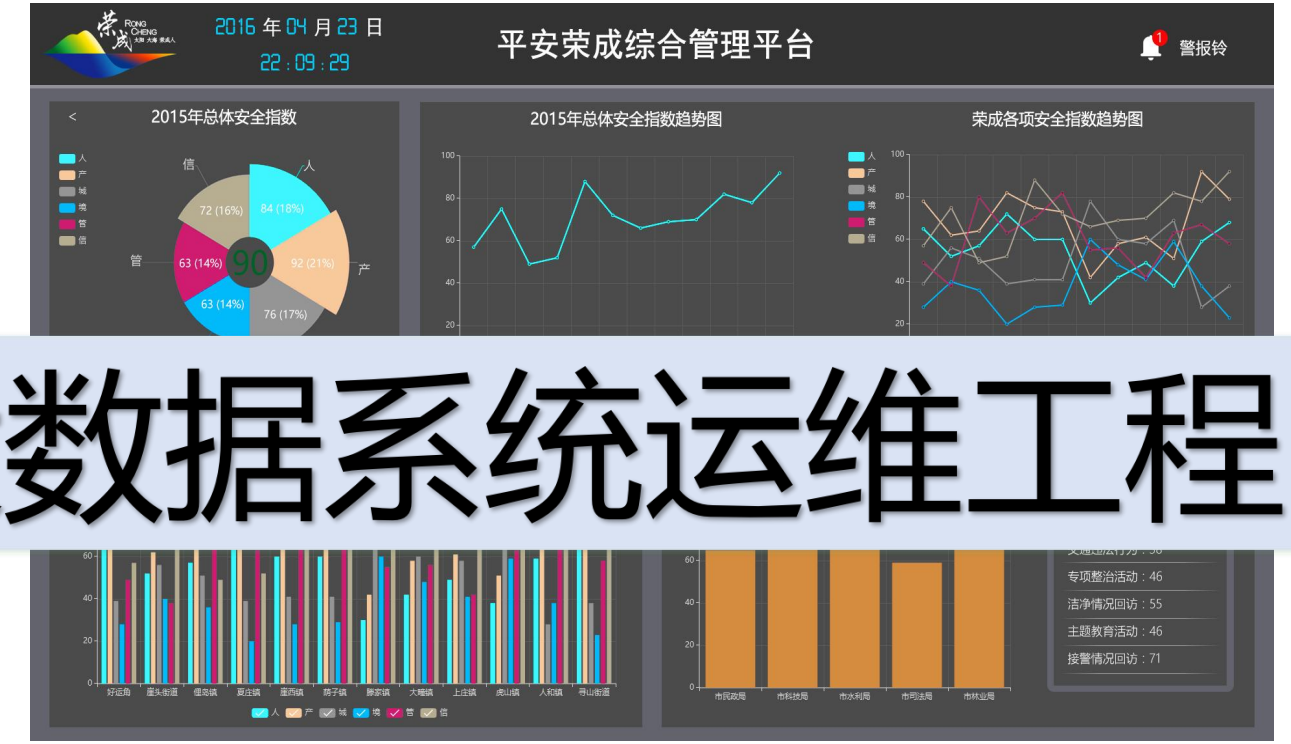
分类算法

聚类算法

- Kmeans
- DBSCAN
- ……

大数据系统开发工程师

从系统到稳定运行的系统



大数据系统运维工程师

2017 2018 2019 2020 2021 ……

培养目标

面向世界科技前沿和国家重大需求，瞄准数据科学与大数据技术的未来原创性、革命性、颠覆性、交叉性技术，秉承“规格严格、功夫到家”的校训，着力培养信念执著、品德优良、知识丰富、本领过硬、具有国际视野、引领未来发展的新时代杰出人才。

以掌握迎接第四次工业革命发展变化所需要的数据科学与大数据技术基础知识和专业技能为核心，注重能力培养和素质提高，培养基础宽厚、能力出众、德智体美劳全面发展的时代新人，具备良好的**学术大师、工程巨匠、业界领袖、治国栋梁**潜质。

毕业五年的目标

- 身心健康，具有正确的世界观、人生观与价值观，恪守工程伦理与职业道德规范，德智体美劳全面发展
- 具备社会责任感、专业使命感，具有在**数据科学、大数据技术及其相关交叉领域**引领未来发展的潜力
- 成为数据科学和大数据技术领域的**助理教授、数据科学家、大数据工程师或系统架构师**、数据科学和大数据技术领域的创业者和企业管理者、**具备数据科学和大数据技术知识的社会管理者**
- 具备面向数据科学和大数据技术的**计算思维和系统思维**能力，能够综合运用**计算机软硬件、数学和数据科学与大数据技术**等方面知识，能够对现实世界和信息化系统中的问题进行**建模**，面向实际应用设计高效的**逻辑和物理数据结构**，分析、设计、实现面向大数据的**算法**和计算**系统**并进行模型、算法和系统的**评估**
- 具备较强的数据科学跨学科交叉创新能力或大数据计算相关的理论与工程创新能力
- 具有国际视野、团队合作、项目管理、跨文化交流、终身学习能力

更长远的目标

毕业十五年左右的杰出人才培养目标为：

成为大数据技术领域中在学术界有一定影响力科学研究人员、高水平系统研发人员、企事业单位管理者、以大数据关键技术为核心的创业者；生物、工业、金融等以数据为核心的重要领域中有影响力的数据科学家、企事业单位管理者或以数据科学驱动交叉领域的创业者；具备数据科学和大数据技术相关知识的社会管理者。

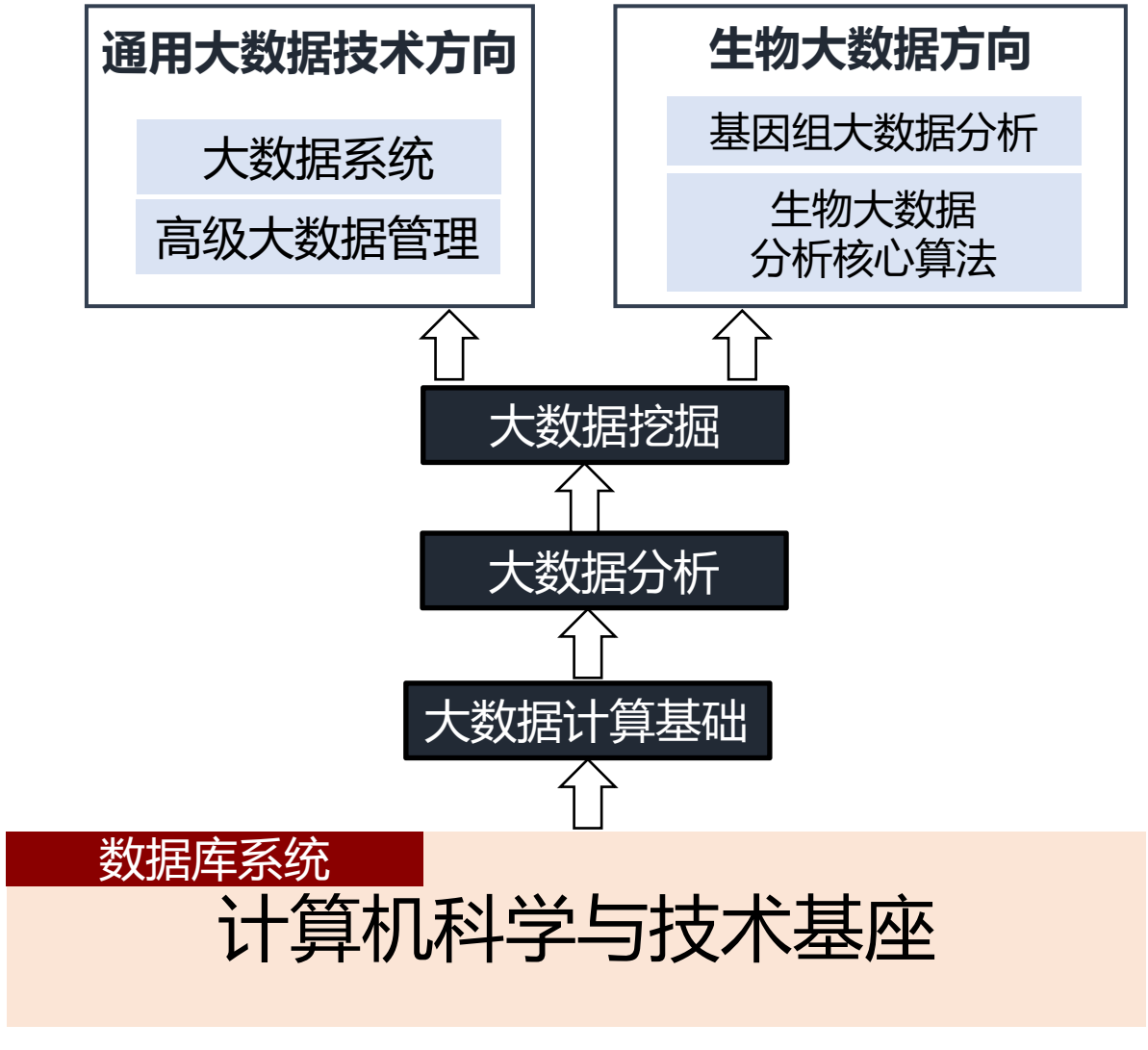
毕业三十年左右的杰出人才培养目标为：

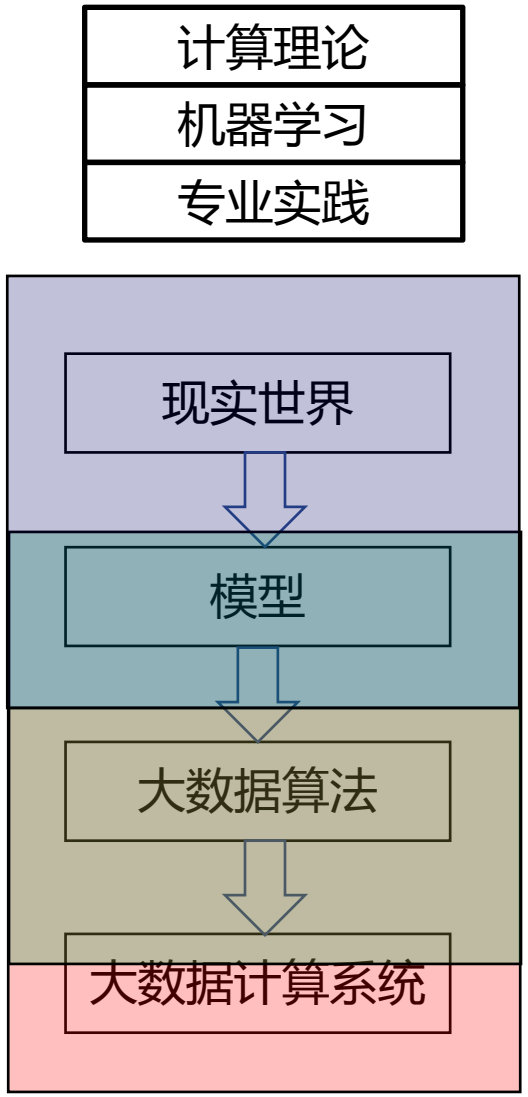
成为引领世界数据科学发展的学术大师，及运用大数据科技解决工业、经济、医疗、健康、能源、环境等重大问题与挑战，带领我国迈向世界强国并引领人类进步的工程巨匠、业界领袖、治国栋梁。

课程体系

	课程	学分	学时
基础能力培养	大学计算机	2	32
	高级语言程序设计	3	48
	集合论与图论	3	48
	数据结构与算法	3	48
	专业解读	1	16
	小计	12	
科学理论培养	数理逻辑与近世代数	3	48
	形式语言与自动机	2	32
	算法设计与分析	2	32
	人工智能	2	32
	小计	8	
系统能力培养	数字逻辑与数字系统设计	2	32
	计算机系统	3	48
	软件工程	3	48
	计算机网络	3	48
	操作系统	3	48
	编译系统	2	32
	小计	16	
实践创新能力	程序设计能力训练	1	24
	软件设计与实践	2	48
	计算机硬件设计与实践	2	
	专业实践	2	48
	实习实训	2	2周
	Pjbr项目开发	1	
	毕业设计	10	一学年
	小计	20	
专业胜任力	大数据计算基础	3	48
	数据库系统	3	48
	大数据分析	3	48
	大数据挖掘	3	48
	专业方向选修I	3	48
	专业方向选修II	3	48
	小计	18	

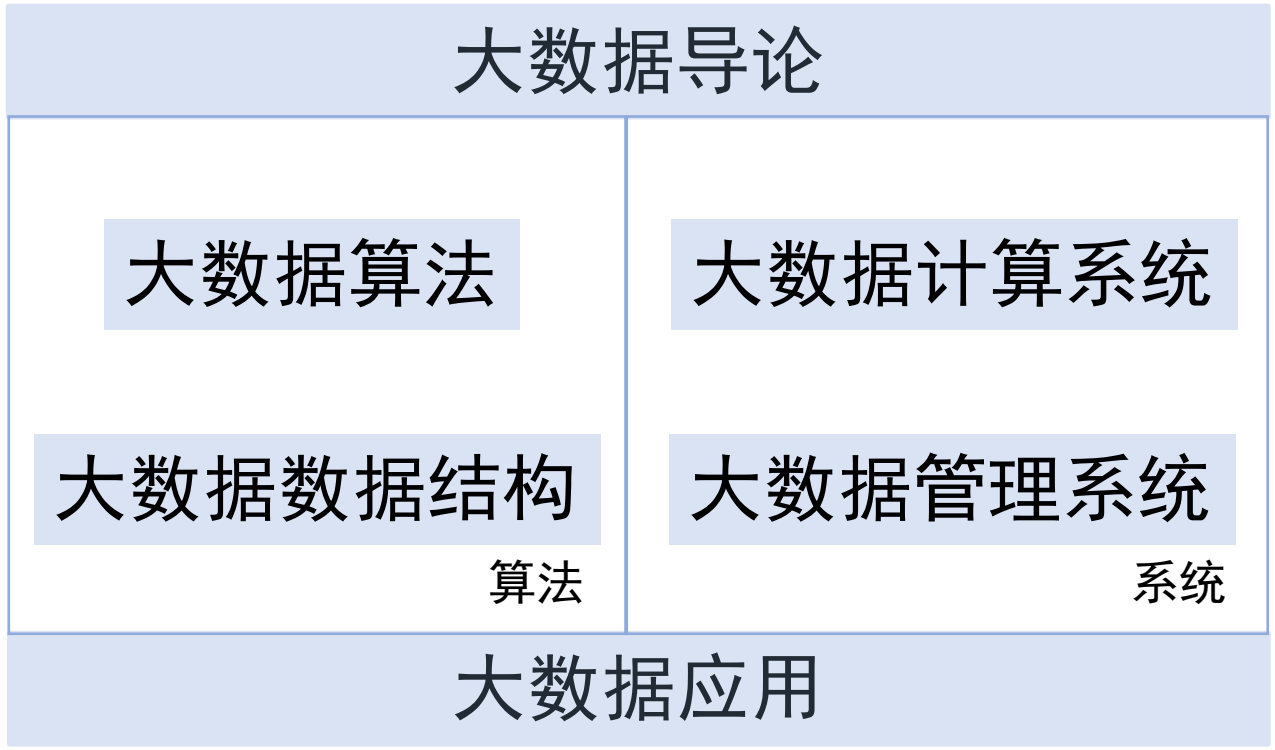
专业课程体系


通用大数据技术方向
大数据系统
高级大数据管理
生物大数据方向
基因组大数据分析
生物大数据
分析核心算法
大数据挖掘
大数据分析
大数据计算基础
数据库系统
计算机科学与技术基座

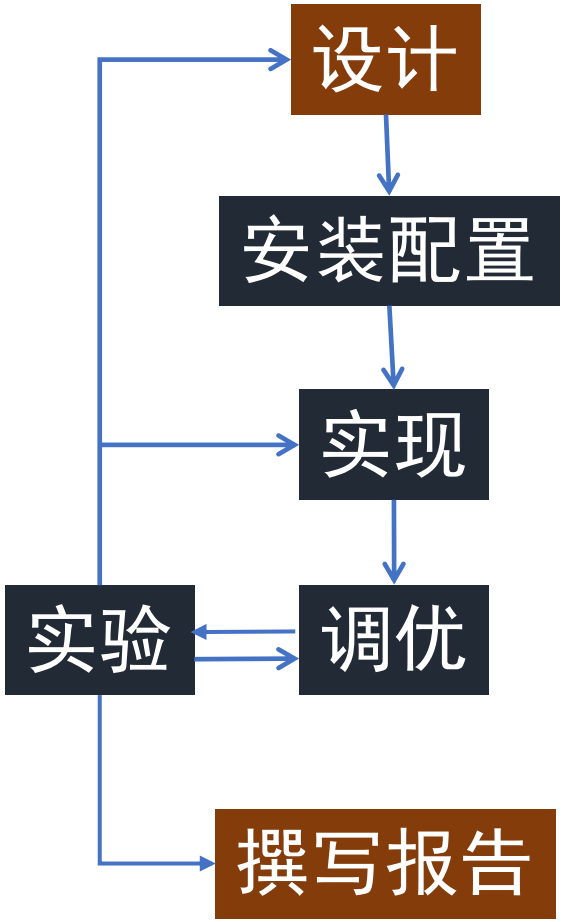

计算理论
机器学习
专业实践
现实世界
模型
大数据算法
大数据计算系统

“大数据计算基础”脉络图

课堂授课



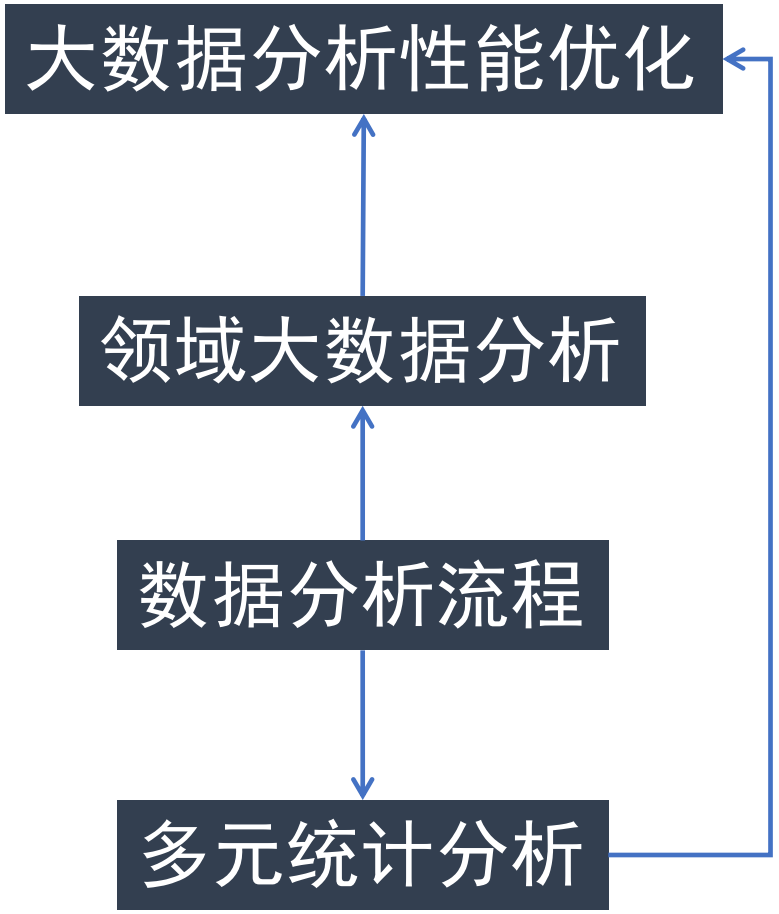
实验与作业



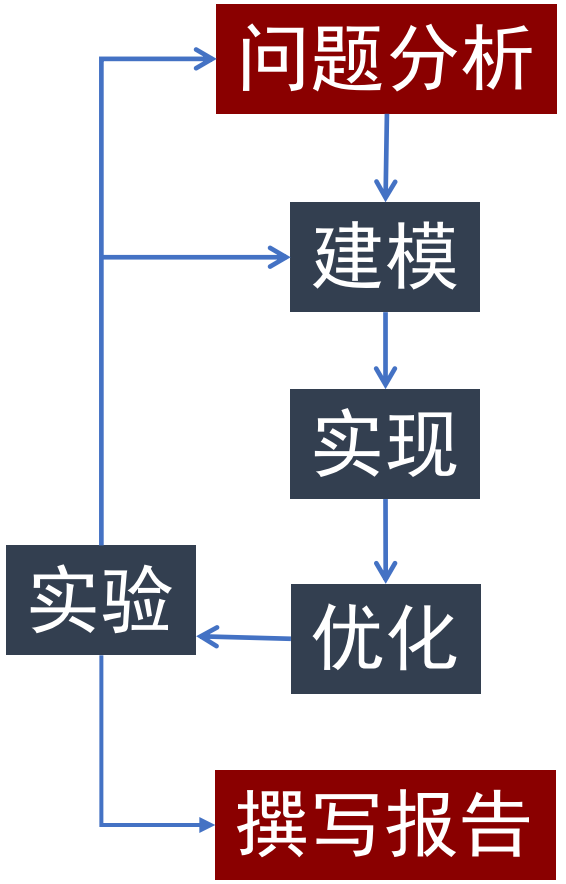
“大数据计算基础”脉络图

大数据分析

课堂授课



实验与作业



数据挖掘

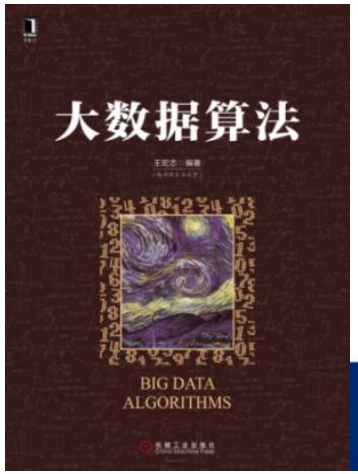
- 数据库、机器学习、统计学、模式识别、人工智能以及高性能计算等技术的融合
- 大数据计算基础和大数据分析课程的总结与延伸

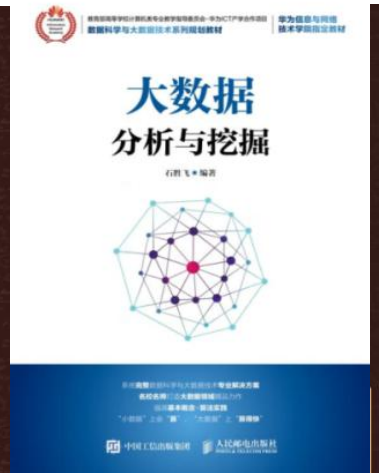
1.数据挖掘概述	
2.数据特征分析与预处理 (1) 数据的类型 (2) 数据的统计特征	(3) 数据预处理 (4) 缺失值的处理 (5) 数据可视化
3.关联规则挖掘	
4.分类算法	
5. 聚类算法	
6. 异常检测	

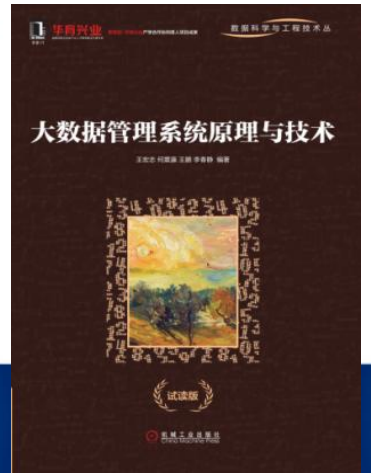
专业建设效果

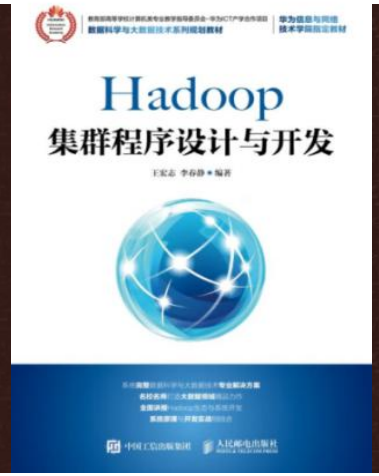
- 2022年获批国家一流专业建设点
- 三门专业核心课程全部进入ACM数据科学示范课程
- 2021年和2022年蝉联软科中国大学排名第一名

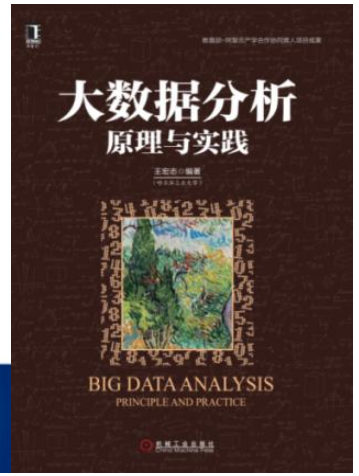
现实世界 → 模型 → 大数据算法 → 大数据计算系统











哈爾濱工業大學

谢谢各位老师

哈尔滨工业大学　王宏志
wangzh@hit.edu.cn
http://homepage.hit.edu.cn/wang