

# 企业的大数据战略

作者：[荷] 马克·冯·里吉门纳姆（Mark van Rijmenam）

## 版权信息

本书纸版由浙江人民出版社于2017年6月出版

作者授权湛庐文化（Cheers Publishing）作中国大陆（地区）电子版发行（限简体中文）

版权归作者所有，看完即删除，勿传播

书名：企业的大数据战略

著者：[荷] 马克·冯·里吉门纳姆（Mark van Rijmenam）

译者：盛杨燕

电子书定价：37.99美元

Think Bigger: Developing a Successful Big Data Strategy for Your Business.

Copyright © 2014 Mark van Rijmenam

## 目录

[版权信息](#)

[中文版序](#)

[每个企业都需要一个大数战略](#)

[01 每个人都身处一个大数时代](#)

[20世纪，大数据的序幕](#)

[21世纪，数据大爆发](#)

[02 你必须知道的大数据7大特征和8大事实](#)

[大数据的“7V”特征](#)

[你早该知道的大数据的8个事实](#)

[大数据对社会的影响](#)

[03 大数据的7大发展趋势](#)

[趋势1，移动大数据](#)

[趋势2，实时大数据](#)

[趋势3，物联网](#)

[趋势4，量化自我](#)

[趋势5，社交大数据](#)

[趋势6，公共大数据](#)

[趋势7，游戏化](#)

[04 你应该了解的大数据技术](#)

[Hadoop HDFS和Map Reduce](#)

[开源工具](#)

[大数据工具和分析类型](#)

[05 如何布局企业大数据战略](#)

[以信息为中心的企业的主要特征](#)

[大数据的8大用途](#)

[大数据和投资回报](#)

[资产负债表中的大数据](#)

[企业内部大数据的3大影响力](#)

[大数据企业落地路线图](#)

[7种最重要的大数据员工](#)

[中小型企业的大数据机会](#)

[管理大数据](#)

[06 大数据落地的18个行业](#)

[农业](#)

[汽车业](#)

[消费品业](#)

[教育](#)

[能源](#)

[金融服务业](#)

[游戏业](#)

[医疗](#)

[法律](#)

[制造业](#)

[非营利组织](#)

[传媒和娱乐产业](#)

[石油和天然气产业](#)

[公共部门](#)

[零售业](#)

[电信业](#)

[交通运输业](#)

[旅游与休闲业](#)

[07 大数据的隐私、道德和安全](#)

[大数据隐私](#)

[大数据道德](#)

[大数据安全](#)

[08 大数据的未来](#)

[商业分析的未来](#)

[迈入波字节时代](#)

# 中文版序

## 每个企业都需要一个大数据战略

对不同的企业和行业来说，大数据的意义各有不同。那么，大数据能够为你的企业做什么呢？这取决于你的企业的类型，你的企业所拥有的数据量，你的企业所处的行业，以及一系列其他变量。我每次给企业提供大数据战略相关的建议时，常会面临的一个主要问题就是：大数据带来了如此多的可能性，如何找出对的用例以发展到概念验证阶段？

### 5大信号，拥抱大数据战略

对中国企业来说，相比要不要使用大数据这个问题，更需要考虑的问题应该是何时开始使用大数据。当然，如果你现在是一家富有市场竞争力的企业，而且不想在接下来的五到十年内退出历史舞台的话，那你的回答应该是毫不犹豫的：马上就开始使用大数据！不过，你的企业何时准备好开始拥抱大数据？你的企业何时能具备成功开展和实施大数据战略的必要条件？你的企业何时准备好迎接大数据带来的挑战？在我担任大数据策略分析师的过程中，我发现，一旦企业内出现了如下5个信号，就说明做好了拥抱大数据的准备。

#### 企业内部已达成共识，认为企业内部存在大数据

关于大数据能为你的企业带来什么，如果企业内部并没有达成共识，那么所谓的开始使用大数据将是毫无意义的。就像上文已经提及过的，大数据对不同的个人、企业和行业来说，意义各有不同，因此大数据能带来众多的可能性。对于大数据能为企业带来什么，如果每个员工都有各自不一样的见解，那么要实现所有员工团结一致或者成功实施大数据战略是很难的，甚至是起码的启动概念验证都很难实现。

#### 商业智能是决策流程的一部分

如果你的企业文化中根本没有大数据驱动基因，那么要拥抱大数据就会很困难。所以，你的企业应该至少运用了某种形式的商业智能，以监督企业的表现。你应该已经建立了能持续监控的KPI指标体系，并对关键流程进行定期分析。也就是说，你已经在你的企业内开始使用一些可用的数据，而且企业内部有一定的数据驱动的文化，要求管理者在决策制定过程中使用商业智能报告。管理者已经对使用的数据有信任感，并且因此能比完全依靠直觉时做出更好的决策。

#### 鼓励实验和创新

如果你想要开始使用大数据，首先要做的就是推进概念验证，以了解大数据究竟是什么，以及大数据究竟能为你的企业带来什么。大数据项目通常需要花费一年半的时间，而且在项目推进的过程中，很难看出结果会如何，但是这不能成为对大数据望而却步的理由。因此，企业必须鼓励实验和创新，从而发现大数据为企业带来益处的方式。如果你没有将一些时间花在未知事务上，就没有做好拥抱大数据的准备。

#### 使用高质量的数据

数据质量差是很多企业拒绝使用大数据的原因之一，因为低质量的数据会让你，尤其是数据分析师的工作变得很难。从以下几方面出发，可以判定数据的质量水平：

◎标准化：对数据元素的定义和理解是否一致

◎完整性：是否所有必需的数据都存在

◎精确性：数据是否确切地代表现实情况或者来源是否可靠

◎有效性：数据的价值是否处于可接受的范围内

◎唯一性：数据是否多次出现

要拥有高质量的数据，首先你的企业应该有正确的主数据管理（Master Data Management），而这正是大数据分析的基础。如果你的数据管理不当，质量不能保证，那么你的数据非但不会成为你的宝贵财富，反而会变成一项危险的累赘。基于低质量的数据所做的任何分析都是毫无意义的。因此，高质量的数据是一个大数据项目取得成功的先决条件。

董事会清楚大数据的价值

正如上文已经提到的，大数据项目的落地需要一段比较长的时间，而且开展大数据项目成本比较大，但是项目结果在初期并不能清楚地预见。因此，如果你想让大数据项目获得成功，取得董事会管理层的支持是至关重要的，因为如若不然，大数据项目很可能会被中途叫停。另外，大数据归根结底还是人的问题，因此文化氛围的改变也很重要，而要改变文化氛围，那就必须要做从上至下的改变。因此，如果你的董事会清楚大数据的价值，你取得大数据成功的胜算就会大大增大。

制订和实施大数据战略并不容易，也并非所有的企业都做好了迎接大数据的准备。不过，如果你的企业已经具备了上述五个方面的条件，那么，也就是时候拥抱大数据了，最终你应该也会取得成功。在没有任何数据背景的前提下开展大数据项目基本上就是做无用功。因此，你最好是先让你的企业适应大数据这种新的工作方式。这可能是一个漫长的过程，但是你终将因此有所收获。

大数据用例框架

为了帮助企业更好地了解大数据不同的可能性，我开发了一个大数据用例框架。该框架将9类大数据用例划分为三个核心：

◎你的客户

◎你的产品

◎你的企业

每一个核心包含三种可以明确定义的大数据用例，这和各行各业的所有企业都密切相关。这个框架如表0-1所示，接下来就让我们逐个分析大数据的用例。

0-1 大数据用例框架

□

360度客户视图

对任何一个企业来说，形成一个对客户的完整视图都是非常重要的，因为这能够帮助企业了解客户的渴望、需求、偏好以及客户习惯的营销方式。当你将多个数据源的数据结合起来，就能得到一个360度的视图。你的内部来源数据，如客户关系管理（Customer Relationship Management，简称CRM）数据、销售数据或呼叫中心数据，可以与外部来源数据，如社交媒体数据或新闻数据相结合。零售巨头沃尔玛就是创建360度客户视图的最佳例子。多亏了在线营销平台，沃尔玛才能够实现360度客户视图的创建。

了解市场

以前，企业如果想更好地了解市场，一般会寻求市场调研企业的帮助。虽然消费者固定样组、焦点小组和问卷调查都能提供市场反应数据，但不幸的是，这些调研方式费时费力，且只能提供对过去的见解而非对未来的见解。有了大数据，我们就没有必要再采用如此费时费力的方法了，因为当你将不同的数据集，如销售数据、市场新闻数据和社交媒体数据结合起来，你就能够实时了解市场对你的产品的评价了。另外，举个例子，当你发布新广告的时候，你也能通过结合多个数据集了解到市场对它的看法。大

数据将市场调研提升到了一个新的高度。

### 寻找新市场

对具有多个数据源的数据，如网站统计和社交媒体数据进行分析，能够帮助你寻找到新市场或你尚未意识到的有潜在需求的客户。使用自然语言处理（Natural Language Processing）或机器学习（Machine-Learning）等技术，你将能更好地预测（潜在）客户正在寻求什么，而模式分析则有可能帮助你寻找到全新的市场。

### 个性化的网站/产品

大数据就是相关性，并通过正确的渠道，在正确的时间，以适当的价格，为特定客户提供正确的产品或服务。谷歌会根据你的个人资料将其搜索结果个性化，亚马逊则基本上会针对不同的用户显示不同的产品推荐主页。最终，它们会通过结合来自不同数据源的数据实现对客户的完全了解，真正清楚客户寻找的是什么样的产品或服务。通过为客户提供个性化的产品推荐而获得成功的企业比比皆是，如洲际酒店集团（Inter Continental Hotel Group）和Spotify音乐服务平台。

### 提升服务

大数据可以帮助你大幅提升服务水平。你可以通过使用深度数据分析来优化客户服务，从而提升客户满意度。美国西南航空公司（Southwest Airlines）的所作所为就是一个很好的例子。它们使用语音分析，实时提炼出现场录制的公司员工与客户之间的对话背后的深层含义和有价值的信息。然后，通过将这些数据与其他来源的数据，如客户档案数据、航班数据以及社交媒体数据相结合，来为客户提供始终如一的高品质服务。

同时，智慧城市能够利用大数据来更好地管理城市，提高市民服务水平。韩国智慧城市松岛（Songdo）就是一个很好的例子，在这里，甚至连垃圾都要受到分析，以提升垃圾清运服务的质量。

### 共同创造和创新

大数据不仅可以为你提供你的客户的信息，也可以为你提供你的产品以及产品使用情况的信息。如果你能够通过传感器和远程信息处理设备监控产品的使用情况，你就能够对如何完善产品有一个深入的了解。此外，使用海量数据以及超级计算机进行的仿真分析能够帮助你极大地提高产品创新的速度。因此，宝洁曾使用仿真分析在数秒内创建成千上万次迭代，以找到最好的一次性尿布。

### 降低风险或减少欺诈行为

利用大数据可以轻松地检测到异常情况和异常值，而这些异常情况和异常值则有可能在提示着欺诈行为存在。万事达卡就是利用大数据来确定某项支付行为是否属于欺诈行为。此外，大数据也可以降低企业所面临的风险。当你对客户有了更好的了解之后，你就能更好地确定其风险状况，无论这个客户是一个人，还是一家正在寻求信贷、抵押贷款或保险服务的企业。英国汽车保险商Insurethebox就是通过让客户自愿接受保险公司监控其驾驶习惯来降低保险风险的。Insurethebox的定价策略是，客户的驾驶习惯越好，保险费用越低。当然，这也降低了保险公司的赔付风险。

### 更好地管理企业

员工在办公室里会产生大量的数据。在办公室的设备上以及整个办公室里安装的传感器，能够为你提供有关员工工作表现的信息。这些信息能够帮助你更好地管理办公场所。例如，卡毕斯特制药公司（Cubist Pharmaceuticals）使用数据后发现，公司办公场所内的咖啡机过多。然后，该公司通过减少咖啡机的数量，并设置集中的咖啡饮用场所，增加了员工之间的非刻意沟通。

同时，你还可以监控所有的非结构化数据，如邮件、文档和会议，以了解哪个员工熟悉哪个话题，以及哪些员工之间有互动。这不应被视为对员工的监视，相反这会帮助员工更快速、更有效地找到他们所

需的信息。

#### 了解你所面临的竞争

当然，你能够为你的企业所做的事情，你的竞争对手也或多或少地能够为他的企业做。不过，如果你对竞争对手的定价策略进行监控，那么大数据就能实时让你了解竞争对手的定价变化，以便你更迅速地做出反应。当然，这种策略并非万无一失，因为当两个算法开始交互的时候，会有一些奇怪的事情发生，例如，可能会显示一本关于苍蝇的书售价高达2 300万美元。

这九个大数据用例只体现出了大数据能做的事情的冰山一角。当然，大数据的具体用途会根据企业自身和所处行业的不同而有所不同，但是这个大数据用例框架应该还是能够为你提供一些如何开始利用大数据等相关指导意见的。

# 01 每个人都身处一个大数据时代

□

在人类史册记载的所有数据中，90%是在近两年创造的。然而，利用和解析大数据的需求却由来已久。事实上，早在7 000年前就曾出现过利用数据追踪和控制生产的情况——这也是最早的相关案例。当时，美索不达米亚人利用基础的会计准则记录农作物的生长情况。之后，会计准则被不断完善。1663年，约翰·格朗特（John Graunt）记录并观察了有关伦敦地区居民死亡情况的所有信息，并为了研究肆虐欧洲的黑死病，建立了警示系统。在首次有记载的统计学数据分析案例中，他将自己的研究成果汇集到了《关于死亡表的自然观察与政治观察》（Natural and Political Observations Made upon the Bills of Mortality）一书中。此书对揭示17世纪人们的死因有巨大的启示作用，格朗特因此被誉为“人口统计学之父”。

19世纪见证了信息时代的开端。1887年，现代数据首次被采集。当时，赫尔曼·霍尔瑞斯（Herman Hollerith）发明了一种计算机，可以通过穿孔卡片来管理人口普查数据。

## 20世纪，大数据的序幕

1937年，富兰克林·罗斯福当政期间，美国推行了第一个大型数据项目——在新颁布的《社会保障法》（Social Security Act）的规定下，政府要记录300多万个用人单位和2 600多万员工的财政贡献。IBM有幸得到了这个项目，为这个庞大的统计工程开发穿孔卡片读卡机。

1943年，英国人开发出了首台数据处理机器“巨人”（Colossus）计算机，用以在第二次世界大战期间破译纳粹密码。它能以每秒5 000字的速度破解截获情报中暗含的信息，极大地缩短了破译纳粹密码所需的时间，是一个巨大的进步。

1952年，美国国家安全局成立。短短10年间，它就签下了12 000多名密码专家。“冷战”期间，这些专家开始采用自动收集和处理想情报信号的方式，信息超载的难题摆在了他们面前。

1965年，美国政府决定设立首个数据中心，来存储收集到的超过7.42亿张纳税申报单和1.75亿枚指纹。数据中心的员工将这些记录全部转移到计算机磁带上，集中存放在一个地方。不过后来，迫于他国反对，该项目没有继续推进，但是它揭开了电子数据存储时代的序幕。

1989年，英国计算机科学家蒂姆·伯纳斯-李（Tim Berners-Lee）发明了万维网（World Wide Web），希望借助一个“超文本”环境促进信息共享。当时，他肯定没有料到，这项发明会对全人类产生这么重大的影响。

从20世纪90年代开始，随着越来越多的设备连上互联网，数据以一种前所未有的速度被创造出来。1995年，第一台超级计算机诞生。过去，一个人用计算器需要花费3万年才能完成的工作，现在，计算机能在1秒钟之内解决。

## 21世纪，数据大爆发

2005年，奥莱利出版社（O'Reilly Media）的罗格·马古拉斯（Roger Mougals）创造了“大数据”一词（此前一年，该出版公司曾提出“Web 2.0”一词），用来指代使用传统商业智能工具难以实现分析、管理的大量数据的集合。

同年，雅虎在谷歌Map Reduce的基础上创造了Hadoop，目的在于为整个万维网编写索引。如今，世界上许多企业都在使用Hadoop这个开源大数据平台来处理大规模数据集。

随着越来越多的社交媒体的涌现以及Web 2.0的蓬勃发展，每天都有越来越多的数据诞生。创新型企业开始慢慢地挖掘这些海量数据，政府也开始推进大数据项目。2009年，印度政府开始对全国12亿人口进行虹膜扫描以及指纹、照片采集，然后将这些数据存储在一个全球最大的生物识别数据库中。

2010年，谷歌CEO埃里克·施密特（Eric Schmidt）在加利福尼亚州太浩湖（Lake Tahoe）举行的科技经济论坛上这样描述信息革命：“现在，我们每两天创造出来的数据量，相当于从文明伊始到2003年创造的

所有数据量的总和……差不多有5艾字节（Exa Byte）<sup>[1]</sup>……”

2011年，麦肯锡发布了一篇备受好评的报告，题为《大数据：创新、竞争和生产力的下一个前沿》（Big Data: The Next Frontier for Innovation, Competition, and Productivity）。该报告得出结论称，到2018年为止，美国将面临14万名数据科学家和150万名数据管理员的巨大缺口，数据科学家也被认为是21世纪最令人向往的职业。

在过去几年间，大数据初创公司如雨后春笋般涌现，都在试图帮助企业理解和利用此次大数据爆炸。就像1993年人们开始慢慢接受互联网一样，如今越来越多的公司开始采用大数据，未来我们势必会见证更多改变。

事实上，数据量爆炸式地增长让我们跨过了十进制系统时代。如今，美国的国家安全局和联邦调查局在计算其数据存储量的时候，使用的单位是尧字节（Yotta Byte）<sup>[2]</sup>。在不久的将来，我们将采用波字节（Bronto Byte）<sup>[3]</sup>作为传感器数据的单位。因此，人们已经创造了新术语来形容在未来几年内有望被创造出的数据量（详见图1-1）。

□

图1-1 波字节资讯图

大数据将彻底改变全球企业和人类社会，而当前可用的全球数据量有望每两年实现翻番。那么，大数据到底是什么？接下来，让我们走进大数据，一探究竟。

## 02 你必须知道的大数据7大特征和8大事实

□

欢迎访问：电子书学习和下载网站 (<https://www.shgis.cn>)

文档名称：《企业的大数据战略》[荷] 马克·冯·里吉门纳姆 (Mark van Rijmenam) 著.epub

请登录 <https://shgis.cn/post/864.html> 下载完整文档。

手机端请扫码查看：

