Data Warehouse Landscape Q4 2014

The data warehouse may be a relatively mature concept, but the technology within it continues to innovate rapidly. A key reason for this is the need to cope with inexorably increasing data volumes. In 2003 the largest data warehouse in the world was 30 TB in size, yet there are numerous examples now of petabyte sized operational data warehouses, a more than 30 fold increase in just a decade. A 2012 Information Difference survey of 209 customers showed that most were experiencing data growth of 20-50% annually. Traditional databases have begun to creak under the strain.

The challenges of such "big data" growth, particularly to handle the volumes of data being generated by machines (such as sensors) and web traffic has led to non-traditional file storage mechanism such as Hadoop. Implementations of this such as those of Cloudera, Hortonworks and MapR are gaining increasing traction. Although in practice most such technologies are aimed at tackling less structured data than is the natural territory of data warehouses, their low cost and scalability has led some to consider them as an alternative to traditional data warehousing. However an in-depth survey on this subject at the end of 2014 by the Information Difference leads us to conclude that the worlds of Hadoop and the data warehouse are, at least for now, quite distinct and complementary. Existing database vendors have moved quickly to offer Hadoop connectors for their products.

The database world has not stood still either, with several NoSQL based databases appearing in the last few years, some specifically aimed at handling analytic query workloads. Such databases are often marketed to customers based on their perceived lower data administration needs due to their non-traditional approaches to the database schema. Hardware is helping too, with greater use of in-memory approaches to data warehousing as memory becomes cheaper. Given that memory is more than an order of magnitude faster than disk storage, its appeal is obvious for challenging workloads, and has been incorporated in existing databases like Teradata as well as in newer ones.

The data warehouse is not immune to the general market trend towards cloud-based solutions rather than on-premise. Although the majority of data warehouses are still on-premise, there is a steady drift towards cloud deployments, with some specialist offerings like Amazon Redshift appearing that are cloud-only. Given the increasingly well-established economic advantages of cloud deployment, it seems inevitable that this trend will continue, as customers become more used to the idea and their concerns about security and reliability recede.

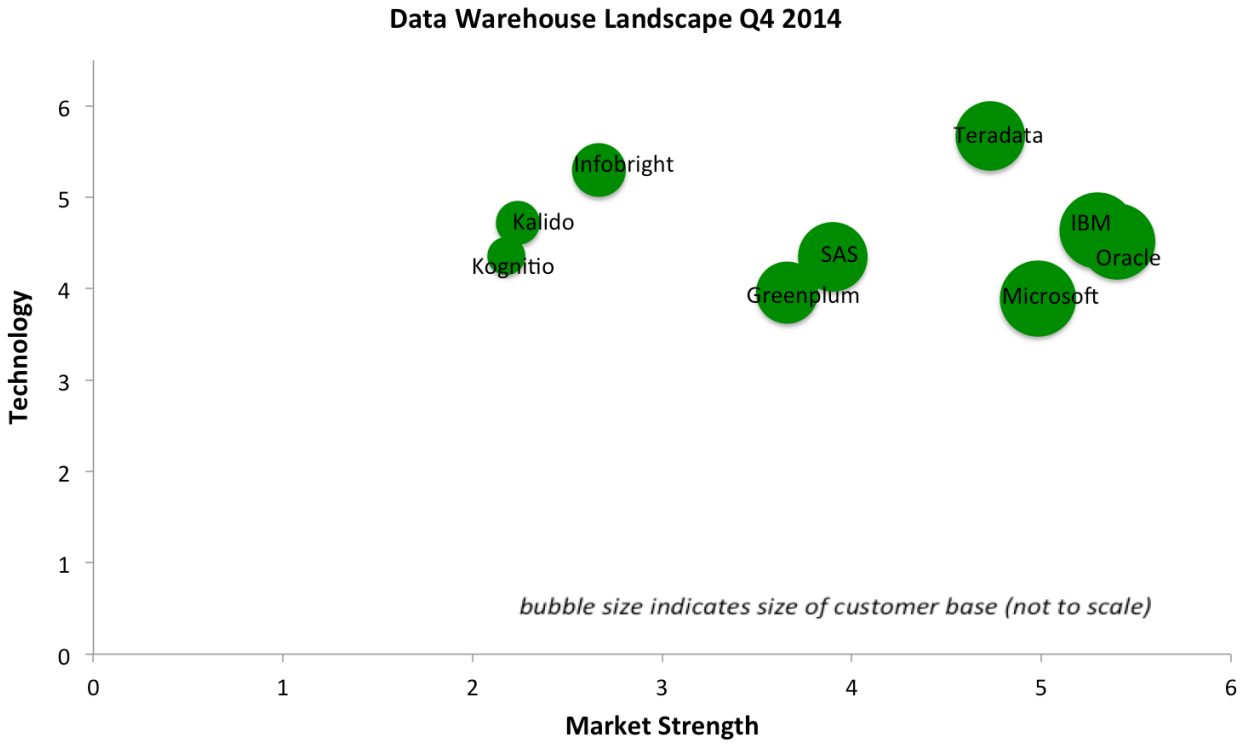
Within the data warehouse world, the largest vendors remain Oracle, IBM, Microsoft and Teradata, with Greenplum (now part of Pivotal) and SAS Institute being other large-scale providers. Assorted niche providers fill out the market, including the data warehouse application of Kalido, the data warehouse appliance of Kognitio and the columnar database of InfoBright, which focuses on handling machine-generated data. Increasingly, but not exclusively, columnar approaches are used for large-scale data warehouses. In general, columnar databases allow greater compression

Copyright © 2014 The Information Difference Ltd. All Rights Reserved.

than row-based and offer faster performance for queries at the expense of slower load times. Some traditional database vendors now offer columnar options “under the covers” for suitable database workloads.

The data warehouse world continues to be surprisingly dynamic for such as a well-established market. This is largely due to the influx of newer competitors in response to the increase in data volumes that have challenged the traditional relational database vendors. The combination of new competitors and customer frustration has forced the major vendors to react, either by connecting to Hadoop stores or by acquiring or partnering with some of the newer technologies. The coming year promises to be an interesting one.

The main vendors in the market are summarised in the diagram below.



The landscape diagram represents the market in three dimensions. The size of the bubble represents the customer base of the vendor, i.e. the number of corporations it has sold data warehouse software to, adjusted for deal size. The larger the bubble, the broader the customer base, though this is not to scale. The technology score is made up of a weighted set of scores derived from: customer satisfaction as measured by a survey of reference customers[1], analyst impression of the technology, maturity of the technology in terms of its time in the market and the breadth of the technology in terms of its coverage against our functionality model. Market strength is made up of a weighted set of scores derived from: data warehouse revenue, growth, financial strength, size of partner ecosystem, (revenue adjusted) customer base and geographic coverage. The Information Difference maintains vendor profiles that go into more detail. Customers are

[1] In the absence of sufficient completed references, a neutral score was assigned to this factor

Copyright © 2014 The Information Difference Ltd. All Rights Reserved.

encouraged to carefully look at their own specific requirements rather than high-level assessments such as the Landscape diagram when assessing their needs.

A significant part of the "technology" dimension scoring is assigned to customer satisfaction, as determined by a survey of vendor customers. In this research cycle the vendors with the happiest customers were Teradata, followed by InfoBright and Kalido. Our congratulations to those vendors.

Below is a list of the significant data warehouse vendors.

Vendor	Brief Description	Website
Actian	Actian's product is an analytic database on commodity hardware.	www.actian.com
Amazon Redshift	Cloud-based data warehouse solution.	aws.amazon.com/redshift/
Exasol	German data warehouse appliance vendor.	www.exasol.com
Greenplum	Appliance vendor aiming at high-end warehouses, now part of Pivotal.	www.pivotal.io/big-data/pivotal-greenplum-database
IBM	Infosphere Balanced Warehouse (formerly DB2) is the data warehouse software offering from the industry giant, which also offers two appliances: PureData for Operational Analytics (based on DB2) and PureData for Analytics powered by Netezza technology.	www.ibm.com
InfoBright	Provides a columnar-database analytics platform.	www.infobright.com
Kognitio	Mature data warehouse appliance, offering its data warehouse as a service.	www.kognitio.com
Kalido	Information management vendor (now part of Magnitude Software) with data warehouse solution.	www.kalido.com
MarkLogic	Enterprise NoSQL database vendor.	www.marklogic.com
Microsoft	As well as its SQL Server relational database, Microsoft acquired Data Allegro and at the end of 2010 launched its Parallel Warehouse based on this technology.	www.microsoft.com
MonetDB	MonetDB is an open-source columnar database system for high-performance applications.	www.monetdb.org
MongoDB	Open source document database.	www.mongodb.org
Neo4j	Open source graph database.	www.neo4j.org

Copyright © 2014 The Information Difference Ltd. All Rights Reserved.

Oracle	Database and applications giant with its own data warehouse appliance.	www.oracle.com
ParStream	Columnar, in-memory, MPP database vendor aimed at analytic processing.	www.parstream.com
Sand	Focuses on allowing customers to-effectively retain massive amounts of compressed data in a near-line repository for extended periods.	www.sand.com
SAP/Sybase	Sybase was a pioneer in column-oriented analytic database technology, acquired in mid-2010 by giant SAP. SAP is also now offering the in-memory database technology HANA.	www.sap.com
SAS Institute	Comprehensive data warehouse technology from the largest privately owned software company in the world.	www.sas.com
1010 Data	Provides column-oriented database and web-based data analysis platform.	www.1010data.com
Teradata	Database giant with its own data warehouse solution.	www.teradata.com
Vertica	Appliance vendor Vertica was purchased by HP in 2011.	www.vertica.com
WhereScape	Not an appliance, but a framework for the development and support of data warehouses.	www.wherescape.com

Copyright © 2014 The Information Difference Ltd. All Rights Reserved.