Microsoft's SQL Server Parallel Data Warehouse Provides High Performance and Great Value

Published by: Value Prism Consulting

Sponsored by: Microsoft Corporation

Publish date: March 2013

Abstract: Data Warehouse appliances may be difficult to compare and contrast, given that the different preconfigured solutions come with a variety of available storage and other specifications. Value Prism Consulting, a management consulting firm, was engaged by Microsoft® Corporation to review several data warehouse solutions, to compare and contrast several offerings from a variety of vendors. The firm compared each appliance based on publicly-available price and specification data, and on a price-to-performance scale, Microsoft's SQL Server 2012 Parallel Data Warehouse is the most cost-effective appliance.

Disclaimer

Every organization has unique considerations for economic analysis, and significant business investments should undergo a rigorous economic justification to comprehensively identify the full business impact of those investments. This analysis report is for informational purposes only. VALUE PRISM CONSULTING MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

CONTENTS

EXECUTIVE SUMMARY

Enterprise Data Warehouse (EDW) solutions have become less expensive and easier to install and manage, especially with vendors providing preconfigured hardware plus software "appliance" solutions.

This whitepaper is meant to be an aid to organizations looking to compare and contrast similar appliances from these vendors, before investing in a data warehouse appliance. Value Prism Consulting, a management consulting firm, was engaged by Microsoft® Corporation to review several data warehouse solutions. The firm conducted a survey of publicly-available price and specification data for each appliance in this study.

Enterprise data warehouse appliances from EMC, IBM, Microsoft, Teradata, and Oracle were reviewed and compared. Price-to-performance comparisons have been collected and summarized across each vendor – two based on storage (compressed and uncompressed user-available storage, as a factor of price) and two based on performance (number of cores and GB standard memory, again as a factor of price). In the Figure 1 chart, results closer to the center show lower cost-options. In all four cases Microsoft has the lowest ratio, showing they are a high-performing and economic data warehousing appliances. Teradata was a close second in many categories; IBM and EMC were next; and Oracle was, in most cases, not just last, but much more expensive based on both total cost and cost ratios.

Care should always be taken in assessing the best solution for your situation. This comparison is based on the public retail price and publicly available specification metrics. Individual vendors offer different discounts and volume price breaks, so results may be different than the ones listed here.

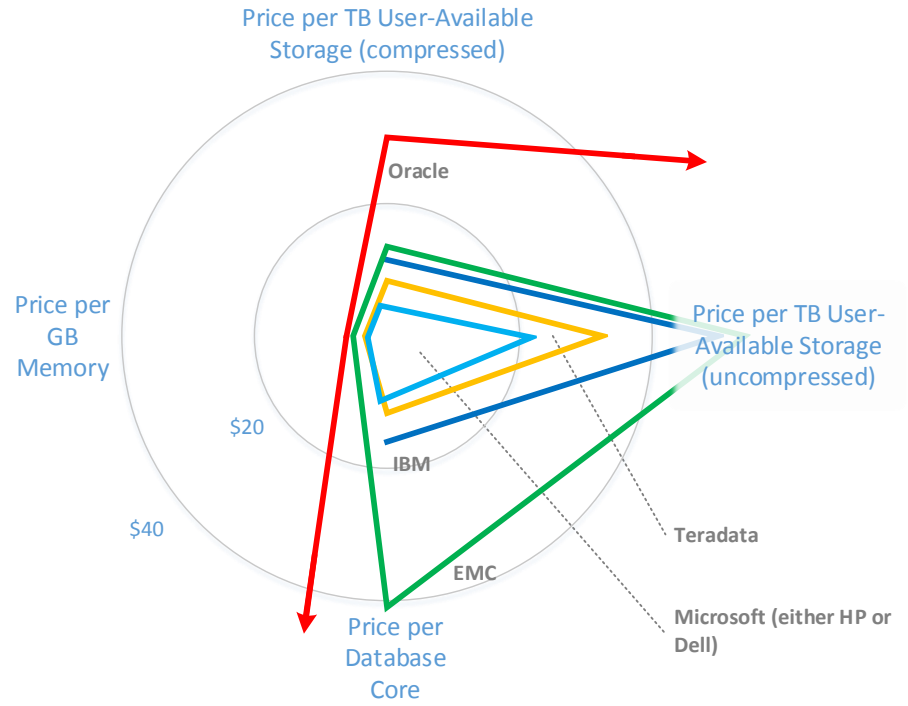


Figure 1: Price-to-Performance Ratios Across Multiple Storage and Performance Metrics (prices in U.S. dollars, in thousands)

VALUE PRISM CONSULTING
VALUATION & DECISION SUPPORT ANALYSIS

INTRODUCTION

Data warehouse appliances provide customers with a simpler way to buy, as a pre-configured, optimized, and often single-priced package of:

- Software required to run the appliance, including server operating system, database software, and data management tools, and
- Hardware components required to run the appliance, including the box, disk drives, memory, network connectivity, and processors.

Most everything is packaged and preconfigured, ready for the customer to plug in and start using immediately (or at least the with an expectation of a minimum of setup).

By-and-large, purchasing is also a simpler process as a fixed set of software and hardware – if not a single SKU, at least a short list of hardware and software products with a minimum of options for easy ordering. Customers can often pick an appliance and expect it will be nearly ready to "plug and play" with much less setup and configuration than a custom solution that could take many months.

In a study commissioned by Microsoft®, several primary Enterprise Data Warehouse (EDW) appliances, each from one of five leading vendors, have been reviewed, summarized, and compared. Each vendor provides via its Website an appliance datasheet that has been used as the primary source for specification data (such as storage, cores, etc.). List pricing and other details are cited specifically, and are also taken from public sources. For each vendor, one leading EDW appliance (if they offer more than one) was selected for comparison – the one exception is Microsoft, with an option to include appliance hardware from either Dell or HP; both appliance specifications are included and discussed. High Performance appliances were selected for consistent comparison. Full-rack pricing and specifications were used to ensure standard comparison. The appliances detailed in this whitepaper are listed in the sidebar.

The full rack appliances from these five vendors provide a range of available storage, cores, compression, and other specifications, and are priced at a wide variety of retail list price:

Appliances included:

EMC Greenplum Data Computing Appliance (Standard Module)
http://www.greenplum.com/sites/default/files/2012_0419_EMC_GP_DCA_Computing_Appliance.pdf

IBM PureData System for Analytics N1001-010
http://public.dhe.ibm.com/common/ssi/ecm/en/imd14400usen/IMD14400USEN.PDF

Microsoft SQL Server® 2012 Parallel Data Warehouse (PDW), with hardware from HP or Dell
http://www.microsoft.com/en-us/sqlserver/solutions-technologies/data-warehousing/pdw.aspx

Oracle Exadata Database Machine X3-2 (SC)
http://www.oracle.com/us/products/database/exadata-db-machine-x3-2-1851253.pdf

Teradata Data Warehouse Appliance 2690 (600 GB disks)
http://www.teradata.com/brochures/Teradata-Data-Warehouse-Appliance-2690/

Vendor and Appliance	Memory (GB)	Total Cores	Compression	User Storage (TB, Compressed)	List Price
EMC Greenplum Data Computing Appliance	768	48	4 to 1	144	$2,000,000
IBM PureData System for Analytics N1001-010	n/a	112	4 to 1	128	$1,599,000
Microsoft SQL Server 2012 Parallel Data Warehouse[1]	2,304	144	5 to 1	340	$1,569,970
Oracle Exadata Database Machine X3-2	2,048	128	10 to 1	450	$13,580,000
Teradata Data Warehouse Appliance 2690	768	96	4 to 1	146	$1,168,000

With the appliance model, it is harder to compare and contrast similar solutions such as these, especially when vendors make various claims in each of their public datasheets (as listed above, with URLs). IBM says it is "a low cost option" with "low total cost of ownership." Greenplum claims the "Best price/performance ratio in the industry." Teradata says the 2690 is "The best price for performance appliance in the marketplace." And Oracle, on its Exadata Web Page, says "Extreme Performance, Lowest Cost."[i] EDW appliances highlight key aggregate metrics, but the underlying hardware and software features can impact purchase decisions.

Summary metrics, list price, total user storage space (compressed and uncompressed), and performance were compared. To make a more accurate comparison, these solutions were also measured against several price-to-performance ratios:

- Price per terabyte of compressed and uncompressed user space is used as a price/value approximation that can help provide more comparison details when the amount of user space and compression ratios are not the same across all appliances.
- Price per database core and gigabyte of memory were also used, as more performance-related metrics.

Based on this comparison, Microsoft's SQL Server 2012 Parallel Data Warehouse appliances provide the most cost-effective solution in terms of total cost, as well as being the top performers in every price-to-performance metric comparison. In addition to being cost-effective, Microsoft's PDW solutions include:

- A proven, scalable, and pre-configured hardware and software solution.
- High performance networking.
- A complete data warehouse solution, with Microsoft BI tools such as HDInsight and Analysis Services, and integration support for many third-party solutions.
- Hardware choice options between HP and Dell.

[1] This is a representative list price that combines Microsoft software list price and hardware list price from Dell or HP. As hardware prices will vary dependent on which hardware vendor chosen, take this list price as guidance and not law.

NOTE: This comparison is based on the list price and publicly available metrics. Each unique customer discount situation will be different, and more information (and customization) is likely available from each vendor so the comparison may be different than the one using only list prices and specifications.

Appliance price is only part of the cost consideration. Customers should carefully evaluate licensing offers as well as installation, migration, and on-going management cost before making a decision.

PRICE COMPARISONS

The appliances have varying numbers of drives, cores, and user storage, so simply looking at total price is not a reasonable comparison. To provide additional comparison perspectives, the following scaled metrics have been calculated and presented. The first two are related to total appliance size, the other two are related to appliance power and performance. While the first (price per terabyte of compressed, user-available storage) is the standard Data Warehouse comparison metric, reviewing and considering all four should provide a more complete analysis. The four metrics are:

- **Price per terabyte of compressed user-available storage**, as a measure of the value of useable storage space. This includes any compression factors that each vendor provides – compression shrinks the amount of space data takes, allowing organizations to make better use of existing hardware so that more data can be stored on each drive and I/O transactions are smaller.
- **Price per terabyte of uncompressed user-available storage**, as an additional measure of the value of useable storage. Since each organization provides different compression technologies and recommendations (and each organization decides on their own compression factor), this allows a comparison regardless of compression.
- **Price per database core**, as one measure of performance. As more cores provide capacity for more concurrent database and server tasks.
- **Price per gigabyte of memory,** as a second measure of performance. Since memory is able to store information in a quickly-accessible place, more memory means more information can be stored without having to write to a disk, which means transactions can be processed more quickly.

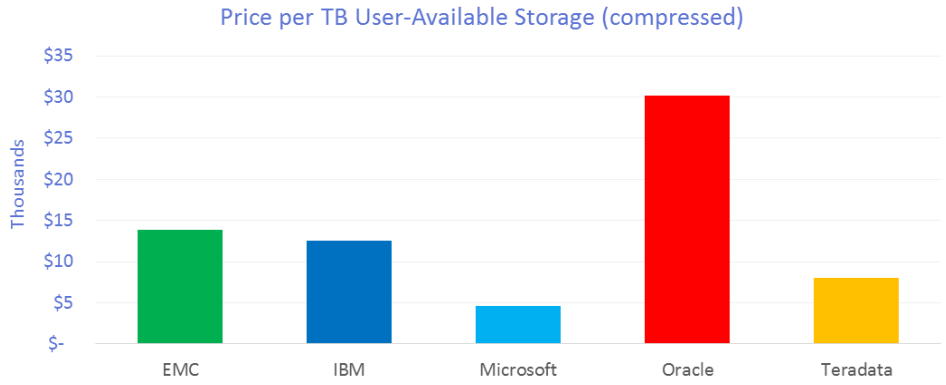
Across all four metrics, Microsoft SQL Server 2012 Parallel Data Warehouse appliances provide the best price-to-performance ratio. The storage, performance, and price metrics are then given individual detail in the following sections.

The conclusions discussed in this section are based on pricing and specification data covered and referenced in later sections. Refer to those for background information and links to source materials.

Price per Terabyte (TB) of Compressed User-Available Storage

For price-to-compressed-storage for full-rack appliances, both Microsoft options are the lowest at a little more than U.S. $4,000. That is nearly two times better than the next option, four times better than average, and seven times better than Oracle, based on retail prices. Oracle's high software price (as detailed in the Total Price section) is greatly offset by its 10-to-1 compression ratio (at least twice as high as any other appliances), but is still the most expensive option on a price/user-available-storage (compressed) comparison (as detailed in the Compression metrics section).

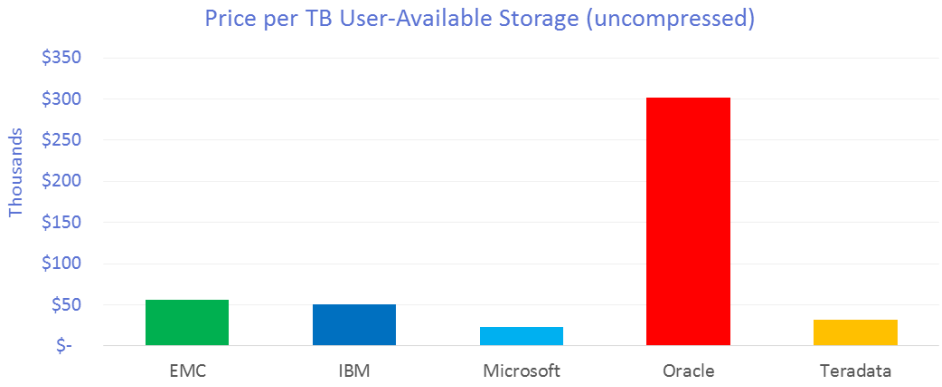
Vendor	User Available Storage (Compressed, in TB)	Price per TB of User-Available Storage (Compressed)
EMC	144	$13,889
IBM	128	$12,492
Microsoft	340	$4,618
Oracle	450	$30,178
Teradata	146	$8,000



Price per Terabyte of Uncompressed User-Available Storage

Uncompressed user-available storage for a full-rack appliance is included to provide perspective of total useful storage provided outside of compression. Microsoft appliances again outdistance the competition with a much better price-per-terabyte ratio, about three to four times better than the median and average, and about seven times better than Oracle, based on retail pricing.

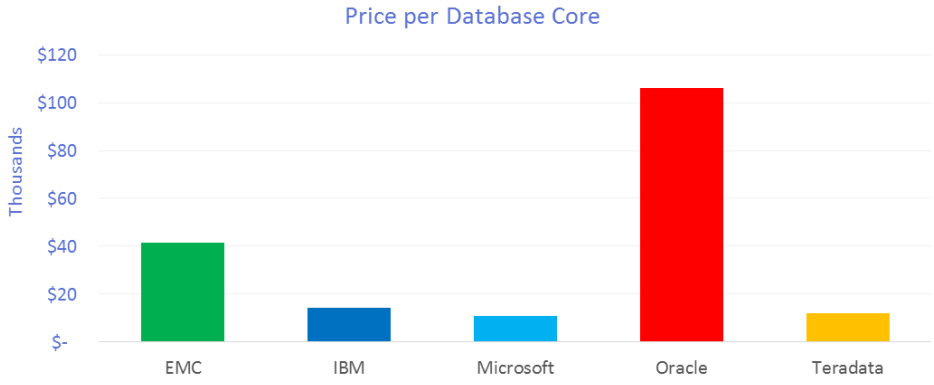
Vendor	User Available Storage (Uncompressed, in TB)	Price per TB of User-Available Storage (Uncompressed)
EMC	36	$55,556
IBM	32	$49,969
Microsoft	68	$23,088
Oracle	45	$301,778
Teradata	37	$32,000



Price per Database Core

With more cores, database applications and virtual machines run more efficiently, meaning that databases can handle a large user load and manage new data collection, while still providing high performance results. So more is better, but with data warehouse appliances that can scale by adding more modules or racks, even more important is price per core. Note that the number of cores is not exactly correlated with higher price – some licenses are not required for every core, but price-per-database-core provides a view into the performance of each appliance as a factor of total price. When viewed in relation to price for full-rack appliances, Microsoft appliances again lead the pack, followed closely by Teradata, and IBM not far behind. Microsoft is nearly three times better than average, and almost ten times better than Oracle, based on retail pricing.

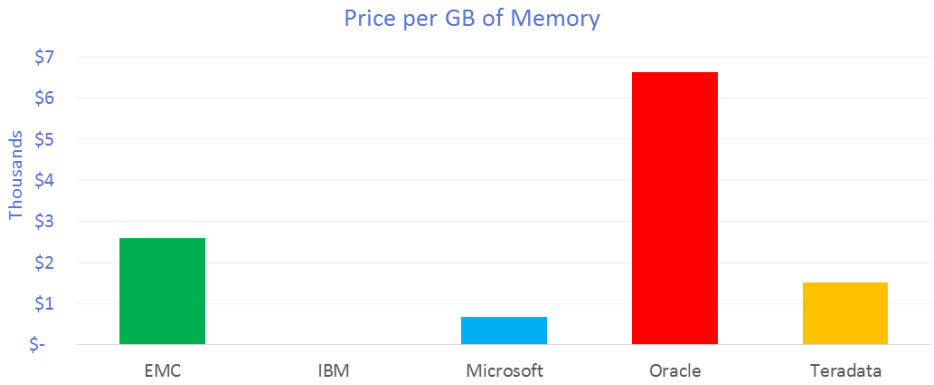
Vendor	Database Cores	Price per Database Core
EMC	48	$41,667
IBM	112	$14,277
Microsoft	144	$10,903
Oracle	128	$106,094
Teradata	96	$12,167



Price per Gigabyte (GB) of Memory

Like cores, the amount of memory included in a full-rack appliance is a significant indicator of potential performance. Large amounts of memory allow for faster processing of data by keeping more information in memory to be processed, instead of having to make read/write calls to the hard drive. Memory is much more expensive than disk drives (for equal units of storage), so in addition to the storage ratios above, it is important to include price/memory ratio comparison as well. Again, Microsoft appliances provide better value over others, at more than two times better than the next best option, four times better than average, and ten times better than Oracle, based on retail pricing. Note: IBM does not provide this metric in its appliance datasheet.

Vendor	GB of Memory	Price per GB of Memory
EMC	768	$2,604
IBM	n/a	n/a
Microsoft	2,304	$681
Oracle	2,048	$6,631
Teradata	768	$1,521



TOTAL PRICE

The total price of each full-rack appliance with high performance disks is based on publicly-available information directly from the vendor, from a reseller that has listed appliance pricing, or if necessary from news or blog organizations that have published price estimates. Total retail price for each appliance, along with the pricing source, is listed here. URLs for each price source are listed in the Appendix. All prices are listed in U.S. dollars.

Oracle Software packages and per-core pricing for each

Vendor	Appliance Retail Price	Appliance Price Source
EMC	$2,000,000	InformationWeek news article from October 2010[ii]
IBM	$1,599,000	Gemini licensing website, an IBM partner (includes install and 12 months support)[iii]
Microsoft	$1,569,970	Microsoft (also includes installation and support)
Oracle	$13,580,000	Oracle's hardware[iv] and software[v] Pricing Sheets
Teradata	$1,168,000	Teradata's Pricing Brochure[vi] (unspecified support time plus installation are included)

Most appliances include both software and hardware – only Oracle and Microsoft break out the retail pricing details between software and hardware. Microsoft SQL Server 2012 Parallel Data Warehouse can be purchased with hardware from HP or Dell.

Vendor	Software Price	Hardware Price
Microsoft	$1,003,970	$566,000
Oracle	$12,480,000 [vii]	$1,100,000[viii]

The price estimates listed above are derived based on the following sources and assumptions:

- EMC does not publish pricing information. At the time the appliance launched, InformationWeek published an article which stated, "Prices will start at [U.S.]$1 million for a half-rack appliance with 18-terabytes of user-available storage."[ix] For a full rack, the EMC Greenplum Data Computing Appliance Cluster comes with 36TB, so the U.S. $1 million estimate is multiplied by two, though software and hardware pricing details were not provided in the article. It is possible that the U.S. $1 million estimate is for a lighter appliance option (as full specs were not listed in the article), so the appliance included in this whitepaper may be more expensive. However, as a conservative estimate the U.S. $1 million (per half-rack) has been used.[x]
- IBM also does not publish pricing information; however, Gemini, an IBM partner, sells the IBM Puredata System for Analytics N1001-010 Appliance and lists "IBM List" pricing as shown above.[xi] While Gemini's site lists slightly discounted pricing, the "IBM List" retail pricing has been used since all other vendor appliances are listed with retail pricing (and even if Gemini's discount was used, it would not impact the results). Gemini's pricing does state that 12 months of support plus installation services are included; it is assumed this is also included in IBM's list price.
- Microsoft's pricing is based on retail software pricing, as well as the hardware appliance cost. Microsoft SQL Server 2012 Parallel Data Warehouse has the option of hardware from two different vendors, Dell and HP.

- Oracle provides a full price list on their Website. An Exadata price list is provided covering appliance hardware[xii], and the Exadata 3-2 Datasheet[xiii] identifies a number of software licenses required to run the appliance. This list was used to refer to the software price list[xiv] to total up software license pricing, as show in the table to the right.
    - Oracle lists software license prices per core, and then references the "Oracle Processor Core Factor Table"[xv] to check if there is a factor to adjust the number of cores based on the specific processor type/family to be considered for software license purposes (factors range from 0.25 to 1.0). For Xeon processors, the core factor is 0.5.
    - The Oracle Exadata 3-2 appliance includes 128 database cores, so the total software license cost is calculated as: U.S. $196,000 x 128 x 0.5 = U.S. $12,544,000.
- Teradata does provide pricing, but only very generally. For the Teradata Data Warehouse Appliance 2690[xvi] (the "2000 Family"), Teradata lists a price of U.S. $32,000 per TB of uncompressed, user available storage[xvii], so U.S. $32,000 x 36.5 TB = U.S. $1,168,000.

Oracle DB Enterprise	$47,500
RAC	$23,000
Partitioning	$11,500
Advanced Compression	$11,500
Advanced Security	$11,500
Active Data Guard	$10,000
GoldenGate	$17,500
RAT	$11,500
OLAP	$23,000
BI Standard Ed	$23,000
Enterprise Manager	$5,000
Oracle Linux	Included
Total Software Cost per Core	**$195,000**

Oracle pricing, as listed in its "Oracle Technology Global Price List," provides details on each software application and add on. The column for "Processor License" is listed above (though actually the pricing is per-core). The above per-core price is multipled by the number of cores, which itself is adjusted based on the Oracle Processor Core Factor Table.

TOTAL USER STORAGE AND COMPRESSION

Each appliance provides about the same amount of total uncompressed storage, around 100TB, except for Microsoft appliances, which come with 288 terabytes of total storage (i.e., the number of hard drives multiplied by the size per hard drive; this includes space taken by the system for software such as the operating system, database, and data tools). But what is more important, and what varies across the appliances, is the amount of user-available storage and the estimate of reasonable compression.

Vendor	User Available Storage Uncompressed (TB)	User Available Storage Compressed (TB)
EMC	36	144
IBM	32	128
Microsoft	68	340
Oracle	45	450
Teradata	37	146

In storage provided in a full rack, Microsoft is the top, or among the top, in both categories and, as described above, significantly outperforms all other vendors in price-to-storage ratios. The storage estimates for each appliance are listed in the table above.

Teradata does not list a specific compression metric in their datasheet specs and marketing materials, but in its pricing datasheet it lists a compression range of "50% to 80%," or in ratio form, from 2:1 to 5:1.[xviii] As 5:1 is the maximum, an efficient (i.e., that

won't likely overly impact performance) compression of 4:1 is used in this whitepaper. The compression estimates for each vendor are listed in the table to the right.

Estimated Compression Ratios for each Vendor	
EMC	4 to 1
IBM	4 to 1
Microsoft	5 to 1
Oracle	10 to 1
Teradata	4 to 1

Oracle lists a 10-15:1 compression based on proprietary compression processes[xix]. Again, an efficient compression (but not detrimental to performance) of 10:1 is used in this whitepaper.

Most user available storage for a full-rack implementation is 125-150 TB, except for Oracle with higher compression providing 450 TB and Microsoft appliances based on larger 1TB high performance drives with a 5-to-1 compression ratio providing 302 TB (for HP) or 340 TB (for Dell).

In addition to the appliance brochure, IBM provides some storage specs in its S-Blade datasheet.[xx]

PROCESSORS AND CORES

Database processors and cores provide the engines for running data management and analysis tasks. Vendors that separate software and hardware pricing license software on a per-core basis, so that is what is detailed here.

As detailed above, Microsoft appliances are as good or outperform all other vendors, and provides as many or more cores than the other vendors. The number of cores included in a full-rack appliance from each vendor is listed in the table to the right.

Vendor	Cores
EMC	48
IBM	112
Microsoft	144
Oracle	128
Teradata	96

INPUT/OUTPUT SPEED

Microsoft again provides high-performance with faster I/O bandwidth as Oracle, but at a much better price-to-performance ratio versus the alternatives (that provide this information).

While I/O Bandwidth was only provided by 3 of 5 vendors, and thus not included in the price-to-performance overview above, it's an important enough performance metric to include and discuss, as an indicator of throughput performance capabilities – both in total and as a factor of price.

Vendor	Input/Output (I/O) Bandwidth (GB/sec)	Price per GB of I/O Bandwidth (per sec)
EMC	24	$83,333
IBM	n/a	n/a
Microsoft	108	$14,537
Oracle	100	$136,440
Teradata	n/a	n/a

CONCLUSION

Overall, Microsoft not only provides data warehouse appliances that are competitively priced compared to other options, but also the best price-to-performance value.

APPENDIX

Methodology:

In a study commissioned by Microsoft, similar EDW appliances from five leading vendors have been reviewed, summarized, and compared, based on publicly-available pricing and solution specification information. They are:

- Microsoft SQL Server 2012 Parallel Data Warehouse, with hardware provided by HP or Dell
- Oracle Exadata Database Machine X3-2 (SC)
- Teradata Data Warehouse Appliance 2690 (600 GB disks)
- EMC Greenplum Data Computing Appliance (Standard Module)
- IBM PureData System for Analytics N1001-010

These solutions were chosen as some of the top vendors and appliances discussed in the February, 2011 Forrester Wave report on enterprise data warehousing platforms.[xxi] Teradata's Data Warehouse Appliance 2690 is used, which is a newer version of Data Warehouse Appliance 2650 listed in the Forrester report.

High Performance or Standard (as opposed to High Capacity) appliance options were selected for consistent comparison.

Full racks were compared, to ensure standard comparisons across large-capacity appliances.

Summary metrics, list price, total user storage space (compressed and uncompressed), and performance (if made public) were compared. Price per terabyte of user space is used as a price/value approximation that can help provide more comparison details when the amount of user space is not the same across all appliances.

Comparisons are based on the list price and publicly available metrics. Each discussion with your vendor and their customer discount situation will be different, and more information (and customization) is likely available from each vendor. Individual results may be different than the ones listed here which use only list prices and specifications.

Appliance price is only part of the cost considerations. Customer should carefully evaluate installation, migration, and on-going management cost before making a decision.

About Value Prism Consulting

Value Prism Consulting is a valuation and management consulting firm that provides services to a broad range of clients worldwide. Our Management Consulting Practice assists clients with business case development and decision support analysis. Our solutions measure the results of various options related to business process improvement, capital and major budget-spending decisions, including infrastructure upgrades, and the value of new product introductions. Visit www.valueprism.com for more information.

Sources:

[i] Oracle. http://www.oracle.com/us/products/database/exadata/overview/index.html

[ii] InfoWeek. (2010, October 13). *EMC Intros Backup-Savvy Greenplum Appliance*. http://www.informationweek.com/software/information-management/emc-intros-backup-savvy-greenplum-applia/227701321

[iii] Gemini. (2011). *PureData System Analytics N1001-010*. http://estore.gemini-systems.com/ibm/software-license/information-management-software/netezza-1000/netezza-1000-12/

[iv] Oracle. (2012, December 4). *Oracle Engineered Systems Price List.* http://www.oracle.com/us/corporate/pricing/exadata-pricelist-070598.pdf

[v] Oracle. (2012, December 18). *Oracle Technology Global Price List.* http://www.oracle.com/us/corporate/pricing/technology-price-list-070617.pdf

[vi] Teradata. (2012). *Teradata Purpose-Built Platform Pricing.* http://www.teradata.com/t/brochures/Teradata-Purpose-Built-Platform-Pricing-eb5496

[vii] Oracle. *Oracle Technology Global Price List.*

[viii] Oracle. *Oracle Engineered Systems Price List.*

[ix] Infoweek

[x] EMC. (2012). *EMC Greenplum Data Computing Appliance*. http://www.greenplum.com/sites/default/files/2012_0419_EMC_GP_DCA_Computing_Appliance.pdf

[xi] Gemini. (2011). *PureData System Analytics N1001-010*. http://estore.gemini-systems.com/ibm/software-license/information-management-software/netezza-1000/netezza-1000-12/

[xii] Oracle. *Oracle Engineered Systems Price List.*

[xiii] Oracle. (2012). *Oracle Exadata Database Machine.* http://www.oracle.com/us/products/database/exadata-db-machine-x3-2-1851253.pdf

[xiv] Oracle. *Oracle Technology Global Price List.*

[xv] Oracle. (2009, March 16). *Oracle Processor Core Factor Table.* http://www.oracle.com/us/corporate/contracts/processor-core-factor-table-070634.pdf

[xvi] Teradata. (2012). *Teradata Data Warehouse Appliance 2690*. http://www.teradata.com/brochures/Teradata-Data-Warehouse-Appliance-2690/

[xvii] Teradata. (2012). *Teradata Purpose-Built Platform Pricing*.

[xviii] Teradata. (2012). *Teradata Purpose-Built Platform Pricing.*

[xix] Oracle. (2012, November). *Hybrid Columnar Compression (HCC) on Exadata.* http://www.oracle.com/technetwork/middleware/bi-foundation/ehcc-twp-131254.pdf

[xx] IBM. (2012). *BladeCenter S Chassis*. http://www-03.ibm.com/systems/bladecenter/hardware/chassis/blades/index.html

[xxi] Forrester. (2011, February 10). *The Forrester Wave™: Enterprise Data Warehousing Platforms*, Q1 2011.