

简译版

解密大数据分析的八个谬误

非官方中文译文·安天技术公益翻译组 译注

文档信息

原文名称	Debunking 8 Big Data and Analytics Myths		
原文作者	Andrew Froehlich	原文发布日期	2017 年 9 月 21 日
作者简介	Andrew Froehlich 拥有十多年的企业网络咨询经验，专门从事企业网络架构和数据中心建设。 https://www.informationweek.com/author-bio.asp?author_id=929		
原文发布单位	InformationWeek		
原文出处	https://www.informationweek.com/big-data/debunking-8-big-data-and-analytics-myths/d/d-id/1329930		
译者	安天技术公益翻译组	校对者	安天技术公益翻译组
分享地址	请浏览创意安天论坛 bbs.antivy.cn 安天公益翻译板块		
免责声明	<ul style="list-style-type: none"> 本译文译者为安天实验室工程师，本文系出自个人兴趣在业余时间所译，本文原文来自互联网的公共方式，译者力图忠于所获得之电子版本进行翻译，但受翻译水平和技术水平所限，不能完全保证译文完全与原文含义一致，同时对所获得原文是否存在臆造、或者是否与其原始版本一致未进行可靠性验证和评价。 本译文对应原文所有观点亦不受本译文中任何打字、排版、印刷或翻译错误的影响。译者与安天实验室不对译文及原文中包含或引用的信息的真实性、准确性、可靠性、或完整性提供任何明示或暗示的保证。译者与安天实验室亦对原文和译文的任何内容不承担任何责任。翻译本文的行为不代表译者和安天实验室对原文立场持有任何立场和态度。 译者与安天实验室均与原作者与原始发布者没有联系，亦未获得相关的版权授权，鉴于译者及安天实验室出于学习参考之目的翻译本文，而无出版、发售译文等任何商业利益意图，因此亦不对任何可能因此导致的版权问题承担责任。 本文为安天内部参考文献，主要用于安天实验室内部进行外语和技术学习使用，亦向中国大陆境内的网络安全领域的研究人士进行有限分享。望尊重译者的劳动和意愿，不得以任何方式修改本译文。译者和安天实验室并未授权任何人士和第三方二次分享本译文，因此第三方对本译文的全部或者部分所做的分享、传播、报道、张贴行为，及所带来的后果与译者和安天实验室无关。本译文亦不得用于任何商业目的，基于上述问题产生的法律责任，译者与安天实验室一律不予承担。 		

解密大数据分析的八个谬误

Andrew Froehlich

2017 年 9 月 21 日

与其它新兴技术一样，大数据分析一直笼罩着一层神秘的面纱。本文将解密大数据分析的八个谬误，帮助您推进分析策略。

毫无疑问，多年来，大数据分析的概念不断遭到误解。早期采用者在许多领域挣扎前行，最终导致高于预期的失败率和投资回报率低下。然而，过去的许多错误早已被克服，但是仍然有一些关于大数据分析概念和实施步骤的谬误。



尽管追踪记录（或观察）不尽如人意，但大数据仍然是一个大问题。IDC 在 2016 年第三季度发布了一项预测，显示大数据和分析市场同比增长率达到两位数。如果这一点属实，那么关于大数据分析的许多观点肯定是错误的，对吧？

谬误之所以能够长久流传，是因为谬误中混杂了一小部分真实内容，让人分不清真假。复杂的技术也是这样，它们常常被夸大，导致最终的采用慢于预期。大数据就是这样一种技术，其它复杂技术包括软件定义的 WAN（SD-WAN）、IT 安全，甚至云计算。然而，如果这些技术适合企业，谬误最终会被打破，真理会浮出水面。

今天，我们将会介绍围绕着大数据和分析的八个谬误。当你翻阅幻灯片时，试着弄清楚

真相在哪里被曲解导致形成了谬误。这是打破谬误回归真理的最佳方法。在大多数情况下，围绕大数据或分析的一些谬误源于早期采用者的错误判断。在其它情况下，谬误源于缺乏运行大数据项目所需的技能和工具的企业 IT 部门。最后，还有一些谬误源自关于大数据架构概念和组件的错误信息和错误沟通。

谬误 1：保存所有数据，总有一天会用得着



在大数据活动的早期，最大的谬误之一是：企业应该保存收集到的所有数据。对于这样做的人来说，他们面临着以最低的成本存储所有数据的任务。许多人寻求基于云的数据存储技术，如 Amazon Glacier 或 Google Coldline Storage。虽然这些技术确实是用于灾难恢复的极好的低成本解决方案，但它们无法进行适当的数据分析。最终，人们发现大数据的真正好处在于：对获取的信息进行实时分析和报告。

谬误 2：大数据分析成本太高



当您与完全不了解大数据分析的人探讨时，您会发现许多 IT 领导认为大数据分析的成本太高昂。这可能是因为大数据首先是在最大的企业中流行起来的，例如 Facebook、微软和沃尔玛，这导致许多人认为只有大型企业才能使用该技术。早期确实是这样的，但是现在，基于云的大数据和数据分析解决方案能够帮助中小型公司分析大数据，并根据需要增加规模，而且启动成本相对较低。

谬误 3：我们已远远落后于大数据曲线



如果你认为已经没有机会使用大数据分析技术来获得竞争优势了,那我有一些消息给你。大数据将长期成为企业 IT 的一部分。如果想绕过大数据,等待下一个大技术浪潮,那你要等很长时间了。事实是,数据分析的下一个创新只不过是许多已经实现的大数据基础的演变。话虽如此,由于过去几年出现的工具和平台的进步,你还是有时间迎头赶上的。在人工智能和自动化领域尤其如此,这可以大大减少大数据分析所花费的时间和精力。

谬误 4：在开始大数据项目之前，确定要问的问题



在开始任何大数据项目之前,确定大致的目标是一个好主意,但没必要确定具体的问题。通常,分析可以帮助您回答问题或解决您从未考虑过的问题。如果您太专注于获得具体的答案,您可能会失去重要的洞察力,而这些可能会比您认为的更加有用。即使这些见解可能是奇怪的,但它们仍然非常有用。

谬误 5: IT 人员不能承担数据分析角色



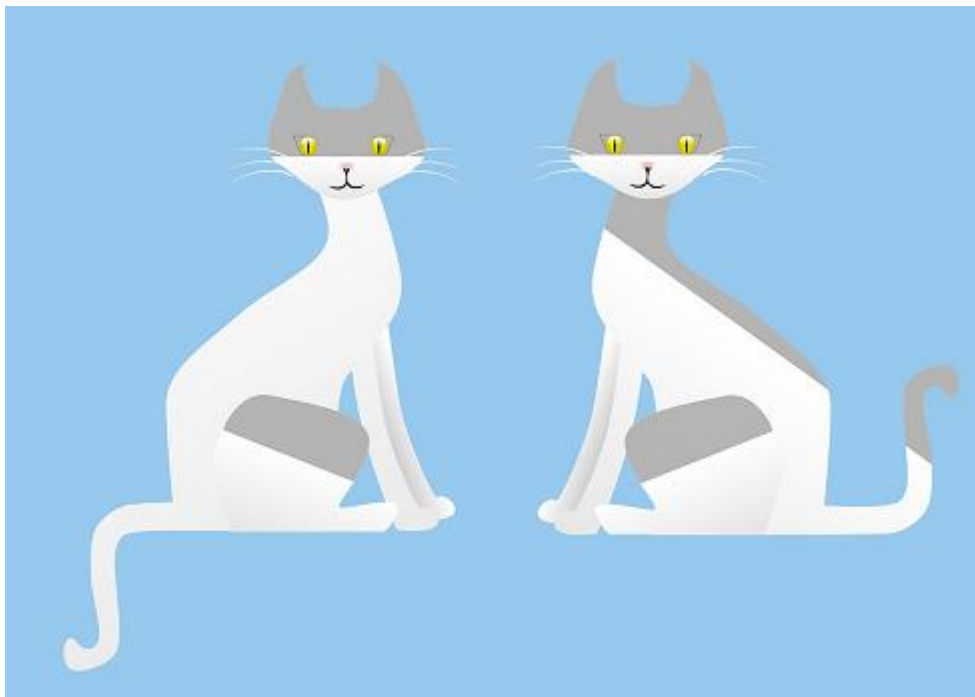
2017 年,围绕大数据项目的最大担忧之一是媒体经常报道的数据科学家短缺问题。这个消息导致许多人认为只有具备数据科学教育背景的人能够担任数据分析师。换句话说,您无法将负责编码或管理虚拟机的 IT 专业人员转变为分析专家。尽管数据分析的学习曲线很陡峭,但是有理由认为,任何有企业 IT 背景的人(也许是内向型)都比其他人更有优势进入数据分析领域。

谬误 6：分析结果只会确认已知内容



正如一句谚语所说：“知之为知之，不知为不知”，数据分析领域也是这样。在许多情况下，企业领导认为他们熟悉企业业务和市场。因此，大数据分析无法给他们提供什么信息。这个问题主要在于打破常规思维模式。事实已经证明，数据驱动型组织能够获得关键的业务洞察力，从而获得真正的竞争优势。

谬误 7：所有分析平台都是一样的



分析平台都是一样的，不同之处在于您寻找什么。这些观点错的离谱，说明人们缺乏对大数据和分析领域的研究。证明这一点的最好方法是：在多个平台上使用相同的数据，并寻找相同的答案。您可能会感到惊讶，结果将会非常不同。您可以使用这些信息找到适合您的数据和业务的平台。

谬误 8：数据湖泊只是数据库的别称



企业数据库已经存在了很长时间。所以，当“数据湖泊”一词出现时，很多人认为这只是企业数据库（EDW）的流行行话。虽然数据湖泊听起来像在营销公司的会议室中制定的术语，但EDW和数据湖泊之间存在明显的差异，主要围绕结构化和非结构化数据存储方法。

结论

在这个快节奏的企业 IT 世界中，能够区分事实与谬误是非常重要的。相信谬误会导致 IT 部门推迟采用能够真正推动企业发展的技术。这就是为什么我们社区致力于识别和打破所有技术（包括大数据）的谬误。