

IBM InfoSphere Streams

2.0 版即時分析資訊，值得仰賴



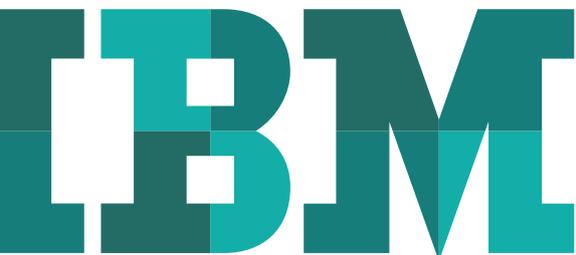
產品特色

- 針對不斷變動的資料，執行即時的複合分析
 - 持續分析大量、多變的結構化與非結構化資料
 - 幾乎各種規模的運算叢集，都能順利擴充現有的應用程式，增加資料採礦功能
-

消費者與企業產每天以前所未見的速度產生資料，幾年前這根本是天方夜譚。零售業龍頭 Walmart 每小時處理超過 100 萬筆的客戶交易，輸入超過 2.5 PB 的資料至資料庫，約為國會圖書館書籍資料總和的 167 倍。¹ 紐約證券交易所從開盤到收盤，每天約擷取 1 TB 的交易資訊，² Google 每天約處理 24 PB 的資料，³ BBC iPlayer 每個月的資料傳輸量達 7 PB，⁴ 每 20 分鐘，使用者就上傳 270 萬張照片至 Facebook。⁵

今日資訊數量龐大、種類多變，遠非過去能夠比擬，企業必須持續奮戰，根據不斷變動的資料，作出明智、即時的商業決策。聰明的公司很快瞭解，更新資料採礦方法，利用非傳統、非結構化的資料來源（如音訊、視訊與電子郵件），便能提升現有系統的價值，創造更大的商業優勢，根據客戶評價迅速回應、找出新商機，導入突破新產品。

IBM® InfoSphere® Streams 是新一代分析處理方法，提供先進的運算平台，協助企業分析瞭解迅速成長的資料，產生有效的商業資訊，據以行動。在 IBM 海量資料 (Big Data) 分析平台中，IBM® InfoSphere® Streams 是一大要件，提供靈活的軟體基礎架構，可隨需求調整，讓企業以前所未有的速度與數量，從數千個資料來源隨時分析各種不斷變動的關聯與非關聯資料。有了 InfoSphere Streams，企業能隨時即刻擷取重要的商業資料，並加以因應。



「自 2003 年起，美國政府便與 IBM 研究中心合作，研發嶄新的資料分析方法，執行快速的異質資料流 (heterogeneous data streams) 複合分析，IBM InfoSphere Streams 就是目前的技術成果。此專案非常成功，美國政府將進一步部署，讓其他機構的未來專案能更順利。」

— 美國政府

資訊處理新典範

InfoSphere Streams 是 IBM 研究中心與美國政府合作的創新成果，提供開發平台與執行階段環境，讓客戶開發與執行應用程式，針對大量連續資料串流，吸收、過濾、分析及建立關聯性。這些資料串流可能來自結構化或非結構化的資料來源，包括各種各樣的數位資訊，如：

- 文字檔案、試算表、影像、影音記錄
- 電子郵件、會談與即時通訊、網路流量、部落格與社群網站
- 金融交易、客服紀錄、警方掃描器資料、系統與應用程式日誌
- 衛星資料、GPS 資料、感應器日誌、識別證刷卡與存取記錄
- InfoSphere Streams 將這些異質資料融合至強大的運算平台，以優異的效能與出色的回應時間，執行複合分析。

平台可調整，從單一伺服器到無數節點均可，並提供各式各樣的分析運算器 (operators) 與一系列的先進開發工具，可整合現有基礎架構及應用程式，企業可善用 InfoSphere Streams 技術，擴展目前的 IT 投資，在熟悉的應用程式中融入全新層次的商業智慧。

舉例來說，金融服務企業可運用 InfoSphere Streams，結合接收的股市報價與內部資料庫的資料（如來自產業或相關交易員的資料），從環境脈絡去瞭解股價變動，並達到更複雜層級的分析與自訂警示。此外，InfoSphere Streams 能即時判斷接收的資料哪些部分最為相關，可儲存在資料庫，供日後分析使用，協助降低儲存與管理成本。

進階工具打造靈活開發環境

InfoSphere Streams 讓開發人員與管理者在應用程式與工作流程中，輕鬆建置出色的資料採礦功能。InfoSphere Streams 具備整合式工具箱與範例應用程式，包括 Data Mining Toolkit、Financial Services Toolkit 與 Standard Toolkit，涵蓋各式各樣常用的運算器 (operators)，能加速開發特定產業或部門的解決方案。

此外，Streams Studio（一種 Eclipse 架構的整合開發環境）支援 InfoSphere Streams 應用程式開發，包括編輯、精靈、應用程式結構圖形與執行階段監視等功能。IBM 開發串流應用程式經驗豐富，才能打造出上述方案，準確分析資料，協助客戶迅速輕鬆地擴充應用程式功能。

海量資料 (Big Data) 分析改頭換面

有了 InfoSphere Streams，企業能即時掌握事件與趨勢，而非只能亡羊補牢。這項技術能洞察龐大、快速變化資料串流的意義，已在眾多產業取得重大斬獲。

舉例來說，一間通訊公司目前使用 InfoSphere Streams，每秒可處理超過 20 萬則通話的詳細記錄，同時建立資料摘要；過去建立摘要要花 12 個小時，現在幾乎能即時取得，讓該公司根據即時通話資訊，作出關鍵決策，影響業務。

在金融服務業，InfoSphere Streams 運用程式每秒分析超過 500 萬筆市場訊息並建立關聯性，執行演算法選項交易平均只有 30 微秒的延遲，讓客戶比同業更能迅速、敏捷地掌握市場趨勢。在醫療保健業，InfoSphere Streams 可用來分析新生兒加護病房多重感應器的資訊，最快能比過去早 24 小時偵測到病患健康狀況的變動，這對治療疾病非常關鍵。

若企業需要即時且精確的分析，並根據即時資訊制訂商業決策，InfoSphere Streams 是您最佳選擇。如欲進一步瞭解 InfoSphere Streams 如何協助貴公司取得競爭優勢，請造訪 ibm.com/software/data/infosphere/streams

「因為 Streams Processing Language 快速靈活，現在 TerraEchos 開發人員研發應用程式的時間比過去快 45%。」

—Alex Philp, CEO, TerraEchos

InfoSphere Streams V2.0: 功能與加強功能

- 可重用現有 Java、C++ 程式碼與 Predictive Model Markup Language (PMML) 模組
- IBM WebSphere® MQ Low Latency Messaging (LLM) 傳輸技術，支援 InfiniBand
- 支援 IPv6 網路通訊
- Streams Studio: 基於 Eclipse 3.6.2 的整合開發環境 (IDE)，快速開發、測試與偵錯串流應用程式，Streams Live Graph 以圖形顯示執行階段叢集上進行的工作與工作元件
- Streams Runtime: 單一伺服器或伺服器叢集，叢集規模不受限制；高可用性功能，包括偵測故障程序元素、重新配置、重啟與選擇恢復狀態
- 工具箱、配接卡與範例：
 - 標準工具箱，具備關聯式、檔案、通訊與公用程式運算器
 - 網際網路工具箱，包含 HTTP、HTTPS、FTP、FTPS、RSS 與檔案來源運算器
 - 資料庫工具箱，包含 ODBC 驅動程式與高速 IBM solidDB® 串流擴充驅動程式
 - PMML Scoring 的資料採礦工具箱
 - 財務工具箱，包含 IBM WebSphere Front Office V3.0.2.1、Financial Information eXchange (FIX) 與 QuantLib Library 功能，運算股市選擇權衍生性商品價值
 - 50 種以上的範例應用程式及 Streams Processing Language 範例構件

InfoSphere Streams: 系統需求

執行階段作業環境

- 支援的作業系統：Red Hat Enterprise Linux (RHEL) Version 5.3、RHEL Version 5.4 與 RHEL Version 5.5；亦支援 Security Enhanced Linux (SELinux) 功能
- IBM Java SE v6.0-9.0 SDK
- Oracle Java SE v6.0.x SDK
- 支援的硬體：Intel/AMD x86 架構（32 或 64 位元）
-

開發作業環境

- 支援的作業系統：RHEL Version 5.3 與 RHEL Version 5.4
- Eclipse Platform SDK v3.6.2
- IBM Java SE v6.0-9.0 SDK
- Oracle Java SE v6.0.x SDK
- 支援的硬體：Intel/AMD x86 架構（32 或 64 位元）
-

支援的資料庫環境

- IBM DB2® for Linux、UNIX 與 Windows 9.5、9.7 或以上
- IBM Informix® v11.5
- IBM Netezza® Twinfin®
- IBM solidDB V6.5
- Oracle 11gR2
- Microsoft SQL Server 2008
- MySQL v5.1
- 支援的瀏覽器環境
- Firefox 3
- Microsoft Internet Explorer 6、7 與 8



© Copyright IBM Corporation 2011

IBM 軟體

技術諮詢熱線：0800-000-700

台北市松仁路 7 號 3 樓

美國

台灣印製

2012 年 3 月

版權所有

IBM、IBM 標誌、ibm.com 及 InfoSphere 是國際商業機器股份有限公司在美國或其他國家的商標或註冊商標。若上述及其他 IBM 註冊名稱在本文首次出現時附有商標符號 (® 或 TM)，代表這類商標在本文發布時已於美國註冊，或為 IBM 擁有的普通法商標。這類商標可能已在其他國家註冊或屬於普通法商標。IBM 最新的商標清單，請造訪 IBM 網站的「版權及商標資訊」，網址為：ibm.com/legal/copytrade.shtml

Netezza 和 Twinfin 是 IBM 旗下公司 Netezza Corporation 的商標或註冊商標。

Intel、Intel 標誌、Intel Inside、Intel Inside 標誌、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美國及其他國家的商標或註冊商標。

Java 和所有以 Java 為基礎的商標和標誌是 Oracle 及/或其子公司的商標或註冊商標。

Linux 是 Linus Torvalds 在美國及/或其他國家的註冊商標。

Microsoft、Windows、Windows NT 和 Windows 標誌為 Microsoft Corporation 在美國及/或其他國家的商標。

UNIX 係 The Open Group 在美國及/或其他國家的註冊商標。

其他公司、產品或服務名稱可能是其所屬公司的商標或服務標誌。

1 「Data, data everywhere」經濟學人。2010 年 2 月 25 日。

www.economist.com/node/15557443?story_id=15557443

2 Henschen, Doug. 「New York Stock Exchange Ticks on Data Warehouse Appliances」Information Week。2008 年 5 月 16 日。

www.informationweek.com/news/software/bi/207800705

3 Dean, Jeffrey and Ghemawat, Sanjay. 「MapReduce: simplified data processing on large clusters」ACM 通訊。Volume 51, Issue 1, January 2008。 <http://portal.acm.org/citation.cfm?doid=1327452.1327492>

4 Lanxon, Nate. 「iPlayer uncovered: What powers the BBC's epic creation?」Crave/CNET UK。
<http://crave.cnet.co.uk/software/iplayer-uncovered-what-powers-the-bbcs-epic-creation-49302215>

5 NW Linux: Linux and Technology News。2010 年 12 月 31 日。

<http://nwlinux.com/facebook-data-and-post-statistics-over-20-minutes-time>



請回收