

# 专业数据人才教育行业生态报告



2018年6月

机密和专有  
未经腾云大学和美世的许可  
任何对此资料的使用严格禁止

# 前言

随着大数据、人工智能、区块链等前沿技术的迅速普及，企业越来越重视前沿技术给企业带来的非凡的竞争优势。根据全球领先人才社区 LinkedIn 的《全球 AI 领域人才报告》，转型技术驱动的企业之间的竞争将最终升级为对核心的人才的竞争（见图 1）。根据著名咨询公司美世的数据显示，大数据/人工智能/区块链等先进行业的核心人才的薪资水平远远高于传统 IT 或咨询分析部门的同等级专业人才水平（见图 2）。企业在日益重视核心人才的引入的同时，也面临着巨大的成本压力。外部高科技人才落地给企业带来巨大竞争力的同时，也意味着风险——海归人才难以适应国内的商业环境 and 企业文化等等——企业技术人员把新兴技术看作可能在未来对他们的就业产生最大影响的威胁等。目前在实现利益和规避风险方面，大多数企业仍然做得不足，企业自主的人才教育成为了当前企业培养自己的前沿技术核心团队的重要方法。

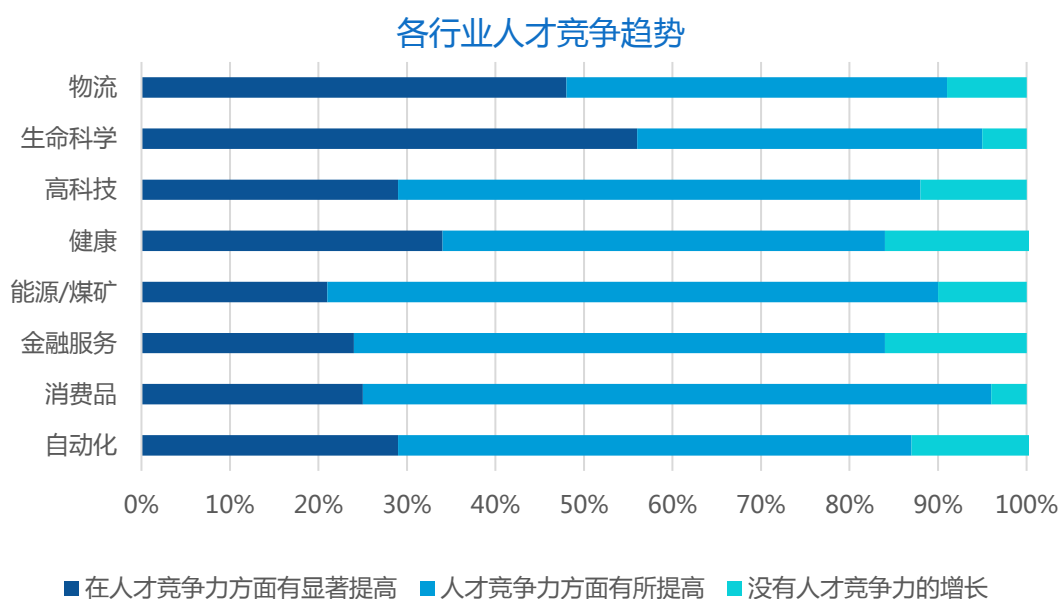


图 1 跨行业竞争 – 人才的竞争  
(资料来源：美世 2017 全球人才趋势报告)

抛开数字化经营，企业就像是在黑暗中对别人微笑。只有自己知道自己在做什么，但是别人不知道。

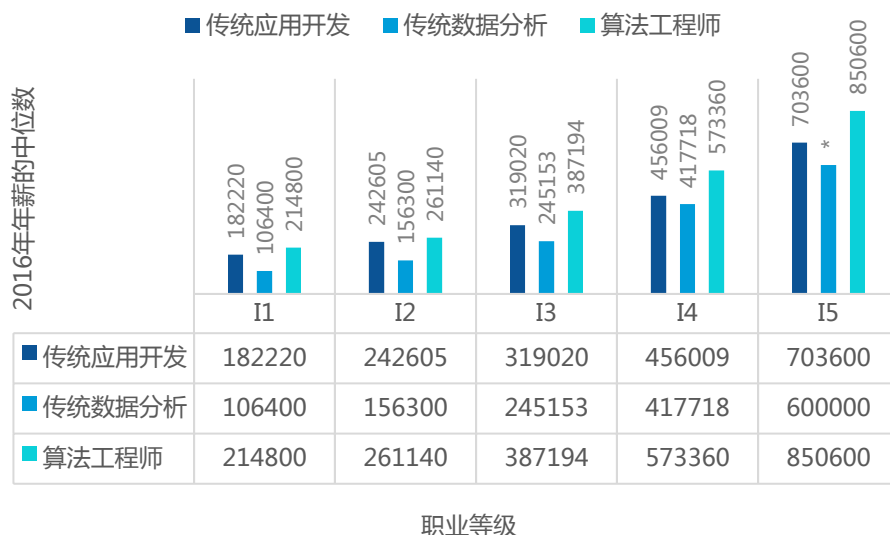


图 2 传统应用开发、传统数据分析与算法工程师的年薪对比。

\*号代表调研数据缺失，后引入专家评论预估数据。

(数据来源：美世咨询 2017 年互联网行业薪酬调研数据)

在当今社会，随着数据行业的变化，更多的机构开始重视海量数据背后蕴藏的无限“财富”，将数据资源作为一种基础性或战略性资源。在国家层面，许多国家将数据资源的开发提升到了战略高度，纷纷出台大数据战略。但是依据 TalkingData University 的访谈，所面向的公司在计划向数据驱动企业转型的过程中，专业数据人才教育的投资比例远远小于对于基础设施和数据源的投资；95%的企业都投入了较大的资源在基础设施、平台和工具、数据源以及算法优化（如下图表 3 所示）。而业内认为，一个好的公司应至少在专业数据人才上的投入要与其他方面的投入等同，甚至高于其他方面的投入。

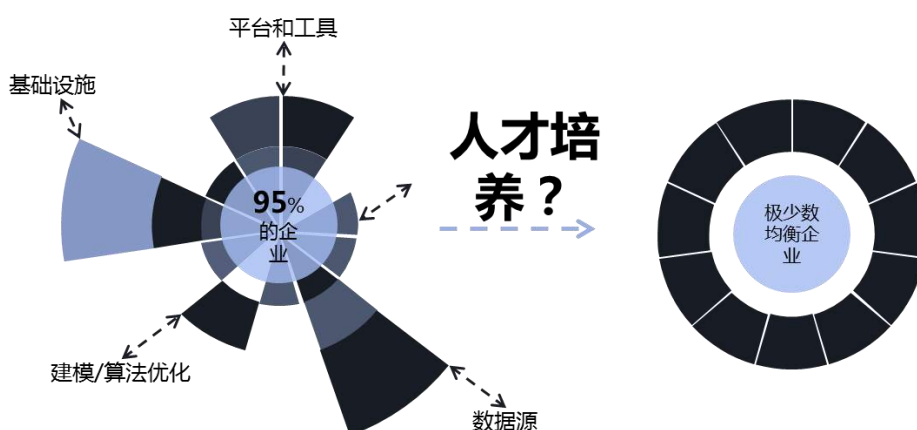


图 3 相比于其他方面企业在专业数据人才上投入过少

(资料来源：TalkingData University)

甚至早在 2015 年 CSDN 的一份报告中就指出，人才的缺乏已经成为制约公司实现重大的数据相关战

略的首要因素（见图 4）。

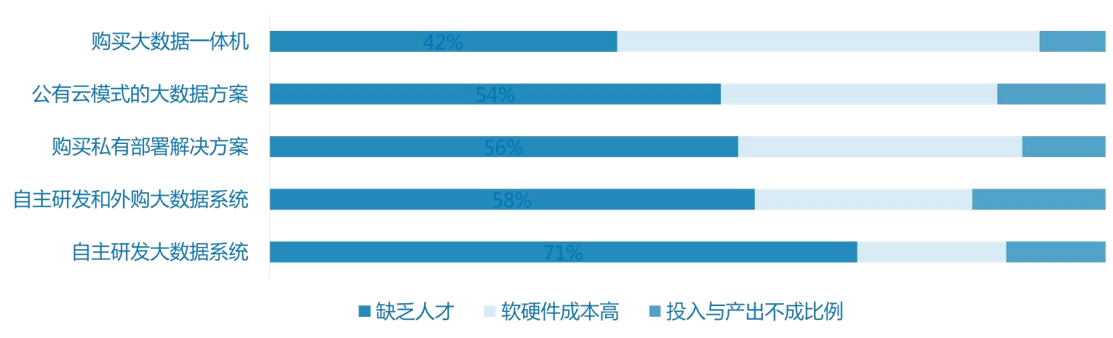


图 4 人才的缺乏已经成为制约公司重大的数据相关战略的首要因素  
(来源：CSDN：《2015 年中国软件开发者白皮书》)

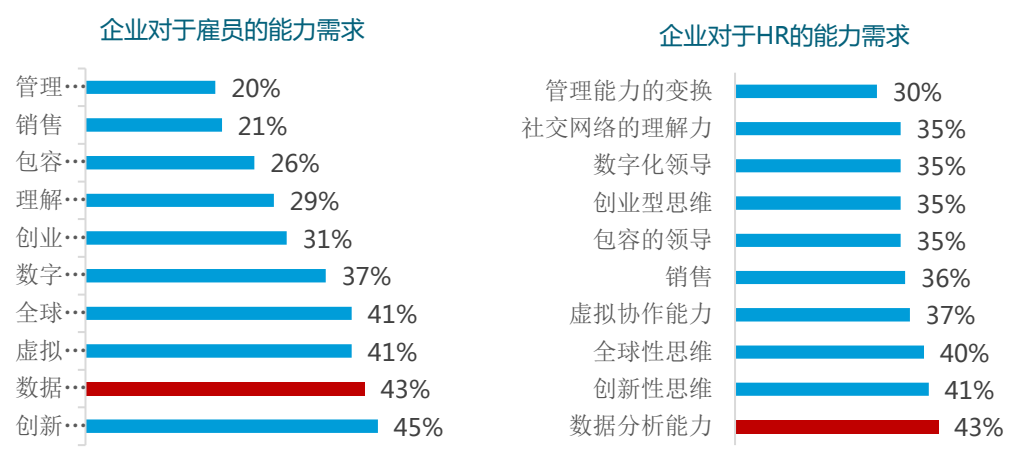


图 5 企业对雇员，及人力资源管理者各项能力需求  
(资料来源：美世 2017 全球高科技人才趋势报告)

现有企业对于人员的能力需求中，数据分析能力排在前列：很多客户都需要能够深入了解他们的业务需求并可以运用数据解决商业问题的人才（见图 5）；同时，数据技术和服务类企业需要能长期创造客户价值的人才以增加客户对产品的粘性，降低客户对服务的依赖以提高投入产出比；同样的，专业数据人才需要深耕某一商业情境，建立清晰的职业发展路径，保障成长资源的到位。培养数据人才成为大势所趋：公司如果想要培养可以应对不断变化工作状况的高效能专业员工队伍，则需要实战训练现有雇员，让他们得到专业化技能的提升；同时让员工们对于现有数据市场进行全新的认知，培养和鼓励员工关于数据的新思维。

内部 我们如何塑造员工体验	我们已经是一个数字化组织			1%	5%
	已取得极大进展		8%	14%	1%
	任重而道远	4%	37%	12%	1%
	尚未启程	8%	6%		
		尚未启程	任重而道远	已取得极大进展	我们已经是一个数字化组织
		外部 我们如何开展业务			

图 6 企业数字化进程  
(资料来源：美世 2017 全球高科技人才趋势报告)

从美世2017年高科技行业报告的企业数字化进程调查结果来看（见图 6），自认为已实现数字化的企业不超过10%，而已经踏上数字化之路的企业倾向于优先关注外部竞争力，其次关注内部的员工培养。本报告旨在需要让企业认识到提高企业数据能力的重要性，提升企业对于专业数据人才和数据教育的关注程度。

本报告由TalkingData University（腾云大学）和Mercer China（美世中国）联合撰写，旨在分析目前专业数据人才教育行业的现状，为数据教育行业的参与者和监管者提出建议，同时达到促进数据人才发展、指导传统企业构建数据团队，从而驱动企业数字化转型的目标。

# Contents

## 目录

### 前言

### 第一部分 ..... 1

#### 专业数据人才教育行业概述 ..... 1

##### 1.1 专业数据人才分类 ..... 1

##### 1.2 专业数据人才特点及其教育范式 ..... 5

##### 1.3 什么是专业数据人才教育 ..... 5

### 第二部分 ..... 7

#### 专业数据人才教育行业的历史沿革 ..... 7

##### 2.1 专业数据人才教育行业的兴起和溯源 ..... 7

##### 2.2 中国专业数据人才教育行业的发展 ..... 8

##### 2.3 数据行业和专业数据人才教育行业的关系 ..... 9

### 第三部分 ..... 12

#### 专业数据人才教育行业现状 ..... 12

##### 3.1 宏观环境分析 ..... 12

##### 3.2 市场规模 ..... 16

##### 3.3 生态图谱产业地图 ..... 17

##### 3.4 数据教育行业面临的挑战 ..... 23

### 第四部分 ..... 26

#### 专业数据人才教育行业发展趋势预测 ..... 26

##### 4.1 数据人才需求继续扩大，尤其以传统行业需求增加为主要方向 ..... 26

##### 4.2 核心人才缺口由数据分析师向数据科学家转变 ..... 27

##### 4.3 高端人才短期和中期还是以海归人才为主，初级人才发展迅速 ..... 27

##### 4.4 大数据行业发展飞快，校企合作颠覆教育分工 ..... 27

##### 4.5 专业数据人才教育行业竞争加剧，行业分化，等待平台 ..... 28

<b>第五部分 .....</b>	<b>29</b>
持续保持专业数据行业的领先地位 .....	29
5.1    吸引和留住未来人才 .....	29
5.2    培养未来人才并做好准备 .....	32
<b>第六部分 .....</b>	<b>34</b>
持续保持专业数据行业的领先地位 .....	34
6.1    行业参与者发展策略和建议 .....	34
6.2    政府部门的政策建议 .....	34
<b>附录：部分专业数据人才教育行业参与者分析.....</b>	<b>36</b>

## 第一部分

# 专业数据人才教育行业概述

### 1.1 专业数据人才分类

专业数据人才是从事数据业务的核心技术人才。依据数据行业的多年实践经验，TalkingData 将专业数据人才分为四类：数据科学家、数据工程师、数据分析师和数据产品经理。

#### 1.1.1 数据科学家

数据科学家是综合运用数据科学领域知识对数据进行采集、处理、挖掘、建模等操作以形成洞察并最终解决问题的专业型人才（见图 7）。数据科学家致力于用数据产生实际的价值。

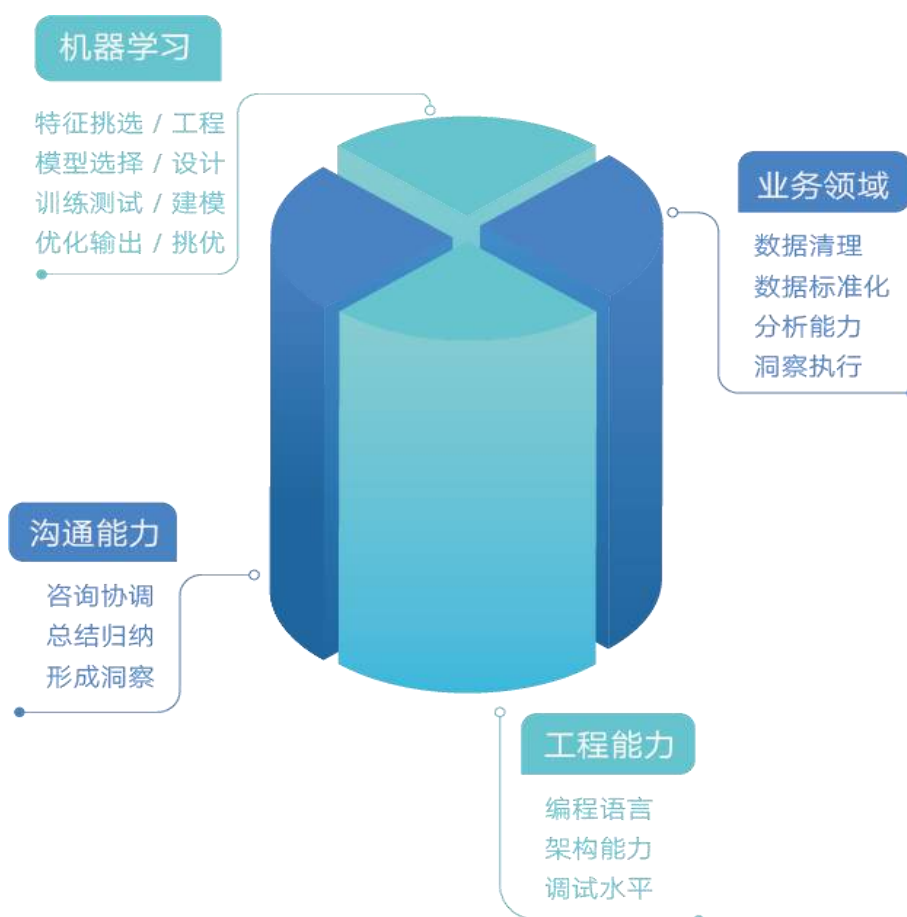


图 7 数据科学家的能力组合

尽管市场上现有数据科学家往往具有不同的教育背景和工作背景，但是在专业角度上，他们一般需要如上图所示四个方面有能力：



**业务领域：**业务领域方面的知识和能力是必备的。数据科学家要对问题所在领域的知识有充分的掌握和理解，例如接触金融数据项目，就需要理解金融行业的趋势和业务模型，能够进行专业化的数据分析。数据科学家对业务领域的了解和知识储备是提升数据价值的关键。

**机器学习：**机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。这些学科都可以用来帮助数据科学家梳理业务领域中遇到的各种数据问题，进行建模并得出指标和预测值来辅助业务。

**工程能力：**工程能力代表的是数据科学家的动手能力，影响数据科学家的工作效率和产出质量。

**沟通能力：**沟通能力可以帮助数据科学家发现并解决问题。数据科学家的沟通能力应该能够将数据的价值以通俗易懂的方式表达出来，以确保数据发挥真正的价值。随着业务领域的深入，数据科学家应该能经常发现和提出新的洞察和建议，以帮助企业实现其目标。

除了上述四点能力外，数据科学家应该具备的其他素质包括：

- 根据需要自定义方法并解决具体的问题，以便最大限度地交付最优化的成果（包括编写新的算法和修改现有算法的能力）；
- 访问和查询多个不同的数据库和数据源（例如，RDBMS、NoSQL、NewSQL 等），以及将数据集成到主要分析驱动的数据源里（例如，OLAP、数据仓库、数据湖等）；
- 查找并选择最佳的数据源和数据特征（变量），包括根据需要构建新的特征指标（功能设计）；
- 了解统计知识，编程和封装可供选择的算法库，并选择最优；
- 保证数据具有很高的质量，并且是以最佳的形式和条件呈现，从而得出准确、可靠的结果；避免处理与问题无关的垃圾数据；
- 选择并实施最佳工具、算法、框架、语言和技术，最大限度地在大规模数据下得出成果；选择正确的性能指标，并最大限度地提高应用性能；发现如何利用数据实现自动化操作；从事跨职能工作，并与公司各部门和团体高效协作；从失败中吸取经验，从而降低潜在风险和经济损失；了解产品（或服务）的客户和/或用户等等。

### 1.1.2 数据工程师

数据项目一般是一个工程项目，负责工程部分的实施人员就是数据工程师。数据工程师负责搭建架构和实现数据工程所需的技术平台和工具（包括数据连接器、数据存储和计算引擎、数据可视化能力、工作流引擎等），保证稳定可靠的数据处理，为数据架构师、数据科学家和数据分析师的工作提供支持。

数据工程师需要具备的能力包括但不限于如下：

- 技术能力：包括编程能力、架构设计能力、工程能力等；
- 业务能力：主要是对于业务的理解能力；
- 实战能力：包括数据意识、结果导向和文档撰写能力；
- 团队协作能力：包括学习辅导、沟通、合作以及影响力。

目前，市场上的数据工程师一般来自团队现有的人员，但是需要扩展数据技术相关的能力。数据工程师的主要职责是维护数据仓库和开发对应的服务。数据工程师在职位上贡献最多的是数据的抽取、转换和上载 (Extraction、Transformation、Load，ETL)，将所需的数据从不同来源不同格式的数据源中提取出来，转换类型以便使用，然后归档入数据仓库。通常，数据工程师需要有一定的计算机科学技术背景，需要能够编程和维护数据仓库以及保证服务正常进行。

### 1.1.3 数据分析师

数据分析师负责将数据中的价值提取、分析并解读出来以指导行动。数据分析师和数据科学家从名字来看十分容易混淆。那么两者之间的区别在哪里呢？

首先，两者使用的工具通常是不同的。数据分析师进行商业智能和分析使用的工具：Microsoft Excel（进行可视化，数据透视表等）的 Tableau、SAS、SAP 和 Qlik。数据分析师考虑的首要因素是快捷地从中间解读结果并验证设想，得出结论；而数据科学家则比较侧重算法选型、模型稳定以及大数据情境下运算的实现。

其次，数据分析师有时会进行数据挖掘和建模工作，倾向于使用数据可视化工具，如 IBM SPSS、Rapid Miner、SAS 和 KNIME。相比较下，数据科学家在执行相同的任务时，偏向于编程类的库，如 R 或者 Python 中相关的库。

总结来看数据分析师通常只针对于公司高层提出的问题和目标进行分解、分析，并最终报告他们的发

现。然而，数据科学家是主动通过数据和现象了解其业务关联性并采用更先进的统计和建模技术实现数据可视化，并更多以驱动业务的方式来进行汇报。

#### 1.1.4 数据产品经理

相较于其他专业数据人才，数据产品经理是等级更高的交叉型人才。

数据产品经理是面向客户特定数据需求和数据处理流程来开发产品的人。为了满足客户需求，数据产品经理需要履行如下的基本职责：

- 结合客户的数据需求，并将其分解成数据领域不同范畴的问题：数据集、人、流程、最终产品形态
- 准备解决问题的数据集，针对客户的特定需求识别数据来源和数据接口，并对数据进行清理、结构化，形成解决问题的完善的数据集；
- 数据产品经理知道选择何种数据人员来满足客户的需求，例如知道选择擅长何种方向技能的数据科学家来解决这个特定的问题，需要哪些数据工程流程，客户的问题在什么领域并和数据分析师讨论，并能够明确这些人员之间的合作界面和职责分配；
- 了解整个数据处理的生命周期，能够帮助梳理从数据集到业务洞察的机会列表，明确关键环节和关键环节上的产出，并明确最终产品形态/交付形态如何呈现给客户。所以，数据产品经理的能力包括但不限于：

项目管理：数据科学项目实战流程，实现 MVDP（最小可行数据产品）；

数据科学：知道建模需要什么计算资源和数据集结构；

数据工程：知道整个平台的结构，流程设计；

数据准备：知道怎么去结构化地准备数据，进行 ETL；

数据分析：如何将数据结果更好的呈现给客户-可视化，分析结果的本身含义

行业知识：包括怎么选择指标，怎么设置 A/B test，怎么优化模型)等等。

所以，一般来说，好的数据产品经理都是由其他三种资深的专业数据人才晋升而来。

## 1.2 专业数据人才特点及其教育范式

	IT coding 能力	思辨和建 模能力	数据处 理能力	Domain 知识	客户沟 通能力	系统工 程能力
数据分析师	弱	强	弱	强	强	弱
数据科学家	中	强	强	中	中	弱
数据工程师	强	弱	中	弱	中	强
数据产品经理	中	中	中	强	强	强

图8 专业数据人才的能力集对比

综上所述，专业数据人才的能力集都是复合型的（见图8）。公司现在需要的是既能涉猎较宽，在某些方面也能非常专业的交叉“大T型”人才。从教育的角度来看，成为一名专业数据人才需要明确的学习路径而且需要长期培养才能成为综合素质人才（需要注重培养的方式，科学和创新的培养方法）。

进一步的，根据 TalkingData 近期发起的一次专业数据人才心理调研<sup>1</sup>，从初步的结果分析中发现专业数据人才有如下特性：在霍兰德职业兴趣测试中，数据科学人才在“研究型”这一类别上的得分显著高于非数据科学相关岗位人员，在“现实型”和“常规型”的得分也相对较高，而在“企业型”这一类别的得分则相对较低。在五大人格测试中，相比于非数据科学的岗位人员，数据科学人才的开放性相对较高，而他们性格外向性，讨人喜欢的程度和对于工作全盘考量的领导力相对较低。虽然更加深入的分析结果还有待进一步研究，但上述结果可以说明专业数据人才有一定的共性，与其他类型的人才有着显著差异，展现了这个群体的独特性与特殊性。但是找寻和挖掘这类人才也并非易事，加之不同企业的数据人才所熟悉的数据结构和商业情境有限，依赖简单的猎头无法完全满足企业的用人需求，必须依靠深度的教育培养，配合相应的教育资源，结合实际情境，将在数据领域有潜力的青年培育成能堪大任，能够在不同数据岗位上发挥用武之地的优秀人才。

## 1.3 什么是专业数据人才教育

在本文中，专业数据人才教育被定义为：为了培养专业数据人才而专门进行的教育行为。根据数据行业业务特点，TalkingData 认为，专业数据人才的教育体系应以数据科学的教育为核心大纲，辐射数据工

<sup>1</sup> 回收问卷总数共计 238 份，有效问卷数 235 份；其中 152 份为专业数据人才，其他为非专业数据行业从业人员。

程和数据分析，因为数据科学已经成了专业数据人才区别于其他核心技术人员的主要技能点。数据科学是一门交叉度极高的学科，要求学生同时掌握数据、统计学、计算机等基础学科的大量知识。而从数据科学教育的学习流程来看，数据科学教育又可分为三个阶段<sup>2</sup>：以数据科学的基础学科为中心的教育；以数据科学的基础理论为中心的教育；以数据科学的应用为中心的教育。

### 以数据科学的基础学科为中心的教育

这一阶段侧重在学习数据科学课程之前的必备知识储备，处于数据科学课程链的上游，一般很少涉及数据科学本身的知识，更加适合在本科低年级阶段开设。相对于其他课程，数据科学对统计学、机器学和数据可视化等基础理论之间的耦合度较高，学习难度较大。因此，此类课程可被视为是数据科学课程的先修课程，其意义在于为学生更好掌握数据科学的知识奠定基础。

### 以数据科学的基础理论为中心的教育

在学生完成了基础学科的知识储备后，需要学习数据科学本身的理念、理论、方法、技术、工具和(或)最佳实践应用，而这正是数据科学课程链的核心。这一阶段的课程，往往安排在本科高年级或研究生阶段。

### 以数据科学的应用为中心的教育

以数据科学的“领域应用”为中心的课程：主要讲解数据科学对某一个学科领域的影响及其应用方法论或(和)最佳实践，处于数据科学课程链的下游。在这一阶段，学生需要在自己的数据科学理论基础上，通过科研项目、数据商业项目实践等方式不断增加行业领域知识，增强对问题的理解，学会疑难问题的解决方法。

---

<sup>2</sup> 资料来源：《全球数据科学课程建设现状的实证分析》，朝乐门

## 第二部分

# 专业数据人才教育行业的历史沿革

### 2.1 专业数据人才教育行业的兴起和溯源

#### 2.1.1 战略背景

专业数据人才的教育行业是在大数据行业，尤其是移动大数据普及之后开始兴起的。美国白宫科技政策办公室在 2012 发布了《大数据研究和发展计划》，并成立“大数据高级指导小组”，用以增强联邦政府收集海量数据、分析萃取信息的能力，进而提升政府预测社会经济发展的准确度。我国在 2015 年发布的“十三五”规划建议中，也提出“实施国家大数据战略，推进数据资源开放共享”。此举被外界视为我国首次提出推行国家大数据战略。在各国大数据战略的推行下，数据产业得以迅速发展，同时，对人才的需求也日益迫切。仅就数据分析师这一典型数据人才而言，麦肯锡全球研究院发布的关于大数据发展趋势的研究报告中预测，到 2018 年美国将会面临 14 万-19 万的大数据分析人才缺口，和 150 万具备数据头脑意识的管理人才缺口。如何满足行业对数据人才的需求，成了教育界面临的一大挑战，但也为“数据科学”这一新兴学科的发展，提供了契机。

#### 2.1.2 英美领跑数据科学领域人才培养历史

面对数据科学领域人才紧缺的现状，高校被赋予了补充人才缺口的首要责任，而在这个方面，发达国家由于技术发展领先以及产学研结合机制较为成熟等原因，在人才培养方面走在了前列。据 DataScience Community 数据（未包含中国大陆相关数据）显示<sup>3</sup>，截至 2018 年 4 月，全球共有 27 个国家的 366 所高校开设了 581 个数据科学相关项目，其中北美洲高校开设数量最多，占比达到 78.4%，其中美国高校的数据科学项目占比高达 77.6%。在亚洲地区，香港特别行政区、韩国、菲律宾、新加坡共开设了 5 个数据科学项目。其中，香港中文大学开设了研究生层次的“数据科学与商业统计”项目，香港科技大学在 2009 年开设了本科生层次的“风险管理与商业智能”项目。

<sup>3</sup> 数据来源：<http://datascience.community/colleges>，数据统计截止时间：2018 年 4 月 22 日。

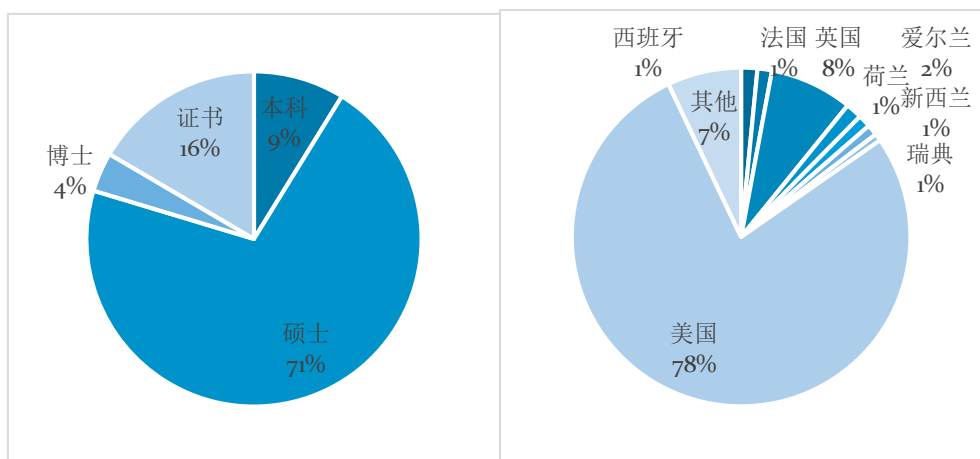


图 9 数据科学相关项目开设情况  
(数据来源: DataScience Community)

纵览全球数据科学项目，可以发现如下特点：

- 在数据科学领域，市场对应用型人才的需求决定了高校的人才培养的主要方向，即目前高校主要处于为了满足市场需求的应用技能型人才调整培养模式的阶段。例如，华盛顿大学西雅图分校在经过对数据科学领域人才需求情况调研后，为其数据科学专业人才设定了 9 个职业发展路径：商业分析师、数据分析师、数据架构师、数据工程师、数据科学家、定量分析师、研究员和统计师。从知识、能力和素养上来看，高校也多以社会需求为导向进行培养方向的设定。
- 人才培养模式的多元化，其中硕士研究生项目占据主流。相较于传统的本科基础教育，数据科学方向的证书类项目约占 16%，证书类项目往往以就业与实际应用为导向，这也反应出了对应用型人才的急迫需求。此外，数据科学类的授课形式不再拘泥于传统校园的线下教学，在线课堂更受欢迎，有 28.7% 的项目可通过在线方式进行学习。
- 商学院在数据科学项目的开设中大放异彩。传统的互联网人才往往集中于软件学院或信息学院等技术类学院，但在专业数据人才的培养上，商学院所占的比重越来越高，根据 DataScience Community 收录的数据显示，全球约 4 成的数据类项目由商学院开设。大数据在商业中的应用、辅助进行商业决策，使商学院成为了数据科学的重要推行者，培养更具商业视角、理解真实业务的数据人才成为不少商学院的培养重点。

## 2.2 中国专业数据人才教育行业的发展

国内的数据行业发展稍晚于国外，在数据人才的培养上，也存在发展进度相对缓慢的情况。与国内的传统学科相比，数据人才培养学科建设尚不完善，但在人才培养模式上，呈现出接近国外的多元化。数据



类学科的强应用性使得很多具有实战经验的工业界人员参与到数据人才的培养上。高校交叉学院的学位教育、项目制认证教育、培训类教育等多方力量共同推动着国内数据行业的发展。

尽管国内的数据人才培养在形式上与国外基本相同，但核心的课程设计并不完善。以数据科学专业为例，国内的数据科学专业往往依托于高校的优势专业存在，但在核心课程如“机器学习”的设置上存在很大的问题，主要表现在课时量设置过少，课程内容过于简单，交叉性不足等。

受教育部政策的影响，数据人才培养的整体趋势呈现由点到面，高校水平不断丰富的特点，目标旨在培养多层次的数据人才。

2015 年 10 月，教育部新设了云计算技术与应用、电子商务技术专业，增设了网络数据分析应用专业方向。

2016 年 2 月，北京大学、对外经济贸易大学和中南大学成为首批成功获批设立“数据科学与大数据技术”本科新专业的高校。2017 年 3 月，经教育部批准，第二批共 32 所高校成功申请「数据科学与大数据技术」本科专业。从两次获批开设「数据科学与大数据技术」本科专业的院校名单来看，专业学制都为 4 年，授予工学或理学学士学位。

## 2.3 数据行业和专业数据人才教育行业的关系

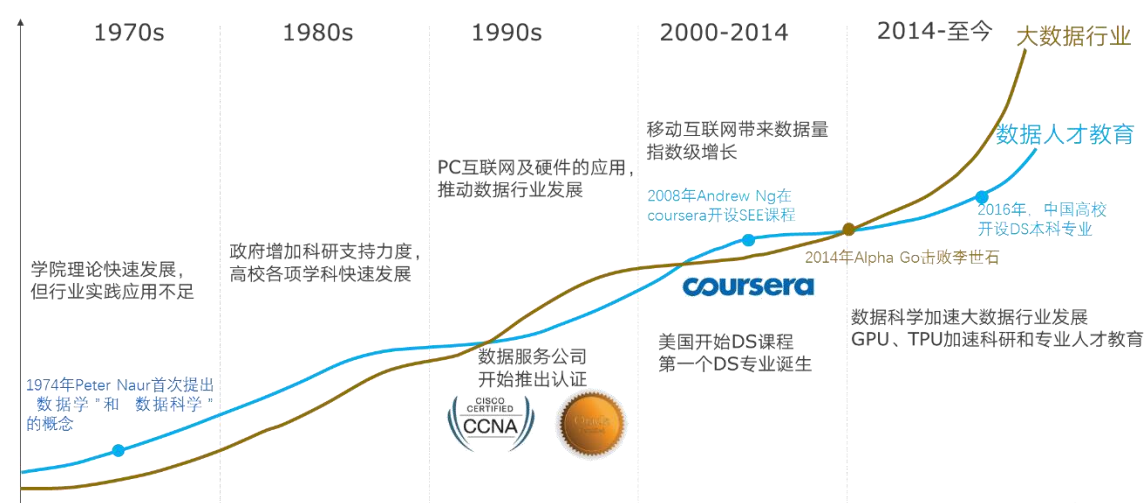


图 10 数据行业和专业数据人才教育行业的关系示意图

（资料来源：TalkingData University）



如图 10 所示，二者之间的关系是相互促进的：在大数据及数据科学专业发展初期，主要依靠高校力量来推动数据技术和教育的发展；随着硬件及数据规模的增加，工业界后来居上，两者相互促进，共同成长。这是由于专业数据人才的教育特点所决定的。专业的数据人才不仅需要交叉学科背景的基础知识，还需要大量的业界数据集和真实问题来供给和培养，因此产学研结合和计算资源的投入成为了专业数据人才教育上升的主要动因。

具体来说，随着大数据行业的发展，专业数据人才教育行业发展分为如下几个阶段：

- 上世纪 70 年代：高校主导的传统专业学历人才培养阶段，此时是算法基础规则的萌芽期。由于政府参与主导，最好的计算资源都集中在各个高校。此阶段专业数据人才的培养还集中在理论部分为主。
- 上世纪 80 年代：与计算相关的资源都集中在高校及相关部门，例如巨型计算机，因此科研的进步推动了教育的大力发展；此阶段与数据相关的各个专业（包括传统统计学、应用数学）也处于百花争鸣的状态。对于数据价值的挖掘的概念在工业界开始萌芽，但是由于计算能力的缺失，并没有形成大潮。
- 上世纪 90 年代：个人电脑和互联网的普及极大的丰富了数据的来源。传统的大型数据服务企业开始兴起，例如 IBM、EMC 和 ORACLE。数据存储、数据仓库和数据挖掘的概念在大型公司中率先兴起。为了推广软硬件产品，专业数据人才教育行业迎来了由大型企业主导的证书培训阶段：例如著名的思科工程师认证，ORACLE 数据库工程师等。
- 2001-2014 年：移动互联网的发展，让大数据的概念走入大众视野。互联网的平台经济特性，催生了一系列直接面向于个人学习者的专业人才教学平台，例如著名的 coursera。在整个专业数据教育行业，平台商、内容提供商、高校等组成的紧密耦合的生态催生了一大批的专业数据人才。并且，进一步的，数据人才之间诞生了分化。不同的数据人才从不同的历史轨迹演化而来。数据人才的能力标准初具雏形，业内定义标准不一。
- 2015 年至今：人工智能、机器学习的普及极大的提高了对专业数据人才的需求，于此同时，GPU/TPU 等一系列专业计算资源的诞生极大的提高了数据的计算效率。由生态类企业主导的全生命周期学习阶段开始，对于专业数据人才的教育竞争，本质上回归到了计算资源、数据集的竞争之上。专业数据人才的不同分工的定义，逐渐呈现统一标准。目前在这个阶段中，专业数据人

才还较为集中在以互联网和数据服务为代表的高科技企业中；在传统企业中，对于数据的投资都还刚刚开始集中在数据的基础设施、数据源、数据服务等；而对于培训和教育的投入，传统企业还没有跟上节奏，因此形成专业数据人才教育市场的窗口期。

## 第三部分

# 专业数据人才教育行业现状

### 3.1 宏观环境分析

#### 3.1.1 政策方面

2015 年 8 月国务院发布《促进大数据发展行动纲要》，这为我国大数据发展进行了顶层设计和统筹部署，这是我国促进大数据发展的第一份权威性、系统性文件。自此，我国发改委、教育部、工信部、科技部等部门密集出台了一系列相关政策（见图 11），推动《纲要》的具体落实。主要集中在以下三个方面：支持专业开设：新增“数据科学与大数据技术专业”；建立多层次人才培养体系：跨校联合、校企合作、企业在职培训；设立专项研究项目。

发文单位	部委公告	行业影响	备注
国务院	《促进大数据发展行动纲要》	行业发展顺应国家政策导向，利好	
教育部	《2015 年度普通高等学校本科专业备案和审批结果》	增加大数据人才供给	获批高校：北大、对外经贸、中南
发改委	《关于请组织申报大数据领域创新能力建设专项的通知》	有利于行业资源集成、分享	
教育部	《2016 年度普通高等学校本科专业备案和审批结果》	增加大数据人才供给	获批高校：人大、北邮、复旦、华东师范、电子科技等 32 所，详细名单见附录

图 11 专业数据人才教育行业发展相关的部委重要公告

#### 3.1.2 基础环境

2016 年全球大数据市场规模达 442 亿美元，超过 2011 年市场规模的 5 倍。大数据百度搜索指数近年来不断攀升表现人们对大数据的关注度逐渐提升（如图 12 所示）。根据公开数据统计，2015 年至 2018 年复合增长率在 20%至 30%之间（如图 13 所示）；2018 年中国大数据市场规模将远超 500 亿元，复合增长率在 40%以上，增长率始终保持在全球复合增长率的两倍左右。



图 12 “大数据” 百度指数<sup>4</sup>

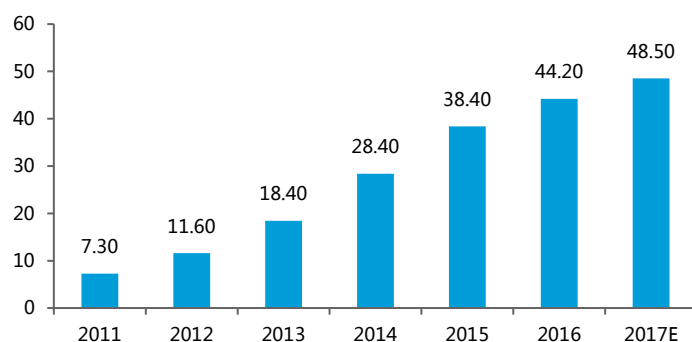


图 13 全球大数据市场规模 (十亿美元)

我国还设立了国家级大数据综合试验区 (图 14)。贵州、内蒙古等，主要提供大数据基础设施环境；天津、河北等偏向于应用环境；而广东偏向提供云计算、物联网等技术。

<sup>4</sup>图表数据来源：百度指数，以“大数据”为词条。

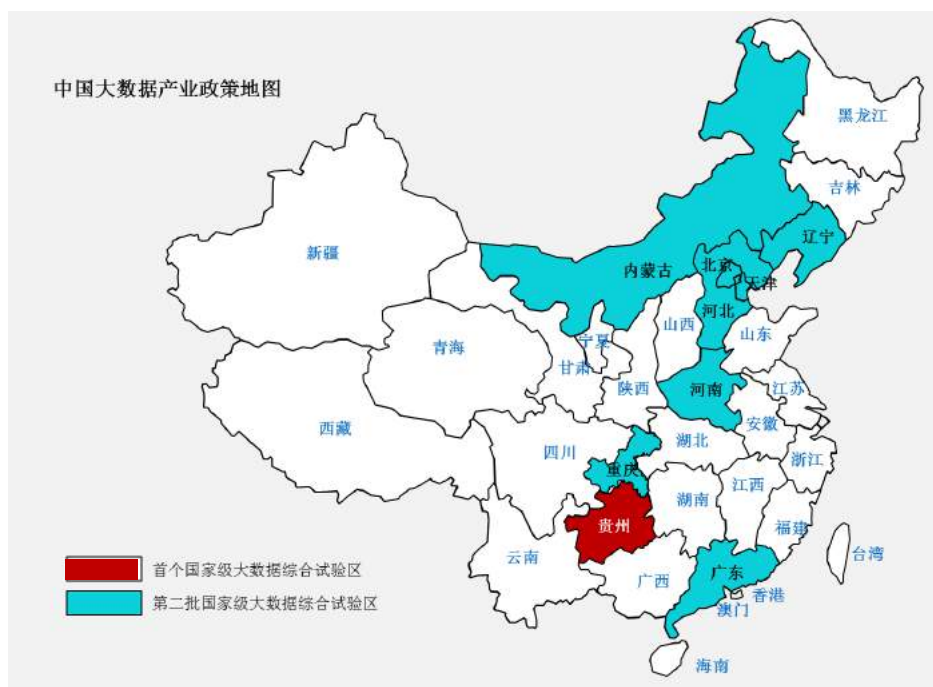


图 14 中国大数据产业政策地图

### 3.1.3 数据市场人才需求

大数据市场发展迅猛，然而全球的人才供给却面临严重不足。根据 TDU 报告，至 2025 年中国大数据人才缺口达到 200 万，而目前的中国大数据人才仅有 30 万左右。LinkedIn 发布的《2016 年中国互联网最热职位人才库报告》<sup>5</sup>表示，当前互联网行业最为热门的六大职位分别是研发工程师、产品经理、人力资源、市场营销、运营和数据分析。其中，数据分析作为新兴的热门职位，人才最为供不应求，供求比仅为 0.05（见图 15）。

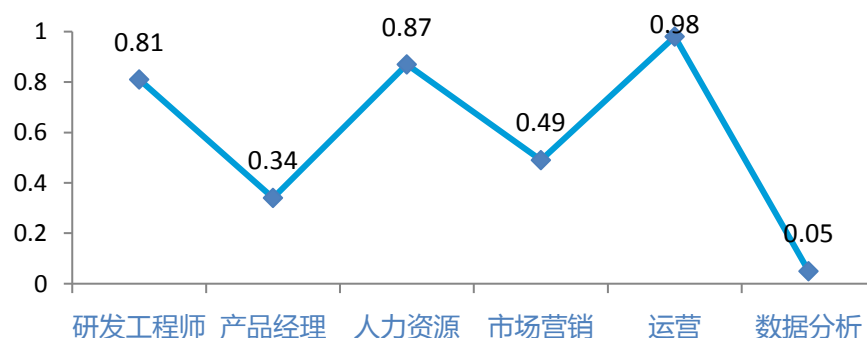


图15 互联网最热职位人才供给指数

<sup>5</sup> 资料来源：<http://www.useit.com.cn/thread-11520-1-1.html>

在中国，研发工程师和数据分析人才的地域分布极为不平衡，北京的人才集中程度尤为突出，77.8%的研发工程师和 88.9%的数据分析人才都集中在北京等前六名城市中（见图 16）。

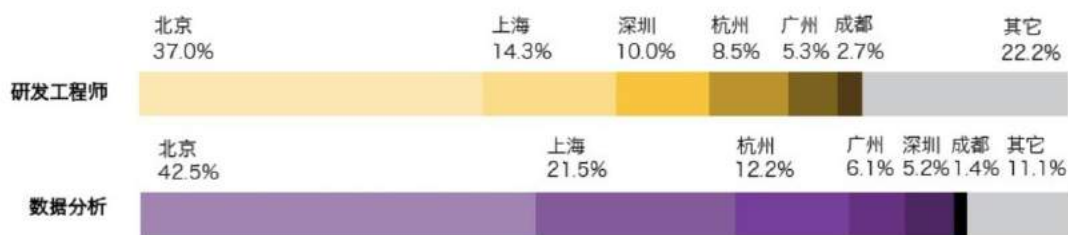


图 16 研发工程师和数据分析师地域分布  
(数据来源：LinkedIn，《2016年中国互联网最热职位人才库报告》)

数据分析岗位对学历要求较高，硕士与博士人才比例都远超其他岗位，且海归人才占比是最高的，达 15.6%，说明该岗位对专业性要求较高，而我国对数据人才的教育水平与国外相比还有一定差距。数据分析人才跳槽速度最快，平均跳槽时间为 19.8 个月，反映该岗位人才争夺较为激烈。

全球范围内，大数据领域人才同样匮乏。RJMetrics<sup>6</sup> 分析了 3.6 亿的领英个人资料后发现，截至 2015 年，全球只有 11,400 个数据科学家，这个数量在 4 年内增长了 52%。全球 55%的数据科学家在美国，印度、荷兰和以色列也有相当数量的数据科学家。对数据进行标准化后发现，以色列数据科学家集中度最高。如图 17 显示，相对于大数据行业的发展，中国的专业数据人才的集中度并不突出，缺口显著（见图 18）。

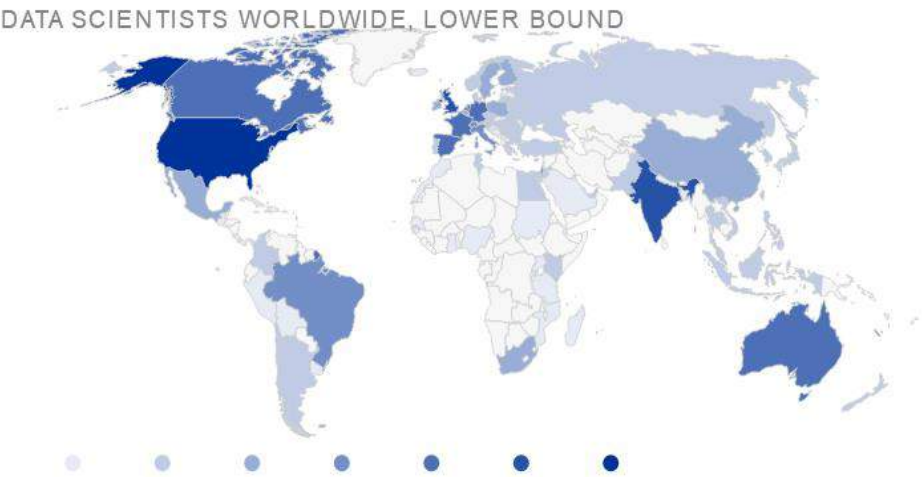


图 17 全球数据科学家集中度  
(资料来源：stitchdata.com)

<sup>6</sup> 来源 <https://www.stitchdata.com/resources/reports/the-state-of-data-science/?thanks=true>

国家	人口数	大数据从业人员	大数据人才缺口
韩国	0.51 亿	30 万	巨大
美国	3.21 亿	千万级	150 万（至 2018 年）
中国	13 亿	预计千万级	200 万+（至 2025 年）

图 18 部分国家大数据人才缺口情况  
(资料来源: stitchdata.com)

除了人才数量上的缺失，由于行业发展初期带来的人才质量上的不足也是较为凸显的。根据 TalkingData 的客户访谈，目前在传统企业中，也就是专业数据人才的需求增长处呈现如下一些特点：

- 核心专业数据人才在数据企业和传统企业都缺乏。增加学习数据的人数才有可能形成数据人才的上升；
- 带头人才获取成本高，而且从高科技企业挖来的人才也面临不适用的风险；不同的原始数据的使用也会让数据人才的产出效率不同；目前市场上仅有的几家专业猎头只适合高级人才的获取，且价格不菲；
- 大多数的企业现有员工的数据知识陈旧，还止步于上一代的 IT 和数仓技能。对于最新的数据科学知识缺乏；缺少数据团队的人才梯队规划；
- 人才成长不配套：数据科学家难以以现有的高校教育来形成，因为其成长需要大量的计算资源和数据集，并且需要全生命周期的终身成长的平台，就职后也需要提供社群平台保证其不断成长和知识更新；
- 行业及企业之间人才流动不充分，不利于知识的扩展，数据团队的整体素养亟待提高。

## 3.2 市场规模

根据爱分析 (ifenxi.com) 的数据，中国大数据行业规模 1000 亿，并能带动万亿级别的应用市场，这其中，主要的行业价值都需要通过专业数据人才进行实现，尤其是行业应用方面。第二，考虑到大数据教育在我国的渗透率极低，如果有千万级的大数据从业人员，以每人每年在培训领域消费 10000 元计算，每当渗透率提升 1%，市场规模就有望扩大 10 亿，发展前景广阔。再参照美国大数据教育的渗透率（50%以上），未来大数据教育市场的规模至少能达到 500 亿以上。

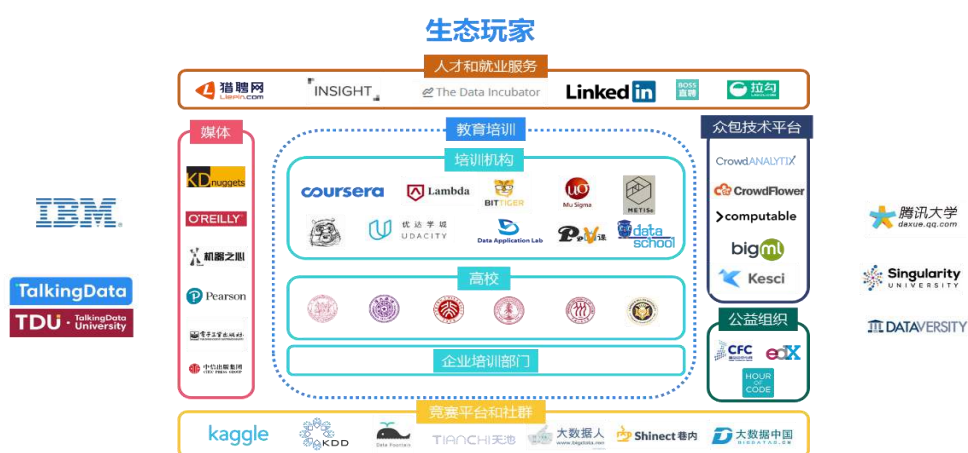
从人才规模来看，根据著名专业数据人才媒体和社群机器之心的相关报道，我国的数据科学泛人群群

体为 50 万左右，大多数都是在校研究生，或对数据科学感兴趣想要转型的在职人员；其中硬核部分，也就是真正可以被称为是数据科学家的是 5-6 万人。与美国不同的是，大多数核心数据科学从业人员在公司里面就职，其中有 70-80%集中在 BATJ、TMD 等大型互联网高科技公司中。这个数据也被著名的数据科学人才社区和竞赛平台 KAGGLE 所证实，KAGGLE 有 150 万注册用户，在其 TOP 100 的顶级数据科学家中，有 30%来自于中国，由此推断整体注册用户中来自于中国的大约有 30-50 万。这仅仅是专业数据人才中的“独角兽”——数据科学家的推断。如果核心科学家花费 10 万用以提高社群交流和技能提高，而平民数据科学家会花费 10 万左右用以获得工作机会，则整个数据科学教育市场约有 500 亿左右的规模。如果数据工程师、数据分析师和数据产品经理的教育市场是同等规模（虽然单人教育产品均价较低，但是市场基数较大）则整个市场的规模约在 2000 亿左右。

综上所述：1）中国的专业数据人才教育行业市场规模约在千亿级别；2）由于专业数据人才群体是有上限的，其教育行业成形后，波动率不会太大。

### 3.3 生态图谱产业地图

在专业数据人才教育市场，主要的参与者类型有媒体、教育培训机构、众包技术平台、公益组织、竞赛平台和社群、人才和就业服务以及生态玩家。



#### 3.3.1 主要参与者和代表企业

##### 媒体（见附录-表 5）

媒体主要为大数据行业提供最新信息和知识梳理。线上媒体通过优质信息、广告或导量等方式吸引流量。线下出版社主要通过图书销售、企业合作的方式获得收益。媒体出版社向社会大众传递行业信息，对



用户群举办线上线下交流活动（如读书会等）进行宣传教育。

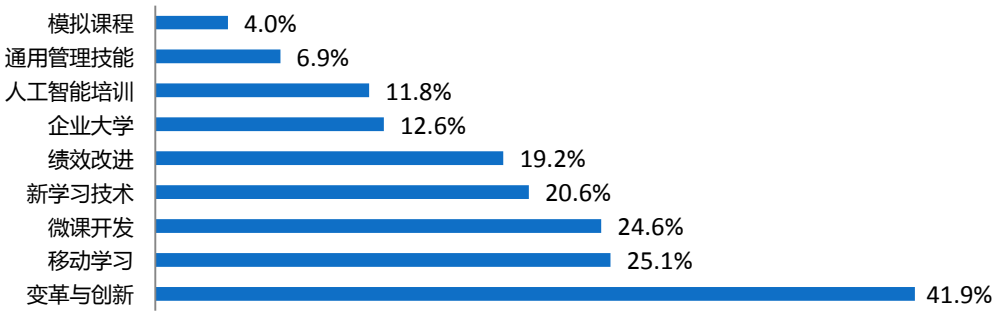
机器之心是国内最著名的数据人才相关的专业领域垂直媒体，是国内领先的前沿科技媒体和产业服务平台，关注人工智能、机器人和神经认知科学；坚持为从业者提供高质量内容和多项产业服务。

Pearson 作为全球最大的教育出版集团在专业数据人才教育领域的基础课程教材是非常具有代表性的，并被国内的高校出版社争相引入。O'Reilly 是世界上在 UNIX、X、Internet 和其他开放系统图书领域具有领导地位的出版公司。O'Reilly 涉足数据技术领域较早，其在大数据领域的涉足较深。O'Reilly 曾经在官网编纂过大数据历程推荐书单，并与 TalkingData 联合出版过 Smart Data Platform 等行业手册或白皮书。在国内，机械工业出版社、电子工业出版社、人民邮电出版社、华章、中信等出版社也陆续出版了大量的大数据教育类书籍。


教育培训

- 培训机构（见附录-表 3）

根据中国培训行业研究报告显示，企业所关注的培训焦点领域中，人工智能培训、移动学习、微课开发等领域占据较大的比例，这意味着部分企业已逐渐意识到数据科学与分析的重要性（图 19）。培训机构也是大数据教育行业最直接的参与者。目前主流的大数据教育机构主要通过线上 MOOC 课程方式授课，盈利主要来自课程贩卖。培训机构一般会和知名企业合作，为企业提供员工培训课程。同时，培训机构也会对优秀的学员提供奖学金或项目孵化等。虽然高校的数据科学教育呈爆炸式增长，但仍难以满足市场对数据科学家的需求。有很多人希望可以用更灵活的方式来安排上课（在职教育）。在这样一种情况下，线上课程网站如 Coursera、优达学城( Udacity )、Udemy ,线下集训营(Bootcamp)如 General Assemble、Insight、The Data Incubator 等应运而生。同时，很多线上课程网站也提供获得纳米学位( Nano Degree )的选项。



- 高校：包括相关院系和交叉学院（见附录-表 1）

市场对大数据下的数据科学领域应用型人才的迫切需求决定了高校人才培养的主要方向， 为市场快速输出应用技能型人才是高校的首要任务。面对数据科学领域人才紧缺的现状，高校被赋予了补充人才缺口的责任。据 DataScience Community 数据显示，全球共有 26 个国家的 356 所高校开设了 551 个数据科学相关项目，其中北美洲高校开设数量最多，占比达到 78.2%。亚洲有中国香港、土耳其、韩国、新加坡、立陶宛和以色列等 6 个国家和地区的 7 所高校开设了数据科学项目。85%的数据科学项目由英美高校开设。

美国最早开设数据科学专业的大学是北卡罗来纳州立大学（North Carolina State University），该校曾开发著名的决策支持大型集成信息系统 SAS（Statistics Analysis System）。2007 年 2 月，北卡罗来纳州立大学成立高级数据分析研究院（Institute for Advanced Analytics, IAA），成为美国首个正式的数据分析硕士研究生学位授予单位。随着这一领域的不断发展，如今，已有包括哈佛大学、麻省理工学院、斯坦福大学、卡内基梅隆大学在内的近 50 所美国高校开设了数据科学相关专业。

在这些学校中，数据科学相关专业主要对应 3 类研究方向：

- 商业数据分析方向（通常设在商学院内）
- 计算机科学主修数据分析方向（通常设在工程学院或信息学院内）
- 商业管理主修数据分析方向（通常设在管理学院内）

进一步分析可发现，许多美国大学会将数据科学与原有特色专业相结合，在本校具有优势的领域中关注数据科学，并且非常重视这一专业的实践性。例如，借助硅谷附近的地域优势和学校顶尖的计算机科学研究实力，斯坦福大学使用亚马逊提供的 EC2 云平台，让学生进行大规模计算的学习实践。而在康涅狄格大学，数据科学相关专业的课程设计建立在学校与美国通用电气公司（General Electric Company）长期的技术合作关系之上，学生在修读过程中，将有机会将科研成果应用于实践，在商业分析和项目管理方面获得扎实的理论基础与实践经验。

而国内高校也纷纷设立了与大数据相关的专业或研究所。在 2014-2015 年期间，多所国内知名高校采取设立“大数据研究院”等交叉学院的形式，探索数据科学专业的未来发展方向。经过国内各高校与互联网企业的研究与探索，“数据科学”成为一门独立专业的时机已经日渐成熟：在国内除了教育部批准的大

数据本科专业培养，越来越多的高校成立大数据研究所、研究中心或实验室，不断加强大数据基础研究，并设立大数据专业，积极培养大数据相关人才。例如：清华成立数据科学研究院，并宣布将推出多学科交叉培养的大数据硕士项目，第一批大数据硕士学位研究生将正式开始培养。北京航空航天大学成立大数据科学与工程国际研究中心，作为布局大数据战略方向的另一重要举措，并创办了国内第一个“大数据科学与应用”软件工程硕士专业。华东师范大学成立云计算与大数据研究中心、厦门大学成立大数据挖掘研究中心并出版《大数据技术基础》教材、广西大学成立复杂性科学与大数据技术研究所等。

### 众包技术平台（见附录-表 7）

最开始在专业数据人才领域有代表性的平台是竞赛技术平台，例如 Kaggle 和 Kesci，对参赛者开放计算能力，一般这类平台背后都有类似于 Google Cloud 等云服务平台支持，以保证计算资源。除此之外，最近诞生的比较多的是 Human-In-The-Loop Crowd Sourcing 平台，主要是为整个领域提供专业数据人才教学、科研或者实践所用的成本较低的标注数据集或者数据科学能力商业化平台，例如数据集标注平台 Crowdfunder、Computable，开放数据探索平台 BigML 等。

### 公益和行业组织（见附录-表 2）

公益组织主要是承担着整个行业的基础教育和通识教育的角色，例如哈佛和 MIT 共同建立的 edX 学习平台，目标就是让更多人免费接受顶级的基础教育。

除此之外，数据行业协会在沟通政府与企业关系上起到承上启下的桥梁作用。以中关村大数据产业联盟为例；自成立之初，它就一直致力于推动大数据产业的发展。如今，联盟已相继开展了技术研发，成果转化，市场对接，咨询培训、资本运作，会展服务，政府委托和国际交流合作等工作，广泛积累了政、产、学、研各界资源，吸纳了一批大数据领袖公司和创新团队，这一举动增加了高端数据人才在行业内获得经验交流和职业发展的机会。

### 竞赛平台和社群（见附录-表 4）

数据领域的竞赛平台的缘起就是解决业界问题和发现人才。对于政府、企业来说，通过开放特定数据举办比赛，一方面可以通过众包的方式，寻找更好的解决问题的方式，同时提供数据开放给更多的数据人才来探索，也是对企业数据能力的一种证明。另一方面，对于互联网企业来说，数据竞赛也可以发现、储备优秀的人才。例如：Kaggle, 天池大数据，Data Fountain, DataCastle。

社区论坛是大部分数据分析师、数据科学家等数据型人才启蒙和学习的首要选择。社区论坛通过知识共享，数据提供等方式吸引平台流量，通过与企业合作或广告方式变现。企业借由平台举办竞赛成为社区

论坛吸引流量，获得优质内容的最好方式。美国比较有名的数据科学家社群有 Data science community, Kaggle。在中国比较有名的数据科学家社群有大数据人和大数据中国。

近期在行业的发展推动下，垂直类数据人才社群逐渐浮现，例如针对于零售行业人才的 CrowdAnalytix 社群（DataX）。

### 人才和就业服务

对于刚刚兴起的独角兽类人群——专业数据人才，就业服务目前是一个火热的商业模式。早在五年前，The Data Incubator 就提供了在此领域的专业人才服务，并积累了一大群业内大型企业客户。更特别的是，Insight 仅仅针对博士人群进行就业培训和服务，帮助已有良好教育背景的人群获取更多高质量就业机会。有很大一部分的项目是和企业合作直接输送高质量人才，也有和猎头合作比如 TalkingData 和猎聘北美联合推荐就业项目“梧桐计划”。

### 生态玩家（见附录-表 6）

大型企业的企业大学往往就是优秀的生态玩家，利用企业现有的生态优势和资源杠杆，进行数据人才方面的全生命周期平台建设，无论是和传统的高校共建教育项目还是开放网上课程，集训营，这些项目和工业界有着极为紧密的联系。企业在中间帮助设计课程，提供实习就业机会。因此有很多公司自己也开展了企业大学的数据科学家项目，例如：微软正在和 edX 合作开展新的线上数据科学家学位项目，同时微软也为本科生提供暑期夏令营活动。EMC 提供线上和线下课程的高级分析师的证书项目。IBM 成立了 Big Data University 线上服务提供免费的大数据相关培训。SAS 则成立了 SAS 数据科学家项目。Data Science 公司则创立了 DS12，这是一个 12 周的免费培训计划培养数据科学家，该项目以 Scala 为编程语言而不是通用的 Python 和 R。腾云大学（TalkingData University）作为 TalkingData 的教育和人才生态运营部门，依托 TalkingData 原有的生态结构和技术平台，为人才提供从教育到入职到社群全生命周期的成长平台。

### 3.3.2 不同参与者商业模式和 fast-win 策略分析

从商业模式来说，可以通过判断基础知识和行业实战训练来分为 2B 和 2C（见图 20）。目前来说，专业数据人才教育行业的商业模式初见定型，2C 模式主要靠流量转化（例如广告，但是这类人群的总量不大，因此收入有瓶颈）、教育培训或者人才就业服务（例如猎头）；而 2B 模式主要依托企业本身的商业模式，例如大型数据服务企业的商业模式。

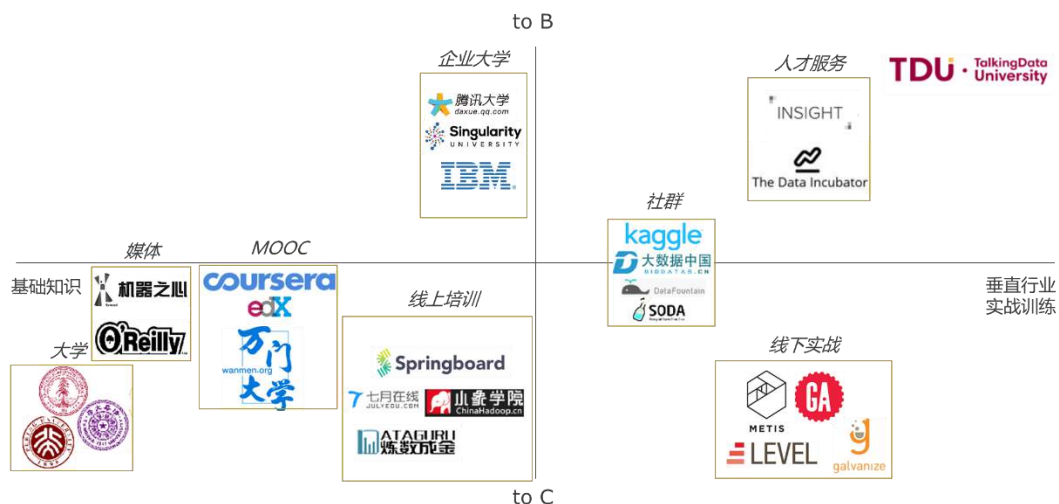


图 20 to B 和 to C 模式企业分布图  
(资料来源：TalkingData University)

参与者	运营模式	商业化路径	行业贡献	FAST-WIN 策略
媒体	与企业合作宣传，发布行业前沿资讯和广告；出版书籍，线下分享沙龙等。	教育，培训，广告，猎头服务	普及大数据行业相关知识和新闻;提供领域内通识和专业知识，组织形成小社群	树立品牌形象，垄断新闻来源 ;专注某一领域书籍出版，形成领域垄断；用持续的营销活动覆盖垂直人群
教育培训	培训机构 线上课程出售 ,2C 主打品牌宣传，与企业签订员工培训协议的 2B 业务盈利	教育内容变为主，其他服务为辅	为行业培训大数据人才。	与大型知名企业签订合作协议，依托企业产品和计算资源 ;与知名高校排他性合作。
	高校 招收学生，收取学费	教育	为学生提供大数据行业通识教育，为行业培养大数据基础人才。	获得教育部批准开设相关专业；率先梳理学科人才培养标准
众包技术平台	公开的可被任何人申请账号的平台；运营与产品一体化	SAAS ; TOKEN 交易	形成大量的数据集以及使得数据的价值能够被更多人探索	数据集的丰富性；技术开源性 ;计算资源开放性 ;资源流通模式设计需激活贡献
公益和行业组织	NGO，挂靠在高校或者行业监管部门底下	主要靠筹资和赞助	为社会提供免费培训和数据资源；支持行业发展维护行业秩序	社会影响力

竞赛平台和社群	举办赛事、会议；线下 meetup、线上论坛社区	比赛佣金、广告公关、PaaS 和 SaaS 服务	免费提供学习资源，提供专业交流平台。	与知名企业合作，发布最新、权威信息；设计合理社群贡献和资源交叉验证机制
生态玩家	以上所有	PaaS 和 SaaS 服务，教育，培训，人才服务，组织咨询	提高整个行业的资源配置效率和引导行业标准建立	生态完整性；有支撑整个平台运作的统一平台和打通账号；有核心人才标准体系

图 21 专业数据人才教育行业商业模式象限

### 3.4 数据教育行业面临的挑战

由于专业数据人才交叉背景、全生命周期学习的特点，专业数据人才的教育向传统教育模式的转变提出三个挑战：

- 师生关系由传统的单向变成双向：由传统的教师面向学生的单向授课方式，逐步向多边的，相互式的学习模式进行转变。随着各类互助学习社区，社群的出现，每一位学生既是学生，也是老师。学生间的相互学习、相互分享、相互交流和讨论，将成为新的社群化教育的趋势；
- 学习形式从原来的线下学习变成线上线下结合，并且在课堂中加重实操部分，将标准的模块化的基础知识通过线上课程的形式完成；
- 学习模式从原来的模块进阶化学习转向生命周期学习：这类人才再也不满足于模块进阶式的学习方式，而是在大量快速的补充基本知识之后，有一个开放式平台让他们充分的交流和练习，更广泛的接触到更多的经验知识；并且在人才的整个职业生涯阶段，都有充分补充知识和交流经验的平台。

由此，专业数据人才教育行业存在如下的现有问题：

#### 3.4.1 研究生为主的学历教育仍是主流模式，然而学历教育只是一部分，需要职业教育补全，同时丰富人才结构

在数据科学人才的培养上，学历教育仍占主导地位，共有 8 成项目是学历教育项目。在学历教育项目中，又以研究生层次的项目最为普遍，达到 7 成。在研究生培养中，数据科学多作为某专业的细分方向。



例如,斯坦福大学在统计学专业上设置了数据科学方向,要求申请者有修习数学和计算机基础课程的经历。南加州大学维特比工学院在研究生阶段开设了计算机科学 - 数据科学方向,要求申请者获得区域认可大学的工程、数学或自然科学本科学位。在本科层次教育的项目上,数据科学多作为统计学院、计算机学院或商学院下的专业。例如,诺丁汉大学的数据科学专业设在计算机科学系,俄亥俄州迈阿密大学的数据科学专业设立在商学院下面。也有部分学校独立设置了数据科学系。例如,根据 Data Community 的统计,伍斯特理工学院在数据科学系下,分别设置了数据科学的 1 个本科项目、2 个研究生项目、1 个博士项目和一个非学历教育的课程认证。课程认证式的非学历教育在数据科学领域的人才培养中也发挥着重要作用。全球有 16.9% 的数据科学项目在以课程认证的形式进行,多招收需要进行紧急技能培训或背景提升的在职人员。

### 3.4.2 近 4 成数据科学项目由商学院开设,需要更多的学院和力量来培养交叉背景人才

尤其是对于数据人才这种新型的复合型人才需要三个方面的知识储备:一是数学统计学的理论指导,二是得有一定的工学和计算机背景,还有很重要是要有一定的商业知识。

通过使用大数据辅助商业决策是数据科学最早、也是最广泛的应用场景之一。在此背景下,企业对同时精通技术和商业运作的复合型人才有着巨大的需求。这也决定了商学院成为了探索数据科学领域人才培养的先驱。据统计,全球近 4 成数据科学项目由商学院开设,占比最大。例如,得克萨斯大学奥斯汀分校的红麦库姆斯商学院开设了“商业分析”的研究生项目,旨在为企业培养兼备数据分析所需技能和企业管理所需能力的人才,德勤、沃尔玛等多家行业领先企业参与该项目的人才培养。

多学科联合培养也是数据科学人才培养的重要特征之一,这也直接体现在课程体系的建设和上,许多高校数据科学项目的课程由多个学科共同参与设计。例如,伍斯特理工学院开设的数据科学本科生项目课程就由商科、计算机科学和数学三个学科共同提供。其中商科课程包括“基于数据分析的商业决策”、“商业数据管理”、“消费者行为学”等,计算机科学则提供“程序设计基础”等课程,数学则提供“线性代数”、“应用统计”等,培养学生的综合知识能力基础。

华盛顿大学西雅图分校的数据科学研究生项目,要求学生培养解决实践问题的能力,学习统计学、机器学习、数据可视化、数据系统和管理等课程,掌握 Python、R、Java 等技能。

### 3.4.3 实践类教育资源稀缺,缺乏教研学一体化平台

由于大数据行业属于新兴行业,发展时间较短,因此既具备专业知识技能、又有较强的业务能力的教师较少,行业整体师资水平参差不齐,阻碍了大数据教育行业的发展进度。对于大数据人才针对性的实践

类教育资源缺失且主要集中在企业，例如真实问题与数据集的 capstone 项目、实习项目、黑客松和应用竞赛、开放探索平台和社群等等，也同时导致了人才引入的转化成本增加。虽然目前有多所高校开展数据科学项目，但是课程建设中还是存在许多瓶颈。一是数据科学课程需要数据分析，尤其是大数据的管理平台；二是缺少应用数据，由于数据科学的特殊性，企业离商业情境近、数据、真实业务痛点问题和经验多，高校少；三是缺少实战经验，学校需要与企业进行跨行合作。所以，目前的数据科学项目缺少贯穿从数据教育机构和企业合作的一体化平台。

#### 3.4.4 业务情景化，不同情景需求能力不同

数据行业应用场景较多，不同的应用场景对于技术要求不同。市场上现在主要以大数据理论和技术教育为主，在不同应用场景上深入不足，使得企业对于专业数据人才需要进行二次培训，造成资源的浪费。而面临数字化转型的传统企业才是痛点的存在方，理论和技术培养并不能满足企业的需求。

#### 3.4.5 缺乏全生命周期的统筹规划，行业之间的上下游配合

如果从一个人才成真的生命周期来看，目前市场上还没有一个培养计划能够覆盖整个人才成长的生命周期。比如说职前培训的高校是比较集中在基础知识的培养上，比如职中公司的人才计划是提升性的培养，比如社群主要是人才交流自我提高。但是如果不把整个用户生命周期和整个人才成长生命周期打通的话，就无法形成一个完整的知识结构，而这些知识的获取就需要一个贯穿人才成长的一个商业生态。

#### 3.4.6 政府对数据安全方面政策收紧，相关行业会受影响

政府对个人隐私的保护政策逐渐收紧，如最高人民法院、最高人民检察院颁布的《关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》，以及监管机构对大数据的内容审核，企业对于数据预处理等要求逐渐提高，对于大数据行业来说压力不断增加。因此对专业数据人才和从业者能够用来提高个人能力的培训资源都将越来越难以获得，这需要政府加强引导和支持。



## 第四部分

# 专业数据人才教育行业发展趋势预测

### 4.1 数据人才需求继续扩大，尤其以传统行业需求增加为主要方向

对于筛选评估、社交媒体、邮件往来甚至可穿戴设备中获得的反馈信息，我们收集到的人才数据相当庞大，只增不减。因此我们面临的问题并不是缺乏数据，而是如何利用数据！极少数企业能够将数据转化为前瞻性洞见，而近四分之一的企业还是只能生成基本的描述性报告和历史趋势分析。虽然生命科学和物流行业的企业在这方面走在前沿，但是在提供影响经理日常决策的可操作建议这一方面，企业仍有很长的路要走。根据麦肯锡 2011 年 5 月发布的《大数据：下一个创新、竞争和生产力的前沿》报告，如果能够富有创造性而有效地利用大数据来提高效率和质量，预计美国医疗行业每年通过数据获得的潜在价值可超过 3000 亿美元，能够使得美国医疗卫生支出降低超过 8%；充分利用大数据的零售商有可能将其经营利润提高 60%以上；通过利用大数据实现政府行政管理方面的运作效率提高，估计欧洲发达经济体可以节省开支超过 1000 亿欧元。根据美世 2017 年全球高科技人才趋势报告预测，大多数行业现在正在面临严重的人才匮乏和稀缺，所有的企业都将在人才方面面临竞争。

近年来各行业企业向数据企业转型，造成了数据人才缺口不断扩大，全社会对数据人才需求不断增加。根据美世 2017 全球高科技人才趋势报告显示（图 22），未来几年内高科技相关技术人才的缺口达 37%，是最缺乏人才的行业。为了快速高效地培养数据型人才，数据教育机构不断涌现，行业加速发展。未来对专业数据人才教育行业的需求将会持续增加。由于现有人才过于集中在互联网服务企业，市场规模更大的传统行业人才比例极其失调，缺口更为明显。



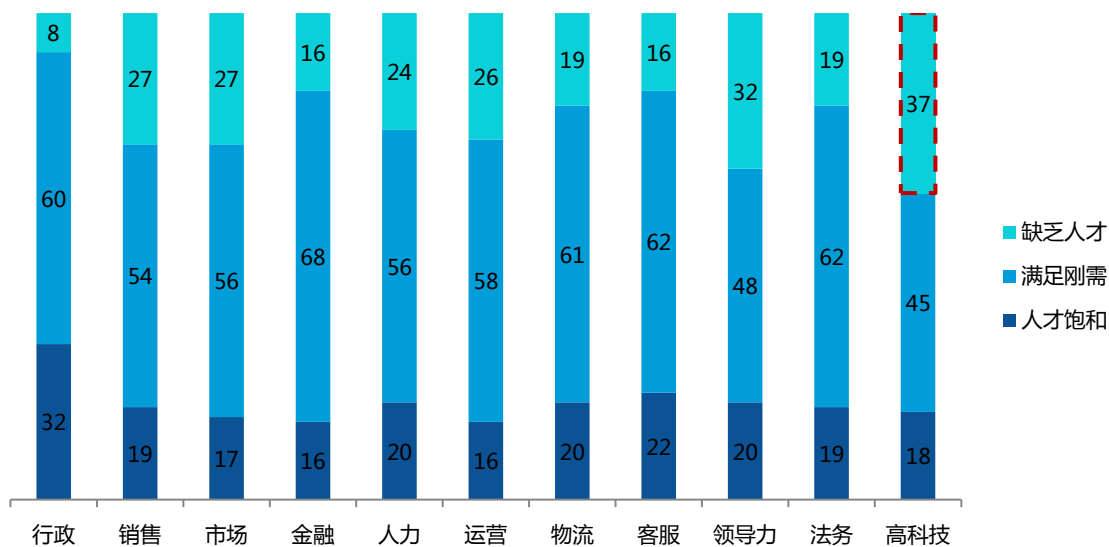


图 22 未来几年在不同行业的人才供给情况  
(资料来源：美世《2017 全球高科技人才趋势报告》)

#### 4.2 核心人才缺口由数据分析师向数据科学家转变

随着大数据行业不断发展，数据分析师的需求逐渐减少。数据分析的工具不断迭代降低了数据分析的门槛，同时在数据来源广泛、体量巨大且实时的情境下，企业自有数据不断增加以及对数据处理、选择、分析等要求不断提高，这对数据科学家的需求不断增加。企业对于大数据的建模、分析、以及商业逻辑判断等，都需要数据科学家的支持。未来数据教育行业将逐渐以培养行业数据科学家为主要目的。同时，数据工程师、数据分析师、数据产品经理也要懂得数据科学知识，才能更好的与其进行协作。

#### 4.3 高端人才短期和中期还是以海归人才为主，初级人才发展迅速

从 LinkedIn 的报告可以看出，北美地区的数据专业人才不仅在结构上发展较合理，并且数据人才属于比较复合的人才，能够兼顾技术、产品和业务，这种人才在中国的教育体制中很难被教育出来，所以短期内高端人才还是以海归为主。而随着中国专业数据人才教育行业的发展，打破国界的初级人才项目将会越来越多，在加强经验交流的同时，也会极快的丰富初级人才市场，提高人才基数。

#### 4.4 大数据行业发展飞快，校企合作颠覆教育分工

由于大数据行业发展飞速，高校的相关领域研究难以紧跟时代发展，而与此同时，许多大数据企业掌握了行业领先的技术，同时拥有一大批优秀数据型人才。高校更加需要与企业深度合作，通过课程共建，人才共建，科研合作等形式倒挂企业以获得研究的进一步成果。通过校企合作优化组合行业教育培训资源，

提高教育培训和考试的科技手段，建立拓展行业引导的数据人才培养基地。

在线课堂成为在专业数据人才教育的校企合作中一个重要的领域。技术的进步促进了教学方式的变革，也催生了众多“互联网+教育”的产品和服务，在线课堂正是被越来越多的大学广泛认可的互联网产物之一。在数据科学领域的人才培养中，全球有两成多数据科学项目将在线课堂作为授课方式。采取这种方式授课的项目中，以研究生层次的学历教育为主，开设项目有 72 个，占全部在线课程类项目的 72%，分布在斯坦福大学、罗格斯大学、纽约大学等高校。传统专业人才教育（例如 MBA 等）需要加强数据能力；而基本的数据能力教育要推动向本科延伸。

数据科学领域巨大的人才缺口，倒逼着数据科学教育行业的发展。数据科学教育已成为了高校课程设计与教学改革的主要关注点之一。数据科学涵盖的基础学科多，学科之间的交叉度高，教学难度大，对前沿理论的要求高，但数据科学教育经验及素材的不足，决定了传统的高校教育无法完全涵盖数据科学的所有教育，也意味着在数据科学教育领域，需要有更多的参与者。

#### 4.5 专业数据人才教育行业竞争加剧，行业分化，等待平台

在专业数据人才教育方面，目前整个教育生态中的玩家包括了高校，媒体、教育培训企业、数据人才服务企业等等，这些企业只覆盖了人才教育全生命周期成长中的一个价值链环节。要么是在就职前教育这个方面，要么是在就职中帮助人才更好地与岗位匹配，要么是在就职后提供一个数据科学方面的沟通平台。各教育培训机构都存在教育孤岛和教育断点。同时每一个类型的企业都在不同程度上的向上下游延伸，但由于资源专用性限制，很少有平台能够打通数据科学人才全生命周期，让专业数据人才能够一站式的解决所有成长和提高的需求。未来，专业数据人才教育生态的打通，贯穿和整合将是未来的主要形式。

## 第五部分

# 持续保持专业数据行业的领先地位

企业为了构建数据能力，构建专业数据人才团队，除了如前文所述的构建全生命周期成长平台、完善以教育为核心的培训体系、高校与企业深度合作之外，还需要保持建设在专业数据人才的吸引和留存的能力。

### 5.1 吸引和留住未来人才

要保持专业数据人才行业的领先地位，首要任务就是吸引和留住未来人才。在人才主导的经济中，员工体验在吸引最优秀、最睿智的人才这一方面至关重要。当今社会，一个多元化的职场通常会包含多五个时代的人，他们都有不同的标准和期望，因而优化员工体验更加棘手，特别是针对稀缺型数据人才。求职者在招聘过程中的互动，员工在职期间如何融入到企业中，以及他们离职后受到怎样的对待——这三个阶段都是塑造员工“体验”至关重要的时机。值得注意的是，一半的员工将他们的求职和招聘流程评为一般或较差，更不用说那些中途就被淘汰的求职者了。

拥有强大的人才数据库对企业来说很重要，尤其是在接触不可多得却又“相当合适”的被动人才时更是如此。与求职者、员工和前任员工互动时，利用好公司的外部品牌，将品牌吸引力的效力发挥至最大，从而吸引优秀人才。同时，科学技术也帮助塑造了这种格局，不仅是为了提高效率、减少招聘时间，更是为了确保求职者得到良好的体验。其中一些例子包括：

聊天机器人——回答求职者的问题以及在不用填写繁琐的表格的情况下收集背景信息，以此创建更灵活愉快的招聘流程。

算法——通过挖掘社交数据，在几秒钟内生成一张高质量的求职者名单，更有针对性地搜集人才。

在线评估——利用纳入员工评价的游戏和预测未来潜力的小型心理测试，做出更加明智的决策。在不断变化的工作格局中，实行着眼于未来的标准，可能会比评估求职者当前能力或过往经验的做法，更有效。

高盛集团目前正利用创新型技术和以能力为基础的面试方法来吸引更多的求职者，同时也继续采用及时通知以及根据数据决定招聘的做法。如今，本科生应聘实习岗位时，在线提交预先录制好的面试视频会作为第一轮评估。求职者针对一组预先确定的问题（与团队合作、分析思维、判断力等核心能力相关）进行回答并录下全程。随后，面试官会评估求职者的回答所展现出来的某种特定能力的程度，对求职者进行排名和比较，确保客观性和一致性始终是招聘的关键要素。

如果员工价值主张（EVP）与企业文化（即我们在公司里的行为方式）相左，那么这种对品牌吸引力的激情将无法转变为对工作的激情。在薪酬和公平工资方面保持领先，或者重点关注健康和弹性工作，这些都能起到一定作用。如果企业想要降低奖励的话，那就不应依赖于行业对标，而是应该选择一两个真正能做到差异化的领域。

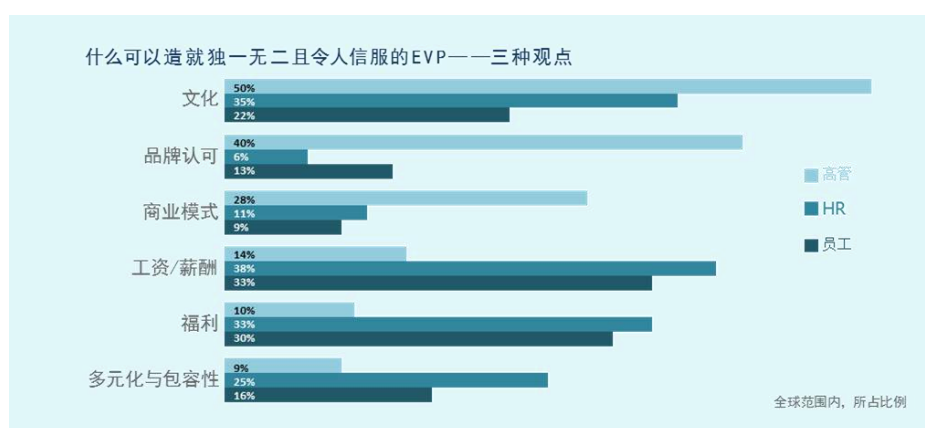


图 23 高管、HR、和员工最不同价值主张的认同度

这三个群体一致认可企业文化的重要性（图 23）。直线经理在塑造员工对企业文化的体验方面发挥了至关重要的作用，他们传达品牌承诺，同时将员工价值主张（EVP）转变为个人价值主张（IVP）。当员工面临敬业度或留任风险时，智能人力资源平台会用人才分析来提醒经理。但员工最终定义他们是否受企业重视，还是依靠经理是否有能力说服他们“留任”并将他们所在的团队纳入未来的规划中。

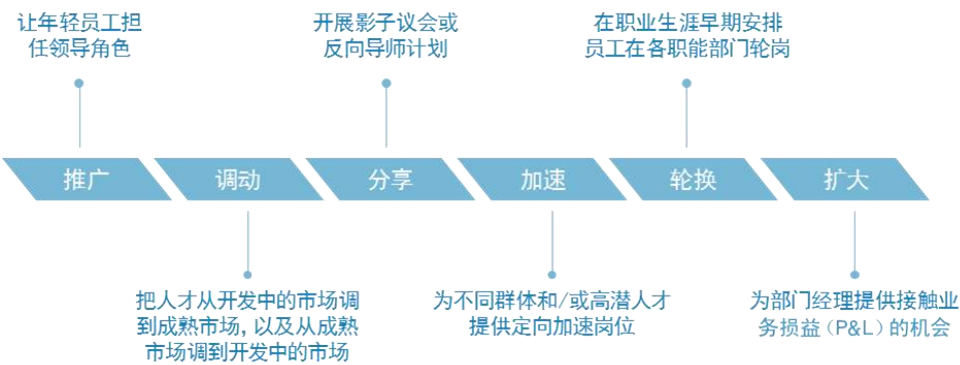
沟通是所有良性关系的基础。综合的沟通策略能激活员工价值主张，从而方便快捷地获取所需的资源，将关键信息送到人们手边。定制的消息应在合适的时间被推送到最需要的人群手中。简便是关键，沟通时

要迅速直奔信息主题，或者不超过三次点击来获取内容。个人提醒功能和操作简单的应用程序可以鼓励员工工作出更优质和明智的选择，以及探索职业的各种可能性。这些解决方案合在一起，提供了一种当今员工所渴望的“客户级别”的工作体验。



图 24 员工寻求最多的职业支持坐标图

在员工看来，薪酬差距和不平衡的晋升率往往伴随着留任挑战，这是职业动力告急的早期指标。这一等式的一端是员工渴望更多的灵活工作。企业应评估每个岗位固有的灵活工作类型和程度，并有意将灵活性融入工作设计之中。等式的另一端则是员工希望了解自己的职业选择以及晋升标准。当我们去询问员工，对于推进职业生涯而言最重要的是什么（图 24）。结果标明，除却薪酬和着眼于未来的培训以及定期的经理谈话之外，所需技能排在第一位，平级调动和轮岗计划似乎不是非常有成效。



很多数据企业已经开始认识到为了吸引和留住未来的人才，要把数据交给可以充分分析数据并得出前

瞻性洞见的组织手里，这种做法所带来的力量通用电气已深有体会。借助分析员工的非敏感信息和数据，HR 可以更容易地获得针对员工群体的了解，以提升人才规划能力。

5.2 培养未来人才并做好准备

未来的工作会变得很不一样，特别是当前的技能、文化和工作模式，三年后很可能就完全改变。对于刚起步的公司来说，更要未雨绸缪。

通常整合的人才策略不仅限于能力计划，同时有助于明确当前与逐渐成型的未来之间的差距。采用着眼于未来的方法，意味着识别能够带领公司前进的员工非常重要，即使他们现在并未在核心岗位上。好消息是全球近四分之三的企业有明确的方法去识别高潜人才，将严格的人才评估方法纳入流程。对人格和认知能力的心理测试能洞悉潜能人才的基本特征，虚拟评价中心则可以回答“谁已经准备好承担有难度的任务并且胜任上一等级的工作”之类的问题。

成功所需的知识和技能基础是不断变化的，动态职业生涯是在每一级组织普及数字技能的关键。“数字技能”不是一项单独的技能，而是每一个职能部门所需要的技能组合。例如，接受生物化学培训的制药行业研究人员现在将需要掌握操作先进的机器人技能来胜任本职工作。

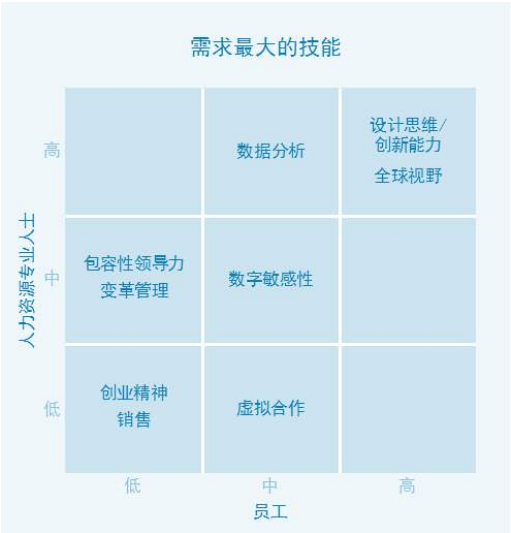


图 25 技能需求方阵

不管是通过外部招募还是内部发展，培养人才掌握多样化技能可以令企业不偏离市场需求，并且打造生机勃勃的未来团队（图 25）。加速创新所需的能力包括创业精神、冒险意识、市场敏锐、挑战现状、风

险意识以及长远眼光。不管是通过内部派遣还是通过沉浸式学习，这些都需要去亲身经历、不断尝试和犯错。通过营造一种培养这些特质的文化，企业可以培养实现未来宏伟目标所需的敏捷度与耐力。

企业应该鼓励员工努力完善自身职业生涯来有意识地培养能力。今年的研究发现，认为自己没有能力实现职业生涯成功的员工倾向于给经理的培训和指导较高的评价，那些感到“职业自主”的员工更喜欢公司的创新支持。



## 第六部分

# 持续保持专业数据行业的领先地位

### 6.1 行业参与者发展策略和建议

对于专业数据人才教育领域的参与者来说，在发展中需要注意以下几点关键因素：

- 该行业的核心在于对于人才标准的定义和培养人才所需的高质量内容体系的构建（包括了教学平台、数据集和问题的完善、内容体系的完整性以及人才能力评估机制），因此一套完整且落地的标准机制是企业占领先机的竞争壁垒；
- 数据集的可获得性、知识的开放性是培养人才的关键，所以社群和生态的做法是符合发展趋势的；数据人才需要享有数据公开、计算成本下降和 AI 公民化等的红利，这其中大公司的责任尤其重大；
- 科研和商业前所未有的紧密合作是推动专业数据人才教育行业发展一个重要因素，需要采取更多灵活的方式支持产研结合的雇佣机制、产权保护机制、利益分配机制；
- 细分市场精耕和生态化是两个必然趋势，前者靠资产（包括知识资产）专用性，后者靠规模优势获得可持续发展；
- 教育重运营，因此社群运营是必需的能力之一，同时需要重视数据知识的沉淀和理念的传播。鼓励及支持为社会输送数据行业专业人才的数据人才教育模式。在形成行业人才的流动生态的同时，更要形成人才社群。

### 6.2 政府部门的政策建议

教育变革是一个渐进的过程，专业数据人才教育还处于刚刚起步的阶段中，对数据人才教育相关的认识也会逐渐更为全面、系统。专业数据人才教育培养工作成了推进大数据产业发展的关键步骤。未来，应以加快大数据在经济、社会等领域应用价值挖掘人才培养为主线，以提升跨界复合型人才能力为主攻方向，积极完善配套措施，积极探索数据科学人才能力评价体系，全面培育数据应用创新型人才。

#### 6.2.1 规范专业数据人才教育体系

从政府层面，鼓励和支持建立规范的培训体系、科学的课程体系、高效的服务体系。初步建成交叉培训、更新技能和再就业人才培训生态链，培育全国领先的数据人才教育组织和机构。优化数据人才培养环

境，将技能考核和工作实践相结合，聚集起一批高潜力的数据人才队伍。

### 6.2.2 构建专业数据科学人才库，用政策引导和鼓励人才教育项目落地

鼓励持续增加专业数据应用人才，打造连贯、系统、实用的知识图谱，在用大数据方法重构大数据学习上实现重大突破，在理论知识的完整联系、行业知识和数据的跨越联系上取得积极进展。

复合型应用人才培养体系进一步规范优化，培养出的人才在企业中既有纵向参与和推进，又有横向对比和维护，具备多学科、多领域的大数据综合能力，成为企业产业升级和转型的骨干力量。

选拔具备专业数据人才教育资质的社会教育机构，以金融、零售、地产等行业应用为课程导向，计算机科学、统计学、机器学习、数据挖掘、可视化、经济学等将通过在这些行业中的广泛应用实现教学。

构建专业数据科学人才库，探索建立数据人才能力评价体系，完善数据人才的认定标准，为培养大数据领域创新型领军人才创造条件，吸引海外数据人才精英。

### 6.2.3 利用政府的号召能力联合行业多方力量，扶持专业数据人才教育项目的孵化

鼓励和支持能够通过国内外的各类创意数据竞赛等平台中挖掘并吸收拥有独特创意的数据人才 / 团队，为大数据领域增添新鲜血液，同时能够为大数据应用领域的发展提供项目创意。

附录：部分专业数据人才教育行业参与者分析

表 1 专业数据人才教育行业参与者--高校

		简述与特点	市场 定位	主要 服务	盈利模式	教学 平台	线下实践	学习周 期	优势	劣势
高 校	北卡 州立 大学	2007 年 2 月，北卡州立大学成立高级数据分析研究院（ Institute for Advanced Analytics , IAA ），成为美国首个正式的数据分析硕士研究生学位授予单位。数据分析软件 SAS 公司参与课程设计并提供数据教学平台	高校 教育	面授 课程	收取学生 课程学费	学校	让不同的企业 提供实习项目， 让学生参与到 实习项目中	10 个月	学期时间短，知 名度高，专业排 名第一，投资回 报率高	学校综合排 名略低，与 SAS 的捆绑 有局限
	康涅 狄格 大学	数据科学相关专业的课程设计建立在学校与美国通用电气公司（ General Electric Company ）长期的技术合作关系之上，学生在修读过程中，能有机会将科研成果应用于实践，在商业分析和项目管理方面获得扎实的理论基础与实践经验。	高校 教育	面授 课程	收取学生 课程学费	学校	无线下实习	24 个月 （ 37 学 分 ）	和 GE 的合作项 目保证就业	时间长，灵活 度不高
	斯坦 福大 学	使用亚马逊提供的 EC2 云平台，让学生进行大规模计算的学习实践	高校 教育	面授 课程	收取学生 课程学费	学校	无线下实习	18 个月 （ 30 学 分 ）	学校知名度高， 就业好	性价比一般/ 学费高
	人大	成立统计与大数据研究院，引进海外优质师资力量，旨在建立界一流水平的统计与数据学科。已推出硕士及博士项目，其中博士项目多与世界数据科学一流	高校 教育	面授 课程	收取学生 课程学费	学校	无线下实习	5 年 （ 硕 博 ）	学校知名度高， 就业好	学期较长

	高校合作						连读 )		
清华	成立数据科学研究院，并宣布将推出多学科交叉培养的大数据硕士项目	高校教育	面授课程	收取学生课程学费	学校	无线下实习	24 个月	学校知名度高，就业好	时间长，灵活度不高
北大	首批申请国家“数据科学与大数据技术”专业的三所高校之一。2016 年开始通过元培学院数据科学方向进行本科生招生，两年将主修数学、计算机科学和统计学方面的基础课程，后两年将结合学生自己的兴趣和能力选修金融、医疗、生物、人文、社科、自然语言处理等多方面数据分析和处理的相关课程并参与实际处理这些数据的一些项目。光华管理学院 2018 年推出了 BA 硕士专业。	高校教育	面授课程	收取学生课程学费	学校	无线下实习	24 个月	知名度高，商业分析方向实用性强	较新的专业，存在不确定性

表 2 专业数据人才教育行业参与者--公益组织

	简述与特点	市场定位	主要服务	盈利模式	教学平台	线下实践	学习周期	优势	劣势
公益组织	edX  由麻省理工和哈佛大学在 2012 年联手创建，是非盈利开源的大型在线课堂平台，两所大学分别资助 3000 万美元。edX 提供高校的线上课程，课程视频全部放在 YouTube 上。截止 2013 年 10 月，edX 共有 29 所教育机构参与，其中北大、清华、香港大学、香港科技大学也在其中。课程方面，人工智能课程和互动实验室是 edX 独有的。	在线课程平台	在线授课	免费 + 收费模式；部分课程向全世界提供免费学习，部分课程收取学费。与组织或企业合作课程收费，有一部分课程设置收费证书。	PC+移动端	1. 有企业提供部分课程的内容设计；2. 优秀学员有 refer 的机会，不保证就业	6-8 周完成一节课程 ( course ) 的学习；成体系的项目 ( program ) 一般包括 2-4 个课程	1. edX 与一些大学联手推出一些项目，提供通过 edX 获得大学认证的正式学分的机会；2. 相对来说课程更贴近大学课程 3. 多种认证证书	1. 课程教少 2. 国内访问不便，很多课程没有中文字幕

表 3 专业数据人才教育行业参与者--培训机构

表 3		简述与特点	市场 定位	主要服 务	盈利模式	教学 平台	线下实践	学习周期	优势	劣势
培 训 机 构	Udacity 优达学城	一所在线的“硅谷大学”，由来自硅谷的技术领袖设计课程，让每个人都能免费在线学习世界最高水平的 IT 课程。优达学城和 Google、Facebook、亚马逊等企业合作推出的“纳米学位”认证项目，将学员培养为世界一流的网站开发者、数据分析师和移动开发者，在当今社会中获得理想工作。 在免费的在线课程之外，优达学城还为学员提供付费的学习导师服务。高度互动课程。提供单项课程，提供微专业(Nanodegree)。可以按照自己的节奏上课。	在线 课程 平台	在线授 课+线 下授课	免费+收费模式；（推出独特的微专业使学员可以从 0 学到 1。微专业提供两周“试用期”，单独课程皆为免费。一年内完成可以领回 50%的学费。）主打和硅谷顶尖企业的合作。每个纳米学位收费不等，以商业数据分析为例：为期两个月的学习收费 1499 元	线上 授课	主打和硅谷 顶尖企业的 合作，证书 有合作企业 认证，含金 量高；但不 提供就业保 证	self- paced。 平均每个 纳米学位 按每周 10 小时 计算，一 般 4-6 个 月可以结 业	1. 部分课程 有人工 review 2. 工 作导向，提 供工业级别 的项目练习 项目 3. 本土 化做的更 好，有微信 客服、讨论 群等	工作导向性 的课程，理 论方面弱化
	Insight Data Science	成立于 2012 年，总部在加州 Palo Alto，这是一个社会团体，提供 7 周密集全职的数据科学培训，有机会获得全额学费奖学金，在顶尖行业数据科学家带领下在项目中学学习，项目结束后可以获得顶尖公司的面试机会。	在 线 课 程 平 台 （ 培	在 线 授 课 + 线 下授课	第一周介绍，第二~四周数据项目学习，第 5~7 周企业访问。项目结束后即参与企业面试、入职。面向学生不收取费	线 下 培训	企业会定向 来招聘，根 据 数 据 来 看，毕业生	7 周左右	1. 课程提供 的技术链新 2. 学生生源 好(名校 PHD	1. 周期较短 2.门槛很高， 只适用于名 校理科毕业

			训 + 中介)		用,学生毕业后,向企业推荐,并收取企业费用。(必须招收博士。)		百分之百成功就业		preferred), 校友强大	的(博士)群体
	<b>Bit-Tiger</b>	是以 project - driven 的项目实战学习平台,主打社区概念,有一些线上线下活动分享内容,打造硅谷华人范科技社区。核心是 bootcamp 培训课程。目前有:系统设计/硅谷程序员直通车/大数据工程师直通车/人工智能/商业分析等课程。	在线课程平台	在线授课	线上授课,在线答疑,以商业分析课程为例,60%左右的内容为数据科学基础课程,40%左右的内容是导师带领完成一个工业界的实战项目。一轮课程时常三个月左右,收费在5000-6000 刀美金。	在线课程	提供优秀学员 refer 的机会,不保证就业	大部分课程周期在2-3 个月	靠近硅谷,教师多有一线工程师背景,课程内容定期有更新	费用偏高
	<b>General Assembly</b>	由早期的沙龙演化的 bootcamp 项目。教授编程,网页设计,数据分析等实用技术,目前在旧金山、悉尼等 14 个城市开课,提供包括软件开发,数据科学等多个方向的培训内容。已经培训超过 24 万名学生。据称其中超过 99% 的人通过培训后,在六个月内找到了相关的工作。	boot camp	线上 + 线下培训结合	全职学生收费 1-1.2 万美金; 兼职学生每一门课收费 3000-4000 美金	公司办公室	会有校友定期回访提供一些就业推荐,并不提供就业保证	3 个月左右	1. 课程贴近实际;2. 全职兼职上课都可自主选择	主要在英语文化圈,对非英语使用者不友好



	<p><b>Mu Sigma 大学</b></p>	<p>是 Mu Sigma 为了培养所谓的 “Decision Scientists” ，创办的企业大学。Decision Scientists = BA + Mathematician+ Programmer。课程包括咨询、应用数学、数据建模、统计学、经济学等。相当于一个 “培养学生定量分析习惯的小型 MBA 课程” 。</p> <p>进入 Mu Sigma 公司的雇员都会先进入 MSU 进行为期 3 个月的培训。来自印度的雇员参与学校培训，他们有专门的项目叫做哥伦布项目。课程结束后员工立即入职，作为 Decision Scientists 服务公司。</p>	<p>企业大学</p>	<p>面授课程</p>	<p>非盈利机构，该项目只开放给公司员工</p>	<p>企业</p>	<p>培养初级的分析师 ,Mu Sigma的企业客户学习处理数据，更加有效的开展业务。</p>	<p>3 个月</p>	<p>1.和企业的合作深化，贴近业务实际；2.主打培养“决策科学家”概念</p>	<p>主要是和印度方面的合作，在印度裔之外的人群中吸引力不足</p>
--	---------------------------	--	-------------	-------------	--------------------------	-----------	---	-------------	--	------------------------------------

表 4 专业数据人才教育行业参与者—竞赛平台和社群

		简述与特点	市场定位	主要服务	盈利模式	教学平台	线下实践	学习周期	优势	劣势
竞赛平台和社群	Kaggle	<p>Kaggle 是由联合创始人、首席执行官安东尼·高德布卢姆( Anthony Goldbloom )2010 年在墨尔本创立的,主要为开发商和数据科学家提供举办机器学习竞赛、托管数据库、编写和分享代码的平台。</p> <p>Kaggle 举办数据科学线上竞赛,吸引大量数据科学家、机器学习开发者,为各类现实中的商业难题开发基于数据的算法解决方案。竞赛的获胜者、领先者,在收获对方公司提供的优厚报酬之外,还将引起业内科技巨头的注意,获得各路 HR 青睐,为自己的职业道路铺上红地毯。该平台已经吸引了 80 万名数据科学家的关注。</p> <p>2017 年 3 月,谷歌收购了 Kaggle,保留了 Kaggle 所有团队,并作为 Google 旗下一个独立品牌继续运营。</p>	社区论坛 / 竞赛平台	<p>提供数据类的竞赛平台,吸引来自世界各地的数据家和相关领域的学习者竞赛;在 B 端,Kaggle 的模式也对接了大批优秀企业,为各类现实中的商业难题探寻算法解决方案。</p>	<p>专注打造社区,盈利模式并不明确,现有的赢利点(企业合作办比赛拿佣金,一般是抽取奖金的 10-20%作为服务费)规模不大</p>	线上	无线下实践	无固定周期	数据竞赛平台的领先者和行业标杆	和企业招聘推荐方面的合作不多

	<p>大数据人社区诞生于 2014 年，其前身是数据分析论坛，于 2015 年 2 月 28 日进行网站调整，更换域名及网站名称。</p> <p>大数据人是一个专注大数据、大数据技术、大数据应用案例和数据可视化的社区型网站。从数据角度出发，讲述大数据健康医疗、互联网金融、电子商务、网络游戏、社交、舆情、教育、征信、物流、税务、工/农/商业等多个领域的应用，并关注科技前沿问题，如：块数据、机器人、互联网+、工业 4.0 等。</p> <p>社区分为工具区——各类数据分析工具板块；资源下载区——包括软件、数据和行业报告等下载；小熊学院——统计、经管、电商、商业智能、大数据、O2O 学习板块；服务器——Linux、Hadoop 板块；编程；数据可视化；考试&amp;比赛区；其他——聊天、招聘、知识普及；以及行业动态板块。</p> <p>大数据人主要通过外部导量，举办活动，以及广告来维持运营。目前社区活跃度较低，多数话题的贴数不超过 100。</p>	社区论坛	社区论坛，会员下载专利等	传统的社群运营模式：提供大数据领域的垂直专业论坛，用户自贡献内容，有专门团队运营，采用会员制。	N/A	N/A	N/A	较早的大数据论坛沉淀出很多有用的信息	社区活跃度低，人才流失严重
大数据中国	<p>大数据中国提供数据交易、求职招聘、学习资料、企业入驻、活动峰会和广告服务。在线人数平均在 100~200，各话题发帖数不一，目前最活跃的话题是数据挖掘，发帖数超过 1000。</p>	社区论坛	社区论坛，会员下载专利等	传统的社群运营模式，和大数据人的相似度很高	N/A	N/A	N/A	行业资讯丰富	社区活跃度一般，

	<b>Shinect</b>	Shinect【巷内】创业社区，是一群硅谷高科技行业精英，因“初创”梦自发成立的非盈利公益组织。Shinect 创业社区实行会员推荐制，目前拥有 700+会员，已走出 100 多位全职“创客”，所创产品涉及大数据，移动互联网，信息安全，医疗，在线教育，互联网金融等多个领域。Shinect 致力于打造一张高质量、高密度的紧密社交网，为成员提供多样化的创业资源，信息和技术的交流平台。定期举办线下活动。	社区论坛	会员福利：人脉拓展，商务合作，提供投资圈的资源等等，相当于一个创业社区+加速器	不对外收费，盈利可能来自于基金管理费、加速器培训项目、活动及会议、合作项目或者赞助、以及教育项目	N/A	N/A	N/A	社区 + 加速器的设置	目前局限于硅谷地区
--	----------------	--	------	---	--	-----	-----	-----	-------------	-----------

表 5 专业数据人才教育行业参与者—媒体/出版社

表 5	简述与特点	市场定位	主要服务	盈利模式	教学平台	线下实践	学习周期	优势	劣势
媒体 / 出版社	<b>Kdnuggets</b> KDNuggets 成立于 2001 年，位于布鲁克林。作为大数据媒体，KDNuggets 专注于提供商业分析、大数据、数据挖掘、数据科学和 ML 的咨询。KDNuggets 的推特在 2013 年被 Big Data Republic 评为最好的大数据推特。KDNuggets 还提供数据挖掘、NLP 等课程，还有企业招聘信息。	媒体	在线课程	课程是免费的，自身是典型科技媒体的盈利模式	线上	无线下实习	14 周	免费，提供录播以及免费的学习资料等	互动性差，知名度不高
	<b>机器之心</b> 机器之心 Synced 创立于 2014 年，是国内首家系统性关注人工智能的科技媒体。创办至今，始终坚持提供专业的、客观严肃的高质量学术及产业内容，向人工智能从业者及爱好者输出真正有价值的信息和知识，借助优质内容的力量来正确引导甚至影响人工智能行业的发展。目前机器之心在各平台拥有共计 50 万高质量用户，微信端日均 PV 5 万，日活跃用户 6 万人。 与此同时，机器之心在全球人工智能重点区域均部署有专职技术分析师、行业分析师和顾问团队，输出原创技术分析文章与产业分析报告，关注全球人工智能行业一手新闻报道。	新媒体	内容分享	内容分享免费，依靠线下精品活动，赞助及广告的盈利模式	线上访谈/线下活动	无线下实习	N/A	人工智能领域影响力大，用户数量庞大，日活率高。分析文章较为专业	以人工智能领域为主，缺乏数据科学方向的内容。无教学平台，以内容分享为主。

	<b>大数据文摘</b>	成立于 2013 年，专注于数据领域资讯、案例、技术，已成为数据行业最具影响力的新媒体。聚集了一批垂直领域订阅者，覆盖读者超过 60 万，在多家具具有影响力的网站、杂志设有专栏，致力于打造精准数据分析社区。	新媒体	在线公开课 / 内容分享	内容分享/免费课程	线上	无线下实习	不固定	覆盖面广，团队成员有较深的计算机和行业背景。数据科学与人工智能领域有影响力。	缺乏数据科学方向的深度内容。有简单的免费课程，但缺乏专业的教学课程
--	--------------	---	-----	--------------	-----------	----	-------	-----	--	-----------------------------------

表 6 专业数据人才教育行业参与者—生态玩家

表 6	简述与特点	市场定位	主要服务	盈利模式	教学平台	线下实践	学习周期	优势	劣势
生态玩家	<b>奇点大学</b> 由谷歌（Google）、美国国家航天航空局（NASA）以及若干科技界专家联合建立的一所新型大学。地址位于美国加州硅谷 NASA 研究中心内，于 2009 年起正式招生。学校设立未来学和预测学、设计学、网络和计算机、生物技术和生物信息学、纳米技术、医学和神经科学、人工智能机器人和认知学、能源和生态科学、空间物理学、政策法规和伦理学、企业家学等学科。 奇点大学校长雷·库兹韦尔（Ray Kurzweil）在 2005 年出版《奇点临近》，学校也因此得名。讲师包括美国国防部及 NASA 顾问 Neil Jacobstein，谷歌无人驾驶项目顾问 Brad Templeto，生物技术和信息学负责人 Raymond McCauley 等。培养目标是面向科学界和商业精英，培养改变世界的未来领军人物。 奇点大学获得 NASA 和 Google 的资助，在课程结束后奇点大学会跟踪孵化优秀项目，截止 14 年 3 月，奇点大学已经投资了 24 个项目。	企业大学	面授课程	提供顶尖专家和课程，课程费用高达 2.5 万美元，课程周期 9 周。	学校	虽然并不提供真正的硕士学位，但是奇点大学本身的孵化器和风投基金以及校友资源都非常强大，这里的毕业生不是以找工作为目的，而是强化人脉，寻求资源等	9-10 周	1.校友资源强大 2.课程内容紧贴时代前沿，每年有多达五次的修改和更新 3.知名度高	几个月的 master，只有课程，没有研究项目，学费奇高。不符合美国政府对于正式授予文凭的定义和要求。
	<b>Facebook</b> Facebook 大学是 Facebook 面向社会开放的一个浸入式项目，官方的定义是一个跨越工业界、商界的实习项目，主要面向在校学生，通过工	企业大学	本质上是实习项目	非盈利，少量项目	公司本部	本部实习工作	9-10 周	紧贴工业界的真实问题，提供	仅限于本科生申请



	<b>k 大 学</b>	业项目实践使学生获得业界的真实实践经历.以 Facebook University for Analytics 为例提供一个付费的,为期八周的项目。第一周教学知识,后七周导师项目经验指导。主要学习 SQL 方面的相关知识。			目收取 一些培 训费用				return offer 的 机会,是本科生 进入 FB 的非常好 的机会	
	<b>腾讯 大学</b>	成立于 2014 年,包括微信学院、开放平台学院、营销学院和游戏学院。提供课程培训,知识分享,以及各类讲座。腾讯大学与众多媒体以及数据咨询公司合作,如 36kr、易观、艾瑞等。目前腾讯大学的人工智能等培训主要通过讲座的形式进行,同时,各类课程也会邀请到各行业 KOL 主讲,包括马化腾本人。	在 线 课 程 平 台	在线授课	免 费 公 开 课 程 + 收 费 精 品 课 程 + 合 作课程	在 线 课 程	腾讯自身的课程无线 下实习。第三方合作 课程中有的课程含有 线下实习环节	不 同 的 课 程 周 期 不 同。	合作伙伴较多, 讲师资源丰富, 课程分类丰富, 内容丰富	课程内容 不够体系 化,较为 散乱。部 分课程收 费较高

表 7 专业数据人才教育行业参与者—众包技术平台

表 7		简述与特点	市场定位	主要服务	盈利模式	教学平台	线下实践	学习周期	优势	劣势
众包技术平台	科赛网	科赛网是智能时代在线数据应用平台，推出的 K-Lab 在线数据分析协作平台，极大提高数据人才工作效率。已吸引国内大量优质数据分析师人才的加入，为企业提供数千份数据算法、数据产品方案（携程、上海联通等）。公司日前获得千万元 PreA 轮投资，投资机构为线性资本。	竞赛平台/bootcamp/在线学习平台	在线数据应用平台，提供大数据实战、竞赛服务，通过开放数据集、在线数据分析协作平台、打造真实的行业实战项目，	竞赛平台作为主要赢利点，依靠竞赛平台抽取企业的佣金。	线上	无线下实习	N/A	AI 和数据科学方向较为细分且垂直。在线数据分析平台应用较为广泛，目前积累了数万名数据科学家和 AI 人才。	偏向于技术支持，缺乏运营能力

## 参考文献

- 调研样本: 回收问卷总数共计 238 份, 有效问卷数 235 份; 其中 152 份为专业数据人才, 其他为非专业数据行业从业人员。
- 美世全球人才趋势报告[R] 全球: 美世咨询公司, 2017
- 美世全球高科技人才趋势报告[R] 全球: 美世咨询公司, 2017
- 中国培训行业研究报告 2016-2017; 《培训》杂志&安德曼研究院

## 本报告主创团队如下（排名不分先后）：

杨慧

孙可意

季青

王俊

谢若涵

此外，感谢谢凯林、宋神桥、马斋、孙一鑫、李明喆、廖彧慧、孙泽洲、叶杰生、刘英在创作中给予的支持。

## 美世联创团队如下（排名不分前后）：

乔鲁诺

孙悦

刘源

杨麟

# 关于 TalkingData 和 TDU

TalkingData 成立于 2011 年，是国内领先的第三方数据智能服务商。

借助以 SmartDP 为核心的数据智能应用生态为企业赋能，帮助企业逐步实现以数据为驱动力的数字化转型。

TalkingData 成立以来秉承“数据改变企业决策，数据改善人类生活”的愿景，逐步成长为中国领先的数据智能服务商。以开放共赢为基础，TalkingData 凭借领先的数据智能产品、服务与解决方案，致力于为客户创造价值，成为客户的“成效合作伙伴”，帮助现代企业实现数据驱动转型，加速各行业的数字化进程，利用数据产生的智能改变人类对世界以及对自身的认知，并最终实现对人类生活的改善。

TalkingData 不仅专注于数据智能应用的研发和实践积累，同时也在积极推动大数据行业的技术演进。早在 2011 年成立初始，TalkingData 就组建了数据科学团队，将机器学习等人工智能技术引入海量数据的处理、加工流程中。通过几年来的不断发展，TalkingData 已在大数据、人工智能领域拥有多项国家专利。此外，TalkingData 还开源了大规模机器学习算法库 Fregata、UI 组件库 iView、地理信息可视化框架 inMap 等项目，在海内外得到广泛支持与认可，使用者和贡献者遍布全球。

目前 TalkingData 设立了包括硅谷边缘计算实验室、人本实验室在内的多个大数据、人工智能实验室，并与 MIT 媒体实验室、斯坦福人工智能实验室、加州理工航天技术实验室等国际顶尖学府、研究机构展开合作，

共同加速大数据、人工智能相关技术的探索和演进，并将国际前沿技术引入高速发展的中国市场，与国内丰富的应用场景相结合，驱动新技术的落地应用与行业的飞跃发展。

腾云大学 ( TalkingData University , 简称 TDU ) 是 TalkingData 在专业数据人才领域的生态运营组织，致力于创新数据科学人才培养模式，创造数据科学价值。腾云大学自创立以来，已与国内外一流高校、专业数据人才社群、教育机构、媒体等形成完整的专业数据人才教育生态。

基于专业数据人才教育和科研平台，腾云大学得以为专业数据人才提供全职业生命周期的培养方案，为专业数据人才包括入职前教育培训、课题研究，求职中就业服务，以及入职后线下活动、职业咨询等。同时，腾云大学推出的数据驱动领导力、行业精品培训课程等产品可以帮助正在数字化转型的企业构建数据团队、提高数据能力、提升数据价值。

访问 TDU 官网：<https://edu.talkingdata.com>

# 关于美世

在美世，我们致力于改变全球 1 亿 1 千多万人的生活，帮助他们改善健康、财富和职业。我们的专业，是为客户以及他们的员工构建更具有安全感和成就感的未来——无论是为他们设计更经济划算的健康计划，帮助他们确保退休后的收入，还是实现员工和劳动力需求的最佳匹配。我们以分析和洞察作为变革的催化剂，预测并理解每一个商业决策对当下以及未来的影响。我们通过创新的视角体察人们现在和将来的需求，而我们提供的每个解决方案，都立足于全面的思考、深度的分析和专业的把握。70 多年来，我们将洞察转化成行动，服务于各种机构，帮助他们的员工过更健康的生活、发展他们的职业，创造更有安全感的未来。在美世，我们把握今天，成就明天。

美世和全球各地的美世分支机构都隶属于上市公司 Marsh & McLennan Companies（威达信集团），威达信集团在纽约、芝加哥和伦敦交易所上市（代码：MMC）。

更多信息，敬请联系各地美世办公室或访问美世中国官网

[www.mercer.com.cn](http://www.mercer.com.cn)





**TD Mail:** [support@tendcloud.com](mailto:support@tendcloud.com)

**TDU Mail:** [tdu@tendcloud.com](mailto:tdu@tendcloud.com)

关注TalkingData、TalkingData University 获取最新资讯