

第五章 大数据人才资源

大数据来势迅猛，仿佛一夜之间大家都在谈大数据、做大数据、用大数据。于是，大数据人才变成了紧缺人才。根据麦肯锡报告，“到 2018 年，美国在‘深度分析’人才方面将面临 14 万至 19 万的人才缺口；在‘能够分析数据帮助公司做出商业决策’方面将面临 150 万的人才缺口”。另一方面，《哈佛商业评论》声称，21 世纪最富挑战的工作是数据科学家。很多企业开始设立数据科学家岗位。美国社交媒体公司 DataSift 的创始人兼 CTO Nick Halstead 认为：“大数据的真正价值就在于‘数据科学家’这一提法的传播”。据 Gartner 预测，到 2015 年，全球将新增 440 万个与大数据相关的工作岗位，且会有 25% 的组织设立首席数据官职位。其中有 190 万个工作岗位将在美国。而每一个与大数据有关的 IT 工作，都将在技术行业外部再创建 3 个工作岗位，这将在美国再创建将近 600 万个工作岗位。但是，Gartner 也同时指出，拥有大数据技能的 IT 专业人员严重短缺，只有 1/3 的新的工作岗位将雇佣到人员。

从广义上讲，大数据人才就是数据科学家和数据工程师，因此，大数据人才的培养就是数据科学家和数据工程师的培养。这从国际上开设的《数据科学》课程、数据科学学位计划、数据科学短期培训班可以看出这一点。

在中国，香港中文大学从 2008 年起设立“数据科学商业统计”科学硕士学位；复旦大学从 2007 年起开设数据科学讨论班，2010 年开始招收数据科学博士研究生，并从 2013 年起开设研究生课程《数据科学》；北京航空航天大学于 2012 年设立大数据工程硕士学位。在美国，加州大学伯克利分校（UC Berkeley）从 2011 年起开设《数据科学导论》课程，并从 2012 年起开设《数据科学和分析》课程；伊利诺伊大学香槟分校（University of Illinois at Urbana-Champaign, UIUC）从 2011 年起举办“数据科学暑期研究班”（Data Sciences Summer Institute program）；哥伦比亚大学（Columbia University）从 2011 年起开设《数据科学导论》课程，2013 年起开设《应用数据科学》课程，并将从 2013 年秋季起开设“数据科学专业成就认证”（Certification of Professional Achievement in Data Sciences）培训项目，并计划从 2014 年起设立硕士学位，2015 年起设立博士学位；芝加哥大学（University of Chicago）开设 3 个月的夏季培训课程；纽约大学（New York

University）将从 2013 年秋季起设立“数据科学”硕士学位；南加州大学（South California University）设立“数据科学”硕士学位；华盛顿大学（University of Washington）从 2013 年 5 月起开设《数据科学导论》课程，并对修满数据科学相关课程学分的学生颁发数据科学证书（Certificate in Data Science）；雪城大学（Syracuse University）也提供数据科学高级研究证书（Certificate Advanced Studies in Data Science）培训项目。在英国，邓迪大学从 2013 年起设立“数据科学”科学硕士学位。

从上述人才的培养计划来看，数据科学家应该系统地掌握数据分析相关的技能，主要包括数学、统计学、数据分析、商业分析和自然语言处理等，具有较宽的知识面，具有独立获取知识的能力，具有较强的实践能力和创新意识。其中，只有复旦大学的课程设置强调了数据科学家是研究数据的科学家，而不仅仅是一个数据工程师或者数据分析师。

目前，大数据人才培养可以分成两个方面：一个是学位培养，另一个职业培训。

5.1 数据科学学位人才培养

该类型人才的培养主要包括本科生、硕士生、博士生，向其颁发数据科学学士、硕士和博士学位，为政府和公司输送数据科学家。而培养数据科学家，除了需要很好掌握数学、计算机科学和应用统计学等基础知识点外，还需深入学习经济、生物、物理、化学等交叉学科业务课程，并在数据获取、数据存储、数据检索等数据工程方面做深入的了解和亲身实践。IBM 公司的“全球大学关系项目”总监、同时也是计算机科学家的吉姆·斯伯热表示，从学术角度看，在一些本来跟数据无缘的学科里，比如社会科学和人文学科的一些分支，大数据也正在发挥重要作用。数据科学学位的人才培养需要关注如下几点：

➤ 数据科学家储备欠缺

大数据职位相关的技能主要包括数学、统计学、数据分析、商业分析和自然语言处理，数据科学家是复合型人才，是对数学、统计学、机器学习等多方面知识的综合掌控。

大数据最关键的部分是数据分析和挖掘数据价值，要获得这些，就需要大量

的数据科学家。数据科学家是复合型人才，是对数学、统计学、机器学习等多方面知识的综合掌控。初级的分析人员只能是对数据进行报表、描述性分析，真正高级的数据科学家需要对数据做出预测性的、有价值的分析。从目前的人才储备来看，这部分的储备欠缺。

➤ 掌握机器学习和知识图谱很重要

从计算机学界的理解来看，大数据的核心技术是机器学习和知识图谱。这是一种框架性的知识，介于基础设施和应用之间的技术。例如大数据应用的代表谷歌公司就有两个大的开发方向，一个是机器学习，另一个是由搜索团队负责的知识图谱。

任何一种大数据方案都不可能适合所有的行业，因此，大数据的核心业务必然是一种扎根于特定行业，综合运用已有的存储、分析、挖掘、展现技术，根据用户需求并融入行业特色技术模型的一站式大数据平台业务。正是由于大数据具有这样的业务特点，所以企业最需要两种人才：一类是复合型人才，另一类是技术专家。一方面，大数据具有强烈的行业特点，这就需要复合型人才，这种人才需要了解行业，了解技术的各个层面，以综合的视角制定确实可行的方案为目的，还必须具有统计学背景，并对数据管理有丰富经验，他们是目前最急缺的人才；另一方面，大数据方案的实现，必须由技术专家来完成，技术专家的能力也直接决定了企业所能制定大数据方案的深度和广度，传统的数据库应用开发，特别是商业智能应用开发人才，以及熟悉 Hadoop 等分布式存储的人才，也都是必不可少的。

➤ 大数据人才培养需要校企合作

企业可以与学校联合培养人才，或建立专门的数据科学家团队，或与专业的数据处理公司合作，以解人才之急。对于企业来说，虽然人才储备有缺口，但是大数据业务还是得做。虽然目前大数据应用比较少，人才也比较少，但是中国的知识积累并不少，例如中国的学术界和产业界在机器学习上也有积累，现在的问题是如何将这方面知识和大数据结合起来。

由企业和大学合作来培养自己所需要的大数据人才，是考虑到大数据的解剖对象是大量的数据，这些数据只有企业才有，而学校并不生产数据。在企业的支持下，学校就能通过针对性的实践训练来培养学生的技能。

5.2 数据科学职业人才培养

该类型人才的培训主要针对大数据在商业和数据分析中的应用、商业智能的管理者、数据库专家和在校欲将数据科学作为未来职业的研究生，提供中短期培训项目，培训合格后向其颁发数据科学培训证书，从而培养符合国民经济发展战略急需的数据工程师和数据分析师。该类型人才的培养除了要初步掌握数学、计算机科学和应用统计学等基础知识点，需着重学习数据获取、数据存储和数据检索等数据工程方面的知识，并根据其所在领域参与商业大数据项目的分析和处理。

大数据分析的广泛应用要求对现有岗位进行再培训，同时也会出现新职位和新技能，比如“首席数据官”就会成为大大小小的企业里司空见惯的职位。这些岗位也并不一定就非要资深数据分析专家才能胜任，那些在各自行业中受过良好教育、经验丰富的非数据专家只要能善于使用大数据工具就能担任。数据科学职业人才就业的行业和岗位需求以及发展趋势如下：

➤ **大数据职业人才就业的主要行业包括：**

- （1）零售、保险、电子商务
- （2）政府数据中心
- （3）医药和银行
- （4）研究性大学
- （5）金融机构
- （6）互联网企业

➤ **典型的大数据专业岗位需求：**

（1）大数据系统研发工程师：负责大数据系统研发工作，包括大规模非结构化数据业务模型构建、大数据存储、数据库架构设计以及数据库详细设计、优化数据库构架、解决数据库中心建设设计问题。他们还负责集群的日常运作、系统的监测和配置、Hadoop 与其他系统的集成。

（2）大数据应用开发工程师：负责搭建大数据应用平台、开发分析应用程序。他们熟悉工具或算法、编程、包装、优化或者部署不同的 MapReduce 事务。他们以大数据技术为核心，研发各种基于大数据技术的应用程序及行业解决方案。

（3）大数据分析师：运用算法来解决分析问题，并且从事数据挖掘工作。他们最大的本事就是能够让数据道出真相；此外，他们还拥有某个领域的专长，

帮助开发数据产品，推动数据解决方案的不断更新。

（4）数据可视化工程师：具备良好的沟通能力与团队精神，责任心强，拥有优秀的解决问题的能力。他们负责在收集到的高质量数据中，利用图形化的工具及手段的应用，一目了然地揭示数据中的复杂信息，帮助企业更好的进行大数据应用开发，发现大数据背后的巨大财富。

➤ **大数据领域从业人员的十个趋势。**

- （1）薪金将继续增长
- （2）大数据人才供不应求
- （3）雇佣外包
- （4）人才团队内出现分歧
- （5）大数据专业人士需要不断进步
- （6）精通大数据的专业人才将成为最重要的业务角色
- （7）大数据领域需要数据科学家
- （8）高校应对大数据人才缺口
- （9）数据驱动的工作令人满意并充满挑战
- （10）大数据专业人士将拥抱未来