



清华大学  
Tsinghua University



# 清华大学大数据人才培养项目

王建民

清华大学数据科学研究院

2016年10月15日

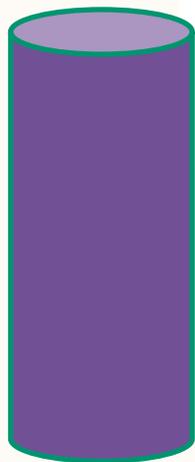


# 大数据人才需求

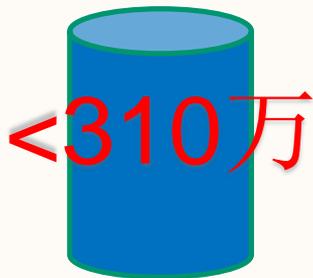
Gartner®

2015年全球大数据人才需求

440万



缺口超过 **130万**



全球大数据人才总数

国内形势

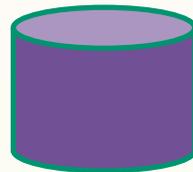
CSDN

全球最大中文IT社区

缺口达到 **150万**

未来3-5年，中国需求

**180万**



**~30万**

目前中国大数据人才总数

国际形势



# 大数据人才分类

## A 应用数据分析师

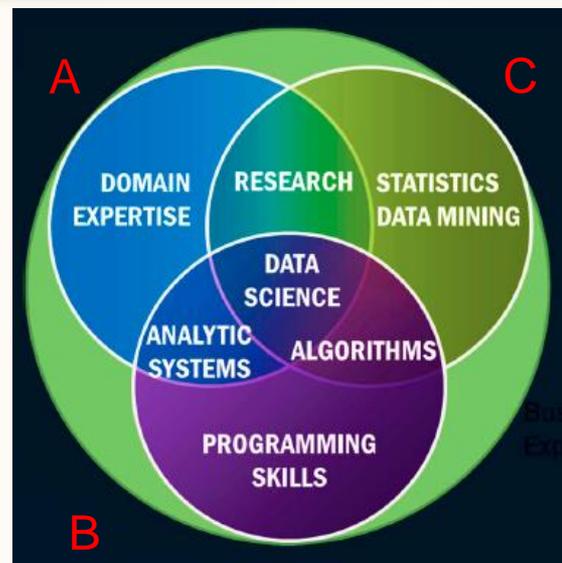
具备行业应用领域知识，提供基于数据挖掘分析的战略决策

## B 数据系统构架师

对数据的存储、分析与处理等技术和系统构架有深入理解

## C 算法设计工程师

注重数据分析算法设计，统计分析为核心的数学能力突出



- ◆ 清华大学大数据硕士项目
- ◆ 清华大学大数据能力提升项目
- ◆ 清华大学数据科学与工程专业硕士项目





# 清华大学大数据硕士项目

## (学位项目)



## ➤ 酝酿提出

2013年12月3日，研究生院召开大数据研究和人才培养研讨会。酝酿由研究生院统筹校内外优势资源，依托清华大学工科和商科优势，以及社会科学在大数据环境下的战略机遇期，以及充分利用清华大学的业界资源，成立数据科学研究院。

## ➤ 目标

培养一批具有**大数据思维、运用大数据思维及分析应用技术**的高层次**大数据人才**，引领中国乃至世界的大数据研究和开发应用。





## ➤ 方针确定

2014年1月2日，陈吉宁校长主持召开清华数据科学研究院筹建会议，全校相关院系和部门等参加。提出“**学校统筹、问题引导、社科突破、商科优势、工科整合、业界联盟**”24字方针。



## ➤ 培养框架建构

- 确定大数据硕士项目总学分不少于30学分，采用统一培养方案，分设数据科学与工程、大数据与国家治理、经济社会数据、互联网金融、商务分析等若干不同的培养方向。
- 设置大数据数据分析学（3学分）、大数据系统基础（3学分）、大数据思维与行为（2学分）、大数据管理与创新（2学分）、大数据治理和政策（2学分）等5门必修课，大数据科学选修课（3学分）和大数据实践（1-3学分）必修环节，核心课程由几个学院分别牵头、共同设计。



## ► 清华大学数据科学研究院成立 启动大数据人才培养

- 2014年4月26日下午，清华大学与青岛市人民政府签署共建合作协议，清华-青岛数据科学研究院正式成立。同时，清华大学宣布将推出多学科交叉培养的大数据硕士项目。
- 2014年9月，第一届大数据硕士项目开始招生，统筹协调7个学院。计划每年招生150人。



## ➤ 大数据硕士项目参与院系和研究方向

- 信息学院、交叉信息研究院、深圳研究生院（数据科学与工程）
- 经管学院（商务分析和金融硕士保险方向）
- 公管学院（大数据与国家治理）
- 社科学院（经济社会数据）
- 五道口金融学院（互联网金融）



## ➤ 课程设置

课程名 (课号)		学分
大数据分析A (60700033) 或B (60240103)	必修	3学分
大数据系统基础A (60470013) 或B (64100033)	必修	3学分
数据伦理 (60250121)	必修	1学分
大数据科学与应用系列讲座 (60250131)	必修	1学分
下面三门课程三选一:		
数据思维与行为 (60700052)	必修	2学分
大数据管理与创新 (60510202)	必修	2学分
大数据治理与政策 (80591342)	必修	2学分
大数据平台核心技术	选修	2学分





# 案例：《大数据系统基础A、B》

## ➤ 课程特色

- ❑ 信息类和非信息类同学交叉混合组队，跨学科团队合作
- ❑ 引入大数据企业实际项目，和真实用户交流
- ❑ 企业导师和授课老师搭配，共同指导学生
- ❑ 大数据系统设计和实现能力体现



## ➤ 课程征集2类项目

### □ 偏重系统架构

- 如何持续、高效地进行大数据分析
- 如何整合不同的数据集

### □ 偏重数据分析和应用

- 如何通过一个数据集高效地分析出结果

2016年《大数据系统基础A、B》共征集16家企业，26个项目



	企业名称	项目名称
1	清华海峡研究院	基于卫星数据获得逐日云量数据实验
2	北京鼎泰智源科技有限公司（法海风控）	行业风险分析
3	北京易格在姆信息技术有限公司	语义查重项目
		场次预编排项目
4	北京屏芯科技有限公司	菜品销量及定价分析
		餐饮商家营业数据（营业额、订单数、客流）预测
5	北京市一路热点信息技术有限公司	用户的出行线路分析
6	广东万丈金数信息技术股份有限公司	文本相似度建模在保险行业的应用
7	璧合科技	NLP(自然语言处理技术)在提升DSP展示广告点击率预估方面应用
8	元码基因科技（北京）有限公司	基于高通量测序数据进行个性化医疗
9	北京地米科技有限公司	基于海量数据的病案首页数据核查平台
10	北京超萌国际文化有限公司	超盟零售数据挖掘
11	优选创新（北京）科技有限公司	北京租房数据分析
12	国家发改委城市和小城镇改革发展中心	城市（区域）竞争力评价体系研究
13	北京奥维云网大数据科技股份有限公司	关于家电评论文本多分类项目
		文本信息的创意挖掘
		人口属性识别
14	腾云天下（TalkingData）	时间序列分析和预测
		app使用生命周期分析和分类
		图分析实践- POI和用户关系挖掘
		用户轨迹分析、预测和可视化
15	清华大学遥感大数据中心	基于机器学习的全国主要农作物提取
		基于深度学习的全球卫星降水产品研制
		基于遥感影像的目标地物提取
16	路通世纪（中国）科技有限公司（汤森路通）	推特（Twitter）情绪分析算法研究
		公司基本面数据与股价相关性研究

## ➤ 实验与考核

- ❑ 由AB课程学生结合的小组进行，完成真实的项目需求
  - ❑ 为每个团队提供5台服务器（虚拟机）
  - ❑ 实现持续、高效地大数据整合、分析等
- ❑ A课程考核：侧重**文档**
  - ❑ 开题报告：需求分析文档，产品背景调研，相关案例
  - ❑ 阶段性报告：代码、实验、数据、团队管理过程文档
  - ❑ 期末报告：系统架构完成程度，系统性能评估，应用展示
- ❑ B课程考核：侧重**实践**
  - ❑ 开题报告：技术路线调研，初步设计
  - ❑ 阶段性报告：代码的正确性与可读性，项目进展
  - ❑ 期末报告：系统架构设计的合理性，系统性能，应用展示

## ➤ 项目答辩



“脑洞”大开，收获巨大





清华大学数据科学研究院  
Tsinghua Institute for Data Science

# 清华大学大数据能力提升项目 (证书项目)



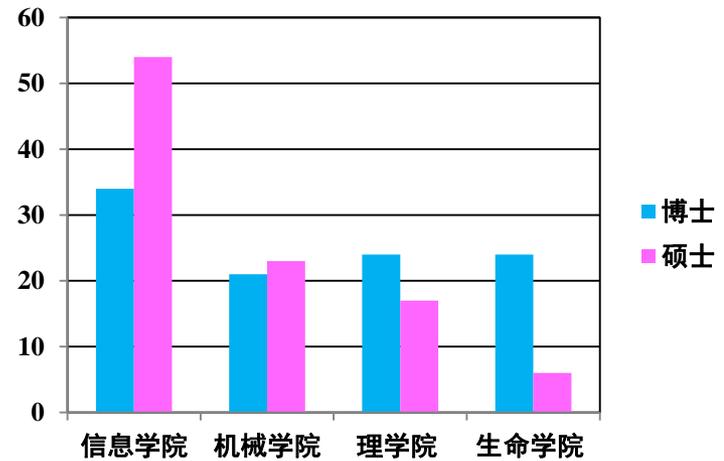
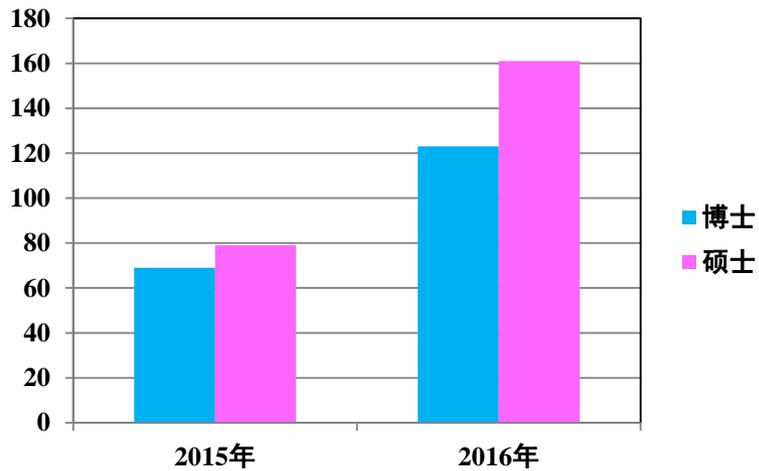


## ➤ 清华大学大数据能力提升项目官方发布

2015年12月4日，清华大学研究生院、清华大学数据科学研究院共同发布清华大学大数据能力提升项目，面向全校在学研究生招生。

截至2016年9月30日统计，共有432名学生报名，其中博士研究生192人，硕士研究生240人，来自全校23个院系。报名人数排名前四名院系分别是信息科学学院、机械工程学院、理学院、生命学院。





2015年，博士69人、硕士79人，2016年，博士123人、硕士161人；

信息学院88人，机械学院44人，理学院41人，生命学院30人



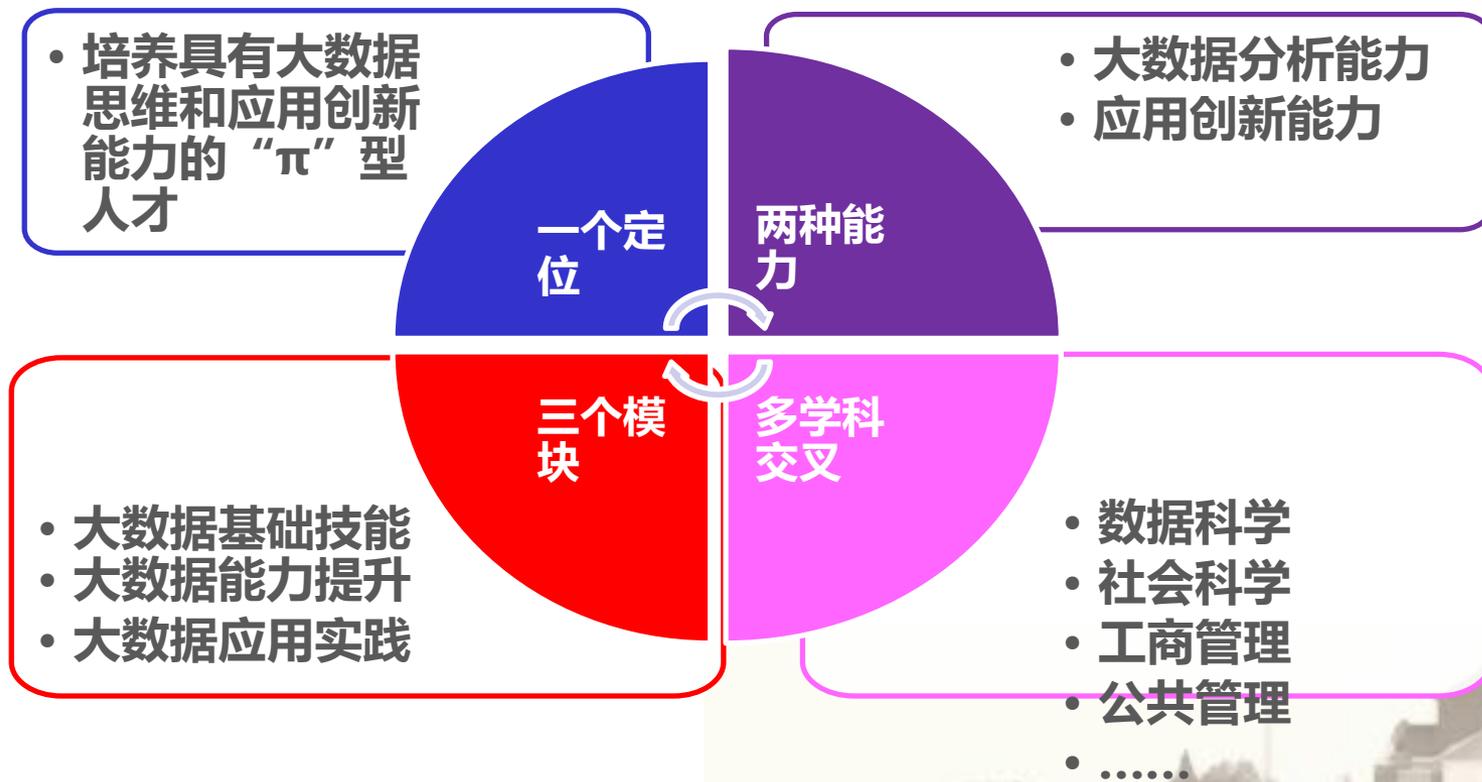
## ➤ 项目特点

- **不限学期和学年**，毕业前一个学期完成要求的学分，可获得证书；
- 突出**创新性和实战性**，引入大数据企业实际项目，提供丰富实践机会；
- 突出跨学科**交叉融合**

➤ **招生对象：** 在学研究生(博士和硕士)



## ➤ 项目线上线下混合式互动教学新法



## ➤ 证书要求

- 总学分 $\geq 10$ 学分

大数据基础技能模块： $\geq 3$ 学分

大数据能力提升模块： $\geq 4$ 学分

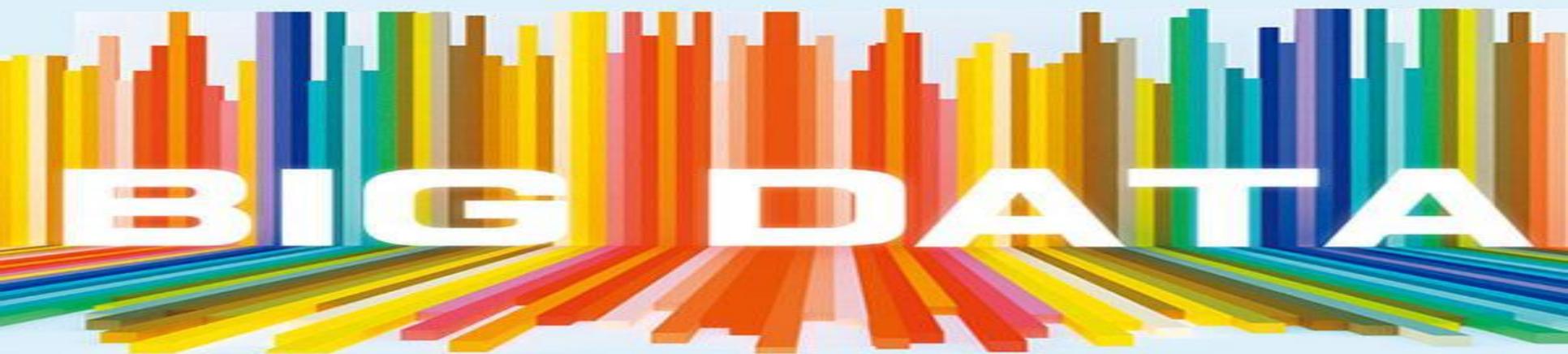
大数据实践模块：必修（3学分）

- 基础技能模块学分不可由能力提升模块替代
- 累计至少参与8次及其以上实践活动





# 清华大学数据科学与工程专业硕士项目



## ➤ 项目目标

**顺应**互联网+时代国家治理、产业经济、政府决策、社会运行等模式变革而诞生；

**培养**具备数据存储、运行监管、智能分析挖掘以及战略决策等技能的专门人才；

**胜任**数据存储管理师、数据分析师、数据系统架构师乃至数据科学家、首席数据官、商务分析师、战略管理者等岗位；



## ➤ 项目启动

- ❑ 国内首个混合式教育学位
- ❑ 2015年5月7日新闻发布会





## ➤ 项目影响力

sina 新闻中心

新闻 人民币干了一件大事

# 清华读研 可用手机上课 “数据科学与工程” 是国内首个混合式教育项目

2015年05月07日14:39 法制晚报

清华混合式模式培养大数据工程硕士  
在线学习也能拿清华学位

【浏览字体：大 中 小】 2015-05-08 文章来源：《中国教育报》 分享

本报北京5月7日讯（记者 万玉凤）通过一根网线，就能见到清华名师，在线学习也能拿清华学位，这个愿望正在变为现实。今天上午，清华大学宣布启动国内首个混合式教育的硕士学位项目——“数据科学与工程”专业硕士学位项目，今年将开始招生，首批50名学生将于2016年入学。

知乎

搜索你感兴趣的内容...



首页 话题 发现

数据挖掘 MOOCs

### 如何看待清华“数据科学与工程”硕士项目？

今天晚上学堂在线时发现了这个新的项目，和传统不同的是线下与线上MOOC结合，并针对企业提供证书项目。地址：[清华大学混合式教育学位项目新闻发布会](#)  
其实还想问，这对于更多非清华的然而完成了学堂在线等MOOC平台的学生而言，会有多大程度上得到更多的国内的面试官了解乃至认可呢？

## 清华首推“数据科学与工程”专硕 专家名师汇聚新项目

➤ 项目特色：基于MOOC的翻转课堂模式



## ➤ 以学习和实践能力为考核标准的招录环节改革

研究生入学  
考试



专业笔试



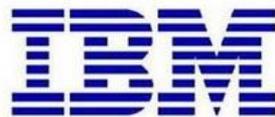
综合面试



2015年, 共有281人报名该项目并参加GCT考试, MOOC学习报名人数80人, 最终共有112人进入复试, 其中MOOC学习者28人。

## ➤ 企业合作

- 清华大学-腾讯科技（深圳）有限公司互联网创新技术联合实验室
- 清华大学-搜狐搜索技术联合实验室
- 清华大学-HP多媒体联合实验室



## ➤ 问卷反馈（大数据系统基础B- 2015-3-7）

- ❑ 大部分同学对课堂氛围还算满意（占比70.59%），但是对于教师在课堂上使用的教学方法只有47.06%认为满意，52.94%的人认为一般
- ❑ 对于多名教师共同授课的方式，52.94%的人认为合理，35.29%的人认为一般，11.76%的人认为不合理。同学反映这种方式课程连续性差，不成体系，前后课程内容会有重复，课堂纪律和课后复习缺乏足够的监督。
- ❑ 关于企业调研方面基本上都认为很有必要。



## ➤ 若干挑战

- 新专业？老专业？
- 学生知识基础差别巨大
- 人才需求尚未定型





## ➤ 致谢

感谢清华大学数据科学研究院的罗国荣老师、软件学院学院宋韶旭老师提供了相关材料。

本PPT版权归清华大学研究生院和数据科学研究院所有。





NDBC2016全国高校大数据教学论坛主页

<http://dmlab.xmu.edu.cn/post/ndbc2016bigdatateachingforum/>



扫一扫访问论坛主页下载更多内容