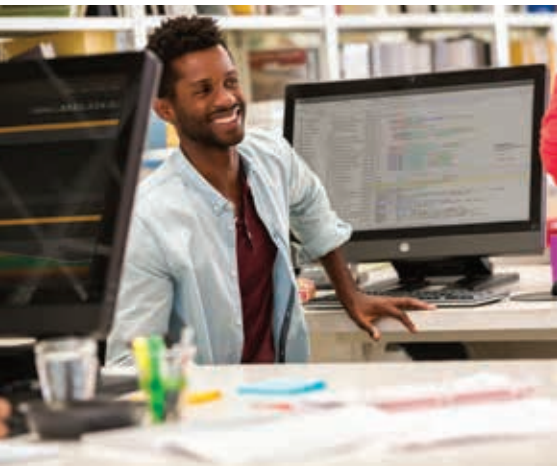
Hewlett Packard Enterprise

# Vertica for SQL on Apache Hadoop

Advanced ANSI SQL analytics for your enterprise



**Benefits**

- Takes advantage of a highly optimized, enterprise-ready SQL engine for Hadoop.
- Leverages existing SQL skillsets, BI tools, and Hadoop deployments.
- Complies with ANSI SQL 99.
- 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.
- Uses advanced analytics with features like Monte Carlo, time-series analysis, and aggregate projections.
- Installs directly on Hadoop cluster with no single point of failure or helper node.

Works with Cloudera Distribution including Apache Hadoop (CDH) 4.6, 5.0, and 5.0.1; Hortonworks Data Platform (HDP) 2.1; MapR 3.1.1 and 3.0.3. Please check online documentation for a detailed description and updates to compatibility information.

## HPE Vertica SQL on Apache Hadoop

HPE 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 opened up our platform to use the full power of the Hadoop cluster. Users can perform analytics regardless of the format of data or Hadoop distribution used.

HPE 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 file reader.** Vertica can quickly and efficiently query ORC files for fast Hadoop data analytics without moving the data. Other formats like Parquet and 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.

- **Secure.** Platform supports Kerberos to provide security and encryption with a single authentication to both Vertica and Hadoop services.
- **No single point of failure.** Unlike other SQL on Hadoop analytics platform solutions, Vertica installs directly on your Hadoop nodes with no need for a helper node.
- **Platform-agnostic.** Hewlett Packard Enterprise is focused on the deep integration of our Hadoop SQL engine with any distribution of Hadoop. Hewlett Packard Enterprise has partnered with Hortonworks, Cloudera, MapR, and others to bring you a flexible Hadoop data analytics platform that performs seamlessly across major distributions.
- **Open platform.** Hewlett Packard Enterprise offers open APIs and developer tools on our community site where users may access our ecosystem of SQL on Hadoop platform enhancements.

With HPE Vertica SQL on Apache Hadoop, businesses can diminish switching costs and learning curves for users with the familiar ANSI SQL syntax. Analysts can continue to use familiar BI/analytics tools that assume and auto-generate ANSI SQL code to interact with any Hadoop distribution. Users can interact with data via Hive and HCatalog, MapReduce, or directly via HDFS. There is even a feature to leverage ORC and Parquet files without having to use the data.

## Real data analytics, not a science experiment

HPE Vertica SQL on 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, HPE Vertica SQL on Apache Hadoop takes full advantage of the Hadoop Big Data analytics cluster and works with YARN to optimize cluster resources.
- HPE 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 HPE Vertica Enterprise edition and take full advantage of our optimizations like compression, columnar storage, and projections.
- HPE Vertica SQL on Apache Hadoop supports use cases for those who want to query data that sits in one or more distributions of Hadoop, HPE Vertica, or both.

## How it fits into your architecture

Aside from using HPE Vertica SQL on Apache Hadoop as your primary query engine, there are many other ways to leverage the benefits of HPE Vertica SQL on Apache Hadoop. Existing HPE Vertica implementations can set up Hadoop storage clusters nodes to move data from existing data lakes into these HPE Vertica Hadoop locations for higher query speeds, or tier down data from highly optimized Vertica storage clusters to open capacity for new critical hot data.

New implementations may also choose to use Hadoop HDFS as the primary storage layer to run HPE Vertica and leverage existing skilled Hadoop staff and infrastructures. Separate Hadoop clusters can then be created as additional storage locations for cost-optimized data placements. Vertica is extremely good at resource management and works well in a YARN environment.

**Table 1.** Enabling technologies

| **Several technologies are part of HPE vertica SQL on apache Hadoop** | |
|---|---|
| **ORC file reader** | Allows users to directly query ORC files and take advantage of their SQL for Hadoop optimizations |
| **Connector for HCatalog** | Allows users to query data stored in Hive using the HPE 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 |
| **Connector for Apache Hadoop MapReduce** | Allows users to create a MapReduce job or Pig script that can read data in Hadoop |

Learn more at
**vertica.com/SQLonHadoop**

**Sign up for updates**



4AA5-7834ENW, August 2016, Rev. 3