April 2013

MANAGEMENT BRIEF

Cost/Benefit Case for IBM PureData System for Analytics

Comparing Costs and Time to Value with Teradata Data Warehouse Appliance

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 Figure

EXECUTIVE SUMMARY

Challenges

In a world of big data and real-time processing, the demands that data warehouses must meet are changing. The general trend is toward use of appliances – i.e., integrated hardware and software packages optimized for high-performance analytical query workloads.

Appliances, however, are no longer a single category of solution. The market has segmented into offerings designed for different types and sizes of workload. This report focuses on the principal competitors in the midrange to high-end appliance bracket, IBM PureData System for Analytics N2001, powered by Netezza technology, and Teradata Data Warehouse Appliance 2700.

Both platforms implement massively parallel processing (MPP). But in other areas of architecture and technology, there are significant differences that affect comparative performance, quality of service and – the focus of this report – economics.

At the most basic level, platform costs may be compared in terms of dollars per terabyte ($/TB) of user data. This metric, however, deals only with acquisition prices. There are typically significant differences in data structures, compression levels and performance between platforms that affect its validity.

A more realistic measurement is offered by cost of ownership. Calculations should allow for acquisition costs as well as maintenance and support, personnel costs for database and system administration, deployment costs – it may take weeks to months of work to bring systems into production – and facilities costs over multi-year periods.

Another type of cost comparison also comes into play. In fast-moving analytics markets, time to value has significant cost implications. Deployment delays may result in lost revenue and/or lost profit opportunities, and other bottom-line effects (e.g., for security applications, higher theft and fraud losses due to deployment delays) may also be experienced. The overall impact may be substantial.

This report applies both sets of metrics to cost comparisons for IBM PureData System for Analytics N2001 and Teradata Data Warehouse Appliance 2700 in four representative installations in digital media, financial services, retail and telecommunications companies. Results are based on input from 17 organizations employing Teradata Data Warehouse Appliances and 21 employing IBM PureData System for Analytics appliances in comparable roles.

Costs of Ownership

Three-year costs of ownership for use of PureData System for Analytics N2001 appliances average 36 percent less than for Teradata equivalents. Comparisons are for comparable applications and workloads. Results may be summarized as follows:

- ***System costs.*** Initial acquisition costs were similar on a $/TB basis although, as figure 1 shows, three-year maintenance costs were higher for Teradata Data Warehouse Appliances. IBM offers one year of support as part of PureData System for Analytics N2001 acquisition prices, and annual costs are also lower on a percentage basis.

  System and facilities costs were calculated for use of latest-generation Teradata 2700 and PureData System for Analytics N2001 models. However, as these were introduced only recently – in October 2012 and January 2013 respectively – personnel and deployment costs were derived from user experiences with earlier models of both platforms.

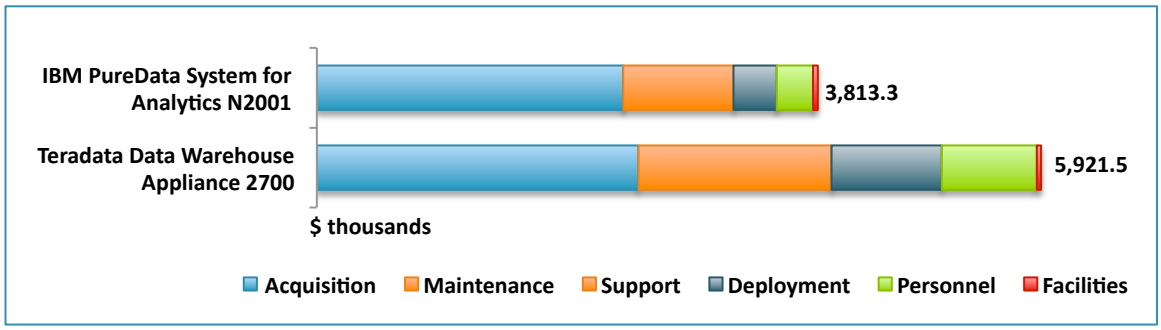

Figure 1: ***Three-year Costs of Ownership for IBM PureData System for Analytics N2001 and Teradata Data Warehouse Appliance 2700 – Averages for All Installations***

- ***Personnel costs***. Full time equivalent (FTE) staffing for database administrators (DBAs) was significantly less for PureData System for Analytics N2001 than for Teradata 2700 appliances. Personnel costs for these averaged 61 percent less.

  Cost disparities reflect use of significantly different system architectures.

  Teradata 2700 systems are positioned as a simplified, lower-cost alternative to the company's flagship Active Enterprise Data Warehouse (EDW). The principal features of EDW architecture are, however, retained. A great deal of DBA time and effort is still required for building and maintaining indexes, performance tuning, management of data models and related tasks.

  The effects are magnified if systems undergo frequent changes in applications, data volumes and/or workloads. Where this is the case, maintaining consistent performance may require sustained tuning over multi-year periods.

  In comparison, PureData System for Analytics data structures are less complex, and require minimal intervention. FTE staffing for DBAs is lower – by wide margins – than for Teradata equivalents, and in all except very large installations the same individuals handle database as well as system administration tasks. Little or no performance tuning is required.

  Among Teradata users surveyed for this report, numbers of FTE DBAs ranged from one to more than five, and averaged 1.7. Among PureData System for Analytics users, administrative overhead ranged from 20 hours per year to – in the case of a company operating more than 30 systems – two FTEs. The average was less than 0.5 FTE.

  Costs were calculated based on average U.S. DBA salary levels, with allowance for benefits, bonuses and related items. Training costs are included in personnel totals.

- ***Deployment costs***. The greater complexity of Teradata environments means that deployment time and effort is – by wide margins – higher than for PureData System for Analytics N2001. Deployment costs, principally for external professional services, averaged 74 percent less for PureData System for Analytics N2001 than for Teradata 2700.

  Deployment costs were calculated based on published vendor rates for appropriate professional services staff, with allowance for normally applicable discounts. Costs include travel and entertainment (T&E) expenses.

Deployment time, in this context, refers to the elapsed time between a decision to deploy a specific platform, and the beginning of production use for a significant part of, or entire user populations (i.e., proof of concept and pilot tests are not regarded as the end point of the deployment process).

In the four installations employed for comparisons, Teradata Data Warehouse Appliance deployment times ranged from four weeks to nine months, compared to four days to three months for PureData Systems for Analytics appliances supporting equivalent applications, user data volumes and workloads.

Among the survey population as a whole, the deployment times shown in figure 2 were reported.

Deployment Time	IBM PureData System for Analytics	Teradata Data Warehouse Appliance
1-2 days	4	–
3-10 days	5	–
10-20 days	7	–
20-50 days	3	2
50-100 days	1	3
100 days to 6 months	1	5
7-12 months	–	4
12 months+	–	4
Total	21	17

Figure 2: ***Reported Deployment Times for IBM PureData Systems for Analytics and Teradata Data Warehouse Appliances***

Among IBM PureData System for Analytics users, more than three quarters – 76 percent – reported deployment times of three weeks or less. In comparison, no Teradata Data Warehouse Appliance users reported deployment times in this range.

Overall, PureData System for Analytics and Teradata Warehouse Appliance deployment times averaged around 23 days and 196 days respectively. (These averages assume 30 days per month – users often reported deployment times of, for example, three months or six months).

Greater Teradata architectural complexity was reflected not only in the amount of time required for construction of data structures, but also in more protracted cycles for testing, tuning, system integration and related tasks.

Numbers of FTE in-house personnel involved in deployments were less easily quantifiable. It was clear, however, that internal staffing and costs were also significantly higher for Teradata Data Warehouse Appliance.

- ***Facilities costs***, principally for energy consumption, were marginally higher for PureData System for Analytics N2001 than for Teradata Data Warehouse Appliance 2700.

Further information on installations, configurations and methodology, along with granular cost breakdowns, may be found in the Basis of Calculations section of this report.

Lost Opportunity Costs

Experience has shown that data warehouse applications may yield significant bottom-line gains, often in a matter of weeks to months. The corollary is that delays in bringing such applications in production may represent significant costs in lost revenue and/or profit opportunities. Competitive position may also be eroded if, in the meantime, others exploit such opportunities.

The impact of such delays is increasing over time. It is a truism that, in recent years, there has been a progressive acceleration of analytical cycle times. Across a wide range of industries, forecasting and planning cycles have declined from months to weeks, to days or even hours. A growing number of organizations are moving to "real time" models.

This shift has been most obvious among digital media companies operating in volatile Internet and social media markets. But few businesses are not impacted by the growth of e-commerce and – increasingly – m-commerce. Immediate response to opportunities is becoming a competitive mandate.

These effects were apparent in the same four installations employed for cost of ownership comparisons. In these cases, lost opportunity costs for deployment of PureData System for Analytics N2001 were significantly lower than for Teradata 2700. Figure 3 illustrates disparities.

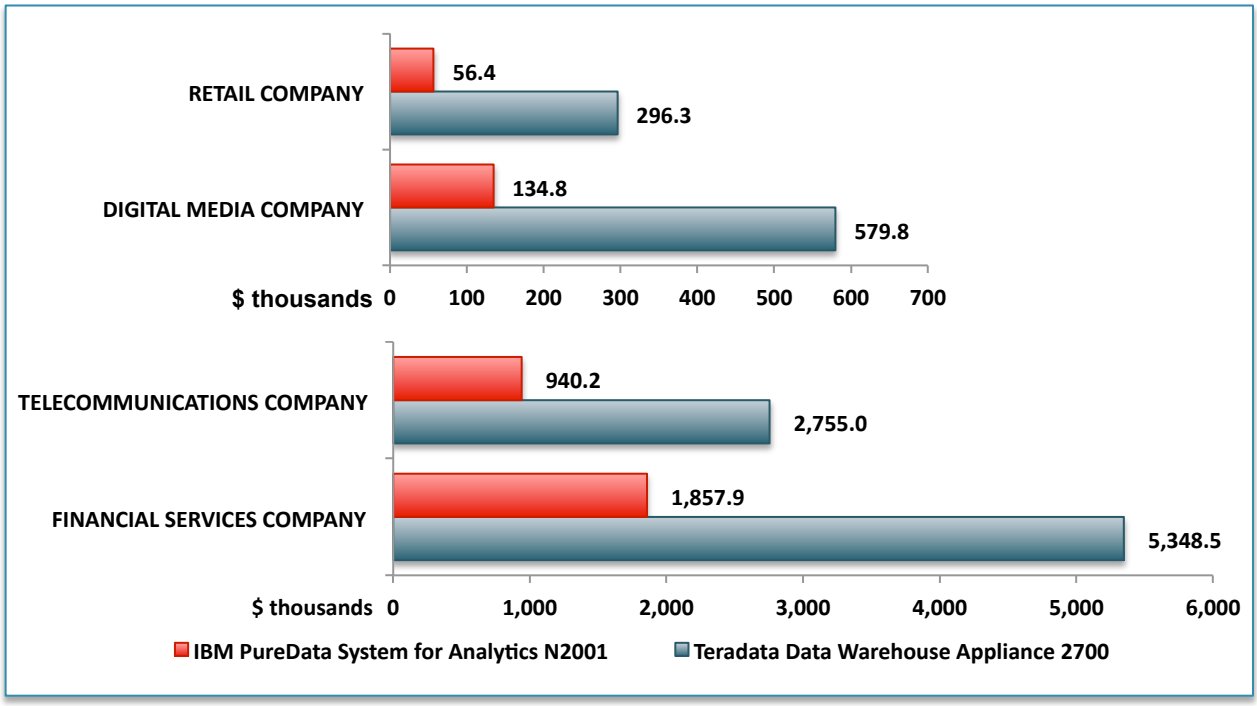

Figure 3: ***Lost Opportunity Costs for Use of IBM PureData System for Analytics N2001 and Teradata Data Warehouse Appliance 2700 – All Installations***

Lost opportunity costs for use of PureData System for Analytics N2001 ranged from 65 to 81 percent less, and averaged 67 percent less than for Teradata 2700 equivalents.

If lost opportunity costs are added to costs of ownership, overall costs for these platforms may be restated as shown in figure 4. Overall costs for use of PureData System for Analytics N2001 averaged 44 percent less than for Teradata 2700.

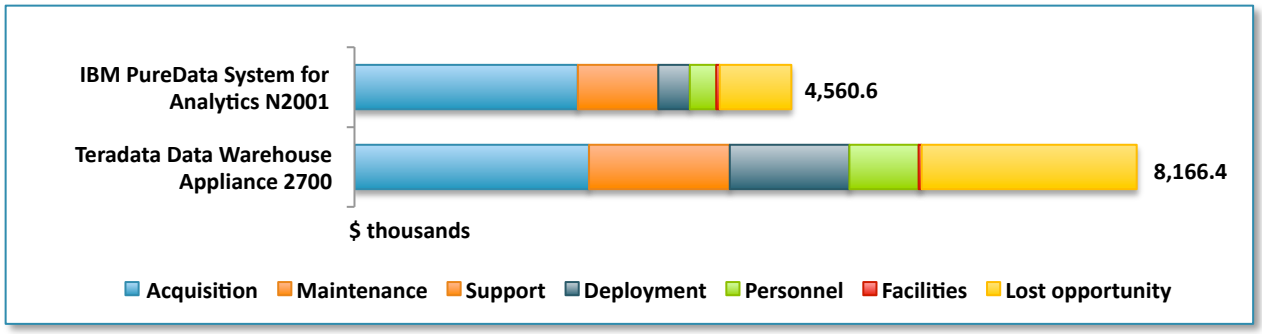

Figure 4: ***Three-year Overall Costs for Use of IBM PureData System for Analytics N2001 and Teradata Data Warehouse Appliance 2700 – Averages for All Installations***

Conclusions

A number of conclusions emerge.

One is that, simply stated, "Time is money." Deployment times affect costs as well as competitive performance over time. From this perspective, the comparative lost opportunity costs cited above understate the case for PureData System for Analytics N2001.

Lost opportunity costs shown in figures 3 and 4 are for initial applications only. In practice, organizations would continue to deploy new applications. The cumulative impact of faster deployment over multi-year periods would be a great deal larger. Disparities in lost opportunity costs would increase, potentially by orders of magnitude.

IBM systems offer a further advantage. The ease with which end users may develop and deploy their own applications not only reduces delays, but also creates the potential for closer business alignment than conventional techniques.

The cost/benefit case for PureData System for Analytics N2001 is thus not simply that it is more cost-effective than, and enables faster deployment than Teradata 2700. It is that the distinctive capabilities of this platform map more closely to the long-term requirements of high-performance data warehousing than any competitor.

USER VIEW

Introduction

This section provides additional detail on user organizations surveyed for this report and on their input on FTE staffing and deployment times. The following section, Technology View, provides further information on technology and vendor positioning for platforms covered in this report.

The final section, Basis of Calculations, outlines installations, and details configurations as well as FTE DBA staffing levels for these. Methodologies and values employed for cost calculations are also presented.

Demographics

PureData System for Analytics as well as Teradata users ranged from recent digital media start-ups with fewer than 200 employees to Fortune 100 and Financial Times Global 100 corporations.

Industry distribution was as shown in figure 5. Comparatively low penetration of Teradata Data Warehouse Appliances among digital media companies appears to reflect overall market demographics.

	Telecom	Digital Media	Financial Services	Retail	Other	Total
IBM PureData System for Analytics	6	5	3	3	4	21
Teradata Data Warehouse Appliances	4	2	4	3	4	17

Figure 5: ***Industry Distribution of Survey Population***

In seven cases (41 percent), Teradata users already employed Active EDW systems. In most cases, Data Warehouse Appliances were used to offload specialized applications from these, or were deployed in complementary roles. Some organizations had migrated data warehouses from Active EDW systems to Data Warehouse Appliances.

These organizations were able to draw upon existing skills and experience with Teradata system architecture. The amount of work required to create and test Teradata data structures was significantly less than would be the case for first-time deployments. Average FTE staffing levels and deployment times for Data Warehouse Appliances thus tend to be understated.

FTE Staffing

Teradata Data Warehouse Appliance

Among Teradata users, three out of 17 reported that less than one FTE DBA was employed to manage systems. The remainder reported between one and five FTE DBAs. The overall average was 1.68 FTEs. Personnel typically had multiple years of experience with Teradata databases.

System administrators/engineers, technical support and other personnel were also typically employed. Figure 6 shows examples.

Because data on non-DBA administrators was often incomplete, these were not included in personnel cost calculations.

Financial Services	Manufacturing	Distribution	Digital Media
2 DBAs 2+ system administration & support	1.5 DBAs 1-2 FTE system administrators	1 DBA 1-1.5 system administrators	1 DBA 1 system administrator

Figure 6: ***Teradata Data Warehouse Appliance FTE Staffing Examples***

The amount of training required for Teradata DBAs varied, depending on whether organizations already employed personnel experienced with Teradata systems. Where this was the case, training was typically required only to update skill bases to Teradata Database 13.10 or 14, and to provide familiarity with changes in systems software and hardware.

Where experienced DBAs were not available, multiple weeks of training were typically required, and even then organizations were reluctant to employ inexperienced personnel to support business-critical systems. If staffing needs could not be met through internal transfers, outside hires or consulting contracts were the norm.

IBM PureData System for Analytics

Among 21 PureData System for Analytics users, 18 reported that they employed less than one FTE administrator. The exceptions were an organization that declined to state the number of systems employed, but described the installation as "over one petabyte" (one FTE was employed); and others reporting more than 20 and more than 30 systems respectively (two FTEs were employed).

Among organizations reporting less than one FTE, 12 (67 percent) estimated than the actual number was less than 0.5. Administration overhead was said to represent "a fraction of one person's time once a week…two hours a week…a couple of hours a week…a few hours a month…less than an hour a day (to administer five systems)…maybe six hours every three months…20 hours a year."

PureData System for Analytics administrators typically handled DBA tasks along with system management, system engineering, development and other functions. Three organizations reported that there was no demand for PureData System for Analytics DBAs in the conventional sense of this term.

In most cases, PureData System for Analytics administrators had previous experience with other business intelligence systems and/or databases, or as system administrators and engineers. Only four organizations reported hiring an external specialist.

A number of reasons were cited for low staffing levels. The most common was that "(end) users interface directly to the system." Two organizations that had migrated from Teradata to PureData System for Analytics, and one that employed both platforms offered more detailed explanations.

One noted that "we don't have to build indexes… users write directly to the system, they don't need to go through a DBA…we work with complete data sets instead of having everything aggregated and summarized first…we don't have to use data models." In comparison with Teradata systems, performance-tuning overhead was said to be "virtually non-existent."

The amount of training required for PureData System for Analytics administrators ranged from none (five cases) to two days. One organization commented that its system had been "up and running for six months before any training was required – and that was for a (system) upgrade." The "learning curve" for PureData System for Analytics administrators was also said to be less steep than for Teradata equivalents.

There were also indications that FTE staffing for PureData System for Analytics developers was typically lower than for Teradata systems. One respondent commented, "(end) users do most of the work. They don't have to go through a developer unless there's something out of the ordinary."

Deployment Times

Start-up time for any appliance depends upon a number of factors. The amount of time and effort required for such tasks as business alignment, identification of data sources, and construction of extract, transformation and load (ETL) mechanisms tends, for example, to be platform-independent. Start-up times are also affected by applications, and by volumes of data that must be loaded and processed.

Survey responses nevertheless indicated that PureData System for Analytics appliances were brought into production more rapidly than Teradata equivalents. With the latter, more time was required for architecture design, construction of data models and indexes, configuration, testing and other tasks. Extensive performance tuning was also the norm even for small deployments.

Deployment time disparities were striking. The fastest reported PureData System for Analytics deployment, for example, involved availability of reporting applications and data to 500+ end users within four days, and full production operations supporting 3,000+ users in less than three weeks.

The fastest comparable Teradata Data Warehouse Appliance deployment involved initial availability in four weeks, and full production operations for 600 reporting users in approximately two months.

At the other end of the spectrum, the longest reported deployment cycles were 3 months and 12 months for PureData System for Analytics and Teradata systems respectively. These cycles were for projects involving near real-time processing of hundreds of terabytes of data.

Deployment times reported by users for smaller Teradata projects are consistent with those cited by the company for its Accelerate program, which offers single-price packages of assistance by the company's Professional Services organization for specific types of deployment.

Figure 7 summarizes principal Teradata offerings. The first two of these, Accelerate for Do It Yourself and Accelerate for Load and Go, do not result in production deployment. The remainder of the offerings have approximate deployment times of 70 to 120 days. Actual deployment times are often longer, and additional professional services fees may be incurred.

Offering	Description	Time	Cost
Teradata Accelerate for Do it Yourself	Initial data load & sample queries. Designed for experienced data warehouse users	10 days	$350,000
Teradata Accelerate for Load & Go	Data loading	30 days	$440,000
Teradata Accelerate for Data Warehousing	Full data warehouse deployment for first-time users	120 days	$830,000
Teradata Accelerate for Finance	Pre-built financial reporting & analysis solution including finance-oriented data models.	70-80 days	$700,000
Teradata Accelerate for Marketing	Pre-built solution for customer segmentation, campaign management, one-to-one targeting & other marketing applications	70 days	$670,000
Teradata Accelerate for Retail	Pre-built solution for sales & inventory analysis, assortment planning, market-basket & store performance analysis & other retail applications.	90 days	$640,000

Source: Teradata

Figure 7: *Principal Teradata Accelerate Offerings*

Additional Accelerate packages are offered by Teradata for platform migration, demand signal processing for consumer packaged goods (CPG) manufacturing and gaming applications.

The company also has a partnership with software company Kalido, which offers accelerated conversion of existing data models to Teradata Data Warehouse Appliance. Customers may, according to Teradata "build or expand a data warehouse in 90 days or less." Typical deployment times again appear, however, to be longer than for comparable PureData Systems for Analytics installations.

TECHNOLOGY VIEW

Teradata

Overview

Teradata is the largest and longest-established player in the data warehouse appliance market. The company, which shipped its first system in 1983, has progressively enhanced its core MPP system and database architecture since that time. The company claims more than 1,500 customers worldwide.

The company's early success in MPP-based data warehousing translated into a large, embedded installed base. Users typically have major commitments to customized systems as well as Teradata-specific applications. In addition, Teradata DBA skills are not easily transferrable to other platforms.

These factors have for a long time allowed Teradata to charge higher prices than smaller and/or newer competitors. Although the company may price systems aggressively in competitive bids, its software subscription and maintenance prices have remained relatively high.

Teradata platforms are built around Teradata Database, of which the most recent versions 13.10 and 14 were introduced in August 2010 and October 2011 respectively. Teradata 13.10 incorporated new spatial and temporal data management features. Teradata 14 added support for columnar data structures and is described by the company as a hybrid columnar and row database.

In hardware terms, Teradata employ OEM Intel processor-based system units, with SUSE Linux Enterprise Server (SLES) and third-party storage, I/O and other hardware components. All systems employ proprietary Teradata BYNET interconnect technology.

While its high-end market position has remained strong, Teradata has faced mounting competition since the mid-2000s from start-ups as well as established vendors. Competitive pressures were largely responsible for Teradata decisions to introduce smaller, less expensive 2000 series Data Warehouse Appliances, and to add other specialized appliance offerings.

The Teradata portfolio currently includes the following platforms.

Active Enterprise Data Warehouse

Teradata 6000 series, a.k.a. Active EDW, is the flagship Teradata platform. The latest generation of 6700 systems, introduced in April 2013, can in principle be scaled to 628 or, in special configurations, up to 2,096 nodes supporting 61 PB of uncompressed user data. Within the IBM appliance portfolio, Active EDW competes primarily with IBM PureData System for Operational Analytics.

The 6700 series is an upgraded version of the 6690 introduced in March 2012. (Teradata's general policy is to update systems annually.) The 6700 series employs six- or eight-core Intel E5, a.k.a. "Sandy Bridge," processors rather than the six-core 5600, a.k.a. "Westmere," processors employed in 6690 models. According to Teradata, performance is "40 percent" higher than for the latter.

The 6700 series is, according to Teradata, designed to exploit new features in Teradata 14. It also incorporates hybrid storage capabilities combining solid-state drives (SSDs) and conventional SAS drives. Hybrid storage implements a variation of automated storage tiering approaches adopted by major disk array vendors.

There are two model groups. Teradata 6700H systems incorporate Intel eight-core 2.6 GHz E5 processors and support hybrid storage. The 6700C models, which are designed for pre-6690 legacy compatibility, incorporate six-core 2.0 GHz processors and may be configured with SAS drives only. Disk drives supported are 300 GB, 450 GB and 600 GB 10K SAS and (for 6700H systems) 400 GB SSDs.

Data compression in Teradata 14 employs multiple software-based compression techniques – algorithmic, block-level and multi-value – introduced in earlier Teradata versions, and for which the company has claimed combined compression rates of 25 to 50 percent. A new capability for columnar compression was introduced in Teradata 14.

Data Warehouse Appliance

Teradata Data Warehouse Appliance systems implement the Active EDW architecture, but are less scalable (up to 162 nodes) and do not support hybrid storage. They are positioned by the company as a lower-cost alternative to Active EDW for organizations with less complex computational requirements.

The current Teradata 2700 system was announced in October 2012, and became generally available in March 2013. It is an upgraded version of the earlier 2690 incorporating Intel eight-core 2.6 GHz E5 rather than the six-core Westmere processors. Systems may be configured with 300 GB, 600 GB or 900 GB 10K SAS drives

According to Teradata, 2700 systems offer "up to two times" higher performance than 2690 equivalents. Data loading speeds are said to be "two to four times" faster.

Systems feature compression technology implemented through application specific integrated circuit- (ASIC-) based engines and exploiting new columnar data structures in Database 14. According to Teradata, columnar structures allow for higher compression levels than row-based techniques. Values published by Teradata indicate 65 to 70 percent compression levels.

Systems support user data capacities of between 6.8 TB and 493.8 TB uncompressed or 22.4 TB to more than 1.6 PB with 70 percent compression. Figure 8 summarizes these capacities.

DRIVE TYPE	300 GB		600 GB		900 GB	
USER DATA	Uncompressed	70% compression	Uncompressed	70% compression	Uncompressed	70% compression
Quarter cabinet	6.8 TB	22.4 TB	13.65 TB	45.5 TB	20.5 TB	67.7 TB
Half cabinet	13.65 TB	45.045 TB	27.3 TB	90.99 TB	41.0 TB	135.3 TB
Three-quarter cabinet	20.6 TB	68 TB	41.1 TB	137 TB	61.7 TB	203.6 TB
Full cabinet	27.2 TB	89.8 TB	54.4 TB	181.3 TB	82.3 TB	271.6 TB
2 cabinets	54.4 TB	179.5 TB	108.8 TB	362.6 TB	164.6 TB	543.2 TB
3 cabinets	81.6 TB	269.3 TB	163.2 TB	544 TB	264.9 TB	874.2 TB
4 cabinets	108.8 TB	359 TB	217.6 TB	725.3 TB	329.2 TB	1,294.3 TB
5 cabinets	136 TB	452.9 TB	272 TB	906.6 TB	411.5 TB	1,358 TB
6 cabinets	163.2 TB	538.6 TB	326.4 TB	1,077.1 TB	493.8 TB	1,629.5 TB

Figure 8: ***Teradata Data Warehouse Appliance 2700 User Data Capacities***

This presentation includes published Teradata user data capacities for certain models, and prorated estimates for the remainder.

According to Teradata, 2700 systems may be extended to 21 racks, although the company has to date not published details of larger configurations.

Other Platforms

Other Teradata platforms include the following:

- ***Extreme Data Appliance 1650*** is designed for analytical applications involving extremely large volumes (hundreds of terabytes to petabytes) of data. This platform was introduced in May 2011.

  Representative applications cited by Teradata include analysis of clickstream data in e-commerce; call detail record (CDR) data in telecommunications; and customer transaction data in retail and financial services.

  The 1650 supports Teradata MPP architecture and Database 13.10 and 14, and can in principle be scaled to 4,096 nodes. Its principal distinguishing characteristic is that it employs slower, higher-capacity 1 TB and 2 TB 7.2K NL-SAS drives.

- ***Data Mart Appliance 670***, introduced in April 2013**,** is a single-node, entry-level platform designed for small-scale production, as well as test and development applications. It supports most capabilities in Teradata Database 13.10 and 14, but employs symmetric multiprocessing (SMP) rather than the MPP architecture of larger Teradata systems.

  Systems may be configured with a single processing unit – the same options are available as for the 6700 series – and NetApp E2600 disk arrays supporting up to 8 TB of uncompressed user data. According to Teradata, data compression levels of 50 to 70 percent may be realized.

The overall Teradata platform portfolio is summarized in figure 9.

	Active Enterprise Data Warehouse 6700	Data Warehouse Appliance 2700	Extreme Data Appliance 1650	Data Mart Appliance 670
Introduced	04/13	10/12	10/10	04/13
Max. nodes	2,048	162	4,096	1
Number of racks	2 to 32	¼ to 21	¼ to 64	½ to 1
Processors/node	2/16 x E5-2600 2.6 GHz (6700H) 2/12 x E5-2600 2.0 GHz (6700C)	2/16 x E5-2600 2.6 GHz	2/12 x x5670 2.93 GHz	2/16 x E5-2600 2.6 GHz (670H) 2/12 x E5-2600 2.0 GHz (670C)
Memory	256GB per node	128GB per node	48GB per node	Up to 256GB
Disk types	300, 450 & 600 GB 10K SAS, 400GB SSD	300, 600 & 900 GB 10K SAS	2 TB 7.2K NL-SAS	300 & 600 GB 10K SAS, 400 GB SSD
Max. user data (uncompressed)	61 PB	1.6 PB (6 racks)	186 PB	8 TB

Figure 9: ***Teradata Platform Portfolio Summary***

In addition, Teradata markets the Aster Big Analytics Appliance, which employs technology from Aster Data Systems, acquired by Teradata in 2011. This is an MPP system supporting petabyte-scale unstructured data applications developed using SQL standard interfaces. Systems can be equipped with additional nodes supporting the Hortonworks distribution of Apache Hadoop.

IBM

Platform Portfolio

IBM has been a player in the data warehouse appliance market since 2005, when it introduced its Balanced Warehouse system. The company currently offers three main platforms:

1. ***IBM PureData System for Operational Analytics*** is the flagship IBM data warehouse appliance. Formerly known as IBM Smart Analytics System 7700, it is designed for large-scale complex data warehouse environments.

   PureData System for Operational Analytics is an MPP platform built around the latest-generation IBM DB2 10 database and the DB2 10-based InfoSphere Information Warehouse framework. This framework incorporates a broad range of discovery and predictive analytics functions including OLAP, data mining, text analytics and Cognos Business Intelligence (BI) tools.

   Hardware infrastructure is provided by IBM Power Systems 16-core servers equipped with 8-core 3.55 GHz POWER7 processors and the AIX operating system. Disk arrays support tiering using mixes of 300 GB or 600 GB 10K SAS drives, and 700 GB SSDs.

   PureData System for Operational Analytics is offered in 1/3 rack to 10-rack configurations, with user data capacities of up to 1.2 PB uncompressed or around 4 PB with 70 percent compression. It is expected that the system, introduced in October 2012, will be updated in the near future.

2. ***IBM PureData System for Analytics N2001*** employs more powerful Intel and field-programmable gate array (FPGA) processors, and higher-capacity, faster disks than previous-generation PureData System for Analytics N1001 (formerly TwinFin) models. Performance is, according to IBM, approximately three times higher than for the latter.

   Currently, half-rack to four-rack models are offered, and it is expected that quarter-rack and larger multi-rack configurations will be added.

   A modified version of the PureData System for Analytics N2001 appliance, ***DB2 Analytics Accelerator for z/OS***, is designed to offload high-performance analytical processing from IBM System z mainframes using the z/OS version of DB2.

3. ***IBM Smart Analytics System 5710*** is the small-footprint (2U to 4U) data warehouse appliance. Designed for small and medium business use, and for departmental applications in larger organizations, it competes primarily with the Teradata Data Mart 560 and comparable offerings from HP, Oracle and others.

   Also built around DB2 10 with InfoSphere Information Warehouse, the system employs IBM System x dual-socket servers equipped with six-core Intel X5675 processors and SUSE Linux Enterprise Server (SLES). Disk arrays support 300 GB or 600 GB SAS 10K drives.

IBM does not currently offer a dedicated “big data” appliance comparable to the Aster Big Data Analytics Appliance. The company’s appliances, however, support the full range of emerging big data standards, and may be employed with a variety of IBM and third-party big data solutions.

Architecture and Technology

Key components of NPS system architecture may be summarized as follows:

- ***MPP*** employs a proprietary compiler that divides query workloads into segments, named "Snippets," that are then executed in parallel by blade-based "Snippet Processors" (S-blades).

- ***Streaming*** allows data to be transferred to and from disks more rapidly – by orders of magnitude – than conventional MPP architectures. Data is moved between disks and S-blades in asynchronous mode (i.e., transmission delays and protocol overheads are minimal).

  In PureData System for Analytics N2001 systems, each S-blade may handle up to 40 simultaneous data streams from as many disks at rates of up to 130 MB/second each. In a full-rack system with 240 active disk drives, aggregate bandwidth is 240 x 130 MB/second = 31.2 GB/second or, with data compression, close to 128 GB/second. IBM cites an overall throughput level of 450 TB per hour per rack.

- ***Filtering*** ensures that data not required for a specific query is screened out before being passed to processor memory. According to IBM, typically 95 to 98 percent of user data is excluded in this manner. Filtering is accomplished using multiple software-based engines, principally those shown in figure 10.

Project Engine	Filters out unnecessary column data based on parameters specified in the SELECT clause of the SQL statement being processed.
Visibility Engine	Filters out rows of data that should not be visible to query being executed – either because the records had been marked deleted by an earlier query, or because they had been added to the database after the start of the current query. Maintains ACID (Atomicity, Consistency, Isolation & Durability) compliance at streaming speeds.
Restrict Engine	Filters out unnecessary row-level data based on WHERE predicate clauses of the SQL statement being processed.

Figure 10: ***Principal PureData System for Analytics N2001 Filter Engines***

  This approach massively reduces internal latency. In a conventional MPP architecture, such as that employed by Teradata Active EDW and 2700 systems, multiple interactions between disks and processor memory occur in a manner that slows the entire process cycle.

- ***Compression*** employs a set of algorithms that compress numeric, integer and temporal (date and time) data written to disk during load, insert and update operations. The system automatically chooses the best compression algorithm to use depending upon data characteristics.

  Earlier PureData System for Analytics models typically realized two to four times compression rates. In the PureData System for Analytics N2001, algorithm enhancements as well as more powerful FPGAs have, according to IBM, increased typical rates by two to three times.

In PureData System for Analytics N2001 as in earlier systems, ***S-Blades*** combine Intel processors executing core NPS logic with FPGA-based filtering and control engines. A separate Compression Engine decompresses data before transferring it to Intel processors. This structure is illustrated in figure 11.

Current S-Blades combine IBM HX5 blade servers with dual Intel E7-2830 eight-core 2.13 GHz Linux processors and 128 GB RAM, and dual eight-core Xilinx FPGAs. Up to seven active S-Blades are supported in a full rack.

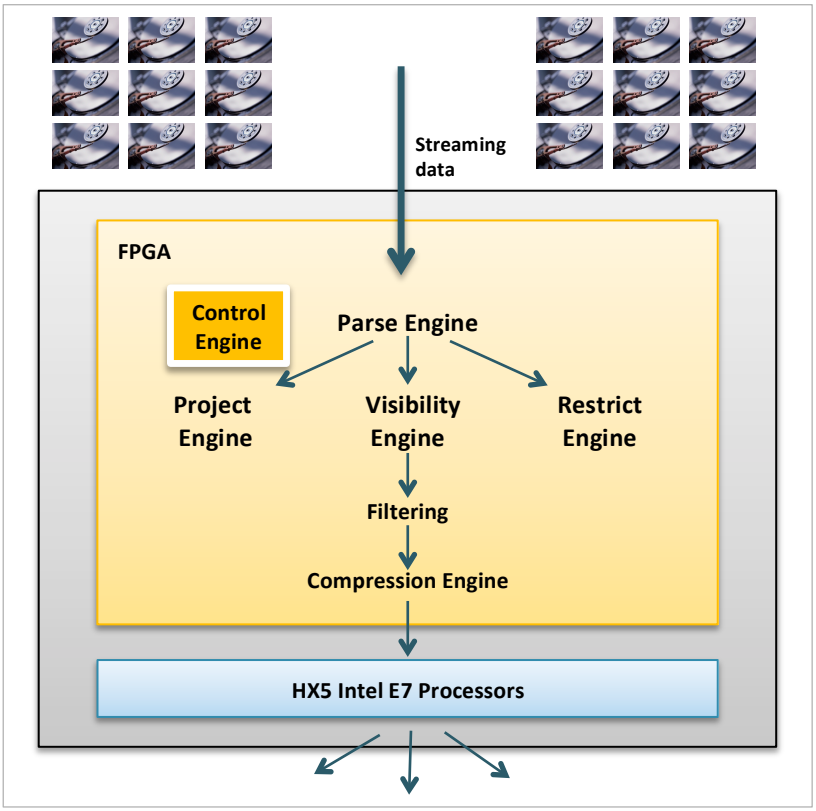

Figure 11: ***PureData System for Analytics S-Blade Structure***

Disk storage is provided by standard 600 GB 10K SAS drives in 12 enclosures per rack. A full single rack system contains 288 drives, of which 240 are active, 14 provide swap/log space and 34 act as spares. Data is striped across primary disks, and duplicated on secondary disks in a RAID 1 configuration.

The current PureData System for Analytics N2001 product line is summarized in figure 12.

Model	005	010	020	040
User data	96 TB	192 TB	384 TB	768 TB
Racks	½	1	2	4
Active disk drives	120	240	480	960
Active S-Blades	4	7	14	28
Intel processor cores	64	112	224	448
FPGA cores	64	112	224	448

Figure 12: ***Current PureData System for Analytics N2001 Product Line***

In addition, dual redundant symmetric multiprocessing (SMP) hosts implement system-wide SQL compiler, query plan, management and optimization, and other functions. In PureData System for Analytics N2001 appliances, IBM HX5 blade servers and Red Hat Linux 6 are employed in this role.

BASIS OF CALCULATIONS

Composite Profiles

The calculations presented in this report are based upon the four composite profiles shown in figure 14.

RETAIL COMPANY	DIGITAL MEDIA COMPANY	TELECOMMUNICATIONS COMPANY	FINANCIAL SERVICES COMPANY
BUSINESS PROFILE			
Multi-channel retailer 70 TB+ user data	Online content provider 200 TB+ user data	Landline, mobile & Internet services provider 300 TB+ user data	Diversified multinational trading services 500 TB+ user data
APPLICATIONS			
Sales & inventory analysis, customer behavior modeling, planning, merchandising, forecasting, campaign management, one-on-one marketing, various	Customer targeting; online advertising & content delivery & other applications for financial services, online services, telecommunications, travel & other businesses	Analysis of call detail record (CDR), social media & billing data for range of applications; e.g., cost & profitability analysis, usage forecasting, churn reduction	Real-time trading analysis for compliance & regulatory applications; identification of abnormal patterns for fraud detection, anti money laundering etc.
IBM PUREDATA SYSTEM FOR ANALYTICS N2001			
½ rack 96 TB user data compressed ***0.5 FTE DBA*** **Deployment time: 4 days**	1 rack 192 TB user data compressed ***0.4 FTE DBA*** **Deployment time: 2 weeks**	2 racks 384 TB user data compressed ***0.5 FTE DBA*** **Deployment time: 2 months**	4 racks 768 TB user data compressed ***1.0 FTE DBA*** **Deployment time: 3 months**
TERADATA DATA WAREHOUSE APPLIANCE 2700			
½ rack 90 TB user data compressed ***1.2 FTE DBAs*** **Deployment time: 4 weeks**	1 rack 180 TB user data compressed ***0.75 FTE DBA*** **Deployment time: 2 months**	2 racks 362 TB user data compressed ***1.25 FTE DBA*** **Deployment time: 6 months**	4 racks 725 TB user data compressed ***2 FTE DBAs*** **Deployment time: 9 months**

Figure 14: *Composite Profiles*

Teradata 2700 systems were configured with Teradata Database 14, and system sizing and FTE staffing calculations reflect use of this version. Teradata 2700 and PureData System for Analytics N2001 appliances were both configured with 600 GB 10K SAS drives.

Cost Calculations

Costs were calculated as follows:

- ***System costs*** are based on discounted acquisition and maintenance (Teradata) or support (IBM) fees for bundled configurations offered by vendors.

- ***Personnel costs*** are for the numbers of FTEs shown above. Costs were calculated using annual average salaries of $112,230 and $95,532 for Teradata 2700 and PureData System for Analytics N2001 DBAs respectively. Salaries were increased by 55.48 percent to allow for benefits, bonuses and related items, and multiplied for three years.

- ***Deployment costs*** for Teradata 2700 installations in the retail, digital media and telecommunications companies were calculated based on pricing for the company's Accelerate for Retail, Accelerate for Marketing and Accelerate for Data Warehousing offerings respectively. Costs for the financial services company were calculated based on applicable Teradata Professional Services skill levels and rates.

  Deployment costs for PureData System for Analytics N2001 in the retail and digital media installations were based on IBM business partner offerings. There is no direct IBM equivalent to the Teradata Accelerate program. Costs for the telecommunications and financial services companies were calculated based on applicable IBM Global Services skill levels and rates.

  For both platforms, costs include travel and education (T&E) for onsite visits by outside professional services personnel.

- ***Training costs*** were calculated for 17 days of DBA classes for Teradata 2700 and 5 for PureData System for Analytics N2001, plus additional online education courses. It was assumed that, in each case, two individuals attended onsite; i.e., no allowance was made for T&E expenses.

- ***Facilities costs*** are for energy consumption by appliances. Calculations are based on vendor specifications and, where appropriate, ITG estimates, and assume near-24/365 operations over a three-year period.

All values are for the United States.

Cost Breakdowns

Costs of ownership breakdowns are presented in figure 15.

	RETAIL COMPANY	DIGITAL MEDIA COMPANY	TELECOMMUNICATIONS COMPANY	FINANCIAL SERVICES COMPANY
IBM PUREDATA SYSTEM FOR ANALYTICS N2001				
Acquisition	701,250	1,229,500	2,475,000	4,922,500
Support	252,450	442,530	891,000	1,772,100
Deployment	98,530	165,985	388,900	663,458
Personnel	222,800	178,240	222,800	445,599
Training	6,000	6,000	11,000	11,000
Facilities	10,839	19,355	38,710	77,421
TOTAL ($)	**1,291,869**	**2,041,610**	**4,027,410**	**7,892,078**
TERADATA DATA WAREHOUSE APPLIANCE 2700				
Acquisition	655,200	1,305,600	2,611,200	5,222,400
Maintenance	393,120	783,360	1,566,720	3,133,440
Deployment	279,502	753,888	1,362,550	2,598,694
Personnel	628,183	392,614	654,357	1,046,971
Training	32,200	32,200	52,800	52,800
Facilities	8,279	17,110	34,220	68,440
TOTAL ($)	**1,996,484**	**3,284,772**	**6,281,847**	**12,122,745**

Figure 15: *Costs of Ownership 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

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 120 major consulting projects, released more than 250 management reports and white papers and more than 1,800 briefings 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 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

WAL12377-USEN-00