



Microsoft®  
**SQL Server® 2012**



# SQL Server 2012 Parallel Data Warehouse

Solution Brief

Published February 22, 2013

**Microsoft**

# Contents

- Introduction..... 1
- Microsoft Platform: Windows Server and SQL Server ..... 2
- SQL Server 2012 Parallel Data Warehouse..... 2
  - Built for Big Data ..... 3
  - Next-Generation Performance at Scale ..... 4
  - Engineered for Optimal Value..... 4
- Conclusion ..... 5
- Call to Action ..... 5

## Key Highlights

- Up to 100 times faster performance than legacy data warehouse queries
- Up to 50 times faster query performance over the prior version of Microsoft SQL Server Parallel Data Warehouse
- Up to 15 times data compression
- Up to 70% more storage capacity
- Up to double the rate of data loading speed over prior versions
- Data capacity requirement variable from smallest (1 terabyte) to largest (8 petabytes)
- 75% smaller footprint by using maximum density storage and a revamped architecture
- Lowest Price/Terabyte in the industry

## Key Customer Benefits

- Get faster insights than ever before with xVelocity updateable columnstore technologies
- Query both worlds—relational and non-relational data from Hadoop—from easy-to-use and familiar tools like Microsoft Excel
- Scale painlessly, avoid forklift upgrades, and successfully expand up to multiple petabytes of data with little disruption
- Start small and grow as needed with a modular approach that minimizes risks and over-investment
- Ensure a successful deployment with key consulting assistance and technical support from Microsoft Services

## Introduction

Today, data is the new currency of the enterprise. This means organizations that know how to maximize the value of data will thrive. Gartner<sup>1</sup> asserts that “By 2015, organizations integrating high-value, diverse, new information types and sources into a coherent information management infrastructure will outperform their industry peers financially by more than 20%.” McKinsey<sup>2</sup> agrees with Gartner, confirming that organizations that use data and business analytics to drive decision-making are more productive and deliver higher return on equity than those that do not.

There has never been a more exciting time in the world of data. Enterprises are seeing the convergence of new information types and sources that are fundamentally transforming the industry. For example, the volume of information a typical organization needs to manage is exploding, and most of the growth is coming from beyond traditional data types in the form of unstructured data. Current software solutions cannot cope with the complexity of this data or the tremendous workload growth it is creating. Transitioning to backend solutions that support non-traditional data sources—like Big Data—is the only way to effectively manage this data, but organizations are failing to make this transition because of vendor lock-in, costly deployments, and time-consuming migrations.

To help organizations successfully transition to this new world of data, Microsoft introduced SQL Server 2012 Parallel Data Warehouse (PDW)—the next version of the scale-out massively parallel processing (MPP) data warehouse appliance. Built

---

<sup>1</sup> Casonato, Regina et al. “Information Management in the 21st Century.” Gartner. 2 Sept. 2011.

[http://download.microsoft.com/download/7/B/8/7B8AC938-2928-4B65-B1B3-0B523DDFCDC7/Big%20Data%20Gartner%20information management in the 21st%20Century.pdf](http://download.microsoft.com/download/7/B/8/7B8AC938-2928-4B65-B1B3-0B523DDFCDC7/Big%20Data%20Gartner%20information%20management%20in%20the%2021st%20Century.pdf)

<sup>2</sup> “Perspectives on Digital Business.” McKinsey & Company. Jan. 2012.

[http://www.mckinsey.com/~media/mckinsey/dotcom/client\\_service/BTO/PDF/MCBT\\_Compedium\\_Perspectives\\_on\\_Digital\\_Business.ashx](http://www.mckinsey.com/~media/mckinsey/dotcom/client_service/BTO/PDF/MCBT_Compedium_Perspectives_on_Digital_Business.ashx)

on industry-standard database and commodity hardware, the PDW solution provides a generic data platform for customers to unlock the value of their data. In short, SQL Server 2012 PDW offers customers a premium solution without a premium price tag.

## Microsoft Platform: Windows Server and SQL Server

Microsoft provides a complete solution suite of technologies to help enable an interoperable, end-to-end infrastructure. This helps IT departments to maximize return on investment and enable mission-critical environments at a low total cost of ownership (TCO).

With Windows Server 2012, Microsoft has transformed the definition of server operating systems by providing enhanced Hyper-V virtualization, storage spaces, runtime memory configuration, and more.

Microsoft's optimal database platform, SQL Server 2012, enables mission-critical environments, offering availability and performance at a low TCO. Customers can benefit from other enterprise class features of SQL Server 2012, including xVelocity and automatic memory tuning.

Microsoft has always focused on delivering best-in-class enterprise solutions—from consulting services to patching to integration within the system, and even the use of different tools like SQL Server Data Tools, SQL Server Integration Services, or other Business Intelligence (BI) tools. Regardless of the solution, Microsoft has delivered it with full authority and almost always offered it as a built-in feature, at no additional cost.

## SQL Server 2012 Parallel Data Warehouse

SQL Server 2012 PDW has evolved to fully embrace the new world of data. It can scale out to handle requirements of virtually any size, perform seamless analysis on datasets sourced from unstructured and semi-structured data from Hadoop, and provide next-generation query performance improvements of up to 100 times over legacy data warehouse deployments.

Simple to deploy, SQL Server 2012 PDW is delivered as a pre-built appliance with software, hardware, and networking components already pre-installed. SQL Server 2012 PDW is also engineered to deliver optimal value by providing the lowest price per terabyte of any high-end warehouse appliance offering in the market today.

### Key Design Elements



- More high density storage and compute power with Microsoft Windows Server 2012 Storage Spaces and Hyper-V
- Increased speed and scalability requiring fewer computing resources and less IT effort
- Up to 15 times compression of data to help save up to 70 percent of storage requirements
- 50-percent reduction in hardware cost along with 50-percent reduction in energy consumption

## Value Proposition

**Built for Big Data:** Enables integrated query across Hadoop and relational data with PolyBase, a fundamental breakthrough on the data processing engine

**Next-generation performance at scale:** Delivers next-generation performance on datasets of nearly any scale—up to 50 times faster than before—with a primary storage engine that includes a new updateable version of xVelocity columnstore, and enables users to linearly scale out from small terabyte requirements to petabytes of data

**Engineered for optimal value:** Redefines the PDW appliance with software innovations such as xVelocity columnstore, PolyBase, Windows Server 2012 Hyper-V, and Storage Spaces, while providing a monitoring pack for integration with existing deployments of Microsoft System Center

*“By extending our current SQL Server 2012 business intelligence implementation to include SQL Server PDW, we’ve enhanced our self-service data portal to accommodate large data sets, as well as daily point-of-sale data, which provides on-demand results from our clients’ sales and marketing efforts. The fast, on-demand data access that SQL Server PDW provides through our self-service portal will give our clients tremendous opportunities to leverage shopper insights into their strategic planning and tactical operations.”*

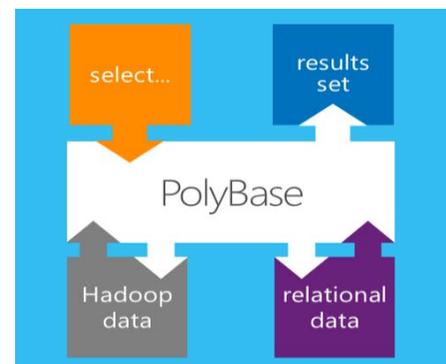
-Jim Norred,  
Executive Vice President and Chief Information Officer,  
CROSSMARK

<http://www.microsoft.com/en-us/news/Press/2012/Nov12/11-13CROSSMARKPR.aspx>

## Built for Big Data

SQL Server 2012 PDW uses PolyBase, a breakthrough in data processing to provide customers integrated query across Hadoop and relational data as well as integration with Microsoft BI.

**Seamless integration of Hadoop data with the data warehouse in a single query:** PolyBase can accept a standard Transact-SQL query that joins tables from a relational source with tables from a Hadoop cluster referencing a non-relational source and seamlessly return the results to the user. IT is not required to pre-load data from Hadoop into the warehouse and users are not required to learn MapReduce to make the query. This gives organizations the ability to perform interactive analysis on data of virtually any velocity, complexity, and size with simplicity. With the export option in PolyBase, organizations can store their historical web log files in a directory on a Hadoop cluster on low-cost commodity hardware, and then process it with PolyBase by specifying the directory location while performing the query.



**BI Integration:** Native Microsoft BI integration enables users to quickly create compelling visualizations and make key business decisions from nearly any type of data (relational or non-relational) from familiar tools such as Microsoft Excel. Power View is built into Excel, enabling users to quickly and easily visualize data, while sharing and collaborating on business insights in a familiar Microsoft SharePoint Server environment. Additionally, Power View eliminates the need to

build cubes and enables organizations to query much more data without any constraints on tabular model limits.

## Next-Generation Performance at Scale

To provide next generation performance at scale, SQL Server 2012 PDW uses the new updateable version of xVelocity columnstore combined with an MPP engine for seamless linear scale out.

**xVelocity columnstore:** SQL Server 2012 PDW uses the newest version of xVelocity memory-optimized columnstore, which is both updateable and clustered. By reorienting data in a column rather than a traditional row store, organizations can see next-generation performance (up to 50 times) by reducing the query time from hours to seconds and high compression on their data (up to 15 times). This is due to the grouping of data in a columnar format (containing values from multiple rows) that allows for higher efficiency in both storage and the returned dataset which ultimately reduces costs and delivers more with existing investments. The ability to permanently convert traditional row store tables with the updateable columnstore of xVelocity is another new feature, which dramatically reduces storage (by up to 70 percent) while still providing real-time data warehouse query performance.

**Massively parallel processing:** SQL Server 2012 PDW uses an MPP scale-out engine that offers powerful distributed computing and scale. This appliance uses a shared-nothing architecture with multiple physical nodes, where each node runs its own instance of SQL Server with dedicated CPU, memory, and storage. As queries are initiated, they are distributed evenly across and run simultaneously over each physical node. The benefit of MPP architecture is that it gives the highest performance at scale through parallel execution. Users need only to add more capacity to continually scale out this implementation up to

petabytes of data and handle complex and concurrent queries without requiring the flattening of existing data models or compromising business relationships. In addition, implementations can accommodate high concurrency and process complex queries at scale. SQL Server 2012 PDW offers proven scalability on all dimensions to enable greater performance and lower the TCO, as well as faster data loading for excellent trending analysis. There is no need to limit the data for performance reasons; instead, customers can ask ad-hoc questions without experiencing compromised performance or loading bottlenecks.

**Add capacity with ease:** With traditional data warehouses that scale up, organizations need to rebuild entire servers once the maximum physical capacity and performance of servers hosting the data warehouse is exceeded. With MPP scale-out architecture, SQL Server 2012 PDW will scale out to virtually any or all future data requirements customers encounter, starting from the smallest requirements (0–15 terabytes) up to the very largest (8 petabytes). With MPP architecture organizations buy only what they need and scale up by adding racks of compute nodes by avoiding forklift upgrades. The base system starts with one rack, for example. A full rack holds 10 nodes, and additional 10-node racks can be added up to a total of 40 nodes.

## Engineered for Optimal Value

**Ease of setup:** SQL Server 2012 PDW is a pre-built appliance that ships to the customer's door as a fully packaged appliance solution. Customers do not need to worry about building the most optimal hardware mix (in terms of CPU, storage, and memory) or figuring out the exact science of installing and tuning the software. They just have to plug the appliance in and start integrating their specific data into the solution.

The fundamental architecture of SQL Server 2012 PDW, combined with automated features, reduces the need for human involvement. PDW gives

database administrators simple database and table definition processes, so they no longer need to toil with detailed space, memory, or cache management. Finally, System Center is included in-the-box, which simplifies the IT management of the entire appliance (hardware, operating system, warehouse, and applications) from a single tool at no additional cost.

**Built-in software features lead to efficient**

**hardware:** SQL Server 2012 PDW has reshaped the very hardware specifications required of an appliance through software innovations that help deliver optimal value to customers. PDW delivers customers an exceptional value through Windows Server 2012 Storage Spaces, which has built-in reliability and performance features. Storage Spaces helps to significantly reduce the cost of storage by ensuring the appliance uses an economical hardware option of high-density direct attached disks.

Hyper-V virtualization, delivered in Windows Server 2012, and hardware design based on homogeneous nodes completely eliminate the "control rack" from the previous architecture. This dramatically lowers the hardware footprint and cost of the appliance.

**One of the lowest on-premises prices per**

**terabyte in the industry:** SQL Server 2012 PDW is the only solution that continues to add value while also lowering the cost of entry to acquire a complete solution. Microsoft lowers the cost by reducing the hardware footprint through virtualization, providing resilient and scalable performance features on the software e, and with compression (up to 15 times) to lower storage usage. All of these features help make SQL Server 2012 PDW the lowest price per terabyte over every other company by a significant margin.

**Choice of hardware vendors:** HP and Dell provide hardware and support options that best fit customers' infrastructure needs with true quarter-rack, half-rack, and full-rack solutions.

**Dell Parallel Data Warehouse Appliance**



**HP Enterprise Data Warehouse Appliance**



	Dell	HP
Servers	Poweredge R620	ProLiant Gen8 DL360
Compute Nodes	Up to 9 per Rack (3 minimum)	Up to 8 per Rack (2 minimum)
Racks	¼ to 6	¼ to 7
Raw Disk Capacity (uncompressed)	0TB-1.2PB	0TB – 1.2PB

*SQL Server 2012 PDW allows you to visualize what you would do if you could query all your data and more within seconds.*

## Conclusion

SQL Server 2012 PDW is ready to meet the demanding challenges customers face—seamlessly integrating structured and unstructured data—with revolutionary performance improvements at an affordable price. With SQL Server 2012 PDW and its enhanced PolyBase and xVelocity technologies, customers can turn their data into valuable insights faster than ever before, with minimal investments.

## Call to Action

For more information on the PDW appliance, go to <https://www.microsoft.com/sqlserver/en/us/solutions-technologies/data-warehousing/pdw.aspx>.