



欢迎参加

第 1 届全国高校大数据教学研讨会

2017年5月12日-13日 厦门大学



第1届全国高校大数据教学研讨会 (BDTS2017)  
大会特邀报告



2017年5月12日至13日，第1届全国高校大数据教学研讨会（BDTS2017）在厦门大学科艺中心音乐厅隆重举行。本届研讨会由教育部高等学校计算机类专业教育指导委员会主办，厦门大学、厦门理工学院、贵州师范大学、人民邮电出版社联合承办，旨在搭建专业的大数据教学交流平台，汇聚全国高校大数据教学精英力量，共同探讨大数据专业和课程体系建设，为加快推进全国高校大数据教学发展贡献力量。来自全国300多所院校的400余名教师参加了本次研讨会。

厦门大学谭绍滨校长助理、人民邮电出版社教育中心营销部肖稳副主任，北京大学、中国科学院、厦门大学、华东师范大学、同济大学等重点院校的6位大数据教学知名专家，以及来自国内知名大数据企业的3名业界专家出席会议并做特邀大会报告。厦门大学林子雨助理教授主持会议。

更多内容请访问大会官网：<http://dbl原因lab.xmu.edu.cn/post/bigdata2017/>



华东师范大学 钱卫宁 教授 在做大会特邀报告

# 数据科学与工程专业课程体系建设 探索与思考



钱卫宁

[wnqian@dase.ecnu.edu.cn](mailto:wnqian@dase.ecnu.edu.cn)

華東師範大學





# 探索过程

6

## □ 2012.6 云计算与大数据研究中心

- ⊙ 学术伙伴、企业赞助、访问学者
- ⊙ **研究生培养**



## □ 2013.9 数据科学与工程研究院

- ⊙ 全国高校第一个大数据研究院
- ⊙ 学科交叉，应用驱动创新
- ⊙ 2014：完善 **“数据科学与工程” 二级学科**
- ⊙ 着手 **“数据科学与工程” 本科专业建设**



## □ 2016.9 数据科学与工程学院

- ⊙ 2017：招收 **“数据科学与大数据技术” 专业本科生**

# 互联网时代

7

搜 狐  
SOHU.com

新浪网  
sina.com

網 易  
NEWS

Baidu 百度

淘宝网  
Taobao.com

新浪微博  
weibo.com



支 支付宝  
ALIPAY

滴滴出行  
滴滴一下美好出行



# 互联网时代

8

搜 狐  
SOHU.com

新浪网  
sina.com

網 易  
NEWS

Bai du 百度

淘宝网  
Taobao.com

新浪微博  
weibo.com



支 付 宝  
ALIPAY

滴滴出行  
滴滴一下美好出行



□ 信息服务

□ 社交

□ 支付

□ 共享经济



# 互联网时代：Online2Offline

9

搜 狐  
SOHU.com

新浪网  
sina.com

網 易  
NEWS

Baidu 百度

淘宝网  
Taobao.com

新浪微博  
weibo.com



支 支付宝  
ALIPAY

滴滴出行  
滴滴一下美好出行



□ 信息服务

□ 广告

□ 社交

□ 电子商务

□ 媒体

□ 支付

□ 银行

□ 共享经济

□ 实体经济

# 互联网时代：Offline2Online

10



# 互联网时代

11

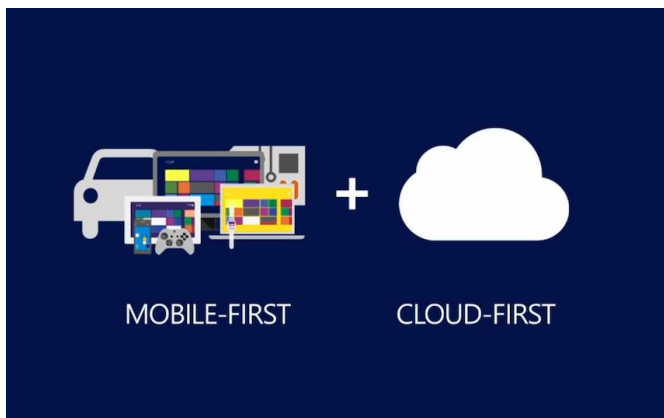
- 2015.3 国务院政府工作报告中提出 **“互联网+”行动计划**
- 2015.6 国务院正式出台  
《关于大力推进**大众创业万众创新**若干政策措施的意见》
- 2015.8 国务院印发《促进**大数据**发展行动纲要》
- 2015.9 国务院出台《关于加快构建**大众创业万众创新**支撑平台的指导意见》
- 2015.11 习近平总书记发表讲话首次提出  
**“供给侧结构性改革”**
- 2016.1 中央一号文件中写入 **“供给侧结构性改革”**

# IT 行业萧条

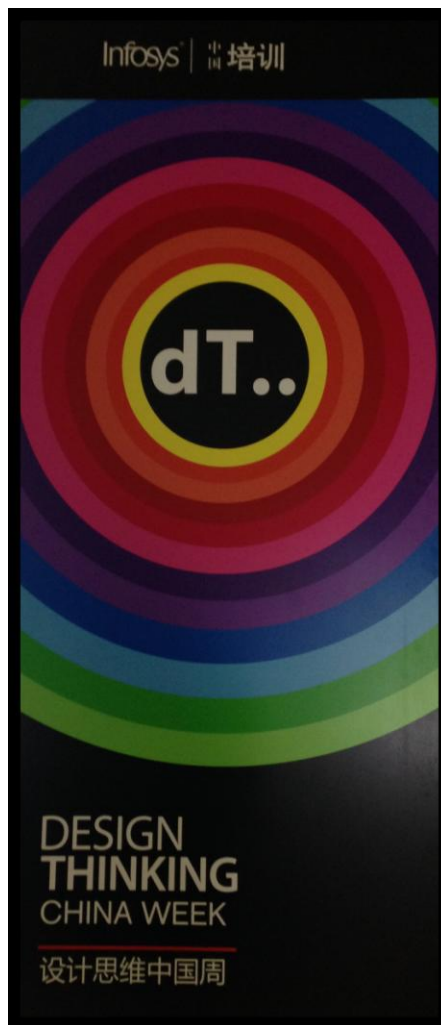
12



去 IOE 热潮  
(IBM,  
Oracle,  
EMC)



微软发展  
战略调整



Infosys  
转型

# IT 发展范型的变化

13

- IT 能力的建设不依赖于传统的IT企业 (表面)
- 整个 IT 界发展模式的变化 (深层)

## □ 传统方式

- ⊙ 垄断IT企业主导、应用单位削足适履
- ⊙ 硬件/软件/系统集成 (咨询、方案、外包)

## □ 互联网时代的方式

- ⊙ 应用驱动、度身定制
- ⊙ 融会贯通、开源系统 (Hadoop是典型代表)

- IT 发展模式的变化：非IT企业推动IT的发展
  - ⊙ 云计算和大数据 (**应用驱动创新**)
  - ⊙ 技术生态的建设 (开源万众创新)

# 我国 IT 发展范型的变化

14

## □ 我国 IT 行业的现状

- ◎ 严重依赖美国的垄断型企业 (IOE)
- ◎ 硬件/软件/系统集成比例失调

## □ 我国 IT 的新需求

- ◎ 夯实基础自成体系  
(摆脱对垄断的依赖)
- ◎ 跨越式发展  
(充分利用新硬件、新技术)
- ◎ 安全可靠、自主可控  
(网络空间安全)

# 我国 IT 发展的机遇

15

- 互联网企业的成功破除了迷信
  - ◎ 解放思想，增强信心
  - ◎ 开源社区，万众创新
- 我国独特且丰富的应用是源头活水
  - ◎ 应用驱动创新
- 硬件技术的进步提供了发展条件
  - ◎ 轻装前进，弯道超车

# IT 系统的变化

16

## 传统 IT 系统

- Enterprise-oriented
- 产品迭代周期长
- 用户群相对固定
- 质量要求高
- 研发成员只了解其中很小一部分
- **充满神秘感**

## 互联网 IT 系统

- Consumer-oriented
- 快速迭代
- 低成本
- 更强的组件化
- 可以从开源索取
- 庞大系统的搭建变成一件**切实可行的事**



# IT 系统的变化

17

## 回归本质，实践至上

- 摈弃了学科细分带来的只见树木不见森林的弊端，解决问题是根本目的
- 将教科书上的知识融会贯通

# “大数据”全景图

18

科学研究

企业管理

社会治理

在线教育

健康医疗

电子商务

搜索引擎

应用

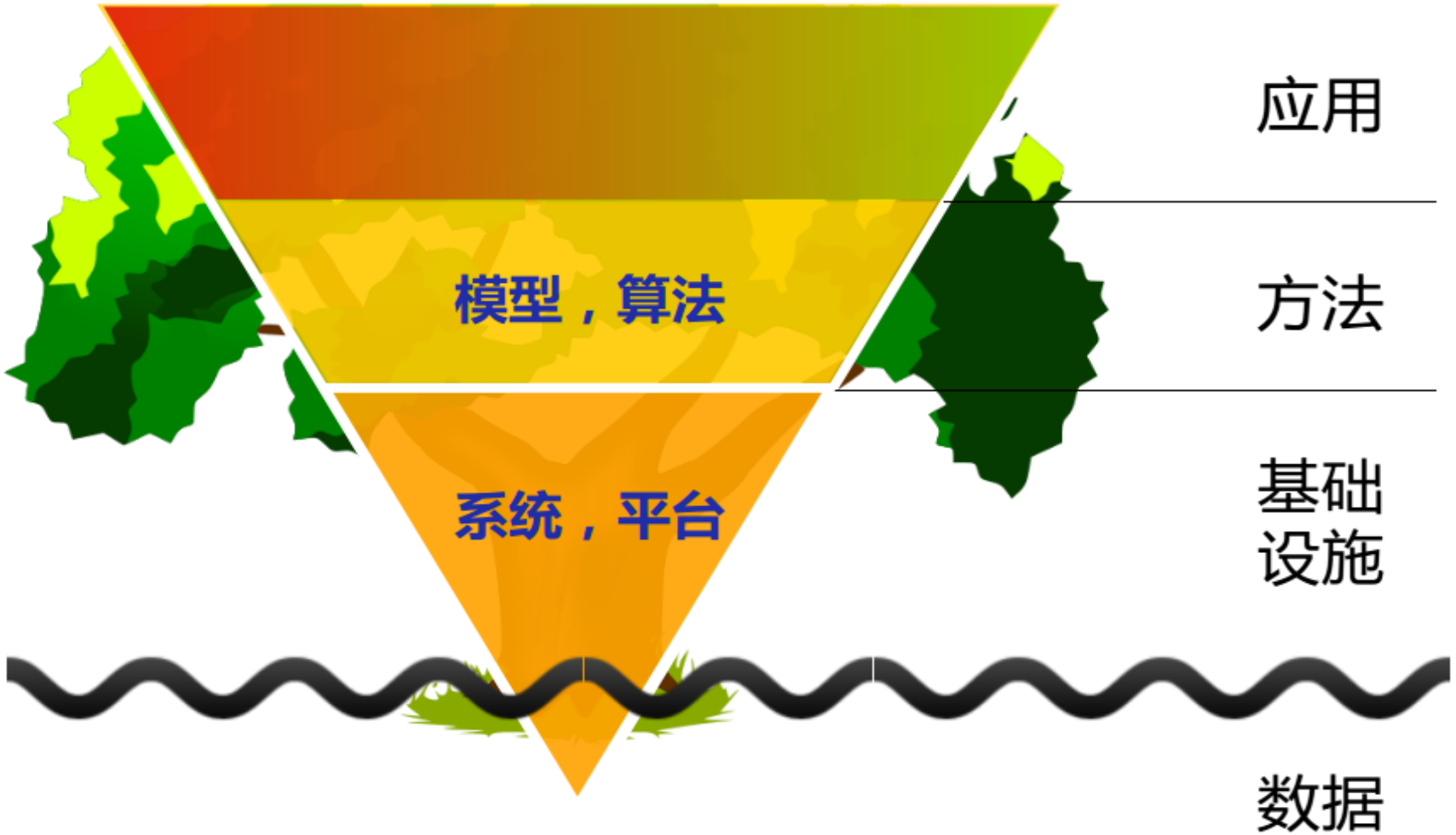
模型，算法

方法

系统，平台

基础设施

数据



# 大数据 vs. 互联网+

19

科学  
研究

企业  
管理

社会  
治理

在线  
教育

健康  
医疗

电子  
商务

搜索  
引擎

应用

模型, 算法

+

计算平台

互联网

# “大数据”人

20

## □ 应用层面

- ⊙ 行业人士、领域科学家等

## □ 算法层面

- ⊙ 人工智能、机器学习、数理统计、矩阵计算、优化等

## □ 基础设施层面

- ⊙ 计算机体系结构、网络架构、编程范式、文件系统、分布并行处理等

# 数据相关人才培养现状

21

- 计算机专业毕业生离市场需求有较大差距
  - ⊙ 企业通过校招找不到合适的人才
  - ⊙ 大量的毕业生找不到钟情的工作
- 计算机/应用数学/统计/信息系统学生对教学内容和方式革新充满期待
  - ⊙ 内容陈旧、远离现实、厌学厌教、恶性循环
- 在计算机专业基础上进行革新最为自然
  - ⊙ 注重系统实现、关注实际应用

# 数据人才培养的意义

22

□ 夯实基础、自成体系

**摆脱依赖**

□ 安全可靠、自主可控

**网络空间安全**

# 新专业：是 xx

23

## 培养目标

- **系统架构师、数据科学家**

## 知识结构

- **计算机+应用数学+信息系统** (领域应用)

## 课程设置

- 加强应用数学训练，**精简计算机课程**，注重系统理念培养

## 教学特色

- 注重应用，践行**设计思维**

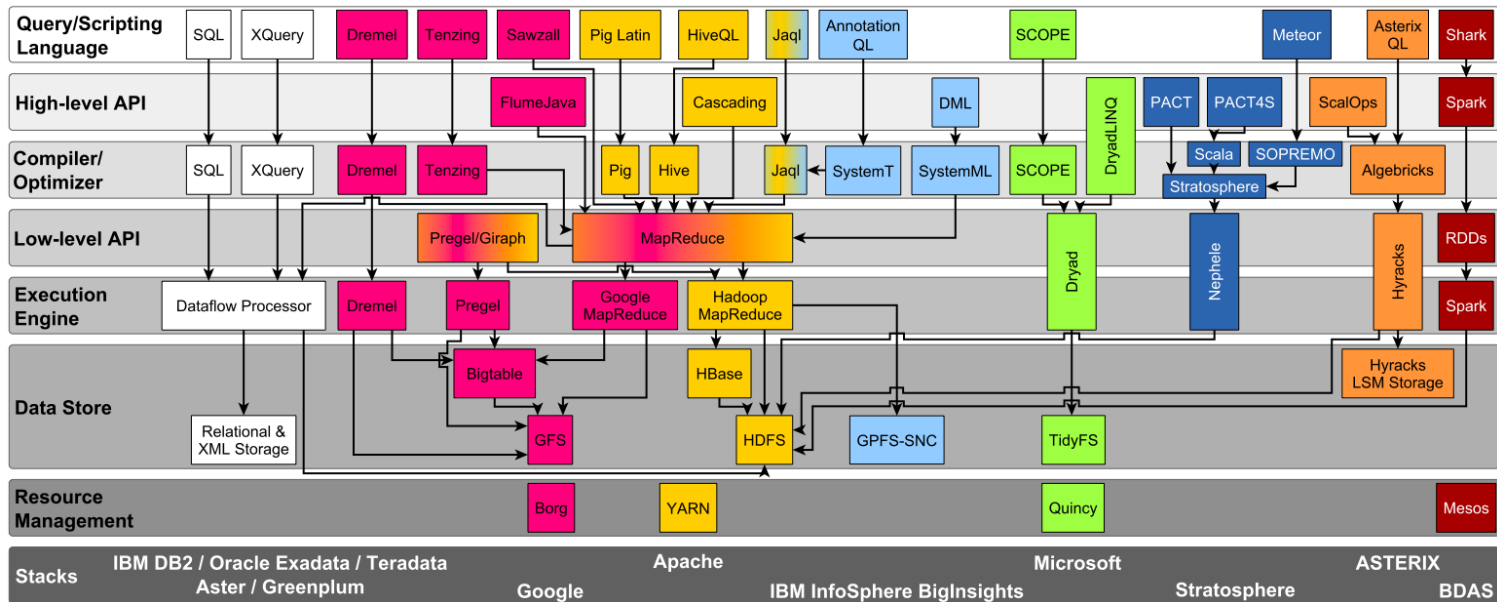
# 新专业：不是 xx

24

~~X 另一个计算机 / 软件工程 / 统计专业~~

- ◉ 原专业培训课程
- + 数据科学选修课 + 数据系统选修课

~~X 技术培训~~





# 新专业 vs. 旧专业

25

## 新时代的计算机专业

- 培养**开源**玩家
- 倒推知识点和课程设置

## 新时代的统计学专业

- 培养建模和算法**实现能力**
- 注重**数学知识**的应用

## 新时代的图书情报档案专业

- 培养信息管理的**全生命周期**理念和**设计思维**
- 注重信息的应用及其实效

# 课程体系：核心课程

26

- 计算机系统
  - ⊙ 计算机组成，体系结构，网络，分布式，云计算
- 操作系统
  - ⊙ 传统 OS，分布式，虚拟化 / 云计算
- 数据管理系统
  - ⊙ OldSQL，NoSQL，NewSQL，云
- 离散数学
  - ⊙ 集合论，数理逻辑，组合数学，图论
- 统计基础
  - ⊙ 数理统计基础，回归，多元统计分析
- 数据挖掘
  - ⊙ Python，数据生命周期全流程，机器学习初步，与大数据系统结合
- 分布式模型与编程
  - ⊙ 大数据系统，Hadoop，Spark，...
- 商业分析
  - ⊙ 含案例实践

# 课程体系：基础课

27

- 数据科学导论
  - ⊙ 数据结构、统计、数据挖掘等课程中的共性内容
- 工科高等数学
- 线性代数
- 离散数学
- 概率论
  - ⊙ 新的示例与习题
- 程序设计
- 数据结构
  - ⊙ 提供数据分析与系统研发两方面的编程知识储备

# 课程体系：专业选修课

28

- 软件工程
- 数据库系统实现
- 数据仓库与OLAP
- 信息检索
- 信息可视化
- 高等统计选讲
- 机器学习
- 统计学习
- 概率图模型
- 深度学习
- 计算机视觉
- 自然语言处理

# 课程体系：专业选修课

29

- 计算广告
- 智慧城市
- 社会计算
- 推荐系统
- 位置服务
- 电子商务
- 设计思维
- 开源软件设计与开发
- 云计算应用与开发
- 前端开发
- 大数据处理工具
- 统计分析软件  
( R语言 )

# 课程体系：研究生课程

30

课程性质	名称	备注
专业基础	数据科学数学基础	介绍数据科学领域的数学基础内容，包括高级统计、计算进阶。
	大数据处理系统	介绍新型的基于集群的大数据框架技术，包括Hadoop, Pig等
	数据科学算法基础	新的算法 / 数据结构，面向分析，包含：随机算法 / 近似算法初步，Bloom filter / Lsh等结构，含应用
专业选修	机器学习	介绍机器学习的高级内容
	高级数据库系统	介绍数据库系统实现以及理论方面的内容
	数据存储系统与技术	介绍与存储相关的技术
	统计推断	介绍统计里面的推演技术
	事务处理	NewSQL
	分享型数据库	区块链、数据库安全等
	海量数据挖掘	数据挖掘算法
	项目管理与开源社区	介绍与项目管理、开源相关内容

# 理论学习和实践教学

31

## □ 多级实验和实践设计

- ◎ 课程实验
- ◎ 课程大作业
- ◎ 企业真实问题实践
- ◎ 研究生科研实践：长期、多轮实践
  - 企业项目实践
  - 纵向课题实践
  - 打破 1 年级修课、2 年级科研、3 年级实习找工作的模式

# 本-硕-博贯通培养

32

- 知识点和能力素质梳理、教学内容裁剪和组合、培养目标和课程设计
- 阶梯状、螺旋式上升

## 示例

- 数据挖掘=>
- 分布式模型与编程=>
- 大数据处理工具=>
- 大数据处理系统  
<=数据管理系统 / 操作系统

## 示例

- 数据科学导论=>
- 统计基础=>
- 高等统计选讲=>
- 统计学习<=机器学习
- 概率图模型<=
- 统计推断<=数据科学的数学基础



# 招生与遴选

33

- 从转专业、插班生中招收本科生
  - 理工科基础
  - 已完成高等数学学习
  - 有一定线性代数、编程基础者优先
  - 笔试：基础知识、思维能力、表达能力
  - 逐一面谈，多方考察
- 通过夏令营招收硕士或长学制研究生
  - 全面考察基础知识、动手实践和综合能力
- 通过考核制招收博士研究生
  - 通过一个较长的周期，考察、选拔学生

# 师资

34

## □ 学科背景

- ⊙ 计算机 / 软件工程
- ⊙ 统计学
- ⊙ 计算数学
- ⊙ 管理信息系统 / 图书情报

## □ 师资来源

- ⊙ 全职
- ⊙ 兄弟院系
- ⊙ 企业

# 小结

35

- 应用驱动创新：面向我国各行业、领域需求
  - ◎ **新工科**：“互联网+”和“大数据”背景下的行业变革
- 培养“**造车的**”，而不仅仅是“驾驶员”
- 以培养“**系统架构师**”和“**数据科学家**”为目标
- 拥抱**云计算**，拥抱**开源**
- **交叉**学科、**学科交叉**

# Thanks!

36



钱卫宁

中国

**DaSE**  
Data Science  
& Engineering

扫一扫上面的二维码图案，加我微信