

OFFERING OVERVIEW

# Vertica Solves Data Silo, Data Science & Hybrid- & Multicloud Challenges

Vertica in Eon Mode Taps Object Stores and Hadoop,  
Separates Compute and Storage Requirements



Doug Henschen  
Vice President and Principal Analyst

# TABLE OF CONTENTS

Executive Summary .....3

Company Overview .....4

About Vertica .....5

Functional Capabilities.....9

Use Cases ..... 14

Pricing ..... 20

Analysis and Observations ..... 20

Recommendations ..... 23

Analyst Bio. .... 25

About Constellation Research ..... 26



## EXECUTIVE SUMMARY

Vertica is a leading analytical data platform, and it's among the most important products in the portfolio of Micro Focus International, Plc., a multibillion-dollar, U.K.-based global software company. Vertica was founded in 2005 and introduced in 2007 as a columnar, scale-out analytical database capable of powering ultra-high-scale analytics. It has since evolved to be more than a database. Today, Vertica is a modern analytical data platform that can unify myriad data sources in warehouse- and data-lake-style approaches; span SQL, multidimensional and data science analytical requirements; and adapt to on-premises, multicloud and hybrid deployment needs with flexibility and agility.

This Offering Overview provides an overview of Vertica, which is one of two top research and development investment priorities for Micro Focus. This report identifies key differentiators, examines functional capabilities, considers Vertica's strengths and weaknesses, and provides use cases, including three customer case studies. Technology buyers should use this report to evaluate Vertica for implementation.

### Business Themes



Data to Decisions



Digital Marketing &  
Sales Effectiveness



Next-Generation  
Customer Experience



Technology  
Optimization

## COMPANY OVERVIEW

Micro Focus is a global software and software consultancy firm headquartered in Newbury, Berkshire, England. The company was founded in 1976, and its earliest success was in bringing the COBOL language, so closely associated with mainframes, to microcomputers. 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.

### Micro Focus International, Plc.

- **Company:** Micro Focus International, Plc
- **Headquarters:** Newbury, England
- **Founded:** 1976
- **Type:** Public (MCRO on London Stock Exchange and MFGP on New York Stock Exchange)
- **2019 Revenue:** \$3.3 billion
- **No. of Employees:** 12,000
- **Website:** [MicroFocus.com](https://www.microfocus.com)
- **Twitter:** [@MicroFocus](https://twitter.com/MicroFocus)

Today, Micro Focus has organized and focused its portfolio around four pillars of digital transformation:

- **Insights**, through analytical capabilities led by Vertica and focused on customer behavior, Internet of Things initiatives, operational insight, and cognitive search and knowledge discovery capabilities delivered by the IDOL platform.
- **Security**, risk and governance, through breach defense, application security, data privacy, data governance and regulatory compliance products, such as ArcSight and Interset.
- **Speed**, through enterprise DevOps release orchestration and automated delivery technologies, such as Micro Focus LoadRunner Cloud.
- **Agility**, through hybrid (in the cloud and on-premises) IT service management products, such as Micro Focus Operations Bridge.

In February 2020, Micro Focus identified the security and insights pillars as the biggest drivers of growth for the company. To underscore this priority, the company's CEO announced that Micro Focus will make incremental investments of \$70 million to \$80 million in Vertica and the security portfolio in 2020, with additional investments planned for 2021 and beyond.

This investment plan, which will be split approximately two-thirds toward research and development and one third toward go-to-market enhancements, underscores the importance of Vertica, the subject of this report, as one of the cornerstones (if not *the* crown jewel) of the Micro Focus portfolio.

## ABOUT VERTICA

### Overview

Vertica was first released in 2007, at the dawn of the big data era, co-founded by renowned database innovator Michael Stonebraker. The database was designed to manage vast data volumes, scaling into the petabytes, while harnessing massively parallel processing and columnar architecture to ensure lightning-fast query performance.

Today, Vertica is much more than a database. Vertica, which operates as an independent business unit of Micro Focus, describes it as a “Unified Analytics Warehouse” because the delivery of analytics matters more than the location of the data sources or the style of the deployment. Vertica can power focused data marts or broader data warehouses, and it can draw on Hadoop or modern object-store-based data lakes. It spans conventional business intelligence (BI) and analytical requirements, while also powering built-in and externally developed machine learning (ML) algorithms. Vertica can be deployed on-premises, on any public cloud (with quick-to-deploy marketplace options on the leading platforms) or in hybrid-cloud and multicloud combinations.

To meet all these needs, Vertica has steadily evolved and matured over the last 15 years. The product was first adapted to work in conjunction with high-scale, Hadoop-based data lake deployments, adding features to load data from Hadoop, work with the leading Hadoop file formats and directly query data in the Hadoop Distributed File System (HDFS).

Over the last five years, capabilities have been added to Vertica to better support data science workloads and cloud deployment. Data science innovations include geospatial, time series and ML analysis, flexible tables for schema-on-read flexibility, flattened tables supporting complex joins and more extensive in-database support for ML algorithms.

Cloud-oriented enhancements include ready-to-deploy images on leading cloud marketplaces and prebuilt integrations with native services on Amazon Web Services (AWS), Microsoft Azure (Azure) and Google Cloud Platform (GCP). Also added were connectors for Amazon S3 and other cloud and on-premises object storage options. Most significantly, a new deployment option called Eon Mode was introduced to separate compute and storage decisions, promising cost savings and elasticity in hybrid-cloud and multicloud deployments.

The Vertica 10 release in early 2020 introduced yet more Eon Mode deployment options along with deep integration with Python and TensorFlow and new model import/export capabilities based on the Predictive Model Markup Language (PMML).

In short, the Vertica team has steadily evolved the product to keep pace with the demands of organizations that seek to do the following:

- Unify diverse data sources and data types, at scales ranging from tens of terabytes to the double-digit petabyte league.
- Meet diverse analytical requirements spanning business intelligence, advanced analytics and machine learning.
- Unify deployments with flexibility and elasticity across hybrid environments, with options to run on bare metal and multiple public clouds.

## Market Segment

Vertica is a high-scale analytical data platform, a segment of the larger database market that has flourished and evolved over the last two decades. This segment was a tiny niche before exploding

in the 2000s as organizations increasingly grappled with rising volumes and varieties of data on one hand and expectations for ever-faster analysis on the other.

Data scale expectations quickly expanded in the 2000s as organizations looked beyond historical information held in internal transaction systems, such as enterprise resource planning and customer relationship management systems. Internet clickstreams, sensor data, log files, mobile data rich with geospatial information and social network data were among the many forms of information that pushed stores into the big data league above 10 terabytes. “Cloud-native” companies were also emerging, building entirely new data-driven business models by tapping into clickstreams and other customer behavior data at cloud-scale.

At the dawn of the big data era, the leading general-purpose relational database management systems (RDBMS) at the time—Oracle Database, Microsoft SQL Server and IBM Db2—were being used for data warehousing, but they had yet to be highly adapted for high-scale analytical use. As a result, data warehouses and even focused data marts could not easily or affordably scale up to crunch tens or hundreds of terabytes, let alone petabyte-scale data.

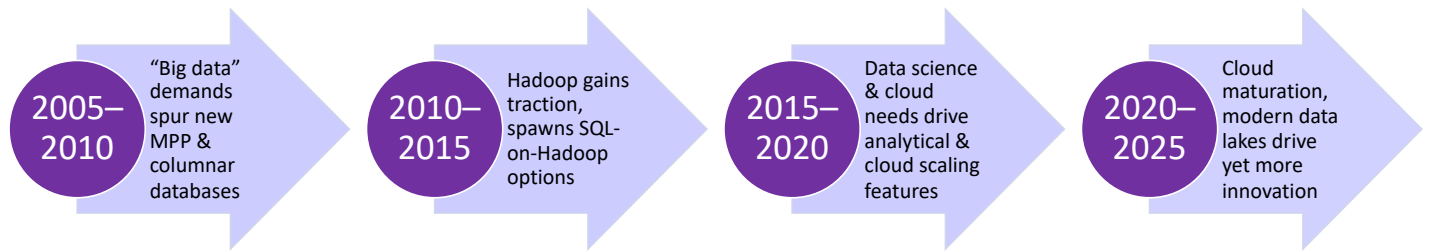
Early pioneers of the analytical database market harnessed massively parallel processing (used by Teradata) and column-store architectures optimized for analytics (used by Sybase IQ). Also known as scale-out architecture, massively parallel processing (MPP) distributes workloads across many nodes for scaling. MPP became the cornerstone of many high-scale, distributed data platforms, including high-scale analytical databases, Apache Hadoop, Apache Spark and many NoSQL databases.

Column-store databases have an advantage when querying selected columns of data, for example the ZIP codes, product SKU numbers or transaction dates that populate a database. Such a query could yield analytics on what sold in which regions within the last month without the need to wade through all the (irrelevant-to-the-query) data stored row by row in a traditional, transactional database—such as customer names, address, account number and so on. Less data queried translates into faster query results.

As shown in Figure 1, high-scale analytical requirements have evolved and continue to change. Growing demand for big data analysis brought a renaissance of new database products to the market by the mid-2000s. Vertica was among the innovators, tapping both MPP and columnar architecture.

## Figure 1. Vertica's Maturation Mirrors the Evolution of High-Scale Analytical Databases

### The Evolution of High-Scale Analytical Databases: 2005–2025



Source: Constellation Research

Starting in 2010 and gathering steam by 2012, organizations looking to exploit high-scale data were drawn by the promise of low-cost storage in Hadoop. Deployments multiplied, but the complexity of this new platform limited access to data. Analytical databases (including Vertica) soon adapted (and yet more products emerged), with new SQL-on-Hadoop options bringing the familiarity of SQL-based querying to data lakes.

Innovations in high-scale analytical databases continued over the last five years, with yet more features introduced for in-database data science capabilities (as exemplified by Vertica's built-in ML algorithms) and separation of compute and storage provisioning (pioneered by Snowflake and implemented in Vertica through Eon Mode in 2018). Market trends have favored cloud-ready software, but dedicated, hardware-based offerings are still available, including Oracle Exadata, various Teradata hardware-based offerings and the IBM Netezza Performance Server.

Throughout this evolutionary period, the market has seen ever-rising demands for scalability, query performance, user- and query-concurrency, and sophisticated analysis capabilities. While some products have fallen by the wayside, the market leaders have remained vibrant by helping organizations to embrace a range of data sources and storage options, to support a broad range of analytical needs extending into data science and to meet a range of evolving deployment requirements, from conventional on-premises to public clouds and a changing, hybrid-cloud and multicloud mix of choices.



## Partnerships and Alliances

Vertica has strong, long-term partnerships on three fronts: technology partners, original equipment manufacturers (OEMs) and systems integrators. Following is a summary of these partnership types along with leading examples.

- **Technology partners:** Software and service providers across many categories integrate with Vertica. Public cloud provider AWS, for example, is a long-term partner, and the AWS Marketplace offers both Vertica by the Hour, a usage-based offering, and ready-to-deploy Vertica BYOL (bring your own license) instances available on Amazon Linux or RedHat. (Vertica by the Hour was made available on GCP in August, and BYOL options are also available on GCP and Azure.)

Other important technology partners include data integration vendors (including Informatica, IBM, SAP, Talend and TIBCO) and BI and analytics vendors (including IBM Cognos, Looker, MicroStrategy, Qlik, SAS and Tableau).

- **OEM partners:** Dozens of OEM partners use Vertica as the analytical database underpinning their software or services. Examples include CA Technologies, CallidusCloud, Cisco, FICO, NetScout, NICE and Vendavo.
- **Systems integrator partners:** These organizations help Vertica customers with systems design and deployment consulting, offering specific knowledge of source systems, Vertica performance and configuration. Leading examples include Deloitte, Hewlett Packard Enterprise, Tata Consultancy Services and Tech Mahindra.

## FUNCTIONAL CAPABILITIES

Vertica's calling cards of vast scalability and superior query performance were designed-in at the product's inception, thanks to the combination of MPP scalability and column-store query-speed and compression advantages. Nonetheless, the database has been extended over the last 15 years to become a versatile analytics platform spanning data sources, analytical requirements, storage choices and deployment choices.

Detailed below are key categories of functionality that help organizations to unify data silos, span advanced analytical and data science needs, and meet hybrid- and multi-cloud deployment requirements. At the conclusion of each section are the latest upgrades introduced in early 2020 with the release of Vertica 10.

## Unify Data Silos

Vertica handles the analysis of conventional structured sources as well as sparse and variable data types common to Hadoop and 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 Hadoop and emerging object-storage-based data lake platforms.
- Parquet export for writing result sets back to external data lake platforms.
- Role- and column-based security for secure integration with source systems.
- Vertica 10 upgrades include:
  - Support for querying complex data types, such as Maps, Arrays and Structs stored in Parquet and common to S3, HDFS and Google cloud storage.
  - Security advances, which include improved LDAP link for user authentication, simplified Transport Layer Security (TLS) certificate administration and Kerberos support for the open source Vertica Python interface.

## Unify Analytics & 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 business intelligence, advanced analytics and data science requirements. Capabilities include:

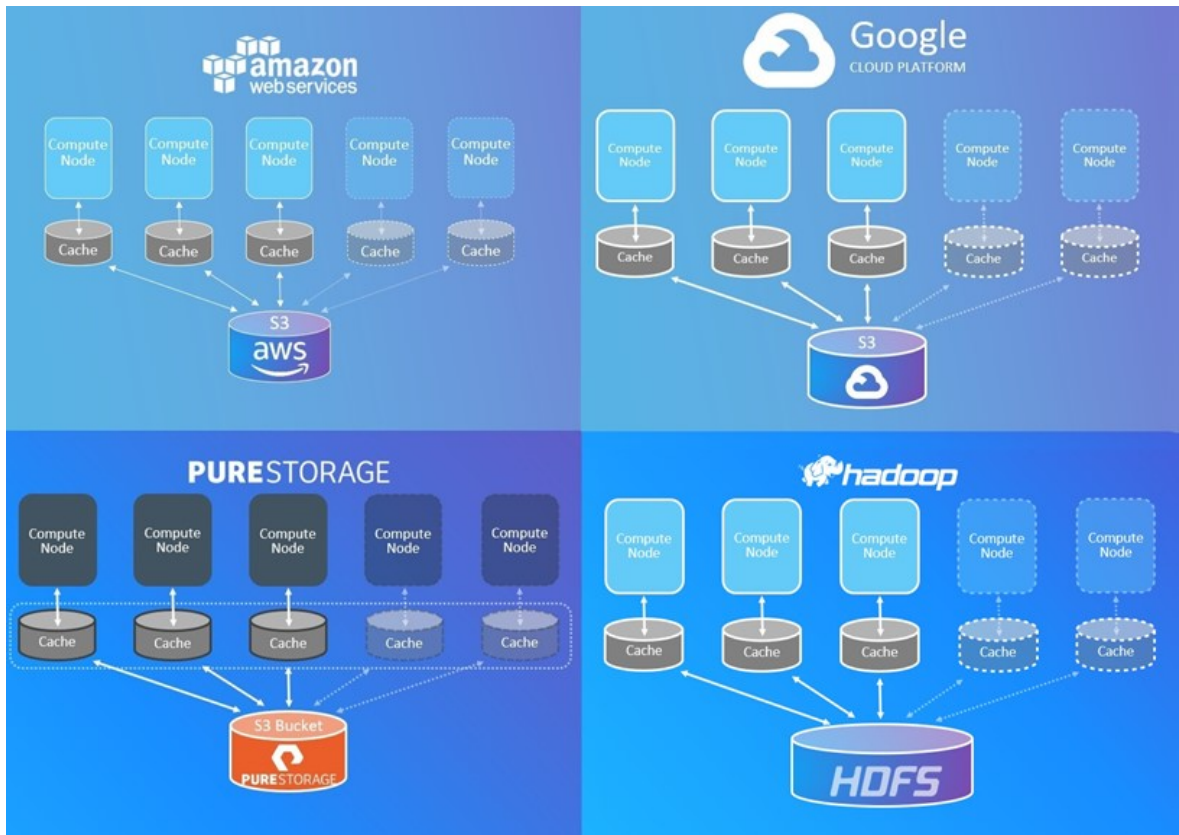
- Advanced SQL functions including pattern matching, time series and geospatial analysis.
- Relational online analytical processing (ROLAP) SQL functions supporting multidimensional analysis include rollup aggregations, grouping sets aggregations, cube aggregations and pivot.
- Predictive analytics and ML functions include outlier detection, linear and logistic regression, k-means, Naïve Bayes, random forest and confusion matrix.
- Vertica UDX (user-defined extensions) framework enables parsers and data loaders written in C++, Java, R and Python to be run in-database as Vertica SQL functions.
- Open source Python client and Python SDK for Vertica support the use of Jupyter notebooks and in-Vertica processing of user-defined functions expressed in Python.
- Vertica 10 upgrades include:
  - PMML model export support for training of models on Vertica using parallel, in-database algorithms and subsequent export of finished models to external systems for scoring.
  - PMML import supports training of models on a choice of external frameworks and languages (such as Apache Spark, R and Python). Models exported from these tools in PMML format can be imported to Vertica, where they can be managed as easily as tables and run in production on very large datasets.
  - TensorFlow model import supports training of custom ML models (including deep neural network models) on a choice of compute platforms (including high-performance compute or GPU-based systems, as required). Models saved in TensorFlow's native graph format can be imported to Vertica and used for distributed prediction on very large datasets.

## Unify Hybrid-Cloud and Multicloud Deployments

Core management and administrative features and the Eon Mode option introduced in 2018 simplify the deployment and use of Vertica in on-premises deployments, on multiple public clouds or in a hybrid mix of these options. Relevant Vertica features include the following:

- Database Designer automatically creates projections and optimizes compression and sorting based on query patterns and performance.
- Storage-agnostic design enables different tables or partitions of a single table to be stored in different locations. Recent data, for example, could be stored on local disk, while older partitions can be placed on storage area networks (SANs), HDFS, cloud-based object stores or a mix of options.
- Vertica in Eon Mode options, shown in Figure 2, separate compute and storage, supporting shared storage for all workloads and introducing independent, elastic scalability and cost savings for compute. The software:
  - 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.
  - Allows compute hardware choices and configurations that can be matched to workloads, an advantage over appliances with limited options.
  - Isolates workloads through a subcluster design that provides dedicated compute capacity and surer compliance with service-level demands.
  - Eliminates redundant copies of data and reduces ETL requirement by enabling multiple subclusters to read a common database.

Figure 2. Vertica in Eon Mode Is Available for AWS, GCP, Pure Storage and Hadoop; Additional Cloud and On-Premises Options Are Planned



Source: Vertica

- Vertica 10 upgrades include:
  - Management Console workload analyzer spots query performance problems and automatically runs recommended fixes.
  - Improved query performance on virtual machines.
  - Redesigned Database Designer slashes resource requirements while improving automated projection design and optimization.
  - New Vertica in Eon Mode options introduced for Google Cloud Platform and, on-premises, HDFS and MinIO. Additional planned options include the Microsoft Azure and Alibaba clouds and Network File System storage.

- Fast, wizard-driven configuration introduced for Eon Mode on AWS and GCP.
- Migration feature lets organizations back up Vertica Enterprise databases and relaunch them in Eon Mode.

## USE CASES

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

- Communication and network analytics
- Customer behavior analytics
- Data warehouse modernization
- Embedded analytics
- Fraud monitoring and risk management
- Internet of Things (IoT) analytics
- Time series analytics

Customers can draw on a rich trove of 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 including IBM Cognos, Informatica, LogiAnalytics, MicroStrategy, Pentaho, Qlik, SAP and Tableau.

Vertica is often used to address the broader challenges of unifying data silos, unifying analytics and data science, and unifying hybrid- and multicloud deployments—challenges that might cut across multiple use cases. Following are three short Vertica customer case studies in which these challenges were successfully met.

### Taboola Unifies Data Silos

Taboola is a content discovery platform that recommends more than 450 billion articles, blogs, videos, products and apps to more than 1.5 billion unique users every month across thousands of websites and mobile networks. Taboola turned to Vertica in 2014 because the conventional relational database

it was using for reporting and analysis wasn't keeping up with the company's rapid growth. Taboola was trying to load as much as 30 terabytes of user behavior data a day, but the database had slowed to a crawl and the deployment could no longer scale.

Taboola needs reporting and analytics to understand and explain campaign performance and interactions with its recommendations. It answers questions such as, what did customers see, what did they click on, and did they convert? The analytics platform analyzes user session data, slicing and dicing by country, platform and other dimensions to help publishers and advertisers understand the clicks and campaign costs and benefits at a detailed level.

The company's Vertica deployment immediately addressed the scale and performance challenges, according to Keren Bartal, director of data engineering at Taboola. The deployment has since expanded beyond loading of clickstream data from HDFS, adding yet more data sources and use cases while continuing to scale up in-depth analytics.

"In the last year, we've turned Vertica into our data lake," Bartal explains, citing Salesforce and Netsuite among a list of new data sources extending into marketing, human resources and finance. "It's much easier for our analysts to have all of our information in a single place, and that makes Vertica more powerful for us."

*"With Vertica in Eon Mode, we plan to centralize all of our data on HDFS and then spin up subclusters [of Vertica] as needed. We won't have to replicate data, and it will cut our operational cost because we can have as many subclusters as we want working on the same data."*

—Keren Bartal, Taboola

Taboola's analysts use Vertica in combination with Tableau, and Bartal says the combination is typically more powerful than the analytical capabilities built into commercial applications. In other cases, analysts were using desktop tools "and that just didn't work as well," says Bartal.

Taboola's core use case remains in-depth analytics at scale, and on that front Vertica now analyzes upward of 300 terabytes of data per day across a total store that exceeds 1 petabyte. The company

is in the process of upgrading to Vertica 10, and Bartal says the company is looking forward to the efficiencies promised by Vertica in Eon Mode.

“We’ve pushed hard for Vertica in Eon Mode because there are times when we need to store more data but we don’t need more compute and vice versa,” says Bartal. “With Eon, we plan to centralize all of our data on HDFS and then spin up subclusters [of Vertica] as needed. We won’t have to replicate data, and it will cut our operational cost because we can have as many subclusters as we want working on the same data.”

## **Bidtellect Unifies Analytics and Data Science**

Bidtellect is a content delivery platform that works with website publishers and advertisers. It uses open auctions and real-time bidding to deliver advertisements in real time. The company relies on a high-scale analytical database to understand and give customers insight into bid results, ad delivery, impressions, conversions, costs and more.

In early 2017, the company was struggling with scalability and performance problems on its analytics platform, which at the time was powered by Greenplum running on-premises. Poor response times and difficulty supporting ad hoc analysis and real-time reporting contributed to high customer turnover. A four-person operations team couldn’t keep up with customer questions.

“We optimized queries as best we could, but the high read and high write demands overwhelmed the database,” says Mike Conway, Bidtellect’s CTO.

After a competitive assessment that included continued use of Greenplum, Amazon Redshift and Vertica, Bidtellect chose Vertica in December 2017. An on-premises deployment of Vertica was completed by March 2018. (Conway says Bidtellect could not afford a public-cloud-based approach at the time because it wasn’t big enough to command significant discounts.)

The Vertica deployment not only eliminated Bidtellect’s analytical performance problems, it quickly scaled up to match the company’s resurgent growth. The company’s revenue has increased at a 30%-per-year pace since mid-2018, increasing analysis requirements from 2 billion auction



transactions per week to 55 billion transactions per week over the same period. Conway says Vertica has helped drive that growth, as performance improvements have “changed the entire user experience.”

“Previously, customers couldn’t make decisions on their campaigns fast enough,” Conway explains. “Once we had Vertica, we could run our price-estimation algorithms by the minute [versus by the hour], so the accuracy of estimation and our optimization process has improved dramatically.” Customer retention has increased from 50% before the deployment to 90% today.

*“Once we installed Vertica, we could run our price-estimation algorithms by the minute [versus by the hour], so the accuracy of estimation and our optimization process has improved dramatically.”*

—Mike Conway, Bidtellect

Bidtellect handles advanced analytics and data science both within Vertica and on top of HDFS using Spark. At this writing, a handful of predictive models written in R are run on Vertica, but the company’s long-term direction is to do more on the platform. In fact, Conway is hoping to do away with HDFS entirely using Vertica in Eon Mode.

“If I can separate compute from storage, it solves about 80% of my problems,” Conway explains. “The amount of information that we’re processing continues to grow, but today it’s a shared system with production analytics, so it’s a scalability challenge.”

Despite Bidtellect’s growth, its Vertica deployment has held steady at five nodes. To handle the growth in data volume, Bidtellect has cut data retention from five years down to two years of historical data. With Vertica in Eon Mode, Bidtellect would have a single shared storage environment and could scale up and scale down compute as needed to meet the fourth-quarter peaks and first-quarter valleys experienced in the advertising business, Conway explains.

In the short term, Conway says Bidtellect is considering the Vertica in Eon Mode option with MinIO (on-premises object storage). Longer term—at least two to three years out—Conway says Vertica in Eon Mode public cloud options would be considered, as continued growth is expected to put the company in a position to command steeper cloud discounts.

“For now, I’m trying to get rid of HDFS, and my long-term goal is to move all of our analytics and ML into Vertica,” Conway concludes.

## Cerner Unifies Hybrid and Cloud Deployment

Cerner is a health informatics software and services supplier that operates around the globe. The company has been using Vertica for seven years, for two petabyte-scale analytical applications supporting the company’s Millennium Electronic Health Records (EHR) application.

One system, the Response-Time Measurement System (RTMS), tracks the uptime, availability and latency of Cerner’s hosted Millennium EHR system for all customers, in exhaustive detail and in real time. Roughly 70% of Cerner customers use the company’s hosted, Citrix-virtualized service—think of it as a private cloud—while the rest run the Millennium software on-premises. RTMS monitoring and ad hoc analysis enables Cerner’s hosting organization to proactively maintain system performance and back-end help-desk support when system performance problems arise.

A second system, the Lights On Network, was added by Cerner to help customers comply with U.S. federal Meaningful Use requirements. This free, add-on service gives customers that join the network measurements on whether and how quickly they are completing tasks required for compliance. The network also gives customers insight into the performance within their domain, how their clinicians are performing, how much time they’re spending on the system and how many orders users are executing.

*“There are customers all around the world that want deployments in their regions. As soon as they get wind that it’s possible [with Vertica in Eon Mode], they’re going to demand that their data also be moved over to AWS.”*

—Dan Woicke, Cerner

The scale of Cerner’s systems built on Vertica is astounding. The company has distributed more than 90,000 software agents across servers, switches and clients around the globe. These agents are monitored at different latency levels, depending on their criticality. RTMS timers on Citrix servers running Millennium instances for Cerner customers,

for example, are polled every minute. Other, less-sensitive measures are updated hourly. All these operational measures are initially stored on Hadoop, with batch loading to Vertica every five minutes.

All told, 1 billion rows of new operational data are committed to Vertica each hour, and the total footprint exceeds 2 petabytes of compressed data. The two separate Vertica-based systems are deployed on-premises in two separate data centers near Cerner's headquarters in Kansas City, Missouri, but the deployments are about to go hybrid.

Cerner has many moving parts, and most of them are headed into the cloud. Certain customers have data residency requirements, for example, so Cerner is moving to support those needs with hybrid options.

One country subject to GDPR and data-sovereignty requirements, for example, is moving its instance of Millennium into its own data center, but it still wants to take advantage of the operational insights of the Cerner RTMS system. To meet this need, operational measures from the on-premises system will securely go to a separate, dedicated instance of Vertica running in Eon Mode on an AWS Availability Zone in Europe. Similarly, one U.S.-based customer with sensitive needs also wants its own instance of Millennium, so a separate RTMS system instance is planned on a separate instance of Vertica in Eon Mode running in the organization's environment on AWS.

"There are customers all around the world that want deployments in their regions," says Dan Woicke, senior director of CernerWorks Engineering, the hosting organization at Cerner. "As soon as they get wind that it's possible [with Vertica in Eon Mode], they're going to demand that their data also be moved over to AWS."

In the short term, Woicke says Cerner plans to start using Vertica in Eon Mode to spin up separate instances of the RTMS and Lights On Network platforms in any AWS region. Over the longer term, Cerner is rewriting the Millennium application to run software-as-a-service style on AWS (rather than as a hosted service in its Kansas City data centers). Once that happens (one to two years from now), Woicke says Cerner will have decisions to make regarding where to run Vertica.

For now, Cerner has a big investment in on-premises infrastructure that's "performing well," Woicke says, but over the long term, proximity to systems running on AWS and the cost advantages of

separating compute and storage could lead to a massive Vertica in Eon Mode deployment on AWS. Either way, Cerner has hybrid- and multicloud options with 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 extending to monthly, one year, two years or three years, with discounts for longer terms.

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 as 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 in the tens of terabytes and scaling up to serve cloud-scale companies with the largest and most demanding requirements on the planet. In addition to Taboola, Bidtellect and Cerner, other high-scale Vertica customers include Intuit, Uber and TheTradeDesk. Intuit uses Vertica to personalize customer interactions with its TurboTax service. Uber uses Vertica for surge pricing analysis on a platform that supports as many as 14 million rides per day. TheTradeDesk is an ad auction platform that has two separate 7-petabyte, 320-node Vertica instances, with deployments on both AWS and on-premises and plans to run on Azure.

Keep in mind that 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. As Taboola and Bidtellect discovered, 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.

## Strengths

Detailed below is Constellation's analysis of the product strengths that would-be Vertica customers should consider (see Figure 3, p. 23).

- **Core architecture is ideal for high-scale analytics.** Vertica was designed expressly for high-scale analytical requirements, exploiting both massively parallel processing, for petabyte-league scalability, and columnar architecture, for query-performance and data-compression advantages. Non-MPP and row-store-based competitors have bolt-on features intended to match these traits, but they tend to introduce complexity and data redundancy.
- **Vertica in Eon Mode offers cost, performance and deployment advantages.** Introduced in 2018, Vertica in Eon Mode met the emerging trend and customer demand to separate compute and storage requirements. The design eliminates copies of data, reduces ETL and data-movement requirements, and lets customers take advantage of cost-saving elastic scaling, workload isolation, hardware customization and dynamic workload management options. Vertica in Eon Mode is also opening up a growing list of on-premises, cloud, hybrid-cloud and multicloud deployment options.
- **Analytical breadth and depth are industry leading.** Extensive analytical capabilities span advanced SQL functions, multidimensional analysis and built-in predictive and ML functions that include outlier detection, linear and logistic regression, k-means, Naïve Bayes and random forest analysis. User-defined functions enable algorithms written in C++, Java, R and Python to be run in-database as Vertica SQL functions, and PMML import and export extend data science capabilities to popular frameworks, including TensorFlow. Vertica has codified hundreds of prebuilt, reusable user-defined extensions.

- **Unifies diverse data sources and data types.** Over 15 years, Vertica has been adapted to connect with myriad data sources and modern data formats, including Parquet and ORC, used in HDFS, Spark and object-store-based data lakes. It also analyzes complex data types, such as maps and arrays, not usually supported in SQL. In short, Vertica is a market leader in its ability support analysis of diverse data sources and types.

## Weaknesses

In some areas, Vertica is responding to, rather than leading, the development of cutting-edge functionality. Here's Constellation's analysis of Vertica weaknesses:

- **Automation features are a work in progress.** Vertica has strong optimization and recommendation features to assist DBAs, but automation is focused on optimizing to preset workload priorities. The Vertica team is working on capabilities to automatically adapt to dynamic workload requirements and to automate more routine maintenance tasks.
- **Vertica in Eon Mode demands careful tuning.** Vertica in Eon Mode separates storage and compute decisions, but in doing so it relies on third-party storage. Storage options and caching plans must be carefully set, particularly in cloud deployments, to ensure performance and compliance with demanding service-level requirements.
- **Containerization support is a work in progress.** Pioneering customers are leading the way on deploying Vertica in containerized deployments. Vertica is catching up on testing, certification and publication of best practices for containerized deployment.
- **Micro Focus data management portfolio is limited.** Micro Focus is a large, global company, but it's best known for its mainframe and IT service management roots. Vertica, acquired in 2017, is the highest-profile data management product in the Micro Focus portfolio. Micro Focus does not offer data integration, data catalog, data quality/prep or master data management offerings. That makes Vertica partnerships, such as those with best-of-breed data integration vendors, important for creating solutions.

Figure 3. Vertica's Strengths and Weaknesses

STRENGTHS	WEAKNESSES
<ul style="list-style-type: none"><li>• Core MPP and columnar architecture provides foundation for demanding, high-scale analytical requirements.</li><li>• Vertica on Eon Mode separates compute and storage, bringing cost-saving elastic scaling in a growing list of cloud and on-premises deployment options.</li><li>• Extensive analytical capabilities span advanced SQL functions, ROLAP, built-in predictive algorithms and use of Python, TensorFlow and other modeling environments.</li><li>• Mature product has been adapted to unify myriad sources and data types, including modern data lake formats.</li></ul>	<ul style="list-style-type: none"><li>• Automated optimization capabilities are a work in progress.</li><li>• Vertica in Eon Mode demands careful tuning of the caching layer to meet interactive query requirements.</li><li>• Vendor-certified containerization standards and best practices are works in progress.</li><li>• Micro Focus data management portfolio is limited, so Vertica partnerships are crucial.</li></ul>

Source: Constellation Research

## RECOMMENDATIONS

Any organization dealing with, or expecting to grow into, high-scale analytical workloads ranging from the tens of terabytes into the petabytes should consider Vertica. With its Vertica in Eon Mode options, Vertica is particularly well suited to spanning hybrid-cloud and multicloud deployments and unifying analysis of high-scale data sources, including legacy Hadoop clusters and modern, object-store-based data lakes.

Vertica also addresses diverse analytical and data science requirements spanning advanced SQL and multidimensional analysis as well as ML and predictive analytics through built-in algorithms. In addition, Vertica offers extended support for operationalizing custom algorithms through user-defined functions and PMML capabilities supporting in-database execution of models developed in R, Java, Python and myriad frameworks, including TensorFlow.

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 CIO 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 using the same software and the same licenses, so you can leverage training and financial investments? 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 the available 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.



## ANALYST BIO

# Doug Henschen

Vice President and Principal Analyst

Doug Henschen is 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 multi-disciplinary 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](https://twitter.com/DHenschen)  [constellationr.com/users/doug-henschen](https://constellationr.com/users/doug-henschen)  [linkedin.com/in/doughenschen](https://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 models 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](http://www.ConstellationR.com)



[@ConstellationR](https://twitter.com/ConstellationR)



[info@ConstellationR.com](mailto:info@ConstellationR.com)



[sales@ConstellationR.com](mailto:sales@ConstellationR.com)

---

Unauthorized reproduction or distribution in whole or in part in any form, including photocopying, faxing, image scanning, e-mailing, digitization, or making available for electronic downloading is prohibited without written permission from Constellation Research, Inc. Prior to photocopying, scanning, and digitizing items for internal or personal use, please contact Constellation Research, Inc. All trade names, trademarks, or registered trademarks are trade names, trademarks, or registered trademarks of their respective owners.

Information contained in this publication has been compiled from sources believed to be reliable, but the accuracy of this information is not guaranteed. Constellation Research, Inc. disclaims all warranties and conditions with regard to the content, express or implied, including warranties of merchantability and fitness for a particular purpose, nor assumes any legal liability for the accuracy, completeness, or usefulness of any information contained herein. Any reference to a commercial product, process, or service does not imply or constitute an endorsement of the same by Constellation Research, Inc.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold or distributed with the understanding that Constellation Research, Inc. is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional person should be sought. Constellation Research, Inc. assumes no liability for how this information is used or applied nor makes any express warranties on outcomes. (Modified from the Declaration of Principles jointly adopted by the American Bar Association and a Committee of Publishers and Associations.)

Your trust is important to us, and as such, we believe in being open and transparent about our financial relationships. With our clients' permission, we publish their names on our website.

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