Vertica Overview

The Vertica Analytics Platform delivers speed, scalability, and built-in machine learning that today's most analytically intensive workloads demand, whether in the public clouds, on-premises, on Hadoop, or any hybrid combination.

Key Benefits

Step Up to the Fastest, Most Flexible Big Data Analytics Platform

What should you look for in a data analytics warehouse to address today and tomorrow's data challenges? Consider the following Vertica capabilities:

- **Unify your analytics, not the data:** Vertica's Unified Analytics Warehouse allows you to combine data siloes that are growing exponentially—without moving the data.

- **Save on both storage and computational charges:** While cloud-based data storage is low cost, analyzing that data can lead to prohibitively expensive compute charges. Vertica in Eon Mode manages dynamic workloads, so you can spin up storage and compute resources as you need them, and spin then down afterward to eliminate unnecessary costs.

- **Meet business expectations:** Users don't want to wait for results. Vertica provides the scalability to meet service level agreements (SLAs) and business needs with the best TCO and fastest ROI, including the ability to dedicate compute resources to individual use cases without replicating the data.

- **Embrace popular tools:** Vertica provides robust and powerful SQL and is certified to work with all of your tools—not just those from your primary vendor or limited to a single infrastructure. Use the extract, transform, load (ETL) tools or SQL-based visualizations of your choosing.

Key Features

At the core of the Vertica Advanced Analytics Platform is a column-oriented, relational database built specifically to handle today's analytic workloads. This powerful analytics platform provides you with:

- Complete and advanced SQL-based analytical functions to provide powerful SQL analytics.

- A clustered approach to storing big data, offering superior query and analytic performance.

- Better compression, requiring less hardware and storage than comparable data analytics solutions.

- Flexibility and scalability to easily ramp up when workloads increase.

- Better load throughput and concurrency with querying.

- In-database machine learning addressing every step in the ML process—algorithms, R support, Python extensibility, and more.

- Analyze data in place and in any format, including complex data types like Maps and Arrays, Structs in Parquet on S3, HDFS, open SQL-based analytics, and new use cases.

- Vertica in Eon Mode provisions dynamic workloads as needed, separates storage and compute, and enables workload

Built-in Functions

Geospatial	Machine Learning
Event Series	Text Analytics
Time Series	Pattern Matching
Real-Time	Regression
Statistics	

Data Visualization

Logi Analytics	ODBC
Looker	JDBC
Power BI	OLEDB
Qlik	
Tableau	

User Defined Functions

R	Python
Java	SQL
C/C++	

Security Integrations

LDAP	FIPS
Kerberos	Voltage

Analyze External Tables in the Right Place

Parquet	HDFS
Amazon S3	Apache ORC

On-Premises Deployment Options

Openstack	Commodity Hardware
MINIO	Pure Storage
Hadoop	

Cloud Deployment Options

Microsoft Azure
Amazon Web Services
Google Cloud Platform

User Defined Loads

Data Transformation—(Spark)
Messaging—(Kafka)
ETL—(Attunity, Informatica)

Figure 1. Vertica's open architecture and rich ecosystem

isolation to serve multiple departments without duplicating the data. Vertica in Enterprise Mode is ideal for stable workloads and regular queries.

- Runs in the clouds, including Google Cloud Platform (GCP), Azure, AWS, VMware clouds; and runs on-premises with commodity hardware and support for a range of object stores, such as Apache Hadoop HDFS for communal storage, MinIO, and on Pure Storage FlashBlade S3.

Figure 2. Vertica built-in machine learning process flow

Product Overview

Vertica provides blazingly fast speed (queries run 10–50X faster), exabyte scale (store 10–30X more data per server), and broad ecosystem integration (use any business intelligence tools, ETL tools, storage, etc.) at a much lower cost than traditional data warehouses or cloud-only data warehouses.

The Power to Handle Today's Massive Data Volumes

Modern businesses must manage more data sources than ever before—no longer just CRM and ERP, systems, but also IoT sensors, social media data, Web logs and data streams, gas and electrical grids, and mobile networks, just to name a few. Organizations that are truly data-driven must manage this explosive data growth, and discover the patterns and trends that can lead to new business opportunities, as well as repeat business from their customers.

Vertica answers these needs. It handles data at exabyte scale, and enables your organization to unify data siloes across multiple cloud and hybrid (cloud and on-premises) environments. Not only can Vertica manage massive data volumes, it keeps you from getting locked into a single cloud vendor. Use the tools of your choice, and take full advantage of the underlying infrastructure you already have in place, with portability across multi-cloud, on-prem, and Hadoop data lakes.
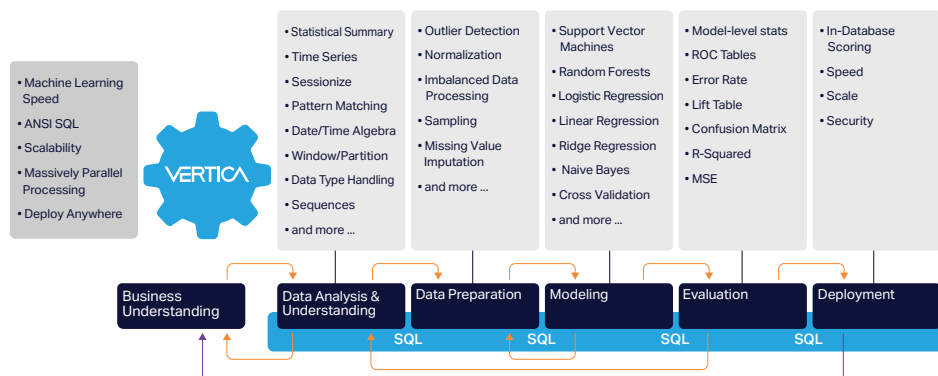
Operationalize Machine Learning— at Scale

Not long ago, data science was limited by the inability to base models on full data volumes, which led to inaccurate predictions. To make matters worse, the majority of machine learning initiatives never make their way into production at all, so only portions of the organizations benefit from the work from data scientists.

With Vertica, you can finally operationalize machine learning, so that you can understand—and act on—what that data is telling you, with the speed and scalability to make a difference. Vertica's in-database machine learning supports the entire predictive analytics process with massively parallel processing and a familiar SQL interface, allowing data scientists and analysts to build their models using their preferred tools and languages to embrace the power of big data and accelerate business outcomes with no limits and no compromises. More details below.

The Technology Big Data Demands

Vertica is built from the ground up to tackle the challenges of big data analytics. Its massively parallel processing system addresses the most demanding analytics use cases in the industry. Its columnar store offers aggressive data compression, so it delivers extremely fast results, reducing query times from hours to minutes, or minutes to seconds—something outdated row-store technologies cannot achieve.

Vertica offers advanced SQL-based analytics—from graph analysis, to triangle counting, to Monte Carlo simulations, to time series and geospatial, and more. All this can be applied to your "hot" data loaded directly into Vertica for the most demanding use cases.

You also get choices. Vertica is the only advanced analytics platform that can analyze data in HDFS, in S3 Object Storage, and within the Vertica data warehouse itself, including the ability to join these disparate data sets into unified analytics. Perhaps most important, Vertica offers the broadest choice of deployment modes—Vertica in Eon Mode for dynamic workloads that benefit from a separation of compute and storage architecture, and Vertica in Enterprise Mode for more predictable workloads on servers with tightly coupled storage—so you can choose the architecture that works for you today.

Unifying Today's Big Data Siloes

Vertica ensures that all the time, money, and effort you've put into storing your data turns

into business value. It provides a unified analytics platform that can analyze data where it resides—HDFS or Cloud Object Storage—and in all popular formats—ORC, Parquet, JSON, or ROS (native Vertica).

Along with eliminating data center maintenance, public clouds have provided the architectural advantage of separating data compute and storage, and provisioning on-demand. But high compute charges from any given cloud vendor can quickly sink your big data budget. Vertica solves this problem when deployed in Eon Mode. Analytics can be performed on data moved to cache, and you can add as many clusters as you need for the task. Then resize those clusters, all the way to full hibernation if you wish, when the work is complete.

Built-in Machine Learning Brings Data Science Projects into Production

Featuring in-database machine learning, Vertica allows teams to become productive with advanced analytics right out of the box. More experienced teams can import models built and trained in other platforms and languages—like TensorFlow, Spark, Python, and SPSS—via the PMML (Predictive Model Markup Language) format. With PMML model export, models created in Vertica can also be exported for scoring in other systems, such as edge nodes for IoT use cases. Data science and data engineering teams can finally operationalize machine learning, while using their tools of choice to fully leverage their data at scale.

Vertica offers more built-in analytics features than any other big data analytics platform:

- End-to-end machine learning management—from data prep to deployment. No need to download and install separate packages.
- Prepare data with functions for normalization, outlier detection, sampling,

imbalanced data processing, missing value imputation and more.

- Create, train, and test advanced machine learning models on massive data sets.
- Evaluate model-level statistics including ROC tables and confusion matrices.
- Revert back to previous model iterations using model management and version control features.
- Massively Parallel Processing (MPP) architecture allows you to build and deploy models at petabyte-scale with extreme speed and performance.
- Simple SQL execution democratizes predictive analytics with user-friendly, SQL-based machine learning functions.
- Support for familiar programming languages lets you develop user-defined extensions (UDx) with C++, Java, Python, or R.
- Built-in machine learning algorithms support classification, clustering, and predictive applications. These include linear regression, logistics regression, k-means, naïve Bayes, support vector machines, and random forest.
- Machine learning functions are included for data prep, evaluation, prediction, and supervised and unsupervised learning.

Open to the Widest Array of ETL, BI, and Visualization Tools

Every release of Vertica is certified and tested with visualization and ETL tools. It supports popular SQL, and Java Database Connectivity (JDBC)/Open Database Connectivity (ODBC). This enables users to preserve years of investment and training in these technologies because all popular SQL programming tools and languages work seamlessly. Leading BI and visualization tools are tightly integrated, such as Tableau and MicroStrategy, as well as popular ETL tools like Informatica, Talend, Pentaho, and more.
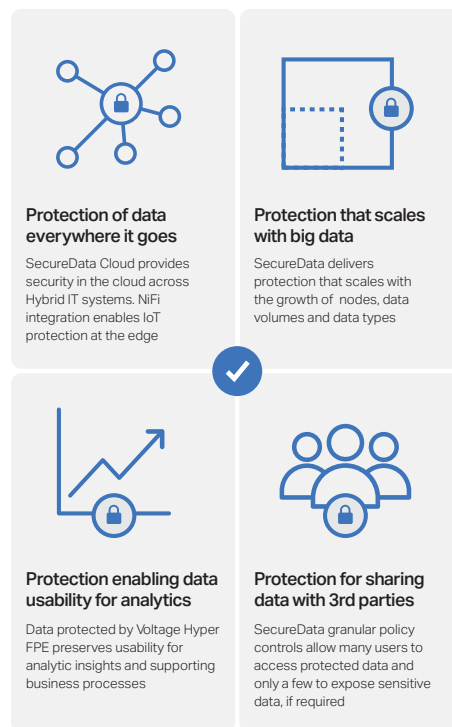

Protection of data everywhere it goes

SecureData Cloud provides security in the cloud across Hybrid IT systems. NiFi integration enables IoT protection at the edge

Protection that scales with big data

SecureData delivers protection that scales with the growth of nodes, data volumes and data types

Protection enabling data usability for analytics

Data protected by Voltage Hyper FPE preserves usability for analytic insights and supporting business processes

Protection for sharing data with 3rd parties

SecureData granular policy controls allow many users to access protected data and only a few to expose sensitive data, if required

Figure 3. Vertica integration with the Voltage product for additional enhanced security

Securing Critical Data

Vertica provides end-to-end security with support for industry-standard protocols and partner solutions such as LDAP, Kerberos, TLS, FIPS 140-2, AWS IAM, and Apache Sentry. Vertica uses a layered security model and provides multiple security authentication authorization mechanisms.

- Uses TLS to establish a secure connection between client machine and server.
- Authentication and access controlled by passwords stored with SHA, LDAP, Kerberos, and SSL certificates.
- Security model built on ANSI standard Role Based Access Control—privileges assigned to roles.

- Access policy mechanism allows fine-grained access control to row and column data, including column masking.
- Logging mechanisms ensure operations audit trail, natively exported to other security domains for analysis and persistence.
- Simplified process for TLS certificate administration, with user authentication and permissions management with LDAP Link, Kerberos for vertica-python and new permissions system tables.
- Support for format preserving encryption via high-performance Voltage SecureData integration.

Vertica's integration with Voltage format-preserving encryption technology ensures end-to-end data protection, securing data in use, in motion, and at rest. Voltage SecureData adds a data-centric layer to Vertica's layered security model, which facilitates compliance with new and emerging data privacy regulations such as CCPA and GDPR. Voltage encryption reduces risk of breach by securing sensitive data while preserving the data's length and character set. Encryption is used to protect structured data such as tax ID, name, address, GPS location, IP address, date of birth, and salary—all typical personal and protected health information types. Decryption of protected data is controlled by policy, including requiring LDAP authentication and authorization if needed. Search can be performed on ciphertext, and

Voltage maintains referential integrity across distributed data sets. With Voltage inside, data can be safely moved into and out of Vertica, for secure access and analytics.

One Vertica, Two Deployment Modes
Whether your workloads are predictable and used for scheduled reporting and analytics, or variable—corresponding to seasonal changes, for example, or designed at varying scale by data science teams—your license supports Vertica in Enterprise Mode, or Eon Mode.

Vertica Database Designer— Moving Toward Full Automation
Offering a newly enhanced user interface for greater ease of use, Vertica Database Designer improves operations and reduces resource consumption, while improving your team's projection designs for consistently faster queries. It analyzes your logical schema, sample data, and, optionally, your sample queries. Vertical Database Designer creates a physical schema design (a set of projections) that can be deployed automatically or manually, and it can be used by anyone, even users without specialized database knowledge.

Try the Vertica Community Edition, for Free
Your big data needs are unique, and always changing. Your analytics data warehouse

Contact us at:
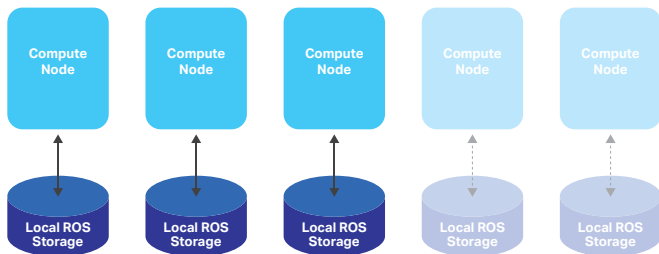**www.vertica.com**

Like what you read? Share it.

should be flexible enough to meet your demands today, while helping you see into your business's future. Try Vertica and make your concept a reality.

Evaluate Vertica today at:
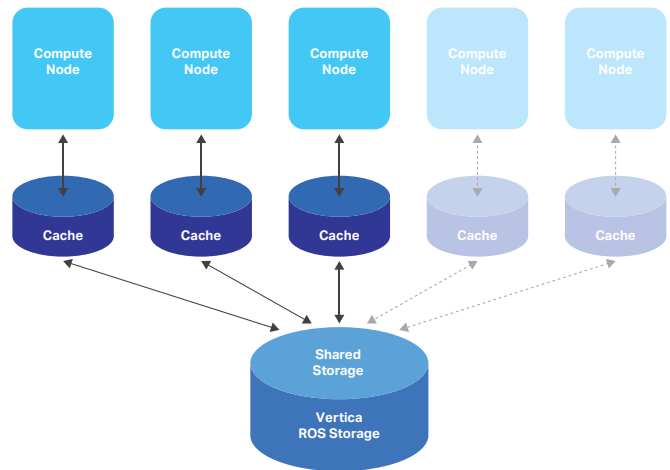**www.vertica.com/try**

Learn more on our website:
www.vertica.com

Find free training and certification courses on Vertica Academy:
https://academy.vertica.com

Vertica in Enterprise Mode

Vertica in Enterprise Mode runs on industry-standard servers with tightly coupled storage, delivering the highest performance for consistent compute capacity.

Vertica in Eon Mode

Vertica in Eon Mode features a cloud-native architecture that separates compute from storage, enabling compute resources to share storage for different purposes.