



**DB2** Information Management Software

## Powerful data warehousing performance with IBM Red Brick Warehouse



---

**Contents**

---

- 1** *Data warehousing for the masses*
- 2** *Single step load processing*
- 3** *Efficient load performance*
- 5** *Large volume data storage*
- 6** *Rapid query response times*
- 8** *Red Brick Warehouse: demonstrated success for data warehousing*

**Data warehousing for the masses**

Industry experts estimate that in the next three years, 92 percent of IT organizations will deploy data warehouses for their companies. As data warehousing has grown in popularity, the technology at its core has become increasingly sophisticated.

The most important component of a data warehouse or data mart application is the relational database management system (RDBMS). The RDBMS stores vast quantities of information, enabling users to quickly and reliably answer a wide range of business questions.

A data warehouse typically provides business intelligence value through an RDBMS optimized for online analytical processing (OLAP), which deals with extracting and viewing data from different viewpoints. An OLAP query can be submitted against a data warehouse to find out, for example, how much revenue a product generated in a particular month this year versus the same period last year.

OLAP applications call for ever-increasing performance levels and functionality. After all, a single query on a data warehouse or data mart can require thousands or millions of times more work (not to mention different kinds of work) than, say, a typical online transaction processing (OLTP) query, which involves activities such as data entry and retrieval.

This paper highlights the performance benefits of IBM Red Brick™ Warehouse, an RDBMS optimized for OLAP and other business intelligence functions. The examples provided in this study come from scalability benchmark tests of star-schema databases. Performed at customer locations, these benchmark tests were run by customers under controlled conditions and were designed to simulate production data warehouse applications.



---

**Highlights**

---

*High performance for low cost*

**Single-step load processing**

Available on the market since the early 1990s, Red Brick Warehouse is an open relational database technology designed specifically for decision-support applications that leverage the evolving standards of client/server computing environments. Red Brick Warehouse has established itself as a high-performance analytical database with a low total cost of ownership. Notes Frank Flynn, manager of Database Operations for Terra Lycos, a leading Internet portal, “Consolidating information on IBM Red Brick Warehouse has enhanced the effectiveness of Terra Lycos’ business intelligence applications, fostering more informed decisions that will better serve our strategic goals.”

When coupled with the cost advantages of UNIX<sup>®</sup>, Linux and the Microsoft<sup>®</sup> Windows<sup>®</sup> Server platforms, Red Brick Warehouse can deliver decision-support, data warehouse and other business intelligence applications. And it can do so for significantly less expense than comparable systems. With a solution such as Red Brick Warehouse, companies of all sizes can implement robust business intelligence applications and achieve substantial returns on their investments.

*Populating the data warehouse.*

Pulling data into a data warehouse—load processing—is a critical process in the initiation and maintenance of a data warehouse application. Optimized load performance yields two major business benefits. All the necessary data can get into the warehouse during the load window without any sacrifice to data quality, and data can be loaded frequently to match the cyclical needs of the business.

The data loading process involves the following tasks, all designed to make sure that only quality data gets into the warehouse:

- Reading data from an input source and writing data into the data warehouse.
- Augmenting internal index structures as new data keys are added.
- Ensuring that referential integrity is not violated by incoming data (for example, if sales of part x are entered, part x must have a part description, source, cost, etc.).
- Performing data conversions, if necessary (i.e., EBCDIC to ASCII, time formats, etc.).
- Maintaining different summary levels of data or hierarchies automatically.



---

## Highlights

---

### *One script for multiple tasks*

Red Brick Warehouse enables businesses to perform all these load processing tasks in just one step. This feature saves development time for business intelligence applications and minimizes associated overhead costs.

### **Efficient load performance**

Load performance is another critical element in a successful data warehouse application. Updating a database used in decision support requires frequent bulk loading of data. The sheer volume of data involved dictates the need for a high-performance data loader. IBM has designed Red Brick's data-loading capabilities to perform the following tasks efficiently:

### *Speedy data loading.*

- Parallel index building, referential integrity checking and load processing.
- Intelligent buffer cache management for efficient memory usage, I/O reduction and optimal load performance.
- Optional automatic row generation for incoming data that has failed the referential integrity check, which significantly reduces the time for reconciling bad data.
- Optimized sorting of data and indexes in memory prior to writing to disk, which reduces the number of disk writes and reads to populate the warehouse.
- Versioning at the block level for continuous database availability, enabling you to submit queries even during high-speed bulk loading.
- Multiple index creation as new data is loaded, so DBAs can build multiple indexes on new data in parallel.



**Highlights**

*The higher Red Brick Warehouse scales, the more efficiently it can load data*

In a joint study of Red Brick Warehouse running on an AlphaServer GS320 System from HP/Compaq, Red Brick Warehouse demonstrated high marks in data loading, querying performance and scalability. The study measured 2TB, 4TB and 8TB data warehouses with a maximum of 1,000 users, and leveraged the Red Brick Warehouse proof of performance and scalability (POPS) benchmark. Overall, as Red Brick Warehouse was scaled up in size, it demonstrated increasing load rates. See the load performance chart in Figure A for a summary of this benchmark test.

Data point	Table name	GB/Hour	# Rows loaded	Elapsed time	Rows/Minute
2TB	Daily sales	200.0	8,483M	10:17:38	13,735,651
4TB	Daily sales	229.8	17,060M	17:59:57	15,797,533
8TB	Daily sales	279.0	34,121M	29:39:27	19,172,922

Figure A. Load performance benchmarks.

*Canon relies on efficient analysis of customer feedback for product enhancements.*

As another example of its load performance, consider Canon, Inc., which uses Red Brick Warehouse in a data warehouse that supports analysis of customer feedback on products. The high-speed data loading and query response of Red Brick Warehouse are key criteria for Canon, which takes customer input into account in product development and quality control. The faster it can forward feedback to product developers, the better the outcome for new and existing products.

**Large volume data storage**

Red Brick data warehouses are scalable from tens of gigabytes to multi-terabytes. In many industries, such as healthcare, financial services and telecommunications, Red Brick data warehouses that grow to tens of terabytes are not uncommon.



**Highlights**

***Red Brick Warehouse was designed to manage vast amounts of data***

***Smaller disk space needs mean lower costs.***

A single table in a data warehouse may contain many hundreds of millions of records. The ability to store vast data sets efficiently can have a dramatic effect on the overall maintenance costs associated with a data warehouse application. Because it is optimized for data warehouse applications, Red Brick Warehouse leverages available data storage very efficiently. Red Brick Warehouse utilizes compressed index structures and minimizes the disk space required to store indexes. Its optimized index-build algorithms fully utilize block space allocations for initial index creation and incremental index updates. And for storing numeric data, it utilizes compact binary numeric data types. The latest version of Red Brick Warehouse also provides compaction utility that reduces the size of the database system catalog, thereby improving query performance.

By utilizing its unique index structures and storage capabilities, users can minimize the amount of disk space required for a data warehouse based on Red Brick Warehouse. This makes Red Brick Warehouse a cost-effective option, as storage subsystems can be costly for larger data warehouses. For more details, see Figure B, which charts the total storage used by Red Brick Warehouse to store 8TB of raw data.

Table name	# of rows	Data storage	Index storage
Daily sales	34,121,091,680	8,271,782,144K	938,052,656K
Daily forecast	3,574,993,950	140,196,072K	38,511,552K
Total		8,411,978,216K	976,604,208K

Figure B. Storage capacity for 8TB of raw data.



---

**Highlights**

---

***Ready for the most complex queries***

Data warehouse applications using OLTP RDBMS technologies are heavily dependent on aggregate data because providing every end user with access to detail data would severely impact system performance. These OLTP RDBMS products can require 200 percent to 300 percent more overhead than Red Brick Warehouse to maintain this aggregate-level information. Red Brick Warehouse boasts an inherent efficiency at processing complex queries against detail data, so aggregate-level query processing is the icing on the cake, compensating for the additional disk storage that aggregate data requires by reducing run-time resources when aggregates are used. Moreover, the Red Brick approach to aggregation is invisible to end users; they query a static, detail-level schema, while the system accesses the aggregate data dynamically to optimize their queries.

**Rapid query response times**

For end users, the query performance of a data warehouse has the greatest impact on its usability and effectiveness. Response times must be fast to sustain an iterative discovery process. As analysts and business managers progress through their discovery process, they drill down on specific qualitative aspects of the business.

***Benefits of the star schema design.***

Consider this typical business question: How did sales for product x perform versus product y, this year compared to last year, in North America? This type of complex query puts extraordinary demands on the optimization algorithms of a relational database. The RDBMS must process joins between many tables.

Red Brick Warehouse has been optimized for complex query processing, providing predictable and fast processing for queries that require multi-table joins. This capability stems from its star schema design, which provides a query-centric view of data. In a star schema, information is classified into two groups: facts and dimensions. Facts are the core data element being analyzed, and dimensions are attributes that describe or give context to the facts.



---

**Highlights**

---

***Defining STARjoin and STARindex technologies***

Red Brick Warehouse features STARjoin and STARindex technologies developed specifically to accelerate the processing of complex queries. STARjoin is a high-speed, single-pass, multi-table join that provides unequalled join processing performance. A STARindex contains highly compressed information that relates the dimensions of a fact table to the rows that contain those dimensions, dramatically accelerating join performance. A STARjoin algorithm can use the STARindex to efficiently identify all the rows required for a particular join. From unexpected queries to widely varying user loads and skewed data, a variety of factors can hamper data warehouse performance. But the star schema processing solutions in Red Brick Warehouse enables it to tackle complete queries efficiently and without delay.

Predictable, linear query performance is beneficial for organizations in their ability to forecast system requirements as data sets grow and the number of users increases. Red Brick Warehouse is designed to ease capacity planning. In multi-table join tests, Red Brick Warehouse consistently returned results for complex queries with sub-minute response times, a function of its STARindex and STARjoin query acceleration algorithms. Dimension tables and their corresponding primary keys participating in a STARjoin plan are cached in memory. Caching improves single query response time, and multi-users benefit from concurrent access, which reduces I/O and memory usage.



---

**Highlights**

---

*A proven platform for business intelligence.*

**Red Brick Warehouse: demonstrated success for data warehousing**

In nearly every industry worldwide, organizations understand that insightful business information can provide as strong a competitive edge as a talented workforce and quality products. This understanding has fueled the popularity of data warehousing and other decision-support technologies. As such, the RDBMS, the core component of any data warehouse, takes center stage. An RDBMS that is optimized for data warehousing exhibits the key qualities that DBAs seek: efficient load processing and performance, rapid responses to queries and the ability to store vast quantities of data. Red Brick Warehouse, an analytical database for business intelligence applications, meets these criteria, as demonstrated in a series of benchmark tests. With Red Brick Warehouse, companies can get the high performance they need with the ease of use and low total cost of ownership that make their investments worthwhile.

**For more information**

Please contact your IBM marketing representative or an IBM Business Partner, or call 1-800-IBM-CALL within the US. Also visit our Web site at **ibm.com/software/data/informix/redbrick**.



© Copyright IBM Corporation 2004

IBM Corporation  
Silicon Valley Laboratory  
555 Bailey Avenue  
San Jose, CA 95141  
U.S.A.

Printed in the United States of America

04-04


All Rights Reserved

The e-business logo, IBM, the IBM logo and Red Brick are trademarks of International Business Machines Corporation in the United States, other countries or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

 Printed in the United States on recycled paper containing 10% recovered post-consumer fiber.

