# Enterprise data warehouses

## Introduction

This paper is one of five. The first, called a Market Report, gives a detailed background for the other four. These four are split into two groups, the first three papers are on analytic data marts and warehouses and the fourth is on enterprise data warehouses (EDW). The difference is that the latter includes functions such as support for master data management and operational BI, which have different support requirements from purely analytic environments, even if the latter are combined with traditional business intelligence. Secondly, the analytic papers are organised by scalability for small (up to 5Tb), medium (5Tb to 50Tb) and large (50Tb+) environments respectively. We have not applied similar scaling principles to the EDW paper because all the companies providing the necessary mixed query workload capabilities also scale well.

This paper discusses enterprise data warehouses. By this we mean an analytic warehouse that also incorporates requirements to support such things as master data management and operational BI where there are a lot of look-ups and small queries running alongside the more complex needs of analytics. This puts a much greater emphasis on the ability to manage mixed query workloads. As a result, we have excluded from this update all products that do meet a certain level of functionality with respect to mixed queries. Note that products thus excluded may be entirely suitable for deployment as an EDW where requirements are simpler or less stringent.

## The market

The EDW market continues to evolve rapidly. We have seen the introduction of the HP/Oracle Database Machine, which Oracle now sees as its major offering in the warehousing market as opposed to Oracle per se, while SAP has moved from a purely row-based architecture to a hybrid one that combines rows with columns in the BW Accelerator, albeit that this is limited to support for InfoCubes for the present. Moreover, in just the last couple of months Vertica has joined the MapReduce bandwagon (following on from Aster Data and Greenplum, though in a rather different way) and released its FlexStore architecture, Exasol has announced new mixed query workload capabilities, and Netezza has moved away from its previous proprietary environment with the announcement of its TwinFin.

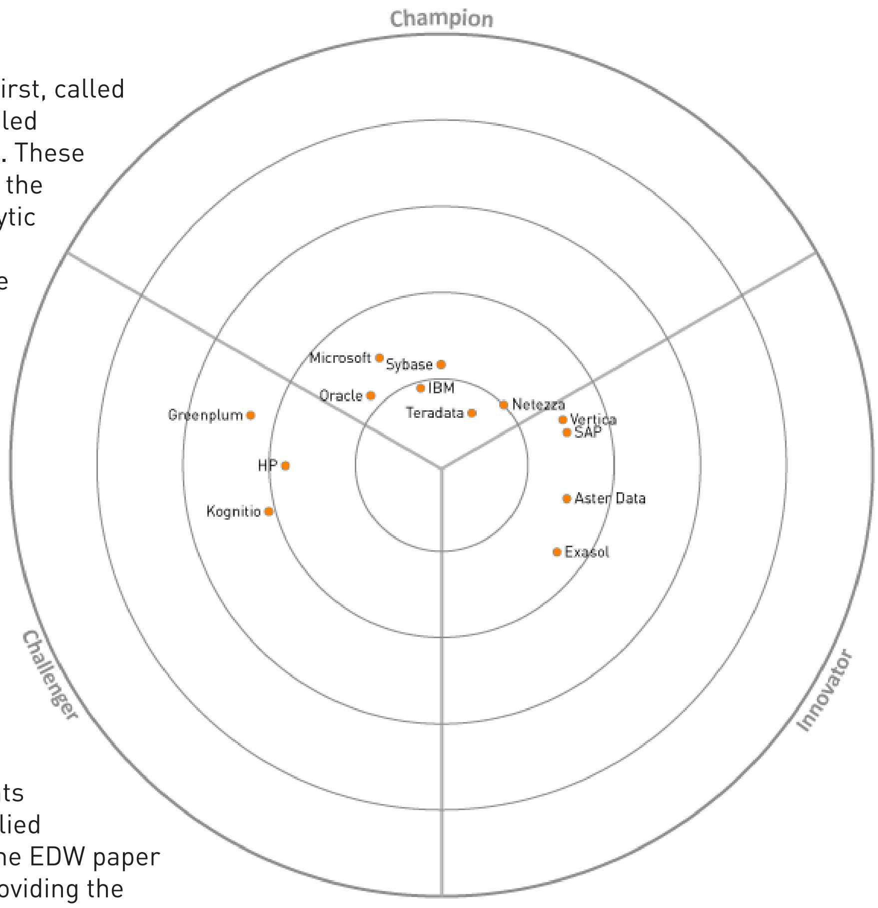


**Figure 1:** The highest scoring companies are nearest the centre. The analyst then defines a benchmark score for a domain leading company from their overall ratings and all those above that are in the champions segment. Those that remain are placed in the Innovator segment if their innovation rating is over 2.5 and Challenger if it is less than 2.5. The exact position in each segment is calculated based on their combined innovation and overall score.

There are also numerous major initiatives waiting to come fruition. Teradata is porting its technology to SAP BW to act as a host to the BW environment in place of Oracle, SQL Server, DB2 or MaxDB; and SAP itself has various plans for more closely integrating the business intelligence, data integration and data quality capabilities of Business Objects. Then there is IBM's Smart Analytic Appliance, Microsoft's Madison project (based on the technology it acquired from DATAllegro), and Netezza is planning some significant additions to its mixed query workload capabilities. We would like to see some major developments from HP because, at the present time, NeoView lacks momentum.

With all this going on it is sometimes difficult to see the wood for the trees. In so far as market positioning is concerned, Teradata seems to have weathered the storm caused by the introduction of appliance vendors, and various companies, as already discussed, have reacted, or are reacting, to counter their threat. At the same time, Netezza is continuing to grow, as are other vendors, most notably Vertica which we rate as having overtaken Greenplum as the second leading supplier amongst the new boys, after Netezza.

In general, there is now greater acceptance of column-based approaches to analytics and the biggest beneficiary of this is arguably Sybase, which is now the fourth largest warehousing vendor in terms of customers and installations after IBM, Oracle and Microsoft (and excluding SAP). Other major trends include MapReduce, support for in-database analytics and mining, cloud computing and warehousing as a service. In this last category, Kognitio is making a big play with its DaaS (data warehousing as a service) offering.

## The vendors

The Bullseye shown is, of course, a snapshot. We see Teradata followed by IBM, and then Netezza followed by Oracle and Sybase, as the leading vendors in this market. However, there may be very good reasons for selecting any particular vendor such as a requirement to support spatial analytics or MapReduce. As one might expect, there are significantly fewer vendors in this update than in our analytic warehousing reports but this will change as more and more vendors add mixed query workload support.

*Philip Howard*
*Data Management*
*September 2009*

For additional information relating to this subject visit http://www.BloorResearch.com/update/1048

Bloor
2nd Floor
145–157 St John Street
London, EC1V 4PY
United Kingdom

Tel: +44 (0)207 043 9750
Fax: +44 (0)207 043 9748
Web: www.BloorResearch.com
email: info@BloorResearch.com