June 2011

MANAGEMENT BRIEF

Cost/Benefit Case for Enterprise Warehouse Solutions

In-depth Comparison of IBM Smart Analytics System 7700, Teradata Active Enterprise Data Warehouse and Oracle Exadata Database Machine

International Technology Group

609 Pacific Avenue, Suite 102
Santa Cruz, California 95060-4406
Telephone: + 831-427-9260
Email: Contact@ITGforInfo.com
Website: ITGforInfo.com

TABLE OF CONTENTS

List of Figures

EXECUTIVE SUMMARY

Setting Parameters

Data warehousing has emerged as one of the IT world's fastest growth areas. New deployments continue to accelerate, and numbers of applications and users within organizations continue to expand. Demand for high-quality, current information and for tools to interpret and exploit it shows no signs of abating. High double-digit growth in data volumes has become the norm.

The business benefits of data warehouse applications are clearly recognized. But, increasingly, users are faced with escalating expenditure not only on data warehouse solutions, but also on underlying platforms. At a time of budgetary pressures, questions are raised about the most cost-effective means of realizing information value.

This is particularly the case for special-purpose platforms offered by IBM (Smart Analytics System, Netezza TwinFin), Oracle (Exadata Database Machine), Teradata (Active Enterprise Data Warehouse) and smaller players. Architectures and technologies of these systems are often unfamiliar to organizations that deploy them. Techniques for measuring comparative performance and cost are rudimentary.

Challenges are compounded by several factors. One is that the performance of different architectures depends on the workloads they execute. Another is that data warehouse usage tends to evolve rapidly – organizations that deploy platforms for specific applications may soon find that they must deal with significantly different environments. A third is that vendor pricing may vary widely between customers.

This report sets some parameters for comparisons. To do this, it takes into account types of workload – in particular, a key distinction is drawn between complex mixed workloads and queries involving large sequential table scans – compares overall three-year as well as acquisition costs, and bases platform calculation on "street" pricing (i.e., discounted prices paid by users).

The report focuses on three platforms: IBM Smart Analytics System 7700, Oracle Exadata Database Machine and Teradata's flagship Active Enterprise Data Warehouse (Active EDW) 6650. Results are based on input from 46 users of these systems and their recent predecessors, on other industry sources, as well as on research and analysis conducted by the International Technology Group (ITG).

Two sets of cost comparisons, based on performance and user data, are presented.

Performance-based Comparisons

Based on relative performance for complex mixed workload environments, IBM Smart Analytics System is a clear leader in cost-effectiveness.

In certain configurations offering approximately the same performance, initial Smart Analytics System 7700 costs – including hardware and software acquisition, and installation – are 11 percent less than for Oracle Exadata Database Machine and 16 percent less than for Teradata Active EDW 6650 equivalents.

However, disparities are significantly larger if three-year costs – including maintenance and support, database administration (DBA) personnel and facilities (energy and computer room occupancy) – are compared. Three-year costs for Smart Analytics System 7700 are 43 and 40 percent less than those for Oracle and Teradata systems respectively.

Larger disparities in three-year costs primarily reflect higher Oracle and Teradata pricing for maintenance and support. Higher personnel and facilities costs also contribute.

Comparisons between the medium Smart Analytics System 7700, full rack Oracle Exadata Database Machine X2-2 and two-node Teradata Active EDW 6650H configurations are shown in figures 1 and 2.

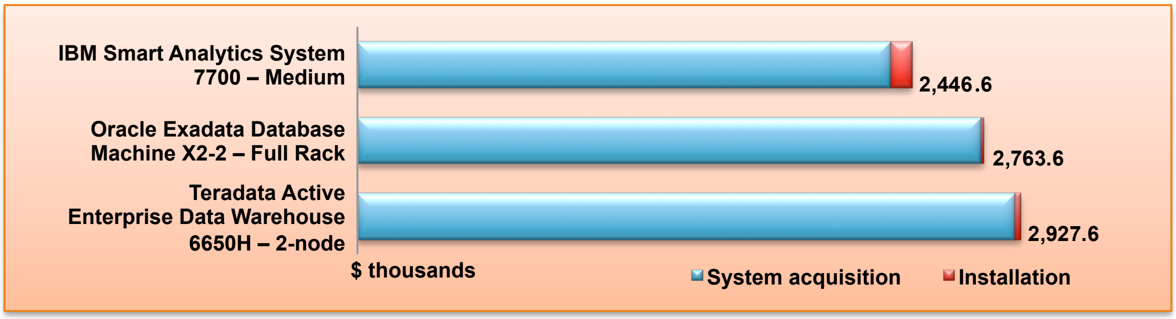


Figure 1: Initial Costs Comparison: Selected IBM Smart Analytics System 7700, Oracle Exadata Database Machine & Teradata Active Enterprise Data Warehouse 6650H Models

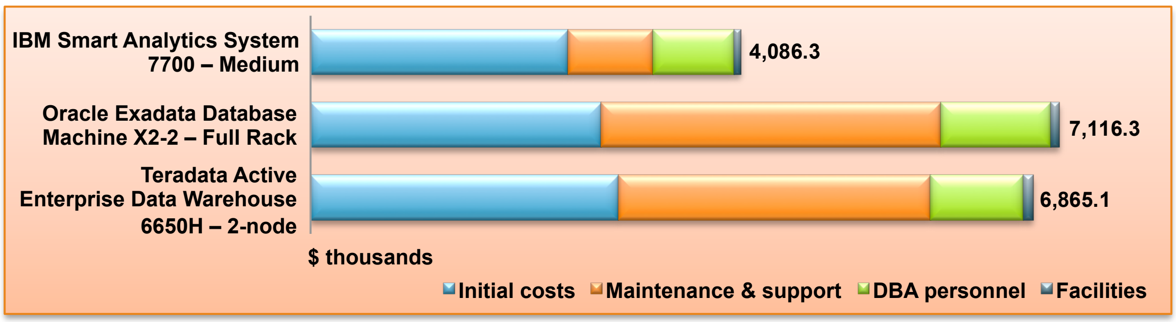


Figure 2: Three-Year Costs Comparison: Selected IBM Smart Analytics System 7700, Oracle Exadata Database Machine & Teradata Active Enterprise Data Warehouse 6650H Models

This picture is largely consistent across other models offered by these vendors. Figures 3 and 4 show initial and three-year costs for the principal Smart Analytics System 7700, Exadata Database Machine and Active EDW 6650 entry-level and midrange models. Models are arranged in approximate order of performance for complex mixed workload environments, as estimated by ITG.

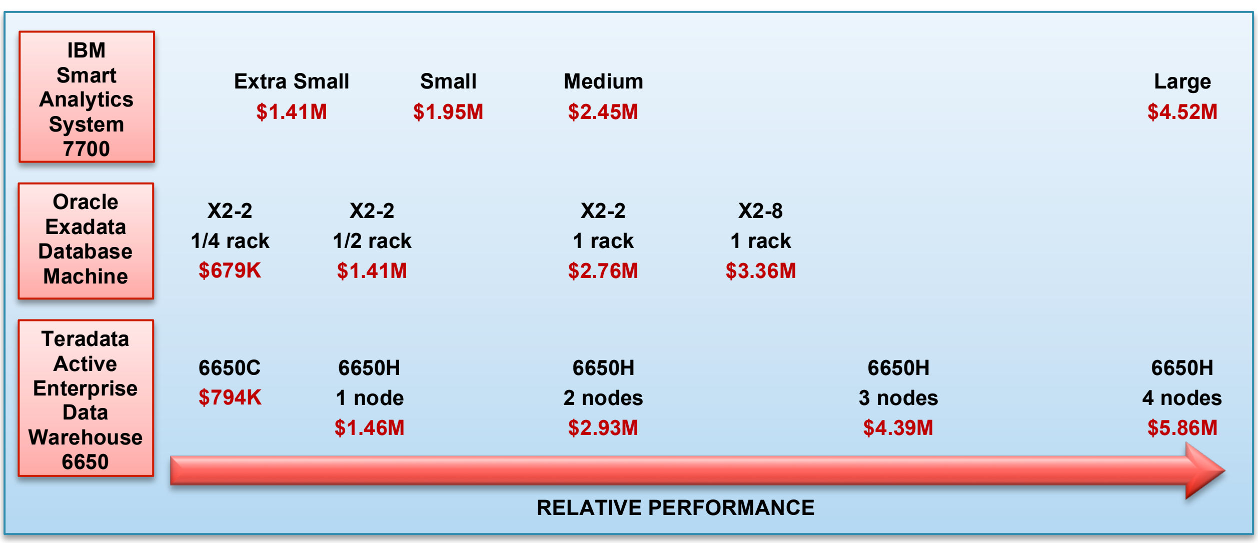


Figure 3: Initial Costs Comparison: Entry-level and Midrange Models

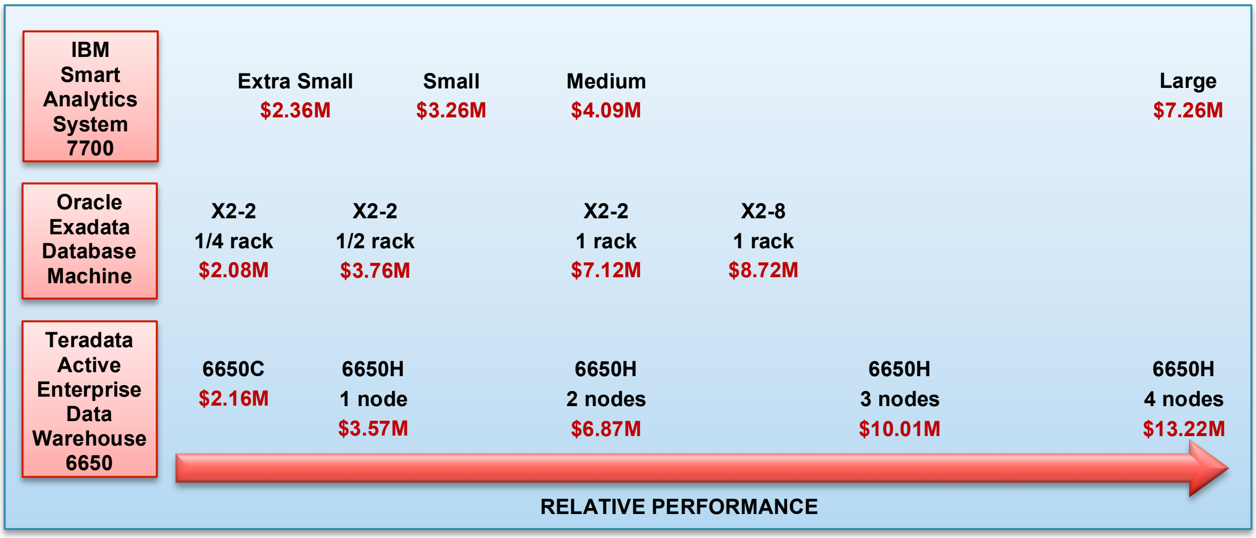


Figure 4: Three-year Costs Comparison: Entry-level and Midrange Models

Initial costs for entry-level Exadata Database Machine X2-2 and Teradata Active EDW 6650C models are comparatively low, although this is not the case for larger models.

All three systems can be configured in larger increments. IBM also offers standard Smart Analytics System 7700 models supporting up to 220TB of user data, and larger non-standard configurations.

Teradata 6650 systems can in principle be configured with up to 4,096 nodes, and there are production installations of earlier Teradata systems with hundreds of nodes. For reasons discussed later in this report, the practical Exadata Database Machine configuration limit is unclear.

User Data Comparisons

A second approach to comparing costs of special-purpose systems is to measure costs per terabyte (TB) of user data, meaning the amount of space available for customer data net of indexes, log files, temporary spaces and other structures.

If user data capacities are measured in non-compressed data, IBM Smart Analytics System 7700 configurations show the lowest three-year cost per terabyte, followed by Teradata. Although Teradata systems tend to be more expensive, user data capacities as reported by the company are higher than for Exadata Database Machine equivalents.

If, however, capacities are measured in compressed data, the gap between Smart Analytics System 7700, and Exadata Database Machine and Active EDW 6650 models widens. Smart Analytics System 7700 costs are two to four times less than for competitive systems.

Smart Analytics System 7700 three-year costs per terabyte for non-compressed data average 57 percent less than for Exadata Database Machines, and six percent less than for 6650 systems. For compressed data, three-year costs per terabyte average 72 and 54 percent less respectively.

Although both have the same user data capacity, the full rack Oracle X2-8 model is equipped with greater database server processing capability than its X2-2 counterpart.

Figures 5 and 6 summarize these results.

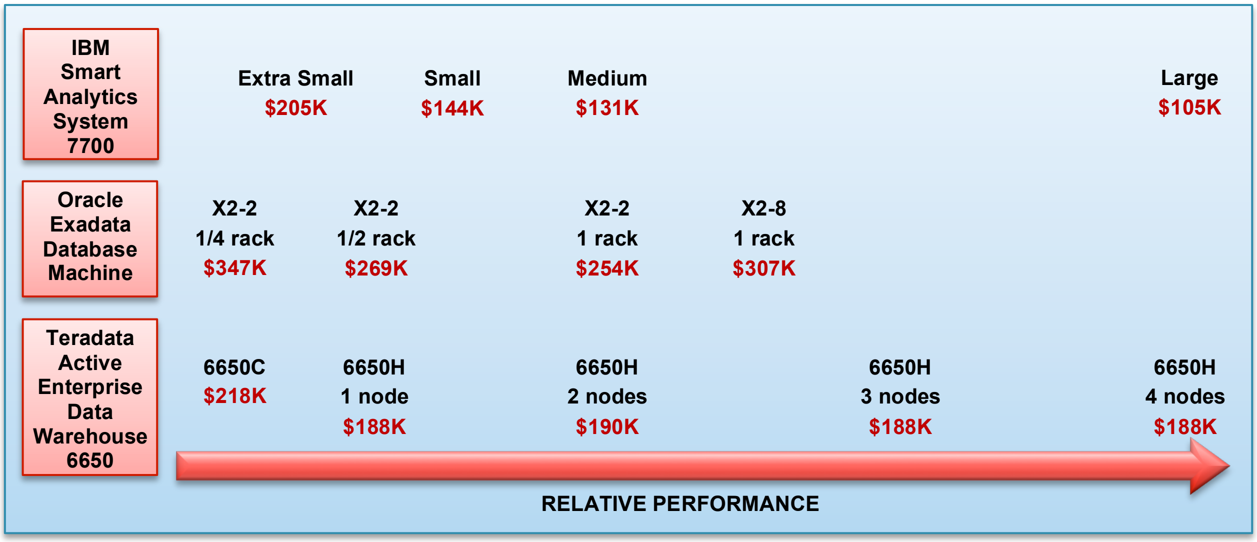


Figure 5: Three-year Costs Comparison per Terabyte of User Data (Non-compressed)

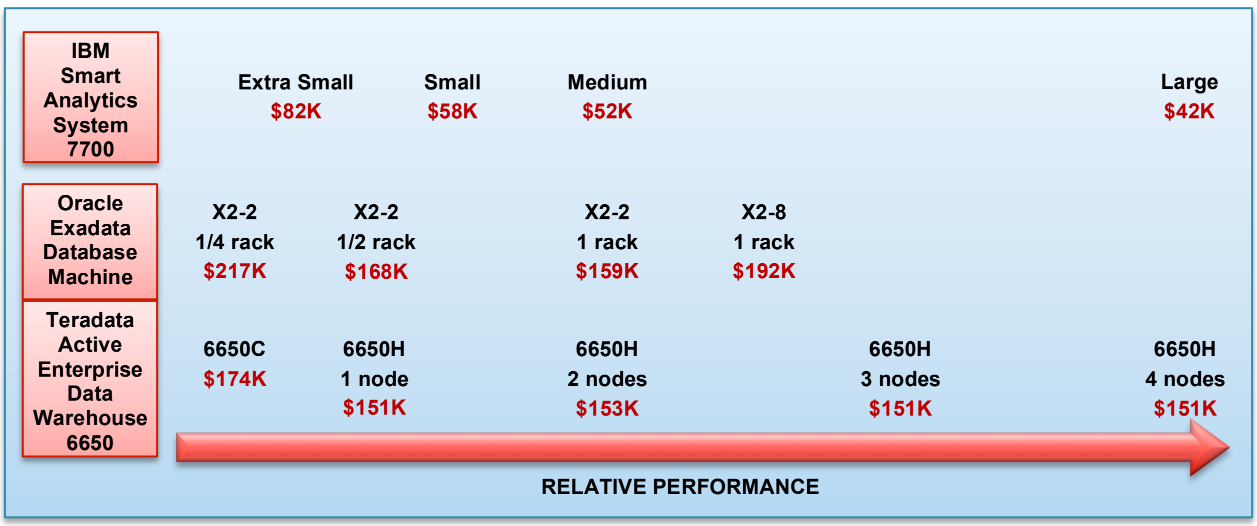


Figure 6: Three-year Costs Comparison per Terabyte of User Data (Compressed)

In these calculations, non-compressed user data capacities for Smart Analytics System 7700 and Active EDW 6650 models are as given by their vendors. Capacities for Exadata Database Machine models were calculated as 55 percent of physical capacity using 600-gigabyte (GB) disk drives.

Compressed user data capacities are based on assumptions of 60 percent compression for Smart Analytics System 7700, 25 percent for Active EDW 6650 and 37.5 percent for Exadata Database Machine models.

For Exadata Database Machines, it is assumed that around 30 percent of database workloads involve table scan-intensive data mining queries, for which a compression level of 3:1 is realized using Exadata Hybrid Columnar Compression (EHCC) technology. For the remaining 70 percent of workloads, 25 percent compression is realized using Oracle Database 11g Advanced Compression. The overall level is thus 37.5 percent. Actual workload mixes and compression levels may vary in practice.

Performance and user data comparisons are for use of conventional hard disk rather than flash drives.

Conclusions

Cost comparisons presented here are based on typical configurations, utilization and staffing levels, along with street prices reported by users for data warehouses characterized by complex mixed workloads. In practice, configurations and applications vary between and in some cases within organizations, and vendors may price more aggressively in genuinely competitive situations.

Certain conclusions may nevertheless be drawn. The economics of special-purpose data warehouse systems are affected not only by pricing, but also by how well architectures handle specific types of workload and by levels of data compression that may be achieved in practice. Personnel costs are also affected by the extent of automation, and facilities costs by the efficiency with which systems operate.

From these perspectives, key distinctions must be drawn among the three platforms that are the focus of this report. In terms of underlying architecture, system design and hardware and software implementation, Smart Analytics System 7700 and Active EDW 6650 are better optimized – by wide margins – to handle complex mixed workloads than Exadata Database Machines.

The level of optimization is highest for Smart Analytics System 7700, which employs a newer architecture. Teradata systems are more constrained by legacy characteristics.

The capabilities of Exadata Database Machines, however, are significantly different from Smart Analytics System 7700 and Teradata Active EDW equivalents. This is particularly the case in the following areas:

- ***Workloads***. Exadata Database Machines deliver their best performance for workloads characterized by large sequential table scans.

  Such workloads are generated by applications that are structurally simple, but require a great deal of processing power; e.g., identifying and collating specific variables in large volumes of records. These applications typically support small numbers of executives and/or analysts.

  The entire Exadata Database Machine architecture is geared to this type of applications and workload. This is the case for the overall system design as well as for three key technologies – (1) Smart Scan, (2) Exadata Hybrid Columnar Compression and (3) Smart Flash Cache – presented by Oracle as Exadata Database Machine differentiators.

  The emphasis on large sequential table scans reflects, at least to some extent, earlier development of the "data warehouse appliance" market. By the mid-2000s, Teradata's dominant market position in special-purpose systems was being eroded by Netezza, which offered lower-cost appliances built around "commodity" components.

  Netezza systems were rapidly adopted by many organizations – by September 2010, the company claimed 373 installed customers – for applications generating this type of workload. When Oracle Exadata Database Machine was introduced, it was generally seen as aimed at Netezza.

  Although Oracle has since positioned Exadata Database Machines more broadly, design for high-volume table scan processing remains a fundamental characteristic.

- ***Compression***. The extent to which systems can compress data without unacceptable performance degradation has a major impact on system capacities and, if measurements are based on user data, comparative costs. System processing may also be accelerated, and I/O throughput times reduced.

  IBM DB2 9.7, which is employed by Smart Analytics System 7700, features one of the industry's most effective across-the-board implementations of data compression. Compression extends to rows as well as indexes, temporary tables, log files, large objects and other data structures. Users have routinely experienced overall compression levels of 55 percent to more than 85 percent.

Among Smart Analytics System 7700 users who contributed to this report, for example, overall compression levels averaged 72 percent. As a general metric, IBM employs a 60 percent ratio.

Teradata expanded compression in version 13.10 of its database. This was introduced only in November 2010, and no Teradata users contacted for this report had practical experience with its compression capabilities.

Oracle employs two compression technologies. Database Machines implement Advanced Compression, which is a feature of Oracle Database 11g. Although higher levels have been achieved, users have found that unacceptable performance degradation typically occurs above approximately 25 percent.

Exadata Storage Servers implement EHCC technology, which is designed to compress large tables, and is most effective when these tables are processed sequentially. Oracle has claimed compression rates of up to 70 times. Among users, rates of two to three times have been reported. EHCC does not, however, have a similar effect for other data structures and types of workload.

- ***Scalability***. Experience has shown that Smart Analytics System 7700 and Teradata systems can scale to very large configurations. Teradata systems routinely contain dozens – in some cases, hundreds – of nodes. Among Smart Analytics System 7700 users, seven production systems were reported to support 40TB to more than 120TB of user data. Larger systems were planned.

  Oracle Database Machines implement RAC architecture, which is built around a "shared everything" model; i.e.; multiple processors share a common memory pool. In comparison, Smart Analytics System 7700 and Teradata systems employ "shared nothing" architecture, in which these components are subdivided into nodes. The Exadata Storage Server also employs a "shared nothing" model.

  "Shared everything" models are generally regarded as more vulnerable to scalability constraints – contention for system resources increases as systems expand. It is unclear whether this is the case for Exadata Database Machines, as deployments to date have been of relatively small configurations. Users have reported installations of 1/2 to 2 racks, with 3TB to 10TB of user data.

- ***Automation***. IBM Smart Analytics Systems and Teradata systems typically require fewer DBAs and other IT personnel than Oracle equivalents. Levels of automation are significantly higher for DB2 9.7 and (to a lesser extent) Teradata 13.10 than for Oracle Database 11g.

  The impact on FTE staffing levels is magnified when databases undergo frequent changes – which is typically the case when numerous applications, diverse user populations and complex workloads must be supported. Users of all types of special-purpose systems noted this effect.

Performance competitiveness and cost-effectiveness of Exadata Database Machines thus tend to be workload-specific. Users have cited "massive table scans and joins," "very large table scans" and similar terms in describing their workloads. For many organizations, however, the viability of this platform will tend to decline as data warehouses evolve towards more complex mixed workload environments.

All Smart Analytics System 7700 and Teradata users who contributed to this report reported that they had either planned for multifunction data warehouses, or that usage had developed in this direction.

There will clearly still be demand for special-purpose systems to handle large sequential table scans. But users should be aware of the limitations of these. Platform decisions should be based not only on short-term requirements, but also on how organizations expect data warehouse demand to evolve in the future.

PLAYERS AND PLATFORMS

Overview

The concept of data warehouses originated in the 1980s, and many users have employed such systems on symmetric multiprocessing (SMP) and massively parallel processing (MPP) servers for decades.

The term "data warehouse appliance" emerged during the 2000s to describe bundled hardware and software systems offered by Netezza, as well as smaller players such as DATAllegro, Greenplum, Kickfire, Kognitio, ParAccel and Vertica. All but two of these niche vendors have been acquired by larger firms – Netezza, for example, was acquired by IBM in 2010.

Teradata responded with its own low-end "appliance" offerings. IBM and Hewlett-Packard (HP) entered the field with Balanced Configuration Units (BCUs) and Neoview MPP system in 2005 and 2007 respectively. HP announced in January 2011 that it would withdraw Neoview systems from the market. IBM BCUs, renamed Smart Analytics System, continue to be marketed. Oracle's Exadata Database Machine was first introduced in 2008.

There are currently three significant players – IBM (including Netezza), Oracle Exadata Database Machine and Teradata. Systems offered by these employ significantly different architectures, demonstrate varying performance and functional characteristics and are offered at a wide range of price points.

Before examining these systems more closely, however, it is important to place their relative capabilities in the context of evolving data warehouse usage. Although there is still significant demand for dedicated table scan processing, there is widespread agreement that a variety of trends are driving a pervasive shift toward increasingly complex mixed workload environments.

Trends

These have included:

1. ***Demand escalation***. In most organizations, demand for high-quality information and analysis expands rapidly once its value is understood. The general industry experience has been that, once data warehouses are put in place, demand soon extends from specialists to larger populations of managers, professionals and front-line users such as sales and customer service representatives.

   Concepts such as "operational business intelligence (BI)" and "active data warehousing" reflect these shifts. The focus of BI strategies moves toward tactically oriented applications supporting front-line user populations.

   Applications are deployed for multiple business units. User tools to locate, analyze and exploit data grow increasingly sophisticated, and are more widely adopted. Drill-down tools, high-performance analytical suites, dashboard-style interfaces and other capabilities become routine requirements for hundreds or thousands of users.

   A growing number of businesses are also extending access to BI information and tools to partners and, in some cases, to customers through the Internet. Agents, brokers, distributors and other types of partner, in particular, tend to find that access to high-quality information and tools represents a major source of business value.

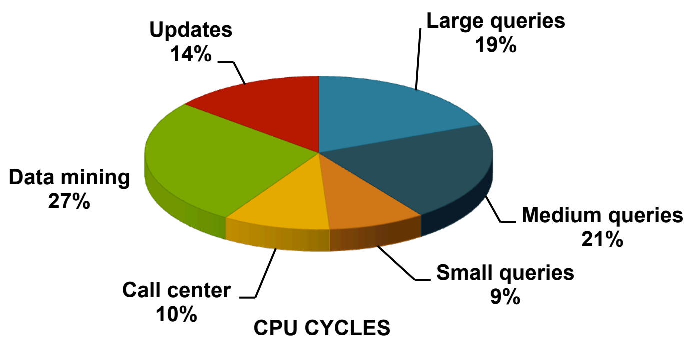
2. ***Query complexity***. Experience has shown that greater exposure to data warehouse capabilities translates into increased query complexity.

A department that starts with, say, an effort to identify customers who fit a certain profile, may move on to more sophisticated exercises such as determining cross-product purchasing patterns or predicting trends in customer behavior. Similarly, financial analysis may evolve from quantifying the profit contribution of individual operating units to developing strategies to maximize profitability across the business as a whole.

In these and other applications, users tend to exploit an increasing number of variables in identifying, retrieving and manipulating information. There is an immediate impact on the complexity of data warehouse workloads.

3. ***Workload diversity***. Growth in numbers of applications, user populations and query sophistication has been reflected in workload mixes. Large organization data warehouses now typically support a wide range of query types and sizes, with varying degrees of time-sensitivity. These commonly range from simple reports to large-scale data mining jobs.

Figure 7, for example, demonstrates one organization's estimate of its overall data warehouse workload during a 24-hour period. Percentages are for numbers of CPU cycles.



CPU EXECUTION TIMES	
Large queries	20 – 60 seconds
Medium queries	15 – 30 seconds
Small queries	0.5 – 10 seconds
Call center	0.2 – 3 seconds
Data mining	Minutes to hours

Figure 7: Distribution of Data Warehouse Workload: Example

In this as in many other cases, data warehouse usage had evolved from relatively simple standardized management reports and ad hoc queries to include large data mining jobs; online analytical processing (OLAP); a variety of strategic and tactical applications for more than 20 business units; operational BI applications supporting call center and sales personnel; and others. No slowdown in the rate of new application growth was anticipated.

In such environments, highly effective workload management capabilities are required. Moreover, organizations are moving toward more aggressive use of prioritization techniques and service level agreements (SLAs) to balance the interests of different user groups.

4. ***Data currency***. Responding to user demands for more current information, data warehouses have experienced a progressive acceleration of refresh cycles.

This trend is reflected in the results of 2008 and 2010 surveys of Fortune 500 data warehouses by ITG. Although monthly cycles continue to be widely employed, the fastest growth has been in update cycles that range from daily to hourly and, in some cases, "near real-time." Figure 8 illustrates these results. Cycles shown are for 2008 and projected for 2011.

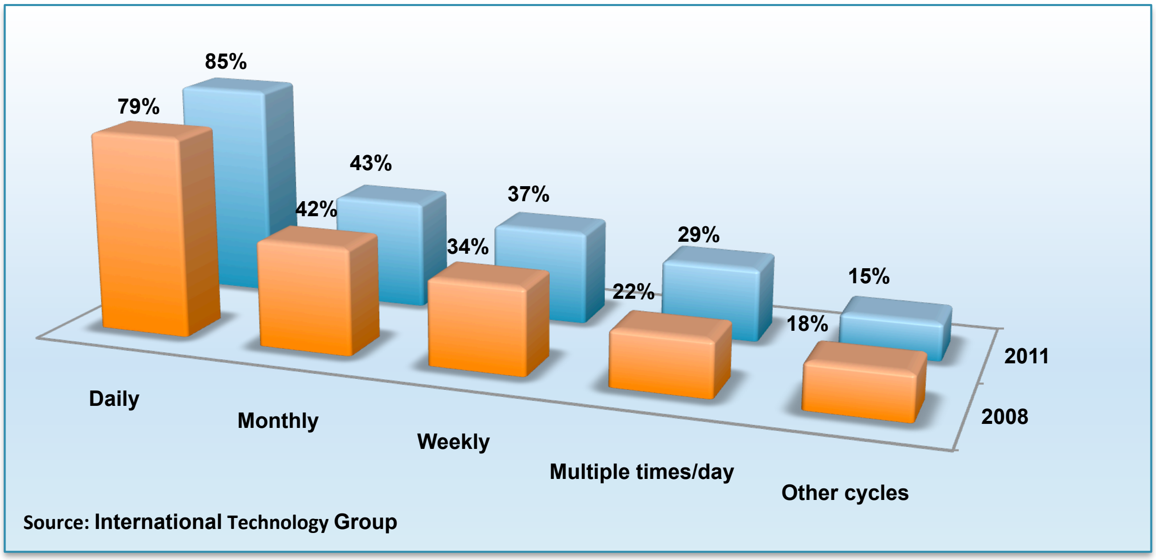


Figure 8: Data Warehouse Refresh Cycles in Fortune 500 Corporations

Pressures for more current data, as well as for faster delivery of query results, have been reinforced by the global economic downturn. Since the onset of recession in 2008, organizations have tended to adopt shorter planning and decision-making cycles that react more closely to changing business conditions. The value of current information has increased accordingly.

These pressures have led to the adoption new update techniques. Approaches include "continuous" batch (e.g., batch ETL operations conducted multiple times per day, rather than overnight), as well as use of staging tables, ongoing trickle feeds and direct input via enterprise application integration (EAI) middleware. Figure 9 illustrates this evolution.

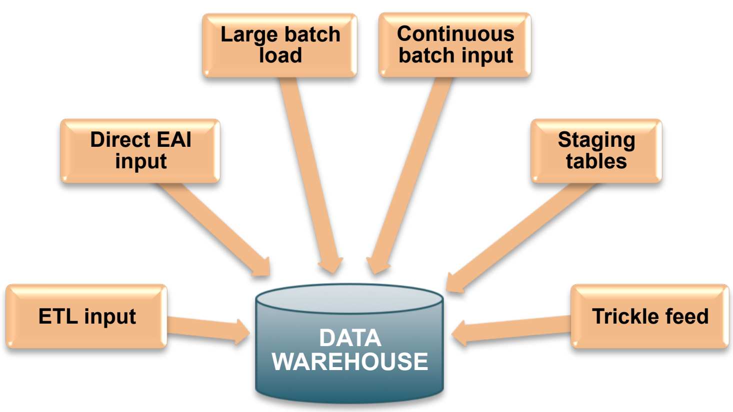


Figure 9: Multiple Data Warehouse Input Methods: Example

The result is that, not only application workloads, but also data movement processes become increasingly diverse, and more spread out over 24-hour cycles. Management challenges increase as this occurs.

As both query and update processes have become more time-sensitive, it has also become increasingly important to avoid unplanned as well as planned system outages. In a growing number of organizations, maintenance of high availability (HA) has become as critical as it is for key transactional systems.

5. ***Data growth***. It is a truism that, in terms of raw data content, data warehouses expand more rapidly than any other major system. High double-digit growth rates are commonly sustained over multi-year periods.

   This trend has been driven not only by the tendency of applications to address larger numbers of records, but also by the growing size and complexity of records themselves. In the early 2000s, for example, banks typically maintained one to two megabytes (MB) of data per customer. By yearend 2010, this had increased to more than 10MB. Other industries have seen similar trends.

   Data growth poses major challenges for data warehouses and the infrastructures that support them. However, these challenges are magnified by the other trends described here. For example, the stresses generated in a complex, continuously updated workload environment are very different from those handled by earlier data warehouse appliances.

Under other circumstances, these trends might have led to proliferation of data warehouse structures. In practice, however, organizations have proved sensitive to the advantages of enterprise-wide data consistency. Another factor has also come into play.

Few businesses can afford the investments in data management; extract, transformation and load (ETL) mechanisms; replication; backup and recovery; and security facilities that would be necessary to support multiple, dispersed data warehouses. Proliferation of underlying database, server, storage and network infrastructures would also soon become prohibitively expensive.

Teradata Active Enterprise Data Warehouse 6650

Overview

Teradata is the largest and longest-established player in the special-purpose systems market. The company, which shipped its first system in 1983, established an early market lead. Since that time, the company has progressively enhanced the Teradata database – version 13 was introduced in 2009, and upgraded with 13.10 in October 2010 – and transitioned to Intel-based hardware.

In the past, the company has claimed more than 1,000 installed customers worldwide, although it currently cites over 900. Most Teradata business is derived from large corporate and government accounts. It enjoys broad support from solution vendors, systems integrators and consulting firms worldwide.

Although Teradata initially focused on high-end analytical applications, it has expanded into a wide range of solutions for customer relationship management (CRM), business performance management (BPM), supply chain management (SCM), enterprise risk management (ERM) and other functions. Overall company strategy, according to management, is to continue to expand its software and services business.

The company has a reputation for robust, reliable platforms along with high-quality software solutions and support. All of these are widely regarded as expensive – the term “expensive” was, in fact, employed by several of the Teradata users surveyed for this report when referring to the company’s offerings.

Revenue streams from established customers tend to be dominated by software subscriptions and services, which accounted for more than half of sales in 2010. The company’s market position enables it to charge higher prices than its smaller and/or newer competitors.

Teradata may price more aggressively for first-time customer placements, although multi-year costs for additional capacity, new model upgrades, software subscriptions and maintenance tend to remain high.

Active EDW 6650 Systems

Introduced in April 2011, Active EDW 6650 systems are the latest versions of the high-end Teradata platform, which the company regularly upgrades with more powerful Intel processors. Until 2010, upgrades typically occurred once a year, but this schedule has recently accelerated.

In October 2010, Teradata introduced 5650 systems based on Intel Xeon 5670 2.93 GHz six-core processors, and in April 2011 followed these with new 6650 and 6680 systems. These employ the same processors, but feature denser packaging and (in the case of 6680 systems) enable mixed use of flash and conventional disk drives.

Current offerings include the entry-level 6650C, which is built around a single Xeon 5670 processor, and flagship 6650H, which is built around two such processors per node. According to Teradata, these typically support 13.8TB and 26TB per node of user data respectively.

These capacities assume use of 450GB Fibre Channel (FC) drives. 6650H configurations can in principle be scaled to 4,096 nodes using the company's BYNET fabric. Teradata also offers high-end disk systems supplied by EMC.

The latest version of the Teradata Database, 13.10, features enhancements in time series and geospatial processing, system management, availability and other areas. The operating system is SUSE Linux.

Compression has also expanded in 13.10. A combination of three techniques – algorithmic, block-level and multi-value is now employed. Teradata has claimed compression rates of from 25 to 50 percent for combined use of these techniques.

Responding to trends among its customers, Teradata has progressively moved to support complex mixed workload environments, and now offers a range of tools for workload and service-level management. Active EDW 6650 systems are equipped with a number of HA features, including "hot standby" – meaning the ability to fail over automatically to an alternate node.

Teradata also offers a number of smaller and specialized platforms. These include the entry-level 560 Data Mart Appliance; midrange 2650 Data Warehouse Appliance; high-density 1650 Extreme Data Appliance; and 4600 Extreme Data Appliance, which is equipped with flash memory.

IBM Smart Analytics System 7700

Overview

Since its entry into the special-purpose systems market in 2005, IBM has employed high performance computing (HPC) clusters to deliver processing power for high-end data warehouse workloads.

Smart Analytics System 7700 includes IBM's DB2 9.7 database, InfoSphere Warehouse 9.7 data warehouse framework and other software. The overall software stack is shown in figure 10.

Software is deployed on clusters of Power 740 servers equipped with eight-core 3.55 GHz and quad-core 3.72 GHz POWER7 processors and the AIX operating system. IBM DS3500 disk arrays are equipped with 300GB SAS drives.

In part, Smart Analytics System 7700 strengths are derived from its individual components. For example, IBM clustering has been widely employed in HPC applications; POWER7-based systems are recognized industry performance leaders; and DB2 9.7 is a recognized industry leader in a number of areas.

These areas include database clustering – DB2 9.7 multidimensional clustering (MDC) is designed for, and provides high levels of efficiency for table-based clustering in large data warehouse environments – as well as compression, workload management and automation.

SYSTEMS SOFTWARE	ENTERPRISE SOFTWARE
AIX 6.1 operating system	InfoSphere Warehouse
General Parallel File System (GPFS)	DB2 Enterprise Server Edition
Reliable Scalable Cluster Technology (RSCT)	DB2 Performance Expert
Systems Director Platform Control for Power	Tivoli System Automation for Multiplatforms
Systems Director Express Edition	Smart Analytics System Control Console
Java Software Development Kit (SDK)	
STORAGE MANAGEMENT	
Remote Support Manager for Storage	DS Storage Manager

Figure 10: IBM Smart Analytics System 7700 Software Stack

Workload management capabilities are derived from mainframe architecture, and draw upon mainframe strengths in managing diverse concurrent workloads. Automation includes use of what IBM characterizes as autonomic computing – meaning application of advanced artificial intelligence technologies to system administration and optimization tasks.

A broader strength, emphasized by IBM, is that the overall system design as well as all components that form part of it are closely optimized to provide high levels of performance and scalability for representative data warehouse workloads. According to IBM, configurations have been extensively tuned to avoid common bottlenecks. User experiences tend to confirm these claims.

System Structure

Systems are built around a core Power 740-based Data Module, with additional modules handling system builds (Management Module), administration (User Module) and failover, and hosting IBM InfoSphere Warehouse and OLAP solutions. Optional modules may be configured to support Cognos Business Intelligence solutions. Figure 11 illustrates this structure.

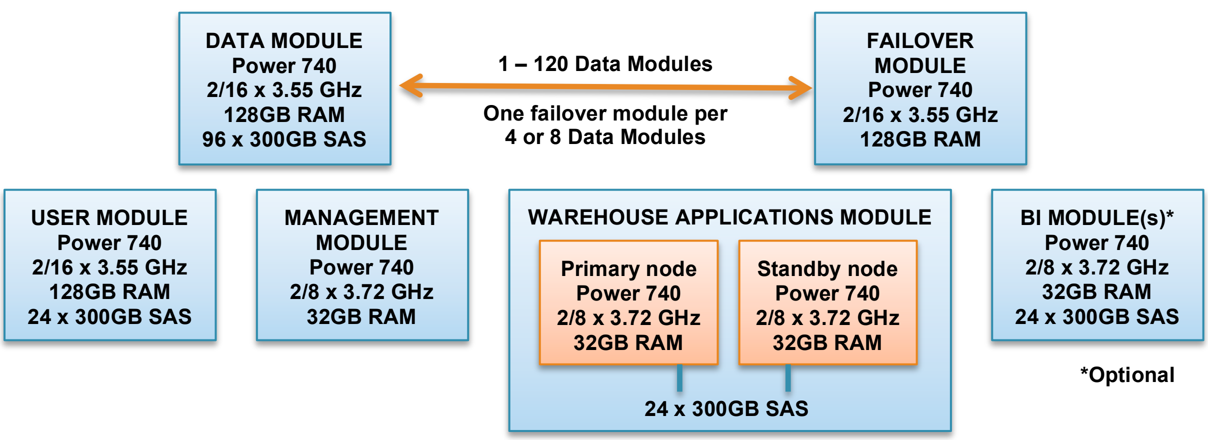


Figure 11: IBM Smart Analytics System 7700 Structure

This presentation shows the total number of processors and cores for each node; e.g., 2/16 x 3.55 GHz refers to two POWER7 3.55 GHz processors with eight cores each. In all configurations except Extra Small, the Warehouse Applications Module is configured with a standby node.

IBM offers six standard configurations, which are summarized in figure 12. Custom configurations with up to 120 Data Modules are also available.

USER DATA	Extra Small	Small	Medium	Large	Extra Large	Extra/Extra Large
Data modules	1	2	3	6	10	20
Racks	2	2	2	3	5	9
SAS disk (TB)	28.8	57.6	86.4	172.8	280	560
User space (TB)	11.5	22.9	34.4	68.8	110.5	220.1
User space – compressed (TB)	28.75	57.5	86.25	172.5	287.5	575
Maximum flash memory (TB)	4.9	9.8	14.7	29.4	49	98

Figure 12: IBM Smart Analytics System 7700 Standard Configurations

Data modules may be equipped with up to seven 700GB flash memory (solid state) drives for a total of 4.9TB. Flash drives are employed to maintain temporary space for processing of intermediate query results. This significantly improves performance, particularly for concurrent query workloads that are characteristic of data warehouses supporting multiple users and applications.

(In comparison, Oracle Exadata Database Machine employs flash memory in a more conventional caching role. The Oracle approach is less optimized for mixed workload environments).

IBM also offers Smart Analytics Systems built around its Intel-based System x servers and System z mainframes (5600 and 9600 respectively).

Oracle Exadata Database Machine

Overview

Oracle Exadata Database Machines have been the subject of a great deal of technical commentary. Advocates tend to focus on the sheer power and bandwidth of Exadata Database Machines; on the advantages of EHCC; and on synergies with broader Oracle enterprise environments. Critics tend to focus on architectural limitations and inefficiencies.

There is little doubt that, in the way in which system resources are used, Exadata Database Machine is a highly inefficient design. The hybridization of a conventional RAC database cluster and the I/O-intensive Exadata Storage Server generates high levels of system overhead.

For example, this is the case for EHCC, which converts Oracle Database 11g rows to columns for compression purposes, then converts these back into rows for processing.

A "greenfield" special-purpose system design would probably have been more efficient. However, management clearly believed that the advantages of incorporating standard Oracle technologies into Exadata Database Machines outweighed the negatives. "Brute force" processing power, Oracle enterprise synergies and off-list discounting would compensate for high system overhead.

Market Response

Market response to Exadata Database Machines, even among the Oracle user base, has clearly been mixed. For example, the company has not fared well in competitive proof of concept (POC) tests against IBM and Teradata systems.

It is also striking that, in many installations for which Oracle has claimed dramatic performance gains, Exadata Database Machines replaced conventional SMP servers and/or RAC clusters, or aging competitive systems. For example, this was the case among 13 of 19 Exadata Database Machine users who contributed to this report; i.e., all organizations that employed systems for data warehousing.

Such comparisons may be misleading. Contrasting the performance of old and new technology systems tends to inflate gains. This is particularly the case when the latter are poorly tuned, inefficiently operated or both.

(The remaining six users had deployed, or were in the process of deploying Exadata Database Machines to support custom Oracle-based transaction processing systems, Oracle FLEXCUBE banking and Oracle Retail, or database consolidation initiatives. Overall demographics are consistent with Oracle's claim that 70 percent of Exadata Database Machine users employ this platform for data warehousing.)

Users were typically major Oracle customers, and in at least some cases, the company appears to have combined Exadata Database Machine purchases with pricing, terms and conditions for broader Enterprise License Agreements (ELAs). Three reported that common Oracle DBA pools were able to support Exadata Database Machine as well as other Oracle-based systems.

Oracle Exadata Database Machine Configurations

Oracle Exadata Database Machines are built around two distinct components:

1. ***Database Machine*** is a cluster of small Sun x86 servers running Oracle Database 11g, Real Application Clusters (RAC), the Enterprise Linux operating system and other Oracle software.

   Current models are built around two, four or eight database servers equipped with Intel Xeon 5670 six-core 2.93 GHz (X2-2) and eight-core Xeon 7560 2.26 GHz (X2-8) processors.

   Oracle Database 11g, RAC cluster architecture and other Oracle software are not unique to Database Machines, and are widely employed by Oracle customers to support other applications on a variety of Intel-based and UNIX platforms.

2. ***Exadata Storage Server*** combines small Sun x86 servers, SAS or SATA disks and a large flash cache to off-load certain SQL tasks and I/O processing from the Database Machine component. Current models employ servers equipped with dual Intel Xeon L5640 six-core processors, 600GB SAS or 2TB SATA drives, and up to 5.2TB of Smart Flash cache.

   Exadata Storage Servers are unique to the Exadata Database Machine, and are not supported for other applications and platforms.

These components are combined in rack configurations. Oracle offers 1/4, 1/2 and full-rack options, which are summarized in figure 13.

Oracle X2-2 1/4 rack	Oracle X2-2 1/2 rack	Oracle X2-2 Full rack	Oracle X2-8 Full rack
• 2 Database Servers	• 4 Database Servers	• 8 Database Servers	• 2 Database Servers
• 4/24 x X5670 2.93 GHz	• 8/48 x X5670 2.93 GHz	• 16/96 x X5670 2.93 GHz	• 16/128 x X7560 2.26 GHz
• 96GB RAM	• 96GB RAM	• 96GB RAM	• 1TB RAM
• 3 Exadata Storage Servers	• 7 Exadata Storage Servers	• 14 Exadata Storage Servers	• 14 Exadata Storage Servers
• 6/36 x L5640 2.26 GHz	• 14/84 x L5640 2.26 GHz	• 28/168 x L5640 2.26 Ghz	• 28/168 x L5640 2.26 GHz
• 36 x 600GB SAS disks	• 84 x 600GB SAS disks	• 168 x 600GB SAS disks	• 168 x 600GB SAS disks
• 1.1TB flash cache	• 2.5TB flash cache	• 5.2TB flash cache	• 5.2TB flash cache

Figure 13: Oracle Exadata Database Machine Options

This presentation shows total number of processors and cores for each configuration; e.g., 4/24 x X5670 four Intel Xeon 5670 2.93 GHz processors with six cores each.

Multiple Exadata Database Machine racks may be interconnected using 40 gigabit per second (Gbps) InfiniBand fabrics.

DETAILED DATA

User Base

User input was provided by 10 organizations that had deployed, or were in the process of deploying Smart Analytics System 7700; 19 organizations that had deployed first- or second-generation Exadata Database Machines; and 17 organizations employing Teradata systems installed since 2007. Input was obtained on applications and workloads, configuration sizes, DBA staffing levels and costs.

Cost Calculations

Costs were calculated as follows:

- ***System costs*** include hardware, systems software licenses and installation (initial costs), along with hardware maintenance, software support and other ongoing services costs for configurations shown in figures 3 through 6.

  Software license and support costs are for database and software stacks typically offered by or required by vendors, but do not include applications software. In the case of Exadata Database Machines, costs include separately licensed Oracle database products normally deployed on this system. These are shown in figure 14.

Database 11g Enterprise Edition	Tuning Pack
Real Application Clusters	Diagnostics Pack
Partitioning	Advanced Compression
	Provisioning & Patch Automation Pack

  Figure 14: Oracle Exadata Database Machine Software

  Smart Analytics System 7700 and Active EDW 6650 costs include Failover and Standby Modules respectively. There is no Oracle equivalent – RAC clusters provide comparable capability for Database Machine configurations. Costs do not include optional modules.

- ***Personnel costs*** are for the numbers of FTEs shown in figure 15.

IBM SMART ANALYTICS SYSTEM 7700		ORACLE EXADATA DATABASE MACHINE		TERADATA ACTIVE ENTERPRISE DATA WAREHOUSE 6650	
Extra Small:	1.05 FTEs	X2-2 1/4 rack:	1.25 FTEs	6650C:	1.1 FTEs
Small:	1.4 FTEs	X2-2 1/2 rack:	1.45 FTEs	6650H–1 node:	1.2 FTEs
Medium:	1.7 FTEs	X2-2 full rack:	2.25 FTEs	6650H–2 node:	1.825 FTEs
Large:	2.45 FTEs	X2-8 full rack:	2.55 FTEs	6650H–3 node:	2.15 FTEs
				6650H–4 node:	2.6 FTEs

  Figure 15: Numbers of Full Time Equivalent DBAs Employed in Personnel Cost Calculations

  Costs were calculated using annual average salaries of $99,864 for DB2 9.7 DBAs; $102,439 for Oracle Database 11g DBAs; and $106,615 for Teradata 13.10 DBAs. Salaries were increased by 51.2 percent to allow for benefits, bonuses and related items, and multiplied for three years.

- ***Facilities costs*** include energy consumption and computer room occupancy by server and storage configurations for each platform. Calculations are based on vendor specifications and, where appropriate, ITG estimates, and include prorated values for infrastructure equipment – including power distribution, cooling and air conditioning systems – supporting systems.

  Energy costs were calculated using a conservative assumption for average price per kilowatt/hour, and assume near-24/7 operations over a three-year period.

  Occupancy costs were calculated using a standard assumption for cost per square foot for existing facilities (i.e., costs of new construction are not included) over a three-year period. Calculations allow for service clearances and aisles as well as space occupied by rack cabinets.

All values are for the United States.

Cost Breakdowns

Cost breakdowns are detailed in figure 16.

IBM SMART ANALYTICS SYSTEM 7700				
	Extra Small	Small	Medium	Large
Initial costs	1,405,244	1,952,969	2,446,582	4,519,738
Maintenance & support	451,081	633,420	807,659	1,508,855
Personnel	475,632	634,176	770,071	1,109,809
Facilities	24,808	43,320	61,988	117,066
TOTAL	**2,356,765**	**3,263,885**	**4,086,300**	**7,255,468**

ORACLE EXADATA DATABASE MACHINE				
	X2-2 1/4 rack	X2-2 1/2 rack	X2-2 full rack	X2-8 full rack
Initial costs	679,461	1,412,544	2,763,644	3,356,044
Maintenance & support	794,664	1,630,728	3,225,456	4,083,408
Personnel	580,829	673,762	1,045,492	1,184,891
Facilities	25,236	44,699	81,736	97,913
TOTAL	**2,080,190**	**3,761,733**	**7,116,328**	**8,722,256**

TERADATA ACTIVE ENTERPRISE DATA WAREHOUSE 6650					
	6650C	6650H – 1 node	6650H – 2 node	6650H – 3 node	6650H – 4 node
Initial costs	794,132	1,463,825	2,927,650	4,391,474	5,855,299
Maintenance & support	803,364	1,482,221	2,964,442	4,446,663	5,928,884
Personnel	531,966	580,326	882,580	1,039,752	1,257,375
Facilities	29,541	45,213	90,426	135,639	180,852
TOTAL	**2,159,003**	**3,571,585**	**6,865,098**	**10,013,528**	**13,222,410**

Figure 16: Three-year Cost Breakdowns

About the International Technology Group

ITG sharpens your awareness of what's happening and your competitive edge
. . . this could affect your future growth and profit prospects

The International Technology Group (ITG), established in 1983, is an independent research and management consulting firm specializing in information technology (IT) investment strategy, cost/benefit metrics, infrastructure studies, deployment tactics, business alignment and financial analysis.

ITG was an early innovator and pioneer in developing total cost of ownership (TCO) and return on investment (ROI) processes and methodologies. In 2004, the firm received a Decade of Education Award from the Information Technology Financial Management Association (ITFMA), the leading professional association dedicated to education and advancement of financial management practices in end-user IT organizations.

The firm has undertaken more than 100 major consulting projects, released approximately 160 management reports and white papers, and delivered nearly 1,800 briefs and presentations to individual clients, user groups, industry conferences and seminars throughout the world.

Client services are designed to provide factual data and reliable documentation to assist in the decision-making process. Information provided establishes the basis for developing tactical and strategic plans. Important developments are analyzed and practical guidance is offered on the most effective ways to respond to changes that may impact or shape complex IT deployment agendas.

A broad range of services is offered, furnishing clients with the information necessary to complement their internal capabilities and resources. Customized client programs involve various combinations of the following deliverables:

Status Reports	In-depth studies of important issues
Management Briefs	Detailed analysis of significant developments
Management Briefings	Periodic interactive meetings with management
Executive Presentations	Scheduled strategic presentations for decision-makers
Email Communications	Timely replies to informational requests
Telephone Consultation	Immediate response to informational needs

Clients include a cross section of IT end users in the private and public sectors representing multinational corporations, industrial companies, financial institutions, service organizations, educational institutions, federal and state government agencies as well as IT system suppliers, software vendors and service firms. Federal government clients have included agencies within the Department of Defense (e.g. DISA), Department of Transportation (e.g. FAA) and Department of Treasury (e.g. US Mint).

International Technology Group
609 Pacific Avenue, Suite 102
Santa Cruz, California 95060-4406
Telephone: + 831-427-9260
Email: Contact@ITGforInfo.com
Website: ITGforInfo.com