



Improving Your Time-to-Market with the Hitachi Business Intelligence Solution

By Pierre Raymond, Ralph Occhipinti, and John Nguyen

Contents

I. Business Intelligence Systems	1
Today's Business Environment	1
The Emergence of Business Intelligence Systems	2
Data Marts—The Business Intelligence System of Choice	2
Assessing the Price/Performance of Business Intelligence Systems	3
The Hitachi Business Intelligence Solution	4
Hitachi VisionBase Server	4
Hitachi Freedom Storage	5
Business Case—ROI Example of Improved Time-to-Market	6
II. Performance Analysis of Typical Business Queries	7
TPC-D Query 1—Pricing Summary Report	7
TPC-D Query 6—Forecasting Revenue Change	10
TPC-D Query 9—Product Type Profit Measure	11
III. The Hitachi Business Intelligence Solution Improves Time-to-Market	12
Appendix A. The Transaction Processing Performance Council	14
Appendix B. TPC-D Query Relevance by Industry Segment	15
Appendix C. TPC-D Performance and Price/Performance Metrics	16
Appendix D. Glossary	16

Improving Your Time-to-Market with the Hitachi Business Intelligence Solution

By Pierre Raymond, Ralph Occhipinti, and John Nguyen

I. Business Intelligence Systems

Today's Business Environment

Enterprises are struggling to keep abreast of change—in their markets, in their competitors' offerings and strategies, in their customers' buying patterns, in their product sales and profitability, in their financial performance, in their employees' productivity. But the rate of change is accelerating and the patterns of change are becoming more complex.

At the same time, the amount of data collected throughout the enterprise is growing at an ever-increasing rate. Data is generated by internal applications such as enterprise resource planning (ERP), customer billing systems, on-line sales, order processing, and human resources. It is also generated by a variety of external sources such as on-line databases and the Internet.

With all of the data available, it's easy to miss the most critical issue for staying ahead of the competition: **time-to-market** decisions about whether to launch new products, service offerings, or marketing campaigns. The ability to analyze enterprise data quickly and to speed up time-to-market can be the difference between success and failure for new programs and products. (See Figure 1.)

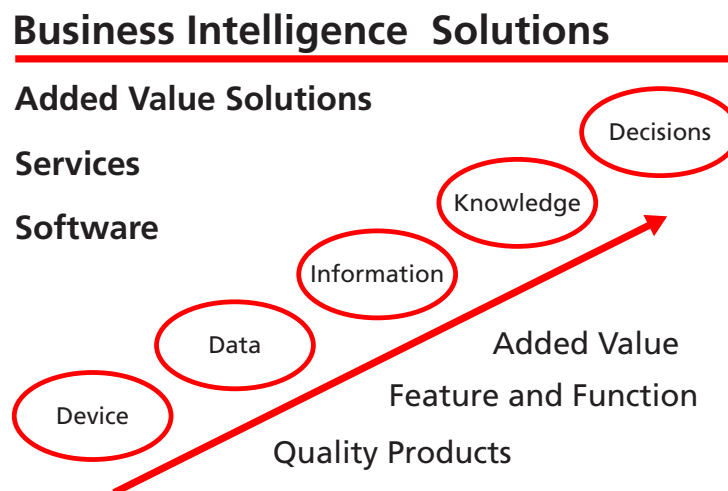


Figure 1: Business Intelligence Solutions

Timely decision making, then, is a necessary condition for maintaining the appropriate time-to-market schedule. This paper outlines how Hitachi Data Systems can help you stay ahead of the competition by improving your time-to-market edge.

The Emergence of Business Intelligence Systems

Business managers understand that information which can be extracted from enterprise data is essential to analyze changes in the business environment and to support day-to-day decision making. They have started implementing business intelligence systems—a combination of decision support tools that include data warehouse, data mart, and data mining software—to turn vast volumes of accumulated enterprise data into useful information for effective decision making.

The acceptance of the business intelligence concept is evident in the new positions that have been created to manage the enterprise's strategic information assets; for example, chief information officer, database administrator, data analyst, and data "miner." Such positions deal directly with protecting and analyzing enterprise data.

As we approach the third millennium, successful enterprises will spend millions of dollars to implement various types and sizes of business intelligence systems.

The business intelligence process involves the following steps:

- Detect and understand what changes are taking place
- Analyze and determine the nature of the changing relationships and their potential sensitivity to proposed actions
- Make a decision and initiate a course of action.

Business intelligence systems can help expedite the data analysis portion of the process, providing managers with a precise understanding of the enterprise's changing markets, products, and customers. This understanding makes the difference between the enterprise successfully taking advantage of market opportunities versus failing to meet business targets. Managers can make timely and better-informed decisions about:

- Launching new products
- Providing new services
- Initiating marketing campaigns and sales programs
- Targeting direct mail, advertising, and so on.

Data Marts—The Business Intelligence System of Choice

A current trend in business intelligence systems is to implement multiple small-to-medium departmental data marts instead of the single enterprise-wide multiterabyte data warehouse. A decision support system (DSS) storing around 30GB of business information defines a typical data mart. Data marts are preferred because, in general, they are quicker to implement, faster to deploy, and cost a fraction of data warehouses.

Hitachi Data Systems believes that the departmental data mart represents the fastest growing portion of the business intelligence market. In fact, *Data Management Review's* survey of more than 200 large enterprises (August 1998) shows that over 60 percent of all business intelligence systems are equal to or smaller than 30GB. The survey also points out that this category is growing much faster than all the other database sizes combined.

In step with this trend is the requirement to analyze departmental data as quickly as possible in order to improve end-user access to important business information. The

support of the process that translates information into business intelligence is crucial to improving time-to-market.

Data mart searches occur in the main accesses of sums, averages, or trend information. Two types of searches must be supported: iterative, and one-time requests.

Iterative uses. Typically a data mart end-user develops a query and may use it only once. When the results are obtained, the user adjusts query parameters and resubmits it in search of a particular result or business intelligence indicator. This iterative process to produce decision support information represents a “tuning” of information. Superior query performance is necessary to support an effective iterative process in departmental data marts.

One-time requests. Superior query performance is an absolute must to support ad-hoc requests from senior management. This class of data mart use is particularly important, since it provides immediate reinforcement or an “ROI” rationale to upper management about the value of the departmental data mart to the business.

Has anyone in a senior management role ever mentioned, “I’m not sure what I want but I’ll know it when I see it!”? This can be part of the fabric that holds any business process together, even at a departmental level.

Assessing the Price/Performance of Business Intelligence Systems

Time is of the essence if a business is to survive and thrive. Regardless of the type of end-user query, the need to lead the competition and the ability to make the best business decision as quickly as possible are fundamental aspects of doing business in today’s economy. The business intelligence system must support fast, cost-effective data analysis.

Enterprise managers can ensure that the business intelligence system they are purchasing will meet their performance and cost objectives by relying on industry-standard benchmarks to measure decision support system performance.

The **Transaction Processing Performance Council (TPC)** is a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable TPC performance data to the industry. While TPC benchmarks certainly involve the measurement and evaluation of computer functions and operations, the TPC regards a transaction in the same way as it is commonly understood in the business world: as a commercial exchange of goods, services, or money. A typical transaction, as defined by the TPC, would include the updating of a database system for such things as inventory control (goods), airline reservations (services), or banking (money).

The TPC-D benchmark is an accepted industry-standard measure for decision support system performance. The TPC-D consists of a suite of business-oriented ad-hoc queries and concurrent data modifications. The queries and the data populating the database were chosen for their broad industry-wide relevance and relative ease of implementation. This benchmark illustrates decision support systems that examine large volumes of data, execute queries with a high degree of complexity, and provide answers to critical business questions.

The TPC-D benchmark evaluates the performance of various decision support systems by the execution of sets of queries against a standard database under controlled conditions. TPC-D queries measure both the server and the storage performance of a complete solution. Since this is an industry-standard benchmark, vendors can be

ranked based on their TPC-D performance and price/performance results. (Appendix A presents information about other TPC benchmarks. Appendix B lists query relevance by specific industry segment. Appendix C provides an overview of the metrics used in reporting TPC-D benchmark results.)

The DSS workload is highly sequential by nature. However, indexing techniques can be used to improve query elapsed time. The DSS environment is somewhat the opposite of the OLTP one (tested with the TPC-C benchmark), in the sense that DSS uses a high-level sequential prefetch and large indexes combined with various complex sort and join operations. This usually means that disk I/O will be saturated when CPU resource is highly utilized, especially with a very large database (VLDB).

The TPC-D benchmark is currently composed of twenty-two unique complex queries and two major update functions called UF1 (which performs inserts) and UF2 (which performs deletes). Some queries are not necessarily relevant to all industry segments. However, by looking at each query, one can determine the relative importance for each major industry segment. The table in Appendix B shows that Query 1, Pricing Summary Report, applies across all industry segments, while Query 12, Shipping Modes and Order Priority, is more appropriate for retail, telco, and banking, than for brokers, utilities, and airlines. This table also demonstrates that the TPC has designed the TPC-D benchmark to ensure an accurate representation of the typical industry environments.

Since decision support systems are now growing into two distinct categories (Business Reports and Ad Hoc queries), the TPC is breaking the TPC-D benchmark into two new benchmarks: TPC-H for Ad Hoc queries, and TPC-R for Business Reports.

The Hitachi Business Intelligence Solution

The Hitachi Business Intelligence Solution includes two major components: the Hitachi VisionBase™ Server and Hitachi Freedom Storage™.

Hitachi VisionBase Server

This Intel-based Windows NT server (see Figure 2) offers many options that enable large enterprises to support various data mart sizes and all kinds of queries, from simple to very complex. The Hitachi VisionBase Server provides unsurpassed performance for the decision support system market. The design of the VisionBase SMP board reflects the same Hitachi engineering used in SMP processors such as the Hitachi Skyline Trinium™ high-end S/390®-compatible servers. The two mandatory design objectives in the plan for the VisionBase Server were to achieve outstanding performance and to exceed the standards set for Hitachi's legendary reliability.

Hitachi VisionBase 4-Way High-End

- Up to four 450MHz Pentium® II Xeon Processors
- Up to 8GB ECC Memory
- Up to 12 hot-swappable drives, RAID 5 support



Figure 2: Hitachi VisionBase Server

Hitachi Freedom Storage

The Hitachi Freedom Storage 5800 subsystem (Figure 3) offers outstanding RAID performance with high levels of connectivity. The sequential and random access performance of the 5800 is optimized for departmental data mart implementations.



Figure 3: Hitachi Freedom Storage

On February 25, 1999, Hitachi Data Systems announced the results of a 30GB TPC-D benchmark. The benchmark test was run on a configuration of a Hitachi AD450NX NT Server (renamed VisionBase 8460) with four 400MHz Intel Pentium II Xeon processors, attached via high-speed fibre channel connections to the Hitachi Freedom Storage 5800 subsystem with 582GB capacity, running the IBM DB2 Universal Database Version 5.2.0 and the Windows NT Server Version 4.0/SP4 operating system.

The TPC-D benchmark test was performed in the Hitachi Open Systems Laboratory in Santa Clara, California. It achieved the following industry-leading results:

TPC-D Power Metric (QppD@30GB):	2,261.2
TPC-D Throughput Metric (QthD@30GB):	325.9
TPC-D Performance Metric (QphD@30GB):	858.4
TPC-D Price/Performance Metric (\$/QphD@30GB):	\$233

This benchmark result exceeded the next-best Power metric by more than three times and the next-best Price/Performance metric by more than seven times.

Business Case—ROI Example of Improved Time-to-Market

Let's examine a business case to demonstrate how these TPC-D results can translate into a significant reduction in time-to-market.

Telephone companies are extensively using data mining and artificial intelligence tools to understand customers' calling habits and call patterns. Based on the information gained from these tools, they hope to bring to the marketplace new and better-targeted long distance programs well in advance of their competitors—perhaps in six months instead of one year. The issue is time-to-market in order to increase market share.

With the Hitachi Business Intelligence Solution, time-to-market can be improved significantly. Figure 4 presents an I/O-bound data mining query where 85 percent of total elapsed time is included in the I/O activity. Hitachi believes that its Freedom Storage offerings can deliver at least a 30 percent improvement in I/O response time over its competitors. The 30 percent improvement will impact the entire query response time, as explained below:

New I/O content: $85\% \times 70\% = 59.5\%$

New query response time = $600 \times ((10+5+59.5)/100) = 447$ seconds

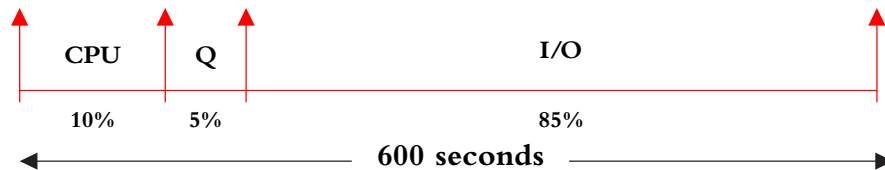


Figure 4: Typical Data Mining Query

Assuming that all data mining queries are essentially of the same nature as described in Figure 4 and that they are running ten hours a day and five days a week on the data mart (a mixture of user-driven and system-driven queries), the total number of queries to complete the study in six months would be:

$$(3600 \text{ seconds} \times 10 \text{ hours} \times 5 \text{ days} \times 4 \text{ weeks} \times 6 \text{ months}) / 600 = 7200 \text{ queries}$$

Now, with the new response time—assuming that the same number of queries (7200) are required to make the decision on a new long distance program—the total duration of the project would be:

$$7200 \text{ queries} \times 447 \text{ seconds} / (3600 \times 10) = 89.4 \text{ days. @ 20 days per month} = 4 \text{ months and 9.4 days}$$

By introducing the program 1.5 months earlier, just the interest on the money coming in earlier would pay multiple times for a business intelligence solution such as the one proposed by Hitachi. Based on our example, the program will help capture 500,000 new customers per month. Each customer will spend an average of \$75 per month in long distance calls. If the interest rate is 9 percent per annum, then:

$$(1.5 \text{ month}) \star (500,000 \text{ customers}) \times (\$75) \times (.09 \star (1.5/12)) = \$703,125$$

In contrast, the total Hitachi solution as reported in TPC-D for a 30GB data mart costs less than \$200,000.

*The real benefit of a faster time-to-market in our example resides in the fact that the actual revenue of the program will be much higher than originally projected. This is achieved by capturing more new customers, **earlier**, with an attractive and innovative long distance program. The company enjoys increased market share, increased revenue, and increased profits.*

Time-to-market will be even more important as we enter the third millennium, and even faster-changing market conditions. As demonstrated in the example above, the Hitachi Business Intelligence Solution can help enterprises achieve it.

II. Performance Analysis of Typical Business Queries

This section describes in detail the relevant queries for most of the industrial segments, and discusses how the Hitachi Business Intelligence Solution solves business problems and improves time-to-market. Our analysis of three typical business queries used in the TPC-D benchmark shows that the Hitachi Business Intelligence Solution delivers excellent query times. Using the Hitachi VisionBase Server and Freedom Storage in departmental data mart environments can help the enterprise achieve its time-to-market objectives.

TPC-D Query 1—Pricing Summary Report

The most classic query of all business industries is Query 1, the Pricing Summary Report query. Most executives, business analysts, and sales managers use this query every day without knowing how much it is affecting their time, their finances, and, most of all, their ability to make fast, effective business decisions.

The question stated by Query 1 is: “What is the amount of business that was billed, shipped, and returned as of a specific date?” The Structured Query Language (SQL) inquiry would be set up as shown in Figure 5.

```

***** Query 1 *****
SELECT
L_RETURNFLAG,
L_LINESTATUS,
SUM(L_QUANTITY) AS SUM_QTY,
SUM(L_EXTENDEDPRICE) AS SUM_BASE_PRICE,
SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)) AS SUM_DISC_PRICE,
SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)*(1+L_TAX)) AS SUM_CHARGE,
AVG(L_QUANTITY) AS AVG_QTY,
AVG(L_EXTENDEDPRICE) AS AVG_PRICE,
AVG(L_DISCOUNT) AS AVG_DISC,
COUNT(*) AS COUNT_ORDER
FROM TPCD.LINEITEM
WHERE L_SHIPDATE <= DATE('1998-12-01') - 91 DAYS
GROUP BY L_RETURNFLAG, L_LINESTATUS
ORDER BY L_RETURNFLAG, L_LINESTATUS

***** Query 1 Result *****
L_RETURNFLAG  L_LINESTATUS  SUM_QTY  SUM_BASE_PRICE  SUM_DISC_PRICE
-----
A              F              1133937013.000  1700303652064.620  1615269944777.963
N              F              29632123.000   44442020110.810   42218880032.734
N              O              2231348409.000  3345844093451.869  3178551620791.149
R              F              1133820213.000  1700118095894.120  1615115765863.198

SUM_CHARGE    AVG_QTY    AVG_PRICE    AVG_DISC    COUNT_ORDER
-----
1679884347984.782  25.503    38241.015    0.050      44462828
43908921788.353   25.494    38235.089    0.050      1162336
3305704374401.567  25.502    38239.137    0.050      87497898
1679720660097.126  25.500    38236.481    0.050      44463247
*****

```

Figure 5: Query 1—Pricing Summary Report

Looking at this query, one might ask:

- “What exactly does this query do to support my decision?”
- “How quickly can I review these summary results?”
- “How rapidly can I submit the same inquiry with different L_SHIPDATE predicate before my morning coffee gets cold?”

This classic query is CPU-bound. It will consume CPU resource as fast as possible while saturating the front-end disk I/O subsystem, regardless of whether the channel is a 17MB/sec ESCON[®], 80MB/sec Ultra-SCSI, or 100MB/sec fibre channel. To all DBMS optimizers, this query is a single table space scan with maximum number of pages per sequential prefetch I/O (16 8K-pages or 32 4K-pages).

Traditionally, this single table space scan goes at the slowest speed of the CPU host, the front-end channel bandwidth, or the subsystem disk I/O sequential prefetch. The faster the slowest speed, the faster the query scans. The Hitachi VisionBase Server and Freedom Storage offer just that—the fastest single table space query scan that delivers intelligent data for faster business decision making.

Figure 6 compares the performance of Hitachi and Sun Microsystems in the equivalent TPC-D 30GB benchmark.

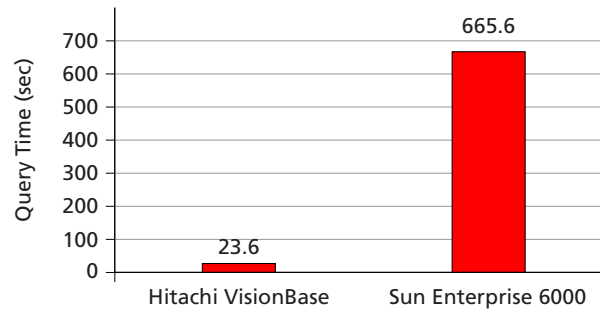


Figure 6: Comparison of Hitachi and Sun Performance for TPC-D Pricing Summary Report Inquiry

How does Hitachi accomplish this performance level? The use of RAID 5 ensures that all data is nicely striped through four data disks in a RAID group. Upon request of a sequential prefetch I/O, several pages are prestaged at the speed of four disks. The use of high-performance 18GB disk drives, combined with the ability to pre-size LUN prefetch buffer size, provides higher locality and optimizes the speed of the disk I/O sequential staging.

The LUN prefetch buffer size can be set at different values (512, 256, 1024, 2048, 4096), depending on the application and the RDBMS used. For DB2 in a DSS environment, with 8K blocksize and 16 blocks per prefetch operation, it has been determined that a value of 2048 is the most effective for Hitachi Freedom Storage.

The Freedom Storage dual controller supports a double number of LUN partitions. This results in a twofold increase in disk I/O parallelism. Each controller is assigned to serve I/O requests for specific assigned LUNs to maximize the bus bandwidth. Figure 7 shows the general design of the Hitachi VisionBase Server and Freedom Storage.

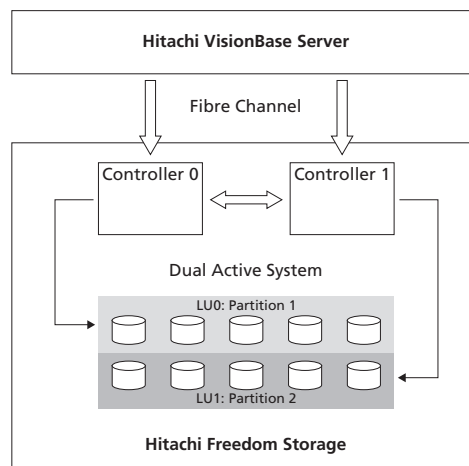


Figure 7: Hitachi VisionBase Server and Freedom Storage Design

The Hitachi VisionBase Server and Freedom Storage use a pre-scan (AST) technique from DB2 UDB V5.2. This reproduces the same kind of results achieved in the pricing summary query, and helps better decision making. Customers are able to plot these data points quickly to gain a better picture of their entire business. Other RDBMS vendors also offer similar functions.

TPC-D Query 6—Forecasting Revenue Change

Another classic business query is Query 6, Forecasting Revenue Change. This query is relevant to all business industry segments and is probably the most important because it is the backbone query of a dynamic business, helping to answer questions such as:

- “What is the return on my investment?”
- “Could I increase the profit margin by 3%?”
- “What would happen to sales revenue if I do this?”
- “With the same business pattern, will I be able to compete financially within the next couple of years?”

These questions could surface from a wide range of related departments. The people who benefit most from such decision support capabilities are strategic business planners and company executives.

The question in Query 6 is: “How much would my total revenue have increased during 1995 if the discount range of 8% to 10% had been eliminated for line items where the quantity is smaller than 24?” The Structured Query Language is shown in Figure 8.

```

***** Query 6 *****
SELECT
SUM(L_EXTENDEDPRICE*L_DISCOUNT) AS REVENUE
FROM TPCD.LINEITEM
WHERE L_SHIPDATE >= DATE('1995-01-01')
AND L_SHIPDATE < DATE('1995-01-01') + 1 YEAR
AND L_DISCOUNT BETWEEN 0.09 - 0.01 AND 0.09 + 0.01
AND L_QUANTITY < 24

***** Query 6 Result *****
REVENUE
-----
5551161383.511
*****

```

Figure 8: Query 6—Forecasting Revenue Change

Query 6 calculates the possible revenue increase for all line items shipped within a given year if a certain discount had not been offered. Chances are that this kind of query will involve a very large table with the return of a single column answer. Because of the predicates applied on index columns with small ranges of index scan, the DBMS optimizer will most likely choose a single index scan to return the answer set.

Understanding the importance of complex business inquiry, Hitachi offers the Hitachi VisionBase Server and Freedom Storage, the fastest, most robust, and the best price/performance DSS for 30GB data marts.

As demonstrated in the TPC-D benchmark with DB2 UDB V5.2, the Hitachi VisionBase Server and Freedom Storage execute a business inquiry like Query 6 considerably faster than any other business intelligence system in this category.

Figure 9 shows how the Hitachi solution compares against its nearest competitor in a 30GB DSS arena.

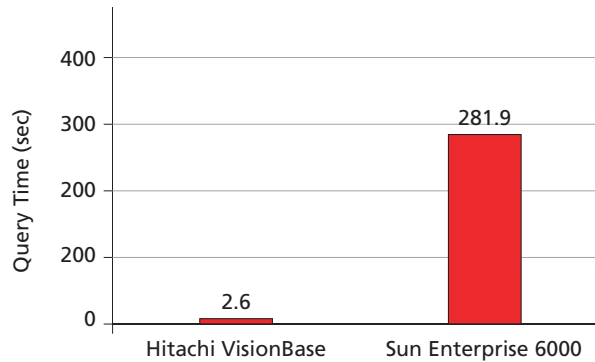


Figure 9: Comparison of Hitachi and Sun Performance in Forecasting Revenue Change Inquiry

TPC-D Query 9—Product Type Profit Measure

Query 9 is most crucial to all business segments and also the most complex one in the TPC-D benchmark. The query’s objective is to find the possible hidden profit for a certain product bought from a certain vendor and sold to every national region in a targeted year. The results assist executives in making better decisions on future product development, profitable product/customer identification, and vendor selection.

The relevant business question posed by Query 9 is: “How much profit did we make on a certain product shipped to a particular national region on a given year where the product name includes the color magenta?” For example, from this answer set, the product name with color “magenta” did very well in the following markets: China, Saudi Arabia, and Vietnam.

```

***** Query 1 Result *****
SELECT NATION, YEAR,
SUM(AMOUNT) AS SUM_PROFIT FROM
(SELECT N_NAME AS NATION,
YEAR(O_ORDERDATE) AS YEAR,
L_EXTENDEDPRI*(1-L_DISCOUNT)-PS_SUPPLYCOST*L_QUANTITY AS AMOUNT
FROM TPCD.PART, TPCD.SUPPLIER, TPCD.LINEITEM, TPCD.PARTSUPP, TPCD.ORDERS,
TPCD.NATION
WHERE S_SUPPKEY = L_SUPPKEY AND PS_SUPPKEY = L_SUPPKEY
AND PS_PARTKEY = L_PARTKEY AND P_PARTKEY = L_PARTKEY
AND O_ORDERKEY = L_ORDERKEY AND S_NATIONKEY = N_NATIONKEY
AND P_NAME LIKE '%magenta%') AS PROFIT
GROUP BY NATION, YEAR
ORDER BY NATION, YEAR DESC

```

```

***** Query 9 Result *****
NATION          YEAR          SUM_PROFIT
-----
CHINA           1998          815931738.009
CHINA           1997          1397343330.547
CHINA           1996          1398078233.656
SAUDI ARABIA   1997          1388869287.856
VIETNAM        1996          1383204022.691
*****

```

Figure 10: Query 9—Product Type Profit Measure

Query 9 is the most difficult, yet also the best of breed of all complex database I/O operations. The most complex data access path is likely to be chosen for this query. A six-table join operation selection will result in several steps of table join and sort with both sequential prefetch of multiple 8K-pages and multiple random index scan of single 8K-page. Workfile is used to aid sort/join operations. Both CPU and/or disk subsystem I/O are likely to cause major bottlenecks.

With the RAID 5 architecture, the Hitachi Business Intelligence Solution offers the best configuration for this complex workload. As mentioned previously, the use of RAID 5 on high-performance disk drives, together with the dual controller partitioning scheme, yields the best sequential prefetch performance. The large cache available on the Hitachi Freedom Storage 5800 minimizes the response time of write activity for the workfile tablespace. Also, in this query, there is a high degree of index scan operation against a small number of 8K-pages for each I/O operation. Superior response time is achieved through the use of sophisticated caching algorithms.

Figure 11 compares the performance of the Hitachi and Sun solutions.

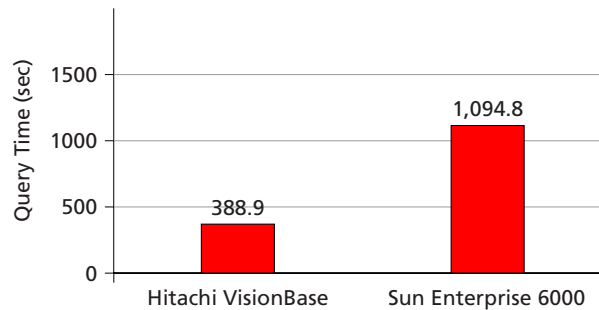


Figure 11: Comparison of Hitachi and Sun Performance in Product Type Profit Measure Inquiry

III. The Hitachi Business Intelligence Solution Improves Time-to-Market

As enterprises prepare marketing and sales campaigns to make a successful entry into the next millennium, time-to-market is the crucial factor. The enterprise that discovers new relationships in customers' buying patterns, and reaches potential buyers first through creative and targeted marketing and sales campaigns, will increase market share and reap great economic benefits.

The Hitachi Business Intelligence Solution enables enterprises to obtain the best response time and throughput from their departmental data marts. The high I/O subsystem performance is essential for enterprises to achieve their time-to-market objectives. Both the Hitachi VisionBase Server and the Hitachi Freedom Storage 5800 offer superior reliability—the best in the industry.

We have demonstrated that the industry-leading TPC-D results achieved with the Hitachi Business Intelligence Solution can lead directly to improved time-to-market. With the Hitachi Data Systems solution, the DSS user is able to make a more effective decision **sooner**. The Hitachi Business Intelligence Solution helps dynamic enterprises enter the next millennium with a high degree of confidence.

Hitachi Data Systems will continue to enhance Hitachi Freedom Storage. These enhancements will be tested and then tied to improvements in the Hitachi VisionBase Server. The benefits of their combined performance will be evaluated and measured so that Hitachi Data Systems can continue to provide superior value and faster time-to-market.

Appendix A. The Transaction Processing Performance Council

The Transaction Processing Performance Council (TPC) is a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable TPC performance data to the industry. While TPC benchmarks certainly involve the measurement and evaluation of computer functions and operations, the TPC views a transaction as it is commonly understood in the business world: a commercial exchange of goods, services, or money. A typical transaction, as defined by the TPC, would include the updating to a database system for such things as inventory control (goods), airline reservations (services), or banking (money).

There are four types of benchmarks. Each addresses a specific computer workload.

1. TPC-B

This benchmark is designed to measure a batch environment, but is not used anymore.

2. TPC-C

This benchmark is designed to measure the Online Transaction Processing (OLTP) environment. It is structured to measure server performance. The benchmark queries tend to be CPU-bound. Normally, an OLTP workload uses intensive indexing. OLTP transactions such as random select, update, insert, and delete will, most of the time, result in small numbers of pages being moved per transaction. As a result, OLTP will not produce a large number of disk I/Os, but a high number of concurrent transactions can consume a substantial amount of CPU power. Therefore, the OLTP environment associated with high overhead of multiple concurrent transactions usually saturates the server's CPU resource before it stresses the I/O subsystem.

3. TPC-D

This benchmark focuses on decision support systems. TPC-D queries measure both the server and the storage performance of a complete solution. This is the most popular benchmark and, as a result, any vendor can be ranked based on their TPC-D results. The DSS workload is one that is highly sequential by nature. However, indexing techniques can be used to improve query elapsed time.

The DSS environment is somewhat the opposite of the OLTP one in the sense that DSS uses a high-level sequential prefetch and large indexes combined with various complex sort and join operations. This usually means the disk IO will be saturated while CPU resource is highly utilized, especially with a very large database (VLDB).

Since decision support systems are now growing into two distinct categories (Business Report and Ad Hoc queries), TPC is considering breaking down the TPC-D benchmark into two new benchmarks: TPC-H for Ad Hoc queries and TPC-R for Business Reports.

4. TPC-W

A new web e-commerce benchmark (TPC-W) is expected to be available from TPC in 1999. This benchmark is designed to represent any business (retail store, software distribution, airline reservation, electronic stock trades, and so on) that markets and sells over the Internet. It also represents Intranet environments that use Web-based transactions for internal operations. It measures the performance of systems that support users in browsing, ordering, and conducting transaction-oriented business activities. Security (including user authentication and data encryption) and dynamic page generation are important elements of TPC-W.

Appendix B. TPC-D Query Relevance by Industry Segment

	Description	Retail	TelCo	Brokers	Utility	Airlines	Banking
Query 1	Pricing Summary Report	X	X	X	X	X	X
Query 2	Minimum Cost Supplier	X	X		X	X	
Query 3	Shipping Priority	X	X		X		
Query 4	Order Priority Checking	X	X	X	X	X	
Query 5	Local Supplier Volume	X	X		X	X	
Query 6	Forecasting Revenue Change	X	X	X	X	X	X
Query 7	Volume Shipping	X	X		X		
Query 8	National Market Share	X	X	X	X	X	X
Query 9	Product Type Profit Measure	X	X	X	X	X	X
Query 10	Returned Item Reporting	X	X			X	X
Query 11	Important Stock Identification	X	X				X
Query 12	Shipping Modes & Order Priority	X	X				X
Query 13	Customer Distribution	X	X	X	X	X	X
Query 14	Promotion Effect	X	X	X	X	X	X
Query 15	Top Supplier	X	X		X	X	
Query 16	Parts / Supplier Relationship	X	X		X	X	
Query 17	Small Quantity Order Revenue	X	X	X			
Query 18	Large Volume Customer	X	X	X	X		
Query 19	Discounted Revenue Query	X	X	X		X	X
Query 20	Potential Part Promotion	X	X	X	X	X	X
Query 21	Suppliers Who Kept Orders Waiting	X	X	X	X	X	X
Query 22	Global Sales Opportunity	X	X	X	X	X	X

Appendix C. TPC-D Performance and Price/Performance Metrics

The Transaction Processing Performance Council presents the TPC-D benchmark results using multiple metrics. A brief overview of these metrics is presented here.

The first metric is the power metric (QppD), which compares the raw power of the solutions being measured. It starts by running a series of inserts to the main tables (line item and orders) and then runs the 22 TPC-D queries in a very specific sequence. At the end, it runs a series of deletes, again, on the main tables.

The second metric is the throughput metric (QthD). This metric measures the throughput capabilities of the solution being measured. It runs multiple series of the power test, again, in a very specific order.

The third metric is the (QphD) or queries per hour that the total solution can support. This value is calculated as follows:

$$\sqrt{QppD \times QthD}$$

The fourth metric is the \$/QphD or \$ per query per hour. This value is calculated by dividing the total solution cost by the QphD metric. The lower this value, the better the overall price/performance for the total solution will be. The objective here is to provide a solution that will deliver the maximum performance with the lowest possible budget.

Appendix D. Glossary

AST – Automatic Summary Table.

DSS – Decision Support System.

ESCON – Enterprise Systems CONnectivity, IBM's formerly proprietary, serialized, mainframe channel technology.

Fibre Channel – The name given to a very high-speed serial interface standard developed as an ANSI standard, which can be implemented over optical fiber as well as copper wires.

LUN – Logical Unit Number. This is the lowest addressable increment of storage used by SCSI file systems.

Page – A database unit of storage consisting of rows of data.

Prefetch – Operation consisting of reading sequential pages ahead of time.

QppD – The power performance metrics of the TPC-D benchmark that measure the raw query execution power of the system.

QthD – The throughput performance metrics of the TPC-D benchmark that measure the ability of the system to process the most queries in the least amount of time.

QphD – A composite query-per-hour rating based on the QppD and QthD metrics.

\$/QphD – The price/performance metric of TPCD-D benchmark that measures the maximum performance at the lowest possible cost.

RAID – Redundant Array of Independent Disks. A storage system that contains several physical disks, which operates input/output instructions in parallel or independently.

RDBMS – Relational Database Management System.

SF – Scale Factor of the database size of the TPC-D benchmark.

SMP – Symmetrical Multiprocessors.

Table-join – A relational database mechanism that retrieves information (rows) from multiple tables based on matching columns.

TPC – Transaction Processing Performance Council.

TPC-B – A batch environment benchmark that is developed and owned by the Transaction Processing Performance Council.

TPC-C – An Online Transaction Processing benchmark consisting of a suite of business queries, developed and owned by the Transaction Processing Performance Council.

TPC-D – A decision support benchmark consisting of a suite of business queries and updates developed and owned by the Transaction Processing Performance Council.

TPC-W – A World Wide Web benchmark to measure the number of Web hit pages. It is being developed by the Transaction Processing Performance Council and will be available in the near future.

UDB – A Universal Database for different platforms (NT, AIX,[®] OS/390[®]) developed and owned by IBM.

SCSI – Small Computer System Interface.

VLDB – Very Large Database.

WWW – World Wide Web.

Hitachi Data Systems

www.hds.com

Corporate Headquarters
750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
(408) 970-1000
info@hds.com

Asia Headquarters
39-09 Tower One
Lippo Centre
89 Queensway
Hong Kong
2525-2385
infoasia@hds.com

Australia/New Zealand Headquarters
11-17 Khartoum Road
North Ryde NSW 2113
Australia
02-9325-3300
info@hds.com.au

Canada Headquarters
380 Saint-Antoine Street West
Suite 7000
Montreal, Quebec H2Y 3X7
Canada
(514) 982-0707
info@hdscanada.com

Europe Headquarters
Sefton Park
Stoke Poges
Buckinghamshire SL2 4HD
United Kingdom
01753-61-8000
info@hds.co.uk

Latin America Headquarters
750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
(408) 970-7447
lad@hds.com

U.S. Headquarters
750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
(408) 970-1066
ussales@hds.com

Hitachi Data Systems is registered with the U.S. Patent and Trademark Office as a trademark and service mark of Hitachi, Ltd. The Hitachi Data Systems logotype is a trademark and service mark of Hitachi, Ltd.

VisionBase, Freedom Storage, and Skyline Trinium are trademarks of Hitachi Data Systems Corporation.

S/390, OS/390, ESCON, and AIX are registered trademarks of International Business Machines Corporation. AIX is being used under license.

All other brand or product names are or may be trademarks of, and are used to identify, products or services of their respective owners.

Notice: This document is for informational purposes only, and does not set forth any warranty, express or implied, concerning any equipment or service offered or to be offered by Hitachi Data Systems. This document describes some capabilities that may be configuration-dependent, and features that may not be currently available. Contact your local Hitachi Data Systems sales office for information on feature and product availability.

©1999, Hitachi Data Systems Corporation.
All Rights Reserved/LD3M069/CORP-296