

数据学院在数据科学与工程教育上的 的实践与思考



周烜

xzhou@dase.ecnu.edu.cn

華東師範大學



探索过程

2

□ 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**



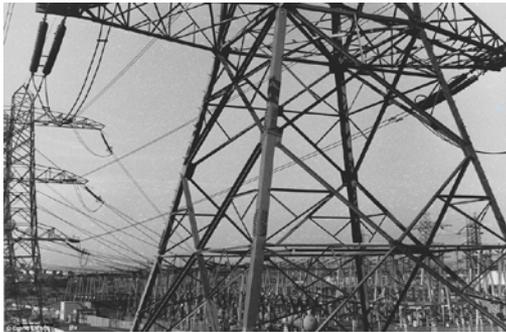
□ 2016.9 数据科学与工程学院

- ⊙ 2017：招收 **“数据科学与大数据技术” 专业本科生**

Vision: Data is the New Power

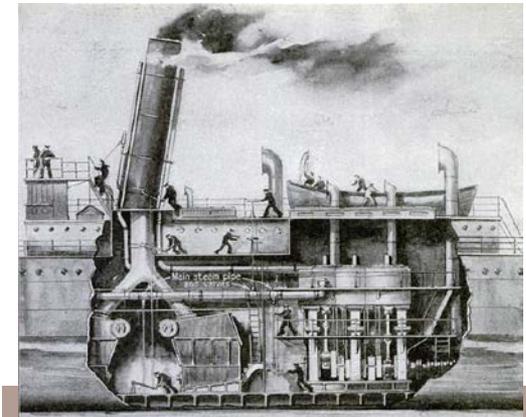
3

1880s - 1990s
electric power
电能



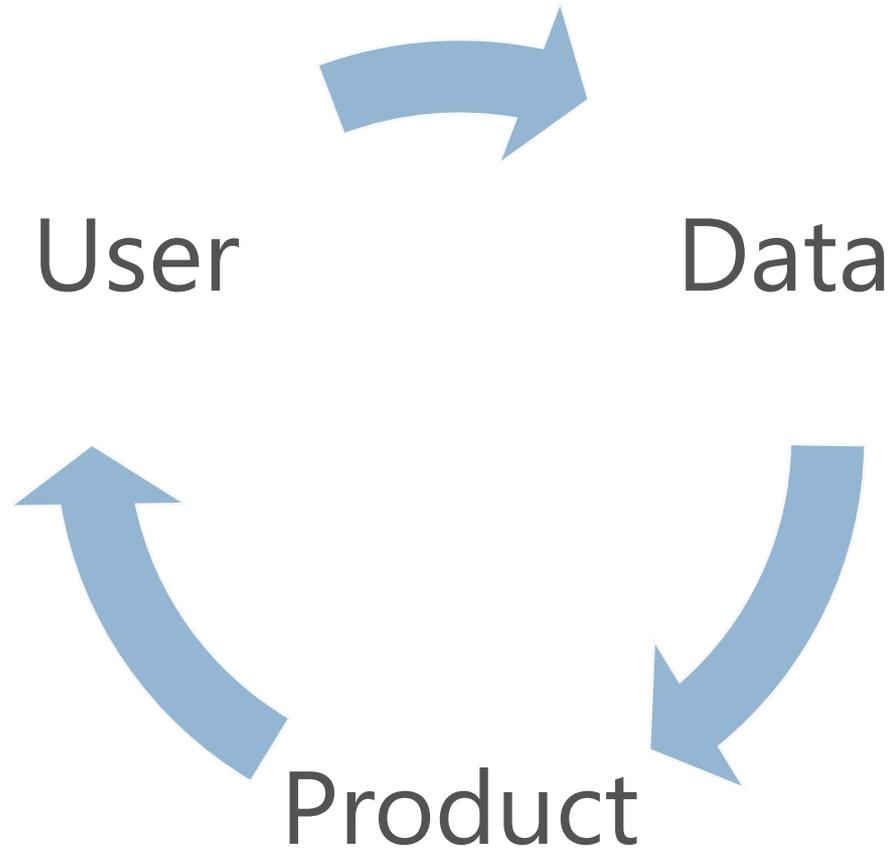
2000s - 未来
data power
数据能

1760s - 1870s
steam power
蒸汽能



Andrew Ng: AI Company Model

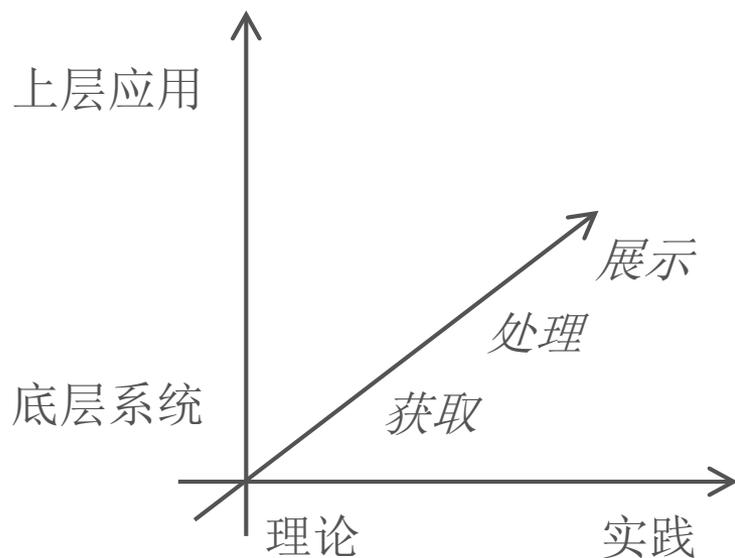
4



数据专业的目的

5

- 培养学生运用“数据能”的能力
培养学生实现数据价值的能力



数据应用的体系

6

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

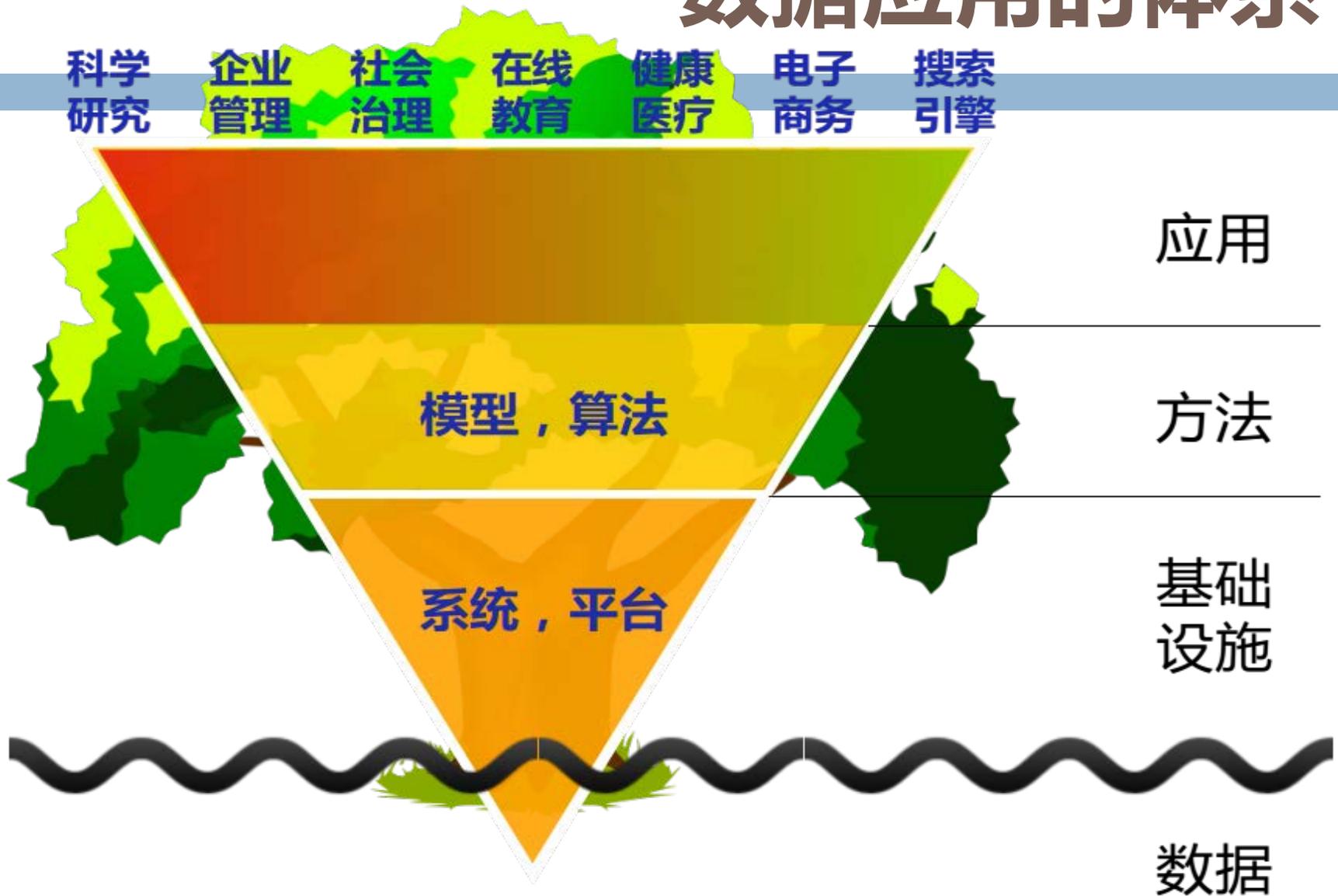
模型，算法

方法

系统，平台

基础
设施

数据



数据的生命周期 – 过程

7

- 数据获取
- 数据治理
- 数据管理
- 数据处理
- 数据可视化

理论到实践

8

- 数学基础
- 统计理论知识
- 计算机理论知识
- 算法基础
- AI理论
- 程序设计
- 软件开发
- 系统工具的使用
- 系统工具的实现
- 应用的理解
- 领域的理解
- 产品的设计与规划

Vision: IT 发展范型的变化

9

- 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)

互联网时代：Online2Offline

10

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 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

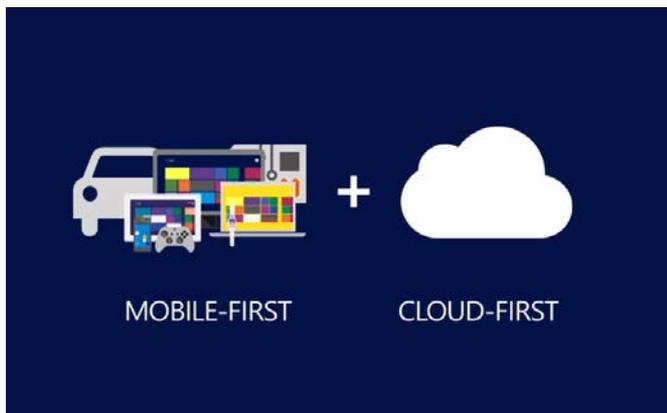
□ 实体经济

IT 行业萧条

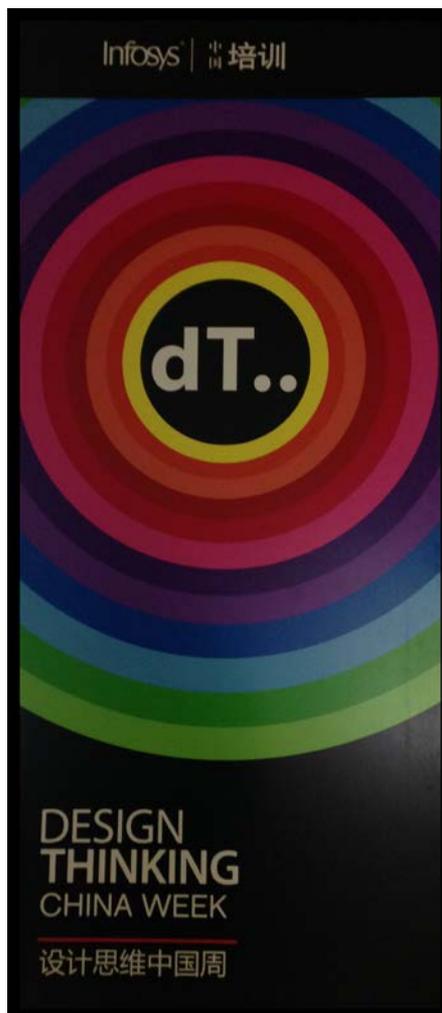
11



去 IOE 热潮
(IBM,
Oracle,
EMC)



微软发展
战略调整



Infocys
转型

IT 系统的变化

12

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**

数据相关人才培养现状

13

- 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

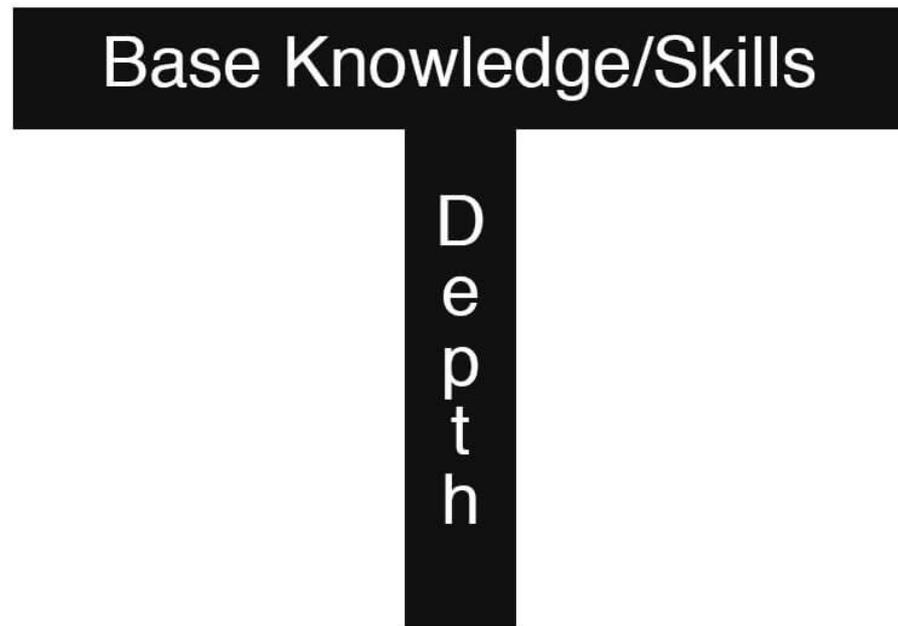
14

回归本质，实践至上

- 摒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
- 全栈式人才

数据人才培养

15



T-Shaped Person

数据学院的培养体系设计

16

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 vs. 旧专业

17

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

18

- 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
- 数据科学的数学基础
 - ◎ 集合论, 数理逻辑, 图论, 运筹学, ...
- 统计基础
 - ◎ 数理统计基础, 回归, 多元统计分析
- 数据分析
 - ◎ Python, 数据生命周期全流程, 机器学习, 大数据处理
- 分布式模型与编程
 - ◎ 大数据系统, Hadoop, Spark, ...
- 应用
 - ◎ 案例实践、数据伦理

理论学习和实践教学

19

- 多级实验和实践设计
 - ◎ 课程实验
 - ◎ 课程大作业
 - ◎ 企业实践
 - ◎ 研究生科研实践：长期、多轮实践

招生与遴选

20

- 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

我们的经历

21

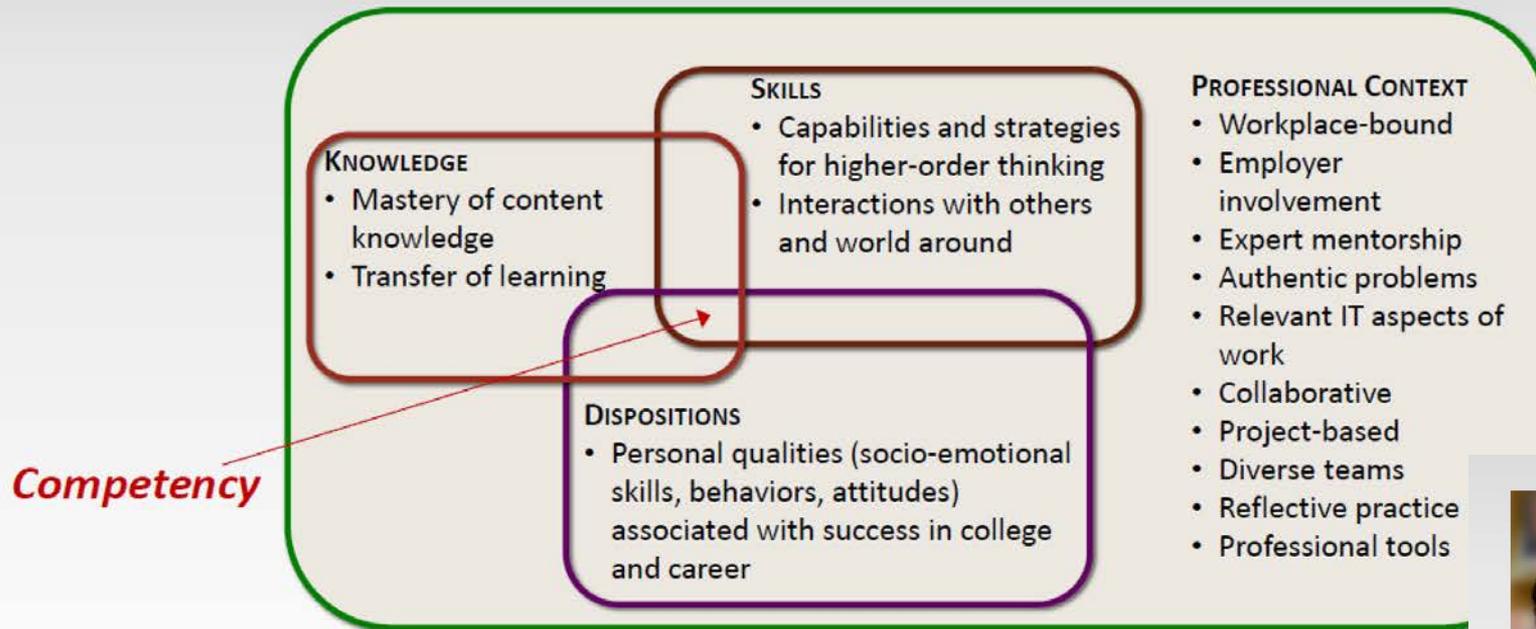
- 2017年不到30人报名，录取24人，一半同学不具备编程基础
 - ◎ 暑期程序设计强化训练
 - ◎ 第一学年高负荷、高要求的课程
 - ◎ 学生倍感压力（1人中途退出）
 - ◎ 收获多、成就感强
- 2018年超过60人报名，录取34人

一些实际的普遍的问题

22

- 大部分教师为学术背景，工程、应用经验欠缺
 - 我们强调工程能力、应用驱动创新的能力。然而缺少相应的教学素材。
- 协作能力的培养缺少办法
 - 处理数据需要理解数据，需要领域专家的帮助
 - 理解数据的含义，理解数据的价值所在
 - 如何与领域专家协作？有方法论吗？

Competencies in IT Context



John Impagliazzo

小结

24

- 教学目标：**数据能**的运用
- **应用驱动创新**：面向我国各行业、领域需求
 - ⊙ “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**

Thanks!

25

