

## 在支离破碎的大数据分析世界中发掘价值 信息中介如何创建新的数据生态系统

作者

Bill Gerhardt

Kate Griffin

Roland Klemann

2012年6月



思科网络解决方案部 (IBSG)

## 在支离破碎的大数据分析世界中发掘价值 信息中介如何创建新的数据生态系统

### 数据就是“新石油”

早在 2006 年，市场研究员 Clive Humby 就提出了数据就是“新石油”这一观点。<sup>1</sup>如今，随着数据不断演变并进入日益扩大的大数据领域，越来越多的专家团体开始接受这一观点。与 19 世纪的构想家曾在深埋地下的天然黑色软泥中所感觉到的创造财富的力量一样，这些专家在大数据中也发现了同样的变革。

正如一个多世纪前石油曾点燃人们的梦想一样，如今数据已成为经济和技术创新愿景的推动力。如果可以对“未经加工的”数据进行提取、提炼并输送到可影响实时决策的地方，该类数据的价值将会猛增。通过分析 — 揭示复杂的行为、模式及所发生的事件（如果以前未曾出现）- 发掘数据中蕴含的重要见解后，数据将实现其固有的力量。此外，如果可以在整个生态系统中合理共享数据，且能在分析最能发挥作用的地方访问数据，那么数据将会成为真正的游戏规则变革者，它会改变我们生活、工作、娱乐、学习、购买、销售及旅行的方式 — 无一例外。

但大数据也会带来巨大的挑战。目前的问题在于日益增加的海量信息 — 从千吉字节到拍字节再到百亿亿字节 — 大有在未经筛选、非结构化、未经处理和看似不可收拾的信息井喷中使人们穷于应付之势。就目前情况来看，数据生态系统非常地支离破碎。数据创建者和数据潜在价值提取者之间存在着一个充满复杂性、差异性和误传性的迷宫。如果想要让分析成为未来新的“提炼厂”，则需要通过更大的连接性、信任和效率来解决那种支离破碎的场面。

谁能够担起这个责任，谁又将担起这个责任？Cisco® 网络解决方案部 (IBSG) 相信，这项包罗万象的任务将由一个新兴阶层来完成，即数据信息中介。这些实体将参与进来，并将数据创建者与数据受益者联系到一起，同时授权给数据创建者，并推进数据价值链上的每一个环节。在这一过程中，他们会为所有参与者创造机遇，以此来协调和实现大数据的巨大潜力并将其货币化。

这一过程的关键要素在于大数据的传输和存储基础设施。那些高瞻远瞩的参与者能够跨网络和数据中心收集和控制动态数据，再将不同的大数据参与者联合为顺利运作的整体，从而让自己处于优势地位。

本文将研究大数据的一些广泛影响、其变革承诺以及尝试驾驭大数据过程中的一些难点。最重要的是，我们将探索数据信息中介在激发大数据作为真正变革力量 — 当代“新石油”的潜力这一过程中的机会。

<sup>1</sup>[http://ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](http://ana.blogs.com/maestros/2006/11/data_is_the_new.html)

## 大数据的本质

为了寻找到更明智的决策方法，许多组织都在运行数据仓库并采用传统的数据分析方法，其目的是减少客户流失、增强宣传活动有效性以及反欺诈（此处仅举几个实例）。大数据代表从传统数据分析向前迈进的革命性一步，其特征就是它的三个主要要素：多样性、数量和速度。

- 数据的多样性来自两个方面：结构化和非结构化。结构化数据能进入已标记的数据仓库，可轻松分类。但是，当今绝大多数数据都是非结构化的，主要由 Facebook、Twitter 及视频内容等来源提供。这类数据具有随机性，难以分析，且数量惊人。
- 数量庞大的大数据淹没了正常的数据库。例如，据 Facebook 报道，其用户每天会发表 27 亿个赞和评论。对于许多人来说，如此庞大的数据规模简直令人生畏：他们无法跟上数据的发展步伐，就别提对数据进行分类和分析以及从中提取价值了。
- 对于许多参与者而言，如果所有数据都以远远超出其处理能力的速度泛滥，那么对这些数据的管理将面临巨大的挑战性。为了让大数据成为游戏规则变革者，其分析速度需要与信息进入数据库的较快速度相匹配。对于某个特定数据位是否值得捕获以及它与其他数据结合使用时是否具有关联性的问题，人们必须以微秒为单位作出决策。例如，数千张图像中一张熟悉的罪犯的面容会触发“停止”操作；某种信用欺诈模式可能会引发警告；而表明客户流失越来越多的迹象可能会赋予商家提供优惠券的灵感。无论何时，只要数据中出现重要的异常情况，都需要有人在存储数据之前指出该异常情况，以便采取实时行动。

## 由网络支持

就其本质而言，网络流量就是大数据。仅在移动网络这一部分网络中，全球就有 60 亿移动用户，每天会发送 100 亿条短信。<sup>2</sup> 据全球市场情报公司 International Data Corp. 预测，2011 年，全球生成的数据总量高达 1.6 万亿千兆字节。

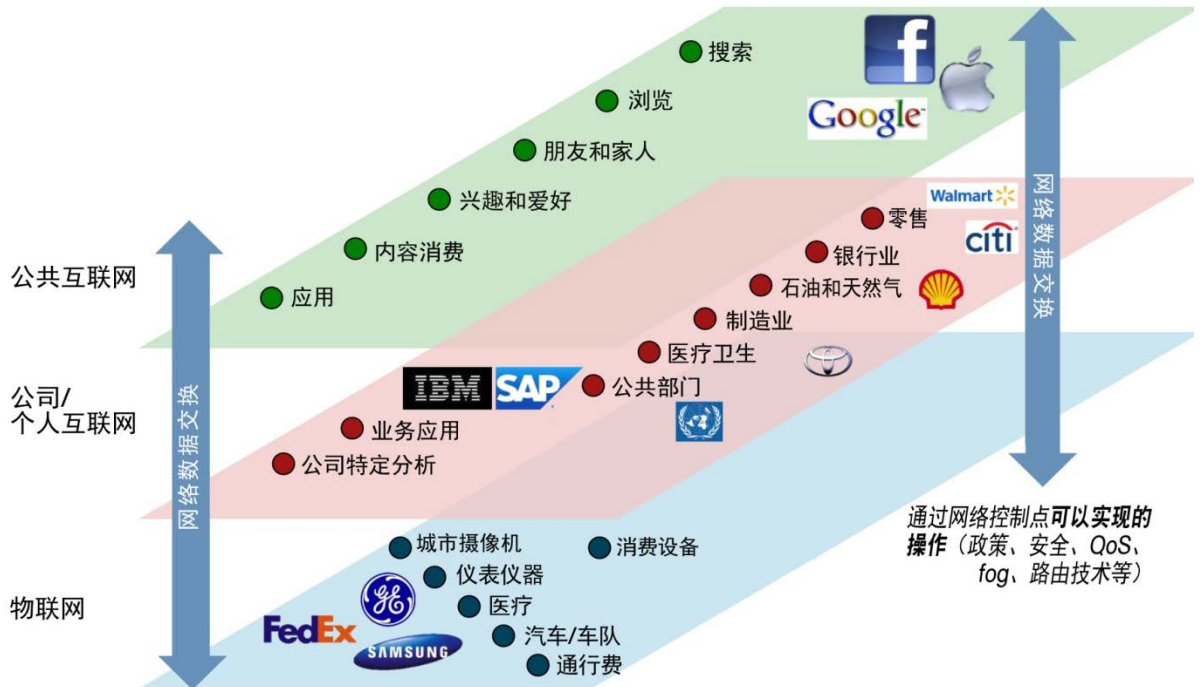
整个网络的数据量还会继续暴增。到 2020 年，将有 500 亿台设备连接到网络和互联网。<sup>3</sup> 但是网络绝不只是仅传输这些海量数据的“无管道”。除支持庞大的数据量以外，网络在增加大数据对于企业的潜力方面还扮演着两个有价值的角色（请参见图 1）。

首先，网络能够以较快速度收集数据并提供环境信息。它往往能够在“网络时间”内捕捉来自设备、专用网和公共广域网的数据。此时便可以在事件发生时以实时或近乎实时的方式提取价值并采取行动。通过网络收集的数据，网络系统可以确定环境信息。该环境信息可包含匹配位置、身份或数据的在线状态信息。此外，已收集数据的实际内容也具有价值。了解的内容不仅包括人们浏览过的网站，而且还包括他们在网站上的操作，这样就可以提供重要的见解。

<sup>2</sup> 国际电讯联盟，2012 年。

<sup>3</sup> “物联网：互联网的下一步发展如何改变一切”，思科 IBSG，2011 年。

图 1. 网络与数据域进行连接。



通过网络可以看到上下文 (位置、身份、存在等)

资料来源: Cisco IBSG, 2012 年

其次，通过网络，企业可以立即根据自身对数据的见解采取行动。通过一系列实时控制点，网络可以到达客户或影响客户体验。这些控制点可在政策、隐私、偏好、安全、服务质量以及重选路线或本地处理方面带来变化。

拥有网络环境信息及控制方面的信息之后，那些掌握了网络端到端优势的组织便能获得广泛而深刻的视野。

### 大数据如何改变生活

如果大数据的潜力能够得到充分实现，那么便会对广泛的复杂和普通日常任务及活动带来革命性的影响（请参见图 2）。

堵车就是其中的一个例子。它们似乎是一种不可避免的灾难，这已是既成事实。但想象一下，如果堵车情况大幅减少乃至消失，将会怎样？人们能减小的压力和节省的时间（当然，这意味着金钱）会相当之大。生活质量会得到提高，因为上下班时间减少了，而与家人呆在一起的时间相应会增多。这个答案并非是要加快汽车的驾驶速度，而是要对交通流量进行实时预测分析，从而在交通堵塞开始之前对交通信号、数字信号和路线选择进行即时更改。

钱包里塞满的零售商和银行纸质收据又能怎么处理呢？试想一下，如果能将它们替换为电子记录，会怎样呢？企业可以通过环境信息和比较信息来丰富这些记录。例如，个人可以通过预算管理 and 健康咨询应用对这些数据进行管理、分享、货币化和利用。

图 2. 大数据的关联性。



资料来源：Cisco IBSG，2012 年

大数据带来的医疗进步就是能够根据患者独有的健康问题制定个性化的治疗方案。我们可以预测、跟踪并减轻传染病疫情或食品污染。目前，Google 流感趋势工具已整合了有关健康信息的搜索项数据，以预测全球范围内的疾病爆发情况。

在全球范围（甚至更广范围）内，大数据能够监控和分析从气候变化、天体物理学及能源消耗到地缘政治和社会经济等各方面的大量信息。不管什么情况，大数据分析都能够在预测未来趋势的同时推动对重要事件背后关键原因的深入理解。

### 让大数据利于商界运作

大数据也可以改变我们与企业互动的方式。消费者每时每刻与周围世界的互动会产生经常被忽略的副产品：大量的个人数据。这些包括天气搜索、价格比较、购买以及成千上万的其他日常选择和行为。通过分析所有此类活动产生的数据，大数据能够提供让消费者和卖家互动方式彻底变革的机会。

某些企业已经开始采取行动了。以下是几个例子：

- Harrah 通过分析客户细分数据，使收入由原来的 8% 提高到了 10%，<sup>4</sup>而 Amazon 曾经报道称，其分析推荐引擎为其带来了 30% 的收入。<sup>5</sup>
- Wal-Mart 的库存管理实施了无线射频识别 (RFID) 技术，可将供应商与其零售环节数据仓库的实时信息联系起来。此过程可将缺货情况减少大约 16%。<sup>6</sup>

<sup>4</sup> Gartner，2005 年。

<sup>5</sup> McKinsey & Company，2011 年。



- 根据贯穿 46,000 多个配送和供应链位置的发货和消费者数据，FedEx 实现了实时可见性。
- Kaiser Permanente 通过整合其电子病历以及住院病人和门诊病人管理系统，将运营成本减少了 40%。整合临床数据和成本数据后，该企业发现了万络的有害影响，随后从市场上撤回了该药物。<sup>7</sup>

因此，大数据已经成为首席信息官们首先要考虑的对象。<sup>8</sup> 但是，思科 IBSG 针对数据价值链设想了一种更高层次的数据分析和互动。我们对许多垂直产业中数百个潜在的大数据应用进行分析，得到了一些使用案例，然后又从中确定了 8 种不同的类别。

1. **收入保障。**整合公司内部数据有助于磨练自身的识别能力，并能在欺诈发生前进行预防。医疗等欺诈密集型行业会受益良多。
2. **风险降低。**网络每天都会为企业、政府和消费者传送大量重要信息，这样便带来了日益增加的入侵和安全攻击风险。跨越更广泛的地域及网络覆盖范围的数据联合有助于识别可疑模式，同时发出需要立即行动的信号。
3. **客户生命周期。**企业可以在客户不满的情况下调整归零，并进行即时响应，从而改进消费者体验并减少客户流失。所有重视客户关系的服务型行业都将从中获益。
4. **市场执行。**大数据可以通过分析带来更佳的市场服务，同时增大交叉销售和追加销售的机会。银行业和互联网商务即是突出的潜在受益者。
5. **产品创新。**消费者的投入对于产品开发而言至关重要，而当下的许多公司都急于进一步了解客户的好恶。整合社交网络源等非公司数据来源能够在消费者的产品感知方面提供更全面的看法，可能有助于在对某种新产品进行设想或设计之前揭示其需求。
6. **业务模式。**新的定价算法将创造新的获利机会，并提升与消费者互动的效率。零售商尤其可以从这种与消费者位置和行为关联的全新定价模式中获益。
7. **运营管理。**大数据几乎可以帮助所有组织更好、更有效地运行。服务供应商可以改进其网络的日常运营。零售商可以创造更有效且利润丰厚的销售点互动。此外，几乎每一条供应链都能更顺畅地运营。总的说来，一个常见的信息结构就能提高流程效率并提供全面的资产情况。
8. **高级广告。**企业对消费者行为了解得越多，就越容易将了解到的信息反映到广告设计中去。eMarketer 发现，基于数据的先进定位能够让广告的效果翻倍。千人成本 (CPM) 可降低 50% 以上。若要通过大数据分析进一步利用这些优势，就需要使用信任机制来减轻隐私约束，并确保具有针对性的广告能与整个网络的消费者信息保持一致。

## 抓住大数据演变的浪潮

这些使用案例的全部潜力将在三次产业演变浪潮中实现（请参见图 3）。每次浪潮中的分析活动都可以在数据平面、控制平面和消费者平面描绘。在数据平面上，我们可以看到所用数据的来源。在控制平面上，我们可以看到数据的使用方式、见解的产生方式，以及出现某些裁断或决定时会发生的事情。在消费者平面上，我们可以看到消费者和数据分析过程之间的关系，同时牢记消费者往往就是数据的创始者，偶尔才会成为数据的拥有者。

<sup>6</sup> “Does RFID Reduce Out-of-Stocks?”，阿肯色大学，2005 年；“Opening Up the Big Box: Measuring the Wal-Mart Effect”，*The Economist*，2006 年 2 月 23 日。

<sup>7</sup> McKinsey & Company，2011 年。

<sup>8</sup> 一份 2011 年的 IBM 调查显示，83% 的高管都会选择大数据作为其重点课题，而不是移动解决方案、虚拟化和云计算。

图 3. 大数据产业演变。



资料来源：Cisco IBSG，2012 年

- 如今，大多数产业均稳固扎根于第一波浪潮之中。数据大多封闭在组织的特定业务单元内。这些数据主要用于撰写事后报告。事后，组织会对数据进行收集，但是这些报告对所有决策的影响都是间接的。对于消费者来说，他们要么对自己的数据没有控制权，要么根本就没有意识到这些数据的存在。他们主要担心数据的安全性和完整性，同时会有数据可能会被滥用的模糊感觉弥漫心头。
- 在第二波浪潮中，数据平面从封闭状态转为公司内部。数据可以在组织内跨单个孤岛移动，并进入公用存储库中。此后，有更广泛的收集数据集可供分析和共享，可进行更有力的决策制定。并且，在控制平面上，新的见解会催生行动。与针对一个数据变量获得的简单见解不同，多个变量和注意事项会提供可产生变化的提示（例如，库存产品已订购，或已向客户发放优惠券）。从消费者的角度来看，在这个阶段，他们越来越能意识到，向组织授予访问个人数据的权限是有回报的。消费者可以在自愿性质的选择加入选项中做出选择，以确定其数据的使用方式。
- 第三波浪潮会跨整个生态系统收集、共享和处理数据。企业不仅可以利用自身的数据，还可以利用无数的第三方数据，包括天气型态、经济指标、“匿名”销售额和社交媒体流量等。在此阶段，组织会通过日益复杂的流程使用数据，以执行预测分析，同时在事件发生之前感知事件。一旦形成此类见解，组织就可能在无人工干预的情况下采取自动化的实时行动（例如，重新路由流量，重新订购存货）。在第三波浪潮中，消费者深信存在着从个人数据中创造价值的机会。他们设定了数据使用策略，并相信可以在整个生态系统中实施这些策略。整个生态系统的透明度日益增加，这让消费者更加安心，因为他们可以看到数据是如何使用的。许多人都主动加入到这个市场中来，在这里，他们的数据都是具有价值的产品，最终结果就是为消费者带来了实实在在的好处。

这些观点一直在不断蔓延，同时，据思科 IBSG 所见，它们已经产生了一定的影响。在从第一波浪潮向第三波浪潮转变的过程中，回报更大了，但难点也变得越来越尖锐。

### 通往大数据世界之路的难点 — 如何克服这些难点

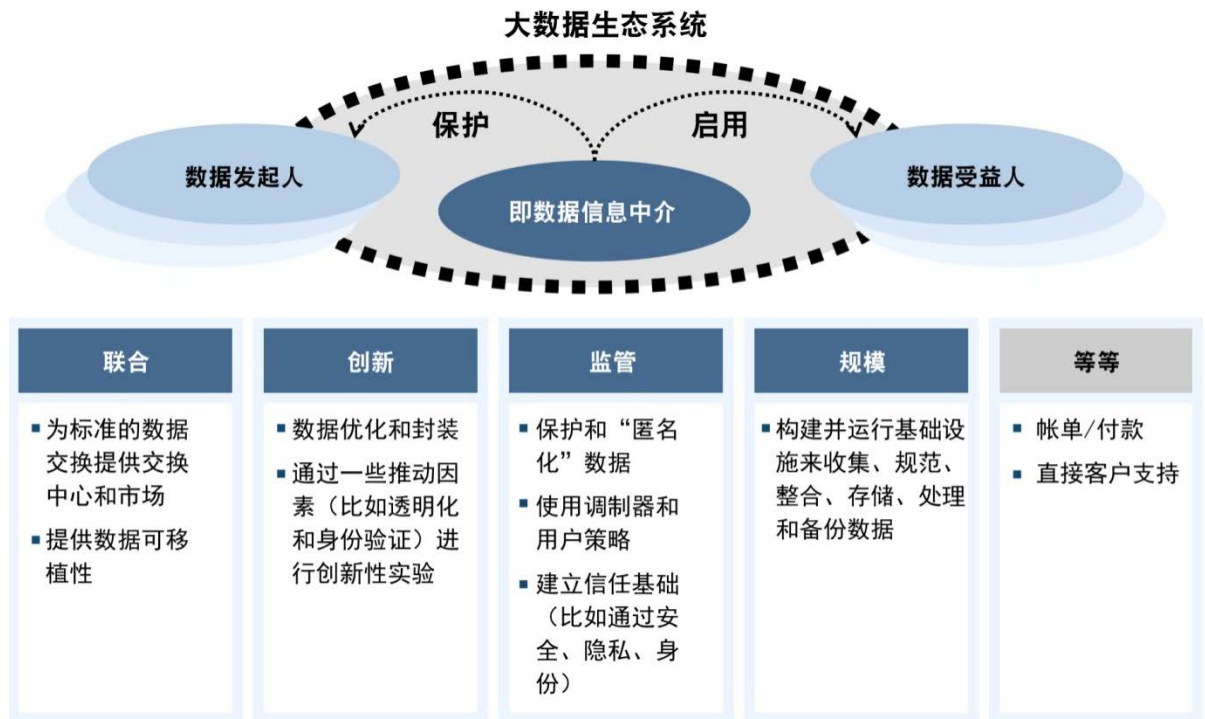
大数据的最终目标是收集数据、存储数据、寻找数据的意义并从数据中提取价值。就目前情况来看，大数据生态系统出现了裂缝。但是还有足够的潜力来协调这个生态系统的方方面面，进而使创始者和受益者带着更高的信任度协调工作。

让我们看看金融交易结算、蜂窝漫游、电子数据交换、自动化道路收费员和电子病历等其他复杂生态系统成功的关键因素。我们的研究表明，复杂的生态系统需要外部实体执行增值功能，才能实现联合、创新、监管和规模。

在某些生态系统中，“中介”的出现实现了这些功能，并能使相应产业蓬勃发展，这些中介包括银行领域的 TransUnion、移动漫游领域的 Syniverse Technologies 以及供应链管理领域的 RosettaNet。相比之下，欧洲的自动化道路收费及美国的电子病历等其他行业一直举步维艰。在这些情况下，没有什么实体愿意或有能力承担信息中介的角色，其结果就是无法取得进展。

进入“数据信息中介领域”。思科 IBSG 相信，在大数据分析生态系统内，如果产业要发挥其极限潜力并实现所有有关大数据的宣传内容，中介或信息中介将必不可少。我们认为数据信息中介必须发挥四种作用（请参见图 4 及下文）：

图 4. 数据信息中介的作用。



资料来源：Cisco IBSG, 2012 年



1. **联合：实现可行的生态系统。**数据需要从创建的地方到达它能发挥价值的地方，同时在整个过程中减轻隐私和安全问题。但现在并不存在可以分享各种数据来源并从中获益的可行市场。尽管如此，创新者们正在发掘变化数据的价值。例如，Target 已针对孕期购物者的消费习惯进行了细致的数据挖掘，因为分娩前后有很大的机会可以建立新的品牌忠诚度。*Forbes* 杂志表示，Target 可能会在合适的情况下及适当的市场内与其他公司共享这个极富价值的信息数据库。<sup>9</sup> 此外，许多保险公司都在通过第三方数据库增加自身的索赔数据，这些数据库包含社交媒体及全行业来源，可进行异常检测，以针对性地识别新的欺诈模式。

但是，在如今的各种情况下，企业仍然需要独立制定使用外部数据的方法。另一方面，数据信息中介能够识别价值、找到来源、制定出合法/商业交易，并对数据格式进行整理和标准化，使其能与内部系统共同发挥作用。数据信息中介能够推动数据在整个生态系统中的共享，同时为行业内和跨行业的数据交换创造交流中心和市场。信息中介还可以大大简化首批期望受益于大数据的公司所面临的繁琐过程，尤其是极少数已大力投资大数据的一流公司之外的小公司。要推动这项工作，数据信息中介可以制定共同标准，并确定谁愿意购买数据以及他们如何从中创造价值。

2. **创新：从一片混乱中获得价值。**对于最终用户而言，在海啸般的数据中获取价值犹如海底捞针。许多组织甚至在用大量数据进行实验方面都困难重重，更不用说从闪烁不定的少量孤立见解中创造价值了。通过承担身份管理和认证的角色，数据信息中介能够让公司更容易进行实验。通过验证之后，数据会便于携带，并可随时随地共享。例如，从消费者角度，存储在 Facebook 上的数据可以转移到 Myspace 或 Amazon，无需每一次都重新创建。
3. **监管：资源的内外部统一。**需要有人将目前这种支离破碎的大数据生态系统整合到一起。在数据通过各种内部和外部中间站的过程中，协调和监管不足阻断了价值形成之路。为了解决这个问题，数据信息中介可以整合内外部的共有及私有资源，从而在共同的云架构下营造和谐氛围。这种方法可以避免数据遭到破坏，同时也能使政府机构确信这些数据得到了正确使用。例如，最近对 Facebook 的审查表明其隐私问题尤为突出。信息中介可以帮助设定业界标准，确保 Facebook 的数据在根据消费者的意愿使用时不具破坏性，而是具有价值性。通过赢得消费者和大数据参与者等的信赖，数据信息中介能提供更多可供分析的数据，同时促进进一步的实时行动。
4. **规模：应对计算需求。**大数据的处理成本令许多组织望而却步，这就需要第三方服务来完成这些困难的工作。如果不具备大规模的内部或外部容量，数据管理就会立即出现瓶颈。此外，即使内部存储容量已扩展为可处理拍字节，但数据分析的并发请求仍可能会超出系统容量极限。数据信息中介在处理庞大的数据量方面起着至关重要的作用。信息中介可提供更具扩展性和经济性的存储方法，并帮助进行最重要的数据处理。如今的 Apple、Amazon 和 Facebook 们都在独立投资这些资源，并且一直在获取实时分析数据（自创建后）的能力。

许多参与者都能担任数据信息中介角色，例如服务提供商、IT 公司、银行、企业、Web 2.0 公司、政府资助的机构或者新的创业公司。如果这些参与者具备了共同克服上述难点的能力，未来的大数据要么会显著成功，要么会彻底失败。

<sup>9</sup> “How Companies Learn Your Secrets”， Charles Duhigg, *The New York Times*, 2012年2月16日；“How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did”， Kashmir Hill, *Forbes*, 2012年2月16日。

## 深入大数据的奇幻梦境：对决策者的行动呼吁

企业领导人们开始想知道从哪里进入大数据领域，以及进入大数据奇幻梦境需要冒多久的险。以下列出了决策者需要考虑的一些选项：

- **在第一波浪潮中进行实验。**组织可以继续有选择性地为单个孤立应用（如定价）收集数据，然后在可能时从中提取价值。
- **向上步入第二波浪潮。**在这里，组织可以选择转移到更广阔的平面，从战术思想转变为战略数据授权。该战略可能依然局限于内部运作，但要作出通过数据分析和主动解决方案差异化服务体验的承诺，并进行相应投入。
- **进行外部思考。**在此阶段，组织会意识到，那些目前处理效率如此之高且如此深刻的大量数据也许只对他人具有价值。因此，生态系统中的一方可以将其位置数据出售给另一方，并获得其服务对应的交易收益率。例如，Wal-Mart 可能会选择将其零售数据出售给 Virgin Media；Amazon 可能将某些多余的云计算能力租赁给那些原本无法访问此类基础设施的中小企业。
- **成为数据信息中介。**某些组织可能会渴望成为行业中最值得信赖的参与者 — 数据信息中介。在此阶段，数据信息中介可以帮助消除行业障碍、连接各个域、提供新服务并提供保护和安全。

## 由此及彼：如何通过大数据获胜

思科 IBSG 勾勒出大数据的未来演变之后，还会定义组织应考虑什么样具战略意义的权衡措施最大限度地发挥其潜在优势。尤其，思科 IBSG 认为，大数据的重要性迅速扩大为潜在的数据信息中介带来了巨大的机会。此外，随着市场逐步接近从关注内部的分析向联合/环境模式转变的关键转折点，现在是时候采取行动了。

即使组织选择保留关注内部的战略，它们仍可以对其数据分析进行变革。但是，许多组织都可以从更关注外部的战略及角度受益，并在整个生态系统中共享数据和见解。其他组织，尤其是网络服务提供商（端对端接入网络）可能会提升为起到数据信息中介的作用，从而在大数据生态系统迷宫中播下和谐、信任、创新和效率的种子。

从思科 IBSG 的角度，这并不是大多数组织是否应该对大数据的未来进行投资的问题，而是他们应该持续投资多久的问题。就像那些曾衡量过发展石油业务的成本和潜在利益的准大亨们一样，今天的决策者们面临的挑战是，要在他们选择的任何平面从即将到来的大数据时代获取最大价值。

这是因为在 19 世纪末，石油的潜在收益规模使它不容忽视。这一点也同样适用于大数据。

## 致谢

作者们在此向 Neeraj Arora、Marc Latouche、Thomas Renger、Rob Rowello 和 Evgenia Ryabchikova 对本文的撰写过程做出的宝贵贡献致谢。

---

### 详情

思科 IBSG（互联网业务解决方案部）拥有业界开拓性思想领先地位，提供 CXO 级别的咨询服务以及创新的解决方案设计和孵化，从而为客户创造市场价值。作为可信任的顾问，思科 IBSG 通过将战略、进程和技术相结合帮助客户进行转型决策，将绝佳的想法转变为实现的价值。

有关 IBSG 的更多信息，请访问 <http://www.cisco.com/ibsg>



美洲总部  
Cisco Systems, Inc.  
加州圣何西

亚太总部  
Cisco Systems (EE. UU.) Pte. Ltd.  
新加坡

欧洲总部  
Cisco Systems International BV Amsterdam.  
荷兰阿姆斯特丹

Cisco 在全球设有 200 多个办事处。思科网站 [www.cisco.com/go/offices](http://www.cisco.com/go/offices) 中列有各办事处的地址、电话和传真。

思科和思科徽标是思科和/或其附属公司在美国和其他国家/地区的注册商标。要查看思科商标的列表，请访问 [www.cisco.com/go/trademarks](http://www.cisco.com/go/trademarks)。  
本文提及的第三方商标均归属其各自所有者。文中的“合作伙伴”用词并不代表思科与任一公司有合作伙伴关系。(1110R)