

Vertica for SQL on Apache Hadoop

Vertica SQL on Apache Hadoop offers the fastest and most enterprise-ready way to perform SQL queries on your Hadoop data. We've leveraged our years of experience in the big data analytics marketplace and now offer the same technology that powers the Vertica database to command a query engine for data stored in HDFS. Users can perform analytics regardless of the format of data or Hadoop distribution used.



Product Highlights

Vertica SQL on Apache Hadoop handles your mission-critical analytics projects by merging the best of our analytics platform with the best that Hadoop data analytics can offer. The principles below help us to deliver on these promises:

- **Data lake or daily analytics.** The SQL engine supports data discovery on your Hadoop data lake as well as highly optimized analytics for even the most demanding SLAs.
- **Unified analytics engine.** The engine is flexible enough to perform analytics on data no matter where it lives—Hadoop, native Vertica, or in the cloud.
- **Complete SQL support.** Get full ANSI SQL 99 compliance that is able to execute 100 percent of the TPC-DS benchmarks without modification.
- **Fast ORC and Parquet file readers.** Vertica can quickly and efficiently query ORC and Parquet files for fast Hadoop data analytics without moving the data. Other formats like AVRO are also supported.
- **Workload management.** Convenient, graphical application supports Ambari to check the health of both the Vertica and Hadoop clusters and their queries. It also supports storage labels for resource allocation in YARN.

Benefits

- Takes advantage of a highly optimized, enterprise-ready SQL engine for Hadoop, complying with ANSI SQL 99.
- Leverages existing SQL skillsets, BI tools, and Hadoop deployments.
- Supports your current visualization tools like Tableau, Looker, QlikView, IBM Cognos, and MicroStrategy with certified connectors.
- Completes 100 percent of the TPC-DS benchmark queries with no modification, while others can complete less than 80% of the benchmarks.
- Advanced analytics with features like Geospatial, Monte Carlo, time-series analysis, and aggregate projections.
- Supports in-database machine learning like logistic/linear regression, k-means, random forest and more.
- Installs directly on Hadoop cluster with no single point of failure or helper node.
- Supports common Hadoop file formats like ORC and Parquet with fast, efficient code.

Real Data Analytics, Not a Science Experiment

Vertica SQL for Apache Hadoop is great for both data discovery and high-performance analytics. It supports the complete information lifecycle, from data capture to cold storage.

- Users can explore live data on their own as it arrives on their Hadoop cluster instead of spending weeks or months on data preparation, modeling and ETL, and subsequent schema management.
- Once explored, Vertica for SQL on Apache Hadoop takes full advantage of the Hadoop Big Data analytics cluster.
- Vertica offers enterprise-ready, advanced analytics that can take you from hindsight to insight to foresight with analytical features like Monte Carlo, time-series analysis, and aggregate projections.
- Analytics are easy to create through a complete and mature SQL for Hadoop engine that is fully certified for many industry-standard visualization tools.
- When you need to boost performance even further, materialize data to Vertica Enterprise edition and take full advantage of our optimizations like compression, columnar storage, and projections.
- Vertica SQL on Apache Hadoop supports use cases for those who want to query data that sits in one or more distributions of Hadoop, Vertica, or both.

Unified Platform for Data Analytics

Our SQL on Apache Hadoop engine is the same new-age data analytics platform designed from the ground up for business analytics at the scale of big data. Whether you deploy on Hadoop, on premise or in the clouds, the platform provides an ideal environment for in-database analytics functions, as well as tight integration with R and other advanced analytics libraries.

The wide-ranging analytical functions in Vertica include standard SQL-99 conventions, value-added analytics with SQL, user-defined extensions, and big data advanced analytics using custom logic. In addition to leveraging the platform's built-in capabilities, you can develop your



own next-generation analytics functions using the platform's C++, Java, and R SDKs. Vertica seamlessly connects with your favorite visualization tools so that you can create dashboards that show metrics on billions of rows of data.

Ultimately, Vertica enables your organization make better-informed decisions, compete more effectively, and gain a real return on information. These ideas are at the heart of the analytics-driven enterprise.

Learn More At
www.vertica.com

Several Technologies Are Part of Vertica SQL on Apache Hadoop

ORC and Parquet file readers

Allows users to directly query ORC and Parquet files and take advantage of their SQL for Hadoop optimizations

Parquet writer

Store analysis in Parquet format

Connector for HCatalog

Allows users to query data stored in Hive using the Vertica native SQL syntax

Connector for HDFS

Allows users to load structured data from the Hadoop Distributed File System (HDFS) and create an external table based on structured data stored in HDFS

Storage location for HDFS

Allows users to store Vertica-formatted data on HDFS

www.vertica.com



Vertica Headquarters

150 Cambridgepark Drive
Cambridge, MA 02140

Learn more at: www.vertica.com