



白皮书

企业中的大数据：世界巨变

赞助商：Oracle 和 Intel

Dan Vesset
2014 年 2 月

Carl W. Olofson

IDC 观点

我们用于描述核心业务流程的不少词汇与商业本身一样古老。各个企业仍在生产产品和提供服务，争取和留住客户，保护他们的资产和遵守法规。但是这些核心任务以及其他核心任务的数字化使这些企业发生了巨大变化：

- 客户交互的渠道增多，这些交互的数字化无处不在，这些交互产生的各种数据激增，客户期望因新的创新型消费者导向实践而发生了变化。这些趋势正从根本上改变销售、营销和客户服务流程。
- 数字化还扩展到了实物对象 — 汽车、建筑、桥梁、电器、服装和设备 — 它们产生了大量数据流，可用于分析、决策和优化操作。我们将这种现象称为“物联网”，它正从根本上影响生产、物流、资产管理和采购流程。
- 人与物之间交互的数字化也在迅速扩大，这使得大数据及分析需求变得更加复杂。同时，大数据及分析技术的重大进展大幅改善了获取、管理和分析大量多结构化数据的性价比特征。

最终结果是实现了以极快速度生成的各种大量数据的可用性。所谓的“大数据”为从根本上改变业务流程，创建新的业务模式和推动创新提供了巨大机会。但也对大多数组织提出了新的挑战。大数据已成为大业务，但近期对大数据的炒作，不但使人们困惑于选择什么样的技术，而且还产生了不切实际的期望，并且经常提出对于大多数组织而言既不可行也无效的建议。

2013 年，大量的大数据及分析技术及其服务市场首次突破了 1000 亿美元大关。大数据及分析代表了各种信息管理和分析技术，然而最近的焦点主要围绕一些特定的技术，如 Hadoop 和高级分析软件。这些技术在大多数组织的架构中都有重要的地位。然而，仅依赖这些技术只能满足一部分需求。

如今，有越来越多的证据表明大数据的非理性繁荣已经结束。2014 年开启了务实采购和部署各种大数据及分析技术及其服务的长期（漫长）趋势。这种实用主义不仅实现了关系和非关系大数据及分析技术共存的需求，而且达成了这样一个事实：这些技术共同促使启用全新的方式来开展业务、服务和保护公民、照顾病人、教学和管理自然资源。

为充分发挥大数据的价值，组织需要将现有的 IT 架构从经过实践检验的关系技术扩展到非关系技术。只有关系和非关系的数据、技术以及它们支持的分析技术相结合才能推动组织对大型和多样化大数据集进行分析而制定基于事实的决策，以及实施优化对业务事件的响应并具有一致性和合规性的决策管理。

对于 IT 决策者而言，为应对不断发展的业务需求而需要进行选择的技术和制定的决策越来越多，这一趋势似乎势不可挡。然而，结构化评估下列需求可以帮助制定一个长期的大数据及分析战略和一个迭代执行计划，即利用关系和非关系大数据及分析技术组合来赢得竞争优势：

- 实验和绩效管理
- 高管和经理、业务分析师和数据科学家、操作和面向客户的员工
- 战略、运营和战术决策

实体经济中的大数据 — 越来越多的证据证明获得的优势

大数据意味着，企业应采用经济实惠且有效的技术以前所未有的规模来收集、处理、存储和使用多样化的数据。也意味着监视快速移动的动态数据。还意味着组合结构化和非结构化数据，以及在人员、组织和实物对象之间的交互越来越数字化的推动下实现全面的数据准备、分析和分发流程。

但更重要的是，大数据意味着改变。人们需要知道如何提出新问题，使用新数据、新分析和新量度。组织不只是通过数据的三个 V — 海量、速度和多样化 — 的视角了解大数据，还需要将其视为使现有技术和新技术进行组合来拥抱变化和推动创新的契机。

IDC 的 2013 年大数据及分析成熟度基准测试调查表明，46% 的组织在过去 12-24 个月开始使用新的分析技术。同一时期，41% 的组织开始使用新量度或新关键绩效指标 (KPI)，27% 的组织开始使用新的数据类型。作为我们研究的一部分，我们对取得高成就的组织（从大数据及分析项目获得的好处达到或超出预期的组织）和取得低成就的组织（获得的好处未达到预期或根本没获得好处的组织）实现的效果进行了比较。对这两组实现的效果进行的对比结果表明，维持现状会对从大数据及分析解决方案实现期望效果产生负面影响。

这只是一个近期示例，该示例将大数据及分析方面的投资与获得的积极成果关联在一起。还有越来越多基于调查或基于案例的市场调研的其他示例。基于调查的示例包含以下等信息来源：

- 2011 MIT 调查“数字的力量：数据驱动的决策如何影响企业绩效？”表明，数据驱动的组织相比于竞争对手更具盈利能力，生产效率更高。
- 在另一项 2013 年 IDC 和 *Computerworld* 联合进行的调查中，70% 的受访者表示，他们从大数据以及分析项目获得了切实的收益，而其中的 90% 表示，收益的价值达到或超过了他们的预期。

鉴于大数据及分析解决方案的好处，组织应满足哪些要求呢？

- 他们是否应首先关注数据科学家或高管的需求？
- 他们是否应在对试用数据提供支持或由最佳绩效管理实践主导的标准化和控制上过度投资？
- 他们是否应为战略决策提供支持或使战术决策过程实现自动化？

答案是肯定的。在现实中，需要满足所有这些需求，当然无法一蹴而就，但至少要在将所有这些需求考虑在内的大数据及分析战略的指导下来实现。

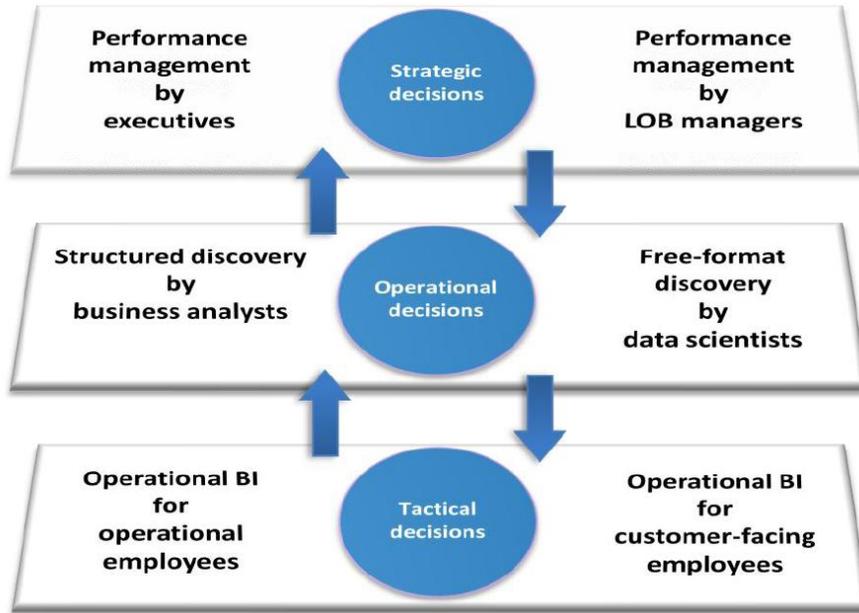
大数据及分析解决方案需求

大多数有关决策的讨论都认为只有高管制定决策，或只有高管的决策非常重要。
这是一个严重的错误。— Peter Drucker

大数据及分析解决方案覆盖范围广泛，影响到组织中从高管到面向客户的员工的各个角落。然而，这种广泛的适用性也意味着用户组的数量、决策过程、分析方法以及其他与数据的交互差别很大。图 1 显示了 IDC 用于识别和评估一个典型组织中各个成员的需求的框架以及大数据及分析对这些需求的影响。该图描述了各个决策过程、多组决策者以及每组中各个成员的分析或信息访问需求类型。该图将需求分为三种，分别规定了更广泛的一组更具体的需求。

图 1

IDC 的决策管理框架



信息来源：IDC，2014

战略决策

这些决策确立了组织的长期发展方向、产品、服务或计划，形成了需要考虑运营决策的指导方针。这些决策由高管和业务线 (LOB) 经理制定，通常通过绩效管理方法和应用得到支持，重点关注基于细分 KPI 和体验的决策。

高管和 LOB 经理通常参与我们所说的绩效管理。这些经理们对他们需要在任意位置按需访问的几个相关 KPI 感兴趣。这些用户通常不自行执行大量数据分析，而是依靠分析师来深入了解审核 KPI 时可能出现的问题。

在大数据及分析环境中，经验、判断和领导的价值没有消失。然而，战略决策过程正被新数据源、新分析、新量度和更强大技术的推出所颠覆。例如，基于更广泛的内部运营数据、外部行业或经济数据、竞争情报和客户舆情快速遍历各个风险变量模型的能力使高管能够对组织的未来制定更好的风险调整决策。可用数据的广度以及处理和数据分析的速度都是前所未有的。

在其他示例中，几个备受关注的案例显示了忽略或未适时访问从社交媒体或电子邮件等非结构化内容源挖掘出来的信息造成的负面影响，这些信息经常作为不良信息的前导指标，之后通过分析历史数据这些不良信息才显露出来。高管可以通过确保其组织的 IT 和分析能力能够满足制定战略决策所需的大数据及分析需求来消除信誉风险、使用数据驱动的模式评估运营风险以及规避监管风险。

运营决策

这些决策侧重于特定的项目或流程。这些决策经常涉及到需要更改这些流程或项目以实施战略目标。这些运营决策形成了用于确定何时应考虑制定战术决策的指导方针。它们受业务分析师、数据科学家和从事结构化或即席数据发现的中层管理者的支持。

分析师大致划分为业务分析师和数据科学家。两者都负责即席分析，但前者通常在多维分析、交互式报告或电子表格的结构化框架中进行分析。后者执行自由格式的即席分析，这种分析通常要求同时具备统计和编程语言专业知识。

在大数据及分析环境中，业务分析师和数据科学家的工作发生了根本性的变化。后者现在可以使用直观、面向消费者的可视化数据挖掘工具。自助服务功能降低或摆脱了分析师访问和操作数据时对 IT 的依赖。数据科学家能够通过使用一系列结构化、非结构化和半结构化数据、结合商业和开源高级分析及开发工具（如，Java）持续体验。数据采样方式将成为历史，运行多小时或夜间模拟的方式也将成为历史，取而代之的是大量不同数据的并行处理。

IDC 采访的一家组织就在结合使用移动设备位置数据、时间和日期信息、最新天气数据、历史客户交易数据、在线浏览行为和客户服务记录来确定为其客户进行产品推荐的下一最佳操作，以及确定何时何地将该产品推荐发送到客户的移动设备。在某种程度上，该公司通过掌握的信息评估客户的情绪——下雨的星期一早上，在移动网络信号较差的地区，不适合向最近对您的某个产品留下差评的客户发送电子邮件。

在其他示例中，零售商已经开始结合使用店内视频分析和库存数据来优化店面布局。甚至还使用了面部识别技术，从而可以分析在零售店中买家对推销的反应，以帮助创建大量数据，数据科学家可使用这些数据更好地进行分析以提供个性化产品。

战术决策

这些决策侧重于交易，即涉及到收入或成本的具体情况。通过识别和修复超出运营决策所制定的指导方针范围之外的交易，战术决策可帮助确保满足运营目标。这些决策受面向客户的员工或操作员工（包括 IT 人员）的支持，在某些情况下完全自动化。

在大数据及分析环境中，战术决策日益呈现出以下特点：将分析师推断出的智能嵌入到业务应用中。一线人员不是分析师，也不应该是分析师，但这些一线人员每天都从实施战术决策自动化技术获得的成果中受益，这些成果有建议、指导方针、规则和脚本等等。当今的技术使组织能够在同一个数据库上运行事务和分析负载，从而减少数据移动量，获得快速的战术决策支持。流处理方法监视高速的事件数据和将业务规则应用于这些数据的补充性规则引擎，并在适当的地点于适当的时间将信息传送给适当的人。在某些情况下，从数据监视到执行操作的完全自动化将改变组织的人力资源需求。

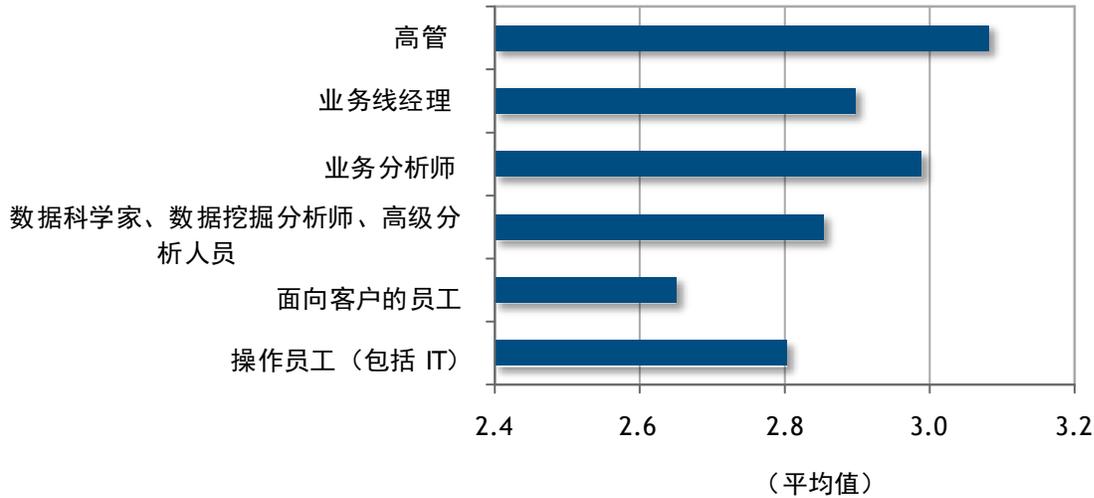
例如，一家农用设备制造商监测它所生产的拖拉机在田间干活的性能和位置，并结合这些流数据以及从土壤样品和天气预报得出的数据，向农民提供战术建议，指导他们如何对拖拉机进行操作和维护。同样，飞机发动机和汽车制造商监测远程信息处理和持续性能数据，以实现预测性维护，从而有助于避免成本高昂的灾难性维修。在另一个示例中，保险公司监视事故索赔并从文本和音频内容中提取信息，然后将这些信息输送到实时欺诈检测应用程序中。如图 1 所示，三种类型的决策不会孤立产生。汽车制造商以及物流公司不应只监测从车辆、集装箱或包裹获得的流数据，他们还应存储这些细粒度事件数据以执行离线纵向分析。换句话说，动态数据用例和静态数据用例越来越相互交织在一起。

尽管在过去的几年里数据科学家在大数据及分析市场中较受重视，但现在从更广泛的角度了解所有决策者是非常重要的。数据科学家在从大数据发掘新见解中发挥着重要作用，但他们的工作成果本身并不是最终目标，而需要与组织内其他人的分析和决策过程共存。如图 2 所示，我们的研究表明，通过组织中的当前大数据及分析解决方案功能，面向客户的员工的决策支持需求最难以满足。公平地说，数据科学家的需求未得到满足的情况要好得多。

图 2

满足最终用户需求

问题：您的组织所拥有的大数据及分析技术满足以下用户组的决策支持需求的程度如何？



n = 330

注意：平均分数是按从 1 分到 5 分进行评分得出的，1 分 = 达到非常有限的程度，5 分 = 达到所需的最大程度。

信息来源：IDC 和 Computerworld 的大数据分析调查，2013

尽管事实是大数据及分析的预期好处之一是改善战术决策，但接受调查的大多数企业都表示，本应在这方面获得最大帮助的那些员工（即，面向客户的一线员工）在组织当前使用大数据及分析时获得的好处却最少。大多数组织在通过将大数据及分析解决方案集成到各级应用中来完全实现大数据及分析带来的好处方面还有很长的路要走。

不要做重复劳动：大数据及分析技术需求

在当今的环境中，不仅要及时获取信息，而且还要能及时分析信息以及据此采取行动，组织需要具备这种能力以便在市场中创造竞争优势，实现社区和自然资源的可持续管理，以及促进交付适当的社交服务、医疗保健服务和教育服务。由于有这么多的机会带来积极的变化以及从大数据获得价值，IT 决策者在从许多技术选择中做出适当决策方面面临着越来越大的压力。

本白皮书中所述的用户组及其需求突出表明需要一系列合适的大数据及分析技术功能以及各种功能的组合和集成。如图 1 所示，即席和结构化发现、绩效管理和在应用程序中嵌入运营智能的企业级大数据及分析需求需要得到一系列关系和非关系技术的支持。如图 3 所示，大数据及分析平台由三个主层组成。

大数据捕获、集成和移动

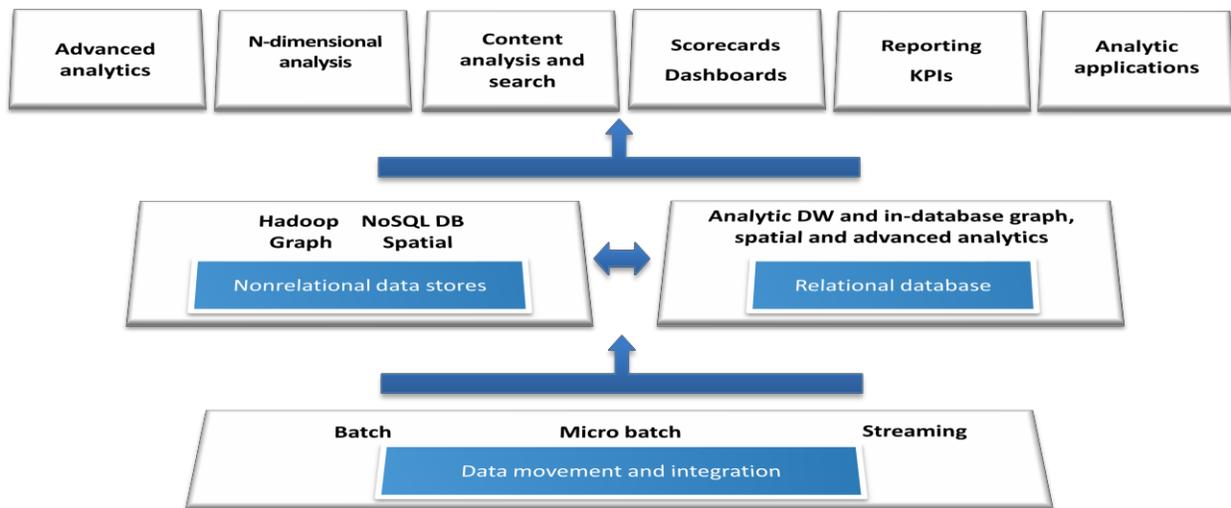
在参考架构中，最低技术层由数据捕获、集成和移动工具构成。其中一些工具提供批处理数据提取、转换、加载，而一些工具使数据流入目标数据存储。大数据及分析需求给这一层的平台带来了巨大的可伸缩性压力。移动多 TB 的数据集或处理数百万个流事件对于许多组织而言并不是一种反常现象。

此外，该软件还用于在各个数据信息库之间移动数据。换句话说，可能需要捕获消费者网络点击流数据，将这些数据移至 Hadoop 集群进行处理，然后将一部分数据移至数据仓库，或可能需要监视来自传感器的流数据，同时捕获事件模式并将其移至数据仓库进行纵向分析。

一方面，大数据及分析平台必须能够提供高度可靠的性能，以及在复杂的多结构化数据转换过程中确保高水平的数据质量。另一方面，该平台必须是动态可伸缩的，以满足涉及使用新数据类型和数据源或现有数据类型和数据源的新组合进行试验的项目的意外需求——特别是来自数据科学家的需求。

图 3

大数据及分析技术平台



信息来源：IDC，2014

大数据管理和处理

大数据及分析平台的中间层是数据管理和处理层，该层由两种技术构成：关系技术和非关系技术。尽管技术（如 Hadoop 和 NoSQL 数据库）已成为最近市场的主要关注点，但关系数据库在大数据及分析平台中的地位仍很牢固，且在可预测的未来仍将保持牢固地位。一方面，关系技术和非关系技术可解决负载的不同“适用场景”。另一方面，它们必须在越来越多的场景中共同运作。例如，组织将关系数据仓库用作“单一数据源”——可信、受控和安全的消息源，来进行绩效管理和结构化即席分析。同时，组织对自由格式的发现使用 Hadoop 数据库或 NoSQL 数据库（通常取决于所需的数据和分析类型）。

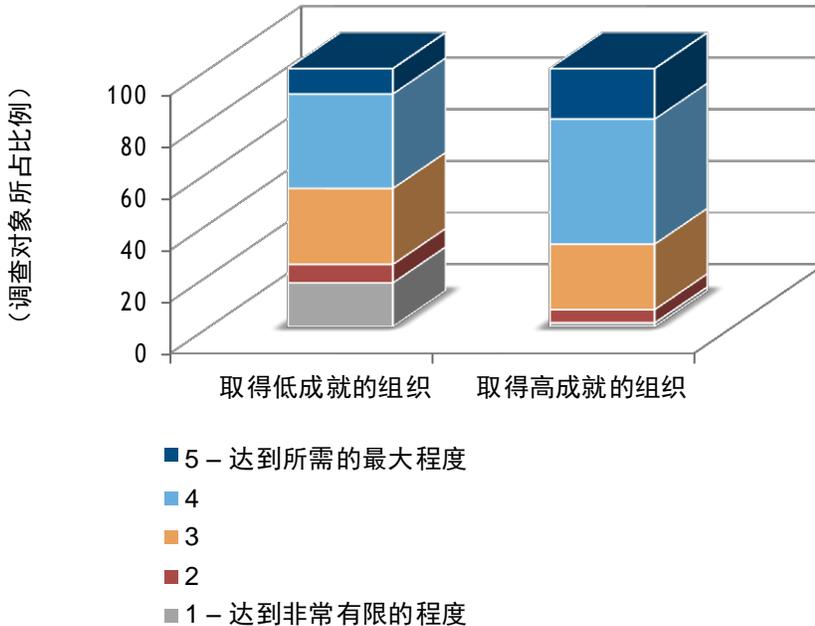
但是，越来越多的实际部署和用例需要同时利用关系技术和非关系技术来移动、处理和分析数据。例如，一家多渠道零售公司利用其非关系数据管理技术对主要由点击流数据构成、同时结合关系数据仓库中的客户参考数据的数据集进行试验。一家保险公司使用音频到文本转换技术，然后使用文本分析，再结合客户的交易数据，从所有的客户交互源获得洞察，从而增强其欺诈检测（预防）模型。一家在线媒体公司使用 Hadoop 存储和预处理网络点击流数据，然后将相关数据子集移至关系数据仓库，业务分析师可从数据仓库中获得这些数据。

最新的 IDC 研究表明，关系技术和非关系技术结合使用时效果更佳。最近 IDC 就以下问题对 701 家大型组织开展了一项调查，图 4 显示了该项调查的结果：您的组织利用 Hadoop、NoSQL 数据库和图形数据库等技术来补充通用 RDBMS 技术的程度有多大。当我们根据大数据及分析项目结果将用户群分为取得高成就的组织 and 取得低成就的组织两组时，我们发现取得高成就的组织更可能拥有结合使用关系技术和非关系技术的大数据及分析平台。在取得高成就的组织中，有 68% 的组织利用其他大数据及分析技术补充通用关系技术达到了很大程度或达到了所需的最大程度。如果从不同的视角来分析同一数据，结果表明，在利用非关系技术补充关系技术达到 *所需的最大程度* 方面，取得高成就组织的数量是取得低成就组织的两倍。

图 4

关系与非关系大数据及分析技术结合使用所产生的影响

问题：您的组织利用任何适用的专用大数据及分析技术（如 Hadoop、NoSQL 数据库、图形数据库、可伸缩的 MPP 数据库）补充通用 RDBMS 的程度如何？



n = 701

注意：对调查对象的评分采用 1 到 5 分制，1 分 = 达到非常有限的程度，5 分 = 达到所需的最大程度。

信息来源：IDC 大数据及分析成熟度基准测试调查，2013

大数据分析和应用程序

大数据及分析技术体系的上层包括各种分析和业务智能工具以及预先打包的分析应用程序。其中一些工具和应用程序已经有些年头了，但也有一些新一代的工具和应用程序。其中一个主要的新功能是可同时访问关系数据和非关系数据这一功能。其他功能则包括以使用者为中心的数据可视化与交互功能的整合，它可以促进分析和 BI 工具的更普遍使用。一些工具帮助组织从客户评论、社交媒体交互、文档和电子邮件等形式的非结构化文本发掘含义。这些功能与事务或运营数据分析相结合，能够支持客户心声分析、保修管理、预测性维护、欺诈检测和预防等一系列新应用。

举一个例子，一家高科技制造商结合使用大数据及分析工具以及实时流数据处理、规则管理、机器学习分析和数据可视化等应用程序，从根本上改善了客户交互以及产品追加销售和交叉销售流程。例如，它

对过去的支持问题和响应进行分析，提炼出发送到客户以响应服务请求的部分。在另一个用例中，新的大数据及分析解决方案基于之前进行的深入分析在交互时为客户服务人员提供追加销售建议。再举一个例子，保险公司使用非结构化文本分析、预测分析以及规则和流数据管理来决定个性化营销活动的最佳内容。通过将实时分析和历史分析相结合，这家公司甄别出了哪些客户最有可能对搞笑性质的消息进行回复，哪些最有可能对特惠商品或其他选项的消息进行回复。

ORACLE 在大数据及分析市场中的位置

IDC 基于软件收入（请参见“*全球业务分析软件 2013 至 2017 年预测以及 2012 年供应商份额*”，IDC #241689，2013 年 6 月）目前将 Oracle 列为世界上最大的业务分析供应商，Oracle 提供一系列广泛的大数据及分析技术，支持静态数据和动态数据用例以及涉及结构化数据和非结构化内容（以及两者同时采用）的用例。如图 5 所示，大数据及分析技术范围极广 — 从关系数据库（包含磁盘和内存优化选项）到 Hadoop 和 NoSQL 数据管理软件。

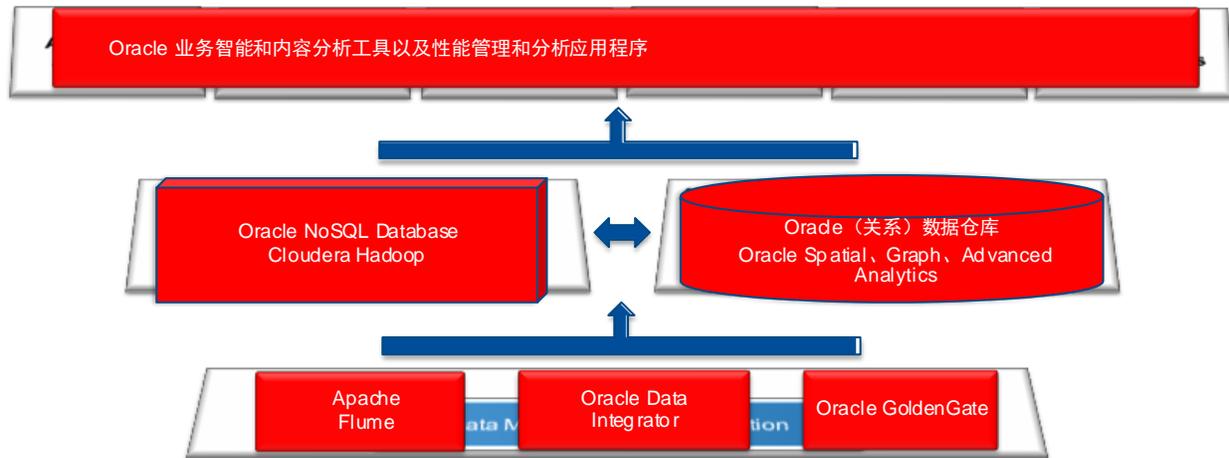
一些技术以纯软件形式的内部部署来提供，一些在云中提供，一些以 Oracle 称其为工程化系统的软件设备的形式提供。这些工程化系统包括 Oracle 大数据机、Oracle Exadata 数据库云服务器、Oracle Exalogic 中间件云服务器和 Oracle Exalytics 商务智能云服务器。除了自己的技术之外，Oracle 还在生产工程化系统方面与 Intel 紧密合作，这些工程化系统采用了 Intel Xeon 处理器 E5 和 E7 系列。Oracle 将这些产品看作一个整体，功能上远远高于各部件之和。Oracle 的大数据及分析解决方案包括以下关键部分：

- **静态大数据：数据储藏库和数据仓库。** Oracle 大数据机（公司将其称为数据储藏库）随 Cloudera Hadoop、Oracle NoSQL 数据库和数据仓库的连接器一起预加载。数据仓储配置中的 Oracle Exadata 非常适用于数据库整合，并能对因数据增长而导致性能落后的仓库进行提速。这两个相互协调工作 — 仓库将数据整合到储藏库以鼓励新发现，储藏库支持使用新数据组合进行试验以发现仓库应提供的新量度。
- **动态大数据：快速捕获和处理流数据。** 对于实时中间件和应用程序，公司推荐使用 Oracle Exalogic。例如，Oracle Event Processing 对从传感器和 Web 交互流入的数据使用基于规则的处理，以快速加载 NoSQL 数据库以及根据 Oracle Coherence 等分布式内存缓存中的实时数据执行数据驱动的操作。一家 Oracle 客户利用该技术有针对性地为消费者提供时间敏感的移动广告。
- **大数据分析：数据发现和业务智能。** Oracle Exalytics 是用于 Oracle Endeca Information Discovery 和 Oracle Business Intelligence Foundation 的工程化系统。Oracle 已将这两个产品相集成，包括将 BI 元数据的自动索引集成到 Oracle Endeca 中，从而在关系报告与非关系发现之间架起了桥梁。BI Foundation 与 Hive 的原生集成也表明了 Oracle 为一起部署关系环境与非关系环境所做的努力。

- **大数据应用程序：数据驱动的行动。** Oracle 提供了包含嵌入式分析的各种应用程序，包括从用于公用事业的智能仪表管理解决方案到营销活动管理以及人力资本管理等应用程序。

图 5

Oracle 的大数据及分析技术



注意：图中所示的 Oracle 大数据及分析技术只是为了便于说明，并非涵盖所有内容。

信息来源：IDC，2014

建议

为保持竞争优势而需要越来越关注大数据及分析解决方案对于大多数组织而言既是一个机会也是一项挑战。更好、更快地制定数据驱动型决策的承诺促使各个组织将大数据及分析功能列入执行议程的首位。然而，许多组织还没有成熟的大数据及分析解决方案来满足利用大数据资产和普遍部署的分析来优化运营决策、战术决策和战略决策所需的一系列技术、人员和过程的要求。此外，各个组织扩展各自大数据及分析功能的起点也是不同的。一些组织拥有现有的基本业务智能功能，一些组织可能具有适当的数据仓库支持，一些组织可能部署了 Hadoop 集群但不具备其他大数据及分析技术功能。您组织的现状将决定从试点和概念验证项目发展为广泛采用和接受的解决方案要采取的下一步措施和方法，从而提高竞争优势或为所有利益相关者创造价值。

IDC 在以下各节中提供了在长期对全球各个组织进行的调研中得出的考虑事项。

阶段：试点项目

如果您的组织处于大数据及分析功能发展过程中的早期阶段，我们建议侧重于概念验证项目或试点项目，在这类项目中，通过新知识以及了解大数据及分析技术组件和它们之间的相互作用定义了初始价值。一种方法可以从您的关系数据仓库和 NoSQL 数据库的功能或 Hadoop 部署开始。首先评估这些技术在您的具体情况下的潜力，然后评估与数据仓库的集成点以及关系组件与非关系组件之间的数据集成或移动功能。在这个阶段，价值体现在哪些技术在您特定的环境中起作用，哪些技术不起作用。最初，业务价值可能非常有限，但在随后的阶段中，随着试点项目发展为生产解决方案并加以实施，知识价值将增长，业务价值机会将显现出来。

例如，一家保险公司想要改善其索赔欺诈检测方案。为此，该公司启动了一个项目，该项目最初侧重于从通过电话提出的车祸索赔中捕获洞察。该保险公司使用新技术捕获音频文件，将这些文件转译为文本，分析文本并使用从非结构化内容分析中获得的新洞察加强其现有的欺诈模型。

从该示例和类似示例中获得的经验教训表明，有必要：

- 制定一项初始的部门级大数据及分析战略或一份意向声明（例如，需要减少索赔欺诈）。为已获得部门或业务单元领导的管理支持的本地化项目做预算。对于大数据项目，没有所谓的“大爆炸”式方法 — 从小规模开始，反复执行。
- 利用现有资源启动一个概念验证项目或试点项目。专注于某特定业务领域内的项目。在该示例中，保险公司不试图一次性改进整个索赔过程 — 只关注于改善其现有的欺诈模型。
- 使用现有数据识别是否不完整以及是否缺少必要的质量，这将需要手动进行数据准备。开始集成来自多个源的数据，使来自大数据及分析系统的信息具有较高的可信度。这家保险公司结合使用音频内容（已转换为文本）和基于关系技术的运营应用程序中的事务数据，以发现新洞察。
- 部署针对特定类型的数据、分析技术和最终用户交互优化的新技术（内部部署或在云中部署）。由于最初已将来自多个源的多结构化数据相集成，该技术将包括关系工具和非关系工具。
- 在建立初始项目验证点之前，不要花大量的时间和精力试图获得管理层支持。向具备专业化大数据及分析技能的同事寻求支持。然后成立一个大数据及分析团队，团队成员均具备现有技术和新部署技术的技能。确保数据科学家可使用新技术和数据进行试验以发现新洞察。

阶段：专业工作

如果您的组织已经完成了基本工作，且在生产环境中已经有了一个或多个大数据及分析解决方案，则可以建立可重复的大数据及分析实践，以便持续使用关系和非关系大数据及分析技术以及相关分析。在大数据及分析过程的这个阶段，业务价值已明确实现，但通常仍限于各个业务部门级别。

我们扩展一下前面启动（实际）的示例，这家保险公司创建改善的欺诈检测模型初见成效后，将这些模型实施到正在运行的欺诈检测应用程序和流程中。这意味着确保适当的系统性能、可用性和安全性，以及确保提供适当流程和技术人员来持续监视和定期改善分析。

从该示例和类似示例中获得的经验教训表明，有必要：

- 制定跨部门、业务单元级的大数据及分析战略。根据业务单元需求做预算。对大数据及分析项目执行本地化成本效益分析。在该示例中，欺诈检测只是扩展的索赔管理过程的一部分，在项目的这个阶段，分析人员和操作人员通力协作来实施新的分析和洞察。
- 继续扩大内部多结构数据源的可用性并对其进行集成。请注意，数据治理策略和过程在单个业务单元级别将很难实施。
- 扩大适应性技术的可用性，并知道最初采用将具有选择性。
- 根据大数据及分析战略分配、培训和雇佣员工。借助外部专业服务提供商加强现有技能。
- 开始监视和记录决策过程和决策成效。确保大数据及分析团队成员包含所有利益相关方的代表，以促进协作。这家保险公司为初步监视基于新分析制定的决策质量以及持续监视系统及其组件的性能分配员工时间。

阶段：获取竞争优势

如果您的组织已在利用一系列大数据及分析技术、技能和流程，则说明它处于强势地位，可以开始从其大数据及分析功能获得竞争优势。为此，您的组织将需要建立管理和优化实践及方法，以便将大数据及分析功能充分整合到组织文化和业务流程中。通过展现出来的大数据及分析 IT 效率，将使结果更容易预测，也更容易将大数据及分析带来的新的产品和服务商机转换为商业计划。从那时起，组织将能够进入一个持续学习和改进的优化状态，在这种状态下，以前无法实现的业务价值会不断产生。

在下一阶段，这家保险公司认识到，改善的欺诈检测模型以及对索赔中报告的潜在损害进行预测意味着，可以将以前由最有经验的员工执行的一些例行调查工作免去或分配给初级员工。从而将最有经验的员工解放出来，使他们仅专注于解决新增自动化操作无法处理的实际异常情况。

从该示例和类似示例中获得的经验教训表明，有必要：

- 制定由首席级高管支持的企业级大数据及分析战略。为大数据及分析项目做预算。为业务案例开发以及绩效及结果度量提供可用的工具和方法。最初的新欺诈检测洞察和索赔过程自动化的相关变更导致了流程的重新设计，也改变了人员配备要求，进而改变了公司与客户的交互行为。此外，该公司还从降低欺诈相关成本以及人力资源成本获得了可量化的收益。
- 将所有相关内部和外部数据源的可用信息提供给具有适当安全权利的用户。建立数据治理的量和度和方法，以及建立衡量大数据及分析流程、员工和结果所依据的量和度。
- 最大限度地使用适应性系统和负载优化的系统、自动系统性能管理以及大数据及分析技术的动态可伸缩性。大多数（如果不是全部的话）解决方案将需要结合使用关系技术和非关系技术。将预测性分析整合到技术性能监视和管理流程中。确保与 IT 供应商协商制定适当的技术定价结构，以便技术得以广泛采用。
- 定期对所有大数据及分析的技术员工、分析员工和业务员工提供培训。最大程度地使大数据及分析员工集中工作，以集中完成数据集成、系统管理以及报告和看板开发等任务，同时确保分析师与业务线密切配合。这家保险公司能够改变员工构成，提高生产力以及根据经验优化员工利用率。
- 确保实验流程、绩效管理流程和运营 BI 流程都有适当人员、技术和资金的强大支持。利用决策管理技术持续改进流程以及将分析集成到业务流程中。

总结

智能设备、社交网络、移动性以及大数据及分析的融合开辟了一个新的经济体制，它重新定义了生产商、分销商和消费者之间的关系。新（大）数据的爆炸式增长、周期更短以及新一代关系和非关系信息管理和分析技术的推出，这一切都彰显出了改变决策制定方式以利用这些新形势在市场中取得优势的需求和机会。

大数据及分析市场发展迅速，关于各种大数据及分析用例的最合适技术的问题比比皆是。此时提出这样的问题正合时宜，以便了解可用选项以及制定大数据及分析战略。然而，所提出的问题不应该是，为满足您组织的大数据及分析要求，是选择关系技术还是非关系技术。答案是您需要同时使用这两种技术。大数据及分析平台以及在上面部署的新一代数据可视化和分析工具及应用程序必须能够：

- 整合结构化数据和非结构化内容
- 处理批量传输和连续流入的数据
- 使用高级且可预测的分析以及企业级、可实施的、严格治理的结构化信息流持续、快速地进行试验
- 将大数据及分析解决方案的成果传达到运营系统和员工中

关于 IDC

International Data Corporation (IDC) 是一家专注于信息技术、通讯和消费技术市场的市场情报、咨询服务以及各种活动的主要全球提供商。IDC 帮助 IT 专业人士、业务主管和投资团体就技术采购和商业战略制定基于事实的决策。IDC 拥有 1000 多名分析师，就全球 110 多个国家/地区的技术和行业机遇及发展趋势提供全球性、地区性和当地的专业建议。48 年来，IDC 一直为客户提供战略洞察，帮助他们实现关键业务目标。IDC 是全球领先的技术媒体、研究和会展公司 IDG 旗下的一家子公司。

全球总部

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com

版权声明

IDC 信息和数据的对外发布 — 在广告、新闻报道或促销材料中使用任何 IDC 信息，均需事先得到 IDC 副总裁或地区经理的书面批准。提出此类请求时，应随附相关文档的草稿。IDC 保留出于任何理由而拒绝批准外部使用的权利。

版权所有 2014 年，IDC。未经书面许可不得复制。



甲骨文（中国）软件系统有限公司

北京远洋光华中心办公室

地址：北京市朝阳区景华南街5号远洋光华中心C座21层
邮编：100020
电话：(86.10) 6535-6688
传真：(86.10) 6515-1015

北京汉威办公室

地址：北京市朝阳区光华路7号汉威大厦10层1003-1005单元
邮编：100004
电话：(86.10) 6535-6688
传真：(86.10) 6561-3235

北京甲骨文大厦

地址：北京市海淀区中关村软件园24号楼甲骨文大厦
邮编：100193
电话：(86.10) 6106-6000
传真：(86.10) 6106-5000

北京国际软件大厦办公室

地址：北京市海淀区中关村软件园9号楼国际软件大厦二区308单元
邮编：100193
电话：(86.10) 8279-8400
传真：(86.10) 8279-8686

北京孵化器办公室

地址：北京市海淀区中关村软件园孵化器2号楼A座一层
邮编：100193
电话：(86.10) 8278-6000
传真：(86.10) 8282-6401

上海名人商业大厦办公室

地址：上海市黄浦区天津路155号名人商业大厦12层
邮编：200001
电话：(86.21) 2302-3000
传真：(86.21) 6340-6055

上海腾飞浦汇大厦办公室

地址：上海市黄浦区福州路318号腾飞浦汇大厦508-509室
邮编：200001
电话：(86.21) 2302-3000
传真：(86.21) 6391-2366

上海创智天地10号楼办公室

地址：上海市杨浦区淞沪路290号创智天地10号楼512-516单元
邮编：200433
电话：(86.21) 6095-2500
传真：(86.21) 6107-5108

上海创智天地11号楼办公室

地址：上海市杨浦区淞沪路303号创智天地科教广场3期11号楼7楼
邮编：200433
电话：(86.21) 6072-6200
传真：(86.21) 6082-1960

上海新思大厦办公室

地址：上海市漕河泾开发区宜山路926号新思大厦11层
邮编：200233
电话：(86.21) 6057-9100
传真：(86.21) 6083-5350

广州国际金融广场办公室

地址：广州市天河区珠江新城华夏路8号合景国际金融广场18楼
邮编：510623
电话：(86.20) 8513-2000
传真：(86.20) 8513-2380

成都中海国际中心办公室

地址：成都市高新区交子大道177号中海国际中心7楼B座02-06单元
邮编：610041
电话：(86.28) 8530-8600
传真：(86.28) 8530-8699

深圳飞亚达科技大厦办公室

地址：深圳市南山区高新南一道飞亚达科技大厦16层
邮编：518057
电话：(86.755) 8396-5000
传真：(86.591) 8601-3837

深圳德赛科技大厦办公室

地址：深圳市南山区高新南一道德赛科技大厦8层0801-0803单元
邮编：518057
电话：(86.755) 8660-7100
传真：(86.755) 2167-1299

大连办公室

地址：大连软件园东路23号大连软件园15号楼502
邮编：116023
电话：(86.411) 8465-6000
传真：(86.755) 8465-6499

苏州办公室

地址：苏州工业园区星湖街328号苏州国际科技园5期11幢1001室
邮编：215123
电话：(86.512) 8666-5000
传真：(86.512) 8187-7838

沈阳办公室

地址：沈阳市和平区青年大街390号皇朝万鑫国际大厦A座39层3901&3911室
邮编：110003
电话：(86.24) 8393-8700
传真：(86.24) 2353-0585

济南办公室

地址：济南市泺源大街150号中信广场11层1113单元
邮编：250011
电话：(86.531) 6861-1900
传真：(86.531) 8518-1133

南京办公室

地址：南京市玄武区洪武北路55号置地广场19层1911室
邮编：210018
电话：(86.25) 8579-7500
传真：(86.25) 8476-5226

西安办公室

地址：西安市高新区科技二路72号西安软件园零壹广场主楼1401室
邮编：710075
电话：(86.29) 8834-3400
传真：(86.25) 8833-9829

重庆办公室

地址：重庆市渝中区邹容路68号大都会商厦1611室
邮编：400010
电话：(86.23) 6037-5600
传真：(86.23) 6370-8700

杭州办公室

地址：杭州市西湖区杭大路15号嘉华国际商务中心810&811室
邮编：310007
电话：(86.571) 8168-3600
传真：(86.571) 8717-5299

福州办公室

地址：福州市五四路158号环球广场1601室
邮编：350003
电话：(86.591) 8621-5050
传真：(86.591) 8801-0330

南昌办公室

地址：江西省南昌市西湖区沿江中大道258号
皇冠商务广场10楼1009室
邮编：330025
电话：(86.791) 8612-1000
传真：(86.791) 8657-7693

呼和浩特办公室

地址：内蒙古自治区呼和浩特市新城区迎宾北路7号
大唐金座19层北侧1902-1904室
邮编：010051
电话：(86.471) 3941-600
传真：(86.471) 5100-535

郑州办公室

地址：河南省郑州市中原区中原中路220号
裕达国际贸易中心A座2015室
邮编：450007
电话：(86.371) 6755-9500
传真：(86.371) 6797-2085

武汉办公室

地址：武汉市江岸区中山大道1628号
武汉天地企业中心5号大厦23层2301单元
邮编：430010
电话：(86.27) 8221-2168
传真：(86.27) 8221-2168

长沙办公室

地址：长沙市芙蓉区韶山北路159号通程国际大酒店1311-1313室
邮编：410011
电话：(86.731) 8977-4100
传真：(86.731) 8425-9601

石家庄办公室

地址：石家庄市中山东路303号石家庄世贸广场酒店14层1402室
邮编：050011
电话：(86.311) 6670-8080
传真：(86.311) 8667-0618

昆明办公室

地址：昆明市三市街六号柏联广场写字楼11层1103A室
邮编：650021
电话：(86.871) 6402-4600
传真：(86.871) 6361-4946

合肥办公室

地址：安徽省合肥市蜀山区政务新区怀宁路1639号平安大厦18层1801室
邮编：230022
电话：(86.551) 6595-8200
传真：(86.551) 6371-3182

广西办公室

地址：广西省南宁市青秀区民族大道136-2号华润大厦B座2302室
邮编：530028
电话：(86.771) 391-8400
传真：(86.771) 577-5500



企业中的大数据：世界巨变

2014 年 2 月

作者：Dan Vesset, Carl W. Olofson

公司网址：<http://www.oracle.com>（英文）

中文网址：<http://www.oracle.com/cn>（简体中文）

销售中心：800-810-0161

售后服务热线：800-810-0366

培训服务热线：800-810-9931

欢迎访问：

<http://www.oracle.com>（英文）

<http://www.oracle.com/cn>（简体中文）

版权©2014 归 Oracle 公司所有。未经允许，不得以任何形式和手段复制和使用。

本文的宗旨只是提供相关信息，其内容如有变动，恕不另行通知。Oracle 公司对本文内容的准确性不提供任何保证，也不做任何口头或法律形式的其他保证或条件，包括关于适用性或符合特定用途的所有默示保证和条件。本公司特别声明对本文档不承担任何义务，而且本文档也不能构成任何直接或间接的合同责任。未经 Oracle 公司事先书面许可，严禁将此文档为了任何目的，以任何形式或手段(无论是电子的还是机械的)进行复制或传播。

Oracle 是 Oracle 公司和/或其分公司的注册商标。其他名字均可能是各相应公司的商标。