OMDIA

UNIVERSE

Publication date
19 Jul 2022

Author
Bradley Shimmin
Chief Analyst

Modern Data Analytics Platform, 2022

Brought to you by Informa Tech

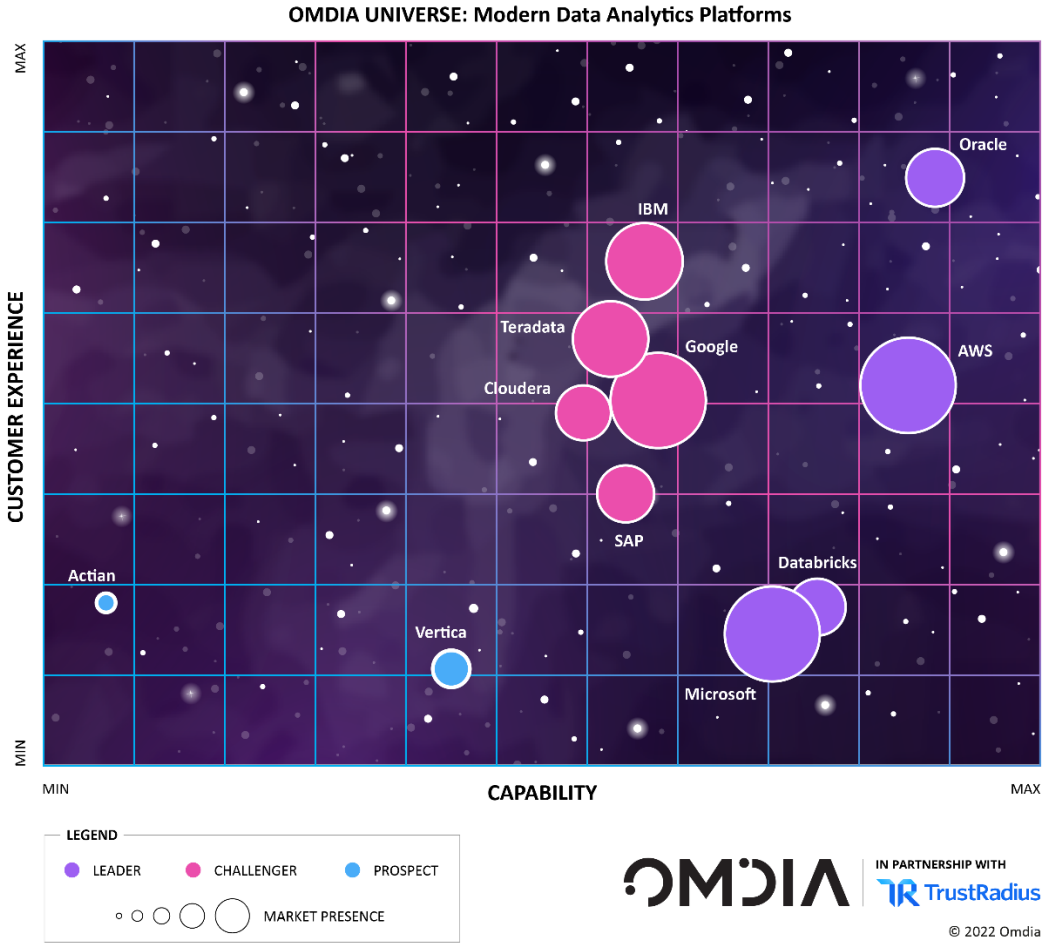
OMDIA

Summary

Catalyst

The rush to turn raw information into impactful business insight has reached a feverish pace as companies continue to fight against unprecedented market disruptions. Beating at the heart of such efforts, tried and true analytical databases are evolving into modern data analytics platforms capable of handling everything from BI dashboards to machine learning (ML)—all with one goal of keeping companies running and allowing them to move forward through innovation.

Figure 1: The Omdia Universe for a modern data analytics platform



Source: Omdia

Omdia view

For more than 40 years, the data warehouse has served the enterprise well, helping companies respond to and anticipate market changes by looking at historical, predominantly tabular data streaming from numerous operational business systems. Its efficacy has been proven repeatedly across all facets of the enterprise marketplace, with use cases ranging from basic quarterly earnings reports to near real-time credit card fraud prevention. To physically support this broad range of capabilities at scale, data warehouses have historically relied upon several highly performant architectural traits, including massively parallel processing (MPP), in-memory processing, and query optimization to deliver data-driven insights at extreme scale.

The arrival of the cloud and the concept of big data in the early 2000s upended this approach by introducing the ability to analyze data that lies beyond traditional structured rows and columns and to do so at an extreme scale. Thanks to cost-effective and flexible storage techniques (e.g., file systems running on object stores), these platforms—branded data lakes—have been instrumental in driving new market advances, particularly in areas such as customer experience management. However, while quite capable, early data lake platforms combining Hadoop, HIVE, and Presto were notoriously difficult to implement and optimize.

Since that time, the data and analytics marketplace has chased after solutions capable of unifying the two seemingly disparate worlds of tabular database queries and object-level file retrieval. Market participants with data lakes and those with more traditional data warehouses have equally sought cloud-native architectural advancements such as separating storage and compute processing to simplify deployment and optimize IT spending. This has given rise to many new cloud-based analytical database platforms capable of scaling to meet any needs while supporting one-click purchasing and provisioning.

Unfortunately, this trend has led to an even more fractured data landscape with the proliferation of departmental data warehouses that are both unmanageable and unable to deliver a single version of the truth for the entire enterprise. Adding to this concern is the continuing proliferation of data sources. The market now expects IT to look beyond the database to incorporate new data types from new, disparate sources, be those manufacturing machine logs or mobile phone geolocation histories. And for good reason. If the COVID-19 pandemic taught humanity any singular lesson, it was that carefully curated historical data is not enough to protect business operations from the vagaries of fate that live well beyond the enterprise firewall.

Today's business demands an analytics database that is more than a centralized data warehouse or departmental data lake. Today's business demands an analytics solution that can answer questions beyond the "what happened?" to reach the "what might happen and what should we do about it?" To answer this kind of question, businesses need an analytical database that can speak the language of both data lake and data warehouse, is fully cloud-native, fueled by artificial intelligence (AI) automation and augmentation, crafted to build AI outcomes, and capable of accommodating disparate data types (documents, geospatial, time-series, etc.).

Analytics databases, in this way, have an identity problem that is actually an opportunity in disguise. As they reach out to support a broad swath of analytics workloads (e.g., time series forecasting,

geospatial analysis, graph analysis, data science, etc.), they are trodding new ground beyond the confines of conventional data warehouses. For the same reason, they are no longer data lakes or even the recently bandied about the term data lakehouse. They are charting new waters as what Omdia refers to as modern data analytics platforms, a solution that goes well beyond analytical databases, transforming to fit the purpose, place, and sweep of each customer demand.

Are there such do-it-all platforms in the market today capable of fully unifying analytics across the enterprise at scale and supporting a wide array of use cases and user requirements? Yes, but also no. That ultimately depends upon how well the customer's needs align with the capabilities of a given solution. In this report, Omdia evaluates several leading data warehouse solutions rapidly evolving into this modern data analytics platform to understand how far along the evolutionary path they have come in delivering across several key capabilities' measures outlined below.

Key messages

- Data fuel all enterprise endeavors; data is housed within an array of disparate, disconnected analytical databases that are hard to manage and even more difficult to use as a source of insight.
- Modern data analytics platforms seek to address this situation by offering a unified analytics platform with a broad deployment footprint and the ability to handle a wide range of analytical use cases—all while meeting high-performance requirements at scale without subverting security, management, and total cost of ownership (TCO) expectations.
- A range of market trends surrounding infrastructure modernization, digital transformation cloud migration, the adoption of AI, process automation, et al., drive modern data analytics platforms.
- Current market solutions address these needs across five key investment areas: supporting core infrastructure capabilities, delivering cloud-native functionality, extending data warehouse workloads, integrating disparate data, and delivering insights to the business.
- While all vendors reviewed in this Universe are in practice able to meet these requirements, each does so in very different ways, according to corporate history, technological expertise/prowess, and market focus.
- Despite very close scoring, market leaders have invested heavily in forward-looking capabilities such as native support for disparate data types (time series, graph, geospatial, document, etc.) and in-database ML model training and execution.
- While still in its infancy and not yet in mainstream usage across all vendors, the use of AI to automate and augment human decision-making will define the next decade of innovation for modern data and analytics platforms.
- In some ways, the platform itself no longer matters, with most solutions supporting hybrid/multicloud deployment or query federation capabilities. However, hyperscale vendors

still enjoy the advantage that comes from a sizable portfolio of adjacent technologies capable of transparently extending solution functionality (e.g., data science platforms, operational databases, data governance tooling, as well as privacy and security solutions).

Analyzing the modern data analytics platform universe

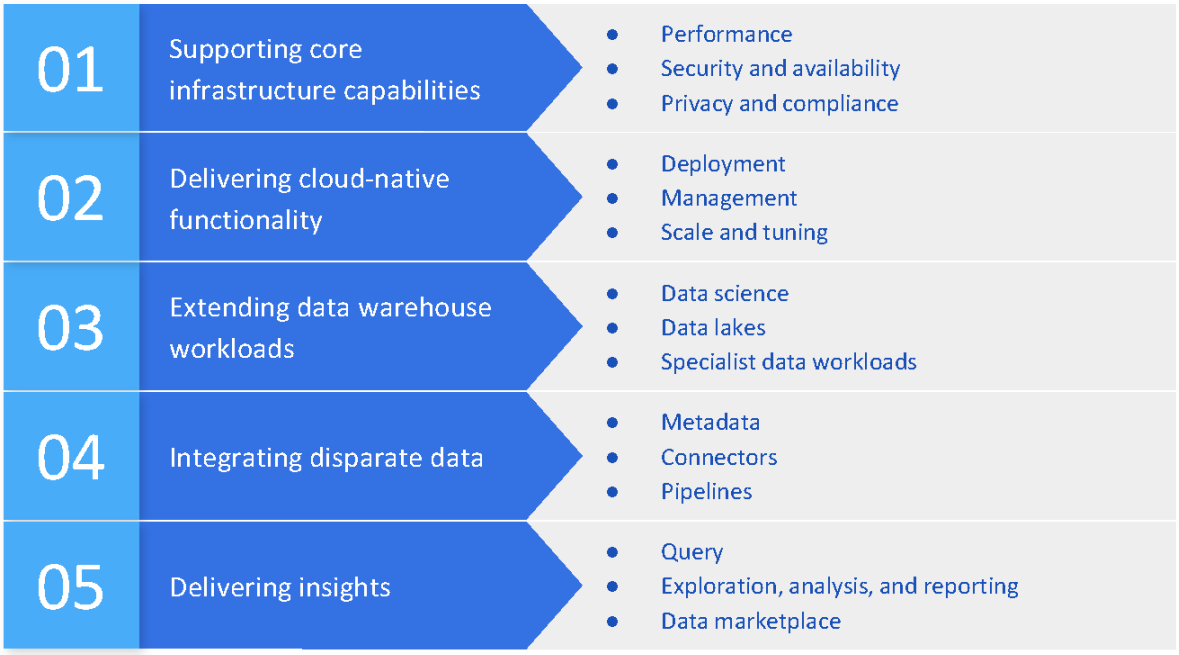
How to use this report

The Omdia Universe report is not intended to advocate an individual vendor but rather to guide and inform the selection process to ensure all relevant options are considered and evaluated efficiently. The report findings gravitate towards the customer's perspective and likely requirements, characteristically those of a medium-large multi-national enterprise (5,000+ employees). Typically, deployments are considered across the financial services, TMT (technology, media, and telecoms), and government sectors on a global basis.

Market definitions

Omdia believes five principal capabilities define a successful modern data analytics platform (see **Figure 2**). Together, these surpass the capabilities found in orthodox data warehouses, data lakes, and other specialized databases. Combined, these constitute a platform capable of delivering valuable insights across the business.

Figure 2: Principal solution criteria



Source: Omdia

Supporting core infrastructure capabilities

A modern data analytics platform is nothing if it cannot deliver data at scale under a high concurrent user load. The challenge for these platforms is extending their intrinsic performance capabilities (MPP, in-memory computing, caching, compression, etc.) into more complex environments, supporting disparate data types, different approaches to indexing data, and new workloads, for example.

Delivering cloud-native functionality

Though on-premises deployments still heavily influence the analytical database marketplace, technology providers have firmly shifted their outlook to the cloud. They did so not just by modernizing databases to run on one or more public clouds but instead by transforming the very architectural nature of the database to function as a cloud-native, containerized fabric of services that enables business practitioners to unify both software and workload across cloud, premises, and multiple-cloud deployments.

Extending data warehouse workloads

In moving beyond basic relational, tabular data, modern data analytics platforms are following two distinct lines of evolution. First, they are seeking to accommodate distinct data types such as documents (JSON files, typically) housing semi-structured information such as user profiles), time-series data, supporting use cases like predictive maintenance for Internet of Things (IoT)-

instrumented devices, and graph analysis to ascribe relationships between entities such as a social network. Second, they are attempting to handle data ingestion historically and in real-time, enabling use cases such as real-time consumer order tracking.

Integrating disparate data

Even though modern data warehouses are evolving to bring disparate data types into a central data repository, they no longer demand such centralization to deliver value. Just the opposite, modern solutions welcome heterogeneity, recognizing that all deployments will require some degree of integration with external operational data sources and analytical platforms. To this end, modern solutions provide a rich set of in-house connectors, working with metadata repositories and federating queries across external data stores to create a unified view of all company data.

Delivering insights

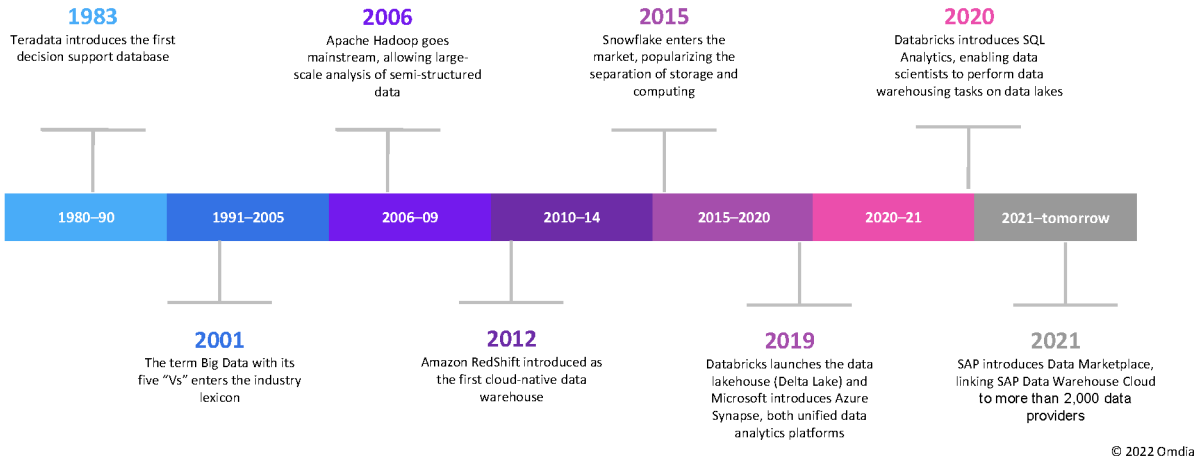
No longer held prisoner by IT departments and database administrators (DBAs), modern data analytics platforms now emphasize self-service access to data for consumption and as a tool of creation. In other words, modern solutions allow users to bring their questions and data. Moreover, solutions now cater to user roles beyond traditional business intelligence (BI) consumers and business users, extending to encompass data engineers, data scientists, data analysts, ML engineers, DataOps professionals, and even software developers through application program interface (API)-based data access.

Market dynamics

Overall, enterprise demand for access to more data in more ways and users define the marketplace for modern data analytics platforms. Data warehouse and data lake vendors, especially those built by or aligned with hyperscale cloud providers—Amazon Web Services (AWS), Microsoft, and Google—predominantly power the market. However, this market's actual shape and scope take its influence from far afield, drawing from adjacencies such as data management and integration, data privacy, data science, and data visualization and discovery. For example, there are BI and data analytics platforms such as Domo, capable of functioning as a full-on modern data analytics platform. Data integration and management solutions like GoodData can create a unified, API-driven data fabric. Distinctive analytics database services, such as the real-time database provider, RockSet, can also meet most modern data analytics platform requirements.

This dynamism can somewhat muddy the waters, blurring the lines between historically distinct solutions. One need only look a few years back to see how rapidly the market has grown from basic data warehousing to embracing new workloads and new opportunities, such as data science and data exchanges, respectively (see **Figure 3**).

Figure 3: Charting the road to modern data analytics platforms

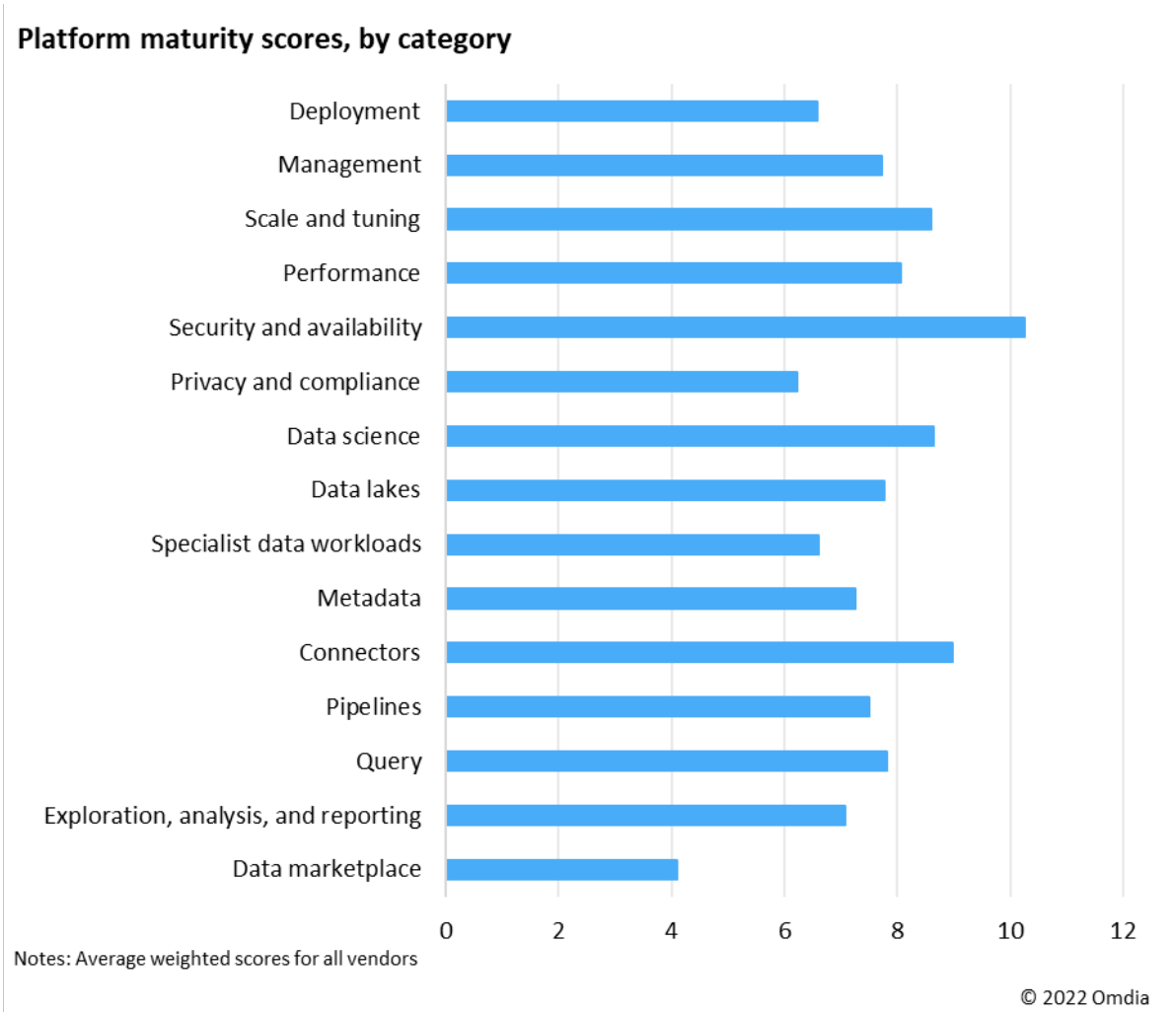


Source: Omdia

With modern data analytics solutions rapidly adopting in-database ML development, training, inferencing, and management capabilities, enterprise buyers may find themselves at a crossroads in selecting between database and data science platforms. The same might hold true for users seeking to build a company-wide data catalog/data hub. Should they purchase something unique from MarkLogic or buy a broader data analytics platform? As with most enterprise market segments, the answer to that question will rest squarely on the shoulders of the buyer. There is no such thing as a completely autonomous or completely extensible do-it-all data analytics platform. Assembly will be required, particularly for companies seeking to tackle hybrid/multicloud operations or accommodate highly specialized workloads that might demand a greater degree of functionality than that available from a broader platform.

That said, some universal truths are shaping the market as espoused by the leading data warehouse and data lake vendors. First, solutions are moving up the enterprise stack from traditional data professionals, targeting a wider array of buyers and users, including data scientists, business analysts, information workers, chief information officers (CIOs), developers, and other business decision makers. Second, solutions are eyeing several highly specialized horizontal use cases spanning data warehouse and data lake modernization, customer analytics, fraud detection, streaming analytics, geospatial analytics, data sharing, and data science. Third, plying their ability to meet the massive scale and performance requirements, solutions are now seeking to empower more business users with low/no-code tools capable of finding, analyzing, and sharing data securely in an IT-governed environment.

Figure 4: Overall vendor performance



Source: Omdia

Such dynamics played a fundamental role in Omdia's ranking of the modern data analytics platforms reviewed within this report, as did basic functionality supporting performance, security, and availability (see **Figure 4**). This is not surprising, given the maturity of all participating vendors. That said, generally, no single solution met all of Omdia's requirements in full. Only a handful of vendors (e.g., SAP, AWS, Microsoft, Google) maintain an in-database data marketplace, supporting publication and subscription. The same can be said for emerging investment areas such as data science and metadata. Here, scoring varied quite a bit, depending on whether the vendor prioritized a supportive investment in this area. Scoring also varied, depending on corporate history, technological expertise/prowess, and market focus.

From this evaluation, Omdia has identified several leaders capable of delivering consistently across all Omdia measures (see **Figure 5**), providing basic functionality and investing in emerging

technologies. Even so, there is plenty of room for challengers and prospects to operate successfully in support of discrete user selection criteria. Only three percentage points separate market leaders from market prospects, indicating a tremendous amount of parity, even Omdia's maturity weightings that prioritize forward-looking capabilities such as using AI to automate solution operations tasks.

What remains is for potential customers to identify the vendor that aligns closest with internal requirements. For example, customers dedicated to a given data processing technology, such as Apache Spark, will be best served to partner with an Apache Spark specialist such as Databricks. Similarly, customers needing a life cycle-complete data science user experience native to the underlying database might opt for a vendor like Cloudera. Alternatively, customers who are not looking for bells and whistles but rather an easy-to-consume solution regarding price/performance might opt for Actian over any competitor. At the end of the day, any one of these offers can stand uniquely as the best fit for a given customer's needs.

Figure 5: Vendor rankings in the modern data analytics platform universe

Vendor	Product(s) evaluated
Leaders	
Oracle	Autonomous Data Warehouse
AWS	RedShift
Databricks	Lakehouse Platform
Microsoft	Azure Synpase
Challengers	
Google	BigQuery
IBM	Db2 Warehouse
SAP	Data Warehouse Cloud
Teradata	Vantage
Cloudera	Cloudera Data Platform
Prospects	
Vertica	Analytic Platform
Actian	Avalance

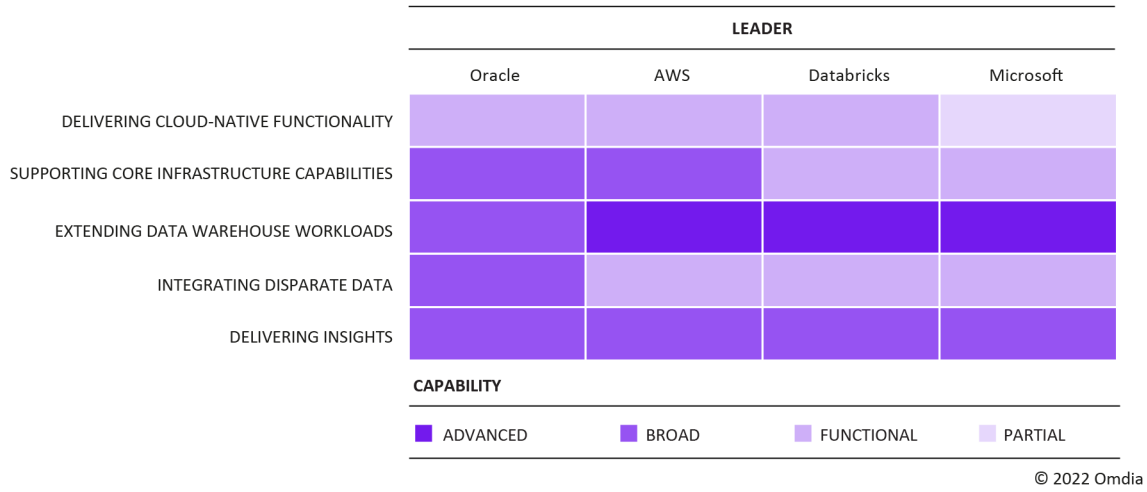
Source: Omdia

Market leaders

Based on the Omdia criteria and analysis—AWS, Databricks, Microsoft, and Oracle—are identified as market leaders in this year's report. This category represents the leading solutions that provide advanced functionality across crucial solution capabilities, namely delivering cloud-native functionality and extending data warehouse workloads. These solutions offer a broad set of features

and are worthy of a place on most technology selection shortlists, regardless of the use case (see **Figure 6**).

Figure 6: Market leaders' capabilities heatmap



Source: Omdia

While overall scoring paints a picture of relative parity across most Universe participants, market leaders AWS, Databricks, Microsoft, and Oracle stood out, generating solution capability scores of 65%. These vendors can be relied upon to bring the technological prowess necessary to meet current market requirements for a modern data analytics platform.

Not surprisingly, three of our four market leaders operate a global hyperscale cloud platform of their own. This enables stand-outs like Oracle, for example, to offer the same hardware found on its public cloud platform to on-premises customers. However, even without its platform data lakehouse pioneer, Databricks brings its differentiation through a partner-led approach to multicloud deployments, creating deployment parity across leading hyperscale platforms.

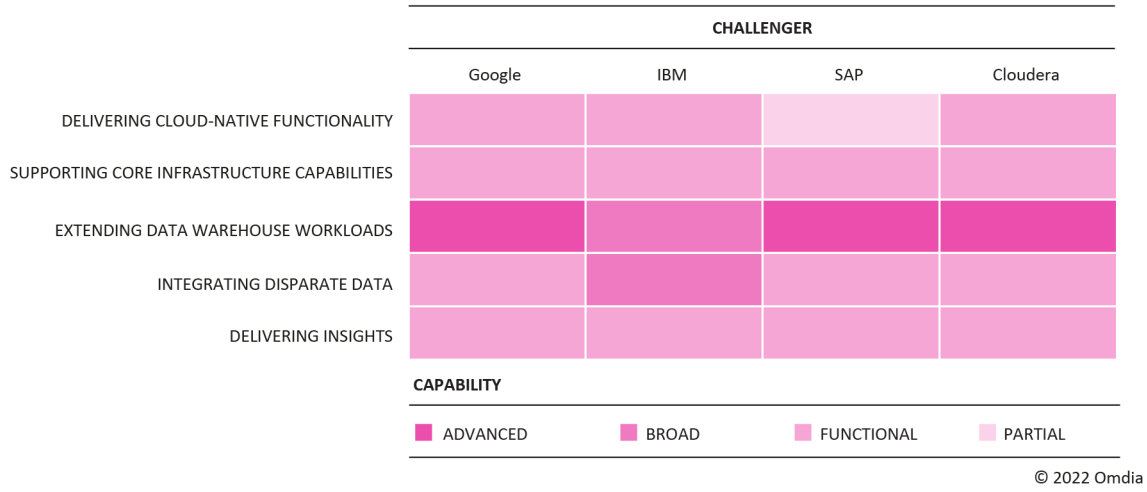
These unique aspects aside, all market-leading solutions scored well across the five Omdia principal capabilities, delivering a solid platform that seeks to bring together disparate data sources in the service of disparate data consumer workloads. This "do-it-all" aspect was particularly evident in how the market leaders approached the challenge of extending data warehouse workloads. Here, vendors have done the most to automate system and resource management, merge data warehouse and data lake functionality, and encourage in-database development, delivery, and management of AI business outcomes.

Market challengers

The solutions in this category offer some advanced capabilities, have appropriate functionality across other areas, and should be considered part of a technology selection process. Omdia has

identified several market challengers—Cloudera, Google, IBM, SAP, and Teradata—all of which can readily deliver at scale across a broad range of use cases (see **Figure 7**).

Figure 7: Market challengers' capabilities heatmap



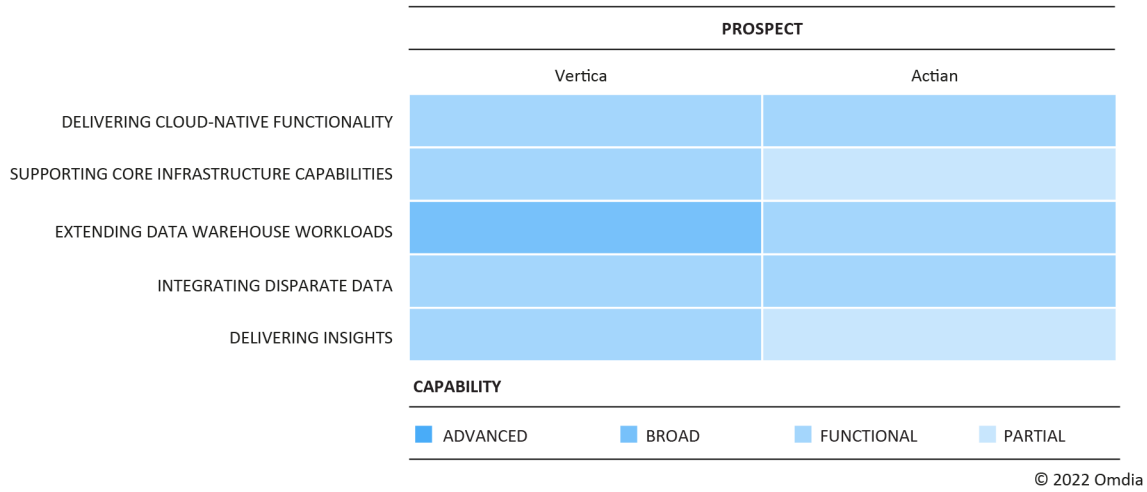
Source: Omdia

This grouping, which scored between 62% and 65% in the Omdia solution capability evaluation, offers buyers a very balanced set of features across Omdia's five principal capabilities. Interestingly, challengers equaled market leaders in several key areas, such as the inclusion of low/no-code tooling for creating and managing data pipelines. In other areas, challengers scored the highest among all competitors. This was also true with regards to managing multicloud deployments.

Market prospects

Solutions in this category offer a good set of capabilities but lack some of the more advanced features and/or capabilities provided by some competitive offerings. Given that only a slim three percentage points mark the difference between a leader and a prospect in this review, the remaining two vendors—Actian and Vertica—are better thought of as "specialists" as they each provide a comparatively unique value proposition and overall philosophy (see **Figure 8**).

Figure 8: Market prospects' capabilities heatmap



Source: Omdia

Honorable mention

While not reviewed comparatively within this report, Snowflake deserves special consideration. Snowflake was early to market data storage and analytics as a true database software as a service (DBaaS), bursting onto the technology landscape in 2014 with one of the first data warehousing offerings to successfully separate storage from computing resources. From its inception through its record-breaking initial public offering (IPO) in 2020 and on to today, the company has espoused a simple strategy—deliver a fast, flexible, secure, and scalable data platform that makes it easy to load, integrate, analyze, and share data.

Like all entrants in this Omdia Universe, Snowflake seeks to unify a wide array of data formats within a single datastore and then open that data up to an equally diverse set of IT and business consumers using their languages and tools of choice. Regarding specific workloads, Snowflake functions as both a data warehouse and a data lake, leveraging an MPP architecture and unified query and development languages (SQL, Python, Java, etc.) across both. As such, Snowflake can readily support data science, data engineering, and software development professionals. Further, it emphasizes the free flow of data within and between organizations, building on its built-in data exchange platform.

Snowflake, however, is unique in its cloud-centricity. While most data warehouse solutions endeavor to accommodate hybrid cloud and premises deployments, Snowflake has focused on delivering fully managed services atop leading hyperscale platforms—AWS, Google Cloud Platform (GCP), and Microsoft Azure. More than this, Snowflake endeavors to enable cross-cloud functionality and is working toward the creation of a single layer of data abstraction that would allow customers to build applications and use cloud services from any cloud provider without having to physically move data or manage data security, governance, and management on each platform independently.

While still a rapidly evolving platform, Snowflake Data Cloud's aspiration aligns nicely with the company's stance as a cloud-first innovator seeking to serve as a cloud-agnostic alternative to data

warehouses provided by the cloud providers upon which Snowflake runs. In other words, Snowflake wants to be seen not as “yet another data silo” but as a genuinely open, global data platform. This endgame is already evident in the company’s rapidly expanding ecosystem of data science, BI, data integration, and security and governance partners. The company’s ultimate success in this endeavor will depend upon its ability to match the advanced capabilities of its more established and mature rivals. Regardless, Omdia feels that Snowflake Data Cloud should be considered alongside the modern data analytics platforms reviewed in this report. For customers looking for a truly cloud-native analytical data platform that effortlessly balances performance, simplicity, and capability, Snowflake stands as a solid option among market leaders.

Market opportunities

Many mature solutions and long-term vendors make up the market for modern data analytics platforms. However, because it is evolving rapidly, enterprise buyers should look at the vendor landscape as if it were an emerging market. Attrition is unlikely, as most players have already established a substantial customer base. That said, disruption is likely as vendors will continue to alter their fundamental platform architectures to simplify data acquisition, processing, analysis, distribution, and understanding. Many of the players with both data warehouses and data lake offerings, for example, have or are in the process of uniting those within a single offering. While creating some disruption for buyers, these changes will ultimately help them more fully leverage data across the business, making new functionality a welcome investment in time and money.

Market threats

AI would usually be considered a strong opportunity for this market. However, vendors are just beginning to take advantage of AI. Therefore, they are still likely to make some missteps, especially given AI’s two-fold impact on modern data analytics platforms. The first is using AI to automate and augment human decision-making in managing these platforms. Second, the ability to build and run AI solutions natively within these data analytics platforms should be a significant differentiator and market driver.

The overall market is still highly immature in using AI to make it easier for administrators to provision, configure, secure, and optimize these platforms. The market has a few select vendors putting this to work within their systems beyond established routines such as query plan optimization. Even so, Omdia believes that using AI to automate and augment human decision-making will define the next decade of innovation for modern data and analytics platforms with early leaders—AWS, Google, SAP, Microsoft, Oracle, and SAP—in building more and more AI-assisted functionality into their solutions. Such innovation will make the most significant difference for IT departments seeking to optimize their infrastructure spending through self-tuning query plans and self-healing data pipelines, for example. More subtle benefits are sure to follow, assuming both technology providers and enterprise practitioners alike further prioritize responsible AI practices.

Relatedly, regarding building AI in the database, the market is still undecided on the ultimate efficacy and value of in-database ML. Yes, it reduces data movement and can simplify the overall ML lifecycle, technology-wise. However, the way vendors deliver this functionality does not align well with established market practices, leaving in-database ML more an interesting area of exploration than a game-changing best practice at this time. As these solutions mature, however, Omdia

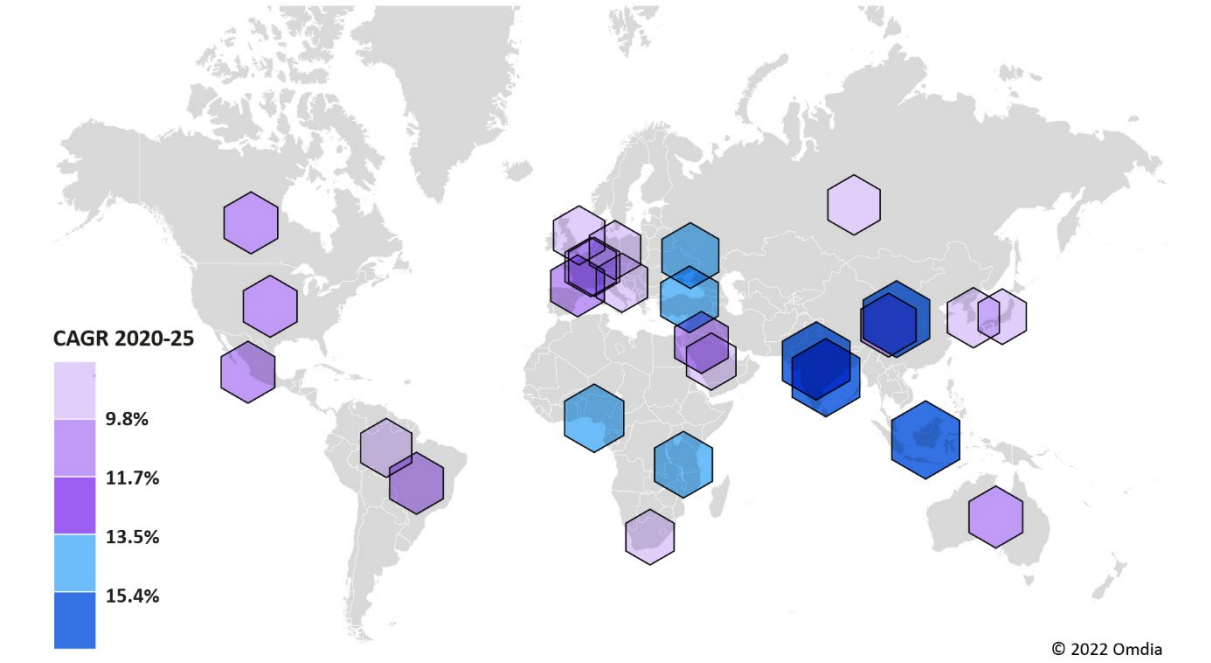
anticipates a general push among technology providers to move ML processing closer to supporting data.

Market outlook

Broadly speaking, as one of the longest-running technology segments, the analytics and data management marketplace has and will continue to enjoy a solid, double-digit growth pattern over the foreseeable future. Omdia finds that early market disruption surrounding the COVID-19 pandemic has diminished and will not have a long-term negative impact on customer optimism. On the contrary, the pandemic has only accelerated customer investments in technologies that directly support cost reduction, business resilience, or operational efficiency. Within the analytics and data management market, this trend translates into an increase in buyer spending on solutions supporting data acquisition, storage, processing, management, and analysis.

Omdia noted that by the end of 2021, the analytics and data management market was worth around $102 billion, with a compound annual growth rate (CAGR) of 10.7% for the 2020–25 forecast period. Geographically, the marketplace will enjoy a strong uptick within the Asia & Oceania region, cooling off elsewhere in regions where heavy investments in big data have already occurred, most notably within western Europe and the Americas (see **Figure 9**).

Figure 9: Global outlook for modern data analytics platforms



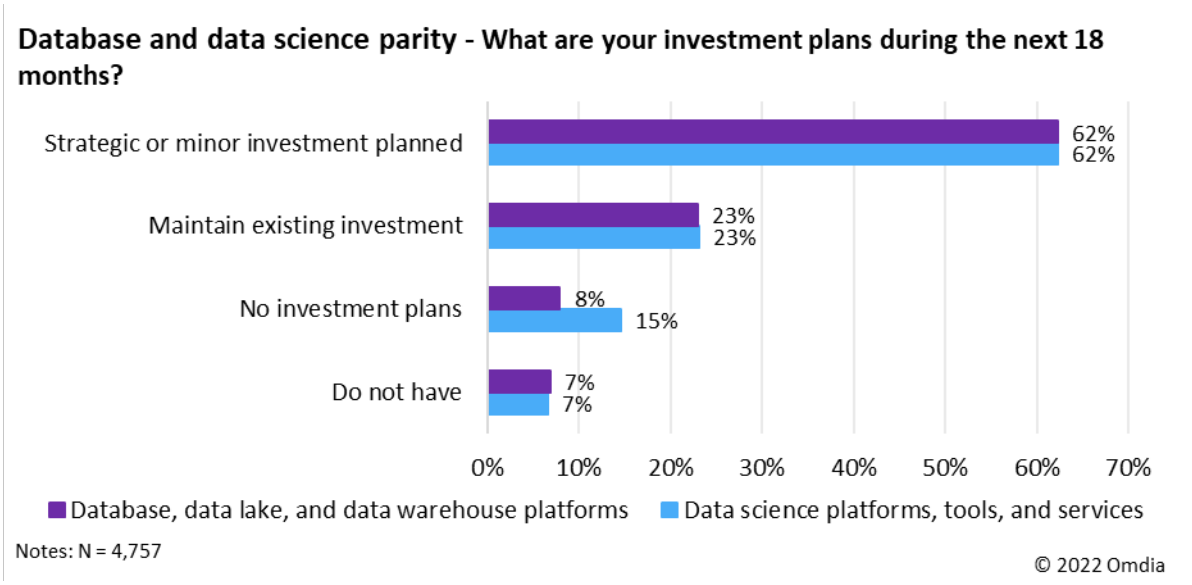
Source: Omdia

The expanding appetite for data-driven insights across enterprise organizations primarily drives the continued growth in the analytics and data management market across industries, especially as more organizations look to Al use cases. Increasing demands such as regulatory compliance reporting requirements and continuing ecological and geopolitical market disruptions further support this appetite. Ever-increasing data volumes, evolving data workloads, and proven data and analytics use cases are promoting adoption among a new audience of business users and domain experts.

Advances in self-service user experiences and ready access to specialized hardware and software infrastructure further propel this demand. In response, analytics and data management solutions are becoming increasingly versatile, allowing the exploration of new implementation areas and emergent use cases supported by new and existing data, analytics, and Al infrastructure, particularly in hybrid and multicloud deployment scenarios.

Down in the trenches, however, Omdia expects a substantial uptick in the various database types that make up a modern data analytics platform, predominantly data lakes and data warehouses. Such investments will rival those of data science platforms, a rapidly growing market that, when viewed globally, will yield a CAGR of 17.17% over the next three years (see **Figure 10**). This investment parity between data analytics platforms and data science technologies speaks to their conjoined value in driving enterprise business transformation projects such as customer 360, supply chain transparency, business process optimization, and many more.

Figure 10: The continued importance of analytical databases



Source: Omdia

Vendor analysis

Actian (Omdia recommendation: Prospect)

The Actian Avalanche cloud data warehouse should appear on your shortlist if you want a straightforward, cost-effective analytics database with a rich history that is fine-tuned to real-time decision-making.

Overview

On the surface, the Actian brand looks relatively straightforward, entering the market in 2018 under the auspices of the HCL Software division, which bought out its former owner, Sumero partners. This change in ownership bodes well for the vendor, with Actian now able to expand its staff (adding 100 roles recently, according to the vendor). However, the company's new ownership also obscures a rich heritage found under the company's previous name, Ingres Corp. Founded in the 1970s by Eugene Wong and Michael Stonebraker (who also founded Vertica), Ingres kick-started the relational database market with the Ingres database, a platform that accentuates computational performance.

Fast forward to today, this legacy of innovation lives on within Actian. The Actian Avalanche is a fully managed hybrid cloud data warehouse that, like its predecessor, is designed to deliver high performance and scale across several dimensions—data volume, user concurrency, and complex query performance. Focusing on North America, the European Union (EU), and UK markets, Actian sells Actian Avalanche as a fully managed hybrid cloud data warehouse service supporting both on-premises and multiple cloud deployment options spanning AWS, Azure, and Google Cloud. Actian partners closely with Google to target several data warehouse cloud migration paths from IBM Netezza, Teradata, and Oracle Exadata to Google Cloud. Interestingly, Actian has worked directly with Google engineers to create pre-configured integrations with key Google services, including Google Cloud Dataproc, Cloud Storage, and Looker; DataFusion, Pub/Sub, and Kubeflow are planned for the future.

Actian currently boasts more than 3,000 customers, including 24 of the Fortune 100. Overall, Actian customers are split between the upper mid-market and enterprise buyers looking to manage highly performant workloads up to approximately 30 terabytes (TB). Actian does not sell verticalized solutions, preferring to accommodate a wide array of horizontal use cases. However, the vendor enjoys a strong customer following across healthcare, retail, and marketing customers. Notable customers include the University of Oxford, Phillips 66, Siemens, HSBC, Sabre, Citi, and Bloomberg.

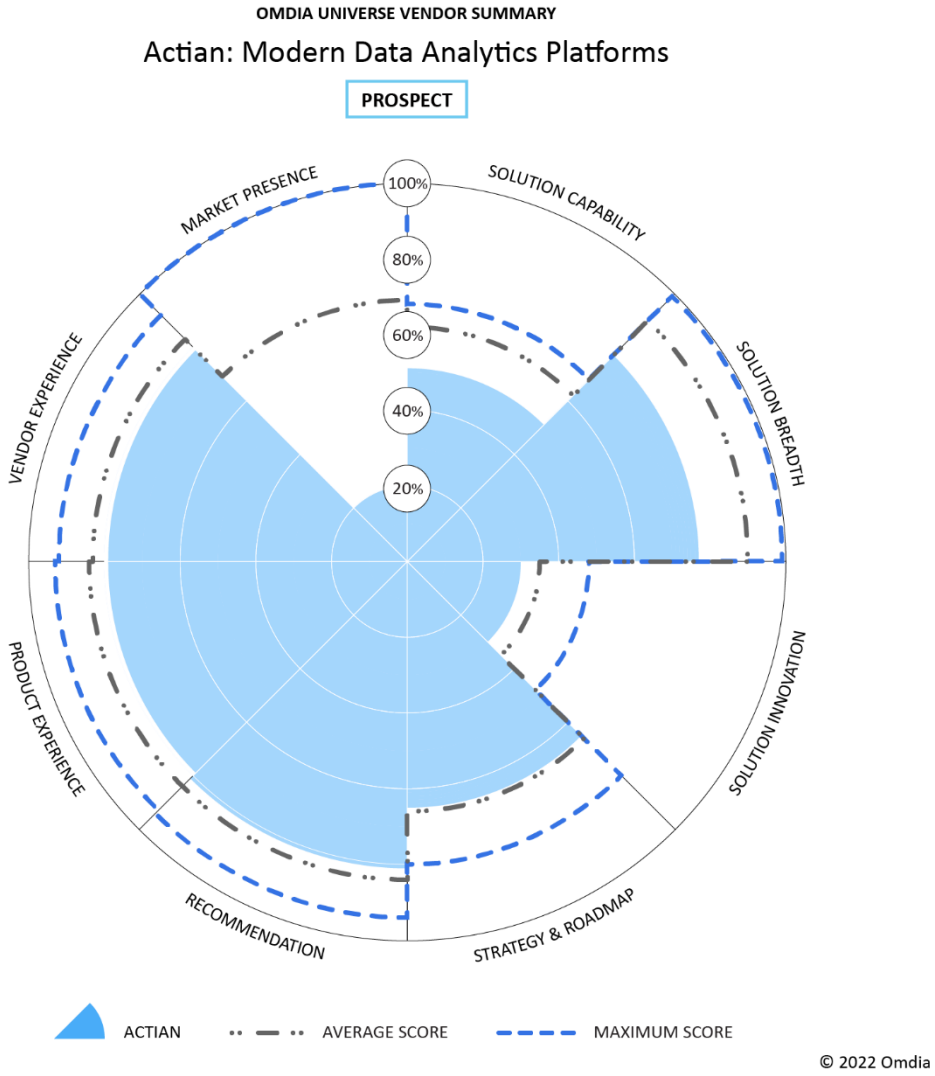
Comparatively, Actian Avalanche does not sport the same bells and whistles as some of the more advanced data warehouses reviewed in this report. However, that is not the company's focus. Instead, Actian seeks to operate as a trusted, flexible, and easy-to-use data platform that can manage a competitor's data warehouse offering alongside its own within the Avalanche platform. In this way, the company is chasing an optimal price/performance ratio for its customers across disparate cloud-borne data solutions, an endeavor very much in line with the needs of enterprise

buyers struggling to achieve the same cost profile online as they had on-premises without sacrificing openness.

Roadmap priorities

- Continue to focus on flexibility options for hybrid deployments and data integration and orchestration capabilities
- Invest in product usability through a substantial investment in a new user experience
- Facilitate the creation of turnkey solutions in support of specific use cases
- Expand Avalanche through the introduction of new Actian products into the Avalanche platform

Figure 11: Omdia Universe ratings — Actian



Source: Omdia

Strengths

- Thanks to its focus on creating an optimal price/performance ratio for its users, Actian Avalanche is a comparatively unique offering among the rivals reviewed in this report. Actian emphasizes real-time insights through its vectorized, columnar database engine on the architectural side of this equation. The engine uses patented technology (positional Delta Trees), aggressive compression ratios around 4–6%, and multi-core parallelism to speed up processing. This vector-based architecture is beginning to pick up momentum in the marketplace because it

can outperform traditional scalar architectures across key use cases like fraud detection, content search, and threat detection.

- On the other side of the price/performance equation, Actian proffers a comparatively simple pricing plan that favors transparency through a basic hourly, on-demand Actian Unit (AU). According to Actian customers, this approach affords a high degree of flexibility to control costs and forecast future spending based on simple AU calculations.

- Actian Avalanche is closing in on true hybrid/multicloud execution and management. First, the solution includes a federated query capacity that optimizes query performance across multiple Avalanche deployments (cloud or premises) without any data movement. Second, the product enables users to manage numerous Avalanche deployments regardless of location, even across multiple hyperscale cloud providers. These and the company's simple licensing approach enable Actian customers to optimize their hybrid/multicloud spending while leaving data in place.

Limitations

- Actian is still building out the finer details of Avalanche as a cloud-native offering. First, though Actian Avalanche can be deployed freely on-premises and in the cloud, the solution does not yet enable multi/hybrid-cloud operations from within a fully managed cloud service. For multi/hybrid support, the vendor recommends self-managing across both cloud and premises. Second, Actian Avalanche as a fully managed service can only be purchased on the GCP. Actian intends to extend this to AWS and Microsoft Azure in late 2022. Third, Actian is still working on virtual private cloud (VPC) deployment options. All told, these growing pains leave Actian favoring GCP over the short term.

- Relatedly, though Actian has fully embraced Kubernetes-managed containerization for Avalanche, the vendor is still working on some of the features enabled by cloud-native architectures. For example, the Actian Avalanche cloud infrastructure can be scaled manually or through a schedule, but the solution does not automate this elasticity as yet. Note, however, that if the server remains idle for a specified amount of time, it will automatically shut down. Also, the solution does not yet support serverless functionality, which is rapidly becoming the norm among competitors.

- Compared with most of the vendors in this review, Actian is not yet pursuing the notion of in-database ML. For example, while Avalanche can execute Python code within a user-defined function (UDF), it cannot store or execute model formats like ONNX, nor does it include any internal ML development or management tooling. It also does not yet integrate with any external ML tools. Support for data science workloads has already become a differentiator for vendors in this space, as with support for a broad set of data formats such as geospatial, time-series, and graph—all of which remain on the vendor's roadmap.

Outlook

For customers seeking a straightforward data warehouse with a unique architecture tuned to real-time decision-making, Actian represents a solid option, particularly regarding the vendor's ability to

simplify pricing and management across hybrid/multicloud deployment scenarios. Feature-for-feature, the vendor has some work ahead if it wants to target current market demand for fully managed multicloud options, data science workloads, and disparate data type support. Given the company's aggressive technology roadmap, which calls for advantageous capabilities like data quality tooling, as an example, Actian will find new market opportunities. This is possible if it can bring the same level of optimization it has built in partnership with Google to other hyperscale cloud platforms.

Amazon (Omdia recommendation: Leader)

Amazon Redshift should appear on your shortlist if you want a highly optimized yet flexible analytics platform capable of synergizing with the company's broad portfolio of data, analytics, and AI tools.

Overview

AWS introduced Redshift in 2012 as a massively parallel, petabyte-scale, columnar data warehouse that was compatible with the popular open-source PostgreSQL database, upon which Redshift is partially based. Since then, Redshift has changed dramatically in scale and scope. Redshift has grown to handle disparate data types (e.g., JSON and geospatial data), evolved to use ML aggressively to optimize operations, and expanded to natively analyze data from across numerous data sources, ranging from operational databases and other Redshift data warehouses to raw data lake object stores.

Operating globally with the capacity to handle exabytes of data, Redshift serves as AWS' principal analytics database, serving tens of thousands of customers globally. AWS markets Redshift as an enabler of several current market trends around embedded BI, operational analytics, data as a service, data marketplaces, collaborative data sharing, data mesh architectures, and predictive analytics. More broadly, the company stresses three key design goals for Redshift.

- Make analytics something anyone can access
- Enable the analysis of any and all data sources/types
- Deliver analytics price-performance at any scale

Like most players reviewed in this report, AWS uses the separation of storage from computing to make Redshift easy to adopt and affordable to maintain. The company does this architecturally by shutting down unused clusters at no cost. However, it also emphasizes adaptive pricing through a mixture of consumption-based, resource-based, on-demand, bounded, and blended models. This includes on-demand and reserved instance pricing, enabling customers to lock a lower price for more predictable workloads.

AWS does not offer Redshift as a transparent component of any verticalized solutions. However, AWS tries to speed time to market for specific use cases through a series of accelerators that leverage professional services to package up Redshift alongside other AWS services. For instance, the company can use professional services and accelerator software to bring Redshift together with AWS SageMaker, EMR, S3, DynamoDB, QuickSight, and Lambda to create a real-time ML model

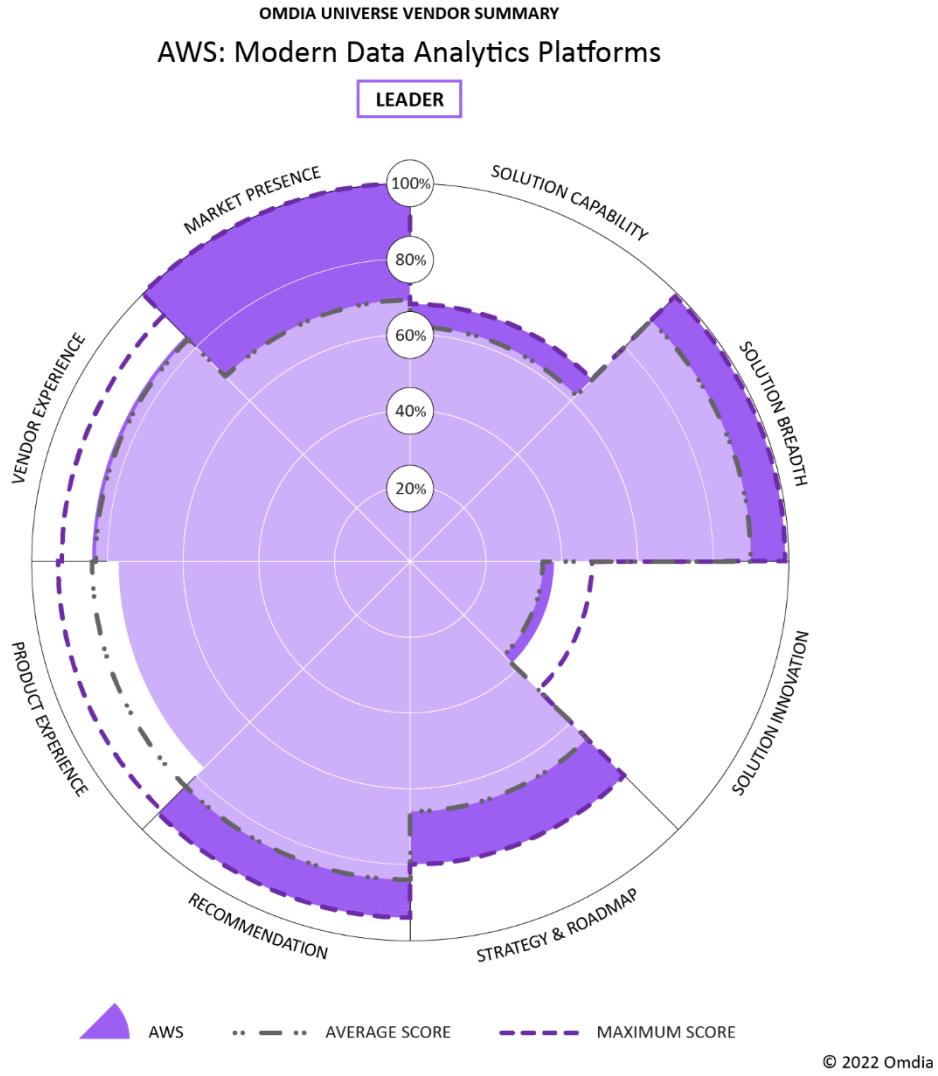
deployment platform for use within logistics companies seeking to gain transparency into their delivery fleet.

This notion of Redshift as a core cast member amidst a much bigger production shines through in how the product has evolved into a true multi-modal database. Though AWS offers more than 15 purpose-built database engines, including relational, key-value, document, in-memory, graph, time series, wide column, and ledger databases, the company is not afraid of equipping Redshift with many of these same capabilities (time-series, document, and geospatial in particular). In this way, the company gives its customers the best of both worlds—a central analytics database to accommodate multiple data types and built-for-purpose databases to accommodate highly demanding, highly specialized workloads.

Roadmap priorities

- Explore opportunities for joint relational/analytical workloads
- Prioritize event-based serverless computation
- Further integration between transactional data, analytic data, and AI workloads
- Shift in customer focus, moving from building on the cloud to building for the cloud
- Continue to make analytics easy to use with the Redshift Serverless and intuitive visual interfaces as with the new Query Editor

Figure 12: Omdia Universe ratings — Amazon



Source: Omdia

Strengths

- Many of the leaders in this comparative review have invested in some form of a data marketplace to help buyers more readily access third-party, public data sources specific to a range of vertical markets and horizontal use cases. AWS Data Exchange (ADX) stands out among these for two reasons. First, it is one of the most mature platforms, featuring more than 3,700 data sources from more than 200 qualified data providers. Second, it works with the company's broader Amazon Marketplace, allowing customers to subscribe to software and data from multiple providers using a single billing system, management console, and API. Additionally,

Amazon Redshift users can query live-shared third-party data without extracting, transforming, and loading files through Redshift's integration with ADX. This allows users to directly operationalize external data with little complexity or overhead.

- As a market leader in operationalizing ML in the enterprise, AWS has a differentiated degree of AI domain expertise at its disposal, which it readily applies to Redshift in two ways. First, like many competitors, AWS enables users to run ML workloads in-database. With Amazon Redshift ML, users with little or no data science experience can build, train, and deploy models using SQL. Users can also bring their ML models, frameworks, and libraries, using Amazon SageMaker as a deployment platform on top of Redshift. More importantly, AWS has used its ML experience and specialized hardware to automate, augment, and optimize Redshift itself. For example, the company's advanced query accelerator (AQUA) uses a distributed and hardware-accelerated cache to speed up Redshift queries, removing/minimizing CPU and network bottlenecks. Further, with automatic table optimization, Redshift can self-tune the design of tables (using sorting and distribution keys) with no user intervention. Additionally, with automatic workload management capabilities, Redshift can dynamically manage computing, memory, and concurrency settings to maximize query throughput.

- Though AWS emphasizes using ML to automate and optimize Redshift operations, the company does not lock users out from taking complete control of any AI-derived functionality. Instead, AWS allows users to augment, extend, and even override automated processes via user experience, command line interface (CLI), or API. With Amazon Redshift workload management and Automatic Table Optimization (ATO), as an example, users can revert fully to a manual mode, selecting specific table design, concurrency, and memory allocations for running Redshift queries. This allows AWS to meet its customers where they are on their data and analytics journey using the same context.

Limitations

- AWS actively targets Redshift optimization (often informed by ML automation) as a means of improving customer return on investment (ROI) by better balancing cost constraints and key performance indicators (KPIs). To that end, the company recently introduced a true serverless deployment option, Redshift Serverless, which allows users to deploy and manage data warehouses without worrying about the underlying infrastructure and pay only for the duration in seconds when their data warehouse is in use. This completely frees users from having to provision supportive clusters, and it will work with RedShift Athena to support serverless access for data lake use cases as well. The company's decision to release this service in November 2021 as a preview speaks to its importance, as AWS generally eschews previews in favor of general availability (GA) software releases. Unfortunately, even with RedShift Serverless available in early release, AWS remains late in chasing the sizable market opportunity for serverless data warehouse functionality.

- Amazon Redshift ML offers a highly automated and capable experience for non-data scientists seeking to build ML models using Redshift. However, AWS' approach does not fully leverage Redshift itself. Users can import and execute a model inside an Amazon Redshift cluster and

invoke a model stored in Redshift using a SageMaker endpoint. However, AWS' preferred architecture for Redshift ML involves moving data out of the data warehouse and into S3 and then into SageMaker, where it can be run through SageMaker Autopilot and SageMaker Neo for a hands-off model building and training. The same goes for manual development work, which still moves data out of Redshift. This translates into additional costs for Redshift customers. Note, however, that this issue centers on model development. AWS does not charge Redshift users extra for inference queries run against models that have been imported into Redshift. Regardless, this architecture allows rivals to position Redshift as favoring only data and model storage and model inferencing in supporting ML development.

- AWS offers users an extensive free tier of more than 60 products, including Redshift and several other databases, counting the company's popular NoSQL database, DynamoDB. Unlike DynamoDB, which is always free to users within certain constraints, Redshift is only available for a two-month free trial with up to 750 hours of use for new customers. While such time-limited trials are not uncommon, the industry norm, especially for hyperscale providers like AWS, is to enable such functionality within a permanently free tier. An always-free tier approach can create long-term momentum, especially in capturing usage among students and future practitioners. When tied to other partially free services such as a data exchange/marketplace, it can help establish the database as a widely used, self-provisioned platform at a departmental level.

Outlook

With Redshift operating as its central analytics database, AWS is well-situated to maintain its dominant position in the analytics and data management marketplace. With Redshift, AWS demonstrates a solid understanding of two important facts. First, there is no one-size-fits-all database. With Redshift, AWS not only offers a flexible platform capable of handling disparate data types, but it also offers specialized databases such as RDS and DynamoDB capable of working alongside Redshift. Second, AWS has aggressively put ML to work within RedShift to optimize and automate operations and applications. This, coupled with AWS' flexible (and now serverless) licensing options, allows Redshift to function as a central yet unseen analytics resource capable of meeting a wide range of enterprise demands cost-effectively.

Cloudera (Omdia recommendation: Challenger)

Cloudera Data Platform (CDP) should appear on your shortlist if you want a cloud-native, flexible, open, and portable data warehouse platform with a storied history in supporting big data workloads at scale.

Overview

Cloudera entered the data and analytics marketplace in the late 2000s with a flourish, riding the big data gold rush ushered in by web giants Facebook, Google, and Yahoo!. At the time, these companies were not just building today's search and social media platforms; they were inventing the software necessary to do so, namely the open-source project, Apache Hadoop. This software, which uses a distributed file system and the MapReduce programming concept, allowed these firms to run many thousands of concurrent analytical queries at speed against massive amounts of data.

Formed by the engineers who built Hadoop, Cloudera launched a commercial distribution of this influential software in 2009 and quickly rose to prominence, going public in 2017. Since that time, the company has endured many financial and market shifts. Still, it has stayed true to its original commitment to open source and big data processing regardless of scale or performance demand.

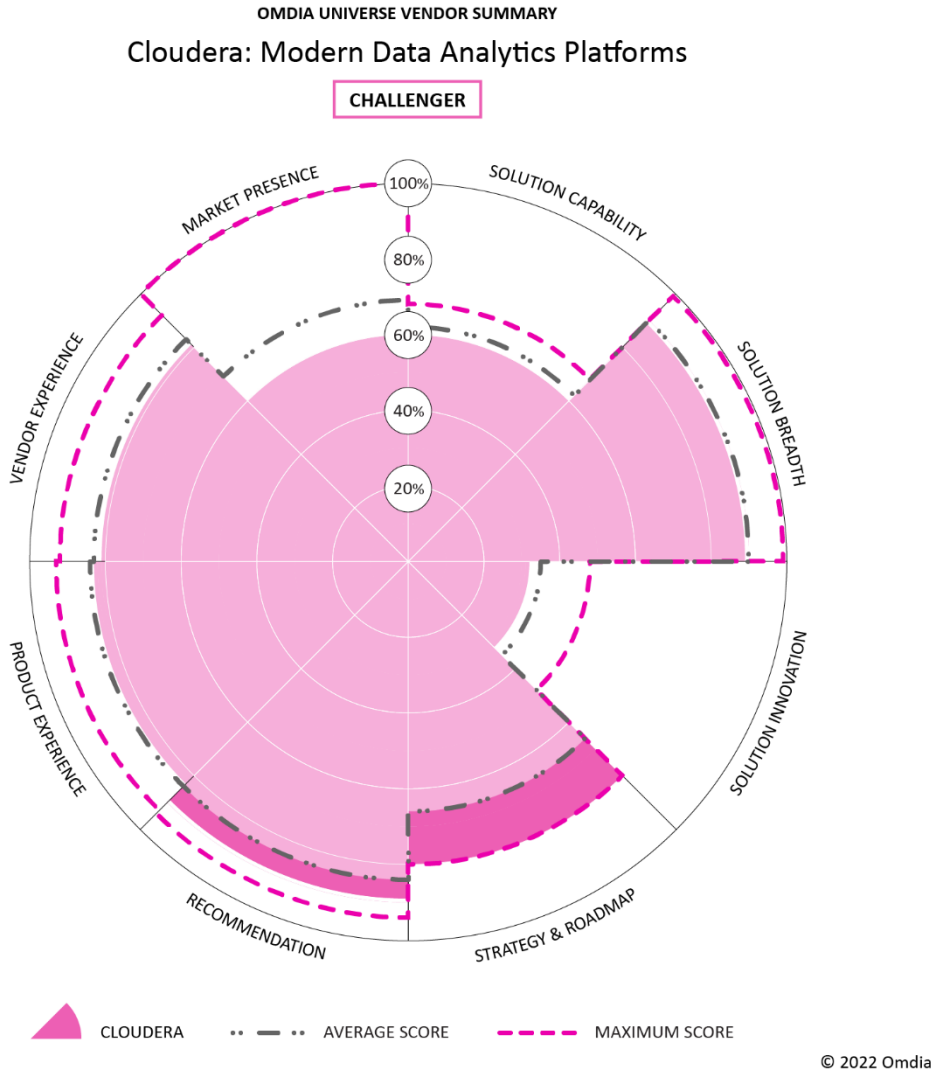
The launch of CDP in 2019 significantly modernized the company's original vision, creating a highly unified, cloud-native, multifunction analytics platform that could run on any cloud platform, public or private. The company runs across AWS, Microsoft Azure, and GCP for its public cloud offering. Regarding reach, Cloudera sells CDP into a wide range of vertical markets globally with no clear preference among them. Across these markets, the company emphasizes creating a wide range of business transformational use cases such as IoT-enabled predictive maintenance, genomics research, fraud detection, and real-time compliance monitoring.

Since the launch of CDP, Cloudera has seen a tremendous uptick in its adoption, with nearly half of the company's 5,000 customers already making the leap from the vendor's legacy software to CDP. The fact that most of these customers are Fortune 5000 companies highlights CDP's ability to offer enterprise-grade performance, scale, security, and the ease of data management expected from cloud platforms. Its support for the entire data lifecycle, an approach to a hybrid cloud that extends beyond basic licensing, a single enterprise-grade security and governance framework across the whole platform, as well as financial governance support makes CDP an easy choice for customers looking for a single platform combining data sharing, data engineering, ML, data warehousing, and operational database capabilities.

Roadmap priorities

- Invest in further automation and optimization across hybrid cloud deployments that simplify how users request, find, use, and share data, regardless of where that data resides
- Improve workload automation and optimization that does not require user intervention but still balances performance KPIs against cost targets

Figure 13: Omdia Universe ratings — Cloudera



Source: Omdia

Strengths

- Cloudera's first-mover status within the big data and analytics marketplace as a predominantly open-source player still serves the company well, particularly in its support for semi-structured and unstructured data. Cloudera's CDP was built explicitly to handle the distributed processing of massive, complex data sets across commodity hardware (i.e., the public cloud) on any cloud for any analytics and data. CDP's open architecture invites flexibility regarding tool selection, offering true multi-engine, multi-function analytics. For example, CDP customers can select the appropriate query engine for the task at hand, picking Hive, Impala, or Spark to perform more

traditional extract, transform, and load (ETL) tasks like schema on write commands or picking other engines for more agile schema on reading tasks. Users enjoy a single management pane and consistent security, encryption, governance, and privacy controls in both instances. In essence, CDP offers users all the benefits of a completely open, self-assembled stack packaged within a highly integrated and unified platform experience across any public or private cloud.

- While all the vendors in this comparison prioritize in-database ML workloads, Cloudera has set its sights more broadly with tooling that spans the entire ML lifecycle. To that end, Cloudera Machine Learning (CML)—licensed separately—delivers an end-to-end ML solution, complete with its own distributed compute resources for Python, R, and Spark-on-Kubernetes workloads. Leveraging Cloudera's security, management, governance, and data cataloging capabilities, this offering eliminates the need to move or duplicate data and does away with technological silos associated with managing external data science technologies.

- Cloudera is building technical expertise around several highly impactful market use cases, including IoT-enabled predictive maintenance, genomics research, fraud detection, and compliance monitoring. Central to these use cases is the ability to provide real-time analysis. The company has been working on this for some time now, surfacing in solutions like Cloudera Dataflow (CDF), a real-time streaming data platform that ingests, curates, and analyzes data in real-time across edge applications and other streaming sources. With differentiated capabilities like data tracking and lineage for streaming data, CDF enables Cloudera to carve out some very lucrative opportunities amidst a crowded market.

Limitations

- Cloudera had undergone significant financial upheaval over the past few years, culminating in June 2021, when the company announced that affiliates of Clayton, Dubilier & Rice (CD&R), and KKR would acquire it and revert it to operating as a privately held company. Building on this move, Cloudera announced in March 2022 that it had made several senior management changes to increase innovation and speed growth. With companies like Dell Computer as a guide, this approach has proven effective for companies seeking to overcome market transitions. However, this approach will also burden Cloudera to prove its worth as a long-term investment, even in the age of annual subscription contracts.

- Among pure-play vendors reviewed in this report, Cloudera enjoys one of the broadest portfolios with many differentiated CDP configurations covering data warehousing, data science, operational database, data flow, and data engineering capabilities. However, the vendor's approach to licensing and provisioning favors a very hands-on sales process with few opportunities for self-service. This philosophy even extends to the vendor's 60-day free trial. While customers choosing to download CDP software can do so on their own, customers selecting to run on top of a supported public cloud platform must wait for direct contact from Cloudera, for assistance, in provisioning the free trial. With the launch of CDP as a service in a couple of months, Cloudera plans to provide self-service evaluation for its potential customers so that more companies can try CDP without assistance. As a result, this will help widen the company's reach outside our existing customers.

- Cloudera may enjoy a long history of addressing significant data challenges. However, the vendor has fallen behind its rivals in natively supporting new and emerging analytic workloads. For example, while CDP enables users to query and chart time-series data, the platform does not incorporate such support for geographic or graph data types. For such help, Cloudera looks to its partner ecosystem, enabling users to select best-of-breed solutions. Still, those solutions function outside the umbrella of CDP, requiring Cloudera and its partners to work together to ease any integration pains.

Outlook

Cloudera has undoubtedly been on a rollercoaster ride since its founding in 2008, on through its IPO in 2017, followed shortly after that by a mega-merger with Hadoop rival Hortonworks in 2019, and recently a return to private status in 2021. Now under a new management structure, Cloudera has set some lofty goals for itself, seeking to lead the way in today's multi/hybrid cloud era with a truly cloud-native data analytics platform that spans the entire data lifecycle—data engineering, data warehousing, transactional data, streaming data, data science, and ML. If history is any guide, Cloudera is making the right moves here by creating some space for innovation as a private firm. In Omdia's opinion, success will depend on innovation and exceeding customer expectations regarding the basics: common security, governance, metadata, replication, and automation. The company's recent partnership with data connector specialist, Datacoral, as but one example, speaks directly to this need and makes room for more future-leaning internal engineering efforts.

Databricks (Omdia recommendation: Leader)

The Databricks Lakehouse Platform should appear on your shortlist if you want a feature-rich and flexible analytics platform steeped in open source and built specifically to unify data warehouse and data lake workloads.

Overview

Though younger than some of the data warehousing giants reviewed in this report, Databricks nevertheless casts a hugely influential shadow across the analytics and data management marketplace. The original creators of Apache Spark and MLflow, two highly influential open-source projects, founded Databricks in 2013. The company launched Databricks in 2015, which was itself an implementation of Apache Spark, to make big data (usually a combination of Hadoop, HIVE, and Spark) not only performant but also accessible to mere mortals.

Extending this foundation, the company introduced the Delta Lake open-source project in 2019, leveraging its expertise in Spark and big data to simplify and democratize analytical data access. With Delta Lake serving as a core, unifying data layer riding on top of many different data lake architectures, Databricks has since built up a sizable portfolio of solutions under the banner, Databricks Lakehouse Platform, a portmanteau combining the terms data warehouse and data lake. The company's platform currently incorporates several closely aligned products:

- Delta Lake
- Data Engineering

- Machine Learning
- Data Science
- SQL Analytics

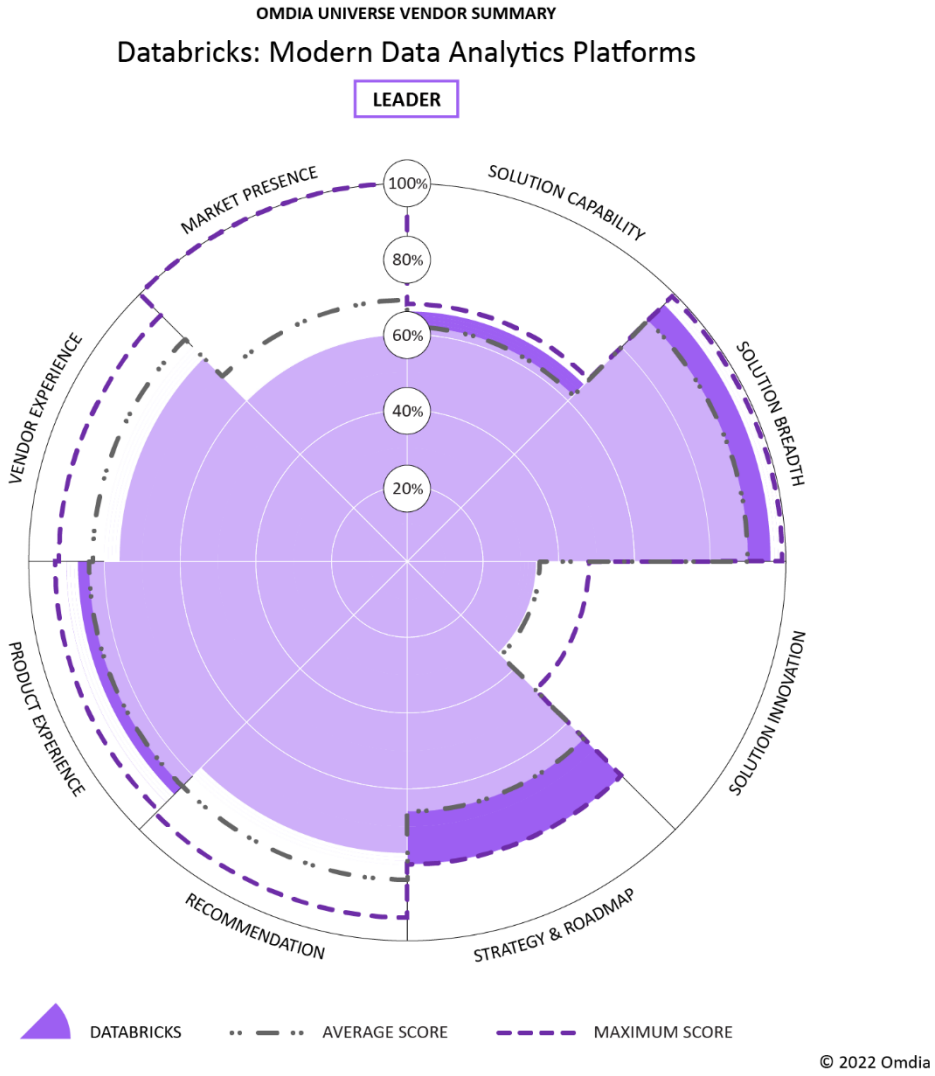
With more than 7,000 customers globally—including J.B. Hunt, Shell, Comcast, H&M, and SEGA—Databricks enjoys a unique position within the marketplace as an independent vendor that does not just run across the three largest cloud platforms (AWS, GCP, and Microsoft Azure). It also benefits from a close partnership with those platform providers with direct and indirect funding.

From a technical perspective, Databricks employs a similarly unique approach, promoting a multicloud platform that provides data warehousing performance, structured semantics, a cross-functional SQL interface, and native data science functionality, all with the economics usually associated with a data lake. However, unlike a data lake, Databricks leverages Delta Lake's Spark prowess to provide ACID transactions, scalable metadata handling, strong governance, and unified streaming and batch processing.

Roadmap priorities

- Transfer many of the benefits found in Databricks to the core open-source platform, Delta Lake, centering on performance and tools to automate workload optimization out of the box
- Embrace important open-source innovations as with the company's work to integrate the popular data pipeline orchestration tool, dbt, into Databricks, which will speed up Photon-enabled workloads and introduce new project orchestration capabilities
- Adopt core CI/CD DevOps practices such as full observability using internal or external tools like Prometheus or Grafana
- Forthcoming functionality will bring together audit logs, query history, real-time stats and alerts, and other measures from across the platform

Figure 14: Omdia Universe ratings — Databricks



Source: Omdia

Strengths

- Databricks is riding right on the current market wave of interest in conjoined data warehouse and data lake functionality, termed the “data lakehouse.” Databricks was early to embrace this trend; an argument can be made that the company invented the term itself, which it now markets as the Databricks Lakehouse Platform. Building on this architecturally unified platform, which provides total ACID transactions for both structured and unstructured data, Databricks has been able to successfully enter and influence the highly lucrative data science marketplace, leveraging its open-source tools, MLFlow and Spark.

- Databricks punches well above its weight in research and development (R&D), introducing leading-edge capabilities such as Delta Live Tables. Now generally available on Microsoft Azure and AWS (GCP will follow shortly), this new service simplifies and operationalizes the process of converting SQL queries into production-ready ETL pipelines. Likewise, with an eye to the future, the company recently introduced MeshaVerse, an augmented reality (AR) interface to the company's lakehouse platform that enables users to work together in a distributed yet fully managed data mesh environment. This tool will enable users across large organizations to more readily find and access data usually locked in data silos.

- Databricks occupies a unique position within the modern data warehouse solutions, operating as an independent provider capable of running seamlessly across all significant hyperscalers while also enjoying the patronage and support from those same hyperscale cloud providers. This gives Databricks the flexibility enjoyed by Snowflake in offering cloud independence. However, unlike Snowflake, Databricks' close alliance with AWS, Google, and Microsoft, enables the company and its customers to operate in a more congenial cooperative atmosphere.

Limitations

- Though Databricks has built a genuinely unified data warehouse and data lake experience, the company has had to make some architectural commitments that favor the underlying data lake file system architecture. For example, because Databricks Delta Lake uses the Parquet file format on top of an object store like the AWS S3, concepts like multi-table transactions, on-the-fly column changes (e.g., drops, name changes), and foreign key queries are more difficult to implement, compared with traditional tabular databases that do not favor transactions at the table level.

- As the early driven and leading proponent of the increasingly popular data lakehouse concept, Databricks occupies a highly enviable position among its rivals, all chasing the same idea of a single analytics data layer. However, operating from this position of strength creates risks and challenges for the company. First, should the concept of the lakehouse fall out of fashion, the company's commitment to that concept will quickly turn from an advantage to a weakness. The same holds true for the company's ownership of Apache Spark itself. Should that technology fall out of favor, as seen with technologies like Hadoop, Databricks will be challenged to evolve its platform architecture to accommodate that market shift.

- Relatedly, Databricks is fortunate because its platform can run on all major hyperscale platforms (Alibaba, AWS, GCP, and Microsoft Azure) in a relatively non-competitive state with those platform providers. Even so, Databricks' dependency on those platform providers can create a long-term challenge for the company as its host providers work to capture the same market use cases and to do natively, not as a third-party product.

Outlook

Databricks finds itself in a highly desirable position within the marketplace. It has built a platform ideally suited to current market demands for a unified analytics platform capable of unifying data engineering, analytics, and ML within a feature-rich and flexible architecture steeped in open source

and built specifically to unify data warehouse and data lake workloads. So long as the company can maintain its value proposition as an independent player renowned for innovative engineering, Databricks will continue to hit well above its weight in helping companies unify their analytics workloads.

Google (Omdia recommendation: Challenger)

Google's BigQuery should appear on your shortlist if you want a highly scalable and feature-rich data warehouse featuring minimal operational overhead and ready access to a broad set of supportive services.

Overview

Google's fundamental mandate of organizing the world's information and making that information universally accessible and useful is nowhere more apparent in product form than in the company's modern data warehouse solution, BigQuery.

Introduced in 2010 as a foundational element of GCP and based on the company's Dremel distributed SQL engine, BigQuery was created as a web-scale analytics platform that could match pace with Hadoop technologies but does so using the industry-standard SQL query language. Since its introduction, BigQuery has steadily grown in both capability and scope. Building on these big data roots, Google now markets BigQuery as a multicloud, serverless, massively scalable, cost-effective data warehouse that delivers business agility.

As one of the "big three" hyperscale cloud platform providers, Google operates from a distinct position of strength in marketing BigQuery, as evidenced by a global customer base for GCP that numbers in the tens of thousands. Key customers of note for BigQuery itself are Procter & Gamble, Major League Baseball, and *The New York Times*. BigQuery is used across various horizontal workloads, including data warehouse modernization, customer analytics, fraud detection, streaming analytics, data lake modernization, BI, geospatial analytics, and data sharing. In support of these workloads, BigQuery incorporates tools and functionality specific to many user personas, including data analysts, data engineers, data architects, data scientists, and business decision makers.

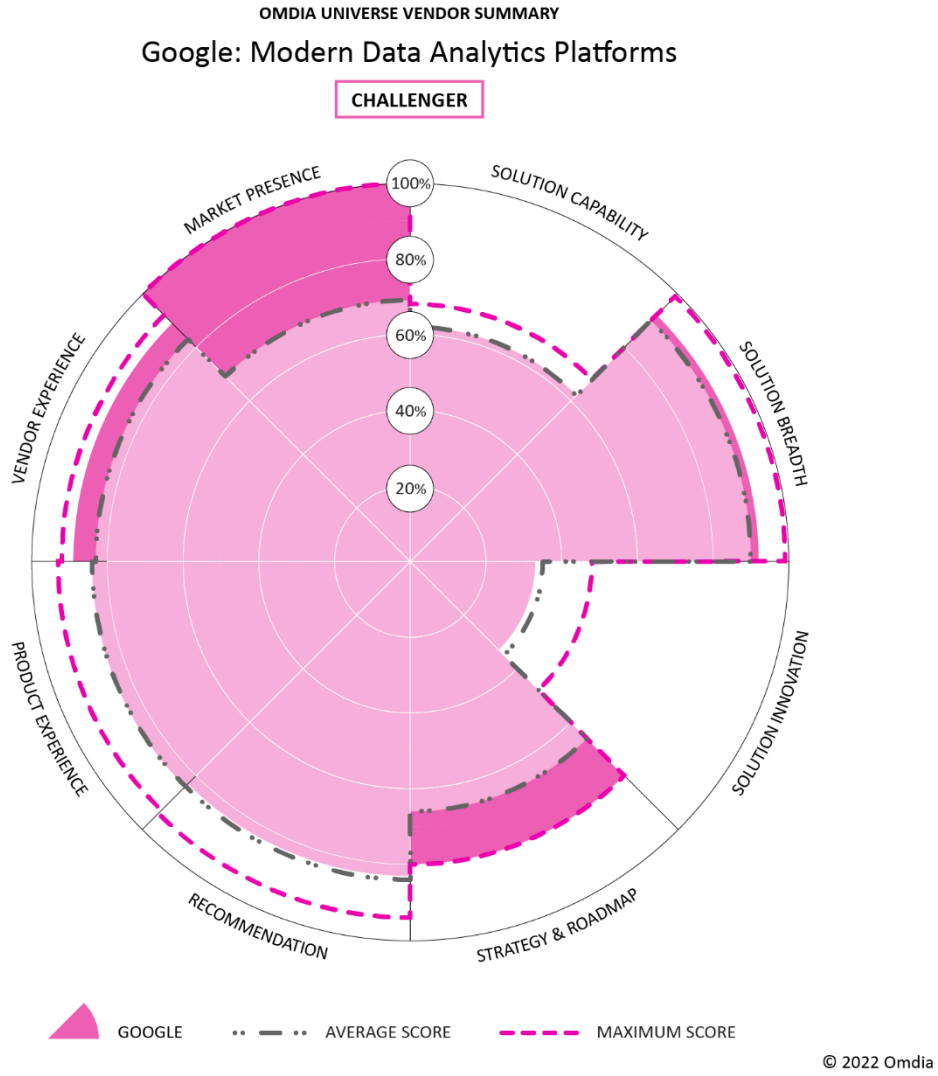
In line with the broader market, Google customers do not have to provision BigQuery via individual instances or virtual machines. Instead, users can choose to pay for computing on-demand or reserve computing capacity ahead of time and receive discounts based on longer terms of commitment. The company also offers a permanently free tier for BigQuery, which it leverages to support its educational resources and community services, most notably Kaggle.

What sets BigQuery apart from many offerings in this review is its central role in supporting Google's broader Data Cloud platform, which includes data integration and processing, analytics, and BI services. Yet, Google has not stripped BigQuery down to force users to purchase ancillary services. Instead, the offering incorporates in-database data processing engines and tools supporting ML and BI workloads. With the BI Engine SQL interface, for example, BigQuery natively delivers distributed, in-memory, in-database analytics services in support of several downstream consumers across API, ETL pipelines, or SQL queries.

Roadmap priorities

- Integrate services further in support of streaming analytics and ML pipelines
- Provide more direct support for data lake use cases through the introduction of BigLake (in preview as of April 7, 2022)
- Improve multicloud capabilities within BigQuery Omni between GCP and rival platforms, AWS and Microsoft Azure
- Offer cross-organizational sharing of data and analytics through BigQuery-based Analytics Hub

Figure 15: Omdia Universe ratings — Google



Source: Omdia

Strengths

- As a part of Google's push to lower operational overhead for BigQuery, the company aggressively incorporates ML into the product, automating the running and optimization of both database infrastructure running analytics queries. Even so, Google does not treat BigQuery as a black box. Instead, the product includes a detailed diagnostic query plan and timing information, broken down into granular query stages. These stages can be viewed in a console or processed via an API call, allowing users to work cooperatively with Google's automated optimizations to

perform further syntax tuning. The objective is to help users transparently optimize their database spending.

- Google BigQuery has been built to serve as a core component of the company's data cloud, driving sales for tools like Looker, Vertex AI, and Cloud Spanner. However, the synergy between BigQuery and Google's broader portfolio is not just in name or licensing alone. Instead, the company has seamlessly incorporated external platform functionality into the database via seamless data visualization with Looker, Google Data Studio and Datalab, and Google Sheets. Moreover, Google is actively building core functionality into BigQuery, which is typically provisioned as an external service. This is the case with automatic Data Loss Prevention (DLP), through which BigQuery users can now automatically identify sensitive data as it enters the database without user intervention.

- Though BigQuery does not run on-premises as a self-managed service, Google has aggressively built out multicloud support for BigQuery through its BigQuery Omni service. This service allows BigQuery users to run queries against AWS S3 or Microsoft Azure Blob Storage services without copying or moving data. Building on BigQuery's ability to separate storage from computing, BigQuery Omni lets users run the database's query engine on compute clusters located on multiple, disparate cloud platforms. Additionally, like BigQuery itself, these clusters are all managed/provisioned transparently by GCP as fully serverless resources. This approach lets Google users transparently access data silos on competing cloud platforms without changing working methods.

Limitations

- From the outset, BigQuery was built as a platform as a service (PaaS) offering. As such, it does not enable customers to take full stewardship of the software itself. Google affords many privacy, security, and sovereignty options regarding the underlying data. For instance, there are provisions for managing data sovereignty via VPC networks. However, the database can only be purchased as a multi-tenant, fully-managed solution. With many independent data warehouse competitors offering both self- and full-managed options on-premises and across multiple clouds, Google BigQuery can be positioned as a comparatively hands-off option with limited opportunities for customers to lock down their implementations within their own data center, controlling when to apply fixes, updates, and security patches, for example.

- Though Google BigQuery natively supports semi-structured data, the solution is now beginning to address market demand for fully unstructured data lake functionality. Previously, customers could remotely access data lakes on AWS and Microsoft Azure through BigQuery Omni, Google Dataproc, and Google's partnership with Databricks. Extending this capability, Google has introduced Google BigLake. Available unsupported as a preview release, BigLake works as a storage engine that unifies data warehouses and lakes. This move brings Google alongside fully early lakehouse proponents, including Databricks, Snowflake, and AWS. However, given the company's propensity to maintain new features in an unsupported manner (Preview), rivals can position Google's portfolio as still in progress within this important competitive arena.

- While Google BigQuery can support semi-structured data through direct support for JSON documents and geospatial data via SQL-specific geographic data types, the product does not yet provide native support for other non-relational data types. Notable gaps include both graph and time-series data and analytics. However, BigQuery users can write user-defined functions (UDF) to approximate some of this functionality. Regardless, because Google has positioned BigQuery as its central data warehouse platform, the vendor will need to bring these technologies to the foreground within BigQuery to remain competitive within a growing market for multi-model databases.

Outlook

Google is well-positioned for the long haul, with BigQuery operating as the vendor's central analytics database. The company can position BigQuery as the data powerhouse behind its substantial data science capabilities with Google Vertex AI and offer BigQuery as a data science platform in its own right via in-database ML capabilities found in BigQuery ML. This same pattern extends to BI workloads and emerging requirements around data cataloging, Spark data processing, and data exchanges. In adding support for these important areas, positioning BigQuery as the engine behind and sometimes in front of new functionality, Google will be able to maintain its influential position in an increasingly infrastructure-agnostic market.

IBM (Omdia recommendation: Challenger)

IBM Cloud Pak for Data should appear on your shortlist if you want a solid performer capable of meeting a wide range of use cases, particularly in support of hybrid/multicloud deployments.

Overview

No stranger to the demands of delivering data-driven insight, IBM has been in the business of selling highly scalable databases for more than 40 years, beginning with its DB2 line of relational database management systems on the IBM mainframe in the early 1980s. Since then, the DB2 database has undergone many changes in both scope and branding, currently finding a home within IBM's tightly integrated suite of services tailored to building a data fabric across siloed data sources, even across hybrid public and private clouds.

In support of data warehousing use cases, IBM introduced DashDB in 2015 as a high-speed, columnar analytics database. It was equipped with in-memory processing capabilities and built to run as a fully managed cloud service. Two years later, the company renamed DB2 as Db2 and rebranded DashDB as Db2 Warehouse and Db2 Warehouse on Cloud. Then in 2019, the company introduced the unified IBM Cloud Pak for Data, a pre-integrated data and AI platform that runs natively on the Red Hat OpenShift Container Platform, running on top of IBM Cloud, AWS, Microsoft Azure, GCP, and any OpenShift cluster, whether on the public cloud or on-premises.

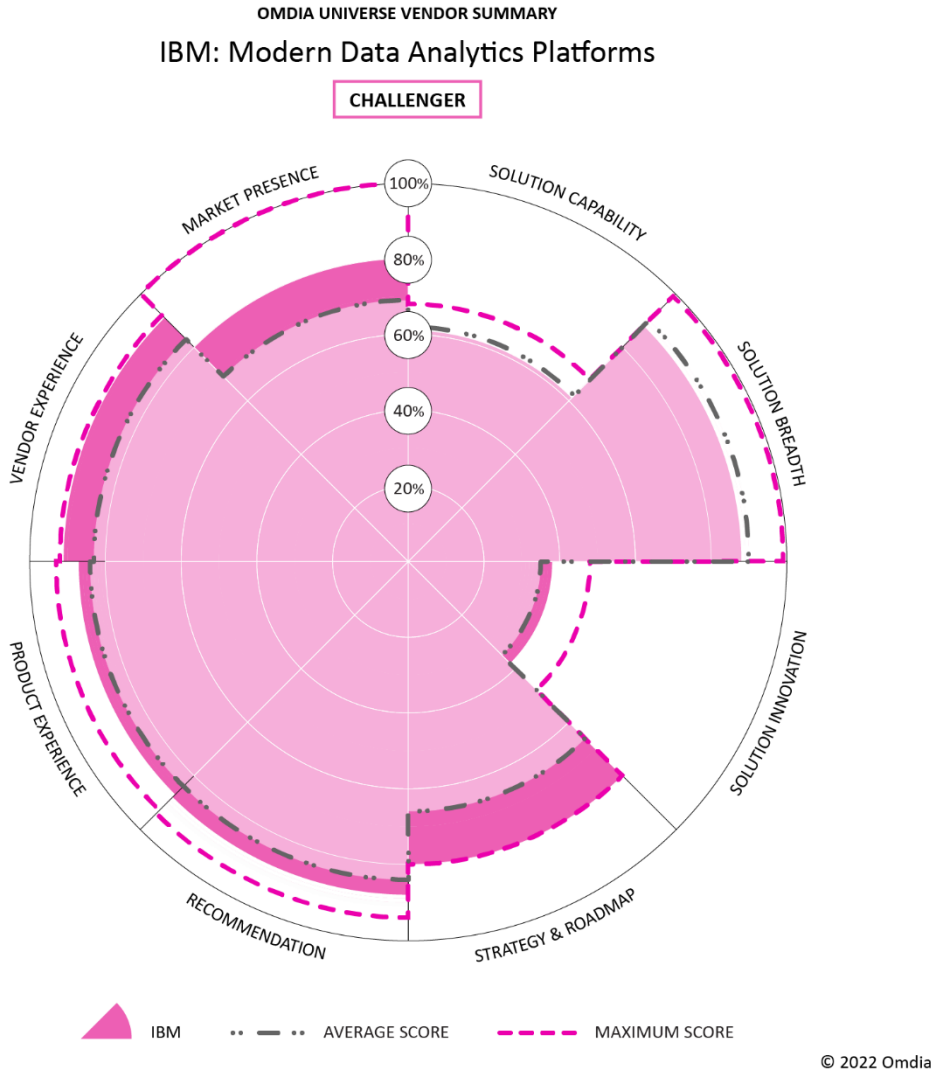
As a unifying container for several synergistic analytics services, IBM Cloud Pak for Data brings together several IBM databases, including Db2, Db2 Big SQL (a SQL-on-Hadoop engine), and Db2 Warehouse. It also unifies third-party solutions such as MongoDB, PostgreSQL, and Elasticsearch. Further, IBM Cloud Pak for Data enables these to integrate seamlessly with numerous data and analytics tools from IBM, featuring AutoSQL (a new, AI-fueled query engine), Cognos Dashboards, SPSS Modeler, and Watson Knowledge Catalog—all within a single management pane.

Db2 Warehouse on Cloud, the principal product under review, functions as a fully-managed, elastic, cloud-native data warehouse that separates storage and compute. Architecturally, the database operates as a columnar data store featuring several synergistic features, including data compression, in-memory processing, data skipping, and single instruction, multiple data processing (SMID)—collectively referred to as IBM BLU Acceleration—to speed query response times at scale and under high concurrency loads. IBM grounds this performance in solid administrative capabilities such as a two-pronged approach to data protection with built-in snapshot backups and geo-replicated disaster recovery backups.

Roadmap priorities

- Simplify user experience in enabling add-on products (REST, Graph, Data replication) via the IBM Cloud Pak for Data user experience
- Offer Schema Level backup and restore services to help users perform backups at the schema level and restore at the table level
- Provide support for Db2's auditing features within Db2 Warehouse on Cloud

Figure 16: Omdia Universe ratings — IBM



Source: Omdia

Strengths

- The role played by IBM Cloud Pak for Data is rapidly evolving. No longer a convenient platform capable of simplifying the purchasing, integration, and management of several related IBM technologies, IBM Cloud Pak for Data can now be considered a drop-and-go data fabric. The product's recent additions, AutoSQL and AutoCatalog, best exemplify this; they work in tandem with tools like Watson Knowledge Catalog to unify disparate data sources via metadata. More than this, because ML algorithms designed to support self-service heavily inform these tools, they enable companies to create a functional data mesh on top of a data fabric that fully unifies

data silos spread across company and even across cloud platforms, enabling users to take ownership of their data wherever it resides without sacrificing central control.

- When it comes to supporting hybrid/multicloud user requirements, IBM maintains a competitive advantage among many rivals in going far beyond fully managed cloud services. First, because IBM Cloud Pak for Data runs entirely within the Red Hat OpenShift containerized services, the solution can be purchased on a month-to-month basis (per virtual processor core) and hosted as a self-managed service across a wide array of platforms, including other hyperscalers like AWS, where the solution can directly access numerous low-level services like virtual private clouds and elastic block storage. Second, building on IBM's legacy of on-premises hardware (PureData Systems and Netezza), IBM Netezza Performance Server for IBM Cloud Pak for Data offers users a single solution that is both an on-premises engineered system and a cloud service running across IBM Cloud, Microsoft Azure, and AWS.

- As an early proponent of and contributor to the Apache Spark open-source project, IBM extensively uses this important data processing and analytics engine within IBM Cloud Pak for Data, particularly in supporting data science workloads. For example, the company uses Spark to create a machine learning model, saving the evaluation information to the Db2 Warehouse on Cloud. Spark can also be used in this context for data transformation, data science, or ML using just Spark job APIs, allowing users to work in conjunction with or entirely outside IBM's Watson Studio service.

Limitations

- IBM Cloud Pak for Data is one of the more portable solutions in this review, thanks to its use of Red Hat OpenShift containerization. This allows customers to self-manage IBM software across various public and private cloud configurations. Customers can install Db2 Warehouse on Cloud on its own without IBM Cloud Pak for Data across disparate public clouds. However, IBM remains comparatively isolated in delivering managed software across global hyperscale platforms, as it is available only as a fully-managed public cloud service on IBM Cloud and AWS. Further, some of the company's solutions are not broadly or consistently available as with IBM's Netezza Performance Server on Microsoft Azure, which is only available in two data centers—one in North America and one in Western Europe. Note that with IBM Cloud Satellite, IBM-managed software from across disparate cloud platforms can access IBM Cloud Pak for Data.

- On a licensing level, IBM Cloud Pak for Data does a great job of simplifying the consumption of numerous interrelated IBM products. When purchased as a fully managed service on IBM Cloud, IBM Cloud Pak for Data somewhat does away with products in favor of use cases like querying data or building a dashboard. In working with IBM Cloud Pak for Data as a self-managed service, however, IBM's comprehensive portfolio of services that support and work alongside Db2 Warehouse on Cloud requires a comparatively high degree of customer involvement in taming the complexities involved in making use of the full range of available features. Moreover, relatively limited documentation and fragmented cloud availability slow the exploration and adoption of IBM Cloud Pak for Data among new customers. This is particularly apparent with the company's free tier on IBM Cloud, which only gives new users an IBM Cloud credit of $200 to use

towards Db2 Warehouse on Cloud. Without a long-term free tier option for Db2 Warehouse on Cloud, IBM will have difficulty attracting new customers just beginning their analytics journey.

- IBM Cloud Pak for Data combines several differentiated capabilities, such as IBM's new AutoSQL query engine, making it easy for users to scale queries across data lake, data warehouse, and transactional databases. However, Db2 Warehouse on Cloud remains comparatively rooted in orthodox data warehousing functionality. For example, though Python and R are both supported, they are not well-integrated alongside SQL in support of analytical queries. This goes for cloud-native functionality as well. For example, while customers can experiment with select serverless cluster functionality within their self-managed instances of IBM Cloud Pak for Data, fully managed implementations do not offer automated scale-to-zero functionality for Db2 Warehouse on Cloud resources.

Outlook

With a robust and scalable data warehousing core beating at the heart of its sweeping IBM Cloud Pak for Data platform, IBM is well-positioned to take on today's modern data and analytics challenges. This is particularly evident in the company's platform-agnostic approach to containerization based on Red Hat's OpenShift and its rapidly maturing emphasis on metadata as a means of not just unifying data silos but enabling true self-service data ownership across those silos (e.g., a data mesh). To remain competitive, IBM will need to modernize Db2 Warehouse on Cloud both functionally and in how customers consume this important data architecture component, making it more ubiquitous and adaptive as a managed cloud service.

Microsoft (Omdia recommendation: Leader)

Microsoft's Azure Synapse Analytics should appear on your shortlist if you want a highly flexible, cloud-native platform that extends traditional data warehousing to encompass big data AI and analytics.

Overview

Microsoft is one of the global giants within the analytics and data management market, thanks in no small part to its long history in delivering market-leading services and solutions for enterprise IT professionals. Plying its expertise and influence in supporting developers with services like GitHub for software version control, Power BI for business intelligence, and SQL Server for data storage and processing, Microsoft finds itself advantageously positioned to build upon these existing affinities to accelerate many enterprise modernization and transformation projects through the addition of Azure Synapse Analytics.

Introduced as a successor to SQL Data Warehouse in 2019, Azure Synapse Analytics combines data integration, data warehousing, log and telemetry analytics, and big data analysis capabilities. Like most solutions in this report, it serves many disparate downstream consumers, including BI and ML practitioners. Functionally, the service includes a significant runtime for SQL, leveraging Microsoft's proprietary SQL extension, Transact SQL (T-SQL). It also consists of an Apache Spark runtime for data integration and processing. These runtimes employ Microsoft Azure Data Lake Storage Gen2 as a storage layer and can draw from a wide array of cloud and on-premises data sources. The primary

tool used to interact with these runtimes and associated tools is Azure Synapse Analytics Studio, which includes data integration, SQL development, solution management, and monitoring capabilities.

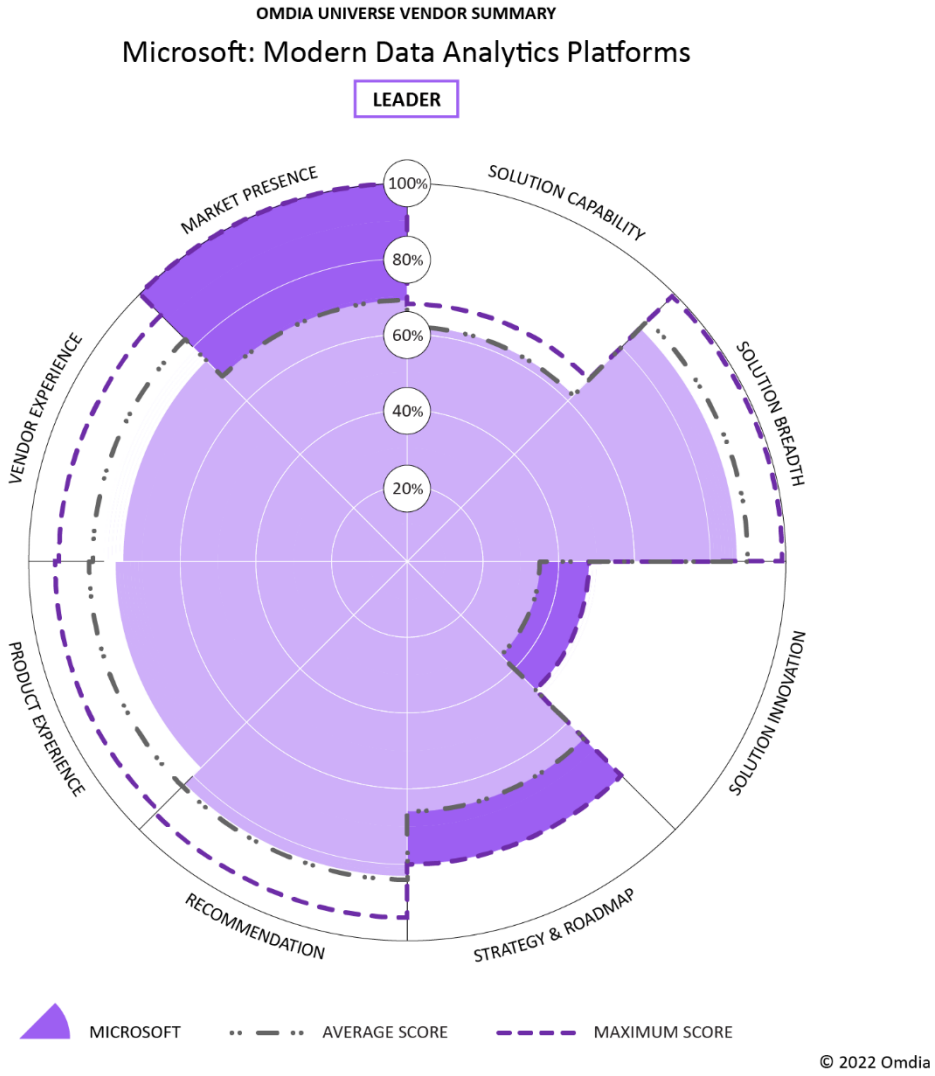
Microsoft's strategy in bringing Azure Synapse Analytics to market hinges on one simple premise: leverage the economies of scale and rich portfolio of services available on Microsoft's global cloud platform, Azure, to provide customers with what the vendor terms "a limitless analytics service." The "limitless" aspect of this offering stems partly from its flexible deployment options and scalability. The solution, for example, can be run in either dedicated or serverless resource models across Microsoft Azure's global footprint, scaling out as required. This idea of limitless analytics comes from the product's synergistic position within Microsoft's broader Azure platform portfolio. Microsoft's Azure Synapse, for example, integrates tightly with and benefits from several Azure services, not only those mentioned above but also Azure Purview (a data governance tool), Azure Machine Learning (for data science workloads), and Cosmos DB (a multi-modal distributed database).

With these capabilities at the ready, Azure Synapse Analytics is a fully self-contained data warehouse capable of supporting many modern analytical data workloads. It is comparatively simple and relies upon external solutions for advanced functionality such as data governance. Even so, its simplicity is a strength, shining in its ability to combine disparate functionality between SQL and Spark environments seamlessly. For instance, users can work entirely within T-SQL across structured and semi-structured data. Still, they can just as readily work within task-specific languages like KQL, Python, Scala, Spark SQL, or .Net. without penalty. Such functionality makes Azure Synapse Analytics an easy decision for Microsoft Azure customers looking to set up a competent analytics platform without much complexity.

Roadmap priorities

- Allow organizations to invest more time creating value with the new Microsoft Intelligent Data Platform that fully integrates databases, analytics, and governance
- Provide continued and increasing alignment of Power BI with Office 365, Microsoft Teams, Excel, and SharePoint
- Accelerate time to market for customers by creating a library of comprehensive data models, each supporting specific data requirements for use cases such as genomics, manufacturing, pharmaceuticals, and automotive industries

Figure 17: Omdia Universe ratings — Microsoft



Source: Omdia

Strengths

- Microsoft's broad investments in AI put the vendor in an enviable position competitively. It can provide a seamless experience for data scientists who do not rely upon third-party integrations' support for complex or demanding workloads. For example, users can splice in Azure Machine Learning pipelines as a fully native and first-class pipeline workflow within Azure Synapse Analytics (and Azure Data Factory). If they work entirely in Azure Synapse Analytics, they can use the T-SQL PREDICT function using PySpark, Scala, or .Net libraries. As with rivals AWS and

Google, Microsoft makes it easy for data warehouse users to access underlying AI hardware acceleration resources.

- Microsoft leads the way among competitors reviewed in this report regarding developer influence. The firm actively leverages market-leading tooling like Visual Studio and the world's largest source code repository, GitHub. The vendor puts these assets to work in supporting Azure Synapse Analytics, creating a highly differentiated level CI/CD orchestration. For example, All Azure Synapse Analytics Pipeline artifacts, notebooks, SQL scripts, and Spark job definitions are transparently managed in Git. Relatedly, Microsoft enjoys a very strong Apache Spark story with this product, seamlessly integrating Apache Spark's big data engine for data preparation, data engineering, ETL, and ML.

- One of Microsoft's greatest comparative strengths is its ability to remove barriers to adoption. This ability spans several areas, such as cloud migration supported by the vendor's aggressive Software Assurance licensing program, which preserves the customer's on-premises investments online. Another and perhaps more impactful area concerns self-service analytics. The company has done a great job of seeking out and filling all the gaps in setting up an analytics pipeline with drag-and-drop functionality and step-by-step guidance. By combining this approach with full customization and code-driven operation, the company has created a highly adaptable platform that meets customers where they are both regarding size and experience.

Limitations

- Microsoft has built Azure Synapse Analytics as a very consistent, end-to-end analytics experience in no small part thanks to its use of Apache Spark as a foundational technology. This, in conjunction with Azure Synapse Analytics, provides many advantages, such as streamlined access to Spark's parallel data processing capabilities. However, from a software development perspective, this decision, coupled with the vendor's use of its proprietary SQL dialect (T-SQL), paints the overall solution as being open-source friendly but at the same time constrained by Microsoft's selection of tools. For example, users looking to stay within Azure Synapse Analytics for data must use Microsoft's rendition of Jupyter notebooks (Synapse Studio) and write using PySpark for Python, Spark for Scala, Spark SQL, and .NET Spark for C#.

- In addressing hybrid and multicloud opportunities, Microsoft and its hyperscalar cloud competitors AWS and Google are actively building the means to run and manage their software across other public and private cloud platforms. For Microsoft, this centers on Azure Arc, which is off to a strong start in supporting operational data workloads with support for Azure SQL and PostgreSQL (preview). However, the vendor has not yet extended this to include analytics workloads with support for Azure Synapse Analytics. This leaves Microsoft playing catch-up with rival Google, which has been working on the same since 2020 with BigQuery Omni, powered by Anthos.

- While Microsoft excels at simplifying adoption through self-service functionality within Azure Synapse Analytics, the solution pulls up short in optimizing the underlying platform and harmonizing with Microsoft's broader data assets. The platform offers some performance

automation, enabling the use of automatic scale up/down on a node-by-node basis. Still, users do not yet have access to full-on automatic performance optimization at the query management level that they have at the ready within other products like SQL Server. Regarding harmonization, the vendor is just beginning to leverage the breadth of its database portfolio beyond basic connectors. With Azure Synapse Link, customers can meld operational and analytics workloads with Microsoft's Cosmos DB database. Thus, Microsoft released the public preview of Synapse Link for SQL in support of SQL Server 2022 and Azure SQL Database.

Outlook

Microsoft offers highly flexible, cloud-native data that extends traditional data warehousing to encompass big data AI and analytics and does so smartly through a highly unified, self-service user experience. With strategic advantages as a leader in AI and software development, Microsoft's biggest challenge going forward will come from within as the vendor seeks to fully leverage its broader portfolio and bulk up its internal functionality. If the company's recent move to improve debugging capabilities within Synapse notebooks is a measure, customers can trust Microsoft to meet their needs in the long term, especially as the company builds out Azure Synapse as an integral component within its broader work on Microsoft Intelligent Data Platform.

Oracle (Omdia recommendation: Leader)

Oracle Autonomous Database should appear on your shortlist if you want a uniquely self-governed data and analytics platform capable of consistently handling a wide array of workloads across hybrid cloud/on-premises deployments.

Overview

Oracle had always forged a unique path in the enterprise technology marketplace, dating back to its early days in 1983 when the company introduced a highly portable and scalable relational database (written in the C programming language). Backed by highly influential customers featuring US government contracts, the company quickly went public in 1986 and found itself in the enviable position of the world's largest database management company just one year later. Since then, Oracle has never looked back on leveraging its commanding market position to expand its scope of operations through engineering and many high-profile acquisitions.

Building on its flagship database offering—Oracle Database—and its experience in delivering high-performance hardware—stemming from its acquisition of Sun Microsystems in 2010—Oracle is currently in the process of carving out a unique proposition within the data and analytics marketplace. It is rapidly shifting to a new, fully self-administering, cloud-native, operational, and analytical database, built to run in the cloud or on-premises on the same optimized hardware/software stack and heavily automated using both heuristics best practices and ML modeling.

Product-wise, the company re-engineered Oracle Database for the cloud by creating a modern, cloud-native, self-driving database—Oracle Autonomous Database, and paired it with Oracle Exadata (based on hardware acquired from Sun Microsystems) available on-premises and in the cloud. Within this solution, customers can find Oracle Autonomous Database for analytics and data warehousing—autonomous data warehouse (ADW), the focus of this review—and autonomous

transaction processing (ATP), creating a seamless experience across analytic and transactional workloads. Regarding deploying Oracle Autonomous Database, customers can select Oracle Cloud Infrastructure (OCI) for shared or dedicated deployments or opt for an on-premises approach via the company's Exadata Cloud@Customer or Dedicated Region Cloud@Customer offerings. Therefore, this makes Oracle a comparatively unique competitor in offering architectural equivalency across the cloud and on-premises.

On top of this highly consistent hybrid cloud/premises architecture, Oracle positions ADW as a converged analytics engine attuned to a broad range of data types, natively supporting structured, unstructured, graph, time series, and geospatial data. The offering includes built-in tools specific to these data types, including data engineering and ETL processing, AI modeling, and traditional data analysis and reporting.

Like other solutions in this review, ADW emphasizes elastic scale to deliver fast query performance on demand without any user intervention. Like its competitors, ADW supports a wide range of analytic workloads—albeit Oracle offers these all within a single database:

- Enterprise data warehouse
- Data lake
- Departmental data warehouses
- Data science and ML platform
- Data and IoT event streams

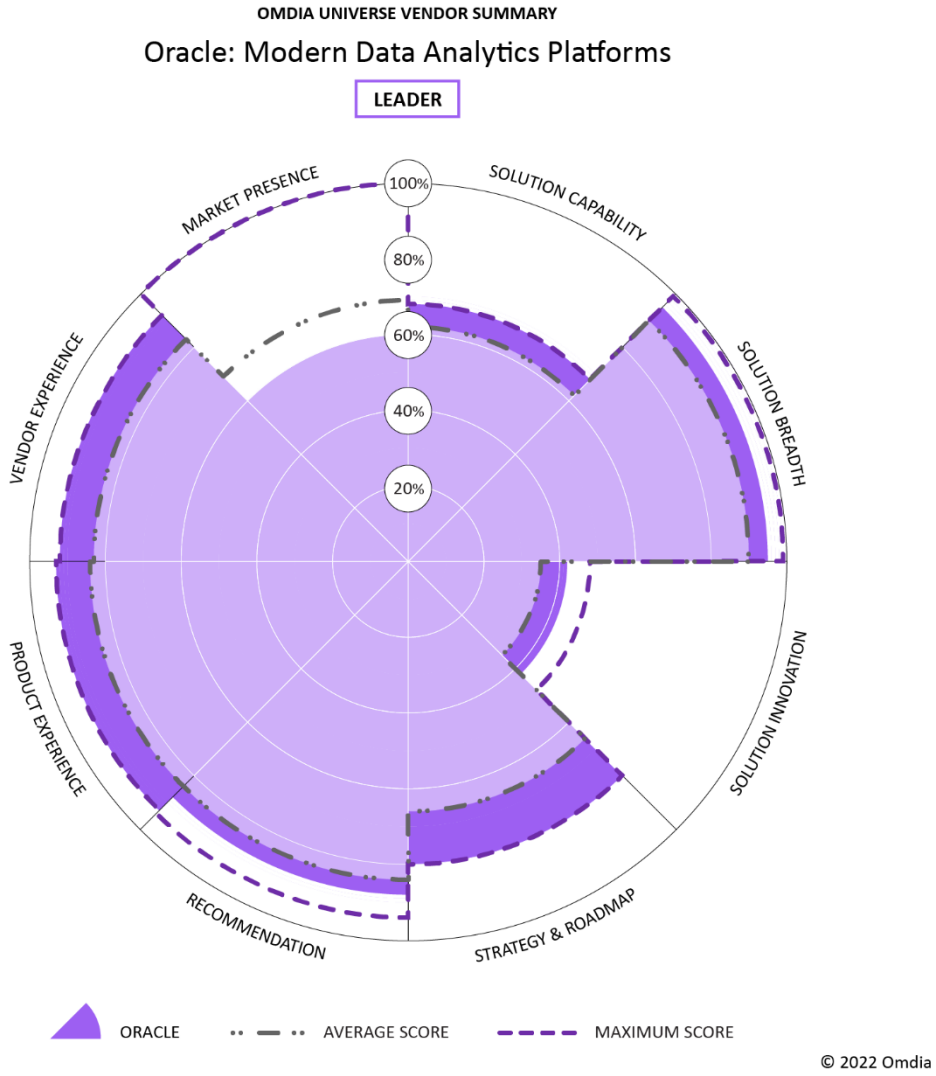
To accommodate data lake use cases, for example, the solution can query and load data from an array of object stores, including Oracle Object Store, Amazon S3, Microsoft Azure Object Store, and GCP. Also, because Oracle can manage data in an object store as external tables, ADW can seamlessly save changes back to those object stores. This way, Oracle's approach aligns with many of the solutions under review. However, where the company significantly diverges from its peers, who still have not offered an alternative, is in its approach to reducing database management overhead through autonomy. Unlike many solutions that encode best practices or use basic heuristics alone, Oracle combines those with ML modeling to provide true autonomous capabilities. For example, the database can automatically protect sensitive and regulated data, apply security vulnerability patches and feature updates with zero downtime, and prevent unauthorized access, all with zero user intervention. These autonomous capabilities fall into three primary areas:

- **Self-driving** - provisioning, tuning, scaling, and backup
- **Self-securing** - patch application and vulnerability detection
- **Self-repairing/healing** - index maintenance and compiling slow routines

Roadmap priorities

- Take on more heterogeneous data management opportunities by extending its data catalog to incorporate metadata on non-Oracle data sources
- This solution can now manage data from sources, including ADW, ATP, MySQL, Oracle Database, and Oracle Object Store
- Open up ADW to integrate more deeply with third-party tooling
- With ML, for example, on-premises and Database Cloud Service, Oracle Database users can install third-party packages to supplement in-database functionality; the company expects to allow Oracle Autonomous Database users to do the same, moving outside of OML4Py to install third-party Python packages

Figure 18: Omdia Universe ratings — Oracle



Source: Omdia

Strengths

- When using AI to automate and eliminate most database administrative tasks, Oracle stands head and shoulders above its rivals. Most vendors offer some form of heuristic query and query planning optimization capabilities. Oracle, however, also uses ML to automate key aspects of query performance, including self-tuning database settings, auto-scaling, automatic indexing, and optimized data compression. The company directly applies this same approach to many other workloads such as data ingestion, providing users with self-service tools that guide the

data discovery process, uncovering data transformation issues, and suggesting appropriate next steps, all within a declarative environment.

- While many cloud-native data warehouse solutions espouse manual or semi-manual elastic scalability at a cluster and/or node level, Oracle offers its customers instantaneous, granular, and hands-free scalability. For example, rather than requiring users to manually provision elastic scale resources using a limited set of “t-shirt” sizes, Oracle enables auto-scaling (up and down) on a CPU core level without requiring that users pre-provision supportive resources. The same goes for storage allocation. For storage and compute, customers can auto-scale up to three times their initially reserved resources with zero downtime or user intervention.

- With ADW, Oracle offers a uniquely unified hybrid cloud and on-premises architecture. Customers can opt for a multitenant managed cloud data warehouse service on shared infrastructure on top of Oracle Cloud, a single-tenant managed cloud data warehouse service on dedicated/isolated Infrastructure also on Oracle Cloud, or a single-tenant managed cloud data warehouse on-premises in the customer’s data center—all with the same software and underlying hardware. This is the key to the company’s elasticity and AI-fueled autonomy. It also enables Oracle to offer some advantageous hybrid public/private cloud functionality, including hybrid partitioned tables, backup services, data integration, and batch and real-time data loading services. In addition, the OCI-Azure Interconnect provides multicloud capabilities that allow customers using Microsoft applications to take advantage of the rich Autonomous Database functionality. In the same vein, AWS customers can connect to Autonomous Database on OCI using an interconnect infrastructure partner like Equinix or Megaport.

Limitations

- Oracle Autonomous Database represents a highly differentiated approach, one that enables companies to see data not as a collection of unruly, disparate databases but rather as a single, self-managing data platform. As such, it represents Oracle's future as a database vendor. However, the company’s sizable customer base remains rooted in Oracle Database, the precursor of Oracle Autonomous Database. It will take time for the company to bring its customer base forward, which will require further work from the vendor to encourage this evolution through functional parity and streamlined migration paths. In the meantime, because Oracle’s approach to autonomy is unique with no direct correlation, its rivals can position Oracle as a company still working toward the ideals espoused by Oracle Autonomous Database.

- True to its roots within Oracle Database, ADW functions like a relational database capable of accommodating semi-structured data via columnar storage and integration with OCI technologies that can be used to build a lakehouse and object stores on other cloud platforms. This foundational architecture, however, keeps the product at arm’s length from providing native data lake functionality. Note that the solution does seek to give users transparent access to both worlds, for instance, using a data catalog to harvest data lake objects, offering hybrid partitioning for tables that combine database storage and object storage, and using predicate pushdown to either Parquet or ORC file formats to minimize data scans.

- Oracle's decision to make ADW an inseparable part of its OCI platform is part and parcel of the solution's ability to seamlessly span cloud/premises, self-manage autonomously, and scale automatically with fine precision. However, this makes ADW a closed system compared with solutions from many rivals, all of which are working to create a layer of abstraction between the underlying infrastructure and the database to facilitate flexible scaling and do away with vendor/platform lock-in and facilitate cross-cloud deployment options. This means that Oracle customers seeking to put ADW to work must commit to OCI as a unifying platform to a greater degree than the commitment required by the company's hyperscale rivals.

Outlook

As a vendor still ostensibly building out its global public cloud platform footprint and encouraging its sizable customer base to move to a new, cloud-native database paradigm, Oracle faces some intense competition within the data warehousing market from global hyperscale rivals AWS, Google, and Microsoft. That said, Oracle's financial solidity, extensive client base, and highly differentiated approach to infrastructure bode well for the company over the long term. Its ability to automate the management and operation of infrastructure and database together as a single entity puts Oracle well in front of its rivals architecturally. What remains is for Oracle to repeat its history by attracting more key customers like ADW and Exadata Cloud users such as Lyft, Deutsche Bank, and FedEx buyers. Doing so will help the broader market conclude that the cost/performance optimization available with this unique approach warrants a commitment to the Oracle ecosystem.

SAP (Omdia recommendation: Challenger)

SAP Data Warehouse Cloud should appear on your shortlist if you want a fully integrated platform capable of breaking down the barriers between the business and IT while unifying a wide array of functionality across databases, data warehouses, and analytics.

Overview

SAP's overall strategy with SAP Data Warehouse Cloud centers on simplicity, compatibility, and collaboration. Available globally as a fully managed, cloud-native solution, SAP Data Warehouse Cloud serves as an end-to-end data warehouse solution, unifying data management, governance, and modeling tools. SAP promotes its warehouse solution as means for on-premises customers to extend and transition their existing investments to the cloud and embrace cloud-native development.

As with most data warehousing solutions, SAP Data Warehouse Cloud emphasizes helping users connect with and analyze analytical data across various data sources without forcing excessive data movement or replication. However, SAP steps out on its own in bringing business users and IT practitioners together through self-service and data democratization capabilities. For SAP, this manifests within a common experience called "spaces," which features no-code tools to connect, model, visualize, and share data without sidestepping IT security and governance.

Also unique to SAP's data warehouse offering is the company's stature as a leading line-of-business software provider. It greatly benefits SAP ERP customers, serving as a unified data layer across the company's finance, human resources, and operations software. As such, SAP Data Warehouse Cloud

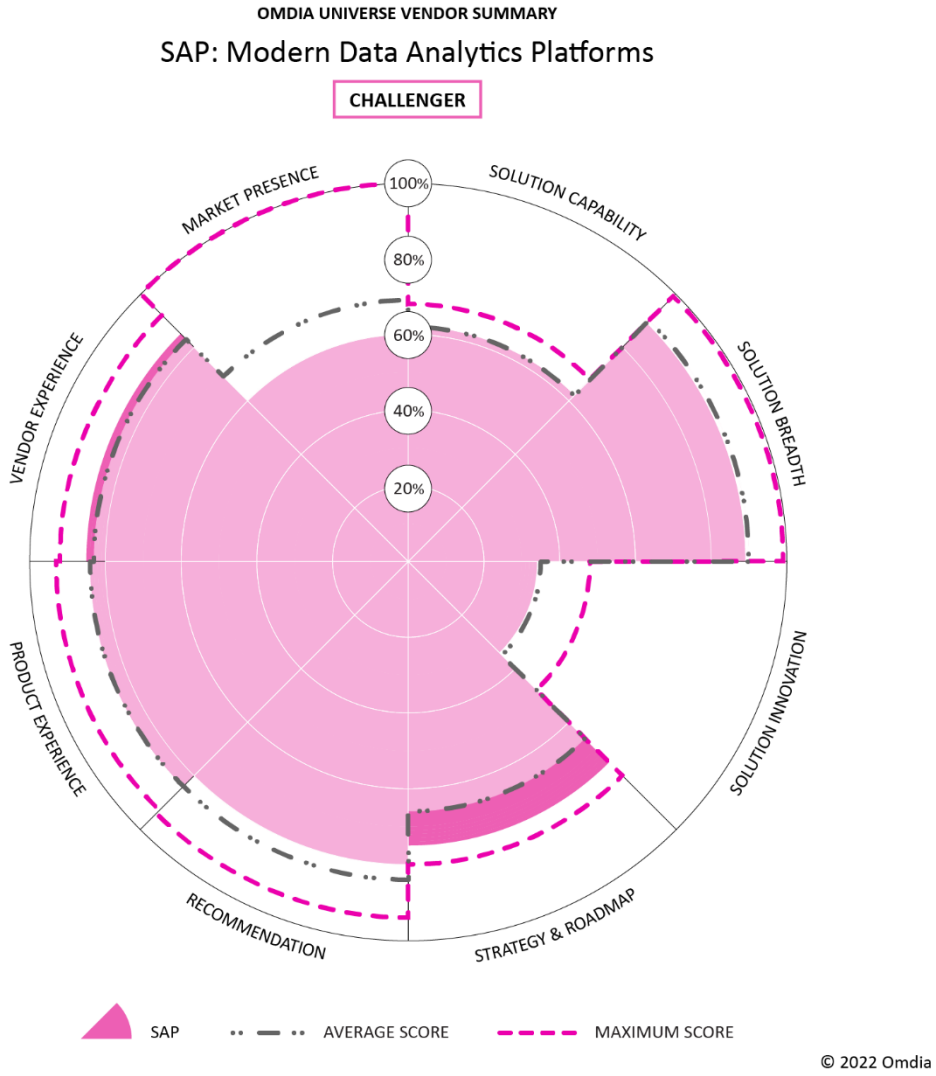
integrates with several solutions, including SAP Business Warehouse and SAP BW/4HANA, to harness existing investments in SAP BW and BW/4HANA.

Overall, SAP Data Warehouse Cloud aligns well with Omdia's core inclusion criteria used in this review. Beyond bringing insights to the business via self-service spaces, the solution fully embraces cloud-native functionality, separating storage and compute resources to provide customers with a flexible pricing structure. It focuses on bringing in external data through a set of extractors that represent external data sources virtually, persisted in real-time, or scheduled regularly as snapshots. SAP also prioritizes the creation of AI outcomes. SAP Data Warehouse Cloud works with SAP Data Intelligence to provide in-database model inferencing runtime capabilities, for example.

Roadmap priorities

- Enhance elasticity to cater to the needs of any company of any size and any workload
- Tight integration between SAP Planning applications and SAP Data Warehouse Cloud
- Help more SAP on-premises customers move to the cloud as with the recently released SAP BW bridge, which lets users bring SAP Business Warehouse functionality to SAP Data Warehouse Cloud
- Continue investing in lifecycle management and authorization management

Figure 19: Omdia Universe ratings — SAP



Source: Omdia

Strengths

- Thanks to the company's pervasive role in managing enterprise resource planning (ERP) and many other line-of-business concerns, SAP operates from a distinct position of strength within the analytics and data management market. According to SAP, 77% of the world’s transaction revenue touches an SAP system. From this foundation, the company markets SAP Autonomous Data Warehouse, a part of its foundational business technology platform (BTP), as a unifying data layer for all lines of business processes, not just those specific to SAP’s ERP portfolio.

- SAP Data Warehouse Cloud is not just cloud-native regarding running on multiple cloud platforms and allowing for flexible storage and compute allocation. Architecturally, the solution does a great job of unifying database, data warehouse, data lake, and data science workloads as a single, fully-managed data platform. For example, it fully automates all database-level operations, including backups, security patching, maintenance, resource allocation, resource optimization, monitoring, and alerting.

- One of SAP's primary differentiators is its focus on empowering business users. This philosophy permeates SAP Data Warehouse Cloud, particularly in taking processes typically centralized in IT and opening them up to departmental users. Like many data warehousing solutions, SAP Data Warehouse Cloud seamlessly integrates analytics, visualization, and reporting functionality (integrating SAP Analytics Cloud). Yet, SAP further introduces a semantic business layer that lets departmental users directly model scenarios without IT involvement, using the semantics they understand. In doing so, they can easily and quickly add external data products to their data models from the SAP Data Marketplace, which features a growing number of data sources from PWC, Refinitiv, Civey, Shippeo, NOW.ai, Accuity, Datazeit, LexisNexis, et al.

Limitations

- Comparatively, the SAP Data Warehouse cloud is more constrained regarding where it can be deployed. The solution is available as a public cloud service only in specific regions on hyperscalers like GCP, AWS, and Microsoft Azure. For example, SAP is still working on data-center availability throughout the US on GCP. Additionally, because Data Warehouse Cloud is offered as a SaaS solution, its hybrid deployment options are limited to integration with SAP BW/4HANA and SAP HANA for SQL data warehousing, both of which are offered as on-premises or hosted private cloud solutions.

- While SAP Data Warehouse Cloud provides data lake functionality, it does so at arm's length compared to many solutions under review. It integrates federated data access to SAP HANA Cloud Data Lake and offers replicated access to external data lake sources such as AWS S3, Google Cloud Storage (GCS), or generic HDFS.

- Though SAP Data Warehouse Cloud is fully cloud-native in its architecture, it does not yet fully equip users with the autonomous and/or self-service tools they need to optimize resource utilization, such as scaling compute and storage sizes. Such functionality, such as autonomous instance resizing, memory increase/decrease, etc., is on the roadmap for later in 2022.

Outlook

Overall, SAP Data Warehouse Cloud is a competent cloud-native data platform built to break down the barriers between business and IT users and fully unify database, data warehouse, data intelligence, data lake, and analytics capabilities. Yet, despite SAP's efforts to position SAP Data Warehouse Cloud as an independent solution that is not interdependent upon SAP's broader BTP or the company's sizable portfolio of business software, SAP must work harder than its independent counterparts to show the broader data and analytics community that it has created a powerful

analytics platform that can stand quite well on its own outside the confines of SAP's established business customers.

Teradata (Omdia recommendation: Challenger)

Teradata should appear on your shortlist if you want a rock-solid performer and trusted partner with more than 43 years of experience managing mission-critical analytical workloads.

Overview

In the mid-1980s, Teradata entered the market with a database purportedly capable of scaling up to terabytes of data, an aspiration that Teradata smashed in 1992 with a database system built for Wal-Mart capable of scaling beyond one terabyte in size. Over the intervening decades, very little has changed for Teradata. After rolling out early innovations around in-memory computing, solid-state disk storage, joint columnar- and row-based tables, and cloud-based analytics services, Teradata has kept its foot pressed down hard on the innovation accelerator.

This rich heritage has allowed Teradata to yield tremendous influence within highly demanding verticals such as financial services, which makes up the bulk of its customer base, followed by retail, telecommunications, and healthcare. Today the company focuses its efforts across these verticals, targeting the top 10,000 largest enterprises in the world that are looking to put data to work at the extremes of performance. Teradata's go-to-market message has steadfastly remained rooted in helping these customers solve complex data challenges and do so at scale and securely across complex, highly concurrent workloads.

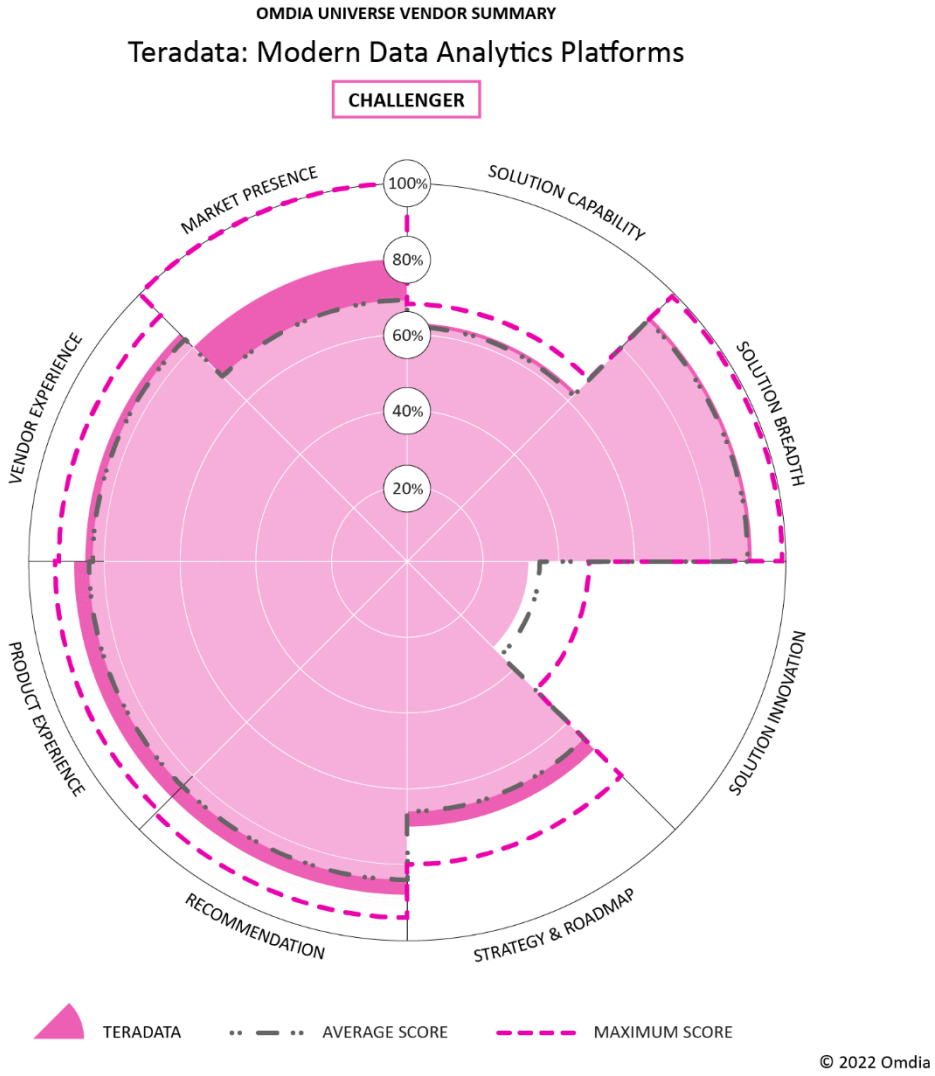
As with most vendors reviewed in this report, Teradata has had to find a way to translate its predominantly premises-centric advantages to the cloud. The company introduced Teradata Vantage in 2018; Teradata Vantage is a cloud-native offering built to help its customers migrate their data warehouse to a cloud-based platform. This holds true for new customer acquisition, with Teradata Vantage serving as their strategic platform going forward. In support of this, the company now dedicates 70% of its budget to cloud innovation.

An interesting aspect of Teradata Vantage is its emphasis on going beyond basic hybrid cloud deployments to support a multicloud ecosystem that spans AWS, Microsoft Azure, GCP, and any cloud-running VMware. Teradata hopes to serve as a connected, consistent data warehouse that encourages users to leverage their tool of choice, be that data science software on AWS or business intelligence software on Microsoft Azure.

Roadmap priorities

- Ease adoption and cloud migration through license portability, ensuring that the same software is enabled in all environments, and enrich workload migration tools and services
- Ensure that customers can obtain greater flexibility to leverage the tools of their choice on top of the Teradata platform

Figure 20: Omdia Universe ratings — Teradata



Source: Omdia

Strengths

- Teradata benefits from a legendary history backed by a highly performant parallel architecture with rich query optimization and workload management, together with a highly differentiated experience supporting highly demanding vertical markets. Throughout its history, the company has repeatedly demonstrated the ability to tackle the largest, most complex, mission-critical workloads.

- Teradata Vantage specializes in delivering nearly unlimited concurrent queries across mixed workloads and query types. Plying its experience in building MPP systems, concurrency is the company's calling card built on deep partitioning, support for numerous indexing methods (e.g., secondary indexes, hash indexes, join indexes), and rigorous compression routines reaching a 15:1 ratio.

- Teradata seeks to deliver a genuinely hybrid/multicloud platform in interoperability and manageability. Through the company's Teradata Viewpoint experience, as an example, Teradata enables holistic management of multiple Teradata Vantage implementations, whether those are on-premises, in the cloud, across multiple clouds, or in a hybrid cloud configuration.

Limitations

- As an independent player focused on the upper end of the market and the finance vertical, Teradata is more vulnerable to market shifts, emerging competitors, and cloud platform providers seeking to displace Teradata directly.

- Teradata's principal technological strength—a focus on highly demanding workloads—presents a significant challenge for the company in marketing against rival solutions that are marketed as being less monolithic; therefore, easier to adopt, maintain, and extend through the adoption of the best of breed tooling.

- Teradata Vantage possesses native data sharing capabilities using open data formats (parquet, CSV, and JSON) or internal and external object stores. However, it does not incorporate any native data sharing functionality such as dataset verification/validation services. Instead, it prefers to look for this functionality externally through integration with AWS Data Exchange and Azure Data Share.

Outlook

Can Teradata succeed in translating its early on-premises success in pushing the envelope of performance to the cloud? If the company’s recent effort to successfully support more than 1,000 nodes and 1,023 active users (each submitting thousands of concurrent queries), all on a single AWS instance is anything to go by, the answer is an indisputable yes. Conversely, as an independent player, this focuses on the upper end of the market and the finance vertical makes Teradata more vulnerable to market shifts, emerging competitors, and cloud platform providers seeking to displace the company directly.

Vertica (Omdia recommendation: Prospect)

Vertica should appear on your shortlist if you want a straightforward, self-contained data warehouse platform with a proven track record in supporting large-scale, complex analytical workloads.

Overview

Founded in 2005 by database luminary Michael Stonebreaker, Vertica entered the data warehousing market with a single guiding principle: bring the highest performance available at scale to bear on a

wide range of analytical use cases and do so at a low cost of ownership and with a rapid ROI. Since that time, a lot has changed for Vertica. The company has undergone some ownership changes, culminating in the acquisition by HP in 2011 and subsequent sale to Micro Focus in 2017. Through those changes, however, Vertica has managed to retain its value, owing greatly to its consistent philosophy of combining high performance with high value.

Today, Vertica occupies a stable and influential position in the modern data warehousing marketplace, supporting more than 1,000 Fortune 500 market leaders such as Philips and The Trade Desk. Forty percent of Vertica's customers claim more than $1 billion in revenue. Interestingly, the company has also found a ready customer base among large-scale data disruptors like Cisco and GoodData, both of which run their data and analytics platforms on Vertica. This aspect of Vertica's business stems from the composable nature of the Vertica platform. More than 30% of its customers embed Vertica in applications into solutions that are then distributed to thousands of end customers.

As with most competitors reviewed in this report, Vertica addresses modern data warehousing requirements, using a SQL-centric, MPP columnar database to support a wide array of workloads data types using SQL as the principal language, including time-series and geospatial data and sensor data analysis. Vertica offers a very straightforward and easily consumed analytics database with just three offerings—all based on the same single product—that does not require additional licensing to support in-database functionality such as ML development, deployment, and management.

- **Vertica Unified Analytics Platform** is a cloud-native analytics database that can be deployed on-premises or in the clouds (public or private) as a self-managed service.
- **Vertica Accelerator** is implementing the Vertica Unified Analytics platform, running on AWS as a SaaS offering, entirely within the customer's private account space.
- **Vertica Community Edition** is a free trial edition with no time limits for storage requirements under 1TB and three nodes.

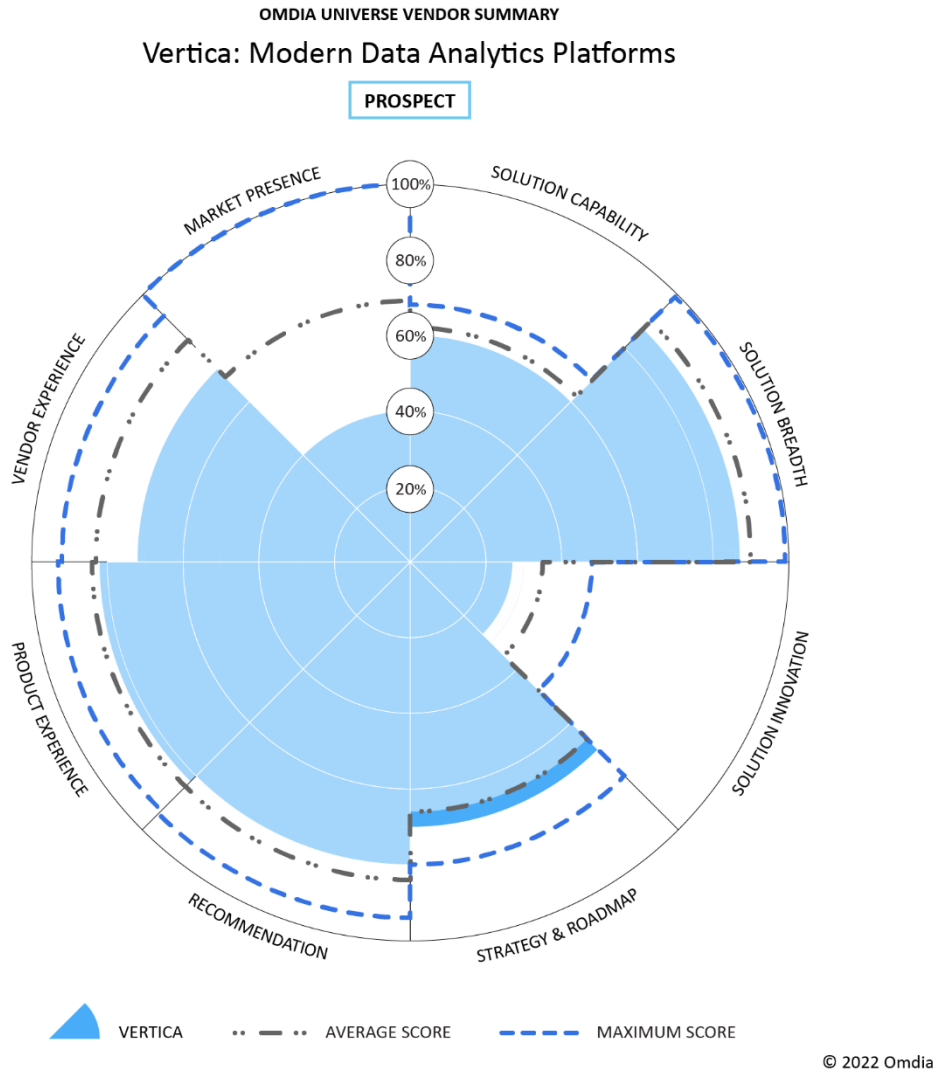
Customers can apply for Vertica's licenses in a highly fungible manner, moving them between cloud platforms and on-premises deployments, supporting hybrid use cases. This way, customers can freely move between AWS, MS Azure, GCP, Alibaba, and private cloud deployments without altering or renewing their contracts.

Roadmap priorities

- Continue to extend deployment options for self-managed offerings while also increasing the use of automation
- Bring Vertica Accelerator to multiple public clouds with all the functionality of Vertica intact as SaaS

- Exploit the use of sub-clusters spanning on-premises and public cloud deployments as a means of keeping customer data within private data centers while leveraging public cloud platforms for computing resources
- Offer a web-based integrated analytics environment and simplified ingestion tools and data pipelines

Figure 21: Omdia Universe ratings — Vertica



Source: Omdia

Strengths

- Like many data warehouse vendors, Vertica understands the value of bringing ML workloads into the analytics database to minimize data movement. What sets the company apart is its complete lifecycle approach to in-database ML using SQL, Python, R, or Java. The database, for example, includes tools specific to data preparation and exploration routines like outlier detection. Likewise, for model evaluation, it contains tools to measure receiver operating characteristic curve (ROC) scores and the ability to revert to previous model iterations using full version control. Also unique to Vertica is the vendor's choice to offer these tools at zero additional charges, making it one of the only providers with a completely self-contained ML platform to emphasize ML lifecycle operationalization (commonly referred to as MLOps).

- Because Vertica functions as both a query engine and a data warehouse, it presents customers with some very differentiated options. For example, in working with data lakes, the solution does not need to rely on connectors to query object stores, nor does it try to shoehorn object storage into relational database architecture. It simply serves as a query engine for a given data lake, using the vendor's Vertica for SQL on data lakes. This also lets the platform function as its own processing engine akin to Apache Spark, Presto, Hive, or Impala. In this way, it can serve not only as a downstream recipient of Kafka streaming data but also as a streaming data producer, retrieving data from Kafka, performing analytics on the data, and then sending the results back to Kafka for consumption by other applications. In this way, Vertica is a perfect partner to processing engines like Spark, Airflow, Kafka, and others.

- As cited by many customers, Vertica is synonymous with performance. Built from the beginning as a columnar MPP platform, Vertica provides rapid query response against enormous data sets with data sets larger than 100TB, which is not uncommon, and multi-petabyte datasets, which are not unusual. For example, one customer, a major insurance provider, described using Vertica to readily compare more than 10 billion geospatial data points as an overlay of customer activity.

Limitations

- Vertica has centered its new logo efforts around the Vertica Accelerator, the company's SaaS offering running on AWS. This approach has dramatically lowered the bar of adoption for new customers through automated setup, administration, and management capabilities. However, until Vertica can bring Accelerator to other hyperscale cloud platforms—Alibaba, GCP, and Microsoft Azure—the vendor will lag a market that is already capable of meeting customers wherever they operate in the cloud. Until then, customers will have to self-provision and manage Vertica on these platforms, licensing Vertica on an hourly, pay-as-you-go basis from GCP, for example. Note, however, that Vertica is betting on the continued importance of on-premises data storage with an architecture that can run the same on any technology, whether Pure/Dell ECS or AWS S3 buckets.

- Despite Vertica's popularity as an embedded solution, the vendor has not actively pursued creating a partner ecosystem focused on data. The company might exchange data sets and other

assets such as ML artifacts with its customers and partners' models. Such markets or exchanges have become a necessary component within the data and analytics market. Through these communities, vendors drive direct revenue and elevate the value of practitioners, which in turn boosts the value of Vertica as a domain of practitioner expertise. Note that if Vertica were to emphasize this aspect of its business, the company would have little trouble creating a differentiated platform, based upon its experience in providing Vertica as an embedded solution, as well as its openness to partnership (both technology and professional services).

- While advantageous for Vertica customers, Vertica's unique architectural aspects can also serve as competitive vulnerability, as rivals can position the vendor as eschewing more customary technological trends. This is the case with in-memory processing, as an example. Vertica does not support such processing on purpose, opting for a more holistic approach that uses every resource, memory, and CPU to its fullest in a memory-first strategy. The same applies to Vertica's decision to use projections instead of materialized views to store query results on disk rather than having to compute them each time they are queried. Such differing views require that Vertica prove to customers the efficacy of its countervailing approaches. It also demands more from practitioners, which must acclimate to the intricacies of the Vertica platform.

Outlook

Vertica's focus on performance, coupled with its emphasis on in-database ML and its straightforward approach to software packaging and consumption, sets the vendor apart from many of its rivals that emphasize integrating several supportive components to fulfill large-scale, complex analytical workloads. As an independent database provider, however, Vertica will have to speed up its move to the public cloud to deliver Vertica Accelerate across more hyperscale platforms and open doors to more hybrid- and multicloud opportunities. The vendor's affinity for working cooperatively with popular technologies will open a new, highly lucrative world of ecosystem opportunities. For example, the recently released Vertica 12 expands Kubernetes support on GCS, Azure Blob Storage, and Hadoop distributed filesystem storage (HDFS).

Appendix

Methodology

Omdia Universe

The process of writing a Universe is long and time-consuming; it involves the following:

- Omdia analysts perform an in-depth review of the market using Omdia's market forecasting data and Omdia's enterprise insights survey data.
- Omdia creates a matrix of capabilities, attributes, and features that it considers important now and in the next 12–18 months for the market.
- Vendors are interviewed and provide in-depth briefings on the current solutions and future plans.
- Analysts supplement these briefings with other information obtained from industry events and user conferences.
- The Universe is peer-reviewed by other Omdia analysts before being proofread by a team of dedicated editors.

Inclusion criteria

To be considered for inclusion in this Universe, the offering must meet the following criteria:

- The solution must be available as a self-hosted or managed service on either the vendor's public cloud platform or upon one or more public cloud platforms from AWS, Microsoft, SAP, Oracle, Google, IBM, et al.
- The solution must directly support hybrid and multicloud deployment scenarios.
- The solution should aggressively use cloud-native capabilities such as running on various containerized architectures, using serverless functionality, and enabling auto-scaling.
- The solution should incorporate a semantic business layer (catalog, hub, etc.) for internal and external data assets.

- The solution should natively incorporate ML capabilities (tooling, artifact management, model storage, etc.).
- The solution should directly support structured, semi-structured, unstructured, and streaming data.
- The solution must incorporate in-product data engineering, management, and analytics capabilities, including data exploration, transformation, optimization, etc.

Further reading

Fundamentals of Modern Data Analytics Platforms – 2022 (May 2022)

Analytics & Data Management Quarterly Briefing – 1Q22 (March 2022)

Analytics & Data Management Forecast Report – 2021 Database (October 2021)

IT Drivers and Technology Priorities – IT Enterprise Insights 2022 (October 2021)

The Rise of DataOps and Continuous Data Delivery (June 2021)

Author

Bradley Shimmin, Chief Analyst, AI Platforms, Analytics, and Data Management

askananalyst@omdia.com

Citation policy

Request external citation and usage of Omdia research and data via citations@omdia.com.

Omdia consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help you. For more information about Omdia's consulting capabilities, please contact us directly at consulting@omdia.com.

Copyright notice and disclaimer

CONTACT US

omdia.com

askananalyst@omdia.com