June 27, 2022

OFFERING OVERVIEW

Vertica Brings Flexibility to Deployment Choices and Analytical Options

Vertica Spans On-Premises, Cloud, and Hybrid Deployment Choices and Analytics and Data Science Needs

Doug Henschen
Vice President and Principal Analyst

Produced exclusively for Constellation Research clients

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Vertica is a leading analytics platform that can be deployed on-premises, in the cloud, or in hybrid fashion and is also available as a database service on Amazon Web Services (AWS). Vertica is one of the most important products in the portfolio of Micro Focus International, a global software company based in the U.K. with annual revenues of $2.9 billion (fiscal year 2021). Vertica, which hit the market in 2007 as a columnar scale-out database capable of powering ultra-high-scale analytics, was acquired by Micro Focus in 2017.

Vertica has evolved from a database into a modern analytics platform that can unify myriad data sources into warehouses and across data lakes. Vertica spans SQL, advanced analytics, and data science requirements such as machine learning (ML). It can also be flexibly deployed on-premises, in multiple clouds, or in hybrid environments as well as consumed as a database service. Vertica is also embeddable into third-party software and software-as-a-service (SaaS) offerings.

This Offering Overview provides an overview of Vertica. It identifies key differentiators; examines functional capabilities; considers Vertica's strengths and weaknesses; and provides use cases, including two customer case studies. Technology buyers should use this report to evaluate Vertica for implementation.

Business Themes

 Data to Decisions

 Technology Optimization

COMPANY OVERVIEW

Micro Focus is a global software and software consultancy firm headquartered in Newbury, Berkshire, England. Founded in 1976, the company had its earliest success in supporting mainframe and microcomputer workloads. Since going public, in 2005, Micro Focus has been highly acquisitive, purchasing multiple software product lines and more than seven companies. Notable acquisitions include NetManage, Borland, Attachmate, and Serena Software. Vertica joined the Micro Focus portfolio through Micro Focus's 2017 merger with the software business unit of Hewlett Packard Enterprise (HPE).

Micro Focus

- **Company:** Micro Focus International, Plc.
- **Headquarters:** Newbury, England
- **Founded:** 1976
- **Type:** Public (MCRO on London Stock Exchange and MFGP on New York Stock Exchange)
- **Fiscal 2021 Revenue:** $2.9 billion
- **No. of Employees:** 12,000
- **Website:** MicroFocus.com
- **Twitter:** @MicroFocus

Today Micro Focus focuses on delivering mission-critical software that helps organizations run their operations while also moving to digital technologies. In 2021 it embraced a strategy to transition to a product group operating model around five portfolios:

- **Application Delivery Management** products focused on streamlining software delivery for shorter time to value
- **Application Modernization and Connectivity** products focused on modernizing core business system to drive transformation
- **CyberRes** products that put security at the core of operations, applications, identity, and data
- **Information Management and Governance** products, including Vertica, focused on advanced analytics; content management; and proven legal, compliance, and privacy solutions
- **IT Operations Management** products focused on simplifying IT transformation

The Information Management and Governance (IM&G) portfolio generated $391 million in revenue in fiscal year 2021—about 13% of the company's total revenue for the year. License revenue in the IM&G category increased by 15.8% during that period, with the increase "primarily driven by growth in Vertica," according to the company's 2021 annual report.

ABOUT VERTICA

Overview

Vertica was cofounded in 2005 by database innovator Michael Stonebraker and seed investor Andrew Palmer and was acquired by HP in 2011 and Micro Focus in 2017. The Vertica relational database management system (RDBMS), introduced in 2007, was designed to manage vast data volumes, scaling into the petabytes and harnessing massively parallel processing (MPP) and columnar storage for fast query performance.

What's now known as the Vertica Unified Analytics Platform has steadily evolved, first adapted to work in conjunction with high-scale data lakes built on Apache Hadoop and then, more recently, in conjunction with object storage. Bulk data loaders and capabilities for direct querying of file formats such as Optimized Row Columnar (ORC) and Parquet have been steadily enhanced for modern data lakes.

Vertica's support for data science workloads also has steadily expanded, now including geospatial, time series, and ML analysis; flex tables for schema-on-read flexibility; and more extensive in-database support for ML algorithms and languages such as Python.

Vertica in Eon Mode, introduced in 2018, separates compute and storage decisions for cost savings and improved elasticity in public or private cloud deployments. Vertica Accelerator, a SaaS offering based on Eon Mode, was made generally available on Amazon Web Services (AWS) in September 2021. An Accelerator service on Microsoft Azure is expected to be added in 2022.

Vertica cloud deployment options go well beyond ready-to-deploy images on leading cloud marketplaces. For example, work continues on extending the platform's hybrid and multicloud

deployment options based on Kubernetes. The platform supports a long list of on-premises object storage options as well as offering prebuilt integrations with cloud-native services on all three major public clouds: AWS, Azure and Google Cloud. The goal is to enable customers to run Vertica anywhere they want, whether on-premises or in any cloud, to address business intelligence, advanced analytics, and data science needs on a single, consistent platform.

Market Segment: Analytical Data Platforms

Market Definition

The high-scale analytical data platforms market has flourished and drastically evolved over the last two decades. Whereas the market once centered on a dozen DBMSs that were deployed on-premises, today most of the attention has turned to DBMSs, lake query engines, and blended lake/warehouse platforms offered as services on public clouds. Still available and still very relevant are choices such as DBMS software that can be deployed on-premises, combined hardware/software (aka appliance) analytical platforms deployed on-premises, and DBMS and lake query engine marketplace offerings that can be deployed by customers in public clouds.

Data lake platforms represent another type of high-scale analytical platform, and today they're increasingly built on cloud-based object stores. Apache Spark and Hadoop are still prevalent, and these platforms are also deployed by customers, either on-premises or in public clouds, or are consumed as cloud-based services run in public clouds.

The new breed of combined lake/warehouse offerings supports data engineering, data science, and data warehouse/business intelligence (BI) workloads against a shared storage environment. These offering's vendors invariably tout the simplicity of having a single security and access control scheme and unified governance of data on one platform.

DBMSs remain the backbone of the vast majority of data warehouses that support BI/analytical workloads. The newcomers to the market are SQL query engines designed to work with data stored in data lakes and/or distributed data fabrics. Their vendors tout the advantages of querying data where it already lives: either in a lake (most frequently) or in distributed stores accessed via a virtualized access/federation approach.

Whether it's an analytical DBMS or a query engine designed to work with lakes, customers will expect it to support the querying required for BI, including scheduled reporting and ongoing refresh of executive and operational dashboards that might have tight service-level agreements (SLAs). Thus, these DBMSs and query engines must offer query tuning, data tiering, and data caching capabilities to support performant querying against high-scale data as well as many concurrent users and queries. DBMSs and query engines might also be taxed with unpredictable ad hoc query-and-analysis workloads, adding yet more workload management challenges on top of the reporting and dashboarding SLAs.

Data lake platforms (and the lake side of combined lake/warehouse offerings) enable organizations to go beyond the structured and semistructured data typically associated with data marts, data warehouses, and SQL-centric querying. Lakes can ingest any data and provide a platform for data transformation and data science analysis at scale. Lakes routinely handle internet clickstreams, sensor data, log files, mobile data rich with geospatial information, and text extracted from customer relationship management (CRM) call records and social network interactions.

The data lake's combination of data type flexibility and lower cost of storage (compared with DBMSs) has become a foundation for innovative and value-driving analyses. What's more, data lakes serve as a platform for data engineering at scale. Lake-based data processing is often used to feed structured data into warehouse platforms. Lakes also support predictive data science and ML workloads that would be difficult; costly; and, in some cases, technically impossible to support on SQL-centric platforms.

Market Trends

The analytic platforms market was sleepy and consolidated before it exploded in the 2000s as organizations increasingly grappled with rising quantities of data, a desire to develop novel insights from unused data, and expectations for ever-faster analysis. The leading general-purpose relational database management systems (RDBMSs) at the time—Oracle Database, Microsoft SQL Server, IBM Db2, and MySQL—were being used for data warehousing, but they had yet to be highly adapted for analytical use. Pioneers of the high-scale analytical data platforms market harnessed MPP—employed

by Teradata beginning in the 1980s—and column-store architectures—harnessed by Sybase IQ (now SAP IQ) in the early 1990s.

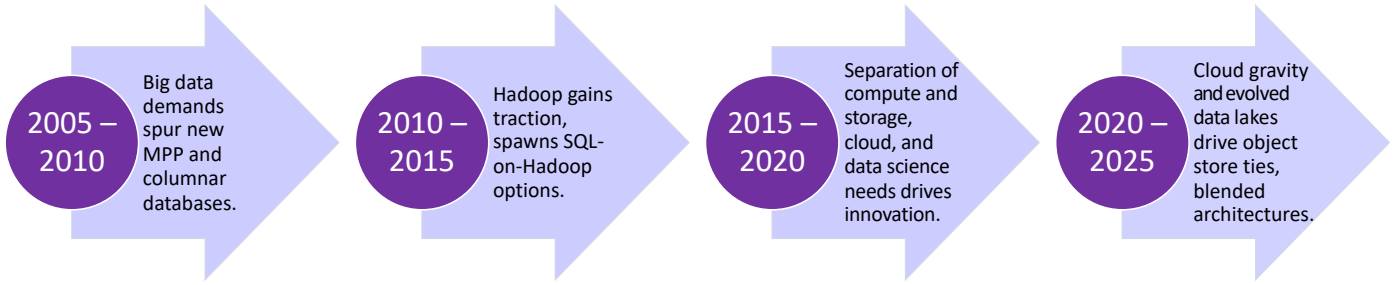
As shown in Figure 1, as the big data era emerged, in the mid-2000s, a raft of new vendors and platforms burst onto the scene, with MPP, columnar architectures, and purpose-built analytic appliances coming to the fore. Vertica, Netezza, and Greenplum were among the vendors that emerged in this era. MPP (also known as scale-out architecture) soon became the cornerstone of many platforms, including yet more DBMS options, Apache Hadoop, Apache Spark, and many NoSQL stores.

Organizations started being captivated by the promise of low-cost storage in Hadoop, and that development was gathering steam by 2012. Deployments multiplied, but the complexity of this new platform limited access to data. Analytical DBMS technologies were soon adapted (and multiplied yet again) to support SQL-on-Hadoop analysis, filling the accessibility void by bringing the familiarity of relational querying to BI-curated datasets within data lakes.

Innovations continued through the second half of the last decade, with yet more features introduced for in-database data science as well as more sophisticated (hot/warm/cold) data tiering and caching schemes aimed at optimal query performance.

The stampede to the cloud has been the most powerful market driver over the last decade, with organizations increasingly moving workloads—and, therefore, data to be analyzed—into public clouds.

Figure 1. The Evolution of High-Scale Analytical Data Platforms: 2005–2025



Source: Constellation Research

Tech flexibility and business and innovation agility have been the key draws to the cloud, although scalability and elasticity are also important for handling big data and spiky analytical workloads.

Cloud advantages and requirements have fueled yet more customer interest in new features, including automated "serverless" scaling, automated systems management capabilities, and separation of compute and storage decisions. The move to the cloud (coupled with the complexity of managing Hadoop) also led to a new generation of data lakes built around low-cost cloud object storage.

With the new generation of object-store-based data lakes, we've seen two additional trends. First, query engine platforms have emerged that are geared to supporting SQL-centric BI workloads directly against data in data lakes. Second, combined lake/warehouse offerings have become available, supporting data engineering, data science, and data warehousing on a single, shared data platform.

As we move toward 2025, Constellation expects to see more extensive use of object storage, including as the foundational storage layer for databases as well as query engines and data lakes. Data fabrics are also gaining ground, with several vendors working on extended capabilities for accessing data where it lives and selectively moving compute to the data or data to the compute, as required, to meet performance demands.

FUNCTIONAL CAPABILITIES

Vertica's combination of high scalability and superior query performance was designed in at the product's inception, thanks to the combination of MPP and column-store query-speed and compression advantages. Over the last 17 years, the database has been extended to become a versatile analytics platform spanning data sources, analytical requirements, storage choices, and deployment choices.

Detailed below are key categories of Vertica Unified Analytics Platform functionality that help organizations unify data silos, span advanced analytical and data science needs, and meet hybrid and multicloud deployment requirements. At the conclusion of each section are the latest upgrades introduced in June 2022 with the release of Vertica 12.

Unification of Data Silos

Vertica handles the analysis of conventional structured sources as well as columnar, sparse, and variable data types common to data lakes built on object storage and Hadoop as well as NoSQL stores. Features include:

- Standard SQL support (ANSI 99)
- JSON and log file format support; Kafka read/write and ORC and Parquet readers for querying object-storage-based data lake platforms and Hadoop
- Parquet export for writing result sets back to external data lake platforms
- Role- and column-based security for secure integration with source systems
- Support for querying complex data types, such as maps, arrays, and structs stored in Parquet and common to Amazon S3 and other object storage options
- Vertica 12 upgrades such as:
    - Faster analysis of Parquet files and extended support for complex data types
    - Federal Information Processing Standards (FIPS) adoption, which broadens single-sign-on capabilities with OAuth 2.0 token authentication support for JDBC and ODBC clients

Unification of Analytics and Data Science

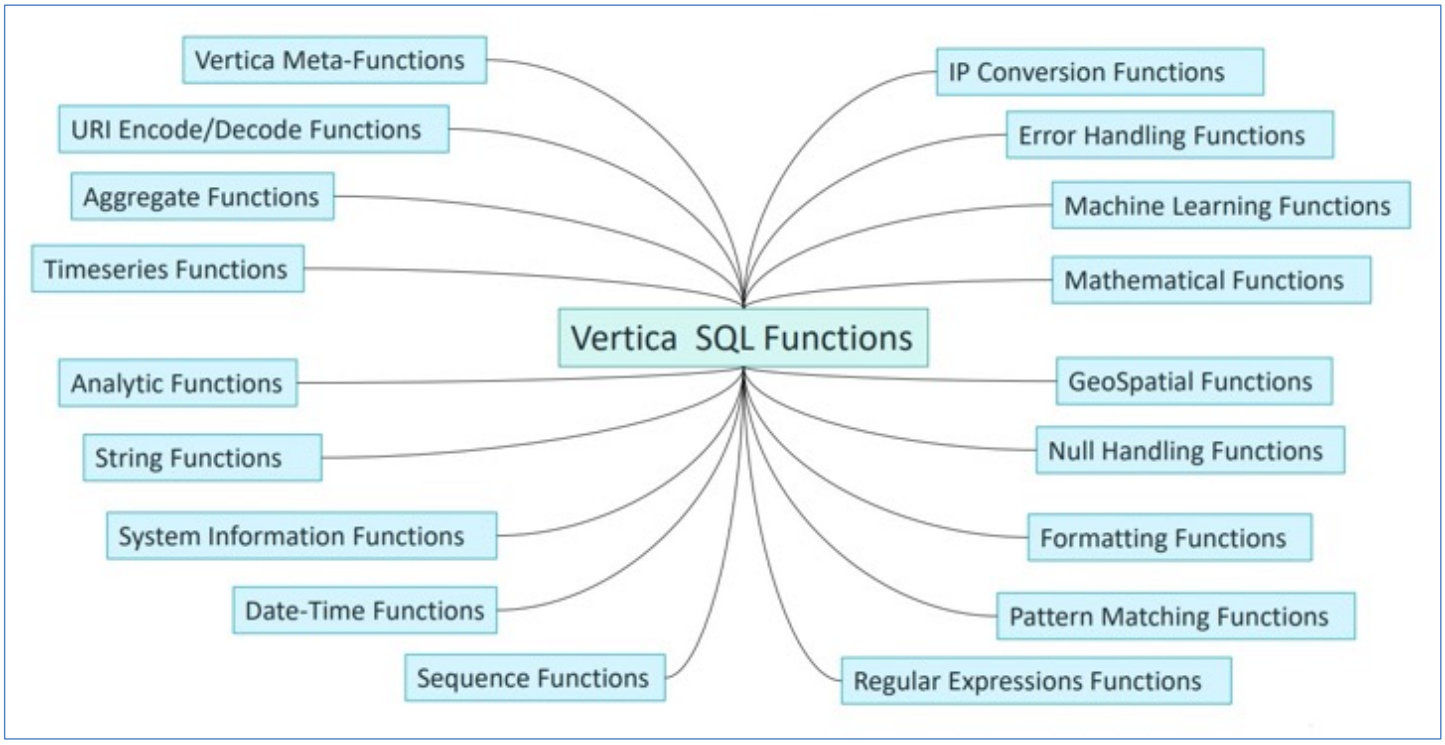
The analytic requirements of leading organizations have extended well beyond standard SQL. Vertica has been steadily enhanced to support a broad range of BI, advanced analytics, and data science requirements. Capabilities include:

- More than 650 advanced SQL functions supporting pattern matching, time series analysis, ML, text, and geospatial analysis (summarized by category in Figure 2) implemented in C++, with API support for Java, R, and Python

- 15 built-in in-database ML models, including regression (linear regression, random forest, vector machine, XGBoost), classification (logistic regression, Naive Bayes, random forest, vector machine, XGBoost), unsupervised clustering and anomaly detection (k-means, outlier detection, isolation forest), and time-series forecasting (ARIMA, SARIMA, VARIMA)

- Vertica user-defined extension (UDx) framework, which enables parsers and data loaders written in C++, Java, R, and Python to be run in-database as Vertica SQL functions

Figure 2. Vertica Offers More Than 650 SQL Analytical Functions in 18 Categories



Data Source: Micro Focus

- Predictive Model Markup Language (PMML) export option, which supports training of models on Vertica, use of parallel in-database algorithms, and export of finished models to external systems for scoring

- PMML import, which supports training of models on a choice of external frameworks and languages and import to Vertica, where models can be scored and then managed as easily as tables and run on very large datasets

- TensorFlow model import, which supports training of custom ML models, including deep learning neural network models, on a choice of compute platforms, including high-performance compute or GPU-based systems (Models saved in TensorFlow's native graph format can be imported to Vertica and run on very large datasets.)

- Open source Python client and Python SDK for Vertica, which support the use of Jupyter notebooks and in-Vertica processing of user-defined functions expressed in Python as well as interactive exploration, experimentation, and visualization in Jupyter

- Vertica 12 upgrades such as:

    - Two new built-in ML models: isolation forest and ARIMA

    - New features for the VerticaPy open source Python/Jupyter interface, including Python Pickle and an enhanced Graphviz package

    - Stored procedures advancements such as new extensions for geometry, geography, and altering of existing procedures

    - Open source Spark connector, which now supports exchange of Parquet data and other columnar data formats on S3, other object storage options, and Hadoop Distributed File System (HDFS)

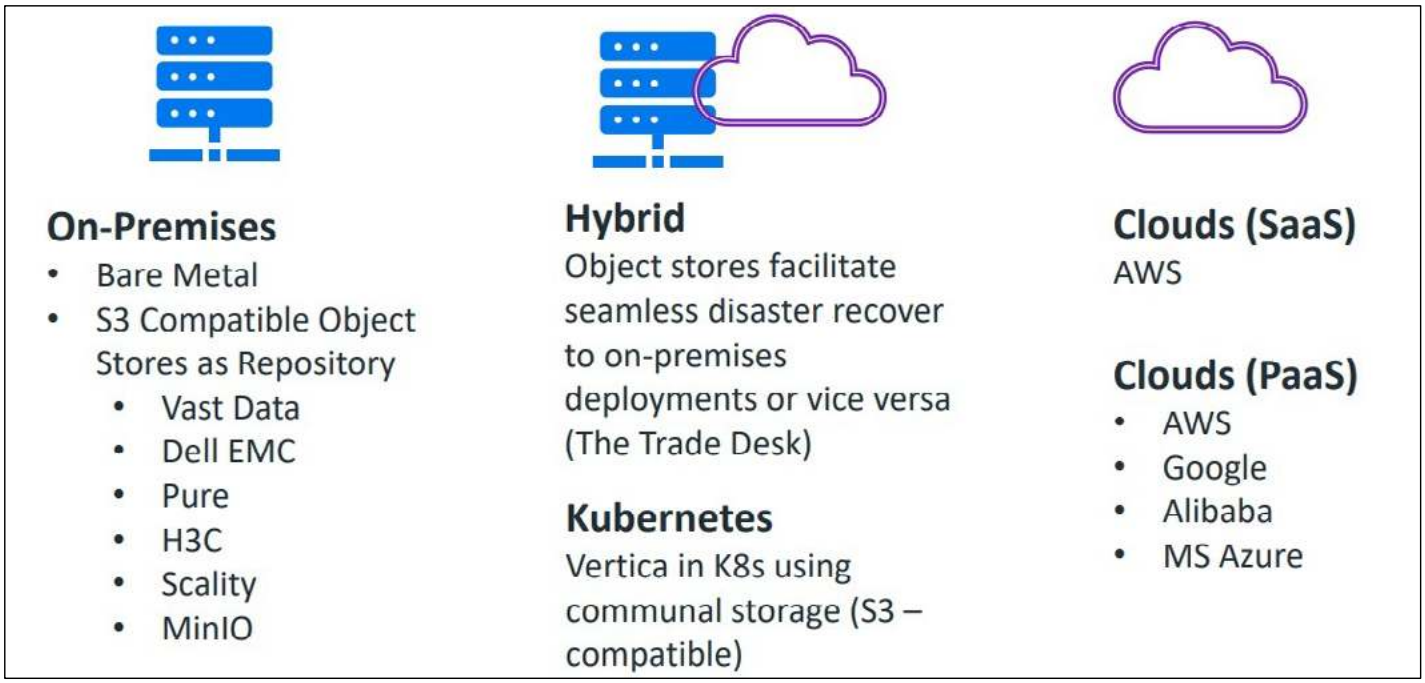
Unification of Hybrid-Cloud and Multicloud Deployments

Core Vertica Unified Analytics Platform management and administrative features and the Eon Mode option simplify the deployment and use of Vertica in on-premises deployments, in multiple public clouds, or in hybrid deployments (as detailed in Figure 3). Relevant Vertica features include:

- Docker images and support for Kubernetes and a growing number of Kubernetes-like orchestration tools, including OpenShift, Azure Kubernetes Service (AKS), Amazon Elastic Kubernetes Service (EKS), and Google Kubernetes Engine (GKE)

- Storage-agnostic design, which enables different tables or partitions of a single table to be stored in different locations (Recent data, for example, can be stored on disk, whereas older data can be placed on object stores, storage-area networks [SANs], HDFS, or a mix of options.)

Figure 3. Vertica Deployment Options Include On-Premises, Hybrid Cloud With Kubernetes, SaaS (Accelerator on AWS), or Cloud With Native Platform Services



Source: Micro Focus

- Vertica in Eon Mode separates compute and storage, supporting shared storage for all workloads and introducing independent elastic scalability and cost savings for compute. The Eon Mode deployment configuration option:

    - Supports dynamic workload management, with rapid elastic scaling of compute subclusters to address fast-emerging and ephemeral workloads (Subclusters can be quickly shut down for cost savings.)

    - Enables compute hardware choices and configurations that can be matched to workloads, an advantage over appliances with limited options

    - Isolates workloads via a subcluster design that provides dedicated compute capacity and more certain compliance with service-level demands

    - Eliminates redundant copies of data and reduces extract, transform, and load (ETL) requirements by enabling multiple subclusters to read a common database

    - Enables organizations to back up Vertica Enterprise databases and relaunch them in Eon Mode via a migration feature

- Vertica 12 upgrades such as:

    - Cloud-optimized architecture improvements for subcluster use for ELT and data ingestion

    - Expanded list of compatible on-premises object stores, supporting options from VAST Data and H3C added to already-supported options from Dell, MinIO, NetApp, Pure Storage, and Scality

    - Expanded Kubernetes storage support beyond S3 to include Google Cloud Storage, Microsoft Azure Blob Storage, and HDFS

USE CASES AND REFERENCE CUSTOMERS

Vertica is used in a broad range of analytical use cases, with leading examples such as:

- Communication and network analytics
- Customer behavior analytics
    - Clickstream analytics
    - Customer churn predictive analytics
- Data warehouse modernization
- Embedded analytics
- Fraud monitoring and risk management
- Gaming data analytics
- Internet of Things (IoT) analytics
- Telecom data analytics
- Utilities data analytics

Customers can draw on case studies in each of the use case categories listed above. Vertica Quickstarts—prebuilt integrations geared to rapid configuration and deployment—are available for partners such as Informatica, Google Looker, MicroStrategy, Qlik, SAP, Tableau, Talend, and TIBCO Spotfire.

When an organization chooses Vertica, it's often because it helps customers address high-scale and high-performance requirements, flexible deployment requirements, or both. Following are two short Vertica customer case studies in which these challenges were successfully met.

The Trade Desk Thrives on Vertica Scalability

The Trade Desk has long been one of Vertica's highest-scale customers. In fact, its demanding requirements have helped drive the Vertica roadmap, including the development of Eon Mode, first introduced in 2018.

The Trade Desk is an ad tech company that helps customers—mostly large ad agencies—place viewer-relevant ads on websites, streaming TV shows, and mobile apps. It makes this connection via real-time auctions that take place within milliseconds. To say that the company operates on a massive scale is almost an understatement: The company drives 8 million to 14 million ad impressions per second. To support this mission, the company ingests and uses Vertica to analyze 700 billion to 1.2 trillion rows of data per day. Long-term retention of this data was also important for reporting purposes, a challenge that begged for the option of using low-cost object storage.

"We were a big supporter of Eon [Mode] when it was initially being designed, because we had a requirement to store five years' worth of data for our customers," explains Bert Corderman, senior manager of engineering at The Trade Desk. "When we priced out what it would cost to support that workload with a traditional MPP database, it would have been astronomical. We would not be where we are today without the separation of compute and storage and the option to use object storage."

The Trade Desk's early-adopter move from Vertica Enterprise deployed on-premises to Eon Mode deployed in the cloud brought multiple benefits. The company uses Vertica for reporting, ad hoc BI, and inputs into some data science workloads, but reporting is a particularly high-scale but spiky workload. The company delivers more than 40,000 reports per day against massive tables approaching 1 petabyte in size. All these reports must be delivered within a six-hour time window. With Eon Mode separation of compute and storage, The Trade Desk can shut down all that compute capacity for 18 hours out of each day, which drives a huge cost savings.

The Trade Desk also needed deployment flexibility. Today the company has two 12-petabyte deployments of Vertica, both in Amazon, in addition to two smaller clusters on Microsoft Azure. Systems are deployed in pairs for the sake of redundancy and resilience. Multicloud deployment flexibility was a must-have, according to Corderman, but the company is also considering using Vertica's hybrid option, in which case the backup system could move to an on-premises data center.

"When I look at the ingest rates into Vertica from Amazon S3, we're pushing more than 100 gigabytes per second every night, which is just amazing," Corderman explains. "One of the challenges of going on-premises will be finding any [on-premises object storage] vendor that can support that requirement." There will be no shortage of options, given that Vertica 12 integrates with six on-premises object storage vendors.

Pulling data from object storage obviously does introduce some latency, but Corderman says having data available locally in Vertica's Eon Mode Depot feature "provides much faster query response time" when and where needed.

The Trade Desk also takes advantage of Projections, a Vertica data optimization feature used for ordering, segmenting, and encoding data without having to build separate pipelines or make copies of data. "We use Aggregate Projection, which supports group-by functions, and Partition Range Projection, which is relatively new," Corderman says. "Our reports are 90% focused on data from the last 30 days, and Partition Range Projection provides a specific data-optimization strategy that's very helpful for that scenario."

Not many companies operate at the scale of The Trade Desk. Nonetheless, Corderman's experience and best advice on database product selection applies to companies of any size.

"Do a proof-of-concept project, and let the vendor help you, but treat it seriously," he advises, noting that The Trade Desk definitely put Eon Mode through its paces before the company migrated from Vertica Enterprise. "Include measurable goals, and be clear with the vendor on your success criteria. The more the vendor sees that you are serious about testing their product, the more they will support you."

ThinkData Works Embeds Vertica Into Its SaaS Offering

More than 30% of Vertica's 2,000-plus customers are independent software vendors (ISVs) and SaaS providers that embed the database into their own software and services. One such customer is ThinkData Works, a Toronto-based SaaS and managed services provider that helps financial institutions and other large firms optimize the use of data with a catalog platform.

Customers turn to ThinkData Works when they need a data catalog solution that strengthens access and governance in equal measure—especially when datasets from multiple sources are spread across large organizations. Banks and insurance companies, for example, take data from hundreds of vendors, government portals, and public sources and then layer that external data on top of their own internal intelligence. They often end up with thousands of datasets, and it's hard to track and efficiently use this information.

ThinkData Works' platform, which can be built on top of Vertica, is used to create data catalogs of each customer's internal and external data sources, often uncovering instances where companies have purchased the same data multiple times across different departments. One large bank, for example, discovered that it was buying the same data through six different departments. ThinkData Works' platform helped the bank immediately save more than $1 million annually via more efficient and effective purchasing of this single data asset.

Even when companies don't have data redundancy problems, the service makes it "much easier to connect to and query data assets," says Brenden Stennett, cofounder and CTO of ThinkData Works. The platform also provides rich data governance capabilities such as a purpose-built ETL tool, data profiling and audit logging functionality, and role-based access control down to the attribute level.

ThinkData Works partnered with Vertica in 2016 for its extensibility. For most ThinkData Works customers, bringing internal and external data together is a priority. "External datasets aren't usually at the vast scale that Vertica customers typically require, but we still needed [Vertica's] instantaneous columnar query performance," Stennett explains.

Deployment flexibility was another important selection criterion. ThinkData Works runs its SaaS service in Google Cloud, but many of the company's large security-sensitive customers want the platform to be deployed as a service (managed by ThinkData Works) within their own data centers. Vertica offers original equipment manufacturer (OEM) licensing arrangements that enable companies such as ThinkData Works to support on-premises deployments.

A final reason ThinkData Works selected Vertica was Vertica's geospatial analysis capabilities. External data sources often include geospatial information. Insurance companies, for example, often use publicly available data on fire hydrant locations, correlating it with home and business addresses to help with risk analysis and underwriting decisions. Geospatial analysis capabilities are now more prevalent, but back in 2016, when ThinkData Works made its database selection, Vertica was one of the few products that could support geospatial analysis.

Summing up why Vertica was selected and remains a preferred vendor and partner for ThinkData Works, Stennett says query performance, deployment flexibility, and geospatial querying were all important but that one other important factor put the icing on the cake.

"We like that it can be an embedded solution that can be packaged with the rest of our technology," Stennett explains. "Cloud-based database vendors don't necessarily give you that option, and we see it as a unique offering for Vertica."

PRICING

Vertica offers a choice of purchasing approaches with either perpetual or subscription-based licensing and per-node or per-terabyte terms. Subscriptions are time-based, starting with a by-the-hour option on AWS and Google Cloud and extending to monthly, one-year, two-year, or three-year options, with discounts for longer terms. Vertica Accelerator subscriptions are time-based per node.

Per-terabyte terms give customers flexibility to scale up and scale down compute nodes without incurring additional Vertica costs. This approach is best for organizations with hot workloads demanding high concurrency and compute elasticity. Per-node licensing is better suited to high-data

volume use cases such as IoT workloads. Customers specify the number of nodes best suited to handle their general workload requirements.

Vertica licenses, whether perpetual or subscription, can be deployed where and when they are needed and split among on-premises and multicloud options. It's the same license and same software, regardless of deployment mode.

ANALYSIS AND OBSERVATIONS

Vertica is a leading choice for high-scale analytics, starting with a single terabyte and scaling up to serve cloud-scale companies with the largest and most demanding requirements on the planet. Companies that now have or that expect to grow into such high-scale, mission-critical requirements should consider the following Vertica strengths and weaknesses (summarized in Figure 4, on page 22):

Strengths

- Vertica has many petabyte-league customers that rely on the database's MPP and columnar architecture for massive scalability and data compression and workload management for fast query performance.

- Vertica in Eon Mode on-premises, cloud, and hybrid deployments and the Vertica Accelerator cloud service separate compute and storage, eliminating copies of data; reducing ETL and data-movement requirements; and adding options for cost-saving elastic scaling, workload isolation, and dynamic workload management.

- Analytical capabilities span advanced SQL functions; multidimensional analysis; and model training, evaluation, and scoring via in-database predictive and ML functions. Analyses include outlier detection, linear and logistic regression, k-means, Naive Bayes, and random forest analysis. Data exploration and preparation functions such as distribution, gap filling, imbalance data handing, and filtering fill out the data science functionality. Vertica provides more than 650 prebuilt SQL extensions for in-database analytics and data science.

Weaknesses

- Optimization and recommendation features (including the Vertica Database Designer, which automatically generates an optimized database design) assist DBAs, but this automation is focused on optimizing known workloads according to user-defined rules and cost and performance thresholds. Automated deployment sizing, scaling, and system and workload management are on the roadmap.

- Customer-managed Vertica-in-Eon-Mode deployments rely on third-party storage. Storage and caching plans must be carefully set to ensure performance and compliance with demanding service-level requirements. The Vertica-managed Accelerator service handles these challenges behind the scenes with autoscaling, autoscheduling, and autoshutdown options. (Accelerator scaling, which can be automated or scheduled with user-defined rules, is currently capped at 48 nodes).

- Vertica offers Docker images on Docker Hub, but it's up to customers to manage and monitor container-based deployments independently. Deployment, monitoring, and management tools spanning hybrid and multicloud deployments are on the roadmap.

Figure 4. Vendor Strengths and Weaknesses

STRENGTHS	WEAKNESSES
• Core MPP and columnar architecture provides foundation for demanding, high-scale analytical requirements. • Vertica in Eon Mode separates compute and storage, bringing cost-saving elastic scaling with an extensive list of cloud, on-premises, and hybrid deployment options. • Extensive analytical capabilities span 650+ advanced SQL functions; built-in ML algorithms; and use of Python, TensorFlow, and other modeling environments. • Product has been adapted to unify myriad sources and data types, including columnar formats common to object-store-based data lakes.	• Automated optimization and system administrative capabilities are a work in progress. • Vertica in Eon Mode demands careful tuning of the caching layer to meet interactive query requirements. • Vendor-certified containerization options are broader and improved since 2019, but it's up to customers to find and master a unified container management option. • Micro Focus data management portfolio is limited, so Vertica partnerships are crucial.

Source: Constellation Research

RECOMMENDATIONS

Vertica should be considered by organizations dealing with (or expecting to grow into) high-scale deployments ranging into the hundreds of terabytes and beyond and preferring to tightly control the management and optimization of their workloads rather than accepting automated configurations. The Vertica Accelerator service eases the burden of administration with autoscaling and autoshutdown options while still providing control over scaling costs.

Vertica supports on-premises, multicloud, and hybrid deployments and works in combination with object storage and Hadoop-based data lakes both on-premises and in the cloud. Deployment flexibility has earned Vertica a place on the Constellation Research ShortList™ for Hybrid-Cloud and Multicloud Analytical Relational Database Management Systems since 2020.

Vertica addresses diverse data science requirements with more than 650 advanced analytical function; 15 built-in ML models; and PMML import and export capabilities supporting in-database execution of models developed in R, Java, Python, and myriad frameworks such as TensorFlow.

Customer examples in addition to The Trade Desk and ThinkData Works—the reference customers profiled in this report—include Optimal Plus, part of National Instruments, which taps 90 terabytes of data in HDFS with Vertica while also analyzing 1.04 petabytes within the database for IoT analyses such as predictive maintenance. MassMutual, another large customer, runs 360,000 queries per hour against a 65-terabyte data warehouse used for customer behavior analysis.

Few, if any, organizations start from scratch with petabyte-scale requirements; high-scale workloads emerge over time. That's why it's important for organizations to plan for the future, even if they're dealing with only a handful of terabytes today. Starting with the wrong database inevitably leads to a time-consuming and potentially painful rewrite of queries, ETL jobs, and more, not to mention the database deployment, configuration, and testing challenges that will be required, whether on-premises or in the cloud.

Based on conversations with dozens of organizations that have ventured into high-scale analytics, Constellation offers the following cautions and suggested best practices:

- **Think big and long-term when assessing deployment requirements.** It's all too common for organizations to outgrow deployments within just two to three years, either through unanticipated organic growth or through business-changing acquisitions. We've also seen on-premises-only policies and commitments quickly give way to cloud adoption. New CEOs and CIOs are famous for resetting agendas. Don't ignore history, but look beyond it to consider future possibilities and plan deployments that will stand the test of time and emerging requirements.

- **Look for deployment consistency and flexibility.** Does the analytical platform you are considering support on-premises deployment as well as cloud, multicloud, and hybrid deployments? Are all deployment modes possible with the same software and the same licenses, so you can still leverage training and financial investments even if

you need to change your deployment choices? Is there flexibility to mix and change deployment modes? In Vertica's case, the answer to all these questions is yes.

- **Be prepared for differences in on-premises and cloud performance.** Don't base cloud configurations and performance expectations on your on-premises experience. Plan for higher capacities to overcome the bandwidth, virtualization, and latency differences that are inevitable when deploying on any public cloud. Consider the guidance available from the vendor, including documentation, best practices, and the level of activity and topics discussed on customer forums and community pages.

- **Consider available skills and training resources.** Database professionals are made, not born, so consider the available talent pool and training resources tied to the products you are considering. There are plenty of SQL-savvy data management professionals out there, but how many have experience deploying and managing the solution you are considering? Most vendors offer training (as does Vertica with its no-cost **Vertica Academy**), but users should determine whether they can draw on advice and best practices from a sizable and active customer community.

Doug Henschen

Vice President and Principal Analyst

Doug Henschen is a vice president and principal analyst at Constellation Research Inc. focusing on data-driven decision-making. His Data to Decisions research examines how organizations employ data analysis to reimagine their business models and gain a deeper understanding of their customers. Data insights also figure into tech optimization and innovation in human-to-machine and machine-to-machine business processes in manufacturing, retailing, and services industries.

Henschen's research acknowledges the fact that innovative applications of data analysis require a multidisciplinary approach, starting with information and orchestration technologies; continuing through business intelligence, data visualization, and analytics; and moving into NoSQL and big data analysis, third-party data enrichment, and decision-management technologies. Insight-driven business models and innovations are of interest to the entire C-suite.

Previously, Henschen led analytics, big data, business intelligence, optimization, and smart applications research and news coverage at *InformationWeek*. His experiences include leadership in analytics, business intelligence, database, data warehousing, and decision-support research and analysis for *Intelligent Enterprise*. Further, Henschen led business process management and enterprise content management research and analysis at *Transform* magazine. At *DM News*, he led the coverage of database marketing and digital marketing trends and news.

@DHenschen **constellationr.com/users/doug-henschen** **linkedin.com/in/doughenschen**

ABOUT CONSTELLATION RESEARCH

Constellation Research is an award-winning, Silicon Valley–based research and advisory firm that helps organizations navigate the challenges of digital disruption through business model transformation and the judicious application of disruptive technologies. Unlike the legacy analyst firms, Constellation Research is disrupting how research is accessed, what topics are covered, and how clients can partner with a research firm to achieve success. Over 350 clients have joined from an ecosystem of buyers, partners, solution providers, C-suite, boards of directors, and vendor clients. Our mission is to identify, validate, and share insights with our clients.

Organizational Highlights

- Named Institute of Industry Analyst Relations (IIAR) New Analyst Firm of the Year in 2011 and #1 Independent Analyst Firm for 2014 and 2015.
- Experienced research team with an average of 25 years of practitioner, management, and industry experience.
- Organizers of the Constellation Connected Enterprise—an innovation summit and best practices knowledge-sharing retreat for business leaders.
- Founders of Constellation Executive Network, a membership organization for digital leaders seeking to learn from market leaders and fast followers.

www.ConstellationR.com

@ConstellationR

info@ConstellationR.com

sales@ConstellationR.com

San Francisco Bay Area | Boston | Colorado Springs | Denver | Ft. Lauderdale | New York Metro
Northern Virginia | Portland | Pune | San Diego | Sydney | Washington, D.C.