Critical Capabilities for Data Management Solutions for Analytics

Published: 18 March 2019   **ID:** G00355667

Analyst(s): Rick Greenwald, Adam Ronthal

Data management solutions for analytics offerings are consolidating, with major vendors able to address a range of use cases and smaller vendors addressing a subset of use cases. Data and analytics leaders can use this research to guide evaluation and initial vendor selection for DMSA offerings.

Key Findings

- **Large vendors are returning —** Large, established vendors are building on core strengths and capabilities to address a full range of use cases.

- **Offerings are expanding beyond core data management —** All vendors are starting to expand their product capabilities to integrate metadata management, data integration, governance and the aspects required for long-term strategic success. This requires users to explore all options with their current solutions before selecting a new vendor for one of these areas.

- **"Best fit" is predominant in the cloud —** Major cloud vendors have introduced a variety of best-fit offerings as a part of their standard architecture, rather than have a best-of-breed approach.

Recommendations

For data and analytics leaders responsible for data management solutions as part of strategizing and planning information infrastructure:

- Evaluate the capabilities of your incumbent solution(s) against new use cases, to determine if existing expertise could be used to reduce development time with a good-enough solution already in place.

- Plan on using a heterogeneous solution landscape overall, but try and reduce duplication of effort by categorizing use cases with regard to their target deployment platform.

- Use a logical data warehouse architecture when you need to integrate separate data repositories efficiently, keeping in mind performance SLAs that may be impacted by remote access.

- Plan for eventual integration with other data silos when scoping the effort needed to implement a specific solution, to avoid crippling overhead caused by proliferating data silos.

What You Need to Know

This document was revised on 14 May 2019. The document you are viewing is the corrected version. For more information, see the Corrections *page on gartner.com.*

Market Trends

Demand for incorporating increasingly varied data sources and their associated use cases continues to expand in the data management solutions for analytics (DMSA) landscape. This pressure is forcing vendors to expand their capabilities. It has contributed to the expansion of the concept of the logical data warehouse (LDW), which allows multiple data types and sources to be accessed through a single logical interface, as well as to the growth in distributed database architectures.

These distributed architectures provide extended capabilities, but also introduce some data challenges. These forces show no sign of slowing, indicating that best-fit solutions addressing a smaller number of use cases will continue to be attractive as an alternative to a single-repository platform play.

The predominance of a best-fit approach in the cloud is aided by the fundamental ability of the cloud to deliver those common management functions automatically, lowering the overhead for multiple best-fit solutions. In addition, cloud providers control the environment in which these services are deployed, making it easier for them to add integration between services. However, a best-fit approach, by definition, can lead to more effort integrating multiple offerings and instances, and the ease of provisioning individual instances may lead to an even greater number of integration candidates.

Cloud vendors can also stream in fixes and new features, making delivery faster for end users. They can also monitor their large fleets of users of each service for information about future fixes and feature upgrades.

The growth of dbPaaS is still the major story in the market; but, as the Key Findings indicate, this story is no longer as disruptive as it was a few years back. Major traditional vendors now have cloud capabilities whose robustness, developed over time, is part of newer cloud offerings. Having a cloud option is no longer a differentiator between cloud-focused vendors, such as Amazon Web Services (AWS) and Google, and traditional on-premises vendors.

Smaller existing vendors are struggling in an environment where buyer interest is directed to new players and platforms. But they still maintain their existing strengths. The time for challengers (other

than cloud providers) to see significant improvement in their market position may have passed, but there is no reason for clients to abandon such products or not consider them where their strengths are appropriate.

Many vendors are starting to supplement their existing products with extended features and newer capabilities. This tendency has led to more overlap between offerings, which can complicate product evaluation cycles.

The Product

Gartner Definition: Data Management Solution for Analytics

A complete software system that supports and manages data in one or many file management systems (most commonly a database or multiple databases). These solutions include specific optimization strategies designed to support analytical processing, including — but not limited to — relational processing, nonrelational processing (such as graph processing), and machine learning (ML) or programming languages (such as Python or R).

Data is not necessarily stored in a relational structure and can use multiple models (relational, document, key value, text, graph, geospatial and others).

At Gartner we state that a DMSA:

- Is a system for storing, accessing, processing and delivering data that is intended for one or more of the four primary use cases that Gartner identifies as supporting analytics (see the Use Cases section).

- Is not limited to a single specific class or type of database management system (DBMS).

- May consist of many different data management technologies in combination. However, any offering or combination of offerings must, at its core, exhibit the ability to provide access to the data under management by open-access tools via commonly used APIs.

- Must include mechanisms to isolate workload requirements and control various parameters of end-user access within managed instances of data.

- Must manage the storage of and access to data residing in a type of storage medium, which may include — but is not limited to — hard-disk drives, flash memory, solid-state drives and DRAM.

Critical Capabilities Research

Coverage

This Critical Capabilities research is aimed at data and analytics leaders. We have focused on the 12 most important functional — critical — capabilities that are required to support the four major use cases we have identified. The research combines analysis of product functions and customer experience to evaluate the support offered by each vendor or products for these critical capabilities.

We evaluated user experience based on the companion Magic Quadrant reference survey, Gartner inquiries, peer insights, in-depth reference calls and interactions with vendors (see the Evidence section). In addition to customer experience, capability ratings include Gartner analysis of differentiating product capabilities (see the Critical Capabilities Definition section).

Gartner took into account both the documented capabilities of the products and the results of the user surveys on the actual adoption of these capabilities. The survey results were given significantly greater weight than the stated capabilities or analyst opinions, as the ultimate proof of use is the end users. Consequently, the results in this Critical Capabilities research should be seen as somewhat lagging — especially for emerging use cases — as organizations need time to implement newer functionality into their environments.

We placed the largest emphasis on data points and trends collected from the survey this year. As customer adoption of new features and technologies is not immediate, the newer use cases (such as real-time data warehouse) have a bias toward incumbent solutions. These are frequently the default choice for new use cases as a market approaches maturity. New solutions are more likely to have new, advanced capabilities, but non-risk-averse initial adopters of these are a smaller part of the market. The main, more-risk-averse part of the market is more likely to use incumbent products that are not yet able to fully implement these new capabilities.

Although this research shares survey results from the 2019 "Magic Quadrant for Data Management Solutions for Analytics," it does not offer an overall estimation of each vendor. Instead, the critical capability ratings focus on how well a specific vendor product addresses one of four use cases. This research focuses on a single product from each vendor, while the Magic Quadrant considers all relevant products or services. Additional products that supported the core functionality of the main product were also considered in this body of research, while similar offerings were not. This approach tended to benefit best-of-breed as opposed to best-fit vendors.

This research does not include all of the criteria that data and analytics leaders should investigate before selecting a particular DMSA vendor, focusing instead on a set of critical capabilities that specifically are used in the four use cases. Many other criteria not included in our analysis will come into play in this research, such as whether the offering is a stand-alone DBMS software package, appliance or cloud solution. Other requirements — pricing, vertical industry offerings, the availability of services and so on — are not included but would need to be part of a formal RFP process (see "Toolkit: RFP Template for Data Warehouse and Data Management Solutions for Analytics"). Such aspects do factor in the evaluations for the Magic Quadrant.

Scoring

Readers should understand that our scores are meant to convey a product's standing in relationship to the market at the time the data was finalized. As such, scores for any capability are not absolute from year to year, but relative and only relevant within the context of this specific yearly report.

As detailed below, a score of 3.0 indicates that a product met the requirements for a particular use case. Although vendors are listed in the order of their relevant ranking (and alphabetically in the case of an equivalent score), be aware of the meaning of the individual ratings.

In some cases, the overall range of these scores may shift from year to year. These changes are the result of both changing market conditions and refinements in the calculations used to evaluate these capabilities. The following have occurred in this year's research:

- Changes to criteria for evaluating support of external data sources

- Replacement of a criterion for repeated queries with a criterion for query optimization
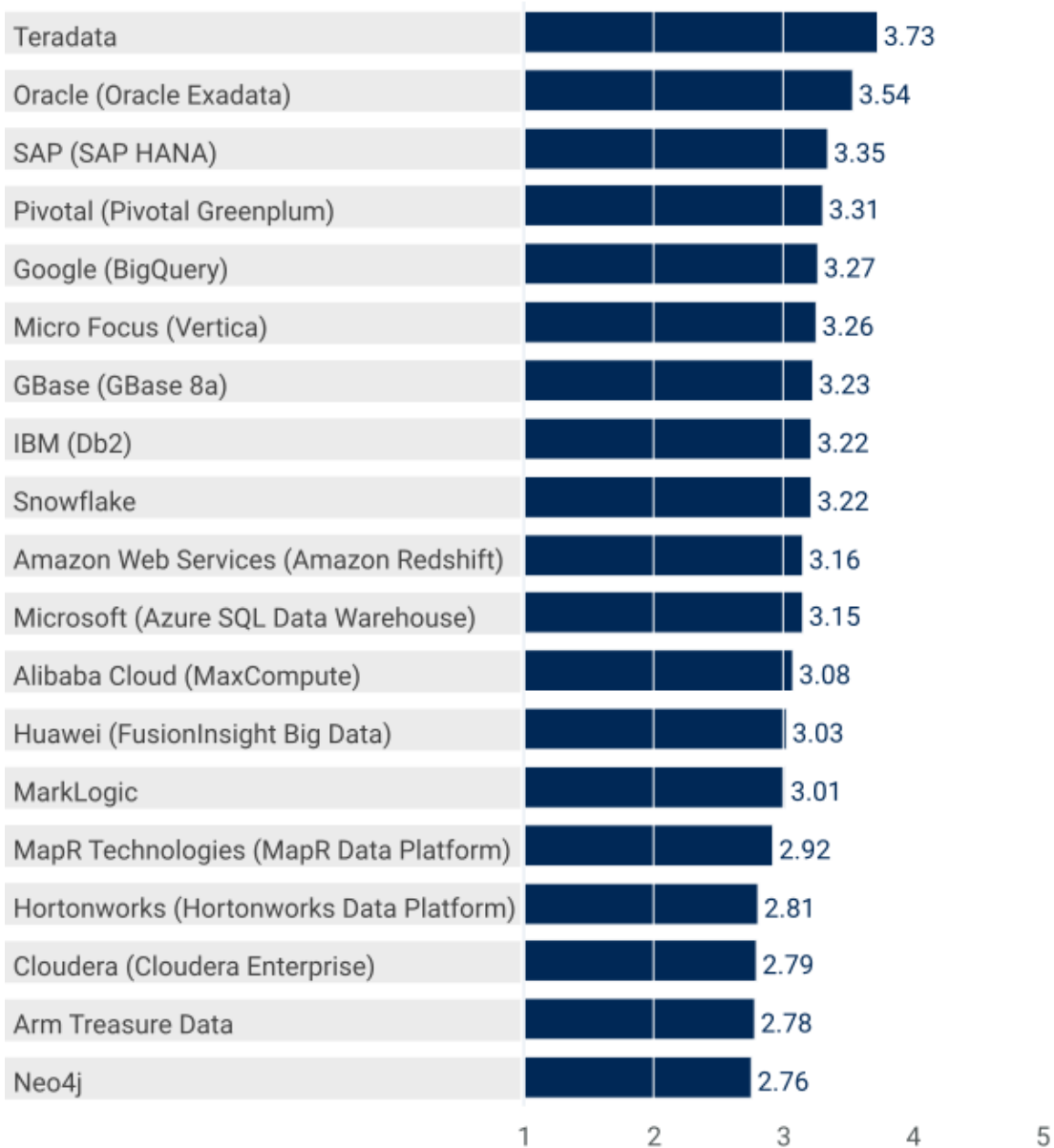
Note: Gartner does not recommend using any rating as the sole or primary basis for product selection, as there are many factors outside the scope of this research that can impact the suitability of a product.

Analysis

Critical Capabilities Use-Case Graphics

Figure 1. Vendors' Product Scores for Traditional Data Warehouse Use Case

Product or Service Scores for Traditional Data Warehouse

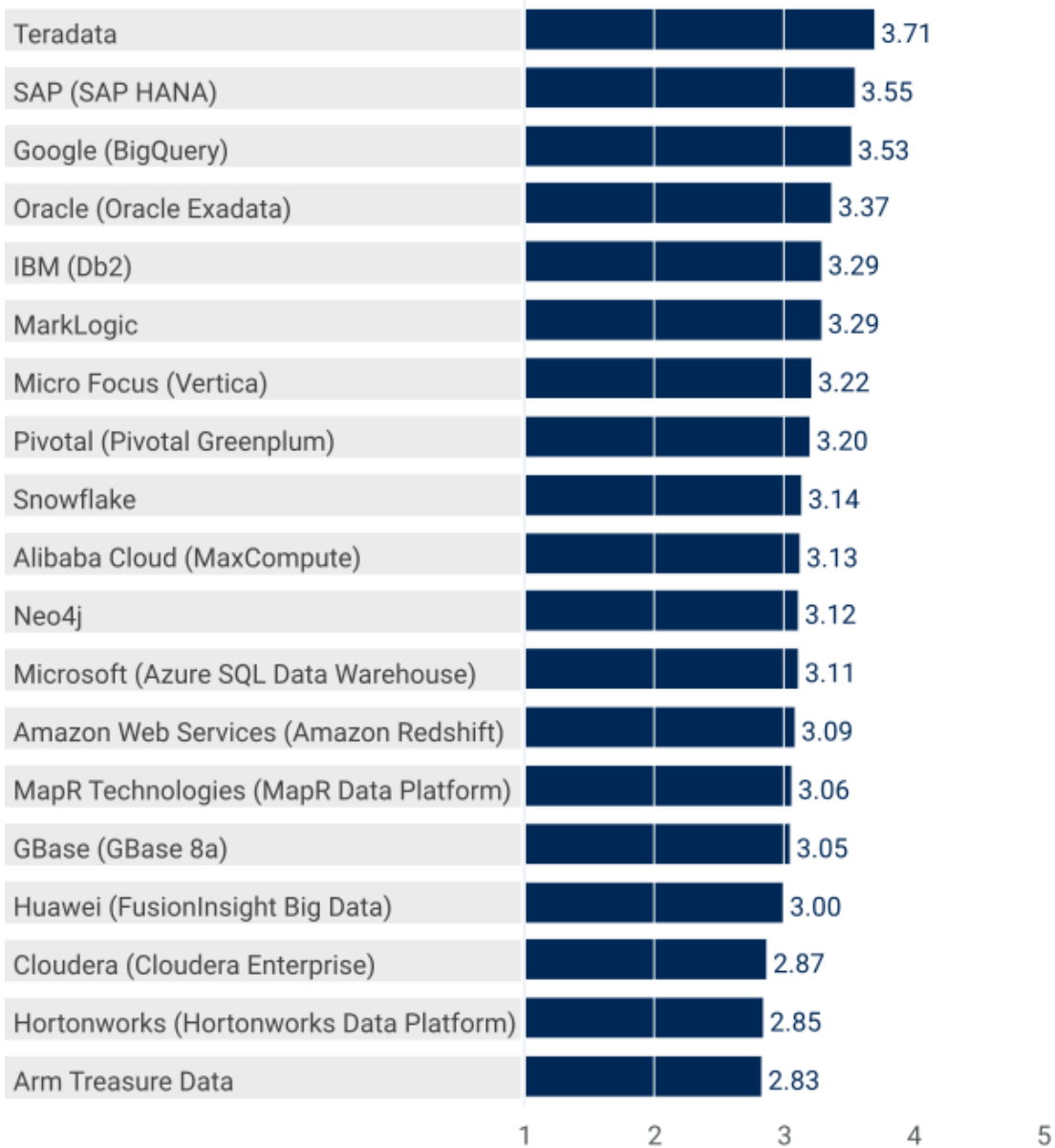| Vendor	Score
Teradata	3.73
Oracle (Oracle Exadata)	3.54
SAP (SAP HANA)	3.35
Pivotal (Pivotal Greenplum)	3.31
Google (BigQuery)	3.27
Micro Focus (Vertica)	3.26
GBase (GBase 8a)	3.23
IBM (Db2)	3.22
Snowflake	3.22
Amazon Web Services (Amazon Redshift)	3.16
Microsoft (Azure SQL Data Warehouse)	3.15
Alibaba Cloud (MaxCompute)	3.08
Huawei (FusionInsight Big Data)	3.03
MarkLogic	3.01
MapR Technologies (MapR Data Platform)	2.92
Hortonworks (Hortonworks Data Platform)	2.81
Cloudera (Cloudera Enterprise)	2.79
Arm Treasure Data	2.78
Neo4j	2.76

As of 21 January 2019   © Gartner, Inc

Source: Gartner (March 2019)

Figure 2. Vendors' Product Scores for Real-Time Data Warehouse Use Case

Product or Service Scores for Real-Time Data Warehouse

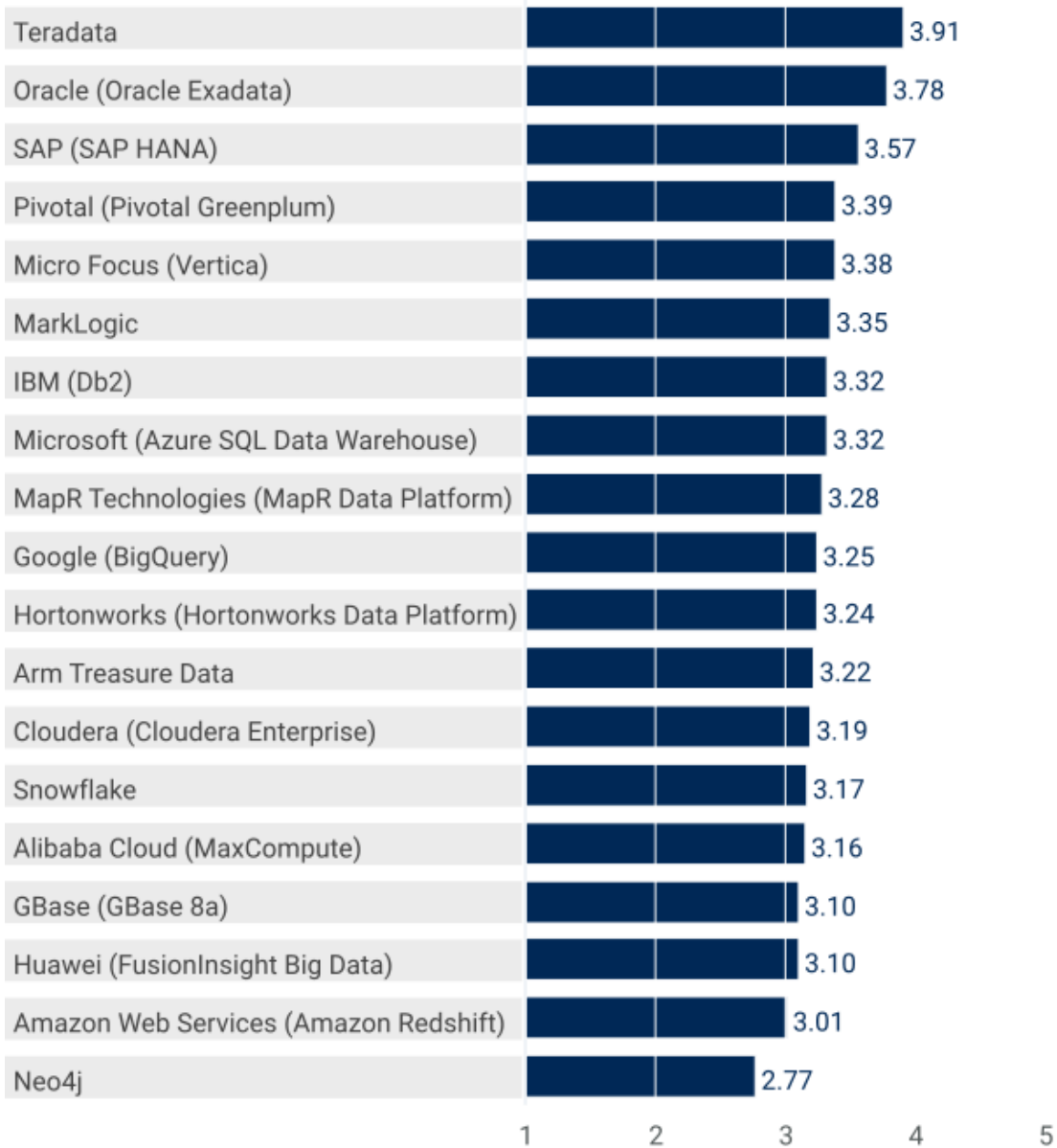| Vendor	Score
Teradata	3.71
SAP (SAP HANA)	3.55
Google (BigQuery)	3.53
Oracle (Oracle Exadata)	3.37
IBM (Db2)	3.29
MarkLogic	3.29
Micro Focus (Vertica)	3.22
Pivotal (Pivotal Greenplum)	3.20
Snowflake	3.14
Alibaba Cloud (MaxCompute)	3.13
Neo4j	3.12
Microsoft (Azure SQL Data Warehouse)	3.11
Amazon Web Services (Amazon Redshift)	3.09
MapR Technologies (MapR Data Platform)	3.06
GBase (GBase 8a)	3.05
Huawei (FusionInsight Big Data)	3.00
Cloudera (Cloudera Enterprise)	2.87
Hortonworks (Hortonworks Data Platform)	2.85
Arm Treasure Data	2.83

As of 21 January 2019   © Gartner, Inc

Source: Gartner (March 2019)

Figure 3. Vendors' Product Scores for Logical Data Warehouse Use Case

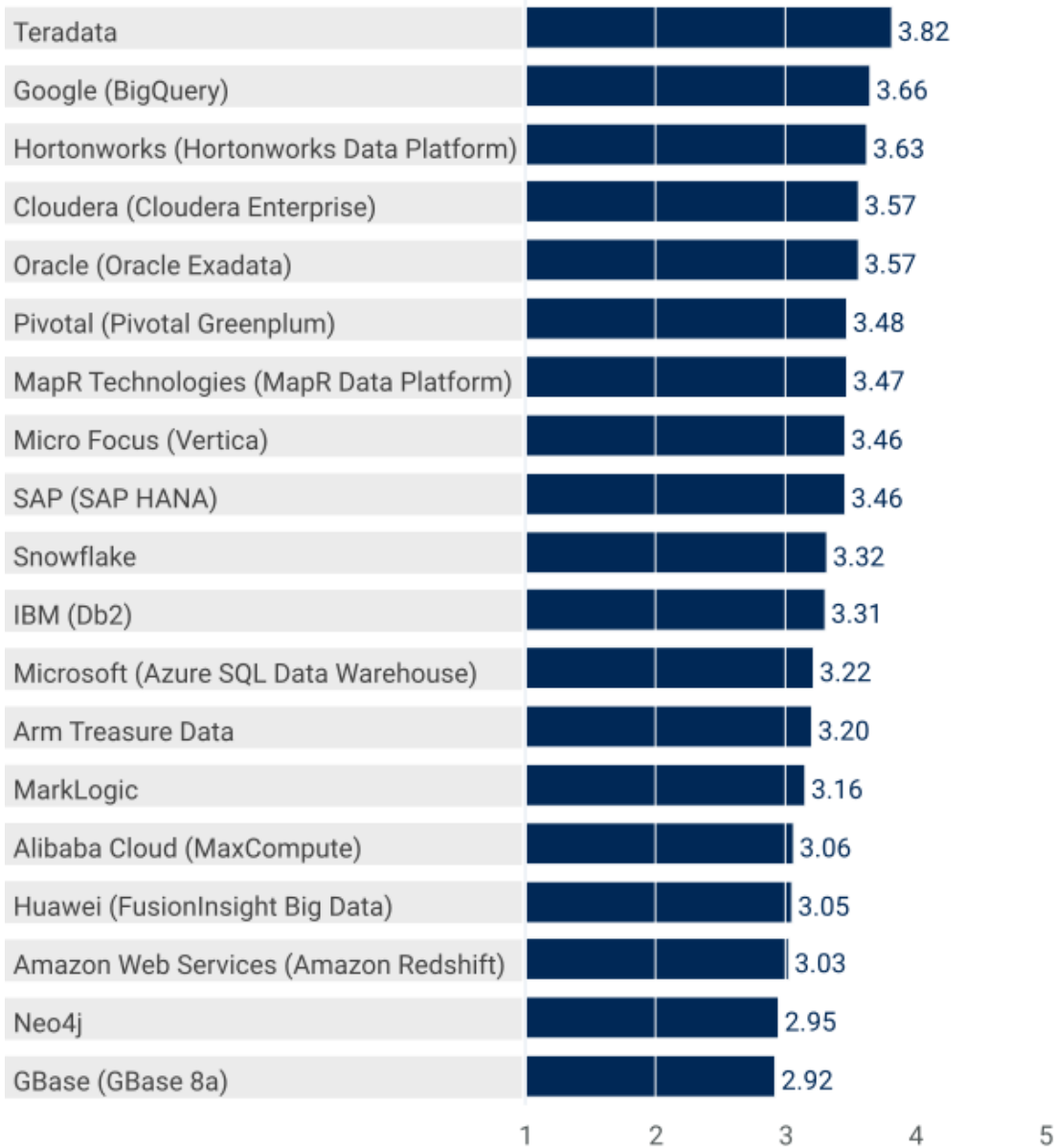Product or Service Scores for Logical Data Warehouse

Vendor	Score
Teradata	3.91
Oracle (Oracle Exadata)	3.78
SAP (SAP HANA)	3.57
Pivotal (Pivotal Greenplum)	3.39
Micro Focus (Vertica)	3.38
MarkLogic	3.35
IBM (Db2)	3.32
Microsoft (Azure SQL Data Warehouse)	3.32
MapR Technologies (MapR Data Platform)	3.28
Google (BigQuery)	3.25
Hortonworks (Hortonworks Data Platform)	3.24
Arm Treasure Data	3.22
Cloudera (Cloudera Enterprise)	3.19
Snowflake	3.17
Alibaba Cloud (MaxCompute)	3.16
GBase (GBase 8a)	3.10
Huawei (FusionInsight Big Data)	3.10
Amazon Web Services (Amazon Redshift)	3.01
Neo4j	2.77

As of 21 January 2019   © Gartner, Inc

Source: Gartner (March 2019)

Figure 4. Vendors' Product Scores for Context-Independent Data Warehouse Use Case

Product or Service Scores for Context-Independent Data Warehouse

Vendor	Score
Teradata	3.82
Google (BigQuery)	3.66
Hortonworks (Hortonworks Data Platform)	3.63
Cloudera (Cloudera Enterprise)	3.57
Oracle (Oracle Exadata)	3.57
Pivotal (Pivotal Greenplum)	3.48
MapR Technologies (MapR Data Platform)	3.47
Micro Focus (Vertica)	3.46
SAP (SAP HANA)	3.46
Snowflake	3.32
IBM (Db2)	3.31
Microsoft (Azure SQL Data Warehouse)	3.22
Arm Treasure Data	3.20
MarkLogic	3.16
Alibaba Cloud (MaxCompute)	3.06
Huawei (FusionInsight Big Data)	3.05
Amazon Web Services (Amazon Redshift)	3.03
Neo4j	2.95
GBase (GBase 8a)	2.92

As of 21 January 2019   © Gartner, Inc

Source: Gartner (March 2019)

Vendors

Alibaba Cloud (MaxCompute)

Alibaba Cloud is the cloud computing division of Alibaba Group Holding, a multinational conglomerate based in Hangzhou, China. It offers a wide variety of services, such as ApsaraDB for RDS (relational database service) for MySQL, SQL Server and PostgreSQL; and HybridDB for PostgreSQL, based on the open-source Pivotal Greenplum Database. It also offers HybridDB for MySQL, AnalyticDB for online analytical processing (OLAP) analysis; MaxCompute for large data warehouse implementations; and E-MapReduce for Hadoop. In addition, Apsara Stack Agility provides an on-premises private cloud implementation.

Alibaba Cloud MaxCompute met requirements for all four defined DMSA use cases, and is a solid choice for a DMSA solution for those strategically invested in Alibaba Cloud. On average, more than half the respondents to our customer reference survey reported using the product for prebuilt analytic queries supported by data marts, views, cubes or semantic-enabled modeling interfaces. The next tier of usage is formed of operational BI queries in support of fixed, repetitive and production-use queries, as well as exploratory and predictive queries associated with data science workloads in the context-independent data warehouse. The lowest reported usage is for ad hoc "train of thought" -type queries.

MaxCompute appears to be well-positioned for real-time, low-latency workloads. All respondents reported data availability from collection to analytics-ready within an hour, and 75% of those reported availability in a minute or less.

MaxCompute received below-average scores for advanced analytics capabilities, performance optimization for exploratory use cases, and flexible scalability. Surprisingly given the product history, it received one of the lowest scores of all the vendors for its ability to accommodate a variety of data types. This indicates that the product is still used primarily for structured, well-known data types that are associated with more traditionally oriented workloads.

Amazon Web Services (Amazon Redshift)

Amazon Web Services (AWS) is a wholly owned subsidiary of Amazon, which is based in Seattle, Washington, U.S. AWS offers Amazon Redshift, a data warehouse service in the cloud. Amazon Redshift includes Amazon Redshift Spectrum, a serverless, metered query engine that uses the same optimizer as Amazon Redshift but queries data in both Amazon Simple Storage Service (S3) and Amazon Redshift's local storage. AWS also offers Amazon S3, a cloud object store; AWS Glue, a data integration and metadata catalog service; and Amazon Elasticsearch Service, a search engine based on the Lucene library. Additional offerings include Amazon Kinesis, a streaming data analytics service; Amazon EMR, a managed Hadoop service; Amazon Athena, a serverless, metered query engine for data residing in Amazon S3; and Amazon QuickSight, a BI visualization tool. Finally, Amazon Neptune is a graph database service.

Amazon Redshift rated above 3.00 ("meets requirements") across all four use cases, and was in the middle of the rankings across those use cases. Reference customer results were low for data ingest, based on a low rate of survey respondents who had used data in near real time. Most

respondents used Amazon Redshift only for their traditional and exploratory workloads — the highest percentage across all the vendor surveys in both categories.

Amazon Redshift had one of the highest percentages of survey respondents who would recommend the product to others, and 80% of its reference customers indicated that they would be purchasing more of the product in the coming year.

Keep in mind this: Our Critical Capabilities requirement to evaluate a single product runs counter to the best-fit strategy of AWS. Products such as Amazon Athena, which can be used against data stored in AWS's cloud object store, or catalog and transfer tool Amazon Glue, are designed to be used with Amazon Redshift, but cannot be considered here. AWS is also centered on data on its own platform and across its services, so suffers in evaluations of an LDW, where many other vendors offer connectivity options outside their own platforms.

Arm Treasure Data

Arm Treasure Data, recently acquired by Arm, is based in Mountain View, California, U.S. It provides Customer Data Platform (CDP), a fully managed DMSA running on AWS infrastructure, with availability in regions of the U.S. and Japan. CDP provides a cloud data lake combined with relational data marts. The ability to ingest data from a wide range of sources, and to feed data to downstream data management platforms and enterprise applications, is a focus of the vendor.

Treasure Data scored above the 3.00 ("meets requirements") threshold for the LDW and context-independent use cases. Reference clients gave high scores for its ability to integrate data from multiple sources and its ingestion capabilities. They praised especially the richness of its API connectors, resulting in a high score for its ability to access multiple data sources.

Treasure Data rated below 3.00 for the other two use cases. In terms of critical capabilities, it scored lowest for managing large volumes of data, with most of its clients having implementation sizes below 50TB.

Since its acquisition by Arm, Treasure Data has been focusing on managing customer data and CRM analytics use scenarios.

Cloudera (Cloudera Enterprise)

Cloudera, which is based in Palo Alto, California, U.S., offers the Cloudera Enterprise platform. Versions of this include Cloudera Enterprise Data Hub (EDH) and Cloudera Data Warehouse (for BI and SQL workloads based on Apache Impala). Additional versions include Cloudera Data Science & Engineering (for data processing and ML based on Apache Spark and Cloudera Data Science Workbench) and Cloudera Operational DB (for real-time data delivery based on Apache HBase and Apache Kudu).

Through its shared data experience technologies, the platform provides unified security, governance and metadata management across these workloads, as well as across deployment environments. Cloudera Workload XM provides tools to efficiently migrate, analyze, optimize and scale analytics workloads. Cloudera's platform is available on-premises, across the major cloud environments

(including native object store support for Amazon S3 and Azure Data Lake Store), and as a managed service under the Cloudera Altus brand.

Cloudera achieved above 3.00 ("meets requirements") across the LDW and the context-independent data warehouse use cases. These scores result from the vendor's ability to manage a variety of data types, plus its high score in the advanced analytics capability. These ratings are aligned with data lake uses of Cloudera's solution by its client base.

Cloudera's scores for the traditional data warehouse and real-time data warehouse use cases were affected by the vendor's relatively low score for the administration and management, workload management and optimized performance (traditional) capabilities. This mix of capabilities plays an important role in a product supporting mixed workloads for traditional use cases. Cloudera EDH must compete with relational DBMSs that are better-suited for these two use cases.

Note: Hortonworks and Cloudera have merged with a commitment to support existing products from both companies for three years. This merger had not been announced at the time of the reference customer survey, so results are reported based on the market at that time.

GBase (GBase 8a)

GBase is a trading name of Tianjin Nanda General Data Technology, which is based in Beijing, China. GBase offers GBase 8a, a relational massively parallel processing (MPP) data warehousing platform; GBase Infinidata 8a, a data warehouse appliance; and GBase HD, a Hadoop distribution based on Apache Hadoop. It also offers GBase UP, an LDW platform supporting data virtualization between GBase 8a, GBase HD and other platforms; and GBase cloud DB (GBase 8a), available in the QingCloud app center.

GBase 8a scored above 3.00 ("meets requirements") for three of the four use cases, with its strongest showing in traditional data warehouse. It scored just below the "meets requirements" threshold in the context-independent data warehouse use case — the realm of data science discovery and exploration. Given GBase 8a's in-database analytic capabilities, this is likely more a reflection of real-world usage as reported in our reference customer survey than any glaring product deficiencies.

On average, more than half the respondents to our reference customer survey reported using the product for prebuilt analytic queries supported by data marts, views, cubes or semantic-enabled modeling interfaces. Indeed, more than a third reported that 80% or more of their query workloads fell into this category. In contrast, almost no respondents reported using the product in support of data science exploratory workloads supporting predictive modeling and forecasting.

Respondents also reported using GBase 8a almost exclusively for batch-oriented traditional data loading activities. Nearly 90% reported data availability from collection to analytics-ready taking an hour or more.

Google (BigQuery)

Google, based in Mountain View, California, U.S., is a wholly owned subsidiary of the Alphabet holding company. Google Cloud is the part of Google that focuses on delivering solutions and services to the business market. Google's dbPaaS offerings in Google Cloud Platform include BigQuery, a serverless, managed data warehouse offering; Cloud Dataproc, a managed Spark and Hadoop service; and Cloud Dataflow, focused on stream and batch processing of data.

Google BigQuery is specifically designed to address the needs of the DMSA market. BigQuery scored above 3.00 ("meets requirements") for all four use cases, with particularly strong showings in the real-time data warehouse and context-independent data warehouse use cases.

In terms of capabilities, BigQuery received high scores for advanced analytics, exploratory use support and data ingest. It ranked above 3.00 for all capabilities except access to multiple sources and workload management. But despite its sub-3.00 score for workload management, reference clients praised the ease of use and performance of BigQuery.

Hortonworks (Hortonworks Data Platform)

Hortonworks is based in Palo Alto, California, U.S. It offers a data management platform called Hortonworks Data Platform (HDP), Hortonworks DataFlow for streaming data delivery and ingestion (powered by Apache NiFi), and the Azure HDInsight service for Microsoft Azure. It also offers the Hortonworks Data Cloud Hadoop service for AWS as well as Hortonworks DataPlane Service, a unified architecture to manage, govern, store, process and access datasets across multiple use scenarios and across multiple hybrid deployment environments, including multicloud and on-premises.

HDP is a Hadoop-based solution that is often used for data lake implementation. Reference customers said they used it primarily for two reasons, mostly in equal proportions: (1) To provide an integrated and consistent dataset across multiple business domains for analysis by all users; (2) As a context-independent data warehouse. These align with data lake implementations in support of experimental uses of data, represented by the vendor's highest rating being for this use case (comfortably meeting requirements). Further evidence of these two uses is that more than 50% of the vendor's reference customers have deployments over 50TB. HDP also rated highly for the advanced analytics capability.

However, for the traditional and real-time data warehouse use cases, HDP received low scores for the traditional use support and workload management capabilities.

Note: Hortonworks and Cloudera have merged with a commitment to support existing products from both companies for three years. This merger had not been announced at the time of the reference customer survey, so results are reported based on the market at that time.

Huawei (FusionInsight Big Data)

Huawei, based in Shenzhen, China, offers the FusionInsight Big Data platform, a data management platform that combines components of Apache Hadoop, Spark and Storm with FusionInsight

GaussDB 200, a proprietary MPP DBMS. Huawei has added industry-specific domain models in some cases and worked with partners in others. It has added proprietary extensions to the Hadoop platform for event stream processing, graph and ML capabilities, and a unified SQL engine that is compatible with its MPP database and runs on Hadoop. Additional enhancements have been made to the Hadoop scheduler with Huawei's Superior Scheduling Engine, and to the supported Hadoop Distributed File System (HDFS) file formats with Apache CarbonData. Huawei's offerings are also available in the vendor's public cloud and through its partners.

Although Huawei FusionInsight Big Data ranked in the bottom half across all four use cases, it scored at least 3.00 ("meets requirements") across all of the use cases. It was in the upper half in terms of supporting a variety of data types.

FusionInsight Big Data scored in the middle of the pack across all critical capabilities. Reference customers scored it toward the bottom of all the vendors in terms of their willingness to recommend the product, but near the top in terms of their intention to purchase more licenses (on par with Hadoop and Chinese vendors).

IBM (Db2)

IBM, which is based in Armonk, New York, U.S., offers stand-alone DBMSs (Db2, Db2 for z/OS, Informix) and appliances (PureData System for Analytics, PureData System for Operational Analytics, Integrated Analytics System, Db2 Analytics Accelerator). It also offers Hadoop solutions (Big SQL), managed data warehouse cloud services (Db2 Warehouse on Cloud) and private cloud data warehouse capabilities (Db2 Warehouse). IBM Db2 Big SQL and Fluid Query provide a consolidated access tier to a wide range of DBMSs and Hadoop distributions. IBM's Db2 Event Store provides a data management foundation for IoT and time series event data.

IBM Db2 meets requirements for all four DMSA use cases, scoring in the top half for all but the context-independent data warehouse use case. While Db2 has strong in-database analytic capabilities based on IBM Netezza capabilities, its scores were below average for ability to access multiple data sources and to accommodate a variety of data types. Only a third of respondents to our reference customer survey reported connecting to data sources outside of their Db2-based DMSA environment.

Survey respondents on average reported that their most frequent use of Db2 is to support operational BI queries, followed by ad hoc queries, then analytic queries supporting prebuilt analytic interfaces (dashboards, data marts, cubes, etc.). Exploratory queries in support of data science workloads comprised, on average, only 10% of the reported usage. However, some users did report using Db2 more heavily in this situation, in some cases for up to a third of their query workloads.

Also of note is that 60% of Db2 reference customers reported low-latency availability of data, with data being available for analytics within a minute of collection. This positions Db2 well as a real-time operational data warehouse engine.

MapR Technologies (MapR Data Platform)

MapR Technologies, which is based in Santa Clara, California, U.S., offers its MapR Data Platform in both open-source and commercial software editions. MapR Data Platform features include performance and storage optimizations using Network File System (NFS) and MapR-XD, a scalable POSIX-compliant data storage tier; and MapR Database, an Apache HBase-compatible, nonrelational DBMS supporting key value, document, wide-column, graph and time series models. It also includes event-streaming capabilities (MapR Event Store for Apache Kafka), high-availability improvements, and administrative and management tools. MapR Edge, a small-footprint edition of MapR Data Platform, extends MapR's reach to edge-processing use scenarios that are common to IoT environments.

MapR did better than the other two main Hadoop vendors, Cloudera and Hortonworks, in all the use cases bar context-independent data warehouse. However, it was still in the lower half of the complete vendor list for two of those cases, and below a 3.00 ("meets requirements") in the traditional data warehouse use case. All of its survey respondents indicated that they would be purchasing more in the coming year (the top score across all vendors).

In terms of capabilities, MapR had the second-lowest survey result for administration and management, and also scored poorly for traditional user support — the latter being common across the Hadoop cohort.

MarkLogic

MarkLogic, which is based in San Carlos, California, U.S., offers a nonrelational multimodel DBMS that it describes as "operational and transactional." The product is available in two editions: Essential Enterprise and a free Developer edition. Essential Enterprise can be deployed on-premises, in the cloud and across hybrid infrastructures, including those of AWS, Microsoft Azure and the Google Cloud Platform, as well as on VMware, Pivotal's Cloud Foundry and Red Hat platforms (the latter recently acquired by IBM). MarkLogic also offers a Data Hub for integrating data, either on-premises or as a cloud service.

MarkLogic met requirements for all four use cases, receiving its best relative ranking in the LDW use case. This is to be expected given its focus on integrating multiple data silos.

MarkLogic also did well in the real-time data warehouse use case, based on receiving the top score of all the vendors from the reference customer survey in terms of data ingest. Respondents also scored the vendor highly in terms of recommending the solution to others. Although most respondents indicated that they would be buying more licenses in the coming year, MarkLogic's result in this area places it in the lower third of all vendors. Few of its reference customers reported having instances in production larger than 100TB.

Micro Focus (Vertica)

Micro Focus, which is based in Newbury, U.K., offers the Vertica analytics platform. This platform is available as Vertica Enterprise, a columnar relational DBMS delivered as a software-only solution for on-premises use. It is also available as Vertica in the Clouds; as machine images from the AWS,

Microsoft Azure and Google Cloud Platform marketplaces; and as Vertica for SQL on Hadoop. Micro Focus also recently announced Vertica in Eon Mode (available on AWS), which enables the separation of compute and storage to capitalize on cloud economics and dynamic workloads by scaling compute resources independently of shared storage.

Vertica met requirements for all four DMSA use cases, ranking in the top eight for all. This reflects one of the major trends in the DMSA market this year: rediscovery. End users are using traditional technologies in order to meet their DMSA requirements rather than utilizing an additional vendor, and Vertica's strong capabilities here as a columnar MPP relational database are well-showcased.

Vertica received its highest relative ranking in the LDW use case. This reflects the vendor's focus on enabling the Vertica compute engine to run with multiple storage architectures, including Hadoop's HDFS storage, and multiple cloud object stores, as well as its native support for file formats such as Parquet and ORC.

On average, Micro Focus reference customers reported running the highest proportion of queries in support of operational BI queries, characterized by fixed, repetitive production reports.

Microsoft (Azure SQL Data Warehouse)

Microsoft, which is based in Redmond, Washington, U.S., offers SQL Server as a software-only solution with certified configurations. It also sells Azure SQL Data Warehouse (fully managed, MPP cloud data warehouse), Azure HDInsight (Hadoop distribution based on Hortonworks), Azure Databricks (Apache Spark-based analytics platform) and Azure Data Lake (big data store and analytics platform) as cloud services. In addition, it offers the Analytics Platform System, an MPP data warehouse appliance.

This is the first year that Azure SQL Data Warehouse has been evaluated in this research. Microsoft achieved scores above the "meets requirements" threshold of 3.00 across all four use cases, with an average position in all relative to all the vendors. The two capabilities that it ranked the lowest for are managing large volumes of data and data ingest.

Reference customers reported mostly having deployments under 50TB and with only a limited portion of the data continuously loaded. Capabilities for workload management and optimized performance for traditional use cases scored below 3.00, which reflects reference client issues with the performance of Azure SQL Data Warehouse Gen1. However, polybase capabilities for accessing data outside Azure SQL Data Warehouse led to a good score for access to multiple data sources.

Neo4j

Neo4j, which is based in San Mateo, California, U.S. and Malmö, Sweden, provides a graph platform that includes the Neo4j native graph DBMS, graph analytics, the Cypher graph query language, data integration, and graph visualization and discovery tools. The company offers the open-source Neo4j Community Edition; Neo4j Desktop, which is free for developers, startups and data scientists; and the paid-for Neo4j Enterprise Edition for production deployments. The company recently released Neo4j Bloom, which provides advanced graph visualization capabilities that enable both experienced and novice users to derive insights from graph processing. It also released

Cypher for Apache Spark (CAPS) for deriving and analyzing graphs from Spark data, and has built a library of nearly two dozen in-database graph algorithms.

Neo4j has the strongest focus of any vendor in this research on a single type of analysis — one based on graphs. As such, it has less applicability across the four (broader) use cases in this research, scoring below a 3.00 ("meets requirements") in all but the real-time data warehouse use case. Its focus on that specific type of analysis was supported by the reference customer survey results, in which every respondent indicated that Neo4j was brought in for a new use scenario.

It received poor survey results in three capabilities: access to multiple data sources, traditional use support and managing large volumes of data. These scores make sense since graph analysis is not a traditional use case; graphs can be run on smaller volumes of data, and data for graphs is usually stored in a graph database for performance reasons.

Survey respondents did give Neo4j very good scores for data ingestion as well as administration and management, which led to its highest ranking being in the real-time data warehouse use case. Neo4j also scored well in delivering performance for exploratory analysis use cases.

Oracle (Oracle Exadata)

Oracle, based in Redwood Shores, California, U.S., provides Oracle Database 18c, Oracle Exadata Database Machine, Oracle Big Data Appliance, Oracle Big Data Management System, Oracle Big Data SQL and Oracle Big Data Connectors. In addition, the Oracle Cloud service provides Oracle Database Cloud Service, Oracle Database Cloud Exadata Service, Oracle Big Data Cloud Service and the Oracle Autonomous Data Warehouse (ADW) Cloud. Oracle's cloud portfolio also includes on-premises solutions in the form of Oracle Database Exadata Cloud at Customer and Oracle Big Data Cloud at Customer.

Oracle did well across all four use cases, ranking among the top vendors for all. It had the highest rating of all the vendors for its ability to access multiple data sources, and among the highest for performance optimization for traditional use cases and workload management, which contributed to those use case rankings. Its lowest capability score was for data ingest.

Two other categories shed light on Oracle's strengths and weaknesses. Although Oracle had one of the strongest responses of all the vendors in terms of reference customers' willingness to recommend the solution to others, it had the weakest for intent to purchase more in the coming year. This result may be due, in part, to the more-regular buying patterns of enterprise customers, who tend to purchase products as part of multiyear deals.

Pivotal (Pivotal Greenplum)

Pivotal, which is based in San Francisco, California, U.S., offers the Pivotal Greenplum database — an open-source MPP database based on PostgreSQL. Available in the AWS, Microsoft Azure and Google Cloud Platform marketplaces, Pivotal Greenplum can also be installed as software on bare metal or virtually with VMware vSphere. Pivotal and Dell have also partnered to provide the

Greenplum Building Block Solution for customers looking to deploy Pivotal Greenplum in an appliancelike configuration of commodity hardware.

Pivotal Greenplum scored well in all four use cases, and was especially strong in the traditional and logical data warehouse use cases. This reflects one of the major trends in the DMSA market this year: rediscovery. End users are turning to traditional technologies in order to meet their DMSA requirements, and Pivotal Greenplum's strong capabilities here as an MPP relational database are well-showcased.

Its lowest position was for the real-time data warehouse — the newest use case in this research and, as such, more fluid than the others. In terms of critical capabilities, Pivotal Greenplum's score for data ingestion were somewhat low, which was the major factor relating to the real-time data warehouse. Note, however, that all survey results reflect existing rather than future use.

SAP (SAP HANA)

SAP is based in Walldorf, Germany. It offers SAP HANA, an in-memory column-store DBMS that supports operational and analytical use cases. SAP also offers SAP BW/4HANA, a packaged data warehouse solution. Both are offered as cloud solutions (for deployment in public and private clouds, and on SAP Cloud Platform), as stand-alone software and as an appliancelike hardware reference architecture. The vendor also offers SAP Cloud Platform Big Data Services, a cloud-based Hadoop distribution; and SAP HANA Vora (offered within SAP Data Hub), a HANA-like engine that can run within the nodes of a Spark cluster.

SAP HANA ranked among the top vendor solutions in three of the four use cases. It is particularly suitable for the real-time data warehouse use case with its in-memory capabilities combined with data ingest. Its "meets requirements" rating for the traditional data warehouse use case reflects its adoption for this purpose among SAP customers. (Three out of four reference customers were using SAP BW on HANA, and one out of four was using SAP BW/4HANA.)

Although rating well for the traditional data warehouse, SAP HANA rated below average across all the vendors for managing large data volumes, indicating that many deployments are below 50TB. Access to multiple data sources is delivered with SAP HANA smart data access, which supports the logical data warehouse use case.

Snowflake

Snowflake, which is based in San Mateo, California, U.S., offers a fully managed data warehouse as a service on AWS and Microsoft Azure infrastructure. It supports ACID-compliant relational processing as well as native support for document store formats such as JSON, Avro, ORC, Parquet and XML. A native Apache Spark connector, R integration, support for user-defined functions, dynamic elasticity, temporal support and data-sharing capabilities round out the core offering. Recently announced partnerships with Qubole and Databricks extend Snowflake's reach to exploratory data lake use cases.

Snowflake meets requirements for all four DMSA use cases. The vendor's middle-of-the-pack ratings across the board reflect the overall maturity of an offering that has been generally available

for less than five years. Snowflake's reference customer survey responses placed it in the lower third of all vendors for the number of concurrent users and overall database size, as well as the length of time that they had been in production. This may also reflect Snowflake's relatively brief time in the market.

End users have reported using Snowflake fairly evenly across operational BI queries, analytic queries (for prebuilt dashboards, cubes and data marts) and ad hoc queries supporting "train of thought" analysis. Use of the product for predictive, exploratory, data-science-focused queries lagged these other scenarios.

Less than one-third of respondents to the reference survey reported low-latency data availability, where data is available to end users within a minute of being collected. This indicates that Snowflake may be used more for traditional, batch-oriented data loads. End users have regularly reported analyzing datasets measured in the 10s of TBs.

Teradata

Teradata is based in San Diego, California, U.S. and delivers data management solutions for analytics across any deployment environment — cloud, on-premises and hybrid. Teradata's offerings include a software-only analytics platform with an underlying SQL engine, ML engine and graph engine; the Teradata IntelliFlex and IntelliBase appliances; and business and analytic consulting services. Teradata IntelliCloud is an "as a service" cloud offering available on public cloud infrastructure (AWS and Microsoft Azure) and on the Teradata Cloud (optimized infrastructure). Support for the LDW comes in the form of Teradata's Unified Data Architecture (UDA). Teradata QueryGrid (part of the UDA) provides multisystem query support via the vendor's own software as well as via open-source Presto. Teradata also offers Hadoop support for Cloudera and Hortonworks distributions.

In October 2018, Teradata announced new packaging and branding for its analytics platform under the Vantage name.

Teradata is the top-ranking vendor for each of the four use cases and has a comfortable margin in each. The gap between Teradata and second-placed Oracle in the traditional data warehouse use case was the largest between any two vendors in this use case.

Teradata had the top reference customer survey score in workload management across all the vendors, and was in the upper half for 10 of the 12 critical capabilities. It had the lowest relative survey score in the impression of value for the money spent.

Context

Overall Performance

This year's Critical Capabilities scores illustrate the increasing breadth of viable solutions for DMSA. Some vendors did significantly better in some use cases than others, based both on their capabilities and on the adoption of their offerings for those use cases.

The results show more variation among vendors across the use cases, especially if you compare the traditional data warehouse use case with the context-independent use case. These two use cases roughly represent the difference between a classic data warehouse and a data lake. The relative disparity between the rankings in these two use cases highlights the need for the logical data warehouse, which allows simpler access across multiple data sources and data types.

Ultimately, most vendors that qualified for this research achieved a "good" rating of 3.00 or above for all the use cases, which indicates that their product "meets requirements." (Although there were more ratings below 3.00 than last year, especially in the traditional data warehouse use case, more than 84% of the ratings were above this.)

This year, no Hadoop-based vendor (Hortonworks, Cloudera or MapR) met requirements for the traditional data warehouse use case, with all being below 3.00. However, they were among top vendors for the context-independent data warehouse use case, which points to survey respondents' main focus for these vendors.

As in previous years, the ability to run in the cloud, on-premises or in a hybrid environment was not considered a critical capability. Certainly, market growth indicates that the overall customer base is moving toward cloud implementations, and every vendor in the research has some type of cloud option. Consequently, cloud deployments are not a distinguishing factor, either for cloud-native or more traditional vendors, at least in terms of DMSA capabilities.

Inclusion in this research should be seen as a significant accomplishment as there are stringent requirements to meet. In a similar manner, many vendors failed to meet a small number of requirements so may still be acceptable alternatives to the vendors in this research, especially for focused or edge scenarios.

Reference Customer Survey

Surveys were sent out to a list of reference customers given to Gartner by the vendors. Different vendors submitted different numbers of names, and not all vendors saw the same response rate from their customers. All vendors did have the same opportunities, and survey responses that were outliers were eliminated from consideration.

Unlike the Magic Quadrant, this Critical Capabilities research judges on the basis of a single offering. This led to reduced sample sizes from vendors with multiple offerings, as is typical of vendors that adhere to a best-fit product strategy. Additionally, best-fit vendors typically spread functionality across multiple products, and the single product focus might affect the evaluation in this research.

Product/Service Class Definition

The various capabilities identified below address the major needs identified above.

Critical Capabilities Definition

Access to Multiple Data Sources

This capability reflects the prevalence of queries across multiple data types and sources by customers across all types of queries, as well as access to data in other sources beyond the DBMS, such as other relational DBMSs or Hadoop distributions.

This capability is also rated on the functionality implemented when accessing external data sources, such as whether some kind of processing (e.g., predicate evaluation) is passed to the external data source for implementation within that source. Additionally, offerings could deliver some of this capability through storing multiple data types within their products.

Administration and Management

This capability demonstrates the product's ease of implementation, upgrade and ease of use, as expressed by customers. It covers overall ease of administration and management, not only during implementation but also during ongoing use and upgrade phases.

Scoring is also affected by the complexity of deployment and by vendor history. Some vendors have recent offerings for which upgrades may not yet have been released.

In addition to customer experience, this capability takes into consideration the completeness of vendor administration capabilities, such as role-based activities, advisors, utilization and capacity planning, resource allocation features and the user interface, as well as complexity of deployment and management.

Advanced Analytics

This capability reflects the product's ability to perform advanced analytic operations within itself. It was evaluated on the basis of what functionality was offered in the current version of the product and what functionality was actually being used by customers, based on their survey responses.

Data Ingest

This capability represents the prevalence of data being loaded continuously by customers. Some use cases more than others require data to be loaded from the operational sources in near real time, making this a key capability in the real-time data warehouse use case.

This capability was evaluated based on survey responses indicating continuous data loading and the amount of data loaded daily, as well as on analyst assessments using briefings and inquiries.

Managing Large Volumes of Data

This capability reflects if the volume of data managed by customers is large. This applies to data of multiple structures and formats.

It plays a role in all use cases but to various degrees, as it may not be equally important for all. In this context, we have defined "small" as being below 10 terabytes (TB) and "large" as being over 150TB, with consideration given to those vendors whose survey respondents reported data stores of 1 petabyte or larger. This year, we considered the mean rather than the median size of survey respondents' data stores, which avoids result skew based on a small number of very large data stores.

In addition to customer experience, this capability takes into consideration the ability of the vendor to address management of query workloads and the availability of price performance optimization options, as well as strategies for query optimization in isolation.

Optimized Performance (Traditional)

This capability reflects the features and functions of a product that was designed to address traditional data warehouse workloads. These features would be more focused on optimization of repeated and complex queries.

Optimized Performance (Exploratory)

This capability reflects the features and functions of a product designed to address exploratory data warehouse workloads, such as those used for building models or prescriptive analytics.

These workloads have a different set of requirements from traditional data warehouse workloads, so were evaluated separately.

Flexible Scalability

This capability reflects the ease with which a product can scale both up and down in response to changing workloads or user specifications.

Different products can deliver this capability in different ways. Cloud-based vendors can scale up with little user effort, although the separation of compute and storage can make it easier for the cloud vendor to implement this capability.

Distributed solutions typically can scale out more easily than nondistributed solutions, although there is significant variation even among distributed architectures in this area.

Variety of Data Types

This capability reflects the ability of an offering to support a variety of data types, either by native storage or by accessing those data types through some type of virtualized interface.

Workload Management

This capability evaluates how well a product manages different types and sizes of workloads.

This ability can significantly contribute to a product being able to handle demanding workloads without an excessive increase in resources, as well as being able to handle varying workloads without a corresponding variance in response times.

Use Support (Traditional)

This capability looks at the overall ability of a product to support traditional data warehouse workloads and their users. These workloads are typically initiated by nontechnical business users and casual users.

In this year's Critical Capabilities calculations, we classified business analysts and casual users as traditional data warehouse users, and data scientists and data miners as discovery users.

The criteria for traditional data warehouse use were based, in large part, on the relative percentage of users classified as traditional data warehouse users. These skill sets were defined as:

- *Business analyst* — Utilizes online analytical processing and dimensional tools to create new objects. Some faculty with computer languages and computer processing techniques.

- *Casual user* — Regularly uses portals and prebuilt interfaces. Minimally capable of designing dimensional analytics (if at all).

We also took into consideration some survey results and product evaluations relating to traditional data warehouse usage.

Use Support (Exploratory)

This capability looks at the overall ability of a product to support exploratory data warehouse workloads and their users, such as model building, predictive analytics and prescriptive analytics. These workloads are typically initiated by data science and data miner users.

Use Cases

Traditional Data Warehouse

This use case involves managing structured historical data coming from multiple sources. Data is mainly loaded through bulk and batch loading.

The traditional data warehouse use case can manage large volumes of data and is primarily used for standard reporting and dashboarding. To a lesser extent, it is also used for free-form ad hoc querying and mining, or operational queries. It requires high levels of capability for system availability as well as administration and management, given the mixed workload capabilities for queries and user skills' breakdown.

Real-Time Data Warehouse

This use case adds a real-time component to analytics use cases, with a goal of reducing latency between when data is generated and when it can be analyzed.

This use case primarily manages structured data that is loaded continuously via microbatching and/or streaming ingest analytics in support of real-time decision support, embedded analytics in applications, real-time data warehousing and operational data stores.

It primarily supports reporting and automated queries, in order to support operational needs or low-latency decision support, and will require high-availability and disaster recovery capabilities to meet operational demands. Managing different types of users or workloads — together with the ability to store large volumes of historical data — will be of less importance. This is because the major driver here is to provide a low-latency, real-time view of — and analytics on — operational data.

Logical Data Warehouse

This use case manages data variety and volume of data for both structured and other content data types, where the DMSA acts as a logical tier to a variety of data sources.

Besides structured data coming from transactional applications, this use case includes other content data types such as machine data, text documents, images and videos. Because such types can drive large data volumes and have specific data persistence requirements, access to data in disparate repositories is an important criterion.

The LDW is also required to meet diverse query capabilities and support diverse user skills. This use case supports queries reaching into other sources than the data warehouse DBMS alone, and may include metadata or data virtualization components.

Context-Independent Data Warehouse

This use case allows exploration of new data values, data form variants and relationships. It supports search, graph and other capabilities to uncover new information models.

This use case is primarily used for free-form queries to support forecasting, predictive modeling or other mining styles, as well as for queries supporting multiple data types and sources. It has no operational requirements and favors advanced users such as data scientists or business analysts, resulting in free-form queries across potentially multiple data types.

Vendors Added and Dropped

Added

- Arm Treasure Data — follows Treasure Data's acquisition by Arm parent company, SoftBank

- Huawei

Dropped

- Actian — did not meet the inclusion requirements for revenue

- MemSQL — did not meet the inclusion requirements for revenue

- Qubole — is a data science exploration platform rather than a DMSA

- Treasure Data — following its acquisition by Arm's parent company, now appears as Arm Treasure Data

Inclusion Criteria

The inclusion criteria represent the specific attributes that analysts believe are necessary for inclusion in this research:

- Vendors must have had DMSA software generally available for licensing, or supported for download, for approximately one year (since 1 December 2017). We do not consider beta releases.

- We use the most recent release of the software to evaluate each vendor's current technical capabilities. For existing solutions, and direct vendor customer references and reference survey responses, all versions currently used in production were considered. For older versions, we considered whether later releases may have addressed reported issues, but also the rate at which customers have or have not moved to newer versions.

- Product evaluations included technical capabilities, features and functionality present in the product or supported for download on 1 December 2018. Capabilities, product features or functionality released after this date could be included at Gartner's discretion and in a manner Gartner deemed appropriate to ensure the quality of our research product on behalf of our nonvendor clients. We also considered how such later releases might reasonably impact the end-user experience.

- Vendors should provide 30 verifiable DMSA production implementations that will exhibit generated revenue from distinct organizations, indicating they are in production, and:

  - A minimum of $40 million in revenue with a 50% growth rate year over year, *or*

  - More than $70 million in revenue.

    (Revenue can be from licenses, support and/or maintenance.)

  - The production customer base must include customers from three or more vertical industries (see Note 1).

  - Customers in production must have deployed DMSAs that integrate data from at least two operational source systems for more than one end-user community (such as separate business lines or differing levels of analytics).

Gartner.

This research note is restricted to the personal use of hugowu@gbase.cn.

- Vendor must demonstrate production customers from at least two distinct geographic regions. This means at least 10% (assessed by customer count or revenue percentage) of the verified production customer base must be outside of the vendor's home geography (see Note 2).

- Any acquired product must have been acquired and offered by the acquiring vendor as of 30 June 2018. Acquisitions after 30 June 2018 will be considered under their preacquisition identity, if appropriate, and represented by a separate dot until publication of the following year's Magic Quadrant.

- Support for the included DMSA products had to be available from the vendor. We also considered products from vendors that control, or contribute specific technology components to, the engineering of open-source DBMSs and their support.

- We included in our assessments the capability of vendors to coordinate data management and processing from additional sources beyond the evaluated DMSA. However, vendors in this Critical Capabilities research need to offer significant value-added capabilities beyond simply providing an interface to data stored in other sources.

- Vendors must provide support for at least one of the four major use cases.

- We considered depth of processing capabilities and variety of analytical processing options (relational and nonrelational) as advantageous in the evaluation criteria.

- Vendors participating in the DMSA market had to demonstrate their ability to deliver the necessary services to support a data warehouse through the establishment and delivery of support processes, professional services, and/or committed resources and budget.

- Products that exclusively support an integrated front-end tool that reads only from the paired data management system did not qualify for assessment.

We also considered the following capabilities when deciding whether products were eligible for inclusion:

- Relational DBMS

- Nonrelational DBMS

- Hadoop distributions

  (No specific rating advantage was given with regard to the type of data store used — for example, relational DBMS, graph DBMS, HDFS, key-value DBMS, document DBMS, wide-column DBMS.)

- Cloud solutions (considered viable alternatives to on-premises solutions)

- Open-source solutions

Gartner may include, at its discretion, additional vendors in cases of known use for classified but unspecified cases.

The following technology categories are specifically excluded:

- Analytical and BI solutions that only offer a DMSA that is embedded or that embeds a DMSA from another provider

- Analytical and BI solutions that only offer a DMSA that is limited specifically to the vendor's own analytical and BI solution, or whose customers exhibit only using the solution within the same vendor stack

- In-memory data grids

- Query service engines

- Prerelational DBMS

- Object-oriented DBMS

Note: Gartner analysts are the sole arbiters of which vendors and products are included in this Critical Capabilities research.

Table 1. Weighting for Critical Capabilities in Use Cases

Critical Capabilities	Traditional Data Warehouse	Real-Time Data Warehouse	Logical Data Warehouse	Context-Independent Data Warehouse
Access to Multiple Data Sources	5%	0%	30%	10%
Administration and Management	20%	20%	10%	10%
Advanced Analytics	0%	5%	5%	15%
Data Ingest	5%	20%	5%	5%
Managing Large Volumes of Data	10%	5%	5%	10%
Optimized Performance (Traditional)	15%	15%	5%	0%
Optimized Performance (Exploratory)	0%	0%	5%	10%
Flexible Scalability	5%	5%	5%	5%
Variety of Data Types	5%	5%	10%	15%
Workload Management	15%	5%	10%	5%
Use Support (Traditional)	20%	20%	5%	0%
Use Support (Exploratory)	0%	0%	5%	15%
Total	**100%**	**100%**	**100%**	**100%**
				As of January 2019

Source: Gartner (March 2019)

This methodology requires analysts to identify the critical capabilities for a class of products/services. Each capability is then weighed in terms of its relative importance for specific product/service use cases.

Critical Capabilities Rating

Each of the products/services has been evaluated on the critical capabilities on a scale of 1 to 5; a score of 1 = Poor (most or all defined requirements are not achieved), while 5 = Outstanding (significantly exceeds requirements).

Table 2. Product/Service Rating on Critical Capabilities

Critical Capabilities	Alibaba Cloud (MaxCompute)	Amazon Web Services (Amazon Redshift)	Arm Treasure Data	Cloudera (Cloudera Enterprise)	GBase (GBase 8a)	Google (BigQuery)	Hortonworks (Hortonworks Data Platform)	Huawei (FusionInsight Big Data)	IBM (Db2)	MapR Technologies (MapR Data Platform)	MarkLogic	Micro Focus (Vertica)	Microsoft (Azure SQL Data Warehouse)	Neo4j	Oracle (Oracle Exadata)	Pivotal (Pivotal Greenplum)	SAP (SAP HANA)	Snowflake	Teradata
Access to Multiple Data Sources	3.6	3.0	3.7	3.2	3.3	2.6	3.3	3.2	3.4	3.4	3.9	3.5	3.8	2.2	4.4	3.4	4.0	2.8	4.4
Administration and Management	3.4	3.3	3.0	2.9	3.3	3.3	3.0	3.3	3.7	2.8	3.3	3.4	3.3	3.9	3.7	3.4	3.2	3.6	3.8
Advanced Analytics	2.5	3.1	2.6	4.2	2.1	4.5	4.5	1.8	3.5	3.8	2.9	3.6	3.1	2.6	2.9	3.4	4.0	3.2	4.8
Data Ingest	3.3	1.8	2.8	3.1	1.9	4.3	2.9	2.9	3.2	3.5	4.5	2.5	2.3	4.2	2.3	2.3	3.9	2.3	3.1
Managing Large Volumes of Data	2.8	1.9	1.4	3.8	2.7	3.7	4.3	2.8	2.4	3.0	1.7	2.9	1.9	1.0	2.5	3.1	2.0	2.9	3.3
Optimized Performance (Traditional)	2.7	3.0	2.1	2.1	3.1	3.1	2.2	2.6	3.3	2.8	2.5	3.3	2.9	3.4	3.8	3.4	3.5	3.1	3.8
Optimized Performance (Exploratory)	3.4	3.6	3.9	4.1	3.6	4.0	4.1	3.8	4.4	4.1	4.0	4.2	4.3	4.3	4.3	4.3	4.1	4.0	4.1
Flexible Scalability	3.1	3.2	3.7	3.6	3.9	4.0	3.6	3.5	3.7	4.0	3.6	4.0	3.4	3.5	3.7	3.7	3.1	4.2	3.5
Variety of Data Types	2.3	2.5	3.6	3.5	2.7	3.0	3.5	3.6	2.8	3.5	3.4	3.1	2.8	2.6	4.0	3.4	3.6	3.4	3.7

Critical Capabilities	Alibaba Cloud (MaxCompute)	Amazon Web Services (Amazon Redshift)	Arm Treasure Data	Cloudera (Cloudera Enterprise)	GBase (GBase 8a)	Google (BigQuery)	Hortonworks (Hortonworks Data Platform)	Huawei (FusionInsight Big Data)	IBM (Db2)	MapR Technologies (MapR Data Platform)	MarkLogic	Micro Focus (Vertica)	Microsoft (Azure SQL Data Warehouse)	Neo4j	Oracle (Oracle Exadata)	Pivotal (Pivotal Greenplum)	SAP (SAP HANA)	Snowflake	Teradata
Workload Management	2.5	2.5	2.5	2.4	2.7	2.9	2.4	2.5	3.0	2.5	2.6	2.8	2.6	2.4	3.2	3.0	2.9	3.0	3.4
Use Support (Traditional)	3.6	4.8	3.3	2.4	4.2	3.2	2.1	3.3	3.2	2.7	3.2	3.6	4.3	1.9	3.9	3.6	4.1	3.3	4.1
Use Support (Exploratory)	3.8	4.2	4.1	3.7	3.2	4.1	3.4	3.2	3.2	3.6	2.6	3.9	3.9	3.5	3.9	3.9	3.3	3.5	3.2
																		As of January 2019	

Source: Gartner (March 2019)

Table 3 shows the product/service scores for each use case. The scores, which are generated by multiplying the use case weightings by the product/service ratings, summarize how well the critical capabilities are met for each use case.

Table 3. Product Score in Use Cases

Use Cases	Alibaba Cloud (MaxCompute)	Amazon Web Services (Amazon Redshift)	Arm Treasure Data	Cloudera (Cloudera Enterprise)	GBase (GBase 8a)	Google (BigQuery)	Hortonworks (Hortonworks Data Platform)	Huawei (FusionInsight Big Data)	IBM (Db2)	MapR Technologies (MapR Data Platform)	MarkLogic	Micro Focus (Vertica)	Microsoft (Azure SQL Data Warehouse)	Neo4j	Oracle (Oracle Exadata)	Pivotal (Pivotal Greenplum)	SAP (SAP HANA)	Snowflake	Teradata
Traditional Data Warehouse	3.08	3.16	2.78	2.79	3.23	3.27	2.81	3.03	3.22	2.92	3.01	3.26	3.15	2.76	3.54	3.31	3.35	3.22	3.73
Real-Time Data Warehouse	3.13	3.09	2.83	2.87	3.05	3.53	2.85	3.00	3.29	3.06	3.29	3.22	3.11	3.12	3.37	3.20	3.55	3.14	3.71
Logical Data Warehouse	3.16	3.01	3.22	3.19	3.10	3.25	3.24	3.10	3.32	3.28	3.35	3.38	3.32	2.77	3.78	3.39	3.57	3.17	3.91
Context-Independent Data Warehouse	3.06	3.03	3.20	3.57	2.92	3.66	3.63	3.05	3.31	3.47	3.16	3.46	3.22	2.95	3.57	3.48	3.46	3.32	3.82
																		As of January 2019	

Source: Gartner (March 2019)

To determine an overall score for each product/service in the use cases, multiply the ratings in Table 2 by the weightings shown in Table 1.

Acronym Key and Glossary Terms

DBMS	database management system
HDFS	Hadoop Distributed File System
ML	machine learning

Gartner Recommended Reading

Some documents may not be available as part of your current Gartner subscription.

"How Products and Services Are Evaluated in Gartner Critical Capabilities"

"Magic Quadrant for Data Management Solutions for Analytics"

Evidence

Our analysis is based on information gathered from interactions with Gartner clients during the 12 months to October 2018, and our survey of the vendors' reference customers (see below).

We also took account of:

- Earlier information and any news about vendors' products, customers and finances that came to light during the time frame for our analysis.

- Information gathered on Alibaba Cloud from the following references:

    - "Alibaba Pulls Back in U.S. Amid Trump Crackdown on Chinese Investment," Bloomberg.

    - "Alibaba Puts the Brakes on U.S. Cloud Expansion," The Information.

- The findings in "Market Share: Enterprise Infrastructure Software, Worldwide, 2017."

Survey of Vendors' Reference Customers

As part of the Magic Quadrant research process, we sought the views of vendors' reference customers (details of whom were supplied by the vendors) via a 35- to 40-minute online survey conducted during September and October 2018. The survey included requests for feedback about:

- *Vendors' product capabilities* — For example, support for large datasets, high-concurrency workloads, analytics capabilities, LDW support, data ingest rates and problems encountered with the products.

- *Vendors' maturity* — For example, support for defined DMSA use cases, ability to support customers, account management, overall perception of customers for experience of doing business with the vendor, pricing, ease of deployment and technical support.

A total of 601 references from 23 vendors completed the survey. More than 540 organizations, representing all the featured vendors' customers, responded to the survey with an average of 26 respondents per vendor. The breakdown of deployments by geography was:

- Asia/Pacific — 41%

- Europe, Middle East and Africa — 27%

- Latin America — 5%

- North America — 49%

Note that the geographic breakdown above does not sum to 100%, because some deployments took place in multiple geographic regions.

The respondents were generally pleased with their vendors and products, but gave relatively low marks in some areas, which we detail in the analysis of each vendor. Some low scores might reflect historical problems, because not all organizations are on the latest product versions.

Gartner's Client Inquiry Service Data

Gartner maintains an extensive database of information about all inquiries to our client inquiry service. Our data management team received more than 4,400 inquiries from end-user clients during the Magic Quadrant research period of November 2017 through October 2018. We used the sentiments apparent from these inquiries to assist in formulating the opinions expressed in this Critical Capabilities.

Note 1 Vertical Industry Sectors

- Accommodation and food services

- Administrative, support, waste management and remediation services

- Agriculture, forestry, fishing and hunting

- Arts, entertainment and recreation

- Construction

- Educational services

- Finance and insurance

- Healthcare and social assistance

- Information

- Management of companies and enterprises

- Manufacturing

- Mining

- Professional, scientific and technical services

- Public administration

- Real estate rental and leasing

- Retail trade

- Transportation and warehousing

- Utilities

- Wholesale trade

Note 2 Geographic Regions

- North America (Canada and the U.S.)

- Latin America (including Mexico)

- Europe (Western and Eastern Europe)

- The Middle East and Africa (including North Africa)

- Asia/Pacific (including Japan)

Critical Capabilities Methodology

This methodology requires analysts to identify the critical capabilities for a class of products or services. Each capability is then weighted in terms of its relative importance for specific product or service use cases. Next, products/services are rated in terms of how well they achieve each of the critical capabilities. A score that summarizes how well they meet the critical capabilities for each use case is then calculated for each product/service.

"Critical capabilities" are attributes that differentiate products/services in a class in terms of their quality and performance. Gartner recommends that users consider the set of critical capabilities as some of the most important criteria for acquisition decisions.

In defining the product/service category for evaluation, the analyst first identifies the leading uses for the products/services in this market. What needs are end-users looking to fulfill, when considering products/services in this market? Use cases should match common client deployment scenarios. These distinct client scenarios define the Use Cases.

The analyst then identifies the critical capabilities. These capabilities are generalized groups of features commonly required by this class of products/services. Each capability is assigned a level of importance in fulfilling that particular need; some sets of features are more important than others, depending on the use case being evaluated.

Each vendor's product or service is evaluated in terms of how well it delivers each capability, on a five-point scale. These ratings are displayed side-by-side for all vendors, allowing easy comparisons between the different sets of features.

Ratings and summary scores range from 1.0 to 5.0:

1 = Poor or Absent: most or all defined requirements for a capability are not achieved

2 = Fair: some requirements are not achieved

3 = Good: meets requirements

4 = Excellent: meets or exceeds some requirements

5 = Outstanding: significantly exceeds requirements

To determine an overall score for each product in the use cases, the product ratings are multiplied by the weightings to come up with the product score in use cases.

The critical capabilities Gartner has selected do not represent all capabilities for any product; therefore, may not represent those most important for a specific use situation or business objective. Clients should use a critical capabilities analysis as one of several sources of input about a product before making a product/service decision.

GARTNER HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit http://www.gartner.com/technology/about.jsp