Big Data Warehouse Landscape Q4 2020

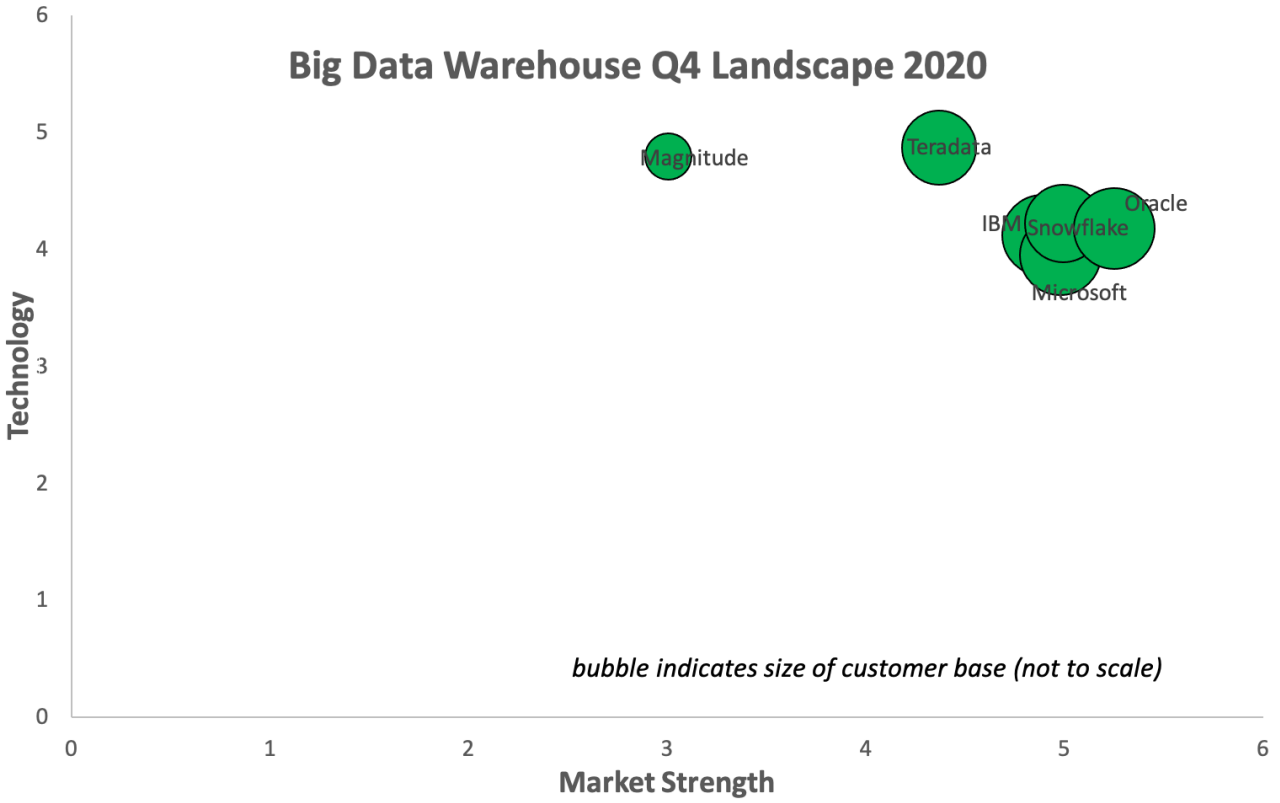
Ever since databases were developed in the 1960s, they have been pressed into service for different purposes. Initially they were used mostly to support the processing of business transactions, but in the 1980s the concept of a "data warehouse" evolved to allow a separate data store that would be dedicated to business analytics rather than transaction processing. The idea was to gather data from the various transaction systems that were in use around an enterprise in order to produce a single, reliable source of analytic data that could be used to monitor business performance. Data warehouses have changed greatly since those early days, in terms of the technologies on which they are based, and the stresses put upon them. As the volume of data that we store has grown then so data warehouses have grown greatly in size to reflect this. 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. This is a trend that has continued to this day, with for example the data warehouse of taxi company Uber weighing in at over 100 petabytes by 2018.

Data warehouses were traditionally row-oriented in their design and mostly relational (SQL based), but increasingly adapted to be columnar in structure, which has many advantages for analytic processing, albeit at the price of the speed of updating data. In recent years a range of different database constructs have emerged, with NoSQL ("not only SQL") databases including graph databases, document databases and more. Data is no longer confined to the enterprise, with organizations wanting to bring in data from suppliers and other third-party providers. The rise of "big data' file systems (Hadoop, Spark) adds a further level of complexity and size to data sources that a data warehouse design has to consider. Data warehouses today have to deal with traditional numeric data but also a wider range of data types and sources, such as text, images, video, time series data and sensor data. Architectures have adapted to spin off "data marts" from a corporate data warehouse, and more recently we see "data lakes" of big data sitting alongside, and potentially acting as feeds into, data warehouses.

The emergence of cloud computing created a new set of challenges and opportunities, with more and more data migrating out of the traditional corporate data centre. By 2019 around half of all corporate data was cloud-based, and this trend allowed the emergence of purely cloud-based data warehouses such as Snowflake and Amazon Redshift. Snowflake's IPO in September 2020 was the largest software company IPO in history, with the company's market cap in December 2020 being higher than IBM. This meteoric rise demonstrates that data warehousing, and the business analytics that depend on it, is far from the mature backwater than some commentators thought just a few years ago. Today it is a market that generates revenues of perhaps $20 billion with compound annual growth of 8-12%, depending on the exact definition of the market and which analyst firm you listen to. This momentum does not seem to have been impacted by the global coronavirus pandemic of 2020. Businesses still need to assess and understand their own performance even if their workforce is mostly working remotely.

Copyright © 2020 The Information Difference Ltd. All Rights Reserved.

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 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, 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, followed by Magnitude. Our congratulations to them.

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

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

Copyright © 2020 The Information Difference Ltd. All Rights Reserved.

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
Magnitude	Part of Magnitude Software, Kalido is an application to automate building and maintaining data warehouses.	magnitude.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
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 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
Snowflake	Cloud-only data warehouse vendor.	www.snowflake.com

Copyright © 2020 The Information Difference Ltd. All Rights Reserved.

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

Copyright © 2020 The Information Difference Ltd. All Rights Reserved.