

企业中使用大数据和 Hadoop 的例证



1 2 3 4 5 6 7 8

简介

医疗保健业

金融服务业

能源与公共
事业

电信业

汽车与制造业

IBM
InfoSphere
BigInsights
的优势

大数据的优势

简介

大数据是指来源众多、格式多样的结构化数据和非结构化数据，在不同的行业中表示不同的内容。尽管各行业对大数据的需求和使用方法不尽相同，但这些行业有一个共同点，即利用大数据对客户、产品、合作伙伴、竞争对手和战略做出更佳、更明智的决策。

例如，在支付方、患者和提供方关系十分松散的医疗业中，大数据和分析能够针对各种模式提供深刻见解，从而提高护理水平、患者护理水平和组织效率。在金融服务业，大数据技术可以对包含动态流数据在内的多个来源的数据进行分析，以快速识别潜在的欺诈行为。通过分析机器传感器和客户记录中的大数据，制造商能够获得从产品质量到供应链优化的全部信息。

随着企业对大数据和分析的潜在功能开始进行探究，他们需要能够快速存储和访问数据的新方法。Apache Hadoop 可以轻松应对这一挑战。过去，数据分析的几大关键方面处理起来过于复杂或成本高昂，而现在，使用 Hadoop 即可轻松解决。例如，Hadoop 不仅能够访问文件、注释、联系人、电子表格和电子邮件等传统应用中的非结构化和半结构化复杂数据，还可访问新闻媒体、web 反馈、视频、图片和传感器数据等创新应用中的非结构化和半结构化复杂数据。



此电子书重点介绍 Hadoop 为若干行业带来的优势并探讨 IBM® InfoSphere® BigInsights™ 如何将开源 Hadoop 与企业级管理和分析功能相结合。

成功部署 Hadoop 的关键因素

任意行业都可以从部署 Hadoop 中获得同样的益处，其中包括增强系统性能以提高分析结果的速度，改进风险管理和运营效率，以及提高新产品和服务的创新速度。

但要获得以上优势，必须在部署基于 Hadoop 的项目的过程中对具体的最佳实践多加注意。请牢记四点。

1 企业不仅应选择数据源随时可供访问且采用电子形式的高投资回报潜在项目，还要制定明确的目标和可量化指标。将结果数据供广大用户群体轻松访问的业务需求应该十分强烈。

2 数据架构应具有可扩展性，以按需添加流数据等其他数据源。

3 在项目进行过程中，创建一个反馈回路，以向其他部门分享与产品、营销和销售相关的见解。此举有助于提升分析价值，形成一种侧重于从分析得出更佳信息的企业文化环境，同时还可逐渐提升对数据可靠性和完整性的信任程度。

4 应将 Hadoop 放在由大数据工具和分析功能构成的强大生态系统中使用。所选 Hadoop 解决方案具备的功能组合越丰富，团队解决问题和提高企业洞察力的自由度就越高。

医疗保健业

如今，医疗保健机构处理数据的速度远不及数据的增长和移动速度，但如果医师能够从这些数据中及时获得恰当的见解，则会对患者的生死存亡产生重大影响。

实时数据分析和分布一直是医疗保健机构难以攻克的难题。患者记录只是众多关键信息源之一，此外还包括索赔、财务记录、客户关系管

理 (CRM) 系统、企业及护理合作企业参考、研究日志以及医师信件等其他信息源。此类数据中的大部分均为非结构化数据，并且在不断发生变化。此外，这些数据通常分布在多个来源和部门中。**访问这些宝贵数据并将其归类为临床分析和高级分析，对于提高护理质量和收入水平、激励患者行为和改善效率来说至关重要。**

通过借助 Hadoop 和流技术成功利用大数据，能够实现医疗保健业转型的几大关键目标，其中包括：

- 构建可持续的医疗保健体系和健康信息交换
- 提高临床治疗效率并降低再入院率
- 减少医疗差错并支持协作
- 检测索赔欺诈和其他滥用医疗资源的行为

例证：亚洲卫生局

亚洲某卫生局希望在蓬勃发展的地区实施远程医疗，同时帮助放射科医生缓解眼疲劳，因而开发了一款医学成像分析系统，用于自动检测异常现象。但是，要适应疾病专用算法与大量数据的处理绝非易事，所以该局选择 Hadoop 并结合 IBM 技术以挖掘这些大型成像数据集（10 GB 到 TB、甚至是 PB 级别不等）。

该医学成像诊断平台预期可允许医师利用其他医师在治疗类似病例时积累的经验并推断预后和治疗结果，从而大幅提高患者的医疗保健质量。该平台还汇集了全部相关数据，允许医师查看达成一致的意见以及不同的替代方案，从而有助于降低诊断的不确定性。



金融服务业

在金融服务业，由多种信息源、客户、合作伙伴和电子反馈生成的高度详细的数据，不仅数量大、类型多，增长速度也十分迅猛，这成为该行业长期以来的显著特征，因此对于大数据的管理挑战并不陌生。但如今，庞大的非结构化数据（包括社交媒体文章）无法永远适用于过去以交易为中心的分析系统，并且交易和客户数据流也容易成为吸引黑客和欺诈者的目标。

金融服务公司如果能利用和保护以交易、实时市场反馈、客户服务记录、通信和社交媒体文章等形式存在的大数据，则可获得前所未有的业务洞察。成功利用和保护大数据可帮助金融服务公司实现三大关键增长目标：

- 通过开发客户的全方位视图来**提高客户洞察力**，从而推行更具针对性的营销活动、定制服务和交叉销售机会，并提高所有沟通渠道的响应速度。

- 通过对流数据进行实时分析，并结合多个数据源来确定可能指示潜在欺诈行为的关系和模式，以此**提高风险管理水平和欺诈检测水平**。
- 通过使用大数据技术，支持那些能够提高效率的流程 - 从更快、更明智的客户服务响应到企业支付中心，从而**提高灵活性并简化运营**。

例证：帝国数据银行

总部位于日本东京的帝国数据银行在企业信用调查、信贷风险管理服务、数据库服务、营销服务和电子商务支持服务等领域为金融服务业客户提供服务。该公司存有 160 万家公司的企业信用报告文件、68 万家公司的财务报表、由 142 万家公司组成的企业档案数据库以及 410 万家公司的其他企业数据。

为提供更加深入的分析，该公司需要将来自其他来源的公开信息整合起来，为此使用 Hadoop 和 IBM 技术创建了一个系统。如今，帝国数据银行能够在 30 分钟内处理几十亿项文本数据，而在过去，该工作需耗时数天才能完成。此外，该公司能够分析的数据量接近客户数据的五倍，因而能以更快的速度响应客户请求。凭借这些先进的功能，帝国数据银行现已在瞬息万变的市场中拥有极具竞争力的差异化优势。



能源与公共事业

对于能源与公共事业领域，流数据不仅带来巨大的挑战，更创造出巨大的机遇。智慧仪表、石油钻塔和发电站上的传感器、工厂效率系统、温度计、水流监测器等，所有这些资源及其他资源每天要进行多次测量并发送多次测量数据，有时每分钟就要进行好几次。**大数据技术可以将智慧仪表和智慧电网项目的信息转化成重要的运营见解以及与公司资产和客户行为相关的见解。**

能源与公共事业公司依靠 Hadoop 分析用户数据并向客户提供相关的见解，从而控制能源消耗。目标：向客户展示行为改变、设备或其他改进措施如何降低能源消耗并改善直接客户沟通。

大数据和分析也可提高运营效率。源自智慧电网的数据可帮助公共事业公司从现有资源中获取更多信息并为主动维护和容量管理作出规

划。对天气和地理数据的分析不仅能帮助公司识别新型环保能源部署（风电场、太阳能系统等）的位置，还可助其优化工作人员的交通情况，以及将原材料或建筑部件及时运送到施工地点和加工工厂。

例证：维斯塔斯风力技术公司

总部位于丹麦奥尔胡斯的维斯塔斯风力技术公司主营风力发电系统的开发、制造、销售和维护业务，现已在六大洲的 66 个国家/地区安装 43,000 余台陆基及海上风力涡轮机。该公司需要精确地放置风力涡轮机以最大限度提高设备性能和使用寿命，因此希望扩充其风力数据库，以收集更长时间段内更大范围的天气数据。

要实现这一目标，该公司需要一个更强大的计算平台来以更快的速度进行全球风力预测。为此采用了 IBM 企业级 Hadoop 和其他 IBM 技术，最终减少近 97% 的风力预测信息响应时间，相当于从数周减少到数小时，从而缩短系统开发时间。Hadoop 平台能够分析更宽、更深的数据库，并借此提高涡轮机安放的准确性。最终，公司借助该平台不仅减少了每千瓦时的客户成本，而且提高了客户的投资回报率，同时减少了 IT 规模 and 成本。



电信业

毋庸置疑，电信业是一个由数据驱动的资本密集性行业。移动网络的部署以及移动设备和社交媒体的普及使客户和市场数据大大增加。要保持竞争力，电信公司必须对瞬息万变的市场环境快速做出响应。

Hadoop 可帮助服务提供商削减基础设施成本，同时提高上市速度。只有确保网络的最佳性能才能扩大并留住客户群。Hadoop 具有极高的可扩展性，能够以合理的成本适应大数据量；它还具有机器学习功能，可帮助防止服务中断。通过对手机和设备数据的有效采集，可为创新产品开发提供支持，并帮助公司对产品故障情况及时做出响应。

Hadoop 还可以处理数十亿条呼叫数据记录，支持以具有成本效益的小规模方式存储数据。通过更全面且更深入地了解客户，电信企业能够提供符合当前利益的服务和内容，同时降低客户流失率。

例证：西尔康公司

总部位于吉隆坡的西尔康公司是亚通集团的子公司，拥有超过 1.9 亿名移动用户。其核心业务是提供预付和后付费形式的移动语音服务，并在移动宽带、移动商务、企业解决方案和大量批发业务这几个细分市场实现了快速发展。由于身处竞争激烈的饱和市场领域，因此西尔康公司需要显著增强对客户了解，以便保留客户群体并增加市场份额。该公司想要利用既有的客户信息和分析功能来管理定向营销活动。

为此，公司已部署 Hadoop 和 IBM 分析工具，以便通过最合适的营销信息锁定各个客户。西尔康公司可为客户提供更合适的计划，这有助于提高每用户平均收入 (ARPU) 率和降低客户流失率，此外还有助于建立客户满意度和忠诚度。借助这些分析工具，该公司能够将新活动的启动时间缩短 80% 以上，并将活动的成效提高 70% 以上。



汽车与制造业

远程信息技术和其他传感器技术的使用日益增多，这会改变汽车和制造企业接触客户以及实现业务战略的方法。此流数据成为内部业务运营、研发活动、供应链管理和企业资源规划 (ERP) 系统中众多数据的一部分，并成为向企业展示巨大商机以获取竞争优势的客户数据的一部分。

通过部署 Hadoop 和其他大数据及分析技术，汽车和制造企业可有效管理所有这些数据。以下三个驱动因素在全球市场中尤为重要：

- **提高客户满意度及保留率：** Hadoop 可帮助企业整合来自呼叫中心业务、社交媒体、服务记录和产品调查等来源的数据，进而掌握客户情绪并对反馈作出合适的响应。营销和销售团队能够更好地确定交叉销售机会，设计团队则能收集针对新提出的设计理念的各种意见。
- **提高产品质量：** 从主动发出服务通知到分析制造工艺和测试组件质量，大数据都具备丰富的洞察力，可使得能精确处理及分析大数据的企业了解制造工作。
- **优化供应链：** 内部供应商、外部供应商、运输提供商和库存会产生大量的复杂数据。Hadoop 可以对供应链经理、RFID 标签和生产车间传感器所收集的结构化数据及非结构化数据进行采集与分析，进而形成有助于防止在流程中出现缺货、故障和延迟现象的交付前分析。

例证：美国汽车制造商

底特律的一家汽车制造商想要系统地获取并分析大量数据、提高系统效率、简化数据分析并加快决策过程。该公司具备以功能为导向且由应用驱动的 IT 和数据管理环境，并且数据在大部分外包的应用程序中处于孤立状态。这些应用还属于单独的大型职能领域，例如设计或营销，不同领域之间几乎不存在可见性和数据共享。

IT 部门部署了 Hadoop 和 IBM 技术，以将数据整合到单一企业数据仓库中，这样各个业务线便都能使用这些数据，从而使公司能使用分析功能解决跨职能问题，以及提高业务绩效和竞争力。这创造出新的分析点，该公司可借此开发能收集来自修复点的经销商数据的项目，这些数据包括销售和营销信息以及服务信息等。通过整合该数据，系统会将其与生产数据和产品设计信息一起进行分析，以帮助制造商更好地确定服务问题，包括保修问题、特定供应商或装配商的问题或者产品设计过程中出现的问题。



IBM InfoSphere BigInsights 的优势

IBM 提供开源 Hadoop 的标准实施，因此，客户可以在利用其丰富的功能的同时，充分利用额外的企业级数据功能。这些功能包括：

- **IBM Big SQL:** SQL-on-Hadoop 具备大规模并行处理查询优化和执行功能，可如实地实施 ANSI SQL 标准并以本地 Hadoop 数据类型进行操作
- **IBM BigSheets:** 适用于业务用户的电子表格风格的数据处理和可视化工具
- **Big R:** 开源 R 语言的集成实施，可在 Hadoop 集群内自动进行并行统计计算

- **自适应 MapReduce:** 与 Apache MapReduce 兼容的性能更好且工作负载更可靠的备用低延迟框架
- **IBM GPFS™ FPO:** 符合 POSIX 的备用 Hadoop Distributed File System (HDFS)

IBM 还提供大量的开发工具，其中包括可用于加速文本分析应用程序开发的 AQL、开发和查询语言 JAQL（基于 JSON）、用于快速构建和部署特定类型应用程序（例如社交媒体或机器数据分析）的开发工具包，以及可简化大数据应用程序的开发和维护过程的基于 Eclipse 的集成开发环境 (IDE)。

借助 IBM InfoSphere BigInsights，可以更方便地使用 Hadoop 获取大数据的价值，以及建立大数据应用程序。它支持开源技术，通过增加以下内容来满足企业需求：

- 针对更高性能的优化工作负载
- 一流的分析功能
- 数据发现、开发和配置功能
- 安全和数据管理的增强特性
- 管理和支持功能

最终为复杂的大规模大数据和分析项目提供更加用户友好的解决方案。

大数据的优势

不论从事哪种行业，各企业都能通过 InfoSphere BigInsights for Hadoop 获益。由于 IBM 已将其关系数据库管理系统工具与 Hadoop 相集成，因此，IT 可凭借熟悉的技术和过程充分利用 Hadoop（例如 SQL、电子表格和 R）。IBM 还能轻松地适应来自各种数据源的数据，从结构化的 CRM 和供应商关系管理系统，到非结构化的电子邮件、社交媒体以及其他行业特定应用程序等，不一而足。此外，IBM 还创建了各种链接，可用于更轻松地从非结构化数据源中提取数据，这有助于弥补企业部署新技术时可能会出现的技术不足。

Hadoop 能够对之前的不同数据源进行整合，从而为公司提供更全面且更深刻的见解。借助 InfoSphere BigInsights 更易于实现 Hadoop 的优势。通过将二者相结合，企业便可以同时利用内部和第三方外部数据，以及进行一流的数据分析。这些先进功能适用于各行各业的各种要求，其中包括流数据的实时分析、为识别潜在欺诈行为而进行的采购数据分析、适用于解决初期支持问题的社交媒体渠道分析，以及用于发现潜在疾病发生风险的医疗记录分析。



通过将 Hadoop 处理速度与 IBM 分析技术及创新相结合，企业可全面实现所有企业数据资源的价值，并能解决之前无法解决的行业特定难题。

资源

如需了解有关 Hadoop 和大数据的更多信息，以及了解开启企业大数据之旅的方式，请访问以下网站：

- www.ibmbigdatahub.com
- www.bigdatauniversity.com
- <https://developer.ibm.com/Hadoop>
- <https://developer.ibm.com/streamsdev>

如需了解有关 Hadoop、大数据分析和 IBM InfoSphere BigInsights 的更多信息，请联系 IBM 代表或 IBM 业务合作伙伴，或者访问：ibm.com/hadoop

还可以通过以下网站下载 IBM InfoSphere BigInsights QuickStart Edition（非生产用途免费）来开始 Hadoop 学习之旅：
ibm.com/software/data/infosphere/biginsights/quick-start

此外，IBM 全球融资部可帮助您以最具成本效益的战略方式购买贵企业需要的软件功能。我们将与有信誉的客户合作，共同定制能够适应业务和发展目标、帮助实现有效的现金管理、降低总体拥有成本的融资解决方案。IBM 全球融资部是您助力重要 IT 投资及推动业务发展的明智选择。欲了解更多信息，请访问：
ibm.com/financing



© IBM 公司版权所有 2014

IBM Corporation
Software Group
Route 100
Somers, NY 10589

美国印制
2014 年 9 月

IBM、IBM 徽标、ibm.com、BigInsights、GPFS 和 InfoSphere 是国际商业机器公司在全球许多司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。当前的 IBM 商标列表请见网站的“版权和商标信息”版块：ibm.com/legal/copytrade.shtml

本文档包含截至发布之日的最新信息，IBM 可能随时更改。并非所有产品或服务在 IBM 开展业务的所有国家/地区均有提供。

援引的性能数据和客户实例仅供说明之用。实际性能结果可能因具体的配置和运行环境而有所不同。本文所载信息按“原样”提供，不做任何明示或暗示的担保，包括对适销性、特定目的的适用性的任何担保，以及针对非侵权的任何担保或条件。IBM 根据产品交付协议中规定的条款和条件为产品提供担保。

客户应负责确保遵守适用法律和法规。IBM 不提供法律建议或声明或保证其服务或产品能够确保客户遵循所有法律或法规。

实际可用存储容量可能表示未压缩和压缩数据容量，会有所不同，可能低于标示容量。



请回收再利用
