



实时组织的数据流分析

实时洞察如何释放全新业务价值

目录

数据流分析的广泛商机	1
数据流分析——从董事会到基层	2
数据流分析的首要步骤	5
英特尔对数据流分析的贡献	6
通过数据流分析帮助您的组织做好准备， 获取实时洞察	6

“在实时的世界里，企业应用
也应做到实时。”

— Mike Gualtieri,
Forrester Research 副总裁兼首席
分析师¹。

数据流分析的广泛商机

由于各行各业的企业希望处理并关联多个来源的数据以实时获取洞察，数据流分析的商机迅速增加。越来越多的企业利用内存数据库和数据流分析来实时处理、关联和分析数据——无论是边缘设备数据、本地数据还是云端数据，数据流分析必将成为主流。

这一趋势具有广泛影响，各层级的决策者将从后顾性决策转向前瞻性决策，因此依靠数据流分析提供洞察的业务需求将随之增长。如今，许多需要实时洞察的大数据解决方案在业务部门 (LOB) 直接产生，而非仅仅来自首席信息官办公室，因为组织意识到获取洞察的速度是企业能否脱颖而出的关键因素，将带来敏捷性和竞争力。

物联网也是推动因素之一。“智能”设备和传感器的普及直接推动了对数据流分析的需求，所有这些设备和传感器都能够生成大量新数据。分析公司 Gartner* 表示，在 2020 年，全球使用的互联网接入设备将达到 200 亿台，超过 65% 的企业 (2017 年为 30%) 将采用物联网产品。²

数据流分析日益重要，其发挥的作用可参考行业关键参与者的经验——如 SAP* 和 SAS*。例如，SAS 事件流处理 (ESP) 已应用于包括预测性分析在内的各种场景。我们将介绍 ESP 如何能够影响运营流程。同时，SAP HANA* 可与内存中的进程协同工作，在整个组织中推动洞察的获取，直达董事会层级。我们还将探讨 Apache Spark* 的潜力，在 2015 年至 2021 年间，预计该技术的年复合增长率 (CAGR) 将达 67%³。这是一项突破性技术，对数据流分析与机器学习和人工智能 (AI) 的关联至关重要。

企业可以通过多种方式转型为“实时洞察企业”，在优化数据流分析工作负载方面，英特尔一直处于前沿。在本白皮书中，我们将分析商机，为您指引方向，寻找数据流分析的未来道路。

数据流分析为何成为热门话题？

传统意义上，数据管理和分析功能主要以批处理为导向，用于处理有限的历史数据集。更大的数据校正、更快的吞吐量和更短的分析时延支持在数秒内生成和交付洞察，从而为新类型的决策制定、新客户体验和更出色的重复任务自动化开辟道路，原因有以下几个融合因素：

- **网络突破性技术**（如 25GbE 和软件定义网络 / 网络功能虚拟化 (SDN/NFV)）使连接性能够满足数据爆炸式增长带来的海量需求。
- **边缘设备**——从物联网 (IoT) 传感器到本地处理——变得更智能、更强大，支持更智能的数据收集、整理和分发。Gartner* 表示，到 2020 年，使用的互联设备将达到 204 亿台²。
- 不管是在数据中心还是在云端，**核心计算的能力**大幅提高，从而创建大规模可扩展的处理架构。
- **内存处理**利用更快、更密集的功能，如英特尔® 傲腾™ 技术，产生革命性飞跃，提高了内存容量，减少了时延，加快了应用速度，可实现快速缓存和存储、降低时延敏感性工作负载事务成本、增加每台服务器的规模。
- **分析软件工具**已进行调整和演进，结合业务部门应用按需提供洞察。例如，Apache Spark* 等大数据分析平台使用机器学习技术来实时处理、关联和分析来自多个来源的流式数据。

数据流分析——从董事会到基层

随着数据流分析被应用于日益广泛的应用案例，从董事会到基层，整个组织都能感受到洞察所带来的优势。以下场景展示了数据流分析解决方案的强大功能。

“机器智能是……一种预测技术，因此经济转型将以预测成本降低为中心……并且有两大已确认的经济影响。首先，我们将利用预测技术探索之前未涉及的领域。第二，用于补充预测技术的其他事物的价值将增加。”⁴

— **Ajay Agrawal 教授**
(多伦多大学经济学家) 和同行

通过 SAS 事件流处理改进运营

软件公司 SAS 专注大数据和分析业务超过 40 年。凭借其高级 ESP 引擎，SAS 物联网分析* (SAS Analytics for IoT) 每秒可处理数百万个事件，通过以下方式推动新运营模型的潜力：

- 识别和评估原本未被发现的问题
- 预测即将发生的失败并生成实时通知
- 筛选、汇总并关联数据，以描绘相关模式

这些功能不仅意味着问题可以得到积极解决，而且能够实现新的工作方式。例如，根据平均预期寿命，公司办公楼的空气过滤器通常每六个月更换一次，而不考虑情境因素（如花粉量）。通过使用主动维护计划，而非任意安排，使用基于传感器的数据和 ESP 可改善空气质量并减少浪费。

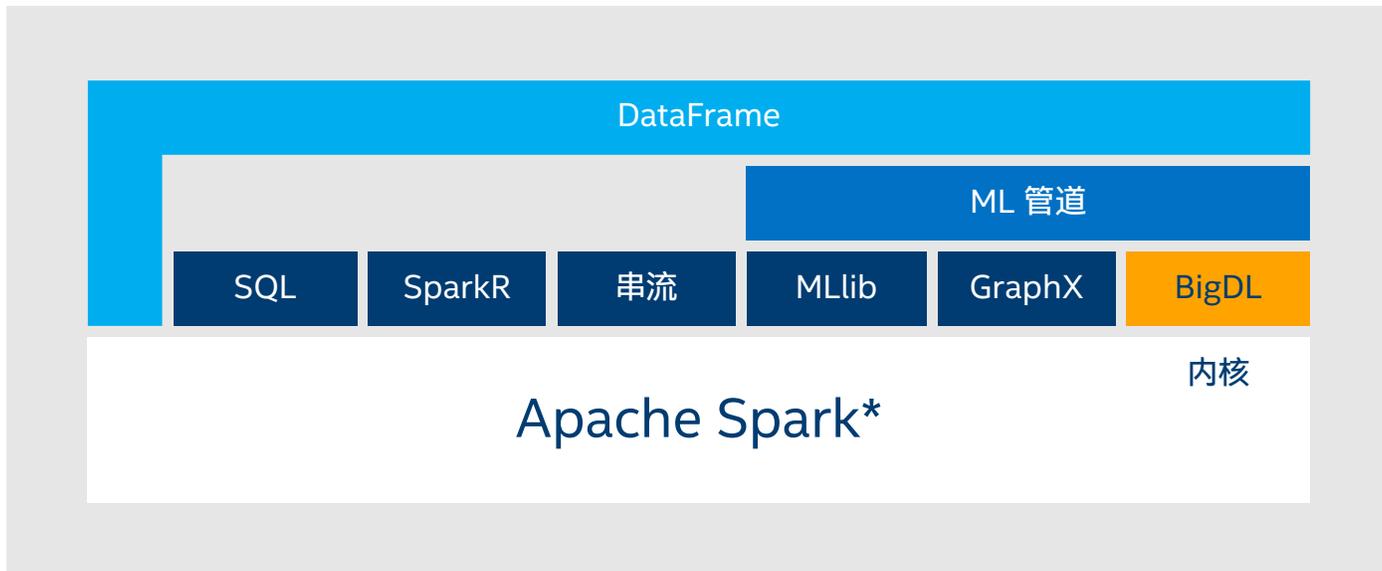


图 1. Apache Spark* 堆栈中的 BigDL

通过 SAP HANA 和内存数据库获取董事会层级的洞察

自 2005 年以来，英特尔和 SAP 一直保持合作，致力于提高 SAP 应用在英特尔® 架构上的性能表现。SAP HANA 平台通过将列式数据存储、大规模并行处理 (MPP) 和内存计算相结合，充分利用三者之优势。

SAP 的 Digital Boardroom 应用可提供下至所有业务部门、上至高层管理人员的实时洞察，在制造和物流、人力资源和采购、财务和销售等部门推动决策制定，确保提高利益相关者价值。该示例说明，数字化转型在整个企业中创造新的数据来源后，组织有机会从自上而下的数据流分析中受益。

随着数据流分析从例外技术场景转变为数据架构和基础设施的核心组成部分，企业可以全方位了解客户和自身。Gartner 将此称为公司“数字孪生”愿景⁵，该愿景的发展有利有弊：虽然“高质量”数据可以增加价值⁶，但根据一项估算，2016 年美国在无效数据上花费高达 3.1 万亿美元⁷。

使用 Apache Spark* 通过深度学习检测欺诈

近年来，Apache Spark 已彻底改变大数据处理。凭借开源内存架构，Spark 可以实时处理和分析批量数据与海量流式数据：高性能来自于执行内存处理的能力，而不是将数据保存到硬盘。

Spark 的一个核心优势在于，它能够处理整个分析流水线，并大规模训练机器和深度学习模型。针对深度学习部署，为了简化该工作流程并提高效率，英特尔提出了名为 BigDL 的开源计划，这是一个分布式库，可借助 Apache Spark 构建深度学习应用。

结合 BigDL，Apache Spark 支持数据流分析传输训练和推理算法。如图 1 所示，与用于深度学习的许多其他库不同，BigDL 衍生自 Apache Spark：结合 BigDL，深度学习应用可作为在现有 Spark 或 Hadoop* 集群上运行的标准 Spark 程序来编写。因此，BigDL 可加快价值实现速度，降低总体拥有成本 (TCO)，提高易用性⁸。

以金融服务为例，基于 Apache Spark 和 BigDL、由数据流分析支持的模型可以扩展到电子贸易之外，从而获得更深入的风险管理、监管合规和欺诈检测洞察。如果可以实时识别欺诈，一旦发生便加以预防，就可大幅降低成本并提高保护级别。

数据流分析应用案例

随着数据流分析的潜力增加，不同行业以全新方式采用该技术，从而为利益相关者交付重要价值，如以下示例所示。



临床医疗

实时分析可帮助监控患者安全、个性化患者结果、评估临床风险并减少患者再入院，从而提升医疗效率并改善患者体验。例如，远程监控和对临床事件的快速响应可显著改进慢性病患者护理质量。⁹



交通

物流公司可以使用传感器中的流式数据预测发动机故障，从而降低在路上进行修理的风险（和大量费用）¹⁰。同时，可以实时监控来自卫星和路旁系统的数据，并将该数据输入到集中仪表盘；或返回给车辆，以驱动高级驾驶员辅助系统。¹¹



零售库存和行为监控

数据流分析可改善实时库存跟踪，了解客户在店内或网上的行为方式，从而推动运营效率、提高销量、改善安全性并增强客户满意度。数据流分析还可以在基于上下文的数据（如市场趋势）中识别会影响业务性能的异常。¹²



呼叫中心分析和客户流失

通过使用呼叫中心日志以及来自传感器、在线互动、交互式语音应答系统和 IT 支持系统的数据，企业可以对客户行为、流失率预测及其他洞察建模。预测模型可帮助提高员工保留率，减少高额招聘开销。



网络流量监控和欺诈检测

网络安全问题是当今企业面临的最严重威胁之一，技术手段日益高超的黑客在不断寻找漏洞，窃取数据。具有机器学习

功能的数据流分析可帮助检测异常，及早发现攻击并发出警告。



金融欺诈检测和信用交易监控

数据流分析可聚合交易数据，用于改善业务状况并获取关于客户行为的洞察。机器学习模型可帮助预测欺诈性交易，加快反应时间，减少风险暴露。¹³



更灵活的保险模式

实时传输数据为保险公司打开了新功能和业务模式的大门，如“按驾驶付费”保险。健身数据也可用作健康保险模式和产品创新的输入，例如，“奖励”客户的锻炼活动。¹⁴



智能运营和供应链

支持分析的预测性维护可大幅减少服务干扰和成本，优化备件和原材料的供应链与前向容量规划。¹⁵ 数据流分析可加快生产与维护周期，提高质量，推动渐进式改进计划。¹⁶



家庭和企业自动化

智能仪表可根据占用情况来管理能源效率，节省量有时可达三分之一¹⁷，并降低家庭和企业的设备风险，例如预测便利店的冰箱故障。它们还可以将数据提供给二级变电站，从而确保能源可满足需求。¹⁸



IT 和业务运营

系统数据分析可改进 IT 响应能力和容量规划，并且可以直接馈送至持续交付和自动化管理流程中。来自设备的数据流分析能够检测并预测故障，在整个基础设施、运营和资源及业务流程中提高 IT 效率。

数据流分析的首要步骤

正如我们所看到的，数据流分析蕴含大量商机。那么，除了拥有更广泛的架构愿景，如何制定可实现数据流分析愿景的策略？根据不同数据流分析阶段的客户的合作经验，我们总结了以下指导准则。

1 在连接中兼顾安全性与质量

第一步是确定如何授权和连接设备，保证以正确且安全的方式报告信息，保持设备更新，让信息符合治理与合规性政策。这些属于 IT 领域，不是业务部门的首要考虑因素。上述问题同样关系到未来安全性，因为端点设备和传感器一旦安装就很难改变。

安全性是头号挑战，因为与您的网络连接的所有设备都可能成为攻击途径。如何安全地部署和维护设备？设备将如何识别自己并证明未被入侵？传感器数据需要保证安全性和加密性，仅在合适的时间对合适的人员可见。

随着时间推移，如何保证数据质量也需要考虑。例如，天气预报公司可能会运行数千个传感器，只有环境保持不变，这些传感器才能稳定地报告数据。然而，周围有可能多出一堵墙，或者传感器在未通知或不知情的情况下被挪动。如果传感器的环境发生变化，将影响所生成的数据，从而影响潜在洞察。通过 Wind River 技术^{*}，英特尔可以帮助保护前端设备数据以及网关与云端或数据中心之间的数据。

2 考虑混合云策略

对于希望测试数据流分析数据架构的组织，基于云的服务尤其有用。尽管您可以设计全面解决方案，但您也可以在云端以较小的规模实施解决方案、验证运行，查看使用情况、带来的挑战以及需要增强的领域。然后，您可以重新定义架构，并在云端逐步扩大使用范围。值得注意的是，一些行业（包括金融业）十分重视数据敏感度或者有监管需求，这给使用云服务增加了难度。

许多组织考虑采用混合策略，其中需要决定在云端和本地存储并处理哪些数据。一些组织希望将云端作为未来的数据中心，从头开始在云端构建全新的数据架构和分析基础设施。由于无需管理数据中心，他们可以专注于业务问题。规模较小的公司也可以从

分析即服务产品中受益，此类产品的部署速度更快，总体拥有成本 (TCO) 更低。

3 着眼大局，脚踏实地

好高骛远可能给组织带来重大风险。要避免这种情况，请挑选投资回报清晰的既定目标：这有助于在公司内开始建立数据流分析的价值，树立可信度并吸收专业知识。例如，在制造环境中，可以重点关注最容易发生故障的设备，而非监控整个工厂。

连接方面也存在相同的挑战，因为 IT 网络需要有效且安全地扩展，覆盖通常未安装计算设备的多个位置。在升级扩展之前，可以通过少量设备、传感器或网关对模型进行测试。需要考虑的额外领域包含运营管理开销：要建立需求，较为简单的做法是先小规模进行测试，这样就不会在全面部署后才发现资源不足。

4 决定并审核要流式传输的数据

考虑到当今世界的海量数据，您无法传输或处理所有内容。虽然不需要保留衍生的数据，但管理潜在瓶颈与数据可以催生洞察（有时甚至出乎意料）的事实需要兼顾。以 DNA 测序为例，起初一些数据被视为不必要，之后人们才发现它们是有用的，这意味着需要重新进行分析。随着机器和深度学习技术识别出数据项之间的新关系，产生改进后的新模型，被视为无关紧要的信息可能会变得非常重要。

模型和数据不断接受测试，以验证假设是否正确，确认结果的置信水平，尤其是数据可能发生变化的情况。新模型可推至边缘设备，以提供更好的结果，从而创造并生成更高质量、更准确的洞察。与此同时，采集尽可能多的经济可行数据十分重要。这一点需要与政策相平衡：例如，在银行业，如果面临诉讼，或者为了证明合规，需要展示什么数据？

5 计划改变业务

数据流分析不仅与收集、存储和分析数据有关，还涉及结果可实施化。向“大众”提供洞察是一回事，大众根据洞察采取行动是另一回事，何况他们不一定做好了准备。例如，某个维护部门可能会根据时间表来安排工作，而非执行被动或预测性维护。

如果您不根据获得的数据来调整工作，剩余步骤将毫无用处。那么，如何把这些结果反馈回给有权调整公司运作方式的人，他们将如何与客户互动或执行维护，以落实节省措施？这可能会改变劳动力的结构。

将数据流分析整合到企业时，必须与专家密切合作，专家了解新技术的构想并具有相关专业背景，知道如何融入您的阶段性转型方法。英特尔等企业也会提供帮助，以保证您实现短期和长期 ROI 目标。

英特尔对数据流分析的贡献

数据流分析需要一个面向未来的平台，该平台应该能够支持复杂的数据管理、高性能、低时延，并且能够大规模运行。正如从示例中看到的，通过将高效的硬件与高效的全新数据分析软件相结合，数据流分析的商机范围可拓展到全新领域。从硬件的角度来看，英特尔仍处于数据流分析发展的前沿：

- 英特尔® 至强® 可扩展处理器可实现出色的资源利用率和敏捷性，凭借时钟速度和英特尔® 技术提供出色的分析性能，无需依赖高内核数的处理器。
- 英特尔® 傲腾™ 技术以独特的方式将英特尔® 3D XPoint™ 内存媒体、英特尔® 内存和存储控制器、英特尔® 互连 IP 和英特尔® 软件组合在一起。对于需要大容量和快存储的工作负载，这些构建块可共同减少时延并加快系统运行速度。
- 英特尔® 内存驱动技术是软件定义的内存 (SDM) 技术，能够透明地扩展系统内存。通过与英特尔® 傲腾™ 固态硬盘 (SSD) 相结合，该技术能够将固态硬盘整合到内存子系统中，使固态硬盘看起来像 DRAM，从而扩大内存容量，突破 DRAM 的限制。

英特尔继续优化硬件、存储和软件，在吸纳、处理和交付数以 TB 计的海量流式数据过程中，兼顾性能与可扩展性。请关注此版块，我们会陆续发布关于如何使用英特尔® 技术实现数据流分析目标的公告。

为进一步加快选择和部署当今广泛分析场景所需的硬件与软件的过程，英特尔® 精选解决方案提供大量灵活、易于部署且可扩展的解决方案，这些解决方案专门针对性能进行了优化。

通过数据流分析帮助您的组织做好准备，获取实时洞察

现代数据的爆发式增长，加上基于实时数据的决策需求增加，使得由机器学习支撑的实时数据流分析成为公司保持竞争力不可或缺的利器。决策信心是总目标，因为提高预测准确性可推动新的业务模式和商机，从而提高竞争力并获得致胜优势。探索不断扩大的[英特尔® 精选解决方案库](#)。

开始跟踪数据流分析也许是您组织可以制定的最重要业务决策之一。我们建议您浏览以下内容，继续了解数据流分析可以带来的优势：

解决方案简介：[边缘商业智能——来自英特尔和 SAP 的端到端物联网产品组合](#)

解决方案简介：[智能机器的智能分析](#)

产品简介：[SAP HANA* 实时分析软件和英特尔® 技术](#)

产品简介：[加速获取洞察，打造难忘的购物体验](#)

英特尔® AI Academy：[BigDL：一种面向 Apache Spark* 的分布式深度学习库](#)

[访问 intel.cn/analytics](https://www.intel.cn/analytics)，探索英特尔的高级分析解决方案

解决方案提供商:



- ¹ Forrester Wave™: 数据流分析, 2017 年第 3 季度 <https://www.forrester.com/report/The+Forrester+Wave+Streaming+Analytics+Q3+2017/-/E-RES136545>
- ² 引领物联网——Gartner 关于如何跻身互联世界前列的洞察, Mark Hung 编辑, Gartner Research, 2017 年, https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf
- ³ 市场分析报告, Apache Spark 市场预测 2017-2020, 2017 年 5 月, <https://www.marketanalysis.com/?p=159>
- ⁴ 机器智能的简单经济学, Ajay Agrawal, Joshua Gans, Avi Goldfarb <https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence>
- ⁵ 准备好应对数字孪生带来的影响, Gartner, <https://www.gartner.com/smarterwithgartner/prepare-for-the-impact-of-digital-twins/>
- ⁶ 实时数据如何促进赢利。 <http://cioinsight.com/it-management/innovation/slideshows/how-real-time-data-boosts-the-bottom-line.html>
- ⁷ 无效数据的成本。 <http://information-age.com/cost-bad-data-123465972>
- ⁸ 在搭载 BigDL 的 Apache Spark 上加速深度学习工作负载, 英特尔, <https://itpeernetwork.intel.com/accelerate-deep-learning-apache-spark-bigdl-2/>
- ⁹ 使用实时数据流分析解决方案提升数据价值, 英特尔, <https://www.intel.cn/content/dam/www/public/cn/zh/documents/solution-briefs/real-time-streaming-analytics-brief.pdf>
- ¹⁰ 推动支持物联网的预测性维护, Cloudera, <https://vision.cloudera.com/driving-iot-enabled-predictive-maintenance/>
- ¹¹ 使用物联网、Kafka 和 Spark Streaming 监测流量数据, <https://www.infoq.com/articles/traffic-data-monitoring-iot-kafka-and-spark-streaming>
- ¹² Ebay 的分析和商业智能, <https://www.intel.cn/content/www/cn/zh/processors/xeon/business-intelligence-analytics-ebay-signals-video>
- ¹³ 针对保险行业利用大数据分析, StackIQ, https://cdn2.hubspot.net/hub/173001/file-18488782-pdf/docs/stackiq_insuranceind_wpp_f.pdf
- ¹⁴ Fitbit 联合创始人称, UnitedHealthcare 与 Fitbit 分发高达 1,500 美元以鼓励用户使用设备, <https://www.cnbc.com/2017/01/05/unitedhealthcare-and-fitbit-to-pay-users-up-to-1500-to-use-devices.html>
- ¹⁵ 英特尔与 Software AG 快速提供可行分析结果, <https://www.intel.cn/content/www/cn/zh/processors/xeon/softwareag-actionable-analytics-video.html>
- ¹⁶ IT@Intel: 大数据在英特尔智能工厂中的应用, <https://www.intel.cn/content/www/cn/zh/it-management/intel-it-best-practices/using-big-data-in-manufacturing-at-intels-smart-factories-paper.html>
- ¹⁷ 审核机构建筑中应用的占用率感应系统和占用率建模方法, Junjing Yang, 新加坡国立大学 https://www.researchgate.net/publication/286455523_Review_of_occupancy_sensing_systems_and_occupancy_modeling_methodologies_for_the_application_in_institutional_buildings
- ¹⁸ 电力企业数字化转型——业务转型通过高级分析功能推动, 英特尔, <https://www.intel.cn/content/dam/www/public/cn/zh/documents/white-papers/iot-digitizing-power-utilities-paper.pdf>

英特尔技术特性和优势取决于系统配置, 并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变化。没有计算机系统是绝对安全的。更多信息, 请见 intel.cn, 或从原始设备制造商或零售商处获得更多信息。

此处提供的所有信息可在不通知的情况下随时发生变更。关于英特尔最新的产品规格和路线图, 请联系您的英特尔代表。

性能测试中使用的软件和工作负荷可能仅在英特尔微处理器上进行了性能优化。诸如 SYSmark 和 MobileMark 等测试均系基于特定计算机系统、硬件、软件、操作系统及功能。上述任何要素的变动都有可能导致测试结果的变化。请参考其他信息及性能测试 (包括结合其他产品使用时的运行性能) 以对目标产品进行全面评估。

关于性能和基准测试程序结果的更多信息, 请访问 <http://www.intel.cn/benchmarks>

英特尔、至强、傲腾、3D XPoint 和英特尔标识是英特尔公司在美国和 / 或其他国家的商标。

* 其他的名称和品牌可能是其他所有者的资产。

© 英特尔公司版权所有

0219/JS/CAT/PDF