

Big Data Warehouse Landscape Q4 2019

Databases originally focused on transaction processing in large enterprises for systems such as sales order processing. Once companies started to want to analyse information about their business performance, it became clear that information would have to be extracted from multiple transaction systems, which usually had incompatible structures and classifications of common data such as customer and product hierarchies. In addition to the complication of having to rationalise and aggregate data from many systems, the underlying databases themselves were unsuited to analytic processing. The relational databases were designed for high performance when there was a high degree of concurrent update from many users, and struggled with queries that stretched across large swathes of the database. Enterprises started to deploy separate databases for such purposes, even if the vendor was the same. As data volumes grew and the processing demands became more demanding, specialist databases started to appear that were optimised for handling analytic style processing, often at the expense of the ability to handle highly concurrent updates, which in the case of data warehouses was a main requirement.

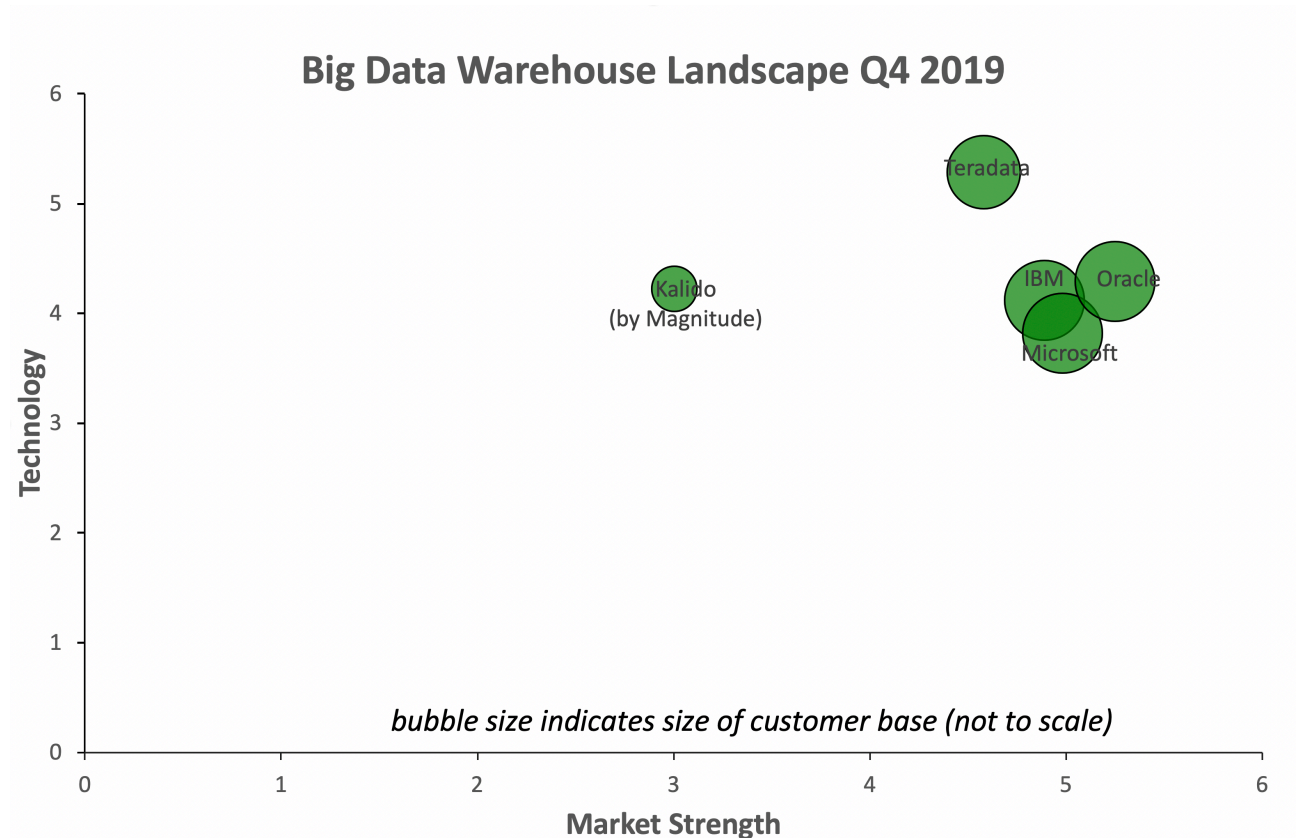
Data warehouses have grown greatly in size. In 2003 the largest data warehouse in the world was 30 TB in size, yet just a decade later there were examples of petabyte sized data warehouses, a 30-fold increase in ten years, a trend that has continued to this day. Database technologies developed to meet these challenges. Parallel processing allowed complex queries to be split across multiple processors, which meant splitting the problem up into smaller bundles, which itself required quite different database optimisation. Another approach was to invert the shape of the database itself. Traditional row-oriented relational databases started to give way to column-oriented databases, which increased query efficiency at the cost of load and update speeds. These columnar databases could often operate in parallel too, bringing more processing power to bear on complex queries across large databases. Some databases now offer a choice of row or columnar deployments, with better optimiser technology that has better understanding of the underlying storage basis. Modern database technologies try to shield the end customer from the mechanics of data warehouse operation as much as possible.

As well as volume, modern data warehouses have to cope with a wider range of data formats beyond just numbers and text. The rise of eCommerce has led to a variety of digital assets to be created, including images and video. Moreover, the proliferation of devices generating data, from sensors in cars and airplanes, through to energy meters, mobile devices and even wearable technology, has further increased the levels of stress on traditional data warehouses, which were not designed with such developments in mind.

In recent years there has been a migration of applications from on-premise to cloud, something that was pioneered in salesforce automation but which had become a much broader secular change. As ever with the huge existing deployments in large enterprises, changes take time. Nonetheless there is no doubt that more and more applications are migrating to deployment in cloud, whether that be the public cloud such as those of Amazon and Microsoft, or private clouds. Data warehouses have started to follow suit, and indeed one of the most interesting and fastest growing newer vendors has been Snowflake, who offer a pure cloud

deployment for data warehouse technology. Established vendors have responded with their own cloud deployment options.

The major 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 customers¹, 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, customer base (revenue adjusted) and geographic coverage. The Information Difference maintains vendor profiles 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 annual research cycle the vendors with the happiest customers were Teradata. Our congratulations to them.

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
Action	Action's product is an analytic database on commodity hardware.	www.action.com
Amazon Redshift	Cloud-based data warehouse solution.	aws.amazon.com/redshift/
Cloudera	Enterprise cloud vendor; now incorporates Hortonworks.	www.cloudera.com
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
HPCC	An open-source, massively parallel platform for big data processing, developed by LexisNexis Risk Solutions.	hpccsystems.com
IBM	DB2 is the data warehouse software offering from the industry giant, now available on cloud as well as on-premise.	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 (by Magnitude)	Kalido (by Magnitude) is an application to automate building and maintaining data warehouses.	magnitude.com/mdm
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
Pivotal	Owners of the Greenplum massively parallel data warehouse solution, now an open-source solution.	pivotal.io/big-data/pivotal-greenplum
Qubole	Markets the Qubole Data Service, which accelerates analytics workloads working on data stored in cloud databases.	www.qubole.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	www.sap.com

	giant SAP. SAP also offers the in-memory database technology HANA.	
SAS Institute	Comprehensive data warehouse technology from the largest privately-owned software company in the world.	www.sas.com
Snowflake	Cloud-only data warehouse vendor.	www.snowflake.com
1010 Data	Provides column-oriented database and web-based data analysis platform.	www.1010data.com
Teradata	Database giant with its own data warehouse solutions.	www.teradata.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