



掌握数据科学的艺术

如何建立具有凝聚力的团队，创造卓越的业务成果

IBM 商业价值研究院

执行报告

业务分析与战略

IBM 大数据和分析

IBM 的大数据和分析实践将管理咨询方面的专业知识与分析科学相结合，助力领先的企业迈向成功。IBM 为客户提供所需的专业知识、解决方案和能力，帮助客户将智能融合到各个业务决策和流程中，从而能够最大程度发挥大数据和分析的潜力；充分利用各种格式的数据和洞察力，更快地采取行动；同时，还通过在安全、监管与合规方面的积极举措，帮助客户建立相互信任的企业文化。如欲了解 IBM 大数据和分析产品/服务的更多信息，敬请访问：
www.ibm.biz/bigdataandanalytics

以智能方式获得关键人才和技能

全球的企业高管都在寻找提升收益、提高运营效率、改善影响分析和降低成本之道，这使得数据科学家的重要性愈发彰显，因为他们能够在决策过程中提供丰富而有价值的信息。成功的数据科学家能够帮助企业领导理解挑战、定义分析解决方案并分析数据，从中发现创新和洞察。如今优秀人才供不应求，因此，如果企业希望建立或扩大数据科学团队，就必须采用智能化的方式。本报告提供了相关的专家建议，探讨如何建立具备相关经验、扎实的学术基础、不同背景和技能、能够设想并努力实现目标的强大人才团队。

执行摘要

在当今的数字时代，数据科学是高管们热议的话题之一。引领数字化颠覆潮流的高管讲述了他们如何应用数据科学解决复杂问题，发掘有价值洞察的故事，这也引起了全球听众的共鸣。但是，对于希望通过数据获得深入洞察和价值的高管而言，必须明白在制定强大的数据科学计划方面的需求和风险。

麦肯锡公司预测，仅在美国，深度分析专业人员职位的供需缺口在 14 万到 19 万个之间，而且他们指出，这种人才短缺是全球性的。该研究公司解释了培养此类人才的难度，估计需要“数年时间才能培训出具备内在数学能力的人才”。¹

全球对数据科学家的需求如此之大，以致招聘公司 Glassdoor 将其列为美国最热门的职位，并在工作满意度和职业发展机会方面都给出很高的评分。² 近期一项针对 LinkedIn 的全球数据库分析（重点考虑代表性的市场领域）结果表明，数据科学家职位缺口超过 60,000 个。³ 而另一项分析结果显示，全球范围内仅有 11,400 名专业人员具备所需技能。⁴ 此外，数据科学家职位的数量有了“令人瞩目的增长”，LinkedIn 上自称“数据科学家”的人中，至少有 52% 在过去四年内获得了这一头衔。⁵



优秀的数学家会**应用技能和科学，获得**应对企业所面临最大挑战所需的洞察。



企业对数据科学家的**需求量极大**，因此人才**争夺战异常激烈**。



数据科学家的薪酬非常可观，**企业聘用以及留住此类人才的成本非常高**。

数据科学家的薪酬非常可观，企业聘用以及留住此类人才的成本也相当高。美国商务部的研究发现，数据工作者的薪资平均比私营企业工作者总体水平高 68%，而且数据科学家则是所有数据工作者中薪资最高的群体。⁶

这种供需之间的不平衡给求贤若渴的高管层带来了风险。要获得深入洞察，解决企业所面临的**最大挑战**，也就是运营优化、创收和创新，就需要真正的数据科学家发挥自身所具备的科学和艺术才能：如果数据科学家经验不足，或者未经过充分培训，那么他们所采用的错误方法或有缺陷的分析可能会导致灾难性后果（见图 1）。⁷

要建立能够实现出色解决方案的成功数据科学计划，招聘数据科学家只是步骤之一。而要实现这些目标，高管还需思考组织架构、工具、途径和结果。

几十年来，拥有实际经验的数据科学家们致力于解决重大的战略性业务挑战，为数以百计的客户团队提供咨询，并对 IBM 内部和外部成千上万有望成为数据科学家的人员进行培训，这样“三管齐下”，使我们取得了不小的成就。

这些经历帮助我们形成有益的经验总结。在这份 IBM 商业价值研究院报告中，我们就企业在建立有效的数据科学计划时需要考虑的关键因素提出了自己的看法 – 以技艺精湛的数据科学家为中心，创造适量的数据科学艺术，实现令人瞩目的成果。

寻求具有相关经验的专业人员

根据 IBM 的定义，数据科学家与业务领导合作，通过理解、准备和分析数据，预测新趋势，提供有助于优化业务结果的建议，从而解决各种业务问题。在这个定义中，包含了数据科学家要取得成功必须具备的五个关键特征：

 **敏锐的商业头脑：**了解企业业务战略和执行；能够倾听领域专家的意见，快速掌握基础业务流程并了解其运作方式；具备将业务问题转化为分析解决方案的专业知识；还应具备企业转型经验。

 **深厚的分析专业知识：**能够确定用于解决各类业务问题的适当分析技术；精通基本和高级的数据挖掘方法，包括回归分析、聚类分析、决策树、神经网络、贝叶斯机器学习方法以及优化、模拟和随机分析。

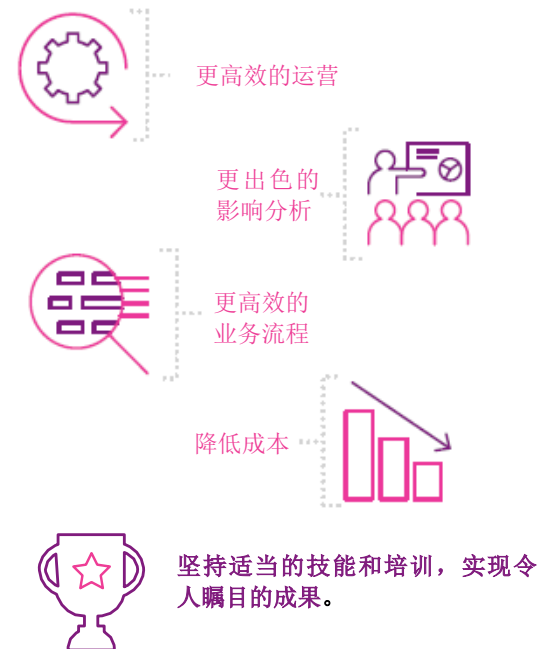
 **高级软件知识：**能够确定要使用的软件包；具备 SPSS Modeler、SPSS Statistics、SAS、R、Python 等关键工具的使用经验；能够设计、开发和应用适当的计算方法来解决业务问题；并且能够创建可重复的自动化流程。

 **21 世纪的数据管理技能：**了解关键的内部和外部数据源以及如何收集、存储和检索数据；拥有处理大量数据（结构化和非结构化、本机和非本机）的经验；熟悉大规模并行平台；熟悉 SQL、NoSQL 和 Hadoop 等工具；熟悉 HDFS 基础架构，如 Pig、Hive、Hue、Sqoop、Hbase 和 Flume；熟悉加速器（如 PureData 或 Exadata）和数据分析语言（如 Groovy）。

图 1

企业需要具有丰富经验与合格资质的数据科学家来应对最严峻的挑战

数据科学家运用自己的能力和技艺，实现至关重要的业务成果：





远见卓识以及讲述技巧：能够有效地将分析部署到企业之中，创造出价值；协助高管重新设计业务流程；利用机器学习、人工智能和认知解决方案自动执行规定性行动和持续学习。

这五大特征只是“入场筹码”；相反，快速区分出色和良好数据科学家的则是软技能：好奇心、科学思维、沟通和可视化技能。数据科学家依靠这些软技能在企业内与其他员工建立协作关系，与利益相关方合作并进行有效的展示。

好奇心可以帮助数据科学家深入其他人尚未探索或未曾考虑的领域，以独特的角度审视业务挑战，拓宽企业的思维。专注于问题的科学思维，不仅仅是研究使用哪些工具和数据，还需要帮助数据科学家着眼大局，制定执行计划。沟通能力和可视化技能体现数据科学的艺术：能够将数据科学化繁为简，体现为切实可行的措施和可预测的结果，并通过各种数据动画和可视化方法、图表和图形呈现出来。

擅长所有这些技能的专家非常罕见（因此许多人称之为“独角兽”）。⁸企业高管应当对这些特征进行优先排序，根据企业的具体情况明确哪些是最重要的，并寻求最合适的候选人，而不是寻找像神话动物般罕见的全才。可以利用大型数据科学团队的专业知识来弥补差距。

需要强大的学术基础来实现深度科学

从根本上说，数据科学是一种扎根于强大学术背景的专业。高管应当寻找接受过定量研究学科（比如统计学、运筹学、机器学习、信息学、计量经济学或物理学）学术培训的候选人。一项对 LinkedIn 全球数据库的分析发现，80% 的数据科学家拥有研究生学位，其中 38% 获得博士研究生学位，42% 获得硕士研究生学位（见图 2）。⁹

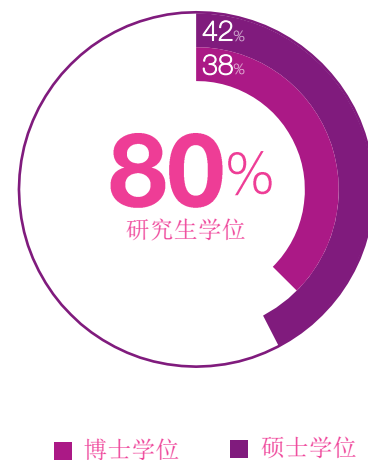
数据科学团队中初级水平的成员应具备定量研究学科领域的学士学位，可能还需要线性代数、应用统计学和机器学习等学科的培训。根据我们的经验，在这些领域中拥有高级学位的人才有能力从事数据科学项目。

现在出现了一个新趋势，那就是企业将商业智能（BI）分析师改头换面，经过粗略的课程和工具培训，就当作数据科学家使用，这一点十分令人担忧。我们认为这是非常危险的权宜之计，高管应当谨慎对待。虽然大多数经过培训的分析师确实可以在职能或部门层面执行探索性或初级数据科学活动，但我们不认为他们有能力取代负责企业使命而且训练有素的数据科学家。我们认为，这些新任命的数据科学家通常准备不足，无法针对战略性的业务挑战提出能够创造价值的解决方案，也可能无力领导成熟的数据科学团队。

数据科学家作为“开路先锋”，发挥着重要的作用，他们需要能够解释最新的技术和数学概念，并使其成为常规工作的一部分。许多此类项目在使用数据、部署的数学方法或所针对的业务问题方面都存在巨大风险，可能导致管理层对不完善的模型下赌注。接受过高水平教育的数据科学家能够有效消除这些风险。

图 2

数据科学家往往接受过高等教育



来源: Stitch, Inc. 2016.

实际经验总结

诀窍是要了解最佳方法

一位数据科学家无意中听到两个统计学专家讨论一个营销难题。

统计学专家创建了一个启发式方法，对数据集执行重复循环，直到基本上对各地区的部门分配了最合理的营销处理。¹⁰ 遗憾的是，这个过程耗时超过 18 个小时，使电脑不堪重负，而最后的结果也没有达到业务目标。

数据科学家在几个小时内就建立了一个整数规划来完成分配。¹¹ 这不仅保证解决方案实现最优，而且在相同的硬件上运行不到一秒钟，实现了迭代和创新，开发出更出色的促销活动。

鼓励采用严格的方法，实现更出色的价值

数据科学家需要深入了解企业所面对的业务挑战。我们认为，只有在自上而下的企业流程以及应用科学对付挑战的过程中采用严格的协作式方法，才能实现数据科学的成功。

数据科学家扮演双重角色：首先，要了解企业面临的业务挑战，定义分析解决方案，自动执行或优化流程；第二，与更广泛的数据科学团队合作，使用基于发现的数据分析方法发掘创新和洞察。企业需要促进数据科学家与业务领导以及主题专家（SME）建立紧密联系，优化他们所创造的价值（见图 3）。¹²

图 3

协作可以改善流程和成果



预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_38897

