Gartner. Licensed for Distribution

Critical Capabilities for Cloud Database Management Systems for Analytical Use Cases

Published 24 November 2020 - ID G00468198 - 41 min read

By Analysts Henry Cook, Donald Feinberg, Merv Adrian, Rick Greenwald, Adam Ronthal

As the move to cloud continues apace, data and analytics leaders can now choose from a wide variety of cloud-based DBMSs for analytical use cases. This research, one of two Critical Capabilities reports providing context for the Magic Quadrant for Cloud DBMS, will help in their evaluation.

Overview

Key Findings

- **There is a rich variety of cloud database management systems (DBMS) for analytical use cases from which to choose.** These include cloud-native DBMS and existing DBMS products that have moved to the cloud. Offerings include relational and nonrelational architectures and open source and proprietary code. This is a good thing, as buyers are more likely to find a solution that fits their needs. On the flip side, it can cause confusion.
- **Organizations can still choose between best-fit and best-of-breed solutions.** Also, cloud analytic ecosystems with cloud DBMS components are becoming available, which provide the integration of a best-of-breed solution by combining multiple best-fit solutions. This can make life easier, but can increase lock-in too.
- **The capability to work across intercloud, multicloud and hybrid is increasingly important.** This will likely become a prerequisite for these systems.

Recommendations

For data and analytics leaders responsible for choosing data management solutions:

- **Choose a product based on the likely longer-term fit to your environment.** This is a fast-moving market. Selecting a product because it has a particular hot feature now can be very shortsighted. While always the case, this is particularly true of this market now. For example, specialist, single-purpose DBMSs may be superseded or subsumed by other, more general-purpose offerings.
- **Identify clearly the processing and data styles you need to support.** Cloud DBMSs for analytical use cases can differ a great deal in the data formats, languages and processing techniques they support. Do not assume that they are interchangeable.
- **Assess the dynamic between your strategic cloud service providers and DBMS providers.** These decisions can be independent — choice of platform need not dictate choice of DBMS. Base your choice on fit to your required use cases.

Strategic Planning Assumptions

By 2022, 75% of all databases will be deployed or migrated to a cloud platform, with only 5% ever considered for repatriation to on-premises.

By 2023, cloud preference for data management will reduce the vendor landscape while the growth in multicloud increases the complexity for data governance and integration.

By 2023, cloud DBMS revenue will account for 50% of the total DBMS market revenue.

What You Need to Know

Data and analytics leaders can use this research to guide evaluation and initial vendor selection for cloud DBMS offerings for analytical use cases. This document replaces Critical Capabilities for Operational Database Management Systems and is one of a family of three documents that should be considered together:

1. Magic Quadrant for Cloud Database Management Systems. This research evaluates selected vendors of DBMSs that run in the cloud — for both analytical and operational use cases. The Magic Quadrant is used to judge the suitability of cloud DBMS vendors for either analytical or operational use, or for both.
2. **Critical Capabilities for Cloud Database Management Systems for Analytical Use Cases** (this document). This evaluates particular cloud DBMS products provided by the vendors in the Magic Quadrant for their suitability to support four analytical use cases, using 13 core capabilities. The findings from this document feed into the evaluations of the cloud DBMS vendors in the Magic Quadrant.
3. Critical Capabilities for Cloud Database Management Systems for Operational Use Cases. This research evaluates particular cloud DBMS products provided by the vendors in the Magic Quadrant for their suitability to support four operational use cases, using 13 core capabilities.

Most of the capabilities are common to the two Critical Capabilities documents, but may be interpreted differently for the analytical and operational use cases. The scores for each capability may also carry different weights in each document. For each vendor in the Magic Quadrant, there may be the same, or two different, products in each of the Critical Capabilities documents. The Critical Capabilities research evaluates individual products that each vendor has chosen.

Previous Gartner research has documented the move of DBMSs to the cloud — the majority of the growth in the DBMS market is taking place in the cloud (see The Future of the DBMS Market Is Cloud).

Enterprises use cloud DBMSs for analytical use cases to provide support for the following capabilities:

- **Tactical and ad hoc analysis of structured information.** The information used by mainstream business processes — such as finance, HR, supply chain, customer relationship management, supplier relationship management — requires the analysis and provision of results to a large number and a wide range of user types. This is typically the domain of the traditional data warehouse.
- **Data science and other advanced analytics.** Artificial intelligence (AI), machine learning (ML) and advanced statistics require the provision of reliable data of known quality. Algorithms may run within the cloud DBMS, or interface to a separate service. Cloud DBMSs can also be used for data engineering, which is a precursor to the analysis.
- **Multiple data types.** It is now common for analysis to be performed on a wide variety of data other than structured information. The Internet of Things (IoT), social media, video, audio, documents and weblogs are commonly used to provide a fuller analytical picture.
- **Federated and virtualized data.** The Gartner concept of the logical data warehouse (LDW) is now firmly established in architectural best practice for analytics (see The Practical Logical Data Warehouse: A Strategic Plan for a Modern Data Management Solution for Analytics). It is now common to find products with preintegration between the data warehouse and data lake, and with access to many remote sources.
- **Real-time or near-real-time analysis.** The need for analytical results in real time or near real time to provide information to mainstream business processes is now the norm. This may be real-time querying of analytical data stores, the mixing of real-time and offline analytics, real-time data ingestion, real-time analytics on an event stream, or feeding real-time data through APIs.
- **Intercloud, multicloud and hybrid operations.** It is increasingly a requirement that analytics be performed on more than one type of cloud, to have collaboration between clouds, to span clouds or to alternate analysis between the cloud and on-premises solutions.

The following trends are also appearing in the marketplace:

- **The emergence of "analytics ecosystems" as the basis for competition.** This is where products or vendors are evaluated not on single product solutions, but on an ensemble of solutions that are integrated to work together. Integration of the data warehouse, data lake and ML services are a few typical scenarios. This research evaluates particular products.
- **The use of a single database product for both transactional and analytical use cases, rather than separate products.** In some cases, the same product may be used but it may be configured differently for the two types of use cases.

How to Use This Research

Data and analytics leaders should use this research to learn how the evaluated cloud DBMS solutions support the 13 capabilities defined by Gartner as relevant to the analytical use cases. They should then consider how those capabilities, in turn, support the four analytical use cases. The interactive version of this document can be used to customize the weighting of the scores. All the while, the organization should take into account the weighting of these factors that reflects what is important to them — the interactive version allows you to vary the weightings to reflect your requirements.

While all the products evaluated address the cloud DBMS analytical use cases, they can be very different in nature and achieve their ends in different ways. In particular, there are relational and nonrelational products, open source and proprietary, and CSP-specific and CSP-independent. Buyers should ensure that they understand the implications for their organization of the type of product that they are acquiring.

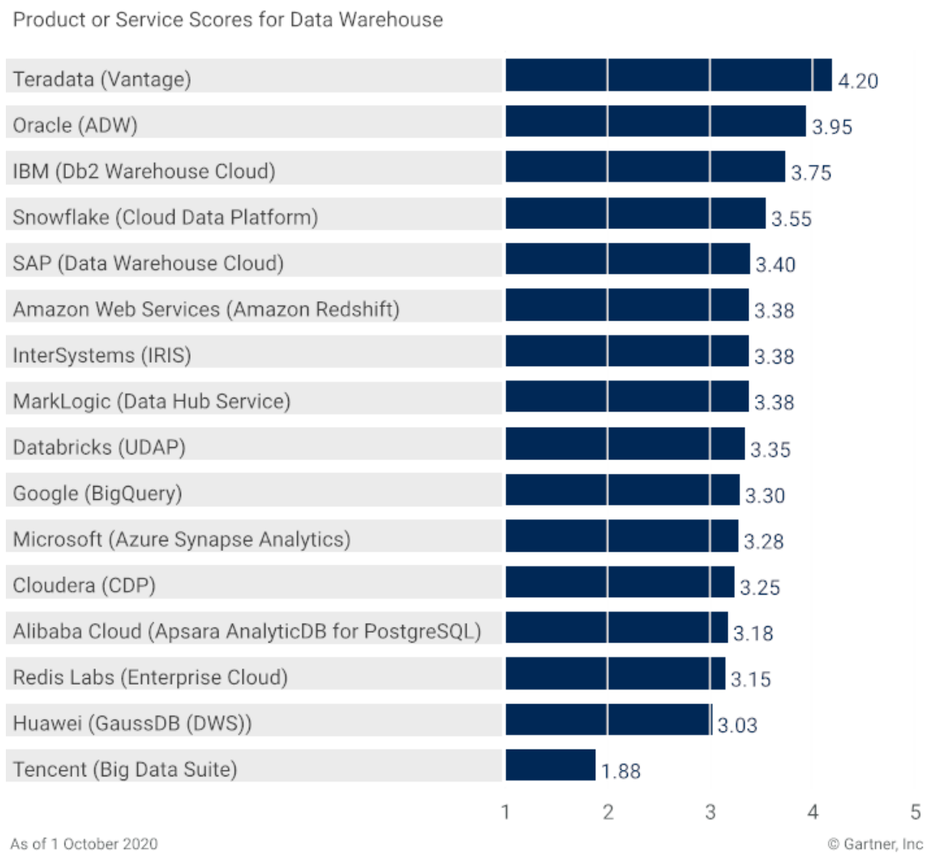
The vendor product scores reflect analyst input combined with Gartner client feedback. However, this does not provide a complete evaluation of either vendor or tool. It is essential to also consider each vendor's market presence, track record, financial and organization strength, availability of skills, product support, and outlook — such as its vision and adaptability to market changes and disruption. In that regard, the research should be used in conjunction with Magic Quadrant for Cloud Database Management Systems.

The critical capabilities assessed in this document represent a subset of the evaluation criteria that Gartner recommends when selecting vendors and tools. Therefore, the positioning of vendors in the graphics and tables do not represent overall vendor positioning in the market, and do not always coincide with positioning of vendors in the corresponding Magic Quadrant.

Analysis

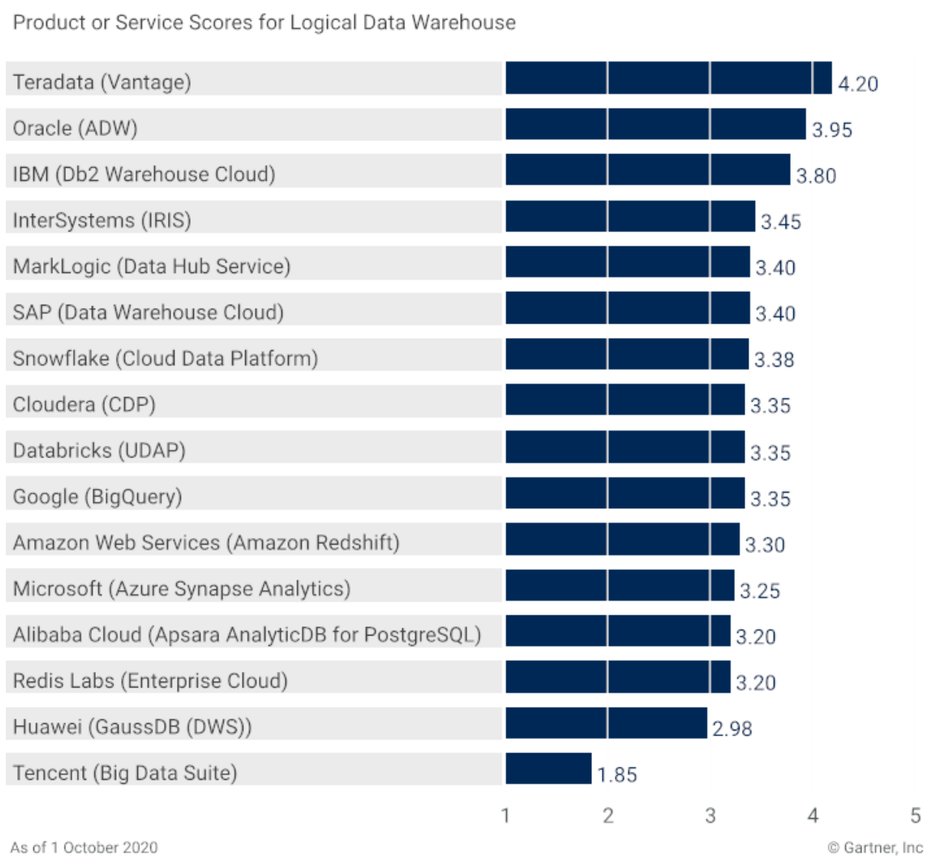
Critical Capabilities Use-Case Graphics

Vendors' Product Scores for Data Warehouse Use Case



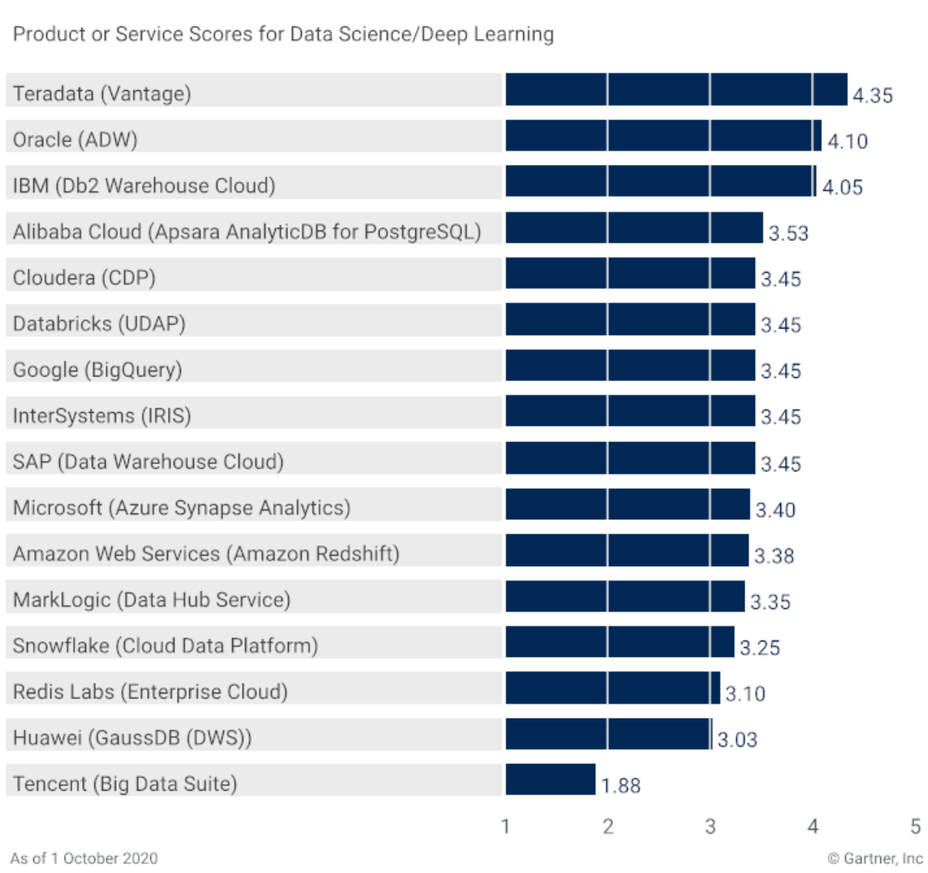
Source: Gartner (November 2020)

Vendors' Product Scores for Logical Data Warehouse Use Case



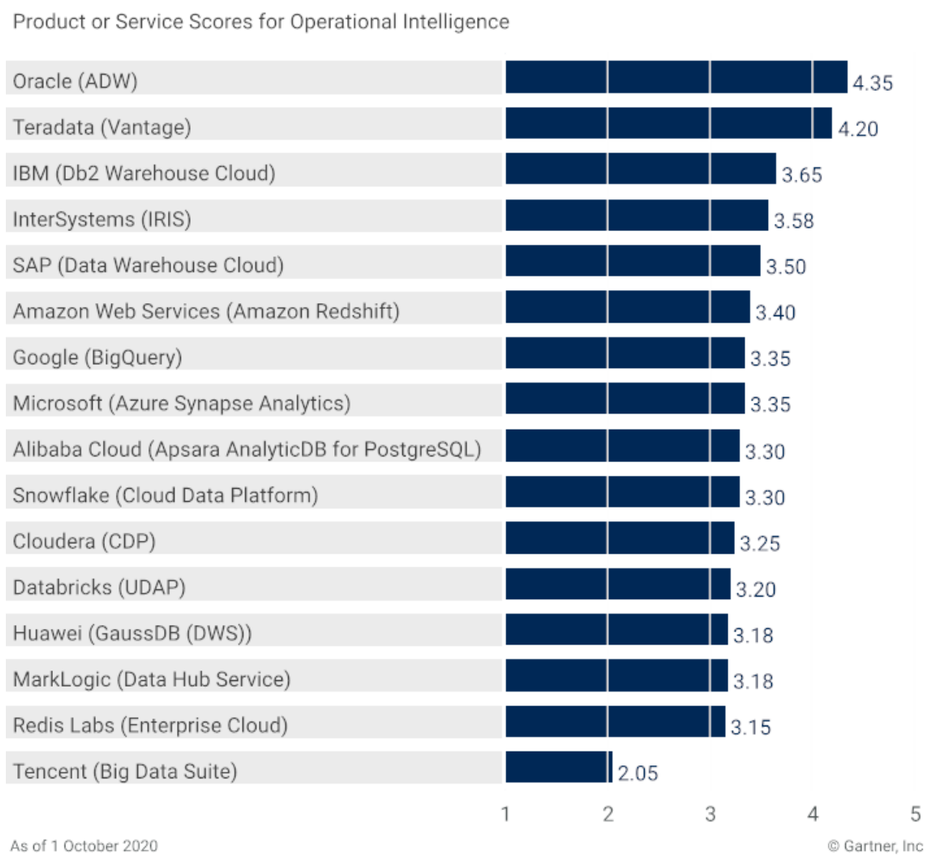
Source: Gartner (November 2020)

Vendors' Product Scores for Data Science/Deep Learning Use Case



Source: Gartner (November 2020)

Vendors' Product Scores for Operational Intelligence Use Case



Source: Gartner (November 2020)

Vendors

Alibaba Cloud (Apsara AnalyticDB for PostgreSQL)

Alibaba Cloud is a global cloud computing company headquartered in Hangzhou, China, with international operations based in Singapore and its largest investor, SoftBank, based in the Cayman Islands. Alibaba Cloud is the largest provider of public cloud platforms in China.

It offers a wide variety of analytic database services for the analytical use cases. These are: ApsaraDB relational database service (RDS) for MySQL, for SQL Server and for PostgreSQL; AnalyticDB for PostgreSQL; AnalyticDB for MySQL and AnalyticDB for online analytical processing (OLAP) analysis; MaxCompute, a large-scale serverless cloud data warehouse; E-MapReduce, an Apache Hadoop-based big data platform; and Data Lake Analytics.

The product considered here is AnalyticDB for PostgreSQL, which is based on the open-source Greenplum database. Alibaba contributes modifications to the Greenplum project.

AnalyticDB for PostgreSQL places well in all four analytical use cases, as it is based on the successful massively parallel processing (MPP) Greenplum database, which has a long history in the analytical DBMS market. The Greenplum database can be chosen because it can run on a variety of platforms as infrastructure as a service (IaaS) in most clouds, or on-premises, either as a direct install or under virtual machines. As such, it offers potential portability. AlibabaDB for PostgreSQL is a full platform as a service (PaaS) implementation.

For the data warehouse use case, AnalyticDB for PostgreSQL meets requirements, scoring in excess of 3 out of a possible score of 5, and is well placed among other cloud data warehouse offerings. While it has a large presence in the China market, it is yet to become well used in the other markets — Asia, North America and Europe — that Alibaba Cloud is expanding into.

For the logical data warehouse use case, AnalyticDB for PostgreSQL meets requirements due to its integration and data interchange with the underlying cloud object store, in conjunction with the main data warehouse. This supports the general trend that Gartner sees in the increasing integration of data warehouses and data lakes.

For the data science and deep learning category, the product places in the top half of vendor solutions evaluated, as it incorporates the MADlib library of over 300 data science algorithms. This allows data scientists to operate directly on the data held within the data warehouse without it having to be extracted, and to leverage the parallel processing AnalyticDB for PostgreSQL server.

AnalyticDB for PostgreSQL again meets requirements for the operational intelligence use case, as it has performance features that allow it to be configured to provide low response time and high throughput for this type of workload.

Amazon Web Services (Amazon Redshift)

Amazon Web Services (AWS), headquartered in Seattle, Washington, offers a range of database management services. They include: services aimed at operational use cases, such as Amazon Aurora and Amazon DynamoDB; services aimed at analytical use cases, such as Amazon Redshift and Amazon EMR; and a number of specialized services, including Amazon Neptune for graph use cases, Amazon DocumentDB for document-based use cases, Amazon Keyspaces (for Apache Cassandra) and Amazon Quantum Ledger Database (QLDB) for ledger use cases.

The product considered here is Amazon Redshift, an analytic platform built on an MPP architecture.

Amazon Redshift meets requirements for all four use cases, and places in the upper half of vendor solutions for three out of the four use cases. Its ratings between 3.30 and 3.40 indicate that it was significantly better than simply meeting requirements.

Amazon Redshift's highest use-case placement was in the data warehouse category, which demonstrates the maturity and adoption of this product in the overall market. Amazon Redshift has been in production in the cloud longer than any other product evaluated here, and its production user base is also the largest of any product evaluated.

Amazon Redshift scored relatively low for its multicloud and intercloud capabilities, which affected its position in the data warehouse and real-time data warehouse use cases. The product was rated as meets expectations or higher for all other critical capabilities.

Cloudera (CDP)

Cloudera is based in Santa Clara, California, U.S. In the cloud, it provides the Cloudera Data Platform (CDP), which includes independently available Cloudera Data Hub, Cloudera Data Warehouse and Cloudera Machine Learning cloud services focused on operational and analytic uses. Cloudera Workload Experience Manager (WXM) is used to migrate, analyze, optimize and scale workloads. Cloudera Shared Data Experience (SDX) provides intercloud and multicloud unified security, governance and metadata management. WXM and SDX are included within CDP Public Cloud services at no extra cost.

Cloudera CDP for Public Cloud was evaluated here; its private cloud versions had not shipped during the period covered.

CDP meets requirements for all four use cases, with particular strength in the data science/deep learning use case, where it was in the top five of solutions evaluated. It did well for the LDW use case, but the rest of the market has begun to catch up, often by using open-source big data technology for the data types Cloudera once led with. Cloudera's strengths in governance and hybrid deployment have not yet become top-of-mind issues.

Databricks (UDAP)

Databricks is headquartered in San Francisco, California, U.S. It provides the Databricks Unified Data Analytics Platform (UDAP) on Microsoft Azure (Azure Databricks), AWS and Alibaba. Gartner expects availability on Google Cloud Platform (GCP) in the near future. Databricks' operations are primarily focused in North America, Europe and APAC, though it does have a global presence. Its customers span a full range of industries and enterprise sizes. Although traditionally associated with Apache Spark and its data science focus, Databricks has invested heavily in the Delta Engine technology that powers its own MPP engine, and indexing capabilities targeted at bringing data warehousing performance to semantically flexible data stores.

UDAP meets requirements for all four use cases. Not surprisingly, given its foundational work on Apache Spark and MLflow, its strongest showing was in the data science/deep learning use case, where it was positioned in the top third of all products evaluated.

The work that Databricks has done on Delta Engine to support relational DBMS workloads of the sort typically seen in data warehousing also yielded positive results — UDAP received a healthy score of 3.35 for this use case.

Notably, the strength in data warehousing also extended to LDW. UDAP's built-in data virtualization capabilities can query a wide range of sources including relational, nonrelational (NoSQL), object stores, files and Hadoop Distributed File System (HDFS).

UDAP's lowest use-case placing was for operational intelligence. This use case generally requires a robust foundation in traditional transaction processing, which is not Databricks' strength, given its historical focus on analytical and data science use cases.

Google (BigQuery)

Google Cloud Platform (GCP), a subsidiary of Google, is located in Santa Clara, California, U.S. GCP supports many database platform as a service (dbPaaS) products, from fully managed versions of products from third-party providers to its own dbPaaS products. These include Google Cloud SQL (PostgreSQL, MySQL and SQL Server), Cloud Spanner, Cloud Bigtable, BigQuery, Dataflow, Dataproc, Cloud Firestore, and Firebase Realtime Database.

GCP has global reach in all regions of the world and customer presence in all vertical markets.

Google BigQuery is the dbPaaS product evaluated here.

Google BigQuery meets requirements in all four use cases, and places in the middle of all products evaluated. It received scores of 3 or 4 for all capabilities, making it a good choice for all four use cases.

Google BigQuery scored high for advanced analytics, high-speed ingest and data distribution, making it a good choice when the project requires large amounts of data. Also, Google BigQuery is the first dbPaaS to use "per-second pricing," enabling financial governance, especially for short-running queries.

Huawei (GaussDB (DWS))

Huawei Cloud is headquartered in Shenzhen, China. It provides GaussDB in a range of targeted offerings. GaussDB Data Warehouse Service (DWS) is evaluated here. All GaussDB offerings are available on Huawei Cloud and Huawei Cloud Stack for on-premises deployment. Huawei Cloud's operations are primarily focused in the Asia/Pacific region, with secondary strength in the Middle East, Africa and Latin America. The public administration, telecom, and finance and insurance sectors account for more than half of

Huawei's industry penetration.

GaussDB (DWS) is a fully PostgreSQL-compatible cloud data warehouse offering that runs on Huawei Cloud infrastructure. It uses a node-based MPP architecture that is designed to handle peta-scale data volume. The offering meets requirements for three out of the four evaluated use cases, falling just short in the LDW use case due to fairly rudimentary capabilities to connect to a wide variety of external data sources (only Huawei Cloud's MapReduce Service and Oracle Database are specifically listed).

For the other three use cases — data warehouse, data science/deep learning and operational intelligence — GaussDB (DWS) is a capable offering. It supports PostgreSQL ecosystem plug-ins — most notably PostGIS for geospatial analysis — and scored well for both high-speed, high-volume processing, and high-speed ingest. As a node-based offering, GaussDB (DWS) provides solid financial governance capabilities that allow for cost predictability, though it may lack some of the flexibility of more modern serverless architectures.

IBM (Db2 Warehouse Cloud)

IBM, based in Armonk, New York, U.S., offers IBM Db2 on Cloud, IBM Db2 Warehouse on Cloud, IBM Cloud SQL Query, IBM Cloudant and the IBM Cloud Databases family. There are also managed services for PostgreSQL, MongoDB, Elasticsearch, Redis, RabbitMQ, DataStax, EDB and etcd. IBM Cloud Object Storage serves as a landing zone and clearinghouse to complete IBM's offerings for operational and analytical use cases, and stream processing with IBM Event Streams.

IBM Cloud Pak for Data, a Red Hat OpenShift integration layer for containerized common software services, supports multicloud environments such as AWS, Microsoft Azure, GCP, IBM Cloud and private cloud deployments. Most offerings are also available on-premises. IBM operations are global, with significant customer penetration in all industries.

Db2 Warehouse on Cloud is the product evaluated here. It placed in the top three of vendor solutions for every use case in this research. Highlights include its alignment with IBM's push into container-based deployment with Red Hat, its leading ratings for advanced analytics and its mature strength in data security.

InterSystems (IRIS)

InterSystems is headquartered in Cambridge, Massachusetts, U.S. It markets IRIS (originally the Caché DBMS). InterSystems has presence across the globe in all regions, primarily in healthcare, but also in other industries like financial services.

IRIS is available as a private, fully managed dbPaaS cloud service today — multicloud on AWS and Microsoft Azure — with the fully managed dbPaaS coming in 2020. The IRIS Enterprise version has additional ML/AI capabilities.

InterSystems IRIS meets requirements in all four use cases and places in the top half of products evaluated, with two top-five placings. This makes it a good choice for all the analytical use cases. Of particular note is its fourth place position for operational intelligence, due to its strength in embedded analytics, ML and AI.

IRIS received only one critical capability rating below "meets requirements" — for dynamic elasticity — a feature not yet implemented in the product. Highlights among its many other capability scores of 4 ("meets or exceeds some requirements") include the product's support for advanced analytics and workload management. Of course, it exceeds requirements for security, which should be expected due to its strength in healthcare.

MarkLogic (Data Hub Service)

MarkLogic has headquarters in San Carlos, California, U.S. It offers the MarkLogic Data Hub Platform, which is delivered on the cloud as the Data Hub Service and is available on the Microsoft Azure and AWS clouds.

MarkLogic focuses on a combination of data management, built around a transactional document store, and an integration hub. The integration hub allows users to access data stored remotely through a universal index, which enables reduced remote data movement through optimization of remote access.

The MarkLogic Data Hub Service did well in this research cycle. Its scores were between 3.35 and 3.40 for three out of the four use cases, significantly better than simply meeting requirements with a 3.0 rating.

The product scored a 3 or higher for every critical capability evaluated. Highlights include financial governance, high-speed ingest, multiple data types/structures and multi/intercloud/hybrid deployment, capabilities for which MarkLogic Data Hub Service scored a 4.

Microsoft (Azure Synapse Analytics)

Microsoft is headquartered in Redmond, Washington, U.S. It provides a broad range of cloud DBMS offerings, including Azure Synapse Analytics, which is evaluated here. Microsoft's operations are geographically diversified, and its customers are spread across a wide range of industries and deployment sizes. Microsoft is focused on delivering a cohesive cloud data management ecosystem that spans all the use cases we have defined for this market, and Azure Synapse is a key component of that strategy.

Azure Synapse meets requirements for all four use cases. Its relatively "middle of the pack" positioning reflects three key factors. First, Azure Synapse is still emerging as a cohesive and comprehensive offering, and Critical Capabilities research — by definition — focuses on product capabilities that are available today. Second, the most mature aspect of Azure Synapse came from its predecessor, Azure SQL Data Warehouse, which is a capable cloud data warehouse offering. However, it does not yet have the track record in scalability and performance of some of its competitors.

Relatively low scores for multi/intercloud/hybrid deployment and automated performance tuning lowered Azure Synapse's position across all evaluated use cases. This was balanced by relatively high scores in dynamic elasticity, high-speed ingest, security, consistency and automated distributed data scaling.

The vision behind Azure Synapse is market leading, and that is reflected in the positioning of Microsoft in the companion Magic Quadrant for Cloud DBMS. Microsoft has a strong foundation to build on in Synapse, but the full scope of this vision has not yet been realized.

Oracle (ADW)

Oracle is based in Redwood Shores, California, U.S. The product analyzed as part of this research is the Oracle Autonomous Data Warehouse (ADW), available in the Oracle Cloud Infrastructure (OCI) and on the Oracle Exadata Cloud@Customer (ExaCC) private cloud. In addition to ADW, the company offers Oracle Autonomous Transaction Processing (ATP), Autonomous JSON Database, Oracle MySQL Database Service, Oracle NoSQL Database and Oracle Big Data Service.

ADW is now over 2 years old and recently available on ExaCC. It has full compatibility with the Oracle Database, with added functionality to manage the database automatically (for example, automated security patching, upgrades and other patches with zero downtime, automated index maintenance, and enhanced optimization using ML).

Oracle ADW placed first for operational intelligence and second for all other use cases, with high scores in all capabilities, making it a good choice for any analytical use case. Although Oracle is a relative newcomer in terms of offering a fully managed service, ADW's high evaluation is due to it being based on the same mature DBMS that has been used on-premises for over 40 years. Reports from Gartner clients verify that Oracle's promise of low cost is proven, although we do continue to recommend a proof of concept to verify the cost profile of different use-case workloads.

Oracle's only low capabilities score (2 out of 5) was for multicloud/intercloud/hybrid deployment. Although ADW has a strong hybrid capability, Oracle completely lacks a multicloud and intercloud vision, with a strategy forcing customers to the vendor's cloud for dbPaaS. Oracle has an agreement with Microsoft for colocation of data centers, allowing the use of Oracle ADW on OCI with tools and applications on Azure, but has no plans to allow ADW on other cloud service providers (CSPs).

Redis Labs (Enterprise Cloud)

Redis Labs is headquartered in Mountain View, California, U.S. It offers the Redis Enterprise Cloud on all three major CSPs (AWS, GCP and Microsoft Azure), and Redis Enterprise Software on-premises. The Redis Labs product is an in-memory data store, which has primarily been used as a cache but is expanding into more traditional DBMS use cases. Redis Enterprise Cloud was the product evaluated in this research.

Redis has normally been used for an in-memory data accelerator. Redis Enterprise Cloud meets requirements for all four use cases. However, it placed in the bottom half of solutions across all four, which is somewhat expected based on its traditional operational usage.

Redis Enterprise Cloud's relatively low score for automatic optimization was in part due to the need for developers to explicitly use this offering, where the in-memory performance characteristics help the most.

SAP (Data Warehouse Cloud)

SAP, which is based in Walldorf, Germany, offers SAP BW/4HANA, a packaged data warehouse solution. It also offers: SAP HANA Cloud, Data Lake; SAP Data Warehouse Cloud; and SAP Data Intelligence, which includes SAP Vora for Apache Spark and Hadoop processing. SAP Data Warehouse Cloud is the subject of this research.

SAP Data Warehouse Cloud is well placed for the data warehouse use case. While it is relatively new as a service, it builds on the track record of SAP HANA when used for data warehousing. Although SAP Data Warehouse Cloud may not be as widely used as other vendors' solutions that have been in the cloud analytical DBMS marketplace for many years, it has been well received by SAP customers. It provides innovation in development through the use of what are termed "spaces," which are dynamic sandboxes that can be created within the data warehouse, making data and analytic processes available to users in an easy-to-use, but well-governed, manner. This has been specifically designed to allow business and IT users to collaborate and to support each other with their respective skills.

SAP is also well placed to support the LDW use case, as it provides several important features that allow collaboration between multiple analytic systems. The underlying SAP HANA Cloud system can connect with a very wide range of SAP and non-SAP data sources, data virtualization is included as a native capability, and the system can also make metadata available for external tools to be able to query, export and load data.

For the data science and deep learning use case, SAP Data Warehouse Cloud is rated well because it can provide a range of different data science capabilities. There are built-in libraries of data science algorithms and the system provides interfaces to a wide variety of analytical and data visualization tools.

SAP Data Warehouse Cloud inherits the transactional performance of its underlying SAP HANA Cloud database instances. This allows a mix of operational transactionlike querying on the same underlying copy of the data used for large-scale analytical processing, leveraging the special technology of SAP HANA that allows a single in-memory copy of the data to be used for both purposes.

Snowflake (Cloud Data Platform)

Snowflake is headquartered in San Mateo, California, U.S. Its Snowflake Cloud Data Platform, available on AWS, GCP and Microsoft Azure, is mainly focused on delivering production-optimized delivery of semantically consistent data in a data-warehouse-as-a-service model. Snowflake Cloud Data Platform also supports relational data lakes, document-style data (for example, JSON), data sharing and a private cloud offering called Virtual Private Snowflake.

Snowflake's operations are mostly focused in North America, but the company has a growing presence in EMEA and Asia/Pacifc and its customers are spread across a wide range of industries and deployment sizes. Snowflake has expanded its focus to extend beyond core data warehousing and has articulated a vision for a broader data platform with a multicloud focus.

Snowflake's traditional area of strength is the data warehouse use case, for which it received its highest score and placing. Its overall "middle of the pack" positioning reflects the highly competitive nature of this market. Less than three-tenths of a point separate the fourth- through tenth-placed solutions. Snowflake finds itself in good company here, scoring better than all three of its ubiquitous competitors: Amazon Redshift, Microsoft Azure Synapse Analytics and Google BigQuery.

While also meeting requirements for the other three evaluated use cases, Snowflake's positioning was hurt by relatively low scores in a number of key capabilities. Specifically, the lack of prebuilt, in-database advanced analytic algorithms hurt the vendor in the data science/deep learning use case. Snowflake does have user-defined function support for a number of languages, but primarily relies on ANSI SQL capabilities and partner integrations to address this use case.

Additionally, relatively low scores for financial governance and workload management — both of which are high to moderately weighted in most use cases — affected Snowflake's overall scores.

Finally, operational intelligence requires a strong foundation in traditional transaction processing, which is not Snowflake's core strength.

Tencent (Big Data Suite)

Tencent is headquartered in Shenzhen, China. The TencentDB managed service offerings include CynosDB (a distributed RDBMS compatible with MySQL and PostgreSQL), TBase (Postgres-based with Oracle compatibility), and MySQL, MongoDB, Redis, PostgreSQL, MariaDB, and SQL Server.

Tencent Big Data Suite, evaluated here, is only available for private cloud, making its results for many of the criteria incomplete. A relatively new offering, it placed at the bottom of vendor solutions for all use cases in this report. Tencent Big Data Suite currently lacks support for core capabilities such as triggers and stored procedures, automated data definition language (DDL) and workload management tools, online patching, disaster recovery, and statistics collection for optimization. Newer features such as support for containers, geospatial, graph and time series support are also absent. Tencent Big Data Suite is likely to develop rapidly in the coming years, especially given the relatively complete feature sets in other TenCent portfolio offerings.

Teradata (Vantage)

Teradata is based in San Diego, California, U.S. It delivers data management solutions for analytics across any deployment environment. Teradata Vantage is successfully established as a cloud-based analytical system and is available on AWS, GCP and Microsoft Azure. It is also available on the Teradata Cloud (optimized infrastructure) and on-premises via the IntelliFlex appliance hardware or customer hardware through the use of virtual machines. Teradata Vantage supports SQL query with spatial and temporal support, a wide variety of ML algorithms and graph processing within the Teradata Hybrid Multi-Cloud Ecosystem, which supports the LDW use case.

Teradata QueryGrid provides multisystem query support via Teradata's own software, as well as via open-source Presto, including support for and integration with Apache Hadoop and Spark.

Teradata Vantage was the top-rated solution for the data warehouse, LDW and data science/deep learning use cases. The solution is an integrated system of analytics capabilities that combines multiple processing engines (relational, nonrelational, spatial, graph, text, temporal, data science algorithms) with multiple access methods and development languages. It provides this integrated system on multiple clouds, on-premises appliances and virtual machines. Teradata Vantage allows multiple analytical techniques to be mixed and matched within an analytical pipeline without having to extract the data to a separate ML or data science service.

Although Teradata Vantage is not a general-purpose transactional system, it does achieve a high score and placing (second) for operational intelligence. This is recognition of its ability to provide a wide variety of analytics into operational business processes, and to support large numbers of users in doing so.

Context

Analytics continues to be a top priority for many organizations. This, combined with the trend toward adoption of the cloud, has meant that the ability of cloud DBMS solutions to address analytical use cases has become very important. There is a wide variety of solutions to choose from. Existing analytics database vendors have moved to the cloud; there are offerings from each of the main CSPs, and also vendors that run on multiple clouds. Cloud DBMSs can be relational DBMSs or nonrelational. They can also vary considerably in their functional richness and maturity. This provides a good degree of choice, but can also make it difficult for data and analytics leaders to compare the offerings. In addition, the market is moving at a fast pace with new features being frequently added — and then matched by competition.

A trend is for these systems to interoperate with other analytic components. In particular, the ability to integrate data warehouses, data lakes, and artificial intelligence and machine learning services is well established. This has given rise to what are termed "analytical ecosystems" that span multiple types of analysis and that are becoming a new basis for competition.

Product/Service Class Definition

Cloud DBMSs for analytical use cases consist of the following types of system:

- Relational DBMS (RDBMS).
- Nonrelational DBMS.
- Hadoop/Spark distributions. (No specific rating advantage was given with regard to the type of data store used — for example, RDBMS, graph DBMS, HDFS, key-value DBMS, document DBMS, wide-column DBMS.)
- Open-source solutions.

Critical Capabilities Definition

As part of the research process for the companion Magic Quadrant, we relied upon ongoing briefings from the vendors selected, an RFI issued in relation to document-specific features, and ongoing interactions with Gartner clients as part of the inquiry process.

Advanced Analytics

The product's ability to perform advanced analytic operations within the dbPaaS. It is evaluated on the basis of what functionality is offered in the current version of the product and what functionality is being used by customers.

In addition, the extent and richness of available in-DBMS analytic libraries of AI/ML algorithms is taken into consideration.

Automated Distributed Data Scaling

The ability to distribute, balance and scale data across nodes as needed to conform to policy-driven performance and cost specifications.

Automated Perf Tuning/Optimization

The ability to optimize performance for queries, transactions and workloads to meet performance SLAs. This will vary according to the workload being optimized.

Consistency

DBMS-guaranteed properties of "atomicity, consistency, isolation and durability" (ACID) to ensure reliable, recoverable database transactions with multiple nodes accepting writes to the same data. Strong consistency is a requirement.

Additionally, the engine can include forms of relaxed, eventual or tunable consistency for specific use cases.

Dynamic Elasticity (Scalable Perf)

The capability to scale both up and down based on policy in response to changing workloads or user specifications to deliver predictable performance and meet SLAs when confronted with workload variability. Also used as an opportunity for cost optimization.

Different offerings can deliver this capability in different ways. The separation of compute and storage can make it easier for the cloud vendor to implement this capability.

Financial Governance

The ability to forecast, budget usage, monitor and control costs by throttling, workload reduction or other means. It can include governing types and numbers of instances used, allowed users or groups, spinning down unused resources, and recommending and implementing less costly storage strategies.

Cost predictability and blended pricing models are beneficial.

High-Speed, High-Volume Processing

The ability to perform computations on, and durably write, large volumes of data for high-speed processing. This includes data with multiple data structures and formats.

High-Speed Ingest

The capability for data to be continuously loaded, often from multiple endpoints and in different formats, and made quickly available. Minimal processing is done here. Important for event processing and real-time data warehousing, it includes forms of stream processing built into the DBMS engine.

Multi/Intercloud/Hybrid Deployment

The ability to deploy databases across on-premises and one or more cloud environments synergistically, synchronizing data transparently. Applications should be able to access data without specifying location.

Multiple Data Types/Structures

Support for data types in addition to structured data (e.g., machine data, documents, images or videos). This includes DBMS-provided functionality for query and integration methods beyond simply providing a field that "can contain anything" and leaving usage to programmers using external languages.

Performance Monitoring and Admin

This capability supports administration and management during implementation and ongoing use. Includes resource utilization, database activity monitoring, role-based activities and advisors. Recommendations may supplement automated optimization.

Security

Includes policy-based DBMS access controls (e.g., row- and column-level authorization), encryption, data masking and obfuscation, separation of concerns, and support for regulatory standards (e.g., GDPR). Integration with database activity monitoring and vulnerability scans are advanced capabilities.

Workload Management

The ability to manage different types and sizes of workloads without an excessive increase in resources; the ability to handle varying and conflicting workloads without a corresponding variance in response times; and the ability to manage the workloads to meet defined service levels.

Use Cases

Data Warehouse

Managing structured historical data from multiple sources, mainly via bulk and batch loading. Data is structured to make it flexibly available to a wide variety of use cases.

Logical Data Warehouse

Managing data variety and volume for structured and other content data types, acting as a logical tier to a variety of data sources.

Data Science/Deep Learning

Exploration of new data values form variants and relationships using search, graph, predictive modeling and other capabilities to uncover information models.

Operational Intelligence

Operational and analytic operations within a business activity in separate process spaces on the same infrastructure and the same physical database.

Vendors Added and Dropped

This is the first iteration of Critical Capabilities for Cloud Database Management Systems for Analytical Use Cases. No vendors were added or dropped, since there is no prior publication.

Inclusion Criteria

This Critical Capabilities research uses the same inclusion criteria as the companion Magic Quadrant.

To qualify for inclusion, vendors had to meet all of the following criteria.

Software availability: Vendors must have cloud DBMS software that has been generally available for licensing or supported download for at least a year, as of midnight, U.S. Eastern Daylight Time on 1 June 2020. This includes any new functionality added to the service(s) by the specified date. We do not consider beta, "early access," "technology preview," or other not generally available functionality or services.

- Any acquired product or service must have been acquired and offered by the acquiring vendor as of 1 June 2020. Acquisitions after this will be considered under their preacquisition identity, if appropriate, and represented separately until publication of the following year's Magic Quadrant and Critical Capabilities reports.

Industry presence: Vendors' cloud DBMS products must have referenceable production presence in accounts in a minimum of three of the following industry sectors:

- Accommodation and food services
- Administrative and support and 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

Use-case support: Each vendor must support at least two of the following four use cases:

- Data warehouse
- Logical data warehouse
- Data science/deep learning
- Operational intelligence

Geographic presence: Each vendor must have market presence in at least three of the following distinct geographic regions. (Regional market presence is defined as a minimum of 5% of revenue of the verified production customer base, and dedicated sales offices or distribution partnerships in a specific region):

- North America (Canada, Mexico and United States)
- Central and South America
- Europe (including Western and Eastern Europe)
- Middle East and Africa (including North Africa)
- Asia/Pacific
- Japan

Market share/revenue: Only named vendors in the DBMS market segment with cloud DBMS products are eligible for inclusion:

- Vendors' inclusion is based on verifiable cloud DBMS revenue from the calendar year 2019. Each vendor must have greater than $20 million in verifiable cloud DBMS revenue from the calendar year 2019. Cloud DBMS revenue is defined in the Gartner report, Market Share: Enterprise Platform as a Service, Worldwide, 2019.
- Only vendors in the DBMS market segment with cloud DBMS products were considered; of those, only vendors with estimated 2019 DBMS revenue that met the above criteria were included.

Software releases and feature availability: Product evaluations include technical capabilities, features and functionality present in the product or supported for download as of midnight, U.S. Eastern Daylight Time on 1 June 2020.

Customers: We drew on publicly available information, feedback from users of our client inquiry service and our industry contacts.

Support: Each vendor must provide support for its CDBMS product(s). For an open-source DBMS, maintenance and support must be available from a vendor that owns or has substantial control over the source code, and offers it under an open-source license, such as the General Public License (GPL) or Apache License. However, only the core DBMS engine needs to be under the open-source license to classify as an open-source DBMS. The products in this report are enhanced with vendor additions, but the vendors also offer and support open-source-only versions of those products.

Services: Vendors participating in the cloud DBMS market must demonstrate their ability to deliver the necessary services to support operational systems via the establishment and delivery of support processes, professional services and/or committed resources and budget.

Excluded products: Vendors marketing only products from the list below are explicitly excluded from this Magic Quadrant and Critical Capabilities research. They include:

- "Streaming" services, whose use cases are dominated by immediate event processing and which are rarely if ever used for subsequent management of the data involved.
- Pre-relational DBMS products.
- Object-oriented DBMS products.
- Data grid products.
- Complex-event processing (CEP) or streaming-data-only services.
- Analytic and BI solutions that only offer an analytical DBMS that is limited specifically to the vendor's own analytic and BI solution, or whose customers exhibit only using the solution within the same vendor stack.

- Query service engines.
- Vendors of data virtualization, data fabric and data federation that do not provide data persistence of its own.

Table 1: Weighting for Critical Capabilities in Use Cases

Critical Capabilities	*Data Warehouse*	*Logical Data Warehouse*	*Data Science/Deep Learning*	*Operational Intelligence*
Advanced Analytics	5%	10%	30%	10%
Automated Distributed Data Scaling	10%	10%	0%	10%
Automated Perf Tuning/Optimization	10%	10%	10%	10%
Consistency	5%	5%	0%	5%
Dynamic Elasticity (Scalable Perf)	15%	10%	20%	5%
Financial Governance	10%	10%	10%	10%
High-Speed, High-Volume Processing	0%	0%	0%	15%
High-Speed Ingest	5%	5%	5%	5%
Multi/Intercloud/Hybrid Deployment	10%	10%	0%	0%
Multiple Data Types/Structures	5%	10%	10%	0%
Performance Monitoring and Admin	5%	5%	5%	10%
Security	5%	5%	5%	10%
Workload Management	15%	10%	5%	10%
As of 1 October 2020				

Source: Gartner (November 2020)

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

Each of the products/services that meet our inclusion criteria has been evaluated on the critical capabilities on a scale from 1.0 to 5.0.

Critical Capabilities Rating

Table 2: Product/Service Rating on Critical Capabilities

Critical Capabilities	*Alibaba Cloud (Apsara AnalyticDB for PostgreSQL)*	*Amazon Web Services (Amazon Redshift)*	*Cloudera (CDP)*	*Databricks (UDAP)*	*IBM (Db2 Warehouse Cloud)*	*Google (BigQuery)*	*Huawei (GaussDB (DWS))*	*InterSystems (IRIS)*	*MarkLogic (Data Hub Service)*	*Microsoft (Azure Synapse Analytics)*	*Oracle (ADW)*	*Redis Labs (Enterprise Cloud)*
Advanced Analytics	4.0	3.0	4.0	4.0	4.0	2.5	5.0	4.0	3.0	3.5	4.0	3.0
Automated Distributed Data Scaling	3.0	4.0	4.0	4.0	4.0	3.0	4.0	4.0	3.0	4.0	4.0	3.0
Automated Perf Tuning/Optimization	3.0	3.0	3.0	2.0	3.0	3.0	3.0	3.0	3.0	2.5	4.5	2.0
Consistency	3.0	3.0	3.0	3.0	3.0	3.0	4.0	3.0	3.0	4.0	5.0	3.0
Dynamic Elasticity (Scalable Perf)	3.5	4.0	3.0	4.0	3.0	3.5	4.0	2.5	3.5	4.0	4.0	3.0
Financial Governance	3.0	4.0	3.0	3.0	4.0	3.5	3.0	3.0	4.0	3.0	4.0	3.0
High-Speed, High-Volume Processing	3.5	3.0	3.0	3.0	3.0	4.0	3.0	4.0	3.0	3.0	5.0	4.0
High-Speed Ingest	4.0	3.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	5.0
Multi/Intercloud/Hybrid Deployment	2.0	2.0	3.0	4.0	3.0	2.0	3.5	3.0	4.0	2.0	2.0	4.0
Multiple Data Types/Structures	3.5	3.0	4.0	3.0	3.0	3.0	4.0	4.0	4.0	3.0	4.0	4.0
Performance Monitoring and Admin	3.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	3.0
Security	3.0	3.0	3.0	3.0	3.0	2.5	4.0	4.0	3.0	4.0	5.0	3.0
Workload Management	3.5	3.5	3.0	3.0	3.0	3.0	4.0	4.0	3.0	3.0	4.0	3.0
As of 1 October 2020												

Source: Gartner (November 2020)

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 (Apsara AnalyticDB for PostgreSQL)*	*Amazon Web Services (Amazon Redshift)*	*Cloudera (CDP)*	*Databricks (UDAP)*	*IBM (Db2 Warehouse Cloud)*	*Google (BigQuery)*	*Huawei (GaussDB (DWS))*	*InterSystems (IRIS)*	*MarkLogic (Data Hub Service)*	*Microsoft (Azure Synapse Analytics)*	*Oracle (ADW)*	*Redis Labs (Enterprise Cloud)*
Data Warehouse	3.18	3.38	3.25	3.35	3.75	3.30	3.03	3.38	3.38	3.28	3.95	3.15
Logical Data Warehouse	3.20	3.30	3.35	3.35	3.80	3.35	2.98	3.45	3.40	3.25	3.95	3.20

Use Cases	Alibaba Cloud (Apsara AnalyticDB for PostgreSQL)	Amazon Web Services (Amazon Redshift)	Cloudera (CDP)	Databricks (UDAP)	IBM (Db2 Warehouse Cloud)	Google (BigQuery)	Huawei (GaussDB (DWS))	InterSystems (IRIS)	MarkLogic (Data Hub Service)	Microsoft (Azure Synapse Analytics)	Oracle (ADW)	Redis Labs (Enterprise Cloud)
Data Science/Deep Learning	3.53	3.38	3.45	3.45	4.05	3.45	3.03	3.45	3.35	3.40	4.10	3.10
Operational Intelligence	3.30	3.40	3.25	3.20	3.65	3.35	3.18	3.58	3.18	3.35	4.35	3.15

As of 1 October 2020

Source: Gartner (November 2020)

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

AI	artificial intelligence
CSP	cloud service provider
DBMS	database management system
EDW	enterprise data warehouse
LDW	logical data warehouse
ML	machine learning
MPP	massively parallel processing
ODS	operational data store
SLA	service-level agreement

Evidence

Our analysis in this Critical Capabilities research is based on information gathered from interactions with Gartner clients during the 12 months to June 2020.

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.
- The findings in:
  - Market Share: All Software Markets, Worldwide, 2019
  - Market Share: Enterprise Platform as a Service, Worldwide, 2019
  - Gartner Peer Insights

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.

© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."