BRYAN, GARNIER & CO

The New Big Data World

GETTING MORE VALUE FROM DATA:
THE RISE OF HIGH-PERFORMANCE DATABASES

TECHNOLOGY WHITE PAPER MARCH 2019

Contents

The database market has seen seismic change in the last decade. Trends including cloud computing, Big Data and IOT, AI and automation have contributed to an explosion in data volumes. This has shaken up the typical enterprise data management ecosystem and created fertile ground for a new generation of startups offering high-volume, high-performance applications. Many mainstream relational database players have reacted by acquiring the newcomers. Other big players such as SAP have championed in-memory databases.

Relational databases still dominate globally. But while the segment is expected to keep growing, driven by factors such as in-memory SQL databases, its global share is forecast to fall steadily. At the other end of the market, NoSQL and Hadoop databases surged by an estimated 66% per year from 2013-18 and are expected to account for 18% of the total database market by 2021. The excitement around new database technologies has been reflected in a wave of pre- IPO funding rounds dominated by Cloudera and Snowflake. However, cloud innovations and plummeting storage costs have driven consolidation in the Hadoop space. NoSQL databases, which offer simplicity and speed in big data applications, and NewSQL, which adds the ACID guarantees of relational databases, have seen significant activity and interest.

To examine some aspect of the evolving database market in more detail, we profile cloud-based data warehousing software vendor Snowflake, a ‘superfunded’ company offering data warehouse as-a-service; and we look at Exasol, which develops in-memory analytics for big data use cases.

In conclusion, we see a continuing tussle between Oracle and Amazon Web Services for the #1 spot in the database market, Hadoop being challenged by its own complexity, acquisitions in new database technologies as incumbents continue to battle AWS, and plenty of potential unicorns in the data management space.

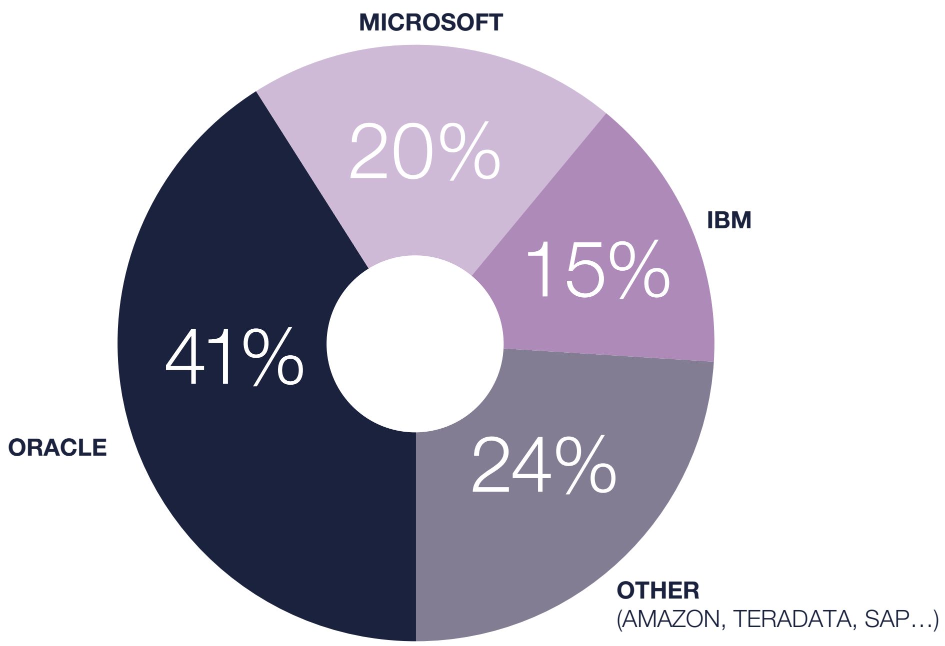
1. Tectonic shifts in the database market

Relational database vendors are challenged

Gartner estimated the size of the database management systems market at USD 38.8bn for 2017, up 12.7%[1], after an 8.1% rise in 2016 and a 1.6% decline in 2015. For more than two decades, this market has been dominated by three players: Oracle, which had 41% market share in 2016, IBM (20%), and Microsoft (15%). Relational databases based on Structured Query Language (SQL) have become the backbone of mainstream enterprise application software like ERP or CRM that flourished in the 1990s and 2000s. (SQL was first developed by IBM in 1974 and made commercially available by Oracle five years later). As they constantly improved in performance, driven by the data warehousing, data mining and business intelligence tools that appeared throughout the 1980s and 90s, relational databases were fast enough to address all of an organisation's data processing needs.

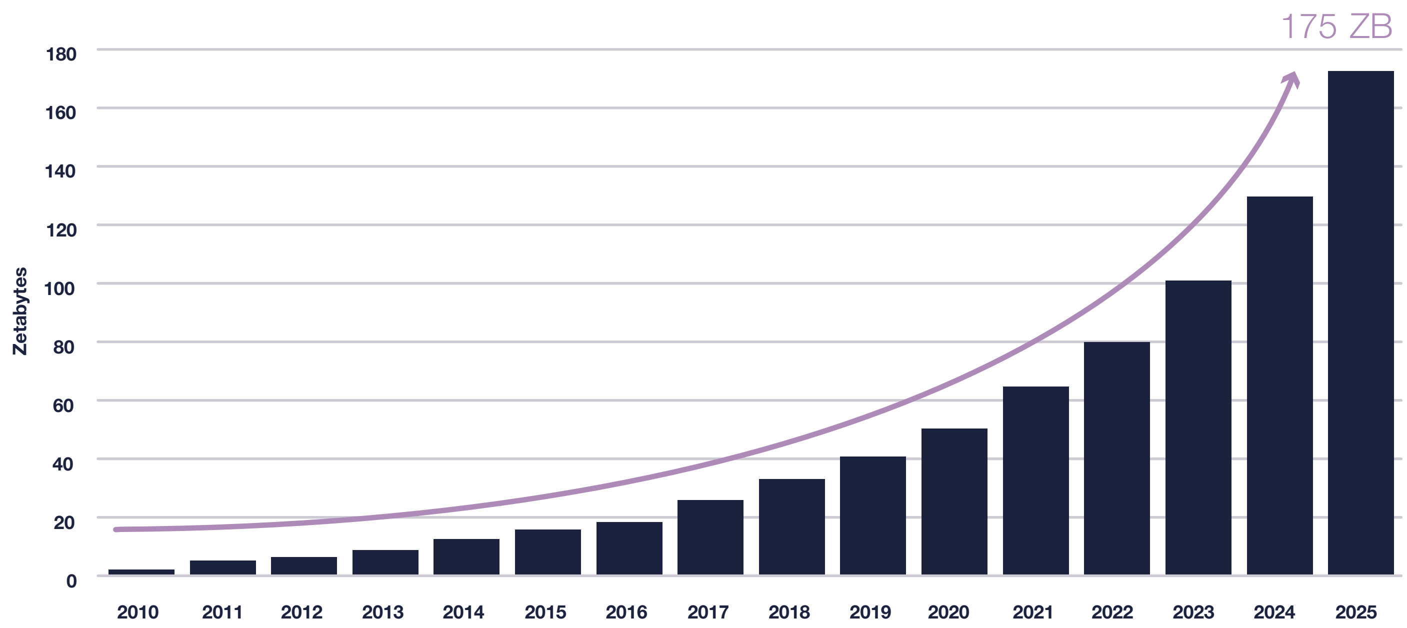
Over the last decade, the database market has undergone massive changes, thanks to the nature of data being generated, the volume, and the processing capability needed to make sense of all the data. According to an IDC white paper published in November 2018[2], the total volume of global data has exploded, rocketing to 33 zettabytes (ZB) in 2018 from less than 5 ZB in 2005. It is projected to reach 163 ZB in 2025. One zettabyte is equivalent to a trillion gigabytes (GB).

FIG. 1: DATABASE MANAGEMENT SYSTEMS MARKET SHARES (2016)



[1]Gartner, State of the Operational DBMS market, 2018, 6th September 2018.
[2]IDC and Seagate, The Digitisation of the World, November 2018.

FIG. 2: ANNUAL SIZE OF THE 'GLOBAL DATASPHERE' IN ZETTABYTES (2010-2025E)



Source: IDC and Seagate, The Digitisation of the World, November 2018

IT departments of large and small companies alike are scrambling to bridge this data infrastructure gap as they failed to anticipate this data explosion, which was primarily caused by six converging trends:

Cloud computing

The public cloud services Amazon Web Services (2002), Google Cloud Platform (2008) and Microsoft Azure (2010) have been designed to facilitate agile projects and provide flexible, demand-oriented infrastructure and services. Using inexpensive technologies such as software-defined storage (SDN), they made data storage and the general use of data analytics accessible to organisations of all sizes and across all departments. These three players developed their own data management systems: 1) Amazon launched Amazon RDS (relational database), Amazon DynamoDB (NoSQL database), Amazon Aurora (relational database) and Amazon Redshift (data warehouse); 2) Google launched Google Cloud SQL (SQL database), Google Cloud BigTable (wide column store), Google Cloud Spanner (NewSQL database), Google Cloud Datastore (NoSQL database) and Google BigQuery (data warehouse); 3) Microsoft launched Microsoft Azure SQL Database and Microsoft Azure Data Warehouse.

Big Data

As we have already seen, data volumes have increased massively. Social media platforms create plenty of it: the overall data footprint of eBay, Facebook and Twitter now is 500, 300 and 300 petabytes, respectively (a petabyte is equivalent to one million GB). And the Internet of Things (IoT) – data from phones, cameras, connected cars, wearables, sensors, operational control centers – generates vast volumes of data and needs new technologies for analysis, more scalable storage, streaming solutions and operational data systems. New applications also need to be created on top of data, for example to offer dynamic pricing or personalised marketing campaigns. The complexity of IoT, where data is processed either on servers, public or private clouds, or at the endpoint or the edge of devices such as cameras or vending machines, has created new challenges for data storage and analysis.

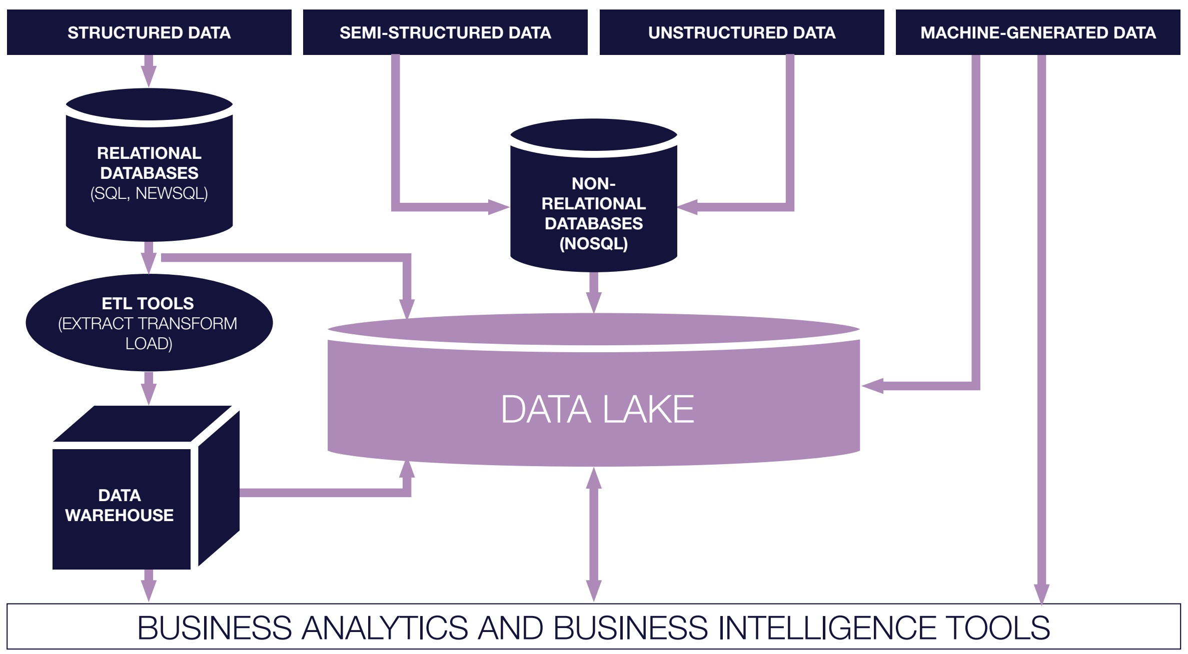
Self-service data analytics

By design, access to legacy data warehouses was restricted to centralised IT departments. The decentralisation of data and the advent of more easy-to-use analysis tools like Tableau, Qlik, or Microsoft Power BI enabled staff to access and analyse data and get more insights from it. User requirements have also changed dramatically. Dashboards can be updated within seconds and include the most recent data, meaning that the patience of users waiting for reports, analyses and results is increasingly short.

Operational business intelligence and automation

Data analytics no longer simply measures historical data. It is being built into all kinds of operational processes and is predictive, deterministic and self-learning. Completely automated and self-learning processes make decisions based on data at a speed and accuracy that humans cannot match.

FIG. 3: TYPICAL MODERN DATA MANAGEMENT ECOSYSTEM



External data

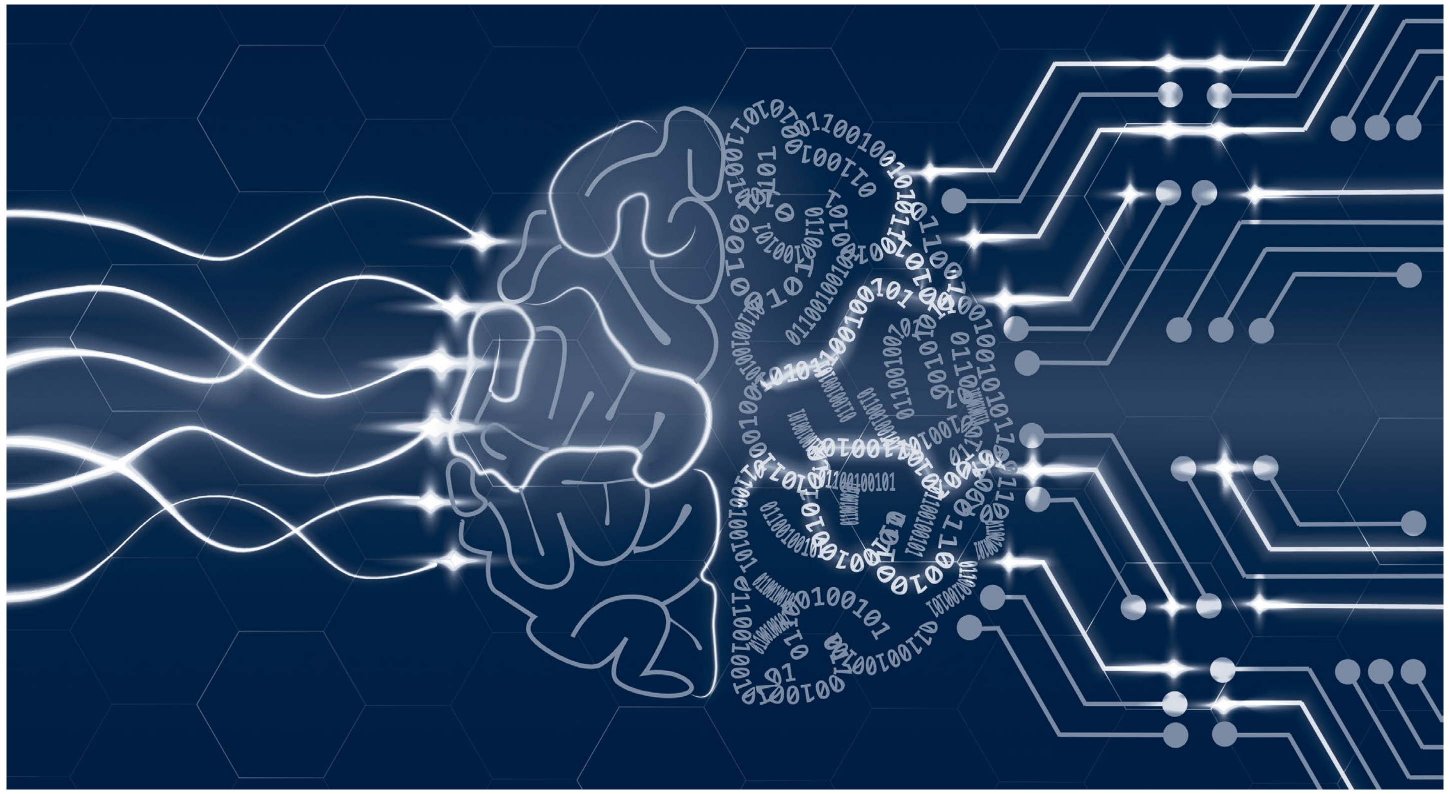
Data is no longer created only within the organisation itself. It can now be easily enriched with all manner of external sources, for example: social media, public statistics, geospatial data, weather forecasts, competition analysis and event-driven data.

Artificial intelligence

With recent advances in infrastructure and tools, data science is much more accessible and now allows for automated decision-making. Applications such as face recognition, self-driving cars and predictive analytics based on self-learning algorithms are radically changing business processes.

These trends have shaken up the typical enterprise data management ecosystem. It is no longer limited to structured data, relational databases, ETL (Extract Transform Load) tools, enterprise data warehouses, and traditional business intelligence tools. In particular, structured data, semi-structured data, unstructured data and machine-based data (IoT, logs, sensors) can be now stored in a 'data lake', which can then be connected to enterprise data warehouses.

Another consequence of these changes is the advent of Amazon Web Services (AWS) as the fourth largest database vendor in the market (with a Cloud database offering). In 2017, AWS had an estimated 9%+ market share of the whole database market and revenues up an impressive 112% in this area.



Mainstream players have reacted with acquisitions

Relational database management software leaders have had to embrace these trends, but at the same time deal with hundreds of thousands of customers who need backwards compatibility with legacy IT environments. This constraint made it practically impossible to replace 30-40-year-old code and rebuild their solutions from scratch. So instead of migrating their database product offerings to the new technology platforms, these large incumbents kept their old platforms and tried to fix their legacy systems by adding elements of the new technologies to their existing products. The result has been a mix of approaches and a hybrid architecture. Alternative solutions emerged, such as column-based data stores (like Sybase IQ in 1995) and in-memory databases, but these were confined to high-end applications.

In the mid-2000s, start-ups including Vertica, Greenplum, Netezza, DatAllegro, Aster Data, ParAccel and Exasol acknowledged that greater performance was going to be needed to deal with increasing data volumes. They developed solutions in the form of large-volume, high-performance data warehousing applications designed and built for analytical use cases. Their products started to see success and most of them were acquired by global IT players (see Fig. 4). ParAccel, which has served as the core component of Amazon Redshift, was acquired by Actian (formerly Ingres) in 2013. These acquisitions enabled global vendors to temporarily address performance issues and consolidate their market positions.

These newly acquired technologies contained their own weaknesses. For example, they are all built on a kernel of the PostgreSQL database (a pre-existing database based on Ingres) which was built in the 1970s, revamped in the 1990s, and difficult to change. The legacy players attempted to lock in the customers by offering 'end-to-end' business solutions (from databases to BI solutions) of which the relational database was only one component. For example, for use in corporate data centers deployed as private clouds, Oracle launched in 2008 its Exadata Database Machine, which is an Oracle database running on dedicated hardware. Exadata was adapted to public clouds in 2015.

FIG. 4: ACQUISITIONS IN DATABASES AND DATA WAREHOUSING

Company	Funding	Price/acquirer
DatAllegro	USD 63m	USD 200m by Microsoft (2008)
Netezza	USD 63m	USD 1,700m by IBM (2010)
Greenplum	USD 97m	USD 300m by EMC (2010)
Vertica Systems	USD 31m	USD 350m by HP (2011)
Aster Data	USD 53m	USD 295m by Teradata (2011)

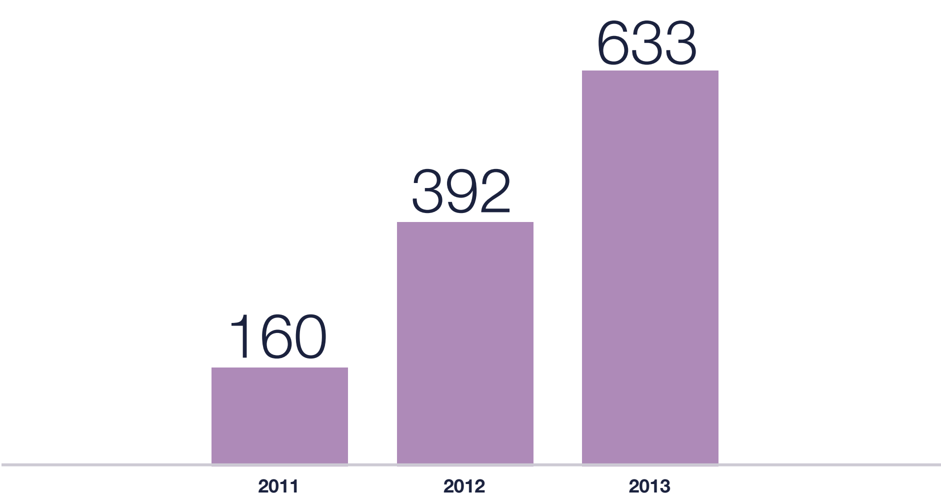
Source: Company Data; Bryan, Garnier & Co ests.

SAP pushes in-memory databases to primetime

Before it launched HANA, SAP was not a database player, even though in 1997 it had acquired Max DB, a relational database with an in-memory engine that was developed in the late 1970s at the Technical University of Berlin. For years, SAP relied on IBM, Oracle and Microsoft to provide the database that ran its enterprise application software products. In 2010, the acquisition of Sybase for USD 5.8bn gave SAP the Sybase ASE relational database (which served as the basis for Microsoft SQL Server) and the Sybase IQ column store.

The launch in November 2010 of the SAP HANA in-memory database, sold as an application running on certified hardware, was a revolution in the database world. SAP now mandated its own in-memory database for running its business applications, at the expense of third-party relational database providers. The initial purpose of SAP HANA was to carry out real-time data analytics and from 2011, SAP used it to power SAP Business Warehouse, its data warehousing system. But SAP HANA was quickly used to manage data transactions, powering SAP Business Suite (2013) then its replacement SAP S/4HANA (2015) and all SAP cloud applications. Today, SAP HANA is the mandatory database for running SAP's Core and Cloud applications.

FIG. 5: SAP HANA REVENUES (2011-2013) (EUR M)



Source: SAP. From 2014 onwards, SAP stopped reporting SAP HANA revenues

The HANA technology had several advantages from the start: 1) speed – the ability to handle several billion records in seconds; 2) scalability - designed to positively respond to the increase in the number of cores per server; 3) in-memory data compression between 5 and 20x; and 4) data management simplification – with no need to create data aggregates, there is no duplication.

In addition, HANA processed in-memory data in both lines and columns, and could handle data management standards for relational databases and OLAP (OnLine Analytical Processing) multi-dimensional analysis. Finally, with its ability to handle transactional and analytical data, SAP HANA was designed eliminate the use of ETL (Extract Transform Load) tools, stored procedures, materialised views and OLAP cubes.

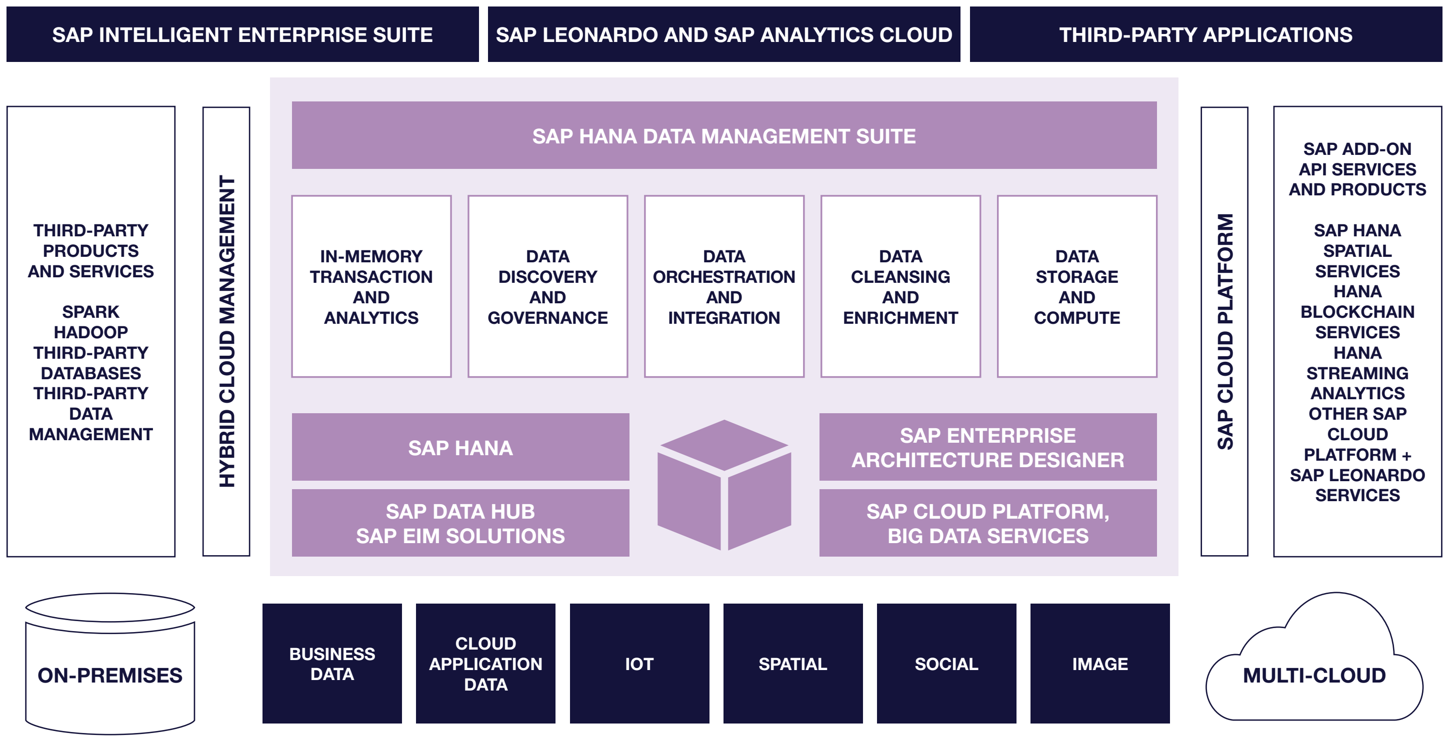
SAP HANA was made possible by the arrival of standard servers at affordable prices, equipped with several terabytes of memory and multi-core processors. With blade servers then sold for EUR 40,000 a unit (now down to EUR 15,000), massively parallel computing became widespread. In-memory data processing technology has therefore long been reduced to the notion of a cache or database accelerator. While all operations take place in-memory, each transaction is automatically captured and copied onto a disk with redundancy and high availability.

In response to the success of SAP HANA, which grew from nothing in 2010 to EUR 633m in revenues by 2013 and now reaches over 21,000 customers, relational database incumbents moved to embrace in-memory technology and protect their installed base. During 2013 and 2014, IBM, Oracle and Microsoft all launched in-memory options on top of their relational databases, rather than releasing an independent in-memory database.

Over the years, SAP HANA has evolved to become an open and governed data platform, with several product extensions and innovations for orchestrating and managing enterprise business data within multi-cloud and hybrid environments. To integrate Big Data stored in Hadoop with enterprise data in SAP HANA, SAP offers a separate in-memory query engine, SAP Vora, which was launched in 2015. And with the acquisition in 2016 of Big Data as-a-service (BDaaS) provider Altiscale, SAP was able to create its SAP Cloud Platform Big Data Services offering based on Hadoop and Spark. This acts as a data refinery, with the ability to cleanse and transform terabytes of raw data – including semi-structured and unstructured external data like clicks, logs, IoT, text, images, and video – before it is surfaced through SAP HANA for further, more detailed analysis.

As of today, the key components of SAP HANA Data Management Suite include SAP Data Hub - which creates a central metadata management catalog; SAP HANA; SAP Enterprise Architecture Designer, which creates architecture documentation; and SAP Cloud Platform Big Data Services.

FIG. 6: ARCHITECTURE OF THE SAP HANA DATA MANAGEMENT SUITE



Source: SAP

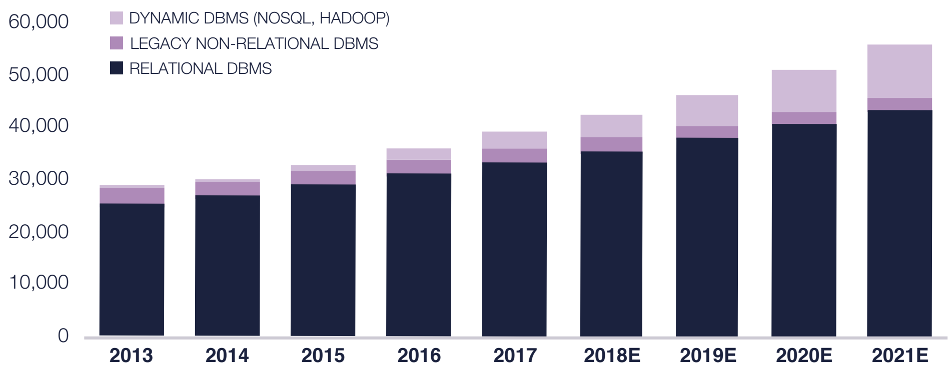
2. New data management technologies

From SQL to NoSQL to NewSQL

Relational databases still account for around 82% of the global market for database management systems. However, this has fallen from 87% in 2013, and IDC anticipates that it will fall further, to 77% by 2021. That said, the relational database segment is still expected to grow by 6-7% per year over the next three years, driven in particular by in-memory SQL databases or in-memory options and the strong growth of NewSQL databases. NoSQL, NewSQL will not replace SQL databases, but co-exist within technology stacks for the foreseeable future. At the other end of the market, NoSQL and Hadoop databases have increased their market share from 1% in 2013 to an estimated 11% in 2018, with revenues up from USD 366m to USD 4.6bn. These databases have surged by an estimated 66% per year on average over the period. IDC anticipates that this segment will see average annual growth of 33% from 2017-2021, reaching USD 10bn or 18% of the total database market by 2021.

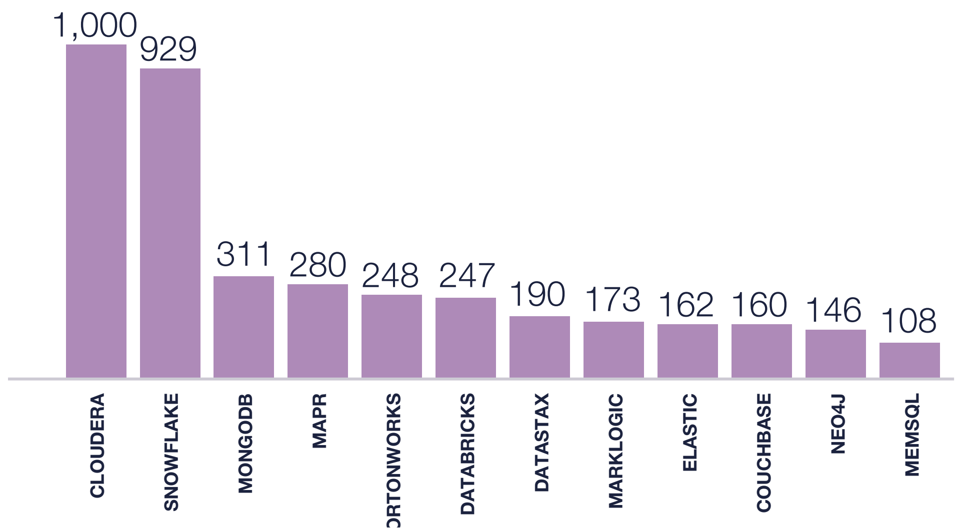
The excitement around NoSQL, NewSQL and Hadoop, which are aimed at replacing relational databases, is reflected by the funding of a significant number of startups in these areas (see Fig. 8).

FIG. 7: DATABASE MANAGEMENT SYSTEMS MARKET (2013-2021E)



Source: IDC

FIG. 8: PRE-IPO FUNDING OF NO/NEWSQL, HADOOP AND CLOUD EDW VENDORS (USD M)



Source: Crunchbase

The table below recaps the financing rounds that have taken place since early 2016 in the database and data warehousing areas. 2018 saw plenty of activity, with USD 713m raised for Snowflake Computing, which is seen as the new 'unicorn' in the space, over two rounds. The other significant rounds for this year include Yellowbrick Data (USD 92m), Neo4j (USD 80m), NuoDB (USD 31m) and MemSQL (USD 30m). In 2017, the two biggest rounds were Databricks (USD 140m), and Snowflake again (USD 105m).

Valuation multiples achieved by Hortonworks, Cloudera, MongoDB and Elastic – the four companies in this space that have gone public since 2014 – show strong investor appetite for these stocks as they generate strong growth, reduce their losses, and become potential champions in their specific area or M&A targets for the likes of IBM, Oracle, SAP and others.

FIG. 9: FINANCING ROUNDS FOR NO/NEW SQL, HADOOP AND CLOUD EDW VENDORS

Date	Company	Round	No. investors	Amount	Lead financing
Dec. 2018	NuoDB	Venture	5	USD 31m	Temenos
Nov. 2018	Neo4j	Series E	5	USD 80m	Morgan Stanley, One Peak
Oct. 2018	Yellowbrick Data	Series B	2	USD 48m	Next47
Oct. 2018	Snowflake Computing	Series F	8	USD 450m	Sequoia
Oct. 2018	Elastic	IPO		USD 252m	Valuation: USD 2.5bn
Aug. 2018	Yellowbrick Data	Series A	5	USD 44m	Menlo Ventures
May 2018	MemSQL	Series D	5	USD 30m	Glynn, GV
Jan. 2018	Snowflake Computing	Series E	3	USD 263m	Altimeter, Iconiq, Sequoia
Oct. 2017	MongoDB	IPO		USD 256m	Valuation: USD 1.6bn
Sep. 2017	MariaDB	Series C	6	USD 27m	Alibaba
Sep. 2017	Snowflake Computing	Series D	7	USD 105m	Iconiq
Sep. 2017	MapR Technologies	Venture	7	USD 56m	Lightspeed
Aug. 2017	Databricks	Series D	4	USD 140m	Andreessen Horowitz
May 2017	MariaDB	Venture	1	EUR 25m	European Investment Bank
May 2017	Cockroach Labs	Series B	5	USD 27m	Redpoint
Apr. 2017	Cloudera	IPO		USD 225m	Valuation: USD 1.9bn
Dec. 2016	Databricks	Series C	3	USD 60m	New Enterprise Associates
Nov. 2016	Neo4j	Series D	4	USD 36m	
Aug. 2016	MapR Technologies	Venture	7	USD 50m	Future Fund
Jul. 2016	Elastic	Series D	2	USD 58m	
Apr. 2016	MemSQL	Series C	7	USD 36m	Caffeinated, REV
Mar. 2016	Cockroach Labs	Series A	4	USD 20m	Index
Mar. 2016	Couchbase	Series F	7	USD 30m	Sorenson
Feb. 2016	NuoDB	Series B	3	USD 17m	

Source: Crunchbase

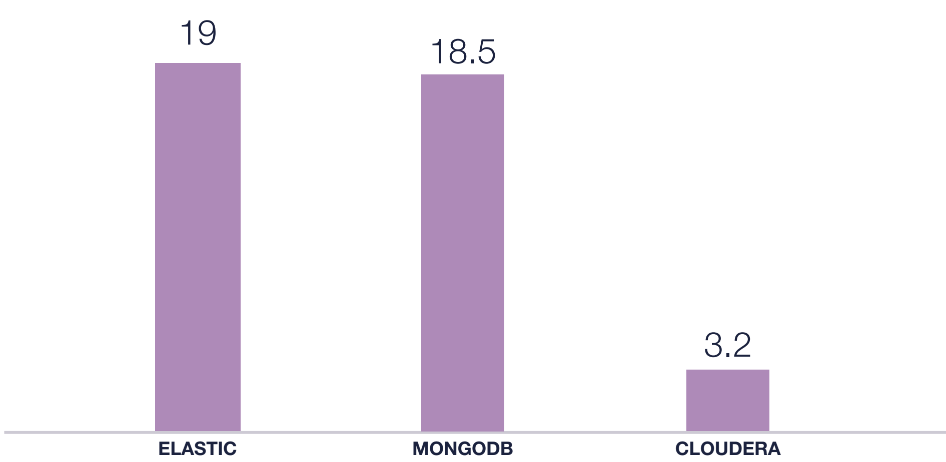
FIG. 10: MARKET CAPITALISATION OF NEW DATA MANAGEMENT TECH VENDORS

Elastic	USD 5.19bn	+107% since IPO
Mongo DB	USD 4.59bn	+181% since IPO
Cloudera	USD 1.72bn	-23% since IPO
Hortonworks	USD 1.24bn*	+88% since IPO
Vertica Systems	USD 31m	
Aster Data	USD 53m	

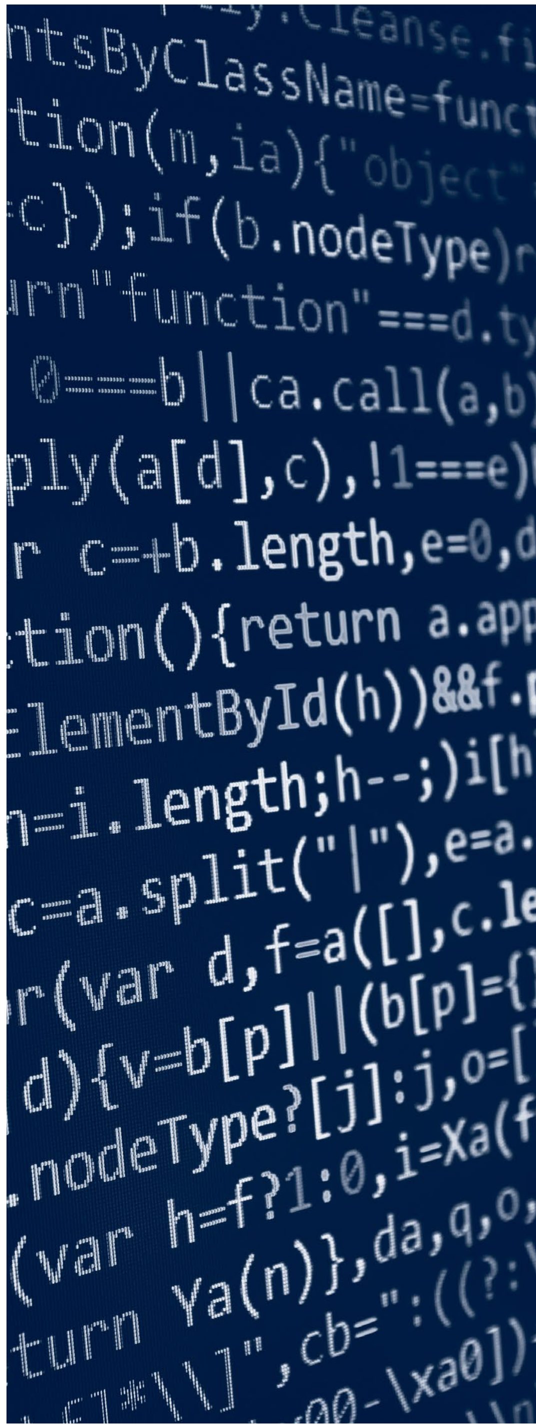
*As of 2 January 2019, when it left the stock market following merger with Cloudera.

Current EV/(N+1) sales multiples are impressive for Elastic and MongoDB:

FIG. 11: EV/(N+1) SALES MULTIPLES AS OF 8TH JANUARY 2019



Source: Company data; Thomson Reuters



Hadoop for big data: the rise and fall?

The initial release of Apache Hadoop took place in 2006. This distributed file system is a collection of open-source software tools that facilitate the use of computer clusters to solve issues involving massive amounts of data. Its core has two elements: storage, Hadoop Distributed File System or HDFS; and processing, the MapReduce programming model. Hadoop splits files into large blocks and distributes them across nodes in a cluster. It then transfers packaged code into nodes to process the data in parallel. This approach allows the dataset to be processed faster and more efficiently than in a conventional supercomputer architecture, which uses a parallel file system where computation and data are distributed via high-speed networking.

FIG. 12: THE HADOOP DATA MANAGEMENT ECOSYSTEM

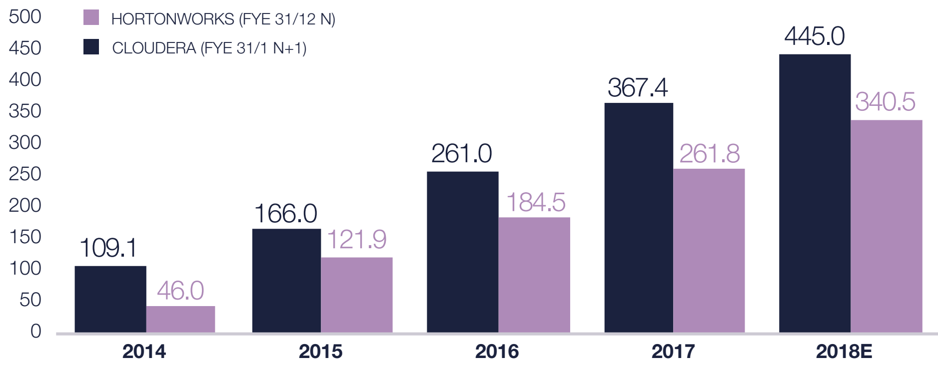
Spark	Cluster-computing framework for data analytics and machine-learning algorithms
Hive	Data warehousing software providing a SQL-like query language, HiveQL
HBase	Non-relational distributed database for storing large quantities of sparse data, for serving data-driven websites
Phoenix	Massively parallel relational database based on SQL, using HBase as its data store
Impala	Massively parallel SQL query engine for analytics
Pig	Platform for creating programmes on Hadoop
ZooKeeper	Distributed computing system
Flume	Log data collection
Sqoop	Data transfer between relational databases and Hadoop
Oozie	Workflow scheduling
Storm	Distributed stream processing computation framework

An ecosystem of additional software packages for managing data has been developed around Hadoop (See Fig.12).

Hadoop can be deployed on-premise, or in a public or private cloud: Microsoft Azure (with Microsoft Azure HDInsight), Amazon EC2 and S3 (with Amazon Elastic MapReduce), Google Cloud Platform (Google Cloud Dataproc), SAP Cloud Platform (SAP Big Data Services), and Oracle Cloud Platform (Oracle Big Data Cloud).

The main Hadoop software distributions are Cloudera, Hortonworks and MapR. Hortonworks and Cloudera have been listed in the US since December 2014 and April 2017, respectively. Their revenues surged at over +40% per year until 2017, after which growth rates slowed in 2018, when the mid-point of company guidance implied +21% for Cloudera and +30% for Hortonworks. Cloudera's shares dropped 40% on 4th April 2018 when it reported its FY18 results due to disappointing guidance for FY19.

FIG. 13: REVENUES OF CLOUDERA AND HORTONWORKS (2014-2018E) (USD M)



Source: Cloudera; Hortonworks

In October 2018, Cloudera announced it was acquiring Hortonworks through an all-stock 'merger of equals', leaving Cloudera shareholders with 60% of the combined company. The transaction was completed in January 2019.

We understand the deal as a signal the Hadoop market could no longer sustain two big competitors, with several trends driving the change:

The shift to public cloud

AWS, Microsoft Azure, and Google Cloud offer their own managed Hadoop/Spark services as part of fully integrated offerings that are cheaper to buy and scale.

Plummeting storage costs

On cloud storage services like AWS S3, Azure Blob Storage, and Google Cloud Storage, a terabyte of cloud object storage costs about USD 20 a month, compared to around USD 100 a month on Hadoop Distributed File System (HDFS).

Faster, better, and cheaper cloud databases

The advent of cloud services like Google BigQuery in 2011 eliminated the need to run Hadoop or Spark. While Apache Spark is used to handle ad-hoc distributed SQL queries, Google BigQuery runs ad-hoc queries on any amount of data stored in its object storage service without having to load it into special storage. Other cloud database services - Snowflake, Google BigTable, Amazon Aurora, and Microsoft Cosmos - are similarly massive-scale, highly flexible, globally distributed 'pay-as-you-go' databases.

Python and R data science running on containers

Hadoop's revolutionary innovation was that it offered both a storage and compute environment for applications written in Java. However, this was out of step with data scientists working on machine learning in Python and R, who need native deployment of Python/R models to iterate and improve machine learning models.

NoSQL databases: simpler than SQL

The concept of 'NoSQL' databases emerged around 2009. The term was introduced by IT developer Johan Oskarsson to label a growing number of non-relational, distributed data stores including Google BigTable, Amazon Dynamo and the newly launched Amazon DocumentDB. Most early NoSQL systems did not attempt to provide the atomicity, consistency, isolation and durability (ACID) guarantees that relational databases usually offer. NoSQL databases enable the storage and retrieval of data modelled in other means than through tabular relations. They are increasingly used in big data and real-time web applications due to design simplicity, simpler scaling to clusters of machines, and finer control over availability. The data structures used by NoSQL databases (key-value, wide column, graph, or document) make some operations faster in NoSQL than in relational databases.

There are four main categories of NoSQL database:

Wide column stores

Data is collected in very large structures for high-performance analytics. By storing data in columns rather than rows, the database an access the data it needs to answer a query more precisely, without the need to scan and discard unwanted data in rows. Wide column stores are well-suited for data warehouses involving highly complex queries.

Document-oriented databases

These are used for managing semi-structured data in document formats such as JSON or XML. Document databases store all the information for a given object in a single instance in the database, and every stored object can be different from every other. This makes mapping objects into the database a simple task and typically eliminates anything similar to an object-relational mapping. This makes document stores attractive for programming web applications, which are subject to continual changes, and where speed of deployment is an important issue.

Key-value (KV) stores

These databases use dynamic data such as session data and machine-generated data used for analysis and systems and device coordination. They treat the data as a single collection that may have different fields for every record. Key-value databases often use far less memory to store the same data than relational databases, which can lead to large performance gains in certain workloads.

Graph databases

These are organised in a graph structure for relationship and pattern analysis, for example fraud detection or semantic text analysis. The graph directly relates data items in the store to a collection of nodes of data and edges representing the relationships between the nodes. The relationships allow data in the store to be linked together directly. Querying relationships within a graph database is fast because they are perpetually stored within the database itself. Relationships can be intuitively visualised using graph databases, making them useful for heavily inter-connected data.

FIG. 14: CLASSIFICATION OF NOSQL DATABASES

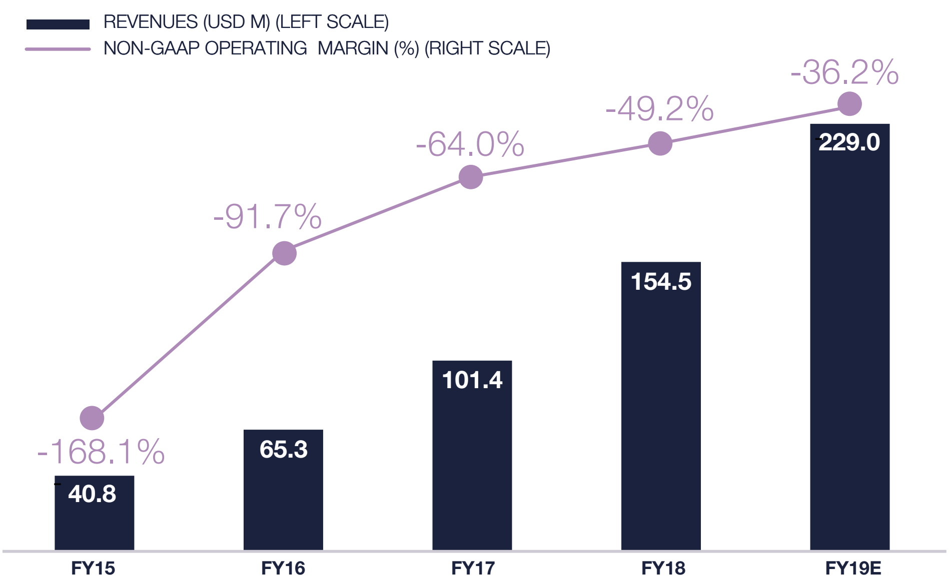
Wide column stores	Document-oriented databases	Key-value stores	Graph databases
Apache Accumulo (based on Google BigTable)	Amazon DocumentDB	Aerospike Database	AllegroGraph
Apache Cassandra (DataStax)	Apache CouchDB	Amazon DynamoDB	Apache Giraph
Apache HBase	ArangoDB	Apache Ignite	ArangoDB
Druid	BaseX	Apache ZooKeeper	Objectivity InfiniteGraph
Micro Focus Vertica	Clusterpoint	ArangoDB	MarkLogic
	Couchbase Server	Basho Riak	Neo4J
	Elasticsearch	Couchbase Server	OpenLink Virtuoso
	Microsoft Azure Cosmos DB	FoundationDB	SAP OrientDB
	IBM Notes/Domino	InfinityDB	
	MarkLogic	Oracle BerkeleyDB	
	MongoDB	SAP OrientDB	
	Qualcomm Qizx	SciDB	
	RethinkDB		
	SAP OrientDB		

Source: Bryan, Garnier & Co

MongoDB was the first listed pure-play NoSQL database software vendor, with a NASDAQ IPO in October 2017. The company is still heavily loss-making, with a non-GAAP operating margin forecasted at -36% at the mid-point of company guidance for FY19. Losses are reducing year-on-year while revenues are growing around 50% annually.

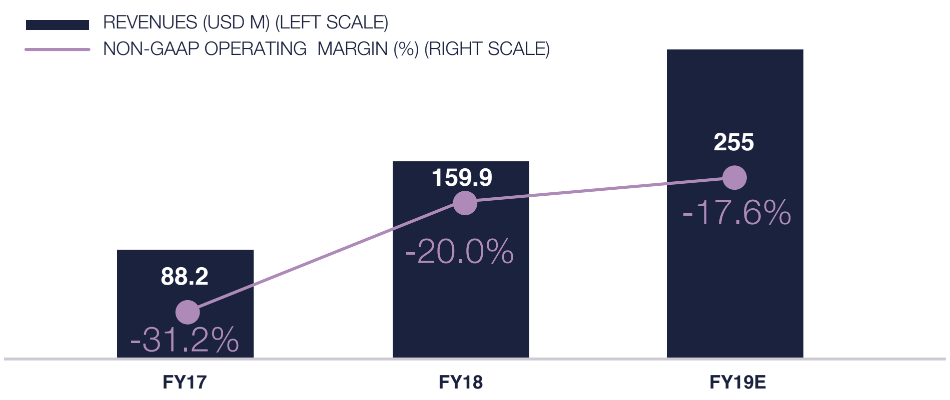
The second NoSQL database pure-player to be listed was Elastic, which develops and sells the Elasticsearch document-oriented database and search and analytics engine. Listed on NYSE in October 2018, this company is also loss-making, with a non-GAAP operating margin of -20% for the fiscal year ending 30th April 2018. Revenues, on the other hand, grew 81% during that fiscal year.

FIG. 15: MONGODB - REVENUES AND NON-GAAP OPERATING MARGIN (FYE 31/1)

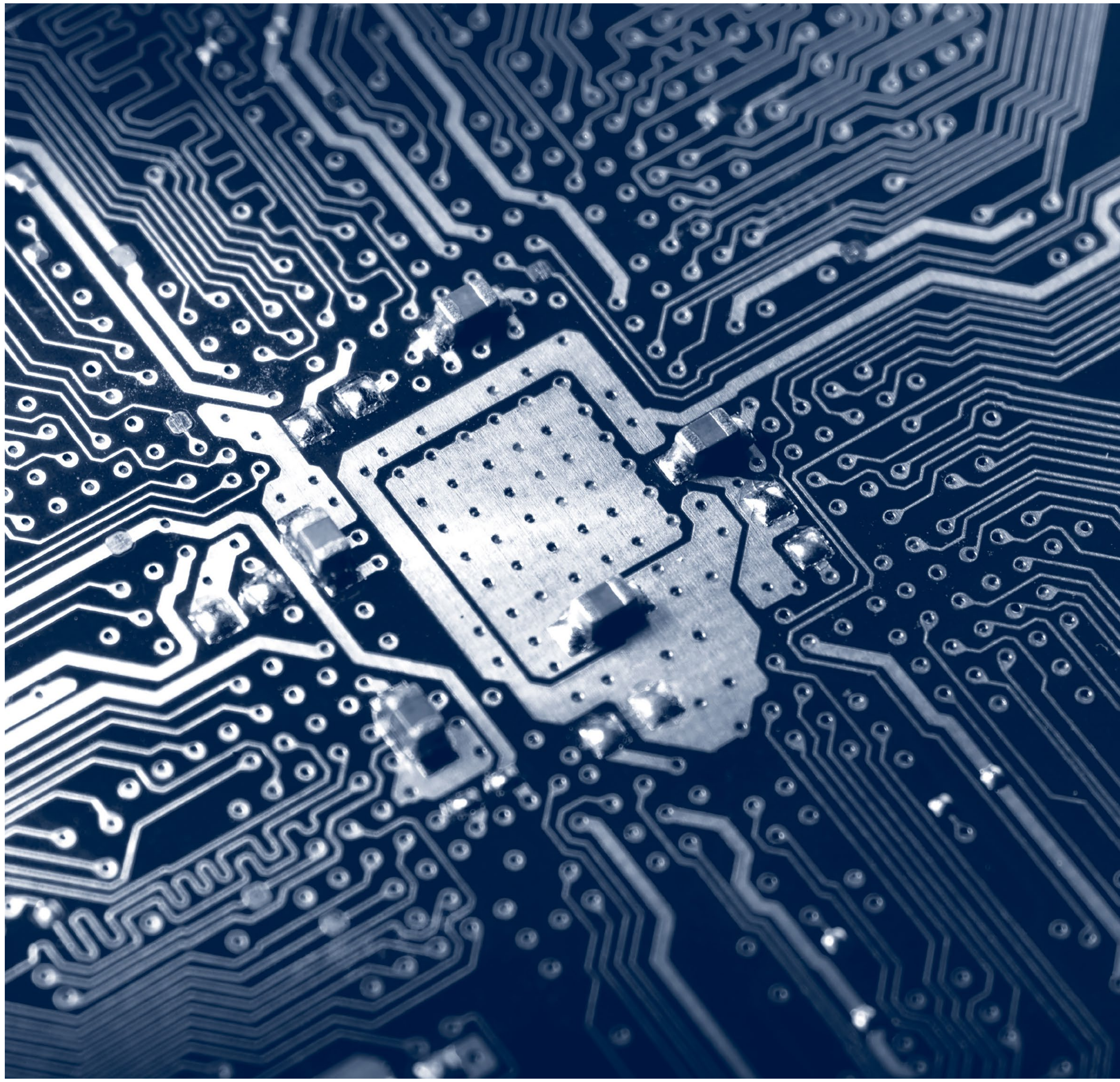


Source: MongoDB

FIG. 16: ELASTIC - REVENUES AND NON-GAAP OPERATING MARGIN (FYE 30/4)



Source: Elastic; Bryan, Garnier & Co



NewSQL databases

One of the reasons why relational databases have been so successful for decades is that they have a set of characteristics that guarantee their validity: Atomicity, Consistency, Isolation, and Durability (ACID). Atomicity guarantees each transaction is treated as a single unit that either succeeds completely or fails completely. Consistency means that any data written to the database must be valid according to all defined rules – this prevents database corruption by an illegal transaction, but does not guarantee a transaction is correct. Isolation ensures that concurrently executed transactions leave the database in the same state which would have been obtained if the transactions were executed sequentially. Finally, durability guarantees that once a transaction has been committed, it will remain committed even in the case of a power outage or crash.

The issue with NoSQL databases is their lack of compliance with the ACID rules when processing 'high-profile' data (e.g. financial data or orders) with strong transactional and consistency requirements. NewSQL – a concept which emerged in 2011 from the 451 Group analyst Matthew Aslett – seeks to offer the same scalable performance of NoSQL databases for transaction workloads while still maintaining the ACID guarantees of a relational database. Both NoSQL and NewSQL support the relational data model and use SQL as their primary interface. Applications typically targeted by NewSQL databases have a large number of short-lived, repetitive transactions that touch a small subset of data. Unlike NoSQL databases, NewSQL databases are not easy to categorise by technical approach as they vary significantly in their internal architecture. Some support only transactional workloads, but others support hybrid transactional/analytical workloads, which some NoSQL databases support too.

There are three main categories of NewSQL database:

Shared-nothing (SN) databases

These are designed to operate in a distributed cluster of 'shared-nothing' nodes in which each node is independent and self-sufficient, and owns a subset of the data, while none of the nodes shares memory or disk storage. Some SN database platforms are in-memory, like Altibase, Apache Ignite, Exasol, GridGain, MariaDB Clustrix, and VoltDB.

Storage engines for SQL

Optimised storage engines for SQL, these databases providing the same programming interface as SQL but scale better than built-in engines.

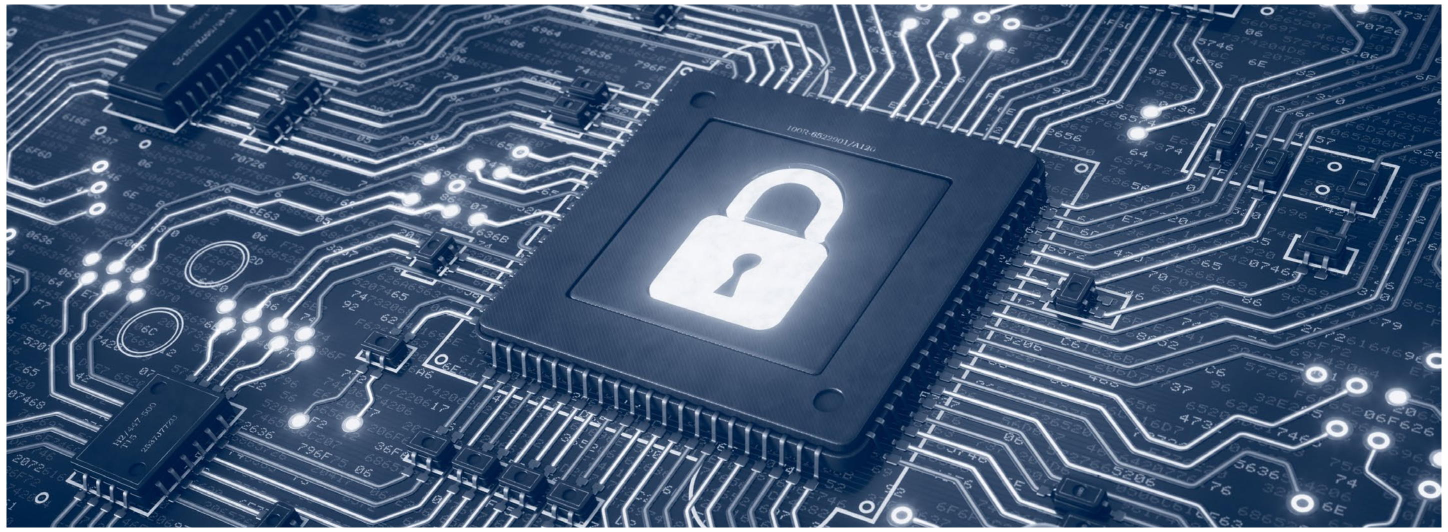
Transparent sharding

These NewSQL databases provide a horizontal data micro-partition layer to automatically split databases across multiple nodes.

FIG. 17: CLASSIFICATION OF NEWSQL DATABASES

Shared-noting databases	Storage engines for SQL	Transparent sharding
Altibase	Infobright	ScaleBase
Apache Ignite	MyRocks	Vitess
CockroachDB	MariaDB Columnstore	
Exasol	Microsoft SQL Server (with ColumnStore and InMemory features)	
Google Spanner	Oracle InnoDB	
GridGain	Oracle MySQL Cluster	
HarperDB	TokuDB	
MariaDB Clustrix		
MemSQL		
Microsoft Azure Cosmos DB		
NuoDB		
Trafodion		
VoltDB		

Source: Bryan, Garnier & Co



Cloud data warehousing

An increasing number of data warehouses are now available in the cloud rather than on-premise. Cloud data warehouses have the advantage of being provisioned in minutes and without any technical expertise. They suit small and mid-sized companies, business analysts, and other non-technical users who want to access, store and process large amounts of data at low cost. Initially focused on simple data loading and reporting, cloud data warehouses now support complex cases such as IoT analytics, real-time analytics, fraud analysis and 360-degree customer views at the hundreds of terabyte (and even petabyte) scale. And while cloud data warehouses were initially only deployed for new projects, some organisations are migrating their on-premise data warehouses to the cloud, whether it is a public, private or hybrid cloud.

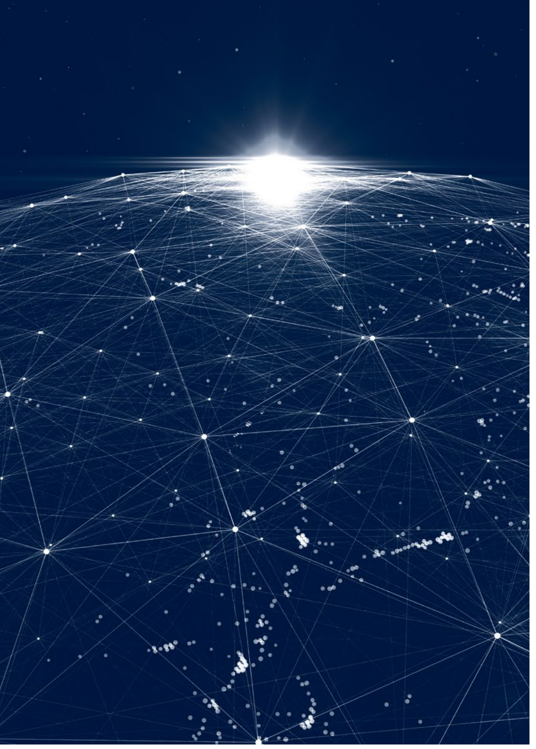
While the on-premise data warehousing software market is dominated by Oracle, IBM, Microsoft, Teradata and SAP, the cloud data warehouse market is led by Amazon Web Services (Redshift), Snowflake, Google (BigQuery), and Oracle (Autonomous Data Warehouse Cloud).

Challengers in this market include Teradata (IntelliCloud), IBM (Db2 Warehouse on Cloud), Microsoft (Azure SQL Data Warehouse), and Cloudera (Hortonworks HDP). Smaller players exist, like MarkLogic, Pivotal, Exasol and Yellowbrick Data, while some large players like Alibaba, Micro Focus (Vertica) and Huawei have a weak presence.

As for access, Amazon and Google provide a fully managed data warehousing service that is available exclusively on their own public cloud infrastructure. Oracle's cloud data warehouse is only available within the Oracle Cloud. IBM is available as a managed cloud service on IBM's cloud infrastructure (SoftLayer), although it can also be deployed on all major cloud providers with the potential 'downside' of the customer having to manage the software themselves. However, most of the other independent players have a multi-cloud approach, using Amazon Web Services, Microsoft Azure, Google Cloud and/or their own data centers to run their software. Finally, Yellowbrick Data - which exited stealth in July 2018 and since then has raised USD 92m - has its data warehouse available on-premise, on private clouds, on colocation and on the edge (IoT), and plans to be available on public clouds in 2019.

3. Cloud-based data warehousing: the Snowflake case

> *"In many cases, Snowflake is the only product that's viable for customers to move to. Customers demand a level of scale that Amazon Redshift just can't do. We can scale to a large number of queries. Because we're just a very large cloud service, we throw the power of the cloud at running things at the same time. Every single day, we run almost 20 million queries a day."*
>
> *Bob Muglia, CEO, Snowflake Computing.*



Snowflake, a 'superfunded' company

The cloud-based data warehousing software vendor Snowflake Computing is, in our view, a good example of the new generation of successful data management companies that have emerged from the cloud revolution. The company develops and sells a cloud-based storage and analytics platform, the Snowflake Elastic Data Warehouse (EDW), which uses a cloud-based infrastructure and has been available since June 2015. As a testimony to its success, large companies like Capital One, Nielsen, and Adobe use Snowflake EDW. New customers in 2018 include leading brands such as Netflix, Office Depot, Netgear, and Yamaha. In our view, Snowflake's success can be attributed to the strategic vision of its founders, the involvement of its CEO Bob Muglia, and a successful funding series.

Snowflake Computing was founded in 2012 in San Mateo, California, by CTO Benoît Dageville, Thierry Cruanes (Product Architect) and Marcin Zukowski. Benoît Dageville spent 15 years at Oracle as the lead architect for parallel execution and a key architect in the SQL Manageability group. Before Oracle, he worked at Bull, where he helped define the architecture and lead database performance efforts for its parallel systems. Thierry Cruanes spent 13 years at Oracle, working on the optimisation and parallelisation layers in Oracle databases. Before Oracle, he spent seven years at the IBM and Sema, working on data mining techniques. Finally, Marcin Zukowski was co-founder and CEO of Vectorwise, an analytical database software vendor sold to Actian in 2010. During its first two years of existence, Snowflake operated in stealth mode, looking to create data warehouse software designed for the cloud.

The company came out of stealth mode in 2014, just after the appointment of CEO Bob Muglia. At Microsoft between 1988 and 2011, he had worked as the product manager for SQL Server, Vice-President Windows NT, Server Application Group and .NET Services Group, then Senior Vice-President Servers & Tools. Then at Juniper Networks he served as EVP Software & Solutions (2011-2013).

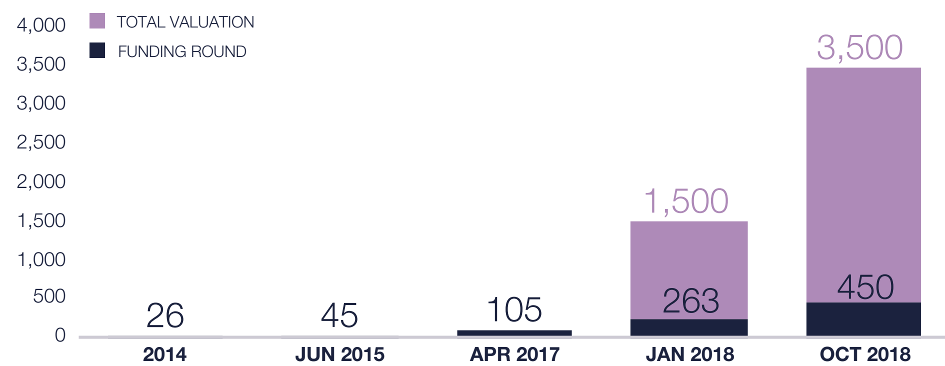
In June 2015, Snowflake announced the general availability of Snowflake EDW, as well as the first in a series of funding rounds (See Fig. 18 and 19).

FIG. 18: SNOWFLAKE COMPUTING – SUMMARY OF FUNDING ROUNDS

Date	Lead investor(s)	Other investors	Capital raised/valuation
June 2015	Altimeter Capital	Redpoint ventures Sutter Hill Wing VC	USD 71m
April 2017	Iconiq Capital Madrona Venture Group	Existing investors	USD 105m
January 2018	Iconiq Capital Altimeter Capital Sequoia Capital		USD 263m/USD 1.5bn
October 2018 (pre- IPO)	Sequoia Capital	Existing investors Meritech Capital	USD 450m/USD 3.5bn

Source: Crunchbase

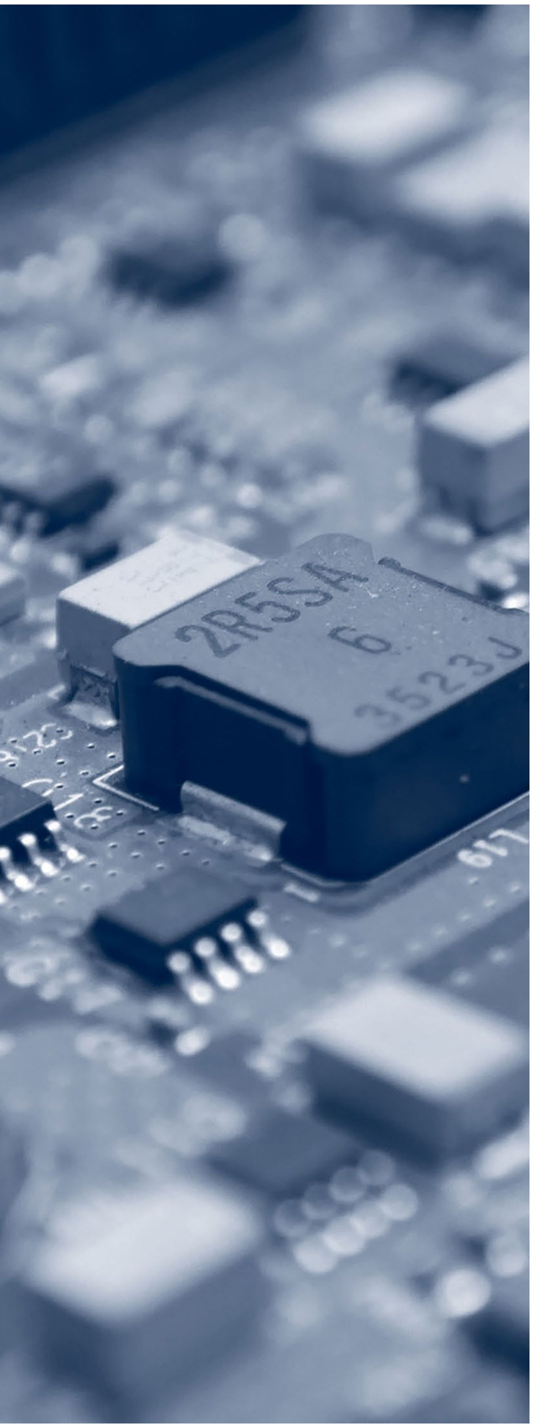
FIG. 19: SNOWFLAKE COMPUTING - TOTAL FUNDING AND VALUATION (USD M)



Source: Snowflake company data

In January 2018, CEO Bob Muglia publicly stated that the USD 263m round would probably be the last before a possible IPO. At that time, the company had more than 1,000 customers, up from 450 in April 2017, and expected this to double over 2018. Snowflake employed 330 staff, up from 175 in April 2017, and expected this to also almost double, to 600, during 2018. The two large funding rounds in 2018 was aimed helping Snowflake in four areas:

1. continuing to expand its multi-cloud strategy beyond Amazon Web Services and Microsoft Azure;
2. growing sales teams across and outside the US to meet increasing demand for the only cloud-built data warehouse;
3. investing further in Snowflake's data warehouse-as-a-service by growing its engineering team in San Mateo, CA, Bellevue, WA, and Berlin, Germany);
4. delivering innovations such as Snowflake Data Sharing, also known as 'The Data Sharehouse', which further consolidates its lead over legacy cloud and on-premise solutions.



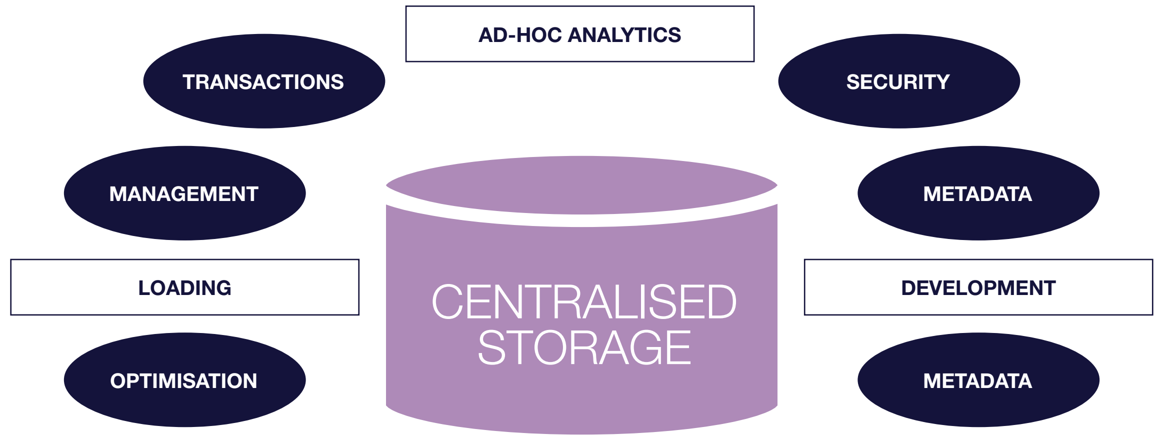
Snowflake's approach: 100% 'as-a-service'

Snowflake's data warehouse-as-a-service uses the SQL language to manage data. It can work with a mix of structured data such as CSV files and tables, as well as semi-structured data like JSON, Avro, XML and Parquet. It uses the Amazon Web Services S3 service as a cloud infrastructure. Since September 2018, Snowflake has also been generally available on Microsoft Azure, integrated with Azure services such as Microsoft Azure Data Lake Store and Microsoft Azure Power BI. The platform also employs several new features, for example limitless storage accounts, accelerated networking and storage soft delete. As of today, Snowflake is available in the Azure East-US-2 region (in Virginia), with more regions to be added in the future[3]. Whether it is running on Amazon Web Services or Microsoft Azure, from the US and Dublin, Snowflake generally costs USD 40 per terabyte of data per month for on-demand storage, and USD 23 per terabyte per month for capacity storage. From Frankfurt, pricing is USD 45 and USD 24.50 respectively.

The Snowflake EDW platform has the following features:

- The cloud storage layer is engineered to scale independently of compute resources. It can process data loading and unloading without impacting running queries. Snowflake uses micro-partitions ('sharding') to store customer data, which are compressed in columns and encrypted.
- On the compute layer, data processing is performed by virtual warehouses, which retrieve the data required from the storage layer to satisfy queries, then cache it locally with computing resources along with the caching of query resources to improve the performance of future queries. Multiple data warehouses can simultaneously operate on the same data.
- The services layer authenticates user sessions, manages, enforces security, performs query compilation and optimisation, and coordinates transactions. It uses a distributed metadata store for global state management. Metadata processing is powered by a separate integrated sub-system, which allows data processing without using query computer resources.

FIG. 20: SNOWFLAKE EDW PRODUCT ARCHITECTURE



Source: Snowflake Computing, company data

The company was initially competing with Amazon Redshift, another cloud-based data warehousing system. However, as an increasing number of large firms move towards cloud adoption and look for solutions that can handle structured, semi-structured and transactional data, Snowflake has been increasingly competing with Oracle, IBM and Teradata. The company often picks up customers disillusioned with Hadoop, and it benefits from the IoT wave as the semi-structured data generated by sensors is complicated to manage with a structured approach.

Bob Muglia's long-term vision for Snowflake is that the company's data lake technology becomes a platform for building data applications. This could be for data sharing between customers or to sell that data. He foresees companies building businesses on top of the data stored in the Snowflake data warehouse much as companies have built businesses on top of the Salesforce platform.

[3] Microsoft Azure has at least 14 cloud regions in the US and 54 globally.

4. In-memory analytics database: the Exasol case

“Exasol’s vision is to provide the most extensible, powerful and platform independent data analytics framework for customers to govern, connect and understand the entirety of the data flowing through their organisations. As they transition to data-driven processes, companies need to see and leverage all the data necessary for the success of their business, wherever the data may be. For this they will need an incredibly powerful and intelligent data framework which will permit new tools, technologies and processes to be implemented on demand, as smoothly and as simply as possible.”

Aaron Auld, CEO, Exasol

Exasol, a hidden champion

Headquartered in Nuremberg (Germany), Exasol develops and sells a high-performance, in-memory, column-oriented relational database management and data warehousing system specifically designed for analytics. It launched its first product in 2005. Either standalone or integrated with Hadoop, Exasol is used in a wide range of Big Data use cases, including accelerating standard reporting, running multi-user ad-hoc analytics, and performing complex modelling. Exasol’s ability to pull in data in real-time allows end-users to compress the multi-step process of data collection and pre-processing into a rapid execution analytics engine. This means users can perform quick ad-hoc analysis and answer critical business questions in minutes.

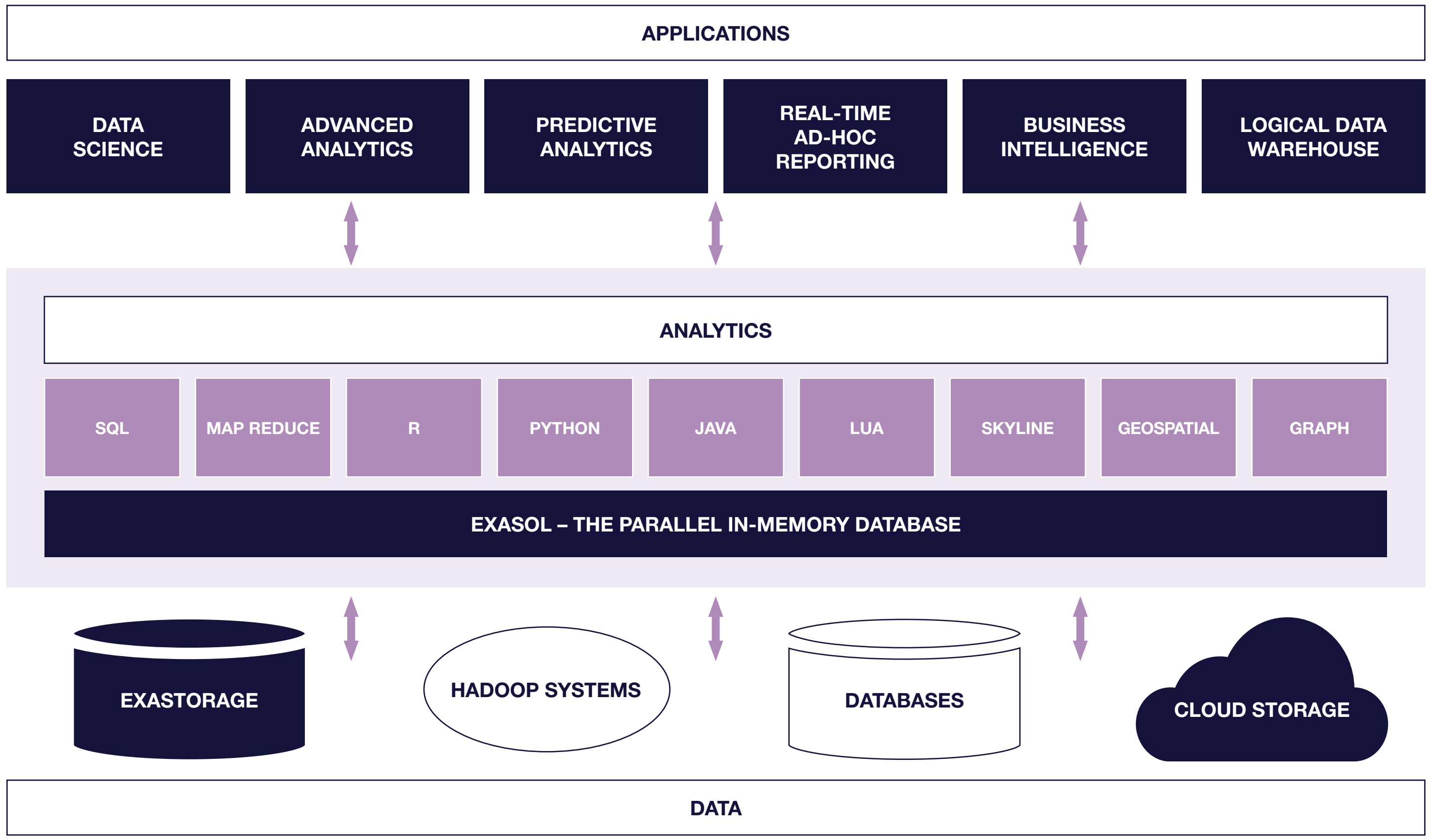
Exasol’s core differentiation lies in its proprietary technology architecture that combines in-memory plus MPP (Massive Parallel Processing) which results in extreme high querying speed at large data volumes. Their platform integrates with a wide range of front-end analytics tools, such as Tableau and Microstrategy.

Exasol has offices in Germany, France, the UK and the US, and sells its products through partners in the rest of Europe, Israel and the USA, but also directly to customers in India, Australia and South America. As of today, more than 140 companies use Exasol to run business intelligence and data analytics, including well-known names such as Adidas, Zalando, GfK, King Digital Entertainment, Vodafone, Xing, Revolut, Dailymotion, and Postbank.

Exasol’s technology was created by one of its founders, Falko Mattasch, in the 1990s. While researching intelligent algorithms in the Parallel Processing Department at the University of Jena (Germany), he showed how data analytics could benefit from parallel processing. After leaving academia, he implemented a new prototype for a parallel database system, which served as the basis for the technology behind Exasol. It took years for the company to develop a complete database, as everything was created in-house without recourse to open source database cores such as PostgreSQL. This led to a software system that offered optimum performance and was highly scalable and ‘tuning-free’. As data was continuously changed and updated, Exasol designed the system so that not all data had to be stored in the main memory.

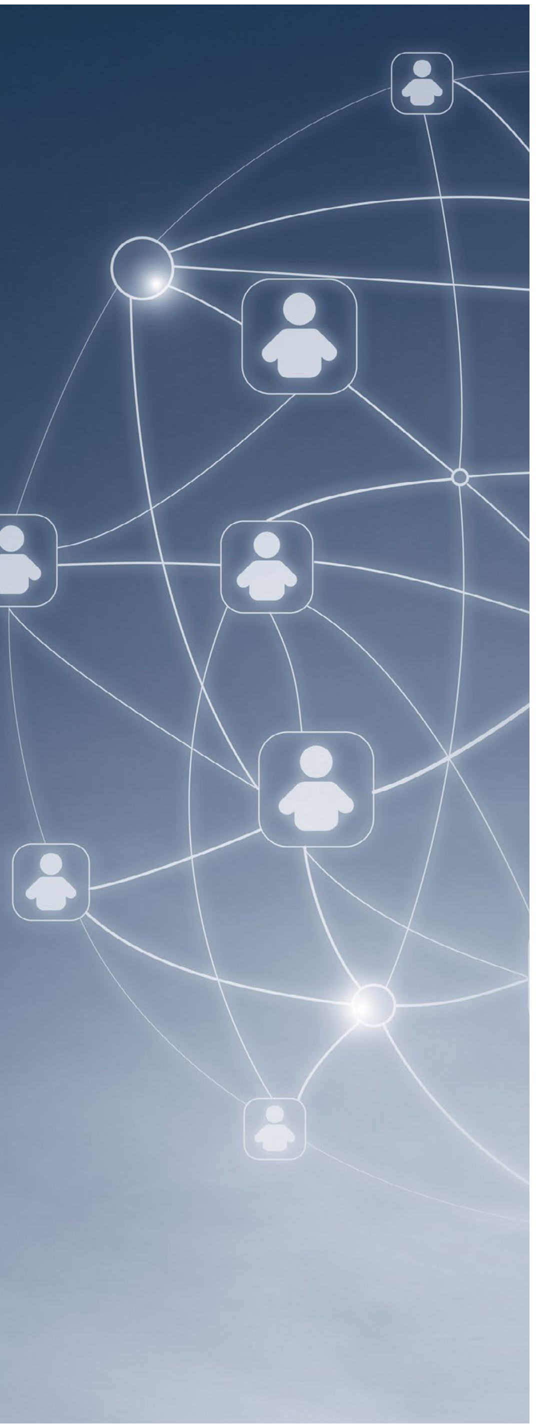
The main memory served as a large cache, from which content was placed in local storage when it was not efficient to fit everything into the cache. That way, it was possible to analyse more data even more quickly than with a pure in-memory database. Current main shareholders are board members, founders, the German State-owned bank KfW and the Swiss early-stage venture firm Mountain Partners.

FIG. 21: EXASOL PRODUCT ARCHITECTURE



Source: Exasol company data

5. Conclusion



In this report, we have reviewed the deep changes happening in the database world, and the new generation of players aiming to challenge longstanding leaders such as Oracle, Microsoft and IBM. The question now is how the competitive landscape will be reshaped ten years on. Will Oracle remain as the #1 database player? Can Amazon Web Services be the new #1 in this market? Is Hadoop definitely a fallen 'star'? How the market will consolidate? What could be the next database 'unicorns'? Let's try to address these questions.

Oracle, in our view, will be increasingly challenged by Amazon Web Services, but this does not mean it will lose its leadership in the database market. Recently, AWS CEO Andy Jassy declared that by the end of 2018 or mid-2019, 88% of Amazon's databases that ran on Oracle would be on an Amazon database instead. He also noted that Amazon moved its data warehouse from Oracle to its own service, Redshift in early November 2018. SAP's long-term ambition is to replace Oracle databases running on its application software with SAP HANA. Other defections may follow. But Oracle has the strength of an installed base (several hundred thousand of customers) which is not fully dependent on Amazon or SAP. Oracle's databases are also evolving, and the new Oracle Autonomous Database is adding resilience. Oracle also has the financial might to acquire cloud database or data warehouse players if they add value to its offerings and for shareholders. So, while it's not dying yet, its competitive position is not as comfortable as it used to be.

AWS can steal the #1 position in databases. But with a market share is at 9%, it could take a long time to get there. However, Amazon's growth comes from the cloud: for every cloud customer it gets, there is a potential customer for DynamoDB, Aurora and Redshift. That said, its database and data warehouse offerings are tied to its cloud, so a Microsoft Azure or Google Cloud customer will not use AWS database or data warehouse services.

Hadoop is losing ground because of its complexity but is not dead in our view. The 2019 merger between Cloudera and Hortonworks means growth is unlikely to be as strong as expected. We think that independent players may well disappear, probably through M&A. However, almost all the large software vendors have made significant investment in Hadoop open source technologies, which suggests it is here to stay and will probably evolve, supported by the open source Apache Software Foundation.

As for market consolidation, we saw a significant wave of M&A deals between 2008 and 2013, before the advent of NoSQL and NewSQL databases and cloud data warehouses. Many small database players have emerged since the beginning of the decade, with only a handful achieving significant scale or funding. Nonetheless, some have become 'unicorns', most recently Snowflake Computing in data warehousing. We believe there will be an opportunity for incumbent database companies to acquire some of these players to stay competitive with AWS. As Elastic and MongoDB demonstrated, EV/sales valuation multiples can be very rich. Consequently, adding a sizeable premium to the valuation of the next IPOs in this space is tempting.

Finally, who are the new unicorns? In data management, it's now MongoDB, Elastic and Snowflake. There are plenty of potential unicorns out there too: we think Neo4j, DataStax (Cassandra), Databricks, Cockroach Labs, MemSQL, Yellowbrick Data, Couchbase, and MariaDB could be the next ones. In Europe, there are only a few players, such as Elastic in The Netherlands and MariaDB in Finland, but Exasol, a German company, has potential to benefit from the cloud data warehouse wave as its technology is based on a solid and powerful in-memory database.

BRYAN, GARNIER & CO

White Paper Authors

Falk Müller-Veerse
Partner
Technology
fmuellerveerse@bryangarnier.com

Gregory Ramirez
Equity Research Analyst
Software & IT Services
gramirez@bryangarnier.com

Berk Kirca
Vice President
Investment Banking
bkirca@bryangarnier.com

Priyanshu Bhattacharya
Vice President
Investment Banking
*pbhattacharya@bryangarnier.*com

Technology Team

INVESTMENT BANKING

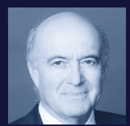
PARIS

Greg Revenu
Managing Partner
grevenu@bryangarnier.com

Olivier Beaudouin
Partner, Technology
& Smart Industries
obeaudouin@bryangarnier.com

Guillaume Nathan
Partner, Digital Media
& Business Services
gnathan@bryangarnier.com

Thibaut De Smedt
Partner,
Application Software
& IT Services
tdesmedt@bryangarnier.com



Philippe Patricot
Managing Director,
Technology
ppatricot@bryangarnier.com

MUNICH

Falk Müller-Veerse
Partner
Technology
fmuellerveerse@bryangarnier.com

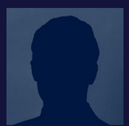
Lars Dürschlag
Managing Director
lduerschlag@bryangarnier.com

EQUITY RESEARCH ANALYST TEAM

Richard-Maxime Beaudoux
Video Games, Payment
& Security
rmbeaudoux@bryangarnier.com

Thomas Coudry
Telecoms & Media
tcoudry@bryangarnier.com

Gregory Ramirez
Equity Research Analyst
Software & IT Services
gramirez@bryangarnier.com



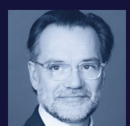
David Vignon
Industrial Software
dvignon@bryangarnier.com

Fréderic Yoboué
Semiconductors
fyoboue@bryangarnier.com

EQUITY CAPITAL MARKETS

Pierre Kiecolt-Wahl
Partner
pkiecoltwahl@bryangarnier.com

EQUITY DISTRIBUTION



Nicolas-Xavier de Montaigut
Partner, Head of Distribution
nxdemontaigut@bryangarnier.com

Nicolas d'Halluin
Partner, Head of US Distribution
ndhalluin@bryangarnier.com

Corporate Transactions

Bryan, Garnier & Co leverage in-depth sector expertise to create fruitful and long lasting relationships between investors and European growth companies.

NETBREEZE	allegorithmic	xerox	intersec	Prima Solutions
Acquired by	Strategic Investment	Acquired by	Private Placement	Acquired by
Microsoft	Adobe	NAVER	HIGHLAND CAPITAL PARTNERS and Cisco	THE CARLYLE GROUP
Undisclosed	Undisclosed	Undisclosed	Undisclosed	Undisclosed
Sole Advisor to the Sellers	Sole Advisor to the Sellers	Sole Advisor to the Seller	Sole Financial Advisor	Sole Advisor to the Sellers

About Bryan, Garnier & Co

Bryan, Garnier & Co is a European, full service growth-focused independent investment banking partnership founded in 1996. The firm provides equity research, sales and trading, private and public capital raising as well as M&A services to growth companies and their investors. It focuses on key growth sectors of the economy including Technology, Healthcare, Consumer and Business Services. Bryan, Garnier & Co is a fully registered broker dealer authorised and regulated by the FCA in Europe and the FINRA in the U.S. Bryan, Garnier & Co is headquartered in London, with additional offices in Paris, Munich and New York. The firm is a member of the London Stock Exchange and Euronext.

Bryan, Garnier & Co Technology Equity Research Coverage

ALBIOMA, altice, altran, ALTEN, ASM, ASML, Atos, axway, Besi, bouygues, Capgemini, Dassault Systèmes, dialog, easyVISTA, faurecia, focus, gemalto, HELLA, indra, iliad, infineon, ingenico, McPhy, orange, Plastic Omnium, SafeCharge, sage, SAP, software, soitec, sopra steria, SWORD, ST, TEMENOS, TIM, UBISOFT, u-blox, Valeo, wirecard, worldline

5 Analysts | 50+ Stocks Covered

With more than 150 professionals based in London, Paris, Munich and New York, Bryan, Garnier & Co combines the services and expertise of a top-tier investment bank with a long-term client focus.

BRYAN, GARNIER & CO

LONDON
Beaufort House
15 St. Botolph Street
London, EC3A 7BB
UK

T: +44 (0) 20 7332 2500
F: +44 (0) 20 7332 2559

Authorised and regulated by the Financial Conduct Authority (FCA)

PARIS
26 Avenue des Champs Elysées
75008 Paris
France

T: +33 (0) 1 56 68 75 00
F: +33 (0) 1 56 68 75 01

Regulated by the Financial Conduct Authority (FCA) and the Autorité de Contrôle prudential et de resolution (ACPR)

MUNICH
Widenmayerstrasse 29
80538 Munich
Germany

T: +49 89 242 262 11
F: +49 89 242 262 51

NEW YORK
750 Lexington Avenue
New York, NY 10022
USA

T: +1 (0) 212 337 7000
F: +1 (0) 212 337 7002

FINRA and SIPC member

IMPORTANT INFORMATION