EMA Radar Report:
Unified Analytics Warehouse

A Guide for Investing in Unified Analytics

May 2021
By John Santaferraro, Research Director

EMA RADAR

Table of Contents

Executive Introduction

2021 is the year of convergence. Most significantly, EMA sees the convergence of the data warehouse and the data lake into the unified analytics warehouse (UAW). The UAW is unified because it adequately analyzes multi-structured and multi-latency data in a single platform. It is an analytics platform because the primary use case for both the data lake and the data warehouse has always been analytics. It is a warehouse because it stores multi-structured data in an organized and accessible manner, ready for a broad range of analytical users and use cases.

UNIFIED ANALYTICS WAREHOUSE
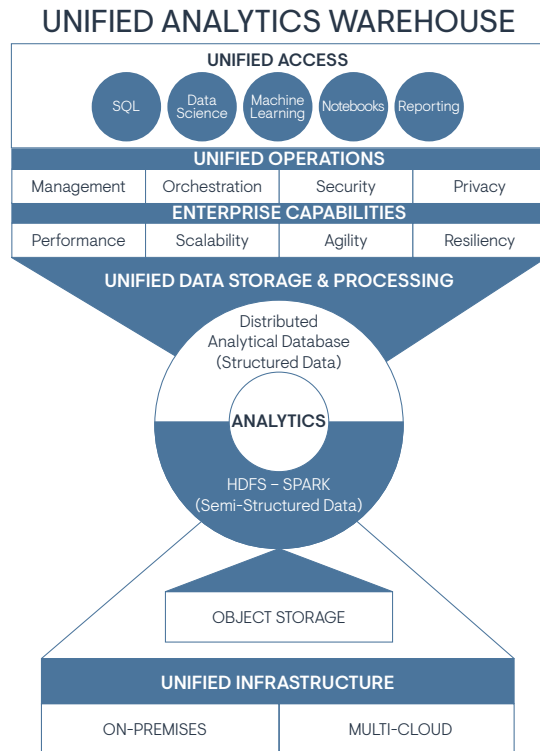
Driven by customer demand, vendors are approaching this unification effort from four different perspectives.

From the data lake perspective: nascent data platform vendors that built their data architectures to handle semi-structured data are working to build database technology, including structured columnar storage formats and SQL query engines, on top of their file-based data storage systems. Because the data is typically stored in a file system, providing access to object storage (like Amazon S3) was either built-in or easy to build. Data lake technology was built for structure-on-read processing, making it readily accessible to the typical discovery analytical work done by data scientists.

From the data warehouse perspective: mature database vendors built for structured data are working to access both semi-structured data and tiered storage. Columnar MPP databases were built for high-speed analytical processing, and some databases have worked to provide notebook and data science access to their platforms. In addition, it is becoming more common to see machine learning algorithms embedded in data warehouse technology for high-performance processing of advanced analytics on structured data.

From the data access perspective: distributed query vendors separate the compute from the storage, creating a data agnostic model in which all data can be queried in a unified manner. Created out of the need to be able to run queries across petabytes of data, these new architectures specialize in optimizing query performance using traditional methods, creative use of memory, and best possible use of distributed computing resources.

From the new world perspective: emerging vendors are completely architecting new unified analytics warehouse platforms to abstract away the complexity of data management and treat data like an operating system.

Core Criteria

In August of 2020, EMA published the first definition of the unified analytics warehouse in the white paper, "The Emergence of the Unified Analytics Warehouse – Data Lakes and Data Warehouses Merge." Initial criteria for this new category included requirements for data, enterprise, infrastructure, cloud, analytical processing, and users. These criteria were expanded to include the following for this research:

- Expansion toward the inclusion of multi-structured and multi-latency data
- Broadening of analytical use cases to support both data science and data analyst workloads
- Unification of multiple data management and analytical capabilities into a single, unified offering
- Support for hybrid and multi-cloud deployments with some consistency across diverse deployments
- Abstraction or automation of complex data management processes to speed and simplify insight delivery
- Strong support for enterprise and regulatory requirements for data and analytics
- Potential use as a strategic investment that may assimilate or consolidate multiple analytical technology investments

A Market in an Accelerated State of Evolution

2020 introduced three major global shifts that accelerate the move to the unified analytics warehouse. First, major economic shutdowns put every organization's digital business models to the test. Many failed, many passed the test, and many adjusted quickly to change to digital business models that worked. Second, digital business models increased the need to analyze multi-structured and multi-latency data and respond intelligently to business events, both inside and outside the enterprise. Third, billions of dollars poured into analytical platform vendors from the venture and private equity world. The speed at which technology was released to the market accelerated exponentially.

The result is a market that is ripe for convergence. The legacy model of separate platforms operating as silos within organizations and competing for resources can no longer be sustained. The decade-old trend of adding niche platforms to already complex data ecosystems is over. At the start of 2021, EMA research indicated that 52% of participants have more than five data platforms under their management. A full 73% of participants indicated that their strategy going forward was consolidation.

Because every data ecosystem is different in terms of data being processed through existing data platforms, there is no single answer to the consolidation question. In like manner, different vendors are approaching unified analytics in different ways. These two trends combined make diversity in the unified analytics market necessary and positive. For this reason, the goal of this Radar Report is not to pick a single winner. Instead, it is to provide analytics organizations with an understanding of the use cases relevant to planned purchases.

In truth, any one of the 12 vendors represented in this report might be the best choice for your organization depending on your current digital business requirements and existing technology. Many of the vendors covered in this report overlap in technology foundations, and several are complementary to each other.

Methodology

EMA first reached out to vendors that were qualified for this Radar in December of 2020. The process required EMA to complete the following steps with each of the 12 vendors covered in this report:

- Finalizing a 22-page questionnaire and sharing it with vendors
- Providing weighting to different questions and answers to support the generation of KPIs, pentagons, and Radar Chart positioning
- Reviewing vendor inputs in sessions that lasted 60-120 minutes
- Entering the data from the questionnaires into Excel to generate KPIs and Radar Charts
- Interviewing customers to validate vendor claims (this step was required)
- Analyzing the results and developing the profiles in March and April of 2021
- Conducting final reviews and writing the report in April 2021

Use Cases and Associated Criteria

EMA evaluated the 12 vendors based on three different use cases that drive the unification of the data warehouse and the data lake:

- Hyperperformance and Enterprise Readiness: The first use case focuses on high-performance analytics with a complete and proven set of enterprise capabilities. Leaders in this category typically leverage database products with years of performance tuning and a long history of enterprise customers. There are some new entrants that have accelerated the delivery of advanced enterprise features and functions.

- Analytical Flexibility and Low Latency: The second use case focuses on extreme flexibility for multiple analytical use cases, including multi-structured data and multi-latency analytics. Leaders in this category typically utilize new memory technology and open data storage to enable a "store-once, analyze many" architecture. Some high-performance, enterprise-focused vendors have extended their platforms for greater data flexibility and lower latency.

- Hyperscale and Platform Leverage: The third use case focuses on maximum scalability and the seamless leverage of existing and low-cost data storage options. Leaders in this category tend to completely separate analytical computing from storage with a focus on extreme, in-memory analytical processing to scale for queries across multiple data sources. Some vendors add these access capabilities to their more traditional data warehouse and data lake platforms.

EMA RADAR™

The core criteria for evaluation are broken down into five categories (Appendix A looks at key performance indicators [KPIs] in detail):

- **Deployment and Administration** – Deployment and administration efficiency is critical to successful implementation and rapid time to value. This category covers everything from presales to ongoing administration. It includes provisioning time and time to production, as well as deployment versatility, acceleration, and resources. For administration, the report looks at resource requirements, breadth of administration, and automation, along with customer support, services, and community.

- **Cost Advantage** – Cost advantage covers more than price or price performance. It targets concepts such as flexibility in pricing, return on investment, and multi-service bundling, as well as cost management and containment capabilities.

- **Architecture and Integration** – Architecture and integration looks specifically at product fit for each of the three use cases identified: hyperperformance and enterprise readiness, hyperflexibility and low latency, and hyperscale and platform leverage. It also assesses architectural excellence for maximum extensibility across hybrid and multi-cloud environments. For integration, it considers breadth of data support for multi-structured and multi-latency data.

- **Functionality** – Product functionality delves into features necessary for the operation of analytics within the modern, digital enterprise. It also looks at the breadth of analytical capabilities across multiple use cases for a diverse set of analytical users. Additional scoring covers automation, business logic, and partner ecosystems.

- **Vendor Strength** – Vendor strength is measured by financial strength, investment in research and development, vision, patents, and geographic coverage.

Some Highlights on What the Data Shows

One of the benefits of doing the Radar was getting data from all 12 vendors on critical areas ranging from deployment and administration to functional and architectural trends in unified analytics.

Deployment, Administration, and Cost Advantage
Deployment
In the area of deployment options, 100% of the UAW vendors are cloud-native or offer a cloud version of their offering. However, only 67% offer a software as a service (SaaS) version of their software. These two cloud figures demonstrate broad cloud adoption and the importance of a move toward SaaS offerings. Almost all of the vendors that do not currently offer SaaS have a service offering in development or on their roadmaps.
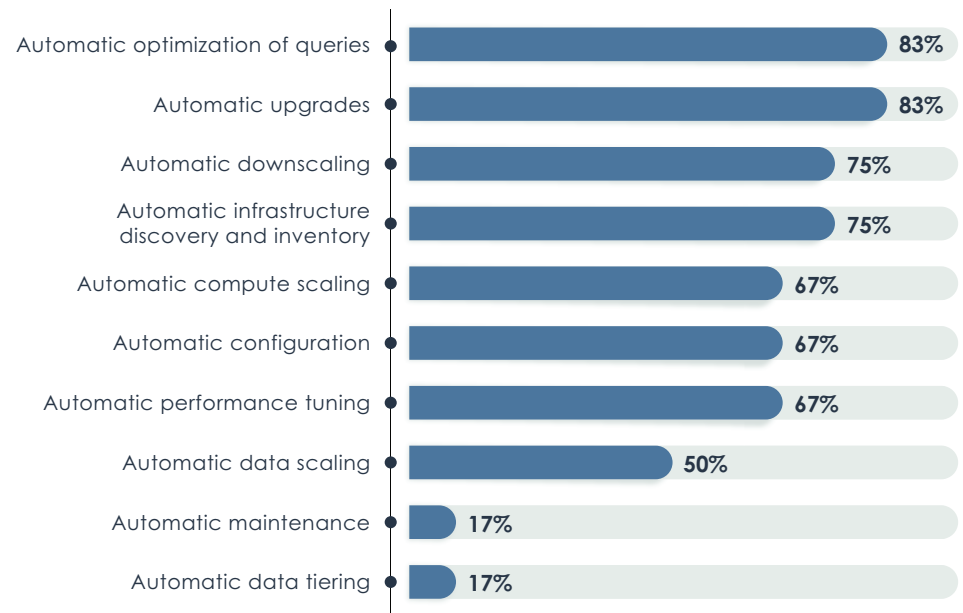
Along with this shift to the cloud comes quicker time to deployment, with a full 58% able to put analytics into production within an hour. Of course, that assumes that a customer has already prepared data and has a good idea of insight that is going to be used in production before they purchase a UAW vendor product. Another 16% of vendors can have analytical insight in production within a day. For organizations that have projects for which time is of the essence, time to production can be a serious competitive advantage.

Administration
In the area of administration, EMA found that, on average, UAW vendors indicated that one full-time employee (FTE) was required for ongoing admin-istration in an enterprise with about 10,000 employees. This did not include initial deployment or any significant extension in breadth of coverage or func-tionality. There was an even split at the average, with 50% of vendors requiring up to one FTE and 50% of the vendors requiring between 1 and 2.5 FTEs. There was no correlation of high and low administrative requirements tied to

cloud-native versus data center-native vendors, nor was there any correlation between a rise in administrative requirements based on offering a broader set of unified analytics capabilities.

There is a direct correlation of automation to low administrative requirements. Vendors that have invested significantly in administrative automation tend to require fewer FTEs to administer their system. EMA asked vendors to identify which of the following administrative functions are currently being automated in their platforms. The result shows significant investment in the use of machine learning and metadata to automate formerly manual processes for the administration of unified analytics.

| Function	Percentage
Automatic optimization of queries	83%
Automatic upgrades	83%
Automatic downscaling	75%
Automatic infrastructure discovery and inventory	75%
Automatic compute scaling	67%
Automatic configuration	67%
Automatic performance tuning	67%
Automatic data scaling	50%
Automatic maintenance	17%
Automatic data tiering	17%

EMA RADAR

While EMA sees maturity in the use of automation for UAW administration, single-pane administration remains an opportunity area for vendors. There are only two vendors that administer all on-premises and multi-cloud deployment in a single pane. Also, with data moving to object and block storage options to reduce cost, automatic data tiering could become a significant competitive advantage. Today, only one of the vendors offers this valuable feature.

Cost Advantage

EMA went beyond price and price performance to evaluate the cost advantage of each vendor. This included looking at flexibility in licensing, built-in cost containment, and the total value provided by each vendor, especially based on the breadth of their offering. For example, 58% of the vendors offer a single license for all cloud and on-premises deployments, leaving 42% with separate licensing and potentially more cost for customers.

EMA also sees a trend toward cost containment, something that is especially critical for cloud computing and pay-per-use pricing. Seventy-five percent of the vendors have capabilities in place to help customers contain cost. Specifically, 83% have capabilities in place to help customers manage cost across a broad number of users or groups on their platform. When EMA asked vendors about the specifics of their cost containment, provisioning data sets and analytics came out on top, followed closely by usage limitations and chargeback.
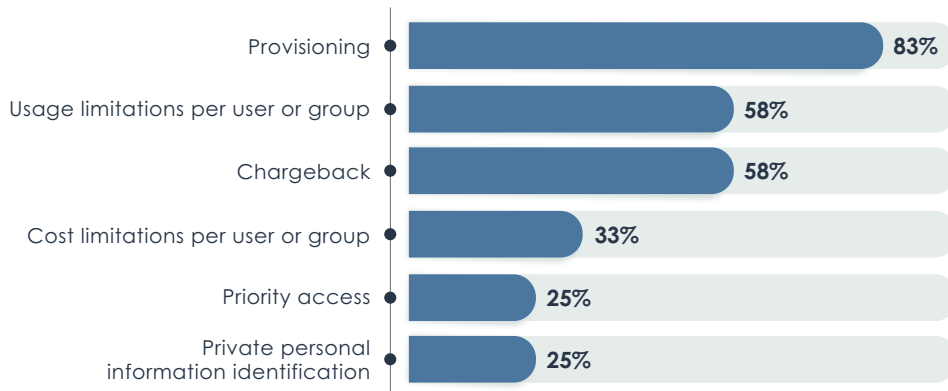
Finally, regarding cost advantage, unified analytics warehouse vendors continue to expand the breadth of their offerings with more capabilities across the entire data lifecycle. The full set of capabilities included by one or more of the vendors includes administration, various data storage types, event streaming, in-memory data storage and processing, data preparation, data pipeline management, data sharing, data governance, embedded machine learning and analytical algorithms, a business semantic layer, logical data models, business intelligence tools, and predefined reports, dashboards, and KPIs by industry and business area.

While not all vendors support MLOps use cases, embedded machine learning algorithms rose to the top of the list, with all but one vendor including this in their unified analytics offering. Every one of these vendors includes more than 50 algorithms in their offering, signifying the importance of machine learning to the unified analytics mix, especially for digitally-oriented organizations.

Surprisingly, there are very few vendors that provide a full set of business accelerators and business logic, either by business area or by industry. This remains an opportunity area for most vendors and sets apart those that provide a rich set of business value in their solution. Even without a major shift toward business value, 83% of the vendors indicated that some customers achieve a return on investment within six months, some as quickly as one week.

EMA recommends looking at cost beyond price or even price performance. It is important to evaluate purchases based on the value being generated by each solution option. This includes all the cost advantage areas evaluated in this report.

Provisioning	83%
Usage limitations per user or group	58%
Chargeback	58%
Cost limitations per user or group	33%
Priority access	25%
Private personal information identification	25%

EMA RADAR™

Architecture and Integration
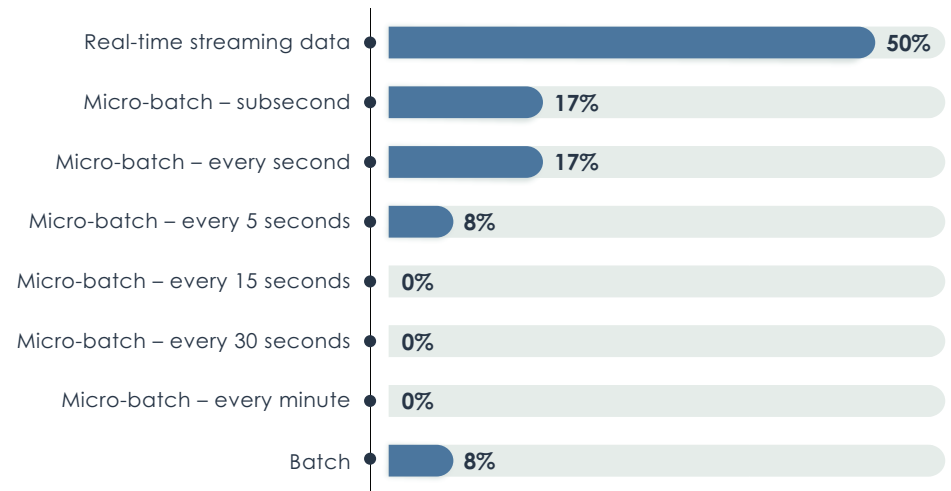## Platform Capabilities

While many of the questions asked of vendors regarding platform capabilities were used to identify their strengths in each of the different use cases, EMA also researched the breadth of different information services unified in their platform offerings. There was a clear line of delineation between capabilities included by two-thirds of the vendors, and less than two-thirds.

Two-thirds of the vendors or more unify the following in their offering: data replication, change data capture, data preparation, data cataloguing, metadata services, master data services, SQL analytics, data science, and machine learning. These clearly represent features that are table stakes for most use cases in unified analytics.

Less than two-thirds of the vendors unify the following in their offering: data integration, data modeling, data sharing, data pipelines, and business intelligence. This indicates an opportunity for vendors to provide full-service features in these areas, or to adopt a "just enough" approach to maintain their partnerships with other vendors in these areas.

Vendor platforms were also scored based on their means of processing analytics and their ability to analyze data in real time. EMA sees a trend toward the complete separation of compute and storage, with 83% of the vendors processing data on an analytical engine completely separate from the storage. This trend opens the door for more flexibility in storage options and the potential to store the data once and analyze it many times.

In the move to unified analytics, there is also a trend toward the inclusion of real-time analytics combined with batch analytics in a single platform. Half of the vendors can query data in motion, and most of the other vendors can query data within five seconds, which meets most use cases for real-time analytics.

Real-time streaming data	50%
Micro-batch – subsecond	17%
Micro-batch – every second	17%
Micro-batch – every 5 seconds	8%
Micro-batch – every 15 seconds	0%
Micro-batch – every 30 seconds	0%
Micro-batch – every minute	0%
Batch	8%

EMA RADAR

Cloud Capabilities

Cloud continues to be a top priority for all analytics, especially for unified analytics. A majority of the vendors provide a common version of software that runs in a like manner on-premises and across the different cloud vendors, such as Google, Amazon, and Microsoft. All but one vendor can now run their software or platform on containers with microservices, such as Kubernetes, RedHat OpenShift, or HPE Ezmeral.

To further support the importance of the cloud, all vendors provide their software solutions as a virtual private cloud, and two-thirds of the vendors provide their cloud in a fully-managed SaaS offering. Given the sheer amount of digital data residing in the cloud, having a cloud offering is no longer an option for unified analytics.
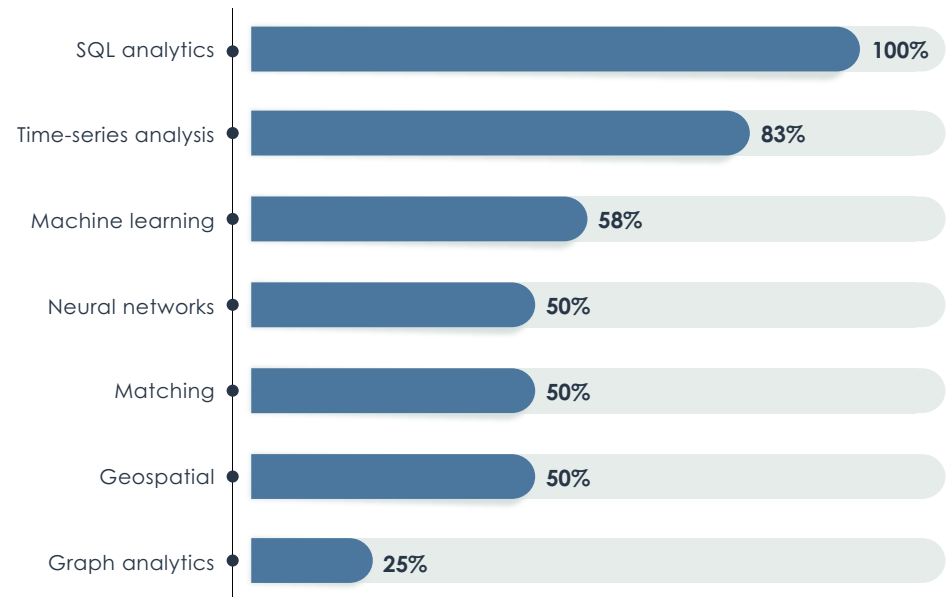
Product Functionality

Enterprise Capabilities

When delving into enterprise capabilities for the unified analytics warehouse vendors, EMA looked at elements like security, availability, scalability, and change management. Since security and availability are expected and necessary for modern analytics, every vendor showed up with both well in place. Even some of the younger vendors have thought through these issues.

Scalability is measured in different ways, including the ability to scale for data size, data volume, query complexity, and query concurrency. For data size, there were three vendors scaling up to 250 petabytes of data, three vendors scaling up to tens of petabytes, four vendors scaling into hundreds of terabytes, and two vendors with evidence up to 100 terabytes, but no customers in that range.

Analytical Capabilities

EMA research also revealed trends in the type of analytical workloads supported by vendors in the unified analytics space. Specifically, we wanted to know which types of workloads ran natively. At the top of the list, SQL remains the lingua franca of analytics, and SQL analytics are the most prevalent in terms of support by vendors. EMA was surprised by 83% support for time-series analysis above machine learning at 58%. This indicates the shift toward more real-time engagement and the importance of digital superiority as a driver for analytics. For clarification, EMA made sure that the vendors answered based on their ability to handle each of these "types" of workloads. We were not looking for equivalence with niche products that do microsecond time-series analysis or complex graph analytics.

Workload	Support
SQL analytics	100%
Time-series analysis	83%
Machine learning	58%
Neural networks	50%
Matching	50%
Geospatial	50%
Graph analytics	25%

EMA RADAR

Since machine learning is highly critical for digital business models, EMA also asked the vendors about their support of the different aspects of machine learning. The question was designed to look for platform support, not necessarily MLOps capabilities. More than two-thirds of the vendors now support almost all platform aspects of machine learning, including model training, scoring model comparison, and machine learning in production. Support tapers off to 50% once the capabilities start to move over into the MLOps side of things with monitoring.

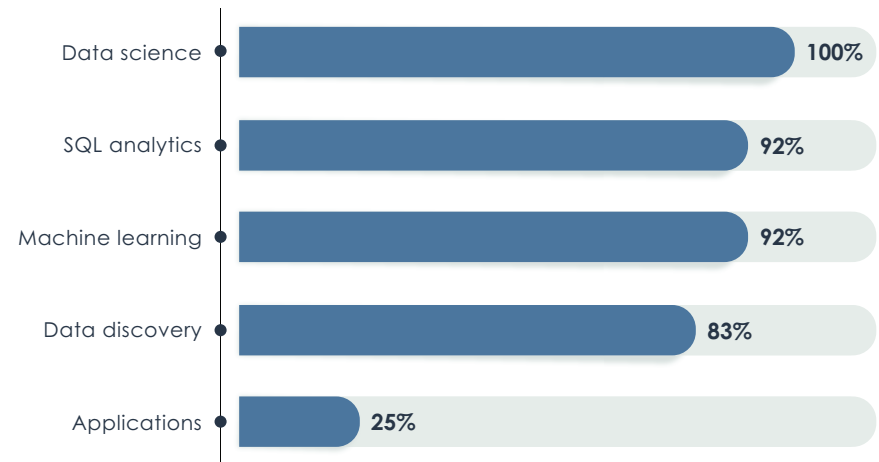| Category	Percentage
Model training	75%
Scoring	75%
Operations (ML in production)	67%
Model comparison	67%
Monitoring	50%

As a final trend, there seems to be rather slow adoption of predictive model markup language (PMML), the machine learning language or protocol that provides interchangeability among all of the different machine learning tools and platforms. Only 50% of the vendors support PMML.

Users and Use Cases

A core capability for the unified analytics warehouse is the ability to support different kinds of users and different kinds of use cases. EMA represented the users by title and by tools used by the different personas.
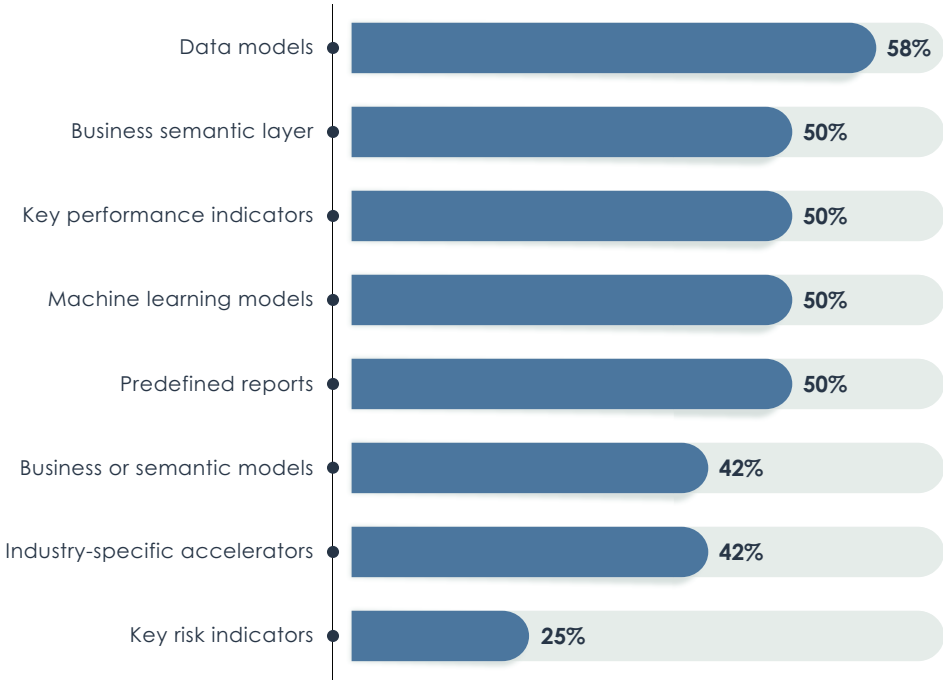
For titles, every vendor supports the administrator, data engineer, data scientist, data explorer, and data analyst. Only three-quarters of the vendors support data sharing and business validation of some kind. For tools and languages used by the different personas, all of the common languages were supported by all or all but one of the vendors. These include ANSI-SQL, ODBC, Python, R, and Notebooks.

EMA is also seeing the need for the development and deployment of smart applications as crucial to success in the digital age. For this purpose, three of 12 vendors support both OLTP and OLAP workloads on their environment. The vendors in the study highly support most other use cases.

Category	Percentage
Data science	100%
SQL analytics	92%
Machine learning	92%
Data discovery	83%
Applications	25%

Business Logic

EMA sees the addition of business semantics and business logic to the unified analytics warehouse as the wave of the future. This is extremely important for the acceleration of business value from an analytics provider. Half of vendors have caught this vision, but only one vendor received a score of 10 out of 10 for their excellence in this area. This is an important area to consider for purchase decisions and for the future.

Category	Percentage
Data models	58%
Business semantic layer	50%
Key performance indicators	50%
Machine learning models	50%
Predefined reports	50%
Business or semantic models	42%
Industry-specific accelerators	42%
Key risk indicators	25%

EMA RADAR

Use Cases: A Closer Look

Reading the Radar Charts

This Radar is specifically intended to provide a useful set of insights into the focus areas and unique strengths of each of the unified analytics warehouse vendors. Only strong offerings were invited to participate, and while there will be some sorting based on *value leader*, *strong value*, and *specific value* for each use case, analytical and platform buyers will be well advised to first define their objectives, then seek out the solution that fits them best—regardless of apparent "rank" or "award." Similarly, the report makes every effort to offer high-level insights into design and function for each vendor so that it, too, can be used as a starting point for planning a unified analytics warehouse investment.

The survey questions covered the five key functions common to all EMA Radar Reports, which include Deployment & Administration, Cost Advantage, Architecture & Integration, Functionality, and Vendor Strength.
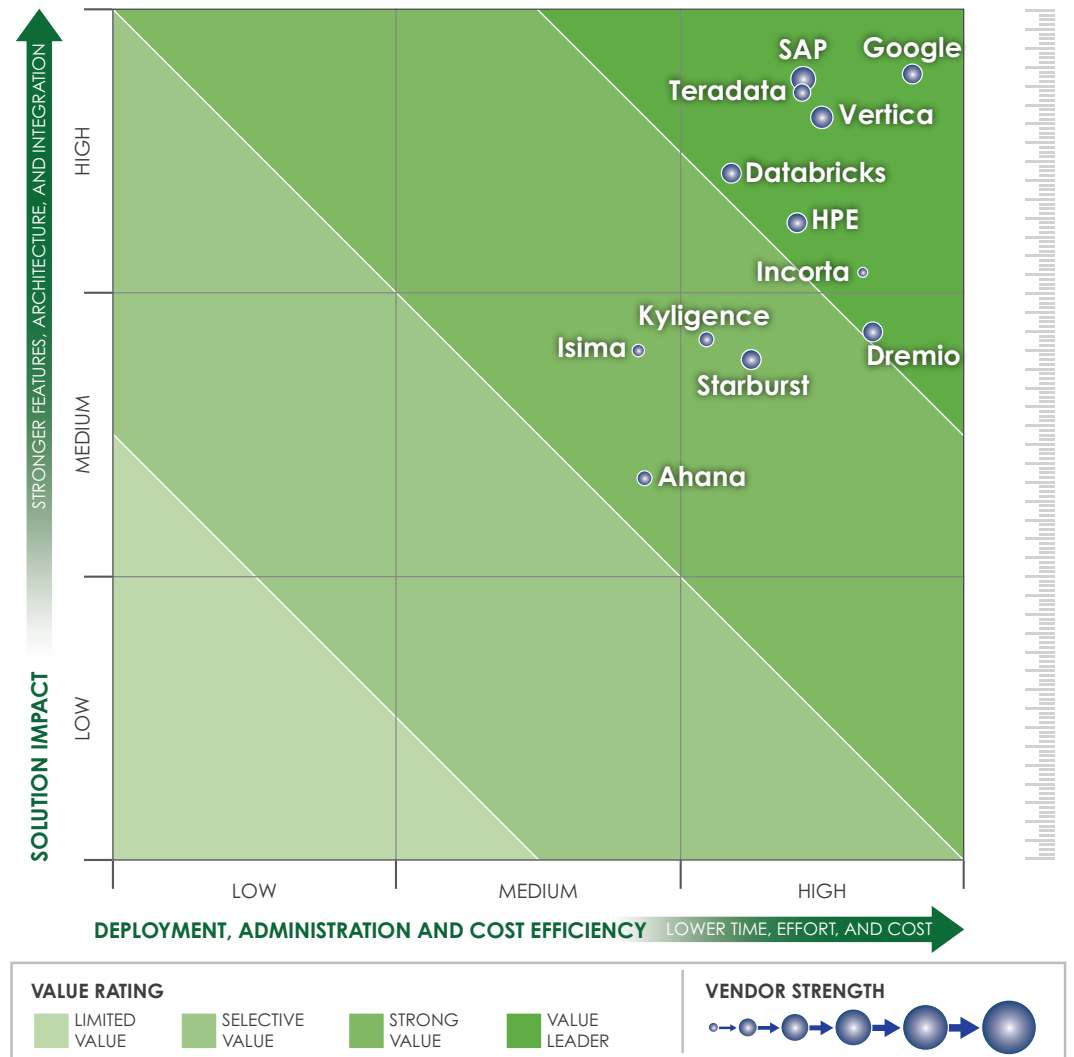
In this Radar for the Unified Analytics Warehouse, there are three Radar Charts to emphasize use case diversity. However, there are some core things to keep in mind as you look at each of these, or look across them in combination:

- The vertical axis, "Product Strength," combines the breadth of product functionality with the foundational nature of the architecture and the ability to integrate with necessary data types and latencies.

- The horizontal axis, "Deployment, Administration, and Cost Advantage," looks at the overall efficiencies of the product, from pre-sales all the way through the entire product lifecycle including purchase, deployment, maintenance, and administration.

- The size of the bubble is reflective of Vendor Strength and represents vision, market strength, and barriers to entry as relevant to the Radar.
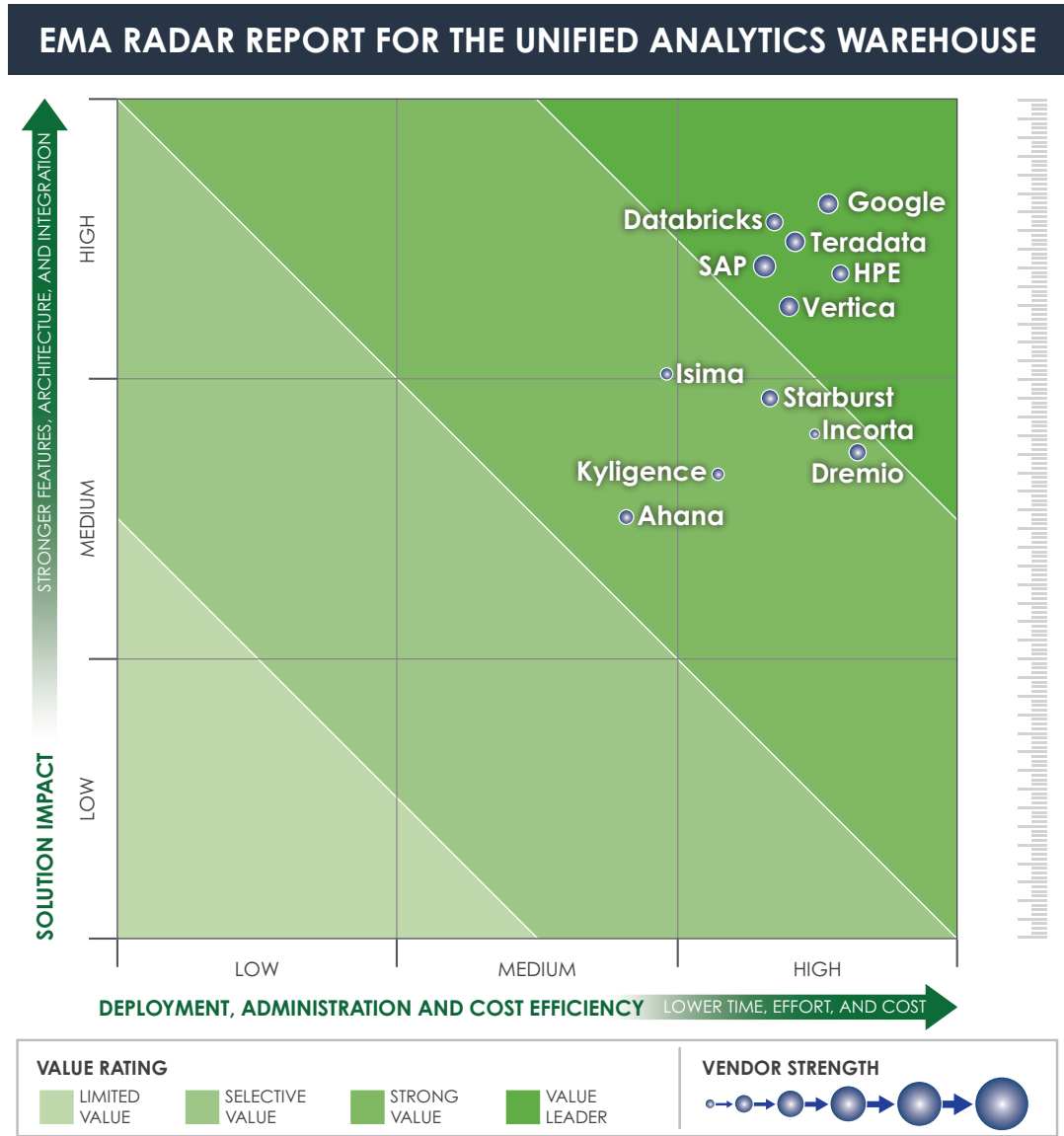
Because of the nature of unified analytics solutions, proximate vendor positions may still reflect radically different advantages. For instance, one vendor's core analytic power may place it next to a vendor at a similar cost advantage, but with weaker core analytics. They may, in fact, represent significantly different types of investment choices. Therefore, no investment decisions should be made without thoroughly considering each vendor's specific profile.

EMA RADAR

Hyperperformance and Enterprise Readiness

EMA RADAR REPORT FOR THE UNIFIED ANALYTICS WAREHOUSE

STRONGER FEATURES, ARCHITECTURE, AND INTEGRATION

SOLUTION IMPACT

HIGH / MEDIUM / LOW

SAP, Google, Teradata, Vertica, Databricks, HPE, Incorta, Kyligence, Isima, Dremio, Starburst, Ahana

DEPLOYMENT, ADMINISTRATION AND COST EFFICIENCY — LOWER TIME, EFFORT, AND COST

LOW / MEDIUM / HIGH

VALUE RATING — LIMITED VALUE / SELECTIVE VALUE / STRONG VALUE / VALUE LEADER

VENDOR STRENGTH

Analytical Flexibility and Low Latency

EMA RADAR REPORT FOR THE UNIFIED ANALYTICS WAREHOUSE

STRONGER FEATURES, ARCHITECTURE, AND INTEGRATION

SOLUTION IMPACT

HIGH

MEDIUM

LOW

Databricks
Google
Teradata
SAP
HPE
Vertica
Isima
Starburst
Incorta
Kyligence
Dremio
Ahana

LOW   MEDIUM   HIGH

DEPLOYMENT, ADMINISTRATION AND COST EFFICIENCY LOWER TIME, EFFORT, AND COST

VALUE RATING

LIMITED VALUE   SELECTIVE VALUE   STRONG VALUE   VALUE LEADER

VENDOR STRENGTH

EMA RADAR

Hyperscale and Platform Leverage

EMA RADAR REPORT FOR THE UNIFIED ANALYTICS WAREHOUSE

STRONGER FEATURES, ARCHITECTURE, AND INTEGRATION

SOLUTION IMPACT

HIGH

MEDIUM

LOW

Starburst
Google
Teradata
HPE
Databricks
Ahana
Dremio
SAP
Vertica
Incorta
Kyligence
Isima

LOW   MEDIUM   HIGH

DEPLOYMENT, ADMINISTRATION AND COST EFFICIENCY LOWER TIME, EFFORT, AND COST

VALUE RATING
LIMITED VALUE   SELECTIVE VALUE   STRONG VALUE   VALUE LEADER

VENDOR STRENGTH

EMA RADAR

Twelve Unified Analytics Warehouse Vendors in Summary

Vendors

Ahana: Ahana, founded in 2020, provides a cloud-native, managed service for Presto, an in-memory distributed SQL engine. Its solution, Ahana Cloud for Presto, includes a SaaS Console that orchestrates and manages Presto clusters in the user's AWS account, a one-step connection or disconnection of data sources, a sandbox with a metastore, object storage, business intelligence, integration with AWS monitoring services, and one-click caching built in to every Presto cluster. Ahana's vision is to provide open data lake analytics to everyone, with unified access.

Databricks: Databricks, founded in 2013, provides the Databricks Lakehouse Platform, a single, open, simple, and collaborative platform to store and manage all data and support all analytics and AI use cases. They also offer Delta Lake, an open and secure data storage and management layer for both streaming and batch, and for structured, semi-structured, and unstructured data. Databricks' combined solution supports business analytics, data science, and machine learning across the entire information lifecycle.

Dremio: Dremio, founded in 2015, provides a data lake engine that enables queries directly against data lake storage and provides a semantic layer that makes the data consumable, consistent, and secure. Its solution is designed around five key features: an open data lake architecture, query optimization and scalability, auto-scaling and shut down, a self-service semantic layer, and data pipeline management.

Google: Google, founded in 1998, offers a fully managed, multi-cloud analytics platform enabling everyone to get insights without constraints for scale, performance, and cost. They do this with BigQuery, a serverless, multi-cloud and real-time data warehouse; Dataproc, a fully managed, highly scalable service for running Apache Spark, Apache Flink, Presto, and 30+ open-source tools and frameworks; Dataflow, a managed real-time and batch streaming data processing platform; and Google Cloud Storage. Google's solution supports five key capabilities: machine learning, in-memory business intelligence, multi-cloud data warehousing, real-time analytics, and interoperability. Google Cloud's vision is to accelerate organizations' ability to digitally transform and reimagine business through data-powered innovation.

HPE: HPE, formed out of the former HP in 2015, offers a unified analytics platform comprised of HPE Ezmeral Data Fabric, a unified platform for multi-use analytics; HPE Ezmeral Container Platform, a unified container platform; and HPE Ezmeral MLOps, a full lifecycle machine learning platform. Its solutions are designed around five key features: enterprise-grade scalability, multi-modal analytical use cases, hybrid deployment including the edge, a broad ecosystem of open-source and independent software vendors, and simple tooling for both developers and data scientists. HPE's vision is to eliminate data silos, enable hybrid deployments, and simplify the developer and data scientist journeys.

Incorta: Incorta, founded in 2013, offers a unified analytics platform that eliminates the need for data integration and comes with built-in business intelligence. Both Incorta and Incorta Cloud are built around five key features: an integrated data and analytics platform, access for technical and business users, direct data mapping to eliminate ETL, speed and agility, and support for a broad set of modern hybrid analytics. Incorta's vision is to enable business teams to innovate data with astonishing agility and speed, maintain their enterprise-class attributes, and drive technical depth and adoption.

Isima: Isima, founded in 2016, offers bi(OS), a hyper-converged analytics platform with a fast path to building feedback loops from ingest to insight for API, AI, and BI. Data analysts using bi(OS) can deploy data discovery, AI, and real-time BI within days and without any data engineering help. Enterprise data teams use bi(OS) in ecommerce, Telco, and finance industries for use cases, such as supply chain optimization, churn reduction, and trade reconciliation. Isima's vision is to help enterprises leapfrog person-years' worth of effort by mature data teams, and to remove time constraints for data practitioners to access data and insight needed to realize their expected business outcomes.

Kyligence: Kyligence, founded in 2016, offers Kyligence Cloud, a big data platform offering data management and analytics services in the cloud. Kyligence is built on Apache Kylin and leverages cloud-native design, data services infrastructure, a unified semantic layer, and an AI-augmented analytics engine to automate and accelerate insight delivery. To round out their offering, Kyligence also offerings Kyligence Insight, their own lightweight business intelligence tool designed to front-end their high-performance OLAP environment. Kyligence's vision is to simplify and automate data analytics on multi-cloud for digital transformation leaders.

SAP: SAP, founded in 1972, offers the SAP Business Technology Platform, allowing users to provide real-time insights through machine learning, AI, business intelligence, and augmented analytics to analyze past and present situations while simulating future scenarios. It includes individual products and product bundles for unified analytics across the entire information lifecycle. At the foundation of SAP analytics offering is SAP HANA, a modern, in-memory database for multi-use analytics. SAP's vision is to help the world run better and improve people's lives.

Starburst: Starburst, founded in 2017, offers Starburst Enterprise, a self-managed solution that provides fast and flexible access to data, making analytics anywhere possible without adding the complexity of data movement and copies. Its solution is built with five key design principles in mind: massive scale and performance, interoperability with data sources, automated management, enterprise security, and world-class customer support. Starburst's vision is to help organizations make faster and better decisions on more complete and accurate data.

Teradata: Teradata, founded in 1979, offers Teradata Vantage, an enterprise analytics software that unifies analytics, data lakes, and data warehouses all on a single platform. Vantage is built on five foundational pillars: hybrid and multi-cloud portability, analytics at any scale, dynamic resource allocation and workload management, modern cloud architectures, and pricing flexibility. Teradata's vision is to be the connected multi-cloud data platform for enterprise analytics from start to scale.

Vertica: Vertica, founded in 2004, is a unified analytics warehouse that unifies HDFS data and object storage data lakes, unifies a company's deployment options spanning multi-cloud and on-premises, and unifies the data science community with the business analyst and IT community on a common platform for machine learning at scale and real-time predictive analytics. Vertica's vision is to provide the highest-performing analytics and machine learning at extreme scale while providing complete freedom from underlying infrastructure and enabling more ways to analyze data.

Awards

Seven of the twelve vendors received awards for their exceptional technology approach to unified analytics. These awards highlight unique or outstanding characteristics or features within the unified analytics landscape, rather than a way to anoint "winners." The goal is to identify and highlight specific solution capabilities with a tie to distinctive customer needs.


Hyperperformance with Multi-Structured Data: Vertica's long commitment to high-performance analytics, combined with their ongoing involvement in open-source technology and complete separation of compute and storage, contributes to their unique ability to analyze multi-structured data in various platforms while maintaining excellent performance.

Multi-Use In-Memory Processing: Databricks has taken the complexity of deploying multi-use, in-memory Spark processing and made it effortless. More than 6,000 customers utilize their platform for a broad range of use cases, including data lakes, data warehouses, data engineering, real-time analytics, and full-service MLOps.


Hyperperformance with Diversified Access: Teradata has a long history as a leader in data warehousing. With Teradata Advantage, they have combined their enterprise-ready analytics engine with a distributed query technology to expand their ability to analyze multi-structured data in various platforms.
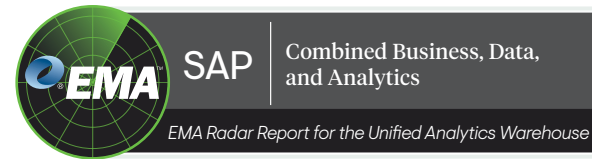
Combined Business, Data, and Analytics: SAP combines more components from across the information supply chain than any other vendor. SAP Data Warehouse Cloud delivers data integration, data preparation, data governance, data sharing, analytics, machine learning, and business intelligence all in a single, fully integrated offering.

STARBURST | Hyperscalability
EMA Radar Report for the Unified Analytics Warehouse
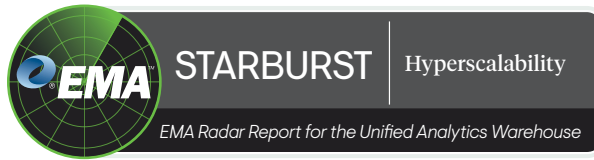
Hyperscalability: Starburst Enterprise for Presto builds on an open-source software technology developed at Facebook and designed to scale up to exabytes of data using SQL queries on distributed data. This level of scalability has yet to be achieved on other data platforms.

GOOGLE | Hyperflexibility at Scale
EMA Radar Report for the Unified Analytics Warehouse

Hyperflexibility at Scale: Google built their cloud platform from the ground up with the customer in mind. This enables them to combine data, analytical, and machine learning services in a way that looks seamless to their customers. Their platform offers the greatest diversity of automation and analytical capabilities in the least complicated manner.

HPE | Enterprise-Ready Cloud Data Lake
EMA Radar Report for the Unified Analytics Warehouse

Enterprise-Ready Cloud Data Lake: HPE Ezmeral has a history of delivering the most enterprise-ready, open-source data lake and analytics platform in the market. HPE's unique combination of a data fabric, open container platform, and support for MLOps adds superior modern cloud capabilities to enterprise analytics and machine learning.

ISIMA | Next-Generation Unified Analytics
EMA Radar Report for the Unified Analytics Warehouse

Next-Generation Unified Analytics: Isima stands out as the most visionary and forward-thinking unified analytics vendor. They are delivering analytics in a way that departs from the standard data approaches of the last four decades by abstracting away the complexities and treating data as an operating system. This is the wave of the future.

Other Vendors Moving Toward the UAW

While 12 vendors make up the heart of the unified analytics warehouse landscape as EMA defines it, the list can never be 100% complete since the unified analytics market is evolving with new and emerging vendors, as well as growing broader capabilities in more established vendors. EMA sees seven additional vendors moving toward the UAW as the market continues to mature.

Amazon Web Services: Amazon Web Services has all of the components for the unified analytics warehouse, but requires engineering resources to put all the services together. AWS did not respond to participate.

Cloudera: Cloudera Enterprise Data Platform knits together open-source technology and adds their own proprietary software to deliver unified analytics.

IBM: IBM has all the components of a unified analytics warehouse and they seem to bring it together in their Cloud Data Pak. IBM did not respond to participate.

Microsoft: Microsoft has all the components of a unified analytics warehouse, but they have not unified them into a single offering.

Oracle: Oracle Autonomous Data Warehouse provides enterprise capabilities necessary for both analytics and data science by providing access to semi-structured data through the Big Data SQL product. Oracle did not respond to participate.

Qubole: The Qubole Open Data Lake platform knits together open-source technology in a unified offering to meet the requirements of the unified analytics warehouse.

Snowflake: Snowflake offers a cloud data warehouse, and they are now attempting to scale for enterprise demands and to include multi-structured data. Snowflake did not respond to participate.

Conclusion

Vendors in the data and analytics platform are clearly moving away from the old paradigm of separation between data lakes and data warehouses. Customer demands are pushing them toward the unified analytics warehouse. The expectation is that a more unified analytics approach will reduce both cost and complexity. In addition, unified analytics should streamline the deployment of analytical applications and yield innovative insight for organizations that pursue the new paradigm.

Business and technology organizations looking to pursue a path forward in unified analytics must clarify their individual business and analytical needs, as well as their current technology landscape, before selecting the right vendor and platform. With vendors pursuing unified analytics from four different per-spectives, the primary barriers to moving forward come from organizational processes, cultures, and commitment to innovation.

The unified analytics warehouse can and should be transformative in enabling intelligent digital business models and providing a competitive advantage against companies stuck in more siloed environments. Which of the three use cases is more relevant to your business requirements and current technology landscape? What are your requirements for multi-structured and multi-latency data and analytics? What are the data requirements in your digital business models for the next two to three years? Which vendor is the most natural fit for your current technology environment and business culture?

Which vendor is likely to bring you the fastest near-term wins? The answer could be any one of the twelve presented in this Radar Report, depending on the answers to these and other questions.

For further consultation on the unified analytics warehouse market, please contact EMA Sales at sales@emausa.com.

EMA RADAR

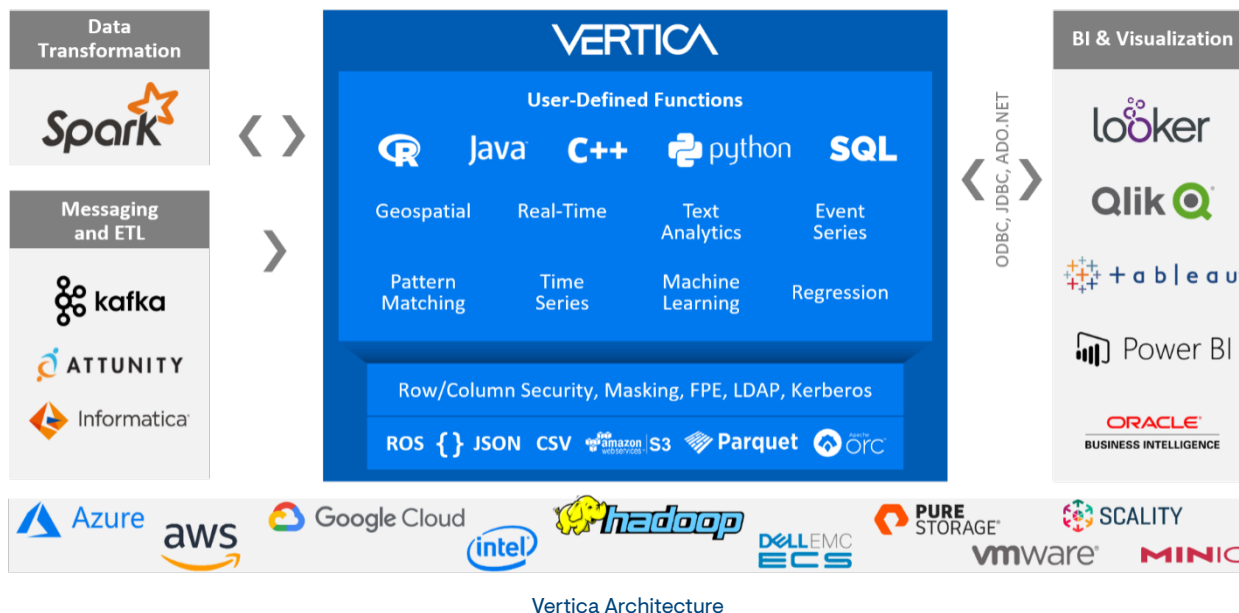**VERTICA**

Solution Description

This EMA Radar Report covers a new category emerging in the data and analytics industry: the unified analytics warehouse. It is a convergence of the data lake and the data warehouse. The UAW is unified because it adequately analyzes multi-structured and multi-latency data in a single platform. It is an analytics platform because the primary use case for both the data lake and the data warehouse has always been analytics. It is a warehouse because it accesses all data in an organized manner, ready for a broad range of analytical users and use cases.

EMA evaluated the 12 vendors based on three different use cases that drive the unification of the data warehouse and the data lake:

- Hyperperformance and Enterprise Readiness: The first use case focuses on high-performance analytics with a complete and proven set of enterprise capabilities.

- Analytical Flexibility and Low Latency: The second use case focuses on extreme flexibility for multiple analytical use cases, including multi-structured data and multi-latency analytics.

- Hyperscale and Platform Leverage: The third use case focuses on maximum scalability and the seamless leverage of existing and low-cost data storage options.

Vertica, founded in 2004, is a unified analytics warehouse that unifies HDFS data and object storage data lakes, unifies a company's deployment options spanning multi-cloud and on-premises, and unifies the data science community with the business analyst and IT community on a common platform for machine learning at scale and real-time predictive analytics. Vertica's vision is to provide the highest-performing analytics and machine learning at extreme scale, while providing complete freedom from underlying infrastructure and enabling more ways to analyze data.

Vertica Analytics Platform is available in two modes: Vertica in Enterprise Mode and Vertica in Eon Mode. Vertica in Eon Mode provisions dynamic workloads as needed, separates storage and compute, and enables workload isolation to serve multiple departments without duplicating the data. Vertica in Enterprise Mode is ideal for fixed workloads, routine reporting, and queries.

Vertica Architecture

EMA RADAR

Unified Analytics Warehouse Highlights

Vertica is the award winner for "Hyperperformance with Multi-Structured Data." Vertica's long commitment to high-performance analytics, combined with their ongoing involvement in open-source technology and complete separation of compute and storage, contributes to their unique ability to analyze multi-structured data in various platforms while maintaining excellent performance.
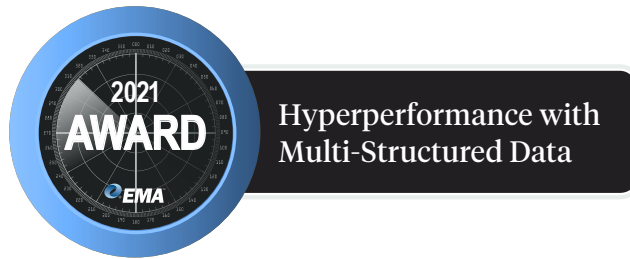
Regarding the three use cases, Vertica is a Value Leader in both the Hyperperformance and Enterprise Readiness and the Analytical Flexibility and Low Latency use cases. They show Strong Value in the Hyperscale and Platform Leverage use cases.

2021 AWARD — Hyperperformance with Multi-Structured Data

2021 VALUE LEADER — Hyperperformance and Enterprise Readiness

2021 VALUE LEADER — Analytical Flexibility and Low Latency

2021 STRONG VALUE — Hyperscale and Platform Leverage

Strengths

Vertica has a long and proven history of providing high-performance analytics with a full array of enterprise capabilities. To bolster analytical performance, they continue to add to their deep offering of analytical functions and machine learning capabilities embedded in their analytical processing engine, speeding time to insight and making it simpler to deploy analytical applications.

Vertica's unified analytics capabilities were greatly enhanced with the release of Vertica in Eon Mode in 2018, which completely separated storage and compute. This move enables them to analyze complex, semi-structured data types and to run queries across multi-structured data stored in object storage and HDFS. It also creates greater flexibility for cloud deployments and delivers its cloud-optimized architecture to on-premises data centers with support for Pure Storage FlashBlade, MinIO, Dell EMC ECS, Scality, and Apache HDFS.

Vertica has also improved their cost-efficiency in two ways. Their all-inclusive pricing model provides a single license for all instances of the platform, both on-premises and in the clouds, for both development and production. In addition, Vertica has a broad set of capabilities built into the product to help enterprise customers manage use and contain cost.
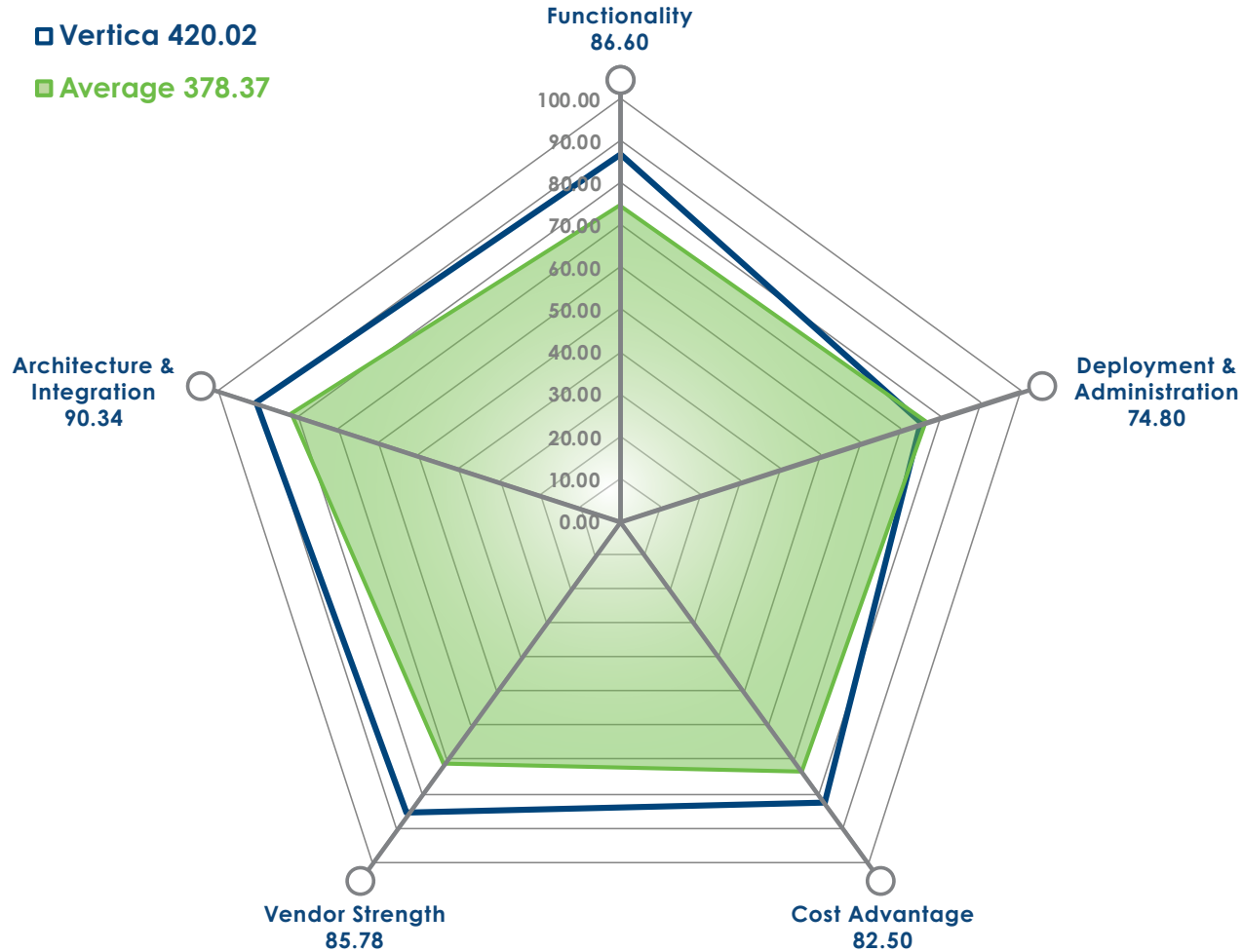
Opportunities

Vertica continues to build out its cloud capabilities around easy-to-deploy offerings via cloud vendors and single-pane administration of all instances on-premises and in multiple clouds. While they have partners who offer Vertica in a SaaS environment, Vertica has an opportunity to develop their own offering.

Vertica also remains on the high end of the number of FTEs required to administer their platform. There remains an opportunity to simplify management, possibly in a SaaS offering, and to automate more of their administrative functions. Lowering the number of FTEs required, combined with their single-pane administration, could give them a nice competitive advantage, especially since there is only one other vendor with full-service, single-pane administrative capabilities.

Finally, Vertica has room for improvement regarding the analysis of streaming data. They can ingest data in micro-batches at subsecond increments, but they are not able to analyze data while it is in motion, or to combine analysis of data in motion with data at rest. This does cover most real-time use cases, but not all.

EMA RADAR

Unified Analytics Warehouse Scoring
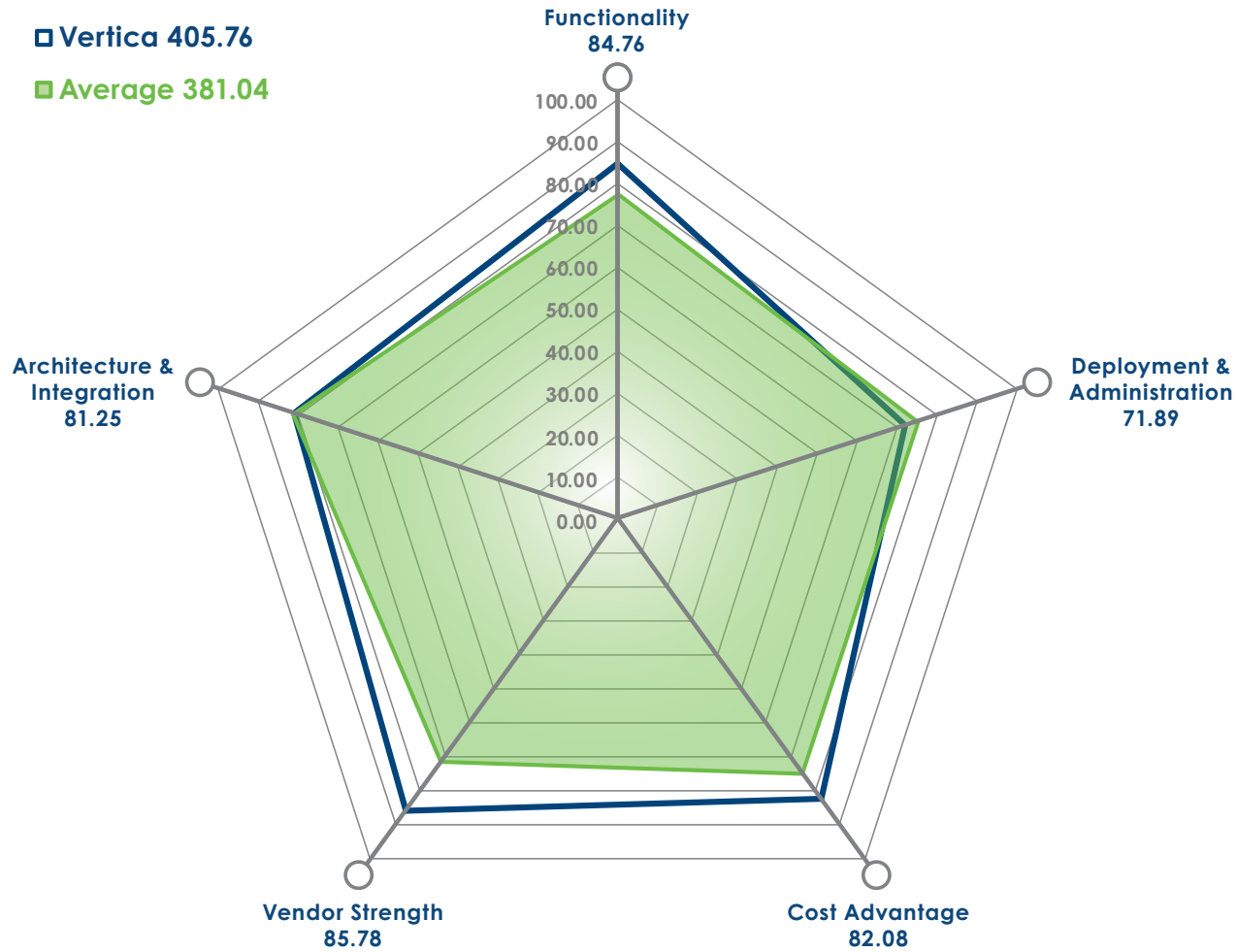
USE CASE: Hyperperformance and Enterprise Readiness

□ **Vertica 420.02**

■ **Average 378.37**

Functionality
86.60

Deployment &
Administration
74.80

Cost Advantage
82.50

Vendor Strength
85.78

Architecture &
Integration
90.34

100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00

Deployment Cost Efficiency	
Deployment & Administration	
Ease of Deployment	
Pre-Sales Deployment	Outstanding
Versatility in Deployment	Outstanding
Deployment Automation	Solid
Time to Production	Strong

Support and Services	
Professional Services	Strong
Customer Support	Strong
User Groups	Strong
Community	Strong

Ease of Administration	
FTEs Required	Strong
Breadth of Administration	Strong
Autonomous Administration	Solid

Cost Advantage	
Licensing	Outstanding
Cost Management	Strong
Estimated Time for ROI	Strong

Product Strength	
Architecture & Integration	
Architecture	
Platform Capabilities	Outstanding
Infrastructure Capabilities	Strong
Cloud Capabilities	Strong

Integration	
Data Capabilities	Outstanding

Functionality	
Product Capabilities	
Enterprise Capabilities	Outstanding
Analytical Capabilities	Outstanding
User Capabilities	Outstanding
Use Cases	Outstanding
Automation	Solid

Business Capabilities	
Business Logic	Strong

Ecosystem	
Partner Ecosystem	Outstanding

Vendor Strength	
Financial Strength	Outstanding
Research & Development	Outstanding
Vision	Strong
Patents	Solid
Geographic Coverage	Outstanding

EMA RADAR
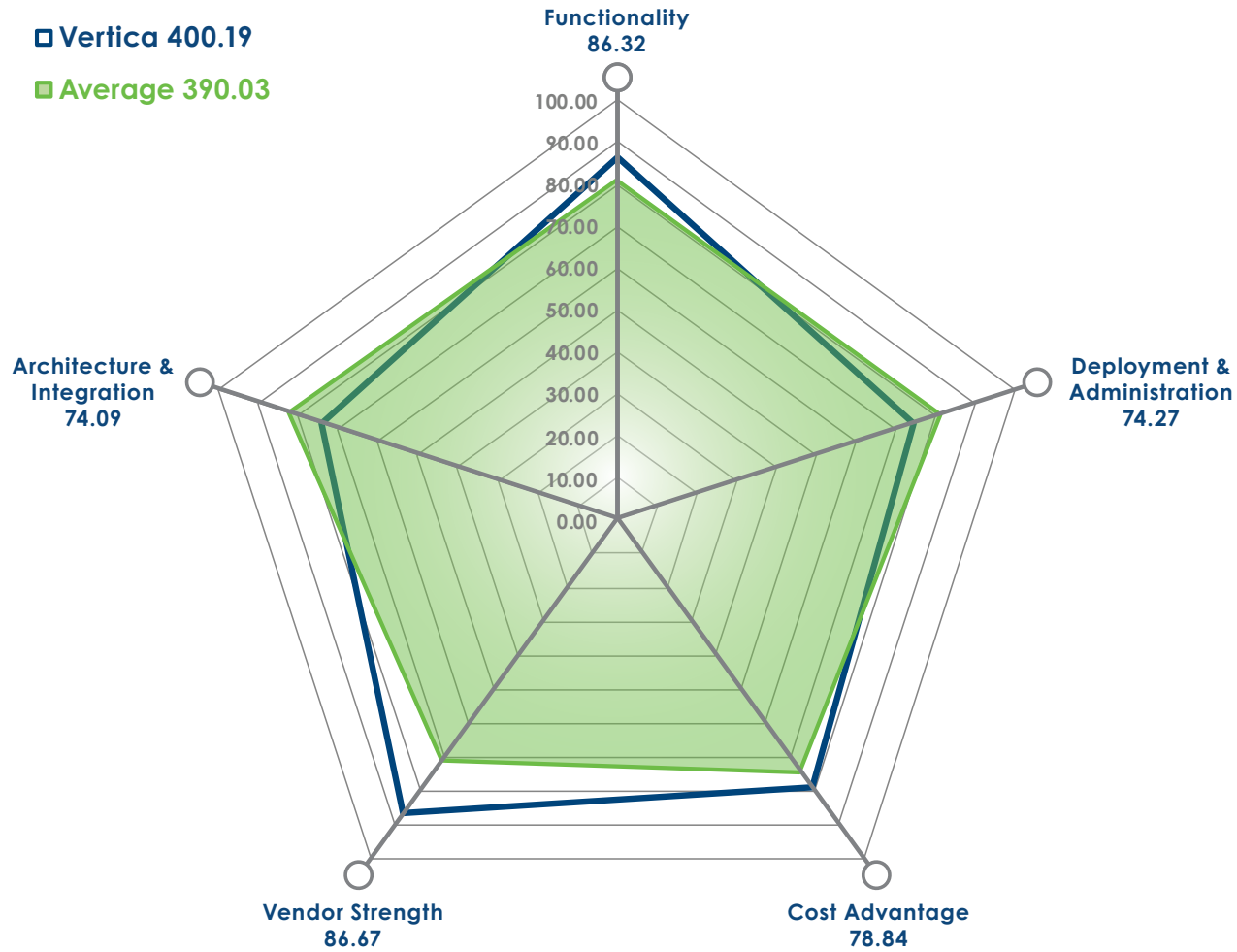
USE CASE: Analytical Flexibility and Low Latency

□ **Vertica 405.76**

▣ **Average 381.04**

Functionality
84.76

Deployment &
Administration
71.89

Cost Advantage
82.08

Vendor Strength
85.78

Architecture &
Integration
81.25

100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00

EMA RADAR

Deployment Cost Efficiency	
Deployment & Administration	
Ease of Deployment	
Pre-Sales Deployment	Outstanding
Versatility in Deployment	Outstanding
Deployment Automation	Solid
Time to Production	Strong

Support and Services	
Professional Services	Strong
Customer Support	Strong
User Groups	Strong
Community	Strong

Ease of Administration	
FTEs Required	Strong
Breadth of Administration	Strong
Autonomous Administration	Solid

Cost Advantage	
Licensing	Outstanding
Cost Management	Strong
Estimated Time for ROI	Strong

Product Strength	
Architecture & Integration	
Architecture	
Platform Capabilities	Strong
Infrastructure Capabilities	Strong
Cloud Capabilities	Strong

Integration	
Data Capabilities	Outstanding

Functionality	
Product Capabilities	
Enterprise Capabilities	Strong
Analytical Capabilities	Outstanding
User Capabilities	Outstanding
Use Cases	Strong
Automation	Solid

Business Capabilities	
Business Logic	Strong

Ecosystem	
Partner Ecosystem	Outstanding

Vendor Strength	
Financial Strength	Outstanding
Research & Development	Outstanding
Vision	Strong
Patents	Solid
Geographic Coverage	Outstanding

EMA RADAR

USE CASE: Hyperscale and Platform Leverage

☐ **Vertica 400.19**

☐ **Average 390.03**

Functionality
86.32

Deployment &
Administration
74.27

Cost Advantage
78.84

Vendor Strength
86.67

Architecture &
Integration
74.09

EMA RADAR

Deployment Cost Efficiency	
Deployment & Administration	
Ease of Deployment	
Pre-Sales Deployment	Outstanding
Versatility in Deployment	Outstanding
Deployment Automation	Solid
Time to Production	Strong

Support and Services	
Professional Services	Strong
Customer Support	Strong
User Groups	Strong
Community	Strong

Ease of Administration	
FTEs Required	Strong
Breadth of Administration	Strong
Autonomous Administration	Solid

Cost Advantage	
Licensing	Outstanding
Cost Management	Strong
Estimated Time for ROI	Strong

Product Strength	
Architecture & Integration	
Architecture	
Platform Capabilities	Solid
Infrastructure Capabilities	Strong
Cloud Capabilities	Strong

Integration	
Data Capabilities	Strong

Functionality	
Product Capabilities	
Enterprise Capabilities	Outstanding
Analytical Capabilities	Outstanding
User Capabilities	Outstanding
Use Cases	Strong
Automation	Strong

Business Capabilities	
Business Logic	Strong

Ecosystem	
Partner Ecosystem	Outstanding

Vendor Strength	
Financial Strength	Outstanding
Research & Development	Outstanding
Vision	Strong
Patents	Solid
Geographic Coverage	Outstanding

EMA RADAR

Measurement Criteria: Appendix A

Research for the Q2 2021 Unified Analytics Warehouse Radar Report took place starting in Q4 2020. Vendor input is included in the process of updating the measurement criteria. EMA used the following requirements to evaluate the participating vendors. Please keep in mind that these categories were weighted differently depending on their importance to a multifunctional, unified analytics warehouse, as well as their relevance to each of the three different use cases addressed: hyperperformance and enterprise readiness, analytical flexibility and low latency, and hyperscale and platform leverage.

Deployment and Cost Efficiency

Deployment and Administration
Ease of Deployment

Pre-sales deployment: Availability of a free version of the software and the limitations placed on the free version compared to the production version

Proof of concept availability: Availability of support for a proof of concept prior to deployment

Versatility in deployment options: Where and how the solution is deployed and made available for testing and production

Deployment automation: Automation for functions such as system configuration, data discovery, data profiling, data tiering, and data modeling

Time to production: Time to ingest data, prepare data for analytics, and start delivering insight in production

Accelerators: Types and breadth of accelerators offered to speed time to production

Support and Services

Breadth of professional services: Breadth of professional services available directly and through partners for deployment and strategic initiatives

Levels of customer support: Different levels of customer support from online to on-premises

User group offerings: Vendor-provided user services including forums, conferences, education, and certifications

Strength of community: Size and nature of the vendor community from loosely knit users of technology to enthusiastic believers

Ease of Administration

FTEs required for administration: Requirements for ongoing administration based on vendor estimates and customer interviews

Breadth of administration: Administrative reach across data stores, security, encryption, access performance, workload management, resource management, application monitoring, infrastructure monitoring, and governance

User and data administration: Capabilities for administrators to manage users and data for their entire organization

Autonomous administration: Automation of formerly manual tasks across all aspects of data management

Cost Advantage

License model: Diversity of licensing models such as subscription, pay per use, and perpetual, as well as the means of calculating cost

Breadth of licensing: Ability to consolidate licenses for different deployment options and different analytics and data management capabilities

Cost management: Ability to estimate cost and manage business requirements

Cost containment: Capabilities to help customers manage cost across a broad number of users and use cases on a unified analytics platform

Estimated time for ROI: Vendor provided estimates for "shortest time to achieve ROI" that are supplemented by case studies and customer interviews

Product Strength

Architecture and Integration
Architecture

Platform strengths: Analysis of platform strengths as a match for each of the three use cases

Platform breadth: Range of capabilities unified in a single solution from across the information lifecycle

Infrastructure capabilities: Breadth of server and storage offering as part of a unified analytics solution

Cloud capabilities: Analysis of cloud readiness for portability, elasticity, and abstraction of cloud complexity

Integrations and Interoperability

Breadth of data: Range of different data types stored or accessed by the unified analytics platform

Multi-structured data analysis: Efficiency of analyzing different types of structured and semi-structured data

Multi-latency data analysis: Ability to ingest and analyze data at low latency or real time

Functionality
Product capabilities

Enterprise capabilities: Support for enterprise and digital business models with security, scalability, agility, and availability

Analytical performance: Speed of analytical processing and support for high query concurrency

Reach of single queries: Ability to run queries across diverse data sets and different data stores

Analysis of complex data types: Ability to analyze different complex data types without having to reformat data

Machine learning: Support for the entire machine learning lifecycle with built-in capabilities

User capabilities: Support for different types of access to unified analytics and different types of users from administrators to business users

Use cases: Support for different types of analytical and operational use cases by a single platform

Automation: Automation for functions such as query routing, ingestion, model generation, model recommendation, data profiling, and metadata generation

Business capabilities

Business logic: Inclusion of business logic and semantics for business areas and industries

Ecosystem

Partner ecosystem: Analysis of partner ecosystem, specifically looking at technical integration, joint sales and marketing, services, and open-source

EMA RADAR™

Vendor Strength

Financial strength: Ongoing revenue and annual growth

Research and development: Percentage and total dollars spent on research and development

Vision: Full product and corporate vision

Patents: Registered and pending patents as a barrier to entry for competitors

Geographic coverage: Geographic customer reach and linguistic support