

: *What is a data warehouse appliance?*

by Dan Power

Editor, DSSResources.com

In general, a data warehouse (DW) appliance is an integrated hardware and software bundled application. A DW appliance includes server hardware and premium storage technology with an installed operating system, database management system and application software tuned for data warehousing. Most DW appliances use massively parallel processing (MPP) architectures to provide fast query performance and platform scalability. An appliance is a purpose specific device.

A data warehouse appliance can range from a small capacity device to a much more powerful platform for storage and querying. Many argue Netezza was the first vendor to offer data warehouse appliances. Netezza started operations in the early 2000s. The company was acquired by IBM in 2010. Netezza's appliances use a proprietary Asymmetric Massively Parallel Processing (AMPP) architecture that "combines open, blade-based servers and disk storage with a proprietary data filtering process using field-programmable gate arrays".

According to Howard (2006), Netezza's AMPP uses a "zonemap. A zonemap allows you to load say, sales by time, and then the zonemap breaks the relevant data down into blocks, storing the details of the first and last record in each block (thus there is a much lower overhead compared to an index). What this means is that when you run a query you only read the blocks that contain the data you are interested in, ignoring all the other blocks. This ability to limit the data you read means that joins are much more effective than would otherwise be the case."

In February 2010, Netezza announced that it had opened up its systems to support major programming models, including Hadoop, MapReduce, Java, C++, and Python models.

According to the IBM Netezza (<http://www.netezza.com>) website, "IBM Netezza data warehouse appliances are purpose-built to make advanced analytics on large volumes of data simpler, faster and more accessible. Now, companies like AOL, eHarmony, Intuit, Nielsen and Acxiom can better understand their customers and increase revenues by delivering the right message to the right audience at the right time. ... IBM Netezza data warehouse appliances are easy to implement, quick to deploy and simple to maintain for optimal efficiency and fast time to value."

Netezza's main competitors include Oracle Exadata and Teradata. Oracle Exadata Database

: *What is a data warehouse appliance?*

Machine (<http://www.oracle.com>) uses Oracle Exadata Storage Servers, which combine smart storage software and industry-standard hardware. Oracle Exadata Storage Servers use a massively parallel architecture to increase data bandwidth between the database server and storage. The Teradata Data Warehouse Appliance (<http://www.teradata.com>) features the Teradata Database, a Teradata platform with dual quad core Nehalem Intel® processors, SUSE® Linux operating system, and enterprise-class storage. Other competitors include Microsoft SQL server bundled with HP technology. The HP Enterprise Data Warehouse appliance is optimized for the SQL Server 2008 R2 Parallel Data Warehouse product.

Data warehouse appliances are cost effective for small, dedicated data warehouses. As data warehouse applications proliferate a company's IT managers may be tempted to install multiple DW appliances which can lead to problems of data duplication and data discrepancies among data marts/warehouses.

References

Howard, P. "Netezza surprises with technical capabilities," The Register, October 3, 2006 at URL http://www.theregister.co.uk/2006/10/03/netezza_annual_conference_roundup/

Morgan, T. "Netezza to bake analytics into appliances," The Register, February 24, 2010 at URL http://www.theregister.co.uk/2010/02/24/netezza_data_analytics/

Power, D. "What is a data warehouse?" DSS News, Vol. 11, No. 24, December 5, 2010 at URL <http://dssresources.com/faq/index.php?action=artikel&id=216> .

Author: Daniel Power

Last update: 2012-02-19 05:21