



## **IBM Red Brick Warehouse 提供強大的資料倉儲效能**

---

## 目錄

---

### 1 普及化的資料倉儲系統

#### 普及化的資料倉儲系統

產業專家預估，未來三年內有 92% 的 IT 部門將為企業部署資料倉儲。隨著資料倉儲日漸普及，其核心技術也愈見精細複雜。

### 2 單一步驟的載入處理程序

資料倉儲或資料超市應用程式的最重要元件為關聯式資料庫管理系統 (relational database management system, 簡稱 RDBMS)，RDBMS 儲存了大量資訊，使用者可迅速確實的回應各種業務問題。

### 3 高效率的載入效能

### 4 大量資料儲存

資料倉儲通常藉由最佳化 RDBMS 的線上分析處理 (online analytical processing, 簡稱 OLAP) 來提供商業智慧價值；OLAP 必須從各種觀點進行資料的萃取與檢視。例如 OLAP 可對資料倉儲提出查詢，比對今年特定月份與去年同期的產品收益。

### 5 迅速回應查詢

### 6 Red Brick Warehouse：展現資料倉儲成功典範

OLAP 應用程式需要不斷提升效能水準及功能。畢竟資料倉儲或資料超市的單項查詢，其作業量是一般線上交易處理 (online transaction processing, 簡稱 OLTP) 查詢的千倍或百萬倍 (更別提其中還牽涉到不同類型的作業)；OLTP 的作業通常只涉及資料輸入與擷取。

本報告強調 IBM Red Brick™ 的效能優點。Red Brick Warehouse 是一個針對 OLAP 與其他商業智慧功能最佳化的 RDBMS。本研究報告所提供的範例取自星狀綱要式資料庫的延展性評比測試，這些評比測試是在控制條件下由客戶執行，用以模擬正式上線的資料倉儲應用程式。

---

## 重要特色

---

### 低成本高效能

#### 單一步驟的載入處理程序

Red Brick Warehouse 於 1990 年代初期上市，是專門為運用主從運算環境發展標準的決策支援應用程式所設計的開放關聯式資料庫技術，被公認為成本低廉且高效能的分析型資料庫。網際網路入口網站的領導廠商 Terra Lycos 公司資料庫作業經理 Frank Flynn 就表示：「利用 IBM Red Brick Warehouse 統合資訊後，不但強化公司商業智慧應用程式的效益，也發展出更符合我們策略性目標的明智決策。」

Red Brick Warehouse 結合 UNIX®、Linux 及 Microsoft® Windows® 伺服器平台的成本優勢，可支援決策支援、資料倉儲及其他商業智慧應用程式，而且成本較同級系統低得多。不論何種規模的企業，只要具備 Red Brick Warehouse 這類解決方案，即可執行強大的商業智慧應用程式，且企業投資也可獲得豐碩的報酬。

### 移入資料倉儲

有關資料倉儲應用程式的初始化及維護，最關鍵部分在於資料倉儲的資料匯入 — 載入處理。最佳化的載入效能可產生兩大商業利益：一是所有必要資料都可在載入期間納入倉儲，而不會影響資料品質；二是可經常性載入資料，以滿足業務上的週期性需求。

資料載入流程包括下列作業，所有設計皆為了確保載入倉儲系統的資料是正確無誤的：

- 從輸入來源讀取資料，將資料寫入資料倉儲。
- 新增資料鍵時，擴增內部索引結構。
- 確保輸入的資料不會違反參考完整性（例如，輸入 x 品目的銷售額前，必須有該品目的描述、來源及成本等等）。
- 必要時可進行資料轉換（例如，將 EBCDIC 轉為 ASCII、或者時間格式轉換等等）。
- 自動維護不同的資料摘要層次或階層。

---

## 重要特色

---

### 以單一腳本執行多重作業

### 快速載入資料

### **Red Brick Warehouse 的規模越大， 載入資料就越有效率。**

企業採用 Red Brick Warehouse 後，只需單一步驟即可執行所有載入處理作業。此特色可節省商業智慧應用程式的開發時間，並將相關成本降至最低。

### 高效率的載入效能

資料倉儲應用程式的另一個成功關鍵就是載入效能。在更新決策支援所需的資料庫時，經常需要載入大量資料，由於處理資料量很大，所以需要高效能的資料載入器。

IBM 所設計的 Red Brick 資料載入功能可有效率的執行下列作業：

- 平行處理索引建立、參考完整性檢查、及載入處理。
- 智慧型緩衝區快取管理，可有效使用記憶體、減少 I/O（輸入/輸出）、並最佳化載入效能。
- 當輸入的資料無法通過參考完整性檢查時，可選擇為該資料自動產生資料列，以大幅減少一致化不良資料所需的時間。
- 寫入硬碟之前在記憶體中進行最佳化的資料和索引排序，以減少移入倉儲時的硬碟讀寫次數。
- 區塊層級的版本控制 (versioning) 功能可維持資料庫的可用性，讓您在進行高速大量載入時仍可提出查詢。
- 載入新資料時建立多重索引，如此資料庫管理員 (database administrator，簡稱 DBA) 就可以平行方式在新資料上建立多重索引。

在以 HP/Compaq 的 AlphaServer GS320 系統上執行 Red Brick Warehouse 的聯合研究中，Red Brick Warehouse 在資料載入、查詢效能和延展性等項目皆獲高分。該研究針對高達一千名使用者進行 2TB、4TB 和 8TB 資料倉儲的效能評估，並且利用 Red Brick Warehouse 的效能及延展性證明 (proof of performance and scalability，簡稱 POPS) 進行評比。整體而言，當 Red Brick Warehouse 的規模變大時，載入效率也隨之提高。表 A 的載入效能表即為此次評比測試之摘要。

## 重要特色

資料點	表格名稱	GB/ 每小時	載入筆數	經歷時間	筆數 / 每分鐘
2TB	每日銷售額	200.0	8,483M	10:17:38	13,735,651
4TB	每日銷售額	229.8	17,060M	17:59:57	15,797,533
8TB	每日銷售額	279.0	34,121M	29:39:27	19,172,922

表 A. 載入效能評比

**佳能 (Canon) 就是仰仗有效率的分析顧客反應來強化產品的。**

證明其載入效能的另一個範例是佳能公司 (Canon, Inc.)，他們使用 Red Brick Warehouse 建置資料倉儲來支援分析顧客對產品的反應。佳能在進行產品開發及查詢控管時會納入顧客的意見，因此對佳能而言，主要的考量點是 Red Brick Warehouse 的高速資料載入及查詢回應功能。越快將顧客的反應傳達給產品研發人員，新產品及現有產品的開發成果就會越好。

### 大量資料儲存

Red Brick 資料倉儲可從數百億位元組延展至數兆位元組。在許多產業像是醫療保健、金融服務和電訊業，Red Brick 資料倉儲成長到數十兆位元組是很常見的。

**Red Brick Warehouse 專為管理大量資料而設計**

資料倉儲的單一表格可能含有數億筆記錄；有效率的儲存大量資料組的能力，可大幅減少資料倉儲應用程式的整體維護成本。由於 Red Brick Warehouse 針對資料倉儲應用程式已經最佳化，可以很有效率地運用可用的資料儲存空間。Red Brick Warehouse 採用壓縮索引結構，並將儲存索引所需的硬碟空間降至最低。最佳化的索引建立運算法會充分利用區塊空間配置，以便初始索引建立以及漸進式索引更新。此外，它以壓縮二進位數字資料類型來儲存數字資料。最新版的 Red Brick Warehouse 還提供壓縮公用程式，可縮小資料庫系統型錄，進而提升查詢效能。

**硬碟空間需求愈小，意謂其所需成本愈低。**

利用 Red Brick Warehouse 獨有的索引結構及儲存功能，使用者可將 Red Brick Warehouse 資料倉儲所需的硬碟空間降至最低。由於適用大型資料倉儲的儲存子系統費用高昂，因此 Red Brick Warehouse 是一個具有成本效益的選擇。詳情請參見表 B，該表列有 Red Brick Warehouse 儲存 8TB 原始資料所用到的總儲存空間。

## 重要特色

表格名稱	筆數	資料儲存	索引儲存
每日銷售額	34,121,091,680	8,271,782,144K	938,052,656K
每日預估	3,574,993,950	140,196,072K	38,511,552K
總計		8,411,978,216K	976,604,208K

表 B. 8TB 原始資料的儲存容量

### 適用於最複雜的查詢

採用 OLTP RDBMS 技術的資料倉儲應用程式極為仰賴聚合 (aggregate) 資料，因為讓每位使用者存取詳細資料將會嚴重影響系統效能。在維護聚合層次資訊方面，這些 OLTP RDBMS 產品較 Red Brick Warehouse 高出 200%-300% 的處理虛耗。Red Brick Warehouse 自詡與生俱來的高效能可處理詳細資料的複雜查詢，因此聚合層次的查詢處理可說是輕而易舉。使用聚合所減少的執行時期資源，可彌補聚合資料所需的額外硬碟儲存空間。此外，一般使用者是看不到 Red Brick 的聚合方式的；他們查詢的是靜態的細部綱要，而系統則是動態存取聚合資料，以最佳化其查詢處理效能。

### 迅速回應查詢

對一般使用者而言，資料倉儲的查詢效能對於其可用性及其有效性的影響最大。回應時間必須迅速，才能支援反覆的探索程序。當分析師及企業經理人進行探索程序時，可深入鑽研特定的業務性質。

### 星狀綱要設計的優點

想想這個常見的商業問題：今年產品 x 相較於產品 y，在北美的銷售額較去年表現如何？這類複雜查詢格外仰賴最佳化的關聯式資料庫運算法，RDBMS 必須處理許多表格之間的聯結。

Red Brick Warehouse 對於複雜查詢的處理已經最佳化，為需要聯結多重表格的查詢，提供可預測且快速的處理。此功能源自於星狀綱要設計，可提供以查詢為中心的資料觀點。星狀綱要將資訊分為兩大類：事實 (fact) 與維度 (dimension)。事實是指要進行分析的核心資料元素，維度則指一些描述事實或說明事實背景的屬性。

---

## 重要特色

---

### 定義 STARjoin 與 STARindex 技術

Red Brick Warehouse 採用專為加快複雜查詢處理所開發的 STARjoin 與 STARindex 技術。STARjoin 是一種高速、單一傳遞 (single-pass)、聯結多重表格的技術，提供優異的聯結處理效能。STARindex 含有高度壓縮的資訊，能夠將事實表格的維度連結至含有此維度的資料列，大幅提升聯結的效能。STARjoin 運算法可利用 STARindex，有效率的識別某特定聯結所需的全部資料列。影響資料倉儲效能的因素有很多，從非預期查詢到各種不同的使用者載入以及資料偏斜等等都是。但是 Red Brick Warehouse 所採用的星狀綱要處理解決方案，可立即而有效率的解決各種查詢。

當資料集與使用者數目增加時，可預測的線性查詢效能將有助於提升企業預測系統需求的能力。Red Brick Warehouse 的設計可減輕容量規劃的負擔。在多重表格聯結的測試中，Red Brick Warehouse 持續以少於一分鐘的回應時間回覆複雜查詢，此即 STARindex 與 STARjoin 的查詢加速運算功能。參與 STARjoin 計畫的維度表格及其對應的主鍵都會快取在記憶體中。快取可縮短單一查詢的回應時間，也可以讓多位使用者同時進行存取，以減少 I/O（輸入 / 輸出）次數與記憶體佔用。

### 備經驗證的商業智慧平台

#### Red Brick Warehouse：展現資料倉儲成功典範

在全球所有的產業中，企業都知道真知灼見的商業資訊就像人才與優質產品一樣，能夠提供強大的競爭優勢。此種理解也加速了資料倉儲與其他決策支援技術的普及化，正因如此，所有資料倉儲的核心元件 RDBMS 也因而位居中樞。針對資料倉儲最佳化的 RDBMS 具有 DBA 需要的重要特質：有效率的載入處理與載入效能、可迅速回應查詢、並具備儲存大量資料的能力。Red Brick Warehouse 是商業智慧應用程式的分析型資料庫，經過一連串的評比測試後，證明其符合這些標準。企業採用 Red Brick Warehouse 可獲得所需的高效能，並且使用簡易且兼具低廉的總持有成本，最具投資效益。

#### 其他相關資訊

如需其他相關資訊，請洽 IBM 業務代表或 IBM 事業夥伴，或者聯絡 IBM 台灣聯合服務中心 0800-016-888，或者造訪我們的網站 [ibm.com/software/data/informix/redbrick](http://ibm.com/software/data/informix/redbrick)



**台灣國際商業機器股份有限公司**

台北市松仁路7號3樓

市場行銷處：0800-016-888 按1

技術諮詢熱線：0800-000-700

© Copyright IBM Corporation 2004

台灣印製

04-04

版權所有，翻印必究

AIX、DB2、IBM、IBM 標誌及 Red Brick 為 International Business Machines Corporation 在美國、其他國家或所有國家的商標。

Microsoft 和 Windows 為 Microsoft Corporation 在美國、其他國家或所有國家的商標。

UNIX 為 The Open Group 在美國及其他國家的註冊商標。

所有其他產品或產品名稱皆為各自所有者的商標或註冊商標。

本文件內提及的 IBM 產品或服務不一定表示在全球 IBM 經銷處皆有出售。