ITG | International Technology Group

Management Report

May 2014

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 Figures

Executive Summary

Challenges

In a world of big data and real-time processing, the demands that data warehouses must meet are increasing. 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 field 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, the IBM PureData System for Analytics N200X, powered by Netezza technology, and Teradata Data Warehouse Appliance 2750.

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, along with 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.

Both sets of metrics are applied to cost comparisons for IBM PureData System for Analytics N200X and Teradata Data Warehouse Appliance 2750 in four representative installations in digital media, financial services, retail and telecommunications companies. Results are based on input from 21 organizations employing IBM PureData System for Analytics appliances and 17 employing Teradata Data Warehouse Appliances in comparable roles.

Costs of Ownership

Three-year costs of ownership for Teradata 2750 systems averaged 1.5 times higher than for PureData System for Analytics 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 N200X acquisition prices, and annual costs are also lower on a percentage basis.

  System and facilities costs were calculated for use of latest-generation Teradata 2750 and PureData System for Analytics N200X models. However, as these were introduced relatively recently – in October 2013 and January 2013 respectively – personnel and deployment costs were derived from user experiences with earlier models of both platforms.

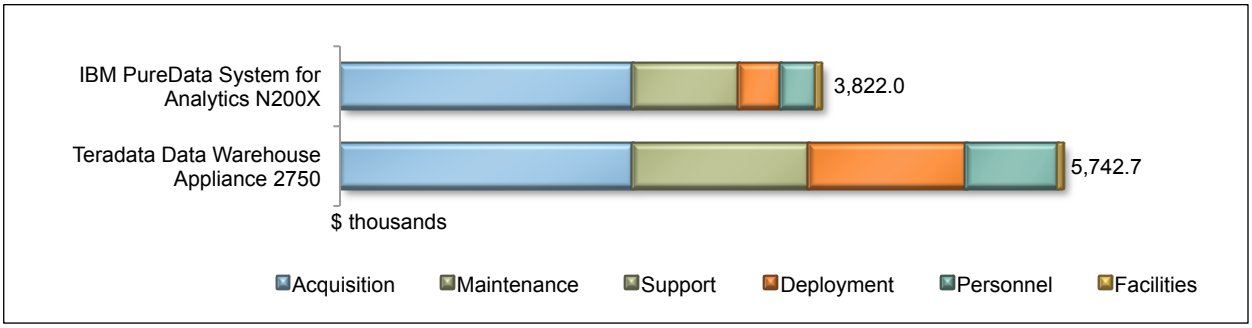

Figure 1: *Three-year Costs of Ownership for IBM PureData System for Analytics N200X and Teradata Data Warehouse Appliance 2750 – Averages for All Installations*

- *Personnel costs.* Full time equivalent (FTE) staffing for database administrators (DBAs) was significantly higher for Teradata 2750 systems. Personnel costs for use of the Teradata 2750 averaged 2.6 times higher than for PureData System for Analytics.

  The Teradata 2750 platform is positioned as a simplified, lower-cost alternative to the company's flagship Active Enterprise Data Warehouse (Active 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 than for Teradata equivalents and, in all but 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; i.e., average FTE staffing for Teradata systems was 3.4 times higher than for use of PureData System for Analytics.

  Personnel 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 N200X. Deployment costs, principally for external professional services, averaged 3.8 times higher for Teradata 2750 systems than for use of PureData System for Analytics N200X.

  Deployment costs were calculated based on published vendor rates for appropriate professional services staff, with allowance for 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 from 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+	–	3
Total	**21**	**17**

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

It is striking that more than three quarters of PureData for Analytics customers completed deployment before any Teradata customer had done so.

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 higher for Teradata Data Warehouse Appliance.

- *Facilities costs.* These were principally for energy consumption, and were marginally higher for PureData System for Analytics N200X than for Teradata 2750.

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 increases 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 has become a competitive imperative.

These effects were apparent in the same four installations employed for cost of ownership comparisons. In these cases, lost opportunity costs for deployment of Teradata 2750 systems were significantly higher than for use of PureData System for Analytics N200X. Figure 3 illustrates disparities.

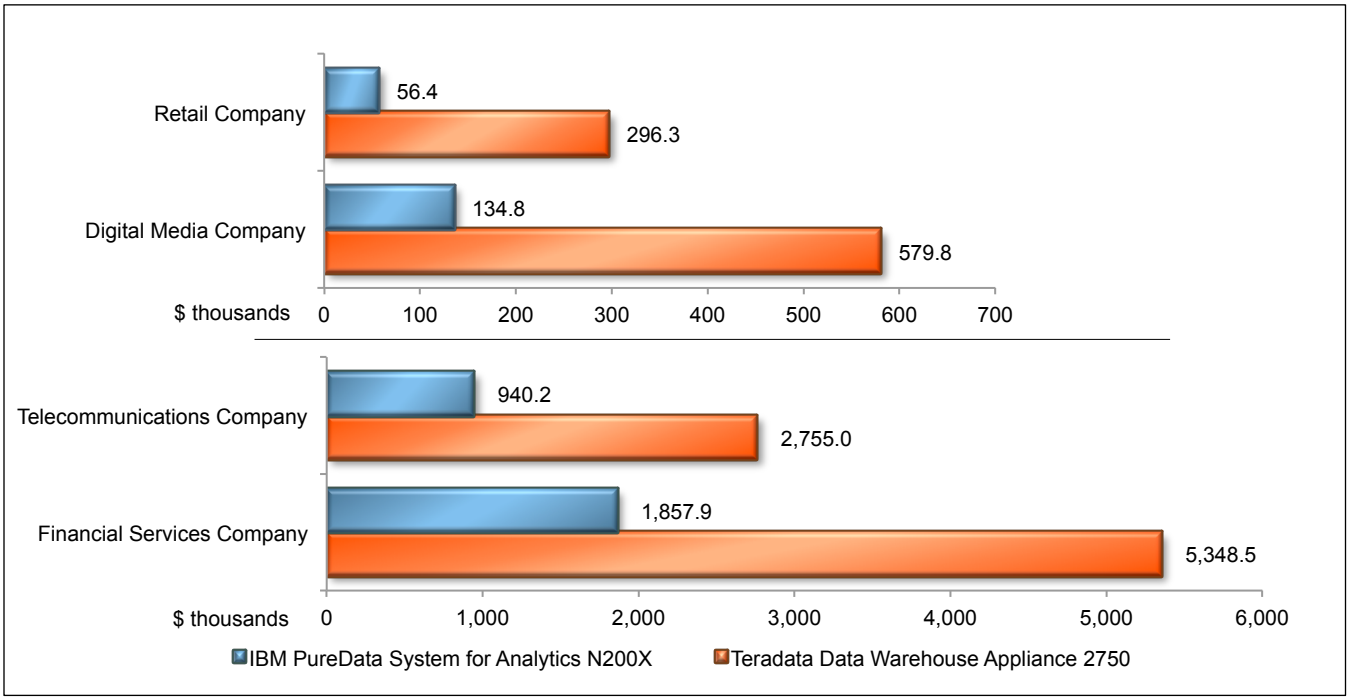

Figure 3: *Lost Opportunity Costs for Use of IBM PureData System for Analytics N200X and Teradata Data Warehouse Appliance 2750 – All Installations*

Lost opportunity costs for use of Teradata 2750 systems ranged from 2.9 to 5.3 times higher, and averaged three times higher than for use PureData System for Analytics N200X.

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 Teradata 2750 systems averaged 1.7 times higher than for use of PureData System for Analytics N200X.

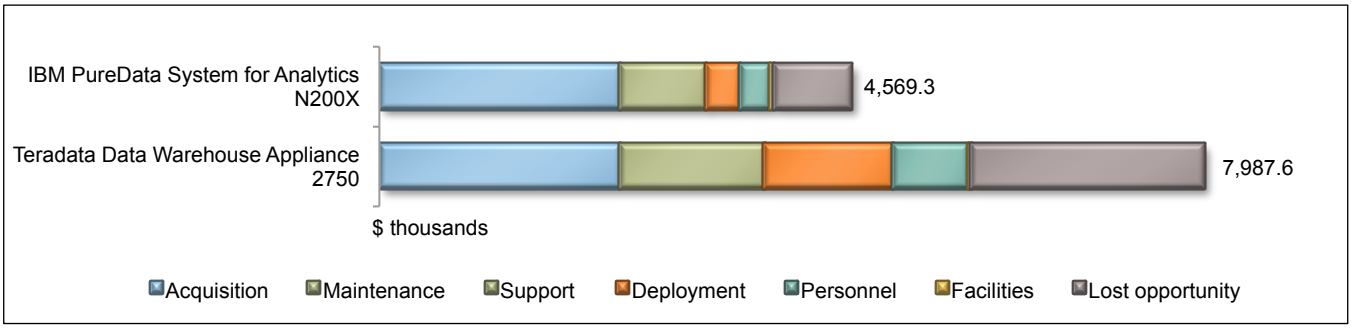

Figure 4: *Three-year Overall Costs for Use of IBM PureData System for Analytics N200X and Teradata Data Warehouse Appliance 2750 – Averages for All Installations*

Conclusions

Conventional SQL-based data warehouses have served users well for more than 20 years. They have, however, grown progressively more complex, and this trend will continue as vendors try to come to terms with new hardware and software technologies. At the same time, pressures for faster application delivery have increasingly frustrated business users.

Simply stated, *time is money*. Deployment times affect costs as well as business performance over time. From this perspective, the comparative lost opportunity costs cited above understate the case for PureData System for Analytics N200X.

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 dramatically.

PureData for Analytics 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 data warehouse techniques.

The cost/benefit case for PureData System for Analytics N200X is thus not simply that it is more cost-effective than, and enables faster deployment than Teradata 2750. 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. 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. 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 1.7 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 were the norm.

It is unclear whether DBA staffing and training requirements will be different for Teradata Database 15. This became available only in April 2014, and there was little user experience with it when this report was prepared.

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 tasks such as business alignment, identification of data sources and construction of extract, transformation and load (ETL) mechanisms tends 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 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 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, Accelerate for Do It Yourself and Accelerate for Load and Go, do not result in production deployment. The remainder offer approximate deployment times of 60 to 120 days. Actual deployment times are often longer, and additional professional services fees may be incurred.

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.

Offering	Description	Time (Days)	Starting Price
Teradata Accelerate for Do it Yourself	Initial data load & sample queries. Designed for experienced data warehouse users	10	$350,000
Teradata Accelerate for Load & Go	Data loading	30	$440,000
Teradata Accelerate for Application Re-Host	Consolidation of data marts & operational reporting systems to data warehouse	120	$550,000
Teradata Accelerate for Data Warehousing	Full data warehouse deployment for first-time users	120	$830,000
Teradata Accelerate for Finance	Pre-built financial reporting & analysis solution including finance-oriented data models.	70-80	$700,000
Teradata Accelerate for Marketing	Pre-built solution for customer segmentation, campaign management, one-to-one targeting & other marketing applications	70	$670,000
Teradata Accelerate for Demand Signal Repository	Pre-built solution implementing demand signal repository for consumer goods manufacturers	60	$400,000
Teradata Accelerate for Gaming	Pre-built solution implementing analytics & reporting applications for gaming companies	90	$600,000
Teradata Accelerate for Retail	Pre-built solution for sales & inventory analysis, assortment planning, market-basket & store performance analysis & related applications	90	$640,000

Source: Teradata

Figure 7: *Teradata Accelerate Offerings*

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'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 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, 14 and 15 were introduced in August 2010, October 2011 and April 2014 respectively.

Teradata 13.10 incorporated new spatial and temporal data management features. Teradata 14 implemented columnar data structures and is described by the company as a hybrid columnar and row database. The new Teradata Database 15 added support for Java Script Object Notation (JSON) and standards commonly employed in NoSQL database and other open source environments.

In hardware terms, Teradata platforms 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.

Recent Developments

While its high-end market position has remained strong, since the mid-2000s Teradata has faced mounting competition 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 (the first of which was introduced in 2008), and to add other specialized appliance offerings.

The company has also sought to integrate new technologies such as columnar data structures, Hadoop and NoSQL, solid-state drives (SSDs) and recently, in-memory computing. Successive versions of Teradata Database added support for these as described above.

In addition, in 2011 Teradata acquired Aster Data Systems, a specialist developer of Hadoop-based MPP analytical software. A line of analytical appliances employing Aster technology was introduced the same year.

Teradata has, in recent generations of systems, also placed a strong emphasis on what the company terms *system-wide temperature management* meaning the use of different types of storage device for *hot* (most frequently accessed), *cold* (infrequently accessed) and intermediate types of data. The approach is comparable to tiering mechanisms employed by storage array vendors.

Key capabilities include Virtual Storage, enabling use of SSDs as well as conventional disk; and Intelligent Memory, allowing exceptionally hot data to be processed in RAM. These were introduced in Database 13 and 14 respectively.

Data compression employs multiple software-based compression techniques – algorithmic, block-level and multi-value – introduced in earlier Teradata versions. High-performance columnar compression is also supported in Database 14 and 15.

Platform Portfolio

Teradata offers a broad portfolio of what the company terms *workload-specific platforms*. These currently include the following:

- *Active Enterprise Data Warehouse* 6000 series is the flagship high-end Teradata platform. The latest-generation 6750, introduced in April 2014, employs Intel Ivy Bridge processors. Standard configurations may be scaled to 2,048 nodes, and larger systems may be implemented on a custom basis.

  The 6750 is, according to Teradata, designed to exploit new features in Teradata 14 and 15, and employs SAS HDDs and SSDs.

- *Data Warehouse Appliance* implements the same core architecture as Active EDW, but does not support use of SSDs. It is positioned by Teradata as a lower-cost alternative to Active EDW systems for organizations with less complex computational requirements; as a development platform; and to support subject-specific data marts.

  Data Warehouse Appliances are also commonly employed as failover systems supporting EDWs. The latest-generation 2750, introduced in October 2013, is discussed in more detail below.

  Data Warehouse Appliance sales, according to Teradata, accounted for 14 percent of the company's product revenues, or more than $170 million during 2013. This does not include consulting and maintenance services revenues.

- *Integrated Big Data Platform 1700* is designed for analytical applications involving extremely large volumes (hundreds of terabytes to petabytes) of data. This platform was introduced in October 2013. It is an upgraded version of the earlier Teradata Extreme Data Appliance.

  The 1700 supports Teradata MPP architecture and Database 14.10 and 15, and can in principle be scaled to 2,048 nodes. It employs slower, higher-capacity SAS 2 TB and 3 TB 7.2K drives.

- *Data Mart Appliance* offers 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 higher, including support for hybrid HDD and SSD configurations, but employs symmetric multiprocessing (SMP) rather than MPP architecture.

  The latest-generation 670, introduced in April 2013, employs Intel Xeon processors 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.

These platforms are summarized in figure 8.

	Active Enterprise Data Warehouse 6750	Data Warehouse Appliance 2750	Integrated Big Data System 1700	Data Mart Appliance 670
Introduced	04/14	10/13	10/13	04/13
Number of nodes	up to 2,048	up to 2,048	up to 2,048	Single
Number of cabinets	2 to 32	Quarter to 32	Quarter to 32	Half to One
Processor nodes	2/24 x Ivy Bridge 2.7 GHz	2/24 x Ivy Bridge 2.7 GHz	2/16 x Sandy Bridge 2.6 GHz	2/16 x 2.6 GHz or 2/12 x 2.0 GHz Xeon
Disk types	300, 450 or 600 GB 10K SAS; 400 GB SSD	300, 600 or 900 GB 10K SAS	2 TB or 3.0 TB 7.2K SAS	300, 450 or 600 GB 10K SAS; 400 GB SSD
Max. user data (uncompressed)	61 PB	21 PB	234 PB	8 TB

Figure 8: *Teradata Workload-Specific Platforms*

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

Teradata also offers the *Teradata Aster Big Data Analytics Appliance*. This combines Aster Hadoop and MapReduce technology with tools enabling users to access Hadoop data using SQL techniques, and to employ SQL-based analytics and extract, transform and load (ETL) tools to exploit Hadoop databases. The appliance is deployed on Intel Sandy Bridge processors with SLES 11.

Teradata has defined a Unified Data Architecture designed to allow integration of Teradata Databases and the Aster environment. According to Teradata, an extension of this architecture, QueryGrid, will allow users to initiate multi-phase analyses of data contained in Teradata and Hadoop databases. QueryGrid will be available in third quarter 2014.

Two other appliances are currently less emphasized: the Teradata Appliance for SAS High Performance Analytics (although many Teradata customers employ SAS on EDWs, the appeal of a dedicated appliance proved to be limited) and the Teradata Extreme Performance Appliance 4600, an all-SSD design aimed at real-time analytics applications.

Teradata takes the position that high levels of performance for data warehouse applications can be delivered more effectively using hybrid HDD and SSD configurations, and in-memory technology.

Data Warehouse Appliance

Data Warehouse Appliance 2750 is an upgraded version of the earlier 2700. Teradata 2750 systems are equipped with 12-core Ivy Bridge 2.7 GHz processors and up to 512 GB of RAM supporting Intelligent Memory, rather than 8-core Sandy Bridge 2.66 GHz processors with 128 GB RAM employed in the 2700.

In other respects, the Teradata 2750 is similar to the 2700. Like the latter, it may be configured with 300 GB, 600 GB or 900 GB 10K SAS drives, and employs application specific integrated circuit- (ASIC-) based compression engines exploiting columnar data structures in Database 14. Values published by Teradata indicate 65 to 70 percent compression levels for typical data warehouse workloads.

The 2750 platform may be configured in increments between one-quarter and 32 cabinets supporting user data capacities between 6.8 TB and more than 21 petabytes (PB) uncompressed or 22.4 TB to more than 70 PB with 70 percent compression. Figure 9 summarizes capacities for one- to six-cabinet configurations.

DRIVE TYPE	300 GB		600 GB		900 GB	
USER DATA	Uncompressed (TBs)	70% compression (TBs)	Uncompressed (TBs)	70% compression (TBs)	Uncompressed (TBs)	70% compression (TBs)
Quarter cabinet	6.8	22.4	13.7	45.7	20.5	67.7
Half cabinet	13.7	45.7	27.4	91.3	41.4	135.3
Three-quarter cabinet	20.5	68.3	41.4	138	61.7	203.6
Full cabinet	27.4	91.3	54.8	182.7	82.3	271.6
2 cabinets	54.8	182.7	109.6	365.3	164.6	543.2
3 cabinets	82.2	274	164.4	548	264.9	874.2
4 cabinets	109.6	365.3	219.2	730.7	329.2	1,294.3
5 cabinets	137	456.7	274	913.3	411.5	1,358
6 cabinets	164.4	548	328.7	1,095.7	493.8	1,629.5

Figure 9: *Teradata Data Warehouse Appliance 2750 User Data Capacities*

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

IBM PureData System for Analytics

Overview

IBM PureData System for Analytics is based on the Netezza Performance Server (NPS) architecture. Netezza, which introduced its first NPS product in 2002, pioneered the data warehouse appliance market during the 2000s. IBM acquired the company in 2010.

The core NPS design employed a unique combination of MPP, filtering, streaming and compression technologies to deliver industry-leading query performance. Use of comparatively low-cost field-programmable gate array (FPGA) processors and commodity disk drives offered highly competitive price/performance levels.

The simplicity of NPS architecture also contributed to its popularity. NPS systems have a longstanding reputation for exceptionally fast deployment and low administration overhead. Over time, this has become an increasingly significant component of their appeal.

The fourth generation of NPS systems, the TwinFin family, was introduced in 2009 and later rebranded as IBM PureData System for Analytics N1001. The fifth generation, N200X, was introduced in January 2013.

N200X systems employ more powerful Intel and FPGA processors, and higher-capacity SAS disks. Performance is, according to IBM, approximately three times higher than for N1001 equivalents. Currently, half-rack to four-rack models are offered.

A modified version, DB2 Analytics Accelerator for z/OS, offloads analytical processing from IBM System z mainframes running the z/OS version of IBM's DB2 database.

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*, which 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 N200X 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/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 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 N200X Filter Engines*

  This approach massively reduces internal latency. In a conventional MPP architecture, such as that employed by Teradata Active EDW and 2750 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 N200X, algorithm enhancements as well as more powerful FPGAs have, according to IBM, increased typical rates by two to three times compared to N1001 systems.

In PureData System for Analytics N200X 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.

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.

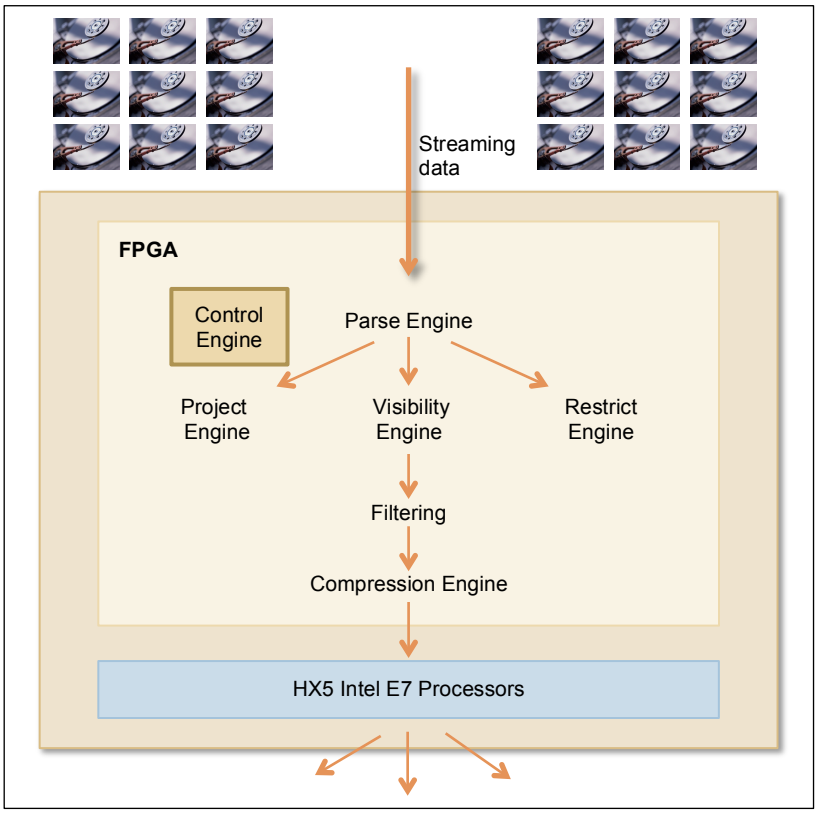

Figure 11: *PureData System for Analytics S-Blade Structure*

The current PureData System for Analytics N200X product line is summarized in figure 12.

Model	002	005	010	020	040
User data	32 TB	96 TB	192 TB	384 TB	768 TB
Racks	Quarter	Half	1	2	4
Active S-Blades	2	4	7	14	28
Intel processor cores	32	64	112	224	448
FPGA cores	32	64	112	224	448

Figure 12: *Current PureData System for Analytics N200X 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 N200X 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 80 TB+ user data	Online content provider 150 TB+ user data	Landline, mobile & Internet services provider 350 TB+ user data	Diversified multinational trading services 600 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 N200X			
Half rack 96 TB user data* 0.5 FTE DBA *Deployment time: 4 days*	Full rack 192 TB user data* 0.4 FTE DBA *Deployment time: 2 weeks*	2 racks 384 TB user data* 0.5 FTE DBA *Deployment time: 2 months*	4 racks 768 TB user data* 1.0 FTE DBA *Deployment time: 3 months*
TERADATA DATA WAREHOUSE APPLIANCE 2750			
Half rack 91.3 TB user data* 1.2 FTE DBAs *Deployment time: 4 weeks*	Full rack 182.7 TB user data* 0.75 FTE DBA *Deployment time: 2 months*	2 racks 365.3 TB user data* 1.25 FTE DBA *Deployment time: 6 months*	4 racks 730.7 TB user data* 2 FTE DBAs *Deployment time: 9 months*

* Compressed

Figure 14: *Composite Profiles*

Teradata 2750 systems were configured with Teradata Database 14, and system sizing and FTE staffing calculations reflect use of this version. Teradata 2750 and PureData System for Analytics N200X systems 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 $114,138 and $97,156 for Teradata 2750 and PureData System for Analytics N200X DBAs respectively. Salaries were increased by 56.7 percent to allow for benefits, bonuses and related items, and multiplied for three years.
- *Deployment costs* for Teradata 2750 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 N200X 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 2750 and 5 for PureData System for Analytics N200X, 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 N200X				
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	228,365	182,692	228,365	456,730
Training	6,000	6,000	11,000	11,000
Facilities	11,446	20,439	40,878	81,757
TOTAL ($)	**1,298,041**	**2,047,146**	**4,035,143**	**7,907,545**
TERADATA DATA WAREHOUSE APPLIANCE 2750				
Acquisition	625,600	1,244,400	2,488,800	4,970,800
Maintenance	375,360	746,640	1,493,280	2,982,480
Deployment	279,502	753,888	1,362,550	2,598,694
Personnel	633,112	395,695	659,492	1,055,186
Training	32,200	32,200	52,800	52,800
Facilities	8,743	18,068	36,136	72,273
TOTAL ($)	**1,954,517**	**3,190,891**	**6,093,058**	**11,732,233**

Figure 15: *Costs of Ownership Breakdowns*

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.

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.

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).

WAL12377-USEN-01