

## IBM InfoSphere DataStage

### 企業級的資料整合、轉換與遞送



#### 產品特色

- 世界級的效能、生產力、協同作業與部署選項
- 高度直覺式、由上而下、隨想隨做的設計介面
- 完整的轉換元件資料庫和架構，支援分享和重複使用
- 領先業界的作業系統、資料庫與企業級應用程式連線能力
- 強大的平行處理與平衡最佳化，具備線性的效能調整能力
- 低影響力、即時的資料來源，可將 ETL 流程最佳化
- IBM® InfoSphere™ Information Server 不可或缺的一環

每天都有大量的資料淹沒著 IT 群組，讓企業經理人頭痛不已。企業需要移轉資料，才能掌握深入見解，進而協助提振公司營收，實現獲利最大化。但即便是斥資數百萬美元，投入企業自動化專案，其中亦不乏企業資源規劃 (ERP)、客戶關係管理 (CRM)、供應鏈管理 (SCM)、商業智慧 (BI) 和資料倉儲解決方案等投資，許多公司仍舊存在著不連貫、「無效」資料的困擾。四處充斥著大量、昂貴、孤立、不相通的資訊，以及未經整合的冗餘系統，無法提供企業單一而值得信賴的觀點。

在部署新的企業應用程式、整合併購企業或追蹤 IT 商業流程計畫期間，組織的第一步便是確保資料的可靠性、相關性和可用性，且在整個企業內部必須妥善維繫格式的一致性與簡潔。

IBM InfoSphere Information Server 能協助組織辨識、清理、轉換並遞送相關資料，專供關鍵商業計畫運用。它能緊密整合全面性的企業資訊，包括了橫跨各種資料來源與目標的多樣資訊。無論是要建構企業資料倉儲、實作 BI 平台，或整合數十個來源系統來支援 CRM、SCM 或 ERP，InfoSphere Information Server 都能協助確保組織擁有可靠、可行的資訊。透過 InfoSphere Information Server，企業得以設計出可靠、可調整且彈性的資料整合基礎架構，以符合當下的動態企業環境。



## 運用 IBM InfoSphere DataStage 整合、轉換並遞送可靠資訊

IBM InfoSphere DataStage® 是 InfoSphere Information Server 資訊整合平台的核心產品模組，它提供了世界級的資料整合及轉換能力，創造卓越的生產力。

### IBM InfoSphere DataStage 主要功能

- 針對橫跨大型主機和分散式系統的作業系統、資料庫和企業和應用級程式，提供卓越的連線能力
- 使用常見的中介資料儲存庫，與其他擁有資料輪廓描繪和資料品質能力的 InfoSphere Information Server 模組緊密整合
- 提供易於使用、由上而下、隨想隨做的設計介面
- 包含完整的轉換元件資料庫，可輕鬆定義常用的整合流程
- 採用完整的工具集，可管理、部署並更新整個資料整合生命週期內的資料流
- 支援服務導向架構 (SOA) 方式，促進了複雜整合資料流的重複使用
- 建構動態的平行處理基礎架構，使用者僅需經過一次設計，便能在執行時期視需要部署，無需變更任何的整合工作
- 提供即時的資料來源，將擷取、轉換、載入 (ETL) 流程最佳化

支援的來源、目標和應用程式包括：

- 文字檔，包含大型主機的檔案
- 複雜的 XML 資料結構
- 企業應用程式系統，例如 SAP、Oracle、Siebel、PeopleSoft、JD Edwards、Hyperion 及 Salesforce.com
- 幾乎所有的資料庫，包含已分割的資料庫，像是任何平台（無論是否具備資

料分割區功能）上的 IBM DB2®、IBM Informix®、Netezza®、Oracle、Sybase（ASE 和 IQ）、Teradata，以及 Microsoft® SQL Server®、MySQL 和 Progress

- Web 服務
- 倉儲環境，例如 Teradata、Netezza 和 IBM InfoSphere Warehouse
- 分析工具，例如 SAS 和 IBM SPSS
- 傳訊和企業應用程式整合 (EAI) 產品，包含 IBM WebSphere® MQ
- 公開和私密的雲端

InfoSphere DataStage 是 InfoSphere Information Server 平台的核心元件之一，並且採用了平台上許多的原生功能，部分的主要功能包括：

### 廣泛的企業連線能力

成功的企業級資訊整合，需要存取各種結構化、半結構化或未結構化，且位在企業內外部的資料來源。InfoSphere DataStage 能連線幾乎所有的異質資料來源、目標和應用程式，且只要一個動作就能完成結合（請參閱圖 1）。

### 進階的開發與維護程序

InfoSphere Information Server 使用強力的結構，協助開發人員在建構、部署、更新與管理其資料整合基礎架構之際，提升速度、彈性和效率至最高等級。InfoSphere DataStage 運用了 InfoSphere Information Server 的生產力強化功能，能縮短學習曲線、簡化管理，並將開發資源的使用最佳化。如此一來，資料整合應用程式的開發及維護，速度與進度也更上層樓。透過 InfoSphere DataStage，組織便能存取可靠的資訊，並在各種應用程式和資料庫之間分享資訊，藉此提高投資報酬率。

## 整合的設計介面及常用的中介資料儲存庫

InfoSphere Information Server 採用 InfoSphere DataStage 和 IBM InfoSphere QualityStage™ 模組所共用的單一設計介面，允許設計人員任意組合資料品質與資料轉換功能，確保在正確的時間整合正確的資料。InfoSphere Information Server 也提供統一的中介資料儲存庫，可供 InfoSphere DataStage 和所有其他模組使用。使用者可以立即存取在資料輪廓描繪、資料清除和整合流程期間開發的技術與流程中介資料，以加快開發腳步，並減少發生錯誤的機率。

商業智慧、主要資料管理、客戶關係管理、  
供應鏈管理、風險和法規遵循管理

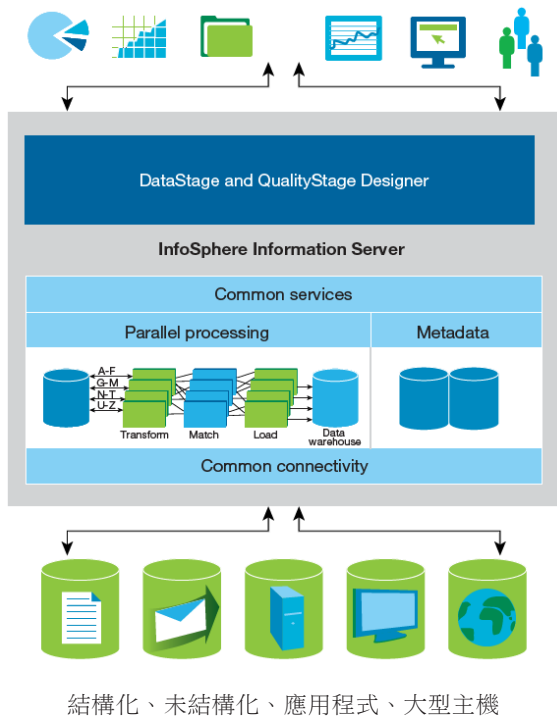


圖 1：InfoSphere DataStage 可存取多重來源系統，整合並轉換特定資料，為關鍵的商業功能提供可靠資訊

## 易用性

InfoSphere DataStage 採用隨想隨做的設計介面（請參閱圖 2）。開發人員可使用由上而下的應用程式設計和執行模型，從而建立虛擬的資料流。功能強大的圖形選用區，協助開發人員利用簡易、GUI 式的拖放設計元件，將整個環境內的資料流轉化為圖形。為了締造最高的生產力，InfoSphere DataStage 加入了超過 50 種預先建構元件，以及數百種的轉換工具。此外還有強大的除錯功能，以及可採用外部程式碼的開放式應用程式設計介面 (API)，堪稱開發人員的福音。

## 生產力與重複使用

InfoSphere DataStage 可促成現有資料整合商業邏輯的重複利用，有助於縮短開發週期。它採用了儲存器 (container) 的概念，將某一儲存器內建立的工作和中介資料提供給其他工作，予以共用和重複使用。此外更具備快速尋找和進階尋找功能，能輕鬆找出物件，在不同的專案中重複使用。強大的工作規格報告功能也提供文件說明，讓其他開發人員可以輕鬆瞭解工作的設計，從而提供額外的支援。

## 時機準確的資料整合

InfoSphere Information Server 架構可讓 InfoSphere DataStage 即時運作，在接獲通知的同時即刻從整合大量資料的同一平台擷取訊息或萃取資料，並且使用相同的轉換規則。資料整合工作可透過 Java™ Message Services、Web 服務和其他方法部署。此種 SOA 方法讓許多的開發人員不需要先瞭解服務所包含的步驟，亦能分享複雜的資料整合流程。所締造的成果：多樣化的資料運用方式，無需高成本的手動撰寫程式碼，可隨需回應組織的資訊整合需求。

## 領先市場的彈性和可調整性

InfoSphere Information Server 促成了大量資料的高效能整合。運用多重處理器硬體平台的平行處理能力，InfoSphere DataStage 可讓企業依據線性擴充資料產量的速度。組織可以擴充轉換工作，從而解決不斷成長的資料量和不斷縮減的批次作業時段等需求。開發可透過循序邏輯完成，部署配置會自動加入所需要的平行化。組織可以在早上採用 2-way 處理，中午改為 32-way 處理，然後晚上改為 128-way 處理；只要輕鬆變更配置檔，就能如願以償。

## 箇中訣竅：分割與動態重新分割

InfoSphere Information Server 平行技術的運作方式，也就是使用「區分—解決」技巧，將最大的整合工作分割為子集（分割區平行化），然後再將這些子集同時散佈到所有可用的處理器（管線平行化）。管線和資料分割區平行化的結合，為 InfoSphere DataStage 帶來真正的線性可擴充性。效能會隨著處理器數量等比例增加，硬體是唯一限制效能的因素。

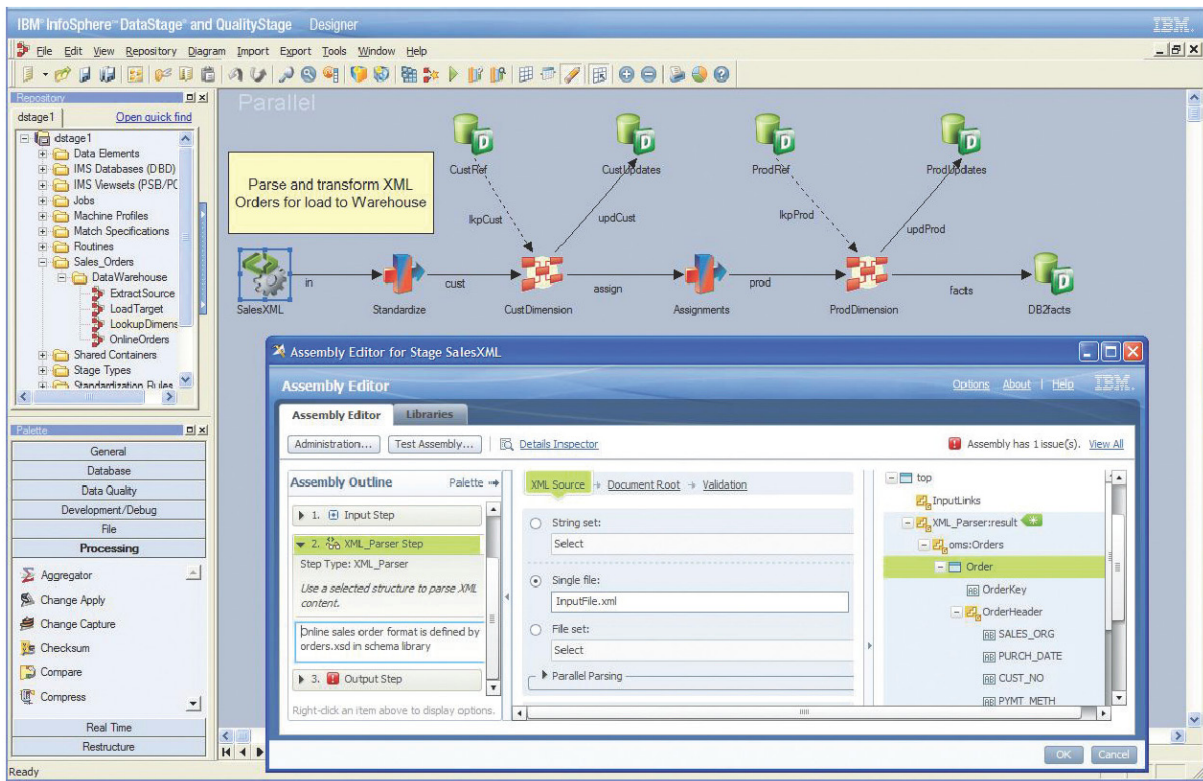


圖 2：InfoSphere DataStage 使用由上而下、隨想隨做的方式，輕鬆規劃企業資料流



試想，根據客戶的姓氏來進行轉換，同時針對家庭用途加入郵遞區號，並加入信用卡號碼，然後將資料載入資料倉儲資料庫平行載入器。藉由動態的資料重新分割，資料便能在流程之中快速重新分割，無需比照其他許多整合產品，被迫採用緩慢且高成本的步驟將資料載入磁碟。

## 廣泛平行支援 SMP、MPP 和網格部署

IBM InfoSphere Information Server 不費吹灰之力，就能從對稱的多重處理器 (SMP) 系統和 SMP 叢集，擴展到擁有上百顆處理器的大量平行處理 (MPP) 伺服器。同樣的這種整合能力，也能使用在低成本伺服器的網格部署上。其對平行處理的廣泛支援，確保了關鍵企業資訊整合工作能隨著企業的需求而擴展。

---

IBM InfoSphere DataStage 可在多種平台上運作。如需完整清單，請造訪 [ibm.com/software/data/infosphere/datastage/requirements.html](http://ibm.com/software/data/infosphere/datastage/requirements.html)

---

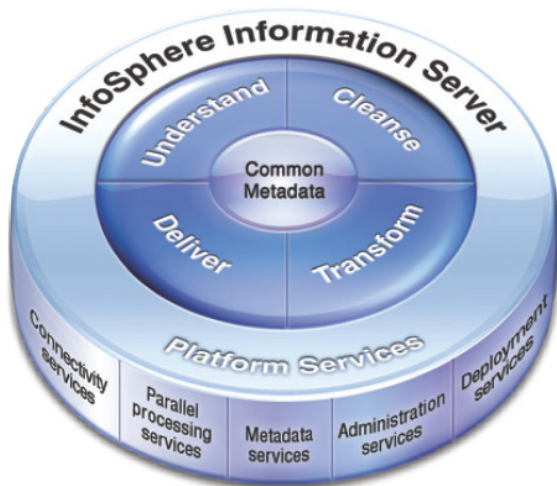


圖 3 : IBM InfoSphere Information Server

## 關於 IBM InfoSphere Information Server

InfoSphere Information Server 是領先業界的資料整合平台，可協助辨識、清理、轉換並遞送相關資料（請參閱圖 3），專供關鍵商業計畫使用。平台提供了整合整個系統內異質資訊所需要的一切工具，包括資訊控管、資料品質、資料轉換和資料同步等功能，從而確保了資訊定義一致、準確呈現、穩定轉換，且獲得定期更新。InfoSphere Information Server 加速並促成了企業及 IT 專業人員之間的合作，確保商業分析、主要資料管理、應用程式合併、移轉及資料倉儲專案等策略計畫，都能使用準確、全面、深入且可即時取得的可靠資訊。

如需 IBM InfoSphere Information Server 系列產品的其他詳細資料，請造訪 [ibm.com/software/data/integration/info\\_server](http://ibm.com/software/data/integration/info_server)

## 進一步資訊

如需 IBM InfoSphere DataStage 相關資訊，請造訪 [ibm.com/software/data/infosphere/datastage](http://ibm.com/software/data/infosphere/datastage)



---

© Copyright IBM Corporation 2011

IBM 軟體事業處  
技術諮詢熱線：0800-000-700  
台北市松仁路 7 號 3 樓

台灣印製  
2011 年 2 月  
版權所有

IBM、IBM 標誌、ibm.com、DataStage 及 InfoSphere 是國際商業機器股份有限公司在美國及／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](http://ibm.com/legal/copytrade.shtml)

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。

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

其他公司、產品或服務名稱各為所屬公司之商標或服務標章。

Netezza 是 IBM 公司旗下 Netezza Corporation 的註冊商標。

本出版品中提及的 IBM 產品或服務，並不代表 IBM 有意將其推展至 IBM 事業營運涵蓋的所有國家。所有關於 IBM 未來方向或意向的陳述，僅供陳述目標及目的之用，均可能改變或撤銷，恕不另行通知。



請回收