



西北工业大学  
计算机学院

# 数据科学与大数据专业 建设思考

尚学群

2018年10月14日





# 目录 CONTENTS

一、背景与意义

二、课程体系设计

三、实践环境搭建



# 一、背景与意义

## 大数据意识润物无声



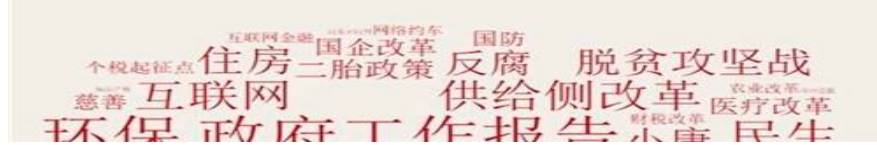
国庆期间全国热门景区TOP10：预计西湖、故宫、天安门排名前三



- 国庆长假期间，各地景区热度增高，预计西湖风景区、故宫博物院、天安门、上海迪士尼度假区、颐和园热度位居全国前五。
- 十一期间，预计风景区、博物馆类景区等将更受游客欢迎。



国庆期间全国热门景区TOP10





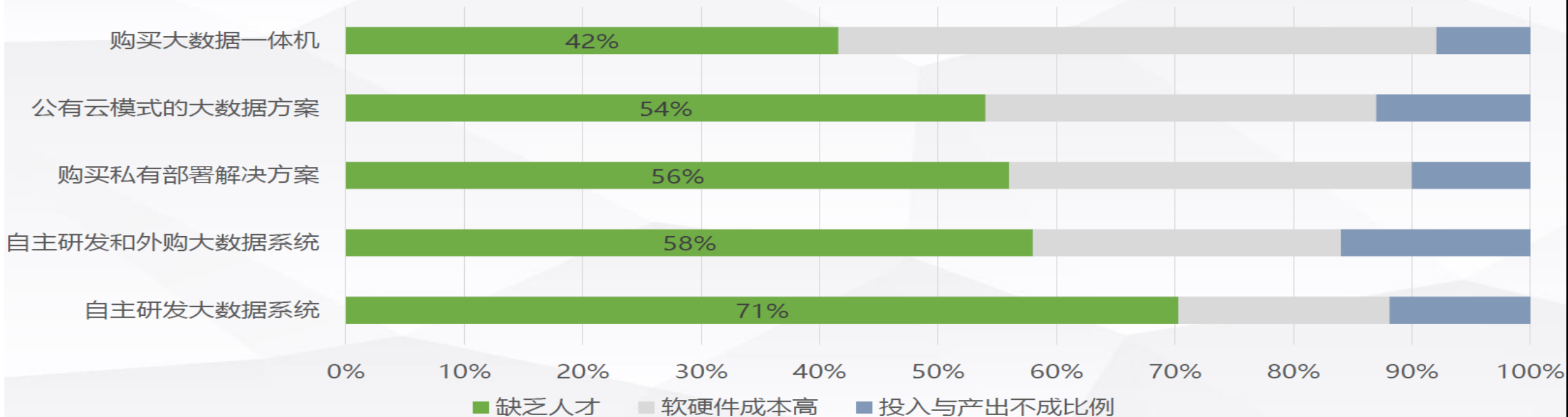
## 数据国家战略发展需求

- ❑ 2015年7月1日《国务院办公厅关于运用大数据加强对市场主体服务和监管的若干意见》（国办发〔2015〕51号）
- ❑ 2015年9月5日《国务院关于印发促进大数据发展行动纲要的通知（国发〔2015〕50号）》
- ❑ 2016年3月，《中华人民共和国国民经济和社会发展第十三个五年规划纲要》提出：
  - 实施国家大数据战略
  - 把大数据作为基础性战略资源，全面实施促进大数据发展行动，加快推动数据资源共享开放和开发应用，助力产业转型升级和社会治理创新
  - 加快政府数据开放共享。统筹布局建设国家大数据平台、数据中心等基础设施



## 2025年中国大数据人才缺口200万

搭建大数据平台应用方式中遇到的主要困难



人才需求数量巨大，高端人才、应用型和专业性人才缺乏



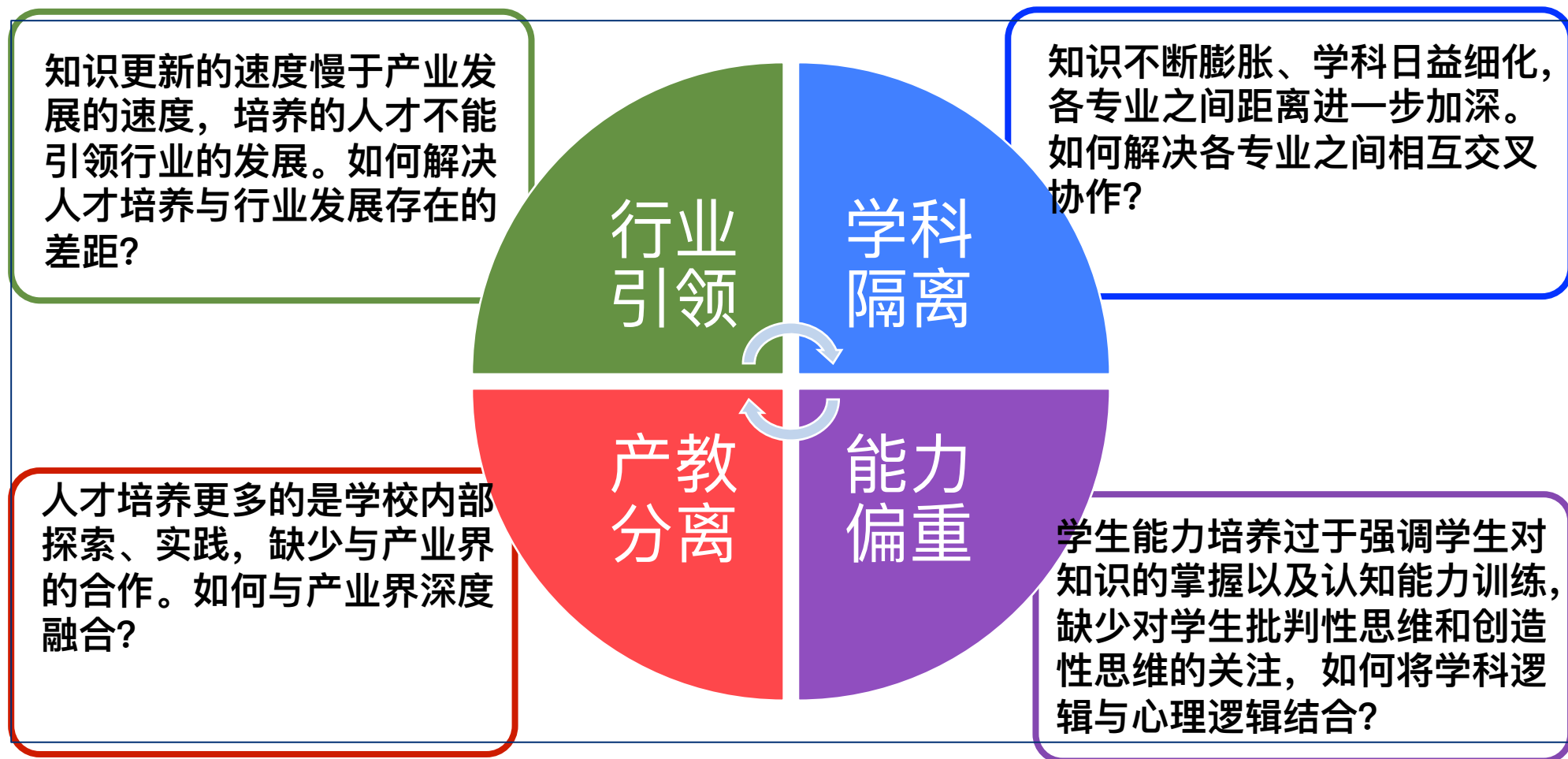
## 数据科学与大数据技术专业建设已成为高校关注的新的热点话题

- 在教育部的新增审批本科专业里，满眼都是“大数据”这样的词汇
  - 第一批（3所）：北京大学、对外经济贸易大学、中南大学
  - 第二批（32所）：中国人民大学、复旦大学、北京邮电大学、华东师范大学、电子科技大学.....
  - 第三批（248所）： .....
- 从申请资料看，主要以统计学和计算机科学与技术专业为基础建设的全新专业
- 从培养目标来看，二专多能复合的跨界人才(有专业知识、有数据思维)。主要分数据科学人才、大数据技术人才和大数据应用人才三个层次
- 从培养方案来看，目前主要集中在大数据分析及挖掘、大数据系统开发

不同办学层次的院校开设此专业，培养模式会有差异



## 技术发展催生下的新兴学科和专业，专业建设方面还存在许多问题





## 国际数据科学专业调研

### □ 调研目标

- 分析世界一流大学数据科学专业的人才培养理念
- 分析世界一流大学数据科学专业的课程体系特点

### □ 调研范围

- 卡耐基梅隆大学 (Carnegie Mellon University, CMU)
- 斯坦福大学 (Stanford University)
- 加州大学伯克利分校 (University of California-Berkeley, UCB)
- 约翰·霍普金斯大学 (Johns Hopkins University, JHU)
- 哥伦比亚大学 (Columbia University in the City of New York)



## 国际数据科学专业人才培养特点

| 学校 | 卡内基梅隆大学<br>CMU                                                               | 斯坦福大学<br>Stanford                                                           | 加州大学伯克利分校<br>UCB                                                  | 约翰·霍普金斯大学<br>JHU                                         | 哥伦比亚大学                                                                      |
|----|------------------------------------------------------------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------------|-----------------------------------------------------------------------------|
| 学位 | 计算数据科学<br>硕士学位                                                               | 统计学：数据科学<br>理学硕士学位                                                          | 信息与数据科学<br>硕士学位                                                   | 数据科学<br>理学硕士学位                                           | 数据科学<br>理学硕士学位                                                              |
| 特点 | 侧重于培养专业中的数据科学家，强调与具体专业学科高度融合的人才的培养。强调跨学科方法，重视统计学、计算机科学和具体应用领域之间的深度融合         | 侧重于培养统计学家，而并非数据科学家。其最突出的特点是将数据科学作为统计学的一个方向，从而培养出面向数据科学的统计学家，强调数据科学与统计学的深度融合 | 旨在培养数据科学领域的领导者，强调数据科学的多学科交叉特点，注重提升学生的基于数据提出和解决问题的能力，强调学生动手实践能力的培养 | 旨在培养“有竞争力”的数据科学家，强调学生对数据科学的理论基础的掌握程度，突出计算机科学和技术、统计学和应用数学 | 旨在培养数据科学家，主要将专业教育与专业认证相结合，专业教育与在线免费课程相结合                                    |
| 课程 | 智能信息系统、大规模数据集的机器学习、搜索引擎和可扩展分析学；操作系统的实现、多媒体数据库与挖掘、并行计算机架构及编程、高级存储系统、云计算及高级云计算 | 数值线性代数、离散数学与算法、优化论、工程中的随机方法、随机算法与概率分析、数据挖掘、基于数据的计算范式、数据驱动型医学                | 面向数据科学的Python语言、面向数据与分析的研究设计、面向数据科学的统计学、数据存储与检索、应用机器学习、试验与因果分析    | 算法基础、统计方法与数据分析、数据库系统原理、数据科学、数据可视化、优化导论、统计模型与回归、计算统计学     | 数据科学原理、面向数据科学的计算机系统、数据科学中的机器学习、数据科学中的算法、概率论、面向数据科学的概率统计、探索性数据分析及可视化、统计推理与建模 |



## 国际数据科学专业人才培养特点

- 最初主要是硕士学位课程，近年来开设本科课程
- 并非是计算机科学和统计学的简单拼凑
- 数据科学与领域知识的高度融合
- 重视解决复杂实际问题能力培养
- 注重与产业界的深度合作
- 不仅是技术和工程的问题，还涉及道德和法律的范畴
- 结合自身的学科优势，定位人才培养目标



### 数据科学与大数据与传统计算机科学与技术的差别

#### 1、基础理论深融合

传统CS课程的数学基础是高数、线代、概率、复变函数、离散数学等；数据科学与大数据还关注统计、序列分析、最优化等

#### 2、编程范型多样化

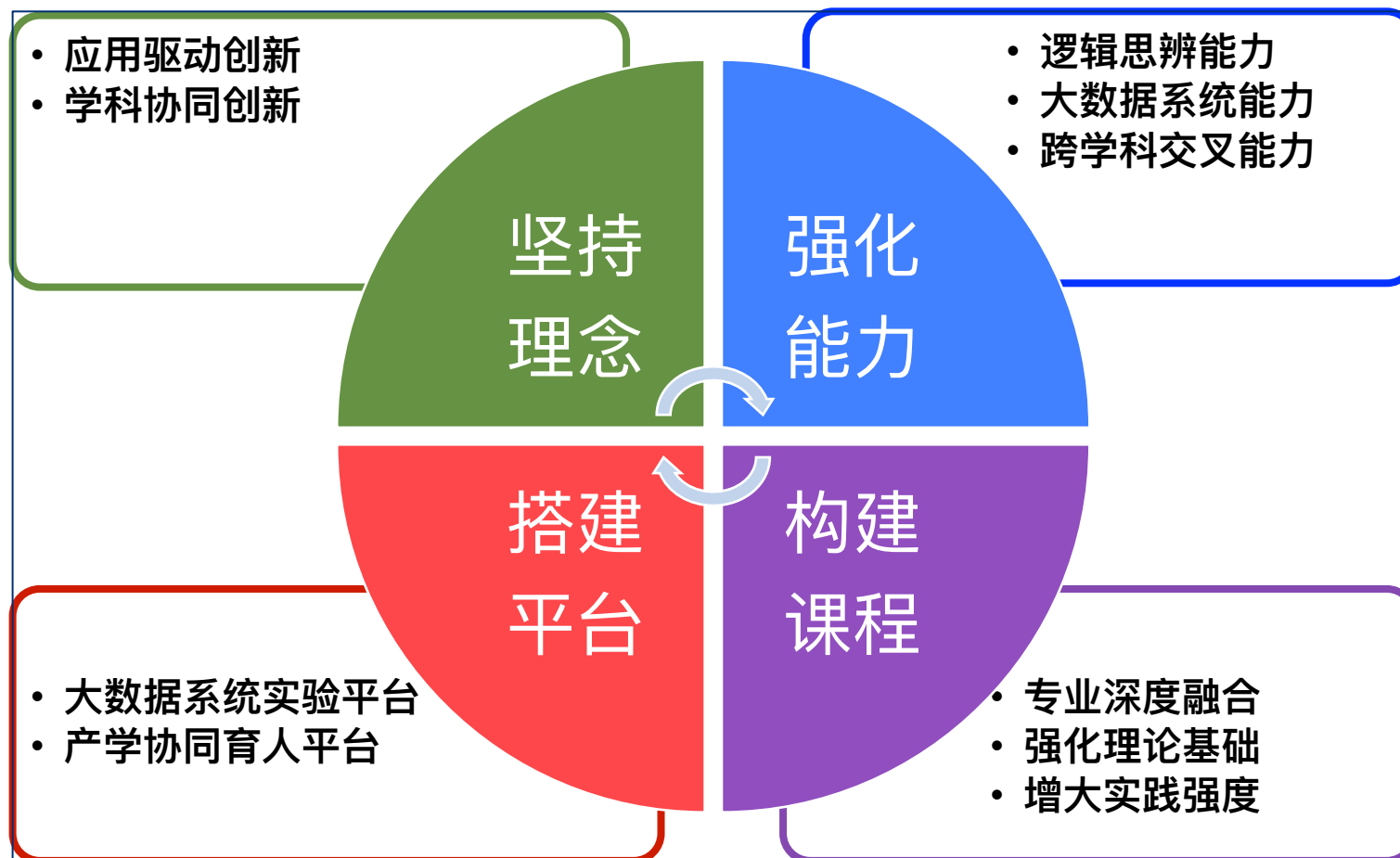
传统CS课程以结构化和面向对象程序设计语言为主。数据科学与大数据中出现了R语言、Scala语言。开发平台的基础向新型语言转变（传统的c更少，新型语言学习能力更为重要）

#### 3、系统平台大型化

大规模系统，不再是单机系统。软硬件综合化程度更高（提高算法、数据结构的性能或效率；同时发挥硬件效率）。

#### 4、领域背景更突出

传统CS注重技术，关心的领域通常也局限在CS内部；数据科学与大数据，则是针对领域要求，提出新的系统或改进现有系统。即系统和技术对应用背景更加敏感 -- 学科交叉性更强。



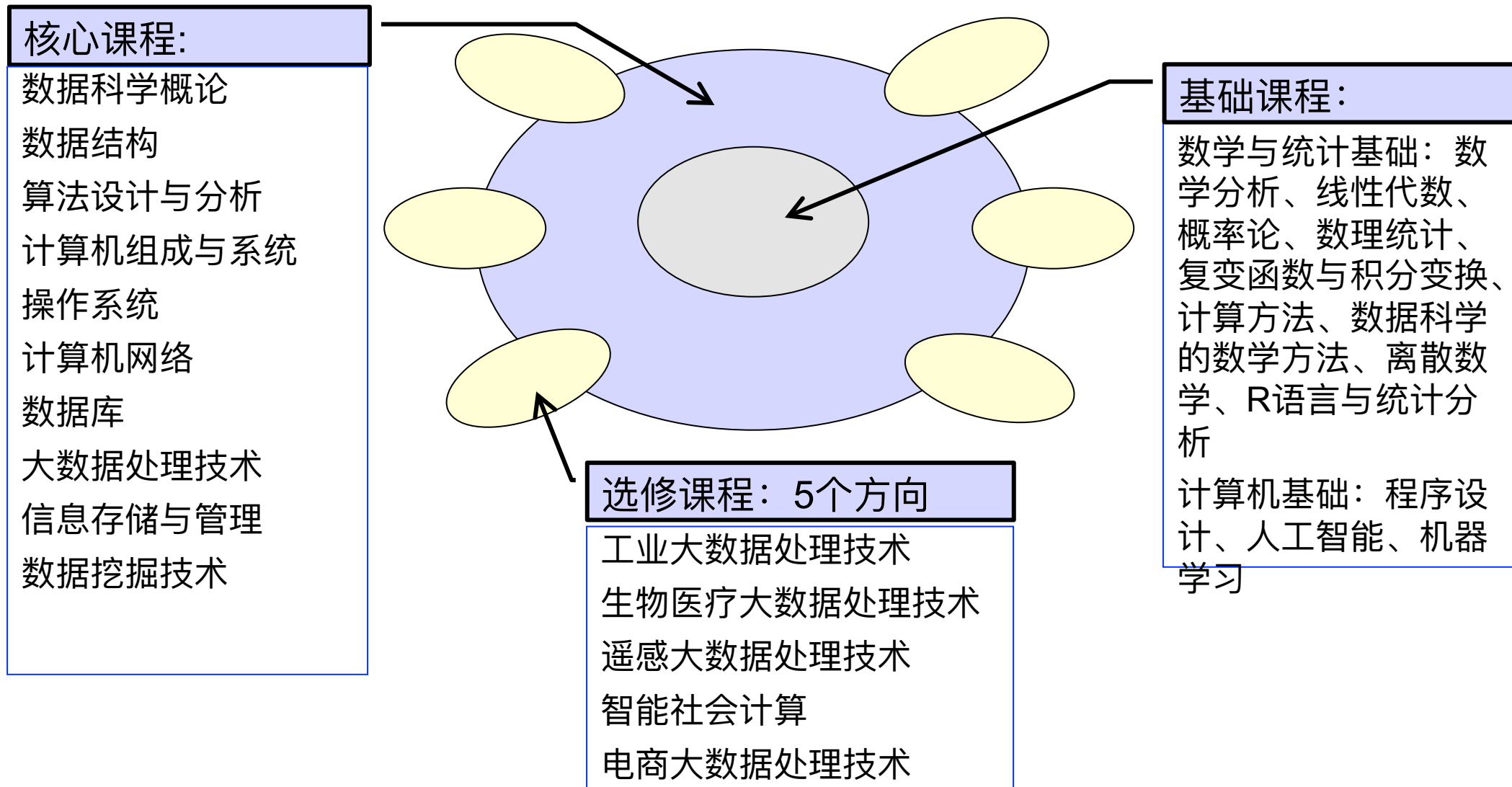


### 数据科学与大数据技术领域的快速发展与交叉性，为课程体系的设计造成了特别的挑战

- 以培养学生哲学思辨、数理逻辑为核心，建立适合数据科学与大数据技术专业的通识课程体系
  - 改变以传授已有“经过简化的、较为完整的学科概论或常识”为主的通识课程，强调学科内部知识之间建立联系
  - 改变只注重解题能力培养，强调对学生数理逻辑能力培养
- 以解决复杂工程问题为驱动，构建适合数据科学与大数据技术专业的专业课程体系
  - 注重计算机、统计、信息系统相关专业核心课程的深度融合
  - 加强应用数学，注重将理论学习与计算机实践的结合
  - 减少计算机核心课程，注重教学内容的系统性，去除冗余，突出重点
  - 坚持问题引导，关注实际应用，增大实践强度
- 以培养跨学科交叉能力为目标，构建支撑数据科学与大数据技术专业的前沿交叉课程体系
  - 强化既有优势，注重跨学科合作与交叉
  - 注重与领域知识的高度融合



### 学科协同创新、面向应用拓展的课程体系





### 开设特色基础课程，探索计算机、数学、统计的融合

#### □ R语言与统计分析

- 掌握经典的R语言编程技术法
- 用R语言解决实际问题的能力
- R语言介绍、数据类型、数据输入、科学绘图、基本数据操作、基本统计方法、参数估计和假设检验、方差分析、回归分析和相关性分析

#### □ 数据科学的数学方法

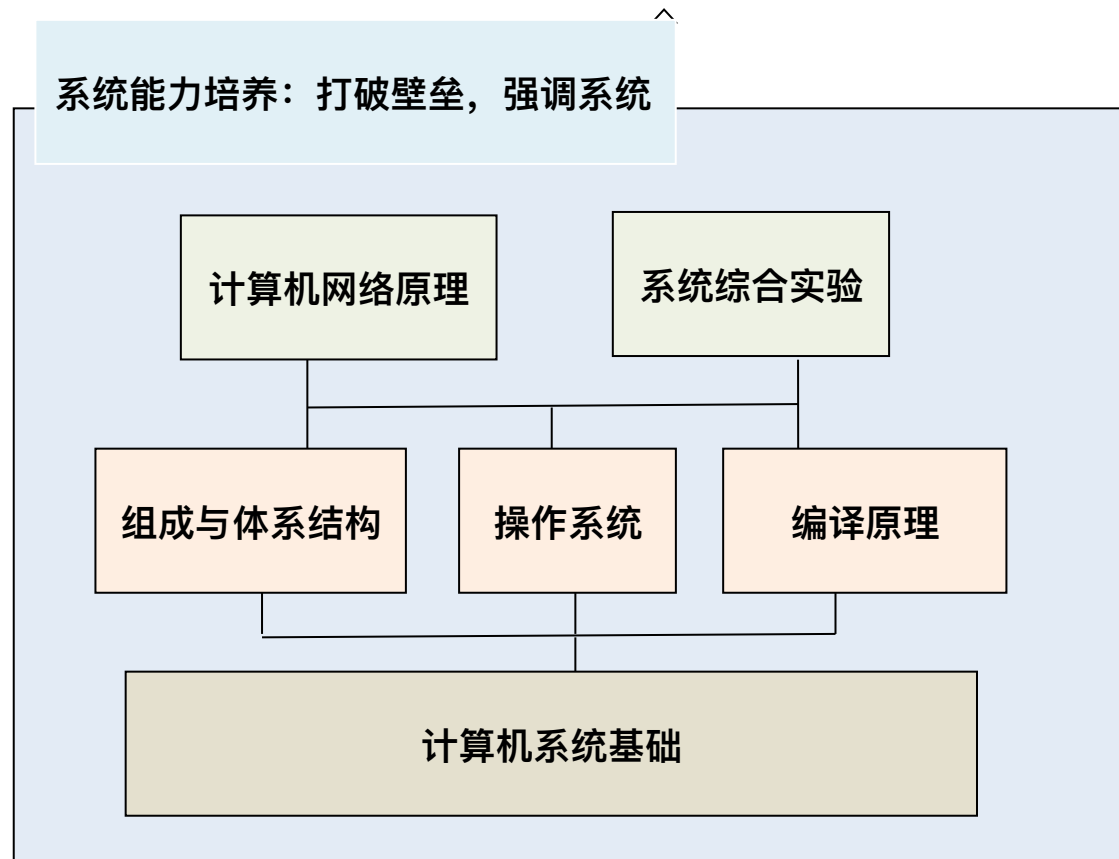
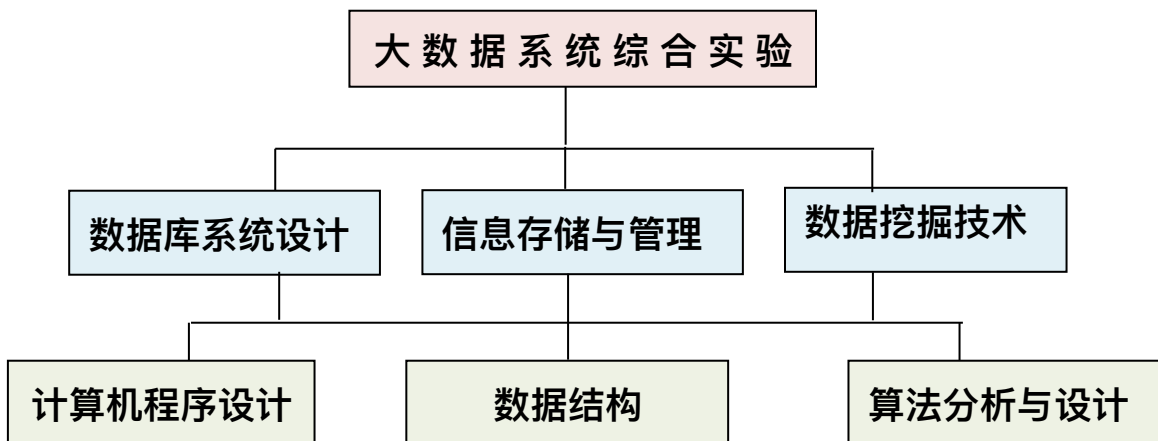
- 理解大数据分析中初级数学理论的基本知识和数学思想
- 掌握有关的计算方法及技巧，提高学生的数学素质，提高科研能力
- 矩阵论与最优化算法基础、概率分布、线性回归模型、线性分类模型、神经网络、核方法、支持向量机、图模型、混合模型和期望最大化、随机模拟技术、隐Markov模型和条件随机场模型



### 减少计算机核心课程，以解决复杂问题能力培养为统一视图，建立层次化、循序渐进、开放式计算机核心课程群

#### □ 重塑计算机核心课程群

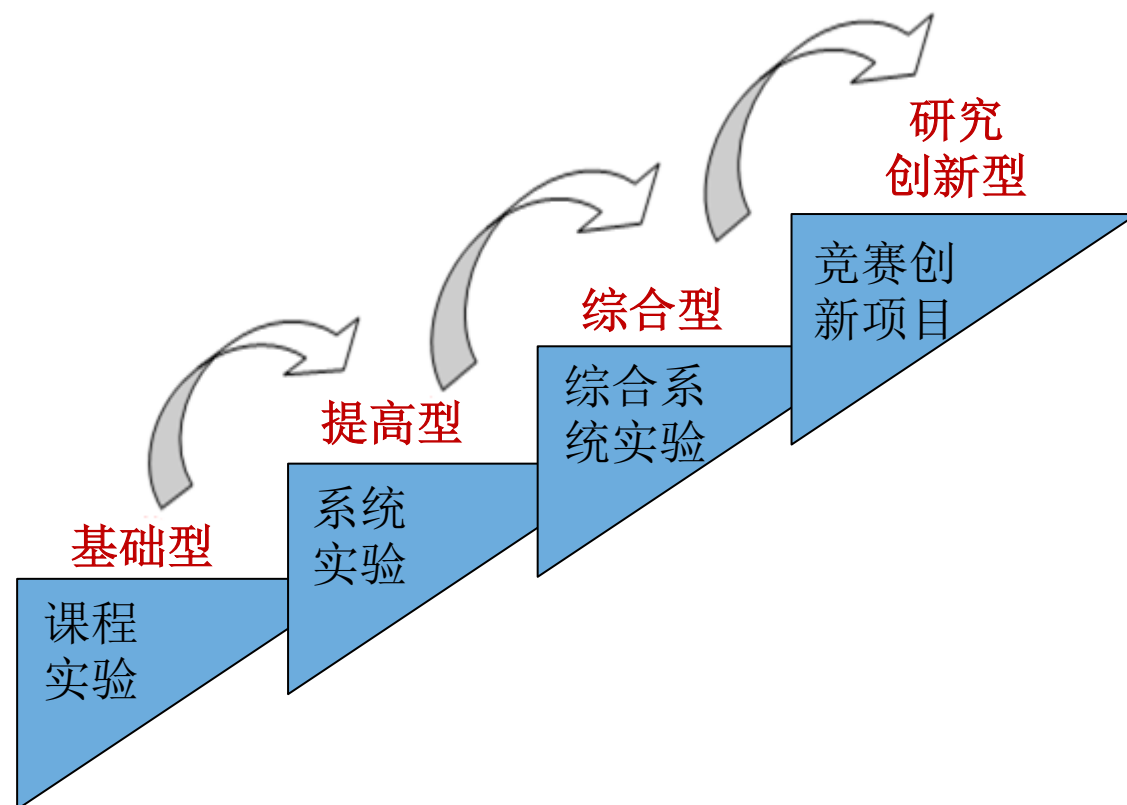
- 研究不同课程之间的逻辑关联性和系统的整体性，去除冗余，突出重点
- 构建构建统一实践教学平台，建立系统性实验课程体系



### 以系统级实验训练提高学生综合运用知识的解决复杂问题的能力

#### □ 突破传统模式，创造全面立体的工程体验环境

- 完善实践教学平台体系：加强校内平台建设。  
把教师的科研、学生的实习、实训有机地结合起来，将所学的理论知识和实验技能最终得到综合应用
- 优化实践教学内容体系：完善以“开放实验平台、创新竞赛结合、自主探索实践”为特色的系列化、层次化、个性化实验教学体系，创新实践教学手段，通过基础型、提高型、综合型、研究型和创新型实验项目的训练，促进学生复杂工程实践能力和创新能力培养





### 搭建产学协同育人平台，探索产教融合新模式

- ❑ 西工大-云创大数据联合实验室
- ❑ 西工大-腾讯媒体联合实验室
- ❑ 西工大-同盾语音联合实验室
- ❑ 西工大-中加设备健康管理实验室
- ❑ 。 。 。 。 。



**路虽远，行则将至！  
事虽难，做则必成！**

**谢谢！ 敬请批评指正！**

