

IBM InfoSphere Streams

实时分析处理, 获得切实有效的洞察



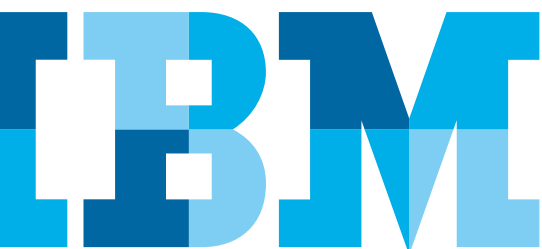
重点内容

- 对运转中的数据执行先进的实时分析
 - 对于来自数千个来源、大容量和种类繁杂的结构化和非结构化流数据, 可以迅速吸收、相互关联并且进行分析
 - 在数据到达时, 可以实时预测和发现
 - 采用拖放操作式开发工具, 轻松实现数据可视化
 - 检测并且立即响应关键事件
 - 通过认知计算了解并且更新模型, 以便用于未来分析和趋势预测
-

每一天, 消费者和企业都在以几年前看似前所未有的速度生成数据。例如, 在普通的一天内, AT&T通过其网络可以传输大约30PB的数据。¹ 汽车每小时可以生成1.3GB的传感器数据, 按照每年6000万辆汽车的产量, 估计仅汽车产生的数据便会达到103EB。² 推特网站每天处理3.4亿条消息, 其中大部分消息都来自移动用户。³ 到2016年, 年互联网流量预计将达到1.3ZB,⁴ 80%的数据将会是非结构化格式。

面对数量不断增长、瞬息万变的数据, 组织迫切需要做出明智、实时的商业决策, 并且在竞争中保持领先地位。然而, 精明的组织快速意识到, 通过更新自身的数据挖掘方法, 支持非传统的非结构化数据源 (例如音频、视频和电子邮件等), 它们可以扩展现有系统的价值, 并且获得显著的商业优势。这可以帮助组织更加快速地应对客户情绪的变化, 揭示全新的市场机遇, 并且推出符合最新趋势的开创性新产品。

作为这种新一代分析处理方法的关键推动要素, IBM® InfoSphere® Streams可以提供最新的计算平台, 帮助组织将日益增长的数据转化为切实有效的信息和商业洞察。InfoSphere Streams是IBM大数据平台的关键组件, 可以交付高度可扩展、灵敏的软件基础架构, 以前所未有的容量和速度对丰富多样的相关和非相关数据类型 (来自数千种的实时来源) 执行运转分析。凭借InfoSphere Streams, 组织可以一直采集数据, 并且及时分析这些关键业务数据。



信息处理的全新范式

InfoSphere Streams是IBM研究院与美国政府共同合作完成的开拓性成果。该软件可以提供开发平台和运行环境，支持组织开发和执行各种应用，以便对海量的持续数据流进行吸收、过滤、分析和相互关联。这些数据流产生自众多结构化和非结构化数据源，可以包括大量丰富的数字信息，例如：

- 文本文件、电子表格、图片、视频和音频录音
- 电子邮件、聊天和即时消息、网络流量、博客和社交网站
- 金融交易、客户关系记录、电话使用记录和日志文件
- 卫星、GPS跟踪、智能设备和网络流量传感器的数据


InfoSphere Streams将这些异构数据类型融合到单一的计算平台之中，这个平台支持强大的数据分析，具有出色的性能和响应时间。

简洁强大的实时分析处理功能

InfoSphere Streams能够解决以下重大的数据挑战：分析大量运转中的数据。这种分析持续生成数据的能力通常对组织而言至关重要，它们必须能够实时应对市场警报或事件，或者有时组织必须过滤大量数据并且识别丰富、高价值信息，以便之后将这些数据转移到数据仓库或Apache Hadoop系统。

为了经济高效地分析流数据，InfoSphere Streams采用了尖端性技术，例如独有的流处理语言(SPL)。InfoSphere Streams应用可以通过C、C++或Java应用加以扩展，重复利用现有逻辑，并且加速获得回报。此外，“入门”(Getting Started)功能仅需几次点击操作，可以简化应用实例的安装、构建、配置和管理。拖放式可视开发有助于减少学习时间并且加速应用开发。

InfoSphere Streams 的关键组件

综合性开发工具	横向扩展架构	成熟的分析工具包和加速器
		
<ul style="list-style-type: none">• Eclipse IDE• Web控制台• 拖放式编辑器• 实例图表• InfoSphere Streams可视化• InfoSphere Streams调试器	<ul style="list-style-type: none">• 集群式运行，近乎无限的容量• Red Hat Enterprise Linux, CentOS, 面向x86和POWER硬件的SUSE Linux Enterprise• 支持InfiniBand和以太网	<ul style="list-style-type: none">• CEP、数据库、金融、IBM InfoSphere Data Explorer、IBM InfoSphere DataStage®、地理空间、互联网、消息队列，包括MQ、MQTT和JMS、挖掘、R、IBM SPSS®、文本、时间序列工具包• 应用加速器：电信事件数据分析、社交数据分析

企业集成: 扩展InfoSphere Streams的能力

InfoSphere Streams可以与IBM大数据平台的其它产品实现即时集成并且具有内置XML支持,可以访问一系列丰富多样的数据来源和系统。

- **IBM InfoSphere BigInsights™**可以让您将流数据存储在企业级的Hadoop环境,以供进一步分析或作为历史记录保存。InfoSphere Streams和InfoSphere BigInsights采用相同的高级文本分析功能,可以简化运转数据和静止数据的自然语言处理应用。除此之外,InfoSphere BigInsights可以用于通过背景信息增强流数据来源,用户可以在InfoSphere BigInsights控制台中实施InfoSphere Streams数据的可视化。
- **IBM InfoSphere Data Explorer**支持用户在InfoSphere Data Explorer CXO仪表板中实施InfoSphere Streams数据的可视化,并且将流数据添加至InfoSphere Data Explorer索引。
- **IBM InfoSphere DataStage**帮助用户利用基础架构其它部件的背景信息和源数据执行深度分析,并且获得更多洞察。
- **XML支持**允许开发者融合更加广泛的传统式和非传统式数据。
- **数据库和储存库支持**让用户可以利用结构化数据源融合并且组合分析流数据和非结构化数据,这些数据源包括IBM DB2、IBM IDS、IBM SolidDB、Microsoft SQL Servicer、Oracle Database、HP Vertica、IBM PureData Systems、Teradata和更多来源。
- **IBM WebSphere MQ、IBM MessageSight和IBM Integration Bus**的消息队列可以提供一种灵活的系统,实现高效的消息和数据传输。

横向扩展架构

InfoSphere Streams可以帮助组织扩展当前IT投资的价值,无需大量的基础架构检修。它可以将单一节点扩展到实质上无限数量的节点,处理任意容量的数据——从TB级到ZB级。InfoSphere Streams提供一种集群式运行环境,每秒可以轻松处理数百万的事件,延迟时间仅有微秒。近乎为零的延迟时间可以获得切实有效的成果。

先进的编译器融合应用的各个组成部分,从而提高性能。然后将融合的部分进行分配,在多种硬件节点之间运行。该编译器还可以支持多种高速传输,包括以太网和InfiniBand。当前应用可以与新应用实现动态扩展,读取相同的数据流,以便进一步充分利用当前投资。

基于web的管理控制台可以简化运行环境和应用的配置和管理,其中包括自动布置各种功能和部署应用组件等。它可以监控应用及其独立元素的状态和性能指标,有助于确保组织达到服务水平协议要求。

InfoSphere Streams包括能够提高可用性和冗余性的多种功能。管理员可以添加或移除集群的处理器节点。无缝添加或移除集群之间的节点,这允许管理员无需关闭InfoSphere Streams应用便可以执行必要的维护操作。这种功能有助于提高环境的整体可用性。此外,当控制系统的管理节点发生故障时,应用节点可以继续运行InfoSphere Streams应用。之后管理员可以重新启动管理节点。

流数据的编程操作可以采用冗余应用组件,轻松创建高度可用的应用。主机池的动态分配以及运行集群的更多定位和隔离限制可以用来隔离冗余应用组件,从而隔离运行时间的各个节点。

面向灵敏开发环境的综合工具

凭借InfoSphere Streams, 开发者可以轻松构建应用, 实时分析大容量、高速度和丰富种类的数据。

InfoSphere Streams Studio是一种基于Eclipse的集成开发环境(IDE), 支持编辑器、向导、应用流图表和运行监控的迅速开发, 能够简化构建和管理InfoSphere Streams应用的流程。这些工具旨在帮助组织通过成熟的实时数据分析, 快速和轻松地扩展应用的各种功能:

- **InfoSphere Streams Debugger**可以提供一种交互式调试器, 用于调试SPL应用, 以便检验SPL出入的数据流。
- **拖放式图形编辑器**允许用户构建应用, 同时可以自动同步图形和SPL源代码视图。这种循环式流程可以帮助开发者创建一种带有文本或图形编辑器的应用, 并且日后可以使用其它编辑器更新。
- **实例图表**, 在InfoSphere Streams Studio和管理控制台中均可使用, 它可以提供应用健康和各项指标的可视监控功能, 让用户采用自定义视图快速识别问题。
- **最新的InfoSphere Streams数据可视化功能**允许用户动态添加新视图, 通过即时提供的各种图表运行应用。

InfoSphere Streams:系统要求

硬件要求	<ul style="list-style-type: none"> • Intel/AMD x86架构(64位)或 IBM POWER7@架构系统 (最低500 MB内存) • 2GB, 内存运行简单应用
软件要求	<ul style="list-style-type: none"> • Red Hat Enterprise Linux (RHEL) Version 5.6 或更高版本, x86架构硬件要求Version 6.1或更高版本 • IBM POWER7架构硬件要求 RHEL Version 6.2或更高版本 • CentOS Version V5.6或更高版本, x86架构硬件要求6.1或更高版本 • x86架构硬件要求SUSE Linux Enterprise Server (SLES) V11.2或更高版本 (SLES不支持 Security Enhanced Linux) • Eclipse Platform SDK V4.2.2 • x86或IBM POWER7架构硬件要求 IBM Java SE V6或V7 SDK • x86架构硬件要求Oracle Java SE v6.0.x 或V7.0 SDK • Mozilla Firefox v10或更高版本 • Microsoft Internet Explorer 9.0或更高版本

工具包和加速器的成熟分析功能

InfoSphere Streams包含集成工具包和示例应用,可以针对特定行业或功能开发各种解决方案。这些包括数据挖掘工具包、金融服务工具包和标准工具包,包含一系列最为常用的操作工具,从而便于和现有基础架构和应用进行集成。

例如,一家金融服务机构可以利用InfoSphere Streams将输入的股市报价与相关交易商的行业数据或信息相结合,一起保留在内部数据库中。将背景信息添加到股市价格信息,可以帮助企业获得远比以往更加复杂的分析和自定义报警。除此之外,InfoSphere Streams有助于实时确定数据库应当存储哪一类数据,以便帮助日后分析、降低存储和管理成本。

InfoSphere Streams的其它工具包还包括:

- **Complex Event Processing (CEP)**, 提供丰富的高性能分析功能,采用各种模式检测简单事件流构成的复合事件。现有应用可以轻松集成,并且迁移到InfoSphere Streams环境中,充分利用高可扩展性的优势,基于相同的硬件每秒处理的事件超出10倍。
- **InfoSphere Data Explorer**, 为InfoSphere Streams带来可视化和存储功能,支持用户将可视化工具从InfoSphere Streams传递到InfoSphere Data Explorer,或者将数据传递到InfoSphere Data Explorer,以便进行索引和日后检索。
- **InfoSphere DataStage**允许用户执行深度数据分析,将其作为信息集成流的组成部分,并且迅速查看结果。InfoSphere Streams工具包可以提供适配器,在InfoSphere Streams和InfoSphere DataStage之间交换数据,从而实现实时分析处理。

- **IBM SPSS**充分发挥了IBM SPSS Modeler的优势,可以开发并构建可预测模型,然后利用SPSS Scoring Operator部署这些模型。SPSS模型可以在InfoSphere Streams中实现动态刷新,无需暂停InfoSphere Streams应用。
- **地理空间**通常用于运输和电信行业。它根据高性能分析和处理地理空间数据,可以支持基于地理位置的营销活动。
- **消息队列**可以让InfoSphere Streams接收或发送IBM WebSphere® MQ、IBM MessageSight和Java Messaging System (JMS)产品的往来数据。
- **时间序列**是一组丰富多样的工具,包括数据生成(合成或提取)、预处理(准备和调节)、分析(统计数据、关联、分解和转换)和认知计算以及建模(预测、回归分析和跟踪)。

InfoSphere Streams还包括应用加速器,有助于加速获得回报,充分利用特定解决方案或使用案例的最佳实践成果。它们可以提供商业逻辑、数据处理和UI/可视化功能、分析操作器或示例应用,可以作为新应用的设计模板使用。

IBM Accelerator for Social Data Analytics可以迅速实施方案,分析大量各种类型的社交媒体数据,并且实时处理。这可以帮助组织获得关键的洞察,开展有效、极具针对性的市场营销活动,并且做出及时的产品和营销决策。它还可以促进产生各种洞察,支持各种竞争性优势的智能工作,制定客户保留和新客户采购计划。

IBM Accelerator for Telecommunications Event Data Analytics 可以为电信公司的呼叫和事件详细记录提供转换和分析功能,它有助于实施实时收入保障和欺诈检测。这可以帮助电信公司根据服务、供应商和业务部门增强关于计费的洞察。它还可以支持各种特色服务、增强客户忠诚度、减少客户投诉、提供集成服务、实现计费个性化服务并且改善客户体验。

InfoSphere Streams V3.2: 特点和增强功能

最新版的InfoSphere Streams包括一系列丰富的新特点和增强功能,可以提高可用性,改进企业就绪能力,并且精简企业集成。

提高可用性

InfoSphere Streams使用简单,它集合了一套强大的开发工具、复杂的分析工具包和加速器,这些都便于开发各类应用,并且有效运行丰富多样的活跃数据。InfoSphere Streams v3.2可以提供面向Windows和Linux的远程开发、用户定义的并行机制和Java简化功能,从而对开发平台加以扩展。它比其它平台更加便于安装、开发、维护、监控和扩展。您如今可以在Windows桌面机中通过本机方式使用Streams Studio来开发应用。Streams Studio可以与基于Linux的集群实例无缝交互,测试并部署各类应用。

InfoSphere Streams v3.2可以提供用户定义的并行机制。开发者可以无数次地复制应用的任意部分,执行并行处理,使用循环或哈希算法分解流记录。并行机制的数量可以在编译时间指定。开发者还可以强制并行通道运行不同的主机,以提高应用性能。最终,Java操作器模型简化工具意味着开发者可以更加轻松地维护实施源代码和模型。

增强企业就绪能力

尽管已经具备企业就绪性,InfoSphere Streams仍然可以提供横向扩展架构和近乎无限容量的集群式运行。InfoSphere Streams v3.2甚至更加便于维持并运行,同时可以动态添加规则和界面。

InfoSphere Streams v3.2采用应用部分的代表状态传输位置(REST)界面,可以支持用户通过编程方式访问应用数据。在InfoSphere Streams v3.2中提供IBM运行决策管理(ODM),可将ODM规则的实施作为应用的组成部分。用户在运行时可以创建、添加、输入并更新ODM规则,无需重启应用。

精简企业集成

InfoSphere Streams可以兼容各种软件和工具。在InfoSphere Streams 3.2中,新的适配器和操作器可以进一步扩展InfoSphere Streams集成,并且更加快速和轻松地实施集成。例如,现在它可以支持消息队列遥测传输(MQTT)适配器。MQTT是一种机器对机器(M2M)物联网(IoT)连接协议,可以作为一种极为轻型的发布和订阅消息传输工具。它可以用于连接需要小型占用空间的远程位置或网络带宽十分宝贵的情况(例如移动设备或汽车车载应用)。InfoSphere Streams v3.2还可以添加BigInsights BigSQL和HP Vertica支持功能。

改变大数据分析的形势

凭借InfoSphere Streams, 众多组织可以查看正在发生的事件和趋势, 从而可以主动应对。通过结合各种难以处理、大量且快速运行的数据流, 该软件已经为各行各业的组织带来了丰厚的真实回报。

例如, 一家通信公司当前正在使用InfoSphere Streams处理每秒超过20万次的详细呼叫记录, 同时还在生成这类数据的摘要。在此之前, 这个摘要流程需要至少12个小时才能完成, 但现在几乎可以即时提供摘要, 该公司能够基于实时信息做出关键的决策。这允许他们根据当前的地理位置, 针对高价值客户推出个性化营销产品, 尽最大可能影响客户。

在金融服务领域, 基于InfoSphere Streams的应用可以每秒分析并关联五百万的市场消息, 执行算法交易, 平均延迟仅为几微秒, 为企业提供灵敏性, 可以比竞争对手更快地充分利用市场趋势。在医疗保健领域, 可以利用InfoSphere Streams分析新生儿重症监护室内多种传感器的信息, 目标在于提前24小时检测患者医疗条件的变化——这是成功治疗病人的一项决定性前沿技术。

欲了解更多详细信息

任何基于实时信息、需要即时、准确分析和业务决策的组织都可以通过InfoSphere Streams获得真正的回报。欲了解关于InfoSphere Streams以及它如何帮助企业获得竞争优势的更多详细信息, 敬请访问以下网址:

ibm.com/software/data/infosphere/streams

立即了解! 实时分析处理功能唾手可得: 隆重推出IBM InfoSphere Streams快速入门版

免费下载、快速入门。为了帮助客户快速获得运转数据的实时处理功能, IBM已经发布了“快速入门”计划, 它可以帮助用户快速获得市场领先的大数据技术。通过免费下载, 这款市场领先的企业级大数据平台可以在其独有的非生产环境中为用户提供数据流计算功能。毫无数据或时间的限制。[阅读关于InfoSphere Streams快速入门版网站的更多详细信息。](#)



© 版权所有IBM Corporation 2014

国际商业机器中国有限公司
北京市朝阳区北四环中路27号
盘古大观写字楼
邮编: 100101

在中国印刷
2014年9月

IBM、IBM徽标、ibm.com、WebSphere、DataPower、DB2、Tivoli、Rational、MQSeries和System z是International Business Machines Corporation在美国和/或其他国家或地区的商标。如果这些商标及其他 IBM 商标术语在本资料中第一次出现时标有商标符号(®或™),这些符号表明在本出版物出版时,由IBM拥有的美国注册商标或普通法商标。此类商标还可能是其他国家或地区的注册商标或普通法商标。IBM商标的最新列表位于以下Web站点的“版权与商标信息”中:

ibm.com/legal/copytrade.shtml

Java和所有基于Java的商标和徽标是Oracle和/或其下属公司的商标或注册商标。

本文档截至其最初出版日期,IBM可能随时更改它。IBM并不将在其运营的所有国家或地区提供这些产品或服务。

本文档内的信息“按现状”提供,不附有任何种类(不论是明示的还是暗示的)的保证,包括适销性保证、具体的适用性和不侵权的保证或条件。IBM产品根据其所属的合同的条款条件获得保证。



请回收利用