Data Warehouse Landscape Q4 2015

The data warehouse concept appeared in the late 1980s, the idea being to maintain a separate copy of transactional system data for purely analytical purposes. This was necessary partly to overcome mixed workload constraints in the databases of the day, and partly to consolidate data from multiple corporate information sources. Database technology has radically changed since those early days, but the notion of maintaining a specialist analytical data source has remained, sometimes based on a different database platform from the core transactional business applications like ERP, CRM and supply chain systems.

Data warehouses have to deal with ever-increasing data volumes. In 2003 the largest data warehouse in the world was 30 TB in size, yet there are many examples now of petabyte sized operational data warehouses, a 30 fold+ increase in just over a decade. A 2012 Information Difference survey showed that most customers were experiencing data growth of 20-50% annually, and there is no sign that the rapid pace of data growth is relenting. Traditional databases have begun to creak under the strain.

Nowadays data warehouses have to adapt to the rise of "big data", which is stored in non-traditional file systems such as Hadoop and may include machine-generated data, text, and images. Vendors have had to co-exist with this emerging source of data, either providing adaptors to big data file systems or acquiring or developing technology that is able to handle such data within their existing platform. An example may be to store such data in a separate physical store but to build an optimiser capable of running queries across these different data stores.

These additional challenges have pushed the traditional SQL-based database to its limits, and we are seeing the rise of newer database technologies (NoSQL) not based on the relational model, sometimes with dynamic rather than traditional fixed schemas. 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. However we expect to see this distinction blur over time.

A further challenge to the traditional models has been the inexorable rise of cloud storage as an alternative to on-premise technology within the enterprise. This approach, which promises more scalable platforms that are simpler to maintain for the end-user, is steadily eroding the traditional boundaries of data being stored within the physical data centres of a company. Data warehouses are expected to be capable of being deployed in either a private or public cloud, the latter being exemplified by the advent of Amazon Redshift into the market.

Within the data warehouse world, the largest vendors remain Oracle, IBM, Microsoft and Teradata, with Greenplum (now ultimately owned by Dell) and SAS Institute being other large-scale providers. Assorted niche providers fill out the market, including the data warehouse application of Kalido. Increasingly, but not exclusively, columnar approaches are used for large-scale data warehouses. In general, columnar databases allow greater compression 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 shows sign of both consolidation and innovation, as the large established vendors acquire innovative technologies in the race to stay ahead of the challenges of the market. Data warehouses are being pulled in several directions, having to cope not just with greater data volumes but with non-traditional data types as well as being expected to cope with a mix of deployment options, both on-premise and cloud. The significant challenges that result are encouraging the advent of innovative start-ups that in time may reshape the data warehouse landscape considerably.

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

Data Warehouse Landscape 2015

bubble size indicates size of customer base (not to scale)
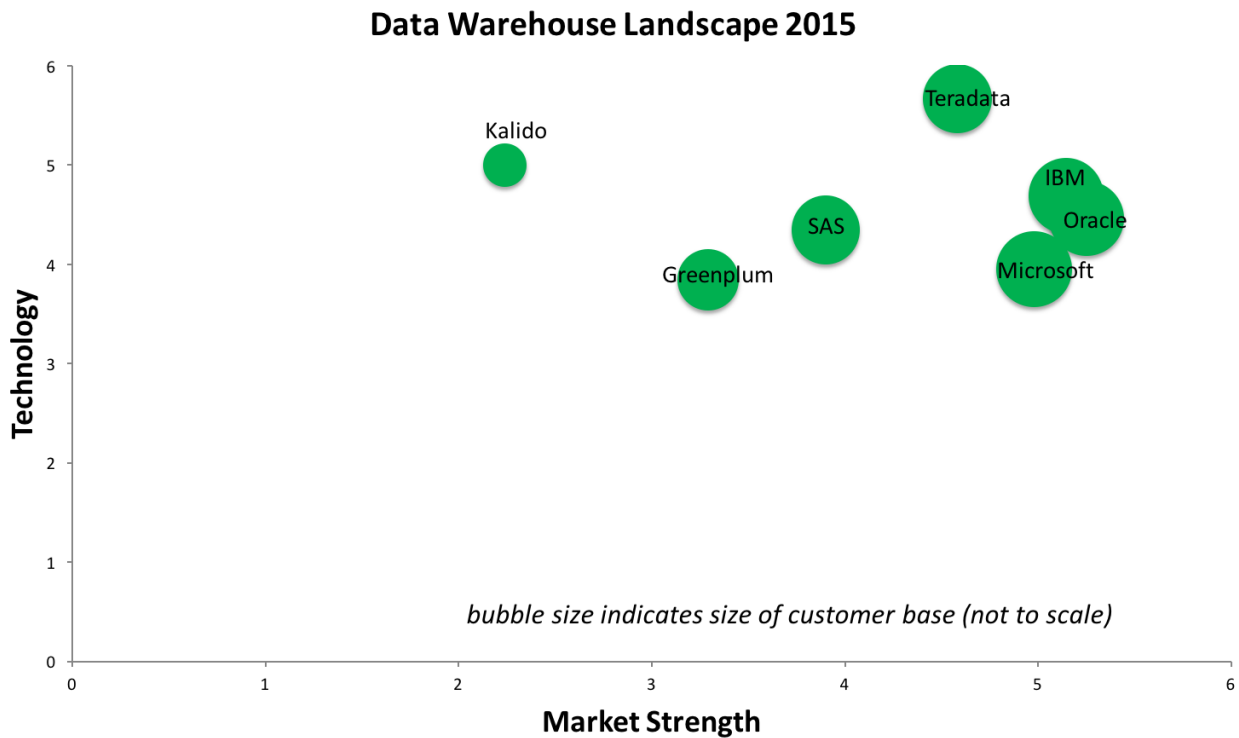
The landscape diagram represents the market in three dimensions. The size of the bubble represents the customer base of the vendor, i.e. the number of corporations it has sold data warehouse software to, adjusted for deal size. The larger the bubble, the broader the customer base, though this is not 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 profiles of vendors that go into more detail. Customers are 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 Kalido. Our congratulations to those vendors.

---

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

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, a subsidiary of EMC, itself acquired by Dell in 2015.	pivotal.io/big-data/pivotal-greenplum
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
jSonar	Boston-based NoSQL data warehouse vendor.	www.jsonar.com
Kognitio	Mature data warehouse appliance, offering its data warehouse as a service.	www.kognitio.com
Kalido	Now part of Magnitude Software, Kalido is an application to automate building and maintaining data warehouses that adapt to change, running on various database platforms.	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.	monetdb.cwi.nl
Neo4j	Open source graph database	www.neo4j.org
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 also offers 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 focused on analytics with its own data warehouse solutions including Teradata Database for Integrated Data Warehouse, Aster Analytics for advanced analytics on big data, and a configurable Hadoop Appliance for data lake.	www.teradata.com www.thinkbiganalytics.com www.claraview.com
Vertica	Appliance vendor Vertica was purchased by HP in 2011.	www.vertica.com
XtremeData	US vendor that provides highly scalable cloud database platform.	www.xtremedata.com
WhereScape	Not an appliance, but a framework for the development and support of data warehouses.	www.wherescape.com