

IBM InfoSphere BigInsights 企業版

以符合成本效益的方式，儲存與管理海量資料 (Big Data)，洞察資料意義



產品特色

- 提供企業級的海量資料 (Big Data) 分析平台
- 整合非 IBM 解決方案與熱門的 IBM 產品及服務 (如 IBM® Netezza® 設備、IBM DB2® 資料庫軟體、IBM InfoSphere® Streams 與 IBM InfoSphere DataStage®)
- 支援結構化、半結構化與非結構化資料，彈性極佳

許多企業的業務活動資訊，無論種類、產生速度與數量都急遽成長，企業疲於管理大量且多元化的資料，無論是傳統結構化、半結構化的資料，或是大量未開發的非結構化資料，這種新一類的資訊，我們稱為「海量資料 (Big Data)」。企業瞭解若能深入分析資料，就能發揮龐大資料潛能，讓決策更迅速、明確與精細，但首先資料管理與分析工具效率必須大幅提升。

IBM InfoSphere BigInsights Enterprise Edition (企業版) 讓企業創造新解決方案，以符合成本效益的方式，洞察大量複雜資料的意義。此分析平台符合企業需求，結合 Apache Hadoop 與 IBM 獨特創新技術，具復原與容錯能力，能處理、分析大量的擴充式 (scale-out) 資料。

重點提要：幫助企業善用大量未開發的資料，以充滿效率、可靈活擴充的最佳方式，洞察資料的意義。



將海量資料 (Big Data) 融入企業

InfoSphere BigInsights 採用開放原始碼 Hadoop，新增企業級功能與整合，滿足關鍵業務需求，企業可利用具成本效益的眾多伺服器，執行大規模且分散的分析工作。此基礎架構運用 Hadoop 的 MapReduce 架構，分解許多節點上的資料，協調大量平行操作的資料處理，以因應大量資料集。原始資料儲存在分散的叢集後，就能有效查詢及分析資料，即時動態解讀資料。

InfoSphere BigInsights 整合式解決方案經過完整的測試，結合最先進技術與成熟功能，符合企業需求，管理者可運用 GUI 型安裝工具，快速啟用執行，並能透過引導，安裝指定的元件並配置平台，安裝進度即時報告，內建的狀態檢查可自動驗證安裝是否成功；這些進階的安裝功能可所短安裝與調整時間，讓管理者能進行其他重要專案。

InfoSphere BigInsights 安裝完畢後，就能發揮企業級的功能，協助簡化工作量與系統管理（請參閱圖 1）。舉例來說，強大的工作管理功能可讓企業精密控制 InfoSphere BigInsights 工作的所有層面，技術人員能輕鬆地建立、送出與取消工作；此外，內建工作狀態顯示、工作日誌與技術器，讓管理者瞭解配置、作業、嘗試與重要資訊，隨時掌握工作進度。InfoSphere BigInsights 也包含各種管理功能，包括 Hadoop Distributed File System (HDFS) 與 MapReduce 管理、叢集與伺服器管理、檢視 HDFS 檔案內容、角色特定檢視等。



圖 1：運用 InfoSphere BigInsights，瞭解工作與連結 Web 主控台。

親身體驗：

下載 InfoSphere BigInsights Basic Edition（基本版）

IBM 提供免費、入門級的 InfoSphere BigInsights 基本版，供開發人員、合作夥伴與有興趣的用戶試用。您可以從屬意的供應商，下載並在雲端執行 BigInsights 基本版。如需相關資訊，請造訪：

ibm.com/software/data/infosphere/biginsights/basic.html

基本版用戶可選購全天候 Elite Support。運用 InfoSphere BigInsights 基本版開發的解決方案，可妥善整合至 InfoSphere BigInsights 企業版。

提升海量資料 (Big Data) 分析的安全性

企業對安全有非常嚴格的要求，InfoSphere BigInsights 可精密調整，維護資料安全與隱私。

身份驗證：

InfoSphere BigInsights 主控台可選擇純文字檔、輕量型目錄存取通訊協定 (LDAP) 或未驗證。有了 LDAP 驗證，InfoSphere BigInsights 安裝程式能與 LDAP 憑證儲存庫聯繫，進行驗證，接著，管理者可根據使用角色成員資格，設定使用者群組存取 InfoSphere BigInsights 主控台的權限。

授權

InfoSphere BigInsights 根據使用角色，提供 4 個層級的使用者授權：系統管理員、資料管理員、應用程式管理員與非管理使用者；使用者存取資料與功能的權限，視其角色而定。

提升效能，簡化工作處理

IBM 新增數種功能，可提高效能，讓 InfoSphere BigInsights 更具彈性且能相容於企業環境。

BigInsights Scheduler 配置工作流程

並非所有工作都優先順序都相同，BigInsights Scheduler（排程器）提供 MapReduce 工作可調整的工作流程配置方案，根據使用者選擇的原則，最佳化處理。排程器是 Hadoop Fair Scheduler 的延伸套件，能為所有工作公平分配叢集資源。

BigIndex 支援大型索引編製

BigInsights 納入 BigIndex 功能，讓 Hadoop 索引編製更為容易。BigIndex 以 Apache Lucene 為基礎，在海量資料 (Big Data) 中提供快速的全文搜尋，在工作流程中運用 BigIndex 模組，能編製、掃描與查詢索引，以及其他複雜功能，如分散式索引編製 (Distributed Indexing) 與多面向搜尋 (Faceted Search)，讓自訂應用程式開發與搜尋技術選擇更有彈性。

Adaptive MapReduce 加速完成工作

在 InfoSphere BigInsights 上執行的工作，經常建立多個小型作業，耗用大量系統資源，有鑑於此，IBM 發明了 Adaptive MapReduce 技術，改變 MapReduce 處理工作的方式，但不改變建立工作的方式，加速完成小型作業；Adaptive MapReduce 完全不影響 MapReduce 與 Hadoop API 作業。

運用分析加速器，深入分析海量資料 (Big Data)

InfoSphere BigInsights 包含眾多分析工具與功能，開箱即用，可迅速找出資料模式，也能建立強大的自訂分析應用程式，洞察資料、發揮成效，滿足特定企業需求。

BigSheets

BigSheets 為革命性資料分析工具，以瀏覽器為操作介面，可探索儲存在 BigInsights 叢集的資料，無需撰寫程式碼就能分析資料（請參閱圖 2），內建分析巨集能滿足一般資料探索需求，進一步改善資料存取。BigSheets 協助企業使用者：

- 從網頁儲存庫整合龐大的非結構化資料
- 從使用者定義的 URL 種子，收集各式各樣非結構化網頁資料
- 分析文字，擷取與擴充網頁資料
- 分析使用者定義環境的資料，以圖形呈現

BigSheets 工具可為海量資料 (Big Data) 進行特定分析，無需 IT 協助，也能深入了解資料。

Year	Quarter	Revenue	Address
2008	Q1	2700000000	http://www-01.ibm.com/services/...
2008	Q2	2720000000	http://www-02.ibm.com/services/...
2008	Q3	2750000000	http://www-03.ibm.com/services/...
2008	Q4	2770000000	http://www-04.ibm.com/services/...
2009	Q1	2800000000	http://www-01.ibm.com/services/...
2009	Q2	2820000000	http://www-02.ibm.com/services/...
2009	Q3	2850000000	http://www-03.ibm.com/services/...
2009	Q4	2870000000	http://www-04.ibm.com/services/...
2010	Q1	2900000000	http://www-01.ibm.com/services/...
2010	Q2	2920000000	http://www-02.ibm.com/services/...
2010	Q3	2950000000	http://www-03.ibm.com/services/...
2010	Q4	2970000000	http://www-04.ibm.com/services/...
2011	Q1	3000000000	http://www-01.ibm.com/services/...
2011	Q2	3020000000	http://www-02.ibm.com/services/...
2011	Q3	3050000000	http://www-03.ibm.com/services/...
2011	Q4	3070000000	http://www-04.ibm.com/services/...

圖 2：BigSheets 瀏覽器工具能快速存取非結構化資料，無需撰寫任何程式碼即能迅速建立查詢。

進階文字分析加速器

(Advanced Text Analytics Accelerator)

BigInsights 包含 IBM 開發的強大文字分析引擎 (Text Analytics Engine)，IBM Watson™ 便是運用此引擎，在 Jeopardy! 機智問答賽中擊敗兩名最優秀的參賽者。開發人員運用無所不包的規則庫 (Library of Rules) (或自訂規則)，迅速查詢、辨識文件及訊息中關注的項目，包括人員、電子郵件位址、地址、電話號碼、網址、合資企業、聯盟等。文字分析引擎支援英文、荷蘭文／法蘭德斯語、法文、德文、義大利文、葡萄牙文、西班牙文、日文與中文。

Jaql

Jaql 由 IBM 開發，為強大、高階的宣告式查詢語言 (declarative query language)，支援開放標準，能處理結構化與非結構化資料。Jaql 介面以類 SQL (SQL-like)，讓熟悉 SQL 語言的開發人員輕鬆上手，整合相關資料庫更輕鬆。InfoSphere BigInsights 內含預先安裝的 Jaql 模組，包括 Lucene 索引、Netezza 資料倉儲應用程式系列、HBase (Hadoop 資料庫) 與工作流程 (包括 InfoSphere BigInsights 內建文字分析功能)。

將海量資料 (Big Data) 融入現有資訊架構

分析海量資料 (Big Data) 對企業資訊供應鏈大有幫助，但必須先與現有系統密切整合。IBM 瞭解這點，因此 InfoSphere BigInsights 企業版提供 IBM Netezza 資料倉儲應用程式系列、IBM DB2、IBM InfoSphere Warehouse 與 IBM Smart Analytics System 的高速連接器，搭配 DB2 或 IBM 倉儲產品，可簡化與加速操作資料。

InfoSphere BigInsights 企業版也隨附標準 Java Database Connectivity (JDBC) 連接器，讓企業能迅速整合各種資料及資訊系統，包括 Oracle、Microsoft SQL Server、MySQL 與 Teradata。

InfoSphere DataStage 工具包含一個連接器，能在 DataStage ETL 工作中運用 BigInsights 資料；IBM InfoSphere Streams 包含一個連接器，終端使用者能讀寫 BigInsights 檔案系統。

InfoSphere BigInsights 也支援開放式標準與 Apache Nutch 網頁搜尋軟體，搜尋企業內外非結構化的資料。

獲得企業級的支援

開放原始碼軟體原本就沒有技術支援，其法律條款可能不適用於部分企業。相形之下，InfoSphere BigInsights 企業版具備標準 IBM 軟體授權與支援協議，企業可根據熟悉的授權條款部署此軟體，將不確定性與風險降到最低，能獲得全天候的支援供應項目、教育與全球專業服務。

IBM 的開放原始碼技術方案功能完備、方便整合且符合企業需求，能將資料分析技術導入現有的資訊供應鏈，幫助企業簡化與加速分析海量資料 (Big Data)。IBM 解決方案效能出色且可靠性高，是企業決策程序不可或缺的一部分。

硬體需求與作業系統支援

Intel x86 伺服器，64 位元，最低 4 GB 記憶體與 30 GB 磁碟儲存空間

Red Hat Enterprise Linux 5.3、5.4、5.5、5.6、6.0S
USE Linux Enterprise 11、11 SP1

附註：請參閱產品文件，瞭解最新作業系統版本要求：
ibm.com/common/ssi/cgi-bin/ssialias?infotype=AN&subtype=CA&htmlfid=897/ENUS211-501&appname=USN

進一步資訊

如需進一步瞭解 InfoSphere BigInsights 企業版的相關資訊，請聯絡 IBM 業務代表，或造訪：ibm.com/software/data/infosphere/biginsights/enterprise.html

此外，IBM 租賃事業部能助您以最符合成本效益與最適切的方式，取得所需的 IT 解決方案。我們會與符合信用資格的客戶合作自訂 IT 租賃解決方案，以符合您的商業目標、啟用有效的現金管理，並降低您的總擁有成本。IBM 租賃事業部資助重要 IT 投資與推動企業向前邁進，是您最智慧的選擇。如需相關資訊，請造訪：ibm.com/financing



© Copyright IBM Corporation 2012

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

台灣印製
2012 年 3 月

IBM、IBM 標誌、ibm.com、DataStage、DB2 和 InfoSphere 是國際商業機器股份有限公司 (IBM) 在全球多個轄區註冊的商標。其他產品和服務名稱，可能是 IBM 或其他公司的商標。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

Intel 是 Intel Corporation 或其子公司在美國及其他國家的註冊商標。

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

Microsoft 是 Microsoft Corporation 在美國及 / 或其他國家的商標。

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

本文為發行當日的最新資訊，IBM 得隨時變動。部份國家可能未提供所有產品與服務。

本文所載資訊僅以「現狀」提供，不包括任何明示或默示之保證，包括未對可售性、符合特定效用及未侵權提供任何保證。IBM 產品保固係根據其隨附合約條款。



請回收