

 新信息时代商业经济与管理译丛

WILEY

大数据制胜

数据驱动型企业成为行业
领导者和创新者之路

[美] Tho H. Nguyen◎著
王燕爽◎译

LEADERS AND INNOVATORS
HOW DATA-DRIVEN ORGANIZATIONS
ARE WINNING WITH ANALYTICS

大数据+分析方法+技术整合=企业未来之路

 中国工信出版集团

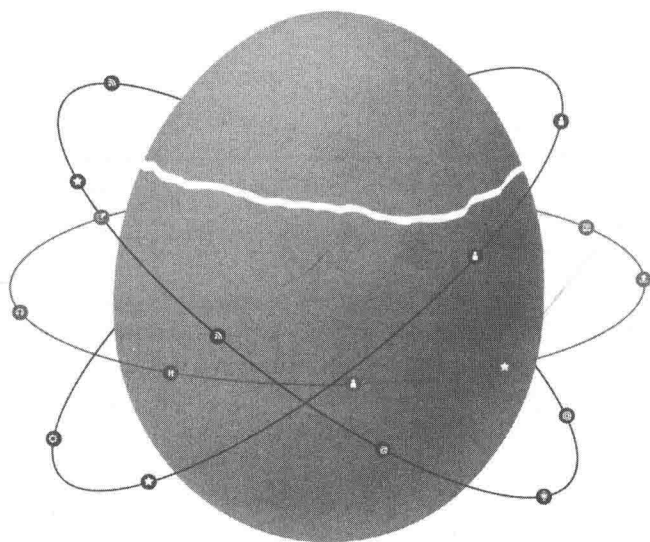
 人民邮电出版社
POSTS & TELECOM PRESS

大数据制胜

数据驱动型企业成为行业
领导者和创新者之路

[美] Tho H. Nguyen◎著

王燕爽◎译



人民邮电出版社

北京

图书在版编目 (C I P) 数据

大数据制胜：数据驱动型企业成为行业领导者和创新者之路 / (美) 阮友寿 (Tho H. Nguyen) 著；王燕爽译。—北京：人民邮电出版社，2017. 10

(新信息时代商业经济与管理译丛)

ISBN 978-7-115-46583-2

I. ①大… II. ①阮… ②王… III. ①企业管理—数据管理—研究 IV. ①F272.7

中国版本图书馆CIP数据核字(2017)第228036号

版权声明

Tho H. Nguyen

Leaders and Innovators: How Data-Driven Organizations Are Winning with Analytics

Copyright ©2016 by John Wiley & Sons, Inc. All rights reserved.

This translation published under license.

Authorized translation from the English language edition published by Wiley Publishing, Inc..

Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

本书中文简体字版由 John Wiley & Sons Ltd 公司授权人民邮电出版社出版，专有版权属于人民邮电出版社。本书封底贴有 Wiley 防伪标签，无标签者不得销售。

-
- ◆ 著 [美] Tho H. Nguyen
 - 译 王燕爽
 - 责任编辑 李 强
 - 责任印制 彭志环

 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 大厂聚鑫印刷有限责任公司印刷

 - ◆ 开本：700×1000 1/16
 - 印张：12.25 2017 年 10 月第 1 版
 - 字数：200 千字 2017 年 10 月河北第 1 次印刷
-
- 著作权合同登记号 图字：01-2016-8289 号

定价：59.00 元

读者服务热线：(010) 81055488 印装质量热线：(010) 81055316

反盗版热线：(010) 81055315

广告经营许可证：京东工商广登字 20170147 号

献词

谨以此书献给 Anh、Ana 和我的家人，他们为此书的出版献出了无数个宝贵的夜晚和周末。

致谢

首先，我想向这本书的读者表示感谢，感谢他们对本书所表现出的兴趣，感谢他们为成为行业内的领导者和创新者所做的努力。我很荣幸能够投身到利用技术和科学来改变世界这项崇高的事业中来，科学技术不仅可以帮助我们消除饥饿，还能推动教育的发展，促进社会变革。

在本书的创作过程中，我得到了很多人真诚的帮助。我要感谢我的同事们，以及来自世界各地的客户，如果没有他们的支持和指引，这本书将永远不会问世。我要特别感谢天睿公司和赛仕软件公司的朋友，谢谢他们鼓励我创作本书，帮助我考证数据方面的专业术语，并努力让本书简单易懂，确保即使是非专业的读者也能轻松阅读。

其次，我还要感谢那些逐章逐节、逐字逐句地审阅本书的人，在此我要特别感谢 Shelley Sessoms、鲍勃·梅西和保罗·西格尔。他们试图通过一页页满是专业术语的文字揣摩我的思想，并将其重新表达出来，这是一项艰巨的任务。这项工作他们完成得相当漂亮，也很高效，感谢他们的耐心。他们真诚的反馈，让我有机会对这些文字进行调整，使其更加严谨准确。

另外，我要衷心感谢这些年我接触过的营销人员、IT 人员和商业人士。谢谢他们对我的接纳，帮助我学习，是他们赋予了我撰写本书的想法和观点。最后，我要感谢我的家人（Nguyen 一家和 Dang 一家），还有圣方济各圣公会的资助和扶轮社（琼斯一家和维尔一家），如果没有他们，就没有我今天所取得的成就。感谢我的妻子和女儿，她们是我此生挚爱，是我的生命之光。

作者简介

Tho Nguyen 1980 年来到美国，和他一同来到美国的还有他的父母、六位姐妹和一位兄弟。Tho 随后受北卡罗来纳州格林斯博罗市圣方济各圣公会的资助，并获得来自美国家庭的帮助和支持，在他们的指导下学会了英语，从此他开启了在美国的精彩生活。

Tho 取得了北卡罗来纳州立大学计算机工程学院的理学学士学位，并获得了英国布里斯托大学国际贸易学院的工商管理学硕士学位。在硕士就读期间，Tho 先后访问过法国国立桥路大学、香港大学和加州大学伯克利分校。Tho 还曾担任过扶轮社的学者大使，这一经历加深了他对世界的认识。

在信息科技领域从业 18 年以来，Tho 同世界范围内的科技合作伙伴、研发人员和客户建立了紧密的合作关系，在数据分析、数据存储和数据管理方案方面，为客户提供增值的商业方案。得益于其在技术和商业领域的工作背景，Tho 获得了在产品管理、全球营销以及商业联盟和战略管理方面的广泛经验。Tho 目前在国际数据分析研究所担任教职员，活跃于各种学术会议，同时撰写数据管理和分析方面的博客。

业余时间，Tho 喜欢和家人待在一起，喜欢旅行、跑步和打网球。此外，他还是一位不折不扣的美食家，希望尝遍世界各地的美食。

序

我从2001年就开始从事高级分析（Advanced Analytics）方面的相关工作了，这些年见证了高级分析的不断发展和成熟。最初的时候，只有银行借助预测性分析（Predictive Analytics）进行风险管理，而现在，所有行业的公司都在使用预测性分析。借助高级分析来管理客户旅程（Customer Journey）现在也成为一种主流，尽管需要分析的数据量在以爆炸式的方式增长，高级分析发展和部署所需的时间依然在不断缩短，由最初的几个月缩短到现在的几秒钟。领域中的领头企业都把数据分析看作是公司的一项核心竞争力，而不只是一个单点解决方案（Point Solution），领域革新者把数据当作是未来创新的源泉。如何向数据驱动型（Data-driven）企业转变，并借助数据分析取得成功，已经成为每个企业管理者的必修课。要实现企业数据驱动型的转变，达到分析创新的目的，就需要制订一个切实可行的实施计划，很多相关书籍可以为计划的制订提供重要指导，其中就包括 Tho 所写的系列书籍。

Tho 和我相识于几年前，我们当时同在国际数据分析研究所（International Institute for Analytics）担任教职员工，并合作了教授分析课程。我们对分析技术和方法都有着浓厚的兴趣。数据分析的应用对各个行业不同规模的企业的发展起到了推动作用，同时，为企业解决相关的数据需求提供了帮助。

所有的企业和机构都拥有数据，而且在我们所生活的这个时代，这些企业和机构还会产生和公布越来越多的数据。而且数字化渠道和设备越来越多，也会产生更多的数据，其中包括客户信息、合伙人信息、供应商信息，甚至是我们的设备信息。政府和第三方所公布的数据越来越多，市场和各种应用程序界面（API）也在为我们提供更多的数据。相应地，我们存储和分析文本、音频、图片和视频数据的能力也随之增强。这些不断增加的数据让我们的数据基础设施达到了上限，但同时也推动了很多新技术的应用，包括数据库分析技术（In-database analytics）、内存分析技术（In-memory analytics）以及 Hadoop 技术（开发和运行处理大规模数据的软件平台）。但是仅仅具备存储和管理数据的能力是远远不够的，要获得成功，我们还需要借助数据来更好地决策。这也就意味着，我们需要理解数据，分析数据，并对获得的数据见解进行部署，即我们

需要具备充分利用数据的能力。这些新技术要想发挥其价值，就必须与端对端数据和数据生命周期分析相融合。

多年来，我接触到的借助数据分析来改善公司决策制定的企业多达数百家，我本人也给上千人培训过，帮助他们获得应用分析技术所需要掌握的相关知识和技能。根据我的个人经验，如果企业能够利用数据推断出相关事项的可行性，在做决策的时候，尤其是有关企业日常运营的相关决策方面，企业就能够保证思想统一，从而做出更加稳妥缜密的决策。以数据为基础的决策制定方法在企业革新方面可谓是首屈一指，然而要让分析性决策发挥作用，为企业带来创新，就需要坚持数据生命周期分析的一贯方法，坚持有效应用数据管理和分析技术。

在本书中，Tho 为我们介绍了要实现以数据为驱动的分析创新所需要的数据生命周期分析和相关技术。Tho 首先为我们介绍了数据生命周期分析、从数据探索到数据准备的过程、分析模式的发展，以及数据在企业决策中的部署，包括如何借助分析将数据转化为具有战略意义的见解。这些介绍为后面章节奠定了基础，Tho 在之后的章节里重点介绍了企业管理和使用数据所需要的重要相关技术。每项技术都在数据生命周期的恰当位置进行了详细分析介绍，同时又辅之以真实的案例做支撑。

在把数据库处理以及高级分析同数据库或数据仓库相融合之后，企业就可以对数据进行就地分析了。数据就地分析不仅省去了数据搬运所需的时间和成本，而且可以在有限的工作时间内分析更多的数据。其结果就是，在提高了分析准确性的同时，也减少了分析所需的时间。

而内存分析技术则可以快速应对复杂的分析问题，从而减少数据分析所需的时间。分析速度加快后，就可以实现更多的迭代次数，获得更多理解数据的途径和提高有益见解的可能性。另外，分析速度提高以后，还可以分析处于快速变化中的流动数据，无须花费时间等待数据先被存储后再进行分析。

Hadoop 大数据生态系统也可以比以前收集和管理更多的数据（包括结构化数据和半结构化数据）。企业之前会把那些认为价值过低的数据处理掉或者存放起来，现在他们可以以较低的成本来存储这些数据了。Hadoop 与传统的数据存储技术相结合，可以帮助企业以更加灵活和广泛的方式来管理数据。

在最后描述协作数据架构和有效生命周期的时候，作者又把这些新分析方

法与部分传统的分析技术结合起来进行整体回顾。本书最后一章探讨云计算、网络安全、物联网（IoT）、认知计算以及“一切皆服务”的业务模式转变对数据和分析所产生的影响。

如果你也是上述领域的从业者，或者是 IT 技术人员，且正在学习如何利用数据来为企业带来创新，成为行业内的领先者，那么你首先需要了解数据管理和分析过程的相关内容，才能利用数据获得成功。本书可以为你提供数据管理和分析的概况，通过一些最佳实践案例介绍，向你展示在现实生活中，企业是如何借助数据取得成功的。

詹姆斯·泰勒

詹姆斯·泰勒是 Decision Management Solutions 公司的 CEO 兼首席顾问，同时也是国际数据分析研究所（International Institute for Analytics）的教职员工。詹姆斯是 *Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics* (IBM Press, 2012) 一书的作者。另外，他还与 Neil Raden 和 Tom Debevoise 分别合著了 *Smart(Enough) Systems* (Prentice Hall, 2007) 和 *The Microguide to Process and Decision Modeling in BPMN/DMN*。詹姆斯是一位在世界范围内都比较活跃的顾问、教育家、演讲家和作家。詹姆斯现居美国加利福尼亚州的帕洛阿尔托市。

前言

我从1998年开始从事数据管理和分析方面的相关工作，自我进入到该行业工作以来，数据管理和分析已经发生了翻天覆地的变化。首先以拍字节（Petabytes）为存储单位的数据量正在以超乎我们想象的速度急剧增加。其次，我们现在收集的数据种类也比以前有所增加，主要是结构化数据和半结构化数据。最后，我们获取数据的速度也在快速增加，对数据更新频率的要求已经提高到了每天甚至是每小时。希腊哲学家赫拉克利特在几个世纪前就曾说过这样一句至理名言：“唯一不变的就是改变。”

为什么要阅读本书

自“大数据”时代兴起以来，数据管理、处理和分析的方式已经发生了重大改变。但不管我们管理和分析数据的方式如何变化，数据最终的目的都是要给决策的制定者提供所需的相关信息。数据的这一终极目的无疑给我们带来了众多挑战，但同时也给我们创造了多种机遇，让我们可以借此去探索和应用新的革新性技术，从而为企业管理和分析数据提供支持。现在就让我们开启本书的阅读之旅吧。

我出席过很多会议，有幸与来自企业和IT行业的众多专业人士分享我的经验，同他们交流如何利用相关技术来更加有效地管理数据，从而创建更加合理的数据生命周期。我从客户那里掌握了他们在数据分析方面所遇到的各种问题，也了解到他们是如何借助灵活的数据分析技术来推动企业创新，从而让企业具备更加有利的竞争优势。以下是最常见的主要问题。

■ “我应该如何把数据管理和分析过程整合到一个统一的环境，从而让整个过程更加高效呢？”

■ “为分析准备所需的数据需要花费好几天甚至几周的时间，我根本没有那么多时间。”

■ “我的分析过程需要几天甚至几周的时间才能完成，等分析过程结束了，信息也过时了。”

■ “我的员工在战术性数据管理任务上花费太多时间了，以致没有足够

的时间用于战略性分析探索。”

■ “有什么办法可以解决员工因为工作没有挑战性而离职的问题呢？”

■ “我的数据太分散了，到处都是。进行数据分析的时候，我应该参照哪些最新数据呢？”

我的一位编辑朋友联系到我，希望我根据这几年在各种演示和白皮书中所表达的观点，撰写一本书，介绍如何利用我所提到的观点来获取成功。我们花费了几个月的时间，整理摘要、大纲，组织章节，最终决定出版此书，书中每个章节都会介绍相关的企业成功案例。我编写本书的主要目的如下。

■ 向读者介绍：在一个有凝聚力的环境中，我们可以借助哪些创新性技术，帮助企业将数据管理和数据分析整合在一起。

■ 在面对重大数据挑战的时候，那些走在最前端的企业借助和使用了哪些技术。

■ 介绍各个领域内的企业案例和成功经验，包括零售、银行、电信、电子商务和交通。

不管是企业管理者还是 IT 人员，想必你都会对现实世界中的最佳实践和使用案例感兴趣，并对书中的这些实践和案例加以应用。书中提到的最佳实践都为企业管理者提供了以数据为驱动的意见，帮助他们更快地进行决策。

能够有机会编写此书，我感到很荣幸。尽管我很期待通过本书，与其他 IT 人员和企业管理者分享我的经验和企业成功案例，但同时也存在以下两点顾虑。

■ 本书出版的时候，书中所谈论到的技术是否还依然具有创新性，是否能够继续为企业管理者提供帮助？

■ 我应该如何为企业创新者和领导者提供他们需要的价值？

“客户互动”（Customer Interactions）对我来说很重要，也是我工作中的一个重点。我跟世界各地的客户交流过，努力去了解他们企业的问题，并给他们推荐合适的技术和恰当的解决方案。我也曾到世界各地去走访，向客户和潜在客户介绍市场上最新的技术，同他们分享那些走在领域前沿的企业是如何利用这些技术来增强自身竞争力，成为统一环境下管理数据和应用数据分析的领导企业的。在介绍本书的详细内容之前，我想先介绍一下本书的主题和书中提到术语的定义，同时介绍一下在竞争激烈的全球经济环境中，需要借助前瞻技

术才能维持自身发展的特殊行业的发展趋势。本书的两个主题是：数据管理和分析，以及如何将两者整合在一起来促进最佳操作、业务发展和企业管理。

相关定义

“数据管理”（data management）一词已经出现很久了，多年来，这一术语也被翻译出了多个流行版本。不过，为了方便说明，我这里选择了“数据管理”，这也是本书的基础所在。根据我的个人定义，数据管理是指数据获取、整合和存储的过程。数据管理通常与 ETL（extraction、transformation 和 load，即提取、转换和加载）过程相关联，为数据库准备数据。ETL 过程已经深深地融入数据管理的大环境之中。ETL 过程的终极目标是要为数据分析者提供可靠、及时的数据。

“分析”（analytics）的释义也有很多版本，而且我们对“分析”的重视程度还在不断增加。“分析”一词开始流行于 20 世纪 90 年代，当时已经有很多行业内的企业意识到了分析的重要价值，并开始通过分析数据来解读过去、现在和未来的发展趋势。广义上的“分析”已经成为各种不同行业内通用的一个术语。根据美国咨询公司高德纳的解释，“分析”是指通过对分散的、整合的统计和数学数据的描述，来解读过去、现在和未来的发展趋势。¹借助分析探索数据的意义，这已经成为关联 IT 和企业的重要桥梁。结合我从客户方面了解到的情况，我把“分析”定义为：通过对大量数据进行分析来获得对企业业务的了解，并以此帮助制定以数据为驱动的决策，从而达到增长业务或改善企业结构的目的的过程。

行业发展趋势和面对的挑战

前面介绍了“数据管理”和“分析”的定义，接下来我们了解一下 IT 行业的发展现状，以及面临的相关挑战。

■ 数据本身就是一种资产和核心竞争力：《福布斯》上的一项研究也认同“数据本身就是一种资产和核心竞争力”的说法。²就 IT 整个行业来说，我们拥有的数据很丰富，但是我们的知识却很匮乏，因为企业根本无法充分解读他们所收集到的数据。对于我们所收集到或者可以获取的数据，我们的分析解读

也都只停留在表面。此外，对数据分析能力的要求也日趋复杂，很多企业都不具备分析数据的有效基础设施和工具。随着数据量的不断增加，我们必须具备管理和分析大数据所需的基础条件。

■ 分析即为一切：对数据进行实时分析，可以帮助公司上至 CEO 下至工厂操作工等各级员工制定以数据为驱动的决策。根据 TechRepublic 近期的一项研究，70% 的受访者借助某种形式的分析来更好地工作和决策。不管是创建新的业务分支还是增设新的产品线，决策的正确与否对企业都有着深远影响，决定着企业的成败。由于现在的业务更加个性化，目标性和开放性也更强，企业的决策也就需要更加具体和透明，并且要基于既定事实（以数据为驱动）。也就是说，企业的所有决策都必须遵循相应的法律法规，并能够体现出恰当的风险管理。

■ “当下”的因素：企业成功所必须具备的“X 因素（未知因素）”似乎已经明朗了，那就是要具备为消费者尽快提供他们所需产品和服务的能力。例如，零售业目前就面临着“当下”这一因素的挑战。单靠绝对低廉的价格和优质的服务已经无法成功吸引到消费者了。企业必须能够在消费者需要的时候，立即满足他们的要求，包括产品的颜色和尺寸等，这才是吸引和维系客户的关键所在。如果企业能够提供满足消费者当下要求的产品，他们是很乐意出高价购买的。《福布斯》一项针对零售业的调查显示，58% 的受访者表示他们更看重产品的即时性而不是价格，且有 92% 的人表示，如果企业不能立即提供他们所需的产品，他们是不愿意花时间等待产品上架的。企业不仅要想方设法超越竞争者，还要和消费者分享产品和服务到位的信息。

上述挑战和机遇已经成为所有行业所面临的问题了。根据我的客户情况，以下 3 方面是他们所面临的巨大挑战。

■ 数据库性能：如果数据库架构达不到数据量所需的要求，就无法充分处理所有的数据集，也不能有效完成对数据的挖掘、分析和可视化。

■ 分析能力：由于获取数据和准备数据所需的时间过长，分析人员往往把精力都放在了数据获取上，从而忽略了对数据进行战略和战术性分析的任务。此外，很多企业还无法根据经济的变化，快速地创建和处理复杂的分析模式。

■ 数据质量和整合：当下数据的种类、独立的数据集和本地化数据的数量过多，导致我们无法准确地了解数据的具体数量和类别。一旦数据的位置和 /

或管理方式发生改变，数据的质量就不再可靠。而一旦数据质量遭到了质疑，分析结果就变得不再确定。

数据是所有企业的战略资产，可以为企业的运营性和战略性决策提供信息依据。但是鉴于我们从网站、社交媒体、移动设备和传感器等设备上收集数据的种类越来越多，收集数据的速度也越来越快，数据量正在以指数方式快速增长。在不到 24 个月的时间里，我接触过的所有客户，他们所拥有的数据量都已经增加了一倍，这已经远远超过戈登·摩尔在 50 年前年做出的摩尔定律的速度了（根据摩尔定律，集成电路上可容纳的晶体管数量每 24 个月就会增加一倍）。随着信息领域变化速度的不断加快，企业也在寻找最新的技术，来满足他们在 IT 和业务发展方面的需求，帮助他们把所遇到的挑战转化为可以促进企业发展，增加利润的机遇。而我深信，当前出现的新技术，如数据库内处理技术、内存分析技术和正在兴起的 Hadoop 技术，都可以帮助企业应对来自大数据管理方面的挑战，帮助他们借助数据分析来发掘新的发展机遇，通过把数据管理和综合分析相结合，提高投资回报率。

谁需要阅读本书

本书可以帮助商业人士和 IT 专家学习数据分析方面的创新技术，帮助他们了解他们的同行是如何在工作中取得成功的；帮助商业分析人员为企业的不同部门提供更加精确睿智的分析见解；帮助数据科学家探索数据分析的新途径；还可以帮助企业经理、主管和管理者借助数据分析，来制定以数据为驱动的决策，从而促进业务的发展，增加企业收入。

推荐以下从业人员阅读本书。

- 执行经理，包括首席执行官（CEO）、首席运营官（CFO）、首席战略官（CSO）、首席营销官（CMO）以及其他希望借助创新性分析技术，提高企业效率，实现战略目标的公司管理层领导。

- 见证了新技术的发展并希望将其引入到公司的企业管理人员。

- 希望把新理念和新技术介绍给企业的销售经理和会计主管。

- 为领导团队决策制定提供分析数据和信息的业务分析人员、项目经理等业务人士。

- 管理数据，确保数据质量和整合、提供分析数据的 IT 专业人员。

如果你希望借鉴他人的成功经验，改善公司的数据管理和分析过程，并通过直接应用数据分析来探索新机遇，那么本书就是你的首选。

如何阅读本书

读者可以按照章节的先后顺序，从前往后阅读。本书的布局按照儿童学步的正常顺序，从爬行、行走、疾走，再到奔跑，依次排列。不过，如果读者对数据库内处理技术、内存分析技术和 Hadoop 技术的概念已经有所了解，也可以跳过前面的章节，直接阅读读者最为关心的章节。如果读者对上述概念都不甚熟悉，那么我还是建议读者从第 1 章开始阅读，因为第 1 章重点介绍了数据分析的生命周期和数据转变为有益信息和见解的历程。紧接着在第 2~4 章（爬行、行走和疾走阶段）中，读者可以了解如何把具体的技术应用到数据分析中去。第 5 章（跑步阶段）则把所有的因素都整合到了一起，同时帮助读者了解各项技术是如何帮助企业处理大数据和高级分析的。第 6 章讨论了数据管理和分析领域最受关注的 5 个话题，同时分析了未来技术的发展。每个章节的主要内容和关注重点见表 1。

开启本书的阅读之旅

企业最重要的资产是客户，而仅次于客户的宝贵资产就是数据，企业可以利用数据来吸引、维系客户，并同客户进行互动，从而达到增加收入的目的。我所接触过的所有企业都拥有大量的数据，这些数据就如同是从四面八方汇聚入海，不仅渠道多样，而且来源广泛。数据无处不在，只要眼睛能看到的地方就存在数据！你的数据库时时刻刻都在接收着汹涌而至的数据。现在就让我们了解一下，如何以一种更加有效的方式分析这些数据，从而帮助企业制定以数据为驱动的决策。

尾注

1. Gartner, IT 词汇：“分析”。
2. Forbes Insight, 《投资大数据》(*Betting on Big Data*) (Jersey City, NJ: Forbes Insights, 2015)。

表 1 章节介绍

章节	介绍	问题思考
第1章 数据分析生命周期	本章主要介绍数据的一般生命周期以及将数据转化为战略见解所涉及的分析阶段（数据探索、数据准备、模型开发和模型部署）	<ul style="list-style-type: none"> ■ 数据分析的生命周期是什么？ ■ 每个生命周期阶段的特征是什么？ ■ 每个阶段所使用的最佳技术是什么？
第2章 数据库处理技术	本章主要为读者介绍数据库处理的概念。数据库处理是指将高级分析和数据库或数据仓库相整合的过程。在融合了数据库或数据仓库之后，分析处理就可以在无须数据复制或搬运的前提下对数据进行就地和并行处理了	<ul style="list-style-type: none"> ■ 什么是数据库处理？ ■ 为什么选择数据库处理？ ■ 哪些分析过程应该利用数据库？ ■ 有哪些最佳实践？ ■ 有哪些使用案例和成功案例？ ■ 使用数据库分析技术有哪些优势？
第3章 内存分析技术	本章主要为读者介绍内存分析的概念。内存分析技术可以借助内存分析引擎快速处理复杂的分析问题，是一项处理大数据的全新方法	<ul style="list-style-type: none"> ■ 什么是内存分析？ ■ 为什么选择内存分析？ ■ 哪些分析过程应该利用内存分析？ ■ 有哪些最佳实践？ ■ 有哪些使用案例和成功案例？ ■ 使用数据库分析技术有哪些优势？
第4章 Hadoop技术	本章主要介绍Hadoop技术的价值。企业现在收集的数据量越来越大，包括结构化数据和半结构化数据，这就给企业带来了前所未有的大数据挑战。企业因而也更加需要采用积极灵活的策略来管理和整合大数据	<ul style="list-style-type: none"> ■ 什么是Hadoop技术？ ■ 为什么要在大数据环境下选择Hadoop技术？ ■ Hadoop技术在现代数据架构中是如何运作的？ ■ 有哪些最佳实践？ ■ 有哪些使用案例和成功案例？ ■ 使用Hadoop技术有哪些优势？
第5章 技术整合	本章主要对前面第2~4章介绍的技术和概念进行总结。介绍如何通过把传统方法和现代新方法结合起来为企业节约时间和资金	<ul style="list-style-type: none"> ■ 数据库处理技术、内存分析技术和Hadoop技术是怎么互补的？ ■ 有哪些使用案例和成功案例？ ■ 整合后的数据管理和分析架构具有哪些优势？
第6章 结语	本章对本书进行总结，介绍端到端数据管理和分析平台在帮助企业制定以数据为驱动的决策方面所发挥的重要作用。本章还对分析技术的未来发展进行了思考	<ul style="list-style-type: none"> ■ 数据管理的未来发展趋势如何？ ■ 分析技术的未来是怎样的？ ■ 数据管理和分析领域的5个主要关注点是什么？

第1章 数据分析生命周期	//001
阶段 1: 数据探索	//003
阶段 2: 数据准备	//004
阶段 3: 模型开发	//005
阶段 4: 模型部署	//006
端到端过程	//008
第2章 数据库处理技术	//009
背景介绍	//010
传统数据分析方法	//011
数据库处理方法	//013
为什么选择数据库处理技术	//014
成功案例	//015
数据库内数据质量	//029
对数据库处理技术进行投资	//037
尾注	//039
第3章 内存分析技术	//041
背景介绍	//042
传统方法	//043
内存分析方法	//044
使用内存分析技术的必要性	//047